主管单位：中华人民共和国工业和信息化部
主办单位：北京航空航天大学

ISSN 1001-5965
CN 11-2625/V

北京航空航天大學学报

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

高超声速再入双锥飞行器横向喷流控制效率数值研究

雷娟棉　朱沛宇

1 2026
Vol.52

北京航空航天大学学报

第 52 卷　第 1 期　（总第 395 期）　2026 年 1 月

目　　次

高超声速再入双锥飞行器横向喷流控制效率数值研究 …… 雷娟棉，朱沛宇（1）

一种融合 MSC 和时空双重注意力的 TCN 航迹预测方法 …… 焦卫东，杨蓓（15）

多模态特征交互与语义引导融合的 RGB-T 人群计数 …… 陈永，张娇娇，董珂（28）

基于改进 YOLOv8 的飞机蒙皮缺陷检测算法 …… 章东平，王杼涛，夏岳键，徐云超，林丽莉（38）

星凸形不规则形状多扩展目标 PMBM 滤波器 …… 陈辉，刘孟波，连峰，韩崇昭（49）

基于图注意力堆叠自编码器微生物-药物关联预测 ……
…… 王波，何洋，杜晓昕，张剑飞，徐靖然，贾娜（61）

基于多通道耦合的时空增强异常行为检测 …… 章东平，潘鑫，马道滨，米红妹，林丽莉（73）

基于邻居边界度的自主决策系统性能边界识别方法 …… 路辉，吕静茹，王诗琪，孙泽斌（80）

面向多无人机物流配送的双层任务规划方法 …… 王飞，杨清平（94）

一种测扭齿的接触式角度测量方法 …… 童晓艳，鲁东生，柏慧，常龙（104）

基于应变能理论的全陶瓷球轴承细观缺陷识别 …… 白晓天，张赵南，石怀涛，张啸尘（110）

火箭倒飞逆向喷流干扰特性数值模拟研究 … 张培红，贾洪印，赵娇，吴晓军，周桂宇，张耀冰（120）

广义未知扰动下多模型最小上限滤波 …… 秦月梅，张荣华，杨衍波，潘泉（129）

地基 cGNSS-R 双天线干涉测高试验研究 …… 洪学宝，李杰，邢进，杨鹏瑜，杨东凯（140）

空地网联集群协同模式识别方法 …… 曲桂娴，周建山，司杨，刘晓静，袁奇雨，马清琳（147）

面向石油物探领域的 HPC 故障与性能分析 … 王明键，赵长海，李超，文佳敏，尚民强，侯红军（157）

基于叠加流场面元法的民用飞机跨声速气动伺服弹性分析 … 吴善强，陈琦，孙亚军，尼早，陈文（167）

智能制造数字孪生模型可变模糊综合评估 …… 王向章，于莉莉（180）

高空大气风切变时空分布特征 …… 杨钧烽，程旋，王建美，张依鸣，胡雄，肖存英（192）

基于 INT-VSMM 算法的目标航迹跟踪和外推 …… 任宣铭，汤新民，刘雨生，鲁其兴（203）

液体火箭注气式蓄压器流动阻尼理论模型 …… 史童，何允钦，梁国柱，潘辉，朱平平（214）

基于不匹配结点对接子结构法的超声速颤振预测 …… 刘晨宇，谢长川（223）

基于深度学习的双域信息 CT 金属伪影抑制方法 ……
…… 海潮，田鑫，张宏，谭大龙，何一新，孟凡勇，杨民（232）

基于旋转单天线的欺骗干扰检测方法 …… 陈世森，倪淑燕，王绍山，王海宁，吕金飞（244）

无对齐实体场景的多语言知识图谱补全 …… 唐榕氚，徐秋程，汤闻易，翟飞飞，周玉（252）
航空发动机盲腔螺母精密拧紧技术 …… 张宗江，李小强，韩玉杰，李兆宇，张铭铭（260）
考虑空间碎片质量的多碎片主动清理任务规划方法 …… 郭甲，张海博，庞兆君，杜忠华（275）
噪声环境下告警冗余设计对装甲车乘员响应绩效的影响 ……
…… 周孙夏，刘双，完颜笑如，冯传宴，王一行，解芳（286）
领域知识内嵌的电磁信号调制方式智能识别 …… 赵宏佳，张多纳，鲁远耀，丁文锐（294）
双磁路旋转洛伦兹力磁轴承设计与分析 …… 庞维坤，王卫杰，樊亚洪，李磊，杨洋，朱宏业（306）
基于 OTFS-LFM 的水下可见光通信感知一体化波形设计 …… 杨新，朵琳，任勇，许渤雨（317）
基于改进的组合赋权和云模型的平流层飞艇健康状态评估 ……
…… 曾盛富，梁浩全，王岩，孙康文，钟梓鸿（326）
有控自旋卷弧翼弹箭角运动的非线性吸引域 …… 周文涛，常思江，展鹏遥，崔慧振（338）
充液航天器固定时间姿态滑模容错控制 …… 范志文，宋晓娟，吕书锋，岳宝增（352）
基于快速傅里叶卷积的电成像测井图像修复 ……
…… 苏乾潇，乔德新，任义丽，冯周，林盛斓，黄睿琦（362）
不同来流下双楔定常/非定常流动的模拟分析与流态预测 …… 张鑫泽，李沁，翁谊辉，尤延铖（371）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 384 * zh * P * ￥50.00 * 400 * 36 * 2026-01

（**编　辑**　张　嵘　孙　芳　万玉敏　卞欢欢　王　茜　李亚泰）

JOURNAL OF BEIJING UNIVERSITY OF
AERONAUTICS AND ASTRONAUTICS

Vol. 52 No. 1 （Sum 395） January 2026

CONTENTS

Numerical study on lateral jet control efficiency of a hypersonic re-entry double-cone vehicle
…… LEI Juanmian，ZHU Peiyu （1）

A TCN trajectory prediction method fusing MSC and spatio-temporal dual attention
…… JIAO Weidong，YANG Bei （15）

Multimodal feature interaction and semantic guided fusion for RGB-T population counting
…… CHEN Yong，ZHANG Jiaojiao，DONG Ke （28）

Aircraft skin defect detection algorithm based on enhanced YOLOv8
…… ZHANG Dongping，WANG Zhutao，XIA Yuejian，XU Yunchao，LIN Lili （38）

Star convex irregular shape multi-extended target PMBM filter
…… CHEN Hui，LIU Mengbo，LIAN Feng，HAN Chongzhao （49）

Prediction of microbe-drug association based on graph attention stacked autoencoder
…… WANG Bo，HE Yang，DU Xiaoxin，ZHANG Jianfei，XU Jingran，JIA Na （61）

Spatial-temporal enhanced abnormal behavior detection based on multi-channel coupling
…… ZHANG Dongping，PAN Xin，MA Daobin，MI Hongmei，LIN Lili （73）

Performance boundary identification method for autonomous decision-making systems based on neighbor boundary degree
…… LU Hui，LYU Jingru，WANG Shiqi，SUN Zebin （80）

Two-layer task planning method for multi-UAV logistics distribution
…… WANG Fei，YANG Qingping （94）

A contact angle measurement method for torsional teeth
…… TONG Xiaoyan，LU Dongsheng，BAI Hui，CHANG Long （104）

Identification of mesoscopic fault of full ceramic ball bearings based on strain energy theory
…… BAI Xiaotian，ZHANG Zhaonan，SHI Huaitao，ZHANG Xiaochen （110）

Numerical simulation research on opposing jet interaction characteristics of rocket inverse flight
…… ZHANG Peihong，JIA Hongyin，ZHAO Jiao，WU Xiaojun，ZHOU Guiyu，ZHANG Yaobing （120）

Multiple model minimum upper bound filter under generalized unknown disturbances
…… QIN Yuemei，ZHANG Ronghua，YANG Yanbo，PAN Quan （129）

Experimental study on interferometric altimetry using a ground-based dual-antenna cGNSS-R
…… HONG Xuebao，LI Jie，XING Jin，YANG Pengyu，YANG Dongkai （140）

Cooperative pattern recognition method for air-ground networked swarms
…… QU Guixian，ZHOU Jianshan，SI Yang，LIU Xiaojing，YUAN Qiyu，MA Qinglin （147）

HPC fault and performance analysis in field of petroleum geophysical exploration
…… WANG Mingjian，ZHAO Changhai，LI Chao，WEN Jiamin，SHANG Minqiang，HOU Hongjun （157）

Transonic aeroservoelastic analysis of civil aircraft based on overset field-panel method
…… WU Shanqiang，CHEN Qi，SUN Yajun，NI Zao，CHEN Wen （167）

Variable fuzzy comprehensive evaluation for intelligent manufacturing digital twin model
…… WANG Xiangzhang，YU Lili （180）

Spatiotemporal distribution characteristics of high-altitude wind shear
…… YANG Junfeng，CHENG Xuan，WANG Jianmei，ZHANG Yiming，HU Xiong，XIAO Cunying （192）

Target trajectory tracking and extrapolation based on INT-VSMM algorithm
…… REN Xuanming，TANG Xinmin，LIU Yusheng，LU Qixing （203）

Theoretic model of flow resistance for gas-filled accumulators in liquid rockets
…… SHI Tong，HE Yunqin，LIANG Guozhu，PAN Hui，ZHU Pingping （214）

Supersonic flutter prediction based on unmatched substructure method
…… LIU Chenyu，XIE Changchuan （223）

A deep learning-based dual-domain information method for CT metal artifact reduction
…… HAI Chao，TIAN Xin，ZHANG Hong，TAN Dalong，HE Yixin，MENG Fanyong，YANG Min （232）

Spoofing detection method based on single rotating antenna
…… CHEN Shimiao，NI Shuyan，WANG Shaoshan，WANG Haining，LYU Jinfei （244）

Multilingual knowledge graph completion without aligned entity pairs
…… TANG Rongchuan，XU Qiucheng，TANG Wenyi，ZHAI Feifei，ZHOU Yu （252）
Precision tightening technology for aeroengine blind cavity nuts
…… ZHANG Zongjiang，LI Xiaoqiang，HAN Yujie，LI Zhaoyu，ZHANG Mingming （260）
Planning method for a multi-debris active removal mission considering space debris mass
…… GUO Jia，ZHANG Haibo，PANG Zhaojun，DU Zhonghua （275）
Influences of multimodal redundant alarm design on reaction performance of armored vehicle crew under noise environment
…… ZHOU Sunxia，LIU Shuang，WANYAN Xiaoru，FENG Chuanyan，WANG Yihang，XIE Fang （286）
Intelligent recognition of electromagnetic signal modulation with embedded domain knowledge
…… ZHAO Hongjia，ZHANG Duona，LU Yuanyao，DING Wenrui （294）
Design and analysis of double magnetic circuit rotating Lorentz force magnetic bearing
…… PANG Weikun，WANG Weijie，FAN Yahong，LI Lei，YANG Yang，ZHU Hongye （306）
Underwater visible light communication sensing integrated waveform design based on OTFS-LFM
…… YANG Xin，DUO Lin，REN Yong，XU Boyu （317）
Health status assessment of stratospheric airship based on improved combination weighting and cloud model
…… ZENG Shengfu，LIANG Haoquan，WANG Yan，SUN Kangwen，ZHONG Zihong （326）
Nonlinear region of attraction of angular motion for a controlled projectile with self-rotating wraparound fins
…… ZHOU Wentao，CHANG Sijiang，ZHAN Pengyao，CUI Huizhen （338）
Fixed-time attitude sliding mode fault-tolerant control for liquid-filled spacecraft
…… FAN Zhiwen，SONG Xiaojuan，LYU Shufeng，YUE Baozeng （352）
Inpainting of electrical imaging logging images based on fast Fourier convolution
…… SU Qianxiao，QIAO Dexin，REN Yili，FENG Zhou，LIN Shenglan，HUANG Ruiqi （362）
Numerical analysis and flow state prediction of double wedge steady/unsteady flow at different in flow
…… ZHANG Xinze，LI Qin，WENG Yihui，YOU Yancheng （371）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0714

高超声速再入双锥飞行器横向喷流控制效率数值研究

雷娟棉*，朱沛宇
(北京理工大学 宇航学院，北京 100081)

摘 要：横向喷流是飞行器进行姿轨控的有效手段，为研究高超声速再入双锥飞行器横向喷流控制效率随飞行条件和飞行姿态角的变化规律，基于三维雷诺平均Navier-Stokes (RANS)方程对再入双锥飞行器横向喷流绕流场进行了数值模拟研究，分析了飞行马赫数、飞行高度、迎角和侧滑角对横向喷流控制效率的影响，并揭示了相应的流动机理。结果表明：在进行横向喷流控制时，飞行马赫数增大或飞行高度降低可显著提升喷流法向力和俯仰力矩的控制效率，随着飞行高度的增加，控制效率均逐渐接近1；迎角为−20°与−30°时，来流与喷流干扰使喷流气体大面积作用在飞行器表面上，导致产生与喷流推力方向相反的吸力，使喷流法向力控制效率降为负值，但会使喷流俯仰力矩的控制效率提高；随着迎角向正转变，喷流法向力控制效率逐渐转为正值，并在10°时达到最佳，喷流俯仰力矩控制效率会向1接近；侧滑角的变化对横向喷流控制效率的影响较小。

关 键 词：横向喷流；再入双锥飞行器；控制效率；高超声速；数值模拟
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2026)01-0001-14

为提升生存性能，高超声速再入飞行器在进入大气层后应具有飞行轨道多变、急速拐弯与下压等大范围机动能力[1]。然而，由于高超声速再入飞行器的飞行速度和高度跨度大，气动特性变化剧烈，进行机动飞行面临诸多难题。在临近空间(海拔高度20～100 km)内，空气密度低、来流动压小，传统的气动控制舵面难以对高超声速再入飞行器进行机动控制，而此时横向喷流是有效可行的控制手段[2]。横向喷流控制是利用布置在飞行器表面的小型火箭发动机或其他射流装置产生直接力对飞行器的飞行姿态或质心运动轨迹进行控制的方式。相较于气动舵面控制，喷流控制受环境的影响较小，在来流动压较小的情况下也有较高的控制效率，并且由于响应时间短，可以更好地满足高超声速再入飞行器的快速机动要求[3]。然而，横向喷流会与高超声速来流发生强烈的气动干扰，横向喷流从启动到关闭的过程是一个具有流场结构剧烈变化的非定常流动过程，横向喷流产生的控制力和力矩不仅取决于发动机产生的净推力大小，还会受到气动干扰产生的附加气动力和力矩的影响[4]，从而改变喷流产生的力或力矩的控制效率。

早期，研究人员主要通过风洞试验对横向喷流进行研究。Kumar等[5]通过风洞试验发现喷流压比、马赫数及迎角等均会对高超声速细长钝锥体横向喷流分离范围造成显著影响。Brandeis和Gill[6-8]对尖拱形头部导弹在超声速和高超声速条件下的横向喷流干扰流场进行了一系列风洞试验研究，结果表明，随着迎角的增大，横向喷流控制效率会得到显著提升。Kurita和Inoue等[9]在高超声速激波风洞中进行的10°半角钝锥大迎角下横向喷流实验结论与Brandeis等[6-8]的实验结果有所差异，结果表明，虽然喷口上游干扰高压区范围随着迎角增大而

收稿日期：2023-11-02；录用日期：2023-12-11；网络出版时间：2024-01-03 17:38
网络出版地址：link.cnki.net/urlid/11.2625.V.20240102.1552.003
*通信作者. E-mail：leijm@bit.edu.cn

引用格式：雷娟棉，朱沛宇. 高超声速再入双锥飞行器横向喷流控制效率数值研究[J]. 北京航空航天大学学报，2026，52(1)：1-14.
LEI J M，ZHU P Y. Numerical study on lateral jet control efficiency of a hypersonic re-entry double-cone vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：1-14 (in Chinese).

增大，但由于背风区整体压力降低，从而导致喷流控制效率下降。王志坚和伍贻兆等[10]在高超声速条件下进行了冷、热喷流干扰影响对比风洞试验研究，结果表明，冷喷与热喷造成的喷流干扰效果相似。以上研究获得了有价值的结论，但风洞试验成本高，受实验技术的限制，可模拟的条件少，且难以进行流动机理研究。

随着计算流体力学 (computational fluid dynamics, CFD) 和计算机技术的发展，对横向喷流的研究逐渐由实验转为数值模拟，并且都取得了较为丰富的结果。Gnemmi 等[11]针对旋成体导弹分别进行了横向喷流干扰特性的风洞试验和数值模拟，通过结果对比验证了数值模拟在雷诺平均 Navier-Stokes (Reynolds-averaged Navier-Stokes, RANS)方法中使用 SST k-ω 湍流模型的有效性。赵法明等[12]通过 RANS 方法进行数值模拟，发现舵面使旋成体弹身横向喷流的缠绕效应减小，对喷流干扰起有利作用。赵弘睿和龚安龙等[13]使用 RANS 方法对某钝头双锥旋成体外形进行了高空稀薄环境下高超声速真实燃气热喷数值模拟，结果表明，在高度较低时，飞行迎角对喷流干扰特性的影响较大，随着高度的增加，迎角的影响逐渐减小。梁杰和阎超等[14]使用直接模拟蒙特卡罗 (direct simulation Monte Carlo, DSMC)算法在高超声速条件下对三维平板稀薄大气环境下的横向喷流气动干扰特性进行了数值模拟研究，结果表明，随着来流稀薄程度的增加，推力和力矩放大系数随着飞行高度的增加而不断减小。Williams 和 Moeller 等[15]使用大涡模拟 (large eddy simulation, LES)方法对三维平板在超声速下的横向喷流进行了模拟计算，给出了流场中明显的大尺度涡系结构。相较于风洞试验，采用数值模拟方法进行横向喷流研究的条件受限较少，可在不同飞行条件下进行横向喷流与来流流场的模拟，并且能够给出干扰流场结构，可根据流场结构进行流动机理分析。目前，关于横向喷流的相关实验与数值研究已取得了大量的研究成果，但对于临近空间高超声速再入飞行器而言，在飞行条件或飞行姿态发生较大变化的情况下，相应的横向喷流控制效率变化规律和流动机理的系统性研究还较为缺乏。

本文针对临近空间高超声速再入双锥飞行器，使用基于三维 RANS 方程和 SST k-ω 湍流模型的数值方法对其横向喷流与来流的干扰流场进行模拟，研究飞行马赫数、飞行高度、迎角及侧滑角等对喷流法向力和俯仰力矩控制效率的影响规律，并对相应的流场结构特征进行分析，为进一步了解临近空间高超声速再入飞行器横向喷流的作用机制和控制方案设计提供参考。

1 数值模拟方法

1.1 流动控制方程

本文重点关注喷流干扰对飞行器气动力特性产生的影响，热化学非平衡效应主要对飞行器的气动热环境产生较大影响，而对气动力的影响较小[16]，因此，本文不考虑高超声速条件下的真实气体效应与热化学非平衡效应，在数值模拟计算时，来流与喷流气体均采用具有完全气体性质及满足连续介质假设的空气。采用有限体积法，基于 RANS 方程进行数值离散求解，其基本流动控制方程在笛卡儿坐标系下可写为

$$\frac{\partial}{\partial t}\iiint_V \boldsymbol{W}\mathrm{d}V+\iint_{\partial V}[\boldsymbol{F}-\boldsymbol{G}]\cdot\boldsymbol{n}\mathrm{d}S=\iiint_V \boldsymbol{Q}\mathrm{d}V \tag{1}$$

式中：$\boldsymbol{W}$ 为守恒变量；$\boldsymbol{F}$ 为无黏通量；$\boldsymbol{G}$ 为黏性通量；$\boldsymbol{Q}$ 为源项；V 为控制体积；$\boldsymbol{n}$ 为法向向量；S 为面积微元。

$$\boldsymbol{W}=\begin{bmatrix}\rho\\ \rho u\\ \rho v\\ \rho w\\ \rho E\end{bmatrix},\boldsymbol{F}=\begin{bmatrix}\rho\boldsymbol{U}\\ \rho u\boldsymbol{U}+p\boldsymbol{i}\\ \rho v\boldsymbol{U}+p\boldsymbol{j}\\ \rho w\boldsymbol{U}+p\boldsymbol{k}\\ \rho E\boldsymbol{U}+p\boldsymbol{U}\end{bmatrix},\boldsymbol{G}=\begin{bmatrix}0\\ \boldsymbol{\tau}_{xi}\\ \boldsymbol{\tau}_{yi}\\ \boldsymbol{\tau}_{zi}\\ \boldsymbol{\tau}_{ij}\boldsymbol{U}_j+\boldsymbol{q}\end{bmatrix} \tag{2}$$

式中：ρ、$\boldsymbol{U}$、E 和 p 分别为密度、速度、单位质量的总能和压强；$\boldsymbol{\tau}$ 为黏性应力张量；$\boldsymbol{q}$ 为热流通量；u、v、w 分别为 x、y、z 方向上的速度分量；$\boldsymbol{i}$、$\boldsymbol{j}$、$\boldsymbol{k}$ 分别为 x、y、z 方向上的单位向量。

单位质量总能 E 和总焓 h_0 的关系为

$$E=h_0-p/\rho \tag{3}$$

$$h_0=h+|\boldsymbol{U}|^2/2 \tag{4}$$

补充完全气体状态方程为

$$p=\rho RT \tag{5}$$

式中：h 为单位质量的焓值；空气气体常数 R=287 J/(kg·K)；T 为气体静温。

将式(1)中的控制方程进行离散后，采用全隐式格式进行求解。空间离散和时间离散均采用二阶迎风格式，对流项采用二阶 AUSM 格式离散，黏性项采用二阶中心差分格式离散。

1.2 湍流模型

Navier-Stokes 方程被时均化时会引入雷诺应力项，使得方程不封闭，需要引入湍流模型对方程进行封闭。选择适当的湍流模型能够比较准确地模拟横向喷流与来流相互干扰产生的一系列复杂流动。SST k-ω 两方程全湍流模型由 Menter 提出[17]，

该模型将 k-ε 模型在外场的自由流无关性与 k-ω 模型在近壁区域的准确性进行了有效结合, 能够精确捕捉到喷流干扰效应诱导的剪切流、非对称逆压梯度流动及非对称流动分离等现象, 被其他学者广泛应用于喷流干扰的数值计算中[11]。因此, 本文选取 SST k-ω 湍流模型进行数值模拟。

2　计算外形、网格与条件

2.1　计算外形与网格

本文以高超声速再入飞行器常用的双锥体外形[18]为对象展开横向喷流数值模拟研究, 其外形参数与文献 [18] 相同, 如图 1 所示。其中, 全弹长 L=3 000 mm, 头部单锥长 L_1=500 mm, 半锥角分别为 16°和 4°, 头部钝度 r=5 mm, 底部弹径 D=636 mm, 质心位于距头部顶点 x_m=1 978 mm 处(相对位置 x/L=0.66), 喷口位于距头部顶点 x=1 500 mm 处(相对位置 x/L=0.5), 喷口直径 ϕ=40 mm。

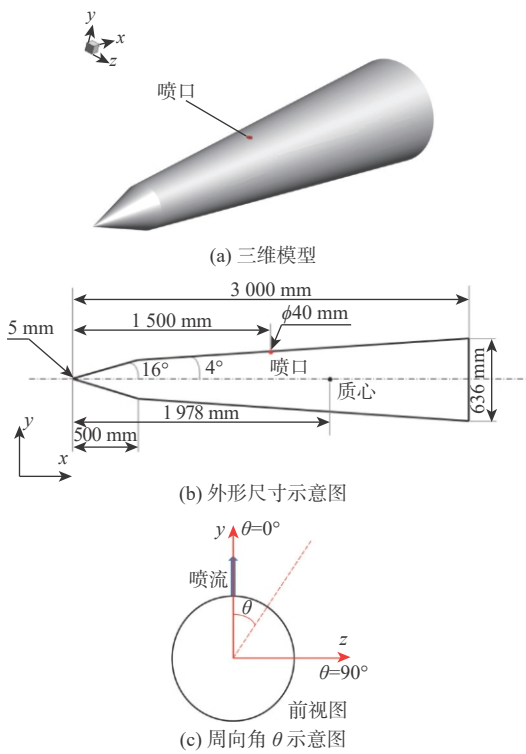


图 1　高超声速再入双锥飞行器外形参数及喷口位置

Fig. 1　Shape parameters and nozzle locations of hypersonic re-entry double-cone vehicle

本文坐标系定义如下: 取飞行器质心为坐标系原点与力矩参考点, 从头部顶点指向飞行器底部中心为 x 轴的正向, 迎角平面内垂直 x 轴向上为 y 轴正向, z 轴正向由右手坐标系确定, 喷口朝向 y 轴正向, 并且定义由飞行器底部向头部前视的 y 轴正向对应的周向角 θ=0°, y 轴负向对应的周向角 θ=180°, z 轴对应的周向角 θ=90°, 如图 1(c) 所示。

如图 2 所示, 本文选取飞行器周围的空气为计算域, 并采取结构化网格对计算域进行离散, 其中, 蓝色部分为流场域, 来流边界位于距头部顶点 L 处, 为压力远场边界条件, 周向边界位于距离弹体轴线 $6D$ 处, 底部出口位于距离弹体底部 $2.5L$ 处, 为压力出口边界条件; 飞行器表面为无滑移壁面边界条件, 喷口为压力入口边界条件。由于横向喷流与来流的相互干扰与壁面附近的流动关系十分密切, 将第 1 层网格高度选取 $\Delta h=5\times10^{-6}$ m 以确保无量纲壁面距离 $y^+\leqslant1$, 同时, 为了精确捕捉到喷口周围流场结构, 对喷口附近的网格进行加密, 如图 2(c) 所示。

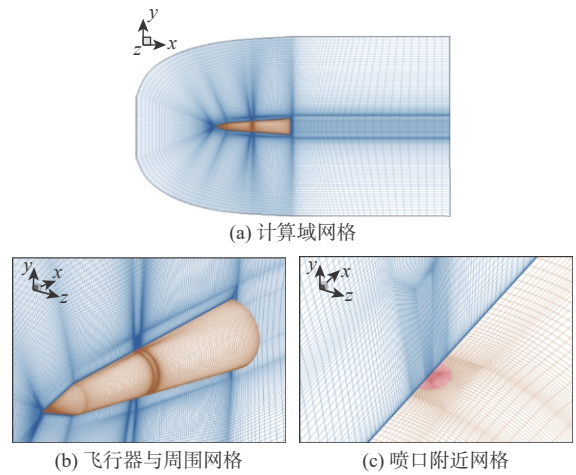


图 2　数值模拟计算网格

Fig. 2　Computational grid for numerical simulation

2.2　计算条件

本文在喷流参数一定的条件下, 研究高超声速再入双锥飞行器的横向喷流控制效率随飞行高度等参数的变化规律。临近空间高度跨度大, 但在 70 km 以上大气压强变化极小, 因此, 选择在 20～70 km 海拔高度的大气条件下进行数值模拟, 在模拟条件中对相应高度下的静压、空气密度及空气温度进行调整, 以达到改变飞行高度的效果。再入飞行器的速域跨度同样较大, 并且在飞行过程中会有一定的迎角和侧滑角, 因此, 分别在马赫数 6～18 的飞行速度、−30°～30°的迎角范围及 0°～18°的侧滑角范围条件下对再入飞行器横向喷流流场进行模拟, 具体条件如表 1 所示。

表 1　飞行模拟条件

Table 1　Flight simulation conditions

飞行参数	数值
飞行高度H/km	20,30,40,50,60,70
飞行马赫数Ma	6,9,12,15,18
迎角α/(°)	−30,−20,−10,0,10,20,30
侧滑角β/(°)	0,3,6,9,12,15,18

为确保在最极端的飞行条件下(H=20 km, Ma=18)喷流气体仍能喷出,经估算选择喷流压比为 H=20 km 大气压的 200 倍,即喷流总压 p_{0j}=1 105 862 Pa,喷流马赫数为 1,介质为空气,喷流的具体参数如表 2 所示。

表 2 喷流具体参数

Table 2 Specific parameters of jet

喷流参数	数值
喷流总压p_{0j}/Pa	1 105 862
出口压力p_e/Pa	584 184.89
喷流马赫数Ma_{jet}	1
喷流介质	空气
喷流总温T_{jet}/K	300

本文在表 1 和表 2 所示条件下对再入双锥飞行器分别进行了有横向喷流和无横向喷流 2 种工况的数值模拟,分别对有无喷流的飞行器表面压力进行积分,得到飞行器的总气动力大小,以及喷流干扰引起的飞行器附加气动力大小。有横向喷流时,飞行器上的附加气动力即为喷流干扰力,是由于喷流干扰效应而导致飞行器表面压力发生变化产生的额外气动力[19]。在本文中,横向喷流仅作用在飞行器的法向,因此,飞行器在有横向喷流和无横向喷流条件下受到的实际控制力变化为喷流净推力与附加气动力之和,产生的法向力控制效率可由喷流干扰法向力放大系数 K_N 来表示[19-21],表达式为

$$K_N = \frac{F_i + F_{jet}}{F_{jet}} \tag{6}$$

式中:F_i 为附加气动力;F_{jet} 为喷流净推力。

若附加气动力 F_i 很小,则 $K_N \approx 1$,此时飞行器受到的实际控制力基本等于喷流净推力;当附加气动力 F_i 与喷流净推力 F_{jet} 方向相同时,则 $K_N > 1$,说明此时喷流干扰会提高横向喷流的控制效率;方向相反时,则 $K_N < 1$,说明此时喷流干扰会削弱横向喷流的控制效率;若 $K_N < 0$,则说明喷流干扰对飞行器产生方向相反的干扰力已大于喷流净推力,得到的实际控制力方向与期望得到的喷流净推力方向完全相反。

同理,实际俯仰力矩控制效率可由喷流干扰俯仰力矩放大系数 K_{M_z} 来表示:

$$K_{M_z} = \frac{M_{z,i} + M_{z,jet}}{M_{z,jet}} \tag{7}$$

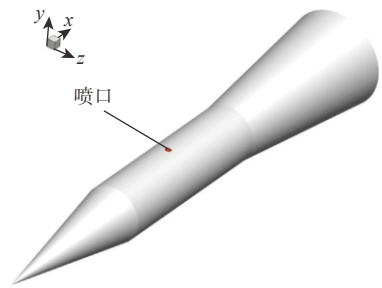
式中:$M_{z,i}$ 为喷流俯仰力矩;$M_{z,jet}$ 为喷流控制俯仰力矩。

3 数值方法验证

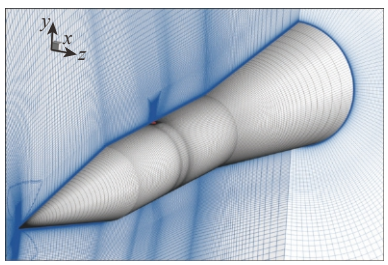
3.1 数值结果与实验数据的对比

在横向喷流干扰流场研究领域,锥-柱-裙 (cone-cylinder-flare, CCF) 导弹外形被研究人员大量应用于风洞试验和数值模拟研究中,该外形有大量的风洞试验数据,可以用于横向喷流控制数值模拟的验证对比。

本文以文献 [22] 进行的 CCF 外形横向喷流控制风洞试验数据为参考,进行数值模拟方法验证。数值模拟计算条件与文献 [22] 相同,即 Ma=3.0, Re=1.9×10^6, 迎角 α=0°,10°,−10°, 喷流压比 P_R=150, 喷流总温 T_{jet}= 280 K。采取与 2.1 节相同的结构化网格离散方法,如图 3 所示。



(a) CCF 三维外形示意图



(b) CCF 网格示意图

图 3 CCF 三维外形与网格

Fig. 3 3D shape and computational grid of CCF

横向喷流对弹身的气动干扰主要体现在对弹体表面压力分布的影响,因此,将数值模拟计算得到的弹体表面压力分布风洞试验结果对比,以验证数值模拟方法的准确性。与喷口同侧的弹体表面压力变化较为明显,因此,图 4 给出了通过数值模拟得到的不同迎角下弹体表面 θ=0°子午线上有横向喷流与无横向喷流时的压力系数差 ΔC_p 曲线,也给出了文献 [22] 的风洞试验结果。可以看出,在各迎角下,数值模拟计算结果与风洞试验数据结果吻合良好,说明本文数值模拟方法能够比较准确地模拟出横向喷流与来流的干扰流场,可用于进行飞行器横向喷流控制流场的数值模拟研究。

3.2 网格无关性验证

为确保网格数量对计算结果不会造成影响,需要进行网格无关性验证。表 3 为进行网格无关性验证使用的 3 套网格在本文的再入双锥飞行器表面及空间上的节点分布情况,网格 A、网格 B 和网格 C 的网格量分别为 213 万、544 万和 770 万。在 Ma=6.0、α=0°和 H=20 km 的大气环境条件下进行有

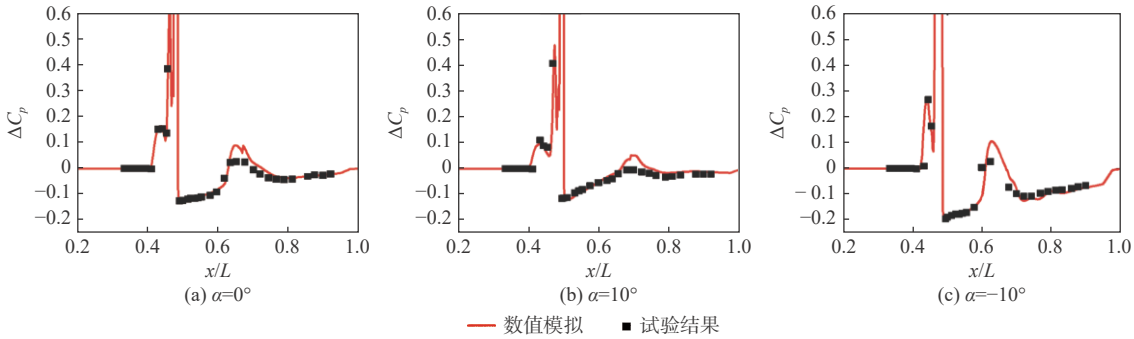


图 4　θ=0°子午线上有无横向喷流的压力系数差计算曲线与试验值对比结果

Fig. 4　Comparison of CFD-computed pressure coefficient difference between lateral jet on and off along θ=0° meridian line against experimental data

表 3　不同网格的节点分布情况

Table 3　Node distribution in different grids

节点方向	网格A	网格B	网格C
轴向(表面)	146	195	210
径向(流场)	130	170	195
周向(表面)	112	164	188

横向喷流下的数值模拟。

表 4 给出了分别使用 3 套网格计算得到的再入双锥飞行器气动系数之间相对误差的绝对值。可以看出, 用网格 A 与其余 2 套网格计算得到的结果, 每项气动系数之间的相对误差为 1%～8%, 而用网格 B 与网格 C 计算得到的双锥飞行器各项气动力系数之间的误差均在 1% 以内, 表明在网格 B 的基础上继续增加网格量对计算结果的影响已很小, 因此, 可认为网格 B 对计算结果已经达到了网格无关, 后续进行的高超声速再入双锥飞行器横向喷流数值模拟计算采用网格 B 进行。

表 4　不同网格计算的气动系数之间的相对误差

Table 4　Relative errors in aerodynamic coefficients between different grid computations　%

对比组	法向力	俯仰力矩	轴向力
网格A与网格B	2.56	3.20	1.97
网格A与网格C	5.73	7.41	4.04
网格B与网格C	0.32	0.79	0.46

4　结果分析

4.1　喷流干扰对流场与飞行器表面压力分布的影响

如 2.2 节所述, 横向喷流控制效率的变化通过喷流干扰法向力放大系数 K_N 和俯仰力矩放大系数 K_{M_z} 的变化表示, 其受到横向喷流与来流干扰流场结构引起的飞行器表面压力分布变化影响, 因此, 需要对有横向喷流时流场结构和飞行器表面压力变化进行分析。图 5 给出了本文再入双锥飞行器在飞行高度 H=30 km、飞行马赫数 Ma=6.0 的条件

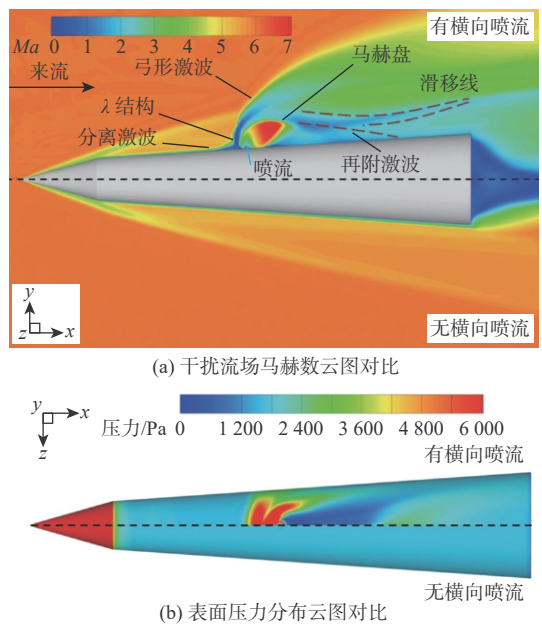


图 5　有无横向喷流条件下 z=0 平面内流场马赫数云图和飞行器表面压力分布云图对比 (Ma=6, H=30 km)

Fig. 5　Comparison of Mach number contours and surface pressure distribution contours in z=0 plane with and without lateral jet flow conditions (Ma=6, H=30 km)

下, 有横向喷流和无横向喷流时 z=0 的纵向对称截面内横向喷流干扰流场马赫数云图和飞行器表面压力分布云图的对比。从图 5(a) 可以看出, 在有横向喷流时, 高超声速来流与横向喷流发生相互作用, 在喷口上游形成了弓形激波。弓形激波与喷口前方的边界层发生干扰, 引起喷口上游一定位置内的边界层分离, 产生分离激波, 分离激波与弓形激波共同形成了喷口上游的 λ 形波系结构。喷流从喷口进入流场后迅速膨胀, 并逐渐向下游方向倾斜, 形成马赫盘。在马赫盘后方生成一道再附激波与 2 条滑移线, 再附激波作用于弹体表面。

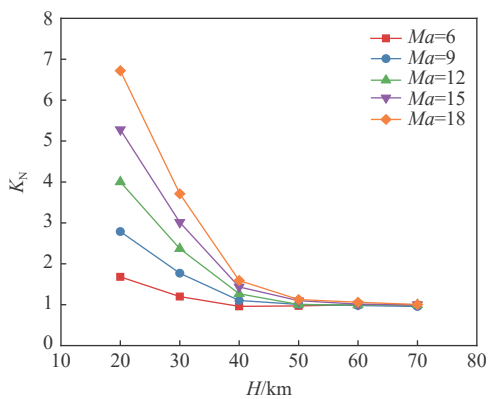
从图 5(b) 可以看出, 横向喷流会对飞行器表面压力分布产生显著影响。在喷口上游来流受到阻碍, 流速降低, 与喷流气体互相压缩, 在飞行器表面

形成高压区；喷口下游喷流气体出喷口后继续膨胀，压力降低，对飞行器表面造成吸力作用，从而形成低压区；喷口前的高压气流绕过喷口在飞行器尾部聚拢，与喷流气体互相产生压缩作用而形成高压区。

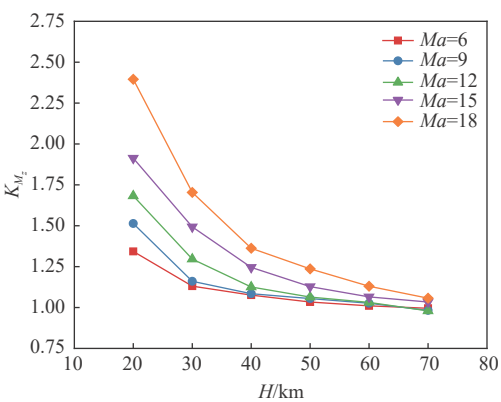
横向喷流会对飞行器绕流场造成强烈干扰，对飞行器表面压力分布造成影响，产生额外的喷流干扰气动力，对横向喷流控制效率造成放大或削减作用。

4.2 飞行高度对横向喷流控制效率的影响

在喷流参数一定且迎角与侧滑角均为0°的条件下，对不同飞行高度条件下高超声速再入双锥飞行器横向喷流控制时的绕流场进行模拟计算，研究横向喷流控制效率随飞行高度的变化规律及其机理。图6给出了不同飞行马赫数条件下，通过数值模拟计算得到的喷流干扰法向力放大系数 K_N 和喷流干扰俯仰力矩放大系数 K_{M_z} 随飞行高度的变化曲线。可以看出，在不同马赫数下，K_N 和 K_{M_z} 均随飞行高度增大而逐渐下降，在 H=40 km 以下的高度，下降速率较快，随着飞行高度增加，下降速率降低，且 K_N 和 K_{M_z} 均逐渐接近于1，说明随飞行高度增大，



(a) 喷流干扰法向力放大系数



(b) 喷流干扰俯仰力矩放大系数

图6 不同飞行马赫数条件下喷流干扰法向力放大系数和俯仰力矩放大系数随飞行高度的变化曲线

Fig. 6 Curves of jet normal force amplification coefficient and pitch moment amplification coefficient varying with flight altitude under different flight Mach numbers conditions

横向喷流控制干扰力对横向喷流控制效率的放大效果在削弱，高海拔飞行时，飞行器受到的喷流净推力及喷流干扰力的合力作用与无横向喷流干扰时的喷流净推力基本一致，喷流干扰力对横向喷流控制效率产生的影响极小。

图7为 Ma=9 时不同飞行高度下有、无横向喷流时飞行器上表面的压力差 Δp 分布云图与极限流线图。可以看出，随着飞行高度的增加，喷口前方分离点明显前移，喷口前高压区范围扩大，但压力上升幅值下降，喷口后方高压区压力上升幅值同样下降，并且向下游移动。因此，与喷流净推力方向相同的压力会随着飞行高度的增加而被削弱，使 K_N 和 K_{M_z} 均减小，降低了横向喷流控制效率。

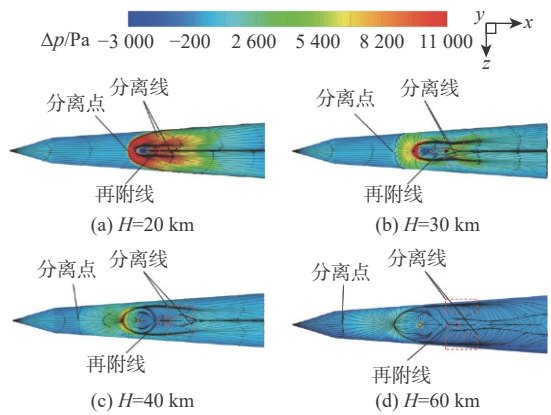


图7 不同飞行高度下飞行器上表面有无横向喷流的压力差分布云图与极限流线图 (Ma=9)

Fig. 7 Pressure difference between lateral jet on and off distribution contours and limit streamlines on vehicle upper surface at different flight altitudes (Ma=9)

为分析出现上述现象的原因，图8给出了 Ma=9 时不同飞行高度下 z=0 平面内飞行器周围流场马赫数云图与飞行器表面极限流线图。可以看出，随着飞行高度的增加，喷流出喷口后膨胀区域明显变大，马赫盘体积增加，弓形激波逐渐直立并

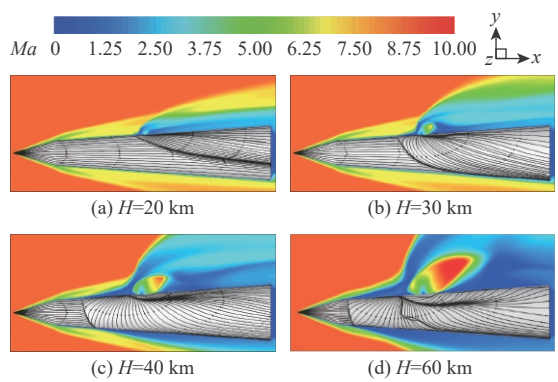


图8 不同飞行高度下 z=0 平面内流场马赫数云图和飞行器表面极限流线图 (Ma=9)

Fig. 8 Mach number contours in z=0 plane and limit streamlines on vehicle surface at different flight altitudes (Ma=9)

远离飞行器表面, 使再附气流作用到飞行器上的位置远离喷口。随着飞行高度的增加, 来流动压减小, 喷流与来流的压比逐渐增大, 造成了更强的逆压梯度, 使喷口前方的分离程度增加, 分离起始点向上游移动, 同时来流动压的减小会导致分离区压力上升幅度下降。

从图 8 中也能看出, 随着飞行高度增大, 气流向弹体下表面偏转程度明显增加, 即增强了“缠绕效应”, 从而对弹体下表面压力分布造成影响, 在 H = 20 km 时, 下表面并未受到“缠绕效应”的影响, H > 20 km 后, “缠绕效应”明显影响到飞行器下表面。图 9 为不同飞行高度下飞行器表面 θ=180°子午线上压力差分布曲线。当 H=20 km 时, 飞行器下表面压力分布不会受到喷流干扰的影响, H>20 km 后, 下表面压力升高, 增强了与实际喷流推力相反的压力, 进一步降低了横向喷流控制效率。

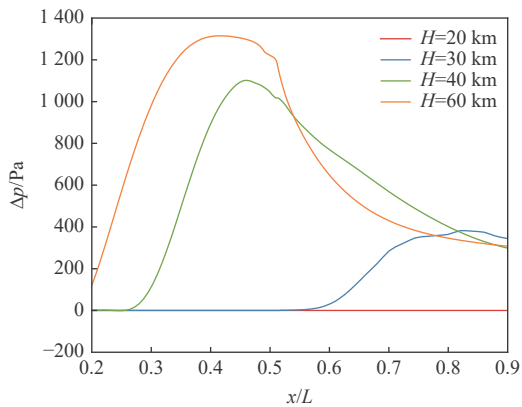


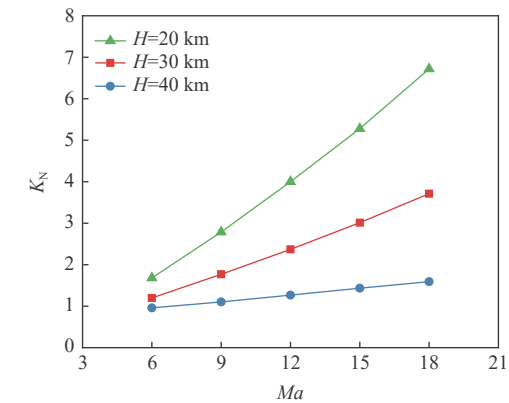
图 9　不同飞行高度下飞行器表面 θ=180°子午线上有无横向喷流的压力差分布曲线 (Ma=9)

Fig. 9　Pressure difference distribution curves between lateral jet on and off on vehicle surface along θ=180° meridian line at different flight altitudes (Ma=9)

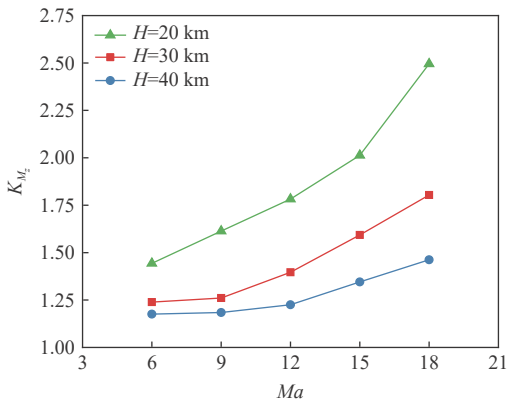
当飞行马赫数一定时, 飞行高度的增加会降低来流动压, 导致喷流与来流压比上升, 产生更为强烈的逆压梯度和黏性阻滞效应, 从而使喷口前方分离区起始位置前传。但由于来流动压降低, 导致分离区压力上升幅值与喷口下游高压区压力上升幅值减小, 削弱了与实际喷流控制方向相同的压力, 使 K_N 和 K_{M_z} 均减小, 降低了横向喷流控制效率; 距离质心更远的喷口前高压区压力下降幅度小于距离质心较近的喷口后高压区下降幅度, 导致喷口前高压区始终对飞行器有较长力臂的作用, 最终使得 K_{M_z} 下降幅度小于 K_N, 因此, 横向喷流俯仰力矩的控制效率下降程度比喷流法向力的控制效率下降程度更小。

4.3　飞行马赫数对横向喷流控制效率的影响

在喷流参数一定且迎角与侧滑角均为 0°的条件下, 对不同飞行马赫数条件下高超声速再入双锥飞行器横向喷流控制时的绕流场进行模拟计算, 研究横向喷流控制效率随飞行马赫数的变化规律及其机理。图 10 给出了不同飞行高度下通过数值模拟计算得到的喷流干扰法向力放大系数 K_N 和喷流干扰俯仰力矩放大系数 K_{M_z} 随飞行马赫数的变化曲线。可以看出, 喷流干扰法向力放大系数 K_N 随飞行马赫数增大基本呈线性增大, 且飞行高度越低, 增长率越高, 在 H=40 km 时, K_N 约为 1～2, 在 H= 20 km 时, K_N 约为 1～7; 喷流干扰俯仰力矩放大系数 K_{M_z} 随飞行马赫数增大, 基本呈非线性增大, 同样, 飞行高度越低, 增长率越高, 在 H=40 km 时, K_{M_z} 约为 1～1.5, 在 H=20 km 时, K_{M_z} 约为 1.25～2.5。



(a) 喷流干扰法向力放大系数



(b) 喷流干扰俯仰力矩放大系数

图 10　不同飞行高度下喷流干扰法向力放大系数和俯仰力矩放大系数随飞行马赫数的变化曲线

Fig. 10　Curves of jet normal force amplification coefficient and pitch moment amplification coefficient varying with flight Mach number under different flight altitudes

对飞行器的表面压力变化与流场结构变化进行分析, 研究飞行马赫数对横向喷流控制效率的影响机理。定义飞行器表面压力差 Δp 为有横向喷流时飞行器表面压力与无横向喷流时飞行器表面压力的差值。图 11 为 H=30 km 时不同飞行马赫数下飞行器表面 θ=0°子午线上压力差 Δp 的分布曲线, x/L 为距飞行器头部顶点的相对距离, 其中, 喷口位

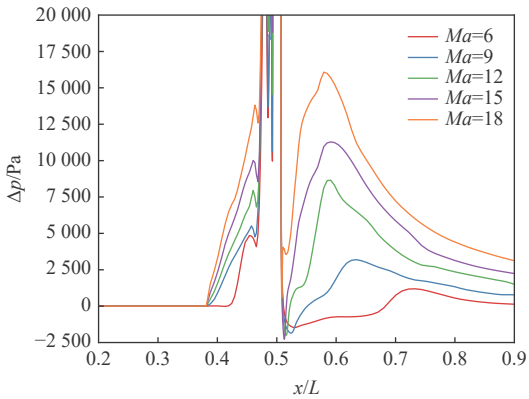


图 11 不同飞行马赫数下飞行器表面 θ=0°子午线上有无横向喷流的压力差分布曲线 (H=30 km)

Fig. 11 Pressure difference distribution curves between lateral jet on and off on vehicle surface along θ=0° meridian line at different flight Mach numbers (H=30 km)

于 x/L=0.5 处。可以看出，在飞行器表面 θ=0°的子午线上，横向喷流导致喷口前方压力升高、喷口后方压力先降低后升高，并随着飞行马赫数的增加，喷口前方和后方的高压区压力幅值均增加，喷口前方高压区位置逐渐向上游移动，后方高压区逐渐靠近喷口。

为分析上述现象出现的原因，图 12 给出了 H=30 km 时，不同飞行马赫数下飞行器上表面有无横向喷流的压力差云图与极限流线图。可以看出，随着飞行马赫数的增加，喷口前高压气流压力上升，分离点逐渐前传，喷口后高压区压力升高，并逐渐靠近喷口。飞行马赫数的增加会使来流动压上升，与喷流气体产生更强烈的压缩作用，导致喷口前高压区压力上升；喷口后的高压区是由喷口前高压气流绕过喷口向下游发展，并逐渐向中心线靠拢，对再附到飞行器表面的喷流气体互相形成压缩

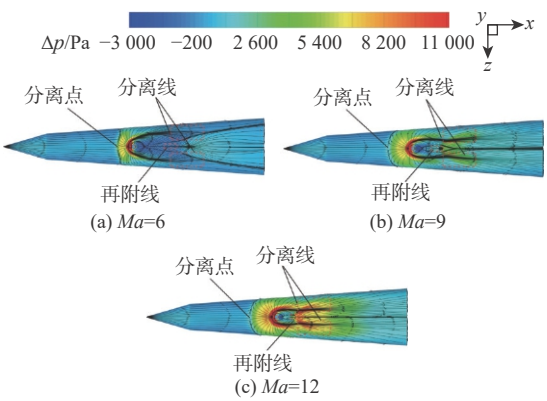


图 12 不同飞行马赫数下飞行器上表面有无横向喷流的压力差云图与极限流线图 (H=30 km)

Fig. 12 Pressure difference distribution contours between lateral jet on and off and limit streamlines on vehicle upper surface at different flight Mach numbers (H=30 km)

作用而产生的，如图 12 中红色方框区域所示，喷流气体再附到飞行器表面形成再附线，并对靠拢的高压气流形成阻碍作用，造成流动分离形成 2 条分离线。因此，飞行马赫数的增加会增强与实际喷流净推力方向相同的压力，使 K_N 和 K_{M_z} 均增大，提升喷流的控制效率。

从流场结构的变化分析造成上述现象的机理。图 13 为 H=30 km 时不同飞行马赫数下 z=0 平面内流场马赫数云图与飞行器表面极限流线图。可以看出，随着飞行马赫数的增加，飞行器头部产生的附体激波角减小、激波更强，受飞行器头部激波及波后气流的干扰影响，喷流出喷口后膨胀区域缩小，马赫盘体积明显减小；喷流前的弓形激波倾斜程度增加，并向飞行器表面靠近，使再附气流作用到飞行器上的位置更靠近喷口，抑制了喷流气体的膨胀作用，导致了再附线向喷口移动，低压区范围减小，高压区向喷口靠近。

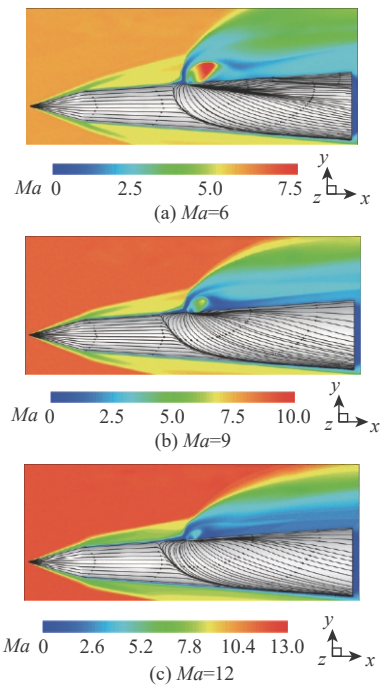


图 13 不同飞行马赫数下 z=0 平面内流场马赫数云图与飞行器表面极限流线图 (H=30 km)

Fig. 13 Mach number contours in z=0 plane and limit streamlines on vehicle surface at different flight Mach numbers (H=30 km)

由于飞行马赫数的增加会导致来流与壁面产生更剧烈的摩擦，使飞行器表面的边界层内气体温度升高，黏性增加，最终导致边界层厚度增加。边界层厚度和黏性的增加会对来流气体产生更强的黏性阻滞效应，使来流更加难以克服由于喷流产生的逆压梯度，从而在喷口前方产生更大程度的分离，使起始位置向上游移动。

综合上述分析, 在一定飞行高度时, 飞行马赫数的增加会使横向喷流喷口前方分离区压力上升, 并且位置提前, 喷口后方高压区压力升高, 并逐渐接近喷口, 增强与实际喷流净推力方向相同的压力; 喷口后方低压区范围与压力下降幅值均减小, 削弱了与实际喷流净推力方向相反的吸力, 与喷口前高压区共同提升横向喷流控制效率, 使 K_N 和 K_{M_z} 均增大。但喷口后方高压区部分位于质心位置 (x/L=0.66) 后, 产生与横向喷流控制力矩相反的抬头力矩, 对 K_{M_z} 产生反作用, 导致其放大效果小于 K_N。因此, 喷流法向力的控制效率提升效果比喷流俯仰力矩的控制效率提升效果更加显著。

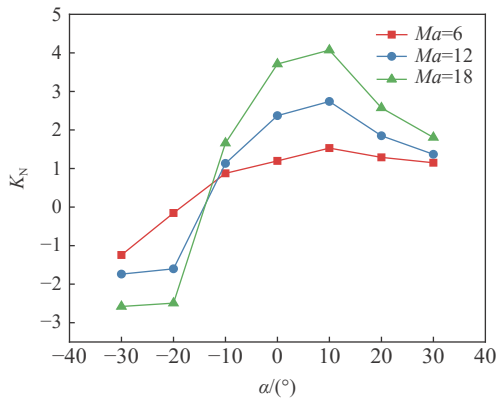
4.4　迎角对横向喷流控制效率的影响

在喷流参数一定、侧滑角 β 为 0°的状态下, 对不同迎角下高超声速再入双锥飞行器横向喷流控制时的绕流场进行模拟计算, 研究横向喷流控制效率随迎角的变化规律及其机理。由于飞行高度和飞行马赫数对横向喷流控制效率影响的机理相似, 都是由来流动压变化引起, 本节仅将飞行马赫数作为动压变化的考量, 飞行高度选取喷流干扰较为强烈的 30 km, 飞行马赫数选取 6、12、18, 迎角变化范围为 −30°～30°(间隔 10°)。

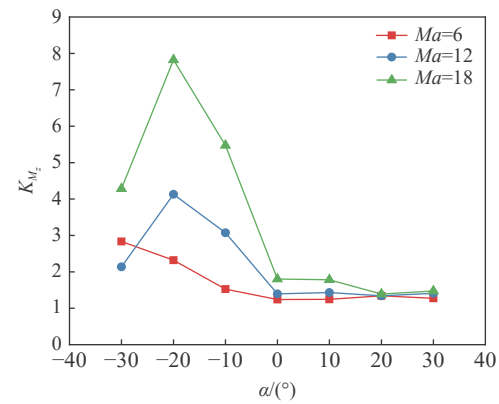
图 14 给出了 H=30 km 时不同飞行马赫数下再入双锥飞行器的喷流干扰法向力放大系数 K_N 和喷流干扰俯仰力矩放大系数 K_{M_z} 随迎角的变化曲线。可得, K_N 和 K_{M_z} 随迎角的变化规律并不一致。当 α=−20°,−30°时, K_N 在所有飞行马赫数下均小于 0, 说明喷流法向力的控制效率为负, 喷流干扰力产生了与实际喷流净推力完全相反的作用, 并且飞行马赫数越大, 反作用效果越强; 在 Ma=6, α=−20°时, $K_N \approx 0$, 说明此时飞行器受到的喷流法向力与实际喷流净推力相互抵消, 除来流产生的法向力外, 受到额外的法向控制力几乎为零。随着迎角的增加, K_N 先增大后减小, 在 α=10°时达到最大值, 喷流法向力的控制效率达到最优, 并且放大效果随着飞行马赫数的增加而增加; 此后随着迎角的继续增大, 在所有飞行马赫数下, K_N 逐渐减小并在 α=30°时向 1 靠近, 说明此时喷流法向力对飞行器产生的作用效果可以忽略不计, 飞行器受到的法向控制力大小接近于实际喷流净推力的大小。

从图 14(b) 中可以看出, K_{M_z} 在所有迎角下均大于 1, 说明喷流干扰产生的俯仰力矩始终对喷流净推力产生的俯仰力矩造成放大效果。在 Ma=12, 18 条件下, 当迎角由−30°逐渐增加时, K_{M_z} 在−20°时达到最大值, 然后减小并逐渐向 1 靠近; 在 Ma=6 条件下, K_{M_z} 在−30°时达到最大值, 并随着迎角的增加而下降并向 1 靠近。综上可以得出, 在所有飞行马赫数下, 喷流俯仰力矩的控制效率在较大的负迎角时达到最大, 随着迎角的增加而下降, 并在较大正迎角时, 喷流俯仰力矩对飞行器产生的作用效果可以忽略, 飞行器受到的俯仰力矩大小接近于实际喷流净推力产生的俯仰力矩大小。

下面通过飞行器表面压力的变化对上述现象进行分析。图 15 给出了 Ma=6 时, α=−30°,0°,10°,



(a) 喷流干扰法向力放大系数



(b) 喷流干扰俯仰力矩放大系数

图 14　不同飞行马赫数下喷流干扰法向力放大系数和俯仰力矩放大系数随迎角的变化曲线 (H=30 km)

Fig. 14　Curves of jet normal force amplification coefficient and pitch moment amplification coefficient varying with angle of attack under different flight Mach numbers (H=30 km)

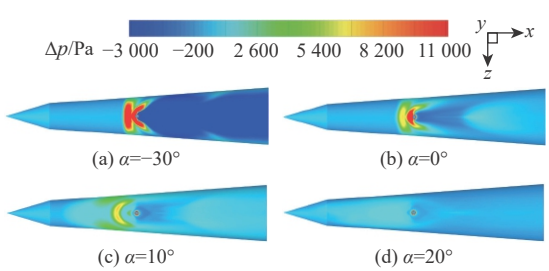


图 15　不同迎角下飞行器上表面有无横向喷流的压力差云图 (Ma=6)

Fig. 15　Pressure difference distribution contours between lateral jet on and off on vehicle upper surface at different angles of attack (Ma=6)

20°条件下飞行器上表面有无横向喷流的压力差云图。可以看出，当 α=−30°时，虽然喷口前方高压区强度较大，但在喷口后方出现了大面积的低压区，产生与喷流净推力相反的吸力，这是导致 K_N 为负值的主要原因；低压区大部分位于质心位置 (x/L=0.66) 后，产生与实际喷流控制俯仰力矩方向相同的低头力矩，从而放大了 K_{M_z}。随着迎角的增加，喷口前方高压区压力幅值下降，范围明显扩大，喷口后方低压区范围减小，低压程度降低，并出现局部压力升高的区域，导致 K_N 增加、K_{M_z} 减小。在 α 由 10°增至 20°时，飞行器上表面压力上升幅度降低，因此，K_N 和 K_{M_z} 开始降低并逐渐接近于 1，喷流干扰力对喷流法向力和俯仰力矩的控制效率几乎不造成影响。

以下通过流场结构的变化揭示出现上述现象的原因。图 16 给出了 Ma=6 时，α=−30°,0°,10°,20° 条件下 z=0 平面内马赫数云图及飞行器表面极限流线图。可以看出，当迎角为负时，喷口上游的弓形激波被飞行器头部产生的斜激波强烈干扰，从而向表面弯曲，对喷流气体产生压制作用，使其出喷口后倾斜程度增大，喷流气体尾迹作用到飞行器表面的范围增大。随着迎角的增加，喷口上游的分离区范围明显增大。弓形激波逐渐远离飞行器表面，对喷流气体的压制作用减小，使其膨胀程度增加且更加直立，喷流气体尾迹几乎不作用到飞行器表面，从图 16(d) 中可以看出，在 α=20°时，喷流尾迹几乎已经完全远离飞行器表面，因此，对喷口后方吸力作用降低，导致低压区范围减少。

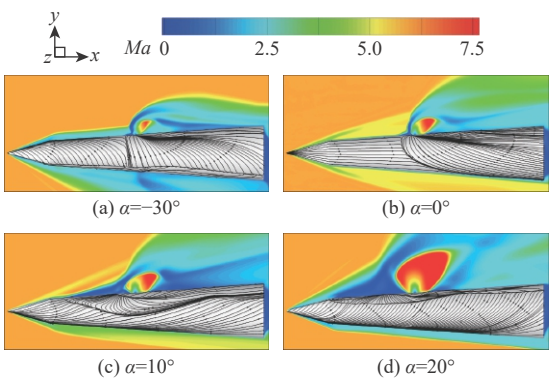


图 16 不同迎角下 z=0 平面内流场马赫数云图与飞行器表面极限流线图 (Ma=6)

Fig. 16 Mach number contours in z=0 plane and limit streamlines on vehicle surface at different angles of attack (Ma=6)

图 16 中极限流线图能够反映飞行器表面的流动情况。当迎角逐渐增大时，气流向飞行器下表面的偏转程度明显降低，即“缠绕效应”降低，对下表面气流的压缩程度大幅度下降；当迎角为正时，气流均不向下偏转，因此，对下表面压力分布不造成任何影响。图 17 给出了 Ma=6 时不同迎角下 θ=180°子午线上有无横向喷流的压力差分布曲线。可知，在迎角由−30°转为 0°时，下表面压力均有不同程度的增加，且压力峰值位置逐渐后移，峰值先增大后减小；在迎角为正时，下表面压力差均为零，即有横向喷流时也不会使下表面压力发生变化。下表面压力的增加会对喷流净推力造成反作用效果，因此，在正迎角时，喷流法向力控制效率不会受到飞行器表面气流偏转的影响，导致 K_N 在正迎角下较大。

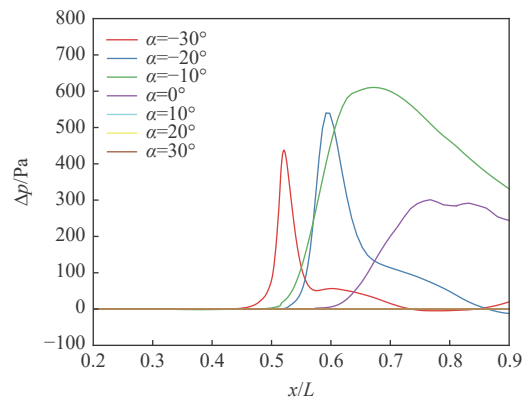
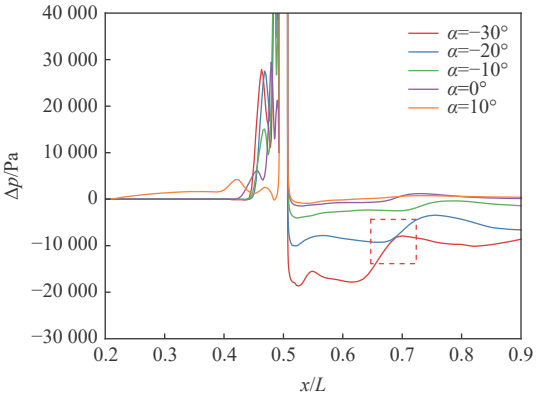


图 17 不同迎角下飞行器表面 θ=180°子午线上有无横向喷流的压力差分布曲线 (Ma=6)

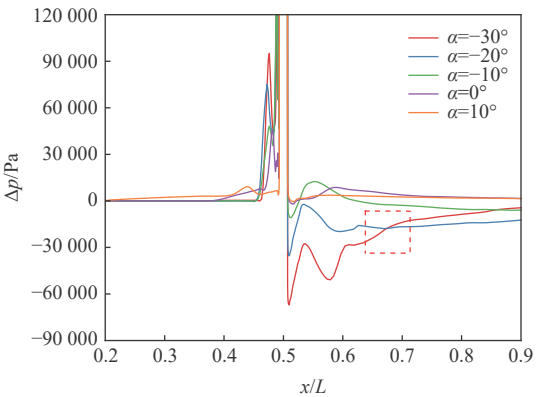
Fig. 17 Pressure difference distribution curves between lateral jet on and off on vehicle surface along θ=180° meridian line at different angles of attack (Ma=6)

下面对 Ma=12,18 条件下 K_{M_z} 在 α=−20°时达到最大值，而 Ma=6 时在 α=−30°达到最大值的原因进行分析，因 Ma=12 与 Ma=18 曲线规律类似，取 Ma=12 的情况进行分析。图 18 为 Ma=6,12 时不同迎角下飞行器表面 θ=0°子午线上有无横向喷流压力差分布曲线。可得，α=−30°, −20°时，喷口后方均为低压区。在质心位置 (x/L=0.66) 前的低压区产生的吸力作用会对弹身产生与喷流净推力产生的低头力矩相反的抬头力矩，因此，会对 K_{M_z} 造成削弱效果；而在质心位置后方的低压区产生的吸力会产生低头力矩，从而对 K_{M_z} 造成增强效果。当 Ma=6 时，迎角从−30°向−20°转变过程中，质心位置前方低压区吸力作用减弱，对 K_{M_z} 的削弱效果减小，但质心后方吸力作用同样程度减弱，如图 18(a) 中红框区域所示，因此，对 K_{M_z} 的增强效果也会减小，综合导致了 K_{M_z} 的减小；而当 Ma=12 时，质心位置后方的吸力作用却没有减弱，甚至比 α=−30°时更大，进而放大了对 K_{M_z} 的增强效果，Ma=18 时同理。

图 19 给出了 H=30 km, Ma=6,12 时，在 α=−30° 条件下 z=0 的纵向对称平面内与飞行器表面的压



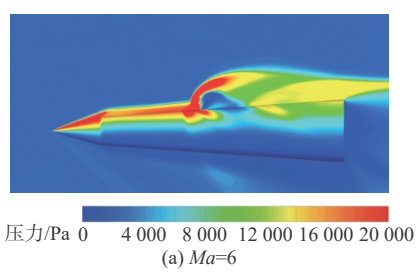
(a) *Ma*=6



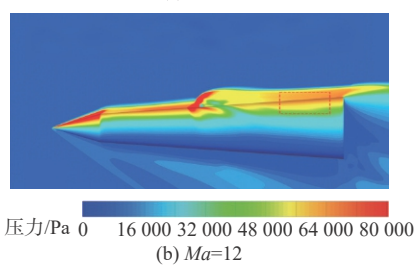
(b) *Ma*=12

图 18　不同迎角下飞行器表面 θ=0°子午线上有无横向喷流的压力差分布曲线 (*Ma*=6,12)

Fig. 18　Pressure difference distribution curves between lateral jet on and off on vehicle surface along θ=0° meridian line at different angles of attack (*Ma*=6,12)



(a) *Ma*=6



(b) *Ma*=12

图 19　α=−30°时 z=0 平面内与飞行器表面压力分布云图 (*Ma*=6,12)

Fig. 19　Pressure distribution contours on z=0 plane and vehicle surface at α=−30°(*Ma*=6,12)

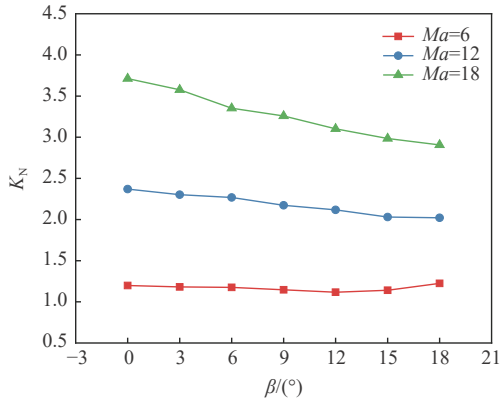
力分布云图。可以看出, 相较于 *Ma*=6 的情况, *Ma*=12 时的弓形激波更贴近飞行器表面, 高压气流向飞行器表面偏转, 并在下游处直接接触飞行器表面, 如图 19(b) 中红框区域所示, 从而抵消了部分喷流气体的吸力作用, 使质心后方低压区压力降低程度下降; 因此, 会削弱质心后方喷流气体吸力作用产生的低头力矩, 导致在 *Ma*=12, α=−30°时 K_{M_z} 还未到达最大值, 随着迎角增大至−20°, 弓形激波逐渐远离飞行器表面后 K_{M_z} 到达最大值, *Ma*=18 时同理; 而在 *Ma*=6, α=−30°时, 弓形激波产生的高压气流对喷口后方的吸力作用抵消效果弱, 因此, *Ma*=6 时 K_{M_z} 的变化规律与 *Ma*=12,18 时不一致。

由上述分析可得, 飞行迎角的变化对喷流法向力和俯仰力矩的控制效率影响规律并不一致。在较大负迎角时, 虽然喷口前方形成了压力上升幅度较高的高压区, 但在喷口后方喷流气体大面积作用在飞行器表面, 产生与实际喷流净推力相反的吸力, 同时飞行器表面气流向下表面偏转形成与实际喷流净推力相反的压力, 与上表面大面积低压区共同作用导致 K_N 为负; 随着迎角向正转变, 喷口前高压区压力上升幅度降低但范围增大, 喷口后方喷流气体逐渐远离飞行器表面, 导致吸力作用减弱, 并且飞行器表面气流向下表面偏转程度降低, 共同作用使 K_N 逐渐转为正值并达到最大, 本文中, 在 *Ma*=6,12,18, α=10°时, 喷流法向力控制效率达到最优; 在较大负迎角时, 喷流气体产生的吸力大量作用在质心位置后, 产生与实际喷流控制俯仰力矩方向相同的低头力矩, 因此, K_{M_z} 达到最大值, 但在不同飞行马赫数下, 喷流气体对飞行器表面产生的吸力效果并不一致, 因此, K_{M_z} 达到最大值时的迎角不同, 本文中, *Ma*=6 条件下, 在 α=−30°时喷流俯仰力矩的控制效率达到最优, *Ma*=12,18 条件下, 在 α=−20°时达到最优。增加至较大正迎角时, 喷流干扰在飞行器上下表面产生的影响均下降, 最终使得 K_N 和 K_{M_z} 逐渐接近 1。

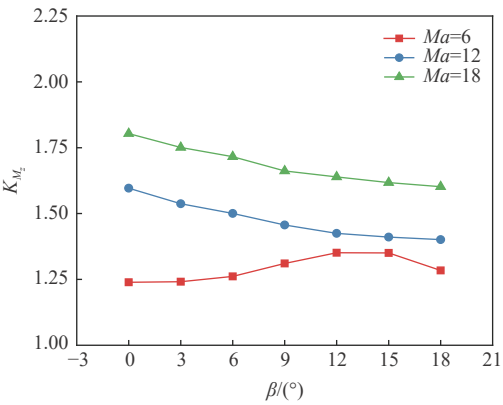
4.5　侧滑角对横向喷流控制效率的影响

在喷流参数一定、迎角 α=0°的状态下, 对不同侧滑角下高超声速再入双锥飞行器横向喷流控制时的绕流场进行模拟计算, 研究横向喷流控制效率随侧滑角的变化规律及其机理。飞行高度选取喷流干扰较为强烈的 H=30 km, 飞行马赫数选取 6、12、18, 侧滑角变化范围为 0°～18°(间隔 3 °)。

图 20 为 H=30 km 时不同飞行马赫数下再入双锥飞行器的喷流干扰法向力放大系数 K_N 和喷流干扰俯仰力矩放大系数 K_{M_z} 随侧滑角的变化曲线。可以看出, K_N 和 K_{M_z} 在所有飞行马赫数及侧滑角条件下均大于 1。当 *Ma*=12,18 时, K_N 和 K_{M_z} 随侧滑角的增大而减小; 当 *Ma*=6 时, K_N 的变化幅度极小, 会随着侧滑角增大而先减小后在 15°处开始增加, K_{M_z} 会



(a) 喷流干扰法向力放大系数



(b) 喷流干扰俯仰力矩放大系数

图 20 不同飞行马赫数下喷流干扰法向力放大系数和俯仰力矩放大系数随侧滑角的变化曲线 (H=30 km)

Fig. 20 Curves of jet normal force amplification coefficient and pitch moment amplification coefficient varying with sideslip angle under different flight Mach numbers (H=30 km)

随着侧滑角的增大而先增大后在15°处开始减小。

下面通过分析喷流前后飞行器表面压力变化来揭示出现上述现象的原因。图21为Ma=6,12时不同侧滑角下飞行器上表面有无横向喷流的压力差云图。可得,在Ma=6条件下,当侧滑角增加,靠近飞行器尾部的高压区顺时针偏转,并逐渐脱离飞行器上表面,飞行器-z侧的低压区顺时针偏转,并转移至飞行器上表面,综合导致与喷流净推力方向相同的压力作用减小,而与喷流净推力方向相反的吸力作用增大,从而削弱了喷流干扰力对喷流净推力的放大作用,导致K_N下降;由于喷口后的低压区基本位于质心位置(x/L=0.66)之后,产生的力矩与喷流净推力产生的低头力矩方向相同,导致K_{M_z}上升;当侧滑角由15°增至18°时,喷口前高压区顺时针偏转并蔓延至飞行器尾部的上表面,导致尾部上表面压力升高,使K_N上升而K_{M_z}下降;在Ma=12条件下,喷口后方低压区范围较小,随着侧滑角的增加总体变化不大,而喷口后高压区则会顺时针偏转,并逐渐脱离飞行器表面,因此,削弱了喷流干扰力对喷流净推力的放大作用,导致K_N下降;并且喷口后高压区主要位于质心位置之前,因此,K_{M_z}同样会随着侧滑角的增加而下降,Ma=18时同理。

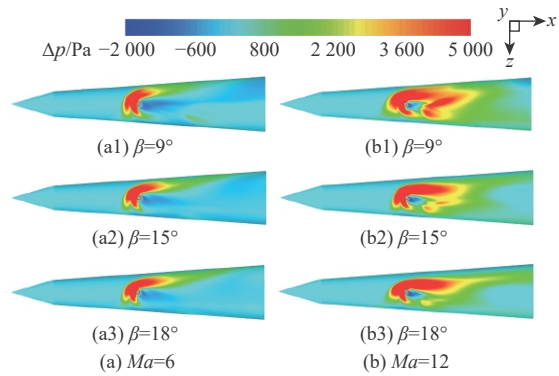


图 21 不同侧滑角下飞行器上表面有无横向喷流的压力差云图 (Ma=6, 12)

Fig. 21 Pressure difference distribution contours between lateral jet on and off on vehicle upper surface at different sideslip angles (Ma=6, 12)

从上述分析可得,在Ma=6时,侧滑角的增加导致喷口后方的高压区和低压区同时偏转,并脱离飞行器表面,但飞行器尾部上表面压力先降低后增大,使K_N先减小后增大,K_{M_z}先增大后减小,但总体变化幅度都较小;在Ma=12,18时,由于喷口后方几乎不存在低压区,当喷口后方的高压区偏转并脱离飞行器表面时,K_N和K_{M_z}均逐渐下降,并且在无侧滑角时,喷流法向力及俯仰力矩控制效率达到最优。此外,有侧滑角时,喷流干扰力不仅作用在法向方向上,还会由于其不对称性产生额外的偏航或滚转方向上的力,因此,应避免在有侧滑角时采取喷流控制。

5 结 论

1) 当飞行高度一定时,增加飞行马赫数会放大喷流法向力和俯仰力矩的控制效率,使其大于1,但喷流俯仰力矩的控制效率放大程度要小于喷流法向力控制效率的放大程度。

2) 当飞行马赫数一定时,增加飞行高度会降低喷流法向力和俯仰力矩的控制效率,飞行高度越高,喷流法向力和俯仰力矩的控制效率越接近于1。

3) 在喷流干扰较为强烈的飞行高度下,不同飞行马赫数下,迎角对喷流法向力和喷流俯仰力矩控制效率造成的影响并不一致。α=−20°,−30°时,所有飞行马赫数下,喷流法向力的控制效率均为负,并随着迎角的增大而增大,在α=10°时均达到最大值;Ma=6条件下,在α=−30°时,喷流俯仰力矩控制效率达到最大值,Ma=12,18条件下,在α=−20°时达到最大值。随着正迎角的增加,喷流法向力和俯仰力矩

的控制效率均向 1 靠近。

4) 在喷流干扰较为强烈的飞行高度下, 不同飞行马赫数下侧滑角对喷流法向力和俯仰力矩的控制效率造成的影响并不一致, 但总体影响较小。*Ma*=6 条件下, 喷流法向力控制效率基本不随侧滑角的变化而变化, 喷流俯仰力矩的控制效率在 β=15°时达到最大值; *Ma*=12,18 条件下, 喷流法向力和俯仰力矩的控制效率均在无侧滑时最大, 并随着侧滑角的增大而减小。

参考文献（References）

[1] 朱广生, 刘文伶. 高超声速轴对称再入机动飞行器气动外形设计与布局研究[J]. 空气动力学学报, 2016, 34(3): 327-332.
ZHU G S, LIU W L. Design for aerodynamic configuration of a hypersonic axisymmetric maneuverable reentry vehicle[J]. Acta Aerodynamica Sinica, 2016, 34(3): 327-332(in Chinese).

[2] 李素循. 近空间飞行器的气动复合控制原理及研究进展[J]. 力学进展, 2009, 39(6): 740-755.
LI S X. Progress in aerodynamics of combination control for vehicles at high speed[J]. Advances in Mechanics, 2009, 39(6): 740-755(in Chinese).

[3] 许晨豪, 蒋崇文, 高振勋, 等. 高超声速飞行器反作用控制系统喷流干扰综述[J]. 力学与实践, 2014, 36(2): 147-155.
XU C H, JIANG C W, GAO Z X, et al. The jet interaction effects of reaction control systems in hypersonic vehicles[J]. Mechanics in Engineering, 2014, 36(2): 147-155(in Chinese).

[4] ROGER R. The aerodynamics of jet thruster control for supersonic/hypersonic endo-interceptors: lessons learned: AIAA 1999-804[R]. Reston: AIAA, 1999.

[5] KUMAR D, STOLLERY J L, SMITH A J. Hypersonic jet control effectiveness[J]. Shock Waves, 1997, 7(1): 1-12.

[6] BRANDEIS J, GILL J. Experimental investigation of side-jet steering for missiles at supersonic and hypersonic speeds: AIAA 1995-316[R]. Reston: AIAA, 1995.

[7] BRANDEIS J, GILL J. Experimental investigation of side-jet steering for supersonic and hypersonic missiles[J]. Journal of Spacecraft and Rockets, 1996, 33(3): 346-352.

[8] BRANDEIS J, GILL J. Experimental investigation of super- and hypersonic jet interaction on configurations with lifting surfaces: AIAA 1997-3723[R]. Reston: AIAA, 1997.

[9] KURITA M, INOUE T, NAKAMURA Y. Aerodynamic interaction due to side jet from a blunted cone in hypersonic flow: AIAA 2000-4518[R]. Reston: AIAA, 2000.

[10] 王志坚, 伍贻兆, 徐翔, 等. 冷/热喷流对飞行器气动特性干扰实验研究[J]. 南京航空航天大学学报, 2009, 41(4): 429-431.
WANG Z J, WU Y Z, XU X, et al. Investigation of lateral jet interaction test for different media[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2009, 41(4): 429-431(in Chinese).

[11] GNEMMI P, SCHÄFER H. Experimental and numerical investigations of a transverse jet interaction on a missile body: AIAA 2005-52[R]. Reston: AIAA, 2005.

[12] 赵法明, 王江峰. 带舵旋成体侧向喷流流场特性分析[J]. 航空动力学报, 2016, 31(3): 726-732.
ZHAO F M, WANG J F. Analysis on flow field characteristics of lateral jet on slender body[J]. Journal of Aerospace Power, 2016, 31(3): 726-732(in Chinese).

[13] 赵弘睿, 龚安龙, 刘周, 等. 高空侧向喷流干扰效应数值研究[J]. 空气动力学学报, 2020, 38(5): 996-1003.
ZHAO H R, GONG A L, LIU Z, et al. Numerical study of lateral jet interaction at high altitude[J]. Acta Aerodynamica Sinica, 2020, 38(5): 996-1003(in Chinese).

[14] 梁杰, 阎超, 李志辉, 等. 稀薄过渡流区横向喷流干扰效应数值模拟研究[J]. 空气动力学学报, 2013, 31(1): 27-33.
LIANG J, YAN C, LI Z H, et al. Numerical investigation of lateral jet interaction effects in rarefied transition flow regime[J]. Acta Aerodynamica Sinica, 2013, 31(1): 27-33(in Chinese).

[15] WILLIAMS N J, MOELLER T M, THOMPSON R J. Numerical simulations of high frequency transverse pulsed jet injection into a supersonic crossflow[J]. Aerospace Science and Technology, 2020, 103: 105908.

[16] 杨建龙, 刘猛, 阿嵘. 高超声速热化学非平衡对气动热环境影响[J]. 北京航空航天大学学报, 2017, 43(10): 2063-2072.
YANG J L, LIU M, A R. Influence of hypersonic thermo-chemical non-equilibrium on aerodynamic thermal environments[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(10): 2063-2072(in Chinese).

[17] MENTER F R. Two-equation eddy-viscosity turbulence models for engineering applications[J]. AIAA Journal, 1994, 32(8): 1598-1605.

[18] 王泽江, 李杰, 曾学军, 等. 逆向喷流对双锥导弹外形减阻特性的影响[J]. 航空学报, 2020, 41(12): 124116.
WANG Z J, LI J, ZENG X J, et al. Effect of opposing jet on drag reduction characteristics of double-cone missile shape[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(12): 124116(in Chinese).

[19] GRANDHI R K, ROY A. Effectiveness of a reaction control system jet in a supersonic crossflow[J]. Journal of Spacecraft and Rockets, 2017, 54(4): 883-891.

[20] SRIVASTAVA B. Computation and validation of asymmetric force and moment coefficients for lateral jet controlled supersonic missile: AIAA 1998-2410[R]. Reston: AIAA, 1998.

[21] SRIVASTAVA B. Lateral jet control of a supersonic missile: computational and experimental comparisons[J]. Journal of Spacecraft and Rockets, 1998, 35(2): 140-146.

[22] STAHL B, ESCH H, GÜLHAN A. Experimental investigation of side jet interaction with a supersonic cross flow[J]. Aerospace Science and Technology, 2008, 12(4): 269-275.

Numerical study on lateral jet control efficiency of a hypersonic re-entry double-cone vehicle

LEI Juanmian*, ZHU Peiyu

(School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China)

Abstract: Lateral jet control is an effective means for vehicle attitude and trajectory control. Based on the three-dimensional Reynolds-averaged Navier-Stokes (RANS) equations, a numerical simulation study of the flow field surrounding the lateral jets of re-entry double-cone vehicles was carried out in order to investigate the variation of jet control efficiency of hypersonic re-entry double-cone vehicles with flight conditions and flight attitude angles. The study analyzed the influence of flight Mach number, altitude, angle of attack, and sideslip angle on jet control efficiency and revealed the corresponding flow mechanisms. The results indicate that increasing the flight Mach number or reducing the flight altitude can significantly enhance the control efficiency of normal force and pitch moment during lateral jet control. As the flight altitude increases, the control efficiency gradually approaches 1. The interaction between the lateral jet and the incoming flow causes a large area of jet gas to act on the vehicle's surface when the angle of attack are −20° and −30°. This results in a negative control efficiency of the jet normal force but an improvement in the control efficiency of the jet pitch moment. As the angle of attack changes towards positive values, the jet normal force control efficiency gradually becomes positive and reaches its optimum at 10°, while the jet pitch moment control efficiency approaches 1. Changes in flight sideslip angle have a relatively small impact on jet control efficiency.

Keywords: lateral jet; re-entry double-cone vehicle; control efficiency; hypersonic; numerical simulation

Received: 2023-11-02; **Accepted**: 2023-12-11; **Published Online**: 2024-01-03 17:38

URL: link.cnki.net/urlid/11.2625.V.20240102.1552.003

* **Corresponding author**. E-mail: leijm@bit.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0717

一种融合MSC和时空双重注意力的TCN航迹预测方法

焦卫东*，杨蓓

(中国民航大学 天津市智能信号与图像处理重点实验室，天津 300300)

摘　　要：针对现有航迹预测模型难以有效提取多尺度时空特征，导致预测精度受限的问题，提出一种融合多尺度卷积(MSC)网络和时空双重注意力(STDA)的时间卷积网络(TCN)航迹预测方法MDAT-Net。MDAT-Net模型包含MSAT、MTAT航迹预测模块及投票模块。在航迹预测模块中，为优化传统时间卷积网络中卷积核大小固定的问题，搭建由不同尺度卷积核组成的多尺度卷积架构，以更好地提取不同尺度的时空特征；为动态挖掘隐藏特征与目标特征之间的潜在相关性，引入空间注意力机制及时间注意力机制来自适应地关注重要信息，跳过次要信息。在投票模块中，决定各维度使用合适模块进行预测，以融合MSAT、MTAT航迹预测模块的优点，实现高精度的航迹预测。实验结果表明：MDAT-Net模型的均方根误差(RMSE)的改进率最高可达83.33%，平均绝对误差(MAE)的改进率最高可达85.85%，具有较高的准确性和鲁棒性。

关　键　词：航迹预测；多尺度卷积；时间卷积网络；注意力机制；时空特征

中图分类号：V355; TP391.9

文献标志码：A　　**文章编号**：1001-5965(2026)01-0015-13

随着空中交通流量的增加，航班延误、空域拥堵等一系列安全问题日益突出，这对现有的空中交通管制(air traffic control, ATC)系统提出了严峻的挑战。鉴于此，国际民用航空组织正在进行新一轮的ATC系统升级。其中，基于航迹的运行(trajectory based operation, TBO)模式作为下一阶段的核心技术，其以精确的4D航迹预测为基础。航迹预测指通过使用飞机当前的位置、速度、航向等信息，预测未来一段时间内飞机的位置和轨迹。精确的航迹预测是优化空中交通管制、减少航空拥堵、提高空中交通效率的重要手段。对于飞行员和空中交通管制人员而言，精确的航迹预测是确保航空运输系统安全高效运行的关键因素之一。

目前，航迹预测模型大致分为3类：动力学模型、状态估计模型及机器学习模型[1]。

利用动力学模型[2-3]进行航迹预测，需考虑航空器的性能参数、飞行意图、运动状态等，对航空器进行动力学建模以完成对目标航空器的航迹预测，但该方法往往需要依赖大量的先验条件。利用状态估计模型[4-5]进行航迹预测，需通过运动方程建立状态方程的状态转移矩阵，过程中不涉及航空器性能参数，而是研究未来时间点的位置与历史时间点的位置、航向角、加速度等状态之间的关系，但通过该方法搭建的预测模型往往难以捕捉航迹数据中蕴含的复杂非线性关系。

近年来，机器学习迅猛发展，为航迹预测提供了新方法。利用机器学习模型进行航迹预测，是将航迹预测问题转化为时间序列预测问题，通过挖掘航迹训练集中航迹点之间的关系来进行建模。目前，航迹预测的经典神经网络模型有长短期记忆(long short-term memory, LSTM)网络[6-10]、门控循环单元(gate recurrent unit, GRU)等循环神经网络

收稿日期：2023-11-02；**录用日期**：2024-02-07；**网络出版时间**：2024-03-08 10:37
网络出版地址：link.cnki.net/urlid/11.2625.V.20240305.1821.006
基金项目：国家重点研发计划(2020YFB1600101)
*通信作者. E-mail：nxjiaowd@sina.com

引用格式：焦卫东，杨蓓. 一种融合MSC和时空双重注意力的TCN航迹预测方法[J]. 北京航空航天大学学报，2026，52(1)：15-27.
JIAO W D，YANG B. A TCN trajectory prediction method fusing MSC and spatio-temporal dual attention[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：15-27 (in Chinese).

(recurrent neural network, RNN)[11-13]变体。然而,RNN在处理时间序列时不能并行处理数据,下一个时间步需要等上一个时间步处理完毕后才能被处理,这种依赖性导致了计算效率低下。此外,RNN还存在梯度消失或爆炸、信息丢失或混淆、难以有效捕捉多尺度时空信息等缺陷,使得挖掘时间序列数据中的长期依赖关系变得困难。近年来,深度学习领域出现了一种用于处理时序问题的新方法,即时间卷积网络(temporal convolutional network, TCN)。与RNN相比,TCN在权值共享和卷积局部感知方面具有显著优势,并且不易出现梯度爆炸问题。TCN不仅可以学习输入时序的长期历史时间相关性,还能像卷积神经网络(convolutional neural network, CNN)一样进行并行计算,在机器翻译、音频合成等序列建模任务中表现出极佳的效果[14]。Huang等[15]提出了一种基于融合注意力机制、TCN及双向GRU的航迹预测模型DSA-TCN-BiGRU,并使用贝叶斯算法对模型超参数进行优化以获得最佳性能。Lin等[16]对传统TCN进行改进,并将改进后的TCN与GRU融合,对船舶轨迹进行预测。Zhang等[17]对鲸鱼优化算法(whale optimization algorithm, WOA)进行改进,将线性收敛因子修改为非线性收敛因子,利用改进WOA-TCN-Attention模型进行船舶轨迹预测。Zhang等[18]使用Douglas-Peucker算法处理各轨迹段以简化和压缩轨迹数据,利用聚类算法对轨迹数据进行聚类,使用TCN进行轨迹预测。Zhou等[19]通过TCN-GRU模型准确预测共享单车需求,为区域共享单车智能调度提供决策依据。尽管上述方法在航迹预测方面表现出一定的有效性,但未能充分提取时间数据的多尺度时空特征信息,导致预测精度受限。

为充分提取多尺度时空特征信息,实现高精度的航迹预测,本文在TCN的基础上融合多尺度卷积(multi-scale convolution, MSC)网络和时空双重注意力(spatio-temporal dual attention, STDA),提出一种新的基于MDAT-Net的航迹预测模型。该模型由3个模块组成,分别为MSAT、MTAT航迹预测模块及投票模块。其中,MSAT表示融合MSC网络、空间注意力(spatial attention, SA)机制和TCN的混合网络模型,MTAT表示融合MSC网络、时间注意力(temporal attention, TA)机制和TCN的混合网络模型。在航迹预测模块中,首先,利用MSC网络提取航迹数据不同尺度的时空特征,并进行有效融合;其次,使用TCN进一步捕捉时序数据之间的长期依赖关系;然后,分别结合空间注意力机制及时间注意力机制赋予输入序列不同的关注程度,给予对预测结果影响较大的特征更大的权重。在投票模块中,对MSAT及MTAT输出各维度的预测误差进行比较,利用多数投票法决定各维度使用合适模型进行预测,以此融合2个航迹预测模块的优点。为验证模型性能,将本文模型与GRU、LSTM网络、LSTM-AT、CNN-LSTM在广播式自动相关监视(automatic dependent surveillance-broadcast, ADS-B)数据集上进行对比实验及消融实验,实验结果表明,MDAT-Net模型具有较高的预测精度。

1 航迹描述和数据处理

1.1 ADS-B及航迹描述

ADS-B是一种航空交通系统中用于航空器位置报告的技术,该技术产生的海量数据蕴含着重要的飞行信息,动态数据暗示了飞机未来的运动趋势。ADS-B报文将每架飞行器在飞行过程中某一天的轨迹信息以一系列离散轨迹点的形式返回,汇总为每架航空器的轨迹数据,设D表示n条航空器的轨迹集合:

$$D=\{D_1,D_2,\cdots,D_i,\cdots,D_n\} \tag{1}$$

式中:D_i为第i(i=1,2,$\cdots$,n)次航班的轨迹,由Q个按照时间顺序排列的离散多维轨迹点$\boldsymbol{p}_{ij}$构成,即

$$D_i=\{\boldsymbol{p}_{i1},\boldsymbol{p}_{i2},\cdots,\boldsymbol{p}_{ij},\cdots,\boldsymbol{p}_{iQ}\} \tag{2}$$

而每个航迹点又包含多维特征:纬度A_{ij}、经度J_{ij}、高度G_{ij}和时间S_{ij},即

$$\boldsymbol{p}_{ij}=[A_{ij},J_{ij},G_{ij},S_{ij}] \tag{3}$$

式中:j=1,2,$\cdots$,Q。

1.2 数据处理

1.2.1 数据清洗与重构

由于通信链路延迟、数据包重传或传输错误等因素,ADS-B数据往往存在数据缺失、数据重复等缺陷,应先对数据进行预处理,以确保可用训练数据正确。

首先,剔除ADS-B数据中时间重复和空间重复(指经度、纬度、高度同时重复)的数据点;然后,使用三次样条插值算法以1 s为单位对数据进行插值处理。经插值处理后的航迹变得完整且平滑,保留了原始轨迹的运动趋势。

1.2.2 数据归一化

为消除不同数据量纲对结果的影响,需要将输入数据归一化到[0,1]区间内。本文采用min-max方法对数据进行归一化处理:

$$X_i'=\frac{X_i-X_{\min}}{X_{\max}-X_{\min}} \tag{4}$$

式中:$X_{\max}$为各维度的最大值;$X_{\min}$为各维度的最小值;X_i为归一化前的数据;X_i'为归一化后的数据。经归一化处理后的数据并未破坏其内在的相关性。在使用归一化数据进行训练及预测后,再对预

测值进行反归一化，便可得到符合实际范围和意义的预测航迹。

2　MDAT-Net 航迹预测模型

本文所提 MDAT-Net 模型融合了 2 个航迹预测模块的优势，能够充分提取多尺度时空特征，实现高精度的航迹预测。本文提出的航迹预测模型如图 1 所示，该模型包含 2 个航迹预测模块及 1 个投票模块，其中，航迹预测模块包括 MSAT 预测和 MTAT 预测 2 个子模块。

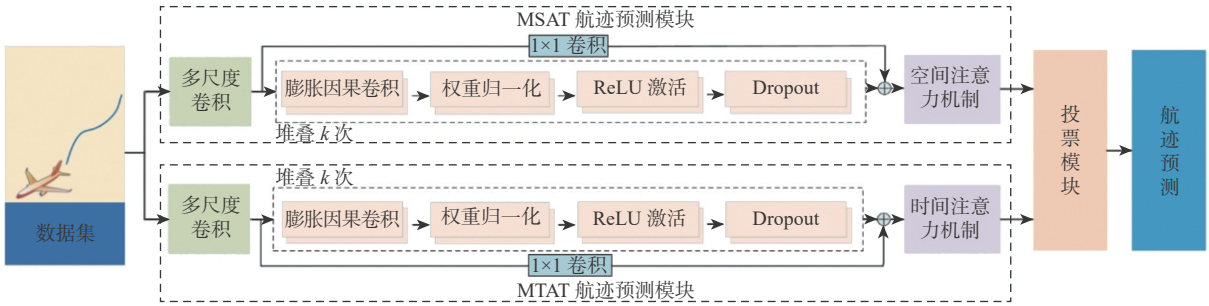


图 1　MDAT-Net 航迹预测模型

Fig. 1　MDAT-Net trajectory prediction model

2.1　航迹预测模块

航迹数据是具有丰富时空特征的多维时间序列，每个变量都呈现出长短期的时间变化模式，并且各变量之间还存在着相互依赖的空间信息。航迹预测的准确性取决于模型对航迹序列特征的提取能力，因此，本文提出一种能够充分提取航迹数据多尺度时空特征且实现高精度航迹预测的模型。

考虑到 TCN 在处理时间序列问题上具有一定的优势，将其作为提取航迹数据长期依赖关系的基础组件。首先，将 MSC 网络与 TCN 进行融合，以优化传统 TCN 因卷积核大小固定而难以提取不同尺度时空特征的问题，降低从输入航迹序列中提取多样化时空特征的难度。MSC 网络利用并行计算的方式在不同尺度上对航迹数据进行卷积操作，可以同时处理不同时间步的信息，捕获 ADS-B 数据不同尺度的时空特征。然后，搭建由不同膨胀系数的残差块堆叠而成的 TCN，TCN 利用膨胀卷积可以在相同尺度下提取不相邻数据的时空特征，扩大了神经网络对航迹序列的感知范围。这意味着 TCN 能够捕捉到航迹序列中远处时刻的相关性，对于航迹预测来说是有帮助的，因为航空轨迹通常具有长期依赖性，远处时刻的信息对于未来轨迹的预测是重要的。MSC 网络与 TCN 的融合可以充分提取多尺度时空特征，挖掘 ADS-B 数据蕴含的运动规律，从而提高预测的准确性。此外，传统的 TCN 在提取输入序列时未体现差异化倾向，可能会导致对重要信息学习不足，而对一些不重要的信息进行过度学习，从而影响预测效果。为解决这一问题，本文提出在级联 MSC-TCN 网络的基础上引入空间注意力机制及时间注意力机制，并分别命名为 MSAT 和 MTAT。空间注意力机制提取各特征之间的空间相关性，而时间注意力机制捕获窗口内所有时间步长之间的依赖关系，从而提高网络的预测精度。最后，通过输出层得到预测结果。航迹预测模块的结构如图 2 所示，包括输入层、MSC 层、TCN 层、空间注意力机制层或时间注意力机制层及输出层。

2.1.1　输入层

航迹预测属于有监督学习，需将数据划分为样本和标签后输入到模型中。使用滑动窗口方法对航迹序列进行分割，既保持了时间序列中相邻状态的连续性，又充分挖掘了每个窗口中相邻点的动态依赖关系，其结构如图 3 所示。图中：L 为时间窗口大小，q 为预测步长，即长度为 L 的航迹数据经过预处理后输入到航迹预测模块中，可用 $\boldsymbol{P}=[\boldsymbol{p}_{i1},\cdots,\boldsymbol{p}_{it},\cdots,\boldsymbol{p}_{iL}]^{\mathrm{T}}$ 表示，以预测未来航迹点 $\boldsymbol{p}'_{i(L+q)}$。持续向后移动时间窗口进行分割操作，直到到达最后一个轨迹点。

2.1.2　MSC 层

输入层以滑动窗口的方式将航迹数据分割成多个子序列，传递给 MSC 网络。每个子序列代表一个时序的局部区域，通过对这些局部区域进行不同尺度的卷积操作，能够捕获航迹序列的长期和短期趋势。本文构建了由 5 个不同尺度卷积核组成的 MSC 架构。鉴于航迹数据的特点，将各卷积层设计为一维卷积，并选取 ReLU 激活函数进行激活。经过 MSC 处理后的航迹数据被映射到隐层特征空间，使用 Concatenate 操作来有效融合各输出。MSC 层的输出特征向量 $\boldsymbol{H}_{\mathrm{c}}$ 可表示为

$$\boldsymbol{C}_i=f(\boldsymbol{P}\otimes\boldsymbol{W}_i+\boldsymbol{b}_i)=\mathrm{ReLU}(\boldsymbol{P}\otimes\boldsymbol{W}_i+\boldsymbol{b}_i)\tag{5}$$

$$\boldsymbol{H}_{\mathrm{c}}=\mathrm{Concatenate}(\boldsymbol{C}_1,\boldsymbol{C}_2,\boldsymbol{C}_3,\boldsymbol{C}_4,\boldsymbol{C}_5)\tag{6}$$

式中：$\boldsymbol{W}_i$ 为权重矩阵；$\boldsymbol{b}_i$ 为偏差；$\boldsymbol{C}_i$ 为卷积层的输出。

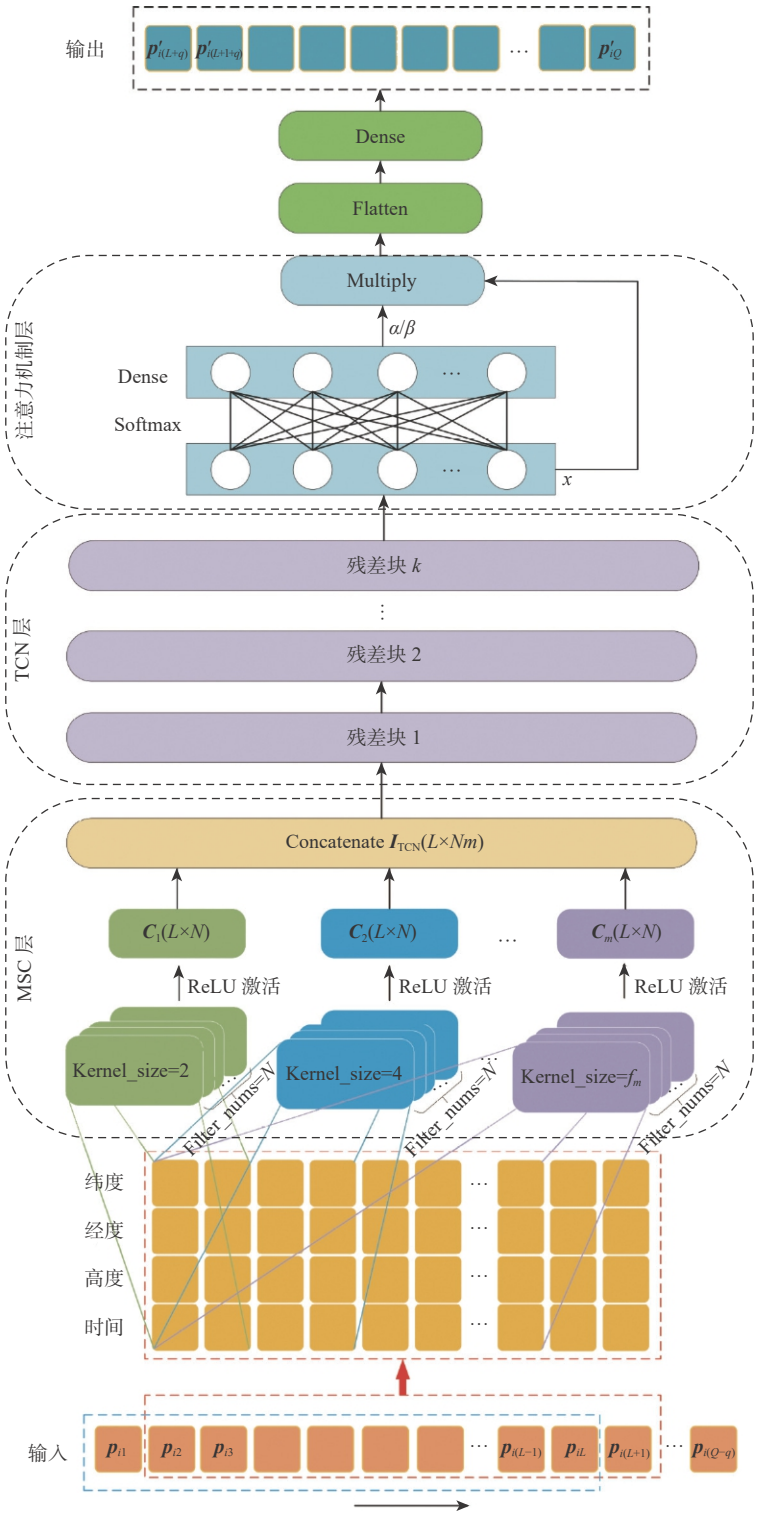


图 2 航迹预测模块结构

Fig. 2 Structure of trajectory prediction module

2.1.3 TCN 层

TCN 由因果卷积、膨胀卷积和残差模块组成。因果卷积是指输出序列的第 t 个航迹点 $\boldsymbol{y}_t$ 仅与当前及之前航迹点 $\boldsymbol{s}_t$, $\boldsymbol{s}_{t-1}$, $\boldsymbol{s}_{t-2}$, …, $\boldsymbol{s}_0$ 有关，而与未来的航迹点 $\boldsymbol{s}_{t+1}$ 无关。因此，因果卷积是单向结构，不能看到未来数据，这与航迹序列的特点相符。其可以表示为

$$\boldsymbol{y}_t = h(\boldsymbol{s}_0, \boldsymbol{s}_1, \cdots, \boldsymbol{s}_t) \tag{7}$$

式中：$h(\cdot)$ 为卷积操作。

因果卷积的感受野与网络深度线性相关，意味着处理更长的时间序列需要更深的网络结构。一个简单的因果卷积无法有效捕获较长时间序列的信息。因此，TCN 中采用膨胀卷积，膨胀卷积的结构如图 4 所示。图中：s 为卷积核大小。

此外，为解决梯度消失问题，TCN 的输出层中引入了残差模块，该模块由 2 层具有相同膨胀系数

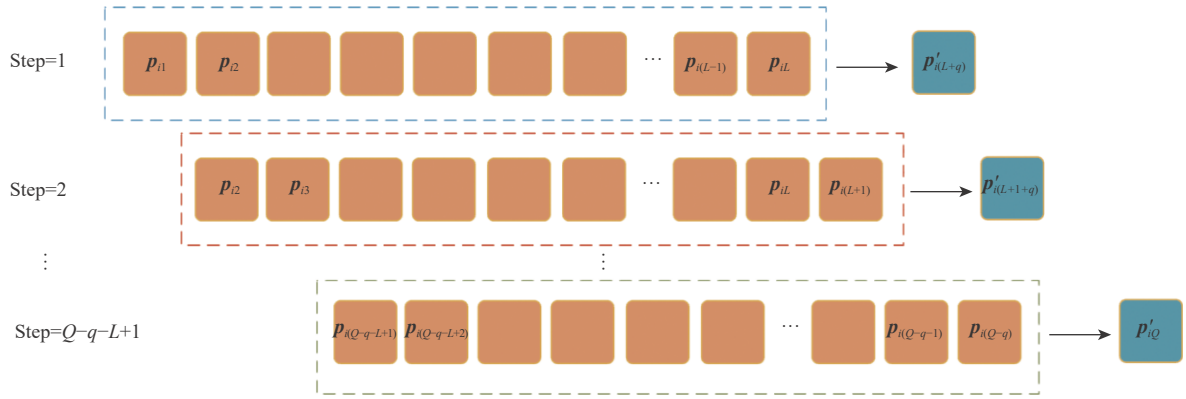


图 3　滑动时间窗口

Fig. 3　Sliding time window

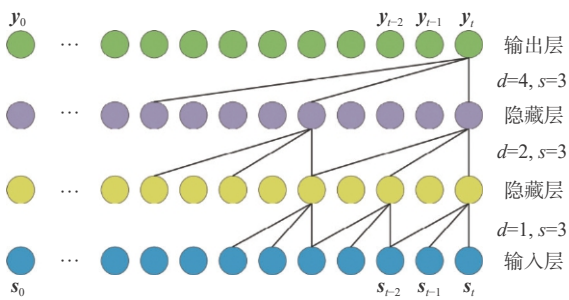


图 4　膨胀卷积结构

Fig. 4　Expansion convolution structure

的卷积和非线性恒等映射组成, 结构如图 5 所示。

使用 TCN 对 MSC 层提取的特征向量进行学习, 通过堆叠不同膨胀系数的残差模块, TCN 能够更深入地捕获航迹序列的长期依赖关系。TCN 层的输出记为

$$\boldsymbol{X}=\mathrm{TCN}(\boldsymbol{H}_{\mathrm{e}},\boldsymbol{W}_{\mathrm{TCN}})=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_t,\cdots,\boldsymbol{x}_L]^{\mathrm{T}}\in\mathbf{R}^{L\times M}\quad(8)$$

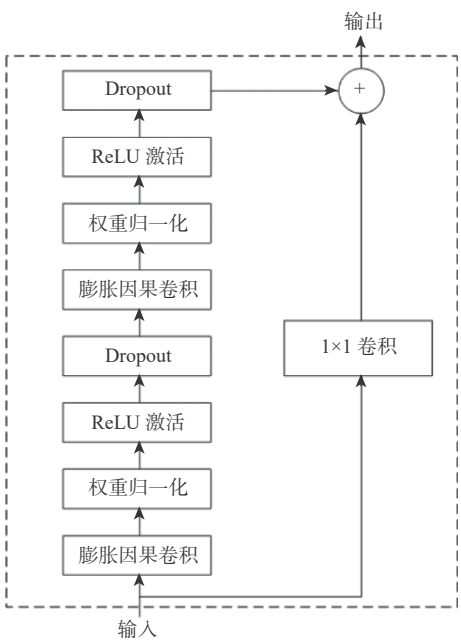


图 5　残差模块

Fig. 5　Residual block

式中: $\boldsymbol{W}_{\mathrm{TCN}}$ 为 TCN 编码计算的参数; TCN 层的输出长度为 L; M 为特征数量。

2.1.4　注意力机制层

TCN 对所有信息的提取未体现差异化倾向, 忽略了特征之间的相关性和时间差异。引入注意力机制能对 TCN 提取到的隐藏特征(如飞机的时间、空间位置、速度和加速度等信息)的高阶表达进行动态分配权重, 以此挖掘隐藏特征与目标特征之间的关联关系, 将注意力集中于重要特征, 减少甚至忽略次要特征, 从而改善模型性能。

空间注意力机制的结构如图 6 所示。以第 t 个航迹点为例, 含 M 个特征的输入特征向量为 $\boldsymbol{x}_t=[x_{1t},x_{2t},\cdots,x_{Mt}]$。通过单层神经网络映射得到注意力权重向量 $\boldsymbol{e}_t$:

$$\boldsymbol{e}_t=\boldsymbol{W}_{\mathrm{e}}\boldsymbol{x}_t+\boldsymbol{b}_{\mathrm{e}}\quad(9)$$

式中: $\boldsymbol{e}_t=[e_{1t},e_{2t},\cdots,e_{it},\cdots,e_{Mt}]$为第 t 个时间步各输入特征对应的注意力权重系数; $\boldsymbol{W}_{\mathrm{e}}$ 为可训练权重矩阵; $\boldsymbol{b}_{\mathrm{e}}$ 为计算空间注意力权重的偏置向量。

使用式 (10) 中的 Softmax 函数对各注意力权重系数做归一化处理, 得到各特征注意力权重 $\boldsymbol{\alpha}_t=[\alpha_{1t},\alpha_{2t},\cdots,\alpha_{it},\cdots,\alpha_{Mt}]$, 其中, α_{it} 表示第 i 个特征的注意力权重值:

$$\alpha_{it}=\exp(e_{it})\Big/\sum_{i=1}^{M}\exp(e_{it})\quad(10)$$

且有 $\sum_{i=1}^{M}\alpha_{it}=1$。使用 $\boldsymbol{\alpha}_t$ 将输入特征向量 $\boldsymbol{x}_t$ 重新计算为加权向量 $\boldsymbol{x}'_t$, 即注意力机制层在 t 时刻的输出。注意力机制层的输出记为 $\boldsymbol{X}'=[\boldsymbol{x}'_1,\cdots,\boldsymbol{x}'_t,\cdots,\boldsymbol{x}'_L]^{\mathrm{T}}\in\mathbf{R}^{L\times M}$, 其中,

$$\boldsymbol{x}'_t=\boldsymbol{\alpha}_t\odot\boldsymbol{x}_t=[\alpha_{1t}x_{1t}\quad\alpha_{2t}x_{2t}\quad\cdots\quad\alpha_{Mt}x_{Mt}]\quad(11)$$

式中: $\odot$表示哈达玛积。

时间注意力机制与空间注意力机制的原理类似, 但其输入采用 $\boldsymbol{x}_i=[x_{i1},x_{i2},\cdots,x_{iL}]^{\mathrm{T}}$的形式。其中, i 为第 i 个空间特征。同样, 原始输入经过类似

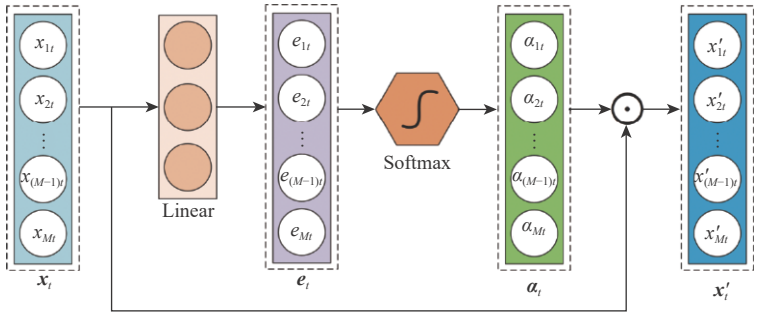


图6 空间注意力机制

Fig. 6 Spatial attention mechanism

式(9)～式(11)的线性变换后得以对各时间信息进行动态加权。

2.1.5 输出层

模块的输出层接收来自注意力机制层的输出作为输入。使用Flatten操作将序列展平，通过Dense层进行特征映射，以依次输出各预测航迹点。预测公式可表示为

$$\boldsymbol{p}'_{ij}=\text{linear}(\boldsymbol{W}_{\text{o}}\boldsymbol{X}'+\boldsymbol{b}_{\text{o}}) \tag{12}$$

式中：$\boldsymbol{W}_{\text{o}}$为权重矩阵；$\boldsymbol{b}_{\text{o}}$为偏差向量。

2.2 投票模块

多数投票法方便且易于理解[20]，在投票模块中，使用该方法对MSAT和MTAT这2个航迹预测模块的各维输出进行投票。具体步骤为：同时训练MSAT和MTAT航迹预测模块，且为了确保投票结果的有效性，需在奇数次实验中进行投票。每次实验均计算2个航迹预测模块各维度输出的均方根误差(root mean square error, RMSE)，选取RMSE较低的模块在输出维度中获得一票。由此通过投票法选择输出各维度使用合适模块进行预测，将最佳模块应用于测试样本，以进行航迹预测。

2.3 航迹预测方案设计

在MDAT-Net模型的基础上，结合数据样本划分和模型测试环节，可得到完整的航迹预测流程，如图7所示。具体步骤如下：

1）使用滑动时间窗将预处理后的航迹数据划分为训练集、验证集和测试集。

2）将训练集输入到MSAT及MTAT这2个航迹预测模块中进行训练，利用Adam优化算法更新参数，使得损失函数值最小。当训练次数达到设定的epoch阈值时，本次实验结束。

3）在投票模块中，比较每次实验中2个航迹预测模块各输出维度的RMSE。RMSE较低的一方在输出维度中获得一票。若实验次数未达到阈值，重复步骤2)；当实验次数达到阈值时，则显示各维度选择合适模块进行预测的结果，并保存最佳模块。

4）将测试集数据样本分别输入到保存的最佳模块中。所设置的实验次数为T，根据投票模块结果，使用MSAT航迹预测模块预测经纬度，使用MTAT航迹预测模块预测高度，以完成航迹预测。

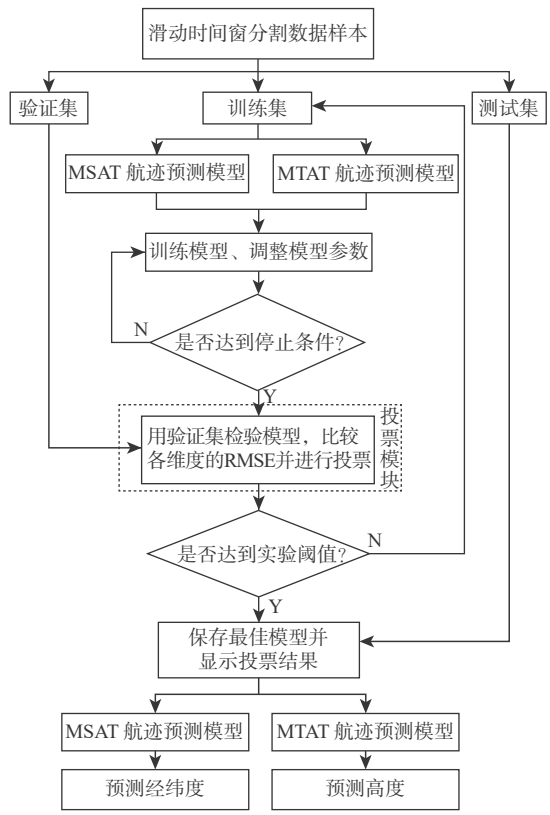


图7 航迹预测流程

Fig. 7 Trajectory prediction flow

3 仿真实验及结果分析

为验证MDAT-Net模型的预测效果，选取广州白云国际机场飞往天津滨海国际机场的CSN3301航班的30条ADS-B数据，每条轨迹包含约1 450个多维轨迹点。该数据的地理范围限定在经度116.23°～117.45°和纬度38.64°～39.29°之间，所使用的特征为经度、纬度、高度及时间间隔。这段数据集以天数为单位按照6∶2∶2的比例划分为训练集、验证集和测试集，以确保充分评估模型的性能。所有实验均在Keras深度学习框架进行模型训

练和验证，硬件环境为一台搭载 Intel CPU i5-10400 2.9 GHz 处理器和 32 GB 内存的计算机。

3.1　MDAT-Net 模型与基线模型对比实验

合理的超参数配置在模型训练中至关重要，直接影响模型的性能和泛化能力。根据 RMSE 和平均绝对误差（mean absolute error, MAE）这 2 个误差评价指标来调整模型的各项参数，包括卷积核尺寸、滤波器数目、TCN 的膨胀系数及学习率等，取训练结果最优时的参数。本文模型的部分参数优选结果如表 1 所示。

表 1　部分参数优选结果

Table 1　Results of some parameter preferences

参数	数值
多尺度卷积网络各层卷积核尺寸	[2,4,8,16,32]
多尺度卷积网络的滤波器个数	16
时间卷积网络的膨胀系数	[1,2,4,8,16]
时间卷积网络的卷积核尺寸	3
时间卷积网络的滤波器个数	64
学习率	0.001
迭代次数	500
一次训练所取样本数	128
投票模块实验次数T	17

滑动窗口大小 L 为 60，预测步长 q 为 10，即使用前 60 s 的航迹数据去预测航空器 10 s 后的空间位置。以测试集中一条航迹数据为例，进行预测结果及误差的可视化，如图 8 和图 9 所示。图 8 展示了本文模型的预测值与真实值之间的三维结果对比，可以清晰地看出，模型预测的三维航迹与航空器实际飞行吻合较好，误差较小。以具体时刻 13∶12∶25 的航迹点为例，真实的空间位置坐标为 [38.651 7°,116.356 4°,2 987.040 0 m]，而模型预测坐标为 [38.649 7°,116.354 4°,2 993.927 7 m]，这表明预

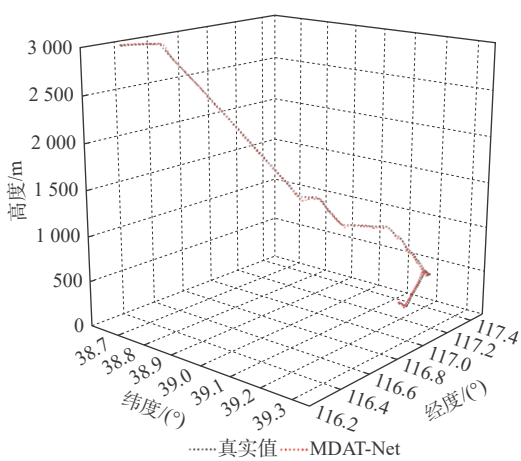


图 8　基于 MDAT-Net 的航迹预测结果

Fig. 8　MDAT-Net based trajectory prediction results

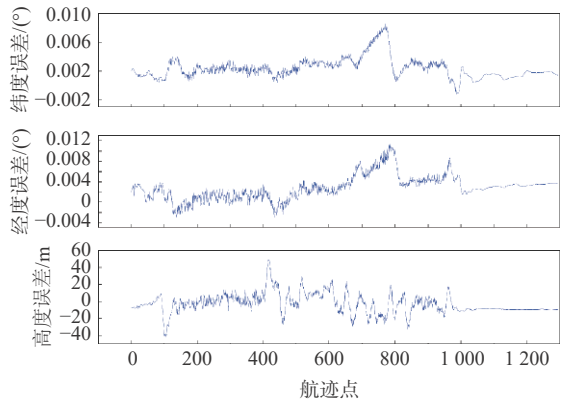


图 9　每个航迹点的纬度、经度和高度的预测误差结果

Fig. 9　Prediction error results for latitude, longitude and altitude at each track point

测结果与实际情况相符。图 9 展示了航迹中每个航迹点的纬度、经度和高度的预测误差结果，表明模型在不同空间维度的预测均取得了较小的误差，进一步验证了本文模型具有较高的准确性。其中，第 700～800 个航迹点的经纬度误差出现峰值，可能是受到复杂因素的影响，如航空交通管制、飞行计划变更等，但误差在 0.012°以内仍表现出较好的性能。

为检验 MDAT-Net 模型的泛化性能，将其与现有较为先进的航迹预测模型进行比较，包括 LSTM、GRU、LSTM-AT、CNN-LSTM，且所有模型均使用相同的 ADS-B 数据集进行评估。将滑动窗口大小 L 设置为 60，各基线模型中，LSTM 与 GRU 滤波器个数设置为 64，并分别用这些模型来预测未来 10 步、20 步、30 步、40 步后的航迹点空间位置。为定量分析预测结果，使用 RMSE 和 MAE 作为误差评价指标，以检测航迹预测的准确度。RMSE 和 MAE 的计算公式分别为

$$y_{\mathrm{RMSE}}=\sqrt{\frac{\sum_{i=1}^{z}(y_i-\hat{y}_i)^2}{z}} \tag{13}$$

$$y_{\mathrm{MAE}}=\frac{\sum_{i=1}^{z}|y_i-\hat{y}_i|}{z} \tag{14}$$

式中：y_i 为实际值；$\hat{y}_i$ 为预测值；z 为航迹点个数。

这 2 个评价指标的值越小，预测轨迹越接近实际轨迹，预测效果越好。各模型的预测误差结果分别如表 2 所示。为了直观地一次性比较不同模型在不同时间步上的性能表现，将各模型预测误差以箱线图的形式展示，如图 10 所示。

从表 2 可以看出，对于经典的时间序列预测模型 LSTM 和 GRU，两者的预测精度是相似的，但与其他预测模型相比，预测精度相对较低。LSTM-

表 2 各模型预测误差结果

Table 2 Prediction error results for each model

特征	预测步长	RMSE/(°)					MAE/(°)				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net	LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
纬度	10	0.009 3	0.008 2	0.009 2	0.007 6	0.003 4	0.007 9	0.006 8	0.005 9	0.007 1	0.002 7
	20	0.017 4	0.016 7	0.012 1	0.010 2	0.006 8	0.015 9	0.010 8	0.009 5	0.008 6	0.005 1
	30	0.018 7	0.018 1	0.015 0	0.013 9	0.010 1	0.016 8	0.014 5	0.012 9	0.012 4	0.008 2
	40	0.019 3	0.018 9	0.017 3	0.015 8	0.011 6	0.018 9	0.015 4	0.014 9	0.013 8	0.009 1

特征	预测步长	RMSE/(°)					MAE/(°)				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net	LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
经度	10	0.022 2	0.019 7	0.018 7	0.019 1	0.003 7	0.021 2	0.015 5	0.015 9	0.016 0	0.003 0
	20	0.023 5	0.020 8	0.020 9	0.021 0	0.007 7	0.022 3	0.017 7	0.017 1	0.016 4	0.006 3
	30	0.041 1	0.037 1	0.036 0	0.029 1	0.008 9	0.029 2	0.027 6	0.023 8	0.022 5	0.006 9
	40	0.047 8	0.043 1	0.042 4	0.036 2	0.015 7	0.035 4	0.031 8	0.031 1	0.030 4	0.013 5

特征	预测步长	RMSE/m					MAE/m				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net	LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
高度	10	39.546 2	31.104 1	29.662 7	29.083 5	14.144 5	27.314 4	23.293 8	21.094 9	18.724 8	10.251 1
	20	44.410 2	40.570 0	30.358 1	31.019 4	22.925 6	30.024 7	28.585 6	22.549 5	21.519 9	16.128 4
	30	56.391 0	48.472 2	44.869 9	41.141 2	34.584 9	38.261 7	32.417 1	31.863 6	26.305 5	22.199 4
	40	67.222 8	64.866 5	57.891 9	56.191 0	41.194 5	46.999 3	43.101 1	40.039 6	40.898 0	31.115 2

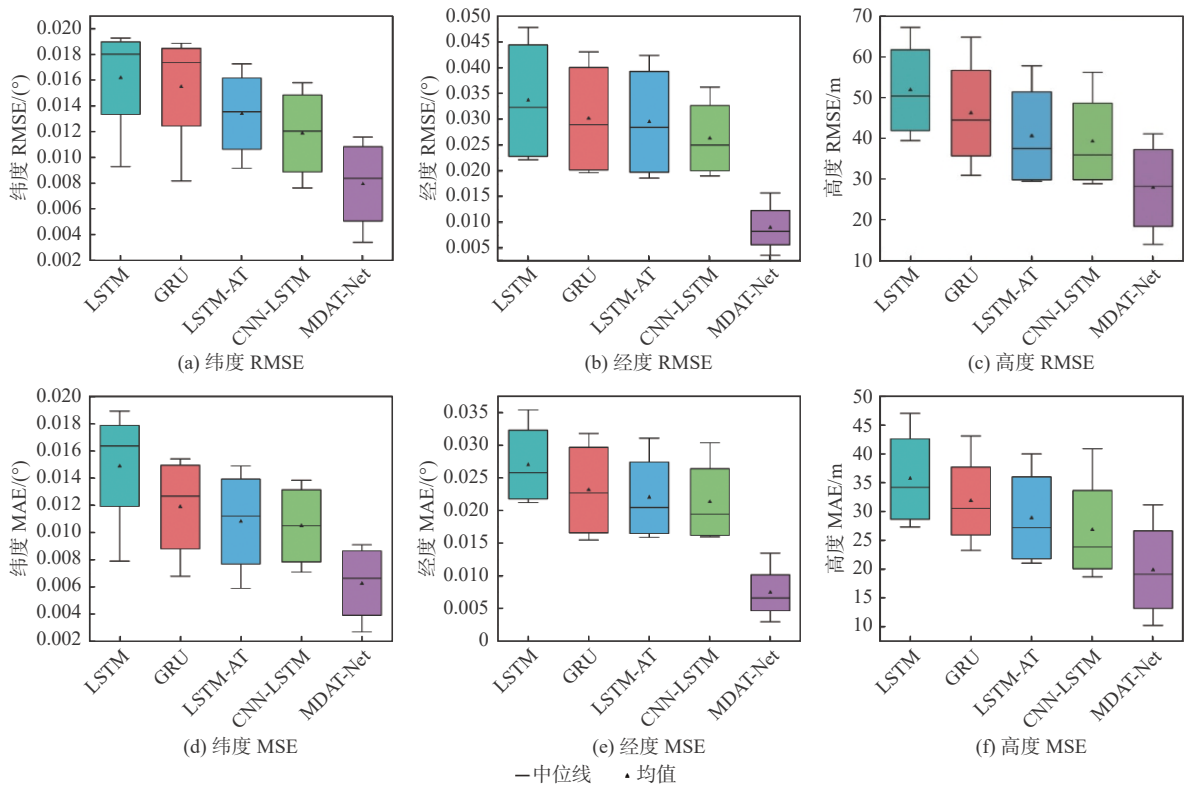


图 10 各模型预测误差箱线图

Fig. 10 Box plots of prediction errors for each model

AT 和 CNN-LSTM 是基于 LSTM 的混合模型。其中，LSTM-AT 模型是在 LSTM 基础上融合了注意力机制，通过注意力机制使得模型可以动态地给予隐藏层中各特征不同的权重，使得模型更关注重要特征，从而提高模型的预测性能。CNN-LSTM 模型使用 CNN 提取轨迹空间特征，LSTM 提取轨迹时间特征，相对于单一的 LSTM 模型，具有较高的预测精度。

实验结果表明，不论是预测 10 步后还是 40 步后的轨迹，MDAT-Net 模型的性能均优于所有基线模型。当使用前 60 个时间步的数据来预测未来

10 步后航迹点的空间位置时, MDAT-Net 模型的误差较小。具体而言, 其纬度、经度、高度的 RMSE 分别为 0.003 4°、0.003 7°、14.144 5 m, MAE 分别为 0.002 7°、0.003 0°、10.251 1 m。相对于误差较大的基线模型 LSTM, RMSE 的改进率分别达到了 63.44%、83.33%、64.23%, MAE 的改进率分别达到了 65.82%、85.85%、62.47%。当使用前 60 个时间步的数据来预测未来 40 步后航迹点的空间位置时, MDAT-Net 模型的纬度、经度、高度的 RMSE 分别为 0.011 6°、0.015 7°、41.194 5 m, MAE 分别为 0.009 1°、0.013 5°、31.115 2 m。相较于误差较大的基线模型 LSTM, RMSE 的改进率分别达到了 39.90%、67.15%、38.72%, MAE 的改进率分别达到了 51.85%、61.86%、33.80%。图 10 中, 每个箱体表示一个模型在不同时间步上的误差数据分布。箱体中间的一条线是数据的中位数, 代表样本数据的平均水平, 较低的中位数通常表示更准确的预测性能。上下四分位数分别对应箱体的上下边界, 较低的上下四分位数表示航迹预测模型的误差在上下边界方面相对较小。从图 10 可以看出, MDAT-Net 模型的中位数和上下四分位数都是最低的, 进一步表明该模型具有较高的预测准确度。本文模型与基线模型的三维、二维对比结果分别如图 11、图 12 所示。此外, 除了使用各种深度学习模型进行航迹预测, 本文还采用了经典传统方法, 即卡尔曼滤波进行航迹预测。其纬度、经度、高度的 RMSE 分别为 0.058 8°、0.094 2°、265.688 5 m, MAE 分别为 0.048 5°、0.083 0°、218.159 0 m。由此可见, 相对于传统方法, 深度学习方法具有显著优势。

为进一步验证本文方法的性能, 使用从成都天府国际机场飞往天津滨海国际机场的 CCA1428 航班的 ADS-B 数据来进行航迹预测实验。不同预测步长下, 各模型的预测误差结果以折线图的形式清

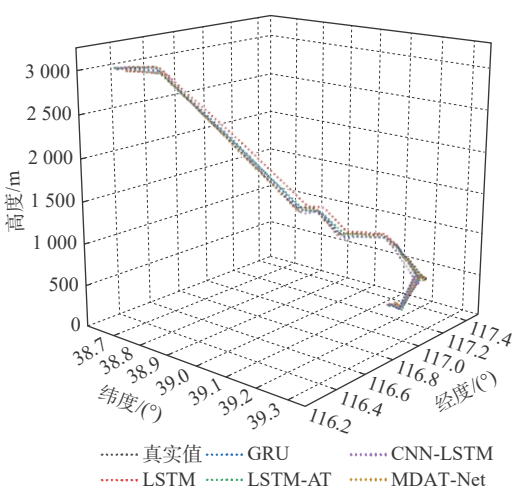


图 11　MDAT-Net 与基线模型的三维对比

Fig. 11　3D comparison of MDAT-Net with baseline model

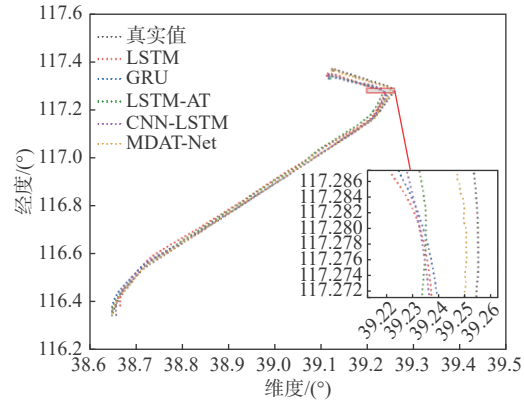


图 12　MDAT-Net 与基线模型的二维对比

Fig. 12　2D comparison of MDAT-Net with baseline model

晰展示, 如图 13 所示。可以看出, 在对其他航班的实际飞行数据进行预测时, MDAT-Net 模型的表现同样优于基线模型, 具有更高的准确性和鲁棒性。同时也能看出, 随着预测时间跨度的增加, 各模型的预测误差也随之增加。这是因为: 当预测时间跨度较小时, 未来航迹路径的不确定性较小; 而当预测时间跨度较大时, 意味着有更多的不确定性和可能的外部干扰。因此, 预测时间距离长的航迹路径通常有更大的误差。

除了预测时间跨度对方法精度的影响外, 时间窗口长短、滤波器个数等参数设置也会对模型的预测准确度产生影响。不同窗长的航迹点输入对预测模型的性能产生不同影响。窗口过短, 可能导致模型难以捕捉到航迹中的长期趋势和复杂关系; 窗口过长, 则可能引入冗余信息, 从而影响实时性和预测精度。滤波器个数同样对模型性能有显著影响。当滤波器个数过多时, 模型可能过度拟合训练数据而失去泛化能力; 反之, 如果滤波器个数太少, 模型可能无法捕捉到数据中的复杂模式和特征。在 q=10 且其他参数不变的情况下, 分别对时间窗口长度和滤波器个数 (指 LSTM/GRU/TCN 的滤波器个数) 进行变化, 相关分析的实验结果如表 3 和表 4 所示。可以看出, 在不同参数设置下, 各模型预测准确度也呈现差异, 因此, 合理选择时间窗口长短和滤波器个数等参数对模型的预测准确性至关重要。而 MDAT-Net 模型在不同参数设置下依然保持高度准确性和稳定的波动性, 展现出较强的鲁棒性。

综上所述, 基于 MDAT-Net 的航迹预测模型表现出良好的性能, 能够充分提取航迹数据的时空特征, 并准确反映航空器的实际飞行情况, 具有出色的泛化效果和较低的误差水平。

3.2　消融实验

在 3.1 节的实验中, 通过与已有航迹预测模型

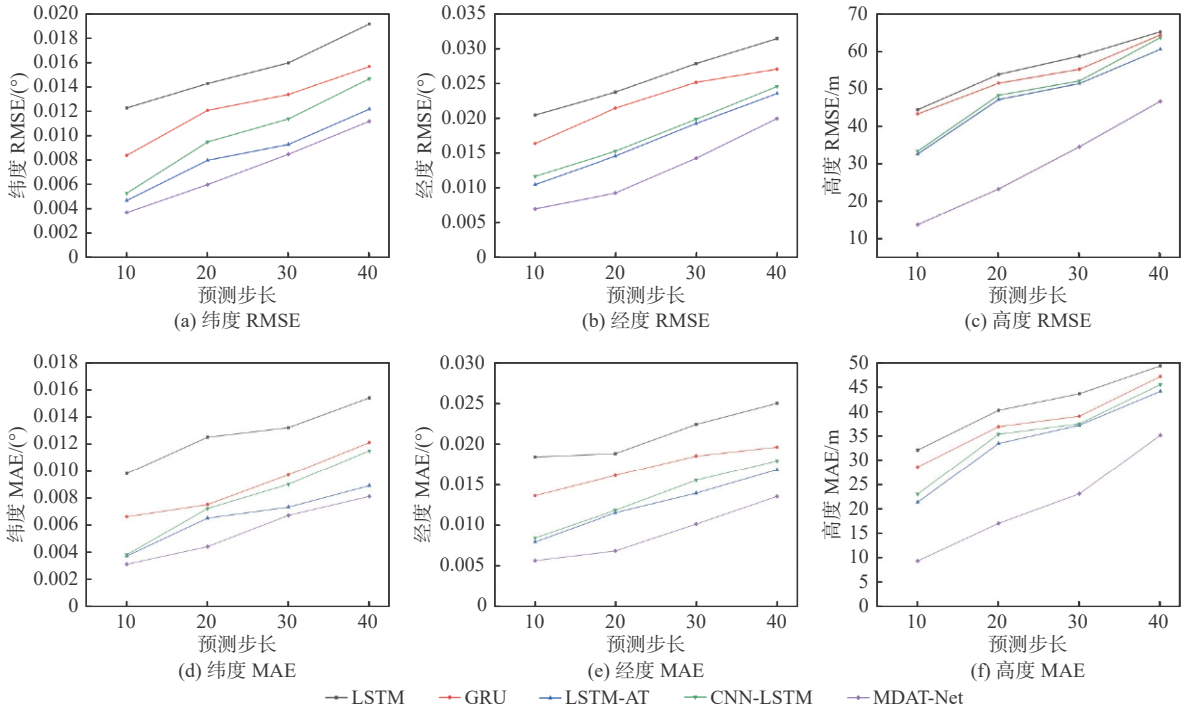


图 13 不同预测步长下各模型误差结果

Fig. 13 Error results of each model with different prediction step lengths

表 3 不同时间窗长下各模型误差结果

Table 3 Error results for each model with different time window lengths

特征	时间窗长	RMSE/(°)				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
纬度	50	0.016 7	0.009 6	0.006 5	0.007 2	0.004 6
	60	0.012 3	0.008 4	0.004 7	0.005 3	0.003 7
	70	0.012 8	0.010 8	0.004 5	0.006 3	0.004 2
特征	时间窗长	RMSE/(°)				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
经度	50	0.019 6	0.018 7	0.011 7	0.013 7	0.006 8
	60	0.020 5	0.016 4	0.010 5	0.011 7	0.007
	70	0.019 7	0.020 1	0.011 8	0.014 2	0.007 4
特征	时间窗长	RMSE/m				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
高度	50	56.448 1	50.135 4	44.870 2	31.397 9	15.792 9
	60	44.503 6	43.381 3	32.667 1	33.42	13.826 3
	70	61.693 9	49.963 5	45.484 9	46.110 2	16.392

的对比，验证了 MDAT-Net 模型的有效性。为了更深入地观察 MDAT-Net 模型中各模块的作用，本节进行了一系列的消融实验。在这些实验中，以 TCN 模型为基础模块，并分别移除不同的模块，构建了 TCN、TCN-SA、TCN-TA、MSC-TCN、MSAT、MTAT 及 MDAT-Net 模型。所有消融实验均在 L=60、q=10 的情况下进行，分别对模型训练 500 次，并计算测试集数据各维度的 RMSE 及 MAE，实验结果如表 5 所示。可以看出，一旦移除 MSC 模块或注意力模块，模型的预测精度都会有一定程度的下降，说明 MSC 模块和注意力模块对本文模型预测性能的提升起到了重要作用。

为更好地理解这些结果，进一步详细分析每个模块在 MDAT-Net 模型中的作用。TCN 作为基础模块，在没有其他扩展的情况下已经比 RNN 表现更好，这是因为 TCN 能够更有效地捕捉时间序列数据中的长期依赖关系。TCN-SA 模型及 TCN-TA 模型在 TCN 的基础上添加了注意力模块，该模块有助于模型更集中地关注关键信息，减少甚至忽略次要信息，从而提高了预测的准确性。MSC-

表 4　不同滤波器数量下各模型误差结果

Table 4　Error results for each model with different number of filters

特征	滤波器数	RMSE/(°)				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
纬度	32	0.010 2	0.011 3	0.005 7	0.007 3	0.004 2
	64	0.012 3	0.008 4	0.004 7	0.005 3	0.003 7
	128	0.015 9	0.008 8	0.004 4	0.006 8	0.004
特征	滤波器数	RMSE/(°)				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
经度	32	0.023 6	0.013 4	0.012 4	0.012 6	0.008 1
	64	0.020 5	0.016 4	0.010 5	0.011 7	0.007
	128	0.021 4	0.016 7	0.011 0	0.013 8	0.007 2
特征	滤波器数	RMSE/m				
		LSTM	GRU	LSTM-AT	CNN-LSTM	MDAT-Net
高度	32	67.627 0	56.949 9	42.498 4	37.440 1	14.178 5
	64	44.503 6	43.381 3	32.667 1	33.42	13.826 3
	128	51.114 5	41.985 4	38.705 6	40.161 6	15.054 9

表 5　消融实验结果

Table 5　Results of ablation experiments

模型	消除模块	纬度RMSE/(°)	经度RMSE/(°)	高度RMSE/m	纬度MAE/(°)	经度MAE/(°)	高度MAE/m
TCN	ALL	0.006 9	0.012 9	22.077 2	0.005 8	0.009 9	18.609 3
TCN-SA	MSC、TA	0.004 7	0.011 0	18.978 8	0.003 9	0.009 5	14.743 0
TCN-TA	MSC、SA	0.005 1	0.012 3	16.966 4	0.004 2	0.009 7	13.488 9
MSC-TCN	SA、TA	0.004 0	0.009 4	16.246 4	0.003 1	0.008 4	12.166 5
MSAT	MTAT	0.003 4	0.003 7	15.970 6	0.002 7	0.003 0	11.295 1
MTAT	MSAT	0.003 8	0.005 6	14.144 5	0.002 9	0.004 5	10.251 1
MDAT-Net		0.003 4	0.003 7	14.144 5	0.002 7	0.003 0	10.251 1

TCN 模型在 TCN 的基础上引入了 MSC 模块, 该模块可以有效处理不同时间尺度的信息, 从而提高了模型对多尺度数据的建模能力。MSAT 模块及 MTAT 模块分别在 MSC-TCN 基础上融合了空间注意力机制和时间注意力机制, 具备了较高的预测精确度。

总体来说, MDAT-Net 模型的成功建立在这些模块的协同作用之上。通过这些消融实验, 证明了每个模块对模型性能的贡献。建立复杂的模型不一定更好, 只有充分发挥各个模块的作用, 才能达到更好的预测效果。

4　结　论

本文融合 MSAT 和 MTAT 这 2 个航迹预测模块的优势, 提出了一种基于 MDAT-Net 的航迹预测模型, 克服了现有方法难以提取多尺度时空特征的局限性和不足, 实现了高精度的航迹预测。通过实验验证了所提模型的有效性, 主要结论如下:

1) MSC 网络与 TCN 的融合能够充分提取航迹数据多尺度时空特征, 挖掘 ADS-B 数据中蕴含的运动规律。时空注意力机制通过从 TCN 提取的特征动态分配权重, 增加对关键特征的关注度, 提高了预测的准确性。

2) 与现有航迹预测模型 LSTM、GRU、LSTM-AT 和 CNN-LSTM 相比, 实验结果表明, MDAT-Net 模型对航迹预测更为准确。相对于误差较大的基准模型 LSTM, MDAT-Net 模型的 RMSE 改进率最高可达 83.33%, MAE 的改进率最高可达 85.85%。

3) MDAT-Net 模型引入了投票模块, 采用多数投票法选择各维度使用合适模型进行预测, 以此融合 MSAT 和 MTAT 2 个航迹预测模块的优势, 能够充分发挥各个模块的作用。消融实验结果表明, 这种融合 MSAT 和 MTAT 2 个航迹预测模块优势的方法可进一步提升预测结果的准确性, 在处理航迹数据的时序性和空间特征方面表现更出色, 具有更高的预测准确率和较低的误差率。

未来可以对模型进一步优化, 考虑更多的影响因素, 以构建一个参数更少且精度更高的航迹预测模型。

参考文献（References）

[1] 徐正凤, 曾维理, 羊钊. 航空器轨迹预测技术研究综述[J]. 计算机工程与应用, 2021, 57(12): 65-74.
XU Z F, ZENG W L, YANG Z. Survey of civil aircraft trajectory prediction[J]. Computer Engineering and Applications, 2021, 57(12): 65-74(in Chinese).

[2] 汤新民, 韩云祥, 韩松臣. 基于混杂系统模型的航空器4D航迹推测[J]. 南京航空航天大学学报, 2012, 44(1): 105-112.
TANG X M, HAN Y X, HAN S C. Aircraft 4D trajectory estimation based on hybrid system model[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2012, 44(1): 105-112(in Chinese).

[3] 张军峰, 葛腾腾, 陈强, 等. 离场航空器四维航迹预测及不确定性分析[J]. 西南交通大学学报, 2016, 51(4): 800-806.
ZHANG J F, GE T T, CHEN Q, et al. 4D trajectory prediction and uncertainty analysis for departure aircraft[J]. Journal of Southwest Jiaotong University, 2016, 51(4): 800-806(in Chinese).

[4] 吕波, 王超. 改进的扩展卡尔曼滤波在航空器4D航迹预测算法中的应用[J]. 计算机应用, 2021, 41(增刊1): 277-282.
LYU B, WANG C. Application of improved extended Kalman filter in aircraft 4D track prediction algorithm[J]. Journal of Computer Applications, 2021, 41(Sup 1): 277-282(in Chinese).

[5] 戴礼灿, 刘欣, 张海瀛, 等. 基于卡尔曼滤波算法展开的飞行目标轨迹预测[J]. 系统工程与电子技术, 2023, 45(6): 1814-1820.
DAI L C, LIU X, ZHANG H Y, et al. Flight target track prediction based on Kalman filter algorithm unfolding[J]. Systems Engineering and Electronics, 2023, 45(6): 1814-1820(in Chinese).

[6] SHI Z Y, XU M, PAN Q, et al. LSTM-based flight trajectory prediction[C]//Proceedings of the International Joint Conference on Neural Networks. Piscataway: IEEE Press, 2018: 1-8.

[7] SAHADEVAN D, HARIKRISHNAN P M, PONNUSAMY P, et al. Ground-based 4d trajectory prediction using bi-directional LSTM networks[J]. Applied Intelligence, 2022, 52(14): 16417-16434.

[8] 代翔. 基于注意力循环网络的4D轨迹预测模型[J]. 电讯技术, 2022, 62(1): 39-45.
DAI X. A 4D trajectory prediction model based on attentional recurrent network[J]. Telecommunication Engineering, 2022, 62(1): 39-45(in Chinese).

[9] 田杉. 基于神经网络的4D航迹预测方法[D]. 天津: 中国民航大学, 2020.
TIAN S. 4D track prediction method based on neural network[D]. Tianjin: Civil Aviation University of China, 2020(in Chinese).

[10] JIAO W D, MA C. Track prediction algorithm based on GA-BPNN[C]//Proceedings of the IEEE 3rd International Conference on Civil Aviation Safety and Information Technology. Piscataway: IEEE Press, 2021: 211-217.

[11] TRAN P N, NGUYEN H Q V, PHAM D T, et al. Aircraft trajectory prediction with enriched intent using encoder-decoder architecture[J]. IEEE Access, 2022, 10: 17881-17896.

[12] 韩萍, 张启, 石庆研, 等. 基于DBSCAN-GRU算法的终端区4D航迹预测[J]. 信号处理, 2023, 39(3): 439-449.
HAN P, ZHANG Q, SHI Q Y, et al. 4D trajectory prediction of terminal area based on DBSCAN-GRU algorithm[J]. Journal of Signal Processing, 2023, 39(3): 439-449(in Chinese).

[13] HAN P, WANG W Q, SHI Q Y, et al. A combined online-learning model with K-means clustering and GRU neural networks for trajectory prediction[J]. Ad Hoc Networks, 2021, 117: 102476.

[14] BAI S J, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[EB/OL]. (2018-04-19)[2023-11-01]. https://arxiv.org/abs/1803.01271.

[15] HUANG J, DING W J. Aircraft trajectory prediction based on Bayesian optimised temporal convolutional network-bidirectional gated recurrent unit hybrid neural network[J]. International Journal of Aerospace Engineering, 2022, 2022: 2086904.

[16] LIN Z, YUE W Q, HUANG J, et al. Ship trajectory prediction based on the TTCN-attention-GRU model[J]. Electronics, 2023, 12(12): 2556.

[17] ZHANG B, WANG S, DENG L W, et al. Ship motion attitude prediction model based on IWOA-TCN-Attention[J]. Ocean Engineering, 2023, 272: 113911.

[18] ZHANG K Y, ZHANG C M, ZHU Q H, et al. Ship trajectory D-TCN prediction method based on satellite remote sensing positioning data[C]//Proceedings of the China Automation Congress. Piscataway: IEEE Press, 2022: 1419-1424.

[19] ZHOU S H, SONG C F, WANG T H, et al. A short-term hybrid TCN-GRU prediction model of bike-sharing demand based on travel characteristics mining[J]. Entropy, 2022, 24(9): 1193.

[20] WU H, LIANG Y, ZHOU B, et al. A Bi-LSTM and AutoEncoder based framework for multi-step flight trajectory prediction[C]//Proceedings of the 8th International Conference on Control and Robotics Engineering. Piscataway: IEEE Press, 2023: 44-50.

A TCN trajectory prediction method fusing MSC and spatio-temporal dual attention

JIAO Weidong*, YANG Bei

(Tianjin Key Laboratory of Advanced Signal Processing, Civil Aviation University of China, Tianjin 300300, China)

Abstract: Aiming at the problem that existing trajectory prediction models are difficult to effectively extract multi-scale spatio-temporal features, which leads to limited prediction accuracy, a new method MDAT-Net for trajectory prediction based on a temporal convolutional network (TCN) fused with a multi-scale convolutional (MSC) network and spatio-temporal dual attention (STDA) is proposed. The MDAT-Net model contains the MSAT, MTAT trajectory prediction module and voting module. First, in the prediction module, a multi-scale convolution architecture is built using different scale convolution kernels to better extract spatio-temporal features at various scales and solve the fixed kernel size issue in the traditional temporal convolutional network. Secondly, in order to dynamically mine the potential correlation between hidden features and target features, spatial attention mechanism and temporal attention mechanism are introduced to adaptively focus on important information and skip secondary information. Finally, the voting module decides which model to apply for each dimension prediction, allowing the benefits of both prediction models to be combined and high-precision trajectory prediction to be achieved. The experimental results show that the MDAT-Net model can improve the root mean square error (RMSE) up to 83.33% and the mean absolute error (MAE) up to 85.85% with high accuracy and robustness.

Keywords: trajectory prediction; multi-scale convolution; temporal convolutional network; attention mechanism; spatio-temporal features

Received: 2023-11-02; **Accepted**: 2024-02-07; **Published Online**: 2024-03-08 10:37

URL: link.cnki.net/urlid/11.2625.V.20240305.1821.006

Foundation item: National Key Research and Development Program of China (2020YFB1600101)

* **Corresponding author**. E-mail: nxjiaowd@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0735

多模态特征交互与语义引导融合的RGB-T人群计数

陈永[1,2,*]，张娇娇[1]，董珂[1]

(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 甘肃省人工智能与图形图像处理工程研究中心，兰州 730070)

摘 要：RGB-T模态人群计数旨在利用可见光RGB和热红外图像的互补性实现人群计数。针对RGB-T多模态人群计数方法在特征提取时，存在模态间信息交互不足、特征融合不充分，导致人群计数结果不准确的问题，提出了一种多模态特征交互与语义引导融合的RGB-T人群计数方法。设计堆叠小尺度卷积核作为主干网络分支，提取各单模态的粗特征；提出多模态特征交互模块，对RGB和热红外各模态进行特征精细提取，实现模态间信息交互，克服信息交互不足的缺点；设计语义引导融合模块，通过全局与局部特征引导融合，增强多模态人群特征语义相关性，以充分融合多元上下文信息，提高人群目标的识别能力；利用回归头生成人群密度图，并输出计数结果。实验结果表明：所提方法在公开RGBT-CC数据集上各评价指标均优于对比方法，相较于CMCRL方法，所提方法的均方根误差降低了31.12%，对不同场景下人群计数具有更高的准确率。

关 键 词：深度学习；RGB-T；人群计数；多模态特征交互；语义引导融合
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2026)01-0028-10

近年来，随着人口数量的快速增加，人群计数已成为智能监控、人群分析和公共安全领域的重要研究内容[1]。在景区、体育场和广场等人群密集区，通过人群计数可以实现人流安全管控、发出预警信息，以避免踩踏等恶性事故的发生。随着基于深度学习视觉技术[2]的发展，基于深度学习的人群计数方法[3-4]逐渐成为主流方法。

目前，人群计数方法大部分研究利用RGB图像的视觉特征进行计数。Zhang等[5]提出了多列卷积神经网络(multi-column convolutional neural network, MCNN)人群计数方法，利用多种感受野卷积操作来提取人群特征。Li等[6]提出了一种拥塞场景识别网络(congested scene recognition network, CSRNet)计数方法，使用空洞卷积来改善接受野并增强特征提取能力。此外，一些计数模型通过引入多尺度信息[7]和视角估计[8]来提高人群计数的性能。上述研究基于RGB图像的光学信息进行特征提取，不同程度提高了人群计数的准确性。然而，RGB图像虽然具有高分辨率和丰富的语义信息，但易受光照变化的影响，在低照度等夜间场景下，存在图像目标不清晰、人群细节缺失等情况，导致基于RGB的计数方法在低照明等场景下人群计数的准确性较低。

与RGB图像不同，热图像利用目标自身发射的热辐射成像，对光照变化不敏感，在低照明条件下也具有较好的成像效果。然而，热图像却存在图像分辨率低且物体边缘、纹理和几何等细节信息易缺失的问题。因此，综合利用RGB图像与热图像的互补信息，研究针对RGB-T多模态的人群计数，对于提高人群目标计数的准确性具有重要的现实意义[9]。

近年来，基于RGB-T多模态的人群计数受到了广泛关注。Liu等[10]提出RGB-T多模态协作表

收稿日期：2023-11-08；**录用日期**：2024-02-23；**网络出版时间**：2024-03-12 12:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20240312.0923.001
基金项目：国家自然科学基金(62462043,61963023)；兰州交通大学基础研究拔尖人才项目(2023JC36)；兰州交通大学重点研发项目(ZDYF2304)
* **通信作者**. E-mail：edukeylab@126.com

引用格式：陈永，张娇娇，董珂. 多模态特征交互与语义引导融合的RGB-T人群计数[J]. 北京航空航天大学学报，2026，52(1)：28-37.
CHEN Y，ZHANG J J，DONG K. Multimodal feature interaction and semantic guided fusion for RGB-T population counting[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：28-37 (in Chinese).

示学习框架, 设计信息聚合模块来学习不同模态之间的互补性, 但其融合结构单一, 存在信息丢失的问题。Liu 等[11] 提出了基于 Transformer 的 RGB-T 多模态人群计数模型, 提取不同尺度特征并进行融合, 但 Transformer 结构适用于全局信息获取, 易出现局部细节特征丢失的问题, 导致计数不准确。Tang 等[12] 提出了三流自适应融合网络 (three-stream adaptive fusion network, TAFNet), 通过附加流提取 RGB-T 的组合特征, 但由于采用卷积运算的感受野较小, 导致模态间信息融合不充分, 存在计数精度受限的问题。Zhou 等[13] 提出了一种双分支增强特征融合网络 (dual-branch enhanced feature fusion network, DEFNet), 利用数据增强模块融合 RGB 和热模态特征, 但特征提取时忽略了模态间信息交互。Wang 等[14] 提出了双分支多模态计数网络, 利用信息交互模块捕获多模态全局信息并生成密度图, 但忽略了模态间的差异性, 导致计数精度不高。

综上所述, 目前, 多模态人群计数模型在特征提取时存在模态间信息交互不足与特征融合不充分, 导致计数结果不准问题。针对上述问题, 本文提出了一种多模态特征交互与语义引导融合的 RGB-T 人群计数方法。首先, 设计堆叠小尺度卷积核来提取各单模态的粗特征, 以提高各单模态特征信息能力。其次, 提出多模态特征交互模块, 对 RGB 与热红外多模态特征进行精细提取, 并实现模态信息间的交互特征。然后, 设计语义引导融合模块, 通过全局与局部特征引导融合, 增强多模态人群特征语义相关性, 以充分融合多元上下文信息, 提高目标人群的识别能力。最后, 利用回归头生成高质量人群密度图, 并输出计数结果。

1　本文方法

1.1　网络整体模型

针对多模态人群计数模型在特征提取时, 模态间信息交互不足与特征融合不充分, 导致计数结果不准的问题[15], 提出了一种基于深度学习的语义引导融合 RGB-T 人群计数模型。所提网络模型主要由粗特征提取模块、多模态特征交互模块和语义引导融合模块及计数输出模块组成。本文方法的整体模型结构如图 1 所示。

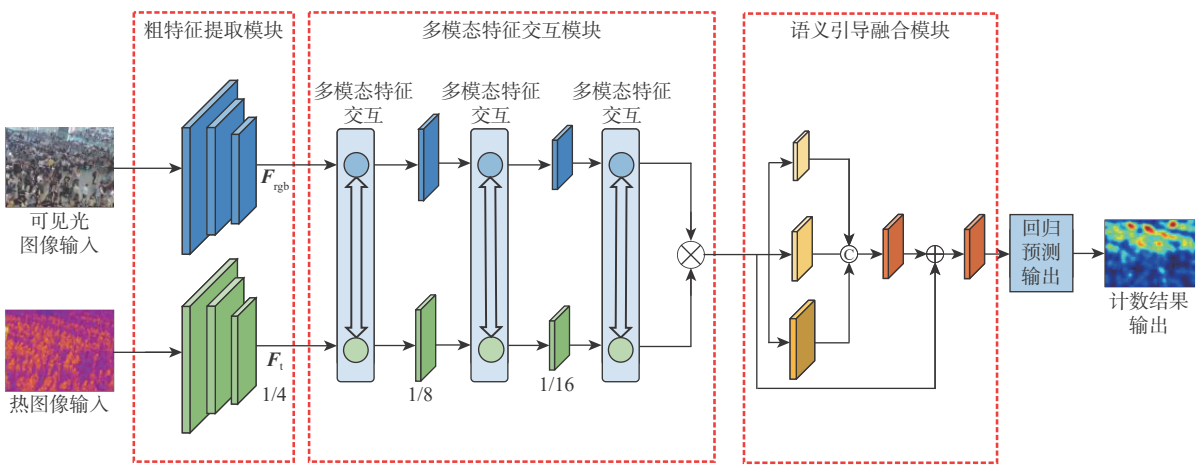


图 1　网络整体结构

Fig. 1　Overall structure of network

模型工作时, 首先, RGB 和热图像分别被输入到粗特征提取主干网络, 用于各自模态特征提取; 其次, 采用多模态特征交互模块进行信息交互和增强, 实现对 RGB 与热红外多模态特征的精细提取及模态信息间交互的特征提取; 然后, 输入语义引导融合模块, 对提取到的特征利用局部与全局引导进行有效融合, 提高目标人群的识别能力; 最后, 通过回归输出密度图及计数结果。

1.2　粗特征提取模块

如图 1 所示, 在本文整体网络模型中设计了堆叠小尺度卷积核网络, 用于 RGB 和热模态的粗特征提取。主干粗特征提取时, 采用双分支结构, 在每个分支采用堆叠小尺度卷积核的设计进行各自模态粗特征提取。卷积层采用大小为 3×3 的小尺寸卷积核, 以便更好地提取不同模态下人群特征, 激活函数为 ReLU 函数, 最后一层为 Softmax 层。具体结构如图 2 所示。图中: Depth 为通道深度。

1.3　多模态特征交互模块

在完成粗特征提取后, 由于热图像与 RGB 图像两者属于不同模态图像, 特征信息差异明显, 为充分利用各模态图像优点, 设计了多模态特征交互模块, 使 RGB 与热红外多模态的信息充分交互, 利用多模态特征交互模块实现对多模态人群特征的精细特征提取, 该模块结构如图 3 所示。图中, 多

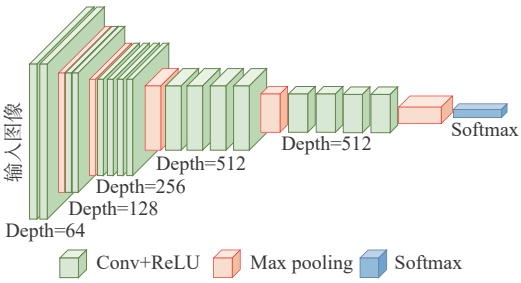


图2 粗特征提取模块结构

Fig. 2 Structure of coarse feature extraction module

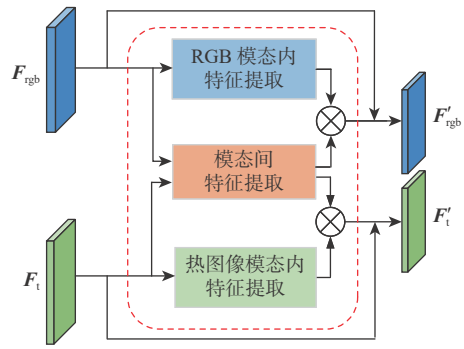


图3 多模态特征交互模块结构

Fig. 3 Structure of multimodal feature interaction module

模态特征交互模块主要由模态间特征提取、RGB模态内特征提取及热图像模态内特征提取构成。

模态间特征提取利用Transformer结构结合卷积操作，其中，Transformer结果可以更好地提取模态间的全局特征，而卷积操作可以更加关注局部特征[16]，将两者结合可以提高RGB图像和热图像之间细节特征的互补能力。

1.3.1 RGB模态内特征提取

图3中，在粗特征提取的基础上，为提高RGB人群计数的细节特征提取能力，设计注意力机制来增加RGB图像中被计数人群的重要区域[17]。本文使用坐标注意力(coordinate attention, CA)[18]来提高可见光RGB图像中的人群细节利用率。坐标注意力以通道注意力为主干，嵌入空间位置信息，在对人群图像中的通道维度进行编码的过程中，也可以关注人物空间信息。坐标注意力结构如图4所示。

由图4可以看出，坐标注意力模块在RGB内人群精细特征提取时，首先，输入高、宽、通道数分别为H、W、C的特征图，在水平和垂直2个方向采用平均池化操作对通道编码后，将其进行空间维度上的拼接，并进行通道数为C/r、大小为1×1的卷积操作(r为控制block大小的缩减变量)；然后，通过批归一化(BN)层和Non-linear激活函数，将变换后的特征图split分离，重新得到2个方向的向量特征，通过卷积重新调整2个方向特征向量的通道数；最后，经过Sigmoid函数得到注意力权重值。

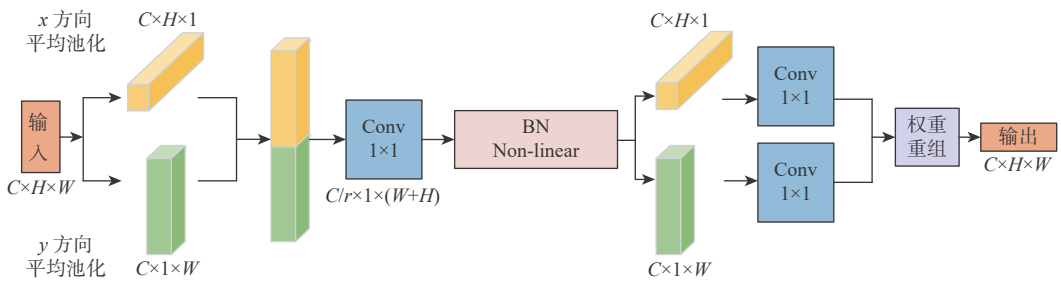


图4 RGB模态内特征提取模块结构

Fig. 4 Structure of RGB in-modal feature extraction module

1.3.2 热图像模态内特征提取

在人群计数时，一般通过增加池化层来增强人群计数特征的平移不变性，但在下采样时会出现目标信息丢失的问题[19]。为解决上述问题，利用三维注意力SimAM[20]提取热图像信息，其主要思想为：通过定义能量函数计算权重值，来提高网络模型的学习能力。三维注意力机制SimAM利用能量函数发掘主要特征的重要性，强化模型对目标人群特征的提取能力，其结构如图5所示。

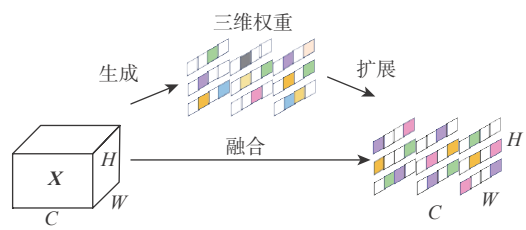


图5 热图像模态内特征提取模块结构

Fig. 5 Structure of feature extraction module within thermal image model

图5中，SimAM通过抑制其他神经元，即通过三维权重重组特征来提高对计数特征的关注。SimAM模块使用能量函数来查找重要神经元，定义如下：

$$e_t(w_t,b_t,y,x_i)=(y_t-\hat{t})^2+\frac{1}{M-1}\sum_{i=1}^{M-1}(y_0-\hat{x}_i)^2 \quad (1)$$

式中：输入特征图$\boldsymbol{X}$表示单一通道中的空间抑制神经元，x_i表示其他神经元；$\hat{t}=w_t t+b_t$和$\hat{x}_i=w_t x_i+b_t$

分别表示线性变换，b_t 和 w_t 分别为偏差和权重的线性变换，t 为目标神经元的值；M 为神经元数量；y 为变量；y_t 为目标神经元的标签值；y_0 为其他神经元的标签值。

将式 (1) 最小化，采用二值标签并添加正则项，得到

$$e_t(w_t,b_t,y,x_i)=\frac{1}{M-1}\sum_{i=1}^{M-1}(-1-(w_tx_i+b_t))^2+(1-(w_tt+b_t))^2+\lambda w_t^2 \tag{2}$$

式中：λ 为系数。

一般，每个通道都有 $H\times M$ 个能量函数，其中，

$$w_t=-\frac{2(t-u_t)}{(t-u_t)^2+2\sigma_t^2+2\lambda} \tag{3}$$

$$b_t=-\frac{1}{2}(t+u_t)w_t \tag{4}$$

式中：$u_t=\frac{1}{M-1}\sum_{i=1}^{M-1}x_i$ 为该通道中除 t 以外的神经元平均值；$\sigma_t^2=\frac{1}{M-1}\sum_{i=1}^{M-1}(x_i-u_t)^2$ 为除 t 以外的神经元的方差。

由此可以得到能量函数的最小值 e_t^*，$1/e_t^*$ 为神经元权重，公式如下：

$$e_t^*=\frac{4(\hat{\sigma}^2+\lambda)}{(t-u_t)^2+2\hat{\sigma}^2+2\lambda} \tag{5}$$

式中：e_t^* 为单对单的神经元能量；$\hat{\sigma}$ 为估计标准差。

一个通道中有多个神经元，可以构成能量矩阵 $\boldsymbol{E}$，经过 SimAM 注意力后得到的特征图 $\tilde{\boldsymbol{X}}$ 如下：

$$\tilde{\boldsymbol{X}}=\mathrm{Sigmoid}\left(\frac{1}{\boldsymbol{E}}\right)\odot\boldsymbol{X} \tag{6}$$

式中：$\boldsymbol{X}$ 为输入注意力机制的特征图。

式(6)根据注意力机制定义对特征增强。Sigmoid 激活函数定义如下：

$$s(x)=\frac{1}{1+\mathrm{e}^{-x}} \tag{7}$$

1.3.3　模态间特征提取

图 3 中，在提取到 RGB 与红外热图内部不同模态各自特征后，进一步设计了模态间特征提取模块，以完成多模态信息交互，得到更加准确的人群特征信息。模态间特征提取模块结构如图 6 所示。

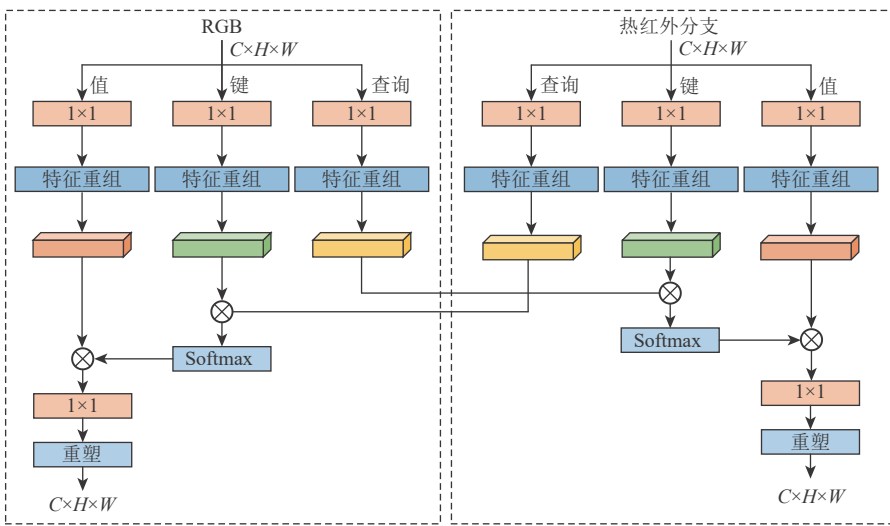


图 6　模态间特征提取模块结构

Fig. 6　Structure of feature extraction module between modalities

图 6 中，在第 $l(l=1,2,\cdots,l_N)$ 个模态间特征提取模块中，输入不同模态的输入 z_{r}^{l-1} 和 z_{t}^{l-1}。其中，输入键值来自相同的模态，但查询键来自另一个模态，通过对不同模态的查询和键值计算注意图，进行特征多模态信息交互，如下：

$$\{z_{\mathrm{r}}^{l-1}\}^c=M_{h=1}^{N_h}(\{Q_{\mathrm{t}}^h\}^c,\{K_{\mathrm{r}}^h\}^c,\{V_{\mathrm{r}}^h\}^c)+z_{\mathrm{r}}^{l-1}=C_{\mathrm{t}\to\mathrm{r}}(z_{\mathrm{r}}^{l-1}) \tag{8}$$

$$\{z_{\mathrm{t}}^{l-1}\}^c=M_{h=1}^{N_h}(\{Q_{\mathrm{r}}^h\}^c,\{K_{\mathrm{t}}^h\}^c,\{V_{\mathrm{t}}^h\}^c)+z_{\mathrm{t}}^{l-1}=C_{\mathrm{r}\to\mathrm{t}}(z_{\mathrm{t}}^{l-1}) \tag{9}$$

式中：$\{\cdot\}^c$ 表示该模块中使用的变量；$M_{h=1}^{N_h}()$ 为多头注意力计算函数，h 为注意力头部，N_h 为注意力头部数量；Q_{t}^h 为热模态注意力头部查询值；K_{r}^h 为 RGB 模态注意力头部键值；V_{r}^h 为 RGB 模态注意力值；$C_{\mathrm{t}\to\mathrm{r}}$ 表示查询从热模态得到，键和值从 RGB 模态得到共同关注；$C_{\mathrm{r}\to\mathrm{t}}$ 表示查询来自 RGB 模态，键和值来自热图像。

将不同子模块得到的特征按元素积进行聚合，聚合公式为

$$\begin{cases}z_{\mathrm{r}}^{l}=\{z_{\mathrm{r}}^{l-1}\}^i\odot\{z_{\mathrm{r}}^{l-1}\}^c\\ z_{\mathrm{t}}^{l}=\{z_{\mathrm{t}}^{l-1}\}^i\odot\{z_{\mathrm{t}}^{l-1}\}^c\end{cases} \tag{10}$$

为验证提出的模态间特征提取模块的有效性，进行模态间特征提取前后特征图的对比，如图 7 所示。其中，图 7(c) 为未使用模态间特征提取特征图，可以看出，其将非人群区域也进行了关注，如地面及树干区域，而该类特征对人群计数有一定的干扰影响。图 7(d) 为使用本文模态间特征模块提取后特征图，与图 7(c) 相比，使用本文方法有效抑制了与人群计数无关的背景干扰，增强了人群待计数主体特征，更有利于输出高质量密度图。

图 7 模态间特征提取模块热力图对比

Fig. 7 Comparison of heatmaps of feature extraction modules between modalities

1.4 语义引导融合模块

在完成多模态特征交互特征提取后，为提高 RGB 与热图像多模态全局信息的利用率，进一步设计基于注意力的语义引导融合模块，以实现两模态间更好的信息融合。在该模块中，设计了 2 种类型的注意力来引导特征融合，即全局注意力 (global attention, GA) 和局部注意力 (local attention, LA)。全局注意力用于增强多模态人群全局上下文语义，局部注意力用于增加局部细节。其结构如图 8 所示。

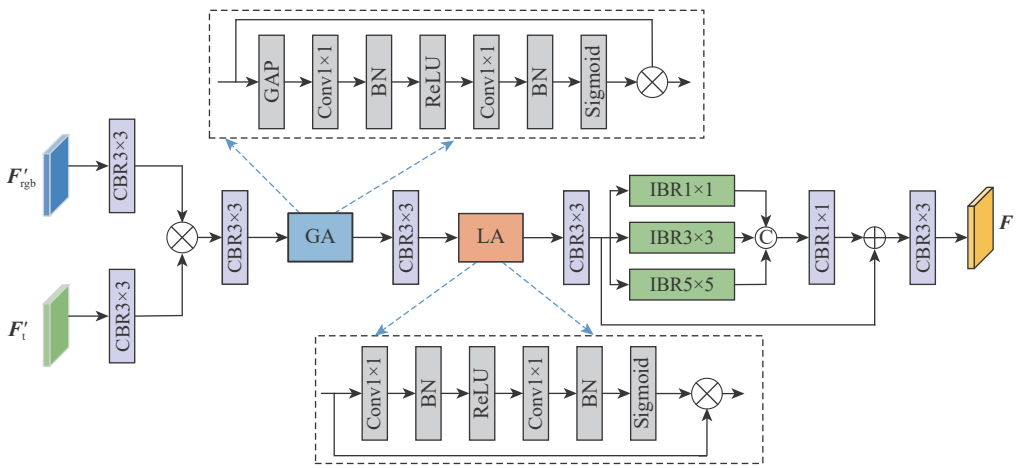


图 8 语义引导融合模块结构

Fig. 8 Structure of semantic guided fusion module

图 8 中，语义引导融合模块有 2 个输入，即模态特定特征与多模态特征间的输出。首先，调整可见光图像特征图 $\boldsymbol{F}'_{\mathrm{rgb}}$ 和热图像特征图 $\boldsymbol{F}'_{\mathrm{t}}$ 的大小，进行 3 × 3 卷积操作，以确保相同的通道数。然后，分别将 $\boldsymbol{F}'_{\mathrm{rgb}}$ 和 $\boldsymbol{F}'_{\mathrm{t}}$ 相乘进行初始融合，并继续进行 3×3 卷积特征更新，该过程如下：

$$\boldsymbol{f}' = F_{3\times3}\left(F_{3\times3}\left(\boldsymbol{F}'_{\mathrm{rgb}}\right)\otimes F_{3\times3}\left(\boldsymbol{F}'_{\mathrm{t}}\right)\right) \tag{11}$$

在得到初始融合特征 $\boldsymbol{f}'$ 后，引入全局注意力和局部注意力来引导特征的互补，其中，全局注意力的结构如图 8 所示，公式如下：

$$\boldsymbol{G}_{\mathrm{att}} = \boldsymbol{f}'\otimes\sigma(B(\mathrm{Conv}_{1\times1}(F_{1\times1}(G(\boldsymbol{f}'))))) \tag{12}$$

式中：$\boldsymbol{G}_{\mathrm{att}}$ 为全局注意力计算结果；$F_{1\times1}$ 表示大小为 1 的卷积操作；$G(\,)$ 为全局平均池化函数；B 表示批归一化层；σ 为激活函数。

将特征全局化操作后，再经过 3×3 卷积操作调整通道数，得到全局注意力特征图 $\boldsymbol{f}_{\mathrm{G}}$。为对局部特征信息进行有效融合，提出了局部注意力部分对其指导融合，结构如图 8 所示，其公式为

$$\boldsymbol{L}_{\mathrm{att}} = \boldsymbol{f}_{\mathrm{G}}\otimes\sigma(B(\mathrm{Conv}_{1\times1}(F_{1\times1}(\boldsymbol{f}_{\mathrm{G}})))) \tag{13}$$

式中：$\boldsymbol{L}_{\mathrm{att}}$ 为局部注意力计算结果。

最后，设计了多尺度卷积块来增强全局与局部特征，并输出最终特征。过程如下：

$$\boldsymbol{F} = F_{3\times3}\left(F_{3\times3}\left(F_{1\times1}\left(\mathrm{Concat}\left(f^{\mathrm{d}}_{1\times1}, f^{\mathrm{d}}_{3\times3}, f^{\mathrm{d}}_{5\times5}\right)\right)\oplus f_{3\times3}\right)\right) \tag{14}$$

式中：$f^{\mathrm{d}}_{1\times1}$、$f^{\mathrm{d}}_{3\times3}$、$f^{\mathrm{d}}_{5\times5}$ 分别表示膨胀率为 1、3、5 的膨胀卷积操作；$f_{3\times3}$ 表示 3×3 卷积操作。

经过上述操作，最终输出融合特征图 $\boldsymbol{F}$。

1.5 损失函数

本文采用贝叶斯损失函数[21] 与标准均方误差

损失函数训练模型, 总损失函数为两者之和, 公式如下:

$$L = L^{\text{baseline}} + L^{\text{mse}} \tag{15}$$

式中: L^{baseline}为贝叶斯损失; L^{mse}为标准均方误差损失。

人群计数中, 真值密度图一般通过高斯核生成, 通过训练模型生成预测密度图并像素求和得到人群计数结果。但在高斯核在生成密度图时, 会存在密集人群的遮挡、透视等因素, 使得计数结果不佳。因此, 引入贝叶斯损失, 对每个像素点的计数期望进行监督训练, 其公式如下:

$$L^{\text{baseline}} = \sum_{m=1}^{M} S\left(D^{\text{gt}}(X_m) - D^{\text{est}}(X_m)\right) \tag{16}$$

式中: $S(\cdot)$为距离函数; $D^{\text{gt}}(X_m)$和$D^{\text{est}}(X_m)$分别为真值密度图和模型预测密度图; X_m为像素位置, $m=\{1,2,\cdots,Z\}$, Z为密度图中的总像素数量。

为训练所提整体模型, 采用标准均方误差(mean squared error, MSE)损失函数, 该函数使用欧氏距离来计算实际真值和预测的人群密度图之间的差距, 公式如下:

$$L^{\text{mse}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{D}}_i - \boldsymbol{D}_i\right\|_2^2 \tag{17}$$

式中: N为训练样本的数量; i为第 i 个训练样本; $\hat{\boldsymbol{D}}_i$为预测的密度图; $\boldsymbol{D}_i$为由实际真值生成的密度图。

2　实　验

2.1　实验数据与实验环境

为验证本文方法的有效性, 在大规模 RGB-T 人群计数数据集 RGBT-CC[10] 上进行对比实验。该数据集使用光学热像仪在不同环境中 (如商场、游乐场、街道、地铁站、火车站) 采集了多幅 RGB-T 图像, 获得了 2 030 对 RGB-T 图像, 其中, 1 013 对图像在照明条件下采集, 其余 1 017 对图像在黑暗条件下采集。在数据集中, 每张 RGB-T 图像的大小为 640 像素×480 像素。从数据集样本中使用 1 030 对用于训练, 200 对用于验证, 其余 800 对用于测试, 其场景也复杂多样。

实验环境为 PyTorch 深度学习框架, 显卡为 NVIDIA GeForce GTX 1660, 采用 Adam 优化器, 初始学习率为 5×10^{-5}, 本文方法及对比方法均以相同环境进行实验。

2.2　评价指标

为客观评价实验结果, 采用均方根误差 (root mean squared error, RMSE) 作为总体评价指标, 此外, 使用网格平均绝对误差 (grid average mean absolute error, GAME) 来评价不同图像区域的预测性能, 将给定的图像分成 4 个不重叠的区域, 并测量每个区域的计数误差。

$$R_{\text{MSE}} = \sqrt{\frac{1}{U}\sum_{i=1}^{U}\left(P_i - \hat{P}_i\right)^2} \tag{18}$$

$$E(j) = \frac{1}{U}\sum_{i=1}^{U}\sum_{l=1}^{4^j}\left|P_i^l - \hat{P}_i^l\right| \tag{19}$$

式中: R_{MSE}为均方根误差值; U为测试图像的个数; P_i和$\hat{P}_i$分别为估计密度图的总数和相应的实际真值; $E(j)$为网格平均绝对误差值, j为网格级别。

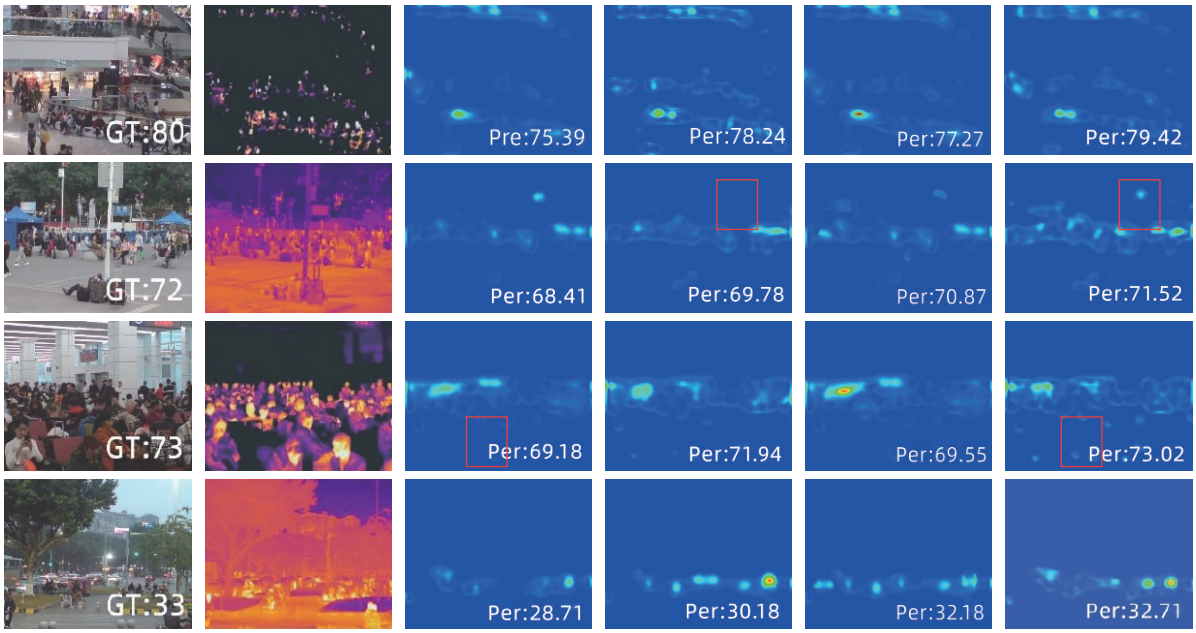
与 RMSE 不同的是, $E(j)$将每张图像划分为 4^j 不重叠区域来测量每个相等区域的计数误差, 其中, P_i^l和$\hat{P}_i^l$分别为图像 i 的区域 l 中估计计数和真值。

2.3　复杂场景下实验结果

为验证本文方法的有效性, 进行复杂场景下人群计数实验。图 9 为本文方法与 CMCRL[10]、SANet[22]、DEFNet[13] 多模态人群计数方法实验比较结果。图 9(a) 为 RGB 图像, 前 3 幅图像分别在商场、车站外、车站等候室场景下拍摄, 都存在遮挡严重问题, 最后 1 幅图像为街景, 由于天气影响, 可以看出拍摄的 RGB 图像中存在远处行人模糊不清问题。图 9(b) 为对应配准后的热图像。图 9(c) 为 CMCRL 计数网络输出的密度图及计数结果, 从实验结果可以看出, 该方法计数准确度较对比方法最低, 这是因为该方法没有充分利用不同模态之间的互补性, 在融合阶段仅利用拼接操作将 2 种模态特征进行融合, 导致人群信息丢失, 精度较低。图 9(d) 为 SANet 计数网络输出的密度图及计数结果, 该方法较 CMCRL 计数准确性有所提高, 但距真值仍有一定误差, 这是因为该方法在学习不同模态之间信息后, 未有效融合特征, 对人群区域的感知较弱, 存在计数误差较大问题, 此外, 对于遮挡及不清晰人群目标存在漏检的情况。图 9(e) 为 DEFNet 网络计数结果, 可以看出, 该方法的部分结果较 CMCRL 和 SANet 方法有所提高, 但仍存在一定的计数误差。图 9(f) 为本文方法结果, 可以看出, 采用多模态特征交互模块与语义引导融合模块后, 其计数精度与行人的关注度置信度相较于对比方法都表现出一定的优势。

2.4　昏暗弱光环境下实验结果

在完成复杂场景下人群计数实验后, 进一步进行昏暗弱光环境下人群计数实验。在该环境下, RGB 图像存在人群特征不明显或缺失现象, 计数难度较大。不同方法实验结果如图 10 所示。图 10(a) 为 RGB 图像, 根据其光学信息, 无法检测出有效的

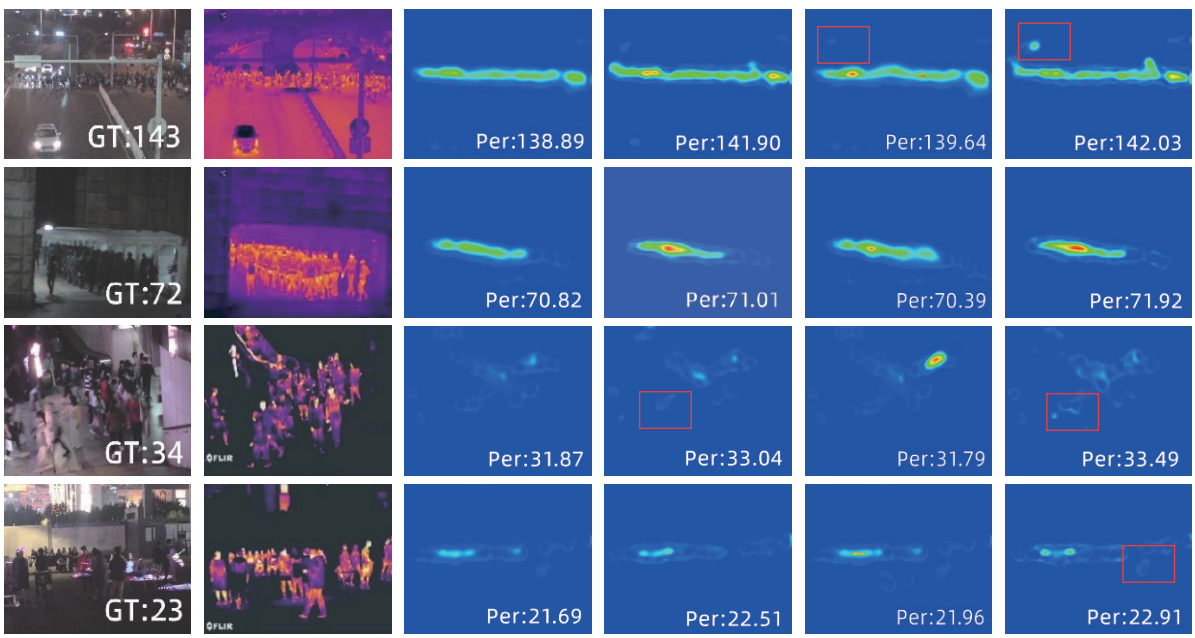


(a) RGB 图像 (b) 热图像 (c) CMCRL[10] (d) SANet[22] (e) DEFNet[13] (f) 本文方法

图 9 复杂场景实验结果

Fig. 9 Experimental results of complex scenarios

行人区域。图 10(b) 为对应的热图像。图 10(c) 为 CMCRL 方法的计数结果，从方框标注可以看出，该方法无法有效关注昏暗状态中的行人目标，检测鲁棒性较差。图 10(d) 为 SANet 方法计数结果，可以看出，该方法相较于 CMCRL 误差有所减小，但如方框标注所示，仍存在部分行人信息未能有效提取的问题。图 10(e) 为 DEFNet 方法计数结果及密度图，可以看出，该方法存在人群特征信息丢失的现象。图 10(f) 为本文方法，可以看出，实验结果较对比方法输出了误差更小的计数结果及密度图。



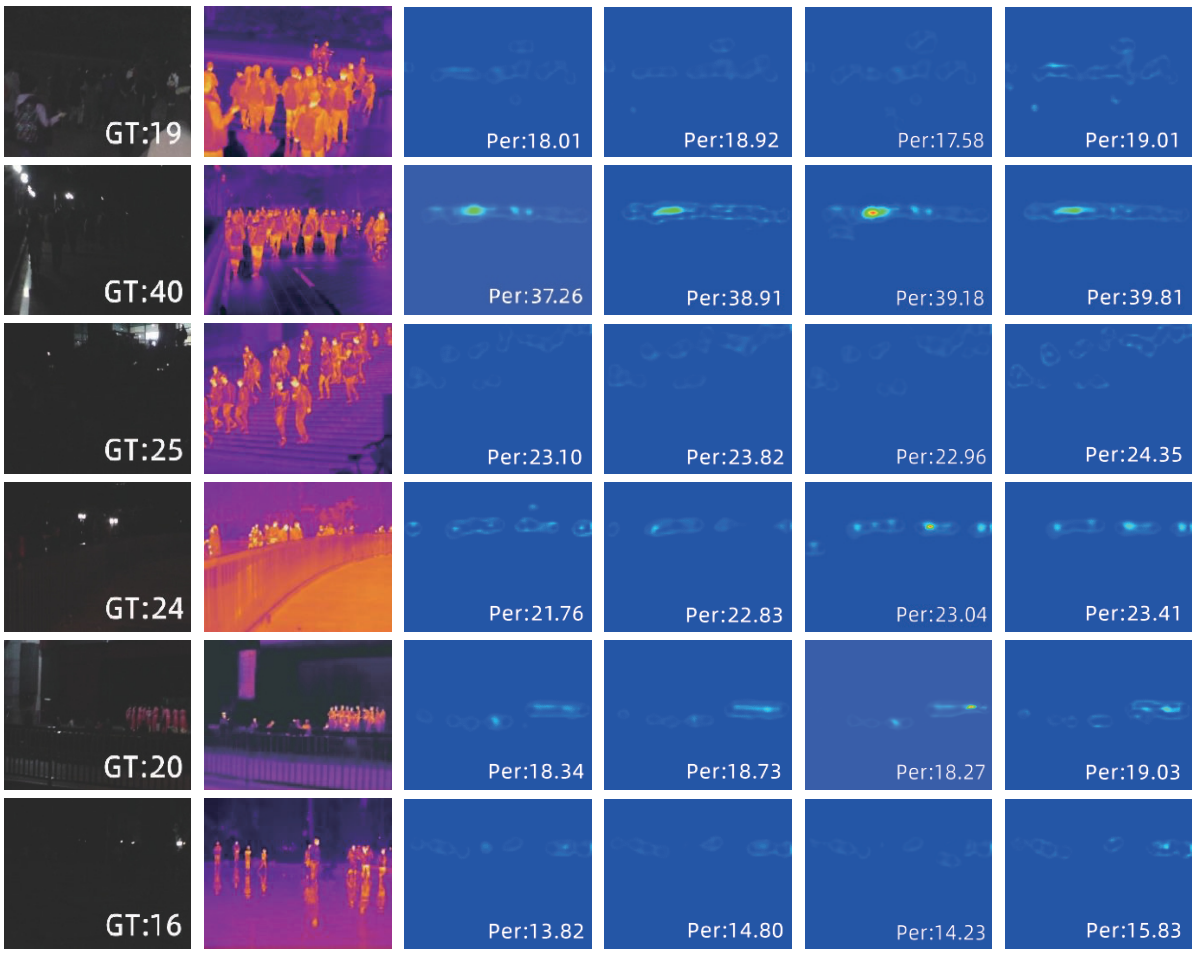
(a) RGB 图像 (b) 热图像 (c) CMCRL[10] (d) SANet[22] (e) DEFNet[13] (f) 本文方法

图 10 昏暗弱光环境实验结果

Fig. 10 Experimental results in dim and low light environment

2.5 夜间环境下实验结果

在夜间场景下，可以看出 RGB 图中行人仅利用光学信息无法区分，该情况下仅利用 RGB 单模态图像无法完成人群计数，结合热图像特点可以实现多模态人群计数。不同方法在夜间环境下人群计数实验结果如图 11 所示。通过对比可以看出，本文方法在夜间环境下较对比方法能够在夜间环境下实现更精准的计数。



(a) RGB 图像　(b) 热图像　(c) CMCRL[10]　(d) SANet[22]　(e) DEFNet[13]　(f) 本文方法

图 11　夜间环境实验结果

Fig. 11　Experimental results of night environment

2.6　客观评价指标对比

表 1 为实验评价指标对比。均方根误差和网格平均绝对误差 $E(j)$ 的值越小，计数精度越高。可以看出，本文方法均优于对比方法，说明本文方法人群计数精度更高。

如表 1 所示，在 RGBT-CC 数据集上，本文方法在网格平均绝对误差 $E(0)$、$E(1)$、$E(2)$、$E(3)$ 和均方根误差 RMSE 这 5 个指标上分别有所改善，与 CMCRL 相比，本文方法在 RMSE 上降低了 31.12%。

表 1　客观评价指标对比

Table 1　Comparison of objective evaluation metrics

方法	$E(0)$	$E(1)$	$E(2)$	$E(3)$	RMSE
CMCRL[10]	15.61	19.95	24.69	32.89	28.18
SANet[22]	14.32	18.91	23.81	32.47	26.01
DEFNet[13]	11.90	16.08	20.19	27.27	21.09
本文方法	10.32	14.28	18.26	25.90	19.41

为进一步验证本文方法对光照条件的鲁棒性，将 800 张图像的 RGBT-CC 测试集分为 406 张高亮度图像和 394 张黑暗图像，测试了该模型在不同光照条件下的性能，结果如表 2 所示。可以看出，与 SANet、CMCRL 和 DEFNet 相比，本文方法在不同光照条件下的性能都优于对比方法。

表 2　不同光照条件下评价结果

Table 2　Evaluation results under varying lighting conditions

数据集	方法	$E(0)$	$E(1)$	$E(2)$	$E(3)$	RMSE
高亮度图像数据集	CMCRL	20.36	23.57	28.49	36.29	32.57
	SANet	15.57	20.65	26.67	36.17	24.25
	DEFNet	12.03	15.39	20.09	27.99	22.96
	本文方法	11.31	14.61	18.92	25.31	21.81
黑暗图像数据集	CMCRL	15.44	19.23	23.79	30.28	29.11
	SANet	14.20	19.20	24.00	31.63	27.50
	DEFNet	12.39	16.16	20.33	27.16	23.16
	本文方法	10.48	14.26	17.79	22.67	19.75

2.7　消融实验

为验证所提网络模块的有效性，进行消融实验。基础网络为 CMCRL，将模态间特征提取模块结构称为 M，RGB 模态内特征提取分支称为 R，热图像模态内特征提取模块称为 T，语义引导融合模块称为 G。消融实验分为 6 组进行比较：CMCRL，M+CMCRL，M+T+CMCRL，M+R+CMCRL，M+R+T+

CMCRL, M+R+T+CMCRL+G。RGBT-CC数据集上的结果如表3所示。

从表3可以看出，加入M模块结构后，误差明显减小；在原始网络模型嵌入M与T模块后，通过特征融合增强热图像目标的特征，将网络模型的误差由28.18降低到22.72；在此基础上加入R模块，在前向训练的过程中继续关注RGB图像的特征感受野，模型误差继续降低；最后，在网络计数后端采用G模块，本文方法的误差降低到19.41。消融实验验证了各模块的有效性。

表3 消融实验指标比较

Table 3 Comparison of ablation experimental indexes

CMCRL	M	R	T	G	RMSE
√					28.18
√	√				25.89
√	√		√		22.72
√	√	√			23.12
√	√	√	√		21.67
√	√	√	√	√	19.41

3 结 论

1）提出一种多模态特征交互与语义引导融合的RGB-T人群计数模型，设计多模态特征交互模块增强RGB与红外热图的人群特征关联性。

2）提出了语义引导融合模块，以充分融合多元上下文特征，提升了不同复杂场景下人群计数精度。

3）实验结果表明，与对比方法相比，本文方法计数精度较高。相较于CMCRL人群计数方法，均方根误差降低了31.12%，本文方法对不同场景人群计数具有更高的准确率。

虽然本文方法在大规模人群计数数据集RGBT-CC上取得了较好的结果，但人群计数时，在不同场景下人群分布、天气状况等方面存在较大的差异，如何提高计数模型的跨场景泛化性是本文后续研究的重点工作，后续将进一步结合迁移学习等方法，提高模型在跨场景下人群计数的准确性和鲁棒性。

参考文献（References）

[1] WANG M J, ZHOU J, CAI H, et al. CrowdMLP: weakly-supervised crowd counting *via* multi-granularity MLP[J]. Pattern Recognition, 2023, 144: 109830.

[2] 陈永, 王镇, 卢晨涛, 等. 红外弱光下多特征融合与注意力增强铁路异物检测[J]. 北京航空航天大学学报, 2023, 49(8): 1884-1895.
CHEN Y, WANG Z, LU C T, et al. Detection of railway object intrusion under infrared low light based on multi-feature and attention enhancement network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(8): 1884-1895(in Chinese).

[3] 杨家林, 林春雨, 聂浪, 等. 面向鱼眼图像的人群密度估计[J]. 北京航空航天大学学报, 2022, 48(8): 1455-1463.
YANG J L, LIN C Y, NIE L, et al. Crowd density estimation for fisheye images[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(8): 1455-1463(in Chinese).

[4] DU Z P, SHI M J, DENG J K, et al. Redesigning multi-scale neural network for crowd counting[J]. IEEE Transactions on Image Processing, 2023, 32: 3664-3678.

[5] ZHANG Y Y, ZHOU D S, CHEN S Q, et al. Single-image crowd counting via multi-column convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 589-597.

[6] LI Y H, ZHANG X F, CHEN D M. CSRNet: dilated convolutional neural networks for understanding the highly congested scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 1091-1100.

[7] LIU W Z, SALZMANN M, FUA P. Context-aware crowd counting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5094-5103.

[8] YANG Y F, LI G R, WU Z, et al. Reverse perspective network for perspective-aware object counting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 4373-4382.

[9] GU S Q, LIAN Z C. A unified RGB-T crowd counting learning framework[J]. Image and Vision Computing, 2023, 131: 104631.

[10] LIU L B, CHEN J Q, WU H F, et al. Cross-modal collaborative representation learning and a large-scale RGBT benchmark for crowd counting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 4821-4831.

[11] LIU Z, WU W, TAN Y C, et al. RGB-T multi-modal crowd counting based on Transformer[C]//Proceedings of the British Machine Vision Conference. [S. l. : s. n.], 2022: 1-14.

[12] TANG H H, WANG Y, CHAU L P. TAFNet: a three-stream adaptive fusion network for RGB-T crowd counting[C]//Proceedings of the IEEE International Symposium on Circuits and Systems. Piscataway: IEEE Press, 2022: 3299-3303.

[13] ZHOU W J, PAN Y, LEI J S, et al. DEFNet: dual-branch enhanced feature fusion network for RGB-T crowd counting[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 24540-24549.

[14] WANG J D, SUN K, CHENG T H, et al. Deep high-resolution representation learning for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3349-3364.

[15] LI H, ZHANG J G, KONG W H, et al. CSA-Net: cross-modal scale-aware attention-aggregated network for RGB-T crowd counting[J]. Expert Systems with Applications, 2023, 213: 119038.

[16] 牛玉贞, 林晓锋, 许煌标, 等. 基于Transformer的多尺度优化低照度图像增强网络[J]. 模式识别与人工智能, 2023, 36(6): 511-529.
NIU Y Z, LIN X F, XU H B, et al. Transformer-based multi-scale optimization network for low-light image enhancement[J]. Pattern Recognition and Artificial Intelligence, 2023, 36(6): 511-529(in Chinese).

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[18] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 13708-13717.

[19] LIU B T, GUAN W L, YANG C J, et al. Transformer and graph convolutional network for text classification[J]. International Journal of Computational Intelligence Systems, 2023, 16(1): 161.

[20] YANG L X, ZHANG R Y, LI L D, et al. SimAM: a simple paraeter-free attention module for convolutional neural net - works [C]//Proceedings of the 38th International Conference on Machine Learning. Piscataway: IEEE Press, 2021: 125-134.

[21] JEWSON J, ROSSELL D. General Bayesian loss function selection and the use of improper models[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2022, 84(5): 1640-1665.

[22] ZHANG Y J, CHOI S, HONG S. Spatio-channel attention blocks for cross-modal crowd counting[EB/OL]. (2022-11-14)[2023-11-01]. https://arxiv.org/abs/2210.10392.

Multimodal feature interaction and semantic guided fusion for RGB-T population counting

CHEN Yong[1, 2, *]，ZHANG Jiaojiao[1]，DONG Ke[1]

(1. School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China；
2. Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing，Lanzhou 730070，China)

Abstract：RGB-T mode crowd counting is designed to take advantage of the complementarity of visible RGB and thermal infrared image to achieve crowd counting. Aiming at the problems of insufficient information interaction between modes and insufficient feature fusion in the feature extraction of the RGB-T multimodal crowd counting method, an RGB-T crowd counting method based on multi-modal feature interaction and semantic guided fusion is proposed. Firstly, a stacked small scale convolution kernel is designed as a branch of the backbone network to extract the coarse features of each single mode. Secondly, in order to address the limited information interaction between the modes, a multi-modal feature interaction module is suggested. This module will extract the features of each RGB and thermal infrared mode and actualize the interactive features of the mode information. Then, a semantic-guided fusion module is designed to enhance the semantic relevance of multi-modal crowd features through global and local feature-guided fusion, so as to fully integrate multi-context information and improve the recognition ability of the target population. Finally, the regression head is used to generate the population density map and output the counting results. Experimental results demonstrate that the proposed method outperforms the comparison algorithms on the open RGBT-CC dataset, with a 31.12% reduction in the root-mean-square error value compared to the CMCRL method and higher accuracy for crowd counting under various scenarios.

Keywords：deep learning；RGB-T；population count；multimodal feature interaction；semantic guided fusion

Received：2023-11-08；**Accepted**：2024-02-23；**Published Online**：2024-03-12 12：49

URL：link.cnki.net/urlid/11.2625.V.20240312.0923.001

Foundation items：National Natural Science Foundation of China (62462043,61963023); Lanzhou Jiaotong University Basic Top-Notch Personnel Project (2023JC36); Key Research and Development Project of Lanzhou Jiaotong University (ZDYF2304)

*** Corresponding author**. E-mail：edukeylab@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0744

基于改进YOLOv8的飞机蒙皮缺陷检测算法

章东平[1,*]，王杼涛[1]，夏岳键[1]，徐云超[1]，林丽莉[2]

(1. 中国计量大学 信息工程学院，杭州 310018；2. 浙江工商大学 信息与电子工程学院，杭州 310018)

摘 要：为解决传统飞机蒙皮缺陷检测依靠人眼观察时，因人眼容易疲劳和个体认知有限导致效率降低的问题，提出一种基于改进YOLOv8的飞机蒙皮缺陷检测算法。对数据增强方式进行改进，提出一种切片推理+马赛克的数据增强方法；集成残差块到特征提取网络，增强网络表达能力的同时，提高模型在飞机蒙皮缺陷检测任务中的精度；应用三分支注意力模块改进特征融合网络，减少小目标样本的误检率和漏检率；优化检测头结构，使网络能够更好地将浅层信息与深度信息有效结合。实验结果表明：相比于YOLOv8算法，改进算法在飞机蒙皮缺陷数据集上的平均精度均值(mAP)和查全率分别提高了3.6%和3.7%，在公开数据集VOC2007上的平均精度均值和查全率提高了2.9%和2.2%。

关 键 词：YOLOv8算法；表面缺陷检测；数据增强；目标检测；注意力机制
中图分类号：V223; TP181
文献标志码：A **文章编号**：1001-5965(2026)01-0038-11

目前，机场地勤人员的目视检查仍是飞机外观检查中使用最广泛的方法。但是，地勤人员的目视检查方法存在一定的弊端[1-2]：①人工成本高；②由于飞机停留时间短，检修时间有限，地勤人员检查压力大；③目视检查可能受到人为经验影响。因此，为确保飞机正常飞行，保护乘客的生命健康安全，针对飞机蒙皮缺陷进行快速且较为准确的检测具有重要意义。

传统的检测算法主要利用边缘、亮度、直方图和光谱特征等基本特征来检测和分割缺陷[3]，但要求操作人员具备较高的专业素质，并且检测效率较低。近年来，随着深度学习和机器学习技术的不断进步，航空工业领域也对自动化检测技术表现出了浓厚的兴趣[4-5]。这些技术不仅减少了人为失误，而且提高了检查的效率和准确性，为航空安全提供了更有力的保证。基于计算机视觉的表面检测已应用于各种工业领域，如飞机制造和维护[6-7]及产品表面检测[8-9]。这些视觉检测算法可大致分为2种主要类型：两阶段和一阶段的目标检测算法。在这些典型的目标检测算法中，赵鹏飞等[10]使用信息包残差结构来改善输出接收器的大小，并引入频道注意工具改善小对象的浅语义特征信息，提出了I-Darknet53网络；张宝朋等[11]利用2个3×3卷积核，以联结思想改善网络感受野，提高了小目标的准确度；张海涛等[12]以SSD网络为基础，为融合低层和高层的特征图，结合通道注意机制和扩张性折叠，提高了小目标的准确性和鲁棒性；候瑞环等[13]提出了YOLOv4-TIA网络，在主网络中加入三分支注意力，并在路径聚合网络(path aggregation network, PANet)中添加跳跃连接，以获得更特殊的信息；蒋镕圻等[14]在YOLOv4的颈部增加空间金字塔池化(spatial pyramid pooling, SPP)模块，并在骨干网络中安装空间与通道挤压激励注意力模块(spatial and channel squeeze & excitation block, SCSE)，以提高小目标的探测准确度。

目前，在目标检测技术的发展轨迹中，YOLO[15]

收稿日期：2023-11-20；**录用日期**：2023-12-01；**网络出版时间**：2024-01-30 17:31
网络出版地址：link.cnki.net/urlid/11.2625.V.20240130.0955.001
基金项目：浙江省重点研发计划(2022C01005,2023C01032,2023C01030)
*通信作者. E-mail：06a0303103@cjlu.edu.cn

引用格式：章东平，王杼涛，夏岳键，等. 基于改进YOLOv8的飞机蒙皮缺陷检测算法[J]. 北京航空航天大学学报，2026，52(1)：38-48. ZHANG D P，WANG Z T，XIA Y J，et al. Aircraft skin defect detection algorithm based on enhanced YOLOv8[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：38-48 (in Chinese).

系列算法是迄今为止发展最快、最好的算法之一，尤其是 2023 年发布的新型 YOLOv8 算法，其精度达到了迄今为止的最高水平。虽然 YOLOv8 擅长检测标准尺寸的物体[16]，但在物体偏离这些标准尺寸的特殊场景中，其性能往往会下降。事实上，还有专为小尺寸物体检测而提出的其他算法[17-18]优于 YOLOv8。

对此，本文将飞机蒙皮常见的 6 类缺陷 (破洞、腐蚀、穿孔、划痕、螺钉生锈、螺钉缺失) 作为检测目标，在 YOLOv8 框架的基础上设计了一种专门用于飞机蒙皮缺陷检测的算法。首先，结合切片推理数据增强方法与马赛克数据增强方法，形成新的飞机蒙皮缺陷数据集。其次，在特征提取网络引入残差块，使网络在更深入地提取图像特征的同时，不会显著增加训练和计算负担。然后，在特征提取网络的最后一层引入三分支注意力机制，使网络能够更精确地定位较小的目标。最后，优化检测头模块，将不同尺度的特征集成到配备了多级注意力机制的检测头中，提高整体检测精度，从而有效检测飞机蒙皮的缺陷。

1　相关工作

近年来，深度学习研究在许多领域取得了重大进展，已经证明了其在自然语言处理和计算机视觉方面的潜在价值。因此，许多研究人员正在探索深度学习在飞机蒙皮检测中的应用。2017 年，文献 [19] 使用 Faster R-CNN 和 SSD 模型对航空发动机表面缺陷样本进行训练，为发动机表面缺陷检测提供了检测平台。随着 YOLO 系列算法的出现，文献 [20] 使用 YOLO 神经网络来识别飞机上 5 种类型的表面缺陷。YOLO 在精度和计算成本方面均优于 Faster R-CNN 和 SSD。文献 [21] 介绍了用于裂纹检测的模型，该模型结合了深度可分离卷积、特征金字塔和 YOLOv3，但其检测精度较低，存在改进空间。

随着机器人的普及，文献 [22] 开发了一种使用二维图像处理方法进行飞机外部自动检测的自主移动协作机器人模型，采用 5×5 核掩模去除噪声，并增强飞机表面的纹理，使用霍夫变换和边缘检测圆算法提取飞机表面物体的几何形状。文献 [23] 提出了一种用于飞机远程访问无损评估的蛇臂机器人系统，该机器人可以进入机翼的狭小空间进行裂纹和缺陷检查。因此，基于机器人和增强深度学习的应用为自动化飞机蒙皮缺陷检测提供了前景广阔的新途径，具有重要的研究潜力。

2　改进的 YOLOv8 算法

2.1　改进的数据增强方法

由专业人员和巡检机器人绕机拍摄的高分辨率飞机蒙皮缺陷图像，通常具有缺陷目标尺寸小且伴随着复杂背景环境的问题。为解决该问题，受文献 [24-25] 的启发，本文设计了一种将切片推理与马赛克相结合的数据增强方法。如图 1 所示，与传统单步骤的数据增强方法不同，本文方法先将收集到的高质量飞机蒙皮缺陷图像输入到切片推理模块，将其分为较小的块或片段进行处理，这一步骤有助于更精确地捕捉小目标的细节特征；再将与任务无关的干扰信息 (如不包含蒙皮缺陷的图像) 从数据中剔除；最终通过引入马赛克数据增强方法将多个图像组合成新的训练样本，从而增加背景复杂性，使模型在真实场景中更加鲁棒的运行，增加模型的准确性。

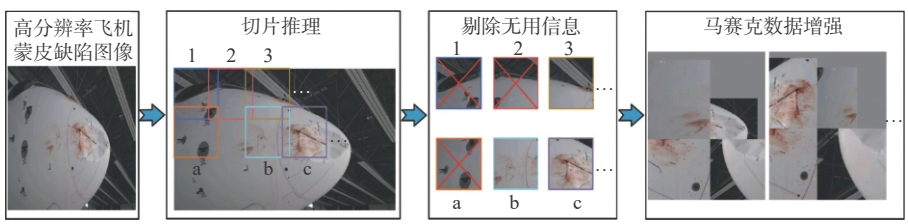


图 1　改进的数据增强流程

Fig. 1　Flow of improved data augmentation

2.1.1　切片推理数据增强方法

切片推理数据增强方法通过采用尺寸为 $M\times N$ 的固定窗口，并以一定的步幅在图像上滑动，将整个图像分割成多个重叠的区域。其中，M 和 N 分别为从预定义范围 $[M_{\min},M_{\max}]$ 和 $[N_{\min},N_{\max}]$ 中选择的超参数。在每个区域内，均保留了图像的局部特征，并通过多次滑动，获得了多个包含丰富信息的子图。该数据增强方法不仅有助于提取图像中的细节信息，而且通过引入多尺度感知，使得模型更具鲁棒性，能够更好地适应不同尺度的目标。

2.1.2　马赛克数据增强方法

马赛克数据增强方法在深度学习目标检测领域具有重要应用价值。该方法通过将 4 张不同的图像巧妙拼接成一张大图，并随机选取裁剪区域的方式，有效引入更多的场景变化和物体组合，从而提高了模型的鲁棒性和泛化能力。在训练过程中，随机选择的图像组合在保留各自特征的同时，模拟了真实世界中复杂场景的多样性，使得目标检测模

型更加灵活且适应性更强。

2.1.3 结合的数据增强方法

尽管切片推理数据增强方法+马赛克数据增强方法引起了额外的计算和存储成本，但目标检测的性能已大大提高。为证明该方法相对于其他数据增强方法的优势，对相同的飞机蒙皮缺陷数据集进行实验。如图2所示，在足够数量的迭代训练和计算下，本文网络的查全率提高了4.4%，平均精度均值（mean average precision，mAP）提高了3.6%，明显优于其他数据增强方法的模型。图2中，传统数据增强方法包括随机旋转、随机裁剪和添加噪声等方法；Mixup表示将2张图像按随机权重进行线性插值生成新样本；Cutout表示在图像随机位置遮挡一块区域以增强模型对缺失信息的鲁棒性；随机翻转表示对图像进行随机调整大小、裁剪、平移和旋转。

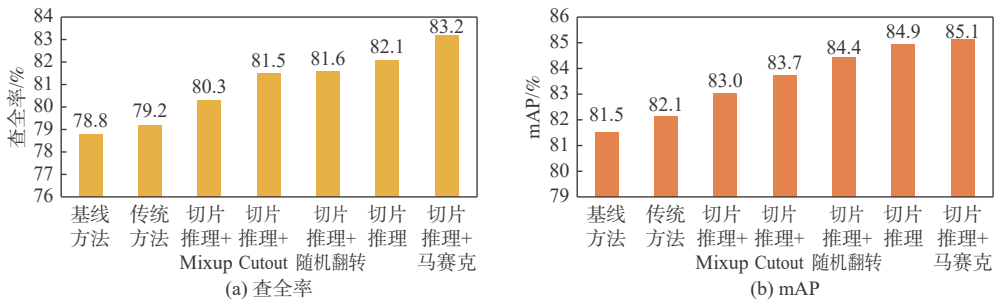


图2 不同数据增强方法对比

Fig. 2 Comparison of different data augmentation methods

2.2 三分支注意力模块

计算机视觉中的注意力机制旨在模拟人类视觉系统，能够有效捕捉复杂场景中的显著区域，从而在多种视觉任务中取得显著进展。与常见的注意力网络结构（如SE-Net和CBAM[26]需要一定数量的可学习参数来建立通道间的依赖关系）不同，三分支注意力模块的核心思想是：通过旋转操作和残差变换，从输入的张量建立维度间的依存关系，并以可忽略的计算开销对通道间和空间信息进行编码，以提高网络学习重要特征的能力。如图3所示，三分支注意力模块由3个平行分支组成，其中，2个分支负责捕获通道C与高度H或宽度W之间的跨维交互，第3个分支专注于构建空间注意力。最终，3个分支的输出使用平均方法进行聚合，形成统一的注意力表示。

由于飞机蒙皮缺陷小目标在图像中像素较小，

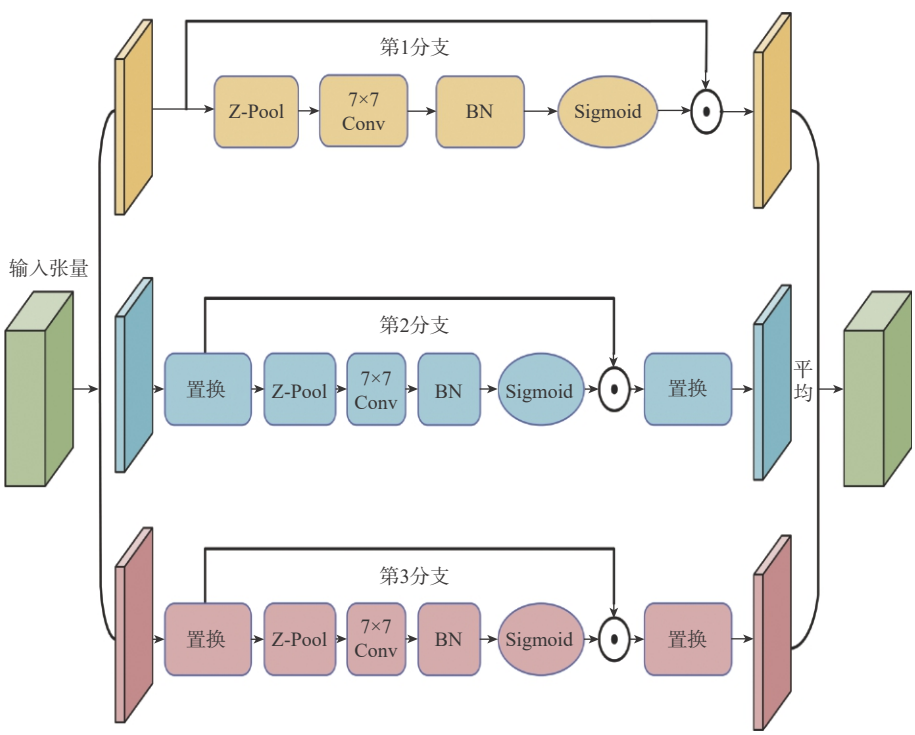


图3 三分支注意力模块结构

Fig. 3 Structure of triplet attention module

骨干网的特征提取难度较大。在骨干网中嵌入三分支注意力模块,使模型在特征提取阶段更容易关注小目标的特征,提高了小目标的检测精度。通过实验比较,本文决定在骨干网的第 5 层增加三分支注意力模块。

2.2.1　第 1 分支

如图 4 所示,输入特征先通过Z-Pool操作进行处理,这一步骤有助于在特定维度上汇总信息;再通过卷积 (Conv) 和批处理归一化层 (BN) 进行运算,进一步提取和精炼特征;最终通过 Sigmoid 激活函数生成通道注意力权重。Z-Pool 的公式如下:

$$\text{Z-Pool}(x)=[\text{MaxPool}(x),\text{AvgPool}(x)] \tag{1}$$

式中:MaxPool表示最大池化;AvgPool表示平均池化。

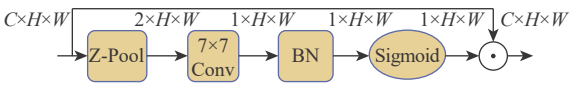


图 4　第 1 分支的网络结构

Fig. 4　Network structure of the first branch

2.2.2　第 2 分支

如图 5 所示,输入特征先通过置换操作,将其维度重新排列为 $H\times C\times W$ 的形式,这一步骤为后续操作提供了合适的数据结构;再在 H 维度上依次执行 Z-Pool 操作、卷积运算、批处理归一化和 Sigmoid 激活函数,这一系列操作可用于有效捕获和强化在 H 维度上的特征依赖和结构信息;最终再次通过置换操作将特征的维度重新排列为 $C\times H\times W$ 的形式,以便与网络的后续层次兼容。

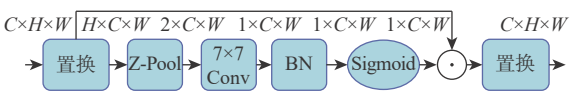


图 5　第 2 分支的网络结构

Fig. 5　Network structure of the second branch

2.2.3　第 3 分支

如图 6 所示,输入特征先通过置换操作,将维度重新排列为 $W\times H\times C$ 的形式;再在 W 维度上采取与第 2 分支类似的操作流程,以进一步处理和精炼特征;最终再次通过置换将特征的维度重新调整为 $C\times H\times W$,从而使其能够与其他分支特征进行相加。

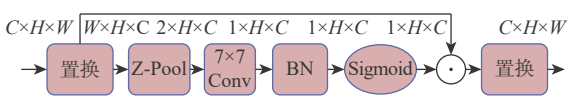


图 6　第 3 分支的网络结构

Fig. 6　Network structure of the third branch

2.3　残差块

主干网络作为 YOLOv8 的核心组件,用于提取图像特征以供后续检测任务使用。虽然在一般的目标检测场景中,YOLOv8 已展现出卓越的性能,但在处理如飞机蒙皮缺陷数据集这类具有复杂几何和纹理特征的目标时,现有主干网络的复杂度可能不足以捕捉足够复杂的特征。然而,单纯增加网络深度可能引发梯度消失或梯度爆炸等问题,从而妨碍网络训练和收敛。

为解决上述问题,本文提出采用残差块替换 YOLOv8 主干网络的策略。如图 7 所示,残差块的结构由主路径和捷径组成,其中,主路径专注于学习输入与目标之间的残差映射,捷径则通过跳跃连接直接连接输入和输出。该结构使梯度可以无阻碍地反向传播,有助于解决深度网络中的梯度消失问题,使网络更深入地学习特征表示,从而改善对目标复杂特征的捕捉能力,进一步提高目标检测的准确性。同时,由于其特殊结构,部分计算可以在训练过程中并行执行,即使模型参数量增加,训练时间也未相应大幅增加。

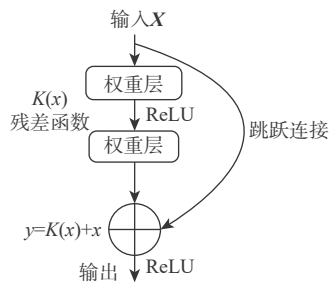


图 7　残差块的网络结构

Fig. 7　Network structure of residual block

在构建新主干网络的过程中,本文不仅选择了合适的残差块深度,还对其进行了调整,以确保与 YOLOv8 的输入和输出兼容。这一设计考虑了增加网络复杂度的需求与计算负担之间的平衡,致力于实现性能提升与计算效率的均衡。

2.4　改进的检测头模块

检测头作为目标检测框架中的关键组件,负责从卷积特征图中提取目标的位置和类别信息。当前主流的 YOLOv8 检测头模块已经采用流行的解耦头结构,并从 Anchor-Based 转变为 Anchor-Free,这使得模型更具有鲁棒性和适应性,能够更好地适应各种复杂的场景和目标特点,为实际应用提供了更可靠的解决方案,但仍缺乏统一的视角来解决检测问题。

受文献 [27] 启示,本文将注意力机制融入目标检测模块中,借助多个自注意力机制在特征层之间的协同交互,实现了尺度感知、空间感知及任务感知的能力,大幅提高了目标检测模块的表

示效果，并且不增加任何计算成本。同时，经过深入的实验研究，本文决定不增加第4个检测头至检测部分。

如图8所示，将图像经过主干网络和颈部网络得到的3层不同尺度的特征调整为统一的特征 $F \in \mathbf{R}^{L\times S\times C}$；将层级特征通过尺度感知注意力模块处理，以融合不同尺度的特征，并加强特征学习表达，使得特征图对前景对象的尺度差异展现出更高的敏感度；空间特征通过空间感知注意力模块来提升特征图中不同位置的特征学习，使得特征图变得更加稀疏，并集中于前景对象的辨别空间位置；将通道特征通过任务感知注意力模块处理以提升特征表达能力。最终，这些特征经过原有的检测头网络处理，输出最终结果。

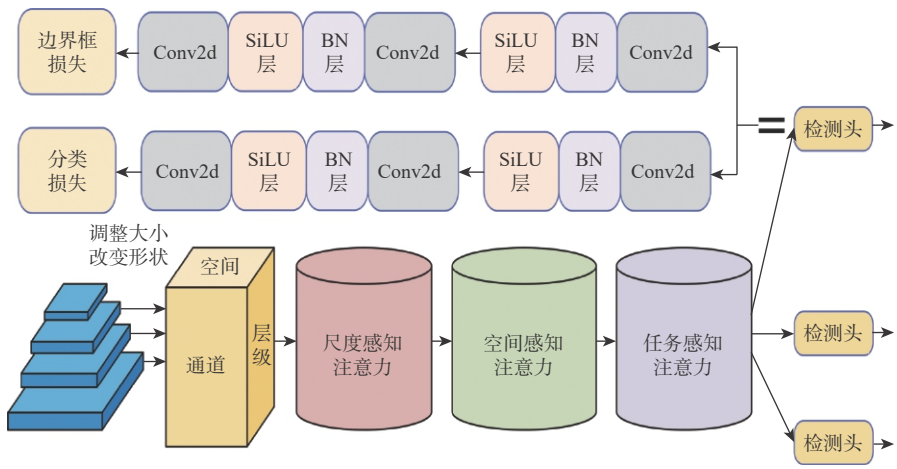


图8 改进的检测头网络结构

Fig. 8 Network structure of improved detection head

2.4.1 尺度感知注意力模块

在层级维度上执行特征，不同层级的特征图对应不同的目标尺度，以基于语义重要性动态融合不同尺度的特征，表达函数为

$$\pi_L(F)\cdot F=\partial\left(f\left(\frac{1}{SC}\sum_{S,C}F\right)\right)F \tag{2}$$

式中：π_L表示尺度感知注意力机制；$\partial()$表示hard-sigmoid函数；$f()$表示线性函数，用1×1卷积实现；$S=H\times W$；F为输入特征；$\frac{1}{SC}\sum_{S,C}F$表示平均池化操作。

2.4.2 空间感知注意力模块

考虑到空间层级的高维度，将空间感知注意力模块分解为2个步骤：①使用可变形卷积使注意力学习变得稀疏；②进行特征跨尺度集成。表达函数为

$$\pi_S(F)\cdot F=\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{K}w_{l,k}F(l;p_k+\Delta p_k;c)\Delta m_k \tag{3}$$

式中：π_S表示空间感知注意力机制；p_k为初始采样位置；c为通道维度；K为稀疏采样位置数；$w_{l,k}F(l;p_k+\Delta p_k;c)$表示可变形卷积操作，$\Delta p_k$为学习到偏差，$w_{l,k}$为可变形卷积运算的参数；$\Delta m_k$为关于位置的$p_k$可自学习的重要性度量因子。

2.4.3 任务感知注意力模块

任务感知注意力模块部署在通道上，其根据来自目标的不同卷积核响应，以max函数的形式，动态地开/关特征通道来选择不同的任务(包括分类、框回归和中心/关键点学习)，促进联合学习与目标表达能力的泛化性。表达函数为

$$\pi_C(F)\cdot F=\max\left(\alpha^1(F)\cdot F_C+\beta^1(F),\alpha^2(F)\cdot F_C+\beta^2(F)\right) \tag{4}$$

式中：π_C表示任务感知注意力模块；F_C为第C通道的特征切片；$[\alpha^1,\beta^1,\alpha^2,\beta^2]$为超函数，用于学习控制激活阈值。

2.5 YOLOv8算法的结构优化

本文提出的改进YOLOv8网络架构如图9所示。其中，Conv2d表示卷积，Bottleneck表示瓶颈结构，Maxpool 2D表示最大池化，SiLU层表示Sigmoid加权线性单元层，DyDetect_Head表示改进的检测头，C2f表示改进型跨阶段部分连接模块，SPFF表示快速空间金字塔池化模块。

该网络结构使用残差块替换主干段，使模型能够提取更深层次的特征。利用三分支注意力模块对特征进行处理，并根据各自的分辨率将合并后的特征传递给检测头模块，实现目标的高精度定位和分类。

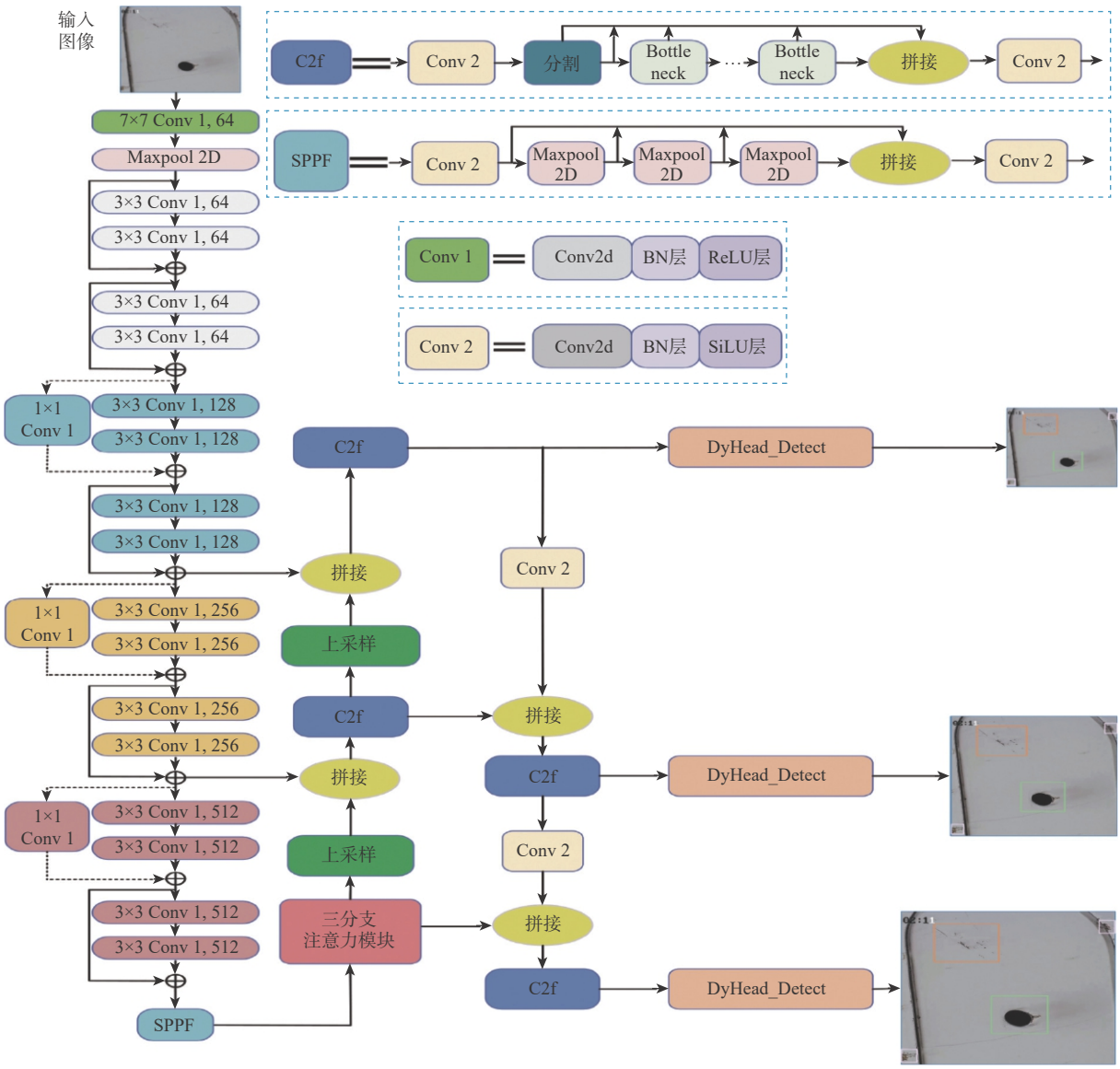


图 9　改进的 YOLOv8 网络结构

Fig. 9　Structure of improved YOLOv8 network

3　实验与分析

3.1　实验环境

硬件配置包括 4 个 NVIDIA GeForce RTX 2080 Ti GPU 和 1 个 Intel(R) Xeon(R) Silver 4210 CPU @ 2.20 GHz, 其中, GPU 的内存为 11 GB RAM。软件方面, 本文使用 Python3.7、PyTorch 1.7.0 和 Cuda11.3 深度学习框架。在实验中, 使用 SGD 优化器, 将初始学习率设置为 0.01, 批处理大小为 32, 输入图像尺寸为 640 像素×640 像素。

3.2　数据采集

由于航空工业的飞机蒙皮质量至关重要, 并且需要严格的检测和评估, 目前还没有为其量身定制的公共数据集。图 10、图 11 显示了捕捉飞机蒙皮的常见缺陷。本文采用巡检机器人和人工拍摄的方式, 遵循巡检路径, 在飞机机场和维修基地中收集不同光线和角度下的飞机蒙皮缺陷照片和视频。

图 10　巡检机器人巡检示意图

Fig. 10　Schematic of inspection by inspection robot

然而, 不是所有的图像都适合用作深度神经网络的训练样本。因此, 本文基于以下准则对图像进行筛选: ①图像中的飞机蒙皮应相对干净, 没有过多污垢、鸟撞留下的血迹或其他污渍; ②当拍摄距离较近时, 飞机蒙皮上的缺陷类型应清晰可见, 且

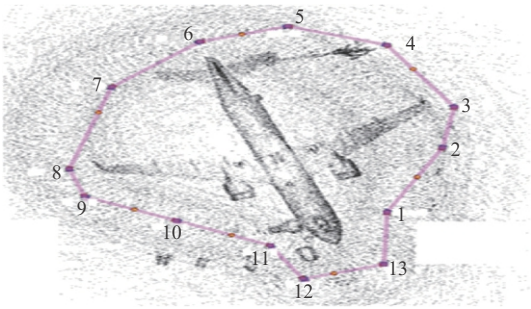
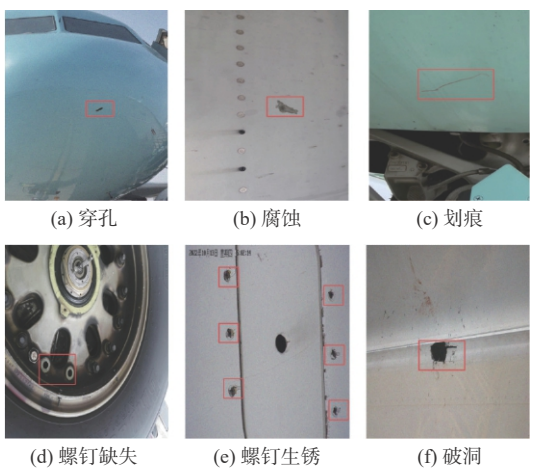


图 11 飞机巡检路线图

Fig. 11 Aircraft inspection route map

损坏不应过于密集或过多；③图像应清晰，没有焦点模糊的问题。在对本文拍摄的数据和从网络上搜集到的数据进行筛选后，发现蒙皮缺陷相对较少。为扩充数据集，进一步提取了一些独立的缺陷，并将其粘贴到正常的蒙皮图像上。通过应用切片推理+马赛克的数据增强技术，创建了一个专门针对飞机蒙皮缺陷检测的数据集。如图 12 所示，数据集标签包括 6 类：穿孔、腐蚀、划痕、螺钉缺失、螺钉生锈和破洞。一共包含 10 343 张图像，35 150 个标签，其中，穿孔 5 383 个、腐蚀 7 912 个、划痕 5 785 个、螺钉缺失 4 437 个、螺钉生锈 6 556 个、破洞 5 077 个。



(a) 穿孔 (b) 腐蚀 (c) 划痕

(d) 螺钉缺失 (e) 螺钉生锈 (f) 破洞

图 12 各类缺陷样本的示意图

Fig. 12 Schematic diagram of various defect samples

为进一步测试模型的泛化能力，在 VOC2007 数据集上进行实验。VOC2007 数据集是目标检测和图像分类领域的权威数据集之一，其由 9 963 张图像组成，覆盖 20 个对象类别，如人、车、动物等，每张图像都带有详细的边界框和类别标注。本文将这 2 个数据集分别随机划分为 80% 的训练集和 20% 的验证集。

3.3 实验结果分析

3.3.1 三分支注意力模块的实验比较

为确切展示三分支注意力模块在性能上的优势，选择相同的飞机蒙皮缺陷数据集，并与其他流行的注意力模块进行比较实验，旨在使所有其他条件（如网络结构、优化策略和超参数等）都保持一致，以确保公正性。实验结果如表 1 所示[28-29]。实验结果表明，本文网络的查全率提高了 1.2%，平均精度均值 (mAP0.5) 提高了 0.9%。这不仅证明了三分支注意力模块的有效性，还显示出其在复杂的飞机蒙皮缺陷检测任务中的实用价值。与此同时，其他使用不同注意力模块的网络在同一数据集上的表现相对较差，进一步显示了三分支注意力模块在挖掘和利用特征方面的出色能力。

表 1 添加不同注意力算法的结果比较

Table 1 Comparison of results in different attention methods %

算法	查全率	mAP0.5
原始的YOLOv8算法	83.2	85.1
改进算法(+Bi-Level Routing[29])	83.2	85.1
改进算法(+CBAM)	83.9	85.5
改进算法(+SENet)	83.5	86.0
改进算法(+SimAM[28])	84.2	86.0
改进算法(+三分支注意力)	84.4	86.0

同时，为确定如何最佳地将三分支注意力模块集成到 YOLOv8 中，对注意力模块的位置和数量进行了不同参数组合的实验，结果如表 2 所示。可知，将三分支注意力模块集成到骨干网络的第 5 层效果最佳。当三分支注意模块集成到其他位置或增加数量时，可能会降低模型的检测效果。本文推测这一现象可能有 2 个原因：①增加注意力模块的数量可能会导致数据冗余，从而影响检测效果；②骨干网络的较低层捕捉到的是简单、局部的特征，而较高层捕捉到的是更复杂、全局的特征。在这种情况下，将注意力模块放在较高层使网络更加关注于复杂特征的关键部分。

表 2 添加注意力模块不同位置和数量的结果比较

Table 2 Comparison of results at different positions and quantities for attention module %

添加数量	添加位置	查全率	mAP0.5	mAP0.5:0.95
0	基线模型	83.2	85.1	43.3
1	第3层	79.0	82.9	39.9
1	第4层	82.8	85.7	41.5
1	第5层	84.4	86.0	42.4
2	第3、4层	80.4	85.3	40.5
2	第3、5层	80.6	83.8	40.6
2	第4、5层	82.4	85.3	41.6
3	第3、4、5层	81.6	85.4	41.2

3.3.2 主干网络与检测头模块的实验比较

在不断追求模型性能提升的过程中，选择合适

的主干网络显得尤为关键。为确保本文选择能够达到预期的效果, 进一步与其他常见网络结构进行对比实验。如表 3 所示[30-31], 经过测试, 采用残差块作为主干网络的模型在各项评估指标上都展现了卓越的性能, 查全率提高了 2.5%, mAP0.5 提高了 2.3%, 超越了其他常用网络结构。这一实验结果进一步证实了残差块在处理深度神经网络中的信息流动和梯度消失问题时的高效性。

表 3　添加不同主干网络的结果比较

Table 3　Comparison of results with different backbone networks　%

算法	查全率	mAP0.5	mAP0.5:0.95
原始的YOLOv8算法	83.2	85.1	43.3
改进算法(+EfficientViT[30])	79.9	82.0	36.0
改进算法(+Fasternet[31])	79.8	80.8	38.4
改进算法(+残差块)	85.7	87.4	48.2

同时, 本文还对 YOLOv8 的检测头数量进行了深入的探究和实验研究, 分别在 2 个数据集上针对 4 个和 3 个改进检测头的配置进行了对比测试。如表 4 所示, 尽管理论上使用 4 个改进检测头的配置为本文带来了更小目标的检测能力, 但从结果上看, 3 个改进检测头的查全率和平均精度均值比 4 个改进检测头要高。这一发现为本文后续模型选择和优化提供了重要的参考依据, 也证明了增加模型复杂度不一定就能带来更好的性能, 简化和精细调整可能是更佳的选择。

表 4　不同检测头数量的结果比较

Table 4　Comparison of results with different numbers of head counts

数据集	方法	查全率/%	mAP0.5/%	检测速度/(帧·s^{-1})
飞机蒙皮缺陷数据集	基线模型	83.2	85.1	169
	3个改进检测头	83.2	86.2	136
	4个改进检测头	82.5	83.6	90
VOC2007数据集	基线模型	71.9	77.5	312
	3个改进检测头	72.0	78.0	222
	4个改进检测头	71.7	77.8	126

3.3.3　消融实验比较

为验证各模块在网络中的优化效果, 进行了消融实验, 结果如表 5 和图 13 所示。图中, 图例中数字表示平均精度均值。从表 5 可以看到, 对于复杂场景下的飞机蒙皮缺陷小目标检测, 改进算法在各个阶段均展现了优化效果。数据证明, 提出的 3 种改进算法都极为有效: ①改进的特征融合方法有效避免了在整个学习过程中因位置信息而错失小目标的情况; ②主干网络优化有效地在增加网络复杂度需求与计算负担之间找到了平衡, 使网络能够更深入地学习特征表示, 不过多增加训练和计算时间; ③检测头的结构优化将注意力机制与任务需求相结合, 动态调整参数, 得到更丰富的特征表示。实验结果进一步证实, 对算法各阶段的改进均有助于提高模型的学习能力。

表 5　消融实验结果比较

Table 5　Comparison of results of ablation experiment　%

残差块	三分支注意力模块	检测头	查全率	mAP0.5	mAP0.5:0.95
			83.2	85.1	43.3
√			85.7	87.4	48.2
	√		84.4	86.0	42.4
		√	83.2	86.2	44.3
√	√		85.7	87.6	48.5
√		√	84.6	86.6	47.3
	√	√	83.4	87.3	45.1
√	√	√	86.9	88.7	49.0

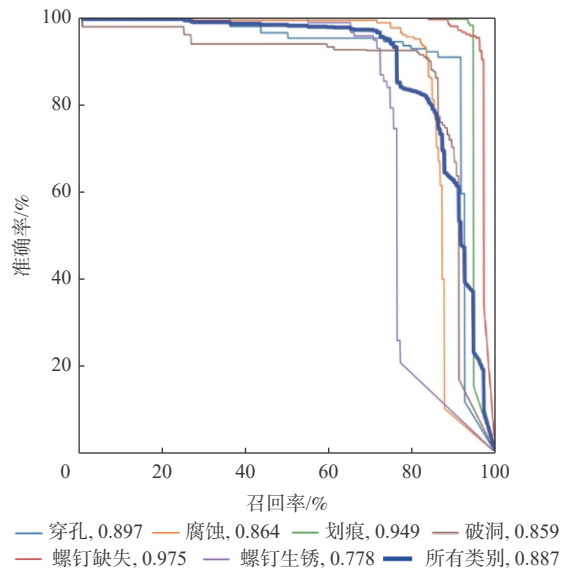


图 13　改进模型在飞机蒙皮缺陷数据集上的准确率-召回率曲线

Fig. 13　Precision-recall curves of improved model on aircraft skin defect dataset

3.3.4　其他算法的实验比较

使用 5 种流行的检测模型 YOLOv3、YOLOv5、YOLOv6、YOLOv7 和 YOLOv8, 在飞机蒙皮缺陷数据集和 VOC2007 数据集上与本文提出的改进模型进行对比实验, 结果如表 6[32-34] 和图 14 所示, 表中, RTD-YOLOv8 表示本文提出的改进模型。在所有参与比较的模型中, 本文算法表现最为出色, 其在 2 个关键指标上均达到了非常高的水平。在飞机蒙皮缺陷数据集上获得了 88.70% 的平均精度均值和 86.90% 的查全率, 而在 VOC2007 数据集上, 这 2 项指标也分别达到了 80.4% 和 74.1%, 比 YOLOv8 算

表 6 与其他算法的结果比较

Table 6 Comparison of results with other algorithms

数据集	模型	查全率/%	mAP0.5/%	mAP0.5:0.95/%	检测速度/(帧·s^{-1})
飞机蒙皮缺陷数据集	YOLOv3-tiny[32]	73.0	75.0	32.8	300
	YOLOv5-s	84.8	80.0	43.0	240
	YOLOv6[33]	80.0	82.6	37.8	243
	YOLOv7-tiny[34]	74.4	76.5	30.5	260
	YOLOv8-n	83.2	85.1	43.3	135
	YOLOv8-s	86.5	88.2	48.6	100
	RTD-YOLOv8	86.9	88.7	49.0	60
VOC2007 数据集	YOLOv3-tiny[32]	55.8	56.8	28.3	322
	YOLOv5-s	71.9	77.0	53.0	263
	YOLOv6[33]	67.8	70.3	46.8	270
	YOLOv7-tiny[34]	54.8	54.8	29.3	357.1
	YOLOv8-n	71.9	77.5	55.8	312.5
	YOLOv8-s	73.4	79.9	57.9	196
	RTD-YOLOv8	74.1	80.4	58.0	167

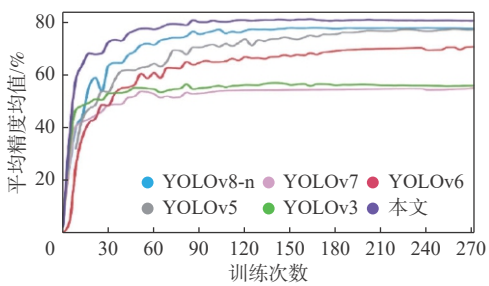


图 14 VOC2007 数据集平均精度均值对比

Fig. 14 Comparison of mAP on VOC2007 dataset

法提高了 2.9% 和 2.2%。这些数据充分证明了改进模型在飞机蒙皮检测任务上的优越性能和高效准确度，并且具有较好的泛化能力。

4 结 论

1）提出切片推理数据增强方法与马赛克数据增强方法结合的技术，得到相对较大的飞机蒙皮缺陷数据集，解决数据不足和目标过小且密集的问题。

2）通过将残差块引入主干部分进行特征提取，有效提升模型的检测性能，无需大幅增加额外的训练时间；通过三分支注意力模块优化模型层次结构，降低小目标样本的误检率和漏检率；在检测头部分将不同维度的信息进行整合，提高难目标样本的识别准确率。平均精度均值由 85.1% 提升至 88.7%，提高了 3.6%；查全率值由 83.2% 提升至 86.9%，提高了 3.7%。

3）在检测精度方面，改进后的 YOLOv8 算法检测的平均精度均值和查全率高于 YOLOv3、YOLOv5、YOLOv6、YOLOv7 和 YOLOv8 算法，更适合用于检测飞机蒙皮缺陷。

参考文献（References）

[1] 李婉婷. 基于关键点的飞机表面伤痕检测及后处理算法研究[D]. 北京: 北京邮电大学, 2021.
LI W T. Research on aircraft surface damage detection and post-processing algorithm based on key points[D]. Beijing: Beijing University of Posts and Telecommunications, 2021(in Chinese).

[2] 张研, 田军委, 刘青, 等. 基于视觉显著性的飞机蒙皮缺陷检测方法[J]. 激光杂志, 2022, 43(3): 61-64.
ZHANG Y, TIAN J W, LIU Q, et al. Aircraft skin defect detection method based on visual saliency[J]. Laser Journal, 2022, 43(3): 61-64 (in Chinese).

[3] TAO X, ZHANG D P, MA W Z, et al. Automatic metallic surface defect detection and recognition with convolutional neural networks [J]. Applied Sciences, 2018, 8(9): 1575.

[4] GAROUSI V, TASLI S, SERTEL O, et al. Automated testing of simulation software in the aviation industry: an experience report[J]. IEEE Software, 2018, 36(4): 63-75.

[5] HAGGERTY J, DEFER E, DE LAAT A, et al. Detecting clouds associated with jet engine ice crystal icing[J]. Bulletin of the American Meteorological Society, 2019, 100(1): 31-40.

[6] SHANG J Z, SATTAR T, CHEN S W, et al. Design of a climbing robot for inspecting aircraft wings and fuselage[J]. Industrial Robot, 2007, 34(6): 495-502.

[7] DONG X H, TAYLOR C J, COOTES T F. Automatic aerospace weld inspection using unsupervised local deep feature learning[J]. Knowledge-Based Systems, 2021, 221: 106892.

[8] MEISTER S, WERMES M A M, STÜVE J, et al. Review of image segmentation techniques for layup defect detection in the automated fiber placement process[J]. Journal of Intelligent Manufacturing, 2021, 32(8): 2099-2119.

[9] TABERNIK D, ŠELA S, SKVARČ J, et al. Segmentation-based deep-learning approach for surface-defect detection[J]. Journal of Intelligent Manufacturing, 2020, 31(3): 759-776.

[10] 赵鹏飞, 谢林柏, 彭力. 融合注意力机制的深层次小目标检测算法[J]. 计算机科学与探索, 2022, 16(4): 927-937.
ZHAO P F, XIE L B, PENG L. Deep small object detection algorithm integrating attention mechanism[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(4): 927-937(in Chinese).

[11] 张宝朋, 康谦泽, 李佳萌, 等. 轻量化的 YOLOv4 目标检测算法[J]. 计算机工程, 2022, 48(8): 206-214.
ZHANG B P, KANG Q Z, LI J M, et al. Lightweight YOLOv4 target detection algorithm[J]. Computer Engineering, 2022, 48(8): 206-214(in Chinese).

[12] 张海涛, 张梦. 引入通道注意力机制的 SSD 目标检测算法[J]. 计算机工程, 2020, 46(8): 264-270.
ZHANG H T, ZHANG M. SSD target detection algorithm with channel attention mechanism[J]. Computer Engineering, 2020, 46(8): 264-270(in Chinese).

[13] 候瑞环, 杨喜旺, 王智超, 等. 一种基于 YOLOv4-TIA 的林业害虫实时检测方法[J]. 计算机工程, 2022, 48(4): 255-261.

HOU R H, YANG X W, WANG Z C, et al. A real-time detection method for forestry pests based on YOLOv4-TIA[J]. Computer Engineering, 2022, 48(4): 255-261(in Chinese).

[14] 蒋镕圻, 彭月平, 谢文宣, 等. 嵌入 scSE 模块的改进 YOLOv4 小目标检测算法[J]. 图学学报, 2021, 42(4): 546-555.
JIANG R Q, PENG Y P, XIE W X, et al. Improved YOLOv4 small target detection algorithm with embedded scSE module[J]. Journal of Graphics, 2021, 42(4): 546-555(in Chinese).

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2023-11-01]. https://arxiv.org/abs/2004.10934.

[16] 赵恩兴, 王超. 基于 YOLOv8 的交通信号灯识别[J]. 人工智能与机器人研究, 2023, 12(3): 246-254.
ZHAO E X, WANG C. Traffic signal recognition based on YOLOv8 [J]. Research in Artificial Intelligence and Robotics, 2023, 12(3): 246-254(in Chinese).

[17] LIU H Y, DUAN X H, CHEN H N, et al. DBF-YOLO: UAV small targets detection based on shallow feature fusion[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2023, 18(4): 605-612.

[18] LIU H Y, SUN F Q, GU J, et al. SF-YOLOv5: a lightweight small object detection algorithm based on improved feature fusion mode[J]. Sensors, 2022, 22(15): 5817.

[19] 旷可嘉. 深度学习及其在航空发动机缺陷检测中的应用研究[D]. 广州: 华南理工大学, 2017.
KUANG K J. Deep learning and its application in aero-engine defect detection[D]. Guangzhou: South China University of Technology, 2017(in Chinese).

[20] ZHANG D Y, WEI P H, TAN M W, et al. Investigation of aircraft surface defects detection based on YOLO neural network[C]// Proceedings of the 7th International Conference on Information Science and Control Engineering. Piscataway: IEEE Press, 2020: 781-785.

[21] LI Y D, HAN Z Q, XU H Y, et al. YOLOv3-lite: a lightweight crack detection network for aircraft structure based on depthwise separable convolutions[J]. Applied Sciences, 2019, 9(18): 3781.

[22] JOVANCEVIC I, LARNIER S, ORTEU J J, et al. Automated exterior inspection of an aircraft with a pan-tilt-zoom camera mounted on a mobile robot[J]. Journal of Electronic Imaging, 2015, 24(6): 061110.

[23] GAO Q J, WANG W J, NIU G C. Design bionic structure and analysis of kinematics for aircraft fuel tank inspection robot[J]. Applied Mechanics and Materials, 2013, 278-280: 594-598.

[24] AKYON F C, ALTINUC S O, TEMIZEL A. Slicing aided hyper inference and fine-tuning for small object detection[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2022: 966-970.

[25] ALIN A Y, YUANA K A. Data augmentation method on drone object detection with YOLOv5 algorithm[C]//Proceedings of the 8th International Conference on Informatics and Computing. Piscataway: IEEE Press, 2023: 1-6.

[26] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.

[27] DAI X Y, CHEN Y P, XIAO B, et al. Dynamic head: unifying object detection heads with attentions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 7369-7378.

[28] YANG L X, ZHANG R Y, LI L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural network[C]// Proceedings of the International Conference on Machine Learning. San Diego: JMLR, 2021, 139: 11863-11874.

[29] ZHU L, WANG X J, KE Z H, et al. BiFormer: vision Transformer with bi-level routing attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 10323-10333.

[30] CAI H, LI J Y, HU M Y, et al. EfficientViT: multi-scale linear attention for high-resolution dense prediction [EB/OL]. (2022-05-29) [2023-11-01]. https://arxiv.org/abs/2205.14756.

[31] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 12021-12031.

[32] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL].(2018-04-08)[2023-11-01].https://arxiv.org/abs/1804.02767.

[33] LI C, LI L, JIANG H, et al. YOLOV6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-07) [2023-11-01]. https://arxiv.org/abs/2209.02976.

[34] WANG C Y, BOCHKOVSKIY A, LIAO H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2023: 7464-7475.

Aircraft skin defect detection algorithm based on enhanced YOLOv8

ZHANG Dongping[1,*], WANG Zhutao[1], XIA Yuejian[1], XU Yunchao[1], LIN Lili[2]

(1. College of Information Engineering, China Jiliang University, Hangzhou 310018, China;
2. School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China)

Abstract: In order to solve the problem that traditional aircraft skin defect detection relies on human eye observation, which leads to reduced efficiency due to easy fatigue of the human eye and limited individual cognition, an aircraft skin defect detection algorithm based on improved YOLOv8 is proposed. Improve the data improvement strategy and propose a new one that combines slice reasoning with mosaic. Integrate the residual block into the feature extraction network to enhance the network expression ability and improve the accuracy of the model in aircraft skin defect detection tasks. Use the triplet attention module to strengthen the feature fusion network and lower the false and missed detection rates of small target samples. Optimize the structure of the detection head so that the network can better effectively combine shallow information with depth information. On the aircraft skin defect data set, experimental results indicate that the revised algorithm's mean average precision (mAP) and recall rate have increased by 3.6% and 3.7%, respectively, in comparison to the most recent YOLOv8 algorithm. The mAP and recall rate on the public data set VOC2007 increased by 2.9% and 2.2%, respectively.

Keywords: YOLOv8 algorithm; surface defect detection; data augmentation; object detection; attention mechanism

Received: 2023-11-20; **Accepted**: 2023-12-01; **Published Online**: 2024-01-30 17:31

URL: link.cnki.net/urlid/11.2625.V.20240130.0955.001

Foundation items: Zhejiang Key R & D Project of China (2022C01005,2023C01032,2023C01030)

* **Corresponding author**. E-mail: 06a0303103@cjlu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0766

星凸形不规则形状多扩展目标PMBM滤波器

陈辉[1,*]，刘孟波[1]，连峰[2]，韩崇昭[2]

(1. 兰州理工大学 自动化与电气工程学院，兰州 730050；2. 西安交通大学 自动化科学与工程学院，西安 710049)

摘 要：针对复杂不确定环境下具有不规则形状的多扩展目标跟踪问题，提出一种星凸形不规则形状多扩展目标泊松多伯努利混合(PMBM)滤波算法。分别利用泊松点过程(PPP)和多伯努利混合(MBM)对未知目标集合和现有目标集合进行建模，有效表示潜在目标信息，并建立高效的多目标密度递推形式；采用随机超曲面模型对任意星凸形扩展目标的量测源分布进行精准建模，利用最优非线性滤波求解高度非线性的伪量测方程，在推导扩展目标泊松多伯努利混合(ETPMBM)滤波器的基础上，详细推导并提出高斯混合星凸形不规则形状多扩展目标PMBM滤波算法；该算法能够建立更紧凑的多伯努利全局假设，从而高效递推包含不规则形状等多特征信息的多扩展目标概率密度。通过多扩展目标和多群目标跟踪仿真实验验证了所提算法的有效性。

关 键 词：多扩展目标跟踪；随机超曲面模型；泊松多伯努利混合；最优非线性滤波；高斯混合

中图分类号：TP274

文献标志码：A **文章编号**：1001-5965(2026)01-0049-12

扩展目标跟踪(extended target tracking, ETT)[1-3]不同于传统雷达“点目标跟踪”问题[4-5]，可借助目标的多源量测信息对目标进行更深层次的识别。此外，现代军事和民用的群智协同使得群目标跟踪(group target tracking, GTT)问题被广泛关注，群目标通常由具有相似运动属性且按照一定规律协同运动的多个子目标组成，如飞机编队、舰艇编队、地面编队等。如果把群目标看成一个整体，只关注群的形态演变和群的整体运动特性，则GTT和ETT就非常类似。多目标跟踪(multiple target tracking, MTT)问题求解的最大困难在于复杂的数据关联决策，而由于每采样时刻传感器采集到扩展目标的多个量测，进一步加剧了MTT关联关系的模糊化和复杂性，为研究设计稳健的多扩展目标跟踪(multiple extended target tracking, METT)算法带来了非常大的困难和挑战。因此，如何在合并有杂波、漏检、噪声量测、目标运动和扩展形态随机演变及多扩展目标(multiple extended target, MET)共存等复杂不确定性环境下，研究更为有效的鲁棒自适应METT滤波器，对于现代目标跟踪系统的性能提升无疑具有重大的现实意义。

基于有限集统计(finite set statistics, FISST)理论的MTT算法的提出，给目标跟踪领域带来了新的生机活力。基于FISST的MTT算法可依据后验概率密度的共轭先验特性进行归类。其中，不满足共轭先验的经典矩递推滤波算法包括多目标密度矩近似的概率假设密度(probability hypothesis density, PHD)滤波器[6-8]及势概率假设密度(cardinalized probability hypothesis density, CPHD)滤波器[9]、传播后验概率密度的多伯努利(multi-Bernoulli, MBer)滤波器[10-11]。这些算法的计算复杂程度较低，并有效避免了数据关联的复杂问题，但由于总体的多目标密度的近似性带来了信息的丢失，一定程度上降低了MTT精度。满足共轭先验的滤波算法是目前

收稿日期：2023-11-24；录用日期：2023-12-15；网络出版时间：2024-01-03 12:02
网络出版地址：link.cnki.net/urlid/11.2625.V.20240102.1351.001
基金项目：国家自然科学基金(62163023)；甘肃省基础研究创新群体(25JRRA058)；中央引导地方科技发展资金项目(25ZYJA040)；甘肃省重点人才项目(2024RCXM86)

*通信作者. E-mail：huich78@hotmail.com

引用格式：陈辉，刘孟波，连峰，等. 星凸形不规则形状多扩展目标PMBM滤波器[J]. 北京航空航天大学学报，2026，52(1)：49-60.
CHEN H，LIU M B，LIAN F，et al. Star convex irregular shape multi-extended target PMBM filter[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：49-60 (in Chinese).

FISST-MTT研究方法的主流，主要包括广义标签多伯努利(generalized labeled multi-Bernoulli, GLMB)滤波器[12-14]及泊松多伯努利混合(Poisson multi-Bernoulli mixture, PMBM)滤波器[15-16]。相较于GLMB滤波器，PMBM滤波器由于具有更加高效的滤波结构和较少的全局假设，在整体跟踪性能和稳定性等方面的提升使得其在解决MTT问题上脱颖而出。

对多扩展目标形状的有效估计是METT的核心问题。扩展目标可以被建模为圆[17-18]、椭圆、棒形[19]、矩形[20]等基本轮廓，而随机矩阵模型(random matrix model, RMM)[21-22]是其中应用最为广泛的方法。RMM刻画了目标的椭圆形轮廓，通过高斯逆Wishart联合分布建模其运动过程与空间分布特性[23]，能有效获取目标的大小和方向信息。但是，不规则形状显然更具辨识度。随机超曲面模型(random hypersurface model, RHM)[24-27]是一种能够对不规则形状进行建模的有效方法，该方法假设量测源位于利用尺度因子进行缩放的目标轮廓上，用径向函数对任意不规则形状目标进行描述，并利用傅里叶级数展开实现不规则轮廓建模的参数化，其可以根据不同的应用场景，动态自适应集群目标的不规则外形，可以跟踪形状先验未知的目标。为了更好地解决多扩展目标的跟踪问题，已有学者将基于FISST的MTT算法应用到扩展目标中[28-31]，并取得了不错的研究成果。本文在多扩展目标PMBM滤波器的基础上拓展RHM的鲁棒自适应不规则目标外形的能力，深入研究星凸不规则形状METT滤波器，这为复杂环境下多集群目标的精确跟踪与深度辨识提供了坚实的科学理论依据。

鉴于此，本文首先利用泊松点过程(Poisson point process, PPP)和多伯努利混合(multi-Bernoulli mixture, MBM)分别对未知目标集合和现有目标集合进行建模；其次，利用RHM建模不规则星凸形扩展目标轮廓，利用泊松似然模型和RHM联合推导提出星凸形多扩展目标PMBM滤波器，并提出该滤波器的高斯混合(Gaussian mixture, GM)实现方法；最后，通过设计复杂环境下具有不规则形状的多扩展目标和多群目标跟踪实验，验证了本文算法的有效性和实用性。

1 随机超曲面模型

在扩展目标中，量测建模分为2步：建立合适的量测源模型，产生多个量测源，量测源合并量测噪声生成目标的多个量测。量测建模过程如图1所示。传感器量测方程表示为

$$\boldsymbol{z}_{k,l}=\hat{\boldsymbol{z}}_{k,l}+\boldsymbol{v}_{k,l} \tag{1}$$

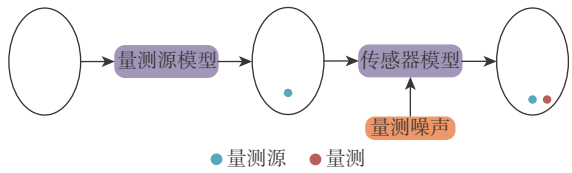


图1 量测建模过程

Fig. 1 Measurement modeling process

式中：$\boldsymbol{z}_{k,l}$为k时刻位置随机的第l个量测；$\hat{\boldsymbol{z}}_{k,l}$为$k$时刻位置随机的第$l$个量测源；$\boldsymbol{v}_{k,l}$为量测噪声且满足$\boldsymbol{v}_{k,l}\sim N(0,\boldsymbol{R}_k)$，$\boldsymbol{R}_k$为量测噪声协方差矩阵。

RHM的基本思想是：假设量测源位于扩展目标形状边界经由尺度因子$s_{k,l}$得到的缩放版本上，并将$s_{k,l}$建模成随机变量，RHM建模如图2所示。本文主要利用RHM对星凸形扩展目标量测源进行准确建模。

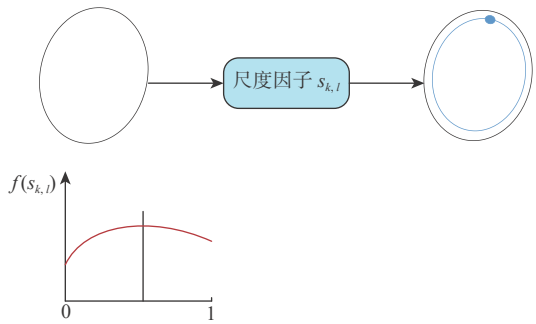


图2 随机超曲面模型建模

Fig. 2 Random hypersurface model modeling

星凸形定义为：若一个集合S是以$\boldsymbol{m}$为中心的星凸集合，S中任意点$\boldsymbol{x}$到$\boldsymbol{m}$的线段仍包含在该集合中，则集合S所描述的目标形状为星凸形，如图3所示。特别地，星凸形$S(\boldsymbol{p}_k)$(无论角长是否对称)可以由一维径向函数$r(\boldsymbol{p}_k,\phi)$进行描述，$\boldsymbol{p}_k$和ϕ分别为扩展目标形状参数和对应角度，径向函数示意图如图4所示。由n阶傅里叶级数展开得到的径向函数$r(\boldsymbol{p}_k,\phi)$表示如下：

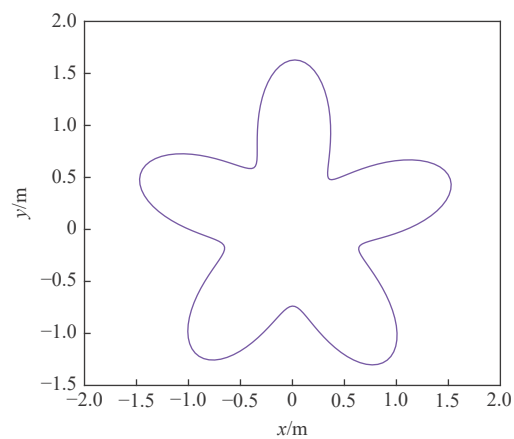


图3 星凸形

Fig. 3 Star convex shape

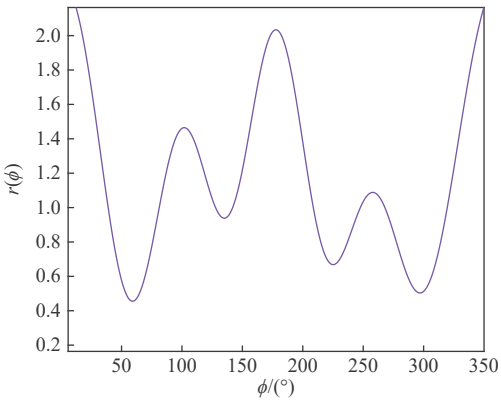


图 4　径向函数

Fig. 4　Radial function

$$r(\boldsymbol{p}_k,\phi)=\frac{1}{2}a_k^{(0)}+\sum_{j=1}^{n}\left(a_k^{(j)}\cos(j\phi)+b_k^{(j)}\sin(j\phi)\right)=\boldsymbol{B}(\phi)\boldsymbol{p}_k \tag{2}$$

式中: $a_k^{(j)}$为 k 时刻傅里叶级数展开的第 j 次谐波余弦分量的系数; $b_k^{(j)}$为 k 时刻傅里叶级数展开的第 j 次谐波正弦分量的系数; $a_k^{(0)}$为 k 时刻傅里叶级数展开的直流分量。

$$\boldsymbol{B}(\phi)=\left[\frac{1}{2},\cos\phi,\sin\phi,\cdots,\cos(n\phi),\sin(n\phi)\right] \tag{3}$$

$$\boldsymbol{p}_k=\left[a_k^{(0)},a_k^{(1)},b_k^{(1)},\cdots,a_k^{(n)},b_k^{(n)}\right]^{\mathrm{T}} \tag{4}$$

选择傅里叶级数展开来对表述星凸外形的径向函数进行表示, 是因为这类方法计算复杂程度低、结构清晰且具有从粗到细的描述能力, 是一种经典的形状描述方法。从粗到细的描述能力是指傅里叶级数的低阶系数描述了星凸形状的粗略形状特征, 而高阶系数则表征了星凸形状的细节形状特征。正因为傅里叶级数展开描述径向函数的诸多优点, 可以借由此类方法来表示一般意义下的星凸轮廓物体。

在 ETT 算法中, 扩展目标状态向量可以表示为 $\boldsymbol{x}_k=\left[\boldsymbol{m}_k^{\mathrm{T}},(\boldsymbol{x}_k^*)^{\mathrm{T}},\boldsymbol{p}_k^{\mathrm{T}}\right]^{\mathrm{T}}$, $\boldsymbol{m}_k$为目标质心位置, $\boldsymbol{x}_k^*$为目标运动参数, $\boldsymbol{p}_k$为目标形状参数。给定一个星凸集合 $S(\boldsymbol{p}_k)$, 利用 RHM 对该扩展目标量测源进行建模, 得到量测源模型的数学表达为

$$\hat{\boldsymbol{z}}_{k,l}=\boldsymbol{m}_k+s_{k,l}\overline{\boldsymbol{S}}(\boldsymbol{p}_k)=\boldsymbol{m}_k+s_{k,l}r(\boldsymbol{p}_k,\phi_{k,l})\boldsymbol{e}(\phi_{k,l}) \tag{5}$$

式中: 随机变量 $s_{k,l}\in[0,1]$为尺度因子; $\overline{\boldsymbol{S}}(\boldsymbol{p}_k)$为星凸形 $S(\boldsymbol{p}_k)$的边界; $\phi_{k,l}=\angle(\hat{\boldsymbol{z}}_{k,l}-\boldsymbol{m}_k)$为量测源到质心的向量与 x 轴形成的夹角; $\boldsymbol{e}(\phi_{k,l})=[\cos\phi_{k,l},\sin\phi_{k,l}]^{\mathrm{T}}$。

式 (5) 中的 $\phi_{k,l}$通常不能直接获取, 可以用 $\boldsymbol{z}_{k,l}$到质心估计 $\hat{\boldsymbol{m}}_k$的向量与 x 轴之间的夹角来给出, 即 $\phi_{k,l}\approx\hat{\phi}_{k,l}=\angle(\boldsymbol{z}_{k,l}-\hat{\boldsymbol{m}}_k)$。则由式 (1)、式 (2) 和式 (5) 可得

$$\boldsymbol{z}_{k,l}-\boldsymbol{m}_k=s_{k,l}\boldsymbol{B}(\hat{\phi}_{k,l})\boldsymbol{p}_k\boldsymbol{e}^{\mathrm{T}}(\hat{\phi}_{k,l})+\boldsymbol{v}_{k,l} \tag{6}$$

由式 (6) 可得

$$\left\|\boldsymbol{z}_{k,l}-\boldsymbol{m}_k\right\|^2=s_{k,l}^2\left\|\boldsymbol{B}(\hat{\phi}_{k,l})\boldsymbol{p}_k\right\|^2+\left\|\boldsymbol{v}_{k,l}\right\|^2+2s_{k,l}\boldsymbol{B}(\hat{\phi}_{k,l})\boldsymbol{p}_k\boldsymbol{e}^{\mathrm{T}}(\hat{\phi}_{k,l})\boldsymbol{v}_{k,l} \tag{7}$$

可将式 (7) 整理为非线性伪量测方程, 如下:

$$\begin{aligned}0=h(\boldsymbol{x}_k,\boldsymbol{v}_{k,l},s_{k,l},\boldsymbol{z}_{k,l})=&\\ s_{k,l}^2\left\|\boldsymbol{B}(\hat{\phi}_{k,l})\boldsymbol{p}_k\right\|^2+\left\|\boldsymbol{v}_{k,l}\right\|^2-\left\|\boldsymbol{z}_{k,l}-\boldsymbol{m}_k\right\|^2+&\\ 2s_{k,l}\boldsymbol{B}(\hat{\phi}_{k,l})\boldsymbol{p}_k\boldsymbol{e}^{\mathrm{T}}(\hat{\phi}_{k,l})\boldsymbol{v}_{k,l}&\end{aligned} \tag{8}$$

伪量测方程建立了 $\boldsymbol{x}_k$、$\boldsymbol{v}_{k,l}$、$s_{k,l}$、$\boldsymbol{z}_{k,l}$之间的关系, 为目标形状估计奠定了基础。

2　基于 RHM 的泊松多伯努利混合滤波器

标准的扩展目标 PMBM (extended target PMBM, ETPMBM)[32] 的递推过程未将扩展目标形状的建模和递推估计考虑其中。本节结合 RHM 对具有复杂不规则外形的 METT 问题展开研究。利用 RHM, 在 MET-PMBM 的基础上提出星凸形不规则形状 METT 算法, 并最终提出其 GM 实现形式。由于 RHM 的伪量测方程的非线性, 本文利用滤波精度和数值稳定性俱佳的容积卡尔曼滤波 (cubature Kalman filter, CKF)[33] 求解伪量测方程。下面详细给出 RHM-GM-ETPMBM 滤波算法的递推过程。

给出一般假设:

1) 有效检测概率 $\overline{p}_{\mathrm{D}}(\boldsymbol{x})$和存活概率 $p_{\mathrm{S}}(\boldsymbol{x})$为常数, 即

$$\begin{cases}\overline{p}_{\mathrm{D}}(\boldsymbol{x})=\overline{p}_{\mathrm{D}}\\ p_{\mathrm{S}}(\boldsymbol{x})=p_{\mathrm{S}}\end{cases} \tag{9}$$

2) 目标动力学遵循线性高斯模型, 即

$$\begin{cases}\boldsymbol{x}_k=\boldsymbol{F}\boldsymbol{x}_{k-1}+\boldsymbol{\varepsilon}_k\\ \boldsymbol{\varepsilon}_k\sim N(\boldsymbol{\varepsilon}_k;0,\boldsymbol{Q})\end{cases} \tag{10}$$

式中: $\boldsymbol{F}$为状态转移矩阵; $\boldsymbol{\varepsilon}_k$表示均值为 0、协方差为 $\boldsymbol{Q}$的高斯白噪声。

3) 新生目标强度函数为 GM 形式, 即

$$D^{\mathrm{b}}(\boldsymbol{x})=\sum_{i=1}^{n_k^{\mathrm{b}}}w^{\mathrm{b},i}N(\boldsymbol{x};\boldsymbol{\mu}^{\mathrm{b},i},\boldsymbol{P}^{\mathrm{b},i}) \tag{11}$$

式中: $w^{\mathrm{b},i}$、$\boldsymbol{\mu}^{\mathrm{b},i}$和 $\boldsymbol{P}^{\mathrm{b},i}$分别为新生目标高斯分量的权重、均值和协方差。

4) 未知目标强度函数为 GM 形式, 即

$$D_{k-1}^{\mathrm{u}}(\boldsymbol{x})=\sum_{i=1}^{n_{k-1}^{\mathrm{u}}}w_{k-1}^{\mathrm{u},i}N(\boldsymbol{x};\boldsymbol{\mu}_{k-1}^{\mathrm{u},i},\boldsymbol{P}_{k-1}^{\mathrm{u},i}) \tag{12}$$

式中: $w_{k-1}^{\mathrm{u},i}$、$\boldsymbol{\mu}_{k-1}^{\mathrm{u},i}$和 $\boldsymbol{P}_{k-1}^{\mathrm{u},i}$分别为未知目标高斯分量的权重、均值和协方差。

2.1　预 测 步

1) 将先验 PMBM 密度中 PPP 分量及 MBM 分

量参数化表示为

$$D_{k-1}^{\mathrm{u}},\{W_{k-1}^{j},\{r_{k-1}^{j,i},f_{k-1}^{j,i}\}_{i\in I^{j}}\}_{j\in J} \tag{13}$$

式中：D_{k-1}^{u}为$k-1$时刻PPP分量的后验强度；$r_{k-1}^{j,i}$和$f_{k-1}^{j,i}$分别为$k-1$时刻第j个全局假设对应的第i个伯努利分量的存在概率和概率密度；全局假设的个数为$|J|$；W_{k-1}^{j}为$k-1$时刻第j个全局假设的权值；I^{j}为第j个全局假设对应的伯努利索引集。

2）未知目标PPP分量强度预测。$k-1$时刻PPP分量的个数为$i=1,\cdots,n_{k-1}^{\mathrm{u}}$，则有

$$w_{k|k-1}^{\mathrm{u},i}=p_{\mathrm{S}}w_{k-1}^{\mathrm{u},i} \tag{14}$$

$$\boldsymbol{m}_{k|k-1}^{\mathrm{u},i}=\boldsymbol{F}\boldsymbol{\mu}_{k-1}^{\mathrm{u},i} \tag{15}$$

$$\boldsymbol{P}_{k|k-1}^{\mathrm{u},i}=\boldsymbol{F}\boldsymbol{P}_{k-1}^{\mathrm{u},i}\boldsymbol{F}^{\mathrm{T}}+\boldsymbol{Q} \tag{16}$$

3）新生目标预测。将其并入到PPP分量中，新生目标分量个数为$i=1,\cdots,n_{k}^{\mathrm{b}}$，有

$$w_{k|k-1}^{\mathrm{u},i+n_{k-1}^{\mathrm{u}}}=w^{\mathrm{b},i} \tag{17}$$

$$\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i+n_{k-1}^{\mathrm{u}}}=\boldsymbol{\mu}^{\mathrm{b},i} \tag{18}$$

$$\boldsymbol{P}_{k|k-1}^{\mathrm{u},i+n_{k-1}^{\mathrm{u}}}=\boldsymbol{P}^{\mathrm{b},i} \tag{19}$$

4）预测步得到的PPP强度函数中高斯分量个数为$n_{k|k-1}^{\mathrm{u}}=n_{k-1}^{\mathrm{u}}+n_{k}^{\mathrm{b}}$，PPP分量的预测强度可重新表示为

$$D_{k|k-1}^{\mathrm{u}}=\sum_{i=1}^{n_{k|k-1}^{\mathrm{u}}}w_{k|k-1}^{\mathrm{u},i}N(\boldsymbol{x};\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i},\boldsymbol{P}_{k|k-1}^{\mathrm{u},i}) \tag{20}$$

5）MBM分量预测，对于$i=1,\cdots,n_{k-1}$，有

$$W_{k|k-1}^{j}=W_{k-1}^{j} \tag{21}$$

$$r_{k|k-1}^{j,i}=p_{\mathrm{S}}r_{k-1}^{j,i} \tag{22}$$

$$f_{k|k-1}^{j,i}(\boldsymbol{x})=N(\boldsymbol{x};\boldsymbol{F}\boldsymbol{\mu}_{k-1}^{j,i},\boldsymbol{F}\boldsymbol{P}_{k-1}^{j,i}\boldsymbol{F}^{\mathrm{T}}+\boldsymbol{Q}) \tag{23}$$

2.2 更新步

标准扩展目标测量模型中，被检测到的扩展目标的单个量测$\boldsymbol{z}\in Z$建模为PPP且强度为$\gamma(\boldsymbol{x})\varphi_{z}(\boldsymbol{x})$，当量测集非空（$|Z|>0$）时，扩展目标的量测集似然为

$$l_{Z}(\boldsymbol{x})=p_{\mathrm{D}}(\boldsymbol{x})\mathrm{e}^{-\gamma(\boldsymbol{x})}\prod_{z\in Z}\gamma(\boldsymbol{x})\varphi_{z}(\boldsymbol{x}) \tag{24}$$

式中：$\gamma(\boldsymbol{x})$为扩展目标产生量测个数的期望；$\varphi_{z}(\boldsymbol{x})$为单个量测似然函数。

设检测概率为$p_{\mathrm{D}}(\boldsymbol{x})$，有效检测概率为扩展目标至少产生一个量测的概率，可以表示为$\overline{p}_{\mathrm{D}}(\boldsymbol{x})=p_{\mathrm{D}}(\boldsymbol{x})(1-\mathrm{e}^{-\gamma(\boldsymbol{x})})$，则漏检的概率可以表示为$\overline{q}_{\mathrm{D}}(\boldsymbol{x})=1-p_{\mathrm{D}}(\boldsymbol{x})(1-\mathrm{e}^{-\gamma(\boldsymbol{x})})$。

在扩展目标跟踪过程中，同一时刻，传感器采样到扩展目标的多个量测，进一步加剧了数据关联的复杂程度。采用距离划分法有效划分所有量测，结合泊松似然模型以量测集（簇）的形式对目标状态进行更新。

未关联到量测下的PPP目标更新为

$$D_{k}^{\mathrm{u}}(\boldsymbol{x})=\sum_{i=1}^{n_{k|k-1}^{\mathrm{u}}}(1-\overline{p}_{\mathrm{D}})w_{k|k-1}^{\mathrm{u},i}N(\boldsymbol{x};\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i},\boldsymbol{P}_{k|k-1}^{\mathrm{u},i}) \tag{25}$$

关联到量测的PPP目标更新为

$$L_{\boldsymbol{C}}^{\mathrm{u}}=\sum_{i=1}^{n_{k|k-1}^{\mathrm{u}}}w_{k|k-1}^{\mathrm{u},i}p_{\mathrm{D}}l_{\boldsymbol{C}}^{\mathrm{u},i} \tag{26}$$

$$l_{\boldsymbol{C}}^{\mathrm{u},i}=\mathrm{e}^{-\gamma(\boldsymbol{x})}\prod_{z\in\boldsymbol{C}}\gamma(\boldsymbol{x})\varphi_{z}(\boldsymbol{x}) \tag{27}$$

$$r_{\boldsymbol{C}}^{\mathrm{u}}=\begin{cases}1 & |\boldsymbol{C}|>1\\ \dfrac{L_{\boldsymbol{C}}^{\mathrm{u}}}{\kappa+L_{\boldsymbol{C}}^{\mathrm{u}}} & |\boldsymbol{C}|=1\end{cases} \tag{28}$$

$$f_{\boldsymbol{C}}^{\mathrm{u}}(\boldsymbol{x})=\frac{\sum_{i=1}^{n_{k|k-1}^{\mathrm{u}}}w_{k|k-1}^{\mathrm{u},i}p_{\mathrm{D}}l_{\boldsymbol{C}}^{\mathrm{u},i}N(\boldsymbol{x};\boldsymbol{\mu}_{k}^{\mathrm{u},i,\boldsymbol{C}},\boldsymbol{P}_{k}^{\mathrm{u},i,\boldsymbol{C}})}{\sum_{i=1}^{n_{k|k-1}^{\mathrm{u}}}w_{k|k-1}^{\mathrm{u},i}p_{\mathrm{D}}l_{\boldsymbol{C}}^{\mathrm{u},i}} \tag{29}$$

式中：$r_{\boldsymbol{C}}^{\mathrm{u}}$和$f_{\boldsymbol{C}}^{\mathrm{u}}(\boldsymbol{x})$分别表示PPP项更新为MBM项的存在概率和概率密度；$L_{\boldsymbol{C}}^{\mathrm{u}}$为量测划分单元$\boldsymbol{C}$的似然函数；$l_{\boldsymbol{C}}^{\mathrm{u},i}$为未知目标的单个似然函数；$\kappa=\lambda c(\boldsymbol{z})$为杂波强度，$\lambda$为泊松过程杂波平均数，$c(\boldsymbol{z})$为杂波的空间概率密度。

MBM部分漏检更新，即

$$L_{\varnothing}^{j,i}=1-r_{k|k-1}^{j,i}+r_{k|k-1}^{j,i}(1-\overline{p}_{\mathrm{D}}) \tag{30}$$

$$r_{\varnothing}^{j,i}=\frac{r_{k|k-1}^{j,i}(1-\overline{p}_{\mathrm{D}})}{1-r_{k|k-1}^{j,i}+r_{k|k-1}^{j,i}(1-\overline{p}_{\mathrm{D}})} \tag{31}$$

$$f_{\varnothing}^{j,i}(\boldsymbol{x})=N(\boldsymbol{x};\boldsymbol{\mu}_{k|k-1}^{j,i},\boldsymbol{P}_{k|k-1}^{j,i}) \tag{32}$$

关联到量测的MBM部分更新，即

$$L_{\boldsymbol{C}}^{j,i}=r_{k|k-1}^{j,i}p_{\mathrm{D}}l_{\boldsymbol{C}}^{j,i} \tag{33}$$

$$l_{\boldsymbol{C}}^{j,i}=\mathrm{e}^{-\gamma(\boldsymbol{x})}\prod_{z\in\boldsymbol{C}}\gamma(\boldsymbol{x})\varphi_{z}(\boldsymbol{x}) \tag{34}$$

$$r_{\boldsymbol{C}}^{j,i}=1 \tag{35}$$

$$f_{\boldsymbol{C}}^{j,i}(\boldsymbol{x})=N(\boldsymbol{x};\boldsymbol{\mu}_{k}^{j,i,\boldsymbol{C}},\boldsymbol{P}_{k}^{j,i,\boldsymbol{C}}) \tag{36}$$

$$W_{k}^{j}=\frac{W_{k|k-1}^{j}\prod_{\boldsymbol{C}\in A}L_{\boldsymbol{C}}}{\sum_{j'\in J}\sum_{A'\in A^{j'}}W_{k|k-1}^{j'}\prod_{\boldsymbol{C}'\in A'}L_{\boldsymbol{C}'}} \tag{37}$$

式中：$A\in A^{j}$为数据关联空间某个具体的关联。

下面详细讨论关联到量测的PPP目标和关联到量测的MBM目标状态均值$\boldsymbol{\mu}_{k}^{\mathrm{u},i,\boldsymbol{C}}$、$\boldsymbol{\mu}_{k}^{j,i,\boldsymbol{C}}$和协方差$\boldsymbol{P}_{k}^{\mathrm{u},i,\boldsymbol{C}}$、$\boldsymbol{P}_{k}^{j,i,\boldsymbol{C}}$及似然函数$\varphi_{z}(\boldsymbol{x})$的求解。由于RHM中伪量测方程的高度非线性，采用CKF对其处理并

进行序贯更新。以求解关联到量测的 PPP 目标的均值和协方差为例, 详细求解步骤如下:

步骤 1　基于三阶球面径向容积准则, 计算基本容积点ξ_i:

$$\xi_i=\sqrt{\frac{n_{\mathrm{c}}}{2}}[1]_i \qquad i=1,2,\cdots,n_{\mathrm{c}} \tag{38}$$

式中: n_{c}为容积点总数, n_{c}为状态维数n_x的 2 倍, 即$n_{\mathrm{c}}=2n_x$; $[1]_i$表示对称点集 [1] 中的第i个点。

已知每个关联到量测的 PPP 目标的高斯分量预测均值和协方差分别为$\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i,C}$、$\boldsymbol{P}_{k|k-1}^{\mathrm{u},i,C}$, 计算容积点为

$$\bar{\boldsymbol{\xi}}_{k|k-1}^{i}=\boldsymbol{S}_{k|k-1}\xi_i+\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i,C} \tag{39}$$

式中: $\boldsymbol{S}_{k|k-1}=\mathrm{chol}(\boldsymbol{P}_{k|k-1}^{\mathrm{u},i,C})$, chol(·)表示 Cholesky 分解。

步骤 2　计算容积点$\bar{\boldsymbol{\xi}}_{k|k-1}^{i}$经过非线性伪量测方程$\boldsymbol{h}(\cdot)$传播后的容积点, 即

$$\boldsymbol{\zeta}_{k|k-1}^{i}=\boldsymbol{h}(\bar{\boldsymbol{\xi}}_{k|k-1}^{i}) \qquad i=1,2,\cdots,n_{\mathrm{c}} \tag{40}$$

步骤 3　计算传递后容积点的预测均值$\hat{\boldsymbol{z}}_{k|k-1}$、协方差$\boldsymbol{P}_{xz}$和方差$\boldsymbol{P}_{zz}$, 即

$$\hat{\boldsymbol{z}}_{k|k-1}=\frac{1}{n_{\mathrm{c}}}\sum_{i=1}^{n_{\mathrm{c}}}\boldsymbol{\zeta}_{k|k-1}^{i} \tag{41}$$

$$\boldsymbol{P}_{xz}=\frac{1}{n_{\mathrm{c}}}\sum_{i=1}^{n_{\mathrm{c}}}\bar{\boldsymbol{\xi}}_i\boldsymbol{\zeta}^{i\mathrm{T}}-\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i,C}\hat{\boldsymbol{z}}_{k|k-1}^{\mathrm{T}} \tag{42}$$

$$\boldsymbol{P}_{zz}=\frac{1}{n_{\mathrm{c}}}\sum_{i=1}^{n_{\mathrm{c}}}\boldsymbol{\zeta}_{k|k-1}^{i}(\boldsymbol{\zeta}_{k|k-1}^{i})^{\mathrm{T}}-\hat{\boldsymbol{z}}_{k|k-1}\hat{\boldsymbol{z}}_{k|k-1}^{\mathrm{T}}+\boldsymbol{R}_k \tag{43}$$

步骤 4　计算增益矩阵$\boldsymbol{K}_k$且进行滤波更新, 即

$$\boldsymbol{K}_k=\boldsymbol{P}_{xz}\boldsymbol{P}_{zz}^{-1} \tag{44}$$

$$\boldsymbol{\mu}_{k}^{\mathrm{u},i,C}=\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i,C}+\boldsymbol{K}_k(\boldsymbol{0}-\hat{\boldsymbol{z}}_{k|k-1}) \tag{45}$$

$$\boldsymbol{P}_{k}^{\mathrm{u},i,C}=\boldsymbol{P}_{k|k-1}^{\mathrm{u},i,C}-\boldsymbol{K}_k\boldsymbol{P}_{zz}\boldsymbol{K}_k^{\mathrm{T}} \tag{46}$$

步骤 5　计算 k 时刻似然函数$\varphi_z(\boldsymbol{x})$, 即

$$\varphi_z(\boldsymbol{x})=N(0;\hat{\boldsymbol{z}}_{k|k-1},\boldsymbol{P}_{zz}) \tag{47}$$

根据得到的似然函数表达式 (47) 计算似然函数更新式 (27), 则新检测到目标的均值和协方差的更新由式 (45) 和式 (46) 得到。已检测到目标的状态均值和协方差的求解与上述流程类似, 不再赘述。

为了更好地说明本文算法的递归过程, 给出了算法实现流程, 如图 5 所示。同时, 给出了该算法的部分实现伪代码, 如下所示。

输入: $k-1$时刻 PPP 强度D_{k-1}^{u}, $k-1$时刻 MBM 参数集$\{W_{k-1}^{j},\{r_{k-1}^{j,i},f_{k-1}^{j,i}\}_{i\in I^j}\}_{j\in J}$, 量测集$Z_k$。

输出: k时刻 PPP 强度D_k^{u}, k时刻 MBM 参数集$\{W_k^{j},\{r_k^{j,i},f_k^{j,i}\}_{i\in I^j}\}_{j\in J}$。

1) 预测步

PPP 分量的预测

1: for $i=1:n_{k-1}^{\mathrm{u}}$ do

2: 由式 (14)～式 (16) 得$w_{k|k-1}^{\mathrm{u},i},\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i},\boldsymbol{P}_{k|k-1}^{\mathrm{u},i}$

3: end for

新生目标的预测, 并入到 PPP 中

4: for $i=1:n_k^{\mathrm{b}}$ do

5: 由式 (17)～式 (19) 得$w_{k|k-1}^{\mathrm{u},i+n_{k-1}^{\mathrm{u}}},\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i+n_{k-1}^{\mathrm{u}}},\boldsymbol{P}_{k|k-1}^{\mathrm{u},i+n_{k-1}^{\mathrm{u}}}$

6: end for

PPP 目标高斯分量个数及强度函数为: $n_{k|k-1}^{\mathrm{u}}=n_{k-1}^{\mathrm{u}}+n_{k-1}^{\mathrm{b}},D_{k|k-1}^{\mathrm{u}}$

MBM 分量的预测

7: for $i=1:n_{k-1}$ do

8: 由式 (21)～式 (23) 得$W_{k|k-1}^{j},r_{k|k-1}^{j,i},f_{k|k-1}^{j,i}(\boldsymbol{x})$

9: end for

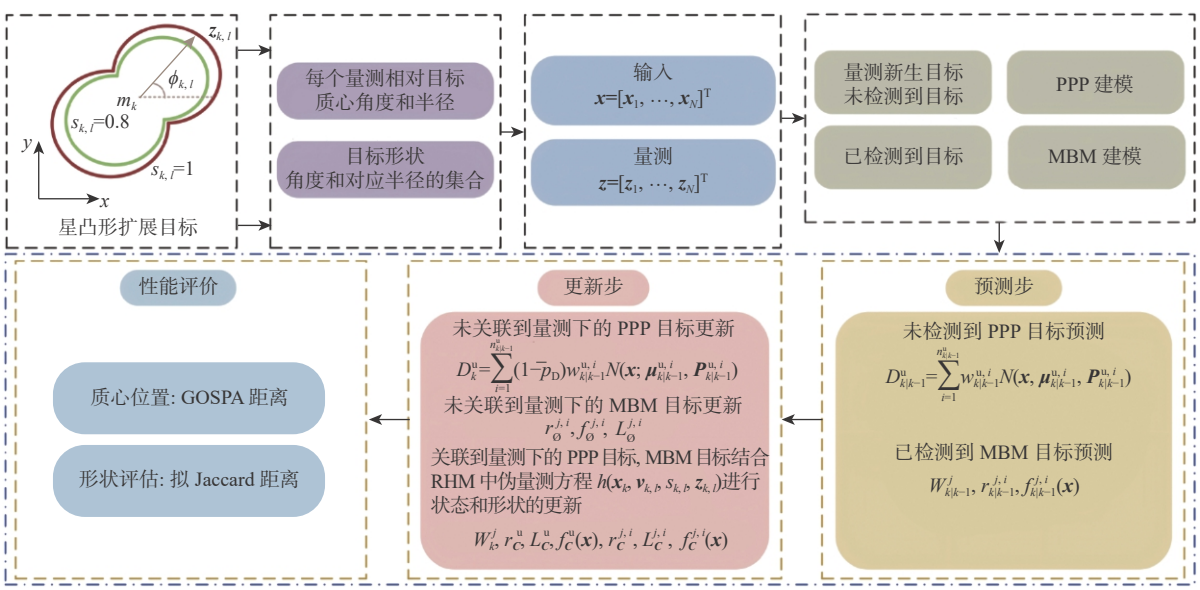


图 5　RHM-GM-ETPMBM 滤波算法流程

Fig. 5　Flowchart of RHM-GM-ETPMBM filtering algorithm

2) 更新步

未关联到量测的 PPP 分量更新

10: for $i=1:n_{k|k-1}^{\mathrm{u}}$ do

11: 由式 (25) 得 D_k^{u}

12: end for

13: for $j\in J$ do

14: for each $A\in A^j$ do

未关联到量测的 MBM 分量更新

15: for $i=1:n_{k|k-1}$ do

16: 由式 (30)～式 (32) 得 $L_{\varnothing}^{j,i},r_{\varnothing}^{j,i},f_{\varnothing}^{j,i}(x)$

17: end for

18: for $\boldsymbol{C}\subset A$ do

关联到量测的 MBM 分量更新

19: for $i=1:n_{k|k-1}$ do

20: $\boldsymbol{\mu}_k^0=\boldsymbol{\mu}_{k|k-1}^{j,i,\boldsymbol{C}},\boldsymbol{P}_k^0=\boldsymbol{P}_{k|k-1}^{j,i,\boldsymbol{C}}$

21: for each $\boldsymbol{z}_k\in\boldsymbol{C}$ do

22: 由式 (38)～式 (47) 计算得 $\boldsymbol{\mu}_k^0,\boldsymbol{P}_k^0,\varphi_z(\boldsymbol{x})$

23: end for

24: $\boldsymbol{\mu}_k^{j,i,\boldsymbol{C}}=\boldsymbol{\mu}_k^0,\boldsymbol{P}_k^{j,i,\boldsymbol{C}}=\boldsymbol{P}_k^0$

25: 由式 (33)～式 (37) 得 $L_{\boldsymbol{C}}^{j,i},r_{\boldsymbol{C}}^{j,i},f_{\boldsymbol{C}}^{j,i}(\boldsymbol{x})$

26: end for

关联到量测的 PPP 分量更新

27: for $i=1:n_{k|k-1}^{\mathrm{u}}$ do

28: $\boldsymbol{\mu}_k^1=\boldsymbol{\mu}_{k|k-1}^{\mathrm{u},i},\boldsymbol{P}_k^1=\boldsymbol{P}_{k|k-1}^{\mathrm{u},i}$

29: for each $\boldsymbol{z}_k\in\boldsymbol{C}$ do

30: 由式 (38)～式 (47) 得 $\boldsymbol{\mu}_k^1,\boldsymbol{P}_k^1,\varphi_z(\boldsymbol{x})$

31: end for

32: $\boldsymbol{\mu}_k^{\mathrm{u},i,\boldsymbol{C}}=\boldsymbol{\mu}_k^1,\boldsymbol{P}_k^{\mathrm{u},i,\boldsymbol{C}}=\boldsymbol{P}_k^1$

33: end for

34: 由式 (26)～式 (29) 计算得 $L_{\boldsymbol{C}}^{\mathrm{u}},r_{\boldsymbol{C}}^{\mathrm{u}},f_{\boldsymbol{C}}^{\mathrm{u}}(\boldsymbol{x})$

35: end for

36: $L\to W_k^j$

37: end for

38: end for

39: $W_k^j=W_k^j\Big/\sum_{j\in J}W_k^j$

3 算法性能评价指标

本文引入广义最优子模式分配 (generalized optimal subpattern assignment, GOSPA) 距离[34]和拟 Jaccard 距离[30]分别对目标运动状态估计和扩展状态估计进行有效评价。

3.1 GOSPA 距离

GOSPA 距离可以理解为最优子模式分配 (optimal subpattern assignment, OSPA) 距离[35]的非归一化形式。当归一化参数 $\alpha=2$, GOSPA 距离能够描述每个目标第 p 阶的位置误差、漏检误差和虚警误差。当 $\alpha=2$, GOSPA 距离定义如下:

$$d_p^{(c,2)}(X,Y)=\left[\min_{\theta\in\Theta}\left(\sum_{(i,j)\in\theta}d(\boldsymbol{x}_i,\boldsymbol{y}_j)^p+\frac{c^p}{2}(|X|+|Y|-2|\theta|)\right)\right]^{\frac{1}{p}} \tag{48}$$

式中: θ 表示由 X、Y 得到的赋值集合; Θ 表示包括所有可能的 θ 的集合; c 为位置误差的截断距离。

3.2 拟 Jaccard 距离

本文采用拟 Jaccard 距离对形状估计进行评估, 建立扩展目标估计形状和真实形状之间的关联矩阵, 进而获取真实形状与估计形状的相似度, 分别计算 k 时刻单个扩展目标的拟 Jaccard 距离, 计算公式如下:

$$J_{k,i}^q=1-\frac{\sum_{j=1}^{n}\left[\min\left(\left\|\hat{\boldsymbol{x}}_{k,j}^{\mathrm{sh}}-\boldsymbol{m}_k\right\|,\left\|\boldsymbol{x}_{k,j}^{\mathrm{sh}}-\boldsymbol{m}_k\right\|\right)\right]^2}{\sum_{j=1}^{n}\left[\max\left(\left\|\hat{\boldsymbol{x}}_{k,j}^{\mathrm{sh}}-\boldsymbol{m}_k\right\|,\left\|\boldsymbol{x}_{k,j}^{\mathrm{sh}}-\boldsymbol{m}_k\right\|\right)\right]^2} \tag{49}$$

式中: 等间隔获取真实形状点集 $X_{k,i}^{\mathrm{sh}}=\{\boldsymbol{x}_{k,j}^{\mathrm{sh}}\}_{j=1}^n$; 根据 RHM 计算得出 k 时刻扩展目标估计形状点集 $\hat{X}_{k,j}^{\mathrm{sh}}=\{\hat{\boldsymbol{x}}_{k,j}^{\mathrm{sh}}\}_{j=1}^n$。

4 仿真实验

ETPMBM 滤波算法是目前基于随机有限集 (random finite set, RFS) 的滤波器研究及其应用的主流, 是 RFS 多目标滤波器中一种综合性能极具竞争力的先进算法。为使对比实验更具针对性和公平性, 对照组也应是基于 RHM 和 RFS 的多扩展目标滤波跟踪经典算法, 故选择经典的基于 RHM 和 RFS 的 RHM-GM-ETPHD[36](方案 1)、RHM-GM-ETCBMeMBer[30](方案 2) 与本文提出的 RHM-GM-ETPMBM 滤波算法 (方案 3) 来验证算法的有效性。

假设在实验环境中存在 3 种不规则形状扩展目标, 量测采样周期 $T=1$ s, 目标跟踪时间长度为 50 s, 场景中设置带有杂波的观察区域为 [−600, 600] m×[−100,400] m, 在该跟踪区域内, 存在服从 $\lambda=5$ 的泊松分布产生的随机数目的杂波。设量测噪声是协方差为 $\boldsymbol{R}_k=\mathrm{diag}\{[0.02^2,0.02^2]\}$ 的高斯白噪声, 目标运动模型为匀速 (CV) 运动模型, 则 k 时刻第 i 个扩展目标的状态演化过程可以表示为

$$\begin{cases}\boldsymbol{x}_k^{(i)}=\boldsymbol{F}_k\boldsymbol{x}_{k-1}^{(i)}+\boldsymbol{\varepsilon}_k^{(i)}\\ \boldsymbol{\varepsilon}_k^{(i)}\sim N(\boldsymbol{\varepsilon}_k^{(i)};0,\boldsymbol{Q}_k)\end{cases} \tag{50}$$

式中: $i=1,\cdots,N_{k-1}$ 为扩展目标的数目; $\boldsymbol{F}_k$ 为状态转移矩阵; $\boldsymbol{Q}_k$ 为过程噪声协方差。

$$\begin{cases}\boldsymbol{F}_k=\mathrm{diag}\{[\boldsymbol{F}_k^{\mathrm{cv}},\boldsymbol{I}_{n_t}]\}\\ \boldsymbol{Q}_k=\mathrm{diag}\{[\boldsymbol{P}_k^{\mathrm{cv}},0.1\boldsymbol{I}_{n_t}]\}\end{cases} \tag{51}$$

$$\boldsymbol{F}_k^{\mathrm{cv}}=\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}\otimes\boldsymbol{I}_2 \tag{52}$$

$$\boldsymbol{P}_k^{\mathrm{cv}}=\begin{bmatrix}\dfrac{T^3}{3} & \dfrac{T^2}{2}\\ \dfrac{T^2}{2} & T\end{bmatrix}\otimes\begin{bmatrix}\sigma^2 & \\ & \sigma^2\end{bmatrix} \tag{53}$$

式中: 过程噪声标准差$\sigma=1.5\ \mathrm{m/s^2}$; $\boldsymbol{I}_{n_i}$和$\boldsymbol{I}_2$分别为n_{f}阶单位阵和二阶单位阵; ⊗表示克罗内克积。本文中, 选取傅里叶级数参数个数为$n_{\mathrm{f}}=11$。每个扩展目标产生量测数目的期望为$\gamma=40$, 设$s_{k,l}\sim N(s_{k,l};2/3;1/18)$, 目标存活概率和传感器检测概率分别为$p_{\mathrm{S}}=0.99$, $p_{\mathrm{D}}=0.99$。新生目标的初始运动状态分别为$\boldsymbol{\mu}_{\Gamma}^{(1)}=[50\ \mathrm{m},0\ \mathrm{m},-8\ \mathrm{m/s},10\ \mathrm{m/s}]^{\mathrm{T}}$、$\boldsymbol{\mu}_{\Gamma}^{(2)}=[-50\ \mathrm{m},0\ \mathrm{m},16\ \mathrm{m/s},10\ \mathrm{m/s}]^{\mathrm{T}}$、$\boldsymbol{\mu}_{\Gamma}^{(3)}=[-50\ \mathrm{m},-50\ \mathrm{m},-16\ \mathrm{m/s},10\ \mathrm{m/s}]^{\mathrm{T}}$, 协方差为$\boldsymbol{P}_{\Gamma}^{(1,2,3)}=\mathrm{diag}\{[100^2\ \mathrm{m}^2,100^2\ \mathrm{m}^2,1\ \mathrm{m}^2/\mathrm{s}^2,1\ \mathrm{m}^2/\mathrm{s}^2]\}^{\mathrm{T}}$。

多扩展目标初始参数设置如表 1 所示。将未知目标初始参数与新生目标参数设定相同, 设置 GOSPA 距离参数$c=3$, $p=1$, $\alpha=2$。拟 Jaccard 距离的轮廓点个数$n=50$。仿真场景中, 目标真实运动轨迹如图 6 所示。

表 1　多扩展目标初始参数

Table 1　Initial parameters of MET

目标	新生时间/s	消亡时间/s	位置/m	速度/(m·s^{-1})
1	1	41	$[50,0]^{\mathrm{T}}$	$[-8,10]^{\mathrm{T}}$
2	15	51	$[-50,0]^{\mathrm{T}}$	$[16,10]^{\mathrm{T}}$
3	25	51	$[-50,-50]^{\mathrm{T}}$	$[-16,10]^{\mathrm{T}}$

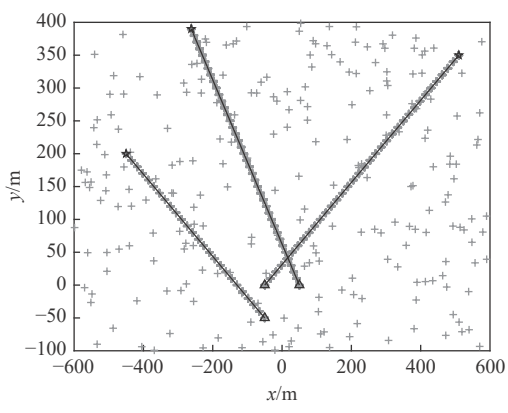


图 6　目标真实轨迹

Fig. 6　Actual trajectories of target

4.1　星凸形多扩展目标跟踪实验

根据设置的跟踪场景, 针对图 7 中所示 3 种不同形状的扩展目标, 本文采用 3 种方案进行跟踪, 并进行了 100 次独立的蒙特卡罗(Monte Carlo, MC)仿真实验, 用以对比各类滤波器的跟踪性能。

为更好地比较各滤波器对质心位置估计和形状估计的效果, 计算得出 3 种滤波器在 50 s 时间步长的 GOSPA 距离和拟 Jaccard 距离均值, 如表 2 所示。

图 8 展示了各滤波器跟踪区域内对各扩展目标的整体跟踪效果。可知, 3 种滤波器均具备高效精确地对多个扩展目标进行跟踪的能力。

图 9 为 3 个目标在运动过程中接近消亡时刻的

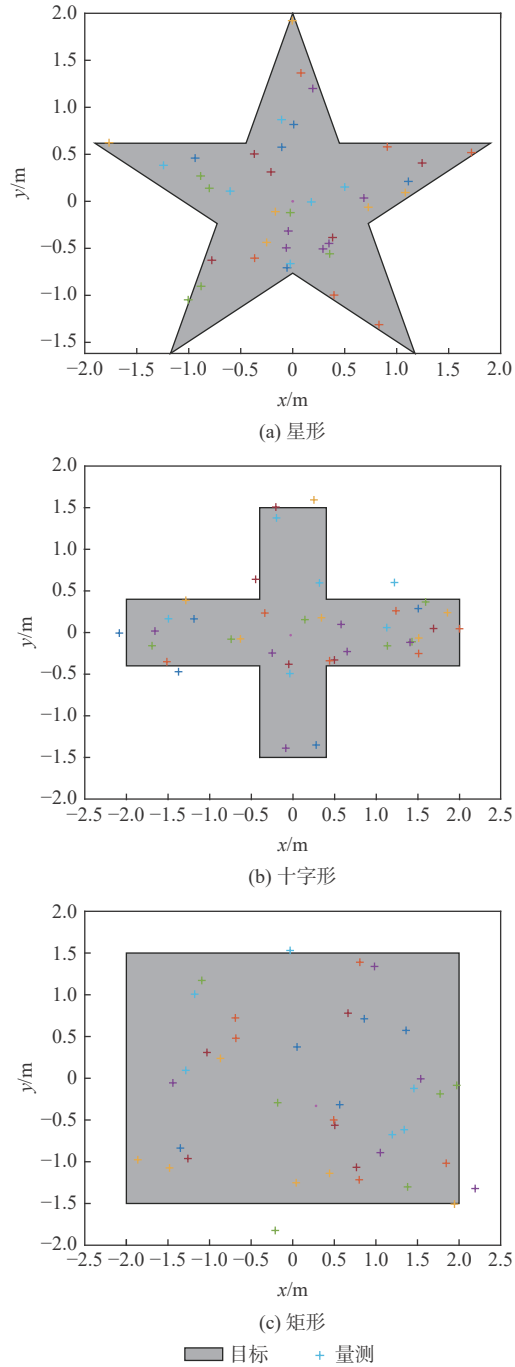


图 7　不同形状的扩展目标

Fig. 7　Extended targets of different shapes

表 2　多扩展目标评价结果均值

Table 2　Evaluation results mean of MET

实验方案	GOSPA距离均值/m	拟Jaccard距离均值/m
方案1	0.298 3	0.426 8
方案2	0.291 6	0.409 2
方案3	0.279 4	0.385 9

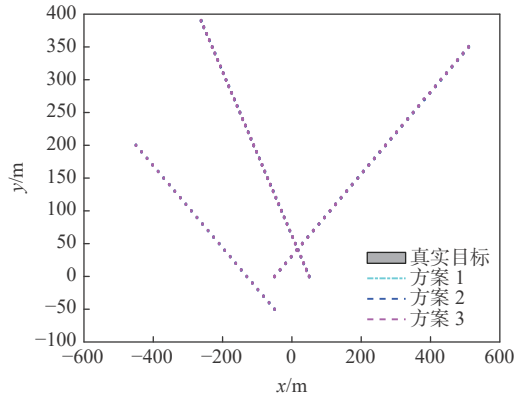
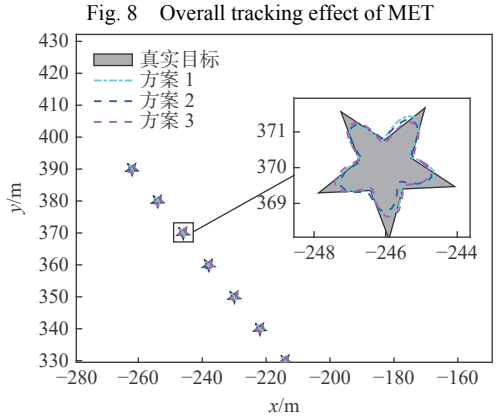
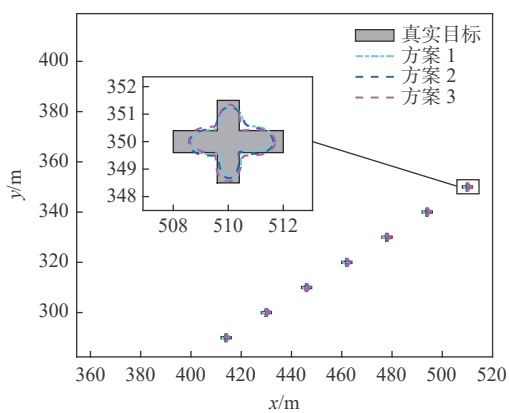


图 8 多扩展目标的整体跟踪效果

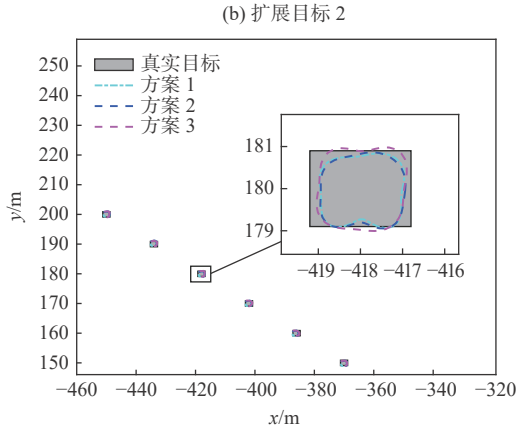
Fig. 8 Overall tracking effect of MET



(a) 扩展目标 1



(b) 扩展目标 2



(c) 扩展目标 3

图 9 扩展目标的形状估计细节放大效果图

Fig. 9 Detail enlarged effort diagram of shape estimation for extended targets

形状估计局部放大图，用以突出各滤波器对目标形状的估计效果。可以看出，3 种算法均能够较好地估计目标真实形状，通过观察图中目标局部轮廓估计的放大细节，可见本文提出的滤波算法在多扩展目标形状估计方面表现最为优异。

图 10 为 100 次独立 MC 仿真实验下得到的势估计。可知，3 种滤波器在杂波干扰的环境中对目标的势估计都表现出很好的跟踪性能。

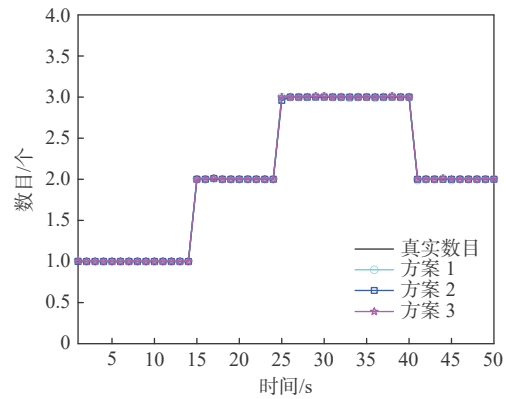


图 10 扩展目标的势估计

Fig. 10 Cardinality estimation of extended targets

图 11 为 3 种滤波器对目标质心估计的 GOSPA 距离结果对比。可知，3 种滤波器都能十分精确地跟踪目标的质心位置，这是因为每个扩展目标都对应了多个量测值，联合估计之下提高了滤波器对目标质心状态估计的性能。在图 11 中，当有新生目标出现时 GOSPA 距离出现上升趋势，这是由于新生目标的突然出现，其质心位置不确定所造成的。根据表 2 中给出的结果及图 11 的整体趋势可以得出，本文提出的滤波算法在目标质心位置估计方面的性能最优。

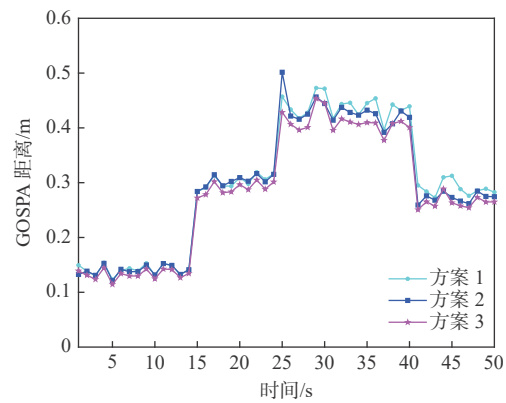


图 11 扩展目标的 GOSPA 距离

Fig. 11 GOSPA distance of extended targets

图 12 为表示滤波器形状估计性能的拟 Jaccard 距离，刚开始估计误差偏大是因为目标形状先验信息未知，但随着时间的推移，各滤波器对扩展目标的形状估计趋于收敛，表现出了稳定的估计性能。

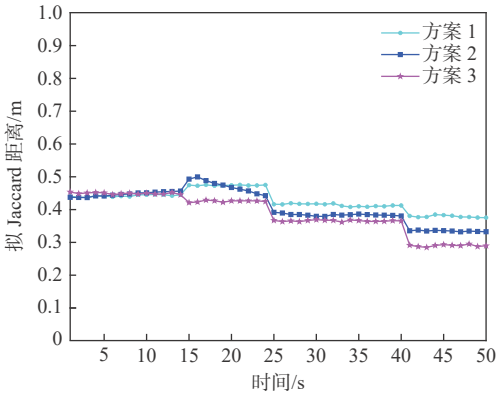


图 12　扩展目标的拟 Jaccard 距离

Fig. 12　Quasi-Jaccard distance of extended targets

且通过观察图 12, 以及表 2 中给出的拟 Jaccard 距离的均值可知, 本文提出的滤波算法在多扩展目标形状估计方面的性能更优。

4.2　星凸形多群目标跟踪实验

为了验证本文所提滤波算法在多群体目标跟踪性能方面的表现, 选择对多个星凸形群目标进行跟踪, 每个群目标由 20 个子目标组成, 其中假定每个子目标都有相同的运动特性, 2 种群目标形状如图 13 所示。为了更好地比较各滤波器的跟踪性能, 表 3 给出了 2 种评价指标在跟踪时间步长内的

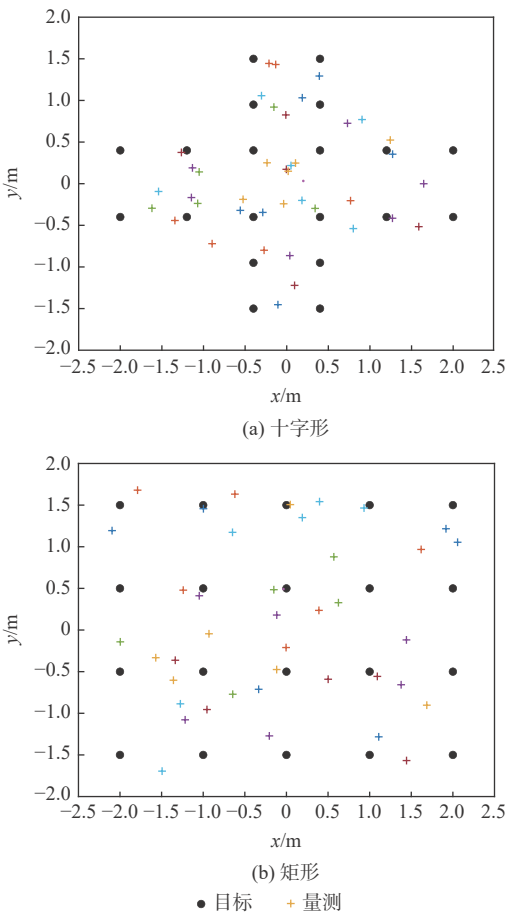


图 13　不同形状的群目标

Fig. 13　Group targets of different shapes

表 3　多群目标评价结果均值

Table 3　Evaluation result mean of multiple group target

实验方案	GOSPA距离均值/m	拟Jaccard距离均值/m
方案1	0.281 0	0.568 3
方案2	0.271 5	0.555 1
方案3	0.264 1	0.453 7

均值。

图 14 反映了各滤波器对多群目标的整体跟踪效果。可以看到, 3 种滤波器都能对多群目标的运动轨迹进行准确的跟踪。图 15 表示每个群目标在接近消亡时刻的形状估计局部放大效果。可知, 3 种滤波器在对群目标形状估计方面都表现出了较好的性能。但是, 从局部放大的轮廓细节也可以看出, 本文算法的估计轮廓还是更贴近于真实目标形状, 算法性能更加优越。

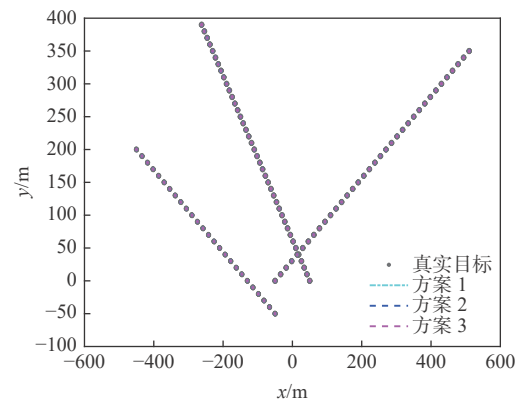


图 14　多群目标的整体跟踪效果

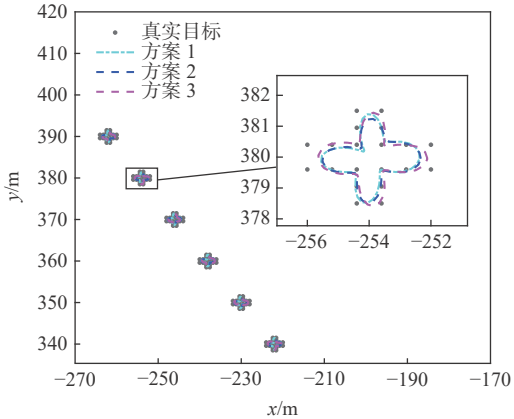
Fig. 14　Overall tracking effect of multiple group target

图 16 为 100 次独立 MC 仿真实验下得到的多群目标的势估计。可以看出, 3 种滤波器在杂波跟踪环境中都能较好地估计多群目标的数目变化。

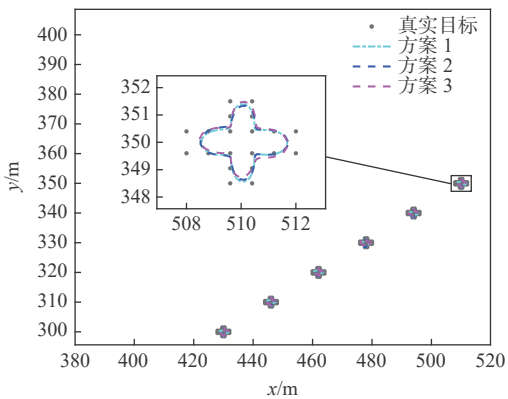
通过图 17 中表示的 GOSPA 距离可知, 3 种滤波器都能较好地跟踪群目标的位置, 同时结合表 3 中得到的 GOSPA 距离均值可以看出, 本文算法跟踪性能表现更加出色。

由图 18 表示的 Jaccard 距离及表 3 得到的 Jaccard 距离均值可知, 3 种算法都能够反映出各个群目标的真实轮廓特征, 并且随着滤波的进行, 估计逐渐趋于稳定。但整体而言, 相较另外 2 种滤波算法, 本文算法的综合性能更优。

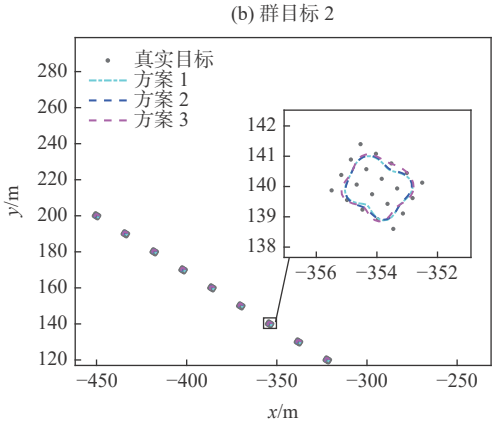
综上可知, 在含有噪声干扰、杂波存在及目标漏检的复杂环境下, 本文提出的滤波算法在与已有的 RHM-GM-ETPHD 算法和 RHM-GM-ETCBMeMBer 算法进行跟踪对比时, 采用了更为先进的共轭先验假设, 将尚未检测到的扩展目标信息考虑其中, 可以对潜在目标的假设数量进行有效管理, 使得滤波递推过程更加完善, 故本文算法在运动状态估计和



(a) 群目标 1



(b) 群目标 2



(c) 群目标 3

图 15 群目标的形状估计细节放大效果图

Fig. 15 Detail enlarged effect diagram of shape estimation for group targets

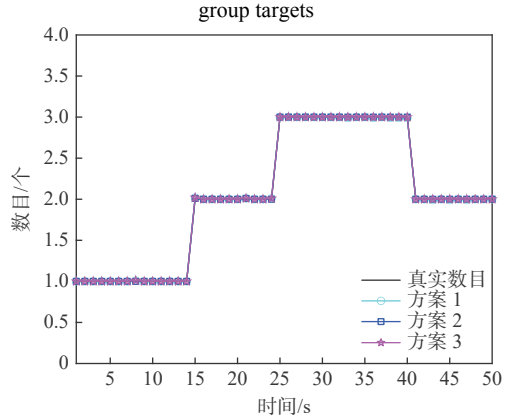


图 16 群目标的势估计

Fig. 16 Cardinality estimation of group targets

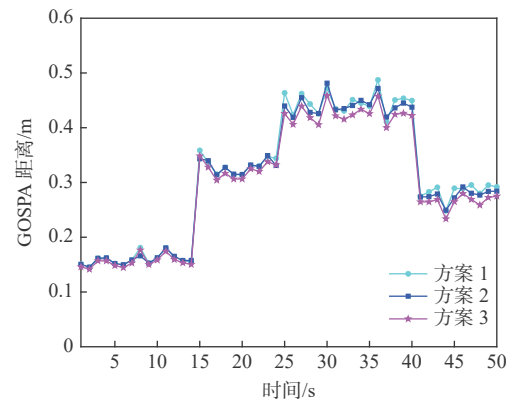


图 17 群目标的 GOSPA 距离

Fig. 17 GOSPA distance of group targets

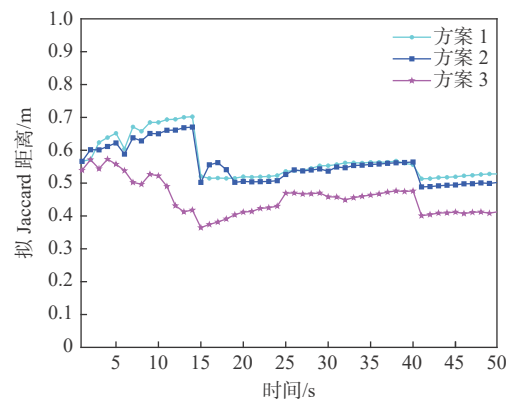


图 18 群目标的拟 Jaccard 距离

Fig. 18 Quasi-Jaccard distance of group targets

形状估计等方面均表现出更加出色且稳定的性能。

5 结 论

1）本文提出了星凸形不规则形状的多扩展目标 PMBM 滤波器，能同时跟踪目标运动状态与复杂外形，超越了传统点目标跟踪仅能提供位置信息的局限。

2）该滤波器建立了多扩展目标概率密度的高效递推形式，能够实时地根据新的量测数据处理复杂场景下的 METT 问题。

3）该滤波器在运动状态估计和形状估计等方面相较于传统算法表现更加优越。

未来将聚焦于非星凸形或更复杂形状的扩展目标、更加科学有效的形状建模方法等方面的研究。

参考文献（References）

[1] XIA Y, GARCÍA-FERNÁNDEZ Á F, MEYER F, et al. Trajectory PMB filters for extended object tracking using belief propagation [EB/OL]. (2023-09-19)[2023-11-01]. https://arxiv.org/abs/2207.10164.

[2] GARCÍA-FERNÁNDEZ Á F, WILLIAMS J L, SVENSSON L, et al. A Poisson multi-Bernoulli mixture filter for coexisting point and

extended targets[J]. IEEE Transactions on Signal Processing, 2021, 69: 2600-2610.

[3] XIA Y X, GRANSTRÖM K, SVENSSON L, et al. Poisson multi-Bernoulli approximations for multiple extended object filtering[J]. IEEE Transactions on Aerospace and Electronic Systems, 2022, 58(2): 890-906.

[4] GARCÍA-FERNÁNDEZ Á F, SVENSSON L, WILLIAMS J L, et al. Trajectory Poisson multi-Bernoulli filters[J]. IEEE Transactions on Signal Processing, 2020, 68: 4933-4945.

[5] GARCÍA-FERNÁNDEZ Á F, MASKELL S. Continuous-discrete multiple target filtering: PMBM, PHD and CPHD filter implementations[J]. IEEE Transactions on Signal Processing, 2020, 68: 1300-1314.

[6] MAHLER R P S. Multitarget Bayes filtering via first-order multi target moments[J]. IEEE Transactions on Aerospace and Electronic Systems, 2003, 39(4): 1152-1178.

[7] VO B N, MA W K. The Gaussian mixture probability hypothesis density filter[J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4091-4104.

[8] VO B N, SINGH S, DOUCET A. Sequential Monte Carlo methods for multitarget filtering with random finite sets[J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(4): 1224-1245.

[9] MAHLER R P S. PHD filters of higher order in target number[J]. IEEE Transactions on Aerospace and Electronic Systems, 2007, 43(4): 1523-1543.

[10] MAHLER R P S. Statistical multisource-multitarget information fusion[M]. Norwood: Artech House, 2007.

[11] VO B T, VO B N, CANTONI A. The cardinality balanced multi-target multi-bernoulli filter and its implementations[J]. IEEE Transactions on Signal Processing, 2009, 57(2): 409-423.

[12] REUTER S, VO B T, VO B N, et al. The labeled multi-Bernoulli filter[J]. IEEE Transactions on Signal Processing, 2014, 62(12): 3246-3260.

[13] VO B T, VO B N. Labeled random finite sets and multi-object conjugate priors[J]. IEEE Transactions on Signal Processing, 2013, 61(13): 3460-3475.

[14] VO B N, VO B T, PHUNG D. Labeled random finite sets and the Bayes multi-target tracking filter[J]. IEEE Transactions on Signal Processing, 2014, 62(24): 6554-6567.

[15] GARCÍA-FERNÁNDEZ Á F, WILLIAMS J L, GRANSTRÖM K, et al. Poisson multi-Bernoulli mixture filter: direct derivation and implementation[J]. IEEE Transactions on Aerospace and Electronic Systems, 2018, 54(4): 1883-1901.

[16] WILLIAMS J L. Marginal multi-Bernoulli filters: RFS derivation of MHT, JIPDA, and association-based MeMBer[J]. IEEE Transactions on Aerospace and Electronic Systems, 2015, 51(3): 1664-1687.

[17] PETROV N, MIHAYLOVA L, GNING A, et al. A novel sequential Monte Carlo approach for extended object tracking based on border parameterisation[C]//Proceedings of the 14th International Conference on Information Fusion. Piscataway: IEEE Press, 2011: 1-8.

[18] BAUM M, KLUMPP V, HANEBECK U D. A novel Bayesian method for fitting a circle to noisy points[C]//Proceedings of the 13th International Conference on Information Fusion. Piscataway: IEEE Press, 2010: 1-6.

[19] BAUM M, FAION F, HANEBECK U D. Modeling the target extent with multiplicative noise[C]//Proceedings of the 15th International Conference on Information Fusion. Piscataway: IEEE Press, 2012: 2406-2412.

[20] DE FREITAS A, MIHAYLOVA L, GNING A, et al. Autonomous crowds tracking with box particle filtering and convolution particle filtering[J]. Automatica, 2016, 69: 380-394.

[21] LAN J, LI X R. Tracking of extended object or target group using random matrix—Part Ⅱ: Irregular object[C]//Proceedings of the 15th International Conference on Information Fusion. Piscataway: IEEE Press, 2012: 2185-2192.

[22] LAN J, LI X R. Tracking of maneuvering non-ellipsoidal extended object or target group using random matrix[J]. IEEE Transactions on Signal Processing, 2014, 62(9): 2450-2463.

[23] FELDMANN M, FRÄNKEN D, KOCH W. Tracking of extended objects and group targets using random matrices[J]. IEEE Transactions on Signal Processing, 2011, 59(4): 1409-1420.

[24] BAUM M, HANEBECK U D. Random hypersurface models for extended object tracking[C]//Proceedings of the IEEE International Symposium on Signal Processing and Information Technology. Piscataway: IEEE Press, 2009: 178-183.

[25] BAUM M, HANEBECK U D. Shape tracking of extended objects and group targets with star-convex RHMs[C]//Proceedings of the 14th International Conference on Information Fusion. Piscataway: IEEE Press, 2011: 1-8.

[26] BAUM M, HANEBECK U D. Extended object tracking with random hypersurface models[J]. IEEE Transactions on Aerospace and Electronic Systems, 2014, 50(1): 149-159.

[27] STEINBRING J, BAUM M, ZEA A, et al. A closed-form likelihood for particle filters to track extended objects with star-convex RHMs[C]//Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. Piscataway: IEEE Press, 2015: 25-30.

[28] GRANSTROM K, ORGUNER U. A PHD filter for tracking multiple extended targets using random matrices[J]. IEEE Transactions on Signal Processing, 2012, 60(11): 5657-5671.

[29] LUNDQUIST C, GRANSTRÖM K, ORGUNER U. An extended target CPHD filter and a gamma Gaussian inverse Wishart implementation[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(3): 472-483.

[30] 陈辉, 杜金瑞, 韩崇昭. 基于星凸形随机超曲面模型多扩展目标多伯努利滤波器[J]. 自动化学报, 2020, 46(5): 909-922.
CHEN H, DU J R, HAN C Z. A multiple extended target multi-Bernouli filter based on star-convex random hypersurface model[J]. Acta Automatica Sinica, 2020, 46(5): 909-922(in Chinese).

[31] GRANSTRÖM K, FATEMI M, SVENSSON L. Gamma Gaussian inverse-Wishart Poisson multi-Bernoulli filter for extended target tracking[C]//Proceedings of the 19th International Conference on Information Fusion. Piscataway: IEEE Press, 2016: 893-900.

[32] GRANSTRÖM K, FATEMI M, SVENSSON L. Poisson multi-Bernoulli mixture conjugate prior for multiple extended target filtering[J]. IEEE Transactions on Aerospace and Electronic Systems, 2020, 56(1): 208-225.

[33] ARASARATNAM I, HAYKIN S. Cubature Kalman filters[J]. IEEE

Transactions on Automatic Control, 2009, 54(6): 1254-1269.

[34] RAHMATHULLAH A S, GARCÍA-FERNÁNDEZ Á F, SVENSSON L. Generalized optimal sub-pattern assignment metric[C]// Proceedings of the 20th International Conference on Information Fusion. Piscataway: IEEE Press, 2017: 1-8.

[35] SCHUHMACHER D, VO B T, VO B N. A consistent metric for performance evaluation of multi-object filters[J]. IEEE Transactions on Signal Processing, 2008, 56(8): 3447-3457.

[36] HAN Y L, ZHU H Y, HAN C Z. A Gaussian-mixture PHD filter based on random hypersurface model for multiple extended targets[C]//Proceedings of the 16th International Conference on Information Fusion. Piscataway: IEEE Press, 2013: 1752-1759.

Star convex irregular shape multi-extended target PMBM filter

CHEN Hui[1,*], LIU Mengbo[1], LIAN Feng[2], HAN Chongzhao[2]

(1. School of Automation and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, China;
2. School of Automation Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: This work suggests a star convex irregular shape multi-extended target Poisson multi-Bernoulli mixture (PMBM) filtering algorithm to address the challenge of monitoring multiple extended targets with irregular shapes in complicated and uncertain situations. First, the Poisson point process (PPP) and multi-Bernoulli mixture (MBM) are used to model the unknown and existing target sets, effectively representing potential target information while establishing an efficient multi-target density recursive form. The measurement source distribution of any star convex extended target is accurately modeled by the random hypersurface model, and the best nonlinear filter solves the highly nonlinear pseudo measurement equation. On the basis of deriving the extended target Poisson multi-Bernoulli mixture (ETPMBM) filter, the Gaussian mixture star convex irregular shape multi-extended target PMBM filter algorithm is derived and proposed in detail. In order to effectively and recursively estimate the probability density of numerous extended targets with multiple feature information, including irregular forms, this approach can create more compact multi-Bernoulli global hypotheses. Finally, the effectiveness of the algorithm proposed in this paper is verified through simulation experiments of multiple extended target tracking and multiple group target tracking.

Keywords: multiple extended target tracking; random hypersurface model; Poisson multi-Bernoulli mixture; optimal nonlinear filter; Gaussian mixture

Received: 2023-11-24; **Accepted**: 2023-12-15; **Published Online**: 2024-01-03 12:02

URL: link.cnki.net/urlid/11.2625.V.20240102.1351.001

Foundation items: National Natural Science Foundation of China (62163023); Gansu Provincial Basic Research Innovation Group of China (25JRRA058); Central Government's Funds for Guiding Local Science and Technology Development of China (25ZYJA040); Gansu Provincial Key Talent Project of China (2024RCXM86)

* **Corresponding author**. E-mail: huich78@hotmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0730

基于图注意力堆叠自编码器微生物-药物关联预测

王波[1,2,*]，何洋[1]，杜晓昕[1,2]，张剑飞[1,2]，徐靖然[1]，贾娜[1]

(1. 齐齐哈尔大学 计算机与控制工程学院，齐齐哈尔 161006；2. 齐齐哈尔大学 黑龙江省大数据网络安全检测分析重点实验室，齐齐哈尔 161006)

摘　　要：传统方法发掘微生物与药物新关联主要通过生物实验完成，耗费时间且开销极大。基于此，提出基于图注意力堆叠自编码器微生物与药物关联预测方法GATSAE。建立微生物与药物异构网络，丰富关联信息；通过图卷积网络(GCN)提取多层潜在特征，得到微生物和药物的卷积融合矩阵；采用改进的堆叠自编码器学习有意义的高阶相似特征的无监督低维表示，在堆叠自编码器的基础上追加图卷积和注意力机制，进一步优化高阶特征信息的提取；将低维特征与关联特征串联，使用多层感知机(MLP)对最终的微生物-药物进行评分预测。通过效能评估，GATSAE方法的受试者工作特征曲线下面积(AUROC)及精确率-召回率曲线下面积(AUPR)分别达到0.961 9和0.957 7，优于经典的机器学习方法和常见的深度学习方法。案例研究表明，GATSAE方法能够准确预测到与SARS-CoV-2、大肠杆菌相关的候选药物，以及与阿司匹林相关的候选微生物。

关 键 词：微生物与药物；关联预测；堆叠自编码器；注意力机制；图卷积网络；多层感知机

中图分类号：TP391

文献标志码：A　　**文章编号**：1001-5965(2026)01-0061-12

微生物在人体中扮演着重要的角色，被称为人体“被遗忘”的器官，它们对人们的健康起着保护作用，包括抵御病原体、改善代谢能力和增强免疫力[1]。然而，微生物群落的不平衡可能对健康和疾病产生重要影响。许多药物不仅仅是对人类细胞产生作用，也可以对微生物产生影响。治疗某些疾病时，通过药物修复有益微生物功能缺失，消除有害微生物的活动，成为一种方法[2-4]。抗生素是最典型的例子，它们被设计用来抑制或杀死细菌[5]。然而，由于抗生素的过度使用，微生物耐药性成为一个严峻的问题[5]，其不仅降低了抗生素的疗效，还给人类健康带来威胁。寻找相关联的新药物和药物重定位成为解决微生物耐药性问题的策略。

微生物除调节疾病和健康的作用外，还可能参与调控药物的活性和毒性[6]。一些研究表明，微生物群落对药物的活性和毒性起着关键作用。例如，某些微生物增加地高辛的毒性[7-8]，降低吗啡的清除率，使某些铜绿假单胞菌的吗啡浓度-时间曲线下面积(area under curve, AUC)升高，使辛伐他汀的AUC升高221%[9-11]，改变华法林的活性[12]，增加伊立替康的毒性[13]等。

目前，人们对微生物与药物之间的关联机制了解还不够。由于微生物具有个体差异性，需要进一步研究为何某些药物在某些人群中表现出良好的反应，而其他药物可能会引起副作用。此外，还有许多潜在的人类药物与微生物之间的关联有待发现。因此，需要继续深入研究微生物与药物之间的相互作用，以便更好地理解它们在人体中的作用，并有助于开发更个体化和高效的治疗方法。

现有的预测微生物-药物关联关系的方法可分为基于网络传播、基于机器学习和基于深度学习3类。

1）基于网络传播的方法。基于网络传播的方

收稿日期：2023-11-07；**录用日期**：2024-01-12；**网络出版时间**：2024-03-09 07:24
网络出版地址：link.cnki.net/urlid/11.2625.V.20240306.0934.001
基金项目：黑龙江省省属高等学校基本科研业务费国自然培育一般项目(145409324)
*通信作者. E-mail：bowangdr@qqhru.edu.cn

引用格式：王波，何洋，杜晓昕，等. 基于图注意力堆叠自编码器微生物-药物关联预测[J]. 北京航空航天大学学报，2026，52(1)：61-72.
WANG B，HE Y，DU X X，et al. Prediction of microbe-drug association based on graph attention stacked autoencoder[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：61-72 (in Chinese).

法关注拓扑信息，通过不同的网络构建和基于微生物-药物关系来推断相关性。通常，使用已知的微生物-药物关联、药物相似性和微生物相似性来构建异质网络，以预测新的微生物-药物关系。Zhu等[14]设计了HMDAKATZ模型，利用药物的化学结构来识别潜在的人类药物-微生物关联，以及与疾病的关系，该模型通过简单的度量方法成功预测了药物-微生物关联，但不适用于预测没有已知药物关联的新微生物或没有已知微生物关联的新药物。

2）基于机器学习的方法。传统的机器学习方法需要手工提取节点特征，不太适用于大规模网络，但具有可解释性和易理解等优点，仍然是关联预测的重要方法。Ma等[15]提出了基于加权超图学习的广义矩阵分解(generalized matrix factorization based on weighted hypergraph, WHGMF)来预测潜在的微生物-药物关联，该方法综合多组数据，计算微生物和药物的多种特征，并引入超图正则化以提高特征表示能力。Zhu等[16]设计了最近邻注意网络(nearest neighbor attention network, NNAN)，其由4个部分组成，用于生成微生物和药物的嵌入表示，并作为二元分类器来区分未标记的微生物-药物的潜在关联。然而，这些方法同样在揭示没有已知药物关联的新微生物或没有已知微生物关联的新药物方面存在一些限制。

3）基于深度学习的方法。近年来，深度学习方法在微生物-药物关联预测方面取得了重要进展，这些方法可以对大规模的微生物和药物数据进行处理，从而挖掘出微生物和药物之间的潜在联系。Long等[17]利用丰富的生物学信息构建了微生物与药物的异质网络，并提出了一个新的基于图卷积网络(graph convolutional network, GCN)框架的模型GCNMDA，用于预测人类微生物-药物关联，在GCN的隐藏层，进一步利用条件随机场(conditional random field, CRF)确保相似节点(即微生物或药物)具有相似的表示，为更准确地聚合邻域表征，在CRF层设计了一种注意机制，GCN提取非线性特征，解决了传统方法中成本高和耗时长的难题，但忽略了线性特征的重要性。Long等[18]提出了一种新颖的图注意力网络集成框架EGATMDA，并采用分层注意机制来预测构建的多个微生物-药物图的微生物-药物关联，针对每个输入图，该模型设计了一个具有节点级别注意力的GCN来学习节点(即微生物和药物)的嵌入，为了有效聚合来自多个输入图的节点嵌入，实现了图级别注意力来学习不同输入图的重要性。Long等[19]提出了一个异构图注意力网络框架HGATDVA来预测药物和微生物之间的关联，通过引入药物靶点和微生物宿主之间的蛋白质相互作用网络来预测药物-微生物关联，加入注意力机制，关注重要信息，忽略或减少其他信息，但未捕获异质网络中复杂的非线性关系。

现有关于微生物与药物关联预测的深度学习方法大多存在数据利用不充分、数据噪声较大且未考虑数据线性特征的问题，为此，本文提出了基于图注意力堆叠自编码器微生物-药物关联预测模型GATSAE。首先，通过整合微生物与药物的信息，构建异构网络，丰富微生物与药物的关联信息；其次，利用GCN提取数据的多层潜在特征，但由于GCN在提取潜在特征时会减少对图中度较小的节点的关注，从而丢失部分信息，本文在GCN模块加入了高斯核函数用以减小信息的损失；然后，采用改进的堆叠自编码器学习有意义的高阶相似特征的无监督低维表示，通过引入图卷积和注意力机制，进一步优化高阶特征信息的提取，同时减少数据噪声对模型性能的影响；最后，将得到的微生物与药物的低维特征与关联特征串联起来，通过结合非线性与线性特征，进一步提高模型的性能。本文使用多层感知机(multi-layer perceptron, MLP)对最终的微生物-药物进行评分预测。通过效能评估，GATSAE模型的受试者工作特征曲线下面积(area under receiver operating characteristic curve, AUROC)及精确率-召回率曲线下的面积(area under the precision-recall curve, AUPR)分别达到了0.961 9和0.957 7，优于经典的机器学习方法和常见的深度学习方法。通过案例研究分析表明，GATSAE模型兼具优越性和可行性。

1 材料与方法

1.1 数据集整理

本文采用MDAD[20]、aBiofilm[21]、DrugVirus[22]这3种关联信息数据集评估模型的预测能力。

1）MDAD来自Sun等[20]的研究数据，其包含大量生物学、化学和药物学信息，采用多种方法进行药物物质受体预测，包括结构因素组学、序列分析、分类器(可从http://www.chengroup.cumt.edu.cn/MDAD/获得)。在处理完数据后，得到包括627种药物和142种微生物的1 152种已知关联。

2）aBiofilm数据库是专门用于细菌生物被膜研究的数据库[21]，其汇集了大量关于细菌被膜形成、结构和功能的数据，包括多种分子生物学、生物化学和显微技术的实验数据，以及生物医药膜的基础组学、转录组学和代谢组学分析结果(可从http://bioinfo.imtech.res.in/manojk/abiofilm/获得)。aBiofilm

包含了 1 720 种药物和 140 种微生物的 2 884 种关联数据。

3）DrugVirus 数据库是专门用于病毒和药物相互作用研究的数据库, 其提供了关于病毒和药物相互作用及其生物学效应的大量数据和信息[22](可从 https://drugvirus.info 获得), 记录了包括 SARS-CoV-2 在内的多种人类病毒的相关化合物的活性和发展。

最终, 从 95 种病毒和 175 种药物中选取了 933 种已知关联信息作为实验数据。表 1 记录了 3 种数据集的详细信息。

表 1　3 种数据集的数据记录

Table 1　Data recording for three datasets

数据集	药物种类	微生物种类	已知关联关系种类
MDAD	627	142	1 152
aBiofilm	1 720	140	2 884
DrugVirus	175	95	933

1.2　异构网络

推断人类微生物-药物网络中的新关系可以看作是一种生物二部图网络预测, GATSAE 模型整体框架如图 1 所示。本文将微生物与药物表示为 2 种不同类型的节点, 定义了一个二维矩阵来存储微生物-药物的关联信息, 具体如下:

$$A(i,j)=\begin{cases}1 & \text{如果}\boldsymbol{d}_i\text{与}\boldsymbol{m}_j\text{之间存在关系}\\ 0 & \text{否则}\end{cases} \tag{1}$$

式中: 微生物-药物关联矩阵元素 $A(i,j)\in\mathbf{R}^{N_\mathrm{d}\times N_\mathrm{m}}$, N_d 为药物的种类数量, N_m 为微生物的种类数量, $A(i,j)=1$ 表示药物 $\boldsymbol{d}_i$ 与微生物 $\boldsymbol{m}_j$ 存在关联, 反之, $A(i,j)=0$ 表示没有关联, $i\in(1,2,\cdots,N_\mathrm{d})$, $j\in(1,2,\cdots,N_\mathrm{m})$, $\boldsymbol{d}_i$ 和 $\boldsymbol{m}_j$ 分别为第 i 个药物和第 j 个微生物。

1.2.1　药物结构相似性

SMILES 是一种用于表示化学分子结构的字符串表示法, 其可以用一行符号方式准确地描述所有药物化合物的化学结构。实验中, 使用 SMILES 的化学开发工具包[23-24]计算药物间的相似度; 从 DrugBank 中收集所有药物的规范 SMILES, 并使用 Chemical Development Kit 工具计算所有药物的化学指纹图谱; 根据药物化学指纹图谱的 Tanimoto 分数测量药物间的结构相似度, 用 $\boldsymbol{D}_\mathrm{s}\in\mathbf{R}^{N_\mathrm{d}\times N_\mathrm{d}}$ 表示药物间结构相似性, $D_\mathrm{s}(\boldsymbol{d}_i,\boldsymbol{d}_j)$ 表示药物 $\boldsymbol{d}_i$ 和 $\boldsymbol{d}_j$ 间的相似度。

1.2.2　微生物序列相似性

本文使用与文献 [25] 中类似的方法计算微生物的序列相似度。从 STRING v11 数据库中检索蛋白质-蛋白质功能相互作用网络(https://string-db.org), 通过检索每个微生物的蛋白序列, 计算微生物间的隐藏关系(更多有关微生物序列相似计算的信息可从文献 [25] 中查找)。用 $\boldsymbol{M}_\mathrm{s}\in\mathbf{R}^{N_\mathrm{m}\times N_\mathrm{m}}$ 表示微生物的序列相似性, $M_\mathrm{s}(\boldsymbol{m}_i,\boldsymbol{m}_j)$ 表示微生物 $\boldsymbol{m}_i$、$\boldsymbol{m}_j$ 间的相似度。

1.2.3　高斯相互作用谱核相似性

高斯相互作用谱核(Gaussian interaction profile, GIP)常用来计算生物实体间的相似度, 其能够较好地捕捉分子之间的相互作用关系, 包括结构、功能等, 如 microbe-disease 关联[26]、lncRNA-disease 关联[27]、drug-disease 关联[28-29] 等。受上述研究启发, 功能相似的药物可能与相似的微生物有关联。GIP 核相似度可用来衡量药物和微生物的相似度, 从而补充了药物和微生物的相似信息。药物 $\boldsymbol{d}_i$、$\boldsymbol{d}_j$ 间的 GIP 核相似度计算式如下:

$$D_\mathrm{g}(\boldsymbol{d}_i,\boldsymbol{d}_j)=\exp(-\eta_\mathrm{d}\|\boldsymbol{A}_{\boldsymbol{d}_i}-\boldsymbol{A}_{\boldsymbol{d}_j}\|^2) \tag{2}$$

式中: $D_\mathrm{g}(\boldsymbol{d}_i,\boldsymbol{d}_j)$ 为药物 $\boldsymbol{d}_i$、$\boldsymbol{d}_j$ 间的 GIP 核相似度; $\boldsymbol{A}_{\boldsymbol{d}_i}$ 表示微生物-药物关联矩阵 $\boldsymbol{A}$ 的第 i 个行向量作为药物 $\boldsymbol{d}_i$ 的相互作用谱核; $\boldsymbol{A}_{\boldsymbol{d}_j}$ 为药物 $\boldsymbol{d}_j$ 的相互作用谱核; η_d 为归一化的核带宽, 其控制着相似性计算的平滑程度, 定义为

$$\eta_\mathrm{d}=\eta'_\mathrm{d}\Big/\left(\frac{1}{N_\mathrm{d}}\sum_{i=1}^{N_\mathrm{d}}\|\boldsymbol{A}_{\boldsymbol{d}_i}\|^2\right) \tag{3}$$

其中: η'_d 为 GIP 原始核带宽。

与药物类似, 微生物 $\boldsymbol{m}_i$、$\boldsymbol{m}_j$ 间的 GIP 核相似度计算式如下:

$$M_\mathrm{g}(\boldsymbol{m}_i,\boldsymbol{m}_j)=\exp(-\eta_\mathrm{m}\|\boldsymbol{A}_{\boldsymbol{m}_i}-\boldsymbol{A}_{\boldsymbol{m}_j}\|^2) \tag{4}$$

式中: $M_\mathrm{g}(\boldsymbol{m}_i,\boldsymbol{m}_j)$ 为微生物 $\boldsymbol{m}_i$、$\boldsymbol{m}_j$ 间的 GIP 核相似度; $\boldsymbol{A}_{\boldsymbol{m}_i}$ 表示微生物-药物邻接矩阵 $\boldsymbol{A}$ 的第 i 个列向量作为微生物 $\boldsymbol{m}_i$ 的相互作用谱核; $\boldsymbol{A}_{\boldsymbol{m}_j}$ 为微生物 $\boldsymbol{m}_j$ 的相互作用谱核; η_m 为归一化的核带宽, 定义为

$$\eta_\mathrm{m}=\eta'_\mathrm{m}\Big/\left(\frac{1}{N_\mathrm{m}}\sum_{i=1}^{N_\mathrm{m}}\|\boldsymbol{A}_{\boldsymbol{m}_i}\|^2\right) \tag{5}$$

其中: η'_m 为 GIP 原始核带宽。

1.2.4　集成相似性

由于单一的药物结构相似性、微生物序列相似性存在某些关联缺陷, 通过整合多源数据及不同的表示方法互补信息, 采用集成相似性能更好地表示微生物间和药物之间的关联度, 同时克服某些关联缺陷和稀疏性。2 种药物 $\{\boldsymbol{d}_i,\boldsymbol{d}_j\}$ 之间的集成相似矩阵如下:

$$D_\mathrm{F}(\boldsymbol{d}_i,\boldsymbol{d}_j)=\begin{cases}\dfrac{D_\mathrm{s}(\boldsymbol{d}_i,\boldsymbol{d}_j)+D_\mathrm{g}(\boldsymbol{d}_i,\boldsymbol{d}_j)}{2} & D_\mathrm{s}(\boldsymbol{d}_i,\boldsymbol{d}_j)\neq 0\\ D_\mathrm{g}(\boldsymbol{d}_i,\boldsymbol{d}_j) & \text{否则}\end{cases} \tag{6}$$

式中: $\boldsymbol{D}_\mathrm{F}\in\mathbf{R}^{N_\mathrm{d}\times N_\mathrm{d}}$。

同样, 2 种微生物 $\{\boldsymbol{m}_i,\boldsymbol{m}_j\}$ 之间的集成相似矩阵为

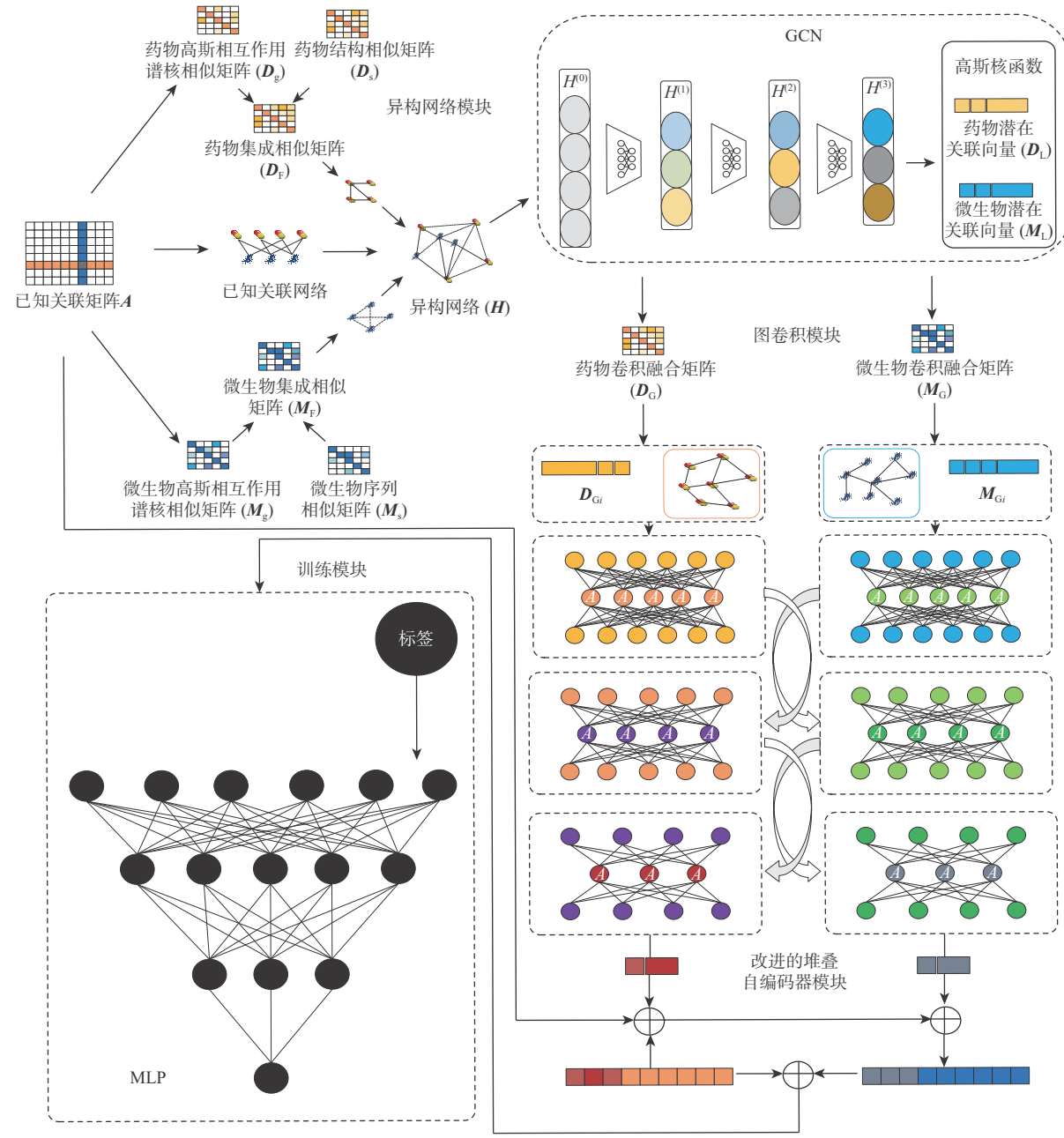


图 1 GATSAE 模型框架

Fig. 1 Framework of GATSAE model

$$M_F(\boldsymbol{m}_i,\boldsymbol{m}_j)=\begin{cases}\dfrac{M_s(\boldsymbol{m}_i,\boldsymbol{m}_j)+M_g(\boldsymbol{m}_i,\boldsymbol{m}_j)}{2} & M_s(\boldsymbol{m}_i,\boldsymbol{m}_j)\neq 0\\ M_g(\boldsymbol{m}_i,\boldsymbol{m}_j) & \text{否则}\end{cases} \tag{7}$$

式中：$\boldsymbol{M}_F\in\mathbf{R}^{N_m\times N_m}$。

1.2.5 特征网络

相比于单一的相似网络，异构网络将不同类型的节点和链路结合起来，这些节点和链路具有不同的属性、关系和语义信息，在微生物-药物关联预测中丰富数据的多样性，提供多视角信息用以挖掘隐藏关联关系。异构网络如下：

$$\boldsymbol{H}=\begin{bmatrix}\boldsymbol{D}_F & \boldsymbol{A}\\ \boldsymbol{A}^T & \boldsymbol{M}_F\end{bmatrix} \tag{8}$$

式中：微生物与药物的关联异构网络 $\boldsymbol{H}\in\mathbf{R}^{(N_d+N_m)\times(N_d+N_m)}$。

1.3 本文方法

1.3.1 图卷积网络

GCN 是一种可以学习潜在特征的神经网络，主要用于图结构，其能够解决节点分类、图分类、链路预测等问题[30]，提取到不同层的低维特征，从而捕获图网络中不同结构的信息，第 1 层获取直接邻居信息，更高层通过迭代更新低维特征来捕获多阶邻居信息，但由于 GCN 在提取潜在特征时会减少对图中度较小的节点的关注，从而丢失部分信息。因此，在 GCN 模块引入高斯核函数，通过在 GCN 的节点间关系计算中引入高斯核函数，调整了节点之间的相似性度量。高斯核函数考虑了节点之间

的距离, 赋予距离较近的节点更高的权重, 使得 GCN 能够减小对图中度较小的节点的忽略。

利用式 (8) 得到的异构网络作为卷积网络学习的原始数据, 从中学习并提取出潜在关系, 具体定义如下:

$$D_{\mathrm{L}}(\boldsymbol{d}_i,\boldsymbol{d}_j)=\exp(-\gamma\|\boldsymbol{X}_{d_i}-\boldsymbol{X}_{d_j}\|^2) \tag{9}$$

$$M_{\mathrm{L}}(\boldsymbol{m}_i,\boldsymbol{m}_j)=\exp(-\gamma\|\boldsymbol{X}_{m_i}-\boldsymbol{X}_{m_j}\|^2) \tag{10}$$

式中: $\boldsymbol{D}_{\mathrm{L}}$ 为 $\boldsymbol{X}_{\boldsymbol{d}}$ 经过 GIP 核函数得到的药物潜在关联矩阵; $\boldsymbol{M}_{\mathrm{L}}$ 为 $\boldsymbol{X}_{\boldsymbol{m}}$ 经过 GIP 核函数得到的潜在微生物关联矩阵; $\boldsymbol{X}_{d_i}$ 为 $\boldsymbol{X}_{\boldsymbol{d}}$ 药物潜在特征矩阵中第 i 个药物特征向量; $\boldsymbol{X}_{d_j}$ 为第 j 个药物特征向量; $\boldsymbol{X}_{m_i}$ 为 $\boldsymbol{X}_{\boldsymbol{m}}$ 微生物潜在特征矩阵中第 i 个微生物特征向量; $\boldsymbol{X}_{m_j}$ 为第 j 个微生物特征向量; γ 为核带宽, 其定义与式 (3)、式 (5) 类似。

潜在特征矩阵 $\boldsymbol{X}$ 的具体定义为

$$\boldsymbol{X}^{(l)}=f(\boldsymbol{X}^{(l-1)},\boldsymbol{H})=\sigma(\boldsymbol{D}^{-1/2}\boldsymbol{H}\boldsymbol{D}^{-1/2}\boldsymbol{X}^{(l-1)}\boldsymbol{W}^{(l-1)}) \tag{11}$$

式中: $\boldsymbol{X}^{(l)}$ 表示节点的第 l 层嵌入; $\boldsymbol{D}=\mathrm{diag}\left(\sum_{i=j=1}^{N_{\mathrm{d}}+N_{\mathrm{m}}}H_{ij}\right)$ 为 $\boldsymbol{H}$ 的对角节点度矩阵; $\boldsymbol{W}^{(l)}\in\mathbf{R}^{(N_{\mathrm{d}}+N_{\mathrm{m}})\times k_l}$ 为第 l 层 GCN 权重矩阵, k_l 为第 l 层 GCN 嵌入维度; $\sigma(\cdot)$ 为 ReLU 非线性激活函数, 其函数形式为 $f(x)=\max(0,x)$。设置 $l=3$, 即 3 层 GCN。对于 GCN 的第 1 层嵌入, 表示如下:

$$\boldsymbol{X}^1=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{A}\\ \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{0}\end{bmatrix} \tag{12}$$

为提高药物和微生物的表达能力, 综合考虑拓扑结构和相似信息, 将其潜在关联矩阵与集成相似矩阵加权融合得到卷积融合矩阵, 表示为

$$\boldsymbol{D}_{\mathrm{G}}=w_1\boldsymbol{D}_{\mathrm{L}}+w_2\boldsymbol{D}_{\mathrm{s}} \tag{13}$$

$$\boldsymbol{M}_{\mathrm{G}}=w_3\boldsymbol{M}_{\mathrm{L}}+w_4\boldsymbol{M}_{\mathrm{s}} \tag{14}$$

式中: $\boldsymbol{D}_{\mathrm{G}}$ 为药物卷积融合矩阵; $\boldsymbol{M}_{\mathrm{G}}$ 为微生物卷积融合矩阵; w_1、w_2、w_3、w_4 为对应权重参数, 设置 $w_1=w_2=w_3=w_4=0.5$。

构建 GCN 损失函数, 选择 Adam 作为优化器, 损失函数为

$$L_{\mathrm{D}}=\sum_{i,j=1}^{N_{\mathrm{d}}}|D_{\mathrm{G}}(i,j)-D_{\mathrm{F}}(i,j)|^2 \tag{15}$$

$$L_{\mathrm{M}}=\sum_{i,j=1}^{N_{\mathrm{m}}}|M_{\mathrm{G}}(i,j)-M_{\mathrm{F}}(i,j)|^2 \tag{16}$$

1.3.2　改进的堆叠自编码器

传统堆叠自编码器通过堆叠多个自编码器来形成深层网络, 这种层堆叠的结构允许模型逐渐学习数据的层次性特征和抽象表示, 从而提高学习的表达能力。SAEMDA[31] 和 SMALF[32] 等方法利用堆叠自编码器提取潜在特征, 以挖掘新的关联。相似地, SGAEMDA[33] 在堆叠自编码器的基础上引入图卷积, 增强了自编码器在处理图数据方面的能力。然而, 这些方法忽略了数据噪声可能带来的影响。

基于此, 本文采用改进的堆叠自编码器, 通过结合图卷积和注意力机制, 使 GCN 专注于邻居节点的信息传播, 图注意力机制根据节点重要性自适应地分配权重, 减小数据噪声的影响, 从而更好地融合局部和全局信息, 提高特征提取能力。

药物和微生物的特征信息是高维的, 这往往会影响预测模型的精度, 为了更好地提取有用信息, 使用图注意力堆叠自编码器对数据降维。

具体来说, 图注意力堆叠自编码器定义如下:

$$\mathrm{Enc}(\boldsymbol{Y},\boldsymbol{\delta})=\tanh(\boldsymbol{Y}\cdot\mathrm{ReLU}(\boldsymbol{Y}\boldsymbol{\delta}\boldsymbol{\phi}^1)\boldsymbol{\phi}^2) \tag{17}$$

$$\boldsymbol{\tau}(\boldsymbol{Y},\boldsymbol{\delta})=\sum\mathrm{sigmoid}(\mathrm{LeakyReLU}(\boldsymbol{\omega}^{\mathrm{T}}[\boldsymbol{Y}\|\boldsymbol{\delta}]))\cdot \mathrm{Enc}(\boldsymbol{Y},\boldsymbol{\delta}) \tag{18}$$

$$\mathrm{Dnc}(\boldsymbol{Y},\boldsymbol{\tau})=\mathrm{sigmoid}(\boldsymbol{Y}\cdot\mathrm{ReLU}(\boldsymbol{Y}\boldsymbol{\tau}\boldsymbol{\phi}^3)\boldsymbol{\phi}^4) \tag{19}$$

式中: Enc 表示编码器; Dnc 表示解码器; $\boldsymbol{Y}$ 为微生物和药物卷积融合矩阵 $\boldsymbol{D}_{\mathrm{G}}$、$\boldsymbol{M}_{\mathrm{G}}$ 的邻接矩阵; $\boldsymbol{\delta}$ 为卷积融合矩阵的节点特征矩阵; "||" 表示将 2 个向量拼接; $\boldsymbol{\tau}$ 为注意力矩阵; $\boldsymbol{\omega}^{\mathrm{T}}$ 为可学习的注意参数矩阵; $\boldsymbol{\phi}^1,\cdots,\boldsymbol{\phi}^4$ 为自编码器参数矩阵。

使用图注意力堆叠自编码器提取药物的低维特征表示, 如下:

$$\begin{cases}\boldsymbol{E}_{\mathrm{d}}^{(l)}=\mathrm{Enc}(\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{E}_{\mathrm{d}}^{(l-1)})\\ \boldsymbol{\tau}_{\mathrm{d}}^{(l)}=\boldsymbol{\tau}(\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{E}_{\mathrm{d}}^{(l)})\\ \boldsymbol{D}_{\mathrm{d}}^{(l)}=\mathrm{Dnc}(\boldsymbol{Y}_{\mathrm{D}},\boldsymbol{\tau}_{\mathrm{d}}^{(l)})\end{cases} \tag{20}$$

式中: $\boldsymbol{E}_{\mathrm{d}}^{(l)}$ 为第 l 层图注意力堆叠自编码器提取的药物低维特征表示, 设置 $l=3$, 当 $l=1$ 时, $\boldsymbol{E}_{\mathrm{d}}^{0}$ 表示药物卷积融合矩阵 $\boldsymbol{D}_{\mathrm{G}}$; $\boldsymbol{\tau}_{\mathrm{d}}^{(l)}$ 为第 l 层图注意力堆叠自编码器得到的药物注意力矩阵; $\boldsymbol{D}_{\mathrm{d}}^{(l)}$ 为重构特征矩阵。

同样, 微生物低维表示为

$$\begin{cases}\boldsymbol{E}_{\mathrm{m}}^{(l)}=\mathrm{Enc}(\boldsymbol{Y}_{\mathrm{M}},\boldsymbol{E}_{\mathrm{m}}^{(l-1)})\\ \boldsymbol{\tau}_{\mathrm{m}}^{(l)}=\boldsymbol{\tau}(\boldsymbol{Y}_{\mathrm{M}},\boldsymbol{E}_{\mathrm{m}}^{(l)})\\ \boldsymbol{D}_{\mathrm{m}}^{(l)}=\mathrm{Dnc}(\boldsymbol{Y}_{\mathrm{M}},\boldsymbol{\tau}_{\mathrm{m}}^{(l)})\end{cases} \tag{21}$$

通过堆叠 3 个图注意力自编码器来提取药物和微生物的低维特征表示, 如图 2 所示。具体来说, 将第 1 个图注意力堆叠自编码器生成的特征表示作为第 2 个自编码器的输入, 第 2 个自编码器生成另一个维数更低的特征表示, 以此类推, 直到构建 3 个图注意力自编码器。堆叠自编码器是基于重构损失函数协同训练的, 如下:

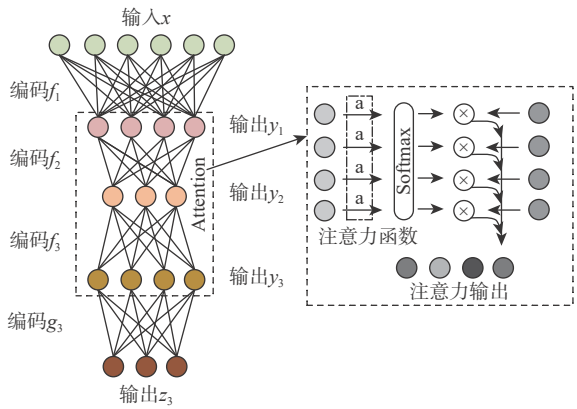


图2 图注意力堆叠自编码器结构

Fig. 2 Structure of graph attention stacking autoencoder

$$\begin{cases} L_{\mathrm{d}} = \sum_{l=1}^{L} \|\boldsymbol{E}_{\mathrm{d}}^{(l-1)} - \boldsymbol{D}_{\mathrm{d}}^{(l)}\|^2 \\ L_{\mathrm{m}} = \sum_{l=1}^{L} \|\boldsymbol{E}_{\mathrm{m}}^{(l-1)} - \boldsymbol{D}_{\mathrm{m}}^{(l)}\|^2 \end{cases} \tag{22}$$

1.4 串联数据训练MLP

为使药物和微生物的特征数据中包含更多潜在的信息，将药物和微生物的低维表示 $\boldsymbol{E}_{\mathrm{d}}$、$\boldsymbol{E}_{\mathrm{m}}$ 与关联矩阵 $\boldsymbol{A}$ 串联，如下：

$$\boldsymbol{Z}_{\mathrm{d}} = \mathrm{connecting}(\boldsymbol{E}_{\mathrm{d}}, \boldsymbol{A}) \tag{23}$$

$$\boldsymbol{Z}_{\mathrm{m}} = \mathrm{connecting}(\boldsymbol{E}_{\mathrm{m}}, \boldsymbol{A}^{\mathrm{T}}) \tag{24}$$

式中：$\boldsymbol{Z}_{\mathrm{d}}$ 和 $\boldsymbol{Z}_{\mathrm{m}}$ 分别为药物和微生物的最终嵌入特征。

随机选取与正例样本相同大小的负例样本组成训练集，具体如下：

$$T_{ij} = \mathrm{connecting}(\boldsymbol{Z}_{\boldsymbol{d}_i}, \boldsymbol{Z}_{\boldsymbol{m}_i}) \tag{25}$$

式中：$\boldsymbol{T} \in \mathbf{R}^{(1\,152\times2)\times(691+206)}$，$T_{ij}$ 为药物 $\boldsymbol{d}_i$ 和微生物 $\boldsymbol{m}_j$ 的关联嵌入，1 152×2 表示正例与负例样本分别为1 152对，691表示微生物嵌入维度，206表示药物嵌入维度。

将训练集放入MLP中进行学习训练，对未知微生物-药物进行预测，具体如下：

$$\boldsymbol{T}^{(l)} = \mathrm{ReLU}(\boldsymbol{T}^{(l-1)}\boldsymbol{w}^{(l)} + \boldsymbol{b}^{(l)}) \tag{26}$$

$$\hat{y} = \mathrm{sigmoid}(\boldsymbol{T}^{2}\boldsymbol{w}^{3} + \boldsymbol{b}^{3}) \tag{27}$$

式中：$l=2$ 表示隐藏层的层数；$\boldsymbol{T}^{(l)}$ 为第 l 层隐藏层的输出；$\boldsymbol{w}^{(l)}$ 为第 l 层隐藏层可学习参数矩阵；$\boldsymbol{b}^{(l)}$ 为第 l 层隐藏层偏置；$\hat{y}$ 为微生物-药物关联预测分数。

使用Adam优化器和最小化二元交叉熵损失函数的误差来训练MLP，损失函数为

$$L_{\mathrm{f}} = -\frac{1}{N}\left[\sum_{i,j\in y^{+}} y_{ij}\ln\hat{y}_{ij} + \sum_{i,j\in y^{-}} (1-y_{ij})\ln(1-\hat{y}_{ij})\right] \tag{28}$$

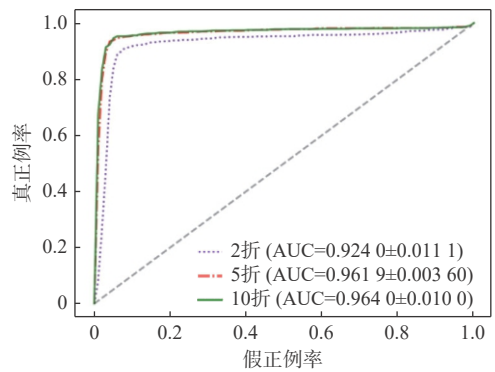
式中：y_{ij} 为正例标签；$\hat{y}_{ij}$ 为预测是正例的概率；$1-y_{ij}$ 表示负例标签；$1-\hat{y}_{ij}$ 表示预测为负例的概率；N 为训练集中微生物-药物关联对数目。

2 实验与结果分析

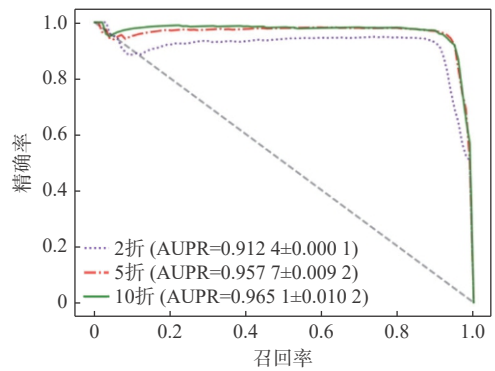
2.1 评估指标

k 折交叉验证是一种常用的模型评估方法，被广泛用于评估模型的性能。在 k 折交叉验证中，原始数据集被分成 k 份大小相似的子集。模型被训练 k 次，每次使用 $k-1$ 份作为训练集，剩下的1份作为验证集。每次训练后，模型的性能指标会被评估，并最终得到平均性能指标作为模型的评估结果。选择AUROC和AUPR作为评价指标。AUROC是以假正例率为横坐标、真正例率为纵坐标的接收者工作特征(receiver operating characteristic, ROC)曲线下面积，AUPR是以召回率为横坐标、精确率为纵坐标的精确召回率(precision-recall, PR)曲线下面积。2折、5折和10折交叉验证得到GATSAE的ROC曲线和PR曲线如图3所示。

由图3可知，GATSAE模型在2折交叉验证下，AUC和AUPR值比5折、10折交叉验证低了约4%，10折和5折交叉验证相比提高的并不多，考虑到模型运行效率和内存消耗，以下实验均选择5折交叉验证作为基准。



(a) GATSAE在2,5,10折交叉验证下的ROC曲线



(b) GATSAE在2,5,10折交叉验证下的PR曲线

图3 GATSAE交叉验证下的ROC曲线和PR曲线

Fig. 3 ROC and PR curves under GATSAE cross validation

2.2 参数选择

参数分析能帮助科研人员优化模型性能、评估

模型的稳定性和复杂性, 并提供模型解释性[34]。本文中, 卷积核带宽 γ、卷积层层数 l 及堆叠自编码器的学习率 l_r 对模型的影响至关重要。所有参数实验均在 MDAD 数据集上进行。

2.2.1　卷积核带宽选择

不同卷积核带宽会导致不同的潜在关联矩阵 $\boldsymbol{D}_L$、$\boldsymbol{M}_L$, 影响模型的预测能力[34]。取 $\gamma \in \{2^{-10}, \cdots, 2^{0}, \cdots, 2^{10}\}$, 图 4 显示了 γ 对模型 AUPR 值的变化。可以看到, AUPR 在 $\gamma=2^{-3}$ 时达到最大值, 较大或较小的 γ 对模型性能有负面影响。

根据文献 [30] 分析, 较小的 γ 值会导致潜在特征矩阵接近于 1, 这意味着药物间和微生物间的相似度差值不大, 影响本文模型的预测效果。

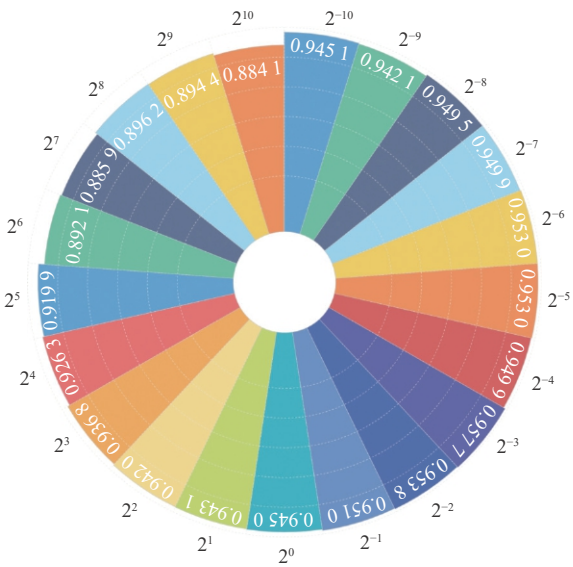


图 4　不同卷积核带宽 γ 对应模型 AUPR 值

Fig. 4　Bandwidth of different convolutional kernels γ Corresponding model AUPR value

2.2.2　卷积层层数选择

卷积层层数决定了提取数据的质量, 以发掘未知微生物-药物关联关系[30]。取卷积层层数 $l \in \{1,2,3,4,5\}$, 图 5 展示了不同卷积层层数 l 对非卷积模型的 AUPR 值增长比率。可以看到, 仅有 1 层卷积的模型效果最好, 比非卷积模型的 AUPR 值提高了 4.53%。但是, 1 层卷积只提取近邻信息而忽略更深层次潜在特征, 不适用于模型泛化。另外, 随着卷积层层数的叠加, 模型的效果反而越来越低, 可能是因为 GCN 过度学习导致。因此, 选择 3 层卷积层, 即 $l=3$。

2.2.3　堆叠自编码器学习率选择

学习率的选择对模型的训练优化过程有着重要的影响。自编码器的损失越小, 编码-解码过程丢弃的信息就会越少, 模型的预测精准度就会越高。取 $l_r \in \{0.000\,1,\cdots,0.001,\cdots,0.03\}$, 图 6 显示了不同学习率对训练损失的变化, 设置训练轮数为 10 的倍数, 共 200 次。可以看到, 较低的学习率增加了模型的训练时间, 消耗大量内存; 而较高的学习率导致了较大的损失波动, 使得训练不稳定。基于此, 选择 $l_r = 0.01$ 作为堆叠自编码器的学习率。

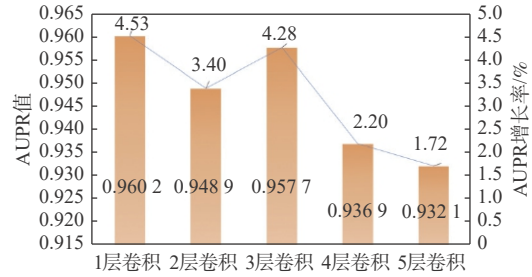


图 5　不同卷积层层数 l 对非卷积模型 AUPR 值的增长率

Fig. 5　Growth rate of AUPR values in non convolutional models with different number of convolutional layers l

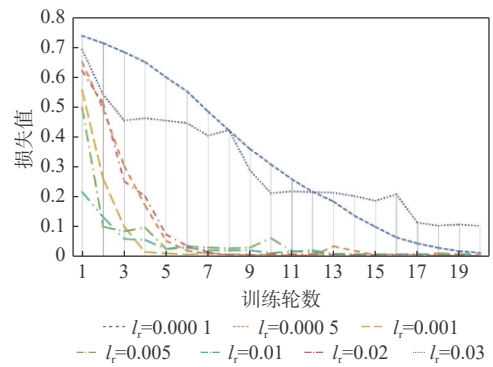


图 6　不同学习率对训练损失的变化值

Fig. 6　Variation value of training loss under different learning rates

通过上述参数选择分析, 当选择卷积核带宽 $\gamma = 2^{-3}$、卷积层层数 $l = 3$ 及堆叠自编码器的学习率 $l_r = 0.01$ 时, GATSAE 可获得最佳预测性能。

2.3　消融实验

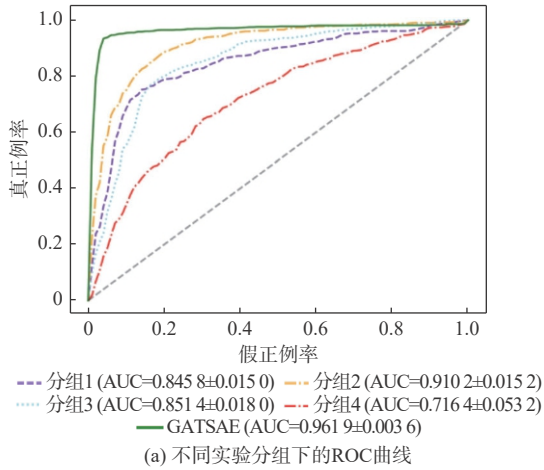
为验证 GCN、图注意力堆叠自编码器及数据串联对模型 GATSAE 预测精度的影响, 做了 5 组对比实验, 对比实验如表 2 所示。

表 2　对比实验分组信息

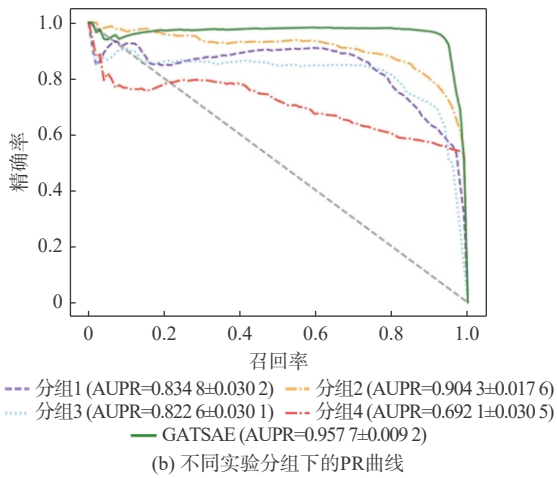
Table 2　Comparison experiment grouping information

实验组	GCN	图注意力堆叠自编码器	数据串联
分组1	引入	无	无
分组2	无	引入	无
分组3	无	无	引入
分组4	无	无	无
GATSAE	引入	引入	引入

对上述 5 组实验进行 5 折交叉验证, 得到 ROC 曲线和 PR 曲线, 如图 7 所示。可以看到, GATSAE 模型无论在 AUC 上或是在 AUPR 上都明显优于其



(a) 不同实验分组下的ROC曲线



(b) 不同实验分组下的PR曲线

图 7 消融实验下的 ROC 和 PR 曲线对比

Fig. 7 Comparison of ROC and PR curves under ablation experiments

他对照实验。其中，第 1 组实验的 AUC 值比第 4 组高 0.129 4，AUPR 值高 0.142 7，表明模型引入 3 层 GCN 后获得了更多潜在微生物和药物的关联信息，提高了模型的预测能力；第 2 组的 AUC 值比第 4 组高 0.193 8，AUPR 值高 0.212 2，说明图注意力堆叠自编码器能够在众多关联信息中找到最有价值的信息，减少对非重要关联关系的依赖，对模型的训练优化有很大的帮助；第 3 组实验的 AUC 值比第 4 组高 0.135，AUPR 值高 0.130 5，揭示了数据串联能够增加数据的多样性，实现数据增强，进而提高模型的预测精度。综上表明，引入 GCN、图注意力堆叠自编码器及数据串联对 GATSAE 预测精度的提高至关重要。

2.4 不同数据集对比

为验证模型的适用性和鲁棒性，选择 3 种数据集，分别为 MDAD[20]、aBiofilm[21]、DrugVirus[22]，详细信息如表 1 所示。可以发现，MDAD 中药物数比 aBiofilm 少 2 倍，同时比 DrugVirus 多 3 倍；MDAD 关联信息比 aBiofilm 少 1 倍。经过 5 折交叉验证后，3 种数据集下 AUC 和 AUPR 值如图 8 所示。

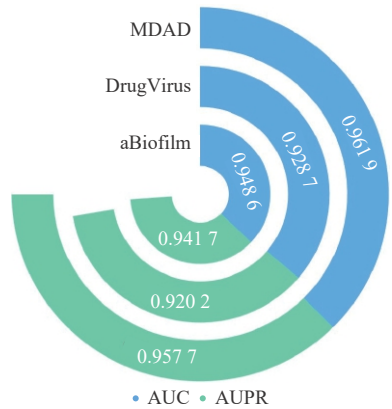


图 8 不同数据集下 GATSAE 预测结果对比

Fig. 8 Comparison of GATSAE prediction results under different datasets

由图 8 可知，GATSAE 在 MDAD、aBiofilm、DrugVirus 数据集上的 AUC 值分别为 0.961 9、0.948 6、0.928 7；AUPR 值分别为 0.957 7、0.941 7、0.920 2。实验结果表明，GATSAE 在不同大小数据集上 AUC 和 AUPR 值均高于 0.92 的预测结果，说明本文模型有较强的鲁棒性。

2.5 分类模型对比

在分类问题中，AUC 是评价模型整体性能的重要方法，对于非平衡数据集，AUPR 比 AUC 更能评价模型。为了更全面地评估 GATSAE 模型的性能，采用准确度 A_{cc}、精度 P_{re}、召回率 R_{ec} 和 F_1 分数等指标，其计算方法如下：

$$A_{cc}=\frac{T_P+T_N}{T_P+T_N+F_P+F_N} \tag{29}$$

$$P_{re}=\frac{T_P}{T_P+F_P} \tag{30}$$

$$R_{ec}=\frac{T_P}{T_P+F_N} \tag{31}$$

$$F_1=\frac{2P_{re}R_{ec}}{P_{re}+R_{ec}} \tag{32}$$

式中：T_P、T_N、F_P、F_N 分别为真阳性、真阴性、假阳性、假阴性。

为验证 MLP 在 GATSAE 模型中的优越性，在 MDAD 数据集下进行了 5 折交叉验证，并且与决策树（decision tree, DT）、随机森林（random forest, RF）、k 近邻（K-nearest neighbors, KNN）、支持向量机（support vector machine, SVM）这 4 种常见的分类模型作比较，结果如图 9 和表 3 所示。从表 3 中可以看到，MLP 在 5 个指标上均高出其他 4 个分类器，表明 GATSAE 模型选用 MLP 作为分类器能更好地预测出微生物-药物的潜在关联。

2.6 与现有方法对比

在相同数据集下，将 GATSAE 与深度学习模

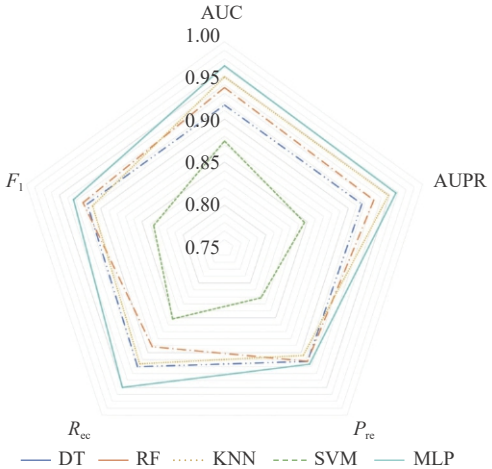


图 9　MLP 与常见分类模型比较

Fig. 9　Comparison between MLP and common classification models

表 3　MLP 与常见分类模型评价指标

Table 3　Evaluation indicators of MLP and common classification models

分类器	AUC	AUPR	P_{re}	R_{ec}	F_1
DT	0.916 3	0.916 4	0.912 6	0.920 3	0.916 2
RF	0.936 6	0.930 9	0.912 4	0.891 5	0.921 4
KNN	0.948 9	0.948 6	0.904 1	0.916 1	0.909 8
SVM	0.874 4	0.846 9	0.821 2	0.851 6	0.835 7
MLP	0.961 9	0.957 7	0.916 6	0.950 0	0.932 9

型 GCNMDA[17]、EGATMDA[18]、HGATDVA[19] 及最近研究较为新颖的微生物与药物关联预测深度学习模型 HNERMDA[35]、NIRBMMDA[36] 做了对比评估实验。HNERMDA 利用多种来源的生物医学数据构建了微生物与药物的异构网络, 同时引入了一种异构网络表征学习方法 metapath2vec, 以学习微生物和药物的低维嵌入表征, 并设计了一种偏置双向网络投影推荐算法, 以提高微生物-药物预测的准确性[35]。NIRBMMDA 提出了一种基于邻域推理(neighborhood inference, NI)和受限玻尔兹曼机(restricted Boltzmann machine, RBM)的计算模型, 利用综合微生物相似性、综合药物相似性和已知微生物-药物关联来预测潜在的微生物-药物关联[36]。

将对比模型的参数调到最佳, 但由于编译环境和设备的不同, 导致模型性能未达到最佳, 但性能差距不大。模型指标对比值如图 10 和表 4 所示。

可以看到, GATSAE 在各个指标上都优于 GCNMDA、EGATMDA、HGATDVA、HNERMDA 和 NIRBMMDA 模型, 其中, AUC 值分别提高了 0.037 9、0.018 5、0.046、0.064 2、0.113 4, AUPR 值分别提高了 0.045 3、0.024 2、0.056 9、0.055 1、0.125。由此可以看出, 本文模型在预测微生物-药物新关联方面有较好的效果。

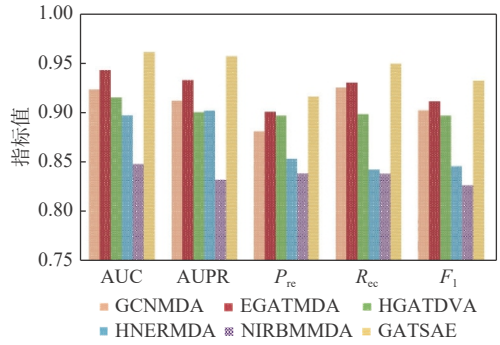


图 10　模型指标对比

Fig. 10　Comparison of model indicators

表 4　模型指标对比值

Table 4　Comparison value of model indicators

模型	AUC	AUPR	P_{re}	R_{ec}	F_1
GCNMDA	0.924 0	0.912 4	0.881 7	0.925 7	0.902 9
EGATMDA	0.943 4	0.933 5	0.901 5	0.921 3	0.911 8
HGATDVA	0.915 9	0.900 8	0.897 6	0.898 9	0.897 1
HNERMDA	0.897 7	0.902 6	0.853 7	0.843 0	0.846 3
NIRBMMDA	0.848 5	0.832 7	0.839 1	0.838 8	0.826 8
GATSAE	0.961 9	0.957 7	0.916 6	0.930 0	0.932 9

3　案例分析

本节在 3 个数据集上进行了案例研究, 进一步验证 GATSAE 的预测效果。在 MDAD 数据集上验证了与大肠杆菌 (Escherichia coli) 相关排名前 10 的药物; 在 aBiofilm 数据集上验证了与阿司匹林 (Aspirin) 相关排名前 10 的微生物。由于新型 SARS-CoV-2 病毒引起的大流行影响了全世界, 迫切需要有效的预防和治疗药物, 也在 DrugVirus 数据集上做了关于 SARS-CoV-2 相关排名前 20 的药物。3 个数据集上的验证结果如表 5～表 7 所示。

可以看到, 在 MDAD 上预测与大肠杆菌相关排名前 10 的药物及在 DrugVirus 上预测与 SARS-CoV-2 相关排名前 20 的药物都得到了验证, 在 aBiofilm 上预测与阿司匹林相关排名前 10 的微生物也仅有 2 种没有文献支持。

针对与 SARS-CoV-2 相关的前 20 名预测药物中, 其中一些已经进入临床试验或文献证明可以抑制及治疗 SARS-CoV-2。Wang 等[37] 报道, 氯喹 (Chloroquine, 排名 1) 和瑞德西韦 (Remdesivir, 排名 13) 在体外控制 COVID-19 感染方面有显著效果。法维匹拉韦 (Favipiravir, 排名 3) 作为核苷酸类似物选择性抑制病毒 RNA 依赖性 RNA 聚合酶, 掺入病

表 5 与大肠杆菌相关排名前 10 的药物

Table 5 Top 10 drugs related to Escherichia coli

排名	相关药物	PMID
1	Ceftizoxime	6 299 968
2	Aminosalicylic Acid	33 468 700
3	Citral	35 776 056
4	Clozapine	25 448 498
5	Palmitic acid	29 719 215
6	Esculin	15 137 927
7	Azlocillin	7 033 199
8	Azidocillin	4 563 142
9	Aspirin	30 658 983
10	Glipizide	32 995 125

表 6 与阿司匹林相关排名前 10 的微生物

Table 6 Top 10 microbes related to Aspirin

排名	相关微生物	PMID
1	Candida albicans	33 242 673
2	Pseudomonas aeruginosa	25 088 031
3	Staphylococcus epidermidis	12 555 346
4	Human immunodeficiency virus	28 480 270
5	Streptococcus mutans	unconfirmed
6	Staphylococcus aureus	34 692 677
7	Mycobacterium tuberculosis	23 997 233
8	Escherichia coli	30 658 983
9	Clostridium perfringens	31 865 684
10	Human herpesvirus	unconfirmed

表 7 与 SARS-CoV-2 相关排名前 20 的药物

Table 7 Top 20 drugs related to SARS-CoV-2

排名	相关微生物	PMID
1	Chloroquine	35 859 449
2	ABT	37 414 987
3	Favipiravir	33 108 587
4	BCX	35 062 212
5	Luteolin	32 389 723
6	Amodiaquine	32 486 229
7	Cyclosporine	34 081 806
8	Emetine	33 302 852
9	Gemcitabine	32 432 977
10	Hydroxychloroquine	32 373 993
11	Amiodarone	36 426 888
12	Obatoclax	34 989 664
13	Remdesivir	33 436 624
14	Chlorpromazine	32 773 341
15	Nelfinavir	35 390 430
16	EIPA	37 632 140
17	Arbidol	32 955 901
18	Niclosamide	35 348 204
19	Dasatinib	36 704 839
20	Eflornithine	34 055 746

毒 RNA 后引起致命突变，被证实可用于治疗 COVID-19[38]。木犀草素 (Luteolin, 排名 5) 被证实可以用来缓解 COVID-19 通过致病性和应激刺激的肥大细胞释放激活小胶质细胞的介质，从而导致下丘脑发炎的病症[39]。Choy 等[40] 研究显示，依米丁 (Emetine, 排名 8) 和瑞德西韦 (Remdesivir, 排名 13) 对 SARS-CoV-2 病毒具有抗病毒作用，它们联合在体外表现出协同效应。此外，还有一些预测药物目前尚未得到文献支持，但已被列入广谱抗病毒药物 (broad-spectrum antiviral agents, BSAAs) 名单[41]，值得考虑对新型 SARS-CoV-2 病毒进行进一步研究。

这些预测结果表明，GATSAE 在微生物-药物关联预测方面有良好的性能，具有一定的应用价值。

4 结 论

为解决传统方法挖掘微生物-药物新关联耗时且费力的问题，本文提出了一种基于图注意力堆叠自编码器微生物-药物关联预测方法。

1）GATSAE 模型实现了较为优异的预测性能，其 AUC、AUPR、精度、召回率、F_1 分数均超过了 0.91，并且在 3 种不同数据集上的 AUC、AUPR 超过了 0.92，说明模型兼具优异性和鲁棒性。

2）通过异构网络和 3 层 GCN 挖掘出微生物-药物的潜在关联，解决了传统方法微生物和药物的信息利用不充分的问题，提高模型的预测性能，同时引入高斯核函数解决 GCN 在提取特征时忽略部分信息的问题。

3）在堆叠自编码器的基础上追加了图卷积和注意力机制，使得自编码器在实现数据降维的同时选择保留更有意义的特征，进一步优化了高阶特征信息的提取，减少了数据噪声对模型性能的影响，有利于模型的训练。

4）数据串联丰富了数据的多样性，同时补充了自编码器在学习高阶特征的无监督低维表示时部分信息丢失的问题，提高预测精度的同时能加快模型收敛。

今后将整合更多的生物信息进一步提升模型的性能。

参考文献（References）

[1] 杨博图. 基于相似性信息的微生物-药物关联关系预测方法研究[D]. 长沙: 中南大学, 2022: 69.
YANG B T. Study on prediction method of microbial-drug correlation based on similarity information[D]. Changsha: Central South University, 2022: 69(in Chinese).

[2] SHREINER A B, KAO J Y, YOUNG V B. The gut microbiome in health and in disease[J]. Current Opinion in Gastroenterology, 2015,

31(1): 69-75.

[3] LEY R E, BÄCKHED F, TURNBAUGH P, et al. Obesity alters gut microbial ecology[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(31): 11070-11075.

[4] TURNBAUGH P J, RIDAURA V K, FAITH J J, et al. The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice[J]. Science Translational Medicine, 2009, 1(6): 6ra14.

[5] GOLDMAN E. Antibiotic abuse in animal agriculture: exacerbating drug resistance in human pathogens[J]. Human and Ecological Risk Assessment: an International Journal, 2004, 10(1): 121-134.

[6] VRBANAC A, DEBELIUS J W, JIANG L J, et al. An elegan(t) screen for drug-microbe interactions[J]. Cell Host & Microbe, 2017, 21(5): 555-556.

[7] AARNOUDSE A L H J, DIELEMAN J P, VISSER L E, et al. Common ATP-binding cassette B1 variants are associated with increased digoxin serum concentration[J]. Pharmacogenetics and Genomics, 2008, 18(4): 299-305.

[8] HAISER H J, SEIM K L, BALSKUS E P, et al. Mechanistic insight into digoxin inactivation by *Eggerthella lenta* augments our understanding of its pharmacokinetics[J]. Gut Microbes, 2014, 5(2): 233-238.

[9] ONG F S, DEIGNAN J L, KUO J Z, et al. Clinical utility of pharmacogenetic biomarkers in cardiovascular therapeutics: a challenge for clinical implementation[J]. Pharmacogenomics, 2012, 13(4): 465-475.

[10] VOORA D, SHAH S H, SPASOJEVIC I, et al. The SLCO1B1*5 genetic variant is associated with statin-induced side effects[J]. Journal of the American College of Cardiology, 2009, 54(17): 1609-1616.

[11] RAMSEY L B, JOHNSON S G, CAUDLE K E, et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update[J]. Clinical Pharmacology & Therapeutics, 2014, 96(4): 423-428.

[12] VIOLI F, LIP G Y, PIGNATELLI P, et al. Interaction between dietary vitamin K intake and anticoagulation by vitamin K antagonists: is it really true? : a systematic review[J]. Medicine, 2016, 95(10): e2895.

[13] GUTHRIE L, GUPTA S, DAILY J, et al. Human microbiome signatures of differential colorectal cancer drug metabolism[J]. NPJ Biofilms and Microbiomes, 2017, 3: 27.

[14] ZHU L Z, DUAN G H, YAN C, et al. Prediction of microbe-drug associations based on chemical structures and the KATZ measure[J]. Current Bioinformatics, 2021, 16(6): 807-819.

[15] MA Y J, LIU Q Q. Generalized matrix factorization based on weighted hypergraph learning for microbe-drug association prediction[J]. Computers in Biology and Medicine, 2022, 145: 105503.

[16] ZHU B, XU Y, ZHAO P C, et al. NNAN: nearest neighbor attention network to predict drug-microbe associations[J]. Frontiers in Microbiology, 2022, 13: 846915.

[17] LONG Y H, WU M, KWOH C K, et al. Predicting human microbe-drug associations via graph convolutional network with conditional random field[J]. Bioinformatics, 2020, 36(19): 4918-4927.

[18] LONG Y H, WU M, LIU Y, et al. Ensembling graph attention networks for human microbe-drug association prediction[J]. Bioinformatics, 2020, 36(Supplement_2): i779-i786.

[19] LONG Y H, ZHANG Y, WU M, et al. Predicting drugs for COVID-19/SARS-CoV-2 via heterogeneous graph attention networks[C]// Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. Piscataway: IEEE Press, 2020: 455-459.

[20] SUN Y Z, ZHANG D H, CAI S B, et al. MDAD: a special resource for microbe-drug associations[J]. Frontiers in Cellular and Infection Microbiology, 2018, 8: 424.

[21] RAJPUT A, THAKUR A, SHARMA S, et al. aBiofilm: a resource of anti-biofilm agents and their potential implications in targeting antibiotic drug resistance[J]. Nucleic Acids Research, 2018, 46(D1): D894-D900.

[22] ANDERSEN P I, IANEVSKI A, LYSVAND H, et al. Discovery and development of safe-in-man broad-spectrum antiviral agents[J]. International Journal of Infectious Diseases, 2020, 93: 268-276.

[23] STEINBECK C, HOPPE C, KUHN S, et al. Recent developments of the chemistry development kit (CDK): an open-source Java library for chemo- and bioinformatics[J]. Current Pharmaceutical Design, 2006, 12(17): 2111-2120.

[24] WEININGER D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules[J]. Journal of Chemical Information and Computer Sciences, 1988, 28(1): 31-36.

[25] KAMNEVA O K. Genome composition and phylogeny of microbes predict their co-occurrence in the environment[J]. PLoS Computational Biology, 2017, 13(2): e1005366.

[26] CHEN X, HUANG Y, YOU Z H, et al. A novel approach based on KATZ measure to predict associations of human microbiota with non-infectious diseases[J]. Bioinformatics, 2017, 33(5): 733-739.

[27] CHEN X. KATZLDA: KATZ measure for the lncRNA-disease association prediction[J]. Scientific Reports, 2015, 5: 16840.

[28] DENG L, HUANG Y B, LIU X J, et al. Graph2MDA: a multi-modal variational graph embedding model for predicting microbe-drug associations[J]. Bioinformatics, 2022, 38(4): 1118-1125.

[29] JIANG H J, HUANG Y, YOU Z H. Predicting drug-disease associations via using Gaussian interaction profile and kernel-based autoencoder[J]. BioMed Research International, 2019, 2019: 2426958.

[30] YANG H P, DING Y J, TANG J J, et al. Inferring human microbe-drug associations via multiple kernel fusion on graph neural network[J]. Knowledge-Based Systems, 2022, 238: 107888.

[31] WANG C C, LI T H, HUANG L, et al. Prediction of potential miRNA-disease associations based on stacked autoencoder[J]. Briefings in Bioinformatics, 2022, 23(2): bbac021.

[32] LIU D Y, HUANG Y B, NIE W J, et al. SMALF: miRNA-disease associations prediction based on stacked autoencoder and XGBoost[J]. BMC Bioinformatics, 2021, 22(1): 219.

[33] WANG S D, LIN B Y, ZHANG Y Y, et al. SGAEMDA: predicting miRNA-disease associations based on stacked graph autoencoder [J]. Cells, 2022, 11(24): 3984.

[34] LI J, ZHANG S, LIU T, et al. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction[J]. Bioinformatics, 2020, 36(8): 2538-2546.

[35] LONG Y H, LUO J W. Association mining to identify microbe drug interactions based on heterogeneous network embedding representation[J]. IEEE Journal of Biomedical and Health Informatics, 2021,

25(1): 266-275.

[36] CHENG X L, QU J, SONG S B, et al. Neighborhood-based inference and restricted Boltzmann machine for microbe and drug associations prediction[J]. PeerJ, 2022, 10: e13848.

[37] WANG M L, CAO R Y, ZHANG L K, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro[J]. Cell Research, 2020, 30(3): 269-271.

[38] GHASEMNEJAD-BERENJI M, PASHAPOUR S. Favipiravir and COVID-19: a simplified summary[J]. Drug Research, 2021, 71(3): 166-170.

[39] THEOHARIDES T C, CHOLEVAS C, POLYZOIDIS K, et al. Long-COVID syndrome-associated brain fog and chemofog: Luteolin to the rescue[J]. BioFactors, 2021, 47(2): 232-241.

[40] CHOY K T, WONG A Y, KAEWPREEDEE P, et al. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro[J]. Antiviral Research, 2020, 178: 104786.

[41] DHAR J, SAMANTA J, KOCHHAR R. Corona virus disease-19 pandemic: the gastroenterologists' perspective[J]. Indian Journal of Gastroenterology, 2020, 39(3): 220-231.

Prediction of microbe-drug association based on graph attention stacked autoencoder

WANG Bo[1, 2, *], HE Yang[1], DU Xiaoxin[1, 2], ZHANG Jianfei[1, 2], XU Jingran[1], JIA Na[1]

(1. College of Computer and Control Engineering, Qiqihar University, Qiqihar 161006, China;
2. Heilongjiang Key Laboratory of Big Data Network Security Detection and Analysis, Qiqihar University, Qiqihar 161006, China)

Abstract: A graph attention stacking autoencoder approach for predicting the association between microorganisms and drugs, known as GATSAE, is proposed in response to the conventional method of finding new associations between microorganisms and drugs, which is primarily accomplished through biological experiments, which is highly costly and time-consuming. Firstly, establish a heterogeneous network of microorganisms and drugs to enrich the associated information. Secondly, the convolutional fusion matrix of microorganisms and drugs is obtained by extracting multi-layer latent features through graph convolutional network (GCN). Once again, an improved stacked autoencoder is used to learn unsupervised low dimensional representations of meaningful high-order similar features. Graph convolution and attention mechanisms are added to the stacked autoencoder to further optimize the extraction of high-order feature information. Finally, the low-dimensional features are concatenated with associated features, and a multi-layer perceptron (MLP) is used to score and predict the final microbial drug. According to performance evaluation, GATSAE subjects' area under the receiver operating characteristic curve (AUROC) and area under the precision-recall curve (AUPR) were 0.961 9 and 0.957 7, respectively. These results are better than those of popular deep learning techniques and traditional machine learning techniques. Case studies have shown that GATSAE can accurately predict candidate drugs related to SARS-CoV-2 and Escherichia coli, as well as candidate microorganisms related to aspirin.

Keywords: microbes and drugs; correlation prediction; stacked autoencoder; attention mechanism; graph convolutional network; multi-layer perceptron

Received: 2023-11-07; **Accepted**: 2024-01-12; **Published Online**: 2024-03-09 07:24

URL: link.cnki.net/urlid/11.2625.V.20240306.0934.001

Foundation item: General Program for National Natural Science Foundation Cultivation under the Basic Research Fund for Provincial Universities in Heilongjiang (145409324)

* **Corresponding author**. E-mail: bowangdr@qqhru.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0752

基于多通道耦合的时空增强异常行为检测

章东平[1,*]，潘鑫[1]，马道滨[1]，米红妹[1]，林丽莉[2]
(1. 中国计量大学 信息与工程学院，杭州 310018；2. 浙江工商大学 信息与电子工程学院，杭州 310018)

摘 要：针对现有异常行为检测模型在特征提取方面存在局限、对动态时序特征建模不足的问题，提出一种基于多通道耦合的时空增强异常行为检测方法。该方法以SlowFast网络为基础，在慢路径中引入多通道耦合的空间增强模块以强化静态特征建模，在快路径中引入多通道耦合的时间增强模块以提升动态时序特征的判别能力。在Violent Flow、Hockey Fight和Real-life Violence Situations 3个基准数据集上的实验表明，所提方法的预测准确率分别达到95.3%、97.3%和94%，优于现有主流方法。结果验证了所提方法在异常行为检测任务中具有更强的特征表达能力和泛化性能。

关 键 词：SlowFast；时空增强；异常行为检测；多通道耦合；注意力机制
中图分类号：V328.3; TP181
文献标志码：A **文章编号**：1001-5965(2026)01-0073-07

检测多种形式多媒体内容(如视频、图像和音频)中的异常行为是当今世界的一项关键任务。随着互联网上该类内容的日益普及，开发能够自动检测和理解异常行为的智能系统变得至关重要。异常行为检测对于公共安全、内容审核、执法及积极的预防措施具有重要意义。

在深度学习领域，视频中的异常行为检测方法大致分为2类：基于特征提取分类器的方法和基于深度学习神经网络的方法。基于特征提取分类器的方法通常使用基于特征提取的分类器，该分类器依赖手工制作的特征，直接从帧中提取，如颜色直方图、纹理描述符或运动矢量，以检测异常行为[1-6]，该方法通常难以捕捉复杂和微妙的暴力行为，检测准确度和稳健性有限。基于深度学习神经网络的方法在各种计算机视觉应用中表现出卓越的性能。最初，研究人员采用由2D卷积神经网络(convolutional neural network, CNN)和长短时记忆(long short-term memory, LSTM)网络组成的联合架构，用于视频中的时空特征学习和序列建模，以进行视频异常检测任务[7-9]。2D CNN从单独的帧中提取视觉特征，而LSTM捕捉帧之间的时间上下文和动态。随后，研究人员提出一种围绕3D CNN的新型网络架构，将标准的2D卷积扩展到时间维度[10-13]。基于3D CNN的网络从视频剪辑中联合学习时空特征并有效建模运动信息。最近，基于Transformer的网络在视频分析任务中引起了相当大的关注[14-18]。基于自注意机制，Transformer在建模视频中的远距离依赖性方面表现出强大的能力。基于深度学习的神经网络在各种研究中的显著性能巩固了其作为视频中异常行为检测的主导技术的地位。

SlowFast网络[19]的出现解决了运动和细粒度特征提取不足、模型资源消耗大、模型泛化能力弱的缺点，其是为解决视频中时空信息不平衡而设计的双分支网络。Slow分支处理较慢的时间信息，主要捕获背景和全局上下文，以学习场景的整体结构；Fast分支处理更快的时间信息，主要捕获运动物体和细节，提取更多的局部和细粒度特征，以学习运动物体的详细信息。通过将Slow分支和Fast

收稿日期：2023-11-20；录用日期：2023-12-01；网络出版时间：2024-02-05 12:29
网络出版地址：link.cnki.net/urlid/11.2625.V.20240204.1401.001
基金项目：浙江省重点研发计划 (2023C01030,2022C01082,2021C03192)
*通信作者. E-mail：06a0303103@cjlu.edu.cn

引用格式：章东平，潘鑫，马道滨，等. 基于多通道耦合的时空增强异常行为检测[J]. 北京航空航天大学学报，2026，52(1)：73-79.
ZHANG D P，PAN X，MA D B，et al. Spatial-temporal enhanced abnormal behavior detection based on multi-channel coupling[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：73-79 (in Chinese).

分支相结合，SlowFast网络可以充分利用视频中的时空信息，提高视频理解性能。

为进一步提高SlowFast网络的性能，本文在慢速和快速通路中分别引入了多通道耦合的空间增强模块和时间增强模块，以提取更多丰富信息的时空信息。这些模块有助于网络在处理视频数据时准确地关注重要的时空特征。

本文提出了一种基于多通道耦合的时空增强异常行为检测方法，可有效解决视频中静态与动态信息判别不准确的问题，主要工作如下：

1）提出了多通道耦合的时间增强模块，通过融合通道信息和时间信息来捕捉视频中人类行为的动态，能够自适应地对不同时间间隔的动态特征进行加权，从而更好地捕捉视频中的运动信息。

2）提出多通道耦合的空间增强模块，将多通道信息与空间信息融合，增强网络在空间维度上对静态特征的建模能力，从而提高异常行为检测的准确性。

3）提出一种基于多通道耦合的时空增强异常行为检测方法，基于SlowFast网络，在慢速和快速路径中分别引入多通道耦合的空间增强模块和时间增强模块，以提取更多丰富信息的时空信息。在多个异常行为数据集上的实验结果表明，提出的基于多通道耦合的时空增强异常行为检测方法在异常行为检测任务中的表现优于对比方法。

1 本文方法

SlowFast是一种双路径网络结构，其中，慢路径以低帧率处理视频中的空间语义信息，快路径以高帧率处理视频中的运动信息。如图1所示，在通道维度上，慢路径中的通道数记为C，而快路径中的通道数记为βC，其中，β=1/8。该设计可以使Fast通路以更高的速度运行，同时使Slow通路能够捕获更大量的空间信息；在时间维度上，慢路径的时间维度为T，而快路径的时间维度为αT（α>1为时间尺度因子），从而增强快路径捕捉运动信息的能力，本文实验中设α=4；在空间维度上，慢路径和快路径的特征图具有相同的高度H和宽度W，以支持跨路径的特征融合。

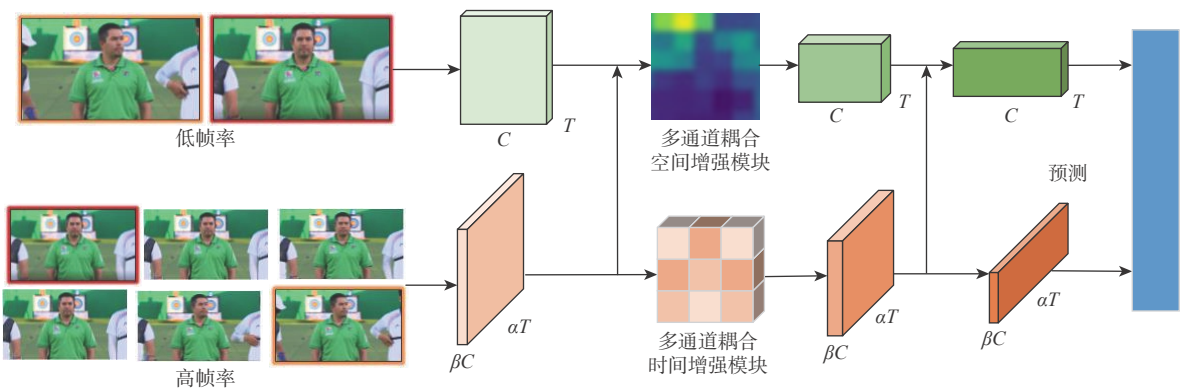


图1 多通道耦合的时空增强异常行为检测网络结构

Fig. 1 Structure of multi-channel coupled spatio-temporal enhanced abnormal behavior detection network

为融合来自双路径的特征信息，SlowFast利用多个横向连接。在融合过程中，来自Fast通路的特征形状$\{\alpha T, H\times W, \beta C\}$被转化为$\{T, H\times W, \alpha\beta C\}$。将融合后的信息送入分类器进行多标签分类预测。

1.1 多通道耦合时空增强模块

多通道耦合时空增强模块由2部分组成：位于慢路径的多通道耦合空间增强模块和位于快路径的多通道耦合时间增强模块。

多通道耦合空间增强模块结构如图2所示，其由通道注意模块和空间增强模块串联构成。

多通道耦合空间增强模块的输入特征为Slow路径中间层的输出，是一个5维特征映射，本文将其称为空间输入特征图$F_1\in\mathbf{R}^{C\times T\times H\times W}$。将空间输入特征图$F_1$通过多通道耦合空间增强模块来捕获通道和空间特征信息，表示为

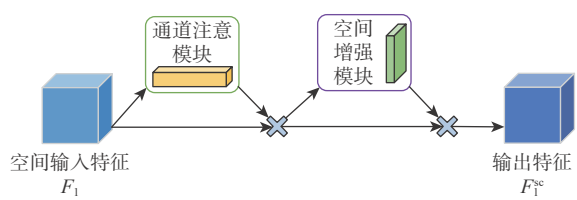


图2 多通道耦合空间增强模块

Fig. 2 Multi-channel coupled spatial enhancement module

$$F_1^{c}=M_c(F_1)\otimes F_1 \tag{1}$$

$$F_1^{sc}=M_s(F_1^{c})\otimes F_1^{c} \tag{2}$$

式中：$M_c\in\mathbf{R}^{C\times1\times1\times1}$为通道特征权值；$M_s\in\mathbf{R}^{1\times1\times H\times W}$为空间特征权值；⊗表示逐元素相乘。这2次乘法产生$F_1^{sc}$作为最终的输出特征图。

多通道耦合时间增强模块的结构如图3所示，其由通道注意模块和时间增强模块串联构成。

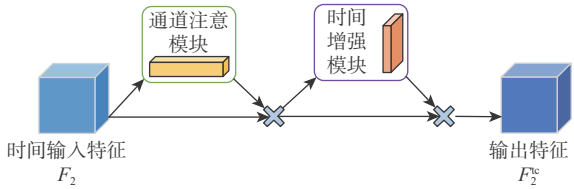


图 3　多通道耦合时间增强模块

Fig. 3　Multi-channel coupled temporal enhancement module

多通道耦合时间增强模块的输入特征为快速路径中间层的输出, 是一个 5 维特征映射, 本文将其称为时间输入特征图 $F_2 \in \mathbf{R}^{C\times T\times H\times W}$。将时间输入特征图 F_2 通过多通道耦合时间增强模块来捕获通道和时间特征信息, 表示为

$$F_2^{c} = M_c(F_2)\otimes F_2 \tag{3}$$

$$F_2^{tc} = M_t(F_2^{c})\otimes F_2^{c} \tag{4}$$

式中: $M_t \in \mathbf{R}^{1\times T\times 1\times 1}$ 为时间特征权值。2 次乘法得到 F_2^{tc} 作为最终输出的特征。

1.2　通道注意模块

通道注意模块通过计算输入特征映射中每个通道的重要性, 并根据任务自适应耦合多通道特征的权重, 从而提高特征表示能力。通道注意模块的结构如图 4 所示。

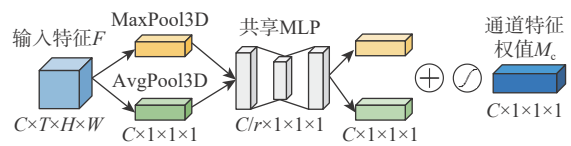


图 4　通道注意模块

Fig. 4　Channel attention module

首先, 通过三维全局平均池化(AvgPool3D)和三维全局最大池化(MaxPool3D)对输入特征 $F \in \mathbf{R}^{C\times T\times H\times W}$ 的时空维度进行压缩, 得到 2 张 $C\times 1\times 1\times 1$ 特征图。然后, 将其输入多层感知器(multilayer perceptron, MLP)网络, 其中, 通道维度被 $1/r$ (缩减比) 压缩, 扩展到原始维度, 以允许通道维度上输入数据的重要性加权。通过 MLP 网络后, 对得到的 2 个特征图进行逐元素相加, 再通过 Sigmoid 激活函数得到通道特征权值 $M_c \in \mathbf{R}^{C\times 1\times 1\times 1}$, 其计算式如下:

$$\begin{cases} M_c(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool3D}(F)) + \\ \qquad \mathrm{MLP}(\mathrm{MaxPool3D}(F))) \\ \mathrm{MLP}(F) = W_1(W_0(F)) \end{cases} \tag{5}$$

式中: σ 表示 Sigmoid 激活函数; MLP 表示多层感知器网络; $W_1 \in \mathbf{R}^{C\times C/r}$ 和 $W_0 \in \mathbf{R}^{C/r\times C}$ 为 MLP 网络的权重系数, W_0 后应用 ReLU 激活函数。

最后, 将得到的通道特征权值 M_c 与输入特征映射 F 逐元素相乘, 生成输出特征 F^c。计算式如下:

$$F^{c} = F\otimes M_c \tag{6}$$

1.3　空间增强模块

空间增强模块通过选择性地关注输入特征图中的特定空间区域, 更好地捕获空间结构信息。空间增强模块如图 5 所示。

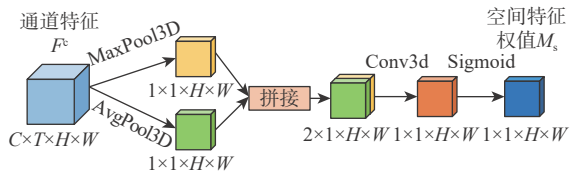


图 5　空间增强模块

Fig. 5　Spatial enhancement module

首先, 将通道注意模块的输出特征 F^c 作为空间注意力的输入特征, 在其上进行基于通道的三维全局最大池化和三维全局平均池化操作, 得到 2 个特征 F_{avg} 和 $F_{\max}$, 沿着通道维度进行串联。然后, 通过卷积模块对通道进行降维。为减少参数数量, 采用伪 3D 卷积, 用 $1\times 7\times 7$ 卷积和 $7\times 1\times 1$ 卷积的串联结构代替原来的 $7\times 7\times 7$ 卷积。通过 Sigmoid 激活函数传递降维特征, 得到空间注意特征 $M_s \in \mathbf{R}^{1\times T\times H\times W}$, 其计算式如下:

$$M_s(F^c) = \sigma(f^{1\times 7\times 7} * f^{7\times 1\times 1}([\mathrm{AvgPool3D}(F^c); \\ \mathrm{MaxPool3D}(F^c)])) \tag{7}$$

式中: $f^{1\times 7\times 7}$ 表示 $1\times 7\times 7$ 卷积; $f^{7\times 1\times 1}$ 表示 $7\times 1\times 1$ 卷积; "$*$" 表示卷积操作。

最后, 将得到的空间特征权值 M_s 与输入特征映射 F^c 逐元素相乘, 生成输出特征 F^{sc}, 计算式如下:

$$F^{sc} = F^{c}\otimes M_s \tag{8}$$

1.4　时间增强模块

时间增强模块可以看作是一种时间选择的动态机制, 通过学习时间间隔之间的相关性和重要性, 可以更好地捕获和理解时间序列数据中的关键信息。时间增强模块如图 6 所示。

首先, 将通道注意模块的输出特征 F^c 作为时间增强模块的输入特征, 输入特征 $F^c \in \mathbf{R}^{C\times T\times H\times W}$ 的通道维度和空间维度分别经过三维全局平均池化和三维全局最大池化, 得到 2 个 $1\times T\times 1\times 1$ 特征图。然后, 将其输入到一个由 2 个卷积层组成的卷积块中, 卷积核大小为 $1\times 1\times 1$, 并采用 ReLU 激活函数, 其中, 时间维度通过 $1/r$ 的降维比率压缩, 再扩展回原始维度。经过卷积块处理后, 获得的 2 个特征图进行逐元素相加, 通过 Sigmoid 激活函数, 得到时间特征权值 $M_t \in \mathbf{R}^{1\times T\times 1\times 1}$, 其计算式如下:

$$\begin{cases} M_t(F^c) = \sigma(\mathrm{MLP}(\mathrm{AvgPool3D}(F^c)) + \\ \qquad \mathrm{MLP}(\mathrm{MaxPool3D}(F^c))) \\ \mathrm{MLP}(F) = \mathrm{Conv3d}(\mathrm{ReLU}(\mathrm{Conv3d}(F))) \end{cases} \tag{9}$$

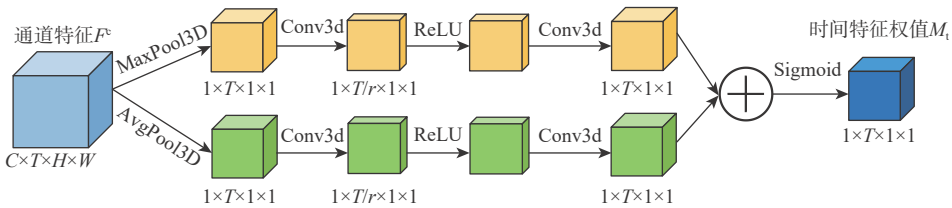


图 6 时间增强模块

Fig. 6 Temporal enhancement module

最后，获得的时间特征权值M_t与输入特征图F^c进行逐元素相乘，生成输出特征F^{tc}，计算式如下：

$$F^{tc}=F^c\otimes M_t \tag{10}$$

2 实验与分析

2.1 数据集

选择 Violent Flow (VF) [2]、Hockey Fight (HF) [20] 和 Real-Life Violence Situations (RLVS) [7] 数据集来评估本文方法，数据集相关信息如表 1 所示。

表 1 实验数据集信息

Table 1 Experimental datasets information

数据集	高度/像素	宽度/像素	暴力视频数	总视频数	帧数
VF	240	320	123	246	50～150
HF	288	360	500	1 000	50
RLVS	224～1 080	224～1 920	1 000	2 000	60～210

2.2 数据增强

数据增强过程如图 7 所示。在该示例中，使用高度为 240 像素、宽度为 320 像素的帧图像，并将其在宽度和高度上分别放大 1.25 倍。

采用随机数据增强方法。由于这些转换依赖于随机值，其在每个时期的每个批次中的每个视频样本中都会被刷新。这种随机增强引入了变化，并增强了训练过程的鲁棒性，因为模型在每个时期观察到同一视频数据的不同视图，扩大了训练数据集的有效大小，而不需要额外的视频样本。数据增强的具体步骤如下：

1）随机短边缩放。以输入尺寸$H\times W$调整视频帧图像的大小，具体操作为：选择1.2H和1.4H之间的随机值作为新的高度，同时保持原始纵横比，按比例增加宽度。在该过程中，使用插值填充图像，得到放大的帧图像。

2）随机裁剪。在随机裁剪过程中，从缩放后的视频帧中随机裁剪出尺寸为 256 像素×256 像素的部分，裁剪的随机位置确保了训练数据的多样性，有助于模型从帧的不同区域学习。

3）随机镜像翻转。以一定的概率水平翻转每个视频帧，意味着一些帧将被水平镜像，而其他帧保持原样。

2.3 训练细节

实验基于 PyTorch 库在一台装有 GeForce RTX 2080 Ti GPU 的计算机上进行。模型从头开始训练，没有使用任何预训练模型。随机从完整视频中抽样一个片段（包含 32 帧），并使用随机数据增强技术将帧图像统一调整为 256 像素×256 像素的尺寸。将这 32 帧作为快通道的输入，并均匀采样 8 帧作为慢通道的输入。在训练阶段，不使用数据增强操作。

2.4 多通道耦合时空增强模型的性能

结果如表 2 所示，ViF 和 OViF 为基于特征提取

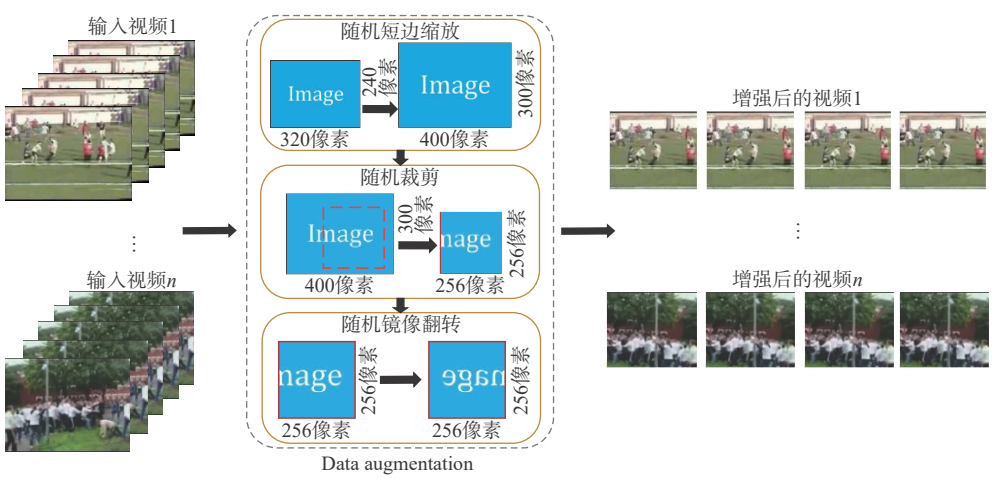


图 7 数据增强流程

Fig. 7 Data augmentation workflow

表 2　与其他方法的结果比较

Table 2　Comparison of results with other algorithms

方法	准确率/%		
	VF	HF	RLVS
ViF[2]	82	83	
OViF[5]	78	87	
CLSTM[9]	94.5	97	
Temporal Fusion CNN+LSTM[21]			91
VGG16+LSTM[7]	90	95.1	88.2
SlowFast	89.1	93.6	91.1
STCA-SlowFast	95.3	97.3	94

的方法, 准确性较低。相比之下, 基于深度学习的方法性能更好。结合 2D CNN 和 LSTM 的方法, 如 CLSTM[9]、Temporal Fusion CNN+LSTM[21]、VGG16+LSTM[7], 在准确性方面表现出显著的改进。本文模型 STCA-SlowFast 不仅优于原 SlowFast 模型, 还在 3 个数据集上获得了最高的预测准确率, 分别为 95.3%、97.3% 和 94%, 表明本文方法在暴力行为检测准确性方面优于现有的基于特征提取的方法[2,5]和基于深度学习的方法[7,9,21]。

为进一步评估模型的性能, 在 VF、HF 和 RLVS 数据集上分别绘制了本文模型与 SlowFast 模型的 ROC 曲线, 并计算了相应的 AUC 值。从图 8～图 10 中可以看出, 与原 SlowFast 模型相比, 本文模型表现出更优的检测效果, 所有 AUC 值均超过 0.98。实验比较结果如图 11 所示。

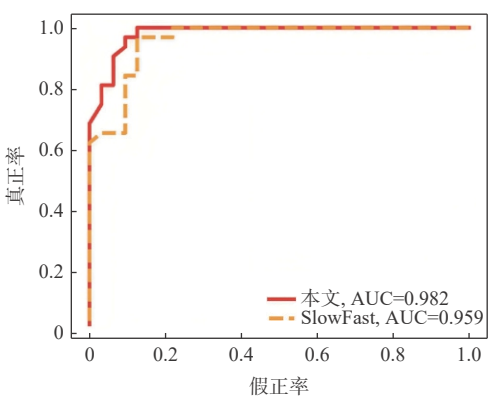


图 8　在 VF 数据集上 ROC 曲线及 AUC 值

Fig. 8　ROC Curve and AUC Value on VF dataset

2.5　消融实验

为确定如何最有效地将多通道耦合时间增强模块和多通道耦合空间增强模块集成到 SlowFast 中, 本文对时空增强模块的位置进行了多组实验。实验结果如表 3 所示, 将时空增强模块添加到 SlowFast 网络的第 3 层效果最好, 因为可以实现更均衡的信息传递, 从而在不同层级上更全面地捕捉特征关系。

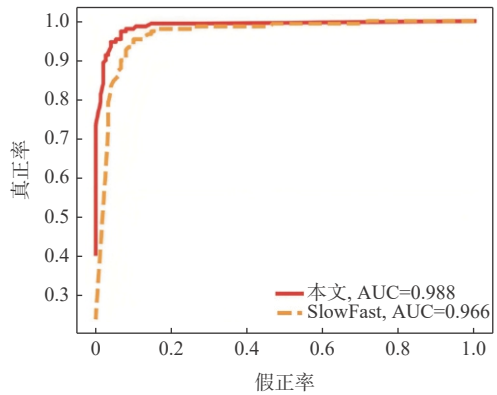


图 9　在 HF 数据集上 ROC 曲线及 AUC 值

Fig. 9　ROC Curve and AUC Value on HF dataset

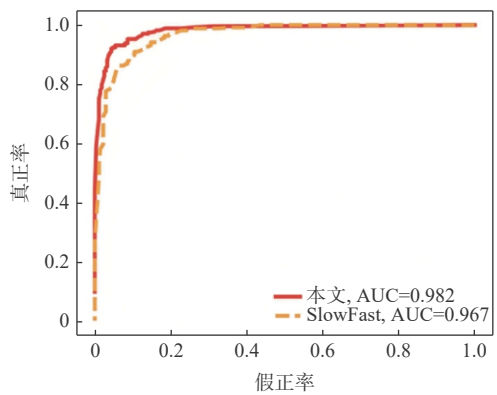


图 10　在 RLVS 数据集上 ROC 曲线及 AUC 值

Fig. 10　ROC Curve and AUC Value on RLVS dataset

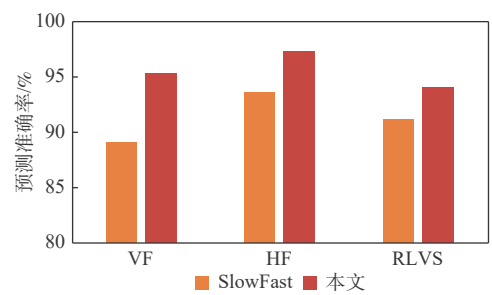


图 11　在不同数据集上的预测准确率

Fig. 11　Prediction accuracy on different datasets

同时, 还进行了多组实验以探究多通道耦合时空增强模块中每个子模块的有效性。所有模块均添加到模型的第 3 层。实验结果如表 4 所示。可

表 3　在不同位置添加多通道耦合时空增强模块的结果比较

Table 3　Comparison of results with addition of multi-channel coupled spatio-temporal enhancement modules at different positions

添加位置	准确率/%		
	VF	HF	RLVS
第2层	93.8	96.3	92.8
第3层	95.3	97.3	94
第4层	95.3	97	93.5
第5层	93.8	96.7	93.2

表4 不同子模块的实验结果比较

Table 4 Comparison of results from different sub-modules

时间增强	空间增强	准确率/%		
		VF	HF	RLVS
		89.1	93.6	91.1
√		93.8	96.7	93.4
	√	92.2	96.3	93.2
√	√	95.3	97.3	94

以看到，本文提出的时空增强模块在提高准确性方面是有效的。

3 结 论

1）提出的基于多通道耦合的时空增强异常行为检测方法在 Violent Flow、Hockey Fight 和 Real-life Violence Situations 这 3 个基准数据集上分别实现 95.3%、97.3% 和 94% 的准确率，AUC 均超过 0.98，显著优于原始 SlowFast 及多种主流异常检测模型。

2）多通道耦合时间增强模块有效提升了模型对快速动态行为的时序建模能力，消融实验显示单独引入时间增强模块后，模型在 VF、HF 和 RLVS 上的准确率分别提升了 4.7%、3.1% 和 2.3%；而空间增强模块则分别带来 3.1%、2.7% 和 2.1% 的提升。当两者联合部署后，模型准确率进一步提升至 95.3%、97.3% 和 94%，相较于基线 SlowFast 模型分别提高了 6.2%、3.7% 和 2.9%，显著优于任一单一模块的增益，表明时空增强具有强互补性与协同效应。

未来工作将从以下方面展开：①探索无监督或弱监督学习范式，降低对大规模精细标注视频的依赖；②将模型扩展至更广泛的异常行为类别（如跌倒、聚集、遗留物等），提升方法的通用性；③结合轻量化网络设计或知识蒸馏技术，优化计算效率，推动其在边缘设备和实时监控系统中的实际应用。

参考文献（References）

[1] CHEN L H, HSU H W, WANG L Y, et al. Violence detection in movies[C]//Proceedings of the8th International Conference Computer Graphics, Imaging and Visualization. Piscataway: IEEE Press, 2011: 119-124.

[2] HASSNER T, ITCHER Y, KLIPER-GROSS O. Violent Flows: Real-time detection of violent crowd behavior[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2012: 1-6.

[3] LI W X, MAHADEVAN V, VASCONCELOS N. Anomaly detection and localization in crowded scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(1): 18-32.

[4] COLQUE R V H M, CAETANO C, DE ANDRADE M T L, et al. Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(3): 673-682.

[5] GAO Y, LIU H, SUN X H, et al. Violence detection using oriented VIolent Flows[J]. Image and Vision Computing, 2016, 48: 37-41.

[6] BILINSKI P, BREMOND F. Human violence recognition and detection in surveillance videos[C]//Proceedings of the 13th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE Press, 2016: 30-36.

[7] SOLIMAN M M, KAMAL M H, EL-MASSIH NASHED M A, et al. Violence recognition from videos using deep learning techniques[C]//Proceedings of the 9th International Conference on Intelligent Computing and Information Systems. Piscataway: IEEE Press, 2019: 80-85.

[8] MEDEL J R, SAVAKIS A. Anomaly detection in video using predictive convolutional long short-term memory networks[EB/OL]. (2016-11-15)[2023-11-01]. https://arxiv.org/abs/1612.00390?context=cs.

[9] SUDHAKARAN S, LANZ O. Learning to detect violent videos using convolutional long short-term memory[C]//Proceedings of the 14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE Press, 2017: 1-6.

[10] LV Y, DUAN Y, KANG W, et al. Human activity recognition based on deep belief network and 3D convolutional neural network [J]. Multimedia Tools and Applications, 2018, 77(11): 12787-12801.

[11] MAQSOOD R, BAJWA U I, SALEEM G, et al. Anomaly recognition from surveillance videos using 3D convolution neural network [J]. Multimedia Tools and Applications, 2021, 80(12): 18693-18716.

[12] DUBEY S, BORAGULE A, JEON M. 3D ResNet with ranking loss function for abnormal activity detection in videos[C]//Proceedings of the International Conference on Control, Automation and Information Sciences. Piscataway: IEEE Press, 2019: 1-6.

[13] TAN Z H, XIA Z C, WANG P F, et al. FTCF: full temporal cross fusion network for violence detection in videos[J]. Applied Intelligence, 2023, 53(4): 4218-4230.

[14] ABDALI A R, AGGAR A A. DEVTrV2: enhanced data-efficient video Transformer for violence detection[C]//Proceedings of the 7th International Conference on Image, Vision and Computing. Piscataway: IEEE Press, 2022: 69-74.

[15] ZHOU L. End-to-end video violence detection with Transformer[C]//Proceedings of the 5th International Conference on Pattern Recognition and Artificial Intelligence. Piscataway: IEEE Press, 2022: 880-884.

[16] LI S, LIU F, JIAO L C. Self-training multi-sequence learning with Transformer for weakly supervised video anomaly detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2): 1395-1403.

[17] PURWANTO D, CHEN Y T, FANG W H. Dance with self-attention: a new look of conditional random fields on anomaly detection in videos[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 173-183.

[18] AHMED W, YOUSAF M H. A deep autoencoder-based approach for suspicious action recognition in surveillance videos[J]. Arabian Journal for Science and Engineering, 2024, 49(3): 3517-3532.

[19] FEICHTENHOFER C, FAN H Q, MALIK J, et al. SlowFast networks for video recognition[C]//Proceedings of the IEEE/CVF In-

ternational Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6201-6210.

[20] NIEVAS E B, SUAREZ O D, GARCÍA G B, et al. Violence detection in video using computer vision techniques[M]//REAL P, DIAZ-PERNIL D, MOLINA-ABRIL H, et al. Computer analysis of images and patterns. Berlin: Springer, 2011: 332-339.

[21] LIMA J P D, FIGUEIREDO C M S. Temporal fusion approach for video classification with convolutional and LSTM neural networks applied to violence detection[J]. Inteligencia Artificial, 2021, 24(67): 40-50.

Spatial-temporal enhanced abnormal behavior detection based on multi-channel coupling

ZHANG Dongping[1,*], PAN Xin[1], MA Daobin[1], MI Hongmei[1], LIN Lili[2]

(1. College of Information Engineering, China Jiliang University, Hangzhou 310018, China;
2. School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China)

Abstract: To address the limitations of existing abnormal behavior detection models—particularly their inadequate feature representation and insufficient modeling of dynamic temporal characteristics—this paper proposes a multi-channel coupled spatio-temporal enhanced anomaly detection method. Built upon the SlowFast network architecture, the proposed approach integrates a multi-channel coupled spatial enhancement module into the slow pathway to strengthen static feature modeling, and a multi-channel coupled temporal enhancement module into the fast pathway to improve the discriminability of dynamic temporal features. Extensive experiments on three benchmark datasets—Violent Flow, Hockey Fight, and Real-life Violence Situations—demonstrate that the proposed method achieves prediction accuracies of 95.3%, 97.3%, and 94%, respectively, outperforming current state-of-the-art approaches. The results validate the superior feature representation capability and generalization performance of the proposed method in abnormal behavior recognition tasks.

Keywords: SlowFast; spatio-temporal enhanced; abnormal behavior detection; multi-channel coupling; attention mechanism

Received: 2023-11-20; **Accepted**: 2023-12-01; **Published Online**: 2024-02-05 12:29

URL: link.cnki.net/urlid/11.2625.V.20240204.1401.001

Foundation items: Zhejiang Key R & D Project of China (2023C01030,2022C01082,2021C03192)

* **Corresponding author**. E-mail: 06a0303103@cjlu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0767

基于邻居边界度的自主决策系统性能边界识别方法

路辉[1,*]，吕静茹[2]，王诗琪[1]，孙泽斌[3]

(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 中国移动通信集团北京有限公司，北京 100007；
3. 北京航空航天大学 可靠性与系统工程学院，北京 100191)

摘 要：性能边界是度量自主决策系统鲁棒性的重要表征，可以反映自主决策系统对抗扰动的能力。针对性能边界数据多空间分布、增量生成等特点，提出一种基于邻居边界度的自主决策系统性能边界识别方法。面对性能边界搜索空间复杂、全空间尺度不统一的难点，设计邻居边界度指标解决绝对尺度度量问题，并提出基于邻居边界度的性能边界识别流程；考虑利用增量数据结合原有识别结果进一步精确刻画性能边界，提出基于邻居边界度的增量性能边界识别方法，实现对增量数据的高效处理；为解决增量过程中的近邻搜索和反向近邻搜索的效率问题，提出改进局部敏感哈希的近似近邻搜索优化方法；分别以标准测试函数、路径规划系统作为典型的自主决策系统，开展理论研究工作的验证和分析。实验结果表明：基于邻居边界度的性能边界识别方法具有很好的方法参数泛化能力，在路径规划系统上，该方法比对比方法识别准确度高出13.68%，运行时间减少91.57%。

关 键 词：自主决策系统；鲁棒性测试；性能边界识别；增量识别；局部敏感哈希
中图分类号：V221$^+$.92; TP274$^+$.5
文献标志码：A **文章编号**：1001-5965(2026)01-0080-14

随着人工智能技术的发展，自主决策系统在工业、农业等各个领域得到广泛应用。自主决策指在没有人为参与的情况下，系统主动收集信息并进行数据分析与策略决策[1]。自主决策系统是用来实现自主决策功能的系统，能够综合外界和自身的各种情况进行分析决策。例如，路径规划系统、图像识别系统、目标检测系统等。由于人工智能算法具有一定的脆弱性，使得自主决策系统的实际运行性能面临潜在的结果不可控等鲁棒性问题。例如，在对输入图像施加微小扰动，基于深度神经网络的图像分类模型会将大熊猫识别成长臂猿。输入的微小“噪声”，很可能就是致使自主决策系统失效的“幕后黑手”。因此，鲁棒性测试成为保障自主决策系统可靠性、安全性的重要环节。尤其在自动驾驶、医疗卫生、空中交通管理等对可靠性与安全性要求十分严苛的领域。

自主决策系统鲁棒性指在有噪声存在的条件下，系统是否可以保持决策正确性的能力。学者将自主决策系统决策边界称为性能边界区域[2]。位于自主决策边界的测试用例，即存在微小扰动引起决策性能模式阶跃变化的一类测试用例，被称为性能边界样本。生成和筛选这类性能边界样本是一项重要的鲁棒性测试手段。

自主决策系统的测试数据中含有很多边界样本和非边界样本，性能边界样本往往比其他样本具有更高的数据价值。了解系统的性能边界样本，能够认识系统行为决策发生转变的重要因素和特征。从大批量、冗余的测试数据中，寻找自主决策

收稿日期：2023-11-24；**录用日期**：2024-06-07；**网络出版时间**：2024-07-08 11:48
网络出版地址：link.cnki.net/urlid/11.2625.V.20240705.1512.004
基金项目：国家自然科学基金 (62371030)
*通信作者. E-mail：mluhui@buaa.edu.cn

引用格式：路辉，吕静茹，王诗琪，等. 基于邻居边界度的自主决策系统性能边界识别方法 [J]. 北京航空航天大学学报，2026，52（1）：80-93. LU H，LYU J R，WANG S Q，et al. Performance boundary identification method for autonomous decision-making systems based on neighbor boundary degree [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：80-93 （in Chinese）.

系统的性能边界样本, 能够有助于提升自主决策系统的鲁棒性。

相关研究中, 最为广泛的是对抗样本生成方法。Szegedy 等发现对于原始样本添加微小扰动后会造成深度神经网络识别错误[3]。Goodfellow 等对这一现象进行分析和解释[4], 并提出“对抗样本”的概念和对抗样本生成方法。近年来, 研究学者根据模型决策偏差和鲁棒性提出各类对抗样本生成方法, 应用于知识蒸馏等更广泛的研究领域[5-7]。

此外, 许多研究人员对自主决策系统测试数据进行过滤挑选性能边界样本。Pan 等为减少支持向量机在处理大数据集时的时间和空间成本, 提出选择训练集的子集进行学习[8], 为搜索位于数据分布边界附近的样本, Pan 等提出在固定邻域范围内, 分类内部样本比边界样本具有更多邻居, 基于此, 可以对边界样本进行筛选。Mullins 等提出的对抗性规划工具(range adversarial planning tool, RAPT) 是一种新的实验设计过程[2]。其利用软件在环仿真和自适应采样方法为任何自主决策系统生成测试场景, 可以帮助理解决策过程。Mullins 等将各性能模式最近邻居距离小于边界距离阈值的样本识别为边界样本[2]。通常情况下, 良好的邻域范围或边界阈值的设定需要大量的实验尝试或测试场景的先验知识。在实际应用中, 测试场景先验知识难以获得, 大批量实验带来高昂计算成本, 该方法存在一定的局限性。

Zykin 等提出通过多维空间中的一组固定参数确定区域边界的算法, 该算法能够确定诊断对象的临界状态、确定接近最佳状态的度量等问题[9]。Cao 等为提高结构优化算法的计算效率, 提出了一种边界识别方法来识别搜索空间的可行区域边界[10]。该边界识别方法包括虚拟采样技术、改进拉丁超立方采样方法和支持向量机分类器。在医学界, 性能边界识别方法也有广泛研究, Zhou 等为解决超声图像中内膜与中层边界模糊问题, 提出一种端到端边界显著性多分支网络, 使用并行线性结构学习模块和边界细化模块来利用先验形状知识和解剖依赖性[11]。

同时, 考虑到自主决策系统测试过程连续产生的测试数据具有增量生成特点, 在原有性能边界识别结果的基础上, 增量数据可能会更新性能边界或探索新空间的性能边界。然而, 简单地将增量数据和原有数据结合, 重新进行性能边界识别会耗费更多计算时间和计算成本。因此, 为充分利用增量数据信息, 在保证识别精度的基础上提升识别效率, 需要研究更适用于增量数据的性能边界识别方法。

增量学习是指能够不断处理现实世界中连续的信息流, 在吸收新知识的同时保留甚至整合、优化旧知识的学习方式[12-15]。在聚类研究中, 增量聚类算法由 Ester 等在基于密度的噪声应用空间聚类(density-based spatial clustering of applications with noise, DBSCAN) 的增量聚类算法中提出[16], 他们提出了用于更新的种子对象的概念, 确定了新增数据点的影响范围和更新聚类的核心对象。处理大型增量数据集的一种方式是按顺序处理分块后的小型子集中的数据, 按顺序处理一次所有数据块即可将所有的数据对象聚类在一起, 如单通增量聚类算法[17]。

在自主决策系统的性能边界识别研究领域, 目前主要为面向静态数据的非增量处理方法, 对于增量数据的性能边界识别尚没有有效的处理方法。由于增量学习问题的复杂性, 通常仅讨论特定设置下的增量学习。因此, 现有的增量学习方法和性能边界识别方法不适用于增量数据的性能边界识别过程。

综上, 现有的性能边界识别方法大多依赖于测试数据的距离和边界分布等信息。虽然能够定义性能边界, 但当测试系统模型未知时, 先验知识很难获得, 难以设置合理的绝对尺度阈值, 导致性能边界识别过程效率和精度降低。为此, 本文采用邻居边界度识别指标, 结合性能模式内分簇流程, 提出基于邻居边界度(neighborhood boundary degree, NBD)的性能边界识别方法。

1　性能边界识别框架

为提升性能边界识别方法在面对多个测试问题时的综合表现能力、提升性能边界识别的精度和效率, 解决方法参数对被测系统先验知识的依赖、多分布和边界宽度不一致带来的识别能力差等问题, 本文面向搜索空间复杂、全空间尺度不统一的性能边界识别问题, 研究基于邻居边界度的性能边界识别方法框架, 具体思路如图 1 所示。在框架中设计邻居边界度识别指标解决绝对尺度度量问题, 并基于邻居边界度提出结合性能模式内分簇的性能边界识别流程; 为解决增量数据性能边界识别效率低的问题, 提出基于邻居边界度的增量性能边界识别方法。在该过程中, 为进一步优化近邻识别的效率, 提出改进局部敏感哈希的近似近邻方法, 搜索被测样本在测试样本集合中的近邻集合; 最后, 给出自主决策系统性能边界识别结果。

在性能边界识别算法的验证阶段, 设计标准测

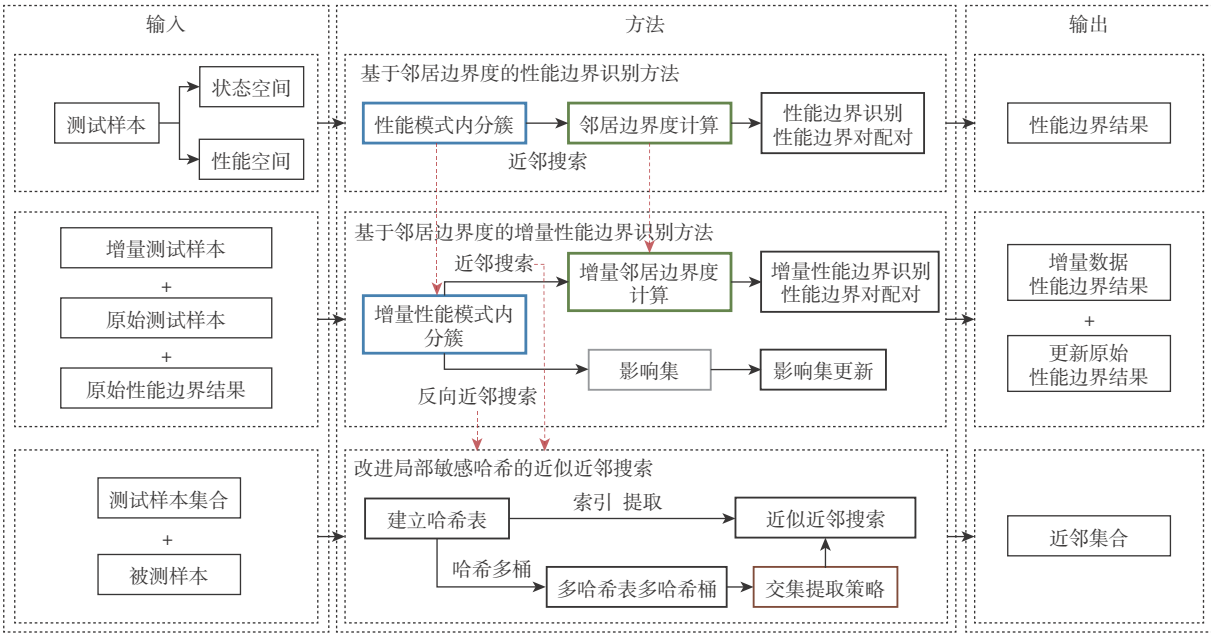


图1 基于邻居边界度的性能边界识别方法框图

Fig. 1 Framework for performance boundary identification method based on neighbor boundary degree

试函数、路径规划系统等自主决策系统，测试性能边界识别方法针对不同类型、不同范围的测试问题的性能边界识别能力。本文基于邻居边界度的性能边界识别方法与其他方法相比对方法参数的敏感性更低，并且具有更高的识别精度和识别效率，本文增量性能边界识别方法能够在保证识别精度的前提下实现提升识别效率的需求。

2 性能边界定义

对于自主决策系统的性能边界有如下定义。

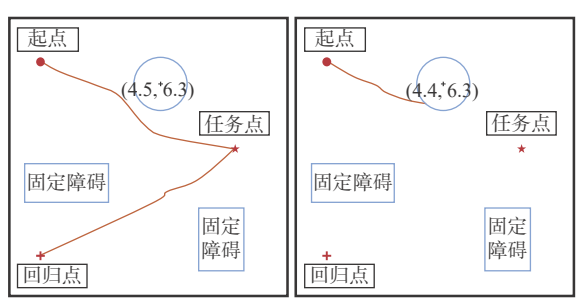
1）状态空间。状态空间 X^n 包含 n 个元素 $\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n\}$。状态空间向量中的每个元素代表被测系统场景配置中的一个变量。状态空间是一个连续度量空间，其中，样本之间的距离表示了在被测系统场景配置中的相似性。

2）性能空间。性能空间 Y^m，包含 m 个元素 $\{\boldsymbol{Y}_1,\boldsymbol{Y}_2,\cdots,\boldsymbol{Y}_m\}$。性能空间向量中的每个元素代表被测系统自主决策的性能度量结果。性能空间是一个离散度量空间，性能空间中的离散取值可以称为性能模式 P，代表被测系统的不连续性能决策类型。

3）性能边界区域。性能模式 P_{a} 和 P_{b} 间的性能边界区域 $B_{\mathrm{a,b}}$，定义为 $\forall \boldsymbol{X}_{i,\mathrm{a}}\in B_{\mathrm{a,b}},\exists \boldsymbol{X}_{j,\mathrm{a}}\in B_{\mathrm{a,b}},\mathrm{s.t.}\left|\boldsymbol{X}_{i,\mathrm{a}}-\boldsymbol{X}_{j,\mathrm{b}}\right|<D_{\varepsilon}$，其中，$D_{\varepsilon}$ 为性能边界区域的宽度。

4）性能边界样本对。性能边界样本对 $\boldsymbol{b}_{i,j}=\left[\boldsymbol{X}_i,\boldsymbol{X}_i,\boldsymbol{Y}_i,\boldsymbol{Y}_j\right]\in B_{\mathrm{a,b}}$ 是由2个测试样本组成的向量，其中，$\left|\boldsymbol{X}_i-\boldsymbol{X}_j\right|=D_{ij}<D_{\varepsilon}$，$\boldsymbol{X}_i,\boldsymbol{X}_j\in X^n$，$\boldsymbol{Y}_i\in P_{\mathrm{a}},\boldsymbol{Y}_j\in P_{\mathrm{b}}|\mathrm{a}\neq\mathrm{b}$。

以路径规划系统为例，系统输入包含起点、任务点、回归点和障碍物信息。其输出为离散化的路径规划决策结果，被称为性能模式。障碍物信息的变化，可能导致路径发生障碍物碰撞、规划时间过长等情况，将这类路径规划决策结果归为规划失败性能模式。因此，路径规划系统决策结果被离散化为规划成功、规划失效2类性能模式。例如，图2中的测试对象为圆形障碍物。测试任务为固定大小的圆形障碍在二维路径规划空间不同位置下的路径规划结果。圆形障碍的二维位置坐标被称为状态空间，路径规划决策结果为性能空间。在测试样本中，存在一些如图3所示的样本对，在状态空间发生微小变化，在性能空间发生阶跃变化。这类样本便是性能边界对，性能边界对分布接近性能决策边界。搜索路径规划系统各场景中的性能边界对，即生成引起路径规划系统决策结果突变的规划场景，能够帮助发现路径规划算法的失效区域。



(a) 路径规划成功场景 (b) 路径规划失败场景

图2 路径规划系统性能边界对示例

Fig. 2 Examples of path planning system performance for boundary pairs

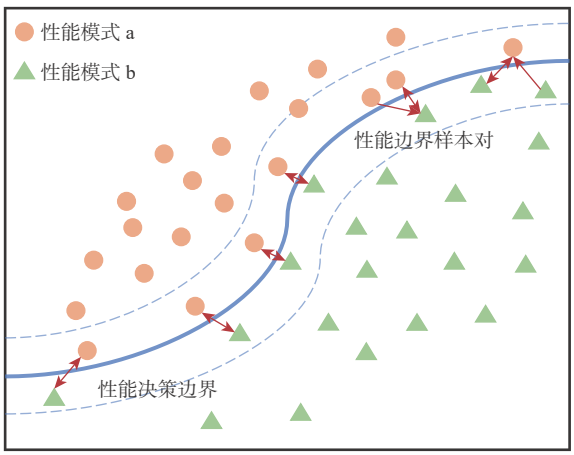


图 3　路径规划系统性能边界示意图

Fig. 3　Performance boundary diagram of path planning system

3　基于邻居边界度的性能边界识别方法

自主决策系统的测试样本具有多空间分布的特点, 即便是单一性能模式下的测试样本也具有多类簇的分布。多分布测试样本难以使用一组绝对距离、密度等尺度标准进行全局衡量, 性能边界识别的标准阈值设定成为难题。目前的性能边界识别方法没有充分考虑性能模式内测试样本分簇分布的特点, 通常采用距离这类绝对尺度信息作为识别标准。因此, 性能边界识别方法的识别准确性较低、方法参数与具体问题具有较大的关联性, 从而导致泛化性差的问题。

为解决上述问题, 本文提出基于邻居边界度的性能边界识别方法, 整体识别过程如图 4 所示。

首先, 设计性能模式内分簇方法, 挖掘性能模式内部测试样本的分布特征以精细化性能边界搜索。在这一过程中, 需要重点考虑性能模式内测试样本分布多样性的问题。分簇数量未知和分簇形状多样的特性, 要求性能模式内分簇过程自适应确定分簇数量并能够突破传统聚类方法只能发现类圆形分簇的问题。借鉴基于密度的聚类方法, 对所有测试样本涵盖的性能模式, 提取单一性能模式下所有测试样本进行分类簇识别。分类簇识别过程中, 以邻域内样本数量表示密度, 以密度连通为聚类标准, 实现通过密度判别的性能模式内分簇过程。

其次, 提出邻居边界度识别指标。一般情况下, 靠近性能边界的样本周围会存在更多的异类样本, 而靠近分布内部远离性能边界的样本周围则存在更多的同类样本。这一理论把测试样本的性能边界识别转为测试样本和其近邻范围内的样本状态, 能够有效避免使用绝对尺度作为识别标准的问题。

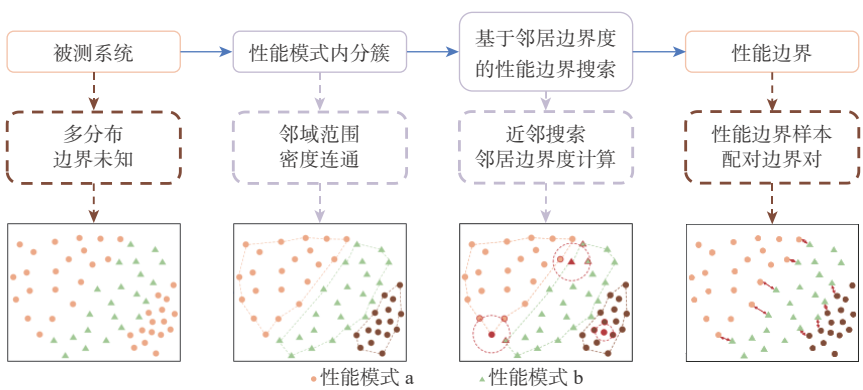


图 4　基于邻居边界度的性能边界识别方法示意图

Fig. 4　Diagram of neighbor boundary degree-based performance boundary identification method

对自主决策系统测试样本 $D=\{\boldsymbol{X},\boldsymbol{Y}\}$, 基于邻居边界度的性能边界识别流程具体步骤如下。

1) 性能模式内分簇。对每个性能模式内测试样本进行分簇, 生成测试样本 D 的一维分簇空间 $\boldsymbol{C}$。

2) 近邻搜索。对于测试样本 D 中每一样本 $\boldsymbol{X}_i$, 搜索其在测试样本 D 中的前 k 个最近邻样本集合 $N_i=\{\boldsymbol{X}_{n1},\boldsymbol{X}_{n2},\cdots,\boldsymbol{X}_{nk}\}$。

3) 邻居边界度计算。对于测试样本 D 中每个样本, 根据其近邻样本集合按照下式计算邻居边界度:

$$B_i=\frac{\left|N_i^{\text{bound}}\right|}{|N_i|}\qquad N_i^{\text{bound}}=\{\boldsymbol{X}_{ni}|\boldsymbol{X}_{ni}\in N_i,C_{ni}\neq C_i\}\tag{1}$$

4) 性能边界识别和性能边界对配对。对测试样本 $\boldsymbol{X}_i$ 的邻居边界度 B_i 进行判断, 若样本 $\boldsymbol{X}_i$ 的邻居边界度 B_i 大于等于邻居边界度阈值 q_B, 则将样本 $\boldsymbol{X}_i$ 标记为性能边界样本, 并从 N_t 中选取距离样本 $\boldsymbol{X}_i$ 最近异簇样本配对为性能边界样本对。

4 基于邻居边界度的增量性能边界识别方法

面对增量数据时，在充分利用增量数据进行性能边界的更新基础上，为进一步提升处理效率，开展对增量性能边界识别方法的研究。同时，邻居边界度指标只与测试样本的邻居分布有关，增量数据添入带来的影响是局部且已知的。因此，在此基础上提出基于邻居边界度的增量性能边界识别方法。

为实现对增量数据的新空间性能边界探索，并利用增量数据对已有性能边界精细刻画更新，设计了增量数据的性能边界识别和增量数据影响集的识别及更新，2个步骤结合形成完整基于邻居边界度的增量性能边界识别方法流程。

4.1 增量数据的性能边界识别

基于邻居边界度的增量性能边界识别方法主要从增量性能模式内分簇和增量邻居边界度计算、识别这2个角度进行分析，具体如图5所示。首先，对增量数据设计以基于密度的增量聚类思路实现增量性能模式分簇过程；需要在同样的分簇邻域范围和密度要求条件下，获取增量数据的性能模式内分簇归属，并以增量数据为中心更新原性能模式内分簇结果。其次，对增量数据进行邻居边界度计算和增量性能边界识别；邻居边界度指标的计算只与被识别样本和其在测试样本中的近邻集合相关，具备一定的局部特征。因此，对增量数据进行性能边界识别时只需要考虑其自身邻居边界度的计算。通过近邻搜索获取增量数据的近邻集合，再计算邻居边界度并进行性能边界识别和性能边界对配对。

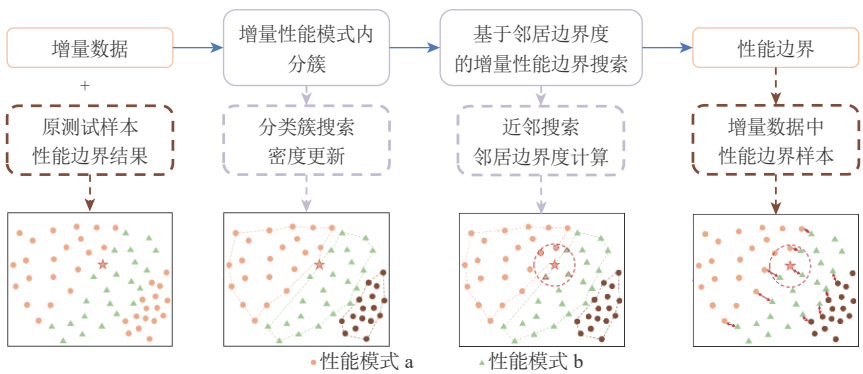


图5 增量数据的性能边界识别示意图

Fig. 5 Illustration of identifying performance boundaries for incremental data

对于原测试样本集 D 和增量数据集 D_{incre}，遍历 D_{incre} 中每个增量测试样本 $\boldsymbol{X}_{\text{incre}}$，搜索其在原测试样本集 D 中的 k 近邻集合 N_{incre}；依据近邻集合 N_{incre} 计算样本 $\boldsymbol{X}_{\text{incre}}$ 的邻居边界度，结合阈值 q_B 进行性能边界状态判断；如满足性能边界识别阈值标准被识别为性能边界样本，则在近邻集合 N_{incre} 中选取样本 $\boldsymbol{X}_{\text{incre}}$ 的最近邻异类样本进行性能边界样本对配对。

4.2 增量数据影响集的识别及更新

增量数据在添入测试数据时，将性能边界状态受到影响的测试样本集合成为影响集。根据性能边界识别方法的局部分布特征，添入增量数据对原测试数据结果的影响范围是有限的。利用这一特点，可以通过预处理方式缩小更新范围，减少增量更新时计算浪费情况。

根据提出的邻居边界度计算特点，被测样本的性能边界状态与其近邻样本集合有直接关系。由此可知，增量数据的添入只对其造成了近邻集合影响的样本带来了性能边界状态的影响。因此，近邻集合中包含增量数据的样本集合即增量数据的反向近邻集合为可能受到性能边界状态影响的集合范围。

对增量数据的反向近邻集合进一步分析，反向近邻集合内的样本与增量数据对比可以分为4种类别：同一性能模式的非边界样本、同一性能模式的边界样本、不同性能模式的非边界样本和不同性能模式的边界样本。根据邻居边界度的定义，非边界样本的邻居边界度指标不会因为同一性能模式的近邻数据添入而升高，其性能边界状态不会受到影响；边界样本的邻居边界度指标可能会因为同一性能模式的近邻数据添入而降低，不满足性能边界识别阈值标准，但考虑性能边界的多样性，保留其为性能边界样本。对增量数据的反向近邻集合根据性能模式分类进行再次精细筛选，生成影响集。对于筛选缩减后的影响集，利用其在增量数据添入后数据集内的近邻集合进行邻居边界度更新计算，再以性能边界识别标准判断是否为性能边界样本，

即可完成更新。

4.3　增量过程的近邻搜索优化

在本文提出的邻居边界度计算和增量过程的影响集识别中, 近邻搜索是必不可少的步骤。近邻搜索在增量过程中被频繁使用, 成为整个过程中较为耗时的步骤, 通过提升近邻搜索的效率来提升整体增量处理效率。局部敏感哈希(locality sensitive Hashing, LSH)利用哈希冲突建立数据近邻的快速检索, 是一种有效的近似近邻搜索方法[18-20]。为提升方法效率, 根据本文性能边界识别方法的特点, 设计改进局部敏感哈希技术进行近似近邻搜索优化。

在原始 LSH 中, 建立哈希表、提取哈希桶内数据是很快速的过程; 耗时超过一半的步骤是对提取出的哈希桶内数据全部进行距离计算并排序筛选。结合本文性能边界识别方法的近邻样本处理特点, 求取样本近邻集合是为进行局部样本分布特征的量化。因此, 性能边界识别方法需求的近邻集合可以省略距离计算排序的步骤, 很大程度上提升近邻搜索的效率, 但同时对哈希分桶的准确性提出更高要求。

不同距离度量下的哈希映射函数决定了哈希映射的概率分布。针对不同的问题, 可以通过多桶策略对哈希映射概率进行放缩。一方面为降低哈希桶内相似度较小样本的比例, 另一方面为降低所有哈希同桶样本的数量, 本文设计多哈希表取交集的策略进行改进。

单个哈希表中两样本映射在同一桶内的概率为 p。哈希表的数量为 k。多哈希表取并集策略意味着只要一张哈希表中两样本在同一桶内则可被提取。其同桶概率可以写为 $\ddot{p}=1-(1-p)^k$。多哈希表取交集策略意味着在 k 张哈希表中两样本均在同一桶内。其同桶概率可以写为 $\ddot{p}_2=p^k$。由于省略了对同桶数据的精细筛选, 希望初步提取的同桶数据中正确性更高。对比图 6 中的概率曲线, 在同样的数据分布情况下, 取交集策略能够减少提取的哈希近邻数量。并且能够有效减少低映射概率样本被提取为哈希近邻的概率。

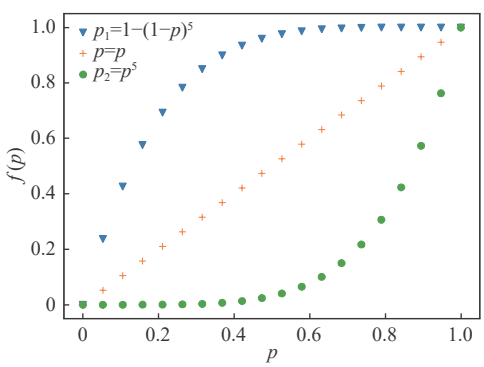


图 6　哈希同桶概率关系

Fig. 6　Relationship between same bucket probability and Hash probability

5　实验结果与分析

5.1　被测系统

为验证本文性能边界识别方法的有效性, 设计 2 类被测系统进行方法验证和对比分析。

1) 标准测试函数。为提高验证效率, 本文构造标准测试函数作为被测系统, 相比实际自主决策系统能够快速决策。为验证本文性能边界识别方法面对不同测试样本时的识别性能, 从真实性能边界比例、边界分布类型和问题范围等角度出发, 构造 6 种不同的标准测试函数。不同问题范围的标准测试函数能够验证本文性能边界识别方法在不同条件下参数调整的影响与识别性能。

2) 路径规划系统。为验证设计本文性能边界识别方法在实际自主决策系统中的应用, 选择基于人工势场算法的路径规划系统作为本文的被测系统。在路径规划系统中, 构造 4 个二维路径规划系统场景和 4 个三维路径规划系统场景, 具有不同的固定场景数量、信息和移动障碍信息; 设置规划场景内均包含一个移动障碍物。以上路径规划系统的性能边界状态空间分别为二维和三维, 较低维度的状态空间便于进行高效的验证实验。

对以上被测系统, 设计基于均匀采样的真实边界生成方法。首先, 采用蜂窝均匀采样获取均匀采样样本集合, 在该集合中筛选周围均匀采样点发生性能模式阶跃变化的样本为均匀采样中真实边界样本; 最后, 以蜂窝六边形边长为半径、每个均匀采样真实边界样本为圆心画圆, 能够包含在任意圆范围内的测试样本标记为采样样本中的真实边界样本。标准测试函数和路径规划系统的具体信息与样本分布情况如表 1 所示。

表 1　被测系统信息

Table 1　Tested system information

标准测试函数	函数状态空间范围	函数真实边界比例/%	路径规划系统	系统状态空间范围	系统真实边界比例/%
1	[−4,4]	26.20	1	[0,500]	41.84
2	[−10,10]	76.96	2	[0,500]	51.68
3	[−50,50]	22.21	3	[0,500]	48.05
4	[−500,500]	39.80	4	[0,500]	50.68
5	[−10,10]	58.71	5	[0,500]	15.35
6	[−10,10]	40.62	6	[0,500]	20.26
			7	[0,500]	20.54
			8	[0,500]	17.14

5.2 评价指标

为定量对比性能边界识别方法，构造合理的性能评价指标对各性能边界识别方法的识别结果进行量化表示。针对性能边界识别方法识别出的性能边界样本和真实性能边界样本研究准确性与覆盖度指标。准确性与覆盖度指标往往相互制约，为统一衡量两项指标，利用F-measure指标对准确性与覆盖度指标进行调和平均。为评价性能边界识别方法的识别效率，采用统一实验条件下处理同一数据的运行时间进行衡量。

将真实边界样本集合记为S_{T}，将测试数据中通过性能边界识别方法筛选的性能边界样本集合记为S_{I}。

以上评价指标的具体定义如下。

1）准确性

$$A=\frac{S_{\mathrm{T}}\cap S_{\mathrm{I}}}{S_{\mathrm{I}}} \tag{2}$$

准确性表示在识别出的性能边界样本中，真实性能边界样本所占比例。准确性指标越高，表明性能边界识别方法的识别越精确。

2）覆盖度

$$P=\frac{S_{\mathrm{T}}\cap S_{\mathrm{I}}}{S_{\mathrm{T}}} \tag{3}$$

覆盖度表示真实性能边界中被性能边界识别方法识别出的比例。覆盖度指标越高，表明性能边界识别方法对真实边界的识别覆盖性越好。

3）F-measure

$$F=\frac{2PA}{P+A} \tag{4}$$

F-measure指标用于对准确性和覆盖度的调和平均。F-measure指标越高，表明性能边界识别方法整体识别效果越好。

4）运行时间

在同样的实验环境下，对比不同性能边界识别方法对同一被测系统处理的运行时间，用于对比性能边界识别方法的识别效率；同样条件下运行时间越短说明识别方法效率越高。

从准确性、覆盖度、F-measure和运行时间4个指标对各方法的性能进行评估，并将实验结果与Mullins等提出基于边界阈值（boundary threshold criteria, BTC）标准的性能边界识别（以下简称BTC）方法[2]和Pan等提出的固定邻域范围（fixed neighborhood sphere, FNS）的性能边界识别（以下简称FNS）方法[8]进行对比，进一步分析验证本文性能边界识别方法的有效性。

在后续实验中，图的横坐标均代表被测系统（无量纲），纵坐标F代表F-measure的数值（无量纲）。

5.3 实验结果

5.3.1 基于邻居边界度的性能边界识别结果

在基于邻居边界度的性能边界识别方法研究过程中，考虑了性能边界识别方法的泛化性和性能边界识别效果。为验证并对比本文基于邻居边界度性能边界识别（以下简称NBD）方法的泛化性和识别性能，进行参数敏感性分析实验和性能边界识别方法性能对比分析实验。

1）参数敏感性分析实验结果

根据对性能边界识别方法的分析，每个方法中都具有直接影响性能边界识别标准的方法参数，而性能边界识别标准直接影响着性能边界识别的效果。NBD方法参数邻居数量k和邻居边界度阈值q_B，分别为正整数和0～1之间的有理数。BTC方法边界距离阈值参数D_ε，取值为正有理数。FNS方法参数固定邻域半径r和边界比例阈值q_r，取值为正有理数和0～1之间的有理数。

首先，在标准测试函数上进行少量预实验，分析性能边界识别方法与方法参数的大致关系。从算法设计角度分析，方法参数在不同被测系统中都具有大致相同的变化趋势。当方法参数k、D_ε、r、q_r较小和q_B较大时，意味着对于性能边界的搜索范围小、限制要求高，在指标计算时存在准确性较高但覆盖度较低的情况；反之，意味着对于性能边界的搜索范围大、限制要求低，存在覆盖度高但准确性较低的情况。根据以上理论分析，综合准确性和覆盖度2个指标的F-measure指标会随着方法参数的变化先增大后减小。

如图7所示，观察F-measure评价指标，可以清晰地看到，随着3个性能边界识别方法各项参数的增大，该指标的变化趋势均为先升高后下降，与理论分析的预期相符。因此，能够通过进一步细致的参数组合实验，搜索到该实验条件下的最优F-measure性能参数。

根据预实验的初步结果与分析，进行精细参数范围、更小步长的批量实验，以搜索性能边界识别方法的最优F-measure性能下的方法参数。

如图8所示，在标准测试函数4上，最优F-measure的参数k明显低于其他标准测试函数；在其他标准测试函数上，最优F-measure的参数k和q_B较为相似。分析表1中的结果，可以明显发现标准测试函数4中边界分布更为复杂。每个局部性能边界区域中测试样本较少，参数k过大则会明显扩大搜索空间为性能边界识别带来误差。如图9和图10所示，BTC方法和FNS方法最优F-measure表

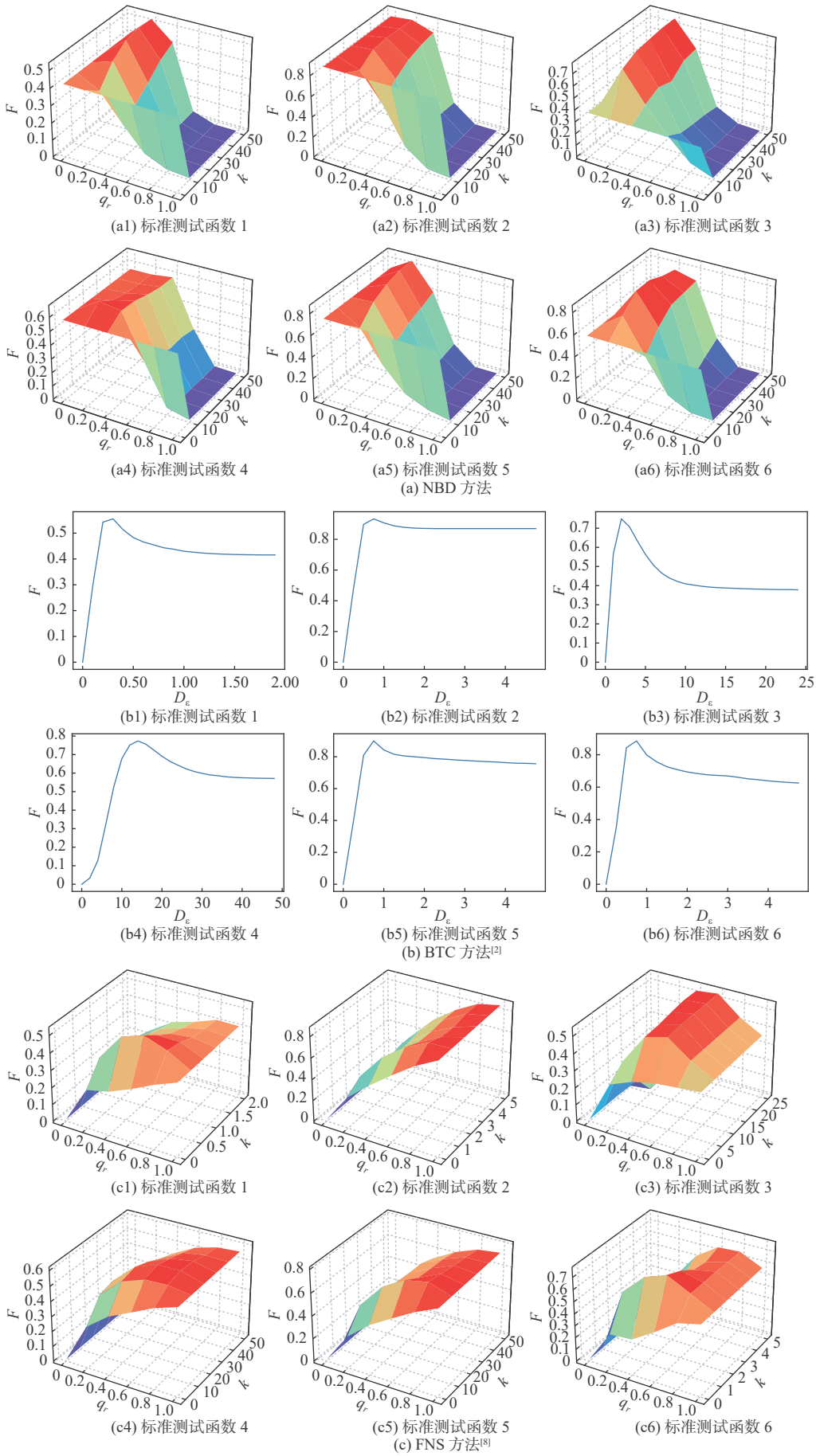
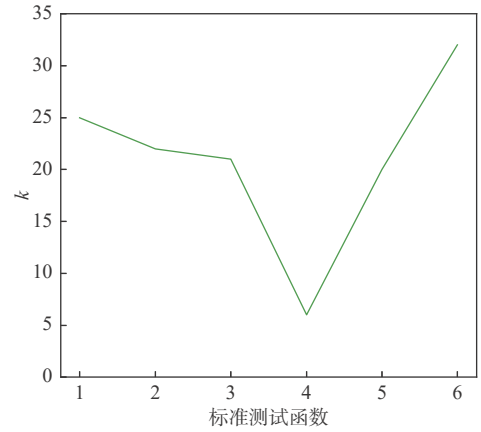
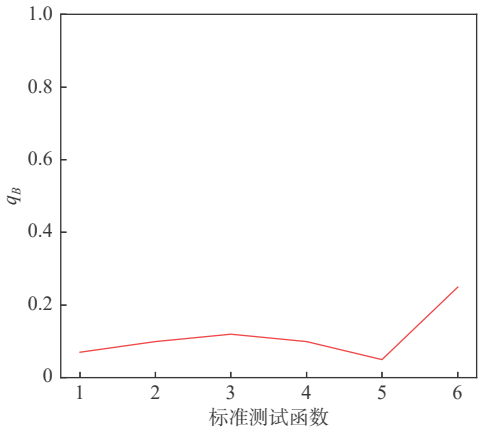


图 7　标准测试函数预实验结果

Fig. 7　Standard test functions preliminary test results



(a) NBD 方法最优性能下 k 取值



(b) NBD 方法最优性能下 q_B 取值

图 8 NBD 方法最优性能参数

Fig. 8 Optimal performance parameters of NBD method

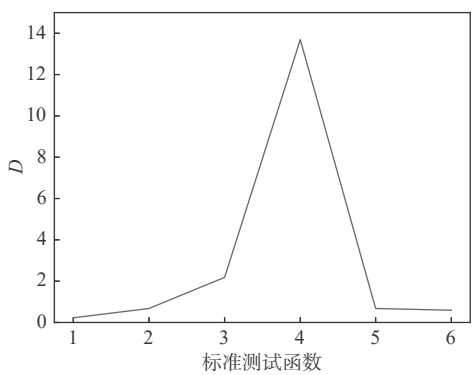


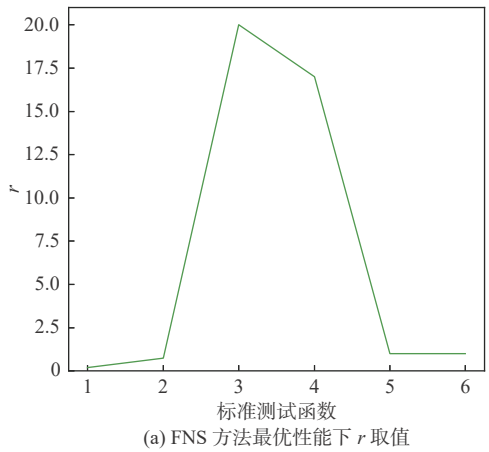
图 9 BTC 方法最优性能参数

Fig. 9 Optimal performance parameters of BTC method

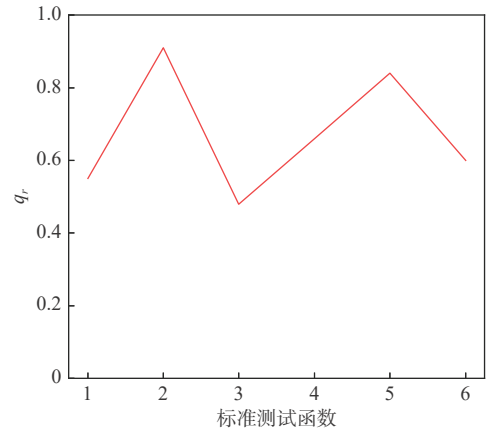
现下的方法参数在各标准测试函数间差别较大，没有明显的关联关系和规律性。

综上所述，NBD 方法的参数相比 BTC 方法和 FNS 方法更具有泛化性。为验证各性能边界识别方法参数的泛化能力，进一步开展方法参数验证。考虑到标准测试函数 4 的边界分布较为特殊，以其他 5 个标准测试函数的最优参数平均值开展实验。

实验结果对比如图 11 所示，使用最优 F-measure 的平均参数进行实验，NBD 方法在各标准测试函数上的 F-measure 相比最优 F-measure 平均减少约



(a) FNS 方法最优性能下 r 取值



(b) FNS 方法最优性能下 q_r 取值

图 10 FNS 方法最优性能参数

Fig. 10 Optimal performance parameters of FNS method

2.03%，BTC 方法相比平均减少 11.55%，FNS 方法相比平均减少 14.44%。由此项实验结果可以看出，NBD 方法的最优 F-measure 参数具有更好的泛化性，方法性能受参数影响较小。

根据以上方法参数敏感性分析的所有实验，可以说明 NBD 方法具有更优的方法参数泛化能力；对于不同的被测系统，NBD 方法能够更便捷、耗时更短的搜索到对应的最优方法参数。

2）性能边界识别方法性能对比分析实验结果

对于标准测试函数，在 5.3.1 节的参数敏感性分析实验的基础上，纵向对比各性能边界识别方法的最优 F-measure 和运行时间。

如图 12 所示，NBD 方法对比 BTC 方法在 6 个标准测试函数上的表现，F-measure 平均提升了 0.32%，运行时间平均减少了 93.51%；NBD 方法对比 FNS 方法，F-measure 平均提升了 16.07%，运行时间平均减少了 28.98%。这也说明了 NBD 方法的准确性和高效性。

将路径规划系统看作未知被测系统，使用各性能边界识别方法在标准测试函数上得到的平均最优参数进行实验，验证实验研究得出的性能边界识别方法平均最优参数的泛化性。同时，对路径规划

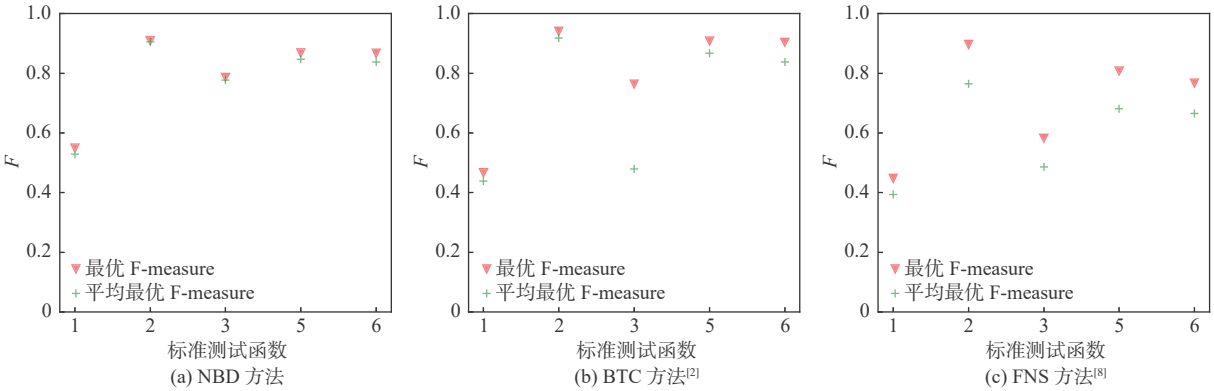


图 11　性能边界识别方法最优 F-measure 与平均最优 F-measure 对比

Fig. 11　Comparison between optimal F-measure and average parameter F-measure for performance boundary identification methods

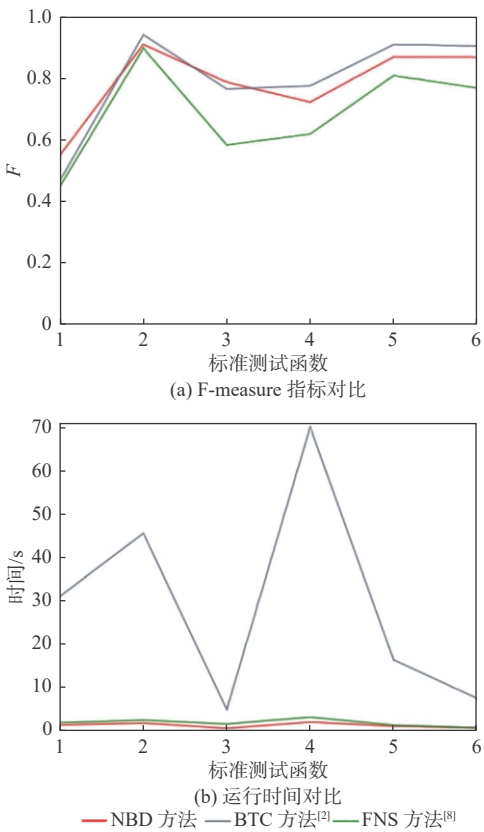


图 12　标准测试函数中各性能边界识别方法指标对比

Fig. 12　Comparison of performance boundary identification methods and indicators in standard test functions

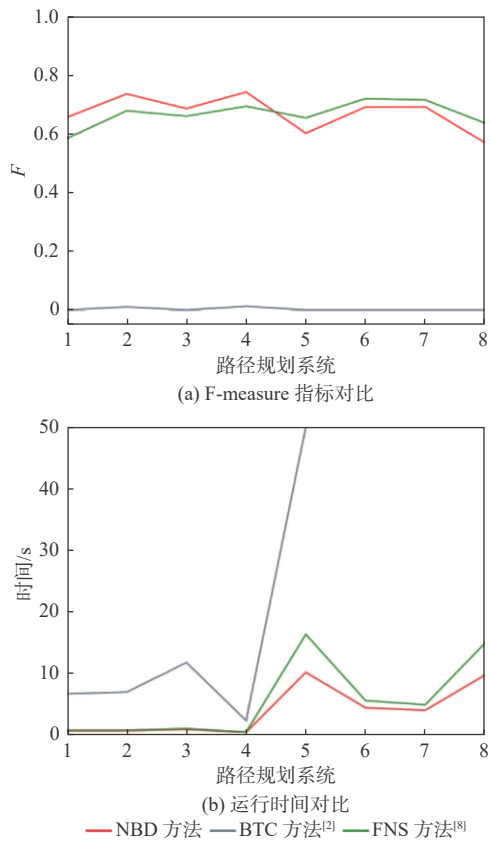


图 13　平均最优参数下路径规划系统中各性能边界识别方法指标对比

Fig. 13　Comparison of performance boundary identification methods and indicators in path planning systems under average optimal parameters

系统也进行参数敏感性实验搜索最优性能, 进一步对比各性能边界识别方法的识别性能。

如图 13 所示, 在平均最优参数下, NBD 和 FNS 方法仍具有有效的性能边界识别性能, BTC 方法则基本完全失效; 进一步说明 NBD 和 FNS 方法相比 BTC 方法具有更优的参数泛化性能。与 BTC 方法相比, NBD 方法的 F-measure 和运行效率均明显更优。与 FNS 方法相比, NBD 方法的 F-measure 比 FNS 方法高出约 0.75%, 在运行时间上平均减少了 18.47%。使用平均最优参数的情况下, NBD 方法面对未知的路径规划系统具有更优的性能边界识别效果。

同样, 采用预实验确定大致范围再批量实验搜索最优性能的实验思路, 针对路径规划系统搜索各种方法参数组合下的最优 F-measure 指标和对应的运行时间。如图 14 所示, NBD 方法相比 BTC 方法, F-measure 指标平均高出 13.68%, 运行时间平均减少 91.57%; NBD 方法相比 FNS 方法, F-measure 指标平均高出 4.49%, 运行时间平均减少 18.19%。在路径规划系统上的最优性能指标对比也可以说明 NBD 方法的边界识别性能。

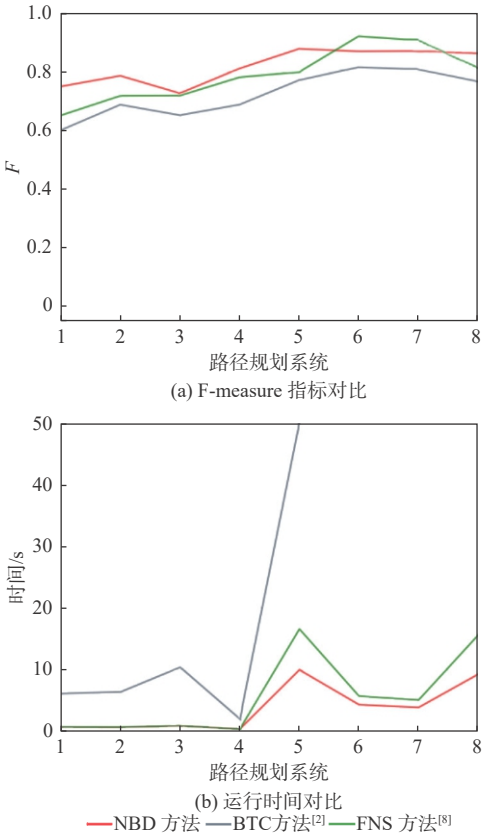


图 14 路径规划系统中性能边界识别方法最优 F-measure 指标对比

Fig. 14 Comparison of optimal F-measure for performance boundary identification methods in path planning systems

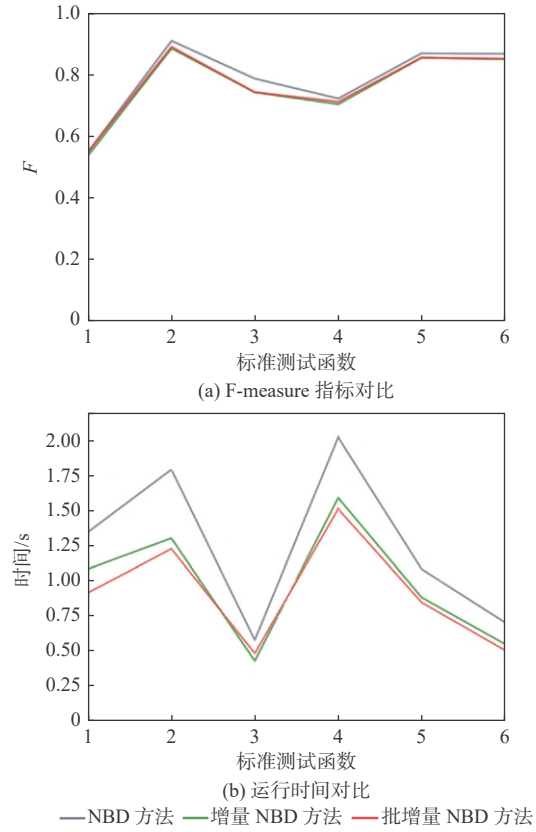


图 15 标准测试函数中增量性能边界识别方法指标对比

Fig. 15 Comparison of indicators for identifying incremental performance boundaries in standard test functions

5.3.2 基于邻居边界度的增量性能边界识别结果

1) 增量性能边界识别方法实验结果

在 5.3.1 节实验中获得的标准测试函数和路径规划系统最优 F-measure 性能参数下，进行增量性能边界识别方法实验，并与不进行增量的 NBD 方法进行评价指标对比分析。

实验结果如图 15 和图 16 所示，对具体实验结果数据分析，在标准测试函数上，增量 NBD 方法和 NBD 方法相比平均降低 F-measure 指标 2.78%，减少运行时间 22.44%。在路径规划系统上，增量 NBD 方法和 NBD 方法相比平均降低 F-measure 指标 3.69%，减少运行时间 30.67%。

2) 近邻搜索方法实验结果

在增量过程中，使用 NBD 方法、增量 LSH 和增量的改进 LSH 方法进行实验，结果如图 17 和图 18 所示。综合分析，增量 LSH 方法和 NBD 方法相比平均降低 F-measure 指标 1.45%，减少运行时间 24.91%；增量的改进 LSH 方法和 NBD 方法相比平均降低 F-measure 指标 0.29%，减少运行时间 35.95%。实验结果表明，改进 LSH 方法能够在较小程度影响增量性能边界识别效果的同时提升识别效率，能够有效应用于增量性能边界识别的近邻搜索优化方法。

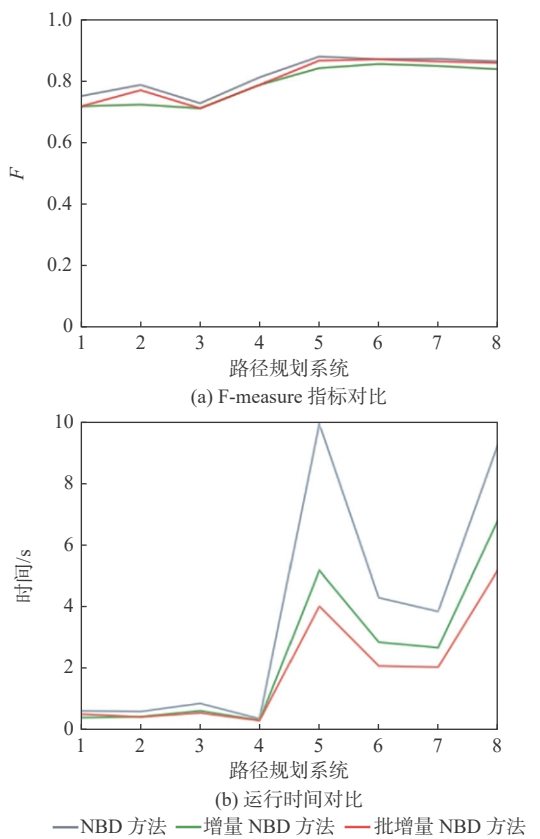


图 16 路径规划系统中增量性能边界识别方法指标对比

Fig. 16 Comparison of indicators for incremental performance boundary identification methods in path planning systems

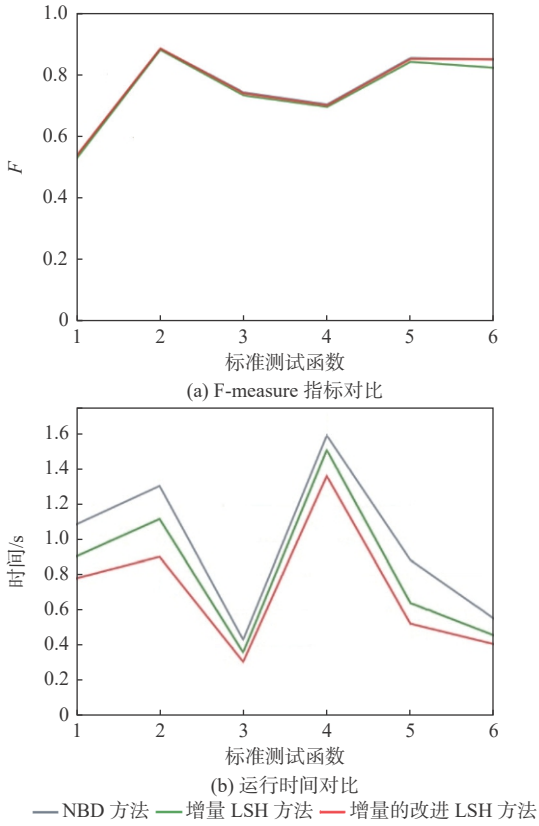


图 17　标准测试函数中近邻搜索优化方法评价指标对比

Fig. 17　Comparison of evaluation indicators for nearest neighbor search methods in standard test functions

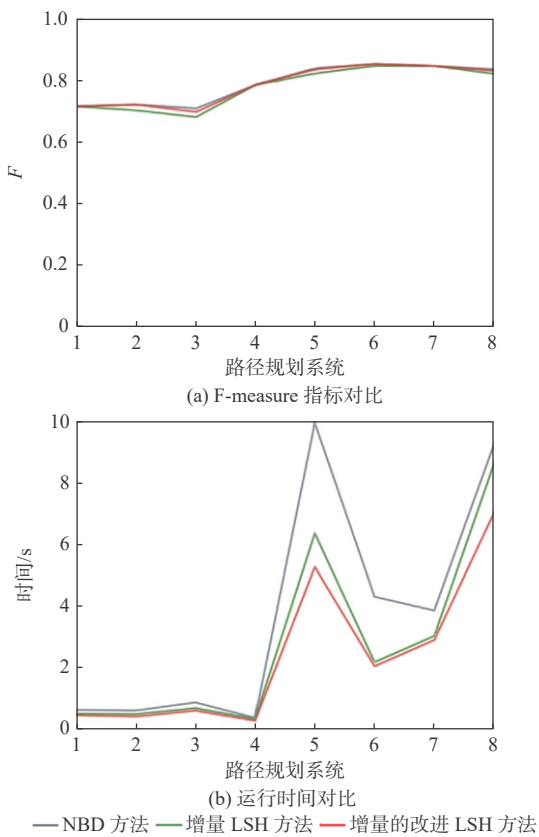


图 18　路径规划系统中近邻搜索优化方法评价指标对比

Fig. 18　Comparison of evaluation indicators for nearest neighbor search methods in path planning systems

6　结　论

1）针对自主决策系统测试数据多分布和未知导致识别标准难以设计的问题，采用邻居边界度识别指标，结合性能模式内分簇流程提出基于邻居边界度的性能边界识别方法，该方法能够提高性能边界识别的准确度、覆盖度和识别效率。

2）针对自主决策系统测试数据增量生成带来的性能边界更新、识别效率降低的问题，考虑在对增量数据识别性能边界样本的同时利用其更新原性能边界结果，提出基于邻居边界度的增量性能边界识别方法，保证识别效果的基础上提高对增量数据的识别效率。

3）为提升增量过程中近邻搜索和反向近邻搜索效率，采用局部敏感哈希的近似近邻搜索方法并对局部敏感哈希提取数据策略进行改进。

本文提出基于邻居边界度的性能边界识别方法和基于邻居边界度的增量性能边界识别方法，可以有效提升自主决策系统性能边界识别的识别性能。在后续的研究工作中将继续优化基于邻居边界度的性能边界识别方法，以解决需求更复杂的性能边界识别问题。

参考文献（References）

[1] HORN P. Autonomic computing: IBM's perspective on the state of information technology[EB/OL]. (2001-3-27)[2023-11-15]. http://www.ibm.com/research/autonomic.

[2] MULLINS G E, STANKIEWICZ P G, HAWTHORNE R C, et al. Adaptive generation of challenging scenarios for testing and evaluation of autonomous vehicles[J]. Journal of Systems and Software, 2018, 137: 197-215.

[3] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[EB/OL]. (2013-12-21)[2023-11-15]. https://arxiv.org/abs/1312.6199v4.

[4] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[EB/OL]. (2015-7-16)[2023-11-15]. http://arxiv.org/abs/1412.6572.

[5] SONG Y, KIM T, NOWOZIN S, et al. PixelDefend: leveraging generative models to understand and defend against adversarial examples[EB/OL].(2018-5-23)[2023-11-15].http://arxiv.org/abs/1710.10766.

[6] ILYAS A, SANTURKAR S, TSIPRAS D, et al. Adversarial examples are not bugs, they are features[J]. Advances in Neural Information Processing Systems, 2019, 32: 132-138.

[7] HEO B, LEE M, YUN S, et al. Knowledge distillation with adversarial samples supporting decision boundary[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 3771-3778.

[8] PAN D Y, YIN Y J, SUN Y J, et al. Sample selection in support vector machines: a fixed neighborhood sphere approach[C]//Pro-

ceedings of the 3rd International Conference on Information Science and Control Engineering. Piscataway: IEEE Press, 2016: 1089-1093.

[9] ZYKIN S, ZYKIN V. The boundary identification algorithm for a process control[C]//Proceedings of the Dynamics of Systems, Mechanisms and Machines. Piscataway: IEEE Press, 2022: 1-4.

[10] CAO H Y, LI H Y, SUN W, et al. A boundary identification approach for the feasible space of structural optimization using a virtual sampling technique-based support vector machine[J]. Computers & Structures, 2023, 287: 107118.

[11] ZHOU G Q, WEI H, WANG X Y, et al. BSMNet: boundary-salience multi-branch network for intima-media identification in carotid ultrasound images[J]. Computers in Biology and Medicine, 2023, 162: 107092.

[12] SOWJANYA A M, SHASHI M. Cluster feature-based incremental clustering approach (CFICA) for numerical data[J]. International Journal of Computer Science and Network Security, 2010, 10: 73-79.

[13] 欧阳秋宝. 基于局部分布的自组织增量聚类算法[D]. 南京: 南京大学, 2012.
OUYANG Q B. Self-organizing incremental clustering algorithm based on local distribution[D]. Nanjing: Nanjing University, 2012(in Chinese).

[14] ZHENG L W, HUO H, GUO Y Y, et al. Supervised adaptive incremental clustering for data stream of chunks[J]. Neurocomputing, 2017, 219: 502-517.

[15] 张煜莹, 陆艺, 赵静. 基于增量学习的数控机床故障诊断系统[J]. 计量学报, 2022, 43(11): 1456-1463.
ZHANG Y Y, LU Y, ZHAO J. Fault diagnosis system of numerical control machine based on incremental learning[J]. Acta Metrologica Sinica, 2022, 43(11): 1456-1463(in Chinese).

[16] ESTER M, KRIEGEL H P, SANDER J, et al. Incremental clustering for mining in a data warehousing environment[C]//Proceedings of the 24th International Conference on Very Large Data Bases. New York: Morgan Kaufmann Publishers Inc, 1998: 323-333.

[17] HORE P, HALL L O, GOLDGOF D B. Single pass fuzzy C means[C]//Proceedings of the IEEE International Fuzzy Systems Conference. Piscataway: IEEE Press, 2007: 1-7.

[18] 夏雨轩, 曲海成, 关茜文. 一种面向图像相似度计算的改进哈希算法[J]. 智能计算机与应用, 2022, 12(4): 110-113.
XIA Y X, QU H C, GUAN Q W. An improved hash algorithm for image similarity calculation[J]. Intelligent Computer and Applications, 2022, 12(4): 110-113(in Chinese).

[19] DATAR M, IMMORLICA N, INDYK P, et al. Locality-sensitive hashing scheme based on p-stable distributions[C]//Proceedings of the Twentieth Annual Symposium on Computational Geometry. New York: ACM, 2004: 253-262.

[20] GIONIS A, INDYK P, MOTWANI R. Similarity search in high dimensions via hashing[C]//Proceedings of the 25th International Conference on Very Large Data Bases. Edinburgh: Morgan Kaufmann, 1999: 518-529.

Performance boundary identification method for autonomous decision-making systems based on neighbor boundary degree

LU Hui[1,*], LYU Jingru[2], WANG Shiqi[1], SUN Zebin[3]

(1. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China;
2. China Mobile Group Beijing Co., Ltd., Beijing 100007, China;
3. School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China)

Abstract: The ability of an autonomous decision-making system to withstand disruptions is reflected in its performance boundary, which is a crucial indicator of its resilience. A performance boundary identification approach based on neighbor boundary degrees is proposed for autonomous decision-making systems, taking into account the features of multi-space distribution and incremental creation of performance boundary data. In order to solve the absolute scale measurement problem, we first design the neighbor boundary degree index. Then, we propose a performance boundary identification process based on neighbor boundary degree, which addresses the challenges of a complex performance boundary search space and non-uniform scale throughout the space. Secondly, the incremental performance boundary Identification method based on neighbor boundary degree is proposed by combining incremental data with the original identification results in order to accurately describe the performance boundary and to achieve efficient incremental data processing. An approximate nearest neighbor search optimization technique that enhances local sensitive hashing is then suggested in order to address the efficiency issue of nearest neighbor search and reverse nearest neighbor search that arose in the incremental phase. Finally, benchmark systems and path planning systems are used as autonomous decision-making systems to carry out verification and analysis of theoretical research work. Experimental results show that the performance boundary identification method based on neighbor boundary degree has good generalization ability of algorithm parameters. In the experiment on the path planning system, this method has a 13.68% higher boundary recognition accuracy and a 91.57% shorter running time compared to the comparison method.

Keywords: autonomous decision-making system; robustness test; performance boundary identification; incremental identification; locality sensitive Hashing

Received: 2023-11-24; **Accepted**: 2024-06-07; **Published Online**: 2024-07-08 11:48
URL: link.cnki.net/urlid/11.2625.V.20240705.1512.004
Foundation item: National Natural Science Foundation of China (62371030)
* **Corresponding author**. E-mail: mluhui@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0719

面向多无人机物流配送的双层任务规划方法

王飞*，杨清平
(中国民航大学 空中交通管理学院，天津 300300)

摘　　要：多无人机任务协同规划与配送路径规划是城市无人机物流配送的核心内容，两者相互耦合，需要进行一体化研究。为保障安全、高效完成多无人机物流配送任务，采用栅格法对三维城市超低空间进行环境建模，阐述了栅格危险度计算方法。构建一种无人机配送线路及航迹协同规划的双层规划模型，在上层规划模型中，考虑无人机载重及最大航程约束，以延迟惩罚代价最小为目标，引入遗传算法来确定无人机配送顺序；在下层规划模型中，考虑无人机性能约束，以时效性代价最小、无人机高度变化及栅格危险度最小为目标，提出一种综合改进粒子群优化（CIPSO）算法，求解无人机飞行路径。进行算例仿真分析，结果表明：与粒子群优化（PSO）算法、改进加速因子粒子群优化（ICPSO）算法相比，CIPSO算法总代价分别下降了65.00%和38.41%，所建模型与所提算法是可行的和有效的。

关 键 词：物流无人机；任务分配；路径规划；双层规划模型；改进粒子群优化算法
中图分类号：V355
文献标志码：A　　**文章编号**：1001-5965(2026)01-0094-10

城市居民消费能力迅猛发展带来了电子商务快速发展，物流业爆发式增长，但同时也伴随着城市交通拥堵，快递延误等问题。无人机(unmanned aerial vehicle, UAV)具有机动性强、灵活度高及成本低廉的优点，为解决城市物流“末端一公里”提供了新的途径，在国内外诸多城市已经得到了探索性实践[1]。多无人机物流配送的核心任务有2个方面：①配送任务分配，即根据客户需求给出货物配送顺序；②配送路径规划，也称航迹规划或路线规划，即给出无人机到每个任务点的飞行路径。现有研究大多将这2个任务分开独立研究，较少从整体上进行一体化研究。

多无人机配送任务分配方面：李翰等[2]综合考虑无人机运行可靠性、飞行性能等约束条件，以安全风险最小和物流成本和为目标函数，建立物流运输中多无人机协同任务分配模型；张瑞鹏等[3]构建一种综合考虑飞行航程、任务收益及任务完成时间窗口的无人机任务分配模型；Devi等[4]通过二进制矩阵编码实现UAV的离散化，应用粒子群优化(particle swarm optimization, PSO)算法求解多UAV任务分配问题；Jiang等[5]在物流无人机任务分配问题上建立综合考虑配送时间窗等多约束模型，并利用改进的PSO算法求解模型；Liu等[6]以分配无人机数最小、总航程最短为目标建立多目标函数解决多无人机巡航中路径协同问题，并应用多目标进化算法求解该模型。

多无人机配送路径规划方面：在单无人机路径规划研究基础上[7-9]，多无人机路径协同规划也取得了一些研究成果。Han等[10]以每飞行小时在地面的死亡概率来定义各栅格的安全系数，建立基于地面风险评估的最优航路规划模型；单文昭等[11]将粒子群和Hook-Jeeves搜索算法相融合，先分别求解

收稿日期：2023-11-03；**录用日期**：2024-03-01；**网络出版时间**：2024-04-03 20:23
网络出版地址：link.cnki.net/urlid/11.2625.V.20240402.1802.001
基金项目：天津市应用基础多元投入基金重点项目(21JCZDJC00840)；中央高校基本科研业务费专项资金(3122019129)
* **通信作者**. E-mail：feiwang@cauc.edu.cn

引用格式：王飞，杨清平. 面向多无人机物流配送的双层任务规划方法[J]. 北京航空航天大学学报，2026，52(1)：94-103.
WANG F，YANG Q P. Two-layer task planning method for multi-UAV logistics distribution[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：94-103 (in Chinese).

单无人机路径信息, 再通过集中航迹规划层协调到达时间, 从而实现多无人机协同航迹规划; 郑锴等[12]针对多无人机多目标侦察任务的需求, 对传统 A*算法、K-means 算法和深度遍历法进行改进, 根据实时变化动态调整任务分配; 潘楠等[13]提出基于 Levy 飞行策略的混合差分乌鸦搜索算法, 将剪枝处理和 Logistic 混沌映射机制加入快速遍历随机树算法中, 并通过改进的快速扩展随机树(rapidly exploring random tree, RRT)算法进行航迹初始化, 以此提高异构无人机航迹规划的效率。

任务分配和路径规划一体化研究方面: 余婧等[14]设计了双层互耦的任务规划求解策略, 用嵌套的 2 层模型解决二维环境中多军用无人机的任务分配、打击顺序和路径规划问题; 魏明等[15]设计一种基于深度强化学习 (deep reinforcement learning, DRL) 的 2 阶段混合算法, 在第 1 阶段利用 DRL 算法生成多架 UAV 访问客户顺序的配送线路, 在第 2 阶段应用 A*算法搜索每架 UAV 的最短路径。目前, 多无人机协同路径规划研究集中于二维空间, 鲜有涉及城市三维空间环境[16]。

本文针对 120 m 以下的城市超低空三维环境, 采用栅格法进行环境建模, 考虑障碍物约束, 提出一种无人机任务分配及路径协同规划的双层规划模型。上下 2 层模型互相嵌套, 相互影响。鉴于 2 层模型都属于 NP-hard 难题, 本文引入遗传算法求解上层规划模型, 设计一种综合改进的 PSO (comprehensive improved PSO, CIPSO)算法解决底层模型问题。最后, 通过仿真分析本文算法、本文模型的有效性。

1　城市超低空环境建模

环境建模是将无人机实际飞行的规划空间抽象为数学模型的过程来进行数学描述。城市环境建模中常见的方法包括栅格法、拓扑图法及可视图法等[17]。其中, 栅格法在无人机运行的城市空间建模中应用较为普遍, 其原理是将无人机运行空间分割成栅格单元, 每个栅格被赋予了表示相应运行代价的数值。

本文采用栅格法进行空间建模。如图 1 所示, 其中, S 和 G 分别为轨迹的起点和终点, 假设一个立方体 $OABC-O'A'B'C'$ 的长度为 x、宽度为 y、高度为 z, 每个栅格的像元粒度(边长)为 l_{grid}, 该立方体区域以 O 为原点, 将无人机运行空间划分为 $m\times n\times h$ 个栅格单元, $m=\text{int}(x/l_{\text{grid}})$, $n=\text{int}(y/l_{\text{grid}})$, $h=\text{int}(z/l_{\text{grid}})$, int()表示向下取整运算, 每个栅格用其中心点坐标表示。

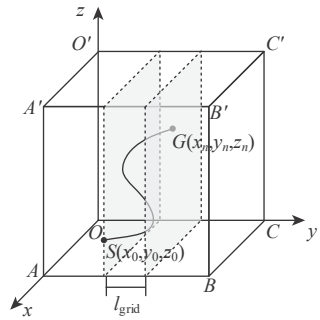


图 1　三维环境建模

Fig. 1　3D environment modeling

栅格像元粒度大小受计算机性能及空间分辨率影响。栅格粒度过大会对无人机路径搜索产生影响, 使算法难以收敛到最优路径; 粒度过小将会显著增加计算时间。城市超低空环境中常见的静态障碍物包括高大树木、建筑物及构筑物等, 往往在环境建模中简化为长方体来表示。

传统 0-1 赋值法简单易用, 包含障碍物的栅格被赋值为 1, 不存在障碍物的栅格被赋值为 0。缺点是没有考虑无障碍物的栅格也会产生潜在的风险, 无法保证对飞行路线安全性精细化度量。因此, 本文在环境建模中引入栅格危险度[18]来量化可通过栅格的危险性。如图 2 所示, 灰色栅格的危险度为 1, 表示其中存在障碍物, 白色栅格中不存在障碍物, 其危险度为与其相邻栅格中存在障碍物的栅格与周边 26 个栅格的比值。白色栅格危险度 ϕ_k 表达式为

$$\phi_k = N_{\text{obstacle},k}/N_{\text{surrounding},k} \tag{1}$$

式中: k 为栅格序号; $N_{\text{obstacle},k}$ 为与栅格 k 相邻且存在障碍物的栅格数量; $N_{\text{surrounding},k}$ 为与栅格 k 相邻的栅格总数。

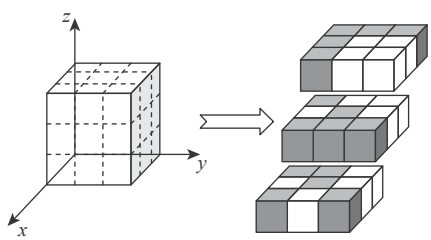


图 2　栅格危险度示意图

Fig. 2　Grid hazard diagram

2　多无人机配送问题模型与算法

针对多无人机物流配送问题, 本文建立了双层规划模型。上层规划模型用于解决任务分配问题, 即主要解决无人机对各个客户的配送顺序及货物的分配方案; 下层规划模型用于解决多无人机路径规划问题, 即根据任务分配方案计算多无人机配送

的最优路径。上层规划模型和下层规划模型相互联系和相互影响，规划过程如图3所示。

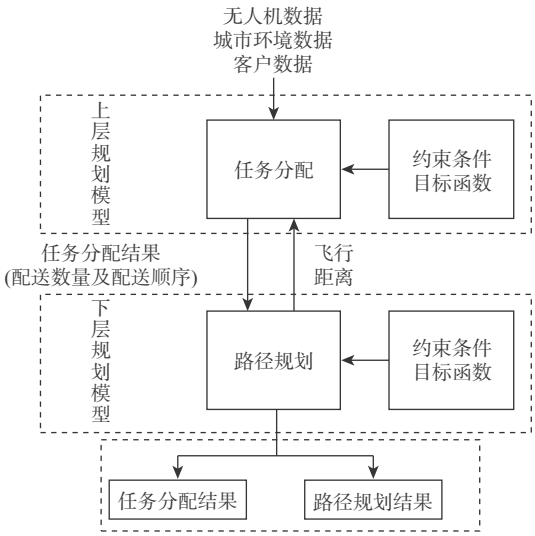


图3 双层任务规划过程

Fig. 3 Process of bi-level mission planning

从图3可以看出，上层规划模型计算的任务分配方案是下层规划模型规划路径的前提，而下层规划模型计算的多无人机飞行路径长度又影响上层规划模型的决策目标。具体而言，上层规划模型给定一个任务分配方案，下层规划模型在此方案基础上规划多无人机路径，计算出路径长度，并将路径长度反馈到上层规划模型，上层规划模型计算目标函数，并根据优化目标进一步调整任务分配方案并输出给下层规划模型，多轮交替循环后输出最终的任务分配及路径规划结果。

2.1 基于遗传算法的多无人机任务分配

如图4所示，为提升物流无人机配送效率减少配送成本，在分配过程中以所有无人机物流配送的延迟惩罚代价最小为目标，考虑无人机载荷限制、最大航程限制、客户时间窗限制，计算最优的配送任务分配方案。

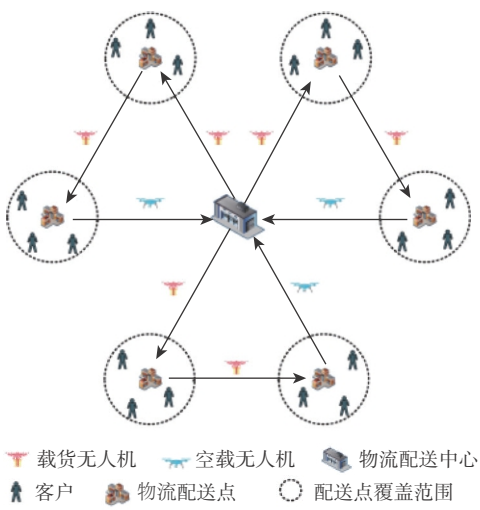


图4 城市环境物流无人机配送示意图

Fig. 4 Urban environmental logistics UAV distribution diagram

模型假设如下：

1）配送所采用的无人机异构，其速度、载重、航程不同；

2）假设该地区仅有一个配送中心；

3）无人机向各个客户配送的货物质量不同，卸货时间不同。

2.1.1 上层规划模型目标函数

假设共有N架物流无人机$U=\{U_1,U_2,\cdots,U_N\}$和N_T个任务$T_{\text{ask}}=\{T_1,T_2,\cdots,T_{N_T},T_0\}$，其中，$T_0$表示无人机完成第$T_{N_T}$个任务后必须返回配送中心。

模型任务分配决策变量x_{ij}为

$$x_{ij}=\{0,1\} \tag{2}$$

当任务j分配给无人机U_i时，x_{ij}为1，否则为0。执行配送任务时，无人机U_i首先从配送中心A出发，依次执行配送序列$\Delta_i=<M_1^i,M_2^i,\cdots,M_m^i,M_{m+1}^i>$中的任务，其中，$M_{m+1}^i$为$U_i$返回配送中心。对于$M_\delta^i\in\Delta_i$，令$t_i(M_\delta^i)$为无人机$U_i$完成任务$M_\delta^i$的时刻，计算过程如下：

$$t_i(M_\delta^i)=\begin{cases}\dfrac{d(A,M_1^i)}{v_i}+p(M_1^i) & \delta=1\\ \dfrac{d(A,M_1^i)+\sum\limits_{n=1}^{\delta-1}d(M_n,M_{n+1})}{v_i}+\sum\limits_{n=1}^{\delta}p(M_n^i) & 1<\delta\leqslant m_i\end{cases} \tag{3}$$

式中：v_i为无人机U_i速度；$p(M_\delta^i)$为任务M_δ^i完成需要的时间；$d(\cdot)$为实际飞行距离。

物流无人机需要尽可能在顾客指定时间内完成配送任务。设顾客为某个任务设置的时间窗为$[0,t_j]$，t_j表示顾客为本次配送设置的时间限制。若无人机完成任务时间晚于t_j则需要付出额外经济成本，假设每架无人机的单位时间延迟成本是相同的，延迟代价则可以用延迟时间来表示，C_i表示第i架无人机的延迟惩罚代价，如下：

$$C_i=\begin{cases}\sum\limits_{j=1}^{N_T}(t(M_\delta^i)-t_j) & t(M_\delta^i)>t_j\\ 0 & t(M_\delta^i)\leqslant t_j\end{cases} \tag{4}$$

2.1.2 上层规划模型约束条件

1）载重约束

物流无人机配送过程中受自身性能限制单次载重不得超过自身运行载荷限制。无人机U_i单次从配送中心起飞装载质量$L_{\text{D}i}$，与最大载荷能力$L_{\text{C}i}$，应满足如下约束：

$$L_{Di} \leqslant L_{Ci} \qquad \forall U_i \in U \tag{5}$$

2) 最大航程约束

考虑到能源供应, 无人机的航程是不容忽视的问题。U_i飞行距离L_i与其最大航程$L_{i(\max)}$满足如下约束:

$$L_i = d(A, M_1^i) + \sum_{\delta=1}^{m_i} d(M_\delta^i, M_{\delta+1}^i) \tag{6}$$

$$L_i \leqslant L_{i(\max)} \tag{7}$$

式中: L_i为无人机U_i的配送路径长度; m_i为无人机U_i分配到的配送任务数量。

综上所述, 上层规划模型中面向物流无人机任务分配问题的多目标函数为

$$\begin{cases} \mathrm{Min}\sum_{i=1}^{N} C_i \times x_{ij} \\ \text{s.t.} \begin{cases} L_{Di} \leqslant L_{Ci} & \forall U_i \in U \\ L_i \leqslant L_{i(\max)} & \\ \sum_{i=1}^{N} x_{ij} = 1 & \forall M_j \in M \\ x_{ij} \in \{0,1\} & \forall U_i \in U,\ \forall M_j \in M \end{cases} \end{cases} \tag{8}$$

本文上层规划模型采用标准遗传算法(genetic algorithm, GA)[19]输出最优任务分配方案, 在编码过程中染色体设置是核心内容。单个染色体是由 2 个部分组成, 第 1 部分是任务分配方案, 即每架无人机分配的任务序号, 第 2 部分是任务配送顺序, 即每架无人机配送货物的顺序。GA 的选择、交叉、变异等操作都采用标准形式。

2.2　基于改进粒子群算法的多无人机路径规划

在上层规划模型解决了多无人机配送访问客户次序之后, 下层规划模型根据时效性、高度变化、栅格危险度为目标优化设计出连接配送中心与客户、客户与客户的无人机飞行航迹。多无人机运行会面临时空重叠产生冲突的问题, 这属于运行控制层面问题, 不属于规划层面问题, 因此, 本文在路径规划过程中不考虑无人机之间的路径冲突问题。

2.2.1　下层规划模型目标函数

1) 时效性代价

配送总耗时T_{total}可表示为

$$T_{\text{total}} = \sum_{i=1}^{N} t_i(M_{m_i}^i) \tag{9}$$

2) 飞行高度变化代价

在配送过程中, 频繁变化的无人机飞行高度会导致无人机的能耗增加, 从而提高了配送运营的成本[8]。飞行高度变化可表示为

$$H = \sum_{i=1}^{N} \sum_{n=1}^{P_n} \Delta z_{i(P_{n-1},P_n)} \tag{10}$$

式中: $\Delta z_{i(P_{n-1},P_n)}$为第 i 架无人机在第 P_n 个航迹点与第 P_{n-1} 个航迹点的高度差。

3) 栅格危险度

无人机运行安全性由栅格危险度代价表示, 即无人机飞行路径经过的栅格对应危险度的累加, 计算式为

$$D = \sum_{i=1}^{N} \sum_{k=0}^{n} \phi_k \tag{11}$$

综上所述, 本文模型目标函数为

$$\min(\alpha_1 T_{\text{total}} + \alpha_2 H + \alpha_3 D) \tag{12}$$

2.2.2　下层规划模型约束条件

1) 运行高度约束

物流无人机配送高度 z 应维持在最大运行高度$z_{\max}$、最小运行高度$z_{\min}$之间以此满足运行风险考量。约束关系如下:

$$z_{\min} \leqslant z \leqslant z_{\max} \tag{13}$$

2) 爬升角约束

物流无人机运行中爬升角受到其性能的限制, 其爬升角度约束关系如下:

$$\beta_i = \tan^{-1} \frac{z_{i+1} - z_i}{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}} \leqslant \beta_{\max} \tag{14}$$

式中: β_i为无人机于第i个航迹点爬升角; $\beta_{\max}$为最大爬升角, $i = 1,2,\cdots,N+1$。

2.2.3　改进粒子群算法设计

为提高粒子群最优解搜索速度, 本节在传统 PSO 算法基础上做以下改进构建 CIPSO 算法: 利用 Singer 映射以改进粒子初始分布状态, 对加速因子数值、粒子最大速度数值做线性改进, 对惯性权重做自适应改善, 提出改进的位置更新策略。

1) 混沌初始化

提高粒子初始均匀度以改善其遍历性随机性, 在改善搜索能力的同时加快算法收敛速度。与随机初始化相比, 混沌初始化[20]被认为是一种有效的策略, 可以更好地分散粒子群, 增加种群多样性。Singer 模型是一种常见的产生混沌序列的数学模型[21], 其过程如下:

$$x_{n+1} = \mu(7.68x_n - 23.31x_n^2 + 28.75x_n^3 - 13.3x_n^4) \tag{15}$$

式中: μ为控制参数, 其取值范围在 [0.9,1.08] 之间; x_n为序号为 n 的混沌变量。

2) 线性调整加速因子

传统 PSO 算法加速因子往往设为常数, 本文采用线性方法自适应调整加速因子c_1、c_2, 以此改善粒子搜索能力避免陷入局部最优解, 计算式为

$$c_1 = c_{\max} - (c_{\max} - c_{\min})t/T \tag{16}$$

$$c_2 = c_{\min} + (c_{\max} - c_{\min})t/T \tag{17}$$

式中：t为当前迭代次数；T为最大迭代次数；$c_{\max}$为最大加速因子，$c_{\min}$为最小加速因子，设为常值且$c_{\max} > c_{\min} > 0$。$c_{\max}$、$c_{\min}$满足$(c_{\max} + c_{\min})/2 = c_1^{\text{std}} = c_2^{\text{std}}$，$c_2^{\text{std}}$、$c_1^{\text{std}}$为标准PSO算法加速因子，设为常值。$c_1$、$c_2$分别线性减小和线性增大，且$c_1 + c_2 = c_{\max} + c_{\min}$，以此满足算法收敛、搜索能力不变。迭代前半程加速因子满足$c_1 > c_2$，以此提升粒子搜索能力；迭代后半程加速因子满足$c_1 < c_2$，以此提高收敛速度。

3）线性调整粒子最大速度

传统PSO算法中最大速度$v_{\max}^p$往往设置为常数，抑制了粒子搜索能力。本节如式(18)所示，对$v_{\max}^p$做相应线性改进。

$$v_{\max}^p = v_1^p - (v_1^p - v_2^p)t/T \tag{18}$$

式中：v_1^p、v_2^p为粒子p的更新速度，其均大于0且满足$(v_1^p + v_2^p)/2 = 2v_{\max}^{\text{std}}$，$v_{\max}^{\text{std}}$为标准PSO算法最大更新速度。

4）位置更新策略

本节在算法运行中执行一次PSO算法后，将粒子适应度排序，如式(19)和式(20)所示，对粒子进行位置更新，以此改善粒子收敛能力。

$$P_{\text{large}} = P_{\text{small}} + ar_3 \tag{19}$$

$$v_{\text{large}} = v_{\text{small}} \tag{20}$$

式中：P_{large}和P_{small}分别为适应度较大和较小的一半粒子；V_{large}和V_{small}分别为粒子的最大和最小速度；r_3为一个随机值，其取值范围于[0,1]之间；a为体现突变粒子位置偏移程度的常量。

5）自适应调整惯性权值

传统粒子群中惯性权值常设为固定值，为避免在迭代中提前进入局部最优，本节对其随着迭代过程的进行动态调整，调整策略如下：

$$\varpi_{t+1} = \varpi_t - \frac{0.75t^3}{T^3}(P_{t\max} - P_{t\min})^3 \tag{21}$$

式中：$P_{t\max}$、$P_{t\min}$分别为当前迭代中最大、最小惯性权值；ϖ_t为第t次迭代中的惯性权值。

本文算法的具体步骤如下：

算法 改进的粒子群优化算法

初始参数：粒子数量N；迭代次数T；最大、最小加速因子$c_{\max}$、$c_{\min}$；最大、最小更新速度v_1、v_2；位置更新常数a。

中间变量：最优解对应的速度V_{large}；最优解对应的位置P_{large}。

1: 初始化粒子种群

2: For $t = 1$ to T do

3: 计算适应度值：

$$x_{n+1} = \mu(7.68x_n - 23.31x_n^2 + 28.75x_n^3 - 13.3x_n^4)$$

4: 线性调整加速因子：$c_1 = c_{\max} - \dfrac{(c_{\max} - c_{\min})t}{T}$

5: 线性调整加速因子：$c_2 = c_{\min} + \dfrac{(c_{\max} - c_{\min})t}{T}$

6: 线性调整粒子最大速度：$v_{\max} = v_1 - \dfrac{(v_1 - v_2)t}{T}$

7: 执行粒子群算法，并按适应度值大小重新排列粒子

8: For 每个粒子 i do

9: 更新位置：$P_{\text{large}} = P_{\text{small}} + ar_3$

10: 更新速度：$v_{\text{large}} = v_{\text{small}}$

11: End for

12: 自适应调整惯性权值：$\varpi_{t+1} = \varpi_t - \dfrac{0.75t^3}{T^3}(P_{t\max} - P_{t\min})^3$

13: End for

15: Return v_{large}、P_{large}

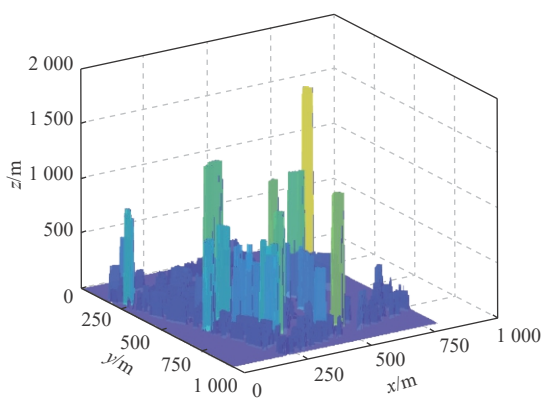
3 仿真分析

3.1 仿真环境与参数设置

选择天津市和平区五大道作为环境建模背景，如图5所示。假设物流配送点有5架异构无人机，性能参数如表1所示；需知执行的任务点有12个，任务参数如表2所示；其余参数设置如表3所示。

在实验进行时，采用了min-max标准化方法用以消除多目标函数3个子目标代价值之间可能存

(a) 卫星图



(b) 模拟图

图5 城市环境建模

Fig. 5 Urban environment modeling

表 1　无人机参数设置

Table 1　UAV's parameters setting

无人机编号	v_i/(m·s^{-1})	$L_{i(\max)}$/km	L_{Ci}/kg
U1	5.0	3.0	3.5
U2	6.0	4.0	4.0
U3	6.0	4.0	4.0
U4	6.0	4.0	5.8
U5	8.0	6.0	6.5

表 2　任务参数设置

Table 2　Parameters setting of task

任务序号	任务坐标/m	时间窗/s	卸货时间/s	货物质量/kg
T1	(39,22,1)	[0,120]	13	0.9
T2	(91,54,2)	[0,70]	8	0.4
T3	(54,43,2)	[0,80]	10	1.2
T4	(72,148,2)	[0,85]	12	2.1
T5	(93,146,1)	[0,110]	13	2.4
T6	(118,145,1)	[0,60]	8	0.8
T7	(120,119,1)	[0,95]	13	1.2
T8	(143,96,2)	[0,125]	20	2.7
T9	(43,93,1)	[0,90]	9	0.6
T10	(68,58,2)	[0,200]	18	1.2
T11	(36,148,1)	[0,160]	23	1.6
T12	(138,39,2)	[0,145]	15	1.3

表 3　其余参数设置

Table 3　Other parameter Setting

参数	数值	参数	数值
配送中心A坐标/m	(81,81,1)	α_1	0.1
α_2	0.4	α_3	0.5
l_{grid}/m	5	路径节点个数	5
$z_{\max}$/m	120	$z_{\min}$/m	0
$\beta_{\max}$/(°)	90		

在的量级差异, 将这些值缩放到 [0,1] 的范围内, 表达式为

$$F(x)=\frac{f(x)-f(x)_{\min}}{f(x)_{\max}-f(x)_{\min}} \tag{22}$$

3.2　仿真结果

上层规划模型 GA 种群大小设置为 150, 亲代选择概率设置为 0.8, 杂交概率设置为 0.8, 变异概率设置为 0.5, 迭代 100 次, GA 适应度值随迭代次数变化如图 6 所示。最终的无人机配送任务分配方案如表 4 所示, 与之对应的配送时序如图 7 所示。

由表 4 和图 7 可知, 12 个任务分配给了 4 架无人机, U1 未分配任务, 这是因为根据表 1, 在 5 架 UAV 中, U1 的运行速度最慢、最大航程最短、最大载重最小, 若将任务分配给 U1, 会导致式(1)的目

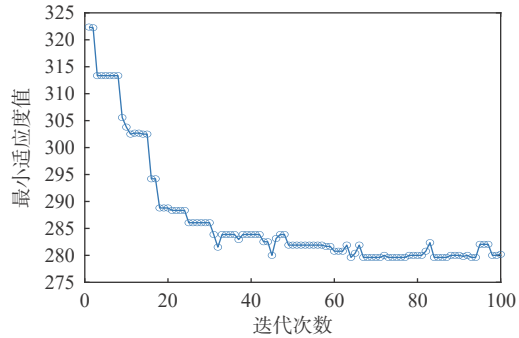


图 6　GA 适应度值随迭代次数变化

Fig. 6　GA fitness value varies with number of iterations

表 4　任务分配方案

Table 4　Task allocation scheme

无人机	到达各客户点的时间/s
U1	
U2	9.299(t_8)
U3	7.771 (t_3)-22.08 (t_1)-42.78 (t_{10})
U4	4.802 (t_2)-28.79 (t_4)-46.79 (t_{11})-79.03 (t_9)
U5	8.262 (t_5)-24.39 (t_6)-35.65 (t_7)-58.9 (t_{12})

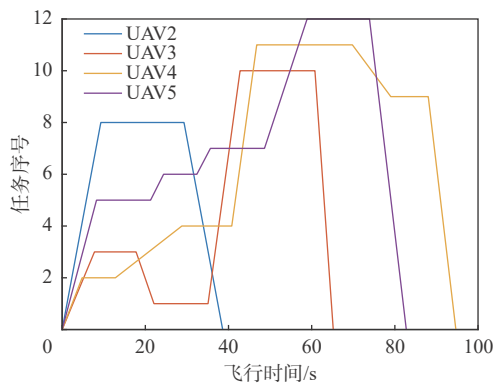
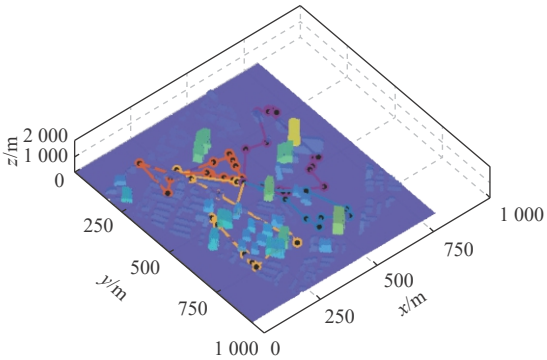


图 7　无人机飞行时序

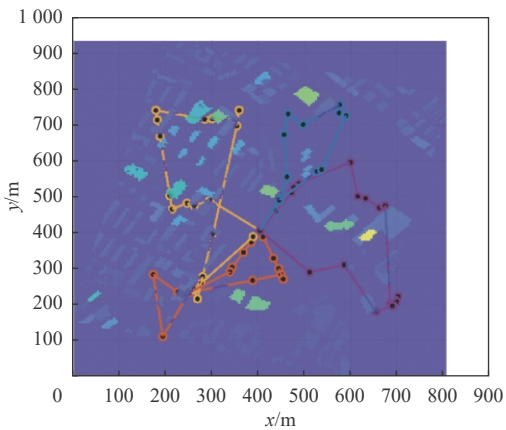
Fig. 7　UAVs flight timing

标函数变大, 在寻优过程中将被排除。各架无人机均满足自身航程和载重方面约束的要求。

下层规划模型的粒子群算法中, 最大加速因子$c_{\max}$设置为 3.5, 最小加速因子$c_{\min}$设置为 0.5, 最大更新速度$v_{\max}^{\text{std}}$设置为 0.3, 更新速度v_1、更新速度v_2分别设置为 0.5 和 0.1, 位置更新常数a设置为 0.2。分别采用 CIPSO 算法、PSO 算法、改进加速因子粒子群优化 (improved acceleration coefficients particle swarm optimization, ICPSO)[22] 这 3 种算法求解模型, 并比较算法性能。所有算法迭代次数和种群规模保持一致, 求解 1 次双层模型仅需 50 s 左右, 完全能够满足实际应用要求。最终路径规划结果如图 8 所示, 具体数值结果如表 5 所示。从表 5 和图 8 可以看出, 3 种算法的适应度值分别为 0.237 7、0.392 2、0.329 0。CIPSO 算法总代价值与 PSO 算



(a) 三维图



(b) 俯视图

图 8 规划结果

Fig. 8 Planning results

表 5 算法实验结果比较

Table 5 Comparison of algorithm experiment results

算法	目标函数值	时效性	飞行高度变化/m	栅格危险度	目标函数中位数	目标函数下4分位数	目标函数上4分位数
CIPSO	0.237 7	0.317 7	1.432 5	0.182 7	0.256 2	0.182 9	0.371 2
PSO[5]	0.392 2	0.652 9	2.090 5	0.319 4	0.626 8	0.390 6	0.700 5
ICPSO[22]	0.329 0	0.364 5	2.128 0	0.244 6	0.384 9	0.250 3	0.684 6

法、ICPSO 算法相比分别下降了 65.00% 和 38.41%，且在时效性代价、飞行高度变化两方面均取得了更优的结果，在栅格危险度取得次优结果。

为验证规划路径是否满足约束，提取各个节点的高度、俯仰角结果，分别如图 9 和图 10 所示。为验证 3 种算法差异性，进行了 Kruskal-Wallis H 检验，结果如图 11 所示。

从图 9 和图 10 可以看出，无人机在各自运行中各个节点的高度小于 120 m，俯仰角小于 90°。同时 4 架无人机飞行距离分别为 741.057 6，918.551 4，1 634.098 8，2 465.676 8 m，均小于 5 000 m 的最大航程限制。经 Kruskal-Wallis H 检验，渐近显著性 P 值为 0.02，小于 0.05 显著性水平，显示 3 种算法的中位数存在显著差异，表明 CIPSO 算法对 PSO 算法的改进有显著的影响。

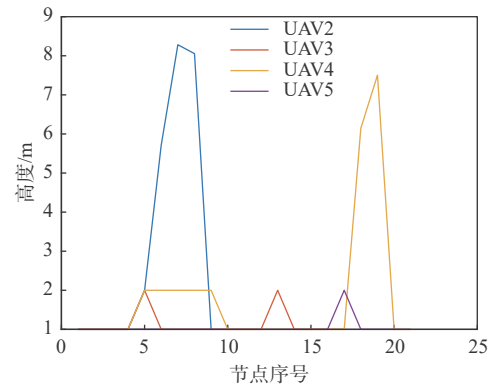


图 9 无人机高度

Fig. 9 UAV altitude

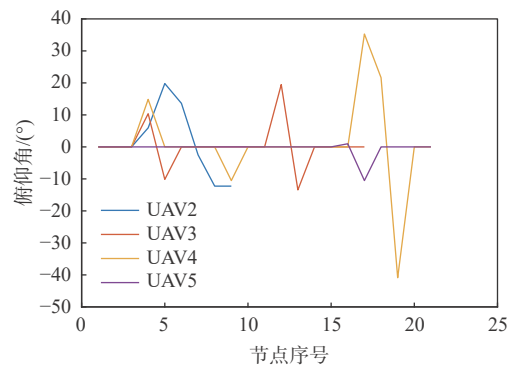


图 10 无人机俯仰角

Fig. 10 UAV pitch angle

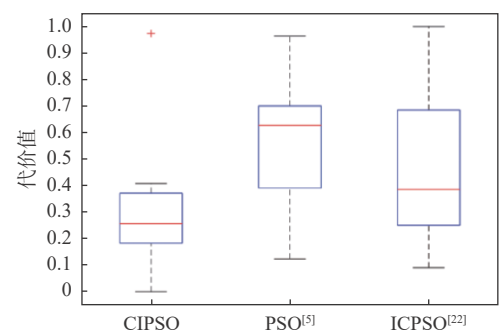


图 11 数据结果箱线图

Fig. 11 Data result box pilot

综合来看，CIPSO 算法在多无人机协同路径规划中是可行的。

3.3 参数灵敏性分析

模型求解中代价权重{α_1，α_2，α_3}会对总代价值造成影响。本节敏感性分析中做 8 组对照实验，每组实验重复 30 次，取平均值并记录于表 6 和图 12，可以看出，当$\alpha_1 = 0.05$，$\alpha_2 = 0.45$，$\alpha_3 = 0.5$时，路径规划的总代价值达到最优。

各个权重组合的 Kruskal-Wallis H 检验结果如图 13 所示。渐近显著性 P 值小于 0.05，拒绝各组分布相同的假设，不同的组别之间存在某种差异，说明权重系数的不同对结果产生了显著影响。

表 6　代价权重影响

Table 6　Influence of cost weighting

实验组别	α_1	α_2	α_3	时效性代价	高度代价	栅格危险度	总代价值	总代价中位数	总代价下四分位数	总代价上四分位数
1	0.05	0.45	0.5	0.537 4	0.124 0	0.259 8	0.212 6	0.083 3	0.069 7	0.151 4
2	0.10	0.40	0.5	0.490 9	0.245 1	0.357 8	0.326 0	0.226 3	0.188 8	0.423 9
3	0.15	0.35	0.5	0.496 2	0.168 0	0.329 2	0.297 8	0.334 4	0.253 8	0.365 3
4	0.20	0.30	0.5	0.461 0	0.243 5	0.384 8	0.357 7	0.378 0	0.343 4	0.442 3
5	0.25	0.25	0.5	0.528 2	0.150 1	0.303 1	0.321 1	0.461 9	0.446 9	0.492 2
6	0.30	0.20	0.5	0.618 9	0.168 5	0.406 2	0.422 5	0.571 2	0.541 1	0.694 6
7	0.35	0.15	0.5	0.559 6	0.236 4	0.369 3	0.416 0	0.694 1	0.630 2	0.727 1
8	0.40	0.10	0.5	0.420 9	0.203 9	0.324 7	0.351 1	0.703 2	0.685 7	0.727 0

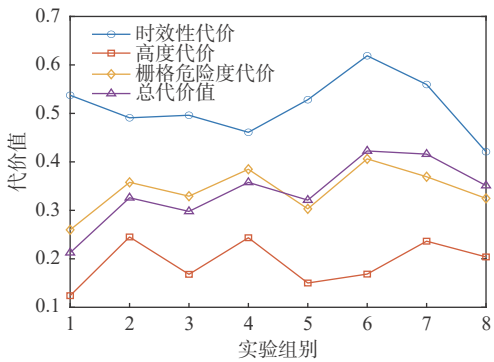


图 12　权重值影响

Fig. 12　Influence of weight value

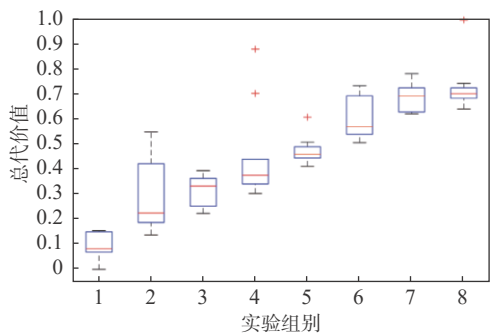


图 13　参数灵敏性分析箱线图

Fig. 13　Parametric sensitivity analysis box diagram

4　结　论

1) 多无人机任务分配和路径规划, 并不是孤立存在, 必须考虑两者耦合关系, 从而进行一体化研究。

2) 本文构建了相互嵌套的双层模型能有效表征上下两层间的交互关系, 设计了 GA 和 ICPSO 算法分别求解上下层规划模型, 能够获得有效的任务分配和路径规划方案, 为无人机物流配送实践提供理论指导。

本文着重考虑的是规划层面的飞行路径, 并没有涉及到运行层面的飞行冲突问题, 未来将重点解决这一问题。

参考文献（References）

[1] HOMOLA J, MARTIN L, CENCETTI M, et al. UAS traffic management (UTM) technical capability level 3 (TCL3) flight demonstration: concept tests and results[C]//Proceedings of the IEEE/AIAA 38th Digital Avionics Systems Conference. Piscataway: IEEE Press, 2019: 1-8.

[2] 李翰, 张洪海, 张连东, 等. 城市区域多物流无人机协同任务分配[J]. 系统工程与电子技术, 2021, 43(12): 3594-3602.
LI H, ZHANG H H, ZHANG L D, et al. Multiple logistics unmanned aerial vehicle collaborative task allocation in urban areas[J]. Systems Engineering and Electronics, 2021, 43(12): 3594-3602(in Chinese).

[3] 张瑞鹏, 冯彦翔, 杨宜康. 多无人机协同任务分配混合粒子群算法[J]. 航空学报, 2022, 43(12): 326011.
ZHANG R P, FENG Y X, YANG Y K. Hybrid particle swarm algorithm for multi-UAV cooperative task allocation[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(12): 326011(in Chinese).

[4] DEVI J, KRUIS F E. A fast Monte Carlo GPU based algorithm for particle breakage[C]//Proceedings of the 4th International Conference on Control, Decision and Information Technologies. Piscataway: IEEE Press, 2017: 784-789.

[5] JIANG X W, ZHOU Q, YE Y. Method of task assignment for UAV based on particle swarm optimization in logistics[C]//Proceedings of the International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence. New York: ACM , 2017: 113-117.

[6] LIU X F, PENG Z R, CHANG Y T, et al. Multi-objective evolutionary approach for UAV cruise route planning to collect traffic information[J]. Journal of Central South University, 2012, 19(12): 3614-3621.

[7] 张洪海, 李翰, 刘嶂, 等. 城市区域物流无人机路径规划[J]. 交通运输系统工程与信息, 2020, 20(6): 22-29.
ZHANG H H, LI H, LIU H, et al. Path planning for logistics unmanned aerial vehicle in urban area[J]. Journal of Transportation Systems Engineering and Information Technology, 2020, 20(6): 22-29(in Chinese).

[8] 徐伟华, 聊士超, 张根瑞, 等. 改进 Theta*算法的物流无人机城域三维路径规划[J]. 计算机工程与应用, 2023, 59(17): 334-340.
XU W H, LIAO S C, ZHANG G R, et al. 3D path planning of logistics UAV based on improved Theta* algorithm in metropolitan area[J]. Computer Engineering and Applications, 2023, 59(17): 334-

340(in Chinese).

[9] 张宏宏, 甘旭升, 李双峰, 等. 复杂低空环境下考虑区域风险评估的无人机航路规划[J]. 仪器仪表学报, 2021, 42(1): 257-266.
ZHANG H H, GAN X S, LI S F, et al. UAV route planning considering regional risk assessment under complex low altitude environment[J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 257-266(in Chinese).

[10] HAN P, ZHAO Y F. An Air Route planning model of unmanned aerial vehicles under constraints of ground safety.[J]. Transactions of Nanjing University of Aero nautics and Astronautics, 2021, 38(2): 298-305.

[11] 单文昭, 崔乃刚, 黄蓓, 等. 基于PSO-HJ算法的多无人机协同航迹规划方法[J]. 中国惯性技术学报, 2020, 28(1): 122-128.
SHAN W Z, CUI N G, HUANG B, et al. Multiple UAV cooperative path planning based on PSO-HJ method[J]. Journal of Chinese Inertial Technology, 2020, 28(1): 122-128(in Chinese).

[12] 郑锴, 尹栋, 殷少锋, 等. 基于改进A*算法的多基地多无人机分阶段任务规划方法[J]. 中国惯性技术学报, 2022, 30(2): 248-256.
ZHENG K, YIN D, YIN S F, et al. Multi-stage mission plan method of multi-UAVs deployed in different bases based on improved A* algorithm[J]. Journal of Chinese Inertial Technology, 2022, 30(2): 248-256(in Chinese).

[13] 潘楠, 张淼寒, 韩宇航, 等. 面向城市巡防的多无人机协同航迹规划[J]. 信息与控制, 2022, 51(4): 411-422.
PAN N, ZHANG M H, HAN Y H, et al. Coordinated path planning of multi-UAV for unban patrol[J]. Information and Control, 2022, 51(4): 411-422(in Chinese).

[14] 余婧, 雍恩米, 陈汉洋, 等. 面向多无人机协同对地攻击的双层任务规划方法[J]. 系统工程与电子技术, 2022, 44(9): 2849-2857.
YU J, YONG E M, CHEN H Y, et al. Bi-level mission planning method for multi-cooperative UAV air-to-ground attack[J]. Systems Engineering and Electronics, 2022, 44(9): 2849-2857(in Chinese).

[15] 魏明, 孙雅茹, 孙博, 等. 基于深度强化学习的无人机线路及航迹协同规划[J]. 中国安全科学学报, 2023, 33(8): 68-76.
WEI M, SUN Y R, SUN B, et al. UAV distribution route and flight path collaborative planning based on deep reinforcement learning [J]. China Safety Science Journal, 2023, 33(8): 68-76(in Chinese).

[16] MUÑOZ G, BARRADO C, ÇETIN E, et al. Deep reinforcement learning for drone delivery[J]. Drones, 2019, 3(3): 72.

[17] CHE H L, WU Z Z, KANG R X, et al. Global path planning for explosion-proof robot based on improved ant colony optimization[C]// Proceedings of the Asia-Pacific Conference on Intelligent Robot Systems. Piscataway: IEEE Press, 2016: 36-40.

[18] 张启钱, 许卫卫, 张洪海, 等. 复杂低空物流无人机路径规划[J]. 北京航空航天大学学报, 2020, 46(7): 1275-1286.
ZHANG Q Q, XU W W, ZHANG H H, et al. Path planning for logistics UAV in complex low-altitude airspace[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(7): 1275-1286(in Chinese).

[19] WU X L, YIN Y N, XU L, et al. Multi-UAV task allocation based on improved genetic algorithm[J]. IEEE Access, 2021, 9: 100369-100379.

[20] TIAN D P, SHI Z Z. MPSO: modified particle swarm optimization and its applications[J]. Swarm and Evolutionary Computation, 2018, 41: 49-68.

[21] GANDOMI A H, YANG X S. Chaotic bat algorithm[J]. Journal of Computational Science, 2014, 5(2): 224-232.

[22] SHAO S K, PENG Y, HE C L, et al. Efficient path planning for UAV formation via comprehensively improved particle swarm optimization[J]. ISA Transactions, 2020, 97: 415-430.

Two-layer task planning method for multi-UAV logistics distribution

WANG Fei*, YANG Qingping

(College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In urban UAV logistics distribution, the two main components that must be merged are multi-UAV task collaborative allocation and distribution path design. In order to ensure the safety and efficiency of multi-UAV logistics distribution, the grid method is used to model the ultra-low space environment of a three-dimensional city, and the grid risk calculation method is described. Secondly, a two-layer programming model of UAV distribution route and flight path collaborative planning is constructed. In the upper layer model, considering the constraints of UAV load and maximum range, a genetic algorithm is introduced to determine the UAV distribution order with the goal of minimum delay penalty cost. In the lower model, a comprehensive improved particle swarm optimization(CIPSO) algorithm is proposed to solve the flight path of the UAV by considering the performance constraints of the UAV and aiming at the minimization of timeliness cost, UAV height variation, and grid risk. Last but not least, the simulation results demonstrate that CIPSO's overall cost is lower than that of particle swarm optimization(PSO)and improved acceleration coefficients particle swarm optimization(ICPSO)algorithm by 65.00% and 38.41%, respectively. This suggests that the constructed model and the proposed algorithm developed in this study are both practical and efficient.

Keywords: logistics unmanned aerial vehicle; task allocation; path planning; two-layer programming model; improved particle swarm optimization algorithm

Received: 2023-11-03; **Accepted**: 2024-03-01; **Published Online**: 2024-04-03 20:23

URL: link.cnki.net/urlid/11.2625.V.20240402.1802.001

Foundation items: Tianjin Application Basic Diversified Investment Fund Key Project (21JCZDJC00840); The Fundamental Research Funds for the Central Universities (3122019129)

* **Corresponding author**. E-mail: feiwang@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0721

一种测扭齿的接触式角度测量方法

童晓艳[1,*]，鲁东生[2]，柏慧[3]，常龙[4]

(1. 北京航空航天大学 宇航学院，北京 100191；2. 航空工业哈尔滨飞机工业集团有限责任公司 飞机设计研究所，哈尔滨 150060；
3. 中国航发南方工业有限公司，株洲 412002；4. 上海宇航系统工程研究所，上海 201108)

摘 要：测扭齿角度的测量在直升机动力系统的制造和性能测试中很常用。为提高装调设备制造中测扭齿角度测量的精度，设计一种利用磁性编码器测量测扭齿角度的接触式测量方案，分析利用接触式测针测量测扭齿角度过程中由于测针与直驱电机不对心导致的安装误差、测扭齿装夹时与直驱电机不同心导致的误差及二者耦合对测扭齿角度测量的影响；提出一种利用多余观测进行角度测量的方法，消除测扭齿和测针的安装误差对角度测量结果的影响，推导角度计算式；在设计并搭建的测量装置中利用所提方法测量了测扭齿的角度，并和三坐标机测量结果进行比较。结果表明：所提方法不仅降低对测量系统的安装和调试要求，还提高测扭齿角度测量的准确度，亦可用于类似的角度精密测量中。

关 键 词：角度测量；接触式测量；安装误差；测扭齿；直升机传动系统
中图分类号：TB22
文献标志码：A **文章编号**：1001-5965(2026)01-0104-06

通过测量角度来获得扭矩，是工程中常用的测量方法，在航空、纺织、医疗等行业都有应用[1-4]，随着制造业水平的提高，对角度测量准确性的要求也提高了。直升机动力系统的功率是其性能的主要参数，可以通过测量扭矩和转速后经计算间接得到功率；扭矩的测量则利用其和动力涡轮传动轴组件上的测扭齿之间的夹角关系，通过测扭齿的角度测量得到，所以测扭齿之间的角度测量是直升机等动力系统的制造和测试中的基本问题。目前测扭齿之间的角度值不能准确测到，一般采用三坐标机测量角度的相对变化值[5]。角度测量有接触式[6]和非接触式[7-14]两类；对于不同类型的测量对象，角度测量的方法不尽相同。

文献[2]利用自由转子转过的角度和时间的关系来测量纱线的扭矩，自由转子转过的角度是用光电传感器对其转动状态进行扫描并将捕捉到的光电信号传输至计算机得到的。对于喷油嘴和涡流室轴线所形成的空间角度，文献[6]用三坐标测量机进行了测量。文献[7]利用激光位移传感器的漫反射测量特性，结合三坐标机和位置敏感探测器，搭建非接触测量装置进行面角度测量。文献[8]对非接触三点内径测量法采用4个观测位置联立方程进行了位移传感器角度安装误差校准，并通过仿真分析确认观测位置数大于4时即可实现误差校准。文献[9]利用单目摄像头从侧面拍摄柱状旋转体周向的靶标图像实现角位移的无接触测量。文献[10]利用双目视觉技术通过在机翼上粘贴识别块实现可变后掠角机翼的角度的无接触测量。文献[11]针对四象限探测器对空间通信中不规则光斑偏转角度测量精度不足的问题，利用信息融合算法，将低分辨率的电荷耦合元件获取的光斑形状信息与四象限探测器对规则光斑高精度角度测量融合，获得了异形光斑光束偏转角度的高精度测量。文献[12]对雷达主瓣干扰下的目标角度测量问题，

收稿日期：2023-11-03；**录用日期**：2023-11-13；**网络出版时间**：2023-11-20 15:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20231117.1420.002
*通信作者. E-mail：tongxiaoyan@buaa.edu.cn

引用格式：童晓艳，鲁东生，柏慧，等. 一种测扭齿的接触式角度测量方法[J]. 北京航空航天大学学报，2026，52(1)：104-109.
TONG X Y，LU D S，BAI H，et al. A contact angle measurement method for torsional teeth[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：104-109 (in Chinese).

考虑等距均匀线阵, 经过滑窗处理形成多个子阵; 然后, 在子阵内进行自适应数字波束形成, 抑制主瓣干扰; 最后, 联合稀疏恢复来估计目标角度。文献 [13] 利用相互垂直的双相机采集斜齿轮端面与侧面图像, 并进行图像处理, 根据斜齿轮齿顶圆螺旋角与分度圆螺旋角的关系计算出斜齿轮的螺旋角大小。文献 [14] 设计了基于霍尔传感器的交叉正交结构来定位角度, 利用一种新的角度标定算法, 降低了偏置电压、幅值不等和附加谐波引起的误差, 通过一个专用的现场可编程门阵列(field programmable gate array, FPGA)装置来测量人工磁场的旋转角度。

通过以上分析可以看出, 虽然角度测量是理论研究和工程测量中的常见问题, 但是, 针对不同的角度测量对象, 角度测量的原理和方法各不相同, 角度测量方法和装置的专用性很强; 这些方法和装置的多样性也带来了测量误差来源的多样性和针对性及误差消除方式的针对性。角度测量和平差的实现在不同的测量对象和测量要求下有不同的实现形式, 需要针对具体需求灵活设计。通过分析测量精度要求和被测对象的几何结构, 本文对测扭齿角度进行了接触式测量, 建立基本的测量模型, 利用测针获得的数据计算得到测扭齿角度的数值; 然后, 分析测针和测扭齿安装对角度测量结果的影响; 进一步地, 提出基于多余观测的测扭齿角度测量方法, 推导角度计算式, 理论推导表明此种测扭齿的角度测量方法消除了角度测量中安装误差的影响, 能够提高测量精度。由于角度测量精度对工件和测具的安装误差不敏感, 本文方法也降低了测量系统调试的难度和时间及对工件安装的同轴度要求, 有利于提高生产效率。

1　测扭齿角度测量方法

1.1　测扭齿角度测量装置

工程中用到的测扭齿如图 1 所示。由于其齿面窄小, 周围有部分遮挡, 而且在制造中角度测量精度要求较高, 在比较了几种非接触式和接触式角度测量方案后, 本文设计了如图 2 所示的接触式角度测量装置。采用旋转接触式测量, 选择带有绝对式磁性编码器的直驱电机作为转台; 转台上安装位移台实现测针进给运动; 触发式测针安装在位移台上, 并且控制电机的转动与停止。测扭齿的轴水平安装在夹具上。

1.2　角度测量的基本方法

采用旋转测量法测量测扭齿 T_1 和 T_2 间的角度。触发式测针的测球依次接触测扭齿的 4 个齿面, 得到 4 个接触位置的绝对角度, 再计算出测扭齿的角度。电机带动测针工作时的照片如图 3 所示。

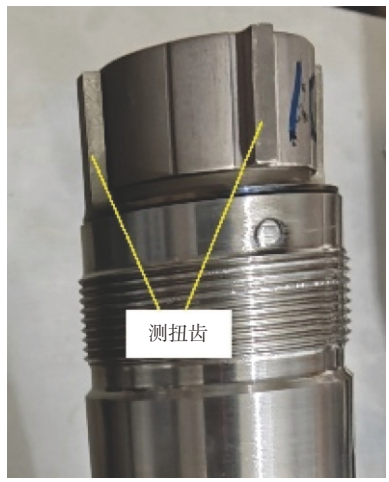


图 1　动力涡轮传动轴上的测扭齿

Fig. 1　Torque-measuring gear on power turbine drive shaft

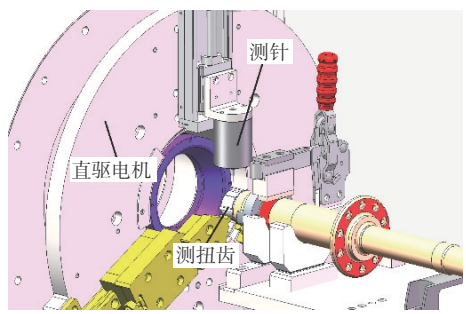


图 2　安装于角度测试台上的测扭齿

Fig. 2　Screenshot of torque-measuring gear and its angle measuring system

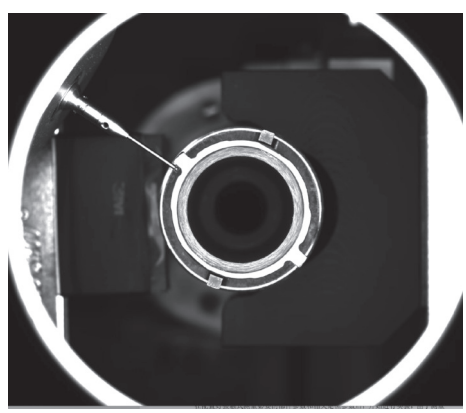


图 3　测针与测扭齿接触时的照片

Fig. 3　Photo of probe contacting with torque-measuring gear

当触发式测针的测球接触测扭齿表面时, 控制电机停机并记录第 1 个测点在磁性编码器中的绝对角度编码 p_1, 而后电机转动使测针与另一测扭齿面接触, 记录第 2 个绝对角度编码 p_2, 这样依次对测扭齿进行 4 次测量, 得到 $p_1 \sim p_4$, 然后, 计算出角度, 图 4 是计算的模型图。设编码器的编码值按照

逆时针方向增加，编码圆的刻度线即编码数目为N。在图4中，跟齿面相切的小圆球代表测针上的测球，测针的位置用其轴线的位置代表。取一个扭齿如T_2的对称线为坐标y轴，电机编码圆的圆心为坐标系圆心O，建立图4所示坐标系，测针对圆心安装，测扭齿T_1和T_2间的待测角度为$\angle T_1OT_2$。

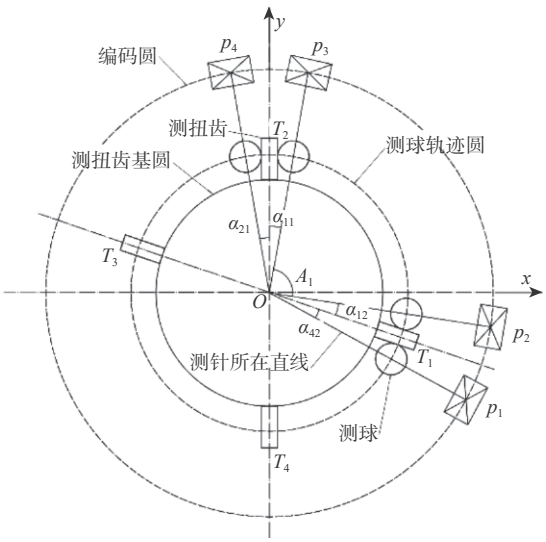


图4 计算模型

Fig. 4 Calculation model

可算出2个测扭齿之间的角度为

$$\angle T_1OT_2=\left(p_3-p_2+\frac{p_4-p_3+p_2-p_1}{2}\right)\frac{360^\circ}{N} \tag{1}$$

2 角度测量准确性的影响分析

2.1 测针安装对角度测量准确性的影响分析

在角度测量过程中，测针的安装不可能正好指向编码圆的圆心，而是存在一定偏心，此种情况的角度测量模型如图5所示。

假设在直驱电机旋转过程中，测针被紧固于电机上，和电机的相对位置保持不变，这样测球在编码圆中的相对位置不变，但测球轨迹仍为与编码圆同心的圆。在图5中测球接触测扭齿位置，测针不偏心时所在直线与编码圆的交点记为$p_{10}\sim p_{40}$。

若测针紧固于电机上，由全等三角形的几何特性可知：$p_1-p_{10}=p_2-p_{20}=p_3-p_{30}=p_4-p_{40}$。

令$p_2-p_{20}=\beta$，则可得

$$\begin{cases}p_2+p_4=p_{20}+p_{40}+2\beta\\ p_1+p_3=p_{10}+p_{30}+2\beta\end{cases}$$

于是，可得

$$\angle T_1OT_2=\left(p_{30}-p_{20}+\frac{p_{40}-p_{30}+p_{20}-p_{10}}{2}\right)\frac{360^\circ}{N}=\left(p_3-p_2+\frac{p_4-p_3+p_2-p_1}{2}\right)\frac{360^\circ}{N} \tag{2}$$

由于式(1)和式(2)相同，因此，仍然可以直接

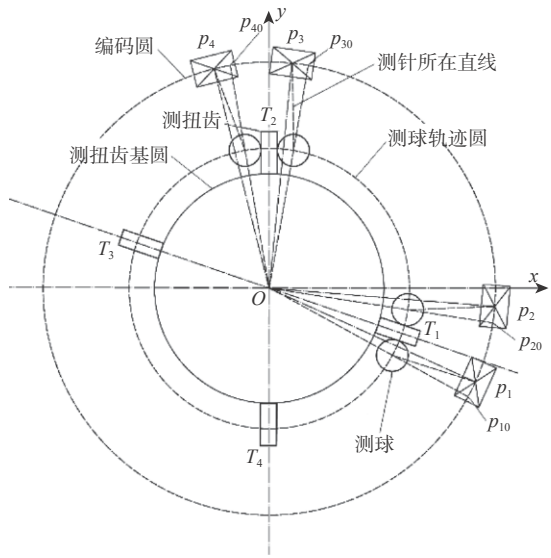


图5 测针安装不对心时测扭齿角度计算模型

Fig. 5 Angle calculation model of torque-measuring gear when probe is installed out of alignment

用测针在编码圆上测得的角度计算出测扭齿的角度；此时测针安装的偏差不会影响测扭齿角度测量的准确性。

2.2 测扭齿和电机安装不同轴对角度测量的影响分析

测扭齿和电机(编码圆)的旋转轴线总是不可能完全同轴，会存在误差，此情况下的测扭齿角度计算模型如图6所示。

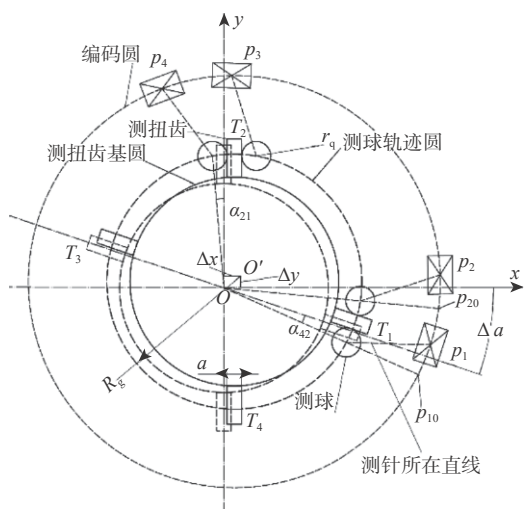


图6 不同轴时测扭齿角度计算模型

Fig. 6 Calculation model for noncoaxial torque-measuring gear

坐标系xOy建立在编码圆的圆心O上，测球运动轨迹圆的圆心也为O，测扭齿基圆的圆心为O'，设O'在xOy中的坐标(即测扭齿安装的偏心量)为$(\Delta x,\Delta y)$。为便于计算，将测扭齿及基圆平移$(-\Delta x,-\Delta y)$，如图6中虚线所示，并记平移后的假想测扭齿记为T_1和T_2，可见，需要测量的测扭齿角度

仍为$\angle T_1OT_2$。

记测扭齿的宽度为 a, 测球半径为 r_q, 测球轨迹圆半径为 R_g, 其他相关角度标示于图 6 中, α 表示测球圆心和测球轨迹圆圆心之间的连线与测扭齿中心线的夹角。2 个测扭齿之间的角度计算式为

$$\begin{cases}\alpha_{21}=\arcsin\dfrac{r_q+a/2-\Delta x}{R_g}\\ \alpha_{42}=\arcsin\dfrac{r_q+a/2-(\Delta x\sin\Delta\alpha+\Delta y\cos\Delta\alpha)}{R_g}\\ \angle T_1OT_2=(p_4-p_1)\dfrac{360^\circ}{N}-\alpha_{21}-\alpha_{42}\end{cases}\tag{3}$$

在测扭齿安装偏心时, 用测针测得齿面角度, 再用式 (1) 计算测扭齿间角度是不准确的, 所带来的测量误差和测球半径、测球轨迹圆半径、测扭齿宽度和测扭齿间角度、测扭齿轴安装的偏心量都有关。

若 $\Delta x=\Delta y=0$, 则 : $(p_4-p_1)\dfrac{360^\circ}{N}-\alpha_{21}-\alpha_{42}=\left(p_3-p_2+\dfrac{p_4-p_3+p_2-p_1}{2}\right)\dfrac{360^\circ}{N}$, 即用式 (3) 和式 (1) 算出的测扭齿角度是一样的。

在一般的扭齿轴安装情形下, $\Delta x\neq\Delta y\neq0$, $(p_4-p_1)\cdot\dfrac{360^\circ}{N}-\alpha_{21}-\alpha_{42}\neq\left(p_3-p_2+\dfrac{p_4-p_3+p_2-p_1}{2}\right)\dfrac{360^\circ}{N}$。例如, 在如图 6 所示的扭齿轴安装情形中, $\Delta x>0,\Delta y>0$, $(p_4-p_1)\dfrac{360^\circ}{N}-\alpha_{21}-\alpha_{42}>\left(p_3-p_2+\dfrac{p_4-p_3+p_2-p_1}{2}\right)\cdot\dfrac{360^\circ}{N}$。即此时用式(1)计算出的测扭齿角度会偏小; 而若要用式(3)准确计算角度, 需要测量测扭齿实际宽度、测球半径、测球轨迹圆半径等几何参数, 以及测扭齿轴线和编码圆轴线间的偏心量, 这些参数的获得都是困难或不可能的。而忽略这些参数的影响、采用简化计算就会带来测量的误差。

3　本文方法

为准确且方便地得到测扭齿间的角度, 本文提出一种基于多余观测的测扭齿角度测量方法, 依次在 8 个齿面间测量角度, 得到 p_1～p_8, 再计算测扭齿间的角度。计算模型如图 7 所示。编码圆的圆心 O 为坐标系原点, 测扭齿基圆圆心为 O', 其坐标为 $(\Delta x,\Delta y)$, 测扭齿与基圆的实际位置用实线表示。为方便计算, 将测扭齿与基圆平移 $(-\Delta x,-\Delta y)$至与编码圆同心的假想位置, 如图 7 中用虚线表示(为视图的清晰, 图中夸大了Δx 和Δy 的值)。测针的测球与各齿面接触时的绝对角度编码值记为 p_1～p_8。2.2 节已经证明了本文测量方式下, 测针安装不对编码圆心不会影响最后算出的测扭齿角度, 所以在本节的推导中不再考虑测针安装不对心的因素。

多余观测的测量方法下, 2 个测扭齿之间的角

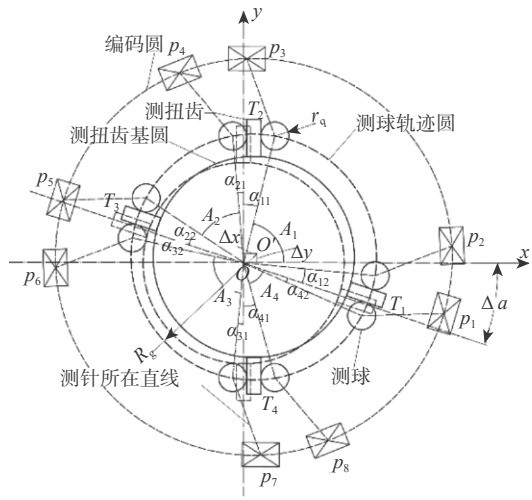


图 7　多余观测的测扭齿角度计算模型

Fig. 7　Angle calculation model of torque-measuring gear for redundant observation

度为

$$\alpha_{11}=\alpha_{41}=\arctan\frac{r_q+a/2+\Delta x}{R_g}$$

$$\alpha_{12}=\alpha_{22}=\arcsin\frac{r_q+a/2+(\Delta x\sin\Delta\alpha+\Delta y\cos\Delta\alpha)}{R_g}$$

$$\alpha_{11}=\alpha_{41}=\arctan\frac{r_q+a/2-\Delta x}{R_g}$$

$$\alpha_{32}=\alpha_{42}=\arcsin\frac{r_q+a/2-(\Delta x\sin\Delta\alpha+\Delta y\cos\Delta\alpha)}{R_g}$$

则 4 个测角分别为

$$\begin{cases}A_1=90^\circ+\Delta\alpha-\alpha_{11}-\alpha_{12}\\ A_2=90^\circ-\Delta\alpha-\alpha_{21}-\alpha_{22}\\ A_3=90^\circ+\Delta\alpha-\alpha_{31}-\alpha_{32}\\ A_4=90^\circ-\Delta\alpha-\alpha_{41}-\alpha_{42}\end{cases}\tag{4}$$

将 α_{11}、α_{12}、α_{21}、α_{22}、α_{31}、α_{32}、α_{41}、α_{42}分别代入式(4), 可以解出

$$\Delta\alpha=\frac{1}{4}(A_1-A_2+A_3-A_4)\tag{5}$$

于是, 可得

$$\angle T_1OT_2=90^\circ+\frac{1}{4}(A_1-A_2+A_3-A_4)=90^\circ+\frac{1}{4}(p_3-p_2+p_5-p_4+p_7-p_6+p_1-p_8+N)\frac{360^\circ}{N}\tag{6}$$

进一步可得

$$\angle T_1OT_2=90^\circ-\frac{1}{4}(p_3-p_2+p_5-p_4+p_7-p_6+p_1-p_8+N)\frac{360^\circ}{N}$$

$$\angle T_3OT_4=\angle T_1OT_2$$

$$\angle T_4OT_1=\angle T_2OT_3$$

此时即使存在测针和测扭齿的安装误差，但是其并不会影响测量计算出的测扭齿间的角度；只是利用测针在不同齿面位置的编码读数，就可以准确地计算出测扭齿间的角度，消除了测针的安装误差、测扭齿轴的安装误差的影响，提高了测扭齿测量角度的准确度。这是利用多余观测消除系统误差的测量平差方法[15]在测扭齿角度测量中的一种应用形式。

对多个测扭齿试件，在图2的测试装置中分别用1.2节中的基本测量方法、本文方法进行了测量，并用精度为2.6 μm的温泽LH108三坐标机的测量结果作为比对，结果如表1所示。

表1 不同测量方法的角度测量结果比较

Table 1 Angle measurement results of different measurement methods

试件编号	测扭齿角度		
	基本方法	本文方法	三坐标机测量
1	90°57′36″	89°59′27″	89°59′26″
2	88°47′16″	89°58′20″	89°58′19″
3	92°9′10″	91°15′00″	91°15′00″

实际测量结果显示，直接测量被测角度的基本方法受到工件和测量装置的安装误差的影响，所得到的角度误差较大；本文方法提高了测量结果的准确性。

通过对比测量式(3)和式(6)所对应的合成标准不确定度计算式可以发现，式(3)的不确定度大于式(6)的不确定度；前者由于图6中α_{21}和α_{42}的存在使得不确定度稍大，本文测量方法的不确定度要小。采用本文的测量方法得到的角度和三坐标机测得的角度更接近、测量结果更准确。

经过计算与分析，并比较图7和图3的测量可以得出进一步推广的结论：对于旋转式的角度测量，通过将被测对象旋转90°或180°而增加测量次数，可以消除安装误差的影响而提高测量角度的准确度。

4 结 论

1）在基本的测量方式中，测扭齿的角度测量结果受测针和测扭齿安装误差的影响。

2）测扭齿间的角度可以通过多余观测来测量更多齿面间的角度，然后计算得出，本文方法得到的角度不受测针不对心、测扭齿和编码圆不同轴等安装误差的影响，降低了安装调试的难度，同时测量的准确度更高。

3）测具的几何尺寸、测具和工件的安装误差会影响角度的测量准确性；利用多余的观测，可以消除这些因素对测量结果的影响，提高测量的准确性。

参考文献（References）

[1] 陈竞炜, 徐鲁兵, 蒋晓伟. 涡轴发动机的输出扭矩测量装置: 中国, CN102607751B[P]. 2014-06-11.
CHEN J W, XU L B, JIANG X W. Output torque measurement device for turboshaft engines: China, CN102607751B[P]. 2014-06-11(in Chinese).

[2] 洪焱, 沈小林, 田野, 等. 基于自由转自的纱线扭矩的检测方法[J]. 纺织学报, 2021, 42(10): 61-66.
HONG Y, SHEN X L, TIAN Y, et al. Method for detecting yarn torque based on free rotation[J]. Journal of Textile Science, 2021, 42(10): 61-66(in Chinese).

[3] 杨坚, 廖东华, 高松, 等. 大鼠食道扭转生物力学特性[J]. 航天医学与医学工程, 2004, 17(2): 126-129.
YANG J, LIAO D H, GAO S, et al. Biomechanical torque properties of rat esophagus[J]. Space Medicine & Medical Engineering, 2004, 17(2): 126-129(in Chinese).

[4] 贾敏, 杨嘉亮, 王涛, 等. 角度和扭矩同步测量装置的研究[J]. 中国机械, 2020(8): 128-130.
JIA M, YANG J L, WANG T, et al. Research on synchronous measurement device for angle and torque[J]. China Machinery, 2020(8): 128-130(in Chinese).

[5] 王聚存, 周斌, 李孝锋. 测扭齿夹角调整工装: 中国, CN203981319U[P]. 2014-12-03.
WANG J C, ZHOU B, LI X F. Torsion gear included angle adjusting tool: China, CN203981319U[P]. 2014-12-03(in Chinese).

[6] 徐敏, 高文玲. 用三坐标测量机间接测量空间角度及误差分析[J]. 计量技术, 1993, 37(3): 16-19.
XU M, GAO W L. Indirect measurement of spatial angle with CMM and error analysis[J]. Measurement Technique, 1993, 37(3): 16-19(in Chinese).

[7] 彭希锋, 陈爽, 李海星, 等. 基于激光位移传感器的面角度测量技术研究[J]. 仪器仪表学报, 2017, 38(11): 2735-2743.
PENG X F, CHEN S, LI H X, et al. Study on the plane angle measurement technology based on aser displacement sensors[J]. Chinese Journal of Scientific Instrument, 2017, 38(11): 2735-2743(in Chinese).

[8] 钟亦林, 马有为, 丁国清, 等. 非接触三点内径测量法中角度安装误差的校准方法研究[J]. 仪表技术与传感器, 2023(4): 107-112.
ZHONG Y L, MA Y W, DING G Q, et al. Calibration of angle setting error in non-contact inner diameter measurement by three-point method[J]. Instrument Technique and Sensor, 2023(4): 107-112(in Chinese).

[9] 吴常铖, 杨德华, 曹青青, 等. 基于视觉和最大长度序列的旋转角度测量系统[J]. 中国测试, 2022, 48(1): 122-128.
WU C C, YANG D H, CAO Q Q, et al. Rotation angle measurement system based on visual and maximum length sequence[J]. China Measurement & Test, 2022, 48(1): 122-128(in Chinese).

[10] 王馨悦, 徐志伟. 基于双目视觉的可变后掠角机翼角度测量技术[J]. 国外电子测量技术, 2022, 41(11): 114-119.
WANG X Y, XU Z W. Variable swept wing angle measurement technology based on binocular stereo vision[J]. Foreign Electronic

Measurement Technology, 2022, 41(11): 114-119(in Chinese).

[11] 邹炎升, 张乃元, 袁辉, 等. 异形光斑的高精度光束偏转角度测量[J]. 中国空间科学技术, 2022, 42(1): 107-113.
ZOU Y S, ZHANG N Y, YUAN H, et al. High precision deflection angle measurement of irregular light spot[J]. Chinese Space Science and Technology, 2022, 42(1): 107-113(in Chinese).

[12] 周必雷, 李荣锋, 曾丽, 等. 一种主瓣干扰下基于滑窗子阵的稀疏测角方法[J]. 北京航空航天大学学报, 2023, 49(7): 1623-1629.
ZHOU B L, LI R F, ZENG L, et al. A sparse estimation method for radar target direction with sliding-window subarray configuration in mainlobe jamming[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(7): 1623-1629(in Chinese).

[13] 宋栓军, 王启宇, 贾秀海, 等. 基于斜齿轮端面与侧面图像处理的螺旋角测量方法[J]. 西安工程大学学报, 2021, 35(3): 81-85.
SONG S J, WANG Q Y, JIA X H, et al. Helix angle measurement method based on end face and side image processing of helical gear[J]. Journal of Xi'an Polytechnic University, 2021, 35(3): 81-85 (in Chinese).

[14] CHEN X, HE X R, ZHANG Y, et al. Angle measurement method based on cross-orthogonal hall sensor[C]//Proceedings of the IEEE 16th International Conference on Electronic Measurement & Instruments. Piscataway: IEEE Press, 2023: 153-158.

[15] 金日守, 戴华阳. 误差理论与测量平差基础[M]. 北京: 测绘出版社, 2011.
JIN R S, DAI H Y. The basis of error theory and surveying adjustment[M]. Beijing: Sino Maps Press, 2011(in Chinese).

A contact angle measurement method for torsional teeth

TONG Xiaoyan[1,*], LU Dongsheng[2], BAI Hui[3], CHANG Long[4]

(1. School of Astronautics, Beihang University, Beijing 100191, China;
2. Aircraft Design and Research Institute, Aviation Industry Harbin Aircraft Industry Group Co., Ltd., Harbin 150060, China;
3. Aecc South Industry Company Limited, Zhuzhou 412002, China;
4. Shanghai Aerospace System Engineering Institute, Shanghai 201108, China)

Abstract: The angle measurement of torsional teeth is commonly used in the manufacture and performance test of a helicopter power system. The impact of their coupling on the measurement of torsional teeth angle is examined, as is the influence of installation error caused by the non-alignment between the contact measuring probe and the direct drive motor during the measurement of torsional teeth angle, the error caused by the non-concentricity between the measuring probe and the motor during the installation of torsional teeth, and a contact measuring scheme for measuring torsional teeth angle using a magnetic encoder. A method of angle measurement using redundant observations is proposed and verified by practical angle measurement, which eliminates the influence of the installation errors of torsional teeth and probe on the measurement results, and deduces the angle calculation formula. The results show that the proposed method not only reduces the installation and debugging requirements of the measuring system but also improves the accuracy of angle measurement for the torsional teeth, making it suitable for similar precise angle measurements.

Keywords: angular measurement; contact measurement; installation error; torque-measuring gear; helicopter transmission system

Received: 2023-11-03; **Accepted**: 2023-11-13; **Published Online**: 2023-11-20 15:34
URL: link.cnki.net/urlid/11.2625.V.20231117.1420.002

* **Corresponding author**. E-mail: tongxiaoyan@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0715

基于应变能理论的全陶瓷球轴承细观缺陷识别

白晓天，张赵南，石怀涛*，张啸尘
(沈阳建筑大学 机械工程学院，沈阳 110168)

摘 要：针对全陶瓷球轴承外圈内部存在亚毫米级甚至毫米级的细观缺陷，传统钢制轴承外圈缺陷动力学模型无法表征的问题，基于应变能理论建立针对全陶瓷球轴承外圈亚表面细观缺陷识别的动力学模型，研究细观缺陷尺度下不同缺陷深度对全陶瓷球轴承外圈运行状态的影响，通过仿真，证实了当轴承外圈亚表面存在细观缺陷时，振动时域信号中会存在小波峰现象，在频域信号中转频附近存在与其对应的特征频率，以经验模态分解(EMD)三阶分量幅值正向最大值与其对应的转频峰值的比值判断缺陷演化程度，通过实验验证了所建模型的有效性。所建模型实现了对全陶瓷球轴承外圈是否存在亚表面细观缺陷及缺陷演化程度的判断，为全陶瓷球轴承的缺陷诊断提供了新的思路，为全陶瓷轴承转子系统的安全稳定运转提供理论参考。

关 键 词：全陶瓷球轴承；细观缺陷；动力学建模；应变能理论；经验模态分解
中图分类号：V229+.2; TH133
文献标志码：A **文章编号**：1001-5965(2026)01-0110-10

全陶瓷球轴承是许多回转设备的重要部件，其运行状态对整个系统来说至关重要[1]。如在航空发动机、食品加工机械、人工置换髋关节及印刷机械领域中，由于零磨粒的需求，具有硬度高、耐磨损等特点的全陶瓷球轴承更受青睐。然而，全陶瓷球轴承多采用热等静压工艺烧结而成，轴承各部件内部存在亚毫米级甚至毫米级的亚表面细观尺度初始缺陷等[2]，在不均匀应力场内极易扩展并形成宏观剥落缺陷，会对设备造成重大影响。同时，由于全陶瓷球轴承的缺陷位置与破坏机理与钢制轴承均不同，传统钢制轴承外圈缺陷动力学模型不能直接移植应用在全陶瓷球轴承上面。考虑到陶瓷材料对局部损伤更为敏感，因此，对全陶瓷球轴承亚表面缺陷识别进行研究意义重大。

而对于轴承各部件缺陷的研究，国内外学者大多采用建立轴承动力学模型的方法。牛蔺楷等[3]提出一种考虑润滑牵引等多种力学因素的滚动轴承局部表面缺陷的动力学模型构建方法，并在时域和频域中对动力学响应进行了分析。关贞珍等[4]考虑轴承间隙及滚珠与滚道的非线性接触力等因素，建立了四自由度非线性振动模型用于轴承缺陷的研究。对于钢制轴承亚表面缺陷，Deng等[5]研究滚动接触疲劳下的次表面疲劳裂纹扩展，基于应力强度因子计算，评估疲劳裂纹扩展情况，Liu等[6]提出一个有限元模型来研究水平和倾斜的次表面裂纹对滚子轴承接触特性的影响，得到了接触变形与裂缝尺寸(长度、深度和角度)之间的关系。Danielsen等[7]研究了陶瓷轴承上的白色蚀刻裂纹的破坏模式，并进行了测试，表明应力在缺陷区附近具有显著差异。文献[8-10]对全陶瓷球轴承的裂纹进行了深入探讨，认为裂纹是由设备运行过程中滚动体与轴承套圈之间的接触应力引起的。但是，研究中没有考虑到滚动体在通过外圈缺陷位置的过程中对外圈施加的弯曲应力，而弯曲应力也可能是内部

收稿日期：2023-11-02；录用日期：2024-03-22；网络出版时间：2024-03-28 13:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20240327.1033.001
基金项目：国家自然科学基金(52275119,52075348,52005352)；辽宁省教育厅基金(LJKZZ20220078)
*通信作者. E-mail：sht@sjzu.edu.cn

引用格式：白晓天，张赵南，石怀涛，等. 基于应变能理论的全陶瓷球轴承细观缺陷识别[J]. 北京航空航天大学学报，2026，52(1)：110-119.
BAI X T，ZHANG Z N，SHI H T，et al. Identification of mesoscopic fault of full ceramic ball bearings based on strain energy theory[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：110-119 (in Chinese).

缺陷产生和扩展的原因。文献 [11-12] 对接触应力下的裂缝形成机制进行了深入分析。在断裂力学原理的基础上, 通过计算裂纹引起的应变能, 可以分析出弯矩引起的裂纹对刚度的影响[13-15]。胡爱军等[16] 提出时变接触变形与时变刚度耦合的滚动轴承缺陷非线性动力学模型, 得到了不同缺陷尺寸下滚动轴承缺陷冲击振动响应特征变化规律。考虑到全陶瓷球轴承外圈破坏机理与钢轴承存在差异, 现有的识别滚动轴承外圈的局部缺陷的数据分析方法也需要改进, 这些方法只能识别外圈内滚道的表面缺陷, 难以准确识别全陶瓷球轴承外圈亚表面细观缺陷是否存在。

结合上述学者的研究, 参考钢制轴承故障动力学模型及断裂力学和应变能理论, 本文提出一种针对全陶瓷球轴承外圈亚表面细观缺陷的非线性动力学模型, 可用于分析外圈在细观缺陷下轴承的振动响应。该模型引入脆性材料断裂力学理论与应变能理论, 基于不同缺陷深度下应变能变化趋势分析轴承动刚度变化情况, 获取轴承外圈缺陷径向深度对轴承动态特性影响情况, 从而准确识别全陶瓷球轴承外圈亚表面细观缺陷是否存在及存在的尺度范围。

1　包含亚表面细观缺陷动力学模型

1.1　外圈细观缺陷处刚度削弱系数与应变能释放率

与钢制轴承和混合陶瓷轴承外圈不同, 全陶瓷球轴承外圈由于制造工艺原因, 在其内部存在微米级甚至毫米级的亚表面细观缺陷, 在不均匀应力场内极易发生扩展, 影响轴承正常运转, 如图 1 所示。

结合轴承外圈的载荷分布, 每个亚表面缺陷都位于脆性材料介质的椭圆区域范围内。由于制造工艺引起的亚表面缺陷可分为 2 种情况: ①导致材料开裂的横向扩展气孔缺陷; ②导致刚度减弱的纵向扩展气孔缺陷[17]。本文主要讨论了气孔缺陷引起的全陶瓷球轴承外圈刚度削弱情况, 主要针对纵向缺陷进行研究。由于亚表面细观缺陷, 特别是纵向扩展缺陷, 不能直接观察和测量, 因此, 可以省去对缺陷尺寸直接的量化过程, 取而代之利用刚度削弱系数 μ_c 来代替出现不同纵向缺陷扩展深度对轴承系统动态特性的影响, 计算式为

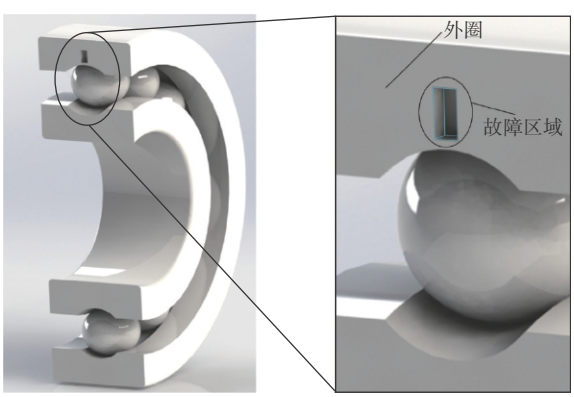


图 1　外圈亚表面缺陷和局部截面示意图

Fig. 1　Schematic diagram of subsurface fault and local section of outer ring

$$EI_C = \mu_c EI \tag{1}$$

式中: E 为陶瓷材料的弹性模量; I 为轴承外圈截面惯性矩; I_C 为缺陷轴承外圈截面惯性矩。

结合不同缺陷深度对轴承外圈刚度的削弱不同, 可以得到细观缺陷的复合刚度削弱系数 μ_c 为

$$\mu_c = \left[\frac{1}{(1/\mu_1)^{2/3} + (1/\mu_2)^{2/3}}\right]^{3/2} \tag{2}$$

式中: μ_1 和 μ_2 分别为全陶瓷球轴承外圈仅含裂纹和仅含气孔缺陷时的刚度削弱系数, 其中, $\mu_1 = EI_{C1}/EI$, $\mu_2 = EI_{C2}/EI$, I_{C1} 和 I_{C2} 分别表示全陶瓷球轴承外圈仅含裂纹和仅含气孔缺陷时的截面惯性矩。

根据陶瓷材料断裂力学及应变能理论, 由于不同缺陷深度对于轴承外圈接触刚度的削弱不同, 考虑到轴承座对轴承的支承作用, 忽略外圈拉应力的影响, 本文假设亚表面细观缺陷产生的主要原因是弯曲载荷, 施加的弯矩不仅会促进细观缺陷以尖端裂纹方式发生扩展趋势[18], 而且会增加结构的应变能。

由于在纯弯矩作用下, 本文提及的引发裂纹扩展趋势均为张开型裂纹, 即 I 型裂纹。全陶瓷球轴承外圈在外部附加载荷的作用下, 如图 2 所示, 其应变能释放率可以表示为

$$S = K_I^2/E \tag{3}$$

式中: K_I 为开口位移法计算出的裂纹的强度系数, K_I 可以表示为

$$K_I = M_e S_b \sqrt{\pi d/Q} \tag{4}$$

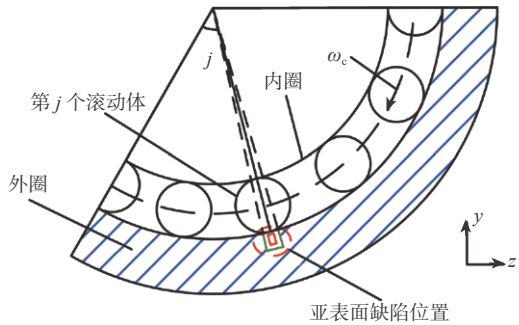
其中: M_e 为表面校正系数; S_b=3M/ba^2, a 为缺陷顶端至外圈外表面长度, b 为缺陷中心至外圈侧面的长度, M 为表面所受弯矩; d 为亚表面缺陷径向深度; Q 为缺陷的形状系数, 其表达式为

$$Q = 1.464 + (a/l)^{1.65} \tag{5}$$

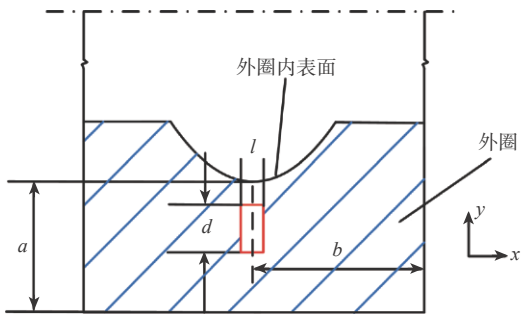
其中: l 为亚表面缺陷的宽度。

1.2　全陶瓷球轴承外圈接触载荷

本文将全陶瓷球轴承在运行过程中所承受的载荷设置为径向载荷, 该载荷通过滚动体在轴承的内圈和外圈之间传递, 滚动体与外圈之间近似满足



(a) 径向截面示意图



(b) 轴向界面示意图

图2 外圈亚表面细观缺陷处局部截面示意图

Fig. 2 Schematic diagram of local cross-section at sub surface fault of the outer ring

赫兹接触理论，在应用赫兹接触理论前做出如下假设：①全陶瓷球轴承在运行过程中各部件之间的接触变形仅为弹性变形；②全陶瓷球轴承各部件之间的相互作用力仅与接触面垂直；③与每个部件的曲率半径相比，相互接触的全陶瓷球轴承部件之间的接触面的尺寸可以忽略不计。工业陶瓷材料表面接触引起的弹性变形很小，但全陶瓷球轴承各部件之间的接触变形仍在弹性变形范围内，因此，赫兹接触理论可以用来分析全陶瓷球轴承的动态特性，其缺陷形貌如图3所示。

基于赫兹点接触理论，接触力是由每个滚动体上的接触分力f_x、f_y累积求得

$$f_x=\sum_{j=1}^{N_b}G(\delta_j)Q_j\cos\theta_j \tag{6}$$

$$f_y=\sum_{j=1}^{N_b}G(\delta_j)Q_j\sin\theta_j \tag{7}$$

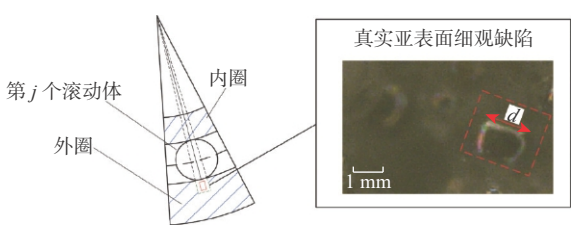


图3 外圈亚表面细观缺陷处局部示意图

Fig. 3 Partial schematic diagram of subsurface defects on outer ring

式中：x、y分别表示轴承的轴向和径向方向；N_b为滚动体个数；$G(\cdot)$为Heavisde函数，正参数值取值为1，负参数值或零参数值取值为0；θ_j为第j个滚动体的角位置。

将滚动体与轴承外圈的受力建模为

$$Q_j=K_c\delta_j^n \tag{8}$$

式中：K_c为接触刚度；n为接触变形系数，对于滚动轴承一般选取n=1.5[19]；δ_j为

$$\delta_j=\sqrt{\left(x\cos\theta_j\right)^2+\left(y\sin\theta_j\right)^2}-\delta_0 \tag{9}$$

式中：δ_0为全陶瓷球轴承径向间隙。

假设滚动体与内外滚道之间没有滑动摩擦，滚动体是可变形的，滚动体的角位置可表示为

$$\theta_j=\omega_c t+2\pi(j-1)/N_b \tag{10}$$

式中：ω_c为保持架转速，保持架转速与转子速度ω_s有关，而ω_s由全陶瓷球轴承的组成结构有关，ω_c可表示为

$$\omega_c=B\omega_s \tag{11}$$

式中：B=D_i/(D_i+D_o)，对于全陶瓷球轴承，D_i为轴承内圈直径，D_o为轴承外圈直径。

当全陶瓷球轴承沿垂直方向受到力时，力的作用方向是垂直向下的。考虑到本研究中选择的全陶瓷球轴承型号为6304，在径向载荷下，滚动体对轴承上部的作用力可以忽略，轴承外圈分布载荷如图4所示。

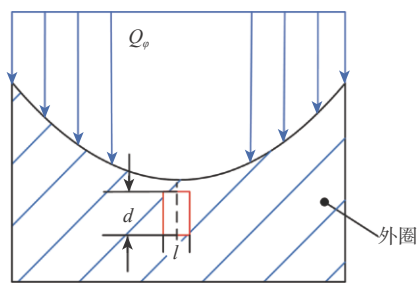


图4 轴承外圈分布载荷示意图

Fig. 4 Schematic diagram of load distribution on outer ring of bearing

轴承外圈上的载荷分布可以表示为

$$Q_\varphi=\begin{cases}Q_{\max}\left[1-\dfrac{1}{2k_L}(1-\cos\varphi)\right]^{3/2} & \varphi\in\varphi_{load}\\ 0 & \text{其他}\end{cases} \tag{12}$$

式中：Q_{max}为最大载荷分布密度；φ为载荷区域中的任意位置角；k_L为载荷分布系数，可表示为

$$k_L=\frac{1}{2}\left(1-\frac{\delta_0}{2\delta_{max}}\right) \tag{13}$$

式中: δ_{max} 为轴承的最大径向偏移。

1.3　全陶瓷外圈缺陷位置的时变刚度与动力学模型

根据陶瓷材料的应变能理论, 假设作用在亚表面缺陷处的外部作用力矩不变, 外圈上产生的应变能 W 可分解无缺陷外圈的应变能和细观缺陷扩展产生的应变能之和[20], 可表示为

$$W = N_c + \Delta U = 2\Delta U \tag{14}$$

$$N_c = \Delta U \tag{15}$$

式中: N_c 为细观缺陷扩展释放的能量; ΔU 为外圈出现缺陷时弹性应变能的增加量, 因此, 陶瓷轴承外圈出现缺陷时的最终应变能为

$$U_c = U + \Delta U = U + N_c \tag{16}$$

式中: U 为全陶瓷球轴承外圈在无缺陷情况下的应变能。

对于弯曲刚度为 EI 的无缺陷轴承外圈, 受到纯弯矩 M 的作用时, 轴承外圈的应变能为

$$U = \frac{1}{2}\int \frac{M^2}{EI}\mathrm{d}x \tag{17}$$

根据载荷分布, 可以计算出亚表面细观缺陷位置的弯矩:

$$M = \int Q_\varphi r\cos\varphi \mathrm{d}\varphi \tag{18}$$

根据断裂力学的考虑, 缺陷位置处的扩展所释放的能量:

$$N_c = \int S\mathrm{d}A \tag{19}$$

结合式 (14)～式 (19), 外圈亚表面细观缺陷处的时变刚度与外环无裂纹位置的刚度之间的关系为

$$\frac{1}{2}\int \frac{M^2}{EI_C}\mathrm{d}x = \frac{1}{2}\int \frac{M^2}{EI}\mathrm{d}x + \int S\mathrm{d}A \tag{20}$$

根据结构动力学, 弯矩 EI 相当于外圈的接触刚度

$$K_o = f(EI) \tag{21}$$

因此, 可得

$$K_c = \mu_c K_o \tag{22}$$

结合式 (20)～式 (22) 可以得到细观缺陷的实际接触刚度:

$$K_c = \left[\frac{1}{(1/\mu_1 K_o)^{2/3} + (1/\mu_2 K_o)^{2/3}}\right]^{3/2} \tag{23}$$

本文所用的二自由度质量-弹簧转子-轴承系统模型是 Sunnersjö提出的[21]。转子被简化为中心的集中质量点, 两端有 2 个相同的轴承。假定轴承外圈和内圈分别通过过盈配合固定在刚性支承中并与转轴刚性固定。全陶瓷球轴承简化为基于赫兹接触的弹簧和阻尼系统。忽略轴承座和基础的自然挠度及转轴的弯曲挠度, 忽略陀螺力矩、离心力和轴承部件之间的润滑牵引力等。将轴承设置为垂直放置, 承受垂直向下的径向力, 并且径向力的作用线穿过轴承的中心。滚动体与陶瓷外圈之间的接触变形可以看作是一种完全弹性变形的状态, 因此, 赫兹接触理论仍然适用于全陶瓷球轴承外圈的建模。根据载荷分布和运行时滚动体与外圈之间的接触变形, 根据赫兹接触理论, 得到了全陶瓷球轴承运动时的广义接触力, 如图 5 所示。结合缺陷外圈应变能释放率、载荷分布和时变刚度[22], 建立了二自由度非线性动力学模型:

$$m\begin{bmatrix}\ddot{x}\\ \ddot{y}\end{bmatrix} + c\begin{bmatrix}\dot{x}\\ \dot{y}\end{bmatrix} + \begin{bmatrix}\sum_{j=1}^{N_b} G(\delta_j)Q_j\cos\theta_j\\ \sum_{j=1}^{N_b} G(\delta_j)Q_j\sin\theta_j\end{bmatrix} = \begin{bmatrix}0\\ F_r\end{bmatrix} \tag{24}$$

式中: m 为全陶瓷球轴承外圈质量; c 为陶瓷材料阻尼系数; f_x、f_y 分别为 x、y 方向上的接触载荷; F_r 为外部载荷。结合不同缺陷深度对接触刚度的削弱不同, 量化缺陷各阶段的权重, 将接触刚度 K_c 代入到二自由度动力学模型中, 即可得到全陶瓷球轴承外圈亚表面细观缺陷振动模型, 技术流程如图 6 所示, H 为相对峰值高度, e 为相对深度。

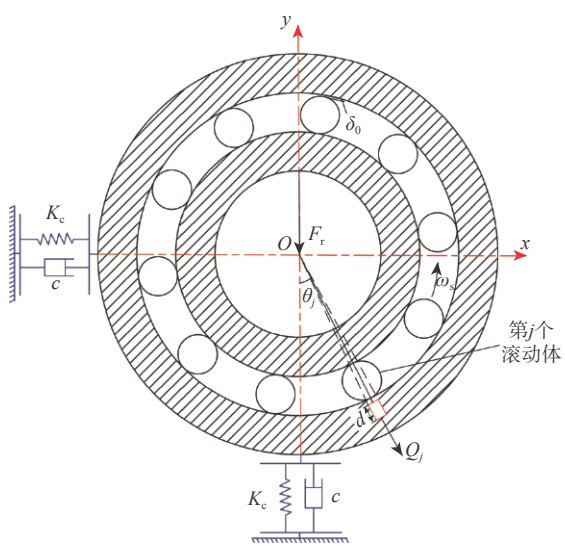


图 5　二自由度非线性动力学模型示意图

Fig. 5　Schematic diagram of a two-degree-of-freedom nonlinear dynamic model

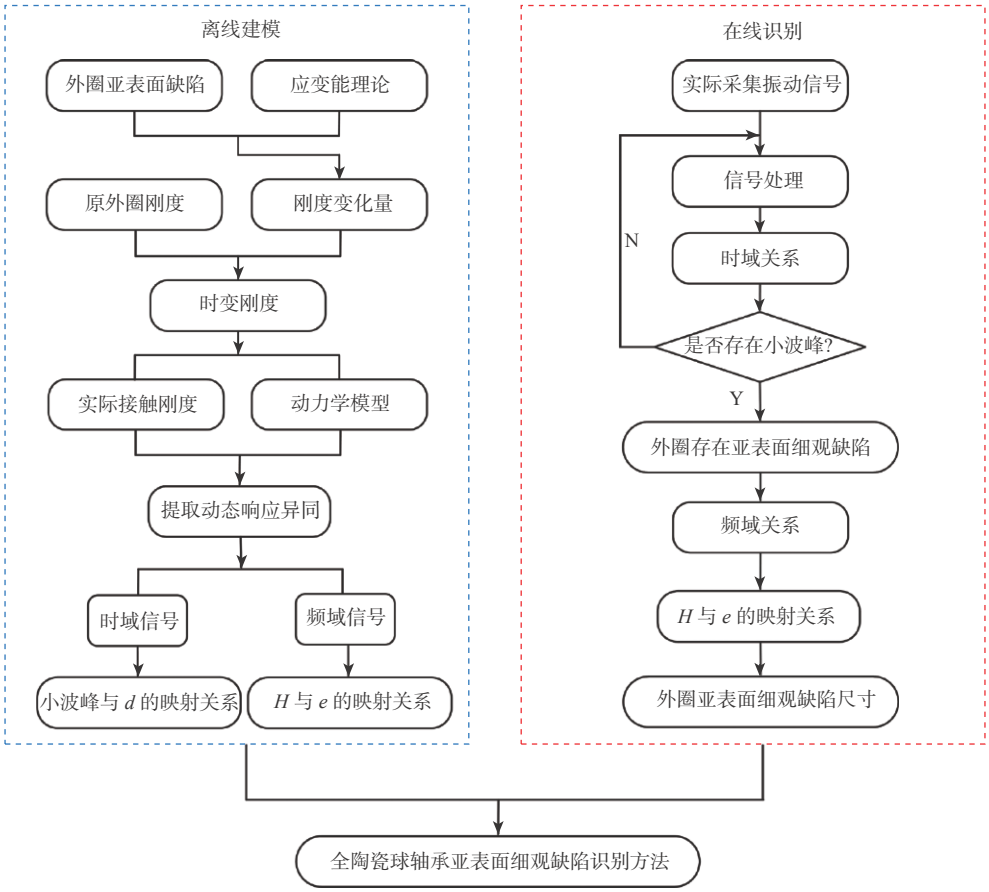


图 6 全陶瓷球轴承外圈亚表面细观缺陷研究技术流程图

Fig. 6 Technical flow chart of research on subsurface meso-fault of outer ring of full ceramic ball bearing

2 数值模拟

模型模拟计算选择全陶瓷 6304 深沟球轴承，轴承主要承受径向载荷 F_r=300 N，轴承外圈缺陷径向深度 d=0～3.6 mm，步长为 0.05 mm。使用四阶 Runge-Kutta 算法来求解二自由度动力学微分方程，分析全陶瓷球轴承外圈缺陷在不同径向深度下振动位移响应变化情况。

根据式 (1) 和式 (2) 可以求得刚度削弱系数变化曲线，为更加直观的量化刚度削弱系数的变化情况，可以用其变化率的绝对值$|h|=|\Delta\mu/\Delta d|$来表征，如图 7 所示。

轴承外圈刚度削弱系数曲线整体呈现出缓慢下降-急剧下跌-趋于平稳的变化趋势。在 d=0.3～1.0 mm 阶段，由于轴承外圈开始出现细观缺陷，相较于正常无缺陷外圈，细观缺陷的出现，对轴承外圈的刚度开始出现削弱，刚度削弱曲线出现明显下降趋势；在 d=1.0～3.0 mm 阶段，由于细观缺陷沿外圈径向深度不断增大，对刚度削弱情况越发明显，刚度削弱曲线整体大致呈现出线性下降的趋势；在 d=3.0～3.6 mm 阶段，此阶段缺陷的变化对刚度的削弱依然存在，只是在此阶段削弱的程度稍降低，但缺陷对外圈刚度的削弱依旧存在，削弱曲线依旧呈下降的趋势。

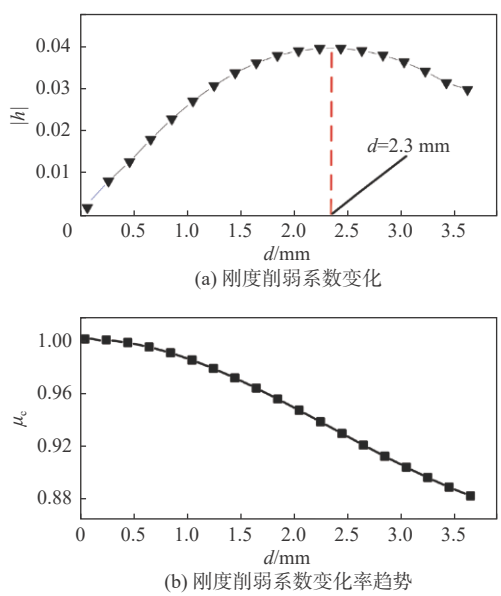


图 7 刚度削弱系数及其变化率趋势

Fig. 7 Change in stiffness weakening coefficient and trend of change rate

整体来看，随着径向深度的不断增大，刚度削弱系数的变化率先增大后减小，表明随着缺陷深度

的不断增大, 亚表面缺陷对轴承外圈刚度的削弱表现出先增大后降低的趋势, 但整体来看削弱曲线依旧呈现出下降的趋势, 当 d 达到 2.3 mm 处时, 刚度削弱变化率达到峰值, 随即开始减弱。

根据式 (21)～式 (23) 可以求得实际接触刚度变化曲线, 为更加直观的量化实际接触刚度的变化情况, 可以用其变化率的绝对值$|i|=|\Delta K_c/\Delta d|$来表征, 如图 8 所示。

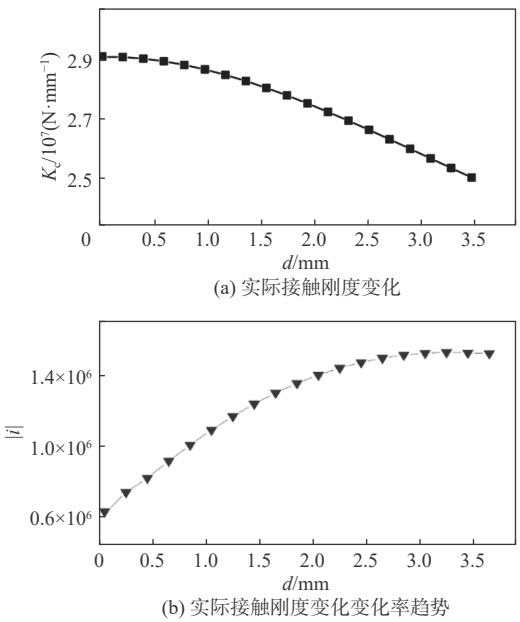


图 8　实际接触刚度及其变化率趋势

Fig. 8　Change in actual contact stiffness and trend of change rate

轴承外圈实际接触刚度曲线整体呈现先缓慢下降后急剧下跌的变化趋势。在 d=0.3～1.0 mm 阶段, 由于全陶瓷球轴承外圈开始出现小尺寸细观缺陷, 相比于正常无缺陷轴承, 这种从无到有的过程导致对轴承外圈接触刚度有较大程度削弱, 实际接触刚度曲线在一开始呈现出明显下降趋势; 在 d=1.0～3.6 mm 阶段, 亚表面细观缺陷继续沿外圈径向深度增大, 导致实际接触刚度不断变小, 曲线大致呈现出线性下降趋势。整体来看, 随着缺陷深度的不断增大, 实际接触刚度呈现下降趋势, 说明滚动体与外圈之间在径向力的作用下, 抵抗变形的能力不断减小。

通过求解二自由度动力学微分方程, 可以得到最终的全陶瓷球轴承系统的动态响应, 选取 3 个具有代表性的特征点进行详细分析。

通过仿真计算得到全陶瓷球轴承在无外圈缺陷时的振动时域信号及外圈存在缺陷时的振动时域信号, 由于外圈亚表面细观缺陷的存在, 导致其接触刚度变小, 滚动体与轴承外圈内滚道的接触力发生波动, 所以在缺陷区域内会出现一个额外的附加位移, 在时域信号中体现在最大值与最小值之间的极小值附近出现了一个信号波峰, 其呈现出周期性冲击现象, 如图 9 所示。

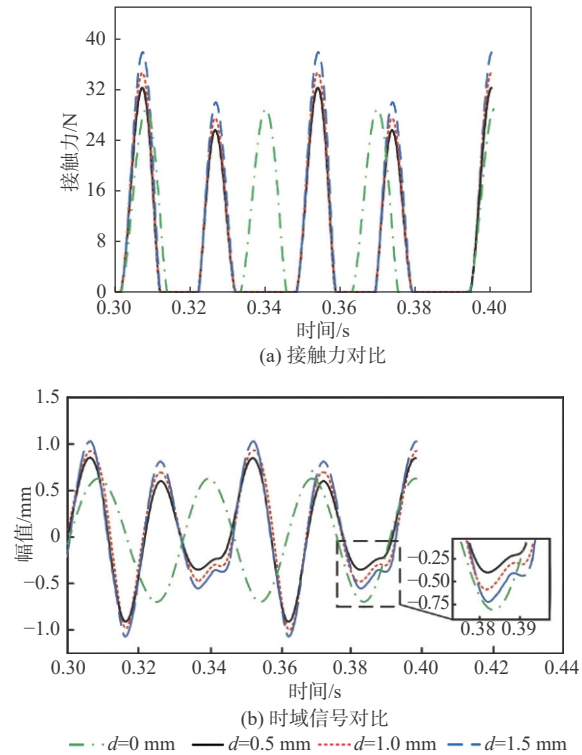


图 9　模拟无缺陷与存在缺陷时接触力、时域信号对比

Fig. 9　Comparison of contact force and time-domain signal between simulated defect-free and defect-present

模拟无缺陷情况下的时域信号波形近似于正弦函数, 其滚动体与轴承外圈内滚道之间的接触力是周期性且等幅值波动, 而模拟缺陷信号在细观缺陷处的时域信号中则显示出明显的小波峰周期性冲击现象, 其滚动体与轴承外圈内滚道的接触力是周期性且幅值不等的波动, 当缺陷深度增大时, 轴承振动响应的信号增加更多, 这表明缺陷深度对全陶瓷球轴承外圈的运动具有更显著的影响, 因此, 可以将时域信号中有无小波峰出现作为全陶瓷球轴承外圈是否出现亚表面细观缺陷的评价标准。

但由于缺陷信号的小波峰现象只能在时域中粗略地看到, 因此, 为对全陶瓷球轴承外圈缺陷位置处进行更准确的定量分析, 对图 8 中仿真计算得到的缺陷时域信号进行傅里叶变换及经验模态分解 (empirical mode decomposition, EMD), 得到全陶瓷球轴承外圈缺陷位置处的频域信号和第三阶模态分解, 如图 10 和图 11 所示。

由于全陶瓷球轴承的外圈存在亚表面细观缺陷, 导致其外圈刚度削弱, 滚动体与外圈内滚道接触力波动, 出现附加位移, 导致振动信号中存在小波峰现象, 经过傅里叶变换转到频域内, 小波峰处

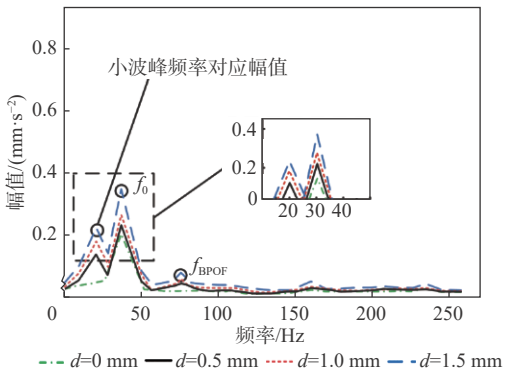
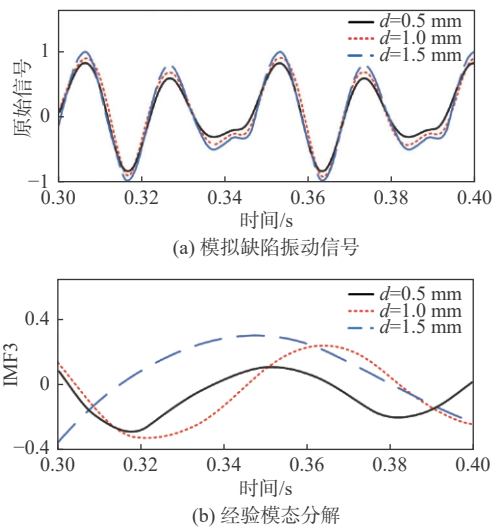


图 10 模拟缺陷振动信号频域

Fig. 10 Frequency domain of simulated fault vibration signal



(a) 模拟缺陷振动信号

(b) 经验模态分解

图 11 模拟缺陷振动信号和经验模态分解

Fig. 11 Simulated defect vibration signal and empirical mode decomposition

对应出现一个峰值。如图 10 所示,其中,f_0 为转频,f_{BPOF} 为轴承外圈缺陷频率,在转频附近出现一个较高幅值,并且随着缺陷深度不断增大,缺陷频域信号中小波峰频率所对应的幅值发生增加。

通过对振动时域信号进行 EMD,其存在三阶分量,本征模函数 (intrinsic mode function, IMF) 分别代表了原始信号中的各频率分量,并按照从高频到低频的顺序依次排列,IMF3 在频率分量中属最低频率,其波峰幅值与频域中新频率的幅值相符,因此,IMF3 对应时域信号中小波峰处频率,将存在不同亚表面缺陷尺寸的时域信号分别进行 EMD,IMF3 分量如图 11 所示。

有量纲指标在进行分析时,得到的结果不仅与被测轴承的运行状态(是否存在亚表面缺陷)有关,而且与被测轴承的运行工况(转速、负载)有关。相反无量纲特征参数只与被测轴承的状态相关,对转速、负载等运行参数的变化不敏感。同时,在时域内小波峰处振幅与信号的峰值存在一定比例关系,经过傅里叶变化后,在频域内这种比例关系仍然存在,因此,取三阶分量幅值正向最大值与其对应的转频峰值的比值 H 为纵坐标,取缺陷深度与轴承外圈厚度 (约 4 mm 厚) 的比值 e 为横坐标,其中,$e=d/a$,建立相对峰值高度 H 与相对深度 e 之间的一种映射关系,如图 12 所示。

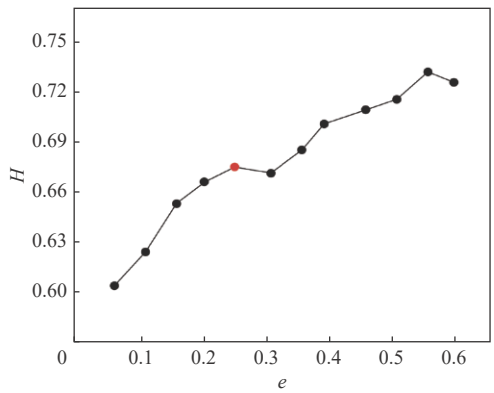


图 12 相对峰值高度 H 与相对深度 e 映射关系

Fig. 12 Mapping relationship between relative peak height H and relative depth e

由图 12 可知,H-e 的映射关系是非线性变化,外圈缺陷相对深度 e 在从 0.05 增加到 0.20 与从 0.35 增加到 0.55 过程中,前者增大幅度明显大于后者。早期外圈内部缺陷对轴承外圈刚度的削弱十分明显,进而导致在故障初期的相对峰值高度 H 的变化率较为显著。这种映射关系可以成为判断全陶瓷球轴承外圈细观缺陷的演化趋势情况的一个重要指标,对全陶瓷球轴承的缺陷演化提供参考。

3 实验验证

通过对实验数据的分析,验证本文模型的准确性和可行性,本实验中使用的 MFS-MG 机械缺陷综合模拟实验台,如图 13 所示。

将 6304 全陶瓷球轴承安装在轴承缺陷实验台上,实验的主要参数如表 1 所示。为获得亚表面细观尺度缺陷对全陶瓷球轴承运行性能的影响,将 2 个健康的全陶瓷球轴承(氧化锆)样品设置为在 300 N 的径向力和 1 800 r/min 的运行速度下连续运行 50 h。然后,通过超声波检测设备对 2 个缺陷轴承进行无损检测。将缺陷轴承放置在“MFS-MG 机械缺陷综合模拟实验台”进行测试。结果如图 14 所示。

考虑到研究对象为外圈亚表面细观缺陷,通过送检超声波扫描显微镜,实验陶瓷轴承在外圈存在缺陷深度 d=1.0 mm(即 e=0.25) 的亚表面细观缺陷。将所测得外圈细观缺陷的实验结果与本文模型的

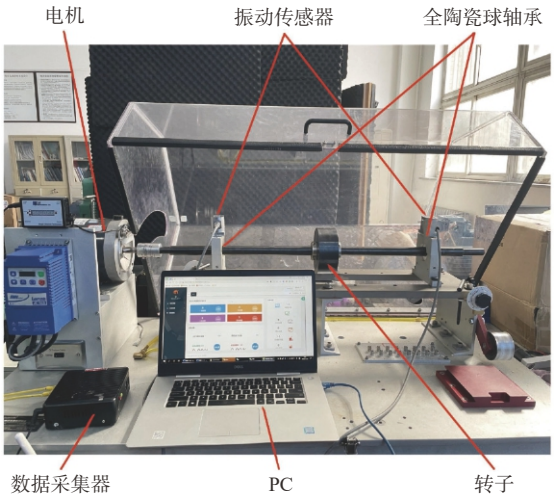


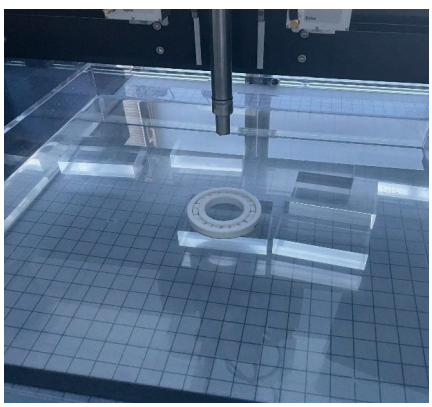
图 13　MFS-MG 机械缺陷综合模拟实验台

Fig. 13　MFS-MG mechanical defect synthetic simulating experiment bench

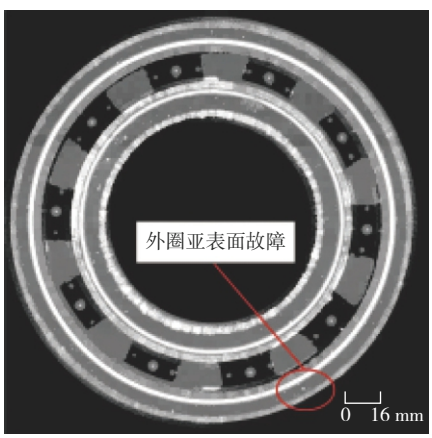
表 1　全陶瓷球轴承外圈亚表面细观缺陷实验参数

Table 1　Full ceramic ball bearing outer ring sub-surface mesoscopic fault experimental parameters

主轴转速ω_s/(r·min^{-1})	径向载荷F_r/N	轴承型号	采样频率/Hz
1 800	300	6304	4 096



(a) 超声波扫描显微镜检测



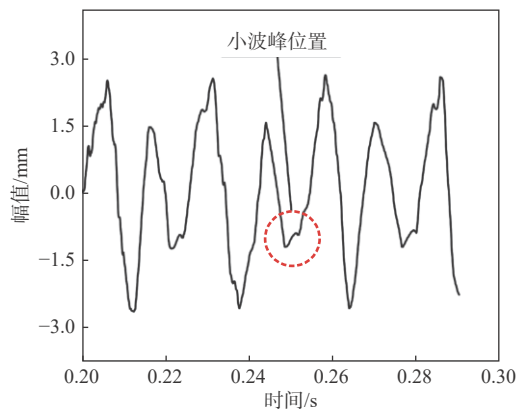
(b) 全陶瓷球轴承检测结果

图 14　超声波扫描显微镜检测及结果

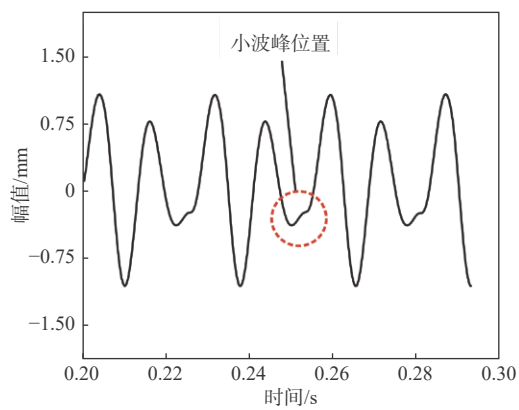
Fig. 14　Ultrasonic scanning microscope test and test results

仿真信号进行对比, 如图 15 所示。

由图 15 可知, 在缺陷深度 d=1.0 mm(即 e=0.25)时, 在实验缺陷时域信号中出现周期性小波峰, 但由于在实际信号采集的过程中还包含着其他背景噪声信号, 因此, 实验信号时域幅值与仿真信号幅值不完全一致, 但在实验信号中存在小波峰, 可以判断出在轴承外圈存在亚表面细观缺陷, 可以验证本文模型的准确性。另外将模拟缺陷时域信号与实验缺陷时域信号进行傅里叶变换, 得到模拟与实验的频域信号, 如图 16 所示。



(a) 实验振动信号



(b) 仿真振动信号

图 15　细观缺陷轴承外圈振动响应

Fig. 15　Fine view of vibration response of outer ring of faulty bearing

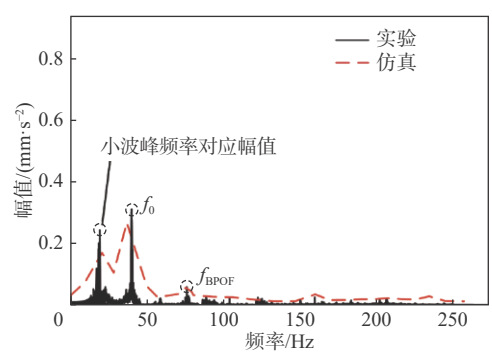


图 16　实验和仿真亚表面缺陷振动信号频域对比

Fig. 16　Frequency domain comparison of vibration signals from experimental and simulated subsurface defects

由图16可知，在频域图中实验与仿真信号的频率成分组成基本一致，小波峰所对应的频率幅值较为突出。在相对深度 e=0.25 的条件下，模拟相对峰值高度 H 约为 0.68，而实验数据中 H 约为 0.77，仿真结果与实验结果较为吻合，准确率近 90%，准确率良好，进一步验证了本文模型的准确性。

4 结 论

1）全陶瓷球轴承外圈存在亚表面缺陷及缺陷存在扩展趋势时，通过量化分析轴承外圈应变能的变化情况，得到应变能与外圈刚度之间的函数关系，借助刚度削弱系数表征亚表面缺陷对轴承外圈刚度的劣化程度，建立基于陶瓷材料应变能理论的全陶瓷球轴承外圈亚表面细观缺陷识别的动力学模型。

2）结合全陶瓷球轴承外圈存在亚表面缺陷时，在时域信号中出现小波峰现象这一特征指标，可通过这一指标判断全陶瓷球轴承外圈是否存在亚表面细观缺陷。实验结果与仿真结果较为一致，可以验证该判断指标的准确性。

3）当全陶瓷球轴承外圈中存在亚表面缺陷时，在其频域信号转频附近会出现一个与小波峰对应的特征频率。可根据频谱 EMD 的三阶分量幅值正向最大值与其对应的转频峰值的比值 H 的变化来判断缺陷演化的程度。实验结果与仿真结果吻合良好，误差率控制在合理区间。

参考文献（References）

[1] HARRER W, DELUCA M, MORRELL R. Failure analysis of a ceramic ball race bearing made of Y-TZP zirconia[J]. Engineering Failure Analysis, 2014, 36: 262-268.

[2] JAURIQUI L, HUNTER L. A more comprehensive NDE: PCRT for ceramic components[C]//Proceedings of the Mechanical Properties and Performance of Engineering Ceramics and Composites V: Ceramic Engineering and Science. College Park: American Institute of Physics, 2010, 31: 145-156.

[3] 牛蔺楷, 曹宏瑞, 何正嘉. 具有局部表面损伤的滚动球轴承动力学建模[J]. 振动 测试与诊断, 2014, 34(2): 356-360.
NIU L K, CAO H R, HE Z J. Dynamic modeling and vibration response simulation for rolling ball bearings with local surface defects[J]. Journal of Vibration, Measurement & Diagnosis, 2014, 34(2): 356-360(in Chinese).

[4] 关贞珍, 郑海起, 王彦刚, 等. 滚动轴承局部损伤故障动力学建模及仿真[J]. 振动 测试与诊断, 2012, 32(6): 950-955.
GUAN Z Z, ZHENG H Q, WANG Y G, et al. Fault dynamic modeling and simulating of rolling bearing with localized defect[J]. Journal of Vibration, Measurement & Diagnosis, 2012, 32(6): 950-955(in Chinese).

[5] DENG S, QIN X P, HUANG S. A study on the effect of subsurface crack propagation on rolling contact fatigue in a bearing ring[J]. Journal of Mechanical Science and Technology, 2015, 29(3): 1029-1038.

[6] LIU J, LI X B, SHI Z F. An investigation of contact characteristics of a roller bearing with a subsurface crack[J]. Engineering Failure Analysis, 2020, 116: 104744.

[7] DANIELSEN H K, GUTIÉRREZ GUZMÁN F, FÆSTER S, et al. Accelerated white etch cracking (WEC) FE8 type tests of different bearing steels using ceramic rollers[J]. Wear, 2022, 494-495: 204230.

[8] RAGA R, KHADER I, CHLUP Z, et al. Damage initiation and evolution in silicon nitride under non-conforming lubricated hybrid rolling contact[J]. Wear, 2016, 360: 147-159.

[9] KANEMATSU W. A review of rolling contact fatigue behavior of silicon nitride focusing on testing practices and crack propagation analysis[J]. Wear, 2018, 400: 10-20.

[10] NAZIR M H, KHAN Z A, SAEED A. Experimental analysis and modelling of c-crack propagation in silicon nitride ball bearing element under rolling contact fatigue[J]. Tribology International, 2018, 126: 386-401.

[11] HÄRTELT M, FÜNFSCHILLING S, SCHWIND T, et al. Deducing the fatigue crack growth rates of natural flaws in silicon nitride ceramics: role of R-curves[J]. Journal of the American Ceramic Society, 2013, 96(8): 2593-2597.

[12] MA H, ZENG J, FENG R J, et al. Review on dynamics of cracked gear systems[J]. Engineering Failure Analysis, 2015, 55: 224-245.

[13] SAEEDIFAR M, AHMADI NAJAFABADI M, MOHAMMADI K, et al. Acoustic emission-based methodology to evaluate delamination crack growth under quasi-static and fatigue loading conditions [J]. Journal of Nondestructive Evaluation, 2017, 37(1): 1.

[14] STROBL S, SUPANCIC P, LUBE T, et al. Surface crack in tension or in bending——a reassessment of the Newman and Raju formula in respect to fracture toughness measurements in brittle materials[J]. Journal of the European Ceramic Society, 2012, 32(8): 1491-1501.

[15] WANG L Z, MA D J, SHI X Z, et al. Determining the fracture toughness of silicon nitride by Vickers indenter[J]. China Measurement & Test, 2017, 43, 129–135.

[16] 胡爱军, 吉新星, 向玲, 等. 滚动轴承非线性时变参数动力学模型与故障机理研究[J]. 机械工程学报, 2022, 58(19): 139-147.
HU A J, JI X X, XIANG L, et al. Nonlinear time-varying parameter dynamic model of rolling bearing and failure mechanism research[J]. Journal of Mechanical Engineering, 2022, 58(19): 139-147(in Chinese).

[17] DONG Q Q, WEI H J, MA G W. Failure mechanism of S-shaped fissure in brittle materials under uniaxial tension: experimental and numerical analyses[J]. International Journal of Solids and Structures, 2020, 191: 486-496.

[18] SHI H T, LIU Z M, BAI X T, et al. A theoretical model with the effect of cracks in the local spalling of full ceramic ball bearings[J]. Applied Sciences, 2019, 9(19): 4142.

[19] 赵联春. 球轴承振动的研究[D]. 杭州: 浙江大学, 2003.
ZHAO L C. Study on vibration of ball bearing[D]. Hangzhou: Zhejiang University, 2003(in Chinese).

[20] LI T, SHI H T, BAI X T, et al. Experimental analysis and modeling of subsurface cracks with random propagation for ceramic material on rolling contact fatigue[J]. Engineering Failure Analysis, 2024,

155: 107753.

[21] SUNNERSJÖ C S. Varying compliance vibrations of rolling bearings[J]. Journal of Sound and Vibration, 1978, 58(3): 363-373.

[22] BAI X T, ZHANG Z N, SHI H T, et al. Identification of subsurface mesoscale crack in full ceramic ball bearings based on strain energy theory[J]. Applied Sciences, 2023, 13(13): 7783.

Identification of mesoscopic fault of full ceramic ball bearings based on strain energy theory

BAI Xiaotian，ZHANG Zhaonan，SHI Huaitao*，ZHANG Xiaochen

(School of Mechanical Engineering，Shenyang Jianzhu University，Shenyang 110168，China)

Abstract：The traditional steel bearing outer ring dynamics model is unable to describe the issue of sub-millimeter or even millimeter-scale mesoscopic flaws inside the outer ring of complete ceramic ball bearings. Based on the strain energy theory, a dynamic model is established for the subsurface mesoscopic defects of the outer ring of full ceramic ball bearings, and the influence of different defect depths on the running state of the outer ring of full ceramic ball bearings is investigated at the scale of mesoscopic defects. Through simulation, it is confirmed that when a mesoscopic defect exists on the subsurface of the outer ring of the bearing. There is a corresponding characteristic frequency near the intermediate frequency of the frequency domain signal, and the degree of defect evolution is determined by the ratio of the maximum positive value of its empirical mode decomposition (EMD) third-order component amplitude to its corresponding peak value of the intermediate frequency. Finally, the effectiveness of the constructed model is verified through experiments. Finally, the validity of the proposed dynamics model and simulation results is verified by experiments. The established model makes it possible to determine the degree of defect evolution and if sub-surface mesoscopic defects are present in the outer rings of complete ceramic ball bearings. The established model provides a new idea for the defect diagnosis of full ceramic ball bearings and a theoretical reference for the safe and stable operation of full ceramic bearing rotor systems.

Keywords：full ceramic ball bearing；mesoscopic defects；dynamic modeling；strain energy theory；empirical mode decomposition

Received：2023-11-02；**Accepted**：2024-03-22；**Published Online**：2024-03-28 13：55

URL：link.cnki.net/urlid/11.2625.V.20240327.1033.001

Foundation items：National Natural Science Foundation of China (52275119,52075348,52005352); Liaoning Provincial Department of Education Fund (LJKZZ20220078)

* **Corresponding author**. E-mail：sht@sjzu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0710

火箭倒飞逆向喷流干扰特性数值模拟研究

张培红，贾洪印，赵娇，吴晓军，周桂宇*，张耀冰
(中国空气动力研究与发展中心 计算空气动力研究所，绵阳 621000)

摘 要：有动力垂直回收是实现运载火箭可重复使用的关键技术之一。垂直回收过程中，火箭倒飞产生的大分离非规则流动，以及发动机逆向喷流与主流相互作用导致的复杂气动干扰，对数值模拟方法提出了严峻挑战。基于NNW-FlowStar软件和非结构混合网格技术，采用球头逆向喷流标模，以及自主开展的火箭倒飞逆向喷流风洞试验标模，对数值模拟方法进行了验证，研究不同双喷喷流状态对火箭倒飞逆向喷流干扰特性的影响。结果表明：NNW-FlowStar软件可以较好模拟火箭倒飞逆向喷流干扰特性，数值模拟结果与试验结果吻合较好，计算得到的流场结构与风洞试验一致；不同组合喷流形式呈现出不同的喷流干扰特性和流场结构，不同喷流方案轴向力差异明显，对力矩特性影响也较大，甚至会导致全箭本体的纵向静稳定性发生变化，在超声速来流状态采用双喷状态2即2个水平喷管的方案减速效果较好。

关 键 词：火箭倒飞；逆向喷流；干扰特性；NNW-FlowStar；数值模拟
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2026)01-0120-09

实现运载火箭的可重复使用，降低运载火箭发射成本，一直是国内外航天领域研究机构和宇航公司孜孜追求的目标[1]。垂直回收技术是实现运载火箭可重复使用的关键[2-3]。早在20世纪90年代，美国麦道公司就开始了动力反推垂直回收技术的研究，并采用“三角快帆”飞行器完成了4次飞行试验，验证了有动力垂直回收技术的可行性[2,4]。美国太空探索科技公司—Space-X通过实施“蚱蜢”计划，开展了一系列一子级陆上和海上软着陆试验，以验证垂直回收技术[5]。并于2015年12月22日，在世界上首次实现猎鹰-9火箭的第1级有动力垂直回收[6]，为降低火箭发射成本、实现火箭重复使用奠定了技术基础。截止2022年5月，Space-X公司成功进行了118次火箭着陆，B1051助推器成为首枚被回收12次的火箭一子级。Space-X公司还基于垂直回收技术，推出了下一代完全可重复使用运载火箭“星舰(Starship)”，并已开展了2次试飞，引领了重复使用运载器的技术发展方向[7]。国内相关机构和民营公司也相继开展了火箭有动力垂直回收的探索和试验，实现了一些技术突破。2022年5月，深蓝航天有限公司成功完成1 km级垂直起飞和降落飞行试验。

可重复使用火箭返回与着陆过程中，需要准确预测其气动特性和流场结构，以实现高精度的制导与控制。火箭再入过程中底部发动机喷管向前飞行，发动机在高空进行减速控制，其气动特性与流动结构与传统意义的飞行器差异较大[8]。火箭倒飞状态，在低马赫数时，火箭肩部区域会产生流动分离，在高马赫数时，向前凸出的发动机喷管会产生强脱体激波，发动机逆向喷流与主流相互作用会产生复杂气动干扰，较小的数值扰动即可能造成数值模拟不收敛，对网格及数值方法提出了严峻挑战。

收稿日期：2023-08-05；**录用日期**：2023-12-01；**网络出版时间**：2023-12-13 09:53
网络出版地址：link.cnki.net/urlid/11.2625.V.20231212.1501.002
基金项目：国家数值风洞工程
* **通信作者**. E-mail：zph2s@sina.com

引用格式：张培红，贾洪印，赵娇，等. 火箭倒飞逆向喷流干扰特性数值模拟研究 [J]. 北京航空航天大学学报，2026，52(1)：120-128.
ZHANG P H，JIA H Y，ZHAO J，et al. Numerical simulation research on opposing jet interaction characteristics of rocket inverse flight[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：120-128 (in Chinese).

火箭发动机逆向喷流效应属于典型的逆向喷流干扰问题。根据喷流压比的不同，超声速逆向喷流流场结构可分为长穿透模态和短穿透模态[9]。当喷流质量流量较低时，喷流产生的复杂斜激波与来流相互作用产生弱弓形激波，此时流场结构为不稳定状态，属于典型的长穿透模态。随着质量流量的增大，逆向喷流与来流作用产生的弓形激波增强，当质量流量超过一定临界值时，喷管出口斜激波系结构会突然崩溃，喷流在主流中的穿透距离骤减，形成短穿透模态。逆向喷流技术的应用主要集中在钝头体返回舱高速再入阶段的热防护[10-11]、大长细比导弹超声速飞行时的减阻[12]，以及高超声速飞行器在临近空间长时间飞行时的热防护和减阻增升[13]。从 20 世纪 60 年代开始，国内外相关学者对逆向喷流在高超声速飞行器减阻降热中开展了大量的研究工作。文献 [14-15] 通过研究发现，逆向喷流会将鼻锥构型头部产的弓形激波推向物面，同时会转换为弱斜激波结构，从而降低气动阻力。Daso 等[16] 采用风洞试验手段研究了某导弹飞行攻角、质量流率、喷口外形等参数的影响规律，获得了 2 种模态流场高速纹影图像。Hayashi 等[17] 基于风洞试验研究了来流马赫数为 3.98 时不同喷流压比对钝头体头部流场结构和喷流降热效果的影响。Meyer 等[18] 数值模拟研究了来流马赫数为 6.5 时，逆向喷流对钝体前缘激波阻力、表面热流和表面摩擦阻力等影响。Aso 等[11] 通过逆向喷流减阻降热试验揭示了逆向喷流降热的作用机理。Venkatachari 等[9] 深入研究了钝头体驻点附近逆向喷流的流场结构，以及模态转换机制。Li 等[19] 研究了逆向喷流对高超声速乘波体升阻特性的影响，表明逆向喷流可以有效减小头部波阻。Bibi 等[20] 通过对 Apollo 返回舱喷流内外流场完全模拟，研究了逆向喷流对减阻降热的影响。

国内学者针对逆向喷流在高超声速飞行器减阻降热方面的应用也开展了大量研究工作，取得了丰硕的成果。黄伟等[15,21] 针对逆向射流及其组合在减阻防热方面的应用开展了一系列详细的研究。王立强等[22] 采用数值模拟手段研究了喷流质量流量和喷口尺寸对钝头体降热效果的影响，分析了不同模态流场结构形成的机理。黄喜元等[23] 研究了单个喷管和多个喷管对降热效果的影响，表明多个喷管的鲁棒性更好、工程适用性更强。邓帆等[24] 系统梳理了逆向喷流在高超声速钝头体外形和升力体外形上的应用情况。戎宜生等[25] 研究了逆向喷流对降低再入飞行器鼻锥物面热流的影响。王泽江等[26] 研究了喷流马赫数、喷流压比等参数对减阻效果的影响。但目前有关逆向喷流的研究成果主要集中在通过流动控制进行减阻降热方面，针对火箭垂直回收过程火箭倒飞逆向喷流干扰特性的研究，公开发表的相关文献较少。程川等[27] 采用风洞试验手段研究了发动机逆向喷流一子级栅格舵气动特性的影响规律。

前期，笔者项目团队针对火箭子级垂直回收倒飞构型气动特性和非规则钝头体绕流进行了研究，分析了喷管外露对气动特性的影响[8]。在此基础上，本文基于 NW-FlowStar 软件和非结构混合网格技术，进一步开展了火箭倒飞逆向喷流干扰特性研究。采用球头逆向喷流标模，以及自主开展的火箭倒飞逆向喷流风洞试验标模，对数值模拟方法进行验证，研究了不同喷流方案对火箭倒飞逆向喷流干扰特性的影响。

1　数值模拟方法

数值计算采用 NNW-FlowStar 软件[28]。该软件是中国空气动力研究与发展中心自主开发的一款基于非结构混合网格技术的大型通用计算流体力学（computational fluid dynamics, CFD）软件，主要面向航空航天领域飞行器定常、非定常气动特性计算分析和多体分离特性模拟。

流动控制方程采用守恒形式的可压缩雷诺平均 Navier-Stokes（N-S）方程：

$$\frac{\partial}{\partial t}\int_{\Omega}\boldsymbol{W}\mathrm{d}\Omega+\oint_{\partial\Omega}(\boldsymbol{F}_{\mathrm{c}}-\boldsymbol{F}_{\mathrm{v}})\mathrm{d}S=\boldsymbol{0} \tag{1}$$

式中：t为时间；Ω为控制体的体积；$\boldsymbol{W}$为守恒变量；$\boldsymbol{F}_{\mathrm{c}}$为无黏通量；$\boldsymbol{F}_{\mathrm{v}}$为黏性通量。定义如下：

$$\boldsymbol{W}=\begin{bmatrix}\rho & \rho u & \rho v & \rho w & \rho E\end{bmatrix}^{\mathrm{T}} \tag{2}$$

$$\boldsymbol{F}_{\mathrm{c}}=\begin{bmatrix}\rho\boldsymbol{V}\\ \rho u\boldsymbol{V}+\boldsymbol{n}_x p\\ \rho v\boldsymbol{V}+\boldsymbol{n}_y p\\ \rho w\boldsymbol{V}+\boldsymbol{n}_z p\\ \rho H\boldsymbol{V}\end{bmatrix} \tag{3}$$

$$\boldsymbol{F}_{\mathrm{v}}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{n}_x\tau_{xx}+\boldsymbol{n}_y\tau_{xy}+\boldsymbol{n}_z\tau_{xz}\\ \boldsymbol{n}_x\tau_{yx}+\boldsymbol{n}_y\tau_{yy}+\boldsymbol{n}_z\tau_{yz}\\ \boldsymbol{n}_x\tau_{zx}+\boldsymbol{n}_y\tau_{zy}+\boldsymbol{n}_z\tau_{zz}\\ \boldsymbol{n}_x\Theta_x+\boldsymbol{n}_y\Theta_y+\boldsymbol{n}_z\Theta_z\end{bmatrix} \tag{4}$$

$$\begin{cases}\Theta_x=u\tau_{xx}+v\tau_{xy}+w\tau_{xz}+k\dfrac{\partial T}{\partial x}\\ \Theta_y=u\tau_{yx}+v\tau_{yy}+w\tau_{yz}+k\dfrac{\partial T}{\partial y}\\ \Theta_z=u\tau_{zx}+v\tau_{zy}+w\tau_{zz}+k\dfrac{\partial T}{\partial z}\end{cases} \tag{5}$$

式中: x、y、z为笛卡儿坐标系下的坐标分量; u、v、w为笛卡儿坐标系下的速度分量; $\boldsymbol{n}_x$、$\boldsymbol{n}_y$、$\boldsymbol{n}_z$为微元面的单位法向矢量在笛卡儿坐标系下的分量; ρ为密度; p为压力; $\boldsymbol{V}$为速度矢量; H为焓; τ为剪切力; k为热传导系数; T为温度。

对流通量离散采用笔者项目团队发展的HLLE++改进格式, 黏性通量离散采用中心差分方法, 梯度求解采用节点型 Green-Gauss 法, 湍流模型采用剪切应力传输 (shear stress transport, SST) 两方程模型。时间推进采用隐式的下上对称高斯-赛德尔迭代 (lower-upper symmetric Gauss-Seidel, LU-SGS) 方法。

2 数值模拟方法验证

2.1 超声速球头逆向喷流算例验证

采用文献 [29] 中的超声速球头逆向喷流标准算例对数值方法进行验证。球头半径 R=0.025 4 mm, 喷流气体为空气, 来流马赫数 $Ma=2.5$, 来流单位雷诺数 $Re_L=2.76\times10^7$, 喷流总压与主流激波后总压之比 $P_{0j}/P_{0f}=6.54$, 计算状态参数如表 1 所示。

表 1 计算状态参数

Table 1 Parameters of simulation conditions

参数	数值
来流马赫数	2.5
来流总温/K	294.4
来流总压/kPa	275.8
喷流马赫数	1.0
喷流总压比	6.54
喷流总温/K	294.4

网格采用由三棱柱、六面体和四面体组成的非结构混合网格, 网格总量约为 1.6×10^6, 附面层第1层间距 $y^+\approx1$。为更好地模拟附面层和喷流附近的强剪切, 在物面黏性作用区采用大伸展比三棱柱形和金字塔形网格, 在球头喷口附近进行了加密处理, 计算网格如图 1 所示。

计算采用 SST 两方程湍流模型, 图 2 为计算得到的球头表面压力系数分布与试验结果比较, 图中横坐标 θ 在驻点处定义为 0°, 沿对称面往上或往下到垂直面为 90°, 如图 3 所示。图 4 为计算得到的流场结构与试验结果比较。可以看出, 无喷流和有喷流计算得到的球头表面压力分布都与风洞试验结果吻合很好, 计算得到的有喷流流场结构与风洞试验结果一致, 验证了本文方法模拟逆向喷流的可靠性。高压气流从喷口喷出后与来流相互作用, 形成干扰层和清晰的激波交界面。流场中存在复杂的激波结构, 在来流和球头之间有一道强的弓形激波, 喷口有膨胀波, 同时球头物面附近还存在再附激波。自由来流由于逆向射流的阻挡作用, 形成典型的马赫盘结构, 在轴线上的一点形成自由流滞止点。

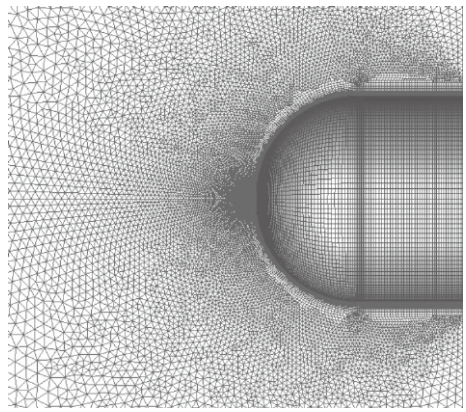


图 1 计算网格

Fig. 1 Computational grid

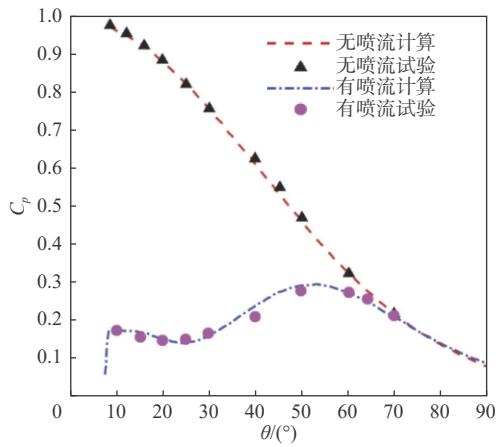


图 2 球头表面压力分布计算与风洞试验对比

Fig. 2 Comparison between calculation and experimental pressure distribution on surface of ball head and wind tunnel tests

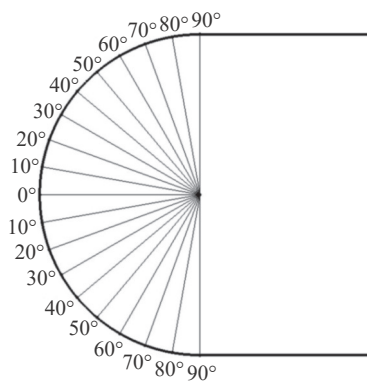
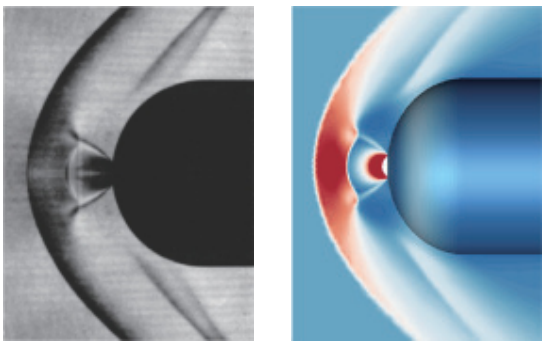


图 3 θ 定义示意图

Fig. 3 Schematic diagram of θ definition



(a) 流场结构试验　　(b) 流场结构计算

图 4　计算与试验流场结构对比

Fig. 4　Comparison between computational and experimental flow field structure

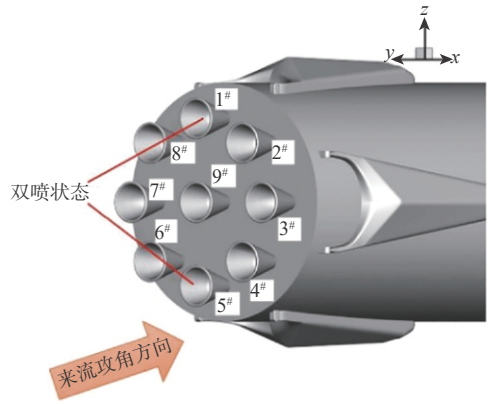
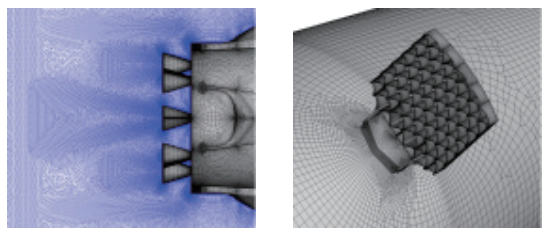


图 6　双喷状态喷流喷口示意图

Fig. 6　Diagram of jet nozzle for double jet state



(a) 喷口附近网格　　(b) 栅格舵附近网格

图 7　火箭子级垂直回收构型网格分布

Fig. 7　Grid distribution of rocket vertical recovery configuration

2.2　某火箭子级逆向喷流干扰算例验证

为进一步验证数值方法模拟倒飞逆向喷流干扰特性的可靠性，本文以某火箭子级垂直回收构型为研究对象，开展了风洞试验和数值模拟对比验证。该火箭子级垂直回收构型基于箭体直径 Φ=3.35 m 火箭一子级设计，4 个弧形栅格舵呈十字形布置，4 个着陆支腿折叠安装在箭体上，发动机喷口布局形式采用 9 喷口环形布置方案，9 台发动机可根据回收过程不同阶段和工况的需要以不同的组合模式工作。风洞试验模型缩比为 1:15，风洞试验模型照片如图 5 所示。风洞试验在中国空气动力研究与发展中心 FL-28 风洞完成，来流马赫数 Ma = 2.0，来流压力 p = 23.879 574 kPa，来流温度 T = 156.14 K，喷流总压 P_{oj} = 8 MPa，喷流压比 P_{0j}/P_{0f} = 335，来流攻角 α = 0°～15°。

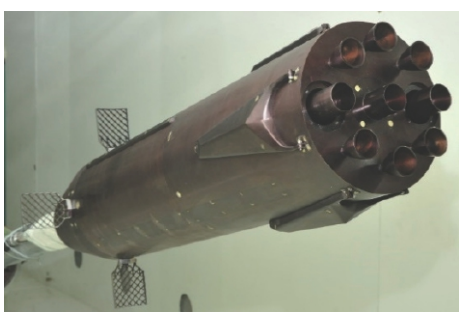


图 5　风洞试验模型照片

Fig. 5　Photo of model for wind tunnel experiment

采用 1#、5#这 2 个喷管有喷流的双喷状态试验数据对数值方法进行验证，双喷状态喷流喷口示意图如图 6 所示。网格生成仍然采用非结构混合网格，附面层第 1 层间距 $y^+ \approx 1$，在分离区、激波区域进行适当加密，考虑逆向喷流干扰流场的复杂性，在逆向喷流干扰区域，采用大量四面体单元进行空间加密，网格总量约为 7.6×10^7。图 7 为火箭子级垂直回收构型典型网格分布。

风洞试验和数值模拟，俯仰力矩参考点均取质心，即箭体长度 50% 处。图 8 为火箭子级垂直回收构型在马赫数为 2.0，1#、5#这 2 个喷管同时工作双喷状态计算与试验结果对比。可以看出，本文计算得到的全箭法向力系数与风洞试验结果具有较好的一致性；在中小攻角范围内，计算得到的轴向力系数与试验结果一致性较好，但攻角较大时计算值略大于试验结果；从力矩曲线可以看出，在攻角小于 2°时，火箭子级为静稳定，攻角大于 2°后变为纵向静不稳定，计算结果和风洞试验结果趋势一致。小攻角时，计算得到的压心系数比风洞试验结果略大，压心更靠后，攻角大于 4°后，计算与试验结果吻合较好。图 9 为攻角 0°，马赫数为 2.0 时，计算得到的流场结构与风洞试验对比。计算得到的波系结构与风洞试验一致性较好，风洞试验纹影图片在流场中间区域能看到明显的脱体激波现象，在数值仿真结果中也有相同的流场结构。综上可知，本文方法可以较好模拟火箭子级垂直回收构型逆向喷流干扰复杂流场结构，计算得到的流场结构与风洞试验一致，计算得到的法向力系数、轴向力系数、俯仰力矩系数和压心系数与风洞试验结果吻合较好。

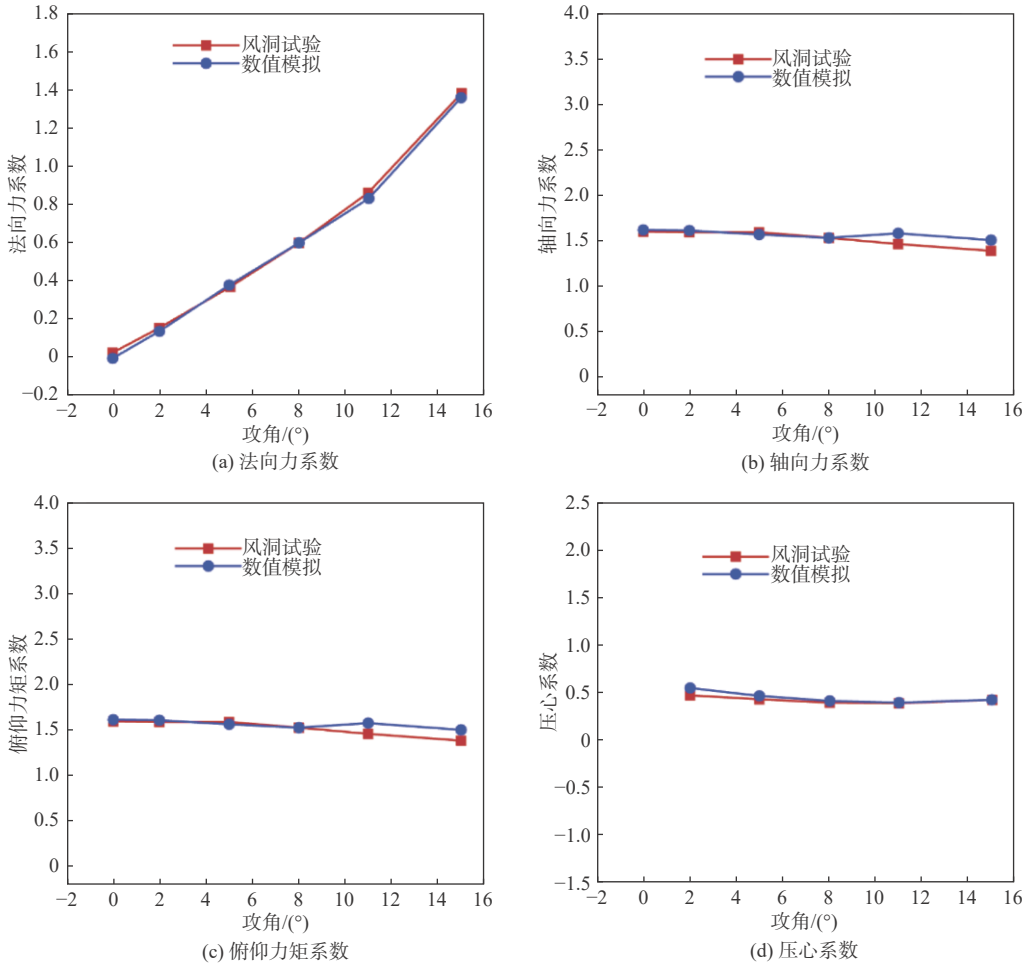


图 8 *Ma*=2.0, 双喷状态计算与试验结果对比

Fig. 8 Comparison between numerical and experimental for double jet state of *Ma*=2.0

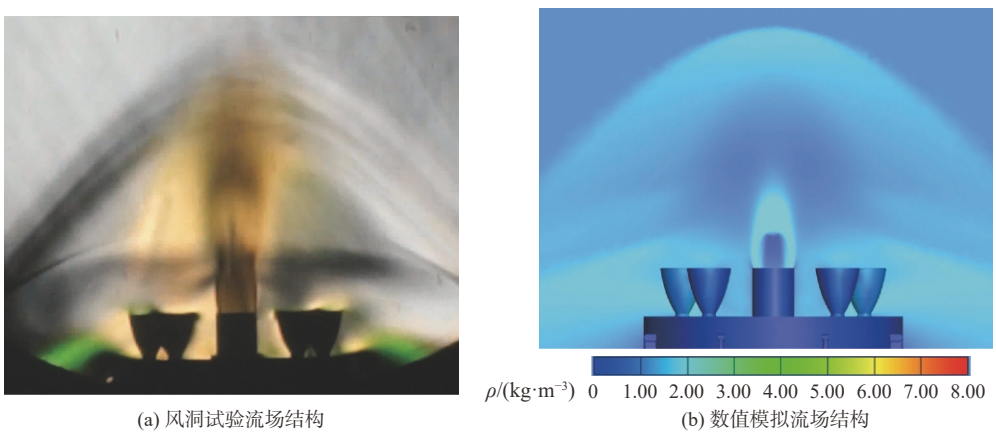
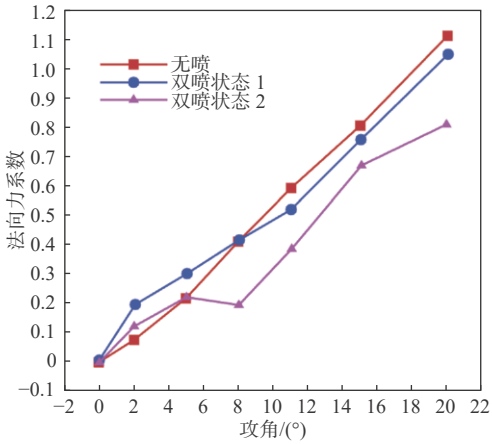


图 9 α=0°时, 计算得到的流场结构与风洞试验比较

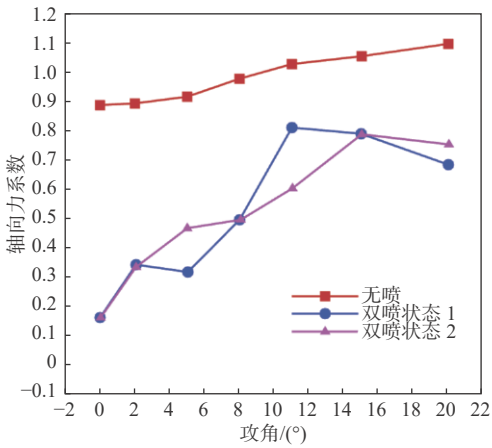
Fig. 9 Comparison between numerical and experimental when α=0° with wind tunnel tests

3 组合喷流对火箭子级干扰特性影响

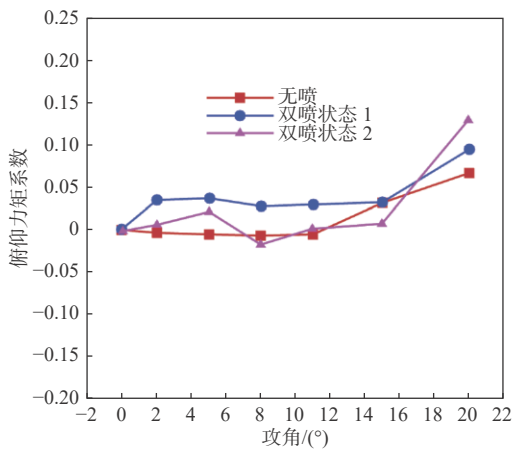
为研究发动机组合喷流对箭子级干扰特性的影响, 以 2.2 节中的火箭子级垂直回收构型为研究对象, 开展了马赫数为 0.4、4.0 时无喷、双喷状态 1、双喷状态 2 等不同工况的数值模拟研究, 其中, 双喷状态 1 和 2.2 节中的算例一样, 1#、5#这 2 个喷管同时有喷流, 双喷状态 2 是 3#、7#这 2 个喷管同时有喷流。图 10 和图 11 分别给出了马赫数为 0.4、4.0 时, 双喷状态喷流对法向力、轴向力和俯仰力矩等气动特性的影响。不同双喷状态对气动特性影响较大, 马赫数为 0.4 时, 法向力干扰特性规律不明显, 马赫数为 4.0 时双喷状态一法向力降低, 而双喷状态 2 法向力增大, 说明组合喷流喷口位置对法向



(a) 法向力系数



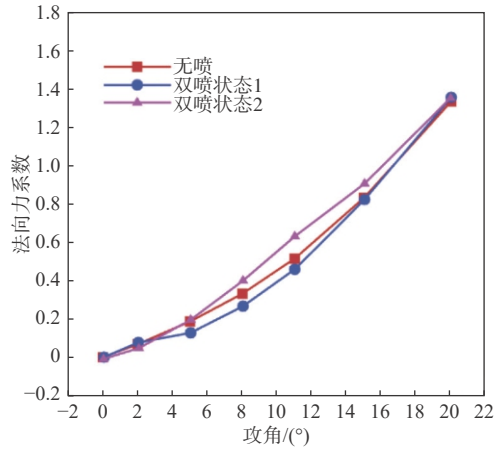
(b) 轴向力系数



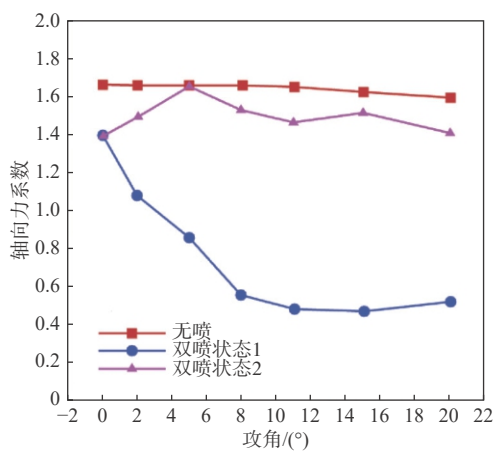
(c) 俯仰力矩系数

图 10　*Ma*=0.4, 双喷状态喷流对气动特性影响

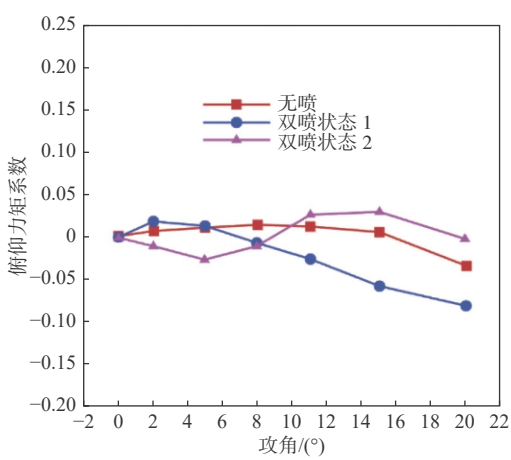
Fig. 10　Influence of double jet state on aerodynamic characteristics of *Ma*=0.4



(a) 法向力系数



(b) 轴向力系数



(c) 俯仰力矩系数

图 11　*Ma*=4.0, 双喷状态喷流对气动特性影响

Fig. 11　Influence of double jet state on aerodynamic characteristics of *Ma*=4.0

力的影响规律有影响。不同马赫数下, 发动机工作后轴向力均呈降低的趋势, 马赫数为 0.4 时, 2 种双喷状态的轴向力差异不大, 而在马赫数为 4.0 时, 2 种双喷状态轴向力相差较大, 双喷状态 2 的轴向力明显高于双喷状态 1, 因此, 在超声速来流状态如果为减速需要发动机采用双喷状态工作, 建议采用双喷状态 2 的组合发动机工作形式。不同发动机组合喷流对力矩特性影响也较大, 甚至会导致全箭本体的纵向静稳定性发生变化, 在控制系统设计时必须给予考虑。

图 12 为 *Ma*=0.4, 双喷状态喷流对流动结构影响。图 13 为 *Ma*=0.4, 双喷状态喷流对迎风端面压力分布影响。亚声速来流时, 喷管出口附近高压喷流在外界环境压力作用下会形成复杂的马赫盘结构, 强剪切流动在空气黏性作用下, 会诱导出 2 个较大分离涡结构。2 种双喷状态组合喷流干扰流场

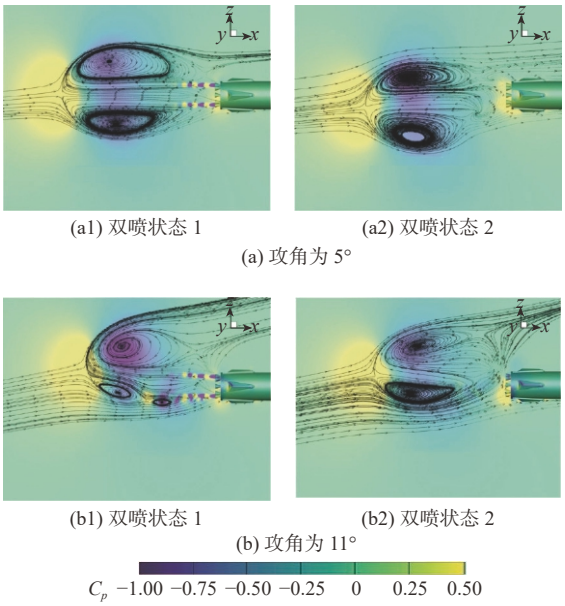


图 12 *Ma*=0.4, 双喷状态喷流对流动结构影响

Fig. 12 Influence of double jet state on flow structure of *Ma*=0.4

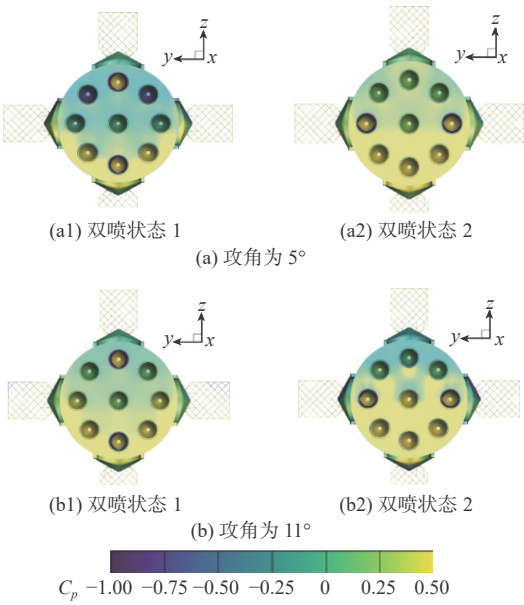


图 13 *Ma*=0.4, 双喷状态喷流对迎风端面压力分布影响

Fig. 13 Influence of double jet state on pressure distribution for windward surface of *Ma*=0.4

差异较大, 随着攻角增大, 双喷状态 1 的干扰流场呈现出了明显的非对称状态, 迎风区域喷流诱导的分离旋涡逐渐减小, 而背风区域的旋涡呈逐渐增强; 对于双喷状态 2, 喷流位于左右对称区域, 左右两侧双喷同样会在对称面上诱导出上下 2 个分离旋涡, 旋涡的强度同样受到攻角的影响。由于逆向喷流带来的遮挡效应, 双喷状态 2 在攻角为 5°时迎风端面上的压力明显高于双喷状态 1, 随着攻角的增大, 2 种双喷状态迎风端面上的压力均有所增大。

图 14 为 *Ma*=4.0, 双喷状态喷流对流动结构影响。图 15 为 *Ma*=4.0, 双喷状态喷流对迎风端面压力分布影响。超声速来流时, 高压气体在喷管出口会急剧膨胀, 在来流的作用下进一步压缩, 形成马赫盘结构, 沿着喷流的膨胀-压缩区域, 会产生强剪切流动, 使气流产生较大范围的分离, 发动机逆向喷流与主流交汇后, 会将钝头体头激波推离火箭发动机, 形成较强的脱体激波。2 种双喷状态下, 火箭迎风端面上的压力分布差异较大, 双喷状态 2 迎风端面上的压力值明显高于双喷状态 1。攻角为 5°时, 上下 2 个喷流的流场呈现出了明显的非对称

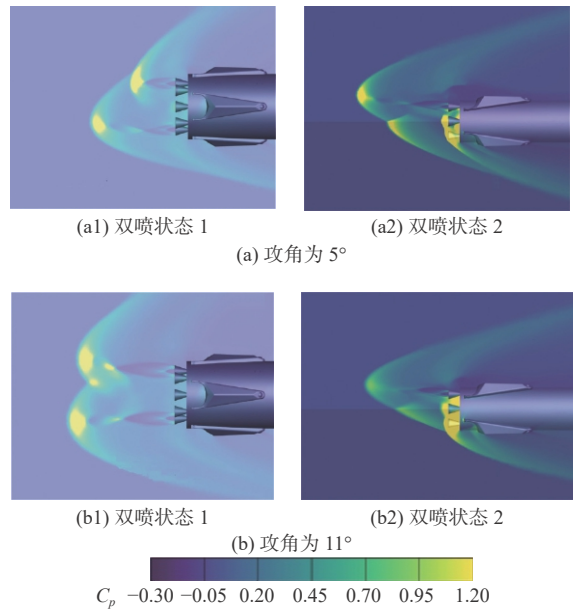


图 14 *Ma*=4.0, 双喷状态喷流对流动结构影响

Fig. 14 Influence of double jet state on flow structure of *Ma*=4.0

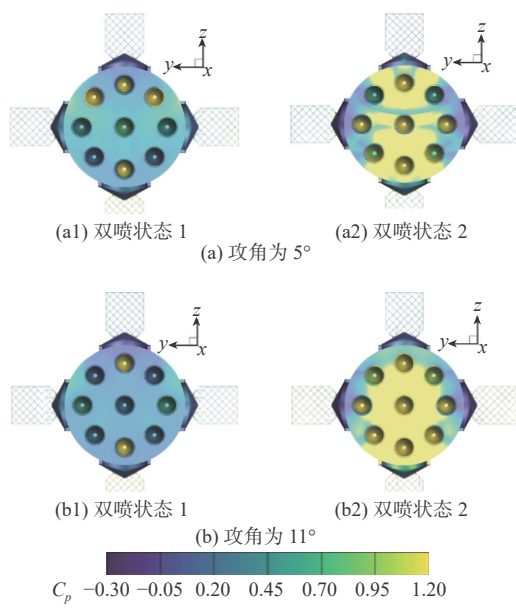


图 15 *Ma*=4.0, 双喷状态喷流对迎风端面压力分布影响

Fig. 15 Influence of double jet state on pressure distribution for windward surface of *Ma*=4.0

性, 下喷管区域的逆向喷流向前穿透的长度明显高于上喷管, 其主要原因是攻角效应使得迎风区域的高压作用到上喷管区域, 从而导致上喷管区域逆向喷流向前穿透的距离减弱; 随着攻角的增大, 上下喷管逆向喷流向前穿透的距离长度差距逐渐变小; 随攻角进一步增大, 下喷管喷流产生的背风侧脱体激波角变大, 从而导致对上喷管的影响变弱, 上下喷流向前传播的距离差距开始逐渐变小。双喷状态 2, 左右两侧双喷在对称面上诱导出上下 2 个分离旋涡, 在攻角的作用下, 左右 2 个喷管的高压会同时作用到上部区域, 使上部区域逆向喷流向前穿透的距离长度较下部区域大; 随着攻角进一步增大, 下部区域脱体激波受来流的作用, 进一步上移, 与上部区域背风侧脱体激波相互作用融为一个脱体激波。

4　结　论

本文采用 NNW-FlowStar 软件对球头逆向喷流标模和自主开展的火箭倒飞逆向喷流风洞试验标模进行了数值模拟, 研究了不同双喷喷流状态对火箭倒飞逆向喷流干扰特性的影响。主要结论如下:

1) 本文计算得到的气动特性数据和流场结构与风洞试验结果吻合较好, 验证了数值模拟方法对火箭倒飞逆向喷流干扰特性研究的有效性和可靠性。

2) 不同组合喷流形式对喷流干扰特性和流场结构影响较大, 甚至会导致全箭本体纵向静稳定性发生变化, 在控制系统设计时必须给予考虑。

3) 对于本文研究的 2 种双喷状态, 在超声速来流状态采用双喷状态 2 减速效果较好。

参考文献 (References)

[1] RICHARDSON M P, HARDY D W F. Economic benefits of reusable launch vehicles for space debris removal[C]//Proceedings of the 68th International Astronautical Congress. Paris: International Astronautical Federation, 2017.

[2] 汪小卫, 张普卓, 吴胜宝, 等. 运载火箭子级回收技术研究[J]. 航天返回与遥感, 2016, 37(3): 19-28.
WANG X W, ZHANG P Z, WU S B, et al. Recovery technology of launch vehicle stage[J]. Spacecraft Recovery & Remote Sensing, 2016, 37(3): 19-28(in Chinese).

[3] 宋征宇, 黄兵, 汪小卫, 等. 重复使用运载器回收技术现状与挑战[J]. 深空探测学报 (中英文), 2022, 9(5): 457-469.
SONG Z Y, HUANG B, WANG X W, et al. Status and challenges of reusable launch vehicle recovery technology[J]. Journal of Deep Space Exploration, 2022, 9(5): 457-469(in Chinese).

[4] 高朝辉, 张普卓, 刘宇, 等. 垂直返回重复使用运载火箭技术分析[J]. 宇航学报, 2016, 37(2): 145-152.
GAO Z H, ZHANG P Z, LIU Y, et al. Analysis of vertical landing technique in reusable launch vehicle[J]. Journal of Astronautics, 2016, 37(2): 145-152(in Chinese).

[5] SEALANDER J. FALCON heavy finally flies again[J]. Spaceflight, 2023, 65(1): 26-27.

[6] MALIK T. Success! SpaceX launches Falcon Heavy rocket on historic maiden voyage[EB/OL]. (2018-02-06)[2023-07-25]. https://www.space.com/39607-spacexfalcon-heavy-first-test-flight-launch.html.

[7] JARED L. SpaceX doubling down on Boca Chica; New permit shows future Starship launch site plans[EB/OL]. (2021-03-09)[2023-07-26]. https://spaceexplored.com/2021/03/09/spacex-doubling-down-on-boca-new-permit-shows-future-starship-launch-site-plans/.

[8] 贾洪印, 张培红, 赵炜, 等. 火箭子级垂直回收布局气动特性及发动机喷管影响[J]. 航空学报, 2021, 42(2): 623995.
JIA H Y, ZHANG P H, ZHAO W, et al. Aerodynamic characteristics of vertical recovery of rocket sub-stage and influence of engine nozzle[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(2): 623995(in Chinese).

[9] VENKATACHARI B S, ITO Y, CHENG G, et al. Numerical investigation of the interaction of counterflowing jets and supersonic capsule flows[C]//Proceedings of the 42nd AIAA Thermophysics Conference. Reston: AIAA, 2011.

[10] KEYES J W, HEFNER J N. Effect of forward-facing jets on aerodynamic characteristics of blunt configurations at Mach 6[J]. Journal of Spacecraft and Rockets, 1967, 4(4): 533-534.

[11] ASO S, HAYASHI K, MIZOGUCHI M. A study on aerodynamic heating reduction due to opposing jet in hypersonic flow[C]//Proceedings of the 40th AIAA Aerospace Sciences Meeting & Exhibit. Reston: AIAA, 2002.

[12] DASO E O, BEAULIEU W, HAGER J O, et al. Prediction of drag reduction in supersonic and hypersonic flows with cunterflow jets [C]//Proceedings of the 11th AIAA/AAAF International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 2002.

[13] SUN X W, GUO Z Y, HUANG W, et al. Drag and heat reduction mechanism induced by a combinational novel cavity and counterflowing jet concept in hypersonic flows[J]. Acta Astronautica, 2016, 126: 109-119.

[14] FINLEY P J. The flow of a jet from a body opposing a supersonic free stream[J]. Journal of Fluid Mechanics, 1966, 26: 337-368.

[15] HUANG W, JIANG Y P, YAN L, et al. Heat flux reduction mechanism induced by a combinational opposing jet and cavity concept in supersonic flows[J]. Acta Astronautica, 2016, 121: 164-171.

[16] DASO E O, PRITCHETT V E, WANG T S, et al. Dynamics of shock dispersion and interactions in supersonic freestreams with counterflowing jets[J]. AIAA Journal, 2009, 47(6): 1313-1326.

[17] HAYASHI K, ASO S, TANI Y. Experimental study on thermal protection system by opposing jet in supersonic flow[J]. Journal of Spacecraft and Rockets, 2006, 43(1): 233-235.

[18] MEYER B, NELSON H F, RIGGINS D W. Hypersonic drag and heat-transfer reduction using a forward-facing jet[J]. Journal of Aircraft, 2001, 38(4): 680-686.

[19] LI S B, WANG Z G, BARAKOS G N, et al. Research on the drag reduction performance induced by the counterflowing jet for

waverider with variable blunt radii[J]. Acta Astronautica, 2016, 127: 120-130.

[20] BIBI A, MAQSOOD A, SHERBAZ S, et al. Drag reduction of supersonic blunt bodies using opposing jet and nozzle geometric variations[J]. Aerospace Science and Technology, 2017, 69: 244-256.

[21] HUANG W, YAN L, LIU J, et al. Drag and heat reduction mechanism in the combinational opposing jet and acoustic cavity concept for hypersonic vehicles[J]. Aerospace Science and Technology, 2015, 42: 407-414.

[22] 王立强, 钱勤建. 钝体逆向喷流减阻降温数值模拟[J]. 弹箭与制导学报, 2019, 39(1): 55-59.
WANG L Q, QIAN Q J. Numerical investigation on counter-flow jets on drag and heat-transfer reduction[J]. Journal of Projectiles, Rockets, Missiles and Guidance, 2019, 39(1): 55-59(in Chinese).

[23] 黄喜元, 李小艳, 杨勇, 等. 高超声速单/多喷管逆向喷流降热特性研究[J]. 导弹与航天运载技术, 2021(4): 45-49.
HUANG X Y, LI X Y, YANG Y, et al. Counter flowing jet/jets on heat flux reduction for hypersonic vehicles[J]. Missiles and Space Vehicles, 2021(4): 45-49(in Chinese).

[24] 邓帆, 谢峰, 黄伟, 等. 逆向喷流技术在高超声速飞行器上的应用[J]. 空气动力学学报, 2017, 35(4): 485-495.
DENG F, XIE F, HUANG W, et al. Applications of counterflowing jet technology in hypersonic vehicle[J]. Acta Aerodynamica Sinica, 2017, 35(4): 485-495(in Chinese).

[25] 戎宜生, 刘伟强. 再入飞行器鼻锥逆向喷流对流场及气动热的影响[J]. 航空学报, 2010, 31(8): 1552-1557.
RONG Y S, LIU W Q. Influence of opposing jet on flow field and aerodynamic heating at nose of a reentry vehicle[J]. Acta Aeronautica et Astronautica Sinica, 2010, 31(8): 1552-1557(in Chinese).

[26] 王泽江, 李杰, 曾学军, 等. 逆向喷流对双锥导弹外形减阻特性的影响[J]. 航空学报, 2020, 41(12): 124116.
WANG Z J, LI J, ZENG X J, et al. Effect of opposing jet on drag reduction characteristics of double-cone missile shape[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(12): 124116(in Chinese).

[27] 程川, 刘阳, 王吉飞, 等. 反向喷流对运载火箭返回段气动特性影响研究[J]. 宇航学报, 2023, 44(3): 379-388.
CHENG C, LIU Y, WANG J F, et al. Investigation of reverse jet effects on aerodynamic characteristics of returned launch vehicle[J]. Journal of Astronautics, 2023, 44(3): 379-388(in Chinese).

[28] 陈坚强, 吴晓军, 张健, 等. FlowStar: 国家数值风洞 (NNW) 工程非结构通用 CFD 软件[J]. 航空学报, 2021, 42(9): 625739.
CHEN J Q, WU X J, ZHANG J, et al. FlowStar: general unstructured-grid CFD software for national numerical windtunnel(NNW) project[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(9): 625739(in Chinese).

[29] ZHOU C Y, JI W Y. A three-dimensional numerical investigation on drag reduction of a supersonic spherical body with an opposing jet[J]. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 2014, 228(2): 163-177.

Numerical simulation research on opposing jet interaction characteristics of rocket inverse flight

ZHANG Peihong，JIA Hongyin，ZHAO Jiao，WU Xiaojun，ZHOU Guiyu*，ZHANG Yaobing

(Computational Aerodynamic Institute of China Aerodynamic Research & Development Center，Mianyang 621000，China)

Abstract：Powered vertical recovery is one of the key technologies to realize the reusable launch vehicles. The rocket's backward flight will create a huge separation irregular flow throughout the vertical recovery process, and complicated aerodynamic interference will result from the engine's reverse jet interacting with the mainstream. These pose serious challenges to numerical simulation methods. Based on NNW-FlowStar software and unstructured hybrid grid technology, the numerical method was validated using the ball head opposing jet model and the self-developed rocket opposing jet wind tunnel test model. The interference characteristics of different double jet states on the opposing jet flow of the rocket are studied. The study demonstrates that the NNW-FlowStar is more capable of simulating the interference features of a rocket's reverse flight, and the findings of the numerical simulations correlate well with the test results. Different combined jet forms show different jet interference characteristics and flow field structures. The axial force of different jet schemes varies significantly, which also has a great impact on torque characteristics and even leads to changes in the longitudinal static stability of the whole rocket body. In the supersonic flow state, the deceleration effect is best in the second double injection state.

Keywords：rocket’s inverse flight；opposing jet；interference characteristics；NNW-FlowStar；numerical simulation

Received：2023-08-05；**Accepted**：2023-12-01；**Published Online**：2023-12-13 09：53

URL：link.cnki.net/urlid/11.2625.V.20231212.1501.002

Foundation item：National Numerical Windtunnel Project

* **Corresponding author**. E-mail：zph2s@sina.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0723

广义未知扰动下多模型最小上限滤波

秦月梅[1,*]，张荣华[1]，杨衍波[2]，潘泉[2]
(1. 西安邮电大学 自动化学院，西安 710121；2. 西北工业大学 自动化学院，西安 710129)

摘 要：针对现有机动目标跟踪算法在广义未知扰动下跟踪峰值误差过大的问题，提出一种多模型最小上限滤波 (MMUBF) 算法。在多模型框架下利用最小上限滤波对不同模式下状态进行递推估计，根据滤波结果及模型后验概率估计加权在线辨识扰动分量，将其引入各模式下似然概率计算以减弱其对模型概率更新的影响。同时，为进一步提升模式匹配精度，利用修正因子自适应调整马尔可夫转移概率矩阵。此外，通过计算每个步骤的浮点运算数量，分析所提算法的计算复杂度。具有时变未知扰动的机动目标跟踪仿真结果表明：相比于现有交互式多模型滤波、自适应交互式多模型滤波、灰狼优化算法改进后的自适应交互式多模型滤波和基于单模型的最小上限滤波算法，所提算法在不同量测噪声、过程噪声、调整系数及概率修正阈值水平下皆具有更小峰值误差和更高估计精度。

关 键 词：最小上限滤波；广义未知扰动；多模型；修正因子；转移概率矩阵
中图分类号：TN953
文献标志码：A **文章编号**：1001-5965(2026)01-0129-11

机动目标跟踪一直是雷达目标跟踪和多传感器数据融合的难点和热点问题之一[1]，被广泛应用于防空预警、多机协同定位跟踪、空管雷达目标跟踪、自动驾驶等军、民用领域[2-4]。为实现快速且精准的机动目标跟踪，以状态估计为核心的机动目标跟踪往往需要同时考虑点目标机动运动建模和高精度滤波器设计两方面[5]。

目标机动运动建模旨在构造动力学方程使其尽量准确匹配目标实际运动。然而，由于目标机动难以预测，构建完全匹配或理想的机动运动模型十分困难。对于弱机动目标而言，利用诸如 Singer 模型、当前统计模型及 Jerk 模型等经典机动统计模型可以取得满意的跟踪效果[6]。然而，随着机动目标的运动轨迹日益复杂，单个状态空间模型难以及时准确地辨识机动参数，多模型 (multiple model, MM) 方法逐渐引起人们的重视[7]。该方法考虑目标在不同运动模式/标称模型之间进行切换，将这种模式演变建模为马尔可夫过程，以更准确地描述目标的运动轨迹[8]。

滤波器的设计是决定机动目标跟踪精度的另一关键因素。对于单模型而言，主要是依据目标运动建模的不同，以卡尔曼滤波为基础，构造相应的自适应滤波算法，如可调白噪声算法、变维滤波算法、输入估计算法等，来在线递推估计目标状态[9-11]。但对于目标强机动运动而言，单一的运动模型通常不足以刻画目标的机动特性，因此，MM 方法成为主流[12]。早期的 MM 方法没有考虑不同模式下估计的交互，只是将所有模式的滤波输出进行加权综合，造成每个模式下的滤波输入并不可靠，存在很大的局限性。相比之下，交互式多模型 (interacting multiple model, IMM) 算法[5] 考虑了目标可能处于多种运动状态，分别建立多个运动子模型，并考虑子模型之间的交互作用，以达到对机动目标的自适应跟踪。IMM 算法实现了状态估计与

收稿日期：2023-11-05；**录用日期**：2024-01-15；**网络出版时间**：2024-01-23 18:11
网络出版地址：link.cnki.net/urlid/11.2625.V.20240119.1110.001
基金项目：国家自然科学基金 (61903299)；陕西省自然科学基础研究计划资助项目 (2020JQ-842)
* **通信作者**. E-mail：qinyuemei@xupt.edu.cn

引用格式：秦月梅，张荣华，杨衍波，等. 广义未知扰动下多模型最小上限滤波 [J]. 北京航空航天大学学报，2026，52（1）：129-139.
QIN Y M，ZHANG R H，YANG Y B，et al. Multiple model minimum upper bound filter under generalized unknown disturbances [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：129-139（in Chinese）.

计算复杂度的有效折中，但其假设模型集先验已知且固定不变，当模型集没有包含目标机动运动的所有可能运动模式时，可能会导致其滤波精度退化。变结构多模型在IMM基础上根据目标机动的实际情况，自适应地调整当前的有效模型集合，进一步增强了目标状态估计的鲁棒性，以适应实际复杂环境[13]。

上述方法在系统建模时都假设所获得的量测仅受与传感器精度相关的高斯白噪声的影响。然而，在实际机动目标跟踪过程中往往会伴随电子对抗、干扰欺骗、甚至传感器故障等多种因素，使得量测总是会受到广义未知扰动(generalized unknown disturbance, GUD)的污染。这类扰动由于缺乏足够的先验信息而难以进行有效建模，往往表征为先验标称量测模型上施加了无先验信息的未知扰动，从而造成跟踪系统量测模型与实际模型失配。此时，无论现有的单模型自适应滤波还是MM滤波方法都不再适用。针对GUD下离散时间随机系统，Qin等通过求解状态预测误差协方差、新息协方差和估计误差协方差矩阵的最小上界，构造了最小上限滤波(minimum upper bound filter, MUBF)，以有效抑制滤波误差的峰值[14-17]，但并未考虑系统建模包含模式不确定、无法直接应用于强机动目标跟踪。考虑将MUBF作为MM方法的子滤波器，但GUD的存在使得模型概率更新难以实现，且固定的转移概率矩阵无法反映目标真实的运动状态，从而导致算法的稳定性和跟踪精度严重下降[18-19]。

基于上述讨论，本文针对GUD和模式不确定耦合共存下机动目标跟踪问题，提出一种多模型最小上限滤波(multiple model minimum upper bound filter, MMUBF)算法。首先，在多模型框架下并行运行MUBF以获得各模式下的高精度状态估计；其次，根据各模式滤波结果求得GUD近似解，将其引入似然值计算提升模型概率更新的准确性；最后，根据后验信息计算修正因子，自适应调整转移概率矩阵以保证模型概率更新的实时性。

1 问题描述

在博弈对抗环境下的机动目标跟踪中，目标机动运动通常建模为跳变马尔可夫过程，而传感器偏差、欺骗干扰和线性化近似误差等导致量测模型中存在时变GUD。此时，跳变马尔可夫参数和GUD共存，这促使考虑GUD下离散时间跳变马尔可夫线性系统：

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{F}_{\Theta_{k-1}}\boldsymbol{x}_{k-1} + \boldsymbol{G}_{\Theta_{k-1}}\boldsymbol{w}_{k-1} \\ \boldsymbol{z}_k = \boldsymbol{H}_{\Theta_k}\boldsymbol{x}_k + \boldsymbol{A}_{\Theta_k}\boldsymbol{\delta}_k + \boldsymbol{D}_{\Theta_k}\boldsymbol{v}_k \end{cases} \tag{1}$$

式中：$\boldsymbol{x}_k \in \mathbf{R}^n$($\mathbf{R}$代表实数向量构成的空间)为$k$时刻的状态向量；$\boldsymbol{z}_k \in \mathbf{R}^m$为$k$时刻的量测向量；$\boldsymbol{F}_{\Theta_{k-1}}$、$\boldsymbol{G}_{\Theta_{k-1}}$、$\boldsymbol{H}_{\Theta_k}$、$\boldsymbol{A}_{\Theta_k}$和$\boldsymbol{D}_{\Theta_k}$为具有合适维数的已知系统矩阵；$\boldsymbol{w}_k \in \mathbf{R}^r$为均值为$\mathbf{0}$，协方差为$\boldsymbol{Q}_k$的高斯白噪声序列；$\boldsymbol{v}_k \in \mathbf{R}^s$为均值为$\mathbf{0}$，协方差为$\boldsymbol{R}_k$的高斯白噪声序列，且对于任意时刻的$k$和$l$，$E(\boldsymbol{w}_k\boldsymbol{v}_l^{\mathrm{T}}) = \mathbf{0}$；$\{\Theta_k\}$表示具有有限状态空间$\{1,2,\cdots,M\}$的跳变马尔可夫链，且其转移概率矩阵$\boldsymbol{\Pi}$的第$(i,j)$个元素$\pi_{ij}$为

$$\pi_{ij} = P\{\Theta_k = j | \Theta_{k-1} = i\} \tag{2}$$

式中：$i,j = 1,2,\cdots,M$。

GUD的统计约束如下：

$$\begin{cases} E\{\boldsymbol{\delta}_k\boldsymbol{w}_{j-1}^{\mathrm{T}}\} = \mathbf{0}_{s\times p} \\ E\{\boldsymbol{\delta}_k\boldsymbol{v}_{j-1}^{\mathrm{T}}\} = \mathbf{0}_{s\times m} \qquad \forall j \geqslant k \end{cases} \tag{3}$$

式中：广义未知扰动$\boldsymbol{\delta}_k$的一般形式可表征为

$$\boldsymbol{\delta}_k = \lambda_1\boldsymbol{f}_{1,k} + \lambda_2\boldsymbol{f}_{2,k} + \lambda_3\boldsymbol{q}_k \tag{4}$$

其中：$\lambda_i \in \mathbf{R}(i = 1,2,3)$为加权系数；$\boldsymbol{f}_{1,k}$为具有动态特性的未知扰动，即以$\bar{\boldsymbol{W}}^{k-1} =: [\boldsymbol{w}_0^{\mathrm{T}}, \boldsymbol{w}_1^{\mathrm{T}}, \cdots, \boldsymbol{w}_{k-1}^{\mathrm{T}}]^{\mathrm{T}}$、$\bar{\boldsymbol{V}}^k =: [\boldsymbol{v}_1^{\mathrm{T}}, \boldsymbol{v}_2^{\mathrm{T}}, \cdots, \boldsymbol{v}_k^{\mathrm{T}}]^{\mathrm{T}}$和$\bar{\boldsymbol{\Delta}}^{k-1} =: [\boldsymbol{\delta}_1^{\mathrm{T}}, \boldsymbol{\delta}_2^{\mathrm{T}}, \cdots, \boldsymbol{\delta}_{k-1}^{\mathrm{T}}]^{\mathrm{T}}$为自变量的线性时变扰动，符号‘=:’表示定义；$\boldsymbol{f}_{2,k}$为任意的确定性时变扰动；$\boldsymbol{q}_k$为与$\bar{\boldsymbol{W}}^{k-1}$、$\bar{\boldsymbol{V}}^k$和$\bar{\boldsymbol{\Delta}}^{k-1}$不相关的随机扰动。

针对式(1)～式(4)，由于未考虑GUD的存在，现有面向机动目标跟踪的MM方法的性能往往会显著下降；同时，由于未考虑模式不确定性，现有针对单模型的最小上限滤波的估计性能会随着模型失配而明显退化。考虑博弈对抗环境下机动目标跟踪面临GUD和模式不确定的耦合共存，本文提出MMUBF算法，在多模型框架下进行目标状态估计以在线匹配目标的运动模式，选择MUBF作为模型条件滤波器以便在GUD下获得高精度滤波结果，通过补偿GUD并修正转移概率矩阵以减弱GUD对运动模式匹配的影响，从而实现机动目标的精准跟踪。

2 多模型最小上限滤波

考虑GUD和模式不确定耦合共存，本文MMUBF算法从子滤波器设计、量测似然构造、模式转移概率修正、模式后验概率更新及输出综合等方面展开设计。现给出MMUBF算法从k～$k+1$时刻的递推步骤。

步骤1 输入交互。假设每个子模式在k时刻的状态估计、估计误差协方差最小上限和模式后验概率分别为$\hat{\boldsymbol{x}}_{k|k}^j$、$\boldsymbol{P}_{k|k}^{j*}$和$\mu_k^j$，其中，$j$表示第$j$个子模式。对每个子模式进行交互计算，得到交互后的状态估计$\hat{\boldsymbol{x}}_{k|k}^{0j}$及其误差协方差矩阵$\boldsymbol{P}_{k|k}^{0j*}$：

$$\hat{\boldsymbol{x}}_{k|k}^{0j} = \sum_{i=1}^{M}\hat{\boldsymbol{x}}_{k|k}^{i}\mu_{k|k}^{ij} \tag{5}$$

$$\boldsymbol{P}_{k|k}^{0j*}=\sum_{i=1}^{M}\mu_{k|k}^{ij}\left[\boldsymbol{P}_{k|k}^{i*}+(\hat{\boldsymbol{x}}_{k|k}^{0j}-\hat{\boldsymbol{x}}_{k|k}^{i})(\hat{\boldsymbol{x}}_{k|k}^{0j}-\hat{\boldsymbol{x}}_{k|k}^{i})^{\mathrm{T}}\right] \tag{6}$$

$$\mu_{k|k}^{ij}=\pi_{ij}\mu_k^i\Big/\sum_{i=1}^{M}\pi_{ij}\mu_k^i \tag{7}$$

式中: M为模式总数; $\mu_{k|k}^{ij}$为当前时刻模型i到模型j的模式预测概率。

步骤 2　模型条件滤波。将交互输入计算的$\hat{\boldsymbol{x}}_{k|k}^{0j}$和$\boldsymbol{P}_{k|k}^{0j*}$作为子滤波器的输入执行 MUBF, 得到状态估计$\hat{\boldsymbol{x}}_{k+1|k+1}^{j}$、滤波残差$\boldsymbol{r}_{k+1}^{j}$、新息协方差上限$\boldsymbol{S}_{k+1}^{j*}$和误差协方差上限$\boldsymbol{P}_{k+1|k+1}^{j*}$。

对于所考虑跳变马尔可夫线性系统式 (1)～式 (4), 在模式j下, 相应卡尔曼滤波器如下:

$$\hat{\boldsymbol{x}}_{k+1|k}^{j}=\boldsymbol{F}_k^j\hat{\boldsymbol{x}}_{k|k}^{0j} \tag{8}$$

$$\boldsymbol{z}_{k+1|k}^{j}=\boldsymbol{H}_{k+1}\hat{\boldsymbol{x}}_{k+1|k}^{j} \tag{9}$$

$$\boldsymbol{r}_{k+1}^{j}=\boldsymbol{z}_{k+1}-\boldsymbol{z}_{k+1|k}^{j} \tag{10}$$

$$\hat{\boldsymbol{x}}_{k+1|k+1}^{j}=\hat{\boldsymbol{x}}_{k+1|k}^{j}+\boldsymbol{K}_{k+1}^{j}\boldsymbol{r}_{k+1}^{j} \tag{11}$$

由于 GUD 的存在, 导致滤波残差$\boldsymbol{r}_{k+1}^{j}$和新息协方差$\boldsymbol{S}_{k+1}^{j}=:E(\boldsymbol{r}_{k+1}^{j}(\boldsymbol{r}_{k+1}^{j})^{\mathrm{T}})$无法准确计算, 进而滤波增益$\boldsymbol{K}_{k+1}^{j}=\boldsymbol{P}_{k+1|k}^{j}\boldsymbol{H}_{k+1}^{\mathrm{T}}(\boldsymbol{S}_{k+1}^{j})^{-1}$无法有效获得。若要在每个时刻将 GUD 从系统方程中解耦, 需要的条件比较严苛, 通常情况下难以满足。考虑到 GUD 的存在会使估计误差协方差变大, 文献 [14-17] 通过构造 MUBF 以寻求估计误差协方差的最小上限, 从而实现估计误差峰值的有效抑制。

如果存在一系列正定矩阵$\boldsymbol{P}_{k+1|k}^{j*}$、$\boldsymbol{S}_{k+1}^{j*}$、$\boldsymbol{P}_{k+1|k+1}^{j*}$满足式 (12)～式 (14), 则如式 (8)～式 (11) 所示的滤波器被称为上限滤波。

$$\boldsymbol{P}_{k+1|k}^{j*}\geqslant\boldsymbol{P}_{k+1|k}^{j}=:E(\tilde{\boldsymbol{x}}_{k+1|k}^{j}(\tilde{\boldsymbol{x}}_{k+1|k}^{j})^{\mathrm{T}}) \tag{12}$$

$$\boldsymbol{S}_{k+1}^{j*}\geqslant\boldsymbol{S}_{k+1}^{j}=:E(\boldsymbol{r}_{k+1}^{j}(\boldsymbol{r}_{k+1}^{j})^{\mathrm{T}}) \tag{13}$$

$$\boldsymbol{P}_{k+1|k+1}^{j*}\geqslant\boldsymbol{P}_{k+1|k+1}^{j}=:E(\tilde{\boldsymbol{x}}_{k+1|k+1}^{j}(\tilde{\boldsymbol{x}}_{k+1|k+1}^{j})^{\mathrm{T}}) \tag{14}$$

式中: $\boldsymbol{P}_{k+1|k}^{j}$、$\boldsymbol{S}_{k+1}^{j}$、$\boldsymbol{P}_{k+1|k+1}^{j}$分别为状态预测误差协方差、新息协方差、估计误差协方差的理论值, $\boldsymbol{P}_{k+1|k}^{j*}$、$\boldsymbol{S}_{k+1}^{j*}$、$\boldsymbol{P}_{k+1|k+1}^{j*}$为相应的上限。其求解过程如下:

$$\boldsymbol{P}_{k+1|k}^{j*}=\boldsymbol{F}_k^j\boldsymbol{P}_{k|k}^{0j*}(\boldsymbol{F}_k^j)^{\mathrm{T}}+\boldsymbol{G}_k^j\boldsymbol{Q}_k(\boldsymbol{G}_k^j)^{\mathrm{T}} \tag{15}$$

$$\boldsymbol{S}_{k+1}^{j*}=\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}^{j*}\boldsymbol{H}_{k+1}^{\mathrm{T}}+\beta_{k+1}^{j\mathrm{opt}}\boldsymbol{A}_{k+1}\boldsymbol{\Phi}\boldsymbol{A}_{k+1}^{\mathrm{T}}+\boldsymbol{D}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{D}_{k+1}^{\mathrm{T}} \tag{16}$$

$$\boldsymbol{K}_{k+1}^{j}=\boldsymbol{P}_{k+1|k}^{j*}\boldsymbol{H}_{k+1}^{\mathrm{T}}(\boldsymbol{S}_{k+1}^{j*})^{-1} \tag{17}$$

$$\boldsymbol{P}_{k+1|k+1}^{j*}=(\boldsymbol{I}-\boldsymbol{K}_{k+1}^{j}\boldsymbol{H}_{k+1})\boldsymbol{P}_{k+1|k}^{j*}(\boldsymbol{I}-\boldsymbol{K}_{k+1}^{j}\boldsymbol{H}_{k+1})^{\mathrm{T}}+\boldsymbol{K}_{k+1}^{j}\boldsymbol{D}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{D}_{k+1}^{\mathrm{T}}(\boldsymbol{K}_{k+1}^{j})^{\mathrm{T}} \tag{18}$$

式中: $\boldsymbol{\Phi}$为设定的正定矩阵, 最优因子$\beta_{k+1}^{j\mathrm{opt}}$的求解可简化为

$$\boldsymbol{S}_{k+1}^{j}\leqslant\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}^{j*}\boldsymbol{H}_{k+1}^{\mathrm{T}}+\beta_{k+1}^{j}\boldsymbol{A}_{k+1}\boldsymbol{\Phi}\boldsymbol{A}_{k+1}^{\mathrm{T}}+\boldsymbol{D}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{D}_{k+1}^{\mathrm{T}} \tag{19}$$

式中: $\beta_{k+1}^{j}\geqslant0$。为求得新息协方差的最小上限, $\beta_{k+1}^{j\mathrm{opt}}$取$\beta_{k+1}^{j}$的最小可行解, 即$\beta_{k+1}^{j\mathrm{opt}}=\min(\beta_{k+1}^{j})$(详细推导过程见文献 [15])。用无偏估计$\hat{\boldsymbol{S}}_{k+1}^{j}=\boldsymbol{r}_{k+1}^{j}(\boldsymbol{r}_{k+1}^{j})^{\mathrm{T}}$近似代替未知的理论新息协方差矩阵$\boldsymbol{S}_{k+1}^{j}$。为保证本文 MMUBF 算法递推求解的实时性, 通过如下特征值求解近似最优因子$\beta_{k+1}^{j\mathrm{opt}}$:

$$\beta_{k+1}^{j\mathrm{opt}}=\max\left(0,\frac{\psi_{\max}(\boldsymbol{S}_{k+1}^{j0})}{\psi_{\min}(\boldsymbol{A}_{k+1}\boldsymbol{\Phi}\boldsymbol{A}_{k+1}^{\mathrm{T}})}\right) \tag{20}$$

式中: $\boldsymbol{S}_{k+1}^{j0}=\boldsymbol{S}_{k+1}^{j}-\boldsymbol{H}_{k+1}\boldsymbol{P}_{k+1|k}^{j*}\boldsymbol{H}_{k+1}^{\mathrm{T}}-\boldsymbol{D}_{k+1}\boldsymbol{R}_{k+1}\boldsymbol{D}_{k+1}^{\mathrm{T}}$; $\psi_{\max}$和$\psi_{\min}$分别为相应的最大和最小特征值。

步骤 3　模型概率更新。由于 GUD 的存在使得量测似然分布未知。为后续计算的简洁性, 采用高斯分布近似量测似然。由于理论新息协方差矩阵无法获得, 故本文利用其最小上限$\boldsymbol{S}_{k+1}^{j*}$代替。同时, 考虑直接计算获得的残差包含 GUD, 对其进行如下修正。在每个子模式下, 利用相应 MUBF 的滤波结果进行二次量测预测并与当前时刻量测作差求得扰动分量, 然后, 根据预测模式概率进行加权得到 GUD 近似解, 并对各模式计算的残差进行修正。具体计算如下:

$$\boldsymbol{z}_{k+1|k+1}^{j}=\boldsymbol{H}_{k+1}\hat{\boldsymbol{x}}_{k+1|k+1}^{j} \tag{21}$$

$$\boldsymbol{\varepsilon}_{k+1}^{j}=\boldsymbol{z}_{k+1}-\boldsymbol{z}_{k+1|k+1}^{j} \tag{22}$$

$$\boldsymbol{\varepsilon}_{k+1}=\sum_{j=1}^{M}\left(\boldsymbol{\varepsilon}_{k+1}^{j}\sum_{i=1}^{M}\pi_{ij}\mu_k^i\right) \tag{23}$$

模型j似然概率

$$\Lambda_{k+1}^{j}=\frac{1}{\sqrt{|2\pi\boldsymbol{S}_{k+1}^{j*}|}}\exp\left[-\frac{1}{2}(\boldsymbol{r}_{k+1}^{j}-\boldsymbol{\varepsilon}_{k+1})^{\mathrm{T}}\times(\boldsymbol{S}_{k+1}^{j*})^{-1}(\boldsymbol{r}_{k+1}^{j}-\boldsymbol{\varepsilon}_{k+1})\right] \tag{24}$$

模型j后验概率

$$\mu_{k+1}^{j}=\frac{\Lambda_{k+1}^{j}\sum_{i=1}^{M}\pi_{ij}\mu_k^i}{\sum_{j=1}^{M}\Lambda_{k+1}^{j}\sum_{i=1}^{M}\pi_{ij}\mu_k^i} \tag{25}$$

步骤 4　模式转移概率修正。计算各模式似然概率时需要真实的新息协方差$\boldsymbol{S}_{k+1}^{j}$, 由于 GUD 的存在, $\boldsymbol{S}_{k+1}^{j}$的理论值无法准确获得。在使用其最小上限$\boldsymbol{S}_{k+1}^{j*}$对其进行代替过程中, 掩饰了相应模型与实际系统之间的差异, 导致系统对模式切换不够敏感。考虑到模式转移概率会影响模型切换的实时性, 根据后验信息计算修正因子[18], 对模式转移概

率矩阵进行自适应修正，以提升本文算法的模式匹配精度和切换速度。

定义修正因子：

$$k_{ij}=\pi_{k+1}^{ji}\Lambda_{k+1}^{j}/(\pi_{k+1}^{ij}\Lambda_{k+1}^{i}) \tag{26}$$

当$i\neq j$时：

$$\pi'_{ij}=(k_{ij})^{\alpha}\pi_{ij} \tag{27}$$

式中：π'_{ij}为修正后的模型转移概率；$\alpha\in[0,1]$为调整系数，修正速度与该值大小正相关。

修正的转移概率矩阵主对角线元素如下：

$$\pi'_{ii}=1-(k_{i1})^{\alpha}\pi_{i1}-\cdots-(k_{iN})^{\alpha}\pi_{iN} \tag{28}$$

为确保主对角线元素占优，设置阈值T_{H}，进行概率修正后，针对小于T_{H}的主对角线元素进一步按照下式调整：

$$\pi'_{ii}=T_{\mathrm{H}} \tag{29}$$

$$\pi'_{ij}=(1-T_{\mathrm{H}})\frac{\pi_{ij}}{1-\pi_{ii}}\qquad i\neq j \tag{30}$$

$$\pi_{ij}=\pi'_{ij} \tag{31}$$

步骤 5 估计综合。最终输出的状态估计$\hat{\boldsymbol{x}}_{k+1|k+1}$和估计误差协方差最小上限$\boldsymbol{P}^{*}_{k+1|k+1}$由更新的后验模式概率和各子模式下状态估计综合获得，即：

$$\hat{\boldsymbol{x}}_{k+1|k+1}=\sum_{j=1}^{M}\mu_{k+1}^{j}\hat{\boldsymbol{x}}_{k+1|k+1}^{j} \tag{32}$$

$$\boldsymbol{P}^{*}_{k+1|k+1}=\sum_{j=1}^{M}\mu_{k+1}^{j}\left[\boldsymbol{P}^{j*}_{k+1|k+1}+(\hat{\boldsymbol{x}}_{k+1|k+1}-\hat{\boldsymbol{x}}^{j}_{k+1|k+1})\times(\hat{\boldsymbol{x}}_{k+1|k+1}-\hat{\boldsymbol{x}}^{j}_{k+1|k+1})^{\mathrm{T}}\right] \tag{33}$$

特别地，当式 (1) 中的$\boldsymbol{\delta}_k=\boldsymbol{0}$($\boldsymbol{0}$是具有适当维数的零矩阵) 且量测方程中没有$\boldsymbol{\delta}_k$时，由于量测噪声的存在，凸优化求解得到的最优因子$\beta_{k+1}^{\mathrm{jopt}}$近似于0，但不等于0，导致计算得到状态预测误差协方差$\boldsymbol{P}^{j*}_{k+1|k}$、新息协方差$\boldsymbol{S}^{j*}_{k+1}$、估计误差协方差$\boldsymbol{P}^{j*}_{k+1|k+1}$仍是上限值，但与理论值非常接近。因此，当$\boldsymbol{\delta}_k=\boldsymbol{0}$时，本文 MMUBF 算法与 IMM 算法近似，但并不等同于 IMM 算法。

为便于理解，图 1 给出了本文 MMUBF 算法的递推结构框图。MMUBF 算法实现步骤如下。

1. 初始化

设置$\hat{\boldsymbol{x}}^{i}_{k|k}$，$\boldsymbol{P}^{i*}_{k|k}$，$\boldsymbol{\Pi}$，$\boldsymbol{F}^{i}_{k}$，$\boldsymbol{H}_{k+1}$，$\boldsymbol{Q}_k$，$\boldsymbol{G}^{j}_{k}$，$\boldsymbol{D}_{k+1}$，$\boldsymbol{R}_{k+1}$，$\boldsymbol{\Phi}$，$\alpha$，$T_{\mathrm{H}}$，$\mu^{i}_{k}$的初始值。

2. 输入交互

a) 计算预测模式概率，由式 (5) 得$\mu^{ij}_{k|k}$。

b) 计算交互后的状态估计和估计误差协方差，由式 (6) 和式 (7) 分别得$\hat{\boldsymbol{x}}^{0j}_{k|k}$、$\boldsymbol{P}^{0j*}_{k|k}$。

3. 模型条件滤波

a) 预测。由式 (8)、式 (9) 和 (15) 分别得$\hat{\boldsymbol{x}}^{j}_{k+1|k}$、$\boldsymbol{z}^{j}_{k+1|k}$和$\boldsymbol{P}^{j*}_{k+1|k}$。

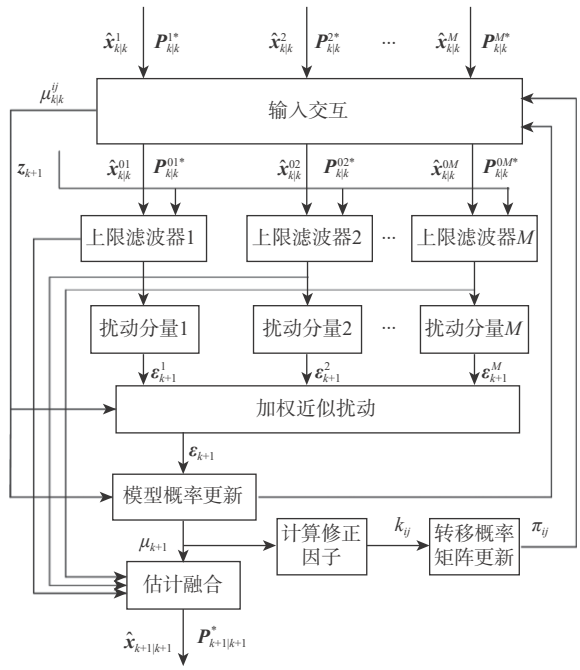


图 1 MMUBF 算法递推结构框图

Fig. 1 Recursive block diagram of MMUBF algorithm

b) 计算滤波残差。由式 (10) 得$\boldsymbol{r}^{j}_{k+1}$。

c) 凸优化。

1) 求解最优因子，由式 (20) 得$\beta_{k+1}^{\mathrm{jopt}}$。

2) 由式 (16) 得$\boldsymbol{S}^{j*}_{k+1}$。

3) 由式 (17) 得$\boldsymbol{K}^{j}_{k+1}$。

d) 状态估计。由式 (11) 和式 (18) 分别得$\hat{\boldsymbol{x}}^{j}_{k+1|k+1}$、$\boldsymbol{P}^{j*}_{k+1|k+1}$。

4. 模式概率更新

a) 近似 GUD。由式 (21)～式 (23) 得$\boldsymbol{\varepsilon}_{k+1}$。

b) 计算似然概率。由式 (24) 得Λ^{j}_{k+1}。

c) 计算模式概率。由式 (25) 得μ^{j}_{k+1}。

5. 模式转移概率修正

a) 计算修正因子。由式 (26) 得k_{ij}。

b) 计算主对角线元素。由式 (28) 和式 (29) 得π'_{ii}。

c) 计算非主对角线元素。由式 (27) 和式 (30) 得π'_{ij}。

d) 更新模型转移概率。由式 (31) 得π_{ij}。

6. 输出综合

由式 (32) 和式 (33) 分别得$\hat{\boldsymbol{x}}_{k+1|k+1}$、$\boldsymbol{P}^{*}_{k+1|k+1}$。

表 1 为从k～$k+1$时刻中每个步骤的浮点运算

表 1 每个时刻 MMUBF 算法各步骤的计算复杂度

Table 1 Computational complexity of each step of MMUBF algorithm at each moment

步骤	浮点运算数量
1	$5M^2+3M^2n^2+3M^2n-3Mn^2-3Mn-3M$
2	$M(8n^3+8mn^2+4m^2n+10m^3+5m^2-2n^2-n-3mn)$
3	$2Mmn+Mm^3+2Mm^2+5Mm+M^2-M-m$
4	$2M^2+2M$
5	$2Mn^2+4Mn-2n^2-4n$

数量, 讨论了本文 MMUBF 算法的计算复杂度[20]。若 $\boldsymbol{A}\in\mathbf{R}^{n\times m}$, $\boldsymbol{B}\in\mathbf{R}^{n\times m}$, $\boldsymbol{A}\pm\boldsymbol{B}$的浮点运算数量为$nm$; 若$\boldsymbol{A}\in\mathbf{R}^{n\times m}$, $\boldsymbol{B}\in\mathbf{R}^{m\times l}$, $\boldsymbol{AB}$的浮点运算数量为$2mnl-nl$。此外, $\boldsymbol{A}\in\mathbf{R}^{n\times n}$的逆运算的浮点运算数量为$n^3$。

3　仿真实验

考虑博弈对抗环境下伴随 GUD 的机动目标跟踪。目标在二维平面 Oxy 上运动, 初始位置为 $[10\,000, 10\,000]^{\mathrm{T}}$ m、初始速度为 $[50, 50]^{\mathrm{T}}$ m/s, 其在 0 ~ 85 s做转弯角速度为 4 (°)/s 的匀速右转弯运动、86 ~ 100 s做匀速直线运动、101 ~ 170 s做转弯角速度为 4 (°)/s 的匀速左转弯运动、171 ~ 200 s做匀速直线运动, 目标运动轨迹如图 2 所示。

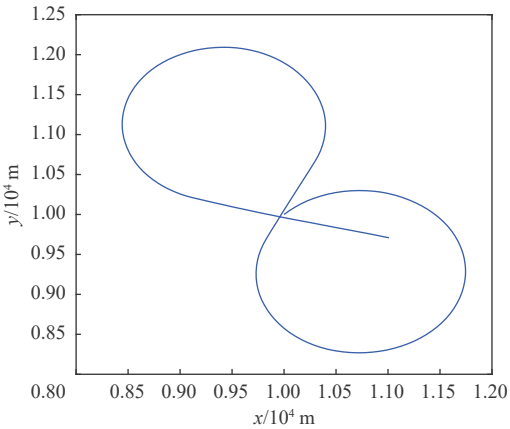


图 2　目标运动轨迹

Fig. 2　Target moving trajectory

在匀速右转弯和匀速左转弯运动模型中, 系统参数如下:

$$\begin{cases}\boldsymbol{F}_{\mathrm{CT}}=\begin{bmatrix}1 & [\sin(\omega T)]/\omega & 0 & -[1-\cos(\omega T)]/\omega\\ 0 & \cos(\omega T) & 0 & -\sin(\omega T)\\ 0 & [1-\cos(\omega T)]/\omega & 1 & [\sin(\omega T)]/\omega\\ 0 & \sin(\omega T) & 0 & \cos(\omega T)\end{bmatrix}\\ \boldsymbol{G}_{\mathrm{CT}}=\boldsymbol{I}_2\otimes\begin{bmatrix}T^2/2\\ T\end{bmatrix}\end{cases}\tag{34}$$

在匀速直线运动模型中, 系统参数如下:

$$\begin{cases}\boldsymbol{F}_{\mathrm{CV}}=\boldsymbol{I}_2\otimes\begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}\\ \boldsymbol{G}_{\mathrm{CV}}=\boldsymbol{I}_2\otimes\begin{bmatrix}T^2/2\\ T\end{bmatrix}\end{cases}\tag{35}$$

式中: ω =4 (°)/s 表示左转弯, $\omega=-4$ (°)/s 表示右转弯; $\otimes$表示克罗内克积; $\boldsymbol{I}$为具有适当维数的单位矩阵。采样周期$T=1$ s, 各模式过程噪声协方差皆为 $\boldsymbol{Q}=q\times\mathrm{diag}\{0.1, 0.1\}$且$q=1$。

该目标被位于平面坐标原点的雷达探测并获得位置量测。本文量测模型参数如下:

$$\begin{cases}\boldsymbol{H}_k=\boldsymbol{I}_2\otimes[1\quad 0]\\ \boldsymbol{D}_k=\boldsymbol{I}_2\end{cases}\tag{36}$$

同时, 量测噪声协方差 $\boldsymbol{R}=\sigma^2\times\mathrm{diag}\{100^2, 100^2\}$, 且$\sigma=1$。进一步地, 考虑博弈对抗环境, 雷达量测受如图 3 所示的时变未知扰动的影响, 且系数矩阵 $\boldsymbol{A}_k=\boldsymbol{I}_2$。时变未知扰动的具体形式为

$$\boldsymbol{\delta}_k=\begin{cases}[2\,000\quad -2\,000] & 0<k\leqslant 15\\ [-2\,000\quad 2\,000] & 30<k\leqslant 40\\ [1\,600\sin(0.06\pi k)\quad -1\,600\cos(0.1\pi k)] & 60<k\leqslant 75\\ [-1\,200\cos(0.2\pi k)\quad 1\,200\sin(0.2\pi k)] & 85<k\leqslant 100\\ [1\,600\sin(0.2\pi k)\quad -1\,600\cos(0.2\pi k)] & 110<k\leqslant 120\\ [-1\,500\cos(0.2\pi k)\quad 1\,500\sin(0.2\pi k)] & 135<k\leqslant 150\\ [1\,500\quad -1\,500] & 170<k\leqslant 180\\ [0\quad 0] & \text{其他}\end{cases}$$

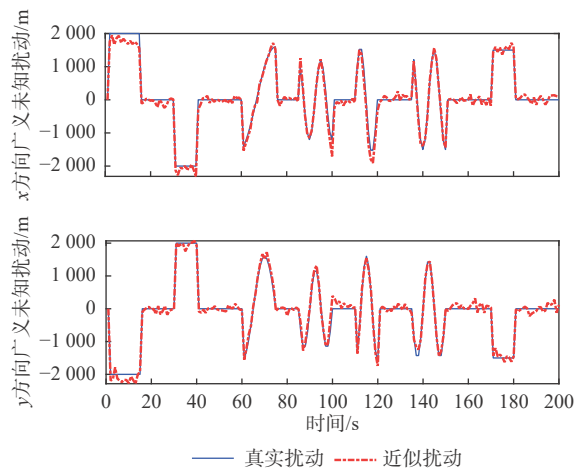


图 3　广义未知扰动真值及估计值示意图

Fig. 3　Illustration of true and estimated values of GUD

本文 MMUBF 算法与标准 IMM 算法、自适应 IMM 算法[18] (简记为 AIMM 算法)、灰狼优化算法改进后的自适应 IMM 算法[19](简记为 AGIMM 算法) 及基于匀速右转弯运动模型[6] 的 MUBF 算法[17] (简记为 rtMUBF 算法) 进行仿真对比。本文假设不同算法的模式集皆选择匀速直线运动、匀速左转弯运动和匀速右转弯运动, 且转弯角速度已知。同时, 在 MMUBF 算法中, 调整系数$\alpha=1$, 概率修正阈值$T_{\mathrm{H}}=0.9$。在 IMM 算法、AIMM 算法、MMUBF 算法和 AGIMM 算法中, 模式初始概率皆为 1/3, 转移概率$\pi_{ii}=0.7$、$\pi_{ij}=0.15(i\neq j)$。

3.1　不同算法滤波精度对比分析

基于 1 000 次蒙特卡罗仿真, 图 4 为各算法位置和速度估计的均方根误差 (root mean square errors, RMSE) 对比结果。表 2 为 x 轴和 y 轴位置、速度估计 RMSE 均值 (average RMSE, ARMSE)。图 3 为单次蒙特卡罗仿真下本文算法估计的扰动近似。

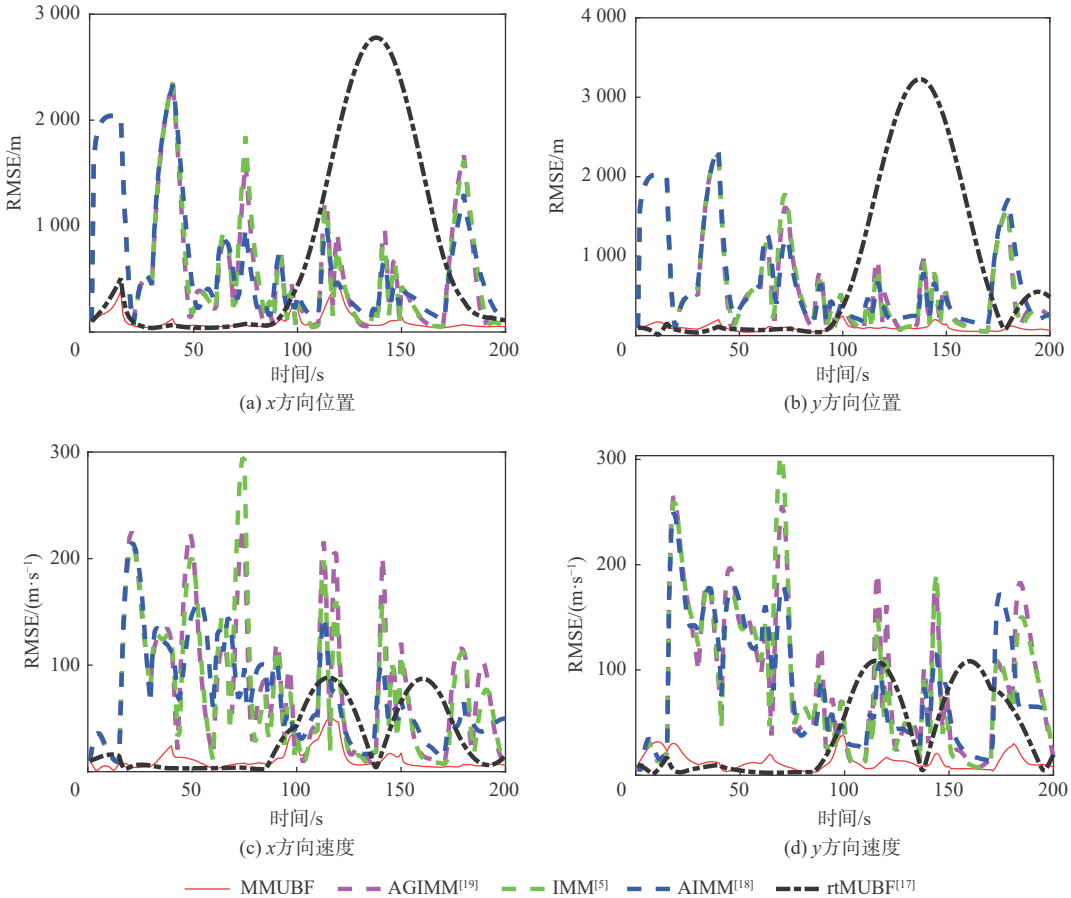


图4 各算法均方根误差对比

Fig. 4 Comparison of RMSE of different algorithms

表2 各算法均方根误差均值对比

Table 2 Comparison of ARMSE of different algorithms

算法	位置ARMSE/m		速度ARMSE/(m·s^{-1})	
	*x*方向	*y*方向	*x*方向	*y*方向
IMM[5]	653.50	637.60	78.17	86.44
AGIMM[19]	642.20	631.20	80.06	82.76
AIMM[18]	623.20	645.80	70.72	80.03
rtMUBF[17]	743.70	782.50	31.82	38.04
MMUBF	119.10	111.40	13.13	14.75

由图4可知，无论是位置还是速度估计，本文MMUBF算法的RMSE都要远低于其他4种对比算法。其中，前85 s单模型下rtMUBF算法的估计误差很小，这是由于目标当前正在做右转弯运动，模型匹配度较高。而85 s后目标发生机动，此时由于rtMUBF算法不具备自适应匹配模型的能力，因此，其估计误差骤增。AGIMM算法、IMM算法和AIMM算法的整体误差都比较大，主要原因在于三者的子滤波器对GUD缺乏鲁棒性。由于同时考虑了GUD和机动模式不确定的耦合影响，本文MMUBF算法具有高精度的跟踪性能。同时，如图3所示，本文MMUBF算法对GUD的近似与真实扰动的拟合度非常高，证明本文算法近似GUD的方法是有效的。

此外，由表2可知，本文MMUBF算法的位置和速度估计的ARMSE均远小于其他4种对比算法，这也证明了本文MMUBF算法可以获得高精度的机动目标状态估计结果。

本文通过统计各算法在单次蒙特卡罗下的平均运行时间完成对算法计算复杂度的验证。如表3所示，本文MMUBF算法相对于其他算法在运行时间无明显增加的情况下大幅提升了跟踪精度，取得了良好的跟踪效果。

表3 各算法单次蒙特卡罗仿真的平均运行时间对比

Table 3 Comparison of average running time of single Monte Carlo simulation for each algorithm

算法	运行时间/ms
IMM[5]	45.005
AGIMM[19]	73.663
AIMM[18]	63.025
rtMUBF[17]	15.160
MMUBF	72.487

3.2 $\delta_k=0$时不同算法滤波精度对比分析

为验证$\delta_k=0$时本文MMUBF算法的性能。基于1 000次蒙特卡罗仿真，图5为各算法位置估计的RMSE算法对比结果。如图所示，当$\delta_k=0$时，本文MMUBF算法、AGIMM算法、IMM算法及AIMM

算法的跟踪误差都很小, 皆远小于 rtMUBF 算法, 这是由于四者的模型匹配度比较高。相比之下, 本文 MMUBF 算法、AGIMM 算法和 AIMM 算法的精度要略高于 IMM 算法, 这是由于三者通过调整模型转移概率可以更快地实现模型切换。

3.3　不同量测噪声水平下滤波精度分析

基于 1 000 次蒙特卡罗仿真, 图 6 为 5 种算法在不同量测噪声水平下的 ARMSE 曲线。其中, $\sigma = 1.1, 1.2, \cdots, 2.0$ 且每次递增 0.1。如图 6 所示, 随着 σ 的增大, 即量测噪声逐渐增大, 各算法的 ARMSE 都有所增加但都比较平缓。这是因为此时量测噪声仍小于 GUD, 算法性能由 GUD 和量测噪声共同决定。同时, 由于考虑了 GUD 和机动模式不确定的耦合共存, 本文 MMUBF 算法的跟踪精度一直都远高于对比算法。这表明本文 MMUBF 算法可以有效解决 GUD 下机动目标跟踪峰值误差过大问题, 且对量测噪声水平的变化具有一定的鲁棒性。

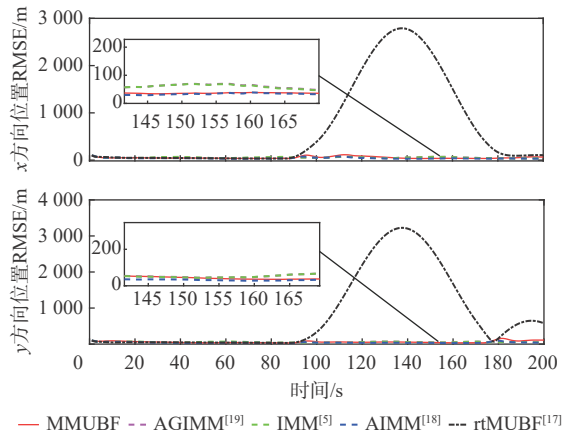


图 5　位置估计均方根误差对比

Fig. 5　Comparison of RMSEs of estimated positions

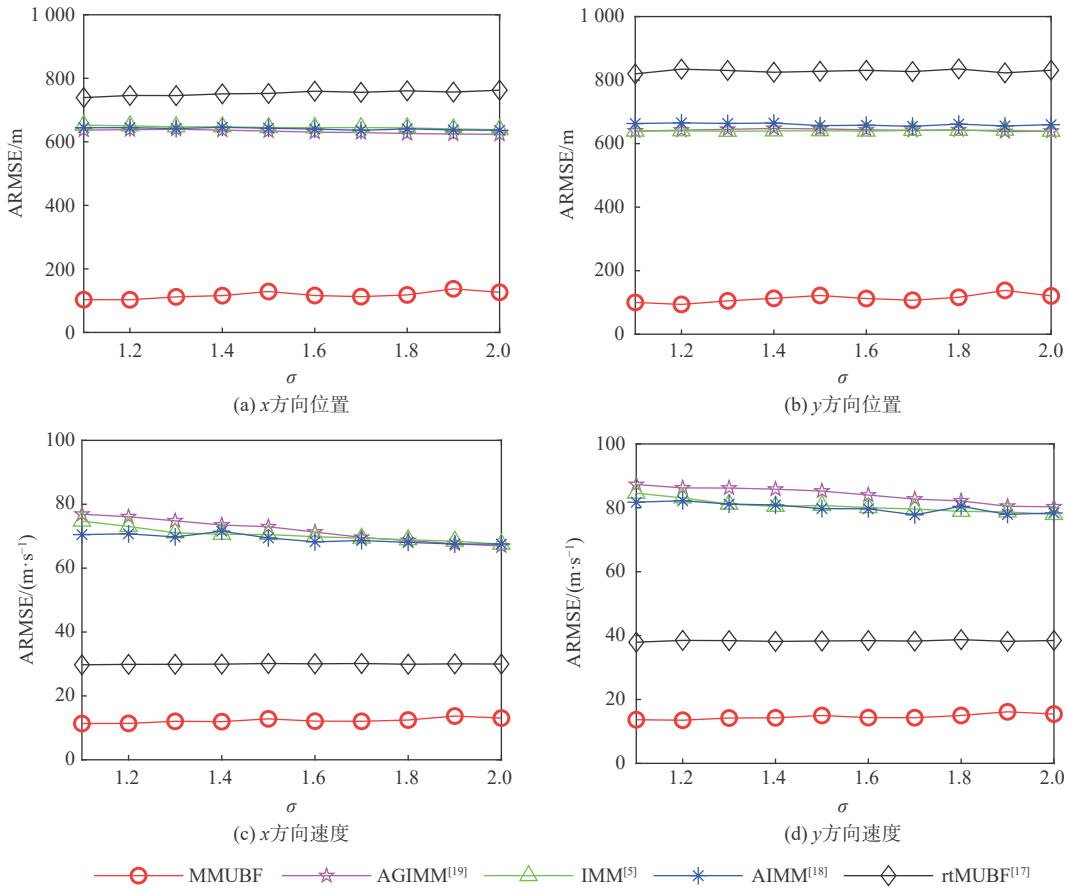


图 6　不同量测噪声水平下各算法均方根误差均值对比

Fig. 6　Comparison of ARMSE of each algorithms with different levels of measurement noises

3.4　不同过程噪声水平下滤波精度分析

进一步地, 基于 1 000 次蒙特卡罗仿真, 图 7 给出了 5 种对比算法在不同过程噪声水平下的 ARMSE 曲线。其中, $q = 1, 2, \cdots, 10$ 且 $q \in \mathbf{N}$。

由图 7 可知, 随着 q 值的不断增大, 即过程噪声整体水平越来越大, 所有算法的 ARMSE 都有所增大, 但本文算法的估计性能始终保持领先地位。这进一步证实了本文算法可以获得高精度的机动目标状态估计结果。

3.5　不同概率修正阈值下滤波精度分析

此外, 基于 1 000 次蒙特卡罗仿真, 图 8 给出了 5 种算法在不同概率修正阈值下的 ARMSE 曲线。其中, $T_{\mathrm{H}} = 0.60, 0.65, \cdots, 0.95$ 且每次递增 0.05。由图 8 可知, 随着概率修正阈值的不断增大, 本文 MMUBF

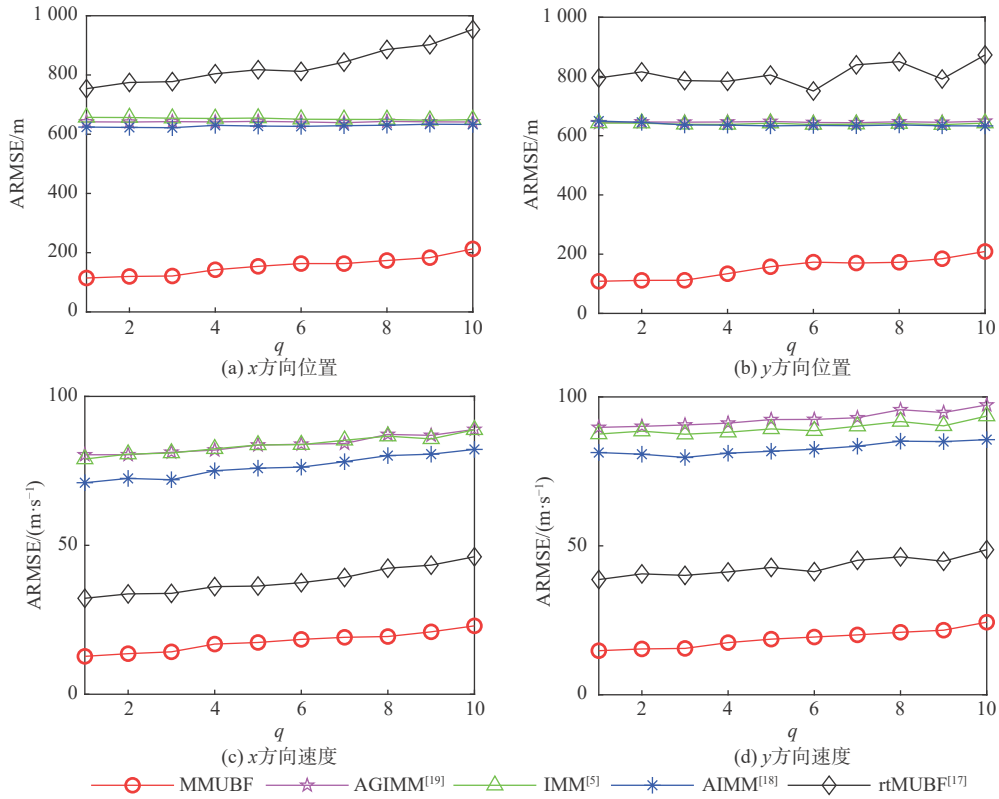


图7 不同过程噪声水平下各算法均方根误差均值对比

Fig. 7 Comparison of ARMSEs of each algorithms with different levels of process noises

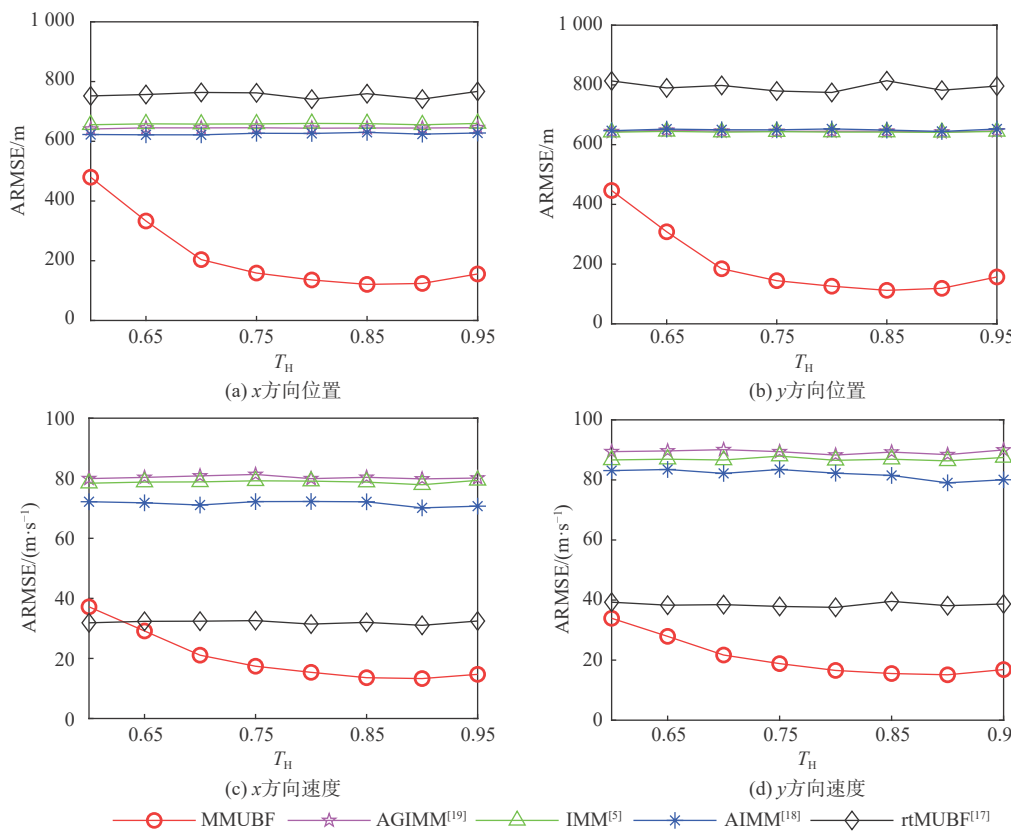


图8 不同模式转移概率修正阈值下各算法均方根误差均值对比

Fig. 8 Comparison of ARMSE of each algorithms with different mode transition probability amending thresholds

的 ARMSE 先减小后增大，在$T_H = 0.90$时精度最高。这是由于修正阈值越大，模式的切换速度就越快，从而跟踪精度就越高。但切换过快也会导致模式失配，故概率修正阈值的选取应在[0.85, 0.95]范围内。

3.6　不同调整系数下滤波精度分析

基于 1 000 次蒙特卡罗仿真, 图 9($T_H = 0.90$)给出了 5 种算法在不同调整系数下的 ARMSE 曲线。其中, $\alpha = 0.10, 0.20, \cdots, 1.0$且每次递增0.10。

由图 9 可知, $T_H = 0.90$时, 概率转移矩阵的主对角线元素对模式切换占主导地位, 非主对角线元素占比很少, 故随着调整系数α逐渐增大, 本文 MMUBF 算法的 ARMSE 曲线没有明显变化。此时, α的选取对算法精度影响很小。

为进一步验证α对模式概率修正速度的影响, 在$T_H = 0.70$时进行不同α下 MMUBF 算法滤波精度分析。如图 10 所示, 随着α逐渐增大, 模式切换越迅速, 模

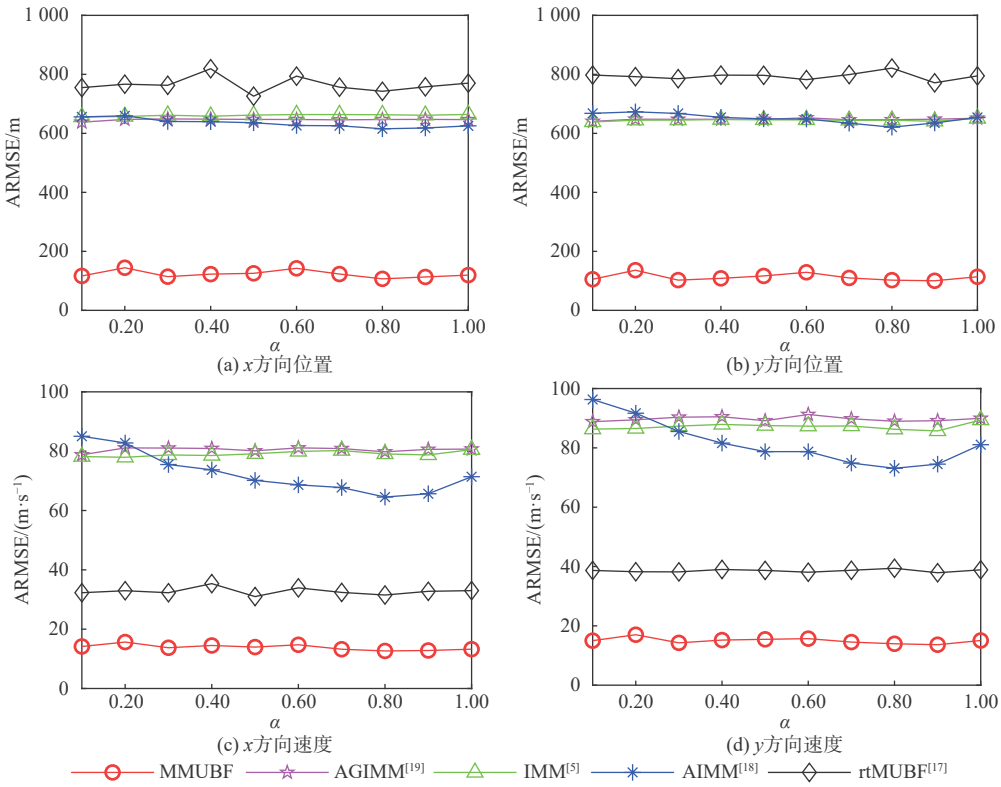


图 9　不同调整系数下各算法均方根误差均值对比

Fig. 9　Comparison of ARMSEs of each algorithms with different adjustment coefficients

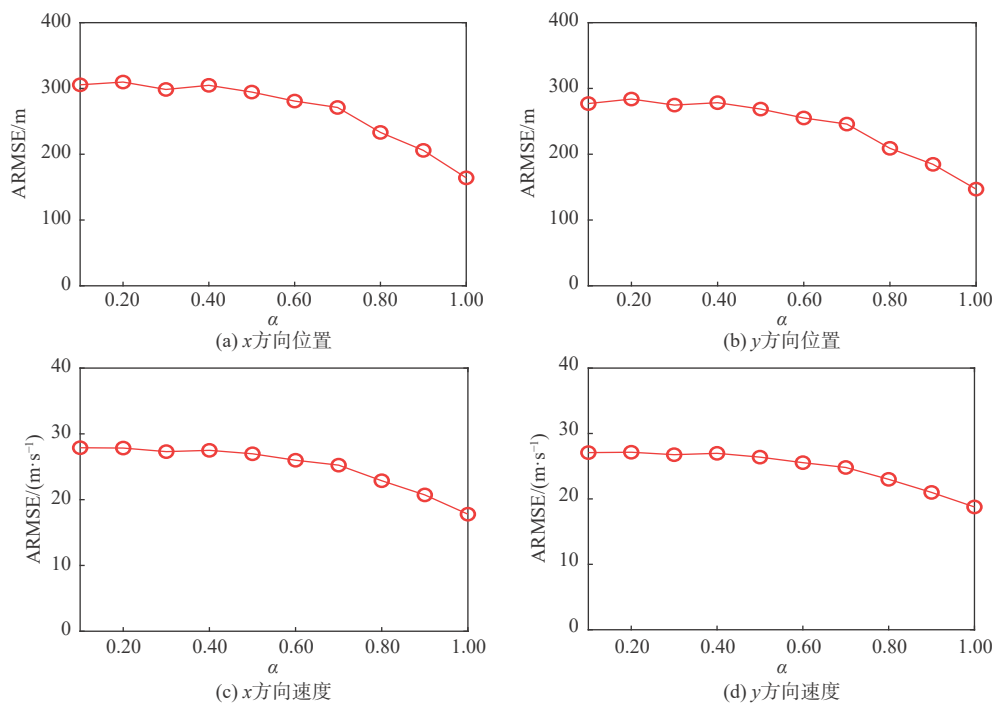


图 10　不同α下 MMUBF 算法均方根误差均值

Fig. 10　ARMSEs of MMUBF algorithm with different values of α

式匹配度越高，本文算法的误差越小。综合图9和图10，α的选取应越大越好。

4 结 论

1）针对GUD下机动目标跟踪峰值误差过大的问题，本文提出一种MMUBF算法。该算法在多模型框架下对不同标称模型进行MUBF，提高了各模型的状态估计精度。

2）利用MUBF估计结果计算扰动分量，结合先验模型概率进行加权得到GUD的近似解，将其引入似然概率计算提高了模型匹配精度。同时，根据后验信息添加修正因子自适应调整转移概率矩阵，提高了本文算法的模式切换速度。

3）经仿真验证，本文算法同时兼顾了MM算法和MUBF算法的优点，在保持计算效率的同时，可以有效解决GUD下机动目标跟踪误差峰值过大问题，获得高精度的状态估计结果。

致谢 感谢西北工业大学刘准钆教授对本文工作中算法模型的构建及论文的修改提出的宝贵意见。

参考文献（References）

[1] 王昱淇, 卢宙, 蔡云泽. 基于一致性的分布式变结构多模型方法[J]. 自动化学报, 2021, 47(7): 1548-1557.
WANG Y Q, LU Z, CAI Y Z. Consensus-based distributed variable structure multiple model[J]. Acta Automatica Sinica, 2021, 47(7): 1548-1557(in Chinese).

[2] SU J, LI Y, ALI W. Underwater passive manoeuvring target tracking with isogradient sound speed profile[J]. IET Radar, Sonar & Navigation, 2022, 16(9): 1415-1433.

[3] ZHENG Z, CAI S C. A collaborative target tracking algorithm for multiple UAVs with inferior tracking capabilities[J]. Frontiers of Information Technology & Electronic Engineering, 2021, 22(10): 1334-1350.

[4] LI B, YANG Z P, CHEN D Q, et al. Maneuvering target tracking of UAV based on MN-DDPG and transfer learning[J]. Defence Technology, 2021, 17(2): 457-466.

[5] MA J, GUO X T. Combination of IMM algorithm and ASTRW-CKF for maneuvering target tracking[J]. IEEE Access, 2020, 8: 143095-143103.

[6] 何友, 修建娟, 刘瑜, 等. 雷达数据处理及应用[M]. 4版. 北京: 电子工业出版社, 2022.
HE Y, XIU J J, LIU Y, et al. Radar data processing with applications[M]. 4th ed. Beijing: Publishing House of Electronics Industry, 2022(in Chinese).

[7] MA Y J, ZHAO S Y, HUANG B. Multiple-model state estimation based on variational Bayesian inference[J]. IEEE Transactions on Automatic Control, 2019, 64(4): 1679-1685.

[8] SHEN C, XU D J, HUANG W, et al. An interacting multiple model approach for state estimation with non-Gaussian noise using a variational Bayesian method[J]. Asian Journal of Control, 2015, 17(4): 1424-1434.

[9] GOFF G M, BLACK J T, BECK J A. Orbit estimation of a continuously thrusting spacecraft using variable dimension filters[J]. Journal of Guidance, Control, and Dynamics, 2015, 38(12): 2407-2420.

[10] LU P. Comparative study of state and unknown input estimation for continuous-discrete stochastic systems[J]. Franklin Open, 2022, 1: 1-8.

[11] ANGELI A, DESMET W, NAETS F. Deep learning of multibody minimal coordinates for state and input estimation with Kalman filtering[J]. Multibody System Dynamics, 2021, 53(2): 205-223.

[12] 肖楚晗, 李炯, 雷虎民, 等. 基于AVSIMM算法的高超声速再入滑翔目标跟踪[J]. 北京航空航天大学学报, 2019, 45(2): 413-421.
XIAO C H, LI J, LEI H M, et al. Hypersonic reentry gliding target tracking based on AVSIMM algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(2): 413-421(in Chinese).

[13] HUANG Y H, WU Y, YAO L X, et al. A class of distributed variable structure multiple model algorithm based on posterior information of information matrix[J]. Journal of Shanghai Jiaotong University (Science), 2022, 27(5): 671-679.

[14] QIN Y M, LIANG Y, YANG Y B, et al. Adaptive upper-bound linear mean square error filter of Markovian jump linear systems with generalized unknown disturbances[C]//Proceedings of the 18th International Conference on Information Fusion. Piscataway: IEEE Press, 2015: 1833-1839.

[15] QIN Y M, LIANG Y, YANG Y B, et al. Minimum upper-bound filter of Markovian jump linear systems with generalized unknown disturbances[J]. Automatica, 2016, 73: 56-63.

[16] YANG Y B, QIN Y M, LIANG Y, et al. Adaptive filter for linear systems with generalized unknown disturbance in measurements [C]//Proceedings of the 16th International Conference on Information Fusion. Piscataway: IEEE Press, 2013: 1336-1341.

[17] QIN Y M, LIANG Y, YANG Y B, et al. Adaptive filter of non-linear systems with generalised unknown disturbances[J]. IET Radar, Sonar & Navigation, 2014, 8(4): 307-317.

[18] 叶瑾, 许枫, 杨娟, 等. 一种改进的时变转移概率AIMM跟踪算法[J]. 应用声学, 2020, 39(2): 246-252.
YE J, XU F, YANG J, et al. An improved AIMM tracking algorithm based on adaptive transition probability[J]. Journal of Applied Acoustics, 2020, 39(2): 246-252(in Chinese).

[19] 游航航, 韩其松, 余敏建, 等. 基于AIGWO-IMMUKF的目标跟踪算法[J]. 北京航空航天大学学报, 2020, 46(10): 1826-1833.
YOU H H, HAN Q S, YU M J, et al. Target tracking algorithm based on AIGWO-IMMUKF[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(10): 1826-1833(in Chinese).

[20] YANG Y B, LIU Z G, QIN Y M, et al. Iteratively distributed instrumental variable-based pseudo-linear information filter for angle-only tracking[J]. ISA Transactions, 2023, 138: 359-372.

Multiple model minimum upper bound filter under generalized unknown disturbances

QIN Yuemei[1, *], ZHANG Ronghua[1], YANG Yanbo[2], PAN Quan[2]

(1. College of Automation, Xi'an University of Posts & Telecommunications, Xi'an 710121, China;
2. College of Automation, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract: This paper presents the multiple model upper bound filter (MMUBF) for maneuvering target tracking, since the tracking error is too big in existing algorithms when encountering with generalized unknown disturbances. In the multi-model framework, the minimum upper bound filter is implemented as the corresponding sub-filter in every mode to realize state update recursively. Then, the unknown disturbance is identified online according to the filtered result and the posterior mode probability, and the resultant estimate of disturbance is adopted to re-calculate the likelihood in each mode to eliminate the effect of the existence of unknown disturbance on the update of the posterior mode probability. Meanwhile, in order to further improve the model matching accuracy, the Markov transition probability matrix is adaptively adjusted using correction factors. In addition, the computational complexity of the algorithm is analyzed by calculating the number of floating-point operations at each step. In the meantime, the Markov transition probability matrix is adaptively modified using correction factors to further increase the model matching accuracy. Additionally, the number of floating-point operations at each step is calculated in order to examine the algorithm's computational complexity. Regarding various levels of measurement noises, process noises, adjustment coefficients, and probability correction threshold, the simulation results of maneuvering target tracking with time-varying unknown disturbances demonstrate that the suggested algorithm effectively suppresses the tracking error and has higher estimation accuracy than the existing interacting multiple model filter, adaptive interacting multiple model filter, improved adaptive interacting multiple model filter by gray wolf optimization algorithm, and single model-based minimum upper bound filter.

Keywords: minimum upper bound filter; generalized unknown disturbances; multiple model; correction factors; transfer probability matrix

Received: 2023-11-05; **Accepted**: 2024-01-15; **Published Online**: 2024-01-23 18:11

URL: link.cnki.net/urlid/11.2625.V.20240119.1110.001

Foundation items: National Natural Science Foundation of China (61903299); Natural Science Foundation of Shaanxi Province, China (2020JQ-842)

* **Corresponding author**. E-mail: qinyuemei@xupt.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0742

地基 cGNSS-R 双天线干涉测高试验研究

洪学宝[1]，李杰[2]，邢进[2]，杨鹏瑜[2]，杨东凯[2,*]

(1. 北京航空航天大学 前沿科学技术创新研究院，北京 100191；2. 北京航空航天大学 电子信息工程学院，北京 100191)

摘 要：地基场景下，传统全球导航卫星系统反射计（cGNSS-R）的直射、反射天线会接收到持续的反射、直射多径，对载波相位测高应用极为不利，但也为双天线干涉测高提供了可能。针对地基 cGNSS-R 双天线干涉测高的可行性问题，开展了水面测高试验研究。采集上视右旋圆极化（RHCP）天线和下视左旋圆极化（LHCP）天线接收的全球定位系统（GPS）L1 C/A 信号数据，处理得到两天线实际接收信号的相关功率序列，利用干涉测高方法处理该序列实现了天线高度反演。结果显示：RHCP 天线和 LHCP 天线干涉测高结果与实测天线高度的相关系数分别为 0.77 和 0.95，测高结果平均绝对误差分别为 0.15 m 和 0.04 m，均方根误差分别为 0.17 m 和 0.04 m。试验证明了传统全球卫星导航系统反射计的 RHCP 天线和 LHCP 天线均具备干涉测高能力，且 LHCP 天线干涉测高在有限开阔场景下的性能更优。

关 键 词：全球定位系统；反射计；干涉测量；水面高度；测高

中图分类号：V19; TP79

文献标志码：A **文章编号**：1001-5965(2026)01-0140-07

全球导航卫星系统反射计(global navigation satellite system reflectometry, GNSS-R)是利用地球表面反射的导航卫星信号进行地表环境参数反演的技术手段，具有优良的对地时-空采样能力，是当前机会源遥感领域的研究热点之一。

水面测高是 GNSS-R 技术的重要应用方向，其基本原理是通过获取反射信号相对于直射信号的延迟信息来反演水面高度[1]，应用场景包括近海岸潮汐监测、河流湖泊水位监测等。根据获取延迟信息所用信号特征的区别，GNSS-R 测高可分为伪码相位测高和载波相位测高 2 种方式[2]，其中，载波相位测高因精度更好而受到广泛关注。

地基 GNSS-R 载波相位测高通常以传统 GNSS-R(conventional GNSS-R, cGNSS-R)方式实现。cGNSS-R 利用特制 GNSS 接收机分立接收、协同处理直反射信号，从反射信号的复相关值中提取反射信号相对于直射信号的载波相位差别信息，从而实现高度反演。文献 [3] 介绍了一套用于潮汐-海态监测的地基 cGNSS-R 系统，测试结果表明该系统可取得约 3 cm 测高精度。文献 [4] 提出一种具有恶劣海况适应性的连续载波相位跟踪算法，并给出了大气传播效应和观测仪器影响校正模型，在不同风速下的实际测高精度优于 5 cm。文献 [5] 利用北斗卫星导航系统的地球同步轨道卫星反射信号实现了海面高度测量，经过 30 min 平滑后测高结果方差优于 0.3 cm。文献 [6] 基于短期地基试验对北斗卫星导航系统 B1I 和 B3I 反射信号测高性能进行了初步评估，结果表明，B3I 反射信号载波相位测高精度略优于 B1I 反射信号载波相位测高。

此外，地基 GNSS-R 载波相位测高也可通过全球导航卫星系统干涉反射测量(GNSS interferometric reflectometry, GNSS-IR)方式实现。GNSS-IR 利用经典 GNSS 接收机接收处理直反射干涉信号，从干涉信号信噪比序列中获取两信号载波相位差别变

收稿日期：2023-11-12；**录用日期**：2024-04-26；**网络出版时间**：2024-05-11 09:19
网络出版地址：link.cnki.net/urlid/11.2625.V.20240510.1703.002
基金项目：国家重点研发计划 (2022YFB3904402)；上海航天科技创新基金 (SAST2020-075)
* **通信作者**. E-mail：edkyang@buaa.edu.cn

引用格式：洪学宝，李杰，邢进，等. 地基 cGNSS-R 双天线干涉测高试验研究 [J]. 北京航空航天大学学报，2026，52（1）：140-146.
HONG X B，LI J，XING J，et al. Experimental study on interferometric altimetry using a ground-based dual-antenna cGNSS-R [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：140-146（in Chinese）.

化率信息, 进而实现高度反演。文献 [7] 提出利用 GPS 观测网络中近海岸跟踪站的测绘型 GNSS 接收机进行潮汐观测, 日均海面高度测量精度优于 3 cm。文献 [8] 提出利用线极化天线接收干涉信号, 实际观测结果表明, 该方式可在 30 min 时间分辨率下取得 4 cm 测高精度。文献 [9] 提出合成垂直阵列概念, 即通过周期性地改变天线高度使干涉信号信噪比快速波动, 初步试验结果表明, 该方式可在 5 min 时间分辨率下取得约 10 cm 的测高精度。文献 [10] 采用二阶动态潮位改正模型和稳健回归策略, 利用法国 BRST 岸基跟踪站 GNSS 观测数据研究了海面高度的季节性变化及年际特征过程。

在上述研究中, cGNSS-R 和 GNSS-IR 测高方法的处理过程互不相通, 这是由二者的工作机制决定的。然而, 实际应用中 cGNSS-R 直射天线和反射天线均可接收到直射信号和反射信号, 地基场景下两天线接收信号功率可能同时具有周期振荡特性[11-12], 这为地基 cGNSS-R 双天线干涉测高提供了可能, 即可以利用 GNSS-IR 高度反演算法分别处理 cGNSS-R 接收机输出的直射天线和反射天线接收信号功率实现对反射面高度的测量。与 cGNSS-R 相比, 这种方式充分利用了直射天线获取的反射信号数据, 也避免了反射天线接收到直射信号对测高结果造成影响。与 GNSS-IR 相比, 这种方式基于 cGNSS-R 体制获取了更多独立的反射信号数据, 对于提高测高性能具有重要意义。为验证这种测高方式的可行性, 本文设计并开展了初步的地基 cGNSS-R 双天线水面干涉测高试验。首先, 本文将对试验情况进行介绍; 然后, 给出数据处理方法; 最后, 给出水面高度反演结果并进行分析和讨论。

1　双天线干涉测高试验

本次试验开展于 2023 年 7 月 5 日和 6 日, 地点在山东省东营市河口区(118°31′33″E, 37°48′52″N)。试验场地为一处水塘, 近似为矩形区域, 东西方向宽度约为 40 m, 南北方向宽度约为 70 m。

试验设备为一套工作在 GPS L1 C/A 频点的 GNSS-R 信号采集系统。该系统主要由一副右旋圆极化(right hand circularly polarized, RHCP)天线、一副左旋圆极化(left hand circularly polarized, LHCP)天线、一台双通道 GNSS 信号采集器和一台笔记本电脑组成。其中: RHCP 天线用于接收直射信号; LHCP 天线用于接收反射信号; 双通道 GNSS 信号采集器用于对两接收信号进行调整、采样和量化; 笔记本电脑用于配置采集器、控制采集过程及存储采集数据。表 1 为所用信号采集系统的部分技术参数。

表 1　信号采集系统技术参数

Table 1　Specifications of the signal acquisition system

增益/dBi		半功率波束宽度/(°)		GNSS信号采集器中频频率/MHz	GNSS信号采集器采样频率/MHz	GNSS信号采集器量化位数/bit
RHCP天线	LHCP天线	RHCP天线	LHCP天线			
5	13	75	43	3.996	16.369	2

试验场地及设备安装情况如图 1 所示。两天线固定于高度可调的三脚架上, 安放在待观测水塘的北侧岸边, 过程中利用水平仪确保三脚架竖直。天线具体架设方式为: RHCP 天线在上, 最大增益方向指向天顶; LHCP 天线在下, 最大增益方向斜指水面, 指向朝南, 天线平面与水平面夹角约为 65°。两天线几何中心高度差约为 0.10 m。

图 1　试验场地及设备安装情况

Fig. 1　Experimental site and instrument installation

试验内容包括信号数据采集和天线高度测量 2 个部分。信号数据采集过程中, 考虑到水塘水位短时间内无显著变化, 本文通过调整三脚架高度来改变天线与水面之间的距离, 共采集了 3 个不同高度下的直射信号和反射信号数据, 每个高度下的数据采集时长超过 2 h。天线高度测量包括 LHCP 天线-地面高度测量和地面-水面高度测量 2 个环节。其中, LHCP 天线-地面高度利用米尺直接测量, 地面-水面高度利用米尺和水平仪通过斜距法进行估计。LHCP 天线-地面高度与地面-水面高度之和为 LHCP 天线实际高度。表 2 为试验采集的信号数据及对应的天线高度情况。

表 2　试验采集的数据及对应天线高度

Table 2　Raw data and antenna heights collected during the experiment

数据块编号	采集时间	LHCP天线高度/m
1	2023-07-05 15:24—17:42	≈3.18
2	2023-07-06 12:13—14:58	≈3.03
3	2023-07-06 15:31—18:16	≈2.83

2 试验数据处理

本节介绍试验数据处理过程，包括原始数据处理和天线高度反演2个环节。

2.1 原始数据处理

利用MATLAB GPS L1 C/A反射信号处理软件对采集到的原始数据进行后处理，流程如图2所示。首先，直射通道对直射信号进行捕获和跟踪，得到直射信号1 ms相关值。在直射通道跟踪直射信号的同时，反射通道直接利用直射信号跟踪环路产生的本地载波和伪码对反射信号进行解调和解扩，得到反射信号1 ms相关值。然后，分别对直射信号和反射信号相关值进行1 000次非相干累加，得到直射信号和反射信号相关功率。最后，利用直射信号相关值获取基本导航信息，计算可见星高度角、方位角及RHCP天线位置坐标。软件输出的数据包括RHCP天线位置坐标及各可见星的方位角、高度角、直射信号相关功率、反射信号相关功率，数据输出频率为1 Hz。

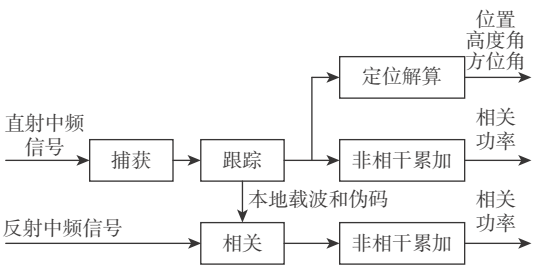


图2 原始数据处理框图

Fig. 2 Block diagram of raw data processing

2.2 天线高度反演

利用GNSS-IR高度反演算法分别处理两天线实际接收信号相关功率序列对LHCP天线高度进行估计。具体步骤如下：

1）观测序列准备。考虑后续频谱分析算法要求时间变量单调递增，在数据处理前需要对观测序列进行准备。首先，将具有非单调卫星高度角的相关功率序列从高度角的极值处进行拆分，保证每个相关功率序列均对应单调的卫星高度角。其次，将具有单调递减高度角的相关功率序列翻转，保证所有相关功率序列对应于单调递增的高度角。

2）数据质量控制。由于天线架设在目标区域北侧，即天线有效观测范围在南侧，因此，对任一相关功率序列，应筛除其中方位角不在90°～270°范围内的相关功率样点；受实际增益分布影响，天线在某些角度下可能无法接收到足够强度的直射或反射信号，使得实际信号的相关功率不具备振荡特性，因此，应筛除无振荡的相关功率序列；考虑样点数量会影响频率估计准确性，进一步筛除了振荡周期数量小于3的序列。

3）去趋势项处理。两信号相关功率是具有缓变趋势项的振荡序列。考虑趋势项的存在会对振荡频率提取造成不利影响，需要在谱分析前予以剔除。时间序列去趋势项方法包括最小二乘曲线拟合、经验模态分解、变分模态分解、小波滤波等。本文采用的是变分模态分解方法[13]。处理过程中，模态数取6，分解得到的本征模态函数分量6与残差之和作为趋势项予以去除。

4）振荡频率估计。以卫星高度角正弦值为时间变量，两信号相关功率可认为是具有幅度和相位调制的余弦序列。考虑卫星高度角正弦值是非等间隔采样的，本文采用Lomb-Scargle周期图法[14-15]对相关功率序列进行频谱分析。获得相关功率序列的功率谱后，提取峰值频率作为相关功率序列的振荡频率。

5）天线高度计算。对于RHCP天线接收信号的相关功率序列，采用下述公式计算LHCP天线高度：

$$H = \lambda f_{\mathrm{r}}/2 - d \tag{1}$$

式中：H为LHCP天线相位中心距离水面的高度；λ为信号波长；f_{r}为RHCP天线接收信号相关功率序列的振荡频率；d为RHCP与LHCP相位中心距离，本次试验中取值为0.10 m。

对于LHCP天线接收信号的相关功率序列，采用下式计算LHCP天线高度：

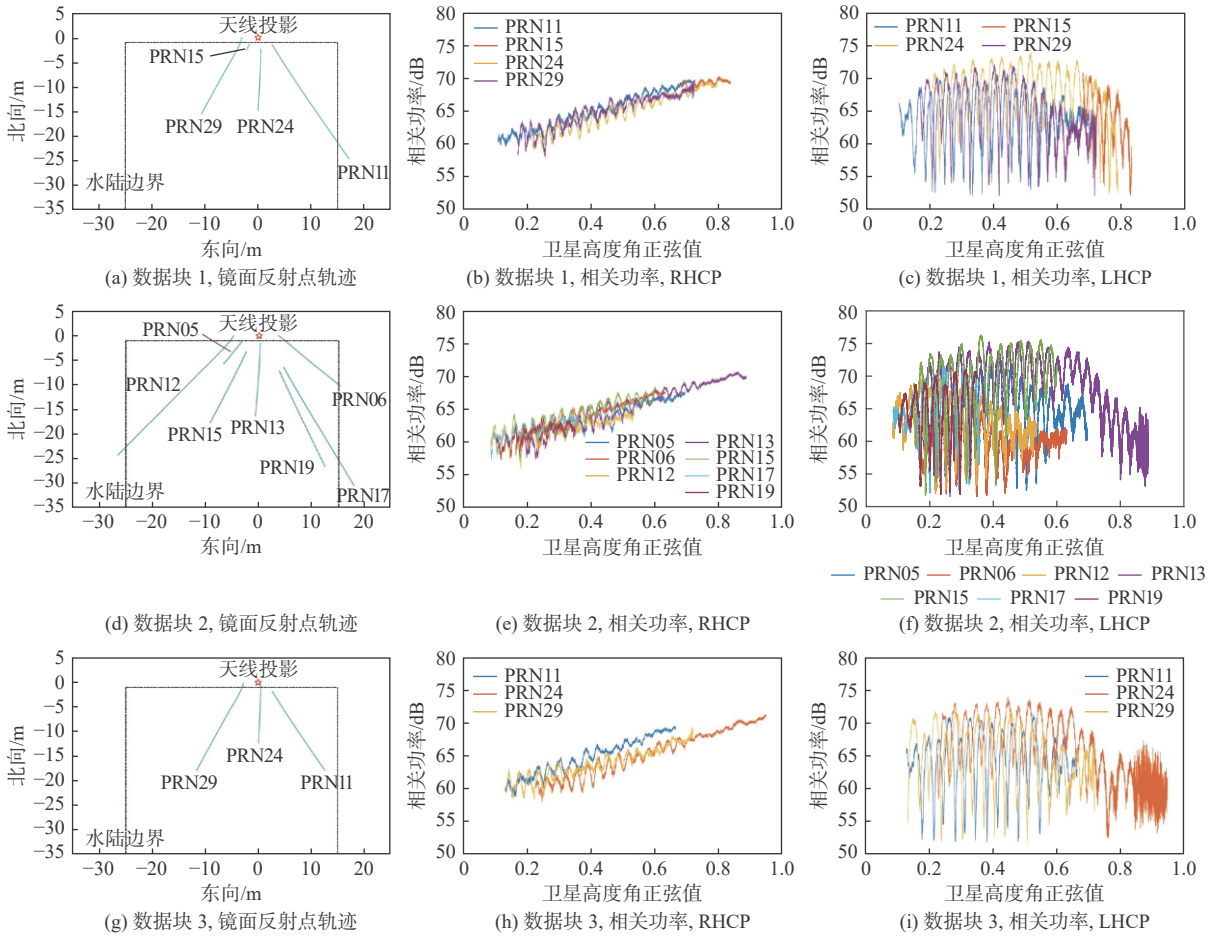
$$H = \lambda f_{\mathrm{l}}/2 \tag{2}$$

式中：f_{l}为LHCP天线接收信号相关功率序列的振荡频率。

3 试验结果分析

图3为经过数据质量控制后各数据块中反射信号镜面反射点轨迹及两天线实际接收信号相关功率序列。其中：图3(a)、图3(d)和图3(g)分别为数据块1、数据块2和数据块3中各可见星反射信号镜面反射点轨迹；图3(b)、图3(e)和图3(h)分别为从数据块1、数据块2和数据块3中获得的RHCP天线接收的各可见星信号相关功率序列；图3(c)、图3(f)和图3(i)分别为从数据块1、数据块2和数据块3中获得的LHCP天线接收的各可见星信号相关功率序列。

本次试验中，两天线接收信号的相关功率序列均具有周期振荡特性，但二者振荡幅度变化情况有显著差别。RHCP天线接收信号相关功率的振荡幅度随着卫星高度角的增大趋于减小，原因在于：



(a) 数据块 1, 镜面反射点轨迹　(b) 数据块 1, 相关功率, RHCP　(c) 数据块 1, 相关功率, LHCP

(d) 数据块 2, 镜面反射点轨迹　(e) 数据块 2, 相关功率, RHCP　(f) 数据块 2, 相关功率, LHCP

(g) 数据块 3, 镜面反射点轨迹　(h) 数据块 3, 相关功率, RHCP　(i) 数据块 3, 相关功率, LHCP

图 3　镜面反射点轨迹与相关功率序列

Fig. 3　Specular points and correlation power waveforms

①随着卫星高度角的增大, RHCP 反射信号对应的表面反射率逐渐减小; ②随着卫星高度角的增大, RHCP 天线对于 RHCP 反射信号的增益趋于减小。LHCP 天线接收信号相关功率的振荡幅度随着卫星高度角的增大呈现先增大后减小的趋势, 原因在于: ①随着卫星高度角的增大, LHCP 反射信号对应的水面反射率逐渐增大并趋于平缓; ②随着反射信号镜面反射点进入和移出天线有效覆盖区域, LHCP 天线对 LHCP 反射信号的增益先增大再减小。对比来看, LHCP 天线接收信号相关功率较 RHCP 天线接收信号相关功率振荡更剧烈, 前者的最大振幅接近 10 dB, 后者的最大振幅接近 3 dB。

对两天线实际接收信号相关功率序列进行谱分析, 提取峰值频率作为相关功率序列的振荡频率, 进而根据式 (1) 和式 (2) 计算 LHCP 天线高度, 结果如图 4 所示, PRN 为伪随机噪声。总体来说, 两天线测高结果与实测结果的基本趋势保持一致, RHCP 天线测高结果较 LHCP 天线测高结果的离散程度更高。

为定量评价两天线的干涉测高性能, 本文计算了反演结果与实测结果的相关系数、反演结果的平均绝对误差(mean absolute error, MAE)和均方根误差(root mean square error, RMSE), 如表 3 所示。可以看到, 本次试验中 RHCP 天线取得的测高结果性能稍差, 而 LHCP 天线取得的测高结果性能大幅优于 RHCP 天线。

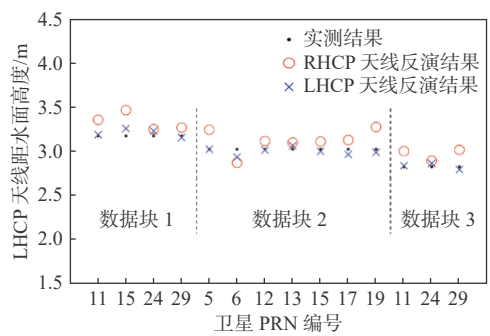


图 4　LHCP 天线高度反演

Fig. 4　Retrieved LHCP antenna heights

表 3　干涉测高性能

Table 3　Performance of interferometric altimetry

天线	相关系数	MAE/m	RMSE/m
RHCP	0.77	0.15	0.17
LHCP	0.95	0.04	0.04

造成二者测高性能较大差别的主要原因在于不同方位角和高度角下两天线接收的反射信号受到场地的影响不同。反射信号的主要能量来自第一菲涅尔反射区，其沿着卫星方位角方向的覆盖范围随着卫星高度角的减小而增大。试验水塘东西窄、南北宽，天线架设在水塘北岸中部，随着卫星方位角由南向东、西偏离，反射信号在卫星高度角较低时更易受到堤岸、植被遮挡和陆地反射的影响。根据本节对相关功率序列振幅变化情况的分析可以推断，RHCP天线测高较LHCP天线测高更依赖于低卫星高度角下的相关功率序列。因此，在卫星方位角偏东或偏西时，RHCP天线测高较LHCP天线测高受到场地影响的可能性更高。为说明这一问题，本文以数据块3为例对去除趋势项的两天线接收信号相关功率序列进行了分段谱分析，具体过程如下：①为避免高度角差异对分析造成影响，截取各可见星高度角交集（约13°～41°）所对应的相关功率序列；②将截取的序列按照高度角正弦值等分为两段，分别定义为Ⅰ段（高度角13°～26°）和Ⅱ段（高度角26°～41°）；③利用Lomb-Scargle方法分别计算各段序列的功率谱。图5为两天线接收的各可见星信号相关功率的分段谱分析结果，其中，参考频率是利用实测天线高度推算的相关功率振荡频率参考值。由图5(a)～图5(c)可见，对于RHCP天线，各可见星Ⅰ段序列振荡频率差别较大，Ⅱ段序列振荡频率差别较小。结合图3(g)可以发现，RHCP天线Ⅰ段序列频率误差最小的PRN24位于南向，误差居中的PRN29位于西南，误差最大的PRN11位于东南，频率误差大小与镜面反射点到水陆边界距离大小基本吻合。而由图5(d)～图5(f)可见，对于LHCP天线，各可见星Ⅰ段序列和Ⅱ段序列振荡频率差别均很小。以上分析表明，低高度角下RHCP天线测高较LHCP天线测高受到场地影响更为严重。

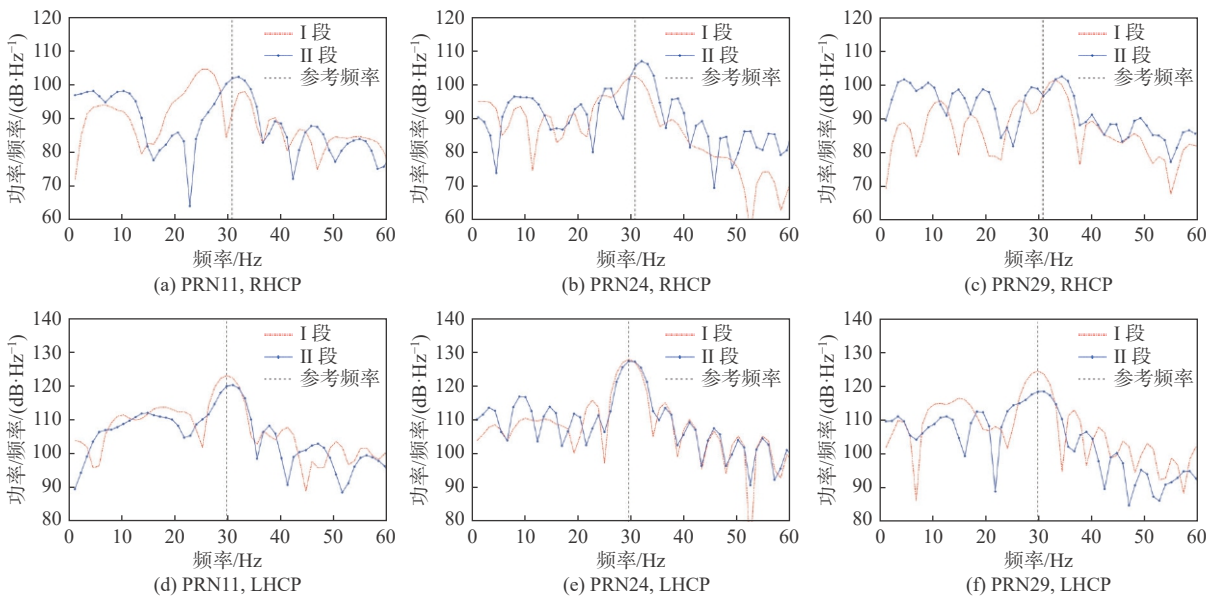


图5 相关功率分段谱分析

Fig. 5 Power spectrum of segmented correlation power

除场地因素外，天线增益大小也是影响测高性能的一个重要原因。本次试验分别采用低增益RHCP天线和高增益LHCP天线接收干涉信号，前者接收的反射信号强度远低于后者，使得前者获得的干涉波形质量明显低于后者，进而导致前者干涉波形振荡频率估值的准确度低于后者，最终造成RHCP天线测高精度较LHCP天线差，如图5(b)和图5(e)所示。

最后需要说明的是，本次试验结果是在未进行二次数据质量控制或数据平滑的情况下取得的，也未对两天线观测结果进行融合，在实际观测中可通过二次数据质量控制、数据平滑和结果融合进一步提升双天线干涉测高精度。

4 结 论

1）本次试验在cGNSS-R系统RHCP天线和LHCP天线实际接收信号相关功率序列中均观测到了周期振荡现象。其中，RHCP天线接收信号相关功率振荡幅度最大值约3 dB，LHCP天线接收信号相关功率振荡幅度最大值接近10 dB。

2）本次试验将GNSS-IR测高方法应用于cGNSS-R两天线实际接收信号相关功率序列，成功实现了双天线干涉测高。其中，RHCP天线干涉测高结果的均方根误差为0.17 m，LHCP天线干涉测

高结果的均方根误差为 0.04 m。

本文论证的 cGNSS-R 双天线干涉测高方法仅是 cGNSS-R 信号接收处理方式与 GNSS-IR 地物参数反演算法结合的一个实例, 该思路对于积雪深度测量、土地沉降监测、土壤湿度反演等应用同样具有参考价值。在具体 cGNSS-R 观测中, 可以通过观察或检测直射、反射天线实际接收信号功率序列的振荡情况来判断直射、反射天线中是否接收到了持续的反射、直射多径, 再视情况选用 cGNSS-R 或 GNSS-IR 方式反演地物参数。

参考文献（References）

[1] MARTIN-NEIRA M. A passive reflectometry and interferometry system (PARIS): application to ocean altimetry[J]. ESA Journal, 1993, 17(4): 331-355.

[2] 杨东凯, 张其善. GNSS 反射信号处理基础与实践[M]. 北京: 电子工业出版社, 2012: 135-142.
YANG D K, ZHANG Q S. GNSS reflected signal processing: fundamentals and applications[M]. Beijing: Publishing House of Electronics Industry, 2012: 135-142(in Chinese).

[3] DUNNE S, SOULAT F, CAPARRINI M, et al. Oceanpal/sup/spl reg// a GPS-reflection coastal instrument to monitor tide and sea-state[C]//Proceedings of the Europe Oceans. Piscataway: IEEE Press, 2005: 1351-1356.

[4] LIU W, BECKHEINRICH J, SEMMLING M, et al. Coastal sea-level measurements based on GNSS-R phase altimetry: a case study at the onsala space observatory, Sweden[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(10): 5625-5636.

[5] WU J M, CHEN Y L, GAO F, et al. Sea surface height estimation by ground-based BDS GEO satellite reflectometry[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 5550-5559.

[6] 张云, 赵乐久, 孟婉婷, 等. 北斗卫星反射信号岸基海面高度反演精度的评估[J]. 北京航空航天大学学报, 2023, 49(5): 999-1008.
ZHANG Y, ZHAO L J, MENG W T, et al. Evaluation of accuracy of shore-based sea surface height inversion based on Beidou satellite reflected signals[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(5): 999-1008(in Chinese).

[7] LARSON K M, RAY R D, NIEVINSKI F G, et al. The accidental tide gauge: a GPS reflection case study from kachemak bay, Alaska[J]. IEEE Geoscience and Remote Sensing Letters, 2013, 10(5): 1200-1204.

[8] ALONSO-ARROYO A, CAMPS A, PARK H, et al. Retrieval of significant wave height and mean sea surface level using the GNSS-R interference pattern technique: results from a three-month field campaign[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(6): 3198-3209.

[9] YAMAWAKI M K, GEREMIA-NIEVINSKI F, GALERA MONICO J F. High-rate altimetry in SNR-based GNSS-R: proof-of-concept of a synthetic vertical array[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 8009005.

[10] 郭斐, 李佰瀚, 张治宇, 等. 利用 GNSS 反射信号监测海面高度变化: 基于法国 BRST 站 2019—2021 年数据[J]. 地球科学与环境学报, 2023, 45(3): 548-558.
GUO F, LI B H, ZHANG Z Y, et al. Change of sea surface height monitored by GNSS reflected signals: based on data from BRST station in France from 2019 to 2021[J]. Journal of Earth Sciences and Environment, 2023, 45(3): 548-558(in Chinese).

[11] 洪学宝, 张波, 杨东凯, 等. 地基 GNSS-R 功率测量应用中的天线方向性影响分析[J]. 北京理工大学学报, 2021, 41(6): 658-664.
HONG X B, ZHANG B, YANG D K, et al. Antenna directivity effect analysis for ground-based GNSS-R power measurement applications[J]. Transactions of Beijing Institute of Technology, 2021, 41(6): 658-664(in Chinese).

[12] 洪学宝, 张波, 阮宏梁, 等. 基于相关功率修正的地基 GNSS-R 土壤湿度反演[J]. 北京航空航天大学学报, 2021, 47(8): 1558-1564.
HONG X B, ZHANG B, RUAN H L, et al. Ground-based GNSS-R soil moisture retrieval based on correlation power correction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(8): 1558-1564(in Chinese).

[13] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition[J]. IEEE Transactions on Signal Processing, 2014, 62(3): 531-544.

[14] LOMB N R. Least-squares frequency analysis of unequally spaced data[J]. Astrophysics and Space Science, 1976, 39(2): 447-462.

[15] SCARGLE J D. Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data[J]. The Astrophysical Journal, 1982, 263: 835-853.

Experimental study on interferometric altimetry using a ground-based dual-antenna cGNSS-R

HONG Xuebao[1]，LI Jie[2]，XING Jin[2]，YANG Pengyu[2]，YANG Dongkai[2,*]

(1. Research Institute for Frontier Science，Beihang University，Beijing 100191，China；
2. School of Electronic and Information Engineering，Beihang University，Beijing 100191，China)

Abstract：The direct and reflected signals can be concurrently received by both antennas of a conventional global navigation satellite system reflectometer (cGNSS-R) in ground-based settings. It is unfavorable for cGNSS-R carrier phase altimetry, but it also provides cGNSS-R the possibility of retrieving reflector height in the way of interferometric reflectometry. An initial ground-based water surface experiment was carried out to investigate the feasibility of cGNSS-R interferometric altimetry. GPS L1 C/A signals were received by an up-looking right hand circularly polarized (RHCP) antenna and a down-looking left hand circularly polarized (LHCP) antenna and then collected by a dual-channel sampler. The raw signals were postprocessed to generate their correlation power sequences. These sequences were then subjected to the GNSS interferometric reflectometry altimetry technique in order to estimate the LHCP antenna heights. Experimental results show that the correlation coefficients between the retrieved values and the true values are 0.77 and 0.95 for the RHCP antenna and the LHCP antenna, respectively. The mean absolute errors of the results are 0.15 m and 0.04 m, and the root mean square errors are 0.17 m and 0.04 m. It indicates that both the RHCP antenna and the LHCP antenna of a conventional GNSS reflectometer can be used for ground-based interferometric altimetry, and the LHCP antenna performs better in this openness-limited scenario.

Keywords：global positioning system；reflectometer；interferometry；water level；altimetry

Received：2023-11-12；**Accepted**：2024-04-26；**Published Online**：2024-05-11 09：19

URL：link.cnki.net/urlid/11.2625.V.20240510.1703.002

Foundation items：National Key Research and Development Program of China(2022YFB3904402); Shanghai Aerospace Science and Technology Innovation Fund(SAST2020-075)

* **Corresponding author**. E-mail：edkyang@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0722

空地网联集群协同模式识别方法

曲桂娴[1, 2, 3]，周建山[4, *]，司杨[5]，刘晓静[1, 2, 3]，袁奇雨[1, 2, 3]，马清琳[2, 3, 6]

(1. 北京航空航天大学 航空发动机研究院，北京 102206；2. 北京航空航天大学 未来航空发动机协同设计中心，北京 102206；3. 天目山实验室，杭州 310023；4. 北京航空航天大学 交通科学与工程学院，北京 100191；5. 北京物资学院 物流学院，北京 101149；6. 北京航空航天大学 能源与动力工程学院，北京 100191)

摘 要：空地网联集群在智慧城市、智慧农林、智慧交通等国民经济生产领域具有巨大的应用潜力，同时在战场态势感知、空地协同打击等军事领域展现出极大的应用价值。面向空地网联集群准确感知与识别复杂环境目标的需求，建立基于模式分类概率的全局似然函数最小化模型，提出空地网联集群的分布式学习与自适应信息融合算法，该算法包括基于梯度下降的信息扩散和基于自适应加权的信息融合2个主要步骤，形成了空地协同的模式识别方法。此外，推导出了空地网联集群协同模式识别方法的平均误差递归方程，理论证明了所提算法的误差收敛性。通过建立空地网联集群网络信息交互拓扑模型，利用雷达实测数据集进行仿真测试。仿真结果表明：集群分布式融合算法对信息估计的平均均方偏差和系统误差可有效逼近理论最优水平。当节点数由10上升至40时，集群分布式融合算法的平均均方偏差由−48.70 dB下降至−53.96 dB，系统误差由−27.42 dB下降至−30.22 dB，接近于误差的理论值。对比实验表明：所提算法较传统方法具有良好的精度，可有力支撑空地网联集群对复杂环境目标的感知与识别。

关 键 词：空地网联集群；协同模式识别；信息交互拓扑模型；分布式融合算法；雷达实测数据

中图分类号：V279

文献标志码：A **文章编号**：1001-5965(2026)01-0147-10

空地网联集群是以空基移动平台(如临近空间有人/无人驾驶航空器、平流层气球/飞艇、低空有人/无人飞机等)、地面载运工具(如无人战车)和基础设施为载体，实现空地一体化信息实时感知、组网传输和计算处理的网络系统[1]，可为海量出行用户提供更加智能化、网联化、泛在化交通信息服务，在应急救援、空间探测等领域发挥重大作用[2-5]，已成为国家安全和大国地位的重要战略保障及大国博弈的战略制高点。2021年2月，党中央、国务院印发《国家综合立体交通网规划纲要》，明确提出“加强智能化载运工具和关键专用装置研发，推进智能网联车辆、智能化通用航空器应用”的建设任务[6]。目前，随着5G大规模商用和蜂窝车联网(Cellular V2X)技术的成熟，空地网联集群已逐渐成为B5G/6G发展的重要愿景[7-8]。由于基于单一传感节点的模式识别方法及传感增强技术的精度不高，因此，具备良好的复杂环境目标感知与识别能力将是空地网联集群成功实施的重要保障。例如，在包含无人机和地面装甲集群的智能集群系统进行协同打击等军事对抗应用场景中，需要无人机和无人车辆对同一目标进行敌我识别。因此，亟需开展空地网联集群协同模式识别方法研究，从而有效

收稿日期：2023-11-04；**录用日期**：2024-03-04；**网络出版时间**：2024-03-18 16:08
网络出版地址：link.cnki.net/urlid/11.2625.V.20240317.1852.001
基金项目：国家自然科学基金(52302510,52202391)；航空发动机及燃气轮机基础科学中心项目(P2023-B-I-005-001)；中央高校基本科研费专项资金资助(YWF-23-Q-1066)；天目山实验室交叉创新研究团队项目(TP-2024-A-001,TK-2024-D-001)
* **通信作者**. E-mail：jianshanzhou@foxmail.com

引用格式：曲桂娴，周建山，司杨，等. 空地网联集群协同模式识别方法[J]. 北京航空航天大学学报，2026，52(1)：147-156.
QU G X，ZHOU J S，SI Y，et al. Cooperative pattern recognition method for air-ground networked swarms[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：147-156 (in Chinese).

提升空地网联集群任务决策与协同控制的精度。

实现空地网联集群的精准目标识别通常面临以下挑战：由于集群中单一节点的传感器性能有限，其感知范围和感知精度会受到外部环境（如采样噪声、电磁噪声、异常气象等）的严重干扰，导致基于单一传感节点的模式识别方法及传感增强技术在实际场景应用受限。此外，由于多个传感节点对同一目标的测量信息中含有噪声，导致整个集群的测量信息具有强随机性和时变性，对基于多节点传感信息融合的模式识别方法带来巨大挑战。当前研究者们围绕这一方向开展了相关研究工作，如Cai等[9]提出一种基于几何关系的三角形表示的认知共享方法，可使无人机和地面机器人在非结构化环境中在不共享外观信息的情况下识别出相同物体。Li等[10]利用图像校正技术对无人机航拍图像进行处理，实现了空地协同系统中的自动地面地图生成。

此外，国内外学者在空地网联集群协同搜索、路径规划、系统架构等方面也开展了相关研究。Bhamidipati等[11]通过分析无人机群网络上GPS接收机记录的前端信号功率的变化，提出同时定位多个干扰机和接收机的算法。Wan等[12]提出一种基于稀疏表示的目标跟踪算法，用以定位无人机视频中的兴趣点，将基于马尔可夫随机场的具有连续遮挡约束的二元支持向量引入到稀疏表示模型中。为挖掘频谱复用潜力并增强协同感知无人机网络的协同感知能力，Chen等[13]提出联合感知通信协同传感无人机网络，采用上限平均协同感知面积作为协同感知度性能的度量。Ebrahimi等[14]采用基于投影的压缩数据采集方法，对密集无线传感器网络中的优化节点聚类、转发树构建、每个簇的簇头选择及用于高能效收集数据的无人机（unmanned aerial vehicle, UAV）轨迹规划的联合问题进行了数学建模和分析。Zheng等[15]研究了单目摄像头对无人机空对空视觉检测问题，提出空对空对无人机检测的数据集，利用该数据集评估了典型深度学习算法的总体性能，分析了环境、目标尺度、视角等因素对目标检测性能的影响。Albanese等[16]研究了基于无人机的自动一体移动网络解决方案，结合机器学习技术有效定位区域中目标位置。Pan等[17]提出飞机和无人机大规模空域感知协同策略，设计了基于数据驱动的空域感知预测方法，确定了无人机群的调度路径。Koutras等[18]提出半分布式算法用以提升对无人机群整体态势感知能力。Liu等[19]基于分布式多智能体深度强化学习方法，提出一种用于节能和分布式车辆导航的分布式控制框架，可有效提升能源利用效率及数据收集率。Jouhari等[20]提出一种针对无人机的深度神经网络分发方法，从而实现设备的数据分类，考虑无人机的移动性模型和资源约束，实现数据收集和决策之间延迟的最小化。

综上所述，当前研究所采用的理论模型和技术架构缺乏智能集群协同感知学习能力，系统无法高效地适应动态的任务和时变的环境条件。鲜有研究在空地网联集群系统中引入自组织自适应的协同学习机制。因此，本文针对现有研究不足，基于分布式理论研究自适应融合算法，从而实现自主智能集群应用的空地协同网络自主模式识别。本文具体研究工作包括3个方面：

1）针对多个空地节点构成的空地网联集群系统，建立基于Logistic函数的目标识别或模式分类的概率表征模型，进而建立相应的似然函数，将空地网联集群对目标模式识别的问题转化为Logistic函数结构参数的最优估计问题，提出空地集群网络的全局似然函数最小化模型；

2）在空地网联集群自组织网络通信基础上，针对全局似然函数优化模型，利用节点实时测量信息和邻域节点共享的估计信息提出空地网联集群分布式学习算法，其包括节点估计信息扩散与融合2大步骤，同时设计了信息融合权重的自适应更新算法，实现空地集群分布式自适应学习和信息融合；

3）利用雷达实测数据开展仿真实验和对比分析，结果表明本文算法比传统的单智能体算法、多智能体分布式算法取得更低的均方偏差及系统误差，本文算法能以更高精度逼近全局理论误差下界，验证了该算法的有效性和性能优势。

1 物理模型

1.1 系统建模与问题描述

本文考虑M个空地节点构成的空地网联集群系统，如图1所示。其中，空中节点和地面节点可分别代表无人机和无人车等具备协同组网、智能环境感知能力的移动集群。假设每个节点v接收传感器数据样本为$\{\gamma_v(t), \boldsymbol{j}_{v,t}\}$，其中，$\gamma_v(t)=\pm 1$，用以指定传感器数据样本$\boldsymbol{j}_{v,t}$的所属类别。例如，在二分类问题中，识别第1类结果可标记为$\gamma_v(t)=1$，识别第2类结果标记为$\gamma_v(t)=-1$。

在空地网联集群模式识别应用场景中，每个节点对于同一目标的识别或分类可以通过一个Logistic函数来实现。每个节点将特征值输入到一个Logistic函数中，Logistic函数将这些特征值映射到一个介

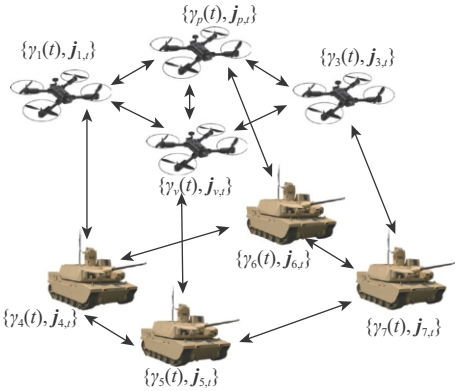


图 1　空地网联集群应用场景

Fig. 1　Application scenario of air-ground networked swarms

于 0 和 1 之间的数值, 这个数值可以被解释为该节点对目标识别分类结果属于某一类别的概率。具体地, 令列向量$\boldsymbol{u}$为 Logistic 函数的结构参数, 但是其对于每个节点是未知的; 用±1 表示目标的类别, 如当目标的测量信号为 1 时, 表示该目标被识别为友军, 当识别信号为–1时, 目标被识别为敌军。对于节点v而言, 用$\boldsymbol{j}_v$表示其测量信号, γ_v表示其识别目标的结果, 则可用 Logistic 函数表示该节点v将目标识别为友军的概率如下:

$$P(\gamma_v = 1 \mid \boldsymbol{j}_v) = \frac{1}{1+\mathrm{e}^{-\boldsymbol{j}_v^{\mathrm{T}}\boldsymbol{u}}} \tag{1}$$

类似地, 节点v将目标识别为敌军的概率如下:

$$P(\gamma_v = -1 \mid \boldsymbol{j}_v) = \frac{1}{1+\mathrm{e}^{\boldsymbol{j}_v^{\mathrm{T}}\boldsymbol{u}}} \tag{2}$$

结合式(1)和式(2)可变为

$$P(\gamma_v; \boldsymbol{u}^{\mathrm{o}} \mid \boldsymbol{j}_v) = \frac{1}{1+\mathrm{e}^{-\gamma_v \boldsymbol{j}_v^{\mathrm{T}}\boldsymbol{u}}} \tag{3}$$

则 Logistic 函数的分类概率 l_{ogit} 可写为

$$l_{\mathrm{ogit}} = \ln \frac{P(\gamma_v = 1 \mid \boldsymbol{j}_v)}{1 - P(\gamma_v = 1 \mid \boldsymbol{j}_v)} = \boldsymbol{j}_v^{\mathrm{T}}\boldsymbol{u} \tag{4}$$

对于单个节点可以通过其对数似然函数来估计 Logistic 模型的全局参数 $\boldsymbol{u}$, 即 l_{ogit} 函数的截距, 该参数对事件发生概率的对数几率具有显著影响。在训练 Logistic 模型时, 可利用带有随机噪声的测量信息建立模式识别的似然函数$C_v(\boldsymbol{u})$, 可表示为

$$C_v(\boldsymbol{u}) \equiv C(\boldsymbol{u}) = \frac{\rho}{2}\|\boldsymbol{u}\|^2 + E\left(\ln\left(1+\mathrm{e}^{-\gamma_v(t)\boldsymbol{j}_{v,t}^{\mathrm{T}}\boldsymbol{u}}\right)\right) \tag{5}$$

式中: $\frac{\rho}{2}\|\boldsymbol{u}\|^2$表示正则化项; $E\left(\ln\left(1+\mathrm{e}^{-\gamma_v(t)\boldsymbol{j}_{v,t}^{\mathrm{T}}\boldsymbol{u}}\right)\right)$表示对数似然函数的期望。

对于任意传感节点 v 给定其测量信息构成的数据集$\{\gamma_v(t), \boldsymbol{j}_{v,t}\}$, 则根据式 (5) 可知, 与 $C_v(\boldsymbol{u})$ 对应的似然函数具体表示为

$$Q\left(\boldsymbol{u}; \gamma_v(t), \boldsymbol{j}_{v,t}\right) = \frac{\rho}{2}\|\boldsymbol{u}\|^2 + \ln\left(1+\mathrm{e}^{-\gamma_v(t)\boldsymbol{j}_{v,t}^{\mathrm{T}}\boldsymbol{u}}\right) \tag{6}$$

所有节点的全局代价 Logistic 函数可写为

$$C^{\mathrm{global}}(\boldsymbol{u}) = \sum_{v=1}^{M} C_v(\boldsymbol{u}) \tag{7}$$

式中: $C^{\mathrm{global}}(\boldsymbol{u})$为全局代价函数; $C_v(\boldsymbol{u}), v = 1,2,\cdots,M$为每个节点的代价函数, 并具有可微及严格凸函数的性质。因此, $C^{\mathrm{global}}(\boldsymbol{u})$是严格凸函数, 空地网联集群全局参数最优估计$\boldsymbol{u}^{\mathrm{o}}$可取唯一最小值。

因此, 空地网联集群实现目标识别需要获得全局 Logistic 函数(见式 (7))的结构参数$\boldsymbol{u}$。所以空地网联集群实现模式识别在于集群协同逼近$\boldsymbol{u}$, 而该问题可转化为网络集群分布式最小化似然函数式 (7) 的最优化问题, 亦即将全局模式识别问题构建为一个空地网联集群全局参数最优估计问题:

$$\boldsymbol{u}^{\mathrm{o}} \in \arg\min_{\boldsymbol{u}} \left\{C^{\mathrm{global}}(\boldsymbol{u})\right\} \tag{8}$$

1.2　集群分布式融合算法设计

为求解集中式优化问题, 设计分布式优化算法, 将全局优化目标求解问题转化为各节点的本地优化求解的总和:

$$C^{\mathrm{global}}(\boldsymbol{u}) = \sum_{p=1}^{M} C_p^{\mathrm{local}}(\boldsymbol{u}) \tag{9}$$

式中: $C_p(\boldsymbol{u})$为节点 p 的本地代价函数, p=1,2,⋯,M。

进一步, 将式(8)全局最优估计问题转化为可分布式求解的全局参数最优估计问题的结构, 即

$$C^{\mathrm{global}}(\boldsymbol{u}) = \sum_{p=1}^{M}\left(C_p(\boldsymbol{u}) + \sum_{v\neq p}^{M} d_{v,p} C_v(\boldsymbol{u})\right) \tag{10}$$

式中: $d_{v,p}$为由节点v分配给节点p的信息权重, 权重总和可写为$\sum_{v=1}^{M} d_{v,p} = 1, v = 1,2,\cdots,M$。令$N_p$表示节点$p$的邻域, 由节点$p$及与节点$p$共享信息的全部节点组成, 当$v \notin N_p$时, $d_{v,p} = 0$。

由于各个$C_v(\boldsymbol{u})$均为凸函数, 因此, $C_p^{\mathrm{global}}(\boldsymbol{u})$为凸函数。根据式 (10), 定义每个节点 p 的代价函数为

$$C_p^{\mathrm{local}}(\boldsymbol{u}) = C_p(\boldsymbol{u}) + \sum_{v\neq p}^{M} d_{v,p} C_v(\boldsymbol{u}) \tag{11}$$

由式 (11) 可知, $C_p^{\mathrm{local}}(\boldsymbol{u})$为凸函数。因此, 集中式优化$C_p^{\mathrm{global}}(\boldsymbol{u})$, 亦即求解式 (8), 可等价于每个节点 p 以分布式优化的方式求解如下问题:

$$\min_{\boldsymbol{u}} C_p^{\mathrm{local}}(\boldsymbol{u}) \qquad p = 1,2,\cdots,M \tag{12}$$

针对式 (10), 进一步地可以通过梯度下降原理设计优化算法。具体地, 给定任意节点$p = 1,2,\cdots,M$在$t-1$时刻对$\boldsymbol{u}^{\mathrm{o}}$的估计值$\boldsymbol{u}_{p,t-1}$, 由传统的梯度下降方法可得到新的估计值$\boldsymbol{u}_{p,t-1}$, 即

$$\boldsymbol{u}_{p,t}=\boldsymbol{u}_{p,t-1}-\lambda_p\nabla\boldsymbol{u}C_p^{\text{local}}\left(\boldsymbol{u}_{p,t-1}\right)=\boldsymbol{u}_{p,t-1}-\lambda_p\left(\nabla\boldsymbol{u}C_p\left(\boldsymbol{u}_{p,t-1}\right)+\sum_{v\in N_p}d_{v,p}\nabla\boldsymbol{u}C_v\left(\boldsymbol{u}_{p,t-1}\right)\right)\tag{13}$$

式中：$\nabla\boldsymbol{u}C_v(\boldsymbol{u})=\rho\boldsymbol{u}-\gamma_v(t)\boldsymbol{j}_{v,t}\left(\dfrac{1}{1+\mathrm{e}^{\gamma_v(t)\boldsymbol{j}_{v,t}\boldsymbol{u}}}\right)$。

在传统梯度下降迭代算法中，无法有效利用邻域节点的新的估计值提升本地节点模式识别精度，为此本文在式(13)的基础上，引入信息融合机制，提出新的分布式算法。引入$\boldsymbol{\psi}_{p,t}$作为节点p在t时刻由梯度下降更新得到的临时估计值，定义线性加权融合参数$l_{v,p}$，当$v\notin N_p\cup\{p\}$时，$l_{v,p}=0$；当$v\in N_p\cup\{p\}$时，$l_{v,p}>0$且满足归一化条件$\sum_{v=1}^{M}l_{v,p}=1$，则本文算法为

$$\boldsymbol{\psi}_{p,t}=\boldsymbol{u}_{p,t-1}-\lambda_p\left(\nabla\boldsymbol{u}C_p\left(\boldsymbol{u}_{p,t-1}\right)+\sum_{v\in N_p}d_{v,p}\nabla\boldsymbol{u}C_v\left(\boldsymbol{u}_{p,t-1}\right)\right)\tag{14}$$

$$\boldsymbol{u}_{p,t}=\sum_{v\in N_p}l_{v,p}\boldsymbol{\psi}_{v,t}\tag{15}$$

本文算法主要的机制是：空地网联集群首先利用无线通信技术建立自组织网络，并通过自组网实现传感信息实时交互与共享，节点个体利用本地探测数据实现基于梯度下降的模式学习，同时根据噪声水平实时调整信息融合的权重，以自适应融合来自其他邻域节点共享的信息。因此，整个网联集群通过节点个体的本地学习和节点间信息共享及自适应融合，协同实现对目标的模式识别。其具体包括2个步骤：①传感节点个体p计算本地代价函数梯度信息，通过通信网络向其他节点共享自己的代价函数梯度信息并汇集其他邻域节点共享的代价函数梯度信息，根据式(14)，基于本地代价函数梯度与邻域节点共享的梯度进行自适应学习，获得本地节点对目标模式的临时估计值$\boldsymbol{\psi}_{p,t}$；②节点个体同样通过通信网络共享自己的临时估计值并汇集其他邻域节点共享的对相同目标模式的临时估计值，根据式(15)，节点个体将自己和邻域节点的临时估计值进行加权融合，最终获得更新的本地估计值$\boldsymbol{u}_{p,t}$。

此外，为实现自适应信息融合，每个节点可根据模式识别偏差动态调整融合权重参数$l_{v,p}$，具体地，令节点v和节点p之间的模式识别偏差定义为

$$\delta_{v,p,t-1}^2\triangleq\left\|\boldsymbol{u}_{p,t-1}-\boldsymbol{\psi}_{v,t}\right\|_2^2\tag{16}$$

则利用一阶滤波器可动态估计该偏差：

$$\delta_{v,p,t}^2=\left(1-\chi_p\right)\delta_{v,p,t-1}^2+\chi_p\left\|\boldsymbol{u}_{p,t-1}-\boldsymbol{\psi}_{v,t}\right\|_2^2\tag{17}$$

式中：χ_p为取值近似为1的遗忘因子。

根据估计的偏差，本文算法中节点间交互链路的权重计算式为

$$l_{v,p,t}\approx\delta_{v,p,t-1}^{-2}\Big/\sum_{n\in N_p}\delta_{n,p,t}^{-2}\qquad v\in N_p\cup\{p\}\tag{18}$$

式中：n为节点p的邻域N_p中的节点，本地节点p融合邻域节点v信息的权重系数$l_{v,p,t}$与该邻域节点v的估计偏差成反比。因此，偏差越大表示估计精度越低，信息质量越差，融合权重系数越小，反之，偏差越小精度越高，信息质量越高，权重系数越大。

1.3 误差收敛性分析

根据式(14)和式(15)，引入误差向量：

$$\begin{cases}\tilde{\boldsymbol{\psi}}_{p,t}\triangleq\boldsymbol{u}^{\mathrm{o}}-\boldsymbol{\psi}_{p,t}\\ \tilde{\boldsymbol{u}}_{p,t}\triangleq\boldsymbol{u}^{\mathrm{o}}-\boldsymbol{u}_{p,t}\end{cases}\tag{19}$$

在$\boldsymbol{u}^{\mathrm{o}}$中减去式(14)和式(15)，可得

$$\begin{cases}\tilde{\boldsymbol{\psi}}_{p,t}=\tilde{\boldsymbol{u}}_{p,t-1}+\lambda_p\left(\nabla\boldsymbol{u}C_p\left(\boldsymbol{u}_{p,t-1}\right)+\sum_{v\in N_p}d_{v,p}\nabla\boldsymbol{u}C_v\left(\boldsymbol{u}_{p,t-1}\right)\right)\\ \tilde{\boldsymbol{u}}_{p,t}=\sum_{v=1}^{M}l_{v,p}\tilde{\boldsymbol{\psi}}_{v,t}\end{cases}\tag{20}$$

通过引入网络误差向量以集合所有节点上的误差量：

$$\begin{cases}\tilde{\boldsymbol{\psi}}_t=\begin{bmatrix}\tilde{\boldsymbol{\psi}}_{1,t}\\ \tilde{\boldsymbol{\psi}}_{2,t}\\ \vdots\\ \tilde{\boldsymbol{\psi}}_{M,t}\end{bmatrix}\\ \tilde{\boldsymbol{u}}_t=\begin{bmatrix}\tilde{\boldsymbol{u}}_{1,t}\\ \tilde{\boldsymbol{u}}_{2,t}\\ \vdots\\ \tilde{\boldsymbol{u}}_{M,t}\end{bmatrix}\end{cases}\tag{21}$$

$$\boldsymbol{R}_{v,p,t-1}=\nabla\boldsymbol{u}^2C_v^{\text{local}}\left(\boldsymbol{u}_{p,t-1}\right)\tag{22}$$

并且引入块矩阵：

$$\begin{cases}\boldsymbol{L}=\left[l_{v,p}\right]\otimes\boldsymbol{I}_N\\ \boldsymbol{D}=\left[d_{v,p}\right]\otimes\boldsymbol{I}_N\end{cases}\tag{23}$$

$$\boldsymbol{Q}=\operatorname{diag}\{\lambda_1,\lambda_2,\cdots,\lambda_M\}\otimes\boldsymbol{I}_N\tag{24}$$

$$\boldsymbol{V}_{t-1}=\sum_{v=1}^{M}\operatorname{diag}\left\{d_{v,1}\boldsymbol{R}_{v,1,t-1},d_{v,2}\boldsymbol{R}_{v,2,t-1},\cdots,d_{v,M}\boldsymbol{R}_{v,M,t-1}\right\}\tag{25}$$

式中：⊗表示内积；$\boldsymbol{I}_N$为N维对角矩阵。

递归式(15)和式(21)，可得

$$\tilde{\boldsymbol{u}}_t=\boldsymbol{L}^{\mathrm{T}}\left(\boldsymbol{I}_{NM}-\boldsymbol{Q}\boldsymbol{V}_{t-1}\right)\tilde{\boldsymbol{u}}_{t-1}\tag{26}$$

取式(26)的期望值，可以得到以下平均误差递归方程：

$$E\left(\tilde{\boldsymbol{u}}_t\right)=\boldsymbol{L}^{\mathrm{T}}\left(\boldsymbol{I}_{NM}-\boldsymbol{Q}\boldsymbol{V}_{\infty}\right)E\left(\tilde{\boldsymbol{u}}_{t-1}\right)\qquad t\to\infty\tag{27}$$

令误差演化动力学方程的系数矩阵$\boldsymbol{A}=\boldsymbol{L}^{\mathrm{T}}\cdot\left(\boldsymbol{I}_{NM}-\boldsymbol{Q}\boldsymbol{V}_{\infty}\right)$的谱半径为$\rho(\boldsymbol{A})$，则有以下定理。

定理　令$\eta_{v,\max}$和$\eta_{v,\min}$分别为$\nabla \boldsymbol{u}^2 C_v^{\text{local}}(\boldsymbol{u})$的最大和最小特征根，且令$\theta_{p,\min}=\sum_{v=1}^{M} d_{v,p}\eta_{v,\min}$，$\theta_{p,\max}=\sum_{v=1}^{M} d_{v,p}\eta_{v,\max}$，可知当步长满足$0<\lambda_p<2/\theta_{p,\max}, p\in\{1,2,\cdots,M\}$时，本文算法可保证均方根的收敛性。

证明　根据$\boldsymbol{u}_t=\boldsymbol{A}\boldsymbol{u}_{t-1}$，若$\boldsymbol{u}_t$收敛，可得$\|\boldsymbol{u}_t\|=\|\boldsymbol{A}\boldsymbol{u}_{t-1}\|\leqslant\|\boldsymbol{A}\|\cdot\|\boldsymbol{u}_{t-1}\|$，因此，需要满足$\|\boldsymbol{A}\|<1$；因此，$\boldsymbol{A}$的谱半径$\rho_A$需要满足$\rho_A\leqslant\|\boldsymbol{A}\|<1$，可得$\rho_A<1$；

根据$\boldsymbol{A}=\boldsymbol{L}^{\mathrm{T}}(\boldsymbol{I}_{NM}-\boldsymbol{Q}\boldsymbol{V}_{\infty})$，$\rho_A$可改写为：$\rho(\boldsymbol{A})\leqslant\{|1-\lambda_p\theta_{\max}|,|1-\lambda_p\theta_{\min}|\}<1$，即需要满足$|1-\lambda_p\theta_{\max}|<1$且$|1-\lambda_p\theta_{\min}|<1$。因此，当步长$\lambda_p$取值满足式 (28) 时，则可以保证$\boldsymbol{A}$的稳定性。

$$0<\lambda_p<\frac{2}{\theta_{p,\max}}\qquad p\in\{1,2,\cdots,M\}\tag{28}$$

因此，对于稳定状态有

$$\lim_{t\to\infty}E(\tilde{\boldsymbol{u}}_t)=0\tag{29}$$

式 (29) 表明，本文算法可有效保证空地网联集群协同识别的均方根误差收敛。

证毕

2　仿真结果及分析

2.1　交互拓扑模型建立

本文建立空地网联集群信息交互拓扑模型，该拓扑模型共包含 10 个节点，其中，5 个节点为低空无人机集群 (unmanned aerial vehicles, UAV)，5 个节点为地面智能网联车辆 (connected and autonomous vehicles, CAV)，如图 2 所示。其中，空地集群通过无线自主组网实现节点间信息交互与共享，从而实现协同目标探测与模式识别。红色圆点代表了空地自主网络节点，节点间信息交互拓扑关系通过节点间连接的有向边链，连接线上的箭头代表信息流的流向。有向边的粗细代表 2 个节点间交互信息的能力大小，例如，较粗的连接线表示 2 个节点间交互的数据量较大，反之则表示 2 个节点间交互的数据量较小。

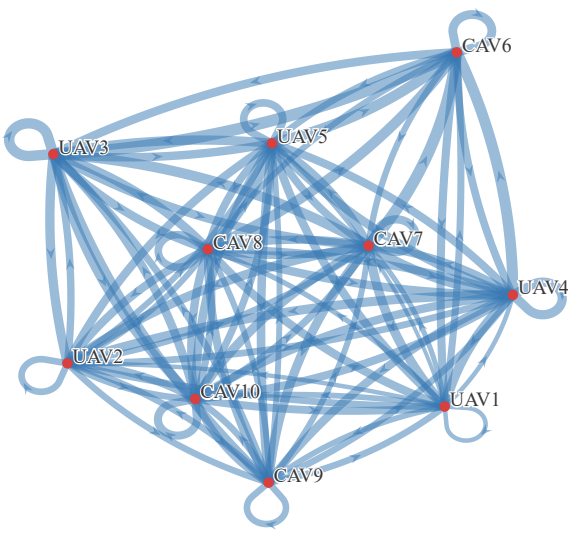


图 2　空地网联集群信息交互拓扑模型(10 节点)

Fig. 2　Air-ground networked cluster information exchange topology model (10 nodes)

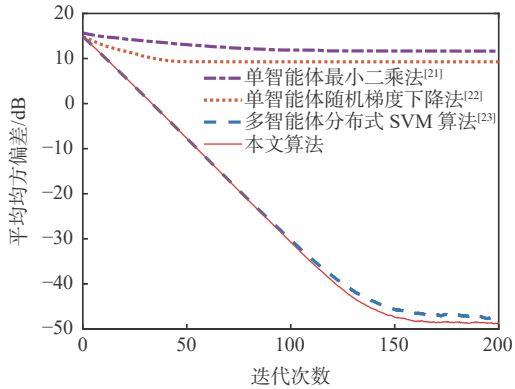
本文基于该仿真场景开展实验，并利用雷达实测数据集根据本文算法进行仿真测试，并与参数估计和模式识别常采用的单智能体最小二乘法[21]、单智能体随机梯度下降法[22]、针对智能网联集群的多智能体分布式支持向量机 (support vector machine, SVM) 算法[23]进行对比。其中，雷达数据集由高频天线相控阵的反射信号数据组成。数据集中共包含 2 类数据，一类包含跟踪目标且精度较高的雷达反射信号，另一类为不包含跟踪目标且精度较低的雷达反射信号，无人机集群和地面智能网联车辆集群通过自组网方式对雷达数据进行分布式融合，协同实现反射信号模式的自主识别，提升对被跟踪目标的检测能力。随机生成自适应融合方法的步长 (最大仿真步长为 0.001)，根据式 (17)，通过采用最大步长为 0.01 的一阶滤波器对样本进行平滑，整个集群网络对信息融合估计的平均均方误差和系统误差如图 3 所示。

由图 3 可知，随着迭代次数的增加，针对空地网联集群的多智能体分布式 SVM 算法和本文算法的平均均方偏差和系统误差随迭代次数的增加不断降低并稳定收敛，而单智能体最小二乘法及单智能体随机梯度下降法的误差曲线无显著变化。本文算法的均方误差和系统误差均显著低于单智能体最小二乘法及随机梯度下降法，略低于多智能体分布式 SVM 算法。当节点数为 10 时，单智能体最小二乘法、单智能体随机梯度下降法、多智能体分布式 SVM 算法和本文算法的平均均方偏差分别为 11.64 dB、9.27 dB、−47.44 dB 和−48.70 dB，系统误差分别为 18.65 dB、16.29 dB、−26.05 dB 和−27.42 dB。与其他算法相比，本文算法的系统误差更逼近理论误差下界，即−30.92 dB，具有良好的精度。

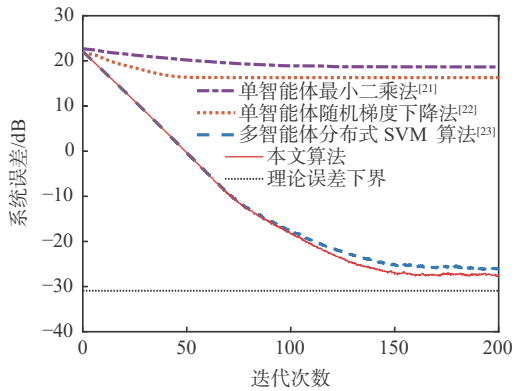
2.2　不同网络节点对网络误差的影响

为研究网络节点对网络误差的影响，本文采用包含 20～40 个节点的空地网联集群信息交互拓扑模型进行对比实验，评估在信息融合估计过程中采用分布式融合算法及传统方法的空地网联集群网络的平均均方误差和系统误差。通过比较不同网络节点下的网络性能，从而量化节点数量对网络误差的影响。

建立包含 20 个节点的空地网联集群信息交互拓扑模型，其中，15 个节点为 UAV，5 个节点为地面智能 CAV，如图 4 所示。整个集群网络对信息融合



(a) 平均均方偏差



(b) 系统误差

图3 空地网络(10节点)信息融合估计误差

Fig. 3 Information fusion estimation error of air-ground swarms with 10 nodes

估计的平均均方误差和系统误差如图5所示。

由图5可知,当节点数为20时,随着迭代次数的增加,针对空地网联集群的多智能体分布式SVM算法和本文算法的平均均方偏差和系统误差随迭代次数的增加不断降低并稳定收敛,而单智能体最小二乘法及单智能体随机梯度下降法的误差曲线无显著变化。本文算法的均方误差和系统误

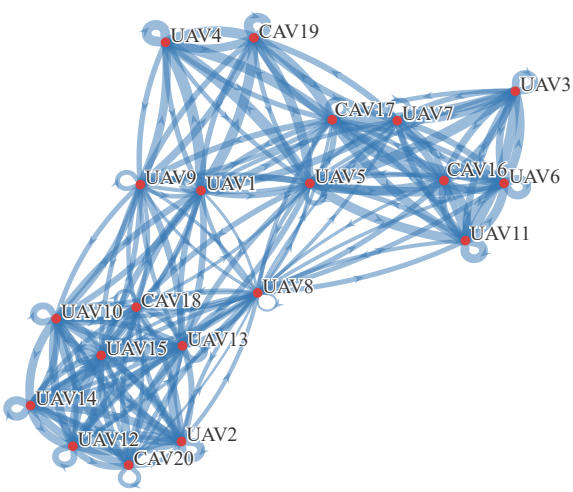
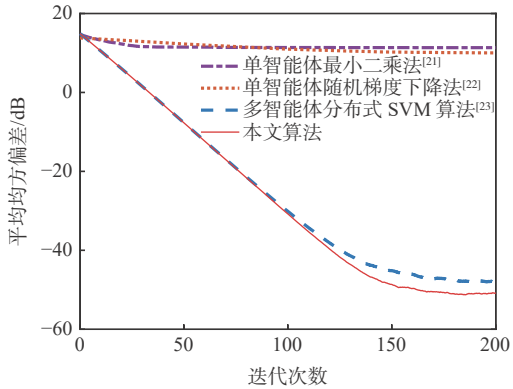
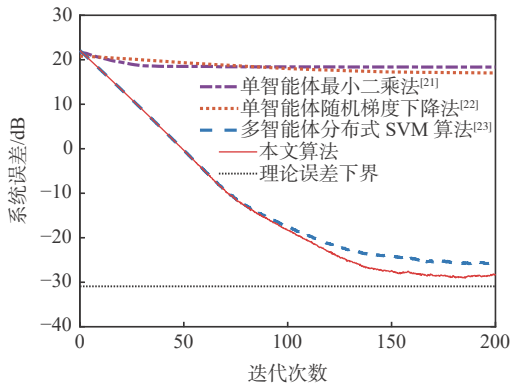


图4 空地网联集群信息交互拓扑模型(20节点)

Fig. 4 Air-ground networked cluster information exchange topology model (20 nodes)



(a) 平均均方偏差



(b) 系统误差

图5 空地网络(20节点)信息融合估计误差

Fig. 5 Information fusion estimation error of air-ground swarms with 20 nodes

差均显著低于单智能体最小二乘法和单智能体随机梯度下降法,略低于多智能体分布式SVM算法。其中,单智能体最小二乘法、单智能体随机梯度下降法、多智能体分布式SVM算法和本文算法的平均均方偏差分别为11.35 dB、10.02 dB、−47.81 dB和−50.80 dB,系统误差分别为18.35 dB、17.03 dB、−25.67 dB和−28.42 dB。与其他算法相比,本文算法的系统误差更逼近理论误差下界,即−30.92 dB,具有良好的精度。

建立包含30个节点的空地网联集群信息交互拓扑模型,其中,25个节点为UAV,5个节点为地面智能CAV,整个集群网络交互拓扑模型如图6所示。集群网络对信息融合估计的平均均方误差和系统误差如图7所示。

由图7可知,当节点数为30时,随着迭代次数的增加,针对空地网联集群的多智能体分布式SVM算法和本文算法的平均均方偏差和系统误差随迭代次数的增加不断降低并稳定收敛,而单智能体最小二乘法及单智能体随机梯度下降法的误差曲线无显著变化。本文算法的均方误差和系统误差均显著低于单智能体法,略低于多智能体分布式SVM算法。其中,单智能体最小二乘法、单智能体随机梯度下降法、多智能体分布式SVM算法和本

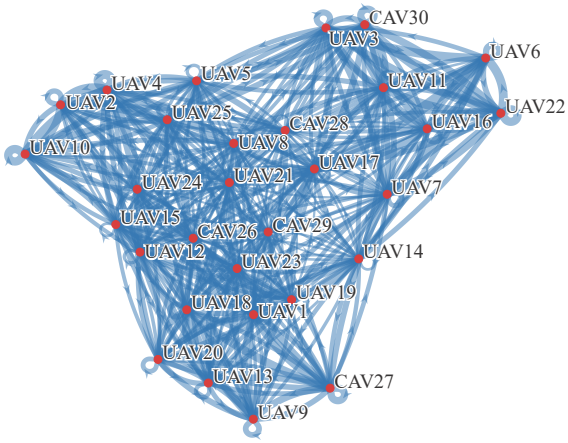
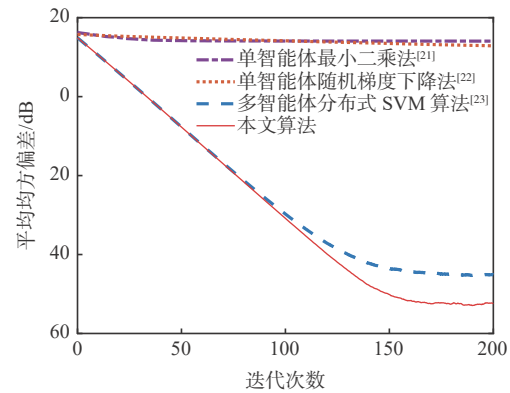
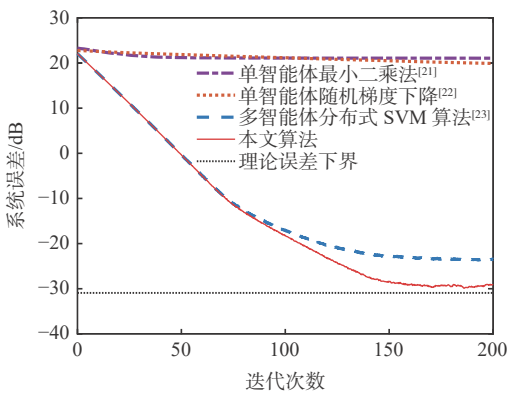


图 6　空地网联集群信息交互拓扑模型(30 节点)

Fig. 6　Air-ground networked cluster information exchange topology model (30 nodes)



(a) 平均均方偏差



(b) 系统误差

图 7　空地网络(30 节点)信息融合估计误差

Fig. 7　Information fusion estimation error of air-ground swarms with 30 nodes

文算法的平均均方偏差分别为 14.06 dB、12.86 dB、−45.17 dB 和−52.45 dB, 系统误差分别为 21.07 dB、19.89 dB、−23.48 dB 和−29.16 dB。与其他算法相比, 本文算法的系统误差更逼近理论误差下界, 即−30.92 dB, 具有良好的精度。

建立包含 40 个节点的空地网联集群信息交互拓扑模型, 其中, 35 个节点为 UAV, 5 个节点为地面智能 CAV, 整个集群网络交互拓扑模型如图 8 所示。整个集群网络对信息融合估计的平均均方误差和系统误差如图 9 所示。

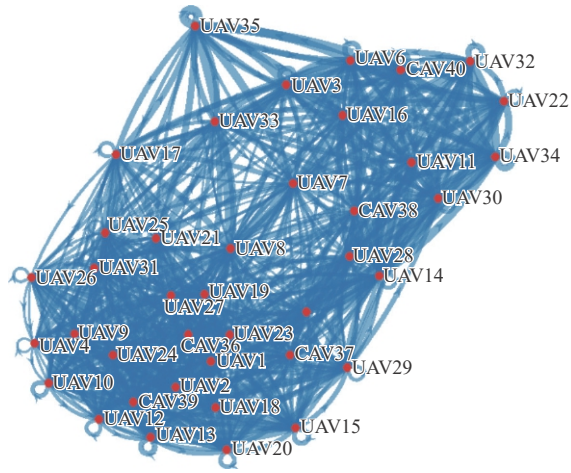
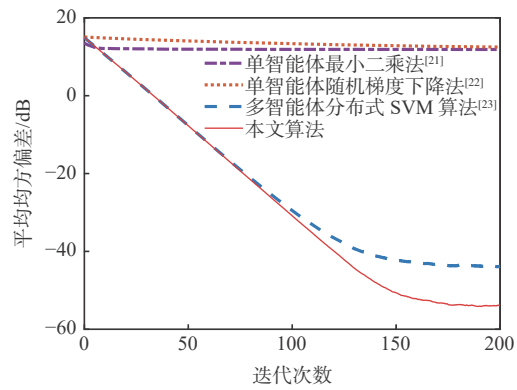
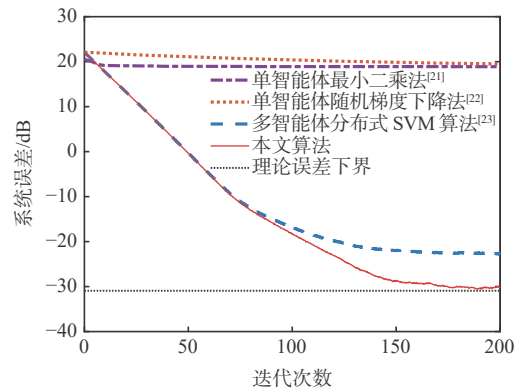


图 8　空地网联集群信息交互拓扑模型(40 节点)

Fig. 8　Air-ground networked cluster information exchange topology model (40 nodes)



(a) 平均均方偏差



(b) 系统误差

图 9　空地网络(40 节点)信息融合估计误差

Fig. 9　Information fusion estimation error of air-ground swarms with 40 nodes

由图 9 可知, 当节点数为 40 时, 随着迭代次数的增加, 针对空地网联集群的多智能体分布式 SVM 算法和本文算法的平均均方偏差和系统误差随迭代次数的增加不断降低并稳定收敛, 而单智能体最小二乘法及单智能体随机梯度下降法的误差曲线无显著变化。本文算法的均方误差和系统误

差均显著低于单智能体法，略低于多智能体分布式SVM算法。其中，单智能体最小二乘法、单智能体随机梯度下降法、多智能体分布式SVM算法和本文算法的平均均方偏差分别为11.86 dB、12.50 dB、−43.89 dB和−53.96 dB，系统误差分别为18.89 dB、19.51 dB、−22.73 dB和−30.22 dB。与其他算法相比，本文算法的系统误差更逼近理论误差下界，即−30.92 dB，具有良好的精度。

由图3～图9可知，随着迭代次数的增加，本文算法的平均均方偏差和系统误差均显著降低，并逐渐稳定收敛。相比之下，传统单智能体法的误差曲线无显著变化。特别地，当节点数分别为10～40时，本文算法的均方误差和系统误差均显著低于单智能体算法，略低于多智能体分布式SVM算法，且更逼近理论最优水平。这一结果表明，在处理空地网联集群的分布式融合问题时，本文算法相较于传统算法具有更高的精度和更优的收敛性能。

2.3 不同学习率对网络误差的影响

为探究学习率对网络误差的影响，本文利用30个节点的空地网联集群信息交互拓扑模型，设计了4种不同的学习率，分别为10^{-4}、10^{-3}、10^{-2}及10^{-1}，以此来评估采用本文算法的空地网联集群网络在信息融合估计过程中的平均均方误差和系统误差。

图10为不同学习率下的空地网联集群网络的

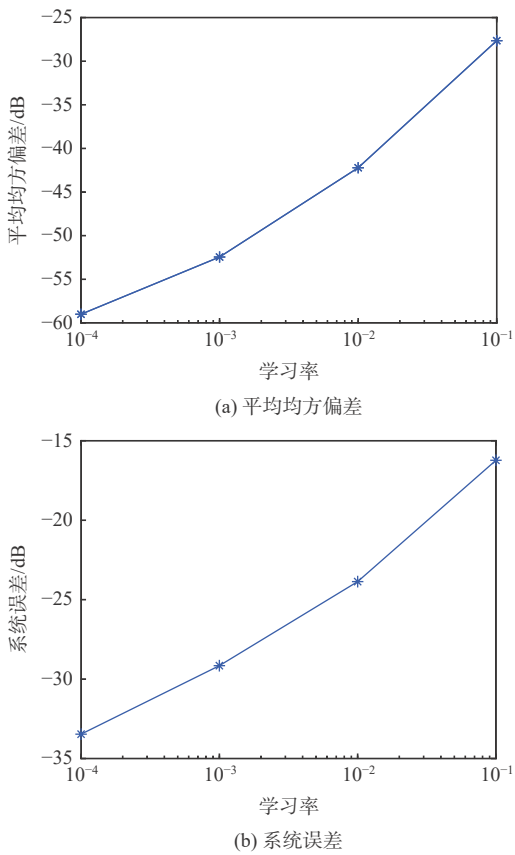


图10 不同学习率下的网络误差

Fig. 10 Network error under different learning rate

误差。由图可知，平均均方偏差和系统误差与学习率呈反比。当学习率为10^{-4}时，平均均方偏差和系统误差最低，分别为−59.01 dB和−33.47 dB；当学习率为10^{-2}时，平均均方偏差为−42.22 dB，系统误差为−23.86 dB；当学习率为10^{-1}时，平均均方偏差为−27.65 dB，系统误差为−16.23 dB。综上所述，学习率降低可提升信息估计的精度，较大的学习率虽然可以提升收敛速度，但会带来较高的估计误差。

3 结 论

为实现空地网联集群协同模式识别，本文在梯度下降学习算法中引入基于空地集群自组织网络的信息扩展与自适应融合机制，提出空地网联集群分布式融合算法，利用雷达数据集开展了仿真测试，结果表明：

1）采用本文算法对信息估计的均方偏差及系统误差均显著低于单智能体算法，略低于多智能体分布式SVM算法，且更逼近理论误差下界最优水平。因此，在处理空地网联集群的分布式融合问题时，本文算法相较于传统算法具有更高的精度和更优的收敛性能；

2）本文算法无需采用中心基站或集中式控制机构，仅通过移动节点分布式协同学习和信息融合即可实现对目标的识别，因此，在基于分布式雷达组网的低空空域监视、基于异构自主智能集群的协同作战等领域具有广阔的应用前景；

3）利用本文算法可有效实现二分类模式识别，可应用于无人机与地面网联车辆等自主组网的空地智能集群的协同探测目标。

在本文的基础上，未来研究工作将设计由多个二分类器组合构成多分类模式识别算法，从而实现在多类别目标识别场景中的应用。

参考文献（References）

[1] SUN J L, LIU F, ZHOU Y Z, et al. Surveillance plane aided air-ground integrated vehicular networks: architectures, applications, and potential[J]. IEEE Wireless Communications, 2020, 27(6): 122-128.

[2] 牛志升, SHEN S, 张钦宇, 等. 面向沉浸式体验的空天地一体化车联网体系架构与关键技术[J]. 物联网学报, 2017, 1(2): 17-27.
NIU Z S, SHEN S, ZHANG Q Y, et al. Space-air-ground integrated vehicular network for immersive driving experience[J]. Chinese Journal on Internet of Things, 2017, 1(2): 17-27(in Chinese).

[3] 中国联通网络技术研究院. 空天地一体化通信网络白皮书[EB/OL].(2020-06-16)[2023-02-05].https://max.book118.com/html/2020/0616/8032036033002117.shtm.
China Unicom Network Technology Research Institute. White paper on space–air–ground integrated communication networks

[EB/OL].(2020-06-16)[2023-02-05].https://max.book118.com/html/2020/0616/8032036033002117.shtm.

[4] YOU X H, WANG C X, HUANG J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts[J]. Science China Information Sciences, 2020, 64(1): 110301.

[5] LIU R, LIU A F, QU Z Z, et al. An UAV-enabled intelligent connected transportation system with 6G communications for Internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(2): 2045-2059.

[6] 中华人民共和国国务院. 国家综合立体交通网规划纲要[EB/OL]. (2021-02-24)[2021-02-25]. http://www.gov.cn/zhengce/2021-02/24/content_5588654.htm.
People's Republic of China State Council. National comprehensive transport network planning outline of the people's republic of China [EB/OL]. (2021-02-24) [2021-02-25]. http://www.gov.cn/zhengce/2021-02/24/content_5588654.htm.

[7] 赵亚军, 郁光辉, 徐汉青. 6G 移动通信网络: 愿景、挑战与关键技术[J]. 中国科学: 信息科学, 2019, 49(8): 963-987.
ZHAO Y J, YU G H, XU H Q. 6G mobile communication networks: vision, challenges, and key technologies[J]. Scientia Sinica (Informations), 2019, 49(8): 963-987(in Chinese).

[8] 张平, 张建华, 戚琦, 等. Ubiquitous-X: 构建未来 6G 网络[J]. 中国科学: 信息科学, 2020, 50(6): 913-930.
ZHANG P, ZHANG J H, QI Q, et al. Ubiquitous-X: constructing the future 6G networks[J]. Scientia Sinica (Informations), 2020, 50(6): 913-930(in Chinese).

[9] CAI Y F, SEKIYAMA K. Geometric relation-based cognitive sharing for flying and ground mobile robot cooperation[J]. Journal of Advanced Computational Intelligence and Intelligent Informatics, 2016, 20(6): 919-927.

[10] LI J Q, DENG G Q, LUO C W, et al. A hybrid path planning method in unmanned air/ground vehicle (UAV/UGV) cooperative systems[J]. IEEE Transactions on Vehicular Technology, 2016, 65(12): 9585-9596.

[11] BHAMIDIPATI S, GAO G X. Locating multiple GPS jammers using networked UAVs[J]. IEEE Internet of Things Journal, 2019, 6(2): 1816-1828.

[12] WAN M J, GU G H, QIAN W X, et al. Unmanned aerial vehicle video-based target tracking algorithm using sparse representation[J]. IEEE Internet of Things Journal, 2019, 6(6): 9689-9706.

[13] CHEN X, FENG Z Y, WEI Z Q, et al. Performance of joint sensing-communication cooperative sensing UAV network[J]. IEEE Transactions on Vehicular Technology, 2020, 69(12): 15545-15556.

[14] EBRAHIMI D, SHARAFEDDINE S, HO P H, et al. UAV-aided projection-based compressive data gathering in wireless sensor networks[J]. IEEE Internet of Things Journal, 2019, 6(2): 1893-1905.

[15] ZHENG Y, CHEN Z, LV D L, et al. Air-to-air visual detection of micro-UAVs: an experimental evaluation of deep learning[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1020-1027.

[16] ALBANESE A, SCIANCALEPORE V, COSTA-PÉREZ X. SARDO: an automated search-and-rescue drone-based solution for victims localization[J]. IEEE Transactions on Mobile Computing, 2022, 21(9): 3312-3325.

[17] PAN Y, LI S N, NING Z Y, et al. auSense: collaborative airspace sensing by commercial airplanes and unmanned aerial vehicles[J]. IEEE Transactions on Vehicular Technology, 2020, 69(6): 5995-6010.

[18] KOUTRAS D I, KAPOUTSIS A C, KOSMATOPOULOS E B. Autonomous and cooperative design of the monitor positions for a team of UAVs to maximize the quantity and quality of detected objects[J]. IEEE Robotics and Automation Letters, 2020, 5(3): 4986-4993.

[19] LIU C H, DAI Z P, ZHAO Y N, et al. Distributed and energy-efficient mobile crowdsensing with charging stations by deep reinforcement learning[J]. IEEE Transactions on Mobile Computing, 2021, 20(1): 130-146.

[20] JOUHARI M, AL-ALI A K, BACCOUR E, et al. Distributed CNN inference on resource-constrained UAVs for surveillance systems: design and optimization[J]. IEEE Internet of Things Journal, 2022, 9(2): 1227-1242.

[21] CHRISTOPHER M B. Pattern recognition and machine learning (information science and statistics) [M]. Berlin: Springer, 2006.

[22] BOTTOU L. Large-scale machine learning with stochastic gradient descent[C]//Proceedings of the COMPSTAT'2010. Berlin: Springer, 2010, 177-186.

[23] SCARDAPANE S, FIERIMONTE R, DI LORENZO P, et al. Distributed semi-supervised support vector machines[J]. Neural Networks, 2016, 80: 43-52.

Cooperative pattern recognition method for air-ground networked swarms

QU Guixian[1, 2, 3], ZHOU Jianshan[4, *], SI Yang[5], LIU Xiaojing[1, 2, 3], YUAN Qiyu[1, 2, 3], MA Qinglin[2, 3, 6]

(1. Research Institute of Aero-Engine, Beihang University, Beijing 102206, China;
2. Aero-Engine System Collaborative Design Center, Beihang University, Beijing 102206, China;
3. Tianmushan Laboratory, Hangzhou 310023, China;
4. School of Transportation Science and Engineering, Beihang University, Beijing 100191, China;
5. Logistics School, Beijing Wuzi University, Beijing 101149, China;
6. School of Energy and Power Engineering, Beihang University, Beijing 100191, China)

Abstract: The air-ground networked swarms have excellent application potential in national economics and production, such as smart cities, smart agriculture and forestry, and intelligent transportation. It also shows an exceptional value of application in the military fields, such as battlefield situational awareness and air-ground cooperative strikes. This paper addresses the need for accurate perception and recognition of complex environmental targets in air-ground networked swarms. We present a distributed learning and adaptive information fusion approach for air-ground network clustering, and we develop a model that minimizes the global likelihood function based on the probability of pattern categorization. This algorithm includes two main steps: information diffusion based on the gradient descent method and information fusion based on adaptive weighting calculation, and forms an air-ground cooperative pattern recognition method. Furthermore, the average error recursive equation for the cooperative pattern recognition algorithm of the air-ground networked clusters is derived, and the error convergence of the algorithm is theoretically proven. The results of the simulation demonstrate that the distributed fusion algorithm is accurate and that it can converge to the theoretical optimal level of the system in terms of both average mean square deviation and system error of information estimation. The simulation results show that the distributed fusion algorithm has good accuracy, and the algorithm's average mean square deviation and system error of information estimation can converge to the system's theoretical optimal level. When the number of nodes increases from 10 to 40, the mean square deviation of the distributed fusion algorithm for air-ground networked swarms decreases from −48.70 dB to −53.96 dB, and the system error reduces from −27.42 dB to −30.22 dB, which is close to the theoretical value of the error. Comparative experiments show that the algorithm proposed in this paper has good accuracy compared to traditional methods and can effectively support the perception and recognition of complex environmental targets for air-ground networked swarms.

Keywords: air-ground networked swarms; cooperative pattern recognition; information interaction topology model; distributed fusion algorithm; measured radar dataset

Received: 2023-11-04; **Accepted**: 2024-03-04; **Published Online**: 2024-03-18 16:08

URL: link.cnki.net/urlid/11.2625.V.20240317.1852.001

Foundation items: National Natural Science Foundation of China (52302510,52202391); Science Center for Gas Turbine Project (P2023-B-I-005-001); The Fundamental Research Funds for the Central Universities (YWF-23-Q-1066); Tianmushan Laboratory Cross-Innovation Research Team Project (TP-2024-A-001,TK-2024-D-001)

* **Corresponding author**. E-mail: jianshanzhou@foxmail.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0768

面向石油物探领域的HPC故障与性能分析

王明键[1]，赵长海[1,*]，李超[2]，文佳敏[1]，尚民强[1]，侯红军[1]

(1. 中国石油集团东方地球物理勘探有限责任公司 物探技术研发中心，涿州 072751；
2. 北京航空航天大学 计算机学院，北京 100191)

摘　　要：如今，石油物探领域地震资料采集技术不断进步，数据规模已达到TB甚至PB级别，随着数据量，运行时间，以及高性能计算（HPC）集群中节点数的增加，集群出现问题的概率和维护难度也随之增加。当集群或节点出现故障时往往需要重新运行计算程序，造成了极大的资源浪费。为解决计算集群中HPC程序可观测性低、故障和性能分析困难的问题，借鉴开放追踪标准(OTF)与分布式链路追踪的思想，提出一种面向生产环境的低侵入式高性能计算集群和程序故障分析方法，该方法不仅能够对生产环境中HPC程序进行高效观测，还具有低侵入性的特点，可以在几乎不修改代码的前提下与现有应用程序结合使用。将所提方法用于生产环境中分布式“抽道集”排序程序进行采集分析，验证了所提方法的有效性，发现了程序中隐藏的软件缺陷和性能薄弱环节。

关　键　词：分布式计算；大规模集群；故障分析；链路追踪；生产环境
中图分类号：TP301.6
文献标志码：A　　**文章编号**：1001-5965(2026)01-0157-10

随着石油物探领域勘探研究开发内容由常规油气藏向复杂油气藏逐渐转变，由常规油气资源向非常规油气资源不断扩展，宽方位、宽频带、高密度的“两宽一高”地震技术不断取得新突破，野外采集得到的地震数据精度不断提高，规模不断增大[1]，普通计算机已经难以对如此大规模数据进行计算处理，使用高性能计算（high- performance compute, HPC）集群进行数据处理已经成为石油物探领域的基本要求。

为解决大规模集群中计算程序难以开发和维护等问题，学术界产生了如消息传递接口（message passing interface, MPI)[2-4]和Hadoop[5-6]这样的分布式计算框架。这些框架在一定程度上接管了分布式计算中的数据传输和节点同步问题。开发者在使用这些框架进行开发时可以专注到计算本身，而无需了解复杂的计算机知识，在一定程度上减少了开发者的负担。

即便如此，在实际生产应用中不难发现，并行程序出现故障从而导致计算失败的情况时有发生，其原因既包括软硬件的固有缺陷，也包括数据规模增长超出软件自身的处理能力。此外，随着时间的增长和集群中硬件的老化，集群中可能会出现一些诸如I/O速度变慢、计算时间变长等非致命性的性能下降，性能下降也有可能诱发软件缺陷导致计算失败。已有研究表明，计算集群中平均每30 min就有可能遇到异常情况[7]。尽管Hadoop、Spark等目前流行的并行计算框架和石油勘探等专有领域也有许多对计算任务运行时容错的探索[8-9]，使得轻微故障不打断程序的正常运行，然而，这不过是集群出现偶然错误时的权宜之计。轻微故障不及时解

收稿日期：2023-11-24；**录用日期**：2024-02-05；**网络出版时间**：2024-03-29 07:22
网络出版地址：link.cnki.net/urlid/11.2625.V.20240328.1001.001
基金项目：中国石油天然气集团公司关键核心技术攻关项目“地震处理解释关键新技术研究与智能化软件研发”子课题“iEco软件平台智能化升级”(2021ZG03-05)；中国石油集团东方地球物理公司技术攻关项目“GeoEast平台功能扩充与优化”(01-01-2023)
* **通信作者**. E-mail：zhao.ch@cnpc.com.cn

引用格式：王明键，赵长海，李超，等. 面向石油物探领域的HPC故障与性能分析[J]. 北京航空航天大学学报，2026，52（1）：157-166.
WANG M J，ZHAO C H，LI C，et al. HPC fault and performance analysis in field of petroleum geophysical exploration[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：157-166（in Chinese）.

决，久而久之随着错误的积累，当集群中大部分节点处于不稳定状态时，无论什么精妙的容错算法也无法包容错误，此时再考虑解决集群中的故障所付出的成本将会成倍增加。

工业界对于并行计算程序不断增长的巨大需求与并行计算程序开发调试困难之间的矛盾催生出了大量相关研究，对于并行程序的性能分析和调优方法也层出不穷，比如常见的TAU[10]和PerfExpert[11]等，这些工具功能强大，应用广泛。但是其在运行时不仅需要消耗大量磁盘空间，还对应用程序的执行效率产生了巨大影响[12]，使得此种工具只能在开发环境中小规模数据集上使用。而开发环境和生产环境间的差异、测试数据与实际数据间的差异往往会导致调优的方向错误，甚至带来额外的问题。

首先，为解决上述问题，本文对分布式程序的计算流程进行了总结，并借鉴开放追踪标准（open tracing format, OTF）[13]规范提出了一种面向HPC的易于理解和使用的分布式程序状态追踪接口—并行编程框架（geoeast parallel programing, GPP）Tracer；现有生产环境中已经拥有大量正在使用的计算程序，为能够采集数据将这些程序的源代码一一修改并不现实。为解决这个问题，本文设计了一种面向生产环境的低侵入式高性能计算集群和程序故障分析方法，能够从链接、运行时库和系统调用3个层面对高性能计算程序实现低侵入式的状态追踪，开发者无需修改代码即可使程序拥有故障与性能分析的能力。本文使用基于国产大型地震勘探软件GeoEast GPP[14] MapReduce的“抽道集”分布式排序程序对本文方法进行了分析和验证。

1 基于OTF的分布式程序状态追踪接口

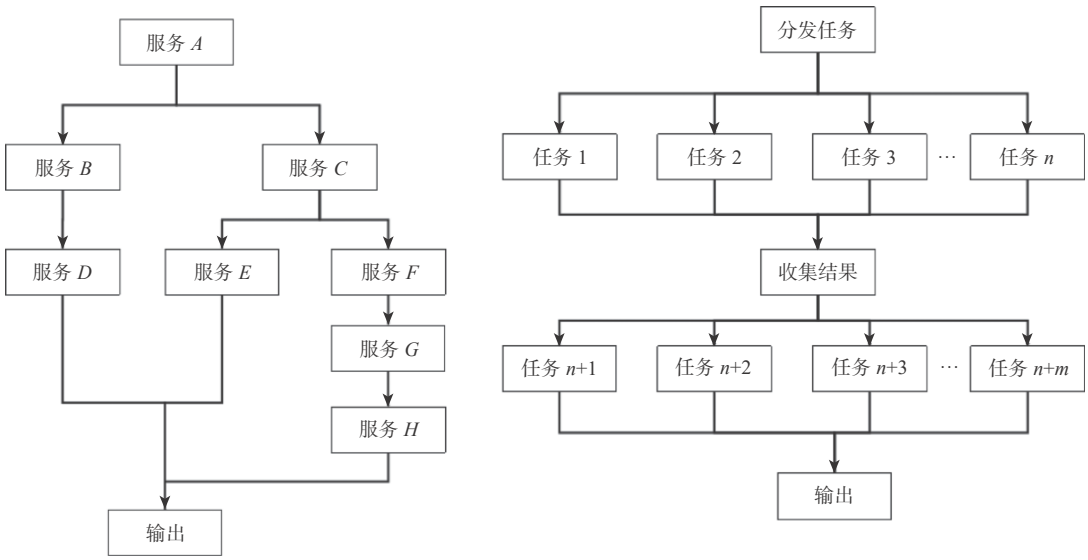
分布式链路跟踪，也称为分布式请求跟踪，是一种用于分析和监视应用程序的方法，特别是使用微服务体系结构构建的应用程序。分布式跟踪技术有助于查明故障发生的位置及导致性能低下的原因,目前在微服务和移动互联网领域十分流行[15-17]。OTF是一种由云原生计算基金会（Cloud Native Computing Foundation, CNCF）提出的建议规范[18-19]，其中，规定了如何定义微服务架构中的各项概念、如何描述一次建立在微服务上的请求及如何对这些请求进行追踪。OTF由应用程序接口（application program interface, API）规范实现了该规范的框架、库及项目文档组成。只要遵循OTF中API规范，开发人员可以使用任何工具向代码中添加自定义的追踪项目。

在OTF的设计中，每次针对微服务系统的调用称为一条Trace，针对具体某个具体微服务的访问称为一个Span[19]。Trace并非被显示定义或创建，而是通过归属于此调用的Span隐性定义，一条Trace可以被认为是由多个Span组成的有向无环图（directed acyclic graph, DAG），Span间的关系被命名为Reference。然而，OTF[18]为在移动互联网中进行微服务链路追踪时使用的接口规范，微服务间的互相调用可以嵌套任意多层，每个服务可能依赖任意多个其他服务，节点间的调用顺序不固定且会随服务需求的变化而更改。而高性能计算任务的计算特征与微服务调用有显著不同。在高性能计算中，往往是由主节点向各个子节点分配任务，子节点计算完成后再将结果返回给主节点或输出到共享磁盘中。因此，难以将面向微服务的OTF链路追踪技术直接套用在高性能计算领域。本文对OTF进行了适当的简化和扩展以适应高性能计算的实际需求。

经调研分析可知，复杂场景下的微服务调用链路较为错综复杂，多个微服务在调用链中呈线性排布，链路较为“深入”，调用链路也可能根据输入数据的不同而变化。而在高性能计算中，以fork-join[20]模型为例，大任务被单一的根节点拆分成若干个较小的计算任务，交由各个子节点进行分布式计算，等待全部子节点计算完成后，根节点再将各个子节点的输出数据进行收集整合，得到最终的输出结果，在此场景下节点间的调用链较为“扁平”，与微服务调用链路有显著差别。微服务和高性能计算2种场景特征如图1所示。

由于微服务调用链路复杂且深入的特性，OTF规范中设计了较为多样的Span关系用于精细地描述微服务调用链路。而在分布式计算领域中，调用链路平坦而简单，OTF复杂精细的接口设计在分布式计算中难以发挥全部功能，反而还会为开发者带来额外的学习成本。因此，为能够针对性地解决高性能计算任务面临的实际问题，降低开发者的学习成本，本文设计的分布式程序状态追踪接口基于OTF规范进行了适当的简化和扩展。

在本文设计的程序状态追踪接口中，主要分为任务开始、结束和报告任务进度3个接口。当计算进程启动后，主节点负责将输入数据进行拆分，生成若干个较小的计算任务并分配给各个子节点进行计算，此时，主节点应该调用任务开始接口。随后，在任务计算的过程中，子节点可以随时调用报告任务进度接口向本程序状态追踪接口报告当前



(a) 微服务调用链路特点　　(b) 分布式高性能程序调用链路特点

图 1　微服务与 HPC 调用链特征对比

Fig. 1　Comparison of call chain characteristic between microservice and HPC

的计算进度, 使得开发者可以观测到此计算任务的内部细节。当子节点计算完成后通知主节点, 再由主节点调用任务结束接口报告此计算任务已经结束。如此便完成最基本的程序运行和追踪过程。如果此计算任务较为复杂, 子节点在计算过程中有必要再获得其他子节点的结果, 也可以对任务开始和结束接口进行嵌套调用。程序状态追踪过程如图 2 所示。

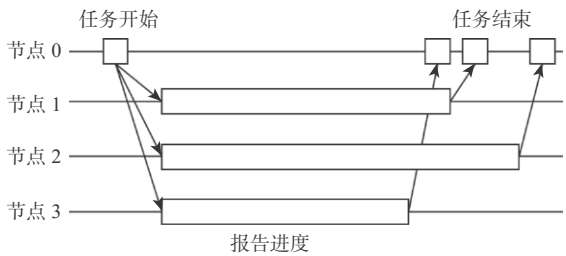


图 2　程序状态追踪过程

Fig. 2　Program status tracking process

本文基于本节的设计思路, 结合分布式高性能计算程序的实际需要, 参考 OTF 规范思想并对 OTF 规范进行适当裁剪和扩展, 设计了一套能够满足分布式高性能程序状态采集的接口。

与 OTF 类似, 本文设计的程序状态追踪接口也拥有 Tracer 数据模型作为程序监控的起点。在高性能计算程序中, Tracer 将作为单例模式存在, 在计算程序中每个进程仅存在唯一的实例, 可以创建 Task 和 IO 子数据模型。Tracer 应具有的功能如表 1 所示。

在本文设计的程序状态追踪接口中, 将 Span 细化为 Task 和 IO 这 2 个具体数据结构, 分别对应运行一个子 Task 和启动一次 IO 操作。本文设计的

表 1　GPP Tracer 接口

Table 1　GPP Tracer interface

方法名	方法说明	返回值	参数列表	参数说明
Init	初始化 Tracer	无	Rank	必填, 本节点的Rank号
			Host	必填, 本节点的主机名
			Name	必填, 本计算程序名称
CreateTask	创建一个Task	Task	Name	必填, 任务名称
			Id	必填, 任务ID
			RankId	必填, 该任务被指派的节点ID
			Info	选填, 任务的其他信息
CreateIO	创建一个IO操作	IO Operate	Type	必填, IO操作的类型
			Fd	必填, 本次操作的文件描述符
			Target	选填, 本IO操作对应的对象
AddLog	添加日志	无	Level	日志等级
			Content	日志内容

接口中不再包含复杂的 Task 间关系, 而只是由任务的“发起节点”和“被指派节点”来描述任务间关系, 父任务用来发起子任务并等待其结果。当调用 Tracer 中的 CreateTask 时, 本 Tracer 自动成为父任务, 其返回的 Task 为子任务, 从而使本接口具有描述复杂嵌套任务的能力。

由于在集群上任务的发起节点和被指派节点往往不是同一台节点, 并且本接口简化了 OTF 中的带内传输功能。为能在子节点上获得 Task 对象, 本接口允许在多个节点上创建同一个 Task, 当 CreateTask 的 RankId 参数与 Tracer 的 Rank 值相同时, 认为本节点为子节点, 此 CreateTask 的操作仅仅是为获得 Task 对象, 并没有真实进行创建任务的操作。

当某节点进行 IO 操作时, 应调用 CreateIO 接

口进行采集。能够采集的IO操作包括对文件和Socket的打开关闭、读取写入等，为缩减采集模块本身的性能开销，应在IO操作出现故障时进行采集。本文设计的Task和IO操作接口如表2所示。

表2 本文设计的Task和IO接口

Table 2 The designed Task and IO interface

接口	方法名	方法说明	返回值	参数列表	参数说明
Task接口	State	更新任务状态	无	State	必填，用于描述任务状态，如当前任务进度等
	Finish	结束该任务	无	Error	选填，用于描述任务错误信息，不填表示该任务无错误
IO接口	Finish	结束IO操作	无	State	必填，IO结束时的状态，如调用IO函数的返回值等
				Info	选填，本操作的自定义信息
				Error	选填，该操作是否出错，不填表示无错误

2 本文方法

2.1 基于构建时的程序状态采集方法

在程序开发中，拥有源代码代表着拥有最丰富最完善的信息，修改源代码即意味着能够实现任何功能。但是在实际生产环境中，对现有源代码进行大量修改并不现实，因此，本文设计了一种基于构建时对程序的特定函数进行动态插桩，从而采集程序状态的方法。

C/C++程序的构建过程分为编译和链接2个部分，编译是指将文本格式的源代码翻译成可执行和链接格式（executable and linkable format, ELF）的二进制文件，而链接是指将多个ELF文件和库文件组合在一起，形成最终的可执行文件[21]。在链接时需要进行符号查找操作，即找到函数和全局变量的地址并填入最终的机器码指令中。在链接器相关功能的支持下[22-23]，此过程可以经人工干预，即当某段代码尝试调用其他文件中的read函数时，链接器可以不将read的地址填入函数调用指令，而是填入预先约定好的__wrap_read函数的地址，在__wrap_read中将本次调用所需的特定信息记录妥当之后再通过__real_read符号调用真实的read函数，完成相应功能，如图3所示。这种方式即可实现在不修改已有源代码的情况下对特定函数进行插桩，通过谨慎控制被插桩函数的数量可以大幅缩减插桩采集对程序带来的运行时负担。

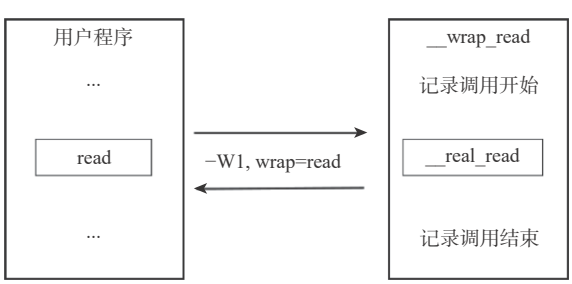


图3 基于构建时符号重定向的程序状态采集方法

Fig. 3 Method for collecting program state based on link time symbol redirection

根据引言中对高性能计算程序和集群故障特点分析可知，虽然在集群和程序上发生的故障经常涉及磁盘、网络、内存、CPU等多个不同方面，但是当故障发生时，其现象往往最先表现在IO上，如IO速度变慢、失败率增高等。因此，为能全面高效地捕捉到高性能计算程序内部状态，快速发现故障，本文使用构建时符号重命名技术对IO相关函数进行插桩采集，选定的函数共包括write、read、connect、accept、open、close、recv、send、readv和writev这10个IO函数，足以涵盖磁盘IO和网络采集的大部分情况。

本文设计的系统利用此方法采集了与网络通信和磁盘IO相关的函数，在程序发生网络通信过程时自动记录通信的时间、大小、内容及通信双方的信息，结合相应的网络流量分析方法可对高性能程序运行状态进行分析，通过对网络流量的分析可发现程序的异常行为，进而推断出程序代码或集群中存在故障。3.3节给出了一个基于网络流量分析发现程序固有缺陷的例子。在程序发生磁盘IO操作并且操作失败时，本系统能够自动记录下失败发生的时间、错误码及相关文件路径等必要信息，方便后续进行故障分析。由于在Linux系统中，网络通信和文件操作共用统一的API，为防止采集功能本身对Linux API的调用被误当成HPC通信过程的一部分，重复采集造成无限递归的死循环。本文设计了一种在打开文件时对文件描述符进行自动注册的机制来解决这个问题。以对文件的写入操作进行采集为例，基于符号重命名的IO函数采集算法如下所示。

子程序： __wrap_open // 对open函数进行插桩

输入： path, flag, mode

输出： fd

fd = __real_open(path, flag, mode) // 调用真实open函数

dir = dirname(path)

if dir != logDir then // 如果文件路径不是采集路

径, 则将此文件描述符加入监控集合

MonitorSet = MonitorSet ∪ fd

子程序： __wrap_write // 对 write 函数进行插桩

输入： fd, buf, size

输出： size

fd = __real_write(fd, buf, size) //调用真实 write 函数

if fd ∈MonitorSet ∧ ret < 0 then

Record // 写入出错, 记录此信息

2.2　基于运行时库的程序状态采集方法

构建时插桩的方法可以解决大部分问题, 但是, 其具有一个较大的缺点就是只能在通过源码进行构建程序时进行插桩。然而, 程序的运行依赖运行时库环境, 如 Glibc[23]、Musl 等 Linux 运行时库。这些第三方的库为方便用户使用, 将许多函数进行了多种不同的封装, 如输出函数就有 printf、fprintf、puts、fputs 等不同的接口, 这些接口将对 read 的系统调用封装到预编译的库中, 用户程序在构建时仅需简单地对库进行链接即可, 无需从库的源码编译。也正因如此, 由于用户程序构建时标准库已被预先编译成二进制的动态或静态链接库格式, 而不是与用于程序一起通过源代码编译, 无法使用 2.1 节所述基于构建时符号重命名技术对库函数进行自动插桩和采集。

为能够统一追踪所有的输出信息, 本文设计的系统采用对 Glibc 进行 Hook 的方式实现该功能[23]。在 Glibc 中, stdout 为指向 FILE 结构体的指针, 而 FILE 结构为_IO_FILE_plus 的第 1 个成员变量, _IO_FILE_plus 除了 FILE 外还拥有一个 vtable 成员变量, 其中包含 IO 相关各个底层函数指针, 因此, 仅需修改此指针指向的函数, 即可实现对程序中全部的输出函数进行插桩, 从而统一追踪所有的输出信息。为实现这个目标, 本文设计了一种能够实现在程序启动时将 vtable->__write 的地址替换成带有数据采集功能的函数 wrapped_write, 经此操作后程序中所有对文件或控制台的写入操作都会经过此自定义的带有数据采集功能的函数。从而实现在不修改代码的前提下无缝记录程序的所有日志输出。基于运行时库的程序状态采集方法如图 4 所示。

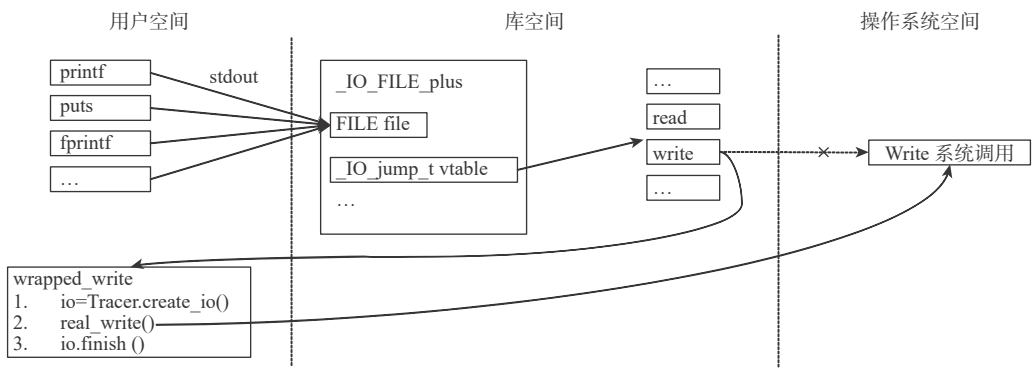


图 4　基于运行时库的程序状态采集方法

Fig. 4　Method for collecting program state based on runtime library

由于在 Linux 操作系统中, 库函数所在内存区域默认只有访问权限, 用户程序并不能修改库空间的虚函数表。因此, 为能够实现动态对 Glibc 进行 Hook 的功能, 还需通过 mprotect 系统调用修改虚函数表区域权限后才能替换虚函数表内容。当用户程序启动时, 首先, 通过 stdout 等结构的地址计算出库空间虚函数表所在的内存页面, 再通过 mprotect 功能为此页面增加可写入权限, 修改虚函数表中的内容为预先准备好的 wrapperd_write 后, 再将此页面恢复只读权限, 防止误操作将其破坏, 如此便完成 Glibc Hook 的全过程。完整的 Glibc Hook 算法如下所示。

子程序： Hooked_write

输入： file, buf, size // 输出的文件和内容

输出： size // 真实输出的长度

If file == stdout ∨ file == stderr then

Record // 采集本次标准输出

size = real_write(f, buf, size)

子程序： hook_glibc //对 Glibc write函数进行 Hook

vtable = ((_IO_FILE_plus*)stdout)->vtable // 找到 Glibc 中的虚函数表 vtable

real_write = vtable->__write // 备份旧的 write 函数

set_writable(vtable) // 设置 vtable 页面为可写

vtable->__write = hooked_write // 将 vtable 中的 write 替换为 hooked_write

set_readonly(vtable) // 恢复 vtable 页面只读

2.3　基于系统调用的程序状态采集方法

计算程序出现故障往往是在进行磁盘或网络等操作时出错, 现代操作系统中访问系统硬件功能均需要通过系统调用。本文通过 Ptrace[23] 功能实现了对系统调用的追踪。Ptrace 是 Linux 内核中提供的功能, 其能够使程序在特定信号或进入系统调用

时暂停，进行一定操作后再继续执行，可用于监控进程行为。使用 Ptrace 进行监控时，首先，使用 fork 调用创建子进程作为监控进程；随后，在子进程中调用 Ptrace_ATTACH 将自身作附加到父进程上，经此操作过后，父进程每当调用系统调用时或从系统调用返回时都会被 SIGSTOP 暂停，而子进程可以使用 waitpid 等待这一事件；最后，可以获得父进程寄存器及内存等相关信息，将这些信息采集妥当后可以通知父进程可以继续执行。使用 Ptrace 进行监控的过程如图 5 所示。

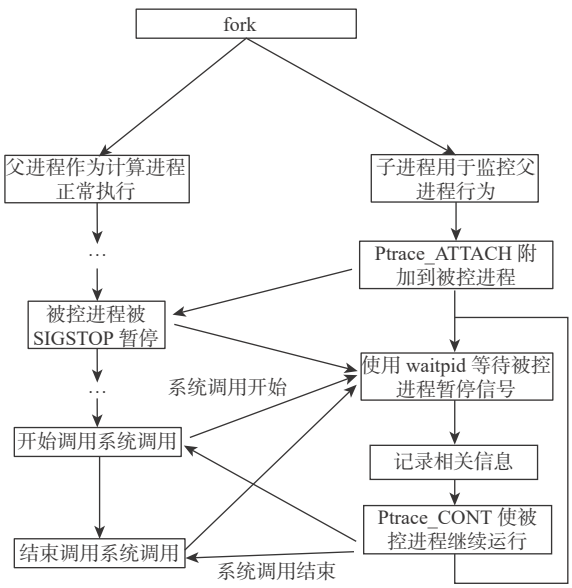


图 5 Ptrace 执行过程

Fig. 5 Process of Ptrace

2.4 程序运行时资源占用情况采集

在程序的故障分析过程中，除程序本身的状态数据具有巨大意义外，程序运行时所消耗的资源情况同样值得关注，程序运行每时每刻所消耗的资源情况可以极大程度反映出程序的计算特征和运行状况。

在 Linux 系统中，通过周期性地读取 /proc/<pid>/stat 文件即可获得该应用程序的统计信息，此文件的格式为空格分隔开的多个字段，每个字段分别代表不同的含义。其中，第 14～第 17 列分别为应用程序在用户态执行的时间（utime）、在内核态执行的时间（stime）、在用户态等待子进程所消耗的时间（cutime）及在内核态等待子进程所消耗的时间（cstime）。通过读取该文件中的以上 4 个字段即可获得该进程所消耗的 CPU 资源。

除此之外，/proc/<pid>/statm 文件中储存了当前进程的内存统计信息，此文件格式与 /proc/<pid>/stat 相同，各字段含义分别为：程序占用内存的总尺寸、内存驻留集尺寸、驻留集的共享页面数、代码段、库尺寸、数据堆栈尺寸及脏页数目。本文设计的程序运行资源采集模块主要关注计算进程申请的全部内存大小，即 statm 文件中的第 6 个字段数据堆栈尺寸。本文对 CPU 和内存占用情况的实现算法如下所示。

子程序： metrics_collect // 采集进程耗费的系统资源

$u_1 = 0$

while True do

sleep t // 采样周期为 t 秒

u_2 = read /proc/<pid>/stat // 获得 CPU 时间

cpu = $(u_2-u_1)/(T\times t\times M\times N)$ // 计算 CPU 占用率

$u_1 = u_2$

memory = read /proc/<pid>statm // 获得内存占用

Record(cpu, memory) // 记录 CPU 和内存占用率

3 基于“抽道集”排序程序验证实际效果

在地震资料处理中，为能够对海量地震数据进行处理，并方便叠加和计算速度谱等操作，先需要将地震道集按特定规则进行抽取排序，称为“抽道集”。对“抽道集”进行排序的操作在地震资料处理中处于十分基础的地位。本文使用一个“抽道集”排序程序对本文方法进行验证。该程序基于自研的 GPP 框架，使用 Map Reduce 技术实现了可自动容错的分布式并行排序，此程序已经在实际生产项目中使用多年，基本能够满足生产的实际需要，但是依然存在一定概率的计算失败和运算效率较低的问题，本文方法对此程序进行了性能和故障分析验证，指出了性能可以继续优化的方面并解决了计算失败的故障问题。本次实验所使用的数据为中国石油集团东方地球物理公司在中东地某区块高密度采集数据，总量共 100 TB。使用的计算集群每节点 CPU 型号为 Intel(R) Xeon(R) CPU E5-2667 v3，内存容量为 512 GB，节点间采用 10 Gbit/s 以太网连接。共 128 节点，每节点上启动 4 个计算进程。

3.1 应用程序 CPU 资源波动采集验证

利用 2.3 节所述高性能计算程序运行时耗费资源度量采集方法对“抽道集”排序运行时各节点 CPU 资源占用波动情况进行采集分析，得到结果如图 6 所示，图中，每条曲线为一个计算进程。本文测试使用的“抽道集”排序程序基于 Map-Reduce 技术开发，图中，可以清晰地看到计算过程整体分为 2 个阶段，在 Map 阶段时，程序所进行的工作主要以读入数据为主，因此，各进程 CPU 占用率均在较低范围波动；当完成 Map 操作，进入 Reduce 阶段后，进程的 CPU 占用率陡然增加，说明程序进入了

计算密集的 Shuffle 排序过程, 等待排序完成后, 仅需向磁盘写入数据, 此时进程的 CPU 占用率再次回归较低水平。由图 6 可知, 存在一个计算进程在其他进程完成之后 CPU 占用率依然保持高水平并持续了较长时间, 当此进程完成计算退出后其他所有进程均随之退出, 此现象可以说明任务运行时此进程拖慢了整个计算任务。该结果提示此计算程序可能存在各计算进程压力分布不均的情况, 或此进程或其所在的节点上可能存在一定故障, 导致计算速度变慢。

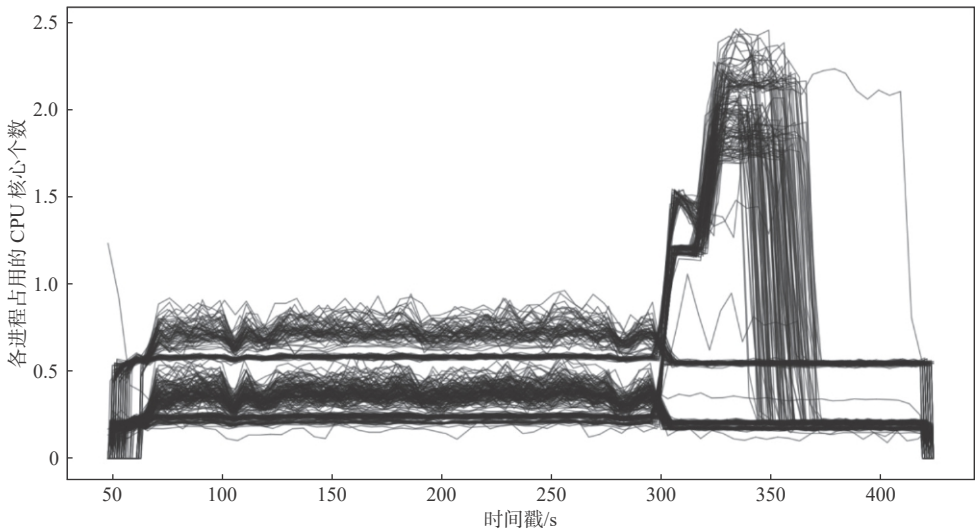


图 6　“抽道集”排序程序各节点 CPU 占用监控

Fig. 6　CPU usage monitor of “trace gather” sorting program

3.2　应用程序数据传输速度采集验证

在 Map-Reduce 程序中, 各 Map 节点计算完成后会将此计算结果发送到相应的 Reduce 节点上, 因此, 节点间的数据传输的速度也是制约 Map-Reduce 程序运行效率的重要因素。

本文使用 2.1 节提到的“基于构建时的程序状态采集方法”和“基于运行时库的程序状态采集方法”2 种方法对“抽道集”排序程序运行时 Reduce 进程接收数据的状况进行了采集。采集结果如图 7 所示, 图中, 横坐标为 Reduce 进程号, 纵坐标为接收到的数据量, 单位为 GB。

在图 7 中可以直观看出, 各 Reduce 节点接收数据量差距较大, 最少的节点接收到了约 90 GB 的数据, 而接收数据量最多的节点接收到了近 900 GB 的数据, 相差 10 倍。不同节点间接收数据量的差异导致部分节点压力过大而部分节点空闲, 计算资源没有很好地利用起来, 使计算效率降低。因此, 可以尝试对 Map 节点的 Partition 操作进行优化, 使数据分布尽可能均匀, 从而平均各节点的计算压力, 从而实现提升整体运行效率的目的。

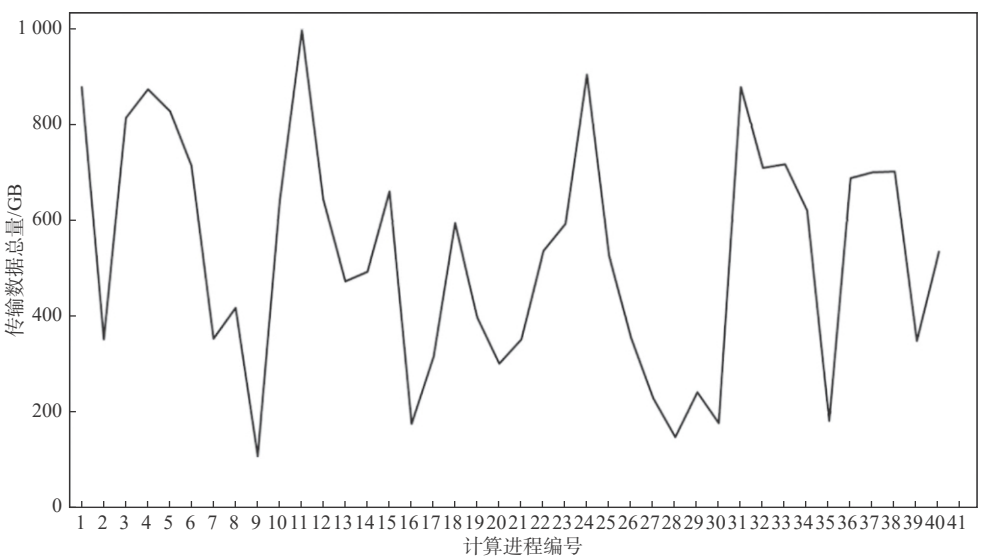


图 7　Reduce 节点接收数据量曲线

Fig. 7　Received data size curve of reduce nodes

此外，本文还对某次 Map-Reduce 计算任务中每个 Reduce 节点接收数据的速率进行了采集，观察发现各个节点接收数据情况都具有相同的特征，其中某个节点的结果如图 8 所示，图中依然可以看到较为明显的 Map 阶段和 Reduce 阶段，但数据的接收过程速度较慢，最快处也仅有 150 MB/s，认为此处的性能瓶颈是由机械硬盘的写入速度导致的。

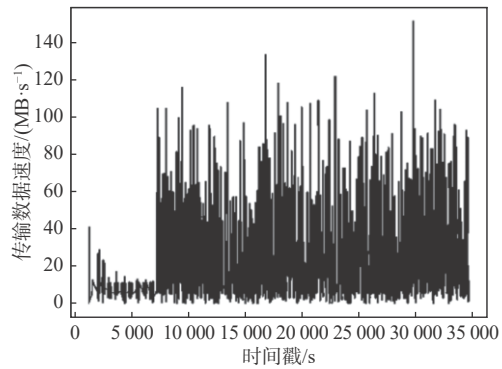


图 8 单一 Reduce 节点接收数据速率曲线

Fig. 8 Received data speed curve of signal reducer

3.3 应用程序网络流量采集和故障分析

使用本文方法对“抽道集”排序程序进行网络流量和传输情况进行采集，经分析后，发现了此程序在运行时节点间通信过程存在异常行为：在节点间进行通信时，有个别节点错误地接收到了本应发往其他节点的数据，如表 3 所示，Rank 6 试图将某数据发往 Rank 9，因此，构建了内容为 c81e728d9d4c2f636f067f89cc14862c，目标地址为 1132396544 的数据包。但通过采集信息得知，尽管此数据包的目标地址为 Rank 9 的本机地址，但却错误地被 Rank 8 接收。

此信息提示程序中通信模块可能存在故障，结合源代码，发现故障原因如图 9 所示。当节点进行通信时，若连接对方节点失败，此时通信节点错误地保留了未打开的文件描述符。后续再尝试连接

表 3 程序通信故障

Table 3 Program communication fault

数据内容	主机名	本机地址	本机端口	计算进程号（rank）	目标地址	目标端口	时间戳/s	数据量/Byte	类型
6a204bd89f3c8348afd5c77c717a097a	hp6-006	1132396544	40593	9	2181118058	47201	1668067183	4	traffic
c81e728d9d4c2f636f067f89cc14862c	hp6-005	2181118053	33601	6	1132396544	52685	1668067323	4	traffic
c81e728d9d4c2f636f067f89cc14862c	hp6-005	0	40593	8	2181118053	33601	1668067323	4	traffic-receive

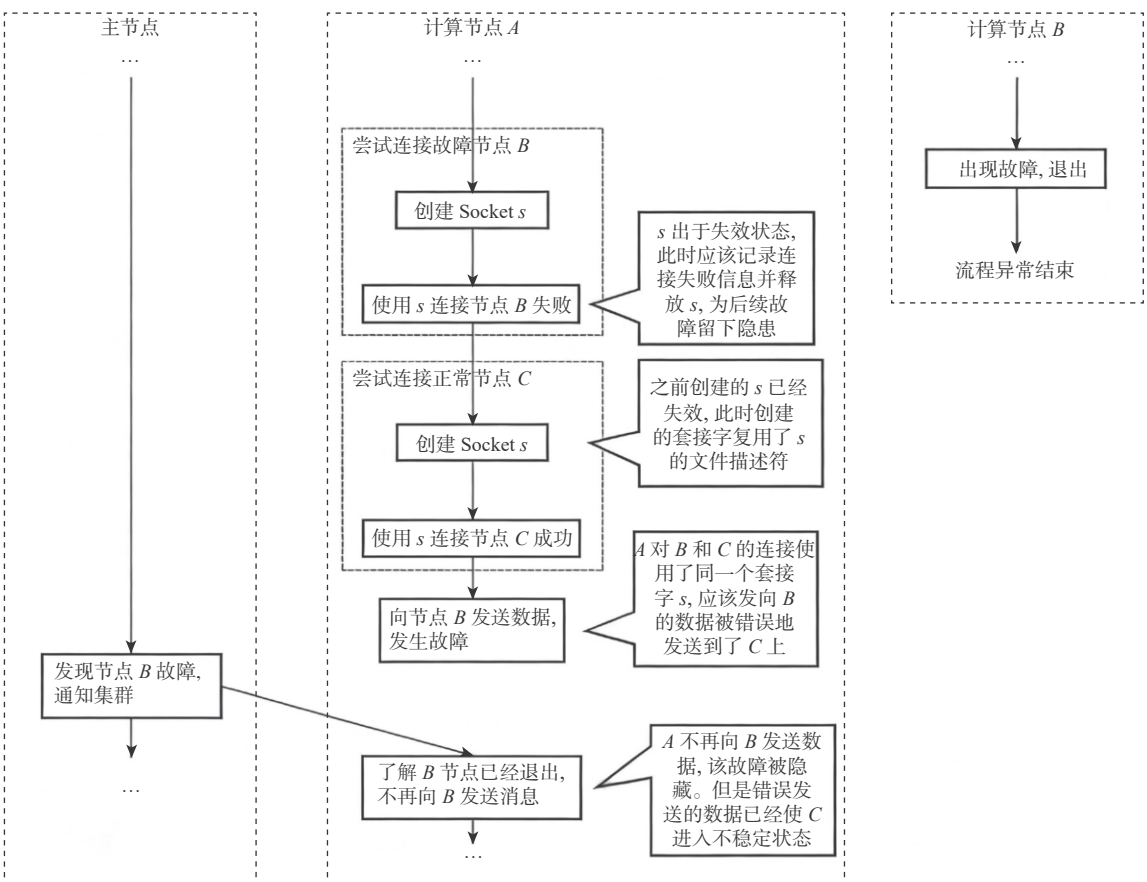


图 9 程序故障原理

Fig. 9 Program fault schematic

其他节点时, 此文件描述符有可能被复用。若再尝试向出错节点发送信息, 则该信息会被发送到后连接的正常节点上, 导致节点状态异常, 程序崩溃退出。

4　结　论

1) 本文针对在石油勘探流域生产环境中高性能计算集群和程序可能出现的故障及其分析方法进行了研究。先从移动互联网领域流行的 OTF 规范出发, 结合石油勘探领域高性能计算的实际情况, 将 OTF 规范进行了适当的裁剪和扩展, 设计了一种用于在分布式程序中进行状态追踪的接口, 通过少量易于解释的 API 将集群和高性能计算程序状态采集规范化, 有利于后续进行故障分析。

2) 本文对设计的接口进行了高效实现, 能够从链接、运行时库和系统调用 3 个层面对应用程序进行低侵入式监控追踪。

3) 将本文方法应用在“抽道集”排序的实际生产程序上, 对此程序的性能热点和故障进行了采集分析, 指出了该程序的计算特征和优化方向, 并解决了该程序的实际故障, 可以说明本文方法具有一定的有效性。

参考文献（References）

[1] 王海龙, 高建虎, 李海亮, 等. “两宽一高”地震勘探技术在油气精细勘探中的应用[J]. 石油地球物理勘探, 2022, 57(S1): 137-144.
WANG H L, GAO J H, LI H L, et al. Application of “two widths and one height” seismic exploration technology in fine oil and gas exploration[J]. Oil Geophysical Prospecting, 2022, 57(S1): 137-144 (in Chinese).

[2] GABRIEL E, FAGG G E, BOSILCA G, et al. Open MPI: goals, concept, and design of a next generation MPI implementation[M]// Recent Advances in Parallel Virtual Machine and Message Passing Interface. Berlin: Springer, 2004: 97-104.

[3] BALASUBRAMANIAN V, BANERJEE P. A fault tolerant massively parallel processing architecture[J]. Journal of Parallel and Distributed Computing, 1987, 4(4): 363-383.

[4] HWANG K, GHOSH J. Hypernet: a communication-efficient architecture for constructing massively parallel computers[J]. IEEE Transactions on Computers, 1987, C-36(12): 1450-1466.

[5] DITTRICH J, QUIANÉ-RUIZ J A. Efficient big data processing in hadoop MapReduce[J]. Proceedings of the VLDB Endowment, 2012, 5(12): 2014-2015.

[6] DEAN J, GHEMAWAT S. MapReduce: simplified data processing on large clusters[J]. Communications of the ACM, 2008, 51(1): 107-113.

[7] SNIR M, WISNIEWSKI R W, ABRAHAM J A, et al. Addressing failures in exascale computing[J]. The International Journal of High Performance Computing Applications, 2014, 28(2): 129-173.

[8] 李超, 赵长海, 晏海华, 等. 一种在复杂环境中支持容错的高性能规约框架[J]. 北京航空航天大学学报, 2018, 44(10): 2115-2124.
LI C, ZHAO C H, YAN H H, et al. A fault tolerant high-performance reduction framework in complex environment[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(10): 2115-2124(in Chinese).

[9] 赵长海, 晏海华, 王宏琳, 等. 面向地震数据处理的并行与分布式编程框架[J]. 石油地球物理勘探, 2010, 45(1): 146-155.
ZHAO C H, YAN H H, WANG H L, et al. Seismic data processing oriented parallel and distributed programming framework[J]. Oil Geophysical Prospecting, 2010, 45(1): 146-155(in Chinese).

[10] SHENDE S S, MALONY A D. The tau parallel performance system[J]. The International Journal of High Performance Computing Applications, 2006, 20(2): 287-311.

[11] BURTSCHER M, KIM B D, DIAMOND J, et al. PerfExpert: an easy-to-use performance diagnosis tool for HPC applications[C]// Proceedings of the ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis. Piscataway: IEEE Press, 2010: 1-11.

[12] WOLF F, MALONY A D, SHENDE S, et al. Trace-based parallel performance overhead compensation[M]//High Performance Computing and Communications. Berlin: Springer, 2005: 617-628.

[13] MALONY A D, NAGEL W E. The open trace format (OTF) and open tracing for HPC[C]//Proceedings of the ACM/IEEE conference on Supercomputing. New York: Association for Computing Machinery, 2006: 24-es.

[14] 李超. 分布式环境中并行地震应用编程模型与性能优化研究[D]. 北京: 北京航空航天大学, 2019: 41-69.
LI C. Research of programming model and performance opti-mization for parallel seismic applications in distributed environments[D]. Beijing: Beihang University. 2019: 41-69(in Chinese).

[15] NADAREISHVILI I, MITRA R, MCLARTY M, et al. Microservice architecture: aligning principles, practices, and culture [M]. Farnham: O’Reilly Media, Inc. , 2016: 101-102.

[16] 林闯, 苏文博, 孟坤, 等. 云计算安全: 架构、机制与模型评价[J]. 计算机学报, 2013, 36(9): 1765-1784.
LIN C, SU W B, MENG K, et al. Cloud computing security: architecture, mechanism and modeling[J]. Chinese Journal of Computers, 2013, 36(9): 1765-1784(in Chinese).

[17] 陈建海, 陈淼, 浦云明. 基于微服务架构 B/S 系统的性能分析[J]. 计算机系统应用, 2020, 29(2): 233-237.
CHEN J H, CHEN M, PU Y M. B/S system performance analysis based on microservices architecture[J]. Computer Systems & Applications, 2020, 29(2): 233-237(in Chinese).

[18] KNÜPFER A, BRUNST H, BRENDEL R. Introducing the open trace format (OTF)[C]//Proceedings of the International Conference on Computational Science. Berlin: Springer, 2006: 526-533.

[19] JANES A, LI X, LENARDUZZI V. Open tracing tools: overview and critical comparison[EB/OL]. (2023-06-23)[2023-11-20] . https://arxiv.org/abs/2207.06875.

[20] KIM C, AGRAWALA A K. Analysis of the fork-join queue[J]. IEEE Transactions on Computers, 1989, 38(2): 250-255.

[21] LU H J. ELF: from the programmer’s perspective[EB/OL]. (1995-01-19)[2023-11-20].https://epdf.pub/elf-from-the-programmers-perspective.html.

[22] JUNG R, GUNDLACH S, HASSELBRING W. Instrumenting C

and fortran software with kieker[C]//Proceedings of the Symposium on Software Performance. Leipzig: CEUR, 2021.

[23] KRENTEL M W. Libmonitor: a tool for first-party monitoring[J]. Parallel Computing, 2013, 39(3): 114-119.

HPC fault and performance analysis in field of petroleum geophysical exploration

WANG Mingjian[1], ZHAO Changhai[1,*], LI Chao[2], WEN Jiamin[1], SHANG Minqiang[1], HOU Hongjun[1]

(1. Research & Development Center, Bureau of Geophysical Prospecting Inc, China National Petroleum Corporation, Zhuozhou 072751, China;
2. School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: Terabytes or even petabytes of data are now available due to the ongoing advancements in seismic data gathering technologies in the field of petroleum geophysical prospecting. As the volume of data, running time, and the number of nodes in high-performance computer (HPC) clusters increase, the probability of cluster issues and the difficulty of maintenance also rise. When a cluster or node fails, it often requires re-running the computing program, resulting in significant waste of resources. Therefore, based on the concepts of open tracing format (OTF) and distributed link tracing, this paper suggests a low invasive fault analysis method for high-performance computing clusters and programs in production environments in order to address the issues of low observability, difficult to analyze fault, and performance of HPC programs in computing clusters. The technique has the advantage of low invasion in addition to being able to monitor HPC programs in a production setting. The suggested method can be applied in production environments programs nearly without changing code. After that, the study applies this technique to the distributed "Trace Gather" sort program in a production setting for data gathering and analysis. It then confirms the efficacy of the approach outlined in this paper and identifies any hidden software flaws or performance issues in the program.

Keywords: distributed computing; huge cluster; fault analysis; link tracing; production environment

Received: 2023-11-24; **Accepted**: 2024-02-05; **Published Online**: 2024-03-29 07:22

URL: link.cnki.net/urlid/11.2625.V.20240328.1001.001

Foundation items: Science and technology Research and Development Project of CNPC (2021ZG03-05); Science and technology Research and Development Project of BGP (01-01-2023)

* **Corresponding author**. E-mail: zhao.ch@cnpc.com.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0729

基于叠加流场面元法的民用飞机跨声速气动伺服弹性分析

吴善强，陈琦，孙亚军，尼早，陈文*
(上海飞机设计研究院，上海 201210)

摘　　要：与亚声速相比，民用飞机进入跨声速飞行时，其气动伺服弹性特性出现明显的频率偏移和稳定裕度下降等问题。对于研发阶段的飞机而言，研发成本昂贵、试飞风险高、设计更改周期长，故在初步设计阶段需充分考虑跨声速区域的气动伺服弹性设计。采用偶极子格网法计算的非定常气动力无法考虑跨声速激波和边界层的影响。针对跨声速非定常气动力计算，将非定常小扰动速度势分为定常项和非定常项，通过计算流体力学求解跨声速定常项，使用非定常流场面元法求解时间线化的跨声速小扰动速度势方程来计算非定常项。将得到的跨声速非定常气动力影响系数代入气动伺服弹性机体频响函数中，并与试飞结果、偶极子气动力模型计算结果进行对比，结果表明：基于叠加流场面元法的跨声速气动伺服弹性分析在频率和稳定裕度等方面与试飞结果更吻合。探索跨声速区域传递函数频率和稳定裕度变化规律，为民用飞机跨声速气动伺服弹性设计提供方法。

关 键 词：民用飞机；气动伺服弹性力学；计算流体力学；面元法；跨声速
中图分类号：V221+.3
文献标志码：A　　**文章编号**：1001-5965(2026)01-0167-13

随着主动控制和电传飞控技术的发展，宽频带伺服控制技术自20世纪70年代起开始应用在民用飞机设计上，其中，气动弹性和控制系统的耦合问题一直被设计人员所关注。飞行器结构弹性振动信号随刚体运动信号一同传入传感器，将信号传递到飞控系统用于驱动舵面运动。舵面在低频偏转的基础上叠加高频偏转，而高频非定常舵面偏转激励引起结构进一步的弹性振动。这种飞机机体和控制系统耦合引起的气动伺服弹性(aeroservoelasticity, ASE)问题对民用飞机的安全与性能起着举足轻重的作用。

以民用飞机为对象的气动伺服弹性力学主要致力于研究飞机结构、非定常空气动力和电传飞控系统之间相互耦合产生的稳定性和响应问题。随着现代飞机结构设计越来越轻柔，飞行速度越来越快，飞行控制系统频带变宽，飞机结构的几何非线性、材料非线性、气动非线性和刚弹耦合等问题的出现加剧了ASE分析与设计的复杂性，ASE研究早期阶段主要在给定飞机设计参数的情况下，通过计算飞机在特定飞行状态下的闭环系统的气动弹性稳定性和响应。随着对气动弹性研究的深入和主动控制技术的发展，通过设计控制律来消除不利耦合以实现减重和增稳等性能目标。这也是当前飞机设计先进理念之一。

在亚声速区，偶极子格网法通常被用于计算飞行器的ASE稳定性，并得到较好的验证[1]。现代民

收稿日期：2024-10-09；录用日期：2025-01-23；网络出版时间：2025-03-14 13:20
网络出版地址：link.cnki.net/urlid/11.2625.V.20250313.1707.002
基金项目：中国商飞公司科技创新专项 (Y23GS29)
*通信作者. E-mail：chenwen1@comac.cc

引用格式：吴善强，陈琦，孙亚军，等. 基于叠加流场面元法的民用飞机跨声速气动伺服弹性分析 [J]. 北京航空航天大学学报，2026，52（1）：167-179. WU S Q，CHEN Q，SUN Y J，et al. Transonic aeroservoelastic analysis of civil aircraft based on overset field-panel method [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：167-179（in Chinese）.

用飞机普遍在跨声速巡航，其中，跨声速飞行中激波的出现使得传统的线性气动力模型不再适用于ASE分析。而跨声速非定常气动力对民用飞机ASE特性影响较大，跨声速效应引起的频率偏移和稳定裕度下降都对ASE设计产生不利影响。跨声速非定常气动力是非线性的，直接将线性化的气动力模型用于ASE建模是不合理的。若要考虑跨声速激波振荡等非线性气动力效应，需要采用计算流体力学（computational fluid dynamics, CFD）进行计算，对于此类非线性气动弹性问题，则需要采用计算流体力学和计算结构动力学相结合的流固耦合计算方法。但此类方法计算效率低，不利于气动伺服弹性控制律设计和闭环动力学分析[2-3]。从控制律设计角度出发，气动非线性给控制律设计带来诸多不确定性因素，对控制系统的鲁棒性提出更高的要求[4]。基于小扰动理论的跨声速小扰动（transonic small disturbance, TSD）方程发展较早，能够较高效率应用于工程计算[5]。1988年，时间精确近似因式分解算法求解三维非定常TSD方程的跨声速气动弹性方法也被提出[6]。2000年和2002年，Kim等[7-8]通过求解TSD方程计算三维机翼的跨声速流动。2008年，Kim等[9]采用了TSD方程计算跨声速非定常气动力，通过将黏性方程与TSD方程耦合，改善了气动力的计算结果。随着CFD和计算机技术的发展，CFD和计算结构动力学（computational structure dyanmics,CSD）耦合技术也被应用于跨声速ASE设计中[10]。但计算流体力学和计算结构动力学直接耦合的求解导致模型阶数过高不利于控制律设计。对此，研究者提出气动力降阶的概念。基于降阶模型（reduced order model,ROM）的气动力模型往往可以取代CFD求解气动弹性问题[11-17]，在保证计算精度的同时也大大减少了计算时间，不过其研究大都处于理论阶段，笔者未见公开报道的工程应用。

本文将时间线化的跨声速小扰动（time-linearized transonic small disturbance, TLTSD）速度势方程分解为定常和非定常两部分，其中，定常部分由CFD计算，非定常部分通过叠加流场面元法求解小扰动速度势方程所得，将定常和非定常结合得到气动力影响系数（aerodynamic influence coefficient, AIC）矩阵。在处理空间流场体单元影响系数矩阵维度灾难时，通过块三对角矩阵近似处理方法降低矩阵维度，提高计算效率。将计算所得非定常气动力和偶极子非定常气动力计算结果对比。将所得气动力影响系数代入弹性飞机动力学方程计算气动伺服弹性频响函数。通过对试验飞机舵面施加扫频激励信号，测试得到机体-电传飞控系统组成的ASE闭环系统的频响函数，与分析结果做对比，验证基于叠加流场面元法的跨声速气动伺服弹性分析方法的准确性，总结了民用飞机机体传递函数在跨声速区频率和稳定裕度随马赫数的变化规律，解决了民用飞机设计阶段在跨声速区域机体-非定常气动力-控制律耦合形成的气动伺服弹性在频率偏移和稳定裕度下降等方面的问题。

1 时间线化的跨声速小扰动速度势方程

时间线化的跨声速小扰动速度势方程为[18]

$$(1-K\Phi_x)\Phi_{xx}+\Phi_{yy}+\Phi_{zz}-2(M_\infty^2/\beta^2)\Phi_{xt}-(M_\infty{}^2/\beta^2)\Phi_{tt}=0 \tag{1}$$

式中：$K=(\gamma+1)M_\infty^2/\beta^2$，$\beta=\sqrt{1-M_\infty^2}$，$M_\infty$为自由来流马赫数；$\Phi$为速度势；$\gamma$为比热容比。

将速度势Φ分解为定常解和非定常解：

$$\Phi=\phi_0+\phi\mathrm{e}^{\mathrm{i}kt} \tag{2}$$

式中：ϕ_0为定常速度势；ϕ为非定常速度势；k为减缩频率。

假设非定常速度势ϕ远小于定常速度势ϕ_0，将(2)式代入式(1)，得到定常跨声速小扰动速度势方程及时间线性化的非定常跨声速小扰动速度势方程：

$$\phi_{0xx}+\phi_{0yy}+\phi_{0zz}=K\frac{\partial}{\partial x}\left(\frac{\phi_{0x}^2}{2}\right) \tag{3}$$

$$\phi_{xx}+\phi_{yy}+\phi_{zz}-2\frac{\mathrm{i}kM_\infty^2}{\beta^2}\phi_x+\frac{k^2M_\infty^2}{\beta^2}\phi=\frac{\partial}{\partial x}(\delta_v) \tag{4}$$

式中：$\delta_v=K\phi_{0x}\phi_x$。

在跨声速情况下存在着混合流场，流场中各点的局部马赫数可以是亚声速、声速和超声速。对应的$1-M_\infty^2$可以小于、等于和大于0。这时式(1)和式(4)有不同的性质，当$1-M_\infty^2<0$时，流场局部亚声速，式(1)和式(4)是椭圆型方程；当$1-M_\infty^2=0$时，流场局部声速，式(1)和式(4)是抛物型方程，这时β^2项需乘到方程两边；当$1-M_\infty^2>0$时，流场局部声速，式(1)和式(4)是双曲型方程。

2 面元法求解非定常跨声速小扰动速度势方程

面元法在计算跨声速非定常气动力方面有较多的应用。1985年，Voss[18]通过速度势方程提出了基于气动力影响系数的非定常面元方法。之后，

Lu 等[19] 使用跨声速偶极子方法来求解跨声速速度势方程。式 (2) 中 ϕ_0 为定常扰动速度势, 被定义为定常背景流, 其考虑了跨声速流动中的非线性激波效应, 本文通过定常 CFD 求解器来求解。与此同时, 式 (4) 的解和结构振动幅值呈线性变化关系, 只是在定常背景流中加入了非线性激波效应。

假设 $\partial(\delta_v)/\partial x$ 在式 (4) 中是一个体积源。时间线化跨声速速度势方程在流场任意一点 (x_0,y_0,z_0) 的解为

$$\phi(x_0,y_0,z_0)=\phi_{\mathrm{s}}+\phi_v+\phi_{\mathrm{shock}} \tag{5}$$

式中: ϕ_{s} 为物面速度势; ϕ_v 为体积源速度势, ϕ_{shock} 为激波速度势, ϕ_{s} 表达式为

$$\phi_{\mathrm{s}}=\frac{1}{4\pi}\int_s\left(\frac{\Delta C_p}{2}\overline{K}-\frac{\partial\phi}{\partial n}G\right)\mathrm{d}s \tag{6}$$

式中: ΔC_p 和 $\partial\phi/\partial n$ 分别为升力面的定常压力差和法向速度, 需要通过 CFD 计算的定常压力分布和法向速度线性插值得到。$\overline{K}$ 为加速度势核函数, 其具体表达式见文献 [20]。G 为非定常核函数。

$$G=\frac{1}{R}\mathrm{e}^{-\frac{\mathrm{i}kM_\infty^2}{\beta^2}\xi}\mathrm{e}^{-\frac{\mathrm{i}kM_\infty}{\beta^2}R} \tag{7}$$

$$R=\sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2} \tag{8}$$

$$\phi_v=-\frac{1}{4\pi}\int_V\frac{\partial\delta_v}{\partial x}G\mathrm{d}V \tag{9}$$

$$\phi_{\mathrm{shock}}=\frac{1}{4\pi}\int_{\mathrm{shock}}\Delta\delta_vG\mathrm{d}s \tag{10}$$

式中: R 为空间距离; $\Delta\delta_v$ 为体积源强度穿越激波的跳跃。

ϕ_v 表达式为

$$\phi_v=-\lim_{\varepsilon\to0}\frac{1}{4\pi}\int_{yz}\delta_v\,G|_{x_{\mathrm{s}}-\varepsilon}^{x_{\mathrm{s}}+\varepsilon}\mathrm{d}A+\frac{1}{4\pi}\int_v\delta_vG_x\mathrm{d}V \tag{11}$$

式中: x_{s} 为激波位置; ε 为无穷小的激波厚度。将式 (10) 和 (11) 相加可得

$$\phi_{\mathrm{shock}}+\phi_v=\frac{1}{4\pi}\int_v\delta_vG_x\mathrm{d}V \tag{12}$$

在物面某节点, 法向扰动速度为[21]

$$\boldsymbol{n}\cdot\nabla\phi=\boldsymbol{W}=\boldsymbol{A}\begin{bmatrix}\Delta C_p\\ \dfrac{\partial\phi}{\partial n}\end{bmatrix}+\boldsymbol{B}\boldsymbol{\sigma}_v \tag{13}$$

在空间单元某节点, 速度势为[22]

$$\phi=\boldsymbol{C}\begin{bmatrix}\Delta C_p\\ \dfrac{\partial\phi}{\partial n}\end{bmatrix}+\boldsymbol{D}\boldsymbol{\sigma}_v \tag{14}$$

式中: $\boldsymbol{W}$ 为下洗速度矩阵; $\boldsymbol{A}$ 和 $\boldsymbol{C}$ 分别为物面源点到物面和空间的影响系数矩阵; $\boldsymbol{B}$ 和 $\boldsymbol{D}$ 分别为空间源点到空间和物面的影响系数矩阵, 其元素表达式分别为

$$A_m=\frac{1}{8\pi}\iint_{\Delta S}\overline{K}(x_0,y_0)\mathrm{d}\xi\mathrm{d}\eta \tag{15}$$

$$B_m=-\beta\mathrm{e}^{\mathrm{i}\varepsilon x}\iiint_{\Delta V}(G_{\xi\zeta}-\mathrm{i}\varepsilon G_\zeta)\mathrm{d}\xi\mathrm{d}\eta\mathrm{d}\zeta \tag{16}$$

$$C_m=\frac{\mathrm{e}^{-\mathrm{i}\varepsilon x}}{8\pi\ell}\iint_{\Delta S}\overline{K}(x_0,y_0)\mathrm{d}\xi\mathrm{d}\eta \tag{17}$$

$$D_m=\iiint_{\Delta V}(G_\xi-\mathrm{i}\varepsilon G)\mathrm{d}\xi\mathrm{d}\eta\mathrm{d}\zeta \tag{18}$$

Voss[18] 提出式 (13) 和式 (14) 可以通过摩尔曼有限差分算子 $\boldsymbol{T}$ 建立联系:

$$\boldsymbol{\sigma}_v=\boldsymbol{T}\phi \tag{19}$$

式中: $\boldsymbol{T}$ 的具体表达式见文献 [23]; $\boldsymbol{\sigma}_v$ 按照摩尔曼格式给出:

$$\boldsymbol{\sigma}_v=(1-\mu_i)\boldsymbol{\sigma}_{vi}+\mu_{i-1}\boldsymbol{\sigma}_{vi-1}+\mu_i\boldsymbol{\phi}_{xi}-\mu_{i-1}\boldsymbol{\phi}_{xi-1} \tag{20}$$

式中: 当 $M_i<1$ 的时, μ_i=0; 当 $M_i\geqslant1$ 的时, $\mu_i=1$, M_i 为第 i 个体积元的当地马赫数; i 为自由来流方向的体单元编号; 摩尔曼格式的 $\boldsymbol{\sigma}_v$ 能够保证正确的数字激波跳跃, 可以从亚声速的中心差分切换到超声速的向后差分。

将式 (19) 代入式 (13) 和式 (14) 得到

$$\boldsymbol{W}=\overline{\boldsymbol{A}}\begin{bmatrix}\Delta C_p\\ \dfrac{\partial\phi}{\partial n}\end{bmatrix} \tag{21}$$

$$\overline{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{B}\boldsymbol{T}\boldsymbol{E}^{-1}\boldsymbol{C} \tag{22}$$

$$\boldsymbol{E}=\boldsymbol{I}-\boldsymbol{D}\boldsymbol{T} \tag{23}$$

式中: $\overline{\boldsymbol{A}}$ 的转置就是 AIC 矩阵, 其求解步骤如图 1 所示; 矩阵 $\boldsymbol{E}$ 为一个满秩的大型矩阵, 其维数由空间单元的数目决定。空间单元由多个子块组成, 矩阵 $\boldsymbol{E}$ 可以表示为

$$\boldsymbol{E}=\boldsymbol{E}_{\boldsymbol{B}}+\boldsymbol{E}_\varepsilon \tag{24}$$

式中: $\boldsymbol{E}_{\boldsymbol{B}}$ 为包括对角矩阵块和其相邻子矩阵块的块三对角矩阵; $\boldsymbol{E}_\varepsilon$ 矩阵包括块三对角矩阵的零元素和对角矩阵非相邻的元素。由于矩阵 $\boldsymbol{E}_{\boldsymbol{B}}$ 中的元素绝对值大于矩阵 $\boldsymbol{E}_\varepsilon$ 中的元素, $\boldsymbol{E}$ 是对角占优矩阵, 所以矩阵 $\boldsymbol{E}$ 的逆可以表示为

$$\boldsymbol{E}^{-1}\approx\boldsymbol{E}_{\boldsymbol{B}}^{-1}-\boldsymbol{E}_{\boldsymbol{B}}^{-1}\boldsymbol{E}_\varepsilon\boldsymbol{E}_{\boldsymbol{B}}^{-1} \tag{25}$$

式中: 矩阵 $\boldsymbol{E}_{\boldsymbol{B}}$ 是块三对角矩阵, 矩阵 $\boldsymbol{E}_{\boldsymbol{B}}$ 的逆 $\boldsymbol{E}_{\boldsymbol{B}}^{-1}$ 可以容易求解, 进一步可求得 $\boldsymbol{E}^{-1}$。

由式 (6) 可知, 求解物面扰动速度势方程需要

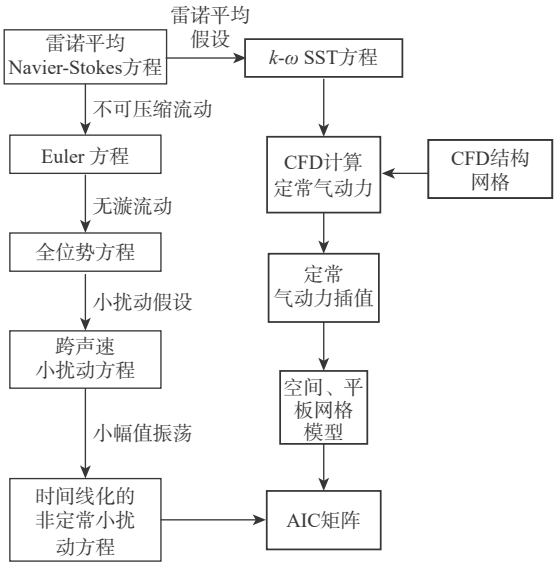


图 1 气动力影响系数求解流程图

Fig. 1 Flow chart for solving aerodynamic influence coefficient

建立物面网格，将机翼简化为平板升力面，机翼平板气动力网格如图 2 所示，机身部分建立包络体网格。由式 (12) 可知，求解空间速度势方程，需要将分布体积源的空间流场离散，建立能够考虑激波对流场影响的空间网格是其中的关键。故在典型民机气动网格建模中，为便于定常气动力插值而在机翼表面建立考虑翼型的物面网格，如图 3 所示。在机翼周围建空间网格，构建物面和空间的重叠网格模型以求解空间速度势，如图 4 所示。由于机身不是升力面，为提高计算效率，机身周围不需要建空间网格。

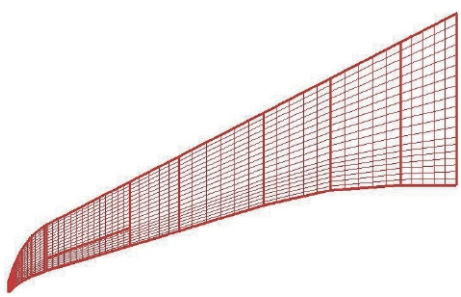


图 2 机翼平板气动力网格

Fig. 2 Wing-shaped slab aerodynamic mesh

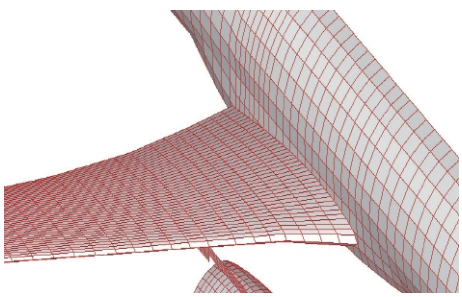


图 3 飞机物面网格

Fig. 3 Aircraft surface grid

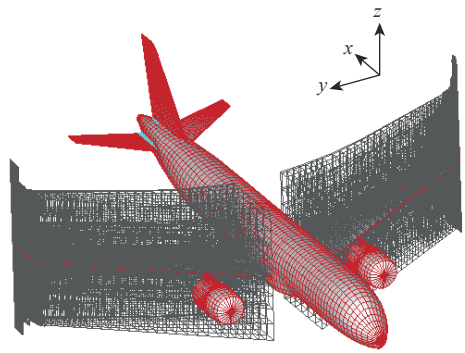


图 4 飞机物面和空间重叠网格模型

Fig. 4 Aircraft surface and space overlapping grid model

3 跨声速非定常气动力计算和对比

将总的扰动速度势分解为定常背景流速度势和非定常小扰动速度势，Liu 等[24] 证明在提供足够精度的定常流动解的情况下，非定常小扰动速度势也可以精确求解。本文采用 CFD 求解跨声速定常气动力，飞机的气动网格如图 5 所示，全流场采用结构网格，通过求解雷诺平均 Navier-Stokes 方程进行数值模拟。无黏通量采用二阶迎风 ROE 格式离散，黏性通量采用二阶中心差分离散，时间推进采用隐式 LU-SGS。湍流模型选择 SST k-ω 模型。为验证计算的可靠性，设计某大型民用客机 1∶18 的缩比模型在跨声速风洞（transonic wind tunnel, TWT）中进行风洞试验，风洞模型设计参数如表 1 所示。

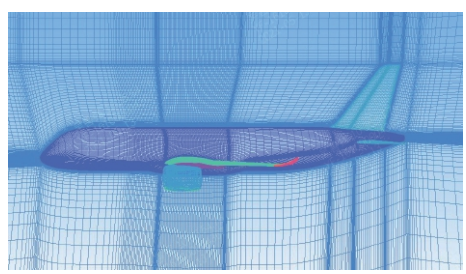


图 5 飞机 CFD 网格示意图

Fig. 5 Schematic diagram of aircraft CFD grid

表 1 飞机风洞模型参数

Table 1 Aircraft wind tunnel model parameters

机翼参考面积/m^2	机翼展长/m	展弦比	平均气动弦长/m
0.35	1.82	9.6	0.25

风洞试验在英国 ARA(Aircraft Research Association)2.74 m×2.44 m 跨声速风洞中进行，风洞湍流度小于 0.1%，风洞压力变化范围为 81.06～121.59 kPa，试验马赫数在 0.4～0.89，对应的雷诺数见表 2。

试验采用电子扫描阀系统测量压力分布，风洞

表 2　不同试验马赫数下试验雷诺数

Table 2　Test Reynolds number at different Mach numbers

试验马赫数	试验雷诺数
0.40	1.88×10^6
0.50	2.23×10^6
0.60	2.51×10^6
0.70	2.74×10^6
0.80	2.90×10^6
0.82	2.93×10^6
0.87	2.98×10^6
0.89	3.00×10^6

试验如图 6 所示。将在跨声速风洞中的试验结果与计算结果进行对比, 试验和计算的对象迎角和侧滑角均为 0。图 7 列出 Ma=0.8 时机翼不同站位剖面的压力系数对比, y/b 表示剖面站位到翼根的距离和机翼展长的比值, d 表示压力观测点到机翼前缘距离和翼形弦长的比值, 显示靠近机翼根部处, 由于机身和发动机短舱的干扰, 计算和试验结果在机翼前缘表现出差异, 越靠近后缘差异越小。在离

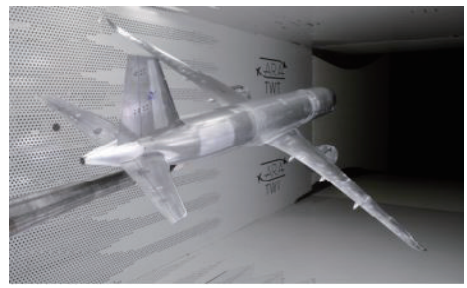


图 6　跨声速风洞试验

Fig. 6　Transonic wind tunnel test

稍部更近的剖面, 计算和试验结果更接近。通过对比, CFD 计算结果可以满足跨声速气动力计算精度要求, 将求解得到的流场速度和压力等数据通过线性插值作为求解定常速度势和非定常速度势的初始条件。

同时为比较偶极子格网法计算得到的非定常气动力与本文计算的跨声速非定常气动力的差异, 分别计算了 Ma=0.8, 且减缩频率 k 等于 0.2 和 0.6 时, 机翼反对称一弯模态和机身垂直一弯模态的机翼从根部到稍部沿展向 35% 和 63% 处的非定常压力系数 $\mathrm{Re}(\Delta C_p)$ 和 $\mathrm{Im}(\Delta C_p)$ 沿弦向 x/c 的分布如图 8 所示。

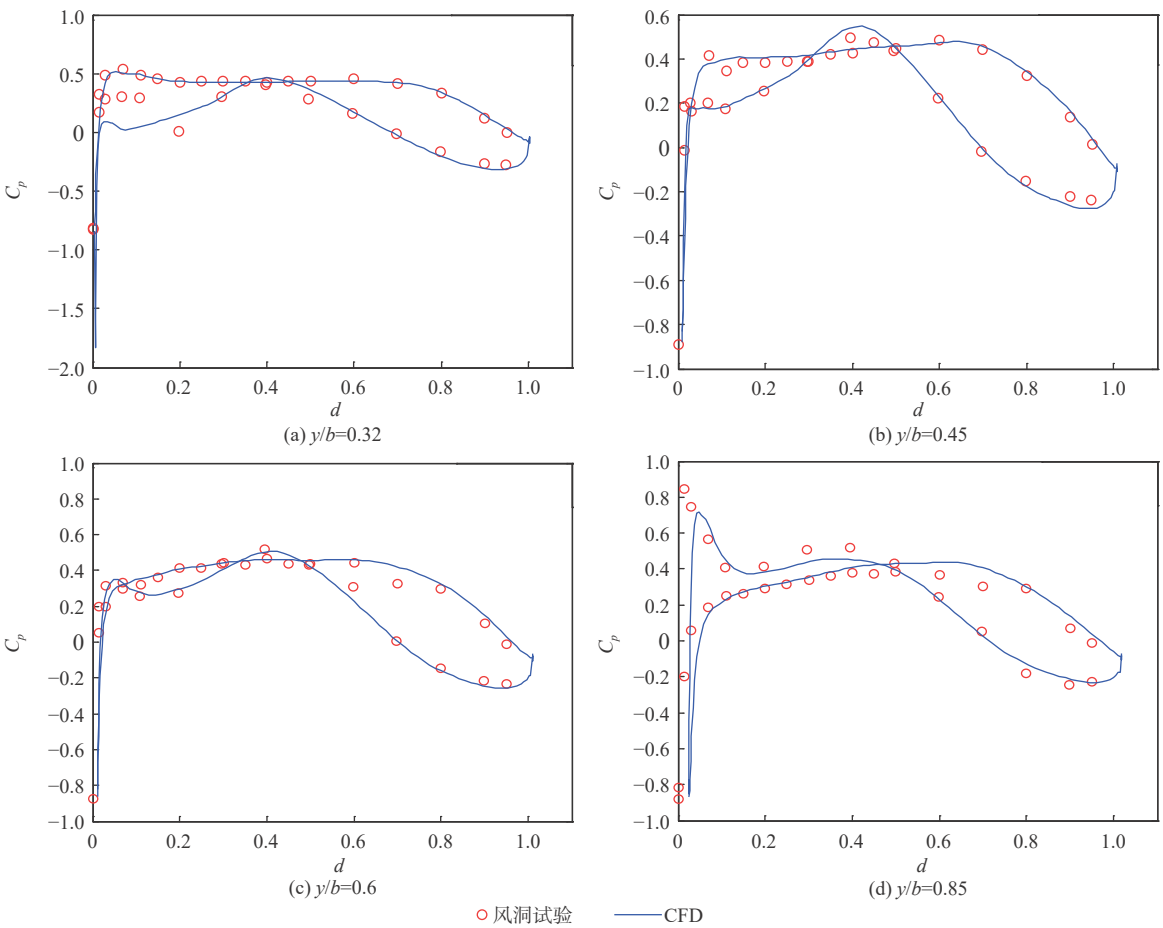
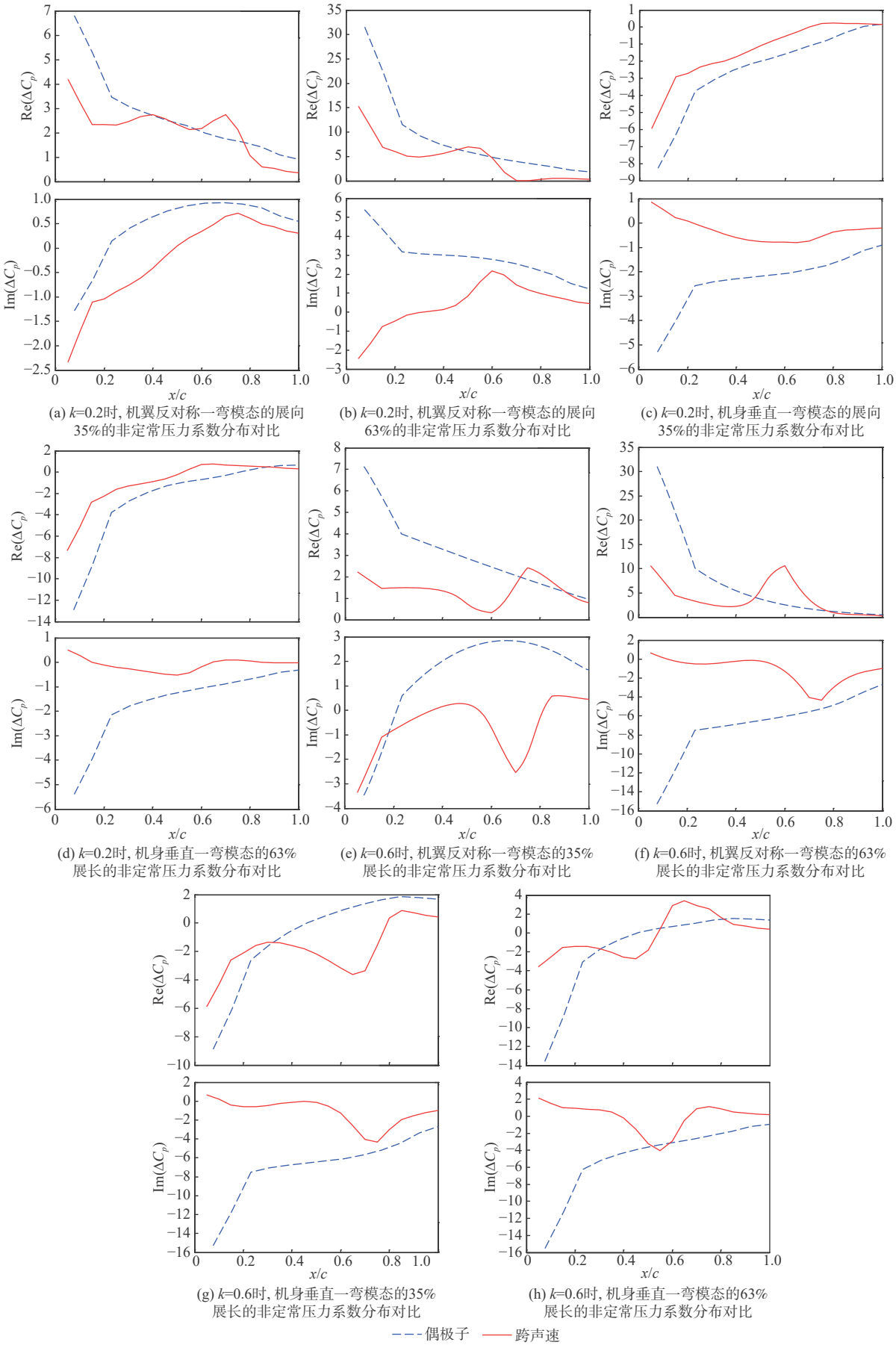


图 7　Ma=0.8 时数值模拟和风洞试验结果对比

Fig. 7　Comparison of numerical simulation and wind tunnel test results with Ma=0.8



(a) k=0.2时, 机翼反对称一弯模态的展向35%的非定常压力系数分布对比

(b) k=0.2时, 机翼反对称一弯模态的展向63%的非定常压力系数分布对比

(c) k=0.2时, 机身垂直一弯模态的展向35%的非定常压力系数分布对比

(d) k=0.2时, 机身垂直一弯模态的63%展长的非定常压力系数分布对比

(e) k=0.6时, 机翼反对称一弯模态的35%展长的非定常压力系数分布对比

(f) k=0.6时, 机翼反对称一弯模态的63%展长的非定常压力系数分布对比

(g) k=0.6时, 机身垂直一弯模态的35%展长的非定常压力系数分布对比

(h) k=0.6时, 机身垂直一弯模态的63%展长的非定常压力系数分布对比

图 8 k=0.2 和 k=0.6 时不同模态下不同站位的非定常压力系数对比

Fig. 8 Comparison of unsteady pressure coefficient of different Root Chord due to different mode with k=0.2 and k=0.6

4　气动伺服弹性机体传递函数计算

弹性飞机结构动力学方程[24]可以表示为

$$\tilde{\boldsymbol{M}}\ddot{\boldsymbol{q}}+\tilde{\boldsymbol{K}}\boldsymbol{q}+\frac{1}{2}\rho V^2\tilde{\boldsymbol{A}}\boldsymbol{q}=\boldsymbol{F} \tag{26}$$

式中: $\boldsymbol{q}=[q_1 \quad q_2 \quad \cdots \quad \beta]$是广义坐标列阵; $\tilde{\boldsymbol{M}}$、$\tilde{\boldsymbol{K}}$和$\tilde{\boldsymbol{A}}$分别为广义质量矩阵、广义刚度矩阵和广义气动力矩阵; V为来流速度; $\boldsymbol{F}$为附加广义外力矩阵, 包括因$\Delta\beta$引起的气动力及惯性力, 即

$$\boldsymbol{F}=-\frac{1}{2}\rho V^2\tilde{\boldsymbol{A}}_{\boldsymbol{q}\Delta\beta}\Delta\beta-\tilde{\boldsymbol{M}}_{\boldsymbol{q}\Delta\beta}\Delta\ddot{\beta} \tag{27}$$

$$\boldsymbol{q}=\boldsymbol{W}^{\boldsymbol{q}}_{\Delta\beta}\Delta\beta \tag{28}$$

$$\tilde{\boldsymbol{A}}_{\boldsymbol{q}\Delta\beta}=\boldsymbol{f}_{\boldsymbol{q}}^{\mathrm{T}}\tilde{\boldsymbol{A}}\boldsymbol{f}_{\Delta\beta} \tag{29}$$

$$\tilde{\boldsymbol{M}}_{\boldsymbol{q}\Delta\beta}=\boldsymbol{f}_{\boldsymbol{q}}^{\mathrm{T}}\tilde{\boldsymbol{M}}\boldsymbol{f}_{\Delta\beta} \tag{30}$$

式中: $\boldsymbol{f}_{\boldsymbol{q}}$为飞机的固有振动模态; $\boldsymbol{f}_{\Delta\beta}$为舵面的旋转模态; $\boldsymbol{W}^{\boldsymbol{q}}_{\Delta\beta}$为由$\Delta\beta$作为输入, $\boldsymbol{q}$作为输出的传递函数矩阵; $\Delta\beta$由飞机弹性振动和控制系统的传递关系决定。

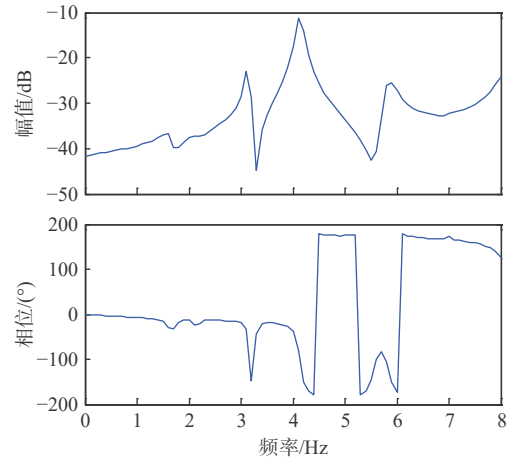
对弹性飞机动力学方程式 (24) 进行拉氏变换, 并结合式 (25) 和式 (26) 得到代数方程组:

$$\begin{aligned}&s^2\tilde{\boldsymbol{M}}\boldsymbol{Q}(s)+K\boldsymbol{Q}(s)+\frac{1}{2}\rho V^2A'\boldsymbol{Q}(s)=\\&\quad -s^2\tilde{\boldsymbol{M}}_{\boldsymbol{q}\Delta\beta}\boldsymbol{D}_{\Delta\beta}(\mathrm{s})-\frac{1}{2}\rho V^2\tilde{\boldsymbol{A}}_{\boldsymbol{q}\Delta\beta}\boldsymbol{D}_{\Delta\beta}(s)\end{aligned} \tag{31}$$

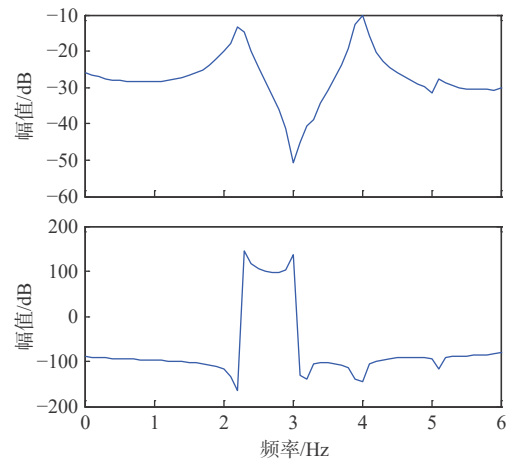
式中: A'为减缩频率k的函数, 且由 AIC 矩阵代入计算所得; $\boldsymbol{Q}(s)$为$\boldsymbol{q}$的拉氏变换; $\boldsymbol{D}_{\Delta\beta}(s)$为$\Delta\beta$的拉氏变换:

$$\frac{\boldsymbol{D}_{\Delta\beta}(\mathrm{j}\omega)}{\boldsymbol{Q}(\mathrm{j}\omega)}=\frac{\tilde{\boldsymbol{M}}_{\boldsymbol{q}\Delta\beta}\omega^2+\tilde{\boldsymbol{K}}+\frac{1}{2}\rho V^2\tilde{\boldsymbol{A}}}{-\tilde{\boldsymbol{M}}_{\boldsymbol{q}\Delta\beta}\omega^2-\frac{1}{2}\rho V^2\tilde{\boldsymbol{A}}_{\boldsymbol{q}\Delta\beta}} \tag{32}$$

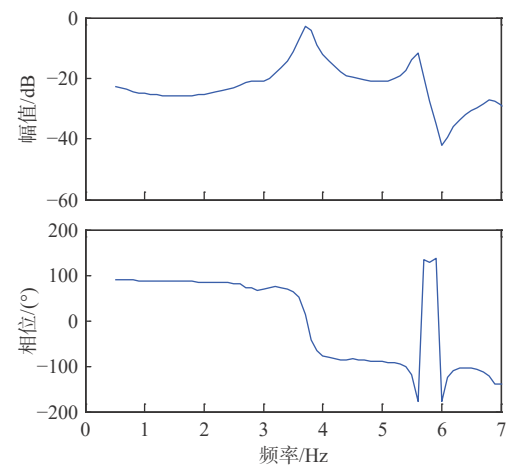
由式 (26) 计算得到$\boldsymbol{W}^{\boldsymbol{q}}_{\Delta\beta}$, 即为弹性飞机的机体传递函数。计算得到以俯仰轴升降舵偏为输入、俯仰角速率为输出的机体传递函数。同理, 得到滚转轴和偏航轴的机体传递函数, 见图 9。飞机部分固有模态和频率见表 3, 在不同频率处, 俯仰轴、滚转轴和偏航轴气动伺服弹性稳定性不一样, 其中俯仰轴传递函数在频率 3.0 Hz 和 3.8 Hz 附近响应较大, 对应的模态分别为发动机对称俯仰和机身垂直一弯(见图 10(a) 和图 10(b))。滚转轴传递函数在频率 2.1 Hz 和 4.0 Hz 附近响应较大, 对应的模态分别为机翼反对称一弯和发动机反对称侧偏(见图 10(c) 和图 10(d))。偏航轴传递函数在频率 3.4 Hz 附近响应较大, 对应的模态分别为机身水平一弯见图 10(e)。



(a) 升降舵偏到俯仰角速率机体传递函数



(b) 方向舵偏到偏航角速率机体传递函数



(c) 方向舵偏到偏航角速率机体传递函数

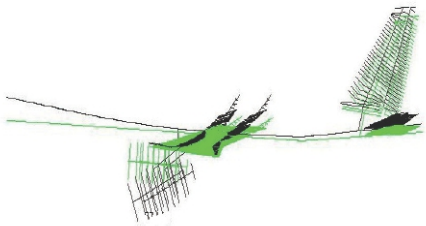
图 9　跨声速机体传递函数

Fig. 9　Transonic aircraft transfer function

表 3　飞机固有模态和频率

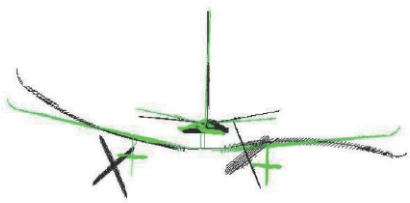
Table 3　Aircraft natural modes and frequencies

模态名称	频率/Hz
机翼反对称一弯	2.1
发动机对称俯仰	3.0
机身水平一弯	3.4
机身垂直一弯	3.8
发动机反对称侧偏	4.0

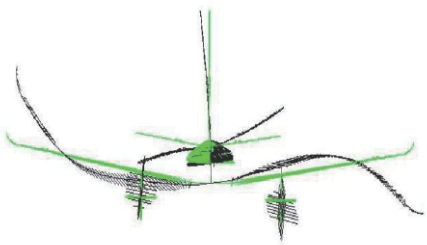


(a) 发动机对称俯仰模态

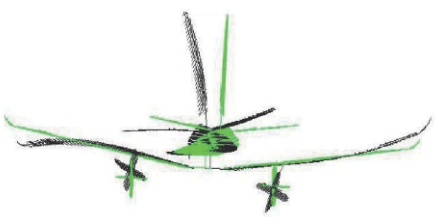
(b) 机身垂直一弯模态



(c) 发动机反对称侧偏模态



(d) 机翼反对称一弯模态



(e) 机身水平一弯模态

图 10 飞机固有模态示意图

Fig. 10 Diagram of aircraft natural modes

5 跨声速机体传递函数计算与试飞结果对比

为验证跨声速气动伺服弹性分析方法的可靠性，以某具有超临界翼型的大型民用客机为试验试飞研究对象，试飞机长为 38 m，翼展为 36 m，重量 50 000 kg，试飞条件为：

1）气象条件要求为无雨、无雪，试验动作区无云，气流平稳；

2）试飞机平飞加速到相应速度点，飞行高度偏差应满足小于 48 m，马赫数偏差应满足小于 0.015。

试飞高度为 10 000 m，在飞行马赫数达到跨声速试飞点时，通过扫频激励系统生成激励信号如图 10 所示，激励飞机舵面运动[25]。激励试验信号如表 4 所示。

表 4 ASE 试飞激励信号

Table 4 ASE flight test excitation signal

激励方法	激励位置	激励方式	激励频带/Hz	激励幅值/(°)	激励时间/s
扫频	升降舵	对称	0.5～40	≤1	≤30
扫频	副翼	反对称	0.5～40	≤1	≤30
扫频	方向舵		0.5～40	≤1	≤30

试验时飞控-机体系统为闭环状态，选择合适的位置注入激励信号并在同一位置取出飞控系统反馈信号，输入信号叠加反馈信号驱动舵面运动[26]，原理如图 11 所示。

使用闭环测试飞机开环传递函数，从而获得飞机的气动伺服弹性机体传递函数特性，激励信号、反馈信号和叠加信号如图 12～图 14 所示。

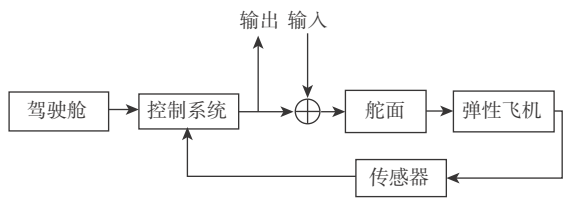


图 11 气动伺服弹性试飞原理示意图

Fig. 11 Schematic diagram of aeroservoelastic flight test principle

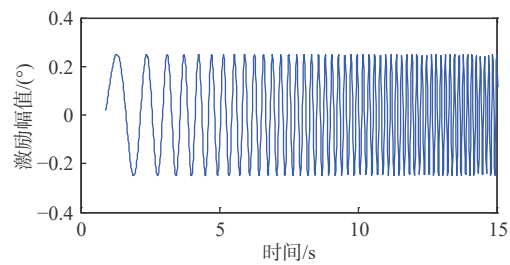


图 12 扫频激励信号

Fig. 12 Sweep frequency excitation signal

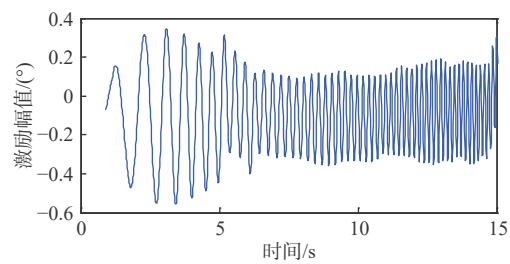


图 13 反馈信号

Fig. 13 Feedback signal

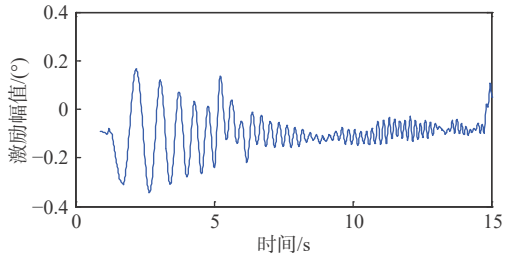


图 14　叠加信号

Fig. 14　Superposed signal

本次试验是用经典的 Bode 图方法来分析系统的稳定裕度。利用式(33)和式(34)计算系统的总开环传递函数：

$$G(\mathrm{j}\omega)=\frac{\sum \boldsymbol{Y}(\mathrm{j}\omega)\boldsymbol{X}^{*}(\mathrm{j}\omega)}{\sum \boldsymbol{X}(\mathrm{j}\omega)\boldsymbol{X}^{*}(\mathrm{j}\omega)}=\boldsymbol{A}(\omega)\mathrm{e}^{\mathrm{j}\varphi(\omega)} \tag{33}$$

$$\boldsymbol{L}(\omega)=20\lg(|\boldsymbol{A}(\omega)|) \tag{34}$$

式中：$\boldsymbol{X}(\mathrm{j}\omega)$和$\boldsymbol{Y}(\mathrm{j}\omega)$分别为系统输入信号和输出信号的拉普拉斯变换，由$\boldsymbol{L}(\omega)$和$\varphi(\omega)$绘制波德图。

现代民机普遍采用超临界机翼设计，机翼厚度相对弦长和展长较小，故在亚声速区将机翼简化成平板采用偶极子非定常气动力模型来分析气动弹性问题是比较合理的。选择高度为 6 000 m，速度为 Ma=0.4 的试飞状态点来验证亚声速偶极子气动力模型的气动伺服弹性分析的精度，和试飞结果对比如图 15 所示。由传递函数对比可知，亚声速时偶极子分析结果和试飞结果在频率上吻合较好，俯仰轴在幅值裕度上非常接近，滚转轴在机翼反对称一弯模态上分析幅值裕度偏大，偏航轴在机身水平一弯模态分析幅值裕度偏大。对比结果显示，在亚声速区，采用偶极子格网法分析民用飞机气动伺服弹性在工程上可接受。

得到飞行高度为 10 000 m，速度为 Ma=0.75，Ma=0.80 时的气动伺服弹性试飞数据之后，选择对应状态点的跨声速气动伺服弹性分析结果与偶极子气动力模型的气动伺服弹性分析结果以及试飞结果进行对比，对比结果如图 16 和图 17 所示。

从对比中可以发现，跨声速分析的机体传递函数频响在频率和幅值上均比偶极子气动力模型的分析更加接近试飞结果。以速度为 Ma=0.80 分析工况为例，在俯仰轴关键模态处跨声速分析的机体传递函数比偶极子气动力模型的分析频率精度提高 2%，幅值裕度精度提高 46%。在滚转轴关键模态处跨声速分析的机体传递函数比偶极子气动力模型的分析频率精度提高 11%，幅值裕度精度提高 44%。在偏航轴关键模态处跨

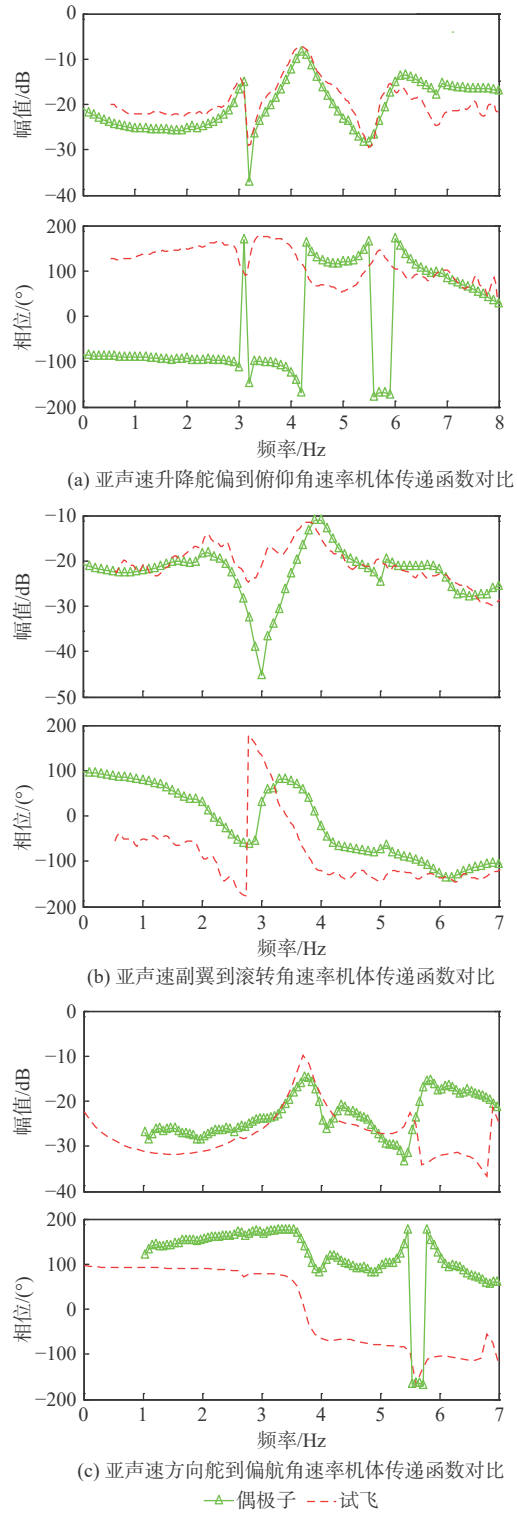
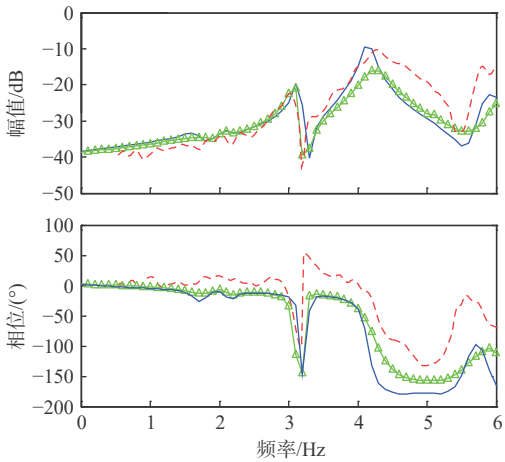


图 15　亚声速偶极子格网法分析与试飞结果对比

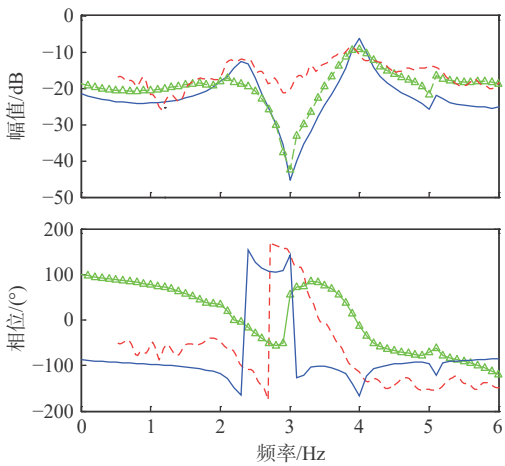
Fig. 15　Comparison of DLM analysis and flight test results of subsonic

声速分析的机体传递函数比偶极子气动力模型的分析频率精度提高 2%，幅值裕度精度提高 58%。结果表明基于叠加流场面元法的跨声速分析方法能够更精确的预测民用飞机气动伺服弹性稳定裕度。

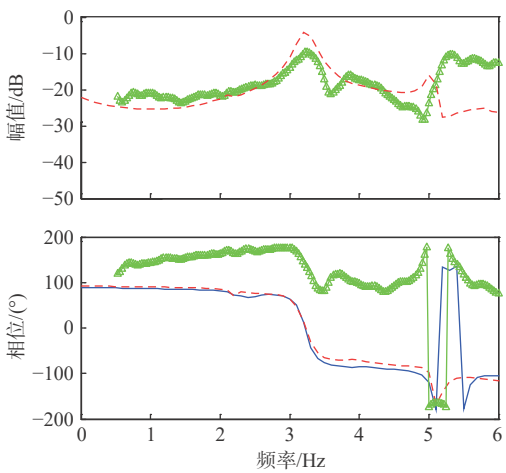
将跨声速计算的飞行高度为 10 000 m，不同马



(a) 升降舵偏到俯仰角速率机体传递函数对比



(b) 副翼到滚转角速率机体传递函数对比

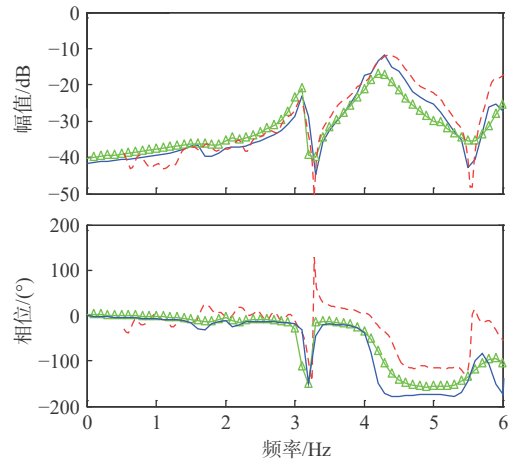


(c) 方向舵到偏航角速率机体传递函数对比

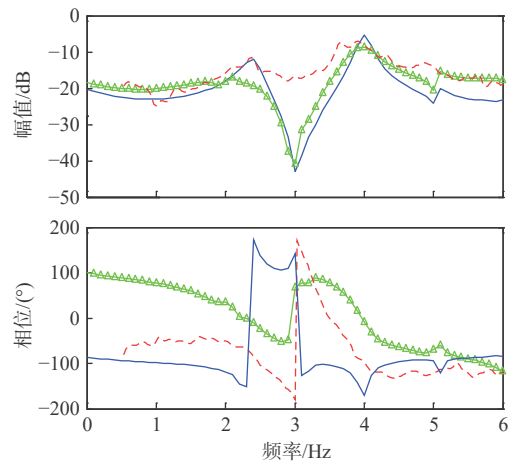
跨声速 偶极子 试飞

图 16 速度为 *Ma*=0.75 时分析与试飞结果对比

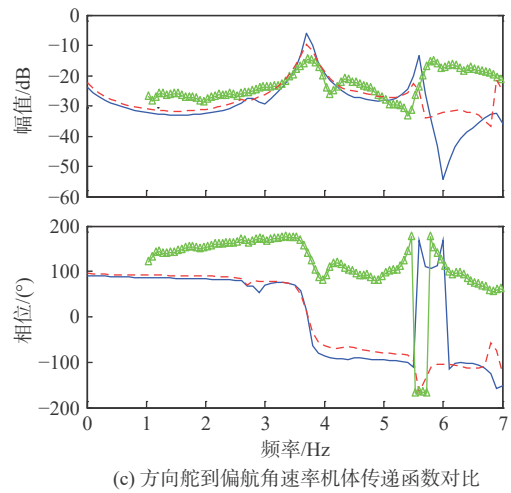
Fig. 16 Comparison of analysis and flight test results with *Ma*=0.75



(a) 升降舵到俯仰角速率机体传递函数对比



(b) 副翼到滚转角速率机体传递函数对比



(c) 方向舵到偏航角速率机体传递函数对比

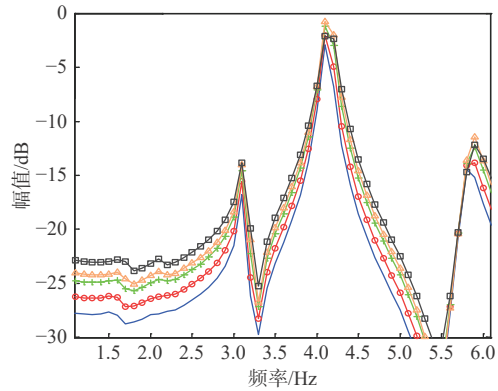
跨声速 偶极子 试飞

图 17 速度为 *Ma*=0.80 时分析与试飞结果对比

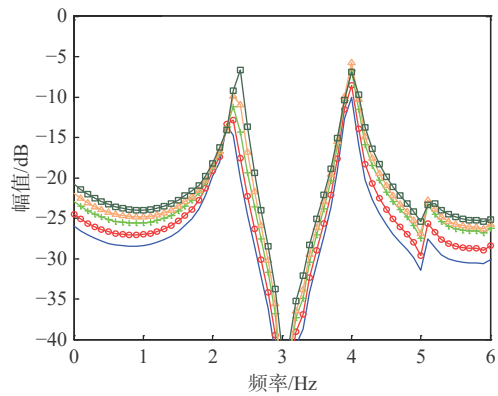
Fig. 17 Comparison of analysis and flight test results with *Ma*=0.80

赫数下的机体传递函数 bode 图放在一起做对比，对比结果如图 18 所示。

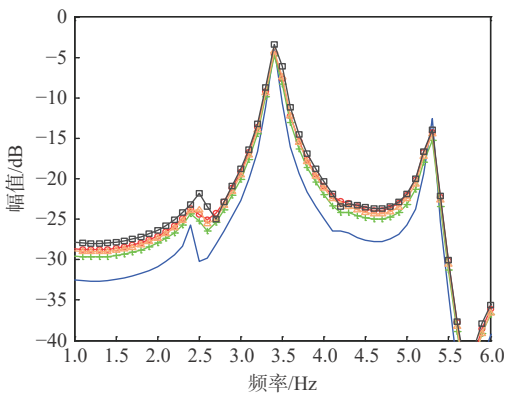
从对比中可以发现，俯仰轴机体传递函数频响在关键模态 3 Hz 处频率没有随马赫数变化，峰值却随马赫数增大而增大，在关键模态 3.8 Hz 处频率没有随马赫数变化，峰值随着马赫数先增大后减小。滚转轴机体传递函数频响在关键模态 2.2 Hz 处频率随马赫数增大而增大，峰值也随马赫数增大而增大，在关键模态 4 Hz 处频率没有随马赫数变化，峰值随着马赫数先增大后减小。偏航轴机体传递函数频响随马赫数变化不大。



(a) 不同马赫数的俯仰轴机体传递函数对比



(b) 不同马赫数的滚转轴机体传递函数对比



(c) 不同马赫数的偏航轴机体传递函数对比

Ma=0.70 *Ma*=0.75 *Ma*=0.80 *Ma*=0.82 *Ma*=0.85

图 18　不同马赫数的机体传递函数对比

Fig. 18　Comparison of aircraft transfer function with different Mach numbers

6　结　论

本文通过叠加流场面元法求解时间线化的跨声速小扰动速度势方程得到跨声速气动力影响系数，将其代入民用飞机气动伺服弹性方程中计算机体传递函数，将考虑跨声速气动力非线性效应的气动伺服弹性机体传递函数与使用偶极子气动力计算的机体传递函数以及试飞结果进行对比，得到以下结论。

1）本文方法计算的机体传递函数在跨声速区域比使用偶极子气动力计算的机体传递函数在频率和幅值上更加接近试飞结果。

2）在此基础上总结了民用飞机跨声速区气动伺服弹性机体传递函数响应的变化规律。这使得在研发阶段能够更准确、高效的对民用飞机跨声速区气动弹性-控制律耦合的多学科综合进行设计。

未来提高跨声速定常气动力计算精度及结构动力学模型的建模精度之后，跨声速气动伺服弹性分析的结果也会更加精确。

参考文献（References）

[1] PITT D, HAYES B, GOODMAN C. F/A-18E/F aeroservoelastic design, analysis, and test[C]//Proceedings of the 44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. Reston: AIAA, 2003.

[2] 黄锐, 胡海岩. 飞行器非线性气动伺服弹性力学[J]. 力学进展, 2021, 51(3): 428-466.
HUANG R, HU H Y. Nonlinear aeroservoelasticity of aircraft[J]. Advances in Mechanics, 2021, 51(3): 428-466(in Chinese).

[3] 胡海岩, 赵永辉, 黄锐. 飞机结构气动弹性分析与控制研究[J]. 力学学报, 2016, 48(1): 1-27.
HU H Y, ZHAO Y H, HUANG R. Studies on aeroelastic analysis and control of aircraft structures[J]. Chinese Journal of Theoretical and Applied Mechanics, 2016, 48(1): 1-27(in Chinese).

[4] 杨超, 黄超, 吴志刚, 等. 气动伺服弹性研究的进展与挑战[J]. 航空学报, 2015, 36(4): 1011-1033.
YANG C, HUANG C, WU Z G, et al. Progress and challenges for aeroservoelasticity research[J]. Acta Aeronautica et Astronautica Sinica, 2015, 36(4): 1011-1033(in Chinese).

[5] EDWARDS J W, BENNETT R M, WHITLOW W J, et al. Time-marching transonic flutter solutions including angle-of-attack effects[J]. Journal of Aircraft, 1983, 20(11): 899-906.

[6] BATINA J T. Efficient algorithm for solution of the unsteady transonic small-disturbance equation[J]. Journal of Aircraft, 1988, 25(7): 598-605.

[7] KIM D H, LEE I. Transonic and low-supersonic aerodynamic analysis of a wing with underpylon/store[J]. Journal of Aircraft, 2000, 37(1): 189-192.

[8] KIM D H, LEE I. CFD-based matched-point transonic and supersonic flutter computations using a modified TSD equation[J]. Computational Fluid Dynamics Journal, 2002, 11(1): 35-49.

[9] KIM J Y, KIM K S, LEE I, et al. Transonic aeroelastic analysis of all-movable wing with free play and viscous effects[J]. Journal of Aircraft, 2008, 45(5): 1820-1824.

[10] 张伟伟, 叶正寅. 非线性伺服气动弹性的时域数值模拟[J]. 振动工程学报, 2005(3): 103-107.
ZHANG W W, YE Z Y. The numerical simulation of nonlinear aero-servo-elasticity in time domain[J]. Journal of Vibration Engineering, 2005(3): 103-107(in Chinese).

[11] LUCIA D J, BERAN P S, SILVA W A. Reduced-order modeling: new approaches for computational physics[J]. Progress in Aerospace Sciences, 2004, 40(1-2): 51-117.

[12] ZHANG W W, WANG B B, YE Z Y, et al. Efficient method for limit cycle flutter analysis based on nonlinear aerodynamic reduced-order models[J]. AIAA Journal, 2012, 50(5): 1019-1028.

[13] KOU J Q, ZHANG W W. Reduced-order modeling for nonlinear aeroelasticity with varying Mach numbers[J]. Journal of Aerospace Engineering, 2018, 31(6): 04018105.

[14] ZHANG W W, KOU J Q, WANG Z Y. Nonlinear aerodynamic reduced-order model for limit-cycle oscillation and flutter[J]. AIAA Journal, 2016, 54(10): 3304-3311.

[15] WAITE J, STANFORD B, BARTELS R E, et al. Active flutter suppression controllers derived from linear and nonlinear aerodynamics: application to a transport aircraft model[C]// Proceedings of the Applied Aerodynamics Conference. Reston: AIAA, 2018.

[16] SILVA W. AEROM: NASA's unsteady aerodynamic and aeroelastic reduced-order modeling software[J]. Aerospace, 2018, 5(2): 41.

[17] WAITE J, STANFORD B, BARTELS R E, et al. Active flutter suppression using reduced order modeling for transonic aeroservoelastic control law development[C]//Proceedings of the AIAA Aviation 2019 Forum. Reston: AIAA, 2019.

[18] VOSS R. Calculation of unsteady transonic flow about oscillating wings by a field panel method[C]//Proceedings of the Third GAMM-Seminar Kiel. Berlin: Springer, 1988.

[19] LU S, VOSS R. TDLM - a transonic doublet lattice method for 3D potential unsteady transonic flow calculation: DLR-FB 92-25[R]. Cologne: DLR, 1992.

[20] CHEN P C, LEE H W, LIU D D. Unsteady subsonic aerodynamics for bodies and wings with external stores including wake effect[J]. Journal of Aircraft, 1993, 30(5): 618-628.

[21] MURMAN E M. Analysis of embedded shock waves calculated by relaxation methods[J]. AIAA Journal, 1974, 12(5): 626-633.

[22] VAN ZYL L H, AEROTEK C, PRETORIA S A. A transonic doublet lattice method for general configurations [EB/OL]. (2016-06-20) [2024-10-07]. https://www.icas.org/icas_archive/ICAS1996/ICAS-96-2.8.1.pdf.

[23] LIU D D, KAO Y F, FUNG K Y. An efficient method for computing unsteady transonic aerodynamics ofswept wings with control surfaces[J]. Journal of Aircraft, 1988, 25(1): 25-31.

[24] FORSCHING H W. 气动弹性力学原理[M]. 沈克扬, 译. 上海: 上海科学技术文献出版社, 1982.
FORSCHING H W. Grundlagen der aeroelastik[M]. SHEN K Y , translated. Shanghai: Shanghai Scientific and Technological Literature Publishing House, 1982(in Chinese).

[25] MICHAEL W K. A historical overview of flight flutter testing: NASA-TM-4720[R]. Washington, D. C. : NASA Technical Memorandum, 1995.

[26] JURISSON A, DE BREUKER R, DE VISSER C C, et al. Aeroservoelastic flight testing platform development for system identification[C]//Proceedings of the AIAA SCITECH 2022 Forum. Reston: AIAA, 2022.

Transonic aeroservoelastic analysis of civil aircraft based on overset field-panel method

WU Shanqiang，CHEN Qi，SUN Yajun，NI Zao，CHEN Wen*

(Shanghai Aircraft Design and Research Institute，Shanghai 201210，China)

Abstract：Compared with subsonic, when a civil aircraft flies in transonic, it's aeroservoelastic characteristics appear as obvious frequency deviation and stability margin decrease. he aeroservoelastic design in the transonic should be thoroughly taken into account during the preliminary design stage since the development cost is high, the test flight risk is high, and the design change cycle is lengthy for aircraft in the development phase. The unsteady aerodynamic calculated by DLM cannot take into account the effects of the transonic shock wave and the boundary layer. The unsteady small disturbance velocity potential is separated into steady and unsteady terms in order to calculate the transonic unsteady aerodynamic. Computational fluid dynamic（CFD）is used to solve the transonic steady term, and the unsteady terms are then calculated by solving the time-linear transonic small disturbance velocity potential equation using the unsteady flow field element method. The obtained transonic unsteady aerodynamic influence coefficient is substituted into the calculation of the frequency response function of the aircraft transfer function. The calculated results are compared with DLM aerodynamic model computation results and flight test results. The transonic aeroservoelastic analysis based on the Over Field-Panel Method is more consistent with the flight test results in terms of frequency deviation and stability margin reduction. Exploring the frequency and stability margin of the transfer function in the transonic region to provide a method for aerosevoelastic design in the transonic region.

Keywords：civil aircraft; aeroservoelasticity; computational fluid dynamic; panel method; transonic

Received：2024-10-09；**Accepted**：2025-01-23；**Published Online**：2025-03-14 13：20

URL：link.cnki.net/urlid/11.2625.V.20250313.1707.002

Foundation item：Scientific and Technological Innovation Project of COMAC (Y23GS29)

* **Corresponding author**. E-mail：chenwen1@comac.cc

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0711

智能制造数字孪生模型可变模糊综合评估

王向章[1,2,*]，于莉莉[2]

(1. 西南交通大学 交通运输与物流学院，成都 610031；2. 广电计量检测集团股份有限公司，广州 510630)

摘 要：针对智能制造数字孪生模型缺乏统一的评估方法和参考标准，采用目标-问题-指标(GQM)方法建立一套系统的数字孪生模型多维评估指标体系，给出指标量化计算方法，并综合可变模糊识别模型和信息熵集结权重算法优势，构建基于改进可变模糊模型的数字孪生模型质量价值评估方法。为解决专家评价的模糊性、不确定性等问题，该方法利用群决策理论改进了可变模糊识别模型，并在考虑专家意见偏好基础上利用信息熵计算指标权重，以某飞机制造工厂数字孪生模型为例，对该模型质量和应用能力进行评估。实例分析表明：该飞机制造工厂数字孪生模型的评估等级为“S_4良好”，但偏向于“S_3合格”，还有优化完善的空间，同时验证了改进可变模糊识别模型的可行性和合理性，为规范智能制造数字孪生模型构建提供了方法支撑。

关 键 词：数字孪生模型；GQM方法；信息熵；群决策；可变模糊识别

中图分类号：TP301.6

文献标志码：A **文章编号**：1001-5965(2026)01-0180-12

数字孪生能够在物理实体与虚拟模型之间实现精准映射[1]，是制造行业实现数字化、智能化转型的最为关键和基础性技术之一。数字孪生的核心要素是数字孪生模型，又称数字孪生体或虚拟模型[2]，其通过精准描述物理实体的几何、物理、行为、规则等多维度属性[3]，在物理实体运行数据驱动下，对物理实体的实际行为和运行状态进行真实刻画，进而实现对物理实体的预测、评估、优化等重要服务。目前，智能制造数字孪生已被航空航天、船舶、汽车等制造行业广泛应用。随着用户需求的不断升级及应用场景的不断变化，智能制造数字孪生模型将更精细化、智能化，并更好的为用户提供服务。然而，当前智能制造行业的数字孪生模型的构建方式缺乏统一的标准规范，对模型的构建校验、重构优化、迁移复用、运行管理等都缺乏系统性的评估理论和方法，而采用传统方法得到的评估结果不具有一致性，使得用户难以量化分析数字孪生模型的质量和应用能力，在一定程度上影响了数字孪生在智能制造行业的应用。

在数字孪生发展过程中，国内外专家学者对数字孪生模型的研究逐渐增多，但主要集中在实践应用、算法研究和模型理论等方面。在实践应用方面，Berri等设计了关于电液伺服阀流体动力学行为的数字孪生模型，并对数字孪生模型精准程度进行了比较和评价[4]。陆剑锋等研究了制造企业数字孪生系统3个种群构建方法及实时交互与动态可视演化过程[5]。Wang等提出一种基于数字孪生模型的城市交通流量预测方法，该方法可以实时、有效地预测城市中不同路段的交通流量[6]。在算法研究方面，Chen等提出一种基于深度神经网络的数字孪生模型，该模型可以在短时间内对工业制造过程进行实时监测和预测[7]；Li等提出一种基于可解释性机器学习的数字孪生模型，该模型能进一步提高模型的可解释性和可靠性[8]。在模型理论方面，陶飞等对已有数字孪生三维模型进行了扩展，提出数字孪生五维模型概念[3]；张辰源等针对数字孪生模型

收稿日期：2023-10-31；录用日期：2023-12-20；网络出版时间：2024-01-03 15:07
网络出版地址：link.cnki.net/urlid/11.2625.V.20240102.1526.002
*通信作者. E-mail：513432965@qq.com

引用格式：王向章，于莉莉. 智能制造数字孪生模型可变模糊综合评估[J]. 北京航空航天大学学报，2026，52(1)：180-191.
WANG X Z，YU L L. Variable fuzzy comprehensive evaluation for intelligent manufacturing digital twin model[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：180-191 (in Chinese).

缺乏评价理论, 构建了一套数字孪生模型评价指标体系, 但未给出模型的综合评估方法[9]; 李翔宇针对如何构建数字孪生模型问题, 提出了数字孪生模型的构造框架[10]。

从以上研究可以看出, 在应用服务和智能化方面, 虽然各领域数字孪生模型的应用场景和采用的智能算法存在较大差异, 但构建的数字孪生模型基本具备了精准模拟、实时交互、连接高效、有效预测及可视交互等特征。在模型理论方面, 多偏向于研究如何描述和构建数字孪生模型等理论, 鲜有开展数字孪生模型评估方法研究。由于国内外市场对智能制造领域所构建的数字孪生均缺少统一的评估准则, 导致市场上的数字孪生千差万别, 同时也带来企业投入与回报不成正比等问题, 因此, 亟需制定有效的智能制造数字孪生模型评估方法标准, 以提高数字孪生模型的质量水平和应用价值。

因此, 本文以智能制造领域数字孪生模型为研究对象, 开展数字孪生模型评估理论方法研究, 以期为数字孪生模型评估的标准化实现提供理论依据和方法参考。首先, 本文采用目标-问题-指标(goal-question-metric, GQM)方法建立一套系统的数字孪生模型多维评估指标体系; 然后, 根据指标体系特点, 采用信息熵集结法计算指标权重, 利用群决策改进可变模糊识别模型构建数字孪生模型评估方法; 最后, 以“飞机制造工厂”数字孪生模型为例, 验证本文评估方法的适用性及泛用性。

1　数字孪生模型评估指标选取

对智能制造数字孪生模型进行综合评估, 可为模型实际应用、优化改进及模型定型提供决策依据。最新研究成果表明[2-3,9,11], 数字孪生模型的构建遵循了精准刻画、实时连接、轻量高效、可重组编排、可视交互及统一规范等准则, 构建过程主要包括建立、组装、验证、校正及管理等过程, 并由反映物理实体的几何模型、行为模型、物理模型及规则模型等组成。对于如此复杂的系统模型, 要准确评估其质量和应用能力, 需考虑的因素较多。本文在参考文献 [12-14] 相关标准条款后, 根据目前智能制造行业数字孪生模型的功能特点和构建要求, 总结多名智能制造数字孪生专家意见, 确定从准确程度、连接程度、高效程度、可视化程度、可编编排程度及规范化程度 6 个维度进行评估。

数字孪生模型的准确程度评估主要是评估孪生模型复现物理实体特效、状态、行为的能力及预测和决策的准确性; 连接程度评估的是数字孪生模型与真实世界物理系统之间连接的稳定性和可靠性; 高效程度评估数字孪生模型的数据处理速度和运算性能, 能够检验数字孪生模型对实时数据的处理能力、运行效率等; 可视化程度评估指的是对数字孪生模型的输出结果进行视觉化展示的程度和质量的评估; 可编排性评估的是特定任务和目标下, 组合数字孪生模型组件, 配置模型参数及确定参数关联关系过程的表现; 规范化程度主要评估数字孪生模型在不同数据集和应用场景下的复用能力, 能够提高数字孪生模型的兼容性和标准化。

因此, 本文遵从指标体系科学性、系统性、通用性、可操作性及可拓展性等原则, 根据数字孪生模型构建流程, 同时参考最新研究成果, 采用 GQM 方法[15], 从上述 6 个维度建立一套科学、系统和可操作的数字孪生模型评估指标体系, 为数字孪生模型的应用和发展提供有力支持。GQM 方法通过定义度量目标, 提出实现目标可能遇到的问题, 并给出每个问题的测量方法, 最后, 将测量方法提炼形成指标。运用 GQM 方法初步识别出 27 个评估指标, 为优化和完善指标体系, 采用李克特量表将每个评估指标的重要性程度划分为“很重要、比较重要、一般重要、不重要、很不重要”5 个等级, 并设计了调查问卷, 并将问卷发放给航空、船舶、精密社备制造等行业数字孪生模型的设计、使用人员。共计发出问卷 122 份, 收回有效问卷 115 份, 有效回收率达到 94.3%。采用 SPSS22.0 对问卷信度进行检验, 若问卷的克隆巴赫系数 α 达到 0.9 及以上, 则问卷可靠度较高, 具有一定参考价值。信度统计结果如表 1 所示。

表 1　信度统计结果

Table 1　Reliability statistics results

α	基于标准化项目的α	项数
0.913	0.934	27

根据调查问卷数据统计, 经计算每个指标重要性评分的平均值、标准差, 得出每个指标的变异系数。若指标重要性平均值大于 3.0, 变异系数小于 0.35, 则该指标可作为入选指标。按照上述筛选标准, 最后剔除了 6 个指标, 再经专家多次会议讨论交流后, 对入选指标进行修订, 得到以数字孪生模型质量和应用能力为评估总目标, 准确程度、连接程度、高效程度等 6 个维度为子目标, 21 个因素为指标的数字孪生模型评估指标体系, 如表 2 所示。

表2 数字孪生模型评估指标体系

Table 2 Evaluation index system of digital twin model

目标(Goal)	问题(Question)	因素(Metric)
准确程度	数字孪生模型中表征物理实体的几何、物理属性或其他可扩展的静态参数是否无重复、无缺漏、无歧义	静态参数的准确性
	数字孪生模型的动态参数是否能够准确表征物理实体行为状态、运行环境等的动态变化过程	动态参数的准确性
	不同参数设置下,数字孪生模型各参数之间的关联关系是否能够反映物理实体整体的准确变化	各参数关联的准确性
连接程度	物理实体状态动态更新时,数字孪生模型参数是否能够实现物理实体状态的同步复刻、监测和优化	模型与物理实体的连接程度
	不同开发模式、部署模式及运行模式下,数字孪生模型之间是否能够可靠连接并实现数据安全、实时的交互共享和同步	模型之间信息交互共享程度
	数字孪生模型与服务对象之间是否能够及时有效的连接,以便准确监控、仿真,并能利用服务对象反馈的数据保证自身的时效性	模型与服务的连接程度
	数字孪生模型是否能安全可靠地连接物理实体数据、运行数据、服务数据、历史数据等孪生数据,并通过数据融合实现精确模拟、控制	模型与孪生数据的连接程度
高效程度	描述物理实体的几何、物理、行为、规则等模型的参数量、层次结构、约束条件、转换规则、规则数量等是否轻量化	模型的轻量化程度
	数字孪生模型是否能够快速处理实时数据,并开展知识挖掘、可视化渲染、预测优化等服务	模型运行的高效程度
	数字孪生模型是否能够利用硬盘、内存、缓存、预取等计算资源、存储资源、网络资源执行服务	模型资源利用的高效程度
可视化程度	数字孪生模型各参数是否能够以三维图像等方式全面展示物理实体属性,并实时展示物理实体的运行状态	模型参数的可视化程度
	数字孪生模型结构的可视化是否能够呈现数字孪生子模型之间的耦合关系及各个参数之间的关联关系	模型结构的可视化程度
	数字孪生模型运行过程是否能够直观反映物理实体的全局运行状态和过程细节	模型运行的可视化程度
	数字孪生模型的演化过程是否能够直观反映物理实体的运行规律,并基于某时刻物理实体参数数据,仿真出物理实体的真实状态	模型演化过程的可视化程度
可编排程度	是否能够根据特定任务和目标,选择结构、内容、功能和服务可调整的模型组件进行自由组合,构建新的数字孪生模型	模型组件的可编排性
	数字孪生模型的参数名、参数描述、参数类型、默认值及校验规则等是否可配置和组合	模型参数的可编排性
	数字孪生模型与其他模型组合时,是否能够根据业务流程对新模型各关联关系进行重新编排	模型关联关系的可编排性
规范化程度	数字孪生模型及其子模型的格式是否符合通用标准要求,并能进行解析	模型格式的一致性
	数字孪生模型各参数的单位是否符合国际单位制,且有助于用户转化、解析	模型参数单位的一致性
	数字孪生模型的数据接口是否具有统一的标准,是否能在不同模型之间能够正确解析、传递和处理数据	模型数据接口的标准化
	各数字孪生模型一起运行时是否相互配合稳定工作,不存在频繁崩溃、卡死等问题	模型之间的兼容性

2 数字孪生模型评估指标测量方法

按照GQM方法,针对本文提出的实现目标的问题,还应为每个问题定义相应的度量方法,用于量化和评估目标的实现程度。因此,本文分别对准确程度、连接程度、高效程度等6个维度下的21个指标设计了量化计算式,以便评估者客观准确的开展评估。

2.1 准确程度

2.1.1 静态参数的准确性

静态参数是指表征物理实体几何、物理等固定属性或特征的参数,描述了物理实体的静态状态,在运行过程中不随时间变化。静态参数准确性的量化计算方法为

$$x_1=\frac{1}{N_1}\sum_{i=1}^{N_1}\left(1-\frac{\left|x_{1i}-x'_{1i}\right|}{x_{1i}}\right) \tag{1}$$

式中:N_1为数字孪生模型映射的静态参数数量;x_{1i}为第i个映射静态参数在物理实体中的值;x'_{1i}为第i个映射静态参数在数字孪生模型中的值。对于非数值参数或不便于量化的静态参数,可凭经验对$1-\left|x_{1i}-x'_{1i}\right|/x_{1i}$进行评估,取值范围为[0,1];对于缺失的第$i$个静态参数,则$1-\left|x_{1i}-x'_{1i}\right|/x_{1i}$赋值为0。

2.1.2 动态参数的准确性

动态参数是描述物理实体行为状态、运行环境等变化的参数,可以通过实时采集的传感器数据、监测数据等来模拟和监测物理实体实时行为和状态。动态参数准确性的量化计算方法为

$$x_2=\frac{1}{N_2}\sum_{i=1}^{N_2}\left(1-\frac{\left|x_{2i}-x'_{2i}\right|}{x_{2i}}\right) \tag{2}$$

式中:N_2为数字孪生模型映射的动态参数数量;x_{2i}为第i个映射动态参数在物理实体中的值;x'_{2i}为第i个映射动态参数在数字孪生模型中的值。对于非数值参数或不便于量化的参数,可凭经验对$1-|x_{2i}-x'_{2i}|/x_{2i}$进行评估,取值范围为[0,1];对于缺失的第$i$个动态参数,则$1-\left|x_{2i}-x'_{2i}\right|/x_{2i}$赋值为0。

2.1.3 各参数关联的准确性

数字孪生模型中各参数之间存在隐性和显性的关联关系,可以帮助用户更好地理解和分析物理实体的运行逻辑,并进行优化、预测和决策制定等。各参数关联的准确性的量化计算方法为

$$x_3 = \frac{1}{N_3}\sum_{i=1}^{N_3} x_{3i} \tag{3}$$

式中：N_3 为数字孪生模型中参数关联数组；x_{3i} 为第 i 组参数关联关系反映物理实体运行逻辑和变化的准确程度，可凭经验或结合实际运行数据进行评估，取值范围为 [0,1]；对于缺失关联关系的参数数组，则 x_{3i} 赋值为 0。

2.2　连接程度

2.2.1　模型与物理实体的连接程度

数字孪生模型各参数与物理实体通过数据采集、传感器、实时监测等技术实现连接，有效的连接可实现物理实体状态的同步复刻、监测和控制。数字孪生模型与物理实体连接程度的量化计算方法为

$$x_4 = \frac{1}{N_4}\sum_{i=1}^{N_4} x_{4i} \tag{4}$$

式中：N_4 为数字孪生模型中与物理实体实时连接的参数个数；x_{4i} 为第 i 个参数与物理实体的连接情况，可根据连接程度进行赋值，取值范围为 [0,1]，取值越大，连接程度越高。不能连接，则 x_{4i} 赋值为 0。

2.2.2　模型之间信息交互共享程度

数字孪生模型之间在不同开发模式、部署模式及运行模式下，可通过网络可靠连接并实现数据安全、信息共享和同步。数字孪生模型之间连接后信息交互共享程度量化计算方法为

$$x_5 = \frac{1}{N_5}\sum_{i=1}^{N_5} x_{5i} \tag{5}$$

式中：N_5 为数字孪生模型与其他数字孪生模型进行信息交互共享的参数个数；x_{5i} 为第 i 个参数与其他数字孪生模型进行有效信息交互共享的情况，可根据实时交互共享情况进行赋值，取值范围为 [0,10]，取值越大，信息实时交互共享程度越高。信息不能实时交互共享，则 x_{5i} 赋值为 0。

2.2.3　模型与服务的连接程度

数字孪生模型与服务有效连接可以更好地模拟和优化物理实体的行为和状态，并利用服务对象反馈的数据保证自身的时效性。数字孪生模型与服务有效连接程度的量化计算方法为

$$x_6 = \frac{1}{N_6}\sum_{i=1}^{N_6} x_{6i} \tag{6}$$

式中：N_6 为数字孪生模型与应用服务有效连接的参数个数；x_{6i} 为第 i 个参数与应用服务进行有效连接的程度，可根据实时连接情况进行赋值，取值范围为 [0,1]，取值越大，与应用服务实时连接程度越高。与服务功能不能实时连接，则 x_{6i} 赋值为 0。

2.2.4　模型与孪生数据的连接程度

数字孪生模型与物理实体数据、运行数据、服务数据、历史数据等孪生数据安全可靠地连接，并通过整合和融合多源数据后，可以实现对物理实体的精确模拟、控制和预测。模型与孪生数据连接程度的量化计算方法为

$$x_7 = (x_{71}w_1 + x_{72}w_2 + x_{73}w_3 + x_{74}w_4) \tag{7}$$

式中：x_{71} 为数字孪生模型与体系中物理实体数据的连接程度，评价打分赋值范围为 [0,1]，分值越高，表示与物理实体数据连接的程度越高；x_{72} 为数字孪生模型与体系运行数据的连接程度，评价打分赋值范围为 [0,1]，分值越高，表示与运行数据的连接程度越高；x_{73} 为数字孪生模型与体系服务数据的连接程度，评价打分赋值范围为 [0,1]，分值越高，表示与服务数据的连接程度越高；x_{74} 为数字孪生模型与历史数据的连接程度，评价打分赋值范围为 [0,1]，分值越高，表示与历史数据的连接程度越高。w_1、w_2、w_3、w_3 分别为 x_{71}、x_{72}、x_{73}、x_{74} 的权重值。

2.3　高效程度

2.3.1　模型的轻量化程度

数字孪生模型轻量化是指对几何、物理、行为、规则等模型的参数量、层次结构、约束条件、转换规则等进行轻量化，以提高模型的效率和可靠性。数字孪生模型的轻量化程度的量化计算方法为

$$x_8 = (x_{8i} - x'_{8i})/x_{8i} \tag{8}$$

式中：x_{8i} 为数字孪生模型在轻量化处理前的参数个数；x'_{8i} 为数字孪生模型轻量化处理后的参数个数。可根据模型轻量化前后所有参数量的变化情况，对数字孪生模型的轻量化程度进行评价打分，赋值范围为 [0,1]，分值越高，表示模型的轻量化程度越高。若 $x'_{8i} > x_{8i}$，则 $(x_{8i} - x'_{8i})/x_{8i}$ 赋值为 0。

2.3.2　模型运行的高效程度

数字孪生模型高效运行是指在实时性要求较高的场景下能快速响应，并处理大规模的数据和复杂的计算及完成可视化渲染、预测优化等任务所用时间。数字孪生模型运行高效程度的量化计算方法为

$$x_9 = (x_{9i} - x'_{9i})/x_{9i} \tag{9}$$

式中：x_{9i} 为数字孪生模型轻量化处理前的同步响应时间；x'_{9i} 为数字孪生模型轻量化处理后的同步响应时间。可根据模型轻量化前后的同步响应时间变化情况，对模型运行的高效程度进行评价打分，赋值范围为 [0,1]，分值越高，表示模型运行越高效。若 $x'_{9i} > x_{9i}$，则 $(x_{9i} - x'_{9i})/x_{9i}$ 赋值为 0。

2.3.3 模型资源利用的高效程度

数字孪生模型利用资源的高效程度是指模型利用系统计算资源、存储资源、网络资源等执行相应服务，确保模型能够充分发挥其功能，并提供准确、可靠的模拟和预测结果。数字孪生模型对资源利用的高效程度的量化计算方法为

$$x_{10}=(x_{10i}-x'_{10i})/x_{10i} \tag{10}$$

式中：x_{10i}为数字孪生模型轻量化处理前，利用计算资源、存储资源、网络资源执行所有服务的响应时间；x'_{10i}为数字孪生模型轻量化处理后，利用计算资源、存储资源、网络资源执行服务的响应时间。可根据模型轻量化前后利用上述资源执行服务的响应时间的变化情况，对模型资源利用的高效程度进行评价打分，赋值范围为[0,1]，分值越高，表示资源利用越高效。若$x'_{10i}>x_{10i}$，则$(x_{10i}-x'_{10i})/x_{10i}$赋值为0。

2.4 可视化程度

2.4.1 模型参数的可视化程度

数字孪生模型参数可视化是指参数能以三维图像等方式全面展示物理实体属性和运行状态，帮助用户理解模型的内部工作原理，发现潜在模式和规律及进行模型优化和调试。模型参数反映物理实体的可视化程度的量化计算方法为

$$x_{11}=p'/N_{11} \tag{11}$$

式中：N_{11}为所有数字孪生模型参数的总数；p'为能够直观呈现物理实体多维属性和运行状态的数字孪生模型参数个数。可根据直观呈现物理实体状态的参数的占比情况，对数字孪生模型参数的可视化程度进行评价打分，赋值范围为[0,1]，分值越高，表示模型参数的可视化程度越高。

2.4.2 模型结构的可视化程度

数字孪生模型结构可视化是指对模型之间的耦合关系及各参数间的关联关系进行呈现，以帮助用户理解模型组成部分、信息流动路径和层次结构。数字孪生模型结构可视化程度的量化计算方法为

$$x_{12}=\frac{q+q'}{N_{12}+N'_{12}} \tag{12}$$

式中：N_{12}为数字孪生模型结构中耦合关系总数；N'_{12}为数字孪生模型结构中参数关联关系总数；q为数字孪生模型结构中能够直观呈现的耦合关系数量；q'为数字孪生模型结构中能够直观呈现的参数关联关系数量。可根据数字孪生模型结构呈现的模型耦合关系和参数关联关系数量占比情况，对数字孪生模型结构的可视化程度进行评价打分，赋值范围为[0,1]，分值越高，表示模型结构的可视化程度越高。

2.4.3 模型运行的可视化程度

数字孪生模型运行可视化是指模型能够直观反映物理实体的运行状态和过程细节，帮助用户更好地理解模型的输入、输出、内部计算过程和运行行为，从而提高对模型的理解和解释能力。数字孪生模型运行可视化程度的量化计算方法为

$$x_{13}=\frac{1}{N_{13}}\sum_{i=1}^{N_{13}}x_{13i} \tag{13}$$

式中：N_{13}为所有数字孪生子模型数量。x_{13i}为第i个数字孪生子模型直观反映物理实体全局运行状态或过程细节的情况，可根据模型运行可视化程度进行赋值，取值范围为[0,1]，取值越高，则模型运行的可视化程度越高。子模型不能反映物理实体运行状态或过程细节，则x_{13i}赋值为0。

2.4.4 模型演化过程的可视化程度

数字孪生模型演化过程可视化是指模型能够直观反映物理实体的运行规律，并依据物理实体相关数据仿真出物理实体的真实状态。数字孪生模型演化过程可视化程度的量化计算方法为

$$x_{14}=s \tag{14}$$

式中：s表示数字孪生模型在不接收物理实体运行数据时的仿真过程，由专家根据某时刻模型仿真物理实体的程度进行赋值，取值范围为[0,1]，取值越高，则模型演化过程的可视化程度越高。

2.5 可编排程度

2.5.1 模型组件的可编排性

数字孪生模型组件的可编排性指模型能够根据特定任务和目标，对不同模型组件进行灵活编排和组合，从而实现更高效、更准确的模拟、预测和优化。数字孪生模型组件可编排性的量化计算方法为

$$x_{15}=t/N_{15} \tag{15}$$

式中：N_{15}为数字孪生模型所有组件的总数；t为能够编排的数字孪生模型组件个数。可根据能够编排的数字孪生模型组件占比情况，对数字孪生模型组件的可编排性进行评价打分，赋值范围为[0,1]，分值越高，表示模型组件的可编排程度越高。

2.5.2 模型参数的可编排性

数字孪生模型参数的可编排性是指参数名、参数描述、参数类型、默认值及校验规则等参数可以灵活编排和组合，从而实现更好的性能和效果。数字孪生模型参数可编排性的量化计算方法为

$$x_{16}=u/N_{16} \tag{16}$$

式中：N_{16}为数字孪生模型参数总数；u为能够编排的数字孪生模型参数个数。可根据能够编排的数

字孪生模型参数占比情况, 对数字孪生模型参数的可编排性进行评价打分, 赋值范围为 [0,1], 分值越高, 表示模型参数的可编排程度越高。

2.5.3　模型关联关系的可编排性

数字孪生模型关联关系的可编排性是指能够根据业务流程, 对新模型之间的关联关系进行重新编排和组合。数字孪生模型关联关系的量化计算方法为

$$x_{17} = v/N_{17} \tag{17}$$

式中: N_{17} 为数字孪生模型中关联关系的的总数; v 为能够编排的数字孪生模型关联关系的个数。可根据能够编排的数字孪生模型关联关系数量占比情况, 对数字孪生模型关联关系的可编排性进行评价打分, 赋值范围为 [0,1], 分值越高, 表示模型关联关系的可编排程度越高。

2.6　规范化程度

2.6.1　模型格式的一致性

数字孪生模型格式的一致性是指各个模型的输入输出格式均符合通用标准要求, 并能进行解析, 以确保数据的正确传递和交互。数字孪生模型中各个子模型格式的一致程度量化计算方法为

$$x_{18} = \frac{1}{N_{18}} \sum_{i=1}^{N_{18}} x_{18i} \tag{18}$$

式中: N_{18} 为数字孪生模型中所有子模型的总数; x_{18i} 为第 i 个子模型格式的一致程度。由专家根据用户解析和使用数字孪生子模型的难易程度进行赋值, 取值范围为 [0,1], 分值越高, 越容易解析和使用。对于完全不能解析或使用过程中难以接受的数字孪生子模型格式, 则 $x_{18i} = 0$。

2.6.2　模型参数量单位的一致性

数字孪生模型参数量单位的一致性指模型参数量应符合国际单位制, 且能转化、解析。数字孪生模型参数量单位的一致程度量化计算方法为

$$x_{19} = \frac{1}{N_{19}} \sum_{i=1}^{N_{19}} x_{19i} \tag{19}$$

式中: N_{19} 为数字孪生模型参数总数; x_{19i} 为第 i 个数字孪生模型参数单位的一致程度。对于符合国际单位制的参数, $x_{19i} = 1$; 对于不符合国际单位制, 但有利于用户转换、解析并使用的参数, 可根据参数单位解析和使用难度进行赋值, 取值范围为 [0,1], 分值越高, 越容易解析使用; 对于既不符合国际单位制又不利于用户使用的参数, 则 $x_{19i} = 0$。

2.6.3　模型数据接口的标准化

数字孪生模型数据接口的标准化是指定义统一的数据接口规范, 以确保不同模型之间能够正确解析、传递和处理数据。数字孪生模型数据接口的标准化程度的计算方法为

$$x_{20} = \frac{1}{N_{20}} \sum_{i=1}^{N_{20}} x_{20i} \tag{20}$$

式中: N_{20} 为数字孪生模型与外界进行数据交互的参数总数; x_{20i} 为第 i 个参数对应接口的标准化程度。x_{20i} 可根据用户解析和使用接口难度进行赋值, 取值范围为 [0,1], 分值越高, 表示越容易解析或接口。对于完全不能解析或使用过程较难接受的模型数据接口, 则 $x_{20i} = 0$。

2.6.4　模型之间的兼容性

数字孪生模型之间的兼容性是指 2 个模型的架构、参数和数据接口都一致, 可相互配合稳定工作, 直接进行数据传递和模型组合, 不存在频繁崩溃、卡死等问题。数字孪生模型之间的兼容程度的计算方法为

$$x_{21} = n'/N_{21} \tag{21}$$

式中: N_{21} 为数字孪生模型中所有模型的总数; n' 为能够与其他数字孪生模型兼容的子模型数量。可根据能够与其他数字孪生模型兼容的子模型数量占比情况, 对数字孪生模型之间的兼容性进行评价打分, 赋值范围为 [0,1], 分值越高, 表示模型关联关系的可编排程度越高。

3　基于群决策可变模糊识别的数字孪生模型评估

从各维度对数字孪生模型质量和应用能力进行准确评估, 需要选择合适的综合评估方法。现有综合评估方法主要分为 2 种, 一种是基于指标体系的评估方法, 主要通过建立评估指标与评估对象之间的关系, 再结合指标权重计算并采用模糊综合评估[16]、灰色评估[17]、集对分析[18] 等算法进行综合评估; 另一种是基于机器学习的评估方法[19-21], 需要建立在大量指标数据基础上, 通过提取指标数据特征, 并对模型进行训练实现综合评估。上述 2 种方法, 前者主要针对复杂系统结构特征, 存在指标权重确定难, 可解释性不强等问题; 后者在指标样本量不足情况下, 难以训练出性能良好的模型。

由于数字孪生模型评估指标大部分为定性指标, 获取指标数据较为困难, 需要根据专家经验进行评估, 而专家存在模糊性、不确定性等问题。对此, 本文采用第 1 种评估方法, 在考虑专家意见偏好基础上利用信息熵计算指标权重, 再采用群决策模型改进可变模糊识别模型构建数字孪生模型评估方法, 并运用该方法对数字孪生模型的质量和应

用能力进行综合评估，最后，根据综合评估结果给出评估结论。数字孪生模型评估流程如图1所示。

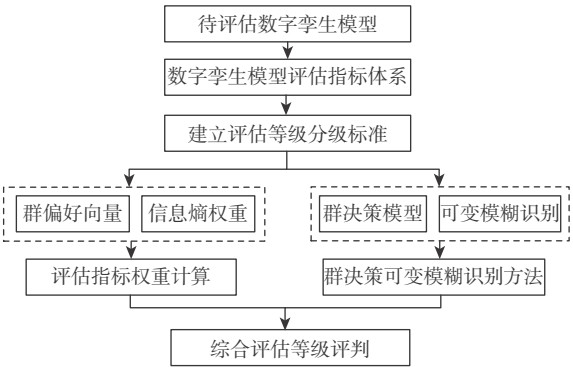


图1 数字孪生模型评估流程

Fig. 1 Assessment flow of digital twin model

3.1 基于信息熵集结法的指标权重计算

由于层次分析法、熵权法、灰色关联法等传统赋权法在计算权重时未考虑不同个体意见和偏好，对此本文结合群决策和信息熵计算指标权重，用于解决个体意见差异和不确定性[22]。该方法通过集结每位决策成员的意见和偏好，并基于信息熵概念计算决策成员的信息熵权重，最后，加权求和得到最终指标权重结果。具体如下：

1）构建专家决策矩阵 $\boldsymbol{H}$

$$\boldsymbol{H}=\left[h_{ijk}\right]_{n\times m} \tag{22}$$

式中：$i=1,2,\cdots,n;j=1,2,\cdots,m;k=1,2,\cdots,s$；$h_{ijk}$ 为第 k 个专家对 i 个目标下第 j 指标的评判结果。决策专家按照1～5标度法对各指标相对于上级目标的重要程度进行评判，其中，1为极不重要，2为不重要，3为一般重要，4为重要，5为极端重要。

2）构建标准化决策矩阵 $\boldsymbol{C}$

$$c_{ijk}=h_{ijk}\Big/\sum_{j=1}^{m}h_{ijk} \tag{23}$$

3）计算群偏好向量 $\boldsymbol{\alpha}$

$$\alpha_{ijk}=\prod_{k=1}^{s}\left(c_{ijk}\right)^{\delta_k}\Big/\sum_{j=1}^{m}\prod_{k=1}^{s}\left(c_{ijk}\right)^{\delta_k} \tag{24}$$

式中：δ_k 为专家的决策权重，取 $\delta_k=1/s$。

4）计算信息熵权重

信息熵权重计算方法是基于信息熵的概念，通过计算每个指标的信息熵来确定其权重。当指标离异程度较大时，该方法计算出的权重更准确客观。该方法的计算过程如下：

$$e_j=-\frac{1}{\ln s}\sum_{k=1}^{s}c_{ijk}\ln\left(c_{ijk}\right) \tag{25}$$

$$v_j=(1-e_j)\Big/\sum_{j=1}^{m}\left(1-e_j\right) \tag{26}$$

式中：e_j 为信息熵；v_j 为差异系数。

5）信息熵权重集结

在信息熵权重的基础上，考虑所有决策成员的意见偏好，求得各个指标的综合权重，设 β_j 为结合后的指标权重，定义为

$$\beta_j=\alpha_j v_j\Big/\sum_{j=1}^{m}\alpha_j v_j \tag{27}$$

3.2 改进可变模糊识别方法

本文构建的数字孪生模型评估指标需要通过专业人员凭经验进行量化。由于经验和知识的局限，单人指标赋值存在模糊性和不确定性，而传统的量化方法无法综合所有参与人员的经验。为克服此问题，本文结合群决策模型和可变模糊识别模型构建数字孪生模型综合评估方法。该方法通过三角模糊数量化指标赋值，解决人员思维的模糊性；通过集结群决策信息[23]，解决部分人员经验知识不足问题。具体步骤如下：

1）确立评语集。评语集 S 作为群决策准则，是由奇数个元素构成的可预先设定的有序语言变量集合。本文将指标重要性程度划分为5个等级，则评语集为：“不合格、待改进、合格、良好、优秀”，即 $S=\{S_1,S_2,S_3,S_4,S_5\}$。

2）对语言变量进行模糊量化。根据模糊数学理论，通过三角模糊数量化评价值大小。语言变量 $S_r(r=1,2,\cdots,5)$ 的三角模糊数表达式如式(28)所示。再根据式(28)，计算对应的三角模糊数，如表3所示。

$$\hat{d}=\left(d^{\mathrm{L}},d^{Z},d^{\mathrm{R}}\right)=\left(\max\left\{\frac{r-2}{4},0\right\},\frac{r-1}{4},\min\left\{\frac{r}{4},1\right\}\right) \tag{28}$$

表3 语言变量的三角模糊数

Table 3 Triangular fuzzy numbers of linguistic variables

S_r	三角模糊数
S_1	(0,0,0.25)
S_2	(0,0.25,0.5)
S_3	(0.25,0.5,0.75)
S_4	(0.5,0.75,1)
S_5	(0.75,1,1)

3）评判信息综合。对单个评判样本进行正规化处理：

$$\begin{cases}xl_{ij}^{P}=(l_{ij}^{P}-\min(l_{ij}^{P}))/(\max(u_{ij}^{P})-\min(l_{ij}^{P}))\\ xm_{ij}^{P}=(m_{ij}^{P}-\min(l_{ij}^{P}))/(\max(u_{ij}^{P})-\min(l_{ij}^{P}))\\ xu_{ij}^{P}=(u_{ij}^{P}-\min(l_{ij}^{P}))/(\max(u_{ij}^{P})-\min(l_{ij}^{P}))\end{cases} \tag{29}$$

式中：l_{ij}^{P}、m_{ij}^{P}、u_{ij}^{P} 分别为第 P 个专家赋值经三角模

糊数后的下界、中值、上界; xl_{ij}^{P}为三角模糊数下界标准化处理值; xm_{ij}^{P}为三角模糊数中值标准化处理值; xu_{ij}^{P}为三角模糊数上界标准化处理值。

计算正规化白化值:

$$\begin{cases} xls_{ij}^{P}=xm_{ij}^{P}/\left(1+xm_{ij}^{P}-xl_{ij}^{P}\right) \\ xus_{ij}^{P}=xu_{ij}^{P}/\left(1+xu_{ij}^{P}-xm_{ij}^{P}\right) \end{cases} \tag{30}$$

式中: xls_{ij}^{P}为下界三角模糊数白化值; xus_{ij}^{P}为上界三角模糊数白化值。

$$x_{ij}^{P}=\frac{\left[xls_{ij}^{P}\left(1-xls_{ij}^{P}\right)+xus_{ij}^{P}xus_{ij}^{P}\right)]}{\left(1-xls_{ij}^{P}+xus_{ij}^{P}\right)} \tag{31}$$

计算各个指标的白化值:

$$d_{ij}^{P}=\min\left(l_{ij}^{P}\right)+x_{ij}^{P}\left(\max\left(u_{ij}^{P}\right)-\min\left(l_{ij}^{P}\right)\right) \tag{32}$$

综合各评判信息:

$$q_{ij}=\frac{1}{m}\sum_{p=1}^{m}d_{ij}^{P} \tag{33}$$

4) 可变模糊识别模型。设综合评判信息矩阵 $\boldsymbol{Q}=[q_{ij}]$, i=1,2,⋯,m; j=1,2,⋯,n, 并设指标分级标准值矩阵 $\boldsymbol{Y}=[y_{ih}]$, h =1,2,⋯,c, 表示按照各指标的 l 个级别识别获得的标准值矩阵。将指标分为递增型和递减型 2 类, 通过式(34)～式(37)对 $\boldsymbol{Q}$ 和 $\boldsymbol{Y}$ 进行规格化处理[24], 得到规格化矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{S}$。

极大型指标:

$$r_{ij}=\begin{cases} 0 & q_{ij}\leqslant y_{il} \\ \dfrac{y_{il}-q_{ij}}{y_{il}-y_{i1}} & y_{il}<q_{ij}<y_{i1} \\ 1 & q_{ij}\geqslant y_{i1} \end{cases} \tag{34}$$

$$s_{ih}=\begin{cases} 0 & y_{ih}=y_{il} \\ \dfrac{y_{il}-y_{ih}}{y_{il}-y_{i1}} & y_{il}<y_{ih}<y_{i1} \\ 1 & y_{ih}=y_{i1} \end{cases} \tag{35}$$

极小型指标:

$$r_{ij}=\begin{cases} 0 & q_{ij}\geqslant y_{il} \\ \dfrac{y_{il}-q_{ij}}{y_{il}-y_{i1}} & y_{i1}<q_{ij}<y_{il} \\ 1 & q_{ij}\leqslant y_{i1} \end{cases} \tag{36}$$

$$s_{ih}=\begin{cases} 0 & y_{ih}=y_{il} \\ \dfrac{y_{il}-y_{ih}}{y_{il}-y_{i1}} & y_{i1}<y_{ih}<y_{il} \\ 1 & y_{ih}=y_{i1} \end{cases} \tag{37}$$

式中: y_{i1}、y_{ih}和y_{il}分别为指标 i 的第 1 级、第 h 级和第 l 级标准特征值; r_{ij}为样本 j 对应指标 i 规格化值; s_{ih}为级别 h 对应指标 i 的规格化参数。

对综合评判信息, 按照构建的标准值矩阵 $\boldsymbol{S}$ 进行识别, 建立可变模糊识别模型[25], 并由式(38)计算相对隶属度矩阵 $\boldsymbol{U}=[u_{hj}]_{l\times n}$。

$$u_{hj}=\begin{cases} \left\{\sum_{k=a_j}^{b_j}\left[\dfrac{\sum_{i=1}^{m}\left(w_i\left|r_{ij}-s_{ih}\right|\right)^{p}}{\sum_{i=1}^{m}\left(w_i\left|r_{ij}-s_{ik}\right|\right)^{p}}\right]^{\frac{a}{p}}\right\}^{-1} & a_j\leqslant h\leqslant b_j \\ 0 & 1\leqslant h<a_j\text{或}b_j<h\leqslant l \end{cases} \tag{38}$$

式中: w_i 为指标 i 的权重; a、p 称为可变模糊模型参数, a 为优化标准参数, p 为距离参数(a 和 p 通常有 4 种组合: 当 a=1, p=1 时, 为模糊综合评价模型; 当 a=1, p =2 时, 为理想模型; 当 a=2, p=1 时, 为神经元激励函数模型; 当 a=2, p=2 时, 为模糊评价优选模型); a_j 和 b_j 分别为第 j 个样本的级别下限值和上限值。

根据应用级别特征值模型, 由式(39)计算出等级特征值 H_j 及其平均值$\overline{H}$, 再根据式(40)进行等级综合判断。

$$[H_j]=[1,2,\cdots,c]\cdot[u_j] \tag{39}$$

$$\begin{cases} 1.0\leqslant\overline{H}\leqslant 1.5 & \text{归属1级} \\ h-0.5<\overline{H}\leqslant h & \text{归属于}h\text{级, 偏}(h-1)\text{级,}\ h=2,3,\cdots,c-1 \\ h<\overline{H}\leqslant h+0.5 & \text{归属于}h\text{级, 偏}(h+1)\text{级,}\ h=2,3,\cdots,c-1 \\ c=0.5\leqslant\overline{H}\leqslant c & c\text{级} \end{cases} \tag{40}$$

4　实例验证

以某航空制造企业正在构建的“飞机制造工厂”为研究对象。该“工厂”是通过数字孪生模型技术, 将现实飞机生产过程与虚拟模型进行同步映射和分析, 实现工厂信息的全面透明化和可视化管理, 并以实时数据、仿真数据和历史数据等多种数据源的驱动, 实现对生产过程的全面监控和分析。对于数字孪生模型的构建, 业内主要遵循精准刻画、实时连接、轻量高效等通用准则, 同时模型的构建及表现形式包含对物理实体的抽象、映射、状态感知、反馈控制及可视化交互、应用服务连接等基本过程, 具体如图 2 所示。因此, 对“飞机制造工厂”数字孪生模型的评估, 也需要在构建准则的基础上, 通过可量化的评估指标, 以对模型构建过程质量及应用能力进行评估。本文根据各指标量化计算式, 制作评估问卷, 通过收集专家评价意见, 采用改进可变模糊识别方法对“飞机制造工厂”数字孪生模型质量和应用能力进行评估。

4.1　建立评估等级分级标准

为对“飞机制造工厂”数字孪生模型质量和应用能力进行分级评估, 将评估结果分为 5 个等级: S_1 为不合格, S_2 为待改进, S_3 为合格, S_4 是为良好,

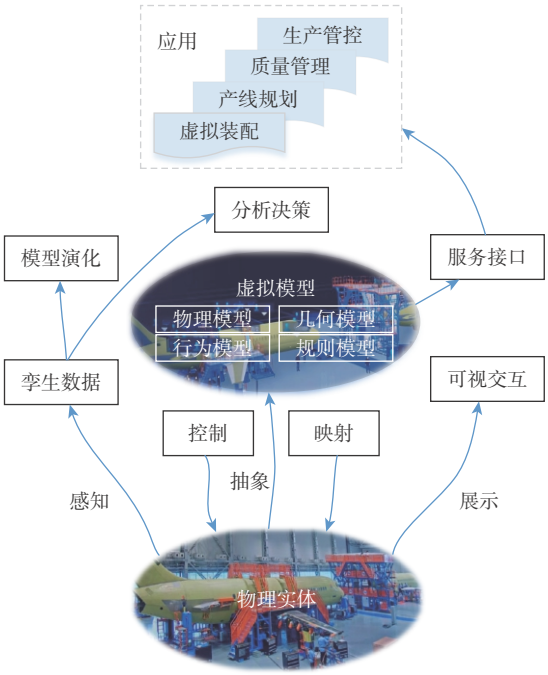


图2 数字孪生模型基本要素及构建过程

Fig. 2 Basic elements and construction process of digital twin models

S_5为优秀，每个评估等级描述如表4所示。

4.2 数据收集和处理

对数字孪生模型质量和应用能力进行评估，可帮助设计、使用者优化和选择合适的模型。首先，本文通过代码查看、实际运行测试等软件测试方法对“飞机制造工厂”数字孪生模型进行了功能、性能等测试，获得各个指标量化公式所需的评估参考值；然后，利用指标量化公式确定指标量化值；最后，设计调查问卷，邀请具有丰富数字孪生技术经验的8位专家组成评估决策小组，由专家根据数字孪生系统测试结果和数字孪生模型设计相关要求，依据表3中的语言变量，给出各个指标评价结果，并采用模糊量化评语集对“飞机制造工厂”数字孪生模型各个指标进行量化赋值。以“准确程度”为例，通过软件测试方法获得的数据，如表5所示。最终，8位专家给出的各个指标评估结果如表6所示。

表4 数字孪生模型评估等级判定标准

Table 4 Evaluation grade criteria of digital twins model

评估等级	等级描述
S_5	数字孪生模型运行效率很高，能够与物理实体、孪生数据、服务等精准连接，具有很强的可解释性，以及很高的规范化程度，能够快速重构优化和迁移复用
S_4	数字孪生模型运行效率较高，能够与物理实体、孪生数据、服务等较好连接，且可解释性较强，规范化程度较高，能够较好重构优化和迁移复用
S_3	数字孪生模型能够稳定运行，准确度、连接程度、可视化、规范化等基本满足需求，可重构优化和迁移复用
S_2	数字孪生模型能够正常运行，但准确度、连接稳定性、可视化、规范化等表现一般，重构优化、迁移复用率较低
S_1	数字孪生模型基本能够实现运行，但准确度、连接稳定性、可视化等较差，规范化程度较低，无法重构优化和迁移复用

表5 “准确程度”各指标测试数据

Table 5 Each indicator test data of “accuracy”

评估指标	测试数据	量化计算值	某专家评估
x_8	$x_{8i}=1\,440$ $x'_{8i}=921$	$x_8=0.36$	S_3
x_9	$x_{9i}=415$ ms $x'_{9i}=175$ ms	$x_9=0.58$	S_4
x_{10}	$x_{10i}=530$ ms $x'_{10i}=210$ ms	$x_{10}=0.60$	S_4

表6 评估数据统计

Table 6 Evaluation data statistics sheet

编号	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}
1	S_5	S_4	S_3	S_2	S_5	S_4	S_4	S_5	S_4	S_5	S_4	S_4	S_3	S_4	S_4	S_3	S_5	S_4	S_4	S_4	S_5
2	S_4	S_4	S_4	S_3	S_2	S_5	S_4	S_5	S_5	S_5	S_3	S_2	S_3	S_5	S_4	S_4	S_4	S_5	S_5	S_4	S_5
3	S_4	S_5	S_5	S_4	S_4	S_4	S_4	S_4	S_5	S_4	S_4	S_4	S_4	S_4	S_4	S_4	S_4	S_5	S_4	S_3	S_5
4	S_5	S_4	S_3	S_2	S_1	S_3	S_5	S_3	S_4	S_4	S_5	S_4	S_4	S_3	S_5	S_3	S_4	S_5	S_5	S_4	S_4
5	S_4	S_4	S_4	S_3	S_3	S_4	S_4	S_4	S_5	S_4	S_5	S_5	S_3	S_4	S_4	S_4	S_5	S_5	S_4	S_5	S_5
6	S_4	S_5	S_4	S_5	S_5	S_5	S_4	S_2	S_5	S_4	S_4	S_3	S_2	S_4	S_5	S_5	S_5	S_4	S_3	S_4	S_4
7	S_4	S_4	S_4	S_5	S_4	S_4	S_4	S_5	S_5	S_5	S_3	S_4	S_4	S_3	S_3	S_5	S_4	S_5	S_5	S_5	S_5
8	S_5	S_3	S_3	S_5	S_4	S_5	S_3	S_4	S_5	S_4	S_5	S_5	S_5	S_5	S_4	S_3	S_2	S_5	S_4	S_5	S_5

4.3 数字孪生模型评估

4.3.1 指标权重计算

决策专家按照1～5标度法，对数字孪生模型的“准确程度、连接程度、高效程度、可视化程度、可编编排程度及规范化程度”下各指标相对于上级目标的重要程度进行评判，并构建专家群决策矩阵。以“准确性”目标为例，专家群决策的标准化矩阵如表7所示，其中，k_1～k_8表示8位参与评估的专家。

根据式(3)可计算8位专家的群偏好向量，并由式(25)和式(26)计算出基于信息熵的指标权重。最后，集结信息熵权重和决策成员意见偏好，计算得到各个指标的综合权重，如表8所示。

表 7　专家群决策标准化矩阵

Table 7　Expert group decision standardization matrix

专家	x_1	x_2	x_3
k_1	0.128	0.135	0.148
k_2	0.128	0.135	0.111
k_3	0.128	0.135	0.148
k_4	0.128	0.108	0.111
k_5	0.128	0.135	0.148
k_6	0.128	0.135	0.111
k_7	0.128	0.108	0.074
k_8	0.102	0.108	0.148

表 8　各评估指标权重

Table 8　Weight of each evaluation index

指标	权重	指标	权重
x_1	0.015	x_{12}	0.035
x_2	0.017	x_{13}	0.012
x_3	0.078	x_{14}	0.115
x_4	0.085	x_{15}	0.017
x_5	0.090	x_{16}	0.042
x_6	0.032	x_{17}	0.095
x_7	0.053	x_{18}	0.029
x_8	0.035	x_{19}	0.032
x_9	0.053	x_{20}	0.078
x_{10}	0.029	x_{21}	0.016
x_{11}	0.042		

4.3.2　综合评估计算

根据评估问卷收集到的 8 名专家打分数据, 由三角模糊数进行量化后, 按照式(29)～式(33)计算步骤, 计算综合评判信息。得到评判信息的综合值 $\boldsymbol{Q}$ = [0.822, 0.763, 0.675, 0.646, 0.793, 0.734, 0.733, 0.907, 0.821, 0.765, 0.705, 0.618, 0.735, 0.765, 0.704, 0.764, 0.907, 0.792, 0.532, 0.793, 0.907]$^{\mathrm{T}}$。

根据智能制造数字孪生评估指标特征, 设定所有指标分级标准为 [0, 0.25, 0.5, 0.75, 1]$^{\mathrm{T}}$, 并建立指标分级标准值矩阵 $\boldsymbol{Y}$=[y_{ih}], 再由式(35)计算得出指标标准的格式化矩阵 $\boldsymbol{S}$。

$$\boldsymbol{S}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.25 & \cdots & 0.25 & 0.25 & 0.25 \\ 0.50 & 0.50 & 0.50 & 0.50 & 0.50 & \cdots & 0.50 & 0.50 & 0.50 \\ 0.75 & 0.75 & 0.75 & 0.75 & 0.75 & \cdots & 0.75 & 0.75 & 0.75 \\ 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 \end{bmatrix}$$

根据式(34)对评判信息的综合值进行规格化, 得到矩阵 $\boldsymbol{R}$ = [r_{hj}]$_{1\times 23}$=[0.822, 0.763, 0.675, 0.646, 0.793, 0.734, 0.733, 0.907, 0.821, 0.765, 0.705, 0.618, 0.735, 0.765, 0.704, 0.764, 0.907, 0.792, 0.532, 0.793, 0.907]$^{\mathrm{T}}$。

根据式(38)计算出的相对隶属度, 得到相对隶属度矩阵 $\boldsymbol{U}$, 如表 9 所示。

表 9　不同 a、p 值下的相对隶属度

Table 9　Relative membership for different a, p values

a	p	u_1	u_2	u_3	u_4	u_5
1	1	0.051	0.077	0.151	0.560	0.161
1	2	0.058	0.086	0.160	0.517	0.179
2	1	0.007	0.016	0.061	0.846	0.070
2	2	0.010	0.022	0.077	0.797	0.095

最后, 根据式(39)的模糊识别模型计算级别特征值及$\overline{H}$, 如表 10 所示, 并根据式(40)的判断准则对安全风险等级进行识别。

表 10　不同 a、p 值下的特征值

Table 10　Eigenvalues for different a, p values

a	p	H_j	$\overline{H}$的相对误差的绝对值/%
1	1	3.703	1.96
1	2	3.673	2.76
2	1	3.956	4.72
2	2	3.945	4.43

4.3.3　评估结果分析

由表 9 可知, 在不同优化准则参数 a 和距离参数 p 下, 计算得到的飞机制造工厂数字孪生模型评估等级特征值不同。但是, 不同等级特征值 H_j 与 $\overline{H}$相对误差的绝对值最大为 4.72%, 最小为 1.96%, 说明计算结果的稳定性趋于良好。同时, 参考模式判断准则, 该智能制造数字孪生模型评估等级$\overline{H}$为 3.777, 在 3.5 <$\overline{H}$< 4 之间, 归属于 4 级, 但偏向于 3 级, 即评估等级为“S_4 良好”, 偏向于“合格”。这意味着该数字孪生模型能够较好的模拟真实飞机制造厂的运行情况, 并为智能化管理提供决策支持和优化建议, 帮助企业降低成本, 提高生产效率。

尽管评估等级为“良好”, 但仍有进一步改进的空间。根据现场验证情况来看, 当描述物理实体状态的数字孪生模型参数发生变化时, 模型无法准确的随动变化; 在执行应用服务期间, 数字孪生模型的几何、物理、行为、规则等模型占用了系统较大资源, 还需要进一步对模型进行轻量化处理, 并提高模型的响应速度和运行效率; 飞机制造工厂是一个高度复杂的系统, 涉及多个工艺和环节, 数字孪生模型的输出结果并未全面覆盖, 甚至缺少部分重要环节。为提高该数字孪生模型的性能和应用效果, 建议进一步对模型进行验证和优化, 包括使用历史数据进行模型验证, 并与实际生产情况进行比较等手段。

5 结 论

1）针对当前智能制造数字孪生模型缺乏统一的评估标准，采用GQM方法构建了一套完整的智能制造数字孪生模型评估指标体系。该指标体系包含准确程度、连接程度、高效程度、可视化程度、可编排程度及规范化程度6个维度，共计23个指标，可作为模型构建、应用、采购参考用。

2）为提高评估指标量化赋值的客观性，本文对每个指标设计了具体的量化计算式，并给出了指标的取值范围，以便专家在对智能制造数字孪生模型进行评估时，能够更加的客观准确。

3）为解决专家评价的模糊性、不确定性等问题，利用群决策理论改进可变模糊识别模型构建了基于群决策可变模糊识别的数字孪生模型评估方法，并在考虑专家意见偏好基础上采用信息熵计算指标权重，为智能制造数字孪生模型量化评估提供算法支撑。

4）利用改进的群决策可变模糊识别模型对“飞机制造工厂”数字孪生模型进行综合评估，评估结果显示该数字孪生模型的评估等级为“S_4良好”，但偏向于“合格”，并结合现场验证检查情况，进一步验证了该评估方法的可行性和有效性。

由于当前软件测试方法和手段限制，部分指标不能通过数据统计方式进行量化评估，只能依赖专家经验判断评估，因此，未来将结合最新软件测试方法，改进指标的量化计算公式，提高评估的准确性。

参考文献（References）

[1] Siemens. The digital twin[M]. Switzerland : Springer Nature , 2015.

[2] 陶飞, 张贺, 戚庆林, 等. 数字孪生模型构建理论及应用[J]. 计算机集成制造系统, 2021, 27(1): 1-15.
TAO F, ZHANG H, QI Q L, et al. Theory of digital twin modeling and its application[J]. Computer Integrated Manufacturing Systems, 2021, 27(1): 1-15(in Chinese).

[3] 陶飞, 刘蔚然, 张萌, 等. 数字孪生五维模型及十大领域应用[J]. 计算机集成制造系统, 2019, 25(1): 1-18.
TAO F, LIU W R, ZHANG M, et al. Five-dimension digital twin model and its ten applications[J]. Computer Integrated Manufacturing Systems, 2019, 25(1): 1-18(in Chinese).

[4] BERRI P C, DALLA VEDOVA M L. A review of simplified servovalve models for digital twins of electrohydraulic actuators[J]. Journal of Physics: Conference Series, 2020, 1603(1): 012016.

[5] 陆剑峰, 夏路遥, 张浩, 等. 制造企业数字孪生生态系统的研究与应用[J]. 计算机集成制造系统, 2022, 28(8): 2273-2290.
LU J F, XIA L Y, ZHANG H, et al. Research and application of manufacturing enterprise digital twin ecosystem[J]. Computer Integrated Manufacturing Systems, 2022, 28(8): 2273-2290(in Chinese).

[6] WANG L, LI Y, ZHANG X. A digital twin-based traffic flow prediction method for urban road network[J]. IEEE Access, 2020(8): 132190-132200.

[7] CHEN J, WANG Y. A real-time monitoring and prediction method for industrial manufacturing process based on deep neural network digital twin model[J]. IEEE Access, 2021(9): 38777-38787.

[8] LI J, WANG L. An explainable machine learning-based digital twin model for industrial manufacturing process[J]. IEEE Access, 2021(9): 104713-104722.

[9] 张辰源, 陶飞. 数字孪生模型评价指标体系[J]. 计算机集成制造系统, 2021, 27(8): 2171-2186.
ZHANG C Y, TAO F. Evaluation index system for digital twin model[J]. Computer Integrated Manufacturing Systems, 2021, 27(8): 2171-2186(in Chinese).

[10] 李翔宇. 数字孪生模型构造技术研究[D]. 沈阳: 沈阳理工大学, 2022.
LI X Y. Research on construction technology of digital twin model[D]. Shenyang: Shenyang Ligong University, 2022(in Chinese).

[11] LUO R P, SHENG B Y, LU Y K, et al. Digital twin model quality optimization and control methods based on workflow management[J]. Applied Sciences, 2023, 13(5): 2884.

[12] 国家市场监督管理总局, 国家标准化管理委员会. 自动化系统与集成 复杂产品数字孪生体系架构: GB/T 41723—2022[S]. 北京: 中国标准出版社.
State Administration for Market Regulation, Standardization Administration of the People's Republic of China. Automation systems and integration——Digital twin architecture of complex product: GB/T 41723—2022[S]. Beijing: Standards Press of China (in Chinese).

[13] 赵强. 数字孪生技术应用指: T/CSPSTC 74-202 [S]. 北京: 中国标准出版社, 2021.
ZHAO Q. Guidance of digital twin technology: T/CSPSTC 74-202 [S]: Beijing: Standards Press of China, 2021(in Chinese).

[14] 杜洪涛. 工业数字孪生系统评估 第2部分: 应用成熟度评估: T/CAMETA 001009.2-2023[S]. 北京: 中国标准出版社, 2023.
DU H T. Assessment of industrial digital twin systems——Part 2 Evaluation to maturity of application: T/CAMETA 001009.2-2023[S]. Beijing: Standards Press of China, 2023.

[15] LI Y R, XU M Y, WEN X, et al. The role of Internet search index for tourist volume prediction based on GDFM model[J]. Tehnicki Vjesnik-Technical Gazette, 2020, 27(2): 576-582.

[16] 张军涛, 李尚生, 王旭坤. 基于灰色关联-模糊综合评判的雷达抗干扰性能评估方法[J]. 系统工程与电子技术, 2021, 43(6): 1557-1563.
ZHANG J T, LI S S, WANG X K. Method of radar anti-jamming performance evaluation based on grey correlation-fuzzy comprehensive evaluation[J]. Systems Engineering and Electronics, 2021, 43(6): 1557-1563(in Chinese).

[17] 高扬, 王向章. 基于灰色模糊层次分析法的航空公司外委维修风险评估[J]. 安全与环境工程, 2016, 23(2): 80-84.
GAO Y, WANG X Z. Risk assessment of air carriers' maintenance outsourcing based on the gray fuzzy analytic hierarchy process[J]. Safety and Environmental Engineering, 2016, 23(2): 80-84(in

Chinese).

[18] 高扬, 王向章. 基于 SPA-Markov 的飞行安全态势评估与预测研究[J]. 中国安全生产科学技术, 2016, 12(8): 87-91.
GAO Y, WANG X Z. Research on assessment and prediction of flight safety situation based on SPA-Markov[J]. Journal of Safety Science and Technology, 2016, 12(8): 87-91(in Chinese).

[19] 陈农田, 满永政, 李俊辉. 基于 QAR 数据的民机高高原进近着陆风险评估方法[J]. 北京航空航天大学学报, 2024, 50(1): 77-85.
CHEN N T, MAN Y Z, LI J H. Risk assessment method for civil aircraft approach and landing at high plateau based on QAR data[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(1): 77-85(in Chinese).

[20] 冯朝辉. 基于机器学习的终端区航空器运行间隔安全评估方法研究[D]. 天津: 中国民航大学, 2022.
FENG Z H. Research on safety assessment method for aircraft operation interval in terminal zone based on machine learning[D]. Tianjin: Civil Aviation University of China, 2022(in Chinese).

[21] 李常刚, 李华瑞, 刘玉田, 等. 大电网动态安全风险智能评估系统[J]. 电力系统自动化, 2019, 43(22): 67-75.
LI C G, LI H R, LIU Y T, et al. Intelligent assessment system for dynamic security risk of large-scale power grid[J]. Automation of Electric Power Systems, 2019, 43(22): 67-75(in Chinese).

[22] 陈华友, 刘春林. 群决策中基于不同偏好信息的相对熵集成方法[J]. 东南大学学报 (自然科学版), 2005, 35(2): 311-315.
CHEN H Y, LIU C L. Relative entropy integration method based on different preference information in group decision making[J]. Journal of Southeast University (Natural Science Edition), 2005, 35(2): 311-315(in Chinese).

[23] 黄景文. 信息安全风险因素分析的模糊群决策方法研究[J]. 山东大学学报 (理学版), 2012, 47(11): 45-49.
HUANG J W. Fuzzy group decision making for information security risk factors analysis[J]. Journal of Shandong University (Natural Science), 2012, 47(11): 45-49(in Chinese).

[24] 王娜娜. 基于可变模糊模型的非毗邻地区经济协调发展研究: 以中俄豫伊两地为例[J]. 数学的实践与认识, 2021, 51(1): 40-46.
WANG N N. Research on coordinated economic development in non-adjacent areas based on variable fuzzy model: take China, Russia, Henan and Iraq as examples[J]. Mathematics in Practice and Theory, 2021, 51(1): 40-46(in Chinese).

[25] 陈勇刚, 熊升华, 贺强, 等. 通用航空机队设备可靠性动态识别模型[J]. 航空学报, 2018, 39(3): 172-183.
CHEN Y G, XIONG S H, HE Q, et al. Dynamic recognition method for reliability of general aviation fleet equipment[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(3): 172-183(in Chinese).

Variable fuzzy comprehensive evaluation for intelligent manufacturing digital twin model

WANG Xiangzhang[1, 2, *], YU Lili[2]

(1. School of transportation and logistics, Southwest Jiaotong University, Chengdu 610031, China;
2. GRG Metrology & Test Group Co., Ltd., Guangzhou 510630, China)

Abstract: A systematic multidimensional assessment index system for digital twin models was created utilizing the goal-question-metric (GQM) method in response to the dearth of reference standards and unified evaluation techniques for intelligent manufacturing digital twin models. By combining the advantages of variable fuzzy recognition model and information entropy aggregation weight algorithm, a digital twin quality value evaluation method based on improved variable fuzzy model was constructed. In order to solve the problems of fuzziness and uncertainty in expert evaluation, the variable fuzzy recognition model is improved by using group decision theory, and the index weight is calculated by using information entropy on the basis of considering expert opinion preference. Finally, the quality, performance and value of a digital twin model of an aircraft manufacturing plant are evaluated. According to the example study, the digital twin model of the aircraft production facility has an evaluation level of "S_4 good," although it tends to be "S_3 qualified," and there is still opportunity for improvement. At the same time, the feasibility and rationality of improving the variable fuzzy recognition model are verified, and the method support is provided for the construction of standardized intelligent manufacturing digital twin model.

Keywords: digital twin model; GQM method; information entropy; group decision; variable fuzzy recognition

Received: 2023-10-31; **Accepted**: 2023-12-20; **Published Online**: 2024-01-03 15:07
URL: link.cnki.net/urlid/11.2625.V.20240102.1526.002

* **Corresponding author**. E-mail: 513432965@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0738

高空大气风切变时空分布特征

杨钧烽[1,2]，程旋[3]，王建美[1,2]，张依鸣[1,2]，胡雄[1,2]，肖存英[4,5,*]

(1. 中国科学院国家空间科学中心 太阳活动与空间天气重点实验室，北京 100190；2. 中国科学院国家空间科学中心 中国科学院空间环境态势感知技术重点实验室，北京 100190；3. 深空探测实验室，北京 100041；4. 北京师范大学 天文与天体物理前沿科学研究所，北京 100875；5. 北京师范大学 天文系，北京 100875)

摘 要：火箭发射阶段容易受到高空大气风场影响，火箭设计需要考虑风切变特性。基于2000—2022年美国国家环境预报中心（NCEP）再分析资料，利用综合矢量风的方法整体分析了东亚区域地面至30 km的高空大气风切变特征，并结合大气环流和高空急流分析了其时空分布成因。研究表明：最多风向、最大风速和风切变都具有显著的季节、高度和水平变化。利用20年的台站气球探空数据进行了差异性分析，再分析资料低估了实际风切变强度。将计算的高空大气风切变应用到两型火箭发射段风攻角的评估。结果表明：风切变在5 km以下和7～18 km高度影响显著。研究结果有助于全面掌握中国风切变分布特征，为火箭风场设计提供参考。

关 键 词：风切变；综合矢量风；高空急流；火箭设计；攻角
中图分类号：V411.8
文献标志码：A **文章编号**：1001-5965(2026)01-0192-11

运载火箭作为空间探测和太空探索的主要工具，成为了许多国家的研究重点，其运载水平和发展规模越来越成为衡量国家航天水平的标准[1]。近年来，世界主要航天强国相继推出了新一代运载火箭，如美国的宇宙神5、德尔塔4，欧洲空间局的阿里安5和日本的H3等[2]。中国航天经过几十年的探索和发展，逐渐形成了自主的长征系列运载火箭，满足了中国近地到深空各种轨道发射任务需求。为航天事业可持续发展，中国仍然在研制与需求匹配的各型运载火箭[3]。

火箭的安全飞行需要考虑大气环境因素复杂变化的影响[4]。大气参量（密度、温度、气压和风场等）变化剧烈，具有随机性和复杂性，影响气动力和气动热环境，给飞行器的弹道规划、气动设计和精确制导与控制带来不确定性[5]。陈闽慷等[6]利用美国热层、电离层、中间层能量学和动力学卫星（thermosphere ionosphere mesosphere energetics and dynamics, TIMED）搭载的宽带发射辐射测量大气探测仪载荷（sounding of the atmosphere using broadband emission radiometry, SABER）探测的环境数据，研究了北半球大气参数随纬度和季节变化对临近空间飞行器气动加热的影响，发现大气波动对热环境有着重要和复杂的影响。程旋等[7]利用TIMED/SABER大气密度观测数据定量分析了全球大气密度时空变化及其对高速飞行器飞行热环境的影响，表明大气扰动状态直接影响飞行器气动加热。程胡华等[8]分析了参考大气密度与真实大气密度的差异特征及对运载火箭最大气动载荷$q\alpha_{\max}$的影响，认为采用真实大气密度分析大气环境对运载火箭飞行影响比参考大气密度更好。针对大气密度的需求，中国学者开展了大量的建模研究[9-11]。

除了大气密度以外，火箭穿过高空大气时，会受到较大的高空风载作用。风载作用可以影响火箭发射精度，威胁安全飞行[12]。火箭设计都需要高

收稿日期：2023-11-09；**录用日期**：2023-12-22；**网络出版时间**：2024-02-02 10:47
网络出版地址：link.cnki.net/urlid/11.2625.V.20240201.1015.002
基金项目：国家自然科学基金(12241101,42174192,11872128)；中国科学院国家空间科学中心“攀登计划”项目
*通信作者. E-mail：xiaocunying@bnu.edu.cn

引用格式：杨钧烽，程旋，王建美，等. 高空大气风切变时空分布特征[J]. 北京航空航天大学学报，2026，52(1)：192-202.
YANG J F，CHENG X，WANG J M，et al. Spatiotemporal distribution characteristics of high-altitude wind shear[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：192-202 (in Chinese).

空风场和切变信息来优化控制系统[13-15]。高空大气风切变对飞行器飞行会产生攻角。风攻角能产生干扰力矩和气动力干扰力, 对姿态控制、姿态调整和制导系统产生影响, 导致实际飞行弹道与标准弹道出现偏差, 也会增加运载火箭的载荷, 对箭体结构产生影响。余梦伦[15]介绍了 CZ-2E 火箭发射中采用的高空风场弹道修正设计方法, 该方法有效降低了火箭飞行的气流攻角。程胡华等[12]分析了高空风预报值与实况风场的偏差对火箭最大气动载荷参数精度的影响。王建明等[16]研究了高空风场和偏差包络对某大型火箭飞行$q\alpha_{max}$值的影响。胡瑞光等[17]提出基于最少参数神经网络的拟合方法在线进行气动风场拟合与预示, 有效提高了拟合精度。

赵人濂等[18]提出利用综合矢量风剖面方法把高空最大风矢量和最大风切变引入到运载火箭设计中。这种根据数理统计理论的风场设计方法采用大子样、大尺度高空风数据的统计分析, 既能真实反映实测数据, 保证危险条件设计, 又不无限保守约束。文献 [19-20] 也分别计算了国内外站点上临近空间 20 km 以上高度风切变特性。

目前风切变研究主要集中在发射场上空, 但是大气风场的剧烈变化体现在时间和空间上, 尤其在中国上空, 受到青藏高原的影响, 高空急流具有独特性[21]。中国主要卫星发射中心高空大气风场受其影响, 具有显著的地域和季节差异, 因此, 分析高空风切变在东亚区域的时空变化规律对深入分析火箭风场设计具有重要的意义[21]。本文利用 2000—2022 年共计 23 年美国国家环境预报中心 (National Centers for Environment Prediction, NCEP) 再分析资料研究中国及周边区域(0～60°N, 70°E～140°E)地面到 30 km 的大气风切变特征, 分析其与大气环流的关系。利用气象站气球探空资料作为实测资料计算得到的风切变, 与 NCEP 资料进行比较。最后, 利用计算得到的高空大气风切变初步分析了其对两型火箭发射段的攻角影响。

1　数据资料和分析方法

1.1　NCEP 再分析资料

开展大气环境研究需要高质量且连续的全球大气资料, 但是站点观测资料时空分布不均匀, 尤其是在地形复杂区域(如沙漠、高原等)较稀疏等缺点, 不能满足研究的需要。因此, 本文采用再分析资料, 可以弥补这些缺陷。NCEP 再分析资料由 NCEP 和美国国家大气研究中心(National Center for Atmospheric Research, NCAR)联合开发的全球大气再分析计划, 采用全球地面、船舶、无线电探空、测风气球、飞机和卫星等多源数据, 利用三维变分同化技术生成的产品, 时间覆盖 1948 年至今[22]。其时间分辨率为 6 h, 水平分辨率为 2.5°(纬度)×2.5°(经度), 高度覆盖地面到 10 hPa(约 32 km), 高度分层采用气压层。NCEP 再分析资料是目前应用最为频繁的全球再分析资料之一, 被广泛应用于中国区域的气候和天气变化[23-24], 因此, 该数据可有效帮助理解东亚区域风场特性。

本文选取 2000—2022 年共 23 年数据, 将纬向风和经向风利用式(1)从气压坐标系转为海拔高度层数据。风场随高度线性变化, 采用线性插值方法计算得到地面到 30 km, 间隔 1 km 的各层纬向风和经向风。

$$Z_g=\frac{H\gamma(\varphi)g_0}{\gamma(\varphi)g(\varphi)-Hg_0} \tag{1}$$

式中: Z_g为几何高度; H为位势高度; g_0取值 9.806 65 m/s^2; φ为纬度; $\gamma(\varphi)$为纬度φ处地球半径, 表示为

$$\gamma(\varphi)=\frac{6\,378\,137}{1.006\,803-0.006\,706\sin^2\varphi} \tag{2}$$

$g(\varphi)$为纬度φ处重力加速度, 表示为

$$g(\varphi)=9.780\,325\times\frac{1+0.001\,931\,85\sin^2\varphi}{(1-0.006\,694\,35\sin^2\varphi)^2} \tag{3}$$

海拔高度Z_s计算式为

$$Z_s=Z_g+Z_e \tag{4}$$

式中: Z_e为大地水准面高度。

1.2　气球探空资料

世界各地的气象站每天进行 2 次气球探空探测, 气球携带无线电探空仪可以探测风速、风向、温度和湿度的数据[25]。每天 2 次气球探测分别在世界时的 00:00 时和 12:00 时左右实施。探空气球数据一般可以覆盖地面到 25～30 km, 但上边界取决于气球爆破高度。利用风速和风向数据得到纬向风和经向风, 同样采用线性插值方法计算得到间隔 1 km 的各高度层纬向风和经向风数据。本文选取 2000—2020 年气象站气球探空数据作为实测数据分析。

1.3　综合矢量风剖面

以某高度为参考高度, 首先, 计算最多风向, 给出一定出现概率下的最大风速。最大风速对其他各高度上产生风切变, 由切变造成的各高度的风速称为“条件风速”。这些风速沿高度分布构成了综合矢量风剖面[21]。该算法使用的前提是风场统计满足二维正态分布的假设条件, 如果一段时间存在双峰或多峰分布, 则会导致风切变计算结果不合

理，该算法不成立。

高空风切变能够对火箭飞行产生风攻角：

$$\begin{cases} T = T_{\mathrm{wp}} + T_{\mathrm{wq}} \\ T_{\mathrm{wp}} = \arctan\left(W_{\mathrm{p}}/V\right) \\ T_{\mathrm{wq}} = \arctan\left(W_{\mathrm{q}}/V\right) \end{cases} \tag{5}$$

式中：T为风攻角；T_{wp}和T_{wq}分别为条件最小风速W_{p}和条件风切变风速W_{q}引起的风攻角；V为飞行器飞行速度。

2 结 果

2.1 NCEP数据风切变特性

风向是指风的来向，北风为0°，顺时针旋转，90°、180°和270°分别代表东风、南风和西风。以中国陆地上空为例，图1为110°E不同纬度（20°N、30°N、40°N和50°N）上空地面到30 km的风向分布。风向在冬季1月份各纬度基本一致，都是盛行西风（概率密度峰值在270°附近）。从20°N也能看出一些差异，在25 km高度以上也出现东风（90°附近出现峰值）。风向在夏季7月份40°N和50°N基本一致，以20 km为界，20 km以下高度盛行西风，20 km以上高度盛行东风。夏季20°N和30°N上空，20 km以上高度风向也盛行东风，但是20 km以下高度略有不同，20°N从地面西南风逆时针转为高层东风，30°N则从地面西南风顺时针转为高层东

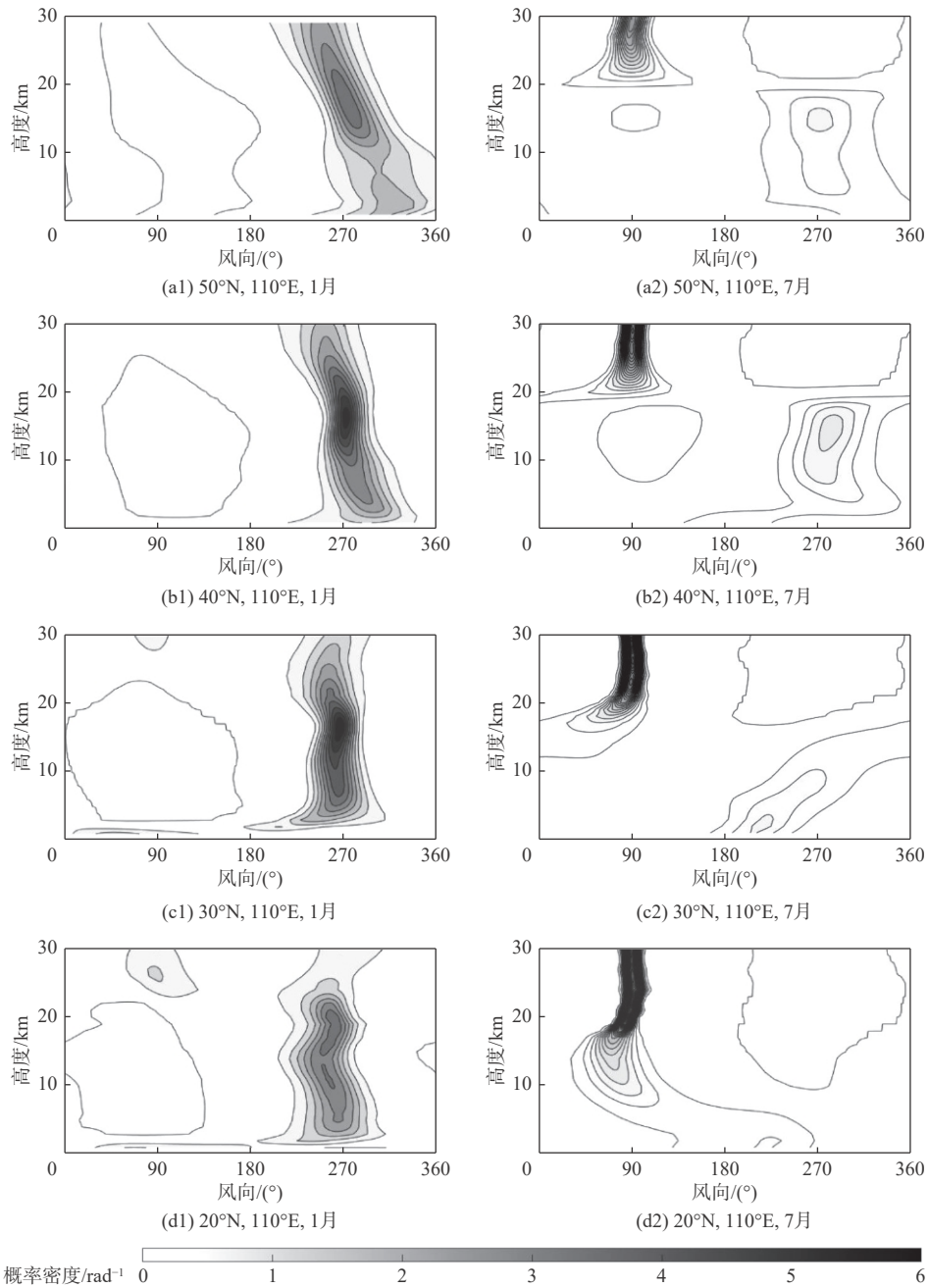


图1 风向概率密度

Fig. 1 Probability density of wind direction

风。由图 1 可知, 高空风场风向具有显著的季节变化特征, 各个高度层风向的概率密度在每个月基本呈正态分布, 满足综合矢量风剖面的假设条件, 因此, 本文采用月分辨率进行统计分析。

针对每个月在最多风向上计算概率最大风速, 该风速对飞行器威胁最大, 运载火箭主要采用 95% 概率。图 2 为 110°E 不同纬度(20°N、30°N、40°N 和 50°N)上空地面到 30 km 的 95% 概率最大风速在不同季节随高度的变化。本文以 1 月、4 月、7 月和 10 月分别代表冬季、春季、夏季和秋季。各纬度最大风速在不同季节不同高度呈现不同特征。最大风速在冬季, 5 km 以下高度差异不大; 7～15 km 高度 30°N 最强可达 81 m/s, 40°N 次之可达 57 m/s, 20°N 和 50°N 最弱; 20 km 以上高度 50°N 最强可达 62 m/s, 40°N 次之(44 m/s), 20°N 和 30°N 最弱。春季比冬季, 最大风速在 30°N 上空 7～15 km 高度和 50°N 上空 20～30 km 高度有所减弱, 其他基本维持不变。最大风速在夏季, 7～15 km 高度 40°N 最强可达 48 m/s, 50°N 次之(38 m/s), 20°N 和 30°N 最弱; 20 km 以上高度 20°N 和 30°N 强于 40°N 和 50°N。7～15 km 高度最大风速在秋季 30°N 和 40°N 明显增强。比较各季节最大风速变化, 整体而言, 冬季的最大风速强于夏季, 春秋季具有过渡特征。

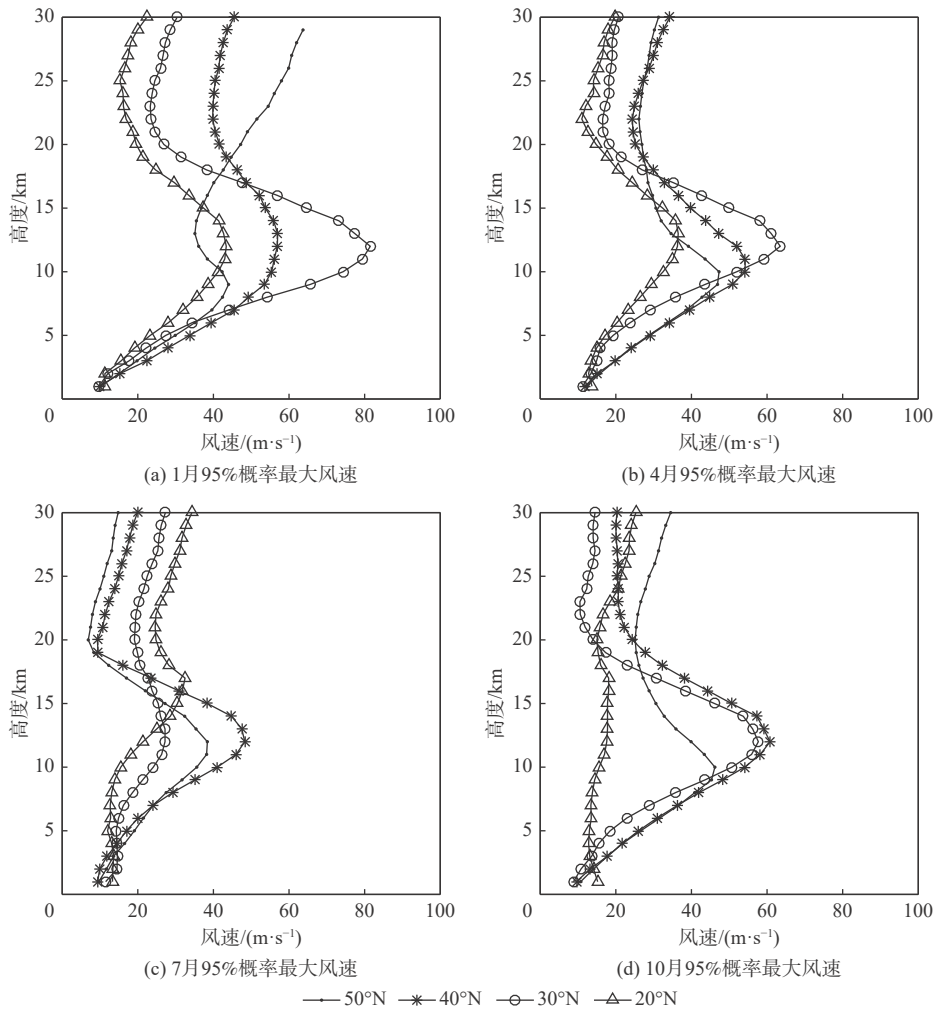
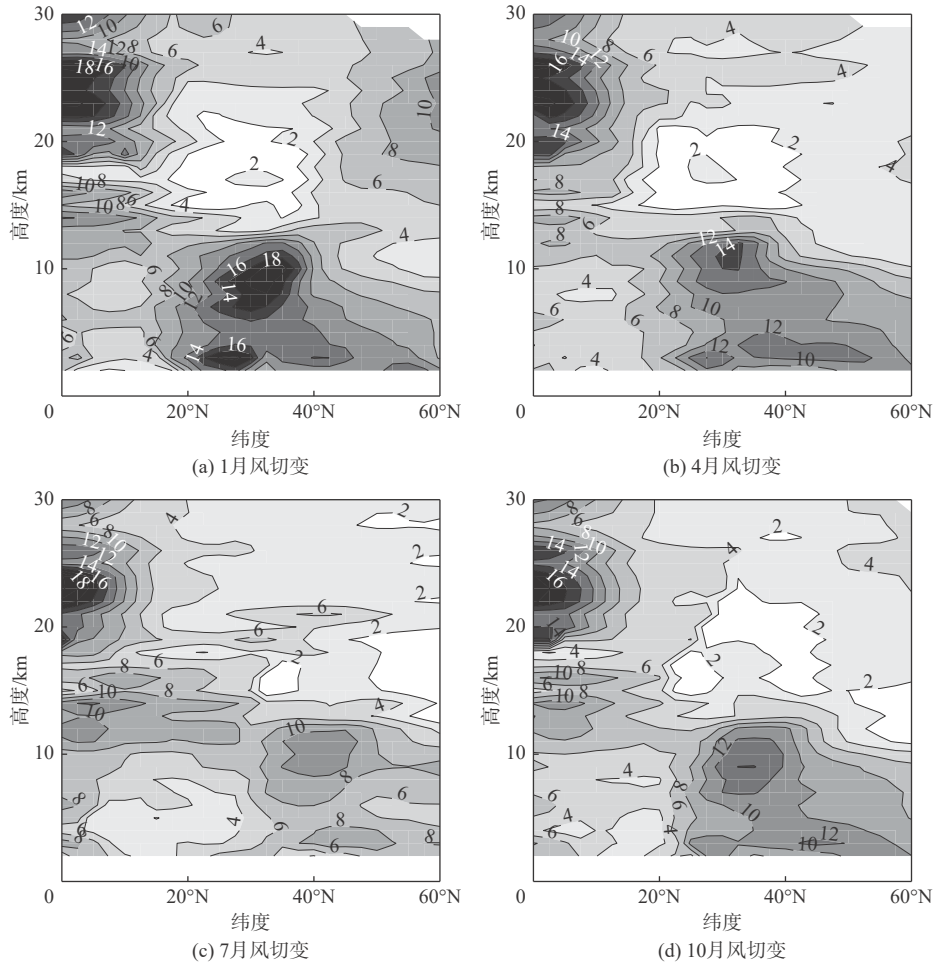


图 2　95% 概率最大风速随高度变化

Fig. 2　Altitude variation of maximum wind speeds with 95% probability

各高度的概率最大风速能够引起其他高度的风切变, 本文主要给出各参考高度最大风速产生该高度以下 1 km 处的切变。图 3 为 110°E 上空风切变随纬度和高度的变化, 图中数值为风切变, 单位为 m/s。由图可知, 以 15°N 为界, 风切变在低纬度和中高纬度的特征明显不同。低纬度地区的风切变极大值主要集中在 20～25 km 高度, 次极大值区域在 10～15 km 高度, 其随季节变化不显著。15°N 以北纬度地区, 风切变极值区域主要在 7～13 km 高度, 但强弱和极值所处纬度随月份变化: 1 月份风切变极大值在 30°N 的 9 km 附近, 约 20 m/s; 4 月份风切变极大值比 1 月份有所减弱, 约 15 m/s; 7 月份风切变极大值北移到 42°N 附近, 约 11 m/s; 4 月份风切变极大值处于 35°N, 约 14 m/s。15°N～50°N



(a) 1月风切变 (b) 4月风切变

(c) 7月风切变 (d) 10月风切变

图 3 110°E 风切变随纬度和高度变化

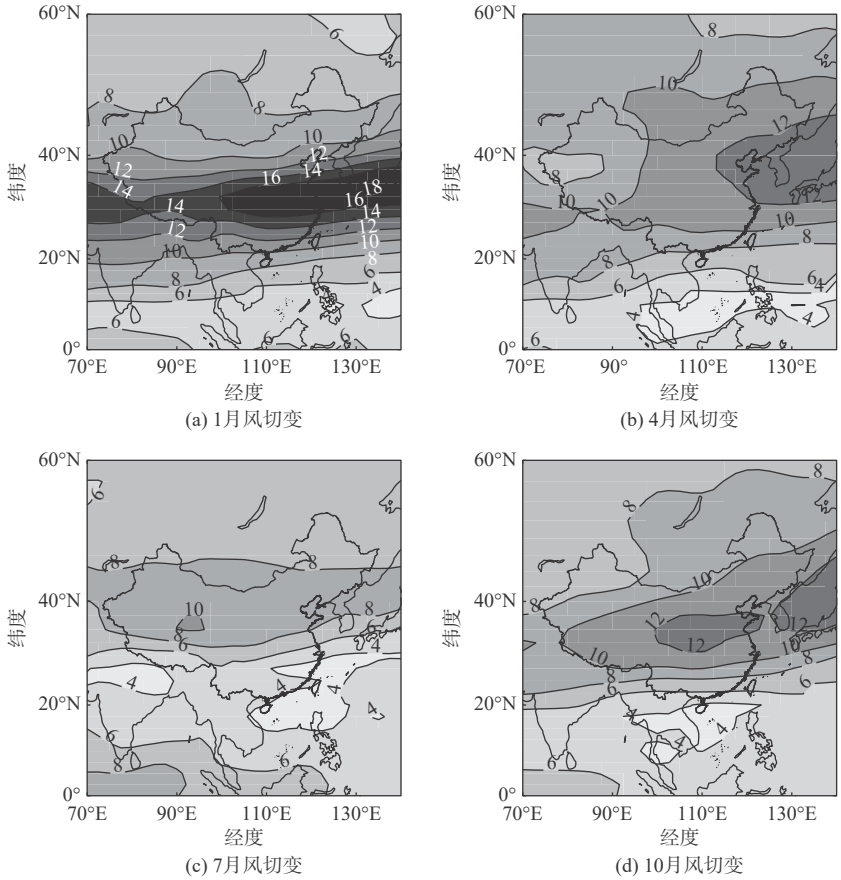
Fig. 3 Latitude-altitude section of wind shear over 110°E

的 13 km 以上风切变较小的原因主要是因为概率最大风速在此高度以上逐渐递减或基本不变造成的。冬季较高纬度地区的 20～25 km 处也存在风切变极值区域，这与图 2(a)中 50°N 该高度 95% 概率最大风速较大有关。

以风切变较大高度 9 km 为例，图 4 为各个季节风切变的水平分布特征，图中数值为风切变，单位为 m/s。由图可知，风切变在不同月份的分布特征差异显著。1 月份，风切变极值区自西(31°N, 70°E)穿过中国青藏高原南侧、江淮流域和东海，到日本的南侧达到峰值 18 m/s，此带 2 次的风切变逐渐减弱。4 月份，西藏高原上空风切变极值区减弱甚至消失，东部的极值区强度减弱，所处纬度略向北移动。7 月份，风切变极值区移动到 40°N 以北，在青藏高原北侧出现最大值 10 m/s。夏季风切变减弱与概率最大风减弱相对应。10 月份，中国东部地区和日本上空风切变增强，在中国中原地区和日本海上空形成极值区。

2.2 风切变与大气环流关系

中纬度地区受大尺度全球环流的影响，对流层顶会形成速度较快的大风区，称为“西风急流”。西风急流不仅具有风速强的特点，还具有强切变的特性。在中国，西风急流受青藏高原的影响会形成南北支急流，从而影响发射场上空的风场。因此，了解中国大气环流特性有助于进一步了解中国高空风切变特性形成原因。由于经向风较弱，风场主要取决于纬向风，图 5 为 110°E 地面到 30 km 高度纬向风随纬度和高度变化，图中数值为纬向风，单位为 m/s。从图 5 可知，北半球的风场随季节变化显著。1 月份，低纬度地区以东风为主，在 15 km 和 25 km 以上高度存在极值区域，20 km 甚至出现零风速区域，这也是导致图 3 中 20～25 km 和 10～15 km 高度出现较大风切变的原因；15°N 上空出现零风速线，以北区域盛行西风，极大值 63 m/s 位于 30°N 上空 12 km 高度，这是引起冬季风切变的高空急流；较高纬度的 25 km 以上也存在极值区，也引起该区域的强切变(见图 3(a))。2 月份，零风速线向北移动，中纬度西风明显减弱，高空急流中心风速为 43 m/s(13 km 高度)。7 月份，0～20°N 的 5 km 以下为弱西风，以上高度以东风为主；零风速线从 20°N 的 5 km 高度迅速到 30°N 的 20 km 高度，30°N～60°N 形成稳定的夏季“准零风层”，该层以下高度

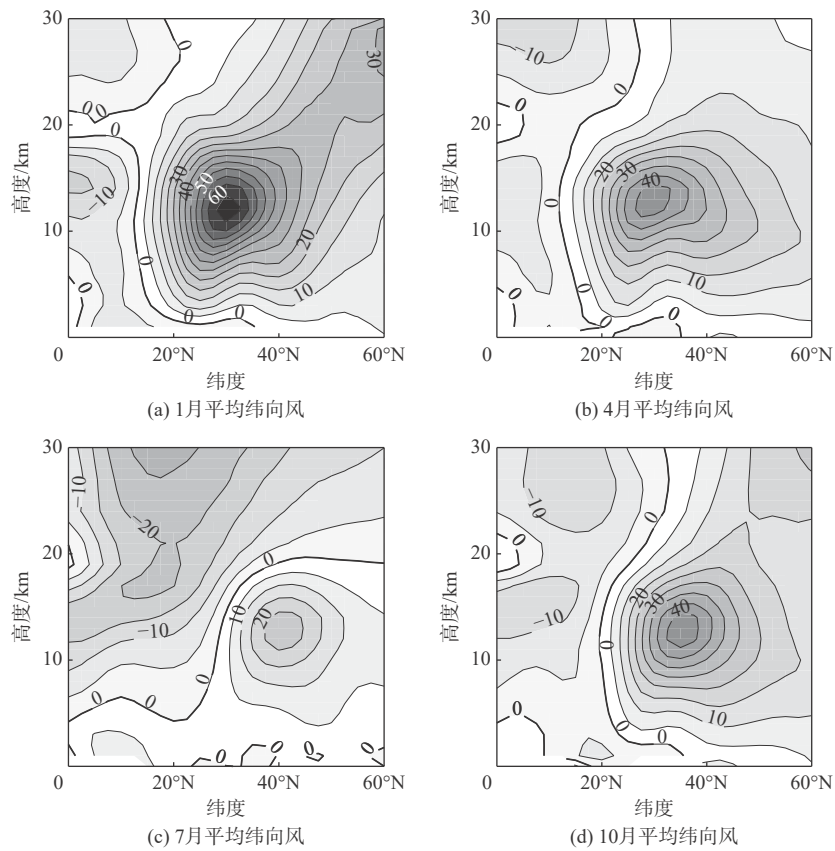


(a) 1月风切变　(b) 4月风切变

(c) 7月风切变　(d) 10月风切变

图 4　9 km 高度风切变水平分布

Fig. 4　Horizontal distribution of wind shear at 9 km



(a) 1月平均纬向风　(b) 4月平均纬向风

(c) 7月平均纬向风　(d) 10月平均纬向风

图 5　110°E 纬向风随纬度和高度变化

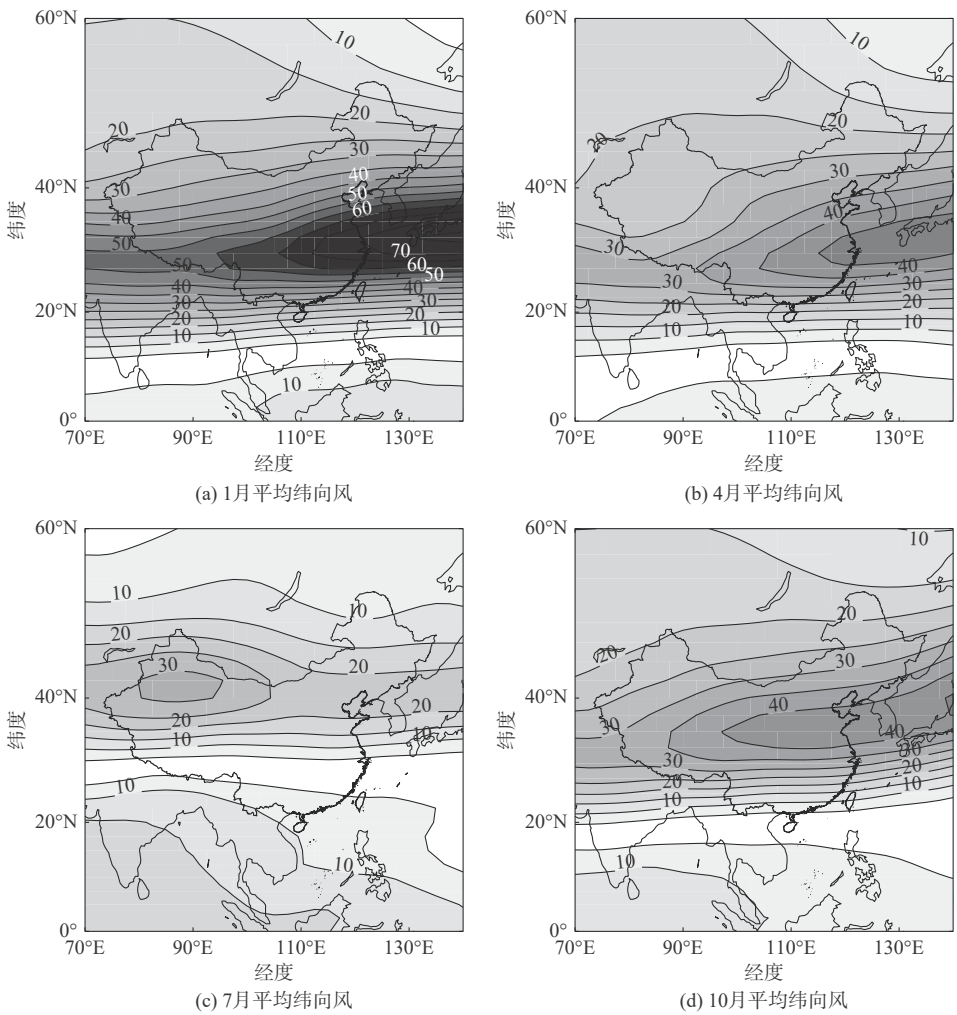
Fig. 5　Latitude-altitude section of zonal wind over 110°E

盛行西风，以上高度盛行东风；西风急流位于40°N的13 km，风速为23 m/s。10月份，零风速线退回到20°N～25°N，向上延伸到30 km，该纬度带以南盛行东风，以北盛行西风，西风急流增强，极大值43 m/s位于35°N的12 km。西风急流的季节变化主要受南北气温差异季节变化的影响[26]。受太阳辐射影响，西风急流的南侧为暖平流，北侧为冷平流，南北两侧存在温差引起急流中心随热量平流的最大经向梯度中心南北移动。在冬半年到夏半年的转换时间，急流中心北侧增温强于南侧增温，南北温差的减小导致急流减弱北进；在夏半年到冬半年，急流中心北侧降温强于南侧降温，南北温差的增大导致急流增强南退。比较图5和图3，纬向风和风切变的结构有很好的对应关系，说明大气环流直接影响风切变。

中国高空急流具有显著的季节突变性，直接影响中国的大气环流、天气和气候系统[24]。图6为各个季节13 km高度纬向风的水平分布特征，图中数值为纬向风，单位为m/s。1月份，高空急流中心轴纬度位于30°N左右，从中国的青藏高原南侧一直向东穿过中国江淮地区、东海到日本南部地区。春季高空急流风速减弱。7月份，高空急流西段在青藏高原南侧消失，北跳到青藏高原北侧，形成风速极值区，夏季急流轴纬度位于40°N附近，从中国的西北地区贯穿中国华北地区、朝鲜半岛到日本中、北部地区。秋季，高空急流西段又从青藏高原北侧南退到高原南侧。由图可知，西风急流的季节性北跳在东侧和西侧具有不一致性，主要与青藏高原大地形动力及热力作用有关。春季青藏高原温度升高，低压下沉气流非绝热加热作用导致了急流中心在7月的西移突变[26]。通过比较图6和图4，风切变的水平分布特征与高空急流基本一致，具有密切的关系。

2.3 站点气球探空数据风切变比较

中国气象观测站的气球探空数据能够提供高分辨的大气参量。本节以气球探空数据作为观测资料计算风场特性，比较其与NCEP资料风切变的差异性。选取酒泉(39.76°N, 98.48°E)，西昌(27.9°N, 102.26°E)和海口(20.03°N, 110.35°E)气象站数据，分别处于不同纬度带，对应于酒泉。图7为3个站

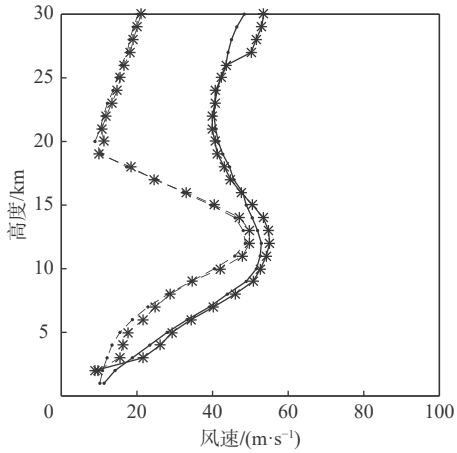


(a) 1月平均纬向风　(b) 4月平均纬向风

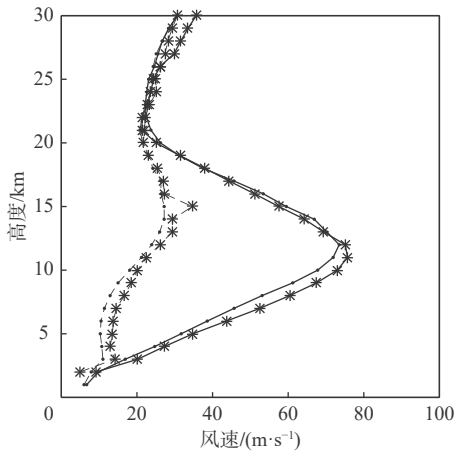
(c) 7月平均纬向风　(d) 10月平均纬向风

图6　13 km高度纬向风水平分布

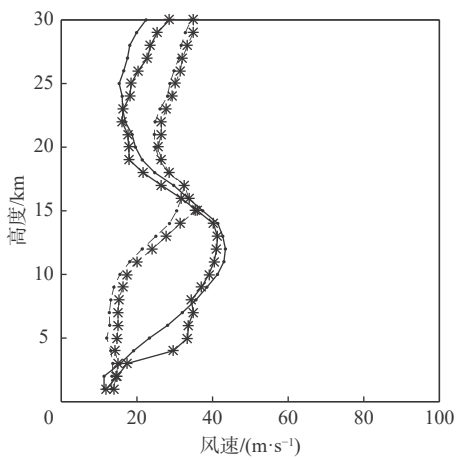
Fig. 6　Horizontal distribution of zonal wind at 13 km



(a) 酒泉95%概率最大风速



(b) 西昌95%概率最大风速



(c) 海口95%概率最大风速

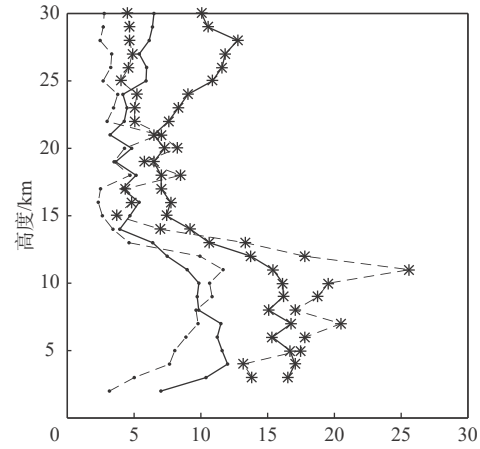
探空-1月　NCEP-1月　探空-7月　NCEP-7月

图 7　气球探空 95% 概率最大风速随高度变化

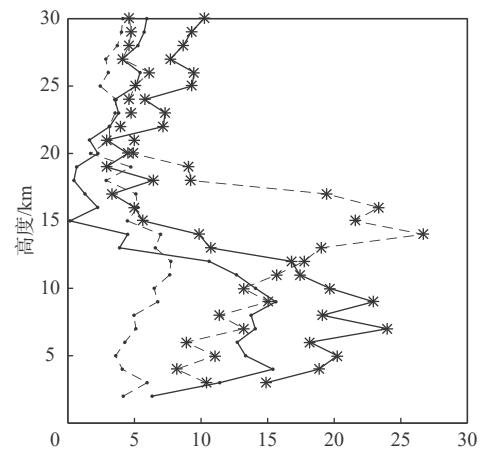
Fig. 7　Altitude variation of maximum wind speeds with 95% probability from radiosonde

点地面到 30 km 高度两者 95% 概率最大风速的比较。对于各个站点, NCEP 中最大风速与实测资料几乎一致, 除在个别高度略有差异。3 个台站上空季节特征各不相同。酒泉上空最大风速在冬季强于夏季。西昌上空最大风速在 20 km 以下高度冬季强于夏季, 20 km 以上高度风速相当。海口上空最大风速在 15 km 以下高度冬季强于夏季, 15 km 以上高度夏季强于冬季。

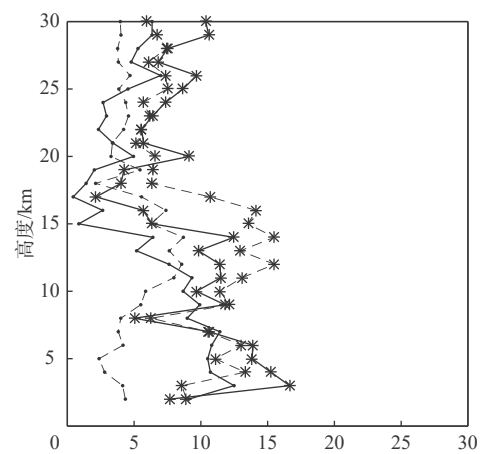
图 8 为气球探空数据中风切变与 NCEP 数据的比较。两者在风切变的差异比最大风速差异显著, 实测数据获得的风切变比再分析资料中风切变偏大, 两者差异夏季较大, 如酒泉上空 11 km 高度偏差可达 14 m/s, 西昌上空 14 km 高度偏差可达近



(a) 酒泉风切变



(b) 西昌风切变



(c) 海口风切变

探空-1月　NCEP-1月　探空-7月　NCEP-7月

图 8　气球探空风切变随高度变化

Fig. 8　Altitude variation of wind shear from radiosonde

20 m/s。NCEP数据中风切变强弱随季节的变化与实测数据一致。海口距离西风急流区域较远，受影响程度较小，风切变强度较弱，实测和再分析资料的差异略小。再分析资料中风切变强度偏弱可能是由于再分析资料同化观测资料时过滤了一些瞬变信息，获得再分析产品相对平滑的原因造成的。程胡华等[12]研究也说明了采用真实大气环境的重要性。采用实测数据，提高测风技术精度，完善风场预报，提供更精确的自然环境研究结果，对火箭设计具有重要的意义。

2.4 风切变对火箭发射段攻角影响

选取两型运载火箭研究其攻角受高空大气风切变的影响：一种是张建栋[27]设计的固体运载火箭上升段轨迹（以“R_1”表示）；另一种是张斌[28]根据猎鹰9号火箭仿真的运载火箭上升段轨迹（以“R_2”表示）。图9为这两型运载火箭高度和速度变化曲线。R_1起飞较快，经过73 s到达30 km，此高度速度达1 600 m/s。R_2起飞较慢，经过106 s到达30 km，此高度速度为852 m/s。因为本文NCEP再分析资料和气球探空资料处理后的数据和计算的综合矢量风信息都为高度分辨率为1 km的数据，因此，图9(c)中给出的两型火箭发射速度也为高度间隔1 km。利用式(5)计算各高度层两型火箭风攻角。

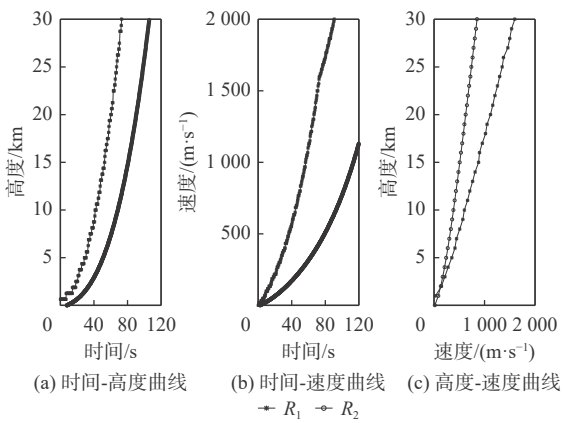


图9 运载火箭高度和速度变化

Fig. 9 Altitude variation of altitude and speed for launch vehicle

以海口为例，图10为气球探空数据和NCEP数据中获得的高空大气风切变对两型运载火箭的风攻角分布。对于R_1速度较快运载火箭，风攻角随高度变化大致可以分为3段：5 km以下高度火箭刚起飞，飞行速度较慢，容易受到风切变影响，风攻角普遍大于2.5°，最大值为1月份3 km高度的9°；7～18 km高度火箭飞行速度提高，风攻角从2.5°逐渐减小到0.5°左右；18 km以上高度风攻角几乎都在0.5°附近。对于R_2速度较慢运载火箭，风攻角比R_1稍大，大约是后者的2倍：5 km以下高度普遍大

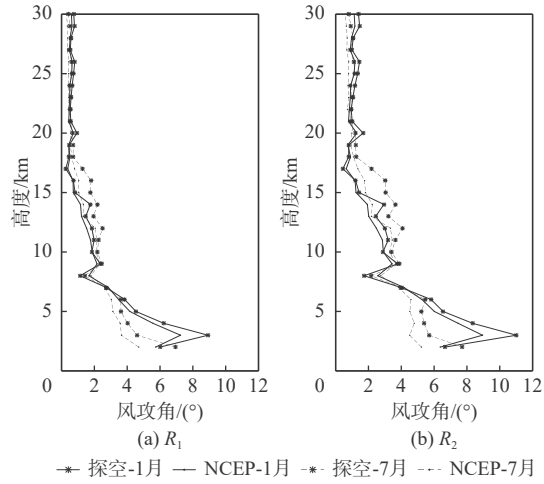


图10 运载火箭风攻角变化

Fig. 10 Altitude variation of wind attack angle for launch vehicles

于4°，最大值为1月份3 km高度的11°；7～18 km高度从4°逐渐减小到1°左右；18 km以上高度主要在1°附近。整体而言，气球探空数据计算风攻角比用NCEP数据更大，受最大风速和风切变的差异影响。

3 结 论

1）中国陆地上空风向具有显著的季节变化，冬季盛行西风，夏季20 km以上高度盛行东风，20 km以上高度在40°N和50°N盛行西风，在20°N和30°N各有不同。冬季最大风速在30°N的7～15 km高度最强可达81 m/s。夏季最大风速在40°N的7～15 km高度最强可达48 m/s。对于各纬度，冬季普遍强于夏季。

2）风切变在15°N以南的低纬度和以北的中高纬度的特征不同。低纬度地区的风切变极大值位于20～25 km和10～15 km高度。中高纬度风切变极值区域主要在7～13 km高度，强弱和极值所处纬度随月份变化，1月份极大值20 m/s在30°N的9 km高度；7月份极大值11 m/s在42°N附近。风切变的季节变化和水平分布直接受中国高空急流的季节特征和水平分布的影响。

3）气球探空数据中的概率最大风速与NCEP数据中基本一致，而前者的风切变明显强于后者，这种差异在夏季中纬度高空急流区附近最为显著。所以在风场设计中采用实测数据非常重要，可以包含更多危险条件。风切变对火箭发射段产生的攻角在5 km以下和7～18 km高度。

致谢 感谢NCEP/NCAR发布的NCEP再分析资料和美国怀俄明州大学大气科学学院发布的气球探空资料。

参考文献（References）

[1] 马艳如, 石晓荣, 刘华华, 等. 运载火箭姿态系统自适应神经网络容错控制[J]. 宇航学报, 2021, 42(10): 1237-1245.
MA Y R, SHI X R, LIU H H, et al. Adaptive neural network fault tolerant control of launch vehicle attitude system[J]. Journal of Astronautics, 2021, 42(10): 1237-1245(in Chinese).

[2] 李东, 王珏, 何巍, 等. 长征五号运载火箭总体方案及关键技术[J]. 导弹与航天运载技术, 2017(3): 1-5.
LI D, WANG J, HE W, et al. The general scheme and key technologies of CZ-5 launch vehicle[J]. Missiles and Space Vehicles, 2017(3): 1-5(in Chinese).

[3] 吴燕生. 中国运载火箭姿态控制技术发展与展望[J]. 宇航学报, 2023, 44(4): 509-518.
WU Y S. Review and prospect of attitude control technologies for China's launch vehicle[J]. Journal of Astronautics, 2023, 44(4): 509-518(in Chinese).

[4] 宋征宇, 刘立东, 陈晓飞, 等. 新一代中型系列运载火箭长征八号的发展及其关键技术[J]. 宇航学报, 2023, 44(4): 476-485.
SONG Z Y, LIU L D, CHEN X F, et al. Development and key technologies of long March 8 family: China's next-generation medium-lift launchers[J]. Journal of Astronautics, 2023, 44(4): 476-485(in Chinese).

[5] 程旋, 肖存英, 胡雄. 临近空间大气环境对高超声速飞行器气动特性的影响研究进展[J]. 飞航导弹, 2018(5): 22-28.
CHENG X, XIAO C Y, HU X. Research progress on the influence of atmospheric environment in near space on aerodynamic characteristics of hypersonic vehicle[J]. Aerodynamic Missile Journal, 2018(5): 22-28(in Chinese).

[6] 陈闽慷, 杜涛, 胡雄, 等. 北半球高空大气参数波动对临近空间飞行热环境的影响[J]. 科学通报, 2017, 62(13): 1402-1409.
CHEN M K, DU T, HU X, et al. Effect of atmosphere parameter oscillation at high altitude in the Northern Hemisphere for near space hypersonic flight aerothermodynamic prediction[J]. Chinese Science Bulletin, 2017, 62(13): 1402-1409(in Chinese).

[7] 程旋, 肖存英, 杜涛, 等. 临近空间大气密度扰动对高超声速飞行器气动热环境的影响[J]. 北京航空航天大学学报, 2021, 47(4): 754-764.
CHENG X, XIAO C Y, DU T, et al. Influence of atmospheric density disturbance on aerothermodynamic environment of hypersonic vehicles in near space[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(4): 754-764(in Chinese).

[8] 程胡华, 王益柏, 蔡其发, 等. 大气密度对运载火箭飞行的 $q\alpha_{max}$ 精度影响及建模分析[J]. 宇航学报, 2021, 42(3): 378-389.
CHENG H H, WANG Y B, CAI Q F, et al. Influence of atmospheric density on $q\alpha_{max}$ accuracy of launch vehicle flight and modeling analysis[J]. Journal of Astronautics, 2021, 42(3): 378-389(in Chinese).

[9] 万田, 刘洪伟, 樊菁. 100km 附近大气密度模型的误差带和置信度[J]. 中国科学: 物理学 力学 天文学, 2015, 45(12): 66-72.
WAN T, LIU H W, FAN J. Error band and confidence coefficient of atmospheric density models around altitude 100 km[J]. Scientia Sinica (Physica, Mechanica & Astronomica), 2015, 45(12): 66-72 (in Chinese).

[10] 肖存英, 胡雄, 杨钧烽, 等. 临近空间 38°N 大气密度特性及建模技术[J]. 北京航空航天大学学报, 2017, 43(9): 1757-1765.
XIAO C Y, HU X, YANG J F, et al. Characteristics of atmospheric density at 38°N in near space and its modeling technique[J]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(9): 1757-1765(in Chinese).

[11] 韩丁, 盛夏, 尹珊建, 等. 临近空间大气参数误差特性分析[J]. 遥感学报, 2017, 21(1): 149-158.
HAN D, SHENG X, YIN S J, et al. Deviation characteristics for atmospheric parameters in near space[J]. Journal of Remote Sensing, 2017, 21(1): 149-158(in Chinese).

[12] 程胡华, 李娟, 肖云清, 等. 风偏差对火箭最大气动载荷精度的影响[J]. 北京航空航天大学学报, 2021, 47(10): 2034-2042.
CHENG H H, LI J, XIAO Y Q, et al. Influence of wind deviation on rocket maximum aerodynamic load accuracy[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(10): 2034-2042(in Chinese).

[13] 程路, 姜长生, 都延丽, 等. 变化风场下近空间飞行器机体/发动机一体化飞行力学建模与分析[J]. 宇航学报, 2012, 33(5): 547-555.
CHENG L, JIANG C S, DU Y L, et al. Flight dynamics modeling of airframe/engine integrated near space vehicle in varying wind field[J]. Journal of Astronautics, 2012, 33(5): 547-555(in Chinese).

[14] 杨伟奇, 许志, 唐硕, 等. 基于自抗扰的运载火箭主动减载控制技术[J]. 北京航空航天大学学报, 2016, 42(1): 130-138.
YANG W Q, XU Z, TANG S, et al. Active disturbance rejection control method on load relief system for launch vehicles[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(1): 130-138(in Chinese).

[15] 余梦伦. CZ-2E 火箭高空风弹道修正[J]. 导弹与航天运载技术, 2001(1): 9-15.
YU M L. CZ-2E ballistic correction for high altitude wind[J]. Missiles and Space Vehicles, 2001(1): 9-15(in Chinese).

[16] 王建明, 林娜, 张博戎. 某大型低温火箭高空风载荷分析[J]. 导弹与航天运载技术, 2021(5): 137-141.
WANG J M, LIN N, ZHANG B R. Analysis on flight wind-load for large cryogenic rocket[J]. Missiles and Space Vehicles, 2021(5): 137-141(in Chinese).

[17] 胡瑞光, 钟文安, 宋征宇, 等. 运载火箭风场高精度拟合方法[J]. 宇航学报, 2023, 44(2): 303-312.
HU R G, ZHONG W A, SONG Z Y, et al. High-precision fitting approach for wind field of launch vehicles[J]. Journal of Astronautics, 2023, 44(2): 303-312(in Chinese).

[18] 赵人濂, 陈振官, 付维贤. 风切变与运载火箭设计[J]. 宇航学报, 1998, 19(2): 105-108.
ZHAO R L, CHEN Z G, FU W X. Wind shear and rocket design[J]. Journal of Astronautics, 1998, 19(2): 105-108(in Chinese).

[19] 黄世勇, 闻悦. 20～85km 高空大气密度与风场研究[J]. 导弹与航天运载技术, 2017(3): 17-21.
HUANG S Y, WEN Y. Reasearch for atmospheric density and wind field from 20km to 85km[J]. Missiles and Space Vehicles, 2017(3): 17-21(in Chinese).

[20] 杨钧烽, 肖存英, 胡雄, 等. 临近空间风切变特性及其对飞行器的影响[J]. 北京航空航天大学学报, 2019, 45(1): 57-65.
YANG J F, XIAO C Y, HU X, et al. Wind shear characteristics in near space and their impacts on air vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(1): 57-65(in Chinese).

[21] 陆日宇, 林中达, 张耀存. 夏季东亚高空急流的变化及其对东亚季风的影响[J]. 大气科学, 2013, 37(2): 331-340.
LU R Y, LIN Z D, ZHANG Y C. Variability of the east Asian upper-tropospheric jet in summer and its impacts on the east Asian monsoon[J]. Chinese Journal of Atmospheric Sciences, 2013, 37(2): 331-340(in Chinese).

[22] KALNAY E, KANAMITSU M, KISTLER R, et al. The NCEP/NCAR 40-year reanalysis project[J]. Bulletin of the American Meteorological Society, 1996, 77(3): 437-472.

[23] 韦芬芬, 汤剑平, 王淑瑜. 中国区域夏季再分析资料高空变量可信度的检验[J]. 地球物理学报, 2015, 58(2): 383-397.
WEI F F, TANG J P, WANG S Y. A reliability assessment of upper-level reanalysis datasets over China[J]. Chinese Journal of Geophysics, 2015, 58(2): 383-397(in Chinese).

[24] 陈钦明, 宋淑丽, 朱文耀. 亚洲地区 ECMWF/NCEP 资料计算 ZTD 的精度分析[J]. 地球物理学报, 2012, 55(5): 1541-1548.
CHEN Q M, SONG S L, ZHU W Y. An analysis of the accuracy of zenith tropospheric delay calculated from ECMWF/NCEP data over Asian area[J]. Chinese Journal of Geophysics, 2012, 55(5): 1541-1548(in Chinese).

[25] ZHANG S D, YI F. A statistical study of gravity waves from radiosonde observations at Wuhan (30 N, 114 E) China[J]. Annales Geophysicae, 2005, 23(3): 665-673.

[26] 况雪源, 张耀存. 东亚副热带西风急流季节变化特征及其热力影响机制探讨[J]. 气象学报, 2006, 64(5): 564-575.
KUANG X Y, ZHANG Y C. The seasonal variation of the east Asian subtrpical westerly jet and its thermal mechanism[J]. Acta Meteorologica Sinica, 2006, 64(5): 564-575(in Chinese).

[27] 张建栋. 固体运载火箭上升段多约束制导方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2021.
ZHANG J D. Research on multi-constraint guidance method for ascending phase of solid launch vehicle[D]. Harbin: Harbin Institute of Technology, 2021(in Chinese).

[28] 张斌. 某型运载火箭上升段制导控制研究[D]. 西安: 西安电子科技大学, 2020.
ZHANG B. Research on guidance and control of ascending phase of a launch vehicle[D]. Xi'an: Xidian University, 2020(in Chinese).

Spatiotemporal distribution characteristics of high-altitude wind shear

YANG Junfeng[1, 2], CHENG Xuan[3], WANG Jianmei[1, 2], ZHANG Yiming[1, 2], HU Xiong[1, 2], XIAO Cunying[4, 5, *]

(1. Key Laboratory of Solar Activity and Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;
2. Key Laboratory of Science and Technology on Environmental Space Situation Awareness, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;
3. Deep Space Exploration Laboratory, Beijing 100041, China;
4. Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 100875, China;
5. Department of Astronomy, Beijing Normal University, Beijing 100875, China)

Abstract: The vehicles at the rising stage are susceptible to being influenced by upper atmospheric wind. The high-altitude wind shear is essential in the rocket design. The synthetical wind vector approach is used to determine the wind shear features from the ground to 30 km in the East Asia region, based on the National Centers for Environment Prediction (NCEP) reanalysis data from 2020 to 2022. The findings indicate a relationship between season, altitude, and horizontal distribution and the maximum wind direction, maximum wind speed, and wind shear. The wind shear from radiosonde observations is used to analyze the difference between true atmospheric wind shear and that from NCEP. NCEP underestimates the wind shear. The attack angles of two simulated launch vehicles in the atmosphere wind are analyzed. The results indicate that wind shear has a significant impact below 5 km and between 7～18 km. These results are helpful to grasp the distribution characteristics of high-altitude wind shear in China and provide a reference for the accurate and safe rocket design.

Keywords: wind shear; synthetical wind vector; upper-tropospheric jet; rocket design; attack angle

Received: 2023-11-09; **Accepted**: 2023-12-22; **Published Online**: 2024-02-02 10:47

URL: link.cnki.net/urlid/11.2625.V.20240201.1015.002

Foundation items: National Natural Science Foundation of China (12241101,42174192,11872128); Pandeng Program of National Space Science Center, Chinese Academy of Sciences

* **Corresponding author**. E-mail: xiaocunying@bnu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0724

基于 INT-VSMM 算法的目标航迹跟踪和外推

任宣铭[1]，汤新民[2,*]，刘雨生[1]，鲁其兴[1]

(1. 南京航空航天大学 民航学院，南京 211106；2. 中国民航大学 交通科学与工程学院，天津 300300)

摘 要：针对常规变结构多模型 (VSMM) 算法计算时间长，难以满足实时性的问题，通过广播式自动相关监视 (ADS-B) 捕获的目标状态和飞行意图信息作为模型先验信息，利用该先验信息以变结构多模型为理论框架，提出一种基于飞行意图的变结构多模型 (INT-VSMM) 算法。将航空器在航路飞行阶段的运动模式分解，建立完备的运动模型集;根据有向图切换原理，设计以“硬”切换为主，以“软”切换为辅的模型集切换算法；采用 INT-VSMM 算法对目标航空器的航迹进行跟踪，并根据目标状态估计进行了短期航迹外推。仿真结果表明:基于 INT-VSMM 算法的目标跟踪性能和计算时间均优于对比算法，外推轨迹在短期内误差较小，可以满足冲突探测需要。

关 键 词：变结构多模型；飞行意图；有向图切换；机动目标跟踪；航迹外推

中图分类号：V249

文献标志码：A **文章编号**：1001-5965(2026)01-0203-11

随着全球航空运输业的急速发展，空域内的航空器密度已经对管制员的处理能力形成挑战，这势必会增加空域内航空器的事故风险[1-2]。为此，使航空器利用自身的空中监视和通讯技术承担一部分冲突探测责任，可以有效缓解管制员的工作负担，保证航空器运行安全。

空中监视技术的进步允许本机通过目标航空器发送的广播式自动相关监视 (automatic dependent surveillance-broadcast, ADS-B) 报文，获得目标航空器的飞行状态[3] 和意图信息。根据这些信息，本机可以对周围航空器进行跟踪[4-6]，并对其在未来一段时间内的飞行轨迹进行外推，为进一步实现冲突探测提供数据支撑。但由于航空器的意图信息难以捕获且不连续，因此，目前的研究主要根据航空器的历史轨迹信息外推航空器在未来一段时间内的飞行航迹[7-9]。

在冲突探测中，目标跟踪与预测依赖于运动模型假设。单模型估计虽简单易实现[10]，但在复杂的航空器轨迹跟踪和预测中，单模型估计容易造成航空器的飞行模式与预测模型失配。文献 [11] 提出一种基于马尔可夫转换系数的交互多模型 (interacting multiple model, IMM) 算法，该算法通过将航空器的飞行动作划分为相应的运动模式，如匀速运动、匀加速运动、匀速转弯，使航空器在每种运动模式下均能匹配与其对应的一种或多种运动模型来描述航空器运动[12-13]。考虑到实际噪声一般符合有色噪声过程，Singer 于 1970 年提出 Singer 模型[14]。Singer 模型是一种全局统计模型[15]，在目标跟踪时无需进行检测，无时间滞后。Jerk 模型在 Singer 模型的基础上增加了一维加速度的倒数[16]，有效消除了高频噪声，从而得到对加速度更加准确的估计，因此，更适用于强机动目标跟踪。但 Singer 模型假设目标加速为零均值随机噪声过程，Jerk 模型则假设目标加速度的倒数为零均值，均与航空器实际飞行情况不符。机动目标“当前”统计模型假设加速度均值为非零的一阶时间相关函数，加速度的“当

收稿日期：2023-11-05；**录用日期**：2023-12-11；**网络出版时间**：2023-12-27 18:10

网络出版地址：link.cnki.net/urlid/11.2625.V.20231227.1212.002

基金项目：国家重点研发计划 (2021YFB1600500)；国家自然科学基金 (61773202,52072174)；中国民航管理干部学院民航通用航空运行重点实验室开放基金 (CAMICKFJJ-2019-04)；高端外国专家引进计划 (G2023202003L)；天津市科技计划项目 (24JCZDJC00090)

* **通信作者**. E-mail：tangxinmin@nuaa.edu.cn

引用格式：任宣铭，汤新民，刘雨生，等. 基于 INT-VSMM 算法的目标航迹跟踪和外推 [J]. 北京航空航天大学学报，2026，52（1）：203-213.
REN X M，TANG X M，LIU Y S，et al. Target trajectory tracking and extrapolation based on INT-VSMM algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：203-213（in Chinese）.

前”概率密度采用修正的瑞利分布表述[17-19]，因此，更加符合航空器的实际飞行情况。

使用IMM进行卡尔曼滤波时，需用固定模型集描述目标航空器机动，但模型增多会导致计算成本激增和模型失配，降低跟踪精度[20]。为解决上述问题，文献[21]提出一种变结构多模型(variable structure multiple model, VSMM)架构，该架构为不同的运动模式建立与其匹配的模型集合，并利用观测数据计算后验的最佳模型集合来描述当前的运动模式。基于VSMM架构，许多专家提出了各自的变结构算法来提高机动目标的跟踪性能[22-24]。文献[25]在当前统计模型中引入模糊隶属度函数，以实现目标加速度的自适应调节。为解决传统跟踪算法在处理强跟踪问题时模型适应性差的问题，文献[26]通过反馈网络实时监控不同模型与目标实际运动的匹配程度并据此动态调整模型参数，优先选择匹配度高的模型参数与跟踪计算，有效提升了模型对目标运动状态突变、遮挡等强跟踪场景的适应性。除基于卡尔曼滤波类算法，专家们还利用最小二乘法将多个数据压缩成一个新的观测值以提高跟踪精度[27-28]，但对非匀速运动效果不理想。曲线插值法[29]和曲线拟合法[30-32]均是通过不断寻找与实际观测数据逼近的函数以实现精确跟踪，适用于非均匀运动场景。但随着计算节点数量的增加，计算量会呈指数级增长。上述算法虽然都利用历史数据估计航空器当前运动状态，但对于快速变化或复杂机动特性的目标，容易产生模型集失配和模型切换迟滞问题，导致算法很难实时准确地反映目标最新的运动状态。

本文基于VSMM算法架构，旨在充分发挥航空器飞行阶段的态势感知能力，利用目标航空器的ADS-B数据，分析目标航空器当前运动模式和预期的飞行意图，在此基础上，针对目标的不同运动模式建立与其对应的运动模型集，随后，利用基于有向图的模型集切换算法匹配与当前运动适配的模型集，从而提高交互多模型算法的计算精度和计算效率。

1 目标运动状态建模

1.1 问题描述

航空器的机载ADS-B输出设备能够将航空器自身的经度、纬度、高度、速度、航向等状态信息及选择高度、选择航向等意图信息编码后以不同类型的报文形式进行广播，而本机则可以通过ADS-B输入设备接收周围航空器的报文并进行解码，从而感知周围航空器的运行状态。对于航空器的航迹外推来说，捕获的他机信息的完整性和准确性，将直接影响航迹的预测精度。本文通过捕获ADS-B空中位置报文、空中速度报文及目标状态和情况报文，分别获得目标航空器的位置信息(经度、纬度、高度信息)，速度和航向信息(东西向速度、南北向速度、爬升率信息)，以及飞行意图信息(选择高度、选择航向信息)，并基于目标信息利用变结构多模型算法结合卡尔曼滤波器进行跟踪，最后，根据航空器的运动趋势进行短期航迹的外推。

1.2 目标运动模型

航空器的运动状态可以用如下离散系统状态方程表示：

$$\boldsymbol{X}(k+1)=\boldsymbol{\Phi}(k)\boldsymbol{X}(k)+\boldsymbol{\Gamma}(k)\boldsymbol{w}(k) \tag{1}$$

式中：$\boldsymbol{\Phi}(k)$为状态转移矩阵；$\boldsymbol{\Gamma}(k)$为噪声驱动矩阵；$\boldsymbol{w}(k)\sim N(\boldsymbol{0},\boldsymbol{Q}_k)$为状态方程白噪声，$\boldsymbol{Q}_k$为过程噪声协方差；$\boldsymbol{X}(k)$为系统在$k$时刻的状态向量。

本文采用ADS-B报文作为观测数据，需要考虑航空器在平飞过程中的高度是不变的，若采用空间直角坐标系会导致外推高度产生偏差[33]。因此，选择大地坐标系为本文目标跟踪的参考系，将目标航空器的状态变量写成三相分量的形式，则目标状态变量$\boldsymbol{X}(k)$可表示为

$$\begin{aligned}\boldsymbol{X}(k)=&\left[\boldsymbol{\alpha}(k),\boldsymbol{\beta}(k),\boldsymbol{\gamma}(k)\right]=\\&\left[x(k),\dot{x}(k),\ddot{x}(k),y(k),\dot{y}(k),\ddot{y}(k),h(k),\dot{h}(k),\ddot{h}(k)\right]\end{aligned} \tag{2}$$

式中：$\boldsymbol{\alpha}(k)$为经度方向分量；$x(k)$、$\dot{x}(k)$和$\ddot{x}(k)$分别为k时刻目标航空器在经度方向的位置、速度和加速度；$\boldsymbol{\beta}(k)$为纬度方向分量；$y(k)$、$\dot{y}(k)$和$\ddot{y}(k)$分别为k时刻目标航空器在纬度方向的位置、速度和加速度；$\boldsymbol{\gamma}(k)$为高度方向分量；$h(k)$、$\dot{h}(k)$和$\ddot{h}(k)$分别为k时刻目标航空器在高度方向的位置、速度和加速度。

1.2.1 匀速直线飞行

假设目标沿经度方向做匀速直线运动，则目标状态向量$\boldsymbol{\alpha}(k)=[x(k),\dot{x}(k)]^{\mathrm{T}}$，$\boldsymbol{w}(k)$为服从$[0,\sigma^2]$的高斯白噪声。为滤波和处理方便，需要将所有运动模型扩充至模型集中的最大维度。考虑到模型集中匀角速度转弯(constant turning, CT)模型的状态转移矩阵为四维矩阵，因此，需将经度方向的匀速直线(constant velocate in longitude, CV-LNG)模型扩充至四维，扩充的维度用0补齐。CV-LNG模型的离散时间状态方程如下：

$$\boldsymbol{\alpha}(k+1)=\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CV}}(k)\boldsymbol{\alpha}(k)+\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CV}}(k)\boldsymbol{w}(k) \tag{3}$$

式中：

$$\boldsymbol{\alpha}(k)=[x(k),\dot{x}(k),0,0]^{\mathrm{T}}$$

$$\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CV}}(k)=\begin{bmatrix}1 & T & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

$$\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CV}}(k)=\begin{bmatrix}T^2/2\\ T\\ 0\\ 0\end{bmatrix}$$

T为采样周期。

目标沿纬度方向和高度方向的匀速直线运动模型, 建模原理与 CV-LNG 一致。因此, 纬度方向的匀速直线 (constant velocate in latitude, CV-LAT) 模型和高度方向匀速直线 (constant velocate in altitude, CV-QNE) 模型的离散状态方程分别为

$$\boldsymbol{\beta}(k+1)=\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CV}}(k)\boldsymbol{\beta}(k)+\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CV}}(k)\boldsymbol{w}(k) \tag{4}$$

$$\boldsymbol{\gamma}(k+1)=\boldsymbol{\Phi}_{\mathrm{QNE}}^{\mathrm{CV}}(k)\boldsymbol{\gamma}(k)+\boldsymbol{\Gamma}_{\mathrm{QNE}}^{\mathrm{CV}}(k)\boldsymbol{w}(k) \tag{5}$$

式中: $\boldsymbol{\beta}(k)=\boldsymbol{\gamma}(k)=\left[y(k),\dot{y}(k),0,0\right]^{\mathrm{T}}$; $\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CV}}(k)=\boldsymbol{\Phi}_{\mathrm{QNE}}^{\mathrm{CV}}(k)=\begin{bmatrix}1 & T & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$; $\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CV}}(k)=\boldsymbol{\Gamma}_{\mathrm{QNE}}^{\mathrm{CV}}(k)=\begin{bmatrix}T^2/2\\ T\\ 0\\ 0\end{bmatrix}$。

1.2.2　匀加速直线飞行

根据 CV 模型的建模原理, 对目标的匀加速直线 (constant accelerated, CA) 模型在 3 个方向上分别建模。目标匀加速直线运动模型的离散时间状态方程为

$$\boldsymbol{\alpha}(k+1)=\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CA}}(k)\boldsymbol{\alpha}(k)+\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CA}}(k)\boldsymbol{w}(k) \tag{6}$$

$$\boldsymbol{\beta}(k+1)=\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CA}}(k)\boldsymbol{\beta}(k)+\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CA}}(k)\boldsymbol{w}(k) \tag{7}$$

$$\boldsymbol{\gamma}(k+1)=\boldsymbol{\Phi}_{\mathrm{QNE}}^{\mathrm{CA}}(k)\boldsymbol{\gamma}(k)+\boldsymbol{\Gamma}_{\mathrm{QNE}}^{\mathrm{CA}}(k)\boldsymbol{w}(k) \tag{8}$$

式中: $\boldsymbol{\alpha}(k)=[x(k),\dot{x}(k),\ddot{x}(k),0]^{\mathrm{T}}$; $\boldsymbol{\beta}(k)=\left[y(k),\dot{y}(k),\ddot{y}(k),0\right]^{\mathrm{T}}$; $\boldsymbol{\gamma}(k)=[h(k),\dot{h}(k),\ddot{h}(k),0]^{\mathrm{T}}$; $\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CA}}(k)=\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CA}}(k)=\boldsymbol{\Phi}_{\mathrm{QNE}}^{\mathrm{CA}}(k)=\begin{bmatrix}1 & T & T^2/2 & 0\\ 0 & 1 & T & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$; $\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CA}}(k)=\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CA}}(k)=\boldsymbol{\Gamma}_{\mathrm{QNE}}^{\mathrm{CA}}(k)=\begin{bmatrix}T^2/2\\ T\\ 1\\ 0\end{bmatrix}$。

1.2.3　“当前”运动

“当前”模型是加速度均值非 0 的一阶时间相关模型, 其目标航空器在下一时刻的加速度取值是基于“当前”加速度的邻域。“当前”模型能够真实地反映目标航空器的机动特性, 但目标加速度在$\left[\frac{4-\pi}{4}a_{\max},a_{\max}\right]$和$\left[\frac{4-\pi}{4}a_{\min},a_{\min}\right]$区间外时, 跟踪效果较差。因此, “当前”模型不适用于跟踪弱机动的目标航空器。

假设目标在 3 个方向上的运动遵循可变加速度, 然后, 基于“当前”统计模型建立目标的可变加速度运动模型, 目标沿经度方向的当前统计 (current statistical in longitude, CS-LNG) 模型、纬度方向的当前统计 (current statistical in latitude, CS-LAT) 模型和高度方向的当前统计模型 (current statistical in altitude, CS-QNE) 模型, 离散状态方程分别为

$$\boldsymbol{\alpha}(k+1)=\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CS}}(k)\boldsymbol{\alpha}(k)+\boldsymbol{\Psi}_{\mathrm{LNG}}^{\mathrm{CS}}(k)\bar{a}(k)+\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CS}}(k)\boldsymbol{w}(k) \tag{9}$$

$$\boldsymbol{\beta}(k+1)=\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CS}}(k)\boldsymbol{\beta}(k)+\boldsymbol{\Psi}_{\mathrm{LAT}}^{\mathrm{CS}}(k)\bar{a}(k)+\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CS}}(k)\boldsymbol{w}(k) \tag{10}$$

$$\boldsymbol{\gamma}(k+1)=\boldsymbol{\Phi}_{\mathrm{QNE}}^{\mathrm{CS}}(k)\boldsymbol{\gamma}(k)+\boldsymbol{\Psi}_{\mathrm{QNE}}^{\mathrm{CS}}(k)\bar{a}(k)+\boldsymbol{\Gamma}_{\mathrm{QNE}}^{\mathrm{CS}}(k)\boldsymbol{w}(k) \tag{11}$$

式中: $\boldsymbol{\alpha}(k)=[x(k),\dot{x}(k),\ddot{x}(k),0]^{\mathrm{T}}$; $\boldsymbol{\beta}(k)=\left[y(k),\dot{y}(k),\ddot{y}(k),0\right]^{\mathrm{T}}$; $\boldsymbol{\gamma}(k)=\left[h(k),\dot{h}(k),\ddot{h}(k),0\right]^{\mathrm{T}}$; $\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CS}}(k)=\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CS}}(k)=\boldsymbol{\Phi}_{\mathrm{QNE}}^{\mathrm{CS}}(k)=\begin{bmatrix}1 & T & (\lambda T+\mathrm{e}^{-\lambda T}-1)/\lambda^2 & 0\\ 0 & 1 & (1-\mathrm{e}^{-\lambda T})/\lambda & 0\\ 0 & 0 & \mathrm{e}^{-\lambda T} & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$; $\boldsymbol{\Psi}$为系统输入控制矩阵, $\boldsymbol{\Psi}_{\mathrm{LNG}}^{\mathrm{CS}}(k)=\boldsymbol{\Psi}_{\mathrm{LAT}}^{\mathrm{CS}}(k)=\boldsymbol{\Psi}_{\mathrm{QNE}}^{\mathrm{CS}}(k)=\begin{bmatrix}-T/\lambda+T^2/2+\left(1-\mathrm{e}^{-\lambda T}\right)/\lambda^2\\ T-\left(1-\mathrm{e}^{-\lambda T}\right)/\lambda\\ 1-\mathrm{e}^{-\lambda T}\\ 0\end{bmatrix}$; $\boldsymbol{\Gamma}_{\mathrm{LNE}}^{\mathrm{CS}}(k)=\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CS}}(k)=\boldsymbol{\Gamma}_{\mathrm{QNE}}^{\mathrm{CS}}(k)=\begin{bmatrix}T^2/2\\ T\\ 1\\ 0\end{bmatrix}$; $\bar{a}(k)$为当前加速度的估计值。

1.2.4　匀角速度转弯

与基于随机过程的模型不同, 转弯模型的建立主要基于目标的运动学特性, 以便更加准确地描述目标的空间轨迹。考虑到民航飞机在垂直方向很少发生机动, 因此, 本节转弯模型采用描述水平方向运动的二维模型, 垂直方向的跟踪滤波采用 CV-QNE、CA-QNE 和 CS-QNE 模型。

由于航空器的转弯机动属于非线性运动, 在状态方程离散化的过程中会造成经度方向和纬度方向状态相互耦合。因此, 在建立经度和纬度方向的转弯模型时, 要对目标状态进行解耦。

1) 经度方向 CT-LNG, 目标航空器经度方向的离散状态转移方程可以表示为

$$\boldsymbol{\alpha}(k+1)=\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CT}}(k)\boldsymbol{\alpha}(k)+\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CT}}(k)\boldsymbol{w}(k) \tag{12}$$

式中:

$$\boldsymbol{\alpha}(k)=[x,\dot{x},0,0]$$

$$\boldsymbol{\Phi}_{\mathrm{LNG}}^{\mathrm{CT}}(k)=\begin{bmatrix}1 & \sin(\tau T)/\tau & 0 & 0\\ 0 & \cos(\tau T) & 0 & 0\\ 0 & (1-\cos(\tau T))/\tau & 0 & 0\\ 0 & \sin(\tau T) & 0 & 0\end{bmatrix}$$

$$\boldsymbol{\Gamma}_{\mathrm{LNG}}^{\mathrm{CT}}(k)=\begin{bmatrix}T^2/2\\T\\0\\0\end{bmatrix}$$

2）纬度方向CT-LAT，目标航空器纬度方向的离散状态方程与CT-LNG一致：

$$\boldsymbol{\beta}(k+1)=\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CT}}(k)\boldsymbol{\alpha}(k)+\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CT}}(k)\boldsymbol{w}(k) \tag{13}$$

式中：$\boldsymbol{\beta}(k)=[y,\dot{y},0,0]$；$\boldsymbol{\Phi}_{\mathrm{LAT}}^{\mathrm{CT}}(k)=\begin{bmatrix}0&0&0&(\cos(\tau T)-1)/\tau\\0&0&0&-\sin(\tau T)\\0&0&1&(\sin(\tau T))/\tau\\0&0&0&\cos(\tau T)\end{bmatrix}$；$\boldsymbol{\Gamma}_{\mathrm{LAT}}^{\mathrm{CT}}(k)=\begin{bmatrix}T^2/2\\T\\0\\0\end{bmatrix}$。

1.2.5 模型先验信息

ADS-B是将GPS设备生成的精确定位和机载飞行管理系统存储的运动状态等信息进行广播，其他空地终端通过数据链接收该广播并进行解码分析，即可实现对空域的监视。本文通过对目标航空器的ADS-B空中位置和速度报文，以及目标状态和情况报文进行捕获和解码，获得航空器当前状态和意图等信息作为变结构交互多模型的先验信息。报文结构和各字段含义可见RTCA DO-260B文件[34]，报文56 bit的ME字段结构如图1所示。

通过对空中位置的解码，可以获得目标航空器的经度、纬度和高度信息，并作为观测值用于更新当前目标航空器位置状态。此外，通过空中速度报文解析出的地速、航向和垂直速率信息经过投影可以转换为3个方向的速度信息。地心坐标系下的观测方程为

$$\boldsymbol{Z}(k)=\boldsymbol{H}(k)\boldsymbol{X}(k)+\boldsymbol{V}(k) \tag{14}$$

式中：$\boldsymbol{H}(k)$为观测矩阵，$\boldsymbol{H}(k)=\begin{bmatrix}1&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&1&0\end{bmatrix}$；$\boldsymbol{V}(k)$为观测噪声，服从$\boldsymbol{V}(k)\sim N(\boldsymbol{0},\boldsymbol{R}_k)$；$\boldsymbol{R}_k$为观测噪声协方差矩阵。

民航飞机一般会在到达特殊航路点或空管指定航路点之前发送空中位置和情报电文，特别是当这些航路点代表重大的航路变化或飞行计划更新时。该信息将包含有关飞机当前状态和飞行意图信息，并将以1.2～1.3 s范围内均匀分布的随机间隔进行广播，以确保其他飞机和地面控制站能够了解飞机的预期机动。通过对目标状态和情况电文的解码可以获得飞行管理系统（flight management system, FMS）选择高度和选择航向数据，该数据可以作为目标航空器的飞行意图信息反映目标在未来一段时间内高度和航向的变化趋势，并可以作为模型集切换的重要依据。

报文类型：5 bit	监视状态：2 bit	附加信息：1 bit	高度：12 bit	时间：1 bit	紧凑格式：1 bit	纬度：17 bit	经度：17 bit

(a) 空中位置报文

报文类型：5 bit	子类型：3 bit	意图改变：1 bit 标识	预留：1 bit	速度导航：3 bit 精度	东/西方向：1 bit	东/西速度：10 bit	南/北方向：1 bit	南/北速度：10 bit
垂直速率：1 bit 来源	垂直速率：1 bit 方向	垂直速率：9 bit	预留：2 bit	气压高度：1 bit 差异	气压高度：7 bit 差值			

(b) 空中速度报文

报文类型：5 bit	子类型：2 bit	数据源完整性补充：1 bit	选择高度：1 bit 类型	选择高度：11 bit	设定气压高度：9 bit	选择航向：1 bit 状态	选择航向：1 bit 符号	选择航向：8 bit
bit置导航精度：1 bit	气压高度完整性：1 bit	数据源完整性：2 bit	高度模式：1 bit	自动驾驶：1 bit	垂直导航模式：1 bit	高度保持模式：1 bit	预留：1 bit	进近模式：1 bit
TCAS运行：1 bit	水平导航模式：1 bit	预留：2 bit						

(c) 目标状态和情况报文

图1 报文解析结果

Fig. 1 Message analysis result

2 模型集自适应切换算法

本文算法包含2个功能部分：①模型集的自适应切换；②模型序列条件估计。本节采用有向图切换算法，将每个模型集作为有向图的一个顶点，利用飞行意图等先验信息和模型集似然概率实现模型集的自适应切换。

2.1 基于有向图的模型集构成

基于有向图的模型切换算法是在能够覆盖目标各种运动模式的模型集合中，根据先验信息或当

前时刻的目标运动状态进行有向集之间的切换。

本文将航空器的运动模式分解为恒定高度下的直线飞行和转弯飞行、恒定航向升降, 以及在升降过程中转弯。以上 4 种飞行模式涵盖了民航飞行器在正常飞行阶段的所有飞行动作, 每种飞行模式采用 6 个子模型进行数学描述。运动模式与模型子集的对应关系如表 1 所示。

表 1　目标运动状态与模型子集的关系

Table 1　Relationship between target motion states and model sub-set

模型集U	运动模式	模型子集C_i
U_1	恒定高度直线飞行	C_{ALNG}、C_{VLNG}、C_{SLNG}、C_{ALAT}、C_{VLAT}、C_{SLAT}
U_2	恒定高度转弯飞行	C_{ALNG}、C_{TLNG}、C_{SLNG}、C_{ALAT}、C_{TLAT}、C_{SLAT}
U_3	恒定航向升降	C_{ALNG}、C_{ALAT}、C_{AQNE}、C_{SLNG}、C_{SLAT}、C_{SQNE}
U_4	升降过程中转弯	C_{TLNG}、C_{TLAT}、C_{VQNE}、C_{SLNG}、C_{SLAT}、C_{SQNE}

目标的状态判断采用事件驱动和模型概率驱动相结合的方式, 充分利用目标先验信息, 以达到更好地跟踪效果。经过状态判断后, 当前模型集将由上一时刻的模型集自适应切换到与目标运动状态更接近的模型集。模型集的有向图示意图如图 2 所示。

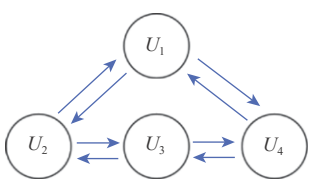


图 2　有向图示意图

Fig. 2　Schematic diagram of directed graph

2.2　模型集切换规则

目前, 常用的有向图切换算法是根据上一时刻系统对目标运动状态的判断, 将当前模型集切换为更加适配目标运动状态的新模型集, 以减小跟踪误差。这种模型集切换算法本质上存在一定的时间滞后, 即需要在目标发生机动的下一时刻才能够做出调整。本文通过对目标状态和情况报文进行解码, 可以提取到目标航空器的选择高度和选择航向等飞行意图信息。航空器的飞行意图反映了航空器在未来一段时间内预期的飞行状态, 其相对于模型的预测结果显然更加准确, 利用飞行意图作为先验信息指导模型集的切换无疑会大大提高模型跟踪精度, 改善固有时滞。因此, 本节融合航空器的飞行意图信息, 提出一种基于目标航空器选择高度和选择航向信息的模型集切换规则, 实现模型集的快速、准确切换。

采用飞行意图驱动模型集切换是一种“硬”切换规则, 虽然可以简单、直接地提高模型计算效率和预测精度, 但在跟踪目标航空器的过程中, 飞行意图信息的获取并不连续。因此, 提供一种“软”切换规则以作为模型集切换的补充手段是必要的。本节通过计算每个模型集的似然概率来匹配当前最优模型, 可以实现模型集的“软”切换。模型集切换原理如图 3 所示。

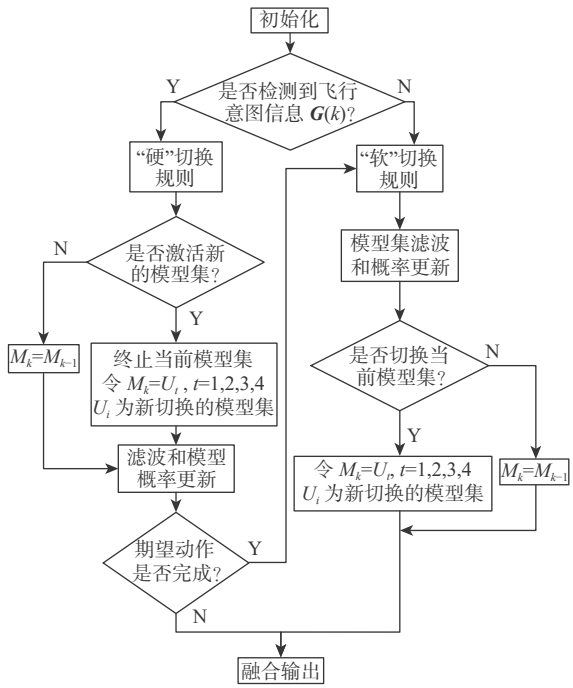


图 3　模型集切换原理图

Fig. 3　Schematic diagram of directed graph

假设系统在 k 时刻获取到飞行意图信息为$\boldsymbol{G}(k)=\left[h(k),\varphi(k)\right]$, 其中, $h(k)$表示选择高度信息, $\varphi(k)$表示选择航向信息, $\boldsymbol{G}(k)$为意图矩阵。基于飞行意图的“硬”切换准则规定如下:

$$M_k=\begin{cases}U_2 & h(k)=0,\varphi(k)\neq 0\\ U_3 & h(k)\neq 0,\varphi(k)=0\\ U_4 & h(k)\neq 0,\varphi(k)\neq 0\end{cases} \tag{15}$$

上述“硬”切换准则可以作为模型集的激活条件用于激活一个单独的模型集, 但何时终止该模型集则需要根据目标是否已完成期望动作作为终止条件。换而言之, 当滤波结果未达到意图设定值时, 说明目标尚未完成期望动作, 此时对滤波结果进行融合输出; 当滤波结果已达到意图设定值时, 说明目标已完成期望动作, 基于意图的“硬”切换结果将不再适用, 此时将采用基于似然概率的软切换方法求解最优模型集。

基于模型集似然概率的模型集切换算法要对所有模型集进行滤波估计, 并根据残差和其协方差计算模型集似然概率。以模型集 U_1 为例, 经过模

型滤波和模型概率更新后，利用残差和协方差计算得到模型集 U_1 的极大似然概率，与其他模型集的极大似然概率循环对比，随后，选择最大似然概率的模型集将其切换为当前模型集，最后，将该模型集的滤波结果混合输出。本文算法相对于传统VSMM算法的优势在于，仅通过一次滤波即可完成模型集的切换和状态估计的输出，大大缩短了计算时间。假设模型集似然概率 Λ_{U_t} 为模型集中每个子模型似然概率 Λ_{C_i} 的累加和，表示为

$$\Lambda_{U_t}(k)=\sum_{C_i\in U_t}\Lambda_{C_i}(k)/\|U_t\| \qquad t\in\{1,2,3,4\} \tag{16}$$

$$\Lambda_{C_i}(k)=\frac{1}{|2\pi \boldsymbol{S}_i(k)|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\boldsymbol{d}_i^{\mathrm{T}}(k)\boldsymbol{S}_i^{-1}(k)\boldsymbol{d}_i(k)\right) \tag{17}$$

式中：$\boldsymbol{d}_i(k)$ 和 $\boldsymbol{S}_i(k)$ 分别表示 k 时刻的残差向量和协方差，其数学表达式将在第3节中进行介绍。

3 目标航空器的航迹跟踪与外推

3.1 本文 INT-VSMM 算法

将第2节讨论的模型集自适应切换算法融入模型集序列条件估计中，对目标航空器的运动模式进行识别。本文INT-VSMM算法由以下4部分组成：建立完备的模型集、选择最优模型集、模型滤波估计、输出交互。本文算法流程如图4所示。

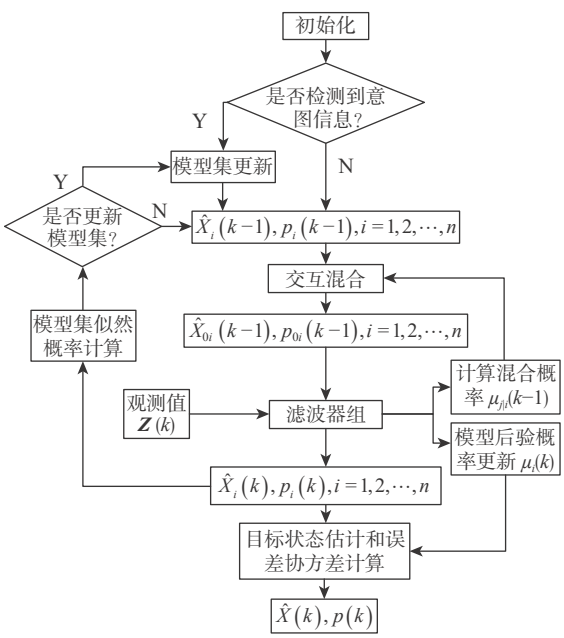


图4 本文算法流程

Fig. 4 Flow of the proposed algorithm

基于INT-VSMM滤波算法的具体步骤如下。

步骤1 模型初始化。

1）建立模型集。建立足以覆盖目标航空器所有可能运动的完备模型集 U。

2）为子模型分配转移概率。假设 k 时刻模型集 M_k 包含 n 个子模型，则从模型 C_i 转移到 C_j（$C_i\in M_k$，$C_j\in M_k$）的模型转移概率矩阵：

$$[p_{ji}]=\boldsymbol{P}\{C_j(k)|C_i(k-1)\} \tag{18}$$

则 k 时刻从模型 C_i 转移到 C_j 的概率可表示为

$$u_j(k-1)=p_{ji}u_i(k-1) \tag{19}$$

步骤2 根据第2节介绍的“硬”切换规则判断选择模型集进行滤波估计。

步骤3 模型滤波估计。

1）滤波初始化。模型 C_i 到模型 C_j 的混合概率：

$$u_{ji}(k-1|k-1)=\frac{1}{c_j(k)}p_{ji}u_i(k-1) \tag{20}$$

式中：$c_j(k)$ 为由其他模型转移到模型的概率。

k–1 时刻，模型 C_j 的混合状态估计：

$$\hat{\boldsymbol{x}}_{0j}(k-1|k-1)=\sum_{i=1}^{n}\hat{\boldsymbol{x}}_i(k-1|k-1)u_{ji}(k-1|k-1) \tag{21}$$

模型 C_j 的混合协方差估计：

$$\begin{aligned}\boldsymbol{P}_{0j}(k-1|k-1)=&\sum_{i=1}^{n}u_{ji}(k-1|k-1)\cdot\big(\boldsymbol{P}_i(k-1|k-1)+\\&\left[\hat{\boldsymbol{x}}_i(k-1|k-1)-\hat{\boldsymbol{x}}_{0j}(k-1|k-1)\right]\cdot\\&\left[\hat{\boldsymbol{x}}_i(k-1|k-1)-\hat{\boldsymbol{x}}_{0j}(k-1|k-1)\right]^{\mathrm{T}}\big)\end{aligned} \tag{22}$$

式中：$\hat{\boldsymbol{x}}_i(k-1|k-1)$ 为 $k-1$ 时刻模型 C_i 对应子滤波器输出的状态估计；$\boldsymbol{P}_i(k-1|k-1)$ 为 $k-1$ 时刻，模型 C_i 对应子滤波器输出的协方差矩阵的估计；$\hat{\boldsymbol{x}}_{0j}(k-1|k-1)$ 为将 $k-1$ 时刻各个子滤波器输出的状态估计进行加权求和得到的 k 时刻模型 C_j 对应的子滤波器的状态输入，权值为各个子模型到模型 C_j 的混合概率；$\boldsymbol{P}_{0j}(k-1|k-1)$ 为将 $k-1$ 时刻各个子滤波器输出的协方差矩阵估计进行加权求和得到的 k 时刻模型 C_j 对应的子滤波器的协方差矩阵输入；$u_{ji}(k-1|k-1)$ 表示 $k-1$ 时刻模型匹配 C_i，k 时刻模型匹配 C_j，以信息 $\boldsymbol{Z}(k)$ 为条件的混合概率：

$$\begin{aligned}u_{ji}(k-1|k-1)=&p\left(C_i(k-1)\middle|C_j(k),\boldsymbol{Z}(k)\right)=\\&p_{ji}\cdot u_i(k-1)\Big/\sum_{i=1}^{n}(p_{ji}\cdot u_i(k-1))\end{aligned} \tag{23}$$

2）模型条件滤波。本文将目标在空间上的非线性运动分解为经度、纬度和高度3个方向上的线性运动，随后为避免不同方向的模型预测混乱，在3个方向上分别采用该方向的卡尔曼滤波器对状态输入进行滤波。

状态一步预测和一步预测的协方差为

$$\hat{\boldsymbol{x}}_j(k|k-1)=\boldsymbol{\Phi}_j(k-1)\hat{\boldsymbol{x}}_{0j}(k-1|k-1) \tag{24}$$

$$\boldsymbol{P}_j(k|k-1)=\boldsymbol{\Phi}_j\boldsymbol{P}_{0j}(k-1|k-1)\boldsymbol{\Phi}_j^{\mathrm{T}}+\boldsymbol{\Gamma}_j\boldsymbol{Q}_j\boldsymbol{\Gamma}_j^{\mathrm{T}} \tag{25}$$

残差信息向量$\boldsymbol{d}_j(k)$和残差协方差$\boldsymbol{S}_j(k)$为

$$\boldsymbol{d}_j(k)=\boldsymbol{Z}(k)-\boldsymbol{H}(k)\hat{\boldsymbol{x}}_j(k|k-1) \tag{26}$$

$$\boldsymbol{S}_j(k)=\boldsymbol{H}(k)\boldsymbol{P}_j(k|k-1)\boldsymbol{H}^{\mathrm{T}}(k)+\boldsymbol{R}(k) \tag{27}$$

卡尔曼滤波的增益矩阵为

$$\boldsymbol{K}_j(k)=\boldsymbol{P}_j(k|k-1)\boldsymbol{H}^{\mathrm{T}}\left(\boldsymbol{H}\boldsymbol{P}_j(k|k-1)\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{R}\right)^{-1} \tag{28}$$

使用量测信息中的观测数据对目标状态和协方差进行更新:

$$\hat{\boldsymbol{x}}_j(k|k)=\hat{\boldsymbol{x}}_j(k|k-1)+\boldsymbol{K}_j(k)\left(\boldsymbol{Z}(k)-\boldsymbol{H}(k)\hat{\boldsymbol{x}}_j(k|k-1)\right) \tag{29}$$

$$\boldsymbol{P}_j(k|k)=\boldsymbol{P}_j(k|k-1)-\boldsymbol{K}_j(k)\boldsymbol{S}_j(k)\boldsymbol{K}_j^{\mathrm{T}}(k) \tag{30}$$

步骤 4　最优模型集切换。

1) 极大似然概率更新。模型似然函数$\Lambda_j(k)$更新的表达式如式(22)所示, 将这一步计算得到的模型似然函数带入到步骤 2 用于模型集的切换。

2) 模型概率更新:

$$u_j(k)=\Lambda_j(k)c_j(k)\Big/\sum_{j=1}^{n}(\Lambda_j(k)c_j(k)) \tag{31}$$

式中: $c_j(k)=\sum_{i=1}^{n}p_{ji}u_i(k)$表示由其他模型转移到模型$m_j$的概率。

3) 选择最优模型集。根据第 2 节介绍的“软”切换规则选择最优模型集。

步骤 5　输出交互。

滤波器的融合输出为对最优模型集中各子模型进行滤波后, 滤波估计结果的加权值, 加权的状态估计$\hat{\boldsymbol{x}}(k|k)$和协方差$\boldsymbol{P}(k|k)$为

$$\hat{\boldsymbol{x}}(k|k)=\sum_{j=1}^{n}\hat{\boldsymbol{x}}_j(k|k)\cdot u_j(k) \tag{32}$$

$$\boldsymbol{P}(k|k)=\sum_{j=1}^{n}u_j(k)\left(\boldsymbol{P}_j(k|k)+\boldsymbol{\delta}_j\cdot\boldsymbol{\delta}_j^{\mathrm{T}}\right) \tag{33}$$

式中: $\boldsymbol{\delta}_j=\hat{\boldsymbol{x}}_j(k|k)-\hat{\boldsymbol{x}}(k|k)$为滤波校正差, 表示模型$m_j$的状态滤波估计值与综合输出的差。

3.2　基于 INT-VSMM 的轨迹外推

根据 3.1 节的计算, 可以跟踪目标航空器在当前时刻的运动状态, 但为能够提前预测出潜在的飞行冲突, 需要对目标航空器的航迹进行外推。由于目标的状态估计会随着观测量不断更新, 因此, 本节在实现目标航空器精确跟踪的基础上, 只对目标航空器的轨迹进行短期外推, 为航空器冲突探测提供数据支撑。

从 3.1 节步骤 3 可知, 模型状态的一步预测结果只与状态向量$\hat{\boldsymbol{x}}(k)$、方差$\boldsymbol{P}(k)$和模型概率$u(k)$相关, 其更新不受观测值的影响。因此, 本文利用某时刻模型状态滤波估计值进一步预测, 从而外推目标航空器下一时刻的飞行状态, 该外推结果将作为输入用于接下来的外推计算。假设当前时刻为k, 由全概率公式可以得到$k+1$时刻子模型C_j的概率:

$$u_j(k+1)=\sum_{i=1}^{n}p_{ji}u_i(k) \tag{34}$$

变结构交互多模型的状态预测值和预测滤波均方值:

$$\hat{\boldsymbol{x}}(k+1)=\sum_{j=1}^{n}\hat{\boldsymbol{x}}_j(k+1)u_j(k) \tag{35}$$

$$P(k+1)=\sum_{i=1}^{n}u_j(k)\left\{P_i(k)+\left[\hat{\boldsymbol{x}}_i(k)-\hat{\boldsymbol{x}}_{0j}(k)\right]\cdot\left[\hat{\boldsymbol{x}}_i(k)-\hat{\boldsymbol{x}}_{0j}(k)\right]^{\mathrm{T}}\right\} \tag{36}$$

式中: $\hat{\boldsymbol{x}}_j(k+1)=\boldsymbol{\Phi}_j(k)\hat{\boldsymbol{x}}_{0j}(k)$为$k+1$时刻模型$C_j$的估计值。

4　仿真实验

4.1　仿真环境及参数设定

本文截取 2022 年 12 月 5 日, 航班号 CXA8830 的飞机在爬升阶段 1 400 s 内的 ADS-B 报文数据进行滤波。首先, 将捕获的报文数据去重并删除异常点, 得到 1 209 个轨迹点, 然后, 以接收到空中位置报文为k时刻, 下一次接收到空中位置报文为k+1 时刻, 离散化时间步长的持续时间为T。在空中位置报文接收周期内将解码出的位置信息和速度信息插入观测矩阵, 将解码出的意图信息插入意图矩阵。根据奈奎斯特采样定理, 采样频率应该至少是信号最高频率的 2 倍。因此, 本文设定$T=0.6$ s。

仿真在 AMD Ryzen 3 2200G 处理器, 主频 3.5 GHz、64 位 Windows10 下, 通过仿真软件 MATLAB 2021b 对本文 INT-VSMM 算法进行仿真计算, 并与传统 VSMM 算法、IMM 算法进行比较实验, 参数设置如表 2 所示。

各模型的过程噪声协方差和观测噪声协方差矩阵$\boldsymbol{\Gamma}(k)=\mathrm{diag}\left(0.1^2,0.1^2,0.1^2,0.1^2\right)$, $\boldsymbol{R}(k)=\mathrm{diag}(2\,500,2\,500,2\,500,2\,000,2\,000,2\,000)$。

4.2　仿真结果及分析

本文设置常规 VSMM 算法采用单一的观测数据进行跟踪, 而 INT-VSMM 算法在目标飞机运行至 873 s 处, 接收意图数据(选择高度、选择航向)。结合传统 IMM 算法, 3 种滤波算法的跟踪轨迹与目标真实轨迹的对比曲线如图 5 所示。

由图 5 可知, 3 种算法都能够较好地实现目标

表 2 模型初始参数

Table 2 Model initial parameters

模型集	模型子集	目标初始状态	初始模型概率	初始模型转移概率
U_1	经度方向$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}_0=[x_0,\dot{x}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\alpha}=[0.3\quad 0.4\quad 0.3]$	$\boldsymbol{p}_{ji}^{\alpha}=\begin{bmatrix}0.8&0.1&0.1\\0.1&0.8&0.1\\0.1&0.1&0.8\end{bmatrix}$
	纬度方向$\boldsymbol{\beta}$	$\boldsymbol{\beta}_0=[y_0,\dot{y}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\beta}=[0.3\quad 0.4\quad 0.3]$	$\boldsymbol{p}_{ji}^{\beta}=\begin{bmatrix}0.8&0.1&0.1\\0.1&0.8&0.1\\0.1&0.1&0.8\end{bmatrix}$
	高度方向$\boldsymbol{\gamma}$			
U_2	经度方向$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}_0=[x_0,\dot{x}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\alpha}=[0.2\quad 0.6\quad 0.2]$	$\boldsymbol{p}_{ji}^{\alpha}=\begin{bmatrix}0.8&0.1&0.1\\0.1&0.8&0.1\\0.1&0.1&0.8\end{bmatrix}$
	纬度方向$\boldsymbol{\beta}$	$\boldsymbol{\beta}_0=[y_0,\dot{y}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\beta}=[0.2\quad 0.6\quad 0.2]$	$\boldsymbol{p}_{ji}^{\beta}=\begin{bmatrix}0.8&0.1&0.1\\0.1&0.8&0.1\\0.1&0.1&0.8\end{bmatrix}$
	高度方向$\boldsymbol{\gamma}$			
U_3	经度方向$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}_0=[x_0,\dot{x}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\alpha}=[0.5\quad 0.5]$	$\boldsymbol{p}_{ji}^{\alpha}=\begin{bmatrix}0.8&0.2\\0.1&0.9\end{bmatrix}$
	纬度方向$\boldsymbol{\beta}$	$\boldsymbol{\beta}_0=[y_0,\dot{y}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\beta}=[0.5\quad 0.5]$	$\boldsymbol{p}_{ji}^{\beta}=\begin{bmatrix}0.8&0.2\\0.1&0.9\end{bmatrix}$
	高度方向$\boldsymbol{\gamma}$	$\boldsymbol{\gamma}_0=[h_0,\dot{h}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\gamma}=[0.5\quad 0.5]$	$\boldsymbol{p}_{ji}^{\gamma}=\begin{bmatrix}0.8&0.2\\0.1&0.9\end{bmatrix}$
U_4	经度方向$\boldsymbol{\alpha}$	$\boldsymbol{\alpha}_0=[x_0,\dot{x}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\alpha}=[0.5\quad 0.5]$	$\boldsymbol{p}_{ji}^{\alpha}=\begin{bmatrix}0.8&0.2\\0.1&0.9\end{bmatrix}$
	纬度方向$\boldsymbol{\beta}$	$\boldsymbol{\beta}_0=[y_0,\dot{y}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\beta}=[0.5\quad 0.5]$	$\boldsymbol{p}_{ji}^{\beta}=\begin{bmatrix}0.8&0.2\\0.1&0.9\end{bmatrix}$
	高度方向$\boldsymbol{\gamma}$	$\boldsymbol{\gamma}_0=[h_0,\dot{h}_0,0,0]^{\mathrm{T}}$	$\boldsymbol{u}_i^{\gamma}=[0.5\quad 0.5]$	$\boldsymbol{p}_{ji}^{\gamma}=\begin{bmatrix}0.8&0.2\\0.1&0.9\end{bmatrix}$

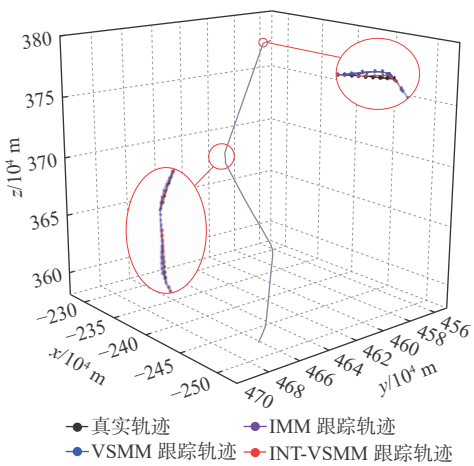


图 5 目标跟踪轨迹曲线

Fig. 5 Curves of target tracking trajectory

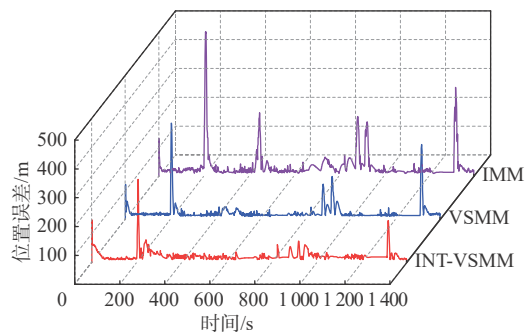


图 6 位置跟踪误差对比

Fig. 6 Comparison of position tracking error

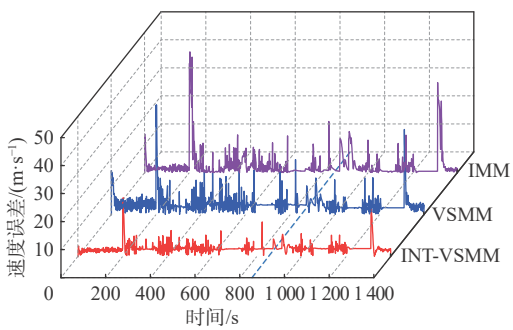


图 7 速度跟踪误差对比

Fig. 7 Comparison of velocity tracking error

跟踪，IMM 和 VSMM 算法在目标运动模式变化时跟踪效果相差不大，而 INT-VSMM 模型跟踪效果明显优于前两者，与真实轨迹更接近。

图 6 和图 7 分别为 3 种滤波算法的位置跟踪误差曲线图和速度跟踪误差曲线。由图可知，在目标航空器运动模式发生变化时，3 种滤波算法的跟踪误差曲线都会出现明显的振荡，这主要是由于模型集或子模型的切换延迟造成模型失配，以及模型预测混乱导致的。INT-VSMM 算法在 873 s 处由于接收到目标航空即将调整航向和高度的意图，其跟踪曲线的振荡明显小于其余 2 种算法，其中，位置和速度最大误差相对于传统 IMM 算法分别提升了 42.44% 和 22.01%，相对于常规 VSMM 算法分别提升了 13.54% 和 8.87%；位置和速度均值误差相对于传统 IMM 算法分别提升了 23.57% 和 25.30%，相对

于常规 VSMM 算法分别提升了 6.84% 和 18.50%。3 种滤波算法的跟踪均值误差和最大误差如表 3 所示, INT-VSMM 算法的各项指标均为最优。

表 3　算法位置和速度误差性能对比

Table 3　Performance comparison of algorithm position and velocity error

算法	最大位置跟踪误差/m	最大速度跟踪误差/(m·s^{-1})	位置误差均值/m	速度误差均值/(m·s^{-1})
IMM[11]	502.212	75.984	44.692	13.077
VSMM[21]	334.424	65.026	34.157	11.985
INT-VSMM	289.142	59.258	31.822	9.768

基于多模型集架构的 VSMM 算法在任意时刻都有多个模型集在运行, 因此, 它相对于固定结构的 IMM 滤波器在计算量上有本质的增加。针对该问题, 本文 INT-VSMM 算法利用意图数据减少了多余模型集的滤波计算, 并优化了滤波算法的流程, 仿真计算时间相对于 VSMM 算法降低了 12.08%。3 种滤波算法的仿真运行时间如表 4 所示。

表 4　轨迹跟踪仿真计算时间

Table 4　Simulation time of trajectory tracking

算法	计算时间/s
IMM[11]	3.522
VSMM[21]	4.900
INT-VSMM	4.308

INT-VSMM 算法在各时刻的最优模型集变化趋势如图 8 所示。图中,黑色虚线表示目标航空器运动的起始位置, 模型集沿顺时针方向变化。由图可知, 由于意图数据对模型集切换的指导作用, INT-VSMM 算法避免了在模式识别过程中模型集的反复切换, 从而提高了模型稳定性和适配性。

选取 $k = 368$ s 的滤波输出作为初始状态, 进行 160 s 的外推。图 9 和图 10 分别为外推轨迹与真实轨迹的对比和误差曲线。从图中外推轨迹可知, 本文 INT-VSMM 算法外推轨迹相对于传统 VSMM 算法具有更好的预测效果, 30 s 外推误差减小了 46%。但由于本文 INT-VSMM 算法只针对目标的轨迹进行短期外推, 外推误差会随着外推时间的延长而增大, 并不能很好地反映目标的长期运动趋势。如图 10 所示, 在 40 s 处外推误差约为 1 000 m。因此, 本文 INT-VSMM 算法适用于对目标进行短期外推, 其外推轨迹可以作为冲突探测的依据。

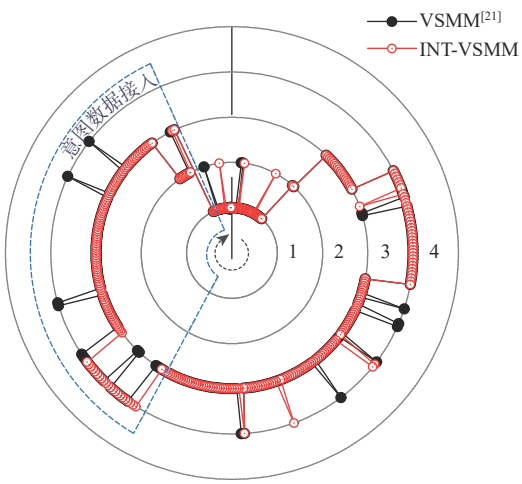


图 8　模型集随时间切换

Fig. 8　Model set switching over time

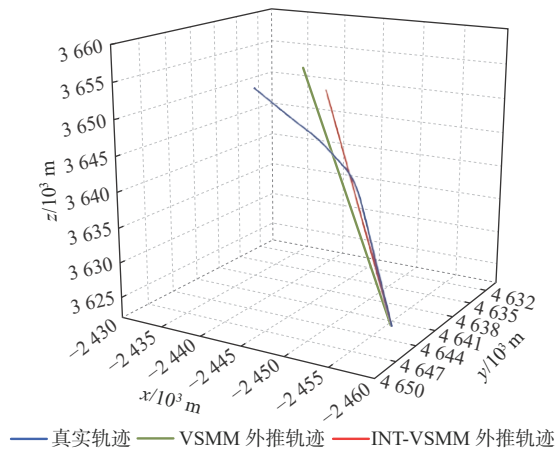


图 9　真实轨迹与外推轨迹曲线

Fig. 9　Real trajectory and extrapolated trajectory curves

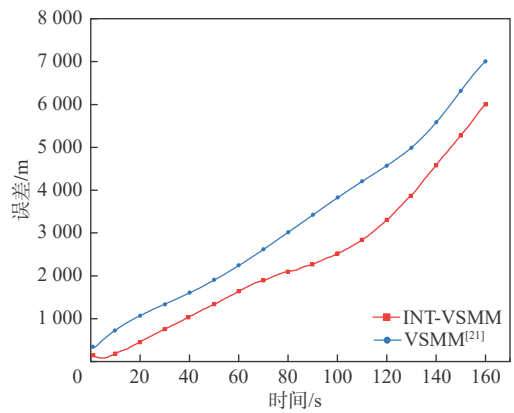


图 10　外推轨迹误差

Fig. 10　Error of extrapolated trajectory

5　结　论

针对航路飞行阶段的目标跟踪和短期航迹外推问题, 本文充分考虑了航空器的空域态势感知能力, 优化 VSMM 架构, 提出 INT-VSMM 算法。根据目标运动模式建立完备的运动模型集, 利用目标飞行意图和模型集似然概率实现各个模型集之间的切换, 并基于滤波估计结果进行了短期航迹外推。仿真结果证明:

1) INT-VSMM 算法相对于常规 VSMM 算法和 IMM 算法, 改善了模型集切换滞后和子模型失

配的问题，极大提高了目标跟踪精度和模型稳定性；

2）INT-VSMM算法的计算时间相对于常规VSMM算法降低了12.08%；

3）外推算法在短期内能够较好地反映目标未来航迹，对冲突探测和间隔保持等工作具有一定的指导作用。

本文INT-VSMM算法只考虑了单目标跟踪，然而实际飞行过程中时常需要对多个目标进行监视。在未来的工作中，将致力于研究多目标跟踪和先进运动模型的整合，进一步提升目标跟踪的性能。

参考文献（References）

[1] YU H C, FANG Z X, MURRAY A T, et al. A direction-constrained space-time prism-based approach for quantifying possible multi-ship collision risks[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(1): 131-141.

[2] 罗亚伦, 廖育荣, 李兆铭, 等. 高超目标强跟踪CKF自适应交互多模型跟踪算法[J]. 北京航空航天大学学报, 2024, 50(7): 2272-2283.
LUO Y L, LIAO Y R, LI Z M, et al. Strong tracking CKF adaptive interactive multiple model tracking algorithm based on hypersonic target[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(7): 2272-2283(in Chinese).

[3] LEE H, LEE H T. Extracting flight plans from recorded ADS-B trajectories[J]. International Journal of Aeronautical and Space Sciences, 2023, 24(2): 581-589.

[4] LI Y F, ZHANG Y, WANG L L, et al. Research on potential ground risk regions of aircraft crashes based on ADS-B flight tracking data and GIS[J]. Journal of Transportation Safety & Security, 2022, 14(1): 152-176.

[5] TONG L, GAN X S, WU Y R, et al. An ADS-B information-based collision avoidance methodology to UAV[J]. Actuators, 2023, 12(4): 165.

[6] WU Y, YU H Y, DU J P, et al. An aircraft trajectory prediction method based on trajectory clustering and a spatiotemporal feature network[J]. Electronics, 2022, 11(21): 3453.

[7] AKHTAR S, HABIBI S. The interacting multiple model smooth variable structure filter for trajectory prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 9217-9239.

[8] DAL SASSO V, DJEUMOU FOMENI F, LULLI G, et al. Planning efficient 4D trajectories in air traffic flow management[J]. European Journal of Operational Research, 2019, 276(2): 676-687.

[9] SEENIVASAN D B, OLIVARES A, STAFFETTI E. Multi-aircraft optimal 4D online trajectory planning in the presence of a multi-cell storm in development[J]. Transportation Research Part C: Emerging Technologies, 2020, 110: 123-142.

[10] AYHAN S, SAMET H. Aircraft trajectory prediction made easy with predictive analytics [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 21-30.

[11] BLOM H A P, BAR-SHALOM Y. The interacting multiple model algorithm for systems with Markovian switching coefficients[J]. IEEE Transactions on Automatic Control, 1988, 33(8): 780-783.

[12] 尹聚祺, 杨震, 罗亚中, 等. 空间机动目标跟踪的改进自适应IMM算法[J]. 系统工程与电子技术, 2021, 43(12): 3658-3666.
YIN J Q, YANG Z, LUO Y Z, et al. Improved adaptive IMM algorithm for space maneuvering target tracking[J]. Systems Engineering and Electronics, 2021, 43(12): 3658-3666(in Chinese).

[13] 游航航, 韩其松, 余敏建, 等. 基于AIGWO-IMMUKF的目标跟踪算法[J]. 北京航空航天大学学报, 2020, 46(10): 1826-1833.
YOU H H, HAN Q S, YU M J, et al. Target tracking algorithm based on AIGWO-IMMUKF[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(10): 1826-1833(in Chinese).

[14] SINGER R A. Estimating optimal tracking filter performance for manned maneuvering targets[J]. IEEE Transactions on Aerospace and Electronic Systems, 1970, 6(4): 473-483.

[15] NING J, CHEN J, WU Q. An adaptive filtering algorithm for singer model based on expectation model[J]. Telecommunication Engineering, 2022, 62(10): 1464-1469.

[16] DAI H, FANG J, TANG L, et al. Strong tracking UKF adaptive interacting multiple-model algorithm based on maneuvering hypersonic-target tracking[J]. Journal of Chinese Inertial Technology, 2018, 26(3): 338-345.

[17] ZHANG Z H, ZHOU G J. Maneuvering target state estimation based on separate modeling of target trajectory shape and dynamic characteristics[J]. Journal of Systems Engineering and Electronics, 2022, 33(5): 1195-1209.

[18] HOU Z W, BU F L. A small UAV tracking algorithm based on AIMM-UKF[J]. Aircraft Engineering and Aerospace Technology, 2021, 93(4): 579-591.

[19] ZHANG Z, GUO H W, HE J X, et al. Adaptive interactive multiple model target tracking algorithm based on Markov matrix with acceleration correction factor[C]//Proceedings of the China Automation Congress. Piscataway: IEEE Press, 2022: 3227-3232.

[20] 肖楚晗, 李炯, 雷虎民, 等. 基于AVSIMM算法的高超声速再入滑翔目标跟踪[J]. 北京航空航天大学学报, 2019, 45(2): 413-421.
XIAO C H, LI J, LEI H M, et al. Hypersonic reentry gliding target tracking based on AVSIMM algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(2): 413-421(in Chinese).

[21] LUKEŠ P, ŘÍHA M. Variable structure interacting multiple model for highly aperiodic sensors with poor measurement precision[J]. IET Radar, Sonar & Navigation, 2021, 15(10): 1181-1194.

[22] HAN S T, RONG Y J, PENG D L, et al. A novel variable structure multi-model approach based on error-ambiguity decomposition[J]. Chinese Journal of Aeronautics, 2020, 33(6): 1731-1746.

[23] 陈星, 李战武, 徐安, 等. 基于目标机动模式识别的VSMM算法[J]. 系统工程与电子技术, 2020, 42(5): 999-1006.
CHEN X, LI Z W, XU A, et al. VSMM algorithm based on target maneuver pattern recognition[J]. Systems Engineering and Electronics, 2020, 42(5): 999-1006(in Chinese).

[24] ZHANG B L, GAO Y X, DUAN Z S. Variable structure multiple model fixed-interval smoothing[J]. Chinese Journal of Aeronautics, 2023, 36(2): 139-148.

[25] GOSWAMI A, GEORGE LEE C S. Design of an interactive mul-

tiple model based two-stage multi-vehicle tracking algorithm for autonomous navigation[C]//Proceedings of the IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2015: 261-266.

[26] ZENG Y, LU W B, YU B, et al. Improved IMM algorithm based on support vector regression for UAV tracking[J]. Journal of Systems Engineering and Electronics, 2022, 33(4): 867-876.

[27] 李松, 程咏梅, 王会宾, 等. 时间偏差校准分布式多传感器多目标跟踪算法[J]. 西北工业大学学报, 2020, 38(4): 797-805.
LI S, CHENG Y M, WANG H B, et al. Distributed multisensor multitarget tracking algorithm with time-offset registration[J]. Journal of Northwestern Polytechnical University, 2020, 38(4): 797-805 (in Chinese).

[28] WANG D Z, HUANG D Q, XU C. An improved method for two-UAV trajectory planning for cooperative target locating based on airborne visual tracking platform[J]. IEICE Transactions on Information and Systems, 2021, E104-D(7): 1049-1053.

[29] 廖加文, 齐春, 曹剑中, 等. 结合自适应特征选择和蕨类分类器的相关滤波跟踪算法[J]. 西安交通大学学报, 2019, 53(6): 101-108.
LIAO J W, QI C, CAO J Z, et al. Tracking algorithm with correlation filtering based on combination of adaptive feature representation and fern classifier[J]. Journal of Xi'an Jiaotong University, 2019, 53(6): 101-108(in Chinese).

[30] 余仁伟, 朱浩, 蔡昌恺. 基于薄板样条函数的无人机多目标跟踪算法[J]. 仪器仪表学报, 2021, 42(3): 168-176.
YU R W, ZHU H, CAI C K. Multi-object tracking algorithm for UAV based on the thin plate spline function[J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 168-176(in Chinese).

[31] ZHANG Z H, ZHOU G J. Maneuvering target state estimation based on separate modeling with B-splines[J]. Aerospace Science and Technology, 2021, 119: 107172.

[32] HE F, WAN L L. B-spline based Extended target tracking technology in cluttered environments[J]. Journal of Physics: Conference Series, 2023, 2589(1): 012029.

[33] 汤新民, 郑鹏程. 基于大地坐标系的 IMM 航空器短期航迹外推[J]. 系统工程与电子技术, 2022, 44(7): 2293-2301.
TANG X M, ZHENG P C. IMM aircraft short-term track extrapolation based on geodetic coordinate system[J]. Systems Engineering and Electronics, 2022, 44(7): 2293-2301(in Chinese).

[34] Radio Technical Commission for Aeronautics . Minimum operational performance standards (MOPS) for 1090MHZ extended squitter automatic dependent surveillance-broadcast and traffic information services broadcast: RICA DO-260B[S]. Washington, D. C.: Radio Technical Commission for Aeronautics, 2009.

Target trajectory tracking and extrapolation based on INT-VSMM algorithm

REN Xuanming[1], TANG Xinmin[2,*], LIU Yusheng[1], LU Qixing[1]

(1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. School of Transportation Science and Engineering, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In order to address the issue of the traditional variable structure multiple model (VSMM) algorithm's lengthy computation time and difficulty meeting real-time constraints, we developed the intent variable structure multiple model (INT-VSMM) algorithm. This algorithm uses the target state and flight intent data collected by automatic dependent surveillance-broadcast (ADS-B) as the model's priori data, combining it with the VSMM theoretical framework. The motion pattern of the target in the flight phase of the flight path is decomposed, and a complete set of motion models is established. According to the principle of directed graph switching, a model set switching method is designed, which is mainly based on "hard" switching and supplemented by "soft" switching. The INT-VSMM algorithm is used to track the trajectory of the target aircraft, and short-term trajectory extrapolation is performed based on the target state estimation. According to the simulation results, the INT-VSMM algorithm outperforms the compared current approaches in terms of target tracking performance and computing time. Additionally, the extrapolated trajectory has a lower short-term inaccuracy, which can satisfy conflict detection requirements.

Keywords: variable structure multiple model; flight intention; digraph switching; maneuvering target tracking; trajectory extrapolation

Received: 2023-11-05; **Accepted**: 2023-12-11; **Published Online**: 2023-12-27 18:10

URL: link.cnki.net/urlid/11.2625.V.20231227.1212.002

Foundation items: National Key Research and Development Program of China (2021YFB1600500); National Natural Science Foundation of China (61773202,52072174); Open Fund of the Key Laboratory of General Aviation Operations, Civil Aviation Management Institute of China (CAMICKFJJ-2019-04); High-End Foreign Experts Recruitment Program (G2023202003L); Tianjin Science and Technology Support Program (24JCZDJC00090)

* **Corresponding author**. E-mail: tangxinmin@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0748

液体火箭注气式蓄压器流动阻尼理论模型

史童[1]，何允钦[1]，梁国柱[1,*]，潘辉[2]，朱平平[2]

(1. 北京航空航天大学 宇航学院，北京 102206；2. 北京宇航系统工程研究所，北京 100076)

摘 要：运载火箭纵向振动稳定性分析过程中需要进行增压输送系统的管道动力学分析。蓄压器流动阻尼计算方法直接影响输送系统动力学分析的准确性。目前，常用方法忽略了推进剂供应主管内流体流动的影响，采用孔板流动模型分析蓄压器流动阻尼，但存在较大偏差。基于此，引入连通孔流动模型，推导考虑推进剂供应主管内流体垂直流动影响的蓄压器非线性阻尼理论计算式；针对某蓄压器通过稳态数值仿真进行验证。结果表明：所提方法能够显著提高预测精度，对不同工况预测偏差在10%以内，而基于孔板流动模型的预测结果偏差超过30%。在此基础上进一步推导准稳态条件下的蓄压器线性阻尼理论计算式，并绘制线性阻尼曲线，发现蓄压器线性阻尼有最小值，其大小与连通孔总通流面积和主管内推进剂流速相关。

关 键 词：液体火箭；增压输送系统；注气式蓄压器；流动阻尼；连通孔

中图分类号：V434^{+}.3

文献标志码：A **文章编号**：1001-5965(2026)01-0214-09

液体火箭在飞行过程中，通常会遇到推进剂增压输送系统压力振荡与箭体纵向振动耦合的动力学不稳定问题，即POGO现象[1]。通常的解决方法是在推进剂增压输送系统中加入蓄压器，以吸收压力脉动能，改变系统的频率特性，其源自液压系统中的蓄能器，因其功能更专一(改变系统频率特性、抑制压力脉动)、环境更严苛(低温推进剂环境)、结构更特殊(安装空间受限)而命名，并在液体火箭领域沿用至今。1970年，Rubin根据Apollo计划期间POGO抑制团队的工作经验，总结了不同类型蓄压器的动态特性，系统阐述了蓄压器的基本工作原理，并指出采用蓄压器抑制POGO是最具可靠性和实用性的一种手段[2]。20世纪70年代以后，蓄压器被广泛安装在世界各国液体火箭的输送系统中，并针对重型氢氧液体火箭的POGO稳定性问题研发了具备大柔度特点的注气式蓄压器[3]，如日本的N-II和H-I运载火箭[4]、法国的Arian运载火箭[5]和美国的Ares运载火箭[6]。

在进行箭体的动力学稳定性分析时，有2种常见的分析方法：基于传递矩阵模型的阻抗分析法和基于动力学建模的状态方程法[7-9]。这2种方法都运用了液电模拟的思想，需要建立流体管道的集中或分布参数模型，是研究流体管路传输和瞬变问题的重要理论依据，也是液体火箭动力学稳定性分析方法的重要基础。

中国从21世纪初开始有关于注气式蓄压器研究成果的公开报道。万屹仑等[10]通过建立注气式蓄压器的集总参数动力学模型，分析了注气质量流量(以下简称“流量”)和温度及排气流量和开始时间对蓄压器工作特性参数的影响，但在建模过程中忽略了连通孔流动阻尼的非线性特征；马方超等[11]采用与文献[10]类似的蓄压器集总参数模型，通过系统仿真方法和地面试验验证了在注气和排气流量匹配的条件下，溢流管能够在满足POGO抑制需求的同时控制注气式蓄压器内液面高度，且指出了发动机起动或助推器分离时，蓄压器可能因连通孔

收稿日期：2023-11-20；**录用日期**：2024-01-26；**网络出版时间**：2024-04-08 14:13
网络出版地址：link.cnki.net/urlid/11.2625.V.20240407.1104.001
*通信作者. E-mail：lgz@buaa.edu.cn

引用格式：史童，何允钦，梁国柱，等. 液体火箭注气式蓄压器流动阻尼理论模型[J]. 北京航空航天大学学报，2026，52(1)：214-222.
SHI T，HE Y Q，LIANG G Z，et al. Theoretic model of flow resistance for gas-filled accumulators in liquid rockets[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：214-222 (in Chinese).

流动阻尼过小而导致气枕容积瞬间膨胀、液位降低至连通孔以下的问题, 故在设计阶段应充分认识到准确计算此种工况下蓄压器流动阻尼的重要性; 张青松等[12]在忽略蓄压器流动阻尼和流动惯性的条件下, 通过对不同注气式蓄压器总体方案建立液体火箭的 POGO 抑制系统动力学模型, 重点分析了注气式蓄压器的总体设计问题; 王楠等[13]基于某含注气式蓄压器的液体火箭 POGO 抑制系统, 通过对比试验结果及忽略蓄压器流动阻尼和惯性的动力学模型计算结果, 发现两者相对偏差为 8%。

然而, 由于流体和电子的本质并不相同, 针对流体传输管道非线性阻抗问题的分析比电气网络更加困难。试验证明, 管道频域特性分析和时域变化预测的准确性依赖于管道元件参数的准确性[14-15]。然而, 这些参数很难根据管道几何结构和工况进行预测, 必须通过试验测定[16]。对于注气式蓄压器中连通孔内的流动问题, 目前既没有充足的针对连通孔动态特性参数的试验测定, 也没有成熟、可靠的预测方法, 因此, 在注气式蓄压器连通孔设计阶段缺乏充分的设计依据; 且当前常用的蓄压器线性模型在稳态点附近失稳时, 主管内脉动压力只能为对称的正弦波, 不仅与实际遥测信号不符, 也无法反映实际流动中的非对称现象[17]。总而言之, 目前对于注气式蓄压器连通孔阻抗的规律性认识、试验测量及理论预测不足以有力支撑蓄压器设计和系统 POGO 稳定性分析。

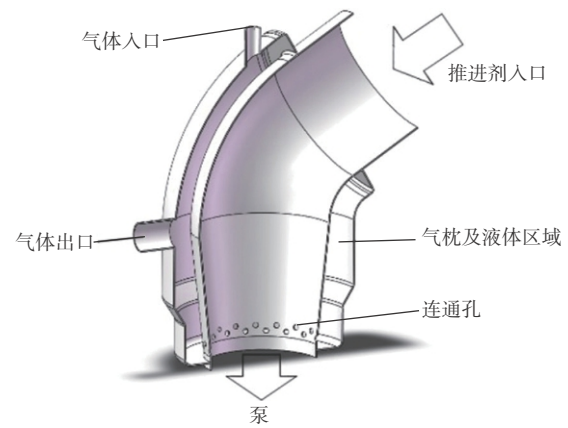
蓄压器的流动阻尼是其重要的动态特性参数之一。Detla 系列火箭曾在 TD-1 任务中因蓄压器入口处流动阻尼(简称入口阻尼)不足而发生高阶模态失稳, 也曾在 ITOS-E 任务中因入口阻尼估算结果远小于实际数值而导致实际飞行过程中因入口阻尼过大而被堵塞, 未能起到作用[18]。目前, 在系统 POGO 稳定性分析过程中, 通常采用忽略蓄压器流动阻尼或对简化后的流动阻尼模型进行线性化处理的方法[10-13,19-20], 其前提条件是忽略推进剂供应主管内流体流动对蓄压器连通孔内流动的影响, 将连通孔流动等效为同尺寸孔板流动, 根据连通孔内流量特性预测蓄压器的流动阻尼。然而, 文献 [21] 指出, 当主管与连通孔内流体流速比大于 1.29 或小于−2.54 时(正负表示流动方向), 连通孔内流量计算偏差已超过 15%。为更准确地预测连通孔的阻抗特性, 在波音公司的资助下, 1999 年, Doiron 假定流入流出工况对称, 在流量时均值为 0 且幅值较小的条件下, 通过能量守恒建立了一种蓄压器线性化流动阻尼(简称线性阻尼)计算方法, 2011 年, Ransom 等在此基础上研究了蓄压器亥姆霍兹频率下的连通孔线性阻尼测量[19]。但该方法理想化条件比较苛刻, 适用范围较模糊, 既未考虑主流纵向流动造成的连通孔横向流入和流出工况的非对称性, 也没有指出微小压力脉动的明确范围, 更无法预测蓄压器在工况大幅变化(如发动机关机前)的工作表现。

为了更直接地研究注气式蓄压器连通孔内的流动现象, 更准确地预测连通孔内的流体流量和蓄压器阻尼, 本文总结目前常用的蓄压器阻尼计算方法, 在此基础上, 基于文献 [21] 中提出的连通孔流动模型给出一种考虑推进剂供应主管流动影响的蓄压器非线性流动阻尼(简称非线性阻尼)计算方法, 并通过稳态计算流体力学(computational fluid dynamics, CFD)仿真方法进行验证; 推导准稳态条件下蓄压器线性阻尼, 绘制蓄压器线性阻尼曲线。

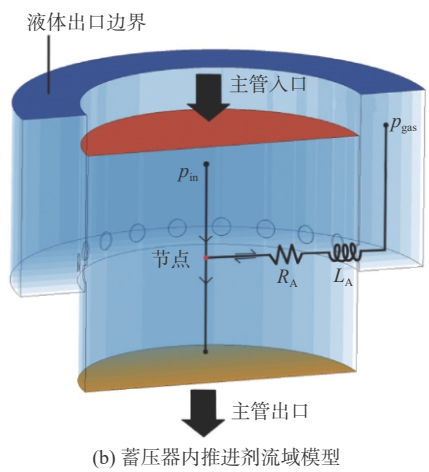
1　蓄压器阻尼的理论计算

1.1　基于孔板流动模型的蓄压器流动阻尼

图 1(a)为注气式蓄压器的结构剖视图。蓄压器的主要功能区域是包裹在推进剂供应主管外的环形空间和连接蓄压器与推进剂供应主管的连通孔。蓄压器内的环形空间内又可根据气液分界面



(a) 注气式蓄压器模型剖面图



(b) 蓄压器内推进剂流域模型

图 1　注气式蓄压器模型

Fig. 1　Model of gas-filled accumulator

分为气枕和液体区域。由于蓄压器试验、设计和加工的需要，通常会把蓄压器环形空间内的这一段推进剂供应主管也包括在蓄压器结构内。

本文的研究对象是蓄压器的流动阻尼特性，可以将蓄压器内液体推进剂部分简化为图1(b)的淡蓝色区域，底部黄色截面为主管出口截面，顶部红色截面为主管入口截面，顶部深蓝色截面为蓄压器内液体出口边界截面。

根据液电模拟法将液体区域简化为图1(b)中的等效电路，蓄压器与推进剂供应主管通过节点(连通孔)连接，推进剂从节点流入蓄压器环形空间内的流动阻尼表示为R_A，流动惯性表示为L_A，研究蓄压器流动阻尼R_A的变化规律。推进剂只能从连通孔流入或流出蓄压器。对蓄压器内液体部分建立如下数学模型[10]：

$$L_A\frac{\mathrm{d}q_l}{\mathrm{d}t}+R_Aq_l=p_{in}-p_{gas} \tag{1}$$

式中：q_l为连通孔内的流体流量，其值可正可负，正值表示流入蓄压器，负值表示流出蓄压器；p_{in}为蓄压器入口(即图1(b)中的主管入口，定义为连通孔中心向主管上游$D/2$处)处压力；p_{gas}为蓄压器内气枕压力。

由于R_A通过总压损失p_s关于q_l的线性化处理得到，且后两者通常又存在如下近似关系[9,22]：

$$p_s=R_sq_l^n \tag{2}$$

式中：R_s为流动摩擦阻尼；n为待定指数，$1<n<2$，通常在高雷诺数下趋近于2。

之前的研究均认为R_s为与流量无关的常数，并近似取n为2，由此推导得出连通孔的非线性阻尼R_{QA}为[9]

$$R_{QA}=\frac{\mathrm{d}p_s}{\mathrm{d}q_l}=2p_s/q_l=2R_sq_l \tag{3}$$

可见，如果能够确定连通孔的流动摩擦阻尼，就可以得到蓄压器流动阻尼。目前，常用方法是将连通孔流动问题简化为孔板流动问题进行分析，即孔前静压p_{in}等于主管压力p_O，忽略垂直方向流动，连通孔流动问题可简化为孔板流动问题，流经连通孔后静压下降至气枕压力p_{gas}。由此可得描述单个连通孔内流动的状态方程为

$$\dot{q}_P=C_DA_P\sqrt{2\rho|\Delta p|} \tag{4}$$

式中：$\dot{q}_P$为单个连通孔内流体流量；C_D为连通孔流量系数；A_P为单个连通孔面积；ρ为流体密度；Δp为推进剂供应主管内与蓄压器气枕的静压之差，且定义其正负表示压力方向(正表示压力方向指向蓄压器，负则背向蓄压器)。

考虑到蓄压器一般含有多个连通孔，假定此时的分流比(连通孔内的流体流量占推进剂供应主管总流体流量的比重)仍然较小，即主管内发生的压力振荡幅值较小，并忽略连通孔之间耦合影响，蓄压器的流动阻尼可以简化为多个单连通孔问题的并联，因此有

$$q_l=nC_DA_P\sqrt{2\rho\Delta p}=C_DA_A\sqrt{2\rho|\Delta p|} \tag{5}$$

式中：A_A为连通孔总面积。

由式(3)和式(5)可得

$$R_{QA}=\frac{|q_l|}{\rho C_D^2A_A^2} \tag{6}$$

1.2 基于连通孔流动模型的蓄压器流动阻尼

在式(6)基础上，如果某一工况下连通孔内推进剂流量近似不变，则可以近似取为常数而成为线性阻尼(同线性电阻元件)。但实际上，蓄压器的流动摩擦阻尼R_s并不是与流量无关的常数，且与流量有很大的关系，连通孔的流量系数与流速比M有如下近似关系[21]：

$$\frac{C_{D,0}}{C_D}=k|M+a|+b \tag{7}$$

式中：$C_{D,0}$为主管流速为0时的连通孔流量系数，近似视为常数(只与孔的几何形状有关)；M为主管内流体的平均流速v与连通孔内流体的平均流速v_P之比，其值为正表示流入蓄压器，为负则表示流出；k、a、b为试验测定的系数。

将式(7)代入式(5)，可求出压差和流量的关系：

$$\Delta p=\left(k\left|\frac{v}{v_P}+a\right|+b\right)^2\frac{q_l^2}{2\rho C_{D,0}^2A_A^2} \tag{8}$$

1) 当$v_P>0$或$v_P<-v/a$时，有

$$\Delta p=\frac{(ak+b)^2q_l^2}{2\rho C_{D,0}^2A_A^2}+\frac{(ak+b)kvq_l}{C_{D,0}^2A_A}+\frac{\rho k^2v^2}{2C_{D,0}^2} \tag{9}$$

$$R_{QA}=\frac{(ak+b)^2q_l}{\rho C_{D,0}^2A_A^2}+\frac{(ak+b)kv}{C_{D,0}^2A_A} \tag{10}$$

2) 当$-v/a<v_P<0$时，有

$$\Delta p=-\frac{(ak-b)^2q_l^2}{2\rho C_{D,0}^2A_A^2}-\frac{(ak-b)kvq_l}{C_{D,0}^2A_A}-\frac{\rho k^2v^2}{2C_{D,0}^2} \tag{11}$$

$$R_{QA}=-\frac{(ak-b)^2q_l}{\rho C_{D,0}^2A_A^2}-\frac{(ak-b)kv}{C_{D,0}^2A_A} \tag{12}$$

为了进一步简化该结果，可以将工况限制在内，采用一个统一的公式描述压差和流量的关系，该式应在正负2个方向上尽可能趋近原式。注意到式(9)和式(11)给出的流量从正负2个方向上趋近于0时，连通孔两侧的压差结果数值相等，符号相反。这一结果可能有合理性，但不便于实际计

算。因此, 合理假设 q_{l}=0 时, 连通孔两侧压差也为 0, 使结果可积。为在正负 2 个方向上趋近于式(9)和式(11), 引入函数:

$$f(x)=\lim_{s\to 0^{+}}\arctan\left(\frac{x}{s}\right)=\begin{cases}1 & x>0\\ 0 & x=0\\ -1 & x<0\end{cases} \tag{13}$$

由此可以得到压差和流量的关系为

$$\Delta p=\frac{akbq_{\mathrm{l}}^{2}}{\rho C_{\mathrm{D},0}^{2}A_{\mathrm{A}}^{2}}+\frac{bkvq_{\mathrm{l}}}{C_{\mathrm{D},0}^{2}A_{\mathrm{A}}}+\frac{2}{\pi}\left(\frac{a^{2}k^{2}+b^{2}}{2\rho C_{\mathrm{D},0}^{2}A_{\mathrm{A}}^{2}}q_{\mathrm{l}}^{2}+\frac{ak^{2}v}{C_{\mathrm{D},0}^{2}A_{\mathrm{A}}}q_{\mathrm{l}}+\frac{\rho k^{2}v^{2}}{2C_{\mathrm{D},0}^{2}}\right)\arctan\left(\frac{q_{\mathrm{l}}}{s}\right) \tag{14}$$

式中: s 取足够小的正数。

对式(14)求导, 可以得到蓄压器的非线性阻尼为

$$R_{\mathrm{QA}}=\frac{2akbq_{\mathrm{l}}}{\rho C_{\mathrm{D},0}^{2}A_{\mathrm{A}}^{2}}+\frac{bkv}{C_{\mathrm{D},0}^{2}A_{\mathrm{A}}}+\frac{2}{\pi}\left(\frac{a^{2}k^{2}+b^{2}}{\rho C_{\mathrm{D},0}^{2}A_{\mathrm{A}}^{2}}q_{\mathrm{l}}+\frac{ak^{2}v}{C_{\mathrm{D},0}^{2}A_{\mathrm{A}}}\right)\arctan\left(\frac{q_{\mathrm{l}}}{s}\right) \tag{15}$$

限制式(15)适用范围的因素主要有如下 2 个: ①式(7)在流速比不大时, 即主管内流体平均流速较小或连通孔内流体平均流速极大时, 有偏差。式(7)代入主管流速 v 为 0 时, 可得 $ak+b$ 的值为 1, 式(7)的图像应与纵轴交于 $C_{\mathrm{D},0}/C_{\mathrm{D}}=1$ 处, 但拟合给出的计算结果往往偏低(文献 [21] 给出的结果为 0.94), 属于模型误差。②式(7)在流出方向上需要满足 $-v/a<v_{\mathrm{P}}$。当 v_{P} 在负向上的绝对值超过该范围时, 式(7)不成立, 但计算结果是否仍然可靠取决于工况。为了进一步验证式(15)的准确性和合理性, 在第 2 节将采用该方法根据单孔 CFD 仿真结果来计算某蓄压器的流动特性, 并将理论预测结果与蓄压器 CFD 仿真结果对比。

2　CFD 仿真方法

2.1　几何模型

根据图 1(b) 建立蓄压器内液体流域的几何模型, 并对模型进行面对称处理以减小仿真计算量, 如图 2 所示。主管外径为 470 mm, 壁厚 2.5 mm, 为控制计算量, 沿连通孔上下游各截取 200 mm 作为流域边界; 连通孔为 20 个薄壁圆孔, 均匀排布在主管壁面一周, 孔径为 30 mm, 距离蓄压器底部高度为 30 mm; 蓄压器内环形液体区域高度为 300 mm, 外径为 695 mm。

2.2　网格划分

使用非结构化网格自动划分方法对蓄压器内部推进剂流域模型进行非结构化网格划分。蓄压器内附面层流域采用五面体棱柱网格, 其他流域采用四面体网格。

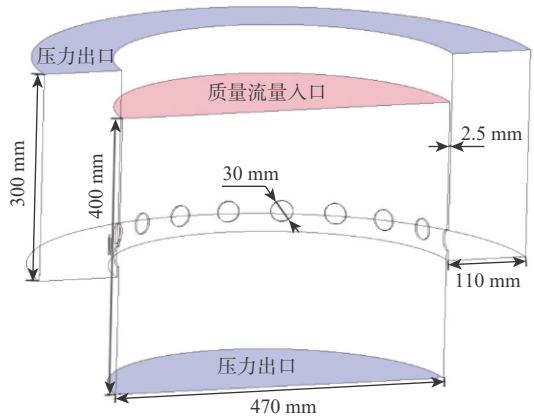
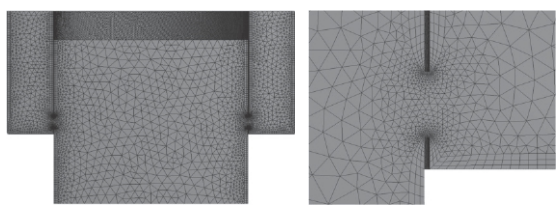


图 2　几何模型示意图

Fig. 2　Geometry model diagram

连通孔附近第 1 层附面层网格的厚度设置为 0.4 mm, 其他流域第 1 层附面层网格厚度为 2 mm, 共计 4 层网格。由于仿真采用的拓展壁面函数方法对壁面函数 y^{+} 值的要求不高, 且根据仿真结果, 各工况下 y^{+} 大于 18, 因此, 认为边界层划分合理。

根据上述条件最终得到的整体及局部网格划分结果如图 3 所示。网格总数为 2 727 351, 节点数为 753 924。最低网格正交质量为 0.405, 平均正交质量为 0.812; 最大网格扭曲率为 0.659, 平均扭曲率为 0.254; 平均雅可比为 1.037。



(a) 网格划分结果　　(b) 局部网格示意图 (连通孔处)

图 3　网格展示

Fig. 3　Mesh display

为验证网格数量无关性, 对所采用的网格进行加密处理, 网格总数由约 2 727 351 增加至 6 341 903, 边界层第 1 层网格厚度和整体网格尺寸缩小约 30%。不改变计算方法和收敛判据, 在相同的模型位置设置监测点, 计算得到的关键截面的压力和流量偏差最大不超过 0.95%, 因此, 认为仿真结果能够满足网格数量无关性要求。

2.3　数学模型

为使模型更具一般性, 并为相关工程试验研究[12]提供必要的理论支持, 根据已有的研究成果[21], 流体工质选用更便于开展地面试验的水, 视为不可压流体, 密度为 998.2 kg/m^{3}, 黏度为 0.001 003 kg/(m·s)。

由于工质为不可压流体, 且不考虑换热及相变带来的影响, 只需对连通孔流域模型建立稳态条件下的连续性方程和动量方程, 求解不含能量方程的

Navier-Stokes 方程组。

Sheikholesiami 等对孔板流量计的仿真研究证明，采用两方程 k-ε 湍流模型和代数应力湍流模型得到的孔板流出系数与经验数据相差在 2% 以内[23]。其他相关研究也表明了在孔板流动数值仿真研究中采用 k-ε 湍流模型和标准壁面函数的可行性和可靠性[24-26]。因此，本文采用更适合低雷诺数流动的RNG k-ε湍流模型，并在壁面处采用拓展壁面函数方法处理，使在绝大部分壁面等效于采用标准壁面函数，且避免了 y^+过小导致的标准壁面函数失真。

2.4 边界条件及数值计算方法

计算域有 1 个入口边界和 2 个出口边界，主管入口边界是质量流量入口，控制主管内流体流速；主管出口边界和蓄压器内液体出口边界是压力出口（回流取当地静压条件）。对于蓄压器的流入工况，蓄压器内流体出口边界的静压条件为 0.1 MPa，主管出口压力在此基础上依次增大 0.01, 0.02, 0.03, 0.04, 0.05 MPa；对于蓄压器的流出工况，主管压力出口边界压力为 0.1 MPa，蓄压器内流体出口边界的回流静压在此基础上依次增大 0.01, 0.02, 0.03, 0.04, 0.05 MPa。主管质量流量入口的推进剂流量为 8.65, 865, 173, 346, 519, 692, 865 kg/s，对应推进剂管路内推进剂流速约为 0.1, 1, 2, 4, 6, 8, 10 m/s。

由于本文中流体被视为不可压，采用压力基求解器求解，并使用 PISO 算法和 PRESTO 压力插值方式，求得稳态过程的解。对湍动能和湍流耗散率采用一阶迎风格式，其他量采用二阶迎风格式，并最终以残差收敛至 10^{-4} 作为计算收敛的判定依据。

2.5 仿真方法验证

对于给定的连通孔形状和面积，由主管入口压力与蓄压器内出口压力之差能够计算出无损流动条件下的理想流量。包含多个连通孔蓄压器的流量系数即为实际流量与该理想流量之比。式(8)的后续推导过程将$C_{\mathrm{D},0}$视为常数，但根据小孔流动的一般结论[27]，流量系数不仅与孔板形状有关，与孔内流动的雷诺数也有关，相较于主流垂直流动带来的影响，这一变化是否可以忽略不计需要通过 CFD 仿真验证。根据已发表的同尺寸单个连通孔的仿真研究[21]验证本文关于蓄压器仿真结果的合理性。

图 4 为主管流速为 0.1 m/s 时的低主管流速比工况，不同压差下仿真得到的蓄压器流量系数与文献 [21] 的单个连通孔流量系数对比。其中，蓄压器流量系数根据主管入口和蓄压器内出口静压差和连通孔内总流量计算得到。

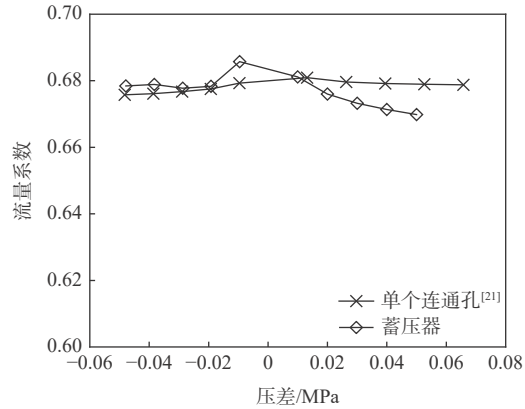


图 4 单个连通孔和蓄压器流量系数对比

Fig. 4 Flow coefficient comparison of single communication-port and accumulator

文献 [21] 给出的单个连通孔流量系数在正负 0～0.05 MPa 压差下最大与最小相差 0.3%，本文仿真结果为 1.7%，相同工况下两者相对偏差小于 1.5%。一方面，佐证了本文仿真方法的合理性及仿真结果的可靠性；另一方面，也佐证了相较于主管流动带来的影响，可以近似在计算过程中将主管流速为 0 时的流量系数$C_{\mathrm{D},0}$作为常数处理，为前文推导过程中的假设提供了仿真数据依据。

3 蓄压器非线性阻尼仿真结果与理论预测的对比分析

式(15)将考虑主管内流体垂直流动影响的连通孔流量系数C_{D}与忽略主管内流体垂直流动影响的流量系数$C_{\mathrm{D},0}$联系了起来，可以根据$C_{\mathrm{D},0}$估计流量的大小。根据图 4，合理取$C_{\mathrm{D},0}=0.68$，分别代入目前普遍使用的式(6)和式(15)（根据文献 [21] 取系数 a=0.643 6, b=0.824 2, k=0.184 5, s=0.01）。假设连通孔内流体流量为 0 时连通孔两侧压差也为 0，对比不同工况下 2 种方法计算得到的蓄压器入口和气枕静压差绝对值与连通孔流量的关系，并与仿真结果对比，如图 5 所示。

当推进剂供应管路内流体流速较低时，采用孔板流动模型能够对蓄压器内流体流动进行较为准确的预测。但随着推进剂供应管路内流体流速的增加，孔板流动模型计算结果与连通孔流动出现越来越大的偏差。一方面，孔板流动模型中，非线性阻尼在流入和流出方向上是对称的，但蓄压器流动阻尼在流入和流出方向上有显著的非对称性；另一方面，在非线性关系更复杂的流出方向上，孔板流动模型首先表现出了较大的偏差，这是由于随着推进剂供应主管内流体流速的增大，流速比也在变大，主导流动状态的流体特性由水平方向上的惯性力和黏性力变为垂直方向的惯性力和黏性力，在推

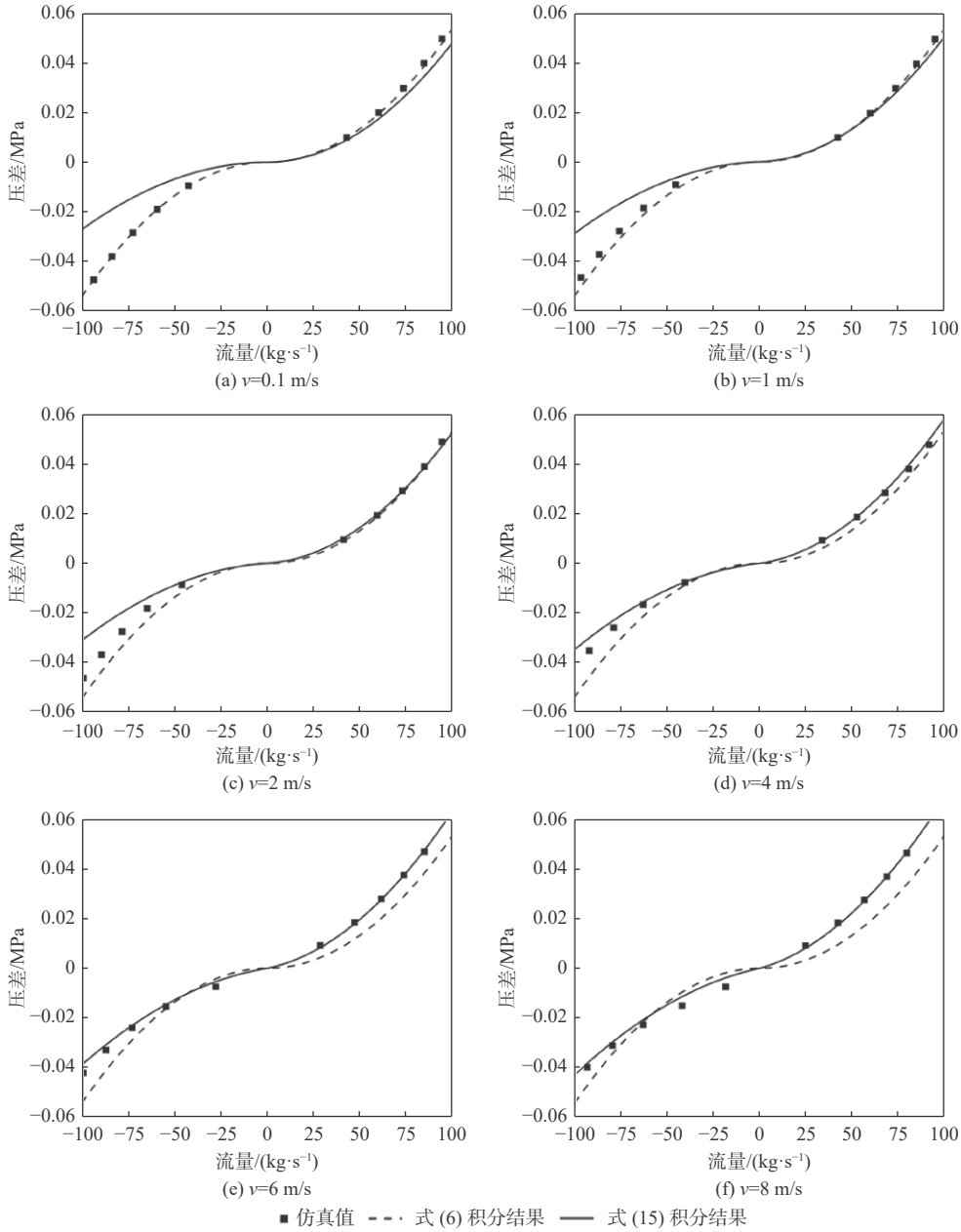


图 5　不同工况下的结果对比

Fig. 5　Comparison of results under different conditions

进剂供应主管内流速达到 2 m/s 时, 通过式(6)来计算该蓄压器的压差-流量特性时已表现出一定的偏差, 向式(15)的积分结果靠近, 而当推进剂供应主管内流速超过 4 m/s 时, 无论是在流入还是流出方向上, 式(6)的偏差都已不可忽略, 式(15)给出的积分结果显然更贴合于仿真结果。

对于低温推进剂而言, 通常最佳流速应控制在 3～9 m/s[27]。液体火箭发动机管路系统设计时, 考虑到尽可能减少管路中的流速损失及管路质量, 通常会将推进剂流速控制在 5～8 m/s[28]。例如, 同样采用注气式蓄压器的日本 H-II 火箭一级发动机液氧供应管路内液氧设计流速为 6.4 m/s[29]。根据仿真结果, 在主管流速为 6～8 m/s 的工况下, 式(6)的最大计算偏差超过 30%, 但式(15)的计算偏差始终在 10% 以内, 能够更准确地描述液体火箭发动机氧路注气式蓄压器内外压差和流入流出流量的关系。

4　蓄压器线性阻尼曲线

通常情况下, 在运载火箭 POGO 稳定性计算中, 需要对蓄压器的流动阻尼做线性化处理, 线性化后的蓄压器流动阻尼记为R_{LA}。假定连通孔内流量振荡的时均值为 0, 流量满足:

$$q_{\mathrm{l}}(t) = W\sin(\omega_0 t) \tag{16}$$

式中: W 为连通孔内流量振荡幅值; ω_0为连通孔内流量振荡。

根据式(15), 代入$q_{\mathrm{l}} = 0$, 给出第 1 种线性阻尼

结果：

$$R_{\mathrm{LA}}(q_1=0)=\frac{bkv}{C_{\mathrm{D},0}^2A_{\mathrm{A}}} \tag{17}$$

第1种线性阻尼是蓄压器流动阻尼拟合公式给出的流量接近0时流入和流出阻尼的均值。在孔板流动模型中，因为孔板不存在流入和流出的不对称性，均值为0；而连通孔仅在一侧垂直于支管的主流流动，因此，相同流量下的流入阻尼大于流出阻尼，存在非对称性。由于流速比较大时，式(7)存在一定偏差，流量系数计算结果偏小，R_{LA}的实际值应大于计算结果，可写为

$$R_{\mathrm{LA},1}=\frac{bkv}{\sigma C_{\mathrm{D},0}^2A_{\mathrm{A}}} \tag{18}$$

式中：σ为修正系数($0<\sigma<1$)。

流量函数q_1经傅里叶变换后记为Q_1，压力函数Δp经傅里叶变换后记为ΔP。如忽略主流垂直流动对连通孔内流动的影响，将描述连通孔两侧压差和流量的关系式(5)代入式(16)，可得蓄压器的第2种线性阻尼为

$$R_{\mathrm{LA},2}=\frac{\Delta P(\omega_0)}{Q_1(\omega_0)}=\frac{4}{3\pi}\cdot\frac{W}{\rho C_{\mathrm{D},0}^2A_{\mathrm{A}}^2} \tag{19}$$

由式(2)、式(4)和式(19)可知，蓄压器的第2种线性阻尼与流动摩擦阻尼R_{s}存在如下关系：

$$R_{\mathrm{LA},2}=\frac{8W}{3\pi}R_{\mathrm{s}} \tag{20}$$

这一结果与文献[19]中通过能量守恒建立的线性阻尼计算结果一致。

第3种线性阻尼计算方式考虑了主流垂直流动对连通孔内流动的影响。式(14)给出了稳态条件下连通孔两侧压差和流量的关系。如果振荡频率足够低，该式也可以用于描述准稳态条件的连通孔工作状态。将式(16)代入式(14)并使s趋向于0，可得蓄压器的第3种线性阻尼为

$$R_{\mathrm{LA},3}=\frac{\Delta P(\omega_0)}{Q_1(\omega_0)}=\frac{bkv}{C_{\mathrm{D},0}^2A_{\mathrm{A}}}+\frac{\left(a^2k^2+b^2\right)W}{\rho C_{\mathrm{D},0}^2A_{\mathrm{A}}^2}\cdot\frac{4}{3\pi}+\frac{\rho k^2v^2}{WC_{\mathrm{D},0}^2}\cdot\frac{2}{\pi} \tag{21}$$

以上3种蓄压器线性阻尼是不同入口流量振荡幅值下的近似结果。第3节中已验证了引入连通孔流动模型进行蓄压器非线性阻尼计算的合理性，因此，由蓄压器非线性阻尼推导得出的准稳态条件下的$R_{\mathrm{LA},1}$、$R_{\mathrm{LA},2}$和$R_{\mathrm{LA},3}$应为不同振荡流动幅值下的线性阻尼。假定蓄压器线性阻尼关于入口流量振荡幅值W连续且可导，采用线性函数连接$R_{\mathrm{LA},1}$和$R_{\mathrm{LA},3}$可以得到线性阻尼关于连通孔内流量振荡幅值W的曲线如图6所示。当蓄压器入口振荡流量进一步增大至可忽略主管内垂直流动影响时，蓄压器线性阻尼符合图6中虚线所示。

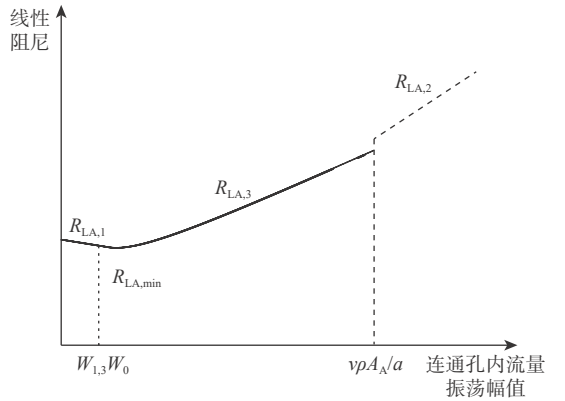


图6 蓄压器线性阻尼曲线

Fig. 6 Linear resistance curve of accumulator

图中，连接点$W_{1,3}$的值为

$$W_{1,3}=\frac{4}{\pi}\cdot\frac{\sigma}{1-\sigma}\cdot\frac{\rho kvA_{\mathrm{A}}}{b} \tag{22}$$

在W_0时，蓄压器具有最小线性阻尼为

$$R_{\mathrm{LA,min}}=\left(b+\frac{4\sqrt{6a^2k^2+6b^2}}{3\pi}\right)\frac{kv}{C_{\mathrm{D},0}^2A_{\mathrm{A}}} \tag{23}$$

根据式(21)可知

$$W_0=\frac{\sqrt{6}}{2}\cdot\frac{\rho kvA_{\mathrm{A}}}{\sqrt{a^2k^2+b^2}} \tag{24}$$

当连通孔内流量振荡幅值低于W_0时，蓄压器线性阻尼随流量振荡幅值增大而从$R_{\mathrm{LA},1}$减至$R_{\mathrm{LA,min}}$；当连通孔内流量振荡幅值大于W_0且小于$v\rho A_{\mathrm{A}}/a$时，线性阻尼随流量振荡幅值增大而增大，可通过式(21)进行估算；当连通孔内流量振荡幅值大于$v\rho A_{\mathrm{A}}/a$时，应采用基于孔板流动模型的式(20)进行线性阻尼估算。

5 结 论

本文推导了考虑推进剂输送系统垂直流速对蓄压器内水平流动影响的注气式蓄压器非线性阻尼的理论计算方法，并通过CFD仿真验证了该方法的可靠性，对基于连通孔流动模型的注气式蓄压器非线性阻尼进行了线性化推导。得到如下结论：

1) 单排连通孔设计中，注气式蓄压器的流动阻尼主要取决于单个连通孔的流动特性。本文单个连通孔的流量系数变化率与注气式蓄压器的流量系数变化率相对偏差小于1.5%，没有观察到明显的耦合流动现象。

2) 导出了考虑推进剂供应主管内流体流动对连通孔内流体流动影响的注气式蓄压器非线性阻尼计算公式，该式反映了注气式蓄压器流动阻尼与

连通孔通流面积及连通孔内推进剂流量的关系。对于本文使用的注气式蓄压器模型和工况, 通过该式得到的流动阻尼与仿真结果偏差在 10% 以内, 而孔板流动模型的偏差超过 30%。

3) 导出了准稳态条件下考虑推进剂供应主管内流体流动对蓄压器连通孔内流体流动影响的蓄压器线性阻尼计算公式和最小线性阻尼计算公式。表明, 在准稳态条件下, 注气式蓄压器的线性阻尼取决于连通孔总通流面积、连通孔内流量振荡幅值及主管内推进剂流速; 注气式蓄压器的最小线性阻尼仅与连通孔总通流面积及主管内推进剂流速有关。

参考文献（References）

[1] LARSEN C E. NASA experience with POGO in human space-flight vehicles: RTO-MP-AVT-152[R]. Washington, D. C. : NASA, 2008: 1-23.

[2] RUBIN S. Prevention of coupled structure-propulsion instability (POGO): SP-8055[R]. Washington, D. C. : NASA, 1970: 1-52.

[3] 王小军, 于子文, 张兵, 等. 国内外运载火箭 POGO 抑制技术研究进展[J]. 中国科学: 技术科学, 2014, 44(5): 492-503.
WANG X J, YU Z W, ZHANG B, et al. Progress of POGO suppression technology of launch vehicles at home and abroad[J]. Scientia Sinica: Technologica, 2014, 44(5): 492-503(in Chinese).

[4] MORI H. POGO analysis based on N-II/H-I vehicle flight data[J]. Journal of the Japan Society for Aeronautical and Space Sciences, 1991, 39(445): 71-80.

[5] HRISAFOVIC M, ESTOUEIG M, 王晋桦. 阿里安运载火箭 POGO 振动的研究[J]. 国外导弹技术, 1980(12): 45-53.
HRISAFOVIC M, ESTOUEIG M, WANG J H. A study of the POGO effect on the Ariane launch vehicle[J]. Missiles and Space Vehicles, 1980(12): 45-53(in Chinese).

[6] SWANSON L, GIEL T. Design analysis of the Ares I POGO accumulator[C]//Proceedings of the 45th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2009: 4950.

[7] RUBIN S. Longitudinal instability of liquid rockets due to propulsion feedback(POGO)[J]. Journal of Spacecraft and Rockets, 1966, 3(8): 1188-1195.

[8] LOCK M H, RUBIN S. Passive suppression of POGO on the space shuttle: CR-132452[R]. Washington, D. C. : NASA, 1974: 1-111.

[9] OPPENHEIM B W, RUBIN S. Advanced POGO stability analysis for liquid rockets[J]. Journal of Spacecraft and Rockets, 1993, 30(3): 360-373.

[10] 万屹仑, 付欣毓, 张黎辉, 等. 注气式蓄压器系统工作特性仿真分析[J]. 火箭推进, 2018, 44(3): 37-42.
WAN Y L, FU X Y, ZHANG L H, et al. Simulation analysis of working characteristics of air-injection accumulator system[J]. Journal of Rocket Propulsion, 2018, 44(3): 37-42(in Chinese).

[11] 马方超, 刘文川, 陈牧野, 等. 注气式蓄压器自由液面控制技术研究[J]. 导弹与航天运载技术, 2020(4): 63-67.
MA F C, LIU W C, CHEN M Y, et al. Study on free liquid level control technology of air-injection accumulator[J]. Missiles and Space Vehicles, 2020(4): 63-67(in Chinese).

[12] 张青松, 范瑞祥, 张兵, 等. 低温液体火箭注气式蓄压器总体方案研究[J]. 导弹与航天运载技术, 2021(3): 61-65.
ZHANG Q S, FAN R X, ZHANG B, et al. Study of gas filled accumulator system scheme for cryogenic rocket[J]. Missiles and Space Vehicles, 2021(3): 61-65(in Chinese).

[13] 王楠, 容易, 胡久辉, 等. 注气式 POGO 抑制系统模型研究[J]. 导弹与航天运载技术, 2020(5): 32-37.
WANG N, RONG Y, HU J H, et al. Study on model of gas injection POGO suppression system[J]. Missiles and Space Vehicles, 2020(5): 32-37(in Chinese).

[14] 苏尔皇. 管道动态分析及液流数值计算方法[M]. 哈尔滨: 哈尔滨工业大学出版社, 1985.
SU E H. Dynamic analysis of pipeline and numerical calculation method of liquid flow[M]. Harbin: Harbin Institute of Technology Press, 1985(in Chinese).

[15] 蔡亦纲. 流体传输管道动力学[M]. 杭州: 浙江大学出版社, 1990.
CAI Y G. Dynamics of fluid transmission pipeline[M]. Hangzhou: Zhejiang University Press, 1990(in Chinese).

[16] JOHNSTON D N. Prediction of hydraulic inductance[C]//Proceedings of the ASME 2002 International Mechanical Engineering Congress and Exposition. New York: ASME, 2008: 85-92.

[17] 任辉, 任革学, 荣克林, 等. 液体火箭 POGO 振动蓄压器非线性仿真研究[J]. 强度与环境, 2006, 33(3): 1-6.
REN H, REN G X, RONG K L, et al. A nonlinear model on the simulation of POGO vibrations in liquid rockets[J]. Structure & Environment Engineering, 2006, 33(3): 1-6(in Chinese).

[18] PAYNE J G, RUBIN S. POGO suppression on the Delta vehicle: TR-74-180[R]. Springfield: National Technical Information Service, 1974: 5-53.

[19] RANSOM D, DOIRON H. Experimentally validated POGO accumulator flow resistance model[C]//Proceedings of the 47th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit. Reston: AIAA, 2011: 5774.

[20] COPPOLINO R N, LOCK M H, RUBIN S. Space shuttle POGO studies: ATR-78(7475)-1[R]. Washington, D. C. : NASA, 1977.

[21] 史童, 梁国柱, 潘辉, 等. 运载火箭注气式蓄压器联通孔流量特性数值仿真研究[J]. 推进技术, 2023, 44(4): 21-31.
SHI T, LIANG G Z, PAN H, et al. Numerical simulation study on flow characteristics of communication hole of air-filled accumulator of launch vehicle[J]. Journal of Propulsion Technology, 2023, 44(4): 21-31(in Chinese).

[22] LAU K K, EDGE K A, JOHNSTON D N. Impedance characteristics of hydraulic orifices[J]. Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 1995, 209(4): 241-253.

[23] SHEIKHOLESIAMI M Z, PATEL B R. Numerical modeling of flow through orifice meters: CREARE-TM-1160[R]. Hanover: Creare Inc. , 1988.

[24] 刘万龙, 梁国柱, 徐铭, 等. 小孔径多级孔板组件节流效应仿真[J]. 北京航空航天大学学报, 2011, 37(2): 241-244.
LIU W L, LIANG G Z, XU M, et al. Simulation of throttling effect of small aperture multi-stage orifice plate assembly[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(2): 241-244(in Chinese).

[25] 陈家庆, 王波, 吴波, 等. 标准孔板流量计内部流场的 CFD 数值模拟[J]. 实验流体力学, 2008, 22(2): 51-55.

CHEN J Q, WANG B, WU B, et al. CFD numerical simulation of internal flow field of standard orifice flowmeter[J]. Journal of Experiments in Fluid Mechanics, 2008, 22(2): 51-55(in Chinese).

[26] 赵奇, 牛志娟, 杨雪峰, 等. 基于 CFD 的非标准孔板流量计的数值模拟[J]. 节能技术, 2015, 33(5): 453-456.
ZHAO Q, NIU Z J, YANG X F, et al. Numerical simulation of non-standard orifice flowmeter based on CFD[J]. Energy Conservation Technology, 2015, 33(5): 453-456(in Chinese).

[27] BARRON R F. Cryogenic systems[M]. New York: McGraw-Hill, 1966.

[28] 廖少英. 液体火箭推进增压输送系统[M]. 北京: 国防工业出版社, 2007.
LIAO S Y. Liquid rocket propellant and pressurization feed systems[M]. Beijing: National Defense Industry Press, 2007(in Chinese).

[29] NAGAI H, NODA K, YAMAZAKI I, et al. Development status of H-II Rocket first stage propulsion system[C]//Proceedings of the 24th Joint Propulsion Conference. Reston: AIAA, 1988: 2933.

Theoretic model of flow resistance for gas-filled accumulators in liquid rockets

SHI Tong[1], HE Yunqin[1], LIANG Guozhu[1,*], PAN Hui[2], ZHU Pingping[2]

(1. School of Astronautics, Beihang University, Beijing 102206, China;
2. Beijing Institute of Astronautical Systems Engineering, Beijing 100076, China)

Abstract: Hydraulic dynamics analysis of the pressurized supply system is needed in the process of longitudinal structure stability analysis of the launch vehicle. The method of calculating the accumulator's resistance directly affects the accuracy of the dynamic analysis of the supply system. At present, the commonly used method is to ignore the influence of the fluid flow inside the propellant supply system and to analyze the accumulator resistance with an orifice flow model. However, this method has large errors. In this paper, by introducing the communication port flow model, the non-linear resistance calculation formula of the accumulator is deduced, where the influence of the vertical flow in the propellant supply system is considered. Then, this theoretical calculation method is verified by conducting a steady-state simulation on an accumulator. Under the various situations, it is discovered that the nonlinear resistance calculation approach suggested in this study may greatly increase prediction accuracy within a variation of less than 10%, whereas predictions based on the orifice flow model have deviations above 30%. The linear resistance theoretical formula under quasi-steady condition is further deduced and the linear resistance curve is drawn. The linear resistance of the accumulator was found to have a minimum value, which is related to the total flow area of the communication port and the velocity of propellant inside the main pipe.

Keywords: liquid rocket; pressurized supply system; gas-filled accumulator; flow resistance; communication port

Received: 2023-11-20; **Accepted**: 2024-01-26; **Published Online**: 2024-04-08 14:13
URL: link.cnki.net/urlid/11.2625.V.20240407.1104.001

* **Corresponding author**. E-mail: lgz@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0749

基于不匹配结点对接子结构法的超声速颤振预测

刘晨宇，谢长川*
(北京航空航天大学 航空科学与工程学院，北京 100191)

摘 要：准确预测超声速飞行器颤振速度边界是超声速飞行器气动弹性研究的重点和难点之一。针对超声速飞行器模型自由度数高且颤振模态耦合形式复杂的问题，提出一种基于不匹配结点对接子结构法的结构降阶建模方法，结合非定常活塞理论和流-固耦合形态插值理论实现气动弹性颤振建模，采用传统模态法、传统子结构法和不匹配结点对接子结构法对某后掠翼模型和壁板模型进行超声速颤振特性分析。结果表明：与商业软件相比，采用本文子结构法建立的降阶模型的主要低阶模态频率相对误差不大于1.1%，颤振速度相对误差不大于1.3%。与传统模态法相比，本文子结构法需要的模态阶数更少。因此，所提方法可高效准确地预测典型超声速飞行器的颤振特性。

关 键 词：子结构；颤振；气动弹性；活塞理论；插值
中图分类号：V215.3+4
文献标志码：A **文章编号**：1001-5965(2026)01-0223-09

模态综合法[1-6]又称为动态子结构法，其将整体结构划分为若干子结构，先对子结构进行降阶，再用对接条件组装起来，理论基础是Raleigh-Ritz法。

20世纪50年代，模态综合技术起源于对飞机及机翼的固有振动分析的需求。Muscolino[7]提出假设模态法，主要思想是假设一组低阶模态作为坐标变换矩阵引入动力学方程，实现方程的降阶，再积分求解。Ritz法也是假设模态法，其优点是避免求解子结构主模态，缺点是假设模态的选取依靠经验。Muscolino[7]同时提出固定界面模态综合法，主要思想是用子结构的刚体模态和对接界面确定的约束模态描述牵连运动，用固定界面主模态描述相对运动。Craig[1]对该方法的适用性进行改进，使之成为工程中经常使用的固定界面模态综合法，并使用剩余柔度作为Ritz基矢量，界面力作为广义坐标，并引入位移协调条件消除广义坐标，显著提高综合效率。向树红等[3]基于变分原理推导，提出双协调自由界面模态综合法，并对高阶主模态截断进行动力学解释。固定界面模态综合法的缺点是：①必须保留所有对接自由度，限制了模型的最低阶数；②固定界面的假设难以与试验结合。而自由界面模态综合法刚好克服了以上缺点。

不同于低速情况，超声速飞行器在结构设计、材料使用和工作环境等方面具有特殊性，模型自由度高，计算资源需求大。子结构法能够对模型进行降阶，降低计算规模。在大型项目建模时，通常将复杂模型划分为多个部件分别进行建模。如果某个部件的模型需要修改，使用子结构法可以将该部分计算量限制在部件自由度内，而不需要再对全机进行重新计算。但是传统子结构法只适用于对接结点匹配的情况，即2个对接界面上的有限元结点都是一一对应的。若出现不同部件在连接处的有限元结点不匹配的情况，传统子结构法不再适用。如在对某个子结构进行重新设计时，每次都要考虑结点匹配问题，显著增加工作量；对于事先设计且划分好网格的标准件，设计人员希望能够直接取用这些标准件组装整体结构，但是不得不反复处理节点匹配问题。温靖等[8]通过建立不匹配的界面结

收稿日期：2023-11-20；录用日期：2023-12-22；网络出版时间：2024-01-15 18:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20240115.1159.002
基金项目：国家自然科学基金(12102027)；1912项目
*通信作者. E-mail：xiechangc@163.com

引用格式：刘晨宇，谢长川. 基于不匹配结点对接子结构法的超声速颤振预测[J]. 北京航空航天大学学报，2026，52(1)：223-231.
LIU C Y，XIE C C. Supersonic flutter prediction based on unmatched substructure method[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：223-231 (in Chinese).

点位移之间的关系处理，提出了非协调动态子结构的直接综合方法，但是该方法只有位移协调条件，缺少力协调条件。

颤振分析研究弹性结构在气流中的稳定性问题，是气动弹性力学主要的研究内容之一，关乎飞行安全。He等[9]使用可以减缩界面刚性连接自由度的固定界面模态综合法和活塞理论分析了均匀热环境下折叠舵的线性和非线性颤振特性。Bai等[10]使用双协调自由界面模态综合法建立了带有间隙非线性的超声速全动舵面的极限环特性。超声速飞行器的颤振耦合模态可能不是常规的二阶低阶模态耦合，而是多阶甚至包含高阶模态参与耦合，模态耦合形式复杂。尤其对于复杂飞行器而言，其颤振模态耦合阶数是未知的，因此，需要让尽可能多的模态参与计算以避免遗漏，但是过多的模态数会增加计算资源。

超声速飞行器的复杂性为其颤振分析带来了难题。对此，本文以典型的超声速后掠翼模型和壁板模型为研究对象，改进双协调自由界面动态子结构法，在对接结点不匹配时的适用性，对结构进行降阶建模。采用活塞理论计算超声速非定常气动力并推导气动弹性方程；对比研究在颤振计算中传统模态法、传统子结构法及不匹配结点对接子结构法的保留模态数对颤振速度的影响。

1 理论方法

1.1 不匹配结点对接子结构法

子结构的动力学方程可以写为

$$\boldsymbol{M}\ddot{\boldsymbol{u}}+\boldsymbol{K}\boldsymbol{u}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{1}$$

式中：$\boldsymbol{M}$为子结构的物理质量矩阵；$\boldsymbol{K}$为子结构的物理刚度矩阵；$\boldsymbol{u}$为物理位移列向量；$\boldsymbol{B}$为布尔矩阵，用来反映界面结点在所有结点中的位置；$\boldsymbol{f}_{\mathrm{J}}$为界面结点对应的载荷列向量。

根据子结构法定义，结构的位移$\boldsymbol{u}$可以用低阶自由界面模态$\boldsymbol{\Phi}_{\mathrm{k}}$和剩余模态$\boldsymbol{\Psi}_{\mathrm{d}}$来近似表示，从而实现自由度的缩减，表达式为

$$\boldsymbol{u}=\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{p}_{\mathrm{k}}+\boldsymbol{\Psi}_{\mathrm{d}}\boldsymbol{f}_{\mathrm{J}} \tag{2}$$

式中：$\boldsymbol{p}_{\mathrm{k}}$为低阶自由界面模态$\boldsymbol{\Phi}_{\mathrm{k}}$对应的模态坐标；剩余模态$\boldsymbol{\Psi}_{\mathrm{d}}$对应的广义坐标为界面结点对应的载荷列向量$\boldsymbol{f}_{\mathrm{J}}$。

1.1.1 子结构自由界面模态的求解

子结构的自由界面模态可通过求解广义特征值方程得到，表达式为

$$\boldsymbol{k}\boldsymbol{\varphi}=\lambda\boldsymbol{m}\boldsymbol{\varphi} \tag{3}$$

式中：$\boldsymbol{k}$为子结构的广义刚度矩阵；$\boldsymbol{m}$为子结构的广义质量矩阵；$\boldsymbol{\varphi}$为广义特征向量，即模态向量；$\lambda=\omega^2$为广义特征值（ω为结构固有频率）。需要注意的是，这里的自由界面模态并不是指结构完全不受约束，而是指在保留原有约束的基础上不再对子结构界面结点施加其他约束。

设子结构的自由度数为n，则可以得到n阶模态。将模态向量按固有频率从低到高排序，按质量归一化可得

$$\begin{cases}\boldsymbol{\Phi}_n^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\Phi}_n=\boldsymbol{\Lambda}_n=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}\\ \boldsymbol{\Phi}_n^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_n=\boldsymbol{I}_n=\begin{bmatrix}1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1\end{bmatrix}\end{cases} \tag{4}$$

式中：$\boldsymbol{\Phi}_n=\begin{bmatrix}\boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_n\end{bmatrix}$为模态矩阵；$\boldsymbol{\Lambda}_n$为广义特征值的对角阵；$\boldsymbol{I}$为单位矩阵。通常对工程问题有意义的是前几阶模态，因此，只截取前η阶模态向量$\boldsymbol{\varphi}_i(1\leqslant i\leqslant\eta)$组成低阶自由界面模态矩阵$\boldsymbol{\Phi}_{\mathrm{k}}$，表达式为

$$\boldsymbol{\Phi}_{\mathrm{k}}=\begin{bmatrix}\boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_\eta\end{bmatrix}\qquad \eta\ll n \tag{5}$$

1.1.2 子结构剩余模态的求解

子结构的剩余模态的作用是修正模态截断引入的误差。由于高阶模态对结构低频振动特性影响较小，因此，通过静平衡方程求解子结构的剩余模态。子结构在界面力作用下的静平衡方程为

$$\boldsymbol{K}\boldsymbol{u}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{6}$$

式(6)等号两边同时左乘$\boldsymbol{\Phi}_n$得到模态坐标下的方程为

$$\boldsymbol{\Lambda}_n\boldsymbol{p}_n=\boldsymbol{\Phi}_n^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{7}$$

式中：$\boldsymbol{p}_n$为全阶广义坐标。

当子结构没有刚体模态时，$\boldsymbol{\Lambda}_n$对角线上的元素均非零，$\boldsymbol{\Lambda}_n$可逆，可以直接求得位移$\boldsymbol{u}$如下：

$$\boldsymbol{u}=\boldsymbol{\Phi}_n\boldsymbol{p}_n=\boldsymbol{\Phi}_n\boldsymbol{\Lambda}_n^{-1}\boldsymbol{\Phi}_n^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{8}$$

类似地，仅使用低阶自由界面模态进行降阶，则得到的位移为

$$\boldsymbol{u}_{\mathrm{k}}=\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{p}_{\mathrm{k}}=\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}^{-1}\boldsymbol{\Phi}_{\mathrm{k}}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{9}$$

式中：下标k表示保留低阶自由界面模态。

为使降阶前后的位移结果一致，应满足：

$$\begin{aligned}\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{p}_{\mathrm{k}}+\boldsymbol{\Psi}_{\mathrm{d}}\boldsymbol{f}_{\mathrm{J}}&=\left(\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}^{-1}\boldsymbol{\Phi}_{\mathrm{k}}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{\Psi}_{\mathrm{d}}\right)\boldsymbol{f}_{\mathrm{J}}=\\&\boldsymbol{\Phi}_n\boldsymbol{\Lambda}_n^{-1}\boldsymbol{\Phi}_n^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{aligned} \tag{10}$$

可解得剩余模态为

$$\begin{aligned}\boldsymbol{\Psi}_{\mathrm{d}}&=\left(\boldsymbol{\Phi}_n\boldsymbol{\Lambda}_n^{-1}\boldsymbol{\Phi}_n^{\mathrm{T}}-\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}^{-1}\boldsymbol{\Phi}_{\mathrm{k}}^{\mathrm{T}}\right)\boldsymbol{B}^{\mathrm{T}}=\\&\left(\boldsymbol{K}^{-1}-\boldsymbol{\Phi}_{\mathrm{k}}\boldsymbol{\Lambda}_{\mathrm{k}}^{-1}\boldsymbol{\Phi}_{\mathrm{k}}^{\mathrm{T}}\right)\boldsymbol{B}^{\mathrm{T}}\end{aligned} \tag{11}$$

当子结构有刚体模态时, 需要通过惯性释放来求解结构的弹性变形。将结构的位移分解成刚体位移(下标 R)和弹性变形(下标 E):

$$\boldsymbol{u}=\boldsymbol{u}_{\mathrm{R}}+\boldsymbol{u}_{\mathrm{E}} \tag{12}$$

结构刚体运动的加速度为

$$\ddot{\boldsymbol{u}}_{\mathrm{R}}=\boldsymbol{\Phi}_{\mathrm{R}}\ddot{\boldsymbol{p}}_{\mathrm{R}}=\boldsymbol{\Phi}_{\mathrm{R}}\left(\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{13}$$

式中: $\boldsymbol{\Phi}_{\mathrm{R}}$为刚体模态矩阵; $\boldsymbol{p}_{\mathrm{R}}$为刚体模态对应的广义坐标。

引入由刚体运动加速度引起的惯性力, 结构的静平衡方程为

$$\begin{aligned}\boldsymbol{K}\boldsymbol{u}&=-\boldsymbol{M}\ddot{\boldsymbol{u}}_{\mathrm{R}}+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}=\\&\left[\boldsymbol{I}-\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\left(\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\right]\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}=\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{aligned} \tag{14}$$

式中: $\boldsymbol{R}=\boldsymbol{I}-\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\left(\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}$。

由于刚体模态的存在, 刚度矩阵$\boldsymbol{K}$奇异, 无法求逆。为了消除结构的刚体模态, 需要施加约束, 表达式为

$$\boldsymbol{u}=\boldsymbol{C}\boldsymbol{u}_{\boldsymbol{C}} \tag{15}$$

式中: $\boldsymbol{C}$为约束矩阵; $\boldsymbol{u}_{\boldsymbol{C}}$为未被约束的位移自由度。

将式 (15) 代入式 (14), 并在等号两边同时左乘$\boldsymbol{C}^{\mathrm{T}}$得

$$\begin{cases}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{C}\boldsymbol{u}_{\boldsymbol{C}}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\\ \boldsymbol{u}=\boldsymbol{C}\boldsymbol{u}_{\boldsymbol{C}}=\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{cases} \tag{16}$$

由刚体模态和弹性模态的正交性可得

$$\begin{aligned}&\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}=\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\boldsymbol{p}_{\mathrm{R}}\Rightarrow\boldsymbol{p}_{\mathrm{R}}=\left(\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}\Rightarrow\\&\boldsymbol{u}_{\mathrm{R}}=\boldsymbol{\Phi}_{\mathrm{R}}\boldsymbol{p}_{\mathrm{R}}=\boldsymbol{\Phi}_{\mathrm{R}}\left(\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{u}\end{aligned} \tag{17}$$

则弹性变形表达式为

$$\begin{aligned}\boldsymbol{u}_{\mathrm{E}}&=\boldsymbol{u}-\boldsymbol{u}_{\mathrm{R}}=\left[\boldsymbol{I}-\boldsymbol{\Phi}_{\mathrm{R}}\left(\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Phi}_{\mathrm{R}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{R}}^{\mathrm{T}}\boldsymbol{M}\right]\boldsymbol{u}=\\&\boldsymbol{R}^{\mathrm{T}}\boldsymbol{u}=\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{aligned} \tag{18}$$

式(18)中已利用质量矩阵$\boldsymbol{M}$的对称性。

类似地, 仅使用低阶自由界面模态进行降阶, 对式 (14) 进行坐标变换:

$$\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\Phi}_{\mathrm{k,E}}\boldsymbol{p}_{\mathrm{k,E}}=\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}} \tag{19}$$

可解得

$$\begin{aligned}\boldsymbol{u}_{\mathrm{k,E}}&=\boldsymbol{\Phi}_{\mathrm{k,E}}\boldsymbol{p}_{\mathrm{k,E}}=\boldsymbol{\Phi}_{\mathrm{k,E}}\left(\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\Phi}_{\mathrm{k,E}}\right)^{-1}\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}=\\&\boldsymbol{\Phi}_{\mathrm{k,E}}\boldsymbol{\Lambda}_{\mathrm{k,E}}^{-1}\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{aligned} \tag{20}$$

其中, 为了使降阶前后的位移结果一致, 应满足:

$$\begin{aligned}&\boldsymbol{\Phi}_{\mathrm{k,E}}\boldsymbol{p}_{\mathrm{k,E}}+\boldsymbol{\Psi}_{\mathrm{d}}\boldsymbol{f}_{\mathrm{J}}=\left(\boldsymbol{\Phi}_{\mathrm{k,E}}\boldsymbol{\Lambda}_{\mathrm{k,E}}^{-1}\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}+\boldsymbol{\Psi}_{\mathrm{d}}\right)\boldsymbol{f}_{\mathrm{J}}=\\&\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{aligned} \tag{21}$$

可解得剩余模态为

$$\boldsymbol{\Psi}_{\mathrm{d}}=\left[\boldsymbol{R}^{\mathrm{T}}\boldsymbol{C}\left(\boldsymbol{C}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{C}\right)^{-1}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{R}-\boldsymbol{\Phi}_{\mathrm{k,E}}\boldsymbol{\Lambda}_{\mathrm{k,E}}^{-1}\boldsymbol{\Phi}_{\mathrm{k,E}}^{\mathrm{T}}\boldsymbol{R}\right]\boldsymbol{B}^{\mathrm{T}} \tag{22}$$

可以证明剩余模态是高阶自由界面模态的线性组合, 低阶自由界面模态关于刚度矩阵和质量矩阵具有正交性。因此, 使用低阶自由界面模态和剩余模态对式 (1) 进行坐标变换, 可以得到模态坐标下的动力学方程如下:

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{I}_{\mathrm{k}}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{M}_{\mathrm{d}}\end{bmatrix}\begin{bmatrix}\ddot{\boldsymbol{p}}_{\mathrm{k}}\\\ddot{\boldsymbol{f}}_{\mathrm{J}}\end{bmatrix}+\begin{bmatrix}\boldsymbol{\Lambda}_{\mathrm{k}}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{K}_{\mathrm{d}}\end{bmatrix}\begin{bmatrix}\boldsymbol{p}_{\mathrm{k}}\\\boldsymbol{f}_{\mathrm{J}}\end{bmatrix}=\\&\begin{bmatrix}\boldsymbol{\Phi}_{\mathrm{k}}^{\mathrm{T}}\\\boldsymbol{\Psi}_{\mathrm{d}}^{\mathrm{T}}\end{bmatrix}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\end{aligned} \tag{23}$$

式中: $\boldsymbol{M}_{\mathrm{d}}=\boldsymbol{\Psi}_{\mathrm{d}}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{\Psi}_{\mathrm{d}}$; $\boldsymbol{K}_{\mathrm{d}}=\boldsymbol{\Psi}_{\mathrm{d}}^{\mathrm{T}}\boldsymbol{K}\boldsymbol{\Psi}_{\mathrm{d}}$。

针对不匹配的界面结点, 需要使用形态插值理论计算位移插值矩阵和力插值矩阵, 具体理论见 1.3 节。求得插值矩阵后, 子结构界面上的位移协调条件和力协调条件表示为

$$\begin{cases}\boldsymbol{u}_{\alpha,\mathrm{J}}=\boldsymbol{G}\boldsymbol{u}_{\beta,\mathrm{J}}\\ \boldsymbol{G}^{\mathrm{T}}\boldsymbol{f}_{\alpha,\mathrm{J}}=-\boldsymbol{f}_{\beta,\mathrm{J}}\end{cases} \tag{24}$$

式中: $\boldsymbol{u}_{\alpha,\mathrm{J}}$和$\boldsymbol{u}_{\beta,\mathrm{J}}$分别为子结构$\alpha$和$\beta$界面结点的位移向量; $\boldsymbol{f}_{\alpha,\mathrm{J}}$和$\boldsymbol{f}_{\beta,\mathrm{J}}$分别为子结构$\alpha$和$\beta$界面结点的力向量; $\boldsymbol{G}$为插值矩阵。

将各子结构在模态坐标下的动力学方程写在一起, 以 2 个子结构(子结构α和β)为例:

$$\begin{cases}\begin{bmatrix}\tilde{\boldsymbol{M}}_{\alpha}&\boldsymbol{0}\\\boldsymbol{0}&\tilde{\boldsymbol{M}}_{\beta}\end{bmatrix}\begin{bmatrix}\ddot{\boldsymbol{p}}_{\alpha}\\\ddot{\boldsymbol{p}}_{\beta}\end{bmatrix}+\begin{bmatrix}\tilde{\boldsymbol{K}}_{\alpha}&\boldsymbol{0}\\\boldsymbol{0}&\tilde{\boldsymbol{K}}_{\beta}\end{bmatrix}\begin{bmatrix}\boldsymbol{p}_{\alpha}\\\boldsymbol{p}_{\beta}\end{bmatrix}=\\\quad\begin{bmatrix}\left(\begin{bmatrix}\boldsymbol{\Phi}_{\mathrm{k}}&\boldsymbol{\Psi}_{\mathrm{d}}\end{bmatrix}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\right)_{\alpha}\\\left(\begin{bmatrix}\boldsymbol{\Phi}_{\mathrm{k}}&\boldsymbol{\Psi}_{\mathrm{d}}\end{bmatrix}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{f}_{\mathrm{J}}\right)_{\beta}\end{bmatrix}\\\tilde{\boldsymbol{M}}_{\theta}=\begin{bmatrix}\boldsymbol{I}_{\mathrm{k}}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{M}_{\mathrm{d}}\end{bmatrix}_{\theta}\\\tilde{\boldsymbol{K}}_{\theta}=\begin{bmatrix}\boldsymbol{\Lambda}_{\mathrm{k}}&\boldsymbol{0}\\\boldsymbol{0}&\boldsymbol{K}_{\mathrm{d}}\end{bmatrix}_{\theta}\qquad\theta=\alpha,\beta\\\boldsymbol{p}_{\theta}=\begin{bmatrix}\boldsymbol{p}_{\mathrm{k}}\\\boldsymbol{f}_{\mathrm{J}}\end{bmatrix}_{\theta}\end{cases} \tag{25}$$

从式 (24) 中消去$\boldsymbol{f}_{\alpha,\mathrm{J}}$、$\boldsymbol{f}_{\beta,\mathrm{J}}$, 得

$$\begin{cases}\begin{bmatrix}\boldsymbol{p}_{\alpha,\mathrm{k}}\\\boldsymbol{f}_{\alpha,\mathrm{J}}\\\boldsymbol{p}_{\beta,\mathrm{k}}\\\boldsymbol{f}_{\beta,\mathrm{J}}\end{bmatrix}=\boldsymbol{T}\begin{bmatrix}\boldsymbol{p}_{\alpha,\mathrm{k}}\\\boldsymbol{p}_{\beta,\mathrm{k}}\end{bmatrix}\\\boldsymbol{T}=\begin{bmatrix}\boldsymbol{I}&\boldsymbol{0}\\-\boldsymbol{E}\boldsymbol{B}_{\alpha}\boldsymbol{\Phi}_{\alpha,\mathrm{k}}&\boldsymbol{E}\boldsymbol{G}\boldsymbol{B}_{\beta}\boldsymbol{\Phi}_{\beta,\mathrm{k}}\\\boldsymbol{0}&\boldsymbol{I}\\\boldsymbol{G}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{B}_{\alpha}\boldsymbol{\Phi}_{\alpha,\mathrm{k}}&-\boldsymbol{G}^{\mathrm{T}}\boldsymbol{E}\boldsymbol{G}\boldsymbol{B}_{\beta}\boldsymbol{\Phi}_{\beta,\mathrm{k}}\end{bmatrix}\\\boldsymbol{E}=\left(\boldsymbol{B}_{\alpha}\boldsymbol{\Psi}_{\alpha,\mathrm{d}}+\boldsymbol{G}\boldsymbol{B}_{\beta}\boldsymbol{\Psi}_{\beta,\mathrm{d}}\boldsymbol{G}^{\mathrm{T}}\right)^{-1}\end{cases} \tag{26}$$

最终得到整体结构在独立广义坐标下的动力学方程为

$$\begin{cases}\overline{\boldsymbol{M}}\begin{bmatrix}\ddot{\boldsymbol{p}}_{\alpha,\mathrm{k}}\\\ddot{\boldsymbol{p}}_{\beta,\mathrm{k}}\end{bmatrix}+\overline{\boldsymbol{K}}\begin{bmatrix}\boldsymbol{p}_{\alpha,\mathrm{k}}\\\boldsymbol{p}_{\beta,\mathrm{k}}\end{bmatrix}=\boldsymbol{0}\\\overline{\boldsymbol{M}}=\boldsymbol{T}^{\mathrm{T}}\begin{bmatrix}\tilde{\boldsymbol{M}}_{\alpha}&\boldsymbol{0}\\\boldsymbol{0}&\tilde{\boldsymbol{M}}_{\beta}\end{bmatrix}\boldsymbol{T}\\\overline{\boldsymbol{K}}=\boldsymbol{T}^{\mathrm{T}}\begin{bmatrix}\tilde{\boldsymbol{K}}_{\alpha}&\boldsymbol{0}\\\boldsymbol{0}&\tilde{\boldsymbol{K}}_{\beta}\end{bmatrix}\boldsymbol{T}\end{cases} \tag{27}$$

1.2 非定常气动力建模

采用传统活塞理论计算超声速非定常气动力。活塞理论是应用最为广泛的超声速非定常气动力的工程计算方法。传统活塞理论[11]诞生于20世纪50年代，由于形式简单，计算简便，且在马赫数2～5范围内有较高的计算精度，在超声速飞行器颤振和热气动弹性分析中得到了广泛的应用。该理论的基本假设是超声速气流中物面上某一点的扰动对其他点的影响可以忽略不计，即认为一点的干扰压力只与该点的下洗有关，这种情况下产生的压力与一维活塞运动十分相似，即活塞在一维管道中运动时所受的压力只与其运动速度有关，如图1所示。

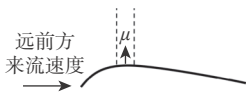


图1 传统活塞理论示意图

Fig. 1 Classical piston theory

管道中的气体方程满足：

$$\begin{cases} \mathrm{d}P=\rho\xi\mathrm{d}\mu \\ \dfrac{P}{P_\infty}=\left(\dfrac{\rho}{\rho_\infty}\right)^{\gamma} \\ \xi^2=\gamma\dfrac{P}{\rho} \end{cases} \tag{28}$$

式中：第1项为牛顿第二定律；第2项为气体等熵条件；第3项为声速公式；P为压强；ρ为密度；ξ为声速；μ为活塞行进的速度；γ为比热比；带下标∞的物理量表示远前方来流参数。

通过推导可以得到活塞理论的表达式为

$$P=P_\infty\left(1+\frac{\gamma-1}{2}\cdot\frac{\mu}{\xi_\infty}\right)^{\frac{2\gamma}{\gamma-1}} \tag{29}$$

式(29)展开后得到活塞理论压强差表达式为

$$\Delta P(x,y,z)=P-P_\infty=\rho_\infty\xi_\infty^2\left[\frac{\mu}{\xi_\infty}+\frac{\gamma+1}{4}\left(\frac{\mu}{\xi_\infty}\right)^2+\frac{\gamma+1}{12}\left(\frac{\mu}{\xi_\infty}\right)^3+\cdots\right] \tag{30}$$

省略高阶项，当只保留一阶项时，称为一阶活塞理论，扰动压力为

$$\Delta P=\rho_\infty\xi_\infty\mu \tag{31}$$

一阶活塞理论属于线性准定常气动理论，适用范围为：$Ma_\infty\gg1$，$Ma_\infty\tau\ll1$。其中，Ma为马赫数，τ为物面角。因此，传统活塞理论适用于小迎角薄翼。

对于工程复杂模型，往往通过有限元方法将结构离散化，通过模态表达结构的振动特性。将模态引入，得到活塞理论的广义气动力$\boldsymbol{F}$的表达式为[12]

$$\boldsymbol{F}=\boldsymbol{a}_1\dot{\boldsymbol{q}}+\boldsymbol{a}_0\boldsymbol{q} \tag{32}$$

式中：$\boldsymbol{a}_1$和$\boldsymbol{a}_0$为活塞理论的广义气动力影响系数矩阵；$\boldsymbol{q}$为广义坐标。

1.3 形态插值理论

在对气动弹性系统进行数学建模时，结构离散化与流场离散化所考虑的因素很不相同，计算方法差异很大，表现在结构与流场界面的网格划分不可能完全匹配和一致。因此，在不同物理域的界面之间如何传递力学信息也是气动弹性研究的重要课题，在数学上就是要研究适用于气动弹性分析的各种插值和拟合方法。在超声速气流中，翼舵、壁板等结构在气动载荷的作用下产生变形，其表面的气动网格也随之变形为空间曲面，因此，气动弹性分析需要使用三维曲面插值来实现气动和结构的耦合。

基于形态插值理论[13-14]，可以实现气动弹性分析中的结构网格与气动网格间的位移和载荷插值。结构结点坐标可以看作三维向量$\boldsymbol{X}_\mathrm{S}$，结构结点的位移对应已知的函数值$\boldsymbol{U}_\mathrm{S}$，气动网格的结点坐标为三维向量$\boldsymbol{X}_\mathrm{A}$，待求气动网格结点位移$\boldsymbol{U}_\mathrm{A}$，具体表示为

$$\begin{cases} \boldsymbol{X}_\mathrm{S}=\begin{bmatrix} x_{\mathrm{S}1} & y_{\mathrm{S}1} & z_{\mathrm{S}1} \\ \vdots & \vdots & \vdots \\ x_{\mathrm{S}l} & y_{\mathrm{S}l} & z_{\mathrm{S}l} \end{bmatrix},\boldsymbol{U}_\mathrm{S}=\begin{bmatrix} u_{\mathrm{S}1} & v_{\mathrm{S}1} & w_{\mathrm{S}1} \\ \vdots & \vdots & \vdots \\ u_{\mathrm{S}l} & v_{\mathrm{S}l} & w_{\mathrm{S}l} \end{bmatrix} \\ \boldsymbol{X}_\mathrm{A}=\begin{bmatrix} x_{\mathrm{A}1} & y_{\mathrm{A}1} & z_{\mathrm{A}1} \\ \vdots & \vdots & \vdots \\ x_{\mathrm{A}m} & y_{\mathrm{A}m} & z_{\mathrm{A}m} \end{bmatrix},\boldsymbol{U}_\mathrm{A}=\begin{bmatrix} u_{\mathrm{A}1} & v_{\mathrm{A}1} & w_{\mathrm{A}1} \\ \vdots & \vdots & \vdots \\ u_{\mathrm{A}m} & v_{\mathrm{A}m} & w_{\mathrm{A}m} \end{bmatrix} \end{cases} \tag{33}$$

可以得到插值系数矩阵为

$$\boldsymbol{A}_\mathrm{S}=\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & 0 & x_{\mathrm{S}1} & x_{\mathrm{S}2} & \cdots & x_{\mathrm{S}l} \\ 0 & 0 & 0 & 0 & y_{\mathrm{S}1} & y_{\mathrm{S}2} & \cdots & y_{\mathrm{S}l} \\ 0 & 0 & 0 & 0 & z_{\mathrm{S}1} & z_{\mathrm{S}2} & \cdots & z_{\mathrm{S}l} \\ 1 & x_{\mathrm{S}1} & y_{\mathrm{S}1} & z_{\mathrm{S}1} & h_1 & b_{12} & \cdots & b_{1l} \\ 1 & x_{\mathrm{S}2} & y_{\mathrm{S}2} & z_{\mathrm{S}2} & b_{21} & h_2 & \cdots & b_{2l} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{\mathrm{S}l} & y_{\mathrm{S}l} & z_{\mathrm{S}l} & b_{l1} & b_{l2} & \cdots & h_l \end{bmatrix} \tag{34}$$

式(34)可以具体写为

$$\boldsymbol{A}_\mathrm{S}\boldsymbol{Z}_\mathrm{g}=\boldsymbol{W}_\mathrm{S} \tag{35}$$

式中：$\boldsymbol{Z}_\mathrm{g}$为待定系数矩阵；$\boldsymbol{W}_\mathrm{S}=\begin{bmatrix}\boldsymbol{0}_{4\times3}\\ \boldsymbol{U}_\mathrm{S}\end{bmatrix}$。

气动网格结点坐标对应的矩阵为

$$\boldsymbol{A}_\mathrm{A}=\begin{bmatrix} 1 & x_{\mathrm{A}1} & y_{\mathrm{A}1} & z_{\mathrm{A}1} & h_1 & b_{12} & \cdots & b_{1m} \\ 1 & x_{\mathrm{A}2} & y_{\mathrm{A}2} & z_{\mathrm{A}2} & b_{21} & h_2 & \cdots & b_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{\mathrm{A}m} & y_{\mathrm{A}m} & z_{\mathrm{A}m} & b_{m1} & b_{m2} & \cdots & h_m \end{bmatrix}$$

可得气动网格的位移为

$$\boldsymbol{U}_\mathrm{A}=\boldsymbol{A}_\mathrm{A}\boldsymbol{A}_\mathrm{S}^{-1}\boldsymbol{W}_\mathrm{S} \tag{36}$$

由于 $\boldsymbol{W}_{\mathrm{S}}$ 前 4 行元素均为零, 因此, 结构结点到气动结点的位移插值关系可进一步写为

$$\boldsymbol{U}_{\mathrm{A}}=\boldsymbol{G}\boldsymbol{U}_{\mathrm{S}} \tag{37}$$

式中: $\boldsymbol{G}=\boldsymbol{A}_{\mathrm{A}}\boldsymbol{A}_{\mathrm{S}}^{-1}$ 为插值矩阵。

气动弹性分析中的另一个插值问题是载荷插值, 需要将气动网格上的载荷等效到结构结点上, 从而进行结构动力学分析。在载荷插值的过程中, 不仅需要满足合力、合力矩等效, 还需要满足虚功原理, 即在结构任意可能的虚位移下 2 套载荷分布造成的虚功相等。假设结构界面的虚位移为 $\delta\boldsymbol{U}_{\mathrm{S}}$, 气动界面的虚位移为 $\delta\boldsymbol{U}_{\mathrm{A}}$, 气动界面的载荷为 $\boldsymbol{F}_{\mathrm{A}}$, 结构界面上的等效载荷为 $\boldsymbol{F}_{\mathrm{S}}$。虚功原理表达式写为

$$\delta\boldsymbol{U}_{\mathrm{A}}^{\mathrm{T}}\boldsymbol{F}_{\mathrm{A}}=\delta\boldsymbol{U}_{\mathrm{S}}^{\mathrm{T}}\boldsymbol{F}_{\mathrm{S}} \tag{38}$$

根据位移插值关系, 气动与结构模型间的位移满足:

$$\delta\boldsymbol{U}_{\mathrm{A}}=\boldsymbol{G}\delta\boldsymbol{U}_{\mathrm{S}} \tag{39}$$

根据虚位移的任意性, 可以推出气动模型与结构模型之间的力插值关系式为

$$\boldsymbol{F}_{\mathrm{S}}=\boldsymbol{G}^{\mathrm{T}}\boldsymbol{F}_{\mathrm{A}} \tag{40}$$

1.4　颤振分析方法

典型的气动弹性方程为

$$\boldsymbol{M}_{\mathrm{W}}\ddot{\boldsymbol{u}}_{\mathrm{W}}+\boldsymbol{K}_{\mathrm{W}}\boldsymbol{u}_{\mathrm{W}}=\boldsymbol{F}_{\mathrm{W}} \tag{41}$$

式中: $\boldsymbol{M}_{\mathrm{W}}$ 为结构物理质量矩阵; $\boldsymbol{u}_{\mathrm{W}}$ 为物理位移; $\boldsymbol{K}_{\mathrm{W}}$ 为结构的物理刚度矩阵; $\boldsymbol{F}_{\mathrm{W}}$ 为实际气动力。

对式(41)用结构模态进行广义化表示, 并在右端项中代入当地流活塞理论的广义气动力表达式, 得

$$\boldsymbol{m}_{\mathrm{W}}\ddot{\boldsymbol{q}}_{\mathrm{W}}+\boldsymbol{k}_{\mathrm{W}}\boldsymbol{q}_{\mathrm{W}}=\boldsymbol{a}_{1\mathrm{W}}\dot{\boldsymbol{q}}_{\mathrm{W}}+\boldsymbol{a}_{0\mathrm{W}}\boldsymbol{q}_{\mathrm{W}} \tag{42}$$

式中: $\boldsymbol{m}_{\mathrm{W}}$ 为结构广义质量矩阵; $\boldsymbol{k}_{\mathrm{W}}$ 为结构广义刚度矩阵; $\boldsymbol{q}_{\mathrm{W}}$ 为广义坐标; $\boldsymbol{a}_{1\mathrm{W}}$ 和 $\boldsymbol{a}_{0\mathrm{W}}$ 为当地流活塞理论的广义气动力影响系数矩阵。

将式 (42) 表示为状态空间形式:

$$\begin{bmatrix}\ddot{\boldsymbol{q}}_{\mathrm{W}}\\ \dot{\boldsymbol{q}}_{\mathrm{W}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{m}_{\mathrm{W}}^{-1}\boldsymbol{a}_{1\mathrm{W}} & \boldsymbol{m}_{\mathrm{W}}^{-1}(\boldsymbol{a}_{0\mathrm{W}}-\boldsymbol{k}_{\mathrm{W}})\\ \boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\dot{\boldsymbol{q}}_{\mathrm{W}}\\ \boldsymbol{q}_{\mathrm{W}}\end{bmatrix} \tag{43}$$

若式(43)的广义化模态是按质量归一的, 则 $\boldsymbol{m}_{\mathrm{W}}=\boldsymbol{I}$, 并且 $\boldsymbol{m}_{\mathrm{W}}^{-1}\boldsymbol{k}_{\mathrm{W}}=\boldsymbol{\Lambda}$($\boldsymbol{\Lambda}$为结构特征角频率平方的对角阵), 则式 (43) 简化为

$$\begin{bmatrix}\ddot{\boldsymbol{q}}_{\mathrm{W}}\\ \dot{\boldsymbol{q}}_{\mathrm{W}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{a}_{1\mathrm{W}} & \boldsymbol{a}_{0\mathrm{W}}-\boldsymbol{\Lambda}\\ \boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\begin{bmatrix}\dot{\boldsymbol{q}}_{\mathrm{W}}\\ \boldsymbol{q}_{\mathrm{W}}\end{bmatrix} \tag{44}$$

根据线性系统理论, 线性时不变系统渐近稳定(即不发生颤振)的充分必要条件是: 系统的状态矩阵 $\boldsymbol{H}$ 的全部特征值均位于复平面左半部, 即

$$\begin{cases}\boldsymbol{H}=\begin{bmatrix}\boldsymbol{a}_{1\mathrm{W}} & \boldsymbol{a}_{0\mathrm{W}}-\boldsymbol{\Lambda}\\ \boldsymbol{I} & \boldsymbol{0}\end{bmatrix}\\ \mathrm{Re}[\lambda_i(\boldsymbol{H})]<0 \qquad i=1,2,\cdots,2n\end{cases} \tag{45}$$

式中: $\mathrm{Re}[\lambda_i(\boldsymbol{H})]$表示取复数 $\lambda_i(\boldsymbol{H})$的实部。

在计算多个工况时, 每个速度状态都能求出一组广义气动力影响系数矩阵 $\boldsymbol{a}_{1\mathrm{W}}$ 和 $\boldsymbol{a}_{0\mathrm{W}}$, 即对应一个状态矩阵 $\boldsymbol{H}$ 及其特征值, 由此可以将速度由低到高的特征值实部和虚部绘制根轨迹图和 VG、VF 图, 并通过式 (45) 寻找首个不稳定的速度点作为颤振速度边界。

但在求解含参数系统特征值时, 常规数值方法并不对特征值参数的演化关系进行跟踪, 因此, 对一系列参数值对应的特征根进行排序时容易发生“窜支”现象, 表现为某条根轨迹曲线失去光滑性, 通常根据分析经验人工调整数据排列得到光顺曲线, 为颤振分析带来极大不便, 因此, 需要应用特征值排序方法。

由于线性气动弹性系统的特征值连续依赖于来流速度, 根据一阶连续性建立特征值随来流速度变化的预测排序算法。当参数变化为小量时, 各阶特征值的变化也为小量[15]。在复平面上, 由第 t–2 个和 t–1 个速度点对应的各特征值及其一阶变化率预测第 t 个速度点对应的各特征值, 并将第 t 个速度点对应的各特征值的实际值与预测值进行最小距离(欧氏距离)匹配, 从而得到第 t–1 个和第 t 个速度点对应的各特征值的匹配关系。将该过程按连续速度点依次进行, 得到所有速度点对应的特征值匹配关系, 如图 2 所示。

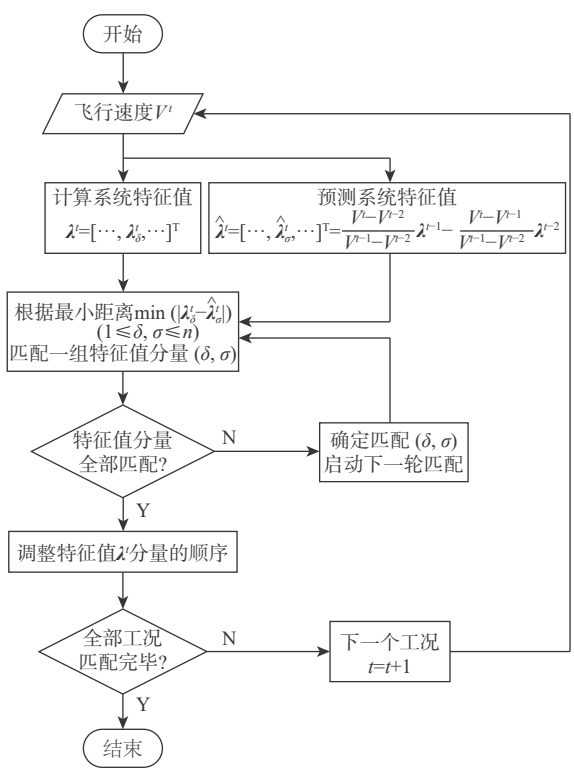


图 2　特征值排序

Fig. 2　Eigenvalue sorting

2 算例分析

2.1 后掠翼模型

2.1.1 后掠翼模型参数

机翼的几何外形和参数如图3所示，机翼后掠角为15°，展弦比为2.67，翼型剖面厚度为1.041 4 mm。翼根采用固支边界条件，模型参数如表1所示。

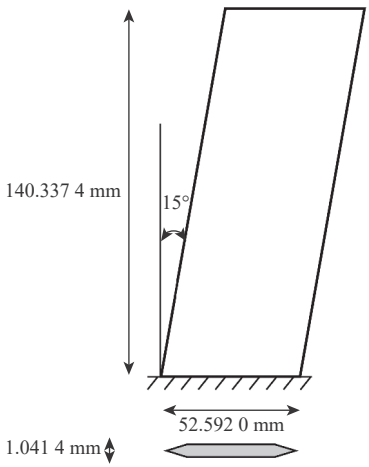


图3 后掠翼模型

Fig. 3 Swept wing model

表1 后掠翼模型参数

Table 1 Parameters of the swept wing model

参数	数值
弹性模量/10^{10}Pa	4.39
密度/10^3(kg·m^{-3})	1.73
泊松比	0.25
空气参考密度/(kg·m^{-3})	1.225
空气密度比	0.391

2.1.2 后掠翼模型不匹配结点对接子结构法建模

将机翼的整体结构划分为2个子结构，为使模型具有一般性，2个子结构在离散化后，其对接界面具有不一致的结点分布，如图4所示。

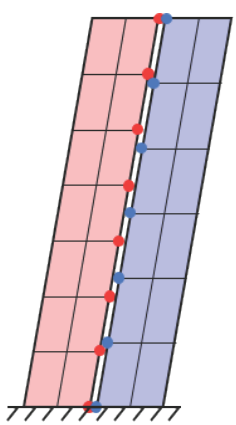


图4 后掠翼子结构划分

Fig. 4 Swept wing substructure division

若2个子结构各取5阶保留模态，最终得到的降阶模型的阶数是10。与完整的机翼有限元模型的前10阶相比，通过子结构法降阶后的频率相对误差分布如图5所示。其中，主要低阶(前5阶)模态频率的相对误差不大于1.1%，证明本文提出的不匹配结点对接子结构法能够对高阶模型进行降阶建模并保留其主要动力学特性。

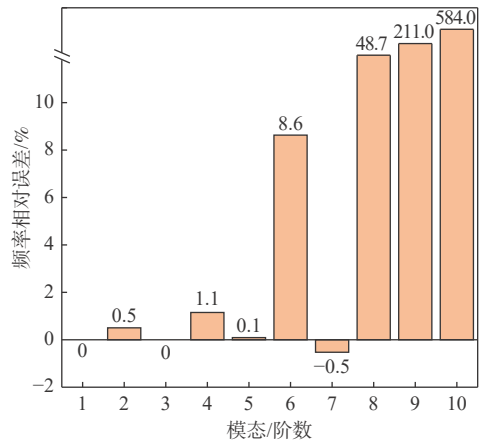


图5 后掠翼子结构频率相对误差

Fig. 5 Frequency relative error of the swept wing substructure

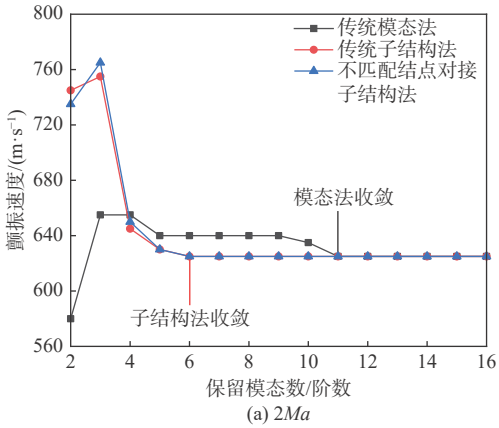
值得注意的是，子结构模型包含了全阶有限元模型中被截断的高阶项的影响，这些截断项的影响集中在子结构模型的高阶项中，在图5中表现为第8、9、10阶的“误差”较大，而这种由于考虑了截断项的影响而引起的“误差”对于颤振分析是有益的。

2.1.3 后掠翼模型颤振计算

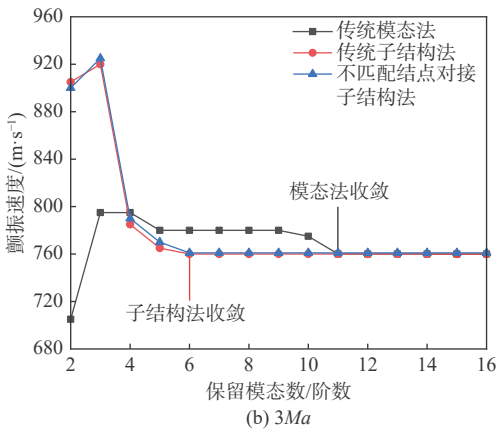
对于一般飞行器模型，由于其颤振模态的耦合阶数通常是未知的，为了研究保留模态数与颤振速度的关系，需要逐渐增加保留模态的数量并计算相应的颤振速度，直到颤振速度不再改变。

图6展示了基于传统模态法、传统子结构法、不匹配结点对接子结构法的颤振速度随保留模态数的变化规律，本文3种方法的颤振速度统一收敛于625 m/s (2*Ma*)和760 m/s (3*Ma*)，与商业软件MSC. Nastran的颤振计算结果630 m/s (2*Ma*)和770 m/s (3*Ma*)吻合较好，相对误差分别为−0.8%和−1.3%，说明对于该模型而言，子结构法对颤振速度的预测结果是准确的。

子结构法的颤振速度比传统模态法更早收敛，需要的保留模态数更少(6阶和11阶)，原因是子结构法建立的低阶动力学模型含有高阶项的影响，于是用较少的阶数就能考虑到某些高阶的颤振耦合模态。在本例中，6阶子结构模型中就能考虑传统的第11阶模态(最高阶颤振耦合模态)，于是，子结构法的颤振速度在第6阶时就收敛了。而传统模态法建立的低阶动力学模型直接对模态截尾，完全



(a) 2*Ma*



(b) 3*Ma*

图 6　3 种方法的后掠翼颤振速度随保留模态数的变化规律

Fig. 6　The variation of swept wing flutter velocity with the number of kept modes of the three methods

舍弃了高阶项, 因此, 必须达到 11 阶才能得到收敛的颤振速度。

此外, 传统子结构法和不匹配结点对接子结构法的结果具有很好的一致性, 说明本文方法在对子结构法做改进时引入的插值误差可以忽略不计, 再次验证本文方法的精度和有效性。

2.2　壁板模型

2.2.1　壁板模型参数

壁板的几何外形和参数如图 7 所示, 厚度为 1.041 4 mm。壁板 4 边采用简支边界条件, 模型参数如表 2 所示。

2.2.2　壁板模型不匹配结点对接子结构法建模

将壁板的整体结构划分为 2 个子结构, 为使模型具有一般性, 2 个子结构在离散化后, 其对接界面具有不一致的结点分布, 如图 8 所示。

若 2 个子结构各取 5 阶保留模态, 最终得到的降阶模型的阶数是 10。与完整的壁板有限元模型的前 10 阶相比, 通过子结构法降阶后的频率相对误差分布如图 9 所示。其中, 主要低阶(前 7 阶)模态频率的相对误差不大于 0.3%, 证明本文提出的不匹配结点对接子结构法能够对高阶模型进行降阶建模并保留其主要动力学特性。

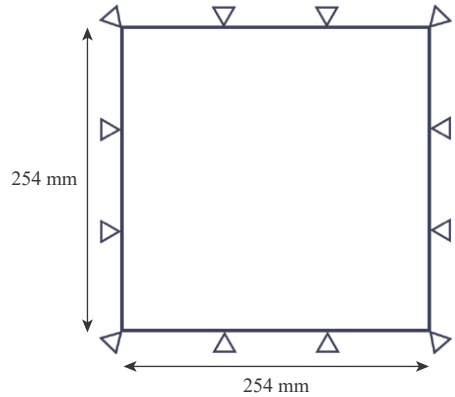


图 7　壁板模型

Fig. 7　Panel model

表 2　壁板模型参数

Table 2　Parameters of the panel model

参数	数值
厚度/mm	1.041 4
弹性模量/10^{10} Pa	7.1
密度/10^3(kg·m^{-3})	2.768
泊松比	0.32
空气参考密度/(kg·m^{-3})	1.225
空气密度比	1.0

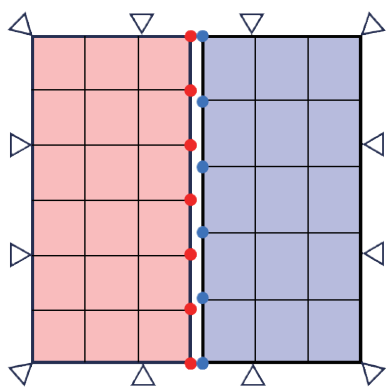


图 8　壁板子结构划分

Fig. 8　Panel substructure division

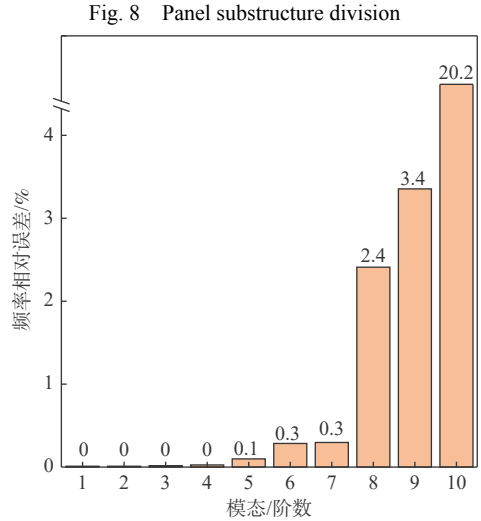


图 9　壁板子结构频率相对误差

Fig. 9　Frequency relative error of the panel substructure

2.2.3 壁板模型颤振计算

图10展示了基于传统模态法、传统子结构法、不匹配结点对接子结构法的颤振速度随保留模态数的变化规律，本文3种方法的颤振速度统一收敛于614 m/s (2*Ma*)和749 m/s (3*Ma*)，与商业软件MSC.Nastran的颤振计算结果612 m/s (2*Ma*)和747 m/s (3*Ma*)吻合较好，相对误差分别为0.33%和0.27%；子结构法的颤振速度比模态法更早收敛，需要的保留模态数更少(6阶和9阶)。与2.1节的结论类似，本算例也验证了方法的有效性。

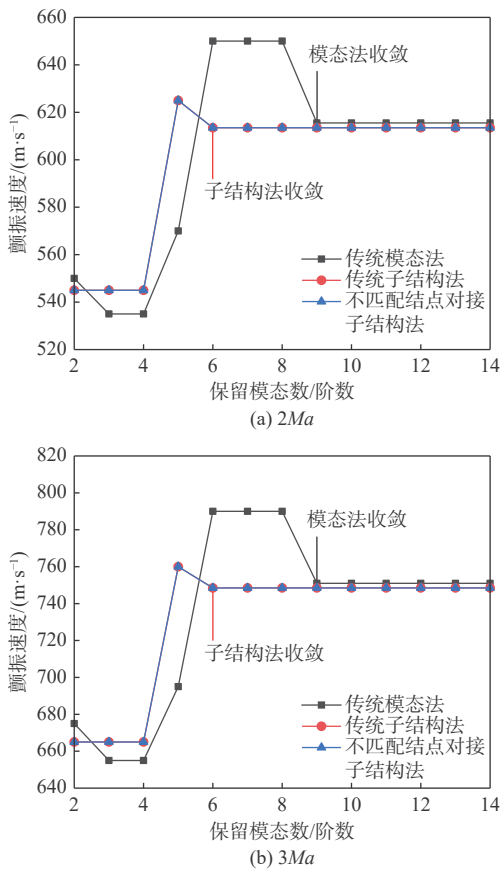


图10 3种方法的壁板颤振速度随保留模态数的变化规律

Fig. 10 The variation of panel flutter velocity with number of kept modes of the three methods

3 结 论

本文改进了传统双协调自由界面动态子结构法在对接结点不匹配时的适用性，并提出了一种基于不匹配结点对接子结构法降阶的超声速颤振分析方法，针对典型后掠翼模型和壁板模型，主要得到以下结论：

1) 改进的不匹配结点对接子结构法能够处理对接结点不匹配的问题，显著降低模型自由度数并保留其主要动力学特性。降阶模型的低阶模态精度非常好(低阶模态频率相对误差不大于1.1%)。改进前后的子结构法对降阶模型的高阶模态做了修正，补偿了原始模型中高阶模态的动力学贡献，而传统模态法在做模态截断后完全不考虑该部分贡献，因此，子结构降阶模型比同阶的传统模态法更准确，这也是子结构法更快收敛的原因。

2) 本文的颤振计算结果与商业软件MSC.Nastran吻合较好(颤振速度相对误差不大于1.3%)，验证了方法的有效性。在颤振分析中，随着保留模态数的增加，子结构法的颤振速度比模态法更快收敛，即子结构法需要的保留模态数(6阶)比模态法(11阶和9阶)更少。因此，对于颤振模态耦合阶数未知或耦合形式复杂的一般飞行器模型而言，采用基于子结构法降阶的颤振计算方法不易遗漏颤振耦合模态，从而减少了迭代计算的工作量。

参考文献（References）

[1] CRAIG R R. Coupling of substructures for dynamic analyses-an overview[C]//Proceedings of the 41st Structures, Structural Dynamics, and Materials Conference and Exhibit. Reston: AIAA, 2000: 1573.

[2] 杜飞平, 谭永华, 陈建华, 等. 航天器子结构模态综合法研究现状及进展[J]. 火箭推进, 2010, 36(3): 39-44.
DU F P, TAN Y H, CHEN J H, et al. Status and progress on substructural modal synthesis techniques for space vehicles[J]. Journal of Rocket Propulsion, 2010, 36(3): 39-44(in Chinese).

[3] 向树红, 邱吉宝, 王大钧. 模态分析与动态子结构方法新进展[J]. 力学进展, 2004, 34(3): 289-303.
XIANG S H, QIU J B, WANG D J. The resent progresses on modal analysis and dynamic sub-structure methods[J]. Advances in Mechanics, 2004, 34(3): 289-303(in Chinese).

[4] KRATTIGER D, WU L, ZACHARCZUK M, et al. Interface reduction for Hurty/Craig-Bampton substructured models: review and improvements[J]. Mechanical Systems and Signal Processing, 2019, 114: 579-603.

[5] YANG N, WANG N, ZHANG X, et al. Nonlinear flutter wind tunnel test and numerical analysis of folding fins with freeplay nonlinearities[J]. Chinese Journal of Aeronautics, 2016, 29(1): 144-159.

[6] 孙玉凯, 杨超, 吴志刚. 含间隙非线性二元翼段的系统辨识[J]. 北京航空航天大学学报, 2021, 47(1): 140-149.
SUN Y K, YANG C, WU Z G. System identification of a z-DOF wing section with freeplay nonlinearity[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1): 140-149(in Chinese).

[7] MUSCOLINO G. Dynamic analysis of structural systems using component mode synthesis[C]//Proceedings of the 6th International Conference on Computational Structures Technology. Edinburgh: Computational Structures Technology, 2002: 255-282.

[8] 温靖, 尚菲菲, 裔英明. 非协调动态子结构的直接综合方法[J]. 工程技术研究, 2022, 7(22): 8-10.
WEN J, SHANG F F, YI Y M. Direct synthesis method of incompatible dynamic substructures[J]. Engineering and Technological Research, 2022, 7(22): 8-10(in Chinese).

[9] HE H N, TANG H, YU K P, et al. Nonlinear aeroelastic analysis of the folding fin with freeplay under thermal environment[J]. Chinese Journal of Aeronautics, 2020, 33(9): 2357-2371.

[10] BAI L Y, WU Z G, YANG C. Nonlinear aeroelastic characteristics of a supersonic all-movable fin with single or multiple freeplay nonlinearities[J]. Applied Sciences, 2023, 13(3): 1262.

[11] MEIJER M C, DALA L. Generalized formulation and review of piston theory for airfoils[J]. AIAA Journal, 2015, 54(1): 17-27.

[12] LIU C Y, XIE C C, MENG Y, et al. Experimental and numerical flutter analysis using local piston theory with viscous correction[J]. Aerospace, 2023, 10(10): 870.

[13] XIE C C, YANG C. Surface splines generalization and large deflection interpolation[J]. Journal of Aircraft, 2007, 44(3): 1024-1026.

[14] 谢长川, 张铎耀, 安朝. 基于动响应数据的大柔性机翼结构降阶方法[J]. 北京航空航天大学学报, 2023, 49(6): 1319-1330.
XIE C C, ZHANG D Y, AN C. Reduced order method for large flexible wing structure based on dynamic response data[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(6): 1319-1330(in Chinese).

[15] 谢长川, 胡薇薇, 杨超. 颤振分析中的特征值排序问题[J]. 数学的实践与认识, 2007, 37(18): 141-146.
XIE C C, HU W W, YANG C. Eigenvalue sorting problem in flutter analysis[J]. Mathematics in Practice and Theory, 2007, 37(18): 141-146(in Chinese).

Supersonic flutter prediction based on unmatched substructure method

LIU Chenyu，XIE Changchuan*

(School of Aeronautic Science and Engineering，Beihang University，Beijing 100191，China)

Abstract：Accurate prediction of the flutter velocity boundary of a supersonic vehicle is one of the key and difficult points in the aeroelasticity research of a supersonic vehicle. Aiming at the problems of a high degree of freedom and complex flutter mode coupling form of the supersonic vehicle model, a structural reduce-order modeling method based on the unmatched substructure method is proposed, and aeroelastic flutter modeling is realized by combining unsteady piston theory and fluid-solid coupling interpolation theory. The conventional mode method, the traditional substructure method, and the unmatched substructure method are used to investigate the supersonic flutter characteristics of a panel model and a swept wing model. The results show that, compared with commercial software, the relative error of the main low-order frequency and flutter velocity of the model is less than 1.1% and 1.3% respectively. Compared with the traditional modal method, the substructure method requires fewer modal orders. As a result, the suggested approach can effectively and precisely forecast the flutter characteristics of common supersonic aircraft.

Keywords：substructure; flutter; aeroelasticity; piston theory; interpolation

Received：2023-11-20；**Accepted**：2023-12-22；**Published Online**：2024-01-15 18：55

URL：link.cnki.net/urlid/11.2625.V.20240115.1159.002

Foundation items：National Natural Science Foundation of China (12102027); 1912 Project

* **Corresponding author**. E-mail：xiechangc@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0753

基于深度学习的双域信息CT金属伪影抑制方法

海潮[1]，田鑫[1]，张宏[2]，谭大龙[1]，何一新[1]，孟凡勇[3]，杨民[1,*]

(1. 北京航空航天大学 机械工程及自动化学院，北京 100191；2. 北京动力机械研究所，北京 100074；
3. 中国科学院过程工程研究所 介科学与工程全国重点实验室，北京 100190)

摘 要：当CT扫描视野中出现金属时，重建图像不可避免地会产生金属伪影，严重影响图像质量。为了抑制金属伪影，提出一种新的深度学习CT金属伪影抑制(MAR)方法，结合正弦图域和图像域的双域信息，采用自适应最优阈值分割方法分割CT图像中的金属，并在正弦图中去除金属污染区域，使用线性插值(LI)初步修复缺失的金属区域，采用正弦图修补网络修复受金属污染的正弦图，利用编码器-解码器网络结构恢复缺失的图像信息。网络输出的正弦图经过滤波反投影(FBP)算法生成CT重建图像。对于初步校正后存在的正弦图信息不一致性问题，使用非局部细化网络在图像域进行修复，减少二次伪影产生。模拟和真实数据实验结果表明：所提方法能有效减少金属伪影，同时保留图像细节信息，显著提高重建图像质量。

关 键 词：图像处理；深度学习；金属伪影抑制；双域信息；Pix2Pix；非局部细化网络
中图分类号：TP391.4
文献标志码：A **文章编号**：1001-5965(2026)01-0232-12

CT技术作为一种主要的无损检测手段，在医疗、安检、航空航天等领域得到广泛应用。然而，当扫描视野中存在金属等高衰减物质时，X射线中低能量的光子会被金属吸收，导致部分光子无法到达探测器，这会使射线变得更容易穿透或变得“更硬”，探测器会接收到大部分高能量的光子，即射束硬化[1]。金属伪影的主要成因是射束硬化和光子饥饿，除此之外，还存在散射、部分容积效应及噪声等其他因素。金属伪影在重建图像中常呈现为放射状或条纹状的明暗伪影[2]，这些条纹会覆盖周围物体的图像信息，导致材料识别和物体边界分割出现错误，对工业CT中的物质辨识和结构判断产生严重影响。

在过去几十年中，学者们提出了许多金属伪影抑制(metal artifact reduction, MAR)方法[3-5]。由于金属伪影产生的原因相当复杂，在CT检测中难以避免，采用物理手段进行金属伪影的校正只能在一定程度上减少伪影的出现，而不能完全消除伪影并修复图像中的其他信息。因此，现有的MAR方法主要集中在投影域和图像域对图像进行处理。

目前，传统MAR方法主要包括投影补全、迭代重建和图像后处理方法。投影补全方法将金属轨迹中的投影数据视为缺失，利用各种插值方法[6]或图像插值技术填充缺失数据。插值方法计算简单，但其边界平滑性[7]难以保证。金属痕迹经过滤波处理后，边界不连续性被放大，会在CT图像中引入新的伪影。为了充分利用角度和能量2个维度的局部信息，Duan等[8]在2009年引入图像嵌入方法进行投影补全。尽管这些方法在一定程度上缓解了不连续性，但重建图像仍然不可避免地带有额外伪影。2010年，Meyer等[9]提出了一种归一化金属伪影抑制(normalized metal artifact reduction, NMAR)

收稿日期：2023-11-20；**录用日期**：2023-12-29；**网络出版时间**：2024-03-14 13:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240313.1434.002
基金项目：国家重点研发计划(2022YFA1604000)；国家自然科学基金(52227802)；介科学与工程全国重点实验室开放课题(MESO-23-D02)
*通信作者. E-mail：minyang.ndt@buaa.edu.cn

引用格式：海潮，田鑫，张宏，等. 基于深度学习的双域信息CT金属伪影抑制方法[J]. 北京航空航天大学学报，2026，52(1)：232-243.
HAI C，TIAN X，ZHANG H，et al. A deep learning-based dual-domain information method for CT metal artifact reduction[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：232-243 (in Chinese).

方法, 以平滑金属和非金属过渡区域并抑制二次伪影。NMAR 方法通过对投影数据进行归一化, 依赖基于插值方法的多阈值分割所得的先验图像。经校正的 CT 图像可通过滤波反投影(filtered back projection, FBP)算法从完整的正弦图中获得。然而, NMAR 方法的结果依赖于先验图像的质量。另外, FBP 算法基于线积分模型, 未考虑测量数据的统计特性, 假设测量数据无噪声, 所有响应权重相同, 与实际情况不符。迭代重建[10]方法采用约束优化, 如最小二乘和最大似然法, 经多次迭代逐步提升图像质量。经典迭代重建 MAR 方法分为 2 种: ①利用金属轨迹外的投影数据, 将其视为非伪影数据[11]; ②采用统计目标函数来减弱受损投影数据[12]。迭代重建具有强大的抗噪性能, 有助于较好地消除金属伪影。然而, 迭代重建方法常常耗时较长, 需要手动调整正则化器参数, 给实际应用带来一定挑战。图像后处理方法[13]的目标在于减少图像中的金属伪影, 无需获取原始投影数据。然而, 由于 CT 图像中的噪声和伪影不遵循特定统计分布, 图像后处理方法通常难以有效抑制伪影, 且可能扭曲原始图像结构[14]。

作为一种新兴的图像处理技术, 深度学习在 MAR 领域展现出巨大的应用潜力。2019 年, Gjesteby 等[15]首次将简单的卷积神经网络(convolutional neural network, CNN)与 NMAR 方法相结合, 建立网络模型, 用于处理含有金属伪影的污染图像和不含伪影的清晰图像, 证明深度学习方法能有效去除伪影。2018 年, Zhang 等[16]提出了 CNN-MAR 框架, 将不同 MAR 方法融合以提升伪影抑制性能。为了消除原始 CT 图像中的金属伪影, 2019 年, Liao 等[17]引入了一种无监督伪影分离网络(artifact disentanglement network, ADN)。Lin 等[18]提出 DuDoNet, 利用拉东反演层连接 2 个域的子网络, 改善了正弦图不一致性问题, 并增强了双域信息。Lee 等[19]提出了基于全 CNN 的双域牙科 MAR 方法。Peng 等[20]设计了一个 MAR 网络, 实现了跨域信息交互, 促进射束硬化污染区域的信息恢复, 进而降低更多的金属伪影。尽管前述基于正弦图域或图像域的深度学习方法取得了重大进展, 但仍存在一些限制[18]。对于基于正弦图域的深度学习方法, 被金属痕迹破坏的投影是局部性的, 但在处理过程中难以保持金属痕迹边界的连续性, 容易在边界产生二次伪影。而对于基于图像域的方法, 由受损的正弦图重建的 CT 图像存在结构性伪影, 覆盖了大部分原始图像的重要细节, 难以完全消除。

针对工业 CT 中金属伪影抑制的难题, 本文提出一种新的双域信息 MAR 方法, 包含 2 个关键步骤, 分别在正弦图域和图像域进行图像修复。首先, 采用自适应最优阈值分割方法得到金属掩膜, 在正弦图域剔除金属污染区域, 并使用正弦图修补网络修复受金属影响的正弦图轨迹。然后, 进行 FBP 重建, 在图像域利用非局部细化网络对校正图像中存在的正弦图不一致伪影进行校正, 提高图像重建质量。

1　本文方法

1.1　结合双域信息的金属伪影抑制方法

图 1 为基于深度学习的 MAR 方法流程, 结合正弦图域和图像域的双域信息, 以减少 CT 图像中的金属伪影。

方法包括正弦图修补网络和图像域细化网络 2 个部分。首先, 利用线性插值(linear interpolation,

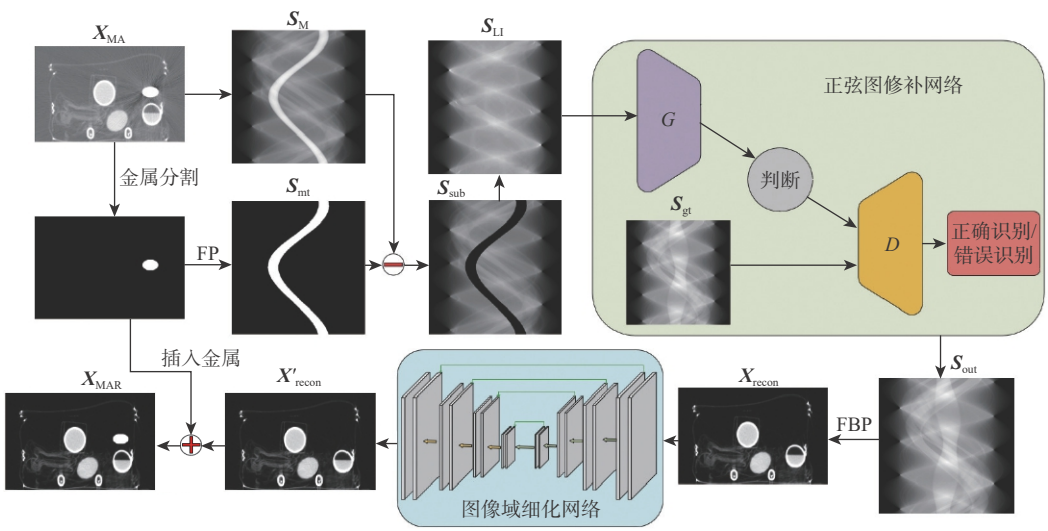


图 1　本文方法流程

Fig. 1　Flow chart of the proposed method

LI)生成金属痕迹内正弦图的初始估计；然后，利用生成对抗网络(generative adversarial network, GAN)填补插值区域，进行FBP重建；最后，利用部分卷积细化网络对重建的CT图像进行细化，从而抑制二次伪影。方法具体步骤如下：

1) 对受金属伪影影响的CT图像$\boldsymbol{X}_{\mathrm{MA}}$进行分割。为了得到CT断层图像中的金属区域，提出一种自适应最优阈值分割方法获得金属区域。从图像的灰度级特征出发，定义前景和背景的类内中心及类内像素的分散度，根据类内中心及分散度，引入类间距离判别函数，从而获得最优的分割阈值。根据该方法得到金属掩膜。

2) 使用前向投影(forward projection, FP)算法获得金属正弦图$\boldsymbol{S}_{\mathrm{mt}}$，其中，正弦图中非0区域为金属污染区域。

3) 原始图像前向投影得到金属伪影正弦图$\boldsymbol{S}_{\mathrm{M}}$，减去步骤2)中金属正弦图$\boldsymbol{S}_{\mathrm{mt}}$，得到$\boldsymbol{S}_{\mathrm{sub}}$，并对正弦图中缺失部分进行插值处理得到$\boldsymbol{S}_{\mathrm{LI}}$。

4) 采用第1个正弦图修补网络对金属污染区域进行修复。

5) 利用FBP算法对修复完成的正弦图$\boldsymbol{S}_{\mathrm{out}}$进行重建。

6) 使用第2个细化网络在图像域调整图像以移除由于步骤5)中正弦图信息不一致导致的二次伪影，得到校正图像$\boldsymbol{X}'_{\mathrm{recon}}$。

7) 将步骤1)中分割出的金属掩膜添加到重建图像中得到最终$\boldsymbol{X}_{\mathrm{MAR}}$。

1.2 自适应最优阈值分割方法

自适应最优阈值分割方法的主要思想是将一幅灰度图像分割成$m\times m$个子图像，针对每幅子图像进行最优阈值的选择及阈值分割。

对于最优阈值的选择，传统的基于类间方差最大的otsu阈值法由于只考虑了类间距而没有考虑类的内聚性，当图像噪声干扰较大时，很难得到理想结果。因此，建立了一种最优的自适应阈值选择方法，从图像的灰度级特征出发，定义前景和背景的类内中心及类内像素的分散度，根据类内中心及分散度，引入类间距离判别函数。具体的阈值选择方法如下：

将灰度图像描述为二维矩阵$\boldsymbol{I}_{M\times N}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})$，其中，$M\times N$为图像大小，$\boldsymbol{I}_{M\times N}=\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})$是像素$(\boldsymbol{x},\boldsymbol{y})$的灰度值，且$\boldsymbol{f}(\boldsymbol{x},\boldsymbol{y})\in\{0,1,\cdots,l-1\}$，$l$为图像的灰度级总数。在图像灰度级$i$出现的次数为$n_i$时，则灰度级$i$出现的概率为

$$p_i=\frac{n_i}{MN} \tag{1}$$

式中：$p_i\geqslant 0$，$\sum_{i=0}^{l-1}p_i=1$。

假设以灰度级t为阈值把全部像素分成2类：B_1(背景类)和B_2(前景类)，分别包括$i\leqslant t$和$i>t$的像素。则B_1和B_2出现的概率为

$$P_1=\sum_{i=0}^{t}p_i \tag{2}$$

$$P_2=\sum_{i=t+1}^{l-1}p_i \tag{3}$$

式中：$P_1+P_2=1$。

为了有效描述2类之间的距离D，定义B_1和B_2的类内中心分别为

$$\omega_1=\sum_{i=0}^{t}ip_i/P_1 \tag{4}$$

$$\omega_2=\sum_{i=t+1}^{l-1}ip_i/P_2 \tag{5}$$

则

$$b=|\omega_1-\omega_2| \tag{6}$$

b在一定程度上能体现B_1和B_2的分割效果，值越大，2类的类间距越大，B_1和B_2分离得越开。

另外，2类中内聚性的好坏也是直接反映分割是否有效的一个重要标志。以类中每个像素到类内中心的距离定义类的分散度：

$$d_1=\sum_{i=0}^{t}|i-\omega_1|\frac{p_i}{P_1} \tag{7}$$

$$d_2=\sum_{i=t+1}^{l-1}|i-\omega_2|\frac{p_i}{P_2} \tag{8}$$

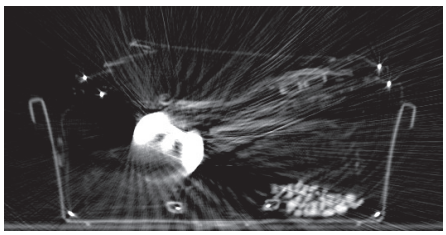
显然，每类的分散度越小，表示其内聚性越好，分类效果越好。因此，综合考虑以上2方面的因素，必须同时确保b值最大且d_1、d_2值最小。这样，每类的内聚性最好，而且2类的类间距离又最大时分类最成功。为此，定义B_1和B_2分类判别函数为

$$H(t)=\frac{P_1P_2b}{P_1d_1+P_2d_2} \tag{9}$$

因此，当H最大时将达到最好的分类效果。

若一幅图像的某一灰度级$\tilde{t}$能使$H(\tilde{t})=H_{\max}(t)$，以$\tilde{t}$为阈值分为B_1和B_2，此时，前景和背景的分类效果最佳。若设置B_1中像素的灰度值为0，B_2中像素的灰度值为1，将图像$\boldsymbol{I}_{M\times N}$二值化，就能使前景和背景最大程度地分开，达到最有效分割图像的目的。因此，$\tilde{t}$是最优的阈值，生成的二值化图像效果最好。

基于最优阈值选择方法，对图2(a)所示的原始CT断层重建图像进行阈值分割，结果如图2(b)所



(a) 原始图像

(b) 自适应最优

图 2　金属图像阈值分割结果

Fig. 2　Images threshold segmentation results of metal

示, 可以看出, 该方法准确地将金属从复杂背景中分割出来, 为之后的伪影抑制提供了良好的数据基础。

1.3　正弦图修补网络

为了完成正弦图的初步修复, 采用生成对抗网络 Pix2Pix[21], 该网络基于 GAN 的图像修复模型, 其优势在于实现端到端的图像生成, 产生细节逼真的修复图像, 适用于多种图像恢复任务。图 3 展示了正弦图修补网络的结构。网络采用插值得到的正弦图 $\boldsymbol{S}_{\mathrm{L1}}$ 和真值目标图像 $\boldsymbol{S}_{\mathrm{gt}}$ 作为网络的输入, 旨在通过网络训练获得初步修补的正弦图 $\boldsymbol{S}_{\mathrm{out}}$。生成器采用全卷积网络结构, 包含 8 个卷积层和 8 个反卷积层, 每层卷积核大小为 4×4, 步长为 2。在下采样过程中, 采用批归一化(batch normalization, BN)代替最大池化操作, 上采样时每层输出后采用 Dropout 随机减少部分输出, 以避免网络过拟合。为避免方差偏移, 在 BN 层之后采用 Dropout[22]。此外, 每个卷积层 k 和其对称的反卷积层 $K-k$ 之间具有跳跃连接, 其中, K 为总层数。判别器同样采用全卷积方式, 包含 5 个卷积模块, 前 4 个模块采用 LReLU 激活函数, 最后使用步长为 1 的卷积层。判别器的训练基于完整的正弦图而不是分块, 以更好地保留正弦图的信息。针对生成器网络输出的图像或真值图像数据, 判别器进行真假判断。

采用的网络整体损失函数可以表示为

$$L = \arg\min_{G}\max_{D} L_{\mathrm{cGAN}}(G,D) + \lambda L_{L_1}(G) \tag{10}$$

式中: $L_{\mathrm{cGAN}}(G,D)$为对抗损失函数; λ 为用于控制表达式中 2 项之间平衡的超参数; $L_{L_1}(G)$为 L_1 正则化损失。

训练过程中生成器 G 和判别器 D 相互对抗, 即 G 试图生成可以欺骗 D 的完美正弦图, 而 D 试图检测假图像。上述对抗损失函数可以表示为

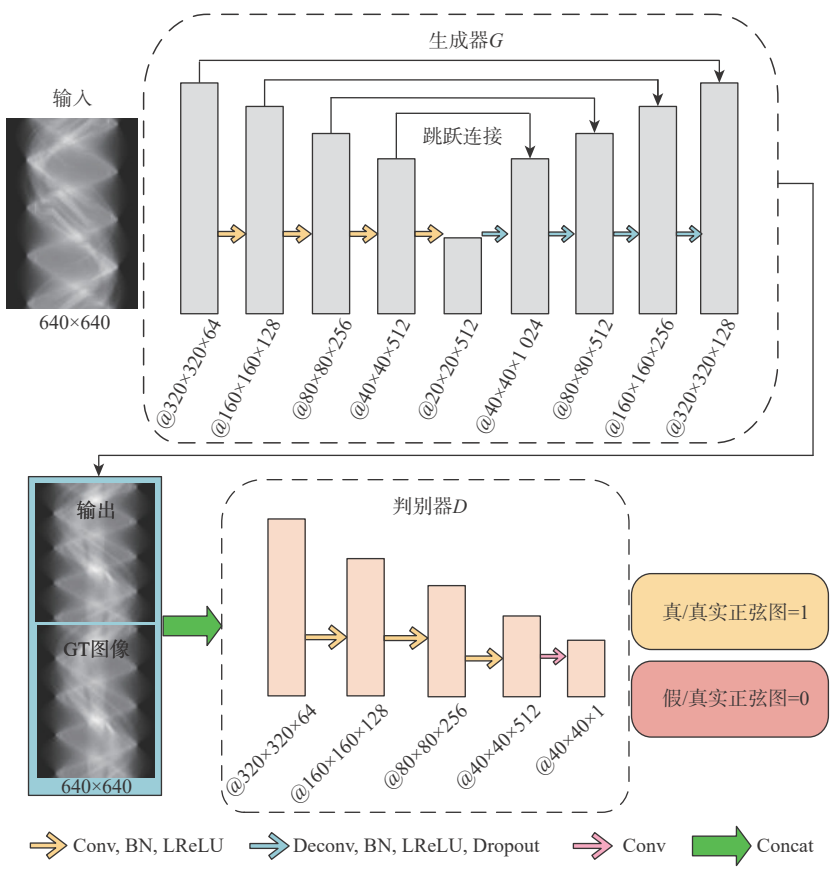


图 3　正弦图修补网络的结构

Fig. 3　The structure of sino-inpainting network

$$L_S(G,D)=E_{x,y}(\ln(D(\boldsymbol{x},\boldsymbol{y})))+E_x[\ln(1-D(\boldsymbol{x},G(\boldsymbol{x})))] \tag{11}$$

式中：$E_{x,y}$为通过训练输入数据$\boldsymbol{x}$和相应的配对真值数据$\boldsymbol{y}$得到的近似数据概率密度的期望；$D(\boldsymbol{x},\boldsymbol{y})$为判别器输出真实样本的概率；$G(\boldsymbol{x})$为生成器生成的完整正弦图数据的估计；$E_x$为不完整数据概率密度的期望。

1.4 基于非局部模块的图像域细化网络

为了减少修补投影数据时产生的二次伪影，并进一步提高图像细节质量，采用以U-Net为主干的图像域网络来细化重建后的CT图像。传统的U-Net模型采用全CNN，而卷积是一种局部操作，其感受野被卷积核的大小所限制。一旦网络的深度不足9，卷积就很难捕捉到图像中相距较远的特征。与此同时，金属伪影在图像中通常表现为非局部特征，往往能够辐射到整幅图像。因此，在图像域网络中引入非局部细化网络[23]，以加强网络对图像全局特征的提取能力。

非局部细化网络源自非局部均值去噪方法[24-25]，与使用相似像素进行加权求和的非局部方法不同，非局部细化网络在整幅图像捕获特征图。通用非局部操作定义为

$$\boldsymbol{y}_i=\frac{1}{c(\boldsymbol{x})}\sum_{j\in s}\boldsymbol{f}(\boldsymbol{x}_i,\boldsymbol{x}_j)\boldsymbol{g}(\boldsymbol{x}_j) \tag{12}$$

式中：$\boldsymbol{x}$为输入图像；$\boldsymbol{y}$为与$\boldsymbol{x}$大小相同的输出图像；$\boldsymbol{x}_i$为需要替换的第i个元素；s为搜索窗口的范围。对于图像中的每个像素$\boldsymbol{x}_i$，都将遍历窗口中的每个像素$\boldsymbol{x}_j$。一元函数$\boldsymbol{g}(\boldsymbol{x})$负责计算$\boldsymbol{x}_j$处输入信号的数值表达，为了简单起见，此处选择线性嵌入的形式：$\boldsymbol{g}(\boldsymbol{x}_j)=\boldsymbol{W}_g*\boldsymbol{x}_j$，其中，$\boldsymbol{W}_g$为可学习的加权矩阵，使用1×1的卷积。成对函数$\boldsymbol{f}$将计算$\boldsymbol{x}_i$与$\boldsymbol{x}_j$之间的相似性，计算公式如下：

$$\boldsymbol{f}(\boldsymbol{x}_i,\boldsymbol{x}_j)=\exp((\boldsymbol{\theta}_1(\boldsymbol{x}_i))^{\mathrm{T}})\exp(\boldsymbol{\theta}_2(\boldsymbol{x}_j)) \tag{13}$$

$\boldsymbol{x}_i$处的响应通过因子$c(\boldsymbol{x})$进行归一化处理，其定义为

$$c(\boldsymbol{x})=\sum_{j\in s}\boldsymbol{f}(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{14}$$

在成对函数f的计算公式中，$\boldsymbol{\theta}_1(\boldsymbol{x}_i)$和$\boldsymbol{\theta}_2(\boldsymbol{x}_j)$为特征图的2个嵌入参数，$\boldsymbol{W}_1$和$\boldsymbol{W}_2$则为2个参数对应的可学习权重矩阵，将随着网络的训练不断调整改进。

非局部细化网络的计算流程和整体结构如图4所示。其中，$\boldsymbol{X}$和$\boldsymbol{Y}$分别为输入和输出特征图，T、C、V和W分别为批处理大小、通道数及图像的高和宽。softmax是归一化指数函数，能够对特征图进行分类操作。为了将非局部模块插入神经网络，采用剩余连接的策略：

$$\boldsymbol{z}_i=\boldsymbol{W}_z\boldsymbol{y}_i+\boldsymbol{x}_i \tag{15}$$

式中：$\boldsymbol{z}_i$为非局部模块输出结果；$\boldsymbol{W}_z$为权重；$\boldsymbol{y}_i$由式(12)计算得到。由于非局部模块需要的计算量较大，因此，若将其放在网络的第1层将会对计算速度造成较大的影响；然而，随着网络层数的深入，特征映射的空间将逐步变小，非局部模块所能捕获的特征不够精确。根据文献[16]的实验结果，将非局部模块放置在网络的第2层和第3层中将取得较好的计算效果。因此，加入非局部模块后的图像域网络结构如图5所示。

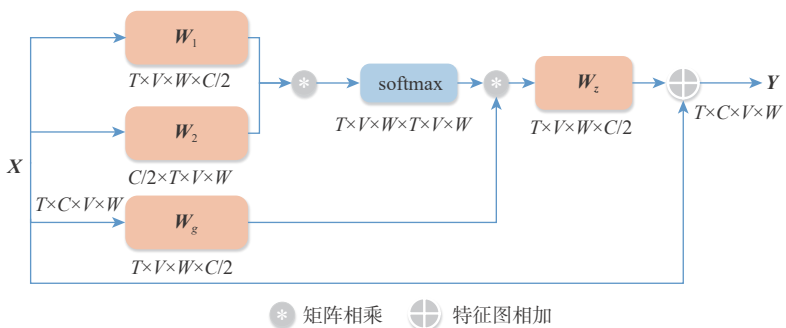


图4 非局部细化网络计算结构示意图

Fig. 4 Schematic diagram of non-local refine network computing structure

该网络的输入为正弦图修补网络修补并重建后的图像$\boldsymbol{X}_{\mathrm{recon}}$，参考图像$\boldsymbol{X}_{\mathrm{gt}}$作为训练的标签。网络的主体部分由编码网络和解码网络2部分组成。编码网络包含5个卷积层，卷积层由卷积、BN和Leaky ReLU激活函数组成，并在第2层和第3层网络中加入了非局部模块；解码网络由5层反卷积层组成，每个反卷积层均由上采样、反卷积、BN和ReLU激活函数组成。编码网络和解码网络直接通过跳跃连接传递信息，跳跃连接是残差学习策略的实现形式，能够有效防止梯度扩散和图像信息损失问题，保证特征学习的完整性和正确性。

图像域网络同样使用L_{im}损失函数进行优化，

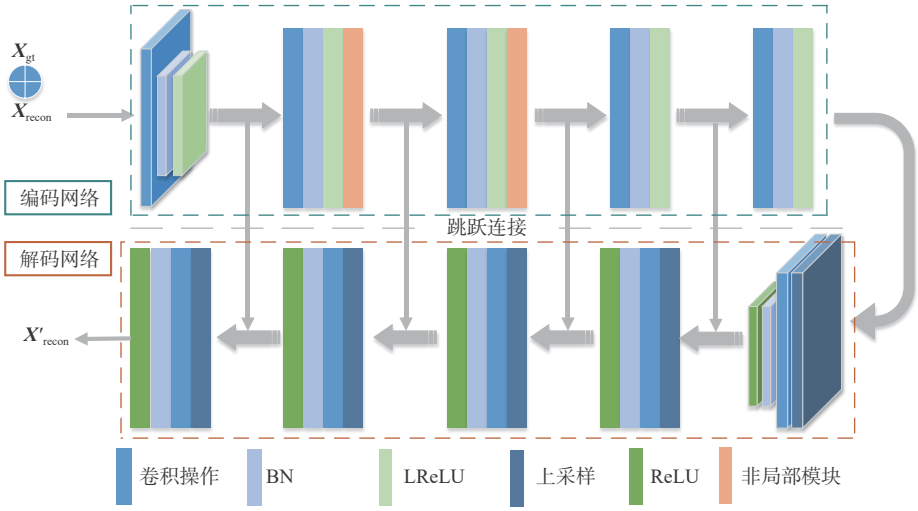


图 5　非局部细化网络结构

Fig. 5　The structure of non-local refine network

其表达式为

$$L_{im} = \left\| \boldsymbol{X}'_{recon} - \boldsymbol{X}_{gt} \right\|_1 \tag{16}$$

正弦图网络与图像域网络之间由 FBP 算法连接, 其损失记为 L_{FBP}。因此, MAR 方法的总目标函数可以表示为

$$L = L_S + L_{FBP} + L_{im} \tag{17}$$

2　实验验证

2.1　数据集获取及实验配置

2.1.1　数据集构建

由于真实数据采集过程很难得到相匹配的金属伪影图像和无金属伪影的参考图像, 因此, 本文采用与文献 [26] 中类似的程序来模拟合成金属污染的正弦图和 CT 图像。多色 X 射线源的光谱如图 6 所示。

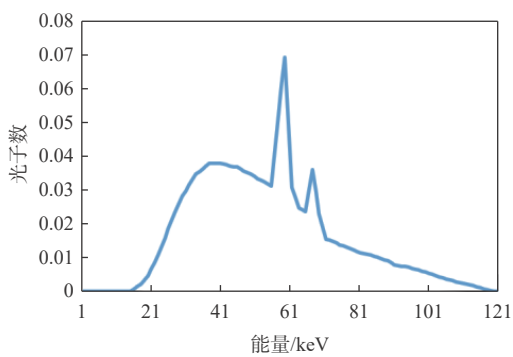


图 6　120 kV 电压下的 X 射线源能谱

Fig. 6　X-ray source energy spectrum at 120 kV voltage

模拟实验基于等角扇束重建系统, 用来获取不包含金属和插入金属后的配对数据。X 射线源峰值电压为 120 kV, 源与旋转中心之间距离为 1 075 mm。重建图像像素尺寸为 512 像素×512 像素, 720 个均匀采样的投影角, 投影角 0.05°, 1 024 个探测器通道生成投影正弦图。选取 3 个样品共 100 张图像, 原始数据中不包含金属; 同时, 选取 90 个不同的金属形状, 每张图像中对应插入一种金属, 共获得 9 000 张金属伪影模拟数据。金属掩膜图像进行了二值化处理, 使用 FP 算法分别对金属插入图像和金属掩膜进行正弦图投影, 相减得到去除金属轨迹的正弦图。采用文献 [27] 中的方法对缺失的正弦图进行插值, 得到的结果与基准图像生成的正弦图作为网络训练时的配对数据, 从中选取 1/10 的数据作为测试集。

采用 MATLAB R2018b 平台来进行金属伪影数据模拟的全过程, 金属掩膜图像由“The 2016 Low-dose CT Grand Challenge”数据集提供, 包含铁、铜、钛等多种金属。模拟生成的部分图像如图 7 所示。

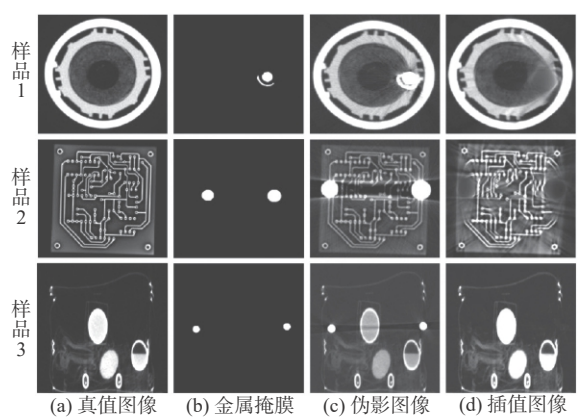


图 7　部分数据集样本图像

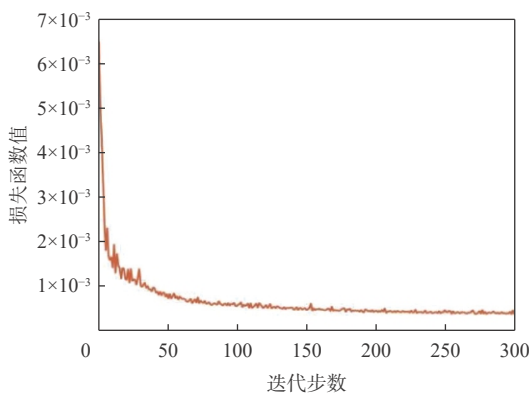
Fig. 7　Some dataset sample images

2.1.2　网络训练

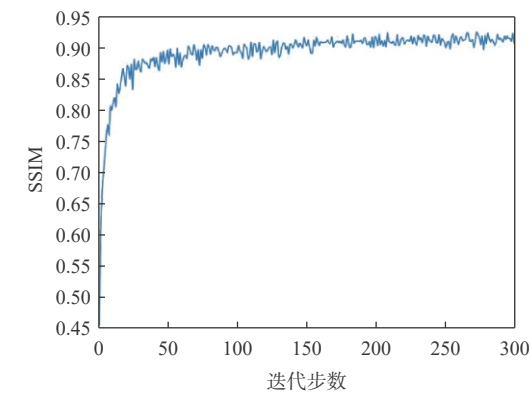
实验配置为: CPU 为 Intel Xeon W-3300, 内存为 32 GB, 显卡为 2 张 NVIDIA Quadro RTX 4000, 操作系统为 Win10, CUDA 版本 10.2, cudnn7.6。基于 Pytorch 深度学习平台完成网络的搭建、训练和验证, 所用编程语言为 Python。

在正弦图修复阶段，Pix2Pix网络训练遵循一般GAN的训练策略，分别在生成器G和判别器D上轮流执行梯度下降步骤，用k表示生成器的梯度下降次数。在最初的4个epoch，生成器每执行1次梯度下降，判别器将执行$6-k$次梯度下降。训练采用Adam优化器，其中，动量参数$\beta_1=0.9$，$\beta_2=0.999$，$\lambda=10$为衡量L_1损失的权重，根据经验设定。设置学习率为0.002，在网络每轮的迭代中，每隔200步打印输出损失函数值，直到D输出概率为0.5时，表明无法区分来自G输出图像的真伪，将该轮迭代保存的模型用作最终训练的模型。

在图像域的卷积细化网络训练阶段，优化算法为Adam，批处理大小为8，训练迭代次数300次，初始学习率设置为0.000 2，每迭代100次，学习率乘以0.5。图8展示了图像域细化网络在训练过程中的损失函数收敛及图像质量提升情况。可以看到，网络的梯度下降较明显，且随着训练的不断深入，输出图像与参考图像之间的结构相似度系数(structural similarity index measure, SSIM)也稳步提升，证明模型的训练过程有效。



(a) 损失函数收敛曲线



(b) 结构相似度系数变化曲线

图8 网络训练损失函数值和SSIM随迭代步数的变化

Fig. 8 Variation of network training loss value and SSIM with number of iterations

2.2 模拟实验结果与分析

为了充分验证本文方法的有效性，分别采用模拟仿真数据和真实数据进行实验，并选取传统LI、射束硬化校正(beam hardening correction, BHC)、NMAR和CNN-MAR这4种MAR方法作处理效果对比。

在图像评价方面，采用客观评价指标对图像质量进行分析。选取SSIM和峰值信噪比(peak signal noise ratio, PSNR)2个评价指标来衡量CT图像中金属伪影抑制质量。PSNR主要衡量图像的噪声去除效果，网络生成图像的PSNR越大，表示图像中的噪声越少，图像质量越好；反之，则图像噪声多，图像质量差。SSIM的作用是衡量参考图像和恢复图像之间的相似程度，SSIM的值越大，表明网络处理后的图像越接近真值图像。

在模拟实验中，选取3类样品及对应的3种金属植入物用作实验数据集，选取不包含金属伪影的CT断层图像作为基准，模拟图像中的金属伪影与训练集中的生成方法相同，但并未出现在训练集中。图9展示了现有的多种传统MAR方法与本文方法的对比结果。同时，本节还将探讨本文方法对不同尺寸金属的伪影抑制效果，图9中案例1～案例3分别对不同断层图像插入了不同形状的金属，金属尺寸均较小。为了能更清晰地观察图像细节，对图中的感兴趣区域(region of interest, ROI)进行放大。

对于图9中的3类图像，从主观视觉方面进行评价，与真值图像相比，金属模拟图像中包含严重的金属伪影，对图像信息的判断产生严重的干扰。从放大图像上可以看到，金属周围存在明显的条纹伪影和暗带。LI方法有效减少了金属伪影，并对金属区域进行了补充，但部分严重的结构化伪影抑制效果不佳，同时也造成了严重的二次伪影，主要沿金属边缘的切线分布，整体处理效果较差；BHC方法对于紧贴金属部位的伪影有较好的清除效果，能够使金属的原始形状更加清晰，但对金属相距较远的伪影抑制效果一般；NMAR方法在处理过程中会剔除金属部分，伪影抑制效果相比BHC方法更好，但保留了很多残留伪影；CNN-MAR方法能较好地去除图像中大部分的金属伪影，但图像纹理细节信息有所损失；本文方法对金属周围的结构恢复得更加清晰，对伪影的校正更为彻底，图像边缘保持完整，并且由于通过自适应阈值分割方法提取了金属区域，金属本身的形状也完整保留。

图10展示了含有较大尺寸金属物体的CT断层图像及各方法处理后的结果对比。从ROI放大图可以看出，大尺寸金属造成的放射状伪影和暗带更为严重，对周围的物体信息有较强的遮挡效果，并导致了部分图像结构的改变。LI方法对金属伪

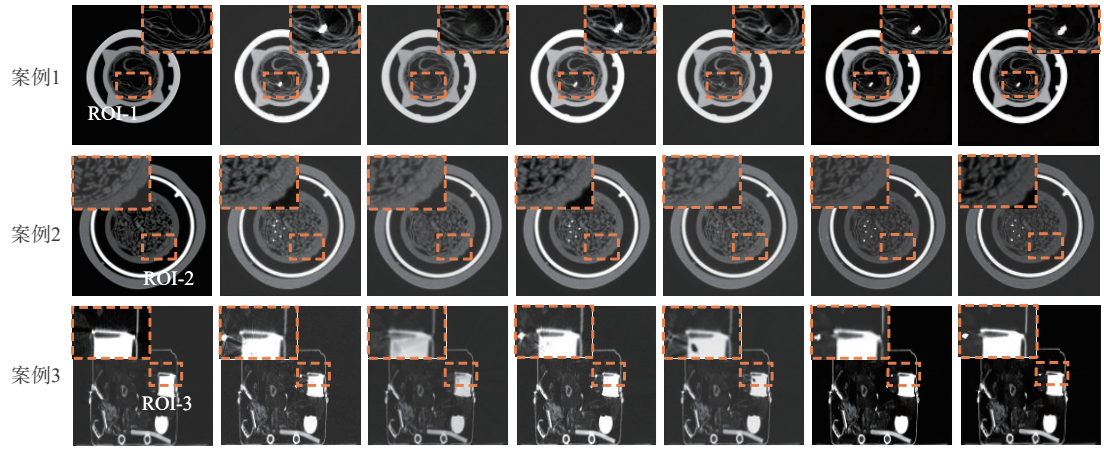


图 9　小尺寸金属植入图像的伪影抑制结果对比

Fig. 9　Comparison of artifact reduction results for small-sized metal implant images

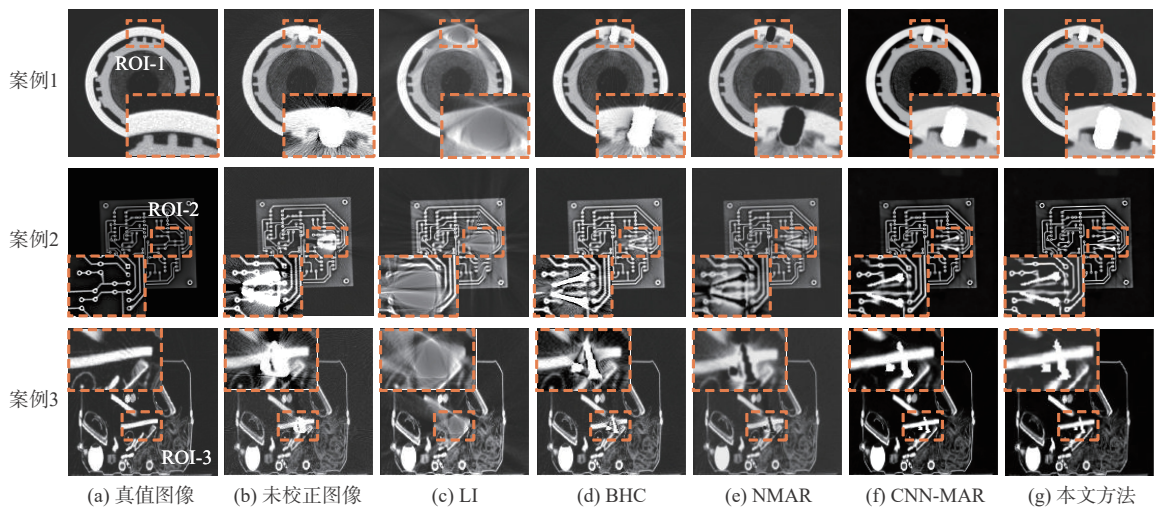


图 10　大尺寸金属植入图像的不同伪影抑制结果对比

Fig. 10　Comparison of artifact reduction results for large-sized metal implant images

影的去除效果微乎其微, 反而会导致强烈的二次伪影; BHC 和 NMAR 方法均能去除部分强度较大的伪影, 但对细小的条状伪影抑制效果较差; CNN-MAR 方法金属伪影抑制效果明显, 但损失了图像细节信息; 通过对比, 本文方法金属伪影抑制效果更为显著, 去除了图像中大部分的条状伪影, 同时图像细节信息也得到了较好的保留, 处理后的图像更加接近参考图像。

使用定量指标对上述方法进行评价。表 1 展示了小尺寸金属 CT 图像的评价指标结果。可以看出, LI 方法由于伪影去除较少导致 PSNR 较高、SSIM 较低; BHC 方法和 NMAR 方法的处理效果则有明显提升, 且 NMAR 方法得到的图像质量略好于 BHC 方法。同时, 案例 3 中的物体最多, 所包含的信息也最多, 伪影对图像质量的影响也更大, 因此, 该案例的数据指标要略低于前 2 个案例。本文方法在不同的情况下, PSNR 和 SSIM 都取得了最优的效果。

表 2 展示了含有大尺寸金属 CT 图像的评价指标数据, 各方法的对比结果与上述小尺寸金属 CT

表 1　小尺寸金属 CT 图像伪影抑制评价指标结果

Table 1　Evaluation index results for artifact reduction in CT images (small-sized metal)

案例	PSNR/dB						SSIM					
	未校正图像	LI	BHC	NMAR	CNN-MAR	本文方法	未校正图像	LI	BHC	NMAR	CNN-MAR	本文方法
1	19.162 8	20.872 5	25.421 0	28.113 2	29.612 3	**30.097 9**	0.818 4	0.910 1	0.959 4	0.960 2	0.975 6	**0.981 5**
2	19.184 6	20.757 3	25.401 4	27.889 2	28.542 2	**30.394 1**	0.812 6	0.915 2	0.956 7	0.961 9	0.976 7	**0.981 2**
3	19.113 2	20.132 4	25.382 3	27.413 7	28.453 1	**29.524 3**	0.801 7	0.900 6	0.948 2	0.958 8	0.964 5	**0.970 8**

注: 加粗数值表示最优结果。

表 2 大尺寸金属 CT 图像伪影抑制评价指标结果

Table 2 Evaluation index results for artifact reduction in CT images (large-sized metal)

案例	PSNR/dB						SSIM					
	未校正图像	LI	BHC	NMAR	CNN-MAR	本文方法	未校正图像	LI	BHC	NMAR	CNN-MAR	本文方法
1	16.421 7	18.372 5	21.421 0	22.721 6	23.541 2	**24.890 9**	0.818 4	0.893 2	0.943 1	0.949 7	0.955 6	**0.976 7**
2	17.327 8	18.157 3	21.473 5	22.790 1	22.984 3	**24.130 9**	0.809 8	0.902 8	0.945 2	0.952 2	0.954 8	**0.968 9**
3	17.043 4	17.918 7	20.473 5	21.790 1	22.005 9	**23.524 3**	0.804 3	0.891 5	0.941 0	0.942 0	0.952 4	**0.960 5**

注: 加粗数值表示最优结果。

图像中的数据情况类似。值得注意的是, 各方法的 PSNR 和 SSIM 相比于小尺寸金属 CT 图像均有所降低, 原因是在图像域中大尺寸金属导致的部分严重的结构化伪影并不能完全去除, 且金属本身占据了一定数量的像素。本文提出的双域信息的校正方法在不同的情况下, PSNR 和 SSIM 都取得了最好的表现。

综上所述, 本文方法在测试结果中 SSIM 值和 PSNR 值最高, 表明图像经过本文方法修复后, 噪声和伪影得到了明显的抑制, 且校正后的图像与参考图像之间的差值小, 图像还原度较高, 图像中被金属伪影损坏的物体结构细节恢复良好, 验证了本文方法校正性能的稳定性和伪影抑制的有效性。

除了定量指标评估外, 还计算了本文方法的处理时间, 如表 3 所示。选择 90 幅测试图像进行验证, 重建图像大小为 512 像素×512 像素, 其中, 本文方法重建图像的平均处理时间为 0.262 s, 虽然 LI 方法用时更少, 但其重建效果最差。

表 3 不同方法所用时间对比

Table 3 Comparison of time consumption of different methods

方法	所用时间/s
LI	0.186
BHC	0.314
NMAR	0.473
CNN-MAR	1.569
本文方法	0.262

2.3 实际实验结果与分析

为了进一步验证本文方法在实际 CT 图像伪影抑制中的有效性, 选取实际采集的含有金属伪影的 CT 图像作为测试数据。对于图像中的金属物体, 采用阈值分割的方法将其提取出来, 并获得金属掩膜正弦图, 由于真实工件无法获取完全与之对应的无伪影图像, 因此, 本节未计算测试图像的客观评价指标。

图 11 展示了各方法的处理效果, 并对 ROI 进行了放大处理。其中, 案例 1～案例 3 为重建后的原始断层图像, 可以观察到图像中含有明显的放射条纹状金属伪影, 严重降低了图像的清晰度, 并遮挡了原有图像信息; 经过 LI 处理后, 条纹状伪影有所减弱, 但从案例 2 中可以看出, LI 对金属物体周围造成了较大破坏, 并产生了严重的二次伪影; BHC 方法在真实案例中对伪影的校正表现一般, 仅能小幅度降低图像的模糊程度; NMAR 方法则可以在一定程度上削弱金属伪影, 图像中亮暗相间的条纹有所减少, 但仍然存在较多的残留伪影; CNN-MAR 方法金属伪影抑制效果较好, 但图像细节有所损失; 本文方法能够去除图像中的绝大部分伪影, 使

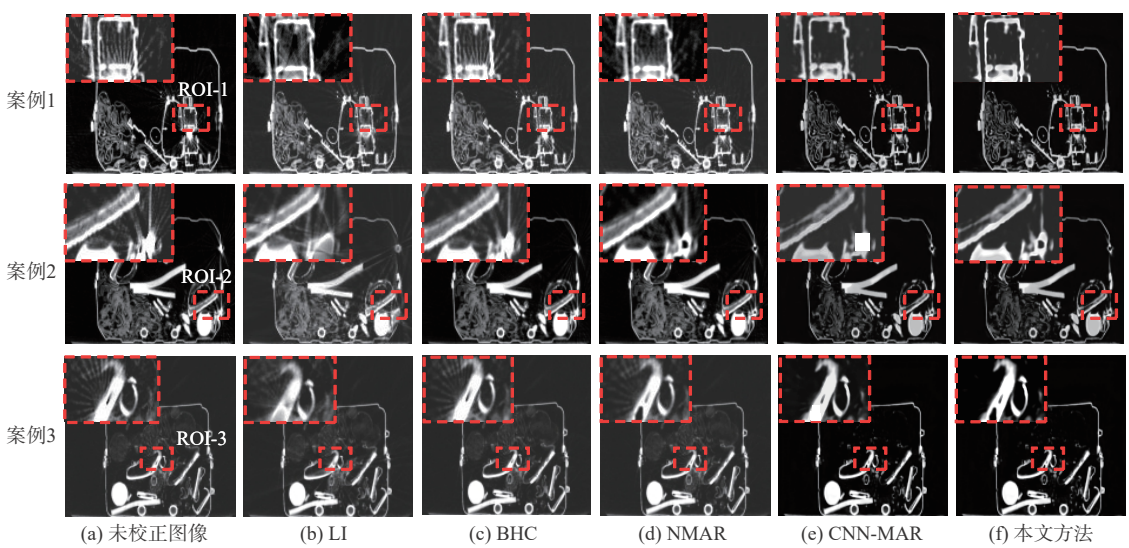
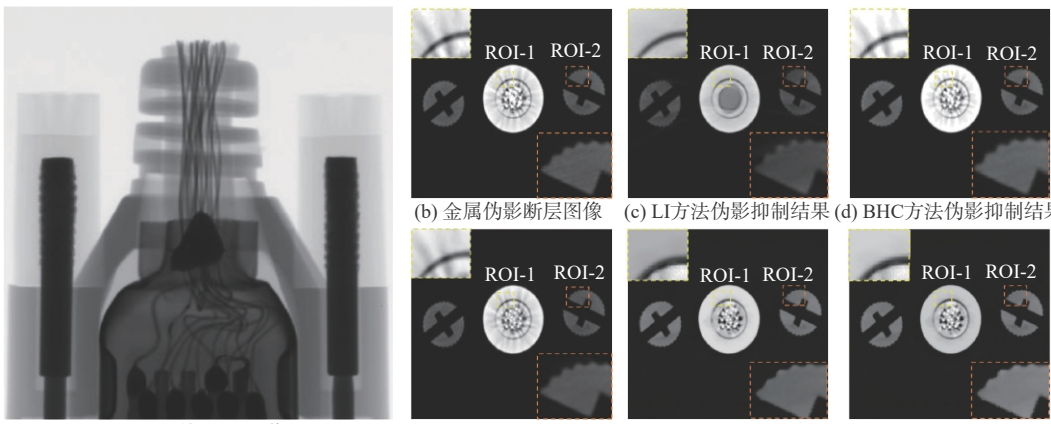


图 11 真实包裹 CT 图像的不同方法效果对比

Fig. 11 Comparison of different methods for real package CT images

图像清晰度得到了明显提升, 对于图像中的细节部分也能较好地恢复和保留。

为了进一步验证本文方法的泛化能力, 选取与训练集图像信息差距较大、伪影严重的断层图像进行 MAR 测试, 结果如图 12 所示。可以看出, 原始图像的中间区域分布有密集的铜质线缆, 导致周围的绝缘层、支撑层及两侧的螺母都出现了大量放射状伪影。LI 方法的金属伪影抑制效果良好, 图像质量得到了有效改善, 绝缘层和支撑层灰度均匀, 仅有少量的二次伪影, 但中心含有金属的部分在插值处理后周围信息被填充, 几乎无法分辨原有形状; BHC 方法和 NMAR 方法对金属伪影的校正效果较弱, 两侧的螺母灰度得到了一定的平滑; CNN-MAR 方法金属伪影抑制效果显著, 但图像对比度略有降低; 本文方法能够较为完整地保留中心区域的非金属特征, 并对全局的金属伪影有较强的校正效果。综上所述, 本文采用的 MAR 方法经过充分训练后具有良好的校正效果, 且有较强的可迁移性。同时, 图像细节信息也得到了较好的保留, 处理后的图像更加接近参考图像。



(a) VGA接口DR图像　(b) 金属伪影断层图像　(c) LI方法伪影抑制结果　(d) BHC方法伪影抑制结果

(e) NMAR方法伪影抑制结果　(f) CNN-MAR方法伪影抑制结果　(g) 本文方法伪影抑制结果

图 12　VGA 接口 CT 断层图像伪影抑制效果对比

Fig. 12　Comparison of artifact reduction effects on VGA interface CT images

3　结　论

为了减轻金属伪影对工业 CT 检测中物质辨识及结构评估的影响, 本文提出一种结合正弦图域和图像域双域信息的深度学习 CT 金属伪影抑制方法。

1) 结合正弦图域与图像域的信息互补, 构建了从数据修复到图像重建的完整抑制流程, 在抑制伪影的同时有效保留图像结构与细节信息。

2) 采用自适应最优阈值分割方法能够可靠定位金属区域, 配合所设计的正弦图修补网络, 显著提升了受损正弦图的修复质量, 为后续高质量重建奠定基础。

3) 通过非局部细化网络对初步重建图像进行二次优化, 有效减轻因数据不一致导致的二次伪影, 进一步提升了图像的视觉质量与结构一致性。模拟数据和真实数据实验结果均表明, 该方法能显著降低金属伪影干扰, 提升重建图像的视觉质量和细节保留能力, 有助于改善 CT 图像中的物质辨识和结构评估准确性。

参考文献 (References)

[1] 李俊江, 胡少兴, 李保磊, 等. CT 图像环状伪影校正方法[J]. 北京航空航天大学学报, 2007, 33(11): 1378-1382.
LI J J, HU S X, LI B L, et al. Ring artifact correction for industrial CT images[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(11): 1378-1382(in Chinese).

[2] 谷建伟, 张丽, 陈志强, 等. 工业 CT 图像的伪影成因和校正方法综述[J]. CT 理论与应用研究, 2005, 14(3): 24-28.
GU J W, ZHANG L, CHEN Z Q, et al. The review of reasons and correction methods for artifacts in ICT images[J]. Computerized Tomography Theory and Applications, 2005, 14(3): 24-28(in Chinese).

[3] 马燕, 余海军, 钟发生, 等. 基于残差编解码网络的 CT 图像金属伪影校正[J]. 仪器仪表学报, 2020, 41(8): 160-169.
MA Y, YU H J, ZHONG F S, et al. CT metal artifact reduction based on the residual encoder-decoder network[J]. Chinese Journal of Scientific Instrument, 2020, 41(8): 160-169(in Chinese).

[4] 马燕, 钟发生, 刘丰林. 基于条件生成式对抗网络的 CT 金属伪影校正研究[J]. 中国体视学与图像分析, 2021, 26(2): 101-112.
MA Y, ZHONG F S, LIU F L. CT Metal artifact reduction based on conditional generative adversarial network[J]. Chinese Journal of Stereology and Image Analysis, 2021, 26(2): 101-112(in Chinese).

[5] 姜世博, 孙跃文, 许硕, 等. 基于生成对抗网络的 CT 图像金属伪影校正[J]. 哈尔滨工程大学学报, 2022, 43(12): 1766-1771.
JIANG S B, SUN Y W, XU S, et al. Metal artifact correction of industrial CT images based on generative adversarial networks[J]. Journal of Harbin Engineering University, 2022, 43(12): 1766-1771 (in Chinese).

[6] MEHRANIAN A, AY M R, RAHMIM A, et al. X-ray CT metal artifact reduction using wavelet domain sparse regularization[J]. IEEE Transactions on Medical Imaging, 2013, 32(9): 1707-1722.

[7] YOUNG J, BEOM R. Metal artifact reduction based on sinogram correction in CT[C]//Proceedings of the 2009 IEEE Nuclear Science Symposium Conference Record. Piscataway: IEEE Press, 2010: 3480-3483.

[8] DUAN X H, ZHANG L, XIAO Y S, et al. Metal artifact reduction in CT images by sinogram TV inpainting[C]//Proceedings of the 2008 IEEE Nuclear Science Symposium Conference Record. Piscataway: IEEE Press, 2009: 4175-4177.

[9] MEYER E, BERGNER F, RAUPACH R, et al. Normalized metal artifact reduction in computed tomography[C]//Proceedings of the 2009 IEEE Nuclear Science Symposium Conference Record. Piscataway: IEEE Press, 2010: 3251-3255.

[10] 武丽君, 孙丰荣, 杨江飞, 等. 极稀疏投影数据的CT图像重建[J]. 北京航空航天大学学报, 2020, 46(12): 2366-2373.
WU L J, SUN F R, YANG J F, et al. CT image reconstruction from ultra-sparse projection data[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(12): 2366-2373(in Chinese).

[11] WANG G, SNYDER D L, O'SULLIVAN J A, et al. Iterative deblurring for CT metal artifact reduction[J]. IEEE Transactions on Medical Imaging, 1996, 15(5): 657-664.

[12] VAN SLAMBROUCK K, NUYTS J. Metal artifact reduction in computed tomography using local models in an image block-iterative scheme[J]. Medical Physics, 2012, 39(11): 7080-7093.

[13] BALLHAUSEN H, REINER M, GANSWINDT U, et al. Post-processing sets of tilted CT volumes as a method for metal artifact reduction[J]. Radiation Oncology, 2014, 9: 114.

[14] MOUTON A, MEGHERBI N, VAN SLAMBROUCK K, et al. An experimental survey of metal artefact reduction in computed tomography[J]. Journal of X-Ray Science and Technology: Clinical Applications of Diagnosis and Therapeutics, 2013, 21(2): 193-226.

[15] GJESTEBY L, SHAN H M, YANG Q S, et al. A dual-stream deep convolutional network for reducing metal streak artifacts in CT images[J]. Physics in Medicine & Biology, 2019, 64(23): 235003.

[16] ZHANG Y B, YU H Y. Convolutional neural network based metal artifact reduction in X-ray computed tomography[J]. IEEE Transactions on Medical Imaging, 2018, 37(6): 1370-1381.

[17] LIAO H F, LIN W A, YUAN J B, et al. Artifact disentanglement network for unsupervised metal artifact reduction[C]//Proceedings of the Medical Image Computing and Computer Assisted Intervention. Berlin: Springer, 2019: 203-211.

[18] LIN W A, LIAO H F, PENG C, et al. DuDoNet: dual domain network for CT metal artifact reduction[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 10504-10513.

[19] LEE D, PARK C, LIM Y, et al. A metal artifact reduction method using a fully convolutional network in the sinogram and image domains for dental computed tomography[J]. Journal of Digital Imaging, 2020, 33(2): 538-546.

[20] PENG C T, LI B, LI M, et al. An irregular metal trace inpainting network for X-ray CT metal artifact reduction[J]. Medical Physics, 2020, 47(9): 4087-4100.

[21] 江民民, 李大朋, 邱昕, 等. 基于CGAN的避扰通信决策网络离线式训练方法[J]. 北京航空航天大学学报, 2020, 46(7): 1412-1421.
JIANG M M, LI D P, QIU X, et al. An offline training method using CGAN for anti-jamming communication decision network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(7): 1412-1421(in Chinese).

[22] LI X, CHEN S, HU X L, et al. Understanding the disharmony between dropout and batch normalization by variance shift[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 2677-2685.

[23] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7794-7803.

[24] 龙超, 金恒, 黎玲, 等. 基于特征融合的非局部均值CT图像降噪[J]. 光学学报, 2022, 42(11): 281-291.
LONG C, JIN H, LI L, et al. CT image denoising with non-local means based on feature fusion[J]. Acta Optica Sinica, 2022, 42(11): 281-291(in Chinese).

[25] BUADES A, COLL B, MOREL J M. A non-local algorithm for image denoising[C]//Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2005: 60-65.

[26] KALENDER W A, HEBEL R, EBERSBERGER J. Reduction of CT artifacts caused by metallic implants[J]. Radiology, 1987, 164(2): 576-577.

[27] XU S Y, DANG H. Deep residual learning enabled metal artifact reduction in CT[C]//Proceedings of the Medical Imaging 2018: Physics of Medical Imaging. Bellinham: SPIE, 2018: 132.

A deep learning-based dual-domain information method for CT metal artifact reduction

HAI Chao[1], TIAN Xin[1], ZHANG Hong[2], TAN Dalong[1], HE Yixin[1], MENG Fanyong[3], YANG Min[1,*]

(1. School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China;
2. Beijing Power Machinery Research Institute, Beijing 100074, China;
3. State Key Laboratory of Mesoscience and Engineering, Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: When metal is present in the field of view of a CT scan, the reconstruction of images inevitably produces metal artifacts, significantly impacting image quality. In order to suppress metal artifacts, we propose a new deep learning CT metal artifact reduction (MAR) method that combines dual domain information from both the sinogram and image domains. Firstly, the adaptive optimal threshold segmentation method is used to segment the metal in the CT image and remove the metal corrosion area in the sinogram. Linear interpolation (LI) is used to preliminarily repair the missing metal area. After the metal-contaminated sinogram domain has been repaired using the sino-inpainting network, further picture information is recovered by employing an encoder-decoder network structure. The sinogram domain output from the network undergoes filtered back projection (FBP) to generate CT reconstructed images. To address inconsistencies in the initially corrected sinogram domain information, a non-local refine network is utilized in the image domain to reduce secondary artifact generation. This technique successfully lowers metal artifacts while maintaining image details, greatly improving the quality of the reconstructed images, according to experimental results using both simulated and real data.

Keywords: image processing; deep learning; metal artifact reduction; dual-domain information; Pix2Pix; non-local refine network

Received: 2023-11-20; **Accepted**: 2023-12-29; **Published Online**: 2024-03-14 13:33

URL: link.cnki.net/urlid/11.2625.V.20240313.1434.002

Foundation items: National Key Research and Development Program of China (2022YFA1604000); National Natural Science Foundation of China (52227802); Open Research Fund of State Key Laboratory of Mesoscience and Engineering (MESO-23-D02)

* **Corresponding author**. E-mail: minyang.ndt@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0764

基于旋转单天线的欺骗干扰检测方法

陈世淼[1]，倪淑燕[1,*]，王绍山[2]，王海宁[3]，吕金飞[1]

(1. 航天工程大学，北京 101416；2. 北京跟踪与通信技术研究所，北京 100094；
3. 北京航空航天大学 人工智能研究院，北京 100191)

摘 要：欺骗干扰具有隐蔽性强、危害性大的特点，严重威胁全球导航卫星系统(GNSS)的安全稳定运行。目前基于运动单天线的欺骗干扰检测方法无法实现多来向欺骗干扰检测。为此，提出一种用于旋转单天线的基于波达方向角度差的欺骗干扰检测方法。该方法的本质是根据接收信号的载波相位估计一对信号之间的波达方向角度差。当没有欺骗信号时，波达方向角度差的估计值应与预测值一致。欺骗信号难以精确模拟导航信号的方向性，会破坏波达方向角度差的估计值和预测值一致性。因此，基于波达方向角度差的估计值和预测值建立检测变量，通过广义似然比检验(GLRT)可以检测多来向欺骗信号。进行软件仿真分析旋转单天线的参数对检测性能的影响，并通过现场试验进行验证和评估。结果表明：所提方法可高效实现多来向欺骗干扰的检测。

关 键 词：欺骗干扰检测；载波相位；波达方向；角度差；旋转天线
中图分类号：TN967.1
文献标志码：A **文章编号**：1001-5965(2026)01-0244-08

全球导航卫星系统(global navigation satellite system, GNSS)的信号强度弱，信号调制方式公开，且部分导航数据可以预测，极易受到欺骗干扰[1-2]，对于导航系统是致命的。因此，开展欺骗干扰检测的研究具有重要意义。在欺骗干扰检测方法研究初期，学者们主要通过寻找单天线接收机被欺骗后信号的变化特征来进行欺骗干扰检测。由于欺骗信号无法完美模拟真实信号的功率，欺骗干扰的加入会引起信号功率[3]、载噪比[4]等的变化。在捕获跟踪阶段，隐蔽欺骗策略下欺骗信号需要进入跟踪环路并缓慢增加功率、改变相位，该过程会引起真实码和欺骗码相位之间无法对准，出现多个相关峰[5]。基于以上发生变化的信号特征均可以进行欺骗干扰检测，这对于简单欺骗干扰具有很好的检测性能，是目前应用最多的欺骗干扰检测方法。惯性导航和GNSS组合可以通过对轨迹、加速度、位置等特征进行一致性检验达到欺骗干扰检测的目的。文献[6]将GNSS接收机估计的轨迹与惯导系统得到的轨迹进行比较。文献[7]对比了加速度计和GNSS接收机计算的加速度结果。当前，许多学者通过密码学的加密认证技术对民用导航信号进行加密，使导航信号难以被预测，并且接收机可以根据接收到的信号判断其完整性，从而更好地抵御欺骗攻击。根据加密认证方法和对象的不同可以将该方法分为导航消息认证[8]和扩频码认证[9]。基于信号加密认证的欺骗干扰检测方法需要修改导航信号体制，实现难度大。

此外，许多学者以欺骗信号和真实信号来向不同及来向不同引起的其他变量的不同为突破点进行欺骗信号的检测。文献[10]提出了一种基于双天线的欺骗检测技术，该方法的本质是根据载波相位观测值和导航信息精确估计一对固定天线之间的到达频率差，当没有欺骗信号时，观察结果应与预测结果一致，否则，由于GNSS卫星和欺骗设备

收稿日期：2023-11-22；**录用日期**：2024-03-19；**网络出版时间**：2024-04-03 19:43
网络出版地址：link.cnki.net/urlid/11.2625.V.20240329.1339.002
*通信作者. E-mail：daninini@163.com

引用格式：陈世淼，倪淑燕，王绍山，等. 基于旋转单天线的欺骗干扰检测方法[J]. 北京航空航天大学学报，2026，52(1)：244-251.
CHEN S M，NI S Y，WANG S S，et al. Spoofing detection method based on single rotating antenna[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：244-251 (in Chinese).

之间的几何和运动差异，欺骗信号会破坏一致性，因此，可以检测多来向欺骗干扰。文献 [11] 利用多个接收机天线的位置信息推导出不同信号的波达方向角度差，基于波达方向角度差的理想值与观测值进行广义似然比检验 (generalized likelihood ratio test, GLRT)，实现了多来向欺骗干扰检测。以上方法的运行均需要阵列天线或多个接收机，成本较高，因此，有学者提出采用运动的单天线进行信号的来向估计。文献 [12] 提出一种基于上下往复运动天线的欺骗干扰检测方法，通过天线上下运动引起的幅值变化与信号入射角的关系判定信号来向，实现欺骗干扰检测。此外，基于旋转单天线的架构，根据旋转单天线下的载波相位[13]、载噪比[14] 和功率[15] 变化趋势进行相似性检验，可以判断导航信号来向是否一致，从而实现欺骗干扰检测。然而，基于运动单天线的欺骗干扰检测方法无法检测来自多个方向的欺骗干扰，实现多个方向的欺骗干扰检测仍然需要多个天线或多个接收机，如何实现基于运动单天线的多来向欺骗干扰检测仍然是一个挑战。

因此，本文将 2 个信号的波达方向角度差应用于旋转单天线，提出了一种针对多来向欺骗干扰的检测方法。多来向欺骗设备难以精确模拟导航信号的方向性，这意味着欺骗信号与真实信号的入射方向会存在偏差。根据旋转单天线获取信号的载波相位，通过极大似然估计得到任意 2 个信号的波达方向角度差的估计值，通过星历信息计算波达方向角度差的预测值，基于预测值和估计值建立检测变量，判定 2 个信号中是否存在欺骗信号。对每组信号均进行 GLRT，可以实现欺骗信号检测。通过仿真对影响检测性能的特征参数进行详细分析。此外，通过现场试验对本文方法的有效性进行验证和评估。本文方法利用旋转单天线的载波相位来进行波达方向角度差的估计，实现多来向欺骗干扰的检测。需要注意的是，该方法中天线处于运动状态，但运动轨迹的圆心应保持不变。

1　系统模型

本节揭示旋转单天线中载波相位的变化规律，提出使用载波相位估计信号波达方向角度差的估计方法。

1.1　旋转单天线模型

旋转单天线模型如图 1 所示，天线以角度β固定于旋转底座，旋转底座以角速度ω进行旋转[13]。由于天线的旋转运动，天线的位置规律性变化，接收机接收到同一导航卫星信号的载波相位也呈规律性变化。对于来自同一方向的不同欺骗信号，其变化规律相同，对于来自不同方向的真实信号，其变化规律不同，因此，可根据以上特性进行欺骗干扰检测。以天线中心旋转圆周的圆心为坐标系原点建立直角坐标系，Ox轴指向 0 时刻原点O到天线相位中心的方向。

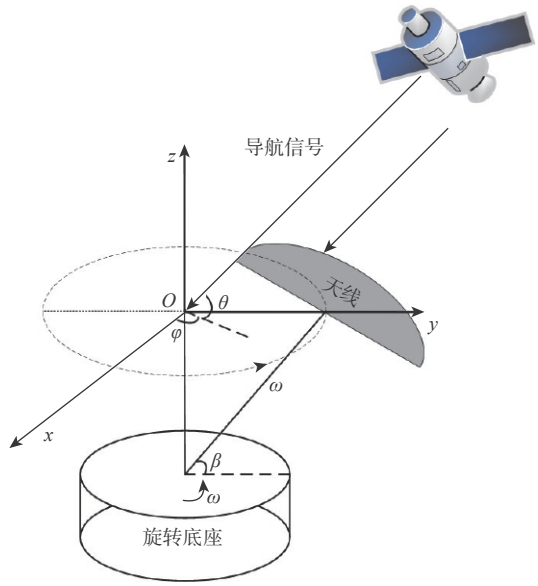


图 1　旋转单天线坐标系

Fig. 1　Rotating single antenna coordinate system

1.2　基于载波相位的极大似然估计

旋转过程中，天线接收到的 I 颗卫星的第 i 颗卫星在不同时刻的 N 个相位测量值为

$$\Phi_i(k)=\rho_i(k)+\varepsilon_i(k)$$
$$i=1,2,\cdots,I;k=0,1,2,\cdots,N-1 \tag{1}$$

式中：$\rho_i(k)$为第 i 颗卫星的载波相位真值；$\varepsilon_i(k)$为第 i 颗卫星的载波相位测量噪声。

由于天线以角速度ω进行旋转，$\rho_i(k)$可以表示为[13]

$$\rho_i(k)=\rho_{i,\mathrm{sat}}(k)+r\cos\theta_i\cos(\omega T_0k+\varphi_i+\varphi_0) \tag{2}$$

式中：$\rho_{i,\mathrm{sat}}(k)$为静止原点 O 处的载波相位真值，其变化主要由卫星相对于原点 O 的运动产生；φ_i和θ_i分别为第 i 颗卫星的信号入射到坐标系的方位角和俯仰角；φ_0为旋转初始相位；r为天线中心到旋转中心的水平投影距离；ω为天线的旋转角速度；T_0为旋转天线所连接接收机的采样周期。

对载波相位测量值进行前向后向差分，可以得到[13]

$$\mathrm{d}\Phi_i(k)=\Phi_i(k+1)-\Phi_i(k-1)=\mathrm{d}\rho_{i,\mathrm{sat}}(k)-$$
$$2r\sin(\omega T_0)\cos\theta_i\sin(\omega T_0k+\varphi_i+\varphi_0)+\delta(k)$$
$$k=1,2,\cdots,N-2 \tag{3}$$

式中：$\delta(k)$为噪声项差分运算得到的新噪声项。

对于处于静止状态的接收机，卫星导航欺骗信号也需要模拟导航卫星运动引起的信号特性变化，即真实信号和欺骗信号由于卫星运动产生的载波相位变化$\mathrm{d}\rho_{i,\mathrm{sat}}(k)$相同，并且可以根据星历数据计算卫星运动的变化量得到$\mathrm{d}\rho_{i,\mathrm{sat}}(k)$。因此，$\mathrm{d}\rho_{i,\mathrm{sat}}(k)$可作为已知量，得到仅包含误差和由天线旋转运动引起的载波相位变化的差分序列：

$$\mathrm{d}\Phi_{i,\mathrm{rcv}}(k)=W_{\Phi,i}(k)-2r\sin(\omega T_0)\cos\theta_i\sin(\omega T_0k+\varphi_i+\varphi_0) \tag{4}$$

对式(4)连续化可以得到

$$\mathrm{d}\Phi_{\mathrm{rcv}}(t)=M\sin(\omega t+\gamma)+W_\Phi=s(t;M,\gamma)+W_\Phi \tag{5}$$

式中：t为时间；下标rcv表示接收机；$M=-2r\sin(\omega T_0)\cdot\cos\theta_i$；$\gamma=\varphi_1+\varphi_0$；$W_\Phi$表示方差为$\sigma_\Phi^2$的零均值加性高斯白噪声。

将载波相位测量差分值表示为$\boldsymbol{x}=[\mathrm{d}\Phi_{\mathrm{rcv}}(0),\mathrm{d}\Phi_{\mathrm{rcv}}(T_0),\mathrm{d}\Phi_{\mathrm{rcv}}(T_1),\cdots,\mathrm{d}\Phi_{\mathrm{rcv}}((N-1)T_0)]^{\mathrm{T}}$，通过极大似然估计计算变量$\boldsymbol{\mu}=[M,\gamma]$，建立代价函数如下：

$$\begin{aligned}J(\boldsymbol{\mu})=&\sum_{n=0}^{N-1}(\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)-M\sin(\omega nT_0+\gamma))^2=\\&\sum_{n=0}^{N-1}(\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)-M\cos\gamma\sin(\omega nT_0)-\\&M\sin\gamma\cos(\omega nT_0))^2\end{aligned} \tag{6}$$

为简化代价函数，假设矩阵：

$$\boldsymbol{\alpha}=[\alpha_1\quad\alpha_2]^{\mathrm{T}}=[M\cos\gamma\quad M\sin\gamma]^{\mathrm{T}} \tag{7}$$

$$\boldsymbol{H}=[\boldsymbol{h}_1\quad\boldsymbol{h}_2]=\begin{bmatrix}0&1\\\sin(\omega T_0)&\cos(\omega T_0)\\\sin(2\omega T_0)&\cos(2\omega T_0)\\\vdots&\vdots\\\sin((N-1)\omega T_0)&\cos((N-1)\omega T_0)\end{bmatrix} \tag{8}$$

将代价函数简化为

$$J(\boldsymbol{\mu})=(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\alpha})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\alpha}) \tag{9}$$

因此，$\boldsymbol{\alpha}$的极大似然估计值为[15]

$$\hat{\boldsymbol{\alpha}}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{x} \tag{10}$$

考虑到向量$\boldsymbol{h}_1$和$\boldsymbol{h}_2$的周期性，矩阵$\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$可以进行简化：

$$\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}=\begin{bmatrix}\boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_1&\boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{h}_2\\\boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_1&\boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{h}_2\end{bmatrix}\approx\begin{bmatrix}N/2&0\\0&N/2\end{bmatrix} \tag{11}$$

将式(11)代入式(10)，$\boldsymbol{\alpha}$的极大似然估计值可以简化为

$$\hat{\boldsymbol{\alpha}}=\begin{bmatrix}\hat{\boldsymbol{\alpha}}_1\\\hat{\boldsymbol{\alpha}}_2\end{bmatrix}\approx\frac{2}{N}\begin{bmatrix}\boldsymbol{h}_1^{\mathrm{T}}\boldsymbol{x}\\\boldsymbol{h}_2^{\mathrm{T}}\boldsymbol{x}\end{bmatrix} \tag{12}$$

由此可以得到极大似然估计的解：

$$\hat{\gamma}=\arctan\left(\frac{\hat{\boldsymbol{\alpha}}_2}{\hat{\boldsymbol{\alpha}}_1}\right)=\arctan\left(\frac{\sum_{n=0}^{N-1}\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)\cos(n\omega T_0)}{\sum_{n=0}^{N-1}\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)\sin(n\omega T_0)}\right) \tag{13}$$

$$\begin{aligned}\hat{M}=&-\sqrt{\hat{\boldsymbol{\alpha}}_1^2+\hat{\boldsymbol{\alpha}}_2^2}=\\&-\frac{2}{N}\sqrt{\left[\sum_{n=0}^{N-1}\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)\cos(n\omega T_0)\right]^2+\left[\sum_{n=0}^{N-1}\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)\sin(n\omega T_0)\right]^2}\end{aligned} \tag{14}$$

定义$\rho=n\omega T_0+\gamma$，求解费歇耳信息矩阵的参数如下：

$$\begin{aligned}[\boldsymbol{I}(\boldsymbol{\mu})]_{MM}=&\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}\left(\frac{\partial s(nT_0;\boldsymbol{\mu})}{\partial M}\right)^2=\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}\sin^2\rho=\\&\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}\left(\frac{1}{2}-\frac{1}{2}\cos(2\rho)\right)\approx\frac{N}{2\sigma_\Phi^2}\end{aligned} \tag{15}$$

$$\begin{aligned}[\boldsymbol{I}(\boldsymbol{\mu})]_{M\gamma}=[\boldsymbol{I}(\boldsymbol{\mu})]_{\gamma M}=&\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}\frac{\partial s(nT_0;\boldsymbol{\mu})}{\partial M}\cdot\frac{\partial s(nT_0;\boldsymbol{\mu})}{\partial\gamma}=\\&-\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}M\cos\rho\sin\rho=-\frac{M}{2\sigma_\Phi^2}\sum_{n=0}^{N-1}\sin(2\rho)\approx0\end{aligned} \tag{16}$$

$$\begin{aligned}[\boldsymbol{I}(\boldsymbol{\mu})]_{\gamma\gamma}=&\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}\left(\frac{\partial s(nT_0;\boldsymbol{\mu})}{\partial\gamma}\right)^2=\frac{1}{\sigma_\Phi^2}\sum_{n=0}^{N-1}M^2\cos^2\rho=\\&\frac{M^2}{\sigma_\Phi^2}\sum_{n=0}^{N-1}\left(\frac{1}{2}+\frac{1}{2}\cos(2\rho)\right)\approx\frac{NM^2}{2\sigma_\Phi^2}\end{aligned} \tag{17}$$

因此，费歇耳信息矩阵为

$$\boldsymbol{I}(\boldsymbol{\mu})=\frac{1}{\sigma_\Phi^2}\begin{bmatrix}N/2&0\\0&NM^2/2\end{bmatrix} \tag{18}$$

根据极大似然估计的渐近性，$\hat{\gamma}\sim N(\varphi+\varphi_0,\sigma_\gamma^2)$，$\hat{M}\sim N(-2r\sin(\omega T_0)\cos\theta,\sigma_M^2)$[15]，其中方差为

$$\sigma_\gamma^2=\left(N\sigma_\Phi^2\right)\Big/\left\{2\left[\left(\sum_{n=0}^{N-1}\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)\cos(n\omega T_0)\right)^2+\left(\sum_{n=0}^{N-1}\mathrm{d}\Phi_{\mathrm{rcv}}(nT_0)\sin(n\omega T_0)\right)^2\right]\right\} \tag{19}$$

$$\sigma_M^2=\frac{2\sigma_\Phi^2}{N} \tag{20}$$

对于来自卫星i和卫星j的导航信号，通过其载波相位可以得到2组信号对应的γ和M，满足：$\hat{\gamma}_i\sim N(\varphi_i+\varphi_0,\sigma_{\gamma,i}^2)$、$\hat{\gamma}_j\sim N(\varphi_j+\varphi_0,\sigma_{\gamma,j}^2)$、$\hat{M}_i\sim N(-2r\cdot\sin(\omega T_0)\cos\theta_i,\sigma_{M,i}^2)$和$\hat{M}_j\sim N(-2r\sin(\omega T_0)\cos\theta_j,\sigma_{M,j}^2)$。

进一步地，$\mathrm{d}\hat{\gamma}_{ij}\sim N(\mathrm{d}\varphi_{ij},\sigma_{\gamma,i}^2+\sigma_{\gamma,j}^2)$，对其进行标准化：

$$\frac{\mathrm{d}\hat{\gamma}_{ij}-\mathrm{d}\varphi_{ij}}{\sqrt{\sigma_{\gamma,i}^2+\sigma_{\gamma,j}^2}}\sim N(0,1) \tag{21}$$

$$\frac{\hat{M}_i+2r\sin(\omega T_0)\cos\theta_i}{\sigma_{M,i}}\sim N(0,1) \tag{22}$$

$$\frac{\hat{M}_j+2r\sin(\omega T_0)\cos\theta_j}{\sigma_{M,j}}\sim N(0,1) \tag{23}$$

以上是导航信号为真实信号时的分析结果。当存在欺骗信号时，以上变量将不符合标准正态分布。

2　欺骗干扰检测方法

分析欺骗干扰对旋转单天线下载波相位观测值的影响，基于观测值构建了假设检验模型，运用 GLRT 方法解决假设检验问题，从而实现对欺骗干扰的有效检测。

2.1　假设检验

通过旋转单天线获取卫星 i 和卫星 j 的载波相位，计算得到 $\mathrm{d}\hat{\gamma}_{ij}$、$\hat{M}_i$ 和 $\hat{M}_j$，建立误差平方和（sum of squares due to error, SSE）统计值 t_{SSE} 来评估以上变量观测值与期望值之间的差异：

$$t_{\mathrm{SSE}}=\frac{\left(\mathrm{d}\hat{\gamma}_{ij}-\mathrm{d}\varphi_{ij}\right)^2}{\sigma_{\gamma,i}^2+\sigma_{\gamma,j}^2}+\frac{\left(\hat{M}_i+2r\sin(\omega T_0)\cos\theta_i\right)^2}{\sigma_{M,i}^2}+\frac{\left(\hat{M}_j+2r\sin(\omega T_0)\cos\theta_j\right)^2}{\sigma_{M,j}^2} \tag{24}$$

如果不存在欺骗信号，以上变量的观测值与期望值应当保持一致，两者之间差值应当是零均值的正态分布。在这种情况下，SSE 统计值应当满足自由度为 3 的卡方分布，记作 $\chi^2(3)$。当存在欺骗信号时，以上变量的观测值与期望值之间将存在偏差，SSE 统计值应满足非中心卡方分布，自由度为 3，偏心量为 Δ。基于以上分析，建立假设如下：

$$\begin{cases}H_0:t_{\mathrm{SSE}}\sim\chi^2(3) & \text{no spoofing}\\ H_1:t_{\mathrm{SSE}}\sim\chi^2(3,\Delta) & \text{spoofing}\\ \Delta=\dfrac{\left(\mathrm{d}\gamma_{ij}-\mathrm{d}\varphi_{ij}\right)^2}{\sigma_{\gamma,i}^2+\sigma_{\gamma,j}^2}+\dfrac{(M_i+2r\sin(\omega T_0)\cos\theta_i)^2}{\sigma_{M,i}^2}+\left(M_j+2r\sin(\omega T_0)\cos\theta_j\right)^2/\sigma_{M,j}^2\end{cases} \tag{25}$$

SSE 统计值在零假设和欺骗假设下的概率密度函数为

$$\begin{cases}p(t_{\mathrm{SSE}}|H_0)=\dfrac{t_{\mathrm{SSE}}\mathrm{e}^{-t_{\mathrm{SSE}}/2}}{3\Gamma(3)} & t_{\mathrm{SSE}}\geqslant 0\\ p(t_{\mathrm{SSE}}|H_1)=\dfrac{\mathrm{e}^{-(t_{\mathrm{SSE}}+\Delta)/2}\sqrt{t_{\mathrm{SSE}}/\Delta}}{2}\mathrm{I}_1(\sqrt{t_{\mathrm{SSE}}\Delta}) & t_{\mathrm{SSE}}\geqslant 0\end{cases} \tag{26}$$

式中：$\Gamma(\cdot)$ 为伽马函数；$\mathrm{I}_1(\cdot)$ 为第 1 类一阶变换的贝塞尔函数。

建立 GLRT：

$$\lambda(t_{\mathrm{SSE}})=\frac{p(t_{\mathrm{SSE}}|H_1)}{p(t_{\mathrm{SSE}}|H_0)}\underset{H_0}{\overset{H_1}{\gtrless}}\eta \tag{27}$$

式中：η 为检测阈值。

根据奈曼皮尔逊准则设置门限值 $t_{\mathrm{SSE,th}}$ 来确定特定虚警概率 P_{fa} 下的检测概率 P_{D}：

$$\begin{cases}P_{\mathrm{fa}}=P\left\{t_{\mathrm{SSE}}>t_{\mathrm{SSE,th}}\mid H_0\right\}=1-\int_0^{t_{\mathrm{SSE,th}}}p_{\chi^2(3)}(t_{\mathrm{SSE}})\mathrm{d}t_{\mathrm{SSE}}\\ P_{\mathrm{D}}=P\left\{t_{\mathrm{SSE}}>t_{\mathrm{SSE,th}}\mid H_1\right\}=\int_{t_{\mathrm{SSE,th}}}^{\infty}p_{\chi^2(3,\Delta)}(t_{\mathrm{SSE}})\mathrm{d}t_{\mathrm{SSE}}\end{cases} \tag{28}$$

基于以上分析，设计基于旋转单天线的欺骗干扰检测方法，详细流程如下：

第 1 阶段　数据采集。采集卫星 i 的载波相位测量值，定义为 $\Phi_i(k)$，$i=1,2,\cdots,I$，$k=0,1,2,\cdots,N-1$。计算由于卫星相对于原点 O 的运动产生的载波相位变化值，对载波相位测量值进行前向后向差分，得到载波相位单差值，记为 $\mathrm{d}\Phi_i(k)$。

第 2 阶段　参数的极大似然估计。

1）根据式 (13) 和式 (19) 计算卫星 i 和 j 的参数 $\hat{\gamma}_i$ 和 $\hat{\gamma}_j$ 的估计值和方差，根据式 (14) 和式 (20) 计算卫星 i 和 j 的参数 $\hat{M}_i$ 和 $\hat{M}_j$ 的估计值和方差。

2）根据星历信息计算卫星 i 和 j 的波达方向角度差 $\mathrm{d}\varphi_{ij}$ 和俯仰角 θ_i、θ_j。

第 3 阶段　似然比检验。

1）根据设定的检测概率和虚警概率确定监测阈值。

2）根据式 (24) 计算 SSE 统计值，如果 SSE 统计值小于阈值，则认定卫星 i 和 j 的信号为真实信号，否则，至少其中一个信号为欺骗信号。

2.2　性能分析

受试者工作特征（receiver operator characteristic, ROC）曲线能够体现阈值变化时检测方法的检测性能。本节以 ROC 曲线为参考，通过仿真分析影响检测性能的因素。除特殊说明外，仿真过程中旋转单天线的参数设置如表 1 所示。

表 1　仿真参数设置

Table 1　Simulation parameter settings

r/m	ω/(rad·s^{-1})	T_0/s	β/(°)	N	σ_ϕ^2/m^2	$\mathrm{d}\varphi_2$/(°)	$\mathrm{d}\theta_2$/(°)
0.2	$\pi/6$	1	70	100	0.02	10	10

图 2 和图 3 分别为参数 $r\sin(\omega T_0)$ 和 N 取值不同时的 ROC 曲线，仿真结果显示，以上参数的变化均会影响 ROC 曲线。对仿真结果进行分析，可以得出以下结论：①采样数据越多，欺骗检测性能越好；②天线中心到旋转中心的水平投影距离越大，检测性能越好；③天线旋转角速度与采样周期的乘积越接近 π，检测性能越好。此外，检测性能随噪声方差

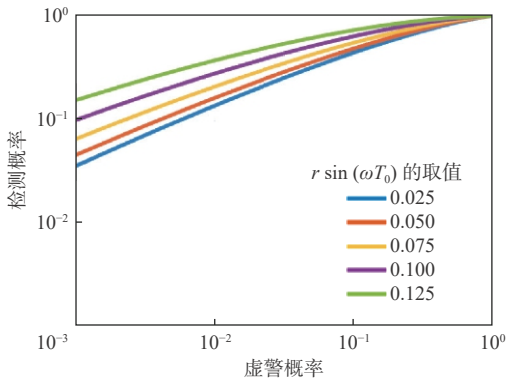


图 2 $r\sin(\omega T_0)$与 ROC 曲线的关系

Fig. 2 Relationship between $r\sin(\omega T_0)$ and ROC curves

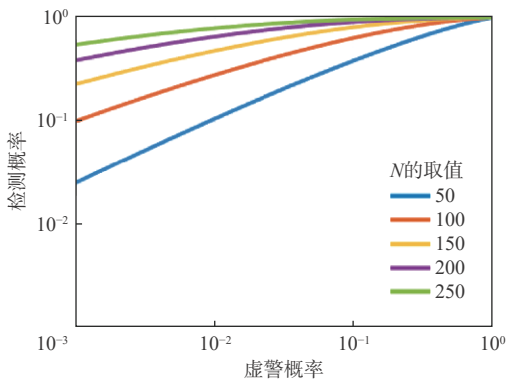


图 3 N与 ROC 曲线的关系

Fig. 3 Relationship between N and ROC curves

的增加而降低。

3 仿真验证

本节对提出的方法进行软件仿真。首先，通过 STK 软件获取当前可视导航卫星的精确位置信息，并据此计算导航卫星相对于旋转天线的俯仰角和方位角，同时，设置欺骗导航信号的角度参数。然后，根据卫星和旋转天线的运动规律，模拟生成载波相位的观测值，并在其中加入白噪声以模拟真实环境中的干扰。其次，基于载波相位的观测值，进一步估计相关参数的值。最后，通过计算 SSE 统计值，并结合 GLRT 方法，准确判断导航信号是否为欺骗信号。仿真过程中旋转天线的参数设置如表 1 所示。

观察式 (25) 可以发现，偏心量越大，检测性能越好，同时，偏心量由 3 组多项式组成，其中，第 1 项与 2 组信号方位差值相关，后 2 项与 2 组信号的俯仰角相关。为全面分析欺骗信号偏离真实信号角度对检测性能的影响，设置了 3 种仿真场景：①欺骗信号仅在俯仰角方向存在偏差；②欺骗信号仅在方位角方向存在偏差；③欺骗信号在俯仰角和方位角方向都存在偏差。图 4 为仿真场景的卫星

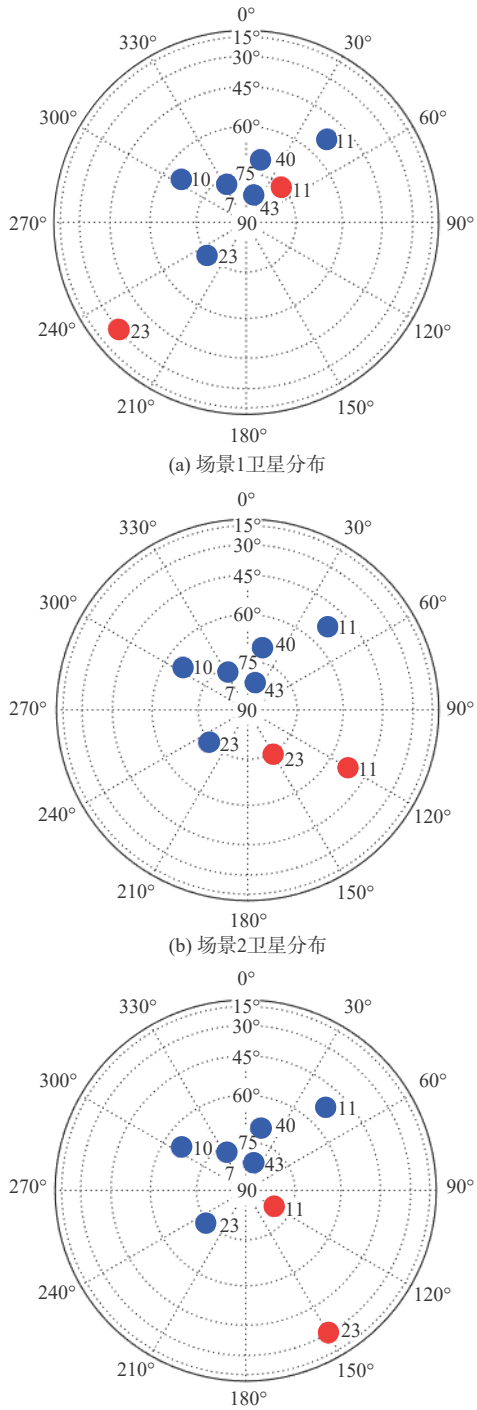


图 4 ENU 坐标系下的卫星分布

Fig. 4 Satellite distribution in the ENU coordinate system

分布图。图中：蓝色为真实信号位置，红色为欺骗信号位置。15°、30°、45°、60°为导航卫星相对于旋转单天线的俯仰角度，0°, 30°, 60°, …, 300°, 330°为导航卫星相对于旋转单天线的方位角度。仿真过程中，接收机接收到的信号为 PRN 7、10、40、43、11、23 的真实信号和 PRN 11、23 的欺骗信号。

图 5 展示了仿真过程中 SSE 在不同卫星组合下的统计值，采用对数表示。图例中欺骗信号的 PRN 编号增加了上划线，包含欺骗信号的 SSE 统计

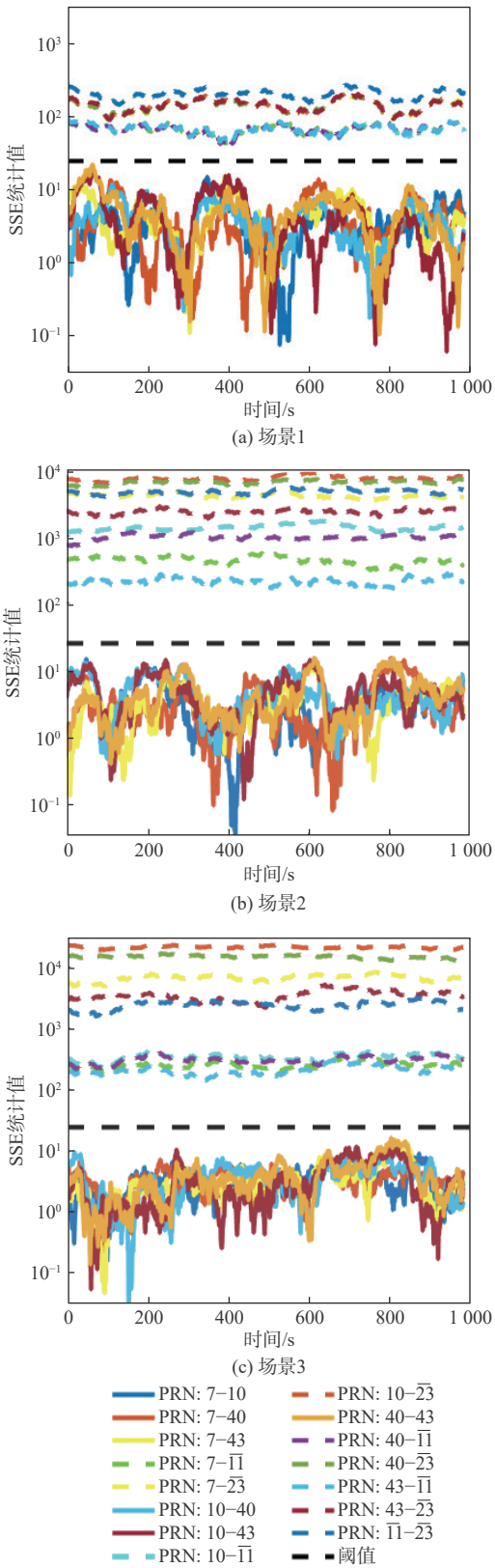


图 5　不同信号组合下的 SSE 统计值

Fig. 5　SSE values under different signal combinations

值用虚线表示，不包含欺骗信号的 SSE 统计值用实线表示，阈值采用黑色虚线表示，设置为 15。仿真结果显示，只要组合信号中包含欺骗信号，其 SSE 统计值均保持在阈值之上，如果组合信号中仅包含真实信号，则 SSE 统计值均保持在阈值之下，说明本文方法可有效实现欺骗干扰的检测。

综上，可以通过以下步骤实现导航欺骗信号的检测：①从接收机读取信号的载波相位；②计算多组组合信号的 SSE 统计值；③根据低于阈值的 SSE 统计值确定真实信号；④根据包含真实信号且高于阈值的 SSE 统计值确定欺骗信号。

4　试验验证

4.1　环境搭建和参数设置

图 6 为欺骗试验场景，由欺骗源、旋转单天线和中频信号采集装置组成。2 个欺骗源模拟多来向欺骗设备发送欺骗信号，中频信号采集装置连接旋转单天线，存储旋转单天线接收到的导航信号。后续通过 FGI-GSRx 软件接收机（https://github.com/nlsfi/FGI-GSRx）处理数据集中的中频信号，将处理后的数据进行欺骗检测方法的验证。

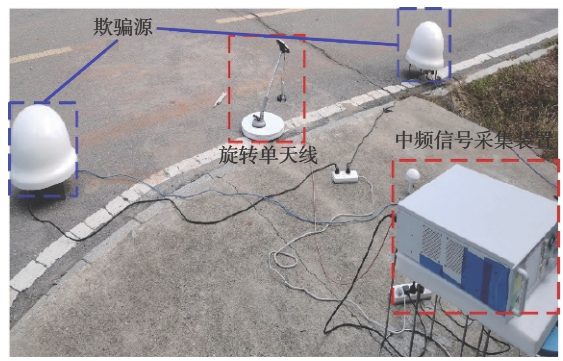


图 6　欺骗试验场景

Fig. 6　Spoofing scenario

旋转单天线旋转半径为 0.15 m，倾斜角度为 30°，旋转角速度为 18 (°)/s，接收机采样频率为 10 Hz。欺骗检测方法进行极大似然估计的样本数为 400。2 个欺骗源记作欺骗源 A 和欺骗源 B。欺骗源 A 发送 PRN 为 10、12、15 的导航欺骗信号，欺骗源 B 发送 PRN 为 18、23、24 的导航欺骗信号。旋转单天线放置于坐标为 (0, 0, 0) m 位置，欺骗源 A 和欺骗源 B 位于本地坐标系的 (0, 2, 0.5) m 和 (0, −2, 0.5) m。整个试验过程持续 490 s，在第 80 s 时开始施加时间欺骗，并且进行持续拉偏。

4.2　方法验证和性能分析

图 7 展示了不同卫星组合下检测变量 SSE 统计值的变化。为了便于观察，两两组合的检测变量同时用颜色和图形进行区分。线的颜色和点的形状分别表示不同的卫星，蓝色表示组合包含 PRN10，黄色和圆形表示组合包含 PRN12，红色和方形表示组合包含 PRN15，绿色和菱形表示组合包含 PRN18，紫色和五角星表示组合包含 PRN23，六角形表示组合包含 PRN24。黑色虚线表示基于奈曼皮尔逊准

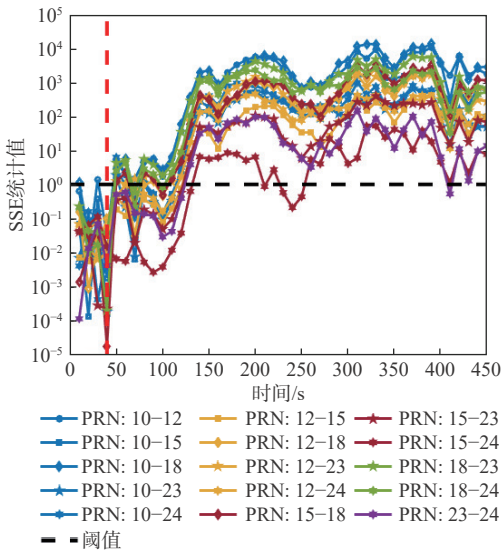


图7 不同卫星组合下检测器值的变化曲线

Fig. 7 The variation curves of detector values under different satellite combinations

则计算得到的最优检测阈值，红色虚线表示欺骗开始的时间。需要说明的是，在使用极大似然估计进行参数估计时需要40 s的数据，因此，检测变量的时间长度为450 s。

观察图7可以看出：①欺骗开始前检测变量的值小于欺骗开始后的检测变量的值，说明欺骗信号出现后检测变量不满足中心卡方分布，与2.1节的分析一致；②欺骗初始阶段（40～80 s）检测变量的值小于后续（80～450 s）检测变量的值，说明本文方法在欺骗初始阶段检测效果较差。分析原因为：初始阶段进行极大似然估计的样本中仍然包含一部分真实信号，极大似然估计的结果不准确。

基于图7中的不同检测变量，通过选取不同的阈值并得到不同阈值下的检测概率和虚警概率，由此可以得到图8所示的ROC曲线。奈曼皮尔逊准则是一种最小错误率准则，用于确定二元分类器的决策阈值，该准则的目标是最小化误报和漏报的总

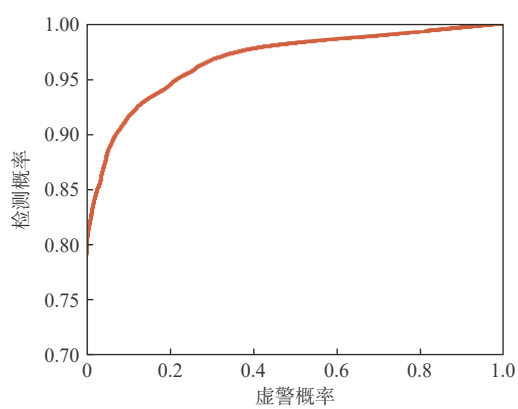


图8 ROC曲线

Fig. 8 ROC curve

错误率。以奈曼皮尔逊准则确定检测阈值，最优阈值下的检测概率可以认为是最优检测概率，此时的虚警概率为最优虚警概率。根据奈曼皮尔逊准则计算得到试验数据中的最优检测概率为0.885 3，最优虚警概率为0.052 3。

综上，对于两两组合的导航信号，使用极大似然估计进行参数估计，将检测变量与阈值进行对比可以判定该组合中是否包含欺骗信号。现场试验结果证明了本文方法的有效性。

5 结 论

本文提出一种应用于旋转单天线的波达方向角度差的欺骗干扰检测方法。基于极大似然估计，使用多个历元的载波相位计算信号波达方向角度差的估计值，结合波达方向角度差的预测值建立欺骗检测变量。通过软件仿真分析旋转单天线参数对检测性能的影响。进行现场试验对本文提出的欺骗检测方法进行验证和评估。根据仿真和试验结果，可以得出以下结论：

1）本文方法可有效实现多来向欺骗干扰的检测，在搭建的欺骗环境中可以达到0.885 3的检测成功率。

2）检测性能受到旋转单天线参数的影响：进行极大似然估计的数据越多，欺骗检测性能越好；天线中心到旋转中心的水平投影距离越大，检测性能越好；天线旋转角速度与采样周期的乘积越接近π，检测性能越好。

3）与其他基于信号来向的欺骗检测方法相比，本文方法仅需要单个旋转天线即可实现多来向欺骗干扰检测。该技术可应用于固定站点接收机的欺骗干扰检测。

未来将以提高本文方法的检测性能为目的进一步展开研究。

参考文献（References）

[1] XIAO L, XIE C X, MIN M H, et al. User-centric view of unmanned aerial vehicle transmission against smart attacks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(4): 3420-3430.

[2] JANSEN K, SCHÄFER M, MOSER D, et al. Crowd-GPS-Sec: leveraging crowdsourcing to detect and localize GPS spoofing attacks[C]//Proceedings of the 2018 IEEE Symposium on Security and Privacy. Piscataway: IEEE Press, 2018: 1018-1031.

[3] JAFARNIA-JAHROMI A, BROUMANDAN A, NIELSEN J, et al. Pre-despreading authenticity verification for GPS L1 C/A signals[J]. Navigation, 2014, 61(1): 1-11.

[4] HU Y F, BIAN S F, CAO K J, et al. GNSS spoofing detection based on new signal quality assessment model[J]. GPS Solutions, 2018, 22(1): 28.

[5] SUN C, CHEONG J W, DEMPSTER A G, et al. Moving variance-based signal quality monitoring method for spoofing detection[J]. GPS Solutions, 2018, 22(3): 83.

[6] BROUMANDAN A, LACHAPELLE G. Spoofing detection using GNSS/INS/odometer coupling for vehicular navigation[J]. Sensors, 2018, 18(5): 1305.

[7] KWON K C, SHIM D S. Performance analysis of direct GNSS spoofing detection with accelerometers for constant velocity[J]. International Journal of Control, Automation and Systems, 2022, 20(8): 2749-2758.

[8] FERNÁNDEZ-HERNÁNDEZ I, ASHUR T, RIJMEN V. Analysis and recommendations for MAC and key lengths in delayed disclosure GNSS authentication protocols[J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(3): 1827-1839.

[9] WU Z J, ZHANG Y, LIU R S. BD-II NMA&SSI: an scheme of anti-spoofing and open BeiDou II D2 navigation message authentication[J]. IEEE Access, 2020, 8: 23759-23775.

[10] HE L, LI H, LU M Q. Dual-antenna GNSS spoofing detection method based on Doppler frequency difference of arrival[J]. GPS Solutions, 2019, 23(3): 78.

[11] CHEN Z Y, LI H, WEI Y M, et al. GNSS antispoofing method using the intersection angle between two directions of arrival (IA-DOA) for multiantenna receivers[J]. GPS Solutions, 2023, 27(1): 11.

[12] HE L, LI H, LU M Q. Global navigation satellite system spoofing-detection technique based on the Doppler ripple caused by vertical reciprocating motion[J]. IET Radar, Sonar & Navigation, 2019, 13(10): 1655-1664.

[13] CHANG H W, PANG C L, ZHANG L, et al. Rotating single-antenna spoofing signal detection method based on IPNN[J]. Sensors, 2022, 22(19): 7141.

[14] GENG Z L, NIE J W, XIAO Z B, et al. A GNSS spoofing detection technique based on signal spatial correlation[C]//Proceedings of the China Satellite Navigation Conference 2017: Volume I. Berlin: Springer, 2017: 959-969.

[15] WANG F, LI H, LU M Q. GNSS spoofing countermeasure with a single rotating antenna[J]. IEEE Access, 2017, 5: 8039-8047.

Spoofing detection method based on single rotating antenna

CHEN Shimiao[1], NI Shuyan[1,*], WANG Shaoshan[2], WANG Haining[3], LYU Jinfei[1]

(1. Space Engineering University, Beijing 101416, China;
2. Beijing Institute of Tracking and Communication Technology, Beijing 100094, China;
3. Institute of Artificial Intelligence, Beihang University, Beijing 100191, China)

Abstract: Spoofing is a serious security risk to the global navigation satellite system (GNSS) because of its strong concealment and high harm features. Multi-directional spoofing detection is not possible with existing spoofing detection techniques that rely on moving single antennas. To achieve this, we provide a spoofing detection technique for a single spinning antenna that is based on the intersection angle between two directions of arrival. The essence of this method is to estimate the intersection angle between two directions of arrival based on the carrier phase difference of the received signal. The projected value and the estimated value of the intersection angle between two directions of arrival should match in the absence of spoofing. Spoofing is difficult to accurately simulate the directionality of navigation signals, which can disrupt the consistency between the estimated and predicted values of the intersection angle between two arrival directions. Therefore, a detection variable is established based on the estimated and predicted values of the intersection angle between two directions of arrival, and the generalized likelihood ratio test (GLRT) can detect multi-directional spoofing. We used on-site experiments to confirm and assess the simulation to examine how single rotating antenna settings affect detection performance. The results indicate that the method proposed in this article can efficiently detect multi-directional spoofing.

Keywords: spoofing detection; carrier phase; direction of arrival; intersection angle; rotating antenna

Received: 2023-11-22; **Accepted**: 2024-03-19; **Published Online**: 2024-04-03 19:43

URL: link.cnki.net/urlid/11.2625.V.20240329.1339.002

* **Corresponding author**. E-mail: daninini@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0709

无对齐实体场景的多语言知识图谱补全

唐榕氚[1,2]，徐秋程[2]，汤闻易[2]，翟飞飞[1,3]，周玉[1,3,*]

(1. 中国科学院 自动化研究所，北京 100190；2. 南京航空航天大学 民航学院 空中交通管理系统全国重点实验室，南京 210007；
3. 北京中科凡语科技有限公司，北京 100190)

摘 要：多语言知识图谱补全(MKGC)旨在利用其他语言知识图谱的信息增强目标语言知识图谱上的链接预测性能。现有方法通常利用不同语言知识图谱之间预先对齐的实体对作为知识迁移的媒介，然而在实际场景中，不同语言知识图谱之间通常没有预先对齐的实体，导致难以实现知识迁移。针对上述无对齐实体场景，提出一种融合预训练语言模型信息的伪对齐实体生成模块，不断迭代生成新的对齐实体进行知识迁移。为区分不同语言知识图谱中信息对目标语言知识图谱的贡献度，提出一种基于多图注意力的图神经网络(MGA-GNN)用于对三元组进行编码，通过该网络输出的嵌入表征计算得到三元组的合理性得分，完成链接预测任务。为验证所提方法的有效性，在2个公开数据集DBP-5L和E-PKG上进行了实验验证，结果表明：所提方法在多个语言知识图谱上链接预测的性能超过了有对齐实体的MKGC方法，证明了该方法在更加实际场景下的优越性能。

关 键 词：多语言知识图谱补全；实体对齐；多图注意力；图神经网络；链接预测
中图分类号：TP391
文献标志码：A **文章编号**：1001-5965(2026)01-0252-08

具有丰富结构化知识的知识图谱(knowledge graph, KG)被广泛应用于各种人工智能领域，如问答系统[1-2]、推荐系统[3-4]和信息抽取[5]等。在知识图谱中，知识以三元组的形式进行存储，包含头实体、关系和尾实体，知识图谱补全(knowledge graph completion, KGC)任务也称链接预测，旨在根据已有知识预测未知三元组中缺失的部分以进行知识的扩充，缓解知识图谱普遍存在的不完备问题。通常，KGC只在一个知识图谱上进行，而考虑到现实世界中存在许多不同语言的知识图谱，这些知识图谱之间存在大量的知识重叠和特有知识。为了利用不同语言的知识图谱信息来增强目标知识图谱的链接预测性能，近年来，多语言知识图谱补全(multilingual KGC, MKGC)引起了许多研究者的关注。在MKGC任务中，对于待补全的目标知识图谱，如何将其他语言的知识迁移到该知识图谱上成为了主要难点。现有工作，如KEnS[6]、AlignKGC[7]、SS-AGA[8]都将不同知识图谱之间存在的预先对齐的实体作为知识迁移的桥梁，通过这些实体将其他语言的知识聚集到目标知识图谱上。然而，在更加符合实际的场景下，通常只能获取不同语言的知识图谱，而无法直接获取这些知识图谱之间实体的对齐情况。但是，很少有研究关注无对齐实体场景下的MKGC任务。本文力图设计一种简单而有效的方法来实现无对齐实体场景下的MKGC，使其能够更好地在实际场景下进行应用。

KGC主要采用基于知识图谱嵌入的方法，如TransE[9]、DistMult[10]、RotatE[11]。上述方法先将实体和关系嵌入到低维向量空间中，为三元组建立不同的几何关系，然后设计不同的评分函数来计算三

收稿日期：2023-10-31；**录用日期**：2023-12-29；**网络出版时间**：2024-03-12 11:05
网络出版地址：link.cnki.net/urlid/11.2625.V.20240311.1607.006
基金项目：空中交通管理系统全国重点实验室开放课题(SKLATM202203)
*通信作者. E-mail：yzhou@nlpr.ia.ac.cn

引用格式：唐榕氚，徐秋程，汤闻易，等. 无对齐实体场景的多语言知识图谱补全[J]. 北京航空航天大学学报，2026，52(1)：252-259.
TANG R C，XU Q C，TANG W Y，et al. Multilingual knowledge graph completion without aligned entity pairs[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：252-259 (in Chinese).

元组的合理性分数。由于知识图谱特有的图结构, 基于图神经网络(graph neural network, GNN)的方法在 KGC 任务上受到了广泛关注[12-13], 具有强大图结构建模能力的 GNN 在此任务上展示出了巨大潜力。

现有 MKGC 相关的方法常常与知识图谱的实体对齐任务紧密关联, 而跨语言知识图谱的实体对齐方法常常利用给定的对齐实体作为监督信号指导模型的训练[14-16], 因此, 当前的 MKGC 任务均在有对齐实体场景下才能实现。KEnS[6] 首次通过集成的方法利用多种不同语言知识图谱的互补知识提升了单语言 KGC 的性能。AlignKGC[7] 在多个语言的知识图谱上联合进行补全、实体对齐和关系对齐, 并提高了 KGC 的准确性和实体对齐性能。SS-AGA[8] 通过使用 GNN 动态添加对齐实体来改进 MKGC 任务。然而, 上述方法均需要预先对齐的实体对, 限制了进一步的应用。

本文提出一种不需要预先对齐实体的 MKGC 方法。首先, 设计了翻译模块和预训练语言模型相结合的对齐实体获取模块, 能够在初始阶段获取高质量的对齐实体。其次, 设计了基于多图注意力的图神经网络(graph neural network based on multi-graph attention, MGA-GNN), 并以此分别构建对齐模型和知识模型, MGA-GNN 能将对齐模型和预训练语言模型的表征结合, 迭代生成新的对齐实体, 使用知识模型将不同来源知识图谱的信息聚合到目标知识图谱上, 提升 MKGC 的性能。

1　本文方法

1.1　预备知识和相关符号

知识图谱可以用$G=(E,R,T)$表示, 其中, E表示实体的集合, R表示实体与实体之间的关系集合, T表示实体与关系组成的三元组集合。每条知识用一个三元组$(h,r,t)\in T$表示, 其中, $h\in E$表示头实体, $r\in R$表示关系, $t\in E$表示尾实体。

KGC 任务的目标是基于当前知识图谱中存在的事实来预测出新的事实三元组。具体而言, 给定$(h,r,?)$或$(?,r,t)$这样的问题, 问号表示缺失的实体, 模型需要推理出缺失的尾实体或头实体。

对于 MKGC 任务, 可以利用多个使用不同语言构建的知识图谱。假设有N个知识图谱$G_1,G_2,\cdots,G_N$, 每个由不同的语言进行构建。在有已知对齐实体的场景下, 任意 2 个知识图谱$G_i=(E_i,R_i,T_i)$和$G_j=(E_j,R_j,T_j)$之间存在一些已知的对齐实体对$\{(\boldsymbol{e}_i,\boldsymbol{e}_j):\boldsymbol{e}_i\in E_i,\boldsymbol{e}_j\in E_j\}$, 其中, $\boldsymbol{e}$表示实体。对于本文所关注的无对齐实体场景, 无法获取真实的对齐实体对。另外, 对于这 2 类场景有一个关于关系集合的约束, 即所有知识图谱中的关系集合都属于一个公共的关系集合R。该约束表示每个知识图谱中的关系集合与语言无关, 即对于任意的$i\in 1,2,\cdots,N$, $R_i\in R$。

1.2　整体框架

目前已有的 MKGC 框架, 如 KEnS[6] 和 SS-AGA[8], 只能用于有对齐实体场景下的 MKGC, 为了实现无对齐实体场景下的 MKGC 任务, 本文构建了无对齐实体的 MKGC 框架, 整体结构如图 1 所示。该框架主要包含 2 个部分: 对齐实体获取模块和 GNN 编码模块。

1.3　对齐实体获取模块

对于 MKGC, 对齐实体起着连接不同知识图谱之间知识的桥梁作用, 但在实际场景中通常无法直接获取真实的对齐实体, 因此, 本文首先构建一个对齐实体获取模块, 目标是寻找出不同知识图谱之间的对齐实体, 进而为后续的知识迁移服务。知识图谱之间实体对齐的任务已经得到广泛研究[17-18], 但是常需要一定数量的真实对齐实体对作为监督信号来训练模型, 这不符合本文研究的场景。为此, 本文提出一个简单而有效的初始对齐实体获取方法。该方法包括 2 个阶段: 对齐实体初始获取和对齐实体迭代获取。

1.3.1　对齐实体初始获取

在任务初始阶段, 无预先给定的对齐实体, 而对于给定的多个不同语言的知识图谱, 每个实体的文字名称和文字描述是容易获取的, 将每个实体的这些文字信息经过预训练语言模型 mBERT[19] 进行编码, 得到实体的文本特征。mBERT[19] 是一个多语言预训练语言模型, 能够捕捉实体文本中丰富的语义信息。随后, 对于任意 2 个知识图谱中的实体$\boldsymbol{e}_i\in E_i$和$\boldsymbol{e}_j\in E_j$, 按照式(1)和式(2)计算它们之间的跨域相似性局部缩放(cross-domain similarity local scaling, CSLS)[20] 距离$\mathrm{CSLS}(\boldsymbol{e}_i,\boldsymbol{e}_j)$, 如果两者在 CSLS 距离上互为最近邻, 说明其语义信息相近, 即很可能是对齐实体, 故将这 2 个实体添加进伪对齐实体对集合中(后文方法获取的对齐实体均指伪对齐实体)。

$$\mathrm{CSLS}(\boldsymbol{e}_i,\boldsymbol{e}_j)=2\cos(\boldsymbol{e}_i^{\mathrm{t}},\boldsymbol{e}_j^{\mathrm{t}})-r_j(\boldsymbol{e}_i)-r_i(\boldsymbol{e}_j)\tag{1}$$

$$r_j(\boldsymbol{e}_i)=\frac{1}{K}\sum_{\boldsymbol{e}_j\in N_j(\boldsymbol{e}_i)}\cos(\boldsymbol{e}_i^{\mathrm{t}},\boldsymbol{e}_j^{\mathrm{t}})\tag{2}$$

式中: $\boldsymbol{e}_i^{\mathrm{t}}$和$\boldsymbol{e}_j^{\mathrm{t}}$分别为$\boldsymbol{e}_i$和$\boldsymbol{e}_j$对应的文本特征表示; $N_j(\boldsymbol{e}_i)$表示在向量空间中使用实体$\boldsymbol{e}_i$的文本特征向量$\boldsymbol{e}_i^{\mathrm{t}}$与知识图谱$G_j$中所有实体的文本特征向量计算距离, 选取$K$个最近邻的实体组成的集合; K为

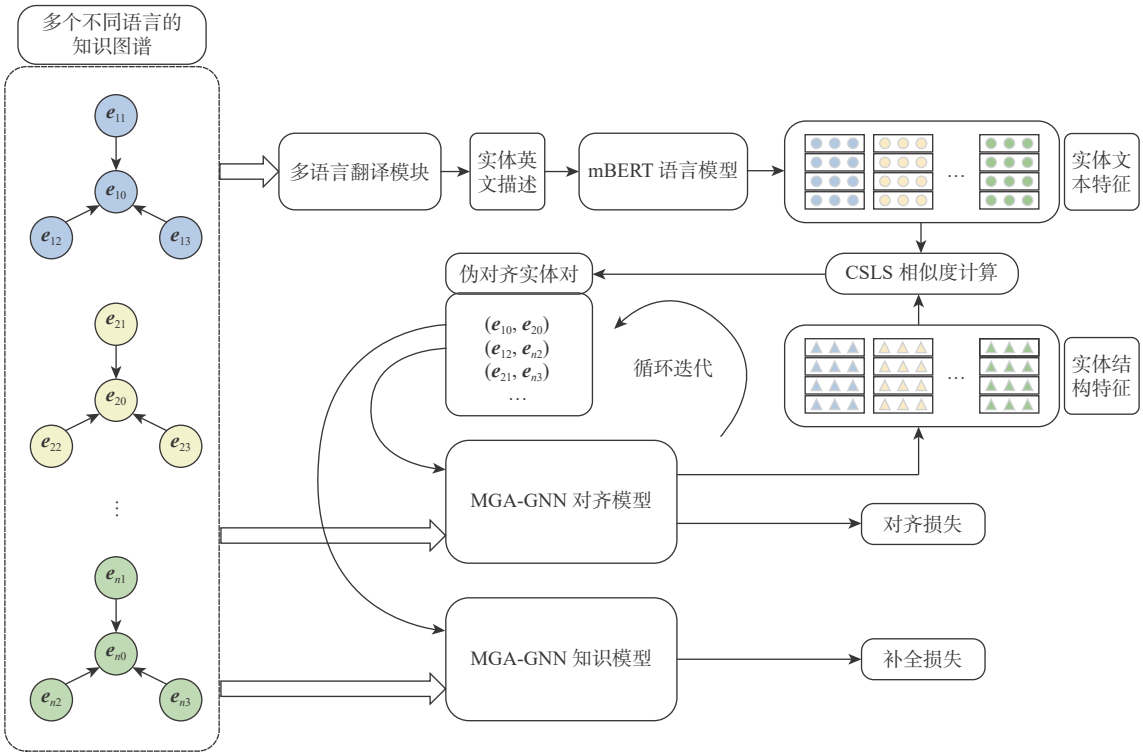


图 1 无对齐实体的多语言知识图谱补全框架

Fig. 1 The framework of multilingual knowledge graph completion without alignment pairs

$N_j(\boldsymbol{e}_i)$中实体的数量，作为超参数进行设置。

虽然经过 mBERT 编码得到的实体向量在不同语言之间已经有一定的对齐，但不足以生成足够的高质量对齐实体，因此，本文使用一个高质量的多语言翻译模块将不同语言知识图谱内实体的文本信息翻译为英文的表述，再将该英文表述经过 mBERT 编码。这里翻译到英文的原因有 2 点：①相比于翻译到其他语言，目前常见的多语言翻译模块翻译到英文的性能更高；②由于训练语料的原因，mBERT 对英文的表征能力更强。通过上述流程，可以大大提高对齐实体初始获取的质量和数量。

1.3.2 对齐实体迭代获取

初始阶段获取的对齐实体数量较少，因此，在模型训练阶段使用一个对齐模型不断迭代生成新的对齐实体。对齐模型的具体结构将在 1.4 节进行详细说明，实体$\boldsymbol{e}_i$和$\boldsymbol{e}_j$经过对齐模型编码后得到对应的结构特征表示记为$\boldsymbol{e}_i^{\mathrm{s}}$和$\boldsymbol{e}_j^{\mathrm{s}}$。随后将文本特征与结构特征相结合，按照式(3)和式(4)计算 2 个知识图谱之间所有实体之间的 CSLS 距离，如果两者在 CSLS 距离上互为最近邻，则将这 2 个实体添加进对齐实体对集合中。

$$\mathrm{CSLS}(\boldsymbol{e}_i,\boldsymbol{e}_j)=2\max(\cos(\boldsymbol{e}_i^{\mathrm{s}},\boldsymbol{e}_j^{\mathrm{s}}),\cos(\boldsymbol{e}_i^{\mathrm{t}},\boldsymbol{e}_j^{\mathrm{t}}))-r_j(\boldsymbol{e}_i)-r_i(\boldsymbol{e}_j) \tag{3}$$

$$r_j(\boldsymbol{e}_i)=\frac{1}{K}\sum_{\boldsymbol{e}_j\in N_j(\boldsymbol{e}_i)}\max(\cos(\boldsymbol{e}_i^{\mathrm{s}},\boldsymbol{e}_j^{\mathrm{s}}),\cos(\boldsymbol{e}_i^{\mathrm{t}},\boldsymbol{e}_j^{\mathrm{t}})) \tag{4}$$

1.4 图神经网络编码模块

GNN 对图结构数据具有强大的编码表征能力，为了更好实现多个知识图谱之间知识互补的能力，本文设计了 MGA-GNN。基于该网络，构建了一个用于实体对齐的模型和一个用于 MKGC 的模型。网络构建分为 2 个步骤：联合建图和 GNN 编码。

1.4.1 联合建图

给定的N个知识图谱为独立不连通的图，无法直接通过 GNN 进行有效的信息传播。本文利用 1.3.1 节中初始获取的对齐实体将N个独立的图连接为一个统一的图。具体而言，在关系集合中增加一个对齐关系，将对齐实体对$(\boldsymbol{e}_i,\boldsymbol{e}_j)$转换为三元组$(\boldsymbol{e}_i,\text{对齐},\boldsymbol{e}_j)$，通过增加的这些三元组将所有知识图谱连接起来，得到大图G。在该统一的大图基础上，对所有知识图谱中的实体和关系在同一个向量空间中进行初始化，得到初始的实体向量表示$\boldsymbol{h}^0$和关系向量表示$\boldsymbol{r}'$，特征维度为d。

1.4.2 GNN 编码

建立图G和初始化实体关系向量后，采用多层 GNN 对实体和关系进行编码。从第l层网络到第$l+1$层网络的传递过程如下：

$$\boldsymbol{h}_i^{l+1}=\boldsymbol{h}_i^{l}+\sigma\left(\sum_{\boldsymbol{e}_j\in N(\boldsymbol{e}_i)}\mathrm{Att}(\boldsymbol{h}_i^{l},\boldsymbol{h}_{j(r)}^{l})\boldsymbol{h}_{j(r)}^{l}\right) \tag{5}$$

式中：$\boldsymbol{h}_i^{l}$和$\boldsymbol{h}_i^{l+1}$分别为中心实体$\boldsymbol{e}_i$在第l层和第$l+1$网络的隐层表示；$\sigma(\cdot)$为 GELU[21] 实现的非线性激

活函数; $N(\boldsymbol{e}_i)$为实体$\boldsymbol{e}_i$的邻域实体集合; $\boldsymbol{e}_j$为实体$\boldsymbol{e}_i$的邻居实体; $\boldsymbol{h}^l_{j(r)}$为实体$\boldsymbol{e}_j$经过与关系r关联之后的隐层表示; $\mathrm{Att}(\boldsymbol{h}^l_i,\boldsymbol{h}^l_{j(r)})$为注意力函数, 用于计算中心实体对邻居实体的注意力权重。

在传统的基于查询-键-值的图注意力网络中, 邻居节点对中心节点的贡献采用同一套查询-键-值的变换矩阵。然而对于 MKGC, 不同语言的知识图谱对于目标知识图谱的知识补充能力应该是不同的, 因此, 本文提出一种多图注意力机制, 将中心实体的邻居实体按照不同来源采取不同的注意力权重计算方式, 用于区分不同来源知识图谱的信息对中心实体的贡献程度。具体而言, 对于每种语言的知识图谱, 各自维护不同的查询-键-值变换矩阵 $\boldsymbol{W}_{\mathrm{q}}$、$\boldsymbol{W}_{\mathrm{k}}$、$\boldsymbol{W}_{\mathrm{v}}$, 随后, 注意力权重计算方式如下:

$$\mathrm{Att}(\boldsymbol{h}^l_i,\boldsymbol{h}^l_{j(r)})=\frac{\exp(\alpha^r_{ij})}{\sum\limits_{\boldsymbol{e}_j\in N(\boldsymbol{e}_i)}\exp(\alpha^r_{ij})} \tag{6}$$

$$\alpha^r_{ij}=\frac{\lambda\beta_r}{\sqrt{d}}(\boldsymbol{W}_{\mathrm{q}i}\boldsymbol{h}^l_i)^{\mathrm{T}}(\boldsymbol{W}_{\mathrm{k}j}\boldsymbol{h}^l_{j(r)}) \tag{7}$$

$$\boldsymbol{h}^l_{j(r)}=\boldsymbol{W}_{\mathrm{v}j}[\boldsymbol{h}^l_j,\boldsymbol{r}'] \tag{8}$$

式中: $\boldsymbol{W}_{\mathrm{q}i}\in\mathbf{R}^{d\times d}$、$\boldsymbol{W}_{\mathrm{k}j}\in\mathbf{R}^{d\times d}$、$\boldsymbol{W}_{\mathrm{v}j}\in\mathbf{R}^{d\times 2d}$分别为实体$\boldsymbol{e}_i$所属知识图谱对应的查询变换矩阵、实体$\boldsymbol{e}_j$所属知识图谱对应的键变换矩阵、值变换矩阵; β_r为可学习的、与关系$\boldsymbol{r}'$相对应的关系变量, 用于衡量关系$\boldsymbol{r}'$的权重; λ用于权衡通过前述方法获取的伪对齐实体的可靠性, 当关系$\boldsymbol{r}'$为对齐关系时, $\lambda=\max(\cos(\boldsymbol{e}^{\mathrm{s}}_i,\boldsymbol{e}^{\mathrm{s}}_j),\cos(\boldsymbol{e}^{\mathrm{t}}_i,\boldsymbol{e}^{\mathrm{t}}_j))$, 否则为 1; $[\cdot]$为向量拼接函数。

由于 1.3 节中得到的对齐实体存在噪声问题, 直接将伪对齐实体的信息传递给中心实体可能会造成误差传递, 为了缓解这类噪声, 在式(7)中引入对齐实体的可靠性参数λ。对于生成的伪对齐实体, 两者的相似度越高, 则认为该实体对对齐的可靠性更高, 进而通过该对齐关系传递给中心实体的信息应该更加可靠, 反之, 则应该抑制该信息传递。

基于上述 MGA-GNN, 本文采取同样的结构分别建立实体对齐模型和知识模型。实体对齐模型用于拉进对齐实体之间的距离, 进而通过训练迭代生成新的对齐实体, 并将新的对齐实体添加进大图G中, 不断提高多个知识图谱之间知识互补的能力; 知识模型用于对所有实体和关系进行编码, 以 KGC 为目标进行训练, 最终预测三元组中缺失的实体。

1.5　模型训练

对齐模型训练的目标是拉近对齐实体之间的距离, 故对于任意 2 个知识图谱G_i和G_j, 设已经生成的它们之间的对齐实体集合为P_{ij}, 则对齐损失函数采用如下对比损失:

$$L^{ij}_{\mathrm{A}}=\sum_{\substack{(\boldsymbol{e}_i,\boldsymbol{e}_j)\in P_{ij}\\(\boldsymbol{e}_{i'},\boldsymbol{e}_{j'})\notin P_{ij}}}\max\left(\gamma_{\mathrm{A}}+\left\|\boldsymbol{e}_i-\boldsymbol{e}_j\right\|_2-\left\|\boldsymbol{e}_{i'}-\boldsymbol{e}_{j'}\right\|_2,0\right) \tag{9}$$

式中: γ_{A}为对比损失的间隔超参数。

知识模型训练的目标是能够预测三元组中缺失的实体, 本文采用 TransE[9] 模型对三元组进行合理性打分, 在知识图谱G_i上的补全损失函数为

$$L^i_{\mathrm{K}}=\sum_{\substack{(h,r,t)\in T_i\\(h',r,t')\notin T_i}}\max\left(\gamma_{\mathrm{K}}-\phi(h,r,t)+\phi(h',r,t'),0\right) \tag{10}$$

$$\phi(h,r,t)=-\|\boldsymbol{e}_h+\boldsymbol{r}-\boldsymbol{e}_t\|_2 \tag{11}$$

式中: γ_{K}为补全损失的间隔超参数。

基于上述损失函数, 将对齐模型和知识模型进行联合训练, 总的损失函数为

$$L=\sum_{i=1}^{N}\sum_{j=i+1}^{N}L^{ij}_{\mathrm{A}}+\sum_{i=1}^{N}L^i_{\mathrm{K}} \tag{12}$$

2　实验准备

2.1　数 据 集

在 2 个数据集上进行实验以评估本文方法的性能, 数据集的统计数据数量如表 1 所示。DBP-5L[6] 数据集是从维基百科知识图谱 DBpedia[22] 抽取获得, 包括 5 个不同语言的知识图谱, 分别为希腊语(EL)、英语(EN)、西班牙语(ES)、法语(FR)和日语(JA)。E-PKG[8] 数据集是一个多语言的、与手机相关的电子产品知识图谱, 包括 6 种语言, 分别为德语(DE)、英语(EN)、西班牙语(ES)、法语(FR)、意大利语(IT)和日语(JA)。2 个数据集中的关系都用英语表示, 并且由不同语言的知识图谱共享。2 个数据集中具有最丰富知识的语言都是英语, 对

表 1　数据集统计数据数量

Table 1　The statistic numbers of datasets

数据集	语言	实体数	关系数	训练数据量	验证数据量	测试数据量
DBP-5L[6]	EL	5 231	111	8 670	4 152	1 017
	EN	13 996	831	48 652	24 051	7 464
	ES	12 382	144	33 036	16 220	4 810
	FR	13 176	178	30 139	14 705	4 171
	JA	11 805	128	17 979	8 633	2 162
E-PKG[8]	DE	17 223	21	45 515	22 753	7 602
	EN	16 544	21	60 310	39 150	10 071
	ES	9 595	21	18 090	9 039	3 034
	FR	17 068	21	47 999	23 994	8 022
	IT	15 670	21	42 767	21 377	7 148
	JA	2 642	21	10 013	5 002	1 688

应知识最少的语言分别为希腊语和日语。需要注意的是，DBP-5L[6]数据集给出了实体对应的名称，而E-PKG[8]数据集中的实体没有实体名称，只有实体相应的具体文字描述。

2.2 基线模型

为了验证本文方法的有效性，选取2类基线模型进行对比，分别为单语言KGC模型和有对齐实体的MKGC模型。对于单语言模型，选取经典模型TransE[9]和RotatE[11]，以及采用预训练语言模型的KG-BERT[23]。对于多语言模型，KEnS[6]首次提出了MKGC方法，并设计了一种集成方法来进行知识迁移；AlignKGC[7]将MKGC同实体和关系对齐进行联合学习；SS-AGA[8]利用已有对齐实体进行实体对齐的自监督学习，进而提升MKGC性能。

2.3 评价指标

通过尾实体预测任务来对模型性能进行评估。首先，对于测试集中的三元组(h, r, t)，给定h和r，将其他所有实体作为候选实体计算三元组合理性得分，根据该得分进行降序排序，以此预测尾实体t。然后，根据该排序计算得到评价指标Hits@1、Hits@10和平均倒数排名（mean reciprocal ranks, MRR）。其中，Hits@1和Hits@10分别表示正确实体在该排序中排名为1和排在前10的比例，MRR表示所有测试样例排名倒数的平均值。另外，在测试阶段，沿用文献[9]中的过滤设置，即将训练集、验证集和测试集中已知三元组的得分从排序中删掉。

2.4 参数设置

实验中使用的翻译模块为谷歌翻译，另外，采用网格搜索法对超参数进行调优，各个超参数的搜索范围为：实体和关系嵌入维度{128,256}，网络层数{2,3}，间隔参数{0.2,0.3,0.5}，学习率{0.000 1,0.001,0.005,0.01}。最终参数设置为：实体和关系嵌入维度为256，网络层数为2，学习率为0.01，在DBP-5L数据集上间隔参数为0.5，在E-PKG上间隔参数为0.3，训练的批量大小为200，使用Adam优化器对网络参数进行优化。

3 实验分析

3.1 实验结果

在数据集DBP-5L和E-PKG上的实验结果分别如表2和表3所示。可以看到，MKGC方法的结果普遍优于单语言KGC方法，这验证了不同语言知识图谱之间可以进行知识互补。可以发现，在DBP-5L数据集上，本文方法在4个语言的知识图谱上取得了最好性能，在E-PKG数据集上也获得了

表2 DBP-5L数据集实验结果

Table 2 Experimental results on the DBP-5L dataset

方法分类	方法	语言	H@1	H@10	MRR
单语言KGC方法	TransE[9]	EL	13.1	43.7	24.3
		EN	7.3	29.3	16.9
		ES	13.5	45.0	24.4
		FR	17.5	48.8	27.6
		JA	21.1	48.5	25.3
	RotatE[11]	EL	14.5	36.2	26.2
		EN	12.3	30.4	20.7
		ES	21.2	53.9	33.8
		FR	23.2	55.5	35.1
		JA	26.4	60.2	39.8
	KG-BERT[23]	EL	17.3	40.1	27.3
		EN	12.9	31.9	21.0
		ES	21.9	54.1	34.0
		FR	23.5	55.9	35.4
		JA	26.9	59.8	38.7
有对齐实体的MKGC方法	KEnS[6]	EL	28.1	56.9	
		EN	15.1	39.8	
		ES	23.6	60.1	
		FR	25.5	62.9	
		JA	32.1	65.3	
	AlignKGC[7]	EL	27.6	56.2	33.8
		EN	15.5	39.2	22.3
		ES	24.2	60.9	35.1
		FR	24.1	62.3	37.4
		JA	31.6	64.3	41.6
	SS-AGA[8]	EL	30.8	58.6	35.3
		EN	16.3	41.3	23.1
		ES	25.5	61.9	36.6
		FR	27.1	65.5	38.3
		JA	**34.6**	**66.9**	**42.9**
无对齐实体的MKGC方法	MGA-GNN	EL	**33.6**	**77.1**	**49.3**
		EN	**19.5**	**59.8**	**33.7**
		ES	**26.1**	**64.7**	**39.3**
		FR	**31.2**	**70.5**	**44.6**
		JA	28.5	60.6	39.8

注：加粗数值表示性能最优。

其中4个语言的知识图谱上最高的MRR和其中5个语言的知识图谱上最高的Hits@1和Hits@10。这主要得益于2点：①对齐实体生成模块提供了足够的对齐实体，使在目标知识图谱上进行补全推理时，可以通过这些对齐实体获取其他知识图谱中的知识；②提出的MGA-GNN充分考虑了邻域实体的来源多样性，而不是按照单一语言进行建模。

3.2 实体对齐分析

DBP-5L和E-PKG数据集本身是给有对齐实体场景下的MKGC建立的，因此，数据集提供了一部

表 3　E-PKG 数据集实验结果

Table 3　Experimental results on the E-PKG dataset

方法分类	方法	语言	H@1	H@10	MRR
单语言KGC方法	TransE[9]	DE	21.2	65.5	37.4
		EN	23.2	67.5	39.4
		ES	17.2	58.4	33.0
		FR	20.8	66.9	37.5
		IT	22.0	63.8	37.8
		JA	25.1	72.7	43.6
	RotatE[11]	DE	22.3	64.3	38.2
		EN	24.2	66.8	40.0
		ES	18.3	58.9	33.7
		FR	22.1	64.3	38.2
		IT	22.5	64.0	38.1
		JA	26.3	71.9	41.8
	KG-BERT[23]	DE	21.8	64.7	38.4
		EN	24.3	66.4	39.6
		ES	18.7	58.8	33.2
		FR	22.3	67.2	38.3
		IT	22.9	63.7	37.2
		JA	26.9	72.4	44.1
有对齐实体的MKGC方法	KEnS[6]	DE	24.3	65.8	
		EN	26.2	69.5	
		ES	21.3	59.5	
		FR	25.4	68.2	
		IT	25.1	64.6	
		JA	33.5	73.6	
	AlignKGC[7]	DE	22.1	65.1	38.5
		EN	25.6	68.3	40.5
		ES	19.4	59.1	34.2
		FR	22.8	67.2	38.8
		IT	24.2	63.4	37.3
		JA	31.2	72.3	46.2
	SS-AGA[8]	DE	24.6	66.3	39.4
		EN	26.7	69.8	41.5
		ES	21.0	60.1	**36.3**
		FR	**25.9**	68.7	**40.2**
		IT	24.9	63.8	38.4
		JA	33.9	**74.1**	48.3
无对齐实体的MKGC方法	MGA-GNN	DE	**24.9**	**68.4**	**40.3**
		EN	**29.3**	**70.4**	**44.4**
		ES	**22.2**	**60.3**	35.6
		FR	24.1	**68.9**	39.6
		IT	**26.1**	**65.9**	**40.9**
		JA	**43.0**	73.2	**54.5**

注: 加粗数值表示性能最优。

分已对齐的实体对, 为了验证本文对齐实体获取模块的有效性, 表 4 和表 5 给出了最终通过该模块获取的对齐实体对数据集中给定的对齐实体的覆盖程度。精确率表示生成的对齐实体对在已给对齐实体对中的正确率, 召回率表示生成的对齐实体对对已给对齐实体对的召回比例。

表 4　DBP-5L 数据集中生成实体精确率和召回率

Table 4　The precision and recall of the generated aligned entities on the DBP-5L dataset　%

语言对	精确率	召回率
EL-JA	97.1	25.4
EL-ES	98.6	68.4
EL-EN	99.2	66.8
EL-FR	99.0	64.6
JA-ES	97.6	22.1
JA-EN	98.0	25.3
JA-FR	98.6	21.2
ES-EN	99.7	74.5
ES-FR	99.6	70.7
EN-FR	99.7	71.1

表 5　E-PKG 数据集中生成实体精确率和召回率

Table 5　The precision and recall of the generated aligned entities on the E-PKG dataset　%

语言对	精确率	召回率
DE-EN	58.5	24.5
DE-ES	51.9	21.0
DE-FR	56.1	21.4
DE-IT	55.6	22.2
DE-JA	60.9	16.3
ES-EN	51.7	23.5
ES-FR	50.1	22.8
ES-IT	55.0	25.2
ES-JA	50.6	16.0
FR-EN	56.1	23.9
FR-IT	58.7	25.2
FR-JA	58.6	16.6
IT-EN	60.3	24.3
IT-JA	61.6	15.8
JA-EN	42.8	17.7

可以发现, DBP-5L 数据集中的对齐精确率都较高, 这表明对齐实体模块获取的对齐实体较为精确, 为不同语言知识图谱之间的知识迁移起了很好的支撑作用。另外可以发现, DBP-5L 数据集中有关日语知识图谱的实体对齐的召回率都比较低, 均不超过 30%, 这是因为从日语翻译到英语的准确率不如其他 3 种语言(希腊语、西班牙语、法语), 导致获取的对齐实体数量不够, 这也解释了在 DBP-5L 数据集上本文方法在日语知识图谱上补全结果不如有对齐实体的方法。

对于 E-PKG 数据集, 从表 5 可以发现, 对齐实体的精确率均在 55% 左右, 远低于 DBP-5L 数据集

中的结果，这是因为E-PKG数据集中没有实体的文本名称，只有对于实体的文本描述，而不同语言中对相同实体的描述侧重不同，引入的语言差异也给翻译模块和文本表征模型带来了困难，导致精确率较低，同样召回率也不高。本文方法在多个知识图谱上的补全结果仍超越了其它模型，证明了本文方法的有效性。

4 结 论

1）本文提出了一种由伪对齐实体生成与多语言GNN编码组成的双模块框架，有效解决了多语言知识图谱补全中缺失实体对齐的关键瓶颈问题。

2）通过融合预训练语言模型与迭代式对齐扩展机制，实现了无需种子对齐实体的跨语言知识自动迁移，提升了方法在实际场景中的适用性。设计的多图注意力编码机制能够区分不同语言知识图谱的信息贡献差异，实现了跨语言知识的选择性融合与针对性表征。

3）在完全无对齐实体的实验设定下，本文方法不仅实现了有效的知识迁移，其性能甚至超越了依赖已知对齐实体的传统多语言知识图谱补全方法，为跨语言知识融合提供了不依赖预设对齐资源的新思路。

参考文献（References）

[1] REN H Y, DAI H J, DAI B, et al. LEGO: latent execution-guided reasoning for multi-hop question answering on knowledge graphs [C]//Proceedings of the 38th International Conference on Machine Learning. San Diego: PMLR, 2021, 139: 8959-8970.

[2] SAXENA A, KOCHSIEK A, GEMULLA R. Sequence-to-sequence knowledge graph completion and question answering[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2022: 2814-2828.

[3] WANG X, HUANG T L, WANG D X, et al. Learning intents behind interactions with knowledge graph for recommendation[C]// Proceedings of the Web Conference 2021. New York: ACM, 2021: 878-887.

[4] WANG X T, LIU K P, WANG D J, et al. Multi-level recommendation reasoning over knowledge graphs with reinforcement learning [C]//Proceedings of the ACM Web Conference 2022. New York: ACM, 2022: 2098-2108.

[5] LIN Q K, LIU J, XU F Z, et al. Incorporating context graph with logical reasoning for inductive relation prediction[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 893-903.

[6] CHEN X L, CHEN M H, FAN C J, et al. Multilingual knowledge graph completion via ensemble knowledge transfer[C]//Findings of the Association for Computational Linguistics: EMNLP 2020. Stroudsburg: ACL, 2020: 3227-3238.

[7] SINGH H, JAIN P, MAUSAM, et al. Multilingual knowledge graph completion with joint relation and entity alignment[EB/OL]. (2021-04-18)[2023-10-29]. https://arxiv.org/abs/2104.08804.

[8] HUANG Z J, LI Z, JIANG H M, et al. Multilingual knowledge graph completion with self-supervised adaptive graph alignment[C]// Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2022: 474-485.

[9] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 2787-2795.

[10] YANG B S, YIH W T, HE X D, et al. Embedding entities and relations for learning and inference in knowledge bases[C]//Proceedings of the International Conference on Learning Representations. Phoenix: ICLR, 2014: 2768038.

[11] SUN Z Q, DENG Z H, NIE J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space[EB/OL]. (2019-02-26)[2023-10-29]. https://arxiv.org/abs/1902.10197.

[12] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the Semantic Web. Berlin: Springer, 2018: 593-607.

[13] ZHANG Y Q, YAO Q M. Knowledge graph reasoning with relational digraph[C]//Proceedings of the ACM Web Conference. New York: ACM, 2022: 912-924.

[14] CHEN M H, TIAN Y T, YANG M H, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[C]// Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York: ACM, 2017: 1511-1517.

[15] GUO L B, ZHANG Q, SUN Z Q, et al. Understanding and improving knowledge graph embedding for entity alignment[C]//Proceedings of the 39th International Conference on Machine Learning. San Diego: PMLR, 2022, 162: 8145-8156.

[16] ZHU Y, LIU H Z, WU Z H, et al. Relation-aware neighborhood matching model for entity alignment[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(5): 4749-4756.

[17] SUN Z Q, WANG C M, HU W, et al. Knowledge graph alignment network with gated multi-hop neighborhood aggregation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 222-229.

[18] YAN Y C, LIU L H, BAN Y K, et al. Dynamic knowledge graph alignment[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(5): 4564-4572.

[19] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. (2019-05-24)[2023-10-29]. https://arxiv.org/abs/1810.04805.

[20] CONNEAU A, LAMPLE G, RANZATO M, et al. Word translation without parallel data[EB/OL]. (2018-01-30)[2023-10-29]. https://arxiv.org/abs/1710.04087.

[21] HENDRYCKS D, GIMPEL K. Gaussian error linear units (GELUs) [EB/OL]. (2023-06-06)[2023-10-29]. https://arxiv.org/abs/1606.08415.

[22] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia[J]. Semantic

Web, 2015, 6(2): 167-195.

[23] YAO L, MAO C S, LUO Y. KG-BERT: BERT for knowledge graph completion[EB/OL]. (2019-09-11)[2023-10-29]. https://arxiv.org/abs/1909.03193.

Multilingual knowledge graph completion without aligned entity pairs

TANG Rongchuan[1, 2], XU Qiucheng[2], TANG Wenyi[2], ZHAI Feifei[1, 3], ZHOU Yu[1, 3, *]

(1. Institute of Automation, Chinese Academy of Science, Beijing 100190, China;
2. State Key Laboratory of Air Traffic Management System, College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210007, China;
3. Beijing Fanyu Technology Co., Ltd., Beijing 100190, China)

Abstract: The goal of multilingual knowledge graph completion (MKGC) is to improve link prediction performance on the target knowledge graph by leveraging data from other language-specific knowledge graphs. Existing methods usually use pre-aligned entities between different knowledge graphs to accomplish knowledge transfer. However, there are usually no pre-aligned entities between different knowledge graphs in practical scenarios, making knowledge transfer difficult to achieve. Considering the MKGC without aligned entity pairs, a pseudo-aligned entity generation module that integrates a pre-trained language model is proposed to iteratively generate new aligned entities for knowledge transfer. It is suggested to use a graph neural network based on multi-graph attention (MGA-GNN) to encode the triples in order to differentiate the information in various language-specific wisdom graphs. Finally, the plausibility of the triples is calculated via the embeddings output by the network to conduct the link prediction task. Experimental results on the DBP-5L and E-PKG datasets show the effectiveness of the proposed method and its superior performance in more practical scenarios.

Keywords: multilingual knowledge graph completion; entity alignment; multi-graph attention; graph neural network; link prediction

Received: 2023-10-31; **Accepted**: 2023-12-29; **Published Online**: 2024-03-12 11:05

URL: link.cnki.net/urlid/11.2625.V.20240311.1607.006

Foundation item: Open Research Projects of State Key Laboratory of Air Traffic Management System (SKLATM202203)

* **Corresponding author**. E-mail: yzhou@nlpr.ia.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0739

航空发动机盲腔螺母精密拧紧技术

张宗江[1]，李小强[1,*]，韩玉杰[2]，李兆宇[1]，张铭铭[1]

(1. 北京航空航天大学 机械工程及自动化学院，北京 100191；2. 北京航天长征飞行器研究所，北京 100076)

摘 要：航空发动机转子狭小盲腔连接质量对转子整体装配质量具有极大的影响，综合国内外现状，由于缺乏盲腔结构高效精准装配工装，导致螺栓连接工艺的精确控制能力不足，进而影响转子盘轴零部件配合表面的连接力学性能，针对某型号盲腔螺母拧紧装配需求，将盲腔螺母拧紧系统分为拧紧力矩传递机构和操作执行机构。在拧紧力矩传递机构设计上，对拧紧工艺扭矩法与扭矩转角法进行比较，确定扭矩转角法作为最终执行工艺，同时，设计力矩传递结构，并对关键部件进行强度和寿命校核，考虑力矩损失建立输入输出力矩模型，确定拧紧工具输入扭矩，实现拧紧力矩的高精度控制；在操作执行机构设计上，对操作执行机构进行机械设计，并对高精度运动部件蜗轮蜗杆进行3DCS仿真，同时，利用MATLAB进行轨迹规划与运动学仿真；基于以上研究开发感知系统及控制系统，进行视觉识别，实现设备的高精度控制，并进行盲腔螺母拧紧技术实验验证，对提高该狭小盘间盲腔螺栓阵列连接质量一致性具有重要意义。

关 键 词：航空发动机；盲腔；扭矩转角法；力矩控制；轨迹规划；偏差分析

中图分类号：V263.2

文献标志码：A **文章编号**：1001-5965(2026)01-0260-15

航空发动机高压压气机是发动机的核心部件，由许多级结构类似的转子零件首尾连接而成[1]。压气机转子盘腔由于轴向结构复杂紧凑，径向结构阶梯起伏，部分连接部位装配开敞性差，容易形成装配空间狭小、目视不可直达的半封闭腔体结构[2-3]（以下简称“盲腔”），且盲腔部位螺栓阵列装配需要满足力矩精度和拧紧顺序要求，导致人工装配难度大、紧固精度差、效率低且连接质量一致性较差，是影响发动机转子单元盘轴零部件结构变形、紧固程度和同心度保持能力的关键因素[4-5]。

近年来，由于航空发动机产业对数字化、自动化的航空发动机盲腔结构高效精准装配工艺和设备的迫切需要[6]，引领了精密装配技术、可视化引导技术、智能防错技术等先进装配技术的快速发展[7]。

GE Aviation、Pratt&Whitney、Rolls Royce、MTU等著名的航空发动机制造企业非常重视装配工艺技术的研发，通过长期积累，已经构建了相对完整的装配工艺技术体系[8-9]。同时，中国商发[10-11]、大连理工大学[12-13]、沈阳航空航天大学[14]等单位针对民用涡扇发动机高压压气机转子盘间盲腔结构的装配，在紧固工艺等方面进行了探索研究[10,12-13,15]。但由于航空发动机高压压气机转子盲腔结构的装配对紧固工艺、关键技术及拧紧设备有很高的要求，给研究带来很大的难度，目前，国内的研究周期较短，成果积累有限，整体还处于探索阶段，尚无可以投入实际应用的技术和设备[16-18]。

基于盲腔螺母拧紧技术发展现状，本文研究某型号发动机转子组件狭小盘间盲腔螺母拧紧技术，用以稳定准确地执行装配工艺，精密控制每颗螺栓的力矩转角精度，对提高该狭小盘间盲腔螺栓阵列

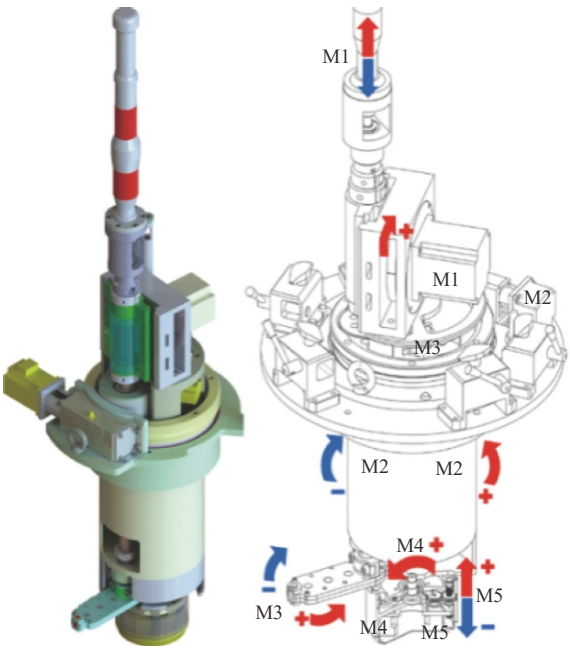
收稿日期：2023-11-09；录用日期：2023-12-01；网络出版时间：2023-12-12 10:15
网络出版地址：link.cnki.net/urlid/11.2625.V.20231211.1143.002
基金项目：国家科技重大专项 (2022-V11-0001-0043)
*通信作者. E-mail：lixiaoqiang@buaa.edu.cn

引用格式：张宗江，李小强，韩玉杰，等. 航空发动机盲腔螺母精密拧紧技术 [J]. 北京航空航天大学学报，2026，52（1）：260-274.
ZHANG Z J，LI X Q，HAN Y J，et al. Precision tightening technology for aeroengine blind cavity nuts [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：260-274（in Chinese）.

连接质量一致性具有重要意义[19-21]。

1　盲腔螺母拧紧设备总体系统设计

盲腔螺母拧紧设备总体可分为硬件和软件 2 个部分。硬件部分主要运动机构如图 1 所示。



M1. 凸轮驱动电机实现拧紧头升降运动; M2. 设备周向转动电机实现装置相对于固连基座发生转动; M3. 拧紧头转动电机实现拧紧头的展开与收缩运动; M4. 螺母盘转动电机实现螺母的分度供给; M5. 螺母顶起电机实现螺母的自动装填。

图 1　设备硬件驱动示意图

Fig. 1　Device hardware driver diagram

盲腔螺母拧紧设备硬件部分总体设计思路为: 通过操作执行机构控制 L 型拧紧臂对航空发动机转子盲腔螺母进行拧靠并限力。因此, 盲腔螺母拧紧设备由 2 部分组成, 分别为拧紧力矩传递机构和操作执行机构, 如图 2 所示。拧紧力矩传递机构功能为通过 L 型拧紧臂对螺母进行高精度扭矩传递, 操作执行机构系统包括用于控制 L 型拧紧机构升降的升降机构、用于偏转 L 型拧紧机构的拧紧臂展开/收缩机构, 以及用于转动 L 型拧紧机构、升降机构和展开/收缩机构的周向转动机构, 还包括用于为 L 型拧紧机构提供螺母的螺母供给机构。

如图 3 所示, 控制系统模块分解为逻辑控制单元、信息采集处理单元、执行器检测反馈单元等。执行器检测反馈单元主要由伺服电机、视觉识别模块、力矩控制单元、报警系统等组成, 接收控制系统的命令, 进行动作的执行与反馈; 信息采集处理单元主要由传感器系统组成, 采集相机监测图像、拧紧力矩、设备角度及光电开关等信息, 实现螺栓中线检测、设备转动角度检测及设备错误检测等。

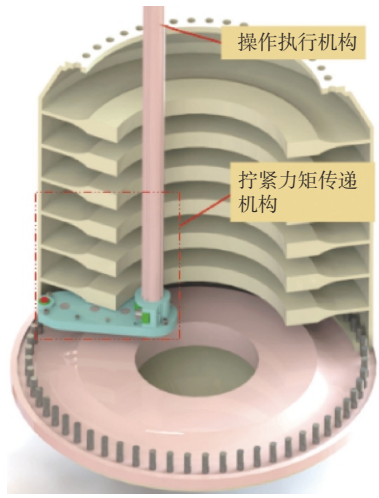


图 2　L 型拧紧臂工作状态

Fig. 2　Working status of L-shaped tightening arm

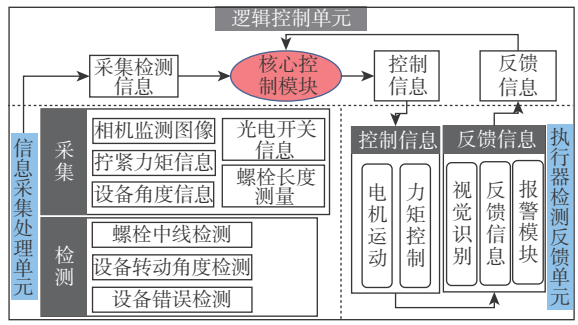


图 3　盲腔螺母拧紧控制系统逻辑控制单元示意图

Fig. 3　Schematic diagram of logic control unit for blind cavity nuts tightening control system

如图 4 所示, 系统采用基于个人计算机(personal computer, PC)的软可编程逻辑控制器(programmable logic controller, PLC)控制, 主控制器端子模块与工控机之间通过串口协议 RS233 直接相连, 网络传输速度快, 总线结构简单。运动控制系统通过数字控制(numerical control, NC)程序使伺服驱动器驱动电机动作, 接收伺服电机自带的绝对值编码器反馈位置信息, 整个运动控制系统构成闭环位置反

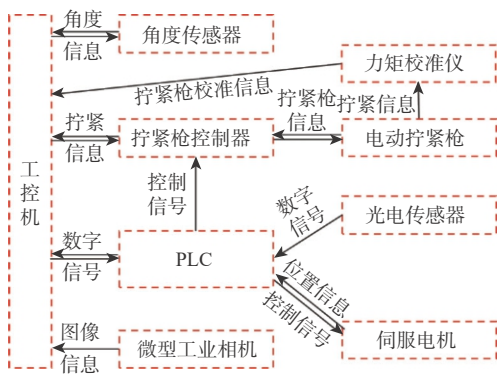


图 4　盲腔螺母拧紧系统控制原理

Fig. 4　Control schematic diagram of blind cavity nuts tightening system

馈。基于上述方案路径进行自动拧紧装置控制系统开发,可提升系统控制稳定性。该控制系统为实现控制系统功能需求,将拧紧枪控制器传输的拧紧信息、PLC传输的数字信号信息、微型工业相机传输的图像信息、角度传感器传输的角度信息均传输至计算机,由计算机进行整合并进行综合分析,与PLC、拧紧枪控制器进行数据交互,实现盲腔螺母拧紧装配控制系统的智能化控制,同时,拧紧枪根据拧紧枪控制器传输的信号进行工作,将拧紧结果反馈回拧紧枪控制器,进而传输至计算机,完成拧紧信息的可视化。同时,PLC接收来自光电传感器及伺服电机的反馈信号,传输至计算机通过软件进行综合分析,进而通过控制信号来完成对伺服电机及拧紧枪控制器的控制。

2 盲腔螺母拧紧技术拧紧力矩传递机构设计

2.1 盲腔螺母拧紧技术拧紧工艺原理

单螺栓拧紧方法一般采用扭矩法和扭矩转角法。其中,扭矩法通过控制螺栓扭矩来间接控制被连接件的预紧力值,当拧紧工具拧紧外力矩达到规定扭矩时停止拧紧。扭矩法操作简单,但影响螺栓摩擦系数的因素很多,在定扭矩条件下,预紧力会受扭矩系数(即摩擦系数)的影响,若摩擦系数不同,扭矩与预紧力的关系一般遵循5-4-1原则,即所施加的扭矩50%需要克服螺栓(螺母)旋转所产生的摩擦力,40%需要克服螺纹副的摩擦力,剩余10%则用于产生预紧力,同时,在相同输入扭矩下,摩擦系数对预紧力的影响变化范围可达30%[4],因此,采用扭矩法对螺栓进行拧紧时易造成预紧力波动量较大,进而影响装配质量。

扭矩转角法通过扭矩和转角2个物理量对螺栓预紧力进行综合控制,在拧紧时,先将螺栓旋拧到贴合扭矩,即螺栓和被连接件保持接触且稳定后,此时,转角与预紧力关系为线性,再转动螺栓(螺母)达到相应的角度,以达到精确控制预紧力的目的[21]。在扭矩转角法中,预紧力大小可以表示为

$$\begin{cases} F_{\mathrm{P}} = F_M + F_\theta \\ F_M = \dfrac{M}{K_M d} \\ K_M = \dfrac{P}{2\pi d} + \dfrac{\mu_{\mathrm{t}} r_{\mathrm{t}}}{d\cos\beta} + \dfrac{\mu_{\mathrm{b}} r_{\mathrm{b}}}{d} \\ F_\theta = \dfrac{k_{\mathrm{B}} k_{\mathrm{J}}}{k_{\mathrm{B}} + k_{\mathrm{J}}} \cdot \dfrac{P}{360}(\theta_{\mathrm{R}} - \theta_{\mathrm{B}}) \\ \theta_{\mathrm{R}} = \dfrac{T_{\mathrm{B}}}{G_{\mathrm{B}}} \end{cases} \tag{1}$$

式中:F_{P}为扭矩转角法下预紧力;F_M为扭矩转角法下旋拧贴合力矩作用下预紧力;F_θ为扭矩转角法下通过转角进行拧紧的预紧力;M为扭矩转角法下旋拧贴合力矩;K_M为扭矩转角法下旋拧螺栓力矩系数;P为螺纹螺距,mm;μ_{t}为螺纹副摩擦系数;μ_{b}为螺母支承面摩擦系数;r_{t}为螺纹副有效摩擦半径,mm;r_{b}为螺母支承面的有效摩擦半径,mm;d为齿轮分度圆直径,mm;β为螺纹轮廓半角;k_{B}和k_{J}分别为螺栓和连接机构刚度;θ_{B}为扭矩转角法下扭矩造成的螺栓扭转变形;θ_{R}为扭矩转角法下沿螺纹轴线旋转角度;T_{B}为端面摩擦力矩;G_{B}为螺栓的扭转刚度。

虽然扭矩转角法得到的预紧力也会受到摩擦系数影响,但其处在弹性变形区域时,若弹性模量恒定,预紧力可以通过拧紧转角来控制,从而降低摩擦系数对预紧力的影响。因此,扭矩转角法可以降低预紧力的分散度,提高对预紧力的控制精度,为了提高盲腔螺母拧紧设备装配精度,本文确定螺栓预紧工艺为扭矩转角法。在实现过程中,首先,通过实验确定初始扭矩值,在此扭矩值下,螺栓与被连接件实现紧密贴合,并且此时螺栓预紧力因摩擦因素所造成的误差最小;然后,根据实验结果确定实际转角与预紧力之间的函数关系,根据计算结果拟合关系因子,最终实现对螺栓预紧力的精确控制。

2.2 盲腔螺母拧紧技术力矩传递结构设计

盲腔螺母拧紧设备通过力矩传递轴与齿轮组进行力矩传递,如图5所示。拧紧枪精准控制输出的力矩,拧紧枪与力矩传递轴上端之间通过阿特拉斯转接头连接,将拧紧枪力矩传递给力矩传递轴,力矩传递轴将力矩传递给内臂输入齿轮,再通过4个齿轮传动对将力矩传递到拧紧头最前端的拧紧套筒,最终加载至螺母。拧紧头最前端设计为套筒齿轮一体化,以适应狭窄的盘间距和盘壁距,便于

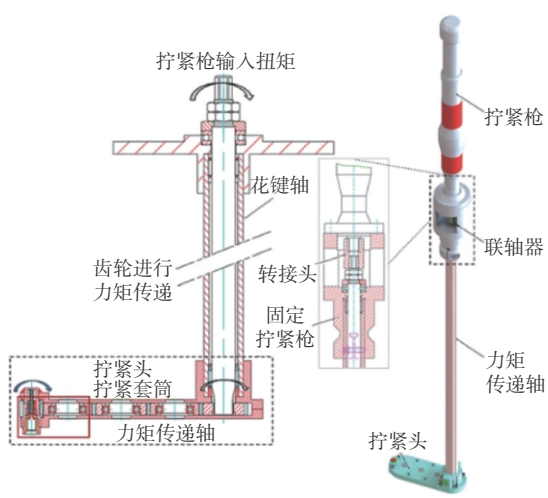


图5 力矩传递结构设计示意图

Fig. 5 Schematic diagram of torque transmission structure design

拆装更换, 拧紧套筒内有弹性卡套及弹簧结构, 弹性卡套可以锁紧螺母, 防止螺母在展开过程中掉落, 弹簧结构在装填螺母和拧紧螺母时起到缓冲作用。

综合分析拧紧头齿轮组各个齿轮的工作状态, 为确保力矩传递模块能够传递足够力矩, 并考虑齿轮的实际工作状态, 进行冲击载荷静强度校核和疲劳强度校核计算。对齿轮传递冲击载荷静强度进行校核时, 首先, 计算齿轮所受切向载荷 F_{cal}(单位: N), 如下:

$$F_{\mathrm{cal}}=\frac{2\,000T_{\max}}{d} \tag{2}$$

式中: $T_{\max}$为齿轮所承受转矩, N·m。

利用齿轮所受切向载荷计算齿轮静强度最大齿面应力及齿根静强度最大齿根弯曲应力, 如下[22]:

$$\begin{cases}\sigma_{\mathrm{Hst}}=\sqrt{K_{\mathrm{V}}K_{\mathrm{H}\beta}K_{\mathrm{H}\alpha}}Z_{\mathrm{H}}Z_{\mathrm{E}}Z_{\varepsilon}Z_{\beta}\sqrt{\dfrac{F_{\mathrm{cal}}}{db}\cdot\dfrac{u\pm1}{u}}\\ \sigma_{\mathrm{Fst}}=K_{\mathrm{V}}K_{\mathrm{F}\beta}K_{\mathrm{F}\alpha}\dfrac{F_{\mathrm{cal}}}{bm_{\mathrm{n}}}Y_{\mathrm{F}\alpha}Y_{\mathrm{S}\alpha}Y_{\varepsilon}Y_{\beta}\end{cases} \tag{3}$$

式中: σ_{Hst}为静强度最大齿面应力; σ_{Fst}为静强度最大齿根弯曲应力; K_{V}为动载系数; $K_{\mathrm{H}\beta}$和$K_{\mathrm{F}\beta}$为齿向载荷分布系数; $K_{\mathrm{H}\alpha}$和$K_{\mathrm{F}\alpha}$为齿间载荷分配系数; Z_{H}为节点区域系数; Z_{E}为弹性系数; Z_{ε}和Z_{β}分别为接触强度计算的重合度系数和螺旋角系数; u为齿轮传动比; b为齿厚, mm; m_{n}为法向模数, mm; $Y_{\mathrm{F}\alpha}$为齿轮齿形系数; $Y_{\mathrm{S}\alpha}$为应力修正系数; Y_{ε}和Y_{β}分别为抗弯强度计算的重合度系数和螺旋角系数。

同时, 为了衡量盲腔螺母拧紧设备力矩传递齿轮强度稳定性, 通过 ABAQUS 对力矩传递齿轮进行应力仿真分析, 结果如图 6 所示, 将结果文件导入 FE-SAFE 软件进行疲劳寿命分析, 其仿真原理是通过给材料施加定循环次数的重复载荷, 仿真模拟齿轮应力应变状态。仿真设置表面粗糙度为 $-16\ \mu\mathrm{m}<R_{\mathrm{a}}<40\ \mu\mathrm{m}$, 循环次数为 1×10^5。

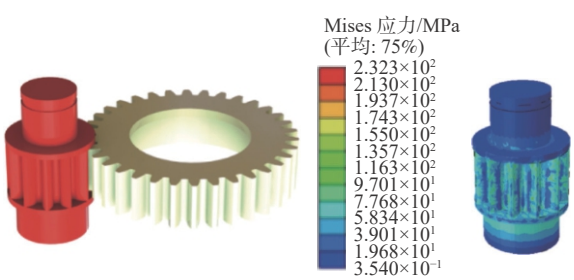


图 6　ABAQUS 力矩传递齿轮应力仿真结果

Fig. 6　ABAQUS moment transmission gear stress simulation results

当前强度安全系数的值基于无限寿命目标计算得出, 如图 7 所示, 在 FE-SAFE 软件中, 若结果中 Infinite 大于 1, 零件在该应力状态下所设置的循环次数满足要求, 零件接近无限寿命, 不会发生疲

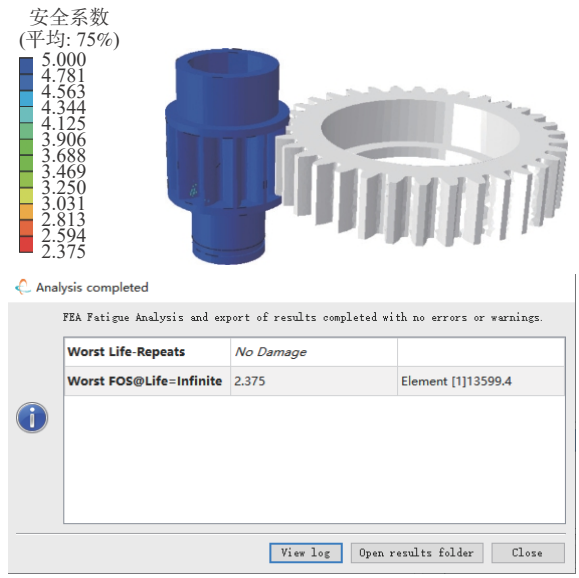


图 7　FE-SAFE 齿轮传递疲劳寿命计算

Fig. 7　Calculation of FE-SAFE gear transmission fatigue life

劳损坏; 若 Infinite 参数小于 1, 零件疲劳寿命为设置的循环次数与 Infinite 数值的乘积。根据图中 Infinite=2.375, 故在循环次数为 1×10^5时, 拧紧套筒没有发生疲劳破坏, 因此, 通过此种方式验证拧紧套筒疲劳强度校验合格性。

根据拧紧位置结构特点, 设置力矩传递结构为 L 型机构, 为了精确控制旋拧扭矩从而最终控制转角预紧精度, 采用齿轮传动对电子扳手所施加的扭矩进行传递, 在此过程中分别对齿轮传动及力矩传递轴进行载荷强度校核及疲劳寿命校核, 对盲腔螺母拧紧设备的力矩传递机构进行闭环设计, 提高设备力矩传递稳定性及疲劳寿命。

2.3　盲腔螺母拧紧技术拧紧力矩精确控制

采用 Atlas 电动扳手 ETD 系列作为输入扭矩端, 其与 Atlas 控制器 Power Focus 6000 采用 STR 线缆进行连接, 通过此种连接方式, 控制器可将力矩控制信息、角度控制信息等传递至电动扳手进行精确控制, 同时, 为了提高拧紧工具控制器一体化应用水平, 构建计算机→控制器→电动扳手扭矩控制指令传递链, 其通信原理如图 8 所示。计算机可以通过 TCP/IP 协议与控制器进行开放协议通信, 将控制器信息与计算机信息进行数据交互(拧紧任务

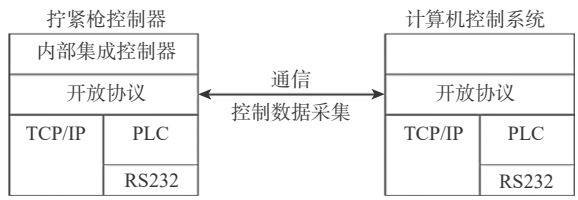


图 8　拧紧控制器扭矩控制通信原理

Fig. 8　Communication principle of tightening controller torque control

选择、拧紧结果反馈、读取拧紧数据库、拧紧开始与停止等)；同时，控制器可以利用内部 PLC 与外部 PLC 连接，将控制器与计算机进行 I/O 口状态交互，对计算机和控制器进行实时监测与控制。

由于发动机转子组件结构限制，采用 L 型拧紧臂进行传递力矩，该种力矩传递方式传送链较长，并且存在齿轮传动损失、传递轴力矩损失、转接头力矩损失、摩擦损失及轴承损失等，力矩衰减不可避免。采用力矩控制系统通过实验测试得到力矩损失百分比，如图 9 所示。力矩传递机构拧紧前通过力矩校准仪 STA 6000 配置 IRTT-B-25A-10A 传感器采集实际输出的力矩及角度，并通过 Power Focus 6000 控制器读取拧紧枪输入扭矩建立力矩输入输出模型。

通过力矩校准仪-检测仪自动计算拧紧枪输出和工装实际输出力矩、角度的对应关系，如表 1 所

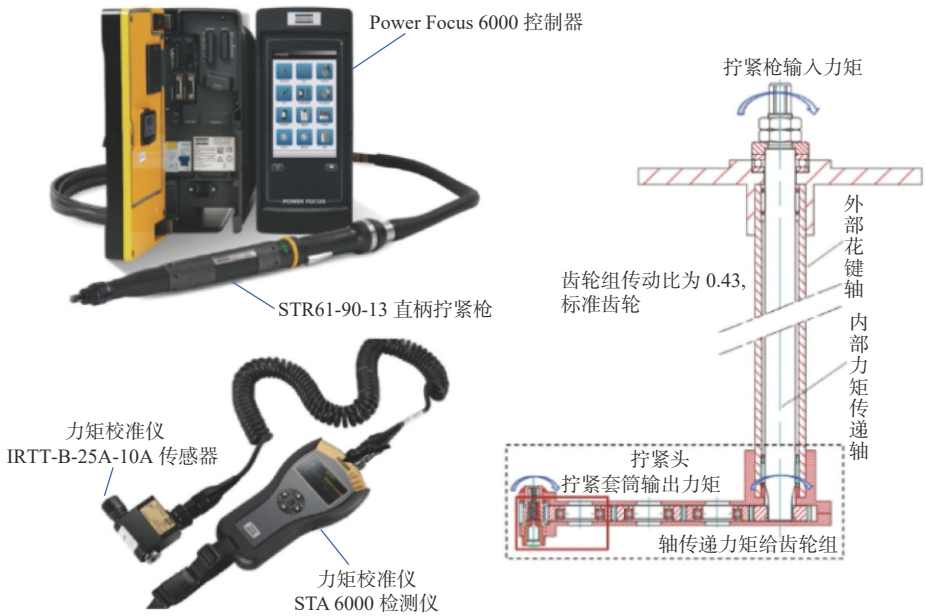


图 9 力矩校准实验逻辑示意图

Fig. 9 Schematic diagram of torque calibration experiment logic

表 1 力矩校核输入输出实验

Table 1 Torque verification input-output experiment

组号	拧紧枪输入力矩/(N·m)	拧紧套筒输出力矩/(N·m)					较输出力矩平均值的最大波动量百分比/%	
		第1次	第2次	第3次	平均值	标准差	最大下波动	最大上波动
1	20	7.59	7.53	7.58	7.57	0.026	−0.48	0.31
2	22	7.87	7.82	7.88	7.86	0.026	−0.47	0.30
3	24	8.44	8.43	8.46	8.44	0.012	−0.16	0.20
4	26	9.04	9.06	9.05	9.05	0.008	−0.11	0.11
5	28	9.61	9.68	9.63	9.64	0.029	−0.31	0.41
6	30	10.30	10.33	10.32	10.32	0.012	−0.16	0.13
7	32	10.98	10.93	10.96	10.96	0.021	−0.24	0.21
8	34	11.65	11.66	11.66	11.66	0.005	−0.06	0.03
9	36	12.23	12.27	12.29	12.26	0.025	−0.27	0.22
10	38	12.96	12.89	12.96	12.94	0.033	−0.36	0.18
11	40	13.65	13.55	13.67	13.62	0.052	−0.54	0.34
12	42	14.29	14.23	14.31	14.28	0.034	−0.33	0.23
13	44	15.04	14.94	14.98	14.99	0.041	−0.31	0.36
14	45	15.52	15.52	15.53	15.52	0.005	−0.02	0.04
15	46	15.80	15.72	15.69	15.74	0.046	−0.30	0.40
16	47	16.12	16.19	16.14	16.15	0.029	−0.19	0.25
17	48	16.30	16.41	16.38	16.36	0.046	−0.39	0.29
18	49	16.61	16.71	16.68	16.67	0.042	−0.34	0.26
19	50	16.96	17.08	17.01	17.02	0.049	−0.33	0.37
20	52	17.64	17.69	17.74	17.69	0.041	−0.28	0.28
21	54	18.15	18.32	18.28	18.25	0.073	−0.55	0.38
22	56	18.93	19.01	18.97	18.97	0.033	−0.21	0.21
23	58	19.52	19.74	19.70	19.65	0.096	−0.68	0.44
24	60	20.16	20.24	20.36	20.25	0.082	−0.46	0.53

示, 拧紧套筒输出力矩与拧紧枪输入力矩的回归关系为 $y=0.328x+0.587\,2$, 且 R^2 为 0.999, 十分接近 1, 如图 10 所示, 故线性拟合直线可信度较高, 且通过斜率 0.328 可知, 拧紧设备的力矩损失具有较恒定的百分比, 为 67.2%。计算出要使拧紧套筒输出 16 N·m 的力矩, 则需要控制拧紧枪输入 46.99 N·m 的力矩; 控制拧紧枪输入 46.99 N·m 的力矩, 重复拧紧 3 次, 记录每组实验后力矩校准仪-检测仪显示的实际拧紧力矩大小, 得到实际输出力矩的标准差为 0.005～0.096, 实际输出力矩较平均输出力矩最大波动量百分比区间为 [−0.68%, 0.53%], 末端执行器扭矩控制精度较高。

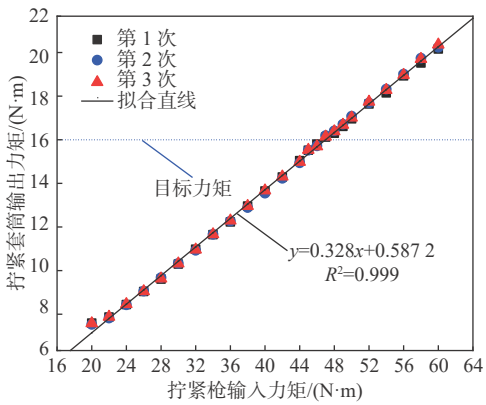


图 10　力矩校核输入输出扭矩关系

Fig. 10　Torque verification input-output torque relationship

3　盲腔螺母拧紧技术操作操作执行机构设计

3.1　盲腔螺母拧紧技术操作执行机构机械设计

发动机转子组件盘间的狭小盲腔结构对装配系统提出很高要求, 不仅要求采用稳定的大力矩、长距离传递机构, 使 L 型拧紧臂顺利到达螺栓阵列预定位置, 而且对拧紧臂的运动轨迹和位置、力矩的输出等多方面提出很高的精度控制需求; 另外, 系统还需要采用全自动化控制方案, 以提升装配效率和智能化水平。为了保证 L 型拧紧臂能够准确到达拧紧位置, 以旋转-平移-旋转(revolute-prismatic-revolute, RPR)关节 SCARA 机器人构型为依据设计操作执行机构。如图 1 所示的设备执行机构示意图, M1 凸轮升降机构为操作执行机构提供移动副, M2、M3 蜗轮蜗杆机构为操作执行机构提供 2 个旋转副, 将其进行串联构建 RPR 机器臂, 精确控制末端执行器的转角和位姿。

发动机转子组件盘间的狭小盲腔结构螺栓拧紧对装配系统提出很高要求, 为保证稳定大力矩、长距离的力矩传递, 操作执行机构末端执行器需准确到达螺栓阵列预定位置。精密蜗杆传动机构模型为盲腔螺母拧紧设备中的一部分, 该传动机构用于整体拧紧设备的旋转。设备的转动有很高的精度要求, 需要对蜗杆传动电机精度进行分析, 以确定其达到设计的精度要求, 基于 SolidWorks 软件建立设备的三维模型, 选取其中包含蜗杆传动机构的部分及单独的蜗轮蜗杆配合的三维模型进行精度分析, 使用 3DCS 偏差仿真软件, 基于蜗轮蜗杆传动模型进行三维偏差仿真建模及分析, 设备的蜗轮蜗杆传动模型如图 11 所示。

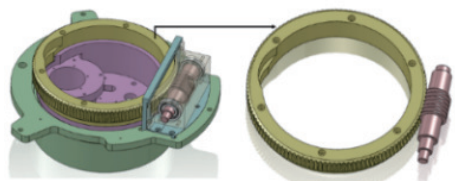


图 11　蜗轮蜗杆传动模型

Fig. 11　Worm gear transmission model

基于仿真模型的零部件几何尺寸、设计公差和定位基准、装配顺序等装配工艺, 利用蒙特卡罗法模拟输入不同公差取值情况下产品的装配过程, 通过图表形式表示测量目标值的分布情况, 并得出各个关键尺寸公差对测量值的贡献率。3DCS 的仿真分析流程如图 12 所示。创建仿真模型关键的 3 个操作为创建装配关系、定义零件公差及创建测量, 这 3 个步骤称为“MTM”[23]。根据实际的装配关系创建模型中零件的装配; 根据上述尺寸链分析, 选择零件公差, 并输入公差的设计值; 在测量侧隙时, 选择蜗轮蜗杆的法向侧隙进行测量。

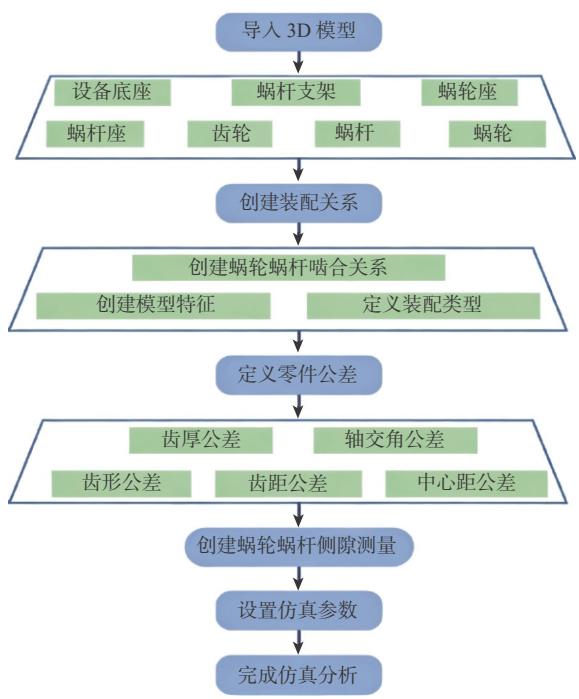


图 12　3DCS 仿真流程

Fig. 12　3DCS simulation process

在仿真前需要对添加完装配关系、公差信息及测量关系的模型进行验证。确定仿真模型没有错误后设置仿真参数，将仿真次数设置为2 000，结果显示贡献度。在当前公差设计值下，蜗轮蜗杆的侧隙范围为0.082～0.248，侧隙不存在超差现象。中心距公差对侧隙的影响程度最大，其次为蜗轮蜗杆的齿厚公差。将实测结果与理论结果进行对比，能够在蜗杆装配开始前，分析部件的制造准确度是否满足侧隙要求，提高零件匹配稳定性，同时对尺寸链环上贡献度较高的影响因子进行公差优化，进一步降低蜗轮蜗杆传动侧隙的超差率，达到预装配分析的目的，减少人工参与次数，提高装配精度及装配效率。

3.2 盲腔螺母拧紧技术操作执行机构轨迹规划

转子盘腔轴向复杂紧凑且径向呈阶梯状分布，如图13所示，L型拧紧臂在操作空间中的动作分为2个阶段，分别为阶段1(到达待拧紧螺栓上方)和阶段2(开始拧紧)。由图13可知，转子内部呈狭窄阶梯状分布且螺栓底面高于转子盘顶面，为了保证螺母能够正确无干涉进行拧紧，须保证螺母初始位置高于螺栓底面。在设计过程中，充分考虑力矩传递机构所要求设计齿轮传动中传递齿轮的极限厚度，在盲腔螺母拧紧设备工装状态轴向分布上形成螺栓、螺母、拧紧臂组合装配，须对盲腔螺母拧紧设备的操作执行机构进行精确轨迹规划。

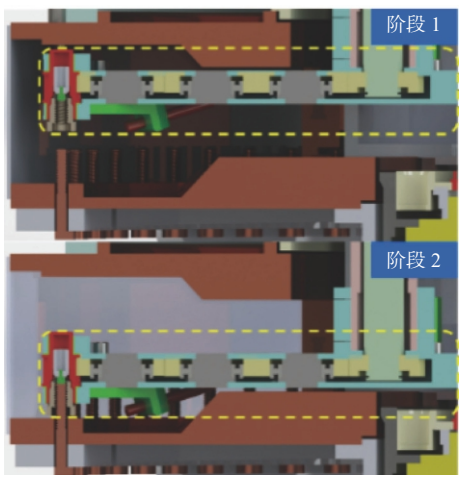


图13 拧紧臂在操作空间工作示意图

Fig. 13 Schematic diagram of the tightening arm working in the operating space

对操作执行机构进行运动学仿真，在操作执行机构及RPR机械臂各个关节处建立坐标系，进而推导出基坐标系与末端坐标系之间的位姿关系，根据所建立的全局坐标系分析机器人正逆运动学的正逆解。

首先，根据机械臂构型建立*DH*坐标系并确定*DH*参数，*DH*法则的本质是坐标变换，将坐标系建立在机器人的移动或转动关节处。然后，进行相应的坐标变换，得到关节变量参数，代入运动方程获得解析解。*DH*坐标系建立公式如下：

$$
\begin{aligned}
{}^{i-1}\boldsymbol{T}_i=&\mathrm{Rot}(x_i,\alpha_{i-1})\mathrm{Trans}(x_i,l_{i-1})\mathrm{Rot}(z_{i-1},\gamma_i)\mathrm{Trans}(z_{i-1},D_i)=\\
&\begin{bmatrix}1&0&0&0\\0&\cos\alpha_{i-1}&-\sin\alpha_{i-1}&0\\0&\sin\alpha_{i-1}&\cos\alpha_{i-1}&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&l_{i-1}\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\cdot\\
&\begin{bmatrix}\cos\gamma_i&-\sin\gamma_i&0&0\\\sin\gamma_i&\cos\gamma_i&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&D_i\\0&0&0&1\end{bmatrix}
\end{aligned}
\tag{4}
$$

式中：$\boldsymbol{T}_i$为坐标转换矩阵；Rot(·)为旋转坐标转换矩阵，$\mathrm{Rot}(x_i,\alpha_{i-1})$表示沿$x_i$轴转动$\alpha_{i-1}$，$\mathrm{Rot}(z_{i-1},\gamma_i)$表示沿$z_{i-1}$轴转动$\gamma_i$；Trans(·)为平移坐标转换矩阵，$\mathrm{Trans}(x_i,l_{i-1})$表示沿$x_i$平移$l_{i-1}$，$\mathrm{Trans}(z_{i-1},D_i)$表示沿$z_{i-1}$轴平移$D_i$。

建立*DH*坐标系后，对操作执行机构进行正运动学计算，记录末端执行器位置，绘制操作执行机构操作空间，如图14所示。依据仿真计算结果，操作执行机构在z轴的操作空间范围为[342,360] mm，即拧紧臂上下运动幅度为18 mm，符合图13所示的空间状态。操作空间在xy平面展开图为同心圆环，表示操作执行机构末端执行器可准确到达以设备中心为基点的分布圆上，符合操作执行机构设计要求。

根据操作执行机构的工作环境和工作任务要求，求得操作执行机构末端位姿变换的时间序列，使机器人末端执行器可以正确地从初始姿态沿着期望的轨迹运动到终止位姿，完成工作任务。对于RPR机械臂，轨迹规划要解决的关键问题是将末端位姿的变换转变为3个关节变量的变换，在轨迹规划中采用五次多项式进行规划，其一般形式为

$$
\begin{aligned}
\theta(t)=&a_0+a_1(t-t_\mathrm{s})+a_2(t-t_\mathrm{s})^2+a_3(t-t_\mathrm{s})^3+\\
&a_4(t-t_\mathrm{s})^4+a_5(t-t_\mathrm{s})^5
\end{aligned}
\tag{5}
$$

式中：θ为关节角度；t为轨迹规划时间；t_s为轨迹规划初始时间；a_0、a_1、a_2、a_3、a_4、a_5为轨迹规划参数。6个未知数需要给定6个约束条件，分别为起点和终点的角位移、角速度和角加速度，如下：

$$
\begin{cases}
\theta(t_\mathrm{s})=\theta_0\\
\theta(t_\mathrm{f})=\theta_\mathrm{f}\\
\dot{\theta}_0=0\\
\dot{\theta}_\mathrm{f}=0\\
\ddot{\theta}_0=0\\
\ddot{\theta}_\mathrm{f}=0
\end{cases}
\tag{6}
$$

式中：θ_0为初始关节角度；θ_f为最终关节角度；t_f为

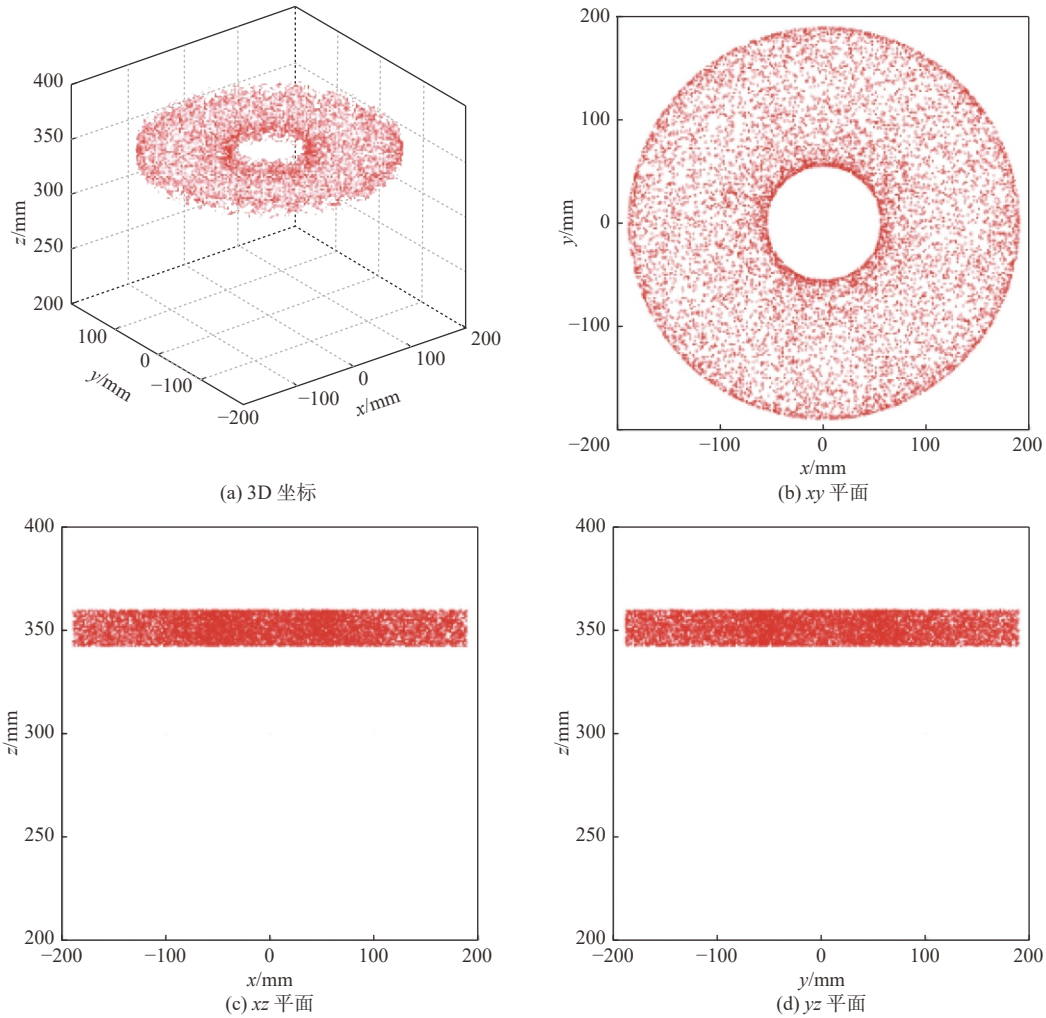


图 14　操作执行机构操作空间仿真计算

Fig. 14　Simulation calculation of the operating space of the operating actuator

轨迹规划终止时间。

将以上约束条件(式(6))代入式(5), 解得

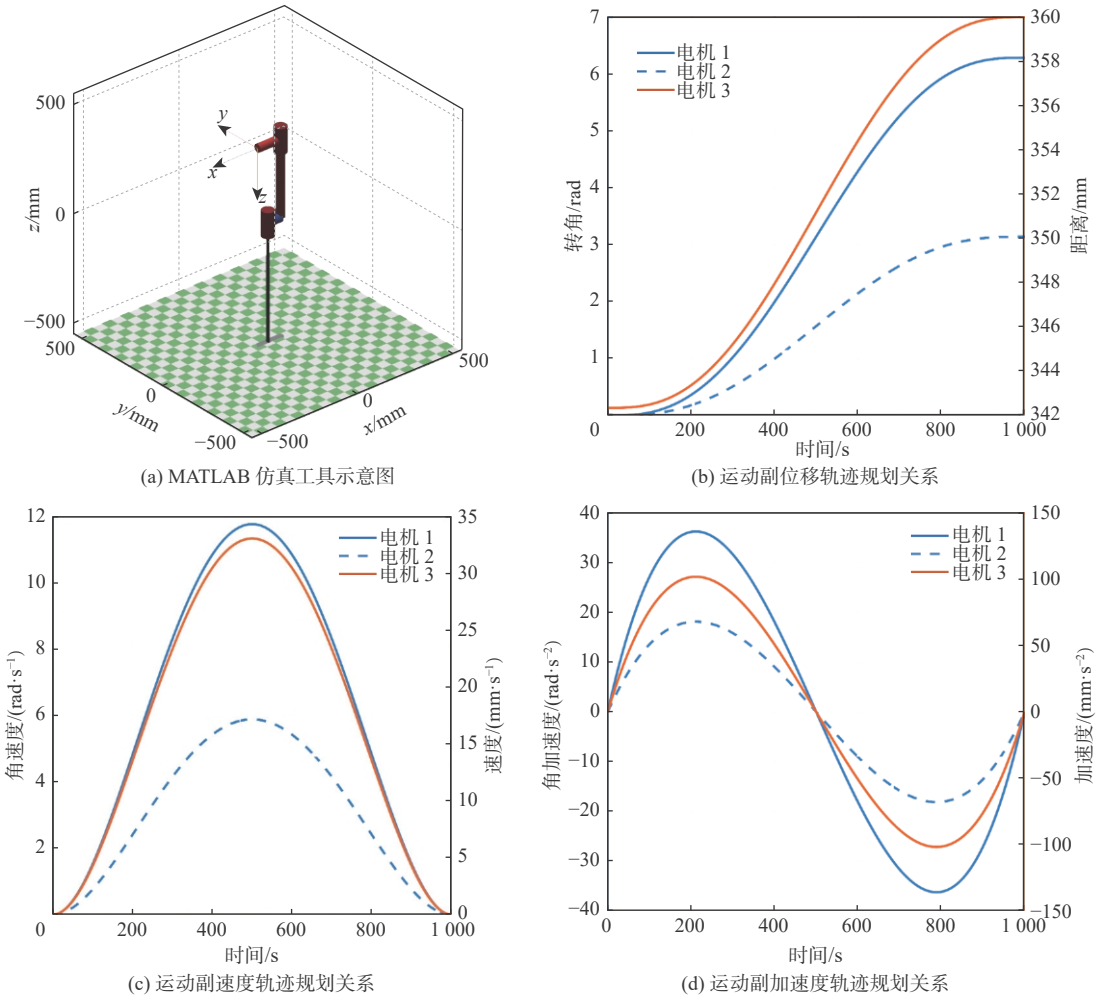
$$\begin{cases} a_0 = \theta_0 \\ a_1 = 0 \\ a_2 = 0 \\ a_3 = 10(\theta_f - \theta_0) \\ a_4 = -15(\theta_f - \theta_0) \\ a_5 = 6(\theta_f - \theta_0) \end{cases} \tag{7}$$

采用 MATLAB 中 Robotics Toolbox 工具对机器人进行逆运动学计算, 确定操作执行机构各个关节运动状态。图 15 为操作执行机构轨迹规划仿真结果, 分别得到各个关节的位姿、速度、加速度, 将轨迹规划信息最终传至控制系统, 实现对执行系统末端执行器高精度位姿控制。将此控制信息由计算机写入操作执行机构控制器, 完成对操作执行机构的高精度计算控制。

3.3　盲腔螺母拧紧技术操作执行机构的高精度控制

根据操作执行机构轨迹规划结果, 选择合适的微型 PLC 控制器、传感器等, 确立相应的输入/输出外围设备对操作执行机构进行操作。将该机电一体化设备分为感知系统、控制系统、驱动系统、机械系统, 并分别将各个系统组合在一起。其中, 感知系统由传感器和微型工业相机组成, 其将信息分别传入 PLC 和 IPC 中, PLC 与 IPC 进行数据交互, 形成对操作执行机构的控制信息整合, 以此控制伺服电机驱动器的输出, 进而控制伺服电机动作。

PLC 与 IPC 进行被动通信, 通信由 IPC 发起, 按照通信协议由 IPC 向 PLC 发送指令。当被动通信时, PLC 与 IPC 的通信内容包括数据读写、状态读写和通信测试。其中, 数据读写是 IPC 向 PLC 某个数据区写数据或 IPC 从 PLC 的某个数据区读数据, 读写不同的数据区的指令不同。状态读写即 IPC 可通过通信命令读或写 PLC 的状态, 如运行状态、监控状态或编程状态, 状态可以使 PLC 停机或开机。通信测试即 IPC 向 PLC 发送通信测试命令, 用以测试通信系统是否正常, 在搜索通信接口时使用。PLC 与 IPC 的通信协议主要是对 PLC 内部元



(a) MATLAB 仿真工具示意图

(b) 运动副位移轨迹规划关系

(c) 运动副速度轨迹规划关系

(d) 运动副加速度轨迹规划关系

图 15 操作执行机构轨迹规划仿真结果

Fig. 15 Simulation results of operation actuator trajectory planning

件 X、Y、M、S、T、C、D 的读写指令，其中，输入继电器 X 是 PLC 从外部的输入开关接收信号的窗口，只能读不能写，输入寄存器必须由外部信号驱动；输出继电器 Y 是 PLC 驱动外部负载的窗口；辅助继电器 M 是 PLC 内部辅助继电器软元件，数据寄存器 D 是 PLC 保存数据的软元件。上位机通过通信协议读写内部元件状态，从而对 PLC 进行控制和读取。

图 16 为 IPC 采用通信协议向 PLC 发送命令的格式。其中，STX 为开始字符，其 ASCII 值 16 进制为 02H；CMD 为命令码，命令码分为读和写，其中，

开始字符	命令码	起始地址	字节数	数据	结束字符	累加和
STX	CMD	ADDR	NUM	DATA1, DATA2,⋯	ETX	SUM

开始字符	命令码	起始地址	数据	结束字符	累加和
STX	CMD	ADDR	DATA1, DATA2,⋯	ETX	SUM

图 16 PLC 通信协议发送命令和响应格式

Fig. 16 PLC communication protocol sending commands and response formats

读的 ASCII 码为 30H，写的 ASCII 码为 31H；ADDR 为起始地址，采用 16 进制表示，占 4 个字节；NUM 为读或写的字节数，占 2 个字节；DATA 为写入的数据，此处填写要写的数据；ETX 为结束字符，其 ASCII 码 16 进制为 03H；SUM 为累加和，以此格式完成 IPC 与 PLC 之间的数据通信。从命令码开始到结束字符停止，各个字符之间的 ASCII 码进行 16 进制累加，结果取累加和低 2 位，其计算方法为 SUM = CMD + ADDR + NUM + DATA1 + DATA2 + ⋯ + ETX。

控制系统上位机软件在 Qt Creator 利用 C/C++ 语言编写上位机程序，Qt Creator 是一种跨平台 C++ 图形用户界面应用程序开发框架，支持 Linux、Windows 和 MacOS 等[24]。

利用 Qt Creator C++中自带的 QSerialPort 类进行上位机与 PLC 串口通信。图 17 为 IPC 与 PLC 通信流程。IPC 通过上位机与 PLC 进行指令交互，同时，在下位机及 PLC 内部程序的编写中，分别用不同的 D 寄存器储存及更改驱动电机脉冲

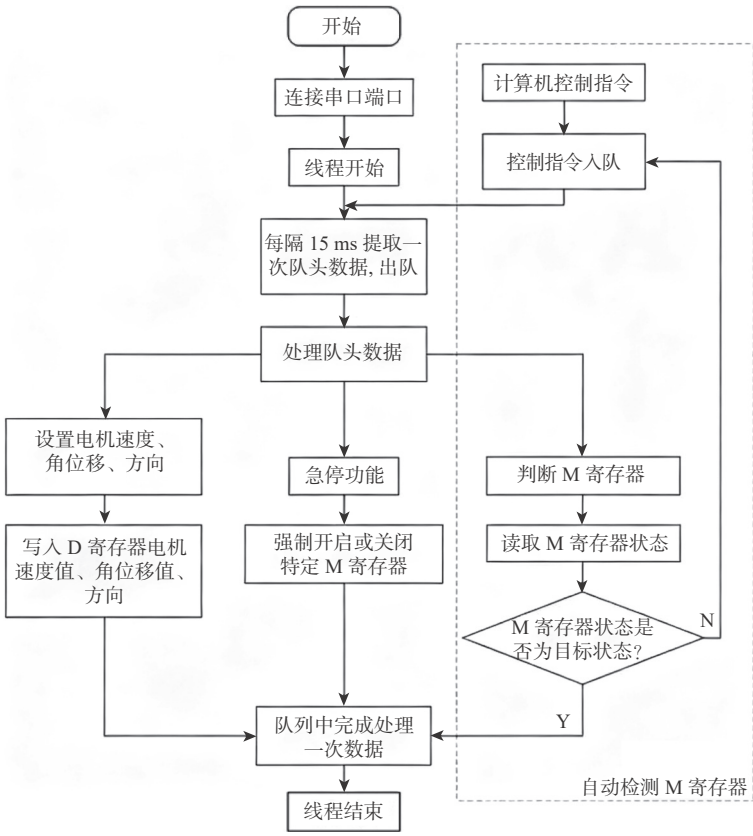


图 17　IPC 与 PLC 通信流程

Fig. 17　Communication flowchart between IPC and PLC

信号的脉冲数和脉冲频率, 进而对伺服电机进行速度、位移设定, 同时, 利用 M 寄存器标记电机工作状态及控制电机方向; 上位机根据对 M 寄存器的实时判断, 监测操作执行机构所处位姿状态及电机运行状态, 从而根据轨迹规划结果将速度指令、位移指令、加速度指令、电机运行方向控制指令发出, 进而驱动操作执行机构智能化、高精度位姿调整, 实现操作执行机构的高精度控制, 提高操作执行机构运行稳定性和螺栓拧紧质量稳定性。

受到狭小盲腔结构特性的限制, 无法直观地对盲腔内部拧紧状态进行观测, 而且航空发动机通常不具备内外对应的结构基准, 难以基于外部基准特征对内部螺栓位置进行定位, 故不能通过设备所处外部位置来对应待拧紧螺栓位置, 因此, 设备中安装有角度传感器, 与整个控制系统构成闭环控制。角度传感器与设定角度值进行比较, 如果在误差范围内则进行螺栓的拧紧, 否则进行微调工装, 直到在误差范围内, 从而保证螺栓识别定位的准确性。由视觉识别模块确定 1 号螺栓位置后, 以 1 号螺栓为基准, 角度传感器在设备转动机构每次运动完成之后获取相对于上一个位置的角度, 提供给上位机进行处理。视觉定位的机理为利用不同条件下的螺栓照片进行机器学习, 确定螺栓中心线的位置, 计算螺栓中线位置与拧紧头的中线位置相差角度, 反馈给上位机处理, 进而控制设备转动直至拧紧头的中心线和螺栓的中心线满足角度精度(1°), 即可准确定位螺栓, 保证拧紧套筒认帽成功。螺栓中心线提取原理如图 18 所示, 基于 OpenCV 库对图像进行滤波、二值化、轮廓提取螺栓左侧线 $x = x_1$、螺栓右侧边缘线 $x = x_2$、轮廓拟合与中心线提取处理, 提高模型稳定性。

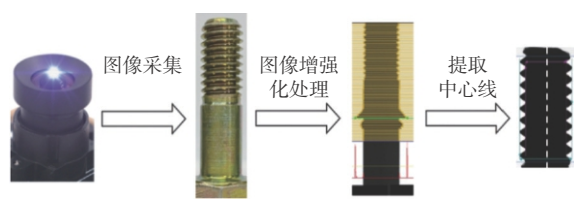


图 18　螺栓中线提取原理

Fig. 18　Principle of bolt centerline extraction

螺栓识别模型学习流程如图 19 所示。首先, 采集足够多的照片(包含不同光照条件、不同位置、不同拧紧状态), 手动标定螺栓轮廓、3 种分类(拧紧状态: 有螺母 0、无螺母 1、未对准 2), 建立机器学习图片库。然后, 使用 PyTorch 开源机器学习库进行模型训练, 模型训练为机器学习中的监督学习方式, 流程是机器通过图像给出螺栓轮廓线位置信息和拧紧状态分类信息, 与标定信息进行比较,

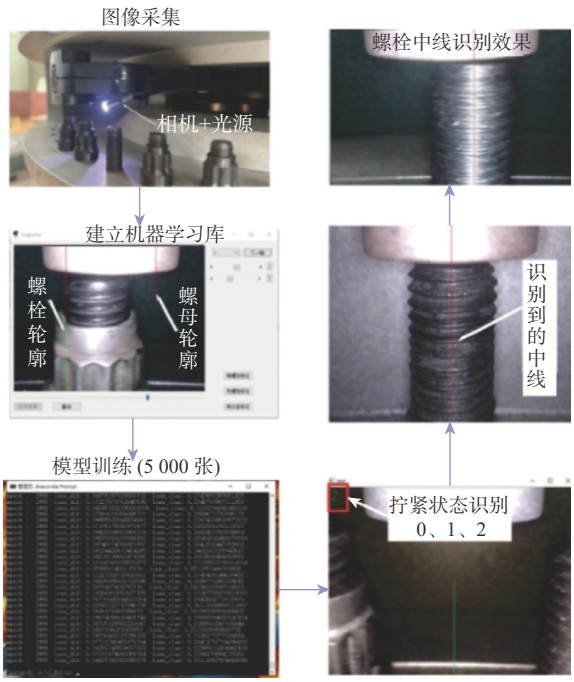


图 19 螺栓识别模型机器学习流程及识别效果展示

Fig. 19 Machine learning process and recognition effect display of bolt recognition model

机器调整参数值。

如图 20 所示，采用区域卷积神经网络（region-based convolutional neural network, R-CNN）算法对螺栓进行边缘识别，创建图像的螺栓候选区域，将候选区域信息和螺栓特征信息通过支持向量机（support vector machine, SVM）分类器进行迭代学习，图中，W 为卷积核的权重参数，B 为偏置参数或全连接层权重参数，C 为全连接层偏置参数。同时，为了适应设备高精度快速识别需求，采用轻量化神经网络 MobileNet v2 作为 backbone 传入 R-CNN 网络，提高模型的工业化应用。

图 21 为机器学习过程中损失函数、识别精度与迭代次数之间的关系。经过 100 次迭代后，损失函数曲线已经趋于稳定，训练集预测精度可达 0.98，说明模型的稳定性和正确性；同时，由于 MobileNet v2 作为本训练模型的 backbone，模型识别螺栓精度误差在 0.2 mm 以内，模型响应时间小于 0.2 s。本次训练的图片数量为 5 000 张，效果已经十分理想，图片越多，训练的次数越多，结果更精确。

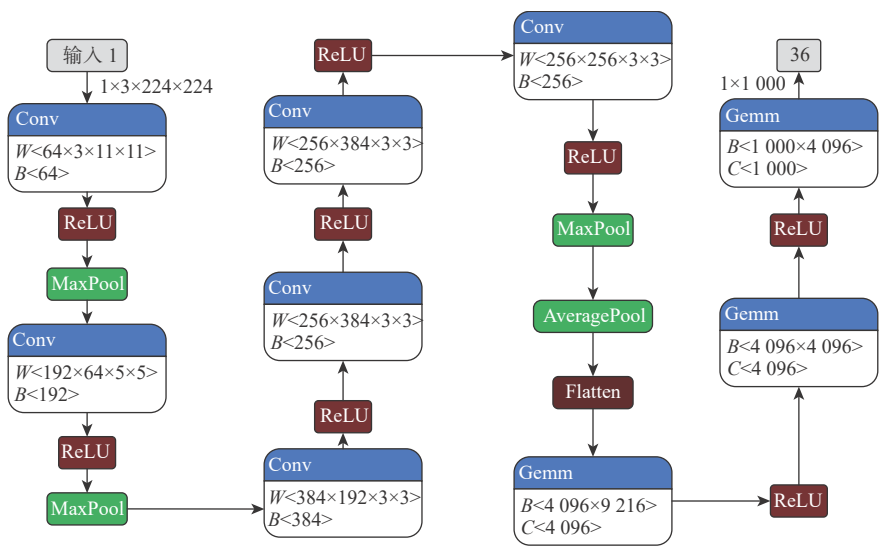
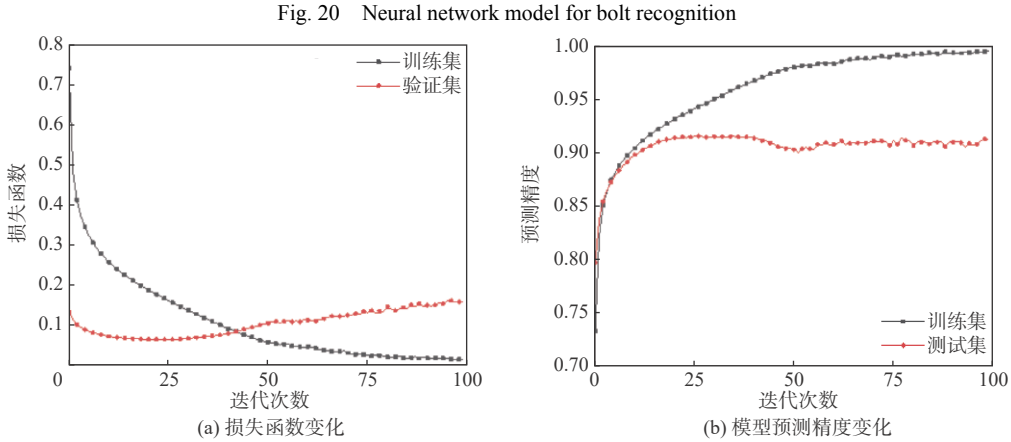


图 20 螺栓识别神经网络模型

Fig. 20 Neural network model for bolt recognition



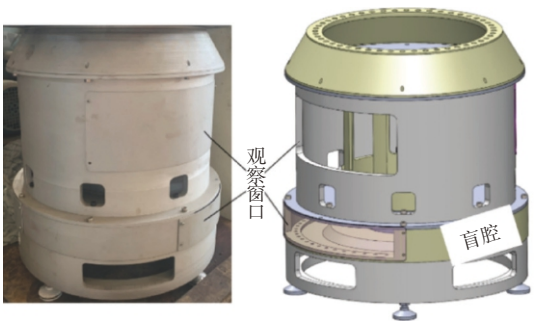
(a) 损失函数变化 (b) 模型预测精度变化

图 21 机器学习过程中损失函数和精度变化

Fig. 21 Loss function and accuracy change in machine learning process

4　盲腔螺母拧紧技术实验验证

针对某航空发动机压气机转子盘间螺栓拧紧, 设计转子盲腔模拟件, 如图 22 所示。模拟件在真实压气机三维模型上简化得来, 保留关键位置尺寸, 简化其他非必要结构, 模拟件盲腔结构、尺寸、相对位置关系与实际发动机待连接盘件盲腔基本一致, 且盲腔的工作环境基本符合真实现场。盲腔螺母拧紧装配控制系统主要针对该模拟件下盲腔内部 60 个螺栓自动化拧紧, 最终在真实发动机转子上进行相应的调试。



(a) 实物图　　(b) 设计模型

图 22　转子盲腔模拟件实物图和设计模型

Fig. 22　Physical image and design model of the rotor blind cavity simulation part

如图 23 所示, 开展操作执行机构方面实验验证。在模拟件上开展轨迹规划验证实验和拧紧拧松实验, 共计完成 60 个螺母的拧紧拧松实验, 实验过程设备与实际转子阶梯状结构无干涉且实现螺母连接过程全程可视化。如图 24 所示, 开展传递机构精度实验验证, 将拧紧设备的拧紧套筒与力矩校准仪-传感器的输入端通过转接头相连, 力矩校准仪-传感器的输出端与模拟螺栓螺母相连, 尽可能地模拟螺母拧紧过程中的旋拧贴合及拧紧至指

(a) 工作模型　　(b) 实物图

图 23　盲腔螺母拧紧设备工作模型和实物图

Fig. 23　Working model and physical image of blind cavity nut tightening equipment

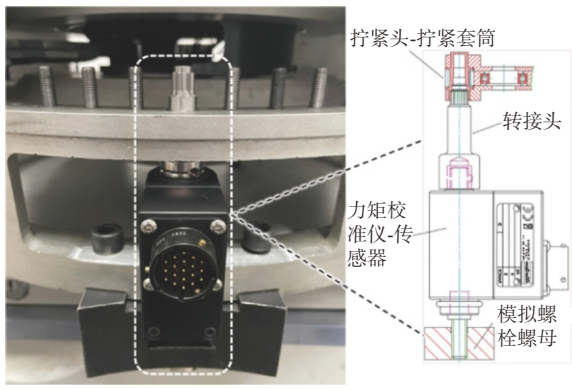


图 24　力矩传递机构精度模拟件测量原理

Fig. 24　Measurement principle of precision analog components for torque transmission mechanism

定力矩。力矩校准仪-传感器与力矩校准仪-检测仪通过线缆连接, 当螺母拧紧至最大力矩后, 力矩校准仪-检测仪的界面即可显示出螺母的实际拧紧力矩和转角信息。

为验证此种控制方式下的扭矩控制精度, 面向现场的操作环境, 将其与现场常用的定力矩扳手拧紧的方式进行对比, 如图 25 所示。经过真机测验, 定力矩扳手输出扭矩分散度为 7.64%, 本文设备输出扭矩分散度可达 0.36%, 表明本文设计力矩控制精度远高于人工通过定力矩扳手拧紧。

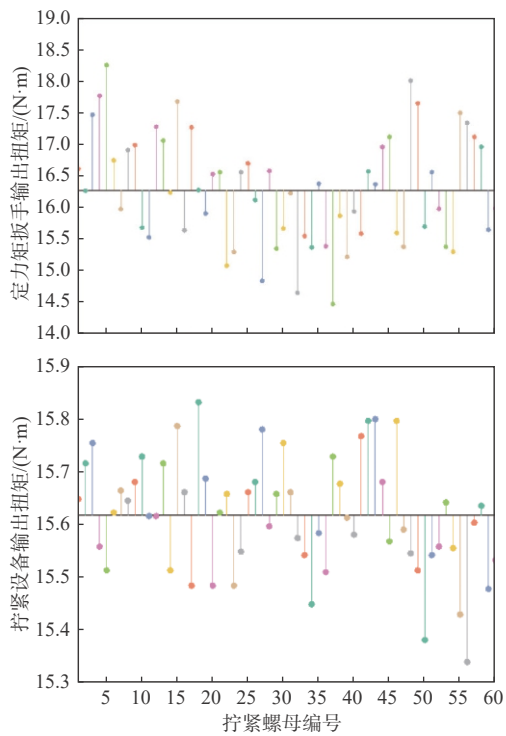


图 25　扭矩控制实验效果对比

Fig. 25　Comparison of torque control test results

5　结　论

本文对盲腔螺母拧紧设备工况进行分析, 分析待拧紧螺栓所处结构特点并给出采用 L 型拧紧臂

进行拧紧螺母的设计思路，由此将盲腔螺母拧紧设备系统架构分为力矩传递机构和操作执行机构。

1）在力矩传递机构设计上，对拧紧工艺进行分析，分析采用力矩法、力矩转角法进行拧紧时对螺栓力矩控制精度的异同，最终确定对螺栓力矩的控制方法为力矩转角法；依据确定的拧紧工艺原理，对力矩传递模块进行综合设计，综合分析力矩传递链上各个零件的强度及疲劳寿命，提高设备的稳定性；基于本文设备对力矩传递结构传递的扭矩进行校核与反馈，实现力矩传递结构的高精度控制。

2）在操作执行结构设计上，依据RPR机械臂构型对操作执行机构进行机械设计，根据拧紧过程中操作执行机构的高精度控制要求，对关键装配部位进行3DCS仿真分析，进而指导零件制造，同时，在操作执行机构设计上由于操作空间的限制，对机械臂进行详细轨迹规划，并通过控制器对驱动电机进行精确控制。

3）航空发动机转子具有内外不对应的结构特点，本文基于大量学习样本，实现视觉检测定位待拧紧1号螺栓的位置，研究基于机器学习的误差辨识算法，得到定位误差与图像之间的网络模型，由此建立“图像-误差-控制”快速关联模型，实现高精度、高效率运动闭环控制。

参考文献（References）

[1] 高一超. 航空发动机转子装配精度预测及相位优化方法研究[D]. 大连: 大连理工大学, 2021: 20-24.
GAO Y C. Research on accuracy prediction and phase optimization of aeroengine rotor assembly[D]. Dalian: Dalian University of Technology, 2021: 20-24(in Chinese).

[2] 方昌德. 航空发动机的发展前景[J]. 航空发动机, 2004, 30(1): 1-5.
FANG C D. Prospective development of aeroengines[J]. Aeroengine, 2004, 30(1): 1-5(in Chinese).

[3] 刘长福, 邓明. 航空发动机结构分析[M]. 西安: 西北工业大学出版社, 2006.
LIU C F, DENG M. Structural analysis of aero-engine[M]. Xi'an: Northwestern Polytechnical University Press, 2006(in Chinese).

[4] 孟亮国. 航空发动机转子装配工艺优化方法研究[D]. 大连: 大连理工大学, 2019.
MENG L G. Research on process optimization of aeroengine rotor assembly[D]. Dalian: Dalian University of Technology, 2019(in Chinese).

[5] 刘鑫. 航空发动机转子装配精度预测及堆叠[D]. 大连: 大连理工大学, 2019: 33-39.
LIU X. Assembly accuracy prediction and stacking of aeroengine rotor[D]. Dalian: Dalian University of Technology, 2019: 33-39(in Chinese).

[6] 魏企业, 石宏, 孙方成, 等. 航空发动机转子装配螺栓自动拧紧机设计与分析[J]. 机械, 2017, 44(6): 67-70.
WEI Q Y, SHI H, SUN F C, et al. Design and analysis of aero engine rotor automatic bolt tightening machine structure[J]. Machinery, 2017, 44(6): 67-70(in Chinese).

[7] 刘君, 吴法勇, 王娟. 航空发动机转子装配优化技术[J]. 航空发动机, 2014, 40(3): 75-78.
LIU J, WU F Y, WANG J. Optimization technique of aeroengine rotor assembly[J]. Aeroengine, 2014, 40(3): 75-78(in Chinese).

[8] 叶永松. 考虑实际接触的螺栓止口连接预紧力及其变形研究[D]. 大连: 大连理工大学, 2019: 41-52.
YE Y S. Study on pre-tightening force and deformation of bolted connection with rabbet considering actual contact[D]. Dalian: Dalian University of Technology, 2019: 41-52(in Chinese).

[9] 韩玉杰. 力矩法下航空发动机螺栓预紧力控制研究[D]. 北京: 北京航空航天大学, 2022: 26-32.
HAN Y J. Research on preload control of aero-engine bolts under moment method[D]. Beijing: Beihang University, 2022: 26-32(in Chinese).

[10] 孙伟, 王连吉, 杨阳, 等. 一种航空发动机连杆机构找准螺母拧紧装置及方法: CN110561097B[P]. 2020-12-11.
SUN W, WANG L J, YANG Y, et al. An aero-engine connecting rod mechanism finding nut tightening device and method: CN110561097B[P]. 2020-12-11(in Chinese).

[11] 李琳, 张渝, 周烁, 等. 一种螺母安放拧紧装置、方法及系统: CN105522537B[P]. 2017-05-17.
LI L, ZHANG Y, ZHOU S, et al. A nut placement tightening device, method and system: CN105522537B[P]. 2017-05-17(in Chinese).

[12] 孙清超, 杨阳, 袁博, 等. 一种航空发动机内部螺母软轴传动式拧紧装置及方法: CN110561098A[P]. 2019-12-13.
SUN Q C, YANG Y, YUAN B, et al. An internal nut soft shaft transmission tightening device and method for aircraft engines: CN110561098A[P]. 2019-12-13(in Chinese).

[13] 孙清超, 袁博, 杨阳, 等. 一种航空发动机内部螺母万向节传动式拧紧装置及方法: CN110497190A[P]. 2019-11-26.
SUN Q C, YUAN B, YANG Y, et al. An internal nut universal joint transmission tightening device and method for aircraft engines: CN110497190A[P]. 2019-11-26(in Chinese).

[14] 王海晨, 陈栋权, 汪俊熙. 转子连接螺母拧紧装置及拧紧方法: CN115229479B[P]. 2023-11-03.
WANG H C, CHEN D Q, WANG J X. Tightening device and tightening method for rotor connecting nuts: CN115229479B[P]. 2023-11-03(in Chinese).

[15] 丁少兵, 刘国瑞, 张荡, 等. 一种基于模糊PID的扭矩分配限滑控制方法及其装置: CN113085574B[P]. 2022-08-16.
DING S B, LIU G R, ZHANG D, et al. A torque distribution limited slip control method and device based on fuzzy PID: CN113085574B[P]. 2022-08-16(in Chinese).

[16] 于海. 中国航发全域数字化赋能平台建设迈上新台阶[J]. 航空动力, 2020(5): 62.
YU H. Digital platform construction of AECC takes a new step[J]. Aerospace Power, 2020(5): 62(in Chinese).

[17] 杨文强, 张丛辉, 刘晓锋, 等. 一种D头螺栓和航空发动机: CN210034116U[P]. 2020-02-07.
YANG W Q, ZHANG C H, LIU X F, et al. A D-head bolt and aviation engine: CN210034116U[P]. 2020-02-07(in Chinese).

[18] 孙贵青, 王彤, 吕玉红. 涡扇发动机先进装配工艺与装备[J]. 航空

制造技术, 2017, 60(22): 72-77.
SUN G Q, WANG T, LYU Y H. Advanced process and equipment for turbofan engine assembling[J]. Aeronautical Manufacturing Technology, 2017, 60(22): 72-77(in Chinese).

[19] 李小冬. 航空发动机转子止口-螺栓装配工艺研究[D]. 大连: 大连理工大学, 2020.
LI X D. Research on assembly process of aeroengine rotor rabbet-bolt joint structure[D]. Dalian: Dalian University of Technology, 2020(in Chinese).

[20] 刘通. 基于扭矩控制的螺纹自动拧紧设备设计开发[J]. 工业仪表与自动化装置, 2022(4): 137-140.
LIU T. Design and development of automatic screw tightening equipment based on torque control[J]. Industrial Instrumentation & Automation, 2022(4): 137-140(in Chinese).

[21] 朱正德, 林湖. 基于螺栓装配技术中扭矩法与扭矩/转角法比较与应用研究[J]. 柴油机设计与制造, 2005, 11(2): 40-43.
ZHU Z D, LIN H. Comparison and application research between torque method and torque/angle method in bolt assembly technology[J]. Design and Manufacture of Diesel Engine, 2005, 11(2): 40-43(in Chinese).

[22] 闻邦椿. 机械设计手册: 第 5 卷. 机械一体化与控制技术[M]. 6 版. 北京: 机械工业出版社, 2018.
WEN B C. Machine design handbook: volume 5. machanical integration and control technology[M]. 6th ed. Beijing: China Machine Press, 2018(in Chinese).

[23] 王鑫, 李兆宇, 王亮, 等. 基于 3DCS 的某型号导弹舱段装配容差分析[J]. 航天制造技术, 2020(4): 48-52.
WANG X, LI Z Y, WANG L, et al. Tolerance analysis of missile cabin assembly based on 3DCS[J]. Aerospace Manufacturing Technology, 2020(4): 48-52(in Chinese).

[24] 霍亚飞. Qt Creator 快速入门[M]. 北京: 北京航空航天大学出版社, 2012.
HUO Y F. Quick start to Qt Creator[M]. Beijing: Beihang University Press, 2012(in Chinese).

Precision tightening technology for aeroengine blind cavity nuts

ZHANG Zongjiang[1], LI Xiaoqiang[1,*], HAN Yujie[2], LI Zhaoyu[1], ZHANG Mingming[1]

(1. School of Mechanical Engineering and Automation, Beihang University, Beijing 100191, China;
2. Beijing Aerospace Long March Aircraft Research Institute, Beijing 100076, China)

Abstract: The connection quality of the narrow blind cavity of the aeroengine rotor has a great impact on the overall assembly quality of the rotor. The precise control ability of the bolt connection process is inadequate given the current domestic and international circumstances, which impact the connection mechanical properties of the mating surface of the rotor disc and shaft parts. This is because there is a lack of accurate and efficient assembly tooling for the blind cavity structure. Based on the above research status, according to the tightening assembly requirements of a certain type of blind cavity nut, the blind cavity tightening system is divided into a tightening torque transmission mechanism and an operation actuator. In the design of the tightening torque transmission mechanism, firstly, the torque method and torque angle method are compared to determine the torque angle method as the final execution process. The input and output torque model is established taking into account the torque loss to determine the input torque of the tightening tool in order to achieve high-precision control of the tightening torque. At the same time, the torque transmission structure is designed and the strength and longevity of important components are examined. In the design of the operating actuator, firstly, the mechanical design of the actuator and the 3DCS simulation of the high-precision moving parts worm gear and worm are carried out. At the same time, the trajectory planning and kinematics simulation are carried out by using MATLAB. Based on the aforementioned research, the blind cavity nut tightening technology is experimentally verified, and a perception system and control system for visual identification are built to achieve high-precision equipment control. It is of great significance to improve the quality and consistency of the blind cavity bolt array connection between the narrow disks.

Keywords: aeroengine; blind cavity; torque angle method; torque control; trajectory planning; deviation analysis

Received: 2023-11-09; **Accepted**: 2023-12-01; **Published Online**: 2023-12-12 10:15

URL: link.cnki.net/urlid/11.2625.V.20231211.1143.002

Foundation item: National Science and Technology Major Project (2022-V11-0001-0043)

* **Corresponding author**. E-mail: lixiaoqiang@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0747

考虑空间碎片质量的多碎片主动清理任务规划方法

郭甲[1]，张海博[2]，庞兆君[1,*]，杜忠华[1]

(1. 南京理工大学 机械工程学院，南京 210094；2. 北京控制工程研究所，北京 100190)

摘　　要：在解决多碎片主动清理问题时，现有基于时间依赖型旅行商问题(TD-TSP)的决策框架仅对空间碎片清理序列和转移时间进行优化，优化效果有限。依据多碎片主动清理任务特点，将空间碎片质量加入决策需考虑的因素，并将空间碎片释放时机引入决策框架，从而实现转移次数与平台负载之间的平衡，以此实现更深层次的优化。设计适用于漂移轨道转移方法的聚类算法，用以从大规模碎片池中选出适宜作为任务目标的空间碎片，实现挑选空间碎片和任务方案优化的解耦，降低计算量。参考铱星-33碎片云的轨道信息，设置固定空间碎片质量及模拟现实中空间碎片质量不定的仿真实验。结果表明：聚类算法可以选出轨道相近、利于转移的目标。所提决策框架得到的解相较于传统决策框架得到的解获得了更优的指标，通过分批捕获空间碎片可以进一步降低任务成本。

关 键 词：空间碎片；任务规划；遗传算法；聚类分析；多目标优化
中图分类号：V448.2
文献标志码：A　　**文章编号**：1001-5965(2026)01-0275-11

随着人类空间活动的增加[1]，在轨的废弃航天器数量也呈持续增长的态势。机构间空间碎片协调委员会(Inter-Agency Space Debris Coordination Committee, IADC)将这些在轨运行或再入大气层的、无法继续工作的人造物体称为空间碎片。进入21世纪，空间碎片已多次对正常空间任务产生影响。目前，大部分人造物体位于低地球轨道(low Earth orbit, LEO)，此外，伴随星链等互联网星座的卫星持续发射入轨，LEO上的人造物体正在急剧增加[2]，LEO空间碎片数量也随之增长。为了处理LEO空间碎片，需将其转移到高度更低的处置轨道，在处置轨道上的空间碎片将在大气阻力的作用下再入大气层烧毁。现在已有多种LEO空间碎片主动清理技术被提出，如空间绳网捕获[3-4]、空间飞矛捕获[5]、离子束推移离轨[6]和激光推移离轨[7]等。部分处理单个空间碎片的技术已达到较高的成熟度，如欧洲主导的RemoveDEBRIS任务在2018—2019年陆续完成了空间绳网捕获立方星、空间目标运动跟踪、空间飞矛捕获、离轨帆展开等4项技术的在轨验证试验[8-9]。在后续发展中，为了减少火箭发射成本并精简空间碎片清理准备工作，提出了能够在一次任务中处理多个空间碎片的多碎片主动清理任务。

多碎片主动清理问题可以看作时间依赖型旅行商问题(time-dependent travelling salesman problem, TD-TSP)[10]的变种。给定若干个空间碎片与其对应的轨道信息，任务平台需要按一定的顺序与各空间碎片交会并将其捕获，转移释放到处置轨道；由于空间碎片在太空中的位置随时间变化而不断变化，碎片间的转移消耗与出发时间、转移时长相关，具有时间依赖属性。

现有的多碎片主动清理问题研究常通过优化空间碎片清理序列和转移时间来降低空间绳系组

收稿日期：2023-11-20；**录用日期**：2024-02-23；**网络出版时间**：2024-03-14 12:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240313.1602.003
基金项目：空间智能控制技术实验室开放基金课题 (HTKJ2021KL502010)
* **通信作者**. E-mail：pangzj@njust.edu.cn

引用格式：郭甲，张海博，庞兆君，等. 考虑空间碎片质量的多碎片主动清理任务规划方法[J]. 北京航空航天大学学报，2026，52(1)：275-285. GUO J，ZHANG H B，PANG Z J，et al. Planning method for a multi-debris active removal mission considering space debris mass[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：275-285 (in Chinese).

合体的燃料消耗和任务时长，且已经有大量相关的研究成果。Yang等[11]以任务的收益作为目标函数，使用分支定界法对任务平台转移路径进行了优化。Madakat[12]和Bérend[13]等使用多目标优化算法对燃料消耗和任务时间进行了优化。Yu等[14-15]则使用混杂优化控制理论对多碎片主动清理任务进行了规划。上述研究都是以空间碎片清理序列和转移时间作为任务变量进行优化，当前通过优化上述2种变量来降低代价函数的潜力已经被挖掘殆尽。若要进一步降低任务成本，需要提出新的优化变量。

相较于传统TD-TSP，多碎片主动清理问题有其自身的特殊性和复杂性，这为本文更深层次的优化提供了可能性。在多碎片主动清理任务中，影响任务成本的不只有空间碎片清理序列和转移时间，还有空间碎片的质量，其以平台负载和转移次数的形式影响任务成本，两者相互影响、相互制衡。增加向处置轨道转移的次数可以及时释放空间碎片，降低平台的负载，但过多的转移次数不利于降低成本。与之相反，减少向处置轨道转移的次数会导致空间碎片质量的积累及平台负载过重的问题，亦会导致任务成本提高。已有学者提出了针对空间碎片这类非合作目标的质量测量方式，因此，本文提出将空间碎片质量加入多碎片主动清理的决策框架，以期实现平台负载和转移次数之间的平衡。

现实中待清理的空间碎片数量规模巨大，随着待选空间碎片数目的增加，多碎片主动清理任务的解空间大小与运算负担呈指数增长，为了减少计算负担，在现有关于多碎片主动清理任务的研究中，作为任务目标的空间碎片往往由学者直接给出。为了补充完善任务决策框架，本文提出一种计算量较小的空间碎片挑选方法，该方法基于聚类算法，结合清理LEO空间碎片常用的漂移轨道转移方法，用以从大规模的碎片池中挑选适宜作为任务目标的空间碎片。

综上所述，本文旨在提出一套完整的考虑空间碎片质量影响的多碎片主动清理任务规划方法，使多碎片主动清理问题的决策框架具备更多的选择维度，以此在现有研究基础上进一步降低任务的成本。同时，设置一系列不同空间碎片质量的工况验证规划方法的有效性，研究空间碎片质量对任务规划的影响。

1 任务模型

1.1 任务描述

假设候选空间碎片数目为N，需从中挑选n个空间碎片作为多碎片主动清理任务的目标，空间碎片清理序列记为向量$\boldsymbol{d}=[d(1),d(2),\cdots,d(n)]^{\mathrm{T}}$，其中，$d(i)$表示第$i$个清理的空间碎片的序号。因为未考虑空间碎片质量的影响，现有研究往往采取以下2种策略之一：①任务平台每捕获一个空间碎片，就将其转移到处置轨道并释放，如图1(a)所示，该方案的优点是不会因捕获空间碎片数量积累导致平台负载过重，缺点是任务平台机动次数过多；②任务平台将n个空间碎片全部捕获完成后再统一释放到处置轨道，如图1(b)所示，其优缺点与方案①相反，优点是任务平台机动次数少，缺点是平台负载会随着捕获空间碎片数量积累而增加。为将空间碎片质量计入决策框架，本文提出一种分批捕获空间碎片的策略，将每次释放前需捕获的空间碎片数目作为变量加入任务模型中。假设空间碎片被分为k批进行清理，第j批需捕获的空间碎片数目记为$n_j(1\leqslant j\leqslant k)$。这样有更加多样的空间碎片释放时机进行选择，有利于实现转移次数与平台负载之间的平衡，进一步降低任务成本。图1(c)演示了$k=2$、$n_1=3$、$n_2=1$时的任务流程。

在采用分批捕获空间碎片的策略时，一旦任务平台完成对空间碎片$d(i)$的捕获，需要完成以下3个步骤：

1) 根据n_j的值决定是否向处置轨道转移，若需

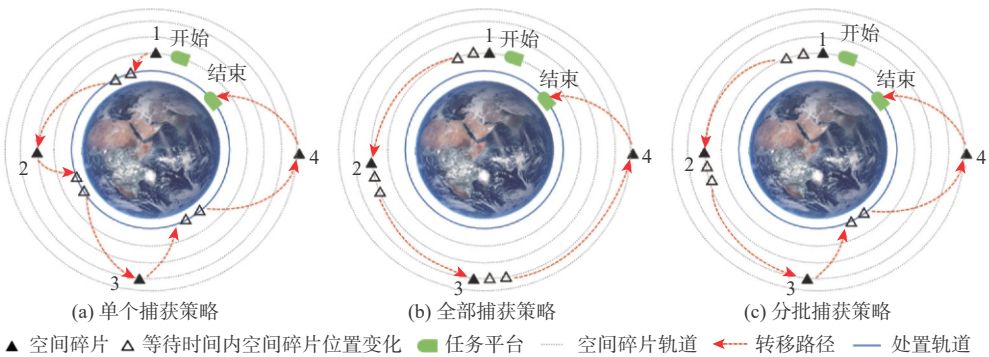


图1 空间碎片捕获策略

Fig. 1 The capture strategy of space debris

要则立即进行转移, 并记转移花费时间为$\Delta t_{\mathrm{h}}(i)$, 若不需要则直接进入步骤 2), 并设$\Delta t_{\mathrm{h}}(i)=0$。

2) 等待转移时机, 若步骤 1)中已向处置轨道转移, 任务平台则在处置轨道上等待, 若未向处置轨道转移, 则在空间碎片$d(i)$的轨道上等待, 等待时间记为$\Delta t_{\mathrm{w}}(i)$, 取值范围为$[\Delta t_{\mathrm{w}}^{\min},\Delta t_{\mathrm{w}}^{\max}]$。

3) 结束等待后, 任务平台开始向空间碎片$d(i+1)$进行转移, 并对其进行捕获, 转移时间记为$\Delta t_{\mathrm{m}}(i)$, 取值范围为$[\Delta t_{\mathrm{m}}^{\min},\Delta t_{\mathrm{m}}^{\max}]$。

完成上述 3 个步骤后, 即可进入新一轮的循环, 直至所有的目标空间碎片被清理完毕。

记t_i为任务平台与空间碎片$d(i)$交会的时刻, 且平台在t_i时刻完成对空间碎片$d(i)$的捕获。假设在任务开始时刻, 任务平台被运载火箭转移到第 1 个需清理的空间碎片处, 并记$t_1=0$。则任务平台捕获空间碎片$d(i+1)$的时刻为

$$t_{i+1}=t_i+\Delta t_{\mathrm{h}}(i)+\Delta t_{\mathrm{w}}(i)+\Delta t_{\mathrm{m}}(i) \tag{1}$$

记任务的总时长为Δt^*, 任务时长上限为$\Delta t_{\max}$, 则有

$$\Delta t^*=\sum_{i=1}^{n-1}\Delta t_{\mathrm{h}}(i)+\sum_{i=1}^{n-1}\Delta t_{\mathrm{w}}(i)+\sum_{i=1}^{n-1}\Delta t_{\mathrm{m}}(i)\leqslant\Delta t_{\max} \tag{2}$$

与其他天体动力学问题不同, 在多碎片主动清理任务中, 平台的干质量会随空间碎片的捕获或释放而发生突变, 使速度增量与燃料消耗不再呈正相关。例如, 已捕获一枚空间碎片的任务平台与空载的任务平台沿相同的转移轨迹进行转移, 两者所需的速度增量相同, 但前者需要消耗更多的燃料。考虑到上述情况, 采用既能反映速度变化也能反映质量变化的冲量增量Δp作为成本函数, 并保持模型的简洁。记平台自身质量为m_{p}, 空间碎片$d(i)$的质量为$m(i)$, 捕获空间碎片$d(i)$后, 平台和空间碎片的总质量为$M(i)$, 则$M(i)$随空间碎片的捕获与释放而变化。任务平台与第 1 个空间碎片交会前有

$$M(0)=m_{\mathrm{p}} \tag{3}$$

在空间碎片$d(i)$被捕获后有

$$M(i)=M(i-1)+m(i) \tag{4}$$

使用霍曼变轨方法转移到处置轨道消耗的速度增量记为$\Delta v_{\mathrm{h}}(i)$, 对应的冲量增量为

$$\Delta p_{\mathrm{h}}(i)=M(i)\Delta v_{\mathrm{h}}(i) \tag{5}$$

在处置轨道释放空间碎片后, $M(i)$重置为

$$M(i)=m_{\mathrm{p}} \tag{6}$$

若空间碎片$d(i)$在被捕获后并未被立即转移到处置轨道, 则令$\Delta p_{\mathrm{h}}(i)=0$, $M(i)$保持不变。在转移至空间碎片$d(i+1)$期间, 消耗的速度增量和冲量增量分别记为$\Delta v_{\mathrm{m}}(i)$和$\Delta p_{\mathrm{m}}(i)$, 则有

$$\Delta p_{\mathrm{m}}(i)=M(i)\Delta v_{\mathrm{m}}(i) \tag{7}$$

设任务消耗的冲量增量总量为Δp^*, 其上限为$\Delta p_{\max}$, 则有

$$\Delta p^*=\sum_{i=1}^{n}\Delta p_{\mathrm{h}}(i)+\sum_{i=1}^{n-1}\Delta p_{\mathrm{m}}(i)\leqslant\Delta p_{\max} \tag{8}$$

多碎片主动清理任务以最小化任务消耗总时长和总冲量增量为目标, 对 2 个目标函数进行归一化处理。采用分批清理空间碎片策略的多目标优化问题设计变量为

$$\begin{aligned}\boldsymbol{\Gamma}=[&n_1,\cdots,n_k,d(1),\cdots,d(n),\Delta t_{\mathrm{w}}(1),\cdots,\\&\Delta t_{\mathrm{w}}(n-1),\Delta t_{\mathrm{m}}(1),\cdots,\Delta t_{\mathrm{m}}(n-1)]\end{aligned} \tag{9}$$

优化问题模型可以表示为

$$\begin{cases}\min\quad f_1=\dfrac{\Delta t^*}{\Delta t_{\max}}\\ \min\quad f_2=\dfrac{\Delta p^*}{\Delta p_{\max}}\\ \text{s.t.}\begin{cases}n_j\in\{1,2,\cdots,n\}\\ \displaystyle\sum_{j=1}^{k}n_j=n\\ d(i)\in\{1,2,\cdots,n\}\qquad i=1,2,\cdots,n\\ \Delta t_{\mathrm{w}}(i)\in\left[\Delta t_{\mathrm{w}}^{\min},\Delta t_{\mathrm{w}}^{\max}\right]\qquad i=1,2,\cdots,n-1\\ \Delta t_{\mathrm{m}}(i)\in\left[\Delta t_{\mathrm{m}}^{\min},\Delta t_{\mathrm{m}}^{\max}\right]\qquad i=1,2,\cdots,n-1\\ \Delta t^*\leqslant\Delta t_{\max}\\ \Delta p^*\leqslant\Delta p_{\max}\end{cases}\end{cases} \tag{10}$$

1.2　漂移轨道转移方法

目前, 有多种方法来计算空间碎片之间转移的消耗, 如 Lambert 转移[12] 和小推力转移[16]。近年来, 学者们多采用 Cerf 提出的漂移轨道转移方法[17], 因为该方法具有以下优点: ①利用 J2 项摄动引起的轨道升交点赤经漂移来实现交会, 有效降低了燃料消耗; ②计算过程中不存在数值积分和解方程的过程, 计算效率高。

假设任务平台将要前往空间碎片$d(i+1)$所在的轨道, 使用漂移轨道转移方法进行转移的步骤如下:

1) 从任务平台当前轨道转移至漂移轨道。

2) 在漂移轨道上等待, 直至漂移轨道的升交点赤经与空间碎片$d(i+1)$所在轨道的升交点赤经相等。

3) 从漂移轨道转移至空间碎片$d(i+1)$所在轨道。

步骤 1)和步骤 3)中采用霍曼转移方法进行转移, 一般霍曼转移仅对轨道高度进行调整, 但任务平台所在轨道与空间碎片$d(i+1)$所在轨道往往存

在倾角差异，为此，Cerf提出将轨道高度与轨道倾角的调整进行合并，如图2所示，所需的速度增量可按式(11)计算[17]。此外，一旦升交点赤经调整完毕，任务平台与空间碎片之间的相位差能够快速校正，交会成本可以忽略不计[17]。

$$\Delta v = \sqrt{v_s^2+v_e^2-2v_sv_e\cos(\Delta I)} \tag{11}$$

式中：v_s为调整前的速度；v_e为调整后的速度；ΔI为轨道倾角变化量。

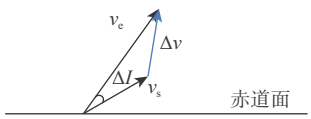


图2 轨道高度与倾角同时调整

Fig. 2 Synchronous adjustment of orbit attitude and inclination

记升交点赤经为Ω，在地球非球形摄动的影响下，升交点赤经的漂移速率$\dot{\Omega}$计算如下[17]：

$$\dot{\Omega} = -C_{J_2}\frac{\cos I}{a^{7/2}(1-e^2)^2} \tag{12}$$

式中：a为轨道半长轴；e为轨道偏心率；I为轨道倾角；C_{J_2}为常数，具体为[17]

$$C_{J_2} = \frac{3}{2}J_2\sqrt{\mu_E}R_E^2 \tag{13}$$

其中：$J_2 = 1.082\,6\times10^{-3}$为地球扁率的谐波常数；$R_E$ = 6 378 137 m为地球半径；$\mu_E = 3\,986\,005\times10^8$ m^3/s^{2}为地球引力常数。

在任务平台开始向空间碎片$d(i+1)$进行转移时，记任务平台所在轨道的升交点赤经为$\Omega_p(t_i+\Delta t_h(i)+\Delta t_w(i))$，在同一时刻，空间碎片$d(i+1)$所在轨道的升交点赤经记为$\Omega_{i+1}(t_i+\Delta t_h(i)+\Delta t_w(i))$，则漂移轨道的漂移速率$\dot{\Omega}_d$应满足以下约束[17]：

$$\begin{aligned}&\Omega_p(t_i+\Delta t_h(i)+\Delta t_w(i))+\Delta t_m(i)\cdot\dot{\Omega}_d=\\&\quad\Omega_{i+1}(t_i+\Delta t_h(i)+\Delta t_w(i))+\Delta t_m(i)\cdot\dot{\Omega}_{i+1}\end{aligned} \tag{14}$$

式中：$\dot{\Omega}_{i+1}$为空间碎片$d(i+1)$所在轨道的升交点赤经漂移速率。

求得$\dot{\Omega}_d$后，漂移轨道的轨道倾角I_d有2种选择：任务平台所在轨道的倾角I_p和空间碎片$d(i+1)$的轨道倾角I_{i+1}，因为轨道倾角确定后，漂移轨道的半长轴也随之确定，可依据上述方法计算出转移所需的速度增量，根据两者速度增量大小，可选择出合适的漂移轨道的轨道倾角I_d及半长轴a_d。

任务平台从目前所在轨道转移到转移轨道，再从漂移轨道转移到空间碎片$d(i+1)$所在轨道需2次霍曼转移，期间发生4次脉冲机动，选4次机动中v_s最小的一次结合倾角调整，并按式(11)计算速度增量，其余3次机动按照正常霍曼变轨计算公式计算速度增量，设4次机动的速度增量为$\Delta v_{m1}(i)$、$\Delta v_{m2}(i)$、$\Delta v_{m3}(i)$、$\Delta v_{m4}(i)$，则式(7)可改写为

$$\Delta p_m(i) = M(i)(\Delta v_{m1}(i)+\Delta v_{m2}(i)+\Delta v_{m3}(i)+\Delta v_{m4}(i)) \tag{15}$$

2 任务规划方法

多碎片主动清理任务规划包括挑选空间碎片和空间碎片清理序列规划2方面内容：①挑选空间碎片，即在大规模的候选碎片池中挑选出一组空间碎片，要求一段时间内在这组空间碎片中任选一个碎片都可以保证转移到其他任一组内空间碎片的代价是可接受的；②碎片清理序列规划是决定空间碎片清理的先后顺序和时间节点，以降低整个多碎片主动清理任务的消耗。

2.1 聚类分析

聚类分析是一种将数据对象的集合分组为由类似的对象组成的多个类的分析过程。通过空间碎片的轨道要素可以定性地分析转移难度，一般轨道要素越相近，转移代价越小，以空间碎片间转移难度作为聚类分析的依据，聚类的过程就可以看作挑选空间碎片的过程，聚类的结果就是可作为任务目标的一组空间碎片。通过聚类分析实现了挑选空间碎片和任务方案优化的解耦，有利于减少规划的计算量。

DBSCAN算法是一种基于密度的聚类算法。在数据空间中给定一组点，将密集在这些点周围的点进行分组，而与所有这些点都远离的点标记为噪声点。该算法有2个基本参数：距离阈值ε_0和邻域样本数阈值δ，在多碎片主动清理任务中，数据点间距离ε表示空间碎片间的转移难度，δ表示在平台到达任意一颗空间碎片后，在其周围还有多少个在燃料消耗允许范围内潜在的下一清理目标[18]。图3为对二维平面内的数据点进行DBSCAN聚类产生的一个簇，其中，核点和边界点属于簇内数据，噪点则不属于这一簇。

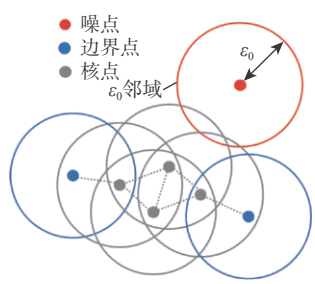


图3 对平面内数据点进行聚类分析的结果($\delta=3$)

Fig. 3 The results of clustering analysis on data points in a plane ($\delta=3$)

解云鹤提出使用相位指示器来初步评估空间碎片间的转移难度[18]，常见的相位指示器包括欧式指示器和轨道指示器。欧式指示器是2个空间碎

片相对位置和相对速度的差值; 轨道指示器在短时间内将 2 个空间碎片和航天器的运动近似为匀速直线运动, 指示器的 2 个坐标分别是转移开始时与空间碎片 1 的相对速度和转移结束时与空间碎片 2 的相对速度。但使用漂移轨道转移方法的转移代价主要取决于 2 个空间碎片轨道半长轴a、偏心率e、轨道倾角I和升交点赤经Ω的差别, 而非 2 个空间碎片相对位置或相对速度的差值。例如, 在同一个轨道不同位置上的空间碎片的位置矢量和速度矢量相差甚远, 但按照漂移轨道转移方法已无需再进行转移。因此, 有必要确定一种适用于漂移轨道转移方法的评价空间碎片间转移难度的量。

因为空间碎片所在的轨道多为近圆轨道, 调整偏心率e的代价可忽略, 评价空间碎片间转移难度的量可以设定为与轨道倾角I、升交点赤经Ω和轨道半长轴a的差值呈正相关的值, 且要求该值可以体现不同轨道根数的调整难度, 因为调整Ω和I花费的代价要明显高于a。根据上述要求, 可以将该量设为调整a、I和Ω的脉冲速度增量。使用脉冲速度增量调整半长轴a即霍曼转移所需的速度增量Δv_a为[19]

$$\Delta v_{a_1}=\sqrt{\frac{\mu_{\mathrm{E}}}{a_1}\cdot\frac{2a_2}{a_1+a_2}}-\sqrt{\frac{\mu_{\mathrm{E}}}{a_1}} \tag{16}$$

$$\Delta v_{a_2}=\sqrt{\frac{\mu_{\mathrm{E}}}{a_2}}-\sqrt{\frac{\mu_{\mathrm{E}}}{a_2}\cdot\frac{2a_1}{a_1+a_2}} \tag{17}$$

$$\Delta v_a=\left|\Delta v_{a_1}\right|+\left|\Delta v_{a_2}\right| \tag{18}$$

式中: Δv_{a_1}和Δv_{a_2}分别为霍曼转移第 1 次和第 2 次脉冲所需速度增量; a_1为调整前轨道半长轴; a_2为调整后轨道半长轴。

调整轨道倾角时, 可在原轨道与预定轨道相交的节点处施加脉冲, 使原轨道绕节线旋转ΔI, 所需的速度增量Δv_I为[19]

$$\Delta v_I=\sqrt{v_1^2+v_2^2-2v_1v_2\cos\Delta I} \tag{19}$$

式中: v_1为调整前任务平台速度大小; v_2为调整后任务平台速度大小。若仅调整轨道倾角, v_1与v_2相等。

调整轨道升交点赤经时, 可在纬度辐角为 90°或 270°时进行脉冲机动, 升交点赤经变化量$\Delta\Omega$对应的脉冲速度增量Δv_Ω计算公式为[19]

$$\Delta v_\Omega=\sqrt{v_1^2+v_1^2-2v_1v_1\cos\Delta\Omega}\cdot\sin I_1 \tag{20}$$

式中: I_1为调整前任务平台所在轨道倾角。

在 DBSCAN 聚类分析中, 空间碎片之间的转移难度评价指标为

$$\varepsilon=w_1(\Delta v_a+\Delta v_I)+w_2\Delta v_\Omega \tag{21}$$

式中: w_1、w_2为权重系数。因为采用了漂移轨道转移方法, 调整轨道倾角和轨道半长轴主要影响冲量增量, 调整升交点赤经则主要影响任务时间, 通过调整w_1和w_2, 可以挑选出更有利于减小冲量增量或任务时长的目标。

2.2　多目标优化

多碎片主动清理任务方案的优化是一个多目标优化问题, 同时也是混合优化问题。多目标优化需要同时优化任务花费的冲量增量和任务总时长 2 个目标, 混合优化问题则体现在优化变量中空间碎片清理序列$\boldsymbol{d}$和释放前捕获空间碎片数目n_j是整数变量, 而等待时间$\Delta t_{\mathrm{w}}(i)$和转移时间$\Delta t_{\mathrm{m}}(i)$则是连续变量。

为了处理多目标优化问题, 采用第二代非支配遗传算法(non-dominated sorting genetic algorithms, NSGA-Ⅱ)[20], 其降低了非劣排序遗传算法的复杂性, 具有运行速度快、解集的收敛性好的优点, 其迭代过程如图 4 所示, NSGA-Ⅱ算法参数的选择以优化得到的结果稳定在一定范围内为准。

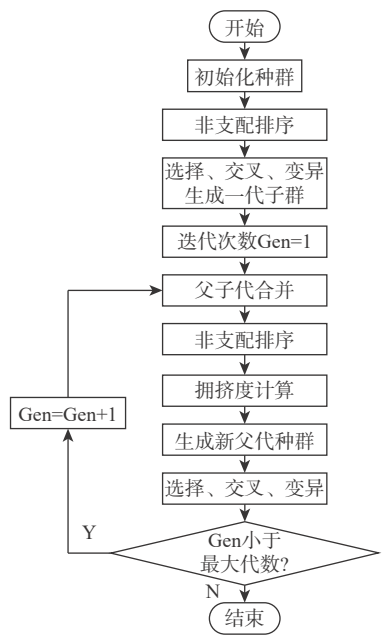


图 4　NSGA-Ⅱ迭代过程

Fig. 4　The iterative process of NSGA-Ⅱ

为了适应混合优化问题, 将多参数级联编码方法与 NSGA-Ⅱ配合使用。多参数级联编码方法是将各参数分别以某种基本编码方法进行编码, 再将其编码按照一定的顺序连接在一起组成表示全部参数的个体编码。需捕获的空间碎片数目n_j转化为n位的二进制编码串, 每一批最后 1 个空间碎片编码为 1, 其余为 0; 空间碎片清理序列$\boldsymbol{d}$是由空间碎片序号组成的向量, 可以直接采用符号编码方法, 即向量$\boldsymbol{d}$本身就是对应的编码, 其符合有意义的积木块编码原则; 对于连续变量出发时间$\Delta t_{\mathrm{m}}(i)$和等待时间$\Delta t_{\mathrm{w}}(i)$, 使用二进制编码难以满足精度要

求，因此，可以采用浮点数编码方法，即个体的基因值用某一范围内的一个浮点数来表示，这样可以改善遗传算法的计算复杂性，提高运算效率。

在未考虑空间碎片质量时，决策框架不包含n_j，多目标优化得到一组 Pareto 解后即优化结束。在考虑空间碎片质量后，决策框架加入n_j，每个n_j的取值方案都会对应产生一组 Pareto 解，因此，在得到多组 Pareto 解后，还需评价对比 Pareto 解以确定n_j的取值方案的优劣。

采用 Hypervolume 指标来评价不同 Pareto 解的优劣。Hypervolume 指标的计算方法如下[21]：给定一个 Pareto 前沿，并设置参考点 H，Pareto 解与参考点 H 围成阴影区域的面积即该 Pareto 解对应的 Hypervolume 指标，如图 5 所示，Hypervolume 指标越大，对应的 Pareto 解越优。

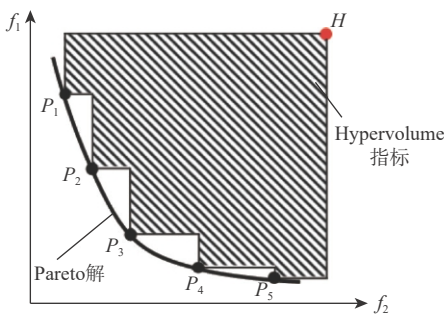


图 5 Hypervolume 指标示意图

Fig. 5 Diagram of the Hypervolume indicator

3 仿真实验

处置轨道设为轨道高度为 400 km 的圆轨道，在该高度的空间碎片会逐渐减速并在 25 年内再入大气层燃烧殆尽[13]，满足 IADC 发布的 *IADC space debrits mitigation guidelines*[22] 的要求。

3.1 计算模型验证

为了验证分批捕获策略的可行性及计算模型的正确性，选取文献 [17] 中的算例进行对比。算例以最小化总速度增量为目标，优化得到的总速度增量为 500.7 m/s。将文献 [17] 所得方案代入本文采用的计算模型，得到的总速度增量与文献 [17] 一致，可判断本文采用的计算模型正确。

设单个空间碎片质量为 1 000 kg，任务平台质量为 3 000 kg，任务时长上限为 730 d，冲量增量上限为6×10^7 kg·m/s。计算方案对应的 2 个目标函数值，将其与 NSGA-Ⅱ算法得到的结果进行比较，如图 6 所示。代表文献 [17] 结果的数据点在 Pareto 解集右侧，即 Pareto 解集中存在解对文献 [17] 结果 Pareto 支配，取 Pareto 解中距离文献 [17] 结果最近的数据点进行对比，如表 1 所示。可见，分批捕获策略得到的结果取得了更低的冲量增量及任务总时长，即分批捕获策略是可行的。

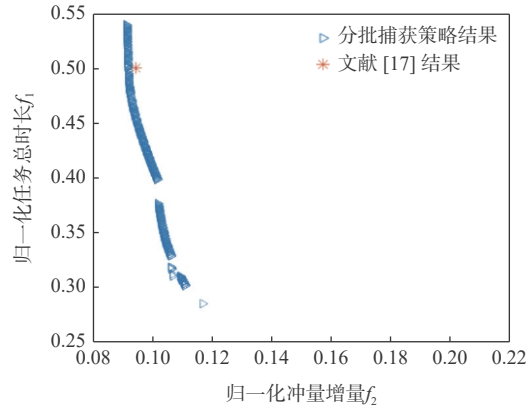


图 6 单个空间碎片质量为 1 000 kg 时的结果对比

Fig. 6 Comparison of results when a single space debris mass is 1000 kg

表 1 优化结果对比

Table 1 Comparison of optimization results

方法	冲量增量f_2	任务总时长f_1
文献[17]	0.094 4	0.501 4
分批捕获策略	0.091 7	0.501 2

3.2 铱星-33 碎片云清理仿真

为了模拟真实的空间碎片轨道，借用铱星-33 碎片云的轨道信息进行仿真实验，以测试包含变量n_j的多碎片主动清理任务的任务模型，并探究空间碎片质量对任务决策的影响。铱星-33 碎片云包含 320 个空间碎片，轨道根数信息来自文献 [23]。任务计划清理 10 个空间碎片，即$N=320$，$n=10$。在任务初始时刻，假设任务平台的轨道根数与空间碎片清理序列中第 1 个空间碎片的轨道根数相同。

3.2.1 聚类结果分析

依据铱星-33 碎片云的轨道信息进行 DBSCAN 聚类分析，距离阈值$\varepsilon_0=80$ m/s，邻域样本数阈值δ=3，权重系数$w_1=1, w_2=1$。结果显示，320 个空间碎片被分为 33 个簇，前 3 簇的结果如表 2 所示，以 1 号簇中的前 10 个空间碎片为任务对象。被选中的空间碎片的轨道根数如表 3 所示。

表 2 DBSCAN 聚类分析部分结果

Table 2 Partial results of DBSCAN cluster analysis

簇代号	簇内空间碎片序号
1	1、21、27、31、36、108、125、167、210、248、300、317
2	3、39、55、86、103、132、133、134、137、152、153、320
3	4、59、123、183、217
⋮	⋮

图 7 绘制了铱星-33 碎片云的分布及被选中的碎片，结合表 3 中的信息可以发现，被选中的空间

表 3　目标空间碎片轨道根数

Table 3　The orbital elements of the target space debrits

空间碎片序号	轨道半长轴 a/m	偏心率 e	轨道倾角 I/rad	升交点赤经 Ω/rad
1	$7.158\,0\times10^{6}$	0.8×10^{-3}	1.507 7	5.308 4
21	$7.154\,0\times10^{6}$	2.5×10^{-3}	1.507 6	5.321 9
27	$7.140\,8\times10^{6}$	1.1×10^{-2}	1.506 4	5.351 7
31	$7.159\,1\times10^{6}$	1.6×10^{-3}	1.507 6	5.313 8
36	$7.143\,1\times10^{6}$	2.5×10^{-3}	1.507 8	5.335 5
108	$7.091\,6\times10^{6}$	6.0×10^{-3}	1.507 4	5.336 4
125	$7.231\,1\times10^{6}$	1.5×10^{-2}	1.504 6	5.320 6
167	$7.151\,7\times10^{6}$	1.8×10^{-3}	1.507 7	5.348 5
210	$7.155\,0\times10^{6}$	2.0×10^{-3}	1.507 5	5.327 3
248	$7.088\,9\times10^{6}$	1.3×10^{-3}	1.508 4	5.307 4

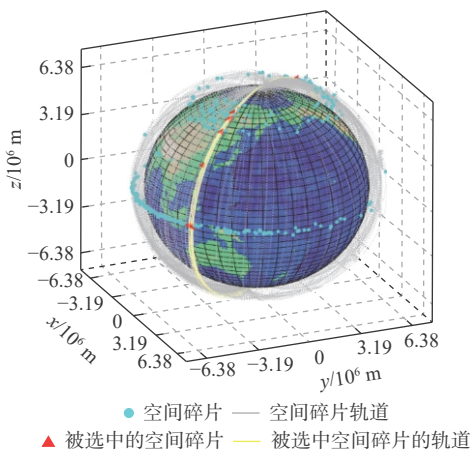


图 7　被选中的空间碎片

Fig. 7　The selected space debris

碎片的轨道面十分接近。使用漂移轨道转移方法时, 2 个轨道间半长轴 a 和轨道倾角 I 的差距需通过消耗燃料机动进行调整, 升交点赤经 Ω 的差距则需要消耗时间进行等待来调整。故被选中的空间碎片半长轴 a 和轨道倾角 I 相近有利于减少任务消耗的总冲量增量 Δp^{*}, 升交点赤经 Ω 相近有利于减少任务消耗时间 Δt^{*}, 被选中的空间碎片半长轴 a、轨道倾角 I 和升交点赤经 Ω 差距最大分别为 1.97%、0.25%、0.83%, 不超过 2%, 可以认为 DBSCAN 聚类分析选择出了合理的空间碎片。

3.2.2　NSGA-Ⅱ 结果分析

本节设计了 2 个仿真实验: 实验 1 设置了一系列固定的空间碎片质量 $m(i)$ 和单批需捕获的空间碎片数目 n_j, 减少随机变量, 有利于总结出 $m(i)$ 和 n_j 改变引起代价函数变化的规律; 实验 2 将针对不同的 $m(i)$ 优化给出对应的 n_j, 更贴近于实际情况, 也可验证提出的分批捕获策略是否有效。两者共同的参数设定如下: 任务时长上限为 2 年, 即 $\Delta t_{\max}=730$ d, 冲量增量上限设为初始种群中冲量增量的最大值。图 8 为半长轴转移时间 $\Delta t_{\mathrm{m}}(i)$、升交点赤经变化量 $\Delta\Omega$ 与转移所需速度增量 Δv 之间的关系。可以发现, 当转移时间 $\Delta t_{\mathrm{m}}(i)$ 小于 50 d 时, 转移所需的速度增量 Δv 急剧上升, 因此, 设转移时间下限 $\Delta t_{\mathrm{m}}^{\min}(i)=50$ d, 其余的变量约束条件如表 4 所示。NSGA-Ⅱ 算法的参数如表 5 所示。

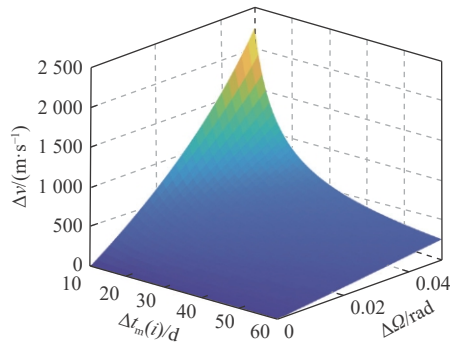


图 8　$\Delta t_{\mathrm{m}}(i)$、$\Delta\Omega$ 与 Δv 之间的关系

Fig. 8　The relationship between $\Delta t_{\mathrm{m}}(i)$, $\Delta\Omega$ and Δv

表 4　变量的约束条件

Table 4　Constraints on variables

约束条件	数值
冲量增量上限 $\Delta p_{\max}$/(kg·m·s^{-1})	8×10^{7}
任务时长上限 $\Delta t_{\max}$/d	730
$\Delta t_{\mathrm{w}}(i)$ 的取值范围/d	[0,10]
$\Delta t_{\mathrm{m}}(i)$ 的取值范围/d	[50,70]

表 5　NSGA-Ⅱ 的参数

Table 5　Parameters of NSGA-Ⅱ

参数	数值
种群大小	100
终止条件	迭代次数≥200
交叉率/%	70
变异率/%	10

任务平台的质量统一设为 $m_{\mathrm{p}}=3\,000$ kg, 任务模型作了适当简化, 暂未考虑任务平台载重上限。

1) 实验 1。固定 $m(i)$ 和 n_j。为了探究空间碎片质量和单批需捕获的空间碎片数目对多碎片主动清理任务的影响, 设置了单个空间碎片质量 $m(i)$ 固定为 1 000 kg、3 000 kg、5 000 kg、7 000 kg、9 000 kg 的 5 种工况, 每种工况下都计算了 $n_j=1$, 2, 3, 4, 5, 8, 10 情况下的 Pareto 解, 当清理过一批空间碎片后, 剩余的空间碎片数目达不到 n_j 的要求时, 只把剩余的空间碎片清理掉即可。计算 Hypervolume 指标时使用的参考点 H 坐标设为(1,1), 每个 Pareto 解的 Hypervolume 指标汇总在表 6 中。

传统的多碎片主动清理决策框架仅在 $n_j=1$ 或 $n_j=10$ 的条件下对空间碎片清理序列和转移时间进行优化。根据表 6 中的数据可以发现, $n_j=10$ 仅在单个空间碎片质量为 1 000 kg 时为最优方案,

表 6 各 Pareto 解的 Hypervolume 指标

Table 6 Hypervolume indicators for each Pareto solution

n_j	单个空间碎片质量为 1 000 kg	单个空间碎片质量为 3 000 kg	单个空间碎片质量为 5 000 kg	单个空间碎片质量为 7 000 kg	单个空间碎片质量为 9 000 kg
1	0.330 2	0.309 3	0.291 2	0.272 4	0.253 8
2	0.342 9	0.320 0	0.293 5	0.281 5	0.253 3
3	0.338 8	0.319 8	0.292 5	0.273 8	0.262 9
4	0.349 5	0.315 9	0.296 7	0.271 3	0.253 5
5	0.344 9	0.324 1	0.288 0	0.276 8	0.245 1
8	0.342 0	0.314 2	0.294 7	0.268 0	0.230 8
10	0.349 6	0.319 5	0.288 7	0.248 6	0.210 7

n_j = 1则在本文所列的所有单个空间碎片质量条件下都并非最优的n_j取值方案。即相较于传统的多碎片主动清理优化方法，通过增加变量n_j可以进一步降低任务成本。

单个空间碎片质量偏小时，如图 9 中$m(i)$ = 1 000 kg的工况所示，总的趋势是每次向处置轨道转移前捕获的空间碎片越多，Hypervolume 指标越优，即在总负载处于较低水平时，影响任务指标的因素主要是转移次数。

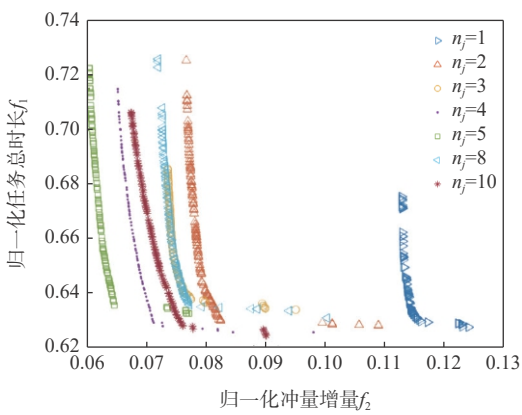


图 9 单个空间碎片质量为 1 000 kg 时的 Pareto 解

Fig. 9 Pareto solution for a single space debris mass of 1 000 kg

单个空间碎片质量偏大时，如图 10 中$m(i)$=9 000 kg的工况所示，总的趋势是每次离轨前捕获的空间碎片越少，Hypervolume 指标越优，此时负载累积对任务的影响大于转移次数，减少转移次数带来的优势无法抵消负载增加带来的劣势。

图 11 为不同单个空间碎片质量的情况下获得最大 Hypervolume 指标的n_j选择。可以看出，随着单个空间碎片质量的增加，最优的n_j总体呈减小趋势。在空间碎片质量较小时，最优的空间碎片清理方案更倾向于一次捕获更多空间碎片，以此减少向处置轨道转移的次数；在空间碎片质量较大时，最优的空间碎片清理方案则更倾向于及时释放空间碎片以减少转移过程中的负载积累。

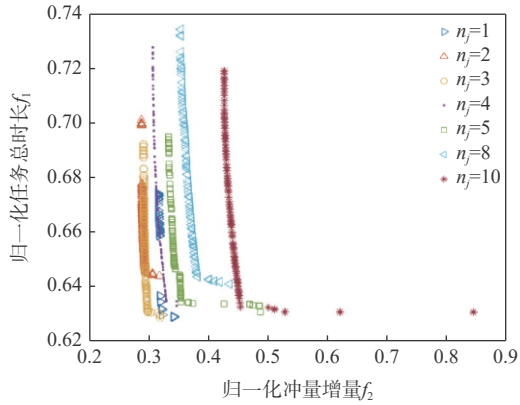


图 10 单个空间碎片质量为 9 000 kg 时的 Pareto 解

Fig. 10 Pareto solution for a single space debris mass of 9 000 kg

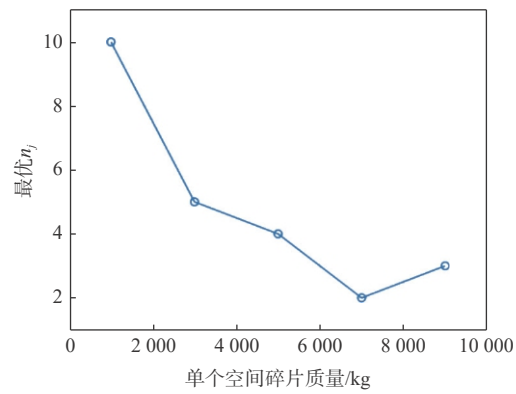


图 11 不同单个空间碎片质量情况下最优的n_j选择

Fig. 11 Optimal n_j for different single space debris mass

2）实验 2。非固定$m(i)$和n_j。在工程实际中，空间碎片质量不定，需根据不同的情况优化给出单批捕获的空间碎片数目。实验随机生成了一组空间碎片的质量，如表 7 所示。

表 7 空间碎片质量

Table 7 Mass of space debris

空间碎片序号	空间碎片质量/kg
1	4 691
21	8 113
27	80
31	6 567
36	2 358
108	8 182
125	7 042
167	1 009
210	1 243
248	7 294

优化加入n_j固定的几种情况作为对比，得到的结果如图 12 所示，对应的 Hypervolume 指标如表 8 所示。以图 12 中n_j不固定方案中取得最小任务时长的解为例，其解的具体内容如表 9 所示，该解将

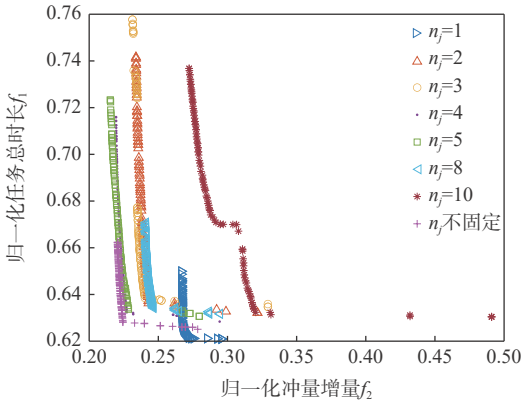


图 12　空间碎片质量不定时的 Pareto 解

Fig. 12　Pareto solutions for uncertain single space debris mass

表 8　Hypervolume 指标汇总

Table 8　Summary of Hypervolume indicators

n_j	Hypervolume指标
1	0.277 9
2	0.281 3
3	0.280 2
4	0.289 6
5	0.289 2
8	0.279 4
10	0.266 7
不固定	0.292 0

表 9　所选解对应的空间碎片清理方案

Table 9　The space debris removal strategy corresponding to the selected solution

清理顺序	批次	空间碎片序号	等待时间/d	转移时间/d
1	1	108	3	50
2		27	5.87	51.09
3	2	210	1.15	52.02
4		167	0	50
5		248	1.9	50.86
6	3	125	2.09	50.04
7	4	1	4.69	51.42
8		21	7.10	50.38
9		31	5.11	50.34
10		36		

空间碎片分为 4 个批次进行清理, 其中, $n_1=2, n_2=3$, $n_3=1, n_4=4$。在第 1 组空间碎片质量条件下, n_j不固定方案取得的 Hypervolume 指标相较于传统的 $n_j=1$和 $n_j=10$方案分别提升了 5.07% 和 9.49%。可以看到, 在n_j不固定的情况下, 任务的变量增多, 可供的选择也更多, 方案的制定更加灵活, 通过计入空间碎片质量影响实现了对多碎片主动清理任务的深度优化。

若需要在 NSGA-Ⅱ 生成的 Pareto 解集中进行选择, 可以从任务需求的角度出发来进行决策。例如, 所清理的空间碎片附近有正常工作的卫星, 为了尽快消除与空间碎片发生碰撞的危险, 应选择任务总时长较短的方案; 若所清理的轨道碎片威胁不大, 对任务时长没有要求, 则可以选择冲量增量较小的方案; 若对任务总时长和冲量增量都有一定的需求, 则可以在 Pareto 解集的基础上通过设置权重并求和的方式来进一步评价 Pareto 解集中各个解的优劣, 进而从一系列解中选择出符合当前权重系数的最优解。

4　结　论

1) 使用聚类算法可以从大规模空间碎片中选出轨道面接近、适宜采用漂移轨道转移方法的任务目标, 实现了空间碎片选择的功能。

2) 设置了一系列的质量梯度来测试空间碎片质量对多碎片主动清理任务的影响, 结果显示, 当空间碎片质量较小时, 选择较大的n_j能够取得较好的效果, 因为单个空间碎片质量较轻, 减少转移次数带来的优势超过了负载累积带来的劣势, 与之相反, 当空间碎片质量较大时, 负载累积增长过快, 减少转移次数带来的优势无法抵消负载累积带来的劣势, 应选择较小的n_j以及时减轻任务平台的负载。

3) 在空间碎片质量各不相同的情况下, 在本文提出的决策框架下得到的结果相较于传统决策框架下的优化结果 Hypervolume 指标有 5%～9% 的提升, 即通过优化n_j可以进一步降低任务成本。

参考文献（References）

[1] 张耀, 刘静. 基于角度转化的空间碎片测角数据精密定轨方法[J]. 北京航空航天大学学报, 2023, 49(7): 1600-1605.
ZHANG Y, LIU J. Precise orbit determination method for angle-only observation data of space debris based on angle conversion theory[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(7): 1600-1605(in Chinese).

[2] 吴树范, 王伟, 温济帆, 等. 低轨互联网星座发展研究[J]. 北京航空航天大学学报, 2024, 50(1): 1-11.
WU S F, WANG W, WEN J F, et al. Review on development of LEO internet constellation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(1): 1-11(in Chinese).

[3] ZHU W L, SI J Y, PANG Z J, et al. Rapid deployment and continuous shape maintenance of tethered-space net robot based on single-pulse action[J]. Advances in Space Research, 2021, 67(5): 1477-1489.

[4] 郭甲, 庞兆君, 岳帅, 等. 空间绳系组合体的继电型控制离轨策略[J]. 航空学报, 2021, 42(12): 324738.
GUO J, PANG Z J, YUE S, et al. De-orbit strategy using Bang-Bang control for space tethered-combination[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(12): 324738(in Chinese).

[5] ZHAO W, PANG Z J, ZHAO Z, et al. A simulation and an experimental study of space harpoon low-velocity impact, anchored debris[J]. Materials, 2022, 15(14): 5041.

[6] 马晓刚. 基于离子束的非接触式空间碎片清除方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017: 1-7.
MA X G. A study of contactless space debris removal based on ion-beam-propelled method[D]. Harbin: Harbin Institute of Technology, 2017: 1-7(in Chinese).

[7] 张品亮, 龚自正, 杨武霖, 等. 激光移除空间碎片过程的三维仿真与建模[J]. 宇航学报, 2017, 38(3): 323-330.
ZHANG P L, GONG Z Z, YANG W L, et al. Three-dimensional simulation and modeling on removing orbital debris with lasers[J]. Journal of Astronautics, 2017, 38(3): 323-330(in Chinese).

[8] FORSHAW J L, AGLIETTI G S, FELLOWES S, et al. The active space debris removal mission RemoveDEBRIS. Part 1: from concept to launch[J]. Acta Astronautica, 2020, 168: 293-309.

[9] AGLIETTI G S, TAYLOR B, FELLOWES S, et al. The active space debris removal mission RemoveDEBRIS. Part 2: in orbit operations[J]. Acta Astronautica, 2020, 168: 310-322.

[10] 方伟, 接中冰, 陆恒杨, 等. 基于多起点和 Mask 策略的深度强化学习算法求解覆盖旅行商问题[J]. 控制与决策, 2024, 39(4): 1160-1166.
FANG W, JIE Z B, LU H Y, et al. Deep reinforcement learning algorithm based on multi-start and Mask strategy for solving the covering salesman problem[J]. Control and Decision, 2024, 39(4): 1160-1166(in Chinese).

[11] YANG J N, HU Y H, LIU Y, et al. A maximal-reward preliminary planning for multi-debris active removal mission in LEO with a greedy heuristic method[J]. Acta Astronautica, 2018, 149: 123-142.

[12] MADAKAT D, MORIO J, VANDERPOOTEN D. Biobjective planning of an active debris removal mission[J]. Acta Astronautica, 2013, 84: 182-188.

[13] BÉREND N, OLIVE X. Bi-objective optimization of a multiple-target active debris removal mission[J]. Acta Astronautica, 2016, 122: 324-335.

[14] YU J, CHEN X Q, CHEN L H. Optimal planning of LEO active debris removal based on hybrid optimal control theory[J]. Advances in Space Research, 2015, 55(11): 2628-2640.

[15] YU J, CHEN X Q, CHEN L H, et al. Optimal scheduling of GEO debris removing based on hybrid optimal control theory[J]. Acta Astronautica, 2014, 93: 400-409.

[16] JORGENSEN M K, SHARF I. Optimal planning for a multiple space debris removal mission using high-accuracy low-thrust transfers[J]. Acta Astronautica, 2020, 172: 56-69.

[17] CERF M. Multiple space debris collecting mission: debris selection and trajectory optimization[J]. Journal of Optimization Theory and Applications, 2013, 156(3): 761-796.

[18] 解云鹤. 基于机器学习的空间多碎片清除策略研究[D]. 哈尔滨: 哈尔滨工业大学, 2019: 13-19.
XIE Y H. Space debris removal strategy based on machine learning[D]. Harbin: Harbin Institute of Technology, 2019: 13-19(in Chinese).

[19] 于志坚. 航天测控系统工程[M]. 北京: 国防工业出版社, 2008: 44-45.
YU Z J. Aerospace TT&C system engineering[M]. Beijing: National Defense Industry Press, 2008: 44-45(in Chinese).

[20] 翟禹尧, 史贤俊, 杨帅, 等. 不可靠测试条件下基于 NSGA-Ⅱ的多目标测试优化选择[J]. 北京航空航天大学学报, 2021, 47(4): 792-801.
ZHAI Y Y, SHI X J, YANG S, et al. Multi-objective test optimization selection based on NSGA-Ⅱ under unreliable test conditions [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(4): 792-801(in Chinese).

[21] 李珂. Hypervolume 指标及其在多目标进化算法中的应用研究[D]. 湘潭大学, 2010: 9-15.
LI K. The research on Hypervolume indicator and its application in multi-objective evolutionary algorithm[D]. Xiangtan: Xiangtan University, 2010: 9-15(in Chinese).

[22] Inter-Agency Space Debris Coordination Committee. IADC space bebris mitigation guidelines[EB/OL]. (2025-01-16)[2025-12-18]. https://www.iadc-home.org/documents_public/view/id/318#u.

[23] JERRETT Y. The raw Iridium 33 debris data with RCS at around 17126UTC[EB/OL]. (2017-10-08)[2023-11-03]. https://github.com/JerrettYang/Iridium-33-17126UTC.

Planning method for a multi-debris active removal mission considering space debris mass

GUO Jia[1]，ZHANG Haibo[2]，PANG Zhaojun[1,*]，DU Zhonghua[1]

(1. School of Mechanical Engineering，Nanjing University of Science and Technology，Nanjing 210094，China；
2. Beijing Institute of Control and Engineering，Beijing 100190，China)

Abstract：In solving the problem of multi-debris active removal, the existing decision framework based on the time-dependent traveling salesman problem (TD-TSP) only optimizes the debris removal sequence and transfer time. And the effect of optimization is limited. In light of the features of multi-debris active removal missions, this study introduces the decision framework's timing for debris release and adds debris quality as a criteria to be taken into account. This achieves a balance between the number of transfers and platform load, thus achieving optimization at a deeper level. A clustering algorithm suitable for drift orbit transfer methods is designed to select debris from a large-scale debris pool that is suitable as mission targets. This decouples the selection of debris from mission planning optimization, thereby reducing computational complexity. With reference to the orbital information of the Iridium-33 debris cloud, simulation experiments were conducted with fixed debris masses as well as with varying debris masses, which is closer to reality. The results indicate that the clustering algorithm can select targets with similar orbits and facilitate transfer. Better performance metrics have been demonstrated by the solutions derived from the decision framework presented in this research than by those derived from more conventional decision frameworks. By capturing debris in batches, the cost of the mission can be further reduced.

Keywords：space debris; mission planning; genetic algorithm; cluster analysis; multi-objective optimization

Received：2023-11-20；**Accepted**：2024-02-23；**Published Online**：2024-03-14 12：33

URL：link.cnki.net/urlid/11.2625.V.20240313.1602.003

Foundation item：Open Fund Project of Space Intelligent Control Technology Laboratory (HTKJ2021KL502010)

* **Corresponding author**. E-mail：pangzj@njust.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0712

噪声环境下告警冗余设计对装甲车乘员响应绩效的影响

周孙夏[1]，刘双[1]，完颜笑如[1,*]，冯传宴[1]，王一行[1]，解芳[2]
(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 中国北方车辆研究所，北京 100072)

摘 要：为探究不同噪声环境下多模态告警冗余设计对装甲车乘员响应绩效的影响，招募24名被试，开展了基于某新型高保真装甲车半实物仿真平台的模拟驾驶实验。采用2种噪声环境（高噪声、低噪声）×3种告警冗余设计（视觉通道冗余设计、听觉通道冗余设计、视听双通道冗余设计）条件下的双因素被试内设计，测量了被试的告警响应操作绩效、眼动指标及NASA-TLX量表得分。实验结果显示，在听觉通道冗余设计和视听双通道冗余设计条件下，乘员对于告警信息的响应错误率和反应时均显著低于视觉通道冗余设计条件；在不同告警冗余设计下，低噪声环境下乘员对于告警信息的响应错误率显著低于高噪声环境条件，但二者反应时之间并无明显差异；对于告警响应操作绩效、眼动指标和NASA-TLX量表得分，噪声条件和告警设计模式之间的交互效应均不显著。研究表明：对于装甲车告警信息设计，与采用视觉通道冗余设计模式相比，采用增加了听觉通道冗余提示的多模态告警模式（包括听觉通道冗余设计模式和视听双通道冗余设计模式），可以显著提高乘员对于告警信息的响应操作绩效；而在听觉通道冗余设计条件下继续叠加视觉通道冗余设计，即采用视听双通道冗余设计模式，却未带来进一步的响应操作绩效增益；低噪声作业环境对于改善乘员的响应操作绩效具有积极作用。

关 键 词：装甲车辆；多模态告警；噪声环境；任务绩效；乘员；人机工效
中图分类号：TJ810.2
文献标志码：A **文章编号**：1001-5965(2026)01-0286-08

研究人的因素对于保障特种车辆的高效与安全作业至关重要。相关研究表明，超过70%的特种车辆事故与人的因素有关[1]，而其中，不良告警设计又占有不可忽视的比例。有数据表明，告警设计在事故诱因占比中最多可达40%[2]，而当车辆驾驶员能够获得有效的告警提示时，事故的发生概率可显著降低30%～43%[3]。因此，如何在紧急情况下采用有效的设计方式警示驾驶员尽快作出正确决策对于保障安全驾驶至关重要。

为提高告警信息设计的可靠性，采用基于视觉、听觉、触觉的多模态告警设计是常用的设计手段之一[4]。研究表明，在高任务负荷作业条件下，多模态告警信息往往能够更有效地捕获驾驶员的注意，从而带来更好的告警响应操作绩效。例如，文献[5]研究表明采用多模态告警可以缩短装甲车乘员的告警响应反应时和错误率。文献[6]在手动驾驶任务条件下采用视觉图标与抽象声音组合的方式，发现驾驶员对视听双模态告警信息的反应比仅采用视觉单模态告警拥有更快的制动开始时间和更小的最大制动幅度。文献[7]在车辆模拟驾驶任

收稿日期：2023-11-01；录用日期：2023-12-20；网络出版时间：2023-12-27 16:31
网络出版地址：link.cnki.net/urlid/11.2625.V.20231227.1026.001
基金项目：国家自然科学基金委员会与中国民用航空局联合基金(U1733118)；国家自然科学基金(71301005)
* 通信作者. E-mail：wanyanxiaoru@buaa.edu.cn

引用格式：周孙夏，刘双，完颜笑如，等. 噪声环境下告警冗余设计对装甲车乘员响应绩效的影响[J]. 北京航空航天大学学报，2026，52(1)：286-293. ZHOU S X，LIU S，WANYAN X R，et al. Influences of multimodal redundant alarm design on reaction performance of armored vehicle crew under noise environment[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：286-293 (in Chinese).

务中研究了单模态、双模态和三模态告警对不同年龄驾驶员操作绩效的影响, 发现采用多模态告警有利于提升告警响应操作绩效并降低驾驶员的主观感知负荷。然而, 也有研究发现多模态告警并不总是有益的, 例如, 文献 [8] 的研究发现, 尽管采用多模态告警的设计方式有助于缩短操作反应时, 但也可能导致更差的用户主观体验。文献 [9] 则发现单一任务条件下的多模态告警设计并未呈现出明显的绩效改善, 因此, 提出低任务负荷条件下的多模态告警设计可能并不比单模态告警设计更具优势。而面向装甲车这一典型任务环境应采用何种告警设计模式以保证绩效最优仍需探索。

此外, 特种车辆乘员通常需要在恶劣的噪声作业环境下执行任务, 而噪声环境会对告警信息设计特别是听觉告警信息设计效果产生直接作用, 因此, 在进行告警信息设计时必须考虑噪声环境的影响。研究表明, 通常情况下, 装甲车乘员对于听觉信息的反应时间比视觉信息更快[10], 但在噪声条件下, 乘员需要耗费更多的努力从背景噪声中感知到听觉告警信息, 从而可能导致信息的获取难度增加[11]。例如, 文献 [12] 研究了噪声环境对 3 种不同声音告警信息响应绩效的影响, 发现背景噪声会明显影响驾驶员对告警信息种类的识别。文献 [13] 发现在高噪声环境下, 操作人员对告警信号的音调感知能力发生显著下降。文献 [14] 则发现背景噪声削弱了操作人员的神经反应, 从而影响到对听觉信号的检测阈值、空间感知和预测能力等。

一般的, 多种引发相同反应的信息被称为冗余信息[15]。现有研究多在民用驾驶领域探讨来自视觉、听觉、触觉等不同组合的多模态冗余告警信息对人员操作绩效的影响[6-8,10,11], 但对于在原始视觉告警信息设计状态下, 采用何种形式的冗余设计模式可以更有效地增强乘员的告警响应操作绩效的对比研究则较有限。此外, 也缺乏多模态冗余告警设计方式在不同噪声作业环境下的设计效果验证, 特别是面向特种车辆设备研制领域的相关设计理论与实验数据更为匮乏。针对上述研究局限性, 本文基于某新型高保真装甲车仿真平台, 面向特种车辆典型驾驶任务开展工效学实验, 研究不同噪声环境下多模态冗余告警设计对装甲车乘员的告警响应操作绩效的影响。研究结果可为优化装甲车辆显控界面的告警信息设计提供理论依据与实验数据支撑。基于上述文献分析并结合本研究实验设计[7,10,12], 提出以下假设:

假设 1　采用听觉通道冗余设计和视听双通道冗余设计可以有效提高乘员对于告警信息的响应操作绩效, 表现为相较于视觉通道冗余设计模式, 听觉通道冗余设计模式和视听双通道冗余设计模式的任务绩效更优。

假设 2　噪声环境将对乘员告警信息的响应操作绩效产生消极影响, 表现为低噪声环境下的任务绩效优于高噪声环境下的任务绩效。

假设 3　噪声条件和告警设计模式之间存在交互效应, 表现为增加了听觉通道冗余提示的多模态告警模式的绩效优势将在高噪声环境下被削弱。

1　实验方法

1.1　实验被试

本文共招募 24 名具有特种车辆模拟驾驶经验的装甲车乘员, 年龄范围为 22～45 岁(平均值为 33.88 岁, 标准差为 7.09 岁), 其中, 男性 22 名, 女性 2 名。所有被试身体健康状况良好, 右利手, 无色盲色弱, 视力或矫正视力正常, 听力正常。实验前一天无剧烈运动, 休息良好。

1.2　实验设计

实验采用完全被试内设计, 自变量为噪声环境和告警冗余设计, 采用 2 噪声环境×3 告警冗余设计。其中, 低噪声水平依据一般室内噪声环境设置为 55 dB, 高噪声水平参照特种车辆作业任务环境噪声参数范围, 并根据预实验对高噪声环境进行了响度的控制, 设置为 95 dB, 模拟噪声经由头戴式耳机在实验过程中向被试播放。

因素 2 为告警冗余设计, 原始视觉告警信号为基线, 并呈现为红色三角形和感叹号的组合图形, 在此基础上叠加呈现不同信息加工通道条件下的冗余设计, 具体包括视觉通道冗余设计、听觉通道冗余设计、视听双通道冗余设计 3 种模式。其中, 视觉通道冗余设计为在原始视觉告警信号基础上叠加呈现频率为 1 Hz 的闪烁信号; 听觉通道冗余设计在原始视觉告警信号基础上叠加呈现符合 GJB1835[16] 规定的语音告警提示音, 如“发动机温度过高”; 视听双通道冗余设计为在原始视觉告警信号基础上同时叠加呈现频率为 1 Hz 的闪烁信号与语音告警提示音。语音告警提示音经由头戴式耳机在实验过程中向被试播放。

本文方案通过了北京航空航天大学生物与医学伦理委员会的审查和批准(批准编号: BM20230003), 实验参数的有效性经设备设计研制单位专家确认。此外, 本实验采用拉丁方设计以消除疲劳效应和练习效应。

1.3　实验仪器设备

实验基于某新型高保真装甲车半实物仿真平台开展, 该仿真平台由装甲车半实物硬件平台和软

件系统构成。其中，硬件平台主要包括显控界面、方向盘和刹车等，显控界面进一步划分为驾驶视景区、地图区、告警显示区、状态显示区；软件系统主要包括装甲车驾驶模拟系统、噪声模拟系统及人机交互数据采集记录系统。本文所采用的装甲车半实物仿真平台软硬件系统构成如图1所示。

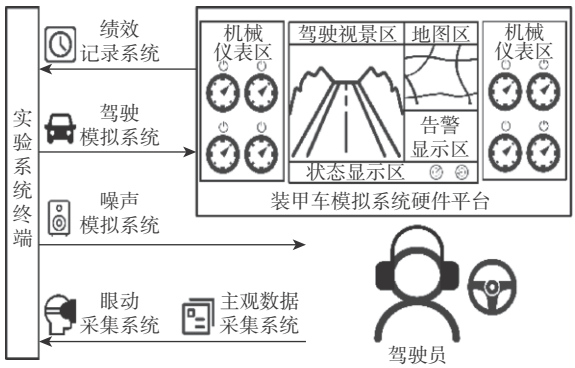


图1 装甲车半实物仿真平台与人机交互数据采集系统

Fig. 1 Armored vehicle semi-physical simulation platform and human-machine interactive data acquisition system

人机交互数据采集系统包括绩效记录系统与眼动采集系统。其中，使用绩效记录系统记录被试在任务执行过程中响应告警信息的错误率和反应时；使用瑞典Tobii技术公司生产的TobiiGlasses 2可穿戴式眼动追踪系统追踪装甲车乘员的眼动信号，眼动追踪系统的校准方式为一点校准法，水平追踪范围为82°，垂直追踪范围为48°，采样频率为50 Hz。

1.4 测量方法与指标

本文采用主观量表评价、作业绩效测量及眼动追踪测量来综合反应装甲车乘员对告警信息的主观感知负荷、响应操作绩效及生理状态反应[5,8,12]。其中，主观评价量表选用美国国家航空航天局任务负荷指数（NASA-Task Load Index, NASA-TLX）量表，其可提供对乘员在执行任务时所承受的认知和心理压力的量化和评估[17]。作业绩效测量指标选取告警响应任务的错误率与反应时，其中，根据实验任务需求分析及预实验样本数据，取4 s为界限对反应时进行处理，反应时大于4 s视为错误响应，较高的错误率可能反映了告警信息设计的不合理[18]。反应时指标被认为与乘员的认知水平有关[19]。眼动测量指标选用平均注视时长和平均扫视幅度[20]，其中，注视时长被认为与任务执行过程中装甲车乘员的信息加工深度相关[21]，而扫视幅度则可反映出装甲车乘员的脑力负荷状态[22]，高脑力负荷状态可能伴随着更大的平均扫视幅度。

1.5 实验任务与流程

本文要求被试在不同实验条件下完成由车辆启动、道路探查、规定路线行驶、车辆熄火4个子任务阶段所构成的完整车辆驾驶任务流程。在车辆启动阶段，被试需要按照要求进行安全检查，选择合适的驾驶模式并完成车辆启动任务；在道路探查阶段，被试需要快速寻找到指定的驾驶路线；在规定路线行驶阶段，被试需要依据驾驶视景沿预定义道路完成车辆驾驶任务；在车辆熄火阶段，被试需要将车辆平稳停泊于目标区域，并按要求完成安全检查及车辆熄火等操作。在整个任务执行过程中，被试除了需要完成车辆模拟驾驶，同时需要尽快准确地响应车辆告警信息以保持安全行驶。告警信息平均呈现频率为4次/min，间隔随机，要求被试通过指定按键进行响应。

被试在参加正式实验前首先接受充分的上岗培训，要求被试在培训过程中熟悉装甲车仿真平台显控界面设计方案与实验任务流程，被试达到培训考核标准后即可开始正式实验。正式实验前要求被试签署知情同意书，佩戴眼动仪和头戴式耳机。每名被试均需要在6种实验条件下各完成一次模拟驾驶任务，如表1所示，高、低分别代表高噪声条件和低噪声条件；视觉、听觉、视听分别代表视觉通道冗余设计模式、听觉通道冗余设计模式和视听双通道冗余设计模式。单次实验任务持续时长约10 min。被试于单次实验结束后填写NASA-TLX量表。

表1 实验变量及拉丁方设计

Table 1 Experimental variables and Latin square design

被试编号	噪声条件	告警模式		
1	高	视觉	听觉	视听
	低	听觉	视听	视觉
2	低	视听	视觉	听觉
	高	视觉	听觉	视听
3	高	听觉	视听	视觉
	低	视听	视觉	听觉
4	低	视觉	听觉	视听
	高	听觉	视听	视觉
5	高	视听	视觉	听觉
	低	视觉	听觉	视听
6	低	听觉	视听	视觉
	高	视听	视觉	听觉

2 实验结果

本文使用IBM SPSS Version 25进行统计分析。在符合正态分布的条件下，进一步针对不同实验条件下的各项测量指标进行“2噪声环境×3告警冗余设计”被试内重复测量的方差分析（analysis of

variance, ANOVA), 统计检验采用 0.05 置信度, 若不满足球形检验, 则采用 Green-Geisser 方法进行校正。

2.1　作业绩效测量结果

F、p 和 η^2 是方差分析的核心指标。F 是检验统计量, 由组间变异(自变量导致的差异)与组内变异(随机误差导致的差异)的比值计算得到, 在本文可理解为: 由自变量“噪声条件”导致的差异与随机误差的比值, 该比值可在一定程度上表征噪声条件的影响大小; p 可理解为在原假设(H_0: 所有样本总体均值相等, 即自变量无影响)成立的条件下, F 统计量取值大于等于当前计算值的概率, 在该案例中表征噪声条件没有影响的概率; η^2 为组间变异占总变异的比例, 是效应量指标, 在该案例中表征噪声条件引起的差异占总差异的比例, 一般来说 η^2 大于 0.14 可认为是大效应。告警响应错误率的描述统计结果如图 2 所示, 重复测量的方差分析表明, 噪声条件和告警冗余设计对告警响应错误率的交互效应不显著($p > 0.05$), 即不同告警冗余设计下的响应错误率并未随着噪声条件的改变而变化。噪声条件的主效应显著($F(1, 23) = 4.562$, $p = 0.044$, $\eta^2 = 0.166$), 表现出低噪声条件下的告警响应错误率($M = 10.40\%$)低于高噪声条件下的告警响应错误率($M = 16.35\%$)。告警冗余设计的主效应显著($F(2, 46) = 7.930$, $p = 0.001$, $\eta^2 = 0.256$)。进一步的事后比较发现, 视觉通道冗余设计模式下的告警响应错误率($M = 21.13\%$)分别显著高于听觉通道冗余设计模式($M = 10.74\%$, $p = 0.014$)与视听双通道冗余设计模式下的告警响应错误率($M = 8.26\%$, $p = 0.002$)。听觉通道冗余设计模式与视听双通道冗余设计模式之间的差异不显著($p > 0.05$)。

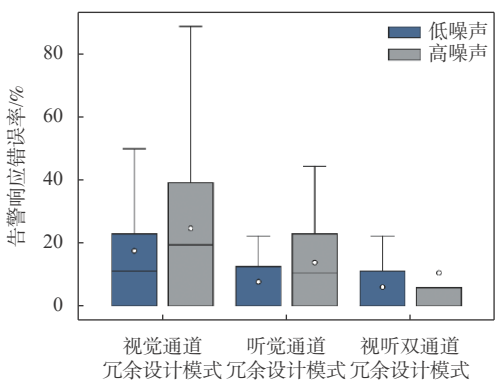


图 2　不同实验条件下告警响应错误率结果

Fig. 2　Results of reaction error rate under different experiment conditions

告警响应反应时的描述统计结果如图 3 所示, 重复测量的方差分析表明, 噪声条件和告警冗余设计对告警响应反应时的交互效应均不显著($p > 0.05$), 即不同告警冗余设计下的告警响应反应时并未随着噪声条件的改变而变化。噪声条件的主效应不显著($p > 0.05$), 即低噪声条件下的告警响应反应时($M' = 2.25$ s)与高噪声条件下的告警响应反应时($M' = 2.46$ s)并无明显差异。告警冗余设计的主效应显著($F(2, 42) = 77.314$, $p < 0.001$, $\eta^2 = 0.786$)。进一步的事后比较表明视觉通道冗余设计模式下的告警响应反应时($M' = 3.22$ s)分别显著长于听觉通道冗余设计模式($M' = 1.96$ s, $p < 0.001$)和视听双通道冗余设计模式($M' = 1.89$ s, $p < 0.001$), 听觉通道冗余设计模式和视听双通道冗余设计模式之间的反应时差异不显著($p > 0.05$)。

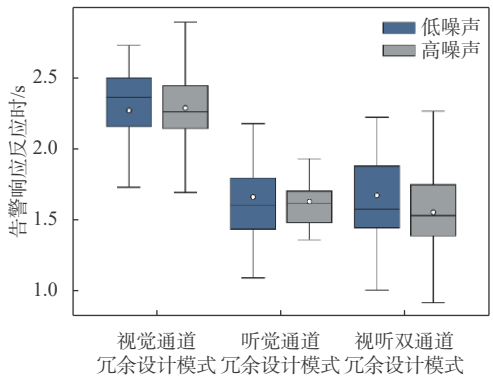


图 3　不同实验条件下的告警响应反应时结果

Fig. 3　Results of reaction time under different experiment conditions

2.2　眼动指标测量结果

平均注视时长的描述统计结果如图 4 所示, 重复测量的方差分析表明, 噪声条件和告警冗余设计对平均注视时长的交互效应不显著($p > 0.05$), 即不同告警冗余设计下的平均注视时长并未随着噪声条件的改变而变化。噪声条件对平均注视时长的主效应不显著($p > 0.05$), 即低噪声条件下的平均注视时长($\overline{M} = 0.53$ s)和高噪声条件下的平均注视时长($\overline{M} = 0.46$ s)之间无明显差异。告警冗余设计对平均注视时长的主效应显著($F(2, 46) = 3.605$, $p =$

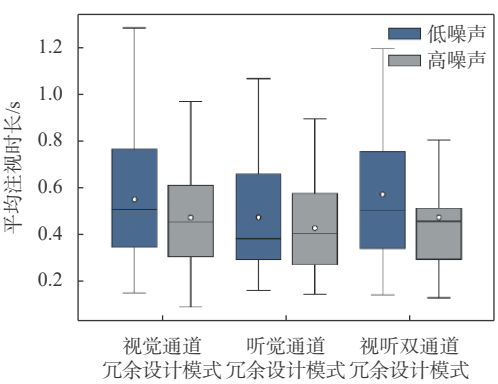


图 4　不同实验条件下平均注视时长结果

Fig. 4　Average duration of fixations under different experiment conditions

0.035, η^2= 0.136)。进一步的事后比较表明听觉通道冗余设计模式的平均注视时长($\overline{M}$ = 0.45 s)显著低于视觉通道冗余设计模式($\overline{M}$ = 0.51 s, p = 0.030), 同时临界显著低于视听双通道冗余设计模式($\overline{M}$ = 0.52 s, p = 0.051), 视觉通道冗余设计模式与视听双通道冗余设计模式之间的差异不显著(p > 0.05)。

平均扫视幅度的描述统计结果如图 5 所示, 重复测量的方差分析表明, 噪声条件和告警冗余设计对平均扫视幅度的交互效应不显著(p > 0.05), 即不同告警冗余设计下的平均扫视幅度并未随着噪声条件的改变而变化。噪声条件对平均扫视幅度的主效应不显著(p > 0.05), 即低噪声条件下的平均扫视幅度(M_s = 3.70°)与高噪声条件下的平均扫视幅度(M_s = 3.64°)并无明显差异。告警冗余设计对平均扫视幅度的主效应显著(F(2, 46)= 3.643, p = 0.034, η^2 = 0.137)。进一步的事后比较发现在视听双通道冗余设计模式条件下的平均扫视幅度(M_s = 3.45°)显著低于视觉通道冗余设计模式(M_s = 3.74°, p = 0.033)与听觉通道冗余设计模式(M_s = 3.80°, p = 0.035), 视觉通道冗余设计模式与听觉通道冗余设计模式之间的差异不显著(p > 0.05)。

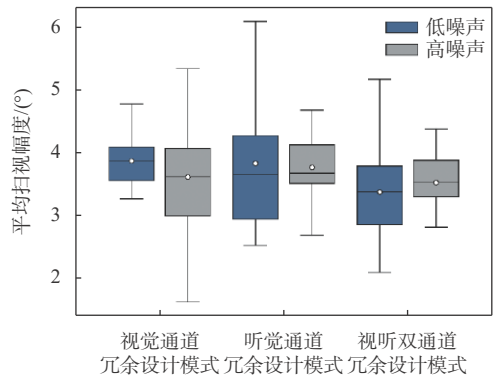


图 5 不同实验条件下平均扫视幅度结果

Fig. 5 Average amplitude of saccades under different experiment conditions

2.3 NASA-TLX 量表结果

不同实验条件下的 NASA-TLX 量表描述性统计结果如表 2 所示, 重复测量的方差分析结果显示, 对于 NASA-TLX 总分及各子维度得分而言, 噪声条件和告警冗余设计之间的交互效应及主效应均不显著(p > 0.05)。

表 2 不同实验条件下的 NASA-TLX 量表结果

Table 2 Results of NASA-TLX subscale ratings under different experiment conditions

实验条件	总得分	脑力需求	体力需求	时间需求	工作绩效	努力程度	挫败感
视觉通道冗余-低噪声	40.38(±18.59)	9.04(±4.40)	7.29(±3.84)	8.25(±4.26)	7.79(±4.56)	9.92(±5.09)	6.88(±3.98)
听觉通道冗余-低噪声	39.35(±18.05)	8.08(±3.99)	7.42(±3.40)	8.25(±4.32)	8.04(±4.29)	9.67(±4.72)	6.38(±4.75)
视听双通道冗余-低噪声	40.55(±17.16)	8.38(±3.66)	7.13(±3.80)	7.92(±3.74)	8.29(±4.30)	9.58(±4.47)	7.38(±4.73)
视觉通道冗余-高噪声	40.33(±19.09)	8.58(±4.62)	8.38(±4.83)	8.58(±4.61)	7.08(±3.87)	10.46(±5.42)	6.83(±4.59)
听觉通道冗余-高噪声	40.04(±16.94)	8.33(±4.23)	7.17(±3.81)	9.29(±4.80)	6.88(±4.88)	10.71(±5.20)	6.13(±3.66)
视听双通道冗余-高噪声	40.16(±17.36)	8.67(±4.70)	8.04(±4.26)	8.63(±4.21)	6.96(±4.51)	10.21(±4.55)	6.83(±4.79)

注: 表中数值形式为均值(均方差)。

3 实验结果讨论

本文基于某新型高保真装甲车半实物仿真平台, 面向车辆启动、道路检查、规定路线行驶、车辆熄火与告警响应的装甲车典型驾驶任务, 以 NASA-TLX 量表及其子维度、告警响应错误率与反应时、平均注视时长与平均扫视幅度为测量指标, 综合采用主观评价、作业绩效测量及眼动测量等多维度方法[23], 开展了噪声环境下不同告警冗余设计对装甲车乘员响应绩效影响的工效学实验研究。

实验结果表明, 装甲车乘员对于增加了听觉通道冗余提示的多模态告警信息的响应操作绩效更优。表现为与视觉通道冗余设计模式相比, 听觉通道冗余设计模式和视听双通道冗余设计模式下的告警响应错误率分别降低了 49.2% 和 60.9%, 反应时分别缩短了 27.7% 和 29.1%, 该结果支持了假设 1。这与已有研究结果一致, 例如, 文献 [24] 表明采用多模态告警设计对于感知阶段的信息加工有促进作用, 从而有利于提升操作人员的任务绩效; 而文献 [25] 也表明增加了听觉冗余信号的多模态告警设计有助于降低驾驶员对于告警响应的反应时和错误率。分析其原因, 一方面, 增加听觉通道冗余信息的多模态告警设计可传递出更加强烈的紧迫感, 从而促使乘员优先处理所感知到的告警信息所指向的事件, 有利于更为迅速、准确地作出反应以规避危险[26]。另一方面, 多模态告警冗余设计可以有效提升乘员对于信息的信任度[27], 即通过听觉通道冗余信号来交叉检验告警信息的真实性与有效性, 从而提升乘员的处置信心与响应操作绩效。

此外, 实验结果还表明, 听觉通道冗余设计模式的平均注视时长最短, 注视时长通常被认为与任务执行过程中乘员的信息加工深度相关, 这表明在

听觉通道冗余设计模式下, 信息加工难度对于乘员而言更低, 从而也有助于提高告警信息的响应操作绩效[28]。本文 NASA-TLX 量表总得分及其各子维度指标在不同实验条件下的差异均不显著, 提示无论采用何种告警冗余设计, 均未给乘员带来更高的感知负荷代价。综上, 通过增加听觉提示, 可以在不给乘员带来更高感知负荷代价的同时, 为乘员提供额外的辅助信息, 提高乘员对关键告警信息的感知与反应能力[12]。

尽管听觉通道冗余设计模式可显著提高乘员的告警响应操作绩效, 但在听觉通道冗余设计条件下继续叠加视觉通道冗余设计, 即采用视听双通道冗余设计模式, 却并没有带来进一步的告警响应操作绩效增益[29], 表现为听觉通道冗余设计模式与视听双通道冗余设计模式之间的告警响应操作绩效差异不显著, 分析其原因, 视觉闪烁告警因高突显性虽然可促进告警信息加工[30], 但同时增加了告警信息的加工复杂度, 两者可能存在抵消效应从而导致响应操作绩效未出现显著提升。文献 [31] 的研究结果表明, 工作负荷较高时, 过高的信息复杂度可能导致乘员忘记告警信号的含义, 从而影响告警响应操作绩效。进一步结合眼动数据分析, 视听双通道冗余设计模式下的平均扫视幅度最小, 且与其他冗余设计模式之间存在显著差异, 表现在该种冗余设计模式下, 乘员的视觉信息扫视范围更小、搜索效率更高的积极趋势[22]。综上, 采用视听双通道冗余设计模式在一定程度上有利于促进告警信息的视觉搜索过程, 但尚不足以带来作业绩效的明显改善。考虑到视听双通道冗余设计模式下的信息复杂度较高, 因此在工程设计中应谨慎采用。

尽管有研究表明, 环境噪声并不一定会恶化告警响应操作绩效[32], 但在本研究中, 面向装甲车典型作业任务, 高噪声作业环境对乘员告警响应操作绩效产生了明显的削弱影响, 表现为高噪声作业环境下乘员的告警响应错误率与低噪声作业环境相比增加了 5.9%, 该结果支持了假设 2。分析其原因, 装甲车辆作业的恶劣噪声环境提高了告警信息的检测阈值[33], 增加了告警信号的获取难度, 从而导致告警信息的响应错误率提高。文献 [34] 的研究也表明, 暴露在噪声环境下时, 操作人员的信息加工过程更容易受到干扰, 从而导致感知到的信息丢失。考虑到噪声对被试身心健康的影响, 本文基于预实验结果, 对高噪声作业环境下背景噪声的频率和响度进行了一定程度的控制, 但仍发现乘员告警响应操作绩效的明显恶化, 考虑到真实任务条件下的噪声环境更为恶劣, 因此, 在进行告警信息设计时需要考虑噪声环境因素的消极影响。此外, 本次实验中, 噪声条件和告警设计模式之间交互效应不显著, 提示不同水平的噪声作业环境并不会削弱多模态冗余告警设计模式的相对优势, 该结果与假设 3 不一致。这可能是因为本文基于模拟环境开展, 相应的噪声设置与真实环境存在一定差异。后续可进一步开展真实作业环境下的告警冗余设计研究, 结合告警响应任务相关的兴趣区进行分析, 从而为特种车辆告警设计提供针对性的指导意见。

4　结　论

1) 与采用视觉通道冗余设计模式相比, 采用增加了听觉通道冗余提示的多模态告警模式, 即听觉通道冗余设计模式和视听双通道冗余设计模式, 可以在不给乘员带来更高感知负荷代价的情况下显著提高乘员对于告警信息的响应操作绩效。

2) 在听觉通道冗余设计条件下继续叠加视觉通道冗余设计, 即采用视听双通道冗余设计模式, 可以在一定程度上优化告警信息的视觉搜索过程, 但未观察到进一步的告警响应操作绩效增益。因此, 在进行特种车辆告警信息设计时, 优先推荐听觉通道冗余设计模式, 但是否进一步增加视觉通道冗余设计应结合实际环境条件择优选择。

3) 高噪声环境会对乘员的告警信息响应操作绩效产生消极影响, 不同水平的噪声作业环境并不会削弱多模态冗余告警设计模式的相对绩效优势。

参考文献 (References)

[1] 傅斌贺, 刘维平, 聂俊峰, 等. 考虑认知行为差异的乘员信息作业绩效研究[J]. 兵工学报, 2019, 40(3): 659-665.
FU B H, LIU W P, NIE J F, et al. Research on crew's information operation performance with the difference of cognitive behavior[J]. Acta Armamentarii, 2019, 40(3): 659-665(in Chinese).

[2] WILLIAMS K W. A summary of unmanned aircraft accident/incident data: human factors implications, DOT/FAA/AM-04/24[R]. Oklahoma City: Federal Aviation Administration, 2004.

[3] LEDEZMA A, ZAMORA V, SIPELE Ó, et al. Implementing a gaze tracking algorithm for improving advanced driver assistance systems[J]. Electronics, 2021, 10(12): 1480.

[4] SARTER N B. Multimodal information presentation: design guidance and research challenges[J]. International Journal of Industrial Ergonomics, 2006, 36(5): 439-445.

[5] 孙晓东, 金晓萍, 解芳, 等. 多模态告警和认知负荷对装甲车辆乘员反应的影响[J]. 兵工学报, 2023, 44(4): 972-981.
SUN X D, JIN X P, XIE F, et al. Effects of multimodal warning and cognitive load on the response of armored vehicle occupants[J]. Acta Armamentarii, 2023, 44(4): 972-981(in Chinese).

[6] NAUJOKS F, KIESEL A, NEUKUM A. Cooperative warning systems: the impact of false and unnecessary alarms on drivers' compliance[J]. Accident Analysis & Prevention, 2016, 97: 162-175.

[7] LUNDQVIST L M, ERIKSSON L. Age, cognitive load, and multimodal effects on driver response to directional warning[J]. Applied Ergonomics, 2019, 76: 147-154.

[8] MA J, LI J T, HUANG H W. Evaluation of multimodal and multi-staged alerting strategies for forward collision warning systems[J]. Sensors, 2022, 22(3): 1189.

[9] SPENCE C. Crossmodal spatial attention[J]. Annals of the New York Academy of Sciences, 2010, 1191(1): 182-200.

[10] WANG Y W, WU B H, MA S, et al. Effect of mapping characteristic on audiovisual warning: evidence from a simulated driving study[J]. Applied Ergonomics, 2022, 99: 103638.

[11] EDWORTHY J. The design and implementation of non-verbal auditory warnings[J]. Applied Ergonomics, 1994, 25(4): 202-210.

[12] ŠABIĆ E, CHEN J, MACDONALD J A. Toward a better understanding of in-vehicle auditory warnings and background noise[J]. Human Factors, 2021, 63(2): 312-335.

[13] KARUNARATHNE B, WANG T Y, SO R H Y, et al. Adversarial relationship between combined medial olivocochlear (MOC) and middle-ear-muscle (MEM) reflexes and alarm-in-noise detection thresholds under negative signal-to-noise ratios (SNRs)[J]. Hearing Research, 2018, 367: 124-128.

[14] ANDERSON S, PARBERY-CLARK A, WHITE-SCHWOCH T, et al. Auditory brainstem response to complex sounds predicts self-reported speech-in-noise performance[J]. Journal of Speech, Language, and Hearing Research, 2013, 56(1): 31-43.

[15] D'AGOSTINO A R, BROWN J, FILLMORE M T. Redundant visual signals reduce the intensity of alcohol impairment[J]. Drug and Alcohol Dependence, 2020, 209: 107945.

[16] 黄琎, 刘雨发, 张松岩, 等. 装甲车辆人-机-环境系统总体设计要求: GJB 1835—1993[S]. 北京: 总参谋部装甲兵部, 1993.
HUANG L, LIU Y F, ZHANG S Y, et al. Overall design requirements for armored vehicle man-machine-environment system: GJB 1835—1993[S]. Beijing: Armored Division of the General Staff Department, 1993(in Chinese).

[17] 冯传宴, 完颜笑如, 陈浩, 等. 基于多资源负荷理论的情境意识模型与应用[J]. 北京航空航天大学学报, 2018, 44(7): 1438-1446.
FENG C Y, WANYAN X R, CHEN H, et al. Situation awareness model based on multi-resource load theory and its application[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(7): 1438-1446(in Chinese).

[18] 刘维平, 聂俊峰, 解芳, 等. 面向装甲车辆乘员信息处理作业的脑力负荷预测模型[J]. 装甲兵工程学院学报, 2017(1): 40-45.
LIU W P, NIE J F, XIE F, et al. Predictive mental workload model for armored vehicle crew's information processing task[J]. Journal of Academy of Armored Force Engineering, 2017(1): 40-45(in Chinese).

[19] 刘双, 完颜笑如, 庄达民, 等. 基于注意资源分配的情境意识模型[J]. 北京航空航天大学学报, 2014, 40(8): 1066-1072.
LIU S, WANYAN X R, ZHUANG D M, et al. Situational awareness model based on attention allocation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2014, 40(8): 1066-1072(in Chinese).

[20] AHLSTROM U, FRIEDMAN-BERG F J. Using eye movement activity as a correlate of cognitive workload[J]. International Journal of Industrial Ergonomics, 2006, 36(7): 623-636.

[21] 冯传宴, 完颜笑如, 刘双, 等. 负荷条件下注意力分配策略对情境意识的影响[J]. 航空学报, 2020, 41(3): 123307.
FENG C Y, WANYAN X R, LIU S, et al. Influence of different attention allocation strategies under workloads on situation awareness[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(3): 123307(in Chinese).

[22] CARDONA G, QUEVEDO N. Blinking and driving: the influence of saccades and cognitive workload[J]. Current Eye Research, 2014, 39(3): 239-244.

[23] 郭司南, 完颜笑如, 刘双, 等. 智能化设计与信息加工通道复杂度对装甲车乘员脑力负荷的影响[J]. 兵工学报, 2021, 42(2): 234-241.
GUO S N, WANYAN X R, LIU S, et al. Influences of intelligent design and information processing modality complexity on occupant mental workload[J]. Acta Armamentarii, 2021, 42(2): 234-241 (in Chinese).

[24] BIONDI F, STRAYER D L, ROSSI R, et al. Advanced driver assistance systems: using multimodal redundant warnings to enhance road safety[J]. Applied Ergonomics, 2017, 58: 238-244.

[25] GEITNER C, BIONDI F, SKRYPCHUK L, et al. The comparison of auditory, tactile, and multimodal warnings for the effective communication of unexpected events during an automated driving scenario[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2019, 65: 23-33.

[26] BLISS J P, ACTON S A. Alarm mistrust in automobiles: how collision alarm reliability affects driving[J]. Applied Ergonomics, 2003, 34(6): 499-509.

[27] CHANCEY E T, BLISS J P, PROAPS A B, et al. The role of trust as a mediator between system characteristics and response behaviors[J]. Human Factors, 2015, 57(6): 947-958.

[28] BELZ S M, ROBINSON G S, CASALI J G. A new class of auditory warning signals for complex systems: auditory icons[J]. Human Factors, 1999, 41(4): 608-618.

[29] OHTSUKA R, JIANWANG, CHIHARA T, et al. Estimation of mental workload during motorcycle operation[J]. Procedia Manufacturing, 2015, 3: 5313-5318.

[30] GIRARD S, PELLAND M, LEPORE F, et al. Impact of the spatial congruence of redundant targets on within-modal and cross-modal integration[J]. Experimental Brain Research, 2013, 224(2): 275-285.

[31] HAAS E C, VAN ERP J B F. Multimodal warnings to enhance risk communication and safety[J]. Safety Science, 2014, 61: 29-35.

[32] MURATA A, KURODA T, KARWOWSKI W. Effects of auditory and tactile warning on response to visual hazards under a noisy environment[J]. Applied Ergonomics, 2017, 60: 58-67.

[33] WIXTED J T. The forgotten history of signal detection theory[J]. Journal of Experimental Psychology Learning, Memory, and Cognition, 2020, 46(2): 201-233.

[34] MURPHY D R, CRAIK F I M, LI K Z H, et al. Comparing the effects of aging and background noise on short-term memory performance[J]. Psychology and Aging, 2000, 15(2): 323-334.

Influences of multimodal redundant alarm design on reaction performance of armored vehicle crew under noise environment

ZHOU Sunxia[1], LIU Shuang[1], WANYAN Xiaoru[1,*], FENG Chuanyan[1], WANG Yihang[1], XIE Fang[2]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. China North Vehicle Research Institute, Beijing 100072, China)

Abstract: In order to investigate the effect of different multimodal redundant alarm design on the reaction performance of armored vehicle occupants under a noisy environment, 24 participants were recruited and subjected to simulation driving experiments utilizing a new type of high-fidelity armored vehicle semi physical simulation platform. Three alarm redundancies were tested: visual channel, auditory channel, and audio-visual dual channel designs, across high and low noise conditions. The findings demonstrate that, in comparison to the visual channel redundancies, drivers' response error rate and reaction time to warnings are substantially reduced in the auditory channel redundancies and audio-visual dual channel redundancies. The response error is generally lower in low noise environments than in high noise environments, although response time remains consistent across noise conditions. The interaction between noise conditions and alarm design does not significantly affect alarm response performance, eye movement indicators, or subjective workload scores. Comparing the multimodal alarm with redundant auditory prompts to the visual-only redundancy, the study finds that the former significantly enhances drivers' reaction to warning information. However, the combination of auditory and visual redundancy does not yield further performance improvements. A low noise environment positively influences response performance enhancement.

Keywords: armored vehicle; multimodal warning; noise environment; task performance; occupant; ergonomics

Received: 2023-11-01; **Accepted**: 2023-12-20; **Published Online**: 2023-12-27 16:31

URL: link.cnki.net/urlid/11.2625.V.20231227.1026.001

Foundation items: National Natural Science Foundation of China and the Civil Aviation Administration of China (U1733118); National Natural Science Foundation of China (71301005)

* **Corresponding author**. E-mail: wanyanxiaoru@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0746

领域知识内嵌的电磁信号调制方式智能识别

赵宏佳[1]，张多纳[1,*]，鲁远耀[1]，丁文锐[2]
(1. 北方工业大学 信息学院，北京 100144；2. 北京航空航天大学 无人系统研究院，北京 100191)

摘 要：电磁环境日益复杂致使无线通信面临重大挑战，电磁信号调制识别作为认知无线电技术的关键环节意义深远。针对传统识别方法表征能力不足，深度学习方法适用性低、可解释性差的问题，结合2种方法的优点，提出领域知识内嵌的调制方式智能识别方法。所提方法将电磁信号高阶信息及频谱机制融入深度神经网络，提升分类性能及网络可解释程度。基于RML2018数据集，所提方法较ResNet调制识别正确率提升6.31%。

关 键 词：电磁特征；深度神经网络；调制方式识别；无线通信；知识内嵌
中图分类号：TN911
文献标志码：A **文章编号**：1001-5965(2026)01-0294-12

电磁信号调制识别是无线通信领域的重要研究方向之一，其应用范围涵盖无线电通信、雷达行为识别、电子对抗等多个方向。在军事领域，通过识别电磁信号从而获取情报信息，为作战决策提供依据。在民用领域，该技术是信道估计、频谱检测、多用户检测等应用的重要环节，对提高通信系统的性能和可靠性意义深远。因此，电磁信号调制识别技术具有重要的理论研究意义和实际应用价值[1]。

电磁信号调制识别技术已得到广泛的研究和应用。模型驱动的调制识别方法包括：①基于似然比检测的调制识别方法；②基于特征提取的调制识别方法。基于似然比检测的调制识别方法通过概率论分析信号特性，基于电磁信号的似然函数，将似然比和设定门限的比较结果作为判决依据。似然比检测方法包括：①平均似然比检测(average likelihood ratio test, ALRT)[2]；②广义似然比检测(generalized likelihood ratio test, GLRT)[3]；③混合似然比检测(hybrid likelihood ratio test, HLRT)[4]。该方法在理想信道下信号分类性能理论最优，但计算复杂度较高，需要完备的信号先验知识且易受到环境干扰因素的影响。基于特征提取的调制识别方法依赖电磁领域专家知识，通过人工设定信号特征提取模型。特征提取将原始信号数据映射到一个特定的特征空间，捕捉电磁信号调制方式之间的差异，并将数据降维，以提高算法的效率。常用的信号特征包括：①瞬时特征[5-7]；②统计特征[8-9]；③变换域特征[10-11]。常见的统计特征主要涉及高阶矩和高阶累积量，针对衰落信道，通过选择适当的阶数、共轭变量和延迟量，统计特征能够抑制相位偏移和衰落影响。与似然比检测相比，人工信号特征对信号的先验信息需求及算法复杂度较低。但该特征局限于针对信号属性单一，特征表征能力有限，复杂电磁环境下鲁棒性较低，多类别电磁信号识别适用性有待提升。

随着深度学习技术的发展，数据驱动的电磁信号调制识别方法逐渐得到广泛关注。数据驱动的调制识别方法包括：①基于人工信号特征和早期机器学习分类器的调制识别方法；②基于深度神经网络的调制识别方法。基于人工特征和分类器的调制识别方法由于早期机器学习分类器不具备特征提取能力，通常分为提取原始信号数据中的人工信号特征和将信号特征输入分类器2个步骤，最终实

收稿日期：2023-11-20；**录用日期**：2024-03-11；**网络出版时间**：2024-03-13 14:41
网络出版地址：link.cnki.net/urlid/11.2625.V.20240313.1402.001
基金项目：国家自然科学基金(62201010)；北京市教育委员会科研计划项目(KM202310009003)
*通信作者. E-mail：zhangduona@buaa.edu.cn

引用格式：赵宏佳，张多纳，鲁远耀，等. 领域知识内嵌的电磁信号调制方式智能识别[J]. 北京航空航天大学学报，2026，52(1)：294-305. ZHAO H J，ZHANG D N，LU Y Y，et al. Intelligent recognition of electromagnetic signal modulation with embedded domain knowledge[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：294-305 (in Chinese).

现电磁信号调制识别。早期机器学习分类器主要包括：①支持向量机[12-14](support vector machine, SVM)；②决策树；③浅层神经网络[15]；④K 最近邻法[16-17](K-nearest neighbor, KNN)。该方法较模型驱动的调制识别相比分类性能得到提升，但其分类效果依赖特征提取质量，在复杂的电磁环境下鲁棒性不足。基于深度神经网络(deep neural network, DNN)的调制识别方法基于原始电磁信号数据直接预测信号标签。基于 DNN 的调制识别方法主要包括：①全连接神经网络 (fully connected neural network, FCNN)；②卷积神经网络[18-19] (convolutional neural network, CNN)；③递归神经网络[20-21] (recursive neural network, RNN)。针对电磁、语音等数据的时序特征，RNN 在网络层内神经元之间建立参数连接，基于不同时刻间的网络参数传递过程构建样本数据时序关系。但该网络训练过程中反向传播梯度为各个时刻的累乘，导致梯度爆炸和梯度消失的现象。与早期机器学习分类器相比，DNN 具有较强的自适应性和鲁棒性，能够适应复杂无线信道和多类调制方式。因此，基于深度神经网络的电磁信号调制识别方法已成为当今主流研究方向。

综上所述，传统人工信号特征基于电磁领域信号先验知识，通过人工设定数据映射模型的方式，采用经典数字信号处理方法实现信号调制分类。该方法依赖专家知识构建特征提取模型，其局限性体现在：①在传统人工信号特征提取过程中，该特征完全依赖于专家知识。在实际应用场景中难以获取较为完备的先验信息且电磁信号调制方式繁多，导致传统人工信号特征属性较为单一、表征能力较差；②传统人工信号特征不具备自学习优化能力，如固定的小波基函数分布及特征数据映射方式，时变无线信道及多种噪声导致该特征鲁棒性较差，仅在理想情况信号调制分类性能较好，当干扰因素较多时，其分类性能急剧下降。随着人工智能技术的快速发展，DNN 在信号调制分类领域已全面优于传统信号特征。该方法通过定义网络结构及损失函数，采用数据驱动的方式进行特征挖掘，由于综合考虑调制类别间差异并将无线信道随机干扰视作网络参数拟合过程的影响因素，因此，特征表征能力显著提升且鲁棒性较强，适用于复杂电磁环境的信号调制分类任务。然而基于可学习网络层堆叠的数据驱动方式属于黑盒测试方式，导致 DNN 缺乏网络可解释性，同时 DNN 在网络架构设计过程中未考虑电磁信号独有属性，针对电磁领域适配度较低。

本文针对 DNN 未考虑电磁信号特有属性，应用于电磁领域可解释程度不足、适配度较低的问题，将电磁领域知识融入网络模型，结合高阶信息及频域机制对模型结构进行改进。提出一种基于领域知识内嵌的电磁信号调制方式智能识别方法，经实验证明，基于 RML2018 数据集，本文方法的平均识别率是 60.22%，较 ResNet 调制识别正确率提升 6.31%。

1　研究方法

本文方法融合高阶信息和频域机制对模型结构进行改进，提出领域知识内嵌网络模型，结构如图 1 所示。其中，高阶卷积模块能够显著提升传统卷积算法的特征提取能力，并抑制背景噪声；多谱通道注意力机制从频域角度出发，针对全局池化运算致使信息缺失问题，从电磁频谱角度实现高效特征筛选。

1.1　残差结构和通道注意力机制

残差结构：针对网络层数过多导致的梯度消失问题，在经典卷积神经网络模型的基础上，残差结构通过跨层参数连接方式，打破传统神经网络训练过程反向梯度仅在相邻网络层间传播的惯例，使梯度实现跨层传播。残差结构输出结果为：$\boldsymbol{F}(x)=\boldsymbol{H}(x)+\boldsymbol{x}$，其中，$x$ 为当前网络层输出特征，$\boldsymbol{H}(x)$为跨层网络特征向量。此结构能够增加网络深度并提升模型正确率。

通道注意力机制：卷积神经网络蕴含高维特征通道，然而多通道特征学习效果差异较大，且模型存在参数冗余问题。因此，压缩和激励(Squeeze-and-Excitation, SE)模块引入轻量级注意力机制，该机制计算通道注意力向量从而实现高维特征数据筛选，增强网络表征能力并去除参数冗余。

为解决通道依赖问题，SE 模块通过全局平均池化(global average pooling, GAP)实现特征信息压缩，并生成特征统计量。$\boldsymbol{Z}_c$为卷积层输出，其特征图维度为 $h\times W\times C$，经过 GAP 运算，特征图映射为 $1\times1\times C$ 的向量，全局平均池化如式 (1) 所示：

$$\boldsymbol{Z}_c=G(\boldsymbol{U}_c)=\frac{1}{W\times h}\sum_{i=1}^{W}\sum_{j=1}^{h}\boldsymbol{U}_{c(i,j)} \tag{1}$$

式中：$\boldsymbol{U}_c$ 是特征张量 $\boldsymbol{U}$ 的第 c 个通道的特征图。卷积特征实现信息压缩后，通道注意力向量 $\boldsymbol{s}$ 由全连接层及激活函数计算生成，如式 (2) 所示：

$$\boldsymbol{s}=\mathrm{Sigmoid}(\boldsymbol{W}_2\times\mathrm{Relu}(\boldsymbol{W}_1\boldsymbol{z})) \tag{2}$$

式中：$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 为权重矩阵；Relu 和 Sigmoid 为非线性激活函数。高维特征注意力筛选过程如式 (3) 所示：

$$\tilde{\boldsymbol{x}}=F_{\mathrm{scale}}(\boldsymbol{U}_c,\boldsymbol{s})=\boldsymbol{s}\cdot\boldsymbol{U}_c \tag{3}$$

通道注意力机制基于高维特征表征效果的差异性，通过赋予特征通道相应的权重系数，实现有效特征增强并降低模型参数冗余。

1.2　高阶卷积模块

传统卷积算法采用局部感受野无法有效捕捉全局信息，且模型难以获取电磁信号高阶特性。本

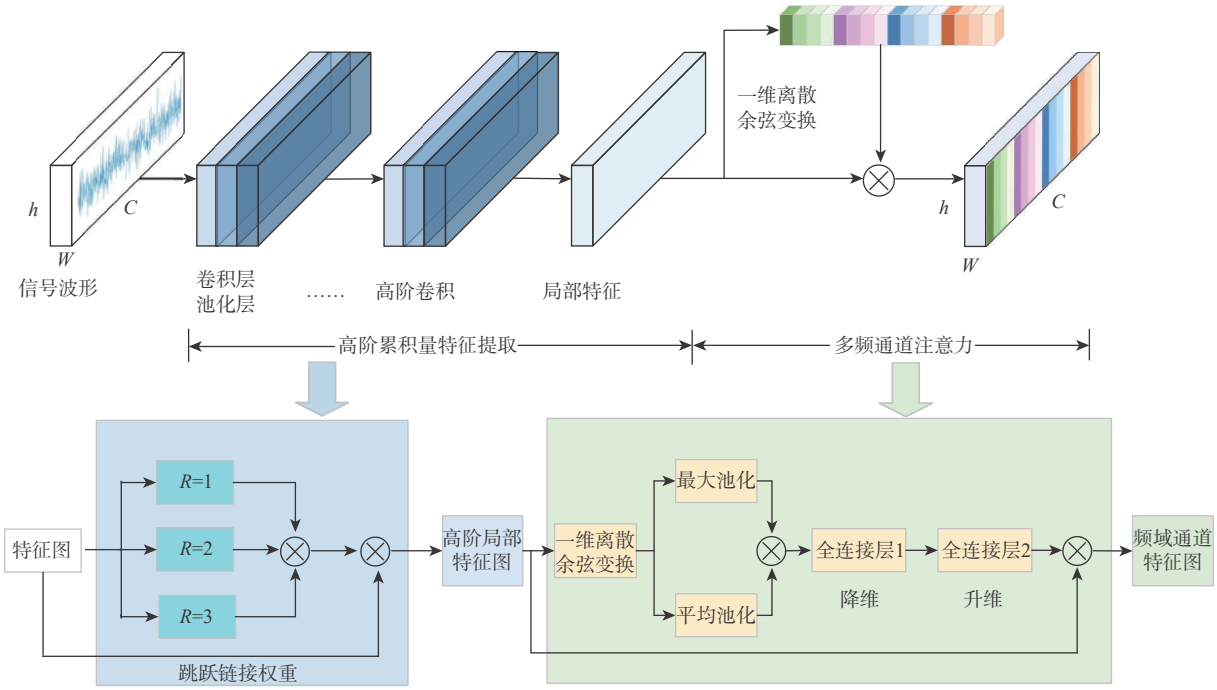


图 1 领域知识内嵌的调制方式智能识别架构

Fig. 1 Intelligent identification architecture of modulation modes embedded in domain knowledge

文提出针对时序信号的高阶卷积模块，将高阶信息融入传统卷积算法，如图 2 所示。相较于传统卷积算法，该模块能够更好地处理非线性关系和复杂特征，显著提升特征提取能力，并抑制背景噪声。

针对电磁信号的周期特性，高阶累积量已被广泛用于信号检测及分类，该特征计算复杂度较低，且对高斯白噪声具有抑制作用。为提取电磁信号高阶累积量，其高阶矩 $\boldsymbol{M}$ 的 p 行 q 列元素 M_{pq} 定义为

$$M_{pq} = E\left[y^{p-q}\cdot(y^*)^q\right] \tag{4}$$

式中：p、q 为高阶矩阶数索引；y 为电磁信号数据；E 和 $*$ 分别为期望值和共轭运算。基于高阶矩定义即可组合实现高阶累积量计算。为将高阶信息融入 DNN，针对网络中电磁信号特征 $\boldsymbol{x}$，基于高阶矩定义其高阶统计量线性多项式：

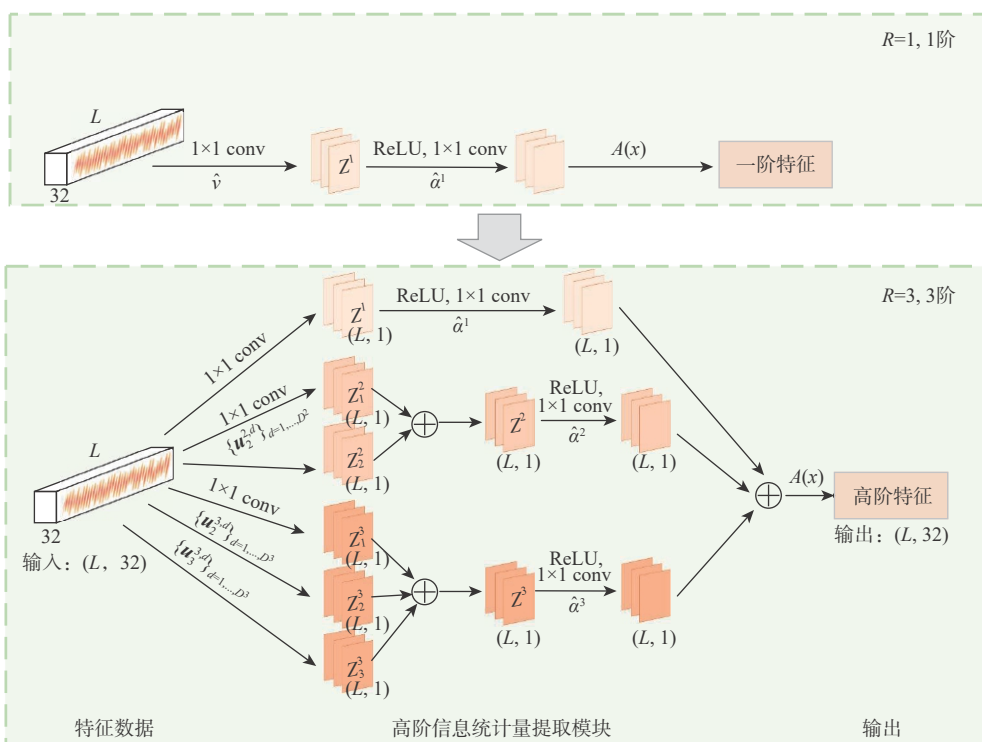


图 2 高阶信息统计量提取模块

Fig. 2 High-order information statistics extraction module

$$\boldsymbol{a}(\boldsymbol{x})=\sum_{r=1}^{R}\left\langle \boldsymbol{W}^{r},\boldsymbol{x}^{\otimes_r}\right\rangle \tag{5}$$

由于电磁信号通常为实序列, 本文针对电磁特征省略共轭运算, 同时采用高阶数据间代数求和简化期望值。其中, $\langle\cdot,\cdot\rangle$为向量的内积; R 为阶数; $x^{\otimes_r}$为 x 的 r 阶外积, $\boldsymbol{W}^r$为 r 阶变量组合权重。为解决参数 r 较大, $\boldsymbol{W}^r$引入过多参数导致过拟合问题。当 r>1 时, $\boldsymbol{W}^r$通过向量分解由 D^r 个 rank-1 向量(1 阶向量)近似, 即 $\boldsymbol{W}^r=\sum_{d=1}^{D^r}\alpha^{r,d}u_1^{r,d}\otimes u_2^{r,d}\otimes\cdots\otimes u_r^{r,d}$ (r >1), 其中, $[u_1^{r,d},u_2^{r,d},\cdots,u_r^{r,d}]$为经简化后的一阶向量, ⊗为外积运算[22], $\alpha^{r,d}$为第 d 个 1 阶向量权重系数。基于上述定义, 式 (5) 可表示为

$$\begin{aligned}\boldsymbol{a}(x)&=\left\langle \boldsymbol{w}^1,x\right\rangle+\sum_{r=2}^{R}\left\langle \sum_{d=1}^{D^r}\alpha^{r,d}\boldsymbol{u}_1^{r,d}\otimes\boldsymbol{u}_2^{r,d}\otimes\cdots\otimes\boldsymbol{u}_{\mathrm{r}}^{r,d},x^{\otimes_r}\right\rangle=\\&\left\langle \boldsymbol{w}^1,x\right\rangle+\sum_{r=2}^{R}\sum_{d=1}^{D^r}\alpha^{r,d}\prod_{s=1}^{r}\left\langle \boldsymbol{u}_s^{r,d},x\right\rangle=\left\langle \boldsymbol{w}^1,x\right\rangle+\sum_{r=2}^{R}\left\langle \boldsymbol{\alpha}^r,\boldsymbol{z}^r\right\rangle\end{aligned} \tag{6}$$

式中: $\boldsymbol{\alpha}^r=\left[\alpha^{r,1},\alpha^{r,2},\cdots,\alpha^{r,D^r}\right]^{\mathrm{T}}$为权重系数, $\boldsymbol{z}^r=\left[\boldsymbol{z}^{r,1},\boldsymbol{z}^{r,2},\cdots,\boldsymbol{z}^{r,D^r}\right]^{\mathrm{T}}$中 $\boldsymbol{z}^{r,d}=\prod_{s=1}^{r}\left\langle \boldsymbol{u}_s^{r,d},x\right\rangle$, 为简化深度网络结构, 将式 (6) 进一步简化为

$$\boldsymbol{\alpha}(x)=\boldsymbol{I}^{\mathrm{T}}\left(\boldsymbol{w}^1\odot x\right)+\sum_{r=2}^{R}\boldsymbol{I}^{\mathrm{T}}\left(\boldsymbol{I}^r\odot\boldsymbol{z}^r\right) \tag{7}$$

式中: 由于矩阵内积和 Hadamard 乘积的关系引入 $\boldsymbol{I}^{\mathrm{T}}$, 即由 1 组成的对角矩阵, ⊙为 Hadamard 乘积, 为在神经网络中实现高阶计算 $\boldsymbol{a}(x)\in\mathbf{R}^C$, 引入辅助矩阵 $\boldsymbol{P}^r$对式(7)进行泛化, 如下所示:

$$\boldsymbol{a}(\boldsymbol{x})=(\boldsymbol{P}^1)^{\mathrm{T}}\left(\boldsymbol{w}^1\odot x\right)+\sum_{r=2}^{R}(\boldsymbol{P}^r)^{\mathrm{T}}\left(\boldsymbol{\alpha}^r\odot\boldsymbol{z}^r\right) \tag{8}$$

式中: $\boldsymbol{P}^1\in\mathbf{R}^{C\times C},\boldsymbol{P}^r\in\mathbf{R}^{D^r\times C}$ (r>1)。由于 $\boldsymbol{P}$、$\boldsymbol{w}$和 $\boldsymbol{\alpha}$均为权重向量, 在神经网络参数训练过程中可进行合并, 从而简化网络结构, 因此, 基于矩阵变换方法将 $\left\{\boldsymbol{P}^1,\boldsymbol{w}^1\right\}$整合为单位矩阵 $\hat{\boldsymbol{w}}^1\in\mathbf{R}^{C\times C}$, 将 $\{\boldsymbol{P}^r,\boldsymbol{\alpha}^r\}$用 $\hat{\boldsymbol{\alpha}}^r\in\mathbf{R}^{D^r\times C}$表示, 如式 (9) 所示:

$$\boldsymbol{a}(\boldsymbol{x})=(\hat{\boldsymbol{w}}^1)^{\mathrm{T}}x+\sum_{r=2}^{R}(\hat{\boldsymbol{\alpha}}^r)^{\mathrm{T}}\cdot\boldsymbol{z}^r \tag{9}$$

为使可学习权重向量易于在网络中实现, $\hat{\boldsymbol{w}}^1$由矩阵 $\hat{\boldsymbol{v}}\in\mathbf{R}^{C\times D^1}$和 $\hat{\boldsymbol{\alpha}}^1\in\mathbf{R}^{D^1\times C}$乘积计算, 即 $\hat{\boldsymbol{w}}^1=\hat{\boldsymbol{v}}\hat{\boldsymbol{\alpha}}^1$, 如式 (10) 所示:

$$\boldsymbol{a}(\boldsymbol{x})=(\hat{\boldsymbol{\alpha}}^1)^{\mathrm{T}}\left(\hat{\boldsymbol{v}}^{\mathrm{T}}x\right)+\sum_{r=2}^{R}(\hat{\boldsymbol{\alpha}}^r)^{\mathrm{T}}\cdot\boldsymbol{z}^r=\sum_{r=1}^{R}(\hat{\boldsymbol{\alpha}}^r)^{\mathrm{T}}\cdot\boldsymbol{z}^r \tag{10}$$

式中: $\boldsymbol{z}^1=\hat{\boldsymbol{v}}^{\mathrm{T}}x$, $\hat{\boldsymbol{\alpha}}^r\in\mathbf{R}^{D^1\times C}$为可训练参数。因此, 基于 $\boldsymbol{a}(x)$即可获取电磁信号特征 x 的高阶统计量。由于电磁信号特征提取过程中, $\boldsymbol{a}(x)$在信号不同时刻共享使用, 因此, 该过程通过卷积及矩阵 Hadamard 乘积运算实现。高阶卷积模块在传统卷积算法上融入高阶统计信息, 实现电磁信号有效特征增强并抑制背景噪声。

本文网络结构参数如表 1 所示。

表 1　网络结构参数

Table 1　Network structure parameters

网络层	卷积核	输出维度
输入层		1 024×2
残差层	32, 3×2	512×32
高阶卷积层	1, 1×1	512×32
多谱注意力层		512×32
残差层	32, 3×1	256×32
残差层	32, 3×1	128×32
高阶卷积层	1, 1×1	128×32
多谱注意力层		128×32
残差层	32, 3×1	64×32
残差层	32, 3×1	32×32
高阶卷积层	1, 1×1	32×32
多谱注意力层		32×32
残差层	32, 3×1	16×32
全连接层		128
全连接层		24

1.3　多谱通道注意力机制

传统通道注意力机制基于 GAP 计算注意力向量, 但该方法忽略电磁信号高频分量, 致使特征频谱信息缺失。因此, 在传统通道注意力基础上, 引入离散余弦变换(discrete cosine transform, DCT)替代 GAP, 构建多谱通道注意力机制, 如图 3 所示。该机制从频域角度出发, 针对 GAP 运算致使特征信息缺失问题, 从电磁频谱角度实现高效特征筛选, 显著提升特征表征能力及可解释程度。

基于数字信号处理技术电磁频谱分析方法启发, 在 SE 模块基础上引入 DCT。其定义如下:

$$\boldsymbol{f}_k=\sum_{i=0}^{L-1}\boldsymbol{X}_i\cos\left(\frac{\pi k}{L}\left(i+\frac{1}{2}\right)\right)\qquad k\in\{0,1,\cdots,L-1\} \tag{11}$$

式中: $\boldsymbol{f}$为电磁信号 DCT 频谱; $\boldsymbol{X}$为卷积特征向量层; L 为一维序列电磁信号长度。令式(11)中 k 为 0, 则

$$\boldsymbol{f}_0=\sum_{i=0}^{L-1}\boldsymbol{X}_i\cos\left(\frac{0}{L}\left(i+\frac{1}{2}\right)\right)=\sum_{i=0}^{L-1}\boldsymbol{X}_i=G(\boldsymbol{X}_i) \tag{12}$$

式中: $\boldsymbol{f}_0$为 DCT 变换后的低频分量, 即 GAP 仅为 DCT 的低频率分量。综上所述并结合 DCT 反变换 $\boldsymbol{X}_i=\sum_{k=0}^{L-1}\boldsymbol{f}_k\cos\left(\frac{\pi k}{L}\left(i+\frac{1}{2}\right)\right),i\in\{0,1,\cdots,L-1\}$, 设 $\boldsymbol{B}_k^i=\cos\left(\frac{\pi k}{L}\left(i+\frac{1}{2}\right)\right)$, 卷积层电磁特征可表示为

$$\begin{aligned}\boldsymbol{X}_i=\sum_{k=0}^{L-1}\boldsymbol{f}_k\cos\left(\frac{\pi k}{L}\left(i+\frac{1}{2}\right)\right)&=\boldsymbol{f}_0\boldsymbol{B}_0^i+\boldsymbol{f}_1\boldsymbol{B}_1^i+\cdots+\\&\boldsymbol{f}_{L-1}^i\boldsymbol{B}_{L-1}^i=G(\boldsymbol{X})L\boldsymbol{B}_0^i+\boldsymbol{f}_1\boldsymbol{B}_1^i+\cdots+\boldsymbol{f}_{L-1}^i\boldsymbol{B}_{L-1}^i\\&i\in\{0,1,\cdots,L-1\}\end{aligned} \tag{13}$$

可见, GAP 仅通过低频分量生成注意力向量, 忽视大部分频谱分量导致电磁信息缺失。在 SE 模

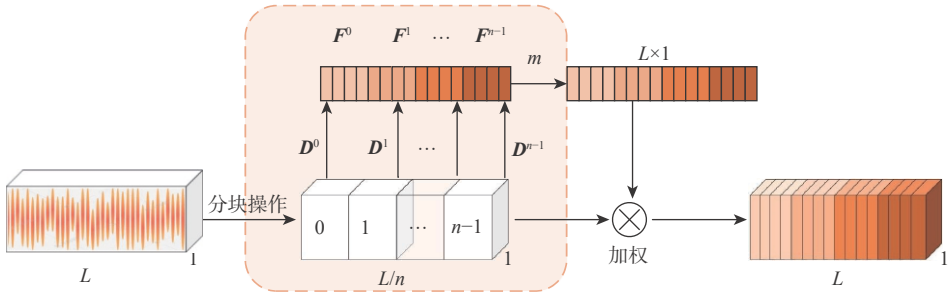


图3 多谱通道注意力机制

Fig. 3 Multispectral channel attention mechanism

块基础上引入DCT，能够从电磁频谱角度提升注意力向量有效性，实现高效电磁特征筛选并显著提升网络性能。在本文中，卷积层电磁特征$\boldsymbol{X}$沿通道维度C切割，电磁特征向量此时可表征为$\boldsymbol{X}_i=[\boldsymbol{X}_0,\boldsymbol{X}_1,\cdots,\boldsymbol{X}_{C-1}]$，此时，特征经过DCT变换后经全连接层及激活函数非线性拟合后得到注意力向量，可表示为

$$\boldsymbol{A}=\mathrm{sigmoid}\left(m\left(\left[\boldsymbol{f}_0,\boldsymbol{f}_1,\cdots,\boldsymbol{f}_{C-1}\right]\right)\right) \tag{14}$$

式中：m函数为多层全连接层与线性变换。m函数输出结果再经过sigmoid函数归一化，最终得到注意力向量。

此时，注意力向量由GAP(低频)和其余频率分量综合考虑计算生成，在频谱分量的角度丰富了电磁信息并提升了网络可解释程度。多谱注意力机制沿特征图通道维度切割向量，对各信道特征分别进行DCT变换，此时虽然向量维度较原始向量未发生变化，但向量已转化为频谱信息，经过非线性拟合后基于频谱信息输出各通道注意力权重，最终通过各通道注意力权重实现多谱通道注意力机制。

2 实验结果及分析

2.1 实验设置

1）RML2018电磁信号数据集

RadioML数据集由O'Shea团队[23]于2019年公开，该数据集面向通用无线电磁信号，包含常规模拟和数字调制方式。RML2018数据集采用开源软件平台GNU Radio生成辐射源基带数字电磁信号，基于已生成的信号数据，通过无线电硬件外设B210实现射频无线收发功能，以上过程在实验室环境下完成。数据集电磁信号生成流程如图4所示。RML2018数据集包含24种调制方式，信噪比范围为−20～30 dB，每条样本采样点数为2×1 024(I/Q信号)，电磁信号样本共2 555 904条，本文实验采用训练集样本20%，验证集样本20%，测试集样本60%。数据集调制方式包含5种模拟调制，19种数字调制方式，具体如表2所示。

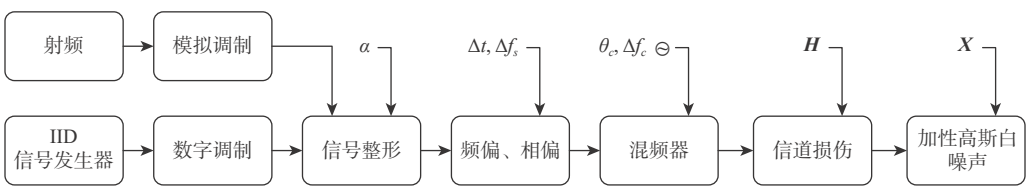


图4 数据集信号生成流程

Fig. 4 Dataset signal generation process

表2 数据集包含调制类型

Table 2 Dataset contains the modulation type

分类	调制方式
模拟调制	AM-SSB-WC/AM-SSBSC/AM-DSB-WC/AM-DSB-SC/FM
数字调制	OOK/4ASK/8ASK/BPSK/QPSK/8PSK/16PSK/32PSK/16APSK/32APSK/64APSK/128APSK/16QAM/32QAM/64QAM/128QAM/256QAM/ /GMSK/OQPSK

该数据集模拟真实无线环境，数据分类难点如下：①调制方式种类较多，包含各种常见模拟调制、数字调制方式，且数据调制阶数较高(如QAM256和APSK256)；②未设定相位跟踪情况下，初始相位不定；③电磁数据星座图差异不明显，为增加分类难度添加时偏、频偏、相偏及热噪声等实际影响因素，如图4所示，在RML2018数据集的信号生成流程中，α为模拟调制调控参数，用于定义调制操作的幅度增益、调制深度等核心特性；Δt、Δf_s分别代表信号传输过程中的频偏、相位噪声系数与初始相位，共同刻画信号在时域与频域的失真特性；θ_c、Δf_c是失真参数的叠加单元，实现频偏、相位噪声向信号的物理层注入；$\boldsymbol{H}$对应信道模型，用于模拟信号传输过程中的信道衰落、滤波等实际传输特性；最终经信道与加性高斯噪声作用后，输出的$\boldsymbol{X}$即为RML2018数据集中的目标信道信号样本，包含了调制、失真、信道及噪声等多维度特征。以上3种因素给调制识别带来了一定的难

度, 也是本领域力图解决的问题。数据集的 24 种调制方式可视化如图 5 所示。

2) 实验环境与平台

电磁信号调制分类实验的软件系统环境为 Windows 10, 深度神经网络训练及测试基于 Google 开发和维护的 TensorFlow 2.0 计算平台进行, 编程语言为 Python 3.10。

实验硬件环境为 Nvidia RTX A2000 的 GPU 图形处理器, 显存为 12 GB、计算机搭载 Intel Core i7-12700K 处理器、64 GB RAM 内存。

此外, 实验使用 NVIDIA 的 GPU 加速库 cuDNN 来加速深度神经网络的计算过程。

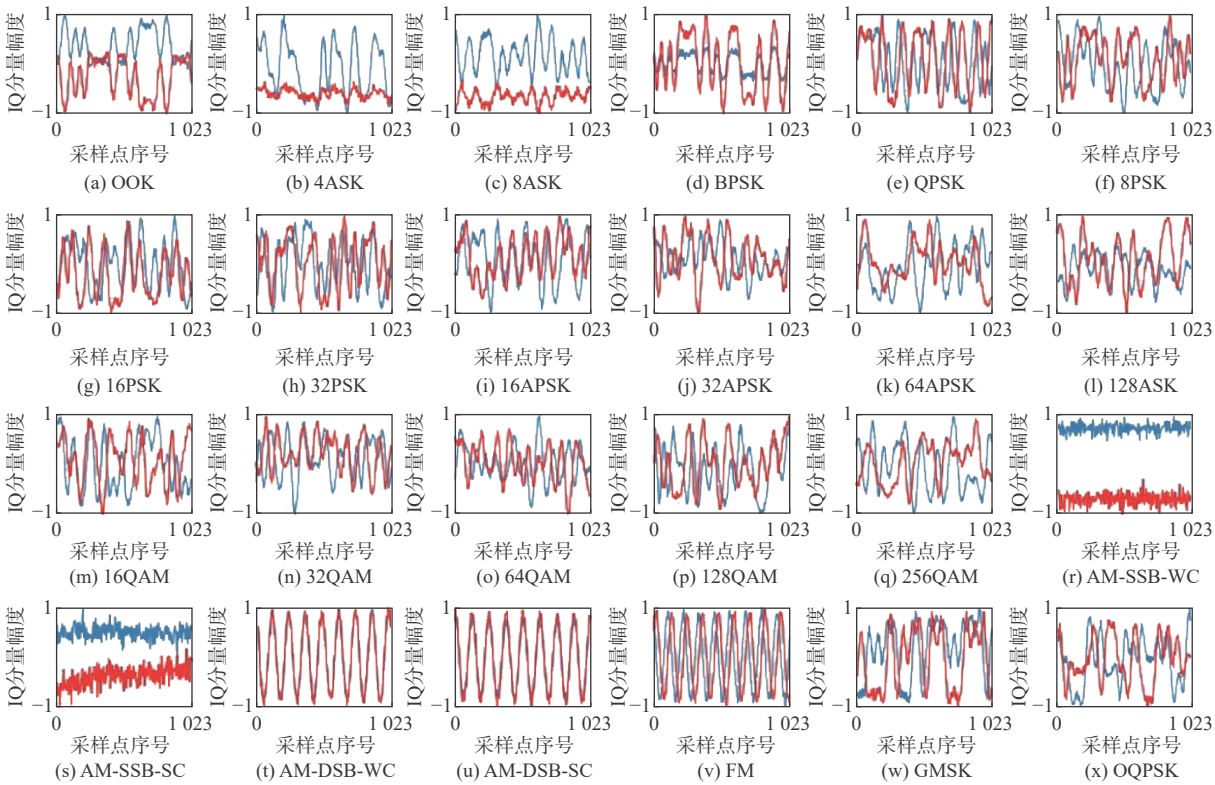


图 5　数据集可视化

Fig. 5　Dataset visualization

2.2　实验分析

2.2.1　模型评估指标

该实验采用平均识别准确率 E_{acc} 作为衡量调制识别性能的指标, 其定义如下:

$$E_{\text{acc}} = n_{\text{correct}}/n_{\text{total}} \tag{15}$$

式中: n_{correct} 为 N 分类中分类正确的样本数; n_{total} 为整体样本数。

2.2.2　模型对比实验

本文基于 4 个深度学习网络模型进行对比试验, 分别是基于 ResNet 模型[23]、SENet 模型[24]、基于稳健自动调制分类的高效 CNN 架构(MSNet 模型)[25]、基于两层门控循环单元网络的 RNN 高效架构(GRU 模型)[26]。通过对比实验证明, 本文改进的领域知识内嵌网络模型识别性能有所上升, 说明对特征进行筛选融合能够有效抑制噪声的影响。进而对比本文的改进方法, 在不明显增加参数量的同时, 使整体的识别准确率相较以上 ResNet 模型、SENet 模型、MSNet 模型、GRU 模型均有不同幅度的提升, 分别提升了 6.31%、3.16%、5.05%、2.44%。

图 6 为对比模型和本文模型在不同信噪比下的识别准确率, 可见曲线有 2 个明显的拐点, 一是在−14 dB 左右, 准确率达到 5% 左右, 二是在 10 dB 以上, 准确率达到 90% 左右。准确率折线图说明: ①对比模型与本文模型的识别准确率均随信噪比的增加而增加, 并且当信噪比大于 6 dB 时识别准

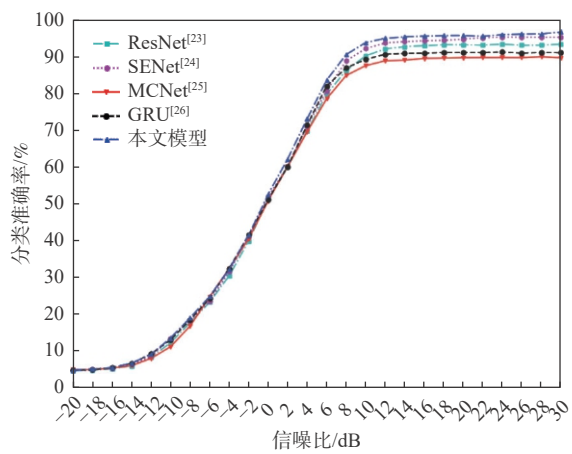


图 6　模型准确率对比图

Fig. 6　Model accuracy contrast chart

确率趋于稳定。随着信噪比的提升，接收数据中有用信息占比增加，模型可以更准确地学习电磁信号特征并进行分类。②图中所有模型在信噪比低于0 dB时，识别准确率均不足50%，识别性能较差。随着信噪比的降低，接收信号含有的有效信息减少，模型难以学习到有效特征并进行准确分类。未来需要进一步研究极低信噪比情况下的识别方法。③2个拐点的区间识别准确率随信噪比近似线性提升，说明信噪比增加会降低噪声对信号的影响且调制信号中包含的特征更为明显，从而提高分类的精度。④在10 dB以上分类准确率可以达到90%以上，高信噪比区间达到比较满意的识别效果。总体而言，在信噪比大于0 dB时，本文模型可以更准确地捕捉和分析信号特征，以较高的精度识别出多数调制方式。

2.2.3 实验混淆矩阵

如图7为ResNet模型测试的混淆矩阵，图8为MSNet模型测试的混淆矩阵，图9为本文模型测试

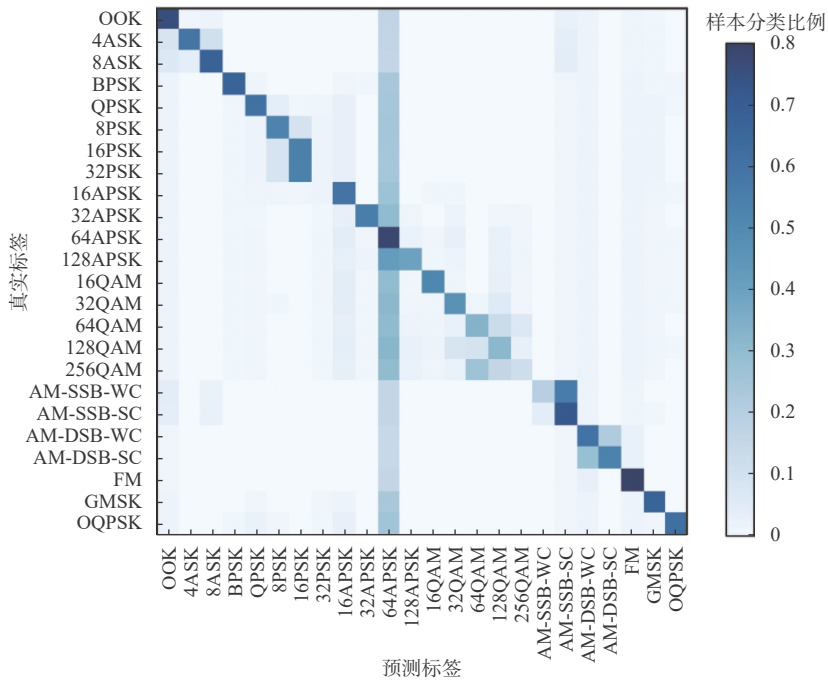


图7 ResNet模型混淆矩阵

Fig. 7 ResNet model confusion matrix

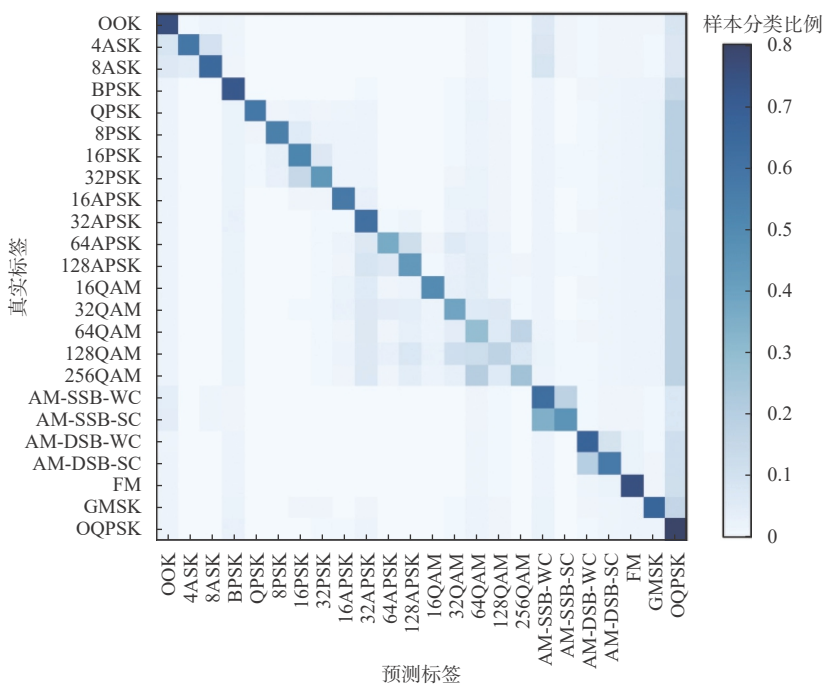


图8 MSNet模型混淆矩阵

Fig. 8 MSNet model confusion matrix

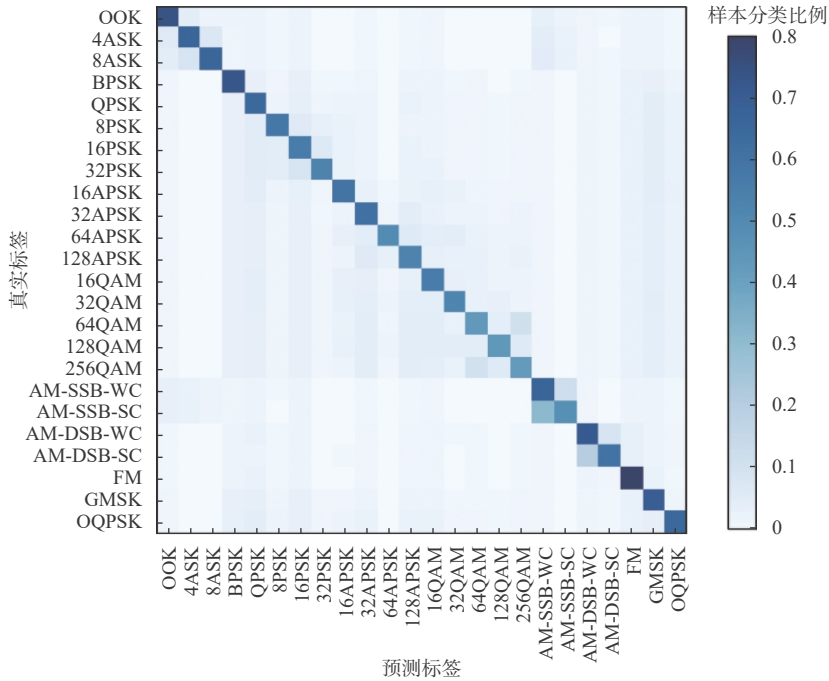


图 9　本文模型混淆矩阵

Fig. 9　The proposed model confusion matrix

的混淆矩阵。其中, ResNet 模型在 PSK 族群和 QAM 族群有明显混淆且对 64APSK 调制信号分类效果最差; MSNet 模型在 APSK 和 QAM 族群调制信号均有混淆; 本文模型的混淆矩阵整体分类效果较好, 能够看到类内(不同阶数的同种调制方式)混淆情况减弱, 但仍然存在; 类间混淆(跨调制方式)情况明显好转。因此, 若想继续提升分类效果, 需减小类内差距, 尤其是高阶相位调制。

为更好地理解限制平均识别准确率的因素, 图 10～图 13 中展示了本文模型在信噪比为 0 dB、14 dB、20 dB、30 dB 下的混淆矩阵。结果表明: ①在低信噪比下, 混淆矩阵对角线均变得不清晰, 整体分类性能不佳, 相位调制混淆程度严重。但本文模型仍以较高的准确率识别出 OOK、BPSK、FM、GMSK 及 OQPSK; ②信噪比为 0 dB 的情况下, 已经能够区分频率调制、幅度调制的不同阶数调制

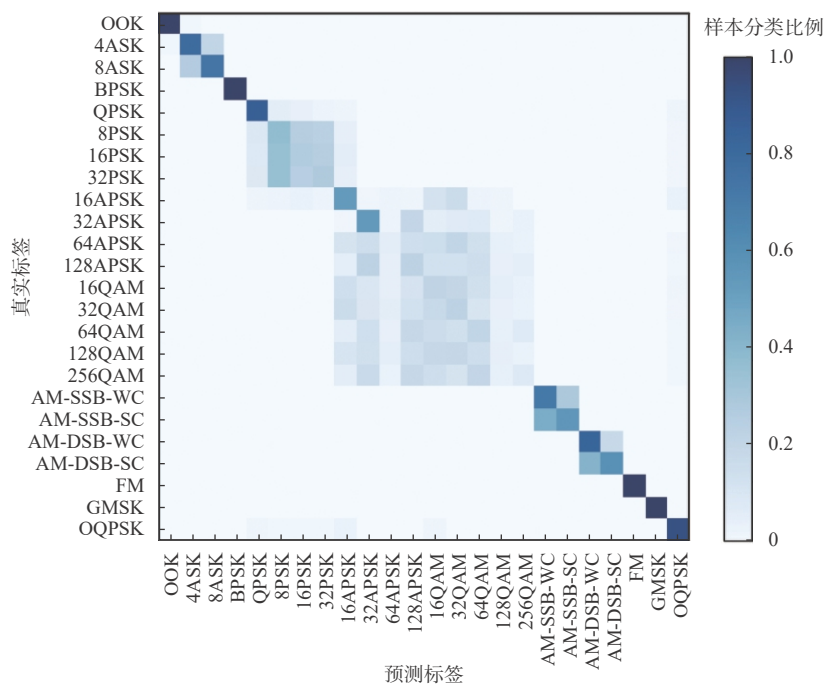


图 10　本文模型信噪比 0 dB 的混淆矩阵

Fig. 10　The proposed model in SNR 0 dB confusion matrix

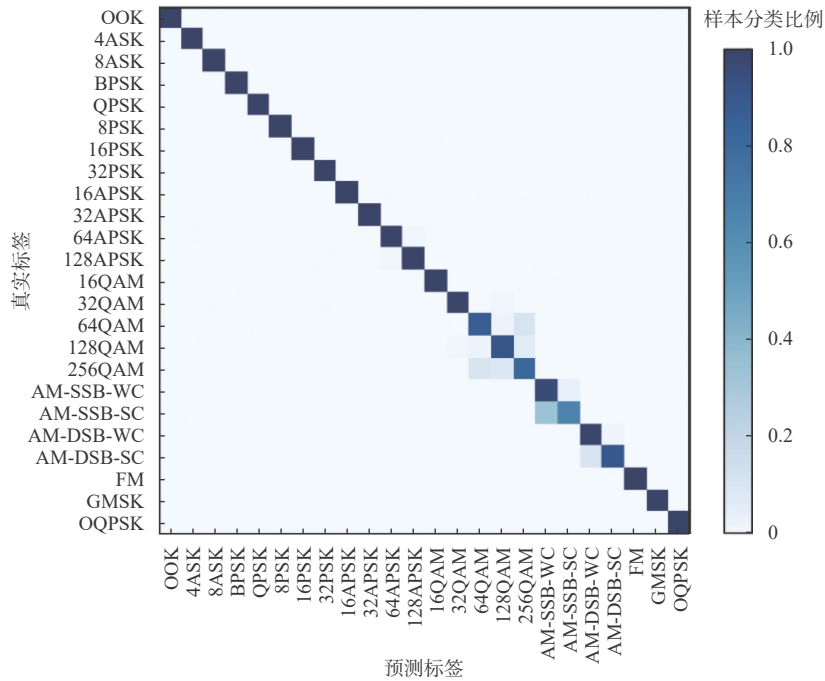


图 11 本文模型信噪比 14 dB 的混淆矩阵

Fig. 11 The proposed model in SNR 14 dB confusion matrix

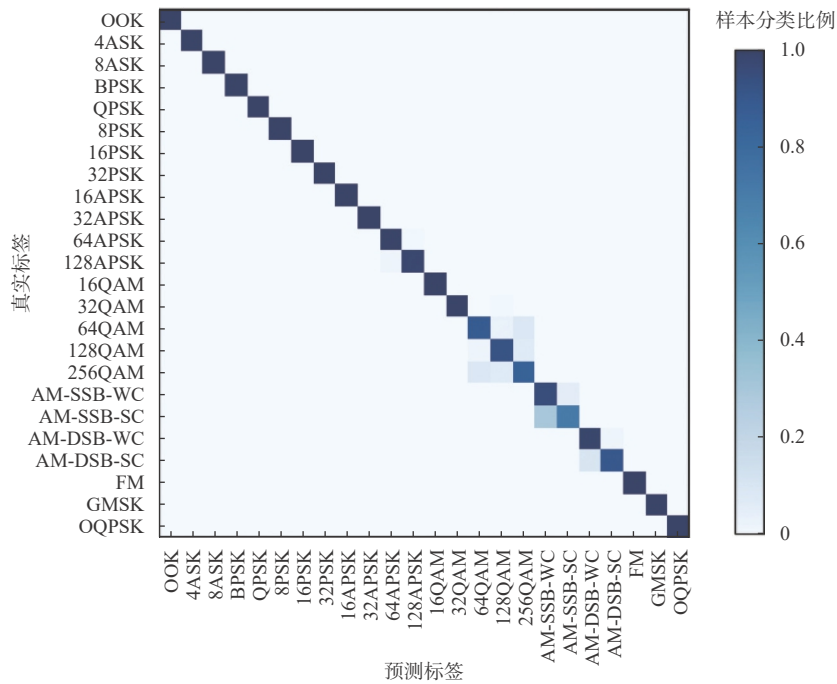


图 12 本文模型信噪比 20 dB 的混淆矩阵

Fig. 12 The proposed model in SNR 20 dB confusion matrix

方式，但分类效果不佳；③信噪比大于 14 dB 情况下，本文模型混淆程度明显减弱，本文模型仅有部分 QAM 调制出现混淆，但总体各类别的分类效果均达到理想效果。

2.2.4 消融实验

实验对比去掉不同模块的识别准确率，如图 14 所示，单独加入高阶卷积模块 (higher order statistics, HOS) 和多谱通道注意力机制 (discrete cosine transform, DCT) 均能提升电磁信号识别效果，但本文模型效果更佳。通过对比消融实验准确率可知，在 2 dB 以下的模型识别性能效果基本相似，但已能轻微区分，8 dB 以上的范围则拉开了相对较大的差距，本文模型较单独融入高阶卷积模块和多谱通道注意力机制分别提升 2% 和 3% 的识别准确率。本文方法能够实现有效特征增强且在一定程度上抑制噪声对网络性能的影响。

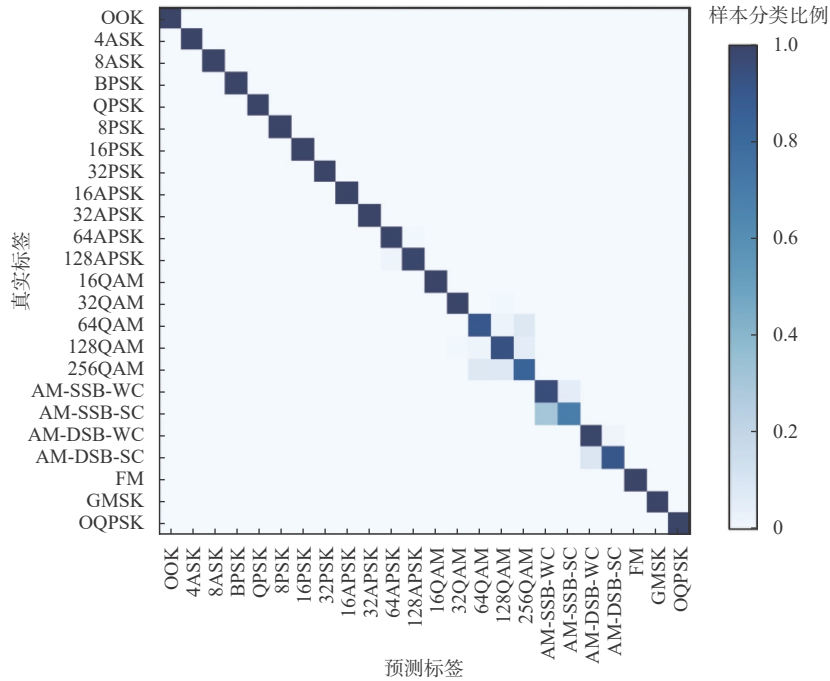


图 13　本文模型信噪比 30 dB 的混淆矩阵

Fig. 13　The proposed model in SNR 30 dB confusion matrix

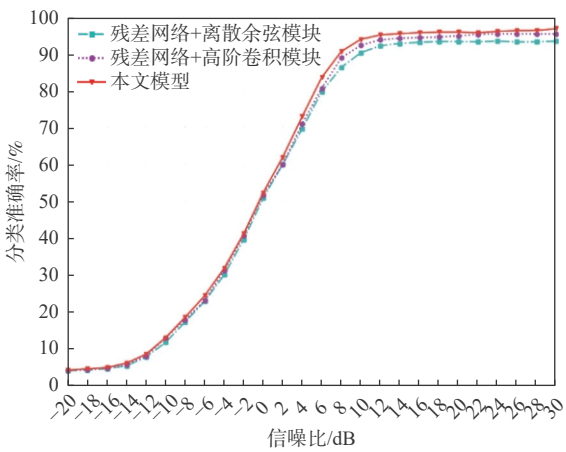


图 14　消融实验

Fig. 14　Ablation experiments

3　结　论

本文针对传统卷积算法采用局部感受野的特性, 将高阶信息融入传统卷积算法有效捕捉全局信息且获取电磁信号高阶特征。为解决传统注意力机制基于 GAP 计算忽略电磁信号高频分量的难点, 在传统通道注意力基础上, 引入 DCT 实现电磁特征高效筛选且显著提升特征表征能力。得到以下结论:

1) 本文算法可实现较为优异的信号调制识别分类性能, 在 RML2018 数据集上验证了本文方法的有效性。

2) 本文算法识别准确率在 10 dB 以上分类准确率可以达到 90% 以上, 高信噪比区间达到比较满意的识别效果。

3) 类内(不同阶数的同种调制方式)混淆情况减弱, 但仍然存在; 类间混淆(跨调制方式)情况明显好转。

综合本文算法性能, 本文方法的平均识别率是 60.22%, 整体的识别准确率较 ResNet 模型、SENet 模型、MSNet 模型、GRU 模型分别提升了 6.31%、3.16%、5.05%、2.44%。在 6～30 dB 下模型对调制信号分类性能最优, 在 10 dB 以上准确率可达到 90% 以上, 在低信噪比下识别效果还有待提升。

参考文献（References）

[1] BKASSINY M, JAYAWEERA S K, LI Y, et al. Wideband spectrum sensing and non-parametric signal classification for autonomous self-learning cognitive radios[J]. IEEE Transactions on Wireless Communications, 2012, 11(7): 2596-2605.

[2] WEI W, MENDEL J M. Maximum-likelihood classification for digital amplitude-phase modulations[J]. IEEE Transactions on Communications, 2000, 48(2): 189-193.

[3] PANAGIOTOU P, ANASTASOPOULOS A, POLYDOROS A. Likelihood ratio tests for modulation classification[C]//Proceedings of the 21st Century Military Communications. Architectures and Technologies for Information Superiority. Piscataway: IEEE Press, 2002: 670-674.

[4] HAMEED F, DOBRE O A, POPESCU D C. On the likelihood-based approach to modulation classification[J]. IEEE Transactions on Wireless Communications, 2009, 8(12): 5884-5892.

[5] LI S T, QUAN D Y, WANG X F, et al. LPI radar signal modulation recognition with feature fusion based on time- frequency transforms[C]//Proceedings of the 13th International Conference on

Wireless Communications and Signal Processing. Piscataway: IEEE Press, 2021: 1-6.

[6] NANDI A K, AZZOUZ E E. Modulation recognition using artificial neural networks[J]. Signal Processing, 1997, 56(2): 165-175.

[7] NANDI A K, AZZOUZ E E. Algorithms for automatic modulation recognition of communication signals[J]. IEEE Transactions on Communications, 1998, 46(4): 431-436.

[8] SHI F N, JING X J, HE Y, et al. Classification model of wireless signals based on higher-order statistics[C]//Proceedings of the IEEE International Symposium on Broadband Multimedia System and Broadcasting. Piscataway: IEEE Press, 2020: 1-5.

[9] 李跃, 郭兴吉, 赵欣. 基于高阶累积量的调制方式识别研究[J]. 西南科技大学学报, 2018, 33(3): 64-68.
LI Y, GUO X J, ZHAO X. Study on modulation recognition based on higher-order cumulants[J]. Journal of Southwest University of Science and Technology, 2018, 33(3): 64-68(in Chinese).

[10] RUAN G X, ZHAO L, LIU Z. Recognition of continuous phase modulation signals based on synchrosqueezing wavelet transform[C]//Proceedings of the International Applied Computational Electromagnetics Society Symposium. Piscataway: IEEE Press, 2022: 1-2.

[11] 谭晓衡, 褚国星, 张雪静, 等. 基于高阶累积量和小波变换的调制识别算法[J]. 系统工程与电子技术, 2018, 40(1): 171-177.
TAN X H, CHU G X, ZHANG X J, et al. Modulation recognition algorithm based on high-order cumulants and wavelet transform[J]. Systems Engineering and Electronics, 2018, 40(1): 171-177(in Chinese).

[12] HAN G, LI J D, LU D H. Study of modulation recognition based on HOCs and SVM[C]//Proceedings of the IEEE 59th Vehicular Technology Conference. Piscataway: IEEE Press, 2004: 898-902.

[13] TENG X Y, TIAN P W, YU H Y. Modulation classification based on spectral correlation and SVM[C]//Proceedings of the 4th International Conference on Wireless Communications, Networking and Mobile Computing. Piscataway: IEEE Press, 2008: 1-4.

[14] MUSTAFA H, DOROSLOVACKI M. Digital modulation recognition using support vector machine classifier[C]//Proceedings of the Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers. Piscataway: IEEE Press, 2005: 2238-2242.

[15] JAGANNATH J, POLOSKY N, O'CONNOR D, et al. Artificial neural network based automatic modulation classification over a software defined radio testbed[C]//Proceedings of the IEEE International Conference on Communications. Piscataway: IEEE Press, 2018: 1-6.

[16] TAYAKOUT H, DAYOUB I, GHANEM K, et al. Automatic modulation classification for D-STBC cooperative relaying networks[J]. IEEE Wireless Communications Letters, 2018, 7(5): 780-783.

[17] WAHLA A H, CHEN L, WANG Y L, et al. Automatic wireless signal classification in multimedia Internet of Things: an adaptive boosting enabled approach[J]. IEEE Access, 2019, 7: 160334-160344.

[18] MA K, ZHOU Y B, CHEN J Y. CNN-based automatic modulation recognition of wireless signal[C]//Proceedings of the IEEE 3rd International Conference on Information Systems and Computer Aided Education. Piscataway: IEEE Press, 2020: 654-659.

[19] 邵敏兰, 周鸿渐, 张浩然. 基于卷积神经网络的调制信号识别算法[J]. 实验室研究与探索, 2021, 40(8): 28-31.
SHAO M L, ZHOU H J, ZHANG H R. Modulation signal recognition algorithm based on convolutional neural network[J]. Research and Exploration in Laboratory, 2021, 40(8): 28-31(in Chinese).

[20] 查雄, 彭华, 秦鑫, 等. 基于循环神经网络的卫星幅相信号调制识别与解调算法[J]. 电子学报, 2019, 47(11): 2443-2448.
ZHA X, PENG H, QIN X, et al. Satellite amplitude-phase signals modulation identification and demodulation algorithm based on the cyclic neural network[J]. Acta Electronica Sinica, 2019, 47(11): 2443-2448(in Chinese).

[21] PANCHAPAGESAN S, PARK D S, CHIU C C, et al. Efficient knowledge distillation for RNN-transducer models[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2021: 5639-5643.

[22] DING Z. Matrix outer-product decomposition method for blind multiple channel identification[J]. IEEE Transactions on Signal Processing, 1997, 45(12): 3053-3061.

[23] O'SHEA T J, ROY T, CLANCY T C. Over-the-air deep learning based radio signal classification[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(1): 168-179.

[24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[25] LIU K, LI F. Automatic modulation recognition based on a multiscale network with statistical features[J]. Physical Communication, 2023, 58: 102052.

[26] SUN S Q, WANG Y Y. A novel deep learning automatic modulation classifier with fusion of multichannel information using GRU[J]. EURASIP Journal on Wireless Communications and Networking, 2023, 2023(1): 66.

Intelligent recognition of electromagnetic signal modulation with embedded domain knowledge

ZHAO Hongjia[1]，ZHANG Duona[1,*]，LU Yuanyao[1]，DING Wenrui[2]

(1. School of Information Science and Technology，North China University of Technology，Beijing 100144，China；
2. Unmanned Systems Research Institute，Beihang University，Beijing 100191，China)

Abstract: With the increasingly complex electromagnetic environment, wireless communication is facing severe challenges, making modulation recognition of electromagnetic signals, which becomes an important aspect of cognitive radio technology. Deep learning techniques have poor interpretability and little applicability, while traditional identification techniques have limited representation capabilities. In this paper, we propose an intelligent modulation recognition method that combines the advantages of both methods by embedding domain knowledge. In order to enhance classification performance and network interpretability, this technique integrates deep neural networks with high-order information and electromagnetic signal spectrum processes. Based on the RML2018 dataset, our method achieves a 6.31% improvement in modulation recognition accuracy compared to the ResNet method.

Keywords: electromagnetic characteristics; deep neural network; modulation recognition; wireless communications; knowledge-embedded

Received：2023-11-20；**Accepted**：2024-03-11；**Published Online**：2024-03-13 14：41

URL：link.cnki.net/urlid/11.2625.V.20240313.1402.001

Foundation items：National Natural Science Foundation of China (62201010); R&D Program of Beijing Municipal Education Commission (KM202310009003)

* **Corresponding author**. E-mail：zhangduona@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0763

双磁路旋转洛伦兹力磁轴承设计与分析

庞维坤[1]，王卫杰[2,*]，樊亚洪[3]，李磊[1]，杨洋[4]，朱宏业[1]

(1. 航天工程大学 研究生院，北京 101416；2. 航天工程大学 宇航科学与技术系，北京 101416；
3. 北京控制工程研究所，北京 100190；4. 河南大学 人工智能学院，郑州 475001)

摘 要：面向复杂航天任务对航天器有效载荷万向敏捷和超精指向性能迫切需求，提出一种洛伦兹力磁悬浮万向稳定平台，开展双磁路旋转洛伦兹力磁轴承设计与分析。采用动圈式转子方案，转子组件4条挂耳形线圈成对绕置并胶装于骨架轴向两侧凹槽内；定子组件成对平行布设共转轴、双环状轴向充磁磁钢，形成均匀性稳定磁密，为载荷舱敏捷机动提供周向双通道对称工作气隙。基于等效磁路法建立气隙磁密模型，从气隙磁密均匀度和波动率2方面定义磁密线性度，进而开展转子旋转动力学建模，构建旋转力矩模型。运用Maxwell有限元法建立旋转磁轴承有限元模型，并进行算例仿真，结果表明：所述旋转磁轴承方案气隙旋转包络中心位置处磁密可达685.624 mT，周向磁密均匀度为99.72%，大幅改善气隙磁密均匀性，避免了径向充磁式方案气隙磁密衰减和纵向磁场扩散的局限，有效提升载荷舱旋转状态下的稳定度和指向精度。

关 键 词：轴向充磁；扇形磁钢环；动圈式转子；磁密线性度；等效磁路法；有限元分析
中图分类号：V414.1；V448.2
文献标志码：A **文章编号**：1001-5965(2026)01-0306-11

当前，高分侦察、激光通信等航天任务对航天器姿态超稳定性、超敏捷性和超静性能提出极高要求，亟须探索发展航天器姿态控制技术新途径。磁悬浮轴承相比传统机械式轴承具有无摩擦、低功耗、长寿命的优势特性[1]，基于磁悬浮的航天器姿态控制已成为热点研究领域。

磁悬浮技术在航天器姿态控制领域应用可分为部件级和系统级2个层面。其中，部件级磁悬浮姿态控制技术即磁悬浮惯性执行机构，以磁悬浮飞轮和磁悬浮控制力矩陀螺为代表[2-4]，技术已趋于成熟，尚有待推广应用。在航天器系统级磁悬浮姿态控制技术方面，航天八院提出了一种双超卫星平台[5]，采用8个洛伦兹力磁轴承实现载荷与星体之间解耦分离；哈尔滨工业大学联合航天五院研制并设计了一种旋转光学卫星[6]，该卫星为提高遥感卫星与其上旋转载荷中高频振动传递，将洛伦兹力磁悬浮轴承作为旋转关节，实现载荷与平台的软连接，提高载荷轴向稳定性。上述2种方案显著提升了卫星载荷的超稳定性和高指向精度，但航天八院设计的双超卫星平台姿态机动方式仍采用传统卫星式，哈尔滨工业大学设计的旋转光学卫星仅具有单自由度机动性能，对未来高精度、高动态响应背景下航天器姿态万向敏捷迫切需求适应性弱。航天五院研制的北京三号卫星[7]为获得三超性能，采用基于洛伦兹力磁悬浮作动器的Stewart平台，构成超静卫星多级复合控制系统[8]，但其平台姿态机动性能较弱。美国Draper实验室研制出一种球面洛伦兹力磁悬浮框架飞轮(magnetically suspended gimballing flywheel, MSGFW)原理样机；德国Teldix公司基于洛伦兹力磁轴承研制了一种MSGFW样机

收稿日期：2023-11-22；**录用日期**：2024-01-02；**网络出版时间**：2024-01-15 21:14
网络出版地址：link.cnki.net/urlid/11.2625.V.20240115.1340.004
基金项目：国家自然科学基金(52075545)
*通信作者. E-mail：wangwjie@126.com

引用格式：庞维坤，王卫杰，樊亚洪，等. 双磁路旋转洛伦兹力磁轴承设计与分析[J]. 北京航空航天大学学报，2026，52(1)：306-316.
PANG W K，WANG W J，FAN Y H，et al. Design and analysis of double magnetic circuit rotating Lorentz force magnetic bearing[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：306-316 (in Chinese).

作为航天器执行机构[9]。Lahdo 等设计的洛伦兹力双执行器为高精度磁悬浮定位系统提供一种结构紧凑、成本效益高的解决方案[10]。Borque 等设计的洛伦兹力演示器可用于测量磁轴承空间致动器中产生的微振动并确定其受力情况[11]。Zhu 等设计的洛伦兹力磁性致动器为低地球轨道小型卫星姿态控制的一种可靠技术[12]。Petit 等设计的洛伦兹力磁转矩器将刚性航天器的角速度稳定为零[13]。综上可知, 基于洛伦兹力磁悬浮技术提升航天器姿控性能已成为当前国内外相关单位的选择, 如何设计并应用是发展洛伦兹力技术是磁悬浮姿态控制技术的重要基础。

文献 [14] 介绍了传统洛伦兹力磁轴承即音圈电机, 但因磁密不均匀而存在干扰力矩。文献 [15-16] 提出对磁轴承输出力矩进行控制补偿, 但未能有效解决气隙磁密不均匀的缺陷。文献 [17] 提出球面梯形磁钢, 改善了气隙磁密均匀性, 但球面磁钢加工工艺复杂, 装配难度大, 难以广泛应用。文献 [18-19] 仅考虑提高偏航角精度而忽视其他重要性能指标, 未能兼顾提升磁密均匀性和磁通密度。为提高磁轴承力矩精度, 文献 [20-21] 提出了混合力构型磁悬浮陀螺飞轮, 采用永磁偏置磁轴承控制径向平动, 洛伦兹力磁轴承控制径向 2 自由度偏转, 轴向则通过被动悬浮支承, 悬浮精度较低。文献 [22] 采用传统洛伦兹力磁轴承, 磁钢两端大面积漏磁影响了磁密均匀性, 进而造成输出力矩精度较低。文献 [23] 提出并设计一种集姿态角速率测量与姿态控制功能于一体的新型磁悬浮控制敏感陀螺, 通过洛伦兹力力矩器感应载体角速率信息, 驱动转子万向偏转输出 2 自由度控制力矩。

综合上述系统级和部件级洛伦兹力磁悬浮发展现状, 航天工程大学王卫杰等[24] 提出一种全洛伦兹力磁悬浮万向稳定平台, 采用 2 级旋转洛伦兹力磁轴承正交布设, 从而驱动载荷沿俯仰、偏航方向快速运动, 实现近 2π 空间万向敏捷机动。其中, 如何设计旋转洛伦兹力磁轴承是研制洛伦兹力磁悬浮万向稳定平台的关键。

针对上述问题, 本文提出并设计一种双闭合磁路、4 并联线圈旋转洛伦兹力磁轴承方案。磁轴承定子组件中成对平行布设共转轴、双环状扇形轴向充磁磁钢, 形成高均匀性稳定磁场; 转子组件成对绕置 4 条挂耳形线圈并胶装于同一绕组骨架轴向两侧凹槽。基于等效磁路法[25-27] 对磁路和气隙磁密进行分析[28-29], 构建转子旋转动力学模型[30-31], 并基于 Maxwell 有限元法仿真验证气隙磁密分布规律及旋转力矩[32] 输出性能。

1　洛伦兹力磁悬浮万向稳定平台构型

洛伦兹力磁悬浮万向稳定平台如图 1 所示, 将传统单体机械式卫星平台拓扑为由星体基座、磁浮舱和载荷舱组成的 3 体 7 自由度磁浮式结构方案。其中, 星体基座通过成对工作的一级旋转洛伦兹力磁轴承驱动磁浮舱和载荷舱整体沿俯仰通道 (x 轴) 敏捷旋转, 磁浮舱通过单个二级旋转洛伦兹力磁轴承驱动载荷舱沿滚动通道 (z 轴) 敏捷旋转。2 级旋转磁轴承转轴正交, 实现载荷近 2π 空间万向敏捷机动。为保证平台结构的可靠性和可控性, 2 级旋转洛伦兹力磁轴承采用相同设计方案。

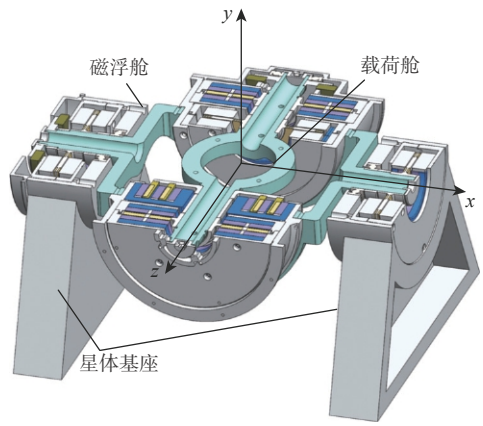


图 1　洛伦兹力磁悬浮万向稳定平台结构

Fig. 1　Lorentz force maglev universal stabilized platform structure

为降低平台质量并缩减其包络尺寸, 平台打破传统对称式结构构型, 除一级旋转磁轴承与双自由度正交磁轴承为对称结构外, 单自由度“双磁钢、双线圈”直线磁轴承、二级旋转磁轴承为非对称式结构。其中, 单自由度直线磁轴承控制载荷舱沿 z 轴平动; 2 个正交磁轴承成对工作, 若 2 者施力方向一致, 则控制载荷舱沿 x 轴和 y 轴 2 自由度平动; 若 2 者施力方向相反, 则形成一对力偶, 抵消载荷舱绕 x 轴和 y 轴的 2 自由度偏转扰动, 完成载荷与平台间物理隔离, 最终实现甚稳超静悬浮。有关直线和正交磁轴承的详细设计方案和具体结构优势不属于本文研究范畴, 故不再赘述。

2　双磁路旋转洛伦兹力磁轴承方案设计

由第 1 节可知, 如何设计旋转洛伦兹力磁轴承是实现洛伦兹力磁浮平台万向敏捷机动的关键, 为此提出一种双闭合旋转洛伦兹力磁轴承结构方案, 并对其结构特点、工作原理、优势特性展开详细设计分析。

2.1 结构设计

旋转洛伦兹力磁轴承主要由定子组件和转子组件组成。如图2所示，定子组件主要包括2块倒马鞍形导磁环、2块隔磁环、4块轴向充磁式扇形磁钢环和2片顺磁贴片；转子组件主要包括4条挂耳形绕组和线圈骨架。

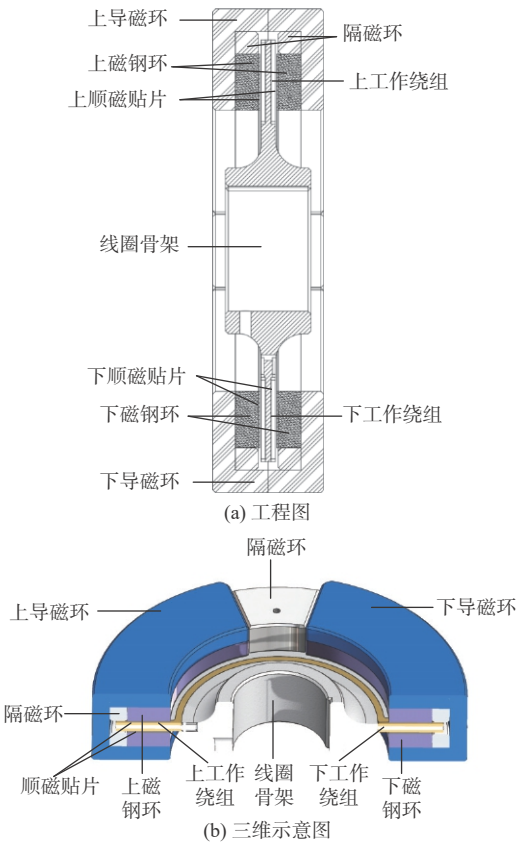


图2 旋转洛伦兹力磁轴承结构

Fig. 2 Structural of a rotating Lorentz force magnetic bearing

定子组件中沿上周向扇形磁钢环、隔磁环、下周向扇形磁钢环依次成对拼装成满环构成磁钢-隔磁部件；沿轴向处气隙左右两侧的磁钢环-隔磁环部件拼接安装在上下相同倒马鞍形导磁环凹槽内，有效避免上下扇形磁钢环间沿径向出现闭合干扰磁路。当转子旋转至两侧边缘时，电流调整大小并换向完成转子加减速；当转子位于稳态旋转时输出稳定转矩。角位移传感器实时监测转子旋转角度及相对位置，随之反馈于控制系统，调整电流参数，改变工作绕组洛伦兹力大小，输出相应转矩，使磁轴承应对外界振动等影响，迅速完成抗干扰应激，确保其平台上载荷甚稳超静悬浮。

为降低磁轴承体积功耗对其性能的影响，如图3所示，提出并设计了“绕组在外、骨架在内”动圈式转子结构方案。如图4可知，4条独立闭合绕组两两对称分布于线圈骨架外侧，相比传统线圈胶装于骨架中心内侧，简化安装步骤并便于后期装配维修。

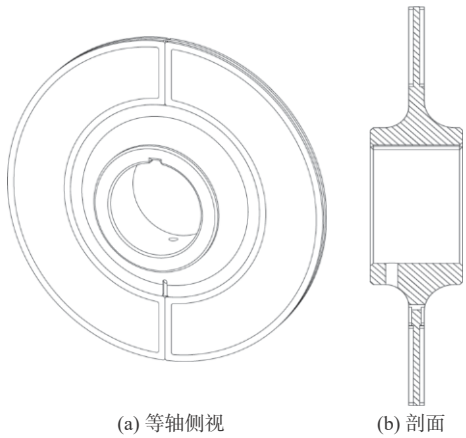


图3 动圈式转子组件工程图

Fig. 3 Moving coil rotor assembly engineering drawing

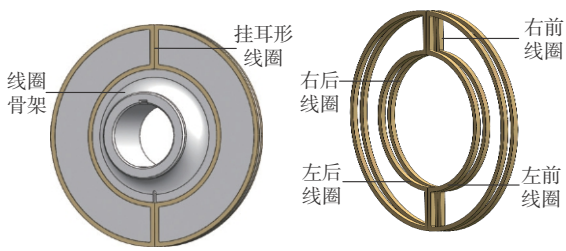


图4 动圈式转子组件示意图

Fig. 4 Schematic of moving coil rotor assembly

挂耳形线圈绕组分为径向工作段和周向非工作段，通电线圈径向工作段绕组置于磁场输出力矩，周向非工作段线圈段置于非磁场区域，避免产生干扰力矩影响转子稳定性。由电磁力产生原理可知，磁密和绕组内电流为输出力矩的直接影响因素，输出相同转矩下，高磁密可缓解响应电流幅值，降低电流功耗，提高转子使用寿命。

本文设计的旋转磁轴承各部件结构参数如表1所示，周向半圆弧段外径大于扇形磁钢环外径尺寸，周向半圆弧段内径小于扇形磁钢环内径尺寸，有效避免了非工作线圈产生干扰力和力矩，提高控制精度。

2.2 工作原理

磁轴承内转子作用于磁气隙中，依据安培力产生原理可知，通电绕组会产生垂直于磁场和电流2个方向的洛伦兹力，可表示为

表1 双磁路旋转洛伦兹力磁轴承结构参数

Table 1 Double magnetic circuit rotating Lorentz force magnetic bearing structure parameters

扇形磁钢环内径/mm	扇形磁钢环外径/mm	扇形磁钢环周向弧度/(°)	扇形磁钢环轴向距离/mm	倒马鞍形导磁环内径/mm	倒马鞍形导磁环外径/mm	线圈径向有效工作长度/mm	单个线圈匝数	绕线规格/mm
50	70	140	8	50	86	16.5	100	0.25×2

$$F = nBIL \tag{1}$$

式中: n 为线圈匝数; B 为气隙磁密; I 为绕组通电电流大小; L 为绕组有效工作长度。当 4 条挂耳形线圈输入相同电流时, 气隙磁密大小和均匀度直接决定绕组输出转子力矩性能。

如图 5 所示, 转子 Oxy 平面内 x 轴正方向处磁场垂直向里, 同理, x 轴负方向处磁场垂直向外。由左手定则可知, 当 4 条并联线圈通入沿 x 轴负方向电流, 上半平面 4 条径向工作段绕组产生沿 y 轴负方向的安培力; 同理, 下半平面 4 条绕组产生沿 y 轴正方向的安培力, 上下端形成力矩, 带动转轴旋转。当平台受外界干扰, 线圈调整电流参数, 输出响应力矩带动转轴补偿平台转动角度。当转子即将到达初/末位置时, 电流瞬时完成换向, 改变线圈受力大小, 进而令转子减速, 转子在稳态旋转时保持电流大小不变使其输出力矩, 沿 Oxy 平面往复运动, 转子在旋转范围内依照平台位置实时动态补偿完成抗干扰。

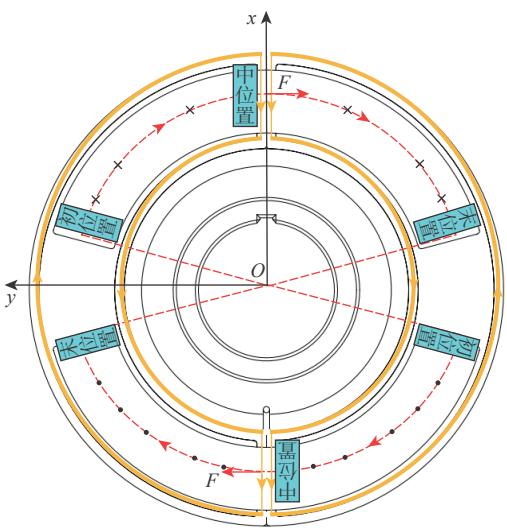


图 5　转子工作原理

Fig. 5　Schematic diagram of the rotor operation

2.3　特点分析

本文旋转洛伦兹力磁轴承设计方案具有 3 个显著特点和优势:

1) 轴向平行充磁方案, 大幅提升气隙磁密均匀性。由磁阻最小原则可知, 磁通沿磁阻最小路径分布, 导磁环、空气和磁钢三者磁导率 (与磁阻成反比[33]) 不同, 合理设计三者结构可抑制气隙磁密纵向扩散趋势, 增大磁通密度并改善磁密均匀性。为提升旋转磁轴承敏捷机动性能, 靠近气隙磁钢表面增贴顺磁贴片, 导磁环与磁钢径向之间安装隔磁环, 倒马鞍形导磁环与扇形磁钢环径向增添隔磁环, 防止单个磁钢环产生自闭合磁路, 最大程度上保证轴向主磁路均匀性好、磁通密度大。为提高旋转磁轴承稳定性能, 分析现有方案中转子结构下磁场分布情况: 同种材质磁钢下轴向充磁式相比同尺寸径向充磁式磁钢气隙磁密均匀性更好、磁通密度更大。

2) 动圈式转子方案, 有效降低磁轴承转子质量, 显著提升负载能力。动圈式转子质量明显低于动磁式, 降低了运动部件质量, 提高了偏转力矩输出精度和快速机动/响应能力, 有效降低结构装配难度并缩减磁轴承包络尺寸, 提高磁轴承运动性能和响应速度。线圈骨架选用陶瓷, 其刚度性能适应于高动态运动下转子往复旋转依旧保持初始性能, 其陶瓷材料作为非导磁材料不存在磁滞及涡流损耗现象, 避免产生干扰力, 从而降低转子启动电流。

3) 圆周方向采用双磁路对称布局方案, 保证扭矩精度的同时, 大幅提升转子旋转范围。为提升旋转磁轴承旋转量程, 沿轴向布设 2 块相同扇形磁钢环安装于导磁环凹槽内, 将单磁路拓扑为双磁路, 增加磁路数量并提高磁密均匀性的同时, 优化闭合中心对称双磁路, 提高气隙磁密。上下气隙周向各处磁密大小一致, 方向相反, 避免对称式外置线圈在旋转过程中因漏磁或磁密不均匀造成转子工作过程中发生振荡, 双磁路布局方案有利于输出高性能、大角度旋转力矩。

3　双磁路旋转洛伦兹力磁轴承动力学建模

在磁轴承方案设计的基础上, 通过构建闭合磁路等效磁路模型, 分析影响磁气隙轴/径向磁通密度因素; 从均匀度和波动率 2 方面定义磁密线性度; 开展旋转磁轴承转子旋转动力学建模, 构建旋转力矩模型。

3.1　特点分析

忽略通电绕组对原磁路产生干扰磁场和边缘漏磁的影响, 为获得真实准确的磁密大小, 将气隙处磁阻分为有效工作处磁阻和非工作处磁阻。

如图 6 所示, 旋转磁轴承磁通回路上闭合磁路中磁感应线从左磁钢 N 极出发经过左顺磁贴片穿过气隙后到达右顺磁贴片、右磁钢 S 极和 N 极后, 磁力线依次通过上导磁环右径向部分、上轴向部分后到达左径向部分形成完整闭合磁路; 同理, 下闭合磁路中磁感应线从左磁钢 S 极出发经过左顺磁贴片穿过气隙后到达右顺磁贴片、右磁钢 N 极和 S 极后, 磁感应线依次通过下导磁环右径向部分、下轴向部分后到达左径向部分形成完整闭合磁路。

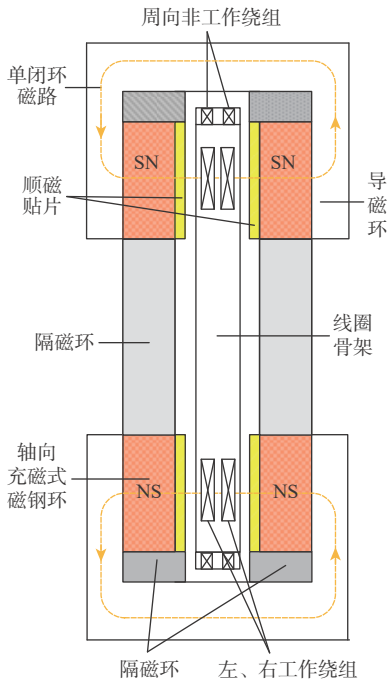


图6 磁轴承磁通回路

Fig. 6 Magnetic circuit diagram of magnetic bearing

基于图6的磁通回路，如图7所示，构建旋转磁轴承等效磁路，上下双磁路间磁力线经过隔磁环 R_{mir1} 和 R_{mir2} 处磁阻无限大，磁通密度近似为零，由此可对其等效双磁路进行简化。

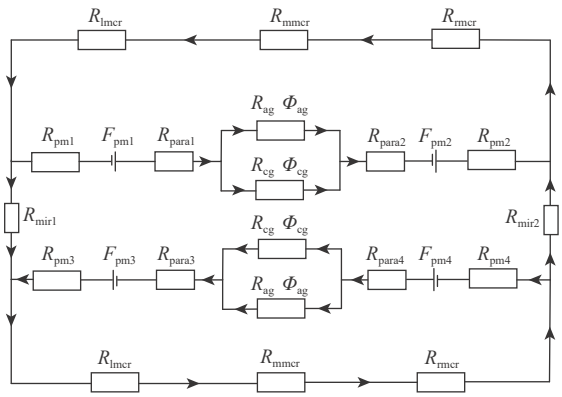


图7 磁轴承等效磁路

Fig. 7 Equivalent magnetic circuit diagram of magnetic bearing

如图8所示，简化后等效磁路中 F_{pm1}、F_{pm2} 为轴向充磁式磁钢的磁动势，R_{pm1}、R_{pm2} 为对应处磁钢磁阻，R_{para1}、R_{para2} 为左右侧顺磁贴片磁阻，R_{rmcr}、R_{lmcr} 为导磁环径向部分磁阻大小，R_{mmcr} 为导磁环轴向部分磁阻大小，Φ_{ag}、Φ_{cg} 分别为线圈有效绕组处磁通和非有效绕组处磁通，R_{ag}、R_{cg} 分别为气隙轴向非工作绕组和工作绕组处磁阻。

磁钢磁动势等于永磁材料矫顽力与充磁长度的乘积，依据图8可得到闭合磁路中总磁动势为

$$F = F_{pm1} + F_{pm2} = H_c l_{pmi} \tag{2}$$

式中：H_c 为磁钢矫顽力；l_{pmi} 为磁钢轴向充磁长度，

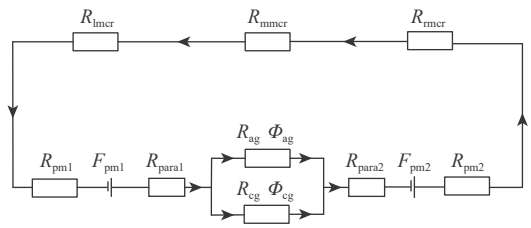


图8 单闭合等效磁路

Fig. 8 Single closed equivalent magnetic circuit

永磁材料均选用钐钴合金且不考虑退磁影响，则 $H_{c1}=H_{c2}$，忽略不同磁钢加工长度误差，即 $l_{pm1}=l_{pm2}$。

可得

$$F_{pm1} = F_{pm2} \tag{3}$$

单闭合磁路中总磁通量 Φ_m 可表示为

$$\Phi_m = \frac{\sum_{i=1}^{2} F_{pmi}}{\sum_{i=1}^{2} R_{pmi} + \sum_{i=1}^{2} R_{parai} + \sum_{i=1}^{2} R_{lmcri} + R_{mmcr} + \dfrac{R_{ag}R_{cg}}{R_{ag}+R_{cg}}} \tag{4}$$

气隙绕组非工作处和工作处磁通量为

$$\boldsymbol{\Phi} = \begin{bmatrix} \Phi_{ag} \\ \Phi_{cg} \end{bmatrix} = \begin{bmatrix} \dfrac{R_{cg}}{R_{ag}+R_{cg}} \\ \dfrac{R_{ag}}{R_{ag}+R_{cg}} \end{bmatrix} \Phi_m \tag{5}$$

气隙非工作绕组和工作绕组处各自磁密为

$$\boldsymbol{B} = \begin{bmatrix} B_{ag} & B_{cg} \end{bmatrix}^{\mathrm{T}} \tag{6}$$

$$\boldsymbol{B} = \frac{\boldsymbol{\Phi}}{\boldsymbol{A}} = \mathrm{diag}(A_{ag}, A_{cg})^{-1}\boldsymbol{\Phi} \tag{7}$$

$$\begin{bmatrix} B_{ag} \\ B_{cg} \end{bmatrix} = \Phi_m \begin{bmatrix} \dfrac{R_{cg}}{A_{ag}(R_{ag}+R_{cg})} \\ \dfrac{R_{ag}}{A_{cg}(R_{ag}+R_{cg})} \end{bmatrix} \tag{8}$$

结合图9结构参数可得扇形轴向充磁磁钢环磁阻：

$$R_{pm1} = R_{pm2} = \frac{r_{pm} - r_{para}}{2\mu_0\mu_r\pi\left(r_{para} + \dfrac{r_{pm}}{2}\right)} \tag{9}$$

式中：r_{pm} 为磁钢的半径；r_{para} 为顺磁贴片的半径；μ 为磁阻率。

顺磁贴片磁阻：

$$R_{para1} = R_{para2} = \frac{r_{para} - r_{cg}}{2\mu_0\mu_r\pi\left(r_{para} + \dfrac{r_{para} - r_{cg}}{2}\right)} \tag{10}$$

式中：r_{cg} 为线圈外侧的长度。

气隙各处磁阻：

$$R_{cg} = \frac{1}{\mu_0\pi h_{cg} r_{cg}} \tag{11}$$

$$R_{ag} = \frac{1}{\mu_0\pi h_{ag} r_{ag}} \tag{12}$$

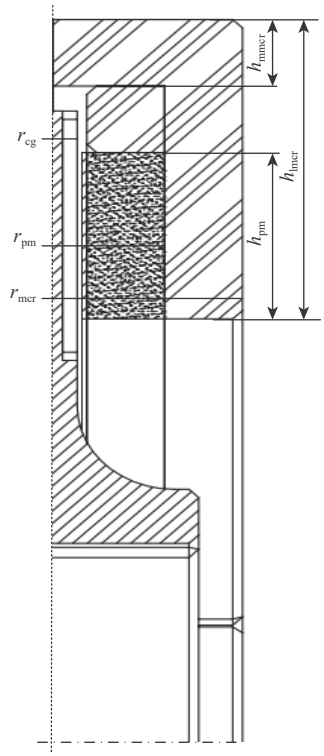


图 9　磁轴承对称上半部件参数尺寸

Fig. 9　Magnetic bearings symmetrical upper part parameter size

式中: h_{cg} 为导磁径向厚度; h_{ag} 和 r_{ag} 分别为有效线组的厚度和距离。

由式(12)求得工作绕组区域下磁密大小:

$$B_{cg}=\frac{\Phi_2}{A_2}=\left(\sum_{i=1}^{2}F_{pmi}R_{ag}\right)\Big/\left\{A_2\left[\sum_{i=1}^{2}R_{pmi}+\sum_{i=1}^{2}R_{parai}+\sum_{i=1}^{2}R_{lmcri}+R_{mmcr}+(R_{ag}+R_{cg})+R_{ag}R_{cg}\right]\right\} \tag{13}$$

3.2　磁密线性度分析

在 3.1 节依照结构搭建等效磁路模型基础上, 为定量分析气隙磁密关键因素, 从气隙磁密均匀度和波动率 2 方面定义磁密线性度。

鉴于气隙磁场存在相应长度, 为确保磁轴承在旋转运动空间内出力线性度良好, 输出稳定转矩, 定义磁密均匀度:

$$\xi=1-\frac{\sqrt{\dfrac{1}{m}\sum_{i=1}^{m}(B_{i1}-B_{avg1})^2}}{B_{avg1}}\times 100\% \tag{14}$$

式中: B_{i1} 为轴向或周向弧段各处磁密大小; B_{avg1} 为区域磁密均值; m 为在轴向间距或周向弧段下取点数量。

为验证双磁路轴向充磁优势方案下气隙中转子上下对应位置磁密对称相等, 确保转子上下出力相同提供稳定转矩, 定义磁密波动率:

$$\eta=\frac{\sum_{i=1}^{m}(B_{i2}-B_{avg2})^2}{m-1}\times 100\% \tag{15}$$

式中: B_{i2} 为径向区域内任意一点的磁密大小; B_{avg2} 为径向区域磁密均值。

3.3　电磁力矩建模

如图 10 所示, 建立 $Oxyz$ 三维直角坐标系。以线圈骨架几何中心为原点 O, 沿线圈上端有效工作段为 x 轴, 垂直于转子平面并指向左侧磁钢为 y 轴, y轴满足右手定则。当有效段位于竖直平面时, 其线圈骨架两侧线圈位于 z 轴正负方向两两对称分布于 Oxy 平面上。选取第 i 组上半部分, 沿轴向选择微元 $\mathrm{d}\theta$, 转子内线圈激励电流大小相等, 同一时刻下电流 I 为恒定值, 则其受力大小为

$$\mathrm{d}f_i=nB_iIL\mathrm{d}\theta \qquad i=1,2,3,4 \tag{16}$$

式中: B_i 为各有效工作绕组处磁通密度; $\mathrm{d}f_i$ 所产生的电磁力矩 $\mathrm{d}\boldsymbol{T}_i$ 位于 Oxy 平面, 可表示为

$$\mathrm{d}\boldsymbol{T}_i=\boldsymbol{r}\times\mathrm{d}f_i=\begin{bmatrix}n\boldsymbol{B}_{i+}IL^2\\ n\boldsymbol{B}_{i-}IL^2\end{bmatrix}\mathrm{d}\theta \tag{17}$$

式中: $\boldsymbol{r}$ 为电磁力 $\mathrm{d}f_i$ 有效作用力臂。则第 i 组有效工作线圈部分上下端产生的有效电磁合力矩为

$$\boldsymbol{T}_i=\int_{\theta_{i1}}^{\theta_{i2}}\begin{bmatrix}n\boldsymbol{B}_{i+}IL^2\\ n\boldsymbol{B}_{i-}IL^2\end{bmatrix}\mathrm{d}\theta=\int_{\theta_{i1}}^{\theta_{i2}}\begin{bmatrix}n\boldsymbol{B}_{i+}IL^2\\ -n\boldsymbol{B}_{i-}IL^2\end{bmatrix}\mathrm{d}\theta \tag{18}$$

式中: θ_{i1} 和 θ_{i2} 分别为第 i 组线圈沿周向弧段的初始和末端角度, 由此可得 4 并联线圈径向一侧产生的有效电磁合力矩为

$$\boldsymbol{T}=\sum_{i=1}^{4}\boldsymbol{T}_i=\frac{8}{3}nIL^2\pi\begin{bmatrix}\boldsymbol{B}_{i+}\\ \boldsymbol{B}_{i-}\end{bmatrix} \tag{19}$$

线圈骨架两侧线圈通入相同方向电流, 上下磁场作用下形成力偶, 有效旋转力矩为

$$\boldsymbol{M}=2\boldsymbol{T}=\frac{16}{3}\pi n\boldsymbol{B}IL^2 \tag{20}$$

从式(20)可以看出, 气隙磁密是决定转子组件输出力矩响应速度快、精度高、稳定性好的直接影

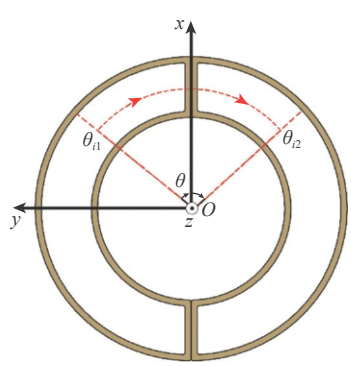


图 10　转子绕组偏转区间

Fig. 10　Deflection interval of the rotor windings

响因素。

将表2相关初始数据代入式(16)和式(20),经数值计算得到旋转磁轴承转子组件中单条闭合线圈平均洛伦兹力大小为3 N,转子输出总转矩为397.3 mN·m。

表2 双磁路旋转洛伦兹力磁轴承转矩参数

Table 2 Double magnetic circuit rotating Lorentz force magnetic bearing torque parameters

单个线圈匝数	通电电流/A	气隙磁密/T	转子旋转角度范围/(°)
100×2	0.65	0.67	120

4 双磁路旋转洛伦兹力磁轴承有限元仿真分析

综合上述研究,开展Maxwell有限元建模仿真,依次对工作气隙轴向、径向和周向磁密情况展开研究分析,通过输出各维度磁密大小并计算分析磁密线性度,验证磁轴承旋转性能。

4.1 Maxwell有限元建模

为保证有限元计算精度,同时输出可靠转矩,采用瞬态建模方式,有利于得到各时刻下磁轴承相应性能数据,同一时间段内不同时刻或时间段作比较可更直观清晰地判断转矩稳定性能。

确定求解区域并创建实体模型,对磁轴承各组件进行材料赋属性,完成结构内各部分差异性网格划分,网格划分直接关系到计算磁密结果及精度,综合考虑零部件实际尺寸、电磁特性、仿真时间等因素,对定子组件采取3.0 mm精度Inside Selection网格剖分,如图11所示;转子工作线圈内外径空心圆面采取0.5 mm精度网格Inside Selection剖分,如图12所示。

4.2 轴、径向磁密特性分析

磁轴承结构中心对称,上下双磁路中气隙磁密大小相等,选取上闭合磁路,基于Maxwell 2D有限元仿真分析其气隙磁密。仿真结果如图13所示,磁通回路各区域未出现磁密过饱和现象。

基于磁轴承上磁路二维平面磁密云图对气隙轴向6 mm间距处磁密进行定量分析。由图14可知,轴向磁密呈左右对称态势,0.5～2 mm和4～5.5 mm左右绕组处最大磁密和最小磁密分别为701.250 mT和681.523 mT,左绕组处磁密均值为696.003 mT,磁密平均值为690.32 mT,磁密均匀度为99.28%,右绕组磁密均值为696.003 mT,磁密平均值为690.43 mT,磁密均匀度为99.29%;2～4 mm线圈骨架处最大磁密和最小磁密分别为690.915 mT和680.023 mT,磁密均值为680.559 mT,磁密均匀度为99.92%,通过

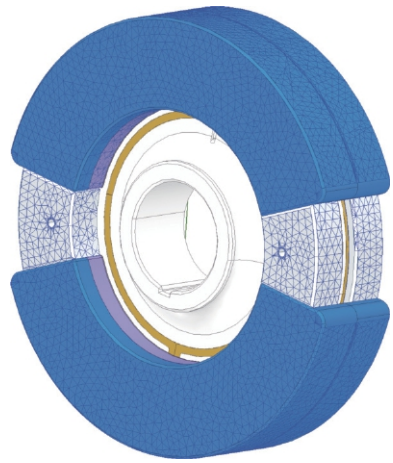


图11 定子网格剖分

Fig. 11 Stator meshing

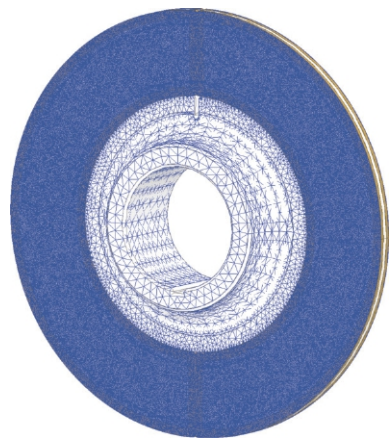


图12 转子细化网格剖分

Fig. 12 Rotor refinement meshing

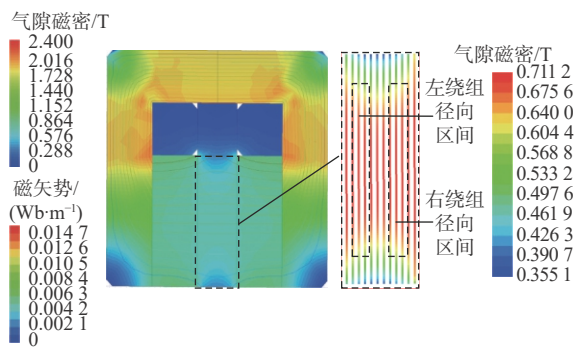


图13 磁轴承2D结构磁密云图

Fig. 13 Magnetic dense cloud diagram of magnetic bearing 2D structure

Maxwell仿真软件输出各处磁密经计算得出轴向磁密均匀度为99.8%。可以得出,气隙轴向磁密均匀性很好,有效保证转子输出稳定旋转力矩。

上磁路径向磁密呈现从两侧向中间聚集,如图15所示,左右绕组径向磁密相比线圈骨架磁密曲线波动更缓,且大小对称一致,保证转子在旋转过程中的稳定性。

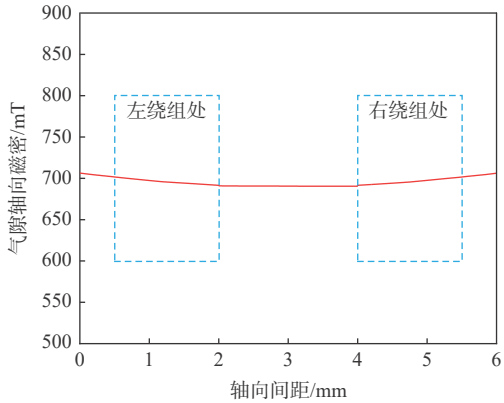


图 14　气隙轴向磁密曲线

Fig. 14　Axial magnetic density curve of the air gap

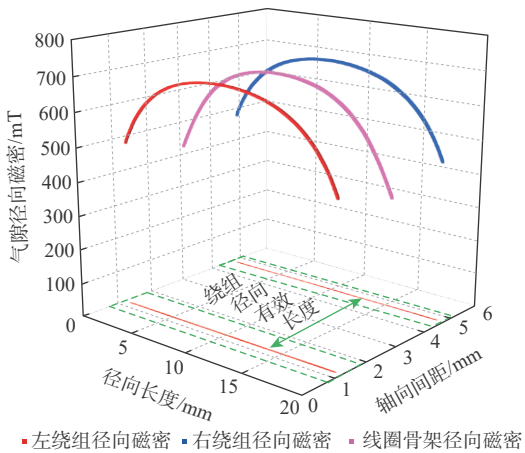


图 15　上磁路径向磁密曲线

Fig. 15　Upper magnetic circuit radial magnetic density curve

磁场气隙径向长度为 20 mm, 磁密分布态势为中间强沿磁极径向上下两端方向降低, 左绕组中心 1.25 mm 处径向最大磁密和最小磁密分别为 435.645 mT 和 692.334 mT, 磁密均值为 660.073 mT, 磁密波动率为 0.568%; 线圈骨架中心 3.00 mm 处径向最大磁密和最小磁密分别为 685.624 mT 和 400.725 mT, 磁密径向均值为 620.158 mT, 磁密波动率为 0.835%; 右绕组中心 4.75 mm 处径向最大磁密和最小磁密分别为 440.274 mT 和 693.657 mT, 磁密均值为 662.024 mT, 磁密波动率为 0.557%。

如图 16 所示, 对上下双磁路气隙径向磁密定量分析可知, 上下闭合磁路中气隙径向磁密曲线整体呈反 “S” 形, 左右绕组径向磁密曲线重合, 上下气隙对应位置处磁密大小相等, 方向相反。

由式(15)可知, 即使单闭合磁路中气隙径向磁密曲线存在波动, 但其转矩为径向端磁密积分效果, 上下磁密对应处相等, 使转子转矩稳定旋转。

定子组件结构采用上下对称磁钢环, 4 条并联线圈作用于上下气隙中, 双磁路中气隙磁密呈现中心对称效果。气隙内同一封闭线圈对应处磁密大小相等, 巧妙消除径向磁密变化所导致的受力不均。径向磁密态势呈现左右绕组处磁密高于骨架位置处, 证明对称式外置挂耳形线圈结构可降低转子旋转工作状态下线圈电流, 从而减小磁轴承功耗。

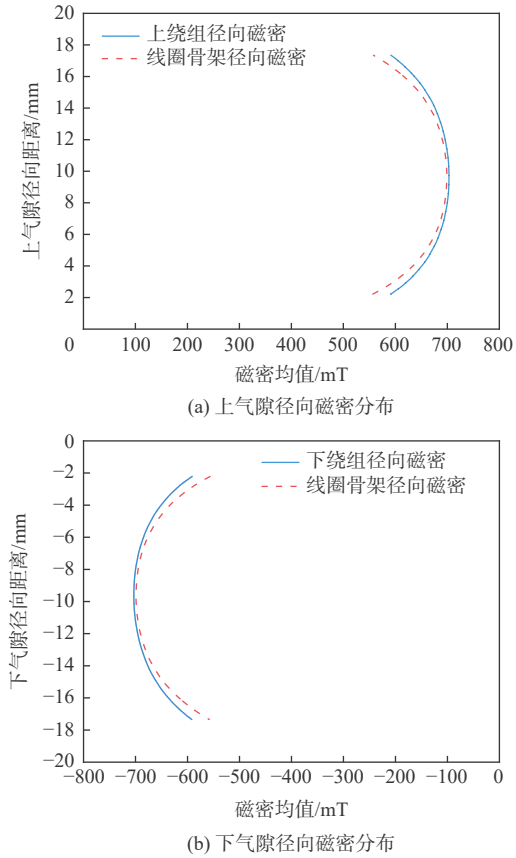


图 16　双气隙径向磁密分布情况

Fig. 16　Radial magnetic density distribution of double air gaps

4.3　周向磁密特性分析

基于 Maxwell 3D 有限元仿真, 磁轴承中转子周向旋转弧段磁密分布如图 17 所示, 选取上磁路气隙磁密, 对扇形磁钢环沿径向由内向外做 3 mm 等间隔 7 条周向圆弧段输出其磁密大小, 并分析其磁密特性。

如图 18 所示, 其中, r_i 为线圈轴心原点位置沿其径向位置的各处距离。定量输出气隙周向弧段区间下磁密大小: 半径 51 mm 弧段内磁密均值约

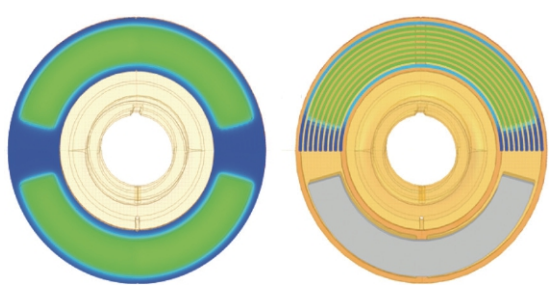


图 17　工作绕组区域周向磁密云图

Fig. 17　Circumferential magnetic dense cloud diagram of the working winding area

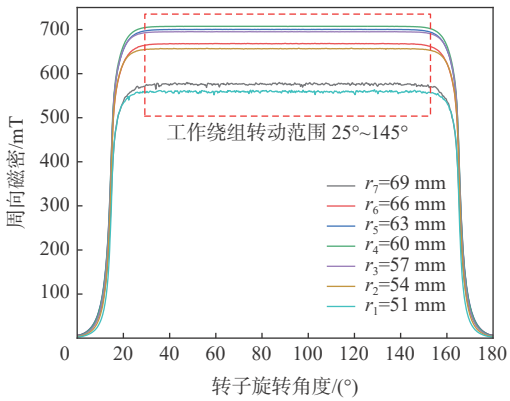


图 18 等间隔周向弧段磁密

Fig. 18 Equivalency circumferential arc segment magnetic density

为0.57 mT;半径54 mm弧段内磁密均值约为0.66 mT,半径57 mm弧段内磁密均值为0.55 mT,半径60 mm和66 mm弧段内磁密均值为0.7 mT,半径69 mm弧段内磁密均值约为0.55 mT。转子旋转扫掠弧段内磁密呈对称态势,初位置磁密变化无震荡起伏,转子旋转扫掠范围下磁密均匀度为99.72%,曲线平滑且弧段内周向磁密波动微乎其微。

4.4 旋转负载能力分析

如图19所示,磁轴承定子内磁钢采用扇形整环状有效避免传统拼接式磁钢存在接缝处磁密振荡现象。

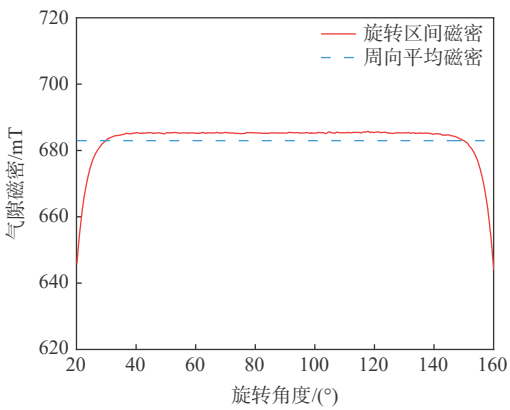


图 19 旋转状态下磁密曲线

Fig. 19 Magnetic density curve in the rotating state

轴向充磁式扇形磁钢的磁密曲线光滑,有利于转子在持续旋转过程中始终保持平衡。同时,转子近1π大角度往复旋转增强了旋转磁轴承的万向机动能力,提高载荷舱敏捷性能。

如图20所示,旋转磁轴承基于Maxwell瞬态条件下150 ms内,转矩随时间推移曲线恒定,大小为382 mN·m。4.3节数学模型结果为397.3 mN·m,两者相差4%,验证数学模型的有效性,有效证明了新型双磁路旋转洛伦兹力磁轴承转矩具有精度高的显著优势,为后期实验加工验证奠定基础。

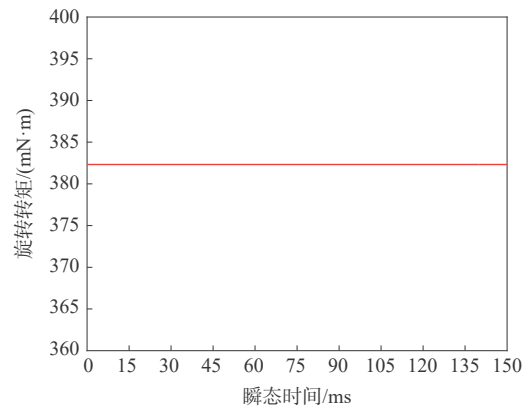


图 20 旋转磁轴承输出工作转矩

Fig. 20 Rotating magnetic bearings output operating torque

5 结 论

旋转洛伦兹力磁轴承作为新型航天器平台姿控系统核心部件,其稳定性和敏捷性影响载荷工作状态。

1)旋转洛伦兹力磁轴承磁钢采用轴向充磁方式,平行布设扇形磁钢环创设高磁密、均匀性气隙磁场;动圈式转子方案显著提升控制精度;定子圆周方向对称布局,大幅提升转子旋转范围,为提升航天器平台姿态敏感精度提供了设计方案和结构支持。

2)依照双磁路旋转洛伦兹力磁轴承上下对称式结构,搭建等效磁路模型并对其简化;定量计算各处磁阻大小,对气隙磁密从均匀度和波动率2方面定义磁密线性度;建立转子动力学模型和电磁力矩模型,明晰转子旋转响应核心要素。

3)在数学建模基础上,基于Maxwell有限元法仿真验证分析气隙磁密,并得到旋转力矩性能。双磁路和4并联线圈方案有效提升气隙中心位置处磁密至683.624 mT;改善气隙磁密均匀性,轴向磁密均匀度为99.8%,径向磁密波动率仅0.568%,周向旋转弧段内磁密均匀度为99.72%,有效保证工作状态下旋转洛伦兹力磁轴承的稳定性。

参考文献(References)

[1] 王曦, 石泳, 樊亚洪, 等. 磁悬浮飞轮用新型异极性永磁偏置径向磁轴承[J]. 宇航学报, 2014, 35(12): 1457-1464.
WANG X, SHI Y, FAN Y H, et al. A new type of heteropolar radial magnetic bearing biased with permanent magnet for magnetically suspended flywheel[J]. Journal of Astronautics, 2014, 35(12): 1457-1464(in Chinese).

[2] 刘彬, 房建成, 刘刚, 等. 磁悬浮飞轮不平衡振动控制方法与试验研究[J]. 机械工程学报, 2010, 46(12): 188-194.
LIU B, FANG J C, LIU G, et al. Unbalance vibration control and experiment research of magnetically suspended flywheels[J]. Journal of Mechanical Engineering, 2010, 46(12): 188-194(in Chinese).

[3] 邓瑞清, 赵岩, 房建成, 等. 磁悬浮飞轮与机械飞轮干扰特性的对比分析[J]. 宇航学报, 2016, 37(8): 917-923.
DENG R Q, ZHAO Y, FANG J C, et al. Disturbance characteristics analysis of magnetically suspended and mechanical flywheels[J]. Journal of Astronautics, 2016, 37(8): 917-923(in Chinese).

[4] 宋玉旺, 韩邦成, 魏学敏, 等. 大中型磁悬浮控制力矩陀螺的框架优化设计方法研究[J]. 载人航天, 2012, 18(4): 1-7.
SONG Y W, HAN B C, WEI X M, et al. Design optimization of the frame for large and middle MSCMG[J]. Manned Spaceflight, 2012, 18(4): 1-7(in Chinese).

[5] 张伟, 赵艳彬, 廖鹤, 等. 动静隔离、主从协同控制双超卫星平台设计[J]. 上海航天, 2014, 31(5): 7-11.
ZHANG W, ZHAO Y B, LIAO H, et al. Design of an active-quiet isolated and master-slave coordination controlled dual-super satellite platform[J]. Aerospace Shanghai, 2014, 31(5): 7-11(in Chinese).

[6] 刘磊. 航天器主动隔振及精确定向控制技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2011.
LIU L. Research on active vibration isolation and precision pointing control for spacecraft[D]. Harbin: Harbin Institute of Technology, 2011(in Chinese).

[7] 陈雄姿, 杨芳, 谢松, 等. 北京三号卫星自主任务规划设计与在轨应用[J]. 空间控制技术与应用, 2022, 48(5): 29-38.
CHEN X Z, YANG F, XIE S, et al. Autonomous mission planning design and onboard application of Beijing-3 satellite[J]. Aerospace Control and Application, 2022, 48(5): 29-38(in Chinese).

[8] 杨芳, 赵键, 姚宁, 等. 北京三号 A/B 卫星总体设计及技术创新[J]. 航天器工程, 2023, 32(3): 7-15.
YANG F, ZHAO J, YAO N, et al. System design and technical innovation of BJ-3A/B satellites[J]. Spacecraft Engineering, 2023, 32(3): 7-15(in Chinese).

[9] 刘强, 赵明师, 韩邦成, 等. 磁悬浮框架飞轮磁轴承技术研究与发展现状[J]. 宇航学报, 2019, 40(11): 1251-1261.
LIU Q, ZHAO M S, HAN B C, et al. Research and development status of magnetic bearing technology on magnetically suspended gimballing flywheel[J]. Journal of Astronautics, 2019, 40(11): 1251-1261(in Chinese).

[10] LAHDO M, STROEHLA T, KOVALEV S. A novel low-power dual-actuator for high precision magnetic levitation systems[C]// Proceedings of the International Conference on New Actuators. Bremen: MESSE, 2016.

[11] BORQUE GALLEGO G, ROSSINI L, ONILLON E, et al. On-line micro-vibration measurement method for Lorentz-type magnetic-bearing space actuators[J]. Mechatronics, 2019, 64: 102283.

[12] ZHU H Y, PANG C K, LAW T L, et al. Design and control of a six degrees-of-freedom magnetically levitated positioning system[J]. IFAC-PapersOnLine, 2016, 49(21): 127-132.

[13] PETIT N, SARRAS I. A generalized control law for uniform, global and exponential magnetic detumbling of rigid spacecraft[J]. IFAC-PapersOnLine, 2020, 53(2): 14876-14881.

[14] LIU Q, FANG J C, HAN B C. Novel electromagnetic repeated launch locking/unlocking device (RLLUD) based on self-locking for magnetic bearing flywheel[J]. Sensors and Actuators A: Physical, 2012, 175: 116-126.

[15] 辛朝军, 蔡远文, 任元, 等. 磁悬浮敏感陀螺动力学建模与关键误差源分析[J]. 北京航空航天大学学报, 2016, 42(10): 2048-2058.
XIN C J, CAI Y W, REN Y, et al. Dynamic modeling and key error sources analysis of magnetically suspended sensitive gyroscopes[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(10): 2048-2058(in Chinese).

[16] 刘强, 房建成, 武登云. 磁悬浮飞轮转子优化设计与实验[J]. 机械设计与制造, 2015(2): 111-114.
LIU Q, FANG J C, WU D Y. Optimization and experiment of magnetic bearing wheel rotor[J]. Machinery Design & Manufacture, 2015(2): 111-114(in Chinese).

[17] 刘虎, 房建成, 刘刚. 基于磁悬浮动量轮微框架能力的卫星滚动-偏航姿态稳定控制研究[J]. 宇航学报, 2010, 31(4): 1063-1069.
LIU H, FANG J C, LIU G. Satellite roll-yaw axis attitude stable control based on magnetically suspended momentum wheel with vernier gimballing capacity[J]. Journal of Astronautics, 2010, 31(4): 1063-1069(in Chinese).

[18] 林竹翀. 磁悬浮惯性稳定平台控制技术研究[D]. 长沙: 国防科学技术大学, 2015.
LIN Z C. Research on the control system of maglev dual-stage intertially stabilized platform[D]. Changsha: National University of Defense Technology, 2015(in Chinese).

[19] 刘强, 赵勇, 曹建树, 等. 新型微框架磁悬浮飞轮用洛伦兹力磁轴承[J]. 宇航学报, 2017, 38(5): 481-489.
LIU Q, ZHAO Y, CAO J S, et al. Lorentz magnetic bearing for novel vernier gimballing magnetically suspended flywheel[J]. Journal of Astronautics, 2017, 38(5): 481-489(in Chinese).

[20] 张钢, 孟庆涛, 钟永彦, 等. 五自由度全永磁轴承系统的稳定悬浮特性分析[J]. 机械工程学报, 2015, 51(5): 56-63.
ZHANG G, MENG Q T, ZHONG Y Y, et al. Stable levitation performance analysis of five degrees of freedom all permanent magnetic bearing system[J]. Journal of Mechanical Engineering, 2015, 51(5): 56-63(in Chinese).

[21] 李丽君, 樊亚洪, 袁军. 磁悬浮万向飞轮在卫星姿态机动中的应用[J]. 机械工程学报, 2015, 51(16): 206-212.
LI L J, FAN Y H, YUAN J. Application of magnetically suspended gimbaling flywheel in satellite attitude maneuver[J]. Journal of Mechanical Engineering, 2015, 51(16): 206-212(in Chinese).

[22] 王卫杰, 任元, 刘强, 等. 球面磁悬浮万向飞轮转子轮盘优化设计[J]. 航空学报, 2016, 37(9): 2874-2883.
WANG W J, REN Y, LIU Q, et al. Optimal design of rotary table for spherical rotor of magnetically suspending gambling flywheel [J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(9): 2874-2883(in Chinese).

[23] 夏长峰, 王小乐, 李智, 等. 磁悬浮控制敏感陀螺转子系统在线动平衡方法[J]. 北京航空航天大学学报, 2024, 50(11): 3417-3425.
XIA C F, WANG X L, LI Z, et al. Field balancing method for rotor system of magnetically suspended control and sensing gyro[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(11): 3417-3425(in Chinese).

[24] 王卫杰, 庞维坤, 任元, 等. 一种洛伦兹力磁悬浮万向稳定平台: CN115789089A[P]. 2023-03-14.
WANG W J, PANG W K, REN Y, et al. Lorentz force magnetic suspension universal stable platform: CN115789089A[P]. 2023-03-14(in Chinese).

[25] 许国锋, 蔡远文, 任元, 等. 洛伦兹力磁轴承磁密均匀度设计与分析[J]. 北京航空航天大学学报, 2016, 43(3): 559-566.
XU G F, CAI Y W, REN Y, et al. Design and analysis on uniformity of magnetic flux density in Lorentz force-type magnetic bearing [J]. Beijing Hangkong Hangtian Daxue Xuebao/Journal of Beijing University of Aeronautics and Astronautics, 2016, 43(3): 559-566.

[26] 胡业发, 周祖德, 江征风. 磁力轴承的基础理论与应用[M]. 北京: 机械工业出版社, 2006: 98-102.
HU Y F, ZHOU Z D, JIANG Z F. Fundamental theory and application of maghetic bearings[M]. Beijing: China Machine Press, 2006:

98-102(in Chinese).

[27] 熊颖, 刘强, 任元, 等. 基于 KF-LESO-PID 洛伦兹惯性稳定平台控制[J]. 北京航空航天大学学报, 2022, 48(6): 1072-1081.
XIONG Y, LIU Q, REN Y, et al. Lorentz inertial stability platform control based on KF-LESO-PID[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(6): 1072-1081(in Chinese).

[28] 高鹏飞, 房建成, 韩邦成, 等. 高速永磁电机的转子涡流损耗分析[J]. 微电机, 2013, 46(5): 5-11.
GAO P F, FANG J C, HAN B C, et al. Analysis of rotor eddy-current loss in high-speed permanent magnet motors[J]. Micromotors, 2013, 46(5): 5-11(in Chinese).

[29] 孙津济. 磁悬浮飞轮用新型永磁偏置主动磁轴承结构与设计方法研究[D]. 北京: 北京航空航天大学, 2010: 133-134.
SUN J J. Study on new permanent magnet biased active magnetic bearing structure and design method in magnetically suspended flywheel application[D]. Beijing: Beihang University, 2010: 133-134(in Chinese).

[30] 孙津济, 房建成, 王曦, 等. 一种永磁偏置外转子径向混合磁轴承的设计方法: CN200610114268.3[P]. 2023-12-25.

[31] 王平, 王华, 任元. 基于磁悬浮控制力矩陀螺的航天器姿态角速率测量方法[J]. 兵工学报, 2015, 36(10): 1907-1915.
WANG P, WANG H, REN Y. Measurement method of spacecraft attitude angular rate based on magnetically suspended control moment gyroscope[J]. Acta Armamentarii, 2015, 36(10): 1907-1915(in Chinese).

[32] WANG W J, REN Y, LIU Q, et al. Optimal design of rotary table for spherical rotor of magnetically suspending gambling flywheel [J]. Acta Aeronauticaet Astronautica Sinica, 2016, 37(9): 2874-2883.

[33] 孙津济, 房建成, 王曦, 等. 一种新型结构的永磁偏置径向磁轴承[J]. 电工技术学报, 2009, 24(11): 53-60.
SUN J J, FANG J C, WANG X, et al. A new permanent magnet biased radial magnetic bearing[J]. Transactions of China Electrotechnical Society, 2009, 24(11): 53-60(in Chinese).

Design and analysis of double magnetic circuit rotating Lorentz force magnetic bearing

PANG Weikun[1], WANG Weijie[2,*], FAN Yahong[3], LI Lei[1], YANG Yang[4], ZHU Hongye[1]

(1. Graduate School, Space Engineering University, Beijing 101416, China;
2. Department of Astronautical Science and Technology, Space Engineering University, Beijing 101416, China;
3. Beijing Institute of Control Engineering, Beijing 100190, China;
4. College of Artificial Intelligence, Henan University, Zhengzhou 475001, China)

Abstract: A Lorentz force magnetic levitation universal stability platform was proposed in response to the pressing need for universal agility and ultra-precise pointing performance of spacecraft payload in complex space missions. The design and analysis of double-magnetic circuit rotating Lorentz force magnetic bearings were conducted, and the moving coil rotor scheme was chosen. The four-hanging lug-shaped coils of the rotor assembly were wound in pairs and glued into the grooves on both sides of the skeleton axial direction. The stator components are arranged in pairs in parallel with a common rotating shaft and a double annular axial magnetized magnetizer to create a uniform and stable magnetic density, and provide a circumferential dual-channel symmetrical working air gap for the agile maneuver of the load compartment. Based on the equivalent magnetic circuit method, the magnetic density model of the air gap was established, and the linearity of the magnetic density was defined from the two aspects of the uniformity and fluctuation rate of the magnetic density of the air gap, and then the rotor rotation dynamics was modeled and the rotational moment model was constructed. Utilize Maxwell's finite element method to establish a finite element model for rotary magnetic bearings and conduct the simulation, the results demonstrate that the magnetic density at the center position of the air gap rotation envelope of the rotating magnetic bearing scheme can reach 685.624 mT, and the circumferential magnetic density uniformity is 99.72%. This significantly improves the uniformity of air gap magnetic density, avoids the limitations of radial magnetization schemes such as air gap magnetic density attenuation and longitudinal magnetic field diffusion, and effectively enhances the stability and pointing accuracy of the payload bay under rotating conditions.

Keywords: axial magnetization; sector-shaped magnetic rings; moving coil rotor; magnetic density linearity; equivalent magnetic circuit method; finite element analysis

Received: 2023-11-22; **Accepted**: 2024-01-02; **Published Online**: 2024-01-15 21:14

URL: link.cnki.net/urlid/11.2625.V.20240115.1340.004

Foundation item: National Natural Science Foundation of China (52075545)

* **Corresponding author**. E-mail: wangwjie@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0770

基于 OTFS-LFM 的水下可见光通信感知一体化波形设计

杨新，朵琳*，任勇，许渤雨
(昆明理工大学 信息工程与自动化学院，昆明 650500)

摘　　要：随着对水下通信需求的不断增加，水下可见光通信 (UVLC) 系统备受关注。通信感知一体化的概念在 UVLC 系统中的实现成为了一个新的研究热点。正交时频空 (OTFS) 技术以其在高多普勒和高时延信道中的卓越性能受到学术界广泛关注，为系统提供了强大的通信支持。与此同时，线性调频 (LFM) 技术由于其对多普勒频移的低敏感性而在无线通信领域广泛应用。将 OTFS 技术与 LFM 技术相融合，设计了一种 UVLC 感知一体化系统。通过实验仿真对比发现，系统在误码率 (BER)、模糊函数及目标速度和距离信息获取方面表现出色。融合 OTFS 技术和 LFM 技术的系统在复杂水下环境中展现出卓越的适应性，为水下通信领域带来新的可能性。同时，考虑 3 种常见的调制方式，进一步分析不同调制对系统性能的可能影响，为系统优化提供有益的参考。

关 键 词：水下可见光通信；通信感知一体化；正交时频空；线性调频；模糊函数
中图分类号：V221+.3; TB553
文献标志码：A　　**文章编号**：1001-5965(2026)01-0317-09

在当前数字化时代，深入了解和有效监测海洋环境变得愈发迫切。水下通信技术作为连接水下设备、实现实时数据传输的关键手段备受学术界和工业界关注。随着水下传感器网络和海洋资源开发的不断发展，传统水下通信感知技术面临诸多挑战，包括通信时延较高、带宽受限及感知目标信息获取精度低等问题。因此，寻找一种在水下环境中具有低时延、高带宽且保密性良好的通信方式显得至关重要。近年来，水下可见光通信[1-2](underwater visible light communication, UVLC) 作为一种新兴的水下通信系统引起了广泛关注。UVLC 利用可见光在水中传播的特性，通过调制可见光信号来实现高效的水下通信与感知。相较于传统水声通信和无线电频段通信，UVLC 具有通信时延低、宽高、保密性好等诸多优势[3]，为水下信息传输领域带来了全新的可能性。然而，水下通信中不可避免的问题也逐渐显露，例如，光在水下传播的衰减、通信方向性的限制及移动性的干扰等都在困扰着研究人员[4]。这些挑战需要深入研究和创新，以进一步推动水下通信技术的发展，满足对海洋环境深度感知和监测的需求。

正交频分复用 (orthogonal frequency division multiplexing, OFDM) 通过将快速的数据流分割成多个慢速子载波，借助正交传输减少子载波之间的干扰，有效地克服多径效应引发的频率选择性衰落。然而，OFDM 在面对高速移动目标的通信能力方面的效率比较低，导致在实际应用中出现捕捉误差。而且，有研究发现 OFDM 在时变信道中可能会导致载波间干扰[5]，破坏载波间的原有正交性，因此，UVLC 系统面临的一个新的挑战是在高多普勒、高时延与高速移动目标的通信和感知。

正交时频空 (orthogonal time-frequency space,

收稿日期：2023-11-30；**录用日期**：2024-04-26；**网络出版时间**：2024-05-31 10:40
网络出版地址：link.cnki.net/urlid/11.2625.V.20240530.1641.002
基金项目：国家自然科学基金 (61962032)；云南省科技厅重大科技专项计划 (202302AD080006)
* **通信作者**. E-mail：duolin2003@126.com

引用格式：*杨新，朵琳，任勇，等. 基于 OTFS-LFM 的水下可见光通信感知一体化波形设计 [J]. 北京航空航天大学学报，2026，52（1）：317-325. YANG X，DUO L，REN Y，et al. Underwater visible light communication sensing integrated waveform design based on OTFS-LFM [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：317-325（in Chinese）.*

OTFS)技术[6-7]是一种在时延-多普勒域(delay-Doppler, DD)中进行调制的新型技术。与传统的OFDM调制不同,OTFS在DD域展现出较强的时延容限和多普勒容限,为可靠通信提供了关键支持。在UVLC系统中,OTFS技术的应用有望填补OFDM技术在复杂快时变水下信道条件下性能下降的缺陷,为提高系统性能提供了新的可能性。目前,OTFS的应用主要集中在无线通信领域,而在UVLC系统研究领域,OTFS技术仍处于初始阶段,当前仍以OFDM为主导技术。这表明在UVLC系统中,引入OTFS技术可能为系统性能的提升开辟新的途径。

线性调频[8](linear frequency modulation, LFM)作为一种扩频通信技术,在无线通信领域得到广泛应用,其特点包括对多普勒频移的低敏感性和强大的抗干扰性[9]。在以往的研究中,已有将LFM与OFDM结合构建通信感知一体化系统[10],取得了显著的成果。受此启发,本文将探讨OTFS与LFM的结合,并基于3种常见不同调制技术(BPSK[11]、MSK[12]、16QAM[13])在UVLC场景下的性能进行分析。同时,在3种调制下进行基于UVLC的通感一体化波形设计。这一研究旨在深化在UVLC系统中对比分析OTFS-LFM在不同通信调制技术下的性能,考虑在这一场景下的通感性能折衷。

本文的主要创新点有3个方面:①通过深入比较OTFS与OFDM在通信和感知性能上的异同,以不同的目标速度进行误码率(bit error rate, BER)计算,以更全面地比较其通信性能。同时,采用多种算法对感知性能进行详尽的对比分析。② 将OTFS与LFM结合,通过考察BER和模糊函数,全面评估不同调制方式的性能差异。该方法为深入理解通信系统的性能提供了新的视角,为选择最合适的调制方式提供了有力支持。③将OTFS-LFM技术与UVLC系统相融合,侧重研究在水下光通信信道中的通感折衷性能。通过系统而深入的实验和分析,有助于更全面地了解OTFS-LFM技术在具体环境下的应用潜力,尤其是在水下通信方面。本文研究可能为未来光通信系统在水下环境中的设计和优化提供新的思路和想法。

1 系统原理与模型

本节介绍UVLC系统的基本原理,并着重介绍OTFS调制原理。同时介绍LFM技术,进一步探讨OTFS-LFM信号模型的相关特性。

1.1 UVLC系统

UVLC系统的构成与传统无线射频通信系统相似,包括发射端、信道和接收端,如图1所示。UVLC系统中,光发射端由光发射二极管[14](light emitting diode, LED)充当,其作用类似于传统无线射频通信系统中的发射天线。LED通过辐射光线将电信号转换为光信号,实现信息的传输。这种光发射端的优点在于其结构简单、体积小、重量轻、能耗低,同时,具有高可靠性、高稳定性以及长寿命等优点。光电检测器[15](photoelectric detector, PD)在UVLC系统中扮演与传统无线射频通信系统中接收天线相似的角色。PD通过捕捉光线并将其转换为电信号,实现信息的接收和检测。同时,PD还具有快速响应、高灵敏度、低噪声和强抗干扰能力等优势。通过LED和PD的组合运用,UVLC系统能够实现信息传输的高速、高可靠性和高稳定性。与传统无线射频通信系统相比,采用光信号作为传输介质有效避免了信号衰减、干扰等问题。

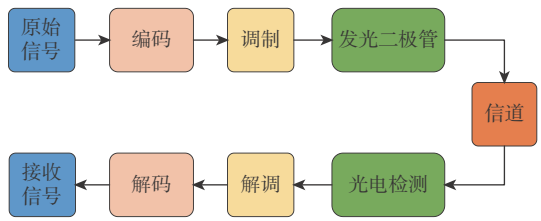


图1 水下可见光通信系统

Fig. 1 Underwater visible light communication system

1.2 OTFS调制原理

OTFS调制是一种利用发射器和接收器上的2个二维变换级联结构的调制方案[16]。在发送端,将DD域中的信息符号经过逆辛有限傅里叶变换(inverse Symplectic finite Fourier transform, ISFFT)映射到时间-频率域(time-frequency, TF):

$$\boldsymbol{X}[n,m]=\frac{1}{\sqrt{MN}}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1}\boldsymbol{x}[k,l]\mathrm{e}^{\mathrm{j}2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)} \tag{1}$$

式中:$\boldsymbol{X}$为时间-频率信号;$\boldsymbol{x}$为发送端发送的多普勒-时延信号;n和m分别为时间轴和频率轴的刻度;N和M分别为多普勒和时延的数量;k和l分别为在多普勒轴和时延轴上的索引。随后,通过海森堡变换和发射脉冲将TF域的信号转换为时域信号$\boldsymbol{s}(t)$,从而在信道中进行传输:

$$\boldsymbol{s}(t)=\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\boldsymbol{X}[n,m]\boldsymbol{g}_{\mathrm{tx}}(t-nT)\mathrm{e}^{\mathrm{j}2\pi m\Delta f(t-nT)} \tag{2}$$

式中:$\boldsymbol{g}_{\mathrm{tx}}(t)$为持续时间为$MT$的发射周期原型脉冲形状;$t$为时域信号的时间刻度;$T$为一个时隙长度;$\Delta f$为子载波间隔。在接收器端,对接收到的数据执行维格纳变换,将时域信号转换为TF域。随后,辛有限傅里叶变换(Symplectic finite Fourier transform, SFFT)将TF域中的信号映射回DD域以进行解调:

$$\boldsymbol{y}[k,l]=\frac{1}{\sqrt{MN}}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1}\boldsymbol{Y}[n,m]\mathrm{e}^{-\mathrm{j}2\pi\left(\frac{nk}{N}-\frac{ml}{M}\right)} \tag{3}$$

式中: $\boldsymbol{y}$ 表示通过转换后到接收端的接收到多普勒-时延信号; $\boldsymbol{Y}$ 表示从信道接收到的 $\boldsymbol{r}(t)$ 时域信号通过海森堡变换转换过来的时间-频率信号。

调制方案如图 2 所示, 通过 ISFFT 和 SFFT 这 2 个关键变换, 实现了 DD 域到 TF 域的映射以及从 TF 域到 DD 域的逆映射。通过仔细观察, 发现通过替换海森堡变换和维格纳变换为逆快速傅里叶变换 (inverse finite Fourier transform, IFFT) 和快速傅里叶变换 (finite Fourier transform, FFT), 可以将内侧虚线框转化为一个 OFDM 系统。在图 2 的下方清晰展示了该转变, 其中, $\boldsymbol{r}(t)$ 表示接收到的时域信号。因此, OTFS 可以被视为在 OFDM 调制系统的发送端添加 ISFFT 模块和在接收端添加 SFFT 模块的一种调制系统。

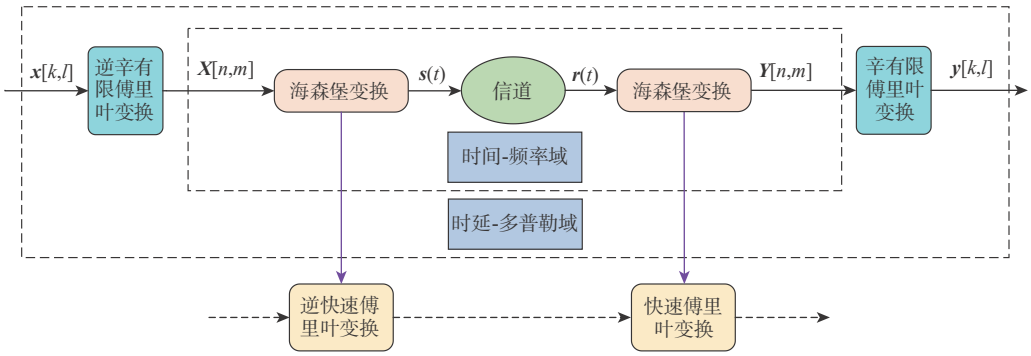


图 2　OTFS 和 OFDM 调制框架

Fig. 2　OTFS and OFDM modulation framework

1.3　OTFS-LFM 信号模型

在通信系统中, 多径传播是指信号在传播过程中经历多个路径到达接收端的现象。这些不同路径导致接收端观察到的信号存在多个版本, 每个版本具有不同的传播时延、相位和幅度。多径传播可能是由于信号在空间中反射、折射、散射等效应引起的。LFM 信号在雷达系统和通信系统中得到广泛应用, 其主要特点在于对多径传播具有较好的抗干扰性。而在雷达系统中, LFM 信号由于具备良好的自相干性质, 有助于区分主要路径和反射路径, 从而降低多径传播引起的干扰。这在提高距离分辨率和改善抗干扰性能方面具有显著优势。LFM 信号的数学模型可表示如下:

$$\boldsymbol{m}(t)=\boldsymbol{u}(t)\mathrm{e}^{\mathrm{j}2\pi\left(f_0t+\mu t^2/2\right)} \tag{4}$$

式中: $\boldsymbol{u}(t)$为矩形窗函数; f_0为 LFM 信号的中心频率; μ为调频斜率。在发射端将信息符号 (如 PSK 符号、QAM 符号) 与 LFM 技术结合后, 将其视为二维的 DD 域信号, 通过 ISFFT 变换转化为 TF 域信号, 并通过海森堡变换将时频域数据转化为能够在时域中传播的连续波形信号, 即实现了 OTFS-LFM 信号的生成:

$$\boldsymbol{z}(t)=\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\boldsymbol{X}[n,m]\boldsymbol{g}_{\mathrm{tx}}(t)\boldsymbol{u}(t-T)\cdot \exp(\mathrm{j}2\pi[m\Delta f(t-nT)+(f_0t+\mu t^2/2)]) \tag{5}$$

通过引入 LFM 技术, OTFS-LFM 信号模型在时域和频域上实现了更为灵活的调制, 从而提升了信号的抗干扰性和传输稳定性。这种调制技术的特性使其在水下环境中更容易适应不同的传播条件, 为实现可靠的水下通信提供了有力支持。为深入研究 UVLC 信号的调制过程, 仿真实验选择通信系统中常见的 3 种调制方式, 即 16QAM、MSK 和 BPSK, 进行进一步分析和研究。下文将展开仿真实验, 以验证所提出调制方式在水下通信环境中的不同性能表现。

2　仿真实验

仿真实验采用 MATLAB R2023a 平台对 OTFS-LFM 一体化波形进行对比分析。首先, 对 OFDM 与 OTFS 调制系统在通信和感知 2 个方面的性能进行详细比较分析。随后, 借鉴 OFDM-LFM 技术, 将 LFM 技术嵌入 OTFS 系统模型, 并对其通感性能展开深入研究。最后, 使用 Gamma-Gamma 分布湍流模型作为 UVLC 的模拟信道模型, 再分别利用 MSK、BPSK、16QAM 3 种不同信号调制方式, 对 3 种一体化波形通感性能进行全面比较分析。对比实验的主要目标是确定在 UVLC 环境中, 采用何种调制方式能够在性能上实现最佳折衷。

2.1　OTFS 与 OFDM 性能对比

本节对 OTFS 和 OFDM 在不同速度下的 BER 进行比较, 以突显 OTFS 的性能优势。同时, 在相同的参考速度下评估其在感知移动目标速度和距离方面的性能差异。仿真参数详见表 1 所示。

表 1 通信系统仿真参数

Table 1 Simulation parameters of communication system

载波频率/GHz	子载波数量	子载波符号数	子载波间隔/kHz	系统带宽/MHz	调制方式	信道估计
1	128	14	15	1	16QAM	理想

在评估通信系统性能时，通常使用 BER 作为重要的参考指标。因此，在一体化设计中对比通信性能时，使用 BER 作为主要比较依据。为了防止因为调制方式不同，导致可能出现的误差。在仿真设定中统一使用 16QAM 作为原始发送信号的调制方式。通过图 3 可以明显看出，OFDM 和 OTFS 的 BER 均随信噪比 (signal noise ratio, SNR) 的升高而降低，且 OTFS 在检测目标速度不同的情况下，其 BER 均比 OFDM 要低。同时，随感知目标移动速度加快，BER 逐渐上升，说明在高速移动状态下，通信性能受较大的影响。

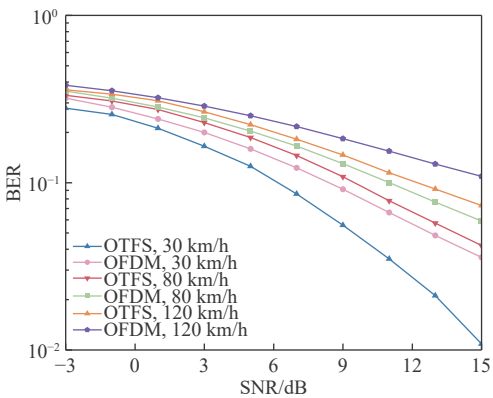


图 3 不同速度下的 OFDM 与 OTFS 误码率对比

Fig. 3 BER comparison between OFDM and OTFS at different speeds

为了比较 OTFS 与 OFDM 的感知性能，仿真实验使用基于循环互相关 (cyclic cross correlation, CCC) 算法分别计算 OTFS 和 OFDM 的模糊函数。CCC 算法通常用于计算 2 个信号之间的相关性，本文将用来估计距离与速度信息。

雷达模糊函数通常用于描述雷达系统中的性能和分辨率[17]。雷达系统通过发射脉冲并接收回波来探测目标，模糊函数可以更好地帮助理解雷达系统在探测和跟踪目标时的精度和灵敏度。为了在仿真计算中防止设定的数值过大，可能造成仿真时间较长，设定移动目标的速度及其与监测点间的距离分别为 30 m/s 和 20 m。图 4 和图 5 展示了 2 种系统的雷达模糊函数，其具体感知数据也在图中呈现。从感知性能来看，两者均能较好地获取移动目标的预设数据。

为了多方面地呈现 OTFS 相较于 OFDM 的不同，使用多重信号分类[18](multiple signal classification, MUSIC) 算法和两阶段感知估计[19] (two-phase sensing estimation, TPSE) 算法分别计算 OFDM 和 OTFS 的感知性能。

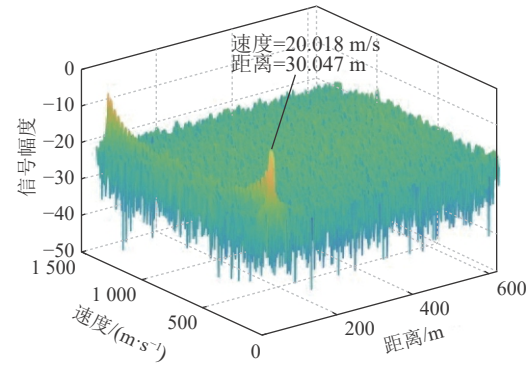


图 4 CCC 算法计算 OFDM 雷达模糊函数

Fig. 4 CCC algorithm calculation of OFDM radar ambiguity function

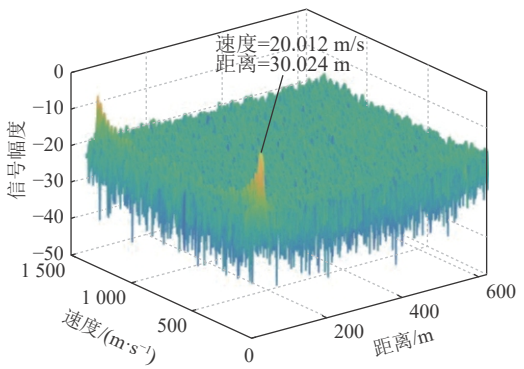


图 5 CCC 算法计算 OTFS 雷达模糊函数

Fig. 5 CCC algorithm calculation of OTFS radar ambiguity function

MUSIC 算法是一种在信号处理中用于方向估计的技术。该算法包括以下关键步骤：

步骤 1 需要从传感器阵列收集数据。接收到的信号通常会进行采样，并表示为数据矩阵。

步骤 2 从接收到的数据矩阵计算协方差矩阵。该矩阵捕获了不同传感器接收到的信号之间的统计关系。

步骤 3 对协方差矩阵进行特征值分解，以获得其特征值和相应的特征向量。这些特征向量构成了信号和噪声子空间的基础。

步骤 4 根据特征值和特征向量，构建信号和噪声子空间。信号子空间通常由对应于最大特征值的特征向量组成，而噪声子空间由剩余的特征向量组成。

步骤 5 利用构建的信号子空间，计算不同角度或频率的 MUSIC 谱。

步骤 6 MUSIC 谱中的峰值对应于信号到达方向的潜在位置。通过识别峰值，算法可以估计信

号源的具体信息。

TPSE 算法是一种用于估计稀疏信号的算法，通常用于压缩感知中，本文使用 TPSE 算法用于 OTFS 感知性能估计。算法主要分为感知阶段和估计阶段，感知阶段中需要对原始信号压缩采样，转换为低维观测向量，获取经过压缩采样后的观测向量；估计阶段利用感知阶段获取的量化测量数据并重构信号，通过已经重构的信号，估计原始信号中带有的距离和速度信息。

在使用 TPSE 算法计算 OTFS 感知性能时，使用 30 次迭代计算出其速度与距离的感知性能。从图 6 和图 7 明显观察到，使用 MUSIC 算法的 OFDM 和使用 TPSE 算法的 OTFS 在获取移动目标信息方面均能准确获取数据。与使用 CCC 算法计算雷达模糊函数得到的数据基本一致，与原始数据相比误差较小。这进一步说明，OFDM 和 OTFS 在雷达感知方面均表现出色。

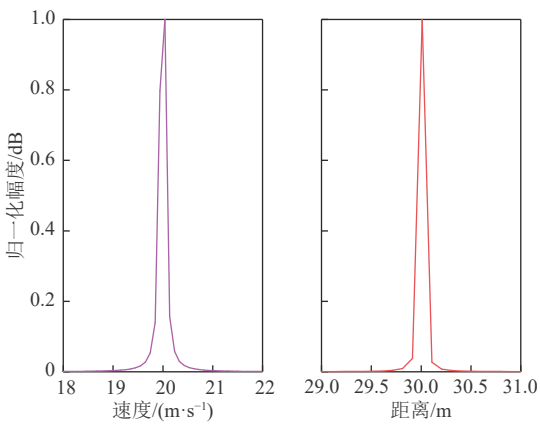


图 6　MUSIC 算法估计 OFDM 感知性能

Fig. 6　Estimating OFDM sensing performance using the MUSIC algorithm

通过对 BER、模糊函数和感知估计等方面进行比较研究，可以得到以下结论：在提高通信性能的同时，OTFS 表现出与 OFDM 雷达感知性能一致的特性。此外，实验发现在与高速移动目标通信的场景下，相较于 OFDM 调制，OTFS 调制表现出更低的 BER，从而展现出更为优越的通信性能。

2.2　OFDM 与 OTFS 在水下可见光信道中的对比

本节旨在比较在 UVLC 信道中采用 OFDM、OTFS 和 OTFS-LFM 这 3 种方式的通信与感知性能。为了减少误差，实验统一采用 BPSK 调制发射信号，并使用 TPSE 算法来估计距离和速度信息。BER 比较结果如图 8 所示，结果表明，OTFS 和 OTFS-LFM 的通信性能均优于 OFDM，且 OTFS 与 OTFS-LFM 之间的差距不大。感知性能的对比进行了 1 000 次实验，并取均值，如表 2 所示。图 9～图 11 展示了较好的感知性能表现，以展示仿真效果。

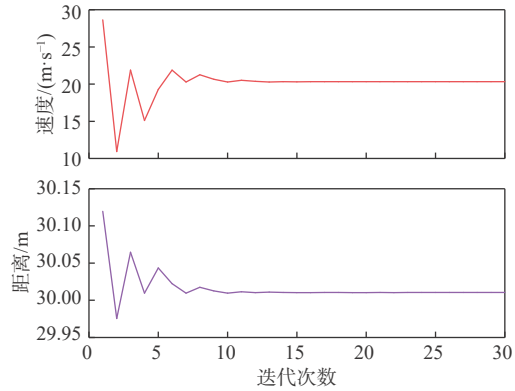


图 7　TPSE 算法估计 OTFS 感知性能

Fig. 7　Estimating OTFS sensing performance using the TPSE algorithm

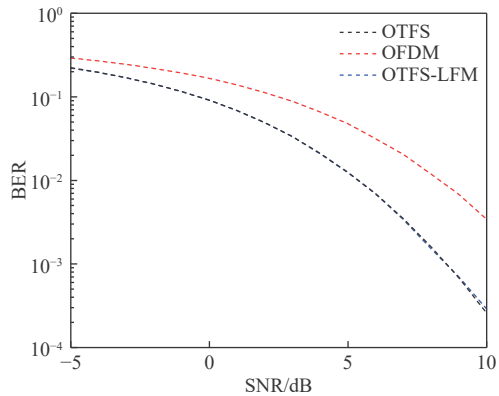


图 8　3 种方式的误码率对比

Fig. 8　The BER comparison of the three methods

表 2　3 种系统下的感知性能

Table 2　Sensing performance under three systems

系统	速度/(m·s⁻¹)	距离/m
OFDM	19.32	30.14
OTFS	20.65	29.88
OTFS-LFM	20.11	30.02

从表 2 可以观察到，OTFS-LFM 在性能上略胜于 OTFS 和 OFDM。结合图 8，可以得出相同的结论。因此，采用 OTFS-LFM，通信和感知 2 方面的性能均能达到最优水平。

2.3　不同调制技术与 OTFS-LFM 的结合

本节使用 MSK、BPSK、16QAM 这 3 种信号调制方式与 OTFS-LFM 的一体化设计，经过对通信和感知 2 方面的分析，对比不同调制方式的影响，为后续仿真实验提供重要的参考依据。表 3 给出了 16QAM-LFM、MSK-LFM、BPSK-LFM 这 3 种一体化波形的通信 BER，基于这些数据绘制相应的 BER 曲线，如图 12 所示。可以清晰地发现，MSK-LFM 和 BPSK-LFM 这 2 种波形在通信性能方面明显优于 16QAM-LFM 波形。值得注意的是，MSK-

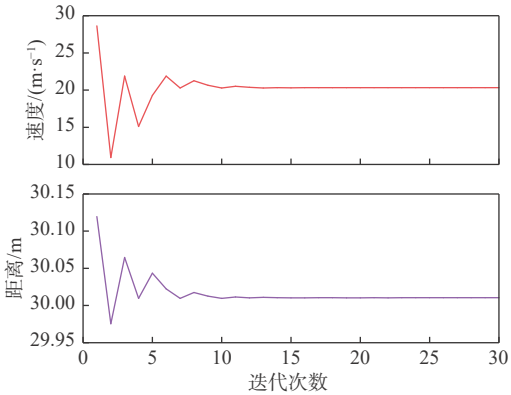


图 9 OFDM 感知性能

Fig. 9 OFDM sensing performance

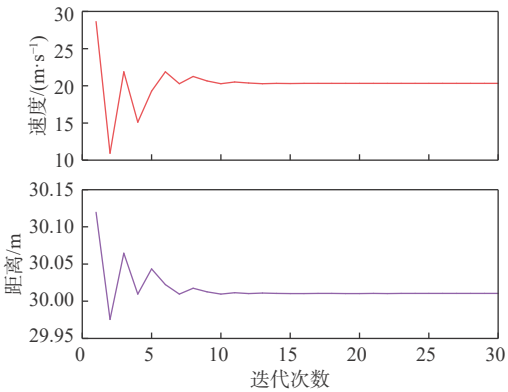


图 10 OTFS 感知性能

Fig. 10 OTFS sensing performance

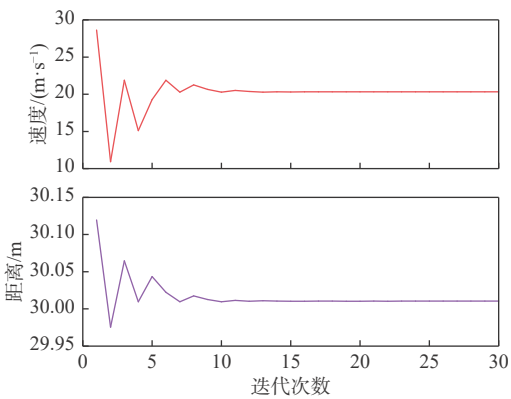


图 11 OTFS-LFM 感知性能

Fig. 11 OTFS-LFM sensing performance

LFM 和 BPSK-LFM 的 BER 曲线几乎重合，与表 3 中计算的 BER 数值一致。该一致性强调了 MSK-LFM 和 BPSK-LFM 在通信性能上的相似性。这些发现为深入理解不同调制方式在 OTFS-LFM 系统中的性能差异提供了实质性的信息。

为了比较感知性能，图 13～图 15 呈现了 3 种不同调制方式生成的雷达模糊函数。通过比较，明显观察到 16QAM 调制在抑制旁瓣方面表现卓越，使整体信号在多普勒和时延分辨率方面具有更出

表 3 不同一体化波形 BER 与接收 SNR 对比

Table 3 Comparison of BER and received SNR of different integrated waveforms

调制方式	BER
BPSK-LFM	$\frac{1}{2}\mathrm{erfc}(\sqrt{S_{\mathrm{NR}}})$
MSK-LFM	$\frac{1}{2}\mathrm{erfc}(\sqrt{S_{\mathrm{NR}}})$
16QAM-LFM	$1-\left(1-\frac{3}{4}\mathrm{erfc}\left(\sqrt{\frac{1}{20}\sqrt{S_{\mathrm{NR}}}}\right)\right)^2$

注：S_{NR} 为接收信噪比；$\mathrm{erfc}(\cdot)$ 表示误差函数。

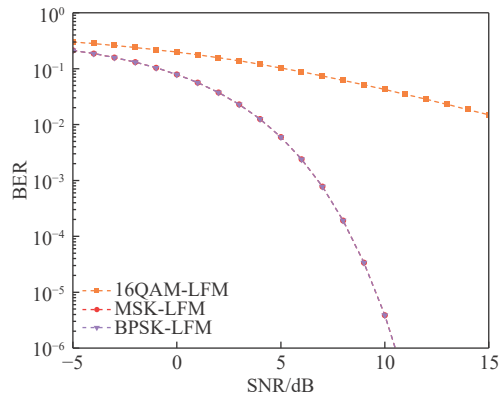


图 12 3 种波形的误码率曲线

Fig. 12 The BER curves of three waveforms

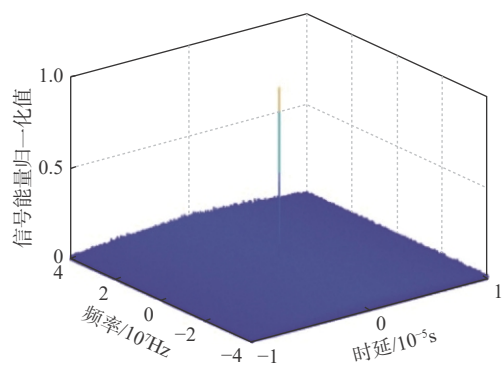


图 13 基于 OTFS 的 16QAM-LFM 信号模糊函数

Fig. 13 Ambiguity function of 16QAM-LFM signal based on OTFS

色的性能。相较于 BPSK 调制和 MSK 调制，这种卓越性能尤为显著。

通过对比 3 种一体化波形的通信和感知性能，可以发现，在不同调制中，OTFS-LFM 在不同性能侧重点时，具有其各自的不同表现。在权衡通感一体化性能时，通信性能和感知性能在不同调制下，达到各自的最优性能需要，再进行不同程度的性能折衷，才能获取一体化系统的最优表现。

2.4 UVLC 感知一体化设计

本节进行基于 OTFS-LFM 的 UVLC 感知一体化波形设计仿真实验。实验参考 2.3 节中与 3 种调制方式的结合，分别对比其通信与感知性能。在信

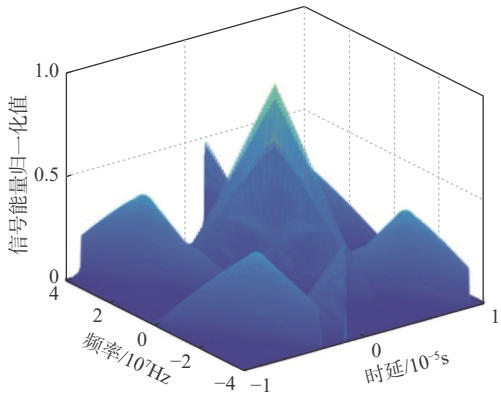


图 14　基于 OTFS 的 BPSK-LFM 信号模糊函数

Fig. 14　Ambiguity function of BPSK-LFM signal based on OTFS

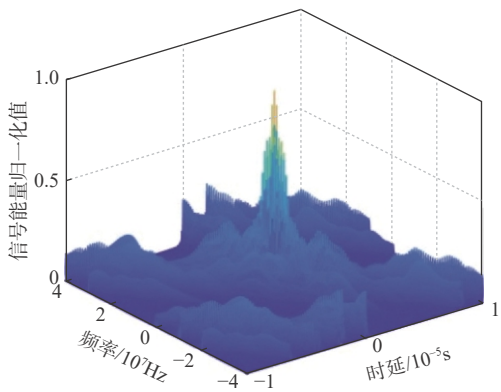


图 15　基于 OTFS 的 MSK-LFM 信号模糊函数

Fig. 15　Ambiguity function of MSK-LFM signal based on OTFS

道设置中, 选用 Gamma-Gamma 分布作为湍流模型[20]。在无线通信系统中, 湍流效应通常由大气中的气象条件引起, 导致信号强度的瞬时变化, 而在水下通信中, 湍流效应主要被用来模拟海水的流动变化。Gamma-Gamma 分布是一种适用于描述湍流信道模型衰落的概率分布, 其可以较好地模拟信号强度的瞬时变化, 同时考虑信号的多路径传播和阴影效应, 因此, 在仿真实验中使用该分布作为信道模型。在光源设置中, 选用 LED 作为光源进行仿真实验, 因其卓越的性能特征, 如低成本、高能效及灵活的部署能力, 使 LED 成为理想的通信系统实验中的光源。表 4 为一体化系统的主要参数设置。

表 4　信道及光源参数设置

Table 4　Channel and light source parameter settings

透镜折射率	LED 视场角/(°)	光学集光器增益	单LED 发射功率/W	LED 个数	海水光衰减系数
1.5	70	2.548 1	1	10	0.056

通信性能使用 BER 作为主要指标进行衡量。使用 MSK、16QAM、BPSK 这 3 种调制方式分别计算 BER, 得到对应的 BER 曲线, 如图 16 所示, 主要表现为 3 种方式 BER 随 SNR 逐渐升高, 且主体 BER 趋势表现为 MSK<BPSK<16QAM。与 2.3 节不同, 此时使用 BPSK 和 MSK 调制后的 BER 不相同, 可能是由于信道模型的改变导致采用 BPSK 调制的 BER 上升。

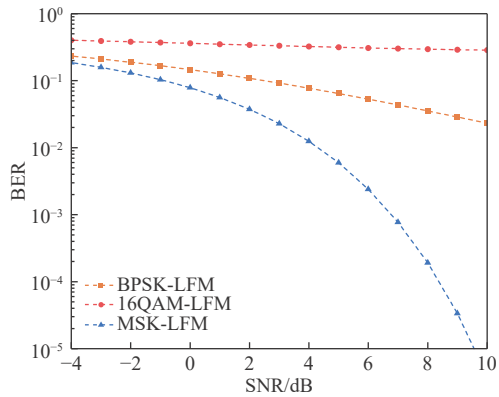


图 16　UVLC 感知误码率

Fig. 16　Underwater visible light communication sensing BER

在对比 3 种调制方式的感知性能时, 参考 2.1 节中使用 TPSE 算法计算 OTFS 感知性能, 仿真实验同样使用该算法对移动目标速度和距离进行感知估计。图 17～图 19 给出了 50 次迭代计算实验的估计计算结果。

同时, 为减小误差, 仿真实验进行 1 000 次重复测算, 并取平均值, 结果如表 5 所示。结果表明, 在感知性能上, 3 种调制方式表现为 16QAM>BPSK>MSK。因此, 结合图 16 和表 5, 分析一体化系统中折衷性能的考量, 在要求通信性能最优时, 可选用 MSK 作为主要的调制方式; 要求感知性能最优时, 可考虑使用 16QAM; 由于 MSK 在通感性能均有良好表现, 通感折衷性能系统可以将 MSK 调制作为主要方式。

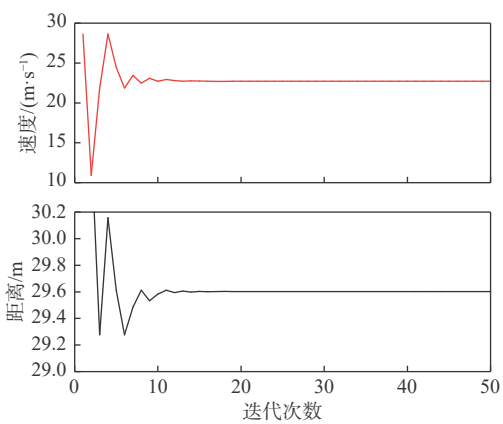


图 17　MSK-LFM 调制感知性能

Fig. 17　MSK-LFM modulation sensing performance

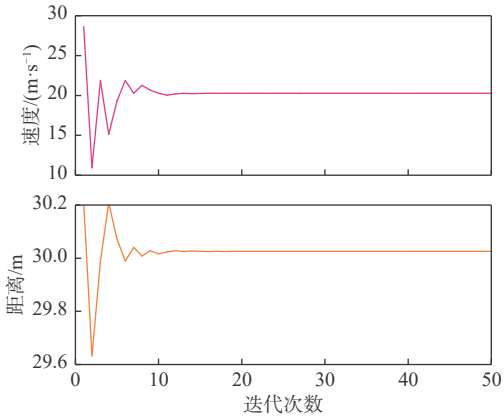


图 18 16QAM-LFM 调制感知性能

Fig. 18 16QAM-LFM modulation sensing performanc

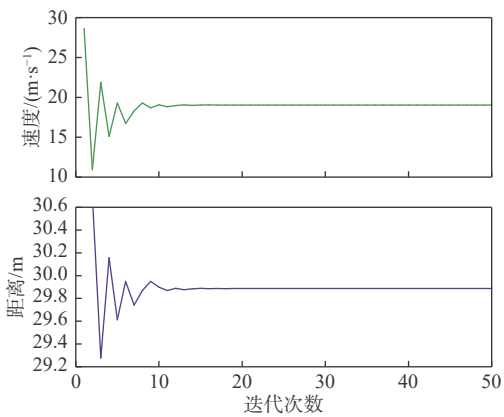


图 19 BPSK-LFM 调制感知性能

Fig. 19 BPSK-LFM modulation sensing performance

表 5 不同调制方式下的感知性能

Table 5 Sensing performance under different modulation modes

调制方式	速度/(m·s^{-1})	距离/m
BPSK-LFM	22.630	29.821
MSK-LFM	19.022	29.778
16QAM-LFM	20.178	30.013

3 结 论

本文在 UVLC 系统中，引入 OTFS 和 LFM 技术进行通信感知一体化研究，并对一体化系统的通感性能进行设计与分析。通过对比不同调制方式的 BER 及目标距离、速度信息，得出以下结论：

1）采用 OTFS 的通信与感知性能在一体化系统中均明显优于传统的 OFDM 调制。结果表明，OTFS 的引入为 UVLC 系统中通感性能的提升提供了有效途径，尤其在目标距离、速度信息的准确度方面取得显著成效。

2）将 LFM 技术嵌入 OTFS 中，通过比较 3 种调制方式的 BER 和模糊函数，进一步分析验证了不同调制在 OTFS-LFM 一体化波形设计中的性能。这有助于更全面地理解不同调制方式在整体系统中的适用性，为系统设计提供了更多选择。

3）在 UVLC 系统中采用 OTFS-LFM 技术设计一体化波形时，在 MSK 调制下一体化波形通信性能相比于其他 2 种方式较为突出；而 16QAM 作为调制方式时，感知性能可达到相对最优，通信性能略逊于 MSK。这说明在实际应用中，调制方式的选择取决于具体需求，需要在通信和感知性能之间做出权衡。

参考文献（References）

[1] KAUSHAL H, KADDOUM G. Underwater optical wireless communication[J]. IEEE Access, 2016, 4: 1518-1547.

[2] ZHU S J, CHEN X W, LIU X Y, et al. Recent progress in and perspectives of underwater wireless optical communication[J]. Progress in Quantum Electronics, 2020, 73: 100274.

[3] ALI M F, JAYAKODY D N K, LI Y H. Recent trends in underwater visible light communication (UVLC) systems[J]. IEEE Access, 2022, 10: 22169-22225.

[4] SONG G B, XU J. Review on long-distance underwater wireless optical communication[C]//Proceedings of the 2021 19th International Conference on Optical Communications and Networks. Piscataway: IEEE Press, 2021: 1-4.

[5] WANG C, YIN J W, DU P Y, et al. Application of orthogonal frequency division multiplexing in cognitive underwater communication[J]. Acoustical Society of America Journal, 2012, 132(3): 2015.

[6] RAVITEJA P, PHAN K T, HONG Y, et al. Interference cancellation and iterative detection for orthogonal time frequency space modulation[J]. IEEE Transactions on Wireless Communications, 2018, 17(10): 6501-6515.

[7] HADANI R, RAKIB S, TSATSANIS M, et al. Orthogonal time frequency space modulation[C]//Proceedings of the 2017 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2017: 1-6.

[8] 梁德永, 李凯, 张必华, 等. 线性调频和 MSK 雷达通信一体化波形设计[J]. 电声技术, 2024, 48(1): 134-136.
LIANG D Y, LI K, ZHANG B H, et al. Integrated waveform design of MSK radar communication based on linear frequency modulation[J]. Audio Engineering, 2024, 48(1): 134-136(in Chinese).

[9] 李惠东, 赵忠凯. 一种针对 LFM 雷达的非均匀间歇采样干扰样式[J]. 应用科技, 2020, 47(3): 37-40.
LI H D, ZHAO Z K. A non-uniform interrupted-sampling jamming mode for LFM radar[J]. Applied Science and Technology, 2020, 47(3): 37-40(in Chinese).

[10] 赵忠凯, 石妙. 基于 OFDM-LFM 的雷达通信一体化波形设计[J]. 应用科技, 2021, 48(3): 73-77.
ZHAO Z K, SHI M. Integrated radar and communication waveform design based on OFDM-LFM[J]. Applied Science and Technology, 2021, 48(3): 73-77(in Chinese).

[11] NOWAK M J, ZHANG Z P, LOMONTE L, et al. Mixed-modulated linear frequency modulated radar-communications[J]. IET Radar, Sonar & Navigation, 2017, 11(2): 313-320.

[12] LIU Z P, CHEN X B, WANG X M, et al. Communication analysis of integrated waveform based on LFM and MSK[C]//Proceedings of the IET International Radar Conference 2015. London: IET, 2016: 1-5.

[13] 曾浩, 吉利霞, 李凤, 等. 16QAM-LFM 雷达通信一体化信号设计[J]. 通信学报, 2020, 41(3): 182-189.
ZENG H, JI L X, LI F, et al. 16QAM-LFM waveform design for integrated radar and communication[J]. Journal on Communications, 2020, 41(3): 182-189(in Chinese).

[14] WANG Y G, TAO L, HUANG X X, et al. 8-Gb/s RGBY LED-based WDM VLC system employing high-order cap modulation and hybrid post equalizer[J]. IEEE Photonics Journal, 2015, 7(6): 7904507.

[15] KHALID A, ASIF H M, MUMTAZ S, et al. Design of MIMO-visible light communication transceiver using maximum rank distance codes[J]. IEEE Access, 2019, 7: 89128-89140.

[16] FARHANG A, REZAZADEHREYHANI A, DOYLE L E, et al. Low complexity modem structure for OFDM-based orthogonal time frequency space modulation[J]. IEEE Wireless Communications Letters, 2017, 7(3): 344-347.

[17] ZHOU C L, WANG C Y, GONG J, et al. Ambiguity function evaluation and optimization of the transmitting beamspace-based FDA radar[J]. Signal Processing, 2023, 203: 108810.

[18] WANG Z S, ZOU J F, ZHANG F R, et al. Cyclic MUSIC algorithm based on dual cyclic frequencies for direction finding [C]//Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications. Piscataway: IEEE Press, 2019: 560-564.

[19] GAUDIO L, KOBAYASHI M, CAIRE G, et al. On the effectiveness of OTFS for joint radar parameter estimation and communication[J]. IEEE Transactions on Wireless Communications, 2020, 19(9): 5951-5965.

[20] CHEN D, HUI J X. Parameter estimation of Gamma-Gamma fading channel in free space optical communication[J]. Optics Communications, 2021, 488: 126830.

Underwater visible light communication sensing integrated waveform design based on OTFS-LFM

YANG Xin，DUO Lin*，REN Yong，XU Boyu

(Faculty of Information Engineering and Automation，Kunming University of Science and Technology，Kunming 650500，China)

Abstract：Underwater visible light communication (UVLC) systems have gained a lot of attention due to the growing need for underwater communication. At present, the concept of communication and sensing integration has emerged, especially the UVLC systems' development has become a new research hotspot. Orthogonal time-frequency space (OTFS) technology has garnered significant attention due to its exceptional performance in high-Doppler and high-delay channels, thereby providing robust communication support for the system. At the same time, linear frequency modulation (LFM) technology is widely used in the field of wireless communication due to its low sensitivity to Doppler frequency shift. In this paper, an UVLC perception integration system is designed by combining OTFS technology with LFM technology. Through experimental simulation comparison, the system performs well in bit error rate (BER), ambiguity function, target speed and distance information acquisition. The system, which combines OTFS technology and LFM technology, exhibits excellent adaptability in complex underwater environments, bringing new possibilities to the field of underwater communication. At the same time, three common modulation methods are also considered, and the possible effects of different modulations on system performance are further analyzed, offering useful insights for system optimization.

Keywords：underwater visible light communication; communication and sensing integration; orthogonal time frequency space; linear frequency modulation; ambiguity function

Received：2023-11-30；**Accepted**：2024-04-26；**Published Online**：2024-05-31 10：40

URL：link.cnki.net/urlid/11.2625.V.20240530.1641.002

Foundation items：National Natural Science Foundation of China (61962032); Major Science and Technology Special Plan of Yunnan Provincial Science and Technology Department (202302AD080006)

* **Corresponding author**. E-mail：duolin2003@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0713

基于改进的组合赋权和云模型的平流层飞艇健康状态评估

曾盛富[1]，梁浩全[2]，王岩[1]，孙康文[1,*]，钟梓鸿[1]
(1. 北京航空航天大学 航空科学与工程学院，北京 100191；2. 北京航空航天大学 无人系统研究院，北京 100191)

摘 要：针对平流层飞艇健康评估需求与评估方法缺失的问题，提出一种结合专家评判系统的组合赋权和云模型的健康状态评估方法，为平流层飞艇健康评估提供技术支持。通过分析常见的平流层飞艇各子系统的可能风险及故障，筛选出评估指标体系；利用专家评判系统结合线性组合赋权法确定综合权重；根据平流层飞艇对象设定合理的评价集和标准云图，分别生成各指标及整体的云模型；采用云相似度、隶属度定量求解各指标健康情况。采用某平流层飞艇试验进行分析验证，结果表明：基于改进的组合赋权和云模型的平流层飞艇健康状态评估方法能有效开展评估，并在专家评判系统的结合中增加了模型对平流层飞艇的适用性，为后续开展平流层飞艇维修、保障及安控等任务提供有利前提。

关 键 词：平流层飞艇；专家评判系统；组合赋权；云模型；健康状态评估
中图分类号：V37
文献标志码：A **文章编号**：1001-5965(2026)01-0326-12

平流层飞艇飞行环境恶劣，且伴随复杂性增加及长时驻空要求的提高，设备面临更加复杂的飞行工况。确定飞艇健康状态并及时对其开展维护、更换，是平流层飞艇持续发展、长久飞行的关键手段。对其做良好的健康监测、保证飞艇长时驻空和正常完成任务是今后实质性提升平流层飞艇长时可靠性、保障完成既定目标、提高经济性的关键，也是未来发展平流层飞艇的必然要求[1]。

近年来，模糊综合评判对复杂系统的适用性使其受到部分学者的青睐。国内外学者将模糊综合评判法应用至大型设备[2-3]、军事[4]及铁路建设[5-6]等领域。在模糊综合评判中，单一的权重求解法存在明显的主观或客观缺陷，因此，出现了组合赋权法，组合赋权法因其融合主观权重与客观权重而避免权重单一的优势被不断应用[7-11]。组合赋权法是将主观赋权法与客观赋权法进行结合的一种权重求解方法。组合赋权法主要有线性组合赋权法、乘法集成法和动态赋权法等。传统的组合赋权方法仅考虑主客观权重数值之间的最优平衡，亦或仅简单将两者对半划分，此类结合方法对分析对象自身特点的考虑略有不足。

此外，传统的模糊综合评判采用设定精确隶属度函数的方式求解指标隶属度，失去了模糊综合评判本身追求模糊的理念。而云模型是处理定性概念与定量描述的不确定转换模型，其拥有比模糊隶属度函数更强的普遍适用性与描述不确定问题的能力[12]，因此，采用云模型替代隶属度函数的方式可有效提高对象分析中的模糊性。云模型与组合赋权法两者各自的优势十分明显，近几年不少学者尝试将云模型与组合赋权法进行结合，由云模型确定隶属度、组合赋权法确定权重。目前该方法已应用于飞行器领域[13]、军事领域[14]和施工风险评估[15]

收稿日期：2023-11-02；录用日期：2024-01-12；网络出版时间：2024-01-23 17:25
网络出版地址：link.cnki.net/urlid/11.2625.V.20240123.0945.001
基金项目：国家自然科学基金 (52302511)
*通信作者. E-mail：09142@buaa.edu.cn

引用格式：曾盛富，梁浩全，王岩，等. 基于改进的组合赋权和云模型的平流层飞艇健康状态评估[J]. 北京航空航天大学学报，2026，52(1)：326-337. ZENG S F，LIANG H Q，WANG Y，et al. Health status assessment of stratospheric airship based on improved combination weighting and cloud model[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：326-337 (in Chinese).

等方面。

截至目前，国内外平流层飞艇的健康管理研究也尚处于初级阶段。而云模型与组合赋权法结合实现健康评估的方式也在近些年得到了验证。结合平流层飞艇自身健康情况模糊的特性，考虑采用云模型与组合赋权的方式实现其健康评估需求。此外，由于云模型与组合赋权结合中仅采用简单结合、均匀划分等形式确定参数，且传统的等级划分与健康等级区间确定不具有特殊性，此结合方式、评价集与标准云缺乏对象针对性。因此，为解决权重求解缺陷、增加求解模糊性及体现对象特殊性，本文引入专家评判系统，提出采用改进的组合赋权和云模型的平流层飞艇健康状态评估法，有效针对平流层飞艇开展健康状态评估。首先，针对平流层飞艇进行总体概况分析，依据分析情况初步筛选平流层飞艇健康评估指标，为避免初步选取指标的数据差异性程度较小导致其无法反馈健康水平，考虑采用变异系数及数值分布实现指标的二次筛选，完成指标体系的构建工作。其次，结合专家评判实现主客观权重及组合权重的计算。再次，依照专家评价结果确定平流层飞艇健康等级划分和标准云模型，并结合权重实现综合云求解。最后，结合某飞艇健康评估情况，验证该评估模型的有效性。

1　平流层飞艇健康评估指标体系构建

1.1　平流层飞艇总体概况

目前，常见平流层飞艇主要包括囊体、尾翼、能源系统、推进系统和吊舱(内含设备)等部件。飞艇往下层级划分一般可以分为：囊体分系统、能源分系统、推进分系统、航电分系统及其他分系统(包括测控、安控、结构和飞控等)。

风险是平流层飞艇实际飞行中一定存在的因素，而风险有导致设备故障、系统故障进而导致飞行终止的可能。平流层飞艇面临各式各样的风险与故障可能。

飞艇囊体承担提供浮力、保持外形、保证超压状态和承力等责任。囊体的主要风险故障为囊体破裂和囊体漏气。能源分系统的主要风险故障为太阳能电池阵列异常、储能锂电池异常和能源管理系统异常等。太阳能电池阵列异常包括太阳能电池板损坏、太阳能电池板脱落；储能锂电池异常可能包括电池过充过放异常、电池温度异常、电池低温低压工作异常；能源管理系统异常包括变换器异常。推进分系统的主要部件为推进电机和螺旋桨。其风险故障包括推进电机失效和螺旋桨失效等。此外，航电分系统的风险故障包括飞艇管理系统失效、采集控制系统失效、图像采集系统失效和电气连接系统失效。测控分系统主要存在链路中断、指令发送失败等问题。飞控分系统主要存在控制发散等问题。安控分系统承担飞艇的降落能力，其主要风险故障在于阀门的开启关闭是否正常与安控逻辑是否失效。结构分系统存在结构强度、刚度不足和不稳定的风险。

1.2　评估指标筛选

根据 1.1 节对平流层可能遭遇风险、故障的分析，可以找寻出多个影响飞艇健康状态的相关指标。为获取平流层飞艇真实健康情况，在遵守客观性、灵活性、适应性、完整性和可衡量性的选取原则上，通过深入各系统分析、咨询相关专家，初步选定以平流层飞艇作为目标层指标，以囊体、压力调节系统、储能锂电池、高压 DC/DC 变换器和推进电机作为对象层指标，此外，还筛选出 16 个指标层指标。初步体系如图 1 所示。

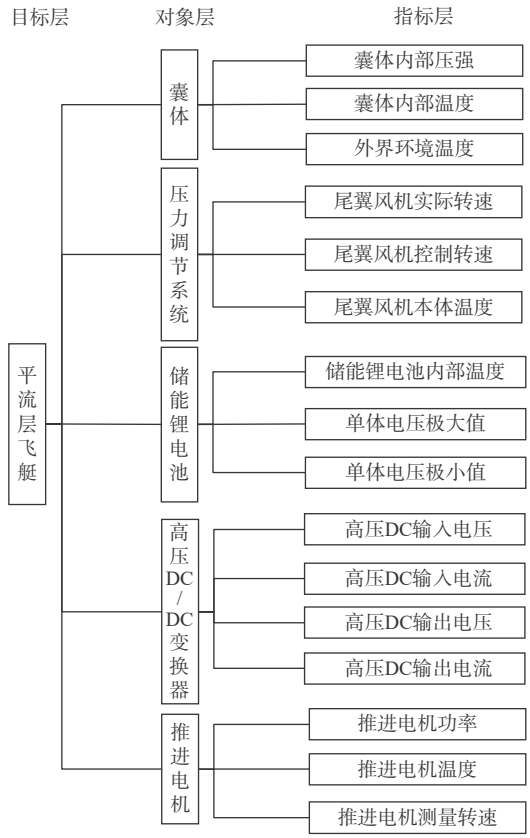


图 1　平流层飞艇评估指标初步体系

Fig. 1　Preliminary evaluation index system of stratospheric airship

但此类指标的反馈数据可能存在差异程度较小的问题。对于分析某时间段的平流层飞艇、部件的健康状态而言，在该时间段内的数据差异程度较小将严重影响客观权重的求解。本节通过计算各指标层指标在某小时时间段内的变异系数及各指标数值数量来判断指标的差异程度，通过求解得到的变异系数筛选出待再筛选指标，再从待再筛选指标中求

解数据分布情况来进一步确定应舍弃的指标。为使后续指标筛选过程表示方便，将各指标用符号代替，其中，囊体内部压强、囊体内部温度和外界环境温度分别以NT1、NT2和NT3表示；尾翼风机实际转速、尾翼风机控制转速和尾翼风机本体温度分别以YLTJ1、YLTJ2和YLTJ3表示；储能锂电池内部温度、单体电压极大值和单体电压极小值分别以CNDC1、CNDC2和CNDC3表示；高压DC输入电压、高压DC输入电流、高压DC输出电压和高压DC输出电流分别以GYDC1、GYDC2、GYDC3和GYDC4表示；推进电机功率、推进电机温度和推进电机测量转速分别以TJDJ1、TJDJ2和TJDJ3表示。

随机选取某平流层飞艇实际飞行过程中的某一天的某时间段数据，计算各指标在该时段的变异系数与数值分布情况，以实现指标筛选。

1.2.1 变异系数初步筛选指标

标准差和方差常用于判断2组数据的离散程度。然而，传统的标准差、方差方法只能对平均值相等或相差不大的情况进行比较。对于2组量纲不同或量级不同的数据而言，传统的方法并不能正确表示数据的离散情况，无法辅助筛选指标。为对所选取的指标进行初步筛选，此处引入变异系数，通过求解各指标在某时间段内的离散程度来初步挑选可能舍弃的指标。求解中增加绝对值以避免数据均值为负数时导致计算结果为负。指标层指标的变异系数求解结果如表1所示。

表1 各指标变异系数

Table 1 Coefficient of variation of each indicators

指标	变异系数/%
NT1	2.95
NT2	1.64
NT3	1.92
YLTJ1	1.27
YLTJ2	0
YLTJ3	31.91
CNDC1	1.03
CNDC2	0.061
CNDC3	0.040
GYDC1	1.07
GYDC2	21.30
GYDC3	0.15
GYDC4	21.53
TJDJ1	15.47
TJDJ2	22.41
TJDJ3	4.61

由表1可知，部分指标的离散程度较小，其可能无法正常处理数据。通过对各回传数据的初步估计，初步筛选变异系数小于或等于3%的指标作为可能剔除指标，并进入二次筛选。

1.2.2 数值分布二次筛选

对于需要求解客观权重的指标而言，若其反馈数据长时间保持在一个数值不变或仅在少量的数据组范围内变化，则其过于集中，将导致指标差异程度较小，难以作为后续开展权重计算的指标，需要进行剔除。此处采用计算该时间段各指标数据出现的数值分布情况（主要计算在某组数据中，共计出现多少种数值，记作数值数量）作为判断依据，来剔除部分指标。其计算结果如表2所示。

表2 二次筛选结果

Table 2 Secondary screening results

指标	数值数量
NT1	135
NT2	23
NT3	3
YLTJ1	130
YLTJ3	1
CNDC1	2
CNDC2	23
CNDC3	16
GYDC1	51
GYDC3	2

根据计算得到的结果及经验分析，对于该时段内的数据而言，数值数量较小的指标变化太不明显，实际求解客观权重中易出现0权重可能，无法加入指标体系。尤其是表2中的NT3、YLTJ3、CNDC1和GYDC3数值出现数量极小，分别为3、1、2和2，与其他指标不在一个量级。这4个指标在该时段内表现出极小的差异程度，说明指标数据不具备良好的状态表征能力，应当剔除。

1.3 评估指标体系构建

根据1.2节对所选取的指标进行筛选后得到共计1个目标层指标为平流层飞艇、5个对象层指标为囊体、压力调节系统、储能锂电池、高压DC/DC变换器和推进电机（D1～D5）及12个指标层指标：囊体内部压强、囊体内部温度、尾翼风机实际转速、尾翼风机本体温度、单体电压极大值、单体电压极小值、高压DC输入电压、高压DC输入电流、高压DC输出电流、推进电机功率、推进电机温度和推进电机测量转速（Z1～Z12）。具体指标体系如图2所示。

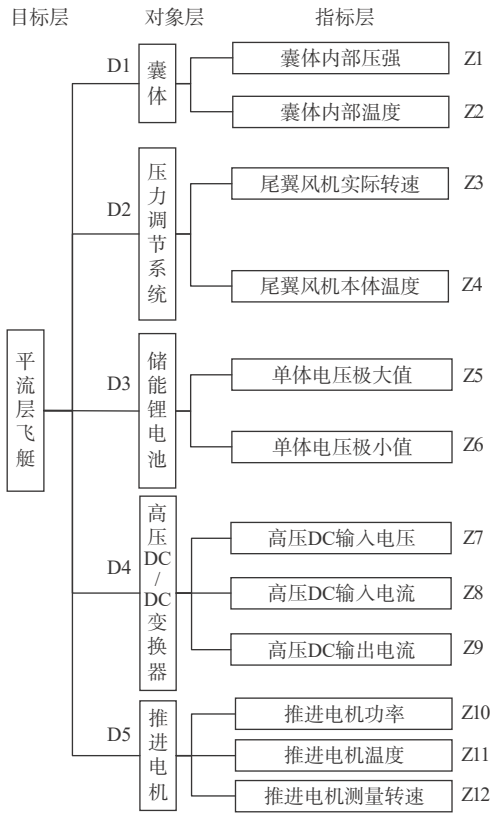


图 2　平流层飞艇评估指标体系

Fig. 2　Evaluation index system of stratospheric airship

2　结合专家评判的组合赋权法

针对平流层飞艇, 采用线性组合赋权法, 通过多位飞艇行业领域的专家对评价指标的主客观重要性进行评分, 提出适用于平流层飞艇的主客观权重系数, 再将其组合成最终权重。引入专家评判系统得到的综合权重更具针对性和可靠性, 同时改善后续确定指标及系统健康状态的准确性。

2.1　主客观权重求解

在平流层飞艇的健康评估中, 计算权重是非常重要的一环。组合赋权法将主客观方法进行结合, 既考虑主观专家的评价, 又考虑客观因素的影响, 充分发挥 2 种方法的优点, 使权重赋值更加准确。本文采用成熟度较高的方法: 层次分析法(主观赋权法)及熵权法(客观赋权法)求解主客观权重。

2.1.1　层次分析法权重求解

层次分析法主要由专家评判系统对各指标进行打分, 再完成数据处理, 最终得到主观权重结果。

层次分析法求解权重步骤如下:

1) 构造判断矩阵。令每位专家以 9 标度法分别对对象层及指标层评分, 从而获取各判断矩阵 $\boldsymbol{P}$:

$$\boldsymbol{P}=[a_{ij}]_{n\times n}\qquad i,j=1,2,\cdots,n \tag{1}$$

式中: a_{ij}为评价值, i和j为序号。

2) 层次单排序。

步骤 1　将判断矩阵按列归一化处理得到 a_{ij}^{norm}:

$$a_{ij}^{\mathrm{norm}}=a_{ij}\Big/\sum_{k=1}^{n}a_{kj}\qquad i,j=1,2,\cdots,n \tag{2}$$

步骤 2　将得到的归一化矩阵按行求和得到向量$\boldsymbol{\sigma}$, 元素σ_i可表示为

$$\sigma_i=\sum_{j=1}^{n}a_{ij}^{\mathrm{norm}} \tag{3}$$

步骤 3　将所得求和向量归一化, 得到的结果即为层次单排序的权重值w_i:

$$w_i=\sigma_i\Big/\sum_{i=1}^{n}\sigma_i \tag{4}$$

$$\boldsymbol{W}^{\mathrm{a}}=(w_1,w_2,\cdots,w_n) \tag{5}$$

$$(\boldsymbol{W}^{\mathrm{a}})^{\mathrm{T}}=(w_1,w_2,\cdots,w_n)^{\mathrm{T}} \tag{6}$$

式中: $\boldsymbol{W}^{\mathrm{a}}$为主观权重; $(\boldsymbol{W}^{\mathrm{a}})^{\mathrm{T}}$为主观权重向量的转置。

步骤 4　计算最大特征根$\lambda_{\max}$:

$$\lambda_{\max}=\sum_{i=1}^{n}\frac{(\boldsymbol{P}(\boldsymbol{W}^{\mathrm{a}})^{\mathrm{T}})_i}{nw_i} \tag{7}$$

式中: $(\boldsymbol{P}(\boldsymbol{W}^{\mathrm{a}})^{\mathrm{T}})_i$为$\boldsymbol{P}(\boldsymbol{W}^{\mathrm{a}})^{\mathrm{T}}$的第$i$个分量。

3) 一致性检验。检验系数C_{R}的求解公式为

$$C_{\mathrm{R}}=\frac{C_{\mathrm{I}}}{R_{\mathrm{I}}} \tag{8}$$

式中: $C_{\mathrm{I}}=(\lambda_{\max}-n)/(n-1)$; R_{I}为随机一致性指标, 通过查表得到。

2.1.2　熵权法权重求解

熵权法确定权重的流程主要包括获取原始数据矩阵、原始数据矩阵归一化、数据修正、计算比重、求解信息熵、计算差异性程度和求解权重。

1) 获取原始数据矩阵。假设某 n 个评价指标共有 m 组数据, 则原始数据矩阵$\boldsymbol{R}$可以表示为

$$\boldsymbol{R}=\begin{bmatrix}r_{11}&r_{12}&\cdots&r_{1n}\\r_{21}&r_{22}&\cdots&r_{2n}\\\vdots&\vdots&&\vdots\\r_{m1}&r_{m2}&\cdots&r_{mn}\end{bmatrix} \tag{9}$$

2) 原始数据矩阵归一化。针对不同指标采用不同归一化方法(包括正向指标、负向指标和适度指标)消除量纲, 再进行平移修正后得到对应矩阵$\boldsymbol{R}_0^{\mathrm{norm}}$:

$$\boldsymbol{R}_0^{\text{norm}}=\begin{bmatrix} r_{11}^{\text{norm}}+\varepsilon & r_{12}^{\text{norm}}+\varepsilon & \cdots & r_{1n}^{\text{norm}}+\varepsilon \\ r_{21}^{\text{norm}}+\varepsilon & r_{22}^{\text{norm}}+\varepsilon & \cdots & r_{2n}^{\text{norm}}+\varepsilon \\ \vdots & \vdots & & \vdots \\ r_{m1}^{\text{norm}}+\varepsilon & r_{m2}^{\text{norm}}+\varepsilon & \cdots & r_{mn}^{\text{norm}}+\varepsilon \end{bmatrix}= \begin{bmatrix} R_{11}^{\text{norm}} & R_{12}^{\text{norm}} & \cdots & R_{1n}^{\text{norm}} \\ R_{21}^{\text{norm}} & R_{22}^{\text{norm}} & \cdots & R_{2n}^{\text{norm}} \\ \vdots & \vdots & & \vdots \\ R_{m1}^{\text{norm}} & R_{m2}^{\text{norm}} & \cdots & R_{mn}^{\text{norm}} \end{bmatrix} \tag{10}$$

式中：r_{mn}^{norm}为归一化后的数据；ε为趋于0的极小值；R_{mn}^{norm}为修正后的数据。

3）求解比重p_{ij}。

$$p_{ij}=R_{ij}^{\text{norm}}\Big/\sum_{i=1}^{m}R_{ij}^{\text{norm}} \tag{11}$$

4）求解信息熵e_j。

$$e_j=-K\sum_{i=1}^{m}p_{ij}\ln p_{ij} \tag{12}$$

式中：$K=1/\ln m$。

5）计算差异性程度G_j。

$$G_j=1-e_j \tag{13}$$

6）求解熵权w_j^*。

$$w_j^*=G_j\Big/\sum_{j=1}^{n}G_j \tag{14}$$

$$\boldsymbol{W}^{\text{b}}=\left(w_1^*,w_2^*,\cdots,w_n^*\right) \tag{15}$$

式中：$\boldsymbol{W}^{\text{b}}$为利用熵权法求解得到的指标权重。

2.2 组合权重求解

组合赋权法选用线性组合赋权法作为权重求解方法，可以通过改变权重系数以适用于平流层飞艇。线性组合赋权法的公式为

$$W_i=\alpha W_i^{\text{a}}+\beta W_i^{\text{b}} \tag{16}$$

式中：W_i为某指标组合权重；W_i^{a}为某指标通过层次分析法求解得到的权重；W_i^{b}为某指标通过熵权法求解得到的权重；α和β分别为主、客观权重系数，$\beta=1-\alpha$。

根据多位专家对评价指标的主客观重要性进行评判给出的线性系数求解得到主观权重系数和客观权重系数。其中，主观权重系数α为0.38，客观权重系数β为0.62。

3 结合专家评判的云模型状态评估

传统健康评价方法采用精确隶属度函数形式求解对象健康水平。利用云模型替代模糊综合评判法中的隶属度函数，有利于保持模糊评价初衷。本文利用云模型替换隶属度函数来确定指标的健康情况。

使用云模型前需要先确定指标的等级划分（即评价集），再生成标准云模型。在等级划分与标准云确定中，结合专家评判系统制定健康等级及等级区间的定值，以适配平流层飞艇这一研究对象。将标准云模型与求解得到的指标云及综合云进行比对，定性确定指标云及综合云的健康状态。引入云隶属度及相似度定量求解指标的健康隶属度。应用云模型实现平流层飞艇健康状态评估更加符合模糊综合评判理念，有效开展对模糊事件的评判。

3.1 正态云

在隶属度函数中，存在三角形隶属度函数、抛物线形函数、梯形函数和正态分布函数等。云模型作为隶属度函数的替代，也存在三角云模型、梯形云模型、泊松云模型和正态云模型等。正态云模型由于充分结合了正态分布和正态隶属度函数的普适性，从而相对于模糊隶属度函数能够更好地刻画具有较强不确定性的设备状态评估指标特性[16]。

云模型的3个特征参数为E、H、H_2，分别表示期望、熵和超熵。期望是最能够代表定性概念的点，为云图中的近似对称轴线所在的横坐标值，是此概念量化的最典型样本；熵属于不确定度的度量，反映了云滴的离散程度，为云图横跨宽度的一种近似度量；超熵即熵的熵，在云图上表示为云的厚度。正态云图如图3所示。

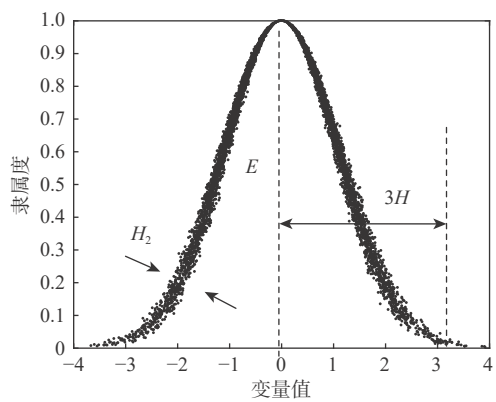


图3 正态云图

Fig. 3 Normal cloud map

3.2 结合专家评判的标准云

在构建标准云模型前，需要对健康等级做出划分。常见的等级划分可以分为2～9级。复杂系统不同于简单单机，复杂系统出现个别内部单机设备异常或故障时，由于系统的安全余度设计及优异的隔离机制等，往往可以通过降额运行使任务安全结束。而针对简单单机而言，出现故障继续运行或停止运行，较少能够以降额状态运行。因此，在等级划分时，简单单机一般划分为较少等级，复杂系统则会划分为更多等级。平流层飞艇是一个较复杂

的系统平台，当部分推进系统故障时，由于自身浮力主要来源于囊体氦气，丧失部分动力不会导致飞艇被迫降落。通过合理隔离，可以操纵飞艇降额飞行，从而完成飞行任务，若超出降额边界，再考虑紧急降落。因此，针对平流层飞艇而言，本文将其健康等级划分为 5 级，分为健康、良好、一般、退化、异常等级。即评价集 $L=\{L_1,L_2,L_3,L_4,L_5\}$，$L_1$ 表示健康，L_2 表示良好，L_3 表示一般，L_4 表示退化，L_5 表示异常。

传统的等级划分区间往往采用均分法或黄金分割法，这些划分方法并不能很好地表征设备或系统在这些等级的实际占比。例如，某些简易单机健康等级划分为正常与故障，但实际评分中存在低于 60 分即故障的评价情况。为避免类似情况，可以通过增加划分等级数和更改健康等级区间边界来解决，然而增加划分等级数会使整个模型更加复杂且流程更加烦琐。对于平流层飞艇而言，其也存在符合自身特点的区间边界。通过邀请多位行业专家根据平流层飞艇自身特点确定评价集内各等级区间，并对结果区间做简易处理，确定各个等级上下界（范围为 0～100），获取更适用于平流层飞艇健康状态的等级划分区间。平流层飞艇健康等级划分结果如表 3 所示。

表 3　平流层飞艇健康等级划分

Table 3　Health classification of stratospheric airships

健康状态	区间范围	健康状态描述
健康	84～100	平流层飞艇飞行状况很好，能够很好地实现可控飞行
良好	61～84	平流层飞艇飞行状况较好，能够较好地实现可控飞行
一般	40～61	平流层飞艇状态一般，部件存在疲劳现象
退化	19～40	平流层飞艇各部件出现问题征兆，需降额飞行
异常	0～19	平流层飞艇关键部件不能工作，需降额飞行或紧急降落

标准云可以采用完整云或半云，针对平流层飞艇的边界等级健康与异常而言，平流层飞艇健康状态位于该等级的概率相对较低，半云形式的标准云可以更好地适应表现健康和表现异常的数据分布情况，并且更加简洁、直观。因此，以半云的形式生成健康和异常 2 个等级的云图，其余等级正常生成完整正态云图。健康和异常 2 个等级的一侧有固定边界（单边定界），而良好、一般和退化 3 个等级初始无明确边界，通过评分给出两侧范围（双边定界），其中，两侧的值分别为该评价等级的上下限。

每个评价等级的上下限可以通过计算转化为云模型的 3 个数字特征，假设双边定界的评价等级的上下限为 $[X_{\min},X_{\max}]$，则转化为数字特征（E、H、H_2）的计算公式为

$$\begin{cases} E=(X_{\max}+X_{\min})/2 \\ H=(X_{\max}-X_{\min})/6 \\ H_2=kH \end{cases} \tag{17}$$

式中：k 为常数，由实际应用情况确定，本文取 0.2。

假设健康的分值范围为 $[X_{\min},100]$，则转化为数字特征的计算公式为

$$\begin{cases} E=0 \\ H=X_{\max}/3 \\ H_2=kH \end{cases} \tag{18}$$

则各等级数字特征如表 4 所示。将各数字特征代入云发生器中，求解得到平流层飞艇的标准云模型如图 4 所示。

表 4　标准云模型数字特征

Table 4　Digital features of standard cloud model

数字特征	异常	退化	一般	良好	健康
E	0	29.5	50.5	72.5	100
H	6.33	3.50	3.50	3.83	5.33
H_2	1.27	0.70	0.70	0.77	1.07

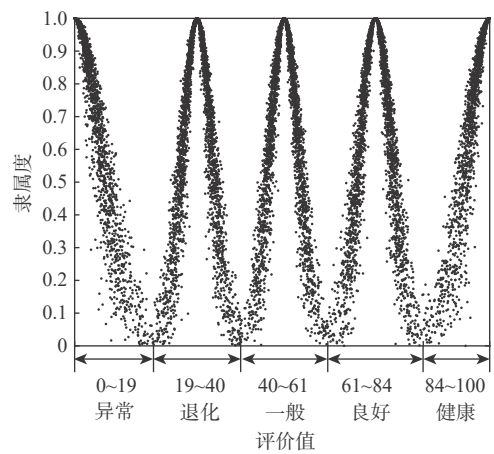


图 4　标准云模型

Fig. 4　Standard cloud model

3.3　评价指标云模型

利用相对劣化度来获取评价指标的云模型参数，可以避免因依赖专家评分获取指标分值而导致整体评价过度主观化。相对劣化度是基于实际回传数据的一种指标状态反馈，通过其回传数据所反馈出的指标劣化情况反推出该指标在标准云下的所处位置，是一种较好的评价指标云的生成方法。相对劣化度表征指标的劣化程度，其范围为 [0,1]，其中当相对劣化度为 0 时，代表指标表现最好，映射至健康水平为最健康；相对劣化度为 1 时，代表指标表现最差，映射至健康水平为最异常。劣化度与归一化类似，主要分为 3 种指标，分别为越大越好型、越小越好型和适度型。

1）越大越好型指标相对劣化度计算如下：

$$d(x)=\begin{cases}1 & x<x_{\min}\\ \dfrac{x_{\max}-x}{x_{\max}-x_{\min}} & x_{\min}\leqslant x\leqslant x_{\max}\\ 0 & x>x_{\max}\end{cases} \tag{19}$$

2)越小越好型指标相对劣化度计算如下:

$$d(x)=\begin{cases}0 & x<x_{\min}\\ \dfrac{x-x_{\min}}{x_{\max}-x_{\min}} & x_{\min}\leqslant x\leqslant x_{\max}\\ 1 & x>x_{\max}\end{cases} \tag{20}$$

3)适度型指标相对劣化度计算如下:

$$d(x)=\begin{cases}1 & x<x_{\min}\\ \dfrac{x_0-x}{x_0-x_{\min}} & x_{\min}\leqslant x\leqslant x_0\\ \dfrac{x-x_0}{x_{\max}-x_0} & x_0<x\leqslant x_{\max}\\ 1 & x>x_{\max}\end{cases} \tag{21}$$

式中:x为指标实际反馈数据;$x_{\min}$为指标下限;$x_{\max}$为指标上限;x_0为指标的最佳值。

相对劣化度的数值与标准云模型中的健康等级评分数值成反比,为方便识别和评估总体健康状态,并适配评分范围,相对劣化度需要进一步映射为健康等级评分。相对劣化度数值越低代表该指标越优异、评分越高;相对劣化度数值越高代表该指标越低劣、评分越低。利用式(22)映射为反映健康水平的分值数据:

$$v(x)=H_{\max}(1-d(x)) \tag{22}$$

式中:$H_{\max}$为健康评估水平分值上限,本文取100。

假设某指标共有P组数据,则通过计算可以得到对应的相对劣化度,再映射为健康水平分值,则该指标的云模型数字特征的计算公式分别为

$$E=\frac{1}{n}\sum_{i=1}^{n}v_i \tag{23}$$

$$H=\sqrt{\frac{\pi}{2}}\cdot\frac{1}{n}\sum_{i=1}^{n}|v_i-E| \tag{24}$$

$$H_2=\sqrt{\left|\frac{1}{n-1}\sum_{i=1}^{n}(v_i-E)^2-H^2\right|} \tag{25}$$

式中:v_i为第i组健康水平分值。

利用求解得到的数字特征代入正向云发生器生成指标云图,并与标准云图形成对比,可以定性获取该指标基于数据反馈的相对劣化度所映射的健康水平。

3.4 综合云求解与量化

综合云通过求解的权重及各指标云参数二次计算得到。综合云为虚拟云[17],即将各个基云的数字特征进行计算得到一个新云[18]。第2节已经根据组合赋权法求解得到了各指标权重W_i,并根据专家评分样本计算得到各评价指标云参数(E、H、H_2)。假定某指标集共有n个指标,则该指标集的综合云参数计算公式如下:

$$E=\sum_{i=1}^{m}W_iE_i \tag{26}$$

$$H=\sqrt{\sum_{i=1}^{m}\left(\frac{W_i^2}{\sum_{j=1}^{m}W_j^2}H_i^2\right)} \tag{27}$$

$$H_2=\sqrt{\sum_{i=1}^{m}\left(\frac{W_i^2}{\sum_{j=1}^{m}W_j^2}H_{2,i}^2\right)} \tag{28}$$

得到综合云的3个数字特征后,可以根据该参数生成各指标集的综合云图,将生成的云图放置到标准云图中进行对比,可以直观地看出该指标集所处健康状态。

虽然直接采用直接看图法对比综合云与标准云图的位置关系,可以比较方便地看出云图所处的健康等级,但对于某些位于不易分辨的靠近2个等级中间位置的云图,用直接看图法确定其健康等级具有一定难度。此外,对于比较相近的2个云图来说,也不易于判断两者之间孰优孰劣。因此,引用定量指标解决此类问题。本文采用云图的隶属度[19]表示:

$$\varphi_l=C(C_l,C_0)\Big/\sum_{l=1}^{5}C(C_l,C_0) \tag{29}$$

式中:φ_l为该指标集下的第l个等级的隶属度;$C(C_l,C_0)$为C_l和C_0这2个云模型之间的相似度。相似度计算公式为

$$C(C_l,C_0)=\frac{1}{2}+\frac{1}{2\theta}-\theta \tag{30}$$

式中:θ的计算公式为

$$\theta=\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{\left(-\frac{x^2}{2}\right)}\mathrm{d}x \tag{31}$$

其中:t的计算公式为

$$t=\frac{|E_l-E_0|}{\sqrt{H_l^2+H_{2,l}^2}+\sqrt{H_0^2+H_{2,0}^2}} \tag{32}$$

通过对云图隶属度的求解可以从数值上清楚看出求解对象的健康隶属情况。对于一级指标对应的部件,求解云图隶属度可以获得该指标(部件)的健康状态,对于整个平流层飞艇,求解云图隶属度可以获取飞艇整体的健康状态。根据飞艇整体的健康情况,可以为后续检修、查因等做出针对性工作。

平流层飞艇健康评估整体采用云模型与组合赋权法并结合专家评判系统的方式实现,总体的技术路线流程如图5所示。

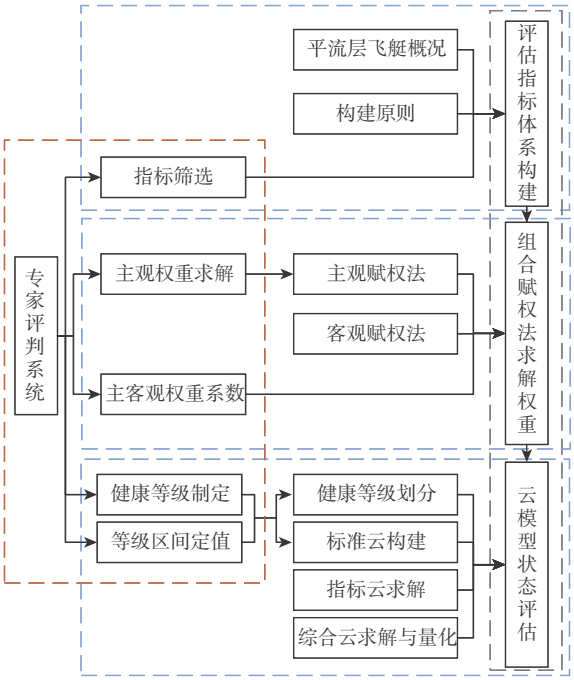


图 5　总体技术路线流程

Fig. 5　Overall technical roadmap

4　试验分析

针对某平流层飞艇的一次试验案例, 对其健康状态做评估。本次评估为了更加凸显平流层飞艇健康状态的变化情况, 选取实际试验过程中连续 4 个周期的试验数据作为原始数据, 该原始数据在第 4 周期时, 推进电机出现故障。

4.1　飞艇指标权重求解

权重由第 2 节所确定的利用主客观权重结合的方法来求解。其中, 主观赋权法采用层次分析法, 客观赋权法采用熵权法。两者的组合方式为评分线性组合赋权法。

层次分析法利用问卷的形式获取数位专家针对指标的评分, 并对所回收问卷做出有效性分析, 提取有效问卷进行数据处理获取主观权重。

客观权重由熵权法求解得到, 在确定每个指标的上限值、下限值和标准值(3 值)后, 对每周期的原始数据归一化、平移修正、计算比重、求解信息熵、计算差异性程度和求解熵权, 最终得到 4 个周期的客观权重。

组合权重由主观权重与客观权重以线性形式结合而成, 2.2 节已通过专家评判给出主观权重系数α=0.38, 客观权重系数β=0.62。因此, 各指标每周期(cy1～cy4)的组合权重如表 5 所示。

表 5　平流层飞艇指标权重

Table 5　Index weights of stratospheric airships

目标层	对象层	指标层	主观权重	主观位次	客观权重				组合权重			
					第1周期	第2周期	第3周期	第4周期	第1周期	第2周期	第3周期	第4周期
平流层飞艇	D1	Z1	0.280	1	0.067	0.108	0.125	0.003	0.148	0.173	0.184	0.108
		Z2	0.152	2	0.103	0.115	0.126	0.003	0.122	0.129	0.136	0.060
	D2	Z3	0.042	10	0.131	0.087	0.160	0.003	0.097	0.070	0.115	0.018
		Z4	0.043	9	0.072	0.031	0.066	0.002	0.061	0.036	0.057	0.017
	D3	Z5	0.106	3	0.004	0.017	0.015	0	0.043	0.051	0.049	0.040
		Z6	0.075	4	0.003	0.011	0.010	0	0.031	0.036	0.035	0.029
		Z7	0.052	8	0.116	0.099	0.103	0.005	0.091	0.081	0.083	0.023
	D4	Z8	0.052	7	0.091	0.080	0.082	0.003	0.076	0.070	0.070	0.021
		Z9	0.074	5	0.018	0.034	0.064	0.002	0.039	0.049	0.067	0.029
	D5	Z10	0.039	11	0.073	0.051	0.061	0.137	0.060	0.046	0.052	0.100
		Z11	0.061	6	0.266	0.169	0.076	0.395	0.188	0.128	0.070	0.268
		Z12	0.025	12	0.056	0.198	0.115	0.447	0.044	0.133	0.081	0.287

4.2　云模型状态评估

4.2.1　指标云求解

根据式(19)～式(22), 求解每周期的相对劣化度。再计算各指标的云模型数字特征(E、H、H_2), 利用云发生器实现数字特征向云图的转换, 并将图像放置于标准云图中进行比较。此处获取的指标云为底层特征, 是生成上级指标云的基础。各指标云如图 6 所示, 可以看出, Z1、Z2、Z5、Z6、Z7、Z8 和 Z9 的指标云随着周期增加, 指标健康状况呈现左移的趋势; Z3 和 Z4 的健康状态随周期变化较小, 与自身特性有关。此外, 部分云图较厚, 表明该数据较离散, 具有较强不确定性; 针对 Z10、Z11 和 Z12 指标云的第 4 周期结果而言, 其健康等级骤降, 且图像呈现大跨度形式, 表明该周期内的各指标回传数据分布异常。

4.2.2　综合云求解

根据式(26)～式(28)将对象层下的各指标云进行整合计算, 获得对应的部件综合云, 部件综合

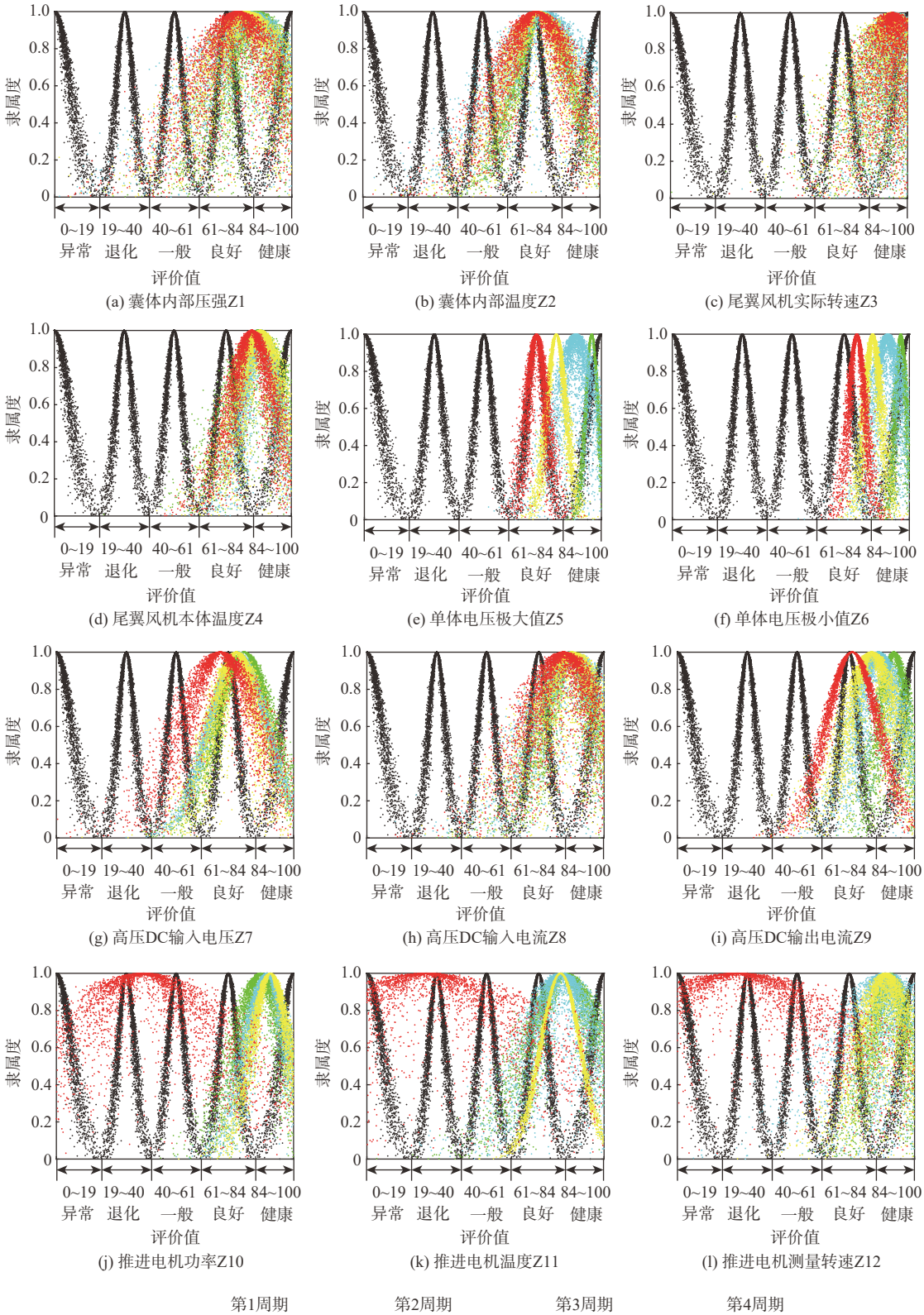


图 6 指标云

Fig. 6 Indicator cloud

云反映了此时部件的健康状态。将计算得到的对象层综合云叠加计算，可以获得平流层飞艇整体的综合云，平流层飞艇整体的综合云表明了此时平流层飞艇的健康状态。各部件综合云如图 7 所示。

部件综合云显示出各部件健康状态随时间的变化趋势，图中显示的变化情况分为 2 类，分别为健康状态随时间逐渐变差、随时间浮动变化。其中，囊体综合云的健康状态呈现由健康向良好轻微移动的趋势，这与其主动释放氦气、氦气泄漏及环境影响的结果一致；压力调节系统综合云图相对较

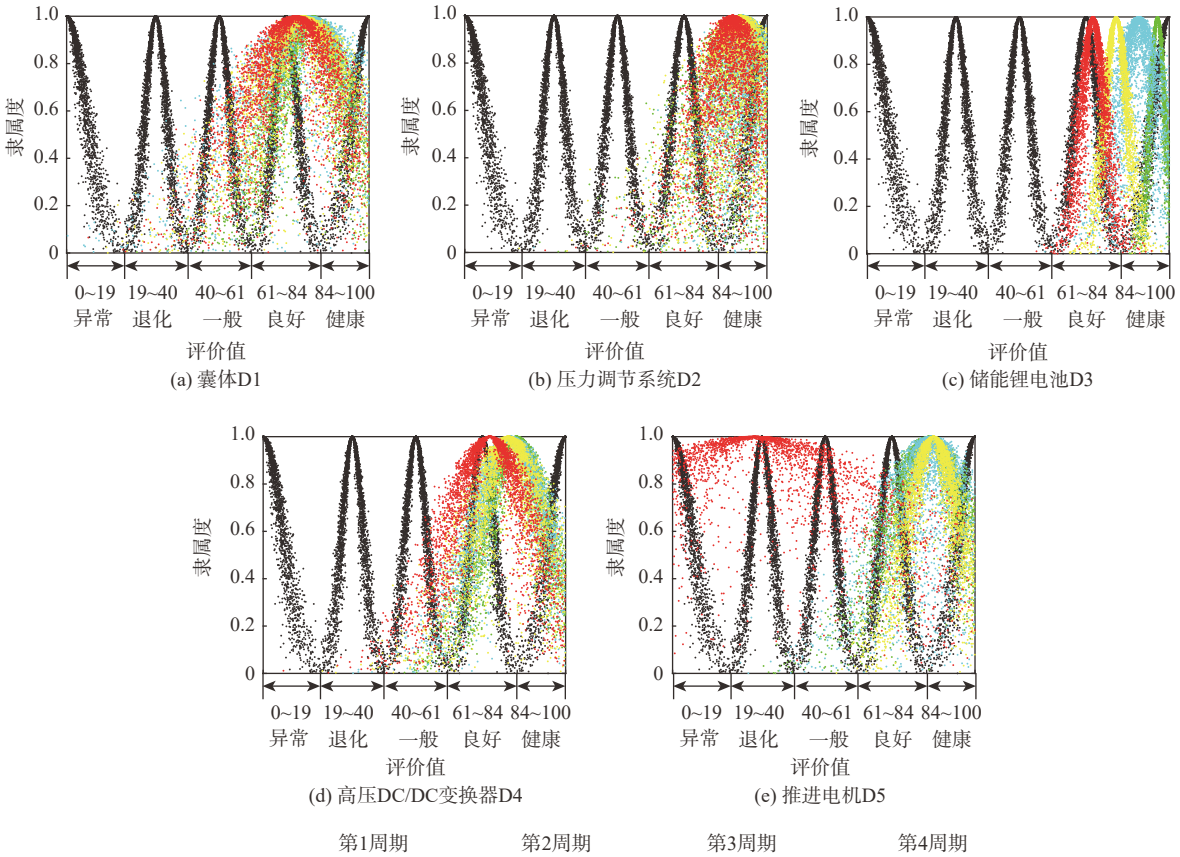


图 7　部件综合云

Fig. 7　Integrated cloud of components

为集中, 表明健康状态变化较小, 与其自身偶发故障居多且无故障运行时间相对较长的特点有关; 储能锂电池部件, 由于存在连续充放电过程, 存在损耗, 故其健康状态呈明显的下降趋势; 高压 DC/DC 变换器健康状态变化情况与其受外界输入电压电流和随时输出电压电流至艇载设备的结果一致; 在实际过程中, 由于推进电机于第 4 周期出现故障, 由推进电机各指标的叠加而生成的推进电机综合云也表现出退化状态, 与实际情况相符。

通过各部件综合云的叠加得到的平流层飞艇综合云如图 8 所示, 可以定性看出, 平流层飞艇前 3 个周期处于健康与良好之间的等级。第 4 周期受推进电机故障的影响较大, 平流层飞艇健康状态位于一般与退化之间, 需要地面人员及时做出反应并考虑采取降额飞行方式以解决问题。

单纯由图像获取的信息仅仅处于定性阶段, 而不能较好地定量比较。为了定量确定综合云的实际健康等级占比以及为避免综合云可能位于 2 个等级之间存在难以区分的情况, 利用云隶属度及相似度公式求解各综合云的云隶属度, 求解结果如表 6 所示, 结果表明, 前 3 个周期中, 平流层飞艇良好水平隶属度随周期变化为 [0.491,0.525,0.576]; 健康水平隶属度变化为 [0.475,0.404,0.369]。表明平

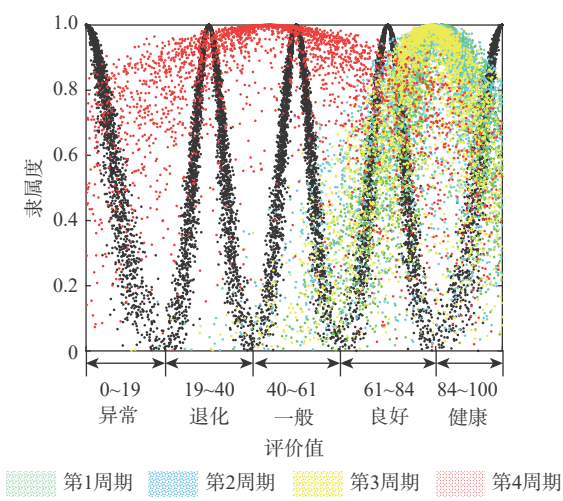


图 8　平流层飞艇综合云

Fig. 8　Integrated cloud of stratospheric airship

表 6　平流层飞艇综合云隶属度

Table 6　Membership degree of stratospheric airship integrated cloud

周期	隶属度				
	异常	退化	一般	良好	健康
1	0	0.001	0.033	0.491	0.475
2	0	0.004	0.067	0.525	0.404
3	0	0.002	0.053	0.576	0.369
4	0.124	0.271	0.335	0.187	0.083

流层飞艇随时间推移，自身健康水平逐渐降低，呈现健康转向良好的趋势。这与平流层飞艇的实际健康变化趋势一致。

第4周期时，推进电机下属的3个指标：推进电机功率、推进电机温度和推进电机测量转速的指标云图出现明显左移，由其组合而成的推进电机综合云也表明其于第4周期的健康状态处于退化与异常之间，说明急需改善该部件状态。平流层飞艇受推进电机影响，第4周期健康状态由良好(0.576)转向一般(0.335)，且退化占比0.271，异常占比0.124，隶属度重心发生转移。说明需注意平流层飞艇试验情况，调整部件状态，避免因部件故障导致试验任务终止。此外，退化与异常的占比也较高，这要求地面人员急需关注飞艇动态，以便做出对应操作。平流层飞艇整体的健康评估结果与实际试验中因推进电机故障导致平流层飞艇健康水平降低结果一致，该评估为地面人员及时采取安全控制措施提供了有效依据。

5 结 论

1）本文针对平流层飞艇构建了健康评估指标体系，体系包括1个目标层指标、5个对象层指标和12个指标层指标。

2）采用线性组合方式将主观权重与客观权重结合，避免了权重单一求解方法的缺陷。考虑平流层飞艇对象的特殊性，引入专家评判法，利用专家评价形式确定主观权重系数为0.38，客观权重系数为0.62，保证了该权重求解方法的适用性。

3）针对平流层飞艇引入专家评判确定了评价集和健康等级划分区间。根据飞艇两侧健康等级特性构建双侧半云的标准云模型。并以相对劣化度所映射的指标评价值生成指标云图，再各自结合生成部件及整体综合云。引入云隶属度定量求解各综合云的健康等级情况。由于评估模型可根据不同飞行任务合理修改等级划分、标准云模型及线性组合赋权系数等，故其适用性较强。

4）实例分析中，对原始数据的求解计算得到的指标云及由指标云叠加生成的综合云所表现的健康状态随时间变化的趋势与实际变化情况相符。经过实例分析得到的结果与实际试验过程中遭遇的情况对比验证表明，基于改进的组合赋权和云模型的平流层飞艇健康评估模型有效。

参考文献（References）

[1] 姚舒雨. 平流层飞艇的建模与控制研究[D]. 南京: 南京理工大学, 2019: 1-11.
YAO S Y. Modeling and control of stratospheric airship[D]. Nanjing: Nanjing University of Science and Technology, 2019: 1-11(in Chinese).

[2] 伏玉明, 刘伯鸿, 宋爽. 基于模糊综合评判的转辙机健康评估研究[J]. 铁道科学与工程学报, 2017, 14(5): 1070-1076.
FU Y M, LIU B H, SONG S. Research on health assessment methods for switch machine based on fuzzy comprehensive evaluation [J]. Journal of Railway Science and Engineering, 2017, 14(5): 1070-1076(in Chinese).

[3] ANDERSON M, SUCHKOV A, EINTHOVEN P, et al. Flight control system design risk assessment[C]//Guidance, Navigation, and Control Conference. Reston: AIAA, 1995: 3197.

[4] CHENG C H. Evaluating naval tactical missile systems by fuzzy AHP based on the grade value of membership function[J]. European Journal of Operational Research, 1997, 96(2): 343-350.

[5] 高焱, 夏晶晶, 耿纪莹, 等. 基于模糊综合评判法的寒区铁路隧道冻害评价体系研究[J]. 铁道标准设计, 2018, 62(7): 124-129.
GAO Y, XIA J J, GENG J Y, et al. Study on the evaluation system of freezing damage of railway tunnel in cold region based on fuzzy comprehensive evaluation method[J]. Railway Standard Design, 2018, 62(7): 124-129(in Chinese).

[6] AN M, CHEN Y, BAKER C J. A fuzzy reasoning and fuzzy-analytical hierarchy process based approach to the process of railway risk information: a railway risk management system[J]. Information Sciences, 2011, 181(18): 3946-3966.

[7] 温廷新, 苏焕博. 基于MICE_RF的组合赋权—极限随机树岩爆预测模型[J]. 黄金科学技术, 2022, 30(3): 392-403.
WEN T X, SU H B. Combined weighting-extremely randomized trees rockburst prediction model based on MICE_RF[J]. Gold Science and Technology, 2022, 30(3): 392-403(in Chinese).

[8] 李进友, 李媛, 冯冰, 等. 基于随机组合赋权模糊评价的风电机组健康状态评估[J]. 太阳能学报, 2022, 43(8): 340-351.
LI J Y, LI Y, FENG B, et al. Wind turbine health state assessment based on stochastic combination weighting fuzzy evaluation[J]. Acta Energiae Solaris Sinica, 2022, 43(8): 340-351(in Chinese).

[9] XU B S, QI N N, ZHOU J P, et al. Reliability assessment of highway bridges based on combined empowerment-TOPSIS method[J]. Sustainability, 2022, 14(13): 7793.

[10] GU Y K, ZHANG M, ZHOU X Q. Fault diagnosis of gearbox based on improved DUCG with combination weighting method[J]. IEEE Access, 2019, 7: 92955-92967.

[11] SHEN J, BAO X H, CHEN X S, et al. Prediction of tunnel earthquake damage based on a combination weighting analysis method [J]. Symmetry, 2022, 14(9): 1922.

[12] 董兴辉, 张鑫淼, 郑凯, 等. 基于组合赋权和云模型的风电机组健康状态评估[J]. 太阳能学报, 2018, 39(8): 2139-2146.
DONG X H, ZHANG X M, ZHENG K, et al. Health status assessment of wind turbine based on combination weighting and cloud model[J]. Acta Energiae Solaris Sinica, 2018, 39(8): 2139-2146(in Chinese).

[13] 周义蛟, 郭基联, 周舟. 基于云模型与组合赋权法的飞机保障性评估研究[C]//2017年首届航空保障设备发展论坛. 北京: 航空工业出版社, 2017: 142-148.
ZHOU Y J, GUO J L, ZHOU Z. A study on military aircraft supportability assessment based on cloud model and game theory

[C]//Proceedings of the 2017 1st Aviation Security Social Development Forum. Beijing: Aviation Industry Press, 2017: 142-148(in Chinese).

[14] 解宝琦, 李英顺, 王德彪, 等. 一种坦克炮长瞄准镜系统状态评估的方法[J]. 兵工学报, 2023, 44(8): 2414-2423.
XIE B Q, LI Y S, WANG D B, et al. Method for evaluating tank gunner's sight system[J]. Acta Armamentarii, 2023, 44(8): 2414-2423(in Chinese).

[15] 廖康君. 基于组合赋权和云模型的 H 物流园 EPC 项目投标风险评价研究[D]. 赣州: 江西理工大学, 2022.
LIAO K J. Research on bidding risk assessment of H logistics park EPC project based on combination weighting and cloud model[D]. Ganzhou: Jiangxi University of Science and Technology, 2022(in Chinese).

[16] 刘锦杰, 叶波, 罗赞琛, 等. 一种基于云模型的 SF_6 高压断路器状态评估方法[J]. 高压电器, 2019, 55(4): 31-38.
LIU J J, YE B, LUO Z C, et al. State assessment method for SF_6 high-voltage circuit breaker based on cloud model[J]. High Voltage Apparatus, 2019, 55(4): 31-38(in Chinese).

[17] 王芳. 基于云模型的高校教师满意度综合评价[J]. 技术与创新管理, 2009, 30(5): 636-639.
WANG F. A comprehensive evaluation method of the university teachers' job satisfaction based on the cloud model[J]. Technology and Innovation Management, 2009, 30(5): 636-639(in Chinese).

[18] 何涛, 马洁. 基于云模型和组合赋权法的 CBTC 系统可靠性评价[J]. 重庆大学学报, 2023, 46(2): 130-139.
HE T, MA J. Reliability evaluation of CBTC based on cloud model and combination weighting method[J]. Journal of Chongqing University, 2023, 46(2): 130-139(in Chinese).

[19] 龚艳冰, 蒋亚东, 梁雪春. 基于模糊贴近度的正态云模型相似度度量[J]. 系统工程, 2015, 33(9): 133-137.
GONG Y B, JIANG Y D, LIANG X C. Similarity measurement for normal cloud models based on fuzzy similarity measure[J]. Systems Engineering, 2015, 33(9): 133-137(in Chinese).

Health status assessment of stratospheric airship based on improved combination weighting and cloud model

ZENG Shengfu[1]，LIANG Haoquan[2]，WANG Yan[1]，SUN Kangwen[1,*]，ZHONG Zihong[1]

(1. School of Aeronautic Science and Engineering，Beihang University，Beijing 100191，China；
2. Institute of Unmanned Systems，Beihang University，Beijing 100191，China)

Abstract: In the context of stratospheric airship health assessment, this paper introduces an innovative methodology that combines composite weighting with a cloud model and incorporates a expert evaluation system. The objective is to provide comprehensive technical support for the evaluation of stratospheric airship health status. First, by examining possible hazards and malfunctions in the different stratospheric airship subsystems, an evaluation criterion system is created. Subsequently, composite weights are determined for these criteria through a combination of a expert evaluation system and the linear composite weighting method. Based on the defined evaluation set and standard cloud model specific to stratospheric airships, cloud models are generated for each indicator and the overall health assessment. Cloud similarity and membership degree computations are used to quantitatively evaluate each indicator's health. Lastly, a case study featuring a stratospheric airship trial is used to validate the methodology. Finally, the methodology is validated through a case study involving a trial of a stratospheric airship. The results demonstrate that the proposed approach, improved combination weighting and cloud model, effectively support the assessment of stratospheric airship health status. Furthermore, the incorporation of a expert evaluation system enhances the model's applicability to the unique characteristics of stratospheric airships, thus providing a solid foundation for subsequent maintenance and safety control tasks in the field of stratospheric airship operations.

Keywords: stratospheric airship; expert evaluation system; combination weighting; cloud model; health status assessment

Received：2023-11-02；**Accepted**：2024-01-12；**Published Online**：2024-01-23 17：25

URL：link.cnki.net/urlid/11.2625.V.20240123.0945.001

Foundation item：National Natural Science Foundation of China (52302511)

* **Corresponding author**. E-mail：09142@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0732

有控自旋卷弧翼弹箭角运动的非线性吸引域

周文涛[1]，常思江[1,*]，展鹏遥[1]，崔慧振[2]

(1. 南京理工大学 能源与动力工程学院，南京 210094；2. 西北工业集团有限公司设计二所，西安 710043)

摘 要：针对有控自旋卷弧翼弹箭在鸭舵控制力与非线性气动力作用下的稳定性问题，开展非线性角运动特性研究。根据该类弹箭的弹道和结构特点，建立七自由度刚体运动模型，据此推导状态空间形式的非线性角运动方程；选取弹道特征点，利用数值仿真和多项式拟合方法得到对应的非线性气动力系数，采用基于平方和规划的吸引域估计方法，定量分析鸭舵控制和非线性气动力系数对该类弹箭稳定域的影响。理论分析及仿真结果表明：鸭舵控制力大小对弹箭角运动的稳定吸引域影响较大，对于给定算例，当舵偏角为20°时，与无控状态相比，吸引域边界减小约16.9%，而控制力方位对其影响却很小；升力系数导数、静力矩系数导数和前、后体马格努斯力矩系数导数的非线性项是影响该类弹箭角运动稳定吸引域的主要因素，且后体非线性马格努斯力矩对吸引域的影响明显大于前体。

关 键 词：自旋卷弧翼；弹箭；角运动；平方和规划；非线性气动力；吸引域

中图分类号：V212

文献标志码：A **文章编号**：1001-5965(2026)01-0338-14

与平直翼相比，卷弧翼可以包覆在弹体外表面，有利于弹箭发射并节省弹内空间；同时，由于卷弧尾翼在零攻角下能产生自诱导滚转力矩，可用于实现弹体自旋稳定，因此，被广泛应用于滚转导弹[1-2]。然而，从气动特性角度，卷弧翼产生的自诱导滚转力矩在亚、跨声速下会发生变号，导致弹体滚转换向，有可能产生显著的侧向力和偏航力矩，从而引起较为复杂的弹体运动[2]。目前，国内外学者已在气动特性[3-5]、动力学特性[2,6]等方面对卷弧翼弹箭开展了深入研究。其中，Morote[7]发现，与固连于弹体的尾翼相比，与弹体形成差动滚转的自旋尾翼能够有效解决弹箭飞行中的共振不稳定问题，并在此基础上对无控自旋卷弧翼弹箭的共振闭锁[2]开展了研究。上述研究都是基于线性化理论开展，而弹箭在各种实际条件下的运动总是具有不同程度的非线性[8]，作用在弹箭上的气动力和力矩系数一般为攻角 δ 的非线性函数。对于采用卷弧翼的导弹或制导炮弹，控制机构作用也将引起弹体动力学响应，当其与卷弧翼、非线性等因素耦合时，有可能产生更为复杂的非线性运动，甚至影响弹箭的飞行稳定性。对此，本文针对一类鸭舵控制的自旋卷弧翼弹箭开展非线性角运动分析，以探明该类弹箭在上述多重因素作用下的运动规律。

自20世纪70年代以来，人们围绕弹箭非线性运动开展了大量研究。Murphy[9]针对有控弹非线性稳定性问题进行了研究；Morote等针对尾翼弹的灾变性偏航[10]、转速闭锁[11]、非线性滚转运动[12]等开展了一系列深入研究。研究结果表明，弹箭非线性运动主要由气动非线性引起[13]，故获取弹箭非线性气动力数据变得十分重要。McCoy[14]通过靶道试验和参数辨识，获取了大量弹箭的非线性气动数据；蒋增辉等[15]采用分段拟合技术获得了弹箭在小迎角高超声速下的非线性气动导数；钱龙等[16]利用数值仿真和多项式拟合方法获得了固定鸭舵双旋

收稿日期：2023-11-07；**录用日期**：2024-01-21；**网络出版时间**：2024-02-29 08:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20240228.1632.002
基金项目：国家自然科学基金 (52202475)
* **通信作者**. E-mail：ballistics@126.com

引用格式：周文涛，常思江，展鹏遥，等. 有控自旋卷弧翼弹箭角运动的非线性吸引域 [J]. 北京航空航天大学学报，2026，52（1）：338-351. ZHOU W T，CHANG S J，ZHAN P Y，et al. Nonlinear region of attraction of angular motion for a controlled projectile with self-rotating wraparound fins [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：338-351 （in Chinese）.

弹的非线性气动系数。

弹箭非线性运动稳定性除了与气动系数等系统参数有关外, 还与初始条件密切相关。早期, 关于弹箭非线性运动的分析, 有学者采用椭圆函数和奇异摄动法[17]求其解析解, 但仅限于弱非线性问题; Chang[18]、李东阳[19]等分别采用同伦分析法和正规形法求取攻角的高精度解析解; 杨树兴等[20]采用相平面分析法对旋转弹锥形运动及动态稳定性进行了研究; 钟扬威[21]、杨杰[22]、赵新新[23]等分别利用分岔理论对弹箭非线性角运动稳定性开展了相关研究。上述研究从一定程度上推动了弹箭非线性运动理论的发展。然而, 在进一步回答“在何种初始条件下弹箭角运动能收敛至平衡点”“初始条件超出什么范围时弹箭会失稳”等问题时, 上述研究却难以给出较为准确的定量结果, 这本质上涉及弹箭角运动的非线性吸引域问题, 也是实际工程应用中人们非常关注的问题。目前, 处理该问题的方法主要有平方和规划方法[24]、Zabov 方法[25]、轨迹逆向积分方法[26]等。相比而言, 基于平方和规划的吸引域估计方法能够有效地定量处理非线性系统的吸引域问题, 该方法目前已被广泛应用于生物[27]、航天[28-29]等领域, 李东阳等[30]将其初步应用于弹箭非线性角运动的稳定性分析, 取得一定效果。

如上所述, 现有文献针对弹箭非线性运动特性的研究, 关于五次方及以上高阶气动力系数对角运动特性影响的研究相对较少, 且绝大部分为定性分析, 未能给出定量结果, 不利于指导弹箭的具体设计。特别是对于本文拟研究的有控自旋卷弧翼弹箭, 关于高阶非线性气动力及控制条件如何影响其角运动稳定性这一重要问题的研究, 目前尚十分缺乏。对此, 本文拟引入有效的定量分析工具, 深入探讨控制力和非线性气动力系数(三次方和/或五次方项)对弹箭角运动吸引域的影响, 并给出定量的分析结果, 以期为该类弹箭的方案设计及工程研制提供必要的理论依据。

因此, 本文首先根据鸭舵控制自旋卷弧翼弹箭的特点, 建立七自由度刚体动力学模型, 推导涉及非线性气动力的角运动方程; 其次, 选取弹道上的典型特征点, 利用平方和规划方法开展角运动的非线性吸引域估计; 最后, 通过计算吸引域边界, 定量分析鸭舵控制条件、诸非线性气动力系数等对自旋卷弧翼弹箭角运动稳定吸引域的影响。

1　动力学建模

1.1　物理模型及工作过程

本文采用的弹箭模型基于 TTCP 卷弧翼模型。在 TTCP 模型的基础上, 本文模型做了如下修改: ①卷弧翼作为一个单独组件和弹身成为前、后两体结构, 两者采用滚动轴承联接, 卷弧翼与弹身可实现差动滚转, 即所谓“自旋卷弧翼”; ②弹身头部增加了一对鸭舵, 用于该弹箭的飞行弹道控制。本文模型如图 1 所示。

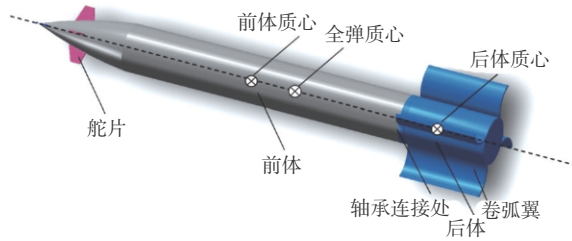


图 1　有控自旋卷弧翼弹箭模型示意图

Fig. 1　The model diagram of controlled projectile with self-rotating wraparound fins

发射前, 鸭舵折叠在前体头部的控制舱内。当弹箭出炮口后, 后体卷弧翼在自诱导滚转力矩的作用下旋转, 前、后体连接轴承产生的黏性阻尼力矩与滚动摩擦力矩带动前体低速旋转, 从而实现差动滚转。当弹箭飞行进入预定区域后, 舵面从控制仓内展开, 通过一定规律的舵面偏转实现弹道控制。

如上所述, 由于弹箭前、后体之间存在差动滚转, 为了较为精确地建模, 有必要在传统六自由度刚体动力学模型的基础上增加一个两体差动滚转的自由度, 即建立七自由度刚体动力学模型。

1.2　坐标系定义

为便于建模, 除了外弹道学[31]中常用的弹道坐标系 $Ox_2y_2z_2$、弹轴坐标系 $O\xi\eta\zeta$ 及弹体坐标系 $Ox_1y_1z_1$ 外, 还需分别建立前体坐标系 $Ox_1y_{1F}z_{1F}$ 和后体坐标系 $Ox_1y_{1B}z_{1B}$, 如图 2 所示。

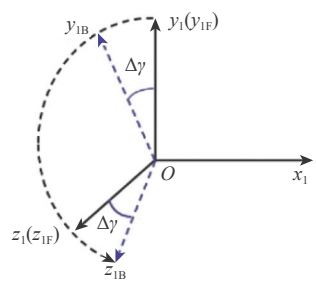


图 2　前体坐标系和后体坐标系

Fig. 2　Precursor coordinate system and postbody coordinate system

前体坐标系 $Ox_1y_{1F}z_{1F}$ 与弹体坐标系 $Ox_1y_1z_1$ 重合, 后体坐标系 $Ox_1y_{1B}z_{1B}$ 是在前体坐标系基础上将 Oy_{1F} 轴绕 Ox_1 轴转过角度 $\Delta\gamma$ 所得, $\Delta\gamma$ 为差动滚转角, 其定义为

$$\Delta\gamma = \gamma_B - \gamma_F \tag{1}$$

式中：γ_{F}、γ_{B}分别为前、后体的滚转角，其中，γ_{F}与弹体滚转角γ重合。

1.3 空气动力和力矩模型

弹箭飞行时在风场中所受空气动力及力矩的大小和方向取决于弹箭相对于空气的速度$\boldsymbol{v}_{\mathrm{r}}$，当考虑风速$\boldsymbol{w}$时，弹箭的飞行速度$\boldsymbol{v}$与$\boldsymbol{v}_{\mathrm{r}}$的关系为

$$\boldsymbol{v}_{\mathrm{r}}=\boldsymbol{v}-\boldsymbol{w} \tag{2}$$

1.3.1 作用于弹体的空气动力

忽略影响较小的马格努斯力，作用于弹体的空气动力主要为阻力$\boldsymbol{F}_x$、升力$\boldsymbol{F}_y$和重力$\boldsymbol{G}$，其矢量表达式为

$$\begin{cases}\boldsymbol{F}_x=\dfrac{1}{2}\rho S c_x|\boldsymbol{v}_{\mathrm{r}}|(-\boldsymbol{v}_{\mathrm{r}})\\ \boldsymbol{F}_y=\dfrac{1}{2}\rho S c_y'\delta\boldsymbol{v}_{\mathrm{r}}\times(\boldsymbol{v}_{\mathrm{r}}\times\boldsymbol{\xi})\\ \boldsymbol{G}=m\boldsymbol{g}\end{cases} \tag{3}$$

式中：ρ为空气密度；$S=\pi d^2/4$为特征面积，d为弹径；c_x为弹体阻力系数；δ为攻角；c_y'为弹体升力系数对δ的导数；$\vec{\boldsymbol{\xi}}$为弹轴坐标系$O\xi$轴上的单位向量；m为弹丸质量；$\boldsymbol{g}$为重力加速度矢量。

非线性阻力系数和升力系数导数分别表示为

$$\begin{cases}c_x=c_{x_0}+c_{x_2}\delta^2\\ c_y'=c_{y_0}+c_{y_3}\delta^2+c_{y_5}\delta^4\end{cases} \tag{4}$$

式中：c_{x_0}和c_{x_2}分别为阻力系数的线性项和平方项；c_{y_0}、c_{y_3}和c_{y_5}分别为升力系数导数的线性项、三次方项和五次方项。

1.3.2 鸭舵控制力

在前体坐标系中，鸭舵控制力可表示为

$$\begin{cases}F_{\mathrm{c},x_1}=-\dfrac{1}{2}\rho S|\boldsymbol{v}_{\mathrm{r}}^2|c_{\mathrm{c},x}\\ F_{\mathrm{c},y_1}=\dfrac{1}{2}\rho S|\boldsymbol{v}_{\mathrm{r}}^2|c_{\mathrm{c},y}'\delta_{\mathrm{c}}\end{cases} \tag{5}$$

式中：F_{c,x_1}和F_{c,y_1}分别为舵面偏转所产生的阻力和升力；δ_{c}为舵偏角；$c_{\mathrm{c},y}'$为鸭舵升力系数导数；$c_{\mathrm{c},x}=c_{\mathrm{c},x_0}(1+\sigma\delta_{\mathrm{c}}^2)$为鸭舵阻力系数，$c_{\mathrm{c},x_0}$为鸭舵零升阻力系数，$\sigma$为攻角系数。

利用坐标转换关系，将式(5)投影到速度坐标系中，可得鸭舵控制力在速度坐标系中的表达式为

$$\begin{cases}F_{\mathrm{c},x_2}=-F_{\mathrm{c},y_1}(\sin\delta_1\cos\gamma_{\mathrm{F}}+\sin\gamma_{\mathrm{F}}\cos\delta_1\sin\delta_2)+\\ \qquad F_{\mathrm{c},x_1}\cos\delta_1\cos\delta_2\\ F_{\mathrm{c},y_2}=F_{\mathrm{c},y_1}(\cos\delta_1\cos\gamma_{\mathrm{F}}-\sin\gamma_{\mathrm{F}}\sin\delta_1\sin\delta_2)+\\ \qquad F_{\mathrm{c},x_1}\sin\delta_1\cos\delta_2\\ F_{\mathrm{c},z_2}=F_{\mathrm{c},x_1}\sin\delta_2+F_{\mathrm{c},y_1}\sin\gamma_{\mathrm{F}}\cos\delta_2\end{cases} \tag{6}$$

式中：δ_1和δ_2分别为弹体的高低攻角和方向攻角；F_{c,x_2}、F_{c,y_2}、F_{c,z_2}分别为鸭舵控制力在速度坐标系中的投影分量。

1.3.3 作用于弹体的空气动力矩

不考虑气动偏心和动不平衡，有控自旋卷弧翼弹箭在飞行过程中，前、后体受到的空气动力矩主要包括静力矩(前体静力矩$\boldsymbol{M}_{\mathrm{F},z}$和后体静力矩$\boldsymbol{M}_{\mathrm{B},z}$)、赤道阻尼力矩(前体赤道阻尼力矩$\boldsymbol{M}_{\mathrm{F},zz}$和后体赤道阻尼力矩$\boldsymbol{M}_{\mathrm{B},zz}$)、极阻尼力矩(前体极阻尼力矩$\boldsymbol{M}_{\mathrm{F},xz}$和后体极阻尼力矩$\boldsymbol{M}_{\mathrm{B},xz}$)、尾翼诱导滚转力矩$\boldsymbol{M}_{xw}$及马格努斯力矩(前体马格努斯力矩$\boldsymbol{M}_{\mathrm{F},y}$和后体马格努斯力矩$\boldsymbol{M}_{\mathrm{B},y}$)，其矢量表达式分别为

$$\begin{cases}\boldsymbol{M}_{\mathrm{F},z}=\dfrac{1}{2}\rho S l|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{F},z}'(\boldsymbol{v}_{\mathrm{r}}\times\boldsymbol{\xi})\\ \boldsymbol{M}_{\mathrm{F},zz}=-\dfrac{1}{2}\rho S ld|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{F},zz}'\boldsymbol{\omega}_1\\ \boldsymbol{M}_{\mathrm{F},xz}=-\dfrac{1}{2}\rho S ld|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{F},xz}'\omega_{\mathrm{F},\xi}\vec{\boldsymbol{\xi}}\\ \boldsymbol{M}_{\mathrm{F},y}=\dfrac{1}{2}\rho S ld\omega_{\mathrm{F},\xi}m_{\mathrm{F},y}''\vec{\boldsymbol{\xi}}\times(\boldsymbol{v}_{\mathrm{r}}\times\vec{\boldsymbol{\xi}})\\ \boldsymbol{M}_{\mathrm{B},z}=\dfrac{1}{2}\rho S l|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{B},z}'(\boldsymbol{v}_{\mathrm{r}}\times\vec{\boldsymbol{\xi}})\\ \boldsymbol{M}_{\mathrm{B},zz}=-\dfrac{1}{2}\rho S ld|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{B},zz}'\boldsymbol{\omega}_1\\ \boldsymbol{M}_{\mathrm{B},xz}=-\dfrac{1}{2}\rho S ld|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{B},xz}'\omega_{\mathrm{B},\xi}\vec{\boldsymbol{\xi}}\\ \boldsymbol{M}_{\mathrm{B},y}=\dfrac{1}{2}\rho S l|\boldsymbol{v}_{\mathrm{r}}|m_{\mathrm{B},y}'\vec{\boldsymbol{\xi}}(\boldsymbol{v}_{\mathrm{r}}\times\vec{\boldsymbol{\xi}})\\ \boldsymbol{M}_{xw}=\dfrac{1}{2}\rho S l|\boldsymbol{v}_{\mathrm{r}}^2|m_{xw}\vec{\boldsymbol{\xi}}\end{cases} \tag{7}$$

式中：l为特征长度；d为弹径；$\boldsymbol{\omega}_1$为弹箭摆动角速度；$\omega_{\mathrm{F},\xi}\approx\dot{\gamma}_{\mathrm{F}}$和$\omega_{\mathrm{B},\xi}\approx\dot{\gamma}_{\mathrm{B}}$分别为前、后体滚转角速度；$m_{\mathrm{F},z}'$、$m_{\mathrm{B},z}'$分别为前、后体静力矩系数对$\delta$的导数，由于其与前、后体转速无关，可合并为全弹的静力矩系数导数$m_z'=m_{\mathrm{F},z}'+m_{\mathrm{B},z}'$；$m_{\mathrm{F},zz}'$、$m_{\mathrm{B},zz}'$分别为前、后体赤道阻尼力矩系数对$d|\omega_1|/|\boldsymbol{v}_{\mathrm{r}}|$的导数，同样可合并为全弹的赤道阻尼力矩系数导数$m_{zz}'=m_{\mathrm{F},zz}'+m_{\mathrm{B},zz}'$；$m_{\mathrm{F},xz}'$为前体极阻尼力矩系数对$\dot{\gamma}_{\mathrm{F}}d/|\boldsymbol{v}_{\mathrm{r}}|$的导数；$m_{\mathrm{B},xz}'$为后体极阻尼力矩系数对$\dot{\gamma}_{\mathrm{B}}d/|\boldsymbol{v}_{\mathrm{r}}|$的导数；$m_{\mathrm{F},y}''$为前体马格努斯力矩系数对$\dot{\gamma}_{\mathrm{F}}d/|\boldsymbol{v}_{\mathrm{r}}|$和$\delta$的联合导数；$m_{\mathrm{B},y}'$为后体马格努斯力矩系数对$\delta$的导数；$m_{xw}$为卷弧尾翼诱导滚转力矩系数。

上述各非线性力矩系数导数可分别表示为

$$\begin{cases}m_z'=m_{z_0}+m_{z_3}\delta^2+m_{z_5}\delta^4\\ m_{zz}'=m_{zz_0}+m_{zz_3}\delta^2+m_{zz_5}\delta^4\\ m_{\mathrm{F},xz}'=m_{\mathrm{F},xz_0}+m_{\mathrm{F},xz_3}\delta^2+m_{\mathrm{F},xz_5}\delta^4\\ m_{\mathrm{B},xz}'=m_{\mathrm{B},xz_0}+m_{\mathrm{B},xz_3}\delta^2+m_{\mathrm{B},xz_5}\delta^4\\ m_{\mathrm{F},y}''=m_{\mathrm{F},y_0}+m_{\mathrm{F},y_3}\delta^2+m_{\mathrm{F},y_5}\delta^4\\ m_{\mathrm{B},y}'=m_{\mathrm{B},y_0}+m_{\mathrm{B},y_3}\delta^2+m_{\mathrm{B},y_5}\delta^4\end{cases} \tag{8}$$

式中：m_{z_0}、m_{zz_0}、m_{F,xz_0}、m_{B,xz_0}、m_{F,y_0}、m_{B,y_0}分别为各力矩系数导数的线性项；m_{z_3}、m_{zz_3}、m_{F,xz_3}、m_{B,xz_3}、m_{F,y_3}、m_{B,y_3}分别为各力矩系数导数的三次方项；m_{z_5}、m_{zz_5}、m_{F,xz_5}、m_{B,xz_5}、m_{F,y_5}、m_{B,y_5}分别为各力矩系数导数的五次方项。

1.3.4　鸭舵控制力矩

鸭舵控制力矩$\boldsymbol{M}_{\mathrm{c}}$在弹轴坐标系中可表示为

$$\begin{cases} M_{\mathrm{c},\xi}=0 \\ M_{\mathrm{c},\eta}=-F_{\mathrm{c},y_1}l_{\mathrm{D}}\sin\gamma_{\mathrm{F}} \\ M_{\mathrm{c},\zeta}=F_{\mathrm{c},y_1}l_{\mathrm{D}}\cos\gamma_{\mathrm{F}} \end{cases} \tag{9}$$

式中: l_{D}为舵面压心到全弹质心的距离。

舵面滚转阻尼力矩为

$$\boldsymbol{M}_{\mathrm{c},xz}=-\frac{1}{2}\rho S\,ld|\boldsymbol{v}_{\mathrm{r}}|m'_{\mathrm{c},xz}\omega_{\mathrm{F},\xi}\vec{\boldsymbol{\xi}} \tag{10}$$

式中: $m'_{\mathrm{c},xz}$为舵面滚转阻尼力矩系数导数。

1.3.5　滚转约束力矩

为便于分析, 将前、后体连接轴承产生的黏性阻尼力矩与滚动摩擦力矩折算为约束力矩M_{V}, 即

$$M_{\mathrm{V}}=c_{\mathrm{v}}(\dot\gamma_{\mathrm{F}}-\dot\gamma_{\mathrm{B}}) \tag{11}$$

式中: c_{v}为滚转约束力矩系数; $\dot\gamma_{\mathrm{F}}$、$\dot\gamma_{\mathrm{B}}$分别为前、后体的转速。

1.4　七自由度刚体动力学模型

由于自旋卷弧翼弹箭的前、后体之间存在差动滚转, 为研究其弹道特性并开展非线性角运动分析, 本节根据上述非线性空气动力和力矩表达式, 建立七自由度非线性刚体动力学模型。

在速度坐标系中建立质心运动的动力学方程组为

$$\begin{cases} \dot v=\left(F_{x_2}+kF_{\mathrm{c},x_2}\right)/m \\ \dot\theta_{\mathrm{a}}=\left(F_{y_2}+kF_{\mathrm{c},y_2}\right)/(mv\cos\psi_2) \\ \dot\psi_2=\left(F_{z_2}+kF_{\mathrm{c},z_2}\right)/(mv) \end{cases} \tag{12}$$

式中: "$\cdot$" 表示对时间 t 的导数; m为弹体质量; v为速度; θ_{a}为弹道倾角; ψ_2为弹道偏角; F_{x_2}、F_{y_2}、F_{z_2}为除控制力外的合外力在速度坐标系中的分量; 弹箭无控飞行时 k=0, 有控飞行时 k=1。

弹箭质心运动的运动学方程与普通尾翼弹质心运动学方程[31]形式相同, 此处不再赘述。

弹箭前、后体的绕心运动动力学方程组为

$$\begin{cases} \dot\omega_{\mathrm{F},\xi}=\left(M_{\mathrm{F},\xi}+M_{\mathrm{V}}\right)/C_{\mathrm{F}} \\ \dot\omega_{\mathrm{B},\xi}=\left(M_{\mathrm{B},\xi}-M_{\mathrm{V}}\right)/C_{\mathrm{B}} \\ \dot\omega_\eta=M_1/A \\ \dot\omega_\zeta=M_2/A \end{cases} \tag{13}$$

其中:

$$\begin{cases} A=A_{\mathrm{F}}+m_{\mathrm{F}}r_{\mathrm{F}}^2+A_{\mathrm{B}}+m_{\mathrm{B}}r_{\mathrm{B}}^2 \\ M_1=-\omega_\zeta(C_{\mathrm{F}}\omega_{\mathrm{F},\xi}+C_{\mathrm{B}}\omega_{\mathrm{B},\xi}-\omega_\zeta\tan\varphi_2A)+M_{\mathrm{F},\eta}+M_{\mathrm{B},\eta} \\ M_2=\omega_\eta(C_{\mathrm{F}}\omega_{\mathrm{F},\xi}+C_{\mathrm{B}}\omega_{\mathrm{B},\xi}-\omega_\zeta\tan\varphi_2A)+M_{\mathrm{F},\zeta}+M_{\mathrm{B},\zeta} \end{cases}$$

式中: $M_{\mathrm{F},\xi}$、$M_{\mathrm{F},\eta}$、$M_{\mathrm{F},\zeta}$分别为前体所受合外力矩在弹轴坐标系中的分量; $M_{\mathrm{B},\xi}$、$M_{\mathrm{B},\eta}$、$M_{\mathrm{B},\zeta}$为后体所受合外力矩在弹轴坐标系中的分量; ω_η、ω_ζ分别为弹箭角速度在弹轴坐标系 $O\eta$轴、$O\zeta$轴上的分量; A_{F}、A_{B}分别为前、后体赤道转动惯量; C_{F}、C_{B}分别为前、后体极转动惯量; m_{F}、m_{B}分别为前、后体的质量; r_{F}、r_{B}分别为前、后体质心到全弹质心的距离。

弹箭前、后体的绕心运动学方程组为

$$\begin{cases} \dot\gamma_{\mathrm{F}}=\omega_{\mathrm{F},\xi}-\omega_\zeta\tan\varphi_2 \\ \dot\gamma_{\mathrm{B}}=\omega_{\mathrm{B},\xi}-\omega_\zeta\tan\varphi_2 \\ \dot\varphi_{\mathrm{a}}=\omega_\zeta/\cos\varphi_2 \\ \dot\varphi_2=-\omega_\eta \end{cases} \tag{14}$$

式中: φ_{a}和φ_2分别为弹轴高低角和弹轴方位角。

2　非线性角运动方程

为便于研究, 假设有控自旋卷弧翼弹箭的七自由度刚体弹道与理想弹道偏差较小, 则弹箭的质心运动满足如下理想弹道方程组:

$$\begin{cases} \dot v_{\mathrm{i}}=-b_xv_{\mathrm{i}}^2-kb_{\mathrm{c},x}v_{\mathrm{i}}^2-|\boldsymbol{g}|\sin\theta_{\mathrm{i}} \\ \dot\theta_{\mathrm{i}}=-|\boldsymbol{g}|\cos\theta_{\mathrm{i}}/v_{\mathrm{i}} \\ \dot x_{\mathrm{i}}=v_{\mathrm{i}}\cos\theta_{\mathrm{i}},y_{\mathrm{i}}=v_{\mathrm{i}}\sin\theta_{\mathrm{i}} \end{cases} \tag{15}$$

式中: θ_{i}为理想弹道倾角, 后文简化表示为θ。

根据文献 [31] 所述方法推导本文研究对象的角运动方程, 该方法有效避免了旋转弹箭高低攻角和方向攻角的耦合问题。因此, 近似认为高低攻角$\delta_1\approx\varphi_{\mathrm{a}}-\theta_{\mathrm{a}}$, 方向攻角$\delta_2\approx\varphi_2-\psi_2$, 则弹箭攻角运动方程可表示为

$$\begin{cases} \dot\delta_1=\omega_\zeta/\cos\varphi_2-\left(F_{x_2}+kF_{\mathrm{c},x_2}\right)/(mv\cos\psi_2) \\ \dot\delta_2=-\omega_\eta-\left(F_{z_2}+kF_{\mathrm{c},z_2}\right)/(mv) \\ \dot\omega_\eta=M_1/A \\ \dot\omega_\zeta=M_2/A \end{cases} \tag{16}$$

为便于表示, 引入如下组合参数: $b_x=\dfrac{\rho S}{2m}c_x$, $b_{\mathrm{c},x}=\dfrac{\rho S}{2m}c_{\mathrm{c},x}$, $b_y=\dfrac{\rho S}{2m}c'_y$, $b_{\mathrm{c},y}=\dfrac{\rho S}{2m}c'_{\mathrm{c},y}$, $k_{\mathrm{c}}=\dfrac{\rho Sl_{\mathrm{D}}}{2A}c'_{\mathrm{c},y}$, $k_z=\dfrac{\rho Sl}{2A}m'_z$, $k_{zz}=\dfrac{\rho Sld}{2A}m'_{zz}$, $k_{\mathrm{F},y}=\dfrac{\rho Sld}{2A}m''_{\mathrm{F},y}$, $k_{\mathrm{B},y}=\dfrac{\rho Sl}{2A}m'_{\mathrm{B},y}$。

将第 1.3 节中的力和力矩表达式代入方程组(16), 经整理可得

$$\begin{cases} \dot\delta_1=\omega_\zeta-v\left[b_y\delta_1+\dfrac{|\boldsymbol{g}|}{v^2}\cos\theta_{\mathrm{a}}-kb_{\mathrm{c},y}\delta_{\mathrm{c}}\cos\gamma_{\mathrm{F}}-(b_x+b_y)\dfrac{w_{y_2}}{v}\right] \\ \dot\delta_2=-\omega_\eta-v\left[b_y\delta_2-\dfrac{|\boldsymbol{g}|}{v^2}\sin\theta\psi_2-kb_{\mathrm{c},y}\delta_{\mathrm{c}}\sin\gamma_{\mathrm{F}}-(b_x+b_y)\dfrac{w_{z_2}}{v}\right] \\ \dot\omega_\eta=-\left(\dfrac{C_{\mathrm{F}}}{A}\dot\gamma_{\mathrm{F}}+\dfrac{C_{\mathrm{B}}}{A}\dot\gamma_{\mathrm{B}}\right)\omega_\zeta+v^2\left[-k_z\delta_2\left(\dfrac{k_{\mathrm{F},y}\dot\gamma_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\delta_1-\right. \\ \qquad\left.\dfrac{k_{zz}}{v}\omega_\eta-kk_{\mathrm{c}}\delta_{\mathrm{c}}\sin\gamma_{\mathrm{F}}-k_z\dfrac{w_{z_2}}{v}+\left(\dfrac{k_{\mathrm{F},y}\dot\gamma_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\dfrac{w_{y_2}}{v}\right] \\ \dot\omega_\zeta=\left(\dfrac{C_{\mathrm{F}}}{A}\dot\gamma_{\mathrm{F}}+\dfrac{C_{\mathrm{B}}}{A}\dot\gamma_{\mathrm{B}}\right)\omega_\eta+v^2\left[k_z\delta_1\left(\dfrac{k_{\mathrm{F},y}\dot\gamma_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\delta_2-\right. \\ \qquad\left.\dfrac{k_{zz}}{v}\omega_\zeta+kk_{\mathrm{c}}\delta_{\mathrm{c}}\cos\gamma_{\mathrm{F}}k_z\dfrac{w_{y_2}}{v}+\left(\dfrac{k_{\mathrm{F},y}\dot\gamma_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\dfrac{w_{z_2}}{v}\right] \end{cases} \tag{17}$$

将方程组(17)的第1式和第2式写成矩阵形式为

$$\begin{bmatrix}\omega_\eta\\ \omega_\zeta\end{bmatrix}=\boldsymbol{B}_1\begin{bmatrix}\dot{\delta}_1\\ \dot{\delta}_2\end{bmatrix}+\boldsymbol{B}_2\begin{bmatrix}\delta_1\\ \delta_2\end{bmatrix}+\boldsymbol{b}+\boldsymbol{w}_1 \tag{18}$$

式中：$\boldsymbol{B}_1=\begin{bmatrix}0&-1\\1&0\end{bmatrix}$；$\boldsymbol{B}_2=v\begin{bmatrix}0&-b_y\\b_y&0\end{bmatrix}$；$\boldsymbol{b}=v\begin{bmatrix}\dfrac{|\boldsymbol{g}|}{v^2}\sin\theta\psi_2+kb_{\mathrm{c},y}\delta_{\mathrm{c}}\sin\gamma_{\mathrm{F}}\\ \dfrac{|\boldsymbol{g}|}{v^2}\cos\theta_{\mathrm{a}}-kb_{\mathrm{c},y}\delta_{\mathrm{c}}\cos\gamma_{\mathrm{F}}\end{bmatrix}$；$\boldsymbol{w}_1=\begin{bmatrix}(b_x+b_y)w_{z_2}\\-(b_x+b_y)w_{y_2}\end{bmatrix}$。

将方程组(17)的第3式和第4式也写成矩阵形式，可得

$$\begin{bmatrix}\dot{\omega}_\eta\\ \dot{\omega}_\zeta\end{bmatrix}=\boldsymbol{D}_1\begin{bmatrix}\omega_\eta\\ \omega_\zeta\end{bmatrix}+\boldsymbol{D}_2\begin{bmatrix}\delta_1\\ \delta_2\end{bmatrix}+\boldsymbol{h}+\boldsymbol{w}_2 \tag{19}$$

式中：$\boldsymbol{D}_1=v\begin{bmatrix}-k_{zz}&-(P_{\mathrm{F}}+P_{\mathrm{B}})\\P_{\mathrm{F}}+P_{\mathrm{B}}&-k_{zz}\end{bmatrix}$；$\boldsymbol{D}_2=v^2\begin{bmatrix}\dfrac{k_{\mathrm{F},y}\dot{\gamma}_{\mathrm{F}}}{v}+k_{\mathrm{B},y}&-k_z\\k_z&\dfrac{k_{\mathrm{F},y}\dot{\gamma}_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\end{bmatrix}$；$\boldsymbol{h}=v^2\begin{bmatrix}-kk_{\mathrm{c}}\delta_{\mathrm{c}}\sin\gamma_{\mathrm{F}}\\kk_{\mathrm{c}}\delta_{\mathrm{c}}\cos\gamma_{\mathrm{F}}\end{bmatrix}$；$\boldsymbol{w}_2=v^2\begin{bmatrix}-k_z\dfrac{w_{z_2}}{v}+\left(\dfrac{k_{\mathrm{F},y}\dot{\gamma}_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\dfrac{w_{y_2}}{v}\\k_z\dfrac{w_{y_2}}{v}+\left(\dfrac{k_{\mathrm{F},y}\dot{\gamma}_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\dfrac{w_{z_2}}{v}\end{bmatrix}$。

联立式(18)和式(19)，并利用导数关系$(\cdot)'=v()'$和$(\cdot\cdot)=v^2()''+\dot{v}()'$，可将方程自变量转化为对弹道弧长$s$的导数，同时，略去一些小量(如$b_x$、$k_{zz}$、$|\boldsymbol{g}|\sin\theta/v^2$等)的乘积项，经推导可得状态空间形式的角运动方程为

$$\begin{bmatrix}\delta_1''\\ \delta_2''\end{bmatrix}=\begin{bmatrix}-(P_{\mathrm{F}}T_{\mathrm{F}}+P_{\mathrm{B}}T_{\mathrm{B}})&-(P_{\mathrm{F}}+P_{\mathrm{B}})\\M&-H\end{bmatrix}\begin{bmatrix}\delta_1'\\ \delta_2'\end{bmatrix}+\begin{bmatrix}M&-H\\P_{\mathrm{F}}T_{\mathrm{F}}+P_{\mathrm{B}}T_{\mathrm{B}}&P_{\mathrm{F}}+P_{\mathrm{B}}\end{bmatrix}\begin{bmatrix}\delta_1\\ \delta_2\end{bmatrix}+\begin{bmatrix}R_y+w_y\\R_z+w_z\end{bmatrix} \tag{20}$$

式中：

$$H=k_{zz}+b_y-b_x-kb_{\mathrm{c},x}-\frac{|\boldsymbol{g}|\sin\theta}{v^2}$$

$$P_{\mathrm{F}}=\frac{C_{\mathrm{F}}\dot{\gamma}_{\mathrm{F}}}{Av},P_{\mathrm{B}}=\frac{C_{\mathrm{B}}\dot{\gamma}_{\mathrm{B}}}{Av},M=k_z$$

$$T_{\mathrm{F}}=b_y-\frac{A}{C_{\mathrm{F}}}k_{\mathrm{F},y},T_{\mathrm{B}}=b_y-\frac{1}{P_{\mathrm{B}}}k_{\mathrm{F},y}$$

$$R_y=k\left\{k_{\mathrm{c}}\delta_{\mathrm{c}}\cos\gamma_{\mathrm{F}}-\left[(P_{\mathrm{F}}+P_{\mathrm{B}})b_{\mathrm{c},y}\delta_{\mathrm{c}}-\frac{b_{\mathrm{c},y}\dot{\gamma}_{\mathrm{F}}\delta_{\mathrm{c}}}{v}\right]\sin\gamma_{\mathrm{F}}\right\}-\frac{\ddot{\theta}}{v^2}-k_{zz}\frac{\dot{\theta}}{v}$$

$$R_z=k\left\{k_{\mathrm{c}}\delta_{\mathrm{c}}\sin\gamma_{\mathrm{F}}+\left[(P_{\mathrm{F}}+P_{\mathrm{B}})b_{\mathrm{c},y}\delta_{\mathrm{c}}-\frac{b_{\mathrm{c},y}\dot{\gamma}_{\mathrm{F}}\delta_{\mathrm{c}}}{v}\right]\cos\gamma_{\mathrm{F}}\right\}+(P_{\mathrm{F}}+P_{\mathrm{B}})\frac{\dot{\theta}}{v}$$

$$w_y=-\left[(P_{\mathrm{F}}+P_{\mathrm{B}})(b_x+b_y)-\left(\frac{k_{\mathrm{F},y}\dot{\gamma}_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\right]\frac{w_{z_2}}{v}+k_z\frac{w_{y_2}}{v}$$

$$w_z=\left[(P_{\mathrm{F}}+P_{\mathrm{B}})(b_x+b_y)-\left(\frac{k_{\mathrm{F},y}\dot{\gamma}_{\mathrm{F}}}{v}+k_{\mathrm{B},y}\right)\right]\frac{w_{y_2}}{v}+k_z\frac{w_{z_2}}{v}$$

进一步令

$$\begin{cases}\boldsymbol{x}=\begin{bmatrix}\delta_1&\delta_1'&\delta_2&\delta_2'\end{bmatrix}^{\mathrm{T}}\\x_1=\delta_1\\x_2=\delta_1'\\x_3=\delta_2\\x_4=\delta_2'\end{cases} \tag{21}$$

则式(20)可表示为

$$\boldsymbol{x}'=\boldsymbol{J}\boldsymbol{x}+\boldsymbol{Z} \tag{22}$$

式中：

$$\boldsymbol{J}=\begin{bmatrix}0&1&0&0\\M&-H&-(P_{\mathrm{F}}T_{\mathrm{F}}+P_{\mathrm{B}}T_{\mathrm{B}})&-(P_{\mathrm{F}}+P_{\mathrm{B}})\\0&0&0&1\\P_{\mathrm{F}}T_{\mathrm{F}}+P_{\mathrm{B}}T_{\mathrm{B}}&P_{\mathrm{F}}+P_{\mathrm{B}}&M&-H\end{bmatrix}$$

$$\boldsymbol{Z}=\begin{bmatrix}0&R_y+w_y&0&R_z+w_z\end{bmatrix}^{\mathrm{T}}$$

式(22)即为自旋卷弧翼弹箭的非线性角运动方程，可用于后续分析。

3 基于平方和规划的吸引域估计方法

3.1 方法简介

采用平方和规划方法来估计有控自旋卷弧翼弹箭角运动的非线性吸引域。

弹箭非线性角运动系统可表示为

$$\dot{\boldsymbol{x}}=f(\boldsymbol{x}) \tag{23}$$

式中：$\boldsymbol{x}(t)\in\mathbf{R}^n$为状态变量；$f$为$n$维的$n$次多项式。

如果使用平方和规划方法，必须满足$f(\mathbf{0})=\mathbf{0}$。然而，对于式(23)，一般$\boldsymbol{x}'(0)\neq\mathbf{0}$，即该弹箭非线性角运动的稳定平衡点不在原点，故在计算前需要先将系统平衡点移至原点。

对此，可以找到一个可证明的吸引域，使所有始于此区域的点最终都能收敛到原点。根据Lyapunov定理[32]，原点吸引域可定义为

$$\boldsymbol{x}_0\in\mathbf{R}^n \tag{24}$$

若$\boldsymbol{x}(0)=\boldsymbol{x}_0$，则$\lim\limits_{t\to\infty}\boldsymbol{x}(t)=0$。

若存在连续可微的标量函数$V(\boldsymbol{x})$：$\mathbf{R}^n\to\mathbf{R}$和标量$\gamma_{\mathrm{R}}\in\mathbf{R}^+$，使得

$$V(\boldsymbol{x})>0\quad\forall\boldsymbol{x}\neq\mathbf{0}\text{且}V(0)=0 \tag{25}$$

$$\boldsymbol{\Omega}:=\{\boldsymbol{x}:V(\boldsymbol{x})\leqslant\gamma_{\mathrm{R}}\} \tag{26}$$

$$\boldsymbol{\Omega}\subseteq\left\{\boldsymbol{x}:(\partial V/\partial\boldsymbol{x})f(\boldsymbol{x})<0\right\}\cup\{0\} \tag{27}$$

则原点渐近稳定。其中，$\boldsymbol{\Omega}$为式(24)的吸引域子集，即弹箭角运动的稳定吸引域，γ_{R}为其水平集，$V(\boldsymbol{x})$为Lyapunov函数。

至此，求解弹箭角运动稳定吸引域的问题便转

化为求解式(25)～式(27)中 $V(\boldsymbol{x})$和 γ_{R}的约束问题。由于式(25)～式(27)中既有不等式约束, 又有集合包含约束, 难以直接求解, 故需通过代数几何学将其全部转化为不等式约束。同时, 为了进一步优化吸引域估计结果, 引入形状函数 $p(\boldsymbol{x})$, 使估计结果更加逼近系统的确切吸引域。

综上所述, 最终将问题转化为一个双线性平方和规划问题, 即 $\max\limits_{s_1,s_2\in\Sigma_n}\beta$, 使得

$$\begin{cases} V-l_1\in\Sigma_n \\ -\left[(\partial V/\partial\boldsymbol{x})f+l_2\right]-(\gamma_{\mathrm{R}}-V)s_2\in\Sigma_n \\ (\gamma_{\mathrm{R}}-V)-(\beta-p)s_1\in\Sigma_n \end{cases} \tag{28}$$

式中: $\beta\in\mathbf{R}$为形状函数 $p(\boldsymbol{x})$ 的水平集; l_1 和 l_2 为小量, 一般取$l_1=l_2=\varepsilon\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$, ε为正的小量, 可取$\varepsilon=10^{-6}$; s_1 和 s_2 为辅助算子, 为一定阶数的平方和多项式; Σ_n表示 n 个变量组成的平方和多项式集合。

为求解上述双线性问题, 将其解耦为 2 个单线性优化问题。采用迭代算法(*V*-*s* 迭代)进行求解, 算法具体步骤简述如下[24,30]:

步骤 1　固定 V, 利用二分法求得γ_{R}的最大值 γ_{R}^*, 并记下此时的s_2, 即$\gamma_{\mathrm{R}}^*:=\max\limits_{s_2\in\Sigma_n}\gamma_{\mathrm{R}}$, 使得

$$-\left[\left(\partial V/\partial\boldsymbol{x}\right)f+l_2\right]-(\gamma_{\mathrm{R}}-V)\,s_2\in\Sigma_n \tag{29}$$

步骤 2　固定V和γ_{R}^*, 利用二分法求得β的最大值β^*, 并记下此时的s_1, 即$\beta^*:=\max\limits_{s_1\in\Sigma_n}\beta$, 使得

$$(\gamma_{\mathrm{R}}^*-V)-(\beta-p)s_1\in\Sigma_n \tag{30}$$

步骤 3　固定 s_1、s_2、β^*、γ_{R}^*, 求解新的满足以下约束的 Lyapunov 函数, 即

$$\begin{cases} -\left[(\partial V/\partial\boldsymbol{x})f+l_2\right]-(\gamma_{\mathrm{R}}^*-V)s_2\in\Sigma_n \\ (\gamma_{\mathrm{R}}^*-V)-(\beta^*-p)s_1\in\Sigma_n \\ V-l_1\in\Sigma_n \\ V(0)=0 \end{cases} \tag{31}$$

步骤 4　$V=V/\gamma_{\mathrm{R}}^*$。

步骤 5　重复以上步骤, 得到最终的 Lyapunov 函数$V=V/\gamma_{\mathrm{R}}^*$和相应的吸引域的估计$\boldsymbol{\Omega}:=\{x\in\mathbf{R}^n:V<1\}$。遇到以下条件中的任意一种, 循环停止: ① 数值计算不可行; ② 相邻 2 次迭代所得的β之间的差小于预设精度, 如 0.001; ③ 达到预定的迭代次数。

为了增强吸引域估计, 可引入高阶 Lyapunov 函数$V(\boldsymbol{x})$, 即在 *V*-*s* 迭代算法中, 可通过待定系数来设定$V(\boldsymbol{x})$的阶次(如 2 阶、4 阶、6 阶 Lyapunov 函数), 阶次越高, 精度越高, 但 Lyapunov 函数越复杂, 计算量也越大。

3.2　计算过程

有控自旋卷弧翼弹箭角运动的非线性吸引域计算过程简要总结如下: ①利用多项式拟合方法建立非线性气动系数的多项式模型; ②将气动系数的多项式模型代入角运动方程(22)中, 可得多项式形式的弹箭非线性角运动模型; ③通过坐标转换, 将系统的平衡点移至原点; ④利用平方和规划方法估计该弹箭角运动的非线性吸引域。

4　非线性吸引域估计及其影响因素分析

攻击地面目标的有制导炮弹或火箭弹一般在降弧段进行弹道控制, 故鸭舵起控点可选为弹道顶点。本节以鸭舵起控点为例, 研究该特征点下自旋卷弧翼弹箭的非线性吸引域及其影响因素。

4.1　弹道仿真

4.1.1　仿真条件

本节利用第 1 节建立的七自由度刚体动力学模型进行弹道仿真, 仿真所用主要物理参数如表 1 所示。

表 1　弹箭物理参数

Table 1　Physical parameters of projectile

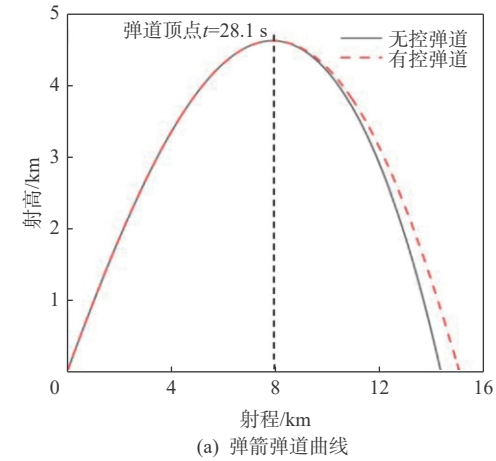
m_{F}/kg	m_{B}/kg	r_{F}/m	r_{B}/m	A_{F}/(kg·m^2)	A_{B}/(kg·m^2)	C_{F}/(kg·m^2)	C_{B}/(kg·m^2)	r/m	d/m	l/m	δ_{c}/(°)
15.10	24.90	0.152 4	0.307 3	1.265	5.767	0.023	0.094	0.619 8	0.101 6	0.101 6	5

仿真初始条件是在文献 [3] 中选取的一个马赫数作为初速, 即 v_0=580 m/s。取射角 θ_0=45°(该弹箭的最大射程角), 其余状态参数(如ψ_2、ω_η、ω_ζ、γ_{F}、γ_{B}等)的初值为 0。

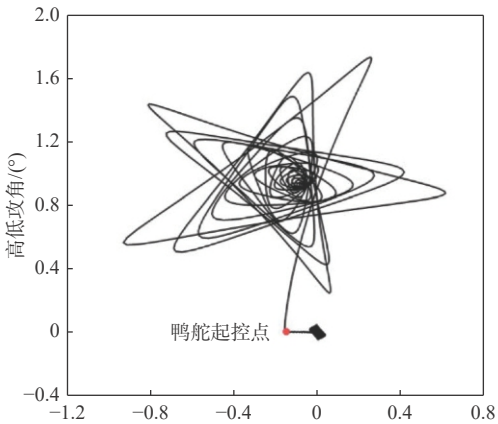
4.1.2　仿真结果

采用 4.1.1 节条件, 对某有控自旋卷弧翼弹箭进行有控和无控弹道特性仿真。假设弹箭无控时, 鸭舵折叠在前体控制舱内, 当弹箭飞至弹道顶点附近(t=28.1 s)时展开鸭舵, 以俯仰舵偏角 δ_{c}=5°进行滑翔飞行直至落地。图 3 给出了弹箭的弹道仿真结果。

由图 3(a) 可知, 无控弹道的飞行时间 t=60.6 s, 飞行距离 x=14.383 km, 末端速度 v_{end}=306 m/s; 有控弹道的飞行时间 t=64.9 s, 飞行距离 x=15.094 km, 末端速度 v_{end}=299 m/s。与无控弹道相比, 有控弹箭飞行距离增加了 711 m, 显示出一定的弹道修正能力。



(a) 弹箭弹道曲线



(b) 攻角相平面

图3 仿真结果

Fig. 3 Simulation results

图3(b)给出了有控弹箭的攻角相平面，可知，鸭舵起控前，弹箭的攻角摆动范围很小，高低攻角在±0.05°内摆动，方向攻角由于受卷弧翼自诱导滚转力矩的影响会发生偏转，最大偏转角度为−0.15°；鸭舵起控后，弹箭攻角运动受到较大影响，攻角摆动范围激增近10～20倍。

4.2 特征点处的非线性气动力系数

选取鸭舵起控点(无控弹道顶点)作为弹道特征点，研究弹箭角运动的非线性吸引域。表2给出了该特征点处的各状态参数。

表2 弹道顶点处的状态参数

Table 2 State parameters at the apex of the trajectory

t/s	v/(m·s^{-1})	Ma	$\dot{\gamma}_{\mathrm{F}}$/(rad·s^{-1})	$\dot{\gamma}_{\mathrm{B}}$/(rad·s^{-1})	θ_{a}/(°)
28.10	230.67	0.714	13.13	182.89	0

为研究有控自旋卷弧翼弹箭角运动的非线性吸引域，将特征点对应的马赫数Ma=0.714、攻角δ=0°～30°范围内各气动系数拟合成关于攻角的非线性函数。表3给出了各气动力系数拟合结果的决定系数R^2的值，可以看出，各气动系数对应的R^2值均接近于1，表明拟合度较好。各气动力系数的表达式为

$$\begin{cases} c_x = 0.489 + 12.900\,5\delta^2 \\ c_y' = 11.979 + 20.199\delta^2 - 82.894\delta^4 \\ m_z' = -17.579 - 61.556\delta^2 + 274.883\delta^4 \\ m_{zz}' = -25.917 - 137.91\delta^2 + 627.86\delta^4 \\ m_{\mathrm{F},y}'' = -25.189 + 1\,383.772\delta^2 - 4\,284.672\delta^4 \\ m_{\mathrm{B},y}' = 0.636 - 15.851\delta^2 + 44.37\delta^4 \end{cases} \tag{32}$$

表3 各气动系数拟合决定系数

Table 3 The coefficient of determination of each aerodynamic coefficient parameter

参数	R^2
c_x	0.999
c_y'	0.995
m_z'	0.994
m_{zz}'	0.998
$m_{\mathrm{F},y}''$	0.983
$m_{\mathrm{B},y}'$	0.976

与传统固定卷弧翼弹箭的不同之处在于，自旋卷弧翼弹箭前、后体的马格努斯力矩系数可分开计算。图4给出了Ma=0.714时，该弹箭前、后体马格努斯力矩系数随攻角δ的变化结果。可知，从正负号上看，前、后体马格努斯力矩系数基本异号，且随着攻角增大，均会出现换向的情形：前体马格努斯力矩系数在攻角δ=8°附近从负值变为正值，其在δ=8°时的值等于0.001 7；后体马格努斯力矩系数在攻角δ=12°附近从正值变为负值，其在δ=12°时的值等于−0.000 3。从数值大小上看，后体马格努斯力矩系数远大于前体(约1个数量级)。

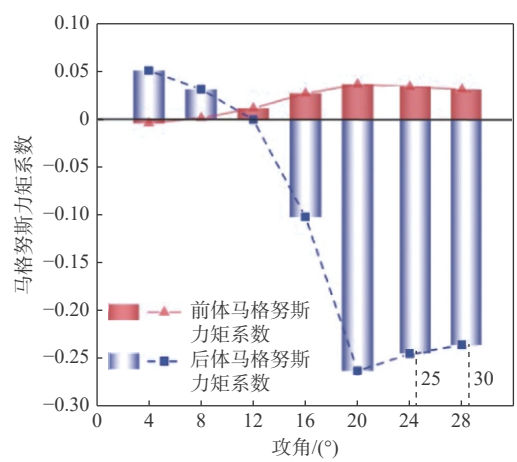


图4 前、后体马格努斯力矩系数随攻角变化

Fig. 4 The Magnus moment coefficient of front and rear bodies varies with the angle of attack

4.3　仿真验证

将非线性气动力表达式(32)代入角运动方程(22), 此时, $\boldsymbol{x}_t=[2\times10^{-5}(°),\ 0(°)/m,\ -2.5\times10^{-3}(°),\ 0(°)/m]^T$ 为该方程的一个稳定平衡点。通过坐标转换可将该平衡点移至原点, 即 $\boldsymbol{x}=\boldsymbol{x}+\boldsymbol{x}_t$, 再利用平方和规划计算其吸引域。选取初始 Lyapunov 函数为

$$V_0=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} \tag{33}$$

式中: $\boldsymbol{Q}$ 为正定矩阵, 由 Lyapunov 方程 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Q}+\boldsymbol{Q}\boldsymbol{A}=-\boldsymbol{I}$ 计算得到, $\boldsymbol{A}=\left(\partial\boldsymbol{x}'/\partial\boldsymbol{x}\right)|_{x=0}$ 为满秩矩阵, $\boldsymbol{I}$ 为单位矩阵。

利用 *V-s* 迭代算法, 分别计算 2 阶、4 阶、6 阶 Lyapunov 函数对角运动方程(22)的吸引域估计, 系统迭代次数设置为 50 次。表 4 给出了不同阶次下 Lyapunov 函数所对应的水平集。

表 4　不同阶次下 Lyapunov 函数所对应的水平集

Table 4　The level sets corresponding to Lyapunov functions of different orders

*V*的阶次	水平集
2	1.000 1
4	1.000 2
6	1.001 2

计算所得的 Lyapunov 函数分别记为 V_2、V_4、V_6。由于 V_4 和 V_6 的解析表达式较冗长, 只给出 V_2 的具体解析表达式, V_4、V_6 给出部分高阶项解析表达式(略去的项用省略号“…”代替):

$$\begin{aligned}V_2=&\ 36.638\delta_1^2-0.094\,969\delta_1\delta_2+0.618\,93\delta_1\delta_1'+\\&3.451\,6\delta_1\delta_2'+36.643\,1\delta_2^2-3.445\,2\delta_2\delta_1'+\\&0.637\,49\delta_2\delta_2'+6.151\,1\delta_1'^2-0.014\,481\delta_1'\delta_2'+\\&6.146\,3\delta_1'^2\end{aligned} \tag{34}$$

$$\begin{aligned}V_4=&\ 4.237\,8\delta_1^4+0.009\,642\,8\delta_1^3\delta_2+0.209\,15\delta_1^3\delta_1'-\\&0.918\,73\delta_1^3\delta_2'+8.429\,1\delta_1^2\delta_2^2+0.917\,85\delta_1^2\delta_2\delta_1'+\\&0.218\,79\delta_1^2\delta_2\delta_2'+0.422\,23\delta_1^2\delta_1'^2-0.081\,598\delta_1^2\delta_1\delta_2'+\\&0.505\,6\delta_1^2\delta_2'^2+0.007\,823\,2\delta_1\delta_2^3+0.191\,82\delta_1\delta_2^2\delta_1'-\\&0.980\,39\delta_1\delta_2^2\delta_2'+0.085\,093\delta_1\delta_2\delta_1'^2-\\&0.185\,98\delta_1\delta_2\delta_1'\delta_2'-0.086\,561\delta_1\delta_2'\delta_2'^2+\\&0.031\,219\delta_1\delta_1'^3-0.149\,16\delta_1\delta_1'^2\delta_2'+\\&0.032\,893\delta_1\delta_1'\delta_2'^2-0.157\,23\delta_1\delta_2'^3+4.197\,1\delta_2^4+\cdots+\\&4.850\,9\delta_1^2-0.002\,398\,9\delta_1\delta_2+0.067\,516\delta_1\delta_1'+\\&0.408\,2\delta_1\delta_2'+4.852\,4\delta_2^2-0.407\,79\delta_2\delta_1'+\\&0.067\,607\delta_2\delta_2'+0.815\,59\delta_1'^2-\\&0.000\,497\,49\delta_1'\delta_2'+0.815\,72\delta_2'^2\end{aligned} \tag{35}$$

$$\begin{aligned}V_6=&\ 2.152\,6\delta_1^6+0.190\,58\delta_1^5\delta_2-0.541\,24\delta_1^5\delta_1'+\\&1.261\,6\delta_1^5\delta_2'+4.570\,3\delta_1^4\delta_2^2-1.295\,7\delta_1^4\delta_2\delta_1'+\\&0.581\,94\delta_1^4\delta_2\delta_2'+0.533\,12\delta_1^4\delta_1'^2-0.376\,85\delta_1^4\delta_1\delta_2'+\\&3.267\,8\delta_1^4\delta_2'^2+0.035\,549\delta_1^3\delta_2^3-0.620\,21\delta_1^3\delta_2^2\delta_1'+\\&2.284\,9\delta_1^3\delta_2^2\delta_2'+0.342\,06\delta_1^3\delta_2\delta_1'^2-5.851\,7\delta_1^3\delta_2\delta_1'\delta_2'-\\&0.296\,21\delta_1^3\delta_2\delta_2'^2+0.050\,52\delta_1^3\delta_1'^3-\\&0.008\,348\,2\delta_1^3\delta_1'^2\delta_2'+0.355\,31\delta_1^3\delta_1'\delta_2'^2+\\&0.048\,204\delta_1^3\delta_2'^3+4.551\,5\delta_1^2\delta_2^4-2.289\,9\delta_1^2\delta_2^3\delta_1'-\\&0.564\,49\delta_1^2\delta_2^3\delta_2'+3.004\,5\delta_1^2\delta_2^2\delta_1'^2+\\&0.009\,418\,9\delta_1^2\delta_2^2\delta_1'\delta_2'+2.987\delta_1^2\delta_2^2\delta_2'^2+\\&0.026\,783\delta_1^2\delta_2\delta_1'^3-0.497\,82\delta_1^2\delta_2\delta_1'^2\delta_2'-\\&0.205\,63\delta_1^2\delta_2\delta_1'\delta_2'^2+0.360\,16\delta_1^2\delta_2\delta_2'^3-\\&0.041\,149\delta_1^2\delta_1'^4-0.014\,999\delta_1^2\delta_1'^3\delta_2'+\\&0.278\,86\delta_1^2\delta_1'^2\delta_2'^2-0.004\,145\,1\delta_1^2\delta_1'\delta_2'^3+\\&0.255\,39\delta_1^2\delta_2'^4-\cdots+0.799\delta_2'^2\end{aligned} \tag{36}$$

利用 V_2、V_4、V_6 计算所得的吸引域分别表示为 $\boldsymbol{\Omega}_{V_2}$、$\boldsymbol{\Omega}_{V_4}$、$\boldsymbol{\Omega}_{V_6}$, 其空间轨迹如图 5 所示, 其在各相平面的轨迹(剖面)如图 6 所示。

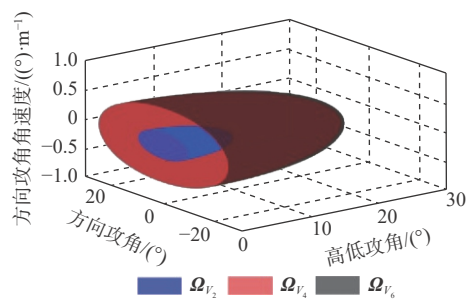


图 5　吸引域估计结果的空间轨迹

Fig. 5　The spatial trajectory of the estimated result of the region of attraction

由图 5 和图 6 可知, V_2 计算所得吸引域 $\boldsymbol{\Omega}_{V_2}$ 明显小于 V_4、V_6, 而吸引域 $\boldsymbol{\Omega}_{V_4}$ 与 $\boldsymbol{\Omega}_{V_6}$ 基本重合。考虑 V_6 的计算量远大于 V_4, 故下文采用 4 阶 Lyapunov 函数进行计算和分析。

图 7 给出了不同初始攻角下, 该有控自旋卷弧翼弹箭的攻角运动轨迹。

为便于比较和分析, 引入参数 χ, 表示攻角吸引域的边界值(即攻角吸引域在纵向或横向上的最大值, 简称吸引域边界)。由图 7 可知, 该特征点处的吸引域边界 $\chi=23.83°$, 当初始攻角(图中的初始攻角 1, $\delta_1=23.49°$, $\delta_2=0°$)在吸引域内时, 弹箭攻角轨迹逐渐收敛到 0°; 对于图中的初始攻角 2, $\delta_1=25.78°$, $\delta_2=25.78°$, 其初始值在吸引域外, 此时, 攻角轨迹运动至 40°并逐渐发散, 这验证了基于平方和规划方法估计弹箭攻角运动稳定范围的有效性。

4.4　鸭舵控制对吸引域的影响

本节研究鸭舵进行俯仰通道控制时, 不同前体

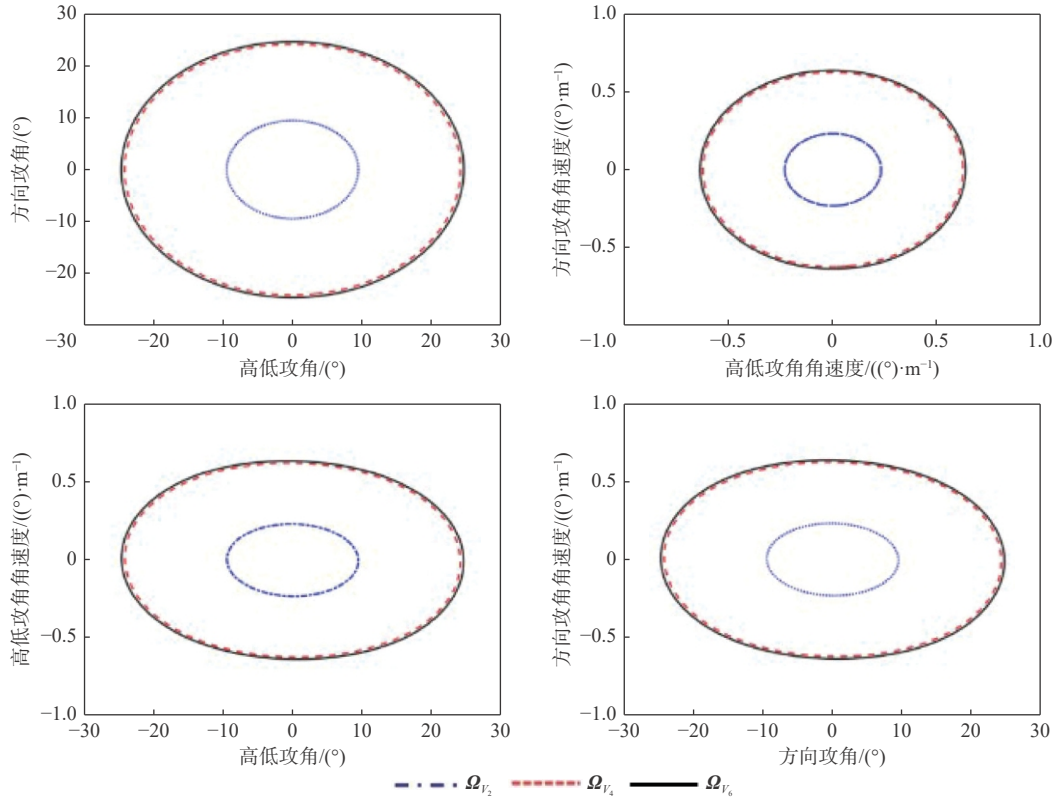


图 6 吸引域估计结果的相平面轨迹

Fig. 6 The phase plane trajectory of the estimated result of the region of attraction

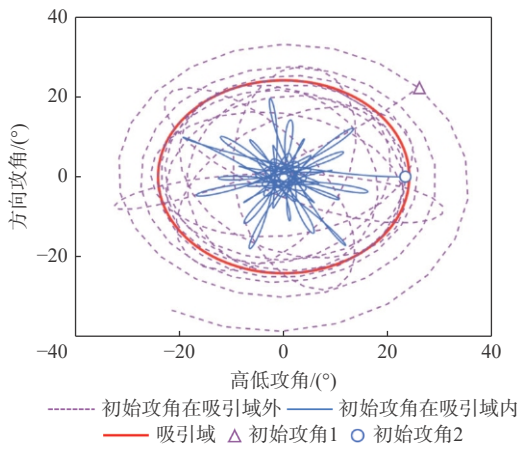


图 7 不同初始攻角下的攻角运动轨迹

Fig. 7 Angle of attack motion trajectories at different initial angles of attack

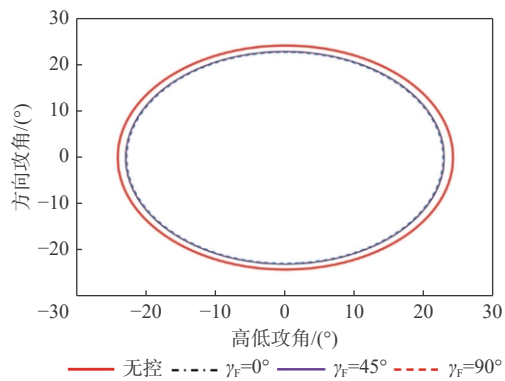


图 8 不同控制方位角下的吸引域估计结果

Fig. 8 The estimation results of region of attraction under different control azimuth

滚转控制方位角γ_F和控制力大小对系统吸引域的影响。图 8 给出了无控和有控条件下的(γ_F=0°, 45°, 90°, δ_c=5°)吸引域估计结果。可知，无控状态下系统的吸引域边界为$\chi=23.83°$。有控状态下，控制方位角$\gamma_F=0°$时，系统吸引域边界$\chi=22.87°$(小于无控状态)，这表明鸭舵控制力会减小弹箭角运动吸引域。而不同控制方位角下的吸引域轨迹曲线基本重合，表明控制方位角对弹箭角运动吸引域的影响很小。

图 9 给出了不同舵偏角对应的吸引域估计结果，可知，当 δ_c=5°时，$\chi=22.87°$，与无控状态相比，χ减小了 4%；当 δ_c=7.5°时，$\chi=22.19°$，与无控状态相比，χ减小了 6.9%；当 δ_c=20°时，$\chi=19.81°$，与无控状态相比，χ减小了 16.9%。因此，随着 δ_c 的增大，弹箭角运动吸引域的范围有所减小，这表明鸭舵控制力大小对弹箭角运动吸引域有一定影响，且呈现一定的线性变化趋势。在弹箭控制时，有必要充分考虑鸭舵控制力和力矩对弹箭非线性稳定性的影响。

4.5 非线性气动力对吸引域的影响

为便于分析三次方气动力系数和五次方气动力系数对弹箭角运动吸引域的影响，将式(32)改写

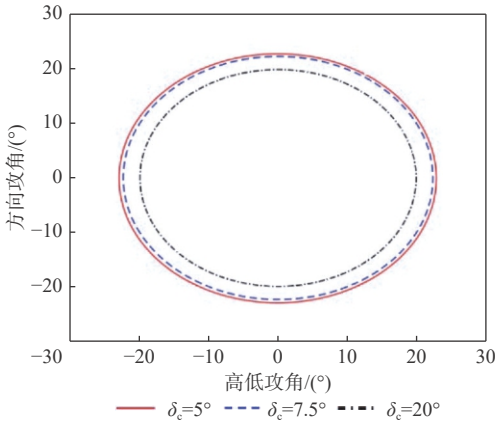


图 9　不同舵偏角对应的吸引域估计结果

Fig. 9　The estimation results of region of attraction corresponding to different rudder angles

为如下形式:

$$\begin{cases} c_x = c_{x_0}(1+\kappa_1\delta^2) \\ c_y' = c_{y_0}(1+\kappa_2\delta^2+\tau_1\delta^4) \\ m_z' = m_{z_0}(1+\kappa_3\delta^2+\tau_2\delta^4) \\ m_{zz}' = m_{zz_0}(1+\kappa_4\delta^2+\tau_3\delta^4) \\ m_{\mathrm{F},y}'' = m_{\mathrm{F},y_0}(1+\kappa_5\delta^2+\tau_4\delta^4) \\ m_{\mathrm{B},y}' = m_{\mathrm{B},y_0}(1+\kappa_6\delta^2+\tau_5\delta^4) \end{cases} \tag{37}$$

式中: $\kappa_i(i=1,2,\cdots,6)$为各气动力和力矩系数(导数)三次方项与其线性项的比值; $\tau_i(i=1,2,\cdots,5)$为各气动力和力矩系数(导数)五次方项与其线性项的比值。

对于本文的算例, 取$\kappa_1=26.3$、$\kappa_2=1.7$、$\kappa_3=3.5$、$\kappa_4=5.3$、$\kappa_5=54.9$、$\kappa_6=24.8$、$\tau_1=6.9$、$\tau_2=15.7$、$\tau_3=24.2$、$\tau_4=171.2$、$\tau_5=69.6$。据此设定κ_i和τ_i的变化范围, 即$\kappa_1\in[-80,80]$, $\kappa_5\in[-200,200]$, $\kappa_6\in[-150,100]$, $\tau_1,\tau_2,\tau_3\in[-100,100]$, $\tau_4,\tau_5\in[-500,500]$。

4.5.1　阻力和赤道阻尼力矩非线性项对吸引域的影响

图 10 给出了弹箭角运动吸引域边界χ随κ_1和κ_4的变化情况。

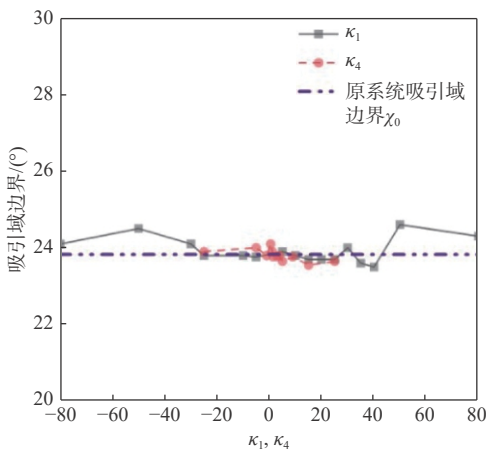


图 10　吸引域边界随κ_1和κ_4变化

Fig. 10　Region of attraction boundaries vary with κ_1 and κ_4

为便于对比, 记 4.3 节算例结果为原系统吸引域边界(即$\chi_0=23.83°$, 如图 7 所示)。由图 10 可知, 随着κ_1和κ_4的变化, 弹箭吸引域边界χ在原系统吸引域边界χ_0附近小幅度波动, 波动幅度不超过 0.8°, 这表明阻力系数和赤道阻尼力矩系数导数的非线性项对弹箭的稳定吸引域几乎无影响, 只起到阻尼作用(即影响弹箭飞行速度), 故下文不对赤道阻尼力矩系数导数五次方项τ_3进行分析。

4.5.2　升力和静力矩非线性项对吸引域的影响

图 11 给出了弹箭角运动吸引域边界χ随κ_2和κ_3的变化情况, 图 12 给出了弹箭角运动吸引域边界χ随τ_1和τ_2的变化情况。

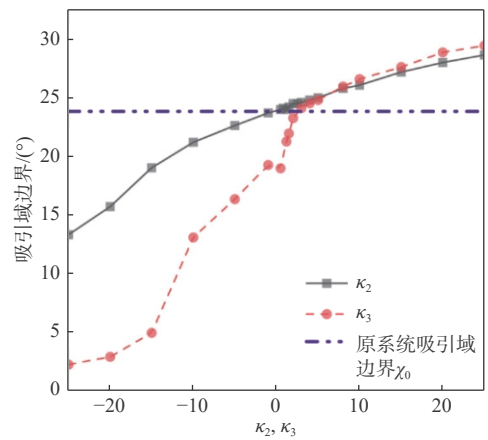


图 11　吸引域边界随κ_2和κ_3变化

Fig. 11　Region of attraction boundaries vary with κ_2 and κ_3

由图 11 可知, 吸引域边界χ随着κ_2和κ_3的增加而增大, 当$\kappa_2<0, \kappa_3<0$时, 吸引域边界χ会快速减小, 特别是当κ_3=−25 时, 吸引域边界χ接近于 0°。这表明, 升力系数导数和静力矩系数导数的三次方项对弹箭稳定吸引域的影响较大, 两者对弹箭吸引域的影响规律较为一致。对于尾翼弹, 由于静力矩起稳定作用, 故当κ_3减小为负值时, 稳定力矩也会随之减小, 从而造成弹箭稳定吸引域范围的减小。

由图 12 可看出, 升力系数导数和静力矩系数导数的五次方项对弹箭吸引域的影响规律差异较大。升力系数导数的五次方项对弹箭吸引域的影响很小, 随着τ_1的变化, 吸引域边界χ与原系统χ_0相比, 最大波动幅值不超过 1.8°, 而静力矩系数导数五次方项对弹箭吸引域的影响较大, 这主要是由于压心的非线性变化较明显, 造成静力矩系数导数与升力系数导数的五次方项对吸引域影响规律的巨大差异。如图 12 所示, 当τ_2趋近于 0 时, 系统吸引域边界χ迅速增大到 47°, 几乎达到原系统吸引域边界χ_0的 2 倍; 当τ_2的绝对值大于 20 时, 系统吸引域边界小于原系统吸引域边界且不断减小, 当$\tau_2=-100$

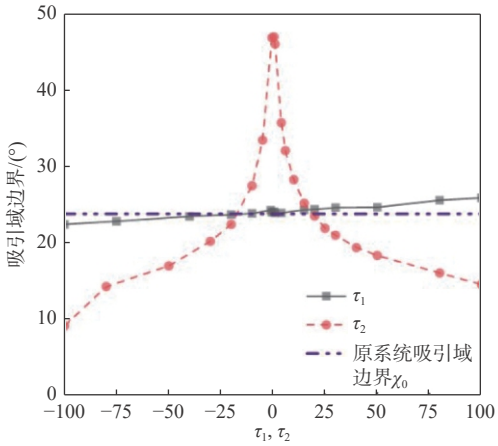


图 12 吸引域边界随τ_1和τ_2变化

Fig. 12 Region of attraction boundaries vary with τ_1 and τ_2

时，$\chi=9.17°$，与χ_0相比减小了61.5%。在图12所示区间内，随着τ_2变化，系统吸引域边界的波动幅值达38°。因此，对于该类弹箭的设计，应重点考虑静力矩系数导数五次方项的影响。

4.5.3 马格努斯力矩非线性项对吸引域的影响

自旋卷弧翼弹箭和传统固定卷弧翼弹箭最主要的差别在于自旋卷弧翼弹箭前、后体是解耦的。前、后体外形参数、卷弧翼的曲率、两体之间的约束力矩等因素均可能造成前、后体转速不同，从而使前、后体马格努斯力矩系数具有一定差异，有可能影响弹箭的稳定性。本节分析前、后体马格努斯力矩系数非线性项对弹箭吸引域的影响。

图13给出了弹箭角运动吸引域边界χ随κ_5和κ_6的变化曲线，图14给出了弹箭角运动吸引域边界χ随τ_4和τ_5的变化曲线。

由图13可知，吸引域边界χ随κ_5和κ_6的变化趋势较为一致。当$\kappa_5\in[-100,10]$、$\kappa_6\in[-30,-10]$时，χ基本维持在χ_0附近，当超过各自边界值时，吸引域边界χ迅速减小。后体马格努斯力矩系数导数三次

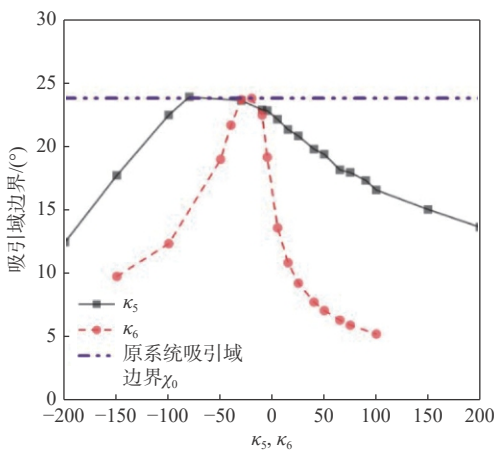


图 13 吸引域边界随κ_5和κ_6变化

Fig. 13 Region of attraction boundaries vary with κ_5 and κ_6

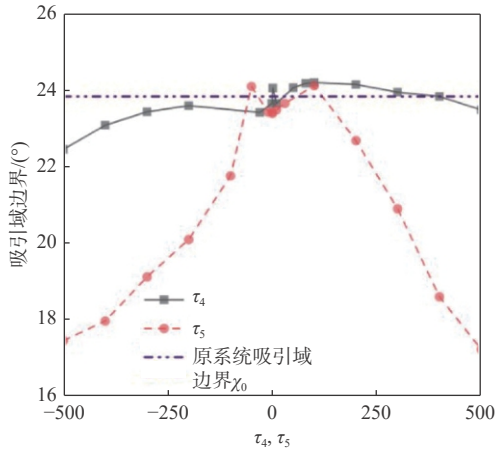


图 14 吸引域边界随τ_4和τ_5变化

Fig. 14 Region of attraction boundaries vary with τ_4 and τ_5

方项对吸引域边界的影响明显大于前体。以$\kappa_5=\kappa_6=100$处为例，当$\kappa_5=100$时，$\chi=16.61°$，与原系统吸引域χ_0相比减小了30.3%；当$\kappa_6=100$时，$\chi=5.27°$，与χ_0相比减小了77.9%。

如图14所示，前体马格努斯力矩系数导数五次方项对系统吸引域边界χ的影响很小，随着τ_4的变化，吸引域边界χ与原系统χ_0相比，最大波动幅值不超过1.37°，而后体马格努斯力矩系数导数五次方项对系统吸引域边界χ的影响很大。当$\tau_5\in[-70,130]$时，吸引域边界χ在原系统吸引域边界χ_0附近小幅度波动，波动幅值不超过0.4°，当超过其边界值时，吸引域边界迅速减小，当$\tau_5=\pm500$时，吸引域边界减小至17.3°左右，与原系统吸引域χ_0相比减小了约27.4%。这主要是由于后体马格努斯力矩远大于前体马格努斯力矩(见图4)。

综上，尽管前体马格努斯力矩系数导数五次方项对系统吸引域边界影响不大，但其三次方项的影响不可忽视。在设计该类弹箭时，要充分考虑卷弧翼外形参数、前后两体之间的滚转约束等与马格努斯效应密切相关的因素，避免马格努斯力矩非线性项超过边界值，从而造成弹箭飞行失稳。

5 结 论

本文针对有控自旋卷弧翼弹箭角运动的非线性吸引域问题开展研究，建立了七自由度刚体动力学模型，推导出状态空间形式的非线性角运动模型。在选取的弹道特征点上，利用基于平方和规划的吸引域估计方法定量分析了鸭舵控制和气动力非线性对该弹箭角运动吸引域的影响，可得到如下结论：

1) 鸭舵控制力增大会引起弹箭角运动吸引域边界χ减小。本文算例中，当δ_c=20°时，与无控状态

相比, 吸引域边界减小了 16.9%。滚转控制方位角对χ的影响很小, 故在弹箭控制时, 应注意对控制力大小有所限制。

2) 升力系数和静力矩系数的非线性项对弹箭吸引域影响较显著。升力系数导数三次方项与线性项的比值κ_2、静力矩系数导数三次方项与线性项的比值κ_3均不能小于 0, 否则, 会造成系统吸引域边界的快速减小; 静力矩系数导数五次方项与线性项比值τ_2的绝对值不能大于某一阈值(如本文算例中为 20), 算例中, 当$\tau_2=-100$时, $\chi=9.17°$, 与原系统吸引域边界χ_0相比减小 61.5%, 影响较大。

3) 自旋卷弧翼弹箭后体非线性马格努斯力矩对弹箭角运动吸引域的影响明显大于前体; 前、后体马格努斯力矩系数导数的三次方项与线性项的比值κ_5和κ_6、后体马格努斯力矩系数导数的五次方项与线性项的比值τ_5在一定区间内变化时, 吸引域边界基本维持在原系统吸引域边界附近, 当超过各自边界值时, 吸引域边界迅速减小, 影响较为明显。

参考文献 (References)

[1] 孙化东, 于剑桥, 梅跃松. 卷弧翼滚转导弹 Lipschitz 自适应轨迹线性化控制[J]. 系统工程与电子技术, 2017, 39(1): 162-167.
SUN H D, YU J Q, MEI Y S. Control of wrap-around fin rolling missiles based on Lipschitz adaptive trajectory linearization[J]. Systems Engineering and Electronics, 2017, 39(1): 162-167(in Chinese).

[2] MOROTE J, LIANO G. Flight dynamics of unguided rockets with free-rolling wrap around tail fins[J]. Journal of Spacecraft and Rockets, 2006, 43(6): 1422-1423.

[3] EASTMAN D, WENNDT D. Aerodynamics of maneuvering missiles with wrap-around fins[C]//Proceedings of the 3rd Applied Aerodynamics Conference. Reston: AIAA, 1985: 4083.

[4] EDGE H L. Computation of the roll moment for a projectile with warp-around fins[J]. Journal of Spacecraft and Rockets, 1994, 31(4): 615-620.

[5] TILMANN C P, BUTER T A, BOWERSOX R D W. Characterization of the flowfield near a wrap-around fin at Mach 2.8[J]. Journal of Aircraft, 1998, 35(6): 868-875.

[6] 赵良玉, 杨树兴, 焦清介. 提高卷弧翼火箭弹圆锥运动渐近稳定性的几个方法[J]. 固体火箭技术, 2010, 33(4): 369-372.
ZHAO L Y, YANG S X, JIAO Q J. Several methods for improving asymptotic stability of coning motion of wrap-around-fin rockets[J]. Journal of Solid Rocket Technology, 2010, 33(4): 369-372(in Chinese).

[7] MOROTE J. Lateral motion of free-rolling tail rockets in free flight[J]. Journal of Spacecraft and Rockets, 2005, 42(5): 873-882.

[8] MURPHY C H. Free flight motion of symmetric missiles: AD Report No. 1216[R]. [s.l.]: Ballistic Research Laboratory, 1963.

[9] MURPHY C H. Instability of controlled projectiles in ascending or descending flight[J]. Journal of Guidance and Control, 1981, 4(1): 66-69.

[10] MOROTE J, MOROTE P. Catastrophic yaw of triform and cruciform-tailed missiles[J]. Journal of Spacecraft and Rockets, 2018, 55(1): 247-250.

[11] MOROTE J. Resonant lock-in of unguided rockets having nonlinear aerodynamic properties[C]//Proceedings of the 44th AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2006: 830.

[12] MOROTE J, LIAÑO G, JIMÉNEZ J. Nonlinear rolling motion of triform finned missiles[J]. Journal of Spacecraft and Rockets, 2017, 54(2): 459-470.

[13] MURPHY C H. Nonlinear motion of a missile with slight configurational asymmetries[J]. Journal of Spacecraft and Rockets, 1971, 8(3): 259-263.

[14] MCCOY R L. Modern exterior ballistics: the launch andflight dynamics of symmetric projectiles[M]. Atglen: Schiffer Military History, 1999.

[15] 蒋增辉, 宋威, 陈农. 小迎角高超声速非旋转钝锥非线性运动分析[J]. 航空学报, 2016, 37(5): 1454-1461.
JIANG Z H, SONG W, CHEN N. Analysis of nonlinear motion for hypersonic non-spinning blunt cones at small angle of attack[J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(5): 1454-1461(in Chinese).

[16] 钱龙, 常思江, 杨文龙, 等. 固定鸭舵修正弹非线性与非定常气动数值模拟[J]. 弹道学报, 2020, 32(1): 31-37.
QIAN L, CHANG S J, YANG W L, et al. Simulation on nonlinear and unsteady aerodynamic characteristics of fixed-canard correction projectile[J]. Journal of Ballistics, 2020, 32(1): 31-37(in Chinese).

[17] 韩子鹏, 常思江, 史金光. 弹箭非线性运动理论[M]. 北京: 北京理工大学出版社, 2016.
HAN Z P, CHANG S J, SHI J G. Nonlinear motion theory of projectile and rocket[M]. Beijing: Beijing Insititute of Technology Press, 2016(in Chinese).

[18] CHANG S J, LI D Y. Homotopy-analysis-method-based solutions of quintic nonlinear angular motion for projectiles[J]. Journal of Spacecraft and Rockets, 2018, 55(3): 669-680.

[19] 李东阳, 常思江, 王中原, 等. 正规形法在弹箭非线性运动分析中的应用[J]. 国防科技大学学报, 2022, 44 (2): 44-54.
LI D Y, CHANG S J, WANG Z Y , et al. Applying the method of normal forms to projectile nonlinear motion analysis[J]. Journal of National University of Defense Technology, 2022, 44(2): 44-54(in Chinese).

[20] 杨树兴, 赵良玉, 闫晓勇. 旋转弹动态稳定性理论[M]. 北京: 国防工业出版社, 2014.
YANG S X, ZHAO L Y, YAN X Y. Dynamic stability of spinning missiles[M]. Beijing: National Defense Industry Press, 2014(in Chinese).

[21] 钟扬威, 王良明, 傅健, 等. 弹箭非线性角运动稳定性 Hopf 分岔分析[J]. 兵工学报, 2015, 36(7): 1195-1202.
ZHONG Y W, WANG L M, FU J, et al. Hopf bifurcation analysis of nonlinear angular motion stability of projectile[J]. Acta Armamentarii, 2015, 36(7): 1195-1202(in Chinese).

[22] 杨杰, 常思江, 魏伟. 带扰流片旋转稳定弹动态稳定性[J]. 兵工学报, 2021, 42(8): 1613-1623.
YANG J, CHANG S J, WEI W. Dynamic stability of spin-stabilized projectile with spoiler[J]. Acta Armamentarii, 2021, 42(8): 1613-1623(in Chinese).

[23] 赵新新, 史金光, 王中原, 等. 鸭舵式双旋弹非线性角运动及其分岔特性[J]. 兵工学报, 2023, 44(10): 3067-3078.
ZHAO X X, SHI J G, WANG Z Y, et al. Nonlinear angular motion and bifurcation characteristics of canard dual-spin projectiles[J]. Acta Armamentarii, 2023, 44(10): 3067-3078(in Chinese).

[24] KHODADADI L, SAMADI B, KHALOOZADEH H. Estimation of region of attraction for polynomial nonlinear systems: a numerical method[J]. ISA Transactions, 2014, 53(1): 25-32.

[25] GENESIO R, TARTAGLIA M, VICINO A. On the estimation of asymptotic stability regions: state of the art and new proposals[J]. IEEE Transactions on Automatic Control, 1985, 30(8): 747-755.

[26] CHIANG H D, THORP J S. Stability regions of nonlinear dynamical systems: a constructive methodology[J]. IEEE Transactions on Automatic Control, 1989, 34(12): 1229-1241.

[27] SIDOROV E, ZACKSENHOUSE M. Lyapunov based estimation of the basin of attraction of Poincare maps with applications to limit cycle walking[J]. Nonlinear Analysis: Hybrid Systems, 2019, 33: 179-194.

[28] CUNIS T, CONDOMINES J P, BURLION L. Sum-of-squares flight control synthesis for deep-stall recovery[J]. Journal of Guidance, Control, and Dynamics, 2020, 43(8): 1498-1511.

[29] 伍强, 徐浩军, 裴彬彬, 等. 基于吸引域与二元极值理论的结冰飞机飞行风险量化评估[J]. 航空学报, 2022, 43(5): 125137.
WU Q, XU H J, PEI B B, et al. Quantitative evaluation of flight risk of icing aircraft based on theory of region of attraction and binary extreme value[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(5): 125137(in Chinese).

[30] 李东阳, 常思江, 王中原. 弹箭角运动的非线性吸引域估计[J]. 兵工学报, 2023, 44(8): 2329-2341.
LI D Y, CHANG S J, WANG Z Y. Nonlinear region of attraction estimation for projectile's angular motion[J]. Acta Armamentarii, 2023, 44(8): 2329-2341(in Chinese).

[31] 韩子鹏, 常思江. 弹箭外弹道学[M]. 2 版. 北京: 北京理工大学出版社, 2022.
HAN Z P, CHANG S J. Exterior ballistics of projectiles and rockets[M]. 2nd ed. Beijing: Beijing Insititute of Technology Press, 2022(in Chinese).

[32] TAN W. Nonlinear control analysis and synthesis using sum-of-squares programming[D]. Berkeley: University of California, Berkeley, 2006.

Nonlinear region of attraction of angular motion for a controlled projectile with self-rotating wraparound fins

ZHOU Wentao[1]，CHANG Sijiang[1,*]，ZHAN Pengyao[1]，CUI Huizhen[2]

(1. School of Energy and Power Engineering，Nanjing University of Science and Technology，Nanjing 210094，China；
2. The Second Design Department，Northwest Industrial Group Corporation，Xi'an 710043，China)

Abstract: A study on the nonlinear angular motion characteristics has been conducted with the goal of addressing the stability issue of a controlled projectile with self-rotating wraparound fins under the influence of nonlinear aerodynamic force and duck rudder control force. According to the ballistic and structural characteristics of this type of sling and arrows, a seven-degree-of-freedom rigid-body motion model is established, according to which the nonlinear angular equations of motion in the state-space form are deduced. The characteristic points of the ballistic trajectory are selected, and the corresponding nonlinear aerodynamic coefficients are obtained by using the numerical simulation and polynomial fitting methods, and the estimation of the domain of attraction based on the sum-of-squares planning is adopted to analyse the effects of the duck rudder control and nonlinear aerodynamic coefficients on the stability domain of this type of slings and arrows in a quantitative manner. Based on the results of the simulation and theoretical analysis, the size of the duck rudder control force has a significant impact on the stable attraction domain of the angular motion of the projectile and arrow. For example, in the paper, the attraction domain boundary decreases by approximately 16.9% when the rudder deflection angle δ_c is 20° compared to the uncontrolled state, while the control force azimuth has barely any effect. The lift coefficient derivative, the static moment coefficient derivative, and the nonlinear terms of the forebody and aft body Magnus moment coefficient derivative are the main factors affecting the stable attraction domain of the projectile and arrow in this type of motion. The nonlinear Magnus moment of the rear body has a significantly larger effect on the region of attraction of the angular motion of the projectile than that of the front body.

Keywords: self-rotating wraparound fins; projectile; angular motion; sum of squares programming; nonlinear aerodynamics; region of attraction

Received：2023-11-07；**Accepted**：2024-01-21；**Published Online**：2024-02-29 08：34

URL：link.cnki.net/urlid/11.2625.V.20240228.1632.002

Foundation item：National Natural Science Foundation of China (52202475)

* **Corresponding author**. E-mail：ballistics@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0736

充液航天器固定时间姿态滑模容错控制

范志文[1,2]，宋晓娟[1,3,*]，吕书锋[2]，岳宝增[4]

(1. 内蒙古工业大学 机械工程学院，呼和浩特 010051；2. 内蒙古工业大学 理学院，呼和浩特 010051；3. 内蒙古自治区机器人与智能装备技术重点实验室，呼和浩特 010051；4. 北京理工大学 宇航学院，北京 100081)

摘　　要：研究存在执行器饱和、执行器故障和外部扰动的液体大幅晃动充液航天器姿态系统，提出一种基于干扰观测器的滑模容错控制策略，实现姿态固定时间稳定。动力学建模过程中，将大幅晃动的液体燃料等效为运动脉动球模型，建立刚-液耦合航天器动力学方程。控制器设计过程中，设计自适应固定时间干扰观测器，保证控制系统中的综合扰动可以在固定时间内被估计到；在此基础上，利用滑模控制理论设计固定时间姿态滑模容错控制策略；此外，通过引入双曲正切函数，克服传统输入饱和的符号函数。该控制策略能够快速实现液体大幅晃动的充液航天器在执行器饱和、执行器故障和外部扰动的影响下姿态高精度稳定。将所提姿态滑模容错控制策略进行数值仿真与对比，验证控制策略的有效性和优越性。

关 键 词：充液航天器；姿态控制；干扰观测器；滑模控制；固定时间控制

中图分类号：V448

文献标志码：A　　**文章编号**：1001-5965(2026)01-0352-10

随着航天技术的蓬勃发展，航天器的探测范围和飞行时间都在明显增加，这也导致航天器所携带液体燃料的比重随之增多。由于航天器的平动和转动，会使液体不断晃动，对航天器产生显著的干扰力和力矩，进而对航天器系统的稳定性和姿态控制造成影响[1]。因此，研究液体晃动建模和充液航天器姿态稳定性有重要意义。

液体晃动是一种复杂且具有强非线性和随机性的流体运动现象。目前，等效力学模型方法是一种很好地研究液体晃动的手段。然而，一旦液体晃动幅度占充液腔体半径的比例达到25%以上，呈现大幅晃动，此时关于液体小幅晃动的等效力学模型[2-3]则不再适用。在此情况下，等效力学模型有运动脉动球模型[4-5](moving pulsating ball model, MPBM)、质心面模型[6]、3-DOF模型[7]等。其中，MPBM不仅被用在Sloshat FLEVO卫星上，而且在不同工况下进行了模拟试验，表明该模型可以准确且有效地等效液体大幅晃动。

近年来，滑模控制因其对外部干扰和模型不确定性具有很强的鲁棒性，成为许多不确定系统高性能控制问题的有效方法。在航天器姿态控制方面也有很多研究，然而大多研究都只能保证姿态系统渐进稳定或指数稳定[8-9]，无法满足航天器执行复杂任务控制的需求。为了克服该缺点，提出了有限时间控制方法。相比于渐进稳定或指数稳定控制，有限时间控制具有收敛速度更快和抗干扰能力强[10-11]的特性。文献[12]针对执行器故障情况下的航天器，设计了自适应非奇异终端滑模控制器，并且引入自适应律估计系统集总干扰的上界，实现有限时间内姿态跟踪并保持期望轨迹。文献[13]针对输入饱和的充液航天器姿态稳定问题，设计了抗干扰有限时间快速终端滑模控制器，并且设计了

收稿日期：2023-11-08；**录用日期**：2024-02-21；**网络出版时间**：2024-03-13 08:57

网络出版地址：link.cnki.net/urlid/11.2625.V.20240312.1520.002

基金项目：国家自然科学基金 (12362004,11962020,12172182,12132002); 英才兴蒙工程团队项目团队 (2025TEL02)；内蒙古自治区高等学校青年科技人才项目 (NJYT23029,NJYT23067)；内蒙古杰出青年基金项目 (2023JQ14)

* **通信作者**. E-mail：xjsong0603@163.com

引用格式：*范志文，宋晓娟，吕书锋，等. 充液航天器固定时间姿态滑模容错控制 [J]. 北京航空航天大学学报，2026，52（1）：352-361. FAN Z W，SONG X J，LYU S F，et al. Fixed-time attitude sliding mode fault-tolerant control for liquid-filled spacecraft[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：352-361 （in Chinese）.*

有限时间积分滑模干扰观测器估计系统的集总干扰。尽管有限时间控制可以获得更快的收敛速度和更好的抗干扰性能, 但其稳定时间上限取决于系统的初始状态, 这也意味着当系统的初始状态未知时, 很难估计到系统收敛时间上限。为了克服该问题, 提出了固定时间控制方法。与有限时间控制相比, 固定时间控制不仅不依赖系统的初始状态, 而且同样可以获得有限时间控制中优良的性能。文献 [14] 针对执行器故障情况下的挠性航天器姿态稳定, 设计了非奇异固定时间终端滑模控制器, 并且设计了固定时间观测器估计系统的集总干扰。文献 [15] 针对输入饱和的柔性航天器姿态稳定问题, 提出了一种改进分数阶非奇异终端滑模控制器。文献 [16] 研究了具有惯量不确定性和外部干扰下的刚性航天器姿态跟踪问题, 提出一种新的基于双极限齐次理论的固定时间滑模面。文献 [17] 研究了角速度不可测下的刚性航天器姿态稳定问题, 设计了一种连续非奇异固定时间滑模控制器。文献 [18] 研究了存在输入饱和的情形下航天器编队飞行的固定时间稳定性, 提出了一种非奇异饱和终端滑模控制器。文献 [19] 研究了具有惯性不确定和外部干扰的柔性航天器, 提出一种基于滑模的固定时间控制器。上述文献中所设计的控制器有的考虑执行器故障, 有的考虑控制器的饱和特性, 有的两者都没考虑。但实际工程中, 执行器故障和输入饱和经常发生, 忽视这些现象会导致意想不到的性能下降或系统不稳定[20], 因此, 所研究的控制器应满足这两者的情况。

为了提高系统控制精度和容错能力, 应对系统综合干扰进行估计。自适应算法估计综合干扰可以有效地抵抗干扰和噪声, 使系统更具鲁棒性。文献 [21-23] 研究了输入饱和下航天器姿态稳定问题, 将执行器故障、外部干扰和其他干扰等效成综合干扰, 采用自适应算法对其估计。然而, 自适应算法可能发生收敛缓慢问题。干扰观测器本质上不改变系统结构, 不依赖于模型, 也可简化控制器的设计。文献 [24] 对执行器故障构成的综合干扰用扩张状态观测器进行估计。文献 [25] 用具有时变增益的超扭观测器进行估计。文献 [26] 用模糊逻辑系统进行估计。然而, 干扰观测器如果估计不准确, 可能无法抵消干扰, 影响系统性能, 并且设计也相对复杂。

此外, 文献 [24-25] 均采用传统的符号函数方法解决出入饱和问题, 虽然符号函数方法可以最大限度地利用执行器输出能力, 但高频的切换指令不仅在实际任务中难以实现, 而且会激发液体晃动, 对系统造成影响。文献 [27] 对航天器姿态控制输入饱和问题的解决方法进行了梳理和归纳。

基于上述分析, 本文旨在为具有外部干扰、执行器故障和执行器饱和的液体大幅晃动充液航天器姿态稳定问题提供一种控制策略, 主要贡献如下: ①考虑自适应估计和干扰观测器的优缺点, 设计一种自适应固定时间干扰观测器。该观测器可以有效地估计由外部干扰、执行器故障、液体干扰构成的综合干扰, 并且消除干扰先验知识的假设, 保证了整个观测器控制系统的稳定性。②基于干扰观测器估计的信息, 提出一种超调量小、更简单的固定时间姿态滑模容错控制器, 并用双曲正切函数限制控制力矩的输出。该控制器的姿态和角速度可以在综合干扰、执行器饱和的影响下, 在任意初始状态下都能在固定时间内稳定。

1　充液航天器的动力学建模

1.1　姿态运动学方程描述

使用修正罗德里格参数 (modified Rodrigues parameters, MRPS) 描述航天器的姿态运动:

$$\dot{\boldsymbol{\rho}} = \boldsymbol{G}(\boldsymbol{\rho})\boldsymbol{\Omega} \tag{1}$$

$$\boldsymbol{G}(\boldsymbol{\rho}) = \frac{1}{2}\left(\boldsymbol{I}_3 + \boldsymbol{\rho}^{\times} + \boldsymbol{\rho}\boldsymbol{\rho}^{\mathrm{T}} - \frac{1}{2}\boldsymbol{I}_3(1+\boldsymbol{\rho}^{\mathrm{T}}\boldsymbol{\rho})\right) \tag{2}$$

式中: $\boldsymbol{\rho}\in\mathbf{R}^3$ 为 MRPS 描述的航天器姿态向量; $\boldsymbol{\Omega}\in\mathbf{R}^3$ 为航天器的角速度; $\boldsymbol{I}_3$ 为单位对角矩阵; "×"表示对应向量的斜对称矩阵, 如 $\boldsymbol{y}=[y_1,y_2,y_3]^{\mathrm{T}}$, 可以得到对应的斜对称矩阵:

$$\boldsymbol{y}^{\times} = \begin{bmatrix} 0 & -y_3 & y_2 \\ y_3 & 0 & -y_1 \\ -y_2 & y_1 & 0 \end{bmatrix} \tag{3}$$

1.2　充液航天器数学模型

如图 1 所示, 使用 MPBM[4] 描述液体大幅晃动的充液航天器。在 MPBM 中, 储箱等效为球形储箱, 液体等效为半径可变的均匀脉动球, 且始终与球形储箱有一个接触点, 在接触点处的晃动力和力矩用于近似液体对储箱产生的晃动力和力矩。在图 1 中, 航天器的本体坐标系为 $Cxyz$, 惯性坐标系为 $Ox_iy_iz_i$, C 为航天器的质心, C_t 为球形储箱的几何中心, S 为脉动球的质心, 在 MPBM 中脉动球可以在储箱内任意移动, 但始终与储箱保持接触, 接触点为 P 点。

本文主要研究充液航天器在姿态机动过程中的姿态控制器设计, 不考虑轨道平移运动。刚性航天器的姿态动力学模型描述为

$$\boldsymbol{J}\dot{\boldsymbol{\Omega}} = -\boldsymbol{\Omega}^{\times}(\boldsymbol{J}\boldsymbol{\Omega}) + \boldsymbol{u} + (\boldsymbol{R}\boldsymbol{e}+\boldsymbol{r}_{\mathrm{t}})^{\times}\boldsymbol{F}_{\mathrm{L}} + \boldsymbol{T}_{\mathrm{L}} + \boldsymbol{d} \tag{4}$$

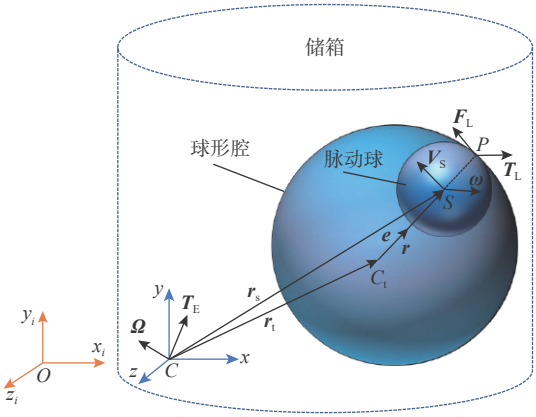


图1 MPBM模型示意图

Fig. 1 Schematic diagram of MPBM

式中：R为球形储箱的半径；$\boldsymbol{J}=\mathrm{diag}\{J_1,J_2,J_3\}$为惯性张量矩阵；$\boldsymbol{u}\in\mathbf{R}^3$为执行器实际输出控制力矩；$\boldsymbol{F}_{\mathrm{L}}\in\mathbf{R}^3$、$\boldsymbol{T}_{\mathrm{L}}\in\mathbf{R}^3$分别为液体对储箱产生的晃动力、力矩；$\boldsymbol{d}\in\mathbf{R}^3$为外部未知干扰力矩；$\boldsymbol{e}\in\mathbf{R}^3$为$C_{\mathrm{t}}$到$S$的单位矢量；$\boldsymbol{r}_{\mathrm{t}}\in\mathbf{R}^3$为$C$到$C_{\mathrm{t}}$的位置矢量。

液体等效为脉动球的平动和转动动力学方程为

$$m[\dot{\boldsymbol{V}}_{\mathrm{S}}+\dot{\boldsymbol{\Omega}}^{\times}(\boldsymbol{r}+\boldsymbol{r}_{\mathrm{t}})+\boldsymbol{\Omega}^{\times}(\boldsymbol{\Omega}^{\times}\boldsymbol{r})+2\boldsymbol{\Omega}^{\times}\boldsymbol{V}_{\mathrm{S}}+\boldsymbol{\Omega}^{\times}(\boldsymbol{\Omega}^{\times}\boldsymbol{r}_{\mathrm{t}})]=-\boldsymbol{F}_{\mathrm{L}} \tag{5}$$

$$\frac{2}{5}m(R-|\boldsymbol{r}|)[-2|\dot{\boldsymbol{r}}|(\boldsymbol{\omega}+\boldsymbol{\Omega})+(R-|\boldsymbol{r}|)(\dot{\boldsymbol{\omega}}+\dot{\boldsymbol{\Omega}}+\boldsymbol{\Omega}^{\times}\boldsymbol{\omega})]=-\boldsymbol{T}_{\mathrm{L}}-(R-|\boldsymbol{r}|)\boldsymbol{e}^{\times}\boldsymbol{F}_{\mathrm{L}} \tag{6}$$

式中：m为脉动球的质量；$\boldsymbol{V}_{\mathrm{S}}\in\mathbf{R}^3$和$\boldsymbol{\omega}\in\mathbf{R}^3$分别为脉动球相对于储箱的速度和角速度；$\boldsymbol{r}\in\mathbf{R}^3$为$C$到$S$的位置矢量。

在MPBM中，液体晃动力矩$\boldsymbol{T}_{\mathrm{L}}$描述为[5]

$$\boldsymbol{T}_{\mathrm{L}}=-0.36m^{\frac{4}{3}}\rho_{\mathrm{s}}^{\frac{1}{6}}\mu^{\frac{1}{2}}\left(\frac{L}{L_{\min}}\right)^2\|\boldsymbol{\omega}\|^{\frac{1}{2}}\cdot[t_{\mathrm{sr}}\boldsymbol{\omega}_{\mathrm{r}}+(1-t_{\mathrm{sr}})(\boldsymbol{\omega}-\boldsymbol{\omega}_{\mathrm{r}})] \tag{7}$$

式中：μ为液体的运动黏性系数；$L=R-|\boldsymbol{r}|$为脉动球的半径；$L_{\min}$为脉动球的最小半径；$\boldsymbol{\omega}_{\mathrm{r}}$为$\boldsymbol{\omega}$沿$\boldsymbol{e}$的分量；$t_{\mathrm{sr}}$为脉动球绕轴转动和滚动转动的权值，满足$0\leqslant t_{\mathrm{sr}}\leqslant 1$；$\rho_{\mathrm{s}}=3m/(4\pi L^3)$为脉动球的密度。

液体晃动力$\boldsymbol{F}_{\mathrm{L}}$由法向力$N\boldsymbol{e}$和摩擦力$\boldsymbol{F}_{\mathrm{b}}$组成[5]，即

$$\boldsymbol{F}_{\mathrm{L}}=N\boldsymbol{e}+\boldsymbol{F}_{\mathrm{b}} \tag{8}$$

$$N=\frac{3m}{8}\left\{|\boldsymbol{r}|\left[\boldsymbol{e}^{\times}\left(\boldsymbol{\Omega}+\frac{\boldsymbol{e}^{\times}\boldsymbol{r}}{|\boldsymbol{r}|}\right)\right]^2+(\boldsymbol{\Omega}^{\times}\boldsymbol{r}_{\mathrm{t}})(\boldsymbol{\Omega}^{\times}\boldsymbol{e})-(\boldsymbol{r}_{\mathrm{t}}{}^{\times}\boldsymbol{e})\dot{\boldsymbol{\Omega}}\right\}+\frac{m}{4}(R-|\boldsymbol{r}|)(\boldsymbol{\Omega}+\boldsymbol{\omega})^2-5\pi\sigma(R-|\boldsymbol{r}|) \tag{9}$$

$$\boldsymbol{F}_{\mathrm{b}}=\frac{5\mu m}{L^2}[\boldsymbol{e}^{\times}(\boldsymbol{e}^{\times}\boldsymbol{V}_{\mathrm{S}}+L\boldsymbol{\omega})] \tag{10}$$

式中：σ为液体表面张力系数。

考虑执行器发生故障和受到饱和值限制2种情形，对存在故障的执行器进行建模，有

$$\boldsymbol{u}=\boldsymbol{E}\boldsymbol{\tau}+\bar{\boldsymbol{\tau}} \tag{11}$$

式中：$\boldsymbol{\tau}$为控制指令力矩；$\bar{\boldsymbol{\tau}}$为故障引起的控制力矩偏差；$\boldsymbol{E}=\mathrm{diag}\{E_1,E_2,\cdots,E_n\}$为执行器的控制效率矩阵，$E_i$为第$i$个执行机构的失效系数，且有$0\leqslant E_i\leqslant 1(i=1,2,\cdots,n)$，当$E_i=0$时，表示执行机构完全失效，当$E_i=1$时，表示执行机构无失效故障，当$0<E_i<1$时，表示执行机构部分失效。

执行器的输入饱和表示为$\mathrm{sat}(\boldsymbol{\tau})=[\mathrm{sat}(\tau_1),\mathrm{sat}(\tau_2),\mathrm{sat}(\tau_3)]^{\mathrm{T}}$，其中，$\mathrm{sat}(\tau_i),i=1,2,3$定义为

$$\mathrm{sat}(\tau_i)=\tau_{i\max}\tanh\left(\frac{\tau_i}{\tau_{i\max}}\right) \tag{12}$$

为了简单起见，输入饱和项重写为

$$\mathrm{sat}(\boldsymbol{\tau})=\Theta(\boldsymbol{\tau})\boldsymbol{\tau} \tag{13}$$

式中：$\Theta(\boldsymbol{\tau})=\mathrm{diag}(\Theta_1(\tau_i),\Theta_2(\tau_i),\Theta_3(\tau_i))$，并且$\Theta_i(\tau_i),i=1,2,3$定义为$\Theta_i(\tau_i)=\dfrac{\tau_{i\max}}{\tau_i}\tanh\left(\dfrac{\tau_i}{\tau_{i\max}}\right)$。

根据执行器故障式(11)和输入饱和式(13)，航天器姿态动力学方程可以进一步写为

$$\boldsymbol{J}\dot{\boldsymbol{\Omega}}=-\boldsymbol{\Omega}^{\times}(\boldsymbol{J}\boldsymbol{\Omega})+\Theta(\boldsymbol{\tau})\boldsymbol{\tau}+\boldsymbol{T}_{\mathrm{d}} \tag{14}$$

式中：$\boldsymbol{T}_{\mathrm{d}}=(\boldsymbol{E}-\boldsymbol{I}_{n\times n})\boldsymbol{\tau}+\bar{\boldsymbol{\tau}}+(R\boldsymbol{e}+\boldsymbol{r}_{\mathrm{t}})^{\times}\boldsymbol{F}_{\mathrm{L}}+\boldsymbol{T}_{\mathrm{L}}+\boldsymbol{d}$看作系统的综合干扰量，由系统的执行器故障力矩、液体晃动对充液航天器姿态系统产生的全部干扰力矩和外部干扰构成。

考虑式(1)和式(14)中的姿态运动学和动力学，可以得到姿态动力学的欧拉-拉格朗日公式，其形式为

$$\boldsymbol{A}\ddot{\boldsymbol{\rho}}=\boldsymbol{B}\dot{\boldsymbol{\rho}}+\boldsymbol{G}^{-\mathrm{T}}\Theta(\boldsymbol{\tau})\boldsymbol{\tau}+\boldsymbol{G}^{-\mathrm{T}}\boldsymbol{T}_{\mathrm{d}} \tag{15}$$

式中：$\boldsymbol{G}$表示$\boldsymbol{G}(\boldsymbol{\rho})$；$\boldsymbol{A}=\boldsymbol{G}^{-\mathrm{T}}\boldsymbol{I}\boldsymbol{G}^{-1}$；$\boldsymbol{B}=\boldsymbol{G}^{-\mathrm{T}}\boldsymbol{I}\boldsymbol{G}^{-1}\dot{\boldsymbol{G}}\boldsymbol{G}^{-1}-\boldsymbol{G}^{-\mathrm{T}}(\boldsymbol{G}^{-1}\dot{\boldsymbol{\rho}})^{\times}\boldsymbol{I}\boldsymbol{G}^{-1}$。

定义$\boldsymbol{x}_1=\boldsymbol{\rho}$、$\boldsymbol{x}_2=\dot{\boldsymbol{\rho}}$为2种状态，则由式(15)得

$$\begin{cases}\dot{\boldsymbol{x}}_1=\boldsymbol{x}_2\\ \dot{\boldsymbol{x}}_2=\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}_2+\boldsymbol{A}^{-1}\boldsymbol{G}^{-\mathrm{T}}\Theta(\boldsymbol{\tau})\boldsymbol{\tau}+\boldsymbol{\delta}\end{cases} \tag{16}$$

式中：$\boldsymbol{\delta}=\boldsymbol{A}^{-1}\boldsymbol{G}^{-\mathrm{T}}\boldsymbol{T}_{\mathrm{d}}$为系统的综合干扰。

因此，将所研究的系统式(1)和式(14)的姿态控制问题转化为系统式(16)的稳定问题。

控制目标：针对航天器系统存在执行器饱和、执行器故障、外部扰动和液体晃动等影响因素，通过设计自适应固定时间干扰观测器估计系统的综合扰动值，进而基于固定时间控制理论设计滑模容错控制器，使系统的姿态可以在固定时间内稳定。

注1 执行器故障形式考虑执行器发生效率损伤和加性故障。

注2 输入饱和函数采用全局可微的双曲正切函数代替传统的饱和函数。这不仅能满足实际饱

和约束, 还能为控制器的设计带来便利。双曲正切函数利用函数值域本身限幅, 饱和约束的输出值略小于实际饱和幅值, 且随着输入增加, 输出能够较快地无限接近于实际幅值。能够充分利用执行机构的输出能力。

2　姿态控制器设计

假设 1　综合干扰$\boldsymbol{\delta}$是有界的, 且存在一个正常数k_3, 满足$\|\boldsymbol{\delta}\| \leqslant k_3$。

引理 1[28]　考虑一个非线性系统$\dot{\boldsymbol{x}}(t)=f(\boldsymbol{x}(t))$, $\boldsymbol{x}(0)=\boldsymbol{x}_0$, 若存在一个连续正定且径向无界的函数$V(\boldsymbol{x})$满足:

$$\dot{V} \leqslant -\alpha_1 V^{\beta_1} - \alpha_2 V^{\beta_2} \tag{17}$$

式中: $\alpha_1>0$, $\alpha_2>0$, $0<\beta_1<1$,$\beta_2>1$。则系统是固定时间稳定的, 稳定时间T_1满足以下不等式:

$$T_1 \leqslant \frac{1}{\alpha_1(1-\beta_1)}+\frac{1}{\alpha_2(\beta_2-1)} \tag{18}$$

作为推广, 如果存在一个 Lyapunov 函数$\dot{V}(\boldsymbol{x})$也满足:

$$\dot{V} \leqslant -\alpha_1 V^{\beta_1} - \alpha_2 V^{\beta_2} + \chi \tag{19}$$

式中: α_1、α_2均为正常数, $0<\chi<\infty$, $0<\beta_1<1$, $\beta_2>1$。此时, 系统也可称为固定时间稳定系统, 引入正常数$0<\theta<1$, 则稳定时间T_2满足以下不等式:

$$T_2 \leqslant \frac{1}{\alpha_1\theta(1-\beta_1)}+\frac{1}{\alpha_2\theta(\beta_2-1)} \tag{20}$$

引理 2[29]　下面的不等式对任意的$\nu>0$和$\eta\in\mathbf{R}$均成立:

$$0 \leqslant |\eta| - \eta\tanh\left(\frac{\eta}{\nu}\right) \leqslant \kappa\nu \tag{21}$$

式中: $\kappa=\exp(-\kappa-1)$。

引理 3[30]　根据$\hat{a}=-\tilde{a}+a$, 当任意正定实数满足$\zeta>0.5$, 可以得到以下不等式成立:

$$\lambda\tilde{a}\hat{a}=\lambda\tilde{a}(-\tilde{a}+a)=\lambda(-\tilde{a}^2+\tilde{a}a) \leqslant \lambda\left(-\tilde{a}^2+\frac{1}{2\zeta}\tilde{a}^2+\frac{\zeta}{2}a^2\right) \leqslant \frac{-\lambda(2\zeta-1)}{2\zeta}\tilde{a}^2+\frac{\lambda\zeta}{2}a^2 \tag{22}$$

式中: $\lambda>0$; $\hat{a}$为a的估计值; $\tilde{a}$为a和$\hat{a}$的误差, 满足: $\tilde{a}=a-\hat{a}$。

引理 4[31]　对于任意实数$\eta_i, i=1,\cdots,n$和$0<b<1$, 以下不等式成立:

$$(|\eta_1|+|\eta_2|+\ldots+|\eta_n|)^b \leqslant |\eta_1|^b+|\eta_2|^b+\ldots+|\eta_n|^b \tag{23}$$

2.1　自适应固定时间干扰观测器设计

考虑集总干扰$\boldsymbol{\delta}$, 引入辅助变量$\boldsymbol{z}$, 自适应固定时间干扰观测器设计如下:

$$\dot{\boldsymbol{z}}=\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}_2+\boldsymbol{A}^{-1}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{\Theta}(\tau)\boldsymbol{\tau}+k_1\mathrm{sig}^{1+\gamma_1}(\boldsymbol{x}_2-\boldsymbol{z})+k_2\mathrm{sig}^{1-\gamma_1}(\boldsymbol{x}_2-\boldsymbol{z})+\hat{k}_3\tanh\left(\frac{\boldsymbol{x}_2-\boldsymbol{z}}{\varepsilon}\right) \tag{24}$$

式中: $k_1>0$; $k_2>0$; $0<\gamma_1<1$; ε为正常数; $\hat{k}_3$为自适应估计综合干扰$\boldsymbol{\delta}$的上界, 其表达式为

$$\dot{\hat{k}}_3=k_4\left[(\boldsymbol{x}_2-\boldsymbol{z})^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{x}_2-\boldsymbol{z}}{\varepsilon}\right)-k_5\hat{k}_3\right] \tag{25}$$

其中: $k_4>0$; $k_5>0$。

定义干扰观测器的误差如下:

$$\boldsymbol{\varpi}=\boldsymbol{x}_2-\boldsymbol{z} \tag{26}$$

对式 (26) 求导可以得到

$$\dot{\boldsymbol{\varpi}}=\boldsymbol{\delta}-\hat{k}_3\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)-k_1\mathrm{sig}^{1+\gamma_1}(\boldsymbol{\varpi})-k_2\mathrm{sig}^{1-\gamma_1}(\boldsymbol{\varpi}) \tag{27}$$

定理 1　考虑系统式 (16), 基于式 (24)～式 (27) 的自适应固定时间干扰观测器可以保证系统是有界的, 干扰观测器误差$\boldsymbol{\varpi}$可以在固定时间收敛到原点附近的一个邻域内。

证明　定义自适应误差$\tilde{k}_3=k_3-\hat{k}_3$, 选择 Lyapunov 函数如下:

$$V_1=\frac{1}{2}\boldsymbol{\varpi}^{\mathrm{T}}\boldsymbol{\varpi}+\frac{1}{2k_4}\tilde{k}_3^2 \tag{28}$$

对式 (28) 求导, 并代入式 (27), 得到

$$\dot{V}_1=\boldsymbol{\varpi}^{\mathrm{T}}\left[\boldsymbol{\delta}-\hat{k}_3\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)-k_1\mathrm{sig}^{1+\gamma_1}(\boldsymbol{\varpi})-k_2\mathrm{sig}^{1-\gamma_1}(\boldsymbol{\varpi})\right]-\frac{1}{k_4}\tilde{k}_3\dot{\hat{k}}_3 \tag{29}$$

利用引理 2 和假设 1, 可以得到

$$\boldsymbol{\varpi}^{\mathrm{T}}\boldsymbol{\delta} \leqslant \sum_{i=1}^{3}|\varpi_i|k_3 \leqslant \sum_{i=1}^{3}k_3\varpi_i\tanh\left(\frac{\varpi_i}{\varepsilon}\right)+\varepsilon_1 = k_3\boldsymbol{\varpi}^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)+\varepsilon_1 \tag{30}$$

式中: $\varepsilon_1=\gamma_0\varepsilon k_3$, $\gamma_0=\exp(-\gamma_0-1)$。

可以将$\dot{V}_1$重写为

$$\begin{aligned}\dot{V}_1 &\leqslant -k_1\boldsymbol{\varpi}^{\mathrm{T}}\mathrm{sig}^{1+\gamma_1}(\boldsymbol{\varpi})-k_2\boldsymbol{\varpi}^{\mathrm{T}}\mathrm{sig}^{1-\gamma_1}(\boldsymbol{\varpi})-\\ &\quad\frac{1}{k_4}\tilde{k}_3\dot{\hat{k}}_3+k_3\boldsymbol{\varpi}^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)-\hat{k}_3\boldsymbol{\varpi}^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)+\varepsilon_1 \leqslant\\ &\quad -k_1\boldsymbol{\varpi}^{\mathrm{T}}\mathrm{sig}^{1+\gamma_1}(\boldsymbol{\varpi})-k_2\boldsymbol{\varpi}^{\mathrm{T}}\mathrm{sig}^{1-\gamma_1}(\boldsymbol{\varpi})-\\ &\quad\tilde{k}_3\left(\boldsymbol{\varpi}^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)-k_5\hat{k}_3\right)+k_3\boldsymbol{\varpi}^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)-\\ &\quad\hat{k}_3\boldsymbol{\varpi}^{\mathrm{T}}\tanh\left(\frac{\boldsymbol{\varpi}}{\varepsilon}\right)+\varepsilon_1 \leqslant -k_1\boldsymbol{\varpi}^{\mathrm{T}}\mathrm{sig}^{1+\gamma_1}(\boldsymbol{\varpi})-\\ &\quad k_2\boldsymbol{\varpi}^{\mathrm{T}}\mathrm{sig}^{1-\gamma_1}(\boldsymbol{\varpi})+k_5\tilde{k}_3\hat{k}_3+\varepsilon_1 \leqslant\\ &\quad -k_1\|\boldsymbol{\varpi}\|^{2+\gamma_1}-k_2\|\boldsymbol{\varpi}\|^{2-\gamma_1}+k_5\tilde{k}_3\hat{k}_3+\varepsilon_1\end{aligned} \tag{31}$$

根据引理 3, 进一步将式 (31) 整理为

$$\dot{V}_1 \leqslant -k_1\|\boldsymbol{\varpi}\|^{2+\gamma_1} - k_2\|\boldsymbol{\varpi}\|^{2-\gamma_1} - \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2+\gamma_1)/2} + \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2+\gamma_1)/2} - \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2-\gamma_1)/2} + \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2-\gamma_1)/2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} + \varepsilon_1 \leqslant -r_1\left[\left(\frac{1}{2}\|\boldsymbol{\varpi}\|^2\right)^{(2+\gamma_1)/2} + \left(\frac{1}{2k_4}\tilde{k}_3^2\right)^{(2+\gamma_1)/2}\right] + \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2+\gamma_1)/2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} - r_2\left[\left(\frac{1}{2}\|\boldsymbol{\varpi}\|^2\right)^{(2-\gamma_1)/2} + \left(\frac{1}{2k_4}\tilde{k}_3^2\right)^{(2-\gamma_1)/2}\right] + \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2-\gamma_1)/2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} + \varepsilon_1 \tag{32}$$

式中：$k_4 = \max\left(\frac{2\zeta r_1^{2/(2+\gamma_1)}}{k_5(2\zeta-1)}, \frac{2\zeta r_2^{2/(2-\gamma_1)}}{k_5(2\zeta-1)}\right)$；$r_1 = 2^{(2+\gamma_1)/2}k_1$；$r_2 = 2^{(2-\gamma_1)/2}k_2$。

根据引理4可以将式(32)写为

$$\dot{V}_1 \leqslant -r_1V_1^{(2+\gamma_1)/2} - r_2V_1^{(2-\gamma_1)/2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} + \varepsilon_1 + \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2+\gamma_1)/2} + \left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2-\gamma_1)/2} \tag{33}$$

选择合适的参数来保证建立的不等式$\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2 \leqslant 1$，则可以得到

$$\left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2+\gamma_1)/2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} \leqslant \frac{k_5(2\zeta-1)}{4\rho_1}\tilde{k}_3^2 + \frac{k_5\tilde{k}_3\hat{k}_3}{2} \leqslant \frac{k_5\zeta}{4}k_3^2 \tag{34}$$

同理可得

$$\left[\frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2\right]^{(2-\gamma_1)/2} + \frac{k_5\tilde{k}_3\hat{k}_3}{2} \leqslant \frac{k_5(2\zeta-1)}{4\zeta}\tilde{k}_3^2 + \frac{k_5\tilde{k}_3\hat{k}_3}{2} \leqslant \frac{k_5\zeta}{4}k_3^2 \tag{35}$$

将式(34)和式(35)代入式(33)，可以将$\dot{V}_1$重写为

$$\dot{V}_1 \leqslant -r_1V_1^{(2+\gamma_1)/2} - r_2V_1^{(2-\gamma_1)/2} + \varepsilon_2 \tag{36}$$

式中：$\varepsilon_2 = \frac{k_5\zeta}{2}k_3^2 + \varepsilon_1$。

根据引理1和式(36)，系统实际是固定时间收敛的，则自适应固定时间观测器的收敛时间满足：$T_3 \leqslant \frac{1}{r_1\theta(1-\upsilon_1)} + \frac{1}{r_2\theta(\upsilon_2-1)}$，其中，$0<\theta<1$，$\upsilon_1 = (2+\gamma_1)/2$，$\upsilon_2 = (2-\gamma_1)/2$。**证毕**

2.2 固定时间姿态滑模容错控制器设计

为了提高收敛性，提出一种新的非奇异固定时间终端滑动面，表示为

$$\boldsymbol{S} = \boldsymbol{x}_2 + H(\boldsymbol{x}_1)(l_1\mathrm{sig}^{1+\gamma_2}(\boldsymbol{x}_1) + l_2\boldsymbol{S}_{\mathrm{C}}) \tag{37}$$

式中：$\boldsymbol{S}\in\mathbf{R}^3$；$0<\gamma_2<1$；$l_1>0$；$l_2>0$；$H(\boldsymbol{x}_1) = 1+(c_1/\pi\tan\mathrm{h}(c_2\|\boldsymbol{x}_1\|^{c_3}))$，$c_1>0$，$c_2>0$，$c_3$为偶数；$\boldsymbol{S}_{\mathrm{C}}\in\mathbf{R}^3$，定义为

$$S_{\mathrm{C}i} = \begin{cases} \mathrm{sig}^{1-\gamma_2}(x_{1i}) & \bar{S}_i = 0 \text{ or } \bar{S}_i \neq 0, |x_{1i}| \geqslant \phi \\ b_1x_{1i} + b_2\mathrm{sgn}(x_{1i})x_{1i}^2 & \bar{S}_i \neq 0, |x_{1i}| < \phi \end{cases} \tag{38}$$

式中：$0<\phi<1$；$b_1 = (1+\gamma_2)\phi^{-\gamma_2}$；$b_2 = -\gamma_2\phi^{(-\gamma_2-1)}$；$\bar{\boldsymbol{S}} = \boldsymbol{x}_2 + H(\boldsymbol{x}_1)(l_1\mathrm{sig}^{1+\gamma_2}(\boldsymbol{x}_1) + l_2\mathrm{sig}^{1-\gamma_2}(\boldsymbol{x}_1))$。

注3 在式(37)中，$H(\boldsymbol{x}_1)$表示收敛速度的时变增益，合理选取γ_2，使其满足引理1的定时稳定性。此外在式(38)中，参数ϕ对系统状态$\boldsymbol{x}_1$、$\boldsymbol{x}_2$的收敛精度有直接影响，ϕ的选取直接影响滑模面奇异性问题。当$\phi=0$时，所提出的滑模面式(38)成为传统的终端滑模面，导致奇异性问题。

定理2 考虑充液航天器系统简化式(14)，一旦非奇异固定时间终端滑动面实现理想的滑模运动($\boldsymbol{S}=\bar{\boldsymbol{S}}=\boldsymbol{0}$)，无论初始姿态如何，系统的状态变量将在固定时间内收敛至零。

证明 如果$\boldsymbol{S}=\bar{\boldsymbol{S}}=\boldsymbol{0}$，则可以得到

$$\boldsymbol{x}_2 = -H(\boldsymbol{x}_1)(l_1\mathrm{sig}^{1+\gamma_2}(\boldsymbol{x}_1) + l_2\mathrm{sig}^{1-\gamma_2}(\boldsymbol{x}_1)) \tag{39}$$

考虑Lyapunov函数$V_2 = \frac{1}{2}\boldsymbol{x}_1^{\mathrm{T}}\boldsymbol{x}_1$，并对其求一阶导，得到

$$\dot{V}_2 = \boldsymbol{x}_1^{\mathrm{T}}\dot{\boldsymbol{x}}_1 = -\boldsymbol{x}_1^{\mathrm{T}}[H(\boldsymbol{x}_1)(l_1\mathrm{sig}^{1+\gamma_2}(\boldsymbol{x}_1) + l_2\mathrm{sig}^{1-\gamma_2}(\boldsymbol{x}_1)] = -H(\boldsymbol{x}_1)\left(l_1\sum_{i=1}^{3}|x_{1i}|^{2+\gamma_2} + l_2\sum_{i=1}^{3}|x_{1i}|^{2-\gamma_2}\right) \leqslant -r_3V_2^{(2+\gamma_2)/2} - r_4V_2^{(2-\gamma_2)/2} \tag{40}$$

式中：$r_3 = 2^{(2+\gamma_2)/2}l_1H(\boldsymbol{x}_1)$；$r_4 = 2^{(2-\gamma_2)/2}l_2H(\boldsymbol{x}_1)$。

根据引理1，充液航天器系统简化式将在固定时间内收敛，其收敛时间为$T_4 \leqslant \frac{1}{r_3(1-\upsilon_3)} + \frac{1}{r_4(\upsilon_4-1)}$，其中，$\upsilon_3 = (2+\gamma_2)/2$，$\upsilon_4 = (2-\gamma_2)/2$。**证毕**

使用非奇异固定时间终端滑动面式(37)和充液航天器简化系统式(14)，有

$$\dot{\boldsymbol{S}} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}_2 + \boldsymbol{A}^{-1}\boldsymbol{G}^{-\mathrm{T}}\boldsymbol{\Theta}(\boldsymbol{\tau})\boldsymbol{\tau} + \boldsymbol{\delta} + \boldsymbol{\Lambda} \tag{41}$$

式中：

$$\boldsymbol{\Lambda} = \dot{H}(\boldsymbol{x}_1)(l_1\mathrm{sig}^{1+\gamma_2}(\boldsymbol{x}_1) + l_2\boldsymbol{S}_{\mathrm{C}}) + H(\boldsymbol{x}_1)\left(l_1(1+\gamma_2)\mathrm{diag}(|x_{1i}|^{\gamma_2})\dot{\boldsymbol{x}}_1 + l_2\dot{\boldsymbol{S}}_{\mathrm{C}}\right) \tag{42}$$

$$\dot{S}_{\mathrm{C}i} = \begin{cases} (1-\gamma_1)\dot{x}_{1i}|x_{1i}|^{-\gamma_2} & \bar{S}_i = 0 \text{ or } \bar{S}_i \neq 0, |x_{1i}| \geqslant \phi \\ b_1\dot{x}_{1i} + 2b_2\dot{x}_{1i}|x_{1i}| & \bar{S}_i \neq 0, |x_{1i}| < \phi \end{cases} \tag{43}$$

为了提高姿态稳定机动的控制精度，设计以下基于干扰观测器的固定时间滑模容错控制律：

$$\boldsymbol{\tau} = -\boldsymbol{G}^{\mathrm{T}}\boldsymbol{A}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}_2 + \hat{\boldsymbol{\delta}} + \boldsymbol{\Lambda} + H(\boldsymbol{S})(g_1\mathrm{sig}^{1+\gamma_3}(\boldsymbol{S}) + g_2\mathrm{sig}^{1-\gamma_3}(\boldsymbol{S}) + g_3\boldsymbol{S})\right) \tag{44}$$

式中: $g_i>0(i=1,2,3)$; $0<\gamma_3<1$。

定理 3　对于充液航天器系统简化式 (14), 基于干扰观测器的固定时间滑模容错控制律式 (44), 可使闭环姿态控制系统在固定时间内到达$\boldsymbol{S}=\boldsymbol{0}$的小邻域内。进一步, $\boldsymbol{x}_1$将被驱动到原点周围的区域, 即充液航天器可以在固定时间内稳定。

证明　选择 Lyapunov 函数 $V_3=\frac{1}{2}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{S}$, 对其求一阶导, 并将式 (41) 和式 (44) 代入可得

$$\dot{V}_3=\boldsymbol{S}^{\mathrm{T}}\dot{\boldsymbol{S}}=\boldsymbol{S}^{\mathrm{T}}\left(\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{x}_2+\boldsymbol{A}^{-1}\boldsymbol{G}^{-\mathrm{T}}\boldsymbol{\Theta}(\boldsymbol{\tau})\boldsymbol{\tau}+\boldsymbol{\delta}+\boldsymbol{\Lambda}\right)\leqslant \boldsymbol{S}^{\mathrm{T}}\left(\boldsymbol{\delta}-\hat{\boldsymbol{\delta}}-H(\boldsymbol{S})(g_1\mathrm{sig}^{1+\gamma_3}(\boldsymbol{S})+g_2\mathrm{sig}^{1-\gamma_3}(\boldsymbol{S})+g_3\boldsymbol{S})\right) \tag{45}$$

由定理 1 可以推导出$\lim\limits_{t\to T_3}\hat{\boldsymbol{\delta}}=\boldsymbol{\delta}$, 因此, 式 (45) 可以进一步简化写为

$$\dot{V}_3\leqslant H(\boldsymbol{S})\boldsymbol{S}^{\mathrm{T}}\left(-g_1\mathrm{sig}^{1+\gamma_3}(\boldsymbol{S})-g_2\mathrm{sig}^{1-\gamma_3}(\boldsymbol{S})-g_3\boldsymbol{S}\right)\leqslant -g_1H(\boldsymbol{S})\left(\|\boldsymbol{S}\|^2\right)^{(2+\gamma_3)/2}-g_2H(\boldsymbol{S})\left(\|\boldsymbol{S}\|^2\right)^{(2-\gamma_3)/2}\leqslant -r_5V_3^{(2+\gamma_3)/2}-r_6V_3^{(2-\gamma_3)/2} \tag{46}$$

式中: $r_5=2^{(2+\gamma_3)/2}g_1H(\boldsymbol{S})$; $r_6=2^{(2-\gamma_3)/2}g_2H(\boldsymbol{S})$。

根据引理 1, 滑模面$\boldsymbol{S}$将在固定时间内收敛到$\boldsymbol{S}=\boldsymbol{0}$的小领域区间范围内。其固定时间为$T_5\leqslant\frac{1}{r_5(1-\upsilon_5)}+\frac{1}{r_6(\upsilon_6-1)}$, 其中, $\upsilon_5=(2+\gamma_3)/2$, $\upsilon_6=(2-\gamma_3)/2$。进一步, 利用定理 2, 可以得到系统状态$\boldsymbol{x}_1$也将在固定时间内稳定。因此, 在固定时间$T_4+T_5$内, 当航天器姿态机动从任意初始条件开始时, 本文提出的基于观测器的控制策略可以使状态$\boldsymbol{x}_1$稳定在零的附近小邻域内。　**证毕**

注 4　本文滑模面的设计受文献 [22] 的启发。式 (37) 中的可调函数$H(\boldsymbol{x}_1)$采用双曲正切函数 tanh进行设计, 而文献 [22] 采用反三角正切函数 arctan设计与其相比, 采用双曲正切函数设计更精确、更平整。而滑模面另一部分的设计与其相比更简单, 参数选择少, 且都具有固定时间收敛性。

3　仿真分析

本文充液航天器的参数参考 Sloshsat 卫星[5], 如表 1 所示。本文控制力矩限制在 1 N·m 内, 执行器故障设置为: 在 50 s 内, x轴执行器同时发生失效和偏差故障, 其中, 失效系数$E_1=0.7$, 偏差故障$\bar{\boldsymbol{\tau}}=0.1$ N·m。自适应固定时间干扰观测器的参数设置为$k_1=0.1$, $k_2=15$, $k_4=1$, $k_5=0.1$, $\gamma_1=0.1$, $\varepsilon=0.01$。控制器的参数设置为$l_1=0.1$, $l_2=0.1$, $\gamma_2=0.3$, $\gamma_3=0.1$, $c_1=0.01$, $c_2=6$, $c_3=2$, $g_1=0.1$, $g_2=0.001$, $g_3=30$, $\phi=0.001$。

表 1　充液航天器的参数

Table 1　Liquid-filled spacecraft parameters

$\mu/(\mathrm{m}^2\cdot\mathrm{s}^{-1})$	$\sigma/(\mathrm{N}\cdot\mathrm{m}^{-1})$	R/m	m/kg	$\boldsymbol{r}_{\mathrm{t}}/\mathrm{m}$	$\boldsymbol{I}_0/(\mathrm{kg}\cdot\mathrm{m}^2)$
10^{-6}	0.000 5	0.64	100	$[-0.05,-0.11,0.11]^{\mathrm{T}}$	diag(20,5,15)
$\boldsymbol{\Omega}(0)/(\mathrm{rad}\cdot\mathrm{s}^{-1})$	$\boldsymbol{\rho}(0)$	$\boldsymbol{r}_{\mathrm{s}}(0)/\mathrm{m}$	$\boldsymbol{V}_{\mathrm{S}}(0)/(\mathrm{m}\cdot\mathrm{s}^{-1})$	$\boldsymbol{\omega}(0)/(\mathrm{rad}\cdot\mathrm{s}^{-1})$	$\boldsymbol{d}/(\mathrm{N}\cdot\mathrm{m})$
$[0.1,0,0]^{\mathrm{T}}$	$[0.282\,0,-0.612\,7,0.402\,9]^{\mathrm{T}}$	$[0,-0.19,0]^{\mathrm{T}}$	$[0,0,0]^{\mathrm{T}}$	$[0,0,0]^{\mathrm{T}}$	$\begin{bmatrix}\cos 0.2t\\ \sin 0.5t\\ \sin 0.1t\end{bmatrix}\times10^{-3}$

为验证控制方案的有效性和优越性, 对以下 2 种情况进行仿真。

1) Case1。采用文献 [22] 的自适应固定时间滑模容错控制器与本文的基于观测器的固定时间滑模容错控制器作对比, 仿真结果如图 2～图 7 所示。

2) Case2。为解决航天器在实际应用中存在的问题, 考虑系统受到随机的外部扰动及敏感器和执行机构输出存在误差, 进行仿真模拟。敏感器和执行机构的输出误差详细见文献 [32-35], 外部扰动设为 ±0.001 N·m的随机数, MRPS 的误差设为 ±0.001 的随机数, 控制力矩的误差设为 ±0.01 N·m的随机数。仿真结果如图 8～图 12 所示。

3.1　Case1 仿真

图 2～图 6 中, AFTDO-AFTSMC 表示本文提出的基于观测器的固定时间滑模容错控制器, AFTSMC 表示文献 [22] 的自适应固定时间滑模容错控制器。

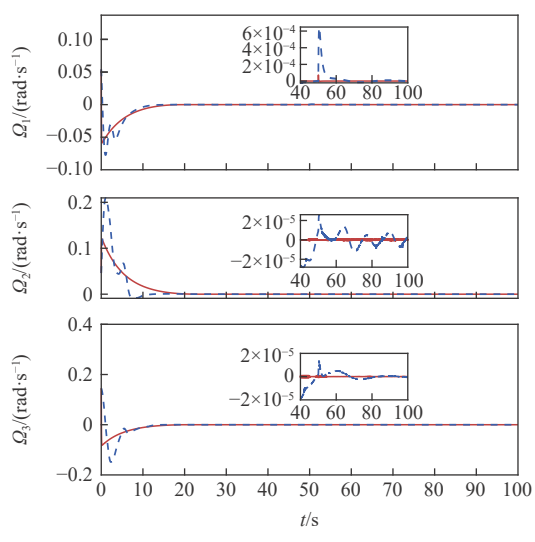


图 2　Case 1 的航天器角速度变化曲线

Fig. 2　Curves of spacecraft angular velocities in Case 1

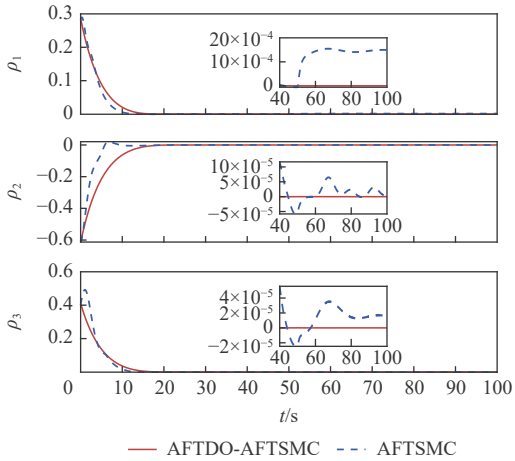


图3 Case 1的航天器MRPS变化曲线

Fig. 3 Curves of spacecraft MRPS in Case 1

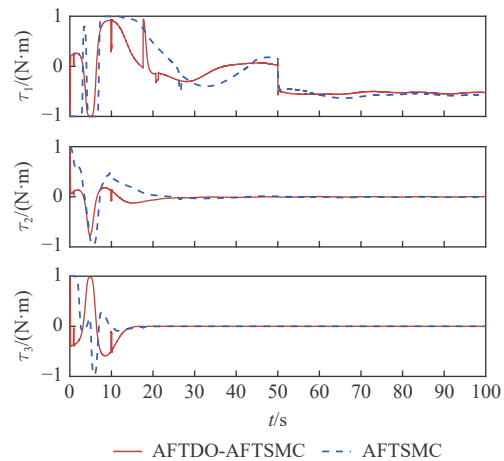


图4 Case 1的航天器控制力矩曲线

Fig. 4 Curves of spacecraft control torques in Case 1

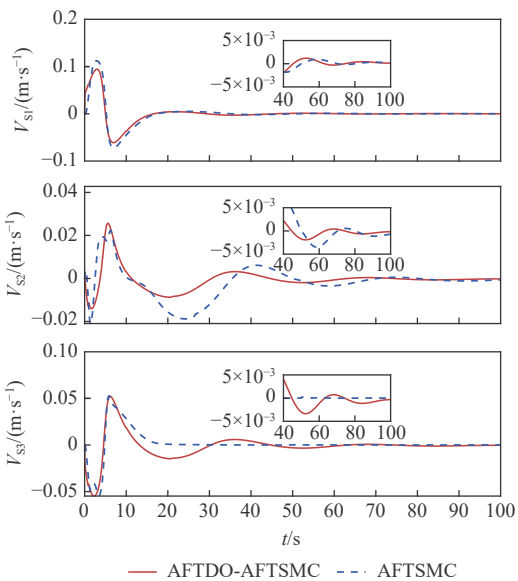


图5 Case 1的脉动球速度变化曲线

Fig. 5 Curves of the pulsating ball velocities in Case 1

图2是航天器角速度响应曲线，可以看出，航天器角速度在2种控制器的作用下均在25 s左右

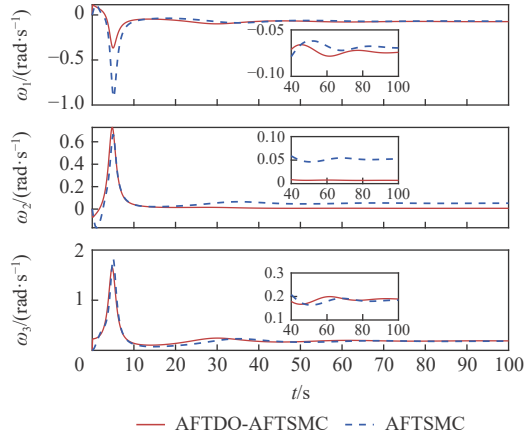
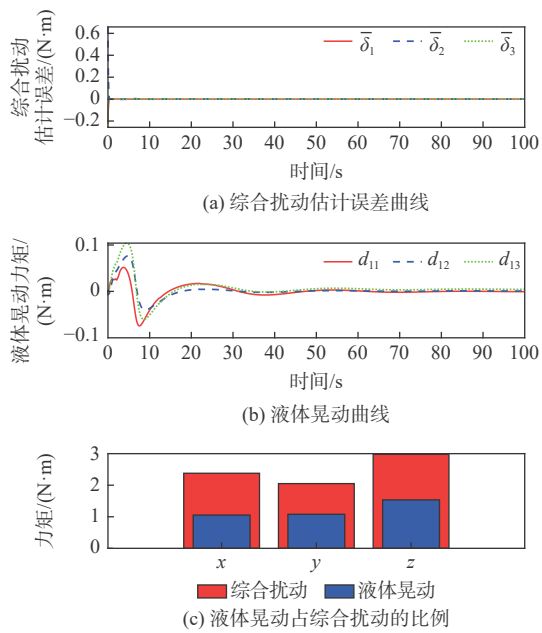


图6 Case 1的脉动球角速度变化曲线

Fig. 6 Curves of the pulsating ball angular velocities in Case 1



(a) 综合扰动估计误差曲线

(b) 液体晃动曲线

(c) 液体晃动占综合扰动的比例

图7 Case 1的综合扰动估计误差曲线、液体晃动力矩曲线和液体晃动占综合扰动的比例

Fig. 7 Curves of comprehensive disturbance estimation error, liquid sloshing and liquid sloshing as a proportion of the comprehensive disturbance in Case 1

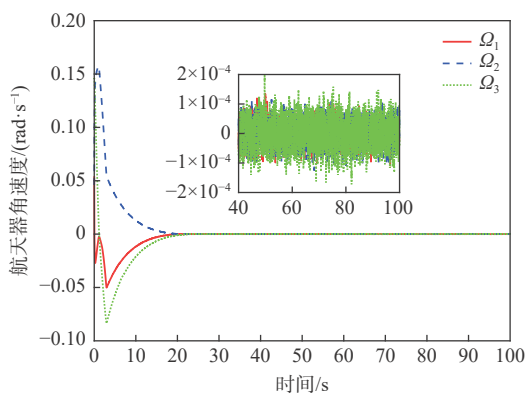


图8 Case 2的航天器角速度变化曲线

Fig. 8 Curves of spacecraft angular velocities in Case 2

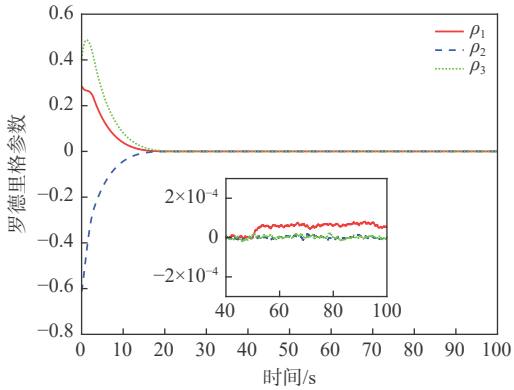


图 9　Case 2 的航天器 MRPS 变化曲线

Fig. 9　Curves of spacecraft MRPS in Case 2

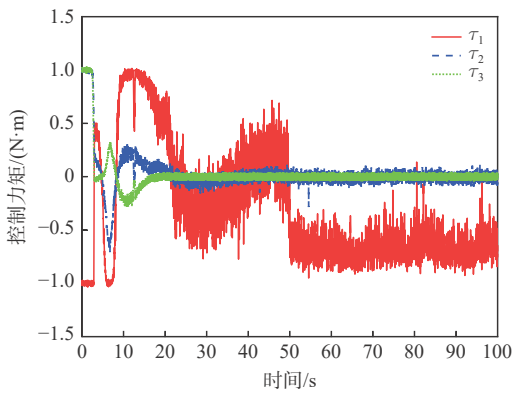


图 10　Case 2 的航天器控制力矩曲线

Fig. 10　Curves of spacecraft control torques in Case 2

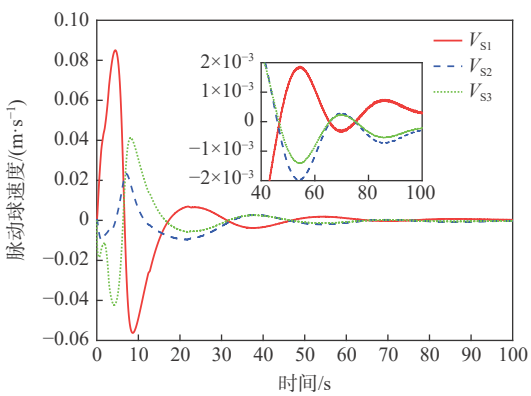


图 11　Case 2 的脉动球速度变化曲线

Fig. 11　Curves of the pulsating ball velocities in Case 2

收敛, 但文献 [22] 提出的控制器使角速度的超调量增大且收敛精度约为2×10^{-5} rad/s, 而本文提出的控制器收敛精度为1×10^{-7} rad/s。

图 3 是航天器 MRPS 响应曲线, 2 种控制策略下, MRPS 均在 25 s 左右收敛, 文献 [22] 控制器的收敛精度约为2×10^{-5}, 本文提出的控制器收敛精度约为8×10^{-8}。图 4 是航天器控制力矩响应曲线, 在执行器故障时输入饱和设置为 1 N·m的情况可以正常工作。

图 5 和图 6 分别是大幅晃动的液体等效为脉动

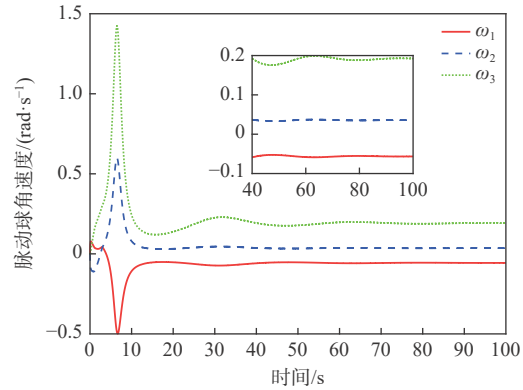


图 12　Case 2 的脉动球角速度变化曲线

Fig. 12　Curves of the pulsating ball angular velocities in Case 2

球的速度和角速度响应曲线, 可以看出, 本文提出的控制器的抑制液体晃动效果明显小于文献 [22] 的控制器。图 7(a) 是本文的自适应固定时间干扰观测器的综合扰动估计误差响应曲线, 可以看出, 扰动可以在短时间内被估计到, 拥有良好的观测性能。图 7(b) 是液体晃动力矩曲线。图 7(c) 是液体晃动占综合扰动的比例, 可以看出, 在综合扰动中液体占比相当大。

3.2　Case2 仿真

航天器角速度、MRPS、控制力矩、脉动球速度和角速度的响应曲线如图 8～图 12 所示。在随机外部扰动及敏感器和控制机构输出存在误差的影响下, 航天器状态变量均能满足控制要求。

4　结　论

1) 设计的自适应固定时间观测器能够在固定时间内估计到由外部扰动、执行器饱和、执行器故障和液体大幅晃动构成的综合扰动, 增强了系统的鲁棒性。

2) 针对存在外部扰动、执行器饱和、执行器故障和液体大幅晃动的充液航天器, 基于干扰观测器提出一种固定时间滑模容错控制方案策略, 实现航天器的姿态稳定控制。

3) 与自适应固定时间滑模控制策略相比, 本文控制策略的控制精度更高, 且有更好的暂态响应。

参考文献 (References)

[1] DONG K, QI N M, WANG X L, et al. Dynamic influence of propellant sloshing estimation using hybrid: mechanical analogy and CFD[J]. Transactions of the Japan Society for Aeronautical and Space Sciences, 2009, 52(177): 144-151.

[2] 宋晓娟, 王宏伟, 岳宝增. 测量不确定的充液航天器自适应鲁棒容错控制[J]. 宇航学报, 2022, 43(5): 638-648.
SONG X J, WANG H W, YUE B Z. Adaptive robust fault tolerant control of liquid-filled spacecraft with measurement uncertainty[J].

Journal of Astronautics, 2022, 43(5): 638-648(in Chinese).

[3] SONG X J, LI X S, LU S F, et al. Adaptive quaternion feedback control and oscillation suppression for liquid-filled flexible spacecraft[J]. Journal of Vibration Engineering & Technologies, 2023, 11(3): 1011-1028.

[4] VREEBURG J P B. Dynamics and control of a spacecraft with a moving pulsating ball in a spherical cavity[J]. Acta Astronautica, 1997, 40(2-8): 257-274.

[5] DENG M L, YUE B Z. Nonlinear model and attitude dynamics of flexible spacecraft with large amplitude slosh[J]. Acta Astronautica, 2017, 133: 111-120.

[6] BERRY R L, TEGART J R. Experimental study of transient liquid motion in orbiting spacecraft: NASA-CR-144003[R]. Washington, D. C.: NASA, 1975: 1-160.

[7] LIU F, YUE B Z, TANG Y, et al. 3DOF-rigid-pendulum analogy for nonlinear liquid slosh in spherical propellant tanks[J]. Journal of Sound and Vibration, 2019, 460: 114907.

[8] SONG X J, LU S F. Attitude maneuver control of liquid-filled spacecraft with unknown inertia and disturbances[J]. Journal of Vibration and Control, 2019, 25(8): 1460-1469.

[9] XING L, ZHANG J Q, LIU C, et al. Fuzzy-logic-based adaptive event-triggered sliding mode control for spacecraft attitude tracking[J]. Aerospace Science and Technology, 2021, 108: 106394.

[10] BHAT S P, BERNSTEIN D S. Finite-time stability of continuous autonomous systems[J]. SIAM Journal on Control and Optimization, 2000, 38(3): 751-766.

[11] LI S H, DU H B Z, LIN X. Finite-time consensus algorithm for multi-agent systems with double-integrator dynamics[J]. Automatica, 2011, 47(8): 1706-1712.

[12] JING C H, XU H G, NIU X J, et al. Adaptive nonsingular terminal sliding mode control for attitude tracking of spacecraft with actuator faults[J]. IEEE Access, 2019, 7: 31485-31493.

[13] 宋晓娟, 王宏伟, 吕书锋. 输入饱和的充液航天器抗干扰有限时间滑模控制[J]. 控制与决策, 2021, 36(5): 1078-1086.
SONG X J, WANG H W, LYU S F. Anti-disturbance finite-time sliding mode control for liquid-filled spacecraft with input saturation[J]. Control and Decision, 2021, 36(5): 1078-1086(in Chinese).

[14] GOLESTANI M, ZHANG W D, YANG Y X, et al. Disturbance observer-based constrained attitude control for flexible spacecraft [J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(2): 963-972.

[15] ALIPOUR M, MALEKZADEH M, ARIAEI A. Active fractional-order sliding mode control of flexible spacecraft under actuators saturation[J]. Journal of Sound and Vibration, 2022, 535: 117110.

[16] ZOU A M, KUMAR K D, DE RUITER A H J. Fixed-time attitude tracking control for rigid spacecraft[J]. Automatica, 2020, 113: 108792.

[17] WANG L, SONG Z Y. Continuous fixed-time sliding mode attitude controller design for rigid-body spacecraft[J]. IEEE Access, 2020, 8: 105399-105410.

[18] ZHUANG M L, TAN L G, LI K H, et al. Fixed-time formation control for spacecraft with prescribed performance guarantee under input saturation[J]. Aerospace Science and Technology, 2021, 119: 107176.

[19] CAO L, XIAO B, GOLESTANI M, et al. Faster fixed-time control of flexible spacecraft attitude stabilization[J]. IEEE Transactions on Industrial Informatics, 2020, 16(2): 1281-1290.

[20] YIN Z Y, LUO J J, WEI C S. Quasi fixed-time fault-tolerant control for nonlinear mechanical systems with enhanced performance [J]. Applied Mathematics and Computation, 2019, 352: 157-173.

[21] 韩治国, 张科, 吕梅柏, 等. 航天器自适应快速非奇异终端滑模容错控制[J]. 航空学报, 2016, 37(10): 3092-3100.
HAN Z G, ZHANG K, LYU M B, et al. Spacecraft fault-tolerant control using adaptive non-singular fast terminal sliding mode[J]. Acta Aeronautica et Astronautica Sinica, 2016, 37(10): 3092-3100 (in Chinese).

[22] GOLESTANI M, ESMAEILZADEH S M, MOBAYEN S. Fixed-time control for high-precision attitude stabilization of flexible spacecraft[J]. European Journal of Control, 2021, 57: 222-231.

[23] 王宏伟, 宋晓娟, 吕书锋. 充液航天器的鲁棒固定时间终端滑模容错控制[J]. 控制理论与应用, 2021, 38(2): 235-244.
WANG H W, SONG X J, LYU S F. Robust fixed-time terminal sliding mode fault tolerant control for liquid-filled spacecraft[J]. Control Theory & Applications, 2021, 38(2): 235-244(in Chinese).

[24] WU X, LUO S B, WEI C S, et al. Observer-based fault-tolerant attitude tracking control for rigid spacecraft with actuator saturation and faults[J]. Acta Astronautica, 2021, 178: 824-834.

[25] WANG X H, TAN C P. Fault-tolerant spacecraft attitude control under actuator saturation and without angular velocity[J]. International Journal of Robust and Nonlinear Control, 2019, 29(18): 6483-6506.

[26] GAO S H, JING Y Q, LIU X P, et al. Finite-time attitude-tracking control for rigid spacecraft with actuator failures and saturation constraints[J]. International Journal of Robust and Nonlinear Control, 2020, 30(5): 1903-1937.

[27] 王靓玥, 郭延宁, 马广富, 等. 航天器姿态控制输入饱和问题综述[J]. 宇航学报, 2021, 42(1): 11-21.
WANG L Y, GUO Y N, MA G F, et al. Review on input saturation in the spacecraft attitude control[J]. Journal of Astronautics, 2021, 42(1): 11-21(in Chinese).

[28] POLYAKOV A. Nonlinear feedback design for fixed-time stabilization of linear control systems[J]. IEEE Transactions on Automatic Control, 2012, 57(8): 2106-2110.

[29] CHEN Z Y, ZHANG H C, XIAO B. Appointed-time nonsingular sliding mode control of spacecraft attitude stabilization[J]. IET Control Theory & Applications, 2023, 17(14): 1944-1954.

[30] ZHU Z, XIA Y Q, FU M Y. Attitude stabilization of rigid spacecraft with finite-time convergence[J]. International Journal of Robust and Nonlinear Control, 2011, 21(6): 686-702.

[31] HARDY G H, LITTLEWOOD J E, POLYA G. Inequalities[M]. Cambridge: Cambridge University Press, 1952: 30-32.

[32] SAFA A, BARADARANNIA M, KHARRATI H, et al. Robust attitude tracking control for a rigid spacecraft under input delays and actuator errors[J]. International Journal of Control, 2019, 92(5): 1183-1195.

[33] SIDI M J. Spacecraft dynamics and control: a practical engineering approach[M]. Cambridge: Cambridge University Press, 1997: 328-401.

[34] 韩聪颖. 小型天文卫星高精度姿态确定和控制技术研究[D]. 西安: 西北工业大学, 2004: 18-23.

HAN C Y. Researches on the attitude determination and control system of a small astronomical satellite[D]. Xi'an: Northwestern Polytechnical University, 2004: 18-23(in Chinese).

[35] 徐芳军. 卫星消旋的姿态稳定控制研究[D]. 哈尔滨: 哈尔滨工业大学, 2017: 22-25.
XU F J. Research on attitude stability control of satellite spin cancellation[D]. Harbin: Harbin Institute of Technology, 2017: 22-25 (in Chinese).

Fixed-time attitude sliding mode fault-tolerant control for liquid-filled spacecraft

FAN Zhiwen[1, 2], SONG Xiaojuan[1, 3, *], LYU Shufeng[2], YUE Baozeng[4]

(1. College of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot 010051, China;
2. College of Science, Inner Mongolia University of Technology, Hohhot 010051, China;
3. Inner Mongolia Key Laboratory of Robotics and Intelligent Equipment Technology, Hohhot 010051, China;
4. School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China)

Abstract: The attitude system of liquid sloshing in liquid-filled spacecraft with actuator saturation, actuator failure, and external disturbance is studied. A sliding mode fault-tolerant control strategy based on a disturbance observer is proposed to achieve attitude fixed-time stability. In the process of dynamic modeling, the liquid fuel with large sloshing is equivalent to the motion pulsating ball model, and the dynamic equation of rigid-liquid coupling spacecraft is established. To guarantee that the complete disturbance in the control system can be calculated in a predetermined amount of time, an adaptive fixed-time disturbance observer is initially created during the attitude controller design phase. On this basis, a fixed-time attitude sliding mode fault-tolerant control strategy is designed by using sliding mode control theory. In addition, by introducing the hyperbolic tangent function, the sign function of traditional input saturation is overcome. Despite external disturbances, actuator failure, and actuator saturation, the control technique may rapidly establish high-precision attitude stability of the spacecraft. Finally, the attitude sliding mode fault-tolerant control scheme proposed in this paper is numerically simulated and compared to verify the effectiveness and superiority of the control strategy.

Keywords: liquid-filled spacecraft; attitude control; disturbance observer; sliding mode control; fixed-time control

Received: 2023-11-08; **Accepted**: 2024-02-21; **Published Online**: 2024-03-13 08:57

URL: link.cnki.net/urlid/11.2625.V.20240312.1520.002

Foundation items: National Natural Science Foundation of China (12362004,11962020,12172182,12132002); Talent Revitalization for Inner Mongolia Project Team (2025TEL02); Inner Mongolia Autonomous Region University Youth Science and Technology Talents Project (NJYT23029,NJYT23067); National Natural Science Foundation of Inner Mongolia (2023JQ14)

* **Corresponding author**. E-mail: xjsong0603@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0754

基于快速傅里叶卷积的电成像测井图像修复

苏乾潇，乔德新*，任义丽，冯周，林盛斓，黄睿琦
(中国石油勘探开发研究院，北京 100083)

摘 要：成像测井是复杂储层测井评价中的重要技术手段。通过成像测井，可以获得井周的电阻率分布二维图像，用于评价井壁缝洞发育和地层沉积构造等。但由于电阻率成像测井仪器的特点，电阻率测井图像上会出现空白条带，这增加了计算机对电成像资料处理的难度。目前的图像修复方法和现有的神经网络图像修复方法在待填充部分占比较大时效果都不好。因此，迫切需要基于深度学习的智能修复方法。以一种基于快速傅里叶卷积的成像测井图像空白条带填充网络为基础，将西南油气田的电成像测井图像构建为数据集，训练得到一种基于快速傅里叶卷积的成像测井图像空白条带智能填充深度学习算法。对比各种算法的时间，结果表明：所提算法在条带宽度大的成像测井图像中修复效果较优，同时修复效率提升明显。通过所提方法实现了成像测井空白条带的快速、准确和智能化修复，实现了全井眼图像的快速生成，并解决了全井眼图像获取的困难。

关 键 词：电成像测井；空白条带填充；神经网络；快速傅里叶卷积神经网络；图像修复
中图分类号：P631.84
文献标志码：A **文章编号**：1001-5965(2026)01-0362-09

微电阻率成像测井是石油勘探过程中复杂储层测井评价的重要技术手段。通过微电阻率探头上的纽扣电极阵列可以测得空间分布的微电阻率数据，经过位置矫正处理再着色后能够得到井周二维图像，能清晰、直观地反映井壁的结构特征，在井壁缝洞发育情况评价、地层沉积构造分析等方面发挥着重要作用。目前，已广泛应用于碳酸盐岩、砂砾岩和火成岩的测井评价[1-2]。

微电阻率成像测井的测量仪器包含承载了纽扣电极的极板，其在井上时折叠，在井下张开贴近井壁进行测量，而极板和极板之间在张开时就存在间隙。这会导致成像图不完整，存在竖条状空白条带。空白条带会造成部分信息缺失，影响地质目标特征的识别，不便于对电成像测井资料进行解释。为使图像能更完整地反映地下地质信息，有必要对图像中信息缺失的空白条带进行填充。

不同于普通自然图像修复，微电阻率电成像测井图像有很多特殊性。首先，微电阻率电成像测井图像空白占比较大，在21.59 cm粗的井眼中，图像覆盖率仅为70%[3]。其次，全井眼图像难以获取，上述一些基于神经网络对井眼图像进行填充的研究都只能使用自然场景图像对神经网络进行训练与测试。并且由于缺少完整的全井眼图像，在测试时只能通过肉眼对填充图像的连续性、地质特征进行评判。

因此，针对这些难点，本文提出一种利用深度学习的方法来智能修复电成像测井图像中的空白条带问题。该方法的核心是基于快速傅里叶卷积的电成像测井图像空白条带填充网络。为了训练该网络，本文创建一个数据集，其中，包括西南油气田的电成像测井图像。经过训练，该深度学习算法能够快速、准确、智能地修复电成像测井图像中的空白条带，从而实现整个井眼图像的快速生成，同时，也克服了全井眼图像获取的难题。

收稿日期：2023-11-20；**录用日期**：2024-01-05；**网络出版时间**：2024-02-04 10:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20240202.1216.002
基金项目：国家自然科学基金(42372175)；国家科技重大专项(2025ZD1401501)
* **通信作者**. E-mail：qdx@petrochina.com.cn

引用格式：苏乾潇，乔德新，任义丽，等. 基于快速傅里叶卷积的电成像测井图像修复[J]. 北京航空航天大学学报，2026，52(1)：362-370. SU Q X，QIAO D X，REN Y L，et al. Inpainting of electrical imaging logging images based on fast Fourier convolution[J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52(1)：362-370 (in Chinese).

1　相关工作

目前, 对电成像测井图像空白条带进行填充的主要手段都是使用图像修复方法, 将空白条带视作待修复区域, 使用算法进行填充。这些图像修复方法包含传统的图像修复方法和近年出现的神经网络算法。

传统的图像修复方法可以在广义上分为基于补丁和基于扩散 2 类[4]。基于补丁的图像修复方法将已有图像剪裁生成补丁, 通过算法挑选最适合的补丁填充待修复区域, 但该方法可能导致像素不连续。而基于扩散的修复方法从缺失区域边界逐步推导出未知区域信息, 适用于较小的缺失区域。随着缺失区域变大, 迭代次数变多, 两者修复效果变差。由于电成像图像之间空白区域较大, 一般基于补丁和扩散的修复方法无法顾及缺失部分的语义信息[5], 不能得到很好的修复效果, 因而不能满足实际工程需要, 需要采用可以利用语义信息进行修复的图像修复算法。

近年来, 神经网络被应用于图像修复任务。起初, 传统的卷积神经网络效果不佳, 但生成对抗网络的出现使基于语义结构的图像修复成为可能[6]。然而, 大部分深度学习修复模型在处理大尺寸待修复部分时面临困难[7-10], 直到 LaMa[11] 等研究取得突破。因此, 对于类似电成像测井图像的人工纹理图像, 需要寻找更适合的修复方法。FcF 神经网络[5]是基于 Encoder-Decoder 结构的神经网络, 由于其很好的抗噪能力得到了广泛应用。MAT 网络[12] 是首个基于 Transformer 网络用于图像修复的网络, 其在 Transformer 网络的基础上优化了多头上下文注意力机制, 可以修复较大范围掩码的图像。LaMa 是本文所基于的神经网络, 在 ResNet 上加入了傅里叶变换的模块, 从而使用更少的卷积层获得更大的感受野。

但这些图像修复算法只有少部分被应用于电成像测井条带填充。目前, CIFLog 软件[13] 和 Techlog 软件[14] 都包含电成像测井条带填充模块, 这 2 款软件的条带填充功能所采用的算法都是基于地质多点统计的 Filtersim 算法。Filtersim 算法利用不同滤波窗口对图像块进行分类, 在填充时根据分类结果选取最相近的图像块对空缺区域进行填补[15]。文献 [16] 使用 Filtersim 算法和传统的反距离加权插值法同时对电成像图像空白条带进行填充, 经过比较认为, Filtersim 算法较反距离加权插值法效果更好。Filtersim 算法在填充均匀性比较好的地层图像时, 图像填充效果较好, 在填充大区域、非均匀性比较严重的地层时, 存在匹配不确定性, 可能导致图像填充结果出现异常[17]。另外, 还有一些研究使用传统算法对井眼图像进行填充。文献 [18] 针对不同地质特征对 Criminisi 算法进行了改进, 通过引入圆弧函数解决了纹理延伸的问题, 使 Criminisi 针对不同地质特征进行了改进。文献 [19] 将图像分解为结构子图和纹理子图, 使用反距离加权插值算法填充结构子图, 使用 Criminisi 算法填充纹理子图, 最后将两者相加, 发现结合算法的填充效果优于单一使用一种算法的效果。文献 [20] 在电阻率图像的某个深度窗内, 对电阻率数据阵进行奇异谱分析, 通过求得低秩电阻率数据矩阵对电阻率数据进行插值, 在空白条带占比较少的图像上进行了测试。

在使用神经网络算法填充电成像测井图像方面, 文献 [21] 使用 Encoder-Decoder 结构的卷积神经网络对砂泥岩剖面和砂砾岩体电成像图像进行空白条带填充, 并与 Filtersim 算法的结果进行对比, 认为 Encoder-Decoder 网络填充结果过渡平滑自然, 能更有效地恢复轮廓边缘信息。文献 [22] 对 ResNet101 网络进行了改进, 将一次卷积过后的低层语义与 ResNet101 网络输出的高层语义融合拼接, 经解码后输出填充图像, 并使用 Places365-standard 自然场景图像数据集对网络进行了训练, 经比较发现其比 Encoder-Decoder、U-Net 网络在自然场景图像下与砂砾岩电成像图像的填充效果都更好。在使用神经网络填充电成像测井图像的研究中, 神经网络填充效果普遍较传统方法更加自然。神经网络法的填充图中对于不明连接部分的处理往往是柔和过渡的, 即使是难以平滑填充的部分也往往是模糊过渡的, 较符合人眼的视觉习惯。这可能是因为神经网络法的填充结果是同步进行计算的。而传统算法由于填充结果是从边缘进行计算的, 在待填充中心的交会位置往往会产生色块和条纹等不符合人眼视觉习惯的缺陷。

综合来说, 传统的基于补丁和基于扩散的方法都不能很好地应对电成像测井图中较大的待填充部分。特别是随着计算力不断进步, 其与神经网络算法的效果差距也越来越大。而神经网络算法同样为了应对较大填充部分的问题, 需要不断增加网络深度以增加感受野, 使得不仅训练参数、模型大小线性地增长, 训练速度也会成倍地减缓。因此, 在针对测井图像的图像填充领域应用新的神经网络算法是非常有必要的。

2　本文方法

图像修复技术的主要目标是填补图像中的缺失信息。对于微电阻率电成像测井图像而言, 与普

通自然图像修复相比有一些特殊性：首先，微电阻率电成像测井图像通常存在较大的空白区域。例如，在直径为21.59 cm的井眼中，图像覆盖率仅为70%[23]，而到31.115 cm的井眼中时，井壁覆盖率只有50%；其次，由于获取全井眼图像的困难，现有研究中基于神经网络的填充算法通常只能使用自然场景图像进行训练和测试。此外，由于缺乏完整的全井眼图像，在测试时只能通过主观的肉眼观察来评估填充后测井图像的连续性和地质特征。

本文提出一种基于快速傅里叶卷积的图像修复神经网络，将其命名为Log-LaMa。快速傅里叶卷积神经网络的网络结构由3层下采样、18个串联的傅里叶卷积模块和3层上采样网络组成。网络结构如图1所示。

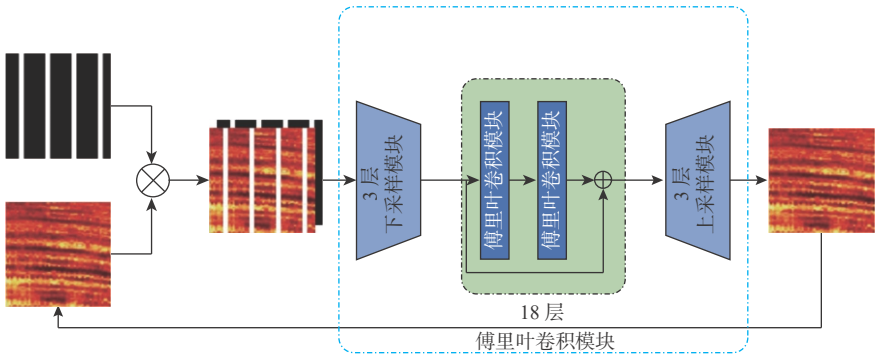


图1 基于快速傅里叶卷积的电成像测井图像修复神经网络

Fig. 1 Inpainting neural network of electrical imaging logging images based on fast Fourier convolution

2.1 下采样及上采样模块

下采样操作可以在缩小数据尺寸时提取特征。Log-LaMa网络的下模块采用3层卷积层来对图片特征做粗提取，同样使用批归一化。每层所采用卷积核个数、卷积核尺寸及通道数如表1所示。

表1 下采样卷积层所采用卷积核个数、尺寸及通道数

Table 1 The number, size, and channels of the convolutional kernels used in the downsampling convolutional layer

层数	卷积核个数	卷积核尺寸	卷积核通道数
1	64	7×7	4
2	128	3×3	64
3	256	3×3	128

上采样可以将输出数据的大小还原至与输入图像一致，一般可以采用2种方法：插值和反卷积。插值一般采用双线性插值，但生成图像效果较差。反卷积过程与卷积类似，通过填充0再进行卷积操作可以使输出尺寸放大。由于也与卷积类似含有卷积核，其也可以被学习，效果较双线性插值好很多。

Log-LaMa网络的上采样模块采用反卷积方式，每层卷积核大小与下采样所采用卷积核大小完全镜像对应，每层卷积核个数则与下采样所采用卷积核通道数完全镜像对应，最后一层采用3个通道的输出。

2.2 傅里叶卷积模块

傅里叶卷积模块会将下采样的结果分为全局和局部2部分通道，如图2所示。前半部分通道信息交由局部部分进行处理，局部部分对于信息的处理与传统卷积神经网络一样，使用3×3的卷积核进行卷积处理。后半部分通道信息交由全局部分进行处理，全局部分不仅会用3×3的卷积核对信息进行卷积处理，还会同时使用频谱转换模块对信息进行处理。局部部分与全局部分处理过后的信息会被线性叠加，输入下一层傅里叶卷积模块中。

频谱转换操作首先会对信息做一次卷积，并经过批归一化和激活函数(BN-ReLU)处理。之后使用二维傅里叶变换将整层信息转化为频域。再进行一次卷积处理，并经过批归一化和激活函数。最后，使用傅里叶逆变换将频域信息变换回，如图3所示。

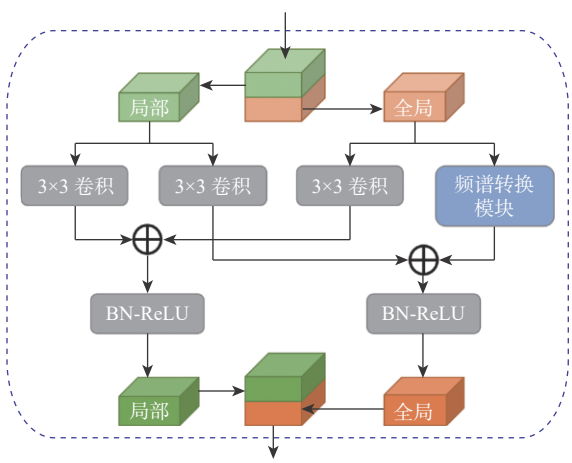


图2 傅里叶卷积模块结构

Fig. 2 Structural diagram of Fourier convolution module

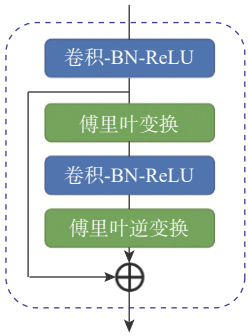


图 3　频谱转换模块结构示意图

Fig. 3　Structural diagram of spectral transform module

图像信息经由上采样模块处理, 将傅里叶卷积模块的输出信息上采样至原输入图像分辨率进行输出。

2.3　损失函数设计

训练时, 损失函数由以下部分组成:

$$L_{\text{final}} = \kappa L_{\text{Adv}} + \alpha L_{\text{HRFPL}} + \beta L_{\text{FM}} + \gamma R_1 \tag{1}$$

式中:

$$L_{\text{Adv}} = \text{sg}_{\theta}(L_D) + \text{sg}_{\xi}(L_G) \to \min_{\theta,\xi} \tag{2}$$

L_{Adv}为对抗损失, 希望鉴别器损失L_D和生成器损失L_G在各自的参数θ和ξ下达到最小, $\text{sg}_{\theta}(\cdot)$表示停止对$\theta$的梯度传播。

$$L_D = -E_{\boldsymbol{X}}\left(\ln D_{\xi}(\boldsymbol{X})\right) - E_{\boldsymbol{X},\boldsymbol{M}}\left(\ln D_{\xi}\left(\hat{\boldsymbol{X}}\right)\odot \boldsymbol{M}\right) - E_{\boldsymbol{X},\boldsymbol{M}}\left[\ln\left(\boldsymbol{A}_1 - D_{\xi}\left(\hat{\boldsymbol{X}}\right)\right)\odot(\boldsymbol{A}_1 - \boldsymbol{M})\right] \tag{3}$$

$$L_G = -E_{\boldsymbol{X},\boldsymbol{M}}\left[\ln D_{\xi}\left(\hat{\boldsymbol{X}}\right)\right] \tag{4}$$

式中: 鉴别器$D_{\xi}(\cdot)$源自文献 [24], 在局部尺度上对图像进行特征提取; $\boldsymbol{A}_1$表示全 1 矩阵; $\boldsymbol{X}$为输入图片; $\hat{\boldsymbol{X}}$为生成的图片; $\boldsymbol{M}$为掩码。

L_{HRFPL}为高感受野损失函数:

$$L_{\text{HRFPL}}\left(\boldsymbol{X},\hat{\boldsymbol{X}}\right) = \text{Mean}\left[\left(\phi_{\text{HRF}}(\boldsymbol{X}) - \phi_{\text{HRF}}\left(\hat{\boldsymbol{X}}\right)\right)^2\right] \tag{5}$$

式中: $\text{Mean}(\cdot)$表示取平均值, 先层内平均, 再层间平均; $(\cdot)^2$表示每个元素逐个平方, 输出结果与输入的维度相同; $\phi_{\text{HRF}}(\boldsymbol{X})$表示使用 16 层空洞卷积神经网络, 同样使用批归一化。

L_{FM}为特征匹配损失[25], 通过将鉴别器各层提取到的特征加权平均, 可以起到稳定训练的作用:

$$L_{\text{FM}} = E_{\boldsymbol{S},\boldsymbol{X}}\left[\sum_{i=1}^{T}\frac{1}{N_i}\left[\left\|D^{(i)}(\boldsymbol{S},\boldsymbol{X}) - D^{(i)}(\boldsymbol{S},G(\boldsymbol{S}))\right\|_1\right]\right] \tag{6}$$

式中: $D^{(i)}(\cdot)$表示鉴别器第i层的输出; T为鉴别器$D(\cdot)$的总层数; N_i为每层中的元素数量; $\boldsymbol{S}$为随机噪声, 用来为生成器$G(\cdot)$提供输入。

R_1为梯度惩罚, 减小过拟合:

$$R_1 = E_{\boldsymbol{X}}\left\|\nabla D_{\xi}(\boldsymbol{X})\right\|^2$$

以上损失中, L_{Adv}和L_{FM}主要负责感知图像局部细节, L_{HRFPL}主要负责全局结构一致。

3　实验结果及分析

3.1　实验数据

选取西南油气田 31 口井的微电阻率电成像测井资料, 共 19 695 m 井段。因为有些井段出现极板之间的纵向错位, 影响网络对地质特征的正常学习, 筛选其中 9 503 m 没有错位、质量较优的微电阻率电成像测井资料, 平均每口井 306 m 微电阻率电成像测井图像。

3.2　训练样本集构建方法

电成像测井的电阻率数据在经过位置矫正后, 生成的数据是单通道的, 并不包含颜色信息。并且由于测井软件在映射过程中会将像素进行缩放, 对色块边缘有模糊处理, 因此, 生成的测井图像与原测井电阻率资料并不一一对应。如果使用色标映射后的三通道彩色图像进行处理, 可能会对图像引入新的误差, 因此, 直接采用单通道的电阻率数据进行填充处理。

将电成像测井的电阻率数据映射到 0～255 的范围内, 按原位置填充入 png 格式的图片中。同时, 新建一张与待填充数据同样尺寸的纯黑图片, 将电阻率数据中无效值数据的对应位置的亮度置为 255, 即可生成对应掩码。

将 0～255 的灰度做逆映射, 可得到原电阻率数据及算法生成的用于填充的电阻率数据。

无论训练还是测试都需要完整真实的测井成像图作为训练集和测试集的模板, 用方形窗口截取数据中单个极板的部分, 如图 4 所示。在填充时, 将数据点长宽比设定为 1∶1, 对于宽度为 250 个点的测井电阻率数据, 同样选取长度为 250 个点的测井电阻率数据映射入图像中。

在 9 503 m 较优的测井资料中, 截取后共筛选出 9 828 个含有有效数据的窗口, 将其按灰度映射到 png 格式的图片, 制成本次实验的数据集。

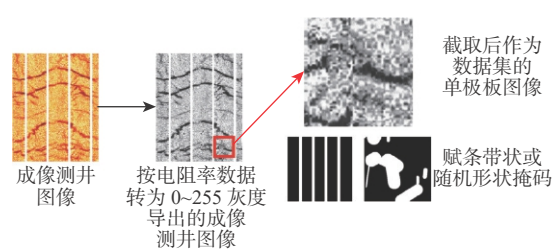


图 4　截取单极板图像

Fig. 4　Capture single pad images

对于数据集的掩码部分，采用竖条状和随机形状2类掩码，竖条状掩码模仿电成像测井图像中空白条带的真实场景，采用占比为10%、20%、30%和40%的4种掩码，如表2所示。而随机形状的掩码则由采用三星人工智能附带的掩码生成器生成，分为细、中、粗3档，如表3所示。将数据集随机分为80%的训练集、10%的验证集及10%的测试集后进行训练，设置训练代数为200。将迁移学习后用于填充井眼图象的网络版本命名为“Log-LaMa”。

表2 竖条状掩码示例

Table 2 Examples of vertical striped masks

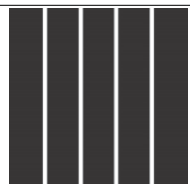
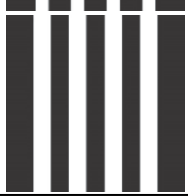
掩码占比/%	掩码
10	
20	
30	
40	

表3 随机形状掩码示例

Table 3 Examples of random shaped masks

掩码类型	掩码
细	
中	
粗	

3.3 实验方法及参数

使用经过Places2和Places Challenge数据集训练的神经网络模型big-LaMa，并使用从西南油气田31口测井的微电阻率电成像测井图像数据中制成的数据集对big-LaMa进行迁移学习，使其能够更好地学习如何修复电成像测井图像，最终得到专为修复电成像测井图像的网络Log-LaMa。

使用电成像测井图像数据集中的测试集部分对Log-LaMa进行测试。同时，对Criminisi算法、Filtersim算法、FcF网络及MAT网络进行测试以对比，其中，神经网络FcF和MAT均经过Places2数据集训练得到。

评价修复结果质量时，采用FID对图像修复质量进行评价。FID距离[26]是从原始图像的计算机视觉特征的统计方面，衡量2组图像的相似度，是计算真实图像与生成图像特征向量之间距离的一种度量。在评价时使用InceptionV3网络分别对真实图像和生成图像进行特征提取，将输出向量作为评价得分。FID在最佳情况下的得分为0，表示2组图像相同。分数越低表示2组图像越相似，或两者的统计量越相似。常用的InceptionV3网络有64、192、768和2048 4种维度的输出，其中，192维的输出最常用，本文在分析时采用192维输出的FID网络。

训练时设置batch size = 10，共100个epoch。而对于损失函数中权重κ、α、β和γ的选择，使用Log-LaMa训练时使用的κ=10、α=30、β=100和γ=0.001。生成器学习率设置为0.001，鉴别器学习率设置为0.000 1。两者的训练方法均使用Adam算法。

3.4 实验结果对比

3.4.1 时间对比

由于神经网络算法在计算时可以采用图形处理器(graphics processing unit, GPU)进行大规模并行计算，因此，虽然其计算量较大，但速度比传统算法更快。但传统算法使用CPU计算，而神经网络可以使用GPU辅助计算，因此，绝对地讨论两者之间的速度差异没有意义。本文使用HUAWEI MateBook 14笔记本电脑进行测试，仅作为示意。此电脑采用Intel i7-8565U处理器、NVIDIA GeForce MX250显卡、16 GB运行内存、2 GB显存。分别使用几种算法对2 000 m长的井段进行电成像测井图像填充，Criminisi算法由于速度过慢仅填充10 m，得到所用时间如表4所示。可以看出，Criminisi算法由于太缓慢而失去实用价值，Filtersim算法由于没有GPU加速显著慢于几种神经网络算法。Log-LaMa网络比MAT网络慢，但对比传统算法，仍然具有较快的速度，具有实用价值。

表 4　不同算法模型填充速度对比

Table 4　Comparison of filling speeds for different algorithm models

算法	填充每10 m电成像测井图像所耗时间/s
Log-LaMa	0.375
Criminisi	26 640
Filtersim	13
FcF	0.6
MAT	0.25

3.4.2　FID 评价

使用 Log-LaMa 网络和 Filtersim 算法同时对测试集数据进行填充, 测试时使用不同占比的条带状掩码, 将填充后的图像与数据集模板图像进行比对。采用 192 维向量输出的 InceptionV3 网络计算 FID, 结果如表 5 所示, 其中, FcF 神经网络和 MAT 神经网络均采用 Places2 数据集训练。Log-LaMa 网络采用 Places2 数据集预训练, 电成像测井数据集训练。可以看出, Log-LaMa 网络对电成像测井图像的填充能力较 Criminisi 算法、Filtersim 算法、FcF 神经网络和 MAT 神经网络都更好, 填充后的图像与原电成像测井图像更接近。虽然随着带填充部分的扩大, FID 有所上升, 但仍处于良好范围。

表 5　不同宽度条带下 Log-LaMa 网络与 Filtersim 算法填充 FID 对比

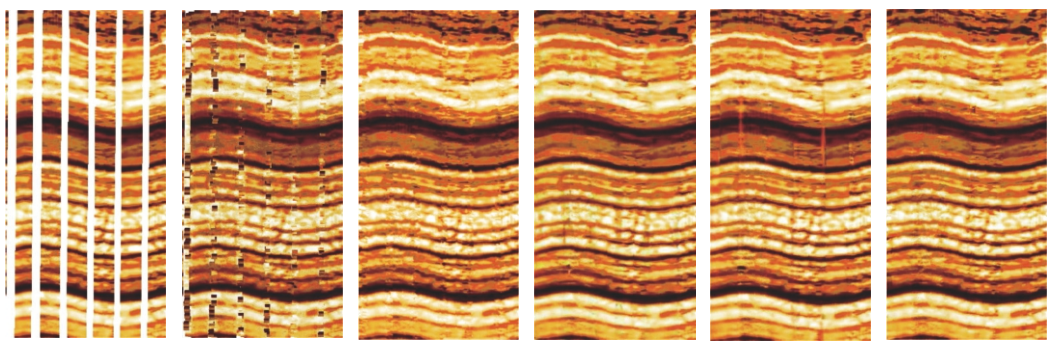
Table 5　Comparison of FID for the filling performance of Log-LaMa network and Filtersim algorithm under different stripe widths

条带占比/%	FID				
	Log-LaMa	Criminisi	Filtersim	FcF	MAT
10	0.150	2.391	0.203	0.156	0.151
20	0.422	25.680	0.650	0.593	0.617
30	0.890	30.010	1.124	0.933	1.015
40	1.511	34.200	1.941	1.477	1.520

3.4.3　效果对比

使用训练好的 Log-LaMa 网络对电成像测井数据进行填充, 对其使用色标赋色, 进行肉眼观察。

图 5(a) 为一段微电阻率成像测井(fullbore formation micro-imager, FMI)电成像测井静态成像, 图像井眼覆盖率为 80% 左右。对比图 5(b)～图 5(f), Criminisi 算法由于没有找到合适的样本块而产生了最多的不连续色块, 极其影响判断; Filtersim 算法填充后像素存在较少不连续现象, 但仍可分辨出一些明显的顿挫; FcF 网络填充效果自然, 但产生了少许竖状暗斑, 略微妨碍观察; MAT 网络填充时出现了竖状条纹, 影响对特征的判读解释; 而 Log-LaMa 网络填充效果非常自然、平滑, 几乎难以找出空白部分的界限。综合判断, Log-LaMa 网络对图像的填充效果最好。



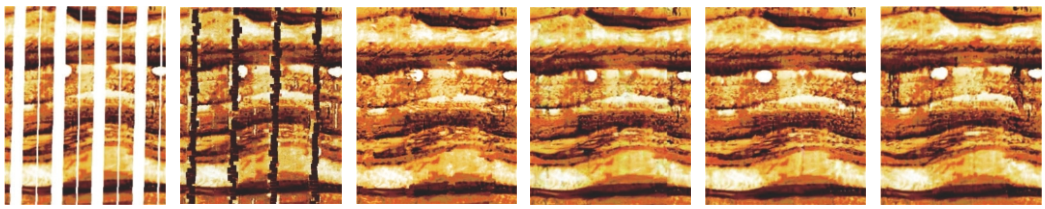
(a) 原始测井图像　(b) Criminisi 算法　(c) Filtersim 算法　(d) FcF 网络　(e) MAT 网络　(f) Log-LaMa 网络

图 5　Log-LaMa 网络与 Criminisi 算法、Filtersim 算法、FcF 网络及 MAT 网络填充测井图像对比(井眼覆盖率为 80%)

Fig. 5　Comparison of the filling performance on logging images by Log-LaMa network with Criminisi algorithm, Filtersim algorithm, FcF network and MAT network (borehole coverage rate is 80%)

图 6 为一段 FMI 电成像测井静态成像, 图像井眼覆盖率为 65% 左右。对比图 6(b)～图 6(f) 可以看出, Criminisi 算法对于极细的空白条带有一定的填充能力, 但在稍宽的条带上出现明显的不一致色块; Filtersim 算法可以较好地完成填充; FcF 网络和 MAT 网络可以较好地完成填充; Log-LaMa 网络则很好地完成了空白条带的过渡。在图中可以明显看出一个高阻颗粒, 传统 Filtersim 算法充填后, 颗粒边缘轮廓的连续性较差; FcF 网络填充后, 图中的圆形颗粒同样产生了形变现象; 而 Log-LaMa 网络有效恢复了轮廓边缘信息, 颗粒圆润自然。综合判断, Log-LaMa 网络对图像的填充效果最好。

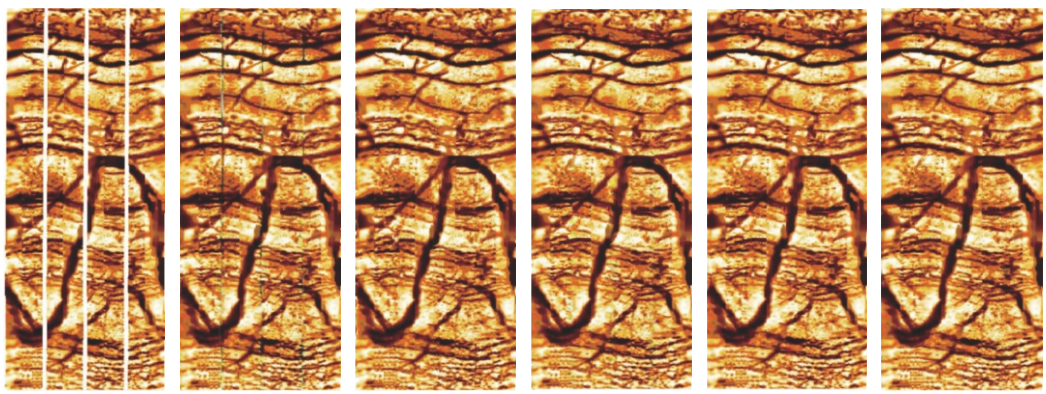
图 7 为一段 FMI 电成像测井静态成像, 图像井眼覆盖率为 90% 左右, 发育一条高角度裂缝。对比图 7(b)～图 7(f) 可以看出, Criminisi 算法对于极细的空白条带填充能力不够好, 出现明显的颜色不一致现象; Filtersim 算法可以较好地完成填充, 但仍存在过渡不柔和的现象; FcF 网络和 MAT 网络可以较好地完成填充, 但出现一些小的噪声斑; Log-LaMa 网络则更好地完成了空白条带的过渡。综合判断,



(a) 原始测井图像 (b) Criminisi 算法 (c) Filtersim 算法 (d) FcF 网络 (e) MAT 网络 (f) Log-LaMa 网络

图 6 Log-LaMa 网络与 Criminisi 算法、Filtersim 算法、FcF 网络及 MAT 网络填充测井图像对比(井眼覆盖率为 65%)

Fig. 6 Comparison of the filling performance on logging images by Log-LaMa network with Criminisi algorithm, Filtersim algorithm, FcF network and MAT network (borehole coverage rate is 65%)



(a) 原始测井图像 (b) Criminisi 算法 (c) Filtersim 算法 (d) FcF 网络 (e) MAT 网络 (f) Log-LaMa 网络

图 7 Log-LaMa 网络与 Criminisi 算法、Filtersim 算法、FcF 网络及 MAT 网络填充测井图像对比(井眼覆盖率为 90%)

Fig. 7 Comparison of the filling performance on logging images by Log-LaMa network with Criminisi algorithm, Filtersim algorithm, FcF network and MAT network (borehole coverage rate is 90%)

Log-LaMa 网络对图像的填充效果最好。

4 结 论

本文为解决填充微电阻率电成像测井图像的空白条带部分时待填充部分占比较大的问题,提出使用一种傅里叶卷积神经网络对空白条带进行填充。同时,为解决缺少训练所用完整图像的问题,截取单极板的电成像测井图像作为训练集的模板。

1)引入傅里叶变换及其逆变换到卷积神经网络中从而允许网络在整张图像的空间频域层面感知图像特征的周期性和变化,进而提高卷积神经网络对测井图像中纵向周期性地层和横向周期性岩块等地质特征的感知能力,使填充部分与两侧的过渡更加自然。

2)该采用微电阻率成像测井图像中的单极板图像作为数据集,对经过自然场景图像训练的神经网络进行迁移学习。该方法改善了模型在成像测井图像的填充任务中的泛化效果,使模型能够更好地发挥潜力。

本文用作训练数据集的单极板测井图像不能过小、分辨率不能过低,因此,从大量的测井数据中筛选出来的高质量数据数量有限,导致该训练方法对测井资料的质量要求较高,不便于推广。如果解决训练图像的要求,该方法的推广能力会得到较大提升。

参考文献(References)

[1] 侯振学, 于雪娴, 李东旭, 等. 电成像测井处理新技术在储层评价方面的应用[J]. 地球物理学进展, 2020, 35(2): 573-578.
HOU Z X, YU X X, LI D X, et al. Application of new processing technology of electrical imaging logging in reservoir evaluation[J]. Progress in Geophysics, 2020, 35(2): 573-578(in Chinese).

[2] 程超, 张亮, 李培彦, 等. 碎屑岩泥质分布形式的测井评价研究进展[J]. 地球物理学进展, 2021, 36(3): 1052-1061.
CHENG C, ZHANG L, LI P Y, et al. Advances in logging evaluation of shale distribution in clastic rocks[J]. Progress in Geophysics, 2021, 36(3): 1052-1061(in Chinese).

[3] ASSOUS S, WHETTON J A, ELKINGTON P A. Microresistivity image inpainting and visualization[C]//Proceedings of the SPE Annual Technical Conference and Exhibition. Pittsburgh: SPE, 2013: D021S039R008.

[4] 李雪涛, 王耀雄, 高放. 图像修复方法综述[J]. 激光与光电子学进展, 2023, 60(2): 29-44.
LI X T, WANG Y X, GAO F. Review of image inpainting methods[J]. Laser & Optoelectronics Progress, 2023, 60(2): 29-44 (in Chinese).

[5] JAIN J, ZHOU Y Q, YU N, et al. Keys to better image inpainting: structure and texture go hand in hand[C]//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2023: 208-217.

[6] YU J H, LIN Z, YANG J M, et al. Generative image inpainting with

contextual attention[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5505-5514.

[7] LAHIRI A, JAIN A K, AGRAWAL S, et al. Prior guided GAN based semantic inpainting[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 13693-13702.

[8] YI Z L, TANG Q, AZIZI S, et al. Contextual residual aggregation for ultra high-resolution image inpainting[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7505-7514.

[9] ZHAO L, MO Q H, LIN S H, et al. UCTGAN: diverse image inpainting based on unsupervised cross-space translation[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 5740-5749.

[10] ZHOU T, DING C X, LIN S W, et al. Learning oracle attention for high-fidelity face completion[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7677-7686.

[11] SUVOROV R, LOGACHEVA E, MASHIKHIN A, et al. Resolution-robust large mask inpainting with Fourier convolutions[C]//Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2022: 3172-3182.

[12] LI W B, LIN Z, ZHOU K, et al. MAT: mask-aware transformer for large hole image inpainting[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 10748-10758.

[13] 陈春. 中国石油新一代测井软件 CIFLog[J]. 石油勘探与开发, 2011, 38(3): 281.
CHEN C. China Petroleum's next-generation well logging software CIFLog[J]. Petroleum Exploration and Development, 2011, 38(3): 281(in Chinese).

[14] 斯伦贝谢. Techlog 2016 九大测井解释新技术[J]. 国外测井技术, 2018, 39(1): 71-73.
Schlumberger. Techlog 2016 nine new logging interpretation technologies[J]. World Well Logging Technology, 2018, 39(1): 71-73 (in Chinese).

[15] ZHANG T F, SWITZER P, JOURNEL A. Filter-based classification of training image patterns for spatial simulation[J]. Mathematical Geology, 2006, 38(1): 63-80.

[16] 孙建孟, 赵建鹏, 赖富强, 等. 电测井图像空白条带填充方法[J]. 测井技术, 2011, 35(6): 532-537.
SUN J M, ZHAO J P, LAI F Q, et al. Methods to fill in the gaps between pads of electrical logging images[J]. Well Logging Technology, 2011, 35(6): 532-537(in Chinese).

[17] 张翔, 张猛, 肖小玲, 等. 复杂地层情况下全井周电成像图像修复方法[J]. 石油物探, 2018, 57(1): 148-153.
ZHANG X, ZHANG M, XIAO X L, et al. Image inpainting for full-bore electrical imaging logging in complex formations[J]. Geophysical Prospecting for Petroleum, 2018, 57(1): 148-153(in Chinese).

[18] 罗歆, 闫建平, 王敏, 等. FMI 测井图像井壁复原方法优化及应用[J]. 测井技术, 2021, 45(4): 386-393.
LUO X, YAN J P, WANG M, et al. Optimization and application of borehole wall restoration method of FMI logging image[J]. Well Logging Technology, 2021, 45(4): 386-393(in Chinese).

[19] 陈长胜, 袁瑞, 王超, 等. 基于图像分解的微电阻率成像测井图像修复方法[J]. 长江大学学报 (自然科学版), 2019, 16(1): 95-99.
CHEN C S, YUAN R, WANG C, et al. Image restoration method for micro-resistivity imaging logging based on image decomposition[J]. Journal of Yangtze University (Natural Science Edition), 2019, 16(1): 95-99.

[20] 李振苓, 沈金松, 李思, 等. 成像测井电导率图像空白带奇异谱插值和缝洞孔隙度分离方法[J]. 测井技术, 2017, 41(1): 33-40.
LI Z L, SHEN J S, LI S, et al. Singular spectral interpolation of blank strips in formation micro-scanner conductivity image and separation of porosity between fracture and Karst cave[J]. Well Logging Technology, 2017, 41(1): 33-40(in Chinese).

[21] WANG Z F, GAO N, ZENG R, et al. A gaps filling method for electrical logging images based on a deep learning model[J]. Well Logging Technology, 2019, 43(6): 578-582.

[22] 袁晓涛, 马旭成, 肖仕军, 等. 融合多层语义特征的测井电成像空白条带填充深度神经网络方法[J]. 西安石油大学学报 (自然科学版), 2022, 37(6): 133-139.
YUAN X T, MA X C, XIAO S J, et al. A deep neural network method for filling blank strips in electrical logging images based on fusion of multi-level semantic features[J]. Journal of Xi'an Shiyou University (Natural Science Edition), 2022, 37(6): 133-139(in Chinese).

[23] 贾文玉, 田素月, 孙耀庭, 成像测井技术与应用[M]. 北京: 石油工业出版社, 2000.
JIA W Y, TIAN S Y, SUN Y T. Imaging logging technology and its application[M]. Beijing: Petroleum Industry Press, 2000(in Chinese).

[24] ISOLA P, ZHU J Y, ZHOU T H, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5967-5976.

[25] WANG T C, LIU M Y, ZHU J Y, et al. High-resolution image synthesis and semantic manipulation with conditional GANs[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8798-8807.

[26] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

Inpainting of electrical imaging logging images based on fast Fourier convolution

SU Qianxiao，QIAO Dexin*，REN Yili，FENG Zhou，LIN Shenglan，HUANG Ruiqi

(Research Institute of Petroleum Exploration and Development，Beijing 100083，China)

Abstract：Imaging logging is an important technique for complex reservoir evaluation. It provides a two-dimensional image of the wellbore, showing the wellbore's structural features and playing an important role in evaluating seam holes and sedimentary structures. However, the lack of a resistivity imaging logging device causes blank streaks to show up on the logging images, which makes computer data processing more challenging and affects visual continuity in manual identification. Currently, traditional image inpainting methods and neural networks do not perform well on filling the well logging images. Therefore, there is an urgent need to research a deep learning-based image inpainting method for blank streaks in imaging logging images. A dataset was constructed using electrical imaging logging images from the LG area of the Southwest Oil and Gas Field. This dataset was used to train a new deep learning algorithm for intelligent restoration of blank streaks based on fast Fourier convolution, based on a fast Fourier convolution neural network for filling blank streaks in imaging logging images. This technique solves the challenge of acquiring entire wellbore photos and makes it possible to quickly, accurately, and intelligently inpaint blank streaks in well logging images.

Keywords：electrical imaging logging；blank streaks filling；neural network；fast Fourier convolutions neural network；image inpainting

Received：2023-11-20；**Accepted**：2024-01-05；**Published Online**：2024-02-04 10：33

URL：link.cnki.net/urlid/11.2625.V.20240202.1216.002

Foundation items：National Natural Science Foundation of China (42372175); National Science and Technology Major Project (2025ZD1401501)

* **Corresponding author**. E-mail：qdx@petrochina.com.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0758

不同来流下双楔定常/非定常流动的模拟分析与流态预测

张鑫泽，李沁*，翁谊辉，尤延铖
(厦门大学 航空航天学院，厦门 361102)

摘　　要：针对高超声速双楔绕流中层流激波相互作用和定常/非定常流动现象，选取不同工况 (Ma_∞和 Re)，采用所发展的三阶精度本质无波动格式开展数值模拟与流动特性研究，总结由参数变化引起的定常/非定常流场激波结构和分离区的变化规律。为了对不同 Ma_∞、Re 下双楔流动出现的定常/非定常流态进行预测，基于最小二乘支持向量机 (LS-SVM) 模型，采用不同核函数（线性、多项式和径向基函数）构建双楔绕流定常/非定常的流态预测模型，并在验证集上对3种预测模型进行验证与评估。结果表明：径向基核预测模型能够正确预测16组工况下的双楔流态，预测结果的准确率和真阳性率为1、假阳性率为0，预测能力最强，具有较优的泛化能力，多项式核预测模型预测能力次之，线性核预测模型预测能力最差。

关　键　词：高超声速；双楔；定常/非定常流动；流态预测；最小二乘支持向量机
中图分类号：V211
文献标志码：A　　**文章编号**：1001-5965(2026)01-0371-14

双楔是高超声速飞行器、超燃冲压发动机和超声速进气道中常见的几何外形之一。例如，在高超声速飞行器中，机身、机翼和控制面等飞机部件常采用双楔简化构型；超燃冲压发动机及进气道中常采用双楔的外部压缩面来压缩空气、降低马赫数（如X-51[1]的入口等）。因此，有必要深入研究双楔绕流流动特性，探索双楔绕流定常/非定常流态的影响因素和机理，建立准确、高效的双楔绕流定常/非定常流态预测模型。

高超声速双楔流动中激波相互作用会引起飞行器表面局部高温和非定常振荡；对于超燃冲压发动机，激波相互作用可能会使相互作用区域附近的流动分离，并进一步增加局部的加热和压力，甚至导致进气道无法启动，影响发动机的正常工作。针对上述问题，学者对双楔绕流中激波相互作用和非定常现象进行了大量探索。Schrijer等[2]使用粒子图像测速法和纹影法研究了马赫数为7的双楔流动，双楔模型的前楔角和后楔角分别为15°和45°，观察到了V型干扰以及弓形激波和分离激波的非定常行为。Hashimoto[3]采用高速纹影成像技术研究了高超声速双楔流动的激波特性，发现当前楔角为25°、后楔角为40°和50°时，流场分别发生VI型和V型干扰，双楔的流动表现为定常，当后楔角为68°时，出现复杂的非定常流动。Swantek等[4]在超声速膨胀管中进行了一组前楔角为30°、后楔角为55°的双楔（30°/55°双楔）实验，获得了时间平均壁面热流数据和纹影图像。Komives等[5]针对Swantek等[4]的双楔模型开展了二维数值模拟，结果显示，壁面热流与实验测量值基本吻合，在激波干扰和分离中存在非定常行为，在一定条件下，数值模拟结果具有周期性。Durna等[6]针对前楔角为30°、后楔角为45°～55°的双楔开展了数值模拟，研究了低焓下双楔绕流周期性流动的复杂相互作用机制，发现对于双楔流动，一定存在一个后楔角的阈值，超过

收稿日期：2023-11-20；**录用日期**：2024-03-22；**网络出版时间**：2024-04-17 16:32
网络出版地址：link.cnki.net/urlid/11.2625.V.20240416.1058.002
*通信作者. E-mail：qin-li@vip.tom.com

引用格式：张鑫泽，李沁，翁谊辉，等. 不同来流下双楔定常/非定常流动的模拟分析与流态预测 [J]. 北京航空航天大学学报，2026，52（1）：371-384. ZHANG X Z，LI Q，WENG Y H，et al. Numerical analysis and flow state prediction of double wedge steady/unsteady flow at different in flow [J]. Journal of Beijing University of Aeronautics and Astronautics，2026，52（1）：371-384（in Chinese）.

该阈值后，流动具有时间周期性。Kumar等[7]通过改变后楔角度和前后楔面长度比对双楔进行了参数化研究，分析了几何参数的变化对非定常模式转变的影响，结果表明，分离区的大小影响激波相互作用，进而影响流动的定常或非定常特性。从以上研究不难发现，大部分学者多关注非定常流场激波结构的演化和几何外形参数对流场定常/非定常特性的影响，研究所采用的来流条件（如马赫数、雷诺数等）局限于Swantek等[4]实验的来流条件，对于不同来流条件对双楔定常/非定常流动的影响还有待研究。值得一提的是，Swantek等[4]指出：在Ma_∞=7.11、以1 m为特征长度的Re=1.1×10^6（后续Re均以此为定义）的工况下实验测得的壁面热流分布与层流状态下的理论热流吻合很好，显示双楔流动为层流流态；由于实验Re较低以及流态显示为层流，针对该工况所开展的数值模拟研究（如Komives[5]、Durna[6]、Kumar[7]及Ninni[8]等）也都采用层流模型。

为了对流态进行预测，需要将机器学习方法与计算流体力学（computational fluid dynamics, CFD）结合。近年来在多学科交叉发展的背景下，机器学习技术与CFD快速融合发展，特别是支持向量机(support vector machine, SVM)、最小二乘支持向量机 (least squares support vector machine, LS-SVM)已经被广泛用于气动力/热建模[9-10]、气动弹性预测[11]等方面。本文基于CFD方法，选取不同来流条件对Swantek等[4]实验所用的30°/55°双楔外形展开高超声速数值模拟和流动特性研究；在此基础上，针对双楔流动出现的层流定常、非定常2种不同流态，采用LS-SVM构建预测模型。

1 数值方法

1.1 控制方程

为模拟双楔模型绕流的定常/非定常流动，采用贴体坐标系下有限体积形式的二维层流可压缩纳维-斯托克斯（Navier-Stokes）方程：

$$\frac{\partial}{\partial t}\int_\Omega \boldsymbol{Q}\mathrm{d}V+\int_{\partial\Omega}(\boldsymbol{F}\boldsymbol{n})\mathrm{d}S=\int_{\partial\Omega}(\boldsymbol{F}_\mathrm{v}\boldsymbol{n})\mathrm{d}S \tag{1}$$

式中：$\boldsymbol{Q}$为守恒变量；$\boldsymbol{n}$为控制体表面外法向向量；$\boldsymbol{F}$和$\boldsymbol{F}_\mathrm{v}$分别为无黏通量和黏性通量，具体形式为：

$$\boldsymbol{Q}=\begin{bmatrix}\rho\\ \rho|\boldsymbol{u}|\\ \rho|\boldsymbol{v}|\\ \rho e\end{bmatrix},\boldsymbol{F}=\begin{bmatrix}\rho V\\ \rho|\boldsymbol{u}|V+n_x p\\ \rho|\boldsymbol{v}|V+n_y p\\ \rho HV\end{bmatrix},\boldsymbol{F}_\mathrm{v}=\begin{bmatrix}0\\ n_x\tau_{xx}+n_y\tau_{xy}\\ n_x\tau_{yx}+n_y\tau_{yy}\\ n_x\Theta_x+n_y\Theta_y\end{bmatrix}$$

其中：

$$V=\boldsymbol{v}\boldsymbol{n}=n_x|\boldsymbol{u}|+n_y|\boldsymbol{v}|,\quad H=E+p/\rho,\quad E=e+(|\boldsymbol{u}|^2+|\boldsymbol{v}|^2)/2$$

$$\begin{bmatrix}\Theta_x\\ \Theta_y\end{bmatrix}=\begin{bmatrix}\tau_{xx} & \tau_{xy}\\ \tau_{yx} & \tau_{yy}\end{bmatrix}\begin{bmatrix}|\boldsymbol{u}|\\ |\boldsymbol{v}|\end{bmatrix}+k\begin{bmatrix}\partial T/\partial x\\ \partial T/\partial y\end{bmatrix}$$

$$\tau_{x_ix_j}=\begin{cases}\mu\left(\dfrac{\partial u_i}{\partial x_j}+\dfrac{\partial u_j}{\partial x_i}\right) & i\neq j\\ 2\mu\dfrac{\partial u_i}{\partial x_i}-\dfrac{2}{3}\mu\nabla\boldsymbol{v} & i=j\end{cases}$$

式中：H为总焓；E为总能；e为比内能；k为热传导系数；μ为黏性系数，由萨瑟兰（Sutherland）公式得到

$$\frac{\mu}{\mu_0}=\left(\frac{T}{T_0}\right)^{1.5}\frac{1+T_S/T_0}{T/T_0+T_S/T_0} \tag{2}$$

其中：$T_0=288.16$ K；$\mu_0=1.789\times10^{-5}$ Pa·s；$T_S=110.4$ K。完全气体状态方程为$p=\rho RT$，其中，p为压力，ρ为密度，T为温度，R为气体常数。

1.2 计算格式

定常计算时间离散格式采用Jameson等[12]提出的LU-SGS(lower-upper symmetric Gauss-Seidel)隐式算法，非定常计算时间离散格式采用基于LU-SGS的双时间步算法[13]，具体计算过程如下：①用双时间步对所选工况进行非定常数值模拟，当始终无法得到一个不随迭代变化的流场时，认为流场为非定常，当能计算出不随迭代变化的流场时，认为该流场为定常；②选用LU-SGS进行定常数值模拟，得到定常的数值模拟结果；③对流项计算采用Steger-Warming通量分裂格式。通量重构采用Li等[14]提出的改进加权本质无波动格式WENO3-Z$_{\mathrm{ES3}}$，该格式在精度、分辨率和鲁棒性方面显示出优良特性。

考虑一维标量双曲守恒律方程：$u'_t+f(u')_x=0$，其中，$\partial f(u')/\partial u'>0$。在$x_j$处的半离散化守恒型差分格式形式如下：

$$(f(u')_x)_j\approx-\left(\hat{f}_{j+1/2}-\hat{f}_{j-1/2}\right)/\Delta x \tag{3}$$

式中：$\hat{f}_{j+1/2}$为数值通量；Δx为步长。对于WENO3-JS格式[15]，其重构算法如下：$\hat{f}_{j+1/2}=\sum_{k=0}^{r-1}\omega_k q_k^r$，其中，$q_k^r=\sum_{l=0}^{r-1}a_{kl}^r f\left(u'_{j-r+k+l+1}\right)$，$r$为格式候选模板的模板点数，$a_{kl}^r$通过使$q_k^r$达到$r$阶精度的方式确定，$\omega_k=\alpha_k\Big/\sum_{l=0}^{r-1}\alpha_l$为与线性权$d_k$对应的归一化非线性权，其中，$\alpha_k=d_k/\left(\varepsilon+\beta_k^{(r)}\right)^2$，$\varepsilon=10^{-6}\sim10^{-7}$，$\beta_k^{(r)}$为光滑度量因子，其定义为$\beta_k^{(r)}=\sum_{m=0}^{r-2}c_m^r\left(\sum_{l=0}^{r-1}b_{kml}^r f\left(u_{j-r+k+l+1}\right)\right)^2$，参数$a_{kl}^r$、$b_{kml}^r$和$c_m^r$的取值见文献[16]。

已有研究表明，传统的WENO-JS格式在一阶临界点会发生降阶问题。为了解决该问题，Borges等[17]

提出了 WENO-Z 方法, 通过引入一个新的非归一化非线性权使精度关系恢复, 其具体形式为 $\alpha_k = d_k\left(1+\left(\tau/\left(\beta_k^{(r)}+\varepsilon\right)\right)^p\right)$, 其中, τ为全局光滑度量因子, $p=1$或2, $\varepsilon=10^{-40}$。根据格式恢复精度的充分条件 $\omega_k^{\pm}-d_k=O(\Delta x^r)$, 其中, 上标“±”对应于网格点 $x_{j\pm1/2}$, 非归一化非线性权应满足: $\alpha_k=d_k(1+O(\Delta x^n)), n\geqslant 2$; 相应地要求:

$$\tau/\beta_k^{(r)}=O\left(\Delta x^{n'}\right)\qquad n'\geqslant 1 \tag{4}$$

Li 等采用提出的区间临界点分析方法[14] 对三阶 WENO-Z 类格式进行了精度分析, 证明在一阶临界点存在时, 式 (4) 无法满足。为解决该问题, Li 等[14] 将全局光滑度量因子的模板由 $\{x_{j-1},x_j,x_{j+1}\}$拓展为 $\{x_{j-1},x_j,x_{j+1},x_{j+2}\}$。但当临界点出现在某些特殊位置时, 式 (4) 仍然无法成立, 需要对光滑度量因子进行类似拓展, 进而得到在一阶临界点存在时始终满足$\beta_{k=0,1}^{(2)*}=O\left(\Delta x^4\right)$的新光滑度量因子。此时$\tau$只需满足: $\tau=O\left(\Delta x^5\right)$, 据此得到满足条件的全局光滑度量因子:

$$\tau_4=\left|\left(2f_{j+1}-3f_j+f_{j-1}\right)\left(f_{j+2}-3f_{j+1}+3f_j-f_{j-1}\right)\right| \tag{5}$$

新的光滑度量因子$\beta_{k=0,1}^{(2)*}$具体形式为

$$\begin{cases}\beta_0^{(2)*}=\beta_0^{(2)}+0.5\left(f_j-2f_{j-1}+f_{j-2}\right)^2\\ \beta_1^{(2)*}=\beta_1^{(2)}+0.15\left(f_{j+2}-2f_{j+1}+f_j\right)^2\end{cases} \tag{6}$$

最终得到 WENO3-Z_{ES3} 格式的非归一化非线性权为: $\alpha_k=d_k\left(1+0.4\left(\dfrac{\tau_4}{\beta_k^{(r)*}+\varepsilon}\right)^2\right)$, 其中, $\varepsilon=10^{-40}$。

1.3　计算验证

采用绕中空圆柱-长裙 (hollow cylinder exterded flave, HCEF)、尖双锥 (sharp double cone, SDC) 和球钝锥 3 种外形对所采用的数值方法进行计算验证。数值模拟的通量重构格式选用 WENO3-JS 和 WENO3-Z_{ES3}, 通量分裂均采用 Steger-Warming 格式。

1.3.1　绕中空圆柱-长裙流动

绕中空圆柱-长裙计算网格如图 1 所示, 其中。L 为中空圆柱-长裙的长度。网格数为该问题典型取值: 256×128, 计算条件为: Ma_∞=11.35、Re=3.596×10^5、来流温度 T_∞=79 K、壁面温度 T_w=294 K。

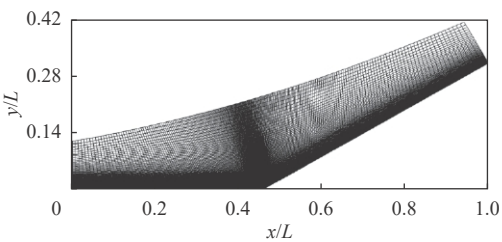


图 1　绕中空圆柱-长裙计算网格

Fig. 1　Computational mesh for the HCEF case

图 2 和图 3 给出了 WENO3-JS 与 WENO3-Z_{ES3} 2 种格式计算得到的绕中空圆柱-长裙壁面热流系数和压力系数分布与实验值的对比[18]。可以看到, WENO3-Z_{ES3} 格式结果与实验值吻合良好, WENO3-JS 格式得到的分离区则过小, 分离区后热流系数峰值过大, 分离区压力平台分布出现较大误差。

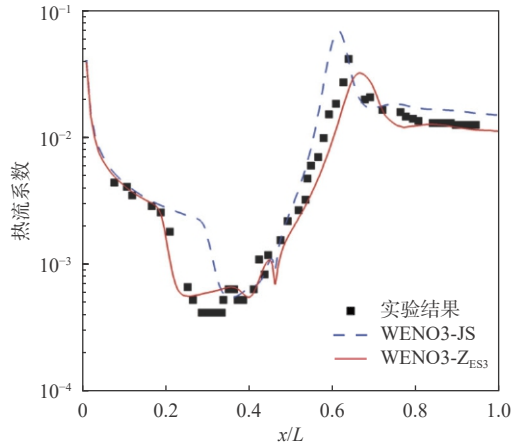


图 2　绕中空圆柱-长裙流动壁面热流系数分布

Fig. 2　The distribution of wall heat coefficient for HCEF

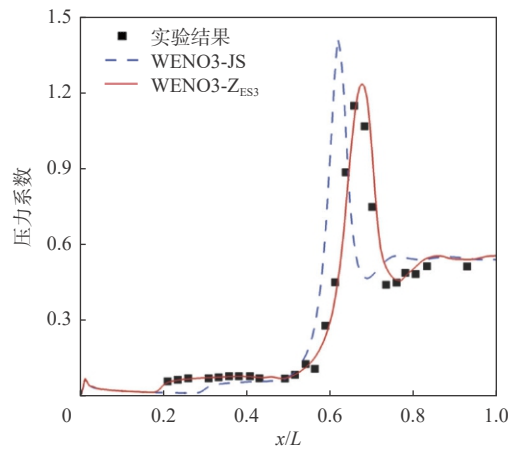


图 3　绕中空圆柱-长裙流动壁面压力系数分布

Fig. 3　The distribution of wall pressure coefficient for HCEF

1.3.2　高超声速绕 25°/55°尖双锥流动

绕尖双锥流动是典型的高超声速绕流问题, 常用于研究激波/激波与激波/边界层干扰现象。Moss 等[19] 进行了绕 25°/55°尖双锥流动的多种工况实验, 本文选择 Run28 工况进行数值模拟, 计算条件为: Ma_∞=9.59、Re=1.44×10^5、T_∞=42.6 K、T_w=293.3 K。网格如图 4 所示, 图中, 1 in=2.54 cm, 网格数为 256×128。

图 5 和图 6 给出了 WENO3-JS 与 WENO3-Z_{ES3} 2 种格式计算得到的尖双锥壁面热流和压力分布与实验值的对比[19]。可以看到, WENO3-JS 格式结果

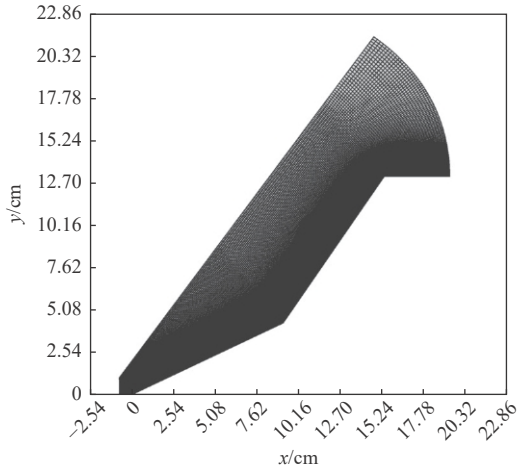


图4 绕尖双锥网格

Fig. 4 Computational mesh for the SDC case

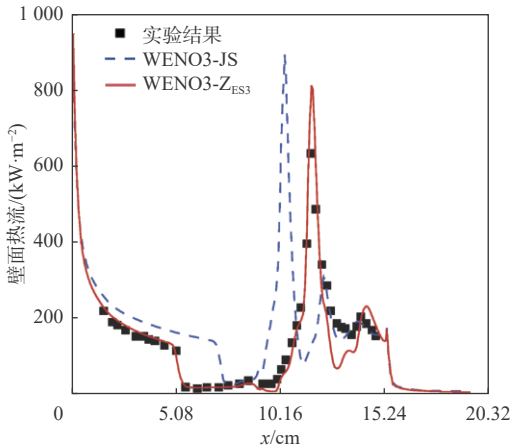


图5 绕尖双锥流动壁面热流分布

Fig. 5 The distribution of wall heat flux for SDC

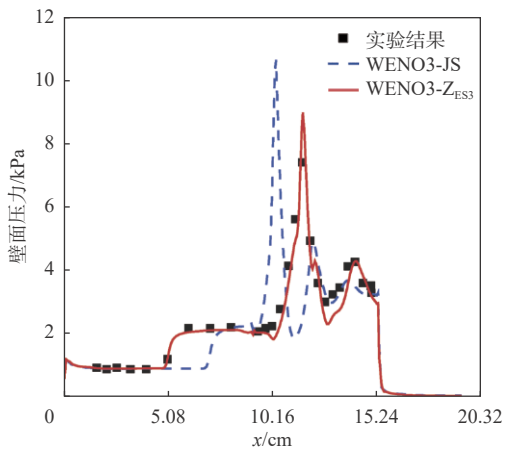


图6 绕尖双锥流动壁面压力分布

Fig. 6 The distribution of wall pressure for SDC

明显偏离实验值，具体表现为分离区偏小、压力峰值提前，而WENO3-Z_{ES3}格式结果均与实验值吻合良好。

1.3.3 Ma_∞=10.6绕球钝锥流动

球钝锥外形如图7所示，其半锥角为15°，头部

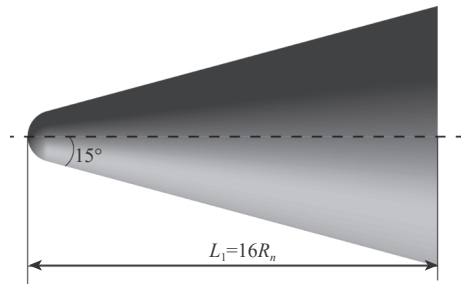


图7 球钝锥外形示意图

Fig. 7 Schematic diagram of the appearance of sphere-cone

球半径R_n=0.028 m，总长度L_1=16R_n。计算条件[20]为：Ma_∞=10.6、Re=4.7×10⁶、T_∞=47.3 K、T_w=294.44 K，来流攻角为20°，计算网格为97×69×181。

图8给出了WENO3-JS与WENO3-Z_{ES3} 2种格式计算得到的球钝锥壁面流向和轴向不同截面的热流分布与实验值对比[20]，其中，θ=0°为背风面，θ=180°为迎风面，参考热流Q_{ref}=2×10⁵ W/m²。可见，

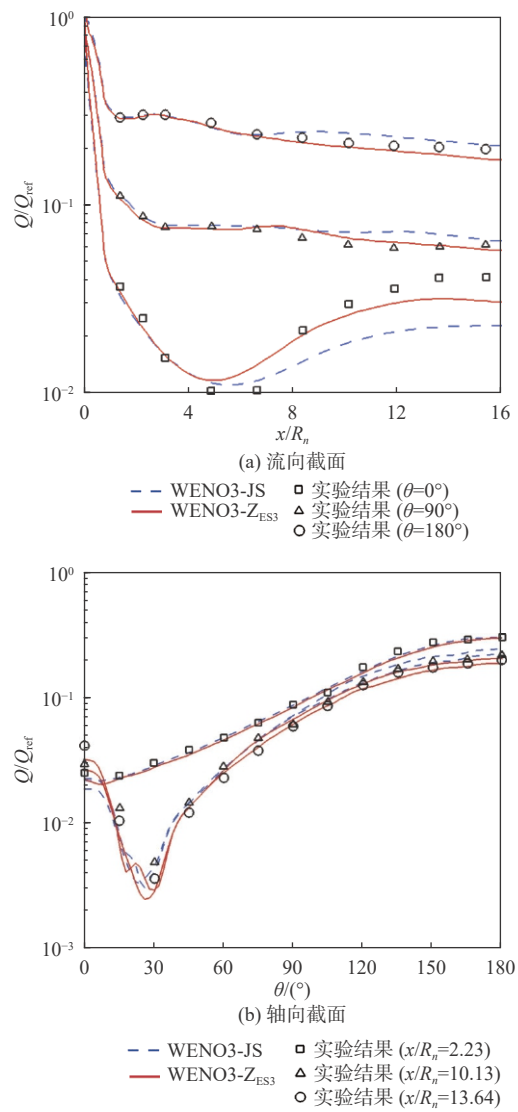


图8 绕球钝锥流动壁面热流分布

Fig. 8 The distribution of wall heat flux for sphere-cone

在球钝锥的尾部背风区, WENO3-JS 格式的热流计算值与实验值误差相对较大, 而 WENO3-Z_{ES3} 格式结果则与实验值较为接近。

在以上 3 个典型高超声速算例中, WENO3-Z_{ES3} 格式结果均明显优于 WENO3-JS 格式, 体现出 WENO3-Z_{ES3} 格式的显著精度优势。

2　不同 Ma_∞、Re 下 30°/55°双楔定常/非定常流动的模拟分析

2.1　计算模型与网格

Swantek 等[4] 采用膨胀管得到不同焓值的来流对双楔构型做了大量实验, 实验模型几何参数如下: 前楔角为 30°, 后楔角为 55°, 前楔面长度为 50.8 mm, 后楔面长度为 25.4 mm, 双楔宽度为 101.6 mm。实验主要针对双楔非定常流动中的激波相互作用、流动分离和热流峰值展开研究, 2011 年, Swantek 等[4] 的双楔模型作为典型外形被 NATO AVT 205 用来评估 CFD 预测高超声速壁面热流和压力分布的能力, 因此, 本文也参考 Swantek 等[4] 的实验模型建立二维计算模型并绘制网格, 计算网格如图 9 所示, 考虑到后楔前方弓形激波后存在较大范围的亚声速区, 在后楔后面接一段长度为 10 mm 的平面采用超声速出口边界条件。计算网格采用结构网格, 对 2 个楔面连接的位置做了过渡处理以确保壁面网格的正交性。

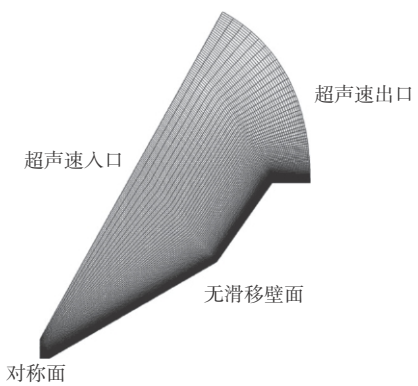


图 9　计算网格示意

Fig. 9　Computational grid diagram

边界条件为: 楔面及其后接的平面为壁面条件, 入口为超声速入口条件, 出口为超声速出口条件, 双楔前端为对称面边界条件。壁面设为无滑移的等温壁, 壁温为 298 K[4]。

2.2　网格收敛性与非定常计算的验证

为了确定合适的网格密度, 选取 2 种工况(具体见表 1), 采用 256×128(Grid 1)、512×256(Grid 2)以及 768×384(Grid 3)3 个网格开展了网格收敛性验证计算, 法向第 1 层网格高度为 2.2×10^{-6} m[21]。

表 1　网格收敛性验证来流条件

Table 1　Inflow conditions of mesh convergence verification

网格类型	Re	Ma_∞	T_∞/K
Case 1	5×10^4	7	191
Case 2	5×10^4	8	191

本文区分定常和非定常流动的标尺是流场的物理量(如密度、速度和压力等)是否随时间的推移而发生改变, 若物理量都不随时间的变化而变化, 就认为该流场是定常流动, 反之就是非定常流动。

计算结果显示, Case 1 和 Case 2 这 2 种来流条件下的双楔流场为定常流动。图 10 给出了 2 种来流条件下 3 个网格计算得到的壁面热流分布。可以看到, Grid 2 和 Grid 3 的计算结果重合, 而 Grid 1 结果存在一定差异, 特别是 x=51～60 mm 时, 热流分布和峰值相对于另外 2 个网格计算结果差异明显。综合数值模拟结果, 可以认为 Grid 2 和 Grid 3 计算达到网格收敛。为了提高计算效率, 在后续

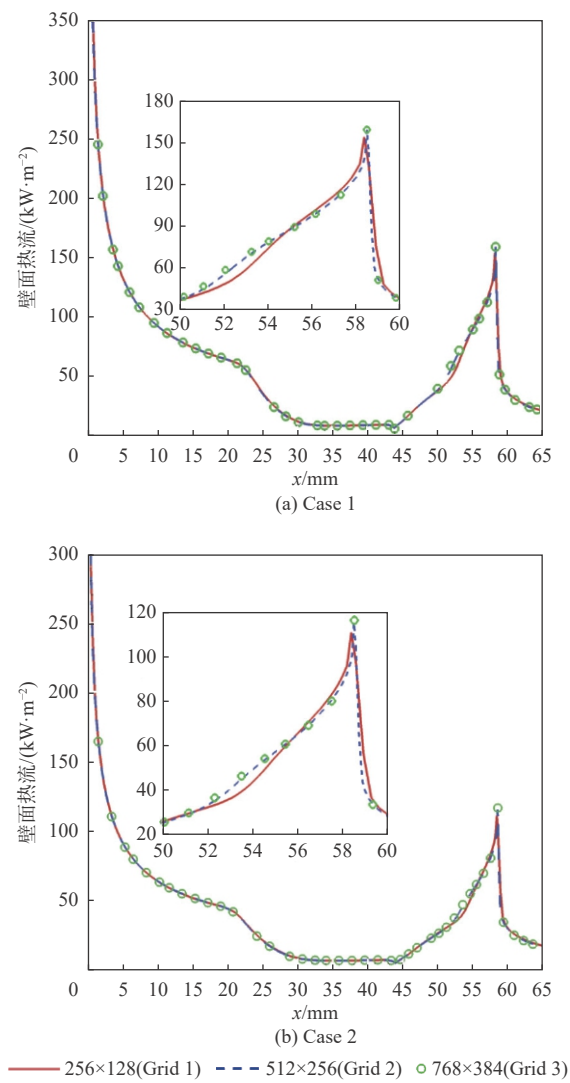


图 10　网格收敛性分析

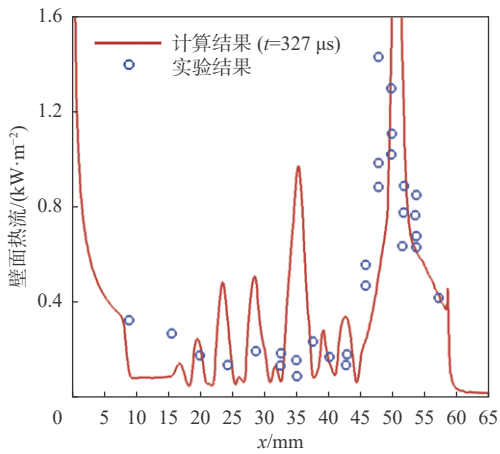
Fig. 10　Analysis of mesh convergence

的计算中选用 Grid 2。

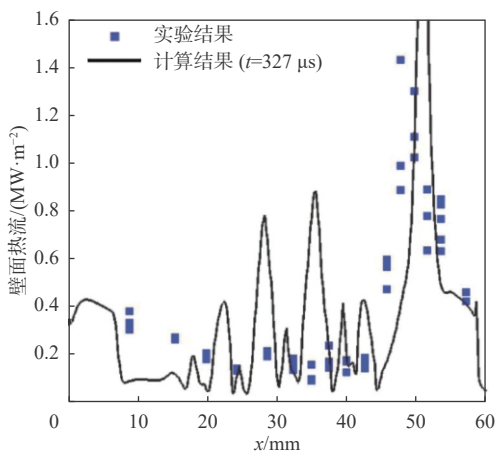
除了定常流动外，双楔还存在非定常流动情况。由文献调研可知：在本文开展研究的 Ma、Re 范围，已开展的数值模拟研究[5-8,22]中的非定常结果显示，双楔流场会出现激波结构在楔面上的非定常运动，但此时流场没有小尺度无规则结构，因此，文献都仍采用层流模型对有限时间范围变化的流动进行模拟分析。此外，本文的主要工作之一是针对定常/非定常流态的转变进行预测，主要关注流场是否出现非定常现象，而对于非定常出现后、流动是否会进一步发生转捩等更复杂演化及其特征尚不作为研究内容。基于以上考虑，在本文的非定常计算中，采用层流模型对有限时间范围内的流动演化进行模拟。

双楔非定常流动计算验证使用 Swantek 等[4]实验的来流条件，具体参数为：Ma_∞=7.11、Re=1.1×10^6、ρ_∞=0.007 1 kg/m^3、T_∞=191 K、来流压力 P_∞=391 Pa、来流速度 u_∞=1972 m/s、h_∞=2.1 MJ/kg，时间步长取 1×10^{-3} μs，与 Durna[23]、Kumar[24]等保持一致。

为考察非定常计算结果的正确性，图 11 给出了 t=327 μs 时本文和 Kninght 等[25]的数值模拟结果。可以看到，两者的结果都显示流场在 x=9 mm 左右出现分离，计算得到的壁面热流在 20 mm< x<44 mm 区域显示出与 Knight 等[25]计算结果一致的振荡行为，并且壁面热流峰值位置也与实验结果较为接近。



(a) 本文计算结果



(b) 文献[25]计算结果

图 11 壁面热流分布 (t=327 μs)

Fig. 11 Wall heat flux distribution at 327 μs

2.3 典型来流条件下双楔定常/非定常流场的数值模拟与分析

对所采用数值模拟方法和网格进行验证后，针对 30°/55°双楔在 Ma_∞=6～10 和 Re=1×10^4～1×10^5 范围内开展了一系列数值模拟，图 12 显示了不同来流条件下流动定常/非定常结果。可见，当 Ma_∞一定时，随着 Re 的增大，流态从定常过渡到非定常；当 Re 一定时，在一定工况下（如 Re=1×10^4～3.5×10^4 和 Re=6×10^4～1×10^5），流态随 Ma_∞的增大不会发生变化，表现为确定的定常或非定常流态；在其余工况下，流态随 Ma_∞的改变呈现出一定的变化规律，其规律将在第 3 节中进一步研究。

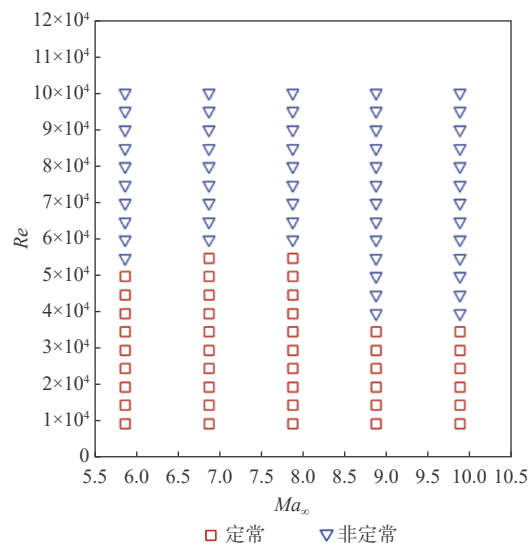


图 12 不同来流条件下的定常/非定常流态

Fig. 12 Steady/ unsteady flow states at different inflow conditions

2.3.1 Ma_∞=8、不同 Re 下双楔定常流场的模拟分析

为了研究不同 Re 对双楔定常流场激波结构和流动分离的影响，选取 Ma_∞=8，Re=1×10^4、3×10^4 和 5×10^4 这 3 种工况进行分析，图 13 给出了 3 种工况下流场的无量纲压力等值线，其中，分离区已用流线显示。结果表明，流场中基本激波结构为前缘激波、分离激波、弓形激波及透射激波。前缘激波和分离激波两者相交合成一道激波，该激波与弓形激波相互作用形成透射激波，透射激波与壁面边界层发生交汇时会产生强烈的逆压

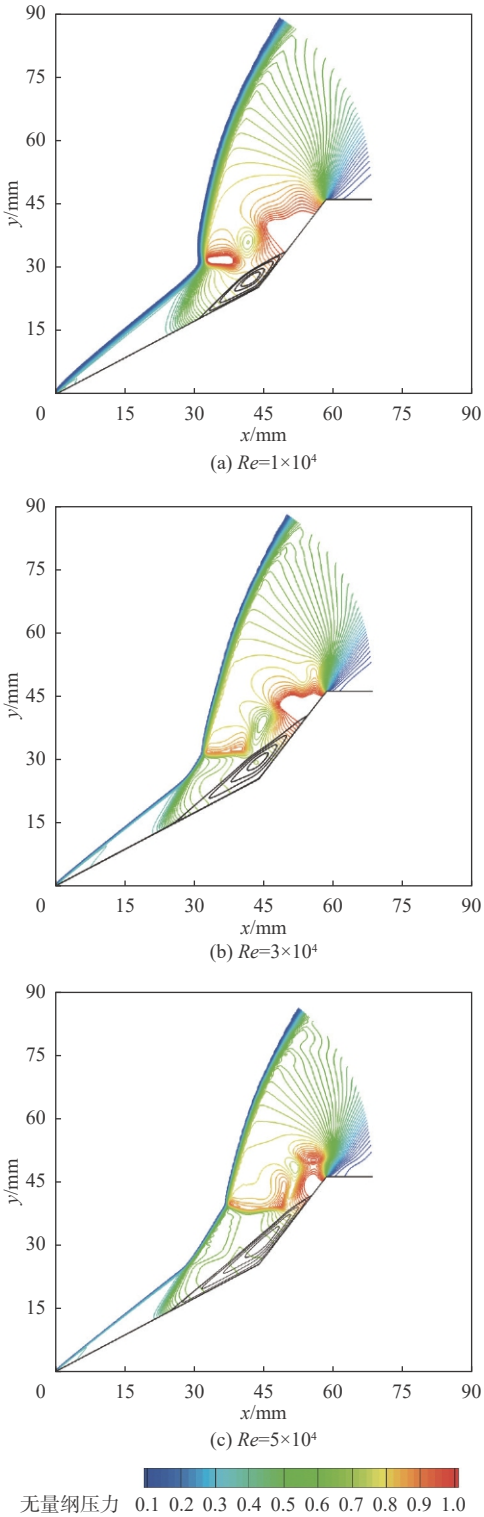


图 13　不同 *Re* 的无量纲压力等值线分布和分离区流线

Fig. 13　No-dimensional pressure contour distribution in the separation region streamline at different *Re*

梯度。逆压梯度通过边界层的亚音速部分同时向上游和下游移动, 形成大范围的流动分离区。

从图 13 可以看出, 随着 *Re* 的增加, 分离激波的波足向上游移动。为了更清晰地显示分离点、再附点的位置和分离区的尺度变化, 将不同 *Re* 下的双楔壁面摩阻系数分布展示在图 14。可以看到, 随着 *Re* 的逐渐增加, 分离点和再附点之间的距离不断增大, 反映出分离区的尺度在逐渐增大。

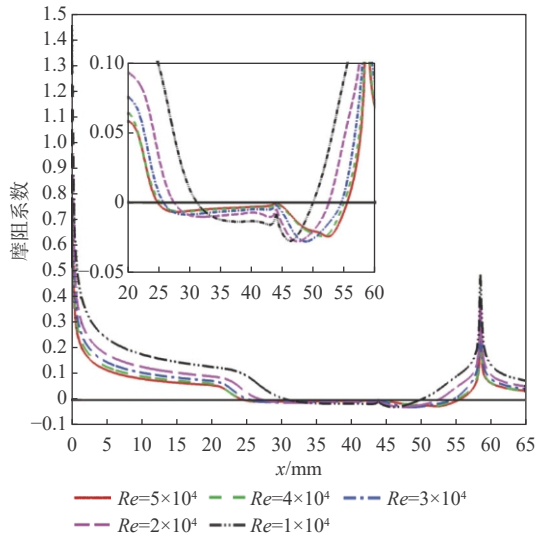


图 14　不同 *Re* 的壁面摩阻系数分布

Fig. 14　Wall friction coefficient distribution at different *Re*

为了定量显示上述变化, 表 2 给出了不同 *Re* 下双楔流动的分离点坐标、再附点坐标及分离区的尺度, 分离区的尺度定义为分离点到楔面折角处再到再附点的距离。

表 2　不同 *Re* 的分离区尺度

Table 2　Dimensions of separation region at different *Re*

Re	分离点坐标/mm	再附点坐标/mm	分离区尺度/mm
1×10^4	(31.178, 18.004)	(49.909, 33.848)	25.110
2×10^4	(27.478, 15.865)	(52.464, 37.499)	33.840
3×10^4	(25.892, 14.981)	(54.460, 40.374)	39.156
4×10^4	(25.027, 14.450)	(54.919, 41.011)	40.955
5×10^4	(24.762, 14.302)	(55.428, 41.733)	42.142

2.3.2　*Re*=3×10⁴、不同 Ma_∞ 下双楔定常流场的模拟分析

为了研究不同 Ma_∞ 对双楔定常流场激波结构和流动分离的影响, 选取 *Re*=3×10⁴, Ma_∞=6, 8, 10 这 3 种工况进行分析, 图 15 给出了 3 种工况下流场的无量纲压力等值线, 其中, 分离区已用流线显示。流场中基本激波结构与 2.3.1 节所描述的大致相同。

从图 15 可以看出, 随着 Ma_∞ 的增加, 分离激波的波足向下游移动。为了更清晰地显示分离点、再附点的位置和分离区的尺度变化, 将不同 Ma_∞ 下的双楔壁面摩阻系数分布展示在图 16。可以看到, 随

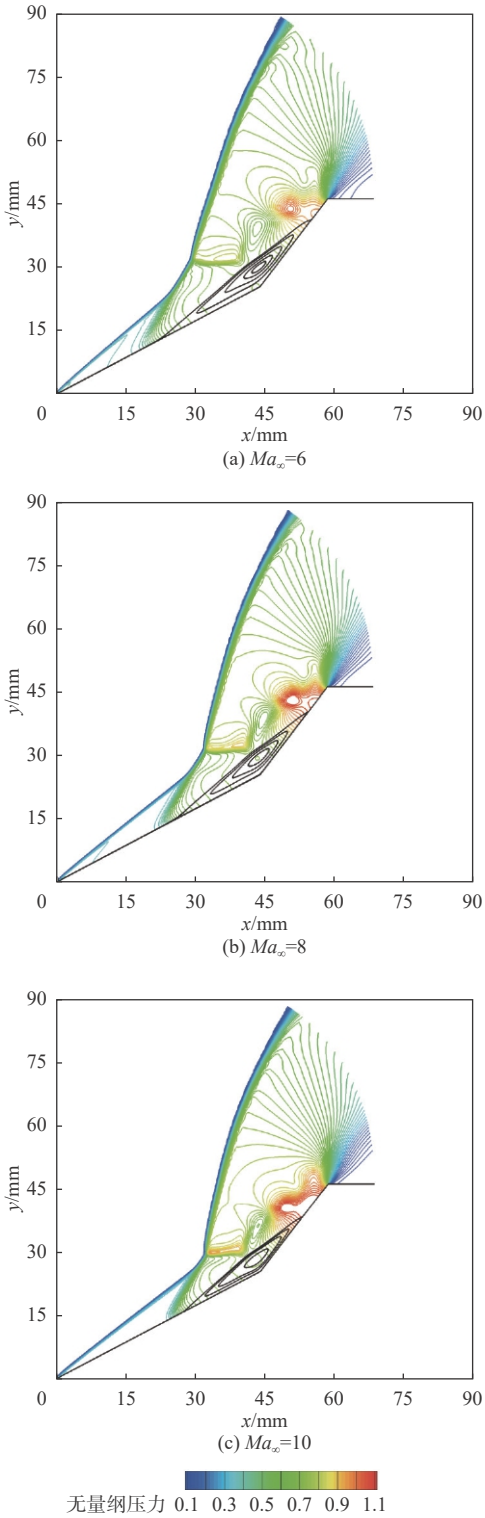


图 15 不同 Ma_∞的无量纲压力等值线分布和分离区流线

Fig. 15 No-dimensional pressure contour distribution and separation region streamline at different Ma_∞

着 Ma_∞的逐渐增加，分离点和再附点之间的距离不断缩小，反映出分离区的尺度在逐渐减小。

为了定量显示上述变化，表 3 给出了不同 Ma_∞下双楔流动的分离点坐标、再附点坐标及分离区的尺度。

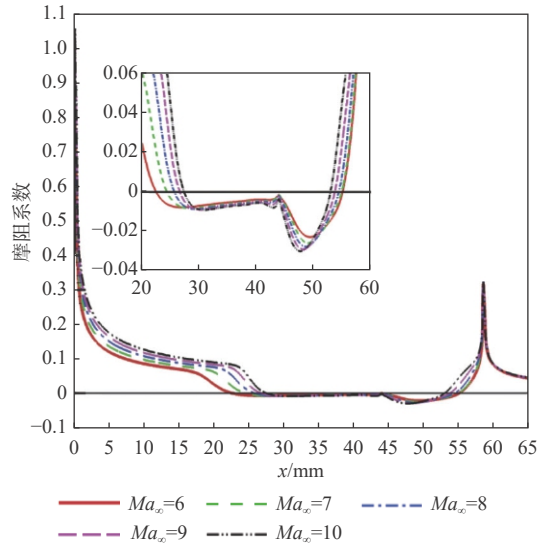


图 16 不同 Ma_∞的壁面摩阻系数分布

Fig. 16 Wall friction coefficient distribution at different Ma_∞

表 3 不同 Ma_∞的分离区尺度

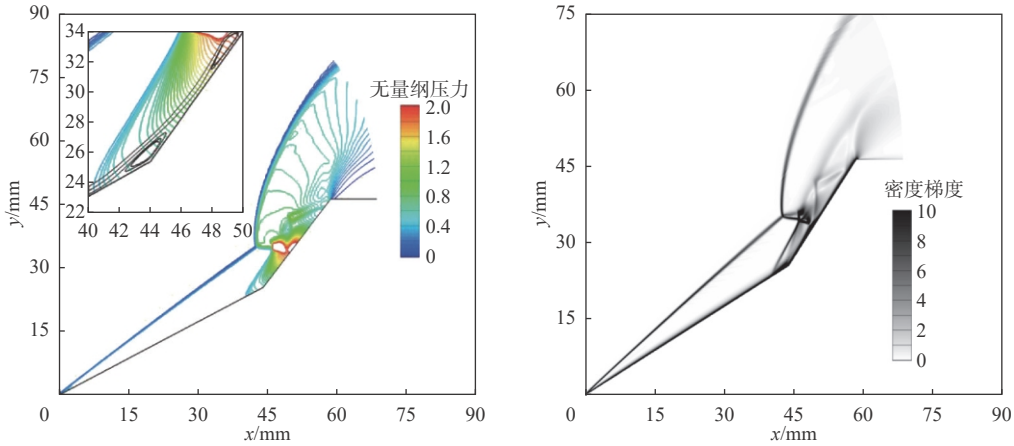
Table 3 Dimensions of separation region at different Ma_∞

Ma_∞	分离点坐标/mm	再附点坐标/mm	分离区尺度/mm
6	(22.506, 13.053)	(55.134, 41.305)	44.201
7	(24.406, 14.096)	(54.978, 41.090)	41.768
8	(25.687, 14.981)	(54.484, 40.389)	39.360
9	(26.656, 15.515)	(53.783, 39.439)	37.073
10	(27.401, 15.857)	(53.023, 38.321)	34.905

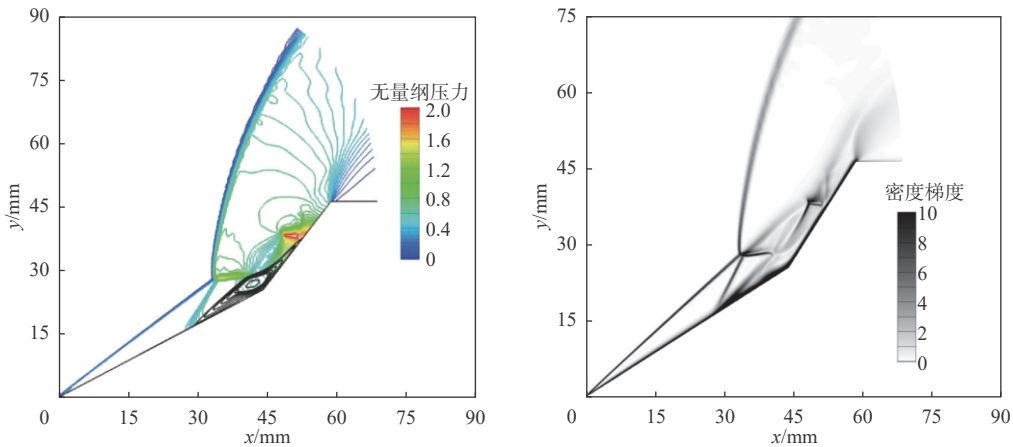
2.3.3 Ma_∞=10、Re=1×10^5 下双楔非定常流场的模拟分析

由于 Swantek 等[4]选取的工况 Re 较大且有较多学者研究，本文选取的非定常计算工况为 Ma_∞=10、Re=1×10^5、T_∞=191 K 及 T_w=300 K，时间步长取 1×10^{-3} μs，与上文取值一致。

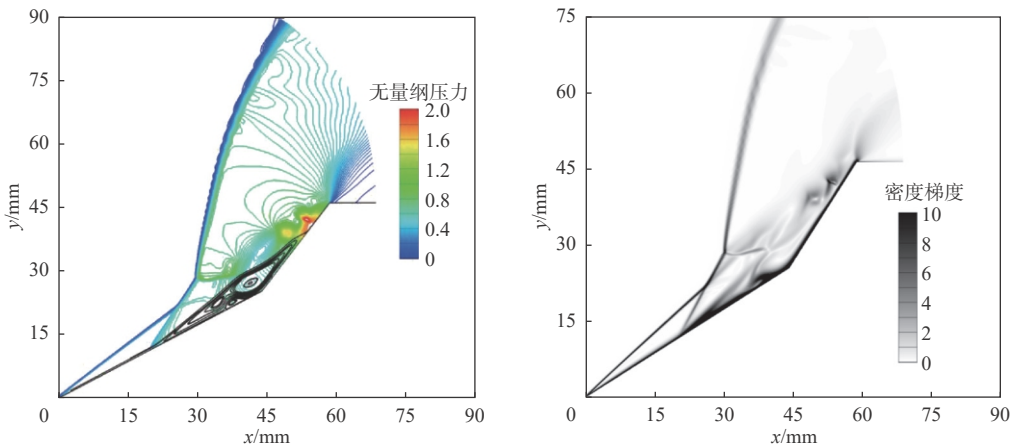
图 17 显示 t=60, 180, 300 μs 时二维模拟得到的无量纲压力等值线分布和密度梯度分布。在 t=60 μs 时，前缘激波和分离激波已基本建立；前缘激波和弓形激波在三波点相互作用，形成透射激波和滑移线；楔面连接处的分离区处于早期演化阶段。由于分离区相对较小，分离激波与透射激波发生干扰。在 t=180 μs 时，分离区变大，三波点向上游移动，激波之间的干扰类型未发生改变。在 t=300 μs 时，随着分离区的不断增大，分离激波继续向上游移动，分离激波不与透射激波相互作用，而是先与前缘激波发生干扰，相交后产生一道激波，该激波与弓形激波发生相互作用产生透射激波，流动分离由单次分离转化为二次分离，并出现 3 个分离涡。



(a) t=60 μs



(b) t=180 μs



(c) t=300 μs

图 17　不同时刻的无量纲压力等值线分布、分离区流线及密度梯度分布

Fig. 17　No-dimensional pressure contour distribution, separation region streamline and density gradient distribution at different instants

为了更清晰地显示分离点、再附点的位置、分离区的尺度变化及壁面热流和压力随时间的演变过程, 图 18～图 20 分别给出不同时刻的双楔壁面摩阻系数、壁面热流和壁面压力的分布。前楔上热流的突然下降和压力的突然增加表明了边界层的分离, 后楔上的流动再附引起了热流和压力峰值。随着时间的推进, 分离点向上游移动, 而热流和压力峰值向下游移动, 分离区有不断变大的趋势。

为了定量显示上述变化, 表 4 给出了不同时刻双楔流动的分离点坐标、再附点坐标及分离区的尺度, 可以看出, 分离点向上游移动, 而再附点向下游移动, 虽然 120 μs 时的再附点与 60 μs 对比是向上游移动, 但分离区总体变化趋势是增大。

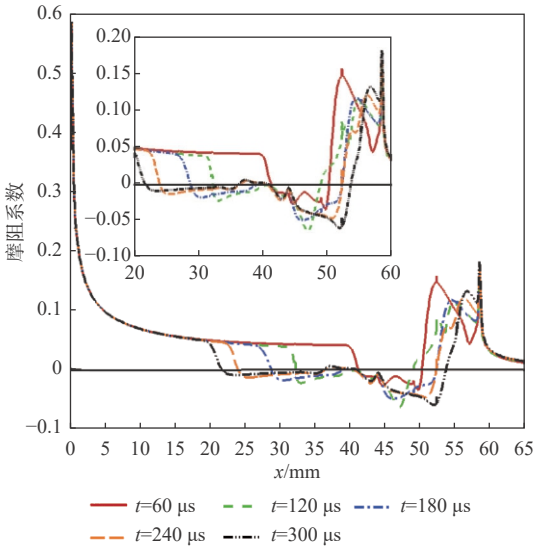


图 18 不同时刻的壁面摩阻系数分布

Fig. 18 Wall friction coefficient distribution at different instants

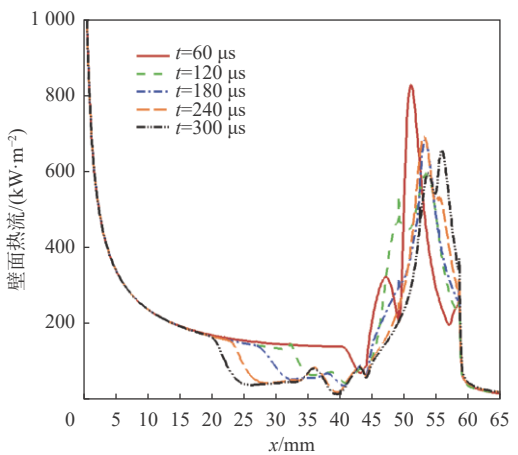


图 19 不同时刻的壁面热流分布

Fig. 19 Wall heat flux distribution at different instants

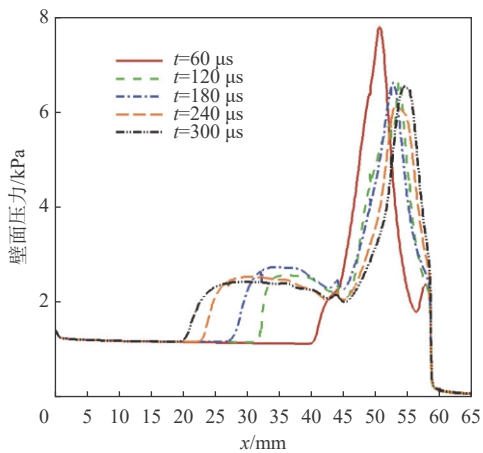


图 20 不同时刻的壁面压力分布

Fig. 20 Wall pressure distribution at different instants

表 4 不同时刻的分离区尺度

Table 4 Dimensions of separation region at different instants

t/μs	分离点坐标/mm	再附点坐标/mm	分离区尺度/mm
60	(41.388, 23.896)	(50.292, 34.401)	13.995
120	(32.172, 18.586)	(48.914, 32.427)	22.223
180	(28.824, 16.647)	(52.321, 37.299)	32.037
240	(24.034, 13.883)	(52.345, 37.331)	37.608
300	(21.760, 12.564)	(53.694, 39.258)	42.588

3 30°/55°双楔流动定常/非定常流态的 LS-SVM 模型预测

通过第 2 节的数值模拟可知，不同 Ma_∞和 Re 下，双楔流场存在定常和非定常 2 种基本流态，如何确定定常/非定常流态边界并根据有限的数值模拟结果对流态定常/非定常进行最大泛化性能的预测是值得研究的课题。本节采用 LS-SVM 方法，使用不同 Ma_∞和 Re 下的双楔流场定常/非定常结果数据对模型进行训练，得到基于 3 种不同核函数的 LS-SVM 双楔流场定常/非定常流态预测模型，选取验证集对预测模型进行验证与评估，通过对比得到相对准确的双楔流动流态预测模型。

3.1 LS-SVM 模型

为了解决机器学习算法在小样本学习时出现的维数灾难和过拟合问题，Cortes 等[26] 在统计学习理论的基础上提出了 SVM 算法，其目的是在一个特征空间中寻找最优边界，从而实现对目标数据的准确高效分类。SVM 具有鲁棒性好、减少高维计算等优点，在图像识别、气动力建模中得到广泛应用。假设样本 $(\boldsymbol{x}_i, y_i), i=1,2,\cdots,N$，$\boldsymbol{x}_i$为真实输入，$y_i \in \{+1,-1\}$为样本种类。SVM 分类超平面可使用函数 $f(\boldsymbol{x}_i)=\boldsymbol{w}\phi(\boldsymbol{x}_i)+b$表示，式中，$\boldsymbol{w}$为权重向量，$b$为偏置项，$\phi(\boldsymbol{x}_i)$为样本$\boldsymbol{x}_i$的非线性映射，其作用是将数据从低维映射到高维空间，使数据具有线性可分性。表示方法为：满足 $f(\boldsymbol{x}_i)=0$的$\boldsymbol{x}_i$构成最优分类超平面；满足 $f(\boldsymbol{x}_i)=1$的$\boldsymbol{x}_i$构成正超平面；满足 $f(\boldsymbol{x}_i)=-1$的$\boldsymbol{x}_i$构成负超平面。

SVM 将最优分类问题转化为如下的二次优化问题[26]：

$$\begin{cases} \min \dfrac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{s.t.} \quad y_i\left[(\boldsymbol{w}\phi(\boldsymbol{x}_i))+b\right]-1 \geqslant 0 \quad i=1,2,\cdots,N \end{cases} \tag{7}$$

通过求解二次优化问题，确定$\boldsymbol{w}$和b。

为求解上述二次优化问题，构造拉格朗日(Lagrange)函数：

$$L(\boldsymbol{w},b,\lambda)=\frac{1}{2}\|\boldsymbol{w}\|^2-\sum_{i=1}^{N}\lambda_i\left\{y_i\left[(\boldsymbol{w}\phi(\boldsymbol{x}_i))+b\right]-1\right\} \tag{8}$$

式中: $\lambda_i \geqslant 0$为 Lagrange 乘子, 对式 (8) 求$\boldsymbol{w}$和b的偏导并令其为 0 可解得: $\sum_{i=1}^{N}\lambda_i y_i \phi(\boldsymbol{x}_i)=\boldsymbol{w}$, $\sum_{i=1}^{N}\lambda_i y_i=0$。利用对偶原理, 结合$\sum_{i=1}^{N}\lambda_i y_i \phi(\boldsymbol{x}_i)=\boldsymbol{w}$, $\sum_{i=1}^{N}\lambda_i y_i=0$可以将式 (8) 转化为仅与$\lambda$有关的对偶问题:

$$\begin{cases}\max\left(\sum_{i=1}^{N}\lambda_i-\frac{1}{2}\sum_{i,j=1}^{N}\lambda_i\lambda_j y_i y_j K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)\right)\\ \text{s.t.}\quad \sum_{i=1}^{N}\lambda_i y_i=0 \qquad \lambda_i\geqslant 0, i=1,2,\cdots,N\end{cases} \tag{9}$$

式中: $K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\phi(\boldsymbol{x}_i)\phi\left(\boldsymbol{x}_j\right)$为核函数, 常见的核函数如表 5 所示。

表 5　常见的核函数

Table 5　Common kernel functions

核函数名称	表达式
线性	$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\boldsymbol{x}_i\boldsymbol{x}_j$
多项式	$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\left(\boldsymbol{x}_i\boldsymbol{x}_j\right)^d$
径向基	$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\exp\left(-\frac{\left\|\boldsymbol{x}_i\boldsymbol{x}_j\right\|^2}{2\sigma^2}\right)$
Sigmoid函数	$K\left(\boldsymbol{x}_i,\boldsymbol{x}_j\right)=\tanh\left(\varphi\left(\boldsymbol{x}_i\boldsymbol{x}_j\right)+\theta\right)$

对偶问题求解得到的最优解为λ, 代入$\boldsymbol{w}$和b, 最终 SVM 分类模型可表示如下:

$$y(\boldsymbol{x})=\operatorname{sgn}(f(\boldsymbol{x}))=\operatorname{sgn}(w\phi(\boldsymbol{x})+b)=\operatorname{sgn}\left(\sum_{i=1}^{N}\lambda_i y_i K(\boldsymbol{x}_i,\boldsymbol{x})+b\right) \tag{10}$$

式中: sgn为符号函数。

LS-SVM 是 Suykens 等[27]在标准 SVM 基础上提出的, LS-SVM 算法将标准 SVM 约束条件中的不等式变为等式, 将优化问题的求解变为线性方程组的求解, 避免了不等式二次规划的复杂性。LS-SVM 优化问题表示如下:

$$\begin{cases}\min\limits_{\boldsymbol{w},b,\gamma}\left(\frac{1}{2}\|\boldsymbol{w}\|^2+\frac{C}{2}\sum_{i=1}^{N}\gamma_i^2\right)\\ \text{s.t.}\quad y_i-\gamma_i=\boldsymbol{w}\phi(\boldsymbol{x}_i)+b \qquad 1\leqslant i\leqslant N\end{cases} \tag{11}$$

式中: γ_i为松弛变量; C为惩罚因子, 控制误差的惩罚程度, 通过交叉验证寻优得到。

引入 Lagrange 乘子构造相应的 Lagrange 函数如下:

$$L(\boldsymbol{w},b,\gamma,\lambda)=\frac{1}{2}\|\boldsymbol{w}\|^2+\frac{C}{2}\sum_{i=1}^{N}\gamma_i^2-\sum_{i=1}^{N}\lambda_i(y_i-\gamma_i-\boldsymbol{w}\phi(\boldsymbol{x}_i)-b) \tag{12}$$

根据卡罗需-库恩-塔克(Karush-Kuhn-Tucker)最优条件求解式 (12) 可得, $\sum_{i=1}^{N}\lambda_i\phi(\boldsymbol{x}_i)=\boldsymbol{w}$, 最终 LS-SVM 的分类模型可表示如下:

$$y(\boldsymbol{x})=\operatorname{sgn}\left(\sum_{i=1}^{N}\lambda_i K(\boldsymbol{x}_i,\boldsymbol{x})+b\right) \tag{13}$$

3.2　基于 LS-SVM 的双楔流动定常/非定常流态预测及验证

对第 2 节的一系列数值模拟结果进行分析, 发现在不同 Ma_∞、Re 下 30°/55°双楔流场呈现定常或非定常 2 种流态, 存在相应的边界, 流场的定常/非定常流态与 Ma_∞和 Re 的组合之间并不是简单的线性关系。本节将应用 LS-SVM, 采用 Ma_∞和 Re 作为模型的输入参数, 将流场的定常与否作为模型的输出参数, 建立预测 30°/55°双楔流场定常/非定常流态的 LS-SVM 机器学习模型。具体模型训练、验证和评估流程如下:

1) 数据收集: 获取来流条件 Ma_∞=6～10, Re=1×10^4～1×10^5 下共 95 个工况的双楔流场定常/非定常流态结果, 建立训练集。

2) 模型训练: 选择核函数, 使用训练集训练 LS-SVM 模型, 并通过交叉验证寻优得到最优参数。

3) 模型验证: 建立验证集并将模型预测结果与 CFD 计算结果对比, 验证准确性。

4) 模型评估: 对模型进行全方面的量化评估。

为了对比不同核函数对预测模型的影响, 采用 3 种核函数训练 LS-SVM 模型, 分别线性核函数、多项式核函数及 RBF 核函数, 具体如表 5 所示。通过交叉验证寻优得到 3 种核函数的最优参数: 线性核函数, $C=56.3$; 多项式核函数, $C=1.99$、$d=5$; 径向基核函数, $C=56.3$、$\sigma=0.285$。经过训练得到的 LS-SVM 机器学习模型如图 21 所示, 蓝色部分为非定常流态, 紫色部分为定常流态, 中间分界线为定常/非定常流态边界。结果显示, 线性核函数训练模型预测相对最为粗糙, 径向基核函数训练模型则能很好地将训练集分类。

为了对模型的预测能力进行检验, 设计了由 16 组数据组成的验证集, 分别使用 3 种预测模型进行预测, 得到的预测结果如图 22 所示, 图中不同线型上方表示预测结果为非定常, 不同线型下方表示预测结果为定常; 符号表示验证集, 其中, 三角形表示 CFD 结果为非定常, 正方形表示 CFD 结果为定常。可以看到, 线性核函数 LS-SVM 预测模型的预测结果最差, 预测结果大部分是错误的; 其次是多项式核函数 LS-SVM 预测模型, 预测结果大部分是正确的, 但也存在少量的错误预测; 径向基核函数 LS-SVM 预测模型表现最好, 预测结果与 CFD 计算

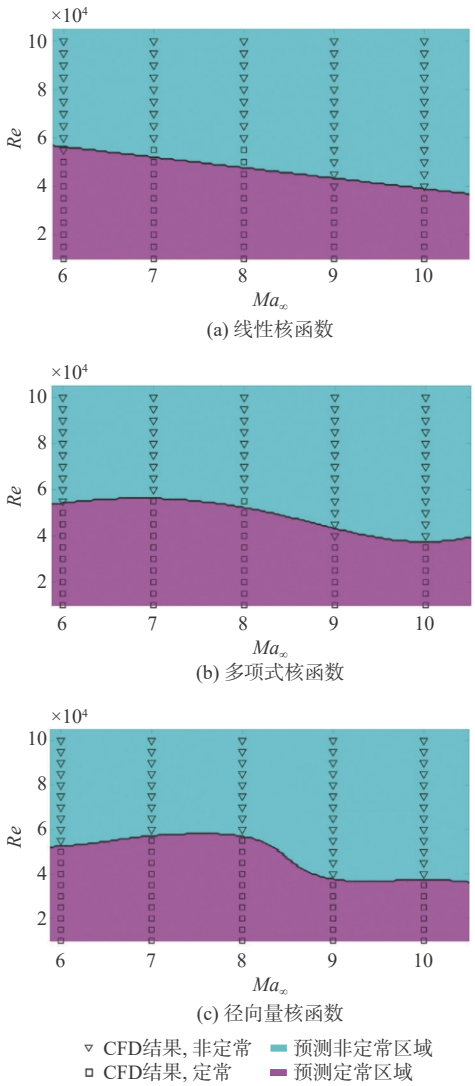


图 21 不同核函数的 LS-SVM 模型的预测结果

Fig. 21 Prediction results of LS-SVM models with different kernel functions

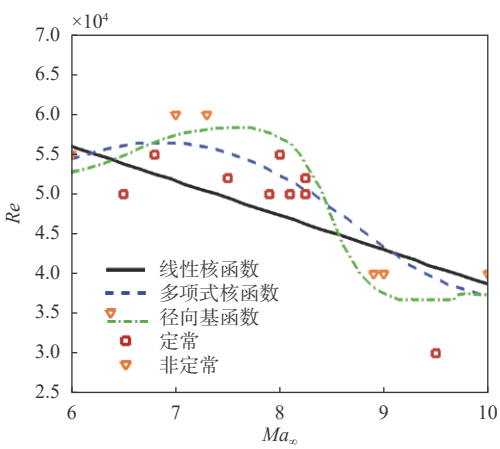


图 22 不同核函数预测结果

Fig. 22 Prediction results of different kernel functions

结果保持一致，没有发生错误预测。

为了进一步评估预测模型的准确性，使用混淆矩阵显示预测结果。混淆矩阵的取值为对应预测结果，CFD 结果的状态个数。3 种核函数预测模型的混淆矩阵如表 6～表 8 所示。

表 6 线性核函数混淆矩阵

Table 6 Confusion matrix of linear kernel function

预测结果	CFD结果	
	定常	非定常
定常	2	3
非定常	7	4

表 7 多项式核函数混淆矩阵

Table 7 Confusion matrix of polynomial kernel function

预测结果	CFD结果	
	定常	非定常
定常	7	2
非定常	2	5

表 8 径向基核函数混淆矩阵

Table 8 Confusion matrix of RBF kernel function

预测结果	CFD结果	
	定常	非定常
定常	9	0
非定常	0	7

预测结果、CFD 结果都为定常称为真阳性 T_P，反之称为真阴性 T_N；预测结果为非定常、CFD 结果为定常称为假阴性 F_N，反之称为假阳性 F_P。根据预测模型正确分类的比例即准确率 A_{cc} 来评估模型预测能力，准确率计算公式为：$A_{cc}=(T_N+T_P)/(T_N+T_P+F_N+F_P)$。表 9 显示了 3 种核函数预测模型的准确率。可以看出，线性核函数 LS-SVM 预测模型准确率最低，预测能力最差；径向基核函数预测模型准确率最高，预测能力最好。

当样本分布不均匀时，准确率由占比大的类别决定，可能会出现较大偏差，为此进一步采用真阳性率 R_{TP} 和假阳性率 R_{FP} 对预测模型进行评估，其中，真阳性率代表 CFD 结果为定常的样本中，预测模型能够正确预测的比例；假阳性率表示 CFD 结果为非定常的样本中，预测模型错误预测为定常的比例。2 种指标计算公式如下：$R_{TP}=T_P/(T_P+F_N)$，$R_{FP}=F_P(F_P+T_N)$。从真阳性率和假阳性率的计算公式可

表 9 不同核函数预测结果的准确率、真阳性率和假阳性率

Table 9 The accuracy, TPR and FPR of different kernel function prediction results

核函数	准确率	真阳性率	假阳性率
线性核函数	0.375	0.222	0.429
多项式核函数	0.75	0.778	0.286
径向基核函数	1	1	0

以看出, 真阳性率越大、假阳性率越小, 模型的预测能力越好。表 9 也给出了 3 种核函数预测模型的真阳性率和假阳性率, 同样显示线性核函数 LS-SVM 预测模型预测能力最差, 径向基核函数预测模型预测能力最好。

4　结　论

本文采用 CFD 方法对 30°/55°的双楔展开数值模拟, 并选取典型结果进行流场分析; 基于 CFD 计算数据, 采用 LS-SVM 方法, 建立高超声速双楔流场定常/非定常流态预测模型并进行验证与评估。得到如下结论:

1) 对于定常流动, 当 Re 不变时, 随着 Ma_∞的增大, 分离激波的波足向下游移动, 再附激波的波足向上游移动, 分离区逐渐减小; 当 Ma_∞不变时, 随着 Re 的增大, 分离激波的波足向上游移动, 再附激波的波足向下游移动, 分离区逐渐增大。

2) 对于非定常流动, 随着时间的推移, 分离点向上游移动, 再附点向下游移动, 分离区总体呈增大的趋势, 但 120 μs 时的再附点相对于 60 μs 时向上游移动。

3) 使用双楔定常/非定常 CFD 计算数据对 LS-SVM 模型进行训练, 分别选取线性核函数、多项式核函数及径向基核函数, 得到 3 种高超声速双楔流场定常/非定常流态预测模型, 并对其进行验证和评估。选取准确率、真阳性率和假阳性率等参数对预测模型进行评估, 在不同的评价体系中, 结果显示出一致性, 即径向基核函数预测模型预测效果最好, 多项式核函数预测模型次之, 线性核函数预测模型预测能力最差。

参考文献（References）

[1] HANK J, MURPHY J, MUTZMAN R. The X-51A scramjet engine flight demonstration program[C]//Proceedings of the 15th AIAA International Space Planes and Hypersonic Systems and Technologies Conference. Reston: AIAA, 2008: 2540.

[2] SCHRIJER F F J, VAN OUDHEUSDEN B W, DIERKSHEIDE U, et al. Quantitative visualization of a hypersonic double-ramp flow using PIV and schlieren[C]//Proceedings of the 12th International Symposium on Flow Visualization. Göttingen: DLR, 2006.

[3] HASHIMOTO T. Experimental investigation of hypersonic flow induced separation over double wedges[J]. Journal of Thermal Science, 2009, 18(3): 220-225.

[4] SWANTEK A, AUSTIN J. Heat transfer on a double wedge geometry in hypervelocity air and nitrogen flows[C]//Proceedings of the 50th AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition. Reston: AIAA, 2012: 284.

[5] KOMIVES J R, NOMPELIS I, CANDLER G V. Numerical investigation of unsteady heat transfer on a double wedge geometry in hypervelocity flows[C]//Proceedings of the 44th AIAA Fluid Dynamics Conference. Reston: AIAA, 2014: 2354.

[6] DURNA A S, CELIK B. Time-periodic shock interaction mechanisms over double wedges at Mach 7[J]. Shock Waves, 2019, 29(3): 381-399.

[7] KUMAR G, DE A. Modes of unsteadiness in shock wave and separation region interaction in hypersonic flow over a double wedge geometry[J]. Physics of Fluids, 2021, 33(7): 076107.

[8] NINNI D , BONELLI F , COLONNA G , et al. Unsteady behavior and thermochemical non equilibrium effects in hypersonic double-wedge flows[J]. Acta Astronautica, 2022, 191: 178-192.

[9] 陈森林, 高正红, 朱新奇, 等. 非稳定动态过程非定常气动力建模[J]. 航空学报, 2020, 41(8): 123675.
CHEN S L, GAO Z H, ZHU X Q, et al. Unsteady aerodynamic modeling of unstable dynamic process[J]. Acta Aeronautica et Sinica, 2020, 41(8): 123675(in Chinese).

[10] 王泽, 王梓伊, 王旭, 等. 一种数据驱动的气动热预示模型[J]. 空气动力学学报, 2023, 41(5): 12-19.
WANG Z, WANG Z Y, WANG X, et al. A data-driven aeroheating prediction model[J]. Acta Aerodynamica Sinica, 2023, 41(5): 12-19 (in Chinese).

[11] 陈志强. 基于参数化降阶模型的高超声速气动弹性问题研究[D]. 南京: 南京航空航天大学, 2022.
CHENG Z Q. Hypersonic aeroelastic analysis based on parameterized reduced-order model[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2022(in Chinese).

[12] JAMESON A, YOON S. Lower-upper implicit schemes with multiple grids for the Euler equations[J]. AIAA Journal, 1987, 25(7): 929-935.

[13] JAMESON A. Time dependent calculations using multigrid, with applications to unsteady flows past airfoils and wings[C]//Proceedings of the 10th Computational Fluid Dynamics Conference. Reston: AIAA, 1991: 1596.

[14] LI Q, YAN P, HUANG X, et al. Improvements to enhance robustness of third-order scale-independent WENO-Z schemes[J]. Advances in Applied Mathematics and Mechanics, 2025, 17(2): 373-406.

[15] JIANG G S, SHU C W. Efficient implementation of weighted ENO schemes[J]. Journal of Computational Physics, 1996, 126(1): 202-228.

[16] LI Q, YAN P, HUANG X, et al. On developing piecewise rational mapping with fine regulation capability for WENO schemes[J]. Journal of Scientific Computing, 2021, 88(3): 75.

[17] BORGES R, CARMONA M, COSTA B, et al. An improved weighted essentially non-oscillatory scheme for hyperbolic conservation laws[J]. Journal of Computational Physics, 2008, 227(6): 3191-3211.

[18] HARVEY J K, HOLDEN M S, WADHAMS T P. Code validation study of laminar shock/boundary layer and shock/shock interactions in hypersonic flow part B: comparison with Navier-Stokes and DSMC solutions[C]//Proceedings of the 39th Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2001: 1031.

[19] MOSS J N, BIRD G A. Direct simulation Monte Carlo simulations of hypersonic flows with shock interactions[J]. AIAA Journal, 2005, 43(12): 2565-2573.

[20] GREENE F A. Application of the multigrid solution solution technique to hypersonic entry vehicles[J]. Journal of Spacecraft and

Rockets, 1994, 31(5): 744-750.

[21] VATANSEVER D, CELIK B. Unsteady shock interaction mechanisms of high enthalpy reacting flows over double wedges at Mach 7[J]. Physics of Fluids, 2021, 33(5): 056110.

[22] DURNA A S, CELIK B. Effects of double-wedge aft angle on hypersonic laminar flows[J]. AIAA Journal, 2020, 58(4): 1689-1703.

[23] DURNA A S, EL HAJJ ALI BARADA M, CELIK B. Shock interaction mechanisms on a double wedge at Mach 7[J]. Physics of Fluids, 2016, 28(9): 096101.

[24] KUMAR G, DE A. Role of corner flow separation in unsteady dynamics of hypersonic flow over a double wedge geometry[J]. Physics of Fluids, 2021, 33(3): 036109.

[25] KNIGHT D, CHAZOT O, AUSTIN J, et al. Assessment of predictive capabilities for aerodynamic heating in hypersonic flow[J]. Progress in Aerospace Sciences, 2017, 90: 39-53.

[26] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.

[27] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300.

Numerical analysis and flow state prediction of double wedge steady/unsteady flow at different in flow

ZHANG Xinze, LI Qin*, WENG Yihui, YOU Yancheng

(School of Aerospace Engineering, Xiamen University, Xiamen 361102, China)

Abstract: This paper focuses on the laminar interaction of shock waves and the phenomenon of steady/unsteady flow in the hypersonic double wedge flow. The different working conditions (Ma_∞ and Re) were selected for numerical simulation and flow characteristics research with the developed third-order WENO scheme; The variation patterns of shock wave structure and separation zone in steady/unsteady flow fields caused by parameter changes were summarized. Based on the least squares support vector machine (LS-SVM) model, we developed a steady/unsteady flow states prediction model for the double wedge flow using various kernel functions (linear, polynomial, and radial basis functions) in order to forecast the steady/unsteady flow states of the double wedge flow at various Ma_∞ and Re. Three prediction models were validated and evaluated on the validation set. With 16 working circumstances, the results demonstrate that the radial basis function prediction model is capable of successfully predicting every double wedge flow state. The accuracy and true positive rate of the prediction results are 1, and the false positive rate is 0. The prediction ability is the strongest and has better generalization ability. The polynomial kernel prediction model takes second place in prediction ability, and the linear kernel prediction model has the worst prediction ability.

Keywords: hypersonic; double wedge; steady/unsteady flow; flow state prediction; least squares support vector machine

Received: 2023-11-20; **Accepted**: 2024-03-22; **Published Online**: 2024-04-17 16:32

URL: link.cnki.net/urlid/11.2625.V.20240416.1058.002

* **Corresponding author**. E-mail: qin-li@vip.tom.com

《北京航空航天大学学报》征稿简则

《北京航空航天大学学报》是北京航空航天大学主办的以航空航天科学技术为特色的综合性自然科学学术期刊(月刊)。本刊以反映航空航天领域研究成果与动态、促进学术交流、培养科技人才和推动科技成果向社会生产力转化为办刊宗旨。本刊为中国自然科学技术核心期刊,并被 Ei Compendex 等国内外权威文献检索系统收录。本刊向国内外公开发行,为进一步提高办刊质量和所刊出文章的学术水平,特制定本简则。

1 论文作者及内容

1.1 本刊面向海内外所有学者。

1.2 主要刊载与航空航天科学技术有关的材料科学及工程、飞行器设计与制造、宇航科学与工程、信息与电子技术、控制技术和自动化工程、流体力学和动力工程、计算机科学及应用技术、可靠性工程与失效分析等领域的研究文章。航空航天科学技术民用方面以及具有航空航天工程背景的应用数学、应用物理、应用力学和工程管理等方面的文章也在本刊优先考虑之列。

2 来稿要求

2.1 论文应具有创新性、科学性、学术性和可读性。

2.2 论文为原创作品,尚未公开发表过,并且不涉及泄密问题。若发生侵权或泄密问题,一切责任由作者承担。

2.3 主题明确,数据可靠,图表清晰,逻辑严谨,文字精练,公式符号标准规范,标点符号正确。

2.4 文稿撰写顺序:中文题名(一般不超过 20 个汉字),作者中文姓名、单位、所在城市、邮政编码(作者顺序始终不得更改),中文摘要(包括目的、方法、结果及结论),中文关键词(5 ~ 8 个),中图分类号,引言,正文,参考文献,英文题名,作者英文姓名、单位、所在城市、邮政编码、国别,英文摘要,英文关键词。首、末页下角注明基金项目名称及编号,作者信息。

2.5 作者请登录本刊网页进行在线投稿。

3 稿件的审核、录用与版权

3.1 来稿须经专家两审和主编、编委讨论后决定刊用与否。

3.2 若来稿经审查后认定不宜在本刊发表,将及时告知作者。如果在投稿满 3 个月后仍未收到本刊任何通知,作者有权改投其他刊。在此之前,请勿一稿多投,否则一切后果自负。

3.3 来稿一经作者签字并在本刊刊出,即表明所有作者都已经认可其版权转至本刊编辑部。本刊在与国内外文献数据库或检索系统进行交流及合作时,不再征询作者意见。

邮寄地址:100191 北京市海淀区学院路 37 号 《北京航空航天大学学报》编辑部
办公地点:北京航空航天大学办公楼西配楼412~416房间
电　　话:(010)82316698,82316981,82315499,82317448
E-mail: jbuaa@ buaa. edu. cn
http://bhxb. buaa. edu. cn
http://www. buaa. edu. cn

集群成员期刊

《北京航空航天大学学报》
第五届编辑委员会

主　任（主　编）：赵沁平

（以下按姓氏笔画为序）

副主任（副主编）：丁希仑　王少萍　孙志梅　李秋实　李焕喜　杨嘉陵
苗俊刚　相　艳　徐立军　钱德沛　曹晋滨

编　委：马殿富　王　琪　王　聪　邓小燕　王青云　王荣明　刘　宇
刘　红　江　洁　刘　强　闫　鹏　朱天乐　刘铁钢　齐铂金
陈万春　邹正平　苏东林　杨世春　沈成平　邱志平　宋知人
杨树斌　张晓林　杨晓奕　杨继萍　李惠峰　吴新开　张瑞丰
杨照华　宋凝芳　周　锐　林宇震　林贵平　战　强　姚仰平
胡庆雷　赵秋红　段海滨　赵巍胜　席　平　郭　宏　徐　洁
徐世杰　郭洪波　康　锐　翟　锦　熊华钢

编辑部主任：张　嵘

编　　辑：孙　芳　万玉敏　王　茜　卞欢欢　李亚泰

北京航空航天大学学报

Beijing Hangkong Hangtian Daxue Xuebao

（原《北京航空学院学报》）

（月刊　1956年创刊）

第52卷　第1期　2026年1月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.52 No.1　January 2026

主管单位　中华人民共和国工业和信息化部
主办单位　北京航空航天大学
主　　编　赵沁平
编辑出版　《北京航空航天大学学报》编辑部
邮　　编　100191
地　　址　北京市海淀区学院路37号
印　　刷　北京科信印刷有限公司
发　　行　《北京航空航天大学学报》编辑部
发行范围　国内外发行
联系电话　(010) 82316698 82316981
82315499 82317448
电子信箱　jbuaa@buaa.edu.cn

Administrated by　Ministry of Industry and Information Technology of the People's Republic of China
Sponsored by　Beijing University of Aeronautics and Astronautics (BUAA)
(Beijing 100191, P. R. China)
Chief Editor　ZHAO Qinping
Edited and Published by　Editorial Board of JBUAA
Printed by　Beijing Kexin Printing Co., Ltd.
Distributed by　Editorial Board of JBUAA
Telephone　(010) 82316698 82316981
82315499 82317448
E-mail　jbuaa@buaa.edu.cn
http://bhxb.buaa.edu.cn

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V

ISSN 1001-5965

国内定价：50.00元／期

微信公众号　超星电子书