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

电力电子变换器拓扑的智能化构建与自动推荐技术专栏



两开关隔离型直流变换器拓扑自生成方法

李　虹　梁一迪　尹成东　郑琼林　张　波

5　2025
Vol.51

北 京 航 空 航 天 大 学 学 报

第 51 卷　第 5 期　（总第 387 期）　2025 年 5 月

目　　次

电力电子变换器拓扑的智能化构建与自动推荐技术专栏

两开关隔离型直流变换器拓扑自生成方法 …… 李虹，梁一迪，尹成东，郑琼林，张波（1415）

非对称电网状态下 MMC-HVDC 改进滑模直接功率控制策略 …… 关天一，王怡博，王睿，赵海乔，郑文婧，孙秋野（1428）

电力电子变换器拓扑同构与拓扑等效辨析 …… 莫李平，陈桂鹏，李志强（1440）

基于图论的变换器冗余开关自动识别方法 …… 莫李平，宋吉祥，陈桂鹏（1449）

基于差分进化算法的 SIMO Buck 变换器最优动态响应搜索 …… 李丹，崔文峰，陈桂鹏（1457）

基于层次分析-熵值法的 DC-DC 变换器拓扑综合评价 …… 张洋，丘东元，张波，陈艳峰（1469）

基于强化学习的单电感多端口变换器调制策略设计方法 …… 白敬波，陈宇，谢诗语，代新维（1480）

论文

锂离子电池故障预测及健康管理技术进展 …… 杨世春，王潇，陈飞，郑一帆，钟益林，周思达（1490）

基于多变量机器学习的 CYGNSS 有效波高反演模型 …… 张云，肖盛，姜丽菲，孟婉婷，杨树瑚，韩彦岭（1503）

嵌入自注意力机制的 U 型高分遥感影像语义分割网络 …… 杨军，张金影（1514）

基于图神经网络的航班地面保障数据插补算法 …… 邢志伟，孙恪，罗谦，刘畅，张涛，乔迪（1528）

高精度动态温度控制系统设计与验证 …… 韩潇，周盈，黄海，邵静怡（1539）

振动对锂离子电池循环及热失控特性的影响 …… 张青松，李东琪，杨娟（1548）

基于级联的航班地面保障动态预测 …… 唐小卫，丁叶，吴政隆，张生润，吴佳琦，叶梦凡（1557）

基于特征融合和分块注意力的无人机跟踪算法 …… 刘芳，杨雨妍，王鑫（1566）

基于时效网络的关键航空器识别方法 …… 王红勇，马丽书，许平（1579）

框架式动量交换装置角动量包络分析方法 …… 王卫杰，汪洲，庞维坤，杨洋（1591）

基于多毁伤等级标准的导弹毁伤效能贝叶斯鉴定试验设计 …… 陈童，刘昊邦，胡涛（1599）

融合全尺度特征与轨迹修正的遮挡车辆跟踪 …… 郭俊锋，张子华（1608）

NNW-FlowStar 软件模拟并联分离特性 …… 张培红，周桂宇，沈盈盈，唐静，赵炜，贾洪印（1620）

基于 Gossip 协议的高效集群数据同步方案 …… 张宏海，崔斌豪，李一鸣，田丰，贾永强，肖傲三（1629）

分布式多旋翼/倾转机翼气弹耦合动特性研究 …… 程毅，赵金瑞，黄水林，余智豪，邓旭东（1637）

低速风洞动压场校核中设备干扰及修正方法 …… 刘江涛，龚小权，周乃春，李明，张耀冰（1651）

考虑材料优选的高速飞行器多层隔热结构优化 …… 郝东，刘建霞，苏杰，贺元元（1662）

双臂空间机器人球形外包络抓捕控制 …… 夏新会，贾英宏，张军（1673）

贮箱自动无级增压可调文氏管的设计与试验 …… 任嘉琳，郭红杰，李文鑫，梁国柱，赵胜（1684）
BDS-3 星载原子钟在轨运行特征分析 …… 王文斌，唐成盼，杨建华，胡小工，曹月玲，周善石（1694）
基于轻量胶囊网络的自监督图像变化检测方法 …… 张益天，罗喜伶，王宇鹏（1705）
基于 COOT 算法的 VMD-HPCA-GRU 超短期风电功率预测 ……
…… 何星月，杨靖，朱兆强，杨斌，覃涛（1716）
板簧橡胶支座刚度与疲劳寿命波动性仿真分析及试验研究 ……
…… 姬娜，刘娟，王浩然，高蕊，卢咏来，李凡珠（1726）
可变拍幅扑翼飞行器模型设计与气动特性分析 …… 梁金泽，潘天宇，郑孟宗，彭连松，曹梦达（1735）
高超变体飞行器再入轨迹罚函数序列凸规划 …… 王仰杰，龙腾，李俊志，徐广通，孙景亮（1747）
重复使用火箭垂直着陆段在线轨迹优化方法 …… 施棋，齐瑞云，佘宇琛，胡存明（1760）
双径向洛伦兹力磁轴承动力学及解耦控制方法 ……
…… 李宗育，王卫杰，王丽芬，任元，樊亚洪，李磊（1770）
AP 虚拟位置感知下的双维度分区定位算法 …… 张阳，秦宁宁（1781）

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 378 * zh * P * ¥ 50.00 * 400 * 35 * 2025-05

（**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. 51 No. 5 （Sum 387） May 2025

CONTENTS

Special Column on Intelligent Construction and Automatic Recommendation Technology of Power Electronic Converter Topologies

Topology self-generating method of isolated DC-DC converters with two switches
…… LI Hong，LIANG Yidi，YIN Chengdong，ZHENG Qionglin，ZHANG Bo （1415）

Improved sliding-mode direct power control strategy for MMC-HVDC under asymmetrical grid state
…… GUAN Tianyi，WANG Yibo，WANG Rui，ZHAO Haiqiao，ZHENG Wenjing，SUN Qiuye （1428）

Discrimination on isomorphism and equivalence of topologies of power electronics converters
…… MO Liping，CHEN Guipeng，LI Zhiqiang （1440）

Auto identification method for redundant switches in converters based on graph theory
…… MO Liping，SONG Jixiang，CHEN Guipeng （1449）

Optimal dynamic response exploration for SIMO Buck converter based on differential evolution algorithm
…… LI Dan，CUI Wenfeng，CHEN Guipeng （1457）

Comprehensive evaluation of DC-DC converters based on analytic hierarchy process and entropy method
…… ZHANG Yang，QIU Dongyuan，ZHANG Bo，CHEN Yanfeng （1469）

Design method for modulation strategy of a single-inductor multi-port converter based on reinforcement learning
…… BAI Jingbo，CHEN Yu，XIE Shiyu，DAI Xinwei （1480）

Thesis Section

Advances in prognosis and health management technologies of lithium-ion batteries
…… YANG Shichun，WANG Xiao，CHEN Fei，ZHENG Yifan，ZHONG Yilin，ZHOU Sida （1490）

Significant wave height retrieval model of CYGNSS based on multivariate machine learning
…… ZHANG Yun，XIAO Sheng，JIANG Lifei，MENG Wanting，YANG Shuhu，HAN Yanling （1503）

U-shaped semantic segmentation network of high-resolution remote sensing images embedded with self-attention mechanism
…… YANG Jun，ZHANG Jinying （1514）

Imputation algorithm for flight ground support data based on graph neural network
…… XING Zhiwei，SUN Ke，LUO Qian，LIU Chang，ZHANG Tao，QIAO Di （1528）

Design and verification of high-precision dynamic temperature control system
…… HAN Xiao，ZHOU Ying，HUANG Hai，SHAO Jingyi （1539）

Influence of vibration on cyclic and thermal runaway characteristics of lithium ion batteries
…… ZHANG Qingsong，LI Dongqi，YANG Juan （1548）

Dynamic prediction of flight ground service based on cascade
…… TANG Xiaowei，DING Ye，WU Zhenglong，ZHANG Shengrun，WU Jiaqi，YE Mengfan （1557）

UAV tracking algorithm based on feature fusion and block attention
…… LIU Fang，YANG Yuyan，WANG Xin （1566）

Critical aircraft identification method based on temporal network
…… WANG Hongyong，MA Lishu，XU Ping （1579）

Angular momentum envelope analysis method of gimbal-type momentum exchange device
…… WANG Weijie，WANG Zhou，PANG Weikun，YANG Yang （1591）

Bayesian identification test design of missile damage effectiveness based on multiple damage grades standards
…… CHEN Tong，LIU Haobang，HU Tao （1599）

Track obscured vehicles by fusing full-scale features with trajectory correction
…… GUO Junfeng，ZHANG Zihua （1608）

Simulation of parallel separation characteristics using NNW-FlowStar software
…… ZHANG Peihong，ZHOU Guiyu，SHEN Yingying，TANG Jing，ZHAO Wei，JIA Hongyin （1620）

An efficient cluster data synchronization scheme based on the Gossip protocol
…… ZHANG Honghai，CUI Binhao，LI Yiming，TIAN Feng，JIA Yongqiang，XIAO Aosan （1629）

Research on dynamic characteristics of distributed multi-rotor/tilting wing aeroelastic coupling
…… CHENG Yi，ZHAO Jinrui，HUANG Shuilin，YU Zhihao，DENG Xudong （1637）

Equipment interference and correction method in dynamic pressure field verification of low-speed wind tunnel
…… LIU Jiangtao，GONG Xiaoquan，ZHOU Naichun，LI Ming，ZHANG Yaobing （1651）

Optimization of multilayer thermal insulation structure for high-speed aircraft considering material optimization
…… HAO Dong，LIU Jianxia，SU Jie，HE Yuanyuan （1662）

Spherical envelope capture control for dual-arm space robots
…… XIA Xinhui，JIA Yinghong，ZHANG Jun （1673）

Design and experiment of adjustable Venturi tube for stepless automatic tank pressurization
…… REN Jialin，GUO Hongjie，LI Wenxin，LIANG Guozhu，ZHAO Sheng （1684）

In-orbit operation characteristics of BDS-3 spaceborne atomic clocks
…… WANG Wenbin，TANG Chengpan，YANG Jianhua，HU Xiaogong，CAO Yueling，ZHOU Shanshi （1694）

Self-supervised image change detection method based on lightweight capsule network
…… ZHANG Yitian，LUO Xiling，WANG Yupeng （1705）

VMD-HPCA-GRU ultra-short-term wind power prediction based on COOT algorithm
…… HE Xingyue，YANG Jing，ZHU Zhaoqiang，YANG Bin，QIN Tao （1716）

Simulation analysis and experimental study on stiffness and fatigue life fluctuation of leaf spring rubber bearings
…… JI Na，LIU Juan，WANG Haoran，GAO Rui，LU Yonglai，LI Fanzhu （1726）

Model design and aerodynamic characteristic analysis of variable-amplitude flapping wing aircraft
…… LIANG Jinze，PAN Tianyu，ZHENG Mengzong，PENG Liansong，CAO Mengda （1735）

Re-entry trajectory planning for hypersonic morphing vehicles using penalty sequence convex programming
…… WANG Yangjie，LONG Teng，LI Junzhi，XU Guangtong，SUN Jingliang （1747）

Online trajectory optimization method for vertical landing phase of reusable launch vehicle
…… SHI Qi，QI Ruiyun，SHE Yuchen，HU Cunming （1760）

Dynamics and decoupling control method of double radial Lorentz force magnetic bearings
…… LI Zongyu，WANG Weijie，WANG Lifen，REN Yuan，FAN Yahong，LI Lei （1770）

Dual-dimensional partition localization algorithm under AP virtual location perception
…… ZHANG Yang，QIN Ningning （1781）

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0483

两开关隔离型直流变换器拓扑自生成方法

李虹[1,2,*]，梁一迪[2]，尹成东[2]，郑琼林[2]，张波[3]

(1. 浙江大学 电气工程学院，杭州 310058；2. 北京交通大学 电气工程学院，北京 100044；
3. 华南理工大学 电力学院，广州 510006)

摘　　要：随着新能源发电、电气化交通等领域的兴起，电能变换的需求日益多样化，需要更多的隔离型直流变换器(即隔离型DC-DC变换器)拓扑。为解决基于人工经验生成拓扑的偶然性和无法进行拓扑优化的问题，提出一种基于顶点素数度的两开关隔离型DC-DC变换器拓扑自生成方法。提出双顶点对变压器模型，利用图论将电路转换为图；将构成电路的元器件的数量和种类作为输入条件，利用所提的拓扑自生成方法遍历生成具有给定元器件的所有可用隔离型DC-DC变换器拓扑。以隔离型$T^1S^2D^2L^1C^1$ DC-DC变换器为例，即变换器电路由1个变压器、2个开关管、2个二极管、1个电感、1个电容、1个输入直流源和1个输出直流源组成，可一次生成87种可用正激变换器和97种反激变换器拓扑，以正激变换器为例，根据电压增益可分为25种升压正激变换器、24种降压正激变换器、32种升降压正激变换器和6种等压正激变换器。所提方法实现了基于计算机编程快速、量化生成新的隔离型DC-DC变换器拓扑，为工程应用提供了拓扑库，为拓扑优化选型提供了条件。

关　键　词：隔离型DC-DC变换器；两开关；拓扑；自生成；顶点素数度
中图分类号：TM461；O157.5
文献标志码：A　　**文章编号**：1001-5965(2025)05-1415-13

由于隔离型DC-DC变换器(如正激变换器)的简单性和高可靠性[1-2]，其被广泛应用于多种功率转换应用场合中，如可再生能源系统、消费电子设备系统等[3-5]。然而，不同于非隔离型DC-DC变换器，可工业应用的隔离型DC-DC变换器的拓扑种类较少，正激变换器仍是低功耗DC-DC变换的行业主力，如经典的正激变换器[6]、带电阻电容二极管(resistor-capacitor-diode, RCD)缓冲器的正激变换器[7]、带电感电容和2个二极管(inductor, capacitor with two diodes, LCDD)缓冲器的正激变换器[8]、具有谐振复位的正激变换器[9]、带有两开关的正激变换器[10]、采用有源钳位复位的正激变换器[11]等。随着新能源发电、电气化交通等领域的兴起，电能变换的需求日益多样化，需要更多的隔离型DC-DC变换器拓扑[12-14]。然而，现有拓扑都是基于人工经验提出，存在偶然性且拓扑数量较少，难以进行拓扑优化选型。为克服基于人工经验生成拓扑的偶然性和无法进行拓扑优化的问题，近年来提出了一些拓扑自动搜索算法，这些算法利用图论将变换器电路转换成图的形式，并借助计算机辅助生成和筛选，快速推导出大量拓扑[15-16]。然而，现有拓扑自动搜索算法输入条件中的器件类型只有开关管、二极管、电感、电容、输入直流源和输出直流源模型，没有变压器模型，因此，现有拓扑自动搜索算法的适用范围仅限于非隔离型DC-DC变换器。

为此，本文通过提出一种双顶点对的变压器模型，进而提出了一种隔离型DC-DC变换器拓扑自生成方法，实现了基于计算机编程快速、量化生成

收稿日期：2023-07-27；**录用日期**：2023-10-06；**网络出版时间**：2023-12-18 14:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20231218.0849.001
基金项目：国家杰出青年科学基金(52325704)；国家自然科学基金重点项目(52237008)
*通信作者. E-mail：hli@bjtu.edu.cn

引用格式：李虹，梁一迪，尹成东，等. 两开关隔离型直流变换器拓扑自生成方法[J]. 北京航空航天大学学报，2025，51(5)：1415-1427.
LI H，LIANG Y D，YIN C D，et al. Topology self-generating method of isolated DC-DC converters with two switches [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1415-1427 (in Chinese).

新的隔离型 DC-DC 变换器拓扑，为工程应用提供了拓扑库，为拓扑优化选型提供了条件。

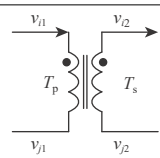
1 基于双顶点对变压器模型的隔离型 DC-DC 变换器

1.1 双顶点对变压器模型

为实现拓扑自生成，本文将隔离型 DC-DC 变换器拓扑结构表示为有向图，提出了变压器的双顶点对模型。如表 1 所示，变压器存在 2 个端口和 4 个端点，且变压器工作时，其一次侧和二次侧不在同一工作回路中。为便于在文献 [16] 中定义的器件矩阵中存储器件信息，本文将变压器 T 分为一次侧 T_p 和二次侧 T_s，其中，T_p 和 T_s 分别使用一对顶点对进行表示，因此，变压器可使用双顶点对进行表示。以表 1 中的变压器为例，T_p 使用顶点对 (v_{i1},v_{j1}) 表示，T_s 使用顶点对 (v_{j2},v_{i2}) 表示，T 可使用双顶点对 $(v_{i1},v_{j1},v_{j2},v_{i2})$ 表示。其中，一次侧电流从顶点 v_{i1} 流入，从顶点 v_{j1} 流出；二次侧电流从顶点 v_{j2} 流入，从顶点 v_{i2} 流出。

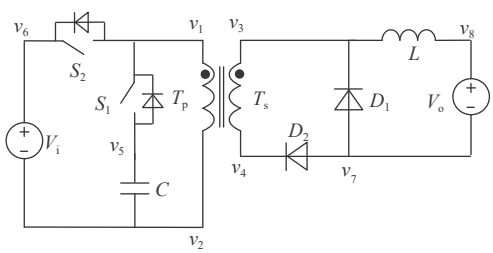
表 1 双顶点对变压器模型

Table 1 Two-vertex-pair transformer model

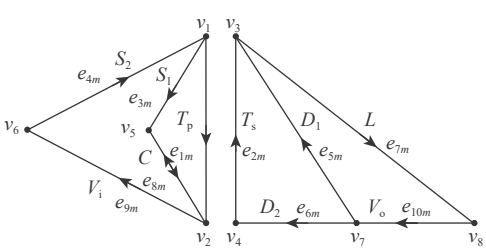
器件	双顶点对	模型
变压器	$T=(T_p,T_s)=(v_{i1},v_{j1},v_{j2},v_{i2})$	

1.2 隔离型 DC-DC 变换器拓扑图形化

以图 1(a) 所示的变换器拓扑为例，根据本文提出的双顶点对变压器模型和文献 [16] 中将变换器拓扑转换成有向图的方法，将拓扑结构转换为有向图，如图 1(b) 所示。将器件转换成带有箭头的边，流经器件的电流方向为箭头的方向。其中，所有器件都是理想器件，不考虑寄生参数[16]。



(a) 变换器拓扑结构



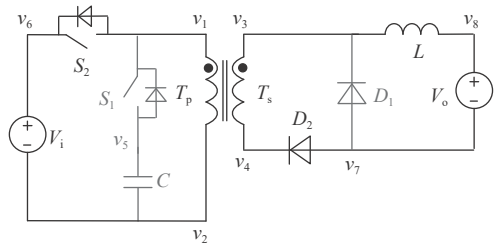
(b) 变换器有向图

图 1 变换器拓扑结构和图形化表示

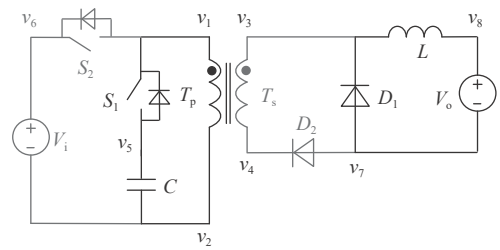
Fig. 1 Topology structure and graphical representation of converter

为便于计算机进行拓扑生成与筛选，可根据文献 [16] 和本文提出的双顶点对变压器模型将变换器拓扑转换成以下矩阵来存储图的信息，即顶点矩阵 $\boldsymbol{V}_m=[v_1,v_2,v_3,v_4,v_5,v_6,v_7,v_8]$、器件矩阵 $\boldsymbol{H}_m=[T_p,T_s,S_1,S_2,D_1,D_2,L,C,V_i,V_o]$、边矩阵 $\boldsymbol{E}_m=[e_{1m},e_{2m},e_{3m},e_{4m},e_{5m},e_{6m},e_{7m},e_{8m},e_{9m},e_{10m}]$ 和简化关联矩阵 $\boldsymbol{C}=[\boldsymbol{E}_1;\boldsymbol{E}_2;\cdots;\boldsymbol{E}_m;\cdots;\boldsymbol{E}_M]^{\mathrm{T}}$，$m=1,2,\cdots,M$。其中，简化关联矩阵的每个边矩阵都对应一个拓扑结构。以图 1(a) 所示变换器拓扑为例，其拓扑结构可以转换成边矩阵 $\boldsymbol{E}_m=[v_1,v_2,v_4,v_3,v_1,v_5,v_6,v_1,v_7,v_3,v_7,v_4,v_3,v_8,v_2,v_5,v_2,v_6,v_8,v_7]$。

由于变换器是通过开关器件的开通和关断进行工作模态切换的，可根据开关组合情况分成 2 个工作模态，如图 2 所示。



(a) 变换器工作模态1



(b) 变换器工作模态2

图 2 变换器工作模态

Fig. 2 Converter working modes

本文通过使用工作环路矩阵来表示变换器的不同工作模态[17]，其中，工作环路矩阵由多条工作环路组成，可将图 2 所示变换器的 2 个工作模态使用 2 个工作环路矩阵进行存储，如式 (1) 所示。在工作模态 1 中，开关管 S_1 不工作，为关断状态，开关管 S_2 工作，为开通状态。因此，开关管 S_2 存储在工作环路矩阵 $\boldsymbol{W}_1$ 中，开关管 S_1 不存储在工作环路矩阵 $\boldsymbol{W}_1$ 中。

$$\begin{cases} \boldsymbol{W}_1 = \begin{matrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \end{matrix} \begin{bmatrix} V_i & S_2 & T_p & 0 \\ T_s & L & V_o & D_2 \end{bmatrix} \\ \boldsymbol{W}_2 = \begin{matrix} \boldsymbol{w}_1 \\ \boldsymbol{w}_2 \end{matrix} \begin{bmatrix} C & S_1 & -T_p & 0 \\ L & V_o & D_1 & 0 \end{bmatrix} \end{cases} \tag{1}$$

根据简化关联矩阵, 给定器件的电流参考方向, 电流由器件的第 1 个顶点流向第 2 个顶点, 流经器件电流方向与参考方向一致, 使用器件本身进行表示, 否则, 使用添加负号的器件进行表示。

1.3　隔离型 DC-DC 变换器拓扑边构建规则

为实现隔离型 DC-DC 变换器拓扑自动生成, 本文使用文献 [16] 提出的顶点素数度法进行拓扑构建。首先, 使用顶点素数度法计算第 m 个图各顶点的顶点素数度[16]。然后, 根据得到的顶点素数度将顶点分为等权顶点和不等权顶点。其中, 等权顶点为具有相同顶点素数度的顶点, 同时, 与顶点相连的邻接顶点也具有相同的顶点素数度; 反之, 则为不等权顶点。最后, 通过边构建规则构建拓扑图。由于文献 [16] 的 3 条边构建规则只针对非隔离型变换器, 进行隔离型变换器拓扑构建时, 没有考虑隔离型变换器的电气隔离需求, 会同时生成隔离型和非隔离型变换器。因此, 本文沿用文献 [16] 中 3 条变换器边的构建规则, 同时新增 1 条变换器边构建规则, 使其只能生成隔离型变换器, 隔离型变换器的 4 条边构建规则如下:

1) 如果等权顶点中存在不少于 2 个顶点, 则从等权顶点中选择顶点编号较小的 2 个顶点构成 1 条边, 如图 3 中“边构建选择 1”所示。

2) 如果不等权顶点中存在不少于 2 个顶点, 则从不等权顶点中任意选择 2 个顶点构成 1 条边, 如图 3 中“边构建选择 2”所示。

3) 如果同时存在等权顶点和不等权顶点, 则从等权顶点中选择 1 个顶点编号最小的顶点和不等权顶点中的任意 1 个顶点构成 1 条边, 如图 3 中“边构建选择 3”所示。

4) 如果选取构成 1 条边的 1 个顶点为两端点器件顶点, 且该顶点与四端点器件一侧的顶点相连, 则选取的另 1 个顶点不允许与四端点器件另一侧的顶点相连, 如图 3 中“边构建选择 4”所示。

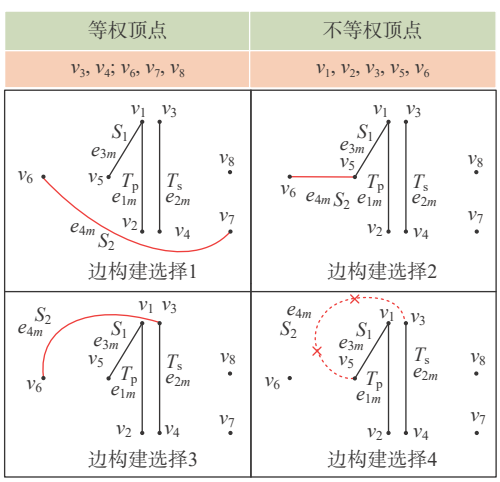


图 3　隔离型 DC-DC 变换器边构建选择

Fig. 3　Selection of edge construction for isolated DC-DC converters

2　隔离型 $T^1S^2D^2L^1C^1$ DC-DC 变换器拓扑自生成方法

本文提出一种隔离型 DC-DC 变换器拓扑自生成方法, 拓扑生成过程如图 4 所示, 分为 5 部分: 参数输入、拓扑图构建、图约束筛选、电气约束筛选和环路冗余拓扑筛选。



图 4　隔离型 DC-DC 变换器拓扑自生成方法

Fig. 4　The topology self-generation method of the isolated DC-DC converters

1) 参数输入。输入构成变换器拓扑的元器件的数量和种类, 生成顶点矩阵 $\boldsymbol{V}_m$ 和器件矩阵 $\boldsymbol{H}_m$。

2) 拓扑图构建。计算各顶点的顶点素数度, 进行顶点分类, 根据变换器的 4 条边构建规则进行拓扑图构建, 生成拓扑图的简化关联矩阵 $\boldsymbol{C}$。

3) 图约束筛选。通过 2 条图约束对拓扑图进行图特性筛选, 生成图特性正常拓扑图的简化关联矩阵 $\boldsymbol{C}_{\text{temp}}$。

4) 电气约束筛选。通过 7 条电气约束对拓扑图进行电气特性筛选, 生成可正常工作拓扑图的简化关联矩阵 $\boldsymbol{C}_{\text{ADGs}}$。

5) 环路冗余拓扑筛选。对可正常工作拓扑图进行环路冗余拓扑筛选, 输出可正常工作无冗余拓扑图的简化关联矩阵 $\boldsymbol{C}_{\text{re}}$。

2.1 参数输入

输入构成变换器拓扑的元器件的数量和种类，以隔离型 $T^1S^2D^2L^1C^1$ DC-DC 变换器为例，该类变换器电路由 1 个变压器、2 个开关管、2 个二极管、1 个电感、1 个电容、1 个输入直流源和 1 个输出直流源组成，因此，输入参数为 T、S_1、S_2、D_1、D_2、L、C、V_i、V_o。根据文献 [16] 和本文所提变压器模型，可以得到顶点矩阵 $\boldsymbol{V}_m=[v_1,v_2,v_3,v_4,v_5,v_6,v_7,v_8]$ 和器件矩阵 $\boldsymbol{H}_m=[T_p,T_s,S_1,S_2,D_1,D_2,L,C,V_i,V_o]$。

2.2 拓扑图构建

本文根据 4 条边构建规则，使用顶点素数度法进行拓扑图构建，以在第 m 个已构建完 3 条边的图的基础上构建第 4 条边 e_{4m} 为例，其构建过程如图 5 所示。

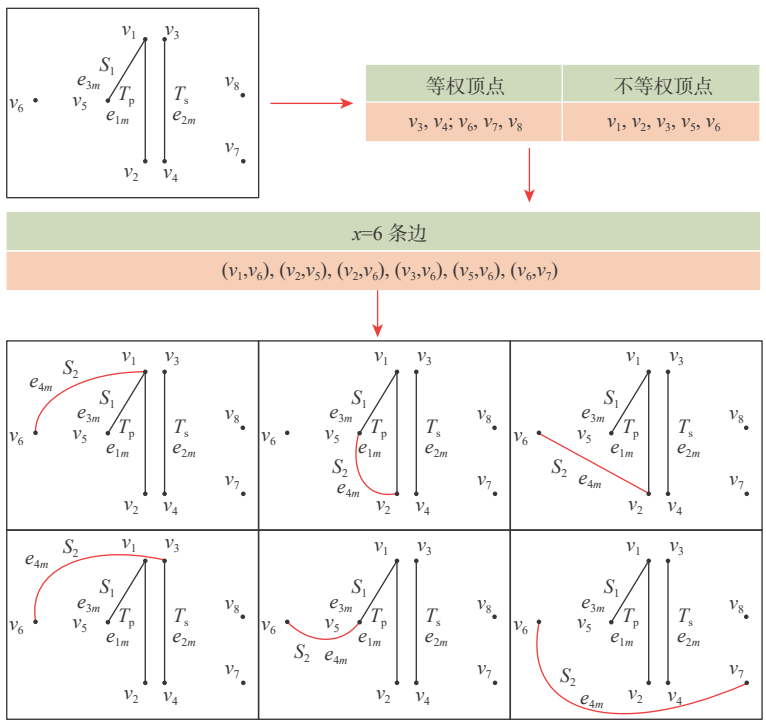


图 5 构建第 4 条边 e_{4m} 的过程

Fig. 5 Process of building the fourth edge e_{4m}

首先，计算第 m 个图的各个顶点的顶点素数度，得到顶点素数度矩阵 $\boldsymbol{P}_m$ = [105,10,10,10,14,0,0,0]。然后，根据顶点素数度矩阵 $\boldsymbol{P}_m$ 对顶点进行分类，将其分为等权顶点和不等权顶点，如图 5 所示，该图的等权顶点为 v_3, v_4; v_6, v_7, v_8, 不等权顶点为 v_1, v_2, v_3, v_5, v_6。最后，根据 4 条边构建规则，生成 x=6 条边，即边 (v_1,v_6), (v_2,v_5), (v_2,v_6), (v_3,v_6), (v_5,v_6), (v_6,v_7), 可依次加入第 m～m+5 个图中，同时将其边对应的顶点存入矩阵 $\boldsymbol{C}$ 中，如式 (2) 所示。

$$\boldsymbol{C}=\begin{array}{c}\\ \vdots\\ \boldsymbol{E}_m\\ \boldsymbol{E}_{m+1}\\ \boldsymbol{E}_{m+2}\\ \boldsymbol{E}_{m+3}\\ \boldsymbol{E}_{m+4}\\ \boldsymbol{E}_{m+5}\\ \vdots\end{array}\begin{array}{c}\begin{array}{cccccccccccccccccccc} T_s/e_{1m} & & T_p/e_{2m} & & S_1/e_{3m} & & S_2/e_{4m} & & D_1/e_{5m} & & D_2/e_{6m} & & L/e_{7m} & & C/e_{8m} & & V_i/e_{9m} & & V_o/e_{10m} & \end{array}\\ \begin{bmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_1 & v_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_2 & v_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_2 & v_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_3 & v_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_5 & v_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_6 & v_7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}\end{array} \tag{2}$$

当器件矩阵 $\boldsymbol{H}_m$ 中所有器件对应的边均完成构建，最终得到 501 647 种拓扑图，其简化关联矩阵如式 (3) 所示。直接采用文献 [16] 的拓扑边构建规则进行拓扑图构建，将得到 2 047 254 种拓扑图，包含较多无用拓扑图，增加后续拓扑搜索时间。因此，本文新增 1 条拓扑构建规则进行拓扑图构建，能够降低后续拓扑搜索时间。

$$\boldsymbol{C}=\begin{array}{c} \\ \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \vdots \\ \boldsymbol{E}_{501\,647}\end{array}\begin{array}{c}\begin{array}{cccccccccccccccccc} T_s/e_{1m} & & T_p/e_{2m} & & S_1/e_{3m} & & S_2/e_{4m} & & D_1/e_{5m} & & D_2/e_{6m} & & L/e_{7m} & & C/e_{8m} & & V_i/e_{9m} & & V_o/e_{10m} & \end{array}\\ \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_1 & v_6 & v_1 & v_7 & v_1 & v_8 & v_2 & v_5 & v_2 & v_6 & v_2 & v_7 & v_2 & v_8 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_1 & v_6 & v_1 & v_7 & v_1 & v_8 & v_2 & v_5 & v_2 & v_6 & v_2 & v_7 & v_5 & v_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_5 & v_7 & v_6 & v_8 & v_5 & v_8 & v_1 & v_6 & v_6 & v_7 & v_2 & v_8 \end{bmatrix}\end{array} \tag{3}$$

2.3　图约束筛选

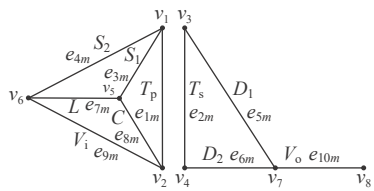
构建出的拓扑图并非全部可用, 需要根据电路基本工作原理进行约束条件筛选, 得到可用变换器拓扑。

首先, 进行图约束筛选。本文对文献 [16] 中变换器的 2 条图约束进行沿用、修改。

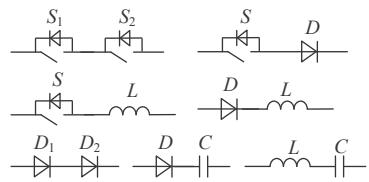
1) 图约束 1。图中不能包含任何只与一条边连接的顶点, 即图不能存在顶点度为 1 的顶点, 如图 6(a) 所示。

2) 图约束 2。图中不能包含 S、D、C、L 任意 2 个器件直接串联连接, 即顶点度为 2 的顶点连接的 2 个器件不能为 S、D、C、L 任意 2 个器件, 如图 6(b) 所示。

由于图约束 1 和图约束 2 只需考虑拓扑的顶点和边之间的连接关系, 不需要通过工作环路矩阵中的环路和器件进行约束判断, 与工作模态无关, 均适用于正激变换器和反激变换器。通过图约束筛选, 从构建出的 501 647 种拓扑图中得到 3 017 种图特性正常拓扑图, 其简化关联矩阵如下:



(a) 存在只与一条边连接的顶点



(b) 器件无效串联结构

图 6　无效图

Fig. 6　Invalid diagram

$$\boldsymbol{C}_{\text{temp}}=\begin{array}{c} \\ \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \vdots \\ \boldsymbol{E}_{3\,017}\end{array}\begin{array}{c}\begin{array}{cccccccccccccccccc} T_s/e_{1m} & & T_p/e_{2m} & & S_1/e_{3m} & & S_2/e_{4m} & & D_1/e_{5m} & & D_2/e_{6m} & & L/e_{7m} & & C/e_{8m} & & V_i/e_{9m} & & V_o/e_{10m} & \end{array}\\ \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_1 & v_6 & v_1 & v_7 & v_2 & v_5 & v_3 & v_8 & v_2 & v_6 & v_5 & v_7 & v_4 & v_8 \\ v_1 & v_2 & v_3 & v_4 & v_1 & v_5 & v_1 & v_6 & v_1 & v_7 & v_2 & v_5 & v_3 & v_8 & v_5 & v_6 & v_2 & v_7 & v_4 & v_8 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 & v_1 & v_5 & v_3 & v_7 & v_2 & v_5 & v_4 & v_8 & v_2 & v_6 & v_4 & v_7 \end{bmatrix}\end{array} \tag{4}$$

2.4　电气约束筛选

式 (4) 中的 3 017 种图特性正常拓扑图并非全部可用, 还需考虑变换器工作时电气特性是否正常。

本文对文献 [16] 中的 3 条变换器电气约束进行沿用、修改, 同时新增 4 条电气约束, 得到 7 条电气约束。

1) 电气约束 1。不允许任何短路环路和 2 个或 2 个以上电压源并联在一个环路中。工作环路矩阵中不能出现图 7 所示 7 种环路, 即 $\boldsymbol{w}\notin(\boldsymbol{W}_1\cup\boldsymbol{W}_2)$。

对于隔离型 $T^1S^2D^2L^1C^1$ DC-DC 变换器, 存在 3 个未知变量, 即 L、C、V_o, 需要至少 3 个独立方程进行求解, 且所有器件正常工作。因此, 对文献 [16] 中的第 2 条电气约束条件进行修改。

2) 电气约束 2。环路矩阵中至少含有 3 条环路, 且所有器件出现在环路矩阵中, 即 $\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 至少存在 3 条环路且 $\boldsymbol{H}_m\in(\boldsymbol{W}_1\cup\boldsymbol{W}_2)$。

如图 8(a)～图 8(d) 所示, 变换器 1 和变换器 2 存在开关管、电容直连情况, 为有源钳位变换器, 能够实现变压器磁复位。对于该变换器, 电容只出

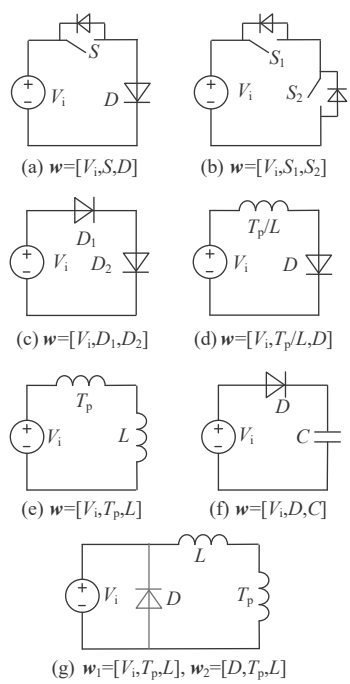
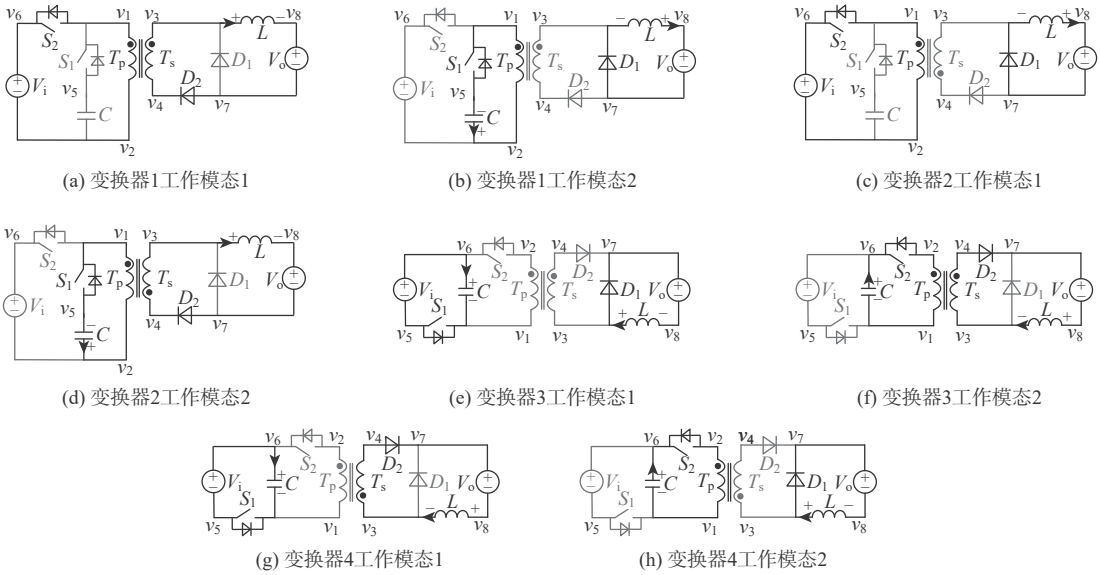


图 7　无效环路

Fig. 7　Invalid loops



(a) 变换器1工作模态1 (b) 变换器1工作模态2 (c) 变换器2工作模态1

(d) 变换器2工作模态2 (e) 变换器3工作模态1 (f) 变换器3工作模态2

(g) 变换器4工作模态1 (h) 变换器4工作模态2

图 8 2 个工作模态下的隔离型 $T^1S^2D^2L^1C^1$ DC-DC 变换器

Fig. 8 Isolated $T^1S^2D^2L^1C^1$ DC-DC converters with two working modes

现在 1 个工作模态中。如图 8(e)～图 8(h) 所示，变换器 3 和变换器 4 不存在开关管、电容直连情况，为普通变换器，对于该变换器，电容出现在 2 个工作模态中，且流经器件的电流方向不一致。因此，本文在添加变压器后，对文献 [16] 中的第 3 条电气约束条件进行修改。

3）电气约束 3。电感出现在 2 个工作环路矩阵中，且符号一致，即 $L\in\boldsymbol{W}_1$ 且 $L\in\boldsymbol{W}_2$。若存在开关管、电容直连情况，且与变压器一次侧构成环路，电容可只出现在 1 个工作环路矩阵中，即 $C\in\boldsymbol{W}_1$ 或 $C\in\boldsymbol{W}_2$。若不存在，则电容出现在 2 个工作环路矩阵中，且符号不一致，即 $C\in\boldsymbol{W}_1$ 且 $-C\in\boldsymbol{W}_2$（$C\in\boldsymbol{W}_2$ 且 $-C\in\boldsymbol{W}_1$）。

考虑到变压器的电气隔离特性，输入直流源 V_i 不能与变压器二次侧 T_s 或输出直流源 V_o 通过任意器件相连，变压器一次侧 T_p 不能与变压器二次侧 T_s 或输出直流源 V_o 通过任意器件相连。

4）电气约束 4。V_i 和 T_s、V_i 和 V_o、T_p 和 V_o、T_p 和 T_s 不能通过任意器件相连，即不能处于同一路径 $\boldsymbol{w}$ 中。

由于隔离型变换器主要分为正激变换器和反激变换器，根据所提出的变压器模型和基本电路原理，给出正激变换器和反激变换器中的变压器模型，如表 2 所示。

对于正激变换器，若不考虑磁复位，变压器一次侧、二次侧同时工作或不工作，若考虑磁复位，变

表 2 变压器在正激变换器和反激变换器中的模型

Table 2 Transformer model in forward converters and flyback converters

变压器	双顶点对	同名端位置	变压器模型	变压器工作情况
正激变换器	$T=(T_p,T_s)=(v_{i1},v_{j1},v_{j2},v_{i2})$	v_{i1},v_{i2}	模态1	
			模态2	
反激变换器	$T=(T_p,T_s)=(v_{i1},v_{j1},v_{j2},v_{i2})$	v_{i1},v_{j2}	模态1	
			模态2	

压器一次侧、二次侧同时工作, 且一次侧进行磁复位时, 二次侧不工作, 其同名端位于双顶点对中的第 1 个顶点和第 4 个顶点的位置。对于反激变换器, 若不考虑磁复位, 变压器一次侧、二次侧不同时工作, 若考虑磁复位, 变压器一次侧工作时, 二次侧不工作, 一次侧进行磁复位时, 二次侧工作, 其同名端位于双顶点对中的第 1 个顶点和第 3 个顶点的位置。

5) 电气约束 5。若存在开关管、电容直连情况, 变压器一次侧 T_p 可在 2 个工作环路矩阵中出现, 二次侧 T_s 只能在 1 个工作环路矩阵中, 该变换器为正激变换器或反激变换器, 如图 8(a)～图 8(d) 所示, 即 $T_p\in\boldsymbol{W}_1$ 或 $T_p\in\boldsymbol{W}_2$, $T_s\in\boldsymbol{W}_1$ 且 $T_s\notin\boldsymbol{W}_2$ ($T_s\in\boldsymbol{W}_2$ 且 $T_s\notin\boldsymbol{W}_1$)。

6) 电气约束 6。若不存在开关管、电容直连情况, T_p、T_s 只能在 1 个工作环路矩阵中出现, 且 T_p 和 T_s 同时在 1 个工作环路矩阵中出现, 则该变换器为正激变换器, 如图 8(e) 和图 8(f) 所示, 即 $T_p\in\boldsymbol{W}_1$ 且 $T_s\in\boldsymbol{W}_1$ ($T_p\in\boldsymbol{W}_2$ 且 $T_s\in\boldsymbol{W}_2$)。

7) 电气约束 7。若不存在开关管、电容直连情况, T_p、T_s 只能在 1 个工作环路矩阵中, 且 T_p 和 T_s 同时出现在不同工作环路矩阵中, 则该变换器为反激变换器, 如图 8(g) 和图 8(h) 所示, 即 $T_p\in\boldsymbol{W}_1$ 且 $T_s\in\boldsymbol{W}_2$ ($T_p\in\boldsymbol{W}_2$ 且 $T_s\in\boldsymbol{W}_1$)。

由于电气约束条件考虑电路基本原理和变换器中变压器一次侧、二次侧工作情况, 通过工作环路矩阵中环路与器件进行电气约束筛选。因此, 电气约束 1～电气约束 7 与工作模态有关。其中, 电气约束 1～电气约束 4 只考虑基本电路原理, 均适用于正激变换器或反激变换器。电气约束 5～电气约束 7 只考虑变换器中变压器一次侧、二次侧工作情况, 其中, 电气约束 5 适用于正激变换器或反激变换器, 电气约束 6 只适用于正激变换器, 电气约束 7 只适用于反激变换器。

如图 9 所示, 拓扑自动生成时, 根据电气约束 1～电气约束 4 判断拓扑是否基本电路原理, 在电气约束 1～电气约束 4 的基础上根据电气约束 5～

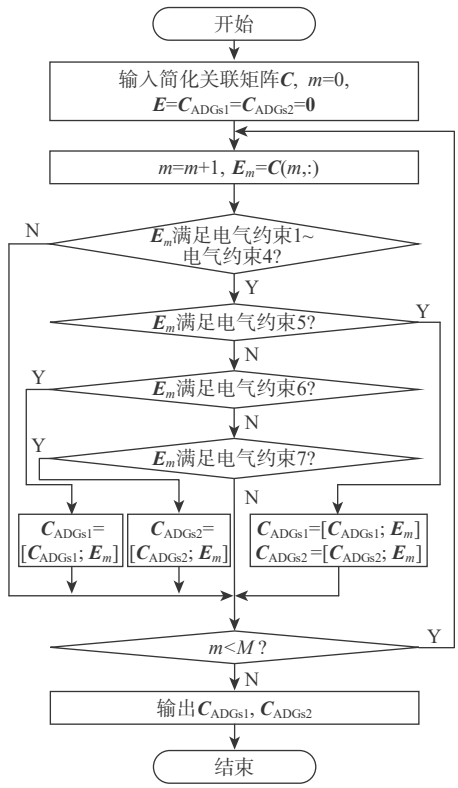


图 9　电气约束筛选流程

Fig. 9　Flow of electrical constraint screening

电气约束 7 分成 2 类进行自动生成: 满足电气约束 5, 可为正激变换器或反激变换器, 满足电气约束 6, 为正激变换器, 其简化关联矩阵为 $\boldsymbol{C}_{\mathrm{ADGs1}}$, 满足电气约束 7, 为反激变换器, 其简化关联矩阵为 $\boldsymbol{C}_{\mathrm{ADGs2}}$。

本文根据开关器件分成 4 种开关组合情况: ①组合 S_1、D_1 在模态 1 工作, S_2、D_2 在模态 2 工作; ②组合 2: S_1、D_2 在模态 1 工作, S_2、D_1 在模态 2 工作; ③组合 3: S_1 在模态 1 工作, S_2、D_1、D_2 在模态 2 工作; ④组合 4: S_2 在模态 1 工作, S_1、D_1、D_2 在模态 2 工作。变换器只需在任一个开关组合下满足电气约束条件, 即可正常工作。对 3 017 种图特性正常拓扑图进行拓扑筛选, 得到 158 种可正常工作正激变换器拓扑图, 其简化关联矩阵如式 (5) 所示。得到 168 种可正常工作反激变换器拓扑图, 其简化关联矩阵如式 (6) 所示。

$$\boldsymbol{C}_{\mathrm{ADGs1}}=\begin{array}{c}\\ \boldsymbol{E}_1\\ \boldsymbol{E}_2\\ \vdots\\ \boldsymbol{E}_{158}\end{array}\begin{array}{c}\begin{array}{cccccccccccccccccccc} T_s/e_{1m} & & T_p/e_{2m} & & S_1/e_{3m} & & S_2/e_{4m} & & D_1/e_{5m} & & D_2/e_{6m} & & L/e_{7m} & & C/e_{8m} & & V_i/e_{9m} & & V_o/e_{10m} & \end{array}\\ \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_6 & v_8 & v_6 & v_4 & v_7\\ v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_8 & v_8 & v_6 & v_4 & v_7\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_5 & v_7 & v_6 & v_1 & v_8 & v_3 & v_7 & v_2 & v_6 & v_7 & v_2 & v_5 & v_4 & v_8 \end{bmatrix}\end{array} \tag{5}$$

$$\boldsymbol{C}_{\mathrm{ADGs2}}=\begin{array}{c}\\ \boldsymbol{E}_1\\ \boldsymbol{E}_2\\ \vdots\\ \boldsymbol{E}_{168}\end{array}\begin{array}{c}\begin{array}{cccccccccccccccccccc} T_s/e_{1m} & & T_p/e_{2m} & & S_1/e_{3m} & & S_2/e_{4m} & & D_1/e_{5m} & & D_2/e_{6m} & & L/e_{7m} & & C/e_{8m} & & V_i/e_{9m} & & V_o/e_{10m} & \end{array}\\ \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_6 & v_8 & v_6 & v_4 & v_7\\ v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_8 & v_8 & v_6 & v_4 & v_7\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_5 & v_7 & v_6 & v_1 & v_8 & v_3 & v_7 & v_2 & v_6 & v_7 & v_2 & v_5 & v_4 & v_8 \end{bmatrix}\end{array} \tag{6}$$

2.5 环路冗余拓扑筛选

在得到的158种可正常工作正激变换器拓扑图和168种可正常工作反激变换器拓扑图中，存在部分拓扑工作性能一致，其电压增益相同，工作环路相同，器件应力相同，这些拓扑称为冗余拓扑，只需保留一种即可。对得到的158种可正常工作正激变换器拓扑图和168种可正常工作反激变换器拓扑图进行冗余拓扑筛选[17]，得到87种可正常工作无冗余正激变换器拓扑图，其简化关联矩阵如式(7)所示，拓扑图见附录图A1；得到97种可正常工作无冗余反激变换器拓扑图，其简化关联矩阵如式(8)所示，拓扑图见附录图A2。其中，图A1(e)、图A1(f)和图A2(c)分别于文献[18-20]中提出。

$$\boldsymbol{C}_{\mathrm{re}}=\begin{array}{c} \\ \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \vdots \\ \boldsymbol{E}_{87} \end{array}\begin{array}{c} \begin{array}{cccccccccccccccccc} T_{\mathrm{s}}/e_{1m} & T_{\mathrm{p}}/e_{2m} & S_1/e_{3m} & S_2/e_{4m} & D_1/e_{5m} & D_2/e_{6m} & L/e_{7m} & C/e_{8m} & V_{\mathrm{i}}/e_{9m} & V_{\mathrm{o}}/e_{10m} \end{array} \\ \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_6 & v_8 & v_6 & v_4 & v_7 \\ v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_8 & v_8 & v_6 & v_4 & v_7 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_7 & v_6 & v_6 & v_3 & v_7 & v_8 & v_8 & v_4 & v_6 & v_8 & v_2 & v_5 & v_4 & v_7 \end{bmatrix} \end{array} \tag{7}$$

$$\boldsymbol{C}_{\mathrm{re}}=\begin{array}{c} \\ \boldsymbol{E}_1 \\ \boldsymbol{E}_2 \\ \vdots \\ \boldsymbol{E}_{97} \end{array}\begin{array}{c} \begin{array}{cccccccccccccccccc} T_{\mathrm{s}}/e_{1m} & T_{\mathrm{p}}/e_{2m} & S_1/e_{3m} & S_2/e_{4m} & D_1/e_{5m} & D_2/e_{6m} & L/e_{7m} & C/e_{8m} & V_{\mathrm{i}}/e_{9m} & V_{\mathrm{o}}/e_{10m} \end{array} \\ \begin{bmatrix} v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_6 & v_8 & v_6 & v_4 & v_7 \\ v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_6 & v_1 & v_2 & v_5 & v_7 & v_3 & v_1 & v_8 & v_5 & v_8 & v_8 & v_6 & v_4 & v_7 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_1 & v_2 & v_3 & v_4 & v_5 & v_1 & v_7 & v_6 & v_8 & v_6 & v_7 & v_8 & v_6 & v_3 & v_3 & v_8 & v_2 & v_5 & v_4 & v_7 \end{bmatrix} \end{array} \tag{8}$$

3 方法验证

以隔离型 $\mathrm{T^1S^2D^2L^1C^1}$ DC-DC变换器为例进行拓扑自生成，得到87种正激变换器和97种反激变换器。为快速了解每个拓扑的具体功能，以87种正激变换器为例，进行电压增益计算。在不考虑变压器匝数比的情况下，根据电压增益将87种正激变换器拓扑分为25种升压正激变换器拓扑(部分示例见图A1(a)～(c))、24种降压正激变换器拓扑(部分示例见图A1(d)～(f))、32种升降压正激变换器拓扑(部分示例见图A1(g)～(i))和6种等压正激变换器拓扑(部分示例见图A1(j)～(l))。

由于拓扑数较多，难以一一进行验证，本文以87种正激变换器拓扑图中的5种含有有源钳位功能的拓扑为例进行验证。这5种含有有源钳位功能的拓扑的简化关联矩阵如式(9)所示，拓扑图如图10所示。

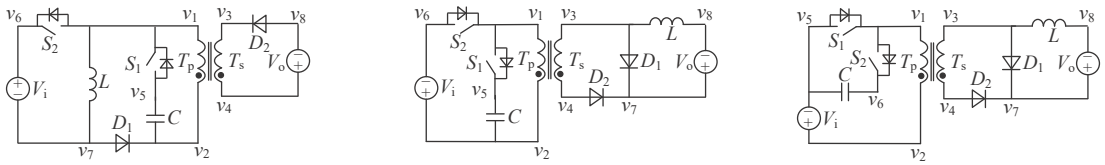
$$\boldsymbol{C}=\begin{array}{c} \\ \boldsymbol{E}_3 \\ \boldsymbol{E}_5 \\ \boldsymbol{E}_6 \\ \boldsymbol{E}_7 \\ \boldsymbol{E}_8 \end{array}\begin{array}{c} \begin{array}{cccccccccccccccccc} T_{\mathrm{s}}/e_{1m} & T_{\mathrm{p}}/e_{2m} & S_1/e_{3m} & S_2/e_{4m} & D_1/e_{5m} & D_2/e_{6m} & L/e_{7m} & C/e_{8m} & V_{\mathrm{i}}/e_{9m} & V_{\mathrm{o}}/e_{10m} \end{array} \\ \begin{bmatrix} v_2 & v_1 & v_3 & v_4 & v_1 & v_5 & v_6 & v_1 & v_7 & v_2 & v_8 & v_3 & v_1 & v_7 & v_2 & v_5 & v_7 & v_6 & v_4 & v_8 \\ v_2 & v_1 & v_3 & v_4 & v_5 & v_1 & v_1 & v_6 & v_3 & v_7 & v_4 & v_7 & v_8 & v_3 & v_2 & v_5 & v_6 & v_2 & v_7 & v_8 \\ v_2 & v_1 & v_3 & v_4 & v_1 & v_5 & v_6 & v_1 & v_3 & v_7 & v_4 & v_7 & v_8 & v_3 & v_5 & v_6 & v_5 & v_2 & v_7 & v_8 \\ v_2 & v_1 & v_3 & v_4 & v_5 & v_1 & v_1 & v_6 & v_3 & v_7 & v_4 & v_8 & v_8 & v_3 & v_2 & v_5 & v_6 & v_2 & v_7 & v_8 \\ v_2 & v_1 & v_3 & v_4 & v_1 & v_5 & v_6 & v_1 & v_3 & v_7 & v_4 & v_8 & v_8 & v_3 & v_5 & v_6 & v_5 & v_2 & v_7 & v_8 \end{bmatrix} \end{array} \tag{9}$$

在5种有源钳位正激变换器拓扑图中，2种拓扑已被提出，其中，图10(b)为1993年提出的高位有源钳位正激变换器[18]，图10(c)为1992年提出的低位有源钳位正激变换器[19]。这2种拓扑存在一定的时间间隔。本文在CPU: Intel I9-13900HX@2.2 GHz, RAM: 16 GB DDR5@5 600 MHz的笔记本电脑下，耗时2 h 41 min 11 s便可搜索出包含这2种拓扑在内的87种正激变换器拓扑和97种反激变换器拓扑，降低了拓扑提出的时间成本。同时，发现了其他3种有源钳位正激变换器拓扑结构。根据电压增益，将图10(a)、图10(d)和图10(e)中的拓扑命名为有源钳位升降压正激变换器。

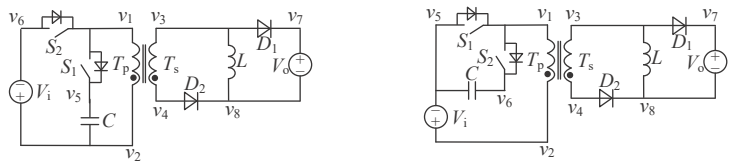
为验证本文方法的正确性，对新发现的3种有源钳位正激变换器进行理论分析和仿真验证，3种拓扑均可正常工作。由于篇幅限制，本文只给出以图10(a)拓扑为例的理论分析和仿真验证。该变换器分为4个工作模态，其中，工作模态2和工作模态4为主要工作模态，如图11所示，其理论波形如图12所示。

1) 0～t_0时刻。如图11(a)所示，在0时刻，开关管S_2体二极管导通，T_{p}通过D_1和L放电，L的电流不能突变而线性增加，S_2的电流反向且线性减小。

2) t_0～t_1时刻。如图11(b)所示，在t_0时刻，



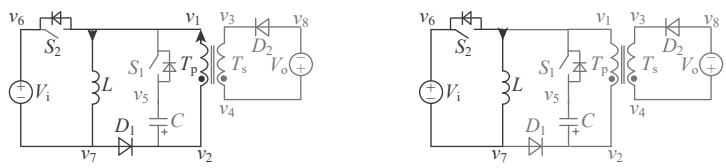
(a) $\boldsymbol{E}_3$拓扑结构　(b) $\boldsymbol{E}_5$拓扑结构　(c) $\boldsymbol{E}_6$拓扑结构



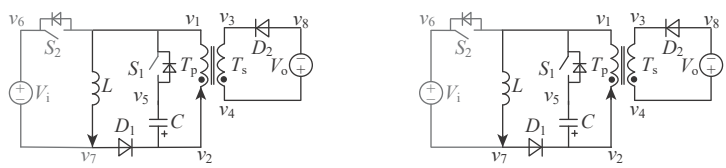
(d) $\boldsymbol{E}_7$拓扑结构　(e) $\boldsymbol{E}_8$拓扑结构

图 10　有源钳位正激变换器的 5 种拓扑结构

Fig. 10　5 topologies of active clamp forward converters



(a) 变换器$\boldsymbol{E}_3$工作模态1　(b) 变换器$\boldsymbol{E}_3$工作模态2



(c) 变换器$\boldsymbol{E}_3$工作模态3　(d) 变换器$\boldsymbol{E}_3$工作模态4

图 11　有源钳位升降压正激变换器 $\boldsymbol{E}_3$ 的工作模态

Fig. 11　Working mode of active clamp Buck-Boost forward converters $\boldsymbol{E}_3$

S_2 的电流小于 0, 开关管 S_2 开通, 实现零电流开关(zero current switching, ZCS)开通。V_i 通过 S_2 和 L 放电, L 的电流不能突变而线性增加, S_2 的电流线性增加。

3) t_1～t_2 时刻。如图 11(c) 所示, 在 t_1 时刻, 开关管 S_1 体二极管导通, L 通过 V_i、C、S_1 和 D_1 放电, L 的电流不能突变而线性减小, S_1 的电流反向且线性减小。

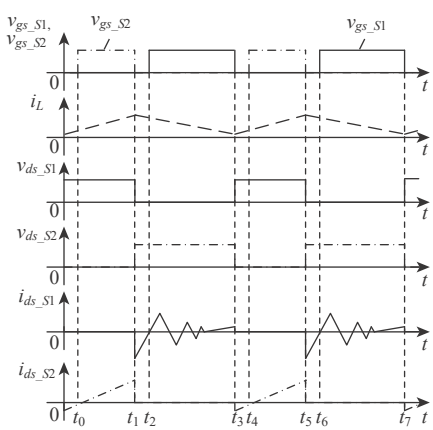


图 12　变换器理论波形

Fig. 12　Theoretical waveform of converter

4) t_2～t_3 时刻。如图 11(d) 所示, 在 t_2 时刻, S_1 的电流小于 0, 开关管 S_1 开通, 实现 ZCS 开通。由于 C 两端电压振荡, S_1 的电流振荡。L 通过 T_p 和 D_1 放电, L 的电流不能突变而线性减小, S_1 的电流增加。

给出表 3 所示仿真参数, 使用 Simulink 对所提变换器进行仿真分析, 设置死区时间为 1×10^{-7} s, 不考虑开关寄生电容。

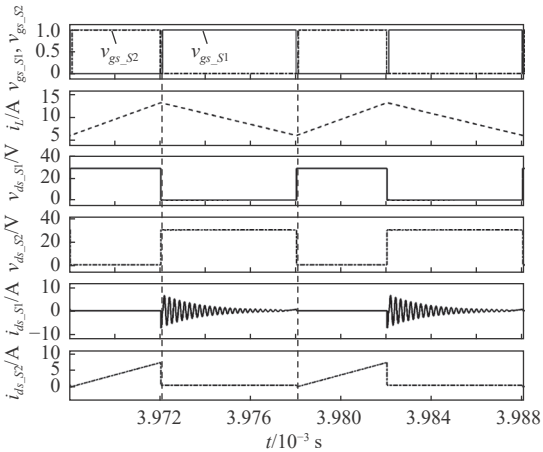
变换器的电压增益为 $M = DN_2/[(1-D)N_1]$, 仿

表 3　仿真参数

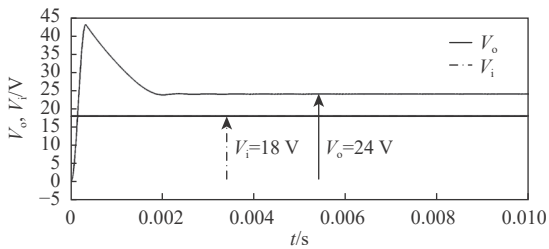
Table 3　Simulation parameters

参数	数值
输入电压V_i/V	18
输出功率P_o/W	24
开关频率f_s/kHz	100
电感L/μH	10
励磁电感L_m/μH	100
输出电压V_o/V	24
占空比D	0.4
变压器匝数比N_1:N_2	1:2
电容C/μF	1
输出电容C_o/μF	100

真结果如图 13 所示。2 个开关管在互补导电模式下工作，当主开关 S_2 的占空比为 0.4 时，输入电压为 18 V，输出电压为 24 V，与理论结果一致。在 S_1 和 S_2 接通前，其电流均先降至 0，因此，S_1 和 S_2 均可实现 ZCS 开通。该拓扑可实现软开关功能，减小开关损耗，提高效率。



(a) 开关管电压与电流波形



(b) 输入电压与输出电压波形

图 13 图 10(a) 拓扑仿真结果

Fig. 13 simulation results of topology in Fig.10(a)

4 结 论

1）本文通过加入双顶点对的变压器模型，提出了一种隔离型 DC-DC 变换器拓扑自生成方法，实现了基于计算机编程快速、量化生成新的隔离型 DC-DC 变换器拓扑。

2）本文方法能够在 2 h 41 min 11 s 内一次生成 87 种可用正激变换器和 97 种可用反激变换器。以正激变换器为例，根据电压增益可分为 25 种升压正激变换器、24 种降压正激变换器、32 种升降压正激变换器和 6 种等压正激变换器，为工程应用提供了拓扑库，为拓扑优化选型提供了条件。

3）本文方法虽然以隔离型 $T^1S^2D^2L^1C^1$ DC-DC 变换器为例进行验证，但不局限隔离型 $T^1S^2D^2L^1C^1$ DC-DC 变换器，还可用于自动生成具有更多元器件或更多变压器的隔离型 DC-DC 变换器拓扑。

4）本文方法虽然能够快速生成多种拓扑结构，但只提供拓扑优化选型条件，后续仍需优化拓扑自生成方法，使其能够自动进行拓扑优化选型。

参考文献（References）

[1] WU H F, XING Y. A family of forward converters with inherent demagnetizing features based on basic forward cells[J]. IEEE Transactions on Power Electronics, 2010, 25(11): 2828-2834.

[2] TAN F D. The forward converter: from the classic to the contemporary [C]//Proceedings of the 17th Annual IEEE Applied Power Electronics Conference and Exposition. Piscataway: IEEE Press, 2002: 857-863.

[3] TSENG S Y, CHEN W C, LI Y J, et al. Buck-Boost converter associated with active clamp forward converter for PV power system [C]//Proceedings of the 7th International Conference on Power Electronics and Drive Systems. Piscataway: IEEE Press, 2007: 209-216.

[4] LI W H, FAN L L, ZHAO Y, et al. High-step-up and high-efficiency fuel-cell power-generation system with active-clamp flyback-forward converter[J]. IEEE Transactions on Industrial Electronics, 2012, 59(1): 599-610.

[5] KIM T S, HAN S K, MOON G W, et al. High efficiency active clamp forward converter for sustaining power module of plasma display panel[J]. IEEE Transactions on Industrial Electronics, 2008, 55(4): 1874-1876.

[6] NAKAHARA M, NINOMIYA T, HARADA K. Surge and noise generation in a forward DC-to-DC converter[J]. IEEE Transactions on Aerospace and Electronic Systems, 1985, 21(5): 619-630.

[7] JOVANOVIC M M, ZHANG M T, LEE F C. Evaluation of synchronous-rectification efficiency improvement limits in forward converters[J]. IEEE Transactions on Industrial Electronics, 1995, 42(4): 387-395.

[8] WITTENBREDER E H, BAGGERLY V D, MARTIN H C. A duty cycle extension technique for single ended forward converters[C]// Proceedings of the 7th Annual Applied Power Electronics Conference and Exposition. Piscataway: IEEE Press, 2002: 51-57.

[9] COBOS J A, GARCIA O, SEBASTIAN J, et al. Resonant reset forward topologies for low output voltage on board converters[C]//Proceedings of the IEEE Applied Power Electronics Conference and Exposition. Piscataway: IEEE Press, 2002: 703-708.

[10] SALAZAR L D, ZIOGAS P D. A high frequency two-switch forward converter with optimized performance[C]//Proceedings of the 15th Annual Conference of IEEE Industrial Electronics Society. Piscataway: IEEE Press, 2002: 60-66.

[11] TSAI F S, NG W W. A low-cost, low-loss active voltage-clamp circuit for interleaved single-ended forward PWM converter[C]//Proceedings of the 8th Annual Applied Power Electronics Conference and Exposition. Piscataway: IEEE Press, 1993: 729-733.

[12] SRIVASTAVA M, TOMAR P S, VERMA A K. Full range soft-switched DC-DC converter with clamping circuit for transportation applications[C]//Proceedings of the IEEE Industry Applications Society Annual Meeting. Piscataway: IEEE Press, 2021: 1-5.

[13] WEI Y Q, LUO Q M, MANTOOTH A. A function decoupling partially isolated high voltage gain DC/DC converter for PV application[C]//Proceedings of the IEEE Transportation Electrification Conference & Expo. Piscataway: IEEE Press, 2020: 1-5.

[14] MURTHY-BELLUR D, PALMER B, IYER K, et al. Commercial DC-DC converters: a survey of isolated low-voltage very-high-current systems for transportation electrification[C]//Proceedings of the IEEE Power and Energy Conference. Piscataway: IEEE Press, 2017: 1-8.

[15] CHEN G P, JIN Z F, LIU Y W, et al. Programmable topology derivation and analysis of integrated three-port DC-DC converters with reduced switches for low-cost applications[J]. IEEE Transactions on Industrial Electronics, 2019, 66(9): 6649-6660.

[16] LI H, LI Y M, YIN C D, et al. Vertex prime degree-based non-isomorphic topology automatic search algorithm for DC-DC converters with two switches[J]. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 2023, 4(4): 1084-1095.

[17] 李虹, 王文财, 李亚敏, 等. 基于图论的 $S^1D^1L^2C^1$ 型 DC-DC 变换器可编程拓扑搜索算法[J]. 中国电机工程学报, 2021, 41(16): 5670-5683.
LI H, WANG W C, LI Y M, et al. Programmable topology searching algorithm for $S^1D^1L^2C^1$ type DC-DC converters based on graph theory[J]. Proceedings of the CSEE, 2021, 41(16): 5670-5683(in Chinese).

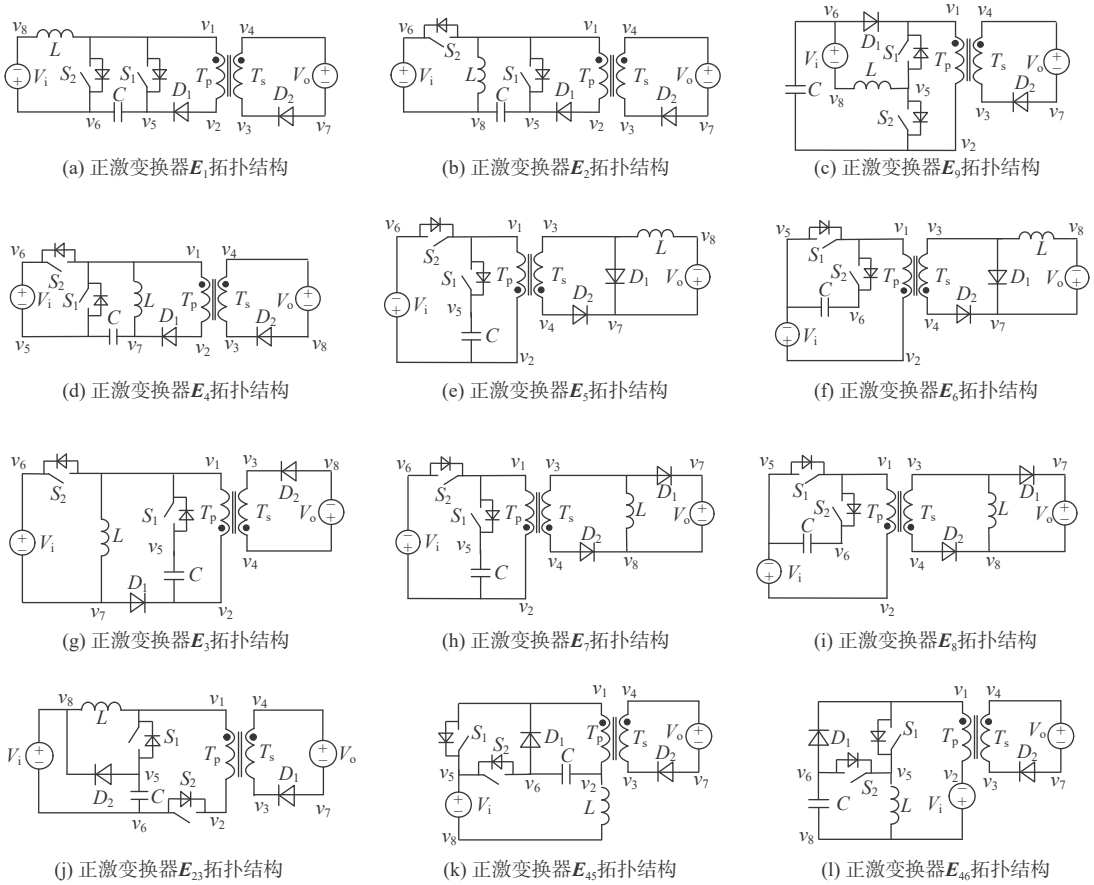
[18] COBOS J A, GARCIA O, SEBASTIAN J, et al. Active clamp PWM forward converter with self driven synchronous rectification[C]//Proceedings of the 15th International Telecommunications Energy Conference. Piscataway: IEEE Press, 1993: 200-206.

[19] KIM H J, LEU C S, FARRINGTON R, et al. Clamp mode zero-voltage-switched multi-resonant converters[C]//Proceedings of the 23rd Annual IEEE Power Electronics Specialists Conference. Piscataway: IEEE Press, 1992: 78-84.

[20] SHIN C H, KIM D Y, KO A Y, et al. The configuration of electric vehicle system using isolated DC-DC converter for a low-voltage and high-current type battery[C]//Proceedings of the 9th International Conference on Power Electronics and ECCE Asia. Piscataway: IEEE Press, 2015: 2796-2801.

附录 A:

图 A1 为 87 种可正常工作无冗余正激变换器的部分拓扑图, 图 A2 为 97 种正常工作无冗余反激变换器的部分拓扑图。



(a) 正激变换器$\boldsymbol{E}_1$拓扑结构　(b) 正激变换器$\boldsymbol{E}_2$拓扑结构　(c) 正激变换器$\boldsymbol{E}_9$拓扑结构

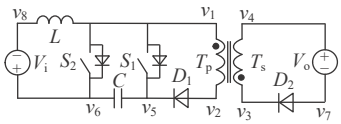
(d) 正激变换器$\boldsymbol{E}_4$拓扑结构　(e) 正激变换器$\boldsymbol{E}_5$拓扑结构　(f) 正激变换器$\boldsymbol{E}_6$拓扑结构

(g) 正激变换器$\boldsymbol{E}_3$拓扑结构　(h) 正激变换器$\boldsymbol{E}_7$拓扑结构　(i) 正激变换器$\boldsymbol{E}_8$拓扑结构

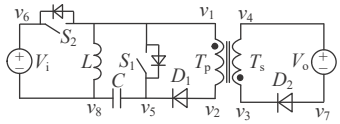
(j) 正激变换器$\boldsymbol{E}_{23}$拓扑结构　(k) 正激变换器$\boldsymbol{E}_{45}$拓扑结构　(l) 正激变换器$\boldsymbol{E}_{46}$拓扑结构

图 A1　87 种正激变换器拓扑的部分电路图

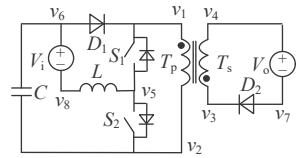
Fig. A1　Partial circuit diagram of 87 types of forward converter topologies



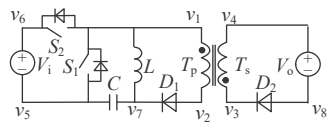
(a) 反激变换器$\boldsymbol{E}_1$拓扑结构



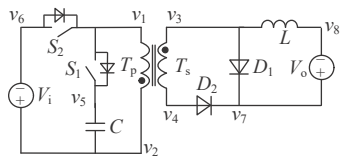
(b) 反激变换器$\boldsymbol{E}_2$拓扑结构



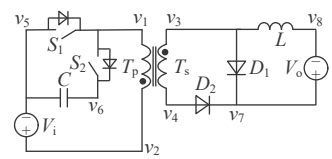
(c) 反激变换器$\boldsymbol{E}_9$拓扑结构



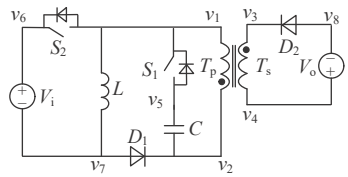
(d) 反激变换器$\boldsymbol{E}_4$拓扑结构



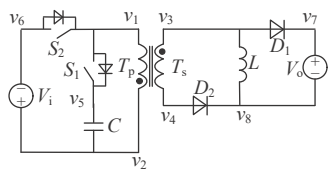
(e) 反激变换器$\boldsymbol{E}_5$拓扑结构



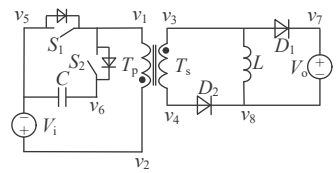
(f) 反激变换器$\boldsymbol{E}_6$拓扑结构



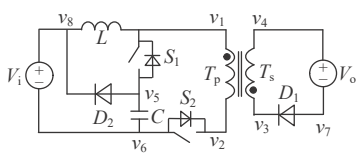
(g) 反激变换器$\boldsymbol{E}_3$拓扑结构



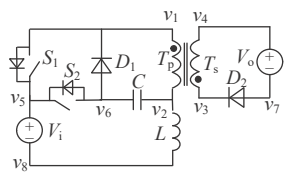
(h) 反激变换器$\boldsymbol{E}_7$拓扑结构



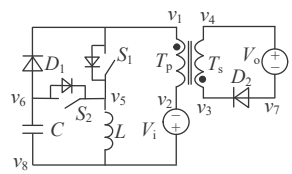
(i) 反激变换器$\boldsymbol{E}_8$拓扑结构



(j) 反激变换器$\boldsymbol{E}_{23}$拓扑结构



(k) 反激变换器$\boldsymbol{E}_{56}$拓扑结构



(l) 反激变换器$\boldsymbol{E}_{57}$拓扑结构

图 A2 97 种反激变换器拓扑的部分电路图

Fig. A2 Partial circuit diagram of 97 types of flyback converter topologies

Topology self-generating method of isolated DC-DC converters with two switches

LI Hong[1, 2, *], LIANG Yidi[2], YIN Chengdong[2], ZHENG Qionglin[2], ZHANG Bo[3]

(1. College of Electrical Engineering, Zhejiang University, Hangzhou 310058, China;
2. School of Electrical Engineering, Beijing Jiaotong University, Beijing 100044, China;
3. School of Electric Power Engineering, South China University of Technology, Guangzhou 510006, China)

Abstract: More isolated DC-DC converter topologies are required as a result of the growing need for different power conversion brought on by the development of new energy generation and electrified transportation. To overcome the randomness and lack of topology optimization in generating topologies based on human experience, this paper proposes a vertex prime degree-based topology self-generation method of the isolated DC-DC converters with two switches. Firstly, the two-vertex-pair transformer model is proposed, and the circuits are converted into graphs using graph theory. Secondly, the numbers and the types of components used to form the circuit are regarded as the input condition of the method. Finally, the proposed self-generation method is utilized to traverse and generate all available isolated DC-DC converter topologies with the given components. An input DC source, an output DC source, two switches, two diodes, one transformer, one inductor, and one capacitor, namely, the isolated $T^1S^2D^2L^1C^1$ DC-DC converters are examples of converter circuits. The 87 forward converters and 97 flyback converters can be generated. Taking the 87 forward converters as examples, they can be further categorized based on voltage gain into 25-boost forward converters, 24-buck forward converters, 32 buck-boost forward converters, and 6 constant voltage forward converters. Therefore, this method enables rapid and quantitative generation of new isolated DC-DC converter topologies through computer programming. It provides a topology library for engineering applications and establishes conditions for topology optimization and selection.

Keywords: isolated DC-DC converter; two switches; topology; self-generation; vertex prime degree

Received: 2023-07-27; **Accepted**: 2023-10-06; **Published Online**: 2023-12-18 14:33

URL: link.cnki.net/urlid/11.2625.V.20231218.0849.001

Foundation items: National Science Fund for Distinguished Young Scholars (52325704); Key Program of National Natural Science Foundation of China (52237008)

* **Corresponding author**. E-mail: hli@bjtu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0277

非对称电网状态下MMC-HVDC改进滑模直接功率控制策略

关天一[1,2]，王怡博[1]，王睿[1,*]，赵海乔[1]，郑文婧[3]，孙秋野[1]

(1. 东北大学 信息科学与工程学院，沈阳 110819；2. 国网辽宁省电力有限公司沈阳供电公司，沈阳 110811；
3. 国网辽宁省电力有限公司营销服务中心，沈阳 110168)

摘 要：基于模块化多电平换流器的高压直流(MMC-HVDC)输电系统是动态功率平衡系统，在非对称电网状态下，MMC的功率控制一般采用基于传统瞬时功率模型的双闭环矢量控制策略，存在控制结构复杂、控制精度低等问题。为此，引入柔性瞬时功率模型，以有功功率和柔性无功功率为控制对象建立通用功率方程，提出一种基于柔性瞬时功率模型的MMC-HVDC改进滑模直接功率控制策略。该策略结合柔性瞬时功率模型和改进滑模控制方法，消除非对称电网状态下MMC输出功率的波动分量，省略了电流内环与功率补偿项，优化了控制结构。仿真和实验结果表明：所提控制策略在非对称电网、参数摄动等运行工况下的动态响应能力、控制精度和鲁棒性更强，更能发挥出柔性瞬时功率模型的优势。

关 键 词：基于模块化多电平换流器的高压直流；非对称电网；滑模控制；柔性瞬时功率模型；鲁棒控制

中图分类号：TM46

文献标志码：A **文章编号**：1001-5965(2025)05-1428-12

模块化多电平换流器(modular multilevel converter, MMC)相比于传统两、三电平电压源型换流器(voltage source converter, VSC)，具有高度模块化、扩展性强、输出波形质量高等优势，已成为高功率电力电子应用中最具前景的多电平换流器拓扑。MMC通过大量堆叠半桥、全桥等不同类型的子模块实现高压电能变换功能，在国内的直流工程应用中呈现上升趋势[1-3]。

MMC是直流输电系统的重要组成部分，其控制策略对整个系统的稳定运行至关重要。目前，基于MMC的高压直流(modular multilevel converter based high voltage direct current, MMC-HVDC)输电系统的控制系统的主流设计方案是以快速电流反馈为特征的直接电流控制策略(矢量控制)，其具体形式为双闭环串级控制，可以获得高品质的电流响应，该方法在同步旋转坐标系(*dq*坐标系)下的实现方案已有较多研究[4-6]。滑模控制(sliding mode control, SMC)作为一种经典的非线性控制方法，具有良好的稳态精度、动态性能和抗干扰性等优点，被广泛应用于PWM整流器、同步发电机、风机系统、AC/DC换流器[7-8]等场景。文献[8]研究了SMC在MMC矢量控制中的应用，提出了一种MMC改进滑模控制策略，具有更好的动态响应能力、抗干扰能力和较小的抖振。然而，上述控制策略均在对称电网状态下进行讨论。工程实际中，由于电网谐波、不对称故障等原因，三相电压经常出现不对称[9-10]，导致MMC输出的有功功率和无功功率出现二倍频波动，严重影响MMC-HVDC的运行性能[11]。

收稿日期：2023-05-25；录用日期：2023-11-15；网络出版时间：2023-12-06 09:39
网络出版地址：link.cnki.net/urlid/11.2625.V.20231205.1012.001
基金项目：国家自然科学基金(U20A20190,62073065)
*通信作者. E-mail：wangrui@ise.neu.edu.cn

引用格式：关天一，王怡博，王睿，等. 非对称电网状态下MMC-HVDC改进滑模直接功率控制策略[J]. 北京航空航天大学学报，2025，51(5)：1428-1439. GUAN T Y，WANG Y B，WANG R，et al. Improved sliding-mode direct power control strategy for MMC-HVDC under asymmetrical grid state[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1428-1439 (in Chinese).

因此，为提高 AC/DC 换流器在非对称电网条件下的暂态控制性能，各国学者对其进行了大量研究。目前，MMC 的主流暂态控制策略分为双闭环矢量控制策略和以有功功率和无功功率为控制对象的直接功率控制策略。双闭环矢量控制策略根据坐标系的不同分为基于 dq 坐标系的 PI 控制策略[12]和基于 $\alpha\beta$ 坐标系的比例谐振（proportional resonant, PR）控制策略[13]，该方法能够分别控制正负序电流，但控制结构较为复杂，且控制效果受锁相环性能影响。相比之下，直接功率控制策略[14-17]以有功功率和无功功率建立控制模型，省略电流内环，简化了控制系统的设计。文献 [14] 结合 $\alpha\beta$ 坐标系提出了一种适用于并网逆变器的直接功率控制策略，通过对功率补偿分量的计算实现对有功功率和无功功率二倍频波动的有效抑制。文献 [15] 针对 PWM 整流器提出一种基于谐振调节器的直接功率控制策略，通过构建直流电压-功率闭环和谐振闭环实现对有功功率和无功功率波动分量的控制。文献 [16-17] 分别采用 SMC 和无差拍控制对 VSC 和 MMC 进行直接功率控制，通过增加功率补偿项保障网侧电流无畸变。文献 [18] 提出了一种功率滑模变结构补偿策略，可在不保证网侧电流无畸变的同时抑制二倍频功率波动分量，但无法完全消除功率的二倍频波动。

上述文献所述控制策略均基于传统的瞬时功率定义，在非对称电网条件下，无法在消除有功功率和无功功率二倍频波动的同时维持网侧电流不畸变。为解决该问题，文献 [19] 提出了一种柔性瞬时无功功率定义，使其适用于非对称电网的运行工况。文献 [20-21] 将柔性瞬时功率模型分别结合预测控制和降阶矢量控制开展应用，其在非对称电网条件下表现出良好的控制性能。

SMC 是一类特殊的非线性控制，可以控制系统按照预定“滑动模态”的状态轨迹运动，具有快速响应、不依赖被控对象数学模型、抗外部干扰能力强等优点[22]，广泛应用于同步发电机、风机系统等系统。然而，抖振问题是 SMC 中无法回避的重要问题，传统滑模控制采用线性滑模面、趋近律等方法，但在保证趋近速率的同时对抖振的抑制能力不足[23]。文献 [24-26] 分别提出了分数阶滑模方法、新型趋近律、准滑模方法等策略抑制抖振现象，但在控制器设计、控制精度等方面仍存在问题。

为进一步探讨柔性瞬时功率模型在 MMC 暂态控制中的应用，并降低常规滑模控制的抖振问题，本文提出一种基于柔性瞬时功率模型的 MMC-HVDC 改进滑模直接功率控制策略。该策略结合柔性瞬时功率模型，可以在非对称电网状态下同时消除 MMC 输出有功功率和无功功率的二倍频波动，并保持交流侧电流无畸变；同时，与 PI 控制策略和传统 SMC 策略相比，该策略在动态响应能力、鲁棒性及抗抖振方面具有明显优势。首先，介绍 MMC 的拓扑结构和数学模型；然后，对比分析柔性瞬时功率模型与传统瞬时功率模型，并推导基于柔性瞬时功率模型的 MMC 功率控制方程，提出基于改进 SMC 策略的 MMC 直接功率控制方法，深入分析改进 SMC 策略的控制性；最后，在 PSCAD/EMTDC 中搭建仿真模型，并构建 MMC 硬件在环实验平台，在电压对称、非对称条件、参数扰动条件下对所提控制策略进行对比仿真和实验验证。

1　MMC 数学模型

MMC 典型的三相拓扑结构如图 1 所示。每相桥臂由 N 个串联的子模块 (sub-module, SM)、桥臂等效电阻 R_0 及桥臂电抗器 L_0 构成，u_{pk} 和 $u_{nk}(k=a,b,c)$ 为各相桥臂的子模块电压之和，i_{pk} 和 i_{nk} 为各相桥臂的电流，每相桥臂输出的相电压中会产生 N+1 个电平，通过最近电平调制策略逼近正弦波。在 MMC 的交流侧，O 和 O' 分别为直流侧和交流侧的中性点（$U_{OO'}$=0），三相等效电阻和电感分别为 R_{ac} 和 L_{ac}，U_{sk} 为交流侧等效电源，i_{vk} 为 MMC 的出口三相电流。

由图 1 可以看出，MMC 换流器的每个桥臂是由 N 个结构相同的子模块和桥臂电抗器串联构成。每个子模块包含一个 IGBT 半桥和一个储能电容 C_0，通过调制策略控制 IGBT 的通断，从而对子模块进行投切，实现对 MMC 的控制。

根据基尔霍夫电压定律，对 MMC 各桥臂回路进行推导，可得到 MMC 的交流侧和直流侧的数学模型为

$$L\dot{i}_{vk}+Ri_{vk}=U_{diffk}-U_{sk} \tag{1}$$

$$L_0\dot{i}_{lk}+R_0i_{lk}=\frac{U_{dc}}{2}-U_{comk} \tag{2}$$

式中：U_{diffk} 为 MMC 的 k 相上下桥臂的差模电压；U_{comk} 为 MMC 的 k 相上下桥臂的共模电压；i_{lk} 为 MMC 内部环流。

$$\begin{cases}U_{diffk}=(u_{nk}-u_{pk})/2\\U_{comk}=(u_{nk}+u_{pk})/2\\L=L_{ac}+0.5L_0\\R=R_{ac}+0.5R_0\\i_{lk}=(i_{nk}+i_{pk})/2\end{cases}$$

对 MMC 交流侧的数学模型式 (1) 进行克拉克变换，将三相静止坐标系下的正弦交流量变换到两相 $\alpha\beta$ 静止坐标系正弦交流量，则 MMC 在两相 $\alpha\beta$ 静止坐标系下的数学模型可以表示为

$$L\dot{i}_{v\alpha}+Ri_{v\alpha}=U_{diff\alpha}-U_{s\alpha} \tag{3}$$

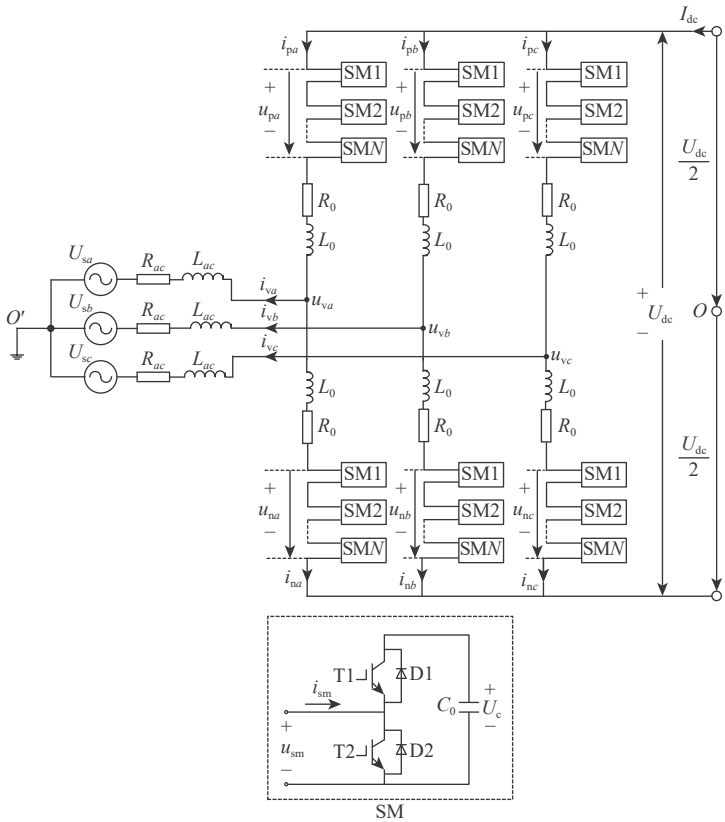


图1 MMC等效电路

Fig. 1 MMC equivalent circuit

$$L\dot{i}_{v\beta}+Ri_{v\beta}=U_{diff\beta}-U_{s\beta} \tag{4}$$

式中：$U_{s\alpha}$、$U_{s\beta}$分别为U_{sk}的α、β轴分量；$i_{v\alpha}$、$i_{v\beta}$分别为i_{vk}的α、β轴分量；$U_{diff\alpha}$、$U_{diff\beta}$分别为U_{diffk}的α、β轴分量。

2 基于柔性瞬时功率模型的MMC功率模型

2.1 柔性瞬时功率模型

传统瞬时功率模型可以表示为[27]

$$S=1.5U_{\alpha\beta}i_{\alpha\beta}^{*} \tag{5}$$

式中："*"表示共轭；$U_{\alpha\beta}$和$i_{\alpha\beta}$分别为电网交流侧电压和电流的α、β轴分量。

在非对称电网条件下，基于传统瞬时功率模型的MMC功率控制策略无法在保证交流侧电流无畸变的情况下，同时消除有功和无功的二倍频波动分量。在不对称的非理想电网条件下，只有瞬时有功功率具备切实的物理和数学意义。因此，本文引入柔性瞬时功率模型，重新定义瞬时无功的表达式，更适用于三相电压不对称的非理想电网。

柔性瞬时功率模型主要针对瞬时无功功率的表达式进行改进，具体表现为电压延迟矢量与电流矢量共轭的乘积的实部，表达式为

$$S^{\text{new}}=P+\mathrm{j}Q^{\text{new}}=1.5\mathrm{Re}(U_{\alpha\beta}i_{\alpha\beta}^{*})+\mathrm{j}1.5\mathrm{Re}(U_{\alpha\beta}^{\text{lag}}i_{\alpha\beta}^{*}) \tag{6}$$

式中：$U_{\alpha\beta}^{\text{lag}}=U_{\alpha\beta}(t-T/4)=U_{\alpha}^{\text{lag}}+\mathrm{j}U_{\beta}^{\text{lag}}$为电压延迟矢量，滞后正常电压90°，$T$为时间周期，值为0.02 s。

当三相电压对称时，负序分量不存在，由式(6)可知，2种瞬时无功功率的表达式相同，即

$$Q=1.5\mathrm{Im}(U_{\alpha\beta}i_{\alpha\beta}^{*})=U_{\beta}i_{\alpha}-U_{\alpha}i_{\beta} \tag{7}$$

$$Q^{\text{new}}=1.5\mathrm{Re}(U_{\alpha\beta}^{\text{lag}}i_{\alpha\beta}^{*})=1.5\mathrm{Re}(-\mathrm{j}U_{\alpha\beta}i_{\alpha\beta}^{*})=U_{\beta}i_{\alpha}-U_{\alpha}i_{\beta} \tag{8}$$

当三相电压不对称时，电压、电流在$\alpha\beta$坐标系下的瞬时矢量可以表示为

$$\begin{cases}U_{\alpha\beta}^{+}=U_{\alpha}^{+}+\mathrm{j}U_{\beta}^{+}=U^{+}\mathrm{e}^{\mathrm{j}(\omega t+\theta_u^{+})}\\ U_{\alpha\beta}^{-}=U_{\alpha}^{-}+\mathrm{j}U_{\beta}^{-}=U^{-}\mathrm{e}^{\mathrm{j}(-\omega t+\theta_u^{-})}\\ i_{\alpha\beta}^{+}=i_{\alpha}^{+}+\mathrm{j}i_{\beta}^{+}=I^{+}\mathrm{e}^{\mathrm{j}(\omega t+\theta_i^{+})}\\ i_{\alpha\beta}^{-}=i_{\alpha}^{-}+\mathrm{j}i_{\beta}^{-}=I^{-}\mathrm{e}^{\mathrm{j}(-\omega t+\theta_i^{-})}\\ U_{\alpha\beta}^{\text{lag}}=U_{\alpha\beta}^{+}(t-T/4)+U_{\alpha\beta}^{-}(t-T/4)=-\mathrm{j}U_{\alpha\beta}^{+}+\mathrm{j}U_{\alpha\beta}^{-}\end{cases} \tag{9}$$

式中：ω为基波角频率；$U_{\alpha\beta}^{+}$、$U_{\alpha\beta}^{-}$和$i_{\alpha\beta}^{+}$、$i_{\alpha\beta}^{-}$分别为电压和电流在$\alpha\beta$坐标系下的正序、负序矢量；I^{+}、I^{-}分别为电流正序、负序分量的幅值；U^{+}、U^{-}分别为电压正序、负序分量的幅值；θ_u^{+}和θ_u^{-}分别表示电压正序和负序分量的初相角；θ_i^{+}和θ_i^{-}分别表示电流正序和负序分量的初相角。

将式(9)代入式(6)，可得$\alpha\beta$坐标系下柔性瞬

时功率模型的表达式为

$$\begin{cases} P = 1.5(P_0 + P_1 + P_2 + P_3) \\ Q^{\text{new}} = 1.5(Q_a + Q_b + Q_c + Q_d) \end{cases} \tag{10}$$

式中：

$$\begin{cases} P_0 = U_\alpha^+ i_\alpha^+ + U_\beta^+ i_\beta^+ \\ P_1 = U_\alpha^- i_\alpha^- + U_\beta^- i_\beta^- \\ P_2 = U_\alpha^+ i_\alpha^- + U_\beta^+ i_\beta^- \\ P_3 = U_\alpha^- i_\alpha^+ + U_\beta^- i_\beta^+ \\ Q_a = U_\beta^+ i_\alpha^+ - U_\alpha^+ i_\beta^+ \\ Q_b = U_\alpha^- i_\beta^- - U_\beta^- i_\alpha^- \\ Q_c = U_\beta^+ i_\alpha^- - U_\alpha^+ i_\beta^- \\ Q_d = U_\alpha^- i_\beta^+ - U_\beta^- i_\alpha^+ \end{cases}$$

结合式(9)和式(10)，可得各功率分量的表达式为

$$\begin{cases} P_0 = U^+ I^+ \cos(\theta_u^+ - \theta_i^+) \\ P_1 = U^- I^- \cos(\theta_i^- - \theta_u^-) \\ P_2 = U^+ I^- \cos(2\omega t + \theta_u^+ - \theta_i^-) \\ P_3 = U^- I^+ \cos(2\omega t + \theta_i^+ - \theta_u^-) \end{cases} \tag{11}$$

$$\begin{cases} Q_a = U^+ I^+ \sin(\theta_u^+ - \theta_i^+) \\ Q_b = U^- I^- \sin(\theta_i^- - \theta_u^-) \\ Q_c = U^+ I^- \sin(2\omega t + \theta_u^+ - \theta_i^-) \\ Q_d = U^- I^+ \sin(2\omega t + \theta_i^+ - \theta_u^-) \end{cases} \tag{12}$$

由式(10)可知，不对称电压条件下，柔性瞬时功率模型中的有功功率和无功功率，在正序分量 P_0 和 Q_a 的基础上分别产生了电压、电流负序分量相互作用的直流量 P_1 和 Q_b，以及电压、电流的正、负序分量相互作用的二倍频波动分量 P_2、P_3 和 Q_c、Q_d，此时，若控制有功功率的二倍频波动为 0，即满足 P_2+P_3=0，则需要满足：

$$\begin{cases} U^+ I^- = U^- I^+ \\ \cos(2\omega t + \theta_u^+ - \theta_i^-) = -\cos(2\omega t + \theta_i^+ - \theta_u^-) \end{cases} \tag{13}$$

即

$$\begin{cases} U^+ I^- = U^- I^+ \\ \theta_u^+ - \theta_i^- = \theta_i^+ - \theta_u^- + \pi \end{cases} \tag{14}$$

将式(14)代入式(12)，可知

$$\begin{aligned} Q_c &= U^+ I^- \sin(2\omega t + \theta_u^+ - \theta_i^-) = \\ &\quad - U^- I^+ \sin(2\omega t + \theta_i^+ - \theta_u^-) = -Q_d \end{aligned} \tag{15}$$

由式(15)可知，在柔性瞬时无功功率定义下，当控制瞬时有功功率的二倍频波动为 0 时，柔性瞬时无功功率的二倍频波动之和 Q_c+Q_d 也为 0，因此，基于柔性瞬时无功功率定义的直接功率控制策略能在保障三相电流不发生畸变的条件下，同时抑制有功功率和无功功率的二倍频波动，相比于传统瞬时功率模型，更适用于电网三相电压不对称的工况。

2.2　MMC 柔性瞬时功率模型

由式(6)可知，瞬时有功功率和柔性瞬时无功功率的表达式分别写为

$$\begin{cases} P = 1.5(U_{s\alpha} i_{v\alpha} + U_{s\beta} i_{v\beta}) \\ Q^{\text{new}} = 1.5(U_{s\alpha}^{\text{lag}} i_{v\alpha} + U_{s\beta}^{\text{lag}} i_{v\beta}) \end{cases} \tag{16}$$

由式(16)可知，瞬时有功功率与柔性瞬时无功功率对时间 t 的导数可以分别写为

$$\frac{\mathrm{d}P}{\mathrm{d}t} = 1.5\left(\frac{\mathrm{d}U_{s\alpha}}{\mathrm{d}t} i_{v\alpha} + U_{s\alpha}\frac{\mathrm{d}i_{v\alpha}}{\mathrm{d}t} + \frac{\mathrm{d}U_{s\beta}}{\mathrm{d}t} i_{v\beta} + U_{s\beta}\frac{\mathrm{d}i_{v\beta}}{\mathrm{d}t}\right) \tag{17}$$

$$\frac{\mathrm{d}Q^{\text{new}}}{\mathrm{d}t} = 1.5\left(\frac{\mathrm{d}U_{s\alpha}^{\text{lag}}}{\mathrm{d}t} i_{v\alpha} + U_{s\alpha}^{\text{lag}}\frac{\mathrm{d}i_{v\alpha}}{\mathrm{d}t} + \frac{\mathrm{d}U_{s\beta}^{\text{lag}}}{\mathrm{d}t} i_{v\beta} + U_{s\beta}^{\text{lag}}\frac{\mathrm{d}i_{v\beta}}{\mathrm{d}t}\right) \tag{18}$$

由式(9)可知

$$\begin{cases} \dfrac{\mathrm{d}U_{s\alpha}^{\text{lag}}}{\mathrm{d}t} = \dfrac{\mathrm{d}U_{s\alpha}^{\text{lag}+}}{\mathrm{d}t} + \dfrac{\mathrm{d}U_{s\alpha}^{\text{lag}-}}{\mathrm{d}t} = \omega U_{s\alpha} \\ \dfrac{\mathrm{d}U_{s\beta}^{\text{lag}}}{\mathrm{d}t} = \dfrac{\mathrm{d}U_{s\beta}^{\text{lag}+}}{\mathrm{d}t} + \dfrac{\mathrm{d}U_{s\beta}^{\text{lag}-}}{\mathrm{d}t} = \omega U_{s\beta} \end{cases} \tag{19}$$

将式(3)、式(4)、式(19)代入式(17)、式(18)，可以得出基于柔性瞬时功率模型的 MMC 功率模型为

$$\frac{\mathrm{d}\boldsymbol{W}}{\mathrm{d}t} = \boldsymbol{X}\boldsymbol{W} + \boldsymbol{Y}\boldsymbol{U} + \boldsymbol{Z} \tag{20}$$

式中：$\boldsymbol{W} = [P \quad Q^{\text{new}}]^{\mathrm{T}}$ 为状态变量，即受控量；$\boldsymbol{U} = \left[U_{\text{diff}\alpha} \quad U_{\text{diff}\beta}\right]^{\mathrm{T}}$ 为控制变量，即输出变量；$\boldsymbol{X} = \begin{bmatrix} -\dfrac{R}{L} & -\omega \\ \omega & -\dfrac{R}{L} \end{bmatrix}$、$\boldsymbol{Y} = \dfrac{3}{2L}\begin{bmatrix} U_{s\alpha} & U_{s\beta} \\ U_{s\alpha}^{\text{lag}} & U_{s\beta}^{\text{lag}} \end{bmatrix}$ 为系数矩阵；$\boldsymbol{Z} = -\dfrac{3}{2L}\cdot\begin{bmatrix} |U_{\mathrm{s}}|^2 \\ U_{s\alpha}^{\text{lag}} U_{s\alpha} + U_{s\beta}^{\text{lag}} U_{s\beta} \end{bmatrix}$ 为扰动变量。

3　基于改进 SMC 的直接功率控制策略

MMC 的直接功率控制策略以有功功率和无功功率为控制对象，通常采用 PI 控制方法，PI 控制虽然可以在稳态下取得较好的控制效果，但其控制结构较为固定，控制性能在参数扰动等暂态工况下表现较差，因此，本文提出一种改进 SMC 策略，并将其应用于 MMC 的基于柔性瞬时功率模型的直接功率控制策略中。

SMC 与常规 PI 控制的根本区别在于控制的不连续性，即控制系统的“结构”是根据其当前的状态有目的地不断变化，迫使系统状态沿预先设计好的状态轨迹作小幅度、高频率的上下运动，直至停留在平衡点，如图 2 所示，系统状态进入滑模面，并在滑模面上运动的过程称为滑动模态。当系统状态进入滑动模态时，其对外界扰动及参数变化不敏感，能够按照既定的设计在滑模面上进行运动，精准实现控制目标，因此，SMC 相比于其他常规控制方法(如 PI 控制策略)具有更佳的响应速度和鲁棒性。

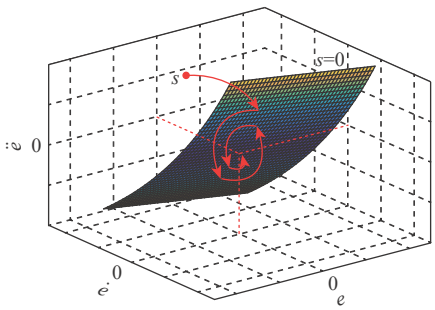


图2 SMC的响应过程

Fig. 2 Response process of SMC

然而，传统SMC策略一般采用线性滑模面和指数趋近律方法，在滑模面附近存在运动惯性，易导致受控量出现抖振现象，影响控制效果。为进一步削弱抖振现象，提高控制精度，减小稳态误差，本文提出一种改进SMC策略，并应用于基于柔性瞬时功率模型的直接功率控制策略。

3.1 改进SMC方法

1）积分滑模面。传统的线性滑模面是对误差函数进行线性化处理，滑模函数的变化速率无约束，收敛速度较慢。因此，本文引入积分滑模面函数，在线性处理的基础上添加误差函数的积分项，减少滑模面函数的切换速率，提升稳态精度，有效削弱抖振现象。积分系数k_s(k_s>0)可以调节积分滑模面函数的变化速率。积分滑模面函数表示为

$$s(x)=e+k_s\int_0^t e\mathrm{d}t \tag{21}$$

式中：e为控制误差。

2）改进指数趋近律。指数趋近律作为一种经典的滑模控制函数，可以优化系统状态在趋近滑模面时的运动轨迹，加快系统状态到达滑模面的速率，但因其表达式中存在等速项，导致系统状态在滑模面附近收敛特性较差，引发一定程度的抖振现象[28]。因此，本文在传统指数趋近律的基础上进行优化，提出一种改进指数趋近律，使其更适用于MMC的直接功率控制模型。

$$\dot{s}=-\varepsilon|e|^p\mathrm{sat}(s)-ks \tag{22}$$

式中：ε和k为控制系数，一般取正常数；sat(s)为饱和函数，其表达式为

$$\mathrm{sat}(s)=\begin{cases}1 & s>\Delta\\ ks & |s|\leqslant\Delta,k=1/\Delta\\ -1 & s<-\Delta\end{cases} \tag{23}$$

饱和函数的控制思想为：设计一个边界层Δ，在边界层之外，采用切换控制sign(s)，使系统状态快速趋近于滑动模态；在边界层之内，采用反馈控制，降低滑模面附近切换控制时引发的抖振现象。

如式(22)所示，本文将控制误差的幂函数引入改进指数趋近律的设计，使系统状态的运动速率与控制误差产生强关联。当系统状态远离滑模面时，控制误差较大，系统状态的趋近速率提高，趋近效应显著增强；当系统状态接近滑模面时，控制误差较小，系统状态趋近速率随之降低，减少了系统状态的运动惯性，使系统状态在滑模面上实现“软着陆”，在提高趋近速度的同时有效抑制滑模趋近过程中产生的抖振现象。同时，将传统切换函数sign(s)替换为饱和函数sat(s)，减少在滑模面上下界运动时的阶跃变换，进一步削弱抖振现象。因此，本文所提改进指数趋近律有效兼顾了趋近速率和抖振削弱的需求。

3）稳定性分析。为确保所提改进SMC策略的稳定性，本文应用李雅普诺夫第二稳定性方法[29]进行稳定性分析。对于式(22)所示的改进SMC方法，可构造一个李雅普诺夫函数为

$$V=Hs^2 \tag{24}$$

式中：H为稳定系数，取正常数。

对于本文所提的改进SMC策略，s代表功率跟踪误差及其积分项的和，当功率误差的平方对时间的导数为负定，则代表MMC输出的有功功率和无功功率可以实现控制目标。对式(24)求取对时间的导数，可得

$$\dot{V}=\frac{\partial V}{\partial s}\cdot\frac{\mathrm{d}s}{\mathrm{d}t}=2H\left[-ks^2-s\varepsilon|e|^p\mathrm{sat}(s)\right] \tag{25}$$

结合式(24)和式(25)可以看出，$V=Hs^2$为正定，而其导数$\dot{V}$在$s\neq0$的情况下均为负定，满足李雅普诺夫第二稳定性方法的判据，可以证明所提改进SMC方法满足渐进稳定条件，可实现功率控制目标。

3.2 基于改进SMC的直接功率控制器设计

根据柔性瞬时功率状态方程式(17)，设定控制目标分别为$P\rightarrow P_{\mathrm{ref}}$、$Q\rightarrow Q_{\mathrm{ref}}$，令$e_1=P_{\mathrm{ref}}-P$，$e_2=Q_{\mathrm{ref}}-Q$，按照积分滑模面的设计方法，分别设计切换函数为

$$s_1=e_1+\int_0^t k_{s1}e_1\mathrm{d}t \tag{26}$$

$$s_2=e_2+\int_0^t k_{s2}e_2\mathrm{d}t \tag{27}$$

式中：k_{s1}、k_{s2}均为正常数。

结合式(22)的改进指数趋近律的设计方法可得

$$\dot{s}_1=\dot{e}_1+k_{s1}e_1=\dot{P}_{\mathrm{ref}}-\dot{P}+k_{s1}(P_{\mathrm{ref}}-P)=-\varepsilon_1|e_1|^p\mathrm{sat}(s_1)-k_1s_1 \tag{28}$$

$$\dot{s}_2=\dot{e}_2+k_{s2}e_2=\dot{Q}_{\mathrm{ref}}-\dot{Q}+k_{s2}(Q_{\mathrm{ref}}-Q)=-\varepsilon_2|e_2|^q\mathrm{sat}(s_2)-k_2s_2 \tag{29}$$

式中：k_1、k_2、k_{s1}、k_{s2}、ε_1、ε_2、p、q均为正常数。

结合式(20)、式(27)、式(28)，可得直接功率控制系统中控制变量差模电压的指令值U_{diffdref}、U_{diffqref}的表达式为

$$\boldsymbol{U}_{\mathrm{ref}}=\boldsymbol{Y}^{-1}\boldsymbol{V}-\boldsymbol{Y}^{-1}\boldsymbol{X}\boldsymbol{W}-\boldsymbol{Y}^{-1}\boldsymbol{Z} \tag{30}$$

式中：

$$\boldsymbol{U}_{\text{ref}}=\left[U_{\text{diff}\alpha\text{ref}}\quad U_{\text{diff}\beta\text{ref}}\right]^{\text{T}}$$

$$\boldsymbol{V}=\begin{bmatrix}\varepsilon_1|e_1|^p\text{sat}(s_1)+k_1s_1+k_{\text{s}1}e_1\\ \varepsilon_2|e_2|^q\text{sat}(s_2)+k_2s_2+k_{\text{s}2}e_2\end{bmatrix}$$

3.3　系统整体控制策略

图 3 为本文所提基于柔性瞬时功率模型的 MMC-HVDC 的改进滑模直接功率控制策略结构框图。图中：NLM 表示最近电平调制。如图 3 所示，通过柔性瞬时功率模型实时计算出有功功率和柔性无功功率偏差，经改进滑模直接功率控制器（式(29)）得到控制变量，即差模电压的 $\alpha\beta$ 轴分量的参考值 $U_{\text{diff}\alpha\text{ref}}$、$U_{\text{diff}\beta\text{ref}}$，经坐标变换，结合环流抑制控制器及最近电平调制策略，得到各子模块中 IGBT 的触发信号，输出恒定的有功功率和无功功率。

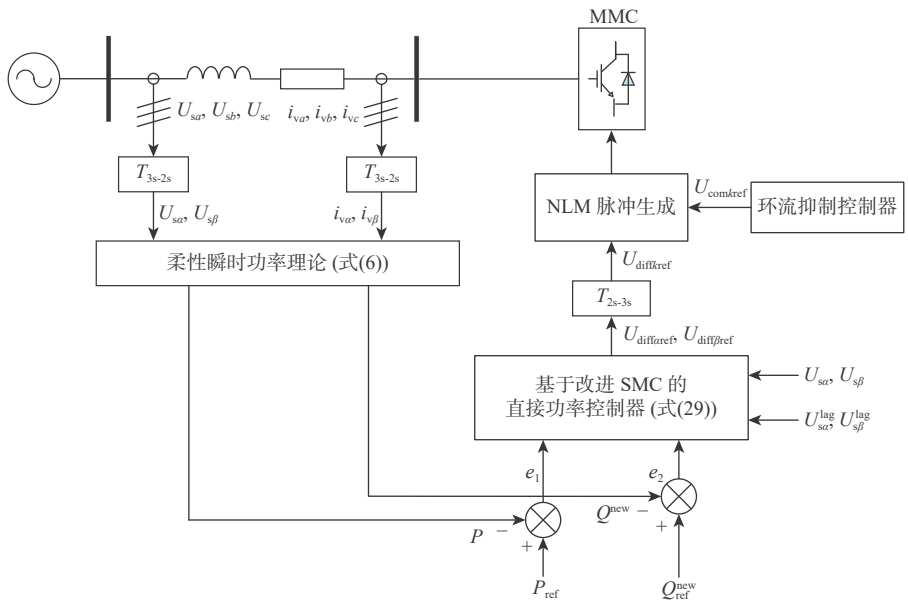


图 3　本文控制系统结构框图

Fig. 3　Block diagram of the proposed control strategy

4　仿真验证

4.1　仿真模型

本文在 PSCAD/EMTDC 平台中搭建了图 4 所示的四端 MMC-HVDC 系统。四端系统采用主从控制策略[30]，MMC4 为主控站，控制系统直流电压，MMC1～MMC3 为从控站，采用定功率控制，其中，MMC2 采用本文所提直接功率控制策略，其仿真参数如表 1 所示。功率正方向定义为交流侧流向直流侧。

为验证本文所提基于柔性瞬时功率模型的 MMC-HVDC 的改进滑模直接功率控制策略的有效性，在电网对称、非对称条件、参数扰动条件下，分别在 MMC2 中应用不同功率定义和不同控制器进行对比性仿真实验，如图 5～图 8 所示。

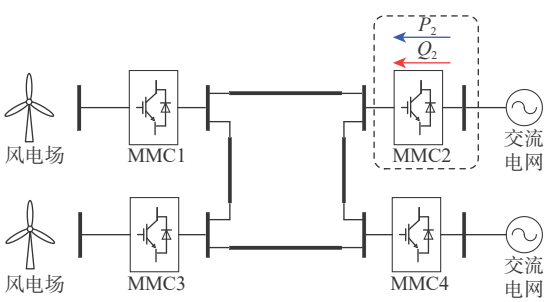


图 4　四端 MMC-HVDC 系统

Fig. 4　Four-terminal MMC-HVDC system

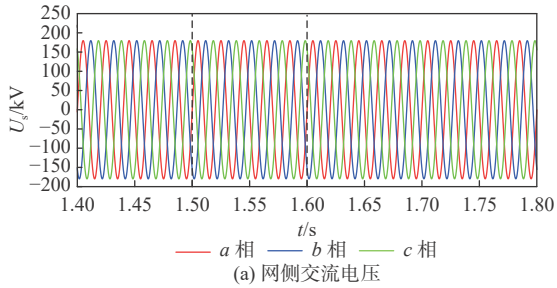
表 1　MMC2 换流站参数

Table 1　Parameters of MMC2 converter station

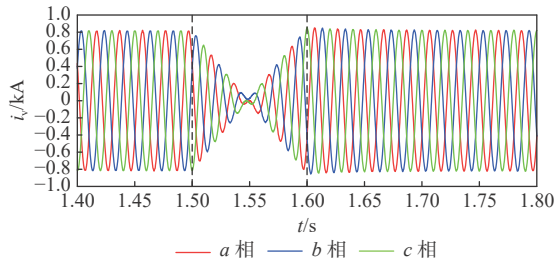
参数	数值
系统直流电压/kV	±200
交流线电压/kV	220
子模块电容/μF	8 950
桥臂电感/mH	41.5
子模块数量/个	120

4.2　对称电网稳态运行情况

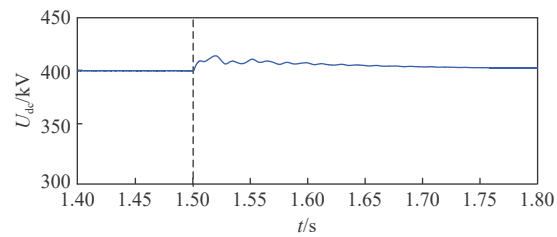
为验证 MMC2 在应用本文所提改进滑模直接功率控制策略后，在不同功率下的稳态及动态性能，本节分析了 MMC2 在电网对称条件下的稳态控制性能，并针对有功阶跃工况进行仿真。仿真过程中，MMC2 有功功率和无功功率的初始参考值分别设定为−200 MW 和 0 Mvar（Mvar 是兆乏，又称兆乏尔，是一种表示无功功率的单位）。在 1.50 s 时设置潮流反转，设置 MMC2 的有功功率指令值由−200 MW 线性变化为 200 MW。从图 5 可以看出，改进滑模直接功率控制策略可以实现有功功率和无功功率参考值的精确跟踪，验证了本文所提控制策略在对



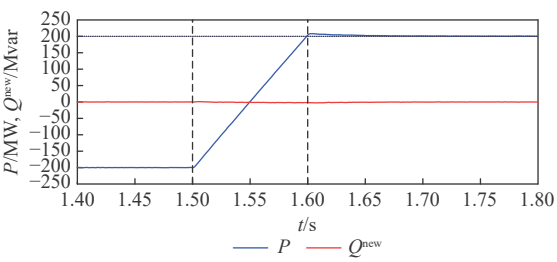
(a) 网侧交流电压



(b) 阀侧交流电流



(c) 系统直流电压



(d) 有功功率和无功功率

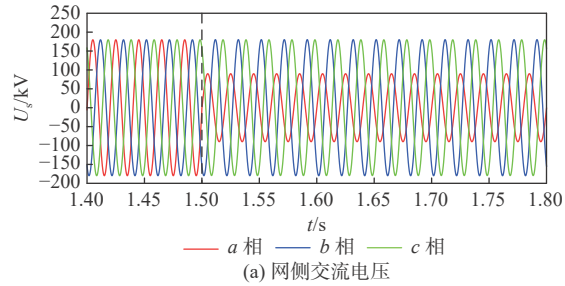
图 5 对称电网条件下 MMC2 的稳态仿真情况

Fig. 5 Steady-state simulation results of MMC2 under balanced grid condition

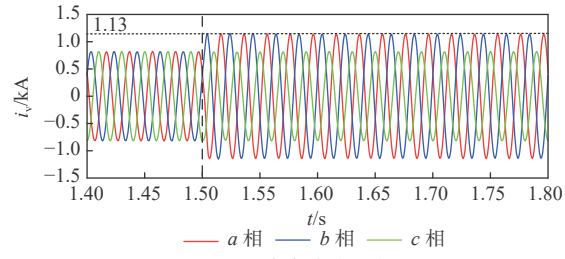
称电网条件下的有效性。在 1.50 s 时，随着有功潮流反转，MMC2 输出的有功功率在变化初期超调量较小，可以精确跟踪功率参考值的变化，同时 MMC2 的阀侧电流趋势平稳，始终保持稳态运行，且系统直流电压在功率波动后逐渐恢复稳定，由此可见，基于改进滑模直接功率控制策略的稳态性能较好。

4.3 非对称电网暂态运行情况

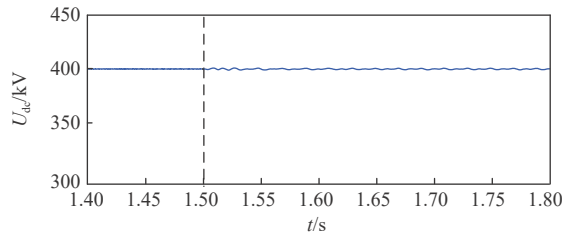
在非对称电网条件下，基于传统瞬时功率模型的直接功率控制策略在抑制有功功率和无功功率的二倍频波动分量方面存在局限性，因此，本节基于柔性瞬时功率模型对 MMC 进行直接功率控制，验证柔性瞬时功率模型的有效性。同时，在对 MMC2 的功率控制过程中分别采用 PI 控制器及改进滑模控制器，对 2 种控制器的控制性能进行比较。仿真过程中，MMC2 的初始参考功率分别设定为−200 MW



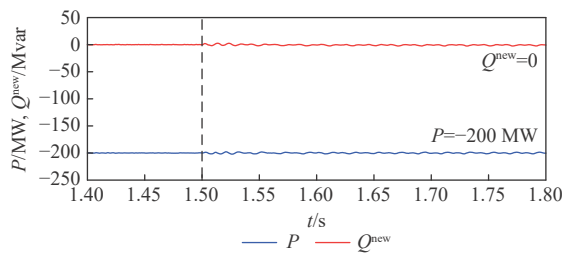
(a) 网侧交流电压



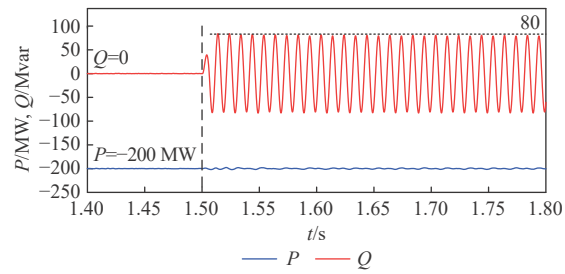
(b) 阀侧交流电流



(c) 四端系统直流电压



(d) 柔性瞬时功率模型运行效果



(e) 传统瞬时功率模型运行效果

图 6 非对称电网条件下 MMC2 的暂态运行情况

Fig. 6 Transient operation results of MMC2 under asymmetric grid condition

和 0 MVar。1.50 s 时，设置 a 相电压出现 50% 的电压跌落，导致交流侧三相电压不对称。

图 6 对比了以抑制非对称电网条件下 MMC2 输出有功功率二倍频振荡为控制目标时，基于不同瞬时功率定义的控制效果。图 6(a)～(d) 为采用本文策略的运行情况，图 6(e) 为采用传统瞬时功率定义时的运行情况。1.50 s 时，a 相电压发生电压跌落，三相电网电压出现不对称，虽然基于传统瞬时

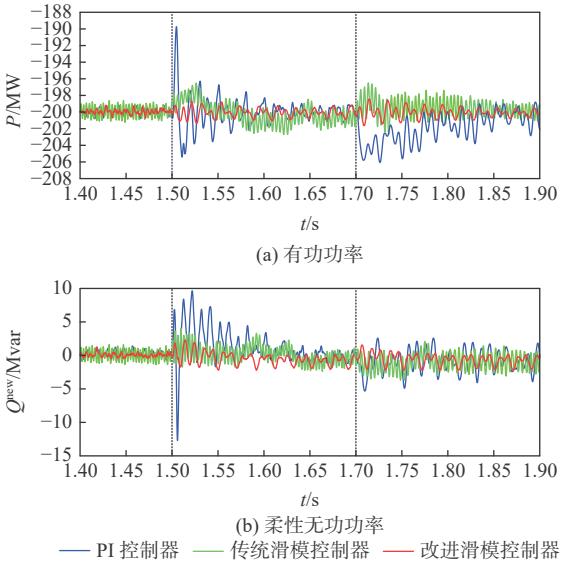


图 7　PI 控制器、传统滑模控制器与改进滑模控制器的动态和鲁棒性控制效果对比

Fig. 7　Comparison of control results of dynamic power responses and robustness using PI controller, conventional SMC and improved SMC

功率模型控制下的有功功率依然保持恒定，但其输出的无功功率中仍存在幅值为 80 MW 的二倍频振荡，说明传统瞬时功率模型无法同时消除电网不对称状态下有功功率和无功功率中的二倍频振荡。相反，基于柔性瞬时功率模型控制下的 MMC2 输出的有功功率和无功功率可以准确地跟踪参考值，并保持阀侧电网电流无畸变，说明基于柔性瞬时功率模型的直接功率控制可以同时消除有功功率和无功功率中的二倍频波动分量。由图 6(c) 可知，由于有功功率控制稳定，a 相电压跌落后仅对四端系统直流电压造成细微波动，在非对称电网条件下，本文所提改进控制策略对四端系统整体影响较小。综上，柔性瞬时无功功率相较于传统瞬时无功功率更适合作为 MMC 恒功率控制中的受控变量。

图 7 对比了 PI 控制器、传统滑模控制器、改进滑模控制器在电网不对称条件下对于直接功率控制的动态控制响应能力和鲁棒性，其中，a 相电压跌落时间与图 5 相同。在相同的系统参数条件下，MMC2 分别采用 PI 控制器及改进滑模控制器，在电压跌落和参数摄动情况下，对其响应性能和鲁棒性能进行横向对比。可以看出，在稳态运行阶段，传统滑模控制器的抖振较为明显，1.50 s 时 a 相电压发生跌落瞬间，PI 控制器所控有功功率的最大偏差在 12 MW 左右，有功超调量接近 6%，0.1 s 后才逐渐进入稳态运行阶段，传统滑模控制器虽然最大偏差缩小至 4 MW，但其抖振现象明显，1.60 s 逐渐稳定后仍存在明显抖振，而改进滑模控制器所控有功的最大偏差仅为 1 MW，超调量在 1% 左右，且 0.05 s 内达到稳态运行，抖振现象明显削弱，说明改

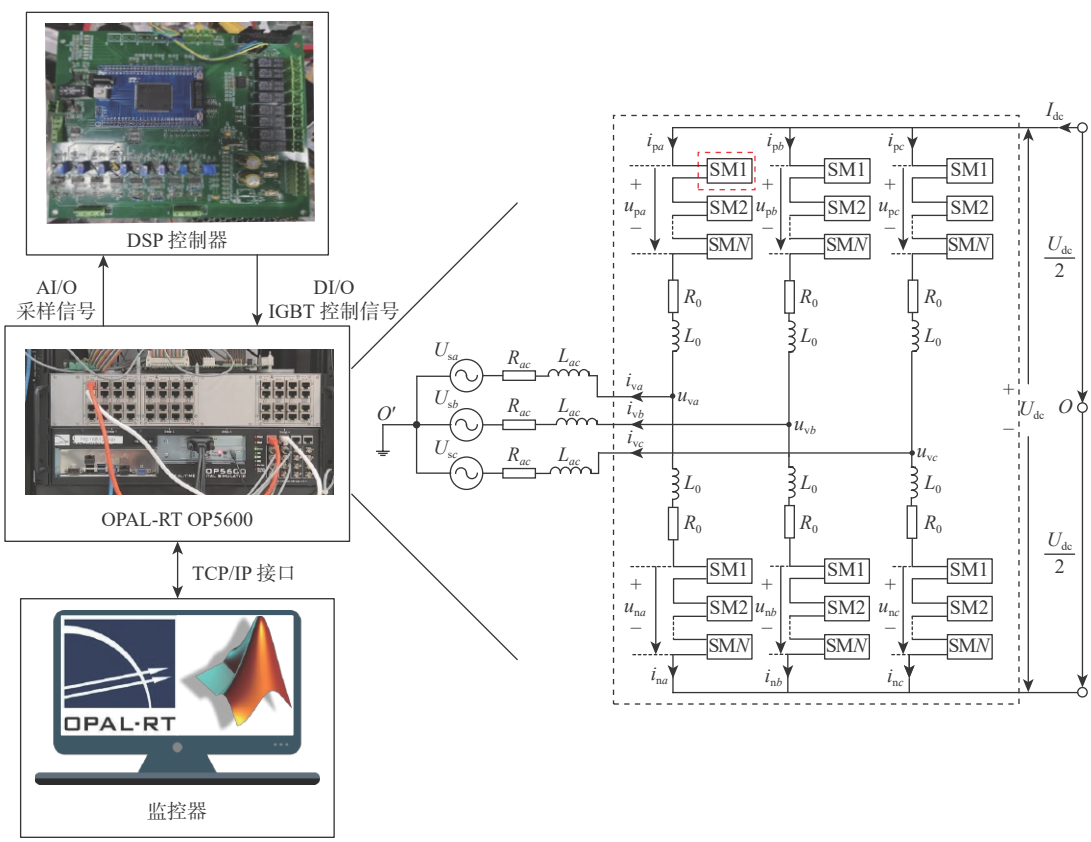


图 8　MMC 硬件在环实验平台

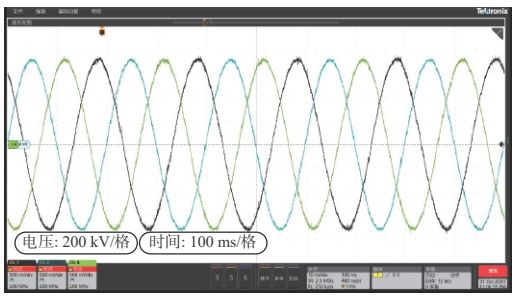
Fig. 8　Hardware-in-the-loop experiment topology of MMC

进 SMC 策略下的有功功率和无功功率的响应速度更快,抖振更小。1.70 s 时,设置桥臂电感值由 41.5 mH 增大至 48 mH,模拟了 MMC 实际运行时的随机参数摄动。由图 7 所示,受桥臂电感值的扰动,对 PI 控制器的影响较大,有功功率瞬间波动至 −207 MW,0.15 s 后才逐渐恢复至正常值,且稳态误差增加,传统滑模控制器虽然受参数扰动波动较小,但抖振频率和幅度较大,最大波幅在 4 MW 左右,而改进滑模控制器所控有功功率波动范围在 2 MW 以内,在 0.1 s 内恢复至稳态运行,有功功率和无功功率依然能够准确跟踪指令值,说明改进滑模控制器在桥臂电感参数变化的干扰下能够保持良好的鲁棒性。因此,当基于柔性瞬时功率模型的直接功率控制器采用改进 SMC 策略时,其有功功率和无功功率动态响应能力更强,且对随机参数摄动的鲁棒性较好,说明改进 SMC 策略在非对称电网条件下可以充分发挥柔性瞬时功率模型的优势。

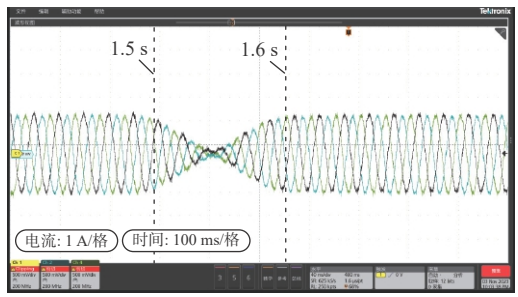
5 实验验证

为验证本文所提基于柔性瞬时功率模型的 MMC-HVDC 改进滑模直接功率控制策略的有效性,本节搭建了图 8 所示的 MMC 硬件在环实验平台。可知,MMC 主电路由 OPAL-RT OP5600 实时仿真系统实现,其运行频率为 10^5 Hz,控制器由数字信号处理器 (digital signal processor, DSP) 实现。其中,OPAL-RT 系统与 DSP 控制器通过 OP8665 作为转接板和采样板进行连接,DSP 控制器的输入为 OPAL-RT 系统输出的交流侧及子模块电压和电流模拟信号,DSP 控制器的输出为根据本文所提改进滑模直接功率控制策略计算出的 IGBT 触发数字信号,用于控制 MMC 中各子模块内 IGBT 的开断,从而实现 OPAL-RT 系统与 DSP 控制器的外闭环连接。将四端 MMC-HVDC 系统的其他元件及其相关控制器嵌入 OPAL-RT OP5600 仿真器及 DSP 控制器中,其余参数同仿真验证部分相同。

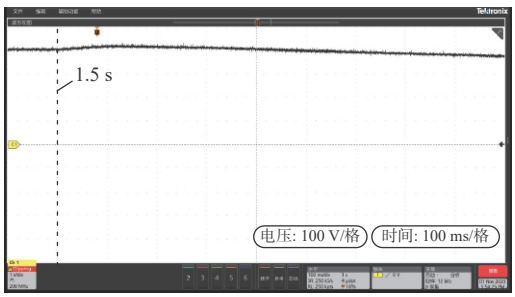
1) 实验 1。选取与图 5 相同的运行工况,MMC2 有功功率和无功功率的初始参考值分别设定为 −200 MW 和 0 MVar。在 1.5 s 时设置潮流反转,设置 MMC2 的有功功率指令值由−200 MW 线性变化为 200 MW。如图 9 所示,稳态情况下,有功功率和无功功率均能够准确跟踪参考值,阀侧电流与系统



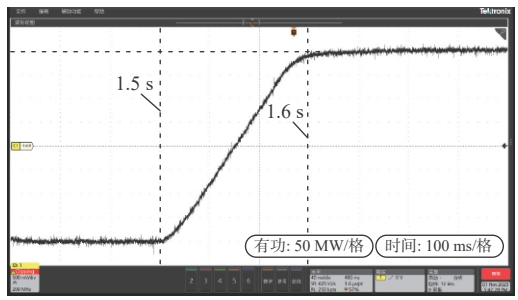
(a) 网侧交流电压



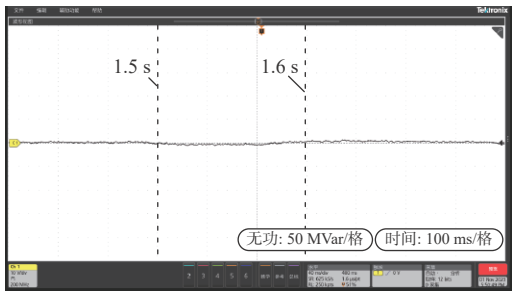
(b) 阀侧交流电流



(c) 系统直流电压



(d) 有功功率



(e) 无功功率

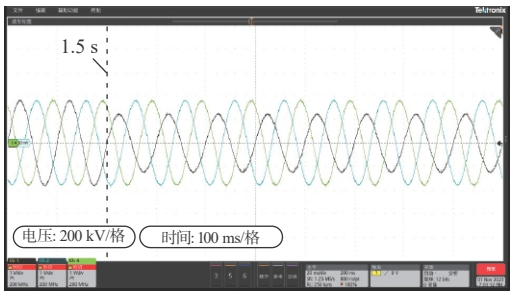
图 9 对称电网条件下 MMC2 的稳态实验结果

Fig. 9 Steady-state experimental results of MMC2 under balanced grid condition

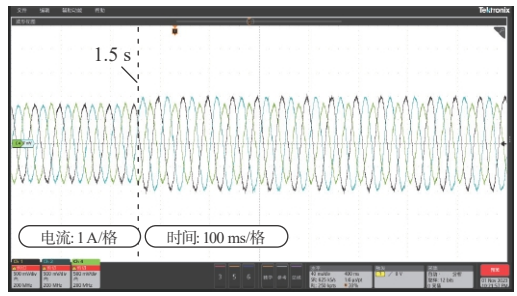
直流电压趋势平稳。通过上述实验结果可知，本文所提改进控制策略可以在电压对称时准确控制有功功率和柔性无功功率。

2）实验 2。选取与图 6 相同的运行工况，MMC2 的初始参考功率分别设定为−200 MW 和 0 MVar。1.5 s 时，设置 A 相电压出现 50% 的电压跌落，导致交流侧三相电压不对称。如图 10 所示，交流电压发生不对称后，系统直流电压较为稳定，有功功率和柔性无功功率均可以准确地跟踪参考值，并保持阀侧电网电流无畸变。通过上述实验结果可知，本文所提改进控制策略可以在电压不对称时准确控制有功功率和无功功率。

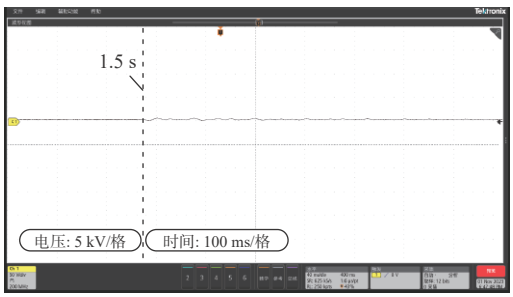
需要注意的是，本文实验部分与仿真部分的运行工况完全相同，所控有功功率和柔性无功功率能跟始终稳定跟踪参考值，仿真结果与实验结果结论相同，进一步说明了本文方法的有效性。



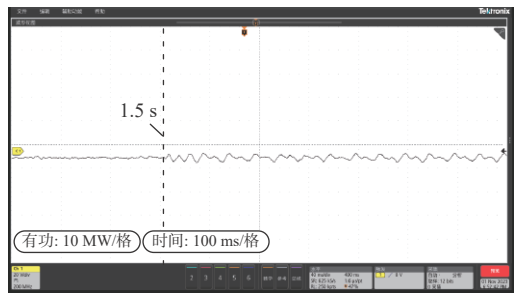
(a) 网侧交流电压



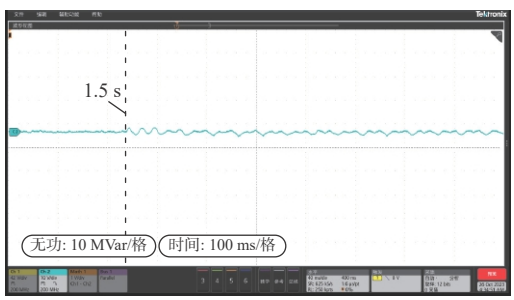
(b) 阀侧交流电流



(c) 系统直流电压



(d) 有功功率



(e) 无功功率

图 10　非对称电网条件下 MMC2 的暂态实验结果

Fig. 10　Transient operation experimental results of MMC2 under asymmetric grid condition

6　结　论

本文提出了基于柔性瞬时功率模型的 MMC-HVDC 改进滑模直接功率控制策略，该策略在柔性瞬时功率模型的基础上，推导出 MMC 的直接功率控制模型，并结合改进 SMC 策略优化控制过程。通过理论分析、模型仿真、实验验证，得出以下结论：

1）柔性瞬时功率模型可以在非对称电网条件下同时抑制有功功率和无功功率的二倍频波动，避免了功率补偿项，优化了功率控制策略。

2）在非对称电网及随机参数扰动条件下，改进 SMC 策略的动态响应能力和鲁棒性较 PI 控制策略更强，更适用于柔性直接功率控制策略的实现。

参考文献（References）

[1] MUTHAVARAPU A K, BISWAS J, BARAI M. An efficient sorting algorithm for capacitor voltage balance of modular multilevel converter with space vector pulsewidth modulation[J]. IEEE Transactions on Power Electronics, 2022, 37(8): 9254-9265.

[2] RAJU M N, SREEDEVI J, MANDI R P, et al. Modular multilevel converters technology: a comprehensive study on its topologies, modelling, control and applications[J]. IET Power Electronics, 2019, 12(2): 149-169.

[3] MA K, HE B, XIN X K, et al. Capacitor voltage control for mission profile emulator of submodule in modular multilevel converter[J].

IEEE Transactions on Power Electronics, 2021, 36(11): 12355-12364.

[4] NAMI A, LIANG J Q, DIJKHUIZEN F, et al. Modular multilevel converters for HVDC applications: review on converter cells and functionalities[J]. IEEE Transactions on Power Electronics, 2015, 30(1): 18-36.

[5] 管敏渊, 徐政. 模块化多电平换流器型直流输电的建模与控制[J]. 电力系统自动化, 2010, 34(19): 64-68.
GUAN M Y, XU Z. Modeling and control of modular multilevel converter in HVDC transmission[J]. Automation of Electric Power Systems, 2010, 34(19): 64-68(in Chinese).

[6] VASILADIOTIS M, CHERIX N, RUFER A. Accurate capacitor voltage ripple estimation and current control considerations for grid-connected modular multilevel converters[J]. IEEE Transactions on Power Electronics, 2014, 29(9): 4568-4579.

[7] 李仪, 郭宏, 蔚永强. 永磁同步电动机大惯量负载的滑模变结构控制[J]. 北京航空航天大学学报, 2007, 33(10): 1208-1211.
LI Y, GUO H, YU Y Q. Sliding mode control of permanent magnet synchronous motor with great inertia load[J]. Journal of Beijing University of Aeronautics and Astronautics, 2007, 33(10): 1208-1211 (in Chinese).

[8] LI B, XIE Y Z, WEN W J, et al. Improved sliding-mode control for MMC in DC power system[J]. IET Renewable Power Generation, 2020, 14(15): 3035-3042.

[9] WANG S D, DRAGICEVIC T, GAO Y, et al. Machine learning based operating region extension of modular multilevel converters under unbalanced grid faults[J]. IEEE Transactions on Industrial Electronics, 2021, 68(5): 4554-4560.

[10] FREYTES J, LI J Q, DE PRÉVILLE G, et al. Grid-forming control with current limitation for MMC under unbalanced fault ride-through[J]. IEEE Transactions on Power Delivery, 2021, 36(3): 1914-1916.

[11] 孔明, 汤广福, 贺之渊, 等. 不对称交流电网下 MMC-HVDC 输电系统的控制策略[J]. 中国电机工程学报, 2013, 33(28): 41-49.
KONG M, TANG G F, HE Z Y, et al. A control strategy for modular multilevel converter based HVDC of unbalanced AC systems[J]. Proceedings of the CSEE, 2013, 33(28): 41-49(in Chinese).

[12] 陈海荣, 徐政. 基于同步坐标变换的 VSC-HVDC 暂态模型及其控制器[J]. 电工技术学报, 2007, 22(2): 121-126.
CHEN H R, XU Z. Transient model and controller design for VSC-HVDC based on synchronous reference frame[J]. Transactions of China Electrotechnical Society, 2007, 22(2): 121-126(in Chinese).

[13] 张建坡, 赵成勇, 敬华兵. 比例谐振控制器在 MMC-HVDC 控制中的仿真研究[J]. 中国电机工程学报, 2013, 33(21): 53-62.
ZHANG J P, ZHAO C Y, JING H B. Simulating research of proportional resonant controllers in MMC-HVDC[J]. Proceedings of the CSEE, 2013, 33(21): 53-62(in Chinese).

[14] SHANG L, SUN D, HU J. Sliding-mode-based direct power control of grid-connected voltage-sourced inverters under unbalanced network conditions[J]. IET Power Electronics, 2011, 4(5): 570-579.

[15] 年珩, 程鹏. 电网电压不平衡时 PWM 整流器的谐振直接功率控制策略[J]. 电工技术学报, 2013, 28(11): 86-94.
NIAN H, CHENG P. Resonant based direct power control strategy for PWM rectifier under unbalanced grid voltage condition[J]. Transactions of China Electrotechnical Society, 2013, 28(11): 86-94 (in Chinese).

[16] 张迪, 魏艳君, 马利轩, 等. 不平衡电网电压下基于滑模变结构控制的双馈风电系统网侧变流器控制策略[J]. 电工技术学报, 2015, 30(10): 266-275.
ZHANG D, WEI Y J, MA L X, et al. Sliding-mode control for grid-side converters of DFIG-based wind-power generation system under unbalanced grid voltage conditions[J]. Transactions of China Electrotechnical Society, 2015, 30(10): 266-275(in Chinese).

[17] 梁营玉, 杨奇逊, 刘建政, 等. 电网电压不平衡时 MMC-HVDC 的无差拍直接功率控制[J]. 电工技术学报, 2015, 30(15): 15-25.
LIANG Y Y, YANG Q X, LIU J Z, et al. Deadbeat direct power control for MMC-HVDC under unbalanced grid voltages[J]. Transactions of China Electrotechnical Society, 2015, 30(15): 15-25(in Chinese).

[18] 刘子文, 苗世洪, 范志华, 等. 不平衡电网电压下柔性直流输电系统功率滑模补偿策略[J]. 电工技术学报, 2018, 33(14): 3296-3305.
LIU Z W, MIAO S H, FAN Z H, et al. Power sliding mode compensation strategy of VSC-HVDC under unbalanced grid voltage[J]. Transactions of China Electrotechnical Society, 2018, 33(14): 3296-3305(in Chinese).

[19] SUH Y, LIPO T A. Modeling and analysis of instantaneous active and reactive power for PWM AC/DC converter under generalized unbalanced network[J]. IEEE Transactions on Power Delivery, 2006, 21(3): 1530-1540.

[20] ZHANG Y C, QU C Q. Model predictive direct power control of PWM rectifiers under unbalanced network conditions[J]. IEEE Transactions on Industrial Electronics, 2015, 62(7): 4011-4022.

[21] 梁营玉. 不对称电网电压下基于降阶矢量谐振器的 MMC-HVDC 直接功率控制策略[J]. 中国电机工程学报, 2017, 37(24): 7294-7303.
LIANG Y Y. Direct power control strategy based on reduced order vector resonant controller for MMC-HVDC under unbalanced grid voltages[J]. Proceedings of the CSEE, 2017, 37(24): 7294-7303(in Chinese).

[22] 刘勇智, 鄯成龙, 林博闻, 等. 基于自适应二阶终端滑模的 SRM 直接转矩控制[J]. 北京航空航天大学学报, 2018, 44(10): 2043-2051.
LIU Y Z, SHAN C L, LIN B W, et al. Direct torque control of switched reluctance motor based on adaptive second-order terminal sliding mode[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(10): 2043-2051(in Chinese).

[23] 李慧洁, 蔡远利. 基于双幂次趋近律的滑模控制方法[J]. 控制与决策, 2016, 31(3): 498-502.
LI H J, CAI Y L. Sliding mode control with double power reaching law[J]. Control and Decision, 2016, 31(3): 498-502(in Chinese).

[24] 高洁, 王华宇, 徐萌. 开关磁阻电机的分数阶终端滑模控制[J]. 电机与控制学报, 2022, 26(2): 142-150.
GAO J, WANG H Y, XU M. Improved terminal sliding mode control strategy of switched reluctance motors based on fractional calculus[J]. Electric Machines and Control, 2022, 26(2): 142-150(in Chinese).

[25] 李逃昌. 基于新型趋近律的非线性积分滑模控制方法[J]. 控制工程, 2019, 26(11): 2031-2035.
LI T C. Nonlinear integral sliding mode control method based on new reaching law[J]. Control Engineering of China, 2019, 26(11): 2031-2035(in Chinese).

[26] 刘金琨. 滑模变结构控制 MATLAB 仿真: 基本理论与设计方法[M]. 4 版. 北京: 清华大学出版社, 2019: 38-39.
LIU J K. Sliding mode control design and MATLAB simulation: the basic theory and design method[M]. 4th ed. Beijing: Tsinghua University Press, 2019: 38-39(in Chinese).

[27] AKAGI H, KANAZAWA Y, NABAE A. Instantaneous reactive power compensators comprising switching devices without energy storage components[J]. IEEE Transactions on Industry Applications, 1984, 20(3): 625-630.

[28] 杨锁昌, 张宽桥, 陈鹏. 带攻击角度约束的自适应终端滑模导引律[J]. 北京航空航天大学学报, 2016, 42(8): 1566-1574.
YANG S C, ZHANG K Q, CHEN P. Adaptive terminal sliding mode guidance law with impact angle constraint[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(8): 1566-1574(in Chinese).

[29] 潘菲, 朱宏玉. 航天器非奇异自适应终端滑模姿轨联合控制[J]. 北京航空航天大学学报, 2020, 46(7): 1354-1362.
PAN F, ZHU H Y. Spacecraft non-singular adaptive terminal sliding mode attitude-orbit coupling control[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(7): 1354-1362(in Chinese).

[30] LU W X, OOI B T. Optimal acquisition and aggregation of offshore wind power by multiterminal voltage-source HVDC[J]. IEEE Transactions on Power Delivery, 2003, 18(1): 201-206.

Improved sliding-mode direct power control strategy for MMC-HVDC under asymmetrical grid state

GUAN Tianyi[1, 2], WANG Yibo[1], WANG Rui[1, *], ZHAO Haiqiao[1], ZHENG Wenjing[3], SUN Qiuye[1]

(1. College of Information Science and Engineering, Northeastern University, Shenyang 110819, China;
2. State Grid Liaoning Electric Power Co., Ltd., Shenyang Power Supply Company, Shenyang 110811, China;
3. State Grid Liaoning Electric Power Co., Ltd., Marketing Service Center, Shenyang 110168, China)

Abstract: The modular multilevel converter based high voltage direct current (MMC-HVDC) is a dynamic power balancing system. The control system of MMC generally adopts a dual closed-loop vector control strategy based on the traditional instantaneous power model under an asymmetric grid state, which has a complex control structure and low control accuracy. This paper introduces a flexible instantaneous power model and establishes a general power equation with active power and new reactive power as control objects. In order to eliminate the twice grid-frequency ripples in both active and reactive power under asymmetric grid states, an enhanced sliding-mode MMC-HVDC direct power control strategy based on the new instantaneous power model is proposed. This approach combines the improved sliding-mode control method with the flexible instantaneous power model. Furthermore, it omits the inner-loop controller and power compensation terms while optimizing the control structure. According to simulation results, the suggested approach can more effectively take advantage of the benefits of the flexible instantaneous power model because it has superior dynamic responsiveness, control precision, and robustness under operating conditions including asymmetric grid state and parameter perturbation.

Keywords: modular multilevel converter based high voltage direct current; asymmetrical grid; sliding mode control; flexible instantaneous power model; robust control

Received: 2023-05-25; **Accepted**: 2023-11-15; **Published Online**: 2023-12-06 09:39
URL: link.cnki.net/urlid/11.2625.V.20231205.1012.001
Foundation items: National Natural Science Foundation of China (U20A20190,62073065)
* **Corresponding author**. E-mail: wangrui@ise.neu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0357

电力电子变换器拓扑同构与拓扑等效辨析

莫李平[1,2]，陈桂鹏[1,2,*]，李志强[3]

(1. 厦门大学深圳研究院，深圳 518000；2. 厦门大学 航空航天学院，厦门 361000；
3. 中国工程物理研究院 电子工程研究所，绵阳 621000)

摘 要：在电力电子变换器拓扑的推演过程中，通过拓扑等效或拓扑同构来识别结构不同但性能相同的拓扑，以避免重复研究，但拓扑等效与拓扑同构的内涵存在显著区别，容易导致混淆。为明确二者的区别与联系，提出准确识别性能相同拓扑的方法。推导出拓扑同构的必要条件为其邻接矩阵的行列式相等，并进一步推导出拓扑等效的充要条件为元件组成相同且简单回路一一对应；从拓扑子图的角度分析两者关系，发现拓扑同构是拓扑等效的充分非必要条件，而拓扑等效是性能相同拓扑的充要条件。基于此，建议实际应用中优先采用拓扑等效识别性能相同的拓扑。通过案例分析验证了理论的正确性和可行性。同时，提出了一种基于深度优先搜索算法实现等效拓扑自动判定的方法，为快速、准确识别性能相同的变换器拓扑提供了技术支持。

关 键 词：电力电子变换器；同构；等效；图论；深度优先搜索
中图分类号：TM46; O157.5
文献标志码：A **文章编号**：1001-5965(2025)05-1440-09

电力电子变换器依靠半导体功率器件可实现电能的高效率转换与控制，已被广泛应用于日常生活和工业生产[1-2]。为满足日益增长的电能应用需求，新型变换器拓扑的构建受到了国内外的广泛关注[3-8]。随着变换器拓扑推演技术的发展，尤其是计算机辅助技术在变换器拓扑推演中的应用[9-14]，大量的新变换器拓扑开始涌现。然而，基于计算机辅助的推演结果，往往包含性能相同(即所有元件工作的电压和电流完全相等)但结构不同的拓扑。

为避免重复研究性能相同的拓扑，学者们将变换器拓扑性能相同与拓扑同构进行等价，并将变换器拓扑同构的识别问题转化为对应的图的同构判定问题[15-17]。图的同构判定问题是图论研究中的基本问题[18]，在电路拓扑、化学分子拓扑等研究领域均有较为广泛的应用[19-20]。文献 [15] 将变换器拓扑建模为其邻接矩阵，通过对比邻接矩阵的特征值和相似变换矩阵是否为置换矩阵辨识同构拓扑。文献 [16] 提出了电路模拟法，通过建模将图同构判定问题转换为判定对应伴随电路是否相同的问题。文献 [17] 在电路模拟法基础上引入了参考节点，进一步提高了拓扑同构判定的效率。基于拓扑同构识别，文献 [11-12] 从生成的大量单开关直流变换器拓扑中剔除了结构不同但性能相同的拓扑，从甄选保留了性能不同的拓扑。

然而，拓扑同构概念完全类比于图论中的图同构，拓扑同构与拓扑性能相同的关系尚未充分验证。值得注意的是，同样以识别性能相同的变换器拓扑为目的，文献 [21] 提出了拓扑等效识别方法。与拓扑同构直接套用图论的概念不同，拓扑等效建立在变换器元件的电压和电流完全相同的基础上。虽然拓扑等效的识别也需要先将拓扑建模为邻接矩阵，但其通过搜索邻接矩阵得到简单回路进行拓扑等效识别。基于等效拓扑识别方法，文献 [9,14] 在推演出的大量拓扑中剔除了性能相同的拓扑，保

收稿日期：2023-06-14；**录用日期**：2023-09-01；**网络出版时间**：2023-09-13 18:55
网络出版地址：link.cnki.net/urlid/11.2625.V.20230913.1129.001
基金项目：广东省基础与应用基础研究基金 (2022A1515011616)；国家自然科学基金 (51907172)
* **通信作者**. E-mail：cgp2017@xmu.edu.cn

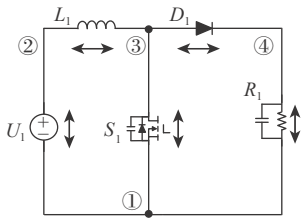
引用格式：莫李平，陈桂鹏，李志强. 电力电子变换器拓扑同构与拓扑等效辨析 [J]. 北京航空航天大学学报，2025，51（5）：1440-1448.
MO L P，CHEN G P，LI Z Q. Discrimination on isomorphism and equivalence of topologies of power electronics converters [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1440-1448 （in Chinese）.

证了推演出的拓扑性能各不相同。

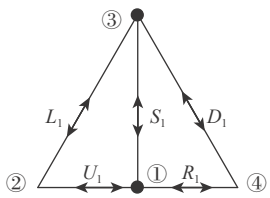
现有研究表明, 目前主要通过拓扑同构和拓扑等效来判定计算机推演的大量拓扑是否具有相同性能。然而, 拓扑同构和拓扑等效的内涵存在差异, 容易导致混淆。因此, 本文以拓扑同构、拓扑等效与拓扑性能相同三者之间的关系为研究重点, 通过理论推导, 明确拓扑同构与拓扑等效的判定条件及其适用范围, 同时揭示三者之间的内在联系。此外, 结合实际需求, 进一步结合深度优先搜索算法, 提出等效拓扑自动判定方法, 以实现性能相同拓扑的高效识别与筛选。

1　图论基本概念

图论是以图为研究对象的数学分支, 图论研究的图是由节点集 V 及连接这些节点的边集 E 组成, 通常表示为 $G(V,E)$[22]。如图 1 所示, 将变换器中的元件视为对应的边, 则变换器可视为对应的图。由于拓扑同构与拓扑等效都与图论密不可分, 有必要对图论的基本概念进行介绍。



(a) 变换器拓扑



(b) 有向有权图

图 1　变换器拓扑与其对应的图

Fig. 1　Converter topology and its corresponding graph

1) 有向有权图。假设图 $G(V,E)$ 由顶点集 $V=\{v_1,v_2,\cdots,v_V\}$和边集 $E=\{e_1,e_2,\cdots,e_E\}$组成, 边的权值 $W=\{w_1,w_2,\cdots,w_W\}$且边有方向, 则图 G 为有向有权图。如图 1 所示, 将变换器中的元件视为有向有权的边, 则变换器可视为对应的有向有权图。

2) 入度和出度。有向图中, 入度是以顶点 v 为终点的有向边的数目, 出度是以顶点 v 为起点的有向边的数目。对应于拓扑中, 节点 v 的入度和出度分别为参考方向流入该节点和流出该节点的元件数量。

3) 子图。对于 2 个图 $G(V, E)$ 和 $G'(V', E')$, 若 $V'\subseteq V$ 且 $E'\subseteq E$, 则称图 G'为图 G 的一个子图。

4) 路径与回路。给定图 $G(V,E)$, 节点 v_x 到节点 v_y 的路径为顶点序列 $\{v_x=v_{i0},v_{i1},v_{i2},\cdots,v_{in}=v_y\}$, 其中, $v_{ik}\in V(0\leqslant k\leqslant n)$, $<v_{ij-1},v_{ij}>\in E(0\leqslant j\leqslant n)$。顶点互不相同的路径称为简单路径。除第 1 个顶点和最后 1 个顶点相同外, 其他顶点互不相同的路径称为简单回路。

5) 连通。在图 $G(V,E)$ 中, 若任意 2 个节点 v_x, $v_y\in V$ 之间都存在从 v_x 到 v_y 的路径, 则图 $G(V,E)$ 是连通的, 称为连通图。

6) 同构[22]。假设 $G(V,E)$ 和 $G'(V',E')$ 是 2 个图, 如果存在一个双射 f: $V\rightarrow V'$, 使得对于任意 2 个节点 $v_x,v_y\in V$, 均有边 $e=<v_x,v_y>\in E$ 等价于边 $e'=<f(v_x),f(v_y)>\in E'$, 则称 $G(V,E)$ 和 G' (V',E') 同构。

7) 邻接矩阵。图论中, 常用邻接矩阵存储图的连接关系。变换器拓扑同样可以建模为对应的邻接矩阵 $\boldsymbol{A}_{\mathrm{dj}}$。当有方向的元件(输入端口、开关管和二极管)连接于节点 i 和节点 j, 且这些元件的参考方向如图 2 所示, 则 $A_{\mathrm{dj}}(i,j)=Z_{\mathrm{pn}}$ 且 $A_{\mathrm{dj}}(j,i)=Z_{\mathrm{np}}$, 其中, $Z\in\{U, S, D\}$; 当无方向的元件(输出端口、电感和电容)连接于节点 i 和节点 j 时, $A_{\mathrm{dj}}(i,j)=A_{\mathrm{dj}}(j,i)=Z$; 当没有元件连接于节点 i 和节点 j 时, $A_{\mathrm{dj}}(i,j)=A_{\mathrm{dj}}(j,i)=0$; 此外, 当 $i=j$ 时, $A_{\mathrm{dj}}(i,j)=1$。

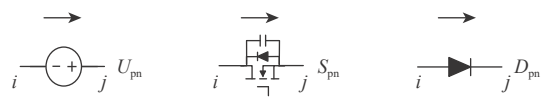


图 2　有方向元件的参考方向

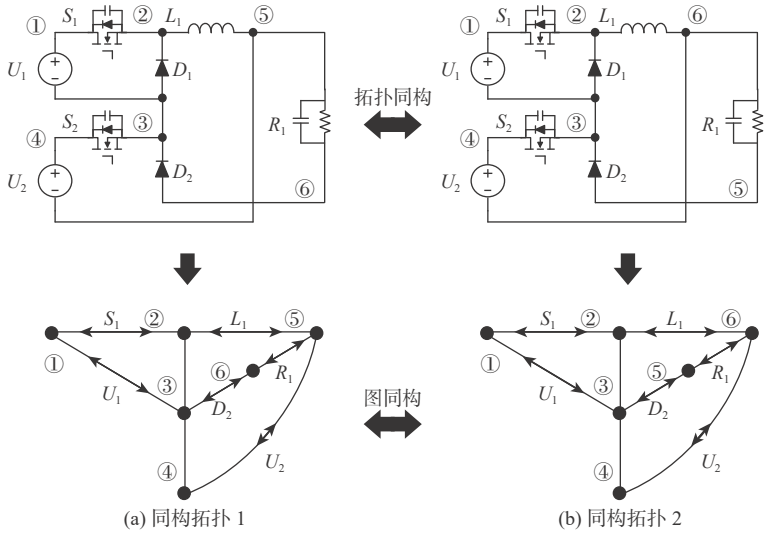
Fig. 2　Reference directions of components with directions

2　变换器拓扑同构

在基于计算机辅助的拓扑推演过程中, 变换器通过邻接矩阵进行数字化表征。由于节点编号的随机性, 常常会出现同一拓扑的邻接矩阵不同。以图 3 为例, 若同一拓扑的 2 个节点编号⑤和⑥对调, 则对应的有向有权图不同, 相应的, 邻接矩阵 $\boldsymbol{A}_{\mathrm{dj}A}$ 和 $\boldsymbol{A}_{\mathrm{dj}B}$ 的第 5 行/列和第 6 行/列都不同, 如下:

$$\boldsymbol{A}_{\mathrm{dj}A}=\begin{bmatrix}1 & S_{\mathrm{pn}} & U_{\mathrm{pn}} & 0 & 0 & 0\\ S_{\mathrm{np}} & 1 & D_{\mathrm{np}} & 0 & L & 0\\ U_{\mathrm{np}} & D_{\mathrm{pn}} & 1 & S_{\mathrm{np}} & 0 & D_{\mathrm{np}}\\ 0 & 0 & S_{\mathrm{pn}} & 1 & U_{\mathrm{pn}} & 0\\ 0 & L & 0 & U_{\mathrm{np}} & 1 & R\\ 0 & 0 & D_{\mathrm{pn}} & 0 & R & 1\end{bmatrix} \tag{1}$$

$$\boldsymbol{A}_{\mathrm{dj}B}=\begin{bmatrix}1 & S_{\mathrm{pn}} & U_{\mathrm{pn}} & 0 & 0 & 0\\ S_{\mathrm{np}} & 1 & D_{\mathrm{np}} & 0 & 0 & L\\ U_{\mathrm{np}} & D_{\mathrm{pn}} & 1 & S_{\mathrm{np}} & D_{\mathrm{np}} & 0\\ 0 & 0 & S_{\mathrm{pn}} & 1 & 0 & U_{\mathrm{pn}}\\ 0 & 0 & D_{\mathrm{pn}} & 0 & 1 & R\\ 0 & L & 0 & U_{\mathrm{np}} & R & 1\end{bmatrix} \tag{2}$$



(a) 同构拓扑 1　　(b) 同构拓扑 2

图 3　2 个同构的变换器拓扑及其对应的图

Fig. 3　Two isomorphic converter topologies and their corresponding graphs

此时，计算机程序会将 $\boldsymbol{A}_{\mathrm{dj}A}$ 和 $\boldsymbol{A}_{\mathrm{dj}B}$ 认为是 2 个不同拓扑的邻接矩阵，但实际上其是性能完全相同的同构拓扑。

根据图论中同构的定义，2 个同构的图本质上是节点编号不同情况下相同的图。据此，2 个同构拓扑的邻接矩阵总可以通过一系列的行初等变换和列初等变换相互转换，如式 (3) 所示，这意味着同构拓扑的邻接矩阵是相似的。

$$\boldsymbol{A}_{\mathrm{dj}A}=\boldsymbol{Q}^{-1}\boldsymbol{A}_{\mathrm{dj}B}\boldsymbol{Q} \tag{3}$$

式中：$\boldsymbol{Q}$ 为初等变换矩阵。

根据相似矩阵的特性可知，相似拓扑应具有相同的行列式。设 2 个变换器拓扑 A 和 B 的邻接矩阵分为 $\boldsymbol{A}_{\mathrm{dj}A}$ 和 $\boldsymbol{A}_{\mathrm{dj}B}$，可得以下推论。

推论 1　拓扑 A 和拓扑 B 同构是其性能相同的充分不必要条件。

推论 2　拓扑 A 和拓扑 B 同构的必要条件是其邻接矩阵的行列式相等，即 $\det(\boldsymbol{A}_{\mathrm{dj}A})=\det(\boldsymbol{A}_{\mathrm{dj}B})$。

3　变换器拓扑等效

文献 [21] 根据 2 个拓扑的元件组成是否完全相同及 2 个拓扑的简单回路是否一一对应，判定 2 个拓扑是否等效。

设由 N 个元件组成的变换器拓扑有 N_{lp} 条回路，每个回路的电流分别为 $\{i_1(t), i_2(t), \cdots, i_{Nlp}(t)\}$，则第 j 个元件的电流 $i_j(t)$ 可通过式 (4) 计算：

$$i_j(t)=\sum_{p=1}^{N_p}k_{j_p}i_p(t)\qquad j=\{1,2,\cdots,N\} \tag{4}$$

式中：当第 j 个元件包含于第 p 个回路时，k_{j_p} 等于 1 或−1，否则，k_{j_p} 等于 0，如下：

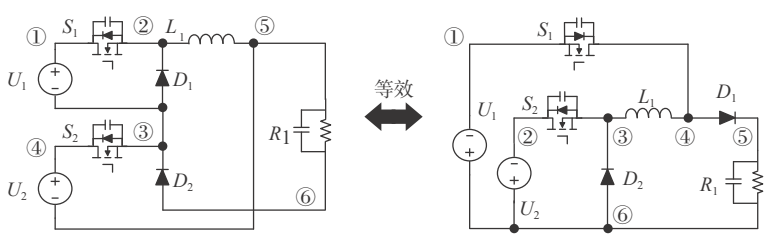
$$k_{j_p},m_{j_p}=\begin{cases}1\text{或}-1 & \text{第}p\text{个回路包含第}j\text{个元件}\\ 0 & \text{第}p\text{个回路不包含第}j\text{个元件}\end{cases} \tag{5}$$

此外，第 j 个元件的电压 $v_j(t)$ 应满足：

$$\sum_{p=1}^{N_{lp}}m_{j_p}v_j(t)=0\qquad j=\{1,2,\cdots,N\} \tag{6}$$

式中：当第 j 个元件包含于第 p 个回路时，m_{j_p} 等于 1 或−1，否则，m_{j_p} 等于 0。

因此，若拓扑 A 和拓扑 B 的元件组成相同且所有简单回路存在一一对应，则这 2 个拓扑所有元件的电压和电流都相等，即 2 个拓扑的性能相同。以图 4 中的 2 个等效拓扑为例，它们的元件组成均为 $\{U_1,U_2,L_1,S_1,D_1,D_2,R_1\}$，其所有简单回路如图 5 所示。其中，回路中对于有方向元件（输入端口、二极管和开关管），当回路方向与其在图 2 中定义的参



(a) 等效拓扑 1　　(b) 等效拓扑 2

图 4　2 个等效的变换器拓扑

Fig. 4　Two equivalent converter topologies

图 4(a) 拓扑的所有简单回路

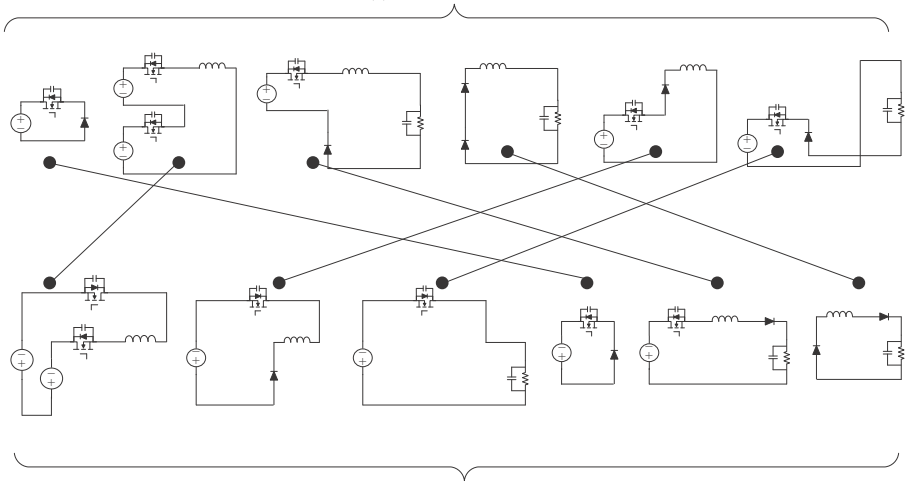


图 4(b) 拓扑的所有简单回路

图 5　图 4 中 2 个拓扑简单回路的对应关系

Fig. 5　One-by-one matching simple loops of two topologies in Fig. 4

考方向相同时, 元件在回路中记为 $Z_{\rm pn}$, 否则记为 $Z_{\rm np}$。如图 5 所示, 图 4(a) 和图 4(b) 拓扑的所有简单回路都一一对应, 因此, 这 2 个拓扑等效。设 2 个变换器拓扑 A 和 B, 可得到以下推论。

推论 3　拓扑 A 和拓扑 B 等效是其性能相同的充分必要条件。

推论 4　拓扑 A 和拓扑 B 等效的充分必要条件是其元件组成相同且所有简单回路一一对应。

4　拓扑同构与拓扑等效辨析

基于推论 2 可知, 当拓扑 A 和拓扑 B 同构时, 其邻接矩阵的行列式相等。若任意矩阵 $\boldsymbol{A}_{\rm dj}$ 为

$$\boldsymbol{A}_{\rm dj}=\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \tag{7}$$

其行列式通过式 (8) 计算:

$$\det\left(\boldsymbol{A}_{\rm dj}\right)=\sum_{k=1}^{K}(-1)^k a_{1,\varphi_k(1)}a_{2,\varphi_k(2)}\cdots a_{n,\varphi_k(n)} \tag{8}$$

式中: φ_k 为序列 $\{1,2,\cdots,n\}$的一种置换。$\Phi=\{\varphi_1,\varphi_2,\cdots,\varphi_K\}$为序列 $\{1,2,\cdots,n\}$上置换的全体, 即序列 $\{1,2,\cdots,n\}$到自身上的一一映射的全体; 而 $\{\varphi_k(1),\varphi_k(2),\cdots,\varphi_k(n)\}$则为序列 $\{1,2,\cdots,n\}$中的元素进行 k 次交换后得到的序列, k 也称为该序列的逆序数。

将式 (8) 中行列式的任何一项 $a_{1,\varphi_k(1)},a_{1,\varphi_k(2)},\cdots,a_{1,\varphi_k(n)}$记为 $Y_{\varphi k}$, 当 $Y_{\varphi_k}\neq 0$ 且 Y_{φ_k} 包含 $a_{x,\varphi_k(x)}$时, 从图论角度可以将 $a_{x,\varphi_k(x)}$视为其所对应的元件从 x 节点流入、从 $\varphi_k(x)$ 节点流出。事实上, n 阶矩阵行列式中的每一项 Y_{φ_k} 都是从矩阵中选取 n 个元素相乘得到, 且保证在每行和每列中都恰好只选取一个元素。因此, 将邻接矩阵所对应的图记为 G, 其行列式中的任意一个非零项 Y_{φ_k} 都正好是图 G 满足所有节点出度和入度都为 1 的一个子图。邻接矩阵的行列式可以视为将所有这样的子图都选取出来后, 并通过将子图的元件进行乘积后求和得到。

以图 3(a) 的变换器为例, 其邻接矩阵如式 (1) 所示。如图 6 所示, 该邻接矩阵行列式的某一项 $Y_{\varphi k}=a_{1,2}a_{2,5}a_{3,1}a_{4,3}a_{5,4}a_{6,6}$, 其中, $a_{1,2},a_{2,5},a_{3,1},a_{4,3},a_{5,4}$ 分别对应于元素 $S_{\rm pn},L,U_{\rm np},S_{\rm pn},U_{\rm np}$。由于 $a_{6,6}$ 的值为 1, 不影响 Y_{φ_k} 的值, $a_{1,2},a_{2,5},a_{3,1},a_{4,3},a_{5,4}$ 这些元素的乘积与简单回路①-②-⑤-④-③-①相对应。显然, 该简单回路是原拓扑的一个子图, 且该子图所有的节点出度和入度都为 1。值得注意的是, 行列式的项不仅包含简单回路, 还包含一些其他类型的子图。例如, 图 7 中, 邻接矩阵的某一项 $Y_{\varphi_k}=a_{1,2}a_{2,1}\cdot a_{3,3}a_{4,4}a_{5,6}a_{6,5}=S_{\rm np}S_{\rm pn}R^2$ 确定的子图为 2 个没有公共节点的元件。基于以上分析, 定义集合 G_{X-1} 为图 G_X 的第一类子

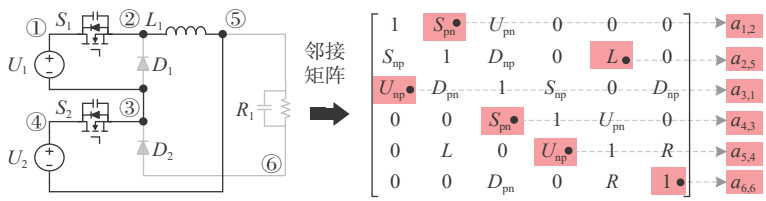


图 6　邻接矩阵行列式某项与简单回路的对应关系

Fig. 6　Corresponding relationship between each item in determinant of adjacent matrix and simple loop

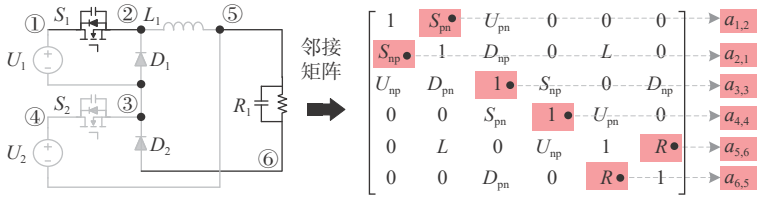


图7 邻接矩阵行列式某项与2个不相连元件的对应关系

Fig. 7 Corresponding relationship between each item in determinant of adjacent matrix and two disconnected components

图集合，任何图 $g\in G_{X-1}$ 的所有节点的出度和入度都为1。从拓扑子图的角度可得到如下推论。

推论5 拓扑 A 和拓扑 B 对应的图分别为图 G_A 和图 G_B。拓扑 A 和拓扑 B 同构的必要不充分条件为图 G_A 和图 G_B 的所有第一类子图一一对应，并表示为 $G_{A-1}=G_{B-1}$。

基于推论4可知，拓扑 A 和拓扑 B 等效的充要条件是其元件组成相同且所有简单回路一一对应。根据简单回路的定义，简单回路实际上就是出度和入度都等于1且节点数量大于2的连通图。实际上，出度和入度都为1且节点数量为2的子图可以视为一个元件。因此，定义集合 G_{X-2} 为图 G_X 的第二类子图集合，任何图 $g\in G_{X-2}$ 是连通的，且其所有节点的出度和入度都为1，则可以得到如下推论。

推论6 拓扑 A 和拓扑 B 对应的图分别为图 G_A 和图 G_B。拓扑 A 和拓扑 B 等效的充要条件为图 G_A 和图 G_B 的所有第二类子图一一对应，并表示为 $G_{A-2}=G_{B-2}$。

设有2个拓扑 A 和 B，其对应的图为 G_A 和 G_B；第一类子图集合和第二类子图集合分别为 G_{A-1}，G_{A-2}，G_{B-1}，G_{B-2}。设定如下命题：

1）P0：拓扑 A 和 B 同构；

2）P1：$G_{A-1}=G_{B-1}$；

3）Q0：拓扑 A 和 B 等效；

4）Q1：$G_{A-2}=G_{B-2}$；

5）R：拓扑 A 和 B 性能相同。

基于以上定义的命题，结合推导得到的推论1、推论3、推论5和推论6，进行进一步分析。分析发现，第一类子图集合必然包含第二类子图集合。这是因为第二类子图在满足第一类子图条件的基础上，还需要额外满足连通性的条件。因此，可以得出结论：P1是Q1的充分非必要条件。

在此基础上，拓扑等效与拓扑同构之间的关系推导为：P0 ⇒ P1 ⇒ Q1 ⇔ Q0 ⇔ R。首先，根据推论1，确定同构拓扑的第一类子图相等；然后，由于第一类子图集合包含第二类子图集合，可以推导出同构拓扑的第二类子图也相等；最后，结合第二类子图相等与拓扑等效的充要关系，可以进一步确定同构的拓扑一定是等效的。可得，拓扑同构是拓扑等效的充分非必要条件。

5 案例分析

5.1 拓扑等效与拓扑同构关系验证

为验证拓扑等效与拓扑同构的关系，对图8的2个三端口变换器拓扑 A_1 和 B_1 进行分析。首先，根据拓扑的连接关系获取其邻接矩阵 $\boldsymbol{A}_{\mathrm{dj}A_1}$ 和 $\boldsymbol{A}_{\mathrm{dj}B_1}$，如下：

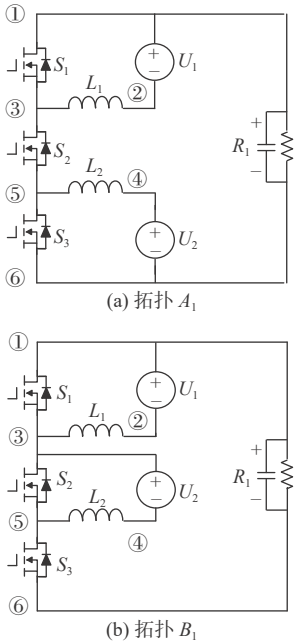
$$\boldsymbol{A}_{\mathrm{dj}A_1}=\begin{bmatrix}1 & U_{\mathrm{np}} & S_{\mathrm{pn}} & 0 & 0 & R\\ U_{\mathrm{pn}} & 1 & L & 0 & 0 & 0\\ S_{\mathrm{np}} & L & 1 & 0 & S_{\mathrm{pn}} & 0\\ 0 & 0 & 0 & 1 & L & U_{\mathrm{np}}\\ 0 & 0 & S_{\mathrm{np}} & L & 1 & S_{\mathrm{pn}}\\ R & 0 & 0 & U_{\mathrm{pn}} & S_{\mathrm{np}} & 1\end{bmatrix} \tag{9}$$

$$\boldsymbol{A}_{\mathrm{dj}B_1}=\begin{bmatrix}1 & U_{\mathrm{np}} & S_{\mathrm{pn}} & 0 & 0 & R\\ U_{\mathrm{pn}} & 1 & L & 0 & 0 & 0\\ S_{\mathrm{np}} & L & 1 & U_{\mathrm{np}} & S_{\mathrm{pn}} & 0\\ 0 & 0 & U_{\mathrm{pn}} & 1 & L & 0\\ 0 & 0 & S_{\mathrm{np}} & L & 1 & S_{\mathrm{pn}}\\ R & 0 & 0 & 0 & S_{\mathrm{np}} & 1\end{bmatrix} \tag{10}$$

其次，通过将 $\boldsymbol{A}_{\mathrm{dj}A_1}$ 和 $\boldsymbol{A}_{\mathrm{dj}B_1}$ 的行列式相减后发现，它们的行列式不相等，如下：

$$\det\left(\boldsymbol{A}_{\mathrm{dj}A_1}\right)-\det\left(\boldsymbol{A}_{\mathrm{dj}B_1}\right)=U_{\mathrm{pn}}^2S_{\mathrm{pn}}^2L^2+\cdots\neq 0 \tag{11}$$

因此，根据推论2，图8中的拓扑 A_1 和拓扑 B_1 是非同构的。然后，获取这2个拓扑所有节点出度和入度都为1的子图，并构造出相应的第一类子图集，如图9所示。为方便展示，将2个方向相反且元件组成相同的简单回路合并为一个图。观察图9发现，拓扑 A_1 的第一类子图集中连通的部分与拓扑 B_1 的第一类子图集中连通的部分一一对应，即拓扑 A_1 和拓扑 B_1 的元件组成相同且简单回路一一对应。因此，根据推论4可以判定，拓扑 A_1 与拓扑 B_1 等效。依据以上子图的等效判定与实际相符，因为拓扑 A_1 中的 S_2 与 $\{L_2, S_3, U_2\}$ 整体串联，所以，将其交换后得到拓扑 B_1。进一步分析图9中拓扑 A_1 与拓扑 B_1 的第一类子图集发现，其非连通的部分并不能一一对应。图9中，拓扑 A_1 的一个由2个简单回路①-②-③-①和⑤-④-⑥-⑤组成的子图无法在拓扑 B_1 的第一类



(a) 拓扑 A_1

(b) 拓扑 B_1

图 8　用于案例分析的三端口变换器拓扑

Fig. 8　Topologies of two three-port converter for case analysis

子图中找到与之对应的子图。因此, 可以判定拓扑 A_1 与拓扑 B_1 不同构。依据以上子图的同构判定与实际相符, 因为拓扑 A_1 和拓扑 B_1 的邻接矩阵行列式不相等, 且式 (11) 中其行列式的差值中包含一项 $U_{\mathrm{pn}}^2 S_{\mathrm{pn}}^2 L^2$, 对应于拓扑 A_1 中由 2 个简单回路①-②-③-①和⑤-④-⑥-⑤组成的子图。基于以上分析可知, 图 8 中的拓扑 A_1 和拓扑 B_1 等效但不同构, 验证了拓扑同构是拓扑等效的充分不必要条件。

5.2　等效拓扑自动判定

基于推论 6 可知, 拓扑等效识别的关键在于得到该拓扑图的第二类子图集, 即出度和入度都为 1 且连通的子图集。本文通过深度优先搜索算法自动获取拓扑的第二类子图集, 实现等效拓扑的自动判定。首先, 任取拓扑的一个节点为根节点 N_{root}, 并以其作为起始节点开始逐层向下深度搜索。然后, 当搜索到 N_{root} 或无法进行深度搜索时, 会进行回溯搜索, 直到拓扑中的所有节点都已经被搜索; 当搜索到 N_{root} 时, 说明找到了一个出度和入度都为 1 的连通子图, 此时在回溯前还需存储该满足条件的子图; 完成某一个节点作为根节点 N_{root} 的子图搜索后, 将该节点从拓扑中删除, 并从剩余的节点中任取一个节点作为新的根节点 N_{root} 继续搜索。最后, 当所有节点都被作为根节点进行搜索后, 可以得到拓扑所有的第二类子图, 即第二类子图集。

以图 8(a) 拓扑为例, 当选取节点①作为根节点时, 第二类子图的搜索过程如图 10 所示。首先, 从节点①开始, 搜索至节点②后回到节点①, 得到第 1 个第二类子图①-②-①, 即元件 S_1; 然后, 回溯至节点②后搜索节点③, 再次搜索到节点①, 得到了第 2 个第二类子图①-②-③-①, 即由元件 $\{U_1, L_1, S_1\}$ 组成的简单回路; 以此类推, 最后, 得到 9 个以节点①为根节点的第二类子图。完成图 10 的搜索后, 删除节点①及与节点①关联的元件后, 从剩余的节点中任选一个作为根节点重新搜索, 直到所有

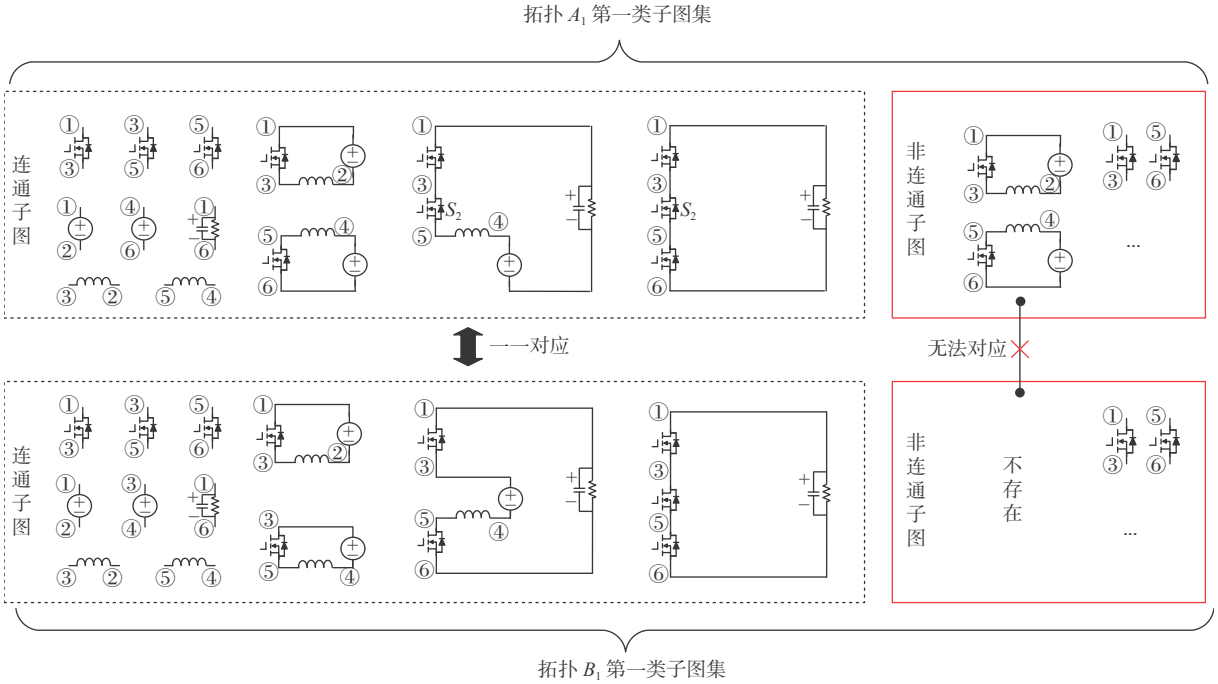


图 9　图 8 中拓扑 A_1 和拓扑 B_1 的第一类子图集

Fig. 9　The first class of subgraph set of topology A_1 and topology B_1 in Fig. 8

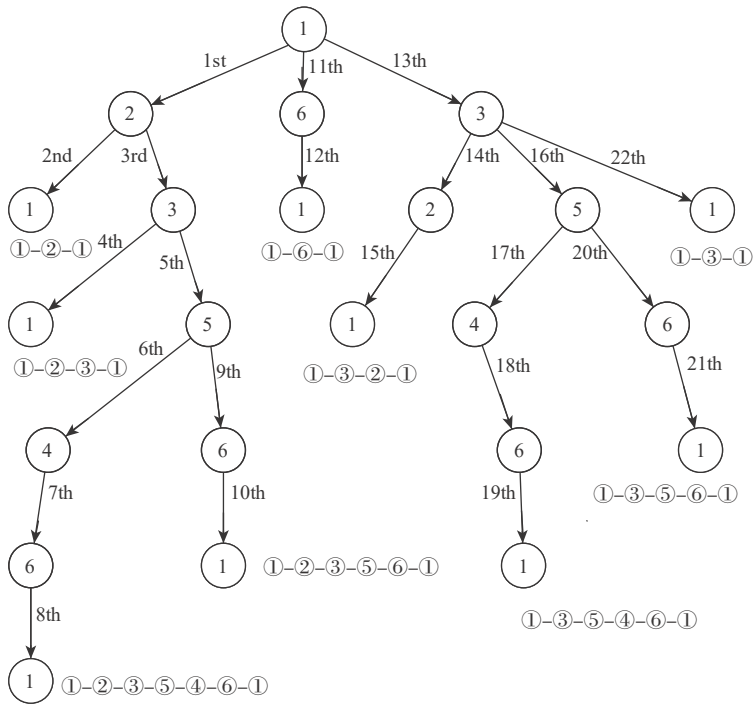


图 10 基于深度优先搜索算法获得的图 8(a) 拓扑以节点①为根节点的第二类子图

Fig. 10 Processes of obtaining the second class of subgraph of topology in Fig. 8(a) with root node based on depth-first-searching algorithm

节点都被选为根节点完成搜索后，可以得到拓扑的第二类子图集。在获取 2 个拓扑的第二类子图集后，通过对比它们是否一一对应，可以实现拓扑等效的自动判定。

6 结 论

1）本文证明了拓扑同构是性能相同的充分非必要条件，而拓扑等效是其充要条件，并提出应采用拓扑等效进行拓扑识别以提高准确性并避免重复研究。

2）基于拓扑等效的充要条件，设计了一种基于深度优先搜索的自动识别方法，成功实现了性能相同拓扑的快速筛选和识别。

3）本文为电力电子变换器拓扑的性能分析与设计优化提供了理论基础和技术支持，未来可通过优化算法效率并结合智能化技术扩展至更复杂的拓扑结构。

参考文献（References）

[1] ZHANG G D, LI Z, ZHANG B, et al. Power electronics converters: past, present and future[J]. Renewable and Sustainable Energy Reviews, 2018, 81: 2028-2044.

[2] ELSAYAD N, MORADISIZKOOHI H, MOHAMMED O A. A new hybrid structure of a bidirectional DC-DC converter with high conversion ratios for electric vehicles[J]. IEEE Transactions on Vehicular Technology, 2020, 69(1): 194-206.

[3] GNANASAMBANDAM K, RATHORE A K, EDPUGANTI A, et al. Current-fed multilevel converters: an overview of circuit topologies, modulation techniques, and applications[J]. IEEE Transactions on Power Electronics, 2017, 32(5): 3382-3401.

[4] 荣德生，孙瑄璋. 高增益耦合电感组合 Buck-Boost-Zeta 变换器.[J]. 中国电机工程学报, 2020, 40(14): 4590-4601.
RONG D S, SUN X J. High gain coupled inductance combination Buck-Boost-Zeta converter[J]. Proceedings of the CSEE, 2020, 40 (14): 4590-4601(in Chinese).

[5] DE ANDRADE J M, SALVADOR M A, COELHO R F, et al. General method for synthesizing high gain step-up DC-DC converters based on differential connections[J]. IEEE Transactions on Power Electronics, 2020, 35(12): 13239-13254.

[6] GANJAVI A, GHOREISHY H, AHMAD A A. A novel single-input dual-output three-level DC-DC converter[J]. IEEE Transactions on Industrial Electronics, 2018, 65(10): 8101-8111.

[7] LI X L, DONG Z, TSE C K, et al. Single-inductor multi-input multi-output DC-DC converter with high flexibility and simple control[J]. IEEE Transactions on Power Electronics, 2020, 35(12): 13104-13114.

[8] KUMAR G G, KRISHNA M V S, KUMARAVEL S, et al. Multistage DC-DC converter using active LC2D network with minimum component[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021, 68(3): 943-947.

[9] MO L P, CHEN G P, HUANG J M. From components to converters: a fundamental topology derivation method for single-inductor multi-input multi-output converters based on graph theory[C]//Proceedings of the IEEE International Conference on Power Electronics, Computer Applications. Piscataway: IEEE Press, 2021: 458-463.

[10] CHEN Y, BAI J B, KANG Y. A nonisolated single-inductor multiport DC-DC topology deduction method based on reinforcement learning[J]. IEEE Journal of Emerging and Selected Topics in

Power Electronics, 2022, 10(6): 6572-6585.

[11] LI H, LI Y, WANG W, et al. Programmable topology deduction algorithm for $S^1D^2C^1L^1$ type DC-DC converters based on graph theory[C]//Proceedings of the 14th International FLINS Conference. Singapore: World Scientific, 2020.

[12] 李虹, 王文财, 李亚敏, 等. 基于图论的 $S^1D^1L^2C^1$ 型 DC-DC 变换器可编程拓扑搜索算法[J]. 中国电机工程学报, 2021, 41(16): 5670-5683.
LI H, WANG W C, LI Y M, et al. Programmable topology searching algorithm for $S^1D^1L^2C^1$ type DC-DC converters based on graph theory[J]. Proceedings of the CSEE, 2021, 41(16): 5670-5683(in Chinese).

[13] CHEN G P, JIN Z F, LIU Y W, et al. Programmable topology derivation and analysis of integrated three-port DC-DC converters with reduced switches for low-cost applications[J]. IEEE Transactions on Industrial Electronics, 2019, 66(9): 6649-6660.

[14] MO L P, HUANG J M, CHEN G P, et al. Computer-aided systematic topology derivation of single-inductor multi-input multi-output converters from working principle[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2022, 69(6): 2637-2649.

[15] 石勇, 杨旭, 何群, 等. 同构混合开关拓扑的辨识[J]. 中国电机工程学报, 2003, 23(11): 116-121.
SHI Y, YANG X, HE Q, et al. Identification of isomorphic hybrid switching topology[J]. Proceedings of the CSEE, 2003, 23(11): 116-121(in Chinese).

[16] SHANG H L, LI F, TANG X D, et al. A new algorithm for isomorphism determination of undirected graphs-circuit simulation method[J]. Circuits, Systems, and Signal Processing, 2011, 30(5): 1115-1130.

[17] 商慧亮, 刘洋, 柳志栋, 等. 改进电路模拟法的应用: 同构混合开关拓扑辨识[J]. 应用科学学报, 2014, 32(2): 199-208.
SHANG H L, LIU Y, LIU Z D, et al. Application of optimized circuit simulation: identification of isomorphic hybrid switching[J]. Journal of Applied Sciences, 2014, 32(2): 199-208(in Chinese).

[18] GROHE M, SCHWEITZER P. The graph isomorphism problem[J]. Communications of the ACM, 2020, 63(11): 128-134.

[19] LI Y Z, KUPRAT J, LI Y W, et al. Graph-theory-based derivation, modeling, and control of power converter systems[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2022, 10(6): 6557-6571.

[20] MERKYS A, VAITKUS A, GRYBAUSKAS A, et al. Graph isomorphism-based algorithm for cross-checking chemical and crystallographic descriptions[J]. Journal of Cheminformatics, 2023, 15(1): 25.

[21] CHEN G P, MO L P, LIU Y W, et al. Computer-aided identification of equivalent power electronics converters[J]. IEEE Transactions on Power Electronics, 2019, 34(10): 9374-9378.

[22] 肖位枢. 图论及其算法[M]. 北京: 航空工业出版社, 1993.
XIAO W S. Graph theory and its algorithm[M]. Beijing: Aviation Industry Press, 1993(in Chinese).

Discrimination on isomorphism and equivalence of topologies of power electronics converters

MO Liping[1, 2], CHEN Guipeng[1, 2, *], LI Zhiqiang[3]

(1. Shenzhen Research Institute of Xiamen University, Shenzhen 518000, China;
2. School of Aerospace Engineering, Xiamen University, Xiamen 361000, China;
3. Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang 621000, China)

Abstract: In the process of deriving topologies for power electronic converters, researchers often use topology equivalence or topology isomorphism to identify topologies with different structures but identical performance, thereby avoiding redundant studies. However, the connotations of topology equivalence and topology isomorphism differ significantly, which can easily lead to confusion. To address this issue, this paper aims to clarify the distinctions and connections between the two concepts and propose a method for accurately identifying topologies with identical performance. First, it is deduced that a necessary condition for topology isomorphism is that their adjacency matrices have the same determinant. Furthermore, it is derived that the necessary and sufficient condition for topology equivalence is that the components are identical and the fundamental loops correspond one-to-one. Subsequently, an analysis from the perspective of topology subgraphs reveals that topology isomorphism is a sufficient but not necessary condition for topology equivalence, while topology equivalence is a necessary and sufficient condition for identical performance. Based on this, the paper recommends prioritizing topology equivalence over topology isomorphism in practical applications to identify topologies with identical performance. The theoretical results are validated through case studies demonstrating their correctness and feasibility. Additionally, this paper proposes a method based on the depth-first search algorithm to automatically determine equivalent topologies, enabling the rapid and accurate identification of converter topologies with identical performance.

Keywords: power electronics converters; isomorphic converter topologies; equivalent converter topologies; graph theory; depth-first-searching

Received: 2023-06-14; **Accepted**: 2023-09-01; **Published Online**: 2023-09-13 18:55

URL: link.cnki.net/urlid/11.2625.V.20230913.1129.001

Foundation items: Guangdong Basic and Applied Basic Research Foundation (2022A1515011616); National Natural Science Foundation of China (51907172)

* **Corresponding author**. E-mail: cgp2017@xmu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0358

基于图论的变换器冗余开关自动识别方法

莫李平[1,2]，宋吉祥[1,2]，陈桂鹏[1,2,*]

(1. 厦门大学深圳研究院，深圳 518000；2. 厦门大学 航空航天学院，厦门 361000)

摘 要：电能变换需求日益增长，推演具有不同性能的新型变换器成为研究热点。当前的推演方法大多未考虑变换器开关冗余的因素，导致推演结果中同时存在含冗余开关的变换器和不含冗余开关的变换器。为从海量的推演结果中快速剔除含冗余开关的变换器，开展了变换器冗余开关自动识别研究。分析变换器冗余开关与其有效开关模态的关系，并在此基础上归纳冗余开关的识别准则；基于图论将识别准则转化为对应的图准则；应用图搜索算法结合图准则建立冗余开关自动识别方法。为验证冗余开关自动识别方法的可行性，将其用于识别现存单电感多端口变换器的冗余开关，识别结果表明：所提方法可以批量、准确地识别出变换器的所有冗余开关，验证了方法的可行性。

关 键 词：图论；变换器；冗余开关；自动识别；开关模态

中图分类号：TM46; O157.5

文献标志码：A **文章编号**：1001-5965(2025)05-1449-08

在“双碳”背景下，电力电子技术依靠半导体功率器件可实现电能的高效率转换与控制，已被广泛应用于日常生活和工业生产，如移动设备、电动汽车、微电网、物联网等[1-4]。为满足各式各样的电能变换需求，推演具有不同特性的新型变换器成为研究热点。文献 [5] 根据对偶原理，将开关电容直流电压型变换器中的电压源、电压型负载和电容分别替换为电流源、电流型负载和电感，同时改变电路连接，推演出对应的开关电感直流电流型变换器。类似地，文献 [6] 根据对偶原理，从电压型多电平变换器推演出电流型多电平变换器。文献 [7] 通过共用 Boost 与 Sepic 变换器输入侧的电感和开关，并将输出端口差分连接，得到了高电压增益 Boost-Sepic 变换器。通过类似的方法，文献 [8-9] 将不同的变换器进行组合，推演出了 Boost-Boost、Sepic-Zeta、Buck-Boost-Zeta 等一系列的高增益变换器。近年来，为提高变换器推演的效率，基于计算机辅助的变换器自动推演方法成为热点[10-17]。这些自动推演方法以图论为桥梁，将变换器的推演问题建模为图的搜索问题，进而在计算机辅助下完成变换器推演。自动推演方法可以实现批量、自动地推演，其推演结果往往包含几十甚至上百个新变换器，极大提升了变换器的推演效率。

然而，目前的推演方法往往只专注于得到有效的变换器，并未考虑推演的变换器是否含冗余开关，导致推演结果中可能存在大量含冗余开关的变换器。冗余开关的存在与否并不影响变换器的主要性能(如电压增益、元件电流应力等)，但会增加半导体开关器件的数量，从而增加变换器的使用成本。识别并简化变换器中的冗余开关，有助于减少半导体开关器件数量，降低成本。当前，变换器的冗余开关往往通过人工分析开关与开关的连接关系进行识别。文献 [15-16] 将 2 个串联开关中的 1 个识别为冗余开关。文献 [17] 将单电感变换器电

收稿日期：2023-06-14；**录用日期**：2023-10-11；**网络出版时间**：2024-01-23 17:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20240119.1136.002
基金项目：广东省基础与应用基础研究基金 (2022A1515011616)；中央高校基本科研业务费专项资金 (20720220083)
* **通信作者**. E-mail：cgp2017@xmu.edu.cn

引用格式：莫李平，宋吉祥，陈桂鹏. 基于图论的变换器冗余开关自动识别方法 [J]. 北京航空航天大学学报，2025，51（5）：1449-1456.
MO L P，SONG J X，CHEN G P. Auto identification method for redundant switches in converters based on graph theory [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1449-1456（in Chinese）.

感两端的开关是否连接于同一节点作为冗余开关的判定依据进行识别。显然，这些方法并未从冗余开关的本质出发确定识别准则，导致无法全面识别变换器中的所有冗余开关。此外，这些方法依赖于人工，不适用于处理现存的大量变换器。

基于此，本文从变换器工作基本原理出发，根据冗余开关与有效开关模态的关系归纳冗余开关的识别准则，并结合图论和图搜索算法实现冗余开关的自动识别。

1 冗余开关识别准则

若变换器中的一个开关器件 S_i 一直处于导通或关断状态，不会影响变换器的基本工作特性，则 S_i 为冗余开关。冗余开关可分为导通冗余开关和关断冗余开关。导通冗余开关可一直导通，并可被简化为导线，如图 1 左侧变换器的冗余开关 S_1 被简化为导线后，得到图 1 右侧与之特性基本相同的变换器。关断冗余开关可一直处于关断，并可被删除，如图 2 左侧变换器的冗余开关 S_1 删除后，可得到图 2 右侧的变换器。本质上，变换器的工作原理是通过周期性地切换若干种有效开关模态实现电能变换[18]。因此，本文根据开关简化/删除后是否影响变换器的有效开关模态种类，来判定开关是否冗余。例如，图 1 左侧变换器的开关 S_1 简化为导线前后的有效开关模态种类相同。因此，图 1 左侧变换器的开关 S_1 是导通冗余开关。类似地，图 2 左侧变换器的开关 S_1 删除前后，变换器均工作于 4 种有效开关模态。因此，图 2 左侧变换器的开关 S_1 为关断冗余开关。

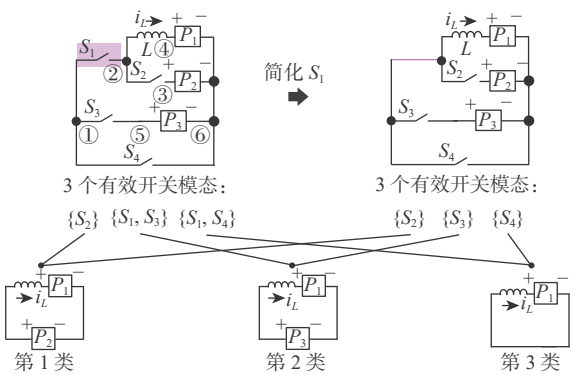


图 1 导通冗余开关简化前后变换器的有效开关模态种类

Fig. 1 Valid switching modes types of converters before and after turn-on redundant switches are simplified

基于以上分析，根据开关简化或删除后变换器有效开关模态种类是否变化，确定冗余开关的识别准则。对于包含 N_S 个理想开关{S_1, S_2, ⋯, S_{N_S}}的变换器，定义该变换器的有效开关模态集（valid switching mode, VSM），该集合中的开关模态互不相同。变换器中的某一开关 S_i 简化为导线后，其有效开关模态集记为 $V_{S_i\text{-on}}$；对应地，当变换器中的某一开关 S_i 删除后，其有效开关模态集记为 $V_{S_i\text{-off}}$。基于以上定义，归纳开关 Si 冗余的判断准则如下：

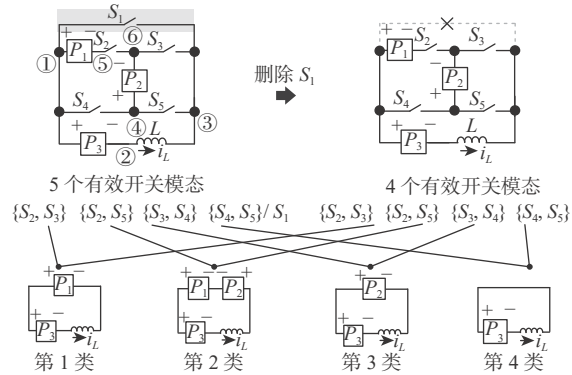


图 2 关断冗余开关删除前后变换器的有效开关模态种类

Fig. 2 Valid switching modes types of converters before and after turn-off redundant switches are deleted

1）准则 1。若 $V = V_{S_i\text{-on}}$，则 S_i 导通冗余。

2）准则 2。若 $V = V_{S_i\text{-off}}$，则 S_i 关断冗余。

2 冗余开关图准则

理论上，基于第 1 节的 2 条准则，通过人工就可以准确且全面地识别出变换器中的所有冗余开关。但是，人工识别耗时耗力，且难以处理大量的变换器。因此，本文将以上 2 条准则转化为对应的图准则。图准则被用于建立基于图论和图搜索算法的冗余开关自动识别方法。

实现冗余开关识别的关键在于获取变换器有效开关模态的种类。值得注意的是，有效开关模态与该模态下的所有回路存在对应关系，即 2 个回路集不同的模态是不同的模态，而 2 个回路集相同的模态则是同一模态。例如，图 2 左侧变换器在开关{S_2, S_3}导通和{S_2, S_5}导通时得到的回路不同，对应的模态为图 2 中第 1 类模态和第 2 类模态；而图 2 左侧变换器在 S_1 导通和{S_4, S_5}导通时的回路相同，均对应图 2 中的第 4 类模态。

基于变换器开关模态和该模态下回路集的对应关系，将冗余开关准则转化为图准则。假设变换器含 K 个有效开关模态：{$v_{\text{sm}1}$, $v_{\text{sm}2}$, ⋯, $v_{\text{sm}K}$}。这些模态用回路集 P_{VSM} 进行表示，如下：

$$P_{\text{VSM}} = \{(p_{\text{vsm}1}(1), p_{\text{vsm}1}(2), \cdots), (p_{\text{vsm}2}(1), p_{\text{vsm}2}(2), \cdots), \cdots, (p_{\text{vsm}K}(1), p_{\text{vsm}K}(2), \cdots)\} \tag{1}$$

式中：$p_{\text{vsm}}(K)$ 为有效开关模态的第 K 个回路。

集合 P_{VSM} 中的每个元素表示对应模态下的所有回路，如 $(p_{\text{vsm}1}(1), p_{\text{vsm}1}(2), \cdots)$ 表示有效开关模态 $v_{\text{sm}1}$ 下的所有回路。对应地，开关 S_i 简化为导线和

删除后的有效开关模态 $V_{S_i\text{-on}}$ 和 $V_{S_i\text{-off}}$ 分别用回路集 $P_{\text{VSM-}S_i\text{-on}}$ 和 $P_{\text{VSM-}S_i\text{-off}}$ 进行表示。准则 1 和准则 2 可以转换为如下图准则:

1) 图准则 1。若 $P_{\text{VSM}}=P_{\text{VSM-}S_i\text{-on}}$, 则 S_i 导通冗余。

2) 图准则 2。若 $P_{\text{VSM}}=P_{\text{VSM-}S_i\text{-off}}$, 则 S_i 关断冗余。

3　冗余开关自动识别

基于第 2 节的图准则, 提出变换器冗余开关自动识别方法, 如图 3 所示。首先, 基于图论将原变换器, 以及开关 S_i 简化和开关 S_i 删除后的变换器, 分别建模为对应的邻接矩阵 $\boldsymbol{A}_{\text{dj}}$、$\boldsymbol{A}_{\text{dj}S_i\text{-on}}$ 和 $\boldsymbol{A}_{\text{dj}S_i\text{-off}}$; 其次, 通过深度优先搜索(depth-first-searching, DFS)算法[19] 获取这 3 个变换器在所有开关模态下的回路, 并构建对应的开关模态回路集 P、$P_{S_i\text{-on}}$ 和 $P_{S_i\text{-off}}$; 然后, 根据有效开关模态约束筛选出有效开关模态, 得到有效开关模态回路集 P_{VSM}、$P_{\text{VSM-}S_i\text{-on}}$ 和 $P_{\text{VSM-}S_i\text{-off}}$; 最后, 根据归纳的图准则判定 S_i 是否为冗余。

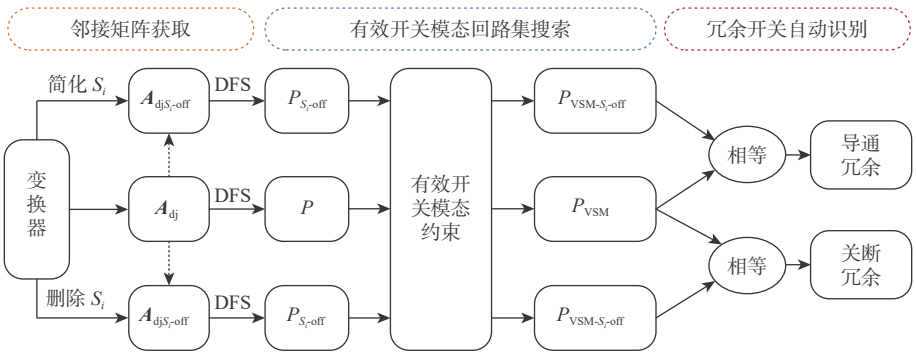


图 3　冗余开关自动识别总体流程

Fig. 3　Main flowchart of auto identification of redundant switches

3.1　变换器图建模

变换器的元件连接信息可以用邻接矩阵进行建模和存储[17-18,20]。在冗余开关自动识别方法建立过程中, 需要先将原变换器、简化某一开关后的变换器和删除某一开关后的变换器建模为对应的邻接矩阵。

如图 4(a) 所示, 以 Buck 变换器为例说明原变换器的邻接矩阵建立方法。首先, Buck 变换器有①、②、③、④这 4 个节点, 初始化邻接矩阵 $\boldsymbol{A}_{\text{dj}}$ 为 4×4 的零矩阵; 然后, 当任意元件连接于节点 i 和节点 j 之间时, $A_{\text{dj}}(i,j)$ 和 $A_{\text{dj}}(j,i)$ 等于该元件, 此外, 对于输入/输出端口和电容还需用下标 pn 和 np 记录元件的方向。例如, 图 4(a) 中, 当输入端口连接于节点①和节点④, 且方向为从节点①到节点④, $A_{\text{dj}}(1,4)=U_{\text{1-pn}}$, $A_{\text{dj}}(4,1)=U_{\text{1-np}}$。基于以上规则, 完成所有元件的赋值后, 可以得到原变换器的邻接矩阵。

在原变换器邻接矩阵 $\boldsymbol{A}_{\text{dj}}$ 的基础上, 可以快速建立开关 S_i 简化或删除后的邻接矩阵 $\boldsymbol{A}_{\text{dj}S_i\text{-on}}$ 和 $\boldsymbol{A}_{\text{dj}S_i\text{-off}}$。如图 4(b) 所示, 当 Buck 变换器中的 S_1 简化为导线后, 只需将 $\boldsymbol{A}_{\text{dj}}$ 中的 S_1 赋值为 1(代表开关 S_1 始终处于导通状态), 可以得到对应的邻接矩阵 $\boldsymbol{A}_{\text{dj}S_1\text{-on}}$。对应地, 如图 4(c) 所示, 当 Buck 变换器中的 S_1 删除后, 只需将 $\boldsymbol{A}_{\text{dj}}$ 中的 S_1 赋值为 0(代表开关 S_1 始终处于关断状态), 就可以得到对应的邻接矩阵 $\boldsymbol{A}_{\text{dj}S_1\text{-off}}$。

3.2　开关模态回路集搜索

在得到原变换器、开关 S_i 简化后和开关 S_i 删除后的变换器邻接矩阵 $\boldsymbol{A}_{\text{dj}}$、$\boldsymbol{A}_{\text{dj}S_i\text{-on}}$ 和 $\boldsymbol{A}_{\text{dj}S_i\text{-off}}$ 后, 需要进一步获取所有开关模态下的回路, 进而建立对应的开关模态回路集 P、$P_{S_i\text{-on}}$ 和 $P_{S_i\text{-off}}$。以邻接矩阵 $\boldsymbol{A}_{\text{dj}}$ 为例, 开关模态回路集 P 的获取方法如下: 首先, 根据变换器的 N_S 个开关, 枚举出 M 个开关模态, 得到开关模态集, 如下:

$$S_M=\{s_{\text{m1}},s_{\text{m2}},\cdots,s_{\text{m}M}\}\qquad M=2^{N_S}\tag{2}$$

然后, 根据每个开关模态下的开关状态, 修改 $\boldsymbol{A}_{\text{dj}}$ 得到所有开关模态下的邻接矩阵集 $\boldsymbol{A}_{\text{djSM}}$, 如下:

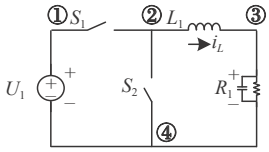
$$\boldsymbol{A}_{\text{djSM}}=\{A_{\text{djsm1}},A_{\text{djsm2}},\cdots,A_{\text{djsm}M}\}\tag{3}$$

任意开关模态 $s_{\text{m}m}$ 下的邻接矩阵 $\boldsymbol{A}_{\text{djsm}m}$ 通过以下方式得到: 若开关模态 $s_{\text{m}m}$ 下的导通开关的集合为 S_{on}, 关断开关的集合为 S_{off}, 将 $\boldsymbol{A}_{\text{dj}}$ 中 S_{on} 的开关赋值为 1, S_{off} 的开关赋值为 0, 就可获取该模态的邻接矩阵 $\boldsymbol{A}_{\text{djsm}m}$。最后, 将式 (3) 中的邻接矩阵逐一输入 DFS 算法, 得到 $\boldsymbol{A}_{\text{dj}}$ 在所有开关模态下的回路集, 如下:

$$\begin{aligned}P=\{&(p_{\text{sm1}}(1),p_{\text{sm1}}(2),\cdots),(p_{\text{sm2}}(1),p_{\text{sm2}}(2),\cdots),\cdots,\\&(p_{\text{sm}M}(1),p_{\text{sm}M}(2),\cdots)\}\end{aligned}\tag{4}$$

式中: $p_{\text{sm}m}(q)$ 为开关模态 $s_{\text{m}m}$ 的第 q 个回路。

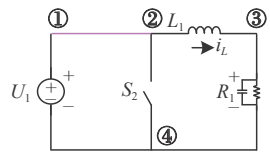
显然, 根据 $\boldsymbol{A}_{\text{dj}}$ 获取原变换器的开关模态回路集的方法同样适用于 $\boldsymbol{A}_{\text{dj}S_i\text{-on}}$ 和 $\boldsymbol{A}_{\text{dj}S_i\text{-off}}$。因此, 通过类似的过程就可以对应得到开关 S_i 简化后和开关



$\boldsymbol{A}_{\mathrm{dj}}$

	①	②	③	④
①	0	S_1	0	$U_{1\text{-pn}}$
②	S_1	0	L_1	S_2
③	0	L_1	0	$R_{1\text{-np}}$
④	$U_{1\text{-np}}$	S_2	$R_{1\text{-pn}}$	0

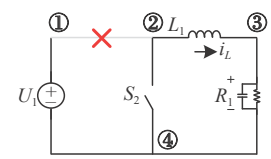
(a) Buck 变换器



$\boldsymbol{A}_{\mathrm{dj}S_1\text{-on}}$

	①	②	③	④
①	0	1	0	$U_{1\text{-pn}}$
②	1	0	L_1	S_2
③	0	L_1	0	$R_{1\text{-np}}$
④	$U_{1\text{-np}}$	S_2	$R_{1\text{-pn}}$	0

(b) 简化开关 S_1



$\boldsymbol{A}_{\mathrm{dj}S_1\text{-off}}$

	①	②	③	④
①	0	0	0	$U_{1\text{-pn}}$
②	0	0	L_1	S_2
③	0	L_1	0	$R_{1\text{-np}}$
④	$U_{1\text{-np}}$	S_2	$R_{1\text{-pn}}$	0

(c) 删除开关 S_1

图 4 原变换器邻接矩阵及简化开关 S_1 和删除开关 S_1 后的邻接矩阵

Fig. 4 Adjacent matrix of original converter and adjacent matrices of converters after switch S_1 are simplified or deleted

S_i 删除后变换器的开关模态回路集 $P_{S_i\text{-on}}$ 和 $P_{S_i\text{-off}}$。

3.3 有效开关模态筛选

在得到原变换器、开关 S_i 简化后和开关 S_i 删除后变换器的所有开关模态回路集 P、$P_{S_i\text{-on}}$ 和 $P_{S_i\text{-off}}$ 后，需要进一步筛选得到对应的有效开关模态回路集 P_{VSM}、$P_{\mathrm{VSM}\text{-}S_i\text{-on}}$ 和 $P_{\mathrm{VSM}\text{-}S_i\text{-off}}$。理论上，变换器的有效开关模态应满足以下电路基本原理，如图 5 所示。

1）基本原理 1。电压源不能短路或并联。

2）基本原理 2。电流源不能开路或串联。

为满足图 5 的基本原理 1，有效开关模态的变换器电路应满足以下 2 个条件：①输入端口、输出端口和电容在任一开关模态下不能短路，如图 6(a)～(c) 所示；②任意数量的输入端口和输出端口不能并联，如图 6(d) 所示。值得注意的是，以上 2 个条件也可以通过变换器的回路进行描述。即在任意有效开关模态下，变换器不允许存在以下回路：①一个输入端口组成的回路；②一个输出端口组成的回路；③一个电容组成的回路；④若干个输入端口和输出端口的回路。为方便自动判定有效开关模态，定义函数 g_C、g_U、g_R 和 g_L 分别用于统计开关模态 s_{mm} 下回路 $p_{\mathrm{smm}}(n_{lp})$ 中的电容数量 $N_{C\text{-smm}}$、输入端口数量 $N_{U\text{-smm}}$、输出端口数量 $N_{R\text{-smm}}$ 和电感数量 $N_{L\text{-smm}}$，如下：

$$\begin{cases} N_{C\text{-smm}}(n_{lp}) = g_C\left(p_{\mathrm{smm}}(n_{lp})\right) \\ N_{U\text{-smm}}(n_{lp}) = g_U\left(p_{\mathrm{smm}}(n_{lp})\right) \\ N_{R\text{-smm}}(n_{lp}) = g_R\left(p_{\mathrm{smm}}(n_{lp})\right) \\ N_{L\text{-smm}}(n_{lp}) = g_L\left(p_{\mathrm{smm}}(n_{lp})\right) \end{cases} \tag{5}$$

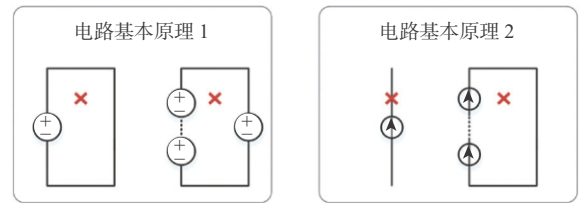
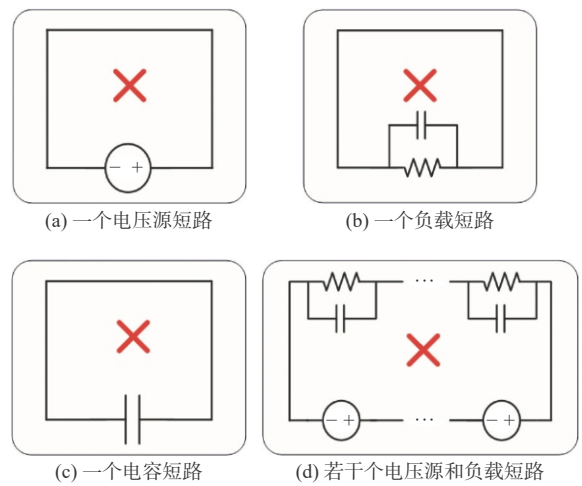


图 5 电路正常工作的基本原理

Fig. 5 Fundamental principles of workable circuits



(a) 一个电压源短路　(b) 一个负载短路

(c) 一个电容短路　(d) 若干个电压源和负载短路

图 6 基本原理 1 对应的 4 种无效开关模态

Fig. 6 Four kinds of invalid switching modes based on fundamental principles of workable circuits 1

对于包含若干个回路的开关模态 s_{mm}，可以通过约束 1 判断该模态的有效性。

1）约束 1。若开关模态 s_{mm} 存在一个回路 $p_{\mathrm{smm}}(n_{lp})$，使得式 (5) 中统计的 $N_{C\text{-smm}}(n_{lp})$、$N_{U\text{-smm}}(n_{lp})$、$N_{R\text{-smm}}(n_{lp})$ 和 $N_{L\text{-smm}}(n_{lp})$ 满足式 (6)，则开关模态 s_{mm} 无效。

$$\begin{cases}\{N_{C\text{-}smm}(n_{lp})=1,\ N_{U\text{-}smm}(n_{lp})=0,\ N_{R\text{-}smm}(n_{lp})=0,\\ \quad N_{L\text{-}smm}(n_{lp})=0\}\\ \{N_{C\text{-}smm}(n_{lp})=0,\ N_{U\text{-}smm}(n_{lp})=1,\ N_{R\text{-}smm}(n_{lp})=0,\\ \quad N_{L\text{-}smm}(n_{lp})=0\}\\ \{N_{C\text{-}smm}(n_{lp})=0,\ N_{U\text{-}smm}(n_{lp})=0,\ N_{R\text{-}smm}(n_{lp})=1,\\ \quad N_{L\text{-}smm}(n_{lp})=0\}\\ \{N_{C\text{-}smm}(n_{lp})=0,\ N_{U\text{-}smm}(n_{lp})+N_{R\text{-}smm}(n_{lp})\geqslant 2,\\ \quad N_{L\text{-}smm}(n_{lp})=0\}\end{cases} \tag{6}$$

为满足图 5 的电路基本原理 2, 有效开关模态下任一电感应包含至少一个回路。因此, 对于含 N_L 个电感的变换器, 若其开关模态 s_{mm} 包含 N_{lp} 个回路, 则可通过约束 2 进一步判断该模态的有效性。

2) 约束 2。开关模态 s_{mm} 中, 若存在一个电感 L_x 不包含于 $\{p_{smm}(1), p_{smm}(2), \cdots, p_{smm}(N_{lp})\}$ 中的任何一个回路, 则开关模态 s_{mm} 无效。

3.4　程序实现

根据 3.1 节～3.3 节获取任意邻接矩阵 $\boldsymbol{A}_{dj}$ 的有效开关模态回路集的流程如图 7 所示。

步骤 1　初始化开关模态编号 m 和有效开关模态编号 k 为 1。

步骤 2　根据式 (2) 枚举出邻接矩阵 $\boldsymbol{A}_{dj}$ 对应变换器的所有开关模态 $\{s_{m1}, s_{m2}, \cdots, s_{mM}\}$, 同时获取所有开关模态下的邻接矩阵 $\{\boldsymbol{A}_{djsm1}, \boldsymbol{A}_{djsm2}, \cdots, \boldsymbol{A}_{djsmM}\}$。

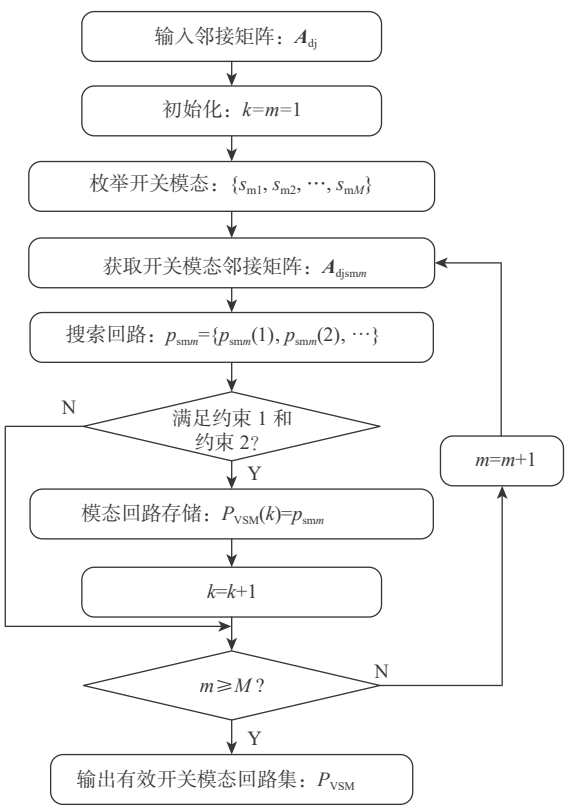


图 7　有效开关模态回路集获取流程

Fig. 7　Flowchart of obtaining loop set of valid switching modes

步骤 3　通过 DFS 算法获取开关模态 s_{mm} 下的所有回路 $p_{smm}=\{p_{smm}(1), p_{smm}(2), \cdots\}$。

步骤 4　若 p_{smm} 满足约束 1 和约束 2, 开关模态 s_{mm} 判定为有效, 同时执行 $P_{VSM}(k)=p_{smm}$。

步骤 5　执行 $k=k+1$, $m=m+1$, 并重复步骤 2 ～步骤 5, 直到 $m\geqslant M$ 时, 输出 P_{VSM}。

基于图 7 的流程建立函数 $h(\boldsymbol{A}_{dj})$, 以邻接矩阵 $\boldsymbol{A}_{dj}$ 作为输入, $h(\boldsymbol{A}_{dj})$ 可以自动获取 $\boldsymbol{A}_{dj}$ 的有效开关模态回路集。以原变换器的邻接矩阵 $\boldsymbol{A}_{dj}$, 以及开关 S_i 简化/删除后的邻接矩阵 $\boldsymbol{A}_{djS_i\text{-}on}$ 和 $\boldsymbol{A}_{djS_i\text{-}off}$ 分别作为函数 h 输入, 可以得到对应的开关模态回路集 P、$P_{S_i\text{-}on}$ 和 $P_{S_i\text{-}off}$。基于以上获取的回路集, 结合图准则可以判定开关 S_i 是否为冗余开关。对于包含 N_S 个开关 $\{S_1, S_2, \cdots, S_{N_S}\}$ 的原变换器, 只需将所有开关循环代入式 (7) 中的 S_i, 并结合图准则, 可以自动识别出原变换器的所有冗余开关。

$$\begin{cases}P_{VMS}=h(\boldsymbol{A}_{dj})\\ P_{VMS\text{-}S_i\text{-}on}=h(\boldsymbol{A}_{djS_i\text{-}on})\\ P_{VMS\text{-}S_i\text{-}off}=h(\boldsymbol{A}_{djS_i\text{-}off})\end{cases} \tag{7}$$

4　冗余开关识别案例

为验证本文提出的冗余开关自动识别方法的可行性, 将其用于识别图 1 左侧变换器的冗余开关。首先, 建立图 1 左侧变换器的邻接矩阵 $\boldsymbol{A}_{dj1}$, 如下:

$$\boldsymbol{A}_{dj1}=\begin{bmatrix}0 & S_1 & 0 & 0 & S_3 & S_4\\ S_1 & 0 & S_2 & L & 0 & 0\\ 0 & S_2 & 0 & 0 & 0 & P_{2\text{-}pn}\\ 0 & 0 & 0 & 0 & 0 & P_{1\text{-}pn}\\ S_3 & 0 & 0 & 0 & 0 & P_{3\text{-}pn}\\ S_4 & 0 & P_{2\text{-}np} & P_{1\text{-}np} & P_{3\text{-}np} & 0\end{bmatrix} \tag{8}$$

然后, 将 $\boldsymbol{A}_{dj1}$ 输入式 (7) 的函数 h, 得到有效开关模态回路集 P_{VSM}。如表 1 所示, P_{VSM} 中包含了 3 种有效开关模态 v_{sm1}～v_{sm3} 的回路 p_{sm1}～p_{sm3}, 且 v_{sm1}～v_{sm3} 分别对应 $\{S_2$ 单独导通$\}$、$\{S_1$ 和 S_3 导通$\}$和 $\{S_1$ 和 S_4 导通$\}$。进一步, 将开关 S_1～S_4 分别简化为导线后, 得到开关简化后的邻接矩阵 $\boldsymbol{A}_{dj1\text{-}S_1\text{-}on}$～$\boldsymbol{A}_{dj1\text{-}S_4\text{-}on}$, 并根据式 (7) 得到这 4 个邻接矩阵对应的有效开关模态回路集 $P_{VSM\text{-}S_1\text{-}on}$～$P_{VSM\text{-}S_4\text{-}on}$。最后, 通过对比表 1 中的 P_{VSM} 和 $P_{VSM\text{-}S_1\text{-}on}$～$P_{VSM\text{-}S_4\text{-}on}$ 发现, $P_{VSM}=P_{VSM\text{-}S_1\text{-}on}$ 且 $P_{VSM}\neq P_{VSM\text{-}S_2\text{-}on}$～$P_{VSM\text{-}S_4\text{-}on}$。由此可以判定图 1 左侧变换器中 S_1 为导通冗余开关, S_2～S_4 不是导通冗余开关。

类似地, 将本文方法用于自动识别图 2 左侧变换器的关断冗余开关。首先, 同样获取该变换器的邻接矩阵 $\boldsymbol{A}_{dj2}$, 如下:

表1 图1变换器的导通冗余开关识别结果

Table 1 Identification results of turn-on redundant switch for converter in Fig. 1

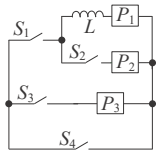
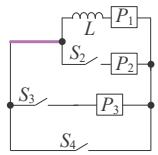
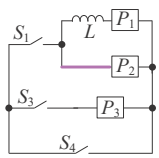
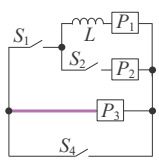
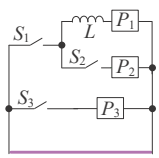
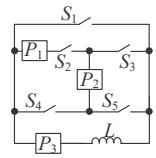
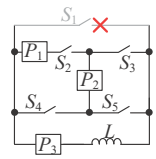
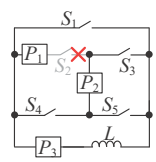
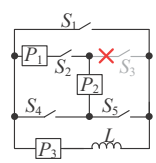
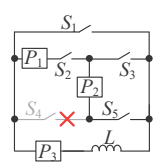
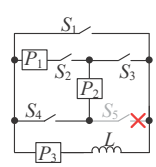
变化器简图	有效开关模态	回路集与识别结果
 原变换器	$v_{sm1}=\{S_2\}$ $v_{sm2}=\{S_1, S_3\}$ $v_{sm3}=\{S_1, S_4\}$	$P_{VSM}=\{p_{vsm1}=(L,P_{1\text{-pn}},P_{2\text{-np}}),$ $p_{vsm2}=(L,P_{1\text{-pn}},P_{3\text{-np}}),p_{vsm3}=$ $(L,P_{1\text{-pn}})\}$
 简化S_1	$v_{sm1}=\{S_2\}$ $v_{sm2}=\{S_3\}$ $v_{sm3}=\{S_4\}$	$P_{VSM\text{-}S_1\text{-on}}=\{p_{vsm1}=(L,P_{1\text{-pn}},P_{2\text{-np}}),$ $p_{vsm2}=(L,P_{1\text{-pn}},P_{3\text{-np}}),p_{vsm3}=$ $(L,P_{1\text{-pn}})\}$ $P_{VMS}=P_{VSM\text{-}S_1\text{-on}}$，$S_1$导通冗余
 简化S_2	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_3\}$ $v_{sm3}=\{S_4\}$	$P_{VSM\text{-}S_2\text{-on}}=\{p_{vsm1}=(L,P_{1\text{-pn}},P_{2\text{-np}}),$ $p_{vsm2}=(L,P_{1\text{-pn}},P_{2\text{-np}}),p_{vsm3}=$ $(L,P_{1\text{-pn}},P_{2\text{-np}})\}$ $P_{VMS}\neq P_{VSM\text{-}S_2\text{-on}}$，$S_2$不是导通冗余
 简化S_3	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_2\}$	$P_{VSM\text{-}S_3\text{-on}}=\{p_{vsm1}=(L,P_{1\text{-pn}},P_{3\text{-np}}),$ $p_{vsm2}=(L,P_{1\text{-pn}},P_{2\text{-np}})\}$ $P_{VSM}\neq P_{VSM\text{-}S_3\text{-on}}$，$S_3$不是导通冗余
 简化S_4	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_2\}$	$P_{VSM\text{-}S_4\text{-on}}=\{p_{vsm1}=(L,P_{1\text{-pn}}),$ $p_{vsm2}=(L,P_{1\text{-pn}},P_{2\text{-np}})\}$ $P_{VMS}\neq P_{VSM\text{-}S_4\text{-on}}$，$S_4$不是导通冗余

表2 图2变换器的关断冗余开关识别结果

Table 2 Identification results of turn-off redundant switch for converter in Fig. 2

变化器简图	有效开关模态	回路集与识别结果
 原变换器	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_2, S_3\}$ $v_{sm3}=\{S_2, S_5\}$ $v_{sm4}=\{S_3, S_4\}$	$P_{VSM}=\{p_{vsm1}=(L,P_{3\text{-pn}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{1\text{-np}}),$ $p_{vsm3}=(L,P_{3\text{-pn}},P_{2\text{-pn}},P_{1\text{-np}}),$ $p_{vsm4}=(L,P_{3\text{-pn}},P_{2\text{-np}})\}$
 删除S_1	$v_{sm1}=\{S_4, S_5\}$ $v_{sm2}=\{S_2, S_3\}$ $v_{sm3}=\{S_2, S_5\}$ $v_{sm4}=\{S_3, S_4\}$	$P_{VSM\text{-}S_1\text{-off}}=\{p_{vsm1}=(L,P_{3\text{-pn}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{1\text{-np}}),$ $p_{vsm3}=(L,P_{3\text{-pn}},P_{2\text{-pn}},P_{1\text{-np}}),$ $p_{vsm4}=(L,P_{3\text{-pn}},P_{2\text{-np}})\}$ $P_{VSM}=P_{VSM\text{-}S_1\text{-off}}$，$S_1$关断冗余
 删除S_2	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_3, S_4\}$	$P_{VSM\text{-}S_2\text{-off}}=\{p_{vsm1}=(L,P_{3\text{-pn}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{2\text{-np}})\}$ $P_{VSM}\neq P_{VSM\text{-}S_2\text{-off}}$，$S_2$不是关断冗余
 删除S_3	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_2, S_5\}$	$P_{VSM\text{-}S_3\text{-off}}=\{p_{vsm1}=(L,P_{3\text{-pn}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{2\text{-pn}},P_{1\text{-np}})\}$ $P_{VSM}\neq P_{VSM\text{-}S_3\text{-off}}$，$S_3$不是关断冗余
 删除S_4	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_2, S_3\}$ $v_{sm3}=\{S_2, S_5\}$	$P_{VSM\text{-}S_4\text{-off}}=\{p_{vsm1}=(L,P_{3\text{-pn}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{1\text{-np}}),$ $p_{vsm3}=(L,P_{3\text{-pn}},P_{2\text{-pn}},$ $P_{1\text{-np}})\}$ $P_{VSM}\neq P_{VSM\text{-}S_4\text{-off}}$，$S_4$不是关断冗余
 删除S_5	$v_{sm1}=\{S_1\}$ $v_{sm2}=\{S_2, S_3\}$ $v_{sm3}=\{S_3, S_4\}$	$P_{VSM\text{-}S_5\text{-off}}=\{p_{vsm1}=(L,P_{3\text{-pn}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{1\text{-np}}),$ $p_{vsm2}=(L,P_{3\text{-pn}},P_{2\text{-np}})\}$ $P_{VSM}\neq P_{VSM\text{-}S_5\text{-off}}$，$S_5$不是关断冗余

$$\boldsymbol{A}_{dj2}=\begin{bmatrix} 0 & P_{3\text{-pn}} & S_1 & S_4 & P_{1\text{-pn}} & 0 \\ P_{3\text{-np}} & 0 & L & 0 & 0 & 0 \\ S_1 & 0 & 0 & S_5 & 0 & S_3 \\ S_4 & 0 & S_5 & 0 & 0 & P_{2\text{-pn}} \\ P_{1\text{-np}} & 0 & 0 & S_{np} & 0 & S_2 \\ 0 & 0 & S_3 & P_{2\text{-np}} & S_2 & 0 \end{bmatrix} \tag{9}$$

其次，将 $\boldsymbol{A}_{dj2}$ 输入式(7)的函数 h，得到有效开关模态回路集 P_{VSM}。与 $\boldsymbol{A}_{dj1}$ 得到的有效开关模态回路集不同，$\boldsymbol{A}_{dj2}$ 得到的有效开关模态回路集包含了4种有效开关模态 v_{sm1}～v_{sm4} 的回路，分别对应{S_1单独导通}、{S_2和S_3导通}、{S_2和S_5导通}和{S_3和S_4导通}，如表2所示。然后，将变换器开关 S_1～S_5 分别删除后，得到开关删除后对应的邻接矩阵 $\boldsymbol{A}_{dj1\text{-}S_1\text{-off}}$～$\boldsymbol{A}_{dj1\text{-}S_5\text{-off}}$。进一步，基于式(7)得到 $\boldsymbol{A}_{dj2}$ 和 $\boldsymbol{A}_{dj1\text{-}S_1\text{-off}}$～$\boldsymbol{A}_{dj1\text{-}S_5\text{-off}}$ 对应的有效开关模态回路集 P_{VSM} 和 $P_{VSM\text{-}S_1\text{-off}}$～$P_{VSM\text{-}S_5\text{-off}}$，如表2所示。最后，根据 $P_{VSM}=P_{VSM\text{-}S_1\text{-off}}$ 且 $P_{VSM}\neq P_{VSM\text{-}S_2\text{-off}}$～$P_{VSM\text{-}S_5\text{-off}}$ 判定 S_1 是关断冗余开关，S_2～S_5 不是关断冗余开关。

为进一步验证冗余开关自动识别方法的有效性，将其用于文献[10]推演的单电感多端口变换器。图8为识别出的20个含有冗余开关的变换器。图中：灰色底纹表示导通冗余开关、虚线框表示关断冗余开关。通过分析这20个变换器的工作模态可知，将导通冗余开关简化为导线、关断冗余开关删除后，不影响变换器的工作性能，验证了冗余开关识别的正确性。在图8的20个变换器中，图8(c)～(e)、(i)～(m)、(r)是只含导通

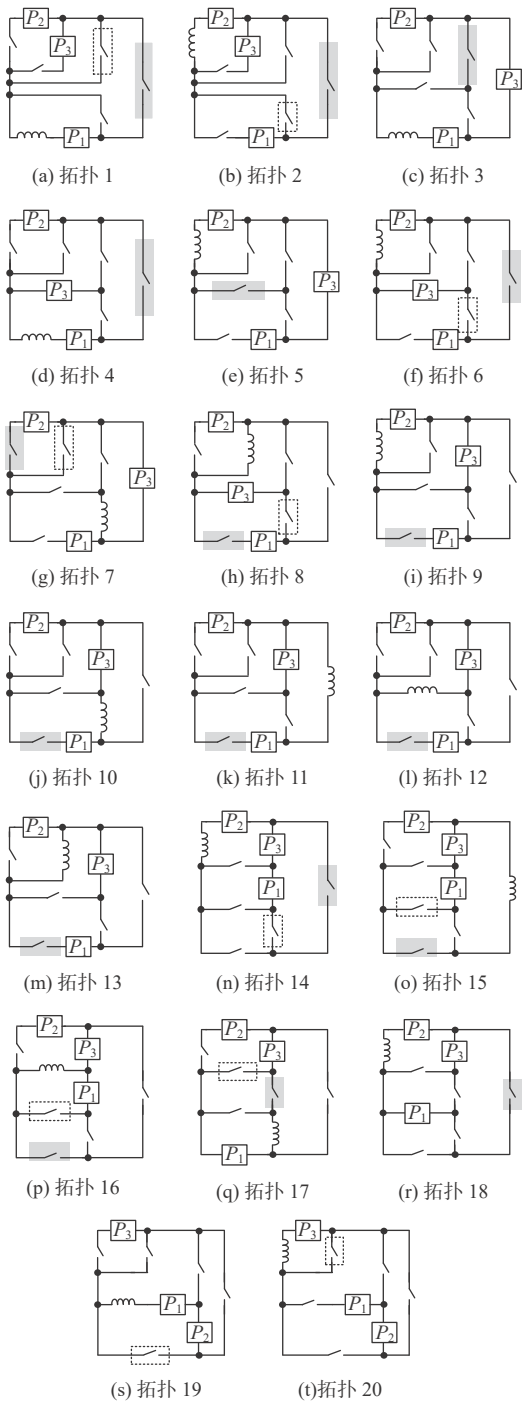


图 8　自动识别出的 20 个含冗余开关的单电感多端口变换器

Fig. 8　20 automatically identified single-inductor multi-port converters with redundant switches

冗余开关的 9 个变换器, 图 8(s)、(t) 是只包含关断冗余开关的 2 个变换器, 图 8(a)、(b)、(f)～(h)、(n)～(q) 是同时包含导通冗余和关断冗余的 9 个变换器。根据图 8 的冗余开关识别结果可知, 本文方法可以在计算机辅助下快速、准确地识别出变换器中的所有冗余开关。

5　结　论

1) 本文提出了一种基于图论的冗余开关自动识别方法, 通过深入分析变换器的基本工作原理, 归纳出冗余开关的识别准则。

2) 基于冗余开关识别准则, 利用图论将识别准则转化为图准则, 并结合深度优先图搜索算法, 实现了冗余开关的快速、批量化识别。

3) 研究结果表明, 该方法能够精准有效地识别变换器中的所有冗余开关, 为变换器拓扑的优化设计和简化提供了理论依据和技术支持。未来研究可进一步优化算法性能, 以适应更加复杂的拓扑结构和大规模数据分析需求。

参考文献（References）

[1] TANG Z T, YANG Y H, BLAABJERG F. Power electronics: the enabling technology for renewable energy integration[J]. CSEE Journal of Power and Energy Systems, 2022, 8(1): 39-52.

[2] EHSANI M, SINGH K V, BANSAL H O, et al. State of the art and trends in electric and hybrid electric vehicles[J]. Proceedings of the IEEE, 2021, 109(6): 967-984.

[3] PENG W Y, DU S J. The advances in conversion techniques in triboelectric energy harvesting: a review[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2023, 70(7): 3049-3062.

[4] KIM J S, YOON J O, CHOI B D. A high-light-load-efficiency low-ripple-voltage PFM buck converter for IoT applications[J]. IEEE Transactions on Power Electronics, 2022, 37(5): 5763-5772.

[5] CHENG K W E, YE Y M. Duality approach to the study of switched-inductor power converters and its higher-order variations [J]. IET Power Electronics, 2015, 8(4): 489-496.

[6] GNANASAMBANDAM K, RATHORE A K, EDPUGANTI A, et al. Current-fed multilevel converters: an overview of circuit topologies, modulation techniques, and applications[J]. IEEE Transactions on Power Electronics, 2017, 32(5): 3382-3401.

[7] 胡雪峰, 王琳, 代国瑞, 等. 单开关高增益 Boost-Sepic 集成变换器[J]. 中国电机工程学报, 2015, 35(8): 2018-2025.
HU X F, WANG L, DAI G R, et al. High step-up Boost-Sepic integrated converters with a single switch[J]. Proceedings of the CSEE, 2015, 35(8): 2018-2025(in Chinese).

[8] 荣德生, 孙瑄璠. 高增益耦合电感组合 Buck-Boost-Zeta 变换器[J]. 中国电机工程学报, 2020, 40(14): 4590-4601.
RONG D S, SUN X J. High gain coupled inductance combination Buck-Boost-Zeta converter[J]. Proceedings of the CSEE, 2020, 40 (14): 4590-4601(in Chinese).

[9] DE ANDRADE J M, SALVADOR M A, COELHO R F, et al. General method for synthesizing high gain step-up DC-DC converters based on differential connections[J]. IEEE Transactions on Power Electronics, 2020, 35(12): 13239-13254.

[10] MO L P, CHEN G P, HUANG J M. From components to converters: a fundamental topology derivation method for single-inductor multi-input multi-output converters based on graph theory[C]// Proceedings of the IEEE International Conference on Power Electronics, Computer Applications. Piscataway: IEEE Press, 2021: 458-463.

[11] DONG M, LIANG R J, YANG J, et al. Topology derivation of multi-port DC-DC converters based on reinforcement learning[J]. IEEE

Transactions on Power Electronics, 2023, 38(4): 5055-5064.

[12] HUANG Z Q, ZHOU D H, SHEN Z W, et al. Directed graph-based topology derivation method for single-stage multiport inverters[J]. IEEE Transactions on Power Electronics, 2023, 38(11): 14614-14627.

[13] CHEN Y, BAI J B, KANG Y. A nonisolated single-inductor multiport DC-DC topology deduction method based on reinforcement learning[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2022, 10(6): 6572-6585.

[14] LI H, LI Y M, WANG W C, et al. Programmable topology deduction algorithm for S1D2C1L1 type DC-DC converters based on graph theory[C]//Proceedings of the 14th International FLINS Conference. Singapore: World Scientific, 2020.

[15] 李虹, 王文财, 李亚敏, 等. 基于图论的 $S^1D^1L^2C^1$ 型 DC-DC 变换器可编程拓扑搜索算法[J]. 中国电机工程学报, 2021, 41(16): 5670-5683.
LI H, WANG W C, LI Y M, et al. Programmable topology searching algorithm for $S^1D^1L^2C^1$ type DC-DC converters based on graph theory[J]. Proceedings of the CSEE, 2021, 41(16): 5670-5683(in Chinese).

[16] CHEN G P, JIN Z F, LIU Y W, et al. Programmable topology derivation and analysis of integrated three-port DC-DC converters with reduced switches for low-cost applications[J]. IEEE Transactions on Industrial Electronics, 2019, 66(9): 6649-6660.

[17] MO L P, HUANG J M, CHEN G P, et al. Computer-aided systematic topology derivation of single-inductor multi-input multi-output converters from working principle[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2022, 69(6): 2637-2649.

[18] ERICKSON R W, MAKSIMOVIĆ D. Fundamentals of power electronics[M]. Berlin: Springer, 2001.

[19] 张永, 李睿, 年福忠. 算法与数据结构[M]. 北京: 国防工业出版社, 2008.
ZHANG Y, LI R, NIAN F Z. Algorithms and data structures[M]. Beijing: National Defense Industry Press, 2008(in Chinese).

[20] XIE B Z, QI C, BEN H Q, et al. The applications of graph theory in electric network[C]//Proceedings of the International Conference on Sensing, Diagnostics, Prognostics, and Control. Piscataway: IEEE Press, 2019: 780-784.

Auto identification method for redundant switches in converters based on graph theory

MO Liping[1, 2], SONG Jixiang[1, 2], CHEN Guipeng[1, 2, *]

(1. Shenzhen Research Institute of Xiamen University, Shenzhen 518000, China;
2. School of Aerospace Engineering, Xiamen University, Xiamen 361000, China)

Abstract: Up to now, massive power electronics converters have been derived to satisfy various demands in practical applications. However, most of these methods focus on discovering effective converters, ignoring the redundant switches in converters. This paper aims to delete converters with redundant switches efficiently. First, the identification principle is introduced based on the connection between legitimate switching modes and redundant switches. After that, the principle is transformed into the corresponding graph principle. Finally, graph theory and graph searching algorithms are used to develop the auto identification method based on the graph principle. The results demonstrate that all redundant switches may be effectively recognized with the use of computers when the suggested method is applied to single-inductor multi-port converters.

Keywords: graph theory; converters; redundant switches; auto identification; switching modes

Received: 2023-06-14; **Accepted**: 2023-10-11; **Published Online**: 2024-01-23 17:35

URL: link.cnki.net/urlid/11.2625.V.20240119.1136.002

Foundation items: Guangdong Basic and Applied Basic Research Foundation (2022A1515011616); The Fundamental Research Funds for the Central Universities (20720220083)

* **Corresponding author**. E-mail: cgp2017@xmu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0356

基于差分进化算法的SIMO Buck变换器最优动态响应搜索

李丹[1, 2]，崔文峰[3]，陈桂鹏[1, 2, *]

(1. 厦门大学深圳研究院，深圳 518000；2. 厦门大学 航空航天学院，厦门 361000；3. 浙江飞航智能科技有限公司，杭州 311100)

摘 要：随着单电感多输出(SIMO)直流变换器的广泛应用，其动态响应问题备受关注。为探究SIMO Buck变换器最优动态响应的理论极限，根据变换器工作原理建立数学模型，利用改进的差分进化(DE)算法进行搜索求解。所提方法可以搜索求解出不同目标下的变换器最优动态响应要求，如最优的自调节或交叉调节，还可以获得不同约束下的最优动态响应，如不同的动态响应时间、峰值电感电流。基于启发式DE算法的最优动态响应理论极限搜索，有助于全面了解SIMO直流变换器的动态性能，并指导控制器设计以优化动态响应过程。

关 键 词：单电感多输出直流变换器；动态响应；差分进化算法；自调节；交叉调节
中图分类号：TM461
文献标志码：A **文章编号**：1001-5965(2025)05-1457-12

许多电子设备需要提供多个输出电压端口，为不同的功能模块供电。相比于应用多个独立的单输出直流变换器实现多路输出的传统方案，单电感多输出(single-inductor multiple-output, SIMO)直流变换器具有功率密度高、体积小、成本低等优点，受到了广泛关注。但是，由于SIMO直流变换器各输出支路共享一个电感，当一条输出支路的负载变化时，不仅会影响其自身输出电压，而且还将通过电感电流导致其他输出支路的电压变化，即存在自调节和交叉调节问题[1-3]。目前，为抑制SIMO直流变换器输出支路间的交叉调节问题，提高动态响应性能，许多文献提出了有效的解决方案。

文献[4-6]提出的分时复用控制技术使输出支路的充放电过程完全隔离，消除了交叉调节影响，但该技术仅对工作于断续导通模式(discontinuous conduction mode, DCM)和伪连续导通模式(pseudo-continue conduction mode, PCCM)的SIMO直流变换器有效。由于连续导通模式(continue conduction mode, CCM)相比DCM和PCCM具有电压、电流纹波小和效率高的优势，工作于CCM的交叉调节影响抑制问题更受关注。文献[7]针对工作于CCM的单电感双输出直流变换器，提出了共模-差模电压型控制技术，改善了交叉调节影响，但动态响应速度较慢。文献[8]采用峰值电流控制技术，通过提高变换器的动态响应速度减小了交叉调节影响。由于电容电流相比电感电流可以更快地反映负载变化，文献[9-10]采用电容电流纹波控制技术，进一步提高了动态响应速度，但负载稳定工作范围有限。文献[11-13]采用解耦控制方法减小交叉调节，其中，文献[11]研究输出电流前馈控制来解耦交叉调节传递函数，文献[12]通过交叉引入输出状态变量的导数来解耦，但控制器的自环和交叉环补偿是针对特定工作点设计的，因此，工作点的改变会对动态响应性能造成不利影响，文献[13]采用多变量数字电压型控制方法，可以同时考虑电

收稿日期：2023-06-14；**录用日期**：2023-09-27；**网络出版时间**：2024-01-12 17:46
网络出版地址：link.cnki.net/urlid/11.2625.V.20240112.1130.001
基金项目：广东省基础与应用基础研究基金(2022A1515011616)；国家自然科学基金(51907172)
* **通信作者**. E-mail：cgp2017@xmu.edu.cn

引用格式：李丹，崔文峰，陈桂鹏. 基于差分进化算法的SIMO Buck变换器最优动态响应搜索[J]. 北京航空航天大学学报，2025，51(5)：1457-1468. LI D，CUI W F，CHEN G P. Optimal dynamic response exploration for SIMO Buck converter based on differential evolution algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1457-1468 (in Chinese).

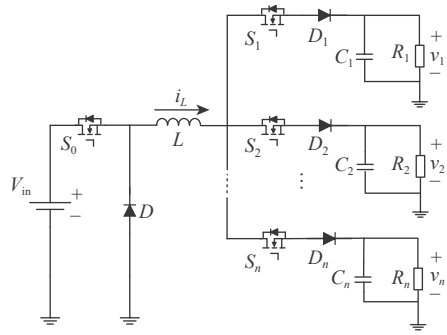
路参数和输入电压变化的多个工作点的解耦，但解耦控制方法实现过程比较复杂。不同于基于PI控制的线性调节，文献[14-15]使用高速电压比较器控制，动态响应非常快，减小了交叉调节影响，但最后一个输出仍需要PI控制来提供足够的能量，因此，交叉调节影响会累积在最后一个输出。预测控制作为一种新型控制算法，适用于非线性、强耦合的控制问题。文献[16-19]将模型预测控制应用于SIMO直流变换器，方便处理各支路的耦合，动态响应快速。但是，为减小在线计算负担，控制信号只有0和1这2种情况，因此，电压波动较大，动态响应效果受到影响。

根据上述分析可知，虽然目前国内外相关研究已经提出了很多方法来提高SIMO直流变换器的动态响应性能，但其最优动态响应效果仍未可知。为明确SIMO Buck直流变换器的最优动态响应理论极限，本文结合数学建模和启发式智能寻优算法进行求解。首先，对SIMO Buck直流变换器的工作原理进行详细的分析，并获得时域解析式，据此建立最优动态响应搜索的数学模型。然后，详细介绍利用改进的差分进化(differential evolution, DE)算法进行求解的具体步骤。最后，以单电感三路输出Buck变换器为例，分别探讨了不同目标、动态响应时间、电感电流约束下的变换器最优动态响应理论极限，验证了本文方法的有效性。

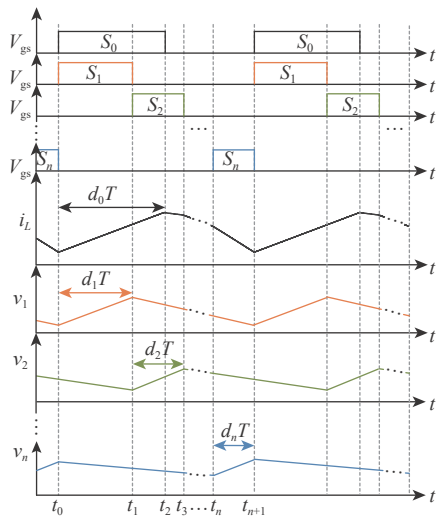
1 SIMO Buck变换器最优动态响应的数学模型

1.1 变换器工作原理

SIMO Buck变换器的电路结构如图1(a)所示，其由输入电压源V_{in}、电感L、主开关管S_0、输出支路开关管$S_1,S_2,\cdots,S_n$、二极管$D,D_1,D_2,\cdots,D_n$、输出电容$C_1,C_2,\cdots,C_n$及负载电阻$R_1,R_2,\cdots,R_n$组成。在任意时刻，支路开关管$S_1,\cdots,S_n$有且仅有一个导通。因此，根据主开关管S_0的不同开通关断状态，当SIMO Buck变换器工作于CCM时，存在$2n$种工作模态，如图2所示。在工作模态$2q-1$(q=1,2,$\cdots$,n)，主开关管S_0和输出支路开关管S_q导通，此时能量从输入传递到电感L和第q路输出v_q上，电感电流i_L上升，输出电压v_q增大，其他路输出电压减小，数学表达式分别如式(1)和(2)所示。在工作模态$2q$(q=1,2,$\cdots$,n)，主开关管S_0关断，此时二极管D和输出支路开关管S_q导通，电感能量传递给第q路输出v_q上，电感电流i_L下降，输出电压v_q增大，其他路输出电压减小，如式(1)和(2)所示。



(a) SIMO Buck变换器



(b) 工作波形

图1 SIMO Buck变换器和工作波形

Fig. 1 SIMO Buck converter and operation waveform

$$i_L(t)=\begin{cases}V_{\text{in}}/R_q+\sum_{g=1}^{2}(A_{g_q}/R_q+C_qA_{g_q}P_{g_q})\mathrm{e}^{P_{g_q}t} & \text{模态}2q-1\\ \sum_{g=1}^{2}(A_{g_q}/R_q+C_qA_{g_q}P_{g_q})\mathrm{e}^{P_{g_q}t} & \text{模态}2q\end{cases} \tag{1}$$

$$v_m(t)=\begin{cases}V_{\text{in}}+\sum_{g=1}^{2}A_{g_q}\mathrm{e}^{P_{g_q}t} & \text{模态}2q-1\text{且}m=q\\ \sum_{g=1}^{2}A_{g_q}\mathrm{e}^{P_{g_q}t} & \text{模态}2q\text{且}m=q\\ v_m(0_+)\mathrm{e}^{-t/(R_mC_m)} & m\neq q\end{cases} \tag{2}$$

式中：

$$P_{1_q}=-1/(2R_qC_q)+\sqrt{[1/(2R_qC_q)]^2-1/(LC_q)}$$

$$P_{2_q}=-1/(2R_qC_q)-\sqrt{[1/(2R_qC_q)]^2-1/(LC_q)}$$

$$A_{1_q}=[P_{2_q}(v_q(0_+)-V_{\text{in}})-(i_L(0_+)R_q-v_q(0_+))/(R_qC_q)]/(P_{2_q}-P_{1_q})$$

$$A_{2_q}=[P_{1_q}(v_q(0_+)-V_{\text{in}})-(i_L(0_+)R_q-v_q(0_+))/(R_qC_q)]/(P_{1_q}-P_{2_q})$$

其中：$v_q(0_+)$和$i_L(0_+)$分别为工作模态输出电压和电感电流初始值。

在一个开关周期内，根据开关管$S_0,S_1,\cdots,S_n$

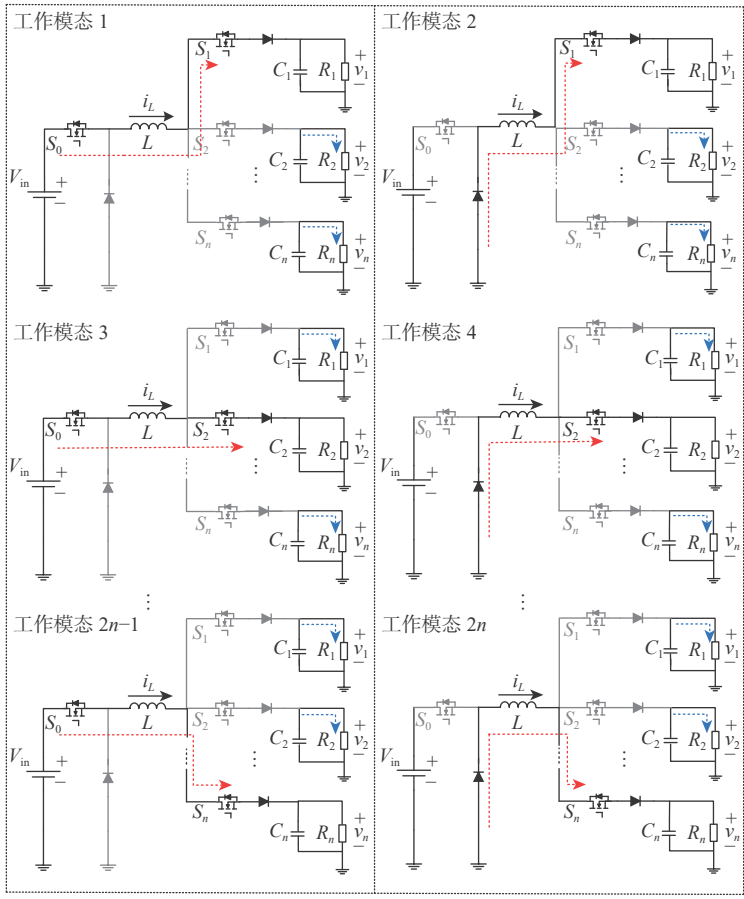


图 2　SIMO Buck 变换器的工作模态

Fig. 2　Operation mode of SIMO Buck converter

导通占空比 $d_0,d_1,\cdots,d_n$ 的大小关系, SIMO Buck 变换器将会出现上述 $2n$ 种工作模态中的 n+1 种。由各工作模态的电感电流和输出电压的时域解析表达式, 可以计算出任何一组占空比 d_0～d_n 下电感电流和输出电压的变化。具体来说, 当 $d_1+d_2+\cdots+d_k<d_0<d_1+d_2+\cdots+d_k+d_{k+1}$($k$=1,2,⋯,$n$−1) 时, 变换器将会出现模态 1,3,⋯,2$k$+1,2($k$+1),2($k$+2),⋯,2$n$。以 k=1 为例, 即 $d_1<d_0<d_1+d_2$, 图 1(b) 给出了开关管的驱动信号及电感电流 i_L 和输出电压 v_1～v_n 的波形示意图。t_0～t_1 和 t_1～t_2 阶段分别对应于工作模态 1 和工作模态 3; t_2～t_3,t_3～t_4,⋯,t_n～t_{n+1} 阶段分别对应于工作模态 4, 工作模态 6, ⋯, 工作模态 2n。因此, 电感电流 i_L 在 t_0～t_2 期间上升, 在其他时间下降; 输出电压 v_1 在 t_0～t_1 期间上升, 输出电压 v_2 在 t_1～t_3 期间上升, 输出电压 v_m(m=3, 4, ⋯, n) 在 t_m～t_{m+1} 期间上升, 在其他时间下降。值得注意的是, 如果电感电流 i_L 在工作模态 2q 中降至 0, 变换器将进入 DCM。此时, 电感的电流保持为 0, 直到再次被充电, 各输出支路的电容放电供给相应的负载, 输出电压解析表达式为 $v_m = v_m(t \mid i_L = 0)\mathrm{e}^{-\tau/R_mC_m}$。

1.2　最优动态响应的数学模型

由 SIMO Buck 变换器电感电流和输出电压的解析表达式 (1)、(2) 可知, 当某一支路负载 R_q(q=1,2,⋯,n) 发生变化时, 不仅该支路的电压 v_q 会发生变化, 其他支路电压也会受影响而变化, 即变换器不仅存在负载调整率问题, 还存在交叉调节影响。理论上, 通过调节占空比 d_0～d_n 来控制电感电流和输出支路电压的变化, 可使变换器回到新稳态。

在不同的占空比调节下, 输出支路电压的调整过程不同, 即变换器的动态响应不同。如图 3(a) 所示, 当采用传统的 PI 控制时, 占空比 d_0～d_n 在负载 R_1 变化后缓慢变化, 输出电压 v_1～v_n 也随之缓慢调整, 直至恢复至额定电压值。当采用文献 [13] 中多变量控制方法, 占空比 d_1～d_n 的初始调整速度相比 PI 控制较快, 因此, 在动态响应过程中, 输出电压 v_1～v_n 的波动减小, 变换器的动态响应得以改善。除此之外, 为改善 SIMO 变换器的动态响应性能, 国内外学者已开展了众多相关研究[4-19], 提出多种控制策略。然而, 由于以上研究均未对变换器动态响应的本质数学模型开展研究, 截至目前, SIMO Buck 变换器的最优动态响应极限尚未清楚。

本质上, SIMO Buck 变换器最优动态响应的搜索即为求解出一组多周期内的占空比 d_0～d_n, 使得变换器在负载变化后回到新稳态, 且整个响应过程输出电压波动最小。数学模型可表述如下: 待求解的占空比如式 (3) 所示, 包含动态响应过程 k 个周

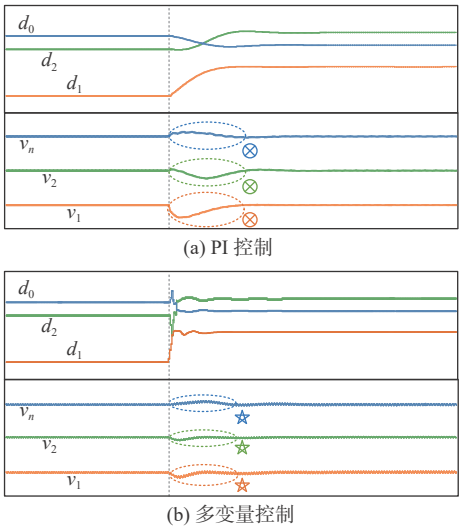


图 3 负载 R_1 变化时使用 PI 控制和多变量控制的占空比 d_1～d_n 和输出电压 v_1～v_n 示意图

Fig. 3 Duty cycles d_1～d_n and output voltages v_1～v_n using PI control and multivariable control when load R_1 varies

期 T 内开关管 S_0～S_{n-1} 的占空比。

$$\text{Solve}: \boldsymbol{S}_{\text{d}}= \begin{matrix} S_0 \\ S_1 \\ S_2 \\ \vdots \\ S_{n-1} \end{matrix} \begin{bmatrix} d_{0,1} & d_{0,2} & d_{0,3} & \cdots & d_{0,k} \\ d_{1,1} & d_{1,2} & d_{1,3} & \cdots & d_{1,k} \\ d_{2,1} & d_{2,2} & d_{2,3} & \cdots & d_{2,k} \\ \vdots & \vdots & \vdots & & \vdots \\ d_{n-1,1} & d_{n-1,2} & d_{n-1,3} & \cdots & d_{n-1,k} \end{bmatrix}_{n\times k} \begin{matrix} \\ \\ \\ \\ \\ \end{matrix} \\ \begin{matrix} T & 2T & 3T & \cdots & kT \end{matrix} \tag{3}$$

求解的约束条件如下：

$$\text{s.t.} \begin{cases} i_{L_t_0+kT} = i_L(1)_{\text{sv}} \\ v_{1_t_0+kT} = v_1(1)_{\text{sv}} \\ v_{2_t_0+kT} = v_2(1)_{\text{sv}} \\ \quad \vdots \\ v_{n_t_0+kT} = v_n(1)_{\text{sv}} \end{cases} \tag{4}$$

$$\sum_{q=1}^{n} d_{q,j} = 1 \qquad j \in [1,k] \tag{5}$$

应保证在这组占空比下变换器在切载时刻 t_0 后经过动态响应时间 kT 回到新稳态，即 t_0+kT 时刻电感电流 i_{L_t0+kT} 和输出电压 v_{1_t0+kT}～v_{n_t0+kT} 分别都达到新稳态的初始值 $i_L(1)_{\text{sv}}$ 和 $v_1(1)_{\text{sv}}$～$v_n(1)_{\text{sv}}$，同时要求每个周期 T 内所有支路开关管的占空比之和为 1；新稳态的初始值 $i_L(1)_{\text{sv}}$ 和 $v_1(1)_{\text{sv}}$～$v_n(1)_{\text{sv}}$ 可通过式 (1) 和式 (2) 进行求解。求解的目标函数为

$$\text{Min}: J = \sum_{q=1}^{n} \omega_q \left(\frac{\max(|v_q(t) - V_{q_\text{ref}}|)}{V_{q_\text{ref}}} \right)^2$$
$$t \in [t_0, t_0+kT] \tag{6}$$

式中：$\max(|v_q(t) - V_{q_\text{ref}}|)$为动态响应过程支路电压瞬时值 $v_q(t)$ 与电压参考值 V_{q_ref} 的最大偏差，反映了动态响应过程支路 v_q 的电压波动；ω_q 为相应的权重。根据对输出支路 v_1～v_n 不同的电压波动要求，设置不同的权重 ω_1～ω_n，可以获得相应的最优动态响应。

综上，SIMO Buck 变换器最优动态响应的数学模型为求解出满足约束条件 (4)、(5) 和最小化目标函数 (6) 的一组占空比 (式 (3))。

值得注意的是，对上述数学模型进行直接求解十分困难。首先，待求解的占空比数量多，约束条件方程多；其次，占空比 d_0 与 d_1～d_n 大小关系不同，SIMO Buck 变换器的工作模态也不同，导致动态响应过程中的电感电流 i_L 和输出电压 v_1～v_n 的数学解析式不同；最后，变换器还可能由 CCM 进入 DCM，使得整体求解过程更为复杂。基于此，本文提出了一种利用改进的 DE 算法来完成 SIMO Buck 变换器最优动态响应问题求解的方法。

2 基于改进 DE 算法的求解策略

经典 DE 算法作为一种高效的全局优化算法，主要操作步骤包括初始化、变异、交叉和选择。但是，经典的 DE 算法还存在后期收敛速度较慢和容易陷入局部最优点等问题。为增加种群多样性，降低陷入局部最优解的危险，本文所用改进 DE 算法采用的改进策略[20-22]包括：在变异步骤中采用自适应更新变异因子的双变异策略，在交叉步骤中采用自适应更新的交叉因子，在选择步骤中采用 ε-约束方法[23-25]对个体进行比较，进化流程如图 4 所示。

2.1 初始化

初始化种群，即在进化代数 G=0 时，在规定的最小和最大边界约束的搜索空间内均匀随机产生 N 个个体。根据式 (3) 中需要求解的变量，初始化的一个个体包括 k 个周期内 n 个开关管的占空比。在第 i 个个体中，初始化的第 j 个周期开关管 S_0 和开关管 S_1 的占空比 $d_{0,j}(i)$ 和 $d_{1,j}(i)$ 为 0～1 之间的随机数。由于约束条件 (5)，剩余开关管的占空比 $d_{2,j}(i)$～$d_{n-1,j}(i)$ 的最大值为 1 减去已经初始化的占空比值，初始化公式如 (7) 所示。另外，初始化的开关管占空比都保留 3 位小数。经过初始化，产生了进化代数 G=0 时的 N 个初始个体，即目标向量 $\boldsymbol{x}_{1,0}$～$\boldsymbol{x}_{N,0}$。

$$\begin{cases} d_{0,j}(i) = \text{round}(\text{rand}[0,1], -3) \\ d_{1,j}(i) = \text{round}(\text{rand}[0,1], -3) \\ d_{2,j}(i) = \text{round}(\text{rand}[0,1] \cdot (1 - d_{1,j}(i)), -3) \\ \quad \vdots \\ d_{n-1,j}(i) = \text{round}\left(\text{rand}[0,1] \cdot \left(1 - \sum_{q=1}^{n-2} d_{q,j}(i)\right), -3\right) \end{cases} \tag{7}$$

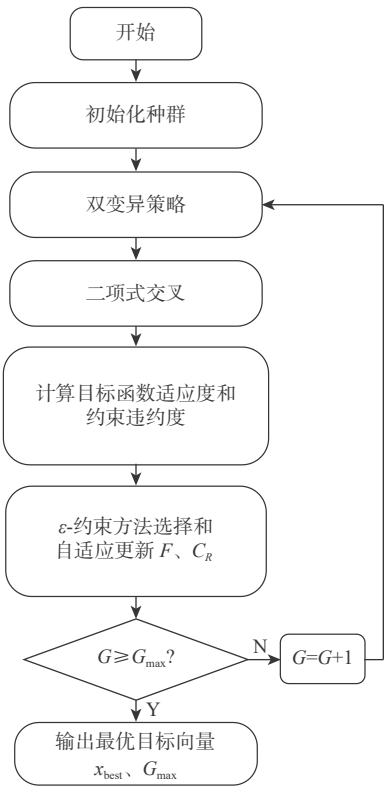


图 4　改进 DE 算法进化流程

Fig. 4　Evolution process of improved DE algorithm

2.2　变　异

变异操作, 即将目标向量 $\boldsymbol{x}_{i,G}$ 通过差分变异操作与自身重组形成变异向量 $\boldsymbol{v}_{i,G}$。由于单一的变异策略很难支持在不同时间对种群的搜索, 多种变异策略应运而生。本文所用的双变异策略为 DE/rand/1 和 DE/rand-to-best/1, 分别如下:

DE/rand/1

$$\boldsymbol{v}_{i,G}=\boldsymbol{x}_{r_1,G}+F_{i,G}(\boldsymbol{x}_{r_2,G}-\boldsymbol{x}_{r_3,G}) \tag{8}$$

DE/rand-to-best/1

$$\boldsymbol{v}_{i,G}=\boldsymbol{x}_{r_4,G}+F_{i,G}(\boldsymbol{x}_{\text{best},G}-\boldsymbol{x}_{r_4,G})+F_{i,G}(\boldsymbol{x}_{r_5,G}-\boldsymbol{x}_{r_6,G}) \tag{9}$$

DE/rand/1 在当前种群中随机性选择 3 个目标向量 $\boldsymbol{x}_{r_1,G}$、$\boldsymbol{x}_{r_2,G}$ 和 $\boldsymbol{x}_{r_3,G}$ 进行变异操作, r_1、r_2 和 r_3 互不相等, 将任意 2 个目标向量差按变异因子 $F_{i,G}$ 进行缩放, 再与第 3 个小目标向量相加产生变异向量 $\boldsymbol{v}_{i,G}$。

由于采用 3 个随机的目标向量, DE/rand/1 这一变异策略具有较强的全局搜索能力, 而 DE/rand-to-best/1 引入最优目标向量 $\boldsymbol{x}_{\text{best},G}$, 增强了局部搜索能力。本文所用的双变异策略将 2 种变异策略以权重 ω_G 组合产生变异向量, 如下:

$$\begin{aligned}\boldsymbol{v}_{i,G}=&(1-\omega_G)[\boldsymbol{x}_{r_1,G}+F_{i,G}(\boldsymbol{x}_{r_2,G}-\boldsymbol{x}_{r_3,G})]+\\&\omega_G[\boldsymbol{x}_{r_4,G}+F_{i,G}(\boldsymbol{x}_{\text{best},G}-\boldsymbol{x}_{r_4,G})+F_{i,G}(\boldsymbol{x}_{r_5,G}-\boldsymbol{x}_{r_6,G})]\end{aligned} \tag{10}$$

式中: ω_G 为当前进化次数除最大进化代数, 从而根据不同时间种群变化调整 2 个变异策略的权重, 满足不同时期对搜索能力和收敛速度的需求。

2.3　交　叉

交叉操作通过交叉部分变异向量 $\boldsymbol{v}_{i,G}$ 和目标向量 $\boldsymbol{x}_{i,G}$ 维度产生试验向量 $\boldsymbol{u}_{i,G}$, 从而提高种群的多样性。本文使用式 (11) 所示的二项式交叉策略。

$$\boldsymbol{u}_{i,G}(:,j)=\begin{cases}\boldsymbol{v}_{i,G}(:,j) & C_{R_i,G}\geqslant \text{rand}(0,1)\ \text{或}\ j=j_{\text{rand}}\\ \boldsymbol{x}_{i,G}(:,j) & \text{其他}\end{cases} \tag{11}$$

式中: $C_{R_i,G}$ 为交叉因子, 当 $C_{R_i,G}$ 大于 [0,1] 之间的随机数时, 试验向量的第 j 列 $\boldsymbol{u}_{i,G}(:,j)$ 选自变异向量的第 j 列 $\boldsymbol{v}_{i,G}(:,j)$, 否则, 选自目标向量的第 j 列 $\boldsymbol{x}_{i,G}(:,j)$; j_{rand} 为随机数, 用来保证变异向量至少有一列信息被保存在试验向量 $\boldsymbol{u}_{i,G}$。

2.4　选　择

选择操作是选择所生成的下一代个体是原目标向量 $\boldsymbol{x}_{i,G}$ 还是经过变异交叉操作之后形成的试验向量 $\boldsymbol{u}_{i,G}$。需要将目标向量 $\boldsymbol{x}_{i,G}$ 和试验向量 $\boldsymbol{u}_{i,G}$ 根据目标函数适应度和约束违反度进行一对一的选择。将目标向量 $\boldsymbol{x}_{i,G}$ 和试验向量 $\boldsymbol{u}_{i,G}$ 的占空比应用于动态过程, 利用式 (1) 和式 (2) 可以获得相应的动态过程的电压 v_1～v_n, 利用式 (6) 可以获得相应的目标函数适应度 $J(\boldsymbol{x})$ 和 $J(\boldsymbol{u})$。基于数学模型中的约束条件式 (4) 和式 (5), 利用式 (12)～式 (14) 求解相应的约束违反度 $\varphi(\boldsymbol{x})$ 和 $\varphi(\boldsymbol{u})$。

$$\begin{aligned}\varphi_1=&\max((i_{L_t_0+kT}-i_L(1)_{sv})^2+\\&\sum_{q=1}^{n}h_q(v_{q_t_0+kT}-v_q(1)_{sv})^2-\delta,0)\end{aligned} \tag{12}$$

$$\varphi_2=\sum_{j=1}^{k}\left(\max\left(\sum_{q=1}^{n-1}d_{q,j}-1,0\right)\right) \tag{13}$$

$$\varphi=\sum_{g=1}^{2}t_g\varphi_g \tag{14}$$

约束条件式 (4) 是为了确保变换器能回到稳态, 因此, 式 (12) 中 δ 为接近 0 的数, 确保经过动态过程中占空比的调节, 电感电流和输出电压非常接近于新稳态的初始值, 而权重 h_q 用来调节电流和电压偏差。约束条件式式 (5) 表明, 每个开关周期所有输出支路开关管的占空比和为 1, 因此, 式 (13) 确保前 n−1 个开关管占空比和小于或等于 1。约束总违反度是式 (12) 和式 (13) 的权重和, 如式 (14) 所示。其中, 权重 t_g 和 h_q 可以根据需求给定。

ε-约束方法用来从目标向量 $\boldsymbol{x}_{i,G}$ 和试验向量 $\boldsymbol{u}_{i,G}$ 中选择较优个体。ε-约束方法将约束违反度 φ 小于 ε 的个体视为可行解，从而充分利用目标函数较好的不可行解信息，具有更好的收敛性。为了最小化目标函数，采用 ε-约束方法的比较准则，如下：

$$f(\boldsymbol{u})_\varepsilon \leqslant f(\boldsymbol{x})_\varepsilon \Leftrightarrow \begin{cases} J(\boldsymbol{u}) \leqslant J(\boldsymbol{x}) & \varphi(\boldsymbol{u}),\varphi(\boldsymbol{x}) \leqslant \varepsilon \\ J(\boldsymbol{u}) \leqslant J(\boldsymbol{x}) & \varphi(\boldsymbol{u}) = \varphi(\boldsymbol{x}) \\ \varphi(\boldsymbol{u}) \leqslant \varphi(\boldsymbol{x}) & \text{其他} \end{cases} \tag{15}$$

当试验向量 $\boldsymbol{u}$ 和目标向量 $\boldsymbol{x}$ 的约束违反度都小于 ε 或 2 个个体约束违反度相等时，如果试验向量的目标函数适应度 $J(\boldsymbol{u})$ 小于目标向量的目标函数适应度 $J(\boldsymbol{x})$，则试验向量 $\boldsymbol{u}$ 优于目标向量 $\boldsymbol{x}$。除此之外，当试验向量的约束违反度 $\varphi(\boldsymbol{u})$ 比较小时，试验向量 $\boldsymbol{u}$ 优于目标向量 $\boldsymbol{x}$。

为确保算法进化终止时，约束能被满足，ε 随着算法进化逐渐减小，如下：

$$\varepsilon_{G+1} = \frac{\varepsilon_G}{1.035} \tag{16}$$

利用 ε-约束方法，从目标向量 $\boldsymbol{x}_{i,G}$ 和试验向量 $\boldsymbol{u}_{i,G}$ 中选择较优个体作为下一代目标向量进行下一次进化，如下：

$$\boldsymbol{x}_{i,G+1} = \begin{cases} \boldsymbol{u}_{i,G} & f(\boldsymbol{u}_{i,G})_\varepsilon \leqslant f(\boldsymbol{x}_{i,G})_\varepsilon \\ \boldsymbol{x}_{i,G} & \text{其他} \end{cases} \tag{17}$$

另外，自适应更新变异因子和交叉因子以提高种群多样性。进化代数 G=0 时的变异因子和交叉因子分别为 0.5 和 0.9，当经过变异交叉操作得到的试验向量劣于目标向量时，就更新变异交叉因子，如下：

$$F_{i,G+1} = \begin{cases} F_{i,G} & f(\boldsymbol{u}_{i,G})_\varepsilon \leqslant f(\boldsymbol{x}_{i,G})_\varepsilon \\ 0.1 + 0.8\text{rand}(0,1) & \text{其他} \end{cases} \tag{18}$$

$$C_{R_i,G+1} = \begin{cases} C_{R_i,G} & f(\boldsymbol{u}_{i,G})_\varepsilon \leqslant f(\boldsymbol{x}_{i,G})_\varepsilon \\ \text{rand}(0,1) & \text{其他} \end{cases} \tag{19}$$

此外，根据 ε-约束方法比较准则式 (15)，从下一代目标向量中选出最优目标向量 $\boldsymbol{x}_{\text{best},G+1}$，用于下一次进化的变异操作。

2.5 终 止

当进化代数达到最大 $G_{\max}$ 时，算法中止。此时，选出的最优目标向量 $\boldsymbol{x}_{\text{best},G_{\max}}$ 是动态过程中占空比的最优解。在这组占空比下，SIMO Buck 变换器可以在切载后回到稳态，并具有最优的动态响应效果。

3 计算验证与分析

为验证利用改进 DE 算法求解最优动态响应的效果，本节以单电感三路输出 Buck 变换器为例进行计算，此时目标函数为

$$\begin{aligned} \min: J = {} & \omega_1 \left[\frac{\max\left(\left|v_1(t) - v_{1_\text{ref}}\right|\right)}{v_{1_\text{ref}}} \right]^2 + \\ & \omega_2 \left[\frac{\max\left(\left|v_2(t) - v_{2_\text{ref}}\right|\right)}{v_{2\text{ref}}} \right]^2 + \\ & \omega_3 \left[\frac{\max\left(\left|v_3(t) - v_{3_\text{ref}}\right|\right)}{v_{3_\text{ref}}} \right]^2 \end{aligned} \tag{20}$$

变换器的开关频率设置为 20 kHz，电路参数如表 1 所示。在改进的 DE 算法中，将种群中个体数量 N 设置为 200，最大进化代数 $G_{\max}$ 设置为 5 000。在算法中，采用双变异策略、ε-约束方法及自适应更新变异因子和交叉因子的改进措施提高搜索寻优能力。值得注意的是，随着输出支路数的增加，最优动态响应的数学模型中，未知数的维数增加，算法的约束条件和目标函数也会更复杂，因此，所需的个体数量和进化代数将随之增加，需要更多的计算资源和时间。

表 1 变换器电路参数

Table 1 Circuit parameters of converter

参数	数值
输入电压V_{in}/V	25
额定输出电压V_{1_ref}/V	5
额定输出电压V_{2_ref}/V	10
额定输出电压V_{3_ref}/V	15
电感L/μH	500
电容C_1, C_2, C_3/μF	1 500
输出负载电阻R_1/Ω	1
输出负载电阻R_2, R_3/Ω	5

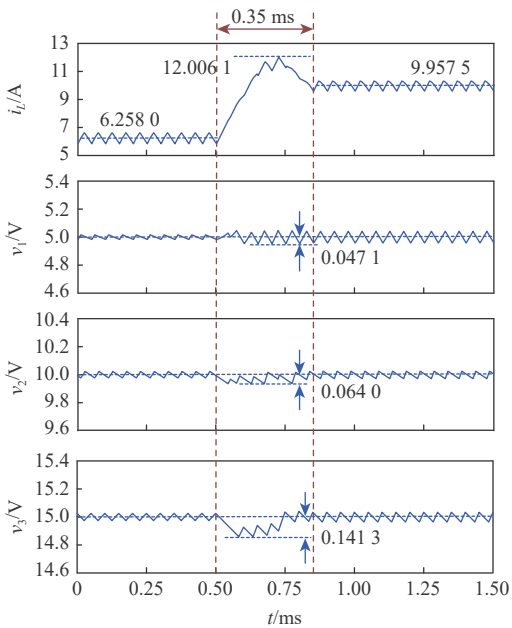
3.1 不同目标的最优动态响应

基于改进的 DE 算法，可以在不同的目标函数 (20) 下求解出相应的最优动态响应和开关管 S_0、S_1 和 S_2 的控制占空比。首先，均衡考虑 3 条输出支路的电压波动，目标函数 (20) 中的权重 ω_1、ω_2 和 ω_3 均设置为 1，动态响应时间设置为 7 个周期，求解第 1 条输出支路负载 R_1 由 4 Ω 跳变至 1 Ω 和由 1 Ω 跳变至 4 Ω 时的最优动态响应。求解得到的具体占空比数据如表 2 所示，最优动态响应时电感电流和输出电压变化波形如图 5 所示。负载 R_1 由 4 Ω 跳变至 1 Ω 时的波形如图 5(a) 所示，在 0.5 ms 前，稳

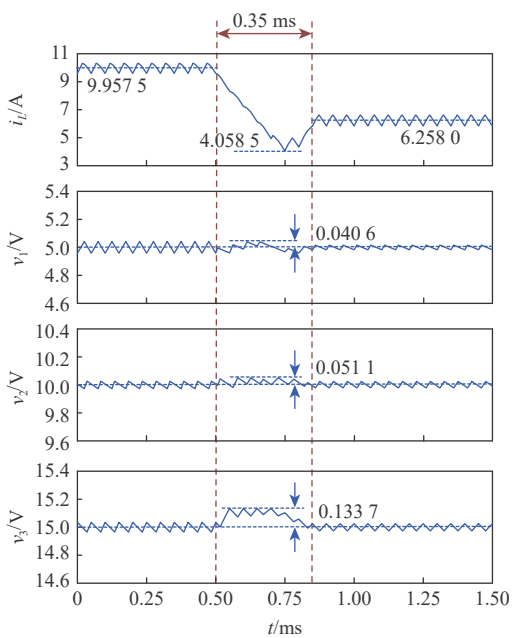
表 2　动态响应时间为 7 个周期时实现最优动态响应的占空比

Table 2　Duty cycles to achieve the optimal dynamic response when dynamic response time is 7 periods

切载	开关管	占空比						
		T	$2T$	$3T$	$4T$	$5T$	$6T$	$7T$
4 Ω → 1 Ω	S_0	1.000	0.997	0.924	0.505	0.490	0.060	0.001
	S_1	0.827	0.438	0.509	0.448	0.435	0.444	0.496
	S_2	0.173	0.231	0.179	0.254	0.091	0.264	0.218
1 Ω → 4 Ω	S_0	0.000	0.000	0.000	0.000	0.140	0.535	0.992
	S_1	0.135	0.251	0.265	0.191	0.020	0.237	0.406
	S_2	0.188	0.348	0.269	0.248	0.504	0.401	0.240



(a) R_1: 4 Ω→1 Ω, ω_1=ω_2=ω_3=1



(b) R_1: 1 Ω→4 Ω, ω_1=ω_2=ω_3=1

图 5　2 种切载情况下, 动态响应时间为 7 个周期的最优动态响应时电感电流和输出电压波形

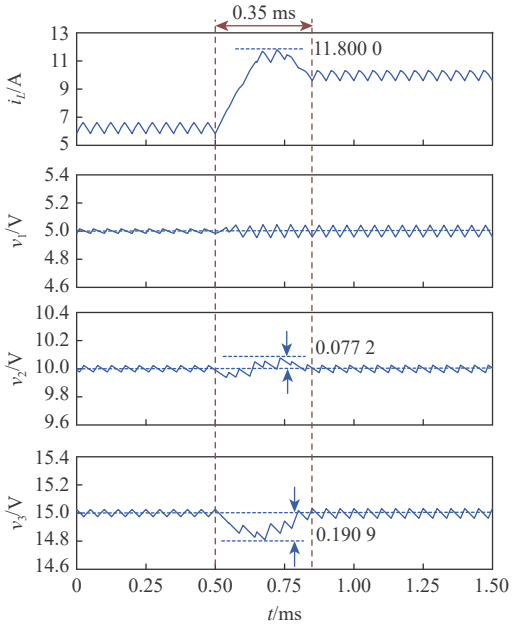
Fig. 5　Inductor current and output voltage when the optimal dynamic response is achieved with 7 periods under two load variations

态占空比 d_0、d_1、d_2 分别为 0.455、0.207、0.309, 此时, 电感电流和支路 v_1、v_2、v_3 的电压平均值分别为 6.285 0 A 和 5 V、10 V、15 V。在 0.5 ms 时, v_1 支路负载 R_1 由 4 Ω 跳变至 1 Ω。经过表 2 所示的 7 个周期占空比调节, 变换器回到稳态, 可以看出, 占空比数值在 0～1 之间自由变化。动态过程中, 支路 v_1、v_2 和 v_3 的电压波动很小, 分别为 0.047 1 V、0.064 0 V 和 0.141 3 V。再次回到稳态后的占空比 d_0、d_1 和 d_2 分别为 0.362、0.498 和 0.197。此时, 电感电流增加至 9.957 5 A, 而输出支路 v_1、v_2 和 v_3 的电压平均值不变。负载 R_1 由 1 Ω 跳变至 4 Ω 时的波形如图 5(b) 所示, 经过求解的占空比调节, 电感电流由 9.957 5 A 变为 6.258 0 A, 动态过程输出电压也比较平稳。

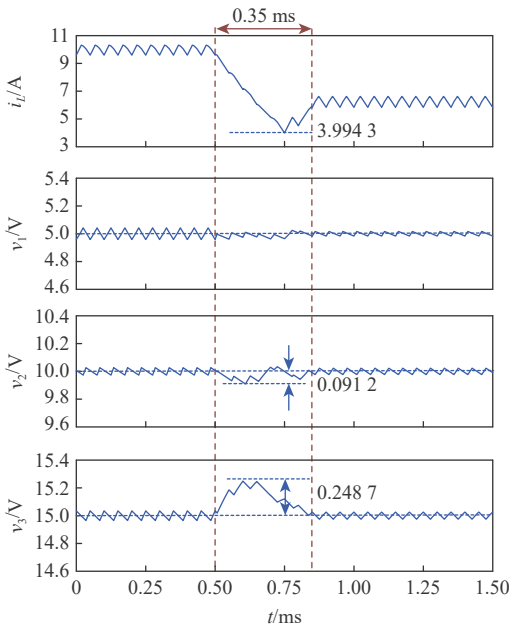
以上求解了目标函数 (20) 的权重 ω_1、ω_2 和 ω_3 都设置为 1 时相应的最优动态响应。此外, 可以将权重 ω_1 设置为 1, ω_2 和 ω_3 都设置为 0, 最小化切载支路 v_1 的电压波动, 即搜索最优的自调节。设置动态响应时间为 7 个周期, 2 种切载情况下, 最优自调节时电感电流和电压的波形如图 6 所示。此时, 相比图 5, 动态过程输出支路 v_1 电压非常平稳, 电压波动集中在输出支路 v_2 和 v_3。另外, 还可以将权重 ω_1 设置为 0, ω_2 和 ω_3 都设置为 1, 最小化输出支路 v_2 和 v_3 的电压波动, 即搜索最优交叉调节影响。设置动态响应时间为 7 个周期, 2 种切载情况下最优交叉调节时电感电流和电压的波形如图 7 所示。此时, 相比图 5, 动态过程输出支路 v_2 和 v_3 的电压波动改善, 而输出支路 v_1 的电压波动增加。

3.2　不同动态响应时间的最优动态响应

变换器在切载后, 可以经过不同的动态响应时间回到稳态。分别将动态响应时间设置为 1～12 个周期, 通过求解后发现, 当动态响应时间小于或等于 5 个周期时, 稳态约束条件无法满足, 即变换器无法回到稳态。当动态响应时间达到 6 个周期时, 经过求解的占空比调节, 变换器可以顺利回到稳态。此时, 2 种切载情况下求解的最优动态响应时电感电流和输出电压变化波形如图 8 所示。



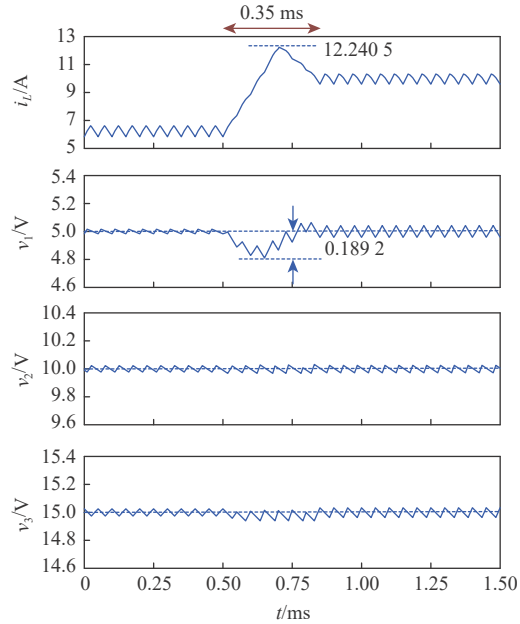
(a) R_1: 4 Ω→1 Ω, ω_1=1, ω_2=ω_3=0



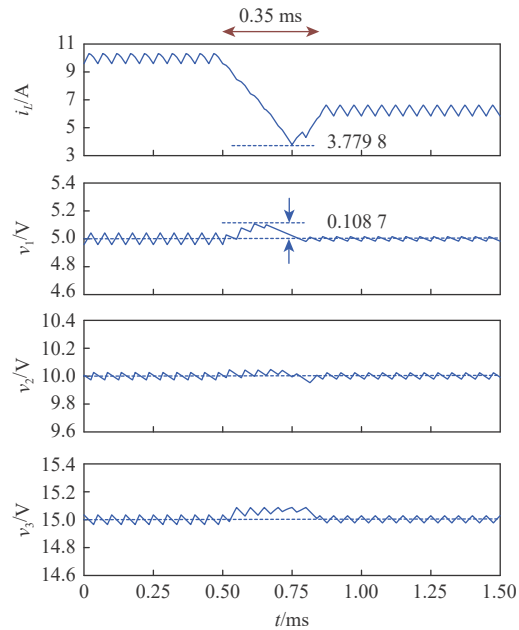
(b) R_1: 1 Ω→4 Ω, ω_1=1, ω_2=ω_3=0

图 6 2种切载情况下,最优自调节时的电感电流和电压波形

Fig. 6 Inductor current and output voltage when the optimal self-regulation is achieved under two load variations



(a) R_1: 4 Ω→1 Ω, ω_1=0, ω_2=ω_3=1



(b) R_1: 1 Ω→4 Ω, ω_1=0, ω_2=ω_3=1

图 7 2种切载情况下,最优交叉调节时的电感电流和电压波形

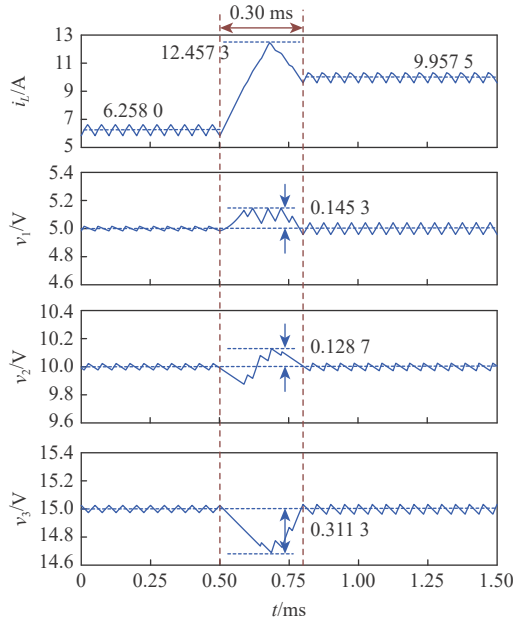
Fig. 7 Inductor current and output voltage when the optimal cross-regulation is achieved under two load variations

相比7个周期时的电感电流和输出电压波形(见图 5),最优动态响应时,3条输出支路的电压波动比较大。

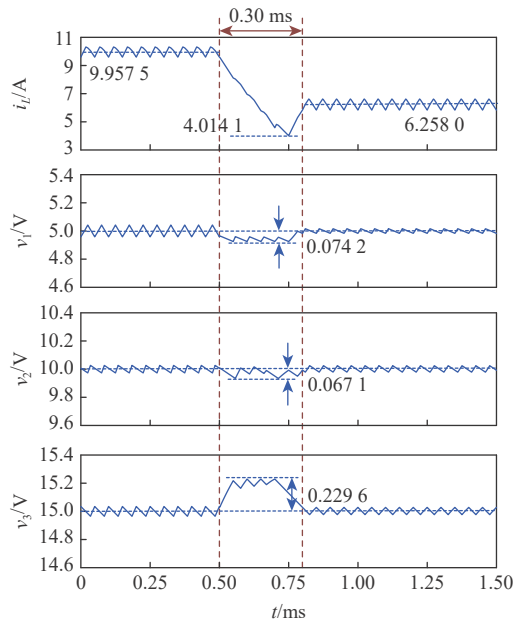
继续增加动态响应时间,求解得到的最优动态响应会发生一定的变化。动态响应时间为6~12个周期时,最优动态响应下输出支路 v_1、v_2 和 v_3 的电压波动变化曲线如图 9 所示。可以看出,2种切载情况下,7个周期相比6个周期,3条输出支路的电压波动都明显减小。而在7个周期后,继续增加动态响应时间,3条输出支路的电压波动没有明显改善。在设计 SIMO 直流变换器的控制器时,可以根据计算结果综合考虑动态响应时间和动态响应效果。

3.3 额外的电感电流约束的最优动态响应

如图 5(a) 所示,在最优动态响应下,3条输出支路 v_1、v_2 和 v_3 的电压波动很小,但电感电流会超调至 12.006 1 A。此时,可以通过在改进 DE 算法中添加约束,限制电感电流的峰值。当电感电流的峰值限制在 11.300 0 A 时,违约度计算公式为



(a) R_1: 4 Ω→1 Ω, ω_1=ω_2=ω_3=1



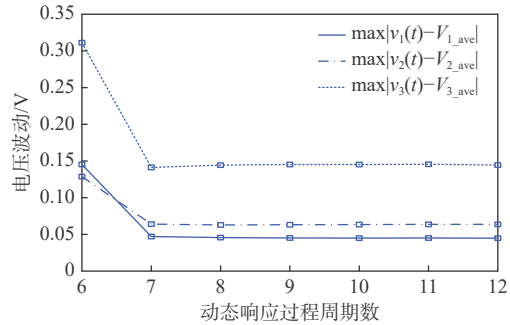
(b) R_1: 1 Ω→4 Ω, ω_1=ω_2=ω_3=1

图 8　2 种切载情况下, 动态响应时间为 6 个周期的最优动态响应时电感电流和输出电压波形

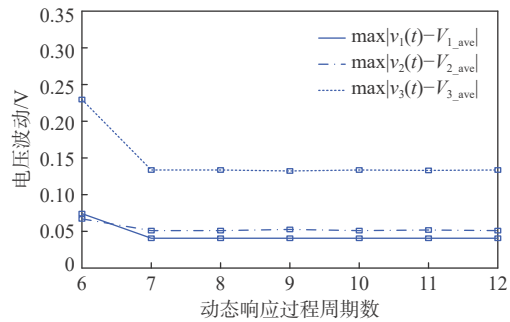
Fig. 8　Inductor current and output voltage when the optimal dynamic response is achieved with 6 periods under two load variations

$$\varphi_3 = \max((i_{L_\max} - 11.300\,0), 0) \tag{21}$$

可以将电感电流约束违约度式 (21) 添加到式 (14) 中计算总违约度。为明确不同电感电流峰值限制下最优动态响应的变化, 求解负载 R_1 从 4 Ω 切载至 1 Ω, 7 个周期动态响应时间时不同电感电流峰值约束下的最优动态响应。分别将电感电流峰值限制为 11.300 0, 11.400 0, 11.500 0, 11.600 0, 11.700 0, 11.800 0, 11.900 0, 12.000 0 A, 此时的 3 条输出支路的电压波动变化如图 10 所示。可以看出, 随着峰值电感电流的增加, 3 条输出支路的电压波动都逐渐减小, 直到达到 11.800 0 A 后, 电压波动基本不变。值得注意的是, 峰值电感电流限制在 11.200 0 A 时, 约束无法满足, 即此时变换器无法在峰值电感电流为 11.200 0 A 时回到稳态。峰值电感电流约束为 11.300 0 A 和 11.800 0 A 时的电感电流和电压波形如图 11 所示。与图 5(a) 无电感电流约束相比, 峰值电感电流限制为 11.300 0 A 时, 电压波动明显增加。而峰值电感电流限制为 11.800 0 A 时, 电压波动基本一样。此时, 电感电流约束对电



(a) R_1: 4 Ω→1 Ω, 不同动态响应时间的最优动态响应时电压波动



(b) R_1: 1 Ω→4 Ω, 不同动态响应时间的最优动态响应时电压波动

图 9　2 种切载情况下, 不同动态响应时间的最优动态响应时电压波动变化曲线

Fig. 9　Voltage fluctuation curves when the optimal dynamic response is achieved with different time under two load variations

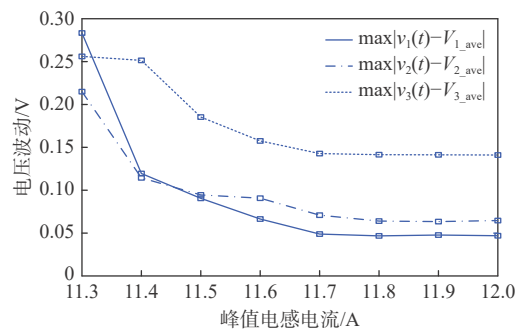
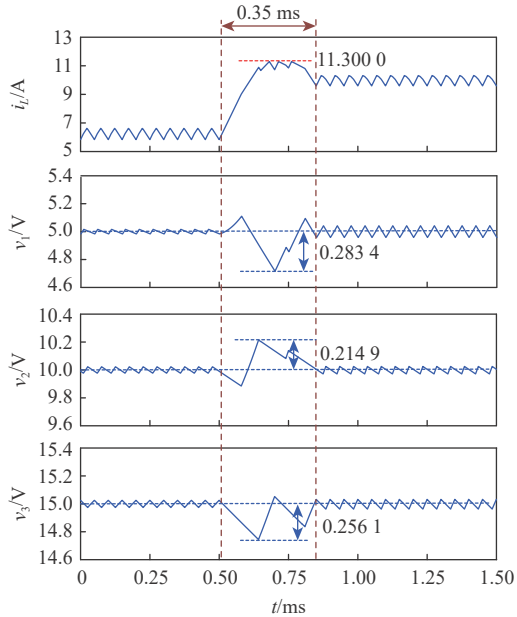
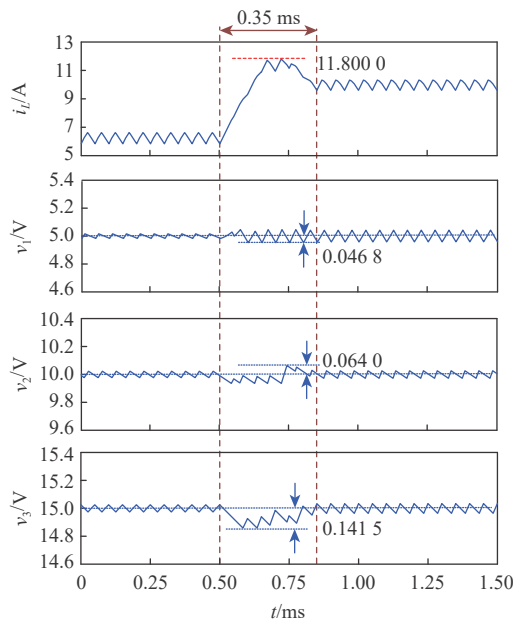


图 10　不同峰值电感电流约束的最优动态响应下电压波动变化曲线

Fig. 10　Voltage fluctuation curves when the optimal dynamic response is achieved with different limited maximum inductor current



(a) 峰值电感电流约束为 11.300 0 A



(b) 峰值电感电流约束为 11.800 0 A

图 11 峰值电感电流约束为 11.300 0 A 和 11.800 0 A 时最优动态响应的电感电流和电压波形

Fig. 11 Inductor current and output voltage when the optimal dynamic response is achieved with maximum inductor current limited to 11.300 00 A and 11.800 00 A

压波动已无影响。

4 结 论

本文通过对 SIMO Buck 变换器的最优动态响应问题进行数学建模，并利用改进的 DE 算法进行求解，成功计算出最优动态响应。以单电感三路输出 Buck 变换器为例进行计算验证，结果表明：

1）在最优自调节目标下，切载支路 v_1 的电压在动态过程中可以保持非常平稳。

2）在最优交叉调节目标下，输出支路 v_2 和 v_3 的电压波动很小。

3）在不同动态响应时间和峰值电感电流约束下，也可求解出最优动态响应。

因此，本文提出的最优动态响应搜索方法有助于深入了解不同条件下 SIMO 直流变换器的动态响应性能。未来可进一步将所提方法推广至其他类型变换器中，为工程应用中的控制器设计提供指导。

参考文献（References）

[1] HUANG M H, CHEN K H. Single-inductor multi-output (SIMO) DC-DC converters with high light-load efficiency and minimized cross-regulation for portable devices[J]. IEEE Journal of Solid-State Circuits, 2009, 44(4): 1099-1111.

[2] XU W W, LI Y, GONG X H, et al. A single-inductor dual-output switching converter with low ripples and improved cross regulation [C]//Proceedings of the IEEE Custom Integrated Circuits Conference. Piscataway: IEEE Press, 2009: 303-306.

[3] JING X C, MOK P K T, LEE M C. A wide-load-range single-inductor-dual-output boost regulator with minimized cross-regulation by constant-charge-auto-hopping (CCAH) control[C]//Proceedings of the IEEE Custom Integrated Circuits Conference. Piscataway: IEEE Press, 2009: 299-302.

[4] MA D S, KI W H, TSUI C Y, et al. Single-inductor multiple-output switching converters with time-multiplexing control in discontinuous conduction mode[J]. IEEE Journal of Solid-State Circuits, 2003, 38(1): 89-100.

[5] MA D S, KI W H, TSUI C Y. A pseudo-CCM/DCM SIMO switching converter with freewheel switching[J]. IEEE Journal of Solid-State Circuits, 2003, 38(6): 1007-1014.

[6] MA D S, KI W H. Fast-transient PCCM switching converter with freewheel switching control[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2007, 54(9): 825-829.

[7] TREVISAN D, MATTAVELLI P, TENTI P. Digital control of single-inductor multiple-output step-down DC-DC converters in CCM[J]. IEEE Transactions on Industrial Electronics, 2008, 55(9): 3476-3483.

[8] WANG W, XU J P, WEI X J, et al. Peak inductor current and differential-mode voltage control of single-inductor dual-output buck converters in continuous conduction mode[C]//Proceedings of the IEEE 6th International Power Electronics and Motion Control Conference. Piscataway: IEEE Press, 2009: 495-499.

[9] WANG Y, XU J P, YIN G. Cross-regulation suppression and stability analysis of capacitor current ripple controlled SIDO CCM Buck converter[J]. IEEE Transactions on Industrial Electronics, 2019, 66(3): 1770-1780.

[10] 王瑶. 电容电流-电容电压纹波控制单电感双输出 CCM Buck 变换器[J]. 中国电机工程学报, 2020, 40(10): 3280-3288.
WANG Y. Capacitor current and capacitor voltage ripple con-

trolled single-inductor dual-output CCM Buck converter[J]. Proceedings of the CSEE, 2020, 40(10): 3280-3288(in Chinese).

[11] WANG Y, XU J P, ZHOU G H. A cross regulation analysis for single-inductor dual-output CCM Buck converters[J]. Journal of Power Electronics, 2016, 16(5): 1802-1812.

[12] PATRA P, GHOSH J, PATRA A. Control scheme for reduced cross-regulation in single-inductor multiple-output DC-DC converters[J]. IEEE Transactions on Industrial Electronics, 2013, 60(11): 5095-5104.

[13] DASIKA J D, BAHRANI B, SAEEDIFARD M, et al. Multivariable control of single-inductor dual-output Buck converters[J]. IEEE Transactions on Power Electronics, 2014, 29(4): 2061-2070.

[14] LE H P, CHAE C S, LEE K C, et al. A single-inductor switching DC-DC converter with five outputs and ordered power-distributive control[J]. IEEE Journal of Solid-State Circuits, 2007, 42(12): 2706-2714.

[15] WOO Y J, LE H P, CHO G H, et al. Load-independent control of switching DC-DC converters with freewheeling current feedback[J]. IEEE Journal of Solid-State Circuits, 2008, 43(12): 2798-2808.

[16] 贺素霞, 张具琴. 基于模型预测控制单电感双输出 Buck 变换器的仿真分析[J]. 可再生能源, 2022, 40(4): 558-563.
HE S X, ZHANG J Q. Simulation analysis of single inductor dual output Buck converter based on model predictive control[J]. Renewable Energy Resources, 2022, 40(4): 558-563(in Chinese).

[17] WANG B F, TAN K T, SO P L. Low cross regulation SIMO DC/DC converter with model predictive voltage control[C]//Proceedings of the IEEE Power & Energy Society General Meeting. Piscataway: IEEE Press, 2013: 1-5.

[18] WANG B F, KANAMARLAPUDI V R K, XIAN L, et al. Model predictive voltage control for single-inductor multiple-output DC-DC converter with reduced cross regulation[J]. IEEE Transactions on Industrial Electronics, 2016, 63(7): 4187-4197.

[19] ZHANG X N, WANG B F, TAN X J, et al. Deadbeat control for single-inductor multiple-output DC-DC converter with effectively reduced cross regulation[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2020, 8(4): 3372-3381.

[20] TIAN M N, GAO X B, DAI C. Differential evolution with improved individual-based parameter setting and selection strategy[J]. Applied Soft Computing, 2017, 56: 286-297.

[21] WANG S H, LI Y Z, YANG H Y, et al. Self-adaptive differential evolution algorithm with improved mutation strategy[J]. Soft Computing, 2018, 22(10): 3433-3447.

[22] HOU Y, ZHAO L, LU H W. Fuzzy neural network optimization and network traffic forecasting based on improved differential evolution[J]. Future Generation Computer Systems, 2018, 81: 425-432.

[23] 李笠, 李广鹏, 常亮, 等. 约束进化算法及其应用研究综述[J]. 计算机科学, 2021, 48(4): 1-13.
LI L, LI G P, CHANG L, et al. Survey of constrained evolutionary algorithms and their applications[J]. Computer Science, 2021, 48(4): 1-13(in Chinese).

[24] NOMAN N, IBA H. constrained differential evolution for economic dispatch with valve-point effect[J]. International Journal of Bio-Inspired Computation, 2011, 3(6): 346.

[25] TAKAHAMA T, SAKAI S. Fast and stable constrained optimization by the ε constrained differential evolution[J]. Pacific Journal of Optimizatin, 2009, 5(2): 261-282.

Optimal dynamic response exploration for SIMO Buck converter based on differential evolution algorithm

LI Dan[1, 2], CUI Wenfeng[3], CHEN Guipeng[1, 2, *]

(1. Shenzhen Research Institute of Xiamen University, Shenzhen 518000, China;
2. School of Aerospace Engineering, Xiamen University, Xiamen 361000, China;
3. Zhejiang Feihang Intelligent Technology Co., Ltd., Hangzhou 311100, China)

Abstract: With the widespread application of single-inductor multiple-output (SIMO) DC-DC converters, their dynamic response has much concern for domestic and foreign researchers. To explore the theoretical limit of the optimal dynamic response for the SIMO Buck converter, this paper first establishes a corresponding mathematical model according to the working principle of the converter, and then an improved differential evolution (DE) algorithm is employed to search out the solution. In addition, the suggested approach can investigate the best dynamic response with various goals, including the best self-regulation or cross-regulation. It is also possible to know the optimal dynamic response with different constraints, such as different dynamic response times and peak inductor currents. In order to completely recognize the dynamic performance of SIMO DC-DC converters, the theoretical limit exploration of optimal dynamic response based on the heuristic DE intelligent algorithm is useful. Ideally, this will direct the controller design in order to obtain an optimized dynamic response.

Keywords: single-inductor multiple-output DC-DC converters; dynamic response; differential evolution algorithm; self-regulation; cross-regulation

Received: 2023-06-14; **Accepted**: 2023-09-27; **Published Online**: 2024-01-12 17:46

URL: link.cnki.net/urlid/11.2625.V.20240112.1130.001

Foundation items: Guangdong Basic and Applied Basic Research Foundation (2022A1515011616); National Natural Science Foundation of China (51907172)

* **Corresponding author**. E-mail: cgp2017@xmu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0291

基于层次分析-熵值法的DC-DC变换器拓扑综合评价

张洋，丘东元*，张波，陈艳峰
(华南理工大学 电力学院，广州 510641)

摘 要：为了更全面地比较拓扑结构和功能相似的一系列DC-DC变换器并实现选优，提出了一种基于层次分析-熵值法的DC-DC变换器拓扑综合评价方法。通过总结DC-DC变换器的常用性能指标，构建分层次的评价指标体系；分别运用层次分析法和熵值法对指标体系中的指标赋权，并采用侧重于层次分析法的策略得到组合权重；进一步运用线性加权综合法得到各DC-DC变换器的评分。以11个电压增益均为D^2的DC-DC变换器拓扑为例，运用所提方法进行综合评价，并对评价结果进行可行性分析，结果表明，所选出的拓扑具有良好性能。

关 键 词：DC-DC变换器；综合评价；指标体系；层次分析法；熵值法
中图分类号：TM461
文献标志码：A **文章编号**：1001-5965(2025)05-1469-11

近年来，随着风力和光伏发电、电动汽车、储能等新能源产业的兴起，DC-DC变换器得到了越来越广泛的应用。为满足不同应用场合的要求，DC-DC变换器拓扑层出不穷。DC-DC变换器的设计从手工设计进入智能设计阶段，科研人员陆续提出了系统性的DC-DC变换器拓扑构造方法[1-5]。例如，利用图论搜索法自动构造出指定元器件类型及数量的所有DC-DC变换器拓扑[1]，利用磁通平衡设计法自动构造出符合给定电压增益表达式的DC-DC变换器系列拓扑[2]，利用强化学习的方法构造出具有特定结构的DC-DC变换器，如多端口变换器[3-4]。

由于运用同一方法构造的一系列DC-DC变换器拓扑基于相同的前提条件，其拓扑结构通常具有相似的性质。在这种情况下，设计者若仅凭简单的观察和分析，难以区分这些拓扑的性能，并从中选出最优拓扑。此外，拓扑比较方法通常只关注少数几个指标，如元器件的类型、数量、电压电流应力等，得出的结论可能不够全面。为解决上述问题，需要研究DC-DC变换器拓扑的综合评价方法。

综合评价属于管理学领域，在各种考核评级、工程活动甚至日常生活中得到了广泛的应用。综合评价的步骤通常包括指标体系构建、数据预处理、指标赋权、性能评价。其中，指标赋权方法可分为主观赋权法、客观赋权法和组合赋权法，性能评价方法包括线性加权综合法[6]、模糊综合评价法[7-9]等。

层次分析-熵值法[10-12]是一种组合赋权法，其将评价者的主观判断所得权重和指标值的分布特点所得权重进行组合，综合了主观赋权法和客观赋权法的优点。文献[10-11]将层次分析-熵值法应用于电力设备的综合评价，但该指标体系包含外观、噪声、继电保护等与电力电子变换器拓扑无关的指标。在性能评价中均采用模糊综合评价法，但隶属函数凭经验选取，难以准确反映指标值的分布特点。此外，上述文献仅对某个对象进行评价，难以发挥熵值法的优点，且缺乏对比。

为此，本文采用层次分析-熵值法对DC-DC变换器进行综合评价，在指标体系构建方面，选择与

收稿日期：2023-05-30；**录用日期**：2023-11-09；**网络出版时间**：2023-11-23 17:44
网络出版地址：link.cnki.net/urlid/11.2625.V.20231122.1713.004
基金项目：国家自然科学基金(52277177)
*通信作者. E-mail：epdyqiu@scut.edu.cn

引用格式：张洋，丘东元，张波，等. 基于层次分析-熵值法的DC-DC变换器拓扑综合评价[J]. 北京航空航天大学学报，2025，51(5)：1469-1479. ZHANG Y，QIU D Y，ZHANG B，et al. Comprehensive evaluation of DC-DC converters based on analytic hierarchy process and entropy method[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1469-1479 (in Chinese).

变换器拓扑结构和工作性能有关的指标；在性能评价方法方面，采用线性加权综合法，无需隶属函数。得到各拓扑的评分后，从中选出性能最优的拓扑。

1 指标体系构建与数据预处理

为实现DC-DC变换器的性能综合评价，需要先构建一个合理的指标体系，再测量DC-DC变换器与指标相关的数据，计算相应的指标值，并对指标值进行一致化和无量纲化处理。

1.1 指标体系的构建

在综合评价中，构建指标体系是首要工作，通常应遵循以下原则[6]：

1）指标在满足评价目标的前提下应精简。

2）指标应具有独立性，不能存在包含关系或因果关系。

3）指标应具有代表性和差异性。

4）指标应具有可行性，数据易于测量和计算。

例如，在DC-DC变换器的综合评价中，可选择输出电压精度、输出电压纹波系数、电压增益、效率、开关管应力、电容电压应力、电感电流应力、器件数量和成本、功率密度等指标。

确定指标体系后，即可围绕各指标对参与评价的DC-DC变换器进行参数测量和指标值的计算，并对所得数据进行处理。

1.2 评价指标类型的一致化

评价指标可分为以下类型[13]：

1）极大型指标。指标值越大越好，如效率。

2）极小型指标。指标值越小越好，如电压和电流应力。

3）中间型指标。指标值越接近某个值越好，如输出电压精度越接近0%越好。

4）区间型指标。指标值越靠近某个区间越好。区间型指标可由中间型指标转化而来，如果认为输出电压纹波系数在区间[0%,1%]均为最佳，则该指标为区间型指标。

如果指标体系中包含不同类型的指标，则必须将其转化为相同类型，即一致化，如将各种类型的指标都转化为极大型指标。

对于极小型指标x，令

$$x' = M - x \tag{1}$$

或

$$x' = 1/x \qquad x \neq 0 \tag{2}$$

式中：M为指标x的允许上界。

对于中间型指标和区间型指标，可通过计算指标数值与最佳值或最佳区间的距离，转化为极小型指标，再转化为极大型指标。

1.3 评价指标值的无量纲化

由于各指标量纲和数量级不同，具有不可公度性，需对指标值进行无量纲化，并转化为同一数量级，以便于比较。常用的无量纲化方法有以下几种：

1）极值处理法。对于极大型指标，处理后的指标值为

$$x_{ij}^* = \frac{x_{ij} - m_j}{M_j - m_j} \tag{3}$$

式中：$M_j = \max_i\{x_{ij}\}$；$m_j = \min_i\{x_{ij}\}$；x_{ij}为第i个被评价对象的第j个评价指标的值。

对于极小型指标，处理后的指标值为

$$x_{ij}^* = \frac{M_j - x_{ij}}{M_j - m_j} \tag{4}$$

经过极值处理后，指标值的最大值为1，最小值为0。

2）归一化处理法。处理后的指标值为

$$x_{ij}^* = x_{ij} \Big/ \sum_{i=1}^{m} x_{ij} \tag{5}$$

式中：m为被评价对象的个数。

归一化处理后，当$x_{ij}>0$时，$x_{ij}^* \in (0,1)$，但没有固定的最大最小值，且$\sum_{i=1}^{m} x_{ij}^* = 1$。

2 赋权方法

在综合评价的指标体系中，各指标的重要性或贡献度不一定相同。为体现指标的相对重要性，需要确定指标的权重，即赋权。赋权方法可分为主观赋权法、客观赋权法和组合赋权法。主观赋权法是评价者基于知识和经验，根据各指标的主观重要程度进行赋权的方法，包括层次分析法（analytic hierarchy process, AHP）[14-16]、序关系法[17]、集值迭代法[18]等。客观赋权法根据已测得的指标值计算权重，包括均方差法[6]、CRITIC法[17]、熵值法[10-11,19]等。组合赋权法将主观赋权法和客观赋权法结合起来，使权重同时反映评价者主观的判断和客观数据的影响。本文采用层次分析-熵值法对DC-DC变换器的指标进行赋权。

2.1 层次分析法

层次分析法是一种将评价者的主观判断进行客观量化的多准则决策方法，该方法将决策问题分解成目标层、准则层、方案层等层次，形成一个层次结构模型。分析比较每层中每2个因素之间相对于上一层所属因素的重要性，形成判断矩阵。根据判断矩阵计算最低层每个因素的权重，按权重排

序后即可作出决策。若仅将层次分析法用于指标体系中的指标赋权, 步骤如下[15]:

步骤 1　将指标体系分为目标层、二级指标层和指标层, 并为每个指标编号。例如, 首先, 将 DC-DC 变换器的拓扑性能定为目标层; 然后, 将拓扑性能分为 5 个方面, 即输出质量 A、效率 B、开关管应力 C、二极管应力 D、电感电容参数 E, 形成二级指标层; 最后, 在二级指标下选取 11 个具体的指标, 如输出电压精度 A_1、输出电压纹波系数 A_2、效率 B_1、开关管电压应力 C_1、开关管电流应力 C_2、二极管电压应力 D_1、二极管电流应力 D_2、电感电流应力 E_1、最小电感值 E_2、电容电压应力 E_3、最小电容值 E_4 等, 形成指标层。

步骤 2　对每层中属于相同类别的指标进行两两比较, 形成判断矩阵。设属于相同类别的 n 个指标分别为 $X_1, X_2, \cdots, X_n$, 则判断矩阵 $\boldsymbol{C}$ 为 n 阶方阵, 如下:

$$\boldsymbol{C}=\begin{array}{c} \\ X_1 \\ X_2 \\ \vdots \\ X_n \end{array}\begin{array}{c} \begin{array}{cccc} X_1 & X_2 & \cdots & X_n \end{array} \\ \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \end{array} \tag{6}$$

式中: c_{ij} 为指标 X_i 相比于指标 X_j 的重要性数值, 常用 9 级标度法确定, 如表 1 所示。

表 1　9 级标度及其含义

Table 1　9-level scale and its meaning

标度值	含义
1	X_i与X_j同等重要
3	X_i比X_j稍微重要
5	X_i比X_j明显重要
7	X_i比X_j强烈重要
9	X_i比X_j极端重要
2, 4, 6, 8	重要性介于1, 3, 5, 7, 9之间
1～9的倒数	X_j比X_i重要, 且$c_{ij}=1/c_{ji}$

步骤 3　一致性校验。判断矩阵涉及多个指标的两两比较, 可能会出现矛盾情形, 使其一致性变差。因此, 引入如下一致性指标:

$$I_{\mathrm{C}}=\frac{\lambda_{\max}-n}{n-1} \tag{7}$$

式中: $\lambda_{\max}$ 为判断矩阵的最大特征值; n 为指标个数, 即矩阵阶数。

进一步, 计算一致性比例:

$$R_{\mathrm{C}}=I_{\mathrm{C}}/I_{\mathrm{R}} \tag{8}$$

式中: I_{R} 为平均随机一致性指标, 取值如表 2 所示。

若 $R_{\mathrm{C}}<0.1$, 则一致性校验通过, 否则需重新调整判断矩阵。

表 2　I_{R} 的取值

Table 2　Value of I_{R}

n	I_{R}
3	0.58
4	0.9
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45

步骤 4　根据判断矩阵, 计算其中指标的权重。先计算判断矩阵各行平均值, 定义算术平均值为

$$\overline{w_i}=\frac{1}{n}\sum_{j=1}^{n}c_{ij} \tag{9}$$

再将各行平均值归一化为权重, 即

$$w_i=\overline{w_i}\Big/\sum_{k=1}^{n}\overline{w_k} \tag{10}$$

以同样方法计算其他指标和二级指标的权重, 得到指标的总权重 w_{s} 为

$$w_{\mathrm{s}}=w_{\mathrm{I}}w_{\mathrm{II}} \tag{11}$$

式中: w_{I} 为指标相对于二级指标的权重; w_{II} 为二级指标相对于目标层的权重。2 个权重来源于不同层次的判断矩阵, 由式 (10) 计算得到。

2.2　熵 值 法

熵值法是基于各指标所含信息量大小的赋权方法, 具有客观性, 可避免权重受到人为因素过多的干扰。在信息论中, 事件所含信息量的大小用信息熵衡量[20]。在综合评价中, 将同一指标的数值归一化并视为概率, 若同一指标的数值差异较小, 即概率分布较均匀, 则意味着“某个数值属于某个被评价对象”这一事件具有较大的不确定性, 信息熵较大, 因此, 赋予该指标较低的权重。反之, 若同一指标的数值差异较大, 则该指标的权重较高[13]。

设决策矩阵为

$$\boldsymbol{D}=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \tag{12}$$

式中: m 为被评价对象的个数; n 为评价指标的个数。

运用熵值法确定指标权重的步骤如下:

步骤 1　计算第 j 项评价指标下第 i 个被评价对象的归一化数值, 得

$$p_{ij}=x_{ij}\Big/\sum_{i=1}^{m}x_{ij} \tag{13}$$

步骤 2 计算第 j 项指标的熵值，得

$$e_j=-\frac{1}{\ln m}\sum_{i=1}^{m}p_{ij}\ln p_{ij} \tag{14}$$

步骤 3 计算第 j 项指标的偏差度，得

$$d_j=1-e_j \tag{15}$$

步骤 4 将第 j 项指标的偏差度归一化，即权重为

$$w_{oj}=d_j\Big/\sum_{k=1}^{n}d_k \tag{16}$$

2.3 组合赋权法

组合赋权法将主观赋权法所得权重与客观赋权法所得权重进行加权平均[13]，即

$$w_j=kw_{sj}+(1-k)w_{oj} \tag{17}$$

式中：w_{sj} 为第 j 项指标的主观权重，由式 (11) 计算；w_{oj} 为客观权重，由式 (16) 计算；k 为系数，取值范围为 $0<k<1$，反映对主观权重的侧重程度。若认为主观权重和客观权重同等重要，则 k=0.5。

3 评价方法

构建指标体系并为指标赋权后，即可选择合适的方法进行综合评价。常用的综合评价方法有线性加权综合法[6]、模糊综合评价法[7-9]、云模型评价法[17,19]、灰色关联分析法[21]、理想点排序法[12]及基于神经网络的评价方法[6,16]等。

3.1 线性加权综合法

线性加权综合法用线性模型来进行综合评价，如下：

$$y_i=\sum_{j=1}^{m}x_{ij}w_j \tag{18}$$

式中：y_i 为第 i 个被评价对象的综合评分值。

线性加权综合法的特点是简单和直观，适用于各评价指标相互独立且可以线性互补的场合。

3.2 模糊综合评价法

模糊综合评价法适用于涉及模糊概念的评价。将指标值按隶属函数划分为若干评价等级，但指标值可以同时属于不同的等级，即不同等级的边界是模糊的[22]。

以效率指标为例，设评价等级的集合为{高，中，低}，某对象的无量纲化效率值为 0.7。对上述 3 个等级分别设置隶属函数，且和为 1。将效率值分别代入 3 个隶属函数，设所得该对象关于效率的评判向量为

$$\boldsymbol{R}_j=[0.8\quad 0.2\quad 0] \tag{19}$$

式 (19) 表示，对于 0.7 的效率值，有 80% 的概率属于“高”，有 20% 的概率属于“中”，体现出模糊的特点。进一步地，将该对象的全部指标的评判向量组成评判矩阵，结合赋权方法获得的各指标权重，通过矩阵运算得出模糊综合评价的结果。

模糊综合评价法可以仅对一个对象进行综合评价。当只有一个对象时，评价的目的是确定该对象所处的等级，而不是与其他对象对比。在这种情况下，利用不同等级之间的模糊属性可实现更精确的评价。然而，隶属函数的选取基于评价者经验，难以准确反映每个指标的值的分布特点。

因此，本文采用线性加权综合法对一系列 DC-DC 变换器进行评价，因为其适用于多个对象的评价，计算量小，且 DC-DC 变换器各评价指标之间的独立性较高。

4 算例分析

运用磁通平衡设计法可以构造出 11 种电压增益为 D^2 的 DC-DC 变换器，如附录 A 中的图 A1 所示。本节基于所提的 DC-DC 变换器综合评价方法对这 11 种变换器进行比较。

4.1 综合评价过程

综合评价流程如图 1 所示，具体步骤如下：

步骤 1 构建指标体系。由于参与评价的 DC-DC 变换器具有相同的元器件和电压增益，可选取

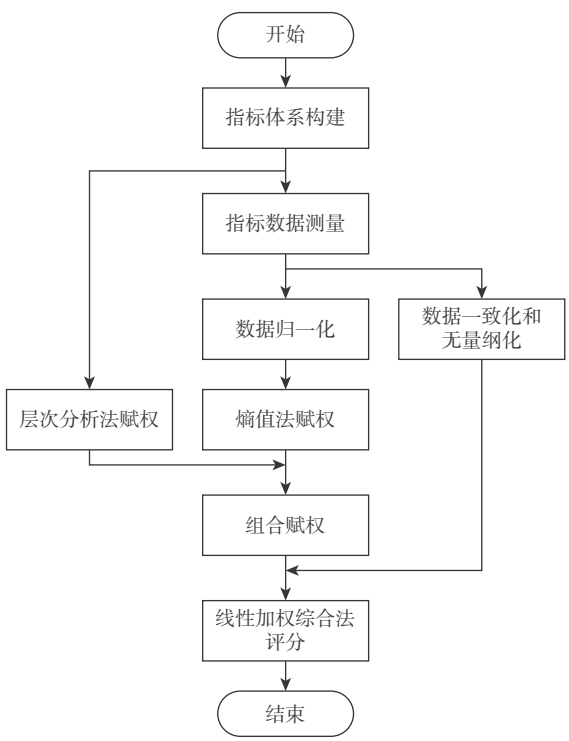


图 1 综合评价流程

Fig. 1 Flow chart of comprehensive evaluation

2.1 节中所述的评价指标。

步骤 2　对参与评价的所有 DC-DC 变换器进行仿真, 测量指标数据。本文以一种常见的降压型电源的规格为参考, 设定输入电压为 48 V, 输出电压为 12 V, 额定功率为 48 W。各拓扑的电感和电容参数采用指标体系中的最小电感值和最小电容值, 计算方式如下:

$$L=\frac{V_{LD}D}{fI_L\Delta I_L} \tag{20}$$

$$C=\frac{I_{CD}D}{fV_C\Delta V_C} \tag{21}$$

式中: f 为开关频率; D 为占空比; V_{LD} 为 D 时段电感 L 的平均电压; I_L 为电感 L 的平均电流; ΔI_L 为电感电流纹波系数, 通常取 20%～40%, 本文取 40%; I_{CD} 为 D 时段电容 C 的平均电流; V_C 为电容 C 的平均电压; ΔV_C 为电容电压纹波系数, 按实际需求选取, 本文取 1%。

除电感和电容外, 其他参数相同, 如表 3 所示。在 MATLAB/Simulink 中进行仿真, 并测量各项指标的数据, 如附录 A 中的表 A1 所示。

表 3　仿真参数

Table 3　Simulation parameters

参数	数值
输入电压/V	48
参考输出电压/V	12
占空比	0.5
开关管导通电阻/Ω	0.05
二极管导通压降/V	0.9
二极管导通电阻/Ω	0.01
负载电阻/Ω	3
开关频率/kHz	100

步骤 3　用极值处理法对数据进行一致化和无量纲化。指标化为极大型, 且最优值为 1, 最差值为 0, 结果如附录 A 中的表 A2 所示。无量纲化的数据即为各个对象中各指标的评分。

步骤 4　运用层次分析法赋权。运用 9 级标度法, 拟定二级指标的判断矩阵, 包括输出质量 A、效率 B、开关管应力 C、二极管应力 D、电感电容参数 E。在 5 个二级指标中, 由于 DC-DC 变换器需要先实现给定的输出, 输出质量最重要。效率作为最常用于比较的指标之一, 重要性与输出质量相当。余下的元器件应力指标中, 可按元器件价格对指标排序。例如, 开关管价格最贵, 则认为开关管应力最重要。对每个二级指标两两比较, 所得判断矩阵如下:

$$\boldsymbol{C}_2=\begin{array}{c} \\ A\\B\\C\\D\\E\end{array}\begin{array}{c}\begin{array}{ccccc}A&B&C&D&E\end{array}\\\begin{bmatrix}1&1&2&3&3\\1&1&2&3&3\\0.5&0.5&1&2&2\\0.333&0.333&0.5&1&1\\0.333&0.333&0.5&1&1\end{bmatrix}\end{array} \tag{22}$$

对式 (22) 进行一致性校验, 该矩阵的最大特征值为 5.013 3。一致性指标 I_C 按式 (7) 计算, 为 0.003 3。一致性比例 R_C 按式 (8) 计算, 为 0.003。因为 R_C<0.1, 所以判断矩阵 (22) 的一致性校验通过。

计算矩阵 (22) 的各行平均值并归一化, 结果即为权重, 如表 4 所示。

表 4　各二级指标判断矩阵的行平均值和指标权重

Table 4　Row averages of judgment matrix and weights of second level indices

行	行平均值	权重
A	2	0.309 3
B	2	0.309 3
C	1.2	0.185 6
D	0.633 3	0.097 9
E	0.633 3	0.097 9

步骤 5　拟定各指标相对于二级指标的判断矩阵, 其中, 二级指标输出质量 A 下有 2 个指标: 输出电压精度 A_1 和输出电压纹波系数 A_2。上述 2 个指标中, 认为输出电压精度最重要。而输出电压纹波系数在选定输出电容时已经设为相同, 因此, 所有拓扑的输出电压纹波应该是接近的, 可认为该指标不重要。由此形成的判断矩阵如下:

$$\boldsymbol{C}_{1A}=\begin{array}{c} \\ A_1\\A_2\end{array}\begin{array}{c}\begin{array}{cc}A_1&A_2\end{array}\\\begin{bmatrix}1&4\\0.25&1\end{bmatrix}\end{array} \tag{23}$$

该矩阵只有 2 阶, 且副对角线的 2 个元素互为倒数, 则该矩阵的一致性校验通过。

计算矩阵 (23) 的各行平均值, 并归一化为权重, 结果如表 5 所示。

表 5　输出质量下各指标的行平均值和指标权重

Table 5　Row averages and weights of indices affiliated to "output quality"

行	行平均值	权重
A_1	2.5	0.8
A_2	0.625	0.2

对于二级指标效率 B, 其下属只有一个指标效率 B_1, 因此, 其权重为 1。

对于二级指标开关管应力和二极管应力, 其各有电压应力和电流应力之分。在本文采用的参数

下，开关管电压应力为几十伏，电流应力为几安。通过调研常用开关管和二极管的参数，发现耐流大的开关器件价格较高，故认为电流应力较重要。经过列写判断矩阵和一致性校验，得到电压应力和电流应力的权重分别为 0.25 和 0.75。

对于二级指标电感电容参数 E，其下属的指标包括电感电流应力 E_1、最小电感值 E_2、电容电压应力 E_3、最小电容值 E_4。由于耐流大、电感值大的电感体积和质量较大，降低了变换器的功率密度，认为电感参数比电容参数重要。进一步，认为电感值和耐流对电感规格的影响相同，电容值和耐压也是如此。经过列写判断矩阵和一致性校验，4 个指标的权重分别为 0.333、0.333、0.167 和 0.167。

步骤 6 按式 (11) 计算总权重，即为层次分析法所得权重，如表 6 所示。

表 6 层次分析法所得总权重

Table 6 Total weights obtained by AHP

指标	权重
输出电压精度A_1	0.247 4
输出电压纹波系数A_2	0.061 9
效率B_1	0.309 3
开关管电压应力C_1	0.046 4
开关管电流应力C_2	0.139 2
二极管电压应力D_1	0.024 5
二极管电流应力D_2	0.073 5
电感电流应力E_1	0.032 6
最小电感值E_2	0.032 6
电容电压应力E_3	0.016 3
最小电容值E_4	0.016 3

步骤 7 运用熵值法赋权。将附录 A 中表 A1 所示的指标数据按式 (13) 归一化，如附录 A 中表 A3 所示。

分别按式 (14) 和式 (15) 计算各指标的熵和偏差度，再按式 (16) 将偏差度归一化，结果即为权重，如表 7 所示。

步骤 8 组合赋权。熵值法虽然是客观的，但容易受到数据波动的影响，有一定的随机性和不确定性。而层次分析法取决于评价者的判断，虽然可能不完全准确，但基本能保证专业性。因此，本文略侧重于主观权重。在式 (17) 中，取 k=0.6 计算组合权重，如表 8 所示。相应雷达图如图 2 所示。

步骤 9 运用线性加权综合法，利用附录 A 中表 A2 所示的各指标无量纲化数据和表 8 所示的权重，按式 (18) 计算各 DC-DC 变换器拓扑的评分值，满分为 1，得到的结果如表 9 所示，相应的雷达图如图 3 所示。

表 7 指标的熵、偏差度和权重

Table 7 Entropy, deviations, and weights of indices

指标	熵	偏差度	权重
输出电压精度A_1	0.994 9	0.005 1	0.055 5
输出电压纹波系数A_2	0.999 9	0.000 1	0.001 3
效率B_1	1	0	0.000 1
开关管电压应力C_1	0.993 1	0.006 9	0.074 8
开关管电流应力C_2	0.995 9	0.004 1	0.044 5
二极管电压应力D_1	0.993 1	0.006 9	0.074 4
二极管电流应力D_2	0.996 4	0.003 6	0.039 5
电感电流应力E_1	0.996	0.004	0.043 9
最小电感值E_2	0.992 2	0.007 8	0.084 1
电容电压应力E_3	0.995 6	0.004 4	0.047 9
最小电容值E_4	0.950 8	0.049 2	0.534

表 8 指标的组合权重

Table 8 Combined weights of indices

指标	权重		组合权重
	层次分析法	熵值法	
输出电压精度A_1	0.247 4	0.055 5	0.170 7
输出电压纹波系数A_2	0.061 9	0.001 3	0.037 6
效率B_1	0.309 3	0.000 1	0.185 6
开关管电压应力C_1	0.046 4	0.074 8	0.057 8
开关管电流应力C_2	0.139 2	0.044 5	0.101 3
二极管电压应力D_1	0.024 5	0.074 4	0.044 5
二极管电流应力D_2	0.073 5	0.039 5	0.059 9
电感电流应力E_1	0.032 6	0.043 9	0.037 2
最小电感值E_2	0.032 6	0.084 1	0.053 2
电容电压应力E_3	0.016 3	0.047 9	0.028 9
最小电容值E_4	0.016 3	0.534	0.223 4

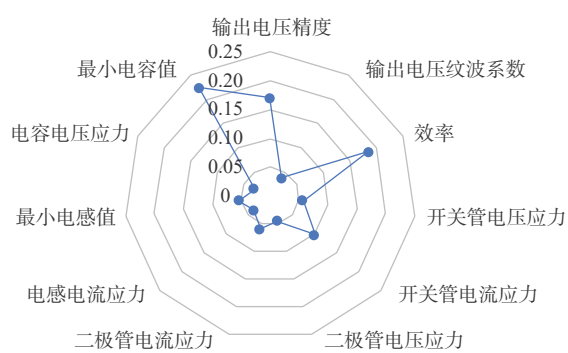


图 2 各指标权重的雷达图

Fig. 2 Radar chart of weights of indices

4.2 权重及评价结果分析

根据表 8 和图 2 所示的各指标权重，输出电压精度、效率、最小电容值等 3 个指标的权重较高，符合预期。其中，最小电容值的权重主要由熵值法贡献，该指标的值最分散，因此，具有最明显的区分度。

根据表 9 和图 3 所示的评价结果，拓扑 1 得到

表 9　各拓扑的评分值

Table 9　Evaluation values of topologies

拓扑	评分值
1	0.795 5
2	0.601 8
3	0.639 8
4	0.473 6
5	0.429 5
6	0.293 9
7	0.458 0
8	0.606 2
9	0.502 5
10	0.737 2
11	0.782 6

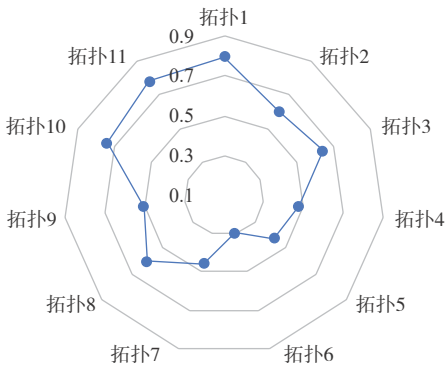


图 3　各拓扑评分值的雷达图

Fig. 3　Radar chart of evaluation values of scores for each topologies

了最高的 0.795 5 分, 为最优拓扑。拓扑 1 的原理图和无量纲化数据分别如附录 A 中图 A1(a) 和表 A2 所示。该拓扑在效率、开关管电流应力、二极管电流应力、最小电感值、最小电容值等 5 个指标中的评分均接近满分。此外, 在高权重的输出电压精度上的表现也较好。

除拓扑 1 之外, 拓扑 10 和 11 的评分也较高, 超过了 0.7 分。分析可知, 上述拓扑在权重最高的输出电压精度、效率、最小电容值等 3 个指标中均表现较好, 另有多项指标接近 1 分, 因此, 评分较高。

拓扑 5 和 6 的评分最低, 原因是其在高权重指标中表现一般, 特别是在权重最高的最小电容值指标中接近 0 分。分析可知, 上述 2 个拓扑中的 2 个电容直接相连, 且同时充放电, 导致充放电电流较大。根据式 (21), 其最小电容值远大于其他拓扑。因此, 若对电容值或电容电压纹波的要求较高, 则应避免采用电容直连且同时充放电的结构。

5　结　论

为评价 DC-DC 变换器的性能, 本文提出了一种基于层次分析-熵值法的 DC-DC 变换器拓扑综合评价方法, 并以电压增益为 D^2 的 11 个 DC-DC 变换器拓扑为例, 对所提出的评价方法进行了论证, 结论如下:

1) 从 DC-DC 变换器拓扑的角度构建了一种包含输出电压精度、输出电压纹波系数、效率等 11 个指标的指标体系, 较为全面地反映了变换器各方面的性能。

2) 通过主观的层次分析法和客观的熵值法, 并采用侧重于主观权重的策略, 得到的组合权重基本符合预期。如果指标权重差距较悬殊, 可调整层次分析法中的判断矩阵, 避免部分指标因权重过低而几乎不发挥作用。

3) 线性加权综合法较好地完成了对多个对象的综合评价, 找出了最优拓扑和最差拓扑。分析最优和最差拓扑的各项指标值和拓扑特点, 可为今后构造新拓扑提供参考。

4) 提出的 DC-DC 变换器综合评价方法有明确的步骤, 可根据流程图编写计算机程序进行评价, 从而实现 DC-DC 变换器拓扑的自动推荐。

参考文献 (References)

[1] 李虹, 王文财, 李亚敏, 等. 基于图论的 $S^1D^1L^2C^1$ 型 DC-DC 变换器可编程拓扑搜索算法[J]. 中国电机工程学报, 2021, 41(16): 5670-5683.
LI H, WANG W C, LI Y M, et al. Programmable topology searching algorithm for $S^1D^1L^2C^1$ type DC-DC converters based on graph theory[J]. Proceedings of the CSEE, 2021, 41(16): 5670-5683(in Chinese).

[2] PANIGRAHI R, MISHRA S K, JOSHI A. Synthesis of an optimum converter topology for a specified voltage conversion ratio[J]. IEEE Transactions on Industry Applications, 2021, 57(4): 3923-3934.

[3] CHEN Y, BAI J B, KANG Y. A nonisolated single-inductor multiport DC-DC topology deduction method based on reinforcement learning[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2022, 10(6): 6572-6585.

[4] MO L P, CHEN G P, HUANG J M, et al. Graph theory-based programmable topology derivation of multiport DC-DC converters with reduced switches[J]. IEEE Transactions on Industrial Electronics, 2022, 69(6): 5745-5755.

[5] DONG M, LIANG R J, YANG J, et al. Topology derivation of multiport DC-DC converters based on reinforcement learning[J]. IEEE Transactions on Power Electronics, 2023, 38(4): 5055-5064.

[6] 何永秀, 刘敦楠, 罗涛, 等. 电力综合评价方法及应用[M]. 北京: 中国电力出版社, 2011: 13-194.
HE Y X, LIU D N, LUO T, et al. Comprehensive evaluation method of electric power and its application[M]. Beijing: China Electric Power Press, 2011: 13-194(in Chinese).

[7] 石宜金, 谭贵生, 赵波, 等. 基于模糊综合评估模型与信息融合的

电力变压器状态评估方法[J]. 电力系统保护与控制, 2022, 50(21): 167-176.
SHI Y J, TAN G S, ZHAO B, et al. Condition assessment method for power transformers based on fuzzy comprehensive evaluation and information fusion[J]. Power System Protection and Control, 2022, 50(21): 167-176(in Chinese).

[8] 朱晓荣, 彭柏, 司羽, 等. 基于知识图谱的配电网综合评价[J]. 现代电力, 2022, 39(6): 677-684.
ZHU X R, PENG B, SI Y, et al. Comprehensive evaluation of distribution network based on knowledge graph[J]. Modern Electric Power, 2022, 39(6): 677-684(in Chinese).

[9] WANG J H, LIU S Q, WANG S X, et al. Multiple indicators-based health diagnostics and prognostics for energy storage technologies using fuzzy comprehensive evaluation and improved multivariate grey model[J]. IEEE Transactions on Power Electronics, 2021, 36(11): 12309-12320.

[10] 颜湘武, 赵帅帅, 董清, 等. 电动汽车充电机性能综合评估[J]. 电力系统保护与控制, 2020, 48(1): 164-171.
YAN X W, ZHAO S S, DONG Q, et al. Comprehensive evaluation of electric vehicle charger performance[J]. Power System Protection and Control, 2020, 48(1): 164-171(in Chinese).

[11] 王旭, 孟润泉, 韩肖清, 等. 基于改进AHP法-熵权法的交直流母线接口变换器综合性能评估[J]. 电网与清洁能源, 2021, 37(3): 8-16.
WANG X, MENG R Q, HAN X Q, et al. Comprehensive performance evaluation of AC/DC bus interface converter based on improved AHP-entropy weight method[J]. Power System and Clean Energy, 2021, 37(3): 8-16(in Chinese).

[12] 曾晔, 加鹤萍, 杨菁, 等. 基于模糊层次分析法–熵权法–逼近理想解排序法的虚拟电厂综合贡献度评估方法[J]. 现代电力, 2024, 41(1): 144-151.
ZENG Y, JIA H P, YANG J, et al. Comprehensive contribution degree evaluation method of virtual power plants based on FAHP-EWM-TOPSIS method[J]. Modern Electric Power, 2024, 41(1): 144-151(in Chinese).

[13] 易平涛, 李伟伟, 郭亚军. 综合评价理论与方法[M]. 2版. 北京: 经济管理出版社, 2019: 12-83.
YI P T, LI W W, GUO Y J. Comprehensive evaluation theory and methods[M]. 2nd ed. Beijing: Economy & Management Publishing House, 2019: 12-83(in Chinese).

[14] 洪晟, 岳天羽, 李新建. 基于层次分析的信息物理网络多维攻击损毁评估[J]. 北京航空航天大学学报, 2022, 48(9): 1827-1835.
HONG S, YUE T Y, LI X J. Multidimensional attack damage assessment of cyber-physical systems based on hierarchical analysis[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(9): 1827-1835(in Chinese).

[15] 付志扬, 王涛, 孔令号, 等. 基于AHP-TOPSIS算法的重要电力客户用电状态评估[J]. 电网技术, 2022, 46(10): 4095-4101.
FU Z Y, WANG T, KONG L H, et al. Power consumption state evaluation of important power customers based on AHP-TOPSIS algorithm[J]. Power System Technology, 2022, 46(10): 4095-4101(in Chinese).

[16] GE Q B, QIAO H Z, LI C X, et al. Real-time charging risk assessment for electric vehicles based on improved broad BP-AHP[J]. IEEE Transactions on Industrial Electronics, 2022, 69(9): 9472-9482.

[17] 李敬强, 樊天辰, 周妍汝, 等. 基于云模型的民航监察员队伍能力综合评价[J]. 北京航空航天大学学报, 2022, 48(12): 2425-2433.
LI J Q, FAN T C, ZHOU Y R, et al. Comprehensive evaluation on capability of civil aviation supervisor team based on cloud model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(12): 2425-2433(in Chinese).

[18] 黄伟, 杨子力, 柳思岐. 基于物元可拓模型的特色小镇能源系统综合评价[J]. 现代电力, 2020, 37(5): 448-457.
HUANG W, YANG Z L, LIU S Q. Comprehensive energy system evaluation of new energy featured towns based on matter-element extension model[J]. Modern Electric Power, 2020, 37(5): 448-457(in Chinese).

[19] 高书垚, 安泰, 宋剑. 基于熵权-正态云模型的智能电能表状态评估研究[J]. 电测与仪表, 2022, 59(1): 190-194.
GAO S Y, AN T, SONG J. Status evaluation of smart meter based on entropy weight-normal cloud model[J]. Electrical Measurement & Instrumentation, 2022, 59(1): 190-194(in Chinese).

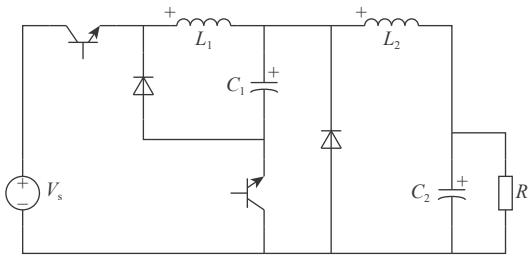
[20] LIU Z Y, WANG Y S, CHENG Q X, et al. Analysis of the information entropy on traffic flows[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18012-18023.

[21] 李智华, 张宇浩, 吴春华, 等. 基于动态指标的光伏组件健康程度诊断[J]. 电网技术, 2024, 48(2): 597-606.
LI Z H, ZHANG Y H, WU C H, et al. Health diagnosis of photovoltaic modules based on dynamic health criteria[J]. Power System Technology, 2024, 48(2): 597-606(in Chinese).

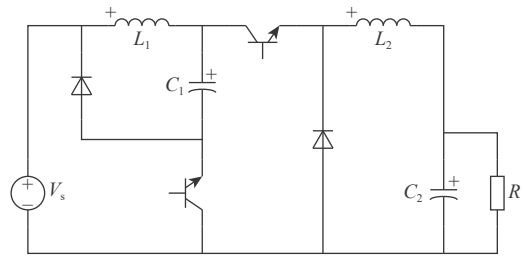
[22] 杜栋, 庞庆华, 吴炎. 现代综合评价方法与案例精选[M]. 3版. 北京: 清华大学出版社, 2015: 37-49.
DU D, PANG Q H, WU Y. Modern comprehensive evaluation method and case selection[M]. 3rd ed. Beijing: Tsinghua University Press, 2015: 37-49(in Chinese).

附录 A

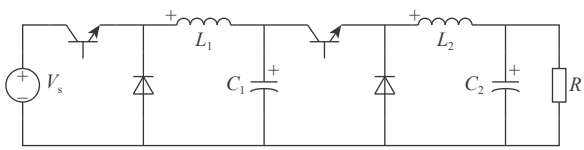
本文算例中11种电压增益为 D^2 的DC-DC变换器的原理图如图A1所示。原始仿真数据及处理后的数据如表A1～表A3所示。



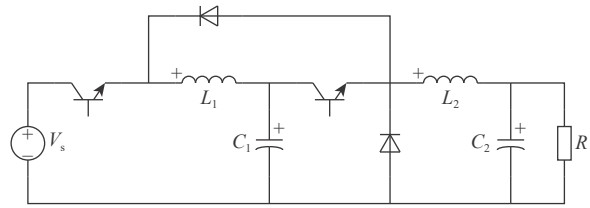
(a) 拓扑1



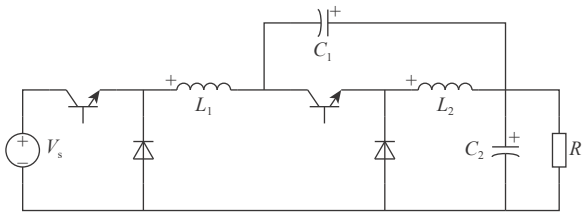
(b) 拓扑2



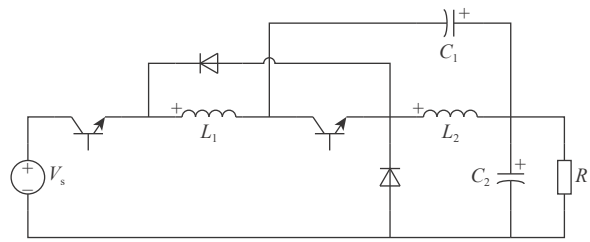
(c) 拓扑3



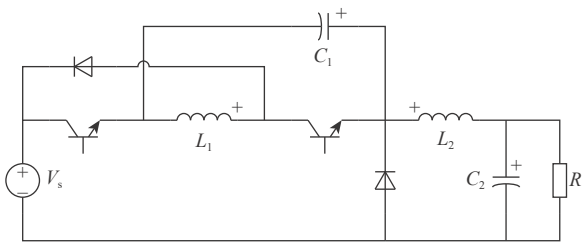
(d) 拓扑4



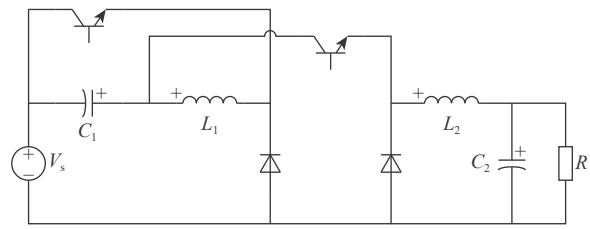
(e) 拓扑5



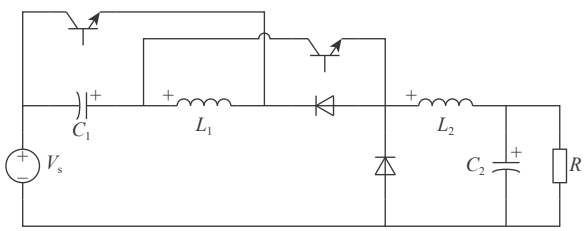
(f) 拓扑6



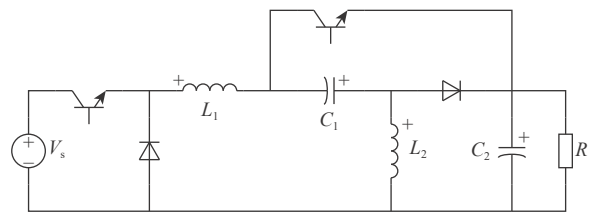
(g) 拓扑7



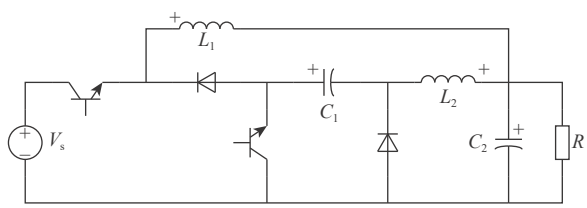
(h) 拓扑8



(i) 拓扑9



(j) 拓扑10



(k) 拓扑11

图 A1　电压增益为 D^2 的 11 个 DC-DC 变换器

Fig. A1　11 DC-DC converters with a voltage gain of D^2

表 A1 各拓扑的各项指标数值

Table A1 Index values of topologies

拓扑	输出电压精度/%	输出电压纹波系数/%	效率/%	开关管电压应力/V	开关管电流应力/A	二极管电压应力/V	二极管电流应力/A	电感电流应力/A	最小电感值/μH	电容电压应力/V	最小电容值/μF
1	6.193	0.987	91.002	49.04	2.265	35.79	3.345	3.386	101.200	17.483	28.370
2	6.454	1.001	90.523	61.21	3.365	35.76	3.34	3.378	101.200	17.470	28.370
3	6.627	0.987	90.395	36.725	3.405	35.685	3.335	3.372	101.200	17.432	28.370
4	8.773	1.001	90.000	36.98	3.3	24.095	4.36	3.288	104.392	17.067	28.355
5	6.688	1.006	90.322	36.725	3.405	35.685	3.335	3.369	101.200	11.802	80.266
6	8.667	0.999	90.065	36.99	3.3	24.095	4.365	3.292	104.392	11.571	80.208
7	8.185	0.958	89.377	37.01	3.3	24.105	4.37	3.290	104.392	18.066	26.908
8	5.621	0.949	89.432	36.775	3.42	35.735	3.345	3.385	101.200	17.961	27.633
9	8.170	1.045	90.180	37.01	3.335	24.105	4.405	3.326	104.392	18.069	26.908
10	5.924	0.984	91.146	36.76	3.43	35.72	3.355	2.259	161.921	17.509	28.370
11	6.170	0.988	91.026	49.04	2.265	35.79	3.345	2.254	161.921	17.483	28.370

表 A2 指标数值一致化和无量纲化

Table A2 Harmonization and dimensionlessness of index values

拓扑	输出电压精度	输出电压纹波系数	效率	开关管电压应力	开关管电流应力	二极管电压应力	二极管电流应力	电感电流应力	最小电感值	电容电压应力	最小电容值
1	0.818 662	0.607 535	0.918 425	0.497 039	1	0	0.990 654	0	1	0.090 316	0.972 584
2	0.735 66	0.458 493	0.647 763	0	0.055 794	0.002 565	0.995 327	0.006 625	1	0.092 243	0.972 584
3	0.680 941	0.606 218	0.575 519	1	0.021 459	0.008 978	1	0.012 367	1	0.098 16	0.972 584
4	0	0.459 766	0.352 407	0.989 585	0.111 588	1	0.042 056	0.086 131	0.947 438	0.154 326	0.972 875
5	0.661 38	0.413 818	0.534 523	1	0.021 459	0.008 978	1	0.014 576	1	0.964 504	0
6	0.033 836	0.481 399	0.388 856	0.989 177	0.111 588	1	0.037 383	0.082 597	0.947 438	1	0.001 078
7	0.186 624	0.913 162	0	0.988 360	0.111 588	0.999 145	0.032 71	0.084 364	0.947 438	0.000 481	1
8	1	1	0.031 351	0.997 958	0.008 584	0.004 703	0.990 654	0.000 883	1	0.016 646	0.986 407
9	0.191 383	0	0.454 215	0.988 36	0.081 545	0.999 145	0	0.052 562	0.947 438	0	1
10	0.903 78	0.636 79	1	0.998 571	0	0.005 985	0.981 308	0.995 141	0	0.086 254	0.972 584
11	0.825 8	0.591 535	0.932 46	0.497 039	1	0	0.990 654	1	0	0.090 201	0.972 584

表 A3 指标数值归一化

Table A3 Normalization of index values

拓扑	输出电压精度	输出电压纹波系数	效率	开关管电压应力	开关管电流应力	二极管电压应力	二极管电流应力	电感电流应力	最小电感值	电容电压应力	最小电容值
1	0.079 9	0.090 5	0.091 6	0.108 0	0.065 1	0.103 3	0.081 8	0.097 9	0.081 1	0.096 1	0.068 8
2	0.083 3	0.091 8	0.091 1	0.134 7	0.096 7	0.103 2	0.081 7	0.097 6	0.081 1	0.096 0	0.068 8
3	0.085 5	0.090 5	0.091 0	0.080 8	0.097 9	0.103 0	0.081 5	0.097 5	0.081 1	0.095 8	0.068 8
4	0.113 2	0.091 8	0.090 6	0.081 4	0.094 9	0.069 5	0.106 6	0.095 0	0.083 7	0.093 8	0.068 8
5	0.086 3	0.092 2	0.090 9	0.080 8	0.097 9	0.103 0	0.081 5	0.097 4	0.081 1	0.064 9	0.194 8
6	0.111 9	0.091 6	0.090 7	0.081 4	0.094 9	0.069 5	0.106 7	0.095 2	0.083 7	0.063 6	0.194 6
7	0.105 7	0.087 8	0.090 0	0.081 5	0.094 9	0.069 6	0.106 8	0.095 1	0.083 7	0.099 3	0.065 3
8	0.072 6	0.087 0	0.090 0	0.081 0	0.098 3	0.103 1	0.081 8	0.097 8	0.081 1	0.098 7	0.067 0
9	0.105 5	0.095 9	0.090 8	0.081 5	0.095 9	0.069 6	0.107 7	0.096 1	0.083 7	0.099 3	0.065 3
10	0.076 5	0.090 2	0.091 7	0.080 9	0.098 6	0.103 1	0.082 0	0.065 3	0.129 8	0.096 2	0.068 8
11	0.079 6	0.090 6	0.091 6	0.108 0	0.065 1	0.103 3	0.081 8	0.065 1	0.129 8	0.096 1	0.068 8

Comprehensive evaluation of DC-DC converters based on analytic hierarchy process and entropy method

ZHANG Yang，QIU Dongyuan*，ZHANG Bo，CHEN Yanfeng

(School of Electric Power Engineering，South China University of Technology，Guangzhou 510641，China)

Abstract: In order to compare and optimize a series of DC-DC converters with similar topological structures and functions, a comprehensive topology evaluation method based on the analytic hierarchy process and entropy method is proposed. A hierarchical evaluation index system is created by compiling the common DC-DC converter performance indices. Analytic hierarchy process and entropy methods are used to assign weights to the indices in the index system, and the strategy focusing on analytic hierarchy process was adopted to get the combined weights. Additionally, each DC-DC converter's evaluation value is obtained using the linear weighted average method. Finally, 11 DC-DC converters with a voltage gain of D^2 are selected as the example to carry out a comprehensive evaluation using the proposed method, and a feasibility analysis of the evaluation result is given. The result indicates that the selected topologies have good performance.

Keywords: DC-DC converter; comprehensive evaluation; index system; analytic hierarchy process; entropy method

Received：2023-05-30；**Accepted**：2023-11-09；**Published Online**：2023-11-23 17：44

URL：link.cnki.net/urlid/11.2625.V.20231122.1713.004

Foundation item：National Natural Science Foundation of China (52277177)

* **Corresponding author**. E-mail：epdyqiu@scut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0302

基于强化学习的单电感多端口变换器调制策略设计方法

白敬波，陈宇*，谢诗语，代新维
(华中科技大学 电气与电子工程学院，武汉 430074)

摘　　要：单电感多端口(SIMP)变换器具有多硅少磁的特性，在诸多领域有很好的应用潜力，但其开关模态多，调制策略设计复杂，目前的设计方法是人为挑选开关模态序列并进行模态分析，设计过程需要电力电子专业知识和经验。基于此，提出一种基于强化学习(RL)的单电感多端口变换器调制策略设计方法，使用神经网络(NN)生成调制策略，该方法将端口电压和变换器结构等已知条件作为神经网络输入，并采用一组简单的规则提供奖励用于训练神经网络，避免繁复的人工设计。通过强化学习，神经网络无需人为干预即可在试错中总结经验，生成不同运行工况下的最优调制策略。对一种单电感多端口变换器进行调制策略设计，并通过实验验证了所提方法的有效性。

关　键　词：单电感多端口变换器；开关模态；调制策略；强化学习；神经网络
中图分类号：TM131.3
文献标志码：A　　**文章编号**：1001-5965(2025)05-1480-10

单电感多端口(single-inductor multi-port, SIMP)变换器具有多硅少磁的特性，其包含一个电感作为储能元件，使用多个半导体(MOSFET或二极管)建立不同的电感电流环路，从而可以在多个端口(输入、输出、双向端口)之间传输功率。SIMP变换器体积小，便于集成，目前已广泛应用于直流微网、电动汽车、消费电子、LED照明[1-4]等需要连接多个端口的场合。

SIMP变换器的开关管数量多，开关模态多，控制方法多样。文献[5]提出了一种SIMP变换器的电流型控制方法，通过斜坡函数补偿系统稳定性；文献[6]提出一种纹波控制方法，通过峰值电流-峰值电压控制减少交叉影响；文献[7]提出一种非线性控制，通过精确反馈线性化提高变换器的抗扰能力；文献[8]提出一种解耦控制策略，通过对控制-输出传递函数进行解耦控制，提高暂态性能。尽管SIMP变换器的控制方法多样，但核心均为调制策略的设计。首先，分析不同开关管的通断状态，得到所有可选的开关模态；然后，从中挑选特定开关模态组成模态序列；最后，计算各模态占空比，使变换器可工作于所期望的工作模式(如多输出模式、多输入模式)[9-11]。当端口数量、开关管数量增多时，开关模态数量大幅增加，可用的模态序列选择多，设计也会变得更加复杂[12-14]。此外，即使在同一工作模式下，针对不同运行需求的调制策略也各不相同。例如，文献[15]设计指标是使电感电流纹波最小，文献[16]设计指标是使半导体元件损耗最少等。由此可见，SIMP变换器的调制策略设计高度依赖于设计人员的专业知识和设计经验，过程复杂繁琐。

SIMP变换器调制策略设计的本质是在庞大的解空间中寻找最优解。针对该问题，人工智能

收稿日期：2023-05-31；**录用日期**：2023-11-01；**网络出版时间**：2023-11-23 21:03
网络出版地址：link.cnki.net/urlid/11.2625.V.20231122.1659.003
基金项目：湖北省青年拔尖人才项目
*通信作者. E-mail：ayu03@163.com

引用格式：白敬波，陈宇，谢诗语，等. 基于强化学习的单电感多端口变换器调制策略设计方法[J]. 北京航空航天大学学报，2025，51(5)：1480-1489. BAI J B，CHEN Y，XIE S Y，et al. Design method for modulation strategy of a single-inductor multi-port converter based on reinforcement learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1480-1489 (in Chinese).

(artificial intelligence, AI)是一种有效的方法。目前, 已有很多学者利用 AI 在巨大解空间中寻找最优解。例如, 文献 [17] 将人工神经网络用于最小化双有源桥变换器的电流应力; 文献 [18] 建立了耦合电感交错 Boost 变换器的人工神经网络模型, 用于提高调制和控制效率; 文献 [19] 提出了一种基于机器学习的永磁同步直线电机多目标优化设计方法等。

在多种 AI 方法中, 强化学习(reinforcement learning, RL)允许在搜索空间中不断探索, 并在试错过程中不断学习经验, 最终高效寻找到最优解, 因此, 可明显减轻设计人员的设计负担。例如, 文献 [20] 利用强化学习生成非隔离的 SIMP 变换器; 为减小电磁干扰, 文献 [21] 将强化学习用于印制线路板过孔设计。基于相同的思路, 本文提出了一种基于强化学习的 SIMP 变换器调制策略设计方法。创新点如下:

1) 将调制策略设计看作是对变换器开关模态排列组合并分配占空比的过程, 使用深度强化学习高效搜索开关模态与占空比的组合, 从而自动生成不同电压工况下的最优调制策略, 避免复杂的人工设计, 提高设计效率。

2) 将电路规则转换为强化学习的奖励, 从而将调制策略设计问题转换为寻找最大奖励之和的强化学习问题, 智能体可以自动探索和学习出能够最大化奖励的调制策略。

3) 利用神经网络(neural network, NN)替代 Q-learning 中的 Q-table, 避免了 Q-table 庞大的状态空间存储需求, 节省了存储空间, 使得所提方法切实可行。

1　SIMP 变换器调制策略设计

1.1　典型 SIMP 变换器拓扑

为解释本文方法, 以非隔离单电感三端口(single-inductor triple-port, SITP)变换器[20] 为例进行讨论。如图 1(a) 所示, 其参数列于表 1 中。该 SITP 变换器包含 1 个电感、3 个开关管、2 个二极管, 可以实现输入端口 V_i、电池端口 V_b 和负载端口 V_o 之间的功率流动。通过恰当地控制该变换器中的 3 个开关管, 可以使 3 个端口工作在 Buck 或 Buck-Boost 模式, 端口电压限制少, 运行范围宽。对于该变换器, $V_b>V_o$ 是需要一直被满足的条件, 因此, 其可以工作在 3 种电压关系: $V_i>V_b>V_o$, $V_b>V_i>V_o$, $V_b>V_o>V_i$。

该 SITP 变换器共计 3 个开关, 由于每个开关均有导通和关断 2 种状态, 共有 2^3=8 种模态, 同时该变换器包含 3 个开关管均断开且电感电流为零的全开路模态, 图 1(b) 中给出了上述 9 种模态。

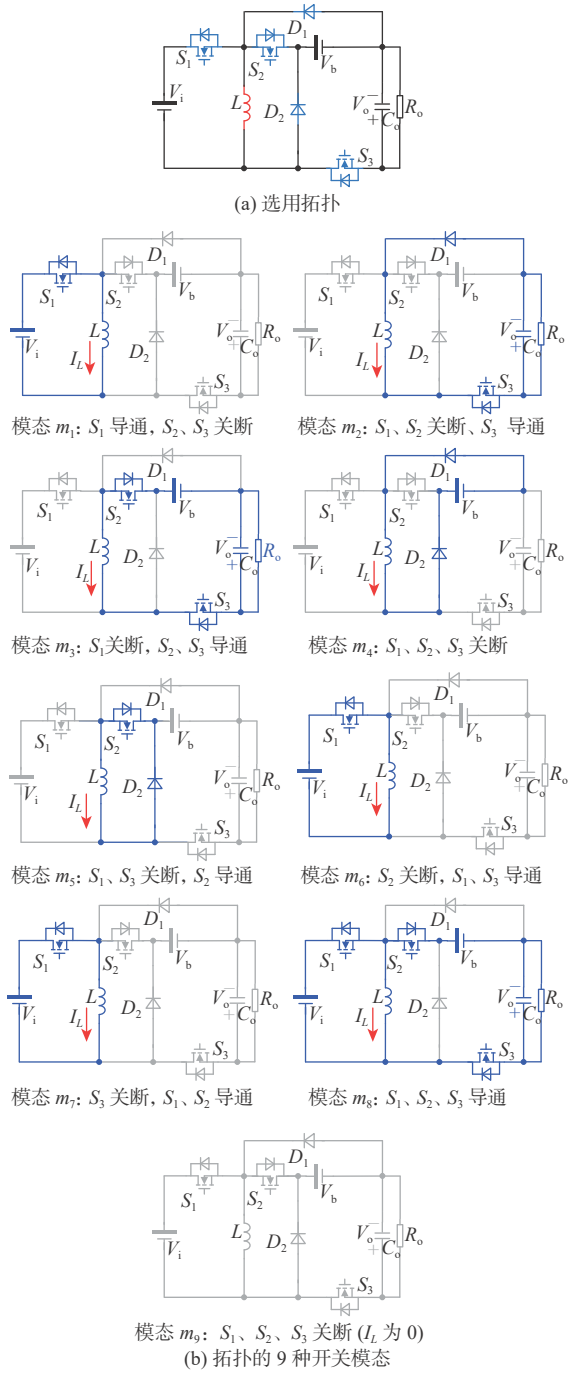


图 1　实际拓扑及其 9 种开关模态

Fig. 1　Actual topology and nine switching modes

需要注意的是, 并不是所有模态均为可用模态, 模态 m_6 中, S_1、S_3 导通, S_2 关断时, D_1 反向截止, 故与模态 m_1 相同; 模态 m_7 中, S_1、S_2 导通, S_3 关断时, D_2 反向截止, 故与模态 m_1 相同; 模态 m_8 中, S_1、S_2、S_3 均导通时, 电源端口短路, 模态不可用。此外, 本文以电感电流连续模式(continuous current mode, CCM)为前提, 故在分析中不考虑模态 m_9。下文设计中直接剔除了上述不可用模态, 以缩减强化学习的策略搜索空间。以最多不超过 6 个开关模态组合的序列为例, 使搜索空间从 8^6=262 144 缩

表1 样机关键参数

Table 1 Key parameters of prototype

元件	数值
光伏输入电压V_i/V	20～40(通常30)
电池电压V_b/V	30～40(通常36)
输出电压V_o/V	24
输出功率P_o/W	0～120
光伏功率P_i/W	0～240
开关频率f_s/kHz	20
电感L/μH	140
MOSFETs (S_1～S_3)	GS61004B
二极管(D_1～D_2)	NTSJ30U80
数字信号处理器(DSP)	TMS320F28335

减到5^6=15 625, 有助于强化学习效率的提升。

对于每个开关模态, 电感电流i_L流过不同的端口, 因此, 可将能量从一部分输送到另一部分。需要注意的是, 电感在每种开关模态下都会被磁化或退磁, 但在一个开关周期内必须令电感满足伏秒平衡条件。因为开关模态多, 模态和占空比组合选择多, 所以, 调制策略设计仍然十分复杂。为实现功率在3个端口之间流动, 有很多可供选择的调制策略, 如下:

$$v_1d_1+v_2d_2+v_5d_5=0 \tag{1a}$$

$$v_1d_1+v_2d_2+v_3d_3=0 \tag{1b}$$

$$v_1d_1+v_2d_2+v_5d_5+v_3d_3=0 \tag{1c}$$

$$v_1d_1+v_2d_2+v_3d_3+v_4d_4=0 \tag{1d}$$

式中: $v_i(i=1,2,\cdots,5)$表示选择图1(b)中模态m_i时电感两端的电压, 其值如表2所示; $d_i(i=1,2,\cdots,5)$为每个模态对应的占空比。

对比式(1a)和式(1b)发现, 不同工况需要使用不同的调制策略, 式(1a)适用于V_i-V_o工况, 式(1b)适用于单电感双输入(single-inductor dual-input, SIDI)工况。对比式(1b)和式(1c)发现, 相同SIDI工况下存在多种可选的调制策略。对比式(1a)和式(1d)发现, 续流模态(m_5)可以用来替换电感被磁化(m_3)和退磁(m_4)模态。

表2 不同开关模态下电感两端电压

Table 2 Voltage at both ends of inductor under different switching modes

开关模态	S_1	S_2	S_3	电感两端电压
m_1	1	0	0	$v_1=V_i$
m_2	0	0	1	$v_2=-V_o$
m_3	0	1	1	$v_3=V_b-V_o$
m_4	0	0	0	$v_4=-V_b$
m_5	0	1	0	$v_5=0$

注: "0"表示开关管关断, "1"表示开关管导通。

由此可见, 该变换器的模态序列多, 并且每个开关模态对应占空比有无穷解。在不同运行工况下, 需要从大量的调制策略中选择性能最优的, 人工选择费时费力, 且不具有通用性。

1.2 强化学习架构

强化学习的主要目标是找到一种策略, 可以最大化累积奖励。当适当的奖励分配给优化目标后, 最大化累积奖励即意味着找到的策略满足尽可能多的优化目标, 从而等效地实现优化。用于调制策略设计的强化学习架构如图2所示, 可将其视为环境与智能体交互的马尔可夫决策过程(Markov decision process, MDP)。智能体根据状态s_t给出动作a_t, 作用到环境, 环境产生奖励r_t并更新状态s_{t+1}, 将状态s_{t+1}和奖励r_t反馈至智能体, 智能体会从奖励r_t中学习经验。重复上述步骤, 最终生成状态与动作组合的序列, 如式(2)所示, 并得到总奖励R_{total}, 如式(3)所示。

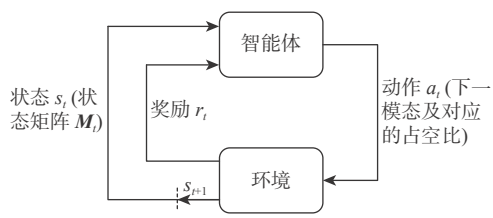


图2 强化学习架构

Fig. 2 Reinforcement learning framework

$$(s_0,a_0)\rightarrow(s_1,a_1)\rightarrow\cdots\rightarrow(s_t,a_t) \tag{2}$$

$$R_{\text{total}}=\gamma^{T-1}R^{(1)}+\gamma^{T-2}R^{(2)}+\cdots+\gamma^{T-k}R^{(k)}+\cdots+\gamma^0R^{(T)} \tag{3}$$

式中: R_{total}为停止一轮学习时累积的总奖励(找到满足要求的调制策略或模态数达到上限$N_{\max}$); $R^{(i)}$为第i步的奖励; T为一轮总步数; γ为折扣因子, 表示较早的动作对调制策略设计的贡献程度。

基于强化学习的特点, 将电路设为环境, 将电路需满足的约束设为奖励, 目标是利用强化学习高效找到可以最大化累积奖励(即R_{total})的策略。该策略不仅是一种可用的调制策略, 更是一种优化的调制策略(其可令累积奖励最大化, 必然满足尽可能多的设计约束)。

1.3 设计方法

用图2所示的强化学习架构实现调制策略设计。在每步中, 智能体的输入为状态矩阵$\boldsymbol{M}_t$, $\boldsymbol{M}_t$将不同工况下的调制策略(包括端口电压、模态序列及其占空比)进行量化, 如图3所示。假定模态序列中模态数量$N_{\max}$最大为8, 当端口电压为V_i、V_b、V_o时, 初始状态矩阵$\boldsymbol{M}_0$如图3(a)所示, 状态矩阵

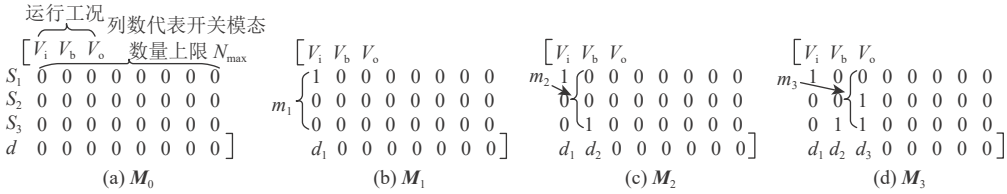


图 3　状态矩阵

Fig. 3　State matrix

第 1 行是 3 个端口的电压, 为智能体提供变换器工作于某种工况的依据, 第 2、3、4 行分别表示开关管 S_1、S_2、S_3 的通断状态, 每一列共同表征某步的模态, 第 5 行表示每步模态对应的占空比。除了端口电压, $\boldsymbol{M}_0$ 中其余元素均为 0, 表明此时还没有选择模态。将 $\boldsymbol{M}_0$ 送入智能体, 智能体选择动作 a_0, 即模态 m_1 和占空比 d_1, 状态矩阵更新为 $\boldsymbol{M}_1$, 如图 3(b) 所示, 同时根据规则生成奖励 r_0, 用来对智能体进行训练, 提高决策水平。同样地, 将 $\boldsymbol{M}_1$ 送入智能体, 智能体选择动作 a_1, 即模态 m_2 和占空比 d_2, 状态矩阵更新为 $\boldsymbol{M}_2$, 如图 3(c) 所示, 生成奖励 r_1。将 $\boldsymbol{M}_2$ 送入智能体, 智能体选择动作 a_2, 即模态 m_3 和占空比 d_3, 生成奖励 r_2, 同时状态矩阵更新为 $\boldsymbol{M}_3$, 如图 3(d) 所示, 最终得到与式 (1b) 相同的调制策略。在每一步中, 智能体先给出一个动作 (选择的开关模态及对应占空比), 得到奖励及下一个状态矩阵, 智能体根据奖励改进调制策略设计的过程。上述设计过程依次重复, 最终可以得到不同工况下奖励最大的调制策略。

2　智能体设计

2.1　神经网络和输出矩阵

智能体采用全连接神经网络, 根据当前状态矩阵 $\boldsymbol{M}_t$, 神经网络决定选择的动作 a_t, 即下一步的开关模态和占空比。如图 4 所示, 状态矩阵 $\boldsymbol{M}_t$ 被送到神经网络的输入层(表示为 1×m 的向量 $\boldsymbol{X}$, 其中, m 为状态矩阵中元素个数, 如图 3 中的状态矩阵会表示为 1×35 的向量), 通过神经网络, $\boldsymbol{X}$ 被映射到一个大小为 1×n 的向量 $\boldsymbol{Q}$:

$$\boldsymbol{Q} = f_\varphi(\boldsymbol{X}) \tag{4}$$

式中: $f_\varphi(\cdot)$ 表示神经网络, 其结构需要根据设计的复杂度合适地选择, φ 为神经网络中的可训练参数。

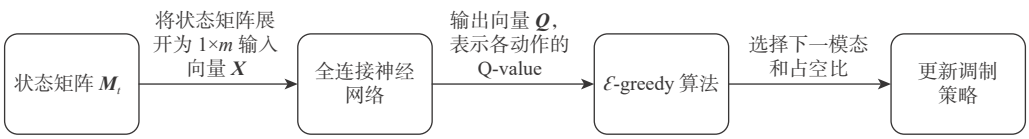


图 4　智能体选择开关模态及占空比的过程

Fig. 4　Process of selecting switch mode and duty cycle by agent

强化学习算法选用深度 Q 网络(deep Q network, DQN)算法, 其是一种 off-policy 时间差分算法[22], 该算法动作空间是离散的。为将调制策略设计问题转换为离散问题, 本文将占空比 0～1 等比例划分为 19 份, 动作空间为开关模态与离散占空比的所有组合。需要注意的是, 占空比离散化对控制没有不利影响, 一方面, 智能体通过离线学习生成调制策略, 实现了开环优化, 在真正电路实现时本身会存在偏差; 另一方面, 智能体生成的调制策略可使电路工作在稳定工作点附近, 在此基础上进行闭环修正, 通过闭环控制调节占空比, 可使电路稳定工作。

因为 SITP 变换器可供选择的开关模态为图 1(b) 所示的 5 种, 占空比被划分为 19 份, 所以, 动作空间大小为 n=95。假定生成包含 5 个模态的调制策略, 有 95^5 种可能性, 使用基于 Q-table 的经典强化学习方式时, Q-table 是非常庞大的, 因此, 选择用神经网络替代 Q-table。$\boldsymbol{Q}$=[q_i], i=1,2,⋯,n, 每个元素值表示“挑选某个开关模态和占空比组合”的动作质量, 称为 Q-value。Q-value 越大, 动作的质量就越高。当 N=19 时, 每份占空比为 0.05, q_5 对应于模态 m_5, 占空比为 0.05; q_{91} 对应于模态 m_1, 占空比为 0.95。假设 q_5=7, q_{91}=4, q_5>q_{91} 意味着选择“模态 m_5 和占空比 0.05”组合比选择“模态 m_1 和占空比 0.95”组合具有更好的质量, 更有利于生成最优调制策略。显然, 不同的神经网络参数 φ 导致 $\boldsymbol{X}$ 和 $\boldsymbol{Q}$ 之间的关系不同, 必然存在一个合适的 φ, 使神经网络可以在输入 $\boldsymbol{X}$ 时输出合适的 $\boldsymbol{Q}$, 并按照最大 Q-value 选择下一步最优开关模态和占空比组合。

2.2　ε-greedy 算法

神经网络的任务是给出动作, 即下一步使用的开关模态和占空比组合。好的设计策略意味着, 通过观察当前的 $\boldsymbol{M}_t$(或 $\boldsymbol{X}$, 因为 $\boldsymbol{X}$ 是从 $\boldsymbol{M}_t$ 展开的), 神

经网络可以给出合适的Q-value来指导下一步。但是，如果“挑选开关模态和占空比组合”的动作总是以最高Q-value为导向，则缺乏随机性，无法完全实现有效探索，在迭代的早期阶段更是如此。因此，本文应用ε-greedy算法，即以一定概率ε，根据最高Q-value选择开关模态和占空比的组合，以概率$1-\varepsilon$随机选择。在学习阶段的早期，神经网络还没有学到好的策略，因此，更随机地选择开关模态和占空比组合来探索更多的可能性；随着学习迭代次数的增加，神经网络可以学会一个足够好的策略，此时，挑选模态和占空比的动作可以更多地依赖于最高的Q-value。

当得到下一开关模态及占空比组合后，会判断其是否满足约束条件，从而获得奖励。

3 奖励设计

奖励设计原则是将电路的设计约束转化为奖励值，其分为必须满足的硬约束和满足后性能会更好的软约束。这一原则将寻找满足所有设计约束的策略问题等价转化为寻找累积奖励最大化的策略问题，从而可以在强化学习框架下进行求解。

3.1 硬奖励

为保证生成的调制策略是可行的，必须满足一些基本约束，这些约束通常关系到变换器能否稳定运行；此类约束统称为硬约束，对应的奖励为硬奖励，由“if…then…”规则给出。奖励分为负奖励和正奖励2种。

3.1.1 负奖励

负奖励用来训练在神经网络之后遇到相同输入$\boldsymbol{X}$时减少对应动作的可能性。

1）开关周期内模态数量达到上限$N_{\max}$。

2）特定端口的功率超过额定功率值。

3）占空比之和大于1，与实际条件矛盾。

4）如果存在2个连续相同模态，是多余的。

在上述任何一种情况下，均会给出一个很大的负奖励R_{N}，同时本轮学习停止，开始下一轮。

3.1.2 正奖励

正奖励用来训练在神经网络之后遇到相同输入$\boldsymbol{X}$时增加对应动作的可能性。

如果生成的调制策略可以使电感L满足伏秒平衡，如式(1a)～式(1d)所示，并且特定端口的功率值与额定一致，则该调制策略是可行的，因此，给定一个很大的正奖励R_{P}，同时停止本轮学习。

3.2 软奖励

硬约束可以保证SIMP变换器稳定可靠运行，但在实际应用中，SIMP变换器的运行效率、使用寿命等也是需要考虑的指标，此类指标在算法中表示为软约束，对应的奖励为软奖励。需要注意的是，软奖励并不需要精确量化各个指标值，只需定性表征优劣即可。

1）开关模态数。如果在一个动作之后没有得到负奖励R_{N}，但也没有找到可用的调制策略，则给定一个小的负奖励R_{SN}来量化这个动作。这样的奖励R_{SN}鼓励神经网络生成具有较少开关模态的调制策略，从而可以使开关损耗最小。

2）电感电流纹波。在变换器运行过程中，希望电感电流纹波尽可能小。如果一个动作之后，电感由被磁化转换为退磁（或者由退磁转换为被磁化），则认为对减少电流纹波有利，会给定一个正奖励R_{SP}。

如1.2节所示，强化学习的学习机制会学习到一种使累积奖励（即式(3)）最大化的策略，在该策略下，如下条件必然同时满足：

1）电路必定可用（获得正奖励R_{P}）。

2）不违背任何设计的硬约束（否则将得到负奖励R_{N}，累积奖励不可能为最大值）。

3）序列中的模态数最少（对应开关耗损最小。序列中每增加一个模态，则会得到一个小的负奖励R_{SN}，因此，使用的模态越多，累积奖励也会越小），或电感在被磁化和退磁2种状态下交替变化的次数最多（得到的小的正奖励R_{SP}最多，对应电流纹波最小），或两者折中（即2项奖励之和最大）。

根据奖励，可以通过DQN算法调整φ，对神经网络进行训练。

4 DQN算法训练神经网络

DQN算法的训练目标是使式(3)中的R_{total}尽可能高，其通用公式为

$$Q_{\mathrm{g}}(s,a)=R_{\mathrm{g}}(s,a)+\gamma\cdot\arg\max_{a'}Q_{\mathrm{g}}(s',a') \tag{5}$$

式中：s表示当前步骤的状态；a表示动作；s'表示动作a完成后的下一步状态；$Q_{\mathrm{g}}(s,a)$表示状态为s时用于评估动作a的Q-value；$R_{\mathrm{g}}(s,a)$为动作a完成后获得的即时奖励；$\arg\max_{a'}Q_{\mathrm{g}}(s',a')$为下一个状态$s'$的最大估计Q-value；$\gamma$为折扣因子。

式(5)表明，$Q_{\mathrm{g}}(s,a)$可以根据2个因素进行更新：第1个是采取动作a后的即时奖励$R(s,a)$，第2个是由a引起的下一个状态s'对应的最高Q-value，即经验。折扣因子γ决定了使用多少经验来更新$Q_{\mathrm{g}}(s,a)$。使用式(5)不断更新迭代，可以生成R_{total}最大的调制策略。强化学习的设计经验表明，奖励值$R_{\mathrm{g}}(s,a)$和折扣因子γ的选择有一定的规律性。例如，一旦所选的模态违背了“硬约束”，则应赋予相对较大的负奖励R_{N}，从而使智能体在下一轮

学习中选择该模态的概率变小；较大的折扣因子 γ 可使奖励传递到更早的步数中，从而便智能体重视以往经验，学会提前规划。但是，奖励值和折扣因子的具体取值需要在实际中进行调整。

在设计中，s 为当前状态矩阵 $\boldsymbol{X}$，a 为“挑选某个开关模态和占空比组合”的动作，$R_g(s,a)$ 为每步的奖励，在奖励设计部分具体给出，s'为 $\boldsymbol{X}$ 新增下一步开关模态及占空比后的状态，即 $\boldsymbol{X}'$，$\arg\max_{a'} Q_g(s',a')$ 对应输入为 $\boldsymbol{X}'$时神经网络的最大输出。式 (5) 可以写为

$$q_k^* = R_x + \gamma \cdot \max f_\varphi(\boldsymbol{X}') \tag{6}$$

式中：R_x 为每步动作后得到的奖励；max(·) 表示选择最大元素；k 为神经网络输出的第 k 个元素，对应于挑选的开关模态和占空比组合。

由于真实的 Q-value 由神经网络给出，有必要调整神经网络参数 φ，使 q_k 可以接近 q_k^*。损失函数定义为

$$f_{\text{Loss}} = \left(q_k^* - q_k\right)^2 \tag{7}$$

通过计算参数 φ 对于 f_{Loss} 的梯度，可以通过以下算法调整 φ：

$$\varphi \leftarrow \varphi - \alpha \frac{\partial f_{\text{Loss}}}{\partial \varphi} \tag{8}$$

式中：α 为神经网络的学习率。

图 5 为调制策略设计系统的流程。可以看出：①$\boldsymbol{M}_t$ 变成 $\boldsymbol{X}$ 后输入到神经网络中；②根据输出 Q-value，ε-greedy 算法采取行动；③更新调制策略后，可对动作进行评估，并给出奖励 R_x；④根据式 (6) 更新 q^*；⑤通过神经网络拟合 q^*，即依据方程 (7) 和方程 (8) 调整神经网络参数。

上述算法可以通过 Tensorflow 等工具实现，本文不再赘述。经过训练后，f_{Loss} 变小，说明神经网络更新了设计策略。

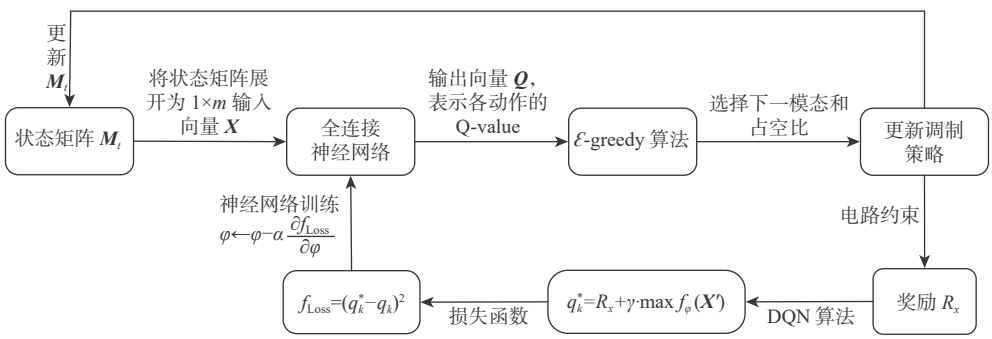


图 5　调制策略设计系统流程

Fig. 5　Flowchart of modulation strategy design system

5　应用实例与实验验证

5.1　SIMP 变换器调制策略设计实例

为验证本文提出的基于强化学习的调制策略设计方法，在 Python 3.7.3 中构建了调制策略设计系统，并使用 Tensorflow 1.14 来构建和训练神经网络。表 3 总结了设置的关键参数。需要注意的是，神经网络的层数和节点数是超参数，其选择需要一定的经验，需要在神经网络的表达能力、占用资源和运算时间之间进行折中。本文尝试表明，一个 3 层、节点数分别为 35-132-95 的全连接网络具有较好的表达能力（即对不同的输入，其输出可以较好地拟合相应的标签），从而在不同的工况下均能输出正确的结果。此外，该神经网络需要存储的权重为 17 445 个，在数字信号处理器（digital signal processor, DSP）进行一次运算需要约 270 ms，因此，训练好的神经网络可以内置于 SIMP 变换器的控制器中实时运行（但训练过程是在电脑上离线进行的，即所谓的“离线训练，在线推理”）。一旦计算完成，在 PWM 中断中将神经网络计算结果（即占空比对应的实际比较值）赋值给 PWM 比较模块，产生驱动

表 3　调制策略设计系统参数

Table 3　Parameters of modulation strategy design system

参数	数值
$\boldsymbol{X}$	1×35
神经网络输入层节点数	35
神经网络隐藏层节点数	132
隐藏层激活函数	ReLU
神经网络输出层节点数	95
训练算法	Gradient Descent
α	0.001
γ	0.9
$N_{\max}$	8
R_{SN}	–10
R_{N}	–300
R_{SP}	20
R_{P}	600

信号，该过程仅需 0.7 μs。因此，SIMP 变换器的控制逻辑如图 6 所示。训练时，在一轮中至少试探 5 步，意味着至少会有 5 种模态进行组合，从而保证电流纹波较小；但最多试探 8 步，若 8 步后未得到预期结果，则停止并给出负奖励，从而减少学习时间。

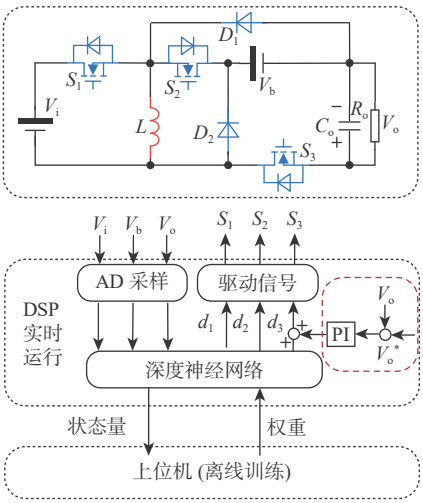


图 6 SIMP 变换器控制逻辑

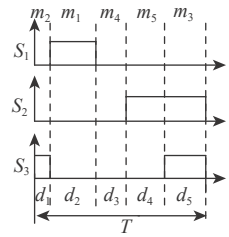
Fig. 6 SIMP converter control logic

针对图 1 所示 SITP 变换器，根据表 1 的样机端口电压变换范围，可以挑选不同的运行工况进行训练。经多次训练结果表明，在任一运行工况下，10 次训练中训练成功次数不低于 9 次，即调制策略设计的成功率约为 90%。本文方法是将 3 个端口的电压范围进行等距采样，获得不同的端口电压组合进行训练。训练完成的智能体(神经网络)可以针对不同的电压组合做出正确的决策。此外，得益于神经网络的拟合能力，只要实际端口电压处于采样区间内，则其输出结果也能基本保证正确。当 V_i = 20 V, V_b = 30 V, V_o = 24 V 时，自动生成的调制策略如图 7(a) 所示，当 V_i = 33 V, V_b = 37 V, V_o = 24 V 时，自动生成的调制策略如图 7(b) 所示。

为进行对比研究，找出了其他可以满足(或基本满足)设计要求的调制策略序列，如图 8 所示。通过对比可知，图 7(a) 和图 8(a) 均具有相对较小的电流纹波，其仿真效率分别为 95.2% 和 94.3%，这是因为：图 7(a) 中的开关模态数更少，因此，开关损耗更少；图 7(b) 和图 8(b) 均具有相同的开关模态数量，其仿真效率分别为 94.5% 和 94.1%，但图 7(b) 中电感状态变化次数更多，故电流纹波更小。综上，强化学习获得的调制策略为更优解。

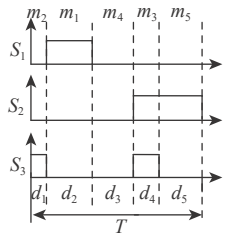
5.2 强化学习生成的调制策略仿真和实验验证

根据表 1 搭建了仿真模型，并搭建了一个 SITP 原型样机，用于验证利用本文提出的调制策略设计系统生成的不同工况下的最优调制策略。当端口电压 V_i = 20 V, V_b = 30 V, V_o = 24 V 时，使用图 7(a)



开关模态	S_1	S_2	S_3	占空比 d
m_2	0	0	1	0.075
m_1	1	0	0	0.275
m_4	0	0	0	0.175
m_5	0	1	0	0.217
m_3	0	1	1	0.258

(a) V_i=20 V, V_b=30 V, V_o=24 V

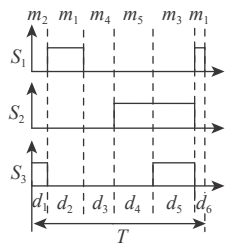


开关模态	S_1	S_2	S_3	占空比 d
m_2	0	0	1	0.075
m_1	1	0	0	0.275
m_4	0	0	0	0.25
m_3	0	1	1	0.152
m_5	0	1	0	0.248

(b) V_i=33 V, V_b=37 V, V_o=24 V

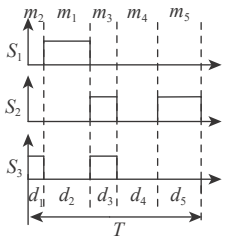
图 7 不同工况下最优调制策略

Fig. 7 The optimal modulation strategy under different working conditions



开关模态	S_1	S_2	S_3	占空比 d
m_2	0	0	1	0.075
m_1	1	0	0	0.225
m_4	0	0	0	0.175
m_5	0	1	0	0.217
m_3	0	1	1	0.258
m_1	1	0	0	0.05

(a) V_i=20 V, V_b=30 V, V_o=24 V



开关模态	S_1	S_2	S_3	占空比 d
m_2	0	0	1	0.075
m_1	1	0	0	0.275
m_3	0	1	1	0.25
m_4	0	0	0	0.152
m_5	0	1	0	0.248

(b) V_i=33 V, V_b=37 V, V_o=24 V

图 8 不同工况下 2 种人为设计的调制策略

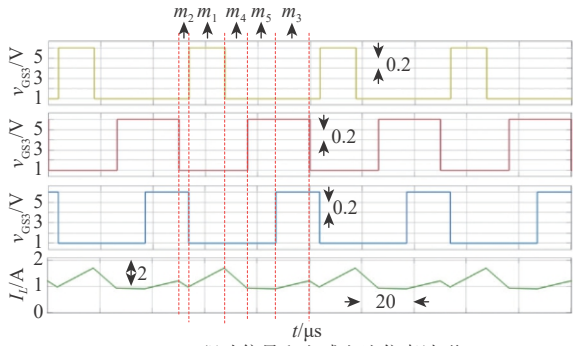
Fig. 8 Two artificially designed modulation strategy under different working conditions

所示的调制策略，开关管驱动信号和电感电流的仿真波形和实验波形如图 9(a) 所示。当端口电压 V_i = 33 V, V_b = 37 V, V_o = 24 V 时，使用图 7(b) 所示的调制策略，开关管驱动信号及电感电流的仿真波形和实验波形如图 9(b) 所示。电感电流实验波形和仿真波形基本一致，可以验证本文所提调制策略设计方法的可行性和可靠性。

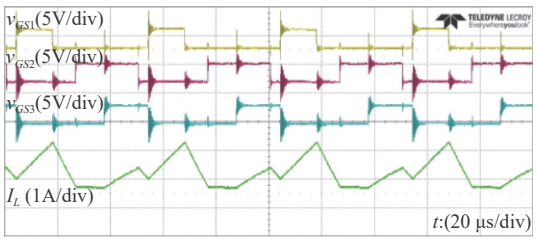
5.3 结果分析

由强化学习的设计过程和验证结果，本文对相关问题作简要讨论：

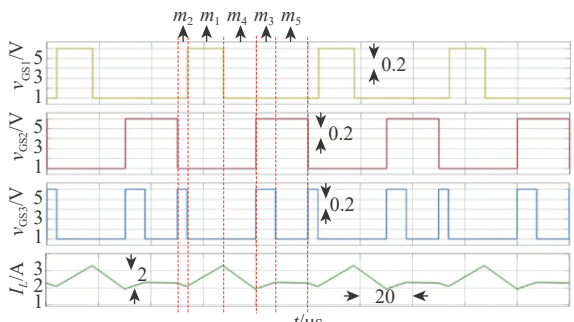
1) 本文以电感电流连续为前提，在理想电感电流连续情况下，端口电压增益仅与模态序列及其占



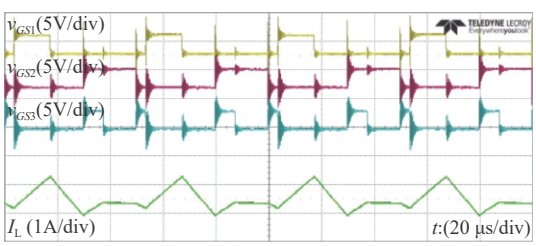
(a1) 驱动信号和电感电流仿真波形



(a2) 驱动信号和电感电流实验波形
(a) V_i=20 V, V_b=30 V, V_o=24 V



(b1) 驱动信号和电感电流仿真波形



(b2) 驱动信号和电感电流实验波形
(b) V_i=33 V, V_b=37 V, V_o=24 V

图 9　开关管驱动信号及电感电流的仿真和实验波形

Fig. 9　Simulation waveform and experimental waveform of switching drive signal and inductance current

空比有关，而与负载无关，因此，在神经网络训练过程中不考虑负载的影响，在相同的端口电压下，所生成的策略适用于所有负载情况。尽管以电感电流连续为例，但本文方法也适用于电感电流断续，此时应修改相关奖励设计。

2）在实际设计时，需要按端口电压和功率校核元件参数，这与一般变换器设计相同，故不作讨论。但需注意的是，在过载工况下，通常需要降低端口电压以限制输出电流（即限功率运行），此时需修改强化学习中设计约束的端口电压值，并将该工况加入训练过程中。需要注意的是，在需要更精细评估导通损耗和开关损耗的影响时，生成最优调制策略过程中应该考虑负载功率。

3）本文的设计将开关损耗大小与使用模态的多少定性关联。事实上，开关损耗还与开关时刻的电压、电流有关，可结合损耗设计公式或电路仿真结果给出更精确的奖励。

4）SIMP 变换器的端口电压变化通常十分缓慢（负载端口一般为定值；光伏电池端口随着光照的变化十分缓慢；电池端口的 SOC 状态也是缓慢变化的，故其端口电压值亦是缓慢变化的），而神经网络的计算时间远快于端口电压的变化。因此，当端口电压的大小关系发生改变时，神经网络可以迅速计算出新的调制策略，系统动态性能良好。但本文方法本质上是一种开环控制，在理想情况下，神经网络生成的最优调制策略用于变换器的实时控制，由于非理想因素的存在，实际输出电压必然与预设值有偏差。例如，在实际电路中，负载电流会在器件上产生压降，因此，电流增加会导致输出电压偏低。如果实际负载端口电压与理论值存在偏差，则可加入图 6 所示的 PI 闭环对 d_3 进行微调，此时无需调节 d_1 和 d_2 就能适应各种负载功率情况。当端口电压增益大幅变化时，神经网络会根据当前输入生成新的最优调制策略，PI 闭环仅在新的调制策略基础上进行微调，依然可以保证调制策略最优。

6　结　论

1）基于所提的强化学习方法，SITP 变换器在 3 种工作电压关系（V_i>V_b>V_o, V_b>V_i>V_o, V_b>V_o>V_i）下的调制策略都可以由强化学习算法自行设计。

2）所提方法生成给定运行工况下的最优调制策略只需要 5 min 左右，极大提高了设计效率。

3）对于某组特定设计约束，调制策略设计的成功率约为 90%，即 10 次学习过程中有 9 次可以生成最优调制策略。

4）由于神经网络强大的泛化能力，对于端口电压处于变化范围内但未被训练的工况，神经网络生成的调制策略亦能基本满足设计要求。

参考文献（References）

[1] KHASIM S R, DHANAMJAYULU C, MUYEEN S M. A single inductor multi-port power converter for electric vehicle applications [J]. IEEE Access, 2023, 11: 3367-3385.

[2] HUANG T W, TSENG W J, HUANG J X. Single-inductor multiple-outputs dimmable LED driver with Buck converter[C]//Proceedings of the International Power Electronics Conference. Piscataway: IEEE Press, 2018: 1626-1633.

[3] CHANG C K, SU C H, CHAO P C. A new single inductor bipolar multiple output (SIBMO) on-chip boost converter using ripple-based constant on-time control for LCD drivers[C]//Proceedings of the 40th Annual Conference of the IEEE Industrial Electronics Society. Piscataway: IEEE Press, 2014: 4122-4125.

[4] KIM J Y, LEE B S, LEE S H, et al. Integrated single inductor converter for a DC grid system connected with battery and load[J]. IEEE Transactions on Industrial Electronics, 2021, 68(12): 12010-12020.

[5] 周述晗, 周国华, 毛桂华, 等. 电流型控制单电感双输出开关变换器稳定性与瞬态特性分析[J]. 电工技术学报, 2018, 33(6): 1374-1381.
ZHOU S H, ZHOU G H, MAO G H, et al. Stability and transient response analysis of current-mode controlled single-inductor dual-output converter[J]. Transactions of China Electrotechnical Society, 2018, 33(6): 1374-1381(in Chinese).

[6] 徐利梅, 王瑶, 张留洋, 等. 用于单电感双输出 Buck 变换器的 PCPV 控制方案[J]. 西南交通大学学报, 2021, 56(1): 182-189.
XU L M, WANG Y, ZHANG L Y, et al. Peak-current and peak-voltage control scheme for single-inductor dual-output Buck converter[J]. Journal of Southwest Jiaotong University, 2021, 56(1): 182-189(in Chinese).

[7] WU J R, LU Y M. Exact feedback linearisation optimal control for single-inductor dual-output Boost converter[J]. IET Power Electronics, 2020, 13(11): 2293-2301.

[8] XU W W, LI Y, GONG X H, et al. A dual-mode single-inductor dual-output switching converter with small ripple[J]. IEEE Transactions on Power Electronics, 2010, 25(3): 614-623.

[9] CHEN G P, LIU Y W, QING X L, et al. Principle and topology derivation of single-inductor multi-input multi-output DC-DC converters[J]. IEEE Transactions on Industrial Electronics, 2021, 68(1): 25-36.

[10] YIN C D, LI H, LI Y M, et al. A new family of non-isolated single-inductor three-port converter based on a storage port switch-commutated unit[C]//Proceedings of the International Power Electronics Conference. Piscataway: IEEE Press, 2022: 2307-2311.

[11] QIAN Z J, ABDEL-RAHMAN O, AL-ATRASH H, et al. Modeling and control of three-port DC/DC converter interface for satellite applications[J]. IEEE Transactions on Power Electronics, 2010, 25(3): 637-649.

[12] BÖHLER J, KRISMER F, SEN T, et al. Optimized modulation of a four-port isolated DC-DC converter formed by integration of three dual active bridge converter stages[C]//Proceedings of the IEEE International Telecommunications Energy Conference. Piscataway: IEEE Press, 2018: 1-8.

[13] QIAN Z J, ABDEL-RAHMAN O, HU H B, et al. A zero-voltage switching four-port integrated DC/DC converter[C]//Proceedings of the IEEE International Telecommunications Energy Conference. Piscataway: IEEE Press, 2010: 1-8.

[14] MA Y H, HUANGFU Y G, XU L C, et al. A novel nonisolated multi-port bidirectional DC-DC converter with high voltage gain for fuel cell hybrid system[C]//Proceedings of the IEEE Transportation Electrification Conference & Expo. Piscataway: IEEE Press, 2021: 376-381.

[15] 王瑶. 电容电流–电容电压纹波控制单电感双输出 CCM Buck 变换器[J]. 中国电机工程学报, 2020, 40(10): 3280-3288.
WANG Y. Capacitor current and capacitor voltage ripple controlled single-inductor dual-output CCM Buck converter[J]. Proceedings of the CSEE, 2020, 40(10): 3280-3288(in Chinese).

[16] 谢锋, 皇金锋. 单电感双输出 Buck 变换器的供能模式及输出纹波电压分析[J]. 电工技术学报, 2020, 35(12): 2585-2595.
XIE F, HUANG J F. Supply modes and output ripple voltage of single-inductor dual-output Buck converter[J]. Transactions of China Electrotechnical Society, 2020, 35(12): 2585-2595(in Chinese).

[17] ZHAO S, BLAABJERG F, WANG H. An overview of artificial intelligence applications for power electronics[J]. IEEE Transactions on Power Electronics, 2021, 36(4): 4633-4658.

[18] WANG S D, DRAGICEVIC T, GONTIJO G F, et al. Machine learning emulation of model predictive control for modular multilevel converters[J]. IEEE Transactions on Industrial Electronics, 2021, 68(11): 11628-11634.

[19] SONG J C, DONG F, ZHAO J W, et al. An efficient multiobjective design optimization method for a PMSLM based on an extreme learning machine[J]. IEEE Transactions on Industrial Electronics, 2019, 66(2): 1001-1011.

[20] CHEN Y, BAI J B, KANG Y. A nonisolated single-inductor multiport DC-DC topology deduction method based on reinforcement learning[J]. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2022, 10(6): 6572-6585.

[21] GU Z M, ZHANG L, JIN H, et al. Deep reinforcement learning-based ground-via placement optimization for EMI mitigation[J]. IEEE Transactions on Electromagnetic Compatibility, 2023, 65(2): 564-573.

[22] SUTTON R S. Learning to predict by the methods of temporal differences[J]. Machine Learning, 1988, 3(1): 9-44.

Design method for modulation strategy of a single-inductor multi-port converter based on reinforcement learning

BAI Jingbo，CHEN Yu*，XIE Shiyu，DAI Xinwei

(School of Electrical and Electronic Engineering，Huazhong University of Science and Technology，Wuhan 430074，China)

Abstract：Single-inductor multi-port (SIMP) converters have great application potential due to their “more silicon less magnetic” feature. However, there are so many switching modes of SIMP converters that the modulation strategy design is complicated. At present, the design of the modulation strategy is mainly to manually select the switching mode sequence and carry out modal analysis, which requires professional knowledge of power electronics. This research proposes a reinforcement learning (RL) based modulation strategy design technique for SIMP converters. The proposed method uses a neural network (NN) to produce modulation strategies, takes known conditions such as port voltages and converter structure as the input of the NN and leverages a set of basic principles to provide rewards for training the NN, avoiding complex artificial design. With RL, the NN sums up experience during trial-and-error without human intervention, and finally generates the optimal modulation strategy under different operating conditions. Based on this method, modulation strategies are designed for a SIMP converter and the effectiveness of the method is verified by experiments.

Keywords：single-inductor multi-port converter; switching mode; modulation strategy; reinforcement learning; neural network

Received：2023-05-31；**Accepted**：2023-11-01；**Published Online**：2023-11-23 21：03

URL：link.cnki.net/urlid/11.2625.V.20231122.1659.003

Foundation item：The Young Top-notch Talent Cultivation Program of Hubei Province

* **Corresponding author**. E-mail：ayu03@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0296

锂离子电池故障预测及健康管理技术进展

杨世春[1,2]，王潇[1]，陈飞[1,*]，郑一帆[1]，钟益林[3]，周思达[1]

(1. 北京航空航天大学 交通科学与工程学院，北京 100191；2. 国家市场监管技术创新中心（新能源汽车数字监管技术及应用），北京 100120；3. 比亚迪汽车工程研究院，深圳 518118)

摘　　要：锂离子电池的故障预测及健康管理（PHM）技术通过对电池可靠性进行分析并综合制定控制策略，降低失效概率、提升失效前服役时间。相比于传统基于状态的电池管理方法，PHM更倾向于分析电池失效及概率，提高对电池可服役性的关注度，是未来电池管理的发展方向之一。锂离子电池PHM技术依赖于对电池失效、故障分析和预测，围绕PHM的关键技术，阐述了PHM技术在端云融合电池管理系统中的总体架构和思路，重点分析了故障分析及失效概率预测、基于故障的数字孪生建模技术、故障诊断及预警算法、可靠性分析及测试方法、可靠性增长技术研究进展，归纳了现有技术的优缺点及不足，并展望了未来PHM技术挑战和发展趋势，以期为相关研究提供参考。

关 键 词：锂离子电池；故障预测及健康管理；可靠性；数字孪生；故障诊断
中图分类号：U469
文献标志码：A　　**文章编号**：1001-5965(2025)05-1490-13

中国已成为全球最大的新能源汽车生产国，连续8年全球产销量第一。新能源汽车的快速发展带动了锂离子电池的需求飞速增长，但锂离子电池因自身材料问题、生产制造缺陷等因素常存在失效、起火问题，引发了国内外学者、用户对电池故障问题的关注。现有的锂离子电池管理方法包含了电池故障检测、热失控预警等功能，但线下依赖于车载电池管理系统（battery management system, BMS）的有限检测数据、云端依赖于历史数据深度分析，缺乏从理论到技术贯通的故障预测和管理方法。故障预测及健康管理（prognosis and health management, PHM）技术从电池失效角度出发，以电池可靠性为管理目标，提出了新的电池管理思路。锂离子电池PHM技术与云平台管理具有较好的适配性，在云端部署电池失效预测模型、建立全生命周期的性能仿真和故障预测模型，并实现可靠性增长，能够为电池管理提供新的系统性思路和方法。

本文围绕锂离子电池PHM技术中的主要关键点研究现状进行评述，归纳了现有技术的优缺点及不足，并展望了未来PHM技术挑战和发展趋势。

1 锂离子电池PHM技术

PHM技术是集故障检测、健康预测与评估及维护、决策于一身的综合技术，通过充分利用传感器采集系统的数据信息，借助于信息技术、人工智能推理算法来监控、管理与评估系统自身的健康状态，在系统发生故障之前对其故障进行预测，并结合现有的资源信息提供一系列的维护保障建议或决策。PHM技术在航天、航空、船舶等领域得到深入研究和广泛应用，并逐渐拓展至锂离子电池管理领域[1]。

传统锂离子电池管理方法以状态估计为主线，

收稿日期：2023-05-30；**录用日期**：2023-09-08；**网络出版时间**：2023-12-25 09:02
网络出版地址：link.cnki.net/urlid/11.2625.V.20231224.1612.001
基金项目：国家重点研发计划 (2021YFB2501700)
*通信作者. E-mail：cf2020@buaa.edu.cn

引用格式：杨世春，王潇，陈飞，等. 锂离子电池故障预测及健康管理技术进展 [J]. 北京航空航天大学学报，2025，51（5）：1490-1502.
YANG S C，WANG X，CHEN F，et al. Advances in prognosis and health management technologies of lithium-ion batteries [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1490-1502（in Chinese）.

依据试验测试制定电池使用边界，通过对荷电状态（state of charge, SOC）、健康状态（state of health, SOH）等进行估计，完成电池的使用管控。而 PHM 技术则以电池失效为主线，通过机理、模型、数据驱动方法对电池的失效模式、失效概率进行预测，建立电池可靠性衰退图谱，通过可靠性增长的方式来优化电池管理边界和目标，降低电池使用故障率。

2000—2023 年，国内外学者围绕锂离子电池 PHM 共发表 SCI 检索论文 157 篇，图 1 通过 VOSViewer 软件统计分析了各论文关键词相关性，依据各论文的关联性和关键词内容，总结出目前国内外围绕 PHM 方法研究内容集中在对于健康度的评价/估算、PHM 总体方法论及核心方法、健康度的主要影响因素 3 个方面。随着对电池 PHM 技术认识的加深与工程实践的探索，未来将会有更多工程应用导向的技术性文章与方法牵引导向的观点性文章产生。Khaleghi 等[2] 提出基于外源性输入的非线性自回归神经网络的电池寿命管理方法并应用至 PHM 中。Kong 等[3] 提出基于电压-温度健康特征提取方法的锂离子电池 PHM 方法。Lyu 等[4] 提出面向 PHM 的间接混合模型，其融合了灰狼优化算法（grey wolf optimizer, GWO）、多核相关向量机（multi-kernel relevance vector machine, MKRVM）、互补集合经验模式分解（complementary ensemble empirical mode decomposition, CEEMD）、长短期记忆神经网络（long short-term memory neural network, LSTM-NN）和前馈神经网络（feedforward neural network, FNN）多种算法来实现电池 SOH 和剩余寿命（remaining useful life, RUL）预测，Zhang 和 Li[5] 研究了深度神经网络在电池 PHM 的现状，并分析了其在电池 SOH 和 RUL 估计中的研究进展。上述研究仍主要着眼于电池寿命估计和预测算法[2,6]，研究的主要目的仍是用于提升电池使用寿命，缺乏对电池故障的考虑。图 2 展示了锂离子电池 PHM 架构，呈现了其主要组件、关键技术及核心功能。锂离子电池 SOH、RUL 及 PHM 相关性如图 3 所示。本文基于航空航天领域中的 PHM 方法，分析车用锂离子电池 PHM 中的关键技术，归纳总结近年来国内外学者对锂离子电池 PHM 技术的研究，分析未来技术发展方向及前景，以期为相关研究提供思路。

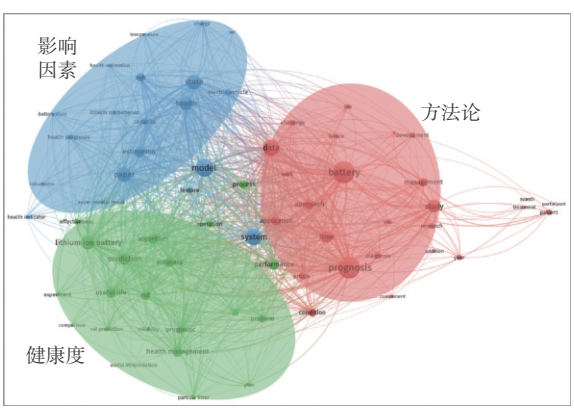


图 1　相关论文检索（2000—2023 年，以 Battery prognosis and health management 为关键词检索）

Fig. 1　Corresponding manuscript analysis (index by Battery prognosis and health management from 2000 to 2023)

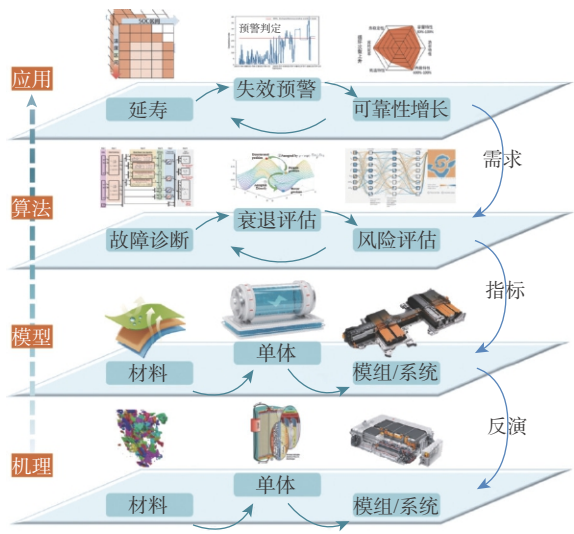


图 2　锂离子电池 PHM 架构

Fig. 2　Architecture of PHM for lithium ion battery

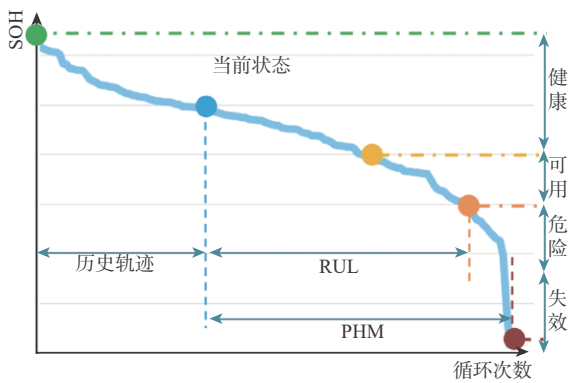


图 3　锂离子电池 SOH、RUL 及 PHM 相关性

Fig. 3　Correlation between SOH, RUL, and PHM of lithium-ion batteries

在建立锂离子电池 PHM 方法过程中，首先，需要考虑的是对于电池失效形式、诱因追溯及当前失效概率的计算评估方法；然后，基于分析得到的电池可靠性或失效概率，形成系统可靠性分析并开展实验测试；最后，将结果进一步应用至电池管理中，特别是故障诊断及预警算法。由于系统可靠性在实际使用过程中不断衰退，为了在使用时提升可靠性，需要考虑 PHM 中的可靠性增长等方法。特别地，数字孪生模型结合了端云融合的优势，能够通过数据和模型融合阐述电池内部参数演化情况及

宏观特征的衰退规律，为PHM提供较好的基础。因此，本文重点分析上述PHM中涉及到的关键技术，包括：故障分析及失效概率研究、数字孪生技术、可靠性分析及测试方法、故障诊断及预警算法，特别地，考虑可靠性领域在系统设计方面的理论，还需研究可靠性建模及增长方法。本文将围绕PHM关键技术的成熟度与现状进行分析和研究。

2 锂离子电池故障分析及失效概率预测

锂离子电池失效机理与故障树模型的构建是PHM“机理-模型-算法-应用”4层技术架构的基石，通过揭示电池在实际使用条件下的失效模式、故障诱因、表现形式等，支撑后续模型构建及算法开发。故障树是较为常用的失效致因追溯与正向推演方法，通过积累大量的数据、经验等对锂离子电池的失效模式进行解析，国内外围绕该领域已开展了大量研究，主要方法包括事故数据分析、实验测试、仿真模拟等，如表1所示[7-18]。

在事故数据分析方面，王芳等[19]基于新能源汽车火灾事故分析，从锂离子电池的关键材料、电芯设计、生产制造、系统集成和应用管控等环节系统梳理了潜在的引发失效的各种诱因，如图4所示。

表1 国内外失效机理研究进展

Table 1 Progress on research for failure mechanisms worldwide

研究项	作者	研究对象	研究内容	研究结论
材料级电池失效机理	Crowley等[7]	铝集流体	容量损失	扩散性锂捕集导致铝集流体容量损失
	Kalnaus等[8]	隔膜	机械损伤	研究了3种商业化隔膜的机械性能，在横向和纵向应力失效形式、敏感度上表现形式不同
	Sarkar等[9]	固态电解质氧化膜(solid electrolyte interphase, SEI膜)	容量损失	在高倍率充电条件下，较薄的SEI膜易发生分解和再生，该过程吸收了部分外电流，从而降低了电池容量和充放电效率
	Wang等[10]	电解液	容量损失	电池电解液泄露时发生更严重的自放电、容量衰减和电阻增长。泄漏的程度越大，自放电程度越严重
	Zhang[11]	负极	热失控	负极与电解液的反应在200℃到350℃之间存在2个热流峰，由锂化石墨和电解液之间的放热作用，以及残余的锂与阳极中的黏合剂发生反应
成品电池失效机理	Mao等[12]	圆柱形电池	热失控	当电池中心位置被穿透时，热失控更为严重。隔膜的收缩及正极和电解液之间的反应会引发电池热失控
	Spielbauer等[13]	圆柱形电池	热失控	在动态负荷下最主要的失效机制是电流截断装置的接触损失，增大了能量密度和安全性之间的矛盾
	Teng等[14]	圆柱形电池	容量损失	低NP比电池的失效是由于锂在负极上的沉积、SEI膜不断形成和破坏，导致电解质逐渐消耗
	Zhang等[15]	18650圆柱形电池	机械损伤	研究了圆柱形电池加载时失效模式，在低速条件下，力的增加是单调的，在中速和高速挤压下，力-位移曲线出现了"上升-下降-稳态-上升"的变化模式，该现象由应力波的传播、反射和叠加造成的
	Ren等[16]	方形电池	热失控	充电电流对电池过充的影响很小，但泄压设计(限制板和小袋上的切口)和良好散热会加剧电池过充，在热失控之前可以承受更大的过充容量和温度
	Sahraei等[17]	方形电池	机械损伤	锂离子电池的冲压试验的恒定单轴应变-失效可以作为宏观层面的有效标准来检测短路的发生
	Wang等[18]	软包电池	机械损伤	负极失效受界面黏合强度的影响，具有均匀粒径分布的活性负极材料在机械性能方面表现良好，并且机械性能与活性材料、集流体的厚度呈正相关

试验测试的失效解析研究是故障溯源和失效机理解析的核心，通过开展宏观层面上的热失控触发、扩展、蔓延测试[20]及电化学阻抗谱、超声波测试[21]和微观机理层面的绝热量热测试等方法，可分析电池材料层级的失效模式并建立数学描述。Hendricks等[22]基于锂离子电池可靠性试验测试及多年研发经验，总结了锂离子电池的故障模式、机制和影响分析(failure mode mechanism and effect analysis, FMEA)。Hao等[23]研究了基于声发射(acoustic emission, AE)和三维数字图像关联(3D-digital image correlation, 3D-DIC)原位实验平台研究了18650锂离子电池在弯曲载荷下不同荷电状态下的力学性能和失效机制，揭示了18650锂离子电池的弯曲模量和刚度与电池SOC呈正相关。王宏伟等[24]研究了锰酸锂电池在实际运行环境下的失效模式，模拟了暴晒、碰撞、追尾等场景下的电池失效模式。王

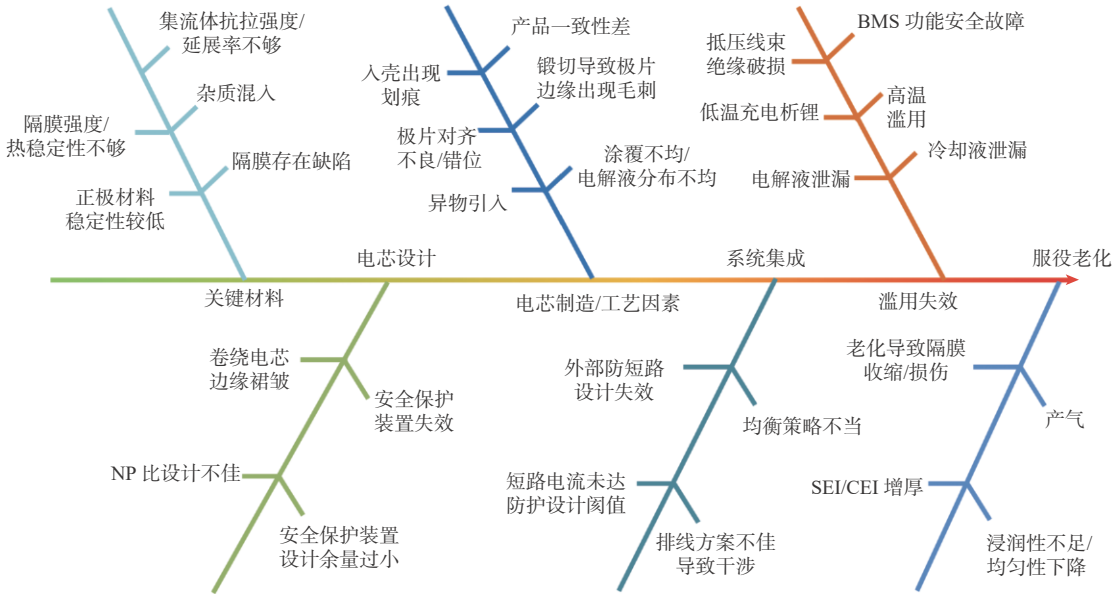


图 4　潜在的引发失效的各种诱因[19]

Fig. 4　Potential failure triggers[19]

震坡等[25]通过过充电试验测试，阐述了锂离子电池在过充热失控发生过程中不同温度点下的主要副反应机理，如图 5 所示。

在仿真模拟方面，Ren 等[26]建立了电池等效电路模型，分析了锂离子电池在极端冲击情况下的失效现象和机理，构建了锂离子电池在高冲击瞬间的机械冲击动力学（mechanical impact dynamics, MID）模型，揭示了隔膜厚度、弹性模量等参数对锂离子电池抗冲击性的影响。Zhang 等[27]采用碎片化分析方法，通过串联不同特征电压下的失效机制，推导出整个过充电热失控过程中的动态失效演化机制，对动态过充电失效机理进行了研究。兰凤崇等[28]研究了方形磷酸铁锂电池内芯的新型本构方程和电池的有限元模型，分析了对平

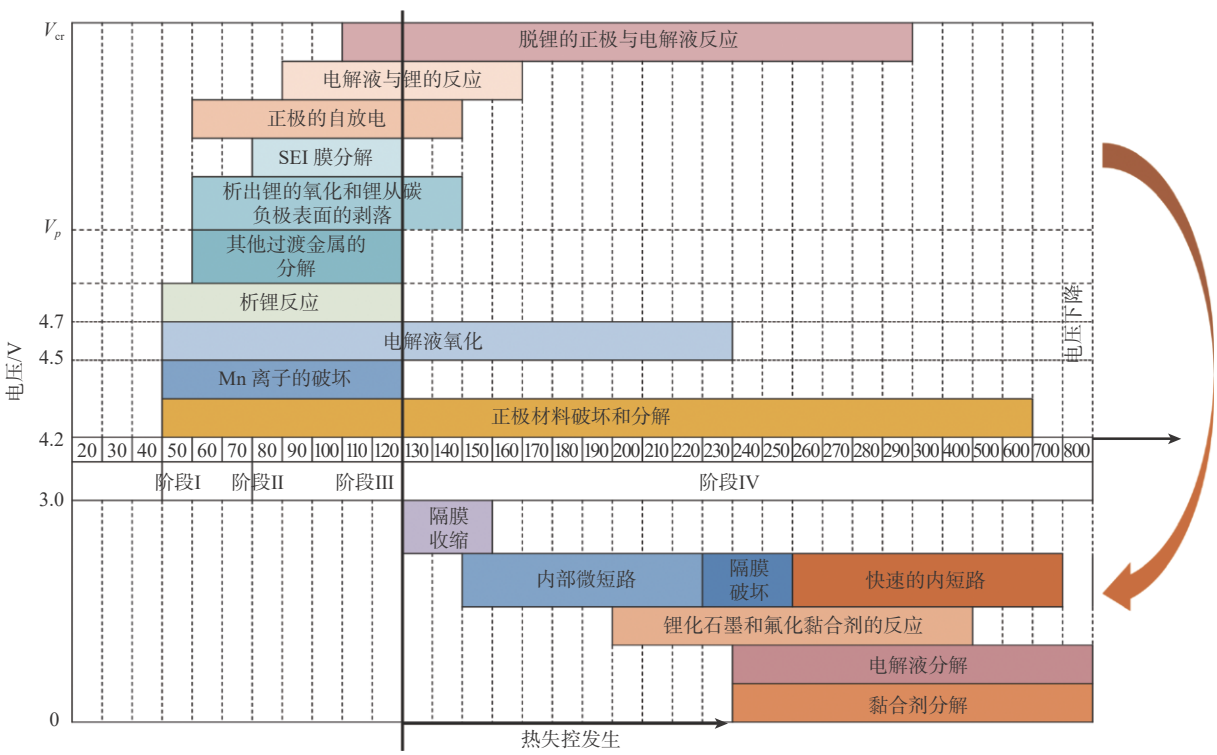


图 5　锂离子电池过充热失控发生过程中不同温度点下的主要副反应机理[25]

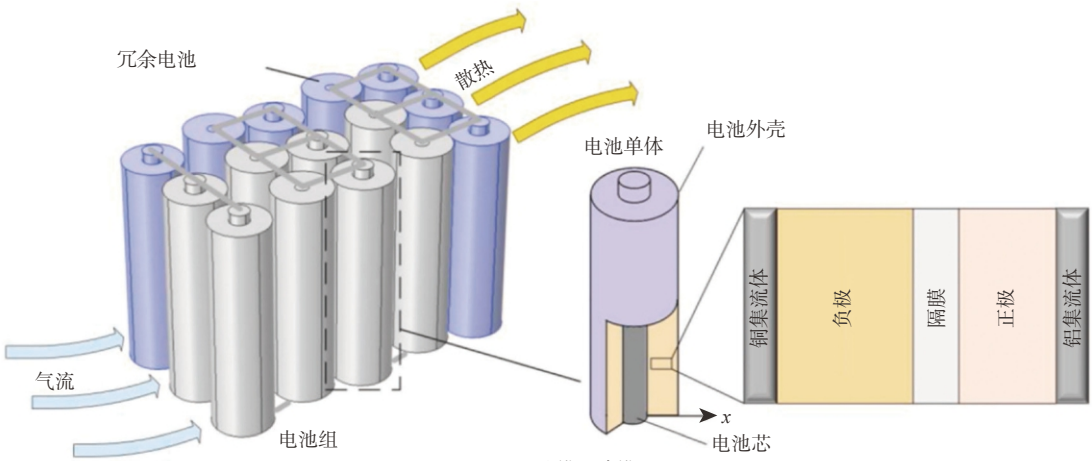
Fig. 5　The main side reaction mechanisms at different temperature points during the occurrence of thermal runaway during overcharging in lithium-ion batteries[25]

面挤压和局部压痕载荷下电池的变形响应和内短路失效。

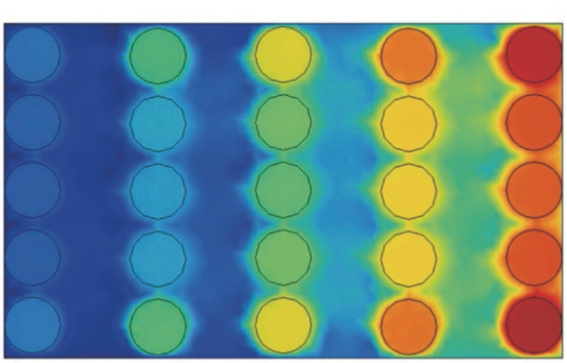
目前，国内外通过试验、仿真等方法，已围绕锂离子电池在颗粒层面、电极层面、单体层面的机械/热/电失效形式与内生机理开展了大量研究，已基本阐明了锂离子电池微观尺度上的失效机理和宏观上的失效模式。但现有研究主要围绕对单体电池、电池材料层面上的失效研究，对于电池成组后的机械、电气失效还缺乏系统性研究，需要和工程经验进一步结合开展深入的试验、仿真研究。

3 单体-系统可靠性分析及测试方法

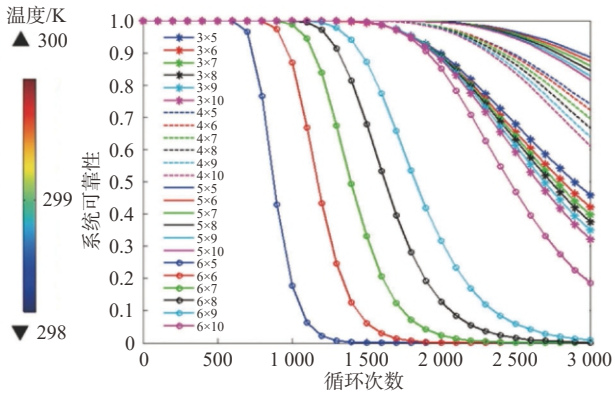
可靠性分析及测试方法是PHM技术的核心，也是以失效为主线的锂离子电池管理的主要体现。通过定义电池的失效行为和表现，量化描述锂离子电池在当前条件下及未来的可能失效概率。通过建立全生命周期的可靠性衰退图谱，可分析锂离子电池的剩余服役寿命及不发生故障的可能性，指导电池的各类使用边界的规划，延长电池服役时间，如图6所示。



(a) 电池模组建模



(b) 5×5排列下电池温度场预测



(c) 不同排列方式下系统寿命可靠性衰退图谱

图6 基于电化学模型的可靠性分析方法[29]

Fig. 6 Reliability analysis method based on electrochemical models[29]

目前，锂离子电池单体-系统的可靠性分析及测试方法主要分为模型预测和试验测试2个方面。模型预测主要用于实现未来工况下可靠性衰退图谱的预测，而试验测试则是从批量化产品失效实验的角度对极端条件下的可靠性进行评定。因此，对于端云融合的电池管理而言，模型预测可靠性图谱是云端部署的主要算法，而试验测试则为其提供了依据和数据支撑，表2总结了锂离子电池可靠性领域相关研究进展。

目前国内外通过建模、试验测试等方法[30]，在SOH估计、RUL预测和可靠性分析等方面开展了大量研究，能够有效对电池的寿命和可靠性进行评估。如Philippot等[31]对镍锰钴($Ni_xCo_yMn_{1-x-y}$, NMC)-钛酸锂氧化物(lithium titanium oxide, LTO)新电池进行测试以开发半经验NMC-LTO电池模型，对老化电芯进行其他测试以评估其二次使用的可行性，分析结果表明，该电池可重新用于固定存储长达408 000 km或10年。但现有研究主要依靠实验室的数据开展研究，工况较为固定，针对实车复杂多变工况下的寿命预测和可靠性分析研究较少，还需

表 2　国内外电池可靠性分析及测试

Table 2　Analysis and testing of battery reliability worldwide

研究项	作者	研究内容	研究方法	误差分析	研究结论
SOH估计模型	Meng等[32]	SOH估计	经验模型	<2%	提出一种考虑电池容量再生的锂离子电池容量精确预测的混合方法, 能有效降低容量再生对预测精度的影响
	Liu等[33]	SOH估计	数据驱动	<3%	提出一种使用健康因子和加权粒子群优化高斯过程回归模型预测不同温度条件下锂离子电池健康状态的方法, 该方法样本量小, 预测精度高, 适用性广
	Guo等[34]	SOH估计和RUL预测	数据驱动	<1%	提出一种Savitzky Golay(SG)滤波器与门控递归单元(gate recurrent unit, GRU)神经网络相结合的SOH和RUL预测模型, 验证结果表明该模型可在各类充电策略下给出准确的预测结果
	Jorge等[35]	SOH估计	数据驱动	<2%	提出一种从充电和放电过程中的电流、电压和温度曲线中提取特征来预测电池SOH演变的方法, 将循环的滑动窗口作为长短期记忆神经网络的输入, 输出SOH的多步预测
	Cheng等[36]	SOH估计	经验模型	<5%	进行加速退化测试来研究存储时间和温度对容量退化的影响, 结果表明, 老化电芯的剩余容量随着存储时间和温度的变化而呈指数级变化
	Henschel等[37]	SOH估计	经验模型		利用在实验室环境中收集的电池测试数据建立容量估计模型, 根据模型获取SOH估计值并与实际实车运行的数据进行比较
寿命预测模型	Ecker等[38]	RUL预测	经验模型	<9%	对代表混合动力电动汽车典型运行条件的一系列加速寿命实验进行多变量分析, 结果表明温度和电荷状态对阻抗上升和容量损失的影响是定量的
	Wang等[39]	RUL预测	经验模型	<5%	采用非破坏性的电化学方法来监测寿命测试期间的容量损失、电压下降、电阻增加、元素损失和活性材料损失
	Wu等[40]	RUL预测	数据驱动	<2%	提出一种基于监测数据的集成学习算法, 以拟合电池的老化模型并预测其RUL, 集成学习算法由5个基本学习器组成, 包括相关性向量机、随机森林、弹性网络、自回归模型和长短期记忆网络, 验证结果表明集成学习算法具有更好的鲁棒性和泛化效果
可靠性模型	Yang等[6]	热可靠性分析	机理模型		使用有限元分析软件建立了锂离子电池的电化学-热耦合模型, 采用自适应克里格法来建立热可靠性性能函数模型和评估锂离子电池的热可靠性
	Philippot等[31]	可靠性优化	机理模型		开发了一个电化学-热流体动力学模型、一个随温度变化的随机退化模型和一个多态性能可靠性模型, 对电池组的布局方案进行了优化
	Zhu等[41]	可靠性分析	经验模型		考虑锂离子电池系统不平衡电流和不均匀冷却的情况, 提出一种整合多物理模型和通用生成函数方法的改进可靠性评估方法
	Bai和Tu[42]	可靠性建模	经验模型		提出一种钛合金疲劳裂纹增长的一阶近似椭圆体非概率可靠性分析方法, 用于解决椭圆体模型的可靠性
	Gong和Frangopol[43]	可靠性建模	经验模型		提出受腐蚀影响的船体大梁可靠性的时间变化的表述, 采用高斯、甘贝尔、克莱顿和弗兰克协同学构建联合概率分布模型
	Heydari和Sullivan[44]	可靠性建模	算法模型		提出基于切割面算法的串联/串并联系统可靠性计算方法
	Byun等[45]	可靠性建模	算法模型		提出了K-out-of-N系统的基于矩阵的系统可靠性方法, 将同质和非同质的k-out-of-N系统组件故障之间的统计依赖性、参数敏感性、相对重要性纳入考虑

进一步扩展应用场景范围。

4　锂离子电池可靠性增长技术

从实际的锂离子电池设计、使用角度来说, 产品落地后的锂离子电池的预期可靠性水平通常难以达到预先设计的要求, 因此, 需要进一步开展可靠性增长的试验、测试、仿真等, 提升系统的可靠性, 排除系统失效原因或故障概率[46], 如图 7 所示。然而, 如何合理的设计可靠性试验测试、降低测试成本并满足可靠性指标需求, 是可靠性增长领域的重要研究内容。可靠性增长试验在航空、汽车、工业等领域是较为常用的方法之一[47]。Nadjafi 和 Gholami[48] 提出一种以硬件为中心、以非均质泊松过程(non-homogeneous Poission process, NHPP)为基础的正态分布系统可靠性增长方法, 并用于某个具有故障数据的航空系统。Huang 等[49] 以非均质泊松过程为基础提出软件可靠性增长模型, 通过考虑不充分的调试、错误和测试变化点, 扩大软件可靠性的实用性。Zou 和 Kolios[50] 提出基于贝叶斯决策分析的信息价值计算方法, 并指出缺陷带来的信息价值不仅有助于降低维护成本, 还有助于降低风险和增长可靠性。

在可靠性增长技术方面, 包括软件可靠性增长和产品可靠性增长研究。前者主要在软件复杂度不断提升时可以更好的检测到软件内部缺陷、部署

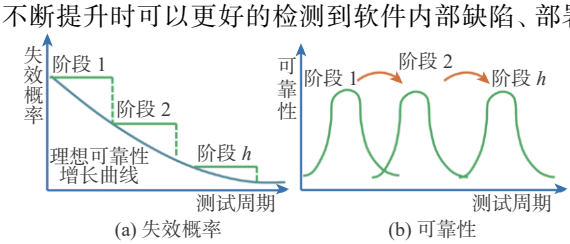


图 7　可靠性增长方法示意图

Fig. 7　Schematic diagram of reliability growth method

环境缺陷等，可用于指导电池管理软件算法设计和测试，而后者则重点关注产品本身的测试项和测试计划，以产品可靠性约束为指标、通过最小化产品成本来合理规划产品的测试计划。对于锂离子电池系统而言，可以通过规划材料-单体-模组/系统不同层级的实验测试内容来提升并达到预期的可靠性目标。但可靠性增长方法在锂离子电池领域尚未展开，未来应围绕电池可靠性增长技术重点开展研究，参考其在航空、航天、船舶等领域的可靠性增长技术实现形式与方法，以提升其在电动汽车、储能电池领域的应用效能。

5 锂离子电池数字孪生技术

锂离子电池数字孪生技术是在现有建模方法基础上发展起来的新的建模思维，在云计算平台的支撑下能够使云端的数字系统随实际物理实体运行而发展进化，使模型无限逼近物理实体并与实体“同时空”的演化，从而实现对实体电池的数字化仿真、监控、预测、优化等。数字孪生模型依据多物理场仿真建模理论，能够模拟电池内部的异常并反应出表征形式，是PHM技术中的关键部分，用于结合失效机理建立失效模型，并反应出失效特征和形式，支撑故障诊断算法的设计。

早期的数字孪生技术起源于航天电子类应用。2010年，美国国家航空航天局 (National Aeronautics and Space Administration, NASA) 在其太空技术路线图中首次引入了数字孪生的表述，有着明确的工程背景，即服务于未来宇航任务的需要。随着对人工智能、大数据引擎算法的深入与突破，数字孪生技术迎来了新的发展契机，通过数据驱动方法建立实体-虚拟模型联系、利用大数据方法辨识模型随实体演化的全生命周期参数，从而实现基于模型预测仿真的电池故障概率、性能衰退可靠性预测等。近年来，随着大数据方法逐步推广应用至锂离子电池领域，数字孪生电池也随之发展而实用。在前期研究中，提出数字孪生模型的3大基本特征：全生命周期建模、对物理实体的逼近、全场景模拟和仿真，如图8所示。

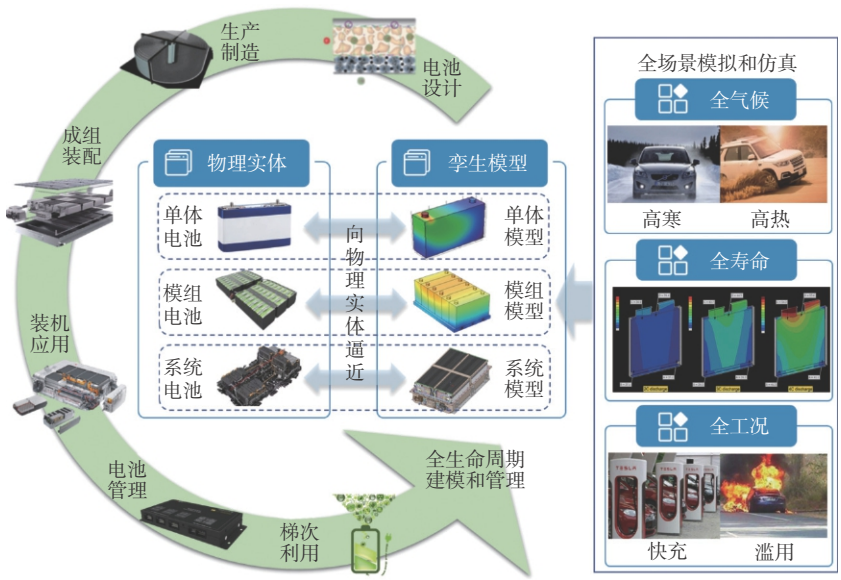


图8 贯穿电池全生命周期的数字孪生模型

Fig. 8 Digital twin model across the battery lifespan

锂离子电池数字孪生模型是指利用仿真技术创建的锂离子电池的虚拟副本，能够反映锂离子电池的内部状态和外部行为，以实现对锂离子电池的监测、诊断和预测。锂离子电池数字孪生模型的建模方法主要被分为基于物理的建模方法、基于数据的建模方法和混合模型的建模方法。基于物理的建模方法利用电化学、热学、力学等理论建立锂离子电池的数学模型，能够较准确地描述锂离子电池的内部机理和外部特性，但是计算量大，参数辨识困难，适用于离线仿真和分析。基于数据的建模方法利用机器学习[51-53]、神经网络[54-55]、回归分析[53]等技术建立锂离子电池的统计模型，能够较快速地拟合锂离子电池的输入输出关系，但是缺乏物理意义，泛化能力差，适用于在线监控和控制。基于混合的建模方法。这种方法结合物理模型和数据模型的优势，采用降阶模型、数字孪生工具、云端部署等技术建立锂离子电池的混合模型，能够在保证精度的同时提高效率，适用于数字孪生应用。数字孪生模型相关研究进展归纳如表3所示。

数字孪生有助于提高对电池性能的控制，利用

表 3　锂离子电池全链条数字孪生建模方法及相关应用的研究进展

Table 3　Research progress on the lifespan digital twin modeling of lithium-ion batteries and related applications

研究项	作者	研究对象	研究内容
数字孪生模型	Ngandjong等[56]	电池生产制造	提出面向电池生产制造过程浆料平衡及其干燥等工艺的建模, 可用于电池全链条数字孪生模型构建
	Husseini 等[57]	电极材料	通过数字孪生基于模型的方法, 来确定电极材料张力控制的最佳方法
	Wang等[58]	电极材料	将基于不对称温度调制的材料不可知方法与热稳定的双盐电解质相结合, 研究这一材料的安全性
	Bilansky等[59]	单体电池	在MATLAB/Simulink中实现电池单体数字孪生模型, 经验SOH模型基于从NR18650-25R电池的SOH≥80% 的电池循环获得的数据。包含了关于整个生命周期的信息
	He等[60]	电池系统结构	建立了一个电池系统架的数字孪生模型-DTMAR。包括数值模型、NN-RSR模型和混合机器学习模型。结果显示, DTMAR模型可以高效、准确地计算和预测结构性能
	Reniers和Howey[61]	电网系统	模拟了一个1 MWh的电网电池系统, 每个电池由单独的电化学模型表示。对单体间变异性、热效应和降解效应的影响进行了长达10 000个周期和10年的模拟
	Pan等[62]	电网系统	提出能源微电网能量分配的数字孪生模型, 学习不确定的环境并根据他们的经验选择最佳的能源管理措施
	Söderäng等[63]	电网系统	建立了数字孪生快速运行模型, 对基于物理学的内燃机模型与包括电池存储在内的发电厂模型进行实时协同模拟
	Tang等[64]	电池系统	提出一个基于数字孪生技术的电池系统, 但整个系统是从物理空间到孪生空间, 但还没有反馈到物理空间
数字孪生应用	Li等[65]	数字孪生应用	提出基于数字孪生模型的电池SOC/SOH联合估计模型, 可监测电池老化过程中的容量衰减和功率衰减
	Semeraro等[66]	数字孪生应用	利用形式化概念分析方法分析了数字孪生的特点, 分析电池系统中数字孪生体功能和架构的趋势和差距
	Semeraro等[67]	数字孪生应用	分析了数字孪生技术在储能领域的应用, 同时评估了应用背景、生命周期阶段、数字孪生功能和数字孪生架构

光谱、物理和化学方法结合, 可以实现数字孪生对于真实物理数据的高度仿真[68]。然而, 数字孪生的高精度和实际实施过程中车载算力的矛盾逐渐凸显, 为在现实世界实现, 通过低成本和便携式的传感技术, 数字孪生电池系统中可以收集到电流、电压和温度。虽然这些数据具有自身的局限性, 但通过适当的数据后处理, 也可以推断出丰富的电化学内涵[69]。除这些指标外, 压力和变形也可以增加车端数字孪生模型的丰富程度。随着获取数据的来源更加丰富和车载电池的工况不断变化, 混合方法可以建立更加符合实际工况条件下的各类电池的数字孪生模型, 如图 8 所示[54,70]。

6　在役电池故障诊断及预警算法

故障诊断及预警算法用于决策锂离子电池的失效判定。可靠性分析仅能为电池提供潜在失效的可能性, 但对于实际运行的锂离子电池而言, 还需要进一步结合诊断及预警算法实现锂离子电池的失效判定。然而, 电池起火速度快、数据传输频率低且内容不全面、起火事故案例少, 导致热失控事故数据可用信息匮乏、可用数据有限, 为在役电池故障诊断及热失控预警算法研发增大了难度。目前, 国内外围绕锂离子电池故障诊断及预警算法已开展了大量研究, 特别是基于云端数据的故障诊断、热失控预警等, 可为综合评估锂离子电池的健康度和可靠性提供技术支撑。

6.1　基于云端数据的故障诊断

在基于云端数据的故障诊断方面, 常见的故障诊断方法包括基于知识、统计、机理这 3 类方法。基于知识的故障诊断方法研究起步早、应用广泛。这类方法不需要借助复杂的数学模型, 通过在概念和处理方法上的知识化, 进行故障推理判断。刘晓俊[71]基于电池常见故障类型结合专家经验, 提出一个基于对象和关系数据库的知识库建立方法, 并基于人工智能模糊推理的综合判断推理机理模型。夏淑英[72]将模糊数学理论与故障树分析方法相结合, 形成基于模糊故障树分析法的汽车故障诊断方法, 建造正确合理的故障树和知识图谱是诊断的核心与关键。Zhu 等[73] 利用北京市电动汽车运行监控系统, 根据三元锂电池充电试验提出基于 3 个阈值规则的过充故障判断。赵洋[74] 利用知识和经验方法, 形成基于远程监控的电池系统故障树, 构建了面向锂离子电池单体电压离群现象的故障诊断方法。

基于统计的故障诊断方法通过对云端电池数据进行特征提取和建模[75], 寻找电池状态和故障特征之间的映射关系与统计规律。刘鹏等[76] 针对故障与正常统计数据, 提出一种基于快速傅里叶变换 (fast Fourier transform, FFT) 和异常系数评估 (anomaly coefficient evaluation, ACE) 的锂离子电池故障统计与诊断方法。Wang 等[77] 使用威布尔参数描述了电压分布的统计特性, 并使用特征电压曲线的数值分析进行了研究, 以理解锂离子电池放电电压的不一

致性。Ma等[78]提出一种基于统计分析的电动汽车锂离子电池串联故障检测，根据统计出的级别故障，提醒用户采取适当的措施。Xue等[79]根据电池电流和速度，阐明电动汽车电池的运行状态，并基于高斯与统计分布确定诊断系数，以突出每个状态下的参数变化。

在机理故障诊断方面，赵帅[80]针对锂离子电池内短路机理与诊断方法开展研究工作。温涛[81]结合锂离子电池内/外短路故障具有相近等效电路的特点，根据短路故障具有的参数效应、耗尽效应以及绝热实验结果，分析短路故障对锂离子电池电热特性的影响。武超[82]研究锂离子电池典型故障机理及诊断方法，建立磷酸铁锂电池的综合故障诊断策略，实现典型故障模式的在线诊断、预警和机理辨析。Hu等[83]对各种故障的机理、特征和诊断进行了全面综述，包括电池内部故障、传感器故障和执行器故障，介绍并讨论了用于更安全的电池系统的故障诊断技术的未来发展趋势。Zhang等[84]从锂离子电池的老化机理和故障原因出发，从宏观层面总结了机理故障诊断的一般方法。Love等[85]通过2个不同的放电截止阈值(单个电池截止和电池组电压截止)来识别0℃下4S电池组内的温度引起的内部机理特性故障。

6.2 热失控预警

在热失控预警方面，朱晓庆等[25]较为具体地阐明了热失控触发条件和发生机理，比较全面地总结了提高锂离子电池系统安全性的方法。陈佳纯[86]以真实的电动汽车工作数据为对象，从机械触发、电触发、热触发3方面对故障扩散途径和触发热失控的单体的位置进行分析。廉玉波等[87]基于混合高斯-隐马尔可夫模型实现对锂离子电池实时热失控检测。Liu等[88]通过测量18650 NCM锂离子电池的放热速率、表面温度、火焰温度、正负极温度和质量损失，分析了不同电流速率下锂离子电池热失控特性。Wang等[89]建立了锂离子电池的电化学热耦合模型，以及热失控的热模型，在车载条件、高速率循环充放电条件和热滥用条件下对锂离子电池进行了研究和预警。Meng等[90]提出一种集成故障树、动态贝叶斯网络和支持向量回归的动态风险预测方法，用于锂电池热失控的风险预警。

7 结 论

锂离子电池PHM技术从电池失效出发，研究了锂离子电池可靠性管理方法，通过可靠性建模、可靠性增长等方式，降低电池使用过程中故障率，及早发现潜在故障可能性。本文围绕PHM关键技术国内外研究现状进行了分析：

1）在电池失效模式及概率预测方面，国内外已通过试验、仿真等方法在颗粒层面、电极层面、单体层面的机械/热/电失效形式与内生机理开展了大量研究，基本阐明了锂离子电池微观尺度上的失效机理和宏观上的失效模式。但现有研究主要集中在对单体电池、电池材料层级的机理性分析，缺乏与工程实际相结合的系统性机械、电气失效分析。此外，由于现有锂离子电池系统失效数据、案例不足，对于电池高层级、耦合性失效研究仍不充分，仍需进一步结合实验或仿真进行研究。

2）在可靠性建模、分析及增长研究方面，国内外围绕SOH估计、RUL预测和可靠性分析等方面开展了大量研究，能够有效对电池的寿命和可靠性进行评估。但现有研究主要依靠实验室标定数据开展研究，缺乏对实际工况条件的考虑，还需进一步扩展应用场景范围。此外，对于锂离子电池可靠性增长方法研究尚不充分，仍需探索或借鉴软件可靠性增长理论，对已部署的锂离子电池可靠性进行优化提升，以实现延寿、预警等效果。

3）在电池故障建模与预警算法方面，国内外已针对单体级的全生命周期数字孪生模型开展了研究，实现部分副反应的模拟、预测和跟踪。然而，现有数字孪生模型仍停留在简单电化学模型或等效电路模型阶段，用于电池深度状态估计、参数辨识、局部特征跟踪等功能，对于数字孪生模型应实现的失效预测、微观参数解析等研究仍不充分。此外，国内外围绕数据驱动的故障诊断、预警算法也开展了大量研究，然而，现有起火事故数据内容不全面、起火事故案例少，导致热失控事故数据可用信息匮乏、可用数据有限，增大了在役电池故障诊断及热失控预警算法研发难度。

参考文献（References）

[1] MENG H X, LI Y F. A review on prognostics and health management (PHM) methods of lithium-ion batteries[J]. Renewable and Sustainable Energy Reviews, 2019, 116: 109405.

[2] KHALEGHI S, HOSEN M S, KARIMI D, et al. Developing an online data-driven approach for prognostics and health management of lithium-ion batteries[J]. Applied Energy, 2022, 308: 118348.

[3] KONG J Z, YANG F F, ZHANG X, et al. Voltage-temperature health feature extraction to improve prognostics and health management of lithium-ion batteries[J]. Energy, 2021, 223: 120114.

[4] LYU Z Q, WANG G, GAO R J. Li-ion battery prognostic and health management through an indirect hybrid model[J]. Journal of Energy Storage, 2021, 42: 102990.

[5] ZHANG Y, LI Y F. Prognostics and health management of Lithium-ion battery using deep learning methods: a review[J]. Renew-

able and Sustainable Energy Reviews, 2022, 161: 112282.

[6] YANG Z, YAN Y, PAK U. Thermal reliability assessment and sensitivity analysis of 18, 650 cylindrical lithium-ion battery[J]. Journal of Energy Storage, 2023, 59: 106504.

[7] CROWLEY P J, SCANLAN K P, MANTHIRAM A. Diffusional lithium trapping as a failure mechanism of aluminum foil anodes in lithium-ion batteries[J]. Journal of Power Sources, 2022, 546: 231973.

[8] KALNAUS S, WANG Y L, TURNER J A. Mechanical behavior and failure mechanisms of Li-ion battery separators[J]. Journal of Power Sources, 2017, 348: 255-263.

[9] SARKAR A, NLEBEDIM I C, SHROTRIYA P. Performance degradation due to anodic failure mechanisms in lithium-ion batteries[J]. Journal of Power Sources, 2021, 502: 229145.

[10] WANG Y B, ZHANG C P, HU J, et al. Investigation on calendar experiment and failure mechanism of lithium-ion battery electrolyte leakage[J]. Journal of Energy Storage, 2022, 54: 105286.

[11] ZHANG Y, CHENG S Y, MEI W X, et al. Understanding of thermal runaway mechanism of LiFePO4 battery in-depth by three-level analysis[J]. Applied Energy, 2023, 336: 120695.

[12] MAO B B, CHEN H D, CUI Z X, et al. Failure mechanism of the lithium ion battery during nail penetration[J]. International Journal of Heat and Mass Transfer, 2018, 122: 1103-1115.

[13] SPIELBAUER M, BERG P, SOELLNER J, et al. Experimental investigation of the failure mechanism of 18650 lithium-ion batteries due to shock and drop[J]. Journal of Energy Storage, 2021, 43: 103213.

[14] TENG J H, TANG X, TANG M Q, et al. Failure mechanism and voltage regulation strategy of low N/P ratio lithium iron phosphate battery[J]. Journal of Energy Storage, 2022, 51: 104588.

[15] ZHANG H J, ZHOU M Z, HU L L, et al. Mechanism of the dynamic behaviors and failure analysis of lithium-ion batteries under crushing based on stress wave theory[J]. Engineering Failure Analysis, 2020, 108: 104290.

[16] REN D S, FENG X N, LU L G, et al. Overcharge behaviors and failure mechanism of lithium-ion batteries under different test conditions[J]. Applied Energy, 2019, 250: 323-332.

[17] SAHRAEI E, BOSCO E, DIXON B, et al. Microscale failure mechanisms leading to internal short circuit in Li-ion batteries under complex loading scenarios[J]. Journal of Power Sources, 2016, 319: 56-65.

[18] WANG L B, DUAN X D, LIU B H, et al. Deformation and failure behaviors of anode in lithium-ion batteries: Model and mechanism[J]. Journal of Power Sources, 2020, 448: 227468.

[19] 王芳, 王峥, 林春景, 等. 新能源汽车动力电池安全失效潜在原因分析[J]. 储能科学与技术, 2022, 11(5): 1411-1418.
WANG F, WANG Z, LIN C J, et al. Analysis on potential causes of safety failure of new energy vehicles[J]. Energy Storage Science and Technology, 2022, 11(5): 1411-1418(in Chinese).

[20] XI S J, ZHAO Q C, CHANG L J, et al. The dynamic failure mechanism of a lithium-ion battery at different impact velocity[J]. Engineering Failure Analysis, 2020, 116: 104747.

[21] APPLEBERRY M C, KOWALSKI J A, AFRICK S A, et al. Avoiding thermal runaway in lithium-ion batteries using ultrasound detection of early failure mechanisms[J]. Journal of Power Sources, 2022, 535: 231423.

[22] HENDRICKS C, WILLIARD N, MATHEW S, et al. A failure modes, mechanisms, and effects analysis (FMMEA) of lithium-ion batteries[J]. Journal of Power Sources, 2015, 297: 113-120.

[23] HAO W F, YUAN Z R, LI D D, et al. Study on mechanical properties and failure mechanism of 18650 Lithium-ion battery using digital image correlation and acoustic emission[J]. Journal of Energy Storage, 2021, 41: 102894.

[24] 王宏伟, 刘军, 邓爽, 等. 锰酸锂动力电池失效分析[J]. 电源技术, 2012, 36(9): 1276-1278.
WANG H W, LIU J, DENG S, et al. Failure analysis of LiMn2O4 power battery[J]. Chinese Journal of Power Sources, 2012, 36(9): 1276-1278(in Chinese).

[25] 朱晓庆, 王震坡, WANG H, 等. 锂离子动力电池热失控与安全管理研究综述[J]. 机械工程学报, 2020, 56(14): 91-118.
ZHU X Q, WANG Z P, WANG H, et al. Review of thermal runaway and safety management for lithium-ion traction batteries in electric vehicles[J]. Journal of Mechanical Engineering, 2020, 56(14): 91-118(in Chinese).

[26] YU D, REN D S, DAI K R, et al. Failure mechanism and predictive model of lithium-ion batteries under extremely high transient impact[J]. Journal of Energy Storage, 2021, 43: 103191.

[27] ZHANG G X, WEI X Z, ZHU J G, et al. Revealing the failure mechanisms of lithium-ion batteries during dynamic overcharge[J]. Journal of Power Sources, 2022, 543: 231867.

[28] 兰凤崇, 郑文杰, 李志杰, 等. 车用动力电池的挤压载荷变形响应及内部短路失效分析[J]. 华南理工大学学报 (自然科学版), 2018, 46(6): 65-72.
LAN F C, ZHENG W J, LI Z J, et al. Compression load-deformation response and internal short circuit failure analysis of vehicle powered batteries[J]. Journal of South China University of Technology (Natural Science Edition), 2018, 46(6): 65-72(in Chinese).

[29] XIA Q, WANG Z L, REN Y, et al. Performance reliability analysis and optimization of lithium-ion battery packs based on multiphysics simulation and response surface methodology[J]. Journal of Power Sources, 2021, 490: 229567.

[30] ZHAO J Y, FENG X N, PANG Q Q, et al. Battery prognostics and health management from a machine learning perspective[J]. Journal of Power Sources, 2023, 581: 233474.

[31] PHILIPPOT M, COSTA D, HOSEN M S, et al. Environmental impact of the second life of an automotive battery: reuse and repurpose based on ageing tests[J]. Journal of Cleaner Production, 2022, 366: 132872.

[32] MENG H X, GENG M Y, XING J D, et al. A hybrid method for prognostics of lithium-ion batteries capacity considering regeneration phenomena[J]. Energy, 2022, 261: 125278.

[33] LIU J Z, LIU X T. An improved method of state of health prediction for lithium batteries considering different temperature[J]. Journal of Energy Storage, 2023, 63: 107028.

[34] GUO F, WU X W, LIU L L, et al. Prediction of remaining useful life and state of health of lithium batteries based on time series feature and Savitzky-Golay filter combined with gated recurrent unit neural network[J]. Energy, 2023, 270: 126880.

[35] JORGE I, MESBAHI T, SAMET A, et al. Time series feature extraction for lithium-ion batteries state-of-health prediction[J]. Journ-

al of Energy Storage, 2023, 59: 106436.

[36] CHENG S J, LI B M, YUAN Z Z, et al. Development of a lifetime prediction model for lithium thionyl chloride batteries based on an accelerated degradation test[J]. Microelectronics Reliability, 2016, 65: 274-279.

[37] HENSCHEL J, HORSTHEMKE F, STENZEL Y P, et al. Lithium ion battery electrolyte degradation of field-tested electric vehicle battery cells——a comprehensive analytical study[J]. Journal of Power Sources, 2020, 447: 227370.

[38] ECKER M, GERSCHLER J B, VOGEL J, et al. Development of a lifetime prediction model for lithium-ion batteries based on extended accelerated aging test data[J]. Journal of Power Sources, 2012, 215: 248-257.

[39] WANG J, PUREWAL J, LIU P, et al. Degradation of lithium ion batteries employing graphite negatives and nickel–cobalt–manganese oxide + spinel manganese oxide positives: Part 1, aging mechanisms and life estimation[J]. Journal of Power Sources, 2014, 269: 937-948.

[40] WU J J, KONG L G, CHENG Z, et al. RUL prediction for lithium batteries using a novel ensemble learning method[J]. Energy Reports, 2022, 8: 313-326.

[41] ZHU W H, LEI F, ZHONG H, et al. An improved reliability assessment method for lithium-ion battery system considering imbalanced current and uneven cooling[J]. Energy, 2023, 276: 127424.

[42] BAI X, TU B Y. Ellipsoid non-probabilistic reliability analysis of the crack growth fatigue of a new titanium alloy used in deep-sea manned cabin[J]. Theoretical and Applied Fracture Mechanics, 2021, 115: 103041.

[43] GONG C Q, FRANGOPOL D M. Time-variant hull girder reliability considering spatial dependence of corrosion growth, geometric and material properties[J]. Reliability Engineering & System Safety, 2020, 193: 106612.

[44] HEYDARI M, SULLIVAN K M. Robust allocation of testing resources in reliability growth[J]. Reliability Engineering & System Safety, 2019, 192: 106020.

[45] BYUN J E, NOH H M, SONG J. Reliability growth analysis of k-out-of-N systems using matrix-based system reliability method[J]. Reliability Engineering & System Safety, 2017, 165: 410-421.

[46] UTKIN L V, COOLEN F P A. A robust weighted SVR-based software reliability growth model[J]. Reliability Engineering & System Safety, 2018, 176: 93-101.

[47] AWAD M. Economic allocation of reliability growth testing using Weibull distributions[J]. Reliability Engineering & System Safety, 2016, 152: 273-280.

[48] NADJAFI M, GHOLAMI P. Expectation-maximization algorithm to develop a normal distribution NHPP reliability growth model[J]. Engineering Failure Analysis, 2022, 140: 106575.

[49] HUANG Y S, CHIU K C, CHEN W M. A software reliability growth model for imperfect debugging[J]. Journal of Systems and Software, 2022, 188: 111267.

[50] ZOU G, KOLIOS A. Quantifying the value of negative inspection outcomes in fatigue maintenance planning: cost reduction, risk mitigation and reliability growth[J]. Reliability Engineering & System Safety, 2022, 226: 108668.

[51] FENG J Q, CAI F, LI H C, et al. A data-driven prediction model for the remaining useful life prediction of lithium-ion batteries[J]. Process Safety and Environmental Protection, 2023, 180: 601-615.

[52] SHU X, SHEN J W, CHEN Z, et al. Remaining capacity estimation for lithium-ion batteries via co-operation of multi-machine learning algorithms[J]. Reliability Engineering & System Safety, 2022, 228: 108821.

[53] WANG S L, TAKYI-ANINAKWA P, JIN S Y, et al. An improved feedforward-long short-term memory modeling method for the whole-life-cycle state of charge prediction of lithium-ion batteries considering current-voltage-temperature variation[J]. Energy, 2022, 254: 124224.

[54] WANG Z, YANG F F, XU Q, et al. Capacity estimation of lithium-ion batteries based on data aggregation and feature fusion via graph neural network[J]. Applied Energy, 2023, 336: 120808.

[55] LIU X H, ZHANG L S, YU H Q, et al. Bridging multiscale characterization technologies and digital modeling to evaluate lithium battery full lifecycle[J]. Advanced Energy Materials, 2022, 12(33): 2200889.

[56] NGANDJONG A C, LOMBARDO T, PRIMO E N, et al. Investigating electrode calendering and its impact on electrochemical performance by means of a new discrete element method model: Towards a digital twin of Li-Ion battery manufacturing[J]. Journal of Power Sources, 2021, 485: 229320.

[57] HUSSEINI K, SCHMIDGRUBER N, HENSCHEL S, et al. Model-based optimization of web tension control for the flexible cell stack assembly of lithium-ion battery cells[J]. Energy Technology, 2023, 11(5): 2200679.

[58] WANG C Y, LIU T, YANG X G, et al. Fast charging of energy-dense lithium-ion batteries[J]. Nature, 2022, 611(7936): 485-490.

[59] BILANSKY J, LACKO M, PASTOR M, et al. Improved digital twin of Li-ion battery based on generic MATLAB model[J]. Energies, 2023, 16(3): 1194.

[60] HE W B, MAO J X, SONG K, et al. Structural performance prediction based on the digital twin model: a battery bracket example[J]. Reliability Engineering & System Safety, 2023, 229: 108874.

[61] RENIERS J M, HOWEY D A. Digital twin of a MWh-scale grid battery system for efficiency and degradation analysis[J]. Applied Energy, 2023, 336: 120774.

[62] PAN M Y, XING Q J, CHAI Z C, et al. Real-time digital twin machine learning-based cost minimization model for renewable-based microgrids considering uncertainty[J]. Solar Energy, 2023, 250: 355-367.

[63] SÖDERÄNG E, HAUTALA S, MIKULSKI M, et al. Development of a digital twin for real-time simulation of a combustion engine-based power plant with battery storage and grid coupling[J]. Energy Conversion and Management, 2022, 266: 115793.

[64] TANG H, WU Y C, CAI Y F, et al. Design of power lithium battery management system based on digital twin[J]. Journal of Energy Storage, 2022, 47: 103679.

[65] LI W H, RENTEMEISTER M, BADEDA J, et al. Digital twin for battery systems: cloud battery management system with online state-of-charge and state-of-health estimation[J]. Journal of Energy Storage, 2020, 30: 101557.

[66] SEMERARO C, ALJAGHOUB H, ALI ABDELKAREEM M, et al. Digital twin in battery energy storage systems: trends and gaps de-

tection through association rule mining[J]. Energy, 2023, 273: 127086.

[67] SEMERARO C, OLABI A G, ALJAGHOUB H, et al. Digital twin application in energy storage: trends and challenges[J]. Journal of Energy Storage, 2023, 58: 106347.

[68] WU B, WIDANAGE W D, YANG S C, et al. Battery digital twins: perspectives on the fusion of models, data and artificial intelligence for smart battery management systems[J]. Energy and AI, 2020, 1: 100016.

[69] MA Y, SHAN C, GAO J W, et al. Multiple health indicators fusion-based health prognostic for lithium-ion battery using transfer learning and hybrid deep learning method[J]. Reliability Engineering & System Safety, 2023, 229: 108818.

[70] XIA L L, WANG S L, YU C M, et al. Joint estimation of the state-of-energy and state-of-charge of lithium-ion batteries under a wide temperature range based on the fusion modeling and online parameter prediction[J]. Journal of Energy Storage, 2022, 52: 105010.

[71] 刘晓俊. 电池故障智能诊断系统的研究与实现[D]. 北京: 北京邮电大学, 2010.
LIU X J. Research and application of intelligent battery fault diagnosis system[D]. Beijing: Beijing University of Posts and Telecommunications, 2010(in Chinese).

[72] 夏淑英. 基于模糊故障树的汽车故障诊断方法及其应用研究[D]. 武汉: 湖北工业大学, 2017.
XIA S Y. Research on the automobile fault diagnosis its method and application based on the Fuzzy Fault Tree[D]. Wuhan: Hubei University of Technology, 2017(in Chinese).

[73] ZHU X Q, WANG Z P, WANG Y T, et al. Overcharge investigation of large format lithium-ion pouch cells with Li(Ni0.6Co0.2Mn0.2)O2 cathode for electric vehicles: thermal runaway features and safety management method[J]. Energy, 2019, 169: 868-880.

[74] 赵洋. 基于大数据的电动汽车动力电池系统故障诊断方法研究[D]. 北京: 北京理工大学, 2018.
ZHAO Y. Research on fault diagnosis method of electric vehicle power battery system based on big data[D]. Beijing: Beijing Institute of Technology, 2018(in Chinese).

[75] SUN Z Y, HAN Y, WANG Z P, et al. Detection of voltage fault in the battery system of electric vehicles using statistical analysis[J]. Applied Energy, 2022, 307: 118172.

[76] 刘鹏, 吴志强, 张照生, 等. 基于电压频域特征和异常系数的动力电池故障诊断方法[J]. 中国公路学报, 2022, 35(8): 89-104.
LIU P, WU Z Q, ZHANG Z S, et al. Fault diagnosis for battery systems based on voltage frequency-domain indicator and abnormal coefficient[J]. China Journal of Highway and Transport, 2022, 35(8): 89-104(in Chinese).

[77] WANG L S, FANG Y Y, WANG L, et al. Understanding discharge voltage inconsistency in lithium-ion cells via statistical characteristics and numerical analysis[J]. IEEE Access, 2020, 8: 84821-84836.

[78] MA M N, WANG Y, DUAN Q L, et al. Fault detection of the connection of lithium-ion power batteries in series for electric vehicles based on statistical analysis[J]. Energy, 2018, 164: 745-756.

[79] XUE Q, LI G, ZHANG Y J, et al. Fault diagnosis and abnormality detection of lithium-ion battery packs based on statistical distribution[J]. Journal of Power Sources, 2021, 482: 228964.

[80] 赵帅. 锂离子动力电池内短路机理与诊断方法研究[D]. 长春: 吉林大学, 2022.
ZHAO S. Study on the mechanism and diagnosis method of internal short circuit in lithium-ion power battery[D]. Changchun: Jilin University, 2022(in Chinese).

[81] 温涛. 锂离子电池内短路故障电热特性等效实验及诊断方法研究[D]. 哈尔滨: 哈尔滨理工大学, 2022.
WEN T. Research on equivalent experiment and diagnosis method of electrothermal characteristics of lithium-ion battery internal short circuit fault[D]. Harbin: Harbin University of Science and Technology, 2022(in Chinese).

[82] 武超. 磷酸铁锂动力电池典型故障机理及诊断方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
WU C. Research on the typical fault mechanism and diagnostic approach for power lithium iron phosphate battery[D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

[83] HU X S, ZHANG K, LIU K L, et al. Advanced fault diagnosis for lithium-ion battery systems: a review of fault mechanisms, fault features, and diagnosis procedures[J]. IEEE Industrial Electronics Magazine, 2020, 14(3): 65-91.

[84] ZHANG Y L, JIANG M, ZHOU Y H, et al. Towards high-safety lithium-ion battery diagnosis methods[J]. Batteries, 2023, 9(1): 63.

[85] LOVE C T, DUBARRY M, RESHETENKO T, et al. Lithium-ion cell fault detection by single-point impedance diagnostic and degradation mechanism validation for series-wired batteries cycled at 0 ℃[J]. Energies, 2018, 11(4): 834.

[86] 陈佳纯. 基于数据分析的锂离子动力电池热失控故障诊断研究[D]. 北京: 北京交通大学, 2020.
CHEN J C. Research on thermal runaway of vehicle lithium-ion battery based on data analysis[D]. Beijing: Beijing Jiaotong University, 2020(in Chinese).

[87] 廉玉波, 凌和平, 王钧斌, 等. 基于混合高斯-隐马尔可夫模型的动力电池实时热失控检测[J]. 汽车工程, 2023, 45(1): 139-146.
LIAN Y B, LING H P, WANG J B, et al. A real-time thermal runaway detection method of power battery based on guassian mixed model and hidden Markov model[J]. Automotive Engineering, 2023, 45(1): 139-146(in Chinese).

[88] LIU Z, GUO X R, MENG N, et al. Study of thermal runaway and the combustion behavior of lithium-ion batteries overcharged with high current rates[J]. Thermochimica Acta, 2022, 715: 179276.

[89] WANG J J, JIA J R, ZHANG Z H. Research on the early warning mechanism for thermal runaway of lithium-ion power batteries in electric vehicles[J]. Security and Communication Networks, 2022, 2022: 8573396.

[90] MENG H X, YANG Q Q, ZIO E, et al. An integrated methodology for dynamic risk prediction of thermal runaway in lithium-ion batteries[J]. Process Safety and Environmental Protection, 2023, 171: 385-395.

Advances in prognosis and health management technologies of lithium-ion batteries

YANG Shichun[1, 2], WANG Xiao[1], CHEN Fei[1, *], ZHENG Yifan[1], ZHONG Yilin[3], ZHOU Sida[1]

(1. School of Transportation Science and Engineering, Beihang University, Beijing 100191, China;
2. Technology Innovation Center of New Energy Vehicle Digital Supervision Technology and Application, State Administration for Market Regulation, Beijing 100120, China;
3. Automotive Engineering & Research Institute, BYD Automobile Industry Co., Ltd., Shenzhen 518118, China)

Abstract: The prognosis and health management (PHM) technology of lithium-ion batteries helps reduce the probability of failure and improve the service time by analyzing current reliability and formulating a comprehensive control strategy. Analysis and forecasting of battery failure are key components of battery PHM technology. This article examines the state of research on the main PHM technologies, such as reliability analysis, testing, and growth technology, failure diagnosis and early warning algorithms, failure model based on digital twin technology, and fault analysis and probability prediction. And the article introduced the general architecture for PHM technology on cloud battery management systems for electric vehicles. The advantages and disadvantages of existing technologies are summarized, and the challenges and future perspectives of PHM technology are prospected.

Keywords: lithium-ion batteries; prognosis and health management; reliability; digital twin; failure diagnosis

Received: 2023-05-30; **Accepted**: 2023-09-08; **Published Online**: 2023-12-25 09:02
URL: link.cnki.net/urlid/11.2625.V.20231224.1612.001
Foundation item: National Key Research and Development Program of China (2021YFB2501700)
* **Corresponding author**. E-mail: cf2020@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0265

基于多变量机器学习的CYGNSS有效波高反演模型

张云[1]，肖盛[1]，姜丽菲[2]，孟婉婷[2]，杨树瑚[1,*]，韩彦岭[1]

(1. 上海海洋大学 信息学院 上海市海洋智能信息与导航遥感工程中心，上海 201306；2. 上海航天电子技术研究所，上海 201109)

摘　　要：旋风全球导航卫星系统(CYGNSS)提供了高质量的全球导航卫星系统反射信号(GNSS-R)技术数据，能可靠地用于有效波高(SWH)的反演。由于CYGNSS的高动态性，导致接收信号很容易受到环境因素的影响，海况的复杂性使简单模型难以准确反演SWH。为解决上述问题，提出一种基于机器学习的多变量SWH反演模型，根据海浪的形成机理及对CYGNSS参数与SWH之间的相关性结果分析选取出相关参数，并设计5参数、9参数和17参数3种训练方案。随后利用随机森林(RF)和卷积神经网络(CNN)对反演模型进行训练和验证，并将SWH反演结果与欧洲中期天气预报中心(ECMWF)的参考值进行比较。最佳的反演模型是17参数CNN反演模型，均方根误差（RMSE）为0.184 0 m，相关系数R^2=0.948 5。与17参数CNN反演模型相比，9参数CNN反演模型减少了24%的训练时间，并且精度损失很小。但9参数反演模型相较17参数反演模型在泛化评估方面表现不佳。因此，为提高模型的泛化能力，将风速作为参数添加到17参数反演模型中，得到了17+1参数泛化模型。其中，最佳的泛化模型是17+1参数RF泛化模型，RMSE为0.497 1 m，R^2=0.584 6。有效地证明了所提模型在SWH反演中具有良好的潜力。

关　键　词：有效波高；旋风全球导航卫星系统；星载GNSS-R；机器学习；卷积神经网络
中图分类号：P715.6
文献标志码：A　　**文章编号**：1001-5965(2025)05-1503-11

有效波高(significant wave height, SWH)是海洋动力环境监测的重要组成部分。准确获取SWH信息对海洋灾害预警和海洋生产安全具有重要意义。近年来，随着全球导航卫星系统(global navigation satellite system, GNSS)的快速发展，GNSS技术不仅可以为用户提供导航、定位和精确的授时信息，还可以提供全天时、全天候、不间断的L波段微波信号源，这使得使用GNSS进行遥感成为可能。全球导航卫星系统反射信号(global navigation satellite system reflectometry, GNSS-R)技术作为一种新的遥感技术，已应用于海洋和陆地遥感领域，如海面风速和风向反演[1-3]、海面高度反演[4-8]、雪深反演[9-10]、海冰探测[11-14]、土壤湿度反演[15-16]和水体检测[17-18]，其已成为陆海遥感领域的一种流行技术手段。由于GNSS-R重访周期短和观测成本低的优点，在反演SWH领域显示出巨大的潜力。

1993年，Martín-Neira[19]发现并提出利用GNSS-R进行海洋测高的方法。2004年，Soulat等[20]在Starla和巴塞罗那港务局进行的Oceanpal岸基海洋实验的基础上推导了干涉复数场(interferometric complex field, ICF)中自相关函数的表达式，并提出使用ICF中有效相关时间函数反演海浪SWH的方法。1982年，Alpers和Hasselmann[21]通过拟合建立了合成孔径雷达回波信号信噪比(signal-to-noise ratio, SNR)的均方根与SWH之间的线性关系，并提出使用图像的SNR来估计SWH。2015年，Alonso-Arroyo等[22]使用GNSS-R干涉图技术(interference pattern technique, IPT)反演SWH，在基于海岸的实验中精

收稿日期：2023-05-23；**录用日期**：2023-08-04；**网络出版时间**：2023-08-30 13:37
网络出版地址：link.cnki.net/urlid/11.2625.V.20230829.1516.001
基金项目：国家自然科学基金(42176175,42271335)；国家重点研发计划(2019YFD0900805)
*通信作者. E-mail：shyang@shou.edu.cn

引用格式：张云，肖盛，姜丽菲，等. 基于多变量机器学习的CYGNSS有效波高反演模型[J]. 北京航空航天大学学报，2025，51(5)：1503-1513. ZHANG Y，XIAO S，JIANG L F，et al. Significant wave height retrieval model of CYGNSS based on multivariate machine learning[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1503-1513 (in Chinese).

度达到6 cm。2021年，Qin和Li[23]通过GNSS-R时延多普勒图(delayed Doppler map, DDM)获得了DDM的SNR（SNR of DDM, DSNR），并结合多颗卫星的仰角差引入了卫星仰角校正，提高了DSNR和SWH之间的相关性。在实验中，精度达到0.167 1 m。

上述实验和方法都是在岸基条件下建立的，地面或机载GNSS-R设备很难实现大规模测量，而ICF和IPT方法的条件苛刻且相对复杂。随着英国UK-DMC卫星、萨里卫星技术有限公司的TDS-1卫星和美国宇航局的旋风全球导航卫星系统(cyclone global navigation satellite system, CYGNSS)卫星接收机的成功发射，开始提供高质量的全球GNSS-R数据，GNSS-R有效波高反演的研究开始向星载方向转移。Clarizia等[24]提出星载GNSS-R DDM可以用于分析海况，并通过UK-DMC卫星验证了DDM与海况之间的一致性，这也为使用星载GNSS-R DDM反演SWH提供了新的思路。Peng和Jin[25]使用星载CYGNSS-R数据提取DDM的SNR，使用DSNR和SWH的半经验公式反演全球海洋的SWH，精度为0.276 1 m，证实了从星载GNSS-R中反演SWH的可行性。Yang等[26]使用DDM数据生成的时延多普勒图平均值(delay Doppler map average, DDMA)和前沿斜率(leading edge slope, LES)来联合反演SWH，精度提高到0.247 m。Bu和Yu[27]通过处理GNSS-R中的DDM来获得归一化积分延迟波形(normalized integral delay waveform, NIDW)和归一化中心延迟波形(normalized central delay waveform, NCDW)，并从中提取了4个GNSS-R观测参数，即NIDW和NCDW的LES，以及NIDE和NCDE的后沿斜率(trailing edge slope, TES)。建立了4个GNSS-R观测变量与SWH之间的关系，其中，NCDW LES变量的精度最高，为0.649 m。Wang等[28]通过拟合解释了风与波的关系，并提取了4组可能影响SWH的参数变量，并将其代入神经网络进行学习。其中，包括归一化双基地雷达截面(normalized bistatic radar cross section, NBRCS)、LES、DSNR、镜面反射点增益、入射角、方位角、经度、纬度、距离和风速的10参数模型精度最高，通过数据滤波器后精度达到0.49 m。

机器学习凭借着能够自动化处理大量数据、模型泛化能力强等优势，已被证明适用于星载GNSS-R，并能获得良好的检索精度。现阶段，对SWH的研究已经朝着机器学习的方向发展。然而，SWH作为一个重要的海况监测因素，受到许多复杂的环境因素的影响，如涌浪和风速。在星载GNSS-R条件下，这种效应更为明显，导致简单的模型无法很好地反演SWH。为解决这一问题，本文提出一种基于多变量机器学习的CYGNSS有效波高反演模型。机器学习使用的参数包括DDM参数、卫星系统参数及外部参数，这些参数是根据海浪的形成机理及CYGNSS参数与SWH之间的相关性结果分析来选取的。

1 原理及方法

1.1 风速和涌浪与SWH的关系

在连续波谱中，将N个海浪从大到小排序后，选择排列后海浪的前1/3求平均值，并将平均值作为SWH值，其单位为m[29]。SWH定义如图1所示。

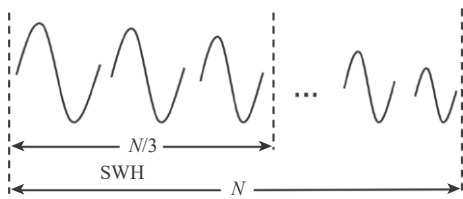


图1 SWH的定义

Fig. 1 Definition of SWH

SWH主要由风浪和涌浪组成。对于完全开放的海洋，风浪SWH可表示为[30]

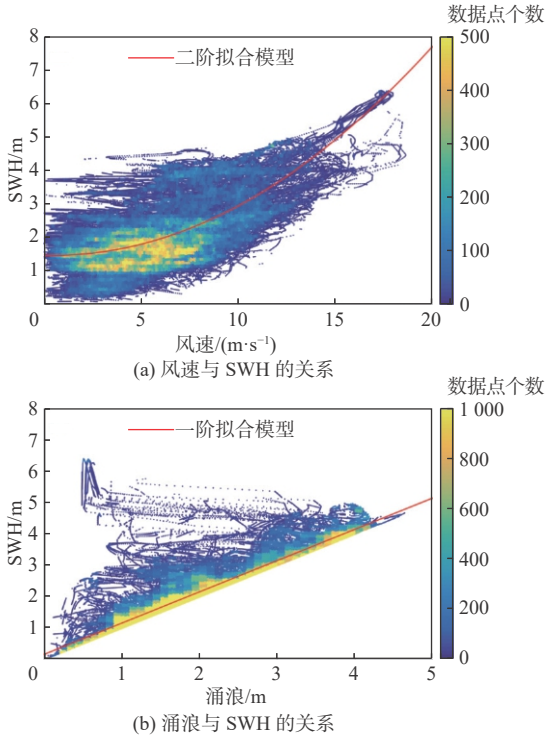
$$S_{\mathrm{WH\,wind}} \approx 0.22u_{10}^2/g \tag{1}$$

式中：u_{10}为距海面10 m高处平均风速；g为重力加速度。已有研究表明，当风速高时，风速和SWH的变化趋势表现出一致性。然而，随着风速的降低，风速和SWH之间的相关性也降低，风速和SWH的变化趋势不断偏离，与此同时，SWH和涌浪之间的相关不断增加，并且变化趋势变得更加密切。这表明，在低风速下，海浪的主要驱动力是涌浪，而在高风速下，海浪的主要驱动因素是风速[31]。

以2020年1月1日的数据为例，图2显示了风速和涌浪与SWH的关系。由图2(a)可知，风速与SWH之间存在二阶关系，相关系数R^2为0.425 9。在高风速区间，风速与SWH密切相关，散射点接近拟合曲线，二阶模型表现更好。然而，二阶模型在低风速范围内表现不佳。由图2(b)可知，涌浪和SWH之间存在一阶线性关系，R^2=0.785 5，并且在每个区间中的相关性是相似的。

1.2 SWH反演原理

对于GNSS-R模型，风浪通常以cm级的特征存在，这些特征决定了海面的粗糙度并控制了海面的漫反射，因此，GNSS-R观测变量对风浪非常敏感。虽然涌浪通常为m级的波浪，但在大多数情况下，涌浪对海面粗糙度的影响很小，通常控制海面的相干反射[32]。对于给定的表面高度光谱，分离大粗糙度和小粗糙度的截止波长为[33]



(a) 风速与 SWH 的关系

(b) 涌浪与 SWH 的关系

图 2　风速和涌浪与 SWH 的关系

Fig. 2　Relationship of wind speed and swell with SWH

$$\lambda_c = 3\lambda/\cos\theta \tag{2}$$

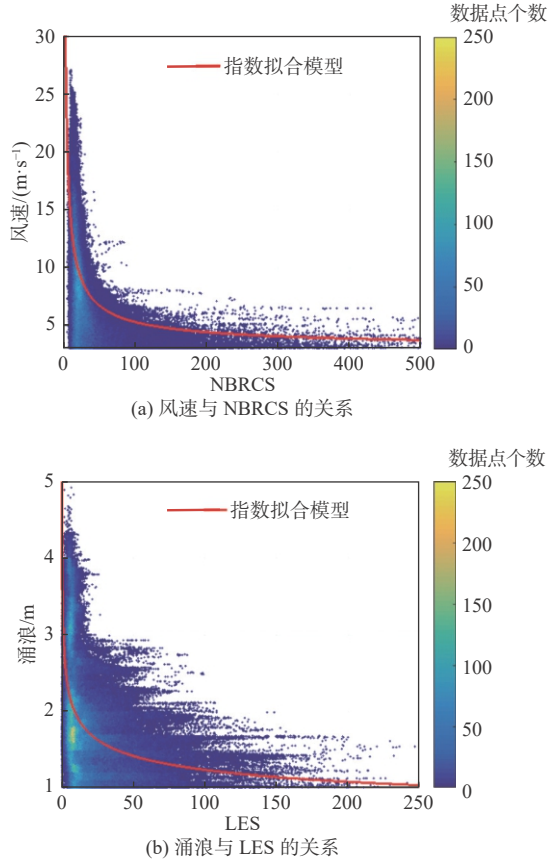
式中：λ 为 GNSS 信号的波长；θ 为 GNSS 信号的入射角。对于 GPS L1 信号，最小截止波长约为 57 cm。因此，GNSS-R 对涌浪敏感。

风浪会影响海面的粗糙度，从而降低接收信号的能量，并可以在 NBRCS 中表现出来[34]。此外，涌浪的存在将导致通过海面接收散射信号出现时间延迟，这可以反映在相关结果的观测量中，如 LES[8]。图 3 为 NBRCS 与风浪及 LES 与涌浪的关系图，由图可知，NBRCS 与风浪之间存在指数关系，其中，R^2=0.621 8，LES 与涌浪之间存在指数关系，R^2=0.538 4。

1.3　机器学习模型

本文使用随机森林 (random forest, RF) 和卷积神经网络 (convolutional neural network, CNN) 进行 SWH 反演。RF 具有学习过程快、泛化能力强和相对稳健的模型策略优点，更适合于解决回归问题[35]。RF 的本质是一种强大的学习方法，其结合了 bootstrap 聚合（bagging）和决策树。多个不相关的决策树共同决定了模型的最终输出。其原理如图 4 所示。

作为深度学习算法的代表性算法，CNN 算法通过卷积池化等操作解决了由于网络层数加深而导致的梯度爆炸问题，可以通过深度学习解决复杂的非线性问题。因此，其适用于研究多变量与结果之间的非线性关系[36]。本文使用的 CNN 结构是一个完整的 CNN 架构，包括输入层、卷积层、批量处理归一化层、ReLU 激活函数、池化层、全连接层和输出层[34]。本文使用的 CNN 模型结构如图 5 所示，其中，C1～C7 为卷积层，F8～F10 为全连接层，M 为输入维度。由于 CYGNSS 中的数据量大，参数之间的关系复杂，因此，星载 GNSS-R 适合使用 RF 和 CNN 进行 SWH 反演。



(a) 风速与 NBRCS 的关系

(b) 涌浪与 LES 的关系

图 3　风速与 NBRCS 和涌浪与 LES 的关系

Fig. 3　Relationship of wind speed with NBRCS and swell with LES

1.4　数据预处理过程

本文使用的数据集包括 CYGNSS 数据集和欧洲中期天气预报中心 (European Centre for Medium-range Weather Forecasts, ECMWF) 数据集。CYGNSS 数据集从 L0 到 L3 分为 4 个级别，本文使用 L1 级别的数据进行分析。本文使用的参考 SWH 数据集为 ECMWF SWH 风速和涌浪的组合数据集。该数据集的时间分辨率为 1 h，空间分辨率为 0.5°×0.5°。为使 2 个数据集在时间和空间上对应，对 ECMWF SWH 数据集和 CYGNSS 数据集进行时间和空间的线性插值匹配[25]。

在数据预处理中，为排除低质量的 CYGNSS 测量数据，本文根据以下标准对数据进行过滤[37]：

1）根据数据中的质量控制 (quality control, QC) 标志进行质量筛选。

2）GNSS-R 观测值及其相应的参考数据不能为空。

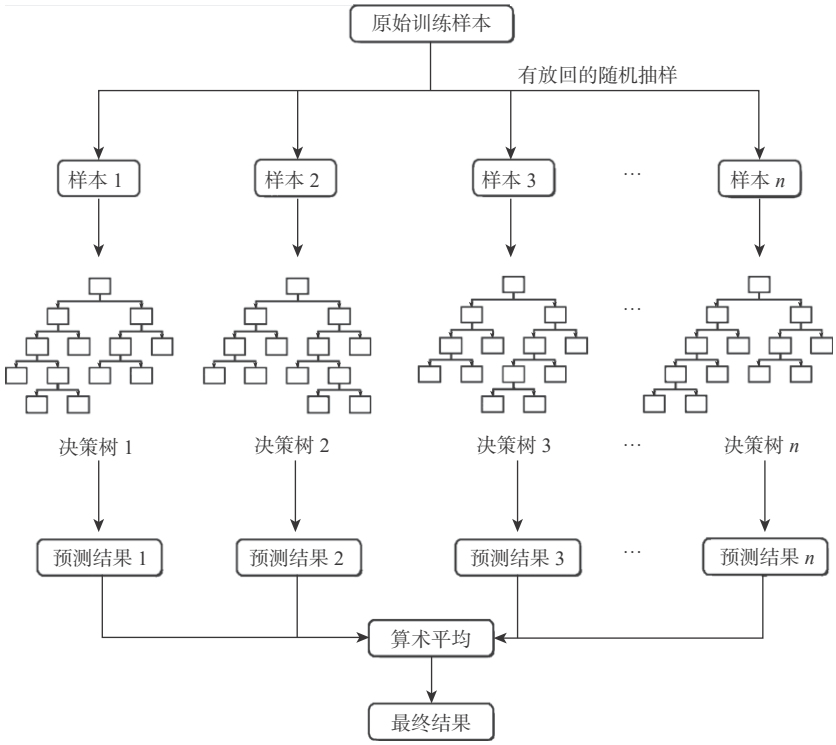


图 4 RF 原理

Fig. 4 RF principle

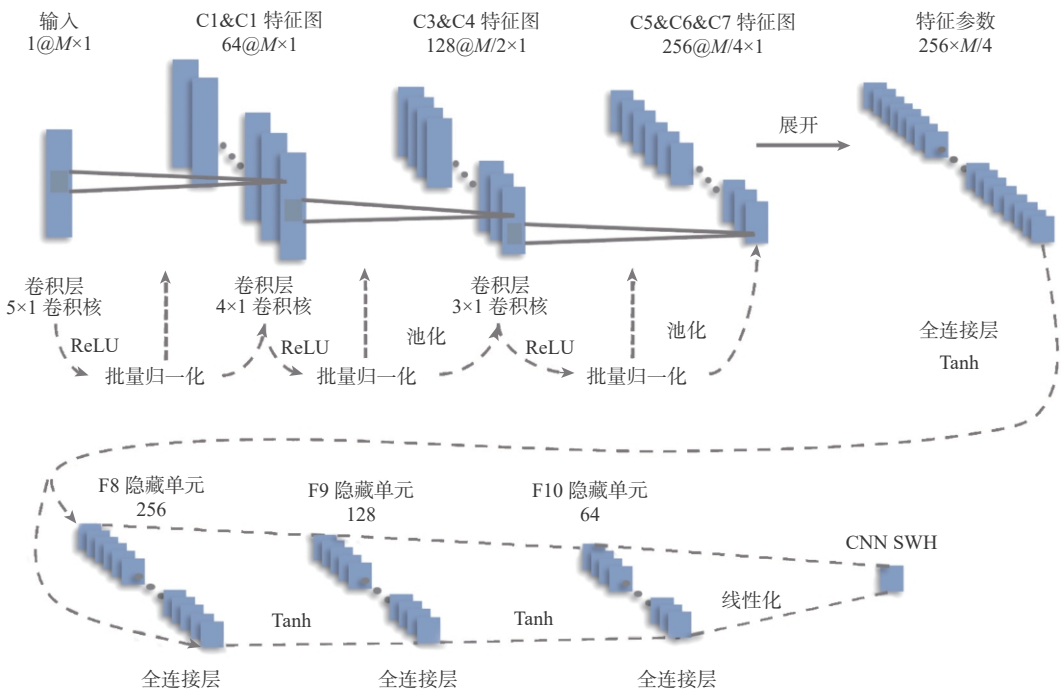


图 5 CNN 结构示意图

Fig. 5 Schematic diagram of CNN structure

3) 当卫星无法被有效跟踪时, 采集的数据将被丢弃。

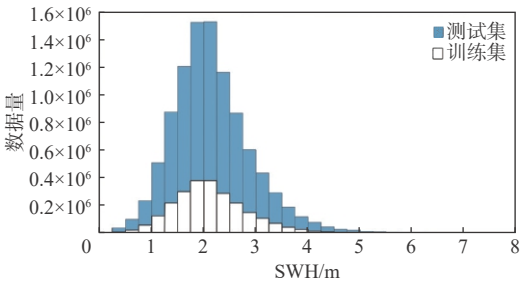
4) 由于陆地污染和近岸浅水的影响, 选择了距离陆地 25 km 以上的镜面反射点数据。

5) 观测数据范围定义为北纬 38°, 东经 0～360°。

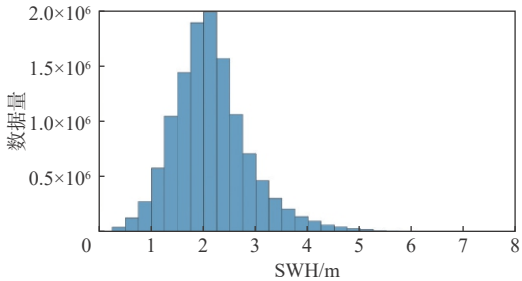
6) 观测数据要求在镜面反射点方向上的接收天线增益大于 0 dBi, 卫星天线的入射角小于 60°, 距离校正增益 (range correction gain, RCG) 大于 10。

本文使用的样本数据集是经过数据预处理后获得的。本文选取 2020 年 1 月 1 日—2020 年 2 月 10 日期间的 CYGNSS 和 ECMWF 数据进行数据匹配和预处理。其中, 对 2020 年 1 月 1 日—2020 年 1 月 31 日的预处理后的数据随机选取 250 万和 1 000 万作为训练集和测试集。使用 2020 年 2 月

1 日—2020 年 2 月 10 日的 1 226 万数据作为泛化测试集来评估反演模型的泛化能力。数据量分布如图 6 所示。



(a) 训练集、反演测试集的 SWH 数据分布



(b) 泛化测试集的 SWH 数据分布

图 6　SWH 数据量分布

Fig. 6　SWH data volume distribution

通过 RF 和 CNN 对训练集进行训练, 得到 2 个反演模型。训练前需要归一化所有数据, 以确保在训练过程中模型收敛。在获得反演模型后, 使用测试集来获得 SWH 反演结果。图 7 为 CYGNSS 反演 SWH 算法的流程。

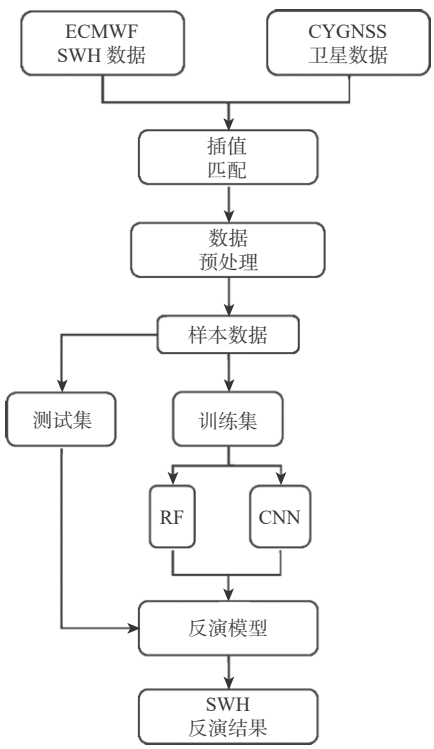


图 7　基于机器学习的 SWH 反演过程

Fig. 7　SWH retrieval process based on machine learning

2　有效波高反演过程

2.1　特征参数选择

如第 1 节所述, 为获得更好的 SWH 反演模型, 有必要尽可能找到对风速或涌浪敏感的参数。本文使用的 17 个参数是在根据海浪的形成机理及对 CYGNSS 参数与 SWH 之间的相关性结果分析基础上得出的。CYGNSS 在 GNSS-R 中有 7 个与 DDM 相关的参数, 包括 NBRCS、LES、DSNR、DDM 峰值延迟行 (DDM peak delay row, DPDR)、DDM 多普勒列 (DDM peak Doppler column, DPDC)、BRCS DDM 镜面反射点延迟行 (specular point delay row, DSDR) 和多普勒列 (specular point Doppler column, DSDC)。这些参数被广泛用于风速反演和海面高度测量。镜面反射点入射角 (specular point incidence angle, SPANG)、天顶信噪比 (zenith signal to noise ratio, ZSNR) 和镜面反射点菲涅尔功率反射系数 (Fresnel power reflection coefficient, FC) 也对 GNSS-R 反演性能有重要影响。此外, 飞机自身的参数, 如 x、y、z 轴的 GPS 速度 V_{tx}、V_{ty}、V_{tz}, x、y、z 轴的航天器速度 V_{sx}、V_{sy}、V_{sz} 及 DDM 天线序列 (DDM antenna, ANT) 也会影响 GNSS-R 的相干结果[34,38]。

为缓解机器学习耗时的过程, 本文计算了每个参数与 SWH 之间的皮尔逊相关系数, 并根据相关系数确定了线性关系。皮尔逊相关系数热力图如图 8 所示, 5 个参数 (DSNR、NBRCS、LES、V_{ty}、V_{sy}) 中的每个参数与 SWH 之间存在良好的线性相关性。此外, 尽管 4 个参数 (SPANG、ZSNR、ANT、FC) 皮尔逊相关系数不高, 但大量小样本实验表明, 这 4 个参数与 SWH 之间存在很强的内在关系。此外, 图 8 显示了本文参数之间存在弱互相关性; 只

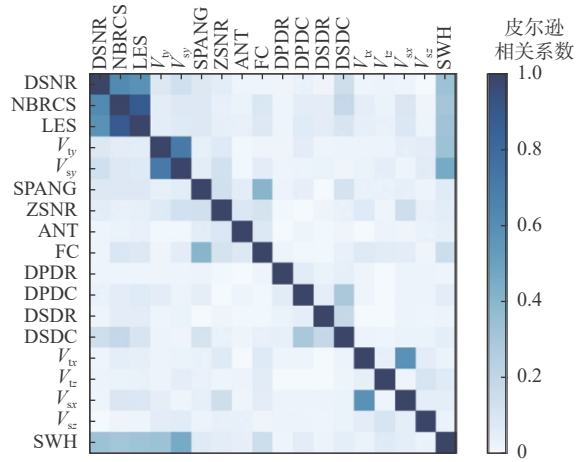


图 8　CYGNSS 参数相关系数热力图

Fig. 8　Heatmap of correlation coefficients of CYGNSS parameters

有部分参数的互相关值 (LES-NBRCS、LES-DSNR、V_{sy}-V_{ty}、V_{sx}-V_{tx} 和 NBRCSNR) 高于 0.6。经过早期的小样本实验，将这些参数同时代入训练模型的性能将比将一个参数单独代入训练模型要好得多。本文使用的 3 种方案如表 1 所示。

表 1 特征参数

Table 1 Characteristic parameters

参数列表	5参数	9参数	17参数
DSNR	○	○	○
LES	○	○	○
NBRCS	○	○	○
V_{ty}	○	○	○
V_{sy}	○	○	○
SPANG		○	○
ZSNR		○	○
ANT		○	○
FC		○	○
DPDR			○
DPDC			○
DSDR			○
DSDC			○
V_{tx}			○
V_{tz}			○
V_{sx}			○
V_{sz}			○

注：○表示使用了该参数。

2.2 反演模型的反演验证

图 9 为 6 个反演模型的反演结果和 ECMWF 提供的参考 SWH 之间的相关性散点图，图中的线为 1∶1 参考线。表 2 为 6 个反演模型的反演结果和 ECMWF 提供的参考 SWH 之间进行比较后的评估指标值。由图 9 和表 2 可知，17 参数 CNN 反演模型具有最高的精度，均方根误差（root mean square error, RMSE）为 0.184 0 m, R^2 =0.948 5。与 5 参数

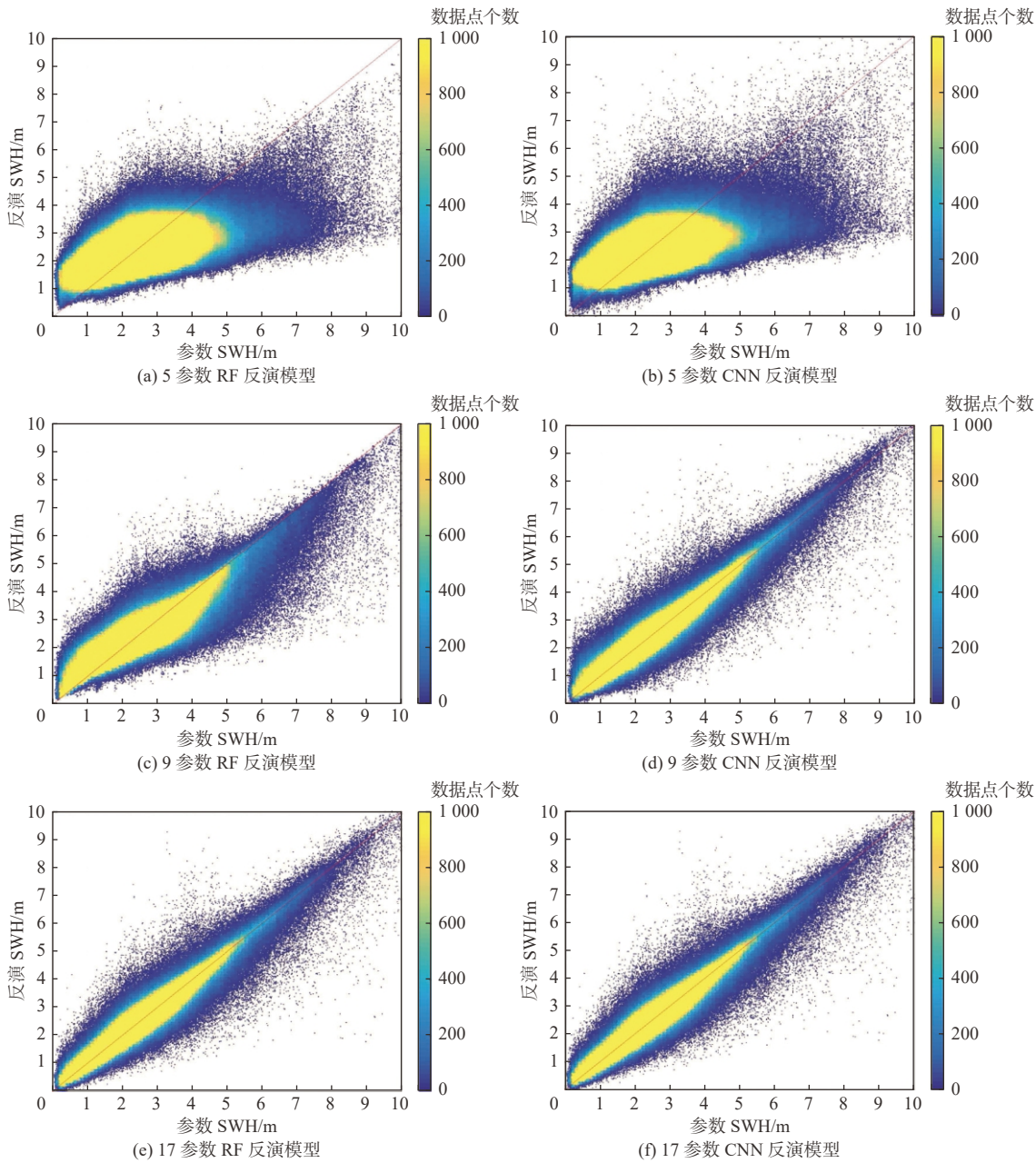


图 9 不同反演模型反演结果的相关散点图

Fig. 9 Scatter plots of retrieval results related to different retrieval models

反演模型相比, 9 参数检索模型的精度有所提高。原因是在 9 参数模型中添加了 4 个参数 (SPANG、ZSNR、ANT、FC), 这些参数通过小样本实验已知与 SWH 具有高度内在相关性。这表明, 仅使用皮尔逊相关系数作为参数选择标准是不够的。此外, 与 17 参数反演模型相比, 9 参数反演模型具有相似的精度, 但所需的运行时间大大减少。本文以 12 GBx 显存的 RTX3090 GPU 为例, 对 9 参数 CNN 反演模型进行 255 次迭代, 模型已完全收敛。与 17 参数 CNN 反演模型相比, 时间减少了约 85 min, 减少幅度为 24%, 精度仅损失了 7.5%

表 2　SWH 反演模型的反演性能分析

Table 2　Retrieval performance analysis of SWH retrieval models

反演模型	RMSE/m	R^2
5参数RF反演模型	0.647 1	0.364 1
5参数CNN反演模型	0.651 7	0.355 0
9参数RF反演模型	0.305 9	0.857 8
9参数CNN反演模型	0.197 8	0.940 5
17参数RF反演模型	0.210 7	0.932 5
17参数CNN反演模型	0.184 0	0.948 5

3　有效波高泛化过程

3.1　反演模型的泛化验证

根据第 2.2 节的评估结果, 与 9 参数反演模型和 17 参数反演模型相比, 5 参数反演模型的精度显著下降, 9 参数反演模型和 17 参数反演模型的精度相似。因此, 本节仅对 9 参数反演模型和 17 参数反演模型进行泛化评估。

从图 10 的相关散点图和表 3 的泛化性能分析中可知, 最佳精度模型是 17 参数 RF 反演模型, RMSE 为 0.517 5 m, R^2=0.549 8。与反演评估结果相比, 9 参数反演模型和 17 参数反演模型的泛化精度有所下降, 可能与特征参数不足有关。此外, RF 的泛化能力优于 CNN, 原因在于 RF 实现了样本随机化和特征随机化, 使 RF 算法具有良好的泛化能力。最后, 训练样本的数据分布不均匀也可能导致泛化能力下降。

3.2　泛化模型的训练和验证

第 1.1 节阐述了风速和 SWH 之间的关系, 风速和 SWH 二阶相关性 R^2=0.425 6。考虑到增加参数可以提高模型的泛化能力, 因此, 在 17 参数反演模型的基础上, 增加了风速作为外部辅助参数。风速数据集也来源于 ECMWF 数据集。在对风速数据集进行插值和匹配后, 将其作为特征参数合并到训练集中。

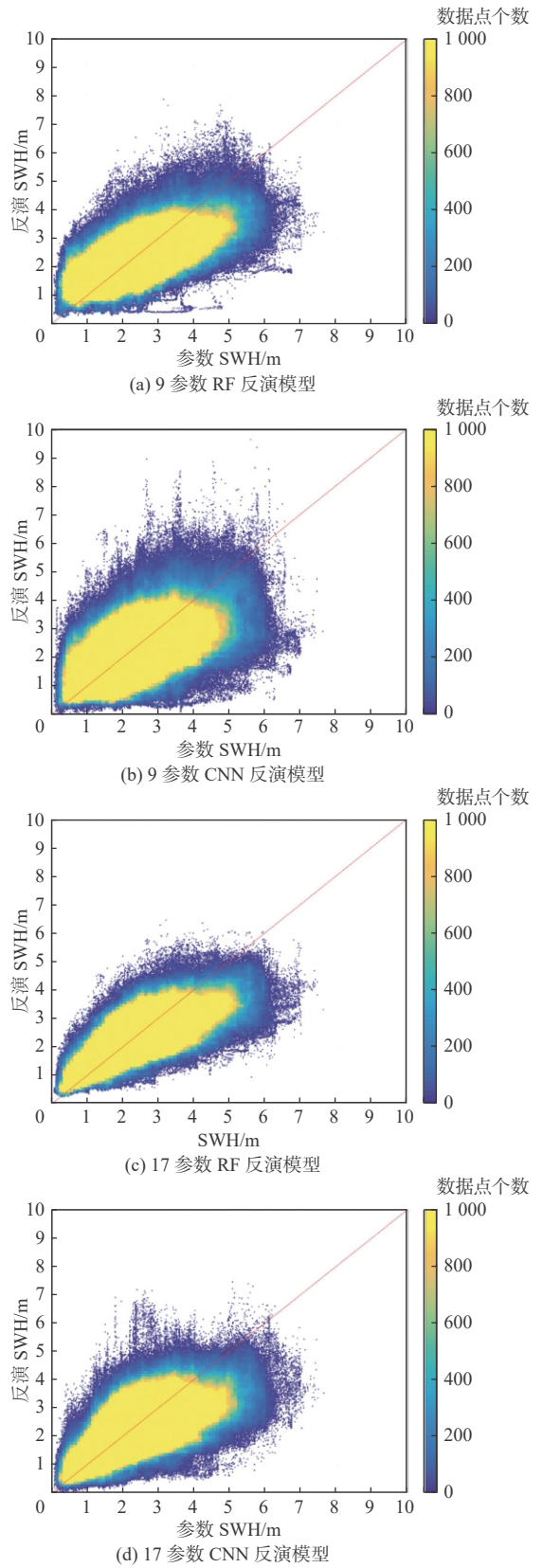


图 10　不同反演模型泛化结果的相关散点图

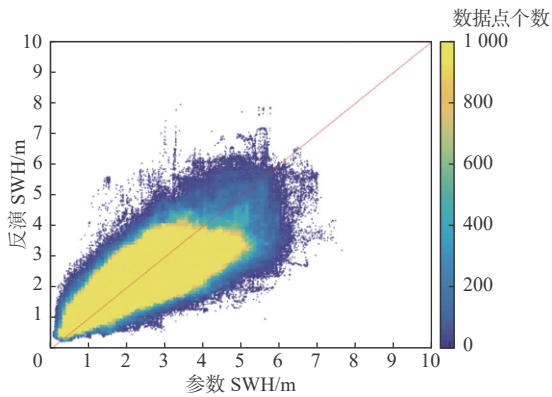
Fig. 10　Scatter plots of generalization results related to different retrieval models

17+1 参数模型的泛化结果如图 11 和表 4 所示。其中, 17+1 参数 RF 泛化模型精度最高, RMSE 为 0.497 1 m, R^2=0.584 6。

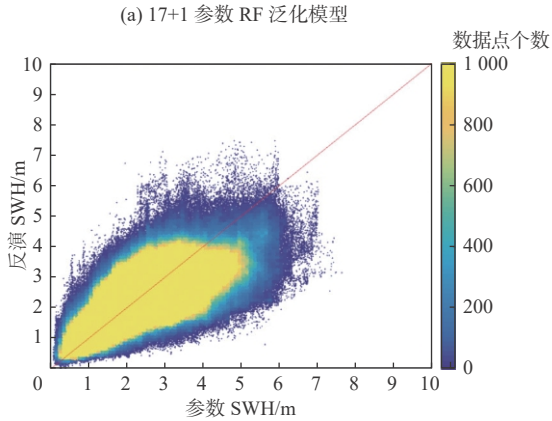
表3 SWH反演模型的泛化性能分析

Table 3 Generalization performance analysis of SWH retrieval models

反演模型	RMSE/m	R^2
9参数RF反演模型	0.578 4	0.437 5
9参数CNN反演模型	0.687 1	0.206 4
17参数RF反演模型	0.517 5	0.549 8
17参数CNN反演模型	0.599 0	0.396 9



(a) 17+1 参数 RF 泛化模型



(b) 17+1 参数 CNN 泛化模型

图11 不同泛化模型泛化结果的相关散点图

Fig. 11 Scatter plots of generalization results related to different generalization models

表4 SWH泛化模型的泛化性能分析

Table 4 Generalization performance analysis of SWH generalization models

泛化模型	RMSE/m	R^2
17+1参数RF泛化模型	0.497 1	0.584 6
17+1参数CNN泛化模型	0.561 7	0.469 6
文献[27]	0.643 0	0.370 0

此外，可以看到，17+1 参数 RF 泛化模型在 2 m 的 SWH 附近表现最好，但在 SWH 大于 4 m 的区域拟合效果不足。为探索泛化模型在不同 SWH 区间下的性能，以 17+1 参数 RF 泛化模型结果为例，将 SWH 分为 3 个区间：0～1 m，1～2 m、2～3 m、3～4 m、4～5 m 和>5 m，表 5 为每个区间 SWH 的泛化性能评估表。从表中可以看出，反演模型中的最佳学习效果在 1～3 m 区间，与表 5 中的数据分布一致。由图 11 和表 5 可知：

1）大部分训练集的 SWH 值集中在 1～3 m 的区间内，因此，泛化模型对这一区间的 SWH 学习得很好。

2）在 0～1 m 和 3～4 m 这 2 个区间中，尽管训练集的数量不如 1～3 m 的区间，但这 2 个区间仍然保持了较多的训练集数量，泛化模型在这 2 个区间仍然保持了较好的学习能力。

3）在大于 4 m 的区间中，训练集的数量非常少，使得泛化模型在该区间学习不足，导致较难准确获得 SWH。

4）加入风速作为辅助参数后，在大于 4 m 的区间中精度有所提高，原因在于随着风速的增大，SWH 也随之增大，且二者相关性不断变强，即在该区间 SWH 的主要驱动力为风速。

5）对比传统拟合模型，本文提出的机器学习模型在全球 SWH 泛化精度上存在优势。

表5 SWH不同区间的泛化精度评估

Table 5 Evaluation of generalization accuracy for different intervals of SWH

区间/m	RMSE/m	训练集数量	测试集数量
0～1	0.433 3	96 626	465 909
1～2	0.392 0	1 032 123	4 972 700
2～3	0.409 0	1 042 745	5 338 182
3～4	0.772 7	262 176	1 146 376
4～5	1.238 6	47 533	267 787
>5	1.659 9	18 797	69 085

4 结 论

为解决星载条件下高动态及复杂海况导致 GNSS-R 简单模型无法精确反演 SWH 的问题，本文提出一种基于多变量机器学习的 CYGNSS 有效波高反演模型：基于海浪的形成机理及相关性结果分析得到了参数，并设计了 5 参数、9 参数和 17 参数 3 种训练方案。利用机器学习 RF 和 CNN 得到反演模型，并且对模型进行反演和泛化评估。为进一步提高模型泛化能力，提出了 17+1 参数泛化模型，证明了本文模型的可行性，结论如下：

1）最佳的反演精度模型是 17 参数 CNN 反演模型，RMSE 为 0.184 0 m，R^2=0.948 5。5 参数反演模型结果不佳，说明仅使用皮尔逊相关结果作为参数选择标准是不够的。9 参数反演模型与 17 参数反演模型相比，减少了 24% 的训练时间，精度仅损失 7.5%。

2）17 参数反演模型泛化精度相较 9 参数反演

模型差距较大，与 9 参数反演模型特征参数不足有关。为进一步提升模型泛化能力，将风速作为辅助参数加入得到泛化模型，最佳泛化精度模型是 17+1 参数 RF 泛化模型，RMSE 为 0.497 1 m，R^2= 0.584 6，精度显著提高。RF 实现了随机样本和特征的随机性，因此，RF 算法相较于 CNN 的泛化效果更好。

由于陆地污染和近岸浅水的影响，卫星雷达高度计数据在近岸误差较大，因此，本文排除了离岸 25 km 以内的数据。本文已经证明了添加风速作为辅助参数来提高 SWH 泛化精度的有效性，但添加真实风速数据的会局限于外部参数的获取，不利于回归计算，但使用 DDM 也可以反演风速。此外，实验证明，泛化结果会受到训练数据样本分布的影响，因此，从 DDM 中提取更多的相关参数作为输入及如何使训练样本均衡将是未来研究的重点。

致谢　感谢 NASA 网站提供的 CYGNSS 数据，感谢北京航空航天大学杨东凯教授和王峰博士在 GNSS-R 星载数据分析中提出的建议，感谢上海航天电子技术研究所高级工程师周勃和秦瑾对 GNSS-R 信号的建议。

参考文献（References）

[1] FOTI G, GOMMENGINGER C, JALES P, et al. Spaceborne GNSS reflectometry for ocean winds: first results from the UK TechDemoSa-1 mission[J]. Geophysical Research Letters, 2020, 42: 5435-5441.

[2] CLARIZIA M P, RUF C S. Wind speed retrieval algorithm for the cyclone global navigation satellite system (CYGNSS) mission[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4419-4432.

[3] ZHANG Y, CHEN X, MENG W T, et al. Wind direction retrieval using support vector machine from CYGNSS sea surface data[J]. Remote Sensing, 2021, 13(21): 4451.

[4] CARDELLACH E, LI W Q, RIUS A, et al. First precise spaceborne sea surface altimetry with GNSS reflected signals[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 13: 102-112.

[5] LI W Q, CARDELLACH E, FABRA F, et al. Assessment of spaceborne GNSS-R ocean altimetry performance using CYGNSS mission raw data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(1): 238-250.

[6] ZHANG Y, LI B B, TIAN L M, et al. Phase altimetry using reflected signals from BeiDou GEO satellites[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(10): 1410-1414.

[7] ZHANG Y, TIAN L M, MENG W T, et al. Feasibility of code-level altimetry using coastal BeiDou reflection (BeiDou-R) setups[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(8): 4130-4140.

[8] ZHANG Y, HUANG S, HAN Y L, et al. Machine learning methods for spaceborne GNSS-R sea surface height measurement from TDS-1[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 15: 1079-1088.

[9] LI Y W, CHANG X, YU K G, et al. Estimation of snow depth using pseudorange and carrier phase observations of GNSS single-frequency signal[J]. GPS Solutions, 2019, 23(4): 118.

[10] YU K G. Theory and practice of GNSS reflectometry[M]. Berlin: Springer, 2021: 1-376.

[11] ZHANG Y, MA D H, MENG W T, et al. Antarctic sea ice detection using a shipborne GPS reflectometry setup[J]. Measurement and Control, 2021, 54(5-6): 618-626.

[12] YAN Q Y, HUANG W M. Spaceborne GNSS-R sea ice detection using delay-Doppler maps: first results from the U. K. TechDemoSat-1 mission[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016, 9(10): 4795-4801.

[13] ZHANG Y, MENG W T, GU Q M, et al. Detection of Bohai bay sea ice using GPS-reflected signals[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 8(1): 39-46.

[14] ZHANG Y, HANG S J, HAN Y L, et al. Sea ice thickness detection using coastal BeiDou reflection setup in Bohai Bay[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(3): 381-385.

[15] CHANG X, JIN T Y, YU K G, et al. Soil moisture estimation by GNSS multipath signal[J]. Remote Sensing, 2019, 11(21): 2559.

[16] YAN Q Y, HUANG W M, JIN S G, et al. Pan-tropical soil moisture mapping based on a three-layer model from CYGNSS GNSS-R data[J]. Remote Sensing of Environment, 2020, 247: 111944.

[17] GERLEIN-SAFDI C, RUF C S. A CYGNSS-based algorithm for the detection of inland waterbodies[J]. Geophysical Research Letters, 2019, 46(21): 12065-12072.

[18] AL-KHALDI M M, JOHNSON J T, GLEASON S, et al. Inland water body mapping using CYGNSS coherence detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(9): 7385-7394.

[19] MARTÍN-NEIRA M. A pasive reflectometry and interferometry system (PARIS) application to ocean altimetry[J]. ESA Journal, 1993, 17(4): 331-355.

[20] SOULAT F, CAPARRINI M, GERMAIN O, et al. Sea state monitoring using coastal GNSS-R[J]. Journal of Geophysical Research, 2004, 31(21): 7385-7394.

[21] ALPERS W, HASSELMANN K. Spectral signal to clutter and thermal noise properties of ocean wave imaging synthetic aperture radars[J]. International Journal of Remote Sensing, 1982, 3(4): 423-446.

[22] 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.

[23] QIN L Y, LI Y. Significant wave height estimation using multi-satellite observations from GNSS-R[J]. Remote Sensing, 2021, 13(23): 4806.

[24] CLARIZIA M P, GOMMENGINGER C P, GLEASON S T, et al. Analysis of GNSS-R delay-Doppler maps from the UK-DMC satellite over the ocean[J]. Geophysical Research Letters, 2009, 36(2):

L02608.

[25] PENG Q, JIN S G. Significant wave height estimation from spaceborne cyclone-GNSS reflectometry[J]. Remote Sensing, 2019, 11(5): 584.

[26] YANG S, JIN S G, JIA Y, et al. Significant wave height estimation from joint CYGNSS DDMA and LES observations[J]. Sensors, 2021, 21(18): 6123.

[27] BU J W, YU K G. Significant wave height retrieval method based on spaceborne GNSS reflectometry[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1503705.

[28] WANG F, YANG D K, YANG L. Retrieval and assessment of significant wave height from CYGNSS mission using neural network[J]. Remote Sensing, 2022, 14(15): 3666.

[29] LV H Q. Fundamentals of physical oceanography[M]. San Jose : Ocean Press, 2012: 178.

[30] PIERSON W J, MOSKOWITZ L. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii[J]. Journal of Geophysical Research, 1964, 69(24): 5181-5190.

[31] CLARIZIA M P, RUF C S, JALES P, et al. Spaceborne GNSS-R minimum variance wind speed estimator[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(11): 6829-6843.

[32] LIN W M, PORTABELLA M, FOTI G, et al. Toward the generation of a wind geophysical model function for spaceborne GNSS-R[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2): 655-666.

[33] BROWN G. Backscattering from a Gaussian-distributed perfectly conducting rough surface[J]. IEEE Transactions on Antennas and Propagation, 1978, 26(3): 472-482.

[34] ZHANG Y, YIN J W, YANG S H, et al. High wind speed inversion model of CYGNSS sea surface data based on machine learning[J]. Remote Sensing, 2021, 13(16): 3324.

[35] LI X H. Application of random forest model in classification and regression analysis[J]. Chinese Journal of Applied Entomology, 2013, 4: 1190-1197.

[36] GUO W F, DU H, GUO C, et al. Information fusion for GNSS-R wind speed retrieval using statistically modified convolutional neural network[J]. Remote Sensing of Environment, 2022, 272: 112934.

[37] CLARIZIA M P, RUF C S. Statistical derivation of wind speeds from CYGNSS data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(6): 3955-3964.

[38] ZHANG Y, LU Q, JIN Q , et al. Global sea surface height measurement from CYGNSS based on machine learning[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2022, 16: 841-852.

Significant wave height retrieval model of CYGNSS based on multivariate machine learning

ZHANG Yun[1]，XIAO Sheng[1]，JIANG Lifei[2]，MENG Wanting[2]，YANG Shuhu[1,*]，HAN Yanling[1]

(1. Shanghai Engineering Research Center of Marine Intelligent Information and Navigation Remote Sensing，College of Information Technology，Shanghai Ocean University，Shanghai 201306，China；
2. Shanghai Aerospace Electronic Technology Institute，Shanghai 201109，China)

Abstract: The cyclone global navigation satellite system (CYGNSS) provides high-quality global navigation satellite system reflectometry (GNSS-R) data that can be reliably used for the retrieval of significant wave height (SWH). Due to the high dynamic nature of CYGNSS, the received signal is easily affected by environmental factors, and the complexity of sea conditions makes it difficult for simple models to accurately retrieve SWH. To address the above issues, this article proposed an SWH retrieval model based on multivariate machine learning. According to the mechanism of wave formation and the analysis of the correlation between CYGNSS parameters and SWH, relevant parameters were selected, and three training schemes were designed, involving five parameters, nine parameters, and 17 parameters, respectively. Random forest (RF) and convolutional neural network (CNN) were used to train and verify the retrieval model, and the SWH retrieval results were compared with the reference values of the European Centre for Medium-Range Weather Forecasts (ECMWF). The best retrieval model among them was the 17-parameter CNN retrieval model, with root mean square error(RMSE)was 0.184 0 m and R^2= 0.948 5. Compared with the 17-parameter CNN retrieval model, the 9-parameter CNN retrieval model reduced training time by 24% and has minimal accuracy loss. However, the 9-parameter retrieval model performed poorly in terms of generalization evaluation compared to the 17-parameter retrieval model. To improve the generalization ability of the model, wind speed was added as a parameter to the 17-parameter retrieval model, resulting in a 17 + 1-parameter generalization model. The best generalization model among them was the 17 + 1 parameter RF generalization model, with RMSE was 0.497 1 m and R^2= 0.584 6. This effectively proves that the model proposed in this article has good potential in SWH retrieval.

Keywords: significant wave height; CYGNSS; satellite-based GNSS-R; machine learning; convolutional neural network

Received：2023-05-23；**Accepted**：2023-08-04；**Published Online**：2023-08-30 13：37

URL：link.cnki.net/urlid/11.2625.V.20230829.1516.001

Foundation items：National Natural Science Foundation of China (42176175,42271335); National Key Research and Development Program of China (2019YFD0900805)

* **Corresponding author**. E-mail：shyang@shou.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0269

嵌入自注意力机制的U型高分遥感影像语义分割网络

杨军[1, 2, *]，张金影[2]

(1. 兰州交通大学 电子与信息工程学院，兰州 730070；2. 兰州交通大学 测绘与地理信息学院，兰州 730070)

摘 要：针对高分辨率遥感影像小目标特征提取困难的问题，结合卷积结构和自注意力机制提出一种双编码器特征融合网络模型，适用于高分辨率遥感影像语义分割。设计一种双编码器结构，同时提取遥感影像的全局和局部细节信息，提高小目标地物的分割准确率；通过特征聚合模块在不同阶段聚合特征信息，以嵌入更多全局上下文信息；利用边缘细化损失模块提高模型对地物目标边缘信息的识别能力。在ISPRS Vaihingen和Potsdam数据集上的分割结果的F_1平均值m_{F_1}分别达到了91.28%和93.16%，与目前主流算法相比，总体分割准确率及汽车类小目标的分割准确率均有提高。所提模型在一定程度上解决了高分辨率遥感影像语义分割中小目标及边缘信息分割不准确的问题。

关 键 词：高分辨率遥感影像；语义分割；深度学习；自注意力机制；特征融合

中图分类号：P237

文献标志码：A **文章编号**：1001-5965(2025)05-1514-14

遥感应用人工智能技术解决影像解译、信息自动提取问题[1]，受到研究人员的广泛关注，其核心任务是对影像中的每个像素点进行类别归属，在图像分割、农业生产、土地覆盖、地物分类等领域发挥了重要作用[2-5]。遥感影像中包含大量地物信息，复杂的城市构造也会导致目标物体之间产生遮挡、阴影等噪声效应，并且遥感影像比自然图像更容易产生物体变形，从而使获取准确的影像信息变得更加困难。此外，由于遥感影像获取过程具有不稳定性，影像中目标物体的边缘信息和小目标物体更容易受到影响，因此，如何准确地捕获小目标物体和地物边缘信息在语义分割任务中至关重要。近年来，深度学习方法在计算机视觉任务中展现了良好的性能，由于其具有高度的泛化能力和快速推理能力，可以有效地提取出具有鲁棒性的特征信息[6]，逐渐被应用在遥感影像分析和处理任务中，如目标检测、实例分割和语义分割等。

卷积神经网络(convolutional neural networks, CNN)的快速发展为图像语义分割提供了技术支撑，与基于机器学习的传统分割方法相比，基于CNN的方法可以更高效地捕获细节局部信息。全卷积网络(fully convolutional networks, FCN)[7]以端到端的方式实现了像素级分类，并进一步促进了深度学习模型在图像分割领域的应用，但是其设计的解码器较简单，没有考虑像素之间的关系，预测结果较为粗糙。U-Net[8]使用编码器-解码器结构和跳跃连接展现出了优异的分割性能，在编码器中使用一系列的卷积层和连续下采样层来提取深层特征，然后，利用解码器融合高级语义信息和低级空间信息以恢复空间分辨率，并将来自编码器的不同尺度的高分辨率特征通过跳跃连接进行融合，以减轻下采样导致的空间信息损失。然而，遥感影像地物信息的复杂性给语义分割带来了新的挑战，具有不同语义类别的地物可能具有相似的光谱特征，难以进行区

收稿日期：2023-05-23；**录用日期**：2023-10-26；**网络出版时间**：2023-11-08 11:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20231107.1027.003
基金项目：国家自然科学基金(42261067)；2025年度甘肃省重点人才项目(2025RCXM031)；兰州市人才创新创业项目(2020-RC-22)
* **通信作者**. E-mail：yangj@mail.lzjtu.cn

引用格式：杨军，张金影. 嵌入自注意力机制的U型高分遥感影像语义分割网络[J]. 北京航空航天大学学报，2025，51(5)：1514-1527.
YANG J，ZHANG J Y. U-shaped semantic segmentation network of high-resolution remote sensing images embedded with self-attention mechanism[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1514-1527 (in Chinese).

分, 并且地物之间的遮挡问题通常会导致语义模糊, 因此, 直接融合从浅层和深层提取的特征会产生噪声并削弱模型的性能。

为解决这个问题, Li 等[9] 引入了级联的空间注意力残差块来捕获长程依赖, 并引入通道注意力层来优化多尺度特征融合, 从遥感影像中提取边界以细化道路。文献 [10] 以改进的残差模型为基础, 引入坐标卷积和全局信息增强模块, 用于增强空间信息和全局上下文信息的感知能力, 提高遥感影像道路分类的一致性, 但是仍然存在由于地物遮挡而漏检道路的现象。Li 等[11] 为细化高级语义信息和低级空间特征图, 设计了深层通道注意力模块和浅层空间注意力模块, 缓解了遥感影像分割中地物类别不平衡的问题。文献 [12] 提出了基于局部注意力的轻量级模块来增强上下文信息的嵌入, 使用注意力嵌入模块耦合遥感影像中的高级语义特征与低级特征信息。上述方法利用注意力机制处理高分辨率遥感影像, 减小了模型在特征提取过程中产生的类内方差, 但是此类方法不是直接编码全局上下文信息, 而是从卷积层获取的局部特征中聚合上下文信息。CNN 在空间位置信息提取方面具有优势, 但是由于卷积运算的局部性, 基于 CNN 的方法很难直接学习到全局上下文信息, 在建模长程依赖关系方面仍有局限性, 因此, 很难从具有复杂背景的遥感影像中获得准确的全局信息, 并且基于 CNN 的模型在特征提取过程中执行特征下采样以减少计算量, 容易导致小目标特征信息的丢失及目标边界模糊问题。

受 Transformer[13] 在自然语言处理 (natural language processing, NLP) 领域取得巨大成功的启发, 文献 [14] 提出的视觉 Transformer(vision Transformer, ViT) 将 Transformer 引入视觉领域执行图像处理任务。Sun 等[15] 提出混合多分辨率和 Transformer 的语义提取网络, 每个像素都能获得全局感受野, 克服了 CNN 感受野较小的局限性, 并且可以增强网络对于遥感影像场景的理解能力。Swin Transformer[16] 提出在局部窗口内计算自注意力的滑窗法, 该结构采用分层 Transformer 捕获不同尺度的特征, 有利于获取空间信息, 并使用移位窗口多头自注意力 (shifted windows multi-head self-attention, SW-MSA) 机制进行全局关系建模, 在图像分类、目标检测和语义分割等各种视觉任务上实现了最先进的性能。

受 U-Net 和 Swin Transformer 网络的启发, 本文针对高分辨率遥感影像小目标特征提取困难及目标边缘分割不准确的问题, 提出结合卷积和自注意力机制的 U 型网络 (convolution and self-attention mechanism UNet, CSUNet) 模型, 实现了能够细化物体边界、减少细节信息丢失的遥感影像语义分割算法。主要创新点有:

1) 基于 Swin Transformer 和卷积结构设计了一种双编码器结构, 可以同时提取遥感影像的全局上下文信息和局部细节信息, 从而提高小目标对象的分割效果。

2) 提出特征聚合模块 (feature aggregation module, FAM), 分别在不同的阶段聚合特征图中的信息, 以嵌入更多的全局上下文特征信息。

3) 设计了边缘细化损失 (edge thinning loss, ETL) 模块, 直接从像素级和图像级的分割预测结果中细化物体边界, 提高本文模型对目标边缘信息的识别能力。

1　相关研究工作

随着遥感传感器技术的快速发展, 高分辨率遥感影像的数据获取量越来越大, 深入研究这些包含丰富空间信息和局部细节特征的高分辨率遥感影像具有重要意义。然而, 高分辨率遥感影像复杂的背景信息、光谱信息和目标结构的多样性给语义分割任务带来了巨大挑战。传统的遥感影像语义分割方法依赖于图像的纹理信息和光谱信息[17-19], 需要投入大量的人力物力。近年来, 深度学习方法的出现为计算机视觉领域开拓了新的研究方向, 与基于机器学习的传统分割方法相比, 基于 CNN 的方法可以捕获图像中更多细粒度的局部信息, 在遥感影像分析和解译中得到了广泛应用。

ResUNet-a[20] 使用 U-Net 编码器-解码器结构, 结合残差连接、空洞卷积、金字塔场景解析池化和多任务学习来处理遥感影像。AFNet[21] 提出多路径注意力融合模块融合多路径特征, 利用细化注意力融合模块融合高层抽象特征和低层空间特征, 解决了遥感影像中低层特征提取不足和高层特征信息丢失严重的问题。Pan 等[22] 提出一种基于条件生成对抗网络的训练样本集改进模型, 用于高分辨率遥感影像的语义分割, 该网络从先前的训练图像中提取高级空间信息的同时主动生成新的样本图像, 改善样本数据的空间信息多样性。然而, 这些模型对遥感影像的背景信息关注不足, 并且对小目标特征的提取效果不佳。FactSeg[23] 提出对称的双分支解码器结构, 由前景激活分支和语义细化分支组成, 通过跳跃连接进行多尺度特征融合, 解决小目标对象和大规模背景之间的样本不平衡问题, 提高了小目标对象分割的准确性。杨军和于茜子[24] 设计了一种结合 Atrous-CNN 的 FuseNet 变体网络

(improved FuseNet with Atrous-CNN, IFA-CNN), 引入空洞卷积调整感受野捕获多尺度信息, 并且提出虚拟融合单元提高了遥感影像语义分割准确率, 但是对于高分辨率遥感影像的边缘还存在分割不准确的情况。Chen 等[25] 提出一种自适应感受野卷积网络, 降低模型的训练难度并避免了过拟合问题, 使网络可以关注到更多的细节信息。此外, 对于地物边界的分割提取也是一个值得关注的方向。文献 [26] 在其编码器和解码器中提出 2 个边缘损失增强模块, 用于中间特征层的边缘细化, 但是网络同时关注边缘和非边缘像素, 这容易产生过度平滑的边缘预测结果, 并且存在计算量大、模型运行速度较慢的问题。HRCNet[27] 设计了轻量级双注意力模块在特征提取阶段获得上下文信息, 并使用特征增强的金字塔结构融合不同尺度的上下文信息, 同时结合边界感知损失改善了对小目标对象的边缘分割效果。GCNet[28] 设计了轻量级的全局上下文模块对长程依赖关系进行建模, 有效减少了计算资源的消耗。上述基于 CNN 的模型在一定程度上解决了空间信息提取、边缘提取、轻量化设计等问题, 但是对于图像全局信息的提取能力仍有待提高。研究人员尝试对全局上下文信息进行建模, 但基于卷积的注意力模块过度依赖卷积运算, 限制了全局特征的提取与表示, 并且单个注意力模块无法在解码器中以多层次语义特征对全局信息进行建模。

Transformer 被提出用于机器翻译任务, 并超越了先前基于复杂递归或神经网络的序列转换模型。受 Transformer 结构在自然语言处理领域取得成功的启发, 研究人员开始尝试将其应用到计算机视觉任务处理中提取全局信息。ViT[14] 是较早将 Transformer 应用于图像分类的方法, 通过分割和扁平化将图像数据转换为一系列空间位置向量, 在大规模数据集的预训练中实现了先进的性能, 但是应用于密集预测任务时需要较大的训练成本。Deformable DETR[29] 将固定形状的卷积过程改进成能够适应物体形状的多尺度可变形注意力模块, 增强网络适应物体形变的能力、加快模型的收敛速度, 并且不需要额外结合特征金字塔 (feature pyramid networks, FPN)[30] 就可以聚合多尺度特征图, 在一定程度上降低了计算复杂度。但会造成无序的内存访问, 因此, 该方法的运行速度仍然比传统卷积稍慢。受此启发, 许多研究人员已经将 Transformer 应用于遥感影像。Wang 等[31] 将 Transformer 架构引入到高分辨率遥感影像的语义分割中, 利用编码器-解码器架构来捕获多尺度特征图, 并设计了密集连接的 FAM 来恢复特征图分辨率, 通过捕获多尺度的语义特征信息提高了遥感影像语义分割准确率。文献 [32] 提出一种双边感知网络, 该网络包含长程依赖分支和纹理分支, 以充分捕捉高分辨率遥感影像中的全局上下文关系和细节特征, 并利用线性注意力机制聚合特征, 有效地融合长程依赖性特征和纹理特征。Gao 等[33] 设计了一个自适应融合模块以充分利用在不同阶段获得的特征信息, 通过融合不同尺度的语义信息获得准确的特征表示。He 等[34] 提出一种空间交互模块, 有效增强被遮挡目标的特征表示能力, 并构造特征压缩模块减少信息丢失。文献 [33] 和文献 [34] 都是通过结合 Transformer 和卷积设计编码器结构提高模型分割性能。虽然这些 Transformer 架构建立了全局关系, 也应用了轻量化设计, 但对于边缘特征的提取能力不足, 而遥感影像的真实标签中存在不规则的边缘特征, 因此, 如何准确识别和提取这些边缘特征是亟待解决的难题。

综上所述, 目前遥感影像语义分割方法主要存在 2 个问题: ①传统基于卷积结构的方法缺乏对全局上下文信息的利用, 不能有效区分目标类别。②Transformer 在建立长程依赖关系方面具有优势, 但是直接应用在遥感影像中会破坏固有结构, 丢失细节和边缘信息。由 CNN 或 Transformer 单独构成的遥感影像分割模型会导致对于局部性或全局性特征捕获能力不足。本文结合卷积结构和自注意力机制提出 CSUNet 模型, 其将卷积在提取底层特征方面的优势和 Swin Transformer 在多尺度建模长程依赖关系方面的优势进行结合, 丰富遥感影像中的局部和全局语义信息, 以解决高分辨率遥感影像语义分割中存在的问题。

2 综合卷积和自注意力机制的 U 型网络

本文 CSUNet 模型结合卷积和 Swin Transformer 构建了双编码器结构, 可以同时捕获整体和细节信息。在编码器中提出 FAM, 融合局部特征和全局表示, 增强了网络特征表示能力; 在解码器中增加 ETL 模块提高模型对目标边缘信息的提取能力, 生成高分辨率遥感影像的语义分割结果。

2.1 CSUNet 模型框架

遥感影像中的语义和空间上下文信息至关重要, CNN 可以捕获细粒度的局部特征, 但受到卷积核固定大小的限制, 难以有效建模影像中的长程依赖关系。而 Swin Transformer 可以通过自注意力计算获得全局语义信息, 但需要将图像扩展为一维的

图像块序列, 这种操作可能会造成图像块内部空间位置信息的丢失。为此, 本文提出 CSUNet 模型, 可以有效地获取遥感影像中的全局上下文信息和空间语义信息。同时使用 Swin Transformer 和卷积结构来处理图像, 并在不同的阶段融合特征图。

与传统的多头自注意模块不同, Swin Transformer 模块是基于移位窗口构造而成的。如图 1 所示, 每个 Swin Transformer 模块由层归一化 (layer norm, LN)、多头自注意力模块、残差连接和具有非线性激活函数的多层感知机 (multilayer perceptron, MLP) 组成。窗口多头自注意力 (window multi-head self-attention, W-MSA) 模块无法在不同窗口间进行信息的传递, 这会影响 Transformer 体系结构对长程依赖建模的能力, 而 SW-MSA 解决了这一问题, 通过将窗口分别向右、向下各偏移 2 个图像块来实现跨窗口的信息传递, 可用来计算图像的全局注意力。式(1)～式(4)表示图像特征传递过程:

$$\hat{z}^{l}=\text{W-MSA}\left(\text{LN}\left(z^{l-1}\right)\right)+z^{l-1} \tag{1}$$

$$z^{l}=\text{MLP}\left(\text{LN}\left(\hat{z}^{l}\right)\right)+\hat{z}^{l} \tag{2}$$

$$\hat{z}^{l+1}=\text{SW-MSA}\left(\text{LN}\left(z^{l}\right)\right)+z^{l} \tag{3}$$

$$z^{l+1}=\text{MLP}\left(\text{LN}\left(\hat{z}^{l+1}\right)\right)+\hat{z}^{l+1} \tag{4}$$

式中: z^{l-1}为 W-MSA 模块的输入特征; $\hat{z}^{l}$为 W-MSA 模块的输出特征, z^{l}和z^{l+1}表示 MLP 模块的输出, $\hat{z}^{l+1}$表示 SW-MSA 模块的输出特征。

Swin Transformer 网络中最主要的就是移位窗口思想, W-MSA 可以减小模型计算量, 但是窗口之间无法进行信息交互, 而基于移位窗口的方法可以解决这个问题。如图 2 所示, 假设窗口大小为 L_{W}, 移位窗口就是将窗口分别向右、向下各偏移 $L_{\text{W}}/2$ 个像素, 得到 9 个窗口之后开始进行移动合并, 将这 9 个窗口整合为和输入一样的 4 个 4×4 窗口, 保证变换前后的计算量是相同的。用 *A*、*B*、*C* 标识移动前后区域变化位置, 便于对比。同时, 为防止窗口移动导致的信息混淆问题, SW-MSA 在实际计算中使用的是 masked MSA, 即带掩码的多头自注意力机制, 可以通过设置掩码隔绝不同区域的信息, 将其和自注意力矩阵进行运算以去除不需要的信息。最后, 对计算完成得到的注意力执行还原操作, 保证和原来的图片相对位置是不变的, 整体的语义信息也没有发生改变。

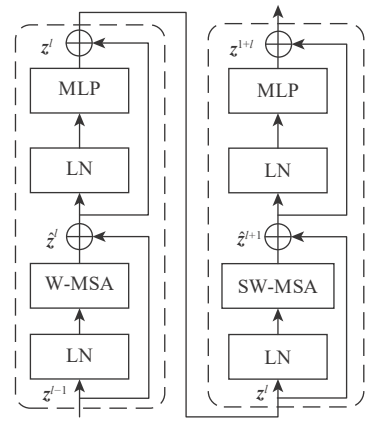


图 1　2 个连续的 Swin Transformer 模块[16]

Fig. 1　Two successive Swin Transformer modules [16]

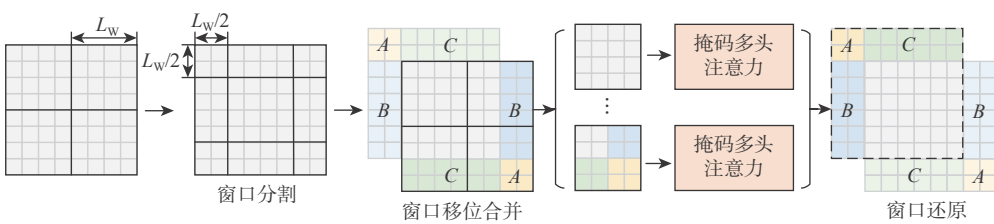


图 2　移位窗口的主要步骤[16]

Fig. 2　Main steps of shifted windows[16]

如图 3 所示, 本文 CSUNet 模型主体是基于 U-Net 的编码器-解码器结构和跳跃连接, 基本单位是 Swin Transformer 模块。特别地, CSUNet 模型构建了基于卷积结构和 Swin Transformer 模块组成的双编码器, 通过 FAM 聚合特征信息, 以充分获得遥感影像的地物信息。将图像 $\boldsymbol{X}\in\mathbf{R}^{H\times W\times C}$ 分别输入卷积网络和 Swin Transformer 结构, 其中, *H* 为图像的高度, *W* 为图像的宽度, *C* 为图像中的通道数, n_{class} 为语义类别数。在卷积结构中, 每个子模块中都包含卷积层、批归一化 (batch normalization, BN) 层、修正线性单元 (rectified linear unit, ReLU) 激活函数、池化模块及用来统一特征通道的 1×1 的卷积操作, 其中, 批归一化层可以加速模型收敛, 防止梯度爆炸和梯度消失, 修正线性单元可以缓解过拟合问题。在自注意力结构中, 为将输入转换为序列嵌入, 编码器通过图像块分割操作将输入图像分割为非重叠块, 线性嵌入层用于将特征维数投影到任意维度。转换后的图像块通过 Swin Transformer 模块和图像块合并层生成分层特征表示, 并使用 2 个连续的 Swin Transformer 模块来构造瓶颈以学习深层

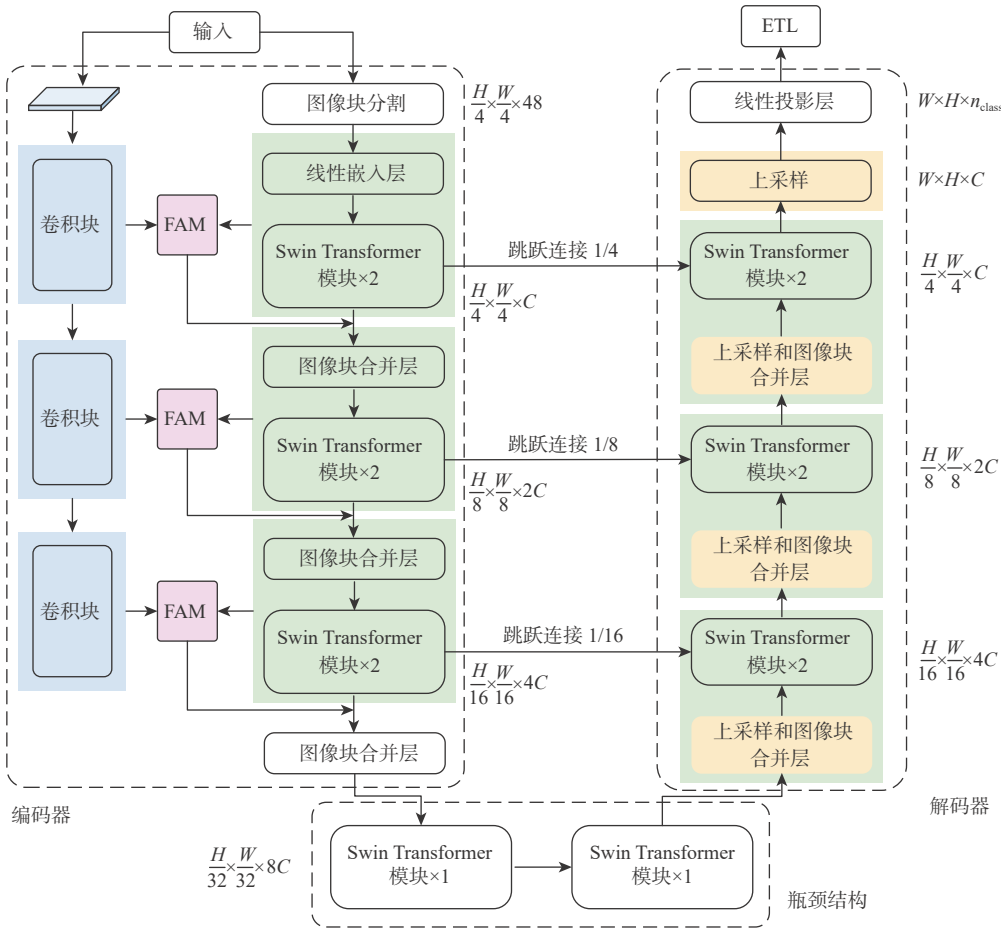


图 3 本文模型网络结构

Fig. 3 Network structure of the proposed model

特征。模型在不同阶段获得不同大小的特征图，可以有效捕获多尺度的特征信息。

为使特征映射能够包含丰富的语义信息和特征细节信息，本文使用分阶段的融合策略，将卷积结构生成的特征图与 Swin Transformer 不同阶段生成的特征图通过 FAM 进行聚合。FAM 在卷积结构和自注意力机制之间建立连接，缓解图像块内部的空间信息丢失问题。受 U-Net 的启发，本文设计了对称的解码器，由 Swin Transformer 模块和上采样操作组成，提取到的上下文特征通过跳跃连接与编码器的多尺度特征融合，以弥补下采样造成的空间信息损失；然后，将聚合后的特征图进行上采样，以将特征图的分辨率恢复到输入分辨率；最后，将这些上采样特征应用线性投影层以输出像素级分割预测。最终的 ETL 模块由一个用于语义分割的标准交叉熵损失函数和一个用于边缘提取的损失函数组成，使用简单的边缘提取器直接从预测的分割图中定位目标边缘。

2.2 特征聚合模块

全局上下文信息和局部细节信息对于遥感影像中复杂场景的语义分割至关重要。为充分利用自注意力编码器获得的全局上下文信息和卷积编码器提取的局部细节信息，本文设计了 FAM，用来对不同图像块的位置信息进行编码，详细结构如图 4 所示。FAM 由全局和局部注意力分支组成，局部分支采用卷积核大小分别为 1 和 3 的并行卷积来提取局部上下文，在级联相加求和操作之前使用批归一化运算加快模型收敛速度。全局分支采用基于窗口的多头自注意力捕获全局上下文，首先，使用标准的 1×1 卷积将输入的二维特征图的通道维数扩大到原始维度的 3 倍；然后，将一维序列分割成查询矩阵 $\boldsymbol{Q}$、键矩阵 $\boldsymbol{K}$ 和值矩阵 $\boldsymbol{V}$。在 $\boldsymbol{Q}$ 和 $\boldsymbol{K}^{\mathrm{T}}$ 之间进行矩阵乘法，使用 softmax 层获得 $\boldsymbol{V}$ 的权重，即注意力矩阵 $\boldsymbol{Y}$，并在原始矩阵和注意力矩阵之间进行矩阵乘法。对相乘结果和原始映射执行逐元素矩阵求和运算，以获得最终表示 $\boldsymbol{Z}$。

具体计算过程：假设输入为 $\boldsymbol{X}\in\mathbf{R}^{H\times W\times C}$，经过 1×1 卷积和窗口分割重塑之后分配给各头部模块，生成 $\boldsymbol{K}\in\mathbf{R}^{H\times W\times C}$、$\boldsymbol{Q}\in\mathbf{R}^{H\times W\times C}$、$\boldsymbol{V}\in\mathbf{R}^{H\times W\times C}$ 以计算自注意力，之后将这几个矩阵重塑为 $\mathbf{R}^{N\times C}$，其中，$N=H\times W$。如式(5)所示，在 $\boldsymbol{Q}$ 和 $\boldsymbol{K}^{\mathrm{T}}$ 之间进行矩阵乘法，并使用 softmax 层获得注意力矩阵 $\boldsymbol{Y}\in\mathbf{R}^{N\times N}$。图

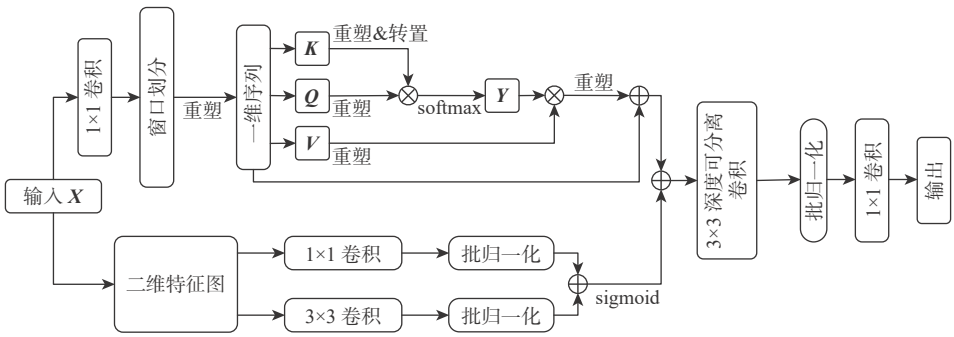


图 4　特征聚合模块

Fig. 4　Feature aggregation module

像块位置 i 对位置 j 的影响 $\boldsymbol{Y}_{ji}$ 为

$$\boldsymbol{Y}_{ji}=\frac{\exp\left(\boldsymbol{Q}_i\cdot\boldsymbol{K}_j^{\mathrm{T}}\right)}{\sum_{i=1}^{N}\exp\left(\boldsymbol{Q}_i\cdot\boldsymbol{K}_j^{\mathrm{T}}\right)} \tag{5}$$

式中: i 为查询向量的位置索引; j 为健向量的位置索引; 计算 $\boldsymbol{Q}_i\cdot\boldsymbol{K}_j^{\mathrm{T}}$ 以提取特征中的长程依赖关系。

注意力机制是通过计算输入特征图中所有位置之间的相似度, 并将相似度作为权重, 也就是模型对于该位置的关注程度, 之后和所有位置的值矩阵进行加权聚合。如式(6)所示, 在 $\boldsymbol{V}$ 和 $\boldsymbol{Y}$ 之间使用矩阵乘法, 然后, 对相乘结果和原始输入执行逐元素矩阵求和运算, 以获得最终表示 $\boldsymbol{Z}\in\mathbf{R}^{H\times W\times C}$。并且在进行求和运算前, 需要将结果重塑为 $\mathbf{R}^{H\times W\times C}$, 保证和原始输入的维度相同, 比例参数 α 的默认值为 0, 通过不断地学习可以为 $\boldsymbol{Y}_{ji}^{\mathrm{T}}\cdot\boldsymbol{V}_i$ 分配更多权重。

$$\boldsymbol{Z}_j=\alpha\sum_{i=1}^{N}\left(\boldsymbol{Y}_{ji}^{\mathrm{T}}\cdot\boldsymbol{V}_i\right)+\boldsymbol{X}_j \tag{6}$$

使用级联操作将从全局分支和局部分支路径中获得的特征图在通道维度上连接在一起, 但是这会导致输出通道数增加。因此, 使用 3×3 的深度卷积 (depthwise convolution, DW Conv)、批归一化操作和标准 1×1 卷积规范最终的全局和局部上下文特征, 将通道数恢复到相应的原始输入大小。

2.3　边缘细化损失模块

遥感影像地物的多样性和复杂性给物体边缘的准确识别带来了挑战, 目标对象边缘过平滑是影响分割质量的一个常见问题, 这会削弱 2 个相邻目标对象之间的边缘, 导致边缘周围的像素容易混淆, 从而难以区分相似的类别。为缓解边缘模糊问题, 本文设计了 ETL 模块, 可以直接从像素级和图像级的分割预测结果中细化物体边界, 提高模型对目标边缘信息的识别能力。如图 5 所示, 其中, L 为总损失函数, ETL 由分割预测边缘图和边缘提取两部分构成, 分割图包含丰富的边缘信息, 在分割预测分支中可以直接用于定位目标边缘。如式(7)所示, 基于像素的交叉熵损失 l_{ce}, 在语义分割结果图中可以提高预测的准确性, 实现像素级的分割预测, 并且收敛速度较快。

$$l_{\mathrm{ce}}=H_x(x')=-\sum_{i=1}^{n}x_i\ln\left(x_i'\right) \tag{7}$$

式中: n 为数据集所含物体的语义类别总数; x_i 为真实标签的第 i 个元素; x_i' 为模型预测 x 属于第 i 个类别的概率, 分类结果越准确, x_i' 越接近 1, 交叉熵损失 $H_x(x')$ 值也越小, 表示模型预测结果越准确。

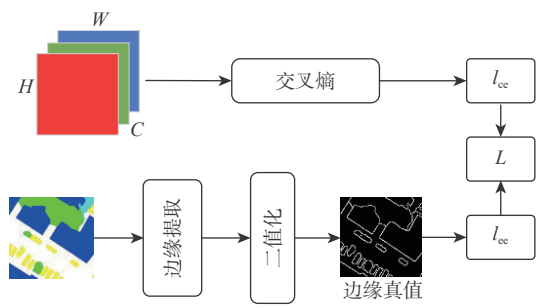


图 5　边缘细化损失模块

Fig. 5　Edge thinning loss module

基于像素的交叉熵损失计算过程中单独处理每个像素, 但没有考虑相邻像素及不同目标对象之间的语义关系, 难以获取到目标对象的整体几何边缘。本文在常用交叉熵损失的基础上提出改进的边缘提取损失, 式(8)和式(9)为边缘提取损失 l_{ee} 的具体计算方法。边缘提取分支使用边缘提取器直接从真实标签中定位目标边缘, 应用滤波器提取符合条件的边缘, 之后使用阈值处理将图像中的高亮边缘提取出来, 并经过二值化处理之后得到更加连续、平滑的边缘真值, 然后, 通过边缘提取损失计算边缘预测误差, 有效增强了边缘周围的特征, 并抑制背景中的噪声, 提高边缘像素分类的准确性。边缘提取损失 l_{ee} 是一种二元分类损失, 包括边缘和非边缘类别。

$$H(x_i)=-p(x_i)\ln p(x_i)-(1-p(x_i))\ln p(x_i) \tag{8}$$

$$l_{ee}=\frac{n_{np}}{n_{ep}+n_{np}}\sum_{i=1}^{n_{ep}}H(x_i)+\frac{n_{ep}}{n_{ep}+n_{np}}\sum_{j=1}^{n_{np}}H\left(y_j\right) \tag{9}$$

式中：$p(x_i)$为x_i像素被分类为边缘像素的概率，$1-p(x_i)$表示x_i像素不属于边缘的概率；$H(x_i)$和$H(y_j)$分别为边缘交叉熵损失和非边缘交叉熵损失；n_{ep}和n_{np}分别为边缘和非边缘像素的数量，其中，n_{np}通常比n_{ep}大得多。因此，边缘部分的像素可以获得更大的损失权重系数，这意味着模型将更加关注边缘部分的学习。

将2条分支的损失函数进行合并计算，得到总的损失函数L：

$$L=l_{ce}+l_{ee} \tag{10}$$

3 实验结果与分析

3.1 数据集与数据增强

实验所使用的数据集是国际摄影测量与遥感学会(International Society for Photogrammetry and Remote Sensing, ISPRS)[35]公开的2个机载高分辨率遥感影像数据集：Potsdam数据集和Vaihingen数据集。

ISPRS Potsdam数据集由38张经过正交校正的航拍图像(6 000像素×6 000像素)组成，其中，24幅影像包含真实标签，覆盖面积为3.42 km^2，空间分辨率为5 cm，具有通过密集图像匹配生成的相应数字表面模型(digital surface models, DSM)，DSM表示树木、建筑物和其他目标地物的高程。每幅航拍影像由4个波段组成，分别为近红外(near infra-red, NIR)、红色、绿色、蓝色，本文使用红色、绿色、蓝色波段。Potsdam是德国的一座历史名城，拥有大型建筑和密集交织的街道，图像标签共分为6类，分别是不透水表面、建筑物、低矮植物、树木、车辆及背景杂物。实验选取其中的24幅图像进行训练，14幅图像进行测试。

ISPRS Vaihingen数据集由33张航拍图像组成，每幅图像有3个波段，分别对应近红外、红色和绿色波段，同样具有相应的DSM和一个归一化数字表面模型(normalization digital surface models, NDSM)数据。这些图像采集于德国Vaihingen 1.38 km^{2}的区域内，空间分辨率为9 cm，像素大小不一，每幅图像的平均大小为2 494像素×2 064像素，标签类别和数量与Potsdam数据集一致。实验选取其中的16幅图像进行训练，17幅图像进行测试。

为适应网络模型结构及硬件环境的限制，对验证集采取滑窗的形式进行256像素×256像素的剪裁处理，但这种操作可能会导致相邻图像块边界的不一致分割，从而导致影像分割准确率降低，为避免剪裁拼接导致的边界模糊现象，将重叠率设置为75%，保留物体的空间结构并提高分割结果的准确率。根据文献[36]可知，随着相邻图像重叠区域大小的增加，网络性能有所提高，重叠大小在75%时分割准确率最高，并且进一步增加重叠大小不会导致性能的显著改善。此外，为提高网络模型的泛化能力并避免过拟合，实验通过水平和垂直翻转、旋转、缩放处理等操作对数据集进行数据增强以扩展数据集，为来自不同环境的图像建立具有高鲁棒性的模型。由文献[37]可知，所谓小目标指的是目标在图像中所占的像素面积小，根据国际光学工程学会(Society of Photo-Optical Instrumentation Engineers, SPIE)的定义，小目标为像素值在256×256的图像中目标面积小于80像素，即像素值小于256×256的0.12%就为小目标，故本文采用数据集中的汽车类地物可作为小目标进行研究。

3.2 实验环境和网络训练

实验环境配置为Intel Core i9-9900K CPU处理器，64 GB内存，NVIDIA GeForce RTX 2080 Ti GPU，11 GB显存，在Ubuntu 18.04.6 LTS操作系统及Python 3.8的环境中运行。网络采用AdamW优化器进行训练以加快模型的收敛速度，学习率初始值设为0.001，动量系数和权重衰减分别设置为0.9和0.01，批归一化尺寸设置为4。

3.3 评价指标

实验采用总体准确率(overall accuracy, OA)、F_1分数和平均交并比(mean intersection over union, mIoU)这3个评价指标对网络分割性能进行评估：

$$O_A=\frac{T_P+T_N}{T_P+F_P+T_N+F_N} \tag{11}$$

$$F_1=\frac{2PR}{P+R} \tag{12}$$

$$m_{IoU}=\frac{1}{n+1}\sum_{i=0}^{n}\frac{T_P}{F_N+F_P+T_P} \tag{13}$$

式中：T_P为真正例，即被模型预测为正类的正样本；F_N为假反例，即被模型预测为负类的正样本；F_P为假正例，即被模型预测为正类的负样本；T_N为真反例，即被模型预测为负类的负样本；$P=T_P/(T_P+F_P)$为预测正确的正样本占整体预测为正样本的比率，即查准率；$R=T_P/(T_P+F_N)$为预测正确的正样本占整体实际为正样本的比率，即查全率。实验中，通过计算F_1分数的平均值m_{F_1}评估网络的分割准确率，表示查准率与查全率的加权调和平均，取值范围从0到1，m_{F_1}的值越大，表示网络越稳健且性能越好；总体准确率O_A越高，表示模型的分割准确率

越高; m_{IoU} 在处理类别不平衡数据时更具优势, 其值越大表示网络分割结果越接近真实分割结果。

3.4　实验结果分析

本文 CSUNet 模型和其他模型在 Vaihingen 数据集上的分割准确率对比结果如表 1 所示, 其中, 沙漏型网络 (hourglass-shaped network, HSN) 的编码器是由主编码器和轻量级分支所构成的。表中加粗显示最优结果, 可以看出, 相较于其他模型, 本文 CSUNet 模型具有较好的表现, 在 m_{F_1} 分数和整体精度 O_{A} 上都获得了最好的分割结果, 其中, 在建筑物类别中取得了最高的分割准确率 96.31%, 因为在 Vaihingen 数据集中建筑类别目标较大并且数据量较多, 并且建筑物类别中可捕获的边缘信息更准确, 模型训练更加稳定。对汽车类的分割准确率也达到了最优, 证明了网络在提取小目标尺度地物方面的优势。而不透水表面和树木的分割准确率相对较低, 因为这 2 类地物容易与建筑表面或低矮植物混淆, 导致不正确的分割结果。但从整体来看, 本文模型在 m_{F_1} 分数、总体分割准确率 O_{A} 和平均交并比 m_{IoU} 上都有一定提升, 证明了 CSUNet 模型的有效性。

表 1　CSUNet 和其他模型在 Vaihingen 数据集上的分割准确率对比结果

Table 1　Comparison of segmentation accuracy between CSUNet and other models on Vaihingen dataset　%

模型	分割准确率					m_{F_1}	O_{A}	m_{IoU}
	不透水表面	建筑物	低矮植物	树木	车辆			
BANet[32]	92.23	95.23	83.75	89.92	86.76	89.58	90.48	81.35
STransFuse[33]	88.25	91.46	79.04	85.51	77.14	78.67	86.07	66.66
ST-UNet[34]	76.36	82.98	57.79	72.53	61.48	82.15	83.99	70.23
HSN[38]	92.40	95.15	78.92	88.76	82.42	87.53	89.67	76.29
HCANet[39]	**94.29**	96.20	83.33	**92.38**	88.86	91.01	88.84	77.46
UFMG-4[40]	91.10	94.50	82.90	88.80	81.30	87.72	89.40	74.91
S-RA-FCN[41]	91.74	94.97	80.63	88.57	87.05	88.54	89.23	80.76
SERNet[42]	91.79	91.64	81.68	81.64	75.68	84.49	88.19	72.69
CSUNet	93.72	**96.31**	**85.59**	90.38	**90.40**	**91.28**	**91.69**	**83.22**

本文模型和其他模型在 Potsdam 数据集上的分割准确率对比结果如表 2 所示, 表中加粗显示最优结果, 可以看出, 相较于其他模型, 本文 CSUNet 模型在 5 种地物类别上的 m_{F_1} 分数和整体分割精度都达到了最优。因为 Potsdam 数据集的样本数量更加充足, 地物信息更加丰富, 满足了 Transformer 结构对于大规模数据的依赖性, 避免因为数据量有限造成的过拟合问题, 所以本文模型在 Potsdam 数据集上的分割准确率更高。低矮植物和树木颜色非常相似, 给分割任务带来了挑战, 并且稀疏的低矮植物易与不透水表面类别混淆, 可能会造成不正确的分割, 而本文模型在这 2 个类别中表现更好。由于 CSUNet 模型可以同时学习整体和细节信息并加以融合, 因此, 对类间差异较小的类别能更好地进行区分, 减少漏报误报现象。与其他几个类别相比, 汽车类目标在整个训练图像中占比较小, 并且容易被建筑物和树木遮挡, 很难提取完整的特征信息进行正确的像素分类, 而 CSUNet 模型在汽车类别上的分割精准确率达到了 96.72%, 这证明了本文提出的双编码器结构和特征融合模块的有效性, CSUNet

表 2　CSUNet 和其他模型在 Potsdam 数据集上的分割准确率对比结果

Table 2　Comparison of segmentation accuracy between CSUNet and other models on Potsdam dataset　%

模型	分割准确率					m_{F_1}	O_{A}	m_{IoU}
	不透水表面	建筑物	低矮植物	树木	车辆			
BANet[32]	93.34	96.66	87.37	89.12	95.99	92.50	91.06	86.25
STransFuse[33]	89.75	93.92	82.91	83.61	88.51	82.08	86.71	71.46
ST-UNet[34]	79.19	86.63	67.89	66.37	79.77	86.13	85.60	75.97
HSN[38]	92.23	95.25	88.24	82.99	89.82	89.71	89.77	78.42
HCANet[39]	92.88	96.90	87.25	88.15	93.88	91.81	90.67	76.64
UFMG-4[40]	90.80	95.60	84.40	84.30	92.40	89.50	87.90	75.35
S-RA-FCN[41]	91.33	94.70	86.81	83.47	94.52	90.17	88.59	82.38
SERNet[42]	91.75	95.30	73.74	82.26	92.13	87.04	90.29	76.76
CSUNet	**93.78**	**96.99**	**88.35**	**89.94**	**96.72**	**93.16**	**91.63**	**87.48**

模型可以更加准确地提取并融合小目标的有效特征信息，从而提高遥感影像的分割精度。

图6为本文模型和其他模型在Vaihingen数据集上的影像分割细节对比。由于遥感影像缺少高度信息，会导致建筑物顶部和不透水表面具有相似的纹理特征。如图6前5行所示，CSUNet模型对建筑的分割效果较好，可以较为准确地区分水泥道路和水泥屋顶的低层建筑这两类相似地物，这主要得益于本文提出的FAM可以从全局和局部上下文信息中学习特征，对建筑物表面可以实现相对完整的分割。同时，本文提出的ETL模块可以更准确地捕获物体边缘细节语义信息，从而获得较好的语义分割效果。从图6第6行和第7行可以看出，树木的阴影遮挡给分割任务带来了困难，主要原因是网络的上下文和空间信息不足，而本文的CSUNet模型对于被树木阴影遮挡的小目标汽车类也能实现较好的分割效果，此外，对于汽车排列较密集的场景，网络通过引入ETL模块也能更好地进行分割，提高边缘分割准确率。与不透水表面相比，汽车在图像中所占的比例很小，而且汽车通常是出现在不透水表面中。由于卷积依赖于固定大小的卷积核捕获特征，当提取小尺度特征时，很容易受到周围特征类别的影响，而CSUNet模型的Transformer分支可以有效地解决这类问题，通过自注意力计算获得权重矩阵，自适应地调整特征值，对像素类别的表示更加准确，并且双编码器结构可以通过分层级联来聚合更多有利于语义分割的特征信息。

图7为本文模型和其他模型在Potsdam数据集上的影像分割细节对比，由图可知，本文CSUNet模型取得了较好的分割结果。如图7前4行所示，本

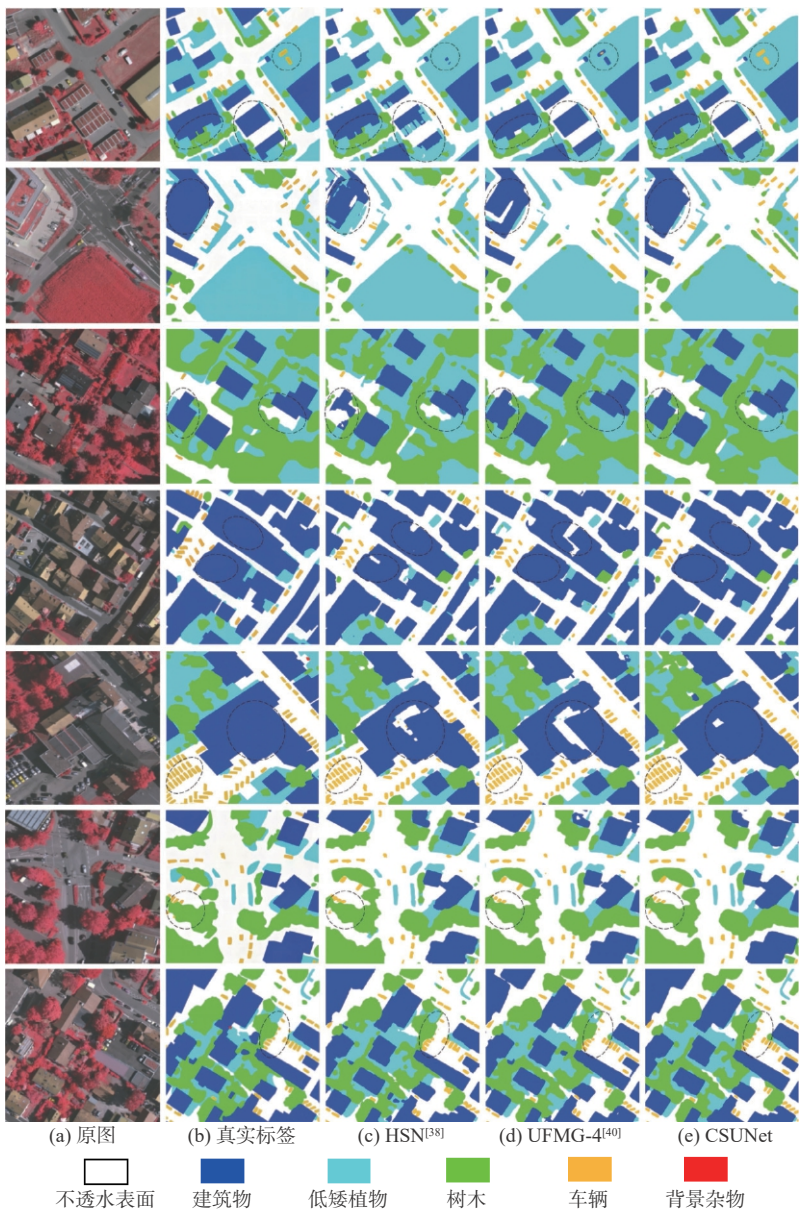


图6 ISPRS Vaihingen数据集分割结果细节对比

Fig. 6 Comparison of detailed segmentation results on ISPRS Vaihingen dataset

文模型对建筑、低矮植物等大规模地物类别的分割效果较好, 边缘细节信息的识别也更加准确, 并且减少了背景杂物类的误报问题。低矮植物与带有屋顶草坪的低层建筑容易混淆, 带有草坪的不透水表面也很难被准确识别, 从图 7 第 1 行和第 3 行可以看出, CSUNet 模型实现了较好的分割结果, 主要得益于自注意力机制对全局上下文信息的提取能力, 可以有效捕获长程依赖关系。从图 7 第 5 行和第 6 行可以看出, CSUNet 模型对于汽车类小目标地物的分割效果更好, 可以准确识别出不同种类的汽车, 并且对于密集排列的汽车类小目标地物的边缘分割也更加准确, 主要因为 CSUNet 模型的双编码器结构可以同时学习全局信息和细节信息并将两者进行融合。

Transformer 网络结构处理图像时, 会将图像块拉伸成一维空间位置向量, 在周围较大目标的像素值影响下, 小目标的像素值将会被分隔到距离较远的位置, 导致小目标像素特征获取困难, 而 CSUNet 模型可以从全局上下文信息中学习特征, 有助于解决 Transformer 对小目标特征提取效果不佳的问题。从图 7 第 7 行可以看出, CSUNet 模型对于细长型地物的识别也较准确, 当低矮植被和树木类相邻时, 如果不能准确地判断出边缘, 高度的类间相似性很容易使得大量的目标像素被错误预测, 而 CSUNet 模型通过 ETL 模块较好地解决了这一问题, 因为清晰的边缘信息可以使目标对象的内部像素更容易被正确分类。

Swin Transformer 与 CSUNet 模型在 Vaihingen 和 Potsdam 数据集上的对比结果分别如表 3 和表 4 所示。从模型参数量对比可以看出, CSUNet 模型

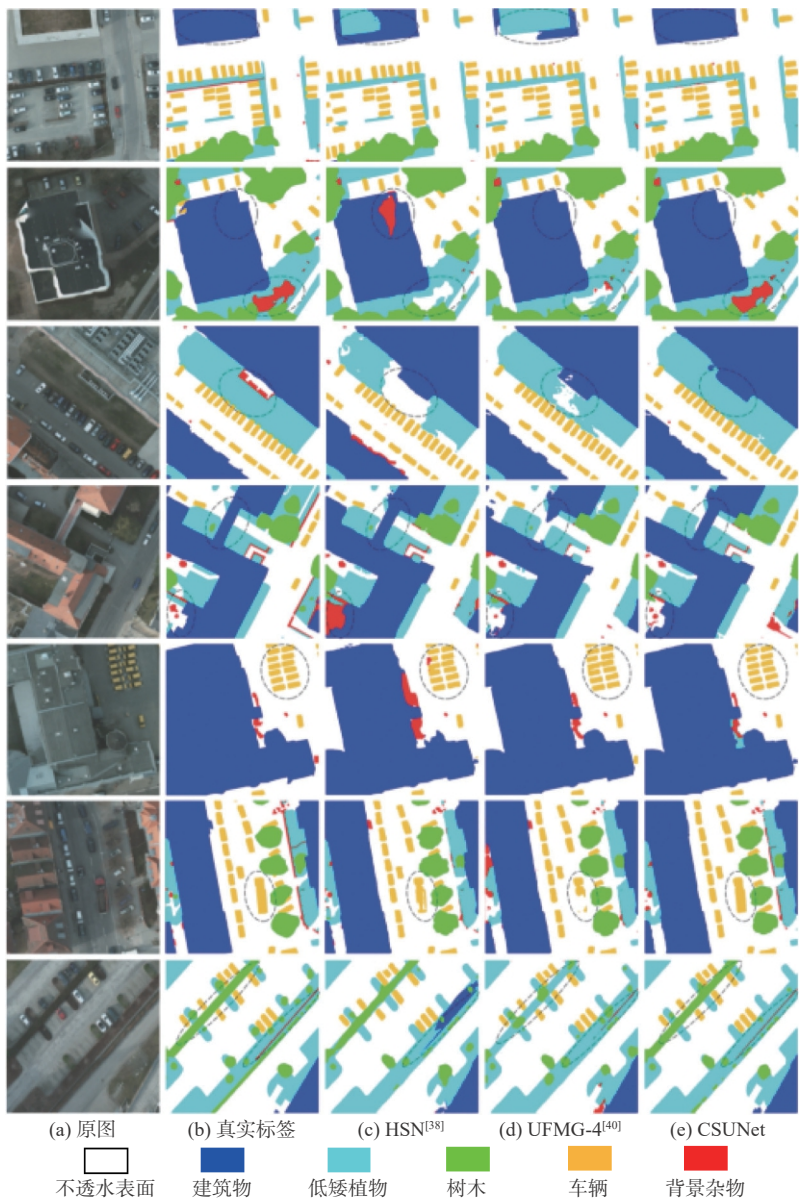


图 7　ISPRS Potsdam 数据集分割结果细节对比

Fig. 7　Comparison of detailed segmentation results on ISPRS Potsdam dataset

在计算复杂度方面并不占优势，并且使用双编码器结构的CSUNet模型训练时间要比单编码器网络的训练时间长。虽然上述问题可能会限制CSUNet模型在某些场景（如小型移动设备）中的应用，但从分割准确率对比结果可以看出，CSUNet模型的结果要明显优于Swin Transformer模型，证明CSUNet模型在遥感影像语义分割领域中具有一定的研究意义。在之后的研究工作中，也会关注如何在保证分割准确率的同时降低计算冗余。

表3 Swin Transformer与CSUNet在Vaihingen数据集的对比结果

Table 3 Comparison of Swin Transformer and CSUNet on Vaihingen dataset

模型	m_{F_1}/%	O_A/%	参数量	时间/min
Swin Transformer	86.21	87.69	59.83×10^6	637
CSUNet	91.28	91.69	87.39×10^6	861

3.5 消融实验

为验证不同模块对CSUNet模型的重要性，本文在Vaihingen和Potsdam数据集上进行消融实验，主要对FAM和ETL进行实验验证，探究每个模块对实验结果的影响，实验结果如表5和表6所示，其中，a表示不添加FAM和ETL的实验结果，编码器结构中只保留Transformer分支，边缘提取分支中只保留“基于像素的交叉熵损失”作为基本的损失函数以训练网络，b表示只添加FAM的实验结果，c表示只添加ETL的实验结果，d表示同时添加FAM和ETL的实验结果。

表4 Swin Transformer与CSUNet在Potsdam数据集的对比结果

Table 4 Comparison of Swin Transformer and CSUNet on Potsdam dataset

模型	m_{F_1}/%	O_A/%	参数量	时间/min
Swin Transformer	90.44	89.39	59.83×10^6	2 158
CSUNet	93.16	91.63	87.39×10^6	2 740

在Vaihingen数据集中，将FAM添加到网络后，m_{F_1}和O_A分别达到了90.23%和90.52%，相比a模型提高了1.52%和1.21%，汽车类小目标的分割准确率提高了2.36%，在使用FAM嵌入更多全局上下文信息后，网络可以自适应地融合不同语义信息的特征，通过聚合上下文和局部特征进一步提高模型性能，小目标对象的分割准确率得到了有效提升。ETL模块的引入有效地减少了模糊边缘的影响，m_{F_1}和O_A分别达到了90.05%和90.39%，相比a模型提高了1.34%和1.08%。同时添加FAM和ETL模块，m_{F_1}和O_A分别达到了91.28%和91.69%，相比a模型提高了2.57%和2.38%。

表5 在Vaihingen数据集上不同模块的消融实验结果

Table 5 Ablation experiment results of different modules on Vaihingen dataset %

模型	FAM	ETL	分割准确率					m_{F1}	O_A
			不透水表面	建筑物	低矮植物	树木	车辆		
a			92.13	93.04	82.90	88.26	87.22	88.71	89.31
b	√		93.60	95.62	84.65	88.79	89.58	90.23	90.52
c		√	93.66	94.32	84.04	89.11	89.12	90.05	90.39
d	√	√	93.72	96.31	85.59	90.38	90.40	91.28	91.69

表6 在Potsdam数据集上不同模块的消融实验结果

Table 6 Ablation experiment results of different modules on Potsdam dataset %

模型	FAM	ETL	分割准确率					m_{F1}	O_A
			不透水表面	建筑物	低矮植物	树木	车辆		
a			91.95	94.84	86.69	88.76	94.11	91.27	90.12
b	√		93.39	96.53	87.37	89.49	95.77	92.31	90.90
c		√	92.55	95.66	87.34	89.63	95.52	92.14	90.78
d	√	√	93.78	96.99	88.35	89.94	96.72	93.16	91.63

Potsdam数据集上的结果变化趋势和Vaihingen大致相同，但由于Potsdam数据集具有更复杂的场景，所以整体变化幅度要小于在Vaihingen数据集上的实验结果。消融实验结果表明，本文提出的FAM和ETL模块都可以有效提高遥感影像语义分割的性能。

4 结 论

1）本文提出一种融合卷积和自注意力机制的高分辨率遥感影像语义分割模型，以提升遥感影像语义分割准确率。

2）在ISPRS Vaihingen和Potsdam数据集上的

实验结果表明, 本文 CSUNet 模型获得了较高的遥感影像语义分割准确率, 优于目前的主流模型。

CSUNet 模型仍然有提升的空间, 未来的研究工作将重点考虑如何融合多模态遥感数据以提升语义分割性能。

参考文献（References）

[1] 龚健雅. 人工智能时代测绘遥感技术的发展机遇与挑战[J]. 武汉大学学报 (信息科学版), 2018, 43(12): 1788-1796.
GONG J Y. Chances and challenges for development of surveying and remote sensing in the age of artificial intelligence[J]. Geomatics and Information Science of Wuhan University, 2018, 43(12): 1788-1796(in Chinese).

[2] ZHAO J Q, ZHOU Y, SHI B Y, et al. Multi-stage fusion and multi-source attention network for multi-modal remote sensing image segmentation[J]. ACM Transactions on Intelligent Systems and Technology, 2021, 12(6): 1-20.

[3] DING L, ZHANG J, BRUZZONE L. Semantic segmentation of large-size VHR remote sensing images using a two-stage multiscale training architecture[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(8): 5367-5376.

[4] SHEIKH R, MILIOTO A, LOTTES P, et al. Gradient and log-based active learning for semantic segmentation of crop and weed for agricultural robots[C]//Proceedings of the IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2020: 1350-1356.

[5] 李道纪, 郭海涛, 卢俊, 等. 遥感影像地物分类多注意力融和 U 型网络法[J]. 测绘学报, 2020, 49(8): 1051-1064.
LI D J, GUO H T, LU J, et al. A remote sensing image classification procedure based on multilevel attention fusion U-Net[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(8): 1051-1064(in Chinese).

[6] MARMANIS D, SCHINDLER K, WEGNER J D, et al. Classification with an edge: improving semantic image segmentation with boundary detection[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 135: 158-172.

[7] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3431-3440.

[8] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]//Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention. Berlin: Springer, 2015: 234-241.

[9] LI S F, LIAO C, DING Y L, et al. Cascaded residual attention enhanced road extraction from remote sensing images[J]. ISPRS International Journal of Geo-Information, 2022, 11(1): 9.

[10] 吴强强, 王帅, 王彪, 等. 空间信息感知语义分割模型的高分辨率遥感影像道路提取[J]. 遥感学报, 2022, 26(9): 1872-1885.
WU Q Q, WANG S, WANG B, et al. Road extraction method of high-resolution remote sensing image on the basis of the spatial information perception semantic segmentation model[J]. National Remote Sensing Bulletin, 2022, 26(9): 1872-1885(in Chinese).

[11] LI X, XU F, LYU X, et al. Dual attention deep fusion semantic segmentation networks of large-scale satellite remote-sensing images[J]. International Journal of Remote Sensing, 2021, 42(9): 3583-3610.

[12] DING L, TANG H, BRUZZONE L. LANet: local attention embedding to improve the semantic segmentation of remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(1): 426-435.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the Annual Conference on Neural Information Processing Systems. New York: ACM, 2017: 5998-6008.

[14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL].(2021-06-02)[2023-05-10].https://arxiv.org/abs/2010.11929.

[15] SUN Z Y, ZHOU W P, DING C, et al. Multi-resolution transformer network for building and road segmentation of remote sensing image[J]. ISPRS International Journal of Geo-Information, 2022, 11(3): 165.

[16] LIU Z, LIN Y T, CAO Y, et al. Swin Transformer: hierarchical vision transformer using shifted window[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 9992-10002.

[17] LADICKÝ L, RUSSELL C, KOHLI P, et al. Associative hierarchical CRFs for object class image segmentation[C]//Proceedings of the IEEE 12th International Conference on Computer Vision. Piscataway: IEEE Press, 2009: 739-746.

[18] ARNAB A, JAYASUMANA S, ZHENG S, et al. Higher order conditional random fields in deep neural networks[C]//Proceedings of the European Conference on Computer Vision. Beilin: Springer, 2016: 524-540.

[19] CHEN B, QIU F, WU B F, et al. Image segmentation based on constrained spectral variance difference and edge penalty[J]. Remote Sensing, 2015, 7(5): 5980-6004.

[20] DIAKOGIANNIS F I, WALDNER F, CACCETTA P, et al. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 94-114.

[21] LIU R, MI L, CHEN Z Z. AFNet: adaptive fusion network for remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(9): 7871-7886.

[22] PAN X, ZHAO J, XU J. Conditional generative adversarial network-based training sample set improvement model for the semantic segmentation of high-resolution remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(9): 7854-7870.

[23] MA A L, WANG J J, ZHONG Y F, et al. FactSeg: foreground activation-driven small object semantic segmentation in large-scale remote sensing imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 5606216.

[24] 杨军, 于茜子. 结合空洞卷积的 FuseNet 变体网络高分辨率遥感影像语义分割[J]. 武汉大学学报 (信息科学版), 2022, 47(7): 1071-1080.
YANG J, YU X Z. Semantic segmentation of high-resolution remote sensing images based on improved FuseNet combined with at-

rous convolution[J]. Geomatics and Information Science of Wuhan University, 2022, 47(7): 1071-1080(in Chinese).

[25] CHEN X, LI Z Q, JIANG J, et al. Adaptive effective receptive field convolution for semantic segmentation of VHR remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(4): 3532-3546.

[26] LIU S, DING W R, LIU C H, et al. ERN: edge loss reinforced semantic segmentation network for remote sensing images[J]. Remote Sensing, 2018, 10(9): 1339.

[27] XU Z Y, ZHANG W C, ZHANG T X, et al. HRCNet: high-resolution context extraction network for semantic segmentation of remote sensing images[J]. Remote Sensing, 2021, 13(1): 71.

[28] CAO Y, XU J R, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2019: 1971-1980.

[29] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[C]//Proceedings of the 9th International Conference on Learning Representations. Virtual: ICLR, 2021.

[30] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 936-944.

[31] WANG L B, LI R, DUAN C X, et al. A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 6506105.

[32] WANG L B, LI R, WANG D Z, et al. Transformer meets convolution: a bilateral awareness network for semantic segmentation of very fine resolution urban scene images[J]. Remote Sensing, 2021, 13(16): 3065.

[33] GAO L, LIU H, YANG M H, et al. STransFuse: fusing swin transformer and convolutional neural network for remote sensing image semantic segmentation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 10990-11003.

[34] HE X, ZHOU Y, ZHAO J Q, et al. Swin transformer embedding UNet for remote sensing image semantic segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 4408715.

[35] KONECNY G. The international society for photogrammetry and remote sensing (ISPRS) study on the status of mapping in the world[C]//Proceedings of the In International Workshop on Global Geospatial Information. Piscataway: IEEE Press, 2013: 4-24.

[36] LIU Y, MINH NGUYEN D, DELIGIANNIS N, et al. Hourglass-ShapeNetwork based semantic segmentation for high resolution aerial imagery[J]. Remote Sensing, 2017, 9(6): 522.

[37] 谢文亮, 朱丹, 佟新鑫. 一种基于视觉注意的小目标检测方法[J]. 计算机工程与应用, 2013, 49(12): 125-128.
XIE W L, ZHU D, TONG X X. Small target detection method based on visual attention[J]. Computer Engineering and Applications, 2013, 49(12): 125-128(in Chinese).

[38] PAN X R, GAO L R, MARINONI A, et al. Semantic labeling of high resolution aerial imagery and LiDAR data with fine segmentation network[J]. Remote Sensing, 2018, 10(5): 743.

[39] LI X, XU F, XIA R L, et al. Hybridizing cross-level contextual and attentive representations for remote sensing imagery semantic segmentation[J]. Remote Sens, 2021, 13: 2986.

[40] NOGUEIRA K, DALLA MURA M, CHANUSSOT J, et al. Dynamic multicontext segmentation of remote sensing images based on convolutional networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(10): 7503-7520.

[41] MOU L C, HUA Y S, ZHU X X. Relation matters: relational context-aware fully convolutional network for semantic segmentation of high-resolution aerial images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(11): 7557-7569.

[42] ZHANG X Y, LI L H, DI D L, et al. SERNet: squeeze and excitation residual network for semantic segmentation of high-resolution remote sensing images[J]. Remote Sensing, 2022, 14(19): 4770.

U-shaped semantic segmentation network of high-resolution remote sensing images embedded with self-attention mechanism

YANG Jun[1, 2, *], ZHANG Jinying[2]

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. School of Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: To reduce the difficulty of extracting small object features from high-resolution remote sensing images, a dual-encoder feature fusion network model based on convolution structure and self-attention mechanism was proposed, which was suitable for semantic segmentation of high-resolution remote sensing images. Firstly, a dual-encoder structure was designed to extract global and local detail information of remote sensing images and improve the segmentation accuracy of small object features. Secondly, a feature aggregation module was used to aggregate feature information at different stages, so as to embed more global contextual information. Finally, an edge thinning loss module was used to improve the recognition ability of the model for object edge information. The m_{F_1} average value of the segmentation results on the ISPRS Vaihingen and Potsdam datasets achieved 91.28% and 93.16%, respectively. Compared with the current mainstream algorithms, the segmentation accuracy of small objects like cars and the overall segmentation accuracy were improved. The proposed model solves the problem of inaccurate segmentation of small objects and edge information in the semantic segmentation of high-resolution remote sensing images to a certain extent.

Keywords: high-resolution remote sensing images; semantic segmentation; deep learning; self-attention mechanism; feature fusion

Received: 2023-05-23; **Accepted**: 2023-10-26; **Published Online**: 2023-11-08 11:34

URL: link.cnki.net/urlid/11.2625.V.20231107.1027.003

Foundation items: National Natural Science Foundation of China (42261067); 2025 Key Talent Project of Gansu Province (2025RCXM031); Talent Innovation and Entrepreneurship Project of Lanzhou City (2020-RC-22)

* **Corresponding author**. E-mail: yangj@mail.lzjtu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0300

基于图神经网络的航班地面保障数据插补算法

邢志伟[1]，孙恪[1,2]，罗谦[2,*]，刘畅[2]，张涛[2]，乔迪[3]

(1. 中国民航大学 电子信息与自动化学院，天津 300300；2. 中国民航局第二研究所 工程技术研究中心，成都 610041；
3. 四川省机场集团有限公司成都天府国际机场分公司，成都 610041)

摘　　要：针对航班地面保障数据缺失问题，提出一种基于图神经网络的数据插补算法。通过降噪编码器降低原始数据中噪声对训练的影响，增强提取特征的可靠性；建立一种图表示学习框架，使用聚合函数聚合采样区间内节点的特征，实现神经网络节点的状态更新，得到第1次嵌入特征；应用长短时记忆网络对航班的时序信息进行第2次嵌入得到隐藏层的状态空间；通过反卷积神经网络进行特征还原，提出一种损失函数实现网络的迭代，在迭代多次后得到最终的航班地面保障数据插补结果。使用西南某机场2018年4～6月份的航班地面保障数据对所提算法进行测试，结果表明：相比于其他算法，所提算法在低缺失率时，插补误差平均降低了约74%；在较高缺失率时，插补误差平均降低了约68%；所提算法迭代次数约在100次，正则化系数约为0.5时，插补误差达到最低。

关 键 词：航班地面保障；图嵌入；缺失值插补；递归神经网络；自动编码器；多元时间序列
中图分类号：V352; TP311.13
文献标志码：A　　**文章编号**：1001-5965(2025)05-1528-11

航班地面保障过程作为过站航班的关键环节，能够反应机场运行效能，为机场协同决策提供重要参考[1]。在航班地面保障实际运行过程中，保障数据采集大多依靠人工采集，这些数据往往在采集、传送和存储过程中出现缺失[2]。缺失的数据不仅不足以正确的表征机场运行态势，而且较差的数据质量对下游的挖掘也会造成较大影响。数据插补方面的研究由来已久，而对航班地面保障数据缺失问题的研究较少，需要进一步探究。

传统的处理缺失值的方法是采用均值插补[3]或删除该样本[4]，而这种方法可能会潜在改变样本的分布，出现过拟合等情况[5]，在样本缺失率较高时，偏差可能会在下游处理任务被放大。因此，如何实现航班地面保障缺失数据精确插补是机场高效运行的关键问题之一。Rubin[6]提出针对数据缺失可能原因的讨论，部分研究者提出期望最大化算法、最邻近算法及链式方程多元插补等基于统计的方法[7-11]。简单的统计方法在应对样本量充足、缺失率较低的情景下会带来较大的方便，而在某些场景中可能并不适用。

数据插补的本质是依据历史数据建模从而通过模型反馈对数据缺失值进行填补，考虑神经网络天生对变量间的非线性关系映射的优势，越来越多的研究者尝试将其应用到数据插补上来[12-14]。Wang等[15]提出改进的编码器算法，通过对特征层正交分解，实现了高效的特征提取。图神经网络由于整合了数据图结构特征的神经网络，相比于传统的网络结构能够进一步把握变量间的关联，Cui等[16]定义网络间流量的转化为一种马尔可夫过程，较好的捕捉了网络的时空效率，提升了预测效率并节省了计

收稿日期：2023-05-30；**录用日期**：2023-10-16；**网络出版时间**：2023-11-21 12:59
网络出版地址：link.cnki.net/urlid/11.2625.V.20231121.1044.001
基金项目：国家重点研发计划项目(2018YFB1601200)；国家自然科学基金青年科学基金(62203452)；中国民航大学研究生科研创新资助项目(2022YJS105)
* **通信作者**. E-mail：cauc_xzw@163.com

引用格式：*邢志伟，孙恪，罗谦，等. 基于图神经网络的航班地面保障数据插补算法[J]. 北京航空航天大学学报，2025，51(5)：1528-1538.*
XING Z W，SUN K，LUO Q，et al. Imputation algorithm for flight ground support data based on graph neural network[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1528-1538 (in Chinese).

算资源; Yao 等[17] 提出一种空间交互图卷积网络模型, 将图形卷积与映射函数结合从图表示学习的角度预测数据流量; Yang 等[18] 提出一种在线学习的框架能够对实时采集的数据进行插补, 通过将数据分解成时间和空间特征, 再分别进行处理, 但其研究的对象仍然是静态下的; Cini 等[19] 提出图递归插补网络(graph recurrent imputation networks, GRIN)模型专门用于对时间序列数据的插补; You 等[20] 通过标签预测的方法尝试使用含有缺失值的数据集直接进行预测, 并提出一种基于图表示学习的插补框架(Grape)。国内也有部分研究者试图由低秩补全理论解决交通和气象等领域的缺失值问题[21-23], 而这些研究缺少对数据时序的讨论。

为解决时序数据的缺失问题, Che 等[24] 提出一种基于门控循环单元(gated recurrent unit, GRU)的用于插补的递归神经网络(recurrent neural networks, RNN)(简称 GRU-D), GRU-D 将缺失机制融入到深度网络中从而能够捕捉到时序的长时依赖性, 并且实现了较精准的预测结果, Tian 等[25] 采用多尺度平滑来推断丢失数据, 以克服采集到的交通数据长度不同、采样不规则等问题, 并提出一种基于长短期记忆(long short term memory , LSTM)网络的新方法(简称 LSTM-M), 学习变量间的时间关联信息, 以实现缺失数据的插补。还有部分针对多元时间序列(multivariate time series, MTS)的插补模型, Luo 等[26] 提出用于 MTS 分布建模的多元时间序列生成对抗网络(multivariate time series-generative adversarial networks, MTS-GAN), 并对 MTS 生成任务制定约束将其应用于 MTS 插补; Spinelli 等[27] 将编码器网络引入标准 GAN 架构中, 该架构避免了对抗阶段中对数据的预处理。而基于 GAN 的插补网络往往需要一个额外的阶段来对模型进行训练[28], 实际情况中数据的缺失满足训练对质量的要求, 现有方法大多是在假设数据集完整的基础上进行模型的验证, 缺少针对实际业务的考虑。

本文在上述研究的基础上, 基于 LSTM 网络提出基于图的时间序列归纳式插补(简称 IGITS-LSTM)算法。首先, 运用降噪自动编码器网络结构, 在编码器阶段加入随机噪声以保证提取到可靠的特征, 规避了在训练阶段对数据的预处理过程; 其次, 结合实际业务流程, 通过保障节点间信息的传递, 聚合采样区间内的节点信息, 更新当前节点的特征嵌入, 在门控神经网络中引入衰退率, 控制节点间传递的信息量; 然后, 应用 LSTM 网络捕捉多航班间的时间相似度, 得到隐藏层的特征向量; 最后, 在反卷积神经网络上将提取到的特征向量还原为原本的维度, 通过提出的损失函数对网络参数多次迭代后, 得到最终插补结果。

1　问题描述

航班地面保障过程是一个存在并行工作流程的动态混杂过程, 根据实际业务流程可大致分为航油补充、客舱服务、货仓装卸、机务巡检等主要并行流程, 各个流程之间不仅存在先后关系, 还有一定的逻辑推理次序。保障资源必须按照机场地面保障运行标准的进离场时间规定按时到位并提供保障服务。图 1 为机场航班地面保障一般作业流程。

本文针对航班地面保障过程中各个保障节点中存在的数据缺失问题, 通过算法模型对历史数据进行分析, 推断网络节点中的缺失数据, 从而提升航班保障数据的质量, 为下游任务提供充足的参考。

假设航班地面保障数据一天采集 n 个航班量, 一个航班包含 D 个保障节点。将航班地面保障数据组织为矩阵形式:

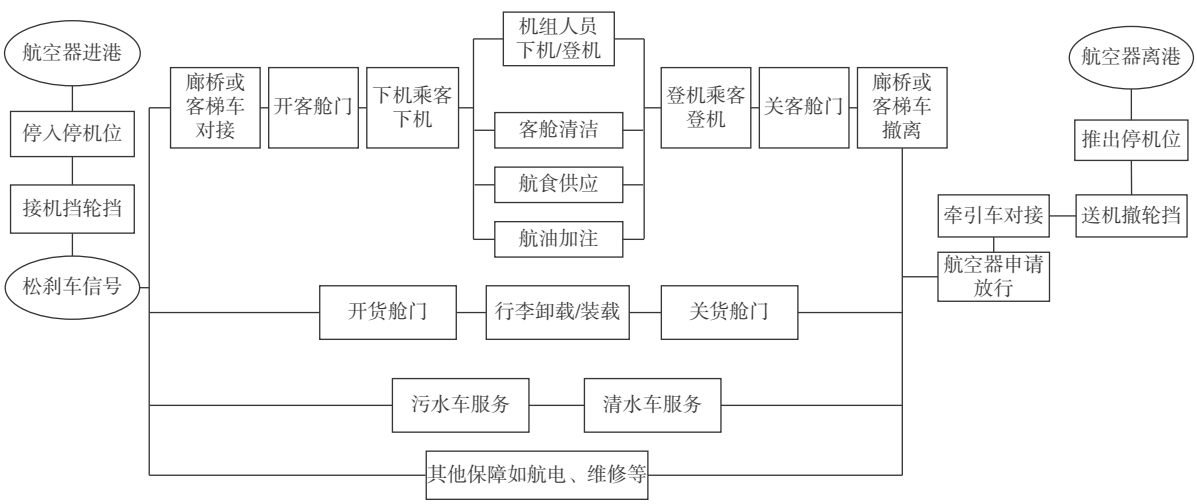


图 1　航班地面保障流程

Fig. 1　Flight ground operation process

$$\boldsymbol{X}=\begin{bmatrix} x_{t_1}^1 & \cdots & x_{t_i}^1 & \cdots & x_{t_n}^1 \\ \text{null} & \cdots & \text{null} & \cdots & x_{t_n}^2 \\ \vdots & & \vdots & & \vdots \\ x_{t_1}^D & \cdots & x_{t_i}^D & \cdots & x_{t_n}^D \end{bmatrix} \tag{1}$$

式中：$\boldsymbol{X}$为采集到的航班地面保障流程原始数据矩阵；null为缺失值；t_i为第i架航班的到达时刻，由于不同航班之间的采样间隔并不是连续不变的，因此，通过航班到达时刻来标记一次保障过程；d为一次保障流程中的第d个保障节点；$x_{t_i}^d$表示t_i时刻到达航班的保障过程中的第d个保障节点。

本文针对数据缺失常见的原因进行分析，一方面航班地面保障数据通常由人工收集[2]，地服保障人员在完成保障工作后可能出现人为的失误，忘记记录时间节点导致该保障节点的数据缺失，另一方面机场数据库积累了大量的保障数据，信号干扰或网络攻击也可能造成数据的大量缺失，而且为节省存储空间对数据进行主动压缩和数据在传输过程中出现的意外缺失都是不可避免的。定义一种遮蔽矩阵$\boldsymbol{M}$来表示航班地面保障数据中存在缺失的部分：

$$\boldsymbol{M}=[\boldsymbol{m}_{t_1},\boldsymbol{m}_{t_2},\cdots,\boldsymbol{m}_{t_n}]=\begin{bmatrix} m_{t_1}^1 & \cdots & m_{t_i}^1 & \cdots & m_{t_n}^1 \\ m_{t_1}^2 & \cdots & m_{t_i}^2 & \cdots & m_{t_n}^2 \\ \vdots & & \vdots & & \vdots \\ m_{t_1}^D & \cdots & m_{t_i}^D & \cdots & m_{t_n}^D \end{bmatrix} \tag{2}$$

其中：$m_{t_i}^d$定义如下：

$$m_{t_i}^d=\begin{cases} 1 & x_{t_i}^d \neq \text{null} \\ 0 & \text{其他} \end{cases} \tag{3}$$

采集的航班保障流程数据缺失率具体定义为

$$\alpha=\frac{\sum_{1}^{n}\sum_{d=1}^{D}(1-m_{t_i}^d)}{nD}\times 100\% \tag{4}$$

定义$\boldsymbol{X}'$为缺失恢复后的航班地面保障数据矩阵：

$$\boldsymbol{X}'=\begin{bmatrix} x_{t_1}^1 & \cdots & x_{t_i}^1 & \cdots & x_{t_n}^1 \\ x_{t_1}^2 & \cdots & x_{t_i}^2 & \cdots & x_{t_n}^2 \\ \vdots & & \vdots & & \vdots \\ x_{t_1}^D & \cdots & x_{t_i}^D & \cdots & x_{t_n}^D \end{bmatrix} \tag{5}$$

因此，航班地面保障数据插补问题可以看作学习一个函数来恢复缺失的保障数据，具体可以描述为

$$F(\boldsymbol{X},\boldsymbol{M})=\boldsymbol{X}' \tag{6}$$

2 基于图嵌入的航班地面保障数据插补算法

本文以图自动编码器框架为基础，分析多航班保障业务间的关联关系，提出IGITS-LSTM算法。首先，考虑不同航班保障过程节点间的特征相似度构建航班保障过程网络模型并通过全连接神经网络对节点的特征维度进行扩充；然后，利用嵌入式-门控循环单元（gated recurrent unit-aggregate, GRU-A）对扩充后的张量模型分割后进行采样聚合，得到第1次嵌入的特征张量；其次，考虑航班间的时间关联关系，用LSTM网络对特征张量进行第2次特征提取，得到最终的中间层表示；最后，在解码层的帮助下将中间层还原为初始维度，并与初始数据进行整合，获得一个完整的数据集。整个数据插补流程示意如图2所示，图中，$\boldsymbol{H}$为特征提取过程中的中间向量。

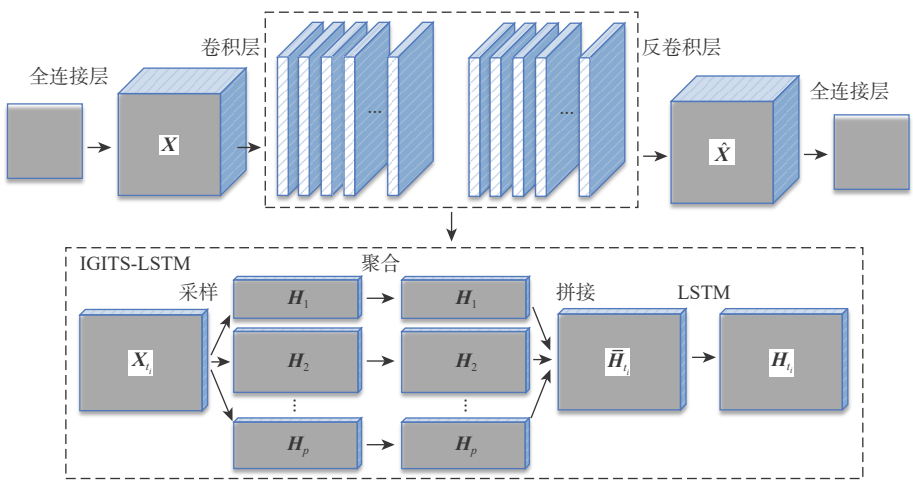


图2 航班地面保障数据插补流程示意图

Fig. 2 Schematic diagram of flight ground operation data imputation process

2.1 GRU-A嵌入

为构建航班地面保障网络结构模型，需要先对保障信息中的图结构信息进行提取，在航班地面保障网络拓扑图中，将每个保障环节看作是图中一个

节点, 拓扑图的邻接矩阵 $\boldsymbol{A}$ 由特征向量的相似矩阵 $\boldsymbol{S}$ 得到, 根据节点特征属性计算两两节点间的欧氏距离得到相似矩阵 $\boldsymbol{S}$, 对于缺失值采用局部平均的计算方式进行初始化:

$$S_i = \text{distance}\left(x_{t_i}^m, x_{t_i}^n\right) \tag{7}$$

式中: S_i 为在 t_i 下的航班保障节点 m 和 n 特征向量的欧氏距离。参考大规模流形正则化算法, 设定距离参考阈值, 当节点间的度量高于该阈值时则进行关联, 得到的邻接矩阵随后用于图嵌入过程。

图嵌入过程与基于空间域的卷积方法(简称 GraphSAGE)相似, 从航班保障流程时序图中选取出具有相似特征结构的节点进行采样。与传统的知识图谱网络由某个节点的发散状拓扑结构不同, 航班地面保障网络呈现的是一张存在唯一起讫点, 从起点到终点为通路状态的拓扑结构, 符合保障过程中节点随时序向前发展的实际业务场景, 并且满足节点的排列不变性。由此, 本文提出一种全新的采样的方法, 如图 3 所示, 其中, p 和 f 分别为不同的时序通道和航班保障节点数据不同的特征维度, 根据网络拓扑结构的特点, 确定航班保障流程的初始节点(上轮挡)和结束节点(撤轮挡), 这 2 个节点往往是唯一且确定的, 依据航班地面保障时序图的邻接矩阵 $\boldsymbol{A}_t$, 采用路径搜索算法, 得到该时序图下从起始保障节点到终止节点的所有通路, 并依据采样到的路径对保障节点进行划分, 确定采样范围。

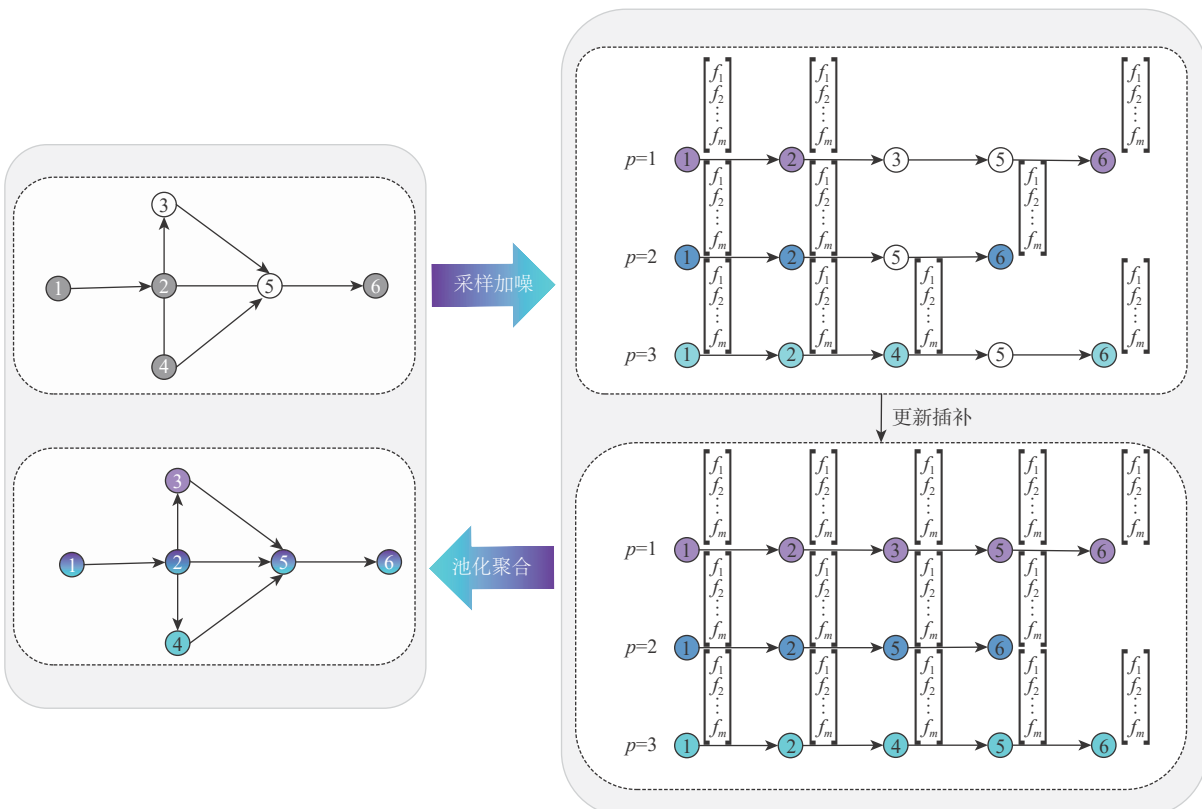


图 3　GRU-A 采样与聚合示意图

Fig. 3　Schematic diagram of GRU-A sample and aggregate

航班保障流程中节点间信息的交互方式依存于时间信息中, 本文提出用于时序插补的 GRU-A 模块, 先引入时滞向量 $\boldsymbol{\delta}$ 以反映当前值与先前时序的时间滞后属性, 定义在当前值缺失状态下时间的滞后程度, 初始状态下该值默认为 0, 具体计算式为

$$\delta_d = \begin{cases} x_t^d - x_t^{d-1} & m_{t_i}^d = 1,\ i > 1 \\ \delta_{d-1} + x_t^d - x_t^{d-1} & m_{t_i}^d = 0,\ i > 1 \\ 0 & t = 1 \end{cases} \tag{8}$$

式中: δ_d 为一次保障流程中第 d 个节点的时滞; x_t^d 为第 d 个保障节点的时间属性。

为进一步对缺失值进行合理的插补, 引入衰退率 γ 对传统的 GRU 网络进行改进, 通过控制上一个时序信息传入当前时序的比例, 实现缺失值的精确插补, 第 d 个保障节点更新时的计算衰退率 γ_d 为

$$\gamma_d = \exp\left(-\max\left(0, W_\gamma \delta_d + b_\gamma\right)\right) \tag{9}$$

式中: W_γ 和 b_γ 为带训练的参数, 指数函数负幂是为使得衰退率 γ_d 在(0,1)范围内单调递减。

具体衰退率 γ 在 GRU-A 模块中的应用存在 2 处, 如式(10)和式(11)所示, 上一序列的数值在经

过全连接层的处理后通过衰退率γ对原始数据中缺失的部分进行插补。

$$\hat{\boldsymbol{x}}_d = \boldsymbol{w}_x \bar{\boldsymbol{h}}_{d-1} + \boldsymbol{b}_x \tag{10}$$

$$\bar{\boldsymbol{x}}_d = \boldsymbol{m}_{t_i} \odot \boldsymbol{x}_d + (\boldsymbol{I} - \boldsymbol{m}_{t_i}) \gamma_d \odot \hat{\boldsymbol{x}}_d \tag{11}$$

式中：$\bar{\boldsymbol{h}}_{d-1}$为上一个节点的隐藏状态；$\hat{\boldsymbol{x}}_d$为用于插补缺失值的状态估计；$\boldsymbol{I}$为单位矩阵；$\boldsymbol{m}_{t_i}$为值全为0或1的向量；$\bar{\boldsymbol{x}}_d$为保障流程中节点$d$插补后完整的特征向量；$\boldsymbol{w}_x$和$\boldsymbol{b}_x$都是带训练的参数向量；⊙表示Hadamard积，其含义表示为

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \odot \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{bmatrix}$$

衰退率γ还作用于网络中的隐藏层，较远间隔的数据对当前节点的嵌入参与度小于关联度较大的采样点附近的数据。GRU-A模块的具体聚合过程如下：

$$\boldsymbol{h}'_d = \gamma_d \bar{\boldsymbol{h}}_{d-1} \tag{12}$$

$$\boldsymbol{r}_d = \sigma(\boldsymbol{W}_{\mathrm{r}} \bar{\boldsymbol{x}}_d + \boldsymbol{U}_{\mathrm{r}} \boldsymbol{h}'_{d-1} + \boldsymbol{V}_{\mathrm{r}} \boldsymbol{m}_{t_i} + \boldsymbol{b}_{\mathrm{r}}) \tag{13}$$

$$\boldsymbol{z}_d = \sigma(\boldsymbol{W}_{\mathrm{z}} \bar{\boldsymbol{x}}_d + \boldsymbol{U}_{\mathrm{z}} \boldsymbol{h}'_{d-1} + \boldsymbol{V}_{\mathrm{z}} \boldsymbol{m}_{t_i} + \boldsymbol{b}_{\mathrm{z}}) \tag{14}$$

$$\tilde{\boldsymbol{h}}_d = \tanh(\boldsymbol{W} \bar{\boldsymbol{x}}_d + \boldsymbol{U}(\boldsymbol{r}_d \odot \boldsymbol{h}'_{d-1}) + \boldsymbol{V} \boldsymbol{m}_{t_i} + \boldsymbol{b}) \tag{15}$$

$$\bar{\boldsymbol{h}}_d = (\boldsymbol{I} - \boldsymbol{z}_d) \odot \boldsymbol{h}'_{d-1} + \boldsymbol{z}_d \odot \tilde{\boldsymbol{h}}_d \tag{16}$$

式中：$\boldsymbol{h}'_d$为被衰退算子处理后的隐藏状态向量；$\boldsymbol{r}_d$为重置门；$\boldsymbol{z}_d$为更新门；$\tilde{\boldsymbol{h}}_d$为候选隐藏状态；$\boldsymbol{W}_{\mathrm{r}}$、$\boldsymbol{W}_{\mathrm{z}}$、$\boldsymbol{W}$、$\boldsymbol{U}_{\mathrm{r}}$、$\boldsymbol{U}_{\mathrm{z}}$、$\boldsymbol{U}$、$\boldsymbol{V}_{\mathrm{r}}$、$\boldsymbol{V}_{\mathrm{z}}$、$\boldsymbol{V}$为待训练的参数矩阵；$\boldsymbol{b}_{\mathrm{r}}$、$\boldsymbol{b}_{\mathrm{z}}$、$\boldsymbol{b}$为待训练的参数向量。

航班保障网络数据经过GRU-A模块特征提取后依然是一个子块，且每个子块之间存在相同的节点，本文采用均值池化的方式对一次航班保障过程的子块进行聚合，对重合的节点特征向量取平均作为聚合后节点的特征表示，如下：

$$\bar{\boldsymbol{H}}_{t_i} = \sigma\left(\boldsymbol{w} \cdot \mathrm{MEAN}\left(\left\{\bar{\boldsymbol{H}}_p, p = \{1,2,\cdots,P\}\right\}\right)\right) \tag{17}$$

式中：$\bar{\boldsymbol{H}}_{t_i}$为$t_i$时刻下航班地面保障数据隐藏状态表征矩阵；$\bar{\boldsymbol{H}}_p$为单航班地面保障网络采样后分割的子块；$\boldsymbol{w}$为待训练的参数矩阵；$\sigma(\cdot)$为激活函数，MEAN为均值函数。

实际航班保障过程是一个针对多航班问题的研究，不同航班由于不同的任务需求，机型大小、乘客数量等都会存在较大区别，因此，多航班地面保障流程图的拓扑网络就是一个随时序变化的动态网络，通过GRU-A对单个时序内的网络节点进行填补，考虑到不同航班之间可能存在时间上的特征关联，额外通过一种基于LSTM网络的时序插补模型对不同航班之间时间维度上可能存在的特征关联进行提取，LSTM网络通过“门控”机制来调节信息流。

经过嵌入后的特征向量$\boldsymbol{h}_d$拼接后形成特征矩阵$\boldsymbol{H}_{t_i}$，再将单个航班保障数据矩阵拼接形成张量数据$\boldsymbol{H}$，用反卷积神经网络对其维度进行还原得到插补后完整的矩阵$\hat{\boldsymbol{X}}$，具体过程为

$$\boldsymbol{H}_{t_i} = \mathrm{concat}(\boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_D) \tag{18}$$

$$\boldsymbol{H} = \mathrm{concat}(\boldsymbol{H}_1, \boldsymbol{H}_2, \cdots, \boldsymbol{H}_D) \tag{19}$$

$$\hat{\boldsymbol{X}} = \mathrm{sigmoid}(\boldsymbol{L}\boldsymbol{H}\boldsymbol{\Theta}) \tag{20}$$

式中：$\boldsymbol{L}$和$\boldsymbol{\Theta}$为待训练的参数矩阵，sigmoid为激活函数。基于IGITS-LSTM算法的航班地面保障数据插补框架如图4所示，图中，$\boldsymbol{G}$表示数据图结构，在LSTM模块中$\boldsymbol{C}_{t-1}$表示上一步的记忆状态，$\hat{\boldsymbol{C}}_t$表示当前时刻下备选状态，$\boldsymbol{h}_t$表示隐藏状态，$\boldsymbol{f}_t$表示遗忘门的状态输出，$\boldsymbol{j}_t$表示输入门的状态输出，$\boldsymbol{o}_t$表示输出门的状态输出。

2.2 损失函数

为让解码器中的$\hat{\boldsymbol{X}}$更接近于原始矩阵$\boldsymbol{X}$，需要在隐藏层中找出最好的中间向量$\boldsymbol{H}$。而如何以更合理的数值替代矩阵中的缺失值是一个值得考虑的问题。受文献[29]的启发，本文提出一种评估插补值精确度的方法。首先，定义两部分损失函数，损失函数的一部分是信息损失，数据插补结果$\hat{\boldsymbol{X}}$应该更加接近初始不完整的时间序列$\boldsymbol{X}$，另一部分是对抗损失，迫使参数训练得到的嵌入表达更加准确；其次，具体描述两部分损失函数的构造细节。

信息损失由原始样本$\boldsymbol{X}$和插补结果$\hat{\boldsymbol{X}}$的平方差定义。由于插补出的数据本身可能受到训练次数的影响，因此，为减小这种不确定因素带来的影响，本文仅仅考虑未缺失部分的方差：

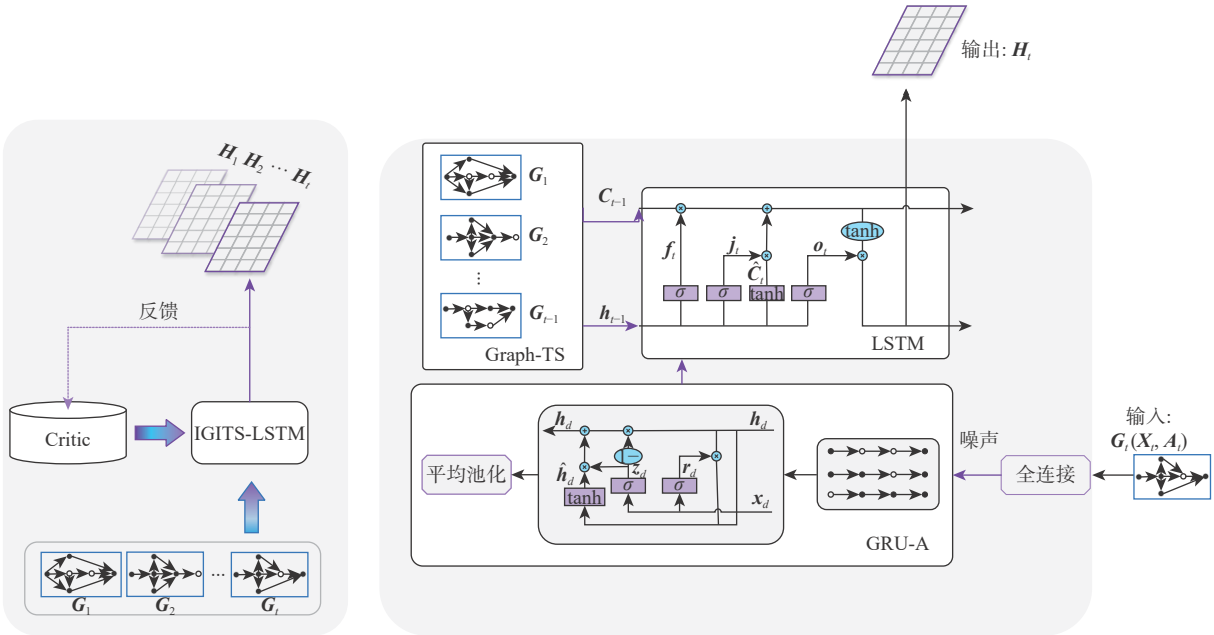
$$L_{\mathrm{info}} = \left\| \boldsymbol{X} \odot \boldsymbol{M} - \hat{\boldsymbol{X}} \odot \boldsymbol{M} \right\|_2 \tag{21}$$

为在无监督环境中学习有用的预测表示，引入一个基于图表示学习的损失函数，如式(22)所示，通过随机梯度下降算法调节矩阵权重和聚合函数中的参数，图嵌入中间层向量$\boldsymbol{h}$将会学习到图更多有用的特征，对抗损失函数会使得采样序列中离节点较近的节点有更加相似的表示，而远离节点的表示则更加不同。

$$L_{\mathrm{G}} = -\lg(\sigma(\boldsymbol{h}^{\mathrm{T}} \boldsymbol{h}_{\mathrm{adj}})) - Q E_{h_{\mathrm{adj}} \sim P(h_{\mathrm{adj}})} \lg(\sigma(-\boldsymbol{h}^{\mathrm{T}} \boldsymbol{h}_{\mathrm{adj}})) \tag{22}$$

式中：$\boldsymbol{h}_{\mathrm{adj}}$为固定长度上与$\boldsymbol{h}$相邻的节点；$P(\cdot)$为负采样分布，$E$表示负采样分布下的期望；$Q$为负采样的数量。

通过损失函数来不断优化插补效果，最后，数



(a) IGITS-LSTM框架　　(b) 编码器框架

图 4　航班保障数据插补框架

Fig. 4　Flight ground support data imputation framework

据插补损失函数由信息损失和对抗损失两部分组成:

$$L = (1-\lambda)L_{\text{info}} + \lambda L_{\text{G}} \tag{23}$$

其中: λ为正则化系数, 在训练中由人工调节, 负责调节信息损失和对抗损失二者的比例。

IGITS-LSTM 算法的伪代码如下所示, 对于每一个初始的时间序列 $\boldsymbol{X}$, 通过一个随机遮蔽矩阵对其中的部分数据置零, 将人工处理缺失后的矩阵输入 GRU-A 模块, 得到第 1 层的嵌入向量; 再通过 LSTM 模块对数据的时间关联特征进行进一步提取, 得到隐藏层的最终向量表示; 最后, 通过解码器对中间向量处理回原本的维度 $\hat{\boldsymbol{X}}$, 通过损失函数 L 将结果反馈到神经网络中, 当数据插补误差不断收敛到一个可以接受的阈值内, 则可以将插补结果替换初始序列中的缺失部分如下:

$$\boldsymbol{X}' = \boldsymbol{X}\cdot\boldsymbol{M} + (\boldsymbol{I}-\boldsymbol{M})\cdot\hat{\boldsymbol{X}} \tag{24}$$

IGITS-LSTM 算法的伪代码如下:

1. **输入 $\boldsymbol{X}$**: 原始航班地面保障数据; $\boldsymbol{A}$: 航班地面保障网络邻接矩阵; epoch: 插补值最大更新迭代次数; $\boldsymbol{M}$: 遮蔽矩阵。

2. **输出 $\boldsymbol{X}'$**: 插补后的航班地面保障数据;

3. $\boldsymbol{X}_{t_i}^d \leftarrow \text{fcn}\left(\left\{x_{t_i}^d, \forall i \in (1,2,\cdots,n), d \in (1,2,\cdots,D)\right\}\right)$ //扩充维度

4. for $v \leftarrow 1$ to epoch do

5.　for $i \leftarrow 1$ to n do

6.　　$\boldsymbol{X}_{t_i} \leftarrow \left\{\boldsymbol{Y}_{t_i}^1, \boldsymbol{Y}_{t_i}^2, \cdots, \boldsymbol{Y}_{t_i}^p\right\}$//依据时序网络图中存在的通路采样出数据子块

7.　　for $p \leftarrow 1$ to P do

8.　　$\tilde{\boldsymbol{Y}}_{t_i}^p \leftarrow \text{noise}\left(\boldsymbol{Y}_{t_i}^p\right)$//对子块添加噪声

9.　　$\bar{\boldsymbol{H}}_p \leftarrow \text{aggregator}\left(\tilde{\boldsymbol{Y}}_{t_i}^p\right)$//对采样到的子块进行第 1 次特征嵌入

10.　end for

11.　　$\bar{\boldsymbol{H}}_{t_i} \leftarrow \text{MeanPool}\left(\bar{\boldsymbol{H}}_1, \bar{\boldsymbol{H}}_2, \cdots, \bar{\boldsymbol{H}}_P\right)$//通过均值池化对子块进行聚合

12.　　$\boldsymbol{H}_{t_i} \leftarrow \text{LSTM1}\left(\boldsymbol{w}_1(\boldsymbol{H}_{t_i-1}, \bar{\boldsymbol{H}}_{t_i}, \boldsymbol{C}_{t_i-1}) + \boldsymbol{b}_1\right)$ //对特征向量进行第 2 次嵌入

13.　　$\boldsymbol{C}_{t_i} \leftarrow \text{LSTM2}\left(\boldsymbol{w}_2(\boldsymbol{H}_{t_i-1}, \bar{\boldsymbol{H}}_{t_i}, \boldsymbol{C}_{t_i-1}) + \boldsymbol{b}_2\right)$

14.　end for

15.　$\boldsymbol{H} \leftarrow \text{concat}(\left\{\boldsymbol{H}_{t_i}, i \in (1,2,\cdots,n)\right\})$

16.　$\hat{\boldsymbol{X}} \leftarrow \text{sigmoid}(\boldsymbol{LH\Theta})$//对特征向量进行反卷积还原回初始维度

17. end for

18. $\boldsymbol{X}' \leftarrow \boldsymbol{X}\cdot\boldsymbol{M} + (\boldsymbol{I}-\boldsymbol{M}) \odot \hat{\boldsymbol{X}}$

3　实验分析

3.1　数据来源与评价指标

为验证本文 IGITS-LSTM 算法的有效性, 选取西南某机场 2018 年 4～6 月的航班地面保障数据进行缺失值插补实验。其 2 个月间的航班数量约为 25 400 个, 保障节点的数据缺失率在 20% 左右。航班保障过程通常包含下客、卸载行李、清洁、燃油、餐食等 42 个保障节点, 但是由于航班机型, 旅客等因素的不同, 其实际保障过程会省略其中部分

节点，因此，在航班保障网络建图过程中以实际记录数据为准。

数据插补的准确性由3个指标进行度量，包括平均绝对误差（mean absolute error, MAE），平均绝对百分比误差（mean absolute percentage error, MAPE）和均方根误差（root mean square error, RMSE）。

3.2 插补误差分析

在实验数据上，应用了7种数据插补算法进行插补，分别是平均值插补（mean imputation, MI）、K最近邻插补 (K-nearest neighbor, KNN)、矩阵分解（matrix factorization, MF）、链式方程多重插补（multiple imputations by chained equations, MICE）及3种基于RNN改进的插补算法GRU-D、LSTM-M、IGITS-LSTM。通过比较MAE、MAPE和RMSE来评估不同算法的插补性能。在模拟随机缺失模式下，分别设定2类缺失率，低缺失率为10%和20%，高缺失率为50%和60%，在每种缺失率下重复实验10次以消除随机性带来的影响，分别计算10次3种指标并取平均值。其最终实验结果如表1和表2所示，并得到如下结论。

表1 低缺失率下每种算法的插补表现

Table 1 Imputation performance of each algorithm with low missing rates

算法	MAE/s		MAPE/%		RMSE/s	
	缺失率为10%	缺失率为20%	缺失率为10%	缺失率为20%	缺失率为10%	缺失率为20%
MI[3]	122.34	130.11	6.38	6.51	176.23	208.18
KNN[9]	63.49	75.32	3.17	3.59	101.33	128.04
MF[11]	55.38	63.91	2.51	3.12	88.60	106.09
MICE[10]	52.32	57.88	2.49	2.89	83.21	92.61
GRU-D[24]	23.42	27.31	1.17	1.14	38.64	45.06
LSTM-M[25]	39.24	42.55	1.96	2.12	66.38	68.93
IGITS-LSTM	**11.32**	**12.71**	**0.55**	**0.64**	**23.74**	**25.97**

表2 高缺失率下每种算法的插补表现

Table 2 Imputation performance of each algorithm with high missing rates

算法	MAE/s		MAPE/%		RMSE/s	
	缺失率为50%	缺失率为60%	缺失率为50%	缺失率为60%	缺失率为50%	缺失率为60%
MI[3]	135.72	141.12	6.78	7.05	217.12	225.81
KNN[9]	88.32	100.33	4.44	4.78	143.05	167.01
MF[11]	68.42	75.86	3.42	3.79	113.54	125.17
MICE[10]	67.71	78.91	3.38	3.76	115.11	128.62
GRU-D[24]	35.62	50.39	1.74	2.52	65.12	83.65
LSTM-M[25]	53.37	67.91	2.71	3.26	90.61	111.37
IGITS-LSTM	**15.22**	**23.21**	**0.73**	**1.77**	**25.42**	**59.86**

1）MI、KNN、MF、MICE这4种算法的插补准确度在高低2种类型的缺失率下，插补精度明显低于另外3种算法，其中，MI算法精度更是远远低于3种常规插补算法，这可能是由于均值插补不考虑节点之间的内在关联，以固定的值插补必然会带来较大的误差，尽管KNN、MICE和MF算法考虑到了部分节点的关联属性，插补精度明显提升，但只是通过笼统的聚类规则对节点特征进行了关联，因此，提升效果有限。

2）基于RNN的改进算法明显优于传统的插补算法，GRU-D、LSTM-M和IGITS-LSTM这3种算法由于其考虑到了数据的时间属性，其对节点特征的嵌入往往能建立更多的关联，捕获到长时序列中的信息传递规律，在对时序数据进行插补时能够带来较好的效果。

3）本文IGITS-LSTM算法明显优于其他算法，表中加粗部分的值表示最优结果，并且在高缺失率下这种差距变得更加明显，在10%和20%缺失率的情况下，与较先进的算法相比，IGITS-LSTM算法在插补精度上提升了约74%，而在50%和60%缺失率情况下，插补精度提升了约68%。这可能是本文算法中的GUR-A模块能够根据实际航班地面保障的业务流程进行建图，并通过两段图嵌入算法对保障节点信息进行深度提取，还原出的数据更能接近业务的真实情况。

4）尽管MI算法的数据插补性能不如其他几种算法，但是该算法在缺失率发生较大变化时，波动率却是最小的，不仅如此，基于RNN的算法在面对

缺失率从 20%～50% 的突变时波动变化均大于传统的数据插补算法, 而且在缺失率达到 60% 时, 所有算法插补精准度均有较大的下降, 这可能是数据量缺失较大对节点间特征关联的嵌入方法带来了一定影响。

3.3　基于插补数据的预测精度分析

为确认插补后的完整数据是否适合下游任务, 在 3.2 节 7 种算法插补后的数据集上训练了不同的回归模型, 包括决策树(decision tree, DT)回归器、线性回归器(liner regressor, LR)、随机森林(random forest, RF)回归器、支持向量机(support vector regression regressor, SVR)回归器和 RNN 回归器。上述回归器均通过 scikit-learn 实现, 并将航班地面保障数据的缺失率设置在 30%, 在上述 7 种数据插补算法对缺失的数据进行补充后, 使用 MAE 对回归器在不同算法插补结果进行评价, 结果如图 5 所示, 并可以得到如下结论。

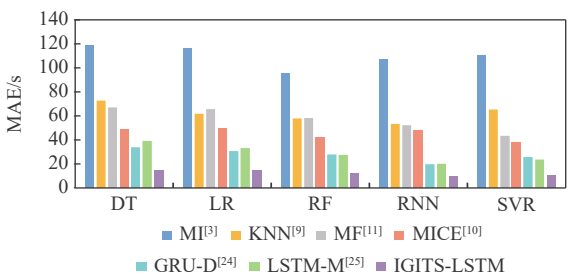


图 5　不同插补数据下不同回归模型训练得到的航班地面保障数据预测 MAE

Fig. 5　MAE of flight ground operation data prediction by different regression models trained on different imputed datasets

1) 基于 RNN 的回归模型预测指标明显整体略低于其他 4 种算法。尽管相同数据下不同预测方法的预测指标变化并不明显, 但说明了 RNN 在相同数据下预测精准度更高, 更适合航班地面保障这种时间属性的数据类型。

2) 在每种回归模型下, 插补精度较高数据大多数情况下会使得预测精度也有所提升。通过插补误差分析实验可以发现, 本文算法模型在大多数情况下能使航班保障数据的插补精度有一定的提升, 图 5 表明插补精度越高的数据下, 相同的回归模型执行预测任务效果整体呈现下降趋势, 这说明提升数据质量在大多数情况下能够对下游任务带来正向的增益。

3) 较高的数据质量, 较好的回归模型在某些情况并不一定能够带来更精准的预测结果。由图 5 可知, 在 SVR 回归模型下, 30% 的数据缺失率下, LSTM-M 插补算法带来了比 GRU-D 算法更精准的插补结果, 因此, 数据预测结果也更加精准, 但是在 DT 回归模型下结果却恰恰相反, 这说明较好的数据质量不一定能够带来较好的数据预测结果, 而在同样采取 MI 算法插补的数据上看 RNN 的预测精度甚至不如 RF, 这说明在较差的数据质量下较好的回归模型也不一定能够带来较精准的结果。出现这种现象的原因可能是在数据插补阶段的图嵌入与在预测阶段的特征提取存在一定的重合, 这种重合可能导致了模型的冗余, 使得模型预测效果进一步降低。

3.4　训练迭代次数的影响性分析

在训练阶段, 训练迭代的次数(epoch)总是会影响训练的结果, 因此, 需要对不同迭代次数下算法的插补精度进行实验。在训练开始前先对数据集进行处理, 分成 20%、30% 和 50% 缺失率下 3 种数据集, 再通过不同迭代次数下的模型对其进行数据插补, 观察精度的变化。如图 6 所示, 数据插补精度确实很大程度上收到了迭代次数的影响。在 20% 和 30% 缺失率下的数据集, 模型大约在 100 次迭代后达到最优精度, 在 50% 缺失率下, 模型大约在 110 次迭代后才达到最优精度, 而在较低缺失率下的模式训练已经出现过拟合; 另外在 50% 缺失下模型的训练在达到最优后呈现上下波动的趋势, 并不像在较低缺失下逐渐偏向过拟合。深度神经网络总是试图先学习数据中的低频信息然后在拟合像噪声一样的高频信息, 因此, 需要在训练阶段相对较早的停止神经网络的训练过程, 实际训练中需要尝试各种不同的 epoch 以获得最佳的训练结果。

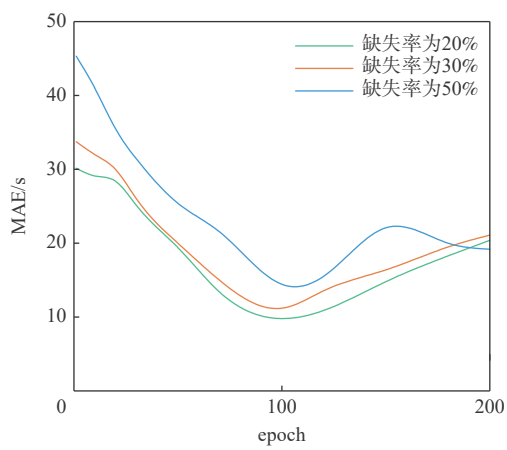


图 6　训练迭代次数的影响

Fig. 6　Effect of influence of training epoch

3.5　正则化系数影响性分析

正则化系数同样是影响模型训练的重要参数之一, 通过尝试不同的参数值来比较数据插补的结果。在 20%、30% 和 50% 数据缺失率下观察数据插补精度随超参数变化而变化的规律。如图 7 所示, 在不同缺失率下, 模型的插补精度均在 0.5 时达

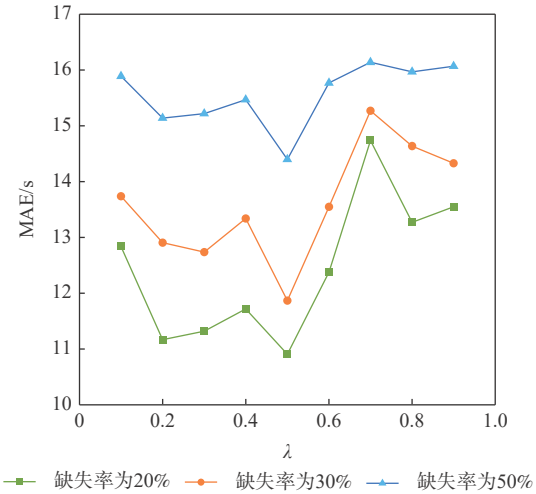


图7 正则化系数λ的影响

Fig. 7 Influence of regularization coefficient λ

到最优结果，而且随着值的增加，模型的插补精度变的越来越差。λ是用于平衡信息损失函数和对抗损失函数比率的参数，损失函数是用于更新模型参数的关键函数，因此，损失函数的偏差可能对训练结果带来巨大的影响，增加的值就相当于在损失函数中考虑了更多关于样本的对抗，而忽视了插补值本身的偏差，这样填补出来的数据也会逐渐偏离实际。

4 结 论

1）本文设计了一种基于降噪自动编码器框架的航班地面保障数据插补框架，通过在训练过程中对原始数据加入随机噪声可以提升隐藏层提取特征的稳定性，同时避免了训练插补模型前对数据的预处理阶段，符合机场航班地面保障数据缺失率较大的场景。

2）综合考虑航班地面保障实际业务过程，提出一种基于图嵌入的数据插补算法，根据航班地面保障网络中节点间的消息传递实现对缺失节点数据值的补充，同时结合多航班时间关联属性，进一步提取了特征向量，应用反卷积神经网络还原特征向量，在迭代多次模型后，可以得到相对准确的完整的航班地面保障数据。

3）使用西南地区某机场2018年4～6月份的航班地面保障数据进行仿真实验，模拟多种缺失率，进行结果对比，仿真结果表明，本文提出的缺失值插补算法相对于较新的算法，在10%～20%缺失率下的插补误差平均降低了约74%，在50%～60%缺失率下的插补误差平均降低了约68%，在模型预测方面，经过该算法插补后的数据在同类回归模型中也表现较好。合适的epoch和λ能够显著提升模型的插补效果。

由于本文模型仅仅考虑了航班地面保障数据，其数据的特征提取与插补机制源于航班保障流程的图结构特征，其适用性存在一定的局限，在未来的研究中，需要考虑如何将模型泛化到更多具有类似MTS属性的数据集中；保障模型在较高缺失率下也不会出现较大的插补偏差，并且由于本文只讨论了随机缺失模式下的插补效果，并没有对另外2种缺失模式进行验证，设计模型也没有考虑到缺失数据自身的属性，因此，挖掘数据缺失过程中的机理模型，可能有助于进一步提升模型的插补精度。

参考文献（References）

[1] SCHMIDT M. A review of aircraft turnaround operations and simulations[J]. Progress in Aerospace Sciences, 2017, 92: 25-38.

[2] DU J H, HU M H, ZHANG W N. Missing data problem in the monitoring system: a review[J]. IEEE Sensors Journal, 2020, 20(23): 13984-13998.

[3] JUNNINEN H, NISKA H, TUPPURAINEN K, et al. Methods for imputation of missing values in air quality data sets[J]. Atmospheric Environment, 2004, 38(18): 2895-2907.

[4] HAWTHORNE G, HAWTHORNE G, ELLIOTT P. Imputing cross-sectional missing data: comparison of common techniques[J]. Australian and New Zealand Journal of Psychiatry, 2005, 39(7): 583-590.

[5] BERTSIMAS D, PAWLOWSKI C, ZHUO Y D. From predictive methods to missing data imputation: an optimization approach[J]. The Journal of Machine Learning Research, 2017, 18(1): 7133-7171.

[6] RUBIN D B. An overview of multiple imputation[C]//Proceedings of the Survey Research Methods Section of the American Statistical Association. Princeton: Citeseer, 1988.

[7] BELLO A L. Imputation techniques in regression analysis: looking closely at their implementation[J]. Computational Statistics & Data Analysis, 1995, 20(1): 45-57.

[8] DEMPSTER A P, LAIRD N M, RUBIN D B. Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1977, 39(1): 1-22.

[9] COVER T, HART P. Nearest neighbor pattern classification[J]. IEEE Transactions on Information Theory, 1967, 13(1): 21-27.

[10] HUGHES R A, HERON J, STERNE J A C, et al. Accounting for missing data in statistical analyses: multiple imputation is not always the answer[J]. International Journal of Epidemiology, 2019, 48(4): 1294-1304.

[11] CHEN X Y, HE Z C, SUN L J. A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation[J]. Transportation Research Part C: Emerging Technologies, 2019, 98: 73-84.

[12] 曾小华, 宋美洁, 宋大凤, 等. 基于车联网信息的公交客车行驶工况数据处理方法[J]. 吉林大学学报 (工学版), 2021, 51(5): 1692-1699.
ZENG X H, SONG M J, SONG D F, et al. Data processing method of bus driving cycle based on vehicular network information[J]. Journal of Jilin University (Engineering and Technology Edition),

2021, 51(5): 1692-1699(in Chinese).

[13] KUPPANNAGARI S R, FU Y, CHUENG C M, et al. Spatio-temporal missing data imputation for smart power grids[C]//Proceedings of the Twelfth ACM International Conference on Future Energy Systems. New York: ACM, 2021: 458-465.

[14] 余嘉茵, 何玉林, 崔来中, 等. 针对大规模数据的分布一致缺失值插补算法[J]. 清华大学学报 (自然科学版), 2023, 63(5): 740-753.
YU J Y, HE Y L, CUI L Z, et al. Distribution consistency-based missing value imputation algorithm for large-scale data sets[J]. Journal of Tsinghua University (Science and Technology), 2023, 63(5): 740-753(in Chinese).

[15] WANG Y X, LI K, GAN S J, et al. Missing data imputation with OLS-based autoencoder for intelligent manufacturing[J]. IEEE Transactions on Industry Applications, 2019, 55(6): 7219-7229.

[16] CUI Z Y, LIN L F, PU Z Y, et al. Graph Markov network for traffic forecasting with missing data[J]. Transportation Research Part C: Emerging Technologies, 2020, 117: 102671.

[17] YAO X, GAO Y, ZHU D, et al. Spatial origin-destination flow imputation using graph convolutional networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(12): 7474-7484.

[18] YANG J M, PENG Z R, LIN L. Real-time spatiotemporal prediction and imputation of traffic status based on LSTM and Graph Laplacian regularized matrix factorization[J]. Transportation Research Part C: Emerging Technologies, 2021, 129: 103228.

[19] CINI A, MARISCA I, ALIPPI C. Filling the g_ap_s: multivariate time series imputation by graph neural networks[EB/OL]. (2022-02-10)[2023-05-08]. https://arxiv.org/abs/2108.00298v3.

[20] YOU J X, MA X B, DING D Y, et al. Handling missing data with graph representation learning[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 19075-19087.

[21] 陈小波, 陈程, 陈蕾, 等. 基于改进低秩矩阵补全的交通量数据缺失值插补方法[J]. 交通运输工程学报, 2019, 19(5): 180-190.
CHEN X B, CHEN C, CHEN L, et al. Interpolation method of traffic volume missing data based on improved low-rank matrix completion[J]. Journal of Traffic and Transportation Engineering, 2019, 19(5): 180-190(in Chinese).

[22] 张波, 宋国君. 大规模空气质量监测数据缺失处理方法实证研究[J]. 中国环境科学, 2022, 42(5): 2078-2087.
ZHANG B, SONG G J. Research on the missing value methods for large-scale online air quality monitoring data[J]. China Environmental Science, 2022, 42(5): 2078-2087(in Chinese).

[23] 陈帅, 赵明, 郭栋, 等. 基于 SVD-KDR 算法的工业监测数据插补技术[J]. 机械工程学报, 2021, 57(2): 30-38.
CHEN S, ZHAO M, GUO D, et al. Missing data imputation using SVD-KDR algorithm in industrial monitoring data[J]. Journal of Mechanical Engineering, 2021, 57(2): 30-38(in Chinese).

[24] CHE Z P, PURUSHOTHAM S, CHO K, et al. Recurrent neural networks for multivariate time series with missing values[J]. Scientific Reports, 2018, 8(1): 6085.

[25] TIAN Y, ZHANG K L, LI J Y, et al. LSTM-based traffic flow prediction with missing data[J]. Neurocomputing, 2018, 318: 297-305.

[26] LUO Y H, CAI X R, ZHANG Y, et al. Multivariate time series imputation with generative adversarial networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 1603-1614.

[27] SPINELLI I, SCARDAPANE S, UNCINI A. Missing data imputation with adversarially-trained graph convolutional networks[J]. Neural Networks, 2020, 129: 249-260.

[28] GUO Z J, WAN Y M, YE H. A data imputation method for multivariate time series based on generative adversarial network[J]. Neurocomputing, 2019, 360: 185-197.

[29] CHEN Y Y, LV Y S, WANG F Y. Traffic flow imputation using parallel data and generative adversarial networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(4): 1624-1630.

Imputation algorithm for flight ground support data based on graph neural network

XING Zhiwei[1], SUN Ke[1, 2], LUO Qian[2, *], LIU Chang[2], ZHANG Tao[2], QIAO Di[3]

(1. Electronic Information and Automation Institute, Civil Aviation University of China, Tianjin 300300, China;
2. Engineering Technology Research Center, The Second Institute of Civil Aviation Administration of China, Chengdu 610041, China;
3. Chengdu Tianfu International Airport Branch, Sichuan Airport Group Co., Ltd., Chengdu 610041, China)

Abstract: A data imputation algorithm based on a graph neural network is proposed to address the issue of missing flight ground support data. Firstly, to reduce the impact of noise in the original data on training denoising autoencoder is applied to enhance the reliability of feature extraction. Secondly, a graph representation learning framework is established to get the first embedding, using aggregation functions to aggregate the features of nodes within the sampling interval to achieve state updating. Furthermore, a long and short-term memory neural network is constructed to embed the temporal feature of flights to obtain the final state space of the hidden layer. Lastly, a loss function is suggested to iterate the deconvolution neural network, which is employed for feature restoration. The final flight ground operation data imputation result was acquired after numerous iterations, and the technique was evaluated using ground operation data from a specific airport in Southwest China from April to June 2018. The results showed that compared to other algorithms, the proposed algorithm imputation error decreased by an average of about 74% at low missing rates. At higher missing rates, the imputation proposed algorithm error decreased by an average of about 68%. When the number of iterations of the proposed algorithm is about 100 and the regularization coefficient is about 0.5, the imputation error reaches the lowest.

Keywords: flight ground operation; graph embedding; missing value imputation; recurrent neural network; autocoder; multivariate time series

Received: 2023-05-30; **Accepted**: 2023-10-16; **Published Online**: 2023-11-21 12:59

URL: link.cnki.net/urlid/11.2625.V.20231121.1044.001

Foundation items: National Key Research and Development Program of China (2018YFB1601200); Young Scientists Fund of National Natural Science Foundation of China (62203452); Civil Aviation University of China Research Innovation Project for Postgraduate Students (2022YJS105)

* **Corresponding author**. E-mail: cauc_xzw@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0297

高精度动态温度控制系统设计与验证

韩潇[1,2]，周盈[2]，黄海[1,*]，邵静怡[2]

(1. 北京航空航天大学 宇航学院，北京 100191；2. 北京卫星环境工程研究所，北京 100094)

摘 要：针对当前温度环境模拟设备难以实现深空探测中温度敏感类科学载荷的温度偏移试验问题，设计、搭建了一种高精度动态温度控制系统，并对其进行试验验证。该系统以液氮冷板作为冷源，通过电加热控温、热管传热及固件导热等方式，实现了300 mm×300 mm载荷接触面上的高精度静态和动态温度控制。试验结果表明：在科学载荷−10～45 ℃工作温度范围内，该温度控制系统与科学载荷的接触面可实现高精度的温度稳定控制，温度波动优于±0.001 ℃；同时，还可实现科学载荷接触面±0.05 ℃/1 000 s～1 ℃/1 000 s的升降温速率线性控制，与温度偏移试验设定的温度曲线相比，全程温度跟踪最大误差分别为0.003 ℃和0.04 ℃。所建温度控制系统具备高精度温度动态偏移试验能力，在深空探测载荷研制试验等任务中具有可观的应用前景。

关 键 词：动态温度模拟；高精度；升降温速率；真空低温；均温板

中图分类号：V416.5

文献标志码：A **文章编号**：1001-5965(2025)05-1539-09

高精度温度控制技术是航天器热控领域重要的研究方向，深空探测等领域的温度敏感类科学载荷对热控系统温度控制要求越来越高[1-4]，如“太极一号”卫星激光干涉测距系统核心载荷的温度稳定度指标为优于±0.1 ℃/1 000 s[5]；“天琴一号”卫星核心载荷温度稳定度指标为优于±0.05 ℃/轨道周期[6]。温度敏感类科学载荷及其热控技术的发展对热真空环境试验设备的温度控制水平提出了更高要求。针对当前地面热真空环境模拟设备难以实现深空探测中温度敏感类科学载荷的研制试验问题，本文设计、搭建高精度动态温度控制系统，对科学载荷温度进行周期性升降控制，以研究热控系统温度偏移对科学载荷的影响。

在常规航天器热真空试验中，通常采用液氮热沉模拟冷黑背景，为科学载荷等航天器被测试件提供低温环境[7-9]。由于液氮的温度一般低于−173 ℃，为实现被测试件的温度控制，除采用辐射方式进行热量传递之外，还可以通过控温冷板的方式，实现被测试件的温度控制。中国空间技术研究院刘春等[10-11]在2020年采用类似插片式散热器的结构，以液氮为冷源，依靠梯形截面的导热基块为热量传递通道，综合考虑控温效果、升降温能力、温度均匀性等参数指标，利用热分析软件仿真验证，实现了被测试件1 ℃/min的快速升降温速率控制，但未实现更小温度波动和更高精度调温的模拟控制。另外，采用高低温氮气作为控温介质也是一种常用的航天器热真空试验控温方式，兰州交通大学刘智勇等[12-13]通过金属狭缝内通入控温后的氮气，实现了控温表面的升降温控制，速率达到1 ℃/min以上，但该方案也未能实现长时间小温度波动的精确模拟。

能达到高精度温度调整的控制方法多用于恒温控温系统，如欧洲宇航局的“赫歇尔”卫星远红外光学系统的温度波动在0.5 ℃以内[14]；欧洲环境卫星(Envisat)的光学温控系统可通过电加热主动

收稿日期：2023-05-30；录用日期：2023-07-18；网络出版时间：2023-08-04 13:35
网络出版地址：link.cnki.net/urlid/11.2625.V.20230803.1828.002
基金项目：中国空间技术研究院杰出青年基金 (R-WY-JQRC-17)
*通信作者. E-mail：hhuang@buaa.edu.com

引用格式：韩潇，周盈，黄海，等. 高精度动态温度控制系统设计与验证 [J]. 北京航空航天大学学报，2025，51（5）：1539-1547.
HAN X，ZHOU Y，HUANG H，et al. Design and verification of high-precision dynamic temperature control system [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1539-1547（in Chinese）.

控温的方式将光学组件温度稳定在−17.9±0.05 ℃[15]。但此类方法是基于 PID 的恒温调节，并不能实现一定区间范围内的温度点的恒定速率升温或降温控制，迄今为止鲜有微小温度偏移波动（如在 1 000 s 内实现 0.05 ～1 ℃ 温度变化）相关模拟设备及其控制方法的公开文献报道。

针对深空探测中温度敏感类科学载荷的高精度温度偏移试验要求[16-18]，本文基于主动电加热控温、辅助导热散热的热控思想，提出一种模拟科学载荷受外部环境影响发生微小温度偏移的高精度动态温度控制方法，据此设计具备高精度动态控温能力的试验系统，实现对科学载荷最小波动度±0.05 ℃/1 000 s 至最大波动度±1 ℃/ 1 000 s 的工作环境温度模拟，建立了高精度温度偏移试验能力。

1 温度控制系统总体设计

航天器热控领域对温度环境模拟设备精度和灵活性的要求越来越严格，被测试件与冷源之间的温差也逐渐增大，同时还要求温度模拟系统具有多区域独立控温能力，可实现被测试件的高精度温度控制及周期性波动控制。

本文根据上述航天器热控领域对温控系统的要求，设计了高精度动态温度控制系统。该系统包括数据采集系统、参数控制系统及传热系统。其中，传热系统的设计问题是关键科学问题，因此，本文仅介绍传热系统的总体设计。

1.1 传热系统总体结构设计

为实现本文微小温度波动的精确模拟，需要精确地控制控温表面的热量流入流出，为解决仅采用辐射换热时散热量不足的问题[19]，通过合理的导热结构设计，可以实现较为可控的冷量供给，同时配合加热片模拟可控的升降温速率。

传热系统主要由不锈钢导冷结构、加热片、均温板及安装板组成，如图 1 所示。

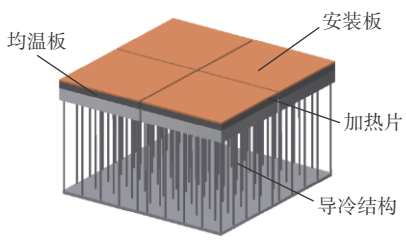


图 1 温控系统结构

Fig. 1 Structure of temperature control system

导冷结构由底板、杆阵列及顶板组成，材料为不锈钢。底板与液氮冷板直接接触，顶板上面为加热片。

加热片上面为均温板，共有 4 块，分别对应于 4 块安装板，每块安装板的温度可以按要求独立控制。

安装板上表面与被测试件底部接触面接触。被测试件为尺寸 300 mm × 300 mm × 300 mm 的立方体，结构如图 2 所示，底部接触面为不锈钢板。

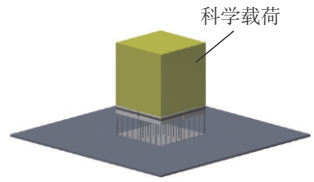


图 2 被测试验件结构

Fig. 2 Structure of test device

被测试件最大发热量为 30 W，需全部由接触面散热。试验时被测试件测试范围为−10～+45 ℃，且可实现接触面的分区控温。

1.2 传热系统结构参数设计

高精度动态温度控制系统传热结构设计的基本原理是以液氮冷板作为被动冷源，利用包括均温板在内的不同结构层之间的热传导，将被测试件的热量通过导冷结构带走，同时通过内部的电加热片进行冷热对抗，使被测试件稳定在目标温度上。

由于大温差和热量传递主要体现在温度控制结构的厚度 x 方向，如图 3 所示，因此，可以将结构的传热视为一维导热问题，导热微分方程为

$$\rho c_{\mathrm{p}} \frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(\lambda \frac{\partial T}{\partial x}\right) + q_{\mathrm{V}} \tag{1}$$

式中：x 为厚度方向的距离，m；t 为时间，s；T 为温度，K；ρ 为密度，kg/m^3；c_{p} 为比热容，J/(kg·K)；λ 为导热系数，W/(m·K)；q_{V} 为内热源，W/m^3，仅加热片层存在，其值由加热片功率决定。

稳定状态时，系统控制方程和边界条件为

$$\lambda \frac{\mathrm{d}^2 T}{\mathrm{d}x^2} + q_{\mathrm{V}} = 0 \tag{2}$$

$$-\lambda \frac{\mathrm{d}T}{\mathrm{d}x} = \alpha (T - T_{\mathrm{f}}) \qquad x = 0 \tag{3}$$

$$-\lambda \frac{\mathrm{d}T}{\mathrm{d}x} = \frac{Q_{\mathrm{h}}}{A} \qquad x = \delta \tag{4}$$

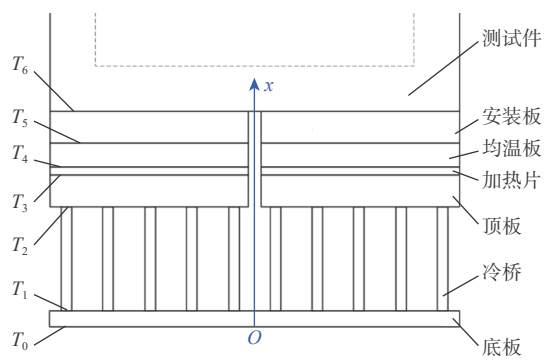


图 3 热传递过程示意图

Fig. 3 Schematic diagram of heat transfer process

式中: α 为液氮冷板传热系数, W/(m^2·K); T_f 为液氮温度, K; δ 为温控系统厚度, m; Q_h 为被测试件发热量, W; A 为被测试件散热面积, m^2。

对式 (2) 进行 2 次积分, 得

$$T=-\frac{q_V}{2\lambda}x^2+c_1x+c_2 \tag{5}$$

式中: 常数 c_1 和 c_2 由式 (3) 和式 (4) 确定, 最后, 可得温控系统的温度分布为

$$T=-\frac{q_V}{2\lambda}x^2+\left(\frac{q_V\delta}{\lambda}-\frac{Q_h}{\lambda A}\right)x+T_f-\frac{q_V\delta}{\alpha}+\frac{Q_h}{\alpha A} \tag{6}$$

根据式 (6) 可知, 通过不同结构层沿厚度方向的位置 x 及每层结构对应的特定的传热参数, 即可求得沿厚度方向任意位置的温度值。反之, 根据每层结构材料特定的设计温度值, 可以设计出结构的几何尺寸及所需的外部换热的边界条件。取不锈钢、铝导热系数分别为 16.2 W/(m·K)、202 W/(m·K), 考虑到控温结构在目标温度范围内均需要满足升降温速率的要求, 得到结构设计参数如表 1 所示, 其中, L、W、H 分别为各部分结构的长、宽、高。

表 1　传热系统结构设计参数

Table 1　Design parameters of temperature control system

部件	L/mm	W/mm	H/mm	数量
导冷结构底部	300	300	4	1
导冷结构中部	5	5	150	100
导冷结构顶部	148	148	4	4
加热片	148	148	0.2	4
均温板	148	148	5	4
安装板	148	148	6	4

2　传热系统主要部件设计

2.1　高效导热均温板设计

相变冷却均温板具有很高的导热性能、优良的等温性、热流密度可变性、热整流特性及环境适应性, 被广泛应用于电子元器件散热, 以及航天热管理系统[20]。

本文温度控制系统采用平板热管[21-22]作为均温板, 其具有非常大的导热系数, 能够迅速地减小温度梯度, 实现均温。

典型的热管由管壳、吸液芯和端盖组成, 将管内绝对压力降至 1.3×10^{-1}～1.3×10^{-4} Pa 后, 充入适量的工作液体, 使紧贴管内壁的吸液芯毛细多孔材料充满液体后加以密封, 管的一端为蒸发段(加热段), 另一端为冷凝段(冷却段), 当热管的一端受热时毛细芯中的液体蒸发汽化, 蒸汽在微小的压差下流向另一端放出热量凝结成液体, 液体再沿着多孔材料靠毛细力的作用流回蒸发段, 如此循环不断。

热管在设计时, 除传热性能应符合设计要求外, 还必须满足与管壳和工质的相容, 即在规定的寿命期内, 工质不应与管壳、管芯及焊接材料发生化学反应产生不凝气体。铝的导热系数和比热均很高, 是均温散热装置基板材料的理想选择。同时, 铝的密度为 2 700 kg/m^3, 约为铜或铁的 1/3, 具有质量轻、机械延展性能好、易于加工、耐低温等优点, 因此, 本文选用 6 系铝合金材料作为基体材料, 工质选用符合工作温度要求且与铝相容的丙酮, 充液率为 70 %, 毛细部分为泡沫镍烧结在管壁表面上, 尺寸设计如图 4 所示。

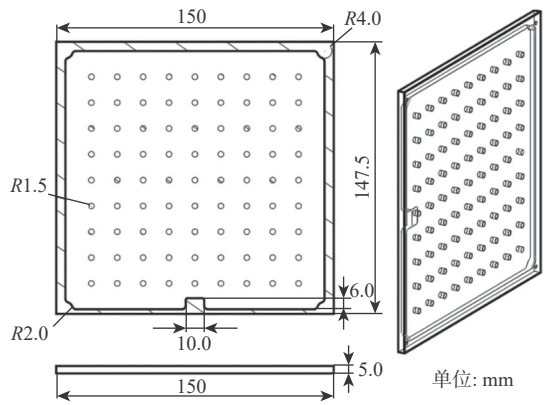


图 4　均温板几何尺寸

Fig. 4　Geometric dimensions of uniform temperature plate

2.2　高效导热均温板性能测试系统设计

图 5 为均温板的性能模拟测试系统。测试过程采用瞬态测试法, 即通过云母加热器模拟工件表面热源加热, 由恒温槽提供恒温载冷剂的冷板模拟冷源冷却, 同时利用恒温箱营造均温板在−20 ℃ 和 50 ℃ 的工作环境。

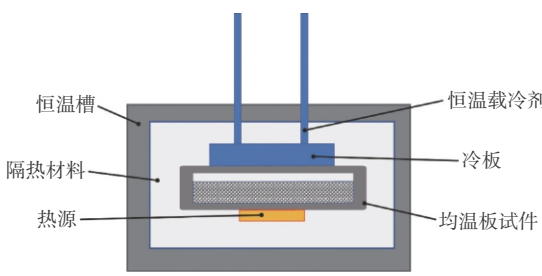
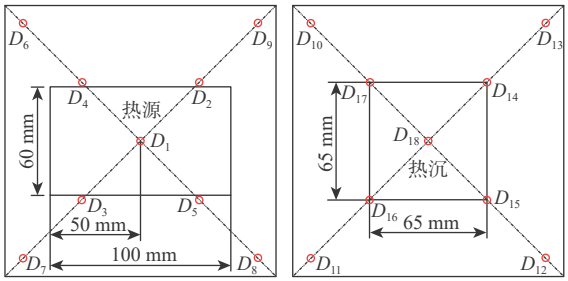


图 5　均温板性能测试系统

Fig. 5　Performance test system of uniform temperature plate

均温板上下面的测温点布置如图 6 所示, 其中, 热源温度为 D_1 测点的温度, 蒸发端平均温度为 D_1～D_9 这 9 个测点温度的平均值, 冷凝端平均温度为 D_{10}～D_{18} 这 9 个测点温度的平均值。

采用性能可靠、测量精度高、响应速度较快的 K 型热电偶, 利用高精度数据采集卡将热电偶测得的各测温点电信号反馈给数据采集仪, 由计算机将数据采集仪各通道读数编译为温度参数, 并对稳定



(a) D_1~D_9 (b) D_{10}~D_{18}

图 6 测温点布置示意图

Fig. 6 Layout of temperature measuring points

后的平均数据进行处理。

3 温度控制系统试验验证

3.1 高效导热均温板性能试验验证

试验过程中，热源的加热功率为 12 W 并保持不变，当恒温箱温度设定为−20 ℃ 和 50 ℃ 时，对应的载冷剂进口温度分别为−15 ℃ 和 30 ℃。最终得到各项试验数据如表 2 所示。

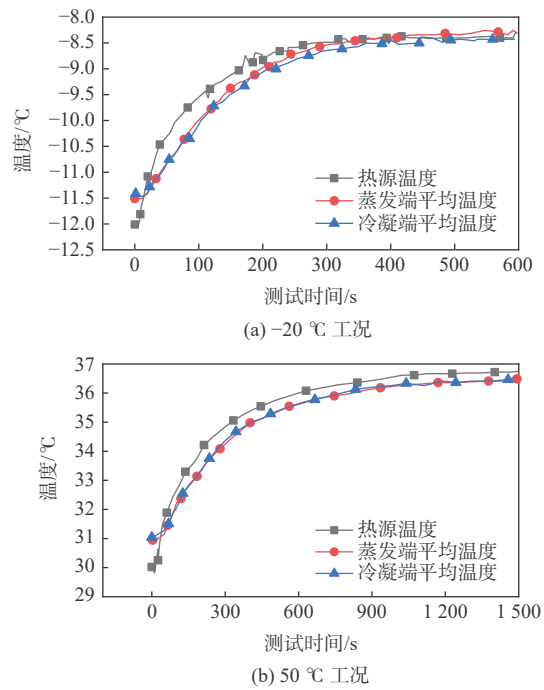
表 2 均温板性能测试

Table 2 Performance test of uniform temperature plate

测温点	测温点温度值 /℃	
	−20 ℃恒温环境	50 ℃恒温环境
D_1	−8.50	36.71
D_2	−7.54	36.98
D_3	−8.27	37.01
D_4	−8.07	36.10
D_5	−8.73	36.01
D_6	−8.39	35.73
D_7	−8.58	36.50
D_8	−8.89	36.55
D_9	−8.07	37.04
D_{10}	−9.23	35.95
D_{11}	−7.55	36.61
D_{12}	−7.40	37.32
D_{13}	−8.71	37.50
D_{14}	−9.81	36.19
D_{15}	−8.10	36.17
D_{16}	−8.21	36.25
D_{17}	−8.24	36.14
D_{18}	−8.48	35.06

由表 2 可知，均温板在−20 ℃ 和 50 ℃ 的环境下蒸发端平均温度分别为−8.34 ℃（标准差为 0.41 ℃）和 36.51 ℃（标准差为 0.48 ℃），冷凝端平均温度分别为−8.42 ℃（标准差为 0.76 ℃）和 36.35 ℃（标准差为 0.73 ℃）。蒸发端和冷凝端的平均温差分别为 0.08 ℃ 和 0.16 ℃，垂直方向当量热阻分别为 0.006 6 K/W 和 0.013 3 K/W，当量导热系数分别为 30 303 W/(m·K) 和 15 038 W/(m·K)，这都反映了均温板上下面较小的温度梯度和良好的均温能力，以及在小温差下实现高效导热的能力。

图 7 为均温板试样在不同工况下的温度启动曲线，由图可知，在−20 ℃ 的环境工况下，经过约 600 s 均温板温度达到稳定，试件热源温升 3.5 ℃。在 50 ℃ 的环境工况下，经过约 1 500 s 均温板温度达到稳定，试件热源温升 6.0 ℃。由此可知，均温板在相同加热功率下，−20 ℃ 环境工况比 50 ℃ 环境工况启动速度更快。同时从蒸发端平均温度和冷凝端平均温度随时间的变化也可以得知，两者之间的变化趋势相同，温差极小，这表明均温板的启动过程的均温性能也很好。



(a) −20 ℃ 工况

(b) 50 ℃ 工况

图 7 试样在−20 ℃ 和 50 ℃ 环境下的启动曲线

Fig. 7 Start curves of test sample at −20 ℃ and 50 ℃

3.2 高精度测量系统性能测试

为验证系统高精度动态温度控制及周期性波动控制的能力，设计了一个高精度温度测量系统。测量系统的分辨率、稳定性和精度取决于热敏电阻值测量的分辨率、稳定性和精度，所以先对测量系统阻值测量能力进行试验测试。

温度测量系统中热敏电阻器的温度特性按 GJB 601B-2018[23] 第 4.6.23 条规定，在−10 ～ 45 ℃ 温度环境下，选择−10 ℃、0 ℃、10 ℃、25 ℃、35 ℃、45 ℃ 共 6 个温度点进行试验。所选温度点的温度场由 Fluke 7012 恒温槽提供，温度特性测试结果如表 3 所示，R_1、R_2、R_3、R_4 分别表示 1#、2#、3#、4#热

表 3　热敏电阻温度特性

Table 3　Thermistor temperature characteristics

设定温度/℃	监测温度/℃	R_1/Ω	R_2/Ω	R_3/Ω	R_4/Ω
−10	−10.257	55 764.50	54 483.77	54 085.23	55 538.28
0	0.046	33 329.43	32 564.70	32 362.74	33 230.61
10	10.426	20 332.18	19 861.57	19 764.81	20 274.07
25	25.078	10 513.74	10 260.68	10 232.95	10 489.02
35	34.941	6 926.73	6 754.94	6 746.21	6 913.22
45	45.092	4 625.92	4 507.66	4 508.22	4 618.57

敏电阻器的实测阻值。

热敏电阻器校准方程采用式 (7) 所示的 Hoge-3 方程[24], 代入表 3 中的数据进行拟合。

$$\frac{1}{T}=\sum_{k=0}^{4}A_k(\ln R_{iT})^k \tag{7}$$

式中: A_k 为待拟合系数, 具体数值如表 4 所示; R_{iT} 为第 i#热敏电阻 (i=1,2,3,4) 在 T 温度下的阻值, Ω; 整数 k 分别取值 0, 1, 2, 3, 4。

表 4　热敏电阻器校准方程拟合系数

Table 4　Fitting coefficients of thermistor calibration equation

热敏电阻器	A_0	A_1	A_2	A_3	A_4
1#热敏电阻器	9.975 828 8	-1.286×10^{4}	7.366×10^{6}	-1.381×10^{9}	9.221×10^{10}
2#热敏电阻器	12.301 928 4	-1.561×10^{4}	8.557×10^{6}	-1.608×10^{9}	1.083×10^{11}
3#热敏电阻器	12.085 213 2	-1.536×10^{4}	8.463×10^{6}	-1.595×10^{9}	1.078×10^{11}
4#热敏电阻器	1.103 653 8	-2.355×10^{3}	2.711×10^{6}	-4.665×10^{8}	2.496×10^{10}

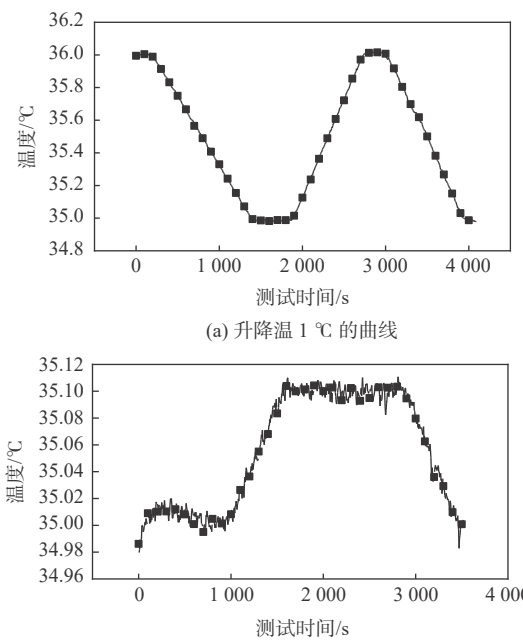
至此, 根据已知的传感器和测量系统稳定度测试, 可以对系统测量方法进行不确定度分析。通过拟合残差分析, 4 个传感器拟合残差最大为 0.914 mK, 传感器在−10～45 ℃ 区间内, 引入的温度不确定度为 U_{Rt} = 0.000 914 ℃。测量系统引入的电阻不确定度优于 0.079 2 Ω, 通过计算, 引入的温度不确定度 U_c = 0.000 1 ℃。电阻测量中使用的 Fluck 742A 标准电阻, 经过计量试验室标定, 该标准电阻实际电阻值为 10 000.02 Ω, 电阻测量引入的电阻误差为 0.001 Ω, 通过计算, 引入的温度不确定度 U_R = 0.001 ℃。根据不确定度合成原理, 如式 (8) 所示, 并综合上述分析结果可知: 温度测量系统合成不确定度为 0.003 ℃, 优于 0.005 ℃ 的温度测量精度要求。同时, 由于传感器测量电阻区间为 4 625.92～55 764.50 Ω, 使用 10 μA 恒流源进行电阻测量, 测量系统中选用的数字万用表电压测量精度为七位半, 即电压分辨率为 1×10^{-7} V, 换算到温度分辨率为 0.000 01 ℃。

$$U_t=K\sqrt{U_c^2+U_R^2+U_{Rt}^2} \tag{8}$$

式中: U_t、U_c、U_R 和 U_{Rt} 分别为系统综合不确定度、测量系统不确定度、标准电阻不确定度和传感器校准不确定度, ℃; K 为置信系数, 认为其符合正态分布, 取 K = 3, 置信水平为 99.7%。

3.3　大气环境下温度控制系统试验验证

系统搭建完成后, 在大气环境下进行功能和初步性能测试, 以验证整套温度控制系统硬件和软件功能状态。同时进行简单工况模拟, 初步验证静态和动态温度控制功能及算法的可行性。控制算法采用 PID 控制算法, 先采用常规的 pt100 温度传感器和数据采集仪器进行测试控制算法功能测试, 验证控制软件功能及算法功能是否满足设计要求。试验环境温度为 27 ℃, 测试温度为 35 ℃, 进行 1 ℃ 和 0.1 ℃ 温度波动控制测试, 结果如图 8 所示。



(a) 升降温 1 ℃ 的曲线

(b) 升降温 0.1 ℃ 的曲线

图 8　pt100 传感器温度波动控制

Fig. 8　Temperature fluctuation control of pt100 sensor

通过控制系统功能测试, 该系统可以完成不同波动度的控制, 波动时间可控, 但由于测量系统的稳定性, 采用普通的 pt100 及相应测温系统, 无法实现极小波动度的波动控制。

控制硬件软件功能测试完成后, 优化控制硬件及软件, 将高精度测量传感器及测量系统安装到试验件上, 进行样机整体大气测试。地面大气条件下 0.1 ℃ 波动的测量与控制曲线如图 9 所示。

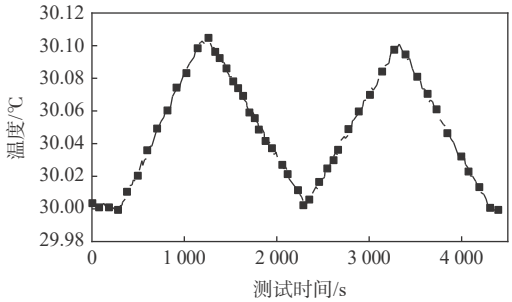


图 9 高精度传感器温度波动控制

Fig. 9 High precision sensor temperature fluctuation control

通过测试，控制系统可以实现周期性波动控制，波动时间可控，且重复性较好。温度测量系统稳定性与常规 pt100 温度测量相比有明显提升，但在温度模拟过程中测到明显的温度扰动，这是由于大气对流环境中，气流会对样机热特性带来微小不规律扰动，该高精度温度测量系统灵敏度极高，所以温度数据会有明显的微小波动。样机还需在真空低温环境下进行全温区的波动控制测试。

3.4 真空低温环境下温度控制系统试验验证

由于动态温度控制系统是在真空低温环境下使用，大气环境下无法准确模拟真空下系统实际温度变化，因此，需要在真空低温条件下验证安装面的静态控制和动态控制功能及控制效果。

图 10 为北京卫星环境工程研究所自研的一体机热真空试验设备，其可用来营造真空低温环境，在一体机内部可搭载调试样机的模拟装置。试验时，模拟装置为 4 个完全独立的可控安装面，为验证各可控边界的独立性，并且缩短试验时间，试验同时进行静态稳定性测试和动态测试。

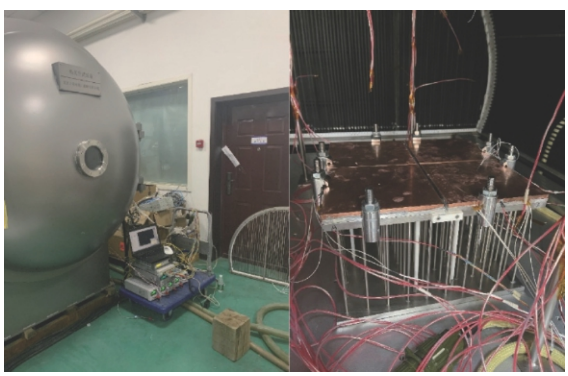


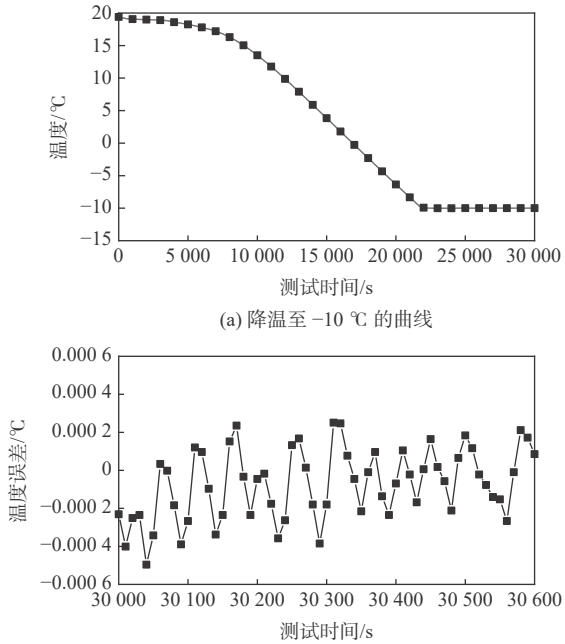
图 10 真空低温试验设备与测试系统

Fig. 10 Vacuum low temperature test facility and test system

3.4.1 静态温度控制验证

静态稳定性测试，选取−10 ℃ 和 45 ℃ 进行温度控制。常温～−10 ℃ 及−10～45 ℃ 的控温曲线分别如图 11 和图 12 所示。

综合静态温度控制曲线和静态控制误差曲线可以看出，安装面在−10 ℃ 和 45 ℃ 温度控制均无超调，各温度稳定后，控制精度优于±0.001℃。

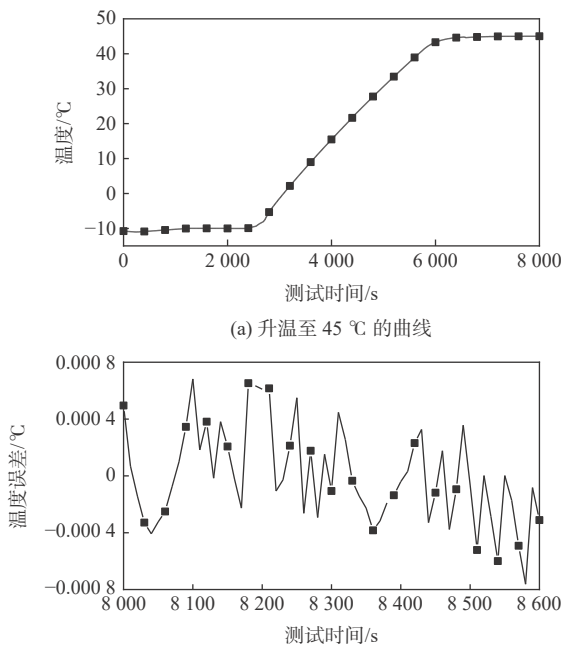


(a) 降温至 −10 ℃ 的曲线

(b) 稳定后与目标温度 (−10 ℃) 的偏差

图 11 降温曲线及−10 ℃ 静态误差

Fig. 11 Temperature drop curve and static stability error at −10 ℃



(a) 升温至 45 ℃ 的曲线

(b) 稳定后与目标温度 (45 ℃) 的偏差

图 12 升温曲线及 45 ℃ 静态误差

Fig. 12 Temperature rise curve and static stability error at 45 ℃

3.4.2 动态温度控制验证

动态温度控制试验主要验证系统在极端温度波动时的控制性能。本文选取±0.05 ℃ / 1 000 s(设计最小波动度)和±1 ℃ / 1 000 s(设计最大波动度)2 个极端波动度指标，分别在−10 ℃ 和 45 ℃ 环境下进行测试。验证结果如图 13～图 16 所示。

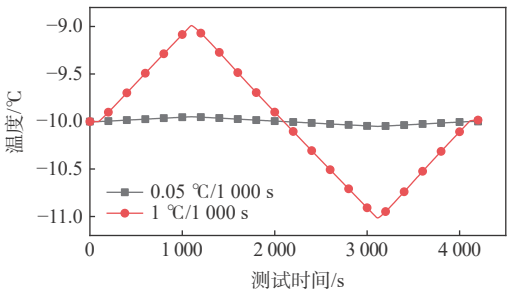
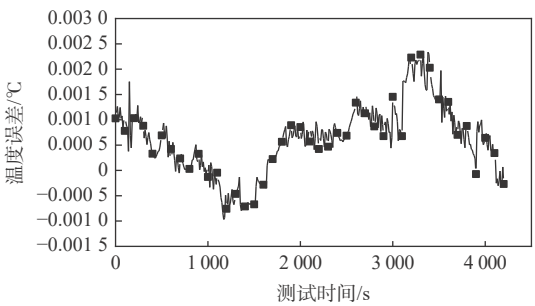
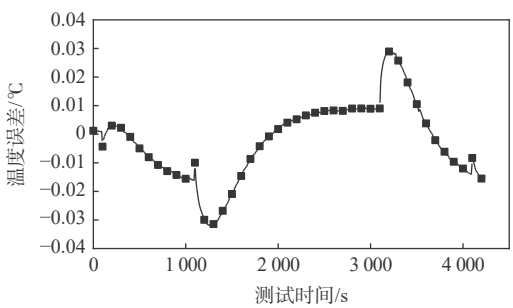


图 13　最小和最大波动度温度曲线(−10 ℃ 工况)

Fig. 13　Curves of minimum and maximum temperature fluctuation degree (−10 ℃ operating condition)



(a) 与目标曲线 (0.05 ℃/1 000 s) 的偏差



(b) 与目标曲线 (1 ℃/1 000 s) 的偏差

图 14　最小和最大温度波动误差曲线(−10 ℃ 工况)

Fig. 14　Curves of minimum and maximum temperature fluctuation error (−10 ℃ operating condition)

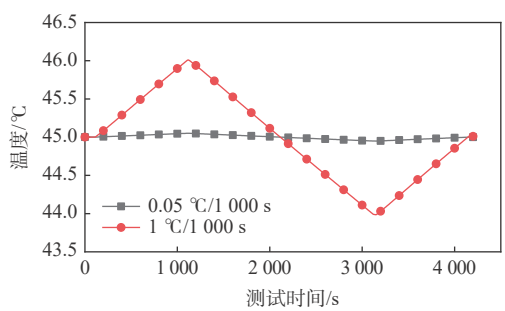
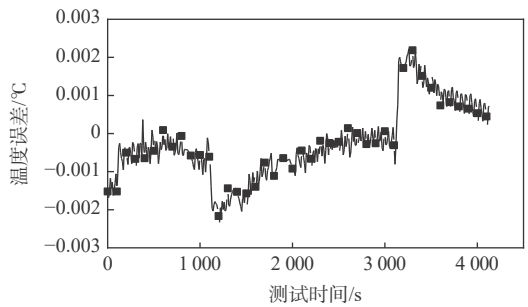


图 15　最小和最大波动度温度曲线(45 ℃ 工况)

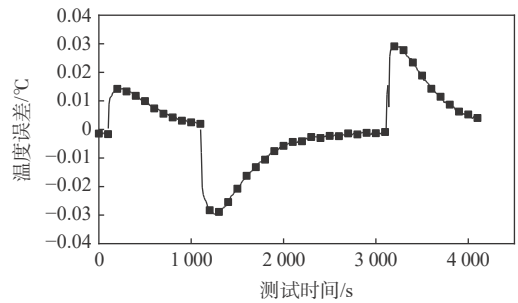
Fig. 15　Curves of minimum and maximum temperature fluctuation degree (45 ℃ operating condition)

由图 13 和 14 可知，该系统在−10 ℃ 工况下可以对最小和最大波动进行有效控制，误差分别在−0.001～0.002 5 ℃ 和−0.04～0.03 ℃ 范围内。

由图 15 和 16 可知，该系统在 45 ℃ 工况下也



(a) 与目标曲线 (0.05 ℃/1 000 s) 的偏差



(b) 与目标曲线 (1 ℃/1 000 s) 的偏差

图 16　最小和最大温度波动误差曲线(45 ℃ 工况)

Fig. 16　Curves of minimum and maximum temperature fluctuation error (45 ℃ operating condition)

可对最小和最大波动进行有效控制，误差分别在−0.003～0.003 ℃ 和−0.03～0.03 ℃ 范围内。

4　结　论

针对当前温度环境模拟设备难以实现深空探测等领域温度敏感类科学载荷的温度偏移试验问题，本文提出一种模拟科学载荷受外部空间环境影响发生微小温度偏移的高精度动态温度控制方法，基于冷桥导冷结构、主动电加热、多域均温板三结合的控温技术途径，建立了具备高精度静态和动态控温能力的试验系统，在大气及真空低温下验证了系统的可行性，实现了 300 mm × 300 mm 接触面上的高精度静态和动态温度控制，具体如下：

1）在真空、温度分别为−10 ℃ 和 45 ℃ 的环境下，进行常温～−10 ℃ 和−10～45 ℃ 的控温曲线验证，静态温度控制精度均优于±0.001 ℃；

2）设定温度曲线为 ±0.05 ℃/1 000 s 温升、温降时，真空、温度为−10 ℃ 环境下温度波动误差在−0.001～0.002 5 ℃ 范围内；真空、温度为 45 ℃ 环境下温度波动误差在−0.003～0.003 ℃ 范围内；

3）设定温度曲线为±1 ℃ /1 000 s 温升、温降时，真空、温度为−10 ℃ 环境下温度波动误差在−0.04～0.03 ℃ 范围内；真空、温度为 45 ℃ 的环境下，温度波动误差在−0.03～0.03 ℃ 范围内。

参考文献（References）

[1] CHUTANI R K, GALLIOU S, PASSILLY N, et al. Thermal man-

agement of fully LTCC-packaged Cs vapour cell for MEMS atomic clock[J]. Sensors and Actuators A: Physical, 2012, 174: 58-68.

[2] WANG B, LOU Y D, LIU J N, et al. Analysis of BDS satellite clocks in orbit[J]. GPS Solutions, 2016, 20(4): 783-794.

[3] WANG W, WANG Y P, YU C, et al. Spaceborne atomic clock performance review of BDS-3 MEO satellites[J]. Measurement, 2021, 175: 109075.

[4] 于峰, 徐娜娜, 赵振明. “高分四号”卫星相机在轨温度分析及热设计优化[J]. 北京航空航天大学学报, 2021, 47(1): 177-186.
YU F, XU N N, ZHAO Z M. On-orbit temperature analysis and thermal design optimization for camera on GF-4 satellite[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(1): 177-186(in Chinese).

[5] 刘红, 张晓峰, 冯建朝, 等. 精密热控技术在太极一号卫星上的应用[J]. 空间科学学报, 2021, 41(2): 337-341.
LIU H, ZHANG X F, FENG J C, et al. Application of precision thermal control techniques in Taiji-1 satellite[J]. Chinese Journal of Space Science, 2021, 41(2): 337-341(in Chinese).

[6] WEI R, ZHOU Y P, TONG Y L, et al. Ground experimental verification of high-stability temperature control system and flight data analysis[C]//Proceeding of the Signal and Information Processing, Networking and Computers. Berlin: Springer, 2023: 583-591.

[7] 龚洁. 基于G-M制冷机的大尺寸冷屏设计方法[J]. 航空动力学报, 2019, 34(2): 341-347.
GONG J. Design method of large dimension cold shield based on G-M cryocooler[J]. Journal of Aerospace Power, 2019, 34(2): 341-347(in Chinese).

[8] YANG S, DU C S, YANG X W, et al. Thermal control of primary mirror of space solar telescope[J]. Optik, 2021, 229: 166290.

[9] MENG Q L, ZHAO Z M, ZHANG T, et al. Experimental study on the transient behaviors of mechanically pumped two-phase loop with a novel accumulator for thermal control of space camera payload[J]. Applied Thermal Engineering, 2020, 179: 115714.

[10] LIU C, LI Z M, LIN B Y, et al. Analytical study on temperature control of fixture with multi-constraints in thermal vacuum test[C]//Proceeding of the IEEE 6th International Conference on Control Science and Systems Engineering. Piscataway: IEEE Press, 2020: 178-183.

[11] 刘春, 林博颖, 武飞, 等. 一种用于单机热真空试验快速升降温的调温平台设计及应用[J]. 航天器环境工程, 2020, 37(2): 166-171.
LIU C, LIN B Y, WU F, et al. Design and application of a temperature control platform used for rapid heating and cooling in thermal vacuum test[J]. Spacecraft Environment Engineering, 2020, 37(2): 166-171(in Chinese).

[12] 祁雪琴, 刘智勇, 颜昌林. 鱼骨式热沉调温系统的数值模拟与试验研究[J]. 低温工程, 2021(2): 25-32.
QI X Q, LIU Z Y, YAN C L. Numerical simulation and experimental research of fishbone heat sink temperature adjusting system[J]. Cryogenics, 2021(2): 25-32(in Chinese).

[13] 冯宁, 刘智勇, 颜昌林, 等. 流速对鱼骨式与板式调温热沉性能的影响[J]. 兰州交通大学学报, 2022, 41(5): 115-122.
FENG N, LIU Z Y, YAN C L, et al. Effect of flow velocity on the performance of fishbone type and plate type temperature regulating heat sink[J]. Journal of Lanzhou Jiaotong University, 2022, 41(5): 115-122(in Chinese).

[14] LEROY C, MAISONNEUVE M, PIAT M, et al. Simulation of the Planck-HFI thermal control system[C]//Proceeding of the Modeling, Systems Engineering, and Project Management for Astronomy III. Bellinham: SPIE, 2008: 701713.

[15] LEMMEN M, KOUWEN J, KOOREVAAR F, et al. In-flight results of the sciamachy optical assembly active thermal control system[C]//Proceeding of the International Conference On Environmental Systems. Warrendale: SAE International, 2004.

[16] 申春梅, 于峰, 刘文凯. FY-3D卫星高光谱温室气体监测仪热控设计及在轨验证[J]. 北京航空航天大学学报, 2020, 46(11): 2026-2038.
SHEN C M, YU F, LIU W K. Thermal control system design and on-orbit verification of hyperspectral greenhouse gas monitor on FY-3D satellite[J]. Journal of Beijing University of Aeronautics and Astronautics, 2020, 46(11): 2026-2038(in Chinese).

[17] 刘中华. 热真空试验标准与方法分析[J]. 电子产品可靠性与环境试验, 2016, 34(4): 16-20.
LIU Z H. Analysis of the standards and methods of thermal vacuum test[J]. Electronic Product Reliability and Environmental Testing, 2016, 34(4): 16-20(in Chinese).

[18] 李春林. 空间光学遥感器热控技术研究[J]. 宇航学报, 2014, 35(8): 863-870.
LI C L. Research on space optical remote sensor thermal control technique[J]. Journal of Astronautics, 2014, 35(8): 863-870(in Chinese).

[19] XIAO H, YING Z, JIA R J, et al. Design of temperature fluctuation test platform based on radiation heat transfer[C]//Proceeding of the MATEC Web of Conferences. Paris: EDP Sciences, 2019, 256: 03002.

[20] 杨涛, 赵石磊, 高腾, 等. 航天分散热源控温用环路热管设计及飞行应用[J]. 宇航学报, 2021, 42(6): 798-806.
YANG T, ZHAO S L, GAO T, et al. Design and in-orbit application of temperature controlled loop heat pipe for aerospace distributed heat sources[J]. Journal of Astronautics, 2021, 42(6): 798-806 (in Chinese).

[21] ZHONG G S, TANG Y, DING X R, et al. Experimental study of a large-area ultra-thin flat heat pipe for solar collectors under different cooling conditions[J]. Renewable Energy, 2020, 149: 1032−1039.

[22] 李德富, 刘小旭, 邓婉, 等. 热管技术在航天器热控制中的应用[J]. 航天器环境工程, 2016, 33(6): 625-633.
LI D F, LIU X X, DENG W, et al. Application of heat pipe technology in spacecraft thermal control[J]. Spacecraft Environment Engineering, 2016, 33(6): 625-633(in Chinese).

[23] 中央军委装备发展部. 热敏电阻器通用规范：GJB 601B-2018[S]. 北京: 工业与信息化标准网, 2018: 19-20.
Central Military Commission Equipment Development Department. General specification for thermistor: GJB 601B-2018[S]. Beijing: Industry and Information Standards Net, 2018: 19-20(in Chinese).

[24] LIU G, GUO L, LIU C L, et al. Evaluation of different calibration equations for NTC thermistor applied to high-precision temperature measurement[J]. Measurement, 2018, 120: 21-27.

Design and verification of high-precision dynamic temperature control system

HAN Xiao[1, 2], ZHOU Ying[2], HUANG Hai[1, *], SHAO Jingyi[2]

(1. School of Astronautics, Beihang University, Beijing 100191, China;
2. Beijing Institute of Spacecraft Environment Engineering, Beijing 100094, China)

Abstract: A high-precision dynamic temperature control system is proposed, developed, and tested with the goal of addressing the issue that the existing ground temperature environment simulation equipment finds it challenging to realize the temperature deviation test of scientific loads with extremely high temperature control accuracy in deep space exploration. The system uses a liquid nitrogen cold plate as the cold source, and it realizes the high precision static and dynamic temperature control on the 300mm×300mm contact surface by means of electric heating, heat pipe heat transfer, and solid structure heat conduction. The test results demonstrate that the temperature control system can accomplish high-precision temperature stability control with a temperature variation of less than ±0.001℃ within the contact surface temperature range of −10～45 °C. In addition, it can also realize the linear control of temperature rise and fall rate of ± 0.05 ℃/1 000 s～1 ℃/1 000 s on the load contact surface. Compared with the set temperature curve, the whole-process temperature tracking error is less than 0.003 ℃ and 0.04 ℃. The established temperature control system has the ability to conduct high-precision temperature dynamic deviation tests. It will have good application prospects in tasks such as deep space exploration payload development and testing.

Keywords: dynamic temperature simulation; high-precision; temperature rise and fall rate; vacuum and cryogenics; uniform temperature plate

Received: 2023-05-30; **Accepted**: 2023-07-18; **Published Online**: 2023-08-04 13:35

URL: link.cnki.net/urlid/11.2625.V.20230803.1828.002

Foundation item: China Academy of Space Technology Science Fund for Distinguished Young Scholars (R-WY-JQRC-17)

* **Corresponding author**. E-mail: hhuang@buaa.edu.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0267

振动对锂离子电池循环及热失控特性的影响

张青松[1,*]，李东琪[1]，杨娟[1,2]

(1. 中国民航大学 民航热灾害防控与应急重点实验室，天津 300300；2. 中国民航大学 工程技术训练中心，天津 300300)

摘　　要：基于交通运输及航空航天等应用场景下机械振动对锂电池安全性的影响，以18650型电池为样本，开展了振动工况对锂电池内部结构、循环老化机制、电气性能及热失控特性影响的研究。结果表明：振动会造成锂电池内部微观结构发生明显变化，长期振动应力积累会导致负极集流片发生形变，从而对电极卷形成挤压；瞬时振动应力会使电极材料与隔膜接触面积发生动态变化，以致锂离子在嵌入和脱嵌过程中不能充分迁移，加剧了库伦效率及能量效率的波动性，造成振动状态下放电不充分现象。热失控实验揭示：随着振动时间的增加，热失控触发时间提前，600 h振动的电池相较不振动电池热失控初爆时间提前了72 s，燃爆时间提前了114 s，初爆及燃爆间隔缩短。同时，燃爆产生的冲击力变化不稳定性增大。

关 键 词：锂离子电池；安全；振动；循环性能；热失控

中图分类号：X951

文献标志码：A　　**文章编号**：1001-5965(2025)05-1548-09

锂离子电池具有高能量密度、寿命长、环保无污染等特点，广泛应用于航天航空、交通运输及便携电子产品等领域。随着世界各国对环境保护要求的提高，锂离子电池的市场需求量逐年增长。在航空领域，锂离子电池不仅用于飞机启动电源[1]，也是航空货运的重要物资。特别是随着垂直起降飞机技术的日趋成熟，锂电池被广泛地用于全电力飞机的驱动能源[2]。基于航空运输对安全的严格要求，以及电池自身的安全风险，锂电池航空安全已经成为航空运输业亟待解决的焦点问题。新能源汽车方面，仅2021年，中国动力电池装车154.5 GWh，同比增长142.8%[3]，然而，因锂电池滥用、机械滥用及电池自身质量问题导致的燃爆事故频发，造成严重的生命、财产损失，因此，锂离子电池应用安全问题也日益受到国内外学者广泛关注，尤其是不同工况、环境、状态影响下锂电池安全性研究，如荷电状态(state of charge, SOC)[4]、老化程度[5]、低压环境[6-7]、涉水环境[8]等。

振动工况方面，无论是应用于航空领域还是新能源汽车，锂电池总处于持续的机械振动条件下，这可能会导致锂电池结构、电气性能等方面发生一定程度的变化，Hooper等[9]通过六自由度多轴振动台对商用电池施加机械振动，发现经历振动后的电池直流电阻在统计学上显著增加，其中，对阻值影响较大的为z轴(垂直)振动。Brand等[10]对锂电池进行了180天的长时振动，结果表明，垂直方向的长时振动会导致内部芯轴发生松动，集流片损坏并发生内部短路迹象。Parasumanna等[11]研究了振动工况前后的磷酸铁锂电池内阻变化及正负极形态变化，发现振动后内阻增加了25 %，同时负极材料中沉积了Fe和P元素，形成了很多塔状结构。Zhang等[12]在95 %的置信水平下对锂电池进行振动，电池的1C容量有所降低，同时基于多特征量，采用聚类算法分析发现，振动工况降低了电池的一致性。

收稿日期：2023-05-23；**录用日期**：2023-07-15；**网络出版时间**：2023-07-25 09:44
网络出版地址：link.cnki.net/urlid/11.2625.V.20230724.1828.002
基金项目：中央高校基本科研业务费专项资金(3122024058)；中国民航大学研究生科研创新资助项目(2022YJS009)
*通信作者. E-mail：nkzqsong@126.com

引用格式：张青松，李东琪，杨娟. 振动对锂离子电池循环及热失控特性的影响[J]. 北京航空航天大学学报，2025，51(5)：1548-1556. ZHANG Q S，LI D Q，YANG J. Influence of vibration on cyclic and thermal runaway characteristics of lithium ion batteries[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1548-1556 (in Chinese).

Lee 等[13] 对 3 种不同的电池进行了振动测试, 对比结果表明, 采用镍钴锰(nickel manganese cobalt, NMC)活性材料的电池相比较而言衰减较小。Somerville 等[14] 对受到振动的 NMC 电池进行了 X 射线光电子光谱学 (X-ray photoelectron spectroscopy, XPS) 测试, 结果显示, 振动导致电池在第 1 周期中形成的表面膜被去除, 并由电解液分解产生的表面膜代替, 而这种表面膜浓度往往高得多, 这种化学机制造成了电池老化水平增加。一些学者对多工况进行了耦合研究, Zhang 等[15] 采用正交分析法研究了温度、振动耦合条件下的放电性能, 指出温度、振动频率及振动方向对电池性能皆有影响, 但温度影响较为显著。总之, 现有研究普遍基于短期振动行为, 缺少长期振动与循环老化耦合实验及长期振动电池热稳定性方面的研究。本文针对锂电池振动工况下的循环老化过程开展了综合性实验研究, 重点阐明了振动工况条件对锂离子充放电性能和效率的影响, 并基于计算机断层扫描(computed tomography, CT)技术的内部结构无损探伤及电池热失控特征参数分析, 提出振动工况下锂电池的安全寿命、安全防控方法与措施, 为航空锂电池安全应用提供技术基础和重要实验数据。同时, 该研究在提升新能源汽车电池组件管理水平和解决航空运输安全技术方面具有指导意义。

1　实　验

1.1　实验流程

选用 18650 型圆柱锂电池, 电池参数如表 1 所示, 电池正负极材料分别为 $LiNi_{0.5}Co_{0.2}Mn_{0.3}O_2$ 和石墨, 电解液由六氟磷酸锂、甲基碳酸乙酯(ethyl methyl carbonate, EMC)、碳酸乙烯酯(ethylene carbonate, EC)和碳酸二甲酯(dimethyl carbonate,DMC)组成。正极封装采用钢制正极帽及铝制安全阀, 电池内部中心具有芯轴结构。为研究振动工况对锂电池的影响机制, 本实验设计了振动工况实验及循环-振动耦合实验, 同时借助 CT 技术、电阻抗扫描成像(electrical impedance scanning, EIS)、热失控实验等对振动工况下电池的结构损伤、电气性能改变及热稳定性进行分析。测试详细流程如图 1 所示。

表 1　实验电池参数

Table 1　Experimental battery parameters

类型	标称电压/V	截止电压/V	容量/(mA·h)	长度/mm	高度/mm
三元锂	3.6	2.75	2 200	65.0	18.4



图 1　实验流程示意图

Fig. 1　Schematic diagram of experimental process

1.2　振动实验

振动实验采用 HT-ZD2-400 型振动平台, 可实现基于调频调幅的程序化控制模式。通过自主设计的刚性电池夹具将实验电池固定在振动平台上。振动方案参考《联合国危险物品运输试验和标准手册》中 UN38.3 部分的 T3 测试[16], 参数如表 2 所示, 其中, A_1、A_2 为加速度幅度, m/s^2, S 为位移幅度, mm; f_1 为交越点频率, 约为 50 Hz。首先, 从 7 Hz 开始, 保持 1g_n 的最大加速度, 其中, g_n 为标准加速度, 直到频率达到 18 Hz; 然后, 将振幅保持在 0.8 mm (总偏移 1.6 mm), 并增加频率直到最大加速度达到 8g_n, 此时交越点频率为 f_1; 最后, 将加速度保持

在 $8g_n$ 直到频率增加到 200 Hz，该标准目前应用较为广泛，涵盖频率宽，加速度范围广、幅值大，能够反映实际中从轻微到恶劣的大范围振动环境。在此标准上进行长时间周期振动，分别设置了振动时间为 0, 100, 200, 300, 400, 500, 600 h 的梯度实验组，为确保后续测试的准确性和重复性，每个组别采用多节电池同时进行。

表 2 振动方案参数

Table 2 Vibration scheme parameters

频率	振动参数	对数扫频循环时间/ min	轴向
7～18	$A_1(1g_n)$	15	三轴
18 Hz～f_1	S(0.8 mm)	15	三轴
f_1～返回至7 Hz	$A_2(8g_n)$	15	三轴

1.3 循环-振动耦合实验

循环设备采用武汉市蓝电电子股份有限公司的 CT3001D 型充放电性能检测仪，通过外接导线连接至振动电池。在焊接连接镍片时为确保机械强度，应增加焊点数量及密度，同时做好绝缘措施。循环试验选取充放电倍率为 1C，模式为恒流恒压(CC-CV)，首先，对电池进行初始放电，以 1C 恒流放电至电池截止电压 2.75 V，静置 10 min，再以 1C 恒流充至电压为 4.2 V，静置 10 min 转恒压充电至充电电流小于 50 mA 为止，以上循环程序设定执行时间为 600 h，实验与振动过程同步进行。

1.4 CT 无损探测实验

CT 探测技术可以在不破坏电池整体结构的情况下观测内部变化，探知关键界面区域的详细情况，目前该技术广泛应用于医疗诊断、无损探伤、结构探测等领域[17]。本实验所用的 X 射线 CT 扫描设备为天津三英仪器公司的 nanoVoxel 显微 CT 装置，针对循环-振动组别电池进行 CT 结构测试，查明电池内部电极卷、芯轴、集流片、电极、安全阀等关键部件受振动影响情况。

1.5 EIS 测试

EIS 测试是一种用于研究电化学系统的测试方法，用以查明电化学系统的动力学特性、界面反应机制及电化学过程的速率限制等，实验采用型号为 BT4560 的电池阻抗测试仪，测试频率范围为 0.1～1 050 Hz。通过对不同振动时间的电池进行阻抗测试，绘制奈奎斯特图，选取合适的等效电路模型，通过拟合得出各部分的交流阻抗，进而分析振动对内部微观变化影响机制。

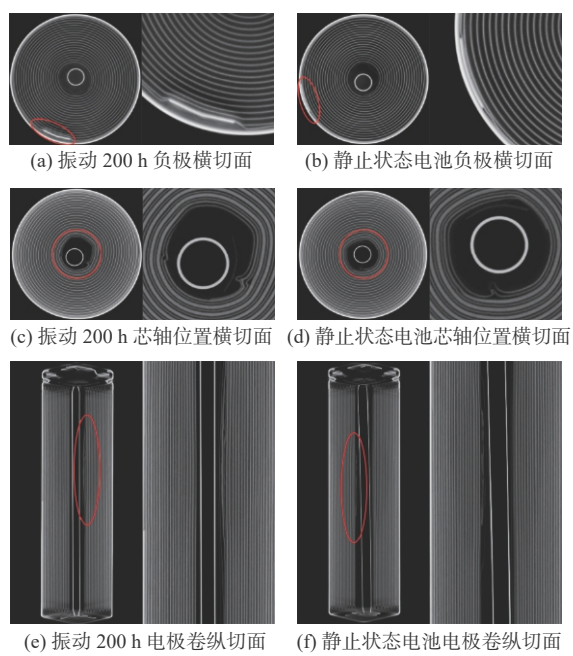
1.6 燃爆实验

对不同振动时间的电池分别进行热失控处理，通过热失控特性参数来分析振动工况对电池热稳定性的影响。实验基于自主搭建燃爆装置，由热失控舱、电池加热部件、K 型热电偶、排气装置、温控设备等主体组成。其中，热失控舱采用不锈钢制 3 L 碟形结构，顶部引出导气管和抽气泵，其具备进气、排气等功能。电池固定在加热座上，表面固定 1 个 K 型热电偶，用于测量电池表面温度变化，电池放置于燃爆舱中央。压力传感器固定于热失控舱顶部，用于测量热失控舱压力变化。热失控舱外接设备还包括了数据记录仪、废气排气装置、加热电源等。实验开始后，首先，通过外部加热(40 W)的方式诱发振动锂离子电池发生热失控，记录燃爆过程的各项参数；然后，待实验舱压力趋于稳定时，将产生的热解气体通过顶部废气排气装置排出；最后，通过收集的数据，对不同振动时间的锂电池燃爆危险程度进行评估。

2 结果与讨论

2.1 振动工况下锂电池内部损伤演变

本文通过 CT 技术，在未破坏电池的情况下，探究了振动引起的内部损伤机制。图 2 为 200 圈振动循环与静止循环下电池横向和纵向的 CT 测试图及对应红圈区域的放大图像，其中，图 2(a)为靠近电池负极区域，可以看出，振动导致了负极集流片松动，从而对电极卷产生了明显的挤压作用，而静止循环组电池如图 2(b)所示，并未发生形变。图 2(c)～图 2(f)分别展现了振动状态和静止状态下电池循环老化对电极卷产生的影响，比较图 2(c)和图 2(d)可知，2 种状态下电极卷靠近中心芯轴位置皆出现



(a) 振动 200 h 负极横切面 (b) 静止状态电池负极横切面

(c) 振动 200 h 芯轴位置横切面 (d) 静止状态电池芯轴位置横切面

(e) 振动 200 h 电极卷纵切面 (f) 静止状态电池电极卷纵切面

图 2 锂电池 CT 图像

Fig. 2 CT images of lithium ion battery

了褶皱变形[18], 但静止状态仅有一处轻微褶皱, 而振动状态则存在 2 处较为严重的褶皱形变, 同时涉及了多层电极, 对比图 2(e)和图 2(f)可知, 振动状态下循环对电极卷的影响程度远大于静止状态, 可见与横向效果相对应, 振动应力同样对电极卷纵向产生较大范围的影响。

2.2　振动对锂电池内阻的影响

图 3(a)为振动后电池的 EIS 曲线, 图像显示, 振动时间低于 400 h 时, 随着振动时间的延续, 半圆曲线与实轴交点变化较小, 但是半圆直径逐步扩大, 当振动时间大于 400 h 后, 与实轴交点开始向右平移, 此时半圆直径有变小趋势, 其中, Z'为阻抗实部, Z''为阻抗虚部。根据曲线特征选取如图 3(b)所示的等效电路, 其中, R_1 为电池的欧姆阻抗, R_1 增大时, 代表着电池内部传导效率的降低, 通常是由于电极腐蚀、溶解、开裂、黏合剂分解及接触失效等[19-20]。R_2 包含了固体电解质界面(solid electrolyte interphase, SEI)膜的阻抗及电荷转移阻抗, SEI 膜的增厚、SEI 膜与隔膜的亲和度降低及石墨层的脱落等现象皆会造成 R_2 的增加。R_w 则是 Warburg 阻抗, 其值一般会随着活性材料损失而变大。CPE 则是代表等效电路中的非理想电容。对各振动时间的阻抗谱曲线进行数值拟合, 得出各阻抗的拟合数值。图 4 为拟合阻抗的变化趋势, 可以发现, 随着振动时间的增加, R_1 在 400 h 之前趋于平稳, 400 h 后显著增大, R_2 则是在 400 h 之前增大, 400 h 之后

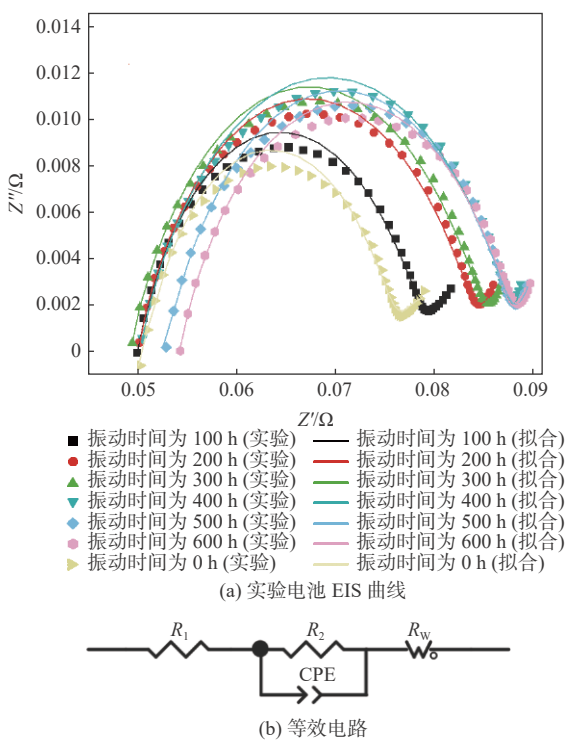


图 3　实验电池 EIS 曲线和等效电路

Fig. 3　EIS curves of experimental battery and equivalent circuit

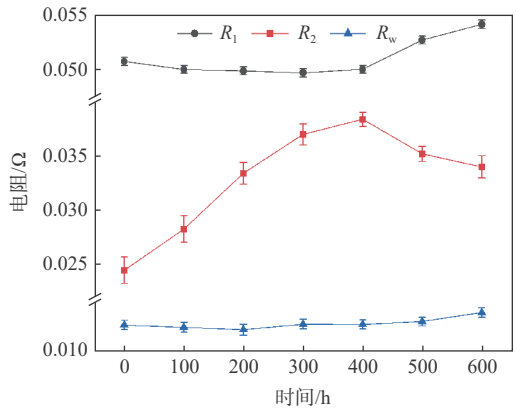


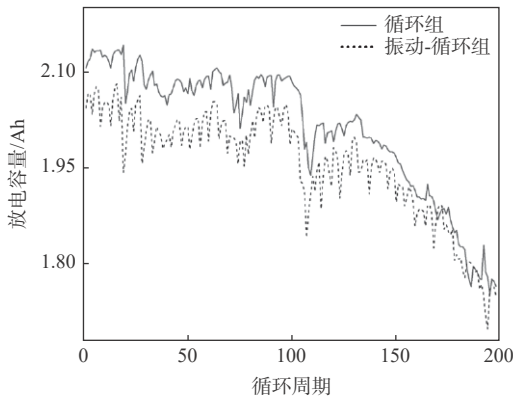
图 4　拟合阻抗数值变化

Fig. 4　Numerical variation of fitted impedance

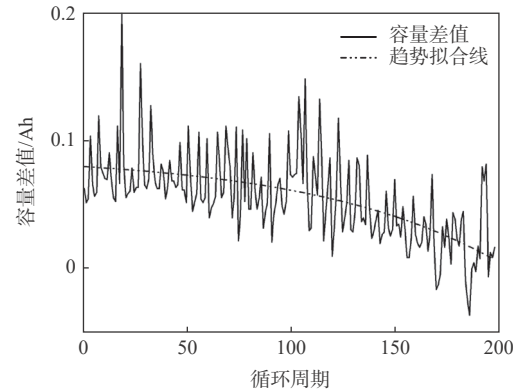
开始降低, 数据表明, 实验中 400 h 是一个关键的临界时间, 小于该时间时, 电池内部形变较小, 由于机械应力造成的结构变化、电极开裂、黏合剂脱落及接触失效现象不显著, 因此, R_1 变化较小, 但该阶段由于振动造成了电极层的松散、脱落及堆积[11], 同时粉化的锂化石墨将进一步与电解质结合[21], 使得 SEI 膜生成和增厚, 从而导致 R_2 增大; 当振动时间大于 400 h 时, 持续振动降低了集流片与电极结合力降低[22], 从而出现了接触失效, 此时 R_1 明显增加, 由于电极卷受到集流片向内形变的轻微挤压, 增加了局部贴合度, 导致 R_2 反而出现降低现象[23]。R_w 在振动过程中受影响较小, 变化较为稳定, 因此, 可以推断振动对于活性材料损失影响较小。

2.3　振动-循环耦合下锂电池关键电气性能

受振动影响, 锂电池的充放电性能发生了明显改变。通过图 5(a)放电容量随循环老化周期变化曲线发现, 在实验前期, 振动工况耦合造成了充放电过程中锂离子嵌入及脱嵌过程的不充分, 从而充放电效率下降, 同时振动时锂电池内部正负极与电解液浸润度降低[22], 接触面积减小, 活性材料出现松散粉化[24], 因此, 使得锂电池振动工况下的放电容量远小于静止状态, 且由于内部成分结构时刻受到机械应力影响, 造成放电曲线的波动程度明显大于静止状态。与前中期现象不同, 后期 2 条曲线呈聚拢趋势, 图 5(b)为放电容量差值曲线, 拟合趋势线直观的反应了 2 条放电容量曲线的靠拢程度, 全周期循环中, 放电容量差值始终呈减小趋势, 在 140 圈前变化较为平稳, 140 圈之后下降率显著增大, 此时静止循环组电池放电容量衰减程度接近振动-循环耦合电池。该现象说明了 2 个实验组在循环全周期中衰减路径存在差异性。锂电池老化机制主要包括了电导率损失(conductivity loss, CL)、活性锂离子损失(loss of active lithium ions, LLI)及活性材料损失(loss of active materials, LAM)[25], 对



(a) 放电容量变化曲线



(b) 放电容量差值曲线

图5 放电容量变化和放电容量差值

Fig. 5 Variation of discharge capacity and discharge capacity difference

于静止状态循环老化的电池，循环前期3种老化机制共同作用，但产生的影响贡献都较小，而后期，由于SEI膜不断增厚及析锂现象的加剧，致使活性锂离子的损失[26]，此时LLI成为了主要影响机制。而对于振动-循环下的锂电池，在循环前期，由于振动应力造成传导效率降低及锂离子嵌入和脱嵌过程不充分，因此，振动循环时容量饱满度受到影响，但并未造成活性锂离子的损失，同时，高容量区间循环老化程度比低容量更为显著，因此，在电池老化后期，静止状态相比振动状态的LLI老化机制影响程度更为严重。结合2.2节中阻抗分析，在振动后期由于机械应力作用造成电极卷轻微挤压形变，压实过程增大了锂离子传导效率，因此，也进一步减缓了放电容量衰减。但该结果并不代表振动对循环老化后期的影响具有正面效果，反而，随着形变继续演化加剧，由此引发的内短路风险将进一步增大[27]。

库伦效率(Coulombic efficiency, CE)是衡量锂电池充放电过程中容量转换的重要效率指标，一般是指放电容量与同循环中的充电容量之比，如下：

$$C_{\mathrm{E}} = C_{\mathrm{discharge}} / C_{\mathrm{charge}} \tag{1}$$

式中：$C_{\mathrm{discharge}}$为放电容量；C_{charge}为充电容量。

通常而言，库伦效率并不是完全恒定的，其往往受到多种因素影响，如环境温度、老化程度、充放电倍率、电解质成分及电极材料结构变化等。图6为振动状态电池与静止状态电池的库伦效率随循环周期变化的分布对比，其中，静止状态循环电池库伦效率收敛于100%效率上下，随着循环周期增长，在140圈后出现降低趋势，前后衰减率为3.4%。振动状态的电池则是在100%效率上下呈明显发散趋势，可以判断振动应力造成了电极材料与隔膜接触面积的动态变化及锂离子在嵌入、脱嵌过程中并不能充分迁移，而导致正负极材料上的锂离子留置，因此，造成库伦效率在96.72%～103.83%区间大幅度离散波动。

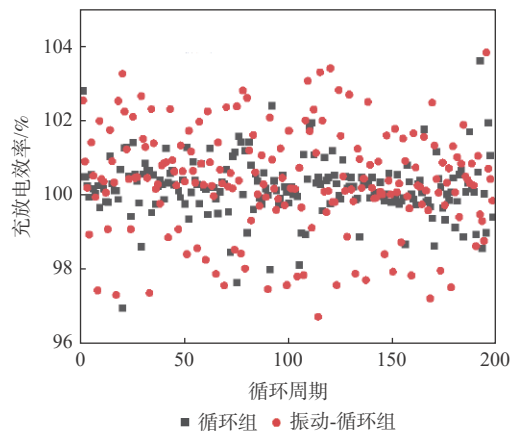


图6 库伦效率变化

Fig. 6 Variationof Coulomb efficiency

能量效率(energy efficiency, EE)是锂电池在充放电过程中用来衡量能量转化效率的指标，表示实际输出能量与输入能量的比值，如下：

$$E_{\mathrm{E}} = E_{\mathrm{discharge}} / E_{\mathrm{charge}} \tag{2}$$

式中：$E_{\mathrm{discharge}}$为放电能量；E_{charge}为充电能量。

对于锂电池，能量效率普遍位于80%～95%之间，具体受使用环境、充放电条件、充放电倍率及电池内部成分改变等因素影响。与库伦效率相比，能量效率关注了整体能量的转化，包含了能量损失的各种形式，而库伦效率侧重关注电荷的转移，主要考虑了电荷的流动和损耗。图7为振动和静止条件下锂电池循环过程的能量效率分布趋势。在全循环周期，2种条件下能量效率皆出现了衰减，静止状态和振动状态的最大衰减率分别为4.99%和6.27%，同时，通过分布点可以清晰发现，静止状态电池的能量效率始终要高于振动状态，且表现出波动范围较小、衰减较为稳定的特点，而振动工况下的能量效率波动范围更为离散，不确定程度更高。由此可知，振动会对电池的能耗性能产生较大影响，增加了系统能耗的同时，使得输出能量的不

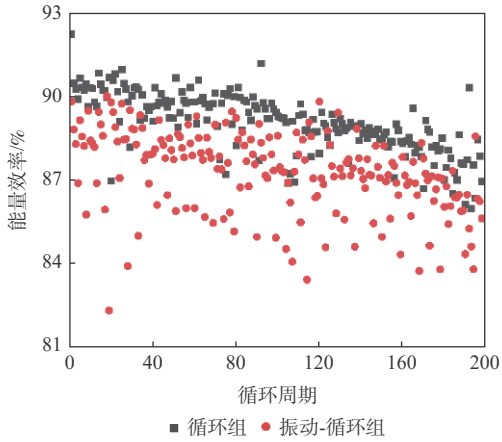


图 7　能量效率变化

Fig. 7　Variation of energy efficiency

稳定性增加。

容量保持率是用来衡量电池整体衰减的关键指标, 其值等于循环 n 圈以后的电池容量与初始容量的比值, 该指标反映了电池整体健康程度变化。图 8 为实验全周期容量保持率变化曲线及对应的保持率差值柱状图。200 圈循环后, 静止状态与振动状态容量保持率分别降至 83.44% 和 84.77%, 振动状态下电池保持率要高于静止电池。但通过曲线发现, 对于循环的不同阶段, 2 种状态的容量保持率变化并不一致。为清晰表现这一变化, 通过绘制容量保持率差值柱状图直观的展现了 2 种状况的差异对比。由图 8 可知, 140 圈循环的前后, 2 种状态表现出了不同的容量保持情况, 140 圈之前, 静止循环组的容量保持率要高于振动组, 而 140 圈之后, 静止循环组表现出了较为严重的老化效应, 其容量保持率逐步低于振动组。

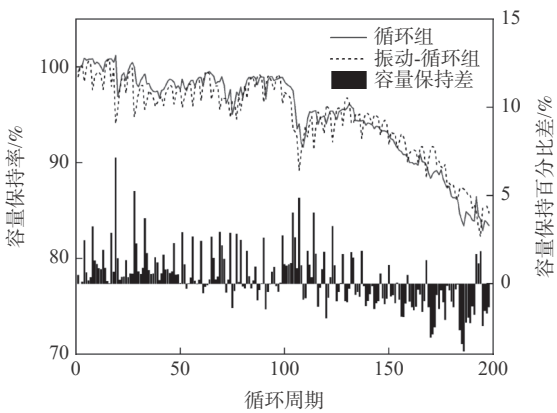


图 8　容量保持率变化

Fig. 8　Variation of capacity retention rate

以上分析可知, 振动电池的电气性能变化与振动应力的瞬时状态及积累作用密切相关。振动瞬时状态直接导致了充放电效率和能量效率不规则发散, 同时加剧了电池电气性能输出量的波动性。由于 140 圈循环所对应的振动时间位于 400～500 h 区间内, 可以判断振动循环耦合和单独振动实验所表现出的机械应力影响区间基本一致, 振动会导致电池内部成分结构时刻发生细微变化, 当应力积累达到一定量级时, 内部结构则会表现出不同程度的突变, 从而反应在电气性能上的变化。因此, 对于电池组设计, 应充分考虑振动对电池的潜在影响, 并采取相应缓振措施以提高电池的可靠性和安全性, 如通过填充减振材料可以一定程度吸收振动应力, 同时加装固定设备及防振支架, 借助螺栓固定、弹簧泄力等方式减小振动的传递效果。在电池运输方面, 使用专门设计的防震包装材料将电池进行包裹, 一定程度上可以保护电池不受外界振动的影响。

2.4　热失控特性

电池在振动后, 内部物质成分结构不同程度产生了改变。为查明振动对电池热安全性的影响, 本文将针对不同振动时间下的电池进行热失控实验, 采集热失控特性曲线, 分析影响机理, 以便对不同振动状态的电芯开展进一步的安全性评估。本文通过等功率的加热底座逐步升温来触发锂电池的热失控, 以此来模拟电池的热滥用。热失控过程划分为初爆和燃爆 2 个阶段, 随着温度的增加, 电池先进入初爆阶段, 初爆期间电池电解液受热不断蒸发, SEI 膜分解等反应产生气体导致电池内压力持续增加, 达到一定阈值时, 安全阀弹开, 内部气体放出, 电池表面温度小幅度降低, 此阶段还未触发电池完全燃爆, 因此, 对于电池热安全预警, 初爆可能是一个关键的风险识别点[28]。之后, 随着温度继续增加到某一值时, 电池内部发生不可控的链式放热反应, 彻底发生燃爆, 此时表面温度骤升, 从正极金属帽喷射出大量气体、粉尘、火焰及少量铝制熔珠, 环境压力大幅增加, 该过程持续时间较短, 但对外释放了大量热能, 待现象结束后, 电池及受限空间逐渐降温, 环境高压趋于平稳。

对不同振动时间的电池进行热失控实验, 图 9 为电池表面温度变化曲线, 7 组实验皆出现了初爆和燃爆过程, 初爆温度在 176 ～186 ℃ 之间, 燃爆温度在 615 ～695 ℃ 之间, 整体差异较小, 但初爆及燃爆触发时间存在较大变化。表 3 为各振动时间触发热失控初爆及燃爆时间。

随着振动时间的增加, 初爆、燃爆发生时间皆逐渐变小, 振动 600 h 的电池相比不振动的电池热失控初爆时间提前了 72 s, 燃爆时间提前了 114 s, 其热稳定性发生了显著降低。燃爆与初爆间隔也呈降低趋势, 尤其对于振动小于 300 h 的电池样本, 振动时间每增加 100 h, 初爆燃爆间隔提前 14 s 左

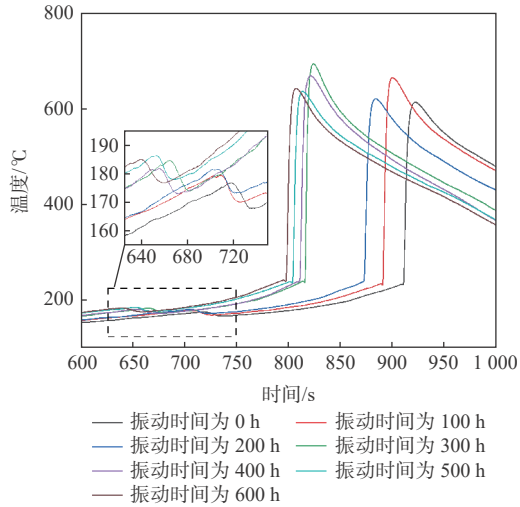


图 9 热失控温度变化

Fig. 9 Temperature variation of thermal runaway

表 3 热失控初爆、燃爆触发时间

Table 3 Trigger time of initial explosion and combustion-induced explosion of thermal runaway

振动时间/h	初爆时间/s	燃爆时间/s	燃爆、初爆时间差/s
0	718	921	203
100	709	900	191
200	705	881	176
300	663	825	162
400	659	822	163
500	653	813	160
600	646	807	161

右，当振动时间大于 400 h 后，该间隔不再随时间发生变化。

图 10 为热失控过程中燃爆舱内压力变化曲线，初爆时由于喷出少量气体，因此，环境压力出现了小幅度的增长，燃爆时由于大量热量及热解气体的喷出，造成环境压力大幅剧增，振动时间为 0，100，300，500，600 h 的电池燃爆压力峰值基本稳定

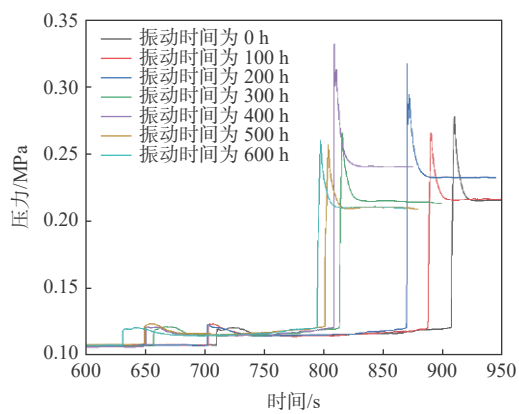


图 10 热失控压力变化曲线

Fig. 10 Pressure variation curves of thermal runaway

于 0.25 MPa 左右，而振动时间为 200 h 与 400 h 的电池样本压力变化增至 0.3 MPa 以上，由于机械振动应力对电池的影响机制具有复杂性，无法直接将振动时间与热失控环境压力进行直接关联，但通过实验样本对比判断，振动一定程度上增加了燃爆压力变化的不稳定性。

3 结 论

本文探究了振动工况对锂电池的影响机理，从结构损伤、电气性能、热失控特性 3 方面深展开研究分析，得出了以下结论：

1）长期振动会造成锂电池内部集流片形变，从而产生对电极卷的挤压。

2）振动工况下锂电池循环老化导致电极卷形成的横向多层褶皱更为明显，纵向影响范围更大。

3）振动工况会对电池阻抗产生影响，振动前期，欧姆阻抗变化较小，SEI 膜阻抗和电荷转移阻抗呈增大趋势，机械应力积累达到临界值时，欧姆阻抗增加，此时 SEI 膜阻抗和电荷阻抗减小。

4）循环老化与振动耦合工况造成了电池电气性能的动态改变，加剧了电池电气性能输出的波动性。

5）振动会造成热失控触发时间提前，初爆、燃爆间隔缩短，同时增加了燃爆对环境冲击压力的不稳定性。

参考文献（References）

[1] 陈农田, 李俊辉, 王志宏, 等. B787-800 飞机锂电池起火事故原因及分析[J]. 电池, 2022, 52(2): 204-207.
CHEN N T, LI J H, WANG Z H, et al. Cause and analysis of lithium battery fire accident of B787-800 aircraft[J]. Battery Bimonthly, 2022, 52(2): 204-207(in Chinese).

[2] 庄淡盛. 锂电池在多旋翼无人机应用中的特性[J]. 电子技术与软件工程, 2018(15): 78-80.
ZHUANG D S. Characteristics of lithium battery in multi-rotor UAV application[J]. Electronic Technology & Software Engineering, 2018(15): 78-80(in Chinese).

[3] 宋文龙, 罗秋月, 何艺, 等. 2021 年我国电池产销情况[J]. 电池工业, 2022, 26(2): 85-89.
SONG W L, LUO Q Y, HE Y, et al. Production and sales of batteries in China in 2021[J]. Chinese Battery Industry, 2022, 26(2): 85-89(in Chinese).

[4] 张青松, 刘添添, 郝朝龙, 等. 锂离子电池热失控气体快速检测及危险性分析方法[J]. 北京航空航天大学学报, 2023, 49(9): 2227-2233.
ZHANG Q S, LIU T T, HAO Z L, et al. Rapid detection analysis method of thermal runaway gas composition and risk of lithium ion battery[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(9): 2227-2233(in Chinese).

[5] 张青松, 曲奕润. 循环老化三元锂离子电池热失控气体毒性研究[J]. 北京航空航天大学学报, 2024, 50(6): 1761-1769.

ZHANG Q S, QU Y R. Study on the toxicity of thermal runaway gas in cyclic aging ternary lithium ion batteries[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(6): 1761-1769(in Chinese).

[6] FU Y Y, LU S, SHI L, et al. Ignition and combustion characteristics of lithium ion batteries under low atmospheric pressure[J]. Energy, 2018, 161: 38-45.

[7] LI Y W, JIANG L H, HUANG Z H, et al. Pressure effect on the thermal runaway behaviors of lithium-ion battery in confined space[J]. Fire Technology, 2023, 59(3): 1137-1155.

[8] 张青松, 李东琪, 连晓雪. 涉水锂离子电池的腐蚀与析氢风险研究[J]. 北京航空航天大学学报, 2024, 50(7): 2083-2092.
ZHANG Q S, LI D Q, LIAN X X. Study of corrosion and hydrogen evolution risk of waded lithium ion-battery[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(7): 2083-2092 (in Chinese).

[9] HOOPER J M, MARCO J, CHOUCHELAMANE G H, et al. Multi-axis vibration durability testing of lithium ion 18650 NCA cylindrical cells[J]. Journal of Energy Storage, 2018, 15: 103-123.

[10] BRAND M J, SCHUSTER S F, BACH T, et al. Effects of vibrations and shocks on lithium-ion cells[J]. Journal of Power Sources, 2015, 288: 62-69.

[11] PARASUMANNA A B K, KARLE U S, SARAF M R. Material characterization and analysis on the effect of vibration and nail penetration on lithium ion battery[J]. World Electric Vehicle Journal, 2019, 10(4): 69.

[12] ZHANG L J, NING Z S, PENG H, et al. Effects of vibration on the electrical performance of lithium-ion cells based on mathematical statistics[J]. Applied Sciences, 2017, 7(8): 802.

[13] LEE P Y, PARK S, CHO I, et al. Vibration-based degradation effect in rechargeable lithium ion batteries having different cathode materials for railway vehicle application[J]. Engineering Failure Analysis, 2021, 124: 105334.

[14] SOMERVILLE L, HOOPER J M, MARCO J, et al. Impact of vibration on the surface film of lithium-ion cells[J]. Energies, 2017, 10(6): 741.

[15] ZHANG L J, MU Z Q, GAO X Y. Coupling analysis and performance study of commercial 18650 lithium-ion batteries under conditions of temperature and vibration[J]. Energies, 2018, 11(10): 2856.

[16] United Nations. UN manual of tests and criteria part III subsection 38.3[S]. New York and Geneva: United Nations, 2009: T3.

[17] SHI L, LIU B D, YU H Y, et al. Review of CT image reconstruction open source toolkits[J]. Journal of X-Ray Science and Technology, 2020, 28(4): 619-639.

[18] WALDMANN T, GORSE S, SAMTLEBEN T, et al. A mechanical aging mechanism in lithium-ion batteries[J]. Journal of the Electrochemical Society, 2014, 161(10): A1742.

[19] PASTOR-FERNÁNDEZ C, WIDANAGE W, MARCO J, et al. Identification and quantification of ageing mechanisms in lithium-ion batteries using the EIS technique[C]//Proceedings of the IEEE Transportation Electrification Conference and Expo. Piscataway: IEEE Press, 2016: 1-6.

[20] ABRAHAM D P, POPPEN S D, JANSEN A N, et al. Application of a lithium-tin reference electrode to determine electrode contributions to impedance rise in high-power lithium-ion cells[J]. Electrochimica Acta, 2004, 49(26): 4763-4775.

[21] HAN X B, LU L G, ZHENG Y J, et al. A review on the key issues of the lithium ion battery degradation among the whole life cycle[J]. eTransportation, 2019, 1: 100005.

[22] 贺元骅, 郭君, 王海斌, 等. 振动环境对锂离子电池热失控危险性的影响[J]. 科学技术与工程, 2020, 20(29): 12242-12246.
HE Y H, GUO J, WANG H B, et al. Impact of vibration on thermal runaway risk of lithium-ion battery[J]. Science Technology and Engineering, 2020, 20(29): 12242-12246(in Chinese).

[23] HUANG P F, LIU S T, MA J, et al. Comprehensive investigation on the durability and safety performances of lithium-ion batteries under slight mechanical deformation[J]. Journal of Energy Storage, 2023, 66: 107450.

[24] 李文华, 邵方旭, 暴二平, 等. 六自由度振动老化条件下锂离子电池的衰退机理诊断与 SOH 预测[J]. 仪器仪表学报, 2021, 42(8): 62-69.
LI W H, SHAO F X, BAO E P, et al. Diagnosis of degradation mechanism and SOH prediction of lithium-ion batteries under 6-DOF vibration and aging conditions[J]. Chinese Journal of Scientific Instrument, 2021, 42(8): 62-69(in Chinese).

[25] DUBARRY M, TRUCHOT C, LIAW B Y. Synthesize battery degradation modes via a diagnostic and prognostic model[J]. Journal of Power Sources, 2012, 219: 204-216.

[26] 任东生, 冯旭宁, 韩雪冰, 等. 锂离子电池全生命周期安全性演变研究进展[J]. 储能科学与技术, 2018, 7(6): 957-966.
REN D S, FENG X N, HAN X B, et al. Recent progress on evolution of safety performance of lithium-ion battery during aging process[J]. Energy Storage Science and Technology, 2018, 7(6): 957-966 (in Chinese).

[27] 杨娟, 牛江昊, 张青松. 循环老化对锂离子电池热失控气体爆炸危险性影响实验研究[J/OL]. 航空学报, (2023-01-10)[2023-04-04]. https://kns.cnki.net/kcms/detail/11.1929.v.20230331.1412.014.html.
YANG J, NIU J H, ZHANG Q S. Experimental study on the influence of cyclic aging on the risk of gas explosion caused by thermal runaway in lithium-ion batteries[J/OL]. Acta Aeronautica et Astronautica Sinica, (2023-01-10)[2023-04-04]. https://kns.cnki.net/kcms/detail/11.1929.v.20230331.1412.014.html(in Chinese).

[28] KONG D P, LV H P, PING P, et al. A review of early warning methods of thermal runaway of lithium ion batteries[J]. Journal of Energy Storage, 2023, 64: 107073.

Influence of vibration on cyclic and thermal runaway characteristics of lithium ion batteries

ZHANG Qingsong[1, *], LI Dongqi[1], YANG Juan[1, 2]

(1. Key Laboratory of Civil Aviation Thermal Hazards Prevention and Emergency Response, Civil Aviation University of China, Tianjin 300300, China;
2. Engineering Technology Training Center, Civil Aviation University of China, Tianjin 300300, China)

Abstract: By analyzing the impact of mechanical vibration on the safety of lithium ion batteries in various application scenarios, such as transportation and aerospace, the 18650 battery was used as a sample, and the.influence of vibration conditions on the internal structure, cyclic aging mechanism, electrical performance, and thermal runaway characteristics of lithium ion batteries was investigated. The results indicate that vibration can significantly induce alterations in the internal microstructure of lithium ion batteries. Prolonged exposure to vibration stress can lead to deformation of the negative electrode collector, resulting in compression on the electrode roll. Additionally, momentary vibration stress can dynamically affect the contact area between the electrode material and the diaphragm, impeding the migration of lithium ions during the embedding and de-embedding processes. Consequently, this exacerbates the fluctuation of Coulomb efficiency and energy efficiency, leading to insufficient discharge under vibration conditions. Furthermore, thermal runaway experiments demonstrate that the time required to trigger thermal runaway shortens as the duration of vibration increases. Compared with that of non-vibrating batteries, the initial explosion time of thermal runaway of batteries subjected to 600 hours of vibration shortens to 72 seconds, and the combustion-induced explosion time shortens to 114 seconds. Meanwhile, the interval between initial and combustion-induced explosions is shortened. Moreover, the instability of the impact force produced by combustion-induced explosion increases.

Keywords: lithium ion battery; safety; vibration; cyclic performance; thermal runaway

Received: 2023-05-23; **Accepted**: 2023-07-15; **Published Online**: 2023-07-25 09:44

URL: link.cnki.net/urlid/11.2625.V.20230724.1828.002

Foundation items: The Fundamental Research Funds for the Central Universities (3122024058); Civil Aviation University of China Postgraduate Research Innovation Project (2022YJS009)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0304

基于级联的航班地面保障动态预测

唐小卫[1,*]，丁叶[1]，吴政隆[2]，张生润[1]，吴佳琦[1]，叶梦凡[1]

(1. 南京航空航天大学 民航学院，南京 211106；2. 新南威尔士大学 航空学院，悉尼 NSW 2052)

摘 要：对航班地面保障过程进行精准预测是实现航班精细化管理、提升机场协同决策（A-CDM）系统管理效能的关键。为此，提出一种基于级联多输出梯度提升回归树模型的航班地面保障多节点动态预测方法。通过搭建级联框架实现了不同保障进度之间预测信息的传递和预测结果的更新，基于可进行多节点预测的梯度提升回归树设计了地面保障过程动态预测算法，以典型繁忙机场为对象构建了航班基础属性与层级信息传递两大类特征集。结果表明：所提方法有效实现了地面保障各关键节点完成时间的动态预测，初始预测各节点±5 min预测精度均达到80%以上，随着保障过程推进模型预测性能逐步提升，超过60%的节点±5 min最终预测精度超过95%，为提升航班运行的可预测性和机场多主体协同决策能力提供有效方法支撑。

关 键 词：航空运输；航班地面保障；机场协同决策；级联；梯度提升回归树；动态预测
中图分类号：V351.11
文献标志码：A **文章编号**：1001-5965(2025)05-1557-09

当前各机场对航班地面运行全流程管理主要依托机场协同决策(airport collaborative decision making, A-CDM)系统进行。A-CDM系统的建设已打破空管、机场和航司等运行主体之间的信息壁垒，但大多数机场只能对地面保障的终末节点“撤轮挡”进行结果监控，属于事后管理的范畴；少数机场虽已实现对“撤轮挡”的主动预测，但单个节点的静态预测无法指导保障部门进行有效的事中干预，其本质原因是地面保障延误萌发时期的特征相对模糊，随着保障过程推进，上游作业的波动会在下游产生累积，继而影响后续离港流程。因此，推动地面保障过程预测研究由静态向动态、由单一节点向全过程节点转变，是实现航班精细化管理、提升A-CDM系统管理效能的关键所在。

航班地面保障作业流程管理研究领域经历了“单一流程单作业→单一流程多作业→多流程多作业”的发展过程。在针对单一流程的研究中，关于登机流程的研究最为丰富[1-3]，近来的研究则更关注疫情防控对地面保障清洁和配餐作业的影响[4]。随着协同决策思想在产业界的不断落地升级，学术界更重视多流程多作业协同管理的相关问题[5]。由于目标撤轮挡时间是A-CDM机制下各方衔接的核心[6]，所以对撤轮挡时间的预测成为研究的重点，研究方法包括分布拟合、仿真建模、网络作业、启发式算法等，如Sheibani拟合保障作业的时间分布并通过关键路径法和蒙特卡罗模拟方法估计保障结束时间[7]，邢志伟等基于人工鱼群算法改进的向量机模型[8]与自适应多层遗传算法改进的BP神经网络模型[9]估计了保障服务时间。上述研究直接对目标撤轮挡时间进行单一节点预测，不利于在延误发生前即对地面保障进行干预。而王立文等提出一种结合高斯核概率密度估计的过站航班地面保障过程贝叶斯网络动态预测方法[10]，初步实现了精

收稿日期：2023-06-01；**录用日期**：2023-11-13；**网络出版时间**：2023-11-21 15:22
网络出版地址：link.cnki.net/urlid/11.2625.V.20231120.1526.002
基金项目：国家自然科学基金(U2333204,U2233208)；2023年度民航安全能力建设项目((2023)155号)；南京航空航天大学科研与实践创新计划资助项目(xcxjh20220715)
* **通信作者**. E-mail：tangxiaowei@nuaa.edu.cn

引用格式：唐小卫，丁叶，吴政隆，等. 基于级联的航班地面保障动态预测[J]. 北京航空航天大学学报，2025，51(5)：1557-1565.
TANG X W，DING Y，WU Z L，et al. Dynamic prediction of flight ground service based on cascade[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1557-1565 (in Chinese).

细化的动态预测，但其预测需要进行大规模概率分布拟合计算，运算复杂度较高，较难在实际保障中应用。在特征构建方面，现有保障时间预测的特征库一般包括机型、机位、运行时段、航空公司、航线性质等因素，较少考虑由机场航班时刻表决定的内在因素。

为此，本文面向航班地面保障过程建立了可在层级间传递信息的级联框架，以此为基础采用多输出梯度提升回归树（gradient boosting regression tree, GBRT）作为不同保障进度下多节点预测的基本模型，并构建了适用于级联模型的特征集，最终结合机场实际保障数据验证了模型的有效性。

1 级联多输出梯度提升回归树模型构建

航班地面保障过程如图1所示，包含开关舱门、上下客货、配餐、清洁、加油等众多具有串并行关系、严格作业次序要求的作业，这些作业的开始与结束时间即为保障中需要关注的节点。航班地面保障多节点动态预测问题是指基于保障进程的实时推进，面向未完成的保障节点采用预测模型逐步更新其完成时间估计值，实现预测精度随任务推进持续提升的目的。本文将构建级联多输出梯度提升回归树（cascaded multiple output-gradient boosting regression tree, CMO-GBRT）模型对地面保障多节点进行动态预测。

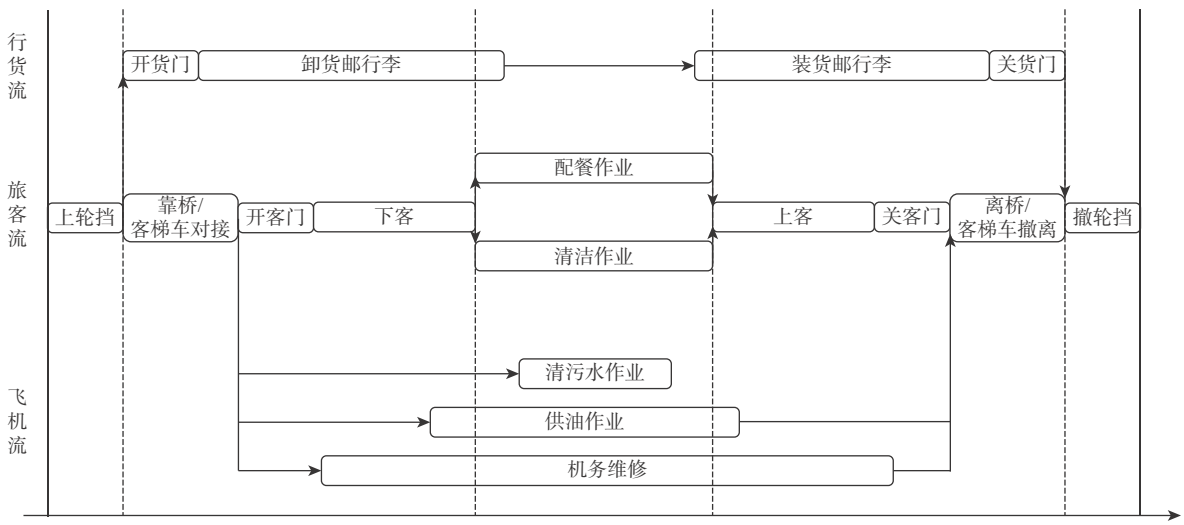


图1 航班地面保障作业流程

Fig. 1 Operation process of flight ground service

1.1 级联框架描述

级联框架是一个层级结构，每层级都需构建单独的学习器，通过学习器的组合来分解复杂问题[11-12]，组合方式可以是串联、并联和串并联结合。由于地面保障过程具有固定的作业次序，因此，本文采取串联组合方式，将多个多节点预测模型按照保障进程顺序进行组合。

级联框架的优点还在于能通过层级之间的信息传递提高系统整体性能。其原理是，构建每层结构时通过训练调优得到预测性能最好的学习器，将原特征、学习结果、学习结果与真实值的偏差作为下一层级的输入特征，下一层级学习器通过额外学习上层学习器的偏差得到改进，进一步提升预测性能。本文新引入了表征上一层级预测值与待修正值的层级信息传递特征，该特征将随着保障过程推进发生动态变化，不断传递修正信息，达到提升总体学习性能的目的。

针对地面保障多节点动态预测问题，若从上轮挡至撤轮挡的某条关键路线上共有$n+1$个具备先后次序的关键节点。当保障进度依次推进到除终末节点撤轮挡以外的n个关键节点时，均需对当前进度下所有未完成节点的完成时间进行更新预测，因此，共需进行n次静态的多节点预测，其预测流程如图2右边框所示，对应n次预测构建n个层级的级联框架，其框架结构如图2左边框所示。

假设保障进度推进到第i个节点，需进行第i次更新预测，则i的取值范围为$1 \leqslant i \leqslant n$，按次序第1至第$i$个节点为已完成节点（图2右边深色节点），第$i+1$至第$n+1$个节点为未完成节点（图2右边白色节点），亦为需更新预测的对象（图2右边虚线框）。令n次预测中节点真实值的集合记为A，节点预测值的集合记为$\bar{A}$，表达式分别为

$$A = \{V_1, \{V_1, V_2\}, \cdots, \{V_1, V_2, \cdots, V_j, \cdots, V_i\}, \cdots, \{V_1, V_2, \cdots, V_n\}\} \quad 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant i \tag{1}$$

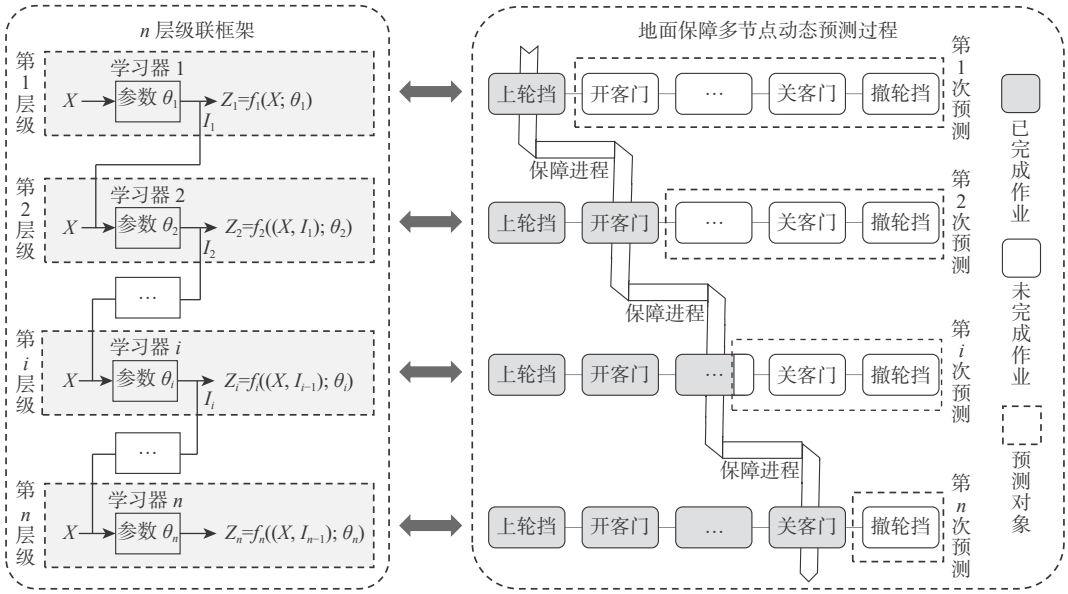


图 2　级联框架与地面保障多节点动态预测过程的映射关系

Fig. 2　Mapping relationship between cascaded framework and multi-node dynamic prediction of flight ground service

$$\bar{A}=\left\{\left\{U_2^1,U_3^1,\cdots,U_{n+1}^1\right\},\left\{U_3^2,U_4^2,\cdots,U_{n+1}^2\right\},\cdots,\left\{U_{i+1}^i,U_{i+2}^i,\cdots,U_j^i,\cdots,U_{n+1}^i\right\},\cdots,U_{n+1}^n\right\}$$
$$1\leqslant i\leqslant n, i+1\leqslant j\leqslant n+1 \qquad (2)$$

式中: i表示预测次数; j表示节点在所有关键节点中的次序; V_j表示第j个节点的真实值, 一旦获知则固定不变; U_j^i表示第i次预测时第j个节点的预测值, 同一节点的预测值随保障进程推进不断更新; A与$\bar{A}$互为补集。

因此, 根据图 2 中预测次数等于级联框架层数的映射关系, n层级联框架定义如下:

定义　n层级联框架 n-cascaded。进行第i次预测的学习器位于级联框架的第i层级, 其目标标签集合为$\left\{U_{i+1}^i,U_{i+2}^i,\cdots,U_j^i,\cdots,U_{n+1}^i\right\}$, 用$Z_i$表示, 第$i$层级的学习器记为$f_i$, 学习器的超参数记为$\theta_i$, 原始输入特征记为$X$, 由第$i-1$层级的学习结果计算得到的层级信息传递特征, 记为$I_{i-1}$, 则第$i$层级的表达式为$Z_i=f_i((X,I_{i-1});\theta_i)$。

1.2　级联多输出梯度提升回归树模型

级联框架搭建完成后需要进一步构建每层级的学习器, 本文选择多输出梯度提升回归树模型进行每层的多节点预测。

输入多个表征预测对象特点的特征值后可以输出多个预测对象的预测值, 即为多输入多输出[13], 借助 python 中 MultiOutputRegressor 函数, 整合封装多个独立单输出模型, 可以构成一个多输出模型, 仅需训练一次即可输出所有对象预测值, 极大减少了模型训练的复杂度与时间成本, 使得预测更加高效; 而梯度提升回归树是以决策树为基本学习器的基于 boosting 框架的集成学习算法[14], 可以灵活地处理连续或离散类型的数据且充分利用数据的全部特征信息[15], 并且对异常值有较好的鲁棒性[14], 在相对较少的运算时间下能获得较高的预测精度[16]。因此, 构建多输出梯度提升回归树进行单层多节点预测, 结合级联框架形成 CMO-GBRT 模型, 并设计航班地面保障多节点动态预测算法如图 3 所示, 具体步骤如下:

步骤 1　获取机场地面保障实际运行数据, 分

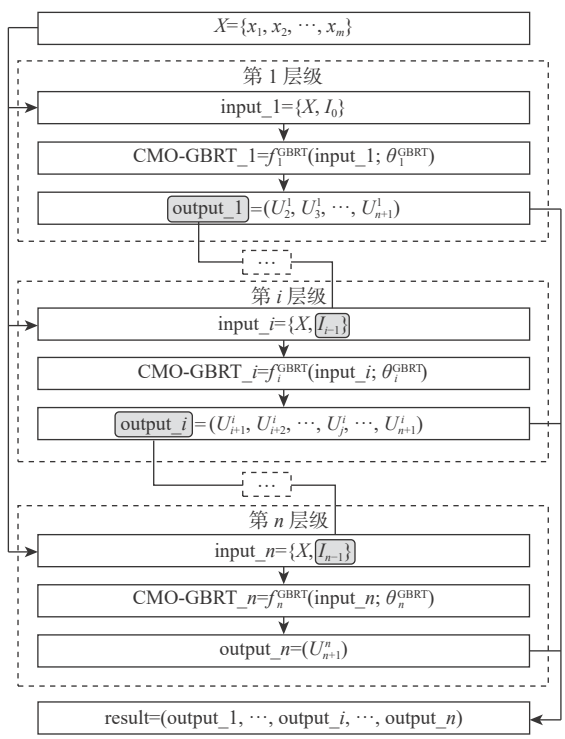


图 3　级联多输出梯度提升回归树算法框架

Fig. 3　Algorithm framework for cascaded multi-output gradient boosting regression tree

析并构建具有m个原始输入特征的原始特征集$X=\{x_1,x_2,\cdots,x_m\}$，判断地面保障关键节点，构建n个层级的CMO-GBRT模型，其中，第i层的子模型记为CMO-GBRT_i。

步骤2 当保障进度推进到起始节点上轮挡时进行第1层级预测。此时无前序层级，取紧前节点完成时间为初始层级传递特征，记为I_0，与X形成第1层级的特征集input_1。

步骤3 通过多次训练调参得到预测性能最优的CMO-GBRT_1模型，输入特征集input_1，得到当前保障进度下所有未完成节点的完成时间预测值output_1。

步骤4 当保障进度推进到第2个关键节点时进行第2层级预测。结合当前进度节点的真实值与上一层级的预测值计算得到层级传递信息特征I_1，与X形成input_2。

步骤5 重复步骤3和步骤4直至完成所有层级预测，其中，CMO-GBRT_i模型输入为input_i，输出为output_i，输入下一层级的层级传递信息特征为I_i。最终获得所有动态预测结果result。

2 数据准备及特征构建

2.1 实验数据及预处理

本文实验数据来源于浦东机场2022年的部分过站客运航班地面保障运行数据，数据字段包括航班基本信息与主要保障节点时刻数据，如表1所示。机位一栏出现2个机位号表示该航班进离港使用了不同机位，存在拖曳的情况。经数据清洗后得到有效数据样本50 770条。

表1 航班地面保障运行数据

Table 1 Data of flight ground service operations

字段	样例	字段	样例
日期	2021-3-1	上轮挡	14:32
航班号	9C6136|9C8569	开客门	14:36
航班属性	正班	下客开始/结束	14:38|14:43
STA	14:35	保洁开始/结束	14:51|15:02
STD	15:50	供油开始/结束	15:24|15:37
机型	A320	配餐开始/结束	15:28|15:29
机位	55|61	登机口开启	15:34
航线性质	国内|国际	登机口关闭	15:50
地面代理	CQH	关客门	15:57
旅客总数	175|165	撤轮挡	16:07

注：STA为计划到港时间，STD为计划离港时间，CQH为国际民用航空组织规定的春秋航空三字代码。

因为时间格式数据不适用于回归预测模型，所以需要进行数据转化，定义某个节点完成时间为地面保障开始至该节点完成的时长：

$$T_{\text{task}}=t_{\text{task}}^{\text{end}}-t_0 \tag{3}$$

式中：T_{task}为节点完成时间，task为节点名；$t_{\text{task}}^{\text{end}}$为该节点实际完成时间；$t_0$为保障实际开始时间，本文以上轮挡时间代替。转化后表1样例节点完成时间如表2所示。

表2 航班地面保障节点完成时间转化结果

Table 2 Conversion result of node completion time in flight ground service

节点	转化后节点完成时间/min	节点	转化后节点完成时间/min
上轮挡	0	配餐结束	57
开客门	4	登机口开启	62
下客结束	11	登机口关闭	78
保洁结束	30	关客门	85
供油结束	65	撤轮挡	95

2.2 特征集构建

2.2.1 基础属性特征

本文构建基础属性特征，对应定义中的原始输入特征X。关于地面保障时间预测的已有研究主要考虑机型、机位、航线性质、地面代理、运行时段等外部特征[10]，而由机场航班时刻表决定的内部特征较少得到关注，包括航班的计划过站时间、最小过站时间及航班密度等，本文在特征集的构建中拟引入这些因素，具体如下。

1）过站时间因素

计划过站时间是航班计划离港时间与计划到港时间的差值，机场保障部门需根据计划过站时间合理调配人力与设备资源。如图4(a)所示，由于每个航班入位后均需立即进行开客门与下客作业，所以计划过站时间长短对开客门与下客结束节点完成时间无显著影响，而其余节点均与计划过站时间密切相关。计划过站时间计算如下：

$$T_{\text{S}}=t_{\text{STD}}-t_{\text{STA}} \tag{4}$$

式中：T_{S}为计划过站时间；t_{STD}为计划离港时间；t_{STA}为计划到港时间。

最小过站时间是指地面服务人员为保证飞机完成后续飞行任务提供所需服务的必要服务时间。浦东机场常见机型的座位数可分为250座以下和250座以上2类，民航局规定的最小过站时间标准分别为65 min和75 min。2022年浦东机场无A380机型执飞的客运航班，其标准为120 min，在本文暂不考虑。如图4(b)所示，本文最小过站时间设

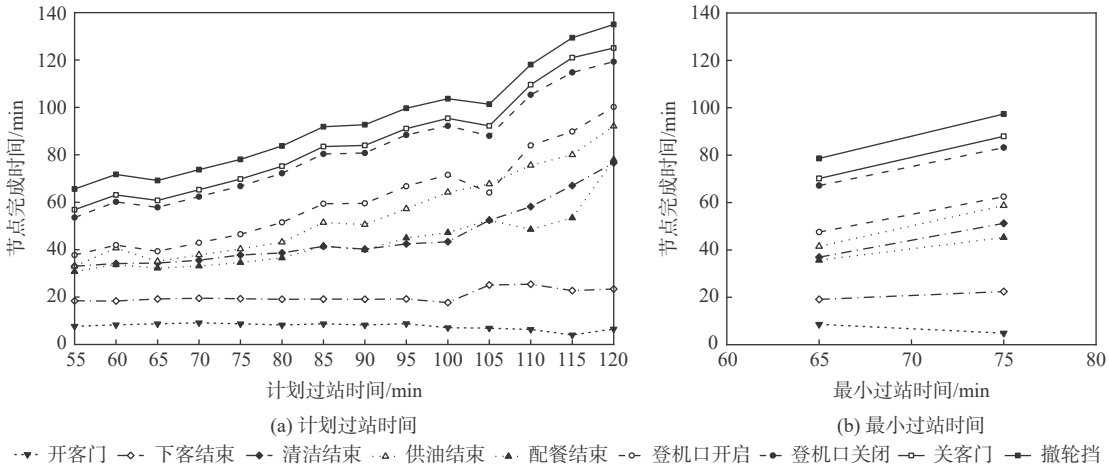


图 4　过站时间对地面保障节点完成时间的影响

Fig. 4　Effect of turn-round time on node completion time in flight ground service

置为分类变量, 基于机型确定 65 min 和 75 min 两类最小过站时间, 分别用 0 和 1 表示。不同的最小过站时间标准也显著影响了节点完成时间(开客门和下客结束除外)。

2) 航班密度因素

航班密度表示在过站航班邻近时间窗内需要竞争地面保障资源的航班数量, 其数值直观反映了机场地面保障的繁忙情况。如图 5 所示, 航班密度越大, 则保障任务越密集, 保障资源使用越紧凑, 促使提高保障作业效率, 进而对除开客门和下客结束外的各节点完成时间产生影响。本文定义航班密度为在当前保障航班实际落地的前后相邻时间窗内落地的航班数量:

$$N_{\mathrm{d}} = (N_{\mathrm{b}}t + N_{\mathrm{a}}t)/2t \tag{5}$$

式中: N_{d}为到达航班密度; N_{b}、N_{a}分别为当前保障航班实际落地前、后相邻时间窗内落地的航班数量; t为时间窗, 本文取 15 min。

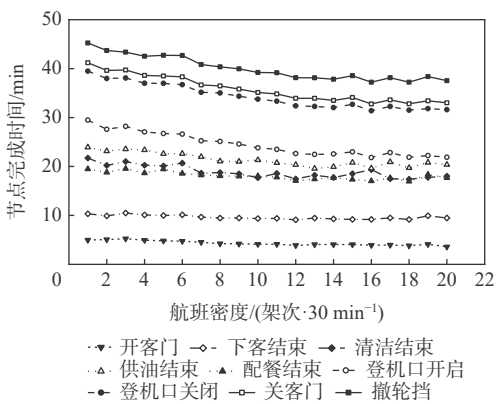


图 5　航班密度对地面保障节点完成时间的影响

Fig. 5　Effect of flight density on node completion time in flight ground service

2.2.2　层级信息传递特征

本文具体构建前序保障预测时间与前序保障预测偏差这 2 个层级信息传递特征, 如图 6 所示, 通过 CMO-GBRT 进行预测时, 保障进度依次推进到每个节点时, 都将对未完成节点的完成时间进行新一层级的预测值更新, 当前进度中刚完成的节点既拥有一个上一层级的预测时间, 还拥有一个新获得的实际时间, 将上一层级的预测值作为前序保障预测时间特征, 将预测值与实际值的偏差作为前序保障预测偏差特征:

$$L_{\mathrm{pre}} = U_i^{i-1} \tag{6}$$

$$L_{\mathrm{bias}} = V_i - U_i^{i-1} \tag{7}$$

式中: L_{pre}为前序保障预测时间特征; U_i^{i-1}为第 $i-1$ 层级中第 i 个节点完成时间的预测值; L_{bias}为前序保障预测偏差特征; V_i为第 i 层级开始时获得的第 i 个节点完成时间的实际值。

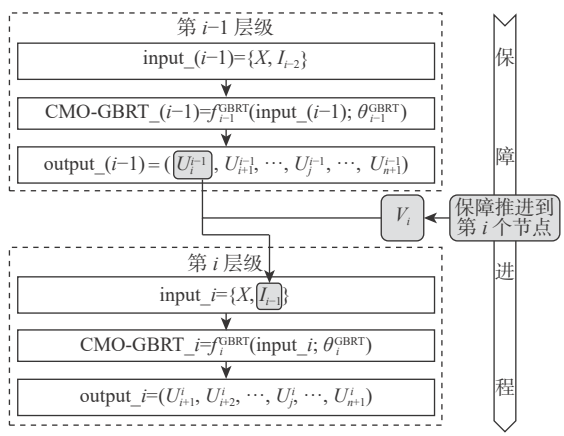


图 6　层级信息传递特征示意图

Fig. 6　Level information transmission feature

2.2.3　最终特征集

结合浦东机场实际运行数据, 考虑基础属性特征与层级信息传递特征的最终特征集。其中, 基础属性特征: 机型 C, 是否拖曳 X_{pull}, 进港机位 S_{arr}, 离

港机位S_{dep}，进港机位类型N_{arr}，离港机位类型N_{dep}，进港航站楼T_{arr}，离港航站楼T_{dep}，进港旅客数P_{arr}，离港旅客数P_{dep}，地面代理G，进港保障时段F_{arr}，离港保障时段F_{dep}，进港航线性质Q_{arr}，离港航线性质Q_{dep}，最小过站时间T_{MTTT}，计划过站时间T_S，到达航班密度N_d。层级信息传递特征：前序保障预测时间L_{pre}，前序保障预测偏差L_{bias}。

3 实验验证与结果分析

进行地面保障多节点动态预测前，需要预先判断保障关键节点以设计预测层级。结合浦东机场实际保障数据计算保障关键路线，得到的关键路线主要由旅客流程及部分其他流程的节点组成，现有研究也表明地面保障能否按时完成与旅客保障流程密切相关[4]。但此时关键路线上的节点数量仍较多，不利于对保障过程进行有效监控，因此，再利用多元线性回归模型进一步分析不同节点的延迟完成时间对最终保障完成时间的影响程度，初步确定上轮挡、开客门、下客结束、供油结束、登机口开启、登机口关闭、关客门与撤轮挡这8个影响最大的节点为初始关键节点。因此，初步构建具有7个层级的CMO-GBRT模型。

由2.2.3节可知每层级特征集各需构建20个特征，包括18个固定的基础属性特征和2个动态变化的层级信息传递特征，将处理好的数据集以7∶3的比例划分为训练集与测试集，设置平均绝对误差（mean absolute error, MAE）、R平方值（R-squared, R^2）和预测精度等预测性能为模型评价指标，其中，预测精度指标为预测结果与实际时间差值在±3 min和±5 min内的样本数量占比。

在保障进行到上轮挡节点时进行初次预测，对应第1层级模型，用紧前节点完成时间进港滑入时长作为初始层级信息传递特征，初始预测结果如表3所示。第1层级预测对象为开客门、下客结束、供油结束、登机口开启、登机口关闭、关客门和撤轮挡共7个节点，除供油结束节点外，其余节点预测结果的MAE均低于3.2 min，表明这些节点的初始预测结果与真实时间的误差较小，符合机场协同决策系统技术规范[17]对关键节点预测准确性的要求；同时R^2较高，包括供油结束节点在内均大于0.7，表明模型对样本数据的拟合度较好，验证了本文构建的特征集的有效性；在预测精度方面，除供油结束节点外，其余节点±5 min的预测精度均达到80%以上，±3 min的预测精度均达到50%以上，其中，开客门与下客结束节点的±3 min预测精度更达到了85.20%与67.32%，表明在第1层级获得的初始预测时间与真实时间的5 min偏离值较小，初始预测能较好地确定节点实际完成时间的范围，在上轮挡开始时即能对资源投入与保障进度安排提供指导。

表3 第1层级关键节点完成时间预测结果

Table 3 Prediction results of key node completion time at Level 1

关键节点	MAE/min	R^2	±3 min预测精度/%	±5 min预测精度/%
开客门	1.60	0.85	85.20	95.39
下客结束	2.46	0.80	67.32	87.98
供油结束	7.66	0.72	26.40	41.61
登机口开启	2.93	0.96	56.36	82.47
登机口关闭	2.96	0.97	54.79	82.37
关客门	2.94	0.97	55.44	83.23
撤轮挡	3.11	0.97	52.97	80.04

注：预测精度是指模型输出的时间与实际滑入时间的差值在某一设定范围内的数量与总预测样本数之比。

在第1层级初始预测的7个节点中，供油结束节点的拟合度较好但预测误差大、预测精度低，表明本文模型在全面学习供油结束节点数据信息的情况下仍不能对其进行准确预测。与其余关键节点不同，实际保障中供油结束后并不一定紧邻后序登机口开启节点，因此，其非延误情况下完成时间范围大、不确定性高，仅在延误时对保障过程产生影响，且可以通过登机口开启节点是否延误进行侧面印证，其预测值对实际保障指导性不高。因此，不再将供油结束节点作为关键节点，最终建立6层级CMO-GBRT模型，层级序号与进行该层级预测时的保障进程对应关系如表4所示。

表4 层级序号与进行该层级时的对应保障进程

Table 4 Level number and service process at corresponding levels

层级序号	对应保障进程	层级序号	对应保障进程
1	上轮挡	4	登机口开启
2	开客门	5	登机口关闭
3	下客结束	6	关客门

为验证GBRT模型的优越性，选择目前常用于地面保障预测的深度学习模型中的卷积神经网络（convolutional neural network, CNN）、经典机器学习算法中的决策树（decision tree, DT）、支持向量机回归（support vector regression, SVR）及同为集成学习模型的随机森林（random forest, RF）进行对比分析，以第1层级撤轮挡节点的预测结果为比较对象，对比如表5所示。预测误差方面，树模型类的DT、RF与GBRT相对较低，其中，GBRT最优；模型拟合方面GBRT的表现也较好；预测精度方面，SVR在±3 min预测精度最高，GBRT处于中间水平，但

表 5　不同模型在第 1 层级时的撤轮挡时间预测结果对比

Table 5　Comparison of off-block time prediction results of different models at Level 1

预测模型	MAE/min	R^2	±3 min预测精度/%	±5 min预测精度/%
CNN	5.77	0.87	34.83	53.64
DT	3.62	0.85	59.58	73.00
SVR	5.31	0.53	63.60	70.23
RF	3.83	0.77	55.03	78.42
GBRT	3.11	0.97	52.97	80.04

是在±5 min 预测精度上, GBRT 具有最高的精度, 远高于 SVR 模型。综上所述, 本文模型误差更小、拟合度更高、预测精度更好, 因此, 本文选择 GBRT 模型作为基准模型是有效的。

随着保障进程推进, 最终所有节点的动态预测结果如表 6 所示, 可知超过 60% 的节点±5 min 预测精度超过 95%, 且各个节点的预测误差均逐渐减小, R^2 均逐渐增大, 预测精度也均获得不同程度的提升。以 MAE 和±5 min 预测精度 2 个评价指标为例, 其变化趋势如图 7 所示, 除登机口开启时间的预测性能没有显著变化外, 随着预测层级的增加, 下客结束、登机口关闭、关客门与撤轮挡节点的 MAE 呈现明显下降趋势, ±5 min 预测精度呈现明显上升趋势。由此可知, 本文模型在上轮挡开始后能对后续各关键节点的完成时间进行动态、定量预测, 而且模型预测性能随着保障过程推进不断提

表 6　所有节点在不同层级的动态预测结果

Table 6　Dynamic prediction results for all nodes at different levels

保障进程	层级	MAE/min	R^2	±3 min预测精度/%	±5 min预测精度/%
开客门	1	1.60	0.85	85.20	95.39
下客结束	1	2.46	0.80	67.32	87.98
	2	1.92	0.87	79.65	95.27
登机口开启	1	2.93	0.96	56.36	82.47
	2	2.93	0.96	57.41	83.06
	3	2.92	0.96	58.37	83.14
登机口关闭	1	2.96	0.97	54.79	82.37
	2	3.02	0.96	55.91	81.11
	3	2.99	0.97	55.69	82.19
	4	2.56	0.97	64.05	88.61
关客门	1	2.94	0.97	55.44	83.23
	2	3.01	0.97	55.89	81.91
	3	2.97	0.97	56.88	82.21
	4	2.5	0.98	66.06	88.93
	5	0.96	0.98	97.04	99.00
撤轮挡	1	3.11	0.97	52.97	80.04
	2	3.11	0.97	54.59	79.96
	3	3.11	0.97	54.02	79.89
	4	2.62	0.98	64.46	86.60
	5	2.11	0.98	75.22	94.54
	6	2.01	0.98	77.34	95.29

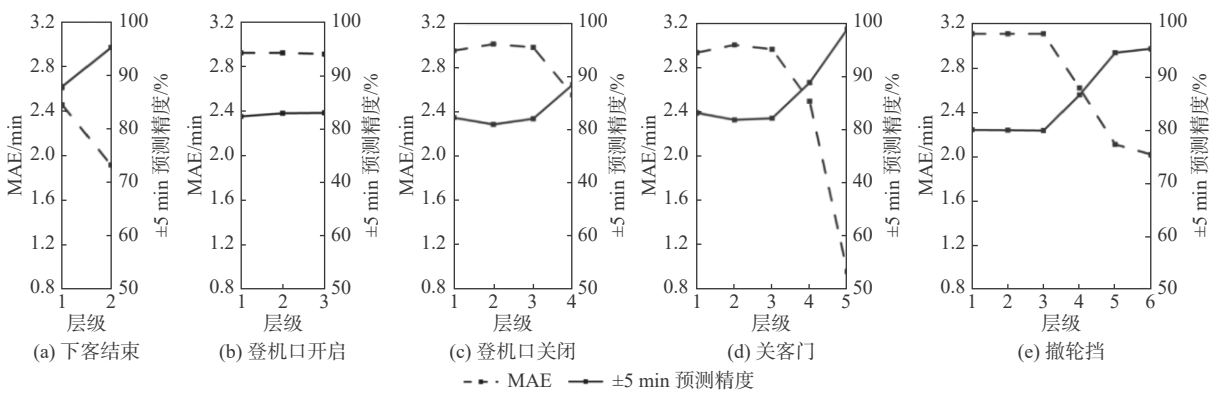


图 7　MAE 与±5 min 预测精度的变化

Fig. 7　Variation of MAE and ±5 min prediction accuracy

升，可以不断指导各保障单位加大或减少资源投入、加快或放慢保障进度等，提高资源的调配效率。值得关注的是，登机口关闭、关客门与撤轮挡节点的预测性能在第4层级得到显著提升，因此，登机口开启是协同决策机制介入地面保障进程、避免航班延误的最佳切入点。

综上所述，本文构建的CMO-GBRT模型在模型误差、拟合度和预测精度上都获得了较好的预测结果，并可与A-CDM系统的相关模块结合以提供航班地面保障关键节点预计完成时间与整体进度，实现A-CDM系统对地面保障进程的实时监控功能，为保障异常情况处置、保障过程干预和保障作业协调提供有力的指导。

4 结 论

本文提出通过级联框架构建多层级模型，实现不同保障进度间的预测信息传递，并结合多输出梯度提升回归树进行保障过程动态预测，研究表明：

1）包含前序保障预测时间与预测偏差的层级信息传递特征是级联结构中连接各层级的核心要素与影响预测结果的关键因素，计划过站时间、最小过站时间和航班密度等由航班时刻表决定的内部特征同样限制了保障进程，因此，引入这些特征构造模型特征集，初始预测±5 min预测精度均可达到80%以上。

2）随着保障过程推进，原层级模型预测结果的指导性降低，层级信息传递特征可为新一层级模型提供修正信息，达成修正预测偏差、提高预测性能的目的，使得60%以上节点的最终预测±5 min精度超过95%，为机场和航司实施干预保障进度提供准确参考依据。

航班地面保障过程是多主体协同、多作业串并行的复杂过程。本文方法仅为串联关系的保障关键节点提供了动态预测，层级信息传递方式未考虑实际保障中可能存在的并行节点情况，尚有不足，未来将在相关数据可得的情况下进一步开展覆盖保障全流程节点的动态预测研究。

参考文献（References）

[1] MILLER T, LU B, STERLING L, et al. Requirements elicitation and specification using the agent paradigm: the case study of an aircraft turnaround simulator[J]. IEEE Transactions on Software Engineering, 2014, 40(10): 1007-1024.

[2] GARCÍA ANSOLA P, GARCÍA HIGUERA A, OTAMENDI F J, et al. Agent-based distributed control for improving complex resource scheduling: application to airport ground handling operations[J]. IEEE Systems Journal, 2014, 8(4): 1145-1157.

[3] 冯霞, 张鑫, 陈锋. 飞机过站上客过程持续时间分布[J]. 交通运输工程学报, 2017, 17(2): 98-105.
FENG X, ZHANG X, CHEN F. Boarding duration distribution of aircraft turnaround[J]. Journal of Traffic and Transportation Engineering, 2017, 17(2): 98-105(in Chinese).

[4] SCHULTZ M, EVLER J, ASADI E, et al. Future aircraft turnaround operations considering post-pandemic requirements[J]. Journal of Air Transport Management, 2020, 89: 101886.

[5] POHLING O, SCHIER-MORGENTHAL S, LORENZ S. Looking into the crystal ball: how automated fast-time simulation can support probabilistic airport management decisions[J]. Aerospace, 2022, 9(7): 389.

[6] EVLER J, ASADI E, PREIS H, et al. Airline ground operations: schedule recovery optimization approach with constrained resources[J]. Transportation Research Part C: Emerging Technologies, 2021, 128: 103129.

[7] SHEIBANI K. Scheduling aircraft ground handling operations under uncertainty using critical path analysis and Monte Carlo simulation[J]. International Journal of Business Strategy and Automation, 2020, 1(1): 37-45.

[8] 邢志伟, 吴兵, 罗晓, 等. 基于AFSA-RVM的航班保障服务时间状态估计[J]. 计算机仿真, 2020, 37(8): 35-39.
XING Z W, WU B, LUO X, et al. State estimation about service time of flight support based on AFSA-RVM[J]. Computer Simulation, 2020, 37(8): 35-39(in Chinese).

[9] 邢志伟, 韩大浩, 罗谦. 基于改进GA的神经网络航班保障时间估计[J]. 计算机工程与设计, 2020, 41(1): 107-114.
XING Z W, HAN D H, LUO Q. Estimation of flight support time based on improved GA neural network[J]. Computer Engineering and Design, 2020, 41(1): 107-114(in Chinese).

[10] 王立文, 李彪, 邢志伟, 等. 过站航班地面保障过程动态预测[J]. 北京航空航天大学学报, 2021, 47(6): 1095-1104.
WANG L W, LI B, XING Z W, et al. Dynamic prediction of ground support process for transit flight[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(6): 1095-1104(in Chinese).

[11] MA Y L, JAMES HOUSDEN R, FAZILI A, et al. Real-time registration of 3D echo to X-ray fluoroscopy based on cascading classifiers and image registration[J]. Physics in Medicine and Biology, 2021, 66(5): 055019.

[12] TIAN M Z, LIU L, LU J Y, et al. Vehicle recognition based on haar features and adaboost cascade classifier[J]. Journal of Physics: Conference Series, 2022, 2303(1): 012052.

[13] 王宁, 曹萃文. 基于XGBoost模型的炼油厂氢气网络动态多输出预测模型[J]. 华东理工大学学报(自然科学版), 2020, 46(1): 77-83.
WANG N, CAO C W. A dynamic multi-output prediction model of the hydrogen network in a real-world refinery based on XGBoost model[J]. Journal of East China University of Science and Technology, 2020, 46(1): 77-83(in Chinese).

[14] ALHAKEEM Z M, JEBUR Y M, IMRAN H, et al. Prediction of ecofriendly concrete compressive strength using gradient boosting regression tree combined with gridsearchCV hyperparameter-optimization techniques[J]. Materials, 2022, 15(21): 7432.

[15] HUANG Y F, LIU Y H, LI C H, et al. GBRTVis: online analysis of gradient boosting regression tree[J]. Journal of Visualization, 2019, 22(1): 125-140.

[16] WANG T, HU S H, JIANG Y. Predicting shared-car use and examining nonlinear effects using gradient boosting regression trees[J]. International Journal of Sustainable Transportation, 2021, 15(12): 893-907.

[17] 中国民用航空局. 机场协同决策系统技术规范: MH/T 6125—2022[S]. 北京: 中国民用航空局, 2022.
Civil Aviation Administration of China. Technical specifications for airport collaborative decision-making system: MH/T 6125—2022 [S]. Beijing: Civil Aviation Administration of China, 2022(in Chinese).

Dynamic prediction of flight ground service based on cascade

TANG Xiaowei[1,*], DING Ye[1], WU Zhenglong[2], ZHANG Shengrun[1], WU Jiaqi[1], YE Mengfan[1]

(1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. School of Aviation, University of New South Wales, Sydney NSW 2052, Australia)

Abstract: Accurate prediction of flight ground service is the key to achieving fine flight management and improving management efficiency of the airport collaborative decision making (A-CDM) system. Therefore, a multi-node dynamic prediction method for flight ground service based on a cascaded multi-output gradient boosting regression tree model was proposed. The cascaded framework was built to realize the prediction information transmission and result updates between different service schedules. The dynamic prediction algorithm of flight ground service was designed based on gradient boosting regression tree which could be used for multi-node prediction. By taking a typical busy airport as an object, a feature set was constructed, covering flight basic attributes and level information transmission. The results show that the proposed method can effectively realize the dynamic prediction of key node completion time in flight ground service. The initial prediction accuracy of each node within ±5 min reaches more than 80%, and the prediction performance gradually improves as the flight ground service continues. The final prediction accuracy of over 60% of nodes within ±5min exceeds 95%. It provides effective method support for improving the flight operation predictability and the collaborative decision making ability of multi-agents in airports.

Keywords: air transportation; flight ground service; airport collaborative decision making; cascade; gradient boosting regression tree; dynamic prediction

Received: 2023-06-01; **Accepted**: 2023-11-13; **Published Online**: 2023-11-21 15:22

URL: link.cnki.net/urlid/11.2625.V.20231120.1526.002

Foundation items: National Natural Science Foundation of China (U2333204,U2233208); 2023 Civil Aviation Safety Capacity Building Project ((2023)No.155); Postgraduate Research & Practice Innovation Program of NUAA (xcxjh20220715)

* **Corresponding author**. E-mail: tangxiaowei@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0281

基于特征融合和分块注意力的无人机跟踪算法

刘芳[1]，杨雨妍[1,*]，王鑫[2]

(1. 北京工业大学 信息与通信工程学院，北京 100124；2. 国网北京丰台供电公司，北京 100161)

摘 要：无人机（UAV）已被广泛应用于各类领域中，目标跟踪是无人机应用的关键技术之一。提出一种基于特征融合和分块注意力的无人机跟踪算法，旨在解决无人机在目标跟踪时面临的外观变化和外界因素干扰等问题。采用 Siamese 网络提取模板图像和搜索图像的方向梯度直方图（HOG）特征、颜色（CN）特征和深度卷积特征，自适应计算 3 种特征权重的大小，增强融合特征的表达能力。采用改进的特征分块注意力机制，增强模板图像特征信息中有效区域的关注度，实现更有效地目标相似度匹配。为降低计算成本，将输出特征向量转换到 YCbCr 空间后进行离散余弦变换（DCT）并保留低频分量，得到特征响应图，进行分类回归得到最终目标位置。实验表明：所提算法可以降低外观变化、外界因素干扰对跟踪性能的影响，提升目标跟踪的准确性。

关 键 词：无人机；目标跟踪；特征融合；特征分块注意力；Siamese 网络

中图分类号：V279; TP391

文献标志码：A **文章编号**：1001-5965(2025)05-1566-13

无人机以其小巧灵活的特点，已经被广泛应用于各种场景，包括但不限于监视、侦察、搜救和环境监测等。目标跟踪技术是无人机应用中的重要技术之一，可以提高无人机的自主飞行能力和监控能力[1]，实现对目标的实时监控和跟踪。然而，无人机在跟踪过程中需要不断的调整飞行角度和飞行高度，经典的目标跟踪算法在跟踪时目标会发生遮挡、形变，并面临相似物、背景等因素的干扰，导致算法的跟踪效果不佳。因此，设计一种有效、准确的目标跟踪算法，解决各种干扰因素带来的影响是当前研究的主要内容[2]。

近年来，深度学习在计算机视觉领域取得了长足的进步，多种深度学习算法已经被广泛应用于目标跟踪领域，其中包括深度空间正则化鉴别相关滤波器 (deep spatially regularized discriminative correlation filter, DeepSRDCF) 跟踪器[3]、孪生全卷机网络 (siamese fully-convolution networks, SiamFC)[4]、相关滤波网络 (correlation filter networks, CFNet)[5]、深度相关滤波网络 (deep correlation filter networks, DCFNet)[6] 等，虽然上述跟踪算法在精度上有明显提升[7]，但是在目标发生外观变化、受到外界因素干扰的情况下，仍存在严重的漂移现象，降低算法的准确率。Li 等提出孪生区域建议网络 (siamese regional proposal networks, SiamRPN)[8]，将浅层网络替换成特征信息较为丰富的深度网络，相对于 SiamFC 而言跟踪准确率有明显提升，但跟踪精度和算法鲁棒性仍有待提高。Li 等提出 SiamRPN++[9]，采用多层特征融合获得更丰富的特征信息，提升了目标跟踪准确率，减轻了外观变化对算法性能的影响，但仍难以满足复杂场景下无人机目标跟踪的实际需求。

深度网络提取的深度特征更关注语义信息，却忽视了在外观刻画方面传统特征具有的优势。而在跟踪目标不清晰的情况下，只能通过外观等信息确认目标身份，因此，不能忽视传统特征在目标特征表达中的作用[10]。此外，深度网络提取的深度特征对于局部刻画、几何不变性等因素的表达能力弱

收稿日期：2023-05-25；**录用日期**：2023-08-01；**网络出版时间**：2023-09-27 18:40
网络出版地址：link.cnki.net/urlid/11.2625.V.20230927.1330.002
*通信作者. E-mail：yangyy115@qq.com

引用格式：刘芳，杨雨妍，王鑫. 基于特征融合和分块注意力的无人机跟踪算法[J]. 北京航空航天大学学报，2025，51(5)：1566-1578.
LIU F，YANG Y Y，WANG X. UAV tracking algorithm based on feature fusion and block attention[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1566-1578 (in Chinese).

于传统特征[11], 传统特征如方向梯度直方图 (histogram of oriented gradients, HOG) 特征[12]、颜色 (color names, CN) 特征[13] 对于深度特征具有很强的互补性。因此, 传统特征和深度特征的融合对于无人机跟踪目标特征的全面表达具有极其重要的意义。

无人机经常在复杂环境中飞行, 存在相似物干扰、背景干扰等问题, 使目标跟踪算法产生漂移, 影响跟踪算法的精度。注意力机制重点关注跟踪范围内重要的目标区域, 降低对非重要区域的关注程度, 有利于网络获取目标的关键特征并捕获特征之间的空间相关性, 提升目标的特征表达能力, 抑制相似物干扰、背景干扰等问题, 提升无人机在复杂环境中的算法跟踪性能。

综上所述, 本文提出一种基于特征融合和分块注意力的无人机跟踪算法, 旨在解决无人机在目标跟踪时面临的外观变化和外界因素干扰等问题。首先, 采用 ResNet-50 构建一个 5 层的 Siamese 网络, 利用 Siamese 网络提取模板图像和搜索图像的 HOG 特征、CN 特征和深度卷积特征, 自适应计算 3 种特征权重的大小, 增强融合特征的表达能力。其次, 采用改进的特征分块注意力机制, 增强模板图像特征信息中有效区域的关注度, 实现更有效地目标相似度匹配。最后, 为降低计算成本, 将输出特征向量转换到 YCbCr 空间后进行离散余弦变换 (discrete cosine transform, DCT) 并保留低频分量, 得到特征响应图, 进行分类回归得到最终目标位置。实验表明, 该算法可以降低外观变化、外部因素干扰对无人机目标跟踪性能的影响, 提升目标跟踪的准确性。

1　本文算法原理

无人机在目标跟踪过程中由于需要频繁的变换姿态和位置, 调整飞行角度和飞行高度, 导致跟踪目标易发生外观变化、受到外部因素干扰的无人机在跟踪过程中出现误跟、漏跟、跟踪目标丢失的状况, 并产生漂移, 影响跟踪的整体精度。为保证无人机在复杂场景下能准确高效的完成目标跟踪任务, 本文提出一种基于特征融合和分块注意力的无人机目标跟踪算法。本文算法整体模型如图 1 所示, 该算法主要分为 3 个部分。第 1 部分为自适应特征融合策略, 通过从模板图像和搜索图像中分别提取 HOG 特征、CN 特征、深度卷积特征, 利用 3 个特征各自的优势, 自适应地融合这些特征, 从而提高目标特征的表达能力。第 2 部分为基于特征分块注意力机制的目标匹配算法, 通过模板图像中目标的关键部分计算注意力权重, 对模板图像的目标特征的不同部分赋予不同的权重, 增强对目标跟踪重要特征的关注并抑制对目标跟踪非重要特征的关注, 实现更有效地目标相似度匹配, 提升目标区域的特征表达能力; 将含权重的特征响应图转换到 YCbCr 通道, 通过 DCT 到频域并舍弃部分高频系数以减小计算成本的同时, 可以保留更多的人类语义信息。第 3 部分为分类回归分支, 将重组后的特征响应图输入到分类回归分支得到最终跟踪位置。

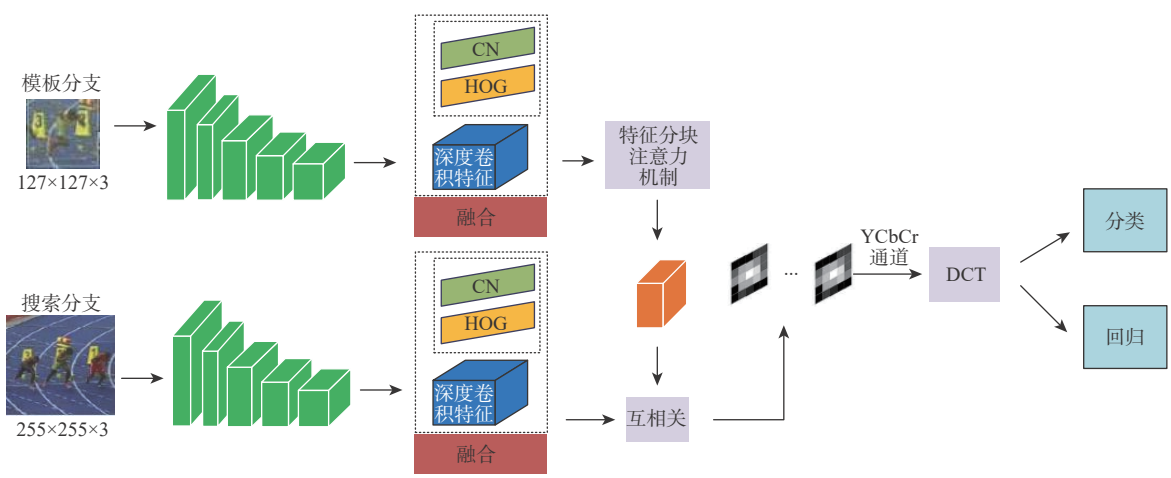


图 1　本文算法模型

Fig. 1　Model of the proposed algorithm

1.1　自适应特征融合策略

无人机在目标跟踪时面临外观变化及外部干扰等问题, 因此, 特征的选择和表达在跟踪过程中起着十分重要的作用。近年来, 深度网络凭借其能够提取到丰富语义信息, 且具有良好的通用性和鲁棒性, 逐渐成为目标跟踪领域研究的热点。但深度网络也存在一些问题, 首先, 基于深度网络的目标跟踪方法往往是通过微调网络模型或预训练的方式对目标进行跟踪, 随着网络模型的加深会产生梯度弥散、梯度爆炸的状况。其次, 由于深度网络对

样本数据量有较高的要求，在一些数据集中容易出现过拟合的现象，从而影响跟踪算法性能。另外，深度网络提取的深度特征更关注高级的语义信息，而局部刻画、几何不变性及光照变化相对于传统特征适应性较弱，忽视了传统特征在外观刻画方面具有的优势。传统特征如HOG特征、CN特征对于局部刻画、几何不变性及光照变化有很强的适应性。而无人机在复杂的自然场景中飞行时，跟踪目标经常面临外观变化、光照变化等问题。因此，传统特征对于无人机目标跟踪也具有重要的应用价值[14]。以往研究大多只采用传统特征或只采用深度网络提取特征的方式，分别在跟踪目标的底层特征和深度特征表达上有不同的优势[15]，但难以将跟踪目标的特征表达发挥充分和全面。综上所述，为提高目标跟踪算法在复杂场景下的适应性和准确性，结合2种特征各自的优势，将深度特征与传统特征进行融合，充分利用深度特征的高层语义信息和传统特征的低层细节信息，得到适应复杂场景下目标跟踪的融合特征。

本文提出一种自适应特征融合策略，旨在解决无人机在目标跟踪时面临外观变化、外部干扰的问题。自适应特征融合策略采用3种不同的特征进行自适应融合得到融合特征，增强特征的表达能力。HOG特征作为一种全局特征，对尺度变化和遮挡具有良好的适应性，并且具有一定的平移、旋转、光照不变性，能够提取出图像的空间颜色分布和光强分布等重要信息，HOG的特征提取方法如图2所示。CN特征也是一种全局特征，是图像最重要的组成部分之一，其对光照变化和噪声具有较好的鲁棒性，同时通过标准化的方法可以避免尺度变化对图像特征的影响，在目标跟踪方面具有较好的性能，CN特征的提取方法如图3所示。深度特征是通过深度网络对图像进行处理，提取出的高层语义信息，每层特征包含不同的目标信息，具有更好的鲁棒性和适应性。传统特征主要关注目标的外观信息，这些信息有利于对目标进行识别；深度特征具有丰富的语义信息，有利于无人机在在背景干扰、尺度变化等复杂场景中进行目标跟踪。为保证提取的特征具有丰富的语义信息，并保留准确的外观信息，从而更好地适应目标的外观变化并且防止模型漂移，因此，将深度特征和传统特征进行融合，利用各自的优势取长补短，增强特征的表达能力，提升跟踪算法的适应性。

具体操作步骤如下：

1）评估3种特征在融合特征中的重要程度，为3种特征自适应地分配融合权重。首先，分别计

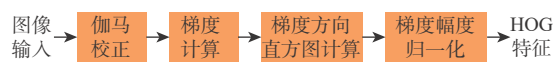


图2 HOG的特征提取方法

Fig. 2 Feature extraction method of HOG

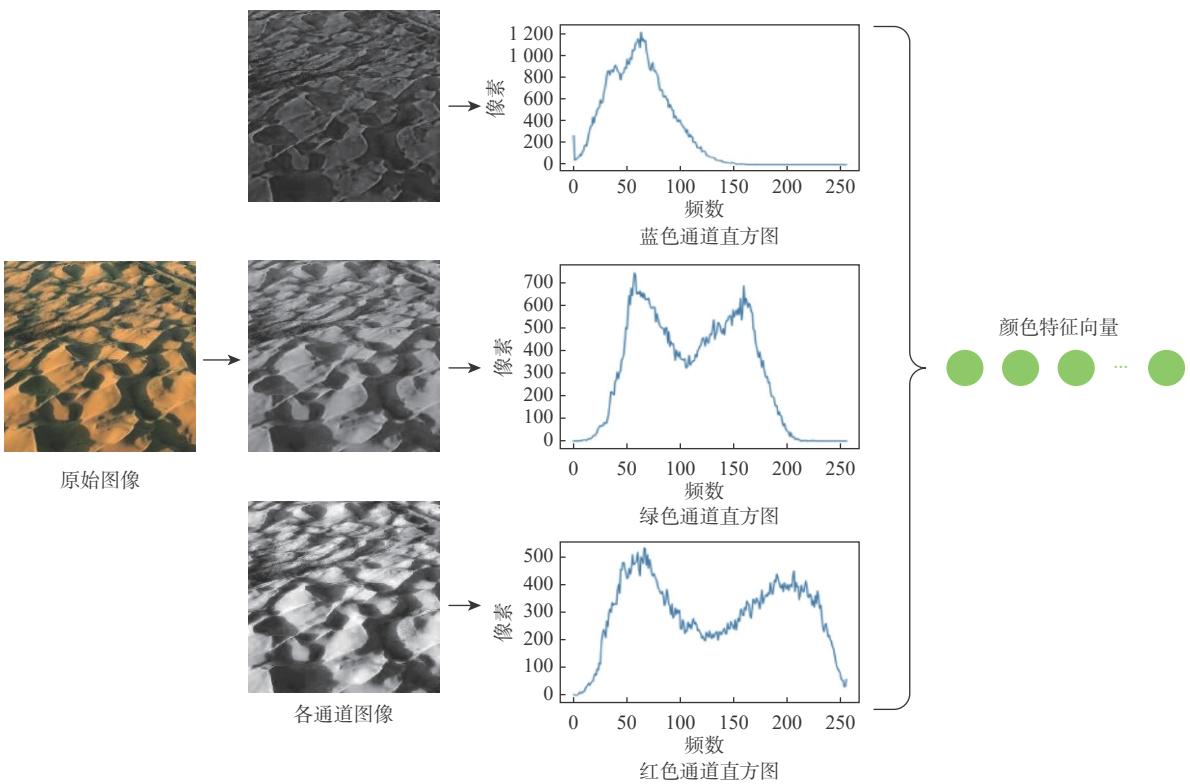


图3 CN的特征提取方法

Fig. 3 Feature extraction method of CN

算 3 种特征的最大特征响应值[16], 然后, 计算每种特征最大响应值的平方, 并将这些平方值相加, 得到 3 种特征响应值平方总和; 最后, 计算每种特征最大响应值的平方在总和中的占比, 得到每种特征的权重。3 种特征的权重分配如下:

$$w_{t\mathrm{HOG}} = \max^2(f_{\mathrm{HOG}}(Z))/(\max^2(f_{\mathrm{HOG}}(Z))+\max^2(f_{\mathrm{CN}}(Z))+\max^2(f_{\mathrm{deep}}(Z)) \quad (1)$$

$$w_{t\mathrm{CN}} = \max^2(f_{\mathrm{CN}}(Z))/(\max^2(f_{\mathrm{HOG}}(Z))+\max^2(f_{\mathrm{CN}}(Z))+\max^2(f_{\mathrm{deep}}(Z)) \quad (2)$$

$$w_{t\mathrm{deep}} = \max^2(f_{\mathrm{deep}}(Z))/(\max^2(f_{\mathrm{HOG}}(Z))+\max^2(f_{\mathrm{CN}}(Z))+\max^2(f_{\mathrm{deep}}(Z)) \quad (3)$$

式中: Z 表示搜索分支样本; $f_{\mathrm{HOG}}(Z)$、$f_{\mathrm{CN}}(Z)$、$f_{\mathrm{deep}}(Z)$ 分别为 HOG 特征、CN 特征和深度特征的特征响应值; t为视频帧数。

最大特征响应值通常与目标区域或目标候选区域相关, 通过将最大特征响应值作为权重分配的依据, 可以使网络在计算相似度时能够更关注目标的信息, 增强无人机在跟踪过程中对目标的关注程度, 抑制相似物干扰、背景干扰, 从而提高目标跟踪的准确性和稳定性。

采用最大响应值的平方计算特征权重更强调对跟踪影响较大的特征并抑制对跟踪影响较小的特征的关注, 更好地平衡多个特征之间的重要性。此外, 如果采用更高的次方值计算权重, 较大值的影响过于突出, 不利于平衡其他特征的影响。

2) 由于无人机处于不断变化的场景中, 因此, 不同帧搜索图像的图像复杂度不同。通过评估图像复杂度, 将图像复杂度高的搜索图像赋予更高的深度特征权重, 将图像复杂度低的搜索图像赋予更高的 HOG 特征和 CN 特征权重, 从而适应不同场景下的跟踪情况。本文采用信息熵估计图像复杂度:

$$H = -\sum P(X)\log_2(P(X)) \quad (4)$$

式中: X 为事件; $P(X)$ 为事件 X 发生的概率。

通过信息熵计算自适应系数L_1、L_2、L_3:

$$L_1 = H\lambda_{\mathrm{HOG}} \quad (5)$$

$$L_2 = H\lambda_{\mathrm{CN}} \quad (6)$$

$$L_3 = H\lambda_{\mathrm{deep}} \quad (7)$$

式中: λ_{HOG}、λ_{CN}、λ_{deep}为图像复杂度权重可调整参数。

3) 初始化搜索分支样本第 1 帧图像的 HOG 特征权重 $w_{t\mathrm{HOG}}$、CN 特征权重 $w_{t\mathrm{CN}}$和深度特征权重 $w_{t\mathrm{deep}}$, 将权重系数设置为 0.3、0.3 和 0.4。计算每帧 HOG 特征、CN 特征和深度特征的特征响应值, 并用 $f_{\mathrm{HOG}}(Z)$、$f_{\mathrm{CN}}(Z)$、$f_{\mathrm{deep}}(Z)$表示。根据特征响应值的大小, 给 3 个特征自适应地分配权重, 得到融合特征, 融合特征的特征响应值 $f_t(Z)$为

$$f_t(Z) = w_{t\mathrm{HOG}}L_1 f_{\mathrm{HOG}}(Z) + w_{t\mathrm{CN}}L_2 f_{\mathrm{CN}}(Z) + w_{t\mathrm{deep}}L_3 f_{\mathrm{deep}}(Z) \quad (8)$$

对于模板分支的特征响应值计算方法与搜索分支相同。通过采用自适应特征融合的方式, 利用 HOG 特征、CN 特征、深度卷积特征各自的优势取长补短, 灵活地调整 3 种特征的权重, 使其在不同场景下发挥最大的作用。

1.2　基于特征分块注意力机制的目标匹配算法

传统的 Siamese 网络通过直接计算模板图像特征和搜索图像特征的距离, 确定目标的具体位置。但是, 面对相似物干扰、背景干扰等外界因素干扰的问题, 仅采用直接距离计算的方式容易产生跟踪漂移, 难以满足无人机复杂场景目标跟踪的实际需求。综上所述, 提出一种基于特征分块注意力机制的目标匹配算法, 旨在解决无人机在目标跟踪时面临的相似物干扰、背景干扰等问题。模板图像特征中的所有特征并不是同等重要的, 深入挖掘模板图像特征信息的相关性和重要性, 将最有价值的信息以较低的计算成本进行计算, 最大可能的发挥最有价值信息的作用, 从而提高跟踪算法的性能。首先, 采用特征分块注意力机制[17], 对模板图像的不同部分赋予不同的权重, 通过增强对目标跟踪重要特征的关注并抑制对非重要特征的关注程度, 实现更有效地目标相似度匹配, 提升目标区域的特征表达能力。其次, 将通过特征分块注意力后的特征响应图转换到 YCbCr 空间后进行 DCT, 保留左上角 15 个低频信息分量, 在保留目标特征关键信息的同时, 减少无用信息对计算成本的影响。

具体操作步骤如下:

1) 在模板特征图像中搜索目标的关键部分: 搜索窗口的长和宽设置为模板特征图像长和宽的 1/2; 从模板特征图像左上角开始搜索, 使用滑动窗口以 1/4 边长为步幅从左至右、从上至下依次搜索目标的关键部分。通过滑动窗口搜索得到 9 个候选目标关键部分, 这些候选目标关键部分可以覆盖模板特征图像的各个区域, 并且降低冗余度, 如图 4 所示。从这些候选目标关键部分中寻找目标关键部分的策略是: 分别对 9 个候选目标关键部分在模板特征图像中覆盖其相应图像, 评价其对模板特征图像的影响, 找出对模板特图像征影响最大的候选目标关键部分; 反之, 则可以证明该候选目标关键部分就是目标的关键部分, 可以覆盖目标的多数特征并保留目标关键的空间和语义信息。

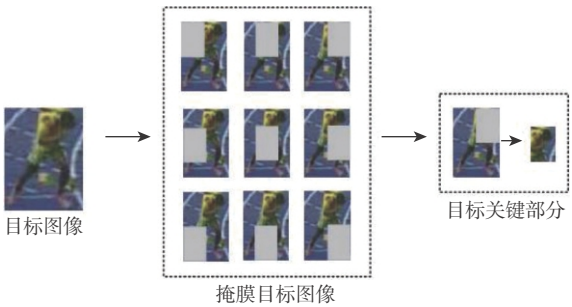


图4 模板关键部分寻找示意图

Fig. 4 Schematic diagram of template key parts to look for

评价候选目标关键部分对模板特征图像影响的方法：通过计算模板特征图像和模板特征图像被候选目标关键部分覆盖后图像之间的内积，找出对模板特征图像影响最大的候选目标关键部分，内积值最小则表明大部分目标关键信息丢失，其对应的候选目标关键部分比其他目标部分更具表现力。由于HOG特征具有不受图像尺寸和旋转影响、有效表达图像纹理信息、低复杂度、不受亮度和对比度变化影响的优点，内积采用目标关键部分的HOG特征作为计算对象，计算式为

$$W = \arg\min \mathrm{HOG}(O) \cdot \mathrm{HOG}(O_{\mathrm{mask}}(i)) \tag{9}$$

式中：$i \in \{1, 2, \cdots, 9\}$表示第$i$个的候选目标关键部分；$O$为模板特征图像；$O_{\mathrm{mask}}(i)$表示被候选目标关键部分覆盖后的模板特征图像。

2）将计算后的9个内积值进行排序，选择最小内积值对应的图像作为模板特征图像的目标关键部分用来计算注意力权重，如图5所示，具体计算方法如下：

①将模板特征图像与目标关键部分特征图像经过卷积变换后，再进行互相关操作，得到注意力权重响应图：

$$\alpha = \mathrm{corr}(f(C), f(E)) \tag{10}$$

②将权重响应图与模板特征图像相结合，对模板特征图像进行权重分配，如式(11)所示。通过突出模板特征图像中重要的部分，并减少不重要的部分来提高匹配性能。

$$c = \alpha f(E) \tag{11}$$

③将权重分配后的模板特征图像和搜索特征图像进行互相关，得到特征响应图：

$$R = \mathrm{corr}(c, f(S)) \tag{12}$$

将式(10)～式(12)合并，得到分块注意力机制下的特征响应图：

$$R = \mathrm{corr}(\mathrm{corr}(f(C), f(E)) \cdot f(E), f(S)) \tag{13}$$

式中：corr表示互相关运算；$f(C)$、$f(E)$和$f(S)$分别为关键部分、模板图像和搜索图像的输出特征图像。

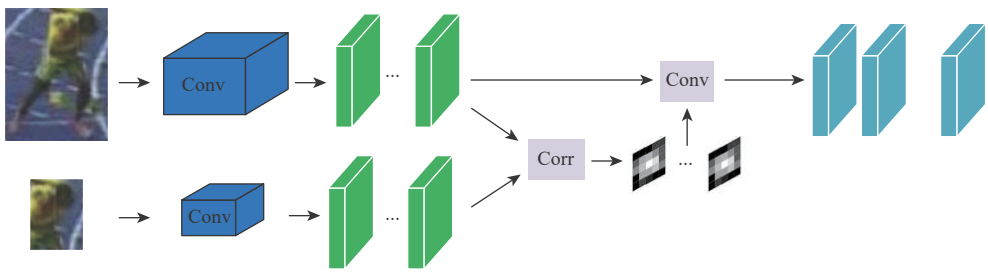


图5 注意力权重计算模型

Fig. 5 Model of attention weight calculation

3）为减少计算成本和计算时间，采用YCbCr-DCT的特征响应图变换方法，保留目标特征的关键信息，减少无用信息对计算成本的影响，模型如图6所示。DCT是一种将时域信号转换为频域信号的线性变换，其特点是能够将信号的能量主要集中在低频部分，通过舍弃部分高频信号的方法可以对信号的能量信息进行压缩。在提取特征时，只需要保留DCT系数矩阵中左上角部分低频系数[18]，这样既可以保证图像关键信息，又可以减少计算量。

具体算法步骤如下：

①将特征响应图转换到YCbCr空间。YCbCr空间由亮度(Y)和色度(Cb、Cr)2个分量构成，更符合人类视觉系统对颜色的感知方式，使用YCbCr空间进行变换、重组和通道选择可以保留更多的语义信息。

②将特征响应图按8×8分块后，将不同分块中相同频率的DCT系数重组为二维DCT矩阵，进而构成三维DCT矩阵。这样可以解决无人机视频帧图像尺寸大、深度学习输入图像尺寸小，图像尺寸重塑操作造成的信息损失。通过这一步骤，将DCT重组后，只需修改输入通道数，就能够以较小的尺寸将更多的图像语义信息输入到模型。

③对三维DCT矩阵根据zigzag操作选取的15个低频分量后进行通道选择，保留绝大多数图像关键信息的同时大幅减小了输入的通道数，从而减少计算复杂度、降低计算所耗时间。

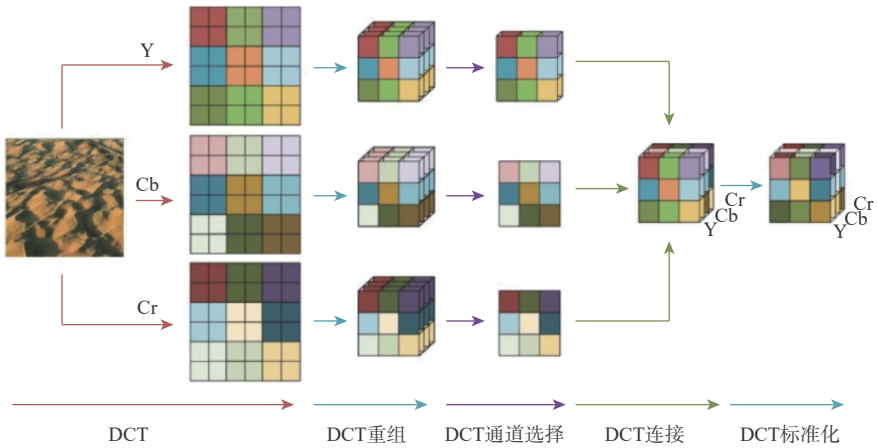


图 6　YCbCr-DCT 的特征响应图变换模型

Fig. 6　Characteristic response graph transformation model of YCbCr-DCT

④将 Y、Cb、Cr 各自三维 DCT 矩阵中选定的通道连接在一起，组成共 45 通道的三维 DCT 矩阵，根据视频帧图像计算对应通道的均值和方差，利用其对每个频率通道进行标准化，得到重塑后的特征响应图。

1.3　分类回归分支

在目标跟踪算法确认目标最终位置时，需要通过大量候选预测框进行筛选，导致网络的计算成本高，降低网络计算效率。为降低网络的计算成本并且提升跟踪精度，本文采用一种自适应边界框分类回归策略确认目标的最终位置。该策略的优点在于不需要设置含有调节参数的边界框，从而降低了网络对预测框尺度属性先验预设的依赖[19]，提升了网络预测框的泛化能力，使其能够更自由地预测目标的位置和大小，而不受先验知识的限制。

在训练过程中，使用 CIoU[20] 损失函数计算回归得分的损失，使用交叉熵损失函数计算分类得分的损失，计算方法如式(14)和式(15)所示。CIoU 损失相对于传统的 IoU[21] 损失包含的因素更加全面，综合考虑了预测框与真实框之间的重叠面积、中心点距离及宽高比等因素，从而更加全面地评估预测框的准确性。另外，CIoU 使预测框和真实框的宽高比尽量保持一致，加强对深度网络的收敛效果，为网络的训练过程提供更加优化的方向。

$$L_{\mathrm{cls}}=-\frac{1}{N}\sum_{i}\sum_{c=1}^{M}y_{ic}\ln p_{ic} \tag{14}$$

$$L_{\mathrm{reg}}=1-L_{\mathrm{IoU}}+\frac{\rho^2(b,b^{\mathrm{gt}})}{c^2}+av \tag{15}$$

式中：L_{IoU}为 IoU 损失函数；$\rho^2(b,b^{\mathrm{gt}})$为预测框中心点 b 与真实框中心点 b^{gt}之间的欧氏距离；c 为包含预测框面积与真实框面积的最小长方形对角线距离；M 为目标类别的数量；y_{ic}为符号函数，如果样本 i 的真实类别等于 c，则取 1，否则取 0；p_{ic}为观测样本 i 属于类别 c 的预测概率；a为一个平衡参数；v用于计算预测框和目标框的高宽比的一致性，本文用 tan 角来衡量：

$$v=\frac{4}{\pi^2}\left(\arctan\frac{\omega^{\mathrm{gt}}}{h^{\mathrm{gt}}}-\arctan\frac{w}{h}\right)^2 \tag{16}$$

其中：ω^{gt}和h^{gt}分别为真实框的宽和高；w 和 h 分别为预测框的宽和高。

根据预测框和真实框之间的 IoU 值来自适应地调整不同部分的权重，使得损失函数更加准确地反映预测框和真实框之间的差异，平衡参数a的计算式为

$$a=\frac{v}{(1-I_{\mathrm{OU}})+v} \tag{17}$$

最后，将负责评估分类代价的分类损失 L_{cls} 与负责评估回归代价的定位损失 L_{reg} 相加，得到总损失 L，计算方法为

$$L=L_{\mathrm{cls}}+L_{\mathrm{reg}} \tag{18}$$

1.4　本文算法步骤

步骤 1　通过模板分支提取目标模板图像的自适应融合特征$\boldsymbol{f}_{\mathrm{e}}$；

步骤 2　通过搜索分支提取视频后续帧图像的自适应融合特征$\boldsymbol{f}_{\mathrm{i}}$；

步骤 3　将模板特征图像分为 9 个候选目标关键部分，通过计算内积寻找影响模板特征图像最大的目标关键部分；

步骤 4　将目标关键部分与模板特征图像进行互相关，得到注意力权重响应图后与模板特征图像进行卷积，得到含权重分配的模板特征图像；

步骤 5　将含权重分配的模板特征图像和搜索特征图像进行互相关，得到特征响应图；

步骤 6　将特征响应图转换到 YCbCr 空间后进

行DCT,保留左上角15个低频分量;

步骤7 将重塑后的特征响应图送入分类回归分支,得到最终目标位置;

步骤8 重复步骤2～步骤7直到视频结束。

本文算法流程如图7所示。

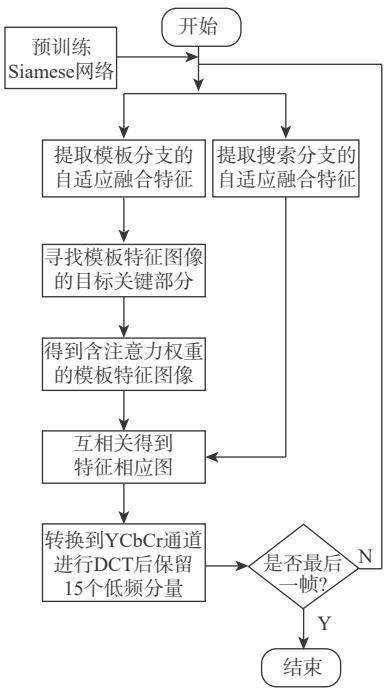


图7 本文算法流程图

Fig. 7 Flow chart of the proposed algorithm

2 仿真实验

实验平台采用Inter Core i7-7700K CPU 4.2 GHZ,图形处理器采用NVIDIA GeForce RTX2080Ti,编程环境采用Python3.7。设置训练的迭代次数为40,小批次训练样本数量为10,每组训练样本数量大小为4×10^5。设置初始学习率为1×10^{-3},权重衰减率为1×10^{-4}。

根据实验需要,本文采用LaSOT[22]作为实验的训练数据集,该数据集有1 400个视频序列,每个视频平均有2 512帧。LaSOT数据集含有70个类别,每个类别由20个视频序列组成,避免了分类不平衡造成算法训练时出现失衡的现象。LaSOT数据集不仅对每帧进行了高准确的密集标注,并且增加了丰富的自然语言描述。为验证算法的有效性,本文采用UAV123数据集[23]作为实验的测试数据集,总共包含123个视频序列和超过110K帧的图像合集。UAV123数据集包含10余种目标跟踪难点,是一个场景复杂的无人机航拍数据集。

实验中,采用中心位置误差取值为20时对应的跟踪精确率和覆盖率取值0.5时对应的跟踪成功率作为目标跟踪算法的评价指标。

其中,成功率衡量了跟踪器在整个跟踪序列中成功跟踪目标的能力。成功跟踪帧数是指在所有跟踪帧中,给定的真实目标框和跟踪所得的目标框面积重叠程度大于给定阈值的视频帧数。

精确率衡量了跟踪器在跟踪过程中对于目标位置的准确性。正确跟踪帧数是在所有跟踪帧中,预测位置与真实目标位置之间平均欧氏距离小于给定阈值的视频帧数占总视频帧数。

2.1 消融实验

为验证自适应特征融合策略和改进的分块注意力机制对算法产生的影响,本文采用UAV123作为验证数据集,设计如表1所示的消融实验。其中,FA为自适应特征融合策略,PA为采用特征分块注意力机制。由表1可知,本文算法相较于只采用Siamese网络进行单一特征提取的算法而言,其成功率提高18.3%,精确率提高14.7%;本文算法相较于未添加自适应特征融合策略的算法而言,其成功率提高12.5%,精确率提高9.1%;本文算法相较于未采用特征分块注意力机制而言,其成功率提高9.2%,精确率提高6.3%。证明本文算法可以有效的提升目标跟踪的精确率。

表1 消融实验

Table 1 Ablation experiment

算法	精确率	成功率
Siamese	0.704	0.672
Siamese+FA	0.792	0.734
Siamese+PA	0.769	0.715
本文算法	0.832	0.771

2.2 定量分析

为多方面验证本文算法的有效性,实验采用UAV123数据集中12种跟踪难点的视频序列对本文算法和主流的无人机目标跟踪算法进行对比。其中,UAV123视频序列包含尺度变换、宽高比变换、低分辨率、快速移动、完全遮挡、部分遮挡、超出视野、背景干扰、光照变化、视角变化、摄像机移动、相似物体这12种不同场景的跟踪难点。与本文算法对比的主流无人机目标跟踪算法包括ECO[24]、ECO-HC[24]、SiamRPN[8]、SiamFC[4]、SiamRPN++[9]、SiamDW[25]、SiamCAR[26]、SiamFC++[27]、SiamAttn[28]。10种算法针对12种不同跟踪难点的跟踪精确率和跟踪成功率对比如图8所示,跟踪成功率对比如图9所示,算法的整体跟踪精确率和成功率如图10所示。

由图8和图9可知,本文算法在背景干扰,完全遮挡,相似物体、光照变化这几种复杂跟踪环境下,跟踪效果相较于其他主流算法效果更好。这是

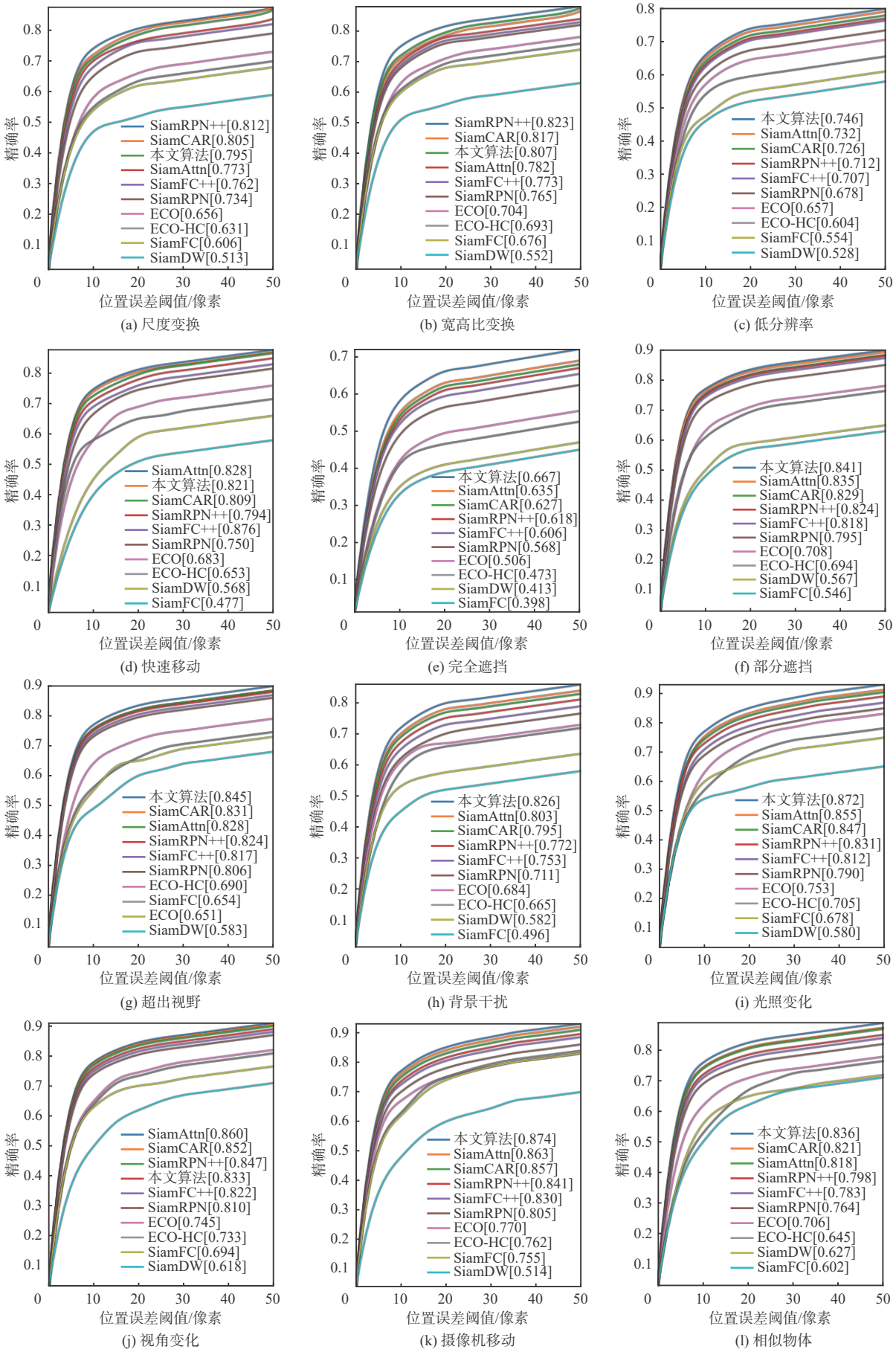


图 8　各算法的跟踪精确率对比

Fig. 8　Comparison of tracking accuracy for various algorithms

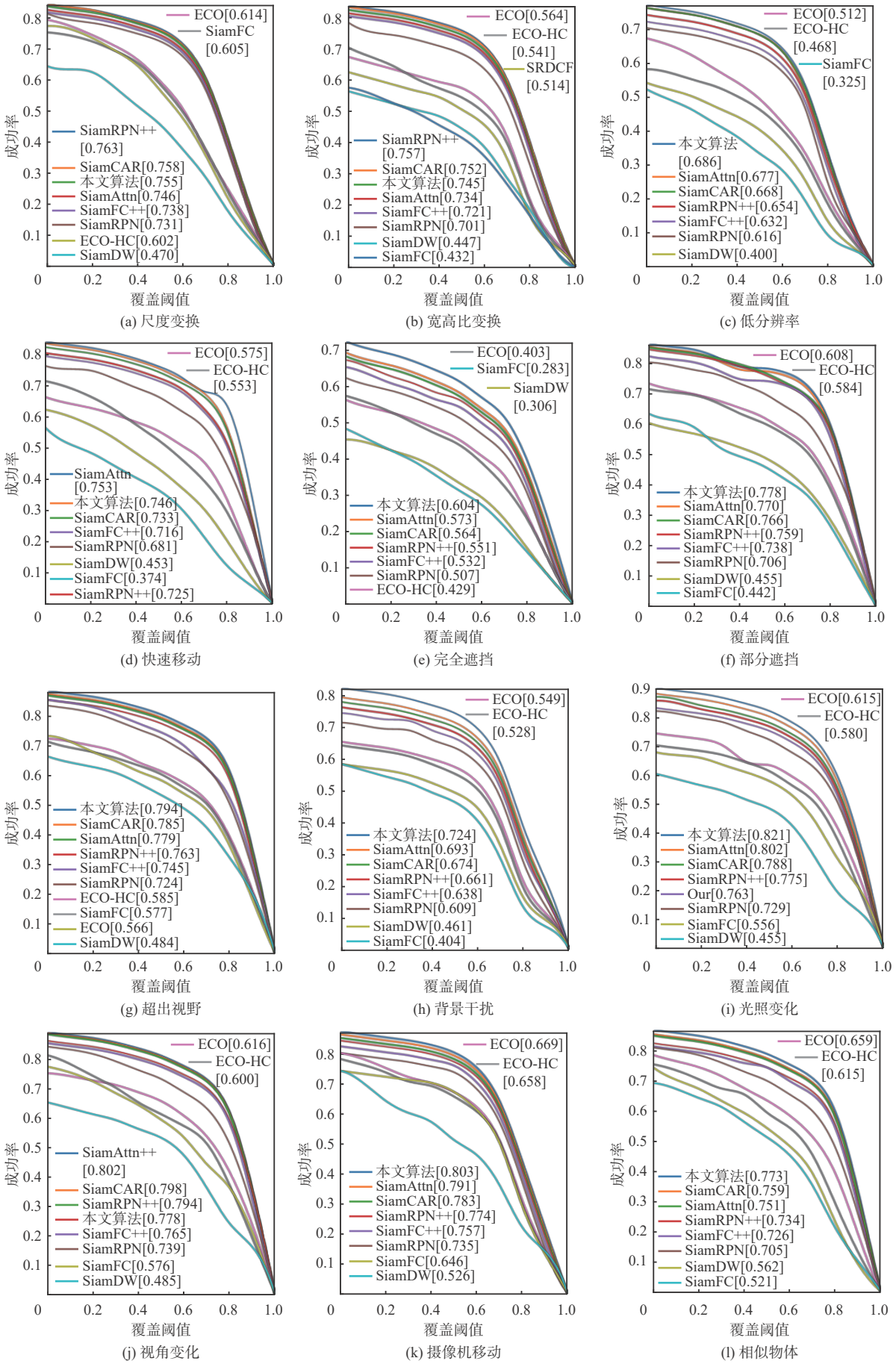
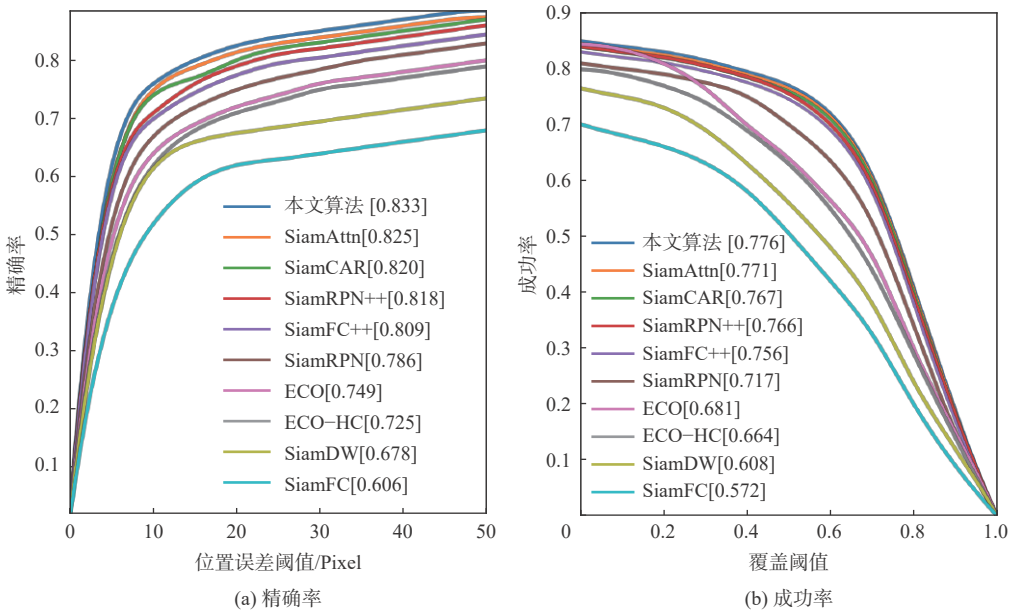


图 9 各算法的跟踪成功率对比

Fig. 9 Comparison of tracking success rates of various algorithms



(a) 精确率　　(b) 成功率

图 10　整体性能比较

Fig. 10　Overall performance comparison

因为这些复杂环境的主要跟踪难点都是考验算法对跟踪目标特征的表达能力及对跟踪目标的识别能力。通过自适应特征融合策略对 3 种表达不同语义信息的特征进行自适应融合得到融合特征, 增强无人机对跟踪目标特征的表达能力。通过采用特征分块注意力机制对目标的不同部分赋予不同的权重, 增强跟踪目标重要的部分并抑制跟踪目标的非重要部分, 从而使无人机在目标跟踪的过程中抓住目标的关键信息, 提升无人机对跟踪目标的识别能力。面对其他的复杂环境, 本文算法也可以进行较好的处理, 这是因为本文主要任务就是针对目标的特征进行操作, 丰富的表达特征和关键特征的选取使得无人机在目标跟踪的过程中适应目标的各种变化及外界的干扰因此能够更好的提升目标跟踪的准确率。

由图 10 可知, 本文算法的整体跟踪精确率和跟踪成功率分别为 0.833 和 0.776, 优于所有参与对比的主流目标跟踪算法。

2.3　定性分析

为更加直观地验证本文算法的跟踪效果, 从 UAV123 数据集中选取 5 种不同跟踪难点的视频序列进行可视化分析, 将本文算法与 SiamFC、SiamRPN、SiamRPN++、SiamCAR、SiamFC++这 5 种跟踪算法进行对比, 可视化分析结果如图 11 所示。

针对光照变化的情况, 由图 11(a) 可知, 本文算法的跟踪效果优于对比算法。只有本文算法能够持续对目标进行正确跟踪, 而其他跟踪器会因光照变换而使预测框偏移。

针对复杂环境内相似物体的干扰和背景干扰的情况, 由图 11(b) 和图 11(c) 可知, 本文算法较对比算法表现更佳。除本文算法外的其他算法会受到周围相似物和背景的干扰, 使预测框完全偏移, 且在一段时间内无法准确回归。同时, 本文算法的特征分块注意力机制可以抓住目标的关键特征且抑制干扰特征, 进而实现全程鲁棒、稳定的跟踪效果。

针对目标部分遮挡、消失重现的情况, 从图 11(d) 和图 11(e) 可以看出, 本文算法较对比算法表现更佳。除本文算法外的其他算法面对遮挡及消失重现的情况下会出现误判目标的位置信息的问题, 造成跟踪目标漂移; 而本文算法由于采用了 3 种特征融合并自适应分配权重, 增强特征的表达能力, 更好地适应目标的外观变化, 防止模型漂移, 实现持续、稳定的跟踪目标, 相较于其他算法具有更高的跟踪成功率。

2.4　实时性和复杂度分析

实验采用画面每秒传输帧数(frame per second, FPS)作为衡量算法检测速度的指标, 表示在 1 s 内该算法能够处理的图像帧数。实验采用每秒浮点运算次数(floating point operations per second, FLOPs)作为衡量算法复杂度的指标, 可反映模型的计算量大小。实验采用期望平均重叠率(expected average overlap, EAO)作为衡量跟踪算法的准确性和稳定性的综合指标, 用于评估算法在不同场景下的平均性能。实验采用了 SiamFC、SiamRPN、SiamRPN++、SiamFC++和 SiamCAR 与本文算法进行比较。

如表 2 所示, 本文算法在衡量跟踪算法准确性和稳定性的综合指标 EAO 上达到 0.543, 优于其他 5 种对比跟踪算法。同时, 相较于其他 EAO 较高的

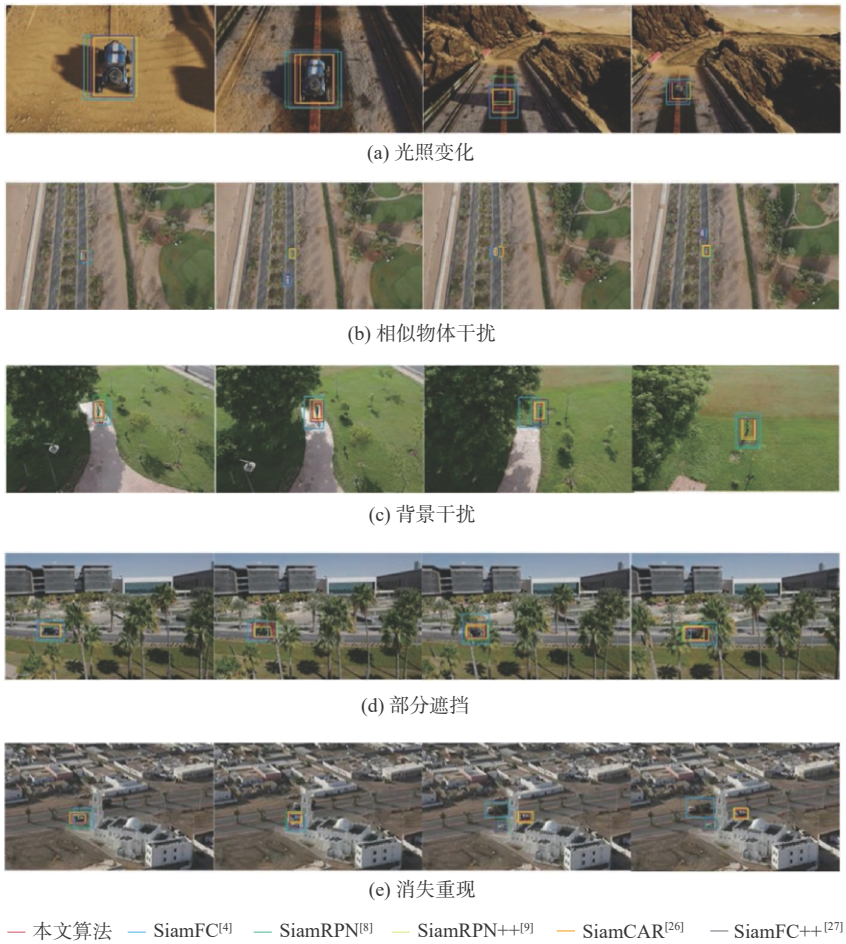


图 11 不同难点可视化分析

Fig. 11 Visual analysis of different difficulties

表 2 不同算法的 3 类指标对比

Table 2 Comparison of three indexes of different algorithms

算法	FPS	FLOPs	EAO
SiamFC[4]	58	4.9×10^6	0.348
SiamRPN[8]	83	6.8×10^6	0.483
SiamRPN++[9]	41	7.8×10^6	0.517
SiamFC++[27]	90	7.4×10^6	0.512
SiamCAR[26]	52	12.1×10^6	0.525
本文算法	60	7.6×10^6	0.543

模型，尤其是与特征处理相关的 SiamRPN++与 SiamCAR 相比，本文算法的计算量指标 FLOPs 相对较低。在反映实际场景下图像处理速度的 FPS 指标上，本文算法达到了 60，完全满足无人机实时跟踪的需求。原因在于：①本文算法采用了 YCbCr-DCT 特征响应图变换方法，保留了目标特征的关键信息，并降低无用信息对计算成本的影响，减少了计算时间，进而在保证跟踪准确率和稳定性的同时，提升图像处理速度；②本文算法采用一种自适应边界框分类回归策略，不需要对含有调节参数的大量边界框进行筛选，从而提升算法的计算效率。

3 结 论

本文提出一种基于特征融合和分块注意力的无人机跟踪算法。旨在解决无人机在目标跟踪时面临的外观变化和外部干扰等问题。

1）本文提出一种自适应特征融合策略，通过融合 HOG 特征、CN 特征和深度特征，自适应计算 3 种特征权重的大小，增强融合特征的表达能力。

2）本文提出一种特征分块注意力机制，对模板特征图像的不同部分赋予不同的权重，增强对目标跟踪重要特征的关注度并抑制非重要特征的关注度，提升目标区域的特征表达能力，从而提升无人机对跟踪目标的识别能力。

实验结果表明，与其他无人机目标跟踪算法相比，本文算法拥有较好的性能，能够在减少计算成本的前提下，显著的提高跟踪精度。

参考文献（References）

[1] 管皓，薛向阳，安志勇. 深度学习在视频目标跟踪中的应用进展与展望[J]. 自动化学报，2016，42(6)：834-847.

GUAN H，XUE X Y，AN Z Y. Advances on application of deep

learning for video object tracking[J]. Acta Automatica Sinica, 2016, 42(6): 834-847(in Chinese).

[2] 李靖, 马晓东, 陈怀民, 等. 无人机视觉导航着陆地标实时检测跟踪方法[J]. 西北工业大学学报, 2018, 36(2): 294-301.
LI J, MA X D, CHEN H M, et al. Real-time detection and tracking method of landmark based on UAV visual navigation[J]. Journal of Northwestern Polytechnical University, 2018, 36(2): 294-301(in Chinese).

[3] DANELLJAN M, HÄGER G, KHAN F S, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision Workshop. Piscataway: IEEE Press, 2015: 621-629.

[4] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional Siamese networks for object tracking[C]//Proceedings of the Computer Vision–ECCV 2016 Workshops. Berlin: Springer, 2016: 850-865.

[5] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 5000-5008.

[6] HU W M, WANG Q, GAO J, et al. DCFNet: discriminant correlation filters network for visual tracking[J]. Journal of Computer Science and Technology, 2024, 39(3): 691-714.

[7] 刘芳, 杨安喆, 吴志威. 基于自适应 Siamese 网络的无人机目标跟踪算法[J]. 航空学报, 2020, 41(1): 323423.
LIU F, YANG A Z, WU Z W. Adaptive Siamese network based UAV target tracking algorithm[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(1): 323423(in Chinese).

[8] LI B, YAN J J, WU W, et al. High performance visual tracking with Siamese region proposal network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8971-8980.

[9] LI B, WU W, WANG Q, et al. SiamRPN: evolution of Siamese visual tracking with very deep networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4277-4286.

[10] 白鑫宇. 基于特征融合的相关滤波目标跟踪算法研究[D]. 重庆: 重庆邮电大学, 2022.
BAI X Y. Research on target tracking algorithm of correlation filtering based on feature fusion[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2022(in Chinese).

[11] 韩江龙. 基于相关滤波框架下多特征融合目标跟踪算法的研究[D]. 秦皇岛: 燕山大学, 2019.
HAN J L. Research on multi-feature fusion target tracking algorithm based on correlation filtering framework[D]. Qinhuangdao: Yanshan University, 2019(in Chinese).

[12] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2005: 886-893.

[13] VAN DE WEIJER J, SCHMID C, VERBEEK J, et al. Learning color names for real-world applications[J]. IEEE Transactions on Image Processing, 2009, 18(7): 1512-1523.

[14] 胡阳光, 肖明清, 张凯, 等. 传统特征和深度特征融合的红外空中目标跟踪[J]. 系统工程与电子技术, 2019, 41(12): 2675-2683.
HU Y G, XIAO M Q, ZHANG K, et al. Infrared aerial target tracking based on fusion of traditional feature and deep feature[J]. Systems Engineering and Electronics, 2019, 41(12): 2675-2683(in Chinese).

[15] 司海平, 万里, 王云鹏, 等. 基于特征融合的玉米品种识别[J]. 中国粮油学报, 2023, 38(12): 191-196.
SI H P, WAN L, WANG Y P, et al. Maize variety recognition based on feature fusion[J]. China Journal of Grain and Oils, 2023, 38(12): 191-196 (in Chinese).

[16] ZHANG W C, CHEN Z, LIU P Z, et al. Siamese visual tracking with robust adaptive learning[C]//Proceedings of the IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification. Piscataway: IEEE Press, 2019: 153-157.

[17] SHEN J B, TANG X, DONG X P, et al. Visual object tracking by hierarchical attention Siamese network[J]. IEEE Transactions on Cybernetics, 2020, 50(7): 3068-3080.

[18] 路丽霞. 基于深度网络的无人机地貌图像分类算法研究[D]. 北京: 北京工业大学, 2018.
LU L X. Research on UAV landform image classification algorithm based on depth network[D]. Beijing: Beijing University of Technology, 2018(in Chinese).

[19] CHEN Z D, ZHONG B N, LI G R, et al. Siamese box adaptive network for visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6667-6676.

[20] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[21] YU J H, JIANG Y N, WANG Z Y, et al. UnitBox[C]//Proceedings of the 24th ACM International Conference on Multimedia. New York: ACM, 2016: 516-520.

[22] FAN H, LIN L T, YANG F, et al. LaSOT: a high-quality benchmark for large-scale single object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5369-5378.

[23] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking[J]. Springer International Publishing, 2016, 9905: 445-461.

[24] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6931-6939.

[25] ZHANG Z P, PENG H W. Deeper and wider Siamese networks for real-time visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 4586-4595.

[26] GUO D Y, WANG J, CUI Y, et al. SiamCAR: Siamese fully convolutional classification and regression for visual tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6268-6276.

[27] XU Y D, WANG Z Y, LI Z X, et al. SiamFC++: towards robust and accurate visual tracking with target estimation guidelines[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2020, 34(7): 12549-12556.

[28] YU Y C, XIONG Y L, HUANG W L, et al. Deformable Siamese attention networks for visual object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6727-6736.

UAV tracking algorithm based on feature fusion and block attention

LIU Fang[1], YANG Yuyan[1,*], WANG Xin[2]

(1. School of Information and Communication Engineering, Beijing University of Technology, Beijing 100124, China;
2. Fengtai Power Supply Bureau of Beijing Power Supply Bureau, Beijing 100161, China)

Abstract: Unmanned aerial vehicle (UAV) has been widely used in various fields, target tracking is one of the key technologies of UAV applications. A UAV tracking algorithm based on feature fusion and segmented attention is proposed to solve the problems of UAV appearance changes and external interference when tracking targets. To improve the expression ability of fusion features, the weights of the three features are adaptively determined after the Siamese network has extracted histogram of oriented gradients (HOG), color names (CN), and deep convolution features from template and search images. Secondly, the improved feature segmentation attention mechanism is used to enhance the attention of the effective region in the template image feature information, so as to achieve more effective target similarity matching. The resultant feature vector is then transformed to YCbCr space to lower the computation cost. The feature response graph is then obtained using the discrete cosine transform (DCT), and classification regression is used to determine the final target location. Experiments show that the algorithm can reduce the influence of appearance change and external factors on tracking performance, and improve the accuracy of target tracking.

Keywords: unmanned aerial vehicle; target tracking; feature fusion; characteristic block attention; Siamese network

Received: 2023-05-25; **Accepted**: 2023-08-01; **Published Online**: 2023-09-27 18:40
URL: link.cnki.net/urlid/11.2625.V.20230927.1330.002
* **Corresponding author**. E-mail: yangyy115@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0259

基于时效网络的关键航空器识别方法

王红勇*，马丽书，许平
(中国民航大学 空中交通管理学院，天津 300300)

摘 要：针对空中交通态势中关键航空器识别问题，现有研究未能充分考虑空中交通实际运行中的时空效应。因此，提出一种基于时效网络的关键航空器识别方法。利用航空器间汇聚关系及其复杂性，通过邻居拓扑重叠系数构建时效网络模型，并基于特征向量中心性确定关键航空器。对关键航空器节点进行网络攻击观察扇区复杂性变化情况，并对比基于静态网络指标下的攻击，采用改进遗传算法为网络攻击删除的航空器节点分配新的进扇区时刻，从而验证关键航空器的选取效果。通过实际数据验证表明：所提方法相较于静态网络攻击在移除关键航空器时更高效地降低了扇区平均复杂性，改进的遗传算法求解关键航空器进扇区时刻分配问题时收敛性更高，使得扇区复杂性在一定时间段内更加平稳。分析关键航空器的控制效果表明：所提方法相较于静态网络更能准确地识别一段时间内对扇区复杂性影响较大的航空器。

关 键 词：航空运输；扇区复杂性；关键航空器；空中交通态势时效网络；网络攻击；遗传算法

中图分类号：V355

文献标志码：A **文章编号**：1001-5965(2025)05-1579-12

随着民航运输业快速发展，空中交通流量大幅增加，空中运行环境日益复杂，给空中交通管制员带来了巨大的工作负荷。基于交通复杂性的空中交通管理可有效评价管制员工作负荷并对其进行合理控制，其基本思路是预先识别对扇区复杂性有较大影响的关键航空器，并通过对关键航空器的调配降低扇区整体复杂性。

空中交通复杂性评估研究，有基于密度[1-2]、动态密度[3-4]、交通流扰动[5-6]和内禀性(相对距离及速度)等多种思路。基于密度和动态密度的复杂性评价方法的核心是通过全面评价管制员为保障空中交通安全而产生的物理和精神上的各种活动，采用线性或非线性的方法确定因子权重，最后，综合相加得出该扇区的复杂程度。虽然易被管制员理解，但为使特定扇区与管制负荷达到最大相关性，必须根据不同扇区选择不同复杂性因子，且不同扇区中复杂性因子的权值都需特别设定，以致该方法一直以来不具普适性。而基于内禀性的复杂性评估方法可以提供更为客观的评估结果，是空中交通复杂性研究的趋势。Delahaye 等提出内禀性的概念[7-8]，并且通过动态系统进一步对空中交通复杂性进行计算分析[9]。张进等在文献 [7-8] 的基础上进一步提出综合迫近、几率和连携因素的复杂性模型[10-11]。徐肖豪等在文献 [10-11] 的模型中进一步拓展，加入了穿越态势，提出三维内禀性的复杂性模型[12]。王红勇和郭宇鹏针对航空器自主运行模式，评估空中复杂性，并验证了该方法在冲突探测和航迹规划中的应用[13]。

在关键航空器识别研究中，复杂网络是研究的重要方法之一。在复杂网络中，人们发现少数节点

收稿日期：2023-05-22；**录用日期**：2023-09-09；**网络出版时间**：2023-10-12 11:10
网络出版地址：link.cnki.net/urlid/11.2625.V.20231010.1553.003
基金项目：国家自然科学基金 (U1833103)；天津市自然科学基金 (21JCZDJC00840)
* **通信作者**. E-mail：hy_wang@cauc.edu.cn

引用格式：王红勇，马丽书，许平. 基于时效网络的关键航空器识别方法 [J]. 北京航空航天大学学报，2025，51（5）：1579-1590.
WANG H Y，MA L S，XU P. Critical aircraft identification method based on temporal network[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1579-1590（in Chinese）.

在网络中起着关键性作用，甚至影响网络的结构和运行[14]。确定网络中的关键节点，可以获得有关实体重要性的先验知识从而预测事情的发展进程[15]。评估节点重要性有静态网络和时效网络2种建模方式。最直接的指标是度中心性，其认为邻居节点数量越多，则该节点的重要性越大[16]。后续研究中，人们发现一些网络中度很小却又很重要的节点，这些节点通常扮演着“桥梁”的角色，在传递信息中担任重要的角色，随之提出接近中心性[17]、衰减效应[18]等。在时效网络中，Taylor等进行了多层耦合网络分析，以基于时效网络的层内关系和层间关系建立了经典的超邻接矩阵(supra-adjacency matrix, SAM)模型[19]，并定义了特征向量中心性指标确定关键节点。Carletti等研究了一类在这种高阶结构上定义的随机游走，并基于微观物理模型，研究多体接近与属于同一超边缘的代理之间的高度可能的交换相关联[20]。Rozenshtein和Gionis改进了PageRank并定义了时效PageRank值，认为其值越大，表明节点越重要[21]。Huang和Yu从传播动力学角度基于易感者-感染者-治愈者(susceptible infective recovered, SIR)模型，提出时效网络中有影响力传播者基于动力学敏感的中心性测度方法[22]。Cai等建立了空中交通延误传播时效网络模型，并以美国和东南亚地区的实际航班数据进行了传播验证[23]。以上研究在复杂网络基础理论上从不同角度研究了节点重要性的评估方法。

对于复杂网络在航空器交互关系中的应用，大量研究表明航空器关系网络中同样存在类似的关键节点。在飞行冲突网络的研究中，黄洋等通过计算每个航空器的边缘权重之和，将关键节点定义为冲突边缘权重之和最大的航空器[24]。刘飞等采用复杂网络理论中的节点度中心性、接近中心性及PageRank指标结合层次分析法设计了关键航空器识别方法[25]。吴明功等在文献[25]的基础上选取节点度、点强、加权聚类系数和介数共4个参数作为评估节点重要度指标，利用层次分析法确定各指标权重，并结合熵权法的思想，对结果进行修正，通过多属性决策的方法计算出节点重要度，确定关键冲突飞机[26]。王兴隆和苗尚飞通过建立扇区网络模型，对比分析不同恢复策略下的韧性指标，制定恢复策略，提高网络韧性[27]。

以上关键航空器的识别研究中从不同角度分析航空器重要性的评价指标，然而研究仅考虑了某一瞬间的静态网络，忽视了网络在未来一段时间内的发展状态。但在实际的运行中，扇区内航空器间关系往往表现出很强的时空性，经常存在某时刻航空器重要性较低，但下一时刻由于飞行冲突激增等原因导致重要性急剧增加。或该时刻航空器重要性很高，但下一时刻重要性骤降。针对以上不足，本文在现有研究的基础上，首先，以SAM无权时效模型为基础，结合航空器间复杂性关系建立加权时效空中交通态势时效网络模型；然后，依据特征向量中心性确定关键航空器节点；最后，在时效网络中移除关键航空器节点并重新计算新的进扇区时刻，再将移除的航空器节点按新的时刻加入时效网络对应时间片中。通过对比调整前后扇区整体复杂性的变化情况验证关键航空器选取的效果，为识别调配关键航空器从而控制扇区复杂性的研究提供一定的基础。

1 建模及方法

1.1 航空器之间的复杂性关系

现有管制模式下，监控和调配航空器的飞行冲突是引起管制员工作负荷产生的重要原因之一。从航空器行为的微观角度来看，主要是航空器的相对位置和速度变化对管制员产生一定的负荷影响，尤其是正在接近的航空器对。因此，从运行角度而言，空中交通复杂性可以表现为一系列航空器汇聚关系的总和。基于内禀性的复杂性建模方法正是从航空器间的微观角度出发，能够更真实地还原空域内运行状态及管制员工作负荷。本文在文献[4-5]的基础上构建基于内禀性的空中交通复杂性模型。由于航空器尺寸相对飞行间隔标准太小，对空中交通复杂性影响可以忽略不计，从而将航空器视为一个节点。记航空器速度矢量和位置矢量分别为$\boldsymbol{V}$和$\boldsymbol{D}$，则航空器i和航空器j的相对速度和相对位置矢量分别为：$\boldsymbol{V}_{ij}=\boldsymbol{V}_j-\boldsymbol{V}_i$，$\boldsymbol{D}_{ij}=\boldsymbol{P}_j-\boldsymbol{P}_i$。航空器汇聚态势的复杂程度与$\boldsymbol{D}_{ij}$和$\boldsymbol{V}_{ij}$有关，其中，$|\boldsymbol{D}_{ij}|$越小或$|\boldsymbol{V}_{ij}|$越大，反映管制调配汇聚关系的需求越强烈。

当2架航空器呈现汇聚态势时，即$(\boldsymbol{D}_{ij},\boldsymbol{V}_{ij})<0$，定义时间函数$T_{ij}$反映2架航空器汇聚的所需时间，可表示为

$$T_{ij}=\frac{|\boldsymbol{D}_{ij}|}{(\boldsymbol{D}_{ij},\boldsymbol{V}_{ij})/|\boldsymbol{D}_{ij}|} \tag{1}$$

式中：$(\boldsymbol{D}_{ij},\boldsymbol{V}_{ij})/|\boldsymbol{D}_{ij}|$为$\boldsymbol{D}_{ij}$单位矢量上$\boldsymbol{V}_{ij}$的分量。

此外，当$|\boldsymbol{D}_{ij}|$小于规定安全间隔D_{sep}时，调配需求紧迫度呈非线性增长，引起的空域复杂性也将倍增。为体现突破安全间隔后带来复杂性非线性增长的现象，引入调节系数λ来体现复杂性变化，表示为

$$\lambda = \left(D_{\text{sep}}/\left|\boldsymbol{D}_{ij}\right|\right)^2 \tag{2}$$

定义航空器对ij的复杂性为r_{ij}, 表示为

$$r_{ij} = \frac{\lambda}{T_{ij}} = -\frac{D_{\text{sep}}^2 \cdot (\boldsymbol{D}_{ij}, \boldsymbol{V}_{ij})}{\left|\boldsymbol{D}_{ij}\right|^4} \tag{3}$$

当 2 架航空器呈现分离态势时, 即$(\boldsymbol{D}_{ij}, \boldsymbol{V}_{ij}) \geqslant 0$, 定义航空器对的复杂性$r_{ij} = 0$。

则任意 2 架航空器的复杂性可表示为

$$r_{ij} = \begin{cases} -\dfrac{D_{\text{sep}}^2 \cdot (\boldsymbol{D}_{ij}, \boldsymbol{V}_{ij})}{\left|\boldsymbol{D}_{ij}\right|^4} & (\boldsymbol{D}_{ij}, \boldsymbol{V}_{ij}) < 0 \\ 0 & (\boldsymbol{D}_{ij}, \boldsymbol{V}_{ij}) \geqslant 0 \end{cases} \tag{4}$$

对于任意时刻, 扇区内航空器(总航空器数为N)邻接矩阵$\boldsymbol{A}$可表示为

$$\boldsymbol{A} = \begin{bmatrix} r_{11} & \cdots & r_{1j} & \cdots & r_{1N} \\ \vdots & & \vdots & & \vdots \\ r_{i1} & \cdots & r_{ij} & \cdots & r_{iN} \\ \vdots & & \vdots & & \vdots \\ r_{N1} & \cdots & r_{Nj} & \cdots & r_{NN} \end{bmatrix} \tag{5}$$

则该时刻扇区复杂性R_t定义为

$$R_t = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}\right) \Big/ 2 \tag{6}$$

扇区复杂性越高, 表明汇聚关系越强烈, 相应的管制操作难度越大。

1.2　空中交通态势时效网络建模

传统基于汇聚关系的网络中, 关键节点选取都基于静态网络, 未考虑飞机未来一段时间内由于汇聚产生的非线性复杂性增加的飞行状态评估, 从而容易忽视在未来几分钟内可能遇到的突发而又严重的飞行冲突状况。针对以上不足, 本文以 SAM 无权时效网络模型为基础, 在单个时间片内, 以航空器为节点, 存在汇聚关系则连边, 航空器对的复杂性为边权, 其表示层内关系。不同时间片上, 采用时序相关系数表示层间关系, 进一步建立加权的空中交通态势时效网络模型, 从而充分挖掘未来一段时间内的飞行状态。

时效网络是一个包含了节点,节点间相互作用为连边及连边按时间顺序相互作用形成层间关系的系统。节点间连边随时间变化且具有时间先后顺序。对于静态网络$G = (N, E)$, 在一定时间T内, 时效网络可以描述为n个离散有序的静态网络层$G_1, G_2, \cdots, G_n$的组合。

如图 1 所示, 在空中交通态势时效网络中, 将航空器视为节点$N = \{n_i\}$, 时间T内网络的边可以描述为$E = (e_1, e_2, \cdots, e_t)$, 其中, $e_t = \{e_{ij}^t\}$, e_{ij}^t表示航空器节点i与j在t时刻存在汇聚关系。扇区内所有航空器的邻接矩阵$\boldsymbol{A}^{(t)} = [r_{ij}^t]$, 其中, r_{ij}^t表示在t时刻航空器i、航空器j对应的复杂性值。

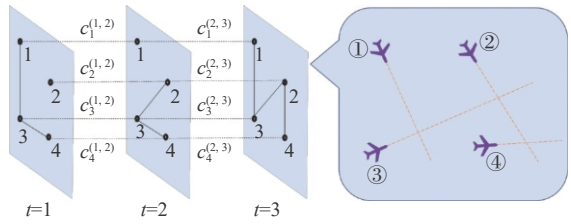


图 1　空中交通态势时效网络示意图

Fig. 1　Temporal network of air traffic

考虑层内关系与层间关系的时效网络建模方法可以更全面的挖掘时效网络的特征。其中, 层内关系对应于各时间切片中网络的邻接关系, 用$\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \cdots, \boldsymbol{A}^{(T)}$表示; 层间关系用时序相关系数$c$表示[28], 表示为

$$c = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T-1} c_i^{(t,t+1)} \tag{7}$$

式中: $c_i^{(t,t+1)}$为节点i的邻居拓扑重叠系数, 表示航空器节点i在时刻t与时刻$t+1$中的层间关系,即节点i在时刻t与时刻$t+1$中共同邻居数量的占比, 具体表示为

$$c_i^{(t,t+1)} = \frac{\sum_j a_{ij}(t)a_{ij}(t+1)}{\sqrt{\left(\sum_j a_{ij}(t)\right)\left(\sum_j a_{ij}(t+1)\right)}} \tag{8}$$

其中, $a_{ij}(t)$表示在t时刻网络中节点i和节点j邻接关系, 若节点i和节点j在t时刻有边相连, 则$a_{ij}(t) = 1$; 否则$a_{ij}(t) = 0$。若节点i和节点j在t和$t+1$时刻均相连, 则$a_{ij}(t)a_{ij}(t+1) = 1$, 否则, $a_{ij}(t)a_{ij}(t+1) = 0$[29]。

邻居拓扑重叠系数刻画了航空器在不同时间片中的度和邻接关系的层间相似性。系数值越大, 表示该航空器连续出现在相邻的时间片中, 且与其他航空器保持稳定的潜在冲突关系。反之, 系数值越小表示航空器未同时出现在相邻的时间片中或相邻时间片中的潜在关系消失。

在一个加权空中交通态势时效网络中, 超邻接矩阵$\boldsymbol{W}$包含所有时间片的节点邻接关系及层间关系, 表示为

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{A}^{(1)} & \boldsymbol{C}^{(1,2)} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{C}^{(1,2)} & \boldsymbol{A}^{(2)} & \boldsymbol{C}^{(2,3)} & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{C}^{(2,3)} & \ddots & \ddots & \boldsymbol{0} \\ \vdots & \ddots & \ddots & \boldsymbol{A}^{(T-1)} & \boldsymbol{C}^{(T-1,T)} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{C}^{(T-1,T)} & \boldsymbol{A}^{(T)} \end{bmatrix} \tag{9}$$

式中: $\boldsymbol{A}^{(T)}$表示$t = T$时航空器对的邻接矩阵; $\boldsymbol{C}^{(T-1,T)}$

表示在$T-1$与T时刻所有航空器节点的层间关系。

1.3 关键航空器识别

空中交通管制员每次调整航空器都会造成当前空中交通态势网络的改变,甚至会对后续一段时间的运行状态造成不同程度的影响。为评估这种航空器造成的“蝴蝶效应”,定义对未来一段时间的空域复杂性影响最大的航空器为关键航空器。实际运行中,可通过对空域内少数关键航空器的调整进而实现空域复杂性的高效控制。

从空中交通管理的实际运行需求角度出发,空中交通态势中关键航空器的识别应该满足如下2个条件:①关键航空器在时效网络相邻时间片内稳定出现,且保持较高的重要性。②关键航空器在网络中的邻居节点的重要性也被考虑在内。因此,本文采用特征向量中心性对空中交通态势关键航空器进行识别。在空中交通态势时效网络中,先对航空器间汇聚关系的复杂性进行时效网络加权,计算特征向量中心性,从而识别在一定时间段内稳定出现、复杂性较高且综合量化邻居节点信息的航空器。

对超邻接矩阵$\boldsymbol{W}$求最大特征值对应的特征向量$\boldsymbol{\alpha}=[\alpha_1,\alpha_2,\cdots,\alpha_{TN}]^{\mathrm{T}}$,记

$$\boldsymbol{Z}=\begin{bmatrix}\alpha_1 & \alpha_{N+1} & \cdots & \alpha_{N(T-1)+1}\\ \vdots & \vdots & & \vdots\\ \alpha_N & \alpha_{2N} & \cdots & \alpha_{TN}\end{bmatrix} \tag{10}$$

式中:$\boldsymbol{Z}$为时效网络的特征值矩阵;N为任意时刻内航空器的总数;T为总的时间片数量。则α_{mn}表示第m架航空器在第n个时间片上的特征向量中心性,则任意航空器m在该时段内的全局重要性I_m定义为

$$I_m=\frac{1}{T}\sum_{i=1}^{T}\alpha_{mn} \tag{11}$$

航空器在空中交通态势时效网络中的全局重要性值越大,航空器越重要。

2 关键航空器验证方法

在实际的空中交通复杂性控制中,研究人员希望能够调配尽量少的航班,且消除较多的汇聚关系,尤其是复杂性值较高的汇聚关系。调整航空器进扇区的时刻是常用的调配方法之一。如某航空器预计t时刻进入扇区,在扇区内共停留Δt时间,调配方案为该航空器推迟1 min进扇区。则此调配方法对应到时效网络中具体体现为2步:①在$[t,t+\Delta t]$时间段的网络切片中删除该航空器节点;②在$[t+1,t+\Delta t+1]$时间段的切片中添加该节点,原始停留时间Δt内所有的预定轨迹信息顺延至$[t+1,t+\Delta t+1]$时间段。对关键航空器验证分为2步:①使用网络攻击在时效网络中删除指定航空器节点,对应管制操作中限制特定航空器在预定时刻进扇区;②进行网络调整,通过改进的遗传算法(genetic algorithm, GA)为删除的航空器节点重新分配进扇区时刻。通过扇区复杂性的变化验证调配关键航空器对扇区复杂性的控制效果。

2.1 网络攻击

在网络攻击中,本文采用随机攻击和蓄意攻击2种方式。蓄意攻击是指将节点按照特定指标进行从大到小排序,并依次攻击指标值大的节点。破坏节点之后,将建立一个新的网络,再重复计算节点重点性并进行新一轮的破坏。而随机攻击中,所有节点被攻击的概率是相等的。本文用随机攻击来模拟随意调配管制扇区内的航空器,蓄意攻击来模拟优先调配关键航空器。为进一步比较时效网络与静态网络选取关键航空器的效果,在蓄意攻击时将选择静态网络中常用的度值、介数、集聚系数指标及本文时效网络中特征向量中心性指标分别进行攻击。通过观察不同攻击方式下扇区复杂性的变化趋势来验证关键航空器对扇区复杂性的影响。具体步骤如下:

步骤1 计算复杂性关系。根据扇区内航空器的汇聚态势计算复杂性关系。

步骤2 构造加权空中交通态势时效网络。其中,以航空器为节点,存在汇聚关系则形成边,边的权值为对应航空器汇聚关系复杂性,从而建立每个时间片上的航空器关系邻接矩阵,以此建立层内关系。计算不同时间片的节点邻居拓扑重叠系数建立层间关系。

步骤3 计算特征向量中心性。根据层内关系及层间关系建立超邻接矩阵$\boldsymbol{W}$,并以此计算特征向量中心性。

步骤4 选择攻击节点。随机选择节点或根据特定指标值排序,选择指标最大值的节点作为攻击节点。

步骤5 网络攻击。删除该节点并建立新的网络,将攻击后的网络重复步骤1~步骤4。

步骤6 结果分析。对比不同攻击方式下扇区复杂性的变化趋势,分析方法的有效性。

2.2 网络调整

网络调整对应于为特定航空器分配一个新的进扇区时刻。在实际空中交通管理工作中,通过配备时间间隔,分配进扇区时刻减少航空器间的潜在冲突从而一定程度上控制扇区复杂性。随着航班量的增加,每次分配都可能仅仅暂时缓解当前的空

中交通复杂态势而造成未来某时刻更复杂的交通态势。在通过调整航空器进扇区时刻控制扇区复杂性时, GA 可以求解一定程度上的满意解。GA 计算时以一定时间段内所有航空器进扇区的允许调整时刻为搜索空间, 通过大量迭代计算不同时间方案下的扇区复杂性寻求满意解。然而直接使用 GA 求解该问题时存在可行解空间大, 计算量大和收敛较慢等不足。为此, 通过预先识别对复杂性影响较大的关键航空器, 优先调整少量关键航空器, 并以此改进 GA 可以使得求解该问题时缩小可行解空间加快收敛。

2.2.1　构造适应度函数

在交通复杂性的控制中, 调配航空器进扇区时刻需要综合考虑扇区复杂性、航空器调整的时间总量、调整进扇区时刻的航空器总数等多个目标。对该多目标问题的求解, 本文采用物理规划的方法将多目标转化为单目标问题。

在物理规划方法中, 首先, 对每个目标设计一个无量纲的偏好函数, 然后, 将各个目标的偏好函数综合以构建多目标的综合偏好函数, 并以此作为物理规划的最终优化目标。本文将简化偏好函数的构造, 即对多目标采用相同的偏好函数以构造适应度函数。在本文求解的多目标问题中, 3 个子目标都是指标值越小越好。该问题对应物理规划中的 Class-1 偏好型问题, Class-1 类设计准则等级函数如图 2 所示。其意义为:

1) 很期望域, $g \in [g_1, g_2]$, 该区间内的目标取值属于很期望的范围;

2) 期望域, $g \in (g_2, g_3]$, 该区间内的目标取值属于期望的范围;

3) 可忍受域, $g \in (g_3, g_4]$, 该区间内的目标取值属于可以忍受的范围;

4) 不期望域, $g \in (g_4, g_5]$, 该区间内的目标取值属于可以忍受的范围;

5) 很不期望域, $g \in (g_5, \infty)$, 该区间内的目标取值属于可以忍受的范围。

本文根据空中交通运行的实际特征, 构造了简易的偏好函数:

$$P(g)=\begin{cases}\lambda(i)+\dfrac{(\lambda(i+1)-\lambda(i))\times(g-g_i)}{g_{i+1}-g_i} & g_i \leqslant g \leqslant g_{i+1}\\ \lambda(i)\times \mathrm{e}^{\frac{g-g_5}{g_5-g_4}} & g \geqslant g_5\end{cases} \tag{12}$$

式中: $P(g)$为偏好函数值, g为目标值, $g_i(i=1,2,\cdots,5)$为偏好区间边界点; $\lambda(i)$为每个边界点的偏好值, 构建如下:

$$\lambda(i)=\sqrt{\mathrm{e}^{i-5}} \qquad i=1,2,\cdots,5 \tag{13}$$

因此, 物理规划的多目标问题转换为单目标问题的总偏好函数P表示为

$$P=\sum_{j=1}^{m} P_j(g) \tag{14}$$

式中: $P_j(g)$为第j个目标的偏好函数值。

在空中交通复杂性控制问题中, 3 个目标分别为扇区复杂性、航空器调整的时间总量、调整进扇区时刻的航空器总数分别最小。

一段时间的扇区总体复杂性R:

$$R=\sum_{t=1}^{T} R_t=\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N} r_{ij}^{(t)}\right)\Big/2 \tag{15}$$

式中: N为任意时刻内航空器的总数。

航空器调整的时间总量S:

$$S=\sum_{i=1}^{N}|s_i|^2 \tag{16}$$

式中: N为任意时刻内航空器的总数; s_i为第i个航空器的进扇区调整时间; s_i为正值表示该航空器延迟$|s_i|$分钟进扇区; s_i为负值表示该航空器提前$|s_i|$分钟进扇区。

调整进扇区时刻的航空器总数F:

$$F=\sum_{i=1}^{N} f_i \tag{17}$$

式中: f_i为第i个航空器调整进扇区时刻标记, 计算如下:

$$f_i=\begin{cases}1 & |s_i|>0\\ 0 & |s_i|=0\end{cases} \tag{18}$$

则, 空中交通复杂性控制问题的适应度函数为

$$\min J=P_1(R)+P_2(S)+P_3(F) \tag{19}$$

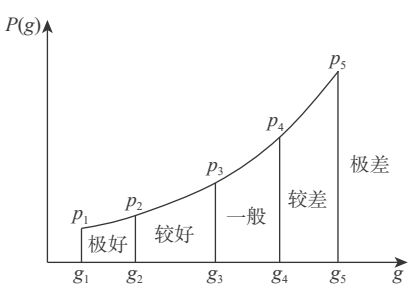


图 2　Class-1 类设计准则等级函数

Fig. 2　Class-1 design criteria level function

2.2.2　改进的遗传算法

使用 GA 求解空中交通复杂性控制问题时, 存在可行解空间较大, 算法的计算量大且收敛性较低的不足。本文通过所确定的关键航空器, 提出一种改进遗传算法(简称为 GA2), 通过优先变异关键航

空器所在的基因即调整高复杂性态势航空器的进扇区时刻,具体步骤如下:

步骤 1 初始化种群。设定种群规模为100,染色体长度为高峰小时内进扇区的航空器总数N,每个航空器进扇区时刻对应于一个基因点位。在初始种群中,每个染色体只调整一个基因点位,设定任意航空器允许调整的进扇区时刻范围为[−3,3],生成初始种群时,随机产生2组在区间[1,N]和[−3,3]的整数D_1和D_2。即种群$M_{(100\times N)}$中第i个染色体D_1基因位置的航空器进扇区时刻调整量$s_i = D_2$,表示为

$$M(i,D_1) = D_2 \tag{20}$$

步骤 2 计算适应度值。通过计算物理规划适应度值进而评价个体。适应度体现了不同个体在计算中达到或接近最优解的优良程度。

步骤 3 选择。为保留上一代中适应度较好的个体,算法采用精英保留策略,将10个适应度值最好的个体直接进入下一代重组中而不用参与交叉变异。

步骤 4 交叉。本文采用单点交叉,将选择进入交叉变异的90个个体随机两两组合,按照交叉概率$Pc = 0.9$决定是否进行交叉操作,如进入交叉操作,则随机生成2个点位进行基因交叉。

步骤 5 变异。关键航空器对扇区复杂性有较大影响,通过优先调配关键航空器可以有效控制扇区复杂性。首先,依照交叉后形成的90个新个体中进扇区的时刻对原始航迹数据更新。然后,将每个个体按1 min为步长建立时效网络,并计算特征向量中心性,以归一化的特征向量中心性值为概率,采用轮盘赌的方式选择变异的基因点位。最后,随机生成一个新的调整时刻替换所选择的基因完成变异。

步骤 6 重组。将精英保留策略的10个个体与交叉变异的90个个体组合成为新的种群。跳转步骤2直至达到迭代次数并输出结果。

3 实例验证

3.1 空中交通态势分析

该高峰小时共有104个架次,即每层网络共有104个节点,时间层共60个。如果在t时刻航空器节点出现在扇区内,则定义该航空器节点为活跃状态;否则,定义该节点为非活跃状态。活跃航空器节点数量及扇区复杂性如图3所示。由图观察到多个不同的时间片内航班数量相同,然而扇区复杂性却差异很大。

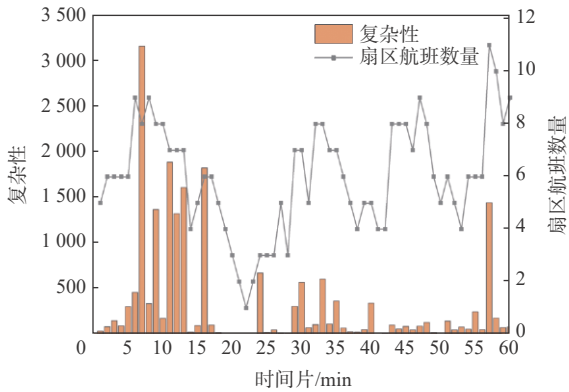


图3 扇区复杂性与航班数量

Fig. 3 Sector complexity and number of flights

为进一步分析深层次原因,本文选取航班数量相同的第7 min和第33 min,对应邻接矩阵如表1和表2所示。

在第7 min,航空器4和航空器8的汇聚关系造成了扇区复杂性的激增。在33 min航空器1和航空器5、航空器1和航空器6、航空器5和航空器6、航空器4和航空器7复杂性值相对其他航空器对偏高。由此看出,在实际运行中,即使扇区内航班数量相同,仍存在一对或多对使扇区呈现不同复杂性态势的关键航空器。而对于管制活动而言,识别并优先调配这些关键航空器能够快速减轻扇区复杂性。

3.2 网络攻击效果分析

针对扇区中存在的关键航空器,时效网络模型

表 1 第 7 min 邻接矩阵

Table 1 Adjacency matrix at 7th minute

航空器	r_{i1}	r_{i2}	r_{i3}	r_{i4}	r_{i5}	r_{i6}	r_{i7}	r_{i8}
1	0	0	8.834	11.14	7.713	25.06	0	2.513
2	0	0	13.74	8.191	11.60	8.162	0	1.545
3	8.834	13.74	0	0	0	6.460	0.414	7.659
4	11.14	8.191	0	0	0	8.792	0	2.878
5	7.713	11.60	0	0	0	5.778	0.197	10.46
6	25.06	8.162	6.460	8.792	5.778	0	142.9	1.440
7	0	0	0.414	0	0.197	142.9	0	0
8	2.513	1.545	7.659	2.878	10.46	1.440	0	0

表 2　第 33 min 邻接矩阵

Table 2　Adjacency matrix at 33rd minute

航空器	r_{i1}	r_{i2}	r_{i3}	r_{i4}	r_{i5}	r_{i6}	r_{i7}	r_{i8}
1	0	4.167	3.349	34.05	100.6	153.1	0.512	0
2	4.167	0	7.662	3.413	0.636	3.776	7.864	45.34
3	3.349	7.662	0	0.529	0	3.332	4.724	12.29
4	34.05	3.413	0.529	0	0	36.84	79.08	0
5	100.6	0.636	0	0	0	90.44	0	0
6	153.1	3.776	3.332	36.84	90.44	0	12.49	0
7	0.512	7.864	4.724	79.08	0	12.49	0	0.254
8	0	45.34	12.29	0	0	0	0.254	0

中在高峰小时内以 1 min 为步长, 某时刻扇区内所有的航空器为节点, 判断该时刻开始未来 5 min 内, 若存在汇聚关系则连边, 以航空器对的复杂性值为边权并建立时效网络, 通过计算特征向量中心性确定关键航空器。随后, 分别进行蓄意和随机网络攻击, 即删除所选取的关键航空器节点, 观察扇区复杂性变化情况。蓄意攻击时, 分别选取时效网络特征向量中心性(即时效网络攻击)和静态网络中以度值、介数、集聚系数为指标确定的关键航空器进行攻击。随机攻击时以相等概率任意选择航空器进行攻击。2 种方式的过程如图 4 所示。

由图 4 可知, 初始状态下, 扇区复杂性最高的时间片出现在第 7 min。其次, 第 9～16 min 及第 57 min 扇区整体处于复杂性较高的状态, 第 30～40 min 内复杂性处于中等水平。随机攻击下, 复杂性最高的第 7 min 及第 9～16 min 在攻击结束后复杂性值仍未下降, 第 57 min 的复杂性值在 16 次攻击后显著下降, 但随机攻击整体效果不佳。静态网络指标中, 度值攻击与介数攻击结果较为相近, 第 5～20 min 内的复杂性均有一定程度的下降, 集聚系数攻击效

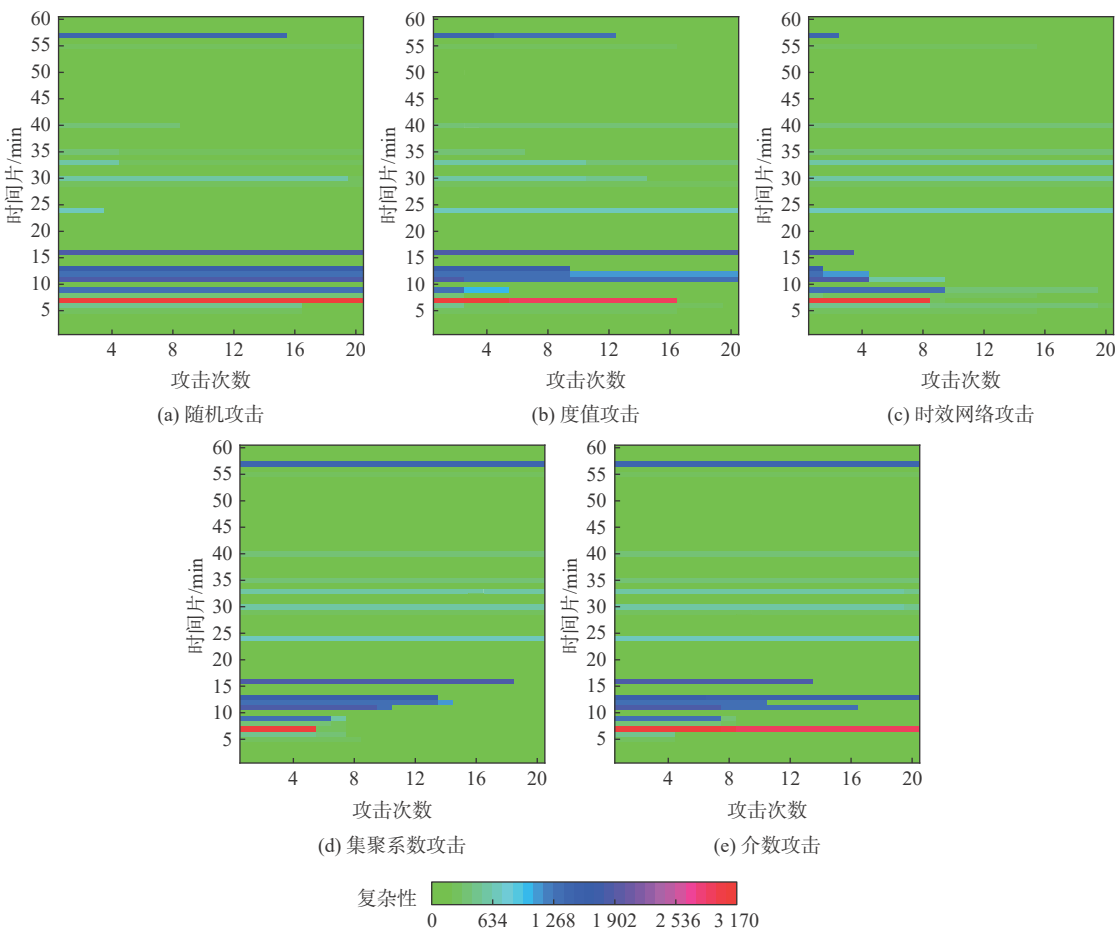


图 4　网络攻击过程

Fig. 4　Process of network attack

果相比静态网络中其他指标对复杂性的控制效果最优。在时效网络中，高峰小时内各时间段的高复杂性时刻均有不同程度的降低。对比时效网络指标与静态网络指标效果最优的集聚系数指标，第5～20 min内的高复杂性态势分别经过10次和19次网络攻击后扇区降为低复杂性态势，结果显示时效网络指标对空中交通复杂性的控制效率更高，且不同时间片上的高复杂性态势均能较高效率的识别与控制。

从扇区角度而言，图5展示了扇区复杂性随攻击的变化趋势。整体来看，攻击过程中，扇区复杂性从高到低分别为随机攻击、介数攻击、度值攻击、集聚系数攻击及时效网络攻击。其中，时效网络攻击后扇区复杂性收敛速率显著。

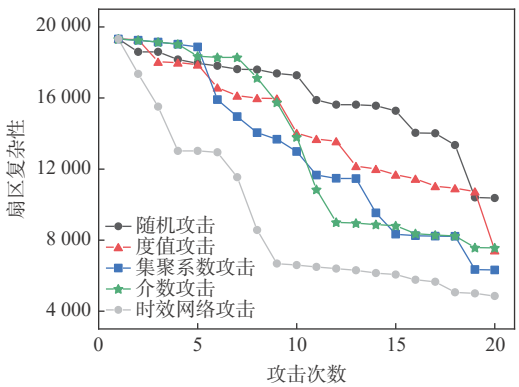


图5 扇区复杂性

Fig. 5 Sector complexity

3.3 网络调整结果

3.3.1 目标函数设置

1）复杂性的偏好区间设置。根据不同时间片复杂性分布情况的历史数据，选取扇区复杂性均值的25%、50%、75%和均值设定复杂性的偏好区间边界点{0,25%×322,50%×322,75%×322,322}，将复杂性的区间偏好点代入偏好函数，得出复杂性的偏好函数，如图6所示。

2）航空器调整时间的偏好区间设置。根据航空器性能要求等设定允许调整的时间范围为[−3,3]，从空管的运行效率而言，通常应当尽量保持其原定航迹运行以加速交通流的移动。因此，设定调整的时间绝对值偏好边界点为{0,1,1.75,2.5,3}。将调整的时间的区间偏好点代入偏好函数，得出复杂性的偏好函数，如图7所示。对所有航空器调整时间的偏好值取均值作为调整时间总量的适应度目标值。

3）调整航空器比例的偏好区间设置。设定调整航空器比例的偏好边界点为{0,0.25,0.5,0.75,1}。将调整航空器比例的区间偏好点代入偏好函数，得出复杂性的偏好函数，如图8所示。

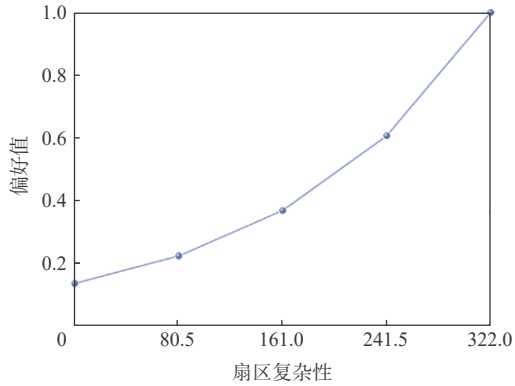


图6 扇区复杂性偏好函数曲线

Fig. 6 Preference function curve of sector complexity

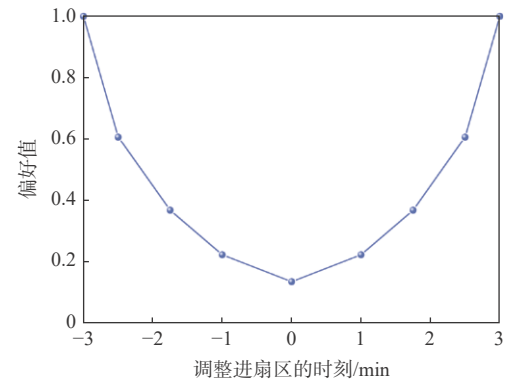


图7 航空器调整时间的偏好函数曲线

Fig. 7 Preference function curve of aircraft adjustment time

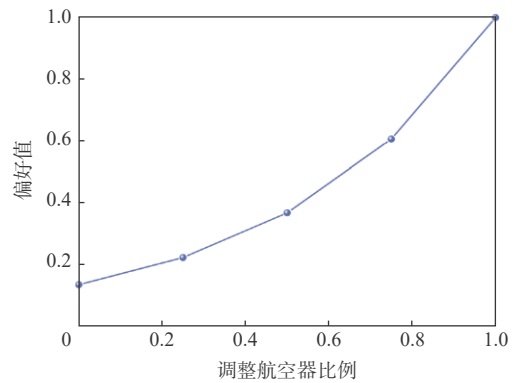


图8 调整航空器比例偏好函数曲线

Fig. 8 Preference function curve of aircraft proportion adjustemnt

3.3.2 遗传算法结果对比

为分析GA2的求解效果，本文同时采用GA直接求解复杂性控制问题，并对比两者的结果。图9为GA2和GA的计算结果。GA2在迭代20次时适应度已经达到0.668 11，而GA在迭代100次的适应度仅为0.891 1。因此，GA2计算结果明显优于GA。从迭代过程而言，由于GA2调整了对复杂性有较长时间影响的关键航空器，使得GA2的收敛性明显优于GA，这对于求解复杂性控制问题的实际应用有重要意义。

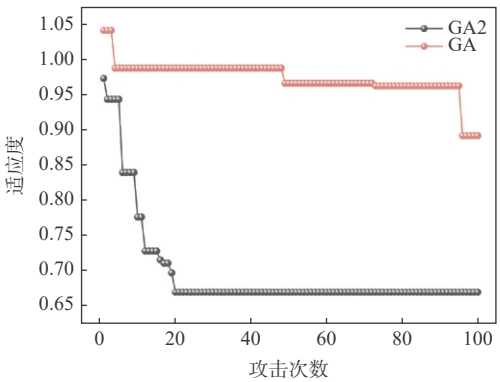


图 9　GA 和 GA2 的计算结果

Fig. 9　The calculation results of GA and GA2

表 3 为 GA2 求解的调整时刻方案，在该高峰小时内共有 84 架航空器，方案共调整了 9 架航空器，包含 7 架提前和 2 架延误，在实际运行中该方案属于可接受的操作方案。

表 3　GA2 调整方案

Table 3　Adjusted scheme of GA2

航空器	原始进扇区时刻/min	时刻调整量/min
CQN2037	3	−1
CES5809	6	3
CES565	8	−1
CQH8954	10	−1
CSZ9880	15	−3
CXA8415	16	−1
CGH1003	29	2
CDG4871	37	−3
CGH1003	51	−2

3.4　网络调整效果分析

为分析时效网络和静态网络选取关键航空器的效果差异，本文对网络调整后的扇区复杂性、扇区汇聚数量、航空器平均复杂性和航空器平均汇聚数量进行了统计分析。

3.4.1　扇区复杂性

经过网络调整之后，扇区的总体复杂性降为原来的 44.82%。图 10 为原始扇区复杂性及 GA 优化结果对比，可以观察到，高峰小时内前 20 min 的原始高复杂性态势经过调整 6 架航空器的时刻得到大幅缓解。调整后的方案复杂性整体更为平缓，避免了较多的高复杂性态势，即减少了管制人员应对高复杂性态势的高压力及高风险。较平缓的空中交通态势也提高了管制的运行效率。

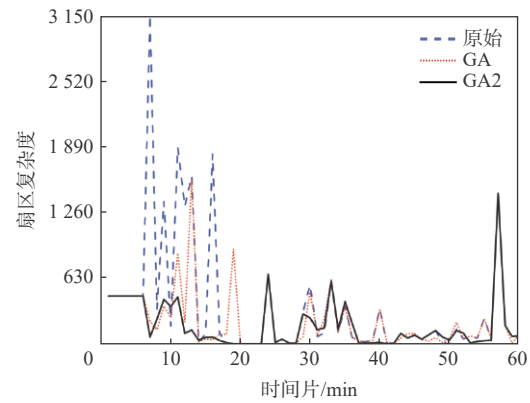


图 10　扇区复杂性对比

Fig. 10　Sector complexity comparison

3.4.2　扇区汇聚数量

t 时刻扇区内航空器汇聚数量 K_t 为该时刻扇区内所有汇聚关系的总和，表示为

$$K_t = \sum_i \sum_j e_{ij}^t \tag{21}$$

式中：e_{ij}^t 为航空器 i 与航空器 j 在 t 时刻存在汇聚关系。

汇聚关系数量增加，表示管制员需要处理的工作量越大。从空中交通管理工作战略角度而言，控制不同时间片内的汇聚数量可以降低扇区整体复杂性，使管制工作量在一定时间内处于稳定状态。

图 11 为不同时间片中的汇聚关系数量及扇区复杂性。初始状态中，不同时间片上汇聚关系总数接近于正态分布。随机攻击后，汇聚关系总数有所降低，但扇区复杂性较高，效果不理想。度值攻击方式下，汇聚关系总数为各攻击方式中最低，且数量稳定。集聚系数攻击、介数攻击和时效网络攻击后，汇聚关系总数呈现多元化，中位线较低，但时效网络中 25%～75% 的汇聚关系更加集中且稳定。由此可见，仅通过降低冲突数量的方法来控制扇区复杂性存在一定的不足，汇聚关系呈多元化的扇区中存在某些对复杂性影响较大的航空器，时效网络对此类关键航空器识别更加准确。

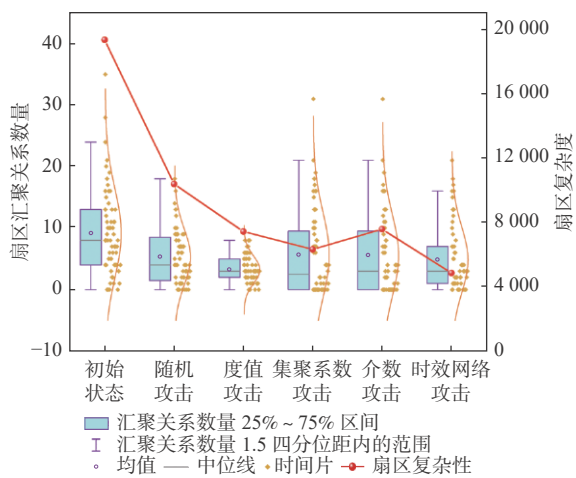


图 11　扇区内汇聚关系数量

Fig. 11　Number of convergence relationships within a sector

3.4.3　航空器平均复杂性

任意航空器 i 在扇区内的平均复杂性 r_i 表示为

$$r_i = \frac{1}{\Delta t} \sum_{t=t_i}^{t_i+\Delta t} \sum_{j=1}^{N} r_{ij}^t \tag{22}$$

式中：Δt为航空器i在扇区内的持续时间；r_{ij}^{t}为t时刻航空器i与航空器j的复杂性。

图12为各攻击方式下航空器平均复杂性分布情况。整体而言，时效网络攻击下航空器复杂性的控制效果优于静态网络，静态网络优于随机攻击。其中，静态网络指标中，集聚系数攻击优于度值攻击，度值攻击优于介数攻击。时效网络攻击后，复杂性高于750的航空器皆被选为关键航空器，消除了更多的高复杂性态势的航空器，且航空器复杂性的均值和标准差最小，说明时效网络选取的关键器能更好的识别持续时间段内的高复杂性态势航空器，对关键航空器的调整使扇区一定时间段内的整体复杂性更为平稳。

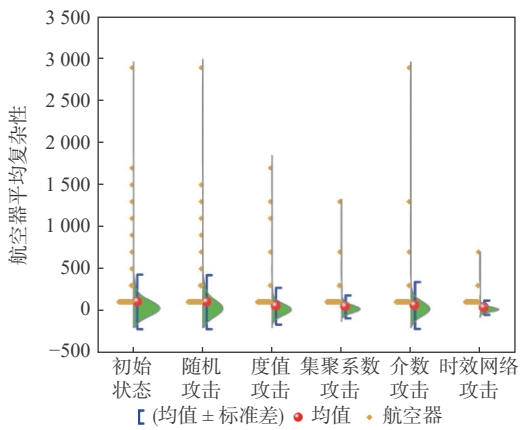


图12 航空器平均复杂性

Fig. 12 Average complexity of aircraft

3.4.4 航空器平均汇聚数量

t时刻扇区内航空器平均潜在汇聚数量k_t为该时刻扇区内所有潜在汇聚数量与航空器总数的比值，表示为

$$k_t = K_t/N \tag{23}$$

同时刻航空器平均汇聚数量越多，表明管制员完成间隔配备或冲突解脱等行为的难度越大。图13为各时间片内航空器平均汇聚数量。度值攻击后航空器平均汇聚数量最低，数据相较其他方式更为集中，表明管制难度较低且相对稳定。集聚系数攻击、介数攻击消除了更多的汇聚关系，但存在航空器平均汇聚数量较大的时间片。时效网络攻击效果介于度值攻击与集聚系数攻击、介数攻击之间，平均汇聚数量相对集中，且消除了一定量的高汇聚关系的航空器。此外，介数攻击后平均汇聚关系的中位数最低，且消除了大量高汇聚关系的航空器，这对研究降低管制难度具有重要意义。

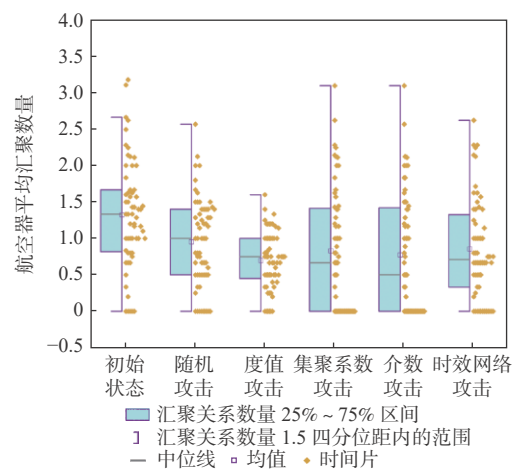


图13 航空器平均汇聚数量

Fig. 13 Average convergence number of aircraft

4 结 论

1）采用网络攻击模拟实际操作中限制航空器按照原预定时刻进入扇区，结果显示对时效网络所确定的关键航空器进行控制相较于静态网络能更高效地降低扇区复杂性。

2）基于关键航空器优先变异改进传统GA的结果表明，该算法对基于调整航空器进入时刻的扇区复杂性控制问题有更快的收敛速度。

3）网络调整效果分析表明，静态网络指标确定的关键航空器无法准确预测持续时间段内对复杂性影响较大的航空器，而时效网络模型通过捕捉航空器未来的运行状态可以更准确地识别持续时间段内复杂性较高的关键航空器。

基于内禀性建立网络的加权关系，关键航空器选取效果可能受不同复杂性评估角度影响，应进一步研究其他复杂性评估角度下的选取效果。

参考文献（References）

[1] SCHMIDT D K. On modeling ATC work load and sector capacity[J]. Journal of Aircraft, 1976, 13(7): 531-537.

[2] HURST M W, ROSE R M. Objective job difficulty, behavioural response, and sector characteristics in air route traffic control centres[J]. Ergonomics, 1978, 21(9): 697-708.

[3] NETJASOV F, JANIĆ M, TOŠIĆ V. Developing a generic metric of terminal airspace traffic complexity[J]. Transportmetrica, 2011, 7(5): 369-394.

[4] KOPARDEKAR P, MAGYARITS S. Dynamic density: measuring and predicting sector complexity[C]//Proceedings of the 21st Digital Avionics Systems Conference. Piscataway: IEEE Press, 2002: 2C4.

[5] LEE K, FERON E, PRITCHETT A. Describing airspace complexity: airspace response to disturbances[J]. Journal of Guidance, Control, and Dynamics, 2009, 32(1): 210-222.

[6] LEE K, FERON E, PRITCHETT A. Air traffic complexity: an input-output approach[C]//Proceedings of the American Control Conference. Piscataway: IEEE Press, 2007: 474-479.

[7] DELAHAYE D, ALLIOT J M, SCHOENAUER M, et al. Genetic algorithms for air traffic assignment[C]//Proceedings of the 11th

European Conference on Artificial Intelligence. New York: ACM, 1994: 33-37.

[8] DELAHAYE D, PUECHMOREL S. Air traffic complexity: towards an intrinsic metrics[C]//Proceedings of the 3rd USA/Europe Air Traffic Management R and D Seminar. Reston:AIAA,2000.

[9] DELAHAYE D, PUECHMOREL S, HANSMAN J, et al. Air traffic complexity map based on non linear dynamical systems[J]. Air Traffic Control Quarterly, 2004, 12(4): 367-388.

[10] 张进, 胡明华, 张晨, 等. 空域复杂性建模[J]. 南京航空航天大学学报, 2010, 42(4): 454-460.
ZHANG J, HU M H, ZHANG C, et al. Airspace complexity modeling[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2010, 42(4): 454-460(in Chinese).

[11] 张进, 胡明华, 张晨. 空中交通管理中的复杂性研究[J]. 航空学报, 2009, 30(11): 2132-2142.
ZHANG J, HU M H, ZHANG C. Complexity research in air traffic management[J]. Acta Aeronautica et Astronautica Sinica, 2009, 30(11): 2132-2142(in Chinese).

[12] 徐肖豪, 黄宝军, 舒勤. 基于内禀属性的扇区复杂性评估[J]. 中国民航大学学报, 2013, 31(2): 22-28.
XU X H, HUANG B J, SHU Q. Sector complexity evaluation based on intrinsic attributes[J]. Journal of Civil Aviation University of China, 2013, 31(2): 22-28(in Chinese).

[13] 王红勇, 郭宇鹏. 基于航空器自主运行的空中交通复杂性建模[J]. 交通运输系统工程与信息, 2022, 22(2): 305-312.
WANG H Y, GUO Y P. Air traffic complexity model based on aircraft self-separation operation[J]. Journal of Transportation Systems Engineering and Information Technology, 2022, 22(2): 305-312 (in Chinese).

[14] MORONE F, MAKSE H A. Influence maximization in complex networks through optimal percolation[J]. Nature, 2015, 524(7563): 65-68.

[15] 陈诗, 任卓明, 刘闯, 等. 时序网络中关键节点的识别方法研究进展[J]. 电子科技大学学报, 2020, 49(2): 291-314.
CHEN S, REN Z M, LIU C, et al. Identification methods of vital nodes on temporal networks[J]. Journal of University of Electronic Science and Technology of China, 2020, 49(2): 291-314(in Chinese).

[16] WANG Z Q, PEI X B, WANG Y B, et al. Ranking the key nodes with temporal degree deviation centrality on complex networks[C]// Proceedings of the 29th Chinese Control and Decision Conference. Piscataway: IEEE Press, 2017: 1484-1489.

[17] SCHOLTES I, WIDER N, GARAS A. Higher-order aggregate networks in the analysis of temporal networks: path structures and centralities[J]. The European Physical Journal B, 2016, 89(3): 61.

[18] ABBAS K, SHANG M S, ABBASI A, et al. Popularity and novelty dynamics in evolving networks[J]. Scientific Reports, 2018, 8(1): 6332.

[19] TAYLOR D, MYERS S A, CLAUSET A, et al. Eigenvector-based centrality measures for temporal networks[J]. Multiscale Modeling & Simulation, 2017, 15(1): 537-574.

[20] CARLETTI T, BATTISTON F, CENCETTI G, et al. Random walks on hypergraphs[J]. Physical Review E, 2020, 101(2): 022308.

[21] ROZENSHTEIN P, GIONIS A. Temporal PageRank[C]//Proceedings of the Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2016: 674-689.

[22] HUANG D W, YU Z G. Dynamic-Sensitive centrality of nodes in temporal networks[J]. Scientific Reports, 2017, 7: 41454.

[23] CAI Q, ALAM S, DUONG V N. A spatial–temporal network perspective for the propagation dynamics of air traffic delays[J]. Engineering, 2021, 7(4): 452-464.

[24] 黄洋, 汤俊, 老松杨. 基于复杂网络的无人机飞行冲突解脱算法[J]. 航空学报, 2018, 39(12): 322222.
HUANG Y, TANG J, LAO S Y. UAV flight conflict resolution algorithm based on complex network[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(12): 322222(in Chinese).

[25] 刘飞, 余敏建, 李佳威, 等. 基于复杂网络理论的飞行冲突关键点识别[J]. 空军工程大学学报 (自然科学版), 2019, 20(4): 19-25.
LIU F, YU M J, LI J W, et al. An identification of flight conflict key nodes based on complex network theory[J]. Journal of Air Force Engineering University (Natural Science Edition), 2019, 20(4): 19-25(in Chinese).

[26] 吴明功, 王泽坤, 甘旭升, 等. 基于复杂网络理论的关键飞行冲突点识别[J]. 西北工业大学学报, 2020, 38(2): 279-287.
WU M G, WANG Z K, GAN X S, et al. Identification of key flight conflict nodes based on complex network theory[J]. Journal of Northwestern Polytechnical University, 2020, 38(2): 279-287(in Chinese).

[27] 王兴隆, 苗尚飞. 空域扇区网络结构特性分析及韧性评估[J]. 北京航空航天大学学报, 2021, 47(5): 904-911.
WANG X L, MIAO S F. Structural characteristics analysis and resilience assessment of airspace sector network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(5): 904-911 (in Chinese).

[28] TANG J, SCELLATO S, MUSOLESI M, et al. Small-world behavior in time-varying graphs[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2010, 81: 055101.

[29] 杨剑楠, 刘建国, 郭强. 基于层间相似性的时序网络节点重要性研究[J]. 物理学报, 2018, 67(4): 279-286.
YANG J N, LIU J G, GUO Q. Node importance idenfication for temporal network based on inter-layer similarity[J]. Acta Physica Sinica, 2018, 67(4): 279-286(in Chinese).

Critical aircraft identification method based on temporal network

WANG Hongyong*, MA Lishu, XU Ping

(College of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China)

Abstract: In view of the problem of critical aircraft identification in air traffic situations, the existing research fails to fully consider the spatiotemporal effect in actual air traffic operation. Therefore, a method of critical aircraft identification based on a temporal network was proposed. Based on the convergence relationship between aircraft and its complexity, the temporal network model was constructed by the neighbor topological overlap coefficient, and the critical aircraft was determined based on the eigenvector centrality. Network attacks on critical aircraft nodes were carried out to observe the changes in sector complexity and compared with attacks based on static network indicators. The improved genetic algorithm was used to assign a new sector entry time to the aircraft node deleted by the network attack, so as to verify the selection effect of the critical aircraft. Actual data verification shows that compared with static network attacks, the proposed method can reduce the average sector complexity more efficiently when removing critical aircraft, and the improved genetic algorithm has higher convergence when solving the time allocation problem of critical aircraft entering the sector, making the sector complexity more stable in a certain period of time. The analysis of the control effect of critical aircraft shows that the temporal network method is more accurate than the static network in identifying the aircraft that has a greater influence on the sector complexity in a period of time.

Keywords: air transportation; sector complexity; critical aircraft; air traffic situation timeliness network; network attack; genetic algorithm

Received: 2023-05-22; **Accepted**: 2023-09-09; **Published Online**: 2023-10-12 11:10

URL: link.cnki.net/urlid/11.2625.V.20231010.1553.003

Foundation items: National Natural Science Foundation of China (U1833103); National Natural Science Foundation of Tianjin (21JCZDJC00840)

* **Corresponding author**. E-mail: hy_wang@cauc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0268

框架式动量交换装置角动量包络分析方法

王卫杰[1,*]，汪洲[1]，庞维坤[2]，杨洋[3]

(1. 航天工程大学 宇航与科学技术系，北京 101416；2. 航天工程大学 研究生院，北京 101416；
3. 河南大学 人工智能学院，郑州 475001)

摘 要：针对单框架控制力矩陀螺(SGCMG)、双框架动量轮(DGMW)、磁悬浮控制敏感陀螺(MSCSG)等典型框架式动量交换装置构型角动量包络分析效率问题，提出一种基于框架等效的框架动量轮角动量包络分析方法。比较基于机械实体框架的SGCMG和DGMW与基于磁悬浮微框架的MSCSG角动量交换原理的异同，研究设计SGCMG与MSCSG角动量径/轴向分量间的等效系数，利用双正交SGCMG实体框架构建MSCSG微框架等效模型，对比MSCSG与DGMW遍历法角动量表达式，分析所提方法在DGMW角动量分析的适用性。量化分析所提方法的误差，证明所提方法计算结果相较于传统遍历法角动量绝对误差和角动量方向误差均在千分之一以下。分析2种方法参数维度，证明所提方法的快速性。分别基于传统遍历法和所提方法对MSCSG和DGMW角动量包络进行仿真对比，证明了所提方法的有效性。所提方法在MSCSG群角动量包络分析中具有更广泛的应用。

关 键 词：单框架控制力矩陀螺；双框架动量轮；磁悬浮控制敏感陀螺；角动量包络；误差分析

中图分类号：V448.22

文献标志码：A **文章编号**：1001-5965(2025)05-1591-08

航天器三轴姿态控制系统大多采用动量交换技术。经过长时间发展，已经拥有了控制力矩陀螺、偏置动量轮、框架动量轮与反作用飞轮等一系列动量交换装置。由于控制力矩陀螺由框架驱动改变转子角动量方向，旋转角度不存在限制。卫星大多采用控制力矩陀螺实现大角度快速机动[1]。但控制力矩陀螺结构复杂，框架控制精度受限，输出力矩精度有待提高。动量轮转子虽然偏转角度受限，但是其控制精度高，在高精度高稳定度姿态控制任务中有广泛应用。

控制力矩陀螺仅能改变转子角动量方向，不能改变角动量大小。而双框架动量轮(double-gimbal momentum wheel，DGMW)可以同时改变转子角动量大小和方向。一个DGMW系统可以实现飞行器的三轴稳定。磁悬浮惯性执行机构与传统滚珠轴承支承[2]的动量轮相比具有无摩擦、高精度、寿命长[3]的优点。国防科技大学研制了一种磁悬浮微框架动量轮[4]，可实现航天器三轴姿态主动控制。北京航空航天大学研制出可输出大力矩的单框架及双框架磁悬浮控制力矩陀螺[5]，设计提出磁悬浮陀螺飞轮方案[6]。磁悬浮控制敏感陀螺(magnetically suspended control and sensing gyroscope, MSCSG)是一种新型的磁悬浮惯性执行机构[7-8]，同时具备力矩输出及姿态敏感功能[9]。以MSCSG为代表的磁悬浮惯性动量轮，通过磁悬浮微框架效应驱动转子万向偏转[10-11]，利用高速电机控制转子转速变化，从而输出三自由度高精度、高带宽控制力矩，与DGMW相比具有相似的力矩输出形式。角动量交换装置

收稿日期：2023-05-23；录用日期：2023-05-26；网络出版时间：2023-09-21 10:27
网络出版地址：link.cnki.net/urlid/11.2625.V.20230920.1623.001
*通信作者. E-mail：wangwjie@126.com

引用格式：王卫杰，汪洲，庞维坤，等. 框架式动量交换装置角动量包络分析方法[J]. 北京航空航天大学学报，2025，51(5)：1591-1598.
WANG W J，WANG Z，PANG W K，et al. Angular momentum envelope analysis method of gimbal-type momentum exchange device[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1591-1598 (in Chinese).

力矩输出能力直接决定了其空间应用场景。因而对角动量交换装置角动量包络进行分析是明确其力矩输出能力的重要一环。

国内外学者[12-17]对实体框架动量交换装置的角动量包络分析方法开展了诸多研究。文献[12]利用奇异可视化的方法分析了单框架控制力矩陀螺(single-gimbal control moment gyroscope, SGCMG)群角动量包络。文献[13]在分析SGCMG[15]群角动量包络的基础上,明确了其奇点分布。文献[14]对SGCMG的奇异问题进行了研究,给出了奇异角动量值的切片图,提供了角动量体内奇异的认识方法。文献[15]详细分析了DGCMG[18-19]的奇异问题,仿真得出了奇异条件下DGCMG的角动量包络,明确了三正交构型下DGCMG角动量奇点分布。但角动量包络分析方法主要集中在框架式控制力矩陀螺。SGCMG结构简单,只能在一个自由度上进行旋转,角动量包络分析方法属于几何运算。文献[16]结合金字塔构型[20]在飞轮调速约束下对变速控制力矩陀螺[21](variable speed control moment gyroscopes, VSCMG)角动量包络特性进行了分析,推导了飞轮调速约束下VSCMG角动量表达式,详细描述了VSCMG的角动量包络,VSCMG可视为传统机械框架动量轮的拓展,将其框架旋转角度限定在一定范围内后,可视为框架动量轮工作模式,因而其角动量分析方法同样适用于框架动量轮。文献[22]研究了冗余轮系航天器角动量包络分析及角动量管理问题。文献[23]给出了DGCMG陀螺转子角动量为标量值时的角动量方程。DGCMG角动量包络分析方法主要采用遍历的方法,基于遍历法对MSCSG构型角动量包络进行分析时会出现参数过多计算量大的问题。上述研究集中在实体框架动量交换装置的角动量包络分析,对于磁悬浮微框架动量轮角动量包络分布特征目前尚未见系统性的分析。MSCSG作为磁悬浮微框架动量轮的代表,通过遍历角度实现角动量包络的分析是常规思路,但对于陀螺构型角动量包络,遍历法存在参数维度急剧增加、分析难度增大的问题。

针对上述问题,本文提出一种基于框架等效的框架动量轮角动量包络分析方法。以SGCMG、DGMW和MSCSG为研究对象,在分析SGCMG、DGMW和MSCSG角动量交换原理的基础上,明确SGCMG与MSCSG、DGMW角动量交换原理之间的区别与联系。用双正交SGCMG对单个MSCSG进行等效,对比MSCSG与DGMW角动量表达式,说明了本文方法在DGMW角动量包络分析中的适用性,推导了该方法下MSCSG角动量包络与DGMW角动量包络的数学模型。同时针对本文方法数学模型相对传统遍历法存在的误差,从方向和大小两方面进行误差分析,证明了该方法在框架动量轮角动量包络与DGCMG局部角动量包络分析中的有效性。

1 框架式动量交换装置角动量交换原理分析

SGCMG的结构与工作原理较为简单,转子驱动电机带动转子进行恒定转速的高速旋转,框架电机驱动框架旋转,改变转子角动量方向,输出力矩。SGCMG角动量交换原理如图1所示,图中,$\boldsymbol{H}_0$为转子初始角动量,δ为转子最大偏转角度,$\boldsymbol{g}$为y轴的单位向量。

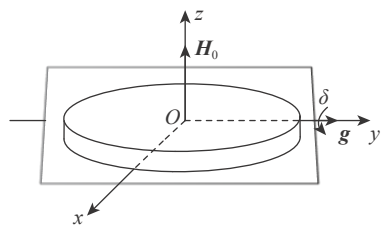


图1 SGCMG角动量交换原理

Fig. 1 Angular momentum exchange principle of SGCMG

DGMW在径向相较于SGCMG增加了一个自由度,内外框架相互垂直,可以分别控制其框架角,输出两自由度控制力矩。DGMW角动量交换原理如图2所示,图中,$\boldsymbol{H}$为角动量,θ_j为外框架角,θ_g为内框架角。

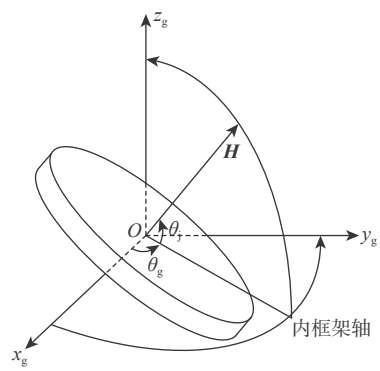


图2 DGMW角动量交换原理

Fig. 2 Angular momentum exchange principle of DGMW

相对于DGCMG,框架动量轮在工作机理与机械结构上有较大的差异,转子偏转角度受限。MSCSG则是利用转子微框架效应输出两自由度陀螺力矩。如图3所示,MSCSG主要由转子组件及定子组件组成。定子组件包括上陀螺房、中陀螺房及下陀螺房。上陀螺房安装有轴向磁轴承上定子、轴向位移传感器、电机定子组件;中陀螺房安装有径向

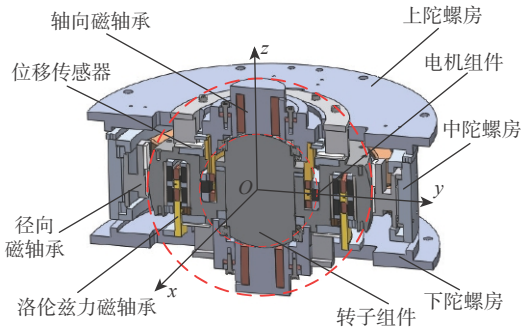


图 3　MSCSG 机械结构示意图

Fig. 3　Mechanical structure of MSCSG

磁轴承定子组件、径向位移传感器; 下陀螺房安装有轴向磁轴承下定子和洛伦兹力磁轴承(Lorentz force magnetic bearing, LFMB)定子组件。在实际应用过程中, 陀螺需要经历起浮、旋转及偏转 3 个阶段。

MSCSG 径向磁轴承和轴向磁轴承结构如图 4 所示。径向磁轴承控制转子径向两自由度 (x、y 方向) 平动, 轴向磁轴承控制转子轴向(z 方向)平动。x、y 方向平动由径向磁轴承力相互垂直的两对磁极进行控制, 两通道不存在耦合。轴向磁轴承与径向磁轴承相互配合, 共同控制转子三自由度平动, 实现转子稳定悬浮。

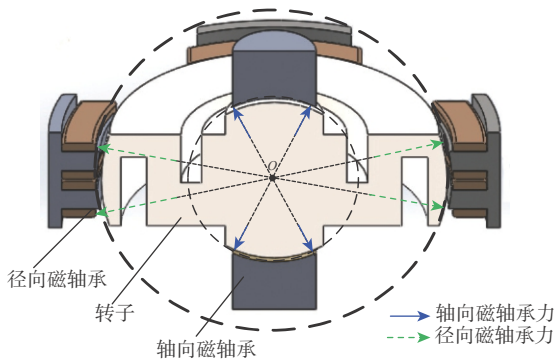


图 4　MSCSG 径向磁轴承和轴向磁轴承结构

Fig. 4　Radial and axial magnetic bearing structures of MSCSG

转子稳定悬浮后, 电机驱动转子高速旋转。在轴向方向, MSCSG 采用偏置动量轮工作模式, 通过改变轴向(z 方向)角动量大小, 输出轴向飞轮力矩。

在径向方向, MSCSG 采用微框架工作模式。转子偏转由 LFMB 控制。LFMB 结构如图 5 所示。LFMB 主要由内磁钢、外磁钢、导磁环、线圈及定子骨架组成。4 组线圈位于定子骨架上, 成对串联使用。磁场方向垂直与线圈方向, 以产生轴向安培力。当 4 个线圈绕组匝数相同时, 产生的安培力大小相等, 方向相反形成力矩, 控制转子的径向 (x、y 方向) 两自由度的偏转, 输出两自由度径向陀螺力矩。

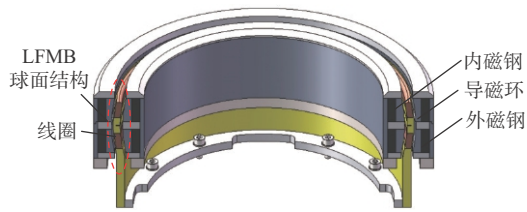


图 5　LFMB 结构示意图

Fig. 5　Structure of LFMB

相较于 SGCMG, MSCSG 通过 2 对洛伦兹力磁轴承实现两自由度径向偏转并且可以改变转子转速。在径向 MSCSG 与 SGCMG 都是通过改变转子角动量方向实现且在偏转过程中二者旋转轴均不发生偏转, 说明二者在角动量交换原理上存在一定的共性。

2　框架等效法角动量包络建模与误差分析

2.1　基于框架等效法 MSCSG 角动量包络建模

如图 6 所示, 定义定子坐标系 $Ox_fy_fz_f$: 与陀螺房固连, 定子坐标系原点与陀螺房质心重合, Ox_f指向虚拟框架旋转方向, Oz_f为虚拟框架处在零位置时转子旋转方向, Oy_f符合右手螺旋定则, 图中, Ω为转子转速, α、β为转子偏转角度。

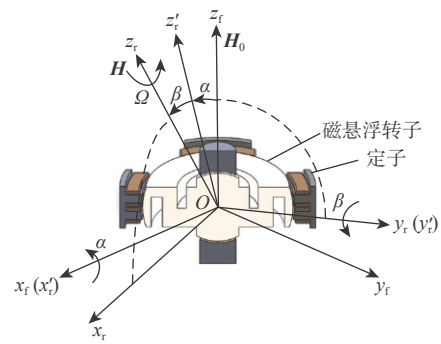


图 6　MSCSG 角动量交换坐标系相对运动关系示意图

Fig. 6　Relative motion relationship of angular momentum exchange coordinate system of MSCSG

定义转子坐标系 $Ox_ry_iz_r$: 与磁悬浮转子固连, 但不随转子绕 Oz_r高速旋转。转子初始静浮未偏转时, 转子坐标系与定子坐标系重合。

磁悬浮转子角动量可以表示为

$$\boldsymbol{H}_r = \boldsymbol{C}_f^r \boldsymbol{H}_0 \tag{1}$$

式中: $\boldsymbol{H}_0 = \begin{bmatrix} 0 & 0 & h_0 \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & J_z\Omega_0 \end{bmatrix}^T$; J_z为转子轴向转动惯量; Ω_0为转子初始转速。

由于虚拟框架轴不随偏转角改变, 所以磁轴承安装坐标系到转子坐标系的转换矩阵 $\boldsymbol{C}_f^r$可以表

示为

$$\boldsymbol{C}_{\mathrm{f}}^{\mathrm{r}}=\begin{bmatrix}1&0&0\\0&\cos\alpha&\sin\alpha\\0&-\sin\alpha&\cos\alpha\end{bmatrix}\begin{bmatrix}\cos\beta&0&-\sin\beta\\0&1&0\\\sin\beta&0&\cos\beta\end{bmatrix}=$$

$$\begin{bmatrix}\cos\beta&0&-\sin\beta\\-\sin\alpha\sin\beta&\cos\alpha&\sin\alpha\cos\beta\\-\cos\alpha\sin\beta&-\sin\alpha&\cos\alpha\cos\beta\end{bmatrix}\tag{2}$$

由于转速变化仅能影响角动量大小，不影响角动量方向。将式 (2) 代入式 (1) 可得 MSCSG 转子角动量为

$$\boldsymbol{H}_{\mathrm{r}}=J_z\Omega\begin{bmatrix}-\sin\beta\\\sin\alpha\cos\beta\\\cos\alpha\cos\beta\end{bmatrix}\tag{3}$$

式 (3) 为 MSCSG 准确角动量表达式。式 (3) 通过转子偏转角度对角动量包络进行建模。当进行构型角动量分析时，随着惯性执行机构数量的增加，参数维度将迅速增大。SGCMG 角动量包络分析则是将转子旋转角度在同一坐标轴系中表示。为减小参数维度，用双正交 SGCMG 等效 MSCSG。MSCSG 两径向角动量由相应 SGCMG 转子旋转后在径向的分量等效，轴向角动量由两 SGCMG 轴向分量标量和的二分之一等效。径向角动量等效仅使用双正交 SGCMG 中的一个陀螺径向分量等效，轴向分量由于 SGCMG 框架轴方向不同，因而利用框架角投影到 z 向相加。在框架角等效偏转范围内，框架角对轴向角动量影响很小，因而利用两 SGCMG 角动量标量和的二分之一等效。因而确立等效系数为 (1,1,0.5)。等效示意图如图 7 所示。

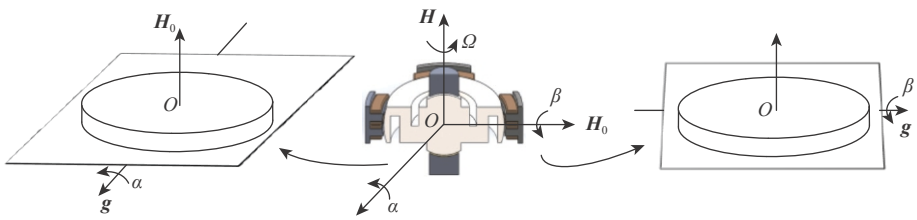


图 7 MSCSG 转子角动量等效示意图

Fig. 7 Angular momentum equivalent of MSCSG rotor

考虑单个 SGCMG 角动量表达式为

$$\boldsymbol{h}=\pm J_z\Omega\frac{\boldsymbol{g}\boldsymbol{u}\boldsymbol{g}}{|\boldsymbol{g}\boldsymbol{u}|}\tag{4}$$

由于 MSCSG 工作在偏置动量模式，角动量方向不会改变。框架等效法下 MSCSG 等效角动量表达式为

$$\boldsymbol{h}=\mathrm{diag}[1\quad 1\quad 0.5]\cdot\left[\sum_{i=1}^{2}J_z\Omega\frac{(\boldsymbol{g}_i\times\boldsymbol{u}_i)\times\boldsymbol{g}_i}{|\boldsymbol{g}_i\times\boldsymbol{u}_i|}\right]\tag{5}$$

式中：$\boldsymbol{u}$表示 MSCSG 期望角动量方向。由于 MSCSG 转子偏转角度存在最大值δ，所以$\boldsymbol{u}$形状为半径等于 1，最大开口圆半径为$\sin\delta$的球冠。

为进行误差分析，根据等效方式，可以得到基于偏转角参数的等效 MSCSG 角动量表达式为

$$\boldsymbol{H}=J_z\Omega\begin{bmatrix}-\sin\alpha\\\sin\beta\\\dfrac{\cos\alpha+\cos\beta}{2}\end{bmatrix}\tag{6}$$

文献 [24] 给出 DGMW 角动量表达式为

$$\boldsymbol{H}_{\mathrm{D}}=J_{\mathrm{D}z}\Omega_{\mathrm{D}}\begin{bmatrix}\cos\theta_{\mathrm{j}}\cos\theta_{\mathrm{g}}\\\sin\theta_{\mathrm{j}}\cos\theta_{\mathrm{g}}\\\sin\theta_{\mathrm{g}}\end{bmatrix}\tag{7}$$

式中：$J_{\mathrm{D}z}$为 DGMW 转子转动惯量；Ω_{D}为 DGMW 转子转速；θ_{j}为外框架角；θ_{g}为内框架角。比较式 (3) 与式 (7)，排除坐标系建立的区别，MSCSG 与 DGMW 角动量表达方式一致。因而框架等效法也适用于 DGMW 的角动量分析。

2.2 框架等效法 MSCSG 角动量包络模型误差分析

首先，分析 MSCSG 角动量模型绝对误差。在 x 轴方向，并不存在误差。在 y 轴方向等效角动量大小为变为原有角动量的 $1/\cos\beta$ 倍。将框架等效法与传统遍历法轴向角动量大小相减，可以得到轴向角动量误差为

$$\Delta=\left|\cos\alpha\cos\beta-\frac{\cos\alpha+\cos\beta}{2}\right|=$$

$$\frac{\cos\alpha+\cos\beta}{2}-\cos\alpha\cos\beta>0\tag{8}$$

求取Δ的 Hessian 矩阵$\boldsymbol{H}(\Delta)$：

$$\boldsymbol{H}(\Delta)=\begin{bmatrix}\dfrac{\partial^2\Delta}{\partial\alpha^2}&\dfrac{\partial^2\Delta}{\partial\alpha\partial\beta}\\\dfrac{\partial^2\Delta}{\partial\beta\partial\alpha}&\dfrac{\partial^2\Delta}{\partial\beta^2}\end{bmatrix}=$$

$$\begin{bmatrix}\cos\alpha\cos\beta-\dfrac{\cos\alpha}{2}&\sin\alpha\sin\beta\\\sin\alpha\sin\beta&\cos\alpha\cos\beta-\dfrac{\cos\beta}{2}\end{bmatrix}\tag{9}$$

由于

$$\frac{\partial^2\Delta}{\partial\alpha^2}=\cos\alpha\cos\beta-\frac{\cos\alpha}{2}=\cos\alpha\left(\cos\beta-\frac{1}{2}\right)>0\tag{10}$$

$$\begin{vmatrix} \dfrac{\partial^2 \Delta}{\partial \alpha^2} & \dfrac{\partial^2 \Delta}{\partial \alpha \partial \beta} \\ \dfrac{\partial^2 \Delta}{\partial \beta \partial \alpha} & \dfrac{\partial^2 \Delta}{\partial \beta^2} \end{vmatrix} = \left(\cos\alpha\cos\beta - \frac{\cos\alpha}{2}\right) \cdot \left(\cos\alpha\cos\beta - \frac{\cos\beta}{2}\right) - \sin^2\alpha\sin^2\beta =$$

$$\cos\alpha\cos\beta\left(\frac{1}{4} - \frac{\cos\alpha}{2} - \frac{\cos\beta}{2} + \cos\alpha\cos\beta\right) - \left(1-\cos^2\alpha\right)\left(1-\cos^2\beta\right) = \frac{\cos\alpha\cos\beta}{4} - \frac{\cos^2\alpha\cos\beta}{2} - \frac{\cos\alpha\cos^2\beta}{2} - 1 + \cos^2\alpha + \cos^2\beta > \frac{\cos\alpha\cos\beta}{4} - \frac{\cos^2\alpha\cos\beta}{2} - \frac{\cos\alpha\cos^2\beta}{2} - 1 + \frac{\cos^2\alpha}{2} + \frac{\cos^2\beta}{2} + \frac{\cos^2\alpha\cos\beta}{2} + \frac{\cos\alpha\cos^2\beta}{2} = \frac{\cos\alpha\cos\beta}{4} + \frac{\cos^2\alpha}{2} + \frac{\cos^2\beta}{2} - 1 > \frac{\cos^2 2\cos 2}{4} + \frac{\cos^2 2}{2} + \frac{\cos^2 2}{2} - 1 > 0 \tag{11}$$

因而 $\boldsymbol{H}(\Delta)$为正定矩阵。由于转子存在最大偏转角$\delta = 2°$, 因而径向偏转角α和β的取值范围均为$(-2°, 2°)$, 令

$$\begin{cases} \cos\alpha\cos\beta - \dfrac{\cos\alpha}{2} = 0 \\ \cos\alpha\cos\beta - \dfrac{\cos\beta}{2} = 0 \end{cases} \tag{12}$$

在径向偏转角α和β取值范围内解的$\alpha = 0°$, $\beta = 0°$。此时误差局部极小值为 0。求取一阶偏导得

$$\begin{cases} \dfrac{\mathrm{d}\Delta}{\mathrm{d}\alpha} = \sin\alpha\cos\beta - \dfrac{\sin\alpha}{2} = \sin\alpha\left(\cos\beta - \dfrac{1}{2}\right) \\ \dfrac{\mathrm{d}\Delta}{\mathrm{d}\beta} = \cos\alpha\sin\beta - \dfrac{\sin\beta}{2} = \sin\beta\left(\cos\alpha - \dfrac{1}{2}\right) \end{cases} \tag{13}$$

因而随着径向偏转角的绝对值增大误差值增大, 因而当偏转角α和β均为2°时, 轴向误差值为

$$\Delta = (\cos 2° + \cos 2°)/2 - \cos 2°\cos 2° \approx 0.000\,6 J_z \Omega \tag{14}$$

除模型绝对误差外还有角动量指向误差, 以 MSCSG 为例, 初始角动量$\boldsymbol{H}_0$方向为$[0 \quad 0 \quad 1]^{\mathrm{T}}$, 已知二者角动量末尾方向。记 2 种角动量表达方式角动量偏转角为δ_1和δ_2。准确偏转角δ_1表达式为

$$\delta_1 = \arccos\left(\frac{\boldsymbol{H}\cdot\boldsymbol{H}_0}{|\boldsymbol{H}||\boldsymbol{H}_0|}\right) = \arccos(\cos\alpha\cos\beta) \tag{15}$$

由于机械结构限制, 准确偏转角$\delta_1 < 2°$。等效偏转角δ_2表达式可写为

$$\delta_2 = \arccos\left(\frac{\boldsymbol{H}\cdot\boldsymbol{H}_0}{|\boldsymbol{H}||\boldsymbol{H}_0|}\right) = \arccos\left[\frac{(\cos\alpha + \cos\beta)/2}{\sqrt{\sin^2\alpha + \sin^2\beta + (\cos^2\alpha + \cos^2\beta)/4}}\right] \tag{16}$$

因而方向误差表达式为

$$\delta_1 - \delta_2 = \arccos(\cos\alpha_2\cos\beta_2) - \arccos\left[\frac{(\cos\alpha_2 + \cos\beta_2)/2}{\sqrt{\sin^2\alpha_2 + \sin^2\beta_2 + (\cos^2\alpha_2 + \cos^2\beta_2)/4}}\right] \tag{17}$$

式 (17) 求极值较为复杂。$\delta_1 - \delta_2$图像如图 8 所示, z 方向坐标为弧度制。由图 8 可以看出, 在转子偏转角范围内, 角动量误差约为真实值的千分之一。且误差随着偏转角的增大而增大。当$\alpha = 2°$, $\beta = 2°$时误差值约为0.000 9°。

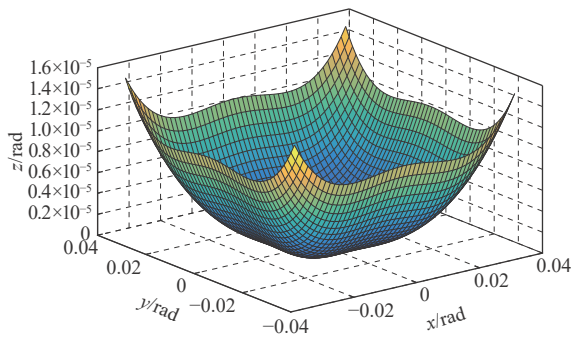


图 8　$\delta_1 - \delta_2$图像

Fig. 8　Image of $\delta_1 - \delta_2$

通过误差分析, 无论是模型绝对误差还是角动量指向误差, 都维持在千分之一的量级, 所以利用 SGCMG 转子角动量等效 MSCSG 转子角动量可行。

2.3　框架等效法快速性分析

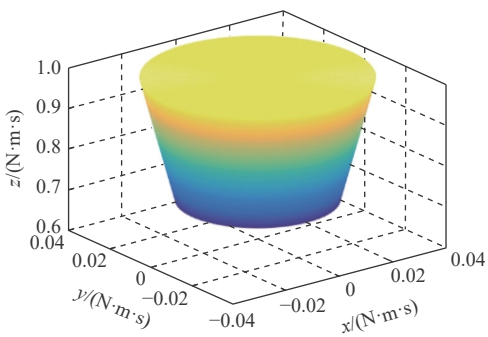
框架等效法将遍历法中的角度参数(利用其在空间中的相对位置), 在同一坐标系中表示, 降低了参数纬度。因而当分析计算量减少时, 应当区分角动量交换装置是否具有飞轮效应。由于所有角度参数都简化成了期望角动量方向, 利用 2 个角度参数即可表示。基于上述分析, 设具有n个角动量交换装置, 当转子转速变化具有飞轮效应时, 遍历法计算参数维度为$3n$, 框架等效法参数为$n+2$, 二者差值为$2n-2$。当转子不具备飞轮效应时遍历法计算参数维度为$2n$, 框架等效法参数维度为2, 二者差值为$2n-2$。所以, 框架等效法相对于遍历法可以减少$2n-2$个参数维度。在三正交构型下, 共有 3 个角动量交换装置。在角动量交换装置具有飞轮效应时, 参数维度减少了 44.4%; 在角动量交换装置无飞轮效应时, 参数维度减少了 66.7%。

3　仿真与结果

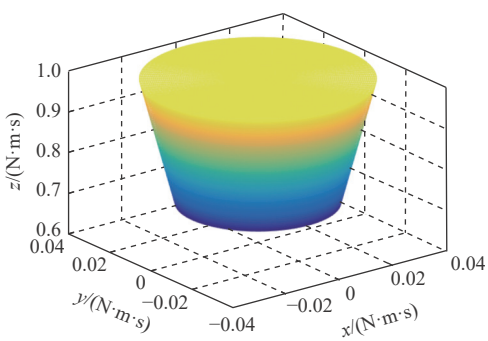
3.1　基于本文方法 MSCSG 角动量包络仿真

在径向角动量大小等效时, 将径向偏转角等

效为2个相互独立的SGCMG框架旋转角，使第2次径向偏转产生误差。在轴向角动量大小等效时，将MSCSG轴向角动量等效为SGCMG转子z方向角动量之和，引入模型绝对误差。以MSCSG为例，y、z方向均存在模型绝对误差，且y、z方向误差明显不成比例，因而也存在角动量指向误差。经过第2节的分析，y方向角动量大小扩大为原角动量的$1/\cos\beta_2$倍。当$\beta_2=2°$时，$1/\cos\beta_2=1.000\,6$。在轴向，当$|\alpha_2|=|\beta_2|=2°$时，等效角动量大小与真实角动量大小的误差约为$0.000\,6J_z\Omega_z$。误差均在千分之一以下。利用传统遍历法及等效框架法实现对单MSCSG角动量包络进行对比，结果如图9所示。



(a) MSCSG 遍历法



(b) MSCSG 等效法

图9 MSCSG角动量包络对比

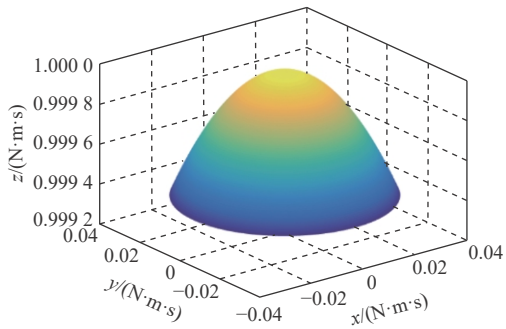
Fig. 9 Angular momentum envelope comparison of MSCSG

由于最大误差值相对大小约为实际大小的千分之一，因而二者区别很小。所以如此等效是合理的。

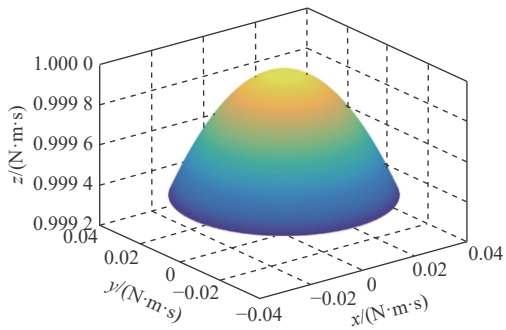
3.2 基于框架等效法DGMW角动量包络仿真

转子初始指向为z方向，设置转子最大偏转2°并进行单位化处理。仿真结果如图10所示。

通过对比可以发现，轴向角动量大小出现了千分之一以下的误差。当外框架角θ_j旋转2°，内框架角θ_g不旋转时，转子旋转角度为2°，此时轴向误差约为0.000 3。径向由于图像分度值远大于千分之一，图10中体现并不明显。



(a) DGMW 角动量包络遍历法



(b) DGMW 角动量包络框架等效法

图10 DGMW角动量包络对比

Fig. 10 Angular momentum envelope comparison of DGMW

4 结 论

1）本文比较了基于机械实体框架(SGCMG和DGMW)与基于磁悬浮虚拟框架的MSCSG角动量交换原理的异同，利用双正交SGCMG角动量径向分量和轴向分量的一半分别等效MSCSG径和轴向角动量，构建了MSCSG微框架等效模型。

2）分析等效模型与角动量表达式，量化分析存在的误差，证明大小误差及方向误差均在千分之一以下。仿真得到利用遍历法及本文方法计算的单MSCSG角动量包络与单DGCMG局部角动量包络，证明了本文方法的有效性。

3）对比MSCSG与DGCMG角动量表达式，本文方法同样适用于DGCMG局部角动量分析。通过计算，当构型中具有n个角动量交换装置时，本文方法相较于遍历法可以减少$2n-2$个参数维度，证明了本文方法具有快速性的特点。因而本文方法在MSCSG群角动量包络分析中具有更广泛的应用。

参考文献（References）

[1] 周军, 刘莹莹. 航天器姿态与轨道控制原理[M]. 西安: 西北工业大学出版社, 2016: 130-133.
ZHOU J, LIU Y Y. Principle of spacecraft attitude and orbit control[M]. Xi'an: Northwestern Polytechnical University Press, 2016: 130-133(in Chinese).

[2] 王巍. 新型惯性技术发展及在宇航领域的应用[J]. 红外与激光工

程, 2016, 45(3): 11-16.
WANG W. Development of new inertial technology and its application in aerospace field[J]. Infrared and Laser Engineering, 2016, 45(3): 11-16(in Chinese).

[3] YIN Z Y, CAI Y W, LIU B, et al. Application of spherical magnetic bearing in magnetically suspended control and sensitive gyro[J]. Mathematical Problems in Engineering, 2020, 2020: 7698794.

[4] 魏静波. 磁悬浮微框架动量轮控制技术研究[D]. 长沙: 国防科技大学, 2016: 19-21.
WEI J B. Research on control technology of momentum wheel of magnetic suspension micro-frame[D]. Changsha: National University of Defense Technology, 2016: 19-21(in Chinese).

[5] 吕奇超, 周一恒, 吕东元, 等. 基于自适应滤波器的磁悬浮控制力矩陀螺内转子振动抑制[J]. 导航与控制, 2021, 20(1): 70-77.
LYU Q C, ZHOU Y H, LYU D Y, et al. Vibration suppression of the inner rotor of a small-sized magnetically suspended control moment gyroscope based on adaptive filter[J]. Navigation and Control, 2021, 20(1): 70-77(in Chinese).

[6] 刘彬, 房建成, 刘刚. 一种磁悬浮陀螺飞轮方案设计与关键技术分析[J]. 航空学报, 2011, 32(8): 1478-1487.
LIU B, FANG J C, LIU G. Design of a magnetically suspended gyrowheel and analysis of key technologies[J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(8): 1478-1487(in Chinese).

[7] REN Y, CHEN X C, CAI Y W, et al. Attitude-rate measurement and control integration using magnetically suspended control and sensitive gyroscopes[J]. IEEE Transactions on Industrial Electronics, 2018, 65(6): 4921-4932.

[8] 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).

[9] YU C M, WANG Z, REN Y, et al. MSCSG two degree of freedom attitude measurement method[J]. Journal of Physics: Conference Series, 2019, 1176: 042054.

[10] 杨倩. 动基座下双框架 MSCMG 磁轴承系统稳定性研究[D]. 北京: 北京航空航天大学, 2011.
YANG Q. Research on stability of magnetic bearing system for double gimbal magnetically suspended control moment gyroscope on moving base [D]. Beijing: Beihang University, 2011(in Chinese).

[11] 任元, 王卫杰, 刘强, 等. 一种磁悬浮控制敏感陀螺: CN104613950B [P]. 2017-06-27.
REN Y, WANG W J, LIU Q, et al. A kind of magnetically suspended control and sensing gyroscopic: CN104613950B [P]. 2017-06-27(in Chinese).

[12] 李佳益. 磁悬浮微框架动量轮动力学建模与仿真研究[D]. 长沙: 国防科技大学, 2012: 2-4.
LI J Y. Research on dynamic modeling and simulation of momentum wheel of magnetic levitation micro-frame[D]. Changsha: National University of Defense Technology, 2012: 2-4(in Chinese).

[13] MARGULIES G, AUBURN J N. Geometric theory of single-gimbal control moment gyro systems[J]. Journal of the Astronautical Sciences, 1978, 26(2): 159-191.

[14] WIE B. Singularity analysis and visualization for single-gimbal control moment gyro systems[J]. Journal of Guidance, Control, and Dynamics, 2004, 27(2): 271-282.

[15] TANG L, XU S J. Geometric analysis of singularity for single-gimbal control moment gyro systems[J]. Chinese Journal of Aeronautics, 2005, 18(4): 2346-2353.

[16] 王磊, 赵育善. 双框架控制力矩陀螺奇异分析及可视化[J]. 宇航学报, 2009, 30(2): 613-619.
WANG L, ZHAO Y S. Singularity analysis and visualization of double-gimbaled control moment gyro systems[J]. Journal of Astronautics, 2009, 30(2): 613-619(in Chinese).

[17] 刘锋. 金字塔构型变速控制力矩陀螺操纵律和参数优化设计[D]. 哈尔滨: 哈尔滨工业大学, 2018.
LIU F. Control law and parameter optimization design of pyramid variable speed control moment gyro[D]. Harbin: Harbin Institute of Technology, 2018(in Chinese).

[18] 李传江, 郭延宁, 马广富. 单框架控制力矩陀螺的奇异分析及操纵律设计[J]. 宇航学报, 2010, 31(10): 2346-2353.
LI C J, GUO Y N, MA G F. Singularity analysis and steering law design for single-gimbal control moment gyroscopes[J]. Journal of Astronautics, 2010, 31(10): 2346-2353(in Chinese).

[19] 侯二永. 磁悬浮控制力矩陀螺结构设计与动力学分析[D]. 长沙: 国防科技大学, 2013: 27-28.
HOU E Y. Structural design and dynamic analysis of magnetic suspension control moment gyro[D]. Changsha: National University of Defense Technology, 2013: 27-28(in Chinese).

[20] 宁欣, 韩邦成, 房建成. 基于干扰观测器的双框架变速率控制力矩陀螺解耦控制[J]. 机械工程学报, 2017, 53(10): 52-59.
NING X, HAN B C, FANG J C. Disturbance observer based decoupling method of double-gimbaled variable speed control moment gyroscope[J]. Journal of Mechanical Engineering, 2017, 53(10): 52-59(in Chinese).

[21] 魏孔明, 吴忠, 刘涛. 单框架控制力矩陀螺构型分析与奇异可视化[J]. 中国空间科学技术, 2013, 33(1): 21-29.
WEI K M, WU Z, LIU T. Configuration analysis and singularity visualization of single gimbal control moment gyroscopes[J]. Chinese Space Science and Technology, 2013, 33(1): 21-29(in Chinese).

[22] 雷拥军, 袁利. 轮系航天器的角动量包络分析及角动量管理[J]. 中国空间科学技术, 2017, 37(6): 1-9.
LEI Y J, YUAN L. Momentum envelope analysis and momentum management for spacecraft with a flywheel array[J]. Chinese Space Science and Technology, 2017, 37(6): 1-9(in Chinese).

[23] 徐洪亮, 翟传润, 战兴群, 等. 双框架控制力矩陀螺伪逆控制律设计及仿真[J]. 上海交通大学学报, 2008, 42(5): 851-855.
XU H L, ZHAI C R, ZHAN X Q, et al. Design and simulation of pseudo-inverse control law for double gimbal control moment gyro[J]. Journal of Shanghai Jiao Tong University, 2008, 42(5): 851-855(in Chinese).

[24] 焉宁. 敏捷机动小卫星姿态控制方法的仿真与实验研究[D]. 北京: 北京航空航天大学, 2011.
YAN N. Research on the simulation and experiment of attitude control method for agile small satellite [D]. Beijing: Beihang University, 2011(in Chinese).

Angular momentum envelope analysis method of gimbal-type momentum exchange device

WANG Weijie[1,*], WANG Zhou[1], PANG Weikun[2], YANG Yang[3]

(1. Department of Aerospace Science and Technology, Space Engineering University, Beijing 101416, China;
2. Graduate School, Space Engineering University, Beijing 101416, China;
3. School of Artificial Intelligence, Henan University, Zhengzhou 475001, China)

Abstract: To improve the angular momentum envelope analysis efficiency of typical gimbal-type momentum exchange devices such as single-gimbal control moment gyroscope (SGCMG), double-gimbal momentum wheel (DGMW), and magnetically suspended control and sensing gyroscope (MSCSG), a gimbal equivalent-based angular momentum envelope analysis method of gimbal-type momentum wheel was proposed. The similarities and differences of angular momentum exchange principles between SGCMG and DGMW based on a mechanical solid gimbal and MSCSG based on a magnetic levitation micro-gimbal. The equivalent coefficients between radial/axial angular momentum components of SGCMG and MSCSG were studied and designed. The biorthogonal solid gimbal of SGCMG was used to construct the equivalent model of the MSCSG micro-gimbal. The angular momentum expressions of ergodic methods for MSCSG and DGMW were compared, and the applicability of the gimbal equivalent method in the angular momentum analysis of DGMW was analyzed. The error of the gimbal equivalent method was quantified to prove that the absolute error and the direction error of angular momentum calculated by the gimbal equivalent method are both less than one thousandth compared with the traditional ergodic method. The parameter dimensions of the two methods were analyzed to prove the rapidity of the gimbal equivalent method. The angular momentum envelopes of MSCSG and DGMW were simulated and compared based on the traditional ergodic method and the gimbal equivalent method, respectively, which proved the effectiveness of the method. This method has a wider application in angular momentum envelope analysis of MSCSG groups.

Keywords: single-gimbal control moment gyroscope; double-gimbal momentum wheel; magnetically suspended control and sensing gyroscope; angular momentum envelope; error analysis

Received: 2023-05-23; **Accepted**: 2023-05-26; **Published Online**: 2023-09-21 10:27
URL: link.cnki.net/urlid/11.2625.V.20230920.1623.001

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0282

基于多毁伤等级标准的导弹毁伤效能贝叶斯鉴定试验设计

陈童，刘昊邦*，胡涛
(海军工程大学 管理工程与装备经济系，武汉 430033)

摘　　要：导弹毁伤效能合格是充分发挥战斗力的重要保障，为检验导弹毁伤效能是否达标，开展毁伤效能鉴定试验研究尤为必要。针对现有“单发毁伤概率”毁伤效能定量表征指标依据单一毁伤标准难以全面描述导弹毁伤效能的局限性，考虑单枚导弹打击目标形成的不同毁伤等级结果，以不同毁伤等级结果发生概率表征毁伤效能，从多毁伤等级标准角度进行导弹鉴定，有利于全面检验导弹毁伤效能。并且为克服小样本导弹整装试验信息不足难题，在鉴定试验设计中利用贝叶斯方法结合体系贡献度融合多源先验信息，提出信息利用充分、双方风险可控的试验方法。研究结果表明：相比单一毁伤标准定量表征指标，所提方法能够全面描述导弹毁伤效能，并与改进后的二项分布假设检验法进行对比，验证了所提方法的优越性。

关 键 词：导弹毁伤效能；鉴定试验；多毁伤等级标准；体系贡献度；贝叶斯
中图分类号：E917
文献标志码：A　　**文章编号**：1001-5965(2025)05-1599-09

导弹毁伤效能合格对于火力打击任务完成及战术规划决策实施至关重要[1]。通过检验导弹毁伤效能是否达标有利于保证导弹使用方接收到合格的导弹，因此，有必要进行导弹毁伤效能鉴定试验研究。

现有导弹毁伤效能研究主要以“单发毁伤概率”和“期望毁伤概率的用弹量”作为定量表征指标[2-4]。赵晓旭等[5]以“单发毁伤概率”为指标对制导杀爆弹毁伤效能评估的应用进行研究。朱莉娜和李翰山[6]通过建立毁伤元对目标侵彻毁伤的评估模型分析智能弹药设备对目标近炸毁伤效能，结果发现随着毁伤元侵彻目标角度增大，毁伤概率变小。高源等[7]基于Monte-Carlo方法对破障火箭弹终点毁伤效能评估进行研究，并以“期望毁伤概率的用弹量”作为评估表征指标。侯鹏等[8]从毁伤效能角度研究破片式战斗部对飞机瞄准点的选择方法。此外为保证导弹使用方接收导弹质量合格，在导弹鉴定试验领域也出现一些研究。贠来峰等[9]建立了一种新的计量型一次抽样检验方法对导弹产品密集度进行检验。王玮等[10]利用贝叶斯方法对导弹可靠性二次抽样鉴定试验方案进行改进。现有导弹鉴定试验主要考虑密集度、可靠性方面，毁伤效能方面的鉴定试验研究较少。

以“单发毁伤概率”或“期望毁伤概率的用弹量”表征导弹毁伤效能时，主要从导弹达到单一毁伤标准概率角度进行衡量。实际上，导弹毁伤效能试验会产生不同程度毁伤效果，从多毁伤等级角度全面鉴定导弹毁伤效能，有利于提升接收导弹毁伤效能的可靠性[11-12]。并且考虑到导弹整装试验受成本、环境制约，难以进行足量外场试验，仅利用小样本外场试验信息难以对导弹毁伤效能做出准确评价。为克服产品鉴定试验中小样本试验信息量不

收稿日期：2023-05-25；录用日期：2023-08-11；网络出版时间：2023-09-11 08:49
网络出版地址：link.cnki.net/urlid/11.2625.V.20230908.0904.002
基金项目：国家自然科学基金 (71501183)
*通信作者. E-mail：D23181102@nue.edu.cn

引用格式：陈童，刘昊邦，胡涛. 基于多毁伤等级标准的导弹毁伤效能贝叶斯鉴定试验设计 [J]. 北京航空航天大学学报，2025，51（5）：1599-1607. CHEN T，LIU H B，HU T. Bayesian identification test design of missile damage effectiveness based on multiple damage grades standards [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1599-1607（in Chinese）.

足的问题，谭尧等[13]在威布尔型产品定时截尾可靠性鉴定试验设计中融入专家信息，设计出试验时间较短、风险可控的试验方案。刘海涛等[14]针对成败型产品生产定型阶段可靠性验收问题，利用贝叶斯方法综合先验信息减少鉴定试验数量。以上研究利用贝叶斯方法在产品鉴定试验中融入先验信息，增加产品信息量。

综合上述分析，为保证导弹使用方接收导弹的毁伤效能达标，有必要对导弹毁伤效能开展鉴定试验。本文针对现有“单发毁伤概率”或“期望毁伤概率的用弹量”定量表征指标衡量导弹毁伤效能较为单一问题，从多毁伤等级标准角度检验导弹毁伤效能更为全面。同时为克服导弹整装发射试验信息量不足问题，利用贝叶斯方法结合体系贡献度融合多源先验信息进行鉴定研究。研究结果有利于提升接收导弹毁伤效能的可靠性，并降低鉴定双方风险，为导弹毁伤效能鉴定试验设计提供借鉴。

1 导弹毁伤效能试验描述

1.1 单一毁伤标准毁伤效能试验描述

导弹毁伤效能体现在打击目标造成的毁伤效果。现有毁伤效能研究主要从单一毁伤标准考虑，以规定的某一毁伤程度效果为标准，达到这一毁伤效果的概率越大，毁伤效能越好。按照打击目标造成的毁伤结果对毁伤效能试验进行分布描述。单一毁伤标准毁伤效能试验可用二项分布进行描述：

$$f(x)=C_n^x p^x(1-p)^{n-x} \tag{1}$$

式中：p为导弹达到单一毁伤标准的概率；n为试验导弹总数；x为达到单一毁伤标准的导弹数。

1.2 多毁伤等级标准毁伤效能试验描述

导弹毁伤效能试验通常产生多个毁伤等级效果，将目标毁伤效果划分为5个等级[15-16]，如表1所示。多毁伤等级标准毁伤效能试验可用多项分布进行描述：

$$f(X\mid n,\phi)=n!\prod_i^5 p_i^{x_i}\Big/\prod_i^5 x_i! \tag{2}$$

式中：$\phi=(p_1,p_2,p_3,p_4,p_5)$为$p_i$的集合，分别代表零毁伤概率、轻度毁伤概率、中度毁伤概率、重度毁伤概率及摧毁歼灭概率；$X=(x_1,x_2,x_3,x_4,x_5)$为x_i的集合，分别代表各毁伤等级事件发生的次数。

多项分布描述毁伤效能试验更为全面，但多项分布公式复杂，不易进行模型求解，此时转化为二项分布进行分析。规定轻度毁伤标准包括轻度及以上毁伤等级效果，中度毁伤标准包括中度及以上毁伤等级效果，重度毁伤标准包括重度及以上毁伤等级效果，摧毁歼灭标准为摧毁歼灭等级效果。

表1 目标毁伤效果等级划分

Table 1 Classification of target damage effect

目标毁伤等级	毁伤情况
零毁伤	目标轻微受损或无损伤，作战效能损失在5%以内
轻度毁伤	目标性能受到轻度损失，如果不及时进行修理，会影响作战性能，作战效能损失5%～20%
中度毁伤	目标损伤较为严重，需要换件修理，作战效能损失20%～50%
重度毁伤	目标严重受损，需要返厂大修，且修理耗时较长，作战效能损失50%～80%
摧毁歼灭	目标完全摧毁，且无法修复，作战效能损失在80%以上

轻度毁伤标准毁伤效能试验表示为

$$f(y_1)=C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1} \tag{3}$$

式中：$\theta_1=p_2+p_3+p_4+p_5$为达到轻度毁伤标准的概率；$y_1=x_2+x_3+x_4+x_5$为达到轻度毁伤标准的导弹数。

中度毁伤标准毁伤效能试验表示为

$$f(y_2)=C_n^{y_2}\theta_2^{y_2}(1-\theta_2)^{n-y_2} \tag{4}$$

式中：$\theta_2=p_3+p_4+p_5$为达到中度毁伤标准的概率；$y_2=x_3+x_4+x_5$为达到中度毁伤标准的导弹数。

同理，可以得到重度毁伤标准、摧毁歼灭标准毁伤效能试验分布。

2 导弹毁伤效能多源先验信息融合

2.1 贝叶斯方法先验信息融合

充分利用已有多源先验信息对于小样本导弹毁伤效能鉴定试验尤为必要。多源先验信息包括专家知识信息、仿真试验信息等[17-18]。贝叶斯方法能够在鉴定试验中融入先验信息增加信息量，可表示为

$$\pi(\theta_i\mid y_i)=\frac{L(y_i\mid\theta_i)\pi(\theta_i)}{m(y_i)}=\frac{L(y_i\mid\theta_i)\pi(\theta_i)}{\int_\Theta L(y_i\mid\theta_i)\pi(\theta_i)\mathrm{d}\theta_i} \tag{5}$$

式中：$\pi(\theta_i|y_i)$为导弹毁伤效能参数θ_i的后验分布；$L(y_i|\theta_i)$为导弹毁伤效能试验样本的似然函数，包含外场射击试验信息；$\pi(\theta_i)$为导弹毁伤效能参数θ_i的先验分布，包含毁伤效能参数的先验信息；$m(y_i)$为试验样本的边缘分布，与参数θ_i无关；Θ表示导弹毁伤效能参数θ_i的积分区域，为参数θ_i的数值范围。式(5)可化简为

$$\pi(\theta_i\mid y_i)\propto L(y_i\mid\theta_i)\pi(\theta_i) \tag{6}$$

由1.2节可知导弹毁伤效能试验样本的似然函数$L(y_i|\theta_i)$服从二项分布，确定$\pi(\theta_i)$即可求出导弹毁伤效能参数的后验分布。贝叶斯先验分布选择方法有多种，包括共轭先验分布法、最大熵法、无信息先验分布等[19-21]。其中，共轭先验分布法使用较

简单, 且能够有效对先验信息进行利用, 采用该方法确定先验分布。选取二项分布的共轭分布 Beta 分布作为先验分布, 则导弹毁伤效能参数θ_i的先验分布表示为

$$\pi(\theta_i)=\theta_i^{a-1}(1-\theta_i)^{b-1}/\mathrm{B}(a,b) \tag{7}$$

式中: a、b为先验分布中的超参数, 反映参数θ_i的先验分布信息。

2.2 基于体系贡献度的多源先验信息融合权重计算

先验分布中包含多种来源先验信息, 不同来源先验信息对导弹毁伤效能鉴定的贡献程度不同。体系贡献度常用来衡量装备对作战体系综合作战能力的贡献程度, 也可用来测算各组成要素对综合体系的贡献程度[22-23]。引入体系贡献度衡量不同来源先验信息在导弹毁伤效能参数先验分布中的融合权重, 由相关领域专家分析不同来源先验信息的可靠性, 信息可靠性越高则对导弹毁伤效能鉴定的贡献度越高, 融合权重占比越大。利用专家认定不同来源先验信息的可靠性信息计算其对导弹毁伤效能鉴定的体系贡献度[24-25]。用$\phi'=\{\delta_1,\delta_2,\cdots,\delta_k\}$表示导弹毁伤效能不同来源先验信息集合。用$\omega_{\delta_j}^d(j=1,2,\cdots,k)$表示专家$\psi_d(d=1,2,\cdots,m)$认定导弹毁伤效能信息来源$\delta_j$的可靠隶属度。用$\upsilon_{\delta_j}^d$表示专家$\psi_d$认定导弹毁伤效能信息来源$\delta_j$的不可靠隶属度。采用直觉模糊评价法对导弹毁伤效能信息来源是否可靠进行表示:

$$\Theta_{\delta_j}^d=\left\{\delta_j,\left(\omega_{\delta_j}^d,\upsilon_{\delta_j}^d\right)\right\} \tag{8}$$

式中: $0\leqslant\omega_{\delta_j}^d\leqslant1$, $0\leqslant\upsilon_{\delta_j}^d\leqslant1$, 且$\omega_{\delta_j}^d+\upsilon_{\delta_j}^d\leqslant1$。

专家看待导弹毁伤效能不同来源先验信息是否可靠时会犹豫, 引入直觉指标表示专家犹豫程度, 表达式为

$$\tau_{\delta_j}^d=1-\omega_{\delta_j}^d-\upsilon_{\delta_j}^d \tag{9}$$

采用犹豫程度的一半对多源先验信息的体系贡献度进行修正, 求出信息来源δ_j的体系贡献度表达式[26]:

$$\varepsilon_{\delta_j}=\frac{1}{m}\sum_{d=1}^{m}\left(\omega_{\delta_j}^d-\upsilon_{\delta_j}^d\right)+\left(1-\frac{1}{m}\sum_{d=1}^{m}\omega_{\delta_j}^d-\frac{1}{m}\sum_{d=1}^{m}\upsilon_{\delta_j}^d\right)\Big/2 \tag{10}$$

进行归一化处理, 得到信息来源δ_j的融合权重:

$$\varepsilon_{\delta_j}^{\tau}=\varepsilon_{\delta_j}\Big/\sum_{j=1}^{k}\varepsilon_{\delta_j} \tag{11}$$

跟据融合权重求出多源先验信息融合后的先验分布:

$$\pi(\theta_i)=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\pi\left(\theta_i\,|\,\delta_j\right) \tag{12}$$

式中: $\pi(\theta_i\,|\,\delta_j)$为信息来源$\delta_j$对应的先验分布。

导弹毁伤效能多源先验信息融合的贝叶斯公式为

$$\pi(\theta_i\,|\,y_i)=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\frac{L(y_i\,|\,\theta_i)\pi\left(\theta_i\,|\,\delta_j\right)}{m(y_i)} \tag{13}$$

该多源先验信息的体系贡献度计算过程能够充分利用专家认定该信息的可靠性与不可靠性信息, 并将专家的犹豫情况考虑进去, 利用专家信息更为充分, 有利于得到更为合理的融合权重结果。

3 导弹生产方风险与使用方风险贝叶斯推断

导弹毁伤效能鉴定试验需要控制导弹生产方风险与使用方风险。利用贝叶斯方法融入多源先验信息对双方风险进行推断。

3.1 导弹生产方风险贝叶斯推断

当导弹毁伤效能达到规定毁伤标准的概率却被误认为不合格时, 此时产生导弹生产方风险, 可表示为

$$\alpha=P(\theta>p_\alpha|R) \tag{14}$$

通过贝叶斯推断可得

$$\alpha=P(R\,|\,\theta>p_\alpha)P(\theta>p_\alpha)/P(R) \tag{15}$$

式中: $P(R)$表示当达到规定毁伤标准的概率为θ时, $m>m^*$的概率, 即判定导弹毁伤效能不合格的概率。p_α为导弹生产方毁伤标准概率的“阈值”。$P(R|\theta>p_\alpha)$表示在$\theta>p_\alpha$的条件下, 判定导弹毁伤效能不合格的概率。按照不同毁伤等级标准计算导弹生产方风险。

1) 按照轻度毁伤标准, 有

$$\begin{aligned}P(R)&=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^1P(m_1>m_1^*)\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1=\\&\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^1\sum_{n-y_1=m_1^*+1}^{n}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1\end{aligned} \tag{16}$$

$$\begin{aligned}P\left(R\,|\,\theta_1>p_{\alpha_1}\right)P\left(\theta_1>p_{\alpha_1}\right)=&\\\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_{p_{\alpha_1}}^1P(m_1>m_1^*)\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1=&\\\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_{p_{\alpha_1}}^1\sum_{n-y_1=m_1^*+1}^{n}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1&\end{aligned} \tag{17}$$

$$\alpha_1=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_{p_{\alpha_1}}^{1}\sum_{n-y_1=m_1^*+1}^{n}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1}{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_{0}^{1}\sum_{n-y_1=m_1^*+1}^{n}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1} \tag{18}$$

式中：m_1为未达到轻度毁伤标准的导弹数；m_1^*为m_1的“阈值”，表示最大允许未达到轻度毁伤标准的导弹数；p_{α_1}为导弹生产方轻度毁伤标准概率的“阈值”。

2）按照中度毁伤标准，可得导弹生产方风险α_2：

$$\alpha_2=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_{p_{\alpha_2}}^{1}\sum_{n-y_2=m_2^*+1}^{n}C_n^{y_2}\theta_2^{y_2}(1-\theta_2)^{n-y_2}\pi(\theta_2\,|\,\delta_j)\mathrm{d}\theta_2}{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_{0}^{1}\sum_{n-y_2=m_2^*+1}^{n}C_n^{y_2}\theta_2^{y_2}(1-\theta_2)^{n-y_2}\pi(\theta_2\,|\,\delta_j)\mathrm{d}\theta_2} \tag{19}$$

式中：p_{α_2}为弹药生产方中度毁伤标准概率的“阈值”。

同理，可以求出重度毁伤标准导弹生产方风险α_3，以及摧毁歼灭标准导弹生产方风险α_4。

3.2 导弹使用方风险贝叶斯推断

当导弹毁伤效能未达到规定毁伤标准的概率却被误认为合格时，此时产生导弹使用方风险，可表示为

$$\beta=P\left(\theta<p_\beta\,|\,A\right) \tag{20}$$

通过贝叶斯推断可得

$$\beta=P\left(A\,|\,\theta<p_\beta\right)P\left(\theta<p_\beta\right)/P(A) \tag{21}$$

式中：$P(A)$表示当达到规定毁伤标准的概率为θ时，$m\leqslant m^*$的概率，即判定导弹毁伤效能合格的概率。p_β为导弹使用方毁伤标准概率的“阈值”。$P\left(A\,|\theta<p_\beta\right)$表示在$\theta<p_\beta$条件下，判定导弹毁伤效能合格的概率。按照不同毁伤等级标准计算导弹使用方风险。

1）按照轻度毁伤标准，有

$$P(A)=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^1 P(m_1\leqslant m_1^*)\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^1\sum_{n-y_1=0}^{m_1^*}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi(\theta_1\,|\,\delta_j)\mathrm{d}\theta_1 \tag{22}$$

$$P\left(A\,|\,\theta_1<p_{\beta_1}\right)P\left(\theta_1<p_{\beta_1}\right)=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^{p_{\beta_1}}P\left(m_1\leqslant m_1^*\right)\pi\left(\theta_1\,|\,\delta_j\right)\mathrm{d}\theta_1=\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^{p_{\beta_1}}\sum_{n-y_1=0}^{m_1^*}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi\left(\theta_1\,|\,\delta_j\right)\mathrm{d}\theta_1 \tag{23}$$

$$\beta_1=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^{p_{\beta_1}}\sum_{n-y_1=0}^{m_1^*}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi\left(\theta_1\,|\,\delta_j\right)\mathrm{d}\theta_1}{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^{1}\sum_{n-y_1=0}^{m_1^i}C_n^{y_1}\theta_1^{y_1}(1-\theta_1)^{n-y_1}\pi\left(\theta_1\,|\,\delta_j\right)\mathrm{d}\theta_1} \tag{24}$$

式中：p_{β_1}为导弹使用方轻度毁伤标准概率的“阈值”。

2）按照中度毁伤标准，可得导弹使用方风险β_2：

$$\beta_2=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^{p_{\beta_2}}\sum_{n-y_2=0}^{m_2^*}C_n^{y_2}\theta_2^{y_2}(1-\theta_2)^{n-y_2}\pi(\theta_2\,|\,\delta_j)\mathrm{d}\theta_2}{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}\int_0^{1}\sum_{n-y_2=0}^{m_2^*}C_n^{y_2}\theta_2^{y_2}(1-\theta_2)^{n-y_2}\pi(\theta_2\,|\,\delta_j)\mathrm{d}\theta_2} \tag{25}$$

同理，可以求出重度毁伤标准导弹使用方风险β_3，以及摧毁歼灭标准导弹使用方风险β_4。

鉴定试验设计需要控制导弹生产方风险α_i和使用方风险β_i，由双方共同商定α_i和β_i的“阈值”。在此条件下求取试验导弹数n和允许最大未达到毁伤标准导弹数m_i^*的值进而设计出导弹毁伤效能鉴定试验方案。

4 鉴定双方风险求解算法

上述2类风险计算的解析式非常复杂。注意到θ_i的先验分布$\pi(\theta_i)$为$\mathrm{B}(a,b)$分布，则计算过程可简化为

$$\begin{aligned}&C_n^{y_i}\theta_i^{y_i}(1-\theta_i)^{n-y_i}\pi(\theta_i)=\\&\frac{n!\Gamma(a+b)}{y_i!(n-y_i)!\Gamma(a)\Gamma(b)}\theta_i^{y_i}(1-\theta_i)^{n-y_i}\theta_i^{a-1}(1-\theta_i)^{b-1}=\\&\frac{n!\Gamma(a+b)}{y_i!(n-y_i)!\Gamma(a)\Gamma(b)}\theta_i^{y_i+a-1}(1-\theta_i)^{n+b-y_i-1}=\\&\frac{n!\Gamma(a+b)\Gamma(a+y_i)\Gamma(b+n-y_i)}{y_i!(n-y_i)!\Gamma(a)\Gamma(b)\Gamma(a+b+y_i)}\cdot\\&\mathrm{B}(a+y_i,b+n-y_i)\end{aligned} \tag{26}$$

令$K=\dfrac{n!\Gamma(a+b)\Gamma(a+y_i)\Gamma(b+n-y_i)}{y_i!(n-y_i)!\Gamma(a)\Gamma(b)\Gamma(a+b+y_i)}$，$K$为常数，双方风险可化简为

$$\alpha_i=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j\sum_{n-y_i=m_i^*+1}^{n}\int_{p_{\alpha_i}}^{1}\mathrm{B}(a_j+y_i,b_j+n-y_i)\mathrm{d}\theta_i}{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j\sum_{n-y_i=m_i^*+1}^{n}\int_{0}^{1}\mathrm{B}(a_j+y_i,b_j+n-y_i)\mathrm{d}\theta_i}=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j\sum_{n-y_i=m_i^*+1}^{n}\int_{p_{\alpha_i}}^{1}\mathrm{B}(a_j+y_i,b_j+n-y_i)\mathrm{d}\theta_i}{(n-m)\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j} \tag{27}$$

$$\beta_i=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j\sum_{n-y_i=0}^{m_i^*}\int_0^{p_{\beta i}}\mathrm{B}(a_j+y_i,b_j+n-y_i)\mathrm{d}\theta_i}{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j\sum_{n-y_i=0}^{m_i^*}\int_0^{1}\mathrm{B}(a_j+y_i,b_j+n-y_i)\mathrm{d}\theta_i}=\frac{\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j\sum_{n-y_i=0}^{m_i^*}\int_0^{p_{\beta i}}\mathrm{B}(a_j+y_i,b_j+n-y_i)\mathrm{d}\theta_i}{(n-m)\sum_{j=1}^{k}\varepsilon_{\delta_j}^{\tau}K_j} \tag{28}$$

这样可以将双方风险复杂解析式转换为求 $\mathrm{B}(a_j+y_i,b_j+n-y_i)$分布在区间$[0,p_{\beta i}]$、$[p_{\alpha i},1]$上的积分问题, 节省大量运算过程。

5　示例分析

为鉴定某型号导弹毁伤效能是否达标, 现对其毁伤效能鉴定试验进行设计。采用现有产品检验的二项分布假设检验法与本文方法进行对比。二项分布假设检验法只利用外场射击试验样本数据, 规定导弹毁伤效能鉴定假设检验为

$$\begin{cases}H_0:p=p_0\\H_1:p=p_0'\end{cases} \tag{29}$$

式中: p_0为导弹达到规定毁伤标准概率的指标值; p_0'为p_0的对比值, 略低于p_0; p_0、p_0'由鉴定双方共同商讨决定。

依据双方风险定义, 构建出双方风险方程表达式:

$$\begin{cases}\alpha=1-\sum_{y=0}^{m^*}\mathrm{C}_n^y p_0^{n-y}q_0^y\\\beta=\sum_{y=0}^{m^*}\mathrm{C}_n^y\left(p_0'\right)^{n-y}\left(q_0'\right)^y\end{cases} \tag{30}$$

式中: $q_0=1-p_0$, $q_0'=1-p_0'$; y为未达到毁伤标准的导弹数量; 控制双方风险可以设计出导弹毁伤效能鉴定试验方案(n,m^*), n为试验导弹总数, m^*为允许最大未达到毁伤标准的导弹数量。

现有二项分布假设检验法可用于单一毁伤标准鉴定试验问题, 即给定某一毁伤等级标准, 导弹打击目标造成的毁伤结果达到这一毁伤等级标准即可认定该枚导弹毁伤效能合格。为进一步与本文方法进行对比, 对现有二项分布假设检验法进行改进, 从多毁伤等级标准角度进行鉴定试验。改进后的二项分布假设检验法的双方风险方程表达式为

$$\begin{cases}\alpha_i=1-\sum_{y_i=0}^{m_i^*}C_n^{y_i}p_i^{n-y_i}q_i^{y_i}\\\beta_i=\sum_{y_i=0}^{m_i^*}C_n^{y_i}(p_i')^{n-y_i}(q_i')^{y_i}\end{cases} \tag{31}$$

式中: α_i、$\beta_i(i=1,2,3,4)$分别为不同毁伤等级标准下的鉴定双方风险; y_i为未达到相应毁伤等级标准的导弹数量; p_i为不同毁伤等级标准概率的指标值, p_i'为p_i的对比值, 略低于p_i, $q_i=1-p_i$, $q_i'=1-p_i'$。

确定导弹毁伤效能鉴定试验的关键参数, 包括改进后的二项分布假设检验法 (方法 1) 参数、本文方法(方法 2)参数及先验分布参数信息, 其中, 先验信息主要包括专家知识信息和仿真试验信息。导弹毁伤效能鉴定试验参数如表 2 所示。

按照体系贡献度打分规则, 邀请 5 位同领域其他专家对专家知识信息δ_1和仿真试验信息δ_2的隶属度进行打分。打分结果如表 3 所示。

表 2　导弹毁伤效能鉴定试验参数

Table 2　Missile damage effectiveness identification test parameters

毁伤等级标准	方法1鉴定试验参数$p_i(p_i')$	方法2鉴定试验参数$p_{\alpha_i}=p_{\beta_i}$	专家知识信息先验分布$\mathrm{B}(a_1,b_1)$	仿真试验信息先验分布$\mathrm{B}(a_2,b_2)$
轻度毁伤标准	0.9 (0.8)	0.9	(4,0.5)	(4.5,0.5)
中度毁伤标准	0.7 (0.6)	0.7	(3,1.5)	(3.5,1)
重度毁伤标准	0.4 (0.3)	0.4	(2.5,3.5)	(1.5,2.5)
摧毁歼灭标准	0.2 (0.1)	0.2	(1.5,4.5)	(1,3.5)

依据专家隶属度打分数据进行融合权重计算, 得到$\varepsilon_{\delta_1}^{\tau}=0.544$, $\varepsilon_{\delta_2}^{\tau}=0.456$。根据隶属度打分结果说明专家认定仿真试验信息$\delta_2$的可信性低于专家知识信息$\delta_1$的可信性, 因此, 仿真试验信息$\delta_2$在进行信息融合时的融合权重占比较小。通过降低可信性不高的先验信息的融合权重, 在一定程度上有利于保证基于先验信息得到的结果是合理可信的。

5.1　给定试验导弹数n设计鉴定试验

假设一共发射 15 枚导弹进行鉴定试验设计。采用方法 1 和方法 2 分别计算导弹生产方风险α_i与使用方风险β_i随m_i的变化情况, 结果如图 1 所示。由图 1 可知, 当给定试验导弹总数$n=15$时, 随着m_i的增加, 导弹生产方风险呈现降低趋势, 使用方风险呈增长趋势。采用本文方法(方法 2), 规定生产方风险“阈值”为 0.2, 使用方风险“阈值”为

表3 先验信息专家隶属度打分

Table 3 Prior information expert membership rating

专家序号	可靠隶属度$\omega_{\delta_i}^d$		不可靠隶属度$\upsilon_{\delta_i}^d$	
	专家知识信息δ_1	仿真试验信息δ_2	专家知识信息δ_1	仿真试验信息δ_2
1	0.5	0.5	0.3	0.4
2	0.6	0.7	0.3	0.2
3	0.8	0.5	0.2	0.4
4	0.4	0.4	0.4	0.3
5	0.6	0.6	0.4	0.4

0.15, 则可设计出如表4所示的鉴定试验方案。依据表4中的鉴定试验方案结果, 在发射15枚导弹情况下, 只有当达到轻度毁伤标准的导弹数不低于14枚、达到中度毁伤标准的导弹数不低于10枚、达到重度毁伤标准的导弹数不低于6枚, 达到摧毁歼灭标准的导弹数不低于3枚时, 使用方才能认定导弹毁伤效能合格。

5.2 给定m_i^*设计鉴定试验

若给定允许最大未达到毁伤标准导弹数m_i^*, 计算生产方风险α_i与使用方风险β_i随n变化情况, 结果如图2所示。生产方风险随n的增加而增加, 使用方风险随n的增加而减少。规定生产方风险“阈值”为0.2, 使用方风险“阈值”为0.15, 采用方法2进行计算可得如表5所示的鉴定试验方案。

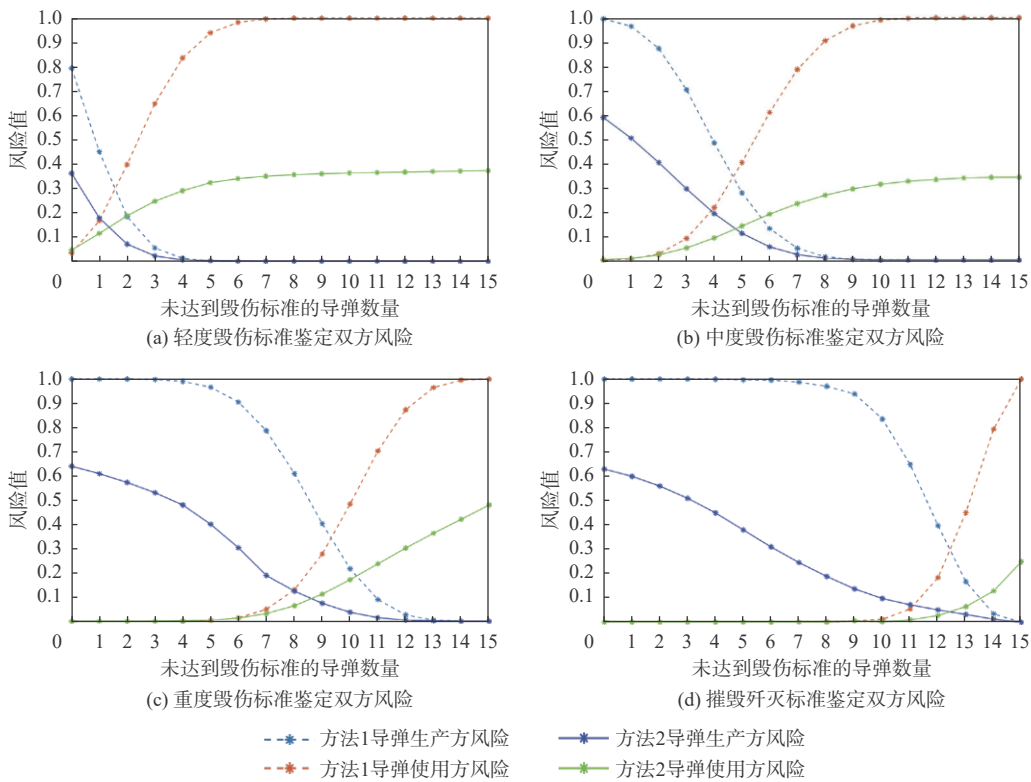


图1 给定$n=15$双方风险随m_i的变化情况

Fig. 1 Given the variation of risk of both parties in $n=15$ with respect to m_i

表4 给定$n=15$鉴定试验方案

Table 4 Given $n=15$ identification test scheme

毁伤等级标准	m_i^*取值	m_i^*最优值
轻度毁伤标准	1	1
中度毁伤标准	4,5	5
重度毁伤标准	7,8,9	9
摧毁歼灭标准	9,10,11,12,13,14	12

依据表5中的鉴定试验设计结果可得: 当在给定未达到轻度毁伤标准导弹2枚情况下, 总发射导弹数不低于21枚; 给定未达到中度毁伤标准导弹为5枚, 总发射导弹数不低于16枚; 给定未达到重度毁伤标准导弹为9枚, 总发射导弹数不低于16枚; 给定未达到摧毁歼灭标准导弹为12枚, 总发射导弹数不低于16枚。此时使用方才能认定导弹毁伤效能合格。

对比图1和图2可知, 方法2与只利用样本信息的方法1得到的鉴定试验设计结果略有不同, 表明先验信息发挥影响作用, 试验设计结果能够将先验信息进行考虑。还观察到方法2比方法1得到的鉴定试验方案双方风险值较低, 表明融入先验信息能够减少鉴定试验不确定性。并且通过划分多毁伤等级标准, 使得导弹毁伤效能得到更全面地鉴定, 有利于导弹使用方接收到毁伤效能更为可靠的导弹。

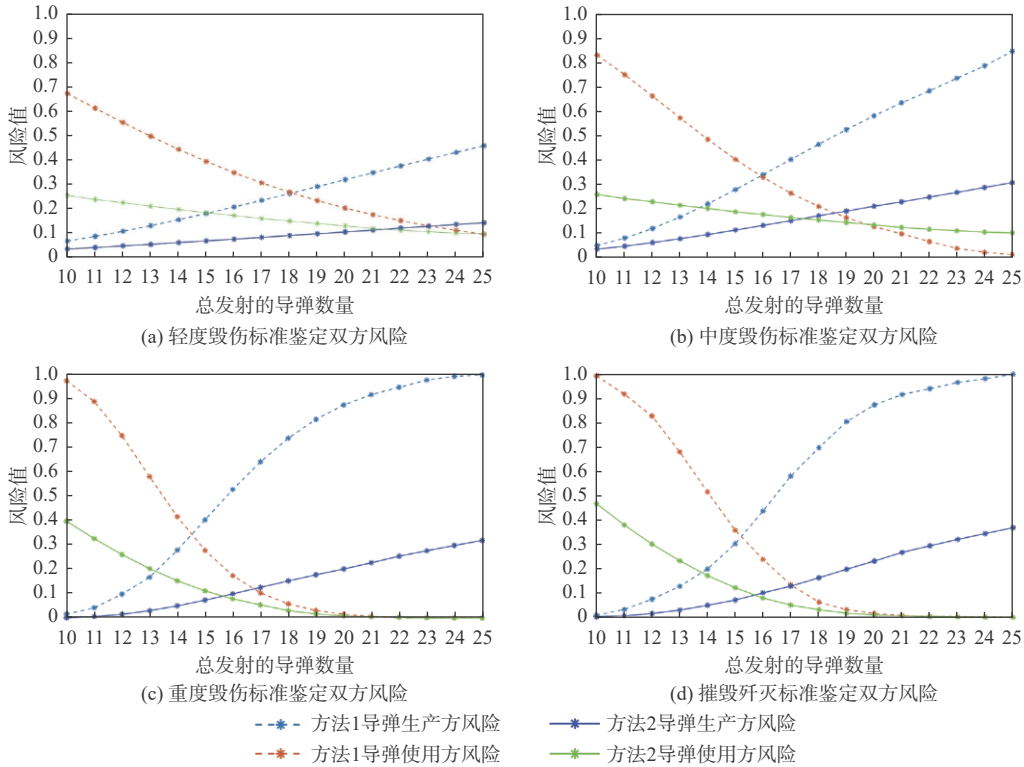


图 2　给定m_i^*双方风险随n的变化情况

Fig. 2　Given the variation of risk of both parties in m_i^* with respect to n

表 5　给定m_i^*鉴定试验方案

Table 5　Given m_i^* identification test scheme

毁伤等级标准	m_i^*	n取值	n最优值
轻度毁伤标准	2	19,20,21,22,23	21
中度毁伤标准	5	19	19
重度毁伤标准	9	15,16,17,18,19	16
摧毁歼灭标准	12	15,16,17,18	16

6　结　论

1) 本文方法考虑到导弹打击目标形成不同毁伤程度结果的实际情况, 从多毁伤等级标准角度设计鉴定试验, 相比依据单一毁伤标准的鉴定试验方法, 对导弹毁伤效能的鉴定更为全面。

2) 本文方法通过贝叶斯方法结合体系贡献度能够充分利用导弹毁伤效能的先验信息, 增加导弹毁伤效能鉴定试验的信息量, 显著降低导弹验收双方风险。

3) 本文方法有利于帮助导弹使用方对接收或拒绝导弹做出判断, 为小样本导弹毁伤效能鉴定试验研究提供借鉴。

参考文献（References）

[1] 徐豫新, 蔡子雷, 吴巍, 等. 弹药毁伤效能评估技术研究现状与发展趋势[J]. 北京理工大学学报, 2021, 41(6): 569-578.
XU Y X, CAI Z L, WU W, et al. Current research and development of ammunition damage effect assessment technology[J]. Transactions of Beijing Institute of Technology, 2021, 41(6): 569-578(in Chinese).

[2] 谷鸿平, 王玉, 舒彬, 等. 制导榴弹对典型舰船的毁伤效能研究[J]. 火力与指挥控制, 2020, 45(10): 39-42.
GU H P, WANG Y, SHU B, et al. Research on damage efficiency of guided grenade against typical ships[J]. Fire Control & Command Control, 2020, 45(10): 39-42(in Chinese).

[3] LI H S, HAO Y, ZHANG X Q. Numerical calculation method of target damage effectiveness evaluation under uncertain information of warhead fragments[J]. Mathematics, 2022, 10(10): 1688.

[4] FAN J, TANG X, WANG B H, et al. Research on the damage effectiveness of AHEAD against the LSS UAV clusters[J]. Journal of Physics: Conference Series, 2023, 2460(1): 012013.

[5] 赵晓旭, 韩旭光, 吴浩, 等. 制导杀爆弹毁伤效能评估的应用研究[J]. 北京理工大学学报, 2019, 39(6): 551-557.
ZHAO X X, HAN X G, WU H, et al. Applied research on damage effectiveness evaluation of guided explosive bomb[J]. Transactions of Beijing Institute of Technology, 2019, 39(6): 551-557(in Chinese).

[6] 朱莉娜, 李翰山. 基于 ANSYS/LS-DYNA 的博弈对抗目标毁伤效能评估[J]. 计算机应用与软件, 2021, 38(6): 66-72.
ZHU L N, LI H S. Target damage effectiveness evaluation of game confrontation based on ANSYS/LS-DYNA[J]. Computer Applications and Software, 2021, 38(6): 66-72(in Chinese).

[7] 高源, 王树山, 梁振刚, 等. 破障火箭弹终点毁伤效能评估研究[J]. 火力与指挥控制, 2020, 45(12): 87-91.
GAO Y, WANG S S, LIANG Z G, et al. Evaluation of terminal damage effectiveness of concrete-piercing rocket projectile[J]. Fire Control & Command Control, 2020, 45(12): 87-91(in Chinese).

[8] 侯鹏, 裴扬, 张睿文, 等. 地空导弹破片式打击军机的瞄准点选择

方法[J]. 北京航空航天大学学报, 2023, 49(6): 1434-1445.
HOU P, PEI Y, ZHANG R W, et al. Selection method of aim point for surface-to-air missile fragment against military aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(6): 1434-1445(in Chinese).

[9] 贠来峰, 缪云飞, 王国平, 等. 一种新的弹药产品密集度计量型抽样检验方法[J]. 弹道学报, 2020, 32(2): 50-55.
YUN L F, MIAO Y F, WANG G P, et al. A new sampling inspection method of variables for dispersion of ammunition product[J]. Journal of Ballistics, 2020, 32(2): 50-55(in Chinese).

[10] 王玮, 王军波, 黄景德. 一个可靠性二次抽样方案的分析及改进[J]. 兵工学报, 2006, 27(2): 360-362.
WANG W, WANG J B, HUANG J D. Analysis and improvement of secondary sampling plan of the reliability[J]. Acta Armamentarii, 2006, 27(2): 360-362(in Chinese).

[11] LI K, SHI X M, LI J, et al. Bayesian estimation of ammunition demand based on multinomial distribution[J]. Discrete Dynamics in Nature and Society, 2021, 2021(1): 5575335.

[12] LIU H B, SHI X M. Damage effectiveness calculation of hitting targets with ammunition based on Bayesian multinomial distribution[J]. Symmetry, 2022, 14(5): 892.

[13] 谭尧, 赵骞, 王文峰, 等. 考虑专家信息的威布尔型产品定时截尾可靠性验收试验方案设计[J]. 系统工程与电子技术, 2022, 44(4): 1409-1416.
TAN Y, ZHAO Q, WANG W F, et al. Type I censored reliability acceptence test plan for Weibull distributed products by considering expert information[J]. Systems Engineering and Electronics, 2022, 44(4): 1409-1416(in Chinese).

[14] 刘海涛, 张志华, 董理. 成败型产品可靠性的 Bayes 验收方案研究[J]. 兵工学报, 2016, 37(3): 565-569.
LIU H T, ZHANG Z H, DONG L. Bayesian acceptance test scheme for reliability of binomial products[J]. Acta Armamentarii, 2016, 37(3): 565-569(in Chinese).

[15] 解亚, 邓安仲, 范振. 基于云模型的机动式基地系统毁伤评估方法研究[J]. 军事运筹与评估, 2022, 36(3): 37-43.
XIE Y, DENG A Z, FAN Z. A study on damage assessment method of mobile base system based on cloud model[J]. Military Operations Research and Assessment, 2022, 36(3): 37-43(in Chinese).

[16] LIU H B, SHI X M, CHEN X J, et al. Bayesian inference of ammunition consumption based on normal-inverse gamma distribution[J]. Discrete Dynamics in Nature and Society, 2022, 2022(1): 6365712.

[17] ZHAO Q, JIA X, CHENG Z J, et al. Bayesian estimation of residual life for weibull-distributed components of on-orbit satellites based on multi-source information fusion[J]. Applied Sciences, 2019, 9(15): 3017.

[18] 李康, 史宪铭, 李广宁, 等. 基于正态-逆伽马分布的反巡航导弹命中概率估计方法[J]. 系统工程与电子技术, 2022, 44(8): 2621-2627.
LI K, SHI X M, LI G N, et al. Estimation method of hit probability of anti-cruise missile weapon based on normal-inverse Gamma distribution[J]. Systems Engineering and Electronics, 2022, 44(8): 2621-2627(in Chinese).

[19] BOURGUIGNON M. An alternative conjugate prior distribution for positive parameters[J]. Annals of Data Science, 2019, 6(2): 237-243.

[20] WANG H, ZHAO C Y, ZANG F, et al. Spatiotemporal patterns of precipitation based on the Bayesian maximum entropy method in a typical catchment of the Heihe River watershed, northwest China[J]. International Journal of Digital Earth, 2022, 15(1): 911-933.

[21] MAIOR C B S, MACÊDO J B, LINS I D, et al. Bayesian prior distribution based on generic data and experts' opinion: a case study in the O&G industry[J]. Journal of Petroleum Science and Engineering, 2022, 210: 109891.

[22] ZHANG X Y, ZHANG Q, ZHENG J F. Research on optimization method of weapon equipment system contribution rate evaluation index based on meta-evaluation[J]. Scientific Journal of Intelligent Systems Research, 2021, 3(11): 87-99.

[23] DU M, CHENG Z H, SHI X W. Contribution rate evaluation of army air defence brigade equipment system based on vector model[J]. Journal of Physics: Conference Series, 2020, 1650(3): 032158.

[24] 赵汝东, 史宪铭, 苏小波, 等. 基于 Bayesian 体系融合的新型弹药消耗预计方法[J]. 兵器装备工程学报, 2020, 41(2): 75-80.
ZHAO R D, SHI X M, SU X B, et al. New ammunition consumption prediction method based on Bayesian system fusion[J]. Journal of Ordnance Equipment Engineering, 2020, 41(2): 75-80(in Chinese).

[25] GU X B, MA Y, WU Q H, et al. The application of intuitionistic fuzzy set-topsis model on the level assessment of the surrounding rocks[J]. Shock and Vibration, 2022(7): 1-12.

[26] ZHAO R D, SHI X M, WANG Q, et al. Bayesian inference for ammunition demand based on Gompertz distribution[J]. Journal of Systems Engineering and Electronics, 2020, 31(3): 567-577.

Bayesian identification test design of missile damage effectiveness based on multiple damage grades standards

CHEN Tong，LIU Haobang*，HU Tao

(Department of Management Engineering and Equipment Economics，Naval University of Engineering，Wuhan 430033，China)

Abstract: Qualified missile damage effectiveness is an important guarantee to give full play to combat effectiveness. The damage effectiveness identification test research is especially important to determine whether the missile damage effectiveness is up to par. In view of the limitations of the existing damage effectiveness quantitative characterization index of "single-shot damage probability", it is difficult to fully describe the missile's damage effectiveness based on a single damage standard. The damage effectiveness is represented by the probability of occurrence of different damage grade results, taking into account the results of multiple damage grades created by a single missile striking the target. The missile damage effectiveness is identified from the perspective of multiple damage grade standards, which are helpful for a thorough examination of the missile damage effectiveness. In addition, in order to overcome the problem of insufficient information on small sample missile assembly tests, the Bayesian method is used to fuse multiple sources of prior information with the system contribution in the design of identification tests, and the test schemes with sufficient information utilization and controllable risk of both parties are designed. The example shows that compared with the quantitative characterization index of a single damage standard, this research describes the missile damage effectiveness more comprehensively, and compares it with the improved binomial distribution hypothesis test method, which verifies the superiority of this method.

Keywords: missile damage effectiveness; identification test; multiple damage grades standards; system contribution; Bayesian

Received：2023-05-25；**Accepted**：2023-08-11；**Published Online**：2023-09-11 08：49

URL：link.cnki.net/urlid/11.2625.V.20230908.0904.002

Foundation item：National Natural Science Foundation of China (71501183)

* **Corresponding author**. E-mail：D23181102@nue.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0288

融合全尺度特征与轨迹修正的遮挡车辆跟踪

郭俊锋，张子华*

(兰州理工大学 机电工程学院，兰州 730050)

摘 要：为改善车辆跟踪过程中因遮挡造成的跟踪漂移及身份切换（IDS）问题，在Deep SORT算法基础上提出一种融合全尺度特征与轨迹修正的遮挡车辆跟踪方法。引入全尺度特征提取网络提取目标不同尺度特征并实现自适应动态融合，增强目标表观特征。提出一种轨迹修正算法对遮挡过程中的跟踪轨迹进行修复，重新更新Kalman滤波参数以减小遮挡过程中累积的线性误差，优化目标运动特征。结合外观特征和运动特征实现遮挡车辆跟踪。通过消融实验与可视化分析验证所提方法的可行性，在KITTI数据集上的实验结果表明：与现有典型方法相比，所提方法取得最高综合得分为78.13%和最低的IDS次数为192，有效改善遮挡车辆跟踪中的IDS问题，提高车辆跟踪鲁棒性。

关 键 词：计算机应用；多目标跟踪；车辆跟踪；全尺度；轨迹修正

中图分类号：TP391

文献标志码：A **文章编号**：1001-5965(2025)05-1608-12

视觉多目标跟踪(multiple object tracking, MOT)是计算机视觉领域的热点问题之一[1-2]，其融合计算机视觉、模式识别、图像处理、深度学习、信号处理、数学优化及计算机应用等多个学科[3]，旨在对视频序列中感兴趣目标实现持续追踪，在自动驾驶、智慧交通、智能机器人、智能监控、人机交互、无人机侦察等领域有着广泛的应用前景及巨大的潜在价值[4]。

MOT的难点在于跟踪目标出现遮挡或姿态变换时的错误跟踪，即目标身份切换(identity switch, IDS)问题[5-7]。人类对感兴趣目标的辨别与跟踪离不开大脑皮层对目标的外观建模与运动建模，类似地，在跟踪器设计中引入外观特征与运动特征是解决IDS问题的一种有效方法，而这也正是MOT方法的常见范式，即基于检测的跟踪(tracking by detection, TBD)。TBD将MOT任务分为检测和数据关联2个单独的子任务，首先，通过前端目标检测器(如R-CNN系列、YOLO系列等)获得每个目标的目标框预测，然后，基于重识别(re-identification, ReID)或运动预测等方法，构建与目标相关的信息以实现帧间匹配，从而形成轨迹[3]。得益于深度学习的快速发展，各种高性能检测器不断更新，因此，TBD方法在MOT算法中处于“统治”地位。SORT[8]算法是典型的TBD目标跟踪法方法，通过Kalman滤波器(Kalman filter, KF)和匈牙利算法分别处理跟踪问题中的运动预测和数据关联，在跟踪速度和准确性方面取得了当时最优。SORT算法提供了一种简单高效的跟踪范式，后续学者基于此范式做出大量研究。Wojke等[9]在SORT框架中引入深度外观特征提出Deep SORT算法，有效改善遮挡情况下的IDS问题。Du等[10]通过为Deep SORT跟踪器配备一些高级组件提出Strong SORT，并且针对MOT中的缺失关联和缺失检测2个固有“缺失”问题，提出一种无外观关联模型AFLink和基于高斯回归过程的高斯平滑插值GSI，AFLink和GSI均为轻量级算法，可以很容易地插入到各种跟踪器中并改善其跟踪性能。Cao等[11]提出一种在线实时的跟踪算法OC-SORT，Maggiolino等[12]在其基础上结合目标外

收稿日期：2023-05-29；**录用日期**：2023-07-28；**网络出版时间**：2023-09-01 14:28
网络出版地址：link.cnki.net/urlid/11.2625.V.20230901.1138.003
基金项目：国家自然科学基金(51465034)
*通信作者. E-mail：zzh15536308429@163.com

引用格式：郭俊锋，张子华. 融合全尺度特征与轨迹修正的遮挡车辆跟踪[J]. 北京航空航天大学学报，2025，51(5)：1608-1619.
GUO J F，ZHANG Z H. Track obscured vehicles by fusing full-scale features with trajectory correction[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1608-1619 (in Chinese).

观信息提出 Deep OC-SORT, 通过自适应加权的外观特征与运动特征来提高跟踪鲁棒性。Aharon 等[13]在 SORT 算法基础上直接估计 KF 状态向量中目标边界框的宽高值, 并结合运动相机补偿, 将外观特征集成到跟踪器中, 提出一种新的鲁棒跟踪器 BOT-SORT, 可以结合外观和运动信息, 具有领先的跟踪性能。此外, Seidenschwarz 等[14]基于传统 TBD 方法设计了一个简单的匈牙利跟踪器, 通过引入新的 spiced-up 外观模型以不同方式处理目标的活动和非活动轨迹, 并应用动态域适应对外观特征和运动特征自适应加权, 在多种数据集上实现了领先性能。Yang 等[15]提出一种级联缓冲交并比(cascaded-buffered intersection over union, C-BIOU)跟踪器, 通过添加缓冲区来扩展检测和跟踪的匹配空间, 并采用级联匹配减小匹配空间过度扩展的风险, 对具有不可区分外观和不规则运动的多个目标对象跟踪效果优于现有方法。

目前 MOT 的研究对象大约 70% 都为行人, 而对车辆目标跟踪的研究较少, 车辆跟踪是智慧交通领域的重要内容之一, 其要求道路摄像头能够对过往车辆实现实时准确地跟踪, 但面对日益复杂的交通场景, 现有跟踪算法在目标被遮挡时效果不佳, 容易出现 IDS 现象, 导致跟踪断裂。因此, 对遮挡车辆的重识别跟踪研究具有重要意义。本文基于 Deep SORT 算法, 提出一种结合全尺度特征与轨迹修正的遮挡车辆跟踪方法。首先, 通过全尺度特征提取网络(full-scale feature extraction network, FSNet)自适应实现多尺度特征融合, 增强目标表观特征, 然后, 通过轨迹修正算法(trajectory correction algorithm, TCA)修复遮挡过程中的 KF 参数线性累积误差, 最后, 融合外观特征与运动特征完成目标跟踪, 提高遮挡车辆跟踪准确性。

1　Deep SORT 算法

Deep SORT 算法是典型的 TBD 目标跟踪方法之一, 其在 SORT 算法基础上将基于深度特征向量最小余弦距离的表观信息和基于马氏距离的运动信息加权融合, 从而形成新的度量匹配信息[16], 有效改善目标被遮挡时的处理能力, 减少 IDS 现象。与众多视觉 MOT 方法相同, Deep SORT 算法在处理目标点迹与航迹的关联过程中也充斥着航迹关联的思想, 其核心为 KF 和匈牙利匹配算法。

1.1　Kalman 滤波器

KF 是一种基于估计算法的递归状态空间模型, 利用前一帧所检测到的目标信息来预测下一帧中对应目标的运动状态, 并基于观测值进行修正。Kalman 滤波算法主要分为预测和更新 2 个过程, 故也被称为 Predictor-Corrector 算法。当目标检测器检测到第 1 帧后, 初始化 KF。设系统状态向量 $\boldsymbol{X}=\left(c_x,c_y,r,q,\boldsymbol{v}_x,\boldsymbol{v}_y,\boldsymbol{v}_r,\boldsymbol{v}_q\right)$, 其中, c_x、c_y 分别为检测框中心点的横、纵坐标, r 为检测框宽高比, q为框的高度, $\boldsymbol{v}_x$、$\boldsymbol{v}_y$、$\boldsymbol{v}_r$、$\boldsymbol{v}_q$分别为 c_x、c_y、r、q相对应的速度分量,假设系统为线性系统, 且噪声分布满足高斯白噪声, 则系统的状态方程和观测方程分别为

$$\boldsymbol{x}_t=\boldsymbol{A}\boldsymbol{x}_{t-1}+\boldsymbol{B}\boldsymbol{u}_t+\boldsymbol{w}_t \tag{1}$$

$$\boldsymbol{y}_t=\boldsymbol{C}\boldsymbol{x}_t+\boldsymbol{v}_t \tag{2}$$

式中: $\boldsymbol{x}_t$ 和 $\boldsymbol{y}_t$ 分别为 t 时刻的系统状态矩阵和实际观测状态; $\boldsymbol{u}_t$ 为系统控制矩阵; $\boldsymbol{A}$、$\boldsymbol{B}$ 和 $\boldsymbol{C}$ 分别为系统的状态转移矩阵、输入控制矩阵和状态观测矩阵; $\boldsymbol{w}_t$ 和 $\boldsymbol{v}_t$ 分别为 t 时刻的系统噪声和观测噪声。

1.1.1　预测过程

利用系统 t–1 时刻的状态预测 t 时刻的状态, 系统状态预测方程和协方差预测方程分别为

$$\boldsymbol{x}_{t|t-1}=\boldsymbol{A}\boldsymbol{x}_{t-1|t-1}+\boldsymbol{B}\boldsymbol{u}_t \tag{3}$$

$$\boldsymbol{P}_{t|t-1}=\boldsymbol{A}\boldsymbol{P}_{t-1|t-1}\boldsymbol{A}^{\mathrm{T}}+\boldsymbol{Q} \tag{4}$$

式中: $\boldsymbol{x}_{t-1|t-1}$ 和 $\boldsymbol{P}_{t-1|t-1}$为 t–1 时刻的准确状态和协方差矩阵; $\boldsymbol{x}_{t|t-1}$ 和 $\boldsymbol{P}_{t|t-1}$ 为 t 时刻状态的预测值和预测协方差; $\boldsymbol{Q}$ 为预测过程噪音矩阵, 满足 $\boldsymbol{p}(\boldsymbol{w})\in N(\boldsymbol{0},\boldsymbol{Q})$。

1.1.2　更新过程

结合当前时刻系统的观测值, 利用 Kalman 增益来修正预测值, 使其逼近真实值, 得到最优估计。系统的状态更新方程和协方差更新方程为

$$\boldsymbol{x}_{t|t}=\boldsymbol{x}_{t-1|t-1}+\boldsymbol{K}_t(\boldsymbol{y}_t-\boldsymbol{C}\boldsymbol{x}_{t|t-1}) \tag{5}$$

$$\boldsymbol{P}_{t|t}=(\boldsymbol{I}-\boldsymbol{K}_t\boldsymbol{C})\boldsymbol{P}_{t|t-1} \tag{6}$$

式中: $\boldsymbol{x}_{t|t}$ 为更新后当前时刻的最优值; $\boldsymbol{P}_{t|t}$ 为更新后的协方差矩阵; $\boldsymbol{K}_t$为 t 时刻的 Kalman 增益, 相当于一个权重项, 决定估计值和观测值之间的权重问题, 其计算式为

$$\boldsymbol{K}_t=\boldsymbol{P}_{t|t-1}\boldsymbol{C}^{\mathrm{T}}\left(\boldsymbol{C}\boldsymbol{P}_{t|t-1}\boldsymbol{C}^{\mathrm{T}}+\boldsymbol{R}\right)^{-1} \tag{7}$$

其中: $\boldsymbol{R}$ 为观测过程噪音矩阵, 满足 $\boldsymbol{p}(\boldsymbol{v})\in N(\boldsymbol{0},\boldsymbol{R})$。

图 1 为车辆跟踪过程中的 Kalman 滤波示例, 首先, 基于前一时刻的估计状态 $\boldsymbol{x}_{t-1|t-1}$ 得到当前时刻状态的预测值 $\boldsymbol{x}_{t|t-1}$, 并结合实际观测值 $\boldsymbol{y}_t$, 通过 Kalman 增益对估计值和观测值加权平均得到该时刻的最优值, Kalman 滤波本质上为一个优化估计算法。

1.2　匈牙利匹配

Deep SORT 算法中通过匈牙利匹配将上一帧的跟踪框与当前帧的检测框关联, 融合表观特征和运动特征构造代价矩阵, 提高跟踪准确度。

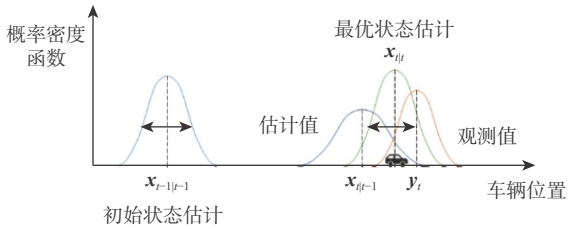
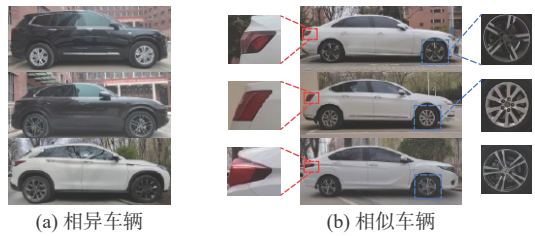


图 1 Kalman 滤波示例

Fig. 1 Kalman filter example



(a) 相异车辆 (b) 相似车辆

图 2 不同车辆外观特征

Fig. 2 Different vehicle appearance features

1.2.1 运动特征

利用当前帧的检测框与上一帧跟踪框之间的马氏距离计算目标位置维度的距离相似度来评价两者的运动匹配程度，其计算式为

$$d_1(i,j)=\left(\boldsymbol{Z}_j-\boldsymbol{M}_i\right)^{\mathrm{T}}\boldsymbol{P}_i^{-1}\left(\boldsymbol{Z}_j-\boldsymbol{M}_i\right) \tag{8}$$

式中：$d_1(i,j)$为跟踪框 i 与检测框 j 之间的马氏距离；$\boldsymbol{P}_i^{-1}$为跟踪框 i 的协方差预测矩阵；$\boldsymbol{Z}_j$为第 j 个检测框的观测向量；$\boldsymbol{M}_i$为第 i 个跟踪框的状态向量。$d_1(i,j)$越小，表示跟踪框和检测框之间的运动特征相似度越高，两者匹配关联度越大。

1.2.2 表观特征

利用跟踪框和检测框之间的最小余弦距离来评价二者的表观匹配程度，计算式为

$$d_2(i,j)=1-\boldsymbol{r}_j\boldsymbol{v}_i\left(\left|\boldsymbol{r}_j\right|\left|\boldsymbol{v}_i\right|\right)^{-1} \tag{9}$$

式中：$\boldsymbol{r}_j$和$\boldsymbol{v}_i$分别表示当前帧检测框 j 和上一帧跟踪框 i 的特征向量。$d_2(i,j)$越小，表示跟踪框和检测框之间的表观特征相似度越高，两者匹配关联度越大。

最后，将运动特征相似度距离和表观特征相似度距离通过超参数 λ 融合，得到最终的综合特征度量值，如式(10)所示，将其作为匈牙利算法的代价矩阵，从而完成确认轨迹与检测响应之间的数据关联。

$$D(i,j)=\lambda d_1(i,j)+(1-\lambda)d_2(i,j) \tag{10}$$

2 融合全尺度特征与轨迹修正的遮挡车辆跟踪

为实现遮挡过后对目标车辆的正确重识别，本文提出一种融合全尺度特征与轨迹修正的遮挡车辆跟踪方法，首先，在 Deep SORT 算法中引入 FSNet，以自适应地学习车辆目标的局部特征与全局特征；然后，提出 TCA，减少 KF 在目标遮挡过程中的线性积累误差，并基于修正后的轨迹重新更新滤波器参数；最后，结合 FSNet 学习到的全尺度外观特征与修正后的运动特征实现遮挡车辆跟踪。

2.1 全尺度特征提取网络

图 2 为不同车辆外观特征，其中，不同品牌车辆具有不同设计语言，通过全局特征即可轻松分辨，而部分同颜色同车型的车辆之间具有极大的外观相似度，则需关注局部特征才能区分。但局部特征必须以全局特征作为参照，因此，如何自适应地提取目标不同尺度外观特征并通过动态融合学习多维特征是提高车辆重识别能力的有效途径。

本文引入 FSNet，如图 3 所示，图中，R_e 表示感受野的大小。首先，通过不同分支提取不同尺度的特征，每个分支关注的特征尺度由维度指数决定，然后，将生成的多个尺度特征图输入到统一的聚合门(aggregation gate, AG)网络，AG 在所有分支之间共享参数，并基于输入图像为每个分支动态分配权重，自适应地捕获单个尺度的判别特征或多个尺度特征的加权融合，最终实现多尺度特征融合，并输出保存有目标特征的一维特征向量，以用于外观相似度匹配计算。此外，每个分支都采用深度可分离卷积(depthwise separable convolution, DBC)减少参数量，维持模型轻量化的同时避免过度拟合。

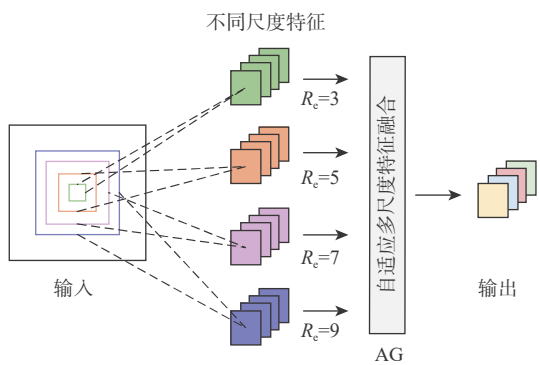


图 3 FSNet 示意图

Fig. 3 Schematic diagram of FSNet

2.1.1 多尺度特征提取

为减少参数量，采用深度可分离卷积代替普通卷积，其核心为逐通道卷积和逐点卷积。图 4(a)为 FSNet 的下采样块，对于大小为$h\times w\times c$的输入图像，经深度可分离卷积采样后，计算成本从$h\times w\times k^2\times c\times c'$降低到$h\times w\times(k^2+c)\times c'$，参数量从$k^2\times c\times c'$减少到$(k^2+c)\times c'$。其中，$h$、$w$、$c$ 分别为输入特征图的高、宽及通道数，k 为卷积核大小，c'为输出通道数。

图 4(b)为 FSNet 中的基础 Bottleneck，其采用残差结构。对于输入φ，经 Bottleneck 后的输出为

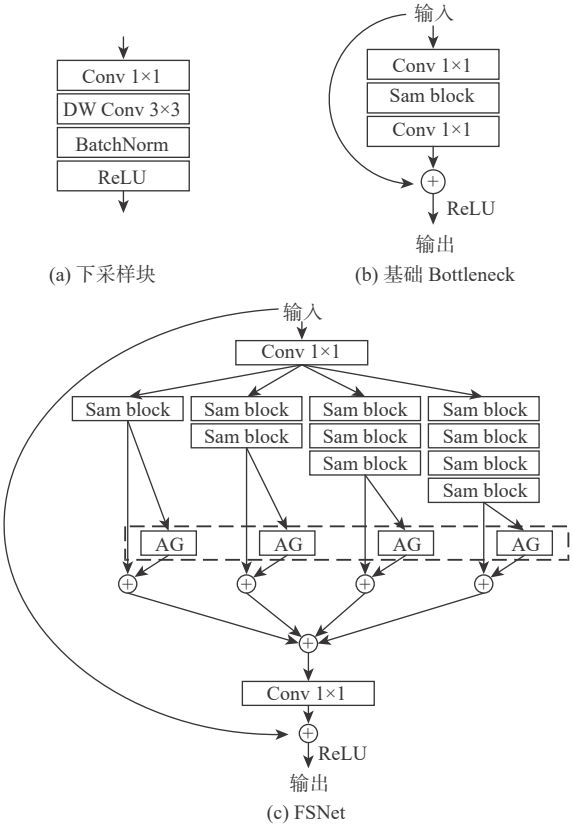


图 4　全尺度特征提取网络结构

Fig. 4　FSNet structure

$$\begin{cases}\phi=\varphi+\varphi'\\ \varphi'=F(\varphi)\end{cases} \tag{11}$$

式中: $F(\varphi)$为学习到的单尺度特征(感受野为 3×3), 首尾两端的 1×1 卷积用于操纵特征维度, 不参与聚合空间信息。

为实现多尺度特征学习, 引入一个新的维度指数 n 来控制特征的尺度, 当$n>1$时, 在不同的并行分支中堆叠 n 个 Sam block 以学习不同尺度的特征, 如图 4(c)所示。此时每个分支的感受野大小为$(2n+1)\times(2n+1)$, 此时学习到的总特征φ'为 N 个增量尺度的总和, 即

$$\varphi'=\sum_{n=1}^{N}F^n(\varphi)\qquad N\geqslant 1 \tag{12}$$

式中: $F^n(\varphi)$为学习到的第 n 个分支的尺度特征。当 N=1 时, 式(12)简化为式(11), 本文设置 N=4, 即分支最大感受野为 9×9, 其中, shortcut 操作允许网络在下一层中保留当前层中学习到的较小规模的特征, 从而使最终特征能够捕捉整个空间尺度。

2.1.2　多尺度特征融合

FSNet 的每个分支都提供特定尺度的特征, 为综合学习多尺度特征, 通过 AG 动态组合不同分支的输出, 即根据输入图像为每个尺度分支自适应分配不同的权重, 从而实现动态尺度融合。此时网络学习到的特征φ'可表示为

$$\begin{cases}\varphi'=\sum_{n=1}^{N}G(\varphi^n)\odot\varphi^n\\ \varphi^n=F^n(\varphi)\end{cases} \tag{13}$$

式中: $G(\varphi^n)$为跨越φ^n的整个通道维度的向量, 用于生成权重项; ⊙表示 Hadamard 乘积。

至此, 对于给定输入图像, 经过 FSNet 不同分支学习不同尺度特征后, 通过 AG 自适应完成动态尺度融合, 从而实现全尺度特征的学习。

2.1.3　全尺度特征提取网络架构

FSNet 的架构如表 1 所示, 网络输入为 256×128 的目标图像数据, 经 FSNet 提取各尺度特征并融合后, 通过全局平均池化层将特征图中的高维度信息降低到一维特征向量中, 以降低数据维度, 减少计算和储存的复杂性。因特征图中的信息被均匀考虑, 因此, 该特征向量也能很好地表达目标特征。

表 1　全尺度特征提取网络架构

Table 1　FSNet architecture

层	输出($h\times w\times c$)	操作
conv1	128×64×64	7×7 conv, stride 2
	64×32×64	3×3 max pooling, stride 2
conv2	64×32×256	bottleneck×2
transition	64×32×256	1×1 conv
	32×16×256	2×2 average pooling, stride 2
conv3	32×16×384	bottleneck×2
transition	32×16×384	1×1 conv
	16×8×384	2×2 average pooling, stride 2
conv4	16×8×512	bottleneck×2
conv5	16×8×512	1×1 conv
gap	1×1×512	global average pooling

2.2　轨迹修正算法

针对 KF 线性运动假设的局限性, 目前已有扩展 Kalman 滤波(extend Kalman filter, EKF)、无迹 Kalman 滤波(unscented Kalman filter, UKF)等相关变体及粒子滤波等方法, 在处理非线性运动时取得良好效果。但 EKF 和 UKF 利用低阶泰勒展开项近似处理非线性运动, 其仍然依赖于近似 KF 假设的高斯先验, 而粒子滤波利用粒子表示概率, 通过基于采样的后验估计解决非线性运动则会引入指数级的计算量, 导致算法复杂度增加, 计算效率低下。现有视觉传感器捕捉视频基本为高帧率, 视频帧之间时间间隔很小, 即使物体运动是全局非线性的, 也可以在短时间内很好地近似为线性, 因此, Kalman 滤波在大部分时间内都能很好地工作, 且其本身可通过实际观测值对参数进行更新。但当目标出现遮挡情况时, 没有观测值可用于 KF 更新, 导致多个时间步长的线性误差积累被放大, 因此, 本文提出 TCA, 构造遮挡过程中的虚拟轨迹, 并通过虚拟轨迹重新更新 KF 参数, 减小遮挡过程中的线

性误差影响，而无遮挡时 KF 参数仍结合观测值更新，在提高算法跟踪效果的同时尽可能地降低计算成本。

假设目标在 t_1–t_2 时刻被遮挡，如图 5 所示，遮挡之前（t_1 之前），目标位置由 KF 预测值与实际观测值加权确定，观测值通过 Kalman 增益修正预测值，使其最接近真实值。当目标被遮挡后（t_1–t_2），其位置完全由 KF 预测值确定，而 Kalman 滤波又基于线性运动假设，导致 KF 参数在此过程中积累大量线性误差。t_2 时刻目标重新出现并关联，此时 Kalman 滤波根据 t_2 时刻的实际观测值与由 t_2–1 时刻确定的 KF 预测值来确定目标位置，而 t_2–1 时刻距 t_1 时刻已经经过多个时间步长，其线性误差已经很大，导致最终确定的目标位置与实际位置之间有着较大偏差，从而使预测轨迹与目标实际运动轨迹之间有着较大的方向差异角，增加后续跟踪难度。

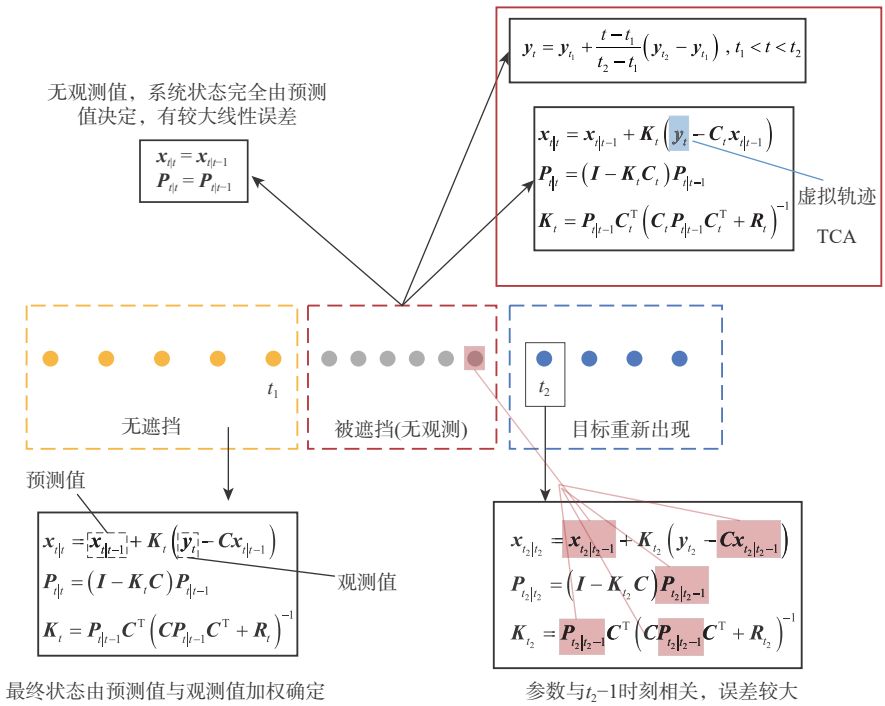


图 5 目标不同时刻参数更新

Fig. 5 Target parameter update at different moments

本文提出的 TCA 通过构造遮挡过程中的虚拟轨迹，并通过虚拟轨迹重新更新 KF 参数，以减小遮挡过程中的线性误差，提高目标跟踪成功率，其伪代码如下所示。如果目标轨迹中有未跟踪到的时刻（t_1–t_2），TCA 会找到目标历史轨迹中最后有观测值的时刻 t_1，并将 KF 参数退回到 t_1 时刻，然后通过式（14）构造遮挡过程中每个 t 时刻的虚拟轨迹，将构造的虚拟轨迹添加到目标的历史轨迹中，并根据新的轨迹重新更新 KF 参数，从而减小线性误差影响，进而减小预测轨迹与真实轨迹之间的方向差异角，提高跟踪成功率。

$$\boldsymbol{y}_t = \boldsymbol{y}_{t_1} + \frac{t-t_1}{t_2-t_1}\left(\boldsymbol{y}_{t_2} - \boldsymbol{y}_{t_1}\right) \qquad t_1 < t < t_2 \tag{14}$$

图 6 为 TCA 对遮挡过程中的轨迹修正示意图，图中带阴影的圆形区域表示预测值和观测值可以匹配的范围。t_1 时刻目标开始被遮挡，t_2 时刻遮挡结束，目标重新出现并重新关联。在原始轨迹图 6(a) 中，由于 KF 的线性假设与目标实际运动不符，预测轨迹与实际轨迹之间存在较大方向差异角 α，导致目标在 t_3 时刻跟踪丢失。

而在修正后的轨迹图 6(b) 中，通过对 KF 参数的重更新，修复 KF 在遮挡过程中的线性累积误差，从而减小方向差异角为 α'，目标处于可匹配范围内，跟踪器在 t_3 时刻仍能正确跟踪。

算法 TCA 伪代码

输入 轨迹集合 $\Gamma = \{\Gamma_t \mid 1 \leqslant t \leqslant T\}$

遍历轨迹工中每帧轨迹 τ:

若 τ 未跟踪到目标 (无观测值):

#执行 TCA 构造虚拟轨迹 #

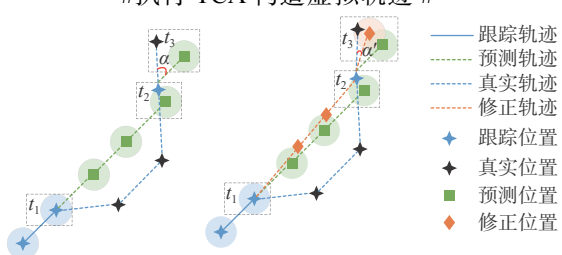


图 6 轨迹修正示意图

Fig. 6 Schematic diagram of trajectory correction

找到轨迹中最后一次观测值z_t^r及其对应的时刻 t

将 KF 参数退回到时刻 t

通过式 (14) 构造遮挡过程虚拟轨迹Z_t^r

沿虚拟轨迹Z_t^r在线平滑 KF 参数

结束

将虚拟轨迹Z_t^r添加到目标历史轨迹中

基于新轨迹重新更新 KF 参数

结束

输出　新的轨迹Γ

2.3　遮挡车辆跟踪流程

融合 FSNet 与轨迹修正的遮挡车辆跟踪流程如图 7 所示, 集合 S 为保存每帧目标轨迹信息的轨迹集合, 假设输入视频共 n 帧, 目标在第 k 帧时被遮挡, 第$k+t$帧时遮挡结束, 目标重新出现。整个网络的跟踪过程如下。

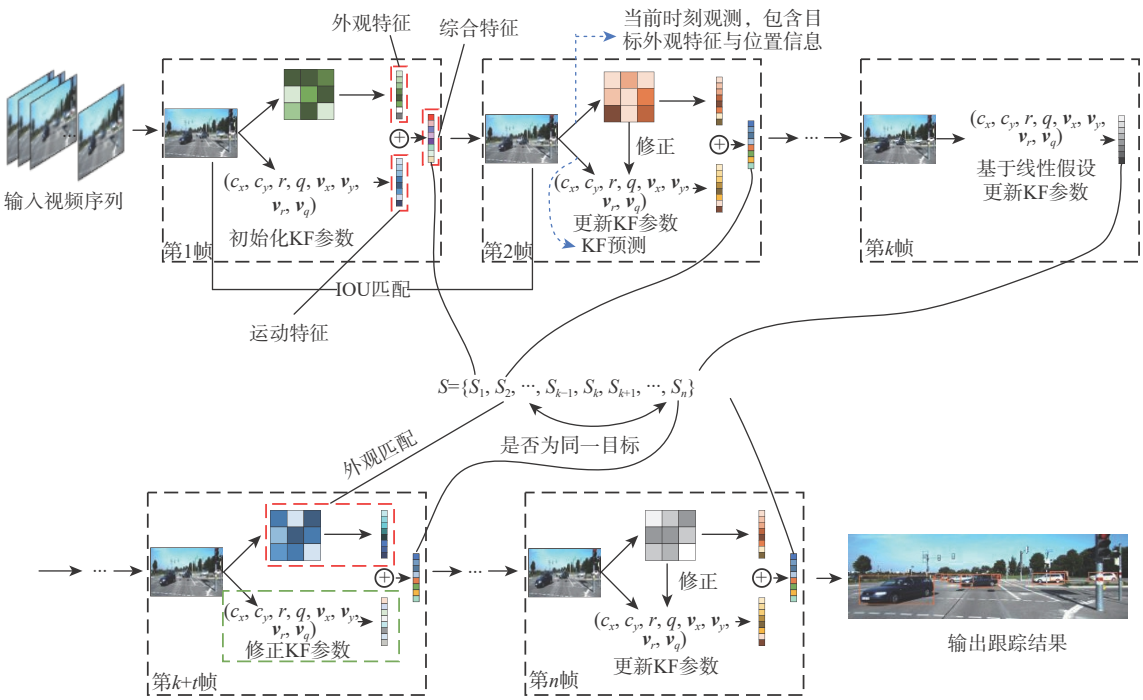


图 7　遮挡车辆跟踪流程

Fig. 7　Obscured vehicle tracking process

第 1 帧: 对于初始视频输入, 目标检测器检测到第 1 帧后, 通过 FSNet 提取目标表观特征, 并初始化 KF, 然后, 将表观特征和运动特征融合后的综合特征保存到第 1 帧对应的轨迹S_1中。

第 2 帧: 将目标检测器获得的检测框与前一帧中的跟踪框做 IOU 匹配, 以确定是否为同一目标, 若为同一目标则继续跟踪, 基于前一帧的信息并结合观测值对 KF 参数进行更新, 将综合特征保存到第 2 帧对应的轨迹S_2中。

⋮

第 k 帧: 此时目标被遮挡, 无法通过观测值修正 KF 参数, KF 参数只能基于线性假设更新, 估计误差开始累积。

⋮

第$k+t$帧: 经过 t 帧遮挡后, 目标重新出现。首先, 将当前帧的综合特征与目标消失前的轨迹$(S_1,S_2,\cdots,S_{k-1})$中保存的综合特征做匹配, 匹配过程遵循优先缺失帧数最少的目标先匹配的原则, 以达到跟踪目标的重识别。目标与轨迹重新关联后, 通过 TCA 对遮挡过程中的轨迹进行修复, 并更新 KF 参数。

⋮

第 n 帧: 完成目标跟踪, 输出跟踪结果。

3　实验与分析

3.1　数据集与实验环境

本文采用 VeRi-776 数据集对 FSNet 进行训练, 该数据集是由华中科技大学研究团队创建的针对车辆重识别任务的大规模数据集, 包含来自 20 个摄像头视角的共 776 辆车的视频序列, 覆盖了城市街道、高速公路、停车场等多个场景, 其中, 每辆车都被多个不同摄像头角度拍摄, 从而提供车辆在不同视角下的多个实例。数据集中有来自 16 678 个视频帧的 37 778 张图像用于训练, 来自 11 579 个视频帧的 27 679 张图像用于测试, 其中, 每张图像都经过精确的实例级标注, 提供了车辆的边界框及车型、颜色等身份信息, 使得车辆重识别任务能够以监督学习的方式进行。

网络的训练过程遵循标准的分类范式，每个目标的ID都被视为唯一一类。预处理阶段将输入图像尺寸调整为256×128，并通过随机翻转、剪裁等数据增强方式提升网络的学习能力。训练阶段，网络获取输入图像的一维特征向量，同时记录每个图像样本的标签信息，将特征向量与标签信息相关联后储存在一个字典中，字典的键为图像索引，字典的值为包含特征向量和标签信息的元组。采用对比损失函数Triplet Loss帮助网络学习区分不同目标的特征表示，通过最大化同一身份图像对的相似度，同时最小化不同身份图像对的相似度来训练网络，其基于图像三元组的概念，数学表达式定义为

$$L(a,p,m)=\max(0,\|\boldsymbol{f}(a)-\boldsymbol{f}(p)\|-\|\boldsymbol{f}(a)-\boldsymbol{f}(m)\|-m_{\mathrm{ar}})$$

式中：a为锚点图像；p为同一身份的正样本图像；m为不同身份的负样本图像；$\boldsymbol{f}(\cdot)$为网络提取的特征向量；$\|\cdot\|$表示向量的范数；m_{ar}为一个预定于的边界值，用于控制正样本和负样本之间的距离差异。通过最小化Triplet Loss，网络被迫将同一身份的图像特征表示更加接近，并将不同身份的图像特征表示分散开来，从而学习到最合适的特征表示。

此外训练过程使用SGD优化器，并采用余弦退火策略动态调整学习率。具体训练参数如表2所示。

表2 全尺度特征提取网络训练参数

Table 2 FSNet training parameters

训练轮次	批处理大小	初始学习率	权重衰减系数
350	64	0.065	0.000 5

采用KITTI数据集作为算法测试数据来源，该数据集由德国卡尔斯鲁厄理工学院和丰田美国技术研究院联合创办，是目前国际上最大的计算机视觉算法测评数据集。KITTI数据集通过车载摄像头、激光雷达等一系列传感器，采集市区、乡村、学校和高速公路等场景的真实图像数据，广泛应用于目标检测、目标跟踪等计算机视觉任务。本文选择其中的二维目标跟踪评估数据集，对象跟踪基准由21个训练序列和29个测试序列组成，包含“Car”和“Pedestrian”2个类别，为方便评估，本文只关注“Car”类别，并且只考虑图像中大于25像素的检测物体。

本实验使用Window11操作系统，处理器为Intel(R) Core(TM) i7-8750H CPU @ 2.20 GHz，使用NVIDIA GeForce GTX 1050Ti显卡进行运算，显存大小为4 GB，所用软件为PyCharm 2021.1.1 x64版本，软件环境为Pytorch，版本为1.12.1，Python语言环境为3.8.13。

3.2 评价指标

为定量分析本文方法对遮挡车辆目标的跟踪性能，选用MOT任务中公认典型评价指标多目标跟踪精度（multiple object tracking accuracy, MOTA）、关联精度（association accuracy, ASSA）、更高级别跟踪精度（higher order tracking accuracy, HOTA）作为主要评价指标；采用IDS数评价跟踪轨迹的连贯性；采用每秒传输帧数（frames per second, FPS）评价跟踪器处理速度。其中，MOTA为检测指标，表示跟踪器对车辆目标的识别能力；ASSA为关联指标，表示跟踪器对同一目标的匹配能力；HOTA对检测和关联进行同等加权，是表示跟踪器跟踪性能的综合指标。

3.3 定量分析

表3为验证本文方法可行性的消融实验研究，模型A为原始Deep SORT算法，模型B在Deep SORT算法中引入FSNet网络，模型C引入TCA算法，模型D为本文方法。通过在KITTI数据集上消融实验证明，本文所提2种方法FSNet和TCA均有助于提高跟踪器的跟踪性能。具体来看，对比模型A与模型B，检测得分提升，说明引入FSNet后，跟踪器对目标外观特征的提取能力加强；对比模型A与模型C，关联得分提升，说明引入TCA后，跟踪器对目标运动轨迹的表达能力加强；模型D同时引入FSNet和TCA，因此，具有最高的综合得分，同时也具有最好的跟踪性能。

表3 消融实验

Table 3 Ablation experimental

模型	FSNet	TCA	MOTA/%	ASSA/%	HOTA/%
A			87.37	73.59	72.18
B	√		89.33	73.77	75.42
C		√	87.31	77.89	75.11
D	√	√	90.81	78.27	78.13

表4为本文方法在KITTI数据集上的测试结果，并与目前先进MOT方法做对比，为保证结果对比的公平性，表内算法均采用相同的目标检测器YOLOX及相同的检测器权重。表内结果显示，在KITTI 2D目标跟踪数据集上，PermaTrack算法具有最高的检测得分，但总体来看，表内算法的检测性能差别不大，StrongSORT和BOT-SORT算法具有较好关联性能，因此，取得了较高的综合得分；本文方法的关联得分和综合得分最高分别为78.27%和78.13%，同时目标IDS次数最少为192，取得优于其他算法的跟踪性能，也印证出对同一目标的正确跟踪是提高算法跟踪性能的关键。Deep SORT算法取得最快的跟踪速度，但跟踪精度较差。本实验环境有限，在更好的硬件平台上，本文方法可以取得更快的跟踪速度。

表 4　不同算法在 KITTI 数据集测试结果

Table 4　Test results of different algorithms on KITTI dataset

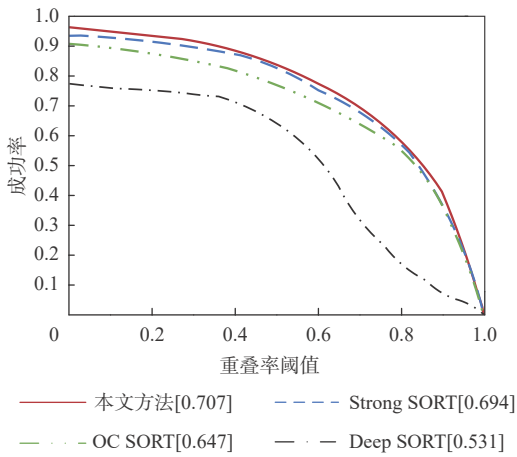
算法	MOTA/%	ASSA/%	HOTA/%	IDS次数	FPS
Deep SORT	87.37	73.59	74.02	334	34.51
PermaTrack	90.85	77.66	77.42	275	11.86
OC SORT	90.28	76.39	76.54	250	29.04
StrongSORT	90.35	78.20	77.75	240	7.09
ByteTrack	90.43	76.14	76.60	361	29.63
BOT-SORT	90.72	78.19	77.95	248	6.58
本文方法	90.81	78.27	78.13	192	17.84

为更直观地验证本文方法对遮挡目标的跟踪性能, 引入成功率曲线下面积(area under success rate curve, AUC)这一评价指标表示跟踪成功率, 面积越大跟踪效果越好, 选取数据集中具有遮挡属性的序列进行测试, 结果如图 8 所示。其中, 重叠率定义为

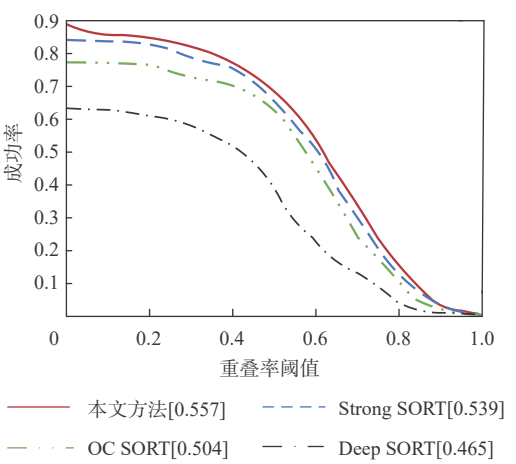
$$O_{\text{rate}} = \frac{\delta_1 \cap \delta_0}{\delta_1 \cup \delta_0} \tag{15}$$

式中: δ_1为算法跟踪框覆盖区域; δ_0为数据集中目标真实覆盖区域; O_{rate}越大表示重叠率越高。

图 8(a)为整个数据集下的测试结果, 本文方法成功率达到 0.707, 相较于 Strong SORT 算法提升了 1.3%。图 8(b)为在具有遮挡属性序列中的测试结果, 本文方法成功率提升了 1.8%, 表明本文方法在目标受到遮挡情况时能改善跟踪效果。综合来看, 本文方法具有更高的跟踪成功率及鲁棒性。



(a) 整体数据集下的成功率



(b) 遮挡属性下的成功率

图 8　不同算法的成功率曲线

Fig. 8　Success rate curves of different algorithms

3.4　可视化分析

图 9 为不同输入下的车辆特征注意力热图, 可直观地表达网络对目标特征的关注程度, 其中, 颜色越深表示该处特征关注度越高。图 9(b)为以同车型不同颜色的车辆 1 与车辆 2 作为输入, 图 9(c)以不同车型不同颜色的车辆 3 与车辆 4 作为输入, 从实验结果看出, 在面对外观相似度低的目标作为输入时, 网络关注的为目标全局特征。图 9(d)和图 9(e)的输入分别为车辆 1 和车辆 3、车辆 4 和车辆 5, 此时输入为同颜色同车型车辆目标, 相互之间有着较大的外观相似度, 从实验结果来看, 网络关注的为车辆轮胎处的局部特征, 原因为上述车辆有着不同样式的轮毂, 因此, 通过轮毂处的局部特征区分。而对比图 9(b)与图 9(d)、图 9(c)与图 9(e)可得出, 即使是同一车辆, 在面对不同的目标对比时网络关注的特征尺度不同, 说明本文方法可自适应地关注目标不同尺度特征, 以达到良好的目标区分能力。

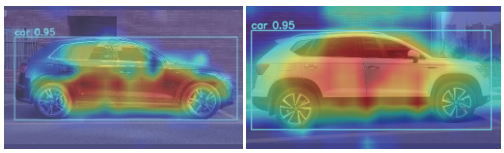
图 10 为 KITTI 数据集中的场景重绘车辆特征热力图, 结果显示: 本文方法对场景 1 和场景 2 的关注区域集中在目标附近且关注度更高, 能有效提高对目标的识别概率; 在场景 3 和场景 4 中, 本文方法虽然提高了对目标的关注效果, 但热力图中显示对背景的关注度也提高, 可能会对场景中的目标识别造成一定影响, 但综合来看, 本文方法能够取得更高的目标特征关注度。

图 11 为不同算法对车辆目标的跟踪轨迹, 黑色轨迹为 KITTI 数据集随机序列中的车辆实际行车轨迹。受 KF 线性假设影响, 原 Deep SORT 算法跟踪轨迹中个别轨迹点与实际轨迹点偏差较大, 通过观测值修正 KF 参数导致整个跟踪过程的滞后性, 且由于目标实际运动非线性, 导致 KF 参数中误差积累, 影响跟踪性能。本文的 TCA 能有效修复 KF 中积累的线性误差, 修复后的轨迹与原始轨迹拟合程度更高, 更新后的 KF 参数对下一帧的预测的准确率提高, 从而提高车辆运动特征表达能力。

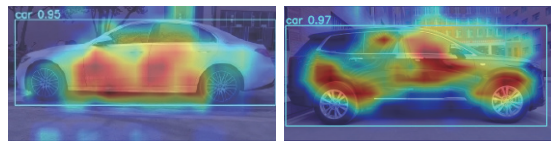
图 12 为不同算法在 KITTI 数据集 tracking-testing 0001 序列中的跟踪结果。106 帧时目标清晰无遮挡, 4 种算法均取得良好跟踪效果。133 帧时, 车辆目标被前方交通标识牌部分遮挡, 本文方法和 Strong SORT 算法仍能取得较好跟踪结果, OC SORT 算法和 Deep SORT 算法只能追踪到目标车辆尾部。157 帧时, 车辆目标遮挡加重, 4 种算法跟踪效果均受影响。192 帧时, 车辆不仅存在遮挡, 且受背



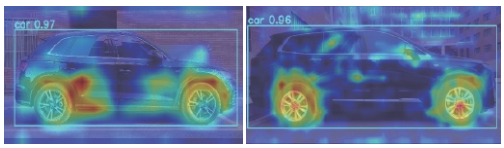
(a) 样例车辆1~车辆5



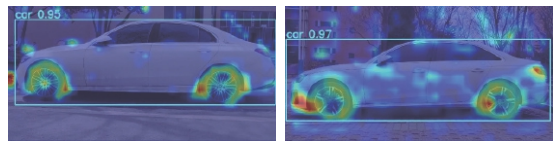
(b) 同车型不同颜色



(c) 不同车型不同颜色



(d) 同车型同颜色 (黑色SUV)



(e) 同车型同颜色 (白色轿车)

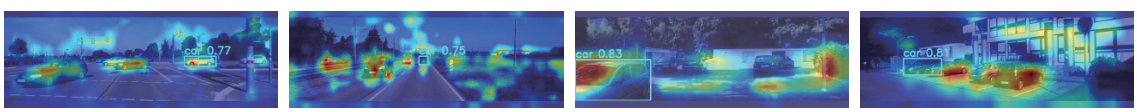
图 9 不同输入下车辆特征注意力热图

Fig. 9 Heat map of vehicle feature attention with different inputs



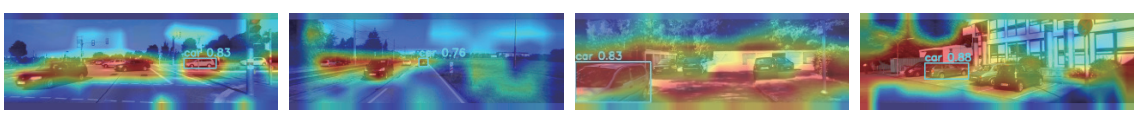
(a1) 场景1 (a2) 场景2 (a3) 场景3 (a4) 场景4

(a) 原始图像



(b1) 场景1 (b2) 场景2 (b3) 场景3 (b4) 场景4

(b) Deep SORT算法

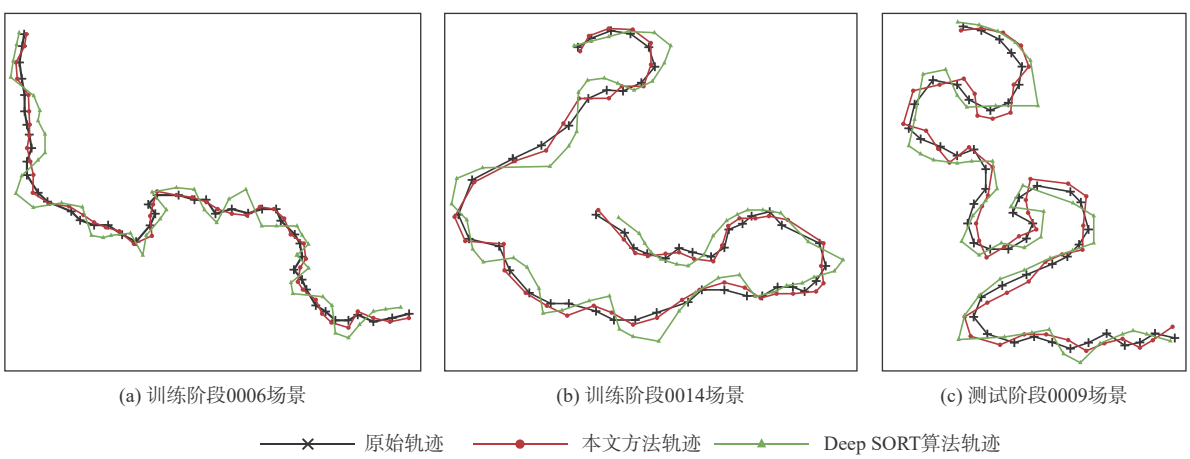


(c1) 场景1 (c2) 场景2 (c3) 场景3 (c4) 场景4

(c) 本文方法

图 10 KITTI 数据集场景车辆特征热力图

Fig. 10 Heat map of vehicle characteristics for dataset scenarios of the KITTI



(a) 训练阶段0006场景 (b) 训练阶段0014场景 (c) 测试阶段0009场景

图 11 车辆跟踪轨迹

Fig. 11 Vehicle tracking trajectory

景影响较大，OC SORT 算法和 Deep SORT 算法跟踪框出现跟踪漂移现象，本文方法和 Strong SORT 算法仍能正确跟踪，但本文方法跟踪框更准确。综上，在面对遮挡及背景影响时，本文方法能够较好地跟踪到目标，具有更好的鲁棒性。

图 13 为本文方法在真实道路场景的跟踪结

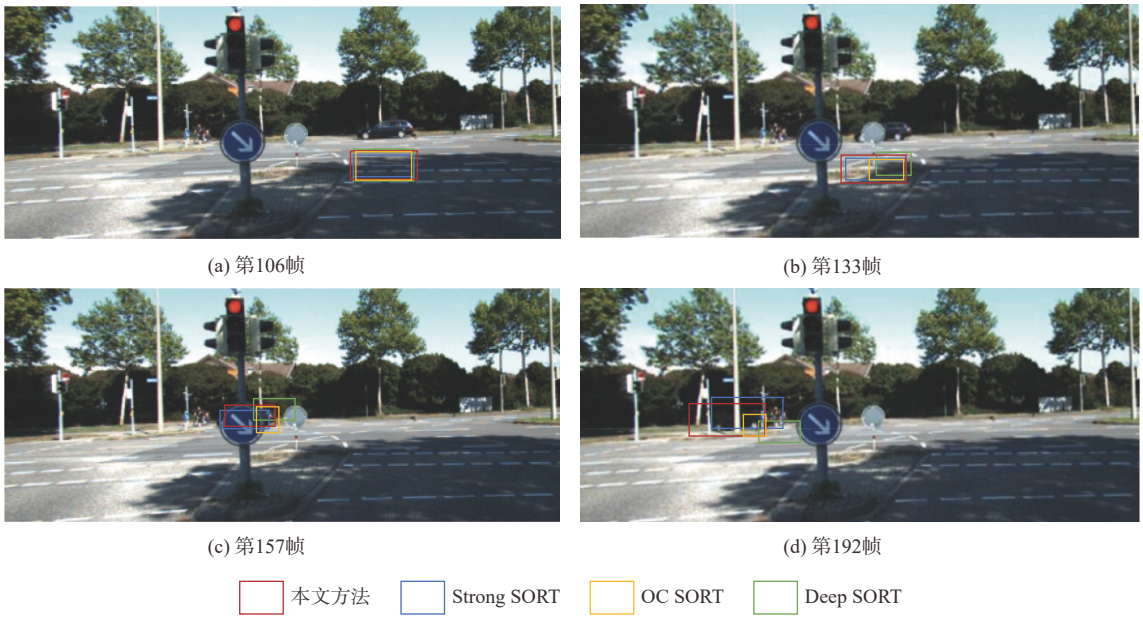
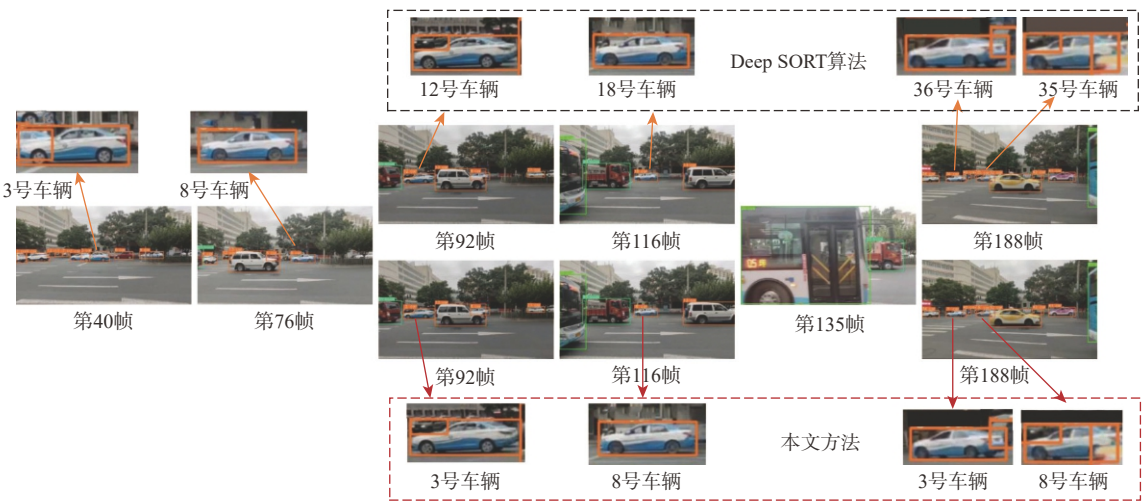
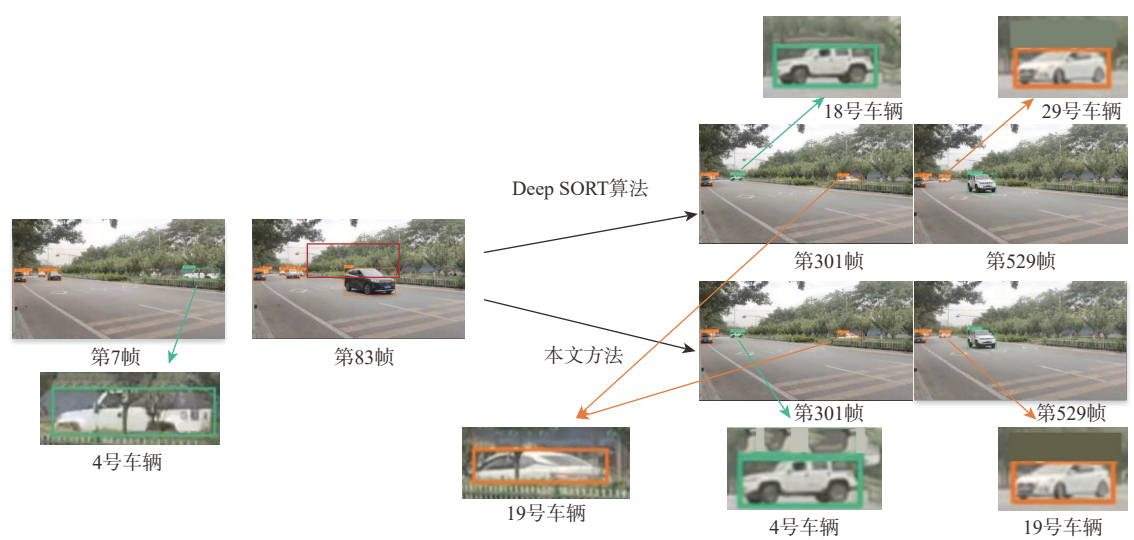


图 12　不同算法跟踪结果对比

Fig. 12　Comparison of tracking results of different algorithms



(a) 目标被道路车辆遮挡



(b) 目标被环境遮挡

图 13　真实道路场景跟踪结果

Fig. 13　Real road scene tracking results

果，并与原 Deep SORT 算法做对比。图 13(a)为目标被道路车辆遮挡的跟踪结果：40 帧时，3 号目标车辆出现；76 帧时 3 号车辆被遮挡，出现与 3 号车辆外观相似的 8 号车辆；92 帧时，3 号车辆重新出现，但 Deep SORT 算法为其分配新的 ID 12，出现 IDS 现象，本文方法能正确识别出其为原 3 号车辆并继续跟踪；同时 8 号车辆在第 92 帧开始被遮挡，116 帧时遮挡结束，Deep SORT 算法为其分配新 ID 18，本文方法正确跟踪；135 帧时，视频画面被大尺度公交车辆完全遮挡，188 帧遮挡结束，虽然原 3 号和 8 号车辆外观相似度较大，但本文方法仍然能够正确区分并跟踪，Deep SORT 算法再次重新分配 ID 35 和 36，出现多次 IDS 现象，影响跟踪准确性。

图 13(b)为目标被周边环境遮挡的跟踪结果：第 7 帧时 4 号目标车辆在上行车道首次出现，第 83 帧时被车道间树木遮挡，第 301 帧 4 号车辆重新出现并驶入下行车道，Deep SORT 算法不能正确识别，为其分配新的 ID 18，导致跟踪断裂，本文方法正确跟踪；同时在 301 帧上行车道出现 19 号目标车辆，在第 529 帧时驶入下行车道，Deep SORT 算法为其分配新的 ID 29，本文方法正确跟踪。

在不同场景下的实验结果表明本文方法能有效改善目标在被遮挡过后的错误跟踪问题，减少 IDS 现象，提高跟踪准确率。

4 结 论

现有 MOT 算法对车辆目标关注较少，车辆跟踪过程中因遮挡造成的跟踪漂移及 IDS 现象仍是一大难点，为克服上述问题，本文在 Deep SORT 基础上提出一种融合全尺度特征与轨迹修正的遮挡车辆跟踪方法，通过仿真实验及真实场景测试对本文方法进行可行性验证。主要结论如下：

1）为提高跟踪器对目标外观特征表达能力，引入一种 FSNet，通过多个分支分别提取不同尺度目标特征，由统一的 AG 实现自适应多尺度动态融合，最终输出目标特征，经 FSNet 提取后的目标特征具有更强的表达能力。

2）传统 KF 受线性假设限制，在目标被遮挡时无法通过观测值更新参数，导致参数误差积累。本文提出 TCA，通过目标被遮挡前最后一帧轨迹与目标重新出现后第 1 帧轨迹对遮挡过程中的轨迹进行修复，并重新更新 KF 参数，减小线性误差影响，修正后的轨迹与真实轨迹拟合程度更高。

3）经仿真实验、可视化分析及真实场景测试，验证了本文方法可以有效避免 IDS 现象，提高遮挡车辆跟踪性能。

4）本文方法为典型 TBD 跟踪范式，通过堆叠多个模块提高跟踪性能，但也带来较大计算量，如何融合外观特征与运动特征构造一体化的 MOT 模型将成为 MOT 领域研究新趋势。

参考文献（References）

[1] PARK Y, DANG L M, LEE S, et al. Multiple object tracking in deep learning approaches: a survey[J]. Electronics, 2021, 10(19): 2406.

[2] LUO W H, XING J L, MILAN A, et al. Multiple object tracking: a literature review[J]. Artificial Intelligence, 2021, 293: 103448.

[3] 周雪, 梁超, 何均洋, 等. 一体化多目标跟踪算法研究综述[J]. 电子科技大学学报, 2022, 51(5): 728-736.
ZHOU X, LIANG C, HE J Y, et al. A survey on one-shot multi-object tracking algorithm[J]. Journal of University of Electronic Science and Technology of China, 2022, 51(5): 728-736(in Chinese).

[4] 韩瑞泽, 冯伟, 郭青, 等. 视频单目标跟踪研究进展综述[J]. 计算机学报, 2022, 45(9): 1877-1907.
HAN R Z, FENG W, GUO Q, et al. Single object tracking research: a survey[J]. Chinese Journal of Computers, 2022, 45(9): 1877-1907(in Chinese).

[5] DENDORFER P, OŠEP A, MILAN A, et al. MOTChallenge: a benchmark for single-camera multiple target tracking[J]. International Journal of Computer Vision, 2021, 129(4): 845-881.

[6] CIAPARRONE G, LUQUE SÁNCHEZ F, TABIK S, et al. Deep learning in video multi-object tracking: a survey[J]. Neurocomputing, 2020, 381: 61-88.

[7] EMAMI P, PARDALOS P M, ELEFTERIADOU L, et al. Machine learning methods for data association in multi-object tracking[J]. ACM Computing Surveys, 2020, 53(4): 1-34.

[8] BEWLEY A, GE Z Y, OTT L, et al. Simple online and realtime tracking[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2016: 3464-3468.

[9] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2017: 3645-3649.

[10] DU Y H, ZHAO Z C, SONG Y, et al. StrongSORT: make DeepSORT great again[J]. IEEE Transactions on Multimedia, 2023, 25: 8725-8737.

[11] CAO J K, PANG J M, WENG X S, et al. Observation-centric SORT: rethinking SORT for robust multi-object tracking[EB/OL]. (2023-03-16)[2023-03-20]. https://arxiv.org/abs/2203.14360v3.

[12] MAGGIOLINO G, AHMAD A, CAO J K, et al. Deep OC-sort: multi-pedestrian tracking by adaptive re-identification[C]//Proceedings of the IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2023: 3025-3029.

[13] AHARON N, ORFAIG R, BOBROVSKY B Z. BoT-SORT: robust associations multi-pedestrian tracking[EB/OL]. (2022-07-07)[2023-03-20]. https://arxiv.org/abs/2206.14651v2.

[14] SEIDENSCHWARZ J, BRASÓ G, SERRANO V C, et al. Simple cues lead to a strong multi-object tracker[EB/OL]. (2022-06-09)[2023-03-20]. https://arxiv.org/abs/2206.04656v7.

[15] YANG F, ODASHIMA S, MASUI S, et al. Hard to track objects with irregular motions and similar appearances? make it easier by buffering the matching space[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2023: 4788-4797.

[16] 马永杰, 马芸婷, 程时升, 等. 基于改进 YOLOv3 模型与 DeepSORT 算法的道路车辆检测方法[J]. 交通运输工程学报, 2021, 21(2): 222-231.
MA Y J, MA Y T, CHENG S S, et al. Road vehicle detection method based on improved YOLOv3 model and deep-SORT algorithm [J]. Journal of Traffic and Transportation Engineering, 2021, 21(2): 222-231(in Chinese).

Track obscured vehicles by fusing full-scale features with trajectory correction

GUO Junfeng，ZHANG Zihua*

(School of Electrical and Mechanical Engineering，Lanzhou University of Technology，Lanzhou 730050，China)

Abstract：An occluded vehicle tracking method that fuses full-scale features with trajectory correction based on the Deep SORT algorithm is proposed to improve the tracking drift and identity switching (IDS) problems caused by occlusion in vehicle tracking. First, a full-scale feature extraction network is introduced to extract features of different scales of the target and to achieve adaptive dynamic fusion to enhance the apparent features of the target. Then, a trajectory correction approach is suggested to fix the tracking trajectory during the occlusion process, and the Kalman filter parameters are adjusted in order to minimize the cumulative linear errors during the occlusion process and optimize the target motion features. Finally, occluded vehicle tracking is achieved by combining the appearance features and the motion features. The feasibility of the proposed method is verified by ablation experiments and visualized analysis. The suggested approach successfully resolves the IDS issue in obscured vehicle tracking and increases the robustness of vehicle tracking, as demonstrated by experimental results on the KITTI dataset, which yield the highest overall score of 78.13% and the fewest number of IDS (192), when compared to typical existing methods.

Keywords：computer application; multiple object tracking; vehicle tracking; full scale; trajectory correction

Received：2023-05-29；**Accepted**：2023-07-28；**Published Online**：2023-09-01 14：28

URL：link.cnki.net/urlid/11.2625.V.20230901.1138.003

Foundation item：National Natural Science Foundation of China (51465034)

* **Corresponding author**. E-mail：zzh15536308429@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0275

NNW-FlowStar 软件模拟并联分离特性

张培红，周桂宇，沈盈盈，唐静，赵炜*，贾洪印
(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：两级入轨(TSTO)飞行器并联分离过程中，一级和二级飞行器之间存在激波多次反射，对飞行器的压力分布、力矩特性和飞行姿态会产生较大影响，甚至可能严重影响飞行器级间分离的安全性。基于自主开发的NNW-FlowStar软件和非结构混合网格自适应技术，对国家数值风洞工程自主设计的某两级入轨飞行器标模开展并联分离特性研究，并与风洞试验数据进行对比分析，从而验证了NNW-FlowStar软件模拟飞行器并联分离特性的可靠性和有效性。研究表明：NNW-FlowStar软件可以较好模拟两级入轨飞行器并联分离特性，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，采用网格自适应技术可以有效提升模拟精度。两级入轨飞行器并联分离过程会经历组合体流动、缝隙流动、小通道流动、大通道流动，最终到自由流动等不同的典型流动阶段，激波结构快速变化，存在激波/激波干扰、边界层/边界层干扰和激波/边界层干扰等复杂的流动现象。

关 键 词：两级入轨；并联分离；数值模拟；网格自适应；NNW-FlowStar
中图分类号：V221+.3; TB553
文献标志码：A **文章编号**：1001-5965(2025)05-1620-09

可重复使用空天飞行器可以有效降低空间出入成本，具有快速、经济、机动、安全等特点[1]。两级入轨(two stages to orbit, TSTO)飞行器是实现可重复使用的重要技术途径，也是当前技术风险较小、技术基础较好、实现难度相对低的技术方案[2]。两级飞行器的安全分离是可重复使用的关键技术之一。相比常规的飞机导弹分离、级间分离问题，两级入轨飞行器并联分离属于2个升力体间的分离，与重力相比，气动力对分离特性的影响相对较大，对数值方法的模拟精度要求更高。飞行器在分离过程中，一级、二级飞行器之间存在复杂的激波结构，二级飞行器的头部激波会在两级飞行器之间产生多次反射，对飞行器的压力分布、力矩特性和飞行姿态等产生严重影响，甚至可能影响飞行器级间分离的安全性[2]。由于两级飞行器都是具有大升力面的气动外形，对流场变化非常敏感，在高速情况下，如果流场预测不精确，控制不当，易发生事故。采用数值模拟手段对两级入轨飞行器开展级间并联分离的设计与评估极其重要。由于两级入轨飞行器并联分离流动的复杂性，对数值模拟方法提出了严峻挑战[3]，基于标模风洞试验数据，确认软件模拟飞行器并联分离特性的可靠性和有效性，非常有必要，且意义重大。

计算流体动力学(computational fluid dynamics, CFD)模拟在应用过程中，面临的最大问题就是模拟结果的可靠性问题，即确认软件的模拟精度[4]。美国国家航空航天局(National Aeronautics and Space Administration, NASA)和美国航天航空学会(American Institute of Aeronautics and Astronautics, AIAA)针对数值模拟软件的确认，开展了一系列研究项目和

收稿日期：2023-05-24；**录用日期**：2023-07-21；**网络出版时间**：2023-08-02 12:24
网络出版地址：link.cnki.net/urlid/11.2625.V.20230801.1727.001
基金项目：国家数值风洞工程
* **通信作者**. E-mail：ethnuaa@163.com

引用格式：*张培红，周桂宇，沈盈盈，等. NNW-FlowStar 软件模拟并联分离特性 [J]. 北京航空航天大学学报，2025，51（5）：1620-1628.*
ZHANG P H，ZHOU G Y，SHEN Y Y，et al. Simulation of parallel separation characteristics using NNW-FlowStar software [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1620-1628 （in Chinese）.

研讨会[5]。NASA 开展了大型气动数值模拟可信度研究国际合作项目 CAWAPI-2[6], 该项目建立了 F-16XL-1 飞机亚/跨声速气动共享数据库, 包括相当完备的风洞试验数据、飞行试验数据和大量的 CFD 计算数据, 十多个国家参加了该项目。NASA 兰利研究中心组织召开了由 7 个国家 75 个单位参加的横向喷流及湍流分离控制 CFD 确认专题研讨会[7]。AIAA 先后组织了 6 次有关运输机构型阻力预测的研讨会(drag prediction workshop, DPW)[5,8-10], 重点关注典型运输机巡航构型阻力的数值模拟精度。并先后举行了 3 次 CFD 高升力预测研讨会(high lift prediction workshop, HiLiftPW)[11], 重点关注典型运输机增升构型升力与力矩的数值模拟精度。关于湍流模型及湍流的数值模拟, NASA 的湍流模型标准算例工作组专门于兰利研究中心设立了湍流模型资源网站[12]。此外, AIAA 还支持组织了高精度格式系列研讨会[13] 等。

随着 CFD 在飞行器设计中的广泛应用, 以及地位和作用日趋突出, 国内也先后多次组织了 CFD 软件的确认工作。中国空气动力研究与发展中心与中国航空工业西安航空计算技术研究所联合发起了“CFD 统一算例研讨”活动[14], 采用统一的标模算例对国内多家 CFD 软件进行对比确认。中国空气动力研究与发展中心举办了 CT-1 标模大迎角气动特性研讨会[15], 基于 CT-1 标模的风洞试验数据对国内 CFD 软件模拟大迎角流动的能力进行确认对比。中国空气动力研究与发展中心联合国内多家单位, 分别组织召开了第 1 届、第 2 届航空 CFD 可信度研讨会[16], 采用 CHN-T1 运输机标模对 CFD 软件进行确认计算, 极大促进了国内 CFD 软件的发展。从 2018 年开始, 中国空气动力研究与发展中心在国家数值风洞(national numerical wind-tunnel, NNW)工程项目的支持下, 联合国内优势研究单位和力量, 共同开展 CFD 软件验证与确认相关研究[17], 在基础理论、应用支持平台开发、验证和确认数据库建设、工程实践等方面都取得了长足进步。国家数值风洞工程验证与确认团队建成的验证与确认数据库, 已于 2023 年 1 月在国家空间科学数据中心正式上线, 为国内 CFD 软件的验证与确认提供可靠的标模数据。

本文基于自主开发的国家数值风洞工程 NNW-FlowStar 软件[18] 和网格自适应技术, 对某两级入轨飞行器标模开展了并联分离特性数值模拟研究, 主要目的是评估 NNW-FlowStar 软件模拟飞行器级间并联分离的能力, 验证软件模拟并联分离流动特性的可信度。通过对比分析采用密网格、粗网格和网格自适应技术, 与实验值比较, 验证了 NNW-FlowStar 软件模拟飞行器并联分离特性的可靠性。

1　数值模拟方法

NNW-FlowStar 软件是在国家数值风洞工程支持下, 项目团队自主开发的一款面向航空航天领域飞行器气动计算分析的大型通用 CFD 计算分析软件。2020 年 9 月, 已面向全国免费发布[18]。

流动控制方程为守恒形式的非定常可压缩雷诺平均 Navier-Stokes 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为时间量; S为面积; Ω为控制体的体积; $\partial\Omega$为控制体封闭面的面积; $\boldsymbol{W}$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:

$$\boldsymbol{W}=\begin{bmatrix}\rho\\ \rho u\\ \rho q\\ \rho w\\ \rho E\end{bmatrix} \tag{2}$$

$$\boldsymbol{F}_{\mathrm{c}}=\begin{bmatrix}\rho V\\ \rho uV+n_x p\\ \rho qV+n_y p\\ \rho wV+n_z p\\ \rho HV\end{bmatrix} \tag{3}$$

$$\boldsymbol{F}_{\mathrm{v}}=\begin{bmatrix}0\\ n_x\tau_{xx}+n_y\tau_{xy}+n_z\tau_{xz}\\ n_x\tau_{yx}+n_y\tau_{yy}+n_z\tau_{yz}\\ n_x\tau_{zx}+n_y\tau_{zy}+n_z\tau_{zz}\\ n_x\Theta_x+n_y\Theta_y+n_z\Theta_z\end{bmatrix} \tag{4}$$

式中: (x,y,z)为笛卡儿坐标系下的坐标分量; (u,q,w)为笛卡儿坐标系下的速度分量; (n_x,n_y,n_z)为微元面的单位法向矢量在笛卡儿坐标系下的分量; E、V、H分别为总能、逆变速度、总焓; τ_{ij}为切应力; Θ_x、Θ_y、Θ_z为源项; ρ和p分别为密度和压力。

对于分离物体的质心运动方程, 直接在机体的体轴系中求解, 机体的体轴系看作惯性系:

$$m\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t}=m\boldsymbol{g}+\sum\boldsymbol{F} \tag{5}$$

$$\begin{cases}m\dfrac{\mathrm{d}U}{\mathrm{d}t}=mg_x+F_x=-mg_x\sin\theta_{\mathrm{initial}}+F_x\\ m\dfrac{\mathrm{d}V}{\mathrm{d}t}=mg_y+F_y=mg_y\sin\varphi_{\mathrm{initial}}\cos\theta_{\mathrm{initial}}+F_y\\ m\dfrac{\mathrm{d}W}{\mathrm{d}t}=mg_z+F_z=mg_z\cos\varphi_{\mathrm{initial}}\cos\theta_{\mathrm{initial}}+F_z\end{cases} \tag{6}$$

式中: m为质量; $\theta_{\mathrm{initial}}$和$\varphi_{\mathrm{initial}}$分别为飞行器初始状

态空间姿态的俯仰欧拉角和滚转欧拉角；$\boldsymbol{F}$为分离物体所受气动力和其他外力的和；F_x、F_y、F_z分别为合力在x、y、z方向的分量；g_x、g_y、g_z分别为重力加速度在x、y、z方向的分量。

网格自适应采用项目团队建立的耦合流场特征的网格自适应技术[19]和提出的“先唯一后同一”策略[20]进行网格单元分布优化过程的并行相容性。以流场速度梯度张量的旋转分量和剪切分量为流场探测特征变量构造网格自适应探测器，探测器判据如下：

$$Q=\frac{\omega^2}{\omega^2+\eta^2} \tag{7}$$

式中：Q为网格自适应判据；ω为流场速度梯度张量的旋转分量；η为流场速度梯度张量的剪切分量。

对流通量采用 Roe 通量差分裂方法进行离散，黏性通量采用中心差分离散，湍流模型采用 SST 两方程模型。

2 计算模型及网格

计算模型采用某两级入轨飞行器标模[21]，如图 1 所示。该标模是在国家数值风洞工程支持下，由中国空气动力研究与发展中心联合中国航天空气动力技术研究院等单位共同自主设计的并联分离标模，用于考核 CFD 动网格技术、非定常流场解算和非定常气动/运动耦合数值模拟方法对高超声速复杂干扰流动结构、气动力干扰效应，以及多体动态分离轨迹的精细模拟和准确预测能力。标模由一级飞行器和二级飞行器组成，采用背负式并联形式。一级飞行器为升力体外形，包含了翼梢下反的大后掠翼、双垂尾、襟翼、副翼等，长为 85 m，高为 15.36 m，翼展为 62.726 m，参考面积为 2 100 m^2；二级飞行器为翼身组合体外形，包含单垂尾、体襟翼、襟翼、副翼等，长为 35 m，高为 4.2 m，翼展为 17.5 m，参考面积为 35 m^2。

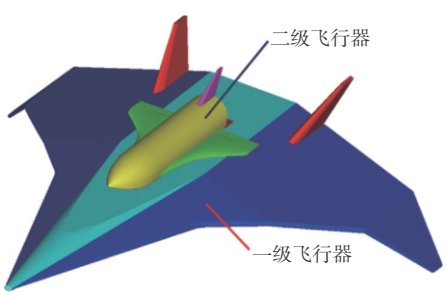


图 1 两级入轨飞行器标模

Fig. 1 Standard model of two stages to orbit vehicles

标模在中国空气动力研究与发展中心直径为 1 m 高超声速风洞（FD-30）完成了马赫数 Ma=6.0 的高超声速并联分离捕获轨迹试验。直径为 1 m 高超声速风洞（FD-30）是一座暂冲型、吹吸式的常规高超声速风洞，具备开展飞行器测力测压试验、舵面铰链力矩试验、多体分离网格测力试验和连续跟踪系统（continuous tracking system, CTS）轨迹捕获试验、喷流干扰试验、模型自由飞试验等能力。图 2 为两级入轨飞行器并联分离标模在 FD-30 风洞中的试验模型。两级飞行器模型均按真实外形 1∶100 进行几何缩比。



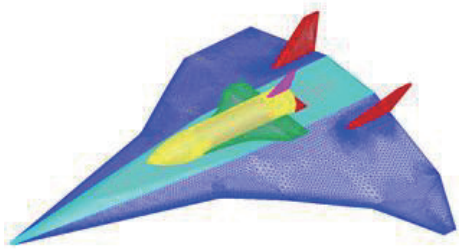
图 2 并联分离风洞试验模型

Fig. 2 Model of parallel separation wind tunnel test

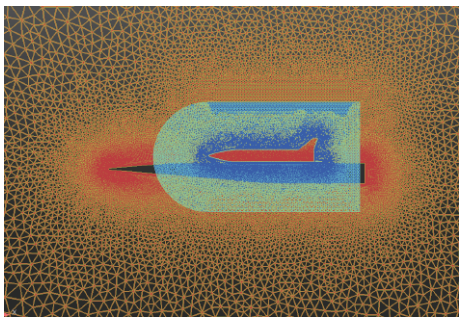
采用非结构混合网格生成策略，在物面附近采用三棱柱网格单元提高附面层流动模拟精度，其他区域采用四面体/金字塔填充，第 1 层网格间距$y^+\leqslant 1$。背景网格计算区域取一级飞行器展长的 8.0 倍作为远场，子网格计算区域取二级飞行器展长的 2.0 倍作为插值边界。计算时分别采用密网格、粗网格和基于粗网格的自适应网格进行了并联分离特性的模拟。密网格包含三棱柱、金字塔和四面体，共计约 2 800 万，其中，一级飞行器网格约 1 430 万，二级飞行器网格约 1 370 万。粗网格包含三棱柱、金字塔和四面体，共计约 510 万，其中，一级飞行器网格约 270 万，二级飞行器网格约 240 万。两级入轨飞行器计算网格如图 3 所示。

网格自适应采用耦合流场特征的网格自适应技术，自适应探测器采用基于拟梯度自适应判据的激波探测器。网格自适应时的并行计算采用项目团队提出的“先唯一后同一”策略进行网格单元分布优化过程的并行相容性。

图 4 为动态优化过程网格量变化曲线。图 5 为动态优化过程典型时刻网格分布。初始时刻两级飞行器距离近，间隙内的流场存在激波多次反射，波系复杂，网格规模增加很多，出现图 4 中的峰值，网格总量达到约 1 800 万，其中，一级飞行器网格量约 600 万，二级飞行器网格量约 1 200 万。随着分离距离增大，激波反射减弱，网格规模逐渐下降，最终大致维持于动态平衡状态，总量约 1 010 万，其中，一级飞行器约 410 万，二级飞行器约 600 万。



(a) 两级入轨飞行器表面网格



(b) 两级入轨飞行器对称面网格

图 3　两级入轨飞行器计算网格(密网格)

Fig. 3　Computation grid of two stages to orbit vehicles (fine grid)

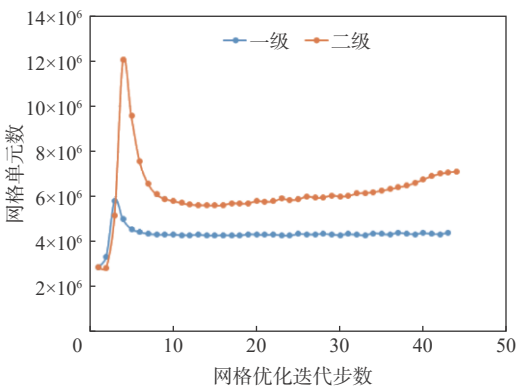
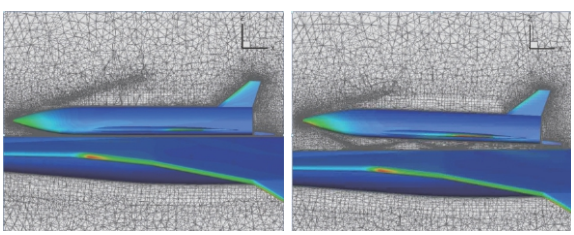
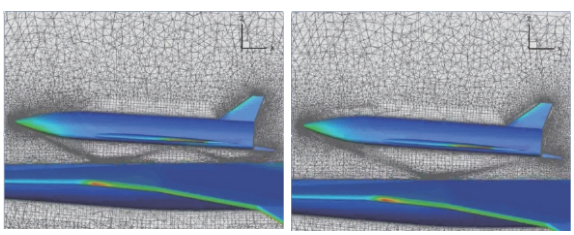


图 4　动态优化过程网格量变化

Fig. 4　Grid quantity change during dynamic optimization



(a) t=0 s　(b) t=0.2 s



(c) t=0.4 s　(d) t=0.6 s

图 5　典型时刻网格分布

Fig. 5　Grid distribution at typical moments

3　计算结果分析

计算状态与试验条件相同: 马赫数 Ma=6.0, 迎角 α=−2º, 侧滑角 β=0º, 来流总温 T_0 = 492 K, 来流总压 $P_0 = 1.89\times10^6$ Pa。两级飞行器仅在气动作用力驱动下实现分离。图 6～图 8 分别给出了密网格、

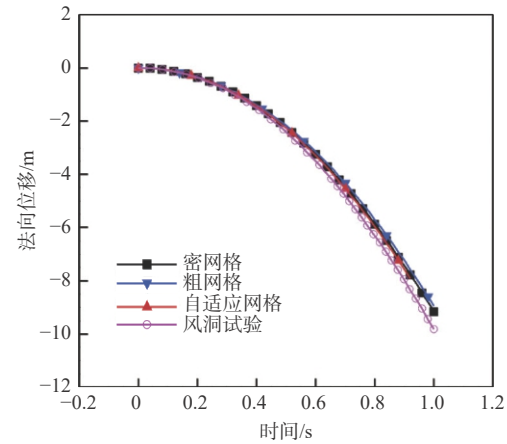


图 6　一级飞行器法向位移比较

Fig. 6　Comparison of normal distance for the primary vehicle

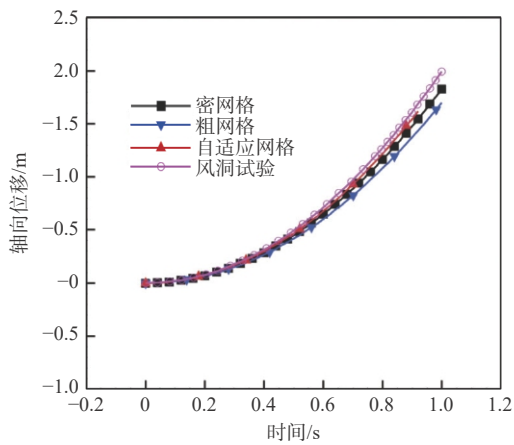


图 7　一级飞行器轴向位移比较

Fig. 7　Comparison of axial distance for the primary vehicle

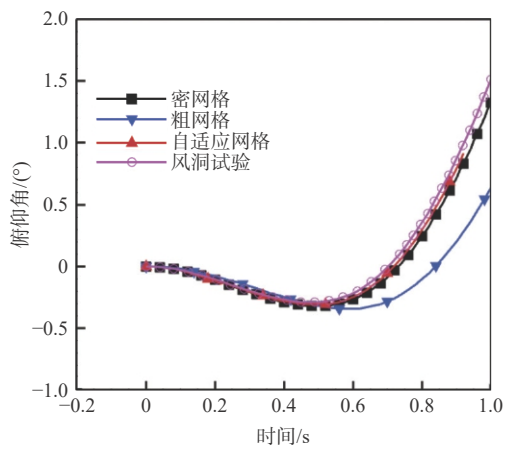


图 8　一级飞行器俯仰角比较

Fig. 8　Comparison of pitch angle for the primary vehicle

粗网格和自适应网格计算得到的一级飞行器法向位移、轴向位移和俯仰角与风洞试验数据比较。图9～图11分别给出了密网格、粗网格和自适应网格计算得到的二级飞行器法向位移、轴向位移和俯仰角与风洞试验数据比较。可以看出密网格和网格自适应计算结果与试验值吻合较好，网格自适应计算结果与试验值吻合最好，采用网格自适应技术后，模拟精度明显提高。

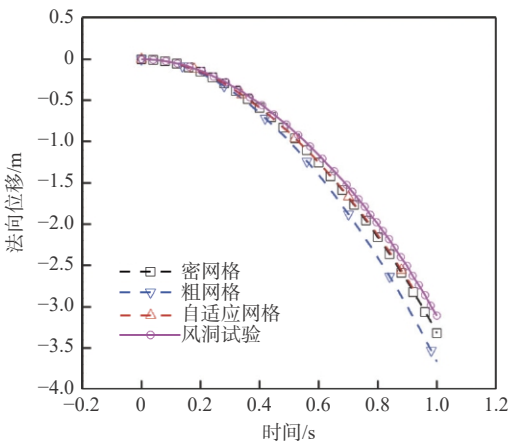


图9 二级飞行器法向位移比较

Fig. 9 Comparison of normal distance for the second stage vehicle

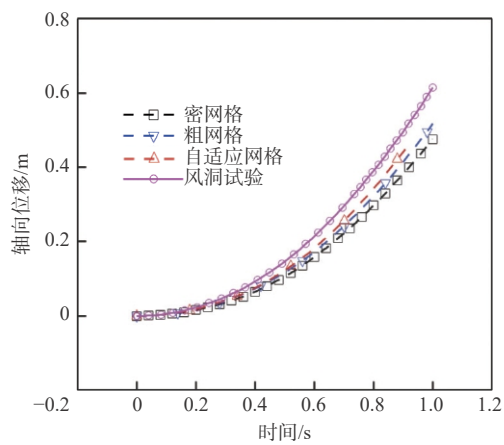


图10 二级飞行器轴向位移比较

Fig. 10 Comparison of axial distance for the second vehicle

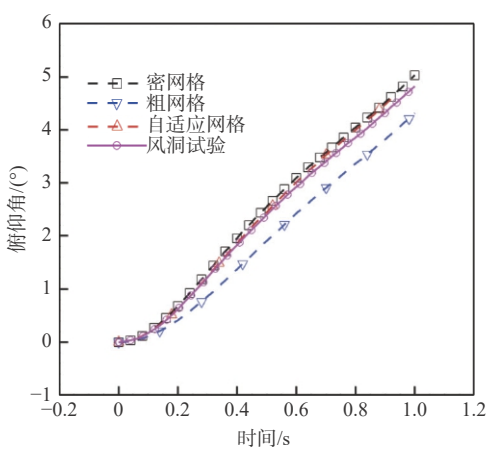


图11 二级飞行器俯仰角比较

Fig. 11 Comparison of the pitch angle for the second vehicle

一级和二级飞行器相对初始位置均逐渐向下运动，由于一级飞行器向下运动的更快，二级飞行器相对一级飞行器呈向上拉开的趋势，0.8 s的真实分离时间里，自适应网格计算得到的两级飞行器法向相对初始位置拉开距离为3.83 m，风洞试验得到的两级飞行器法向相对初始位置拉开距离为4.31 m，误差约为−0.48 m。一级和二级飞行器相对初始位置均逐渐向后运动，但由于一级飞行器向后运动的更快，二级飞行器相对一级飞行器则呈向前运动的趋势，0.8 s的真实分离时间里，自适应网格计算和风洞试验得到的两级飞行器轴向相对初始位置拉开距离均为0.88 m。俯仰方向，一级飞行器呈先低头再逐渐抬头的趋势，0.8 s的真实分离时间里，自适应网格计算得到的一级飞行器俯仰角相对初始位置增加了0.30°，风洞试验算得到的一级飞行器俯仰角相对初始位置增加了0.36°，误差约为0.06°；二级飞行器相对初始位置逐渐抬头且俯仰角呈增大趋势，0.8 s的真实分离时间里，自适应网格计算得到的二级飞行器俯仰角相对初始位置增加了4.00°，风洞试验算得到的一级飞行器俯仰角相对初始位置增加了3.86°，误差约为0.14°。

由图6～图11可知，密网格和自适应网格计算得到的位移和姿态角差异不大。为进一步分析气动特性与流场结构的关系，以及网格对计算结果的影响，对比分析了密网格和自适应网络计算得到的气动特性和不同时刻流场结构。图12为粗网格和自适应网格计算得到二级飞行器上气动特性对比，图13为不同典型时刻，粗网格和自适应网格计算得到流场结构对比。由图13(a)可知，粗网格和自适应网格计算得到的流场结构相近，自适应网格计算得到的头部激波更精细，2种网格计算的得到的两飞行器之间的高压区域结构，激波后局部压力分布和飞行器底部附件压力分布有差异。网格不同产生的初始流场结构的差异，导致二级飞行器上受到的气动力和气动力矩不同，使得二级飞行器分离过程的位移和姿态角产生差异。位移和姿态角的不同，进一步加大了飞行器气动特性的差异和同一时刻流场结构的不同。由图12(c)可知，初始时刻t=0 s到t=0.24 s自适应网格计算得到的抬头力矩相对粗网格计算结果较大，会使飞行器俯仰角较大，t=0.24 s之后，虽然自适应网格计算得到的抬头力矩变小或低头力矩变大，但由于飞行器抬头俯仰角速度的影响，自适应网格计算得到的二级飞行器俯仰角仍然会比粗网格的大，与图11中的结果一致。

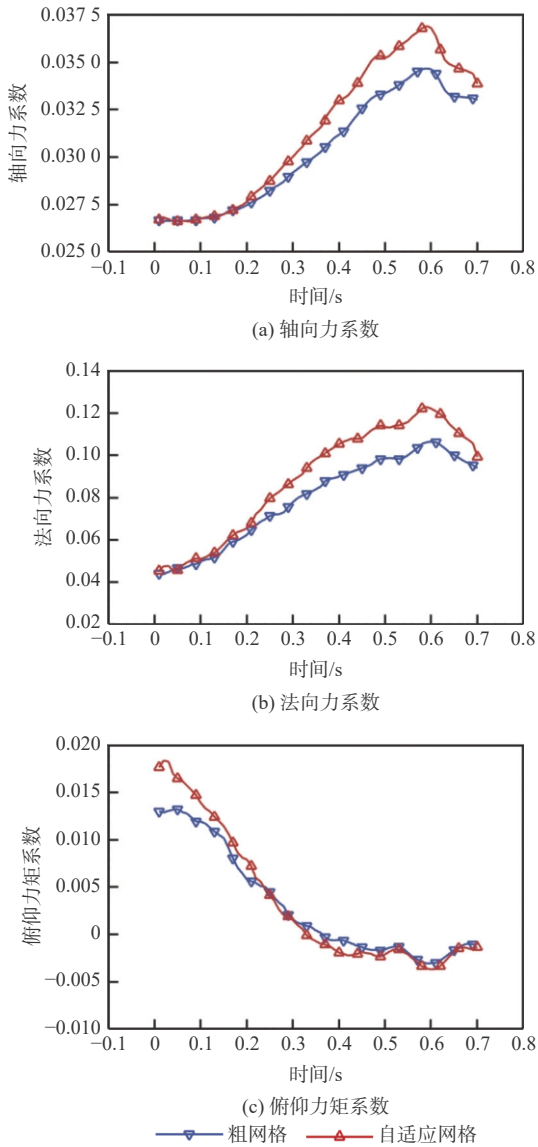


图 12　粗网格和自适应网格计算得到二级飞行器上气动特性

Fig. 12　Aerodynamic characteristics of the second vehicle for coarse and adaptation mesh

图 14 为计算得到的飞行器典型流场结构与风洞试验结果比较。可以看出，计算得到的流场结构与风洞试验结果一致，飞行器在整个分离过程中，经历了组合体流动、缝隙流动、小通道流动、大通道流动和自由流动等典型流动。分离之前为组合体流动，两级尚未分离以组合体形式存在，其流动特征主要为组合体气动外形产生的外部绕流流动特征，包括弓形激波、两级主激波干扰、二级在一级上方由于边界层黏性和逆压梯度引起的头部前缘流动分离而产生的分离激波等。分离启动阶段为缝隙流动，两级飞行器处于分离启动阶段，分离速度较慢，级间通道较窄，两级飞行器之间的缝隙中没有明显的激波反射，但存在边界层/边界层干扰、激波/边界层干扰、激波/激波干扰、流动分离等

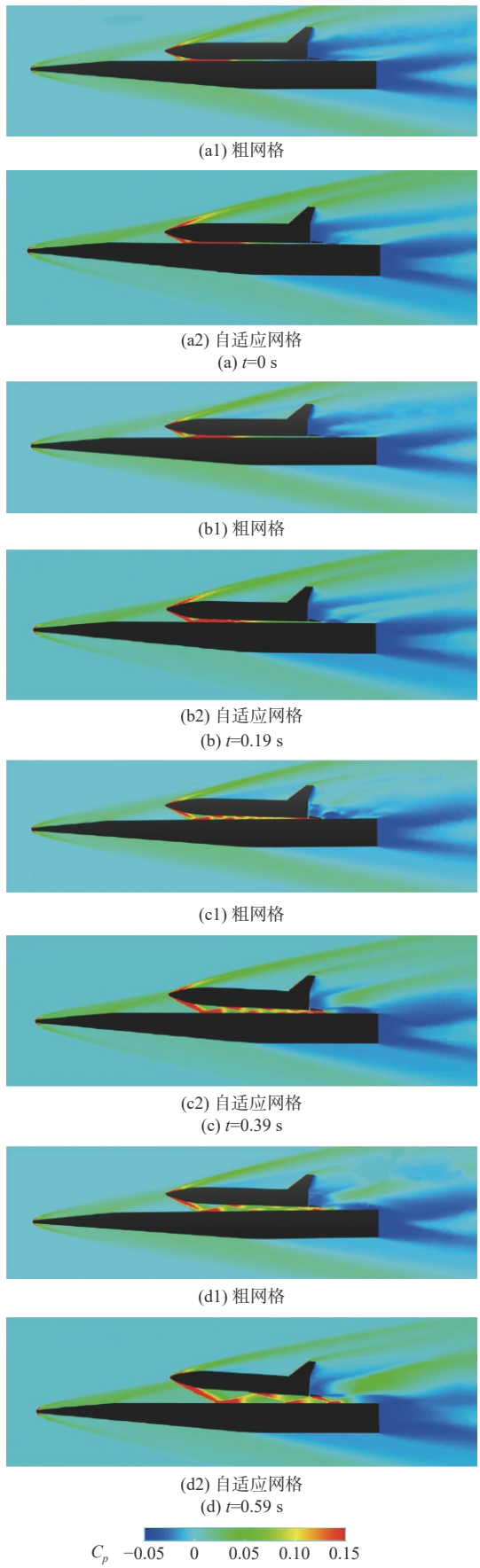
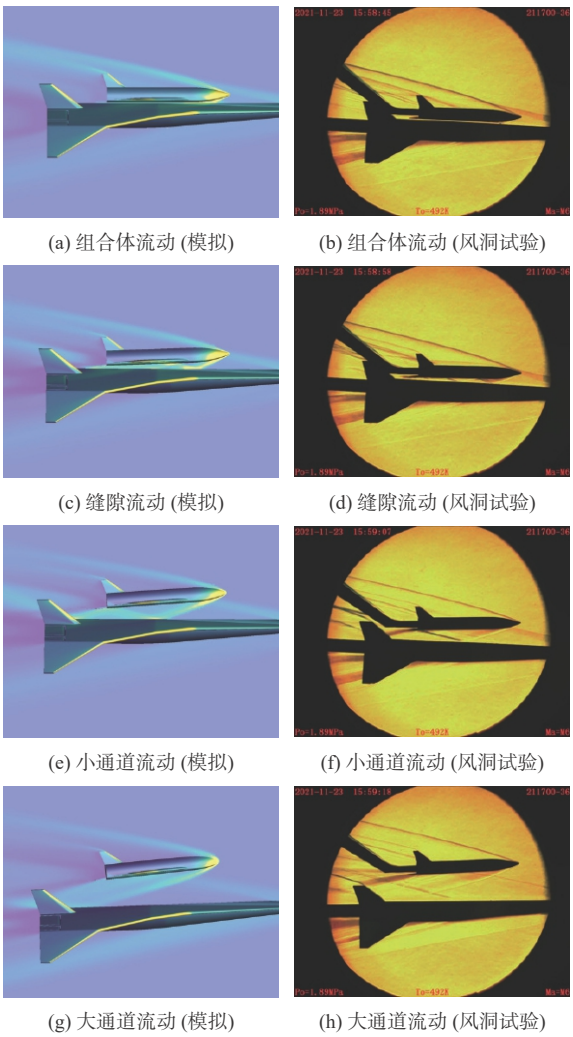


图 13　粗网格和自适应网格模拟得到流场结构

Fig. 13　Flow field structure for coarse and adaptation mesh calculations



(a) 组合体流动 (模拟) (b) 组合体流动 (风洞试验)

(c) 缝隙流动 (模拟) (d) 缝隙流动 (风洞试验)

(e) 小通道流动 (模拟) (f) 小通道流动 (风洞试验)

(g) 大通道流动 (模拟) (h) 大通道流动 (风洞试验)

图 14 飞行器典型流场结构模拟和风洞试验对比

Fig. 14 Typical flow field structure from simulation and wind tunnel test

复杂的缝隙流动。分离初期,随着分离过程中二级飞行器与一级飞行器之间通道高度的增大,两级飞行器之间形成小通道流动,在一级飞行器和二级飞行器之间,存在一系列级间激波反射过程,这些反射激波在一级和二级表面引起流动分离,并随着通道高度的进一步增大,激波反射次数逐渐减少。两级飞行器分离中后期,随着两级飞行器分离姿态的建立,会形成大通道流动,由于两级飞行器之间的高度进一步增大,一级飞行器和二级飞行器头部激波在通道内不能形成激波串,但二级头激波仍可打到一级飞行器上表面,在一级飞行器表面引起流动分离,或者一级飞行器头激波打到二级飞行器下表面,引起当地流动分离。当分离距离再进一步增大,两级之间不再受彼此影响,各自逐渐与自由流的流场结构一致。

通过以上分离特性曲线和典型流场结构的对比分析,可以看出,本文方法可以较好模拟两级入轨飞行器并联分离特性,计算的得到的一级飞行器和二级飞行器位移、姿态角与风洞试验结果吻合很好,计算得到的流场结构与风洞试验结果一致,验证了 NNW-FlowStar 软件模拟并联分离特性的能力。

4 结 论

基于自主设计的两级入轨飞行器并联分离标模和风洞试验数据,开展了 NNW-FlowStar 软件模拟并联分离特性研究。通过粗网格、密网格和自适应网格计算结果与风洞试验结果比较,验证了 NNW-FlowStar 软件模拟飞行器并联分离特性的能力。主要结论如下:

1) 数值模拟结果与风洞试验结果吻合较好,NNW-FlowStar 软件可以较好模拟两级入轨飞行器的并联分离特性。

2) 采用网格自适应技术,可以有效提升并联分离特性的模拟精度。

3) 两级入轨飞行器并联分离过程激波结构快速变化,经历组合体流动、缝隙流动、小通道流动、大通道流动和自由流动等不同流动阶段。

参考文献 (References)

[1] 张国成, 姚彦龙, 王慧. 美国两级入轨水平起降可重复使用空天运载器发展综述[J]. 飞机设计, 2018, 38(2): 1-6.
ZHANG G C, YAO Y L, WANG H. A survey on development of two-stage-to-orbit horizontal-takeoff-horizontal-landing reusable launch vehicle in USA[J]. Aircraft Design, 2018, 38(2): 1-6(in Chinese).

[2] ROBINSON J C, STANLEY D O. Structural and loads analysis of a two-stage reusable manned launch system[J]. La Revue Du Praticien, 2015, 31(5): 821-829.

[3] 王国辉, 王小军, 杨勇, 等. 重复使用运载器气动性能 CFD 研究[J]. 导弹与航天运载技术, 2005(6): 6-11.
WANG G H, WANG X J, YANG Y, et al. CFD study of aerodynamic performance of reusable launch vehicle[J]. Missiles and Space Vehicles, 2005(6): 6-11(in Chinese).

[4] PORTER J L, AGARWAL R, AZAD R S, et al. Guide for the verification and validation of computational fluid dynamics Simulation: AIAA G-077-1998[R]. Reston: AIAA, 1998.

[5] 王运涛. DPW Ⅳ～DPW Ⅵ数值模拟技术综述[J]. 航空学报, 2018, 39(4): 021836.
WANG Y T. An overview of DPW Ⅳ-DPW Ⅳ numerical simulation technology[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(4): 021836(in Chinese).

[6] LUCKRING J M, RIZZI A W. Toward improved CFD predictions of slender airframe aerodynamics using the F-16XL aircraft (CAWAPI-2) [C]//Proceedings of the 52nd Aerospace Sciences Meeting. Reston: AIAA, 2014.

[7] JONES G S, JOSLIN R D. Proceedings of the 2004 NASA/ONR circulation control workshop: NASA/CP-2005-213509/PT2[R]. Washington, D. C. : NASA, 2005.

[8] LEVY D W, ZICKUHR T, VASSBERG J C, et al. Summary of

data from the first AIAA CFD drag predication workshop: AIAA-2002-0811[R]. Reston: AIAA, 2002.

[9] LAFLIN K R, KLAUSMEYER S M, ZICKUHR T, et al. Data summary from second AIAA computational fluid dynamics drag prediction workshop[J]. Journal of Aircraft, 2005, 42(5): 1165-1178.

[10] VASSBERG J C, TINOCO E N, MANI M, et al. Abridged summary of the third AIAA computational fluid dynamics drag prediction workshop[J]. Journal of Aircraft, 2008, 45(3): 781-798.

[11] 洪俊武, 王运涛, 李伟, 等. HiLiftPW-3 高升力构型数值模拟[J]. 航空学报, 2019, 40(3): 122391.
HONG J W, WANG Y T, LI W, et al. Numerical simulation of high-lift configuration from HiLiftPW-3[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(3): 122391(in Chinese).

[12] 郑晓静, 王国华. 高雷诺数壁湍流的研究进展及挑战[J]. 力学进展, 2020, 50: 1-49.
ZHENG X J, WANG G H. Progresses and challenges of high Reynolds number wall-bounded turbulence[J]. Advances in Mechanics, 2020, 50: 1-49(in Chinese).

[13] 姜振华, 阎超. 高超声速热流高精度数值模拟方法[J]. 北京航空航天大学学报, 2011, 37(12): 1529-1533.
JIANG Z H, YAN C. High order accurate methods in numerical simulation of hypersonic heat transfer[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(12): 1529-1533 (in Chinese).

[14] 王运涛, 王光学, 张玉伦. TRIP2.0 软件的确认: DPWⅡ复杂组合体的数值模拟[J]. 航空学报, 2008, 29(1): 34-40.
WANG Y T, WANG G X, ZHANG Y L. Validation of TRIP2.0: numerical simulation of DPWⅡ complex configuration[J]. Acta Aeronautica et Astronautica Sinica, 2008, 29(1): 34-40(in Chinese).

[15] 王运涛, 王光学, 陈作斌. CT-1 标模大迎角静态气动特性数值模拟[J]. 航空学报, 2008, 29(4): 859-865.
WANG Y T, WANG G X, CHEN Z B. Numerical simulation of static aerodynamic characteristics of CT-1 model at high angles of attack[J]. Acta Aeronautica et Astronautica Sinica, 2008, 29(4): 859-865(in Chinese).

[16] 张耀冰, 唐静, 陈江涛, 等. 基于非结构混合网格的 CHN-T1 标模气动特性预测[J]. 空气动力学学报, 2019, 37(2): 262-271.
ZHANG Y B, TANG J, CHEN J T, et al. Aerodynamic characteristics prediction of CHN-T1standard model with unstructured grid[J]. Acta Aerodynamica Sinica, 2019, 37(2): 262-271(in Chinese).

[17] 赵炜, 陈江涛, 肖维, 等. 国家数值风洞 (NNW) 验证与确认系统关键技术研究进展[J]. 空气动力学学报, 2020, 38(6): 1165-1172.
ZHAO W, CHEN J T, XIAO W, et al. Advances in the key technologies of verification and validation system of national numerical windtunnel project[J]. Acta Aerodynamica Sinica, 2020, 38(6): 1165-1172(in Chinese).

[18] 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).

[19] 唐静, 崔鹏程, 贾洪印, 等. 非结构混合网格鲁棒自适应技术[J]. 航空学报, 2019, 40(10): 122894.
TANG J, CUI P C, JIA H Y, et al. Robust adaptation techniques for unstructured hybrid mesh[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(10): 122894(in Chinese).

[20] 唐静, 张健, 李彬, 等. 非结构混合网格自适应并行技术[J]. 航空学报, 2020, 41(1): 123202.
TANG J, ZHANG J, LI B, et al. Parallel algorithms for unstructured hybrid mesh adaptation[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(1): 123202(in Chinese).

[21] 唐伟, 刘深深, 余雷, 等. 用于级间分离研究的 TBCC 动力 TSTO 气动布局概念设计[J]. 空气动力学学报, 2019, 37(5): 698-704.
TANG W, LIU S S, YU L, et al. Conceptual design of TBCC based TSTO configurations for stage seperation investigation[J]. Acta Aerodynamica Sinica, 2019, 37(5): 698-704(in Chinese).

Simulation of parallel separation characteristics using NNW-FlowStar software

ZHANG Peihong，ZHOU Guiyu，SHEN Yingying，TANG Jing，ZHAO Wei*，JIA Hongyin

(Computational Aerodynamic Institute of China Aerodynamics Research and Development Center，Mianyang 621000，China)

Abstract：During the parallel separation process of two stages to orbit (TSTO) vehicles, there are multiple reflections of shock waves between the first and second stage vehicles. The complex flow has a great impact on the pressure distribution, torque characteristics and flight attitude of the vehicles, and may even seriously affect the safety of the separation between stages of the vehicles. The parallel separation properties of the two stages to orbit vehicle model created by the National Numerical Wind Tunnel Project are examined using adaptation techniques for unstructured hybrid mesh and the self-developed National Numerical Wind Tunnel Project software, NNW-FlowStar. The simulation results are compared with the wind tunnel test data, and the reliability and effectiveness of NNW-FlowStar simulation of parallel separation characteristics of vehicles are confirmed. The research shows that the NNW-FlowStar can better simulate the parallel separation characteristics of the two stages to orbit the vehicle. The numerical simulation results are in good agreement with the test results. The calculated flow field structure is consistent with the wind tunnel test. Using the mesh adaptive technology can effectively improve the simulation accuracy. T In order to circle the vehicle, the two stages will separate in parallel and travel through several typical flow stages, including combination flow, gap flow, small channel flow, big channel flow, and free flow. During the whole process, the shock structures change rapidly, and there are complex flow phenomena such as shock wave interference, boundary layer interference and shock wave/boundary layer interference.

Keywords：two stages to orbit; parallel separation; numerical simulation; mesh adaptation; NNW-FlowStar

Received：2023-05-24；**Accepted**：2023-07-21；**Published Online**：2023-08-02 12：24

URL：link.cnki.net/urlid/11.2625.V.20230801.1727.001

Foundation item：National Numerical Windtunnel Project

* **Corresponding author**. E-mail：ethnuaa@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2024.0750

基于Gossip协议的高效集群数据同步方案

张宏海[1,2,*]，崔斌豪[1,2]，李一鸣[1,2]，田丰[1,2]，贾永强[1,2]，肖傲三[1,2]

(1. 中国民航信息网络股份有限公司 研发中心，北京 101318；2. 中国民航信息网络股份有限公司 北京市民航大数据工程技术研究中心，北京 101318)

摘　　要：随着民航客票运价搜索系统业务的快速发展，系统集群内部网络流量规模不断增长。为了在高负载网络流量场景下，实现集群内部数据同步，提出基于Gossip协议的集群数据同步方案。所提方案从网络协议的传输层和应用层着手设计，在传输层使用用户数据报协议（UDP）来减少集群中节点间的连接数量和交互次数，从而实现数据在网络传输过程中的低流量、低耗时。在应用层使用Gossip传播协议来实现数据的最终一致性，保证数据传输的可靠性。通过传输层UDP传输协议和应用层Gossip传播协议相结合，保证了集群监控过程中数据同步的高效性和可靠性。

关 键 词：Gossip协议；数据同步；一致性算法；高性能集群；去中心化
中图分类号：TP393
文献标志码：A　　**文章编号**：1001-5965(2025)05-1629-08

民航运价搜索系统作为互联网时代旅客自助购买机票的基础工具，承载了海量民航旅客用户的机票查询业务，为保障这一关键民航业务的安全运行，需要系统为用户提供低延迟、高稳定、高扩展的搜索服务。为保证系统的稳定运行，集群内部的数据同步方案，特别是监控数据的同步方案显得尤为重要[1]。

民航运价搜索系统的数据同步，面临2个方面的压力：①民航运价搜索系统涉及机票、航班等业务，业务复杂度高，系统存在交互报文大的特点；②系统承载了海量的用户查询量，导致集群内部面临巨大的网络流量压力。如何在高网络负载的情景下，保障系统的低延迟、高稳定、高扩展，是系统当前面临的主要技术障碍。针对上述技术障碍，本文提出一套采用基于用户数据报协议(user datagram protocol, UDP)的Gossip协议 的集群数据同步方案，可以在高负载网络流量的情况下，保障系统高可靠的服务能力。

传统集群数据同步方案通常基于传输控制协议(transmission control protocol, TCP)，在高并发和高网络负载的情况下，采用传统的TCP方式进行数据同步会面临一些瓶颈。TCP是一种面向连接的协议，其缺点包括连接的建立和维护成本较高、相对较重的头部开销及对可靠性的强调会导致较高的延迟。在高并发的场景中，这些问题可能进一步放大，导致系统的性能瓶颈和资源消耗增加。相比之下，基于UDP的Gossip协议在高并发、高网络负载的环境中具有一些优势。UDP是一种无连接、轻量级的传输协议，不涉及连接的建立和维护，具有较低的传输延迟。结合Gossip协议，通过节点之间的分布式信息传播，实现数据同步具有去中心化、容错性强的特点，使系统更具弹性。在大规模网络中，UDP的轻量级和低开销特性有助于降低通信开销，从而更有效地应对高并发和高网络负载的挑战。本文方案采用UDP-Based Gossip Protocol进行数据同步可以在高并发环境中更好地适应系统

收稿日期：2024-10-16；**录用日期**：2024-11-08；**网络出版时间**：2024-11-19 14:17
网络出版地址：link.cnki.net/urlid/11.2625.V.20241119.0932.001
基金项目：国家自然科学基金 (U2033203)
*通信作者. E-mail：hhzhang@travelsky.com.cn

引用格式：张宏海，崔斌豪，李一鸣，等. 基于Gossip协议的高效集群数据同步方案[J]. 北京航空航天大学学报，2025，51(5)：1629-1636. ZHANG H H，CUI B H，LI Y M，et al. An efficient cluster data synchronization scheme based on the Gossip protocol[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1629-1636 (in Chinese).

动态变化，保障业务响应的低延迟，提高系统的扩展性和稳定性[2-3]。

1 Gossip协议实现原理

1.1 Gossip协议发展历史综述

Gossip协议的研究起源可追溯至1972年，Baker和Shostak对“电话问题”进行了描述，即探究最少的电话交流次数，使得每位妇女都能掌握所有流言[4]。1988年，Hedetniemi等对Gossip问题在计算机网络环境中进行了更为精确的概念界定，核心议题转变为信息如何在网络节点间传播[5]。Galton-Watson模型和Bailey的深入分析为Gossip算法的理论基础奠定了基石，而马尔可夫链的引入则进一步丰富了对其分析的方法论[6]。

自20世纪50年代至70年代，学者们致力于构建模型以描述Gossip机制并对其进行分析。随后十年间，该领域的研究热情有所减退[7]。随着新型网络技术的兴起和复杂网络研究的深化，Gossip协议的研究重新受到重视，近十年中取得了显著进展，其应用范围也扩展至数据库复制、故障检测、P2P网络的拓扑构建与维护，以及分布式环境下的聚合计算等多个领域[7]。

1.2 Gossip协议的定义

Gossip协议，又称为Epidemic协议[8]。该协议的核心思想是，当一个节点需要将信息同步到网络中的其他节点时，该节点可以定期地随机选择一些节点作为信息同步的目标。这些接收到信息的节点随后以相同的方式继续传播信息。Gossip协议模拟了信息的自发传播，类似于gossip(流言)和epidemic(流行病)的传播过程[9-10]。

Gossip协议以其简单高效、高扩展性、高容错性和去中心化等特点而著称。其设计追求一致性收敛，收敛速度较快，消息传播速度可达到对数级别的复杂度$\mathrm{Log}_2(N)$。这使得Gossip协议在大规模分布式系统中表现出色。尽管Gossip协议在许多方面具有显著的优势，但需要注意的是，在实时性要求较高的场景中，由于需要进行多轮传播，可能会引入一定的消息延迟[11-12]。

Gossip协议的数据一致性和数据收敛性，是其最显著的优势，以下针对这2点进行着重论证。Gossip协议的收敛性主要体现在其时间复杂度：

$$T = O(\log_2 N) \tag{1}$$

式中：T为收敛所需的时间或轮次；N为网络中的节点数。式(1)表明Gossip协议的收敛时间随着网络规模的增长呈对数增加，能够在对数级别轮次中达到快速收敛。Gossip协议的一致性体现在其消息的传播达到一致性的概率：

$$P_{\text{consistent}} = 1 - \mathrm{e}^{-\lambda N} \tag{2}$$

式中：λ为与Gossip协议参数相关的常数即传播概率；N为网络中的节点数。式(2)表明Gossip协议通过多轮信息交换，使得每个节点知道所有信息的概率逐渐增加，最终趋近于100%，确保了系统的一致性。同时，通过调整Gossip协议的参数，如每个周期内节点交换信息的频率和数量，可以提高达到一致性的概率和速度。

1.3 Gossip协议的通信方式

在Gossip协议下，2个节点之间通信方式主要有推模式、拉模式及推拉模式3种，这些模式在节点之间的通信中起着关键作用，决定了系统的效率和可靠性。推、拉和推拉这3种通信方式的特点，如图1所示。

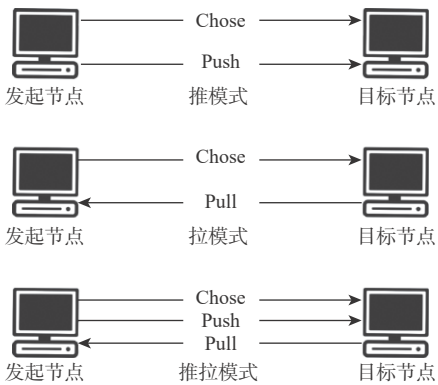


图1 Gossip协议的通信方式

Fig. 1 Communication method of Gossip protocol

1）推模式。信息的发起节点会随机选择其邻近节点作为目标节点，并向其发送自己的消息。这一模式的特点是仅有拥有消息的节点才有资格成为发起节点。

2）拉模式。信息的发起节点会随机选择邻近节点作为目标节点，但这次是从目标节点主动获取对方的消息。这一模式的优势在于，只有那些没有新消息的节点才会作为发起节点。

3）推拉模式。是将推和拉两者结合起来，使消息的发起节点不仅向随机选择的目标节点发送消息，同时也从对方获取消息。这种综合模式综合了推和拉的优点，克服了各自的局限性。通过推拉模式，系统能够更加灵活地进行信息的传递，既能高效推送消息，又能在需要时主动拉取所需的信息，从而提高系统的整体性能。

若将2个节点数据同步一次定义为一个周期，推模式在一个周期内需要进行1次通信，拉模式需要2次通信，而推拉模式则需要3次通信。尽管消息数量增加，但从效果上来看，推拉模式表现出最

高的效率, 理论上推拉模式的收敛速度也是最快的, 能够在一个周期内使得 2 个节点完全一致。

2　系统设计

2.1　系统架构

本文方案的系统架构主要由部署在集群中各节点的协调器组成, 协调器之间的网络通信主要通过 UDP 传输协议实现, 协调器之间的数据同步功能基于 Gossip 协议实现。协调器的主要组件包括 3 部分: 接收器组件、发送器组件、缓存组件, 系统组成逻辑如图 2 所示。

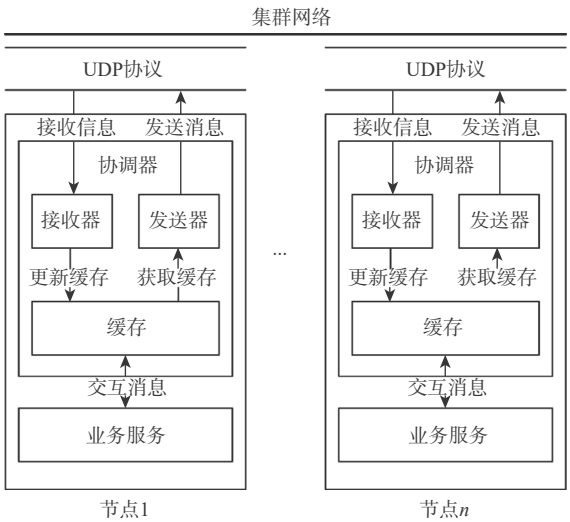


图 2　系统组成逻辑

Fig. 2　System composition logic

1) 发送器组件。负责对其他节点协调器分发请求信息, 将本节点的消息推送给集群中其他节点。

2) 接收器组件。接受其他协调器发来的请求信息, 目的是收集其他协调器发来的服务器信息。

3) 缓存组件。对协调器接受的服务器信息进行收集和存储, 对外提供查询和写入接口, 可使从外部获取服务器集群的相关信息, 节点中其他服务也可以通过接口和缓存之间交互信息。

2.2　传输协议的选择

本文方案根据民航运价搜索系统的业务特点, 最终选择 UDP 协议作为传输协议。选择 UDP 协议, 而放弃 TCP 协议主要出于 2 个方面的考虑: ①民航运价搜索系统的搜索业务对低延迟有强制性的要求, 而 TCP 协议是一种面向连接的协议, 连接的建立和维持成本较高, 存在较高的延迟; ②民航运价搜索业务的报文大并发高, 导致整体的网络负载较高, TCP 协议本身会占用大量网络连接数, 加之每次连接中间存在握手过程产生大量的数据包, 高网络负载的瓶颈会被进一步放大。综上两方面考虑, UDP 协议在本系统中更具优势[13-14]。

2.3　Gossip 传播模型

本节着重介绍 Gossip 协议节点间消息传播用到的传播模型。在讨论传播模型时, 将借鉴流行病学术语。根据术语, 集群中的每个节点可以处于以下 3 种状态之一, 即:

易感态: 节点未获取更新。

感染态: 节点已获取更新并传播。

愈除态: 节点已获取更新, 并且未发生新的更新, 不再参与传播过程。

Gossip 协议主要有 2 种传播模型: 简单流行病 (simple epidemics, SI) 模型和传染病 (susceptible infected recovered, SIR) 模型。

1) SI 模型。所有参与节点有 2 种状态, 分别为易感态和感染态。过程是种子节点会把所有的数据都跟其他节点共享, 以便消除节点之间数据的任何不一致, 其可以保证最终、完全的一致。通常只用于新加入节点的数据初始化[15]。

SI 模型的收敛速度可以通过下式近似:

$$E(S_{t+1}) = S_t\left(1-\frac{1}{N}\right)^{N(1-S_t)} \approx S_t e^{-(1-S_t)} \tag{3}$$

式中: S_t 为在第 t 个周期结束时易感节点的比例; N 为网络中节点的总数; $(1-1/N)$ 为固定感染节点不会感染某个固定易感节点的概率。显然, 如果一个节点在第 t+1 个周期是易感的, 那么其在第 t 个周期是易感的, 并且所有感染节点都选择了其他节点。实际上, 这个近似模型相当准确, 本文将期望值 $E(S_{t+1})$ 作为 S_{t+1} 的良好近似。由式(3)可以看出, 如果等待足够长的时间, 那么最终所有节点都会收到更新。换句话说, 一个特定节点永远不会收到更新的概率为 0。让每个节点都获得更新(被感染)所需的周期数为

$$S_N = \log_2 N + \ln N + O(1) \tag{4}$$

式中: S_N 为传播更新至所有节点所需的周期数。由式(4)可以看出, SI 模型传播效率相当高, 只需要 $O(\log_2 N)$ 个周期就可以让每个节点都获得更新, 这表明 SI 模型具有非常好的可扩展性。

SI 模型算法的伪代码如算法 1 所示。

算法 1　SI 模型算法。

```
loop
    wait(Δ)
    p ← random peer
    if push and in state I then
        send update to p
    end if
```

```
        if pull then
            send update-request to p
        end if
    end loop
    procedure ONUPDATE(m)
        store m.update
    end procedure
    procedure ONUPDATEREQUEST(m)
        if in state I then
            send update to m.sender
        end if
    end procedure
```

2）SIR模型。所有参与节点有3种状态：易感态、感染态和愈除态。过程是消息只包含最新update数据，谣言消息在某个时间点之后会被标记为“removed”，并且不再被传播。通常用于节点间数据增量同步[15]。

SIR模型相较于SI模型引入了“移除”状态，运行协议在所有节点都已接收更新后进行终止，提高了效率。终止条件如下：

$$S^* = \exp[-(k+1)(1-S^*)] \tag{5}$$

式中：S^*为当传播终止时未获得更新的节点比例，理想情况下，S^*趋近于0，意味着所有节点都已接收到更新；k表示平均每个感染节点在停止传播前传播信息的次数，该参数直接影响传播的效率和终止条件。式(5)表明，随着k的增加，S^*的值会迅速减小，因为指数函数的衰减速度非常快。这意味着，增加每个感染节点传播的次数可以显著提高传播的覆盖率。在实际应用中，选择合适的k值是一个权衡过程，需要考虑网络的规模、更新的紧急性及系统的资源限制的。例如，在需要快速传播更新的场景中，可能会选择较大的k值以确保更新能够迅速覆盖大部分节点。而在资源受限的环境中，可能会选择较小的k值以减少网络负载。

SIR模型算法的伪代码如算法2所示。

算法2 SIR模型算法。

```
    loop
        wait(Δ)
        p ← random peer
        if push and in state I then
            send update to p
        end if
        if pull then
            send update-request to p
        end if
    end loop
    procedure ONFEEDBACK(m)
        switch to state R with prob. 1/k
    end procedure
    procedure ONUPDATE(m)
        if in state I or R then
            send feedback to m.sender
        else
            store m.update
        end if
    end procedure
    procedure ONUPDATEREQUEST(m)
        if in state I then
            send update to m.sender
        end if
    end procedure
```

2.4 自研传播协议原理

本文方案在应用层采用Gossip Push和Gossip Pull相结合的方式分别发送增量消息和全量消息，在网络传输层采用UDP单播和UDP多播相结合的方式传输网络报文，有效的减少了网络中的网络负载压力，同时保障了主业务服务的低延迟[16-18]，自研传播协议的时序图如图3所示。

具体步骤如下：

步骤1 启动协调器主进程。集群中的每个节点都部署有协调器服务，依次启动每台节点协调器服务进程。协调器进入初始化状态，并会持续监听节点上的业务服务进程，对业务接口进行实时监听，并将业务服务状态(api_state)写入缓存组件，如果监听到业务服务接口的状态更新，会将更新的数据写入缓存，如果接口状态没有变化则会持续监听。

步骤2 启动监听线程。协调器主进程启动的同时，监听线程随之启动，监听线程主要负责Gossip消息的监听，每个节点都有两类监听进程：Gossip主动监听线程和Gossip被动监听进程。

Gossip主动监听线程：

1）每隔固定时间t触发监听。

2）如果业务服务状态api_state的值为Join，表示业务服务为新加入服务，采用Push-Gossip方式将节点本地数据(LD)加入AckMsg消息当中，通过UDP广播的方式推送给集群中的其他目标节点。如果业务服务状态api_state的值为Update，表示业务服务的数据发生更新，首先随机选取邻居节点，同样采用Gossip push方式将节点本地数据(LD)加入AckMsg消息当中，通过UDP单播的方式推送给选中邻居节点发送消息。

3）更新节点状态node_state为I，表示当前节点易感染。

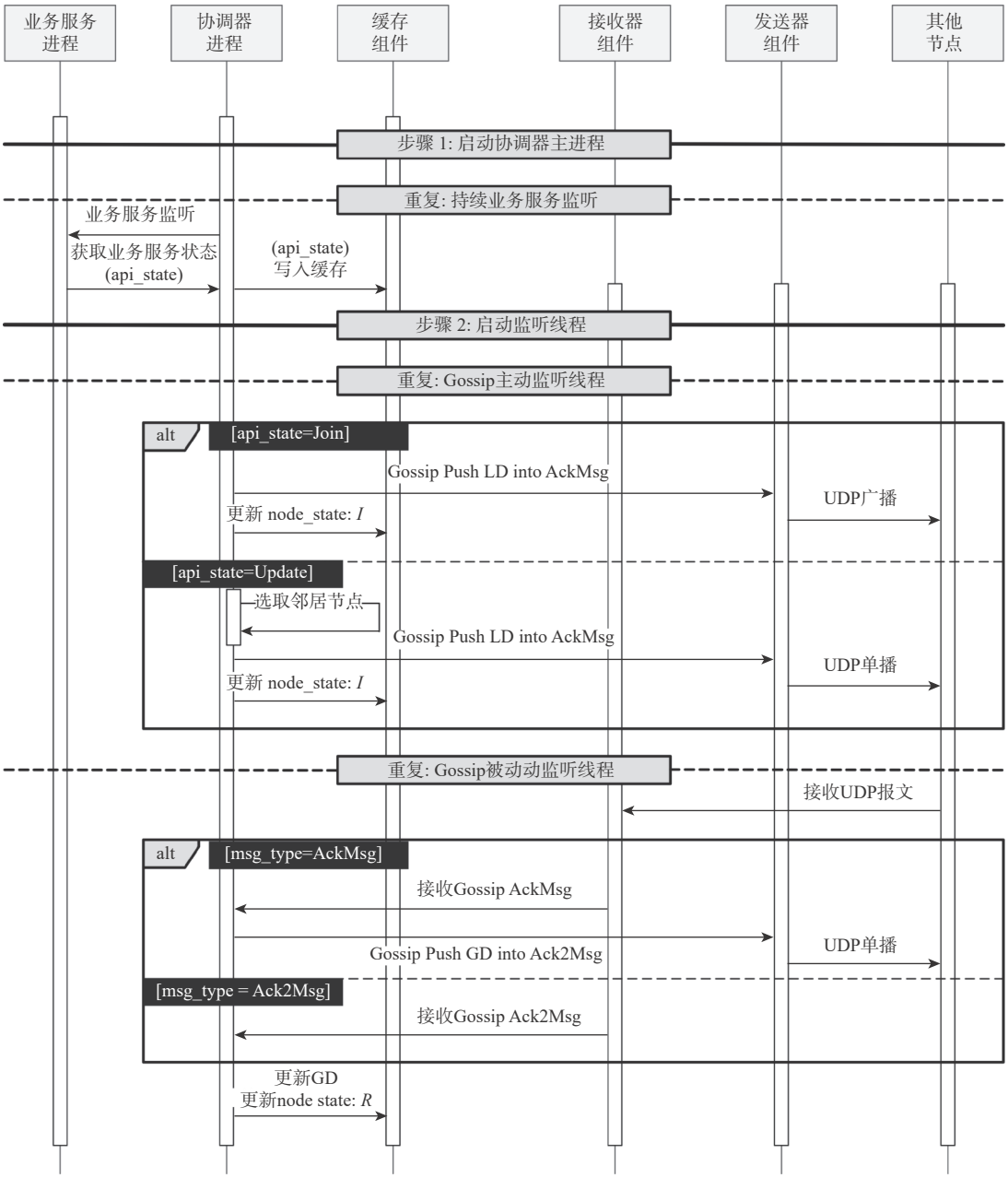


图 3　自研传播协议时序图

Fig. 3　Timing diagram of self-developed propagation protocol

4) 当前节点的被动监听进程进入监听状态, 等待和-Gossip push 消息对应的 Gossip pull 消息的返回。

Gossip 被动监听线程:

1) 每隔固定时间 t 触发监听, 接收网络中 UDP 报文。

2) 如果当前节点是目标节点, 则会接收到源节点发送过来的 AckMsg 消息, 将 AckMsg 中的本地数据(LD)和目标节点自身节点存储的全局数据(GD)进行合并, 并将合并后的全局数据(GD)加入到 Ack2Msg 消息当中, 通过 UDP 单播的方式返回给源节点。

3) 如果当前节点为源节点, 则会接收到目标节点返回的 Ack2Msg 消息, 将 Ack2Msg 中的全局数据(GD)和源节点自身存储的本地数据(LD)进行合并。

4) 重复第 2, 第 3 步, 如果持续 k 次收到同一消息, 将合并后的全局数据(GD)写入缓存组件, 同时将节点状态(node_state)更新为 R, 表示消息传播结束, 终止传播。

3　工程验证及分析

传统数据同步方法包括中心化同步、周期性轮询、时间戳同步、基于版本控制的同步和事务性同步, 其在实现简单性、一致性维护和系统复杂度之间存在权衡。与之相比, 基于 Gossip 协议的 SI 模型和 SIR 模型提供了去中心化、动态自适应性、概

率一致性保证、轻量级设计和系统健壮性等优势。Gossip协议通过优化收敛时间 $O(\log_2 n)$、传输率、通信负载、网络负载率和消息延迟等关键性能指标，在大规模分布式系统中展现出高效性。其概率性传播机制和轻量级通信策略，使得在保持数据一致性的同时，减少了消息冗余和网络负载，提高了数据同步的效率。

基于上述理论依据，系统使用Java语言在工程上进行了编码实现和实践验证，目前系统已部署到生产和开发环境，生产环境的机器和网络配置如表1所示。

表1 验证环境配置

Table 1 Verify environment configuration

参数	配置
编程语言	Java
操作系统	Linux Redhat v6.5
主机类型	X86服务器
CPU核数	16Core
内存	512 GB
硬盘	3T SSD
集群网络	万兆局域网
Node数量	100台

实验针对Gossip协议的数据一致概率、收敛周期、容错率3项关键评估指标进行验证分析。记录了10 000轮的数据同步结果，每1 000轮结果针对3项关键指标求算数平均值。实验设定的传播概率为0.8，即每个节点在每个周期内有80%的概率向随机选择的邻居节点传播信息。信息更新间隔设置为1轮，即每个节点在每个周期都会尝试更新其信息状态。每1 000轮随机故障率设为(0%～0.2%)，实验的统计分析数据如表2所示。

以上实验结果显示，一致概率平均值为99.84%，表明在绝大多数情况下，网络能够在极短时间轮次内达到高度的数据一致性。收敛周期的平均值为2.84轮，这意味着信息在网络中传播并达到一致性状态的速度非常快。容错率的平均值为99.62%，表明即使在部分节点出现故障的情况下，Gossip协议仍能保持极高的一致性，说明该协议具有很强的容错能力。从1 000轮到10 000轮，容错率略有波动，但总体上保持在99%以上，显示出Gossip协议在不同轮次和可能的故障条件下的稳定性。综上所述，Gossip协议在仿真测试中展现出了卓越的性能，尤其是在容错性和快速收敛方面的表现，使其成为分布式系统中数据同步的有力候选方案。

表2 Gossip协议关键评估指标

Table 2 Gossip Protocol Key Performance Metrics

轮次	一致概率/%	收敛周期/轮	容错率/%
1 000	98.8	3.2	99.8
2 000	99.9	2.9	100
3 000	100	2.8	99.9
4 000	99.9	3.1	98.7
5 000	99.6	2.2	99.3
6 000	99.7	2.4	98.9
7 000	100	2.5	98.8
8 000	100	3.0	100
9 000	99.9	2.7	99.9
10 000	100	2.6	100

为验证在高负载网络流量的集群中进行数据同步方案的可行性和有效性，选择了10台与生产机器同等配置的机器进行方案验证。每台机器的网络负载为100 TPS，每次请求的所有报文大小平均为150 kB。通过持续压测这些机器，持续时间达到了24 h，并详细记录了整个过程中每台机器使用传统数据同步协议和UDP-Based Gossip协议时的网络负载率和业务服务耗时。

在实验结束后，对这10台机器的数据进行了平均值计算，得到了2组对比实验数据，如图4和图5所示。通过2组对比数据，可以看到自研的UDP-Based Gossip协议相较于传统数据同步方案在网络负载率和业务耗时两方面均有显著提高，可以

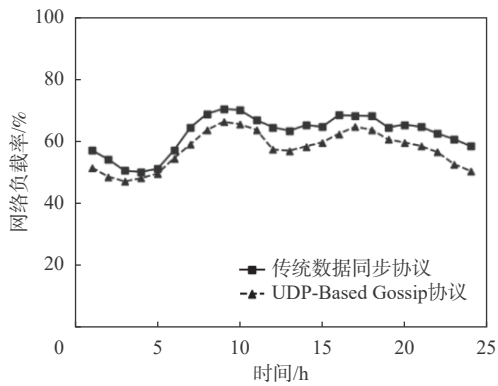


图4 网络负载率对比实验

Fig. 4 Comparison experiment of network load rate

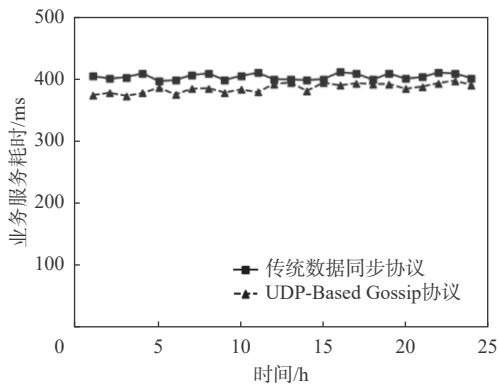


图5 业务服务耗时对比实验

Fig. 5 Comparison experiment of business service time consumption

更好的提升民航运价搜索系统的性能需求[19]。

4　结　论

1) 本文方案设计了一种基于 UDP 传输协议和 Gossip 传播协议的集群状态监控方法。解决了在高负载网络流量的场景下, 集群内业务服务状态的监控和数据同步的问题。

2) 本文方案从网络协议的传输层和应用层着手设计, 在传输层使用 UDP 传输协议来减少集群中节点间的连接数量和交互次数, 从而实现数据在网络传输过程中的低流量、低耗时。在应用层使用 Gossip 传播协议来实现数据的最终一致性, 保证数据传输的可靠性。

3) 通过传输层 UDP 传输协议和应用层 Gossip 传播协议的两方面的结合, 最终保证了集群监控过程中数据同步的高效性和可靠性, 在工程实践领域具有一定的典型性和通用性。

参考文献 (References)

[1] 张宏海, 彭明田, 刘开胜. 基于价格导向的民航运价搜索方案设计[J]. 中国民航大学学报, 2020, 38(3): 40-44.
ZHANG H H, PENG M T, LIU K S. Design of price-based airfare search method[J]. Journal of Civil Aviation University of China, 2020, 38(3): 40-44(in Chinese).

[2] 王成全. 基于 UDP 的消息即时传输系统的研究与设计[J]. 信息技术, 2008(7): 162-164.
WANG C Q. Study and design of real-time message transmission system based on UDP[J]. Information Technology, 2008(7): 162-164(in Chinese).

[3] 刘锦江, 范洪博, 高志伟. 基于 Gossip 网络协议的前置消息验证[J]. 信息技术, 2022(7): 109-113.
LIU J J, FAN H B, GAO Z W. Gossip network protocol-based pre-message authentication[J]. JInformation Technology, 2022(7): 109-113(in Chinese).

[4] BAKER B, SHOSTAK R. Gossips and telephones[J]. Discrete Mathematics, 1972, 2(3): 191-193.

[5] HEDETNIEMI S M, HEDETNIEMI S T, LIESTMAN A L. A survey of gossiping and broadcasting in communication networks[J]. Networks, 1988, 18(4): 319-349.

[6] BOYD S, GHOSH A, PRABHAKAR B, et al. Analysis and optimization of randomized gossip algorithms[C]//Proceedings of the 43rd IEEE Conference on Decision and Control. Piscataway: IEEE Press, 2004: 5310-5315.

[7] WEATHERSPOON H, MIRANDA H, IWANICKI K, et al. Gossiping over storage systems is practical[J]. ACM SIGOPS Operating Systems Review, 2007: 41(5): 75-81.

[8] DEMERS A, GREENE D, HOUSER C, et al. Epidemic algorithms for replicated database maintenance[J]. Acm Sigops Operating Systems Review, 1988, 22(1): 8-32.

[9] MARK J, ALBERTO M, OZALP B. T-Man: Gossip-based fast overlay topology construction[J]. Computer Networks, 2009, 53(13): 2257-2258.

[10] NAOHIRO H, MAKOTO T. Design of the notification system for failure detectors[J]. International Journal of High Performance Computing and Networking, 2009, 6(1): 25-34.

[11] BOYD S, GHOSH A, PRABHAKAR B, et al. Randomized Gossip algorithms[J]. IEEE Transactions on Information Theory, 2006, 52(6): 2508-2530.

[12] GIUSEPPE D, DENIZ H, MADAN J, et al. Dynamo: amazon's highly available key-value store[J]. Symposium on Operating Systems Principles, 2007, 41(6): 205-220.

[13] 李明, 贾智平. 基于 TCP/IP 的分布式监控系统的研究与开发[J]. 计算机工程与应用, 2002, 38(16): 150-153.
LI M, JIA Z P. Research and development of distributed monitoring and control system based on TCP/IP protocol[J]. Computer Engineering and Applications, 2002, 38(16): 150-153(in Chinese).

[14] 王继刚, 顾国昌, 徐立峰, 等. 可靠 UDP 数据传输协议的研究与设计[J]. 计算机工程与应用, 2006, 42(15): 113-116.
WANG J G, GU G C, XUN L F, et al. Research and design of reliable data transfer based on UDP[J]. Computer Engineering and Applications, 2006, 42(15): 113-116(in Chinese).

[15] 刘德辉, 尹刚, 王怀民, 等. 分布环境下的 Gossip 算法综述[J]. 计算机科学, 2010, 37(11): 24-28.
LIU D H, YIN G, WANG H M, et al. Overview of Gossip algorithm in distribute system[J]. Computer Science, 2010, 37(11): 24-28(in Chinese).

[16] 张国印, 李军, 王向辉, 等. 一种基于移动 P2P 改进的 Gossip 算法[J]. 计算机科学, 2013, 40(9): 103-105.
ZHANG G Y, LI J, WANG X H, et al. Improved Gossip algorithm based on mobile P2P networks[J]. Computer Science, 2013, 40(9): 103-105(in Chinese).

[17] 李燚, 顾乃杰, 黄增士, 等. Redis 集群可靠性的研究与优化[J]. 计算机工程, 2018, 44(5): 40-46.
LI Y, GU N J, HUANG Z S, et al. Research and optimization of redis cluster reliability[J]. Computer Engineering, 2018, 44(5): 40-46(in Chinese).

[18] 王峰, 李立新, 曹景源, 等. 发布/订阅系统中的缓存副本一致性研究[J]. 计算机应用, 2016, 36(6): 1510-1514.
WANG F, LI L X, CAO J Y, et al. Research on replication consistency of cache in publish/subscribe systems[J]. Journal of Computer Applications, 2016, 36(6): 1510-1514(in Chinese).

[19] 杨永凯, 彭明田, 王炜东. 集群内高效可靠的数据文件分发方案的设计[J]. 计算机技术与发展, 2019, 29(11): 163-167.
YANG Y K, PENG M T, WANG W D. Design of effcient and reliable data file distribution solution in cluster[J]. Journal of Computer Applications, 2019, 29(11): 163-167(in Chinese).

An efficient cluster data synchronization scheme based on the Gossip protocol

ZHANG Honghai[1, 2, *], CUI Binhao[1, 2], LI Yiming[1, 2], TIAN Feng[1, 2], JIA Yongqiang[1, 2], XIAO Aosan[1, 2]

(1. Research and Development Center, TravelSky Technology Limited, Beijing 101318, China;
2. Beijing Engineering Technology Research Center of Civil Aviation Big Data, TravelSky Technology Limited, Beijing 101318, China)

Abstract: With the rapid development of the civil aviation ticket fare search system business, the internal network traffic scale of the search system cluster continues to grow. Under high-load network traffic, a cluster data synchronization scheme based on the Gossip protocol was developed. This solution is designed from the transport layer and application layer of the network protocol and uses the user datagram protocol (UDP) transmission protocol in the transport layer to reduce the number of connections and interactions between cluster nodes in order to achieve low traffic and low data consumption during network transmission. At the application layer, the Gossip propagation protocol is used to achieve eventual consistency of data and ensure the reliability of data transmission. In the cluster monitoring process, the effectiveness and dependability of data synchronization are guaranteed by the combination of the Gossip propagation protocol at the application layer and the UDP transmission protocol at the transport layer.

Keywords: gossip protocol; data synchronization; consistency algorithm; high-performance cluster; decentralization

Received: 2024-10-16; **Accepted**: 2024-11-08; **Published Online**: 2024-11-19 14:17
URL: link.cnki.net/urlid/11.2625.V.20241119.0932.001
Foundation item: National Natural Science Foundation of China (U2033203)
* **Corresponding author**. E-mail: hhzhang@travelsky.com.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0253

分布式多旋翼/倾转机翼气弹耦合动特性研究

程毅*，赵金瑞，黄水林，余智豪，邓旭东
(中国直升机设计研究所 直升机动力学全国重点实验室，景德镇 333001)

摘 要：基于准线性化隐式气弹建模方法，建立适用于分布式多旋翼/倾转机翼飞行器的耦合柔性多体动力学模型，研究其气弹耦合动力学特性。基于中等变形梁模型和准定常理论，采用Pitt-Peters动态入流模型和Floquet理论，建立求解多旋翼/大展弦比柔性倾转机翼气弹耦合系统动特性的计算方法。在验证理论模型正确性的基础上，研究分布式多旋翼/倾转机翼耦合系结构动特性、回转颤振特性和气弹耦合动响应特性。结果表明：旋翼和短舱对机翼扭转模态影响最大，旋翼与机翼耦合情况下会加大旋翼整体模态振型；增加旋翼个数并将升力桨展开可提高系统低速状态下的稳定性，但增加旋翼个数会降低机翼扭转频率，进而降低颤振速度，增加旋翼有效迎角和机翼攻角可提高系统颤振速度，而增加旋翼转速则会降低系统颤振速度；随前飞速度增加，系统先发生机翼扭转失稳后发生机翼弦向弯曲失稳的回转颤振现象，系统振动响应经历了振动收敛、小幅极限环颤振和大幅多频极限环颤振，其中，机翼颤振形式是垂向、弦向弯曲和扭转运动耦合，其三维耦合效应显著，而旋翼与机翼的模态耦合程度也在不断加深。

关 键 词：分布式推进；柔性多体；气弹动特性；大展弦比；回转颤振
中图分类号：V214+.1
文献标志码：A **文章编号**：1001-5965(2025)05-1637-14

分布式多旋翼/倾转机翼飞行器是一种采用大展弦比机翼与分布式多旋翼[1-2]的新型旋翼飞行器，采用可倾转机翼布局，多个旋翼分布于机翼前缘[3]；构型具备垂直起降与高速前飞能力，旋翼分为升力桨和推进桨，垂直起降时所有旋翼均参与工作，前飞时升力桨收起以减小阻力，由推进桨提供前飞动力。在高速前飞状态，受气动力影响，机翼在升力面方向的结构支撑刚度显著降低，大展弦比机翼结构变形显著，布置在机翼前缘的旋翼受机翼形变或振动的惯性耦合作用将非常强烈，会引起旋翼拉力载荷的脉动，进一步作用于机翼形成强烈的气弹耦合作用，引起经典颤振、回转颤振等典型动力学问题。

NASA于2016年开启分布式电驱动垂直起降技术的探索研究[4-5]，并完成小型验证机GL-10[6-7]的试飞工作。Rothhaar等[8]基于GL-10进行了结构建模和仿真研究，制造了多架验证机开展试飞验证。由于分布式多旋翼/倾转机翼飞行器构型特点，为了布置更多的螺旋桨和降低质量，多采用大展弦比和薄翼型方案，使得机翼扭转频率较低，容易受到螺旋桨和阵风扰动发生颤振失稳，NASA的Helios无人机[9]因为对分布式多旋翼/大展弦比机翼耦合气弹特性的分析不足，导致该无人机在飞行中受阵风扰动后发生解体[10]。

针对大展弦比机翼气弹问题，Patil[11]研究发现，结构大变形引起的几何非线性会改变机翼的振型和频率。Arena等[12]采用梁杆结构模型和修正L-B气动模型对Dowell机翼模型进行气弹分析。

收稿日期：2023-05-17；**录用日期**：2023-07-16；**网络出版时间**：2023-08-02 10:10
网络出版地址：link.cnki.net/urlid/11.2625.V.20230802.0915.002
基金项目：直升机动力学全国重点实验室基金 (2023-HA-LB-067-08)
* **通信作者**. E-mail：chengy012@avic.com

引用格式：*程毅，赵金瑞，黄水林，等. 分布式多旋翼/倾转机翼气弹耦合动特性研究[J]. 北京航空航天大学学报，2025，51(5)：1637-1650. CHENG Y，ZHAO J R，HUANG S L，et al. Research on dynamic characteristics of distributed multi-rotor/tilting wing aeroelastic coupling[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1637-1650 (in Chinese).*

谢长川和杨超[13]研究表明，随着机翼变形幅值的增加，扭转频率降低，弯扭耦合加剧使得机翼颤振临界速度显著较小。付志超等[14]考虑了几何非线性效应的大展弦比机翼颤振模式。许余乾[15]研究了带外挂的大展弦比机翼非线性气弹特性，揭示外挂短舱与机翼间复杂的气动/弹性/惯性耦合机理。

对于旋翼/机翼耦合引起的回转颤振问题，早在1938年，Browne和Taylor从数学解析中发现了该现象，但直到Electra188事故后，回转颤振问题才重新得到关注。倾转旋翼机在高速前飞时也存在回转颤振问题，Yeager和Kvaternik[16]开发的倾转旋翼气弹分析模型，考虑了机翼、短舱、旋翼桨叶自由度，发现机翼气动力可略微提高稳定性，桨叶摆振刚度则对稳定性影响较为明显。Srinivas和Chopra[17]考虑机身运动的影响，建立倾转旋翼全机气弹耦合模型，机身作为柔性体进行有限元建模，考虑机翼对称和反对称模态，研究机翼模态对倾转旋翼机回转颤振稳定性的影响。Slaby和Smith[18]通过改变短舱外形和位置等方法改善倾转旋翼机前飞气弹稳定性。李志权和夏品奇[19]针对倾转旋翼机模型考虑了弹性桨叶、机翼和机身之间的耦合关系，研究了机身自由度、桨毂中心与重心的相对位移、机舱长度等结构参数对系统稳定性的影响。邓旭东和胡和平[20]推导了旋翼/短舱/机翼耦合动力学方程，建立了适用于倾转旋翼构型回转颤振特性分析的配平与特征值求解方法。刘泽宇等[21]建立了一种基于全机配平的无人倾转旋翼机动力学稳定性分析方法。

因此，针对旋翼与机翼间的复杂气弹耦合问题，进行理论模型和气弹分析方法研究，掌握不同飞行状态的气弹耦合模态特性和回转颤振特性，对分布式多旋翼/倾转机翼飞行器动力学设计至关重要。本文基于准线性化隐式气弹建模方法，分析分布式多旋翼/倾转机翼飞行器的结构、气弹耦合动特性及高速前飞状态下的回转颤振特性。在验证了理论模型正确性的基础上，研究分布式多旋翼/倾转机翼耦合系统的气弹耦合动特性，特别是系统回转颤振特性。

1 分析方法

分布式多旋翼/倾转机翼飞行器采用大展弦比机翼设计，其结构变形显著，同时机翼上存在分布式旋翼及短舱引起的集中载荷和质量，在高速前飞时具有复杂的气动/弹性/惯性耦合特性。基于Green应变能张量建立耦合系统应变能和动能方程，并采用Hamilton原理对系统动力学方程进行推导，其表达式为

$$\delta\Pi=\int_{t_1}^{t_2}(\delta U-\delta T-\delta W)\,\mathrm{d}t=0 \tag{1}$$

式中：δU为系统应变能变分；δT为系统动能变分；δW为系统外力虚功。

$$\begin{cases}\delta U=\sum^{N_r}\sum^{N_b}\delta U_{\mathrm{b}}+\delta U_{\mathrm{w}}\\ \delta T=\sum^{N_r}\sum^{N_b}\delta T_{\mathrm{b}}+\sum^{N_r}\delta T_{\mathrm{p}}+\delta T_{\mathrm{w}}\\ \delta W=\sum^{N_r}\sum^{N_b}\delta W_{\mathrm{b}}+\delta W_{\mathrm{w}}\end{cases} \tag{2}$$

式中：下标w、p、b分别表示机翼、短舱、桨叶；N_{b}和N_{r}分别为桨叶和旋翼个数。其中，短舱假定为刚体，不考虑其与机翼和旋翼连接的支撑刚度影响，且忽略气动力只考虑动能。

1.1 坐标系定义

如图1所示，建立耦合系统坐标系，依次建立机翼未变形坐标系（系统惯性坐标系）$x_{\mathrm{I}}y_{\mathrm{I}}z_{\mathrm{I}}$和变形坐标系$x_{\mathrm{w}}y_{\mathrm{w}}z_{\mathrm{w}}$、旋翼短舱坐标系$x_{\mathrm{P}}y_{\mathrm{P}}z_{\mathrm{P}}$、桨毂不旋转坐标系$x_{\mathrm{H}}y_{\mathrm{H}}z_{\mathrm{H}}$和旋转坐标系$x_{\mathrm{R}}y_{\mathrm{R}}z_{\mathrm{R}}$、桨毂平面坐标系$x_{\mathrm{U}}y_{\mathrm{U}}z_{\mathrm{U}}$、桨叶未变形坐标系$x_{\mathrm{B}}y_{\mathrm{B}}z_{\mathrm{B}}$和变形坐标系$\xi_{\mathrm{D}}\eta_{\mathrm{D}}\varsigma_{\mathrm{D}}$。

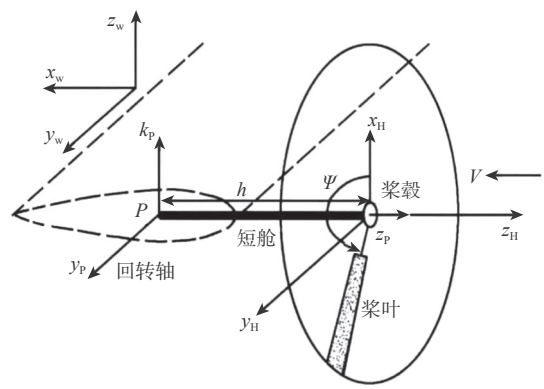


图1 坐标系示意图

Fig. 1 Coordinate system

1.2 结构动力学模型

桨叶或机翼应变能变分表达式为

$$\delta U_{\mathrm{b/w}}=\int_0^L\iint_A(\sigma_{xx}\delta\varepsilon_{xx}+\sigma_{x\eta}\delta\varepsilon_{x\eta}+\sigma_{x\varsigma}\delta\varepsilon_{x\varsigma})\mathrm{d}\eta\mathrm{d}\varsigma\mathrm{d}x \tag{3}$$

式中：σ、ε分别为桨叶或机翼剖面应力、应变，分别由胡克定律和变形协调关系计算得到；η、ς分别为剖面上任意点在剖面坐标系中的坐标；L为旋翼半径或机翼半展长。

桨叶上任一点在惯性坐标系中的矢径达式为

$$\begin{aligned}\boldsymbol{r}_{\mathrm{b}}=&\begin{bmatrix}x_{\mathrm{wh}} & R_{\mathrm{w}P}+y_{\mathrm{wh}} & z_{\mathrm{wh}}\end{bmatrix}+\begin{bmatrix}x_{hP} & y_{hP} & z_{hP}\end{bmatrix}\boldsymbol{T}_{\mathrm{PI}}+\\ &\begin{bmatrix}x+u & v & w\end{bmatrix}\boldsymbol{T}_{\mathrm{UI}}+\begin{bmatrix}0 & \eta & \varsigma\end{bmatrix}\boldsymbol{T}_{\mathrm{DI}}\end{aligned} \tag{4}$$

式中：x_{wh}、y_{wh}、z_{wh}为连接点P处的变形量；$R_{\mathrm{w}P}$为未变形前该点距翼根的长度；x_{hP}、y_{hP}、z_{hP}为桨毂中心到连接点P处的矢径坐标；η、ς为桨叶剖面上任一

点到变距轴的坐标; u、v、w 为桨叶上任一点处的变形量; x 为未变形前该点距桨毂中心的长度; $\boldsymbol{T}_{\mathrm{PI}}$、$\boldsymbol{T}_{\mathrm{UI}}$、$\boldsymbol{T}_{\mathrm{DI}}$ 分别表示短舱坐标系、桨毂平面坐标系、桨叶变形坐标系与惯性坐标系之间的转换矩阵。

桨叶速度为位移矢径对时间求导, 则有 $\boldsymbol{V}_{\mathrm{b}}=\dot{\boldsymbol{r}}_{\mathrm{b}}=\mathrm{d}\boldsymbol{r}_{\mathrm{b}}/\mathrm{d}t$, 桨叶动能变分表达式 $\delta\boldsymbol{T}_{\mathrm{b}}$ 如下:

$$\begin{aligned}\delta\boldsymbol{T}_{\mathrm{b}}=&\int_0^L\left(\iint_A\rho\boldsymbol{V}_{\mathrm{b}}\delta\boldsymbol{V}_{\mathrm{b}}\mathrm{d}\eta\mathrm{d}\varsigma\right)\mathrm{d}x=\\&\int_0^L\left(\boldsymbol{T}_u\delta u+\boldsymbol{T}_v\delta v+\boldsymbol{T}_w\delta w+\boldsymbol{T}_\phi\delta\phi+\boldsymbol{T}_{v'}\delta v'+\right.\\&\boldsymbol{T}_{w'}\delta w'+\boldsymbol{T}_F+\boldsymbol{T}_{x_h}\delta x_h+\boldsymbol{T}_{y_h}\delta y_h+\\&\left.\boldsymbol{T}_{z_h}\delta z_h+\boldsymbol{T}_{\phi_h}\delta\phi_h+\boldsymbol{T}_{\alpha_h}\delta\alpha_h+\boldsymbol{T}_{\psi_h}\delta\psi_h\right)\mathrm{d}x\end{aligned}\tag{5}$$

式中 u、v、w、ϕ 为桨叶自由度; x_h、y_h、z_h、ϕ_h、α_h、ψ_h 为连接点 P 处自由度。

机翼动能变分 $\delta\boldsymbol{T}_{\mathrm{w}}$ 与桨叶类似, 只需考虑机翼自身运动即可; 短舱可视为刚体, 其动能变分 $\delta\boldsymbol{T}_{\mathrm{p}}$ 可用质点惯性力的体积分表达。

1.3　气动模型

桨叶气动力虚功为广义气动力 $\boldsymbol{L}_u$、$\boldsymbol{L}_v$、$\boldsymbol{L}_w$、$\boldsymbol{M}_\varphi$ 相对于桨叶虚位移做的功, 其中, 广义气动力计算是在桨叶变形坐标系中进行的, 翼型剖面相对气流速度来自于: ①来流速度 $\boldsymbol{V}_{\mathrm{f}}$(即前飞速度); ②桨叶运动 $\boldsymbol{V}_{\mathrm{b}}$ 贡献。变换到桨叶变形坐标系下得到剖面相对气流速度 $\boldsymbol{V}_{\mathrm{B}}$ 为

$$\boldsymbol{V}_{\mathrm{B}}=-\boldsymbol{V}_{\mathrm{f}}+\boldsymbol{V}_{\mathrm{b}}=[U_R\quad U_T\quad U_P]\tag{6}$$

式中: U_{R}为桨叶剖面气流速度在垂直于桨盘平面方向的分; U_{T}为桨叶剖面气流速度在桨盘平面切向方向的分; U_{P}为桨叶剖面气流速度在桨盘平面径向方向的分量。

采用片条理论, 对来流速度分量进行推导, 可得翼型剖面的气动合力等于准定常环量(下标 C)和非环量(下标 NC)气动力的叠加, 即

$$\begin{cases}\boldsymbol{L}_u=(\boldsymbol{L}_u)_{\mathrm{C}}\\\boldsymbol{L}_v=(\boldsymbol{L}_v)_{\mathrm{C}}\\\boldsymbol{L}_w=(\boldsymbol{L}_w)_{\mathrm{C}}+(\boldsymbol{L}_w)_{\mathrm{NC}}\\\boldsymbol{M}_{\hat{\phi}}=\left(\boldsymbol{M}_{\hat{\phi}}\right)_{\mathrm{C}}+\left(\boldsymbol{M}_{\hat{\phi}}\right)_{\mathrm{NC}}\end{cases}\tag{7}$$

桨叶气动力虚功变分表达式为

$$\delta\boldsymbol{W}_{\mathrm{b}}=\int_0^L\left[\boldsymbol{L}_u\delta u+\boldsymbol{L}_v\delta v+\boldsymbol{L}_w\delta w+\boldsymbol{M}_{\hat{\phi}}\delta\hat{\phi}\right]\mathrm{d}x\tag{8}$$

机翼气动力虚功与桨叶类似, 但由于分布式多旋翼/倾转机翼飞行器的独特构型, 旋翼相对机翼的位置是固定的, 使得机翼的一部分始终处于旋翼尾流中, 这一段机翼称为滑流区, 其他不受影响的称为自由流区。其中, 自由流区翼型剖面相对气流速度 $\boldsymbol{V}_{\mathrm{w1}}$ 与桨叶一样, 来自于 2 个部分: 来流速度 $\boldsymbol{V}_{\mathrm{f}}$ 和机翼运动速度 $\boldsymbol{V}_{\mathrm{w}}$。

而滑流区需考虑旋翼桨盘入流的影响, 忽略尾流的周向速度, 轴向尾流采用均匀流模拟, 即机翼滑流区旋翼尾流形状为半径收缩的柱形均匀流, 下面给出其半径在滑流区的收缩公式:

$$R_h=R(0.78+0.22\mathrm{e}^{-0.3-2h\sqrt{C_T}-60\sqrt{C_T}})\tag{9}$$

式中: R 和 C_T 分别为旋翼半径和拉力系数; h 为短舱长度; R_h 为滑流区的展向长度。

机翼来流区划分如图 2 所示。

其中, $\boldsymbol{v}_{\mathrm{i}}$ 为诱导入流均值, $\boldsymbol{v}_{\mathrm{i}}$ 为机翼滑流区弦向速度增量。来流满足气体连续性方程, 有以下关系:

$$(U_{Tw}+|\boldsymbol{v}_{\mathrm{i}}|)2\pi R^2=(U_{Tw}+|\boldsymbol{v}_1|)2\pi R_h^2\tag{10}$$

滑流区的剖面来流速度 $\boldsymbol{V}_{\mathrm{w2}}$ 为

$$\boldsymbol{V}_{\mathrm{w2}}=-(\boldsymbol{V}_{\mathrm{f}}+\boldsymbol{v}_1)+\boldsymbol{V}_{\mathrm{w}}\tag{11}$$

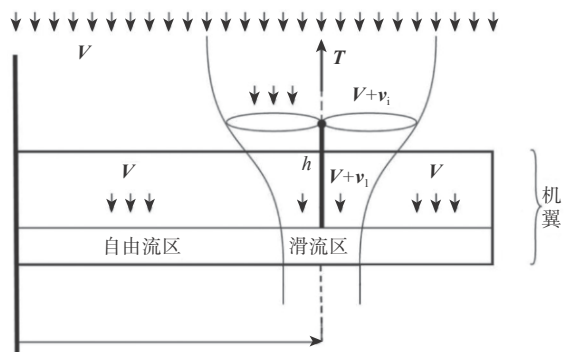


图 2　机翼来流区划分

Fig. 2　Wing inflow zone division

1.4　气动弹性方程

根据 Hamilton 变分原理, 对 δU、δT、δW 线化处理, 耦合结构与气动模型, 再对旋翼进行多桨叶坐标转换, 将各部件矩阵进行组集, 进而推导得到耦合系统的气弹动力学方程为

$$\boldsymbol{M}^{\mathrm{S}}(t)\ddot{\boldsymbol{q}}(t)+\boldsymbol{C}^{\mathrm{S}}(t)\dot{\boldsymbol{q}}(t)+\boldsymbol{K}^{\mathrm{S}}(t)\boldsymbol{q}(t)=\boldsymbol{F}(t)+\boldsymbol{F}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}},t)\tag{12}$$

式中: $\boldsymbol{M}$ 为系统质量阵; $\boldsymbol{C}$ 为阻尼阵; $\boldsymbol{K}$ 为刚度阵上标 S 表示结构阵; $\boldsymbol{q}(t)$ 为系统广义坐标(包含桨叶和机翼自由度); $\boldsymbol{F}(t)$ 为系统常量项力; $\boldsymbol{F}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}},t)$ 为系统结构非线性力项和气动力项。

令方程第 n 次迭代计算后的稳态解为 $\boldsymbol{q}^n(t)$、$\dot{\boldsymbol{q}}^n(t)$、$\ddot{\boldsymbol{q}}^n(t)$, 通过有限差分求得 $\boldsymbol{F}(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}},t)$ 在平衡位置 $\boldsymbol{q}^n(t)$、$\dot{\boldsymbol{q}}^n(t)$、$\ddot{\boldsymbol{q}}^n(t)$的偏导, 其偏导矩阵即为耦合系统的气动 $\boldsymbol{M}$、$\boldsymbol{C}$、$\boldsymbol{K}$ 阵, 用上标 A 表示, 其表达式如下:

$$\begin{cases}\boldsymbol{M}^{\mathrm{A}}(t)=\dfrac{\partial\boldsymbol{F}^n(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}},t)}{\partial\boldsymbol{q}^n}\\\boldsymbol{C}^{\mathrm{A}}(t)=\dfrac{\partial\boldsymbol{F}^n(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}},t)}{\partial\dot{\boldsymbol{q}}^n}\\\boldsymbol{K}^{\mathrm{A}}(t)=\dfrac{\partial\boldsymbol{F}^n(\boldsymbol{q},\dot{\boldsymbol{q}},\ddot{\boldsymbol{q}},t)}{\partial\ddot{\boldsymbol{q}}^n}\end{cases}\tag{13}$$

则由式 (12) 推导可得

$$\boldsymbol{M}(t)\ddot{\boldsymbol{q}}(t)+\boldsymbol{C}(t)\dot{\boldsymbol{q}}(t)+\boldsymbol{K}(t)\boldsymbol{q}(t)=\boldsymbol{F}(t) \tag{14}$$

式中：$\boldsymbol{M}(t)=\boldsymbol{M}^{\mathrm{S}}(t)-\boldsymbol{M}^{\mathrm{A}}(t)$；$\boldsymbol{C}(t)=\boldsymbol{C}^{\mathrm{S}}(t)-\boldsymbol{C}^{\mathrm{A}}(t)$；$\boldsymbol{K}(t)=\boldsymbol{K}^{\mathrm{S}}(t)-\boldsymbol{K}^{\mathrm{A}}(t)$。

则式 (14) 可以写为

$$\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{M}(t)\end{bmatrix}\begin{bmatrix}\dot{\boldsymbol{q}}(t)\\ \ddot{\boldsymbol{q}}(t)\end{bmatrix}-\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}\\ -\boldsymbol{K}(t) & -\boldsymbol{C}(t)\end{bmatrix}\begin{bmatrix}\boldsymbol{q}(t)\\ \dot{\boldsymbol{q}}(t)\end{bmatrix}=\begin{bmatrix}\boldsymbol{0}\\ \boldsymbol{F}(t)\end{bmatrix} \tag{15}$$

式中：$\boldsymbol{I}$ 和 $\boldsymbol{0}$ 分别为 $N\times N$ 阶的单位矩阵和零矩阵。

令 $\boldsymbol{y}(t)=(\boldsymbol{q}(t)\quad \dot{\boldsymbol{q}}(t))^{\mathrm{T}}$，$\boldsymbol{F}(t)$=0，则式 (15) 可写为

$$\begin{cases}\dot{\boldsymbol{y}}(t)=\boldsymbol{A}\boldsymbol{y}(t)+\boldsymbol{G}\\ \boldsymbol{A}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}\\ -[\boldsymbol{M}(t)]^{-1}\boldsymbol{K}(t) & -[\boldsymbol{M}(t)]^{-1}\boldsymbol{C}(t)\end{bmatrix}\end{cases} \tag{16}$$

式 (16) 只有在质量阵 $\boldsymbol{M}(t)$ 不奇异时才能成立。

根据 Floquet 理论，周期方程 (16) 在任意时刻 t 处的解可表示为初始时刻的线性组合：

$$\dot{\boldsymbol{y}}(t)=[\phi(t,t_0)]\,\boldsymbol{y}(t) \tag{17}$$

将式 (17) 代入到式 (16) 有

$$[\dot{\phi}]=[\boldsymbol{A}][\phi] \tag{18}$$

式 (17) 和式 (18) 可组成带初值条件的一阶微分方程，可令 t_0=0，采用形如式 (19) 的 4 阶龙格库塔法求解任意时刻（方位角）对应的转换矩阵 $\boldsymbol{\Phi}(t,0)$。

$$\begin{cases}[\phi_{N+1}]=[\phi_N]+\dfrac{1}{6}(K_1+2K_2+2K_3+K_4)\\ K_1=[\boldsymbol{A}_N][\phi_N]\\ K_2=\left[\boldsymbol{A}\left(t_n+\dfrac{h}{2}\right)\right]\left([\phi_N]+\dfrac{h}{2}K_1\right)\\ K_3=\left[\boldsymbol{A}\left(t_n+\dfrac{h}{2}\right)\right]\left([\phi_N]+\dfrac{h}{2}K_2\right)\\ K_4=[\boldsymbol{A}(t_n+h)]\left([\phi_N]+K_3\right)\end{cases} \tag{19}$$

式中：dt 为时间步长。

周期系统的特征值矩阵为

$$[\boldsymbol{\Lambda}]=\frac{1}{T}\ln[\boldsymbol{\Theta}] \tag{20}$$

式中：$\boldsymbol{\Lambda}$ 为对角阵；T 为系统周期时长；$\boldsymbol{\Theta}$ 为矩阵 $\boldsymbol{\Phi}(T,0)$ 的特征值矩阵，其实部为系统的模态阻尼，虚部为模态频率。

由此，可分析分布式多旋翼/倾转机翼飞行器的气弹动力学特性。

1.5 XV-15 倾转旋翼机

首先，气弹模型，以 XV-15 倾转旋翼机的半展耦合模型系统为验证算例，分析研究机翼在风车状态下的回转颤振特性，其主要动力学参数如表 1 所示。

采用建立的模型和分析方法，对不同前飞速度下机翼模态频率和阻尼进行计算，计算结果如图 3 所示。由图 3(a)可知，机翼一阶弦向弯曲频率和一阶扭转频率随前飞速度增大都略有减小，而一阶垂向弯曲频率变化不大。由图 3(b)可知，机翼一阶垂向、弦向弯曲模态阻尼随前飞速度提高都明显降低，而一阶扭转模态阻尼略有增加；当前飞速度分别大于 165 m/s 和 183 m/s 时，一阶垂向弯曲和一阶弦向弯曲的模态阻尼开始小于零，系统进入失稳状态，机翼发生回转颤振现象。与文献 [22] 对比可知，在计算范围内，模态频率的结果是一致的。一

表 1 XV-15 倾转旋翼机主要动力学参数

Table 1 Main dynamics parameters of XV-15 tilt-rotor

部件	参数	数值
旋翼	片数	3
	洛克数	3.83
	半径/m	3.82
	预锥角/(°)	2.5
	转速/(r·min^{-1})	458
机翼/短舱	一阶垂向弯曲频率q_1/Hz	2.4
	一阶弦向弯曲频率q_2/Hz	5.1
	一阶扭转频率p_1/Hz	5.9

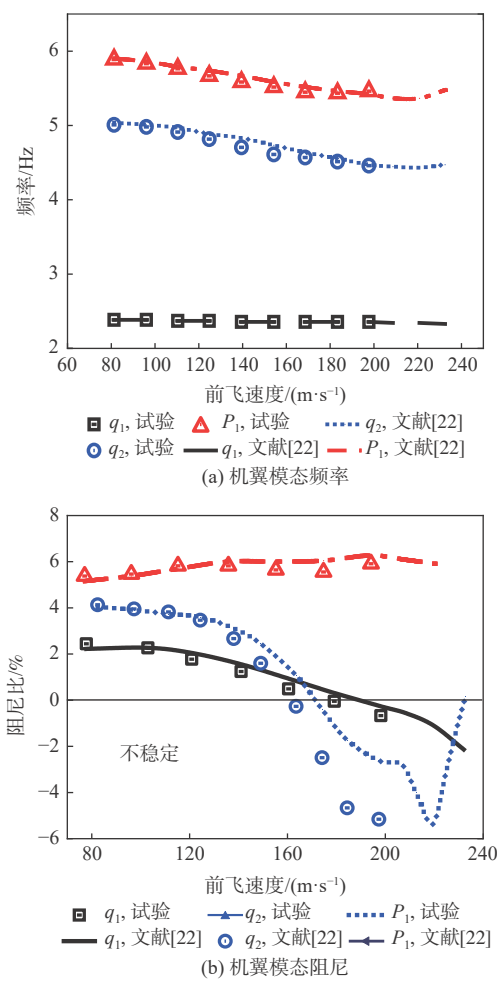


图 3 机翼模态频率和阻尼

Fig. 3 Wing modal frequency and damping

阶垂向弯曲和一阶扭转模态阻尼结果也比较接近; 而一阶弦向弯曲模态阻尼尽管与文献 [22] 在高速状态下有较大偏差, 但其变化趋势是趋同的, 说明建立的模型和分析方法是可靠精确的。

2　分布式多旋翼/倾转机翼飞行器

采用建立的模型和分析方法, 对分布式多旋翼/倾转机翼飞行器半展系统的动力学特性进行分析。系统主要动力学参数如表 2 所示, 各阶耦合频率为机翼/短舱参数。分布式多旋翼/倾转机翼飞行器的半展系统如图 4 所示, 沿机翼展向均匀分布 1 到 3 个旋翼, 旋翼个数分别用 n_1、n_2、n_3 表示, 升力桨展开时用 n_{2z}、n_{3z} 表示。

表 2　耦合系统主要动力学参数

Table 2　Main dynamics parameters of coupling system

参数	数值
旋翼半径R/m	0.7
旋翼片数	4
额定转速Ω/(r·min^{-1})	1 300
短舱质量m_p/kg	3.8
短舱长度h/m	0.35
机翼半展长R_w/m	4
推进桨一阶挥舞β_1^t频率/Hz	33.75
升力桨一阶挥舞β_1^s频率/Hz	42.25
一阶垂向弯曲q_1耦合频率/Hz	3.68
一阶弦向弯曲q_2耦合频率/Hz	9.97
二阶垂向弯曲q_3耦合频率/Hz	19.07
一阶扭转p_1耦合频率/Hz	25.91

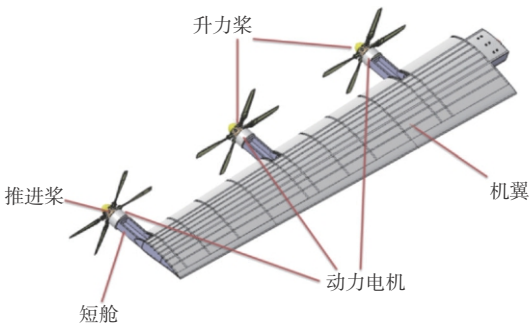


图 4　分布式多旋翼/倾转机翼飞行器

Fig. 4　Distributed multi-rotor/tilting wing aircraft

2.1　耦合系统动力学特性

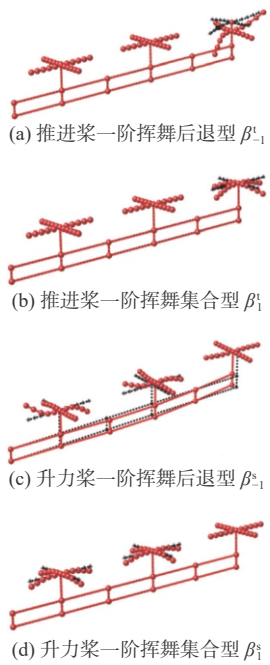
计算耦合系统固有特性, 其耦合系统低阶固有频率如表 3 所示。机翼与旋翼耦合后, 机翼各阶频率都有所下降, 由于旋翼布置于机翼前缘造成机翼重心前移, 使得机翼扭转频率下降最多, 而桨叶各阶频率则变化不大。

旋翼模态振型如图 5 所示。由于桨叶有较大的预扭, 推进桨和升力桨的一阶挥舞模态中均出现了较大挥舞/摆阵模态耦合; 升力桨的一阶挥舞后退型模态与机翼的二阶垂向弯曲模态接近, 故发生了模态耦合, 使得桨叶挥舞后退型模态振型均变大。机翼模态振型如图 6 所示。由于旋翼和短舱在机翼弦向位置存在偏置(布置在机翼前缘), 使得机翼扭转模态与垂向弯曲模态产生负耦合; 而机翼垂向和弦向弯曲模态更易与桨叶摆振和挥舞模态发生耦合。

表 3　半展系统固有频率

Table 3　Natural frequency of half-spread system　Hz

模态	取值
推进桨一阶挥舞后退型β_{-1}^t频率	13.04
推进桨一阶挥舞集合型β_1^t频率	35.77
升力桨一阶挥舞后退型β_{-1}^s频率	18.74
升力桨一阶挥舞集合型β_1^s频率	40.73
机翼一阶垂向弯曲q_1频率	3.337
机翼一阶弦向弯曲q_2频率	9.122
机翼二阶垂向弯曲q_3频率	16.488



(a) 推进桨一阶挥舞后退型 β_{-1}^t

(b) 推进桨一阶挥舞集合型 β_1^t

(c) 升力桨一阶挥舞后退型 β_{-1}^s

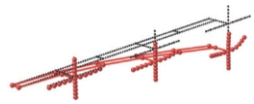
(d) 升力桨一阶挥舞集合型 β_1^s

图 5　旋翼模态振型

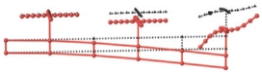
Fig. 5　Rotor mode shape

2.2　耦合系统气弹稳定性

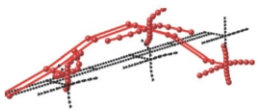
以机翼攻角 3°、桨叶初始有效迎角 6°为基准参数, 在中低速状态, 升力桨展开能提供部分前飞动力, 而随速度增大会将升力桨收起以减小前飞阻力, 计算发现, 旋翼个数和升力桨收放对机翼一阶扭转模态阻尼影响最大。基准参数下, 机翼一阶扭转模态阻尼如图 7 所示。在低速状态下, 增加旋翼个数并将升力桨展开(n_{2z}、n_{3z})会增加扭转阻尼, 这



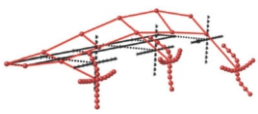
(a) 机翼一阶垂向弯曲 q_1



(b) 机翼一阶弦向弯曲 q_2



(c) 机翼二阶垂向弯曲 q_3



(d) 机翼一阶扭转 p_1

图 6 机翼模态振型

Fig. 6 Wing mode shape

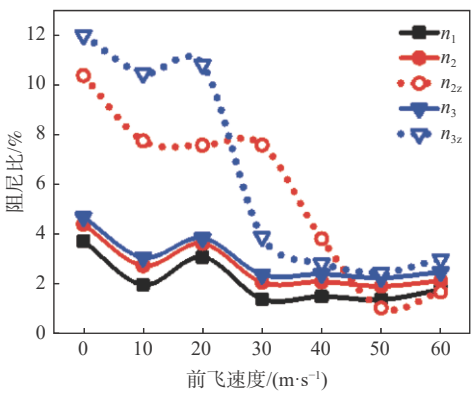
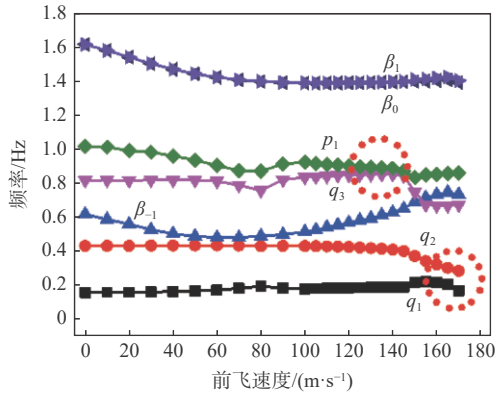


图 7 旋翼个数和升力桨收放对一阶扭转的影响

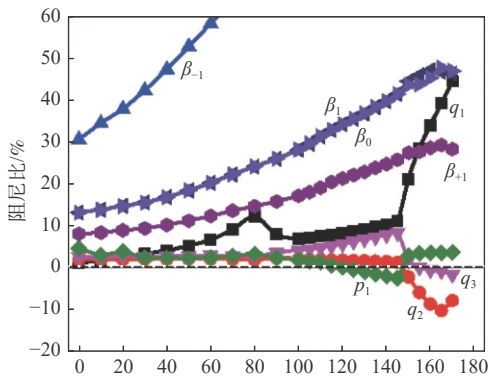
Fig. 7 Influence of number of rotors and lift blade retraction on first order torsion

是由于旋翼个数增加避免了机翼上的载荷和质量过于集中，同时由于滑流区面积增大提高了机翼气动阻尼，但当前飞速度增大、升力桨拉力减小，使得旋翼尾流对滑流区来流速度贡献降低时，升力桨对机翼的增稳作用逐渐减小，机翼扭转稳定性降低。

基准参数下，对有 3 个旋翼且升力桨收起(n_3)的半展系统进行计算，保持桨叶有效迎角不变的情况下，系统随前飞速度变化的模态频率、阻尼和振型如图 8 和图 9 所示。可以看出：①垂向弯曲模态 q_1 和 q_3 的频率和阻尼都随速度增大而增大，而一阶弦向弯曲 q_2 和一阶扭转 p_1 的频率和阻尼都随之减小。②随速度增大，q_3 和 p_1 的频率靠近发生模态耦合，此时机翼弯曲模态 q_1 和 q_2 主振型下的扭转振型都在逐渐增大，而 q_3 和 p_1 模态阻尼则开始减小，在 125 m/s 时，p_1 模态阻尼开始小于 0，机翼发生扭转失稳，系统出现回转颤振现象，此时垂向、弦向弯



(a) 模态频率



(b) 阻尼曲线

图 8 系统模态频率和阻尼

Fig. 8 System modal frequency and damping

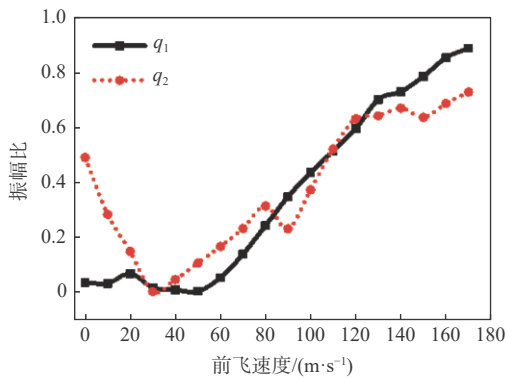
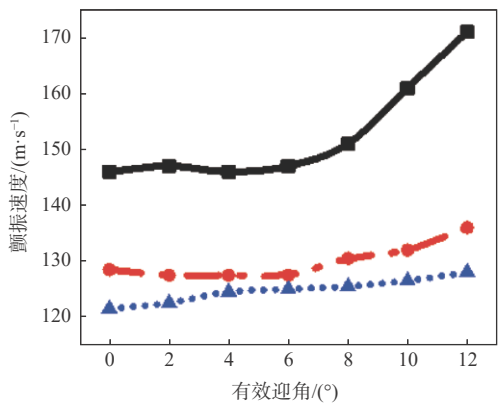


图 9 垂向弯曲主振型下扭转振型

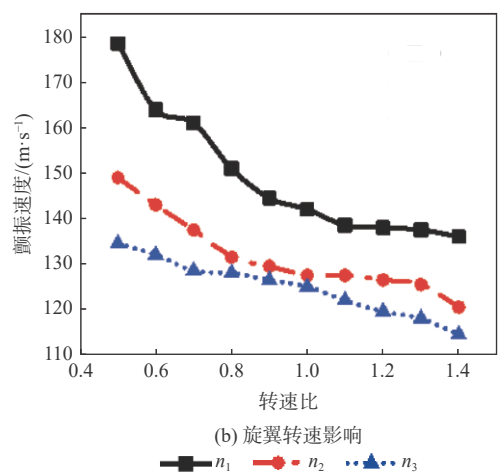
Fig. 9 Torsional mode under vertical bending mode

曲与扭转模态耦合程度明显加剧，机翼颤振的三维耦合效应显著；在 147 m/s 时，q_1 和 q_2 频率靠近发生模态耦合，系统同时出现扭转和弦向弯曲失稳，但随速度增大，q_3 频率逐渐降低并与 p_1 频率拉开，q_3 和 p_1 模态耦合程度降低，p_1 阻尼增大扭转失稳现象逐渐消失。③随速度增大，推进桨一阶挥舞模态阻尼均逐渐增加，集合型 β_1 和无反作用型 β_0 模态频率逐渐减小，而后退型 β_{-1} 的模态频率先减后增加，且在与机翼弦向弯曲模态频率 q_2 靠近时(80 m/s)，使得垂向和弦向弯曲的模态频率和阻尼均出现了拐点。

图 10(a)展示了机翼攻角为 3°时, 不同旋翼有效迎角下系统颤振速度变化。随有效迎角增加, 系统颤振速度也随之增加, 当有效迎角减小到零(类似风车状态)时, 颤振速度最小, 说明旋翼拉力对系统起到增稳作用; 而随着旋翼个数增加, 旋翼有效迎角变化对系统颤振速度影响会减小。图 10(b)为基准参数下不同旋翼转速的颤振速度变化。随旋翼转速增加, 系统颤振速度随之减小, 说明旋翼转速增大会降低系统稳定性。



(a) 旋翼有效迎角影响



(b) 旋翼转速影响

图 10　系统颤振速度随旋翼有效迎角和转速的变化

Fig. 10　Variation of system flutter speed with speed and effective angle of attack of rotor

图 11 展示了旋翼有效迎角为 6°时, 不同机翼攻角下系统颤振速度变化。可知, 随机翼攻角增加, 机翼受到的气动力和系统的气动阻尼均随之增加, 系统颤振速度也随之增加, 且随短舱个数增加, 机翼的扭转频率逐渐降低, 系统更易发生扭转失稳, 颤振速度也随之降低。

2.3　耦合系统气弹响应

分布式多旋翼/倾转机翼飞行器半展系统前飞时的机翼和旋翼的气动力矢量如图 12 所示。

计算基准参数、升力桨展开 n_{3z}、低速(31 ～32 m/s)下的分布式多旋翼/倾转机翼系统气弹时域响应、响应相图和频域。图 13、图 14 分别为机翼

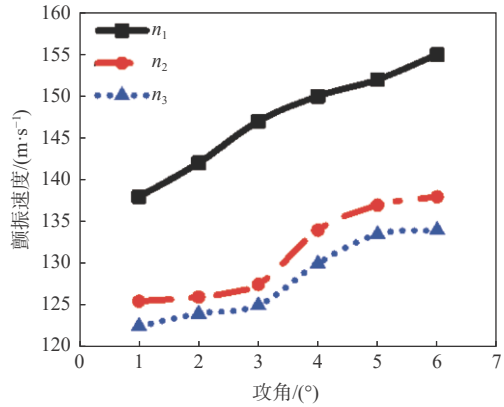


图 11　系统颤振速度随机翼攻角的变化

Fig. 11　Variation of system flutter speed with angle of attack of wing

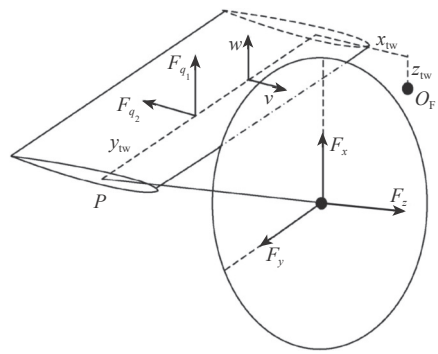
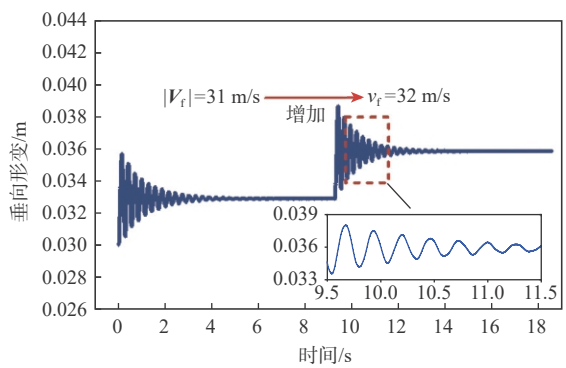
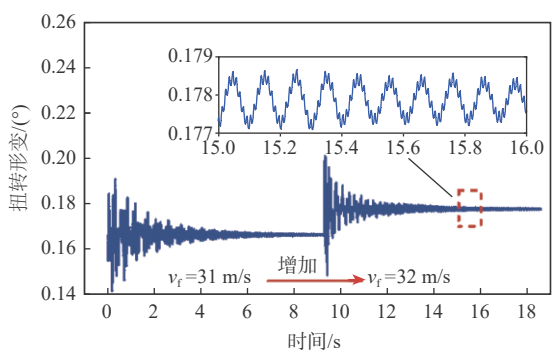


图 12　分布式多旋翼/倾转机翼飞行器前飞气动力

Fig. 12　Aerodynamic forces on distributed multi-rotor/tilting wing aircraft in forward flight



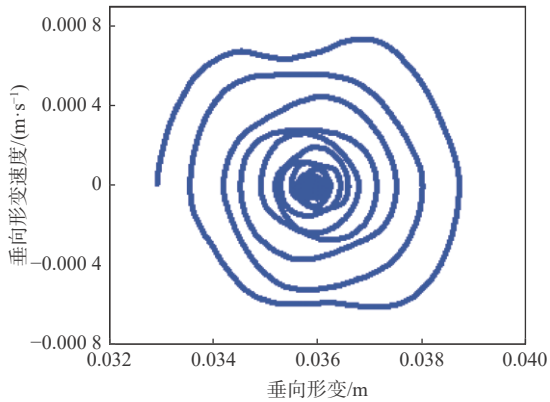
(a) 翼尖垂向弯曲响应



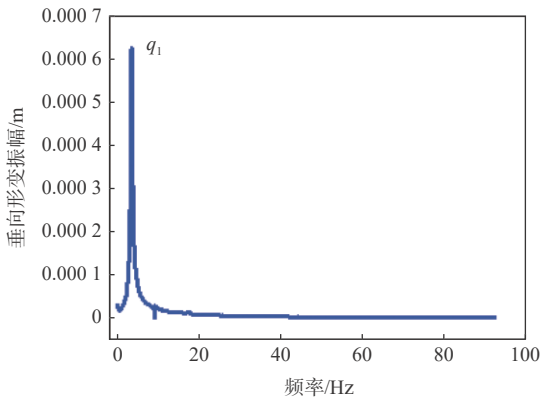
(b) 翼尖扭转响应

图 13　n_{3z} 系统翼尖响应(前飞速度为 31 ～32 m/s)

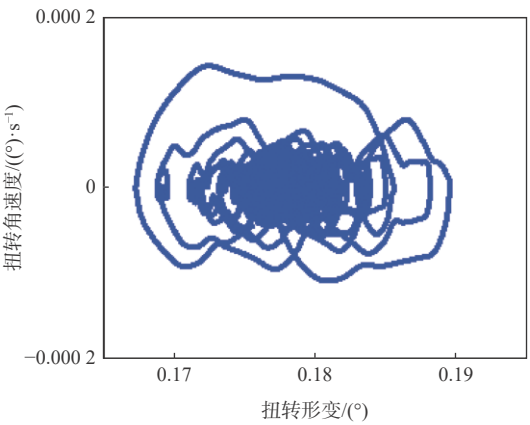
Fig. 13　Wing tip response of n_{3z} system (forward flight speed 31～32 m/s)



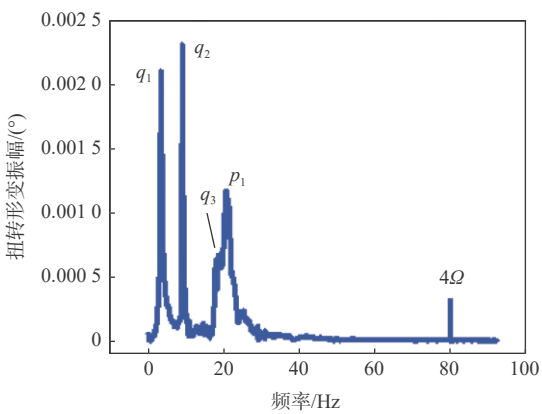
(a) 32 m/s 前飞速度时, 翼尖垂向弯曲响应相图



(b) 32 m/s 前飞速度时, 翼尖垂向弯曲响应 PSD 图



(c) 32 m/s 前飞速度时, 翼尖扭转响应相图



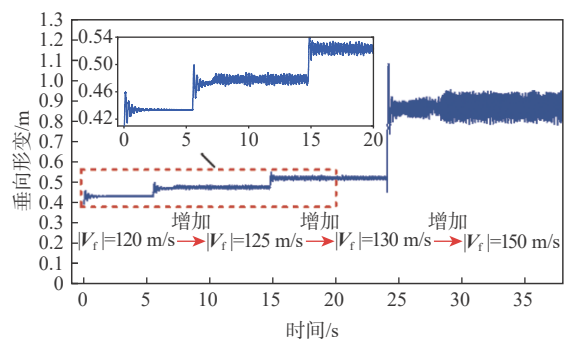
(d) 32 m/s 前飞速度时, 翼尖扭转响应 PSD 图

图 14 翼尖响应相图和 PSD 图

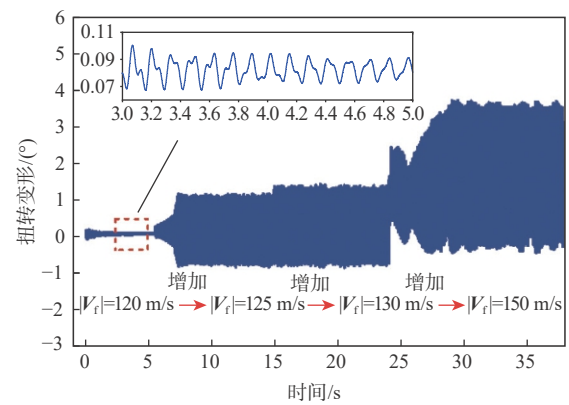
Fig. 14 Phase and PSD diagram of wing tip response

垂向弯曲和扭转时域响应、响应相图、响应频域。可知, 在低速状态下, 机翼垂向弯曲和扭转响应均出现振动收敛; 其中, 机翼垂向弯曲响应振动频率主要为 3.77 Hz 的一阶垂向弯曲模态 q_1, 而机翼扭转响应的振动频谱中有 q_1、q_2、q_3、p_1 和 4Ω 成分, 说明系统振动收敛过程中机翼扭转模态与机翼其余模态及旋翼整体模态发生了耦合。

在高速前飞时, 升力桨因阻力过大收起, 计算基准参数、升力桨收起(n_3)情况下的分布式多旋翼/倾转机翼系统气弹时域响应、响应相图、响应频域。图 15～图 21 分别为机翼和推进桨时域响应、响应相图、响应频域。可知, 当前飞速度(120 m/s)小于回转颤振速度时, 系统响应经历了振动收敛形式, 此时机翼垂向弯曲振动衰减频率主要为一阶垂向弯曲频率, 而扭转振动衰减频谱主要成分却为弦向弯曲频率, 说明此时扭转模态主要与弦向弯曲模态发生了耦合; 此时推进桨桨叶挥舞响应振动频率主要为转速的倍频, 与机翼耦合不明显, 但对于旋翼整体, 其挥舞锥度角和拉力的振动衰减频谱中均有机翼弦向弯曲频率和桨叶挥舞集合型频率的成分, 说明在振动收敛过程中, 旋翼整体模态与弦向弯曲模态发生了耦合。



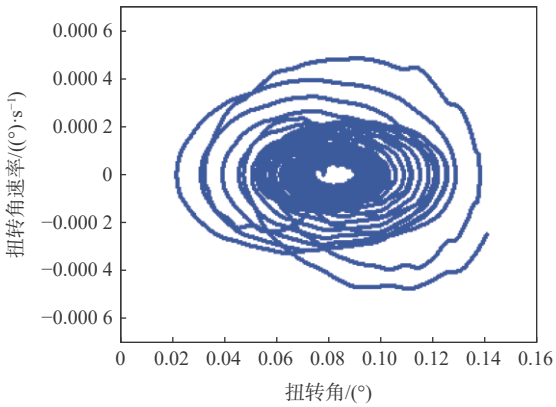
(a) 翼尖垂向弯曲响应



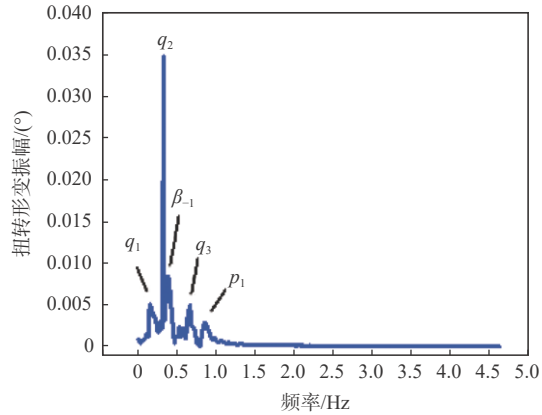
(b) 翼尖扭转响应

图 15 n_3 系统翼尖响应(前飞速度为 120～150 m/s)

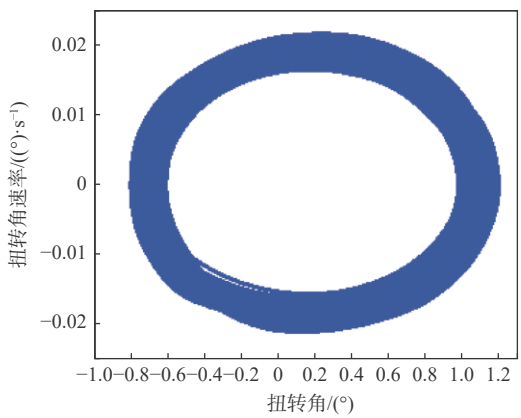
Fig. 15 Wing tip response of n_3 system (forward flight speed 120～150 m/s)



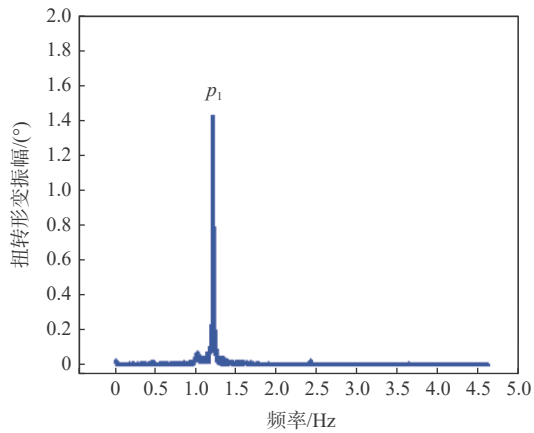
(a) 120 m/s 前飞速度时, 翼尖扭转响应相图



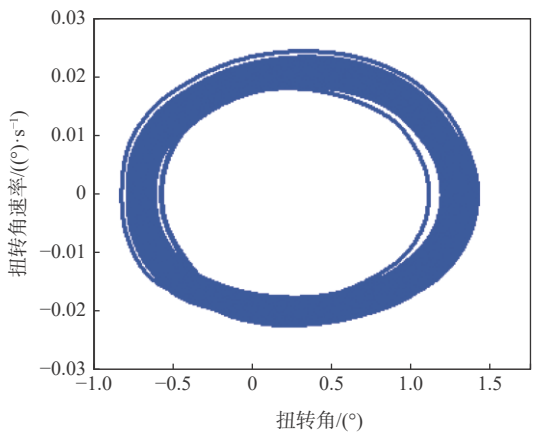
(b) 120 m/s 前飞速度时, 翼尖扭转响应 PSD 图



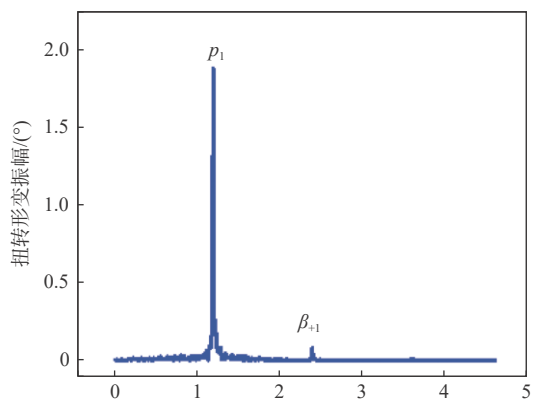
(c) 125 m/s 前飞速度时, 翼尖扭转响应相图



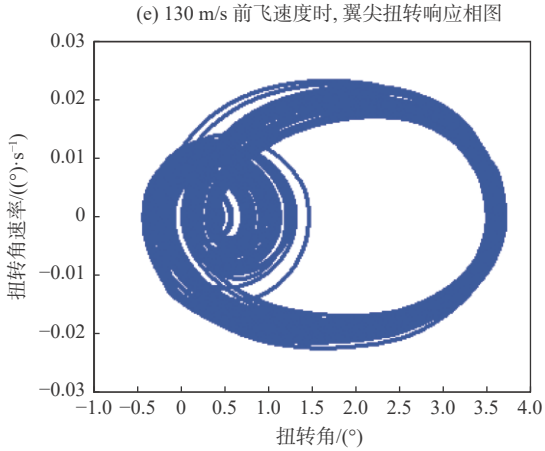
(d) 125 m/s 前飞速度时, 翼尖扭转响应 PSD 图



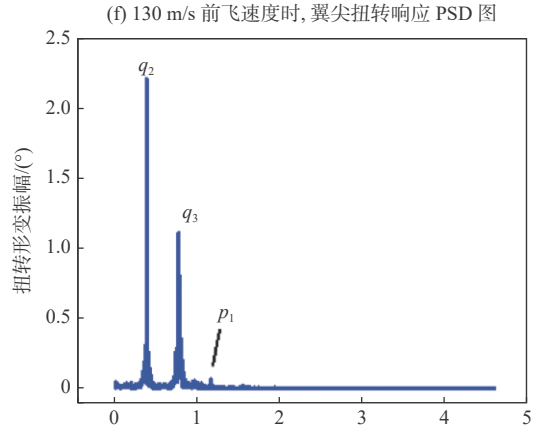
(e) 130 m/s 前飞速度时, 翼尖扭转响应相图



(f) 130 m/s 前飞速度时, 翼尖扭转响应 PSD 图



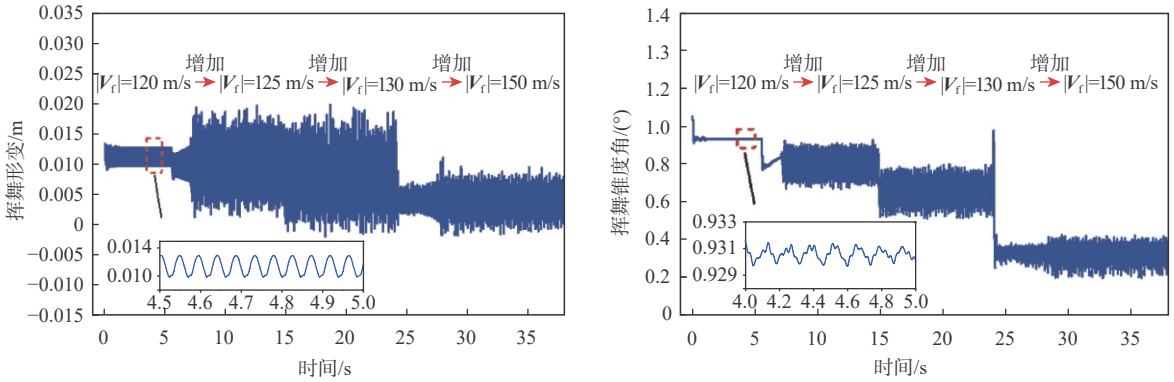
(g) 150 m/s 前飞速度时, 翼尖扭转响应相图



(h) 150 m/s 前飞速度时, 翼尖扭转响应 PSD 图

图 16　扭转响应相图和 PSD 图

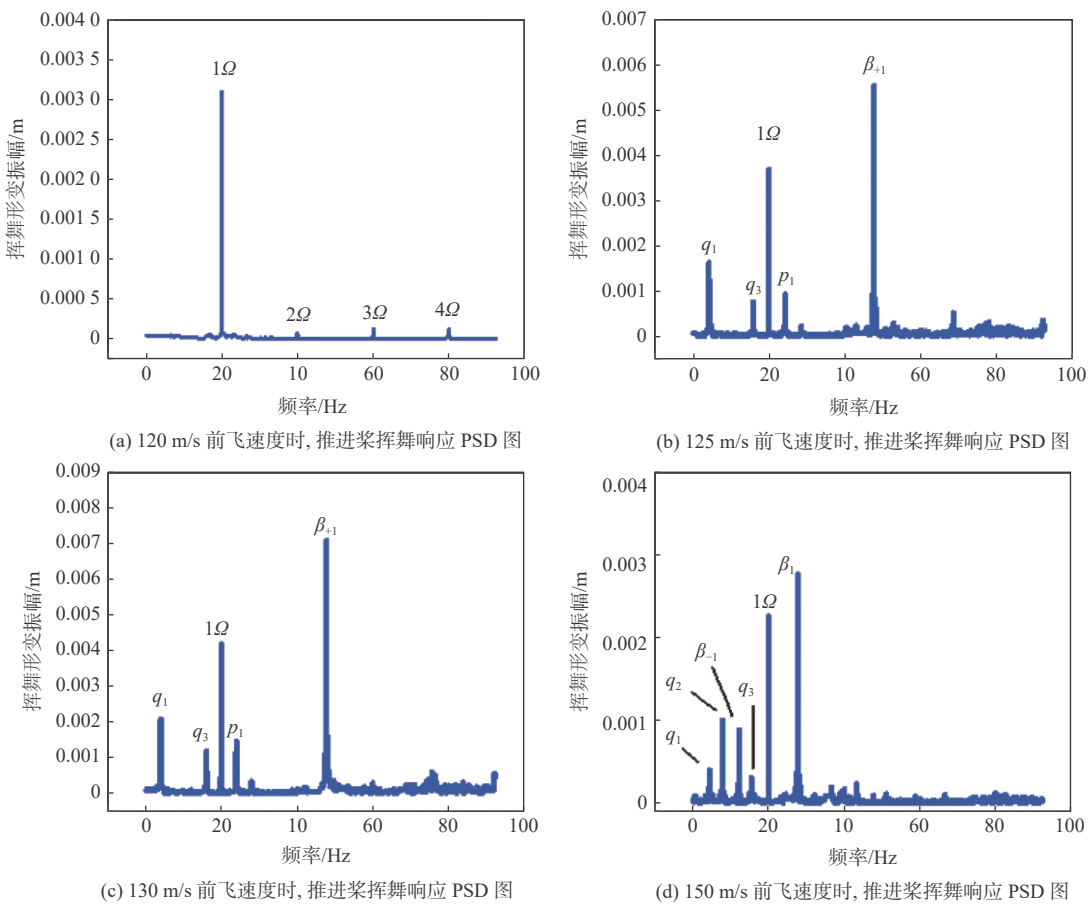
Fig. 16　Torsional response phase and PSD diagram



(a) 翼尖垂向弯曲响应　　(b) 推进桨挥舞锥度角

图 17　推进桨响应(前飞速度为 120～130 m/s)

Fig. 17　Propeller response (120～130 m/s)



(a) 120 m/s 前飞速度时, 推进桨挥舞响应 PSD 图

(b) 125 m/s 前飞速度时, 推进桨挥舞响应 PSD 图

(c) 130 m/s 前飞速度时, 推进桨挥舞响应 PSD 图

(d) 150 m/s 前飞速度时, 推进桨挥舞响应 PSD 图

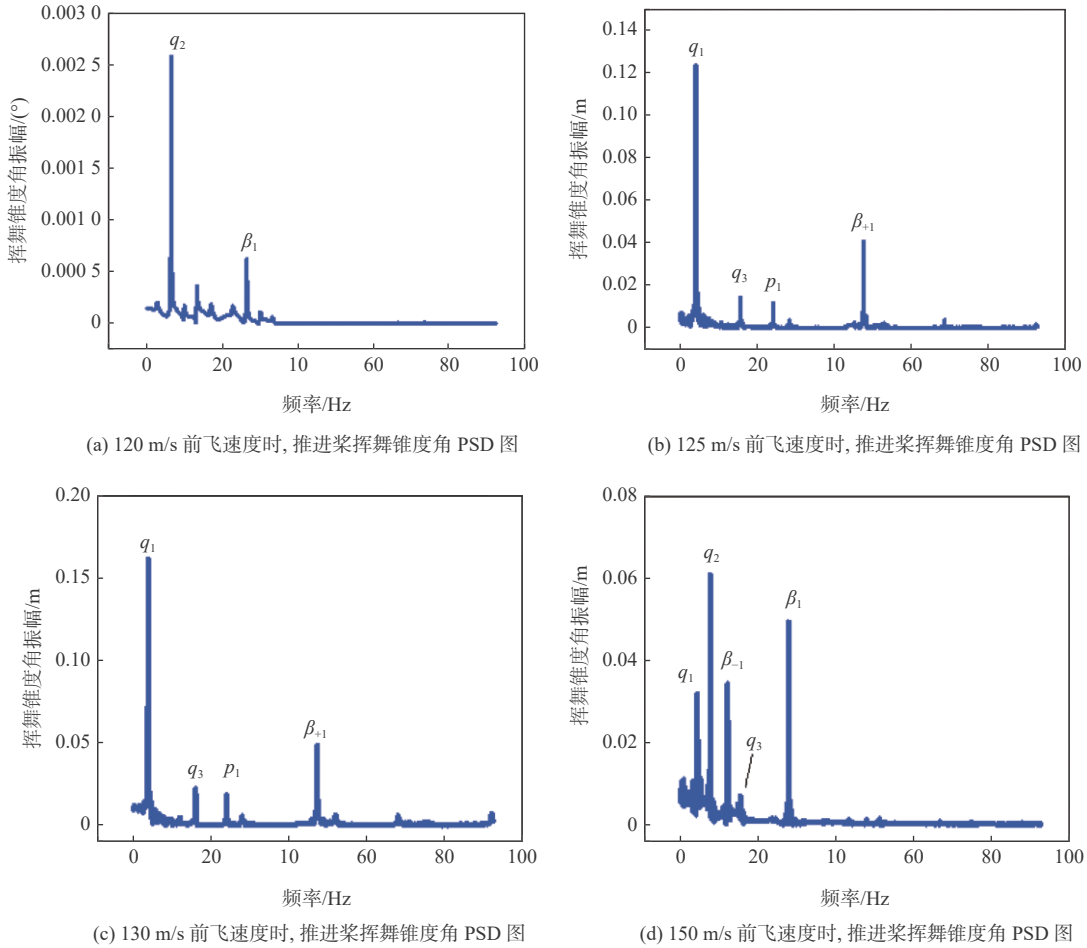
图 18　推进桨挥舞响应 PSD 图

Fig. 18　PSD diagram of propeller flapping response

在前飞速度(125 m/s)等于颤振速度时, 系统发生扭转失稳的回转颤振, 此时机翼响应振动频率主要成分都为扭转频率, 机翼扭转响应也出现了极限环振荡, 机翼颤振呈现了小幅单频的振动形式; 此时推进桨桨叶挥舞响应除了 1Ω 的基频外, 还有挥舞前进型 β^t_{+1} 和机翼的垂向弯曲、扭转频率, 且此时挥舞前进型频率 (51.57 Hz) 恰好约为一阶垂弯曲频率 (26.22 Hz) 的 2 倍, 推进桨挥舞模态与机翼的扭转模态发生倍频耦合; 而推进桨旋翼整体模态频谱成分中同样也包含机翼垂向弯曲和扭转频率, 说明旋翼整体模态同样与机翼模态发生了耦合。

在前飞速度(130 m/s)略大于颤振速度时, 系统仍然发生扭转失稳的回转颤振, 机翼和推进桨响应振动形式和频谱成分与 125 m/s 类似, 只是相应的振动幅值略有变大。

在前飞速度(150 m/s)远大于颤振速度时, 系统发生了以弦向弯曲失稳的回转颤振, 此时机翼

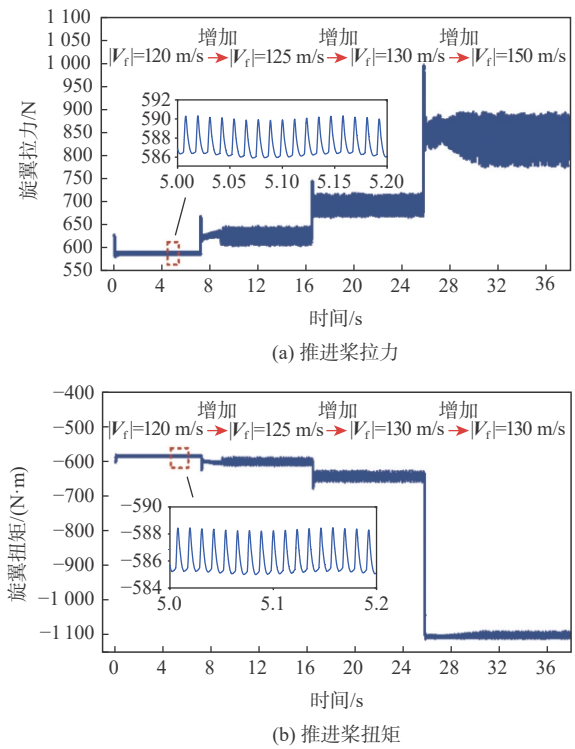


(a) 120 m/s 前飞速度时, 推进桨挥舞锥度角 PSD 图　　(b) 125 m/s 前飞速度时, 推进桨挥舞锥度角 PSD 图

(c) 130 m/s 前飞速度时, 推进桨挥舞锥度角 PSD 图　　(d) 150 m/s 前飞速度时, 推进桨挥舞锥度角 PSD 图

图 19　推进桨挥舞锥度角 PSD 图

Fig. 19　PSD diagram of flapping taper angle of propeller



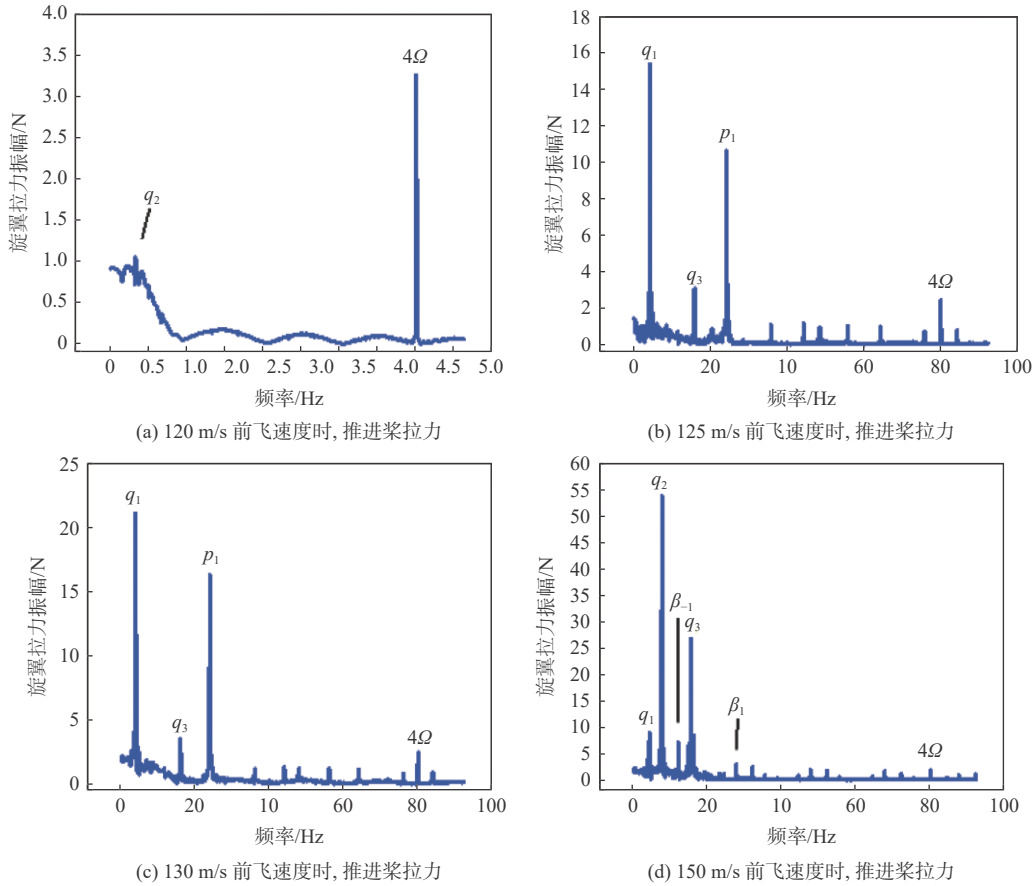
(a) 推进桨拉力

(b) 推进桨扭矩

图 20　推进桨载荷(前飞速度为 120～150 m/s)

Fig. 20　Propeller load (120～150 m/s)

响应的振动频率主要成分都以弦向弯曲频率为主, 且此时二阶垂向弯曲模态频率 (17.07 Hz) 频率恰好约为弦向弯曲频率 (8.43 Hz) 的 2 倍, 机翼二阶垂向弯曲模态与机翼弦向弯曲模态同样也发生了倍频耦合, 使得机翼扭转响应极限环出现分岔, 扭转响应相图中出现 2 个极限环, 同时机翼响应和载荷的振动幅值大大增加, 此时机翼呈现出大幅倍频简谐振动形式。推进桨桨叶挥舞响应除了 1Ω 的基频外, 还有挥舞后退型、集合型和机翼垂向、弦向弯曲频率, 且机翼频率中以一阶弦向弯曲频率为主, 推进桨挥舞模态与机翼的弦向弯曲模态发生耦合; 随着速度增大, 桨叶总距也随之增大到 70°左右, 此时的挥舞刚度大大提高(摆振刚度占主导), 同时由于前飞风阻增加的原因, 尽管旋翼拉力大幅增加, 但挥舞响应反而较 125 m/s 时更小; 而对于推进桨旋翼整体, 模态频谱成分中同样也包含机翼垂向和弦向弯曲频率, 推进桨旋翼整体模态也与机翼弦向弯曲模态发生了耦合。



(a) 120 m/s 前飞速度时, 推进桨拉力

(b) 125 m/s 前飞速度时, 推进桨拉力

(c) 130 m/s 前飞速度时, 推进桨拉力

(d) 150 m/s 前飞速度时, 推进桨拉力

图 21 推进桨拉力 PSD 图

Fig. 21 PSD diagram of Propeller thrust

3 结 论

建立分布式多旋翼/倾转机翼飞行器半展系统动力学理论模型和分析方法, 对半展系统前飞状态气弹耦合动特性进行分析, 掌握了分布式多旋翼/短舱/倾转机翼气弹耦合机理, 并得到以下结论:

1) 旋翼与机翼扭转模态耦合最大, 使得机翼扭转频率下降最多; 旋翼与机翼发生模态耦合会使得耦合振型中桨叶振型变大, 旋翼和短舱的弦向偏置使得机翼扭转与垂向弯曲模态产生负耦合, 机翼垂向和弦向弯曲模态分别更易与桨叶摆振和挥舞模态发生耦合。

2) 增加旋翼个数并将升力桨展开, 可以增加扭转模态阻尼, 降低系统最小前飞速度, 扩大飞行器飞行包线。

3) 垂向弯曲模态频率和阻尼都随速度增大而增大, 而弦向弯曲和扭转的频率和阻尼都随之减小; 系统先发生二阶垂向弯曲与一阶扭转模态耦合的扭转失稳, 后发生一阶垂向、弦向弯曲模态耦合的弦向弯曲失稳; 随前飞速度增大, 旋翼一阶挥舞模态阻尼均逐渐增加, 集合型和无反作用型模态频率逐渐减小, 而后退型的模态频率先减后增加。

4) 旋翼处于风车状态下系统稳定性最差, 增加旋翼有效迎角, 使旋翼带动力工作对系统起到增稳作用; 随着旋翼个数增加, 有效迎角变化对系统颤振速度影响会减小。增加旋翼转速会使系统颤振速度减小, 降低系统稳定性。随机翼攻角增加, 系统颤振速度也随之增加, 且随短舱个数增加机翼扭转频率降低, 系统颤振速度也随之降低。

5) 分析分布式多旋翼/倾转机翼系统时域响应计算结果可得, 随前飞速度逐渐变大, 机翼先发生扭转失稳, 后发生弦向弯曲失稳, 系统经历了振动收敛、小幅单频振动、大幅单频振动、倍频简谐振动形式; 且机翼的垂向、弦向弯曲和扭转运动耦合十分强烈, 机翼颤振的三维耦合效应显著; 旋翼在低速前飞时, 模态频谱中主要有旋翼转速基频的成分, 随前飞速度增大, 桨叶挥舞模态与机翼模态耦合加剧, 其频率成分转为以机翼模态频率为主, 且出现了桨叶挥舞模态的频率。

参考文献 (References)

[1] BORER N K, PATTERSON M D, VIKEN J K, et al. Design and performance of the NASA SCEPTOR distributed electric propulsion flight demonstrator[C]//Proceeding of the 16th AIAA Aviation Technology, Integration, and Operations Conference. Reston: AI-

AA, 2016: 3920.

[2] GOHARDANI A S, DOULGERIS G, SINGH R. Challenges of future aircraft propulsion: a review of distributed propulsion technology and its potential application for the all electric commercial aircraft[J]. Progress in Aerospace Sciences, 2011, 47(5): 369-391.

[3] PATTERSON M D, GERMAN B. Conceptual design of electric aircraft with distributed propellers: multidisciplinary analysis needs and aerodynamic modeling development[C]//Proceeding of the 52nd Aerospace Sciences Meeting. Reston: AIAA, 2014: 0534.

[4] MURPHY P C, LANDMAN D. Experiment design for complex VTOL aircraft with distributed propulsion and tilt wing[C]//Proceedings of the AIAA Atmospheric Flight Mechanics Conference. Reston: AIAA, 2015: 0017.

[5] CHAUHAN S S, MARTINS J R R A. Tilt-wing eVTOL takeoff trajectory optimization[J]. Journal of Aircraft, 2020, 57(1): 93-112.

[6] MURPHY P C, LANDMAN D. Experiment design for complex VTOL aircraft with distributed propulsion and tilt wing[C]//Proceeding of the AIAA Atmospheric Flight Mechanics Conference. Reston: AIAA, 2015: 0017.

[7] BUSAN R C, ROTHHAAR P M, CROOM M A, et al. Enabling advanced wind-tunnel research methods using the NASA langley 12-foot low speed tunnel[C]//Proceeding of the 14th AIAA Aviation Technology, Integration, and Operations Conference. Reston: AIAA, 2014: 3000.

[8] ROTHHAAR P M, MURPHY P C, BACON B J, et al. NASA Langley distributed propulsion VTOL tilt-wing aircraft testing, modeling, simulation, control, and flight test development[C]//Proceeding of the 14th AIAA Aviation Technology, Integration, and Operations Conference. Reston: AIAA, 2014: 2999.

[9] NOLL T E, BROWN J M, PEREZ M E. Investigation of the Helios prototype aircraft mishap[R]. Washington, D.C.: NASA, 2004: 2549.

[10] 杨龙. 大展弦比太阳能无人机结构动力学研究[D]. 长沙: 国防科学技术大学, 2013.
YANG L. Study on structural dynamic of high aspect ratio solar unmanned aerial vehicle[D]. Changsha: National University of Defense Technology, 2013(in Chinese).

[11] PATIL M J. Nonlinear aeroelastic analysis, flight dynamics, and control of a complete aircraft[D]. Atlanta: Georgia Institute of Technology, 1999.

[12] ARENA A, LACARBONARA W, MARZOCCA P. Nonlinear aeroelastic formulation and postflutter analysis of flexible high-aspect-ratio wings[J]. Journal of Aircraft, 2013, 50(6): 1748-1764.

[13] 谢长川, 杨超. 大展弦比飞机几何非线性气动弹性稳定性的线性化方法[J]. 中国科学: 技术科学, 2011, 41(3): 385-393.
XIE C C, YANG C. Linearization method for geometrically nonlinear aeroelastic stability of aircraft with high aspect ratio[J]. Scientia Sinica (Technologica), 2011, 41(3): 385-393(in Chinese).

[14] 付志超, 陈占军, 刘子强. 大展弦比机翼气动弹性的几何非线性效应[J]. 工程力学, 2017, 34(4): 231-240.
FU Z C, CHEN Z J, LIU Z Q. Geometric nonlinear aeroelastic behavior of high aspect ratio wings[J]. Engineering Mechanics, 2017, 34(4): 231-240(in Chinese).

[15] 许余乾. 带外挂细长机翼的非线性气动弹性分析[D]. 哈尔滨: 哈尔滨工业大学, 2019(in Chinese).
XU Y Q. Nonlinear aerolastic analysis for slender wing with store [D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

[16] YEAGER W T, KVATERNIK R G. Contributions of the langley transonic dynamics tunnel to rotorcraft technology and development[C]//Proceadings of the 41st Structures, Structural Dynamics, and Materials Conference and Exhibit, Reston: AIAA, 2000: 1971.

[17] SRINIVAS V, CHOPRA I. Formulation of a comprehensive aeroelastic analysis for tiltrotor aircraft[C]//Proceeding of the 37th Structure, Structural Dynamics and Materials Conference. Reston: AIAA, 1996: 1546.

[18] SLABY J, SMITH E. Aeroelastic stability of folding tiltrotor aircraft in cruise flight with composite wings[C]//Proceeding of the 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. Reston: AIAA, 2011: 2171.

[19] LI Z Q, XIA P Q. Aeroelastic stability of full-span tiltrotor aircraft model in forward flight[J]. Chinese Journal of Aeronautics, 2017, 30(6): 1885-1894.

[20] 邓旭东, 胡和平. 倾转旋翼机螺旋颤振稳定性研究[J]. 空气动力学学报, 2018, 36(6): 1041-1046.
DENG X D, HU H P. Study on whirl-flutter stability of a tiltrotor aircraft[J]. Acta Aerodynamica Sinica, 2018, 36(6): 1041-1046(in Chinese).

[21] 刘泽宇, 招启军, 张夏阳, 等. 倾转机翼无人倾转旋翼机飞行动力学稳定性分析[J]. 飞行力学, 2021, 39(3): 1-7.
LIU Z Y, ZHAO Q J, ZHANG X Y, et al. Analysis of flight dynamics stability of unmanned tilt-rotor aircraft with tilting wings[J]. Flight Dynamics, 2021, 39(3): 1-7(in Chinese).

[22] ACREE C, PEYRAN R, JOHNSON W. Rotor design options for improving XV-15 whirl-flutter stability margins[J]. Journal of the American Helicopter Society, 2001, 46(2): 87-95.

Research on dynamic characteristics of distributed multi-rotor/tilting wing aeroelastic coupling

CHENG Yi*，ZHAO Jinrui，HUANG Shuilin，YU Zhihao，DENG Xudong

(National Key Laboratory of Helicopter Aeromechanics，China Helicopter Research and Development Institute，Jingdezhen 333001，China)

Abstract：Based on the quasi-linear implicit aeroelastic modeling method, a coupled flexible multibody dynamics model suitable for distributed multi-rotor/tilting wing aircraft was established, and its aeroelastic coupling dynamics characteristics were studied. Based on the moderately deformed beam model and quasi-steady theory, the Pitt-Peters dynamic inflow model and Floquet theory were used to establish a calculation method for solving the dynamic characteristics of the aeroelastic coupling system of the multi-rotor/high aspect ratio flexible tilting wing. On the basis of verifying the correctness of the theoretical model, the dynamic characteristics of the distributed multi-rotor/tilting wing coupling system, the whirl flutter characteristics, and the dynamic response characteristics of aeroelastic coupling were studied. The results show that the rotor and the nacelle have the greatest influence on the torsional mode of the wing, and the coupling between the rotor and the wing will increase the overall mode shape of the rotor; increasing the number of rotors and deploying the lift blades can improve the stability of the system at low speed. However, increasing the number of rotors will reduce the torsional frequency of the wings and thus reduce the flutter speed; increasing the effective angle of attack of the rotors and the angle of attack of the wings can increase the flutter speed of the system, and increasing the rotor speed will reduce the flutter speed of the system; as the flying speed increases, the system first occurs the torsional instability of the wing and then the chordwise bending instability of the wing. The vibration response of the system has experienced vibration convergence, small-amplitude limit-cycle flutter, and large-scale multi-frequency limit-cycle flutter. The form of wing flutter is the coupling of vertical and chordwise bending and torsional motion, and its three-dimensional coupling effect is remarkable. The degree of modal coupling between the rotor and the wing is also deepening.

Keywords：distributed propulsion；flexible multi-body；aeroelastic dynamic characteristic；high aspect ratio；whirl flutter

Received：2023-05-17；**Accepted**：2023-07-16；**Published Online**：2023-08-02 10：10

URL：link.cnki.net/urlid/11.2625.V.20230802.0915.002

Foundation item：National Key Laboratory Fund of Helicopter Dynamics (2023-HA-LB-067-08)

* **Corresponding author**. E-mail：chengy012@avic.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0310

低速风洞动压场校核中设备干扰及修正方法

刘江涛[1]，龚小权[2]，周乃春[2,*]，李明[2]，张耀冰[2]

(1. 中国空气动力研究与发展中心 低速空气动力研究所，绵阳 621000；
2. 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：风洞动压场作为风洞流场品质指标之一，直接影响试验结果的精准度，需定期开展动压场校核。动压场校核设备包括五孔探针、排管架及排管架导轨（纵梁和横梁），在分析动压场时需要扣除校核设备影响。基于国家数值风洞工程通用计算流体动力学(CFD)软件NNW-FlowStar，建立了一类用于模拟包含附面层流动的低速压力出口边界条件，采用低速平板湍流边界层验证了该边界条件的残差收敛性及计算精度。基于非结构混合网格及NNW-FlowStar开展了FL-12低速风洞试验段无任何设备、试验段只有排管架导轨、试验段包括排管架和排管架导轨的数值模拟。通过空间动压云图定性展示了各设备的影响量和影响区域，基于空间动压分布曲线定量给出了校核设备对试验段动压的影响量。进一步数值模拟了试验段顶部有无排管架导轨状态，给出了排管架导轨对试验段动压的影响量，排管架导轨对相邻4根探针都有较大影响。通过分解校核设备的动压影响量及对比计算和试验的动压修正曲线，表明排管架导轨对中心探针的动压影响量约为排管架对应量的6倍，验证了试验动压修正方案中忽略排管架影响量的合理性。最后给出了提高动压校核精度的优化建议。

关　键　词：风洞；动压场；影响量；出口边界；修正方法；数值模拟
中图分类号：V211.3
文献标志码：A　　**文章编号**：1001-5965(2025)05-1651-11

风洞流场品质主要是指风洞流场各参数(如压力、温度、密度和气流方向等)分布的均匀程度。针对低速风洞，流场参数主要包括动压、方向场、轴向静压梯度、湍流度、气流温度、动压稳定性和气流噪声等。良好的风洞流场品质是获得可靠试验结果的前提，也是对试验结果进行分析和修正的保证。当风洞洞体有较大变化或者测试仪器有较大更新等情况时，需校核风洞流场。

风洞动压场校核仪器主要是探针排管架，排管架连接于风洞顶部及底部洞壁的导轨横梁上。流场校核时，测量出排管架相对于风洞的安装角，通过移动排管架前后和左右位置获得风洞试验段的动压场及方向场。由于试验段存在排管架及导轨横梁，需要评估导轨横梁、排管架对探针的影响。计算流体动力学(computational fluid dynamics, CFD)方法能够显式提供流场细节，在风洞洞壁干扰、支撑干扰、风洞设计、阻尼网优化等方面发挥了重要作用。王继明等[1]采用CFD方法在高雷诺数风洞中研究了客机腹撑支架干扰影响，获得了支架干扰影响量。代燚等[2]开展了低速风洞全流道数值模拟，分析风洞内部的流动分离，为风洞设计提供了改进方向。李红喆等[3]采用CFD手段研究了连续式跨声速风洞大开角段的阻尼网参数。丛成华等[4]采用CFD手段开展了跨声速风洞开孔壁试验段设计评估优化。

研究人员采用CFD方法开展支撑干扰影响、风洞设计的案例较多，而将CFD运用于风洞动压场及其校核装置影响研究的工作相对较少。本文

收稿日期：2023-06-02；录用日期：2023-09-08；网络出版时间：2023-11-10 18:21
网络出版地址：link.cnki.net/urlid/11.2625.V.20231109.0933.001
基金项目：国家数值风洞工程；国家自然科学基金(11902343)
*通信作者. E-mail：znccxl@qq.com

引用格式：刘江涛，龚小权，周乃春，等. 低速风洞动压场校核中设备干扰及修正方法[J]. 北京航空航天大学学报，2025，51(5)：1651-1661. LIU J T，GONG X Q，ZHOU N C，et al. Equipment interference and correction method in dynamic pressure field verification of low-speed wind tunnel[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1651-1661 (in Chinese).

为准确模拟低速管道流动，在国家数值风洞工程通用 CFD 软件 NNW-FlowStar[5] 上发展了一类修正的低速压力出口边界，并采用低速平板湍流边界层验证这类修正压力出口边界的残差收敛性及计算精度。基于建立的边界和软件数值模拟了中国空气动力研究与发展中心 FL-12 风洞试验段及用于动压场测量的探针排管架和导轨横梁，获得了试验段无设备、试验段只包含导轨、试验段包含排管架/导轨/探针这 3 类模型的流场，通过分析干扰流场，给出排管架影响量。进一步数值模拟了排管架在 y=0, 0.4, 1.0 m 这 3 处，排管架上端有无导轨横梁状态下的风洞流场，给出风洞试验段动压场，在此基础上分析了导轨横梁对风洞动压场的影响，建立了风洞排管架及导轨动压场影响的修正方法，并与试验的动压场修正结果进行对比分析，为风洞试验段动压校核中支架干扰修正提供技术支撑。

1 数值方法

1.1 软件介绍

数值模拟软件为中国空气动力研究与发展中心基于国家数值风洞工程研发的通用 CFD 软件 NNW-FlowStar，其基于非结构格心型有限体积方法和大规模并行计算技术开发。软件内核 MFlow 是基于非结构混合网格的用于求解亚跨超声速流场的专业 CFD 求解器[6-9]，能够处理任意形状的网格单元，具有较强的灵活性。针对低速风洞问题，为保证格式足够的间断和黏性分辨率，无黏通量计算选用 Roe 通量差分分裂格式[10]。非结构单元内引入线性重构，保证空间离散具有二阶精度，单元内变量梯度采用节点型 Gauss 方法[11] 计算，该方法的宽模板特性保证计算鲁棒性和精度。采用 Venkatakrishnan[12] 限制器实现良好的主控方程残差收敛性、计算精度及计算鲁棒性。黏性通量采用中心差分离散。计算假定流场为全湍流，本文研究的流动几乎无分离，因此，选用 NNW-FlowStar 软件中常用的 Spalart-Allmaras 一方程[13] 湍流模型，NNW-FlowStar 软件中的该模型经过了大量的标模和型号应用验证，数据可信度高。主控方程和湍流方程时间迭代均采用 LU-SGS 隐式[14]。软件采用几何多重网格、局部时间步长、变 CFL 数、基于消息传递接口 (message passing interface, MPI) 的大规模并行计算[15] 等一系列加速收敛技术。

软件采用守恒形式的非定常 Navier-Stokes 方程：

$$\frac{\partial}{\partial t}\iiint_V \boldsymbol{Q}\mathrm{d}V+\oiint_{\partial\Omega}\check{\boldsymbol{H}}(\boldsymbol{Q})\check{\boldsymbol{n}}\mathrm{d}S=\oiint_{\partial\Omega}\check{\boldsymbol{H}}_{\mathrm{v}}(\boldsymbol{Q})\check{\boldsymbol{n}}\mathrm{d}S \tag{1}$$

式中：$\boldsymbol{Q}=\left[\rho\ \ \rho u\ \ \rho v\ \ \rho w\ \ \rho E\right]^{\mathrm{T}}$ 为守恒变量，ρ 为密度；u、v、w 分别为笛卡儿坐标系下 3 个方向的速度分量；E 为单位质量的总能；$\check{\boldsymbol{H}}$ 为无黏通量；$\check{\boldsymbol{H}}_{\mathrm{v}}$ 为黏性通量；$\check{\boldsymbol{n}}$ 为控制体单元边界面 $\partial\Omega$ 的外法矢；V 为控制体单元微元体的体积；S 为控制体边界面微元面的面积。

采用格心型有限体积方法离散方程 (1)，基于虚拟单元方法实现固壁、远场、对称面、进出口等边界条件。

1.2 修正的低速压力出口边界

包含附面层的出口边界有 2 类设置方式：①直接将其按亚声速出流边界条件处理[16]；②设置为压力出口边界[17]。两者的主要差别在于虚拟单元体心压力值。本文针对这类低速流动出口边界，在 NNW-FlowStar 软件发展了一种修正的低速压力出口边界。图 1 给出了边界面及边界单元示意图，虚拟单元体心点的压力采用来流的参考压力，来流参考压力设置为无穷远处来流压力。虚拟单元体心的密度和速度采用边界单元的体心值，如下：

$$\begin{cases}\rho_b=\rho_a\\ u_b=u_a\\ v_b=v_a\\ w_b=w_a\\ p_b=p_\infty\end{cases} \tag{2}$$

式中：p_∞ 为来流的静压；a 表示内部单元体心，b 表示远场边界虚拟单元的体心；p 为压力。

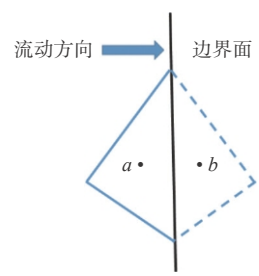


图 1 边界条件示意图

Fig. 1 Schematic diagram of boundary condition

1.3 修正的低速压力出口边界验证

本节采用零压力梯度二维平板来验证建立的修正出口边界条件。计算采用美国航空航天管理局兰利研究中心 (http://turbmodels.larc.nasa.gov/) 提供的标准非结构六面体网格，网格为 grid_quad_545x385_vol.cgns[18]，如图 2 所示。附面层网格第 1 层高度为 5×10^{-7} m，壁面沿流向第 1 个网格单元流向尺寸为 0.001 m。

计算来流马赫数为 0.2，单位雷诺数为 5×10^6，温度为 300.0 K。设置如下边界条件：x=−1/3 m、z=1 网格面均设为远场边界面；网格展向 y=0 和 Y=1 m 网格面设为对称面；壁面 z=0 m 网格面 ($0<x<2$) 为无滑移绝热壁；壁面前端 z=0 m 网格面 ($-1/3<x<0$) 设为对称面；x=2 m 网格面设为修正的压力出口边界面。

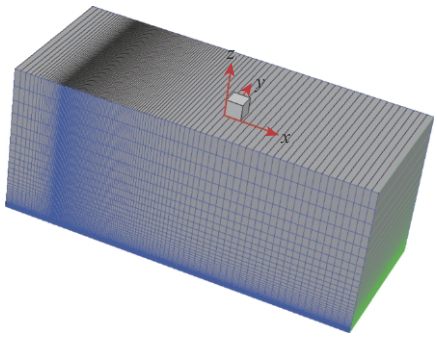


图 2　低速平板湍流边界层计算网格

Fig. 2　Computation grid of low-speed flat plate turbulent boundary layer

图 3 给出了低速平板湍流边界层算例采用修正低速压力出口边界及亚声速出流边界得到的密度残差收敛曲线, 图中给出了不同 CFL 数及 Venkatakrishnan 限制器不同系数 K 的对比曲线。可以看到, 本文建立的修正压力出口边界的密度残差具有很好的收敛性, 而亚声速出流边界的密度残差收敛滞止, 即使减小 CFL 数或减小限制器系数, 仍然不能改变这种现象, 验证了本文建立的修正低速压力出口边界在残差收敛性方面的优势。由图 3 可知, 该算例收敛需要较多的迭代步数, 主要由 2 个方面的原因引起: ①网格的附面层第 1 层间距为 5×10^{-7} m, 间距太小导致数值模拟中附面层发展缓慢; ②展向只有一排单元, 此时 NNW-FlowStar 未开启多重网格功能。

图 4 给出了 NNW-FlowStar、CFL3D[19]、FUN3D[19] 计算的壁面摩擦力系数 C_f 对比结果。可以看出, 三者除壁面起始位置外的绝大部分位置吻合一致。

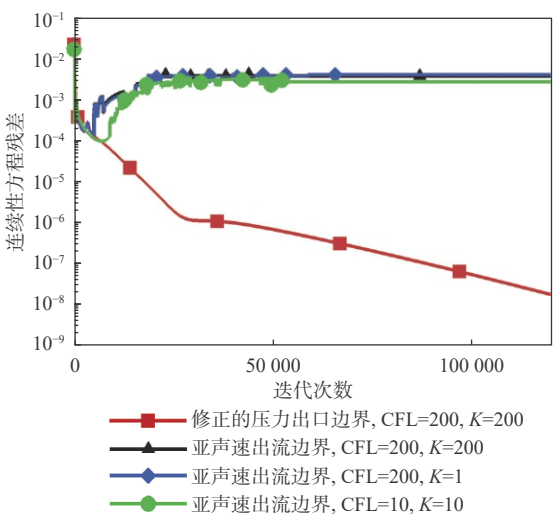


图 3　低速平板湍流边界层不同出口边界下主控方程收敛曲线

Fig. 3　Convergence curves of main control equation under different outlet boundaries of low-speed flat plate turbulent boundary layer

图 5 给出了 NNW-FlowStar、CFL3D 在 x=0.970 08 m 及 x=1.903 34 m 截面处的速度沿 z 向分布, 其中, U 为流向速度, U_{ref} 为来流速度。为进一步比较 2 套软件的法向速度分布, 图 5 纵坐标以对数形式展示了两者在附面层内的速度分布。可知, 2 套软件在 2 个站位的速度分布吻合一致, 验证了本文建立的修正压力出口边界条件及 NNW-FlowStar 软件对该类问题的模拟精度。

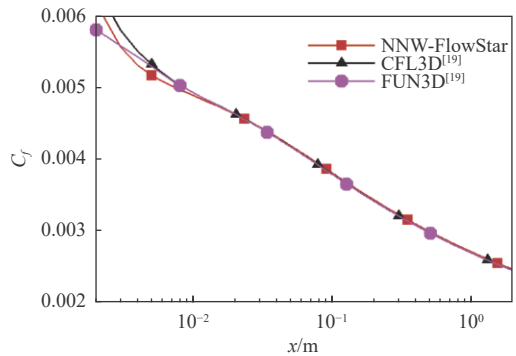


图 4　不同软件的平板摩擦力曲线

Fig. 4　Friction curves of plate by different software

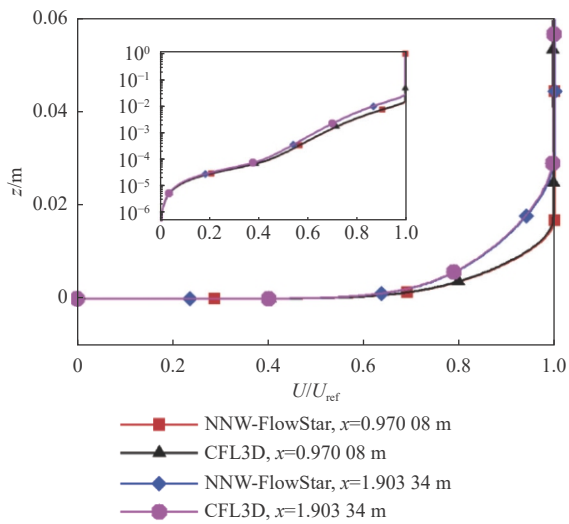


图 5　不同软件在 x=0.970 08 m 及 x=1.903 34 m 截面处的速度特性

Fig. 5　Velocity profiles of different software at x= 0.970 08 m and x= 1.903 34 m

2　低速探针排管架计算模型及网格

2.1　计算模型

4 m×3 m 风洞是中国空气动力研究与发展中心的一座单回流闭口试验段低速风洞, 试验段横截面为切角矩形, 有关该风洞的详细资料可见文献 [20]。为节省网格量, 本文仅模拟了风洞的试验段及试验段前的收缩段, 如图 6 所示。试验段包括上下部导轨、排管架及排管架探针, 排管架上安装 15 根探针, 用于测量空间流场总压、静压, 探针从试验段顶部向底部依次编号。

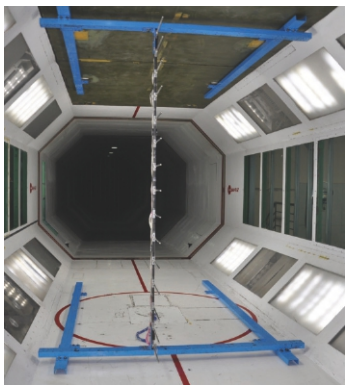


图 6 风洞动压场校核的排管架及导轨

Fig. 6 Pipe rack and guide rail for dynamic pressure field verification of wind tunnel

数值模拟的坐标系定义如图 7 所示。坐标原点位于试验段中心，x 轴沿着试验段气流方向，y 轴指向试验段侧面，z 轴沿地面法线指向试验段顶部。来流从 x 轴负向吹向正向。

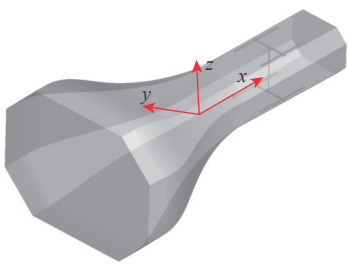


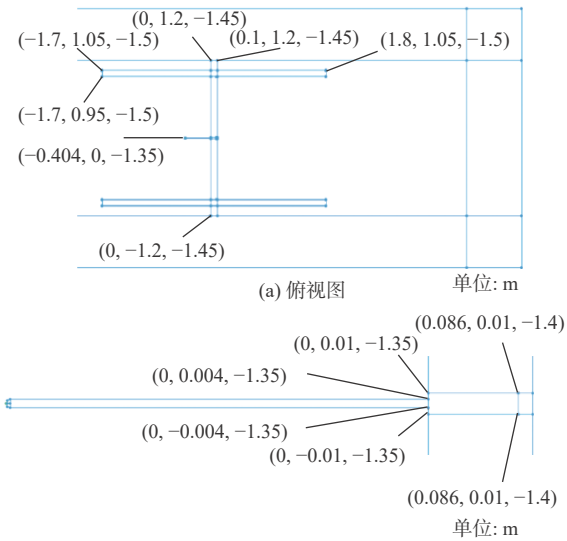
图 7 收缩段、试验段、排管架及导轨模型

Fig. 7 Contraction section, test section, pipe rack, and guide rail model

试验段底部位于 z=−1.5 m，试验段顶部位于 z=1.5 m，导轨宽 0.1 m，高 0.05 m，厚度 0.005 m，导轨前端位于 x=−1.7 m 处，导轨侧面距壁面 0.15 m，排管架宽 0.086 m，高 0.02 m，探针直径 0.008 m，探针最前端位于 x=−0.404 m 处，探针顶端倒圆半径为 0.004 m，探针从上至下依次编号。图 8 给出了探针、导轨、横梁的位置信息。

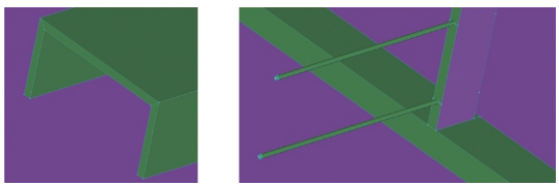
2.2 计算网格

计算网格如图 9 所示。计算中，亚声速入流边界给定入口的总温、总压和速度，总压 P_0=97 184.94 Pa、总温 T_0=286.69 K，来流向速度 V=10 m/s，物面边界采用无滑移等温壁，修正的压力出口边界按照 1.2 节给出。网格总量约为 2 667 万，包括 1 199 万四面体网格、39 万三棱柱网格、1 178 万六面体网格及 251 万金字塔网格单元。为保证网格无关性，本文在生成网格时分为试验段校核设备区域和校核设备以外的其他区域。有无校核设备状态都保证校核设备以外区域的网格完全一致，网格只在试验段校核设备区域不同，同时保证有无校核设备状态都采用相同的加密方式加密空间网格。



(a) 俯视图

(b) 探针与排管架



(c) 导轨前端 (d) 横梁与排管架

图 8 排管架、导轨、探针位置及形状

Fig. 8 Position and shape of pipe rack, guide rail, and probe

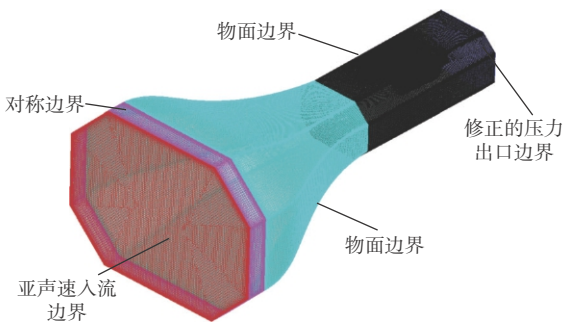


图 9 计算模型边界设置

Fig. 9 Boundary settings of calculation model

从数值模拟角度，根据模拟低速平板边界层的经验，为提高计算鲁棒性，在风洞收缩段前添加对称边界面，相当于增加无黏固壁边界，使气流从无黏逐步过渡到黏性，减弱收缩段最前端奇性，如图 9 中的粉色部分所示。为保证数值模拟计算精度，加密排管架及探针周围的空间流场区域，如图 10 所示。图 11 给出了排管架和探针的表面网格局部放大图。可以看到，其表面网格基本都采用四边形，只在探针头部顶点附近采用各向同性三角形。图 12 给出了试验段对称面空间网格示意图，图中将排管架及探针附近网格放大以便能较清楚观测到探针周围空间局部网格。从图 12 看到，收缩段前部对称面区域网格采用六面体网格，在收缩段、试验段、导轨及排管架附面层采用各向异性的六面体网

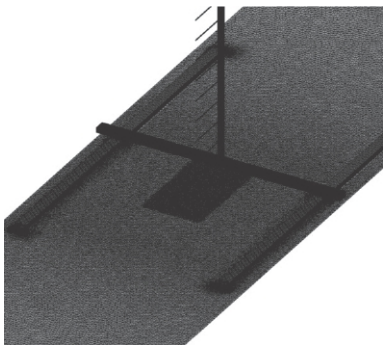


图 10　探针、导轨、排管架计算网格

Fig. 10　Computation grid of probe, guide rail, and pipe rack

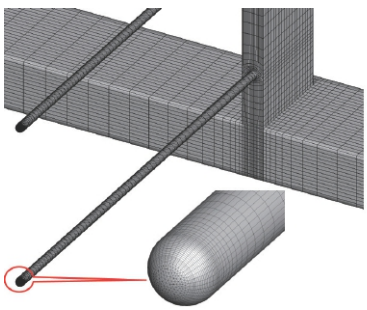


图 11　排管架和探针表面网格局部放大图

Fig. 11　Partial enlarged drawing of grid on pipe rack and probe surface

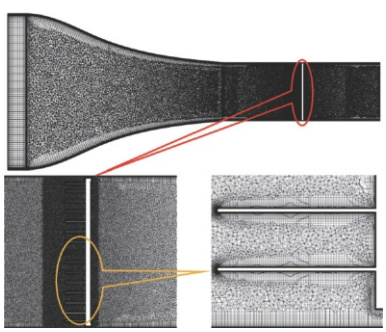


图 12　排管架和探针对称截面空间网格

Fig. 12　Spatial grid of symmetrical profiles of pipe rack and probe

格, 空间各向同性区域采用四面体网格, 两者之间采用金字塔网格过渡。

3　计算分析

数值模拟条件按风洞条件给定, 来流风速 V=70 m/s, 计算高度 H=0.6 km。计算采用 Roe 格式、高斯节点梯度计算方法、Venkatakrishnan 限制器及 LU-SGS 隐式时间迭代, 并采用四重网格。隐式时间迭代过程的 CFL 数在前 1 000 个迭代步逐步从 5 增加到 20, 并采用局部时间步长加速收敛。流场收敛后, 采集 15 根探针头部驻点压力作为测量点的总压, 离探针前端 52 mm 处, 探针表面均布的 8 个点的静压压力平均值作为当地的静压, 探针测量点动压等于总压减去静压。

3.1　低速探针导轨/排管架影响分析

设备干扰分析过程数值模拟了 3 个状态, 分别为试验段包含导轨/排管架/探针 (模型 1)、试验段包含导轨 (模型 2)、试验段无任何设备 (模型 3)。

图 13 给出了数值模拟模型 1 状态的主控方程及湍流方程残差收敛曲线。可知, 主控方程残差下降了 7 个量级, 湍流方程残差下降了 3 个量级, 进一步验证了本文发展的修正压力出口边界较好的残差收敛性。图 14 给出了风洞及排管架的升阻力系数。由于试验段和排管架上下对称, 其升力系数理论上等于 0, 从曲线看到升力系数收敛到 10^{-6}, 从而验证了本文建立的边界条件的计算精度。

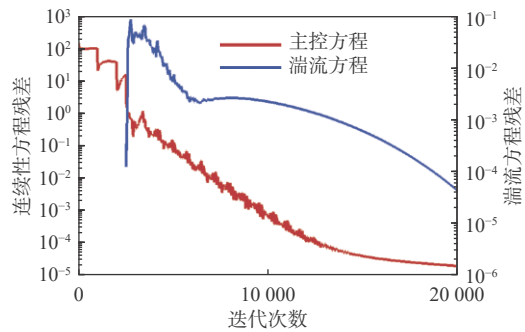


图 13　模型 1 状态的密度残差收敛曲线

Fig. 13　Density residual convergence curves of model 1

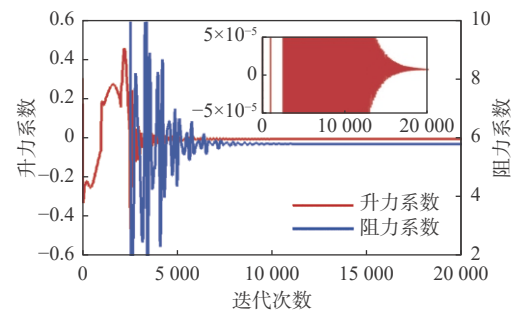


图 14　模型 1 状态的升阻力曲线

Fig. 14　Lift and drag curves of model 1

图 15 给出了模型 1 状态在 y=0 m 截面(左右的中间截面)处的压力分布。可以看到, 排管架及导轨影响其前部试验段气流压力, 越靠近排管架及导轨, 气流压力越大, 靠近导轨的上下各 2 根探针几乎完全处于导轨的高压区内, 因此, 这 4 根探针的测量数据不可用。图 16 给出了模型 1 状态在 z=0 m 截面(上下的中间截面)处空间压力云图及等压线。探针长度 400 mm, 排管架影响区前传至探针中部, 对探针头部影响相对较小。

图 17~图 19 分别给出了模型 1、模型 2、模型 3 在 x=−0.412 m、y=0 m、z=0 m 截面处的动压云图。从图 17 等 x 截面图看, 由于排管架和导轨的阻挡作用, 探针头部的气流动压显著减小, 对整个截

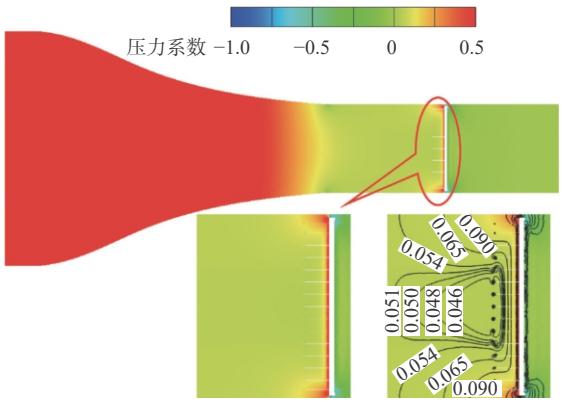


图 15 模型 1 状态 y=0 m 截面处的压力云图及等压线

Fig. 15 Pressure contours and isobaric line of model 1 at y = 0 m

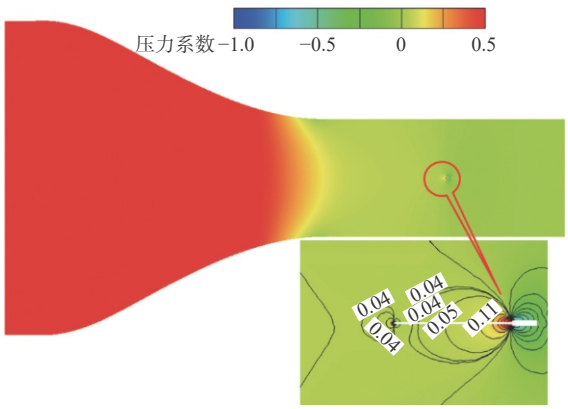


图 16 模型 1 状态 z=0 m 截面处的压力云图及等压线

Fig. 16 Pressure contours and isobaric line of model 1 at z = 0 m

面动压都有一定影响。从图 18 等 y 截面图看到，导轨横梁导致试验段顶部和底部的低速流动区域显著增厚，由于排管架的阻挡，相比模型 2 和模型 3，模型 1 在 y=0 m 截面处的高压区明显向收缩段前传。从图 19 等 z 截面图看到，排管架后部尾迹区明显，模型 3 在该截面处的动压明显高于模型 1 和模型 2，通过改进排管架矩形迎风面外形能够减小扰动。

为分析校核设备对风洞试验段动压的影响，分析了等 y 线及等 z 线上动压随流向变化曲线。等 y 线通过等 z 平面去截取空间流场获得等 z 截面，再针对等 z 截面，采用不同 y 值的线来截取二维面，

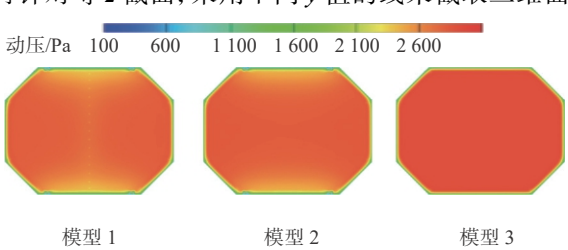


图 17 3 个模型在 x=−0.142 m 截面处的动压云图

Fig. 17 Dynamic pressure contours of three models at x = −0.142 m

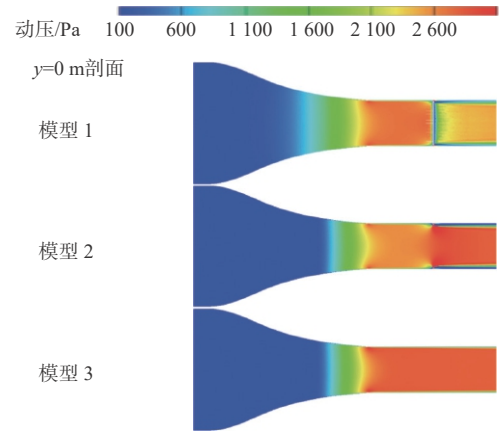


图 18 3 个模型在 y=0 m 截面处的动压云图

Fig. 18 Dynamic pressure contours of three models at y = 0 m

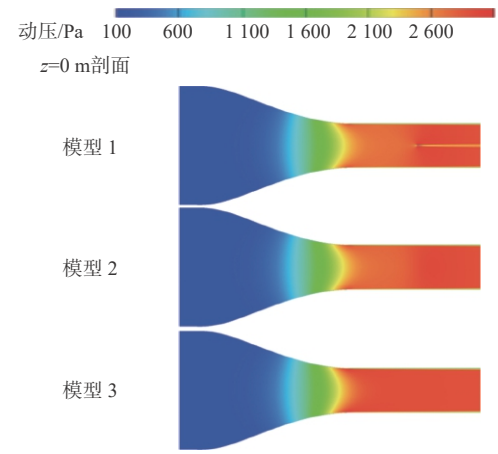


图 19 3 个模型在 z=0 m 截面处的动压云图

Fig. 19 Dynamic pressure contours of three models at z = 0 m

从而获得该截面上的等 y 线。图 20、图 21 分别给出了 3 个模型在 y=0 m、z=0 m 截面处沿流线的动

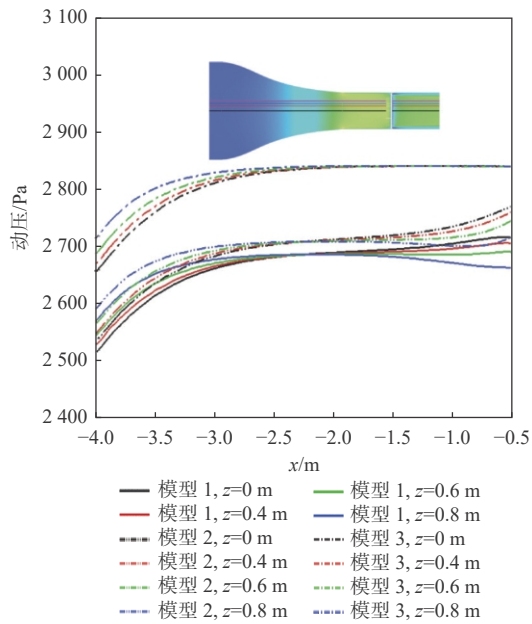


图 20 3 个模型在 y=0 m 截面处沿流线的动压曲线

Fig. 20 Dynamic pressure curves along streamline of three models at y= 0 m

压曲线。可以看出, 对于无任何设备的模型 3, 试验段不同截面的动压都维持在 2 841 Pa 左右。从图 20 看到, 模型 1 和模型 2 的 y=0 m 截面上不同等 z 线的动压略有差别, 等 z 线越靠近导轨横梁, 其所在位置动压越低。沿流线越接近排管架, 不同等 z 线上动压差别越大。由图 21 可知, 模型 1 和模型 2 的 z=0 m 截面上不同等 y 线上动压略有差别, 对于模型 1, 越远离排管架, 动压越大。模型 1 包含导轨/排管架/探针、模型 2 只包含导轨, 由图 20、图 21 的动压曲线看到, 模型 2 和模型 3 相同位置的动压差量明显高于模型 1 和模型 2 相同位置的动压差量, 说明导轨横梁对试验段动压影响更大。

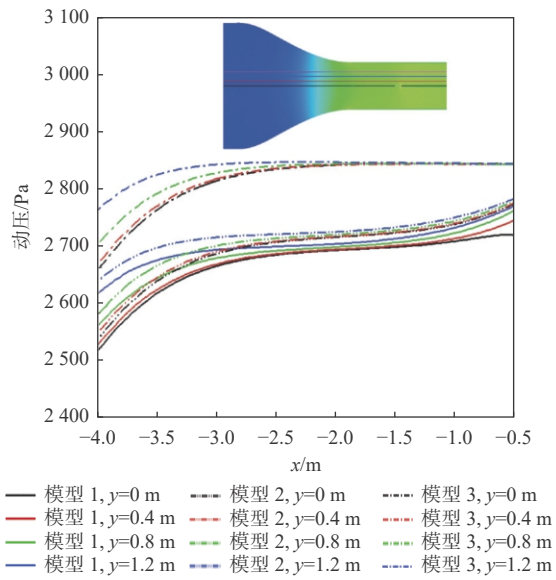


图 21　3 个模型在 z=0 m 截面处沿流线的动压曲线

Fig. 21　Dynamic pressure curves along streamline of three models at z= 0 m

由图 20、图 21 可知, 在 x=−4~−2 m 这段区域, 3 个模型都出现了越靠近壁面速度越大的现象。为分析现象原因, 图 22 给出了模型 3 在 y=0 m 截面处的速度云图。可以看到, 在收缩段和试验段的过渡区域的风洞壁面附近, 速度出现局部极值, 进一步放大该区域看到, 收缩段和试验段连接处, 模型并未完全光滑过渡, 在 x=−4 m 处存在急剧收缩, 导致 x=−4 m 处速度出现局部极值, 从而影响风洞壁面到风洞中心的速度分布。图 23 给出了模型 3 在 y=0 m 截面处不同等 x 线从风洞顶部到底部的速度分布。可以看到, 收缩段和试验段连接处, 不光滑过渡对速度带来一定影响, 导致风洞中心速度低于壁

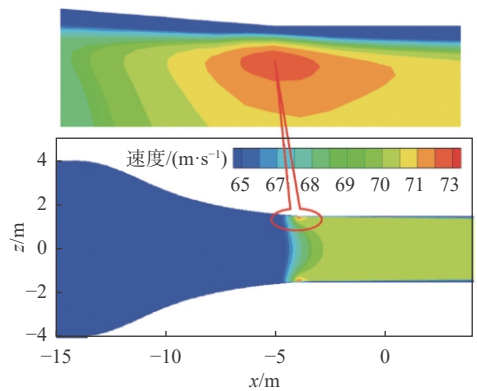


图 22　模型 3 在 y=0 m 截面处的速度云图及收缩段与试验段结合处局部放大图

Fig. 22　Velocity contours of model 3 at y=0 m and partial enlarged view of joint between contraction section and test section

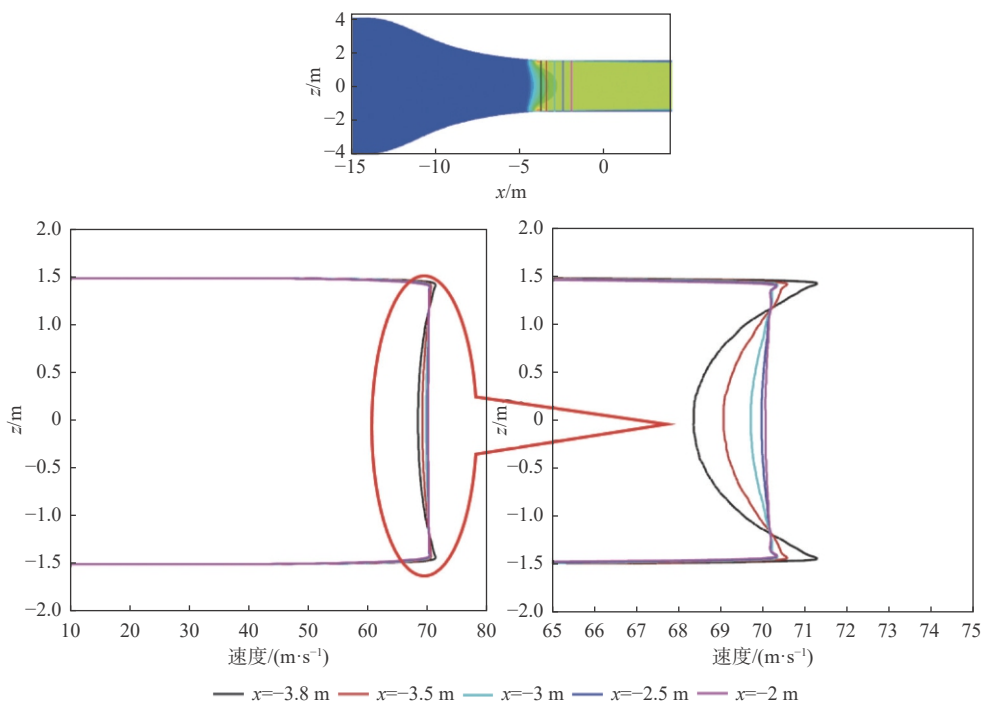


图 23　模型 3 在 y=0 m 截面处不同等 x 线的速度曲线

Fig. 23　Velocity curves of different constant x values of model 3 at y = 0 m

面附面层外的速度，随着等 x 线远离 x=−4 m 拐角处，影响逐步减弱，在 x=−2 m 处，该现象基本消失。

图 24 给出了 3 个模型探针处的动压，模型 3 由于无任何校核设备，其动压约为 2 841 Pa。相比模型 1，模型 2 没有排管架，可知，模型 1 和模型 2 的绝大多数探针处动压差别较小，说明排管架对动压场影响量相对较小，约为 11 Pa。这对试验动压场干扰修正极为重要。

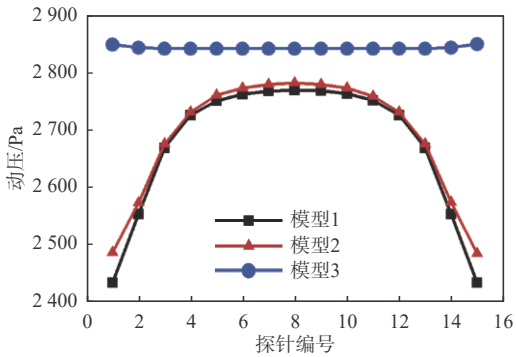


图 24 3 个模型探针处的动压

Fig. 24 Dynamic pressure of three models at probe

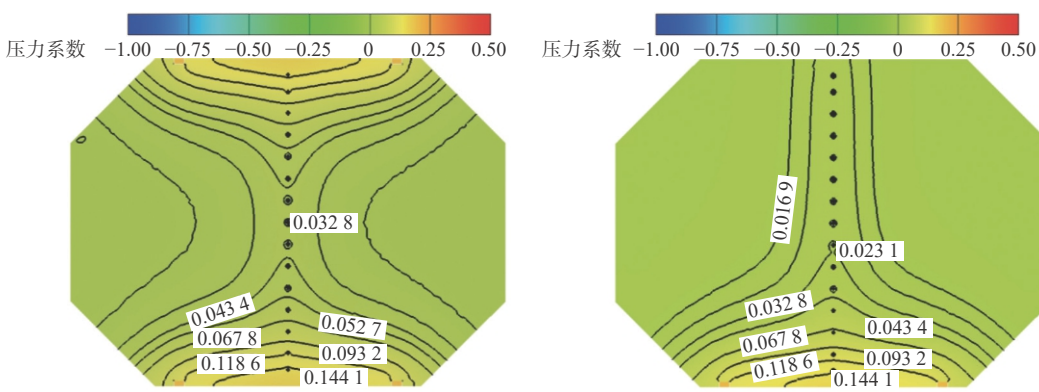
3.2 动压场干扰修正

风洞试验动压场干扰修正通过获取排管架导轨的影响量修正，分 2 步完成，为获得动压场，针对每个位置试验 2 种状态：①状态 1 为排管架上下端都有导轨；②状态 2 为只有排管架下端有导轨，上端直接固连于风洞洞壁。通过这两者探针处动压差量得到上导轨对探针的动压影响量 $\Delta p_i^{\mathrm{up}}(i=1,2,\cdots,14,15)$，下导轨对探针的动压影响量为 $\Delta p_i^{\mathrm{down}}=\Delta p_{16-i}^{\mathrm{up}}(i=1,2,\cdots,14,15)$，排管架对探针影响量为 $\Delta p_i^{\mathrm{const}}(i=1,2,\cdots,14,15)$。由 3.1 节可知，排管架对动压影响较小，因此，忽略排管架对探针处动压的影响量，则试验校核设备的动压场影响量为

$$\Delta p=\Delta p_i^{\mathrm{down}}+\Delta p_i^{\mathrm{up}}\qquad i=1,2,\cdots,14,15 \tag{3}$$

本节数值模拟了探针位于试验段 y=0，−0.4，−1.0 m 这 3 个位置处的流场，用于建立和验证试验动压场分析中的支架干扰影响扣除方法。每个位置同样模拟了状态 1 和状态 2。

图 25 给出了模型 1 排管架位于 y=0 m 位置时上部有无导轨状态的 x=−0.412 m 截面压力云图。可知，排管架导轨对截面动压影响较大，影响范围至相邻的 4 根探针。从图中等压线的数值可以看出，排管架上导轨对最下面的 4 根探针影响很小。



(a) 有导轨状态　　(b) 无导轨状态

图 25 模型 1 排管架位于 y=0 m 时上部有无导轨状态的 x=−0.412 m 截面压力云图

Fig. 25 Pressure contours of model 1 at x = −0.412 m with or without guide rail on upper part when pipe rack is located at y = 0 m

图 26 给出了模型 1 数值模拟和试验获得的探针处动压对比。相比试验测量结果，数值模拟结果对称性更好，不同测压点上数值模拟和试验测量的动压相差很小，进一步验证了软件的计算精度。

图 27 给出了模型 1 排管架上导轨对 3 个等 y 截面的探针位置处动压影响量。可以看出，上导轨对从上到下的 4 根探针影响较大，且影响量逐渐降低，对其余探针影响基本相同。在 y=0 m 截面位置，排管架上导轨对远离导轨的探针动压影响量最大约为 13 Pa。

图 28 给出了计算和试验的探针动压修正曲

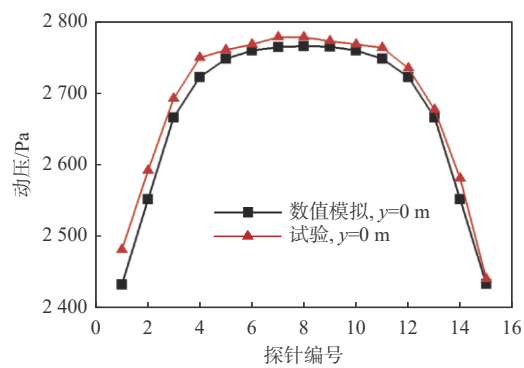


图 26 模型 1 数值模拟和试验的探针处动压对比

Fig. 26 Comparison of dynamic pressures at probe obtained by numerical simulation and test of model 1

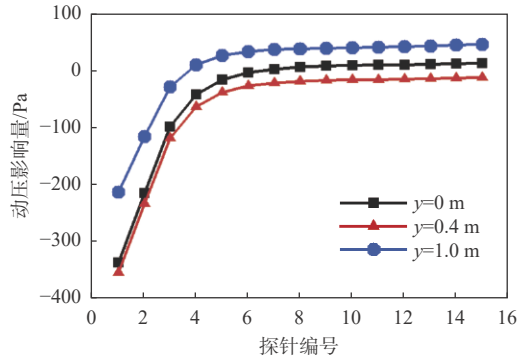


图 27　模型 1 排管架上导轨对 3 个等 y 截面的探针处动压影响量

Fig. 27　Degree of influence of upper guide rail of pipe rack of model 1 on dynamic pressure of probe at three constant y values

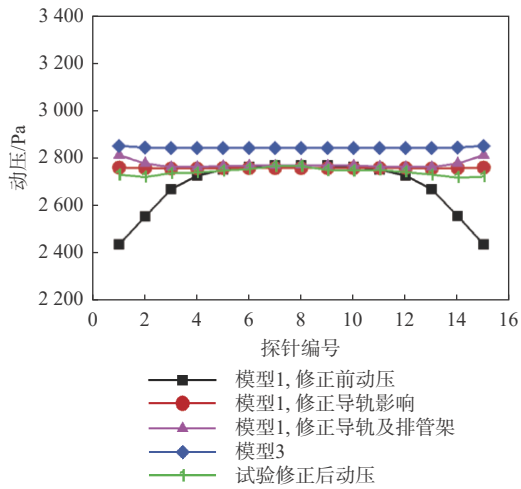


图 28　y=0 m 截面位置的排管架及导轨的动压修正曲线

Fig. 28　Dynamic pressure correction curves of pipe rack and guide rail at y = 0 m

线。通过 CFD 能够获取排管架对动压的影响量，因此，图中同时给出了模型 1 修正导轨和修正导轨及排管架 2 种修正后的动压曲线。可以看到，排管架对靠近洞壁的 1 号和 15 号探针动压场有一定影响，不考虑排管架影响时，计算的修正方法的结果与试验修正后的动压基本保持一致，试验与计算动压最大相差约 20 Pa，相比 2 840 Pa 量级当地动压，误差约 0.7%，从而验证了试验动压修正的合理性。如果需要更进一步精细修正，需要通过计算获得排管架的动压影响量，在该基础上进一步修正试验动压，另外，试验中可进一步优化排管架外形，降低其动压影响量。

4　结　论

本文在国家数值风洞工程通用 CFD 软件 NNW-FlowStar 上发展了一种修正的压力出口边界条件，采用低速平板湍流边界层验证了这类边界的计算收敛性和计算精度，并在此基础上基于非结构混合网格针对 FL-12 风洞流场品质动压场校核问题，开展了探针排管架和导轨横梁对试验段动压场影响分析，获得了探针排管架及导轨对试验段动压场的影响规律。通过数值模拟排管架上端有无导轨获得导轨对动压场影响量，在此基础上建立了动压场校核中的动压场修正方法，并对比了试验的动压修正曲线。通过数值模拟，可以得到以下结论：

1）在 NNW-FlowStar 软件上发展的修正的压力出口边界条件具有很好的残差收敛性及计算精度，适用于包含附面层的低速出口边界。

2）动压场校核中，排管架导轨对风洞试验段动压影响较大，安装探针的排管架对风洞试验段动压相对影响小。

3）通过数值模拟排管架上端有无导轨，能够分析得到排管架导轨对试验段动压场的影响量，基于该影响量获得的动压修正方法能够较为准确地修正试验中排管架导轨的影响量。

4）排管架对试验段动压场影响较小，只对靠近洞壁的探针动压场有一定影响，修正试验动压场时，不考虑排管架影响合理，可以通过优化排管架外形进一步降低其对动压场影响，提高校核试验的精度。

5）排管架导轨对动压场影响大，通过降低导轨高度、改变排管架导轨横梁与纵梁的搭接方式能够减小动压影响量，进一步优化校核结果。

参考文献（References）

[1] 王继明, 高云海, 焦仁山. 大型客机低速构型高雷诺数风洞腹撑支架干扰数值模拟[J]. 航空学报, 2020, 41(4): 123526.
WANG J M, GAO Y H, JIAO R S. Numerical simulation of ventral sting interference in high Reynolds number wind tunnel for civil aircraft low speed configuration[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(4): 123526(in Chinese).

[2] 代燚, 陈作钢, 马宁, 等. 低速风洞内部流场数值模拟[J]. 空气动力学学报, 2014, 32(2): 203-208.
DAI Y, CHEN Z G, MA N, et al. Numerical simulation of flow field inside the low-speed wind tunnel[J]. Acta Aerodynamica Sinica, 2014, 32(2): 203-208(in Chinese).

[3] 李红喆, 廖达雄, 丛成华. 连续式跨声速风洞大开角段整流装置设计数值模拟[J]. 空气动力学学报, 2015, 33(2): 198-203.
LI H Z, LIAO D X, CONG C H. Numerical simulation of flow conditioning device design in wide angle diffuser of continuous transonic wind tunnels[J]. Acta Aerodynamica Sinica, 2015, 33(2): 198-203(in Chinese).

[4] 丛成华, 刘琴, 张志峰, 等. 专用跨声速风洞开孔壁试验段设计数值模拟[J]. 航空学报, 2012, 33(6): 1014-1019.
CONG C H, LIU Q, ZHANG Z F, et al. Numerical simulation of design of transonic wind tunnel perforated test section[J]. Acta Aeronautica et Astronautica Sinica, 2012, 33(6): 1014-1019(in Chinese).

[5] 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).

[6] CHEN J T, ZHANG Y B, ZHOU N C, et al. Numerical investigations of the high-lift configuration with MFlow solver[J]. Journal of Aircraft, 2015, 52(4): 1051-1062.

[7] CHEN J T, ZHANG Y B, ZHAO H, et al. Numerical investigations of the NASA common research model with aeroelastic twist[J]. Journal of Aircraft, 2018, 55(4): 1469-1481.

[8] 张耀冰, 唐静, 陈江涛, 等. 基于非结构混合网格的 CHN-T1 标模气动特性预测[J]. 空气动力学学报, 2019, 37(2): 262-271.
ZHANG Y B, TANG J, CHEN J T, et al. Aerodynamic characteristics prediction of CHN-T1standard model with unstructured grid[J]. Acta Aerodynamica Sinica, 2019, 37(2): 262-271(in Chinese).

[9] 张耀冰, 邓有奇, 吴晓军, 等. DLR-F6 翼身组合体数值计算[J]. 空气动力学学报, 2011, 29(2): 163-169.
ZHANG Y B, DENG Y Q, WU X J, et al. Drag prediction of DLR-F6 using MFlow unstructured mesh solver[J]. Acta Aerodynamica Sinica, 2011, 29(2): 163-169(in Chinese).

[10] ROE P L. Approximate Riemann solvers, parameter vectors, and difference schemes[J]. Journal of Computational Physics, 1981, 43(2): 357-372.

[11] 张培红, 张耀冰, 周桂宇, 等. 面向混合网格高精度阻力预测的熵修正方法[J]. 航空学报, 2018, 39(9): 122030.
ZHANG P H, ZHANG Y B, ZHOU G Y, et al. Entropy correction method for high accuracy drag prediction with mixed grids[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(9): 122030(in Chinese).

[12] 张培红, 张耀冰, 周桂宇, 等. 面向非结构混合网格高精度阻力预测的梯度求解方法[J]. 航空学报, 2018, 39(1): 121415.
ZHANG P H, ZHANG Y B, ZHOU G Y, et al. Gradient calculation method of unstructured mixed grids for improving drag prediction accuracy[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(1): 121415(in Chinese).

[13] VENKATAKRISHNAN V. On the accuracy of limiters and convergence to steady state solutions[C]//Proceedings of the 31st Aerospace Sciences Meeting. Reston: AIAA, 1993: 880.

[14] SPALART P, ALLMARAS S. A one-equation turbulence model for aerodynamic flows[C]// Procadings of the 30th Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 1992: 439.

[15] JAMESON A, YOON S. Lower-upper implicit schemes with multiple grids for the Euler equations[J]. AIAA Journal, 1987, 25(7): 929-935.

[16] 唐静, 李彬, 周乃春, 等. 基于非结构网格流场超大规模并行计算[J]. 空气动力学学报, 2019, 37(1): 61-67.
TANG J, LI B, ZHOU N C, et al. Large scale parallel computing for fluid dynamics on unstructured grid[J]. Acta Aerodynamica Sinica, 2019, 37(1): 61-67(in Chinese).

[17] BLAZEK J. Computational fluid dynamics: principles and applications[M]. 3rd ed. Amsterdam: Elsevier, 2015: 262-264.

[18] NASA Langley Research Center Turbulence Modeling Resource. 2D zero pressure gradient flate plate verification case[EB/OL]. [2023-05-27]. http://turbmodels.larc.nasa.gov/flatplate_grids.html.

[19] NASA Langley Research Center Turbulence Modeling Resource. SA expected results-2D zero pressure gradient flat plate[EB/OL]. [2023-05-27]. http://turbmodels.larc.nasa.gov/flatplate_sa.html.

[20] 中国空气动力研究与发展中心. 4 米×3 米风洞(FL-12)[EB/OL]. [2023-05-27]. http://www.cardc.cn//Dev_Read.Asp?ChannelId=4&ClassId=18&Id=39.
CARDC. 4 m×3 m wind tunnel (FL-12) [EB/OL]. [2023-05-27]. http://www.cardc.cn//Dev_Read.Asp?ChannelId=4&ClassId=18&Id=39.

Equipment interference and correction method in dynamic pressure field verification of low-speed wind tunnel

LIU Jiangtao[1]，GONG Xiaoquan[2]，ZHOU Naichun[2,*]，LI Ming[2]，ZHANG Yaobing[2]

(1. Low Speed Aerodynamics Institute，China Aerodynamics Research and Development Center，Mianyang 621000，China；
2. Computational Aerodynamics Institute，China Aerodynamics Research and Development Center，Mianyang 621000，China)

Abstract: As one of the quality indexes of the wind tunnel flow field, the dynamic pressure field of the wind tunnel directly affects the accuracy of experiment results, and the dynamic pressure field needs to be checked regularly. Dynamic pressure field verification equipment includes a five-hole probe, pipe rack, and guide rail (longitudinal beam and cross beam). The influence of verification equipment needs to be deducted when the dynamic pressure field is analyzed. Based on the national general computational fluid dynamics (CFD) software NNW-FlowStar for numerical wind tunnel engineering, a class of low-speed pressure outlet boundary conditions for simulating boundary layer flow was established. The residual convergence and calculation accuracy of the boundary conditions were verified by the low-speed flat plate turbulent boundary layer. Based on unstructured hybrid mesh and FlowStar, the numerical simulation of the FL-12 low-speed wind tunnel test section without any equipment, the test section only with the guide rail of the pipe rack, and the test section with both the pipe rack and the guide rail was carried out. The degree of influence and affected areas of the equipment were qualitatively displayed through a spatial dynamic pressure cloud diagram. The degree of influence of the verification equipment on the dynamic pressure of the test sections was given quantitatively through the spatial dynamic pressure distribution curve. In addition, the state with and without the guide rail of the pipe rack at the top of the test sections was numerically simulated, and the degree of influence of the guide rail of the pipe rack on the dynamic pressure of the test sections was obtained. The guide rail of the pipe rack had a great influence on the adjacent four probes. By decomposing the degree of influence of the verification equipment on the dynamic pressure and comparing the dynamic pressure correction curves of numerical and experimental results, the results show that the degree of influence of the guide rail of the pipe rack on the dynamic pressure of the central probe is six times that of the pipe rack. This verifies the rationality of the experimental dynamic pressure correction scheme ignoring the degree of influence of the pipe rack, and optimization suggestions to improve the accuracy of dynamic pressure verification are proposed.

Keywords: wind tunnel; dynamic pressure field; degree of influence; outlet boundary; correction method; numerical simulation

Received：2023-06-02；**Accepted**：2023-09-08；**Published Online**：2023-11-10 18：21

URL：link.cnki.net/urlid/11.2625.V.20231109.0933.001

Foundation items：National Numerical Wind Tunnel; National Natural Science Foundation of China (11902343)

* **Corresponding author**. E-mail：znccxl@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0261

考虑材料优选的高速飞行器多层隔热结构优化

郝东，刘建霞，苏杰，贺元元*
(中国空气动力研究与发展中心 空天技术研究所，绵阳 621000)

摘 要：隔热结构是保护高速飞行器安全工作和完成各项任务的关键因素，多层隔热结构因其质量轻、防隔热效果好的特点而备受关注，其性能不仅与物理尺寸有关，还与材料优选、温度影响息息相关。考虑温变影响，建立了多层隔热结构瞬态热传导非线性模型，并提出了热传导方程的数值求解策略。考虑材料优选，引入材料热导率温变特性，以几何尺寸和多层隔热结构等效力学性能为优化约束，建立了多层隔热结构力-热一体化优化模型，并采用蒙特卡罗方法和粒子群优化方法进行优化求解。仿真结果表明：所提优化模型和方法可使结构质量减轻27.4%。温度影响分析表明：温变效应对优化结果有重要影响，若不考虑温度对材料性能的作用，得到的优化设计方案可能出现不满足温度设计约束的情况，导致结构失效。采用有限元模型验证了优化方案的可靠性和合理性。所提的多层隔热结构优化设计方法为高速飞行器的总体设计提供了技术支撑并奠定了优化设计基础。

关 键 词：高速飞行器；隔热；优化模型；蒙特卡罗；粒子群
中图分类号：V416
文献标志码：A **文章编号**：1001-5965(2025)05-1662-11

高速飞行器在飞行过程中将经历稠密大气层中上升、高超声速巡航或大气层外的再入返回，其承受着极端温差变化和严酷的气动力/热环境。极端的温度梯度和热流密度将对飞行器热安全构成直接威胁。为保证飞行器的飞行安全，必须对其采取热防护措施[1]。

隔热结构是保护高速飞行器安全工作和完成各项任务的关键因素，有效的隔热结构可以满足高速飞行器防热和隔热的基本需求，使其免于极端气动热环境的伤害。随着国家高速飞行器技术的发展和升级，传统的隔热结构难以满足低成本、轻量化、防/隔热一体化等方面的要求，因此，多层隔热结构的设计受到各国广泛关注[2]。多层隔热结构一般由耐高温层、隔热层和基层组成[3-4]，其隔热性能主要取决于隔热层材料参数和各层尺寸参数[5]。为获取最佳防隔热设计，国内外通过试验、理论分析和数值仿真等方法做了大量研究。文献[6-10]通过试验的方式获取了各类参数与隔热性能的关系，从而优化出最佳的隔热结构设计方案。文献[11-12]通过理论方法定量分析了多层隔热结构的热性能。Abramenko等建立了泡沫填充的铝合金多层隔热结构热导率理论计算模型[13]，Kostanovskii等考虑稳态传热过程提出了高温下的热导率确定方法[6]，Char等基于积分-差分方法建立了热导率逆确定方法[7]，Domínguez-Muñoz等研究了隔热材料的热导率不确定性问题[8]，吴大方等建立了非线性热环境下的多层隔热结构三维热响应预测模型[9]。Waszkielis等基于两点测量温度提出了一种工程上实用的热导率简化计算模型[10]。总的来说，通过试验手段获取的分析数据具有较高的可靠性，但需要花费大量的时间和金钱成本，且不能揭示传热特性和结构参数之间的关系物理机理。理论分析从物理机理出发，

收稿日期：2023-05-22；录用日期：2023-10-13；网络出版时间：2023-10-30 21:57
网络出版地址：link.cnki.net/urlid/11.2625.V.20231030.1035.002
*通信作者. E-mail：hyy63713@126.com

引用格式：郝东，刘建霞，苏杰，等. 考虑材料优选的高速飞行器多层隔热结构优化[J]. 北京航空航天大学学报，2025，51(5)：1662-1672.
HAO D，LIU J X，SU J，et al. Optimization of multilayer thermal insulation structure for high-speed aircraft considering material optimization[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1662-1672 (in Chinese).

所提出的模型可以较好地计算出材料的热导率和隔热性能，但大部分模型的建模或求解过程复杂，计算成本高，难以直接应用到结构优化中。

数值仿真方法具备一定的理论基础，经过长期发展，其分析数据也得到了广泛认可。在数值仿真方面，当前研究多针对确定多层隔热结构热性能进行预测计算，也有一些学者针对多层隔热结构的优化问题开展研究。马忠辉等对金属热防护系统的复合换热问题进行了研究，优化了隔热系统的设计参数，但在优化中没有考虑温度对材料性能的作用[14]。徐世南和吴催生在确定材料选择下进行了结构热防护优化研究，同样未考虑温度对材料的作用[15]。柴峻对耐高温复合材料作为结构层时的多层隔热结构进行了优化，虽然考虑了材料的热导率温变效应，但针对确定性材料，未考虑工程设计中材料选择的重要性[16]。在多层隔热结构的轻量化优化方面，大部分研究仅考虑确定性材料，且忽略了温度对材料性能的影响，轻量化优化方法距离工程实际应用仍具有一定差距。在工程应用中，热防护结构的隔热性能不仅与各层物理尺寸有关，还与所选择的材料体系紧密相关。材料的力学和热学参数随温度变化剧烈，如材料的热导率会随温度的升高而降低，直接影响隔热结构在高温环境的隔热性能降低；如果在进行隔热结构设计时不考虑温度对材料选择的影响，则无法得到较优的设计方案，导致高温下隔热结构无法满足设计性能指标要求，甚至存在整个结构隔热系统失效的风险。

针对该问题，本文在优化设计中考虑材料的温变效应，以材料优选和多层隔热结构几何厚度参数为设计变量，在满足隔热需要和力学性能的条件下，以质量与隔热率的比值最小为优化目标函数，建立优化数学模型，提出了一种考虑材料优选的高速飞行器多层隔热结构优化设计模型和优化方法。

1　多层隔热结构传热模型

高速飞行器多层隔热结构如图 1 所示。高速飞行器在飞行中所承受的极端高温热环境通过强制对流换热方式对多层隔热结构的耐高温层进行加热，同时耐高温层又向外部空间进行辐射散热，排走部分热量。余下热量中的主要热量被隔热层阻挡，小部分热量由耐高温层自身材料吸收存储和传递给基层，基层结构与内部空气进行自然对流换热。多层隔热结构隔热性能的优劣主要取决于隔热层材料参数和各层厚度尺寸，整体结构的等效力学性能主要由基层材料参数决定，因此，基层有时根据其功能又称为承载层。

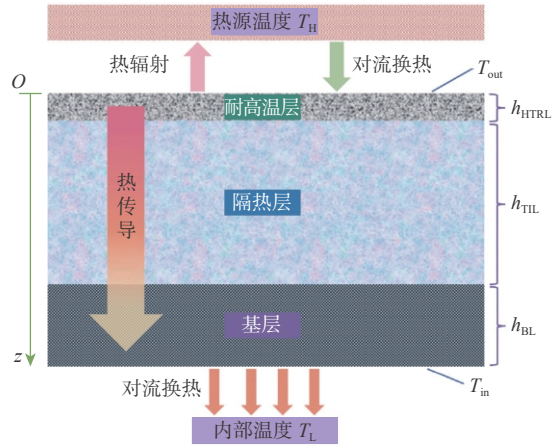


图 1　多层隔热结构示意图

Fig. 1　Multilayer thermal insulation structure

1.1　多层隔热结构瞬态热传导建模

高速飞行器泛指飞行速度超过马赫数 5，采用新型动力方案的空天飞行器，主要分为 3 类：滑翔飞行器、巡航飞行器和重复使用天地往返飞行器。高速飞行器具有较高的气动性能要求，外形型面复杂。为减轻结构质量，飞行器常采用“框+桁+防热承载壳体”的结构组成形式，其中，防热承载壳体为多层隔热结构，承担了飞行器在高速飞行中的防热和承载功能，是飞行器结构系统重要组件之一。

一般来说，高速飞行器壳体结构在轴向和周向的物理尺度远大于厚度方向的物理尺度，因此，高速飞行器局部三维曲面多层隔热结构可以近似为平面多层隔热结构进行研究分析，其厚度方向尺寸远远小于面内横向尺寸，即可将厚度方向的传热过程简化为一维物理模型。在建模中，定义坐标系统原点 O 位于耐高温层外壁面，坐标轴 Oz 从耐高温层指向基层（见图 1），高温热源温度为 T_H（常值），内部空气温度为 T_L（常值），耐高温层外壁面温度为 T_{out}，基层内壁面温度为 T_{in}，耐高温层厚度为 h_{HTRL}，隔热层厚度为 h_{TIL}，基层厚度为 h_{BL}。在进行多层隔热结构瞬态热传导建模之前，做出如下假设：

1）整个结构初始温度场均匀。

2）热量沿垂直于多层隔热结构的方向传递。

3）不同层结构之间紧密接触，热量完全传递。

对于一般的三维瞬态传热问题，传热偏微分方程为[17]

$$\rho c\frac{\partial T}{\partial t}=\frac{\partial}{\partial x}\left(k_x\frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y\frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z\frac{\partial T}{\partial z}\right)+q_v \tag{1}$$

式中：T 为温度变量；t 为时间变量；x、y 和 z 为空间变量；ρ 为材料密度；c 为材料比热容；k_x、k_y 和 k_z 分

别为 x、y 和 z 这 3 个方向的材料热导率；q_v 为内热源强度。

对于各向同性材料，有 $k_x=k_y=k_z=k(T)$。在没有内热源和考虑热导率的温变特性时，依据链式求导法则得到一维瞬态非线性传热控制方程[11]：

$$\rho c\frac{\partial T}{\partial t}=\frac{\mathrm{d}k(T)}{\mathrm{d}T}\left(\frac{\partial T}{\partial z}\right)^2+k(T)\frac{\partial T^2}{\partial z^2} \tag{2}$$

对于不同位置的层结构，式 (2) 中参数取值各不相同，等号右边第 1 项为热导率温变速率影响项，第 2 项为热导率温度非线性影响项。与不考虑温度作用的模型相比，本文传热方程多了等号右边第 1 项，并且等号右边的热导率及其对温度的导数均为温度的非线性函数。对于传热问题，其边界条件有 3 类：

1）第 1 类边界条件为 Dirichlet 条件，规定了结构边界上的温度值。

2）第 2 类边界条件为 Neumann 条件，规定了结构边界上的热流密度值。

3）第 3 类边界条件为 Robin 条件，规定了结构边界上结构与周围流体间的表面换热系数和流体温度。

对于图 1 所示的高速飞行器多层隔热结构物理模型，初始条件为

$$T(z,t)|_{t=0}=T(z,0)=20 \tag{3}$$

耐高温层外壁面受高温高速流体冲刷，存在强制对流换热和热辐射散热过程，其换热方程为[4]

$$-k_{\mathrm{HTRL}}(T)\left.\frac{\partial T}{\partial z}\right|_{z=0}=\psi_{\mathrm{HTRL\text{-}HTF}}(T_{\mathrm{H}}-T_{\mathrm{out}})-\varepsilon\sigma T_{\mathrm{out}}^4 \tag{4}$$

式中：$k_{\mathrm{HTRL}}(T)$ 为耐高温层材料的热导率参数；$\psi_{\mathrm{HTRL\text{-}HTF}}$ 为耐高温层外壁面与高温高速流体之间的强制对流换热系数；ε 为发射率（0～1 之间）；σ 为斯蒂芬-玻尔兹曼常数。

基层内壁面与内部常温流体进行自然对流换热，其换热方程为

$$-k_{\mathrm{BL}}(T)\left.\frac{\partial T}{\partial z}\right|_{z=h_{\mathrm{HTRL}}+h_{\mathrm{TIL}}+h_{\mathrm{BL}}}=\psi_{\mathrm{BL\text{-}NTF}}(T_{\mathrm{in}}-T_{\mathrm{L}}) \tag{5}$$

式中：$k_{\mathrm{BL}}(T)$ 为基层材料的热导率参数；$\psi_{\mathrm{BL\text{-}NTF}}$ 为基层内壁面与常温静态流体之间的自然对流换热系数，基层内壁面温度较低，忽略其热辐射的影响。

式 (2)～式 (5) 共同构成了高速飞行器多层隔热结构的热传导数学模型。

1.2 热传导方程数值求解

在实际应用中，想要求解得到一个具体物理问题的偏微分方程解析解是非常困难的，有时甚至是无法获得的，因此，在误差允许范围内，常常采用数值计算方法得到其近似离散解。本文采用有限差分隐式格式离散偏微分方程，进行迭代求解[12]。有限差分隐式迭代格式求解精度高，无条件稳定收敛，其将离散后的积分空间区域坐标序列用下标 i 表示，离散后的时间区间序列用下标 j 表示，则 $T_{i,j}$ 表示位于空间位置 z_i 且在时刻 t_j 时的温度值。利用有限差分格式代替偏微分求导，将偏微分求导转化为差分计算：

$$\left(\frac{\partial T}{\partial z}\right)_{i,j}\cong\frac{T_{i+1,j}-T_{i,j}}{\Delta z} \tag{6}$$

对 $T_{i+1,j}$ 和 $T_{i-1,j}$ 分别在 $T_{i,j}$ 处进行泰勒展开，则有

$$T_{i+1,j}=T_{i,j}+\left(\frac{\partial T}{\partial z}\right)_{i,j}\Delta z+\left(\frac{\partial^2 T}{\partial z^2}\right)_{i,j}\frac{\Delta z^2}{2!}+\left(\frac{\partial^3 T}{\partial z^3}\right)_{i,j}\frac{\Delta z^3}{3!}+O\left(\Delta z^3\right) \tag{7}$$

$$T_{i-1,j}=T_{i,j}-\left(\frac{\partial T}{\partial z}\right)_{i,j}\Delta z+\left(\frac{\partial^2 T}{\partial z^2}\right)_{i,j}\frac{\Delta z^2}{2!}-\left(\frac{\partial^3 T}{\partial z^3}\right)_{i,j}\frac{\Delta z^3}{3!}+O\left(\Delta z^3\right) \tag{8}$$

由式 (7) 和式 (8) 可得

$$\begin{cases}\left(\dfrac{\partial T}{\partial z}\right)_{i,j}=\dfrac{T_{i+1,j}-T_{i,j}}{\Delta z}+O(\Delta z)\\ \left(\dfrac{\partial T}{\partial z}\right)_{i,j}=\dfrac{T_{i,j}-T_{i-1,j}}{\Delta z}+O(\Delta z)\\ \left(\dfrac{\partial^2 T}{\partial z^2}\right)_{i,j}=\dfrac{T_{i+1,j}-2T_{i,j}+T_{i-1,j}}{\Delta z^2}+O\left(\Delta z^2\right)\end{cases} \tag{9}$$

在得到温度对于空间坐标的离散格式之后，其对于时间的偏微分差分格式可以表达为

$$\left(\frac{\partial T}{\partial t}\right)_{i,j}\cong\frac{T_{i,j}-T_{i,j-1}}{\Delta t} \tag{10}$$

将式 (9) 和式 (10) 代入式 (2)，得到传热方程的数值递推迭代格式为

$$\frac{T_{i,j}-T_{i,j-1}}{\Delta t}=\frac{k'\left(T_{i+1,j}-T_{i,j}\right)\left(T_{i,j}-T_{i-1,j}\right)}{\rho c\Delta z^2}+\frac{k\left(T_{i+1,j}-2T_{i,j}+T_{i-1,j}\right)}{\rho c\Delta z^2} \tag{11}$$

式中：$k'=\mathrm{d}k(T)/\mathrm{d}T$，对于大部分材料 $k'\geqslant 0$。

时间步递推 1 步，式 (11) 中 j 需变为 j+1，有

$$\frac{T_{i,j+1}-T_{i,j}}{\Delta t}=\frac{k'\left(T_{i+1,j+1}-T_{i,j+1}\right)\left(T_{i,j+1}-T_{i-1,j+1}\right)}{\rho c\Delta z^2}+\frac{k\left(T_{i+1,j+1}-2T_{i,j+1}+T_{i-1,j+1}\right)}{\rho c\Delta z^2} \tag{12}$$

联立式 (11) 和式 (12)，得到空间位置 z_i 处的温度在时间尺度上的递推格式：

$$\begin{cases} T_{i,j+1}^2+\theta T_{i,j+1}+\lambda=0 \\ \theta=\dfrac{\rho c\Delta z^2}{\Delta t}+2k+k'\left(T_{i-1,j+1}-T_{i+1,j+1}\right) \\ \lambda=-kT_{i+1,j+1}-(k-k')T_{i-1,j+1}-\dfrac{\rho c\Delta z^2}{\Delta t}T_{i,j} \\ T_{i+1,j+1}\leqslant T_{i-1,j+1} \\ \theta>0 \\ \lambda<0 \end{cases} \tag{13}$$

由式 (13) 可知，当考虑温度影响时，时间尺度递推需要求解二次方程模型。并且由于 θ>0、$T_{i,j+1}$>0，方程有唯一解，得

$$T_{i,j+1}=\frac{\sqrt{\theta^2-4\lambda}-\theta}{2} \tag{14}$$

材料热导率温变特性对温度响应的影响体现在参数 θ 和 λ 中。当不考虑温度的影响时，参数 θ 和 λ 退化为

$$\begin{cases} \theta=\dfrac{\rho c\Delta z^2}{\Delta t}+2k \\ \lambda=-k\left(T_{i+1,j+1}+T_{i-1,j+1}\right)-\dfrac{\rho c\Delta z^2}{\Delta t}T_{i,j} \\ \theta>0 \\ \lambda<0 \end{cases} \tag{15}$$

综上所述，本文求解传热方程的数值计算策略如图 2 所示。在数值求解过程中，材料参数需要根据多层隔热结构不同空间位置处的具体情况选取，不同位置材料参数可能不同，需要实时判定确定。

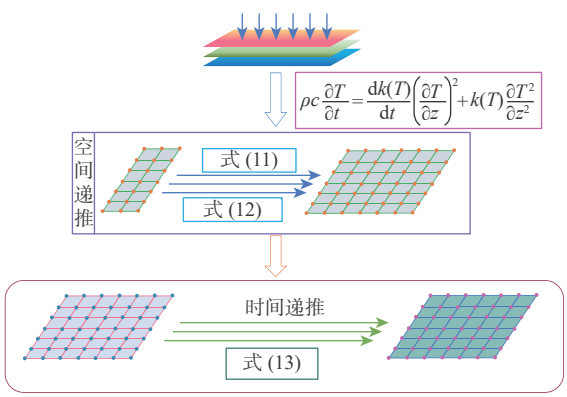


图 2　传热方程求解策略

Fig. 2　Solution strategy of heat transfer equation

2　优化模型建模与求解策略

本文将各层材料优选变量和各层几何尺寸同时作为优化设计变量，并且考虑材料热导率温变特性，建立了多层隔热结构的轻量化优化模型。考虑到设计变量的非连续性，采用智能优化方法进行优化求解。

2.1　设计变量与目标函数

令耐高温层、隔热层和基层的材料选择集变量分别为

$$M_{\mathrm{HTRL}}=\left\{M_{\mathrm{HTRL}}^1,M_{\mathrm{HTRL}}^2,\cdots,M_{\mathrm{HTRL}}^m\right\} \tag{16}$$

$$M_{\mathrm{TIL}}=\left\{M_{\mathrm{TIL}}^1,M_{\mathrm{TIL}}^2,\cdots,M_{\mathrm{TIL}}^n\right\} \tag{17}$$

$$M_{\mathrm{BL}}=\left\{M_{\mathrm{BL}}^1,M_{\mathrm{BL}}^2,\cdots,M_{\mathrm{BL}}^p\right\} \tag{18}$$

式中：M_{HTRL}、M_{TIL} 和 M_{BL} 为材料选择集；m、n 和 p 分别为耐高温层、隔热层和基层候选材料的种类数目。

定义材料优选变量 $\boldsymbol{M}$ 为

$$\boldsymbol{M}=[M_{\mathrm{HTRL}}\quad M_{\mathrm{TIL}}\quad M_{\mathrm{BL}}] \tag{19}$$

将材料优选变量和各层结构物理尺寸统一起来，提出一种混合设计变量式的优化变量 $\boldsymbol{x}$：

$$\boldsymbol{x}=[x_1,x_2,x_3,x_4,x_5,x_6]=[\boldsymbol{M},h_{\mathrm{HTRL}},h_{\mathrm{TIL}},h_{\mathrm{BL}}] \tag{20}$$

由式 (19) 和式 (20) 可知，新的优化设计变量 $\boldsymbol{x}$ 既不是连续性变量，也不是离散型变量，而是一种离散-连续相结合的混合式设计变量。

隔热层材料密度一般较小，是多层隔热结构隔热性能的主要决定因素。因此，在结构总厚度不变的情况下，隔热层厚度 h_{TIL} 越大，隔热能力一般也越好，同时质量也越轻。但是，由于实际情况中结构总厚度的限制，中间隔热层厚度不可能无限大，需要在隔热性能和结构尺寸约束之间均衡考虑。之前研究大多直接将隔热结构质量最小作为结构轻量化优化时的目标函数，这种方式并不能直接体现结构质量与隔热能力之间的关系。因此，本文将结构质量与结构隔热率之间的比值作为优化目标函数，如下：

$$F=\frac{m}{\xi}=\frac{\rho_{\mathrm{HTRL}}h_{\mathrm{HTRL}}+\rho_{\mathrm{TIL}}h_{\mathrm{TIL}}+\rho_{\mathrm{BL}}h_{\mathrm{BL}}}{\xi} \tag{21}$$

式中：m 为面质量密度；ρ_{HTRL}、ρ_{TIL} 和 ρ_{BL} 分别为耐高温层、隔热层和基层材料的质量密度；ξ 为多层隔热结构隔热率，定义为

$$\xi=\frac{T_{\mathrm{H}}-T_{\mathrm{in}}(t_{\mathrm{c}})}{T_{\mathrm{H}}} \tag{22}$$

式中：t_{c} 为高速飞行器工作时间长度。

由式 (21) 可知，F 越小，表示达到相同隔热效率所需要的结构质量越小。

2.2　力-热一体化优化模型

在进行多层隔热结构优化时，不仅需要满足各层厚度的几何尺寸约束，还需要满足结构隔热性能和力学性能要求。由于不同材料力学性能和热性能差距较大，各层材料优选时只能从各层自身材料备选集中选择。在本文中，多层隔热结构的等效力学性能采用等效模量 E_{equ} 进行表征，采用体积加权法进行等效建模：

$$E_{\mathrm{equ}}=\frac{E_{\mathrm{HTRL}}h_{\mathrm{HTRL}}+E_{\mathrm{TIL}}h_{\mathrm{TIL}}+E_{\mathrm{BL}}h_{\mathrm{BL}}}{h_{\mathrm{HTRL}}+h_{\mathrm{TIL}}+h_{\mathrm{BL}}} \tag{23}$$

式中：E_{HTRL} 为耐高温层材料模量；E_{TIL} 为隔热层材料模量；E_{BL} 为基层材料模量。

联立式 (20)～式 (23)，考虑材料选择约束、几何尺寸约束和力学性能约束，得力-热一体化优化模型为

$$\min\ F=F(\boldsymbol{x})=\frac{\rho_{\mathrm{HTRL}}(x_1)x_4+\rho_{\mathrm{TIL}}(x_2)x_5+\rho_{\mathrm{BL}}(x_3)x_6}{\xi(\boldsymbol{x})}$$

$$\text{s.t.}\begin{cases}T_{\mathrm{in}}(t=t_{\mathrm{c}})-T_{\mathrm{in_max}}\leqslant 0\\ E_{\min}-E_{\mathrm{equ}}(\boldsymbol{x})\leqslant 0\\ x_1\in M_{\mathrm{HTRL}}\\ x_2\in M_{\mathrm{TIL}}\\ x_3\in M_{\mathrm{BL}}\\ h_{\mathrm{HTRL_min}}\leqslant x_4\leqslant h_{\mathrm{HTRL_max}}\\ h_{\mathrm{TIL_min}}\leqslant x_5\leqslant h_{\mathrm{TIL_max}}\\ h_{\mathrm{BL_min}}\leqslant x_6\leqslant h_{\mathrm{BL_max}}\end{cases}\tag{24}$$

式中：$T_{\mathrm{in_max}}$ 为基层内壁面最大允许温度；$E_{\min}$ 为等效模量允许最小值；$h_{\mathrm{HTRL_min}}$、$h_{\mathrm{TIL_min}}$ 和 $h_{\mathrm{BL_min}}$ 分别为耐高温层、隔热层和基层厚度最小值：$h_{\mathrm{HTRL_max}}$、$h_{\mathrm{TIL_max}}$ 和 $h_{\mathrm{BL_max}}$ 分别为耐高温层、隔热层和基层厚度最大值。

2.3 优化模型求解策略

对于式 (24) 所示的优化模型，设计变量既有离散部分又有连续部分，无法给出准确梯度信息，因此，不能采用梯度优化方法。对于这类混合型设计变量的优化问题，本文分别采用蒙特卡罗方法和粒子群优化方法进行优化求解，粒子群优化方法优化代码框架参见文献 [18]。式 (24) 的约束条件为隐式类型，需要求解传热方程和计算等效模量才能得到满足约束条件的可行设计变量。蒙特卡罗方法和粒子群优化方法优化流程分别如图 3 和图 4 所示，$T_{\mathrm{in}}(t=t_{\mathrm{c}})$ 通过求解式 (2)～式 (5) 得到。

图 4 中，假设一个由 N 个粒子组成的群体在 D 维空间中搜索，第 i 个粒子的位置表示为：$\boldsymbol{X}_i=(X_{i1},X_{i2},\cdots,X_{iD})$，第 i 个粒子的速度表示为：$\boldsymbol{V}_i=(V_{i1},V_{i2},\cdots,V_{iD})$，第 i 个粒子经过的最优位置表示为：$\boldsymbol{P}_i=(P_{i1},P_{i2},\cdots,P_{iD})$，群体经过的最优位置为：$\boldsymbol{P}_G=(P_{G1},P_{G2},\cdots,P_{GD})$。粒子的位置和速度的更新策略如下：

$$\boldsymbol{V}_i^{k+1}=\boldsymbol{V}_i^k+c_1\xi\left(\boldsymbol{P}_i^k-\boldsymbol{X}_i^k\right)+c_2\eta\left(\boldsymbol{P}_G^k-\boldsymbol{X}_i^k\right)\tag{25}$$

$$\boldsymbol{X}_i^{k+1}=\boldsymbol{X}_i^k+\boldsymbol{V}_i^{k+1}\tag{26}$$

式中：c_1、c_2 为学习因子，通常取值为 2，学习因子使粒子具有自我总结和向群体中优秀个体学习的能力，从而向群体内最优点靠近；$\xi,\eta\in(0,1)$ 为在 (0,1) 区间均匀分布的伪随机数。

粒子的适应值通过式 (24) 中的目标函数进行计算。为进一步验证粒子群优化方法在解决多变量寻优问题时的收敛性和准确性，采用 Ackley 函数、Rosenbrock 函数和 Sphere 函数进行测试，寻优测试结果如表 1 所示。

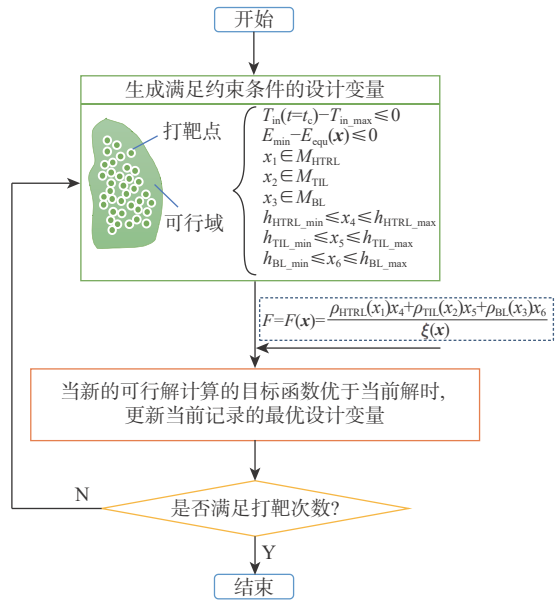


图 3 蒙特卡罗方法优化流程

Fig. 3 Optimization process of Monte-Carlo method

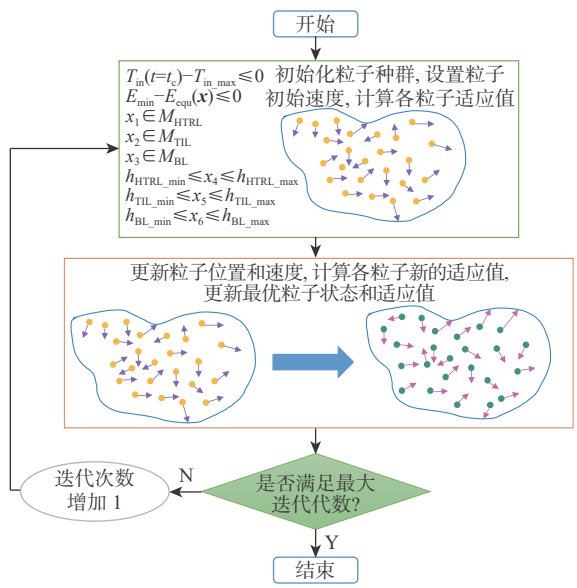


图 4 粒子群优化方法优化流程

Fig. 4 Optimization process of particle swarm method

表 1 寻优测试结果

Table 1 Test results of optimization

函数名称	变量维度	理论最优解	寻优测试解
Ackley	2	(0,0)	$(-2.74\times10^{-17},-2.33\times10^{-16})$
	6	(0,0,0,0,0,0)	$(-6.62\times10^{-16},2.72\times10^{-15},-1.61\times10^{-15},-2.74\times10^{-16},-8.12\times10^{-16},1.75\times10^{-16})$
Rosenbrock	2	(1,1)	(1.00,1.00)
	6	(1,1,1,1,1,1)	(1.00,1.00,1.00,1.00,1.00,1.00)
Sphere	2	(0,0)	$(-1.65\times10^{-25},7.10\times10^{-25})$
	6	(0,0,0,0,0,0)	$(-1.98\times10^{-24},-1.41\times10^{-24},1.06\times10^{-24},-7.47\times10^{-26},2.91\times10^{-24},5.51\times10^{-25})$

3　优化结果分析

以一个典型多层隔热结构方案为算例对象，分别采用蒙特卡罗方法和粒子群优化方法对其进行结构设计优化，得到优化前后的结构隔热效果和减重效果，分析结构优化前后的温度时域响应和温度空间分布，并采用有限元方法对优化结果进行验证。耐高温层、隔热层和基层均选择 3 种材料备选。

3.1　仿真条件与材料参数

引入材料热导率温变特性后，本文所建立的考虑材料优先的高速飞行器多层隔热结构的优化模型如式 (24) 所示，仿真验证参数设置如表 2 所示。耐高温层备选材料为 C/C 复合材料、C/SiC 复合材料和 ZrB_2 复合材料，隔热层备选材料为碳气凝胶、SiO_2 气凝胶和 SiC 气凝胶，基层备选材料为 7075 铝合金、GH4099 高温合金和TC4 钛合金。C/C 复合材料[19-20]、C/SiC 复合材料[21-22] 和 ZrB_2 复合材料[23-24] 的材料参数如表 3 所示，碳气凝胶[25-26]、SiO_2 气凝胶[27] 和 SiC 气凝胶[28-29] 材料参数如表 4 所示，7075 铝合金[30]、GH4099 高温合金[31] 和 TC4 钛合金[32-33] 材料参数如表 5 所示。

表 2　仿真参数

Table 2　Simulation parameters

参数	数值
h_{HTRL_min}/mm	0.5
h_{HTRL_max}/mm	3
h_{TIL_min}/mm	5
h_{TIL_max}/mm	50
h_{BL_min}/mm	2
h_{BL_max}/mm	5
ε	0.8
t_c/s	300
T_H/℃	1 200
T_L/℃	20
T_{in_max}/℃	100
E_{min}/GPa	10
$\psi_{HTRL\text{-}HTF}$/(W·m^{-2}·℃$^{-1}$)	230
$\psi_{BL\text{-}NTF}$/(W·m^{-2}·℃$^{-1}$)	10
M_{HTRL}	C/C复合材料、C/SiC复合材料、ZrB_2复合材料
M_{TIL}	碳气凝胶、SiO_2气凝胶、SiC气凝胶
M_{BL}	7075铝合金、GH4099高温合金、TC4钛合金

表 4 中的 $K_1(T)$、$K_2(T)$ 和 $K_3(T)$ 如式 (27)、式 (28) 和式 (29) 所示：

$$K_1(T)=3.521\,505\times10^{-7}T^2+7.847\,416\times10^{-4}T+1.104\,313 \quad (27)$$

$$K_2(T)=1.135\,885\times10^{-10}T^3-2.256\,084\times10^{-8}T^2+1.640\,566\times10^{-5}T+1.481\,429\times10^{-2} \quad (28)$$

表 3　耐高温层材料参数

Table 3　Material parameters of high temperature-resistant layer

材料名称	密度/(kg·m^{-3})	比热容/(J·kg^{-1}·K^{-1})	热导率/(W·m^{-1}·K^{-1})	弹性模量/GPa
C/C 复合材料	1 600	713	1.6	22.32
C/SiC 复合材料	2 100	1 420	5	49.8
ZrB_2基复合材料	6 300	430	53	490

表 4　隔热层材料参数

Table 4　Material parameters of thermal insulation layer

材料名称	密度/(kg·m^{-3})	比热容/(J·kg^{-1}·K^{-1})	热导率/(W·m^{-1}·K^{-1})	弹性模量/GPa
碳气凝胶	811	870	$K_1(T)$	2.84
SiO_2气凝胶	140	1 050	$K_2(T)$	0.01
SiC气凝胶	9.7	579	$K_3(T)$	0.013

表 5　基层材料参数

Table 5　Material parameters of base layer

材料名称	密度/(kg·m^{-3})	比热容/(J·kg^{-1}·K^{-1})	热导率/(W·m^{-1}·K^{-1})	弹性模量/GPa
7075 铝合金	2 800	962	129.4	71
GH4099 高温合金	8 470	624	18.9	175
TC4 钛合金	4 440	659	9.13	115.85

$$K_3(T)=5.269\,242\times10^{-11}T^3-2.221\,319\times10^{-8}T^2+1.622\,211\times10^{-5}T+2.377\,993\times10^{-2} \quad (29)$$

多层隔热结构的隔热性能主要取决于隔热层材料的热导率，热导率是材料热阻能力的直接表征和体现，表 4 所示的隔热层材料热导率随温度变化，具有明显的温变非线性特征，碳气凝胶、SiO_2 气凝胶和 SiC 气凝胶 3 种材料的热导率与温度的关系曲线如图 5 所示，3 种气凝胶材料的热导率均随温度的升高而增大，即隔热性能降低。从图 5 可知，表 4 所示的材料热导率多项式模型能够表征材料的真实属性，多项式模型拟合的相关系数均大于 0.98。

3.2　不同优化方法对比分析

由于蒙特卡罗方法和粒子群优化方法在产生可行解时具有随机性，2 种优化方法的初始方案并不相同，因此，采用轻量化优化的减重百分比作为比较变量值。在采用蒙特卡罗方法优化时，有效打靶次数设置为 200，优化结果如图 6 所示；在采用粒子群优化方法优化时，种群数量设置为 50，迭代次数设置为 200，优化结果如图 7 所示。由图 6 和图 7 可知，当采用蒙特卡罗方法时，目标函数 F 从 23.286 3 kg/m^2 优化为 13.551 8 kg/m^2，减重 41.8%；当采用粒子群优化方法时，目标函数 F 从 12.741 1 kg/m^2 优化为 9.252 2 kg/m^2，减重 27.4%。

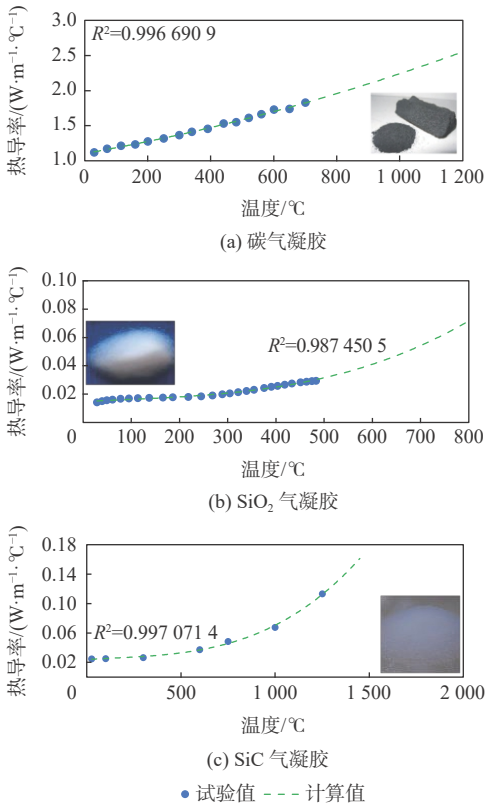


图5 隔热层材料热导率温变特性

Fig. 5 Temperature variation characteristics of thermal conductivity for thermal insulation layer material

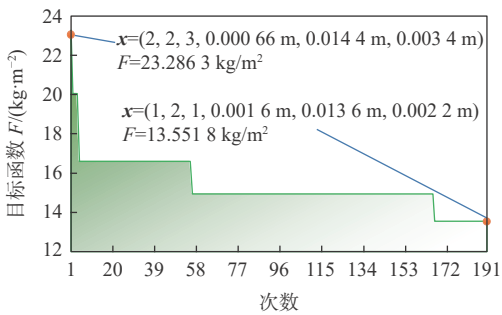


图6 蒙特卡罗方法优化结果(有温度影响)

Fig. 6 Optimization results of Monte-Carlo method (with temperature effect)

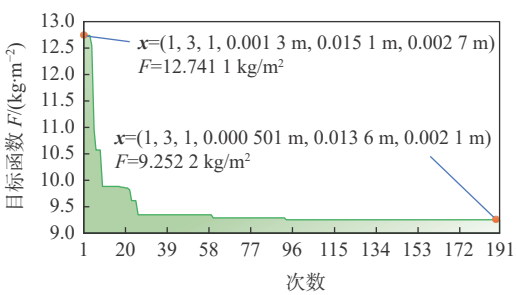


图7 粒子群优化方法优化结果(有温度影响)

Fig. 7 Optimization results of particle swarm method (with temperature effect)

虽然粒子群优化方法优化的轻量化减重百分比小于蒙特卡罗方法,但其优化得到的目标函数 F 绝对值远远优于蒙特卡罗方法得到的目标函数 F 绝对值。

蒙特卡罗方法优化时,优化前各层材料和厚度为:C/SiC 复合材料 0.000 66 m、SiO_2 气凝胶 0.014 4 m、TC4 钛合金 0.003 4 m,优化后的各层材料和厚度为:C/C 复合材料 0.001 6 m、SiO_2 气凝胶 0.013 6 m、7075 铝合金 0.002 2 m。粒子群优化算法优化时,优化前各层材料和厚度为:C/C 复合材料 0.001 3 m、SiC 气凝胶 0.015 1 m、7075 铝合金 0.002 7 m,优化后的各层材料和厚度为:C/C 复合材料 0.000 501 m、SiC 气凝胶 0.013 6 m、7075 铝合金 0.002 1 m。此外,从图 6 和图 7 对比可知,粒子群优化方法的优化收敛速度远优于蒙特卡罗方法,在第 100 次迭代时基本达到收敛。

3.3 温度影响分析

为进一步说明在进行多层隔热结构设计时考虑材料温变影响的必要性,需要将不考虑温度影响的结构优化方案与考虑温度影响的结构优化方案进行比较。此外,由于材料参数的温度特性是其本身固有属性,且在实际应用中也一直存在,需要对不考虑温度影响时得到的优化设计方案引入温变特性进行约束满足性验证。不考虑温度影响时,材料参数为常数,不随温度变化而变化,本文以常用的 25℃ 时的热导率数据进行参数设置,其余参数保持不变。不考虑温度影响时,蒙特卡罗方法和粒子群优化方法的优化结果分别如图 8 和图 9 所示。

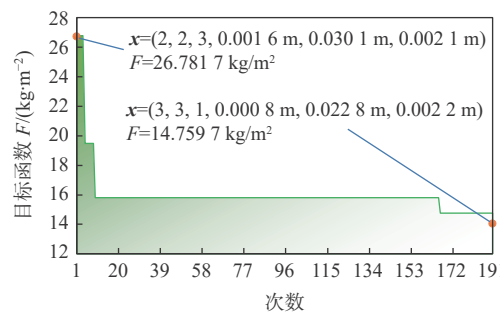


图8 蒙特卡罗方法优化结果(无温度影响)

Fig. 8 Optimization results of Monte-Carlo method (without temperature effect)

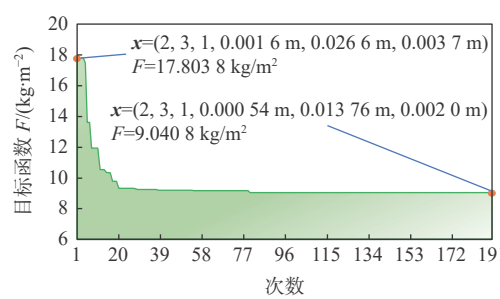


图9 粒子群优化方法优化结果(无温度影响)

Fig. 9 Optimization results of particle swarm method (without temperature effect)

从图 8 和图 9 可知，在不考虑温度影响的条件下，采用蒙特卡罗方法时，目标函数从 26.781 7 kg/m^2 减小至 14.759 7 kg/m^2，减重约 44.9%；采用粒子群优化方法时，目标函数从 17.803 8 kg/m^2 减小至 9.040 8 kg/m^2，减重约 49%。虽然粒子群优化方法的减重百分比低于蒙特卡罗方法，这是由于其初始条件不同而决定；但粒子群优化方法优化后的目标函数绝对值远低于蒙特卡罗方法，即优化效果优于蒙特卡罗方法，因此，后文无特殊说明，均采用粒子群优化方法的结果进行分析。

在得到考虑温度影响和不考虑温度影响 2 种情况的隔热结构优化方案后，采用较优的粒子群优化方法得到的优化设计方案进行对比。因为材料的热导率参数是随温度变化的，为了更符合隔热结构的实际工作环境，应考虑温度对材料参数的影响来分析 2 种优化设计方案的温度响应。有温度影响和无温度影响时，粒子群优化方法得到的多层隔热结构优化方案分别为：(1,3,1,0.000 501 m,0.013 6 m,0.002 1 m) 和 (2,3,1,0.000 54 m,0.013 76 m,0.002 0 m)，对应的耐高温层外壁面温度响应和基层内壁面温度响应分别如图 10 和图 11 所示。从图 10 和图 11 可知，考虑温度影响的隔热结构方案，其耐高温层外壁面温度上升速度快于不考虑温度影响的情况，材料热导率随着温度的升高而增大，导热越快，温升越快，最终趋于稳定，2 种情况的平衡温度相同。不考虑温度影响时的隔热结构方案，其基层内壁面温度在 300 s 时为 137.8℃，而基层内壁面温度设计要求不得高于 100℃，不满足设计约束。这是由于温度升高，材料热导率增大，大于设计时选用的参考值，材料隔热性能降低，导致实际使用时的内壁面温度高于设计值。

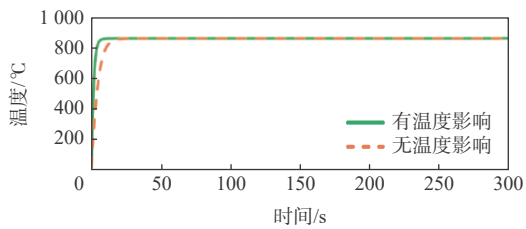


图 10　有无温度影响优化方案的耐高温层外壁面温度响应对比

Fig. 10　Comparison of external wall temperature responses for high temperature-resistant layer with and without temperature effect for the optimized scheme

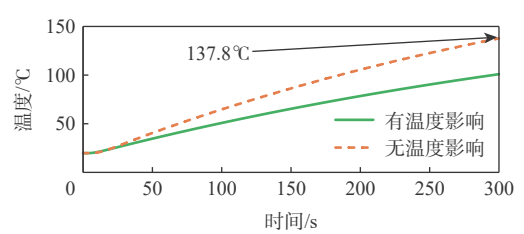


图 11　有无温度影响优化方案的基层内壁面温度响应对比

Fig. 11　Comparison of inner wall temperature responses for base layer with and without temperature effect

综上所述，若在进行高速飞行器多层隔热结构优化设计时不考虑材料参数的温度影响，得到的优化设计方案在实际使用中可能无法满足最初的设计约束条件，导致飞行器内部温度过高，甚至飞行任务失败。因此，在多层隔热结构设计时引入温度的影响，优化材料选择是十分必要的。

3.4　优化方案验证

为进一步验证优化模型的正确性和优化结果的合理性，采用有限元方法建立优化结果对应的传热模型，验证优化得到的设计变量是否满足约束条件，同时将有限元方法计算结果与一维简化模型计算结果进行对比，验证温度响应、结构隔热性能和力学指标的约束条件满足性。

在有限元建模中，耐高温层和基层的材料为线性模型，模型参数均为常数；隔热层材料参数为温度非线性模型，热导率参数随温度变化；单元类型为四边形平面单元；除耐高温层外壁面和基层内壁面外均为绝热边界条件，无外部热流输入；耐高温层厚度方向划分 5 层单元，隔热层厚度方向划分 20 层单元，基层厚度方向划分 5 层单元；横向划分 50 个单元。整个有限元模型的单元 1 510 个，节点 1 581 个。有限元建模采用最优优化结果，即粒子群优化方法的优化结果为：(1,3,1,0.000 501 m,0.013 6 m,0.002 1 m)，有限元模型如图 12 所示。

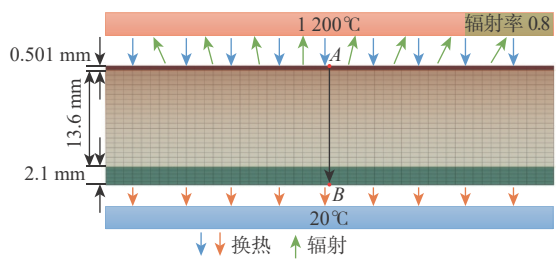


图 12　有限元模型

Fig. 12　Finite element model

耐高温层外壁面和基层内壁面的温度响应分布如图 13 和图 14 所示。耐高温层外壁面温度输出点为图 12 中的 A 点，基层内壁面温度输出点为图 12 中的 B 点，耐高温层外壁面温度在 21 s 时达到稳定，稳定温度为 862.76℃。基层内壁面温度在 21 s 时快速上升，从 24.715℃ 上升到了约 100℃。此外，从图 13 和图 14 中可知，本文一维模型计算温度响应结果与有限元计算结果基本一致，吻合度极好，也验证了本文传热响应计算的正确性。

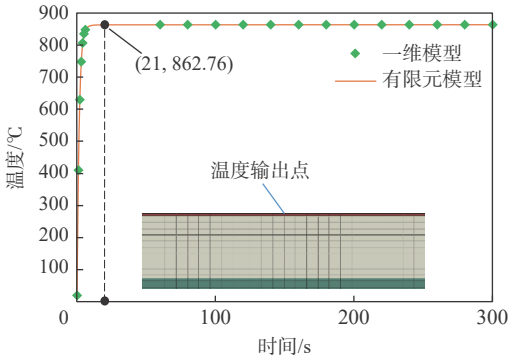


图 13 一维模型和有限元模型耐高温层外壁面温度响应对比

Fig. 13 Comparison of external wall temperature responses for high temperature-resistant layer between one-dimensional model and finite element model

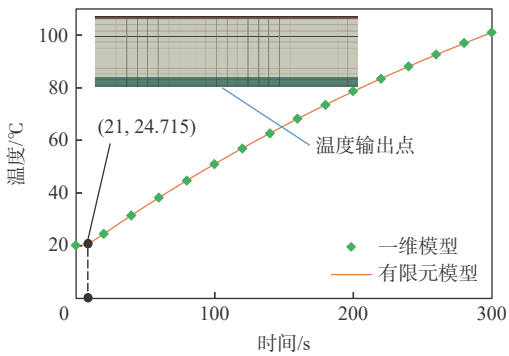


图 14 一维模型和有限元模型基层内壁面温度响应对比

Fig. 14 Comparison of inner wall temperature responses for base layer between one-dimensional model and finite element model

在分析了温度的时域响应之后，对温度的场分布进行分析。50 s 和 300 s 时的温度场分布如图 15 和图 16 所示。耐高温层和基层的温度分布基本均匀，温度梯度非常小；在隔热层部分，温度变化剧烈，温度梯度较大，隔热层在多层隔热结构中起主要防热作用。50 s 时，基层温度为 35℃，当时间为 300 s 时，基层温度上升到约 100℃。此外，当多层隔热结构的参数为 (1,3,1,0.000 501 m,0.013 6 m,0.002 1 m) 时，多层隔热结构的等效模量为 10 GPa，多层结构质量体密度为 420.56 kg/m^3，设计变量约束和力学性能约束均满足式 (24) 的约束条件，本文提出的考

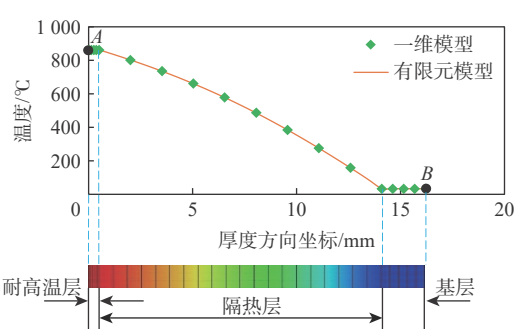


图 15 50 s 时温度场分布

Fig. 15 Temperature field distribution at 50 s

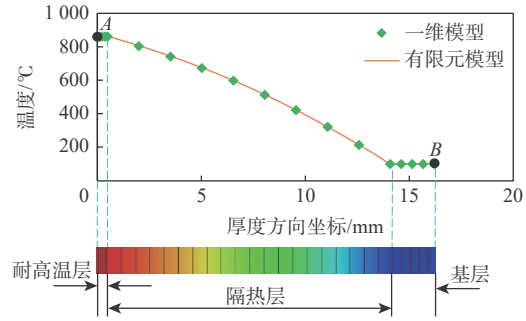


图 16 300 s 时温度场分布

Fig. 16 Temperature field distribution at 300 s

虑材料优选和材料热导率温变特性的高速飞行器多层隔热结构优化设计方法是正确、有效的，能在满足材料选择约束、尺寸约束、力学性能约束和隔热性能约束的条件下对隔热结构进行优化，进而达到隔热结构轻量化设计的目标。

在工程应用中，由于载荷不确定性、界面热阻效应、建模假设和近似、材料参数离散性等因素的存在，在建模中忽略了这些不确定性，得到的理论模型是实际物理模型的一种近似，可能导致本文方法得到的结果与应用对象的实际试验数据产生一定的误差。

4 结 论

隔热结构是保证高速飞行器结构、内部电器和载荷系统免受外部极端热载破坏的重要部件，多层隔热结构是其应用的典型构型方式之一，需要在满足隔热和一定承载能力的前提下进行轻量化优化设计。多层隔热结构的性能不仅与物理尺寸有关，还与材料优选、温度影响息息相关。

1) 本文考虑温变影响，建立了多层隔热结构瞬态热传导非线性模型。热导率的温变特性会导致在进行模型时间尺度离散递推时需要求解一个二阶非线性模型方程，本文针对该问题提出了具体的热传导方程数值求解策略。

2) 基于瞬态传热模型和各层备选材料数据集，引入材料热导率温变特性，建立了考虑材料优选的隔热结构力-热一体化优化模型，采用蒙特卡罗方法和粒子群优化方法分别对优化模型进行求解。

3) 优化结果表明，提出的优化模型和方法可使结构质量减轻 27.4%，隔热结构质量面密度为 9.252 2 kg/m^2。温度影响分析表明，若不考虑温度对材料性能的作用，得到的优化设计方案可能出现不满足温度设计约束的情况，导致结构失效。

4) 为验证结果的可靠性和合理性，将一维模型温度响应结果与有限元模型结果进行比较，两者一致性好，优化后隔热结构设计方案在 300 s 时基层

内壁面温度为 100℃, 最大程度进行了轻量化优化。

5) 从温度的场分布结果可知, 隔热层温度梯度大, 是主要的隔热贡献部分, 也是在总体设计时需要重点考虑的内容。

本文的研究成果能够为高速飞行器的总体设计技术提供支撑, 为加快高速飞行器在工程化道路中的轻量化奠定理论方法基础。

致谢　感谢余婧博士为本文工作提供的帮助。

参考文献（References）

[1] 赵玲, 吕国志, 任克亮, 等. 再入飞行器多层隔热结构优化分析[J]. 航空学报, 2007, 28(6): 1345-1350.
ZHAO L, LU G Z, REN K L, et al. Structural optimization of multi-layer insulation for reentry vehicles[J]. Acta Aeronautica et Astronautica Sinica, 2007, 28(6): 1345-1350(in Chinese).

[2] 程兴华. 高超声速飞行器防热壁板气动热弹性耦合建模与分析[D]. 长沙: 国防科学技术大学, 2012: 1-3.
CHENG X H. Coupled aerothermoelastic modeling and analysis of thermal protection panel for hypersonic vehicles[D]. Changsha: National University of Defense Technology, 2012: 1-3(in Chinese).

[3] 谢永旺, 夏雨, 许学伟, 等. 航天飞行器热防护系统研究概况及其发展趋势[J]. 空天技术, 2022(4): 73-86.
XIE Y W, XIA Y, XU X W, et al. Research status of thermal protection system for spacecraft and its development trend[J]. Aerospace Technology, 2022(4): 73-86(in Chinese).

[4] 杭子轩, 戴婷, 李珺. 基于 SiC 气凝胶的热防护结构隔热性能分析及优化[J]. 导弹与航天运载技术 (中英文), 2023(4): 6-14.
HANG Z X, DAI T, LI J. Analysis and optimization of thermal insulation property of a thermal protection structure based on SiC aerogel[J]. Missiles and Space Vehicles, 2023(4): 6-14(in Chinese).

[5] 叶红, 王志瑾. 高超声速飞行器热防护结构参数优化及对比分析[J]. 航天器环境工程, 2013, 30(5): 516-521.
YE H, WANG Z J. The optimization and comparison of thermal protection structures for hypersonic aircraft[J]. Spacecraft Environment Engineering, 2013, 30(5): 516-521(in Chinese).

[6] KOSTANOVSKII A V, ZEODINOV M G, KOSTANOVSKAYA M E. The determination of thermal conductivity and emissivity of graphite at high temperatures[J]. High Temperature, 2005, 43(5): 793-795.

[7] CHAR M I, CHANG F P, TAI B C. Inverse determination of thermal conductivity by differential quadrature method[J]. International Communications in Heat and Mass Transfer, 2008, 35(2): 113-119.

[8] DOMÍNGUEZ-MUÑOZ F, ANDERSON B, CEJUDO-LÓPEZ J M, et al. Uncertainty in the thermal conductivity of insulation materials[J]. Energy and Buildings, 2010, 42(11): 2159-2168.

[9] 吴大方, 郑力铭, 潘兵, 等. 非线性热环境下高温合金蜂窝板隔热性能研究[J]. 力学学报, 2012, 44(2): 297-307.
WU D F, ZHENG L M, PAN B, et al. Research on heat-shielding properties of superalloy honeycomb panel for non-linear high temperature environment[J]. Chinese Journal of Theoretical and Applied Mechanics, 2012, 44(2): 297-307(in Chinese).

[10] WASZKIELIS K M, BIAŁOBRZEWSKI I, NOWAK K W, et al. Determination of the thermal conductivity of composted material[J]. Measurement, 2014, 58: 441-447.

[11] ZHOU Q, WANG P, WU K D, et al. Performance of high-temperature lightweight multilayer insulations[J]. Applied Thermal Engineering, 2022, 211: 118436.

[12] 周祥发, 冯坚, 肖汉宁, 等. 二氧化硅气凝胶隔热复合材料的性能及其瞬态传热模拟[J]. 国防科技大学学报, 2009, 31(2): 36-40.
ZHOU X F, FENG J, XIAO H N, et al. Performance and heat transfer simulation of silica aerogel composites[J]. Journal of National University of Defense Technology, 2009, 31(2): 36-40(in Chinese).

[13] ABRAMENKO A N, KALINICHENKO A S, BURTSER Y, et al. Determination of the thermal conductivity of foam aluminum[J]. Journal of Engineering Physics and Thermophysics, 1999, 72(3): 369-373.

[14] 马忠辉, 孙秦, 王小军, 等. 热防护系统多层隔热结构传热分析及性能研究[J]. 宇航学报, 2003, 24(5): 543-546.
MA Z H, SUN Q, WANG X J, et al. TPS multi-layer insulation thermal analysis and performance study[J]. Journal of Astronautics, 2003, 24(5): 543-546(in Chinese).

[15] 徐世南, 吴催生. 导弹结构热防护一体化设计与优化[J]. 国防科技大学学报, 2020, 42(2): 78-84.
XU S N, WU C S. Design and optimization of structure-thermal protection integrated for missile[J]. Journal of National University of Defense Technology, 2020, 42(2): 78-84(in Chinese).

[16] 柴峻. 临近空间飞行器多层隔热结构传热分析及优化设计[D]. 哈尔滨: 哈尔滨工程大学, 2016: 36-39.
CHAI J. Heat transfer analysis and optimizing design of multi-layer isolate structure for near space aircraft[D]. Harbin: Harbin Engineering University, 2016: 36-39(in Chinese).

[17] 陈海坤, 姜召阳, 房景臣, 等. 防隔热一体化结构计算模拟研究[J]. 宇航材料工艺, 2009, 39(1): 14-16.
CHEN H K, JIANG Z Y, FANG J C, et al. Numerical study of integrated thermal protection/insulation structure[J]. Aerospace Materials & Technology, 2009, 39(1): 14-16(in Chinese).

[18] 余婧. 航天器在轨服务任务规划技术研究[D]. 长沙: 国防科学技术大学, 2015.
YU J. Research on spacecrafts on-oribt servicing mission planning [D]. Changsha: National University of Defense Technology, 2015 (in Chinese).

[19] SAAD M, BAKER D, REAVES R. Thermal characterization of carbon-carbon composites[C]//Proceedings of the International Mechanical Engineering Congress and Exposition. New York: ASME, 2011: 1233-1241.

[20] SUDHIR A, NAGARAJA A M, GURURAJA S. Effective mechanical properties of carbon-carbon composites[C]//Proceedings of the International Mechanical Engineering Congress and Exposition. New York: ASME, 2014: V001T01A017.

[21] LIAN M R, SU J, HE Y Y. Heat insulation performance of lattice thermal protection structure with reinforced plate[J]. Journal of Physics: Conference Series, 2021, 2044(1): 012003.

[22] LI G D, ZHANG C R, HU H F, et al. Preparation and mechanical properties of C/SiC nuts and bolts[J]. Materials Science and Engineering: A, 2012, 547: 1-5.

[23] PATEL M, PRASAD V V B, JAYARAM V. Heat conduction

mechanisms in hot pressed ZrB_2 and ZrB_2-SiC composites[J]. Journal of the European Ceramic Society, 2013, 33(10): 1615-1624.

[24] NAKAMORI F, OHISHI Y, MUTA H, et al. Mechanical and thermal properties of bulk ZrB_2[J]. Journal of Nuclear Materials, 2015, 467: 612-617.

[25] BOCK V, NILSSON O, BLUMM J, et al. Thermal properties of carbon aerogels[J]. Journal of Non-Crystalline Solids, 1995, 185(3): 233-239.

[26] HSIEH T H, HUANG Y S, SHEN M Y. Mechanical properties and toughness of carbon aerogel/epoxy polymer composites[J]. Journal of Materials Science, 2015, 50(8): 3258-3266.

[27] DENG Z S, WANG J, WU A M, et al. High strength SiO_2 aerogel insulation[J]. Journal of Non-Crystalline Solids, 1998, 225: 101-104.

[28] CHABI S, ROCHA V G, GARCÍA-TUÑÓN E, et al. Ultralight, strong, three-dimensional SiC structures[J]. ACS Nano, 2016, 10(2): 1871-1876.

[29] PARMENTER K E, MILSTEIN F. Mechanical properties of silica aerogels[J]. Journal of Non-Crystalline Solids, 1998, 223(3): 179-189.

[30] MURACA R F, WHITTICK J S. Materials data handbook: aluminum alloy 7075: NASA-CR-123773[R]. Washington, D.C.: NASA, 1972: 45-50.

[31] 《中国航空材料手册》编辑委员会. 中国航空材料手册 (第 2 卷): 变形高温合金、铸造高温合金[M]. 第 2 版. 北京: 中国标准出版社, 2002: 278-293.
China Aeronautical Materials Handbook Editorial Board. China aeronautical materials handbook (volume 2): deformed high temperature alloy, high temperature alloy[M]. 2nd ed. Beijing: Standards Press of China, 2002: 278-293(in Chinese).

[32] 《飞机设计手册》总编委会. 飞机设计手册 (第 3 册): 材料 (上、下)[M]. 北京: 航空工业出版社, 1997: 385-409.
Aircraft Design Handbook General Editorial Board. Aircraft design handbook (volume 3): material (upper and lower)[M]. Beijing: Aviation Industry Press, 1997: 385-409(in Chinese).

[33] WANG B H, CHENG L, WU Y H, et al. Research on ultra-high cycle fatigue performance and reliability life analysis of TC4 by laser shock peening[J]. Journal of Physics: Conference Series, 2022, 2174(1): 012037.

Optimization of multilayer thermal insulation structure for high-speed aircraft considering material optimization

HAO Dong，LIU Jianxia，SU Jie，HE Yuanyuan*

(Aerospace Technology Institute，China Aerodynamics Research and Development Center，Mianyang 621000，China)

Abstract：Thermal insulation structure is the key factor to ensure high-speed aircraft work safely and complete various tasks. The multilayer thermal insulation structure has attracted much attention because of its light weight and excellent thermal insulation effect. The performance of the multilayer thermal insulation structure is not only related to the physical size but also closely related to the material optimization and the temperature influence. In this paper, the transient nonlinear heat transfer model of the multilayer thermal insulation structure was established, and the numerical solution strategy of the heat transfer equation was proposed. Then, by considering the material optimization, introducing the temperature variation characteristic of the thermal conductivity of the material, and taking the geometric dimensions and equivalent mechanical properties of the multilayer thermal insulation structures as the optimization constraints, an integrated mechanical-thermal optimization model of the multilayer thermal insulation structure was established. The Monte-Carlo method and the particle swarm method were utilized to obtain the solution. The simulation results show that the proposed optimization model and method can reduce the structural mass by 27.4%. The temperature effect analysis shows that the temperature variation effect has an important influence on the optimization results. If the temperature effect on the material properties is not considered, the optimal design scheme may not meet the temperature design constraints, which further leads to structural failure. Finally, the reliability and rationality of the optimization scheme are verified by the finite element model. The optimization design method of the multilayer thermal insulation structure proposed in the paper provides technical support and lays a foundation for the overall design of high-speed aircraft.

Keywords：high-speed aircraft; thermal insulation; optimization model; Monte-Carlo; particle swarm

Received：2023-05-22；**Accepted**：2023-10-13；**Published Online**：2023-10-30 21：57

URL：link.cnki.net/urlid/11.2625.V.20231030.1035.002

* **Corresponding author**. E-mail：hyy63713@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0258

双臂空间机器人球形外包络抓捕控制

夏新会[1]，贾英宏[1,*]，张军[2,3]

(1. 北京航空航天大学 宇航学院，北京 100191；2. 北京控制工程研究所，北京 100190；
3. 空间智能控制技术重点实验室，北京 100190)

摘 要：针对双臂空间机器人抓捕目标的控制问题，设计一种利用双臂球形外包络捕获运动目标的抓捕方案。为避免机械臂抓捕时推离目标，进行双臂球形外包络的结构设计，通过几何特性约束目标运动，避免目标逃逸。根据球形外包络与目标的接触碰撞条件，分别建立碰撞检测模型与精确的球形外包络接触碰撞模型，利用球形外包络与目标之间的摩擦降低目标运动速度，为机械臂提供安全的抓捕环境。基于视觉伺服方法设计2部机械臂抓捕运动目标的实时期望轨迹，并基于逆动力学方法完成双臂协调抓捕控制律的设计。完成闭环控制系统仿真实验，自由漂浮双臂空间机器人对立方星目标进行球形外包络抓捕控制，仿真结果验证了所设计的球形外包络抓捕控制器的有效性。

关 键 词：双臂空间机器人；球形外包络；目标抓捕；轨迹规划；碰撞检测；接触力建模
中图分类号：V448.2
文献标志码：A **文章编号**：1001-5965(2025)05-1673-11

自第一颗人造卫星发射升空以来，太空中因航天器发射失败、年久退役、功能失效甚至机体撞击爆炸产生了数量众多的空间碎片[1]。目前，近地轨道中空间碎片的数量已远超现役的在轨航天器，并且随着时间的累积还在不断增加，太空环境的安全性受到严重的威胁[2-3]。国内外研究人员一致认为需要研究主动碎片清除技术以解决空间碎片问题，即利用主动航天器捕获空间碎片，使其脱离轨道或转移至坟墓轨道，从而完成碎片清除任务。

相比于非接触式捕获，利用空间机器人直接接触目标碎片进而实施捕获的优势有很多，如空间操作的可靠性更高、安全性更强、可提供更大的控制力等[4]。然而，随着空间操作任务的多样化和复杂化，单臂空间机器人逐渐无法满足任务需求。相比单臂空间机器人，双臂空间机器人系统负载能力更大、操作灵活性更高、任务适应性更强[5]。因此，利用双臂空间机器人系统捕获目标受到研究人员的广泛关注[6]。Raina 等[7]对双臂空间机器人捕获轨道目标的过程进行了碰撞建模，并针对碰撞后动力学不受控的情况设计了闭环系统的无反作用控制器。Xu 等[8]为规划双臂空间机器人末端执行器在捕获目标上的抓取姿态，将可操纵性指标扩展到双臂协作工作空间，并基于能力图评估了其协同操作的全局可操作性，最大限度地提高了空间机器人的可操作性。李梁等[9]基于反馈线性化与阻抗控制方法设计了自由漂浮双臂空间机器人抓捕移动目标的开链与闭链运动过程。

然而，利用空间机器人捕获目标也存在一定的问题，如机械臂的跟踪能力有限、对旋转的翻滚目标难以直接捕获、容易推离目标使其逃逸等。针对这些问题，常见的解决方法是利用摩擦降低目标的速度，从而提高机械臂的抓捕成功率。Nishida 和

收稿日期：2023-05-18；**录用日期**：2023-06-12；**网络出版时间**：2023-06-30 16:42
网络出版地址：link.cnki.net/urlid/11.2625.V.20230630.1608.005
基金项目：科技创新2030“新一代人工智能”重大项目(2018AAA0103003)
*通信作者. E-mail：jia_yingh@163.com

引用格式：夏新会，贾英宏，张军. 双臂空间机器人球形外包络抓捕控制[J]. 北京航空航天大学学报，2025，51(5)：1673-1683.
XIA X H，JIA Y H，ZHANG J. Spherical envelope capture control for dual-arm space robots[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1673-1683 (in Chinese).

Kawamoto[10]提出了在空间机械臂末端安装毛刷型的末端作用器，通过毛刷与目标之间的摩擦降低目标的旋转速度，逐渐使其进入机械臂的抓捕能力范围之内，完成机械臂对目标的捕获，然而因毛刷质量小、变形复杂，其接触力建模困难，且接触中产生的正压力容易导致目标逃脱。Chen等[11]采用混合阻抗控制方法设计了旋转衰减控制器，通过对摩擦进行控制，降低目标的旋转速度，实现对目标的抓捕。采用柔顺控制策略可实现稳定地降低目标角速度，解决机械臂对旋转目标难以直接捕获的问题，但仍存在将目标推离空间机械臂的可能性。

针对目标逃逸的问题，Rimon[12]提出利用多个刚性手指形成包络，将目标几何约束在封闭区域中，从而避免目标逃脱的方法。Maeda等[13]在该方法的基础上做了改进，即在刚性多指外包裹一层柔性组织，利用刚柔结合的多指机械手合拢包裹住目标，完成抓捕。需要说明的是，上述方法只局限于地面机械手指的应用。孙冲等[14]以3根手指组成的三爪抓捕机构作为末端作用器，安装在单臂空间机器人末端，通过调整三爪与基座的角度形成运动包络来约束目标的运动，实现对抓捕目标的消旋。包络的优势在于能够包裹空间目标，约束其运动，且包络与目标之间的摩擦可有效消除两者之间的相对运动，为成功抓捕目标奠定基础。然而，采用末端作用器组成的包络直接约束目标存在损坏接触表面的风险。

针对上述问题，本文提出了一种利用双臂球形外包络捕获运动目标的抓捕方案，充分结合了双臂空间机器人与运动包络的抓捕优势。首先，基于球形外包络的结构设计，在2部机械臂实施抓捕前，将目标几何约束于球形外包络中。通过建立碰撞检测算法与摩擦接触力模型，利用球形外包络与目标之间的摩擦，降低两者之间的相对运动，避免目标逃逸与损坏末端作用器的风险。其次，利用基于位置的视觉伺服方法设计双臂实时期望轨迹，并完成双臂协调抓捕控制律的设计。最后，选取自由漂浮双臂空间机器人与运动立方体目标为研究对象，通过闭环控制系统仿真验证了该抓捕方案的有效性。

1 球形外包络设计

本文所设计的利用球形外包络抓捕目标的概念图如图1所示。双臂空间机器人由航天器平台、2部刚体机械臂、球形外包络和1根轻质杆组成。球形外包络通过轻质杆固连于平台上，平台、轻质杆与球形外包络可视为一个刚体。由于平台的质量惯量比轻质杆与球形外包络高出多个数量级，轻质杆与球形外包络的质量可忽略不计。

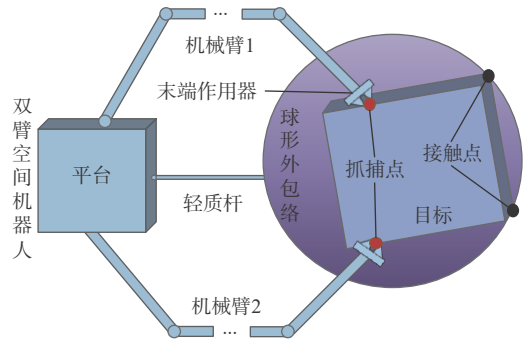


图1 双臂空间机器人球形外包络抓捕概念图

Fig. 1 Spherical cage-based capture concept of a dual-arm space robot

在目标与球形外包络碰撞接触时，轻质杆可以传递目标对球形外包络的反作用力和力矩。因此，球形外包络设计的目的是将目标几何约束于球形外包络中，其优势如下：

1）利用球形外包络笼住目标，避免机械臂将目标推离，便于机械臂实施抓捕。

2）通过球形外包络与目标的接触碰撞，降低目标与平台之间的相对运动，提高抓捕成功率。

设目标位姿的矢量列阵为$\boldsymbol{X}_{\mathrm{t}}$，其自由构型空间为$\boldsymbol{C}_{\mathrm{free}}$，满足$\{\boldsymbol{X}_{\mathrm{t}}\}\subset\boldsymbol{C}_{\mathrm{free}}$。设$\boldsymbol{C}_{\mathrm{all}}$为全集，则目标的逃逸空间为

$$\boldsymbol{C}_{\mathrm{escape}}=\boldsymbol{C}_{\mathrm{all}}\setminus\left\{\bigcup_{\boldsymbol{X}_{\mathrm{t}}}\boldsymbol{C}_{\mathrm{free}}(\boldsymbol{X}_{\mathrm{t}})\right\}\tag{1}$$

基于文献[15]可知，形成球形外包络的充分必要条件为：存在球形外包络的非零封闭空间$\boldsymbol{C}_{\mathrm{closed}}$，满足：

$$\begin{cases}\{\boldsymbol{X}_{\mathrm{t}}\}\subset\boldsymbol{C}_{\mathrm{closed}}\\ \boldsymbol{C}_{\mathrm{closed}}\cap\boldsymbol{C}_{\mathrm{escape}}=\varnothing\end{cases}\tag{2}$$

2 系统建模

2.1 坐标系描述

假设双臂空间机器人第$i(i=1,2)$部机械臂中含有n_i个臂杆，为方便建模，需要先建立体坐标系、安装坐标系、关节参考坐标系及抓捕坐标系4种坐标系，如图2所示。

体坐标系包括航天器平台的本体坐标系$\mathcal{F}_{\mathrm{b}}$、目标体坐标系$\mathcal{F}_{\mathrm{t}}$、球形外包络体坐标系$\mathcal{F}_{\mathrm{env}}$和臂杆体坐标系$\mathcal{F}_{i,j}(i=1,2,j=1,2,\cdots,n_i)$。$\mathcal{F}_{\mathrm{b}}$固连于平台质心，$\mathcal{F}_{\mathrm{t}}$固连于目标体质心，$\mathcal{F}_{\mathrm{env}}$固连于球形外包络的球心，$\mathcal{F}_{i,j}$固连于第$i$部机械臂中第$j$个臂杆上，原点位于第$j$个关节上。

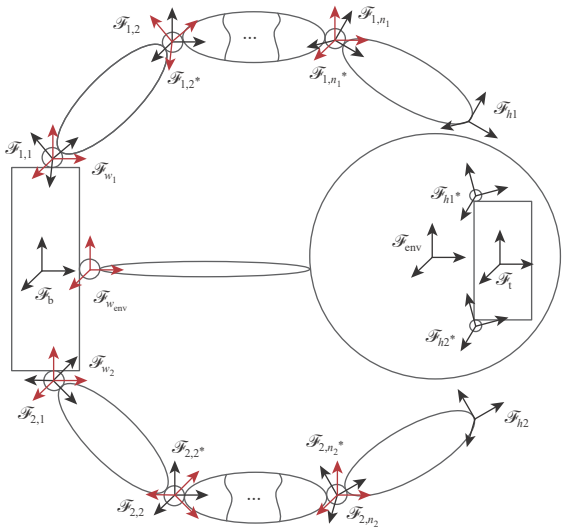


图 2　坐标系示意图

Fig. 2　Coordinate system

安装坐标系包括臂安装坐标系$\mathcal{F}_{w_i}(i=1,2)$和球形外包络安装坐标系$\mathcal{F}_{w_{\text{env}}}$, $\mathcal{F}_{w_i}$位于第i部机械臂与平台的铰接点处, 固连于平台上; $\mathcal{F}_{w_{\text{env}}}$位于轻质杆与平台的连接点处, 同样固连在平台上。

为定义机械臂i的第j个关节角q_j^i, 定义了臂杆的关节参考坐标系$\mathcal{F}_{i,j^*}(i=1,2,j=2,\cdots,n_i)$, $\mathcal{F}_{i,j^*}$固连于第$j-1$个臂杆上, 其转轴与$\mathcal{F}_{i,j}$的转轴$\mathcal{F}_{i,j}$平行, 与$\mathcal{F}_{i,j}$有相同的原点, 将由$\mathcal{F}_{i,j^*}$沿转轴正向转到$\mathcal{F}_{i,j}$的角度定义为q_j^i。

此外, 在机械臂i的末端(第n_i根杆上)固连末端作用器坐标系$\mathcal{F}_{hi}$, $\mathcal{F}_{hi}$在意义上与关节参考坐标系相同。抓捕坐标系$\mathcal{F}_{hi^*}$位于机械臂与目标的抓捕点处, 固连于目标上, 当第i部机械臂完成抓捕时, $\mathcal{F}_{hi}$与$\mathcal{F}_{hi^*}$重合。

2.2　双臂空间机器人建模

选取双臂空间机器人的广义速率为$\boldsymbol{u}_{\text{c}}=\left[\boldsymbol{v}_{\text{b}}^{\text{T}},\boldsymbol{\omega}_{\text{b}}^{\text{T}},\dot{\boldsymbol{q}}^{\text{T}}\right]^{\text{T}}\in\mathbf{R}^{6+N}$, 其中, $N=n_1+n_2$, $\boldsymbol{v}_{\text{b}}$和$\boldsymbol{\omega}_{\text{b}}$分别为平台的线速度和角速度, $\dot{\boldsymbol{q}}=\left[\dot{\boldsymbol{q}}^{1\text{T}},\dot{\boldsymbol{q}}^{2\text{T}}\right]^{\text{T}}\in\mathbf{R}^{N}$为 2 部机械臂的关节角速度, $\dot{\boldsymbol{q}}^i=\left[\dot{q}_1^i,\dot{q}_2^i,\cdots,\dot{q}_{n_i}^i\right]^{\text{T}}(i=1,2)$为第 i 部机械臂各臂杆的关节角速度。参见文献 [16], 可以得到双臂空间机器人的动力学方程为

$$\boldsymbol{H}_{\text{c}}\dot{\boldsymbol{u}}_{\text{c}}+\boldsymbol{c}_{\text{c}}=\boldsymbol{F}_{\text{c}}+\boldsymbol{J}_{\text{h}}^{\text{T}}\boldsymbol{F}_{\text{h}}-\boldsymbol{J}_{\text{env}^*}^{\text{T}}\left[\boldsymbol{f}_{\text{env}}^{\text{T}},\boldsymbol{0}\right]^{\text{T}}\tag{3}$$

式中: $\boldsymbol{H}_{\text{c}}\in\mathbf{R}^{(6+N)\times(6+N)}$为双臂空间机器人广义质量阵; $\boldsymbol{c}_{\text{c}}\in\mathbf{R}^{6+N}$为双臂空间机器人非线性广义惯性力; $\boldsymbol{F}_{\text{c}}=\left[\boldsymbol{F}_{\text{b}}^{\text{T}},\boldsymbol{\tau}_{\text{m}}^{\text{T}}\right]^{\text{T}}\in\mathbf{R}^{6+N}$包含平台的外力/力矩$\boldsymbol{F}_{\text{b}}\in\mathbf{R}^6$与 2 部机械臂的关节控制力矩$\boldsymbol{\tau}_{\text{m}}\in\mathbf{R}^N$; $\boldsymbol{F}_{\text{h}}\in\mathbf{R}^{12}$为 2 部机械臂末端受到的目标接触力/力矩; $\boldsymbol{f}_{\text{env}}\in\mathbf{R}^6$为目标受到的总包络接触力; $\boldsymbol{J}_{\text{h}}\in\mathbf{R}^{12\times(6+N)}$为力雅可比矩阵; $\boldsymbol{J}_{\text{env}^*}\in\mathbf{R}^{3n_{\text{env}}\times6}$为平台与球形外包络碰撞点之间的雅可比矩阵, 可将球形外包络接触力映射到惯性坐标系中。

为得到雅可比矩阵$\boldsymbol{J}_{\text{env}^*}$, 给出目标与球形外包络的相对位置关系, 如图 3 所示。假设目标和球形外包络的中心点分别为o_{t}和o_{env}, 目标与球形外包络接触点的个数为n_{env}, 目标上与球形外包络的第i个接触点为$p_i(i=1,2,\cdots,n_{\text{nev}})$, 对应在球形外包络上的接触点为$p_i^*(i=1,2,\cdots,n_{\text{nev}})$, 则雅可比矩阵$\boldsymbol{J}_{\text{env}^*}$的表达式为

$$\begin{cases}\boldsymbol{J}_{\text{env}^*}=\left[\boldsymbol{J}_{\text{env}^*}^{1\text{T}},\boldsymbol{J}_{\text{env}^*}^{2\text{T}},\cdots,\boldsymbol{J}_{\text{env}^*}^{n_{\text{env}}\text{T}}\right]^{\text{T}}\\ \boldsymbol{J}_{\text{env}^*}^i=\left[\boldsymbol{I}_3\quad\left[{}^{\text{b}}\boldsymbol{p}^{p_i^*}\right]^{\times}\right]\end{cases}\tag{4}$$

式中: ${}^{\text{b}}\boldsymbol{p}^{p_i^*}$为球形外包络接触点$p_i^*$相对平台的位移的矢量列阵, 表达式为${}^{\text{b}}\boldsymbol{p}^{p_i^*}={}^{\text{b}}\boldsymbol{p}^{\text{env}}+{}^{\text{b}}\boldsymbol{A}^{\text{env}}{}^{\text{env}}\boldsymbol{p}^{p_i^*}$, 其中, ${}^{\text{b}}\boldsymbol{A}^{\text{env}}$为平台本体系到包络体坐标系的坐标转换矩阵; 上标 "×" 表示矢量的叉乘反对称阵。

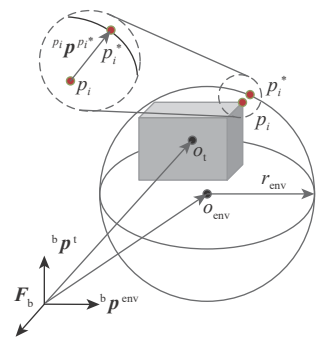


图 3　目标与球形外包络的相对位置关系

Fig. 3　Relative positon of object to spherical cage

2.3　目标建模

基于 Kane 方程, 选取目标的广义速率为$\boldsymbol{u}_{\text{t}}=\left[\boldsymbol{v}_{\text{t}}^{\text{T}},\boldsymbol{\omega}_{\text{t}}^{\text{T}}\right]^{\text{T}}\in\mathbf{R}^6$, 其中, $\boldsymbol{v}_{\text{t}}$和$\boldsymbol{\omega}_{\text{t}}$分别为目标的线速度和角速度, 可建立目标单刚体的六自由度动力学方程为

$$\boldsymbol{H}_{\text{t}}\dot{\boldsymbol{u}}_{\text{t}}+\boldsymbol{c}_{\text{t}}=-\boldsymbol{J}_{\text{h}^*}^{\text{T}}\boldsymbol{F}_{\text{h}}+\boldsymbol{J}_{\text{env}}^{\text{T}}\boldsymbol{f}_{\text{env}}\tag{5}$$

式中: $\boldsymbol{H}_{\text{t}}$为目标体的广义质量阵; $\boldsymbol{c}_{\text{t}}$为目标的非线性广义惯性力, 等式右侧为目标所受到的 2 种碰撞接触力的影响项, 其中, $\boldsymbol{J}_{\text{env}}\in\mathbf{R}^{3n_{\text{env}}\times6}$为平台与目标接触点之间的雅可比矩阵, 其表达式为

$$\begin{cases}\boldsymbol{J}_{\text{env}}=\left[\boldsymbol{J}_{\text{env}}^{1\text{T}},\boldsymbol{J}_{\text{env}}^{2\text{T}},\cdots,\boldsymbol{J}_{\text{env}}^{n_{\text{env}}\text{T}}\right]^{\text{T}}\\ \boldsymbol{J}_{\text{env}}^i=\left[\begin{array}{cc}\boldsymbol{I}_3 & \left[{}^{\text{env}}\boldsymbol{p}^{p_i}\right]^{\times}\end{array}\right]\end{cases}\tag{6}$$

式中: ${}^{\text{env}}\boldsymbol{p}^{p_i}$为目标接触点$p_i$相对球形外包络中心的位移的矢量列阵。

2.4　球形外包络接触力建模

在双臂空间机器人对目标实施抓捕前, 将目标几何约束于球形外包络中, 利用目标与球形外包络之间的碰撞接触力, 可以降低目标与双臂空间机器人之间的相对运动, 提高抓捕成功率。然而, 理论

上接触力是一种极其复杂的非线性项，目前接触力的建模仍是一大难点。

2.4.1 碰撞检测算法

在目标与球形外包络接触碰撞过程中，接触点的位置、个数及接触时刻均不确定，因此，在进行球形外包络接触建模前需要进行碰撞检测。本文做如下假设：

1）球形外包络的材料特性满足使目标始终约束于球形外包络中。

2）目标视为一个六自由度单刚体，几何结构为长方体且只有8个顶点可与球形外包络接触碰撞，即$n_{\text{env}}\in[0,8]$，且n_{env}为整数。

3）同一时刻最多有4个顶点与球形外包络接触碰撞。

假设平台相对惯性坐标系的位置为${}^{\text{e}}\boldsymbol{p}^{\text{b}}$、姿态角为${}^{\text{e}}\boldsymbol{\Phi}^{\text{b}}$、线速度为${}^{\text{e}}\boldsymbol{v}^{\text{b}}$、角速度为${}^{\text{e}}\boldsymbol{\omega}^{\text{b}}$，给定球形外包络中心相对平台的位置${}^{\text{b}}\boldsymbol{p}^{\text{env}}$，$r_{\text{env}}$为给定的球形外包络球半径，${}^{\text{e}}\boldsymbol{R}^{\text{b}}$为平台坐标系相对惯性坐标系的坐标转换矩阵，则球形外包络中心相对惯性坐标系的运动变量为

$$\begin{cases}{}^{\text{e}}\boldsymbol{p}^{\text{env}}={}^{\text{e}}\boldsymbol{p}^{\text{b}}+{}^{\text{e}}\boldsymbol{R}^{\text{b}}\,{}^{\text{b}}\boldsymbol{p}^{\text{env}}\\ {}^{\text{e}}\dot{\boldsymbol{p}}^{\text{env}}={}^{\text{e}}\boldsymbol{R}^{\text{b}}\,{}^{\text{e}}\boldsymbol{v}^{\text{b}}-{}^{\text{e}}\boldsymbol{R}^{\text{b}}\left[{}^{\text{b}}\boldsymbol{p}^{\text{env}}\right]^{\times}{}^{\text{e}}\boldsymbol{\omega}^{\text{b}}\\ {}^{\text{e}}\boldsymbol{\Phi}^{\text{env}}={}^{\text{e}}\boldsymbol{\Phi}^{\text{b}}\\ {}^{\text{e}}\boldsymbol{\omega}^{\text{env}}={}^{\text{e}}\boldsymbol{\omega}^{\text{b}}\end{cases}\tag{7}$$

假设目标体相对惯性坐标系的位置为${}^{\text{e}}\boldsymbol{p}^{\text{t}}$、姿态角为${}^{\text{e}}\boldsymbol{\Phi}^{\text{t}}$、线速度为${}^{\text{e}}\boldsymbol{v}^{\text{t}}$、角速度为${}^{\text{e}}\boldsymbol{\omega}^{\text{t}}$，接触点$p_i$相对目标体质心的位置为${}^{\text{t}}\boldsymbol{p}^{p_i}$，目标本体坐标系相对惯性坐标系的坐标转换矩阵为${}^{\text{e}}\boldsymbol{R}^{\text{t}}$，则在碰撞时刻目标接触点$p_i$与球形外包络接触点$p_i^*$相对包络中心的位置可表示为

$${}^{\text{env}}\boldsymbol{p}^{p_i}={}^{\text{e}}\boldsymbol{p}^{\text{t}}+{}^{\text{e}}\boldsymbol{R}^{\text{t}}\,{}^{\text{t}}\boldsymbol{p}^{p_i}-{}^{\text{e}}\boldsymbol{p}^{\text{env}}\tag{8}$$

$${}^{\text{env}}\boldsymbol{p}^{p_i^*}=r_{\text{env}}\frac{{}^{\text{env}}\boldsymbol{p}^{p_i}}{\|{}^{\text{env}}\boldsymbol{p}^{p_i}\|}\tag{9}$$

设接触点p_i处的总接触力为$\boldsymbol{f}_{\text{env}}^i$，则目标与球形外包络接触点$p_i$处的接触开关条件为

$$\boldsymbol{f}_{\text{env}}^i=\begin{cases}\boldsymbol{0} & r_{\text{env}}>\|{}^{\text{env}}\boldsymbol{p}^{p_i}\|\\ f_{\text{env}}^i & \|{}^{\text{env}}\boldsymbol{p}^{p_i}\|\geqslant r_{\text{env}}\end{cases}\tag{10}$$

即当满足$\|{}^{\text{env}}\boldsymbol{p}^{p_i}\|\geqslant r_{\text{env}}$时，目标与球形外包络在接触点$p_i$处接触碰撞；否则，目标与球形外包络未产生接触碰撞。

2.4.2 接触力建模

本文考虑摩擦因素，将目标接触点p_i处的接触力$\boldsymbol{f}_{\text{env}}^i$分解为法向碰撞力$\boldsymbol{f}_{\text{n}}^i$与切向摩擦力$\boldsymbol{f}_{\text{t}}^i$，如图4所示。

在接触力建模过程中，需要先计算目标接触点p_i与球形外包络接触点p_i^*之间的相对运动状态。

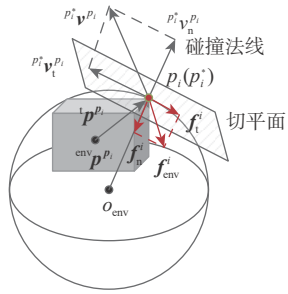


图4 接触点约束力

Fig. 4 Constraining force on contact points

2个接触点相对惯性坐标系的运动变量为

$$\begin{cases}{}^{\text{e}}\boldsymbol{p}^{p_i}={}^{\text{e}}\boldsymbol{p}^{\text{t}}+{}^{\text{e}}\boldsymbol{R}^{\text{t}}\,{}^{\text{t}}\boldsymbol{p}^{p_i}\\ {}^{\text{e}}\boldsymbol{v}^{p_i}={}^{\text{e}}\boldsymbol{v}^{\text{t}}-{}^{\text{e}}\boldsymbol{R}^{\text{t}}[{}^{\text{t}}\boldsymbol{p}^{p_i}]^{\times}{}^{\text{e}}\boldsymbol{\omega}^{\text{t}}\end{cases}\tag{11}$$

$$\begin{cases}{}^{\text{e}}\boldsymbol{p}^{p_i^*}={}^{\text{e}}\boldsymbol{p}^{\text{env}}+{}^{\text{env}}\boldsymbol{p}^{p_i^*}\\ {}^{\text{e}}\dot{\boldsymbol{p}}^{p_i^*}={}^{\text{e}}\dot{\boldsymbol{p}}^{\text{env}}-[{}^{\text{env}}\boldsymbol{p}^{p_i^*}]^{\times}{}^{\text{e}}\boldsymbol{\omega}^{\text{env}}\end{cases}\tag{12}$$

则接触点p_i相对接触点p_i^*的运动变量为

$$\begin{cases}{}^{p_i^*}\boldsymbol{p}^{p_i}={}^{\text{e}}\boldsymbol{p}^{p_i}-{}^{\text{e}}\boldsymbol{p}^{p_i^*}\\ {}^{p_i^*}\boldsymbol{v}^{p_i}={}^{\text{e}}\boldsymbol{v}^{p_i}-{}^{\text{e}}\boldsymbol{v}^{p_i^*}\end{cases}\tag{13}$$

碰撞法线的单位矢阵$\boldsymbol{l}_{\text{n}}$为

$$\boldsymbol{l}_{\text{n}}=\frac{{}^{\text{env}}\boldsymbol{p}^{p_i}}{\|{}^{\text{env}}\boldsymbol{p}^{p_i}\|}\tag{14}$$

相对运动变量${}^{p_i^*}\boldsymbol{p}^{p_i}$与${}^{p_i^*}\boldsymbol{v}^{p_i}$的法向分量为

$$\begin{cases}{}^{p_i^*}\boldsymbol{p}_{\text{n}}^{p_i}=[{}^{p_i^*}\boldsymbol{p}^{p_i}]^{\text{T}}\boldsymbol{l}_{\text{n}}\boldsymbol{l}_{\text{n}}\\ {}^{p_i^*}\boldsymbol{v}_{\text{n}}^{p_i}=[{}^{p_i^*}\dot{\boldsymbol{p}}^{p_i}]^{\text{T}}\boldsymbol{l}_{\text{n}}\boldsymbol{l}_{\text{n}}\end{cases}\tag{15}$$

则切向单位矢阵$\boldsymbol{l}_{\text{t}}$为

$$\boldsymbol{l}_{\text{t}}=\frac{{}^{p_i^*}\dot{\boldsymbol{p}}^{p_i}-{}^{p_i^*}\dot{\boldsymbol{p}}_{\text{n}}^{p_i}}{\left\|{}^{p_i^*}\dot{\boldsymbol{p}}^{p_i}-{}^{p_i^*}\dot{\boldsymbol{p}}_{\text{n}}^{p_i}\right\|}\tag{16}$$

分别采用弹簧阻尼模型与库仑摩擦模型计算目标接触点的法向正压力与切向摩擦力。基于弹簧阻尼模型[17]计算接触点p_i的法向正压力为

$$\boldsymbol{f}_{\text{n}}^i=\begin{cases}\boldsymbol{0} & r_{\text{env}}>\|{}^{\text{env}}\boldsymbol{p}^{p_i}\|\\ -k^{p_i^*}\boldsymbol{p}_{\text{n}}^{p_i}-c^{p_i^*}\boldsymbol{v}_{\text{n}}^{p_i} & \|{}^{\text{env}}\boldsymbol{p}^{p_i}\|\geqslant r_{\text{env}}\end{cases}\tag{17}$$

式中：k和c分别为弹性系数和刚度系数。

根据库仑摩擦模型，摩擦力大小与正压力成正比，则采用库仑摩擦模型，摩擦系数为μ，计算接触点p_i点的切向摩擦力为

$$\boldsymbol{f}_{\text{t}}^i=-\mu\left\|\boldsymbol{f}_{\text{n}}^i\right\|\boldsymbol{l}_{\text{t}}\tag{18}$$

则目标在接触点p_i点处所受的总接触力为

$$\boldsymbol{f}_{\text{env}}^i=\boldsymbol{f}_{\text{n}}^i+\boldsymbol{f}_{\text{t}}^i\tag{19}$$

目标受到的总包络接触力由各碰撞点处的接触力并列组成，即

$$\boldsymbol{f}_{\text{env}}=\left[\boldsymbol{f}_{\text{env}}^{1\text{T}},\boldsymbol{f}_{\text{env}}^{2\text{T}},\cdots,\boldsymbol{f}_{\text{env}}^{n_{\text{env}}\text{T}}\right]^{\text{T}}\tag{20}$$

综上所述，球形外包络接触力建模的计算流程如图5所示。

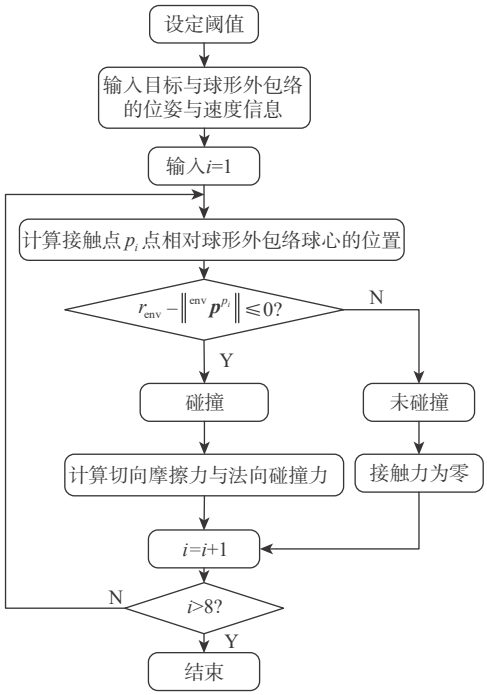


图 5　球形外包络接触建模流程

Fig. 5　Contact modeling process for spherical cage

3　抓捕控制器设计

为验证所设计的球形外包络对双臂空间机器人抓捕目标时的有效性，本节先利用基于位置的视觉伺服方法完成双臂抓捕运动目标的实时轨迹规划，再设计考虑球形外包络时的双臂协调抓捕控制律，使 2 部机械臂准确跟踪上实时的期望轨迹，完成对目标的捕获。系统的闭环控制框图如图 6 所示。

3.1　轨迹规划

在实际工程应用中，空间机器人的每部机械臂末端通常安装有手眼相机和传感器，其作用是通过手眼相机实时对目标上的抓捕点进行拍摄，再经过传感器测量得到目标与相机之间的相对位姿信息。

定义第 i 部机械臂的相机坐标系为$\mathcal{F}_{ci}$，其原点位于末端手眼相机上，指向与末端作用器坐标系$\mathcal{F}_{hi}$相同。在进行轨迹规划前，作如下假设：

1）假设相机坐标系$\mathcal{F}_{ci}$与末端作用器坐标系$\mathcal{F}_{hi}$重合。

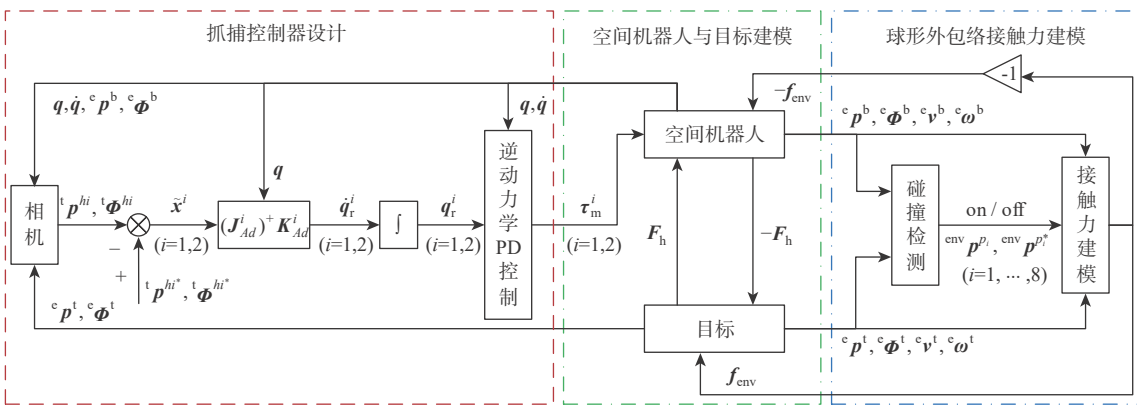


图 6　闭环控制框图

Fig. 6　Closed-loop control block diagram

2）通过图像处理，可以得到目标本体坐标系$\mathcal{F}_{\text{t}}$相对相机坐标系$\mathcal{F}_{ci}$（即末端作用器坐标系$\mathcal{F}_{hi}$）的位姿信息。

为便于表述 2 个坐标系之间的位姿信息，引入齐次变换阵的概念。假设存在 2 个坐标系：B系和C系，设${}^{\text{B}}\boldsymbol{p}^{\text{C}}$和${}^{\text{B}}\boldsymbol{R}^{\text{C}}$分别为$C$系相对$B$系的位置矢阵和坐标转换矩阵，则 2 个坐标系之间的齐次变换阵${}^{\text{B}}\boldsymbol{T}^{\text{C}}$为

$$ {}^{\text{B}}\boldsymbol{T}^{\text{C}}=\begin{bmatrix} {}^{\text{B}}\boldsymbol{R}^{\text{C}} & {}^{\text{B}}\boldsymbol{p}^{\text{C}} \\ \boldsymbol{0}_{1\times 3} & 1 \end{bmatrix} \tag{21} $$

通过 2 个假设可知，目标本体地球坐标系$\mathcal{F}_{\text{t}}$相对末端作用器地球坐标系$\mathcal{F}_{hi}$的齐次变换阵${}^{hi}\boldsymbol{T}^{\text{t}}(i=1,2)$可以通过图像处理得到。由于抓捕点在目标上是固定的，即${}^{\text{t}}\boldsymbol{T}^{hi^*}$是常值矩阵，则第 i 部机械臂末端与其对应的抓捕点之间的齐次变换阵为

$$ {}^{hi}\boldsymbol{T}^{hi^*}={}^{hi}\boldsymbol{T}^{\text{t}}\,{}^{\text{t}}\boldsymbol{T}^{hi^*}=\begin{bmatrix} {}^{hi}\boldsymbol{R}^{hi^*} & {}^{hi}\boldsymbol{p}^{hi^*} \\ \boldsymbol{0}_{1\times 3} & 1 \end{bmatrix} \tag{22} $$

根据式 (22)，可以得到第 i 部机械臂末端作用器与对应抓捕点之间的相对位置${}^{hi}\boldsymbol{p}^{hi^*}$和姿态${}^{hi}\boldsymbol{R}^{hi^*}$。定义第 i 部机械臂末端的抓捕位姿误差$\tilde{\boldsymbol{x}}^i$为

$$ \tilde{\boldsymbol{x}}^i=-\left[\left({}^{hi}\boldsymbol{p}^{hi^*}\right)^{\text{T}},\left({}^{hi}\boldsymbol{\theta}^{hi^*}\right)^{\text{T}}\right]^{\text{T}} \tag{23} $$

式中：${}^{hi}\boldsymbol{\theta}^{hi^*}$为目标抓捕点相对机械臂末端的姿态角，可由${}^{hi}\boldsymbol{R}^{hi^*}$直接得到。

为便于后续进行数值仿真，基于机械臂正运动学，对式 (23) 求导，可得第 i 部机械臂末端的抓捕速度误差$\dot{\tilde{\boldsymbol{x}}}^i$为

$$ \dot{\tilde{\boldsymbol{x}}}^i=-\boldsymbol{J}^i_{Ad}\left(\boldsymbol{q}^i,\tilde{\boldsymbol{x}}^i\right)\dot{\boldsymbol{q}}^i \tag{24} $$

式中：$\boldsymbol{J}_{Ad}^{i}\left(\boldsymbol{q}^{i},\tilde{\boldsymbol{x}}^{i}\right)$为类雅可比矩阵，与关节角$\boldsymbol{q}^{i}$和抓捕位姿误差$\tilde{\boldsymbol{x}}^{i}$有关，其表达式为

$$\boldsymbol{J}_{Ad}^{i}\left(\boldsymbol{q}^{i},\tilde{\boldsymbol{x}}^{i}\right)=\begin{bmatrix}-\boldsymbol{E}_{3} & ({}^{hi}\boldsymbol{p}^{hi*})^{\times}\\ \boldsymbol{0}_{3\times3} & -\boldsymbol{T}^{-1}({}^{hi}\boldsymbol{\theta}^{hi*})\end{bmatrix}\boldsymbol{J}_{\mathrm{m}}^{i}\left(\boldsymbol{q}^{i}\right)\tag{25}$$

式中：$\boldsymbol{J}_{\mathrm{m}}^{i}\left(\boldsymbol{q}^{i}\right)$为第$i$部机械臂相对平台的雅可比矩阵。

为便于后续进行控制律的设计，需要将工作空间中末端作用器的期望轨迹转换成关节空间中各机械臂关节角的期望轨迹。根据式(24)，可选取关节角速度指令为

$$\dot{\boldsymbol{q}}_{\mathrm{r}}^{i}=(\boldsymbol{J}_{Ad}^{i})^{+}\left(\boldsymbol{q}^{i},\tilde{\boldsymbol{x}}^{i}\right)\boldsymbol{K}_{Ad}^{i}\tilde{\boldsymbol{x}}^{i}\tag{26}$$

式中：上标“+”为求矩阵伪逆的符号；$\boldsymbol{K}_{Ad}^{i}$为待选取正定对称阵。

为防止机械臂关节角速度超出限制，选取机械臂关节角速度限制矢量$\dot{\boldsymbol{q}}_{\max}^{i}$对关节角速度指令进行等比缩放。设第$i$部机械臂中超出关节角速度比例最大的关节为第$j$个关节，则使用如下缩放律：

$$\dot{\boldsymbol{q}}_{\mathrm{r}}^{i}\Leftarrow\frac{\dot{q}_{\max,j}^{i}}{\dot{q}_{j}^{i}}\dot{\boldsymbol{q}}_{\mathrm{r}}^{i}\tag{27}$$

式中：“$\Leftarrow$”为赋给符号。

对式(27)进行积分可以得到关节角指令$\boldsymbol{q}_{\mathrm{r}}^{i}$。至此，为实现抓捕，每部机械臂各关节的期望关节角轨迹均已得到。

3.2 控制律设计

考虑双臂空间机器人中各机械臂独立抓捕目标对应的抓捕点，其控制任务是针对每部机械臂设计控制律，使每部机械臂的关节角跟踪关节角指令，即

$$\boldsymbol{q}^{i}\to\boldsymbol{q}_{\mathrm{r}}^{i}\quad t\to\infty\tag{28}$$

假设平台控制不动，即$\boldsymbol{F}_{\mathrm{b}}=\boldsymbol{0}_{6\times1}$。进行控制律设计时，采用简化的具有固定基座的机械臂动力学模型，即对双臂空间机器人的动力学方程(3)进行简化，可得

$$\boldsymbol{H}_{\mathrm{m}}\ddot{\boldsymbol{q}}+\boldsymbol{c}_{\mathrm{m}}=\boldsymbol{\tau}_{\mathrm{m}}\tag{29}$$

式中：$\boldsymbol{H}_{\mathrm{m}}$为2部机械臂的广义质量阵；$\boldsymbol{c}_{\mathrm{m}}$为2部机械臂对应的非线性广义惯性力；$\boldsymbol{\tau}_{\mathrm{m}}$为2部机械臂的关节控制力矩，其表达式分别为：$\boldsymbol{H}_{\mathrm{m}}=\begin{bmatrix}\boldsymbol{H}_{\mathrm{m}}^{1} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{H}_{\mathrm{m}}^{2}\end{bmatrix}$，$\boldsymbol{c}_{\mathrm{m}}=\begin{bmatrix}\boldsymbol{c}_{\mathrm{m}}^{1}\\ \boldsymbol{c}_{\mathrm{m}}^{2}\end{bmatrix}$，$\boldsymbol{\tau}_{\mathrm{m}}=\begin{bmatrix}\boldsymbol{\tau}_{\mathrm{m}}^{1}\\ \boldsymbol{\tau}_{\mathrm{m}}^{2}\end{bmatrix}$。

采用逆动力学PD控制策略，设计控制律为

$$\boldsymbol{\tau}_{\mathrm{m}}^{i}=\boldsymbol{H}_{\mathrm{m}}^{i}\boldsymbol{K}_{\mathrm{Dm}}^{i}\left[\boldsymbol{K}_{\mathrm{Pm}}^{i}\left(\boldsymbol{q}_{\mathrm{r}}^{i}-\boldsymbol{q}^{i}\right)-\dot{\boldsymbol{q}}^{i}\right]+\boldsymbol{c}_{\mathrm{m}}^{i}\qquad i=1,2\tag{30}$$

式中：$\boldsymbol{K}_{\mathrm{Pm}}^{i}$和$\boldsymbol{K}_{\mathrm{Dm}}^{i}$为待选取的正定对称阵。

根据式(24)、式(26)和式(28)，有

$$\dot{\tilde{\boldsymbol{x}}}^{i}+\boldsymbol{K}_{Ad}^{i}\tilde{\boldsymbol{x}}^{i}=\boldsymbol{0}\tag{31}$$

根据式(31)可知，通过恰当选取参数，系统跟踪误差将会收敛。

综上所述，双臂空间机器人球形外包络抓捕控制的闭环控制框图如图6所示。

4 数值仿真

利用数值仿真验证所设计的球形外包络与控制器的有效性。仿真中，选取自由漂浮基双臂空间机器人系统为研究对象，其中，系统平台为边长2 m的正方体，质量为3 000 kg，3轴的转动惯量均为2 000 kg·m^2，不考虑质心偏差。2部机械臂均安装有7个臂杆，各臂杆均可视为圆柱形均质实心杆，其质量特性参数如表1所示。2部机械臂安装在平台的+z平面上，沿$\mathcal{F}_{\mathrm{b}}$的x轴对称，其根部安装位置分别为$[-1,0.85,1]^{\mathrm{T}}$ m和$[-1,-0.85,1]^{\mathrm{T}}$ m。目标星的几何参数、质量特性参数与双臂空间机器人系统的平台一致，目标上的2个抓捕点在$\mathcal{F}_{\mathrm{t}}$中的位置设定为$[0,0.6,-1]^{\mathrm{T}}$ m和$[0,-0.6,-1]^{\mathrm{T}}$ m。

表1 机械臂质量特性参数

Table 1 Mass characteristic parameters of manipulator

编号	质量/kg	静矩/(kg·m)	惯量/(kg·m^2)
1	5	$[0,0,0]^{\mathrm{T}}$	diag[0.02,0.02,0.01]
2	37.5	$[0,0,0]^{\mathrm{T}}$	diag[28.15,0.06,28.15]
3	5	$[0,0,0]^{\mathrm{T}}$	diag[0.07,0.01,0.07]
4	37.5	$[0,0,0]^{\mathrm{T}}$	diag[0.06,28.15,28.15]
5	5	$[0,0,0]^{\mathrm{T}}$	diag[0.02,0.02,0.01]
6	5	$[0,0,0]^{\mathrm{T}}$	diag[0.07,0.01,0.07]
7	5	$[0,0,0]^{\mathrm{T}}$	diag[0.07,0.07,0.01]

仿真的初始条件与控制参数如表2所示。仿真时长设定为30 s。

表2 控制参数与初始条件

Table 2 Control parameters and initial conditions

参数	数值
$\boldsymbol{K}_{Ad}^{i}(i=1,2)$	diag[2,1,3,1,2,2]
$\dot{\boldsymbol{q}}_{\max}^{i}(i=1,2)$	$[10,10,10,10,10,10,10]^{\mathrm{T}}$
$\boldsymbol{K}_{\mathrm{Pm}}^{i}(i=1,2)$	$1.2\boldsymbol{E}_{7\times7}$
$\boldsymbol{K}_{\mathrm{Dm}}^{i}(i=1,2)$	$4.5\boldsymbol{E}_{7\times7}$
μ	0.1
k	500
c	100
${}^{\mathrm{e}}\boldsymbol{p}^{\mathrm{b}}$/m	$[0,0,-5]^{\mathrm{T}}$
${}^{\mathrm{e}}\boldsymbol{\Phi}^{\mathrm{b}}$/(°)	$[0,0,0]^{\mathrm{T}}$
${}^{\mathrm{e}}\boldsymbol{p}^{i}$/m	$[0,0,0]^{\mathrm{T}}$
${}^{\mathrm{e}}\boldsymbol{\Phi}^{i}$/(°)	$[0,0,0]^{\mathrm{T}}$
$\boldsymbol{v}_{\mathrm{t}}$/(m·s^{-1})	$[0,0,0.01]^{\mathrm{T}}$
$\boldsymbol{\omega}_{\mathrm{t}}$/((°)·s^{-1})	$[0,1,0]^{\mathrm{T}}$
$\boldsymbol{q}^{1}$/(°)	$[0,75,0,40,0,-10,0]^{\mathrm{T}}$
$\boldsymbol{q}^{2}$/(°)	$[0,65,0,60,0,-20,0]^{\mathrm{T}}$

为更直观地描述带有球形外包络的双臂空间机器人抓捕目标的运动过程，在 MATLAB 仿真中根据各体的实时位置与姿态进行三维绘图，得到系统的三维仿真模型，如图 7 所示。值得说明的是，在系统三维运动绘图中，系统平台两侧有 2 部太阳帆板，其目的是为了能够更加清晰地观察双臂空间机器人平台的运动状态，在数值仿真过程中并不考虑太阳帆板对系统运动的影响。

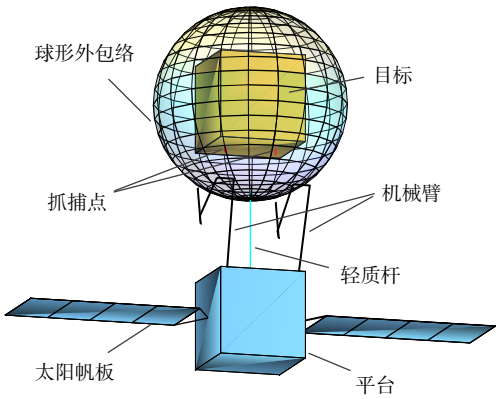


图 7　系统三维仿真模型

Fig. 7　3D simulation model of system

选取判断目标与球形外包络碰撞的指标函数为：$\eta = r_{\text{env}} - \max\limits_{i=1,\cdots,8} \|^{\text{env}}\boldsymbol{p}^{p_i}\|$。

指标函数的变化曲线如图 8 所示，阴影区域（即$\eta < 0$）代表碰撞区，可知，目标与球形外包络先后 3 次发生了碰撞。

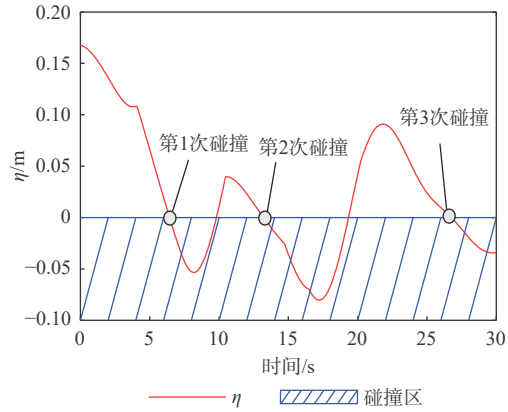


图 8　碰撞指标函数

Fig. 8　Collision index function

在碰撞的瞬间，目标与球形外包络之间会产生接触力和接触力矩。其中，球形外包络施加在目标上的接触力与接触力矩如图 9 所示，其最大值在工程允许范围内。3 次碰撞的时刻分别为 6.4 s、13.2 s 和 26.8 s。在第 2 次碰撞过程中，由于目标上有多个顶点先后与球形外包络发生了不同程度的碰撞，导致该时间段内目标所受接触力矩的方向发生了改变。

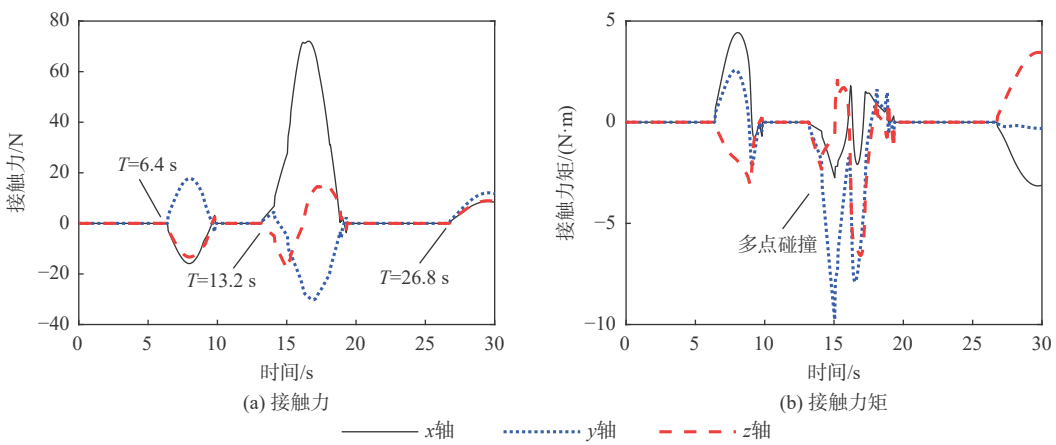


图 9　目标所受球形外包络接触力与接触力矩

Fig. 9　Contact force and contact moment of object subjected to cage

选取 6 个特殊时刻绘制系统的三维运动状态图，如图 10 所示。通过对比不同时刻的系统状态图可以看出，目标在 3 次触碰到球形外包络内壁后又产生回弹。整个仿真过程中，目标一直处于球形外包络内部，未发生逃逸情况。此外，双臂空间机器人上的 2 部机械臂在初始时刻开始跟踪目标，在最终时刻能够抓捕到目标上的 2 个抓捕点。

仿真中系统的最终状态如图 11 所示。根据放大的细节图可知，2 部机械臂的末端作用器可准确到达目标上的 2 个抓捕点，实现对目标的捕获。

为更严谨地得到 2 部机械臂的抓捕精度，还需要获取 2 部机械臂的抓捕位置与姿态误差，如图 12～图 15 所示。通过分析可知，2 部机械臂抓捕的位姿误差均可在 30 s 内收敛到零点附近。2 部机械臂的位置误差精度均为 0.01 m，姿态误

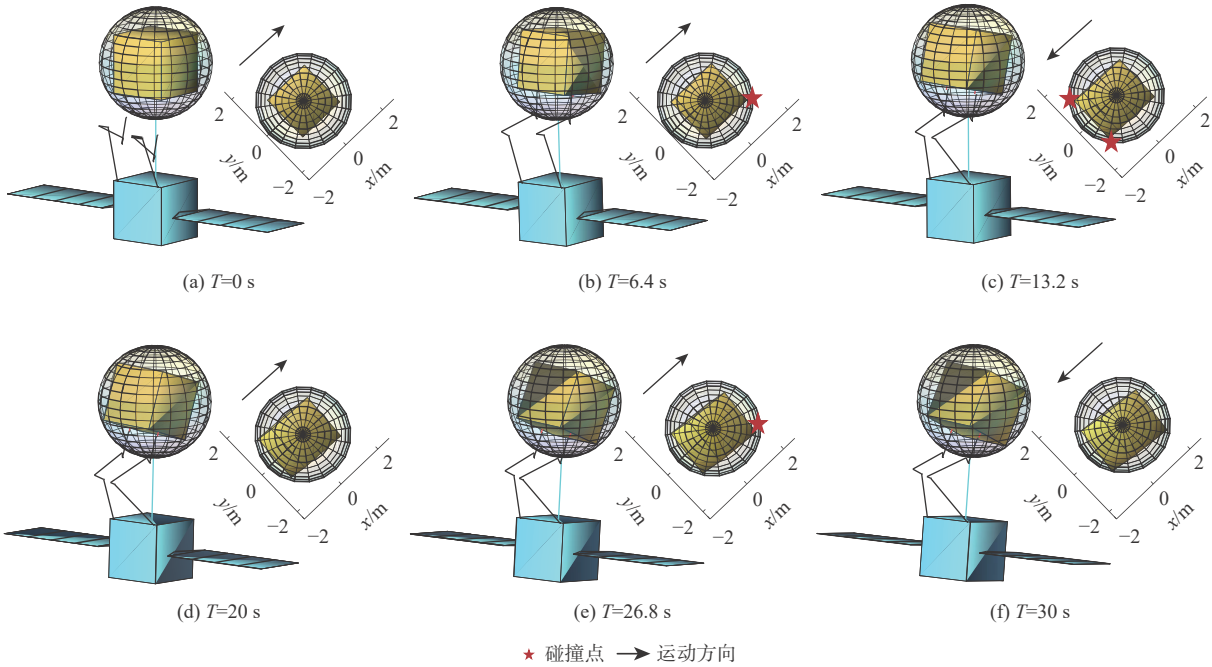


图 10 系统运动状态图

Fig. 10 System motion state

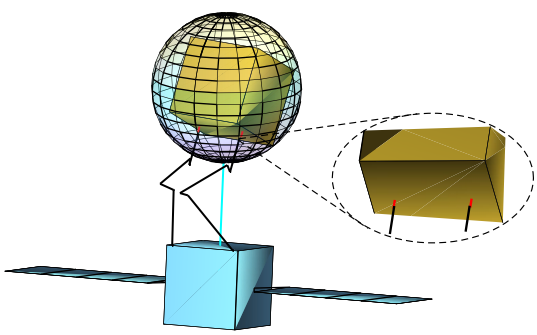


图 11 系统抓捕最终状态图

Fig. 11 Final state of system capture

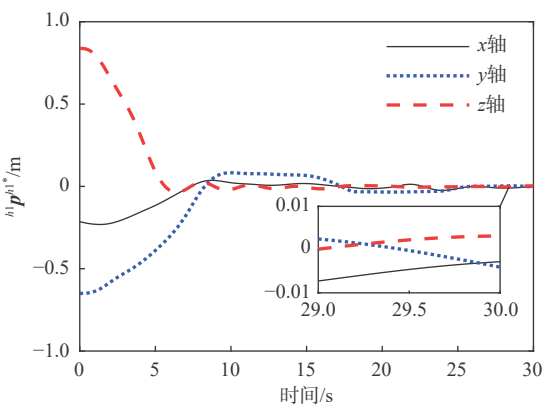


图 12 机械臂 1 抓捕位置误差

Fig. 12 Position error for manipulator 1 during capturing

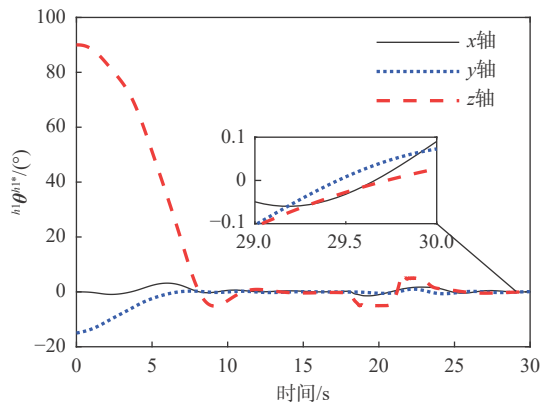


图 13 机械臂 1 抓捕姿态误差

Fig. 13 Attitude error for manipulator 1 during capturing

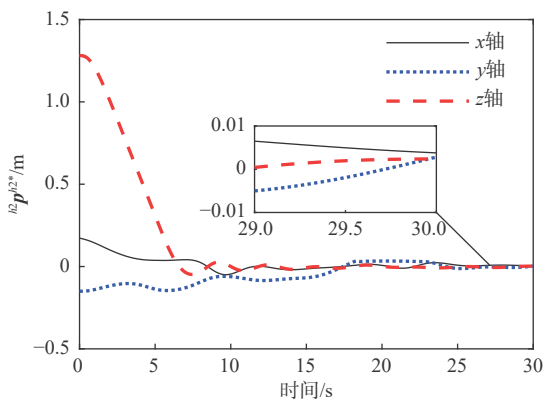


图 14 机械臂 2 抓捕位置误差

Fig. 14 Position error for manipulator 2 during capturing

差精度分别为 0.1°和 0.3°，满足一定的工程应用需求。

在抓捕过程中，2 部机械臂的关节角随时间变化的曲线如图 16 所示。可知，2 部机械臂的关节角均在工程许可范围内，未出现机械臂奇异问题。

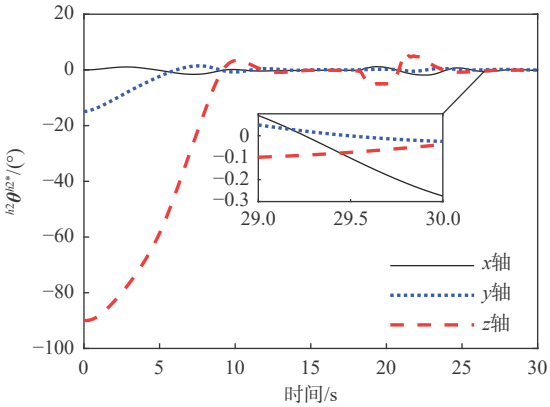
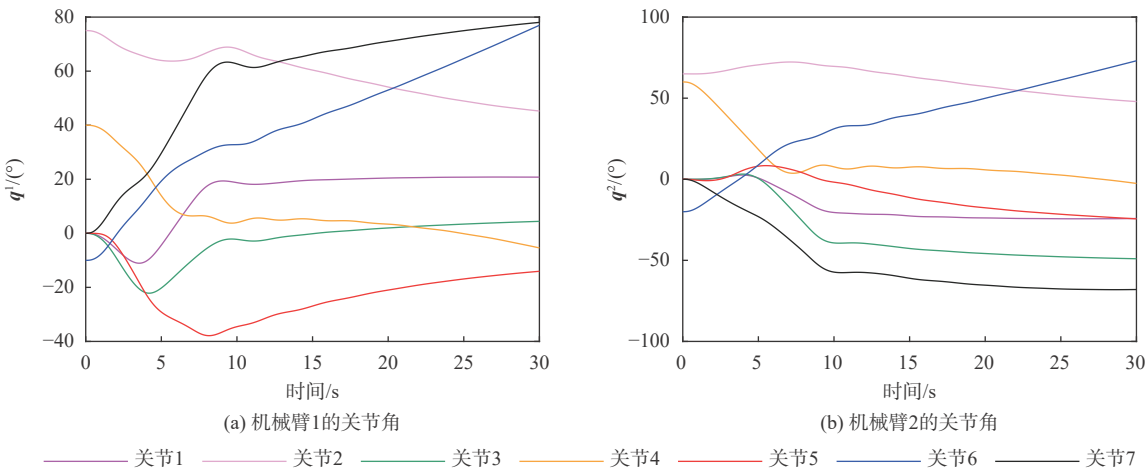


图 15　机械臂 2 抓捕姿态误差

Fig. 15　Attitude error for manipulator 2 during capturing

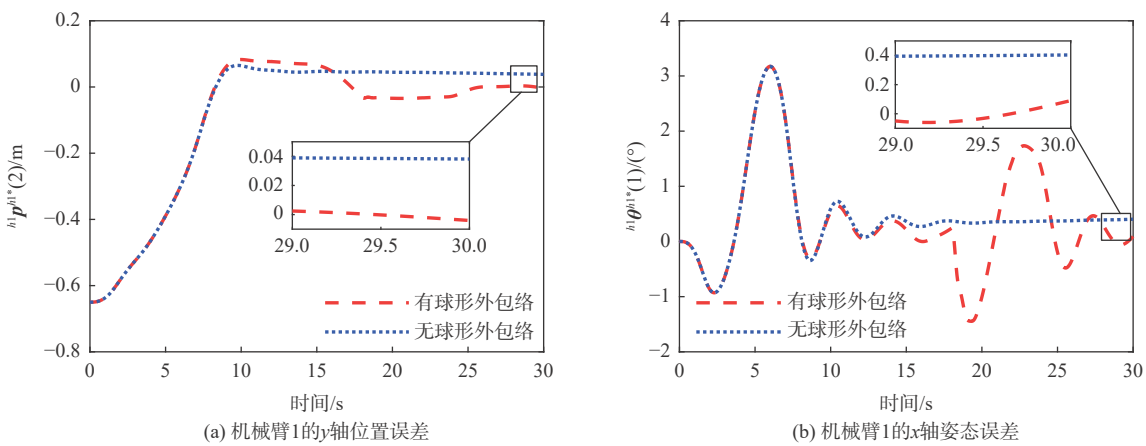
为进一步验证球形外包络在双臂空间机器人抓捕运动目标时的优势，利用常规的双臂空间机器人进行对比仿真实验。在对比仿真实验中，系统不含球形外包络结构，其模型参数、控制参数及初始条件均与双臂球形外包络抓捕仿真实验相同。图 17 为有无球形外包络时机械臂 1 的部分抓捕位姿误差对比曲线。可知，有球形外包络时，其抓捕位姿误差更小。2 部机械臂在 2 组对比实验中的抓捕位姿精度如表 3 所示。可知，球形外包络能大幅提高 2 部机械臂抓捕运动目标的位置精度与姿态精度，充分验证了球形外包络结构的优势与有效性。



(a) 机械臂1的关节角　　(b) 机械臂2的关节角

图 16　2 部机械臂的关节角

Fig. 16　Joint angles of two manipulators



(a) 机械臂1的y轴位置误差　　(b) 机械臂1的x轴姿态误差

图 17　有无球形外包络时抓捕位姿误差对比

Fig. 17　Comparison of position and attitude errors during capturing with or without spherical cage

表 3　有无球形包络时 2 部机械臂抓捕精度对比

Table 3　Comparison of capturing accuracy of dual arms with or without spherical cage

模型	抓捕位置精度/m		抓捕姿态精度/(°)	
	机械臂1	机械臂2	机械臂1	机械臂2
有球形外包络	0.01	0.01	0.1	0.3
无球形外包络	0.05	0.05	0.5	0.5

5　结　论

在主动碎片清除技术中，利用空间机器人直接接触运动目标，可能会出现因相对运动过大而损坏接触表面或因推离目标而导致抓捕失败的问题。针对上述问题，本文提出了一种双臂球形外包络的抓捕模型，并结合双臂空间机器人与球形外包络的

优势，设计了一种基于碰撞检测算法的双臂在线协调抓捕控制方法。通过理论分析与仿真实验可以得出如下结论：

1）所设计的球形外包络可有效约束目标运动，避免了目标的逃逸，提高了抓捕成功率。仿真结果验证了球形外包络结构设计的有效性。

2）通过建立基于碰撞检测算法的摩擦接触力模型，利用球形外包络与目标之间的摩擦，可降低两者之间的相对运动，提高了机械臂抓捕的安全性。

3）实时规划了双臂的期望抓捕轨迹，并设计完成了双臂协调抓捕控制律。仿真结果表明，双臂可准确实现对目标的抓捕，且抓捕位姿精度满足工程应用需求。

参考文献（References）

[1] SVOTINA V V, CHERKASOVA M V. Space debris removal–review of technologies and techniques. Flexible or virtual connection between space debris and service spacecraft[J]. Acta Astronautica, 2023, 204: 840-853.

[2] ZHAO P Y, LIU J G, WU C C. Survey on research and development of on-orbit active debris removal methods[J]. Science China Technological Sciences, 2020, 63(11): 2188-2210.

[3] LEDKOV A S, ASLANOV V S. Active space debris removal by ion multi-beam shepherd spacecraft[J]. Acta Astronautica, 2023, 205: 247-257.

[4] SHAN M H, GUO J, GILL E. Review and comparison of active space debris capturing and removal methods[J]. Progress in Aerospace Sciences, 2016, 80: 18-32.

[5] PENG J Q, XU W F, WANG Z Y, et al. Dynamic analysis of the compounded system formed by dual-arm space robot and the captured target[C]//Proceedings of the 2013 IEEE International Conference on Robotics and Biomimetics. Piscataway: IEEE Press, 2013: 1532-1537.

[6] ZHANG W, LI F, LI J L, et al. Review of on-orbit robotic arm active debris capture removal methods[J]. Aerospace, 2023, 10(1): 13.

[7] RAINA D, GORA S, MAHESHWARI D, et al. Impact modeling and reactionless control for post-capturing and maneuvering of orbiting objects using a multi-arm space robot[J]. Acta Astronautica, 2021, 182: 21-36.

[8] XU R N, LUO J J, WANG M M. Optimal grasping pose for dual-arm space robot cooperative manipulation based on global manipulability[J]. Acta Astronautica, 2021, 183: 300-309.

[9] 李梁, 李剑飞, 张大伟, 等. 自由漂浮空间双臂机器人动目标抓捕控制[J]. 北京理工大学学报, 2019, 39(6): 615-623.
LI L, LI J F, ZHANG D W, et al. Control of free-floating space dual-arm robot for capturing moving targets[J]. Transactions of Beijing Institute of Technology, 2019, 39(6): 615-623(in Chinese).

[10] NISHIDA S I, KAWAMOTO S. Strategy for capturing of a tumbling space debris[J]. Acta Astronautica, 2011, 68(1-2): 113-120.

[11] CHEN G, WANG Y Q, WANG Y F, et al. Detumbling strategy based on friction control of dual-arm space robot for capturing tumbling target[J]. Chinese Journal of Aeronautics, 2020, 33(3): 1093-1106.

[12] RIMON E. Caging planar bodies by one-parameter two-fingered gripping systems[J]. The International Journal of Robotics Research, 1999, 18(3): 299-318.

[13] EGAWA T, MAEDA Y, TSURUGA H. Two- and three-dimensional caging-based grasping of objects of various shapes with circular robots and multi-fingered hands[C]//Proceedings of the 41st Annual Conference of the IEEE Industrial Electronics Society. Piscataway: IEEE Press, 2015: 643-648.

[14] 孙冲, 袁建平, 万文娅, 等. 自由翻滚故障卫星外包络抓捕及抓捕路径优化[J]. 航空学报, 2018, 39(11): 322203.
SUN C, YUAN J P, WAN W Y, et al. Outside envelop grasping method and approaching trajectory optimization for tumbling malfunctional satellite capture[J]. Acta Aeronautica et Astronautica Sinica, 2018, 39(11): 322203(in Chinese).

[15] MAEDA Y, KODERA N, EGAWA T. Caging-based grasping by a robot hand with rigid and soft parts[C]//Proceedings of the 2012 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2012: 5150-5155.

[16] WANG M M, LUO J J, YUAN J P, et al. Coordinated trajectory planning of dual-arm space robot using constrained particle swarm optimization[J]. Acta Astronautica, 2018, 146: 259-272.

[17] ZEBENAY M. Development of a robotics-based satellites docking simulator[D]. Espoo: Helsinki University of Technology, 2014: 7-11.

Spherical envelope capture control for dual-arm space robots

XIA Xinhui[1]，JIA Yinghong[1, *]，ZHANG Jun[2, 3]

(1. School of Astronautics，Beihang University，Beijing 100191，China；
2. Beijing Institute of Control Engineering，Beijing 100190，China；
3. National Laboratory of Space Intelligent Control，Beijing 100190，China)

Abstract: In view of the object capture and control problem of a dual-arm space robot, a new spherical cage-based scheme for dual arms to capture moving objects was designed. Firstly, in order to avoid pushing away from the object when captured by the manipulator, a structure with a spherical cage of dual arms was designed to constrain the moving object geometrically and prevent the object escape. Secondly, the collision detection model and the accurate model of the cage contact and collision model were proposed according to the contact and collision conditions between the cage and the object. The friction between the spherical cage and the object was used to reduce the speed of the object and provide a safe capturing environment for the manipulator. Furthermore, a visual servo method was adopted to design the real-time expected trajectory of dual arms to capture moving objects, and a dual-arm coordination tracking control law based on inverse dynamics was developed. Finally, the closed-loop control system simulation experiments for capturing a moving cubic object by using the free-floating dual-arm space robot based on the spherical cage were carried out. The simulation results demonstrate the capturing effectiveness of the proposed control method based on a spherical cage.

Keywords: dual-arm space robot; spherical cage; object capture; trajectory planning; collision detection; contact force modeling

Received：2023-05-18；**Accepted**：2023-06-12；**Published Online**：2023-06-30 16：42

URL：link.cnki.net/urlid/11.2625.V.20230630.1608.005

Foundation item：National Key Research and Development Program of China (2018AAA0103003)

* **Corresponding author**. E-mail：jia_yingh@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0260

贮箱自动无级增压可调文氏管的设计与试验

任嘉琳[1,2]，郭红杰[1,2,*]，李文鑫[1]，梁国柱[1,2]，赵胜[3]

(1. 北京航空航天大学 宇航学院，北京 102206；2. 北京航空航天大学宁波创新研究院，宁波 315800；
3. 中国航天科技集团有限公司第一研究院，北京 100076)

摘 要：在运载火箭贮箱增压系统中，贮箱压力的稳定控制是至关重要的。基于此，提出了一种基于特型型面调节锥的可调文氏管设计方法，使得阀芯位移与质量流量之间保持良好的线性对应关系，提高调节精度；制定了基于负反馈控制原理的闭式自动增压试验方案及基于可编程逻辑控制器的系统控制与数据采集方案，成功进行了自动增压试验。经试验验证，试验装置在较宽的质量流量范围内能够实现高精度、连续稳定的压力调节。相关工作可为增压系统流量调节装置的设计与改进提供参考。

关 键 词：可调文氏管；型面设计；贮箱增压；流量调节；压力调节
中图分类号：V434
文献标志码：A **文章编号**：1001-5965(2025)05-1684-10

运载火箭贮箱增压系统的主要作用是提供必要的气枕压力，以满足液体火箭发动机泵入口的最低压力要求，同时保持薄壁贮箱结构的强度和刚度要求。对于确定合理的增压压力以提高系统可靠性与运载能力而言，贮箱压力的有效控制具有重要意义。

贮箱增压有2种方案：开式增压和闭式增压。开式增压方案中，增压气体流量是恒定不可调节的，而闭式增压方案需要对增压气体进行人为控制[1-3]，控制模式又可分为定压力和定流量2类，或是流量调节和通断控制2类[2]。为实现这些控制模式，常用的控制元件有孔板、活门、节流圈、电爆阀、电磁阀、减压器、压调器等[4-5]。在航天领域中，根据具体要求使用不同的控制方案，如长征3号火箭三级采用电磁阀-压调器控制增压，长征4号火箭三级、天顶号火箭一二级采用电磁阀-压力信号器控制增压，战神1、土星5、阿里安5、阿特拉斯5等火箭采用电磁阀-压力传感器控制增压等[2,4]。目前，常用的控制模式还存在可改进的方面，如增压过程中采用的有级增压方式，其稳定性和连续性都不够理想。此外，为了实现控制，需要使用的阀门种类与数量较多，使得整个系统的组成变得相对复杂，整体尺寸和质量较大。

目前，研发可重复使用的运载器已成为国际航天技术发展的重要方向之一，仅依靠传统增压系统很难实现大范围流量调节和可重复利用的需求[6]。为了使得增压过程连续、稳定、精度高、易于控制，减轻系统质量并简化系统结构，采用能实现无级增压的新的流量调节装置是重要的技术途径。

在航空航天领域中，文氏管广泛应用于流体调控，其结构简单坚固，精度高，重复性好，表现出较为优越的工作性能。可调汽蚀文氏管在变推力液体火箭发动机、三组元火箭发动机、超燃冲压发动机中可精确控制推进剂组元比，提高推力精度，实现飞行器的稳定发射[7]。在航空发动机的涡流器中，可借助文氏管将气流限制在中心线附近，通过控制内外旋流的混合程度，使燃油集中在回流区域，从而使得油雾分布的均匀性得到提高，雾化质

收稿日期：2023-05-22；**录用日期**：2023-07-20；**网络出版时间**：2023-08-04 13:27
网络出版地址：link.cnki.net/urlid/11.2625.V.20230803.1830.003
*通信作者. E-mail：guohongjie@buaa.edu.cn

引用格式：任嘉琳，郭红杰，李文鑫，等. 贮箱自动无级增压可调文氏管的设计与试验 [J]. 北京航空航天大学学报，2025，51（5）：1684-1693. REN J L，GUO H J，LI W X，et al. Design and experiment of adjustable Venturi tube for stepless automatic tank pressurization [J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1684-1693（in Chinese）.

量也得到显著提升[8]。文氏管还可以应用于航空发动机旋流杯中, 缩小液滴的范围, 使燃油聚集在回流区中, 提高液滴分布的均匀性和雾化的质量, 还可以有效抑制积碳的产生[9]。

本文对用于贮箱线性压力调节的可调文氏管的设计方案进行研究, 与传统的锥形型面可调文氏管相比, 该方案能够实现更高的控制精度; 与采用电磁阀组合进行调压的方案相比, 该方案不仅体积更小, 质量更轻, 结构更为简单, 还能够实现连续无级增压的需求。本文建立了基于负反馈控制原理的闭式自动增压试验系统, 得到相应的控制方案, 进行贮箱自动增压试验, 以获得稳压控制的策略和方法。

1　可调文氏管工作原理及设计计算

1.1　可调文氏管的工作原理

可调文氏管是一种内部含有可同轴移动阀芯的流量控制装置, 阀芯通常为圆柱体或圆锥体, 质量流量大小与入口压力和喉部面积成正比, 喉部面积的调节通过阀芯的移动来实现。流动壅塞时, 出口压力的变化对通过可调文氏管的质量流量大小没有影响, 因此, 即使下游压力发生波动, 可调文氏管也能稳定地保持质量流量不变[10-11], 本文正是运用这一特性来实现贮箱增压。

1.2　设计计算方法

已知来流总压为 p_i^*, 总温为 T_i^*, 入口面积为 A_i, 出口静压为 p_e, 最大质量流量为 Q, 气体比热容比为 k。假设气体为完全气体, 则文氏管内的气体流动为一维定常变截面等熵流动, 由气动函数 $\pi(Ma)$和$q(\lambda)$可得喉部面积A_t和出口面积A_e为

$$A_t=\frac{Q\sqrt{T_i^*}}{p_i^*\sqrt{\dfrac{k}{R}\left(\dfrac{2}{k+1}\right)^{\frac{k+1}{k-1}}}} \tag{1}$$

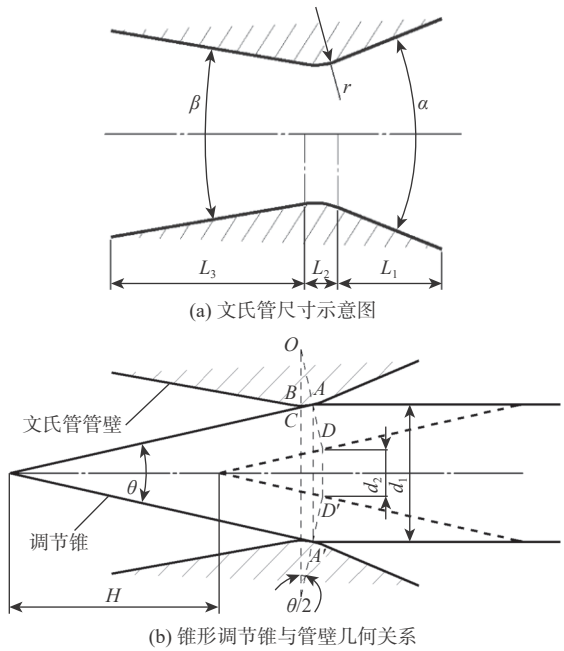
$$\frac{p_e}{p_i^*}=\left(1+\frac{k-1}{2}Ma_e^2\right)\frac{k}{k-1}=\pi(Ma_e) \tag{2}$$

$$A_e=\frac{A_t}{q(\lambda_e)} \tag{3}$$

式中: R 为气体常数; Ma 为马赫数; λ为速度因数。

根据所需工艺尺寸, 需要选择合适的文氏管, 收敛角为 α, 扩张角为 β, 喉部圆角为 r, 调节锥锥角为 θ, 文氏管收敛段、喉部圆弧段、扩张段的长度 L_1、L_2、L_3 可由几何关系计算得出。可调文氏管原理示意图如图 1 所示[12-14]。

图中 A 点为阀门关闭时阀芯与管壁相接点。圆台侧表面母线 AD 长度为



(a) 文氏管尺寸示意图

(b) 锥形调节锥与管壁几何关系

图 1　可调文氏管原理示意图[12-14]

Fig. 1　Principle of adjustable Venturi tube[12-14]

$$\overline{AD}=l=H\sin(\theta/2) \tag{4}$$

式中: H 为调节锥行程。

可调文氏管的最小流通面积即为圆台 $ADD'A'$ 的环状侧表面积 S, 有

$$S=\pi l\frac{d_1+d_2}{2} \tag{5}$$

式中: d_1 为圆台上底直径 AA'; d_2 为圆台下底直径 DD'。

由喉部面积求得喉部直径 d_t, 设

$$b=\pi\sin\left(\frac{\theta}{2}\right)\left[d_t+2r_t\left(1-\cos\left(\frac{\theta}{2}\right)\right)\right] \tag{6}$$

$$a=\pi\sin^2\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right) \tag{7}$$

则

$$S=bH-aH^2 \tag{8}$$

与行程H相对应的质量流量q为

$$q=K\frac{p_i^*}{\sqrt{T_i^*}}S \tag{9}$$

在进入文氏管后, 调节锥先穿过一段非工作区域, 直至最小流通面积与喉部面积相等的位置, 才开始发挥流量调节作用, 进入工作状态。当 q=Q 时, 对应的 H 即为调节锥的总行程。

由锥形型面推导特型型面的几何示意图如图 2 所示。图中, 坐标系的原点为从阀门关闭时阀芯与管壁相接点到轴线作垂线的垂足, 线段 ef 为锥形调节锥节流面母线, n_i 为锥形型面的几何点, H_{n_i} 为对应行程点, m_i 为特型型面的几何点, H_{m_i} 为对应行程点。

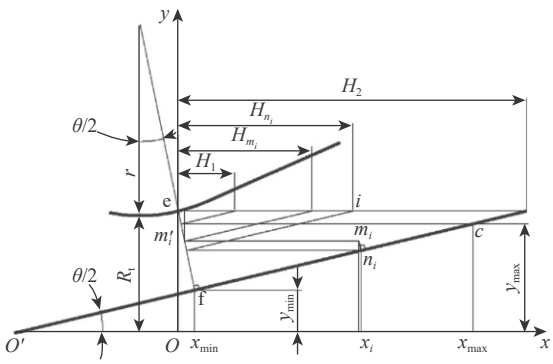


图 2 几何推导示意图

Fig. 2 Geometric derivation

推导过程中，保持节流面母线位置不变，将实际总行程 H 划分为 i 段，根据式 (6)～式 (9)，结合所需质量流量 q 与行程 H 之间的对应关系及型面间的几何关系，将锥形调节锥的每个行程点 H_{n_i} 转化为特型调节锥对应行程点 H_{m_i}，并将 H、ΔH 与 H_{m_i} 代入式 (10) 和式 (11) 进行计算，即可求得特型型面各点 (x_i, y_i) 的坐标：

$$x_i = H - i\Delta H + H_{m_i}\sin^2\left(\frac{\theta}{2}\right) \tag{10}$$

$$y_i = R_t + r_t\left(1-\cos\left(\frac{\theta}{2}\right)\right) - H_{m_i}\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right) \tag{11}$$

简化整理，可得 x_i 与 y_i 的解析关系式形式为

$$x_i = G_1 y_i^2 + G_2 y_i + G_3 \tag{12}$$

1.3 设计案例

根据表 1 所示的初始条件，计算得到文氏管的尺寸如表 2 所示，其中，d_i 为入口直径，d_e 为出口直径，C_p 为定压比热容。每 ΔH=0.001 mm 计算质量流量 q，所得理论流量曲线及调节锥形面对比如图 3 所示。图中，q_z 为锥形型面理论质量流量，q_t 为特型型面理论质量流量，S 为最小流通面积。特型型面理论质量流量与行程应实现的对应关系为

$$q_t = 26.035H \tag{13}$$

表 1 初始条件

Table 1 Initial conditions

参数	数值
Q/(g·s^{-1})	200
k	1.4
p_i*/ MPa	5
T_i*/ K	293
p_e/ MPa	0.35
d_i/ mm	8
C_p/(J·g^{-1}·K^{-1})	1 04
K	0.04

表 2 文氏管设计尺寸

Table 2 Designed size of Venturi tube

参数	数值
d_t/mm	4.822
d_e/mm	7.185
α/(°)	45
β/(°)	20
θ/(°)	25
r/mm	2
L_1/mm	3.469
L_2/mm	1.113
L_3/mm	6.530

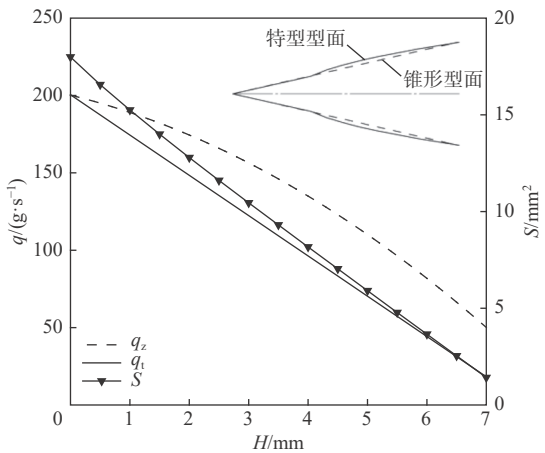


图 3 理论流量曲线及调节锥型面对比

Fig. 3 Comparison between theoretical flow curves and conical pintle profile

根据 1.2 节所述推导方法得到特型型面调节锥，采用步进电机驱动调节锥[15]，所设计的原理试验缩比试验装置如图 4 所示。

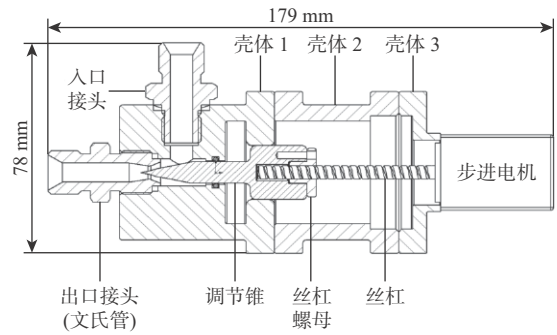


图 4 可调文氏管试验装置结构示意图

Fig. 4 Structure of adjustable Venturi tube test device

2 仿真计算结果与分析

为验证试验装置线性特性，使用 Fluent 对试验装置进行数值仿真，研究喉部面积对试验装置质量流量特性和调节锥所受轴向力大小的影响。为了简化计算，将结构化为二维轴对称形式，以全开位

置为原点, 在上游压力为 1, 2, 3, 4, 5 MPa 的条件下, 对调节锥行程点在 0～7 mm 范围内的试验装置进行计算。采用SST k-ω湍流模型和密度基定常求解器, 壁面为无滑移固体边界, 气体介质设为理想气体, 通过入口条件进行全场初始化, 压力入口设为 1, 2, 3, 4, 5 MPa, 压力出口为大气压。

使用约 4 万和 7 万 2 种不同的网格进行无关性验证, 在距轴线 1 mm 和 2 mm 位置处分别画一条平行于轴线的贯通直线, 计算不同网格数下直线上的马赫数, 结果如图 5 所示。图中: 图例下标中 4 表示模型网格数约 4 万, 7 表示模型网格数约 7 万, 1 表示直线距轴线 1 mm, 2 表示直线距轴线 2 mm。可以看出, 2 组网格计算结果非常相近, 因此, 后续选择约 4 万的网格数进行计算。

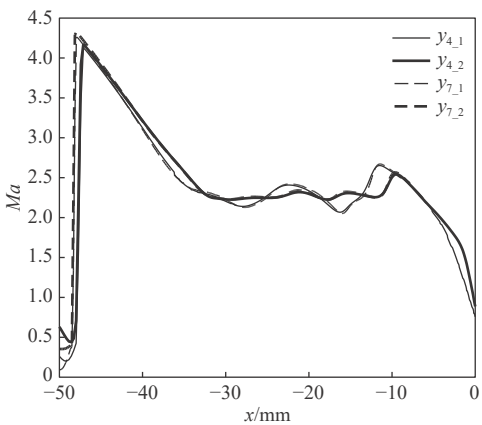


图 5　不同网格数下平行直线上马赫数分布

Fig. 5　Mach number distribution on parallel lines under different grid numbers

试验装置的质量流量特性曲线如图 6 所示。图中: 图例中下标 s 表示仿真质量流量, 下标 t 表示理论质量流量, 下标 1～5 分别表示上游压力 1～5 MPa。对仿真质量流量q_s进行线性拟合, 得出拟合方程分别为

$$q_s = -5.029H + 38.494 \quad (14)$$

$$q_s = -10.064H + 77.100 \quad (15)$$

$$q_s = -15.101H + 115.720 \quad (16)$$

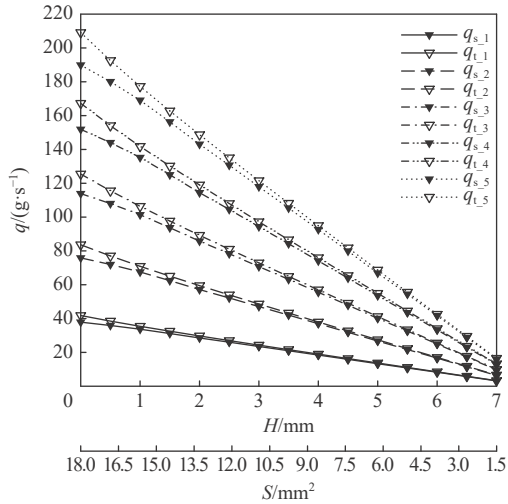
$$q_s = -20.137H + 154.340 \quad (17)$$

$$q_s = -25.174H + 193.970 \quad (18)$$

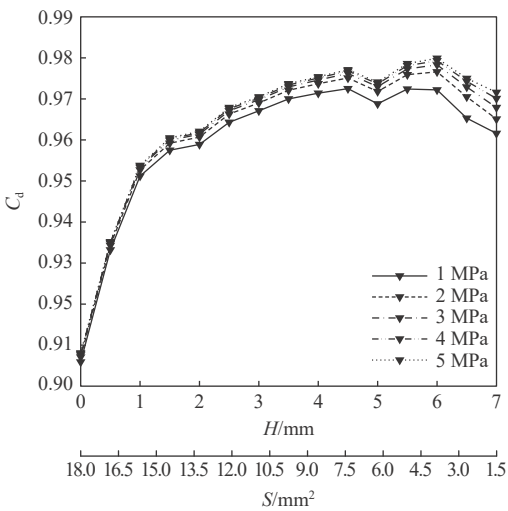
拟合效果参数确定系数 R^2 都达到 0.999 7, 表明线性拟合效果很好, 质量流量特性的线性程度高。求解理论质量流量时, 假设调节锥位置不变, 移动文氏管改变二者的相对位置, 则型面曲线方程为

$$x = 1.436\,925y^2 - 0.221\,695y - 0.457\,086 \quad (19)$$

根据设计结果, 喉部圆弧曲线圆心坐标为



(a) 不同上游压力下质量流量随调节锥行程和喉部面积的变化



(b) 质量流量系数随调节锥行程和喉部面积的变化

图 6　可调文氏管质量流量特性曲线

Fig. 6　Flow characteristic curves of adjustable Venturi tube

(x_c, y_c), 每个行程点 H 所对应的喉部圆弧曲线为 $(x-x_c)^2+(y-y_c)^2=4$。以 2 条曲线之间的最小距离 L 所在的位置为母线进行回转, 其回转面积即为最小流通面积 S, 由式(9)可求得各点理论质量流量 q_t。质量流量系数 C_d 定义为

$$C_d = q_s / q_t \quad (20)$$

由图 6 可知, 在阀门关闭的过程中, 随着喉部截面积的减小, C_d 基本呈现先上升后下降趋势, 当调节锥行程位于 0.5～7 mm 之间时, C_d 位于 0.927 1～0.964 7 之间。当调节锥位于 H=0 mm 时, C_d 则明显低于其他位置, 无论在哪种工况下, 该位置的 C_d 都小于 0.9。图 7 为试验装置喉部马赫数与行程及喉部面积之间的对应关系, 在调节锥行程 H=0～7 mm 之间时, 喉部马赫数位于 0.603 7～0.856 1 的范围内。对比节流面和音速线(Ma=1)可知, H=0 mm 时, 节流面所在位置距临界截面最远, 从而导致 C_d 最小。图 8 为 H=0, 1, 2 mm 时文氏管

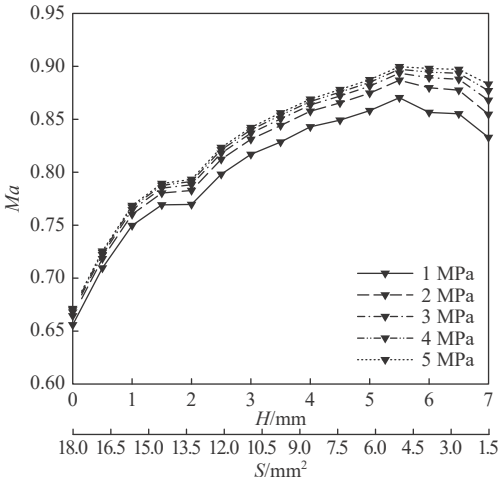


图 7 喉部马赫数随调节锥行程和喉部面积的变化

Fig. 7 Variation of Mach number on throat with conical pintle stroke and throat area

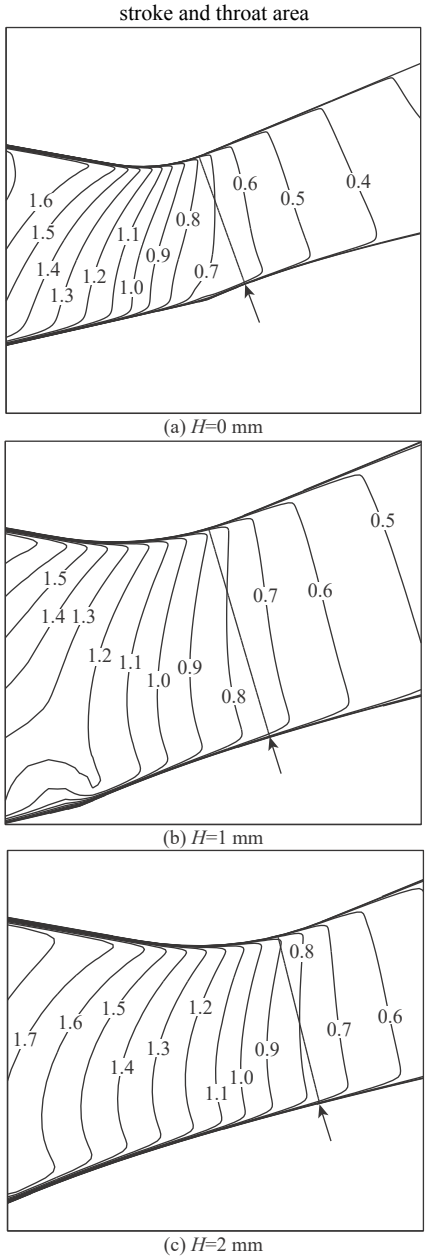


图 8 喉部附近马赫数分布

Fig. 8 Mach number distribution near throat region

节流面附近的马赫数分布，箭头所指位置为节流面所在位置。阀门关闭过程中，节流面始终位于临界截面后方。

轴对称调节锥所受横向力为零，所受轴向力 F 与喉部面积对应关系如图 9 所示。可以看出，调节锥所承受的轴向力随着喉部面积的减小而增大，H=3.5 mm 以后增长速度逐渐加快。调节锥在不同上游压力下所受的最大轴向力 F 及受力增长比例 P_r 如图 10 所示。

可知，在上游压力 1～5 MPa 范围内，调节锥所受轴向力增长比例随上游压力增大而增大，这一比例总体位于 76.84%～ 81.91% 范围内，最大轴向力为 69.18 N，本文选用的步进电机最大可承受轴向力为 100 N。

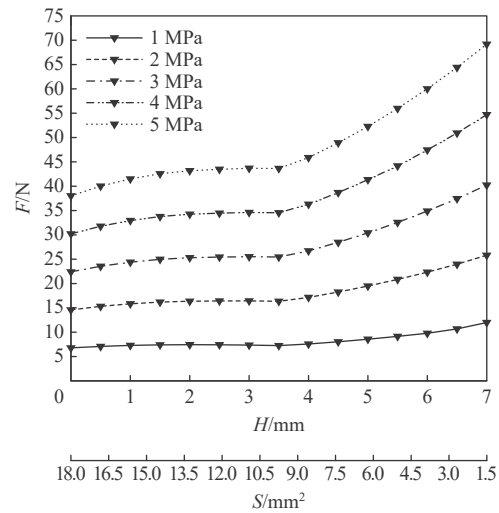


图 9 调节锥所受轴向力随行程和喉部面积的变化

Fig. 9 Variation of axial force of conical pintle with conical pintle stroke and throat area

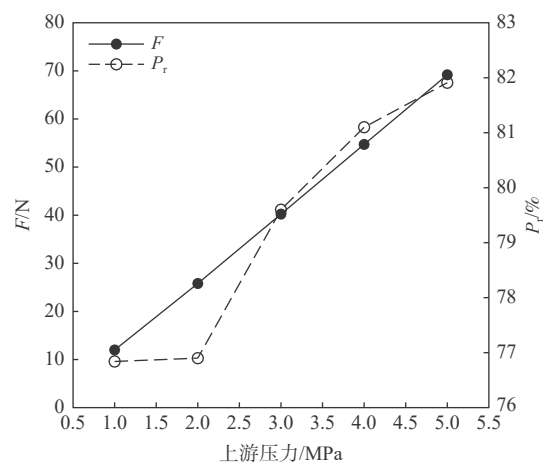


图 10 不同上游压力下调节锥所受最大轴向力及受力增长比例

Fig. 10 Maximum axial force and force growth ratio on conical pintle under different upstream pressures

3　贮箱自动增压试验系统

试验系统工作原理如图 11 所示。上游增压气体为高压氮气, 水作为推进剂的模拟介质, 圆柱体贮箱垂直于地面安装, 总容积为 31.428 L。使用压力表和压力变送器监测文氏管进出口和贮箱的压力变化, 其中, P_1、P_2、P_3 分别为箱压、文氏管出口压力和文氏管入口压力。在下游排液管路中安装脉冲输出涡轮流量计监测模拟推进剂的体积流量变化, 实测值的读取和记录通过上位机中的测控系统来完成。

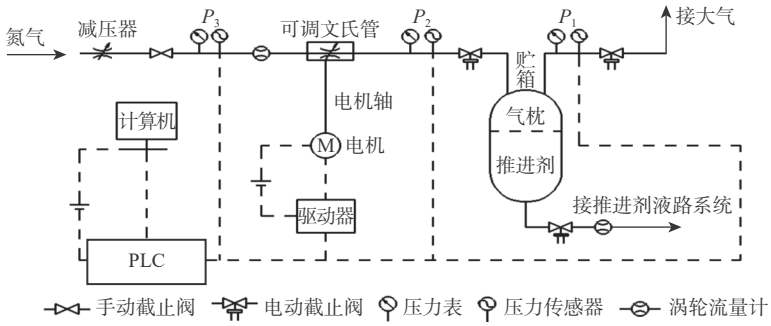


图 11　试验系统原理示意图

Fig. 11　Principle of test system

测控系统采用基于可编程逻辑控制器(programmable logic controller, PLC)的单闭环控制策略, PLC 读取并分析压力变送器传递的反馈信号, 随后输出指令到 UM242 驱动器, 通过向步进电机发送脉冲信号驱动电机轴转动, 安装在轴上的电机螺母将轴的转动转化为螺母的直线运动, 将调节锥固定在电机螺母上即可实现阀芯的往复直线运动。采用 LabVIEW 编写数据采集程序, LabVIEW 通过 OPC 服务器连接到 PLC, 以便在两者之间传输数据。

在 PLC 的方向判断程序中, 对反馈信号的分析体现为实时值和目标数值的比较, 从而决定是否需要改变电机轴的旋转方向。当实时数值低于目标数值时, 驱动器驱动电机轴正转, 电机螺母带动调节锥向增加流通面积的方向移动, 箱压随着增压气体质量流量的增大而升高; 当实时数值高于目标数值时, 驱动器驱动电机轴反转, 调节锥反向移动使得流通面积减小, 下游排液阀门开启时, 箱压随着增压气体质量流量的减小及箱内气体的排出而降低。

4　试验结果与分析

4.1　稳压试验

发动机点火、推力变化和关机状态通过控制下游阀门的开闭来模拟。预增压阶段, 控制程序检测到箱内压力低于目标数值, 由此驱动调节锥向开度增大的方向移动, 使得增压气体进入贮箱, 箱压上升直至高于预设目标数值。此时, 控制程序驱动调节锥向开度减小的方向移动, 直至阀门完全关闭。随后, 下游排液阀门开启, 箱内气体和推进剂一起沿下游管路排出, 导致箱压下降直至低于预设目标数值。此时, 控制程序再次驱动调节锥向开度增大的方向移动, 使得箱压再次提升。之后, 调节锥根据箱压的实时变化持续往复运动, 以保证箱压稳定在目标值附近。

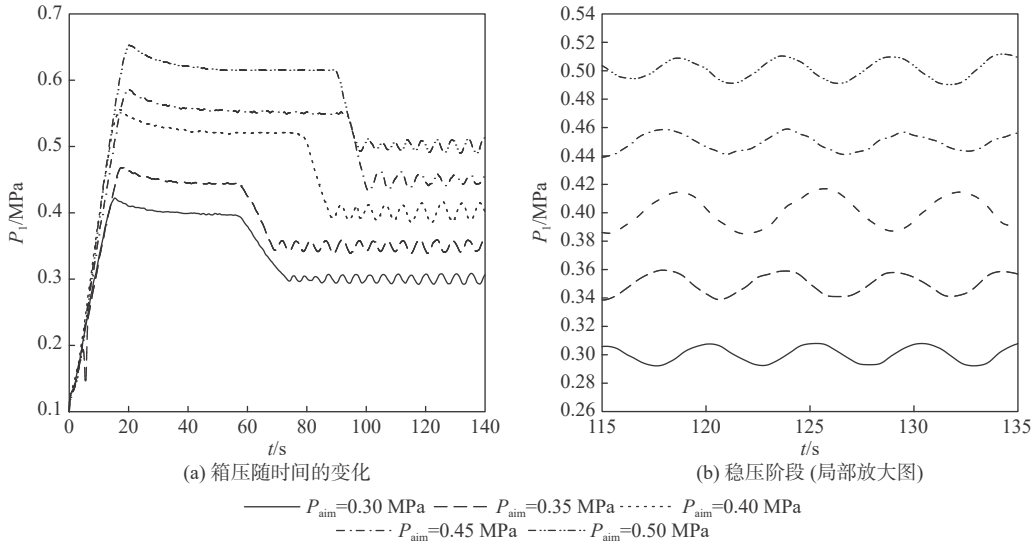
根据运载火箭贮箱常用气枕压力范围(0.2～0.5 MPa)[1,16] 设置 5 个试验工况, 目标箱压 P_{aim} 分别为 0.30, 0.35, 0.40, 0.45, 0.50 MPa, 气源系统出口压力设为 1 MPa。贮箱内未加水时, 各工况箱压随时间变化如图 12(a)所示, 对应超调量分别为 0.123, 0.118, 0.156, 0.137, 0.153 MPa, 控制精度分别为 $0.30^{+0.011}_{-0.010}$, 0.35 ± 0.012, $0.40^{+0.007}_{-0.012}$, $0.45^{+0.012}_{-0.017}$, $0.50^{+0.014}_{-0.010}$ MPa, 压力波动幅度位于 1.75%～3.78% 之间。箱内未加水时, 稳压阶段箱压波动曲线基本表现为类正弦波形式, 如图 12(b)所示。由表 3 可知, 稳压阶段箱压波动曲线平均振幅和频率分别位于 6.50～15.93 kPa 和 0.133～0.200 Hz 范围内。

第 1 次试验中, PLC 脉冲频率为 10 Hz, 第 2 次为 100 Hz, 可知, 脉冲频率恒定时, 箱压波动频率随压力的升高而降低, 同等工况下, 箱压波动频率随脉冲频率的升高而升高。通过调节下游阀门的开度来模拟发动机的变推力工况, 结果表明, 箱压均能稳定地维持在目标范围之内。

在贮箱内加入水, 对充填率 70%、目标箱压 0.5 MPa 的工况进行试验, 得到箱压 P_1 与下游排液质量流量 q_w 随时间的变化曲线, 如图 13 所示。与未加水阶段相比, 加水后预增压阶段超调量明显增加, 稳压阶段压力波动大幅降低, 压力控制精度提高为(0.5±0.003) MPa, 压力波动幅度降为 0.6%。

4.2　超压控制试验

调节文氏管的主要问题在于对贮箱压力超压的精准控制。影响超压量的主要因素包括上游压力、气体质量流量、阀门动作响应时间等。其中, 气体质量流量由阀门的开度控制, 阀门动作响应时间则



(a) 箱压随时间的变化　(b) 稳压阶段 (局部放大图)

图 12　箱压随时间的变化(加水前)

Fig. 12　Variation of tank pressure with time (before water addition)

表 3　平均振幅与频率

Table 3　Average amplitude and frequency

箱压/MPa	振幅/kPa	频率/Hz
0.30	7.75	0.200
0.35	9.00	0.190
0.40(第1次)	13.50	0.159
0.40(第2次)	6.50	0.195
0.45(第1次)	15.93	0.147
0.45(第2次)	7.94	0.190
0.50	8.65	0.133

与电机工作频率有关。一般希望超压量越小越好。

为了解电机工作频率对工作特性的影响,在气枕体积 30%、上游压力 0.7 MPa、目标箱压 0.4 MPa 工况下,分别采用电机工作频率 1 000, 500, 250 Hz 进行试验,结果如图 14(a)所示。试验表明,相同工况下,电机工作频率越高,压力变化速度越快,超压量越小。工作频率为 250 Hz 时,调节锥运动速度较慢,导致压降时间过长,未能在预定时间内达到稳压状态。电机工作频率为 500 Hz 与 1 000 Hz 时,对

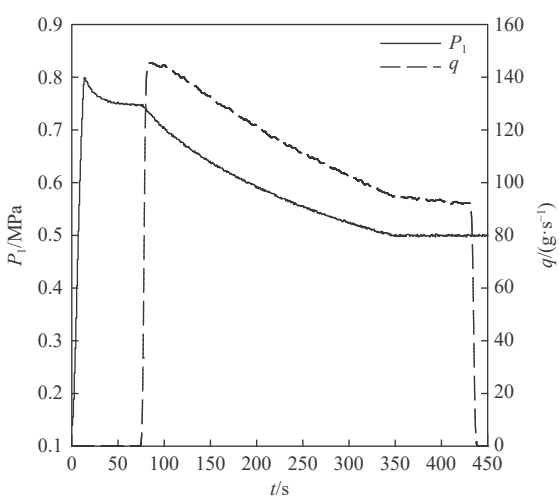
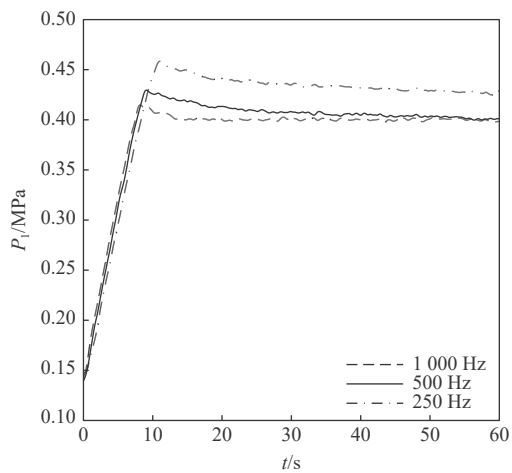
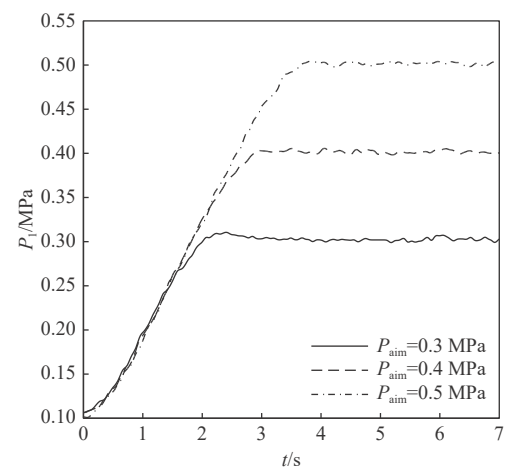


图 13　箱压与质量流量随时间的变化(加水后)

Fig. 13　Variation of tank pressure and flow rate with time (after water addition)



(a) 不同工作频率下箱压随时间变化 (目标箱压 0.4 MPa)



(b) 不同目标箱压下箱压随时间变化 (工作频率 500 Hz)

图 14　电机工作频率对工作特性的影响

Fig. 14　Influence of motor operating frequency on working characteristics

应的控制精度分别为$0.4^{+0.007}_{-0.005}$ MPa 和$0.4^{+0.008}_{-0.005}$ MPa, 压力波动幅度位于 1.25%～2% 之间; 但电机工作频率越高, 振动和压力波动就越剧烈。此外, 上游压力越高, 则超压量越大, 难以快速达到稳压状态。为确保电机平稳运转, 采用电机中等工作频率 500 Hz 的控制方案, 并加入超压控制程序, 以减少超压量并缩短压降时间。超压控制程序中, 在预增压阶段, 若箱压已达到目标箱压的指定百分比, 则提前驱动电机反转, 使得调节锥向开度减小的方向移动, 直至阀门完全闭合。这一过程中, 增压气体质量流量逐渐减少, 使得箱压在阀门闭合时达到略高于目标箱压的位置, 从而减小预增压阶段的超压量。控制程序中使用的判定参数的大小由每次正式试验前的预试验决定。

针对气枕体积 30%、上游压力 1.5 MPa、电机工作频率 500 Hz、目标箱压 0.3, 0.4, 0.5 MPa 的工况开展试验, 结果如图 14(b)所示。结果表明, 各工况下控制精度为$0.3^{+0.008}_{-0.003}$, $0.4^{+0.007}_{-0.004}$, $0.5^{+0.007}_{-0.002}$ MPa, 压力波动幅度位于 0.4%～2.7 % 之间。与加入超压控制程序前的同种工况相比, 预增压阶段超压量几乎为 0。

保持气枕体积 10%、电机工作频率 500 Hz, 分别在上游压力 1, 2, 3, 4 MPa 的情况下对目标箱压 0.3, 0.4, 0.5 MPa 进行试验, 并对上游压力 4 MPa、目标箱压 0.4 MPa 的工况开展长程稳态和长程变质量流量试验。在目标箱压 0.3 MPa 的工况下, 不同上游压力下增压段行程与压力 P_1、P_2 的对应关系如图 15 所示。为与仿真保持一致, 以阀门全开状态下调节锥行程为 0, 箭头所示方向为阀门逐渐打开方向, 图例中下标数字 1, 2, 3 表示上游压力值。

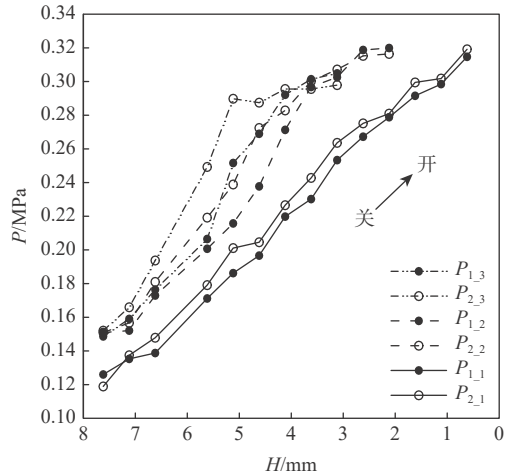


图 15　不同上游压力下箱压随调节锥行程的变化
(预增压阶段, 目标箱压 0.3 MPa)

Fig. 15　Variation of tank pressure with conical pintle stroke under different upstream pressures (pre-pressurization stage, target tank pressure at 0.3 MPa)

可知, 上游压力 1, 2, 3 MPa 下, 对应调节锥行程分别为 7, 5 , 4 mm, 表明上游压力越大, 气流充填速度越快, 调节锥需开启的面积越小; 上游压力越小, 增压段曲线上升越平稳。

试验数据表明, 试验装置在长程稳态工况下的控制精度为$0.4^{+0.007}_{-0.005}$ MPa, 在长程变流量工况下的控制精度为$0.4^{+0.010}_{-0.006}$ MPa, 各短程工况下的相应控制精度如图 16 所示, 下游排液质量流量如表 4 所示。可知各工况下下游排液质量流量较为稳定, 且随目标箱压升高而增加。试验结果表明, 不同工况下, 试验装置均能在所需精度范围(±0.02 MPa)内实现连续稳定的压力调节, 压力波动幅度位于 1.2%～3% 之间。

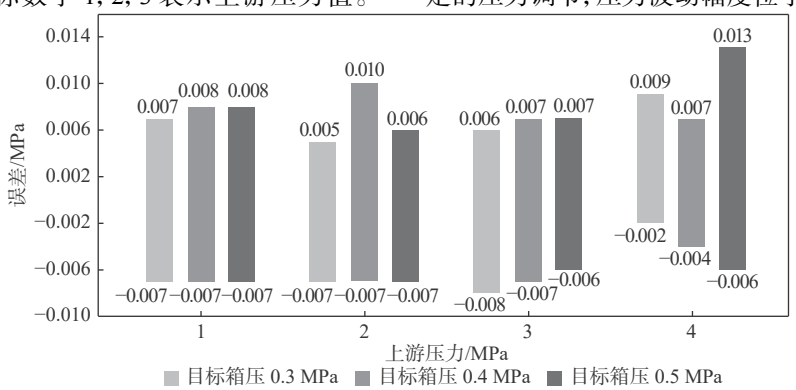


图 16　不同短程工况下压力控制精度

Fig. 16　Precision of pressure control under different short-range working conditions

表 4　不同短程工况下游质量流量

Table 4　Downstream flow rate under different short-range working conditions

上游压力/MPa	下游质量流量/(g·s^{-1})		
	目标箱压0.3 MPa	目标箱压0.4 MPa	目标箱压0.5 MPa
1	32.995	40.746	47.409
2	29.786	38.176	44.848
3	29.722	37.893	44.969
4	30.231	38.233	44.868

5　结　论

本文设计了一种基于可调文氏管的流量调节装置, 通过贮箱自动增压试验, 得出结论如下:

1) 提出了一种基于特型型面调节锥的可调文氏管设计方法, 设计所得调节锥形面曲线为抛物线, 能够使得气体质量流量与调节锥行程之间满足

较好的线性关系。与传统的锥形型面可调文氏管相比，该方法能够实现更高的控制精度；与使用电磁阀进行组合调压方案相比，该方法可以实现连续的无级增压。

2）阀门关闭过程中，调节锥所受轴向力逐步增加，且增加比例随上游压力增大而增大。临界声速截面始终位于喉部截面前端，喉部马赫数与质量流量系数均呈现先上升后下降趋势，喉部马赫数位于0.603 7～0.856 1范围内，质量流量系数位于0.890 6～0.964 7之间。阀门全开状态下，喉部截面所在位置距临界声速截面最远，此时喉部马赫数、质量流量系数及调节锥所受轴向力均为最小值。

3）加入超压控制前，相同工况下电机工作频率越高，压力变化速度越快，预增压阶段超调量越小，但系统的工作稳定性越低。贮箱未加水时，上游压力1 MPa、目标箱压0.3～0.5 MPa工况下的箱压调节精度位于−0.017～+0.014 MPa范围内，箱压波动幅度位于1.75%～3.78%之间。贮箱加水后，稳压阶段波动幅度减小，调节精度提高，但预增压阶段超调量增加。

4）加入超压控制后，能够在保证稳定性的情况下提升电机工作频率，提高增压效率，上游压力1～4 MPa、目标箱压0.3～0.5 MPa、下游排液质量流量0～48 g/s条件下的短程稳态工况（35 s）、长程稳态工况（700 s）和长程变质量流量工况（800 s）中，压力调节精度均位于−0.009～+0.013 MPa范围内，压力波动幅度位于1.2%～3%之间。与加入超压控制前同种工况相比，预增压阶段超压量几乎为0。

参考文献（References）

[1] 范瑞祥, 田玉蓉, 黄兵. 新一代运载火箭增压技术研究[J]. 火箭推进, 2012, 38(4): 9-16.
FAN R X, TIAN Y R, HUANG B. Study on pressurization of new generation launch vehicle[J]. Journal of Rocket Propulsion, 2012, 38(4): 9-16(in Chinese).

[2] 孙礼杰, 金鑫, 程光平, 等. 运载火箭贮箱增压控制技术发展综述[J]. 宇航总体技术, 2017, 1(1): 66-70.
SUN L J, JIN X, CHENG G P, et al. Review on development of pressurization control technology for launch vehicle storage tank[J]. Astronautical Systems Engineering Technology, 2017, 1(1): 66-70 (in Chinese).

[3] 廖少英. 一种新的运载火箭上面级和航天器轻型增压系统[J]. 上海航天, 2005, 22(1): 6-10.
LIAO S Y. The new light-duty pressurization system for upper stage of the launch vehicle and spacecraft[J]. Aerospace Shanghai, 2005, 22(1): 6-10(in Chinese).

[4] 《世界航天运载器大全》编委会. 世界航天运载器大全[M]. 2版. 北京: 中国宇航出版社, 2007.
The Editorial Board of World Spacecraft Encyclopedia. World spacecraft encyclopedia[M]. 2nd ed. Beijing: China Astronautic Publishing House. 2007(in Chinese).

[5] 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).

[6] ZHAO S, DV P E, ZHAO J B, et al. Design of a continuously controlled pressurization system for reusable launch vehicles[C]//Proceedings of the 69th International Astronautical Congress. Bremen: IAF, 2018: 46658.

[7] 韩泉东. 空间变推力液体火箭发动机流量调节及燃烧过程仿真研究[D]. 长沙: 国防科学技术大学, 2006.
HAN Q D. Numerical study on flow control and combustion process of variable thrust liquid propellant space rocket engine[D]. Changsha: National University of Defense Technology, 2006(in Chinese).

[8] 刘冉, 罗莲军, 陈盛, 等. 双级轴向涡流器文氏管长度对流场和喷雾特性影响[J]. 航空动力学报, 2021, 36(6): 1234-1243.
LIU R, LUO L J, CHEN S, et al. Effect of Venturi length on flow field and spray characterization of dual axial swirler[J]. Journal of Aerospace Power, 2021, 36(6): 1234-1243(in Chinese).

[9] ATESHKADI A, MCDONELL V G, SAMUELSEN G S. Effect of hardware geometry on gas and drop behavior in a radial mixer spray[J]. Symposium (International) on Combustion, 1998, 27(2): 1985-1992.

[10] 曹俊章, 邹小舟, 李洪洲, 等. 高温燃料流量调节发拉瓦尔管口特性分析[J]. 机床与液压, 2012, 40(6): 48-53.
CAO J Z, ZOU X Z, LI H Z, et al. Characteristic analysis of Laval nozzle in a fuel flow rate control valve at high temperature[J]. Machine Tool & Hydraulics, 2012, 40(6): 48-53(in Chinese).

[11] DITTAKAVI N, CHUNEKAR A, FRANKEL S. Large eddy simulation of turbulent-cavitation interactions in a venturi nozzle[J]. Journal of Fluids Engineering, 2010, 132(12): 121301.

[12] 高汉如. 液体火箭发动机变推力用流量可调汽蚀文氏管设计中几个问题的分析[J]. 工学学报, 1978(2): 43-63.
GAO H R. Analysis of several problems in the design of cavitation Venturi with adjustable flow rate for variable thrust of liquid rocket engine[J]. Journal of National University of Defense Technology, 1978(2): 43-63(in Chinese).

[13] 高汉如, 李晓滨. 线性可调汽蚀文氏管研究[J]. 国防科技大学学报, 1983, 5(4): 1-16.
GAO H R, LI X B. On linear throttling cavitation Venturi[J]. Journal of National University of Defense Technology, 1983, 5(4): 1-16(in Chinese).

[14] 唐飞, 李永, 耿永兵, 等. 线性可调汽蚀文氏管设计与数值仿真研究[J]. 空间控制技术与应用, 2013, 39(6): 12-16.
TANG F, LI Y, GENG Y B, et al. On design and numerical simulation for linear throttling cavitation venturi[J]. Aerospace Control and Application, 2013, 39(6): 12-16(in Chinese).

[15] 沈赤兵, 吴继平. 可调汽蚀文氏管试验研究[J]. 推进技术, 2004, 25(5): 473-476.

SHEN C B, WU J P. Experimental investigation on the controllable cavitating venturi[J]. Journal of Propulsion Technology, 2004, 25(5): 473-476(in Chinese).

[16] 廖少英. 低温液体推进剂增压分析与计算[J]. 上海航天, 1990, 7(1): 15-22.
LIAO S Y. Analysis and calculation of pressurization of cryogenic liquid propellant[J]. Aerospace Shanghai, 1990, 7(1): 15-22(in Chinese).

Design and experiment of adjustable Venturi tube for stepless automatic tank pressurization

REN Jialin[1, 2], GUO Hongjie[1, 2, *], LI Wenxin[1], LIANG Guozhu[1, 2], ZHAO Sheng[3]

(1. School of Astronautics, Beihang University, Beijing 102206, China;
2. Ningbo Innovation Research Institute of Beihang University, Ningbo 315800, China;
3. First Research Institute of China Aerospace Science and Technology Corporation, Beijing 100076, China)

Abstract: The stable control of tank pressure is crucial in the pressurization system of spacecraft tanks. In this paper, an adjustable Venturi tube based on a special conical pintle was designed, which could maintain a good linear correspondence between the pintle displacement and the flow rate, thereby improving the regulation accuracy. A closed automatic pressurization experiment based on the principle of negative feedback, as well as a system control and data acquisition based on programmable logic controllers were developed. The automatic pressurization experiment was conducted successfully. The results show that a high-precision, continuous, and stable pressure regulation over a wide flow range is accomplished by the device, offering some references for the design and improvement of the flow regulation device in the pressurization system.

Keywords: adjustable Venturi tube; profile design; tank pressurization; flow regulation; pressure regulation

Received: 2023-05-22; **Accepted**: 2023-07-20; **Published Online**: 2023-08-04 13:27
URL: link.cnki.net/urlid/11.2625.V.20230803.1830.003

* **Corresponding author**. E-mail: guohongjie@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0266

BDS-3 星载原子钟在轨运行特征分析

王文斌[1,2]，唐成盼[1,*]，杨建华[1,2]，胡小工[1]，曹月玲[1]，周善石[1]

(1. 中国科学院上海天文台，上海 200030；2. 中国科学院大学，北京 100049)

摘 要：为了评估北斗三号卫星导航系统(BDS-3)星载原子钟的在轨运行特征，基于国际GNSS服务(IGS)分析中心发布的广播钟差参数与精密钟差数据，从卫星广播钟差参数误差、广播钟速参数时变特征、频率准确度和频率漂移率等角度进行了分析。结果表明：在测试期间，BDS-3卫星钟差参数误差均值约为0.41 m，是影响卫星空间信号精度的关键因素；BDS-3广播钟速参数存在高频度跳变，星载原子钟高频率稳定度优势并未充分发挥；BDS-3星载原子钟频率准确度整体较差，但地球静止轨道(GEO)卫星氢钟频率准确度极优；BDS-3星载铷钟频率漂移率过大，导致调频次数过多，影响系统可用性；BDS-3星载氢钟频率漂移率较小，长期稳定性较高，适用于星上自主守时。

关 键 词：广播钟差；BDS-3卫星；原子钟；准确度；漂移率

中图分类号：P228.1

文献标志码：A **文章编号**：1001-5965(2025)05-1694-11

北斗卫星导航系统是中国自主建设、独立运行的卫星导航系统，是全球导航卫星系统(global navigation satellite system, GNSS)的重要成员[1]。2020年，北斗三号卫星导航系统(BeiDou-3 navigation satellite system, BDS-3)完成卫星全球组网，正式开通，其基于星间链路技术，向全球提供导航信号服务，是中国重要信息基础设施，具有极高的应用价值和商业价值[2]。在BDS-3系统中，精密的时间信息是准确定位的基础，时间的测量、转化与维持保障了整个系统的顺利运行。星载原子钟作为星上时间基准，是卫星的重要载荷，更是时间系统的核心组件，其性能的优劣直接决定了导航、定位和授时的质量。BDS-3卫星上搭配了铷原子钟(简称铷钟)和新型的高精度被动型氢原子钟(简称氢钟)[3]。进行卫星星载原子钟在轨运行特征分析与评估，有利于提升系统的服务性能，对系统的运维能力提升具有重大意义。

近年来，国内外学者围绕北斗星载原子钟的性能进行了深入的研究和探讨。其中，毛亚等[4]为分析卫星星载原子钟的时频性能，计算了在轨卫星钟差的频率稳定度、漂移率和准确度，试验结果表明，BDS-3试验卫星各项性能指标普遍优于北斗二号卫星(BDS-2)；廉吉庆等[5]基于iGMAS的钟差产品，分别对北斗卫星导航系统的铷钟和氢钟准确度、漂移率和稳定度进行了评估和对比；王鹏飞等[6]利用北斗卫星导航系统星地双向卫星钟差数据对BDS-3在轨铷钟性能进行了评估；杨宇飞等[7]以星座拟稳时间为观测量进行BDS-3星载原子钟性能及钟差预报精度分析，避免了卫星轨道误差对钟差的影响，得到了更加真实可靠的分析结果；阚昊宇等[8]结合BDS-3系统独特的星间链路和星地时间双向比对体制以及常用的精密轨道与钟差确定体制所估计的精密钟差数据，深入分析评估了BDS-3在轨原子钟服务性能；帅涛等[9]根据地面和在轨测试数据，给出了BDS-3卫星氢钟在轨运行性能评估结果；王宇谱等[10]使用卫星钟差二次多项式拟合模型

收稿日期：2023-05-23；录用日期：2023-07-10；网络出版时间：2023-08-04 11:53
网络出版地址：link.cnki.net/urlid/11.2625.V.20230803.1500.001
*通信作者. E-mail：cptang@shao.ac.cn

引用格式：王文斌，唐成盼，杨建华，等. BDS-3星载原子钟在轨运行特征分析[J]. 北京航空航天大学学报，2025，51(5)：1694-1704.
WANG W B，TANG C P，YANG J H，et al. In-orbit operation characteristics of BDS-3 spaceborne atomic clocks[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1694-1704 (in Chinese).

分析了北斗卫星星载原子钟的相位、频率、频率漂移率及钟差模型噪声的长期变化特性; 丁毅涛等[11]采用德国地学研究中心(German research centre for geosciences, GFZ)、西安测绘研究所分析中心、武汉大学分析中心、产品综合中心提供的精密钟差产品, 分析了全球定位系统(global positioning system, GPS)、北斗卫星导航系统、Galieo 卫星导航系统与 GLONASS 卫星导航系统在轨卫星星载原子钟的频率准确度、稳定度和漂移率; 伏军胜等[12]利用了一年的事后精密钟差数据, 对 BDS-3 和其他全球卫星导航系统的准确度、漂移率和稳定度进行分析, 评估了全球卫星导航系统在轨原子钟的性能。

上述研究以评估 BDS-3 星载原子钟性能为主。高性能的星载原子钟仅为系统在轨运行稳定服务提供了硬件基础, 而评判原子钟在轨运行特征回答星载原子钟的高性能优势是否被充分发挥, 需要以系统运维的角度对 BDS-3 卫星星载原子钟的在轨运行服务特征而非原子钟自身性能进行重新评价。卫星广播钟差参数的特征可在一定程度上反映星载原子钟的在轨运行服务特征。

本文利用 2021 年全年的广播钟差参数数据, 从 BDS-3 卫星的广播钟差参数误差、BDS-3/GPS/Galileo 星载原子钟的广播钟差和钟速的长期变化特征、频率准确度、频率漂移率等指标的角度分析评估了 BDS-3 卫星星载原子钟在轨运行特征, 明确了 BDS-3 氢钟自身性能的领先地位, 若能有效降低双向时间比对的测量误差, 便可显著提高 BDS-3 卫星广播钟差精度。

1　BDS-3 卫星广播钟差参数误差评估

1.1　数据选择

取 2021 年 8 月 29 日至 2021 年 9 月 4 日共 7 d 的数据进行了 BDS-3 广播钟差参数误差评估, 并对轨道和钟差参数误差进行了比较。以 2021 年 9 月 2 日的数据为例, 由中国卫星导航系统管理办公室的 BDS-3 导航电文和 GFZ 的精密钟差分别计算了 1 d 内 BDS-3 卫星的钟差(每 30 s 一次共 2 880 个历元), 做差后进行了差分码偏差(differential code bias, DCB)修正和不同时间系统的差异修正, 从而得到了 BDS-3 卫星的钟差参数误差。

1.2　修正方法

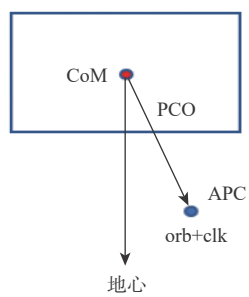
BDS-3 广播星历是基于单频 B3 信号播发的, 而 GFZ/IGS 精密钟差产品则采用 B1/B3 频点的无电离层延迟组合, 因此, 必须使用星历中的电离层群延迟改正参数 B1/B3(T_{GD})进行 DCB 修正, 将钟差从单频归算到双频再进行比较, 如下:

$$t_{B1B3} = t_{B3} - T_{GD}\frac{f_1^2}{f_1^2 - f_3^2} \tag{1}$$

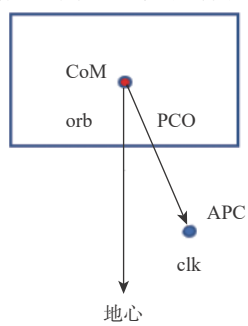
式中: t_{B1B3}为归算到B1/B3双频下的钟差结果; t_{B3}为B3单频下的广播钟差结果; T_{GD}为电离层群延迟改正参数; f_1为B1频点频率; f_3为B3频点频率。

BDS-3 导航电文的时间基准北斗时与 GFZ/IGS 精密产品的时间基准 GPS 时存在 14 s 之外的纳秒级差异, 该差异通常大于时钟解的固有精度, 但会以相同的方式影响一个星座的所有卫星, 并导致系统偏差随时间而变化[13]。因此, 需要先扣除系统时间偏差的影响后再对原子钟本身进行评估, 即先计算每个历元上星座所有卫星的广播钟差与精密钟差的差值的平均值, 再将其减去, 从而得到更精确的广播钟差参数误差。

图 1 为 GNSS 广播电文和精密产品中轨道 (orb) 和钟差 (clk) 参考点的不同约定[13]。图中: CoM 表示质心, APC 表示天线相位中心, PCO 表示相位中心偏差。因此, 计算轨道需要进行相位中心改正, 钟差则不需要。



(a) 广播电文中轨道和钟差参考点相对位置



(b) 精密产品中轨道和钟差参考点相对位置

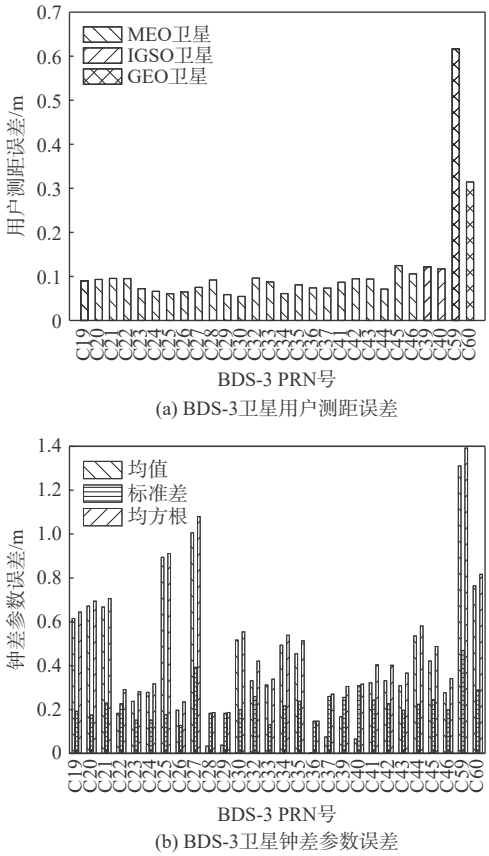
图 1　不同产品中卫星天线参考点示意图[13]

Fig. 1　Satellite antenna reference point in different products[13]

不同 GNSS 系统的接口控制文件指出用户在通过钟差参数计算卫星钟差时需考虑相对论修正, 然而在使用精密钟差进行定位时要进行相同的修正。因此, 进行卫星钟差比较时, 基于广播星历的钟差无需进行相对论修正。

1.3　误差分析

图 2 为不同 BDS-3 卫星轨道和钟差参数误差



(a) BDS-3卫星用户测距误差

(b) BDS-3卫星钟差参数误差

图2 不同BDS-3卫星轨道和钟差参数误差统计结果

Fig. 2 Statistical results of orbital errors and clock parameter errors of different BDS-3 satellites

统计结果。其中，图2(a)显示了不同BDS-3卫星的用户测距误差，包括中圆地球轨道(medium earth orbit, MEO)卫星、倾斜地球同步轨道(inclined geosynchronous orbit, IGSO)卫星、地球静止轨道(geostationary orbit, GEO)卫星；图2(b)显示了不同BDS-3卫星广播钟差参数误差的各项指标，包括均值、标准差、均方根。

由图2(a)可知，BDS-3 MEO和IGSO卫星平均轨道用户测距误差约为0.10 m，而GEO卫星明显大于其他2种轨道卫星。这是由于在利用地面测距数据给GEO卫星定轨处理时，轨道切向存在较大误差，进而导致轨道用户测距误差较大[14]。

由图2(b)可知，BDS-3各轨道类型的卫星中，GEO卫星钟差参数误差最大，MEO卫星次之，IGSO卫星最小，BDS-3卫星钟差参数误差总均值为0.41 m。BDS-3卫星钟差参数标准差约为0.1～0.5 m，其中，GEO卫星钟差参数标准差最大，稳定性也最差，MEO卫星和IGSO卫星基本相当。BDS-3卫星钟差参数均方根误差约为0.2～0.7 m，不同卫星差异较大，GEO卫星钟差参数均方根误差最大，MEO卫星次之，IGSO卫星最小。

由上述分析可知，BDS-3 IGSO和MEO卫星轨道用户测距误差约为0.10 m，而BDS-3卫星钟差参数误差均值约为0.41 m，量级较大，高于Galileo系统(约为0.18 m)[15]，可见BDS-3卫星广播钟差参数误差显著大于广播星历误差，这是影响卫星空间信号精度的关键因素。

2 BDS-3广播钟速长期变化情况分析

2.1 钟差建模

为了分析卫星原子钟的时频特性，可采用包含相位、频率、频率漂移率等指标的二次多项式构建精密钟差模型。该模型可表示为

$$x_i = a_0 + a_1(t_i - t_0) + a_2(t_i - t_0)^2 + \varepsilon_i \tag{2}$$

式中：$i = 1,2,\cdots,n$；t_i为第i个历元的时刻；x_i为该时刻的卫星钟差观测值；ε_i为该时刻的观测噪声误差，可由拟合残差表征；t_0为参考时刻；a_0、a_1、a_2分别表示参考时刻的相位、频率、频率漂移率。

2.2 数据选择

取2021年9月1日至2021年9月30日共30 d的IGS导航电文数据，从中提取出所有BDS-3、GPS、Galileo卫星广播钟速信息(即导航电文文件中的a_1字段)进行分析评估。其中，BDS-3卫星广播钟速更新周期为1 h，GPS卫星广播钟速更新周期为2 h，Galileo卫星广播钟速更新周期为10 min。

分析评估包括广播频率参数(即a_1字段)长期变化、广播频率准确度分析。分析评估前，根据卫星健康字段找出发生相位、频率调整的卫星，将上述卫星剔除，不参与本节分析评估。

2.3 频率长期变化

图3为各系统部分卫星原子钟频率长期变化。对比3个系统可以发现，每小时更新的BDS-3广播钟速剧烈变化，变化序列毛刺点较多，广播钟速的波动与预报策略有关，亦是测量误差导致[16]。GPS广播钟速在周期内保持稳定，量级约为10^{-12} s/s，几天至几十天变化一次，波动量级约为10^{-13} s/s。Galileo广播钟速进行了量化操作，取值有限并发生小幅波动，波动量级约为10^{-13} s/s，具有一定的周期性，该周期性源于卫星钟差参数的快速更新播发策略。Galileo系统充分发挥氢钟短稳高的特性，通过快速更新钟差参数，利用广播钟差参数吸收广播星历周期波动的径向轨道误差，增强卫星轨道和钟差参数误差的耦合性，进而显著降低空间信号测距误差[17]。此外，BDS-3 GEO卫星星载氢

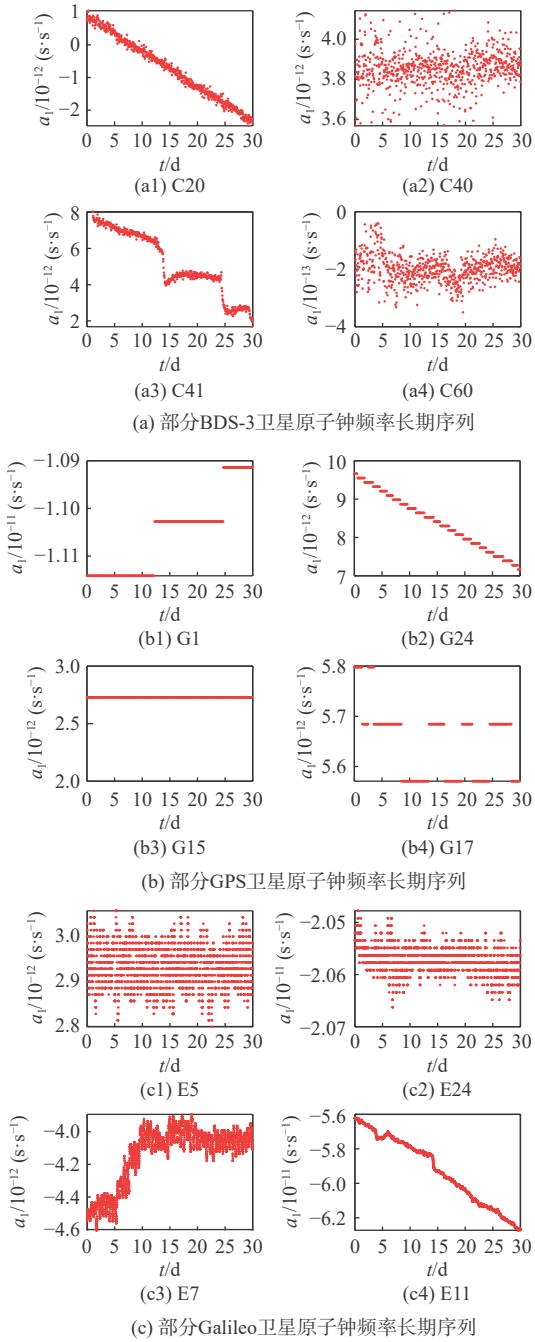


图 3　不同系统卫星原子钟频率长期变化情况

Fig. 3　Long-term frequency variations of spaceborne atomic clocks from different systems

钟广播 a_1 很小, 仅约为 10^{-13} s/s量级。

通过图 3, 对广播钟速波动情况作了定性的描述。为了对其进行定量的分析, 可通过计算标准差来评估星载原子钟广播钟速的稳定性。由于铷钟频率漂移率较大导致广播钟速出现随时间的线性变化, 故先扣除数据中的线性趋势项从而得到残差, 再求标准差来评估广播钟差的离散程度。各星载原子钟广播钟速标准差结果如图 4 和图 5 所示。可以看出, BDS-3 氢钟广播钟速标准差约为 9.513×10^{-14} s/s, 大于 Galileo 氢钟。扣除趋势项后, BDS-3 铷钟广播钟速标准差约为 1.063×10^{-13} s/s, 大于 GPS 铷钟和 BDS-3 氢钟。GPS 铷钟广播钟速稳定性最好, 约为 3.525×10^{-14} s/s。

频率稳定度是衡量原子钟性能优劣的重要指标之一, 常用的评估方法有阿伦方差、时间方差、哈达玛方差等, 方差量级越小, 频率稳定度越高。杨宇飞等[7,18]计算了 BDS-3 部分星载原子钟的阿伦方差和哈达玛方差, 结果表明, BDS-3 星载原子钟万秒稳处于 10^{-14} 量级, 与其他 GNSS 系统星载原子钟基本处于同一水平。然而 BDS-3 星载原子钟广播钟速扣除趋势项后, 其标准差约为 10^{-13} 量级, 大于自身万秒稳指标一个量级, 也比其他系统星载原子钟标准差大一个数量级, 说明 BDS-3 星载原子钟广播钟速的高频度波动与原子钟本身的性能无关, 是由双向时间比对的测量误差导致。

2.4　频率准确度

频率准确度表征的是频标的实际输出频率与其标称频率的一致程度, 即测量值或计算值与理想值的符合情况, 可表示为

$$A=\frac{f_{\mathrm{m}}-f_0}{f_0} \tag{3}$$

式中: A 为频率准确度; f_{m} 为实测频率; f_0 为测量频标的标称频率。

实际操作中, f_0 很难准确测量, 一般取为拟合频率平均值, 只要保证参考频率准确度比实测频率准确度差一个量级即可。采样周期较小时, 可通过时差比对求得星载原子钟相对系统时的频率偏差, 因为系统时频标准确度较高, 故频率准确度可由相对频率偏差近似得到, 即

$$A=\frac{\Delta x}{\Delta T}=\left|\frac{x(t_2)-x(t_1)}{t_2-t_1}\right|=\left|\frac{x(t+\tau)-x(t)}{\tau}\right| \tag{4}$$

式中: $x(t)$ 为 t 时刻卫星钟差; Δx 为钟差观测值的时差; ΔT 为观测时刻的时差; τ 为取样间隔。当采样周期较大、观测数据较多时, 为保证计算结果的准确性, 可直接采用最小二乘法对卫星钟差实测值进行线性拟合, 直线斜率即为频率准确度近似值, 即

$$x(t)=A(t-\bar{t})+\bar{x} \tag{5}$$

$$A=\frac{\sum_{i=1}^{N}(x_i-\bar{x})(t_i-\bar{t})}{\sum_{i=1}^{N}(t_i-\bar{t})^2} \tag{6}$$

式中: N 为取样个数, 取值为正整数; $i=1,2,\cdots,N$;

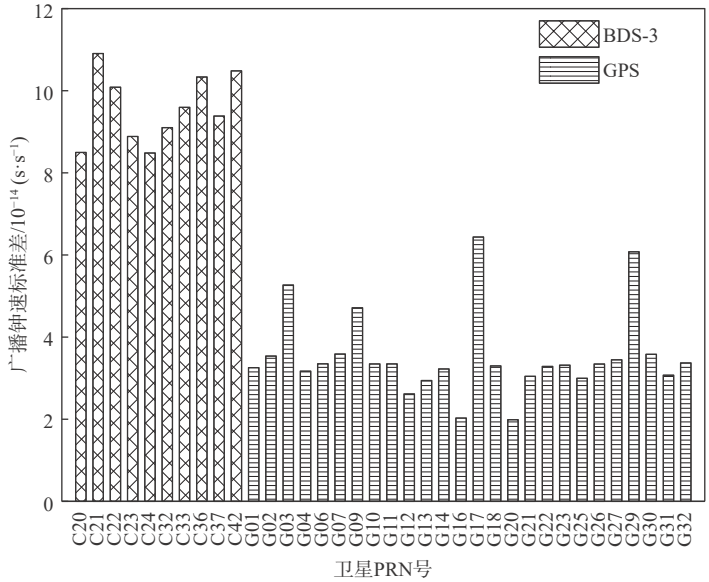


图4 不同系统星载铷钟广播钟速标准差

Fig. 4 Standard deviations of broadcast clock speed of spaceborne rubidium clocks from different systems

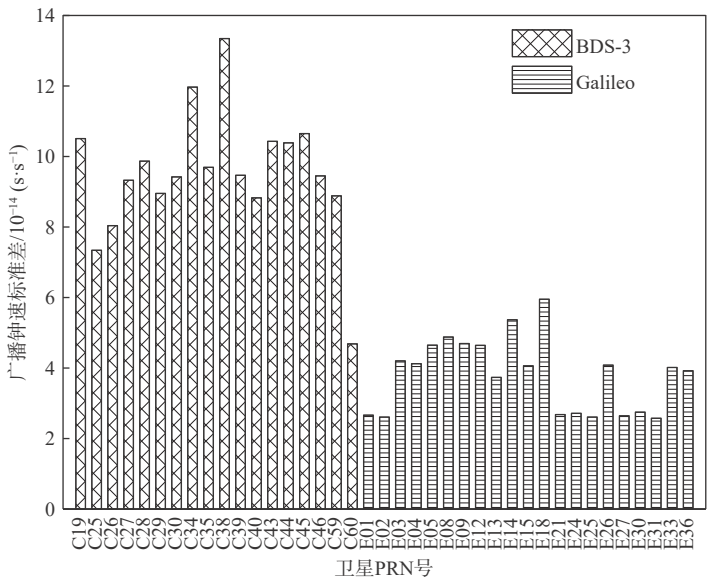


图5 不同系统星载氢钟广播钟速标准差

Fig. 5 Standard deviations of broadcast clock speed of spaceborne hydrogen clocks from different systems

$\bar{t}$为取样时刻平均值，$\bar{t}=\sum_{i=1}^{N}t_i/N$；$\bar{x}$为卫星钟差平均值，$\bar{x}=\sum_{i=1}^{N}x_i/N$。

由于广播钟速就是由一小时钟差拟合一次项系数求得，故本文直接取广播钟差a_1项的绝对值即可得到该取样时间的星载原子钟频率准确度，结果如图6和图7所示。结果表明，BDS-3 GEO卫星原子钟频率准确度最优，均值约为1.445×10^{-13}，IGSO次之，MEO最差，约为$10^{-12}\sim10^{-11}$量级，BDS-3氢钟频率准确度略优于铷钟，前者约为8.261×10^{-12}，后者约为8.867×10^{-12}。BDS-3氢钟频率准确度略优于Galileo氢钟。BDS-3铷钟频率准确度差于GPS铷钟。三大GNSS系统中，GPS铷钟频率准确度最优，均值约为5×10^{-12}，Galileo氢钟频率准确度最差，均值约为1×10^{-11}。

由上述分析可知，BDS-3广播钟速频繁跳变，星载原子钟高频率稳定度优势并未充分发挥，受预报策略和测量误差影响。GPS广播钟速预报周期较长，波动较小，较为稳定。Galileo广播钟速进行了量化操作，取值有限且具有一定的周期性。三大GNSS系统星载原子钟总体来看，GPS铷钟频率准确度最优，BDS-3星载原子钟频率准确度居中，但GEO卫星氢钟频率准确度极高，可达$10^{-14}\sim10^{-13}$量

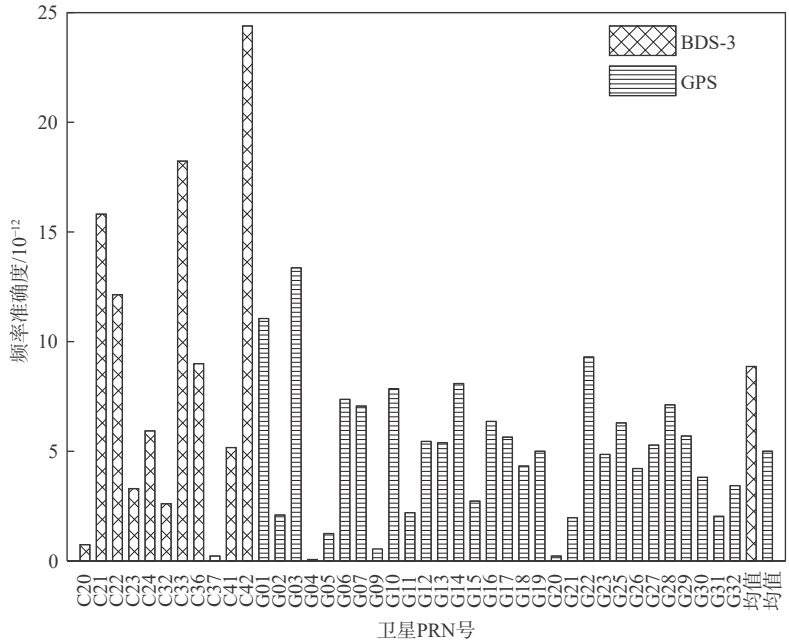


图 6　不同系统星载铷钟频率准确度结果

Fig. 6　Frequency accuracy results of spaceborne rubidium clocks from different systems

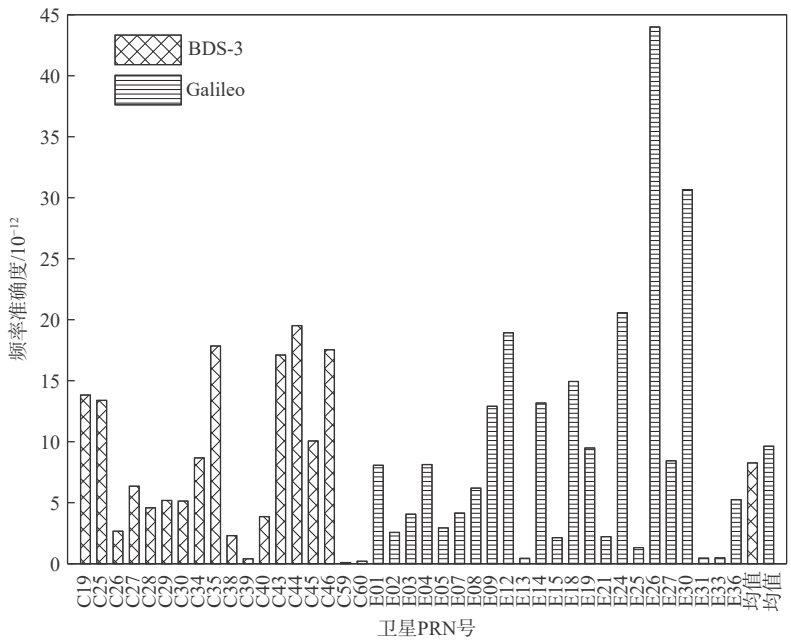


图 7　不同系统星载氢钟频率准确度结果

Fig. 7　Frequency accuracy results of spaceborne hydrogen clocks from different systems

级, Galileo 氢钟频率准确度最差。

3　BDS-3 广播钟差长期变化情况分析

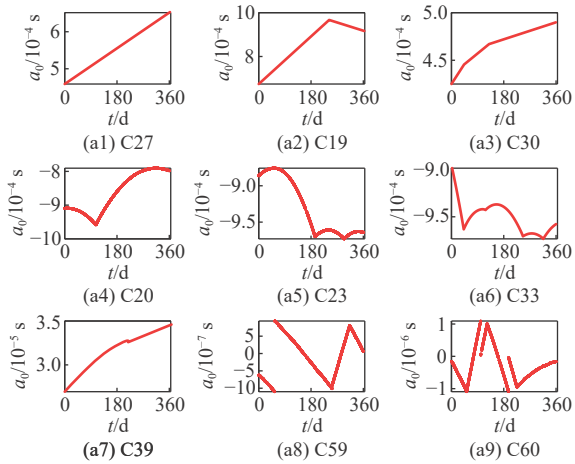
3.1　数据选择

取 2021 年 1 月 1 日至 2021 年 12 月 26 日共 360 d 的 IGS 导航电文数据, 从中提取出所有 BDS-3、GPS、Galileo 卫星广播钟差信息(即导航电文文件中的 a_0 字段)进行分析评估。

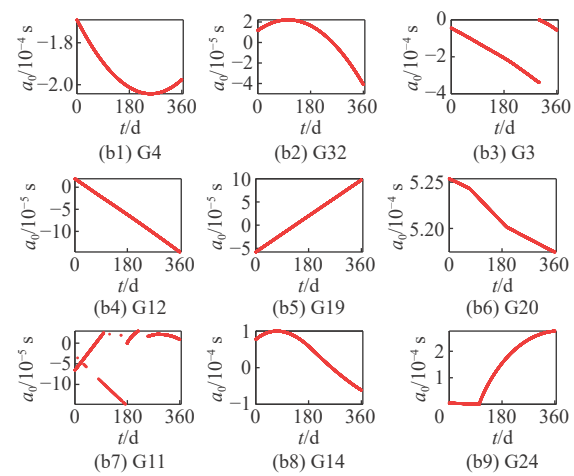
分析评估包括广播钟差参数(即 a_0 字段)长期变化、调频调相情况、广播钟差相位长期稳定性、频率漂移率分析。分析评估时, 以卫星调频调相时刻为节点对 a_0 项进行分段拟合, 进而得到更准确的拟合残差和频率漂移率信息。

3.2　钟差长期变化

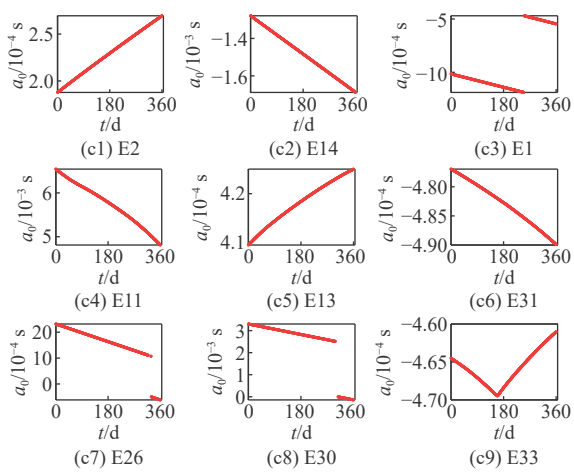
图 8 为各系统部分卫星原子钟相位长期变化。对比 3 个系统可以发现, BDS-3 铷钟相位量级较大, 接近 1×10^{-3} s, 铷钟调频次数较多。GEO 卫星星载氢钟钟差量级最小, 约为 10^{-7} s, 且进行了多次调频调相。GPS 铷钟相位约为 $10^{-4}\sim10^{-5}$ 量级, G11、G24 卫星星载原子钟进行调频调相操作后, 相



(a) 部分BDS-3卫星原子钟相位长期序列



(b) 部分GPS卫星原子钟相位长期序列



(c) 部分Galileo卫星原子钟相位长期序列

图 8 不同系统卫星原子钟相位长期变化情况

Fig. 8 Long-term phase variations of spaceborne atomic clocks from different systems

位从线性变化切换为抛物线变化，G14 卫星星载原子钟钟差随时间呈现类似三次函数图像，频率漂移率项发生了缓慢漂移，即 a_2 项随时间发生了缓慢变化。Galileo星载原子钟连续性较好，调频调相次数最少，但 E11、E14、E30 星载原子钟相位量级较大，可达10^{-3}量级，大于 BDS-3 和 GPS。

3.3 调频调相情况

各系统星载原子钟调频调相情况如图 9 所示。对比不同系统的星载原子钟可以发现，3 个系统中，BDS-3 星载原子钟发生调频调相的次数最多，其中大多为 MEO 卫星铷钟，GEO 卫星氢钟调频调相次数偏多，最高可达 5 次。GPS 仅有少数卫星进行了调频调相，次数一般为 1 次，G11 卫星星载原子钟 a_0 项变化较乱，由于卫星换代调试，导致调频调相较频繁。Galileo 星载原子钟调频调相最少，有 5 颗卫星各进行了 1 次调频调相操作。

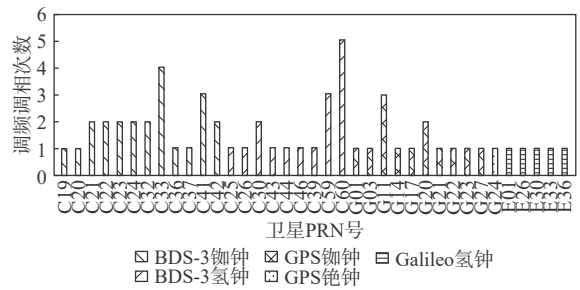


图 9 不同系统星载原子钟调频调相次数

Fig. 9 Frequency and phase modulation times of spaceborne atomic clocks from different systems

BDS-3 铷钟调频次数较多，因为其频率漂移率过大，引起钟速快速改变，进而容易导致相位变化迅速，超出星地接口门限。因此，BDS-3 铷钟面临频繁调频。C33 卫星一年内出现 4 次调频情况，影响单星可用性。

3.4 长期稳定性

取 2021 年 1～360 d 广播钟差相位数据进行二次多项式拟合，各系统星载原子钟相位拟合残差情况如图 10 和图 11 所示。可以看出，BDS-3 氢钟广播钟差长期稳定性高于 BDS-3 铷钟，前者拟合残差均方根约为 40.303 ns，后者约为 209.641 ns。BDS-3 铷钟广播钟差长期稳定性低于 GPS 铷钟，后者拟合残差均方根约为 82.313 ns。BDS-3 氢钟广播钟差长期稳定性高于 Galileo 氢钟，后者拟合残差均方根约为 52.658 ns。在各系统星载原子钟中，BDS-3 氢钟广播钟差长期稳定性最为优越。

星载原子钟钟差的长期稳定性是衡量其自主守时能力的重要指标。自主守时是指在没有地面支持的情况下采用连续运转，高度稳定的星载原子钟建立和保持时标，实现导航卫星时间系统的维持。若原子钟长期稳定性较差，当其处于自主守时状态时，易与协调世界时产生更大的偏差，故星载原子钟长稳越优，星上自主守时能力越强。从各系统星载原子钟长期稳定性的评估结果可以看出，BDS-3 氢钟广播钟差长期稳定性最高，故 BDS-3 星载氢钟具备星上长期自主守时的能力，在各系统星

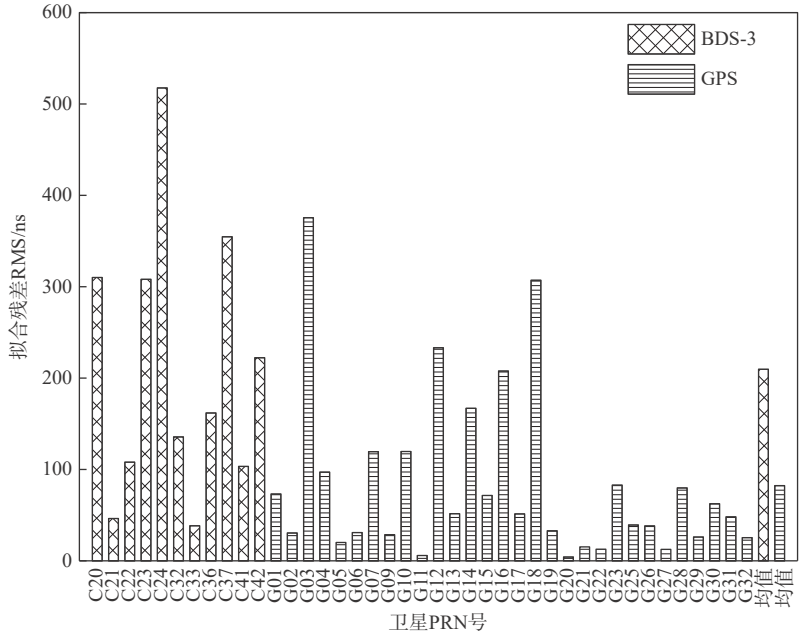


图 10　不同系统星载铷钟广播钟差相位拟合残差

Fig. 10　Fitting residuals of broadcast clock offset phase of spaceborne rubidium clocks from different systems

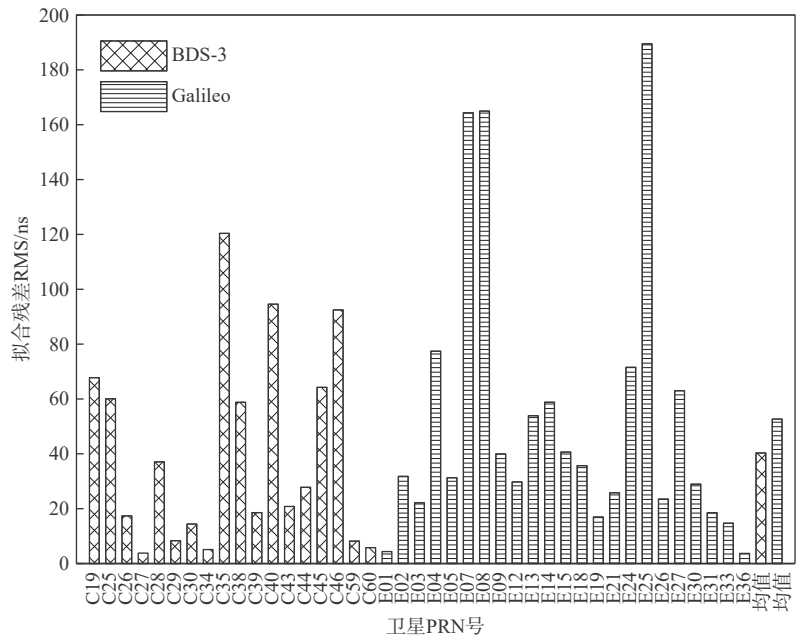


图 11　不同系统星载氢钟广播钟差相位拟合残差

Fig. 11　Fitting residuals of broadcast clock offset phase of spaceborne hydrogen clocks from different systems

载原子钟中居于领先地位。

3.5　频率漂移率

频率漂移率表征的是原子钟输出频率随时间的变化情况。随着卫星在轨时间的增加, 星载原子钟受外部电磁和热环境的变化、控制电路物理部分内部状态的不稳定以及自身元器件的老化等多种因素的影响, 输出频率会呈现单调变化的趋势, 其线性变化率即可视为频率漂移率, 它是衡量原子钟自主守时能力的重要参数。某些原子钟的频率漂移通常是因为其内部高稳石英晶体振荡器随运行时间的增大而发生老化, 此时频率漂移率亦可被称作频率老化率。

频率漂移率的瞬时解可表示为

$$D=\frac{\Delta y}{\Delta T}=\left|\frac{y(t_2)-y(t_1)}{t_2-t_1}\right|=\left|\frac{y(t+\tau)-y(t)}{\tau}\right| \tag{7}$$

式中: D为频率漂移率; $y(t)$为t时刻卫星钟速; Δy为不同时刻的频率差。

频率漂移率长期解可通过对相位长期序列进行二次拟合来求得, 本文取 2021 年 1～360 d 广播钟差相位数据进行二次多项式拟合, 拟合 a_2 项结果

如图12和图13所示。可以发现，BDS-3氢钟频率漂移率远优于铷钟，前者约为5.243×10^{-21} s/s^2，后者约为4.810×10^{-19} s/s^2，相差2个数量级。BDS-3铷钟频率漂移率差于GPS铷钟，后者约为5.036×10^{-20} s/s^2。BDS-3氢钟频率漂移率略优于Galileo氢钟，后者约为5.624×10^{-21} s/s^2。

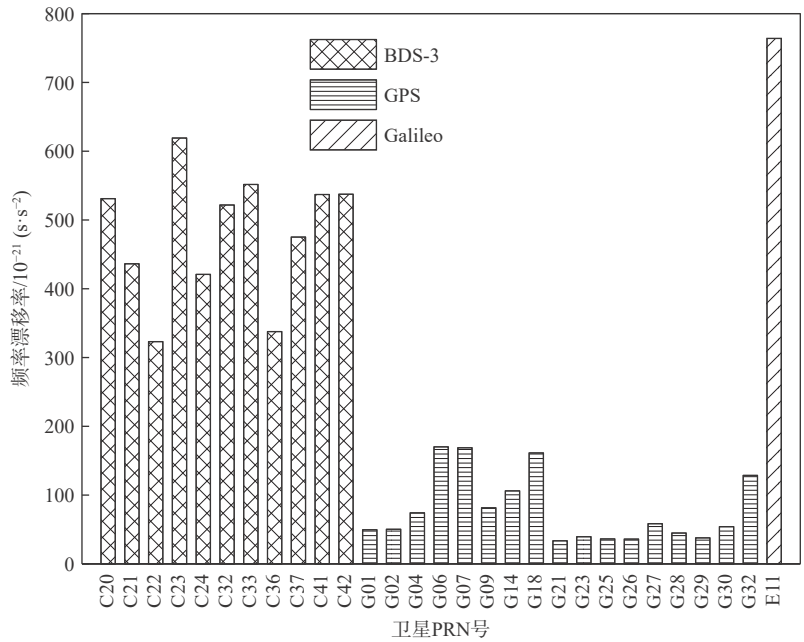


图12 不同系统铷钟频率漂移率结果

Fig. 12 Frequency drift rate results of spaceborne rubidium clocks from different systems

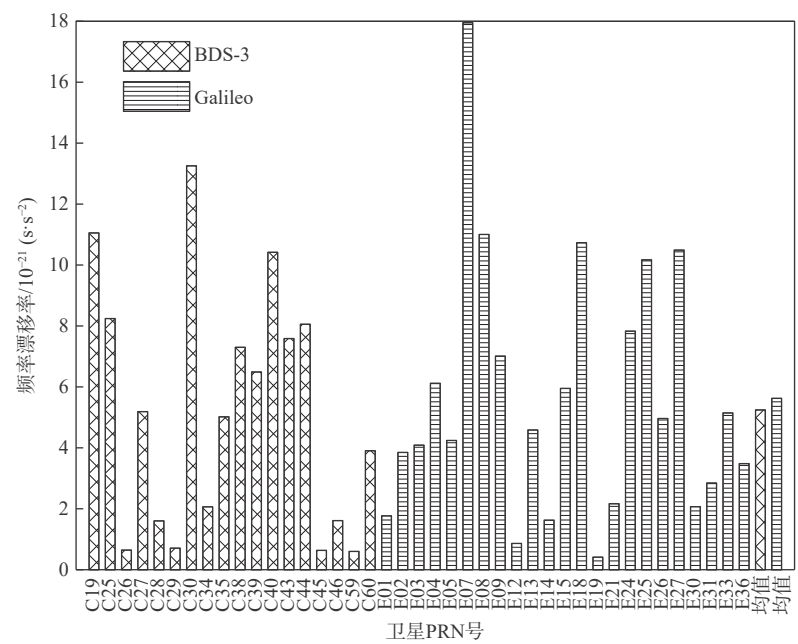


图13 不同系统氢钟频率漂移率结果

Fig. 13 Frequency drift rate results of spaceborne hydrogen clocks from different systems

由上述分析可知，各系统星载原子钟中，BDS-3星载铷钟频率漂移率最差，广播钟差相位长期稳定性最低，调频次数过多，影响系统可用性，BDS-3氢钟频率漂移率最优，相位长期稳定性最高，适用于星上自主守时。GPS铷钟频率漂移率和相位长期稳定性均优于BDS-3铷钟1个数量级，但不如BDS-3氢钟。Galileo氢钟频率漂移率小，调频调相次数最少，相位长期稳定性较高。

4 结 论

1）在试验期间，BDS-3卫星钟差参数均方根误差约为0.2～0.7 m，均值约为0.41 m，量级较大，标

准差约为 0.22 m, 离散程度较大, 稳定性不高, 仍有很大提升空间。BDS-3 卫星广播钟差参数误差显著大于广播星历误差, 是提高卫星空间信号精度的关键瓶颈。

2) BDS-3 氢钟频率准确度略优于 Galileo 氢钟和 BDS-3 铷钟, BDS-3 铷钟频率准确度差于 GPS 铷钟, BDS-3 GEO 卫星氢钟频率准确度量级较小, 可达 10^{-14} ~ 10^{-13}, GPS 铷钟频率准确度最优。

3) BDS-3 广播钟速波动较大, 离散点较多, 受预报策略和测量误差影响, 星载原子钟高频率稳定度优势并未充分发挥; GPS 广播钟速预报周期较长, 波动较小, 较为稳定; Galileo 广播钟速进行了量化, 取值有限且具有一定的周期性。

4) 在试验期间, BDS-3 氢钟频率漂移率略优于 Galileo 氢钟, BDS-3 铷钟频率漂移率差于 GPS 铷钟。三大 GNSS 系统中, BDS-3 星载铷钟频率漂移率最差, 广播钟差相位长期稳定性最低, 调频次数最多, 影响系统可用性, BDS-3 氢钟频率漂移率最优, 相位长期稳定性最高, 适用于星上长期自主守时。

参考文献 (References)

[1] 谭述森. 北斗卫星导航系统的发展与思考[J]. 宇航学报, 2008, 29(2): 391-396.
TAN S S. Development and thought of compass navigation satellite system[J]. Journal of Astronautics, 2008, 29(2): 391-396(in Chinese).

[2] 林夏, 林宝军, 刘迎春, 等. 一种同轨区域集中的北斗卫星自主导航算法[J]. 宇航学报, 2021, 42(1): 113-121.
LIN X, LIN B J, LIU Y C, et al. A coplanar regional centralized autonomous navigation algorithm for Beidou satellites[J]. Journal of Astronautics, 2021, 42(1): 113-121(in Chinese).

[3] 王威雄, 董绍武, 武文俊, 等. 北斗三号卫星共视时间比对性能分析[J]. 宇航学报, 2020, 41(5): 569-577.
WANG W X, DONG S W, WU W J, et al. Analysis of common view time comparison by BDS-3 satellite[J]. Journal of Astronautics, 2020, 41(5): 569-577(in Chinese).

[4] 毛亚, 王潜心, 胡超, 等. 北斗在轨卫星广播星历精度分析[J]. 宇航学报, 2018, 39(9): 1013-1021.
MAO Y, WANG Q X, HU C, et al. Precision analysis of BeiDou satellites broadcast ephemeris[J]. Journal of Astronautics, 2018, 39(9): 1013-1021(in Chinese).

[5] 廉吉庆, 张文玺, 张金海, 等. 基于 iGMAS 的北斗导航卫星原子钟性能评估[J]. 真空与低温, 2022, 28(5): 609-614.
LIAN J Q, ZHANG W X, ZHANG J H, et al. Evaluation of BDS atomic frequency standard based on iGMAS[J]. Vacuum and Cryogenics, 2022, 28(5): 609-614(in Chinese).

[6] 王鹏飞, 王芳, 赵峰, 等. 基于星地双向时间比对数据的北斗三号铷原子钟在轨性能评估[J]. 天文学进展, 2021, 39(4): 555-564.
WANG P F, WANG F, ZHAO F, et al. Performance analysis of BD Ⅲ satellite rubidium atomic clock based on satellite-ground two-way time transfer data[J]. Progress in Astronomy, 2021, 39(4): 555-564(in Chinese).

[7] 杨宇飞, 杨元喜, 陈金平, 等. 北斗三号星座拟稳钟差测定及其预报[J]. 测绘学报, 2021, 50(12): 1728-1737.
YANG Y F, YANG Y X, CHEN J P, et al. Pseudo-stable constellation bias error of BDS-3 and its high-precision prediction[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(12): 1728-1737(in Chinese).

[8] 阚昊宇, 胡志刚, 吕逸飞, 等. 利用不同时间同步体制钟差评估北斗三号星载原子钟性能[J]. 武汉大学学报 (信息科学版), 2023, 48(4): 604-610.
KAN H Y, HU Z G, LÜ Y F, et al. Performance evaluation of BDS-3 spaceborne atomic clock using different time synchronization systems[J]. Geomatics and Information Science of Wuhan University, 2023, 48(4): 604-610(in Chinese).

[9] 帅涛, 林宝军, 张军, 等. 北斗导航卫星氢原子钟性能分析评估[J]. 中国科学: 物理学 力学 天文学, 2021, 51(1): 125-134.
SHUAI T, LIN B J, ZHANG J, et al. Performaces and telemetres analysis of BD satellite passive hydrogen maser[J]. Scientia Sinica (Physica, Mechanica & Astronomica), 2021, 51(1): 125-134(in Chinese).

[10] 王宇谱, 吕志平, 王宁. BDS 星载原子钟长期性能分析[J]. 测绘学报, 2017, 46(2): 157-169.
WANG Y P, LÜ Z P, WANG N. The long-term performance analysis for on-board atomic clocks of BDS[J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(2): 157-169(in Chinese).

[11] 丁毅涛, 郭美军. 基于不同机构钟差产品的 GNSS 星载钟性能分析与评估[J]. 时间频率学报, 2020, 43(1): 72-84.
DING Y T, GUO M J. Performance analysis and evaluation of GNSS satellite clock based on clock products from different institutions[J]. Journal of Time and Frequency, 2020, 43(1): 72-84(in Chinese).

[12] 伏军胜, 贾小林, 刘家龙, 等. BDS-3 卫星与其他 GNSS 系统卫星原子钟性能分析[J]. 真空与低温, 2022, 28(5): 615-622.
FU J S, JIA X L, LIU J L, et al. Performance evaluation and comparison between spaceborne atomic clocks of BDS-3 satellites and other GNSS systems[J]. Vacuum and Cryogenics, 2022, 28(5): 615-622(in Chinese).

[13] MONTENBRUCK O, STEIGENBERGER P, HAUSCHILD A. Broadcast versus precise ephemerides: a multi-GNSS perspective [J]. GPS Solutions, 2015, 19(2): 321-333.

[14] 杜兰. GEO 卫星精密定轨技术研究[D]. 郑州: 解放军信息工程大学, 2006: 44-50.
DU L. Research on precise orbit determination technology of GEO satellite[D]. Zhengzhou: PLA Information Engineering University, 2006: 44-50(in Chinese).

[15] HAUSCHILD A, MONTENBRUCK O. Precise real-time navigation of LEO satellites using GNSS broadcast ephemerides[J]. Navigation, 2021, 68(2): 419-432.

[16] YANG J H, TANG C P, ZHOU S S, et al. High-accuracy clock offsets estimation strategy of BDS-3 using multi-source observations[J]. Remote Sensing, 2022, 14(18): 4674.

[17] 杨建华, 唐成盼, 宋叶志, 等. GNSS 导航电文空间信号测距误差分析[J]. 中国科学: 物理学 力学 天文学, 2021, 51(1): 72-84.
YANG J H, TANG C P, SONG Y Z, et al. Analysis of signal-in-

space ranging error of GNSS navigation message[J]. Scientia Sinica (Physica, Mechanica & Astronomica), 2021, 51(1): 72-84(in Chinese).

[18] YANG Y F, YANG Y X, HU X G, et al. BeiDou-3 broadcast clock estimation by integration of observations of regional tracking stations and inter-satellite links[J]. GPS Solutions, 2021, 25(2): 57.

In-orbit operation characteristics of BDS-3 spaceborne atomic clocks

WANG Wenbin[1,2], TANG Chengpan[1,*], YANG Jianhua[1,2], HU Xiaogong[1], CAO Yueling[1], ZHOU Shanshi[1]

(1. Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai 200030, China;
2. University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: In order to evaluate in-orbit operation characteristics of BeiDou-3 Navigation Satellite System (BDS-3) spaceborne atomic clocks, the broadcast clock offset parameters and precise clock offset data released by the International GNSS Service (IGS) Analysis Center were used to analyze the satellite's broadcast clock offsets, time-varying characteristics of broadcast clock speed, frequency accuracy, and frequency drift rate. The results indicate that the average error of clock offset parameters of the BDS-3 satellite is about 0.41 m, which is the key factor affecting the precision of the satellite space signal. Furthermore, BDS-3 broadcast clock speed parameters fluctuate frequently, so the spaceborne clocks' high frequency stability has not been brought into full play. The frequency accuracy of BDS-3 spaceborne clocks is poor, but that of the hydrogen clock of the geostationary orbit (GEO) satellite is high. The frequency drift rate of the BDS-3 spaceborne rubidium clock is larger, resulting in too many frequency modulation times and affecting system availability. BDS-3 hydrogen clock has a lower frequency drift rate and higher long-term stability, which indicates advantages in on-board autonomous time-keeping.

Keywords: broadcast clock offset; BDS-3 satellite; atomic clock; accuracy; drift rate

Received: 2023-05-23; **Accepted**: 2023-07-10; **Published Online**: 2023-08-04 11:53
URL: link.cnki.net/urlid/11.2625.V.20230803.1500.001

* **Corresponding author**. E-mail: cptang@shao.ac.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0251

基于轻量胶囊网络的自监督图像变化检测方法

张益天[1]，罗喜伶[1,2,*]，王宇鹏[1]

(1. 北京航空航天大学 电子信息工程学院，北京 100191；2. 北京航空航天大学杭州创新研究院，杭州 310051)

摘 要：针对散斑噪声对合成孔径雷达(SAR)图像变化检测精度影响大、现有基于胶囊网络的图像变化检测方法网络模型复杂度高、训练样本丢失大量原始图像信息等问题，提出了一种基于轻量胶囊网络的自监督图像变化检测方法。生成对数比值算子差异图，通过最大类间方差法和模糊C均值聚类算法，获得高置信度的训练样本“伪标签”，为实现自监督学习奠定基础；构造基于两时相SAR图像和对数比值算子差异图的三通道训练样本，最大限度保留样本信息；设计轻量胶囊网络，通过单尺度卷积提取训练样本特征，采用单尺度胶囊网络挖掘特征之间的空间关系；设置对比实验和消融实验，在5个真实SAR数据集上进行测试。实验结果表明：所提方法在降低模型复杂度的条件下，提高了运行效率，获得了更强的鲁棒性特征，抑制了散斑噪声对变化检测效果的不利影响，提升了变化检测效果。

关 键 词：变化检测；胶囊网络；最大类间方差法；模糊C均值聚类法；自监督学习
中图分类号：TP751
文献标志码：A **文章编号**：1001-5965(2025)05-1705-11

现代遥感技术的飞速发展为人类动态监测地面变化提供了便利条件。作为遥感技术领域的一个重要分支，合成孔径雷达(synthetic aperture radar, SAR)图像变化检测的任务是分析和处理同一地区、不同时刻获取的SAR图像，提取其中的变化信息[1-2]。

大量用于图像变化检测的多时相图像通常没有人工标记的标签，而传统的遥感图像标注主要依靠丰富的专家知识和专业经验，对获取的多时相遥感图像进行人眼分辨，其时间成本高、效率低，且易受到主观因素的影响。在样本标记方法的设计上，诸多学者基于模糊C均值聚类(fuzzy C-means clustering, FCM)算法进行改进。Gong等[3]通过相似算子得到2幅遥感图像中像素灰度的相似度，进而得到全局相似度阈值，以控制基于FCM的联合分类器实现样本标记。Gao等[4]利用Gabor小波特征和FCM算法，获得高精度的标记样本。Li等[5]以差异图的计算结果作为特征，结合显著性检测和FCM算法，分割出高置信度的变化区域和未变化区域。Wang等[6]通过稀疏表示将差异图上的像素信息转换为特征空间，再通过FCM算法从差异图中分割出高置信度训练样本。Zhang等[7]提出两阶段中心约束模糊C均值聚类(two-stage centre-constrained FCM, TCCFCM)算法，实现“伪标签”标记，为小区域变化检测问题提出解决方案。尽管上述方法获得的样本标签在模型训练上取得了良好的效果，但其标记结果完全依赖FCM算法的分割结果，未引入多种方法的样本标记结果，不利于构建更高置信度的训练样本集。针对上述问题，可通过多种方法获取多组样本标记结果，再采用图像融合方法，获得高置信度“伪标签”，以提升训练样本集的质量。

胶囊网络(capsule network, CapsNet)[8]是Sabour和Hinton等提出的一种网络模型。与卷积神经网络(convolutional neural network, CNN)[9]中的标量神

收稿日期：2023-05-16；**录用日期**：2023-07-28；**网络出版时间**：2023-08-30 17:34
网络出版地址：link.cnki.net/urlid/11.2625.V.20230830.1621.005
*通信作者. E-mail：luoxiling@buaa.edu.cn

引用格式：张益天，罗喜伶，王宇鹏. 基于轻量胶囊网络的自监督图像变化检测方法[J]. 北京航空航天大学学报，2025，51(5)：1705-1715. ZHANG Y T，LUO X L，WANG Y P. Self-supervised image change detection method based on lightweight capsule network[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1705-1715 (in Chinese).

经元不同，胶囊网络通过向量表示被描述对象的实例化参数，如姿态、纹理和变形，其特征表征更加鲁棒；同时，胶囊网络在胶囊中学习特征变量，最大化保留有价值的信息，因此，使用更少的训练数据就能推断出可能的结果，达到 CNN 的预期效果。针对图像变化检测这一具体任务而言，CNN 模型的缺点为：CNN 的池化操作无法耦合存在大量散斑噪声的 SAR 图像[10]，稀少的样本标签难以满足 CNN 模型的训练需要。从理论上比较而言，胶囊网络比 CNN 更适合用于变化检测任务。同时，事实证明，胶囊网络对 SAR 图像变化检测分析是有效的[11]。Gao 等[12]提出了多尺度胶囊网络（multiscale capsule network, MsCapsNet）用于 SAR 图像变化检测，以描述特征之间的空间关系，实现监督变化检测方法。Wang 等[6]提出了基于稀疏表示和胶囊网络的 SAR 图像自监督变化检测，通过 3 个卷积提取训练样本特征，采用单一尺度初级胶囊提取底层特征，取得了较好的变化检测效果。上述胶囊网络模型复杂度相对较高，执行效率较低。针对该问题，设计轻量胶囊网络模型，降低模型复杂度，以提高方法的效率，缩短检测用时。

自监督学习主要从大规模无监督数据中挖掘监督信息，通过所获取的“伪标签”训练深度学习网络，用于实现目标任务。因此，构造有效的训练样本集对变化检测结果至关重要。对于 CNN 模型而言，Gong[3] 和 Gao[4] 等从两时相 SAR 图像中以某一像素为中心取邻域像素，再将 2 个像素块直接级联或向量化后再级联，以构成训练样本。对于胶囊网络模型而言，Wang[6] 和 Gao[12] 等以某一像素为中心，从邻域比值算子差异图中取单通道像素块，再将其复用为三通道训练样本。上述训练样本构造方式可分为两大类：基于两时相 SAR 图像和基于差异图。尽管这 2 类方法构建出的训练样本集可训练出有效的变化检测方法，但其分别失去了另一种训练样本集的有效信息，影响变化检测效果的进一步提升。针对上述问题，采用基于两时相 SAR 图像和差异图的训练样本构造方式，将 2 类训练样本集的有效信息相结合，以提升变化检测方法的精度。

为了解决 SAR 图像的散斑噪声对变化检测精度影响大、现有基于胶囊网络的图像变化检测方法复杂度高、训练样本构造方式丢失大量信息等问题，本文提出了基于轻量胶囊网络的自监督图像变化检测方法（self-supervised learning lightweight capsule network, SLCapsNet）。首先，设计基于图像融合的图像变化检测标记方法，通过融合自适应阈值和特征聚类的标记结果，获得高置信度的训练样本“伪标签”；其次，构造基于两时相 SAR 图像和对数比值算子差异图的三通道训练样本；最后，设计轻量胶囊网络，经过训练样本集训练后，用于判断 2 幅 SAR 图像每一像素位置的变化检测结果。

1 基于图像融合的图像变化检测标记方法

基于图像融合的图像变化检测标记方法包括图像去噪预处理、差异图生成、自适应阈值分割、特征聚类、Logistic 映射、图像融合和训练样本构造等 7 个部分，实现了训练样本的“伪标签”标注，进而构建低成本、高置信度的训练样本集，用于后续网络的自监督训练，其具体流程如图 1 所示。

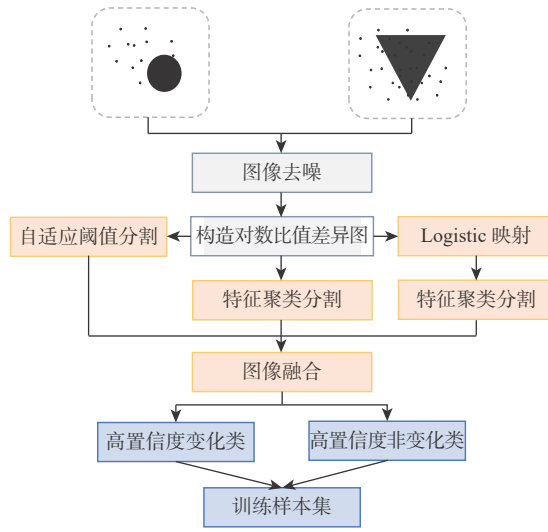


图 1 基于图像融合的图像变化检测标记方法流程

Fig. 1 Flow chart of image change detection and labeling method based on image fusion

1.1 图像去噪及差异图构造

SAR 图像成像机制会引入不可避免的散斑噪声[13]，为了降低 SAR 图像中的噪声污染，设计基于加权平均滤波器[7] 和高斯滤波器[14] 的联合滤波算法，对已配准、矫正后的 2 幅同一地区 t_1 和 t_2 时相图像 $\boldsymbol{I}_1=\{I_1(m,n),1\leqslant m\leqslant M,1\leqslant n\leqslant N\}$ 和 $\boldsymbol{I}_2=\{I_2(m,n),1\leqslant m\leqslant M,1\leqslant n\leqslant N\}$ 及其差异图 $\boldsymbol{D}$ 进行深度处理，其过程如图 2 所示。其中，M 和 N 分别表示 SAR 图像行和列的最大值，$I_1(m,n)$ 和 $I_2(m,n)$ 分别表示 SAR

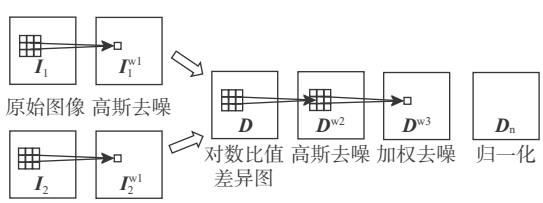


图 2 图像去噪及差异图构造过程

Fig. 2 Image denoising and difference graph construction process

图像$\boldsymbol{I}_1$和$\boldsymbol{I}_2$中(m,n)处的灰度值, (m,n)表示图像灰度矩阵的位置下标。

本文方法将图像去噪预处理和差异图生成过程融合在一起, 有效抑制了噪声, 增强了空域特征, 去除高频能量, 对图像进行平滑处理, 使得最佳阈值能够充分描绘出变化区域轮廓。

首先, 采用高斯滤波对原始图像进行处理。高斯滤波通过二维高斯核与图像进行卷积, 以消除高斯噪声对原始图像的影响, 属于线性平滑滤波器, 现已广泛应用于图像去噪。通过高斯滤波器$\boldsymbol{G}$对给定 2 幅 SAR 图像$\boldsymbol{I}_1$和$\boldsymbol{I}_2$进行卷积, 得到 2 幅池化图像$\boldsymbol{I}_1^{\mathrm{w1}}=\boldsymbol{I}_1*\boldsymbol{G}$和$\boldsymbol{I}_2^{\mathrm{w1}}=\boldsymbol{I}_2*\boldsymbol{G}$, 其中, “*”表示二维卷积运算。

然后, 采用对数比值算子[13]将噪声由乘性转为加性, 对数变换后的差值图像非线性缩小, 未变化区域和变化区域之间的对比度增强。对数比值算子构造的差异图$\boldsymbol{D}$如下:

$$\boldsymbol{D}=\left|\lg\left(\frac{\boldsymbol{I}_1^{\mathrm{w1}}+\varepsilon}{\boldsymbol{I}_2^{\mathrm{w1}}+\varepsilon}\right)\right| \tag{1}$$

式中: ε为非常小的常数。为了避免由于原始图像某一位置的像素值等于 0 而导致式 (1) 无意义的情况, 添加一个非常小的常数ε[15]。

其次, 通过高斯滤波器$\boldsymbol{G}$对差异图$\boldsymbol{D}$进行去噪, 获得差异图$\boldsymbol{D}^{\mathrm{w2}}$。

随后, 采用加权平均滤波器[7]对差异图$\boldsymbol{D}^{\mathrm{w2}}$进行深度处理, 进一步抑制在差异图以高亮的斑点形式存在的散斑噪声, 尽可能避免差异图图像灰度等级的突变。与其他滤波器相比, 加权平均滤波器在抑制散斑噪声的同时保持边缘, 并保持图像的连续性。通过相同加权平均滤波器核函数$\boldsymbol{W}^e$对差异图$\boldsymbol{D}^{\mathrm{w2}}$进行卷积, 得到池化图像$\boldsymbol{D}^{\mathrm{w3}}=\boldsymbol{D}^{\mathrm{w2}}*\boldsymbol{W}^e$。

最后, 使用归一化操作, 获得最终差异图$\boldsymbol{D}_{\mathrm{n}}$。

1.2　自适应阈值

通过对数比值算子生成的差异图与真实参考图像的关系密切, 其灰度值大的位置与真实参考图像的变化区域在形状和位置上具有正相关关系。因此, 采用自适应阈值算法, 通过最大类间方差法[16]实现自动化提取阈值, 从差异图中分割出高概率变化区域和未变化区域。

首先, 对差异图像$\boldsymbol{D}_{\mathrm{n}}$进行尺度放缩, 获得放缩后的图像$\boldsymbol{S}_{\mathrm{n}}$, 其灰度区间为$[1,2,\cdots,g]$。第$i$级灰度的像素数用$t_i$表示。某一灰度的出现概率如下:

$$T_i=\frac{t_i}{\sum_{i=1}^{g}t_i} \tag{2}$$

最后, 通过最大类间方差法确定概率阈值τ, 从而将像素分为变化类和未变化类, 作为生成第 1 组“伪标签”初步预测结果矩阵$\boldsymbol{L}_1$, 为下一步精细化确定边缘轮廓形状奠定基础。

1.3　特征聚类

由于自适应阈值算法在未提取特征或对特征进行分析的情况下直接对差异图进行分割, 未引入更多的判别信息, 导致无法更准确判别变化区域和未变化区域的边缘轮廓, 存在一定局限。针对上述问题, 采用特征聚类算法, 引入更多特征信息, 弥补自适应阈值结果的不足, 提升训练样本集的构造质量。

FCM 算法是一种基于划分的聚类算法, 其使得被划分为同一簇类的对象之间相似度最高, 而不同簇类之间的相似度最低。诸多学者[4,7,17]通过提取 Gabor 小波特征, 获得更鲁棒的 FCM 聚类标记结果。

基于上述思路, 本节采用基于 Gabor 小波特征的 FCM 算法, 生成第 2、3 组“伪标签”初步预测结果矩阵$\boldsymbol{L}_2$和$\boldsymbol{L}_3$, 其区别在于: 提取差异图的 Gabor 小波特征前, 是否跳过对差异图的 Logistic 映射处理环节。

首先, 通过 Gabor 小波变换[7], 提取差异图$\boldsymbol{D}_{\mathrm{n}}$的特征$\boldsymbol{x}=[\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{MN}]^{\mathrm{T}}$, 其中, $\boldsymbol{x}_j$表示特征矩阵, M和N分别表示差异图长度和宽度的像素数, j表示坐标位置(m,n), j的数值为$(m-1)*M+n$的运算结果, M表示原始图像中行的最大值。

然后, 以上述 Gabor 小波特征为 FCM 的输入$\boldsymbol{x}$, 通过迭代优化, 更新u_{dj}和$\boldsymbol{v}_d$直到式(3)收敛, 并基于u_{dj}中的最大值确定每个像素所属类别, 最终将差异图中的像素分为c个类别, 本节中c设置为 2。FCM 聚类的目标函数及约束条件的表达式如下:

$$\begin{cases} J(\boldsymbol{U},\boldsymbol{V})=\sum_{d=1}^{c}\sum_{j=1}^{MN}u_{dj}^{f}\left\|\boldsymbol{x}_j-\boldsymbol{v}_d\right\|^2 \\ \text{s.t.}\quad \sum_{k=1}^{c}u_{dj}=1, j=1,\cdots,MN \end{cases} \tag{3}$$

式中: $\boldsymbol{v}_d$为具有相似特征的一组聚类结果簇; $f\in[1,+\infty)$为控制 FCM 算法模糊程度的参数; $\boldsymbol{V}=[\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_c]$为聚类结果簇的向量矩阵; $u_{dj}\in[0,1]$为第j个像素属于第d个聚类结果簇$\boldsymbol{v}_d$的隶属度; $\boldsymbol{U}=\left[u_{dj}^{f}\right]_{c\times MN}$为$u_{dj}$的模糊划分矩阵; $\|\cdot\|^2$表示输入特征向量$\boldsymbol{x}_j$与第d个聚类结果簇$\boldsymbol{v}_d$之间的差值。

其次, 将聚类结果簇中平均像素值最大的区域设置为变化区域, 其他区域设置为未变化区域。若第j个像素属于变化区域, 则第j个像素的第 2 组“伪标签”初步预测结果L_2^j设置为 1; 反之, 则将L_2^j设置为 0。第j个像素的第 2 组“伪标签”初步预

测结果L_2^j表达式如下：

$$L_2^j=\begin{cases}1 & \underset{d=1,2,\cdots,c}{\operatorname{argmax}}\boldsymbol{M}_{\Omega_d}\\ 0 & \text{其他}\end{cases} \tag{4}$$

式中：Ω_d为第d个聚类结果簇；$\boldsymbol{M}_{\Omega_d}$为第$d$个聚类结果簇$\Omega$中差异图像素的平均值。

最后，通过采用联通分量标记（connection component labeling, CCL）算法[18]，对第2组“伪标签”初步预测结果L_2^j构成的矩阵$\boldsymbol{L}_2$进行微调，将联通分量中所含像素数目较低的变化区域视为置信度较低的变化区域，并将其转为未变化区域，为构建更佳鲁棒的训练样本奠定基础。

更进一步，$\boldsymbol{L}_3$的生成过程如下：

首先，通过Logistic函数进行映射logistic($\boldsymbol{D}_{\mathrm{n}}$,$\alpha$)，生成平滑处理后的差异图。

然后，提取平滑处理后的差异图的Gabor小波特征，通过FCM算法生成第3组“伪标签”初步预测结果矩阵$\boldsymbol{L}_3$。

最后，通过联通分量标记算法，对$\boldsymbol{L}_3$进行微调。

1.4 训练样本的构造

构造基于两时相SAR图像和对数比值差异图的三通道训练样本，最大限度保留训练样本信息，有效提升变化检测精度。其构造过程如下：

首先，采用图像融合方法，对“伪标签”初步标记矩阵$\boldsymbol{L}_1$、$\boldsymbol{L}_2$和$\boldsymbol{L}_3$进行处理，若第j个像素在$\boldsymbol{L}_1$和$\boldsymbol{L}_2$标记结果同为变化类像素，则将其所对应像素位置的高置信度“伪标签”预测结果Y_{D}^j设置为1；若第j个像素在$\boldsymbol{L}_1$和$\boldsymbol{L}_2$标记结果不同为变化类像素且在$\boldsymbol{L}_3$标记结果为未变化类像素，则将其所对应像素位置的高置信度“伪标签”预测结果Y_{D}^j设置为0；若第j个像素不属于上述两种情况，则将其标记为不确定类，其所对应像素位置的高置信度“伪标签”预测结果Y_{D}^j设置为0.5，以获得高置信度“伪标签”预测结果矩阵$\boldsymbol{Y}_{\mathrm{D}}$，其长为$M$，列为$N$。上述过程如下：

$$Y_{\mathrm{D}}^j=\begin{cases}0, & \left(L_1^jL_2^j\right)^2+\left(L_3^j\right)^2=0\\ 1, & L_1^jL_2^j=1\\ 0.5, & \text{其他}\end{cases} \tag{5}$$

式中：L_i^j为第i个“伪标签”初步标记矩阵中第j个像素的二值化标记结果；$\boldsymbol{Y}_{\mathrm{D}}$为由高置信度“伪标签”预测结果$Y_{\mathrm{D}}^j$构成的矩阵。

然后，提取高置信度“伪标签”预测结果矩阵$\boldsymbol{Y}_{\mathrm{D}}$中标记为0和1的像素位置，即未变化区域和变化区域，以像素j为中心，在SAR图像I_1、I_2和差异图$\boldsymbol{D}_{\mathrm{n}}$中取等数目、大小为$k\times k$的邻域像素块$p_j^1$、$p_j^2$和$p_j^3$，将其级联成三通道样本块，其中，$k$为奇数，示意图如图3所示。

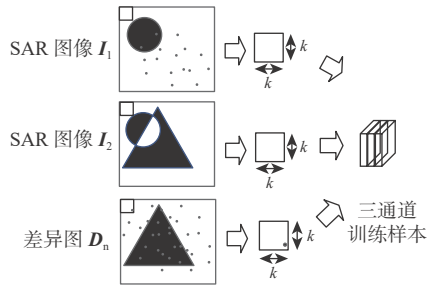


图3 基于两时相SAR图像和差异图的训练样本构造方式示意图

Fig. 3 Training sample construction based on two temporal SAR images and difference maps

其次，通过生成虚拟样本，解决深度学习模型在训练过程中需要大量样本的问题，弥补在图像变化检测中可用样本数量有限的缺陷。一个虚拟样本P'由相同类别的2个样本块以适当的比例生成，如下：

$$P'=\alpha P_1+(1-\alpha)P_2 \tag{6}$$

式中：P_1和P_2分别为来自相同类别的2个训练样本块；α为区间[0,1]上的均匀分布随机值。Gao等[19]在图像变化检测任务中采用了类似的虚拟样本生成方法，结果表明，在训练样本有限的情况下，虚拟样本分类模型的性能优于无虚拟样本分类模型。由于虚拟样本是由同一类的2个样本生成，为虚拟样本P'设置相同的类别标签。

最后，通过随机选取2类真实样本，若某类真实样本数不足500，则使用虚拟样本补足空缺样本，最终构成高置信度训练样本集，其变化类和未变化类样本各500个。

2 基于轻量胶囊网络的图像变化检测方法

轻量胶囊网络通过单尺度卷积提取训练样本特征，采用单尺度胶囊网络挖掘特征之间的空间关系，降低模型复杂度，缩短变换检测用时，抑制散斑噪声对图像变化检测效果的不利影响。轻量胶囊网络由初级胶囊层、卷积-胶囊层和类胶囊层组成，其模型结构如图4所示。

初级胶囊层通过类卷积运算，提取特征向量。轻量胶囊网络基于核数为8、大小为3×3的卷积从输入向量中提取特征，采用类卷积运算和squashing函数得到输出胶囊$\boldsymbol{o}$，squashing函数如下：

$$\boldsymbol{o}=\frac{\|\boldsymbol{s}\|^2}{1+\|\boldsymbol{s}\|^2}\cdot\frac{\boldsymbol{s}}{\|\boldsymbol{s}\|} \tag{7}$$

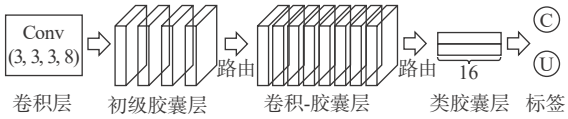


图 4　轻量胶囊网络模型结构

Fig. 4　Lightweight capsule network model structure

式中: $\boldsymbol{s}$为输入向量; $\boldsymbol{o}$为胶囊输出向量。轻量胶囊网络在初级胶囊层中获得了 4 个 8D 胶囊。

卷积-胶囊层利用局部连接和共享变换矩阵在一定程度上减少了参数的数量[20], 采用动态路由策略[8]更新耦合系数。

类胶囊层可以视作全连接层, 设置了 2 个 16D 胶囊, 分别代表不同类别。同样采用动态路由策略[8]更新耦合系数, 通过损失函数进行其他参数迭代更新, 其损失函数如下:

$$F = T_l \max\left(0, m^+ - \|\boldsymbol{o}_l\|\right)^2 + \lambda(1-T_l)\max\left(0, \|\boldsymbol{o}_l\| - m^-\right)^2 \tag{8}$$

式中, $m^+ = 0.9$, $m^- = 0.1$, $\lambda = 0.5$。l表示分类类别: 当l为 0 时, 表示未变化类; 当l为 1 时, 表示变化类。当l存在时, T_l为 1; 否则, T_l为 0。$\boldsymbol{o}_l$是类别l的输出向量。通过逐像素进行分类, 计算得到最终的变化检测结果。

3　实验设置与结果分析

硬件测试平台是处理器为 Intel(R) Core(TM) i7-6700K、主频为 4.00 GHz、内存为 32.0 GB 的台式机, 软件测试平台是 Windows 操作系统、PyCharm 和 MATLAB。

3.1　客观评价指标

为了检验变化检测方法的性能, 采用错检数 V_{FA}、漏检数 V_{MA}、分类正确率 V_{PCC}、Kappa 系数 V_{KC}、F1 分数 V_{F1}和用时作为客观评价指标。上述指标中的前 5 个指标基于混淆矩阵, 混淆矩阵参数含义说明如下: V_{TP}为地面真实情况与检测结果同为变化像素的数目; V_{FP}为地面真实情况与检测结果同为未变化像素的数目; V_{TN}为地面真实结果未变化而检测结果变化的像素数目; V_{FN}为地面真实结果变化而检测结果未变化的像素数目。

错检数V_{FA}是指未变化区域被错检为变化区域的个数, 其表达式为

$$V_{\mathrm{FA}} = V_{\mathrm{TN}} \tag{9}$$

漏检数V_{MA}是指变化区域被漏检为未变化区域的个数, 其表达式为

$$V_{\mathrm{MA}} = V_{\mathrm{FN}} \tag{10}$$

分类正确率V_{PCC}是指在总像素点中变化检测结果与实际结果相一致的百分比, 其表达式为

$$V_{\mathrm{PCC}} = \frac{V_{\mathrm{TP}} + V_{\mathrm{TN}}}{V_{\mathrm{TP}} + V_{\mathrm{TN}} + V_{\mathrm{FP}} + V_{\mathrm{FN}}} \times 100\% \tag{11}$$

然而, V_{PCC}并不能完全反映事实情况, 一般情况下, 由于$V_{\mathrm{TP}} + V_{\mathrm{TN}} + V_{\mathrm{FP}} + V_{\mathrm{FN}}$往往比较大, 不同变化检测方法的分类正确率非常相似, 通过分类正确率很难观察到具体的差异。

Kappa 系数V_{KC}是用于一致性检验的指标, 也用于衡量不同类别检测效果, 其结果介于 0～1 之间。V_{KC}越接近于 1, 则表示二值结果图和参考图的一致程度越高, 整体效果更优, 变化检测性能越好。其表达式为

$$V_{\mathrm{KC}} = \frac{V_{\mathrm{PCC}} - V_{\mathrm{PRE}}}{1 - V_{\mathrm{PRE}}} \tag{12}$$

式中:

$$V_{\mathrm{PRE}} = \frac{(V_{\mathrm{TP}}+V_{\mathrm{FP}})(V_{\mathrm{TP}}+V_{\mathrm{FN}})+(V_{\mathrm{TN}}+V_{\mathrm{FP}})(V_{\mathrm{TN}}+V_{\mathrm{FN}})}{(V_{\mathrm{TP}} + V_{\mathrm{TN}} + V_{\mathrm{FP}} + V_{\mathrm{FN}})^2} \tag{13}$$

F1 分数V_{F1}是用于衡量二分类问题精确度的综合指标, 平衡了分类问题的准确率$V_{\mathrm{precision}}$和召回率V_{recall}, 其表达式为

$$V_{\mathrm{F1}} = \frac{2V_{\mathrm{precision}}V_{\mathrm{recall}}}{V_{\mathrm{precision}} + V_{\mathrm{recall}}} \tag{14}$$

式中:

$$V_{\mathrm{precision}} = \frac{V_{\mathrm{TP}}}{V_{\mathrm{TP}} + V_{\mathrm{FP}}} \tag{15}$$

$$V_{\mathrm{recall}} = \frac{V_{\mathrm{TP}}}{V_{\mathrm{TP}} + V_{\mathrm{FN}}} \tag{16}$$

3.2　数据集

为了检验变化检测方法的性能, 选择旧金山[4]、黄河海岸线[3]、黄河内陆水域[3]、黄河农田[3]和贵州[21]等 5 个真实 SAR 数据集, 用于对比实验和消融实验, 如表 1 所示。

表 1　数据集信息

Table 1　Information for datasets

数据集名称	场景	第1时相	第2时相
旧金山	洪水突发	2003年8月	2004年5月
黄河海岸线	河水冲击	2008年6月	2009年6月
黄河内陆水域	河水冲击	2008年6月	2009年6月
黄河农田	农田耕种	2008年6月	2009年6月
贵州	建筑物变化	2016年6月	2017年4月

3.3　参数设置

分类正确率随邻域大小的变化曲线如图 5 所示。若邻域设置较小, 其包含图像信息较少, 难以提取鲁棒特征; 而邻域设置较大, 则引入过多冗余信息, 导致精度下降。当邻域大小为 9 时, 在旧金山、黄河内陆水域、黄河农田和贵州数据集上, 分

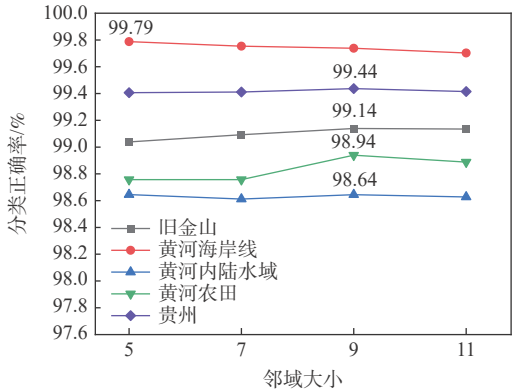


图5 分类正确率随邻域大小的变化曲线

Fig. 5 Curves of correct classification rate with neighborhood size

类正确率最大。由于SLCapsNet出色的性能，较小的邻域块已经足够区分黄河海岸线数据集的变化类和未变化类样本。随着邻域大小的扩大，反而引入干扰信息。因此，当邻域大小为5时，在黄河海岸线数据集上分类正确率最大。

综上，为了最大限度地发挥SLCapsNet性能，对于黄河海岸线数据集，设置邻域大小为5；而对于其他数据集，设置邻域大小为9。

3.4 不同方法的对比实验

设置对比实验，将本文方法SLCapsNet与多种相关的变化检测方法即基于主成分分析和k均值聚类算法的变化检测方法(PCAK)[22]、基于PCANet的变化检测方法(PCANet)[4]、基于卷积小波神经网络的变化检测方法(CWNN)[19]、基于深度学习的SAR图像鲁棒无监督小区域变化检测(TFCWNN)[7]、基于多尺度胶囊网络的变化检测方法(MsCapsNet)[12]进行比较，其中，第1种为传统图像变化检测方法，后4种均为基于深度学习的图像变化检测方法。由表2～表6可知，SLCapsNet显著优于其他方法。在旧金山、黄河海岸线、黄河农田水域和贵州数据集上，SLCapsNet的分类正确率、F1分数和Kappa系数均取得最佳结果。

在旧金山数据集上的实验结果如图6和表2所示，PCANet、CWNN、TFCWNN和MsCapsNet的结果中存在白斑噪声。TFCWNN获得最低的错检数，SLCapsNet取得最佳的漏检数。与MsCapsNet相比，SLCapsNet在F1分数上提升3.39%，在Kappa系数上提升3.65%，在用时上降低56.44%。

表2 对旧金山数据集的客观评价

Table 2 Objective evaluation of San Francisco Dataset

方法	V_{FA}	V_{MA}	V_{PCC}/%	V_{F1}/%	V_{KC}/%	用时/s
PCAK[22]	134	989	98.29	89.02	88.09	0.2
PCANet[4]	359	479	98.72	91.17	90.48	632.9
CWNN[19]	157	909	98.37	89.47	88.59	580.3
TFCWNN[7]	126	713	98.72	91.57	90.88	369.9
MsCapsNet[12]	416	475	98.64	90.55	89.82	429.5
SLCapsNet	317	247	99.14	93.94	93.47	187.1

表3 对黄河海岸线数据集的客观评价

Table 3 Objective evaluation of Yellow River Coastline Dataset

方法	V_{FA}	V_{MA}	V_{PCC}/%	V_{F1}/%	V_{KC}/%	用时/s
PCAK[22]	0	38 217	69.67	6.59	4.62	0.5
PCANet[4]	4	28 335	77.51	8.66	6.76	2 491.7
CWNN[19]	51	20 258	83.88	11.33	9.50	618.0
TFCWNN[7]	137	152	99.77	89.34	89.22	492.4
MsCapsNet[12]	204	102	99.76	88.20	88.08	177.7
SLCapsNet	171	96	99.79	89.81	89.71	84.1

表4 对黄河内陆水域数据集的客观评价

Table 4 Objective evaluation of Yellow River Inland Water Area Dataset

方法	V_{FA}	V_{MA}	V_{PCC}/%	V_{F1}/%	V_{KC}/%	用时/s
PCAK[22]	601	1 406	98.45	78.45	77.65	0.3
PCANet[4]	1 155	847	98.45	75.59	74.79	1 156.6
CWNN[19]	471	2 065	98.08	74.90	73.91	613.2
TFCWNN[7]	854	702	98.80	81.38	80.76	446.3
MsCapsNet[12]	815	711	98.82	81.85	81.24	537.3
SLCapsNet	729	1 022	98.64	80.11	79.41	235.9

表5 对黄河农田数据集的客观评价

Table 5 Objective evaluation of Yellow River Farmland Dataset

方法	V_{FA}	V_{MA}	V_{PCC}/%	V_{F1}/%	V_{KC}/%	用时/s
PCAK[22]	490	1 279	98.01	84.39	83.33	0.2
PCANet[4]	1 345	201	98.26	83.54	82.64	994.2
CWNN[19]	734	969	98.09	84.19	83.18	590.8
TFCWNN[7]	811	213	98.85	89.70	89.09	446.3
MsCapsNet[12]	620	383	98.87	90.27	89.67	468.9
SLCapsNet	771	173	98.94	90.50	89.95	204.4

表6 对贵州数据集的客观评价

Table 6 Objective evaluation of Guizhou Dataset

方法	V_{FA}	V_{MA}	V_{PCC}/%	V_{F1}/%	V_{KC}/%	用时/s
PCAK[22]	99	38 768	75.71	6.69	4.98	0.7
PCANet[4]	106	31 990	79.94	7.95	6.28	2 717.5
CWNN[19]	383	19 069	87.84	10.24	8.65	642.1
TFCWNN[7]	324	701	99.36	69.50	69.18	225.0
MsCapsNet[12]	1 334	118	99.09	17.87	17.63	594.3
SLCapsNet	450	451	99.44	69.82	69.53	248.6

在黄河海岸线数据集上的实验结果如图7和表3所示。PCAK、PCANet和CWNN的变化检测结果中含有大量的漏检像素，TFCWNN、MsCapsNet、

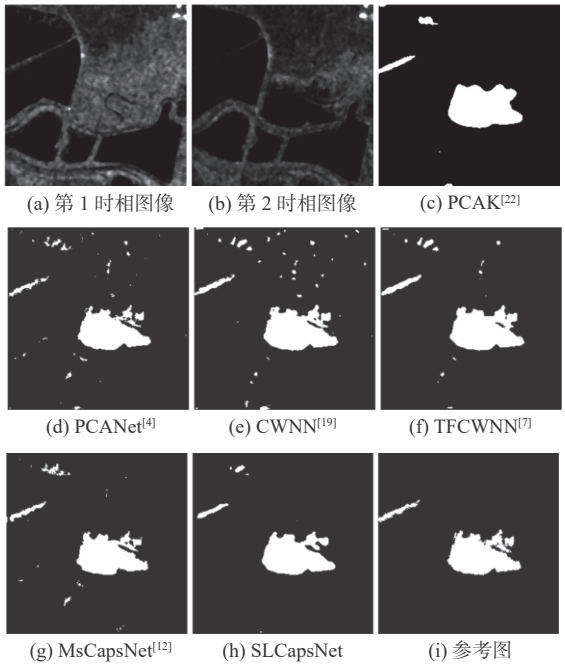


图 6　旧金山数据集中各种变化检测方法的可视化结果

Fig. 6　Visualization results of various change detection methods in San Francisco Dataset

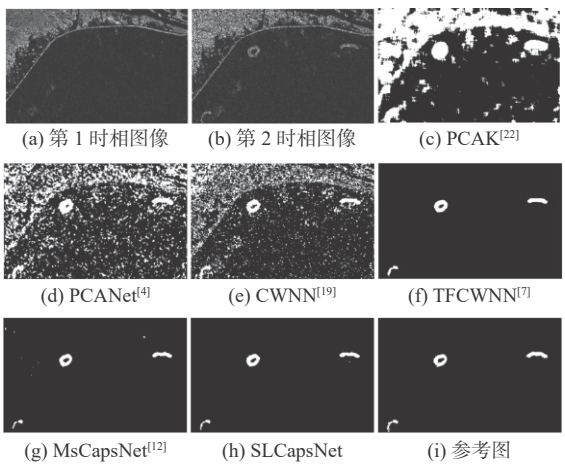


图 7　黄河海岸线数据集中各种变化检测方法的可视化结果

Fig. 7　Visualization results of various change detection methods in Yellow River Coastline Dataset

SLCapsNet 均取得了良好主观视觉效果。依据客观评价指标, 相比于 TFCWNN、MsCapsNet, SLCapsNet 在 Kappa 系数上分别提升 0.49% 和 1.63%, 在 F1 分数上分别提升 0.47% 和 1.61%, 在用时上分别降低 82.92% 和 52.67%。

在黄河内陆水域数据集上的实验结果如图 8 和表 4 所示。TFCWNN 和 MsCapsNet 取得了良好效果, 其 Kappa 系数突破 80%, 但上述方法用时均较长。与 SLCapsNet 相比, 上述 2 种方法在 Kappa 系数上分别提升了 1.35% 和 1.83%, 在 F1 分数上分别提升了 1.27% 和 1.74%, 但在用时上提升了 89.19% 和 127.77%。尽管 SLCapsNet 的变化检测精度未取得最佳效果, 但其用时相对较短。

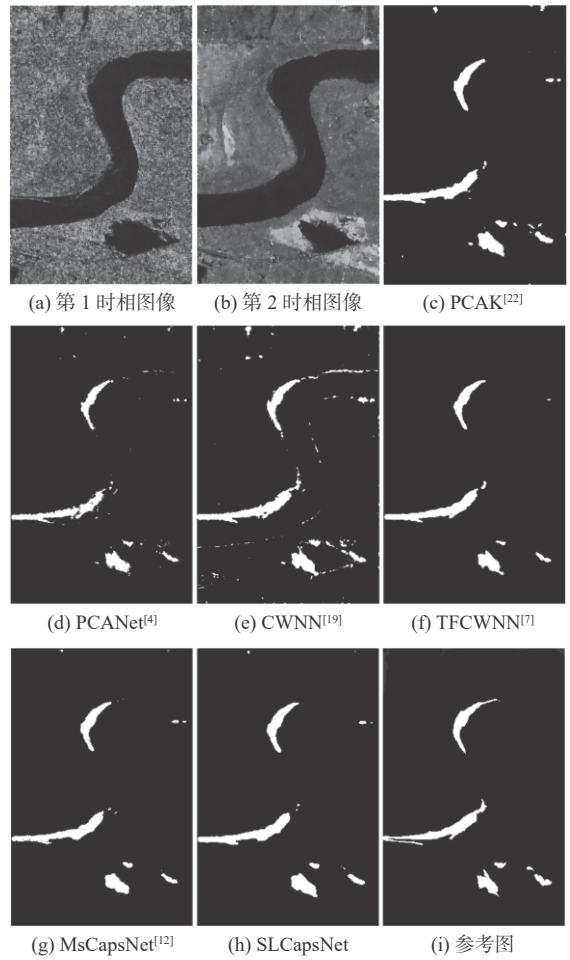


图 8　黄河内陆水域数据集中各种变化检测方法的可视化结果

Fig. 8　Visualization results of various change detection methods in Yellow River Inland Water Area Dataset

在黄河农田数据集上的实验结果如图 9 和表 5 所示。PCAK、PCANet 和 CWNN 方法的错检数和漏检数较大, TFCWNN、MsCapsNet 和 SLCapsNet 方法的 Kappa 系数和 F1 分数精度相当。从用时上看, SLCapsNet 方法相比于其他深度学习方法至少降低 54.20%。

在贵州数据集上的实验结果如图 10 和表 6 所示。PCAK、PCANet、CWNN 方法的检测结果中含有大量的漏检像素, TFCWNN 方法的检测结果中含有大量的错检像素, 而 TFCWNN 和 SLCapsNet 方法的漏检和错检像素之和均相对较低。相比于 TFCWNN, SLCapsNet 在 Kappa 系数和 F1 分数上具有一定优势, 但在用时上增加了 10.49%。

结果表明, 使用 SLCapsNet 方法处理含有大量散斑噪声的 SAR 图像时, 其检测精度较高, 用时较短, 其性能明显优于 MsCapsNet。

3.5　不同训练样本的对比实验

为探究不同训练样本对变化检测精度的影响,

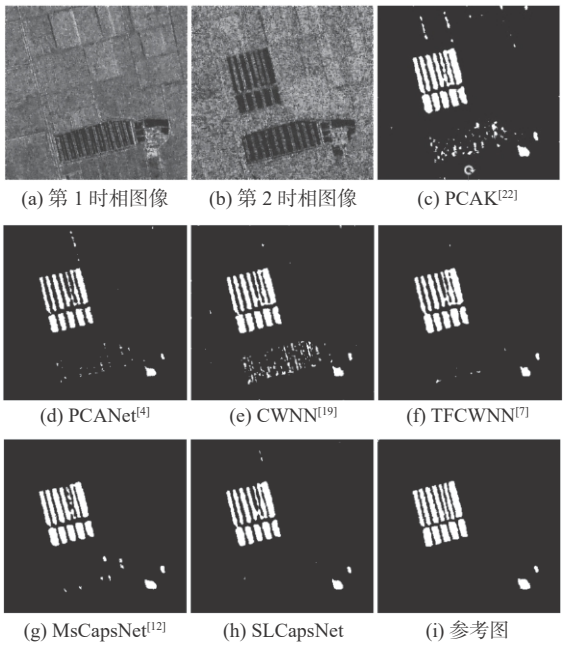


图9 黄河农田数据集中各种变化检测方法的可视化结果

Fig. 9 Visualization results of various change detection methods in Yellow River Farmland Dataset

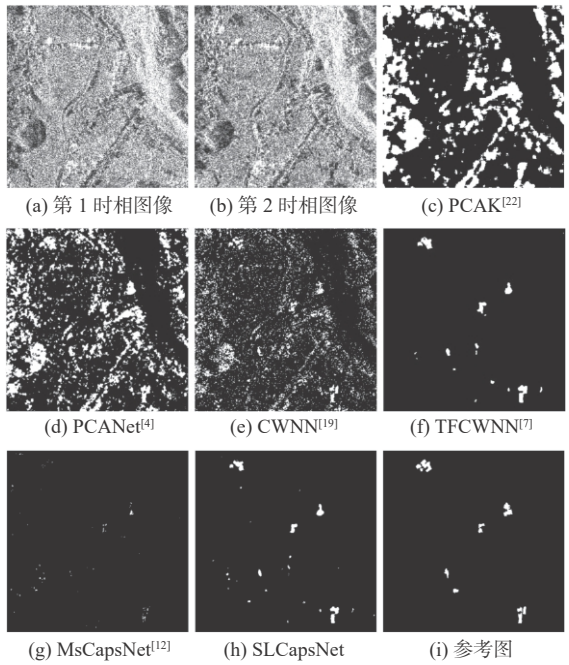


图10 贵州数据集中各种变化检测方法的可视化结果

Fig. 10 Visualization results of various change detection methods in Guizhou Dataset

采用多种训练样本构造方式,得到不同训练样本的Kappa系数和F1分数柱状图,如图11～图14所示。本节构造的训练样本共分为4种:第1种为基于两时相SAR图像和对数比值算子差异图的三通道训练样本(im1im2log),第2种为基于两时相SAR图像和邻域比值算子差异图的三通道训练样本(im1im2nr),第3种为基于对数比值算子差异图的三通道训练样本(logloglog),第4种为基于邻域比值算子差异图的三通道训练样本(nrnrnr)。Gao[12]和Wang[6]等均采用第4种训练样本。本节实验选择的训练样本数为1 000,其中,变化类样本和未变化类样本各500,若某类样本数不足500,则通过虚拟样本补足。

由图11～图13可知,在黄河内陆水域、黄河农田和贵州数据集上,当设置邻域大小为5、7、9和11时,im1im2log的Kappa系数和F1分数取得最佳效果,而nrnrnr取得最差效果。

图14的柱状图基于图11～图13,计算多个数据集在取不同领域大小时Kappa系数和F1分数的平均值。相较于其他3种训练样本,im1im2log在Kappa系数上分别平均提升0.75%、2.77%、4.19%,在F1分数上分别平均提升0.70%、2.71%、4.07%。

综上,基于两时相SAR图像和对数比值算子差异图的三通道训练样本(im1im2log)对SLCapsNet的变化检测精度具有一定的提升作用。

3.6 消融实验

完整轻量胶囊网络方法由SLCapsNet模型、基于图像融合的“伪标签”标记、基于两时相SAR图

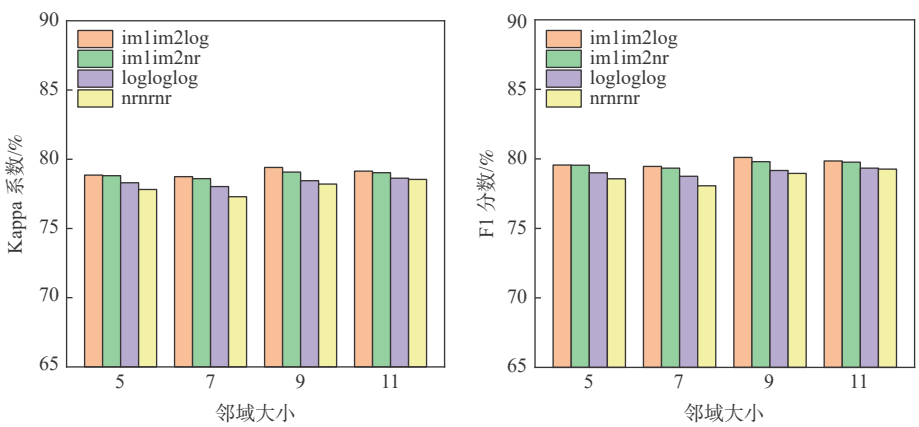


图11 黄河内陆水域数据集上不同训练样本构造方式的Kappa系数和F1分数柱状图

Fig. 11 Histograms of Kappa coefficients and F1 scores for different training sampl construction methods on Yellow River Inland Water Area Dataset

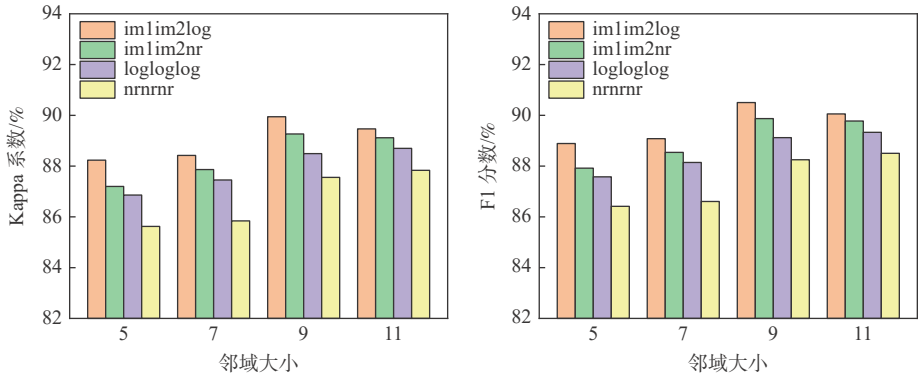


图 12　黄河农田数据集上不同训练样本构造方式的 Kappa 系数和 F1 分数柱状图

Fig. 12　Histograms of Kappa coefficients and F1 scores for different training sample construction methods on Yellow River Farmland Dataset

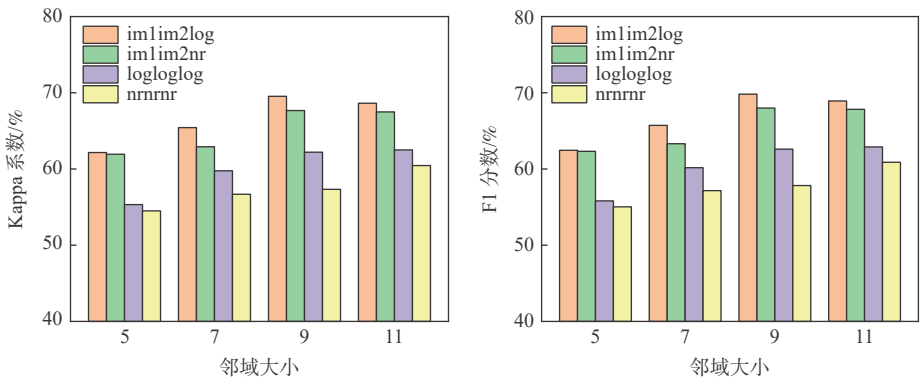


图 13　贵州数据集上不同训练样本构造方式的 Kappa 系数和 F1 分数柱状图

Fig. 13　Histograms of Kappa coefficients and F1 scores for different training sample construction methods on Guizhou Dataset

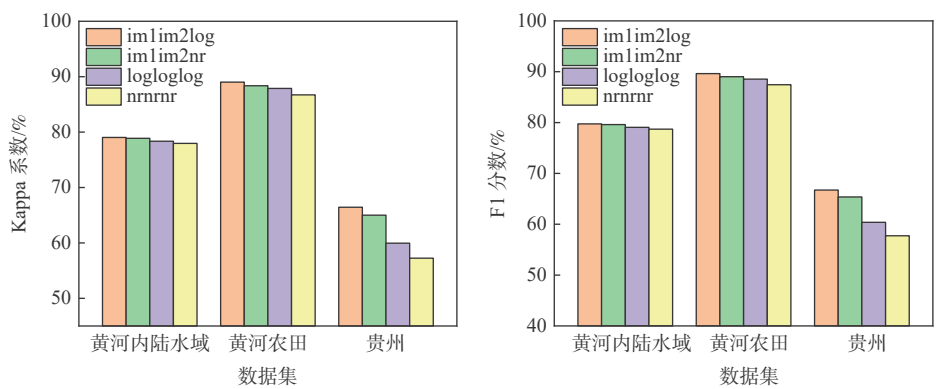


图 14　多个数据集上不同训练样本构造方式的 Kappa 系数和 F1 分数柱状图

Fig. 14　Histograms of Kappa coefficients and F1 scores for different training sample construction methods on multiple datasets

像和对数比值算子差异图的三通道训练样本构建等 3 部分构成。本节消融实验采用不同尺度的胶囊, 验证对变化检测用时的影响; 根据是否采用基于图像融合的“伪标签”标记、是否采用基于两时相 SAR 图像和对数比值算子差异图的三通道训练样本构造方式, 构成 4 种变化检测方法, 验证对变化检测精度的影响。其中, 采用“伪标签”标记的方法属于自监督方法, 不采用“伪标签”标记的方法属于监督方法, 其训练样本的标签直接来自于参考图。监督和自监督方法选择的训练样本数均为 1 000, 其中, 变化类样本和未变化类样本各 500, 若某类样本数不足 500, 则使用虚拟样本补足。消融实验所用的方法及其客观测试结果如表 7 所示。

由表 7 可知, 在黄河海岸线、黄河农田和贵州数据集上, 采用单一尺度胶囊构建轻量胶囊网络模型, 相比于多尺度胶囊构建的 MsCapsNet 模型, 用时平均降低了 56.72%; 通过使用基于图像融合的“伪标签”标记, 方法 3 相比于方法 2 在 Kappa 系数上分别提升 14.79%、3.63% 和 20.84%, 方法 5 相比于方法 4 在 Kappa 系数上分别提升 14.99%、3.11%

和31.72%；通过构造基于两时相SAR图像和对数比值算子差异图的三通道训练样本，方法4相比于方法2在Kappa系数上分别提升0.11%、1.20%和1.32%，方法5相比于方法3在Kappa系数上分别提升0.31%、0.68%和12.20%；通过使用基于图像融合的“伪标签”标记并构造基于两时相SAR图像和对数比值算子差异图的三通道训练样本，方法5相比于方法1在Kappa系数上分别提高1.63%、0.28%和51.90%，方法5相比于方法2在Kappa系数上分别提升15.10%、4.31%和33.04%。

由此可见，与MsCapsNet模型相比，轻量胶囊网络模型降低了变化检测用时，提高方法运行效率；基于图像融合的“伪标签”标记、基于两时相SAR图像和对数比值算子差异图的三通道训练样本均可提升变化检测的精度。

表7 SLCapsNet的消融实验

Table 7 Ablation experiment results of SLCapsNet

方法	模型	伪标签	样本构造方式	平均用时/s	Kappa系数/%		
					黄河海岸线	黄河农田	贵州
1	MsCapsNet	×	nrnrnr	413.6	88.08	89.67	17.63
2	SLCapsNet	×	nrnrnr	179.0	74.61	85.64	36.49
3	SLCapsNet	√	nrnrnr	179.0	89.40	89.27	57.33
4	SLCapsNet	×	im1im2log	179.0	74.72	86.84	37.81
5	SLCapsNet	√	im1im2log	179.0	89.71	89.95	69.53

4 结 论

针对散斑噪声对SAR图像变化检测精度影响大、现有基于胶囊网络的图像变化检测方法网络模型复杂度高、训练样本丢失大量原始图像信息等问题，采用自监督学习方法，提出了基于轻量胶囊网络的图像变化检测方法SLCapsNet，结论如下：

1）提出的基于图像融合的样本标记方法，可获得有效且鲁棒的训练样本，提升了变化检测精度。

2）提出了SLCapsNet方法，降低了模型复杂度，可减少方法用时，提高方法运行效率。

3）基于两时相SAR图像和对数比值差异图构造三通道训练样本，可有效提升SLCapsNet变化检测精度。

参考文献（References）

[1] 刘国强, 房胜, 李哲. 用于遥感图像变化检测的全尺度特征聚合网络[J]. 北京航空航天大学学报, 2022, 48(8): 1464-1470.
LIU G Q, FANG S, LI Z. A full-scale feature aggregation network for remote sensing image change detection[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(8): 1464-1470 (in Chinese).

[2] 周圆, 李祥瑞, 杨晶. 基于混合网络的异源遥感图像变化检测[J]. 北京航空航天大学学报, 2021, 47(3): 451-460.
ZHOU Y, LI X R, YANG J. Heterogeneous remote sensing image change detection based on hybrid network[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(3): 451-460 (in Chinese).

[3] GONG M G, ZHAO J J, LIU J, et al. Change detection in synthetic aperture radar images based on deep neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 27(1): 125-138.

[4] GAO F, DONG J Y, LI B, et al. Automatic change detection in synthetic aperture radar images based on PCANet[J]. IEEE Geoscience and Remote Sensing Letters, 2016, 13(12): 1792-1796.

[5] LI M K, LI M, ZHANG P, et al. SAR image change detection using PCANet guided by saliency detection[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(3): 402-406.

[6] WANG S N, WANG Y N, LIU Y, et al. SAR image change detection based on sparse representation and a capsule network[J]. Remote Sensing Letters, 2021, 12(9): 890-899.

[7] ZHANG X Z, SU H, ZHANG C, et al. Robust unsupervised small area change detection from SAR imagery using deep learning[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 173: 79-94.

[8] SABOUR S, FROSST N, HINTON G E, et al. Dynamic routing between capsules[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 3859-3869.

[9] GONG M G, YANG H L, ZHANG P Z. Feature learning and change feature classification based on deep learning for ternary change detection in SAR images[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2017, 129: 212-225.

[10] FERRETTI A, MONTI-GUARNIERI A, PRATI C, et al. InSAR principles: guidelines for SAR interferometry processing and interpretation[M]. Noordwijk: ESA Publications, 2007.

[11] YU S Q, JIA S, XU C Y. Convolutional neural networks for hyperspectral image classification[J]. Neurocomputing, 2017, 219: 88-98.

[12] GAO Y H, GAO F, DONG J Y, et al. SAR image change detection based on multiscale capsule network[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 18(3): 484-488.

[13] BOVOLO F, BRUZZONE L. A detail-preserving scale-driven approach to change detection in multitemporal SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(12): 2963-2972.

[14] D'HAEYER J P F. Gaussian filtering of images: a regularization approach[J]. Signal Processing, 1989, 18(2): 169-181.

[15] ZHENG Y G, ZHANG X R, HOU B, et al. Using combined difference image and K-means clustering for SAR image change detection[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(3): 691-695.

[16] OTSU N. A threshold selection method from gray-level histograms [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62-66.

[17] LI H C, CELIK T, LONGBOTHAM N, et al. Gabor feature based unsupervised change detection of multitemporal SAR images based on two-level clustering[J]. IEEE Geoscience and Remote Sensing

Letters, 2015, 12(12): 2458-2462.

[18] ROSENFELD A, PFALTZ J L. Sequential operations in digital picture processing[J]. Journal of the ACM, 1966, 13(4): 471-494.

[19] GAO F, WANG X, GAO Y H, et al. Sea ice change detection in SAR images based on convolutional-wavelet neural networks [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(8): 1240-1244.

[20] ZHU K Q, CHEN Y S, GHAMISI P, et al. Deep convolutional capsule network for hyperspectral image spectral and spectral-spatial classification[J]. Remote Sensing, 2019, 11(3): 223.

[21] ZHANG X Z, LIU G, ZHANG C, et al. Two-phase object-based deep learning for multi-temporal SAR image change detection[J]. Remote Sensing, 2020, 12(3): 548.

[22] CELIK T. Unsupervised change detection in satellite images using principal component analysis and K-means clustering[J]. IEEE Geoscience and Remote Sensing Letters, 2009, 6(4): 772-776.

Self-supervised image change detection method based on lightweight capsule network

ZHANG Yitian[1]，LUO Xiling[1, 2, *]，WANG Yupeng[1]

(1. School of Electronic Information Engineering，Beihang University，Beijing 100191，China；
2. Hangzhou Innovation Institute of Beihang University，Hangzhou 310051，China)

Abstract：In response to the significant impact of speckle noise on the detection accuracy of synthetic aperture radar (SAR) image changes, the high network model complexity of existing capsule network-based image change detection methods, and the loss of a large amount of original image information in training samples, this paper proposed a self-supervised image change detection method based on the light capsule network (SLCapsNet). The logarithmic ratio operator difference graph was generated, and the “pseudo label” of training samples with high confidence was obtained through the maximum inter-class variance method and fuzzy C-means clustering method, which laid the foundation for self-supervised learning. The paper constructed a three-channel training sample based on the two temporal SAR images and difference graph of logarithmic ratio operators to maximize the preservation of sample information. Lightweight capsule network was designed to extract training sample features through single scale convolution, and a single scale capsule network was used to mine spatial relationships between features. Comparative experiments and ablation experiments were set up, and tests were conducted on five real SAR datasets. The experimental results show that the advantage of the proposed method is to improve the operational efficiency of the method while reducing model complexity, obtain stronger robust features, suppress the adverse impact of speckle noise on change detection performance, and improve change detection performance.

Keywords：change detection；capsule network；maximum inter-class variance method；fuzzy C-means clustering method；self-supervised learning

Received：2023-05-16；**Accepted**：2023-07-28；**Published Online**：2023-08-30 17：34

URL：link.cnki.net/urlid/11.2625.V.20230830.1621.005

* **Corresponding author**. E-mail：luoxiling@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0255

基于COOT算法的VMD-HPCA-GRU超短期风电功率预测

何星月[1]，杨靖[1,*]，朱兆强[2]，杨斌[2]，覃涛[1]

(1. 贵州大学 电气工程学院，贵阳 520025；2. 中国电建集团贵州工程有限公司 创新研究院，贵阳 550001)

摘　　要：为了提高超短期风电功率的预测精度，提出了一种基于COOT算法优化的变分模态分解(VMD)、分层主成分分析(hierarchical principal components analysis，HPCA)与门控循环单元神经网络(GRU)的组合预测模型。首先，利用能量差值法确定变分模态分解子模态数，从而将具有强非线性的原始功率序列分解为一组相对平稳的子模态。其次，利用灰色关联度分析计算高维气象特征与功率序列的关联度值并进行排序分层，利用主成分分析提取各分层特征变量的第一主成分，实现对高维气象特征的降维。最后，引入COOT算法对门控循环单元预测模型的超参数进行优化，加速模型收敛速度，提高模型预测精度。对贵州某风电场的实测数据进行仿真分析，结果表明：相较于传统GRU模型的预测结果，所提方法的均方根误差、平均绝对误差、平均绝对百分误差分别下降了67.41%、72.25%、45.69%，且预测精度高于其他4种组合预测模型，有效提高了超短期风电功率预测精度。

关 键 词：风电功率预测；变分模态分解；分层主成分分析；COOT算法；门控循环单元
中图分类号：TM614
文献标志码：A　　**文章编号**：1001-5965(2025)05-1716-10

全球经济的快速增长带来了巨大的能源消耗，随着化石能源使用导致的能源危机和环境污染问题不断加剧，清洁能源受到了世界各国的广泛关注。其中，风能因其易获取、无污染、资源储量大等特点在电力系统中的装机比重日益增大[1]。然而，风力发电的固有波动性和随机性导致了供电的稳定性、耗能与储备容量之间如何平衡等诸多问题，给电力系统的安全运行带来了严重隐患。高精度的风电功率预测，可以为电网调度提供可靠的科学凭据，有效地提高风能利用率，缓解电网调峰调频压力，减少弃风现象[2]。

近年来，为了提高风电功率预报的精度，国内外学者开展了众多的研究工作，主要可分为基于物理模型的预测[3]和基于数据统计模型的预测两类。物理模型主要由当前气象参数和地理参数构造，包括温度、地形信息、压力、障碍物等信息。基于物理模型的预测通过求解复杂高维非线性方程组来进行预测。但其对建模参数及计算要求较为苛刻。基于数据统计模型的预测，通过采集大量风电场的历史数据，挖掘风电功率与历史序列之间的非线性映射关系。经典的非线性预测模型包括支持向量机[4-5]、极限学习机[6]、随机森林[7]、人工神经网络[8-9]等。

深度学习模型因具有从原始数据中提取复杂

收稿日期：2023-05-18；**录用日期**：2023-10-21；**网络出版时间**：2023-10-25 17:12
网络出版地址：link.cnki.net/urlid/11.2625.V.20231025.1345.003
基金项目：国家自然科学基金项目(61640014, 61963009)；贵州省教育厅创新群体项目(黔教合KY字[2021]012)；贵州省科技支撑计划(黔科合支撑[2023]一般411, [2022]一般017, [2019]2152)；贵州省教育厅工程研究中心项目(黔教技[2022]043)；中国电建集团科技项目(DJ-ZDXM-2020-19, DJ-ZDXM-2022-44)；贵州省互联网+协同智能制造重点实验室项目

* **通信作者**. E-mail：jyang7@gzu.edu.cn

引用格式：何星月，杨靖，朱兆强，等. 基于COOT算法的VMD-HPCA-GRU超短期风电功率预测[J]. 北京航空航天大学学报，2025，51(5)：1716-1725. HE X Y，YANG J，ZHU Z Q，et al. VMD-HPCA-GRU ultra-short-term wind power prediction based on COOT algorithm[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1716-1725 (in Chinese).

模式的良好能力而备受关注。其中，长短期记忆(long short term memory, LSTM)[10-11]、门控循环单元(gated recurrent unit, GRU)[12]的专门化结构设计，有效规避了模型训练中的梯度爆炸或消失问题[13]，成为风电功率预测等时间序列分析领域最流行的算法之一。但是，在超短期风电功率预测等高频、高噪声实际应用场景下，单一模型的预测精度很难满足应用需求。

当前，通过将不同的模型和算法结合起来，融合多种算法优势的组合模型被广泛提出。冯义等[14]采用集合经验模态分解 (ensemble empirical mode decomposition, EEMD) 对风速序列进行分解，并剔除不规则分量，构建随机森林特征选择算法与极限学习机组合预测模型，实现了对各分量的预测，并取得了良好的预测效果。但 EEMD 分解存在端点效应，子模态中残留分解时添加的白噪声影响了模型对环境特征的提取。Zhao 等[15]采用变分模态分解 (variational mode decomposition, VMD) 来分解原始功率序列，分别使用卷积神经网络和 GRU 网络来提取风电数据的空间特征和时间特征。但 VMD 自适应性较差，分解模态数等参数设置烦琐。当前，采用智能优化算法进行超参数寻优，成为解决上述模型参数设置烦琐的方法之一。Li 等[16]提出了一种基于自适应学习因子和差分进化策略的改进蜻蜓算法，用于优化支持向量机的惩罚因子和径向基核函数宽度。何鑫等[17]通过采用基于 Tent 映射、精英反向学习和指数型旅行距离参数的改进多元宇宙优化算法对极限学习机的权值以及阈值进行优化，并通过仿真验证了模型精度的提升。但是上述智能算法存在搜索空间维度较低、算法结构复杂、内部参数较多、寻优精度有限等问题。

基于以上研究，本文提出了一种基于 COOT 算法的 VMD-HPCA-GRU 超短期风电功率预测模型。该组合模型主要分为三部分：特征处理、模态分解和预测模型。首先，对功率数据以及气象特征进行预处理，对缺失值进行填充，采用基于灰色关联度分析(grey relation analysis, GRA)的分层主成分分析方法提取出气象特征的主成分，完成对高维特征的降维。利用能量差值法确定 VMD 分解模态数，充分提取原始序列信息，分解出若干高、低频分量。将分解所得子模态结合降维气象特征分别构建 GRU 预测模型，利用 COOT 算法对模型的两层隐藏层神经元数、Dropout 丢弃率、初始学习率进行超参数寻优。将各子模态预测结果叠加重构，通过实测数据仿真，验证了所提模型具有更高的预测精度。

1　方法与理论

1.1　变分模态分解及分解模态数

1.1.1　变分模态分解

VMD 是一种非递归变分模式分解方法，其可以将非平稳的风电功率序列分解为具有一定稀疏性质的若干子模态序列，可以有效避免经验模态分解(empirical mode decomposition, EMD)等分解方法所产生的模态混叠等问题。该方法主要框架为求解变分问题。其变分约束表达式为

$$\begin{cases}\min\limits_{\{u_k\},\{\omega_k\}}\left\{\sum\limits_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\dfrac{\mathrm{j}}{\pi t}\right)\cdot u_k(t)\right]\mathrm{e}^{-\mathrm{j}\omega_k t}\right\|_2^2\right\}\\ \text{s.t.}\sum\limits_{k=1}^{K}u_k(t)=f(t)\end{cases}\tag{1}$$

式中：$\delta(t)$ 为狄拉克函数；$u_k(t)$ 为模态分量 k 在 t 时刻的值；ω_k 为各分量的频率中心。

通过加入惩罚因子 α 和拉格朗日算子 λ，将问题转变为非约束变分问题。

$$L(\{u_k\},\{\omega_k\},\lambda)=\alpha\sum_{k=1}^{K}\left\|\partial_t\left[\left(\delta(t)+\frac{\mathrm{j}}{\pi t}\right)\cdot u_k(t)\right]\mathrm{e}^{-\mathrm{j}w_k t}\right\|_2^2+\left\|f(t)-\sum_{k=1}^{K}u_k(t)\right\|_2^2+\left\langle\lambda(t),f(t)-\sum_{k=1}^{K}u_k(t)\right\rangle\tag{2}$$

进行时-频域变换，采用交替方向乘子法 (alternating direction method of multipliers, ADMM) 对 u_k、ω_k 进行迭代更新，直到达到收敛精度或者最大迭代次数则停止。

$$\hat{u}_k^{n+1}(\omega)=\frac{\hat{f}(\omega)-\sum\limits_{i\neq k}\hat{u}_i(\omega)+\hat{\lambda}(\omega)/2}{1+2\alpha(\omega-\omega_k)^2}\tag{3}$$

$$\omega_k^{n+1}=\frac{\int_0^{\infty}\omega|\hat{u}_k(\omega)|\mathrm{d}\omega}{\int_0^{\infty}|\hat{u}_k(\omega)|^2\mathrm{d}\omega}\tag{4}$$

$$\lambda^{n+1}=\lambda^n+\tau\left(f-\sum_k u_k^{n+1}\right)\tag{5}$$

$$\sum_k\left(\left\|u_k^{n+1}-u_k^n\right\|_2^2/\left\|u_k^n\right\|_2^2\right)<\varepsilon\tag{6}$$

式中：n 为迭代数；τ 为更新因子；ε 为收敛精度。

1.1.2　分解模态数

VMD 分解所得子模态序列的正交性是表征分解效果的评价指标之一。当子模态满足正交性时，原始序列的能量与分解所得的 k 个子模态的能量和相等。由于 k 值的选择过大而产生过分解现象时，分解过程将产生虚拟分量，从而导致子模态的能量和大于正常分解下各子模态的能量和。通过

计算各分解模态数下的能量差，选取能量差值最小时的 k 值，为最佳分解模态数。序列的能量及能量差值计算方法如下：

$$E=\sqrt{\sum_{t=1}^{n}f^2(t)\Big/n} \tag{7}$$

$$\eta=\frac{|E_k-E_{k-1}|}{E_{k-1}} \tag{8}$$

式中：E 为序列能量；n 为序列长度；η 为能量差；E_{k-1}、E_k 分别为分解模态数为 k–1 和 k 时，各子模态的能量之和。

1.2 分层主成分分析

通过数值天气预报 (numerical weather prediction, NWP) 系统，可以及时便捷地获取高精度的气象信息，但是庞杂的气象信息不加以处理，直接输入模型，将会加剧模型的运算负荷，降低模型的鲁棒性，甚至会造成维度灾难和过拟合现象的发生。因此，对高维度特征采取降维，合理消除数据冗余分量，保留关键信息，是提高模型学习效率和预测性能的方法之一。

1.2.1 主成分分析

主成分分析（principal components analysis, PCA）是一种对高维度特征预处理方法，对原始数据矩阵的协方差矩阵进行特征值分解，保留特征数据的主要特征，去除数据的非重要特征及噪声。其基本原理如下。

假设有 n 组进行功率预测的气象数据样本，每个样本包含 m 个特征，则整个数据集可以用矩阵表示为

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{9}$$

1）对原始矩阵采用标准化处理，生成的标准化矩阵 $\boldsymbol{X}^*$ 如下式所示：

$$x_{ij}^*=\frac{x_{ij}-\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_{ij}}{\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}\left(x_{ij}-\dfrac{1}{n}\sum_{i=1}^{n}x_{ij}\right)^2} \tag{10}$$

2）计算协方差矩阵 $\boldsymbol{Q}$：

$$\boldsymbol{Q}=\frac{1}{n}(\boldsymbol{X}^*)^{\mathrm{T}}\boldsymbol{X}^*=\begin{bmatrix} q_{11} & \cdots & q_{1m} \\ \vdots & & \vdots \\ q_{n1} & \cdots & q_{nm} \end{bmatrix} \tag{11}$$

3）根据特征方程 $|\lambda\boldsymbol{I}-\boldsymbol{Q}|=0$，求解方程得到特征根 $\lambda_1, \lambda_2, \cdots, \lambda_m$（$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_m>0$），求解出每个特征值对应的单位特征向量 $e_1, e_2, \cdots, e_m$。

4）将特征值按照大小顺序进行排序，并计算出每个特征值的贡献率：

$$\varphi_j=\frac{\lambda_j}{\displaystyle\sum_{s=1}^{m}\lambda_s}\times 100\%, j=1,2,\cdots,m \tag{12}$$

1.2.2 分层主成分分析方法

传统主成分分析降维过程中，数据的损失较为严重，贡献率较低的主成分往往包含着对样本差异的重要特征信息。为此，本文提出一种基于灰色关联度分析的分层主成分分析方法。

GRA 通过参考因素间的发展方向和速率变化趋势，衡量因素与因素之间影响的相对强弱。从而可以对特征序列的影响因素进行有效选择。GRA 的基本原理见文献 [18]。

分层主成分分析（hierarchical principal components analysis, HPCA）的方法首先通过计算输入气象特征与功率序列的灰色关联度，然后将其由大到小划分为 5 层，划分标准如下式所示：

$$h=\begin{cases} 1\ \text{level} & 0.8<r\leqslant 1 \\ 2\ \text{level} & 0.6<r\leqslant 0.8 \\ 3\ \text{level} & 0.4<r\leqslant 0.6 \\ 4\ \text{level} & 0.2<r\leqslant 0.4 \\ 5\ \text{level} & 0<r\leqslant 0.2 \end{cases} \tag{13}$$

式中：r 为灰色关联度值。

其次，对各层进行主成分分析，保留各层的第一主成分，作为降维后的输入特征。

同层特征间灰色关联度值相近，因此对功率序列的影响程度也相似。对同层特征进行主成分分析，在对冗余数据进行降维，保留关键信息的同时，也避免了主成分分析造成对贡献较小成分的损失。HPCA 通过特征分层，考虑特征间的差异性，保留输入特征的关键信息及其多样性。

1.3 COOT 算法

COOT 算法是 Naruei 等[19] 于 2021 年提出的一种新型群体智能算法。COOT 是一种小型水鸟，在水面上有许多不同的群体行为，行为的最终目标是向食物或特定的地点移动。

在水面上，COOT 主要有四种不同的运动行为：随机运动、链式运动、根据领导者调整位置和领导者运动。该算法的具体步骤如下。

普通个体运动：该过程主要由三种运动状态构成，执行机制表示如下：

$$\boldsymbol{X}_{\text{coot}}(i)=\begin{cases} \text{lead} & \text{rand}<0.5 \\ \begin{cases}\text{chain rand}<0.5\ \&\ i\neq 1 \\ \text{random otherwise}\end{cases} & \text{otherwise} \end{cases} \tag{14}$$

式中：lead、chain、random 表示引导运动、链式运动

和随机运动，分别表示如下：

$$\boldsymbol{X}_{\text{coot}}(i)=\boldsymbol{X}_{\text{lead}}(L)+2\times R_1\times\cos(2\pi R)\times(\boldsymbol{X}_{\text{lead}}(L)-\boldsymbol{X}_{\text{coot}}(i)) \tag{15}$$

$$\boldsymbol{X}_{\text{coot}}(i)=\frac{\boldsymbol{X}_{\text{coot}}(i-1)+\boldsymbol{X}_{\text{coot}}(i)}{2} \tag{16}$$

$$\boldsymbol{X}_{\text{coot}}(i)=\boldsymbol{X}_{\text{coot}}(i)+A\times R_2\times(\boldsymbol{P}-\boldsymbol{X}_{\text{coot}}(i)) \tag{17}$$

式中：$\boldsymbol{X}_{\text{coot}}$ 是普通 COOT；$\boldsymbol{X}_{\text{lead}}$ 是领导 COOT；$\boldsymbol{P}$ 是随机生成个体位置；R 是 [−1,1] 中的随机数；R_1、R_2 是 [0,1] 中的随机数，L 为领导者的索引。为普通个体分配领导者如下所示：

$$L=1+(i\ \text{MOD}\ M) \tag{18}$$

式中：i 为普通个体索引；M 为领导者的数量。

领导者运动：该运动根据下式，利用最优个体牵引领导者，帮助种群向最优区域收敛：

$$\boldsymbol{X}_{\text{lead}}(i)=\begin{cases}B\times R_3\times\cos(2\pi R)\times\\(\boldsymbol{g}_{\text{Best}}-\boldsymbol{X}_{\text{lead}}(i))+\boldsymbol{g}_{\text{Best}} & R_4<0.5\\B\times R_3\times\cos(2\pi R)\times\\(\boldsymbol{g}_{\text{Best}}-\boldsymbol{X}_{\text{lead}}(i))-\boldsymbol{g}_{\text{Best}} & R_4\geqslant 0.5\end{cases} \tag{19}$$

式中：$\boldsymbol{g}_{\text{Best}}$ 是当前最优个体位置；R_3 和 R_4 是 [0,1] 中的随机数。

1.4　门控循环单元

GRU 网络是一种传统循环神经网络的变体，适用于处理和预测时间序列等动态非线性序列前后长期依赖问题。相较于 LSTM 神经网络的三门结构，其通过更新门和重置门的两门结构来控制时序数据的保留与更新。因此能够减少内存使用以及提供更快的执行速度。GRU 网络通过重置门 $\boldsymbol{r}_t$ 控制与历史时刻的融合程度，更新门 $\boldsymbol{z}_t$ 控制记忆数据的保留程度。内部具体结构如图 1 所示。计算过程如下列各式所示：

$$\boldsymbol{r}_t=\sigma(\boldsymbol{W}_r\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_r) \tag{20}$$

$$\boldsymbol{z}_t=\sigma(\boldsymbol{W}_z\cdot[\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_z) \tag{21}$$

$$\tilde{\boldsymbol{h}}_t=\tanh(\boldsymbol{W}_h\cdot[\boldsymbol{r}_t\boldsymbol{h}_{t-1},\boldsymbol{x}_t]+\boldsymbol{b}_h) \tag{22}$$

$$\boldsymbol{h}_t=(1-\boldsymbol{z}_t)\boldsymbol{h}_{t-1}+\boldsymbol{z}_t\tilde{\boldsymbol{h}}_t \tag{23}$$

式中：$\sigma(\ )$ 为 Sigmoid( ) 函数；$\boldsymbol{r}_t$ 为重置门；$\boldsymbol{z}_t$ 为更新门；$\boldsymbol{W}_r$、$\boldsymbol{b}_r$ 为重置门权重、偏移量；$\boldsymbol{W}_z$、$\boldsymbol{b}_z$ 为更新门权重、偏移量；$\boldsymbol{W}_h$、$\boldsymbol{b}_h$ 为 tanh 神经层的权重、偏移量；$\boldsymbol{x}_t$ 为当前时刻时序输入；$\boldsymbol{h}_{t-1}$ 为上一时刻隐藏状态。

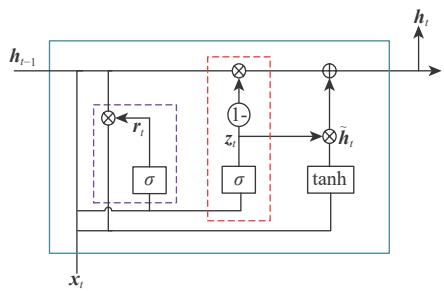


图 1　GRU 网络内部结构

Fig. 1　Internal structure of GRU network

1.5　基于 COOT 算法的 VMD-HPCA-GRU 组合预测模型

鉴于风电功率的波动性和随机性等特点，本文建立了一种基于 COOT 算法的 VMD-HPCA-GRU 超短期风电功率组合预测模型，以提高功率序列预测精度。预测建模流程如图 2 所示。

具体步骤如下：

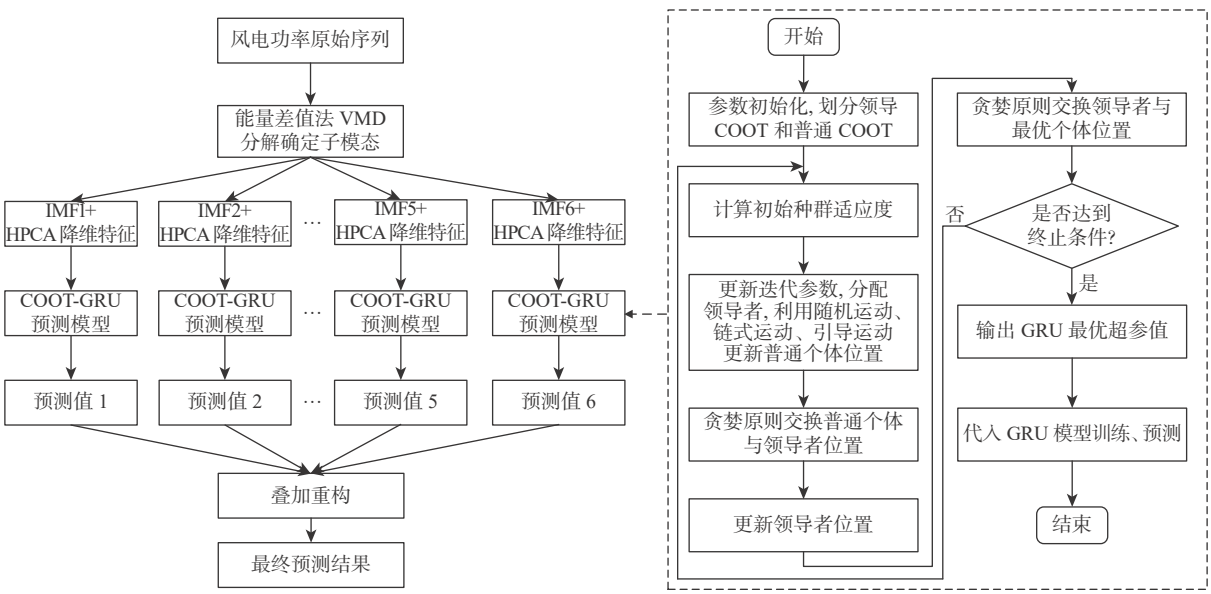


图 2　基于 COOT 算法的 VMD-HPCA-GRU 组合预测流程

Fig. 2　VMD-HPCA-GRU combination prediction process based on COOT algorithm

1）对功率数据及系统采集的气象数据进行预处理，对缺失值进行插值处理。

2）采用能量差值法的VMD分解处理历史功率序列，得到若干子模态序列。

3）采用HPCA方法，对高维气象特征进行降维处理，生成若干保留原始重要特征的低维气象特征。

4）对分解所得的各功率子模态序列，结合HPCA方法所得的新特征序列分别构建COOT-GRU模型。

5）确定COOT算法的迭代次数和种群数。将GRU模型的第一层隐藏神经元个数、第二层隐藏神经元个数、Dropout丢弃率以及初始学习率作为超参数，使用COOT算法进行寻优。

6）将寻优超参数代入GRU模型分别对各子模态进行仿真预测，采用Adam优化器调整模型的权值、偏置参数，提高对各子模态的预测效果。

7）叠加重构各子序列的预测结果，并根据模型评价指标计算模型预测精度。

2 实例分析

本文以贵州某风电场作为研究对象，该风电场装机容量为96MW。选取2020年2月1日至3月28日的实际功率数据（包含10 m处风速、风向，30 m处风速、风向，50 m处风速、风向，轮毂高度处风速、风向，温度，大气压强，湿度共11维气象特征，其中风向取正弦值），共计5 376组数据作为训练样本，3月29日至3月31日共计288组数据作为测试样本。数据采样间隔为15 min，采用迭代式预测方法，预测未来一步，即单步预测。时间窗口大小经过多次实验对比，选择为6。原始功率序列如图3所示。

为了便于模型训练收敛以及减少量纲对预测结果的影响，对功率数据以及气象特征进行最大最小值归一化处理，处理过程如下式所示：

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{24}$$

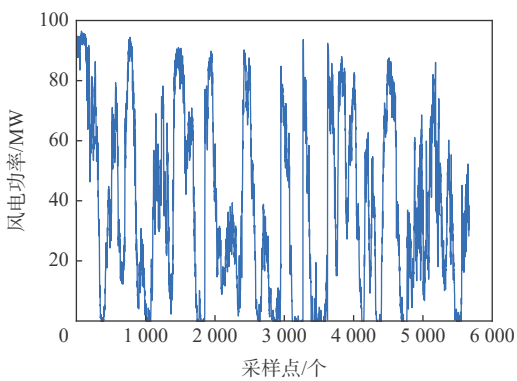


图3 原始功率序列图

Fig. 3 Original power sequence

式中：x_i为原始数据；$x_{\min}$为序列最小值；$x_{\max}$为序列最大值。

2.1 模型评价指标

为更加合理全面地评估本文模型的预测性能，采用均方根误差（root mean square error, RMSE），侧重衡量功率序列在极值点处的预测偏差；平均绝对误差（mean absolute error, MAE），衡量整段序列的预测偏差；平均绝对百分误差（mean absolute percentage error, MAPE），衡量实测值与预测值拟合偏差的程度。上述3种指标均为数值越小，模型效果越佳。3种评价指标的计算公式如下：

$$\sigma_{\mathrm{RMSE}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \tag{25}$$

$$\sigma_{\mathrm{MAE}} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \hat{x}_i| \tag{26}$$

$$\sigma_{\mathrm{MAPE}} = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i - \hat{x}_i|}{x_i} \times 100\% \tag{27}$$

式中：n为样本数；x_i为样本i的实际值；$\hat{x}_i$为样本i的预测值。

2.2 功率序列分解

风电功率序列整体呈现强非线性和波动性。为获取序列深层次特征信息，本文采用VMD算法对原始功率序列进行分解。分解前需要对分解子模态数k进行确定，k值过小，将会造成欠分解，不能很好地分离出短期波动分量。k值过大，将会造成过分解，产生模态混叠现象。本文采用能量差值法确定k值。k值常见取值为2～8，通过式(7)、式(8)计算出各模态数的能量差值，如图4所示。

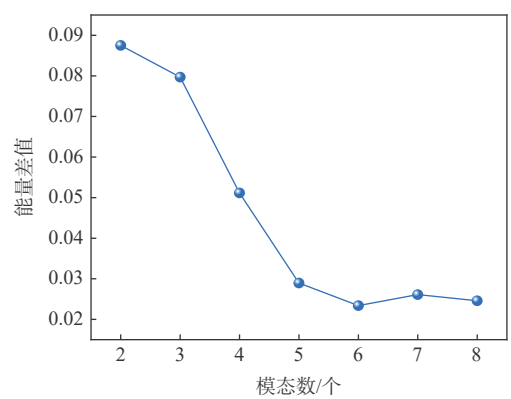


图4 子模态能量差值

Fig. 4 Submode energy difference

图4中，局部极值出现在k=6处，因此针对该功率序列选择VMD分解层数为6。VMD其余各参数按照文献[20]设置，分解结果如图5所示。

从图5可知，经过VMD分解，具有复杂波动的原始序列转化为6个中心频率互异的窄带宽子序

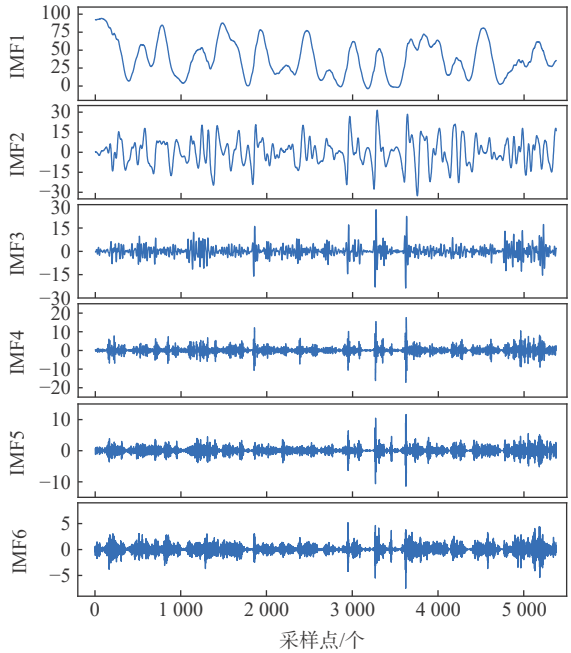


图 5　VMD 分解结果

Fig. 5　VMD results

列。6 个子序列分别包含着风电功率的长期、短期以及其余深层随机特征。各序列内的随机性和复杂性得到降低，增强了模型的可学习性和拟合精度。

2.3　气象特征降维

采用 1.2 节的 HPCA 分析法，对风电场系统采集的气象特征进行降维处理，在本例中气象特征维度 m=11。首先采用 GRA 方法分别计算功率序列与气象特征之间的关联度值。关联度值越大，则该特征对功率序列的影响越大。表 1 为气象特征的灰色关联度值计算及分层结果。

表 1　灰色关联度值计算及分层结果

Table 1　Calculation and hierarchical results of gray relation value

气象特征	关联度	层级	气象特征	关联度	层级
风速(10 m)	0.811 1	1	风速(轮毂)	0.807 5	1
风向(10 m)	0.611 5	2	风向(轮毂)	0.616 9	2
风速(30 m)	0.803 4	1	温度	0.630 0	2
风向(30 m)	0.627 9	2	大气压强	0.676 0	2
风速(50 m)	0.805 7	1	湿度	0.617 1	2
风向(50 m)	0.628 1	2			

从表 1 可知，11 维气象特征根据计算出的灰色关联度被划分在了第 1 层和第 2 层。其中，第 1 层特征维度 m=4，第 2 层特征维度 m=7。利用 PCA 分别对两个层级的气象特征进行降维，保留各层第一主成分作为降维后的新气象特征。

2.4　COOT-GRU 模型超参数优化过程

对 2.2 节分解所得的各子模态分别构建 COOT-GRU 模型。隐藏层数的增加可以提高模型的非线性拟合能力，但也会使得模型训练成本增加，甚至发生过拟合。因此本文设置隐藏层数为 2 层。采用均方误差函数 (mean square error, MSE) 作为模型的损失函数，Adam 算法融合了 Momentum 和 AdaGrad 算法，具有高效搜索参数空间和占用内存小等优点，因此采用 Adam 算法作为模型的训练优化算法。模型的迭代次数(epoch)设置为 50 次，batch-size 设置为 64。

COOT 算法初始种群规模设置为 15，最大迭代次数为 30，使用式 (25) 作为 COOT 算法的适应度函数。将 GRU 网络结构参数编码为 COOT 算法的个体位置，搜索空间为 4 个维度，分别为第 1 隐藏层神经元数、第 2 隐藏层神经元数、Dropout 丢弃率、Adam 初始学习率。4 个超参数的寻优搜索范围如表 2 所示。

表 2　超参数搜索范围

Table 2　Hyperparameter search scope

维度	超参数	搜索范围
1	第1隐藏层神经元数	[2, 128]
2	第2隐藏层神经元数	[2, 128]
3	Dropout丢弃率	[0.1, 0.9]
4	初始学习率	[0.001, 0.1]

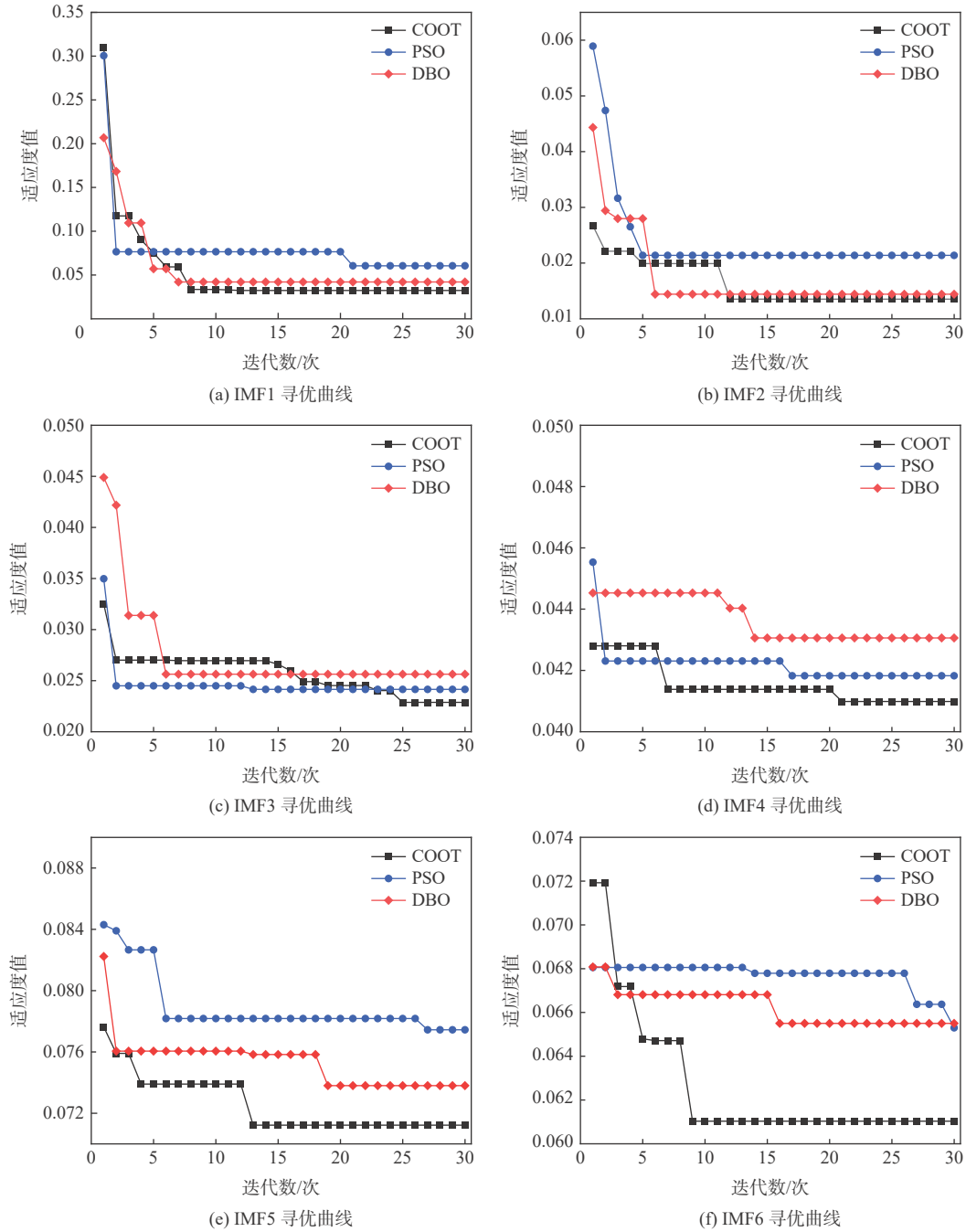
为了验证 COOT 算法的寻优效果，选取粒子群算法 (particle swarm optimization, PSO)、蜣螂优化算法 (Dung Beetle optimizer, DBO) 进行对比实验，初始种群规模和最大迭代次数皆同上。3 种群体智能算法对 VMD 分解所得的 6 个子模态寻优过程的适应度变化曲线如图 6 所示。

从图 6 可以看出，在各个子模态的寻优过程中，COOT 算法的收敛精度更高，寻优效果更佳。在算法搜索后期，跳出搜索空间局部最优的能力最佳。说明 COOT 算法更容易找到适用于 GRU 模型的超参数值，使用 COOT 算法寻优得到的超参数，有助于模型收敛，提高模型鲁棒性和拟合能力。COOT 算法对 6 个子模态寻优所得四维超参数值如表 3 所示。

2.5　预测结果分析

根据 COOT 算法的寻优结果，对各子序列模型参数进行配置。在划分的测试集上进行仿真测试，并与 4 种单一模型，支持向量回归 (support vactor regression, SVR) 模型、后向传播 (back propagation, BP) 模型、GRU 模型、LSTM 模型进行对比。不同单一模型在测试集上的预测结果曲线对比如图 7 所示，3 种误差评价指标对比如表 4 所示。

从图 7 可知，传统单一模型预测曲线出现较明显滞后现象，对剧烈波动情况不敏感。本文提出的



(a) IMF1 寻优曲线　(b) IMF2 寻优曲线

(c) IMF3 寻优曲线　(d) IMF4 寻优曲线

(e) IMF5 寻优曲线　(f) IMF6 寻优曲线

图 6　适应度变化曲线对比图

Fig. 6　Comparison of fitness change curves

表 3　超参数寻优结果

Table 3　Hyperparameter optimization results

子模态	第1维	第2维	第3维	第4维
IMF1	12	44	0.100 5	0.006 3
IMF2	65	57	0.350 6	0.012 4
IMF3	23	49	0.100 1	0.025 9
IMF4	52	116	0.291 2	0.014 8
IMF5	64	65	0.104 2	0.021 7
IMF6	35	69	0.179 6	0.024 4

预测模型拟合曲线更为贴合实测值变化曲线，在实测曲线剧烈变化，反复波动转折之处，追踪效果仍较为准确、灵敏。

从表 4 可知，SVR 模型的预测效果最差，且拟合曲线偏离程度最高。LSTM 模型与 GRU 模型预测误差效果接近，但是 GRU 模型结构更为简单，节约算法内存空间的使用。在处理时间序列预测问题上，GRU 等循环神经网络变体展现了其优越性。GRU 模型的 RMSE 值相较于 SVR 模型和 BP 模型分别降低了 11.40% 和 6.98%，MAPE 值分别降低了 38.33% 和 16.83%。COOT-VMD-HPCA-GRU 组合模型相较于单一 GRU 模型在 RMSE、MAE、MAPE 上分别降低了 67.41%、72.25%、45.67%。相较于其他单一模型，曲线拟合度以及误差值均具有明显优势。

为了进一步验证本文模型的有效性，将其与组

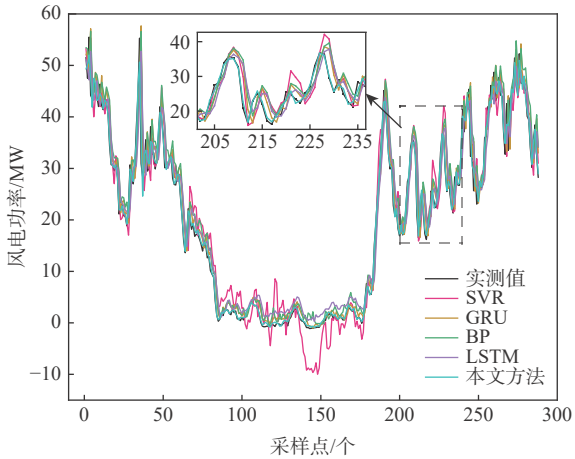


图 7　单一模型预测曲线对比

Fig. 7　Comparison of single model prediction curves

表 4　单一模型预测误差对比

Table 4　Comparison of single model prediction errors

预测模型	RMSE	MAE	MAPE/%
SVR	4.494 6	3.487 0	0.577 1
GRU	3.982 4	3.012 4	0.355 9
BP	4.281 4	3.196 0	0.427 9
LSTM	3.994 0	2.778 7	0.367 7
COOT-VMD-HPCA-GRU	1.298 0	0.835 8	0.193 3

合模型 VMD-GRU、HPCA-GRU 进行对比，并准备了两种相同架构的组合模型，即 PSO-VMD- HPCA-GRU 模型、COOT-EMD-HPCA-GRU 模型。不同组合模型预测结果曲线对比如图 8 所示，相应误差评价指标对比如表 5 所示。

由图 8 可知，COOT-EMD-HPCA-GRU 模型的预测效果最佳，在实测曲线峰谷点等预测难度较大的位置，仍有着较好的拟合度。

由表 5 发现，采用分解算法的组合模型，在 3 项误差评价指标中均有了大幅降低。使用了 VMD 分解算法的 GRU 模型，3 项误差指标分别降

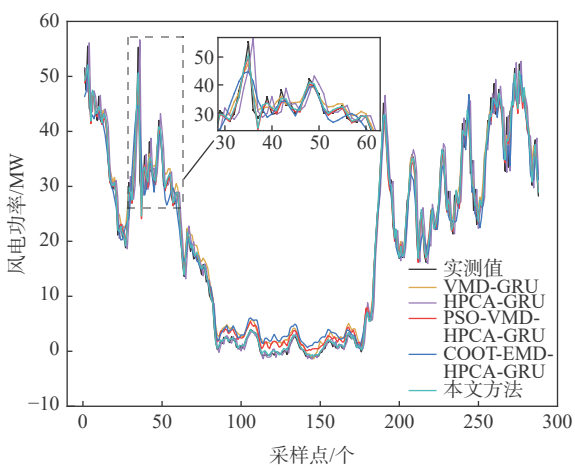


图 8　组合模型预测曲线对比

Fig. 8　Comparison of combined model prediction curves

表 5　组合模型预测误差对比

Table 5　Comparison of combined model prediction errors

预测模型	RMSE	MAE	MAPE/%
VMD-GRU	2.005 1	1.678 4	0.294 5
HPCA-GRU	3.824 6	2.514 2	0.327 3
PSO-VMD-HPCA-GRU	1.602 6	1.265 0	0.273 5
COOT-EMD-HPCA-GRU	2.278 4	1.847 3	0.323 7
COOT-VMD-HPCA-GRU	1.298 0	0.835 8	0.193 3

低了 49.65%、44.28%、17.25%。这是因为分解算法可以将原始功率序列分解为一系列不同频率尺度的稳定子模态，提取出功率序列的长期、短期以及随机特征，降低模型的拟合难度。采用 VMD 分解的组合模型相较于采用 EMD 分解的组合模型 3 项误差指标分别降低了 43.03%、54.76%、40.28%，说明 VMD 分解对细节信息的提取更为准确，可以避免 EMD 分解所存在的模态混叠和端点效应问题。

采用了 HPCA 方法对气象特征进行降维处理的 GRU 模型，3 项误差指标分别降低了 3.96%、16.53%、8.04%。虽然 RMSE 指标并未有较大幅度提升，但是 HPCA 法可以降低模型复杂度，降低维度灾难风险，提高模型学习效率。误差指标并未发生恶化，证明 HPCA 可以分离出高维气象特征的有效信息，建立气象特征和预测功率间的逻辑关联。

引入 PSO 算法和 COOT 算法的 VMD-GRU 模型，RMSE 指标分别降低了 20.07%、35.27%，MAE 指标分别降低了 24.63%、50.20%，MAPE 指标分别降低了 7.13%、34.36%，说明群体智能算法对模型超参数寻优的有效性，可以避免烦琐的人为试探，自适应搜索对应任务的模型超参数。COOT 优化模型相较于 PSO 优化模型的 3 项误差指标分别降低了 19.01%、33.93%、29.32%，说明 COOT 算法的搜索效率更高，寻优能力更优越。

本文模型的预测值与功率实测值的绝对误差值分布如图 9 所示。

从图 9 可以看出，绝对误差整体呈现正态分

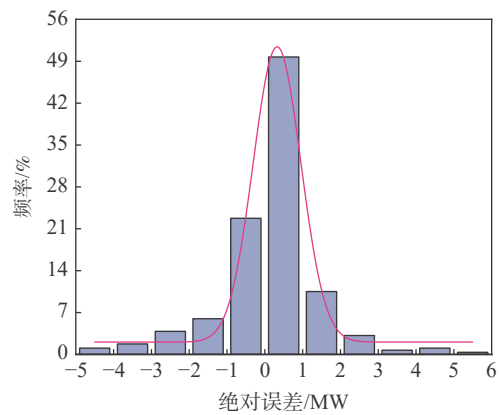


图 9　COOT-VMD-HPCA-GRU 误差分布

Fig. 9　COOT-VMD-HPCA-GRU error distribution

布，有70%分布在[−1,1]，证明模型拟合效果良好，预测效果可信度高，对实际风电场发电并网具有一定指导意义。

3 结 论

鉴于风电功率固有波动性和随机性以及预测难度大等特点，本文提出了一种基于COOT算法的VMD-HPCA-GRU超短期风电功率预测方法。通过实例分析和预测模型对比，总结出以下结论。

1）基于能量差法的VMD分解可以有效提取风电功率深层细节特征，改善传统EMD分解模态混叠以及端点效应问题，提高组合预测模型的预测精度，解决单一预测模型普遍相位滞后问题。

2）采用基于GAR的分层主成分分析方法对模型预测精度提升有限，但其可以有效提取输入特征主成分信息，降低模型内部结构规模，加速模型的收敛速度和学习效率。

3）针对不同预测任务情况下，GRU模型超参数选择困难，取经验值以及人工试探方法烦琐且精度有限问题，采用COOT算法对模型四个超参数进行寻优，相较于PSO算法，RMSE、MAE、MAPE分别降低了19.01%、33.93%、29.32%。

4）本文模型相较于传统的SVR、BP、LSTM、GRU模型，预测精度有了大幅度提升，表明了组合模型方法的可行性。通过与PSO-VMD-HPCA-GRU等组合模型的对比分析，验证了本文超短期风电功率预测的有效性，对提高风电功率预测精度具有较好的理论价值和实践意义。

参考文献（References）

[1] WISER R, RAND J, SEEL J, et al. Expert elicitation survey predicts 37% to 49% declines in wind energy costs by 2050[J]. Nature Energy, 2021, 6: 555-565.

[2] HOSSAIN M A, CHAKRABORTTY R K, ELSAWAH S, et al. Very short-term forecasting of wind power generation using hybrid deep learning model[J]. Journal of Cleaner Production, 2021, 296: 126564.

[3] 宋家康, 彭勇刚, 蔡宏达, 等. 考虑多位置NWP和非典型特征的短期风电功率预测研究[J]. 电网技术, 2018, 42(10): 3234-3242.
SONG J K, PENG Y G, CAI H D, et al. Research of short-term wind power forecasting considering multi-location NWP and uncanonical feature[J]. Power System Technology, 2018, 42(10): 3234-3242(in Chinese).

[4] 王斌, 魏成伟, 谢丽蓉, 等. 基于风速误差校正和ALO-LSSVM的风电功率预测[J]. 太阳能学报, 2022, 43(1): 58-63.
WANG B, WEI C W, XIE L R, et al. Wind power forecasting based on wind speed error corretion and ALO-LSSVM[J]. Acta Energiae Solaris Sinica, 2022, 43(1): 58-63(in Chinese).

[5] DING M, ZHOU H, XIE H, et al. A time series model based on hybrid-kernel least-squares support vector machine for short-term wind power forecasting[J]. ISA Transactions, 2021, 108: 58-68.

[6] 李青, 张新燕, 马天娇, 等. 基于ECBO-VMD-WKELM的风电功率超短期多步预测[J]. 电网技术, 2021, 45(8): 3070-3080.
LI Q, ZHANG X Y, MA T J, et al. Multi-step ahead ultra-short term forecasting of wind power based on ECBO-VMD-WKELM[J]. Power System Technology, 2021, 45(8): 3070-3080(in Chinese).

[7] 曾亮, 狄飞超, 兰欣, 等. 基于CEEMD-CNN-BiGRU-RF模型的短期风电功率预测[J]. 可再生能源, 2022, 40(2): 190-195.
ZENG L, DI F C, LAN X, et al. Short-term wind power prediction based on CEEMD-CNN-BiGRU-RF model[J]. Renewable Energy Resources, 2022, 40(2): 190-195(in Chinese).

[8] COUTO A, ESTANQUEIRO A. Enhancing wind power forecast accuracy using the weather research and forecasting numerical model-based features and artificial neuronal networks[J]. Renewable Energy, 2022, 201: 1076-1085.

[9] YAGHOUBIRAD M, AZIZI N, FARAJOLLAHI M, et al. Deep learning-based multistep ahead wind speed and power generation forecasting using direct method[J]. Energy Conversion and Management, 2023, 281: 116760.

[10] 刘倩倩, 刘钰山, 温烨婷, 等. 基于PCC-LSTM模型的短期负荷预测方法[J]. 北京航空航天大学学报, 2022, 48(12): 2529-2536.
LIU Q Q, LIU Y S, WEN Y T, et al. Short-term load forecasting method based on PCC-LSTM model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(12): 2529-2536(in Chinese).

[11] 王渝红, 史云翔, 周旭, 等. 基于时间模式注意力机制的BiLSTM多风电机组超短期功率预测[J]. 高电压技术, 2022, 48(5): 1884-1892.
WANG Y H, SHI Y X, ZHOU X, et al. Ultra-short-term power prediction for BiLSTM multi wind turbines based on temporal pattern attention[J]. High Voltage Engineering, 2022, 48(5): 1884-1892(in Chinese).

[12] XIAO Y L, ZOU C Z, CHI H T, et al. Boosted GRU model for short-term forecasting of wind power with feature-weighted principal component analysis[J]. Energy, 2023, 267: 126503.

[13] ZHANG Y Z, XIONG R, HE H W, et al. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries[J]. IEEE Transactions on Vehicular Technology, 2018, 67(7): 5695-5705.

[14] 冯义, 刘慧文, 张宝平, 等. 基于集合经验模态分解和特征选择极端学习机的风速预测[J]. 智慧电力, 2018, 46(12): 30-37.
FENG Y, LIU H W, ZHANG B P, et al. Short-term wind speed forecasting using ensemble empirical mode decomposition and extreme learning machine with feature selection[J]. Smart Power, 2018, 46(12): 30-37(in Chinese).

[15] ZHAO Z N, YUN S N, JIA L Y, et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features[J]. Engineering Applications of Artificial Intelligence, 2023, 121: 105982.

[16] LI L L, ZHAO X, TSENG M L, et al. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm[J]. Journal of Cleaner Production, 2020, 242: 118447.

[17] 何鑫, 雷勇, 王进武, 等. 基于变分模态和改进多元宇宙优化的短

期风电功率预测[J/OL]. 现代电力, (2022-12-16)[2023-05-11]. https://doi.org/10.19725/j.cnki.1007-2322.2022.0118.
HE X, LEI Y, WANG J W, et al. Short-term wind power forecasting based on variational modes and improved multiverse optimization[J/OL]. Modern Electric Power, (2022-12-16) [2023-05-11]. https://doi.org/10.19725/j.cnki.1007-2322.2022.0118(in Chinese).

[18] 宋启军, 胡翔奎, 王腾飞, 等. 基于 GRA-SVM 的光伏发电功率预测[J]. 水力发电, 2016, 42(11): 110-113.
SONG Q J, HU X K, WANG T F, et al. Photovoltaic power forecasting based on GRA-SVM model[J]. Water Power, 2016, 42(11): 110-113(in Chinese).

[19] NARUEI I, KEYNIA F. A new optimization method based on COOT bird natural life model[J]. Expert Systems with Applications, 2021, 183: 115352.

[20] 徐岩, 向益锋, 马天祥. 基于粒子群算法优化参数的 VMD-GRU 短期电力负荷预测模型[J]. 华北电力大学学报 (自然科学版), 2023, 50(1): 38-47.
XU Y, XIANG Y F, MA T X. VMD-GRU short-term power load forecasting model based on optimized parameters of partical swarm algorithm[J]. Journal of North China Electric Power University (Natural Science Edition), 2023, 50(1): 38-47(in Chinese).

VMD-HPCA-GRU ultra-short-term wind power prediction based on COOT algorithm

HE Xingyue[1]，YANG Jing[1,*]，ZHU Zhaoqiang[2]，YANG Bin[2]，QIN Tao[1]

(1. College of Electrical Engineering，Guizhou University，Guiyang 520025，China；
2. Innovation Institute，China Power Construction Group Guizhou Engineering Co.，Ltd，Guiyang 550001，China)

Abstract：In order to improve the prediction accuracy of ultra-short-term wind power, a combined prediction model based on variational modal decomposition (VMD), hierarchical principal component analysis (HPCA), and gated recurrent unit (GRU) neural network optimized by COOT algorithm was proposed. Firstly, the submode number of VMD was determined by the energy difference method so that the original power sequence with strong nonlinearity was decomposed into a set of relatively stationary submodes. Secondly, the correlation degree value between high-dimensional meteorological features and power sequence was calculated by gray relation analysis, and the ranking and stratification were carried out. The first principal component of feature variables in each layer was extracted by principal component analysis (PCA) to realize the dimensionality reduction of high-dimensional meteorological features. Finally, the COOT algorithm was introduced to optimize the hyperparameters of the GRU prediction model, accelerate the model convergence speed, and improve the model prediction accuracy. Simulation analysis was carried out on the measured data of a wind farm in Guizhou Province, and the results show that compared with the prediction results of the traditional GRU model, the root mean square error, mean absolute error, and mean absolute percentage error of the proposed method are reduced by 67.41%, 72.25%, and 45.69%, respectively, and the prediction accuracy of the proposed method is higher than that of the other four combined prediction models, which effectively improves the prediction accuracy of ultra-short-term wind power.

Keywords：wind power prediction; variational mode decomposition; hierarchical principal component analysis; COOT algorithm; gated recurrent unit

Received：2023-05-18；**Accepted**：2023-10-21；**Published Online**：2023-10-25 17：12

URL：link.cnki.net/urlid/11.2625.V.20231025.1345.003

Foundation items：National Natural Science Foundation of China (61640014, 61963009); Innovation group of Guizhou Education Department (Qianjiaohe KY[2021]012); Guizhou provincial Science and Technology Projects (Qiankehe Zhicheng [2023]Yiban411, [2022]Yiban017, [2019]2152); Engineering Research Center of Guizhou Education Department (Qianjiaoji[2022]043); Science and Technology Project of Power Construction Corporation of China, Ltd (DJ-ZDXM-2020-19, DJ-ZDXM-2022-44); Key Laboratory of Internet plus Collaborative Intelligent Manufacturing in Guizhou Province

* **Corresponding author**. E-mail：jyang7@gzu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0263

板簧橡胶支座刚度与疲劳寿命波动性仿真分析及试验研究

姬娜[1]，刘娟[3]，王浩然[3]，高蕊[3]，卢咏来[1,2]，李凡珠[1,2,*]

(1. 北京化工大学 有机无机复合材料国家重点实验室，北京 100029；2. 北京化工大学 材料科学与工程学院，北京 100029；
3. 中国重型汽车集团有限公司，济南 250031)

摘 要：板簧橡胶支座是重型卡车中的重要承载和减振部件，在循环加载工况下常出现疲劳寿命低和耐久性差的问题。本工作以板簧橡胶支座为研究对象，计算其垂向静刚度，并与试验刚度进行对照，通过支座的受力分布解释了刚度试验测试曲线呈现折线的原因。基于临界平面法和疲劳裂纹扩展理论，采用热力耦合疲劳寿命分析方法预测板簧橡胶支座在垂向260 kN压缩载荷下的动态温升量和疲劳寿命。结果表明仿真最高温度与试验温度误差为5.0%，疲劳寿命预测值与台架试验均值误差为7.5%。基于橡胶试样的伸张疲劳试验数据拟合出裂纹前兆体尺寸的双参数Weibull分布模型，推算出微裂纹初始尺寸范围为29.8～62.8 μm，中值为38.5 μm。基于裂纹前兆体尺寸的分布数据计算出的板簧橡胶支座疲劳仿真寿命在1.8万～3.9万次，定量解释了台架疲劳试验寿命的波动性。该工作可为减振类橡胶制品疲劳寿命的优化设计提供指导。

关 键 词：板簧橡胶支座；静刚度；疲劳寿命；Weibull分布；有限元分析
中图分类号：V255+.3; TQ333; TB39; TQ339; U467
文献标志码：A **文章编号**：1001-5965(2025)05-1726-09

随着车辆运输行业的快速发展，我国的重型汽车行业得到了大力支持，由于工业生产运输总量的增加与行驶路况、驾驶天气环境的恶劣，车辆在使役工况下将产生循环振动激励。板簧橡胶支座是由橡胶胶料与金属钢板粘合、硫化后，通过中间螺栓锁紧而形成的一种重型卡车用新型橡胶支座产品[1]，具有承载车体重量、阻尼支架底座中由路面诱发的振动，提高车辆的舒适性和安全性的作用[2-3]。重型卡车的行驶路况一般偏差，超载情况时有发生，使得支座会同时受到垂向压缩和切向剪切的冲击作用力[4-5]，导致支座的疲劳寿命短和耐久性差。通过台架疲劳试验发现板簧橡胶支座疲劳寿命的测试结果具有一定的波动性，这通常与初始微裂纹尺寸相关。Li等[6]的研究表明：弹性体材料的裂纹前兆体尺寸(c_0)通常在0.002～0.2 mm。

橡胶疲劳寿命的预测是基于疲劳裂纹扩展理论与临界平面分析方法[7]。采用Abaqus和Endurica联合仿真来预测橡胶制品的热力耦合疲劳寿命成为一种可被接受的方法[3,8-9]，目前许多开创性的工作已经被报道。Mohamedmeki等[10]探究了不同负载下钢层压轴承垫的疲劳寿命，其仿真计算结果与实验结果吻合性较高，建立了疲劳循环次数与施加载荷之间的幂函数关系。Nyaaba等[11]通过可靠的材料参数测试，对矿用卡车轮胎的疲劳寿命进行分析计算，讨论了充气压力、轮胎转速和轴重对轮胎各部位疲劳寿命的影响，其中带束层端部是裂纹成核的重要起始部位。Mars等[12]提出了临界平面分析方法，系统地研究了坦克履带橡胶垫板的自生热与疲劳寿命分布情况，得出了与试验相似的几种破坏模式，并给出加载温度、破坏模式和疲劳寿命的

收稿日期：2023-05-23；录用日期：2023-06-30；网络出版时间：2023-07-21 15:50
网络出版地址：link.cnki.net/urlid/11.2625.V.20230721.1439.001
*通信作者. E-mail：lifz@buct.edu.cn

引用格式：姬娜，刘娟，王浩然，等. 板簧橡胶支座刚度与疲劳寿命波动性仿真分析及试验研究[J]. 北京航空航天大学学报，2025，51(5)：1726-1734. JI N，LIU J，WANG H R，et al. Simulation analysis and experimental study on stiffness and fatigue life fluctuation of leaf spring rubber bearings[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1726-1734 (in Chinese).

真实估计。He 等[13] 系统地比较了三种实心轮胎有限元模型在不同滚动条件下的热力耦合仿真温度与滚动阻力, 发现拉格朗日模型的预测精度最高, 分析误差在 10% 以内。Robertson 等[14] 通过在橡胶材料中添加玻璃微珠和炭黑来模拟未均匀分散的夹杂物, 发现其对试样的裂纹前兆体、抗拉强度以及疲劳寿命的分布有重要影响。Liu 等[15] 基于 Power-law 自生热模型, 对试验测试下橡胶试样裂纹尖端的温度变化进行有限元仿真分析, 并采用热红外分析仪观测试样的温度变化结果, 合理解释了含边缘裂纹的橡胶试样自生热行为。目前, 大多数研究通常是以简单橡胶试样或轮胎等为主要研究对象, 鲜有工作聚焦重型卡车用板簧橡胶支座的疲劳失效研究上。

1　仿真分析方法与流程

本文以中国重型汽车集团有限公司提供的重卡用板簧橡胶支座为研究对象, 实际支座的三维几何结构如图 1 所示。

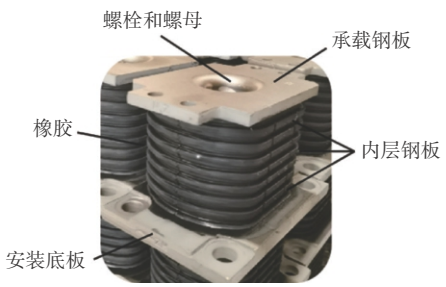


图 1　重卡用板簧橡胶支座部件及其结构组成

Fig. 1　Leaf spring rubber bearings used in heavy truck and its structural composition

板簧橡胶支座仿真分析流程如图 2 所示, 分析流程包含 3 个部分: 变形分析、生热分析和疲劳分析。

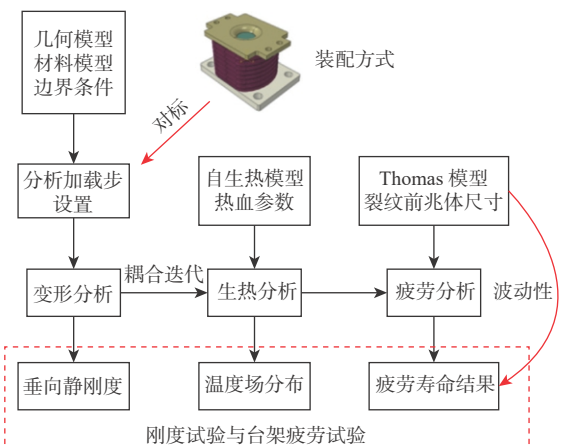


图 2　板簧橡胶支座的刚度及热力耦合疲劳寿命仿真分析流程图

Fig. 2　Simulation analysis flow chart of stiffness and thermo-mechanical coupling fatigue life of leaf spring rubber bearings

在变形分析中, 需要建立板簧橡胶支座的三维几何模型, 通过橡胶材料的拉伸回复试验数据拟合得到超弹性本构方程及相关材料参数。根据板簧橡胶支座的具体装配过程设置分析加载步, 计算静刚度并与试验刚度相比较, 提取出各单元应变分析结果, 用于后续生热分析与疲劳分析。

在生热分析中, 通过测试平面拉伸试样在不同峰值应变下的滞后能与稳态温度, 拟合得到 Power-law 自生热模型参数, 对板簧橡胶支座进行生热分析。在设置热边界条件与热参数后, 通过变形分析与生热分析耦合迭代, 能够得到支座在垂向加载工况下的稳态温度场分布。

在疲劳分析中, 基于疲劳裂纹扩展理论与临界平面法, 通过疲劳裂纹扩展试验(FCGR)与伸张疲劳测试, 拟合 Thomas 疲劳裂纹扩展模型参数并计算初始裂纹前兆体的尺寸。根据中值裂纹尺寸得到板簧橡胶支座的热力耦合疲劳平均寿命值, 并通过台架疲劳试验进行分析验证。

2　试验测试

2.1　试验刚度曲线及垂向静刚度

将注射成型后的板簧橡胶支座静置冷却至室温, 然后将支座预压 5 mm, 并通过螺栓和螺母拧紧焊死, 装配成最终装车用的板簧橡胶支座成品。通过万能电子试验机测试支座在特定下压位移量的反作用力曲线, 如图 3 所示。图 3 统计了 15 个板簧橡胶支座的力-位移曲线, 提取 60～80 kN 范围内第二段折线的斜率计算静刚度, 可知板簧橡胶支座静刚度为 20.4 ± 2.2 kN/mm。

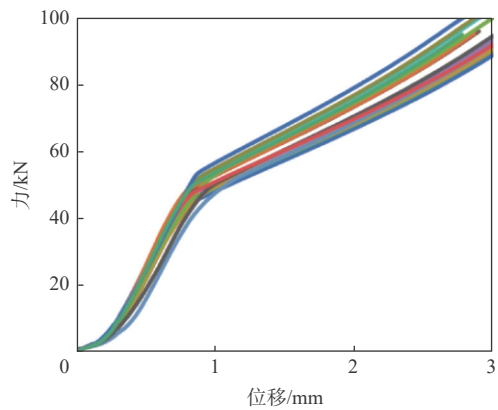


图 3　板簧橡胶支座刚度曲线的试验测试数据

Fig. 3　Test data of stiffness curves of leaf spring rubber bearings

2.2　台架疲劳试验

板簧橡胶支座的台架疲劳测试为单一的垂向循环加载测试。针对支座的垂向动态载荷从 5 mm 预压位移逐渐增大到 260 kN, 加载频率为 1 Hz。记录观察到出现明显裂纹时的循环加载次数即为板

簧橡胶支座的疲劳寿命。同时采用红外热成像仪测量初始环境温度和疲劳失效时部件的表面稳态温度。台架疲劳试验装置和典型的疲劳故障位置如图 4 所示。试验结果发现：板簧橡胶支座的疲劳寿命在 1.6 万～4.1 万次，疲劳寿命均值为 2.8 万次，疲劳开裂现象发生在与第 3、第 4 层内钢板外侧相连的橡胶区域，测量稳态条件下橡胶的外表面温度为 54 ℃。

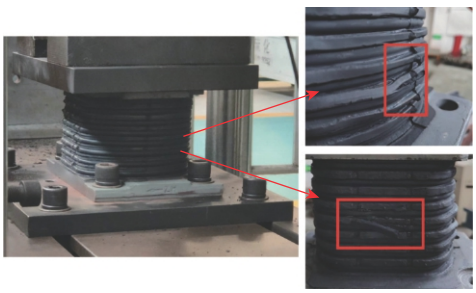


图 4 台架疲劳测试装置与疲劳破坏位置

Fig. 4 Bench fatigue test device and fatigue failure location

3 有限元仿真分析

3.1 板簧橡胶支座三维仿真模型

板簧橡胶支座是由注射胶料和七层不同孔径的金属钢板组合而成的一种橡胶减振制品，由承载钢板、安装底板、内钢板、橡胶部件、螺栓和螺母六大部分组成。板簧橡胶支座的三维几何模型如图 5 所示，其在装载过程中，安装底板被固定在车架上，因此在静态加载分析中，需对安装底板 6 个方向的自由度施加完全固定的约束。承载钢板使用节点耦合受力分析，通过承载钢板在 y 方向上施加压缩位移，模拟台架疲劳试验中的压缩过程。

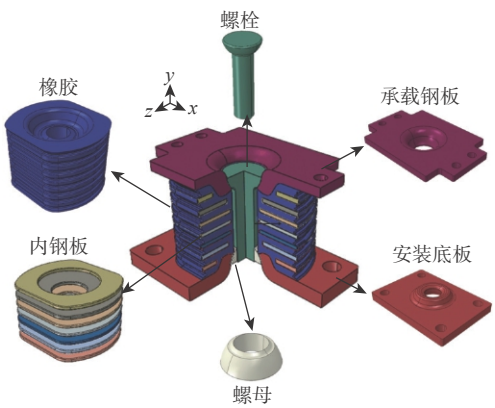


图 5 板簧橡胶支座三维几何模型

Fig. 5 Three-dimensional geometric model of leaf spring rubber bearings

板簧橡胶支座在仿真分析中的相互作用关系较为复杂，结构中存在多个接触面，包括：①橡胶上下表面与安装底板和承载钢板的接触面；②内钢板与橡胶内表面的接触面；③橡胶内表面与螺栓的接触面；④螺栓与承载钢板的接触面；⑤螺母与安装底板的接触面。承载钢板和安装底板均采用离散刚体表示，同时建立了内钢板、承载钢板和安装底板与橡胶部件间的绑定约束。整个模型的单元总数为 43.5 万个，其中橡胶部件占 30.7 万个，284 400 个网格采用六面体减缩积分杂交单元（C3D8RH）构建，23 280 个网格采用楔形杂交单元（C3D6H）构建。板簧橡胶支座的约束条件及几何网格模型如图 6 所示。

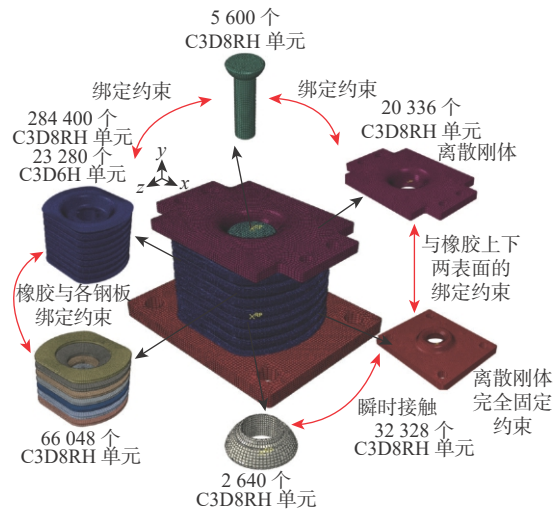


图 6 板簧橡胶支座的约束条件及几何网格模型

Fig. 6 Constraint conditions and geometric mesh model of leaf spring rubber bearings

3.2 仿真模型参数的设置

橡胶材料的本构方程由单轴 (UT)、平面 (PT) 和等双轴 (ET) 拉伸回复试验的应力-应变曲线拟合得到。通过前人[16]对超弹性本构方程的研究，发现下式的三阶 Ogden 本构方程在所有变形模式下，能够再现材料在各类变形模式下的力学响应。本构方程拟合曲线如图 7 所示，可决系数为 0.993 6。

$$W=\sum_{i=1}^{3}\frac{2\mu_i}{\alpha_i^2}\left(\bar{\lambda}_1^{\alpha_i}+\bar{\lambda}_2^{\alpha_i}+\bar{\lambda}_3^{\alpha_i}-3\right) \tag{1}$$

式中：W 为应变能密度；λ_1、λ_2 和 λ_3 分别为 3 个主方向的拉伸比；μ_i 和 α_i 为材料参数。拟合得到的超弹性本构方程参数如表 1 所示。

在动态循环加载下，橡胶材料内部各组分之间产生内摩擦，应变响应落后于应力响应，产生滞后温升现象[17]。本文通过 Power-law 自生热模型建立滞后能与峰值应变和表面温升之间的非线性关系。Power-law 自生热模型[12]如下式所示：

$$h=h_0\left(\frac{\varepsilon_{\mathrm{a}}}{\varepsilon_{\mathrm{a},0}}\right)^{\gamma}\mathrm{e}^{r(\theta-\theta_0)} \tag{2}$$

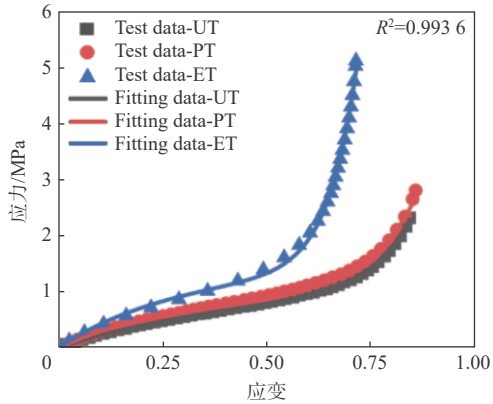


图 7　单轴、平面和等双轴拉伸的试验数据与三阶 Ogden 超弹性本构方程拟合曲线

Fig. 7　Tensile test data under uniaxial, plane, and equibiaxial modes and fitting data by using third-order Ogden hyperelastic constitutive equation

表 1　三阶 Ogden 超弹性本构方程参数

Table 1　Parameters of third-order Ogden hyperelastic constitutive equation

i	μ_i/MPa	α_i
1	1.094 1	0.636 24
2	$8.614\,4\times10^{-3}$	12.028 1
3	$1.416\,0\times10^{-7}$	13.337 1

式中: h为滞后能量密度; ε_a为应变幅值; θ为试样的表面稳态温度; γ为应变敏感参数; r 为温度敏感参数。$\varepsilon_{a,0}$为参比应变幅值; h_0和θ_0分别为参比应变幅值下的滞后能密度和表面稳态温度。仿真分析中, Power-law 自生热模型的相关参数可以通过 DMA+1000 动态热机械分析仪和热红外成像仪结合试验得到[18]。试验拟合的参数如表 2 所示。橡胶和钢板的热导率分别为 0.355 W/(m·K)和 49.8 W/(m·K), 比热容分别为 1 473 J/(kg·K)和 465 J/(kg·K)。

表 2　Power-law 自生热模型材料参数

Table 2　Material parameters of Power-law self-heating model

$\varepsilon_{a,0}$	h_0/(mJ·(mm^{-3}·cycle^{-1}))	θ_0/℃	γ	r/℃$^{-1}$
0.1	0.018 408 59	30.8	1.44	−0.007 55

Thomas 模型[19]是一类常见的橡胶疲劳裂纹扩展速率模型, 其表达式如下:

$$\frac{\mathrm{d}c}{\mathrm{d}N}=r_c\left(\frac{T}{T_c}\right)^F \tag{3}$$

式中: dc/dN为疲劳裂纹扩展速率; T为撕裂能, T_c为临界撕裂能; F为幂律指数; r_c为临界裂纹扩展速率。通过测量不同撕裂能下的疲劳裂纹扩展速率可得到模型参数。Thomas 模型的疲劳特性参数如表 3 所示。

表 3　Thomas 模型的疲劳特性参数

Table 3　Fatigue characteristic parameters of Thomas model

T_c/(kJ·m^{-2})	r_c/(m·cycle^{-1})	F
89.3	1.305×10^{-6}	1.62

4　结果与讨论

4.1　试验刚度与仿真刚度分析对比

试验静刚度的测量分为 3 个加载工况, 包括: ①将注射成型的制品预压 5 mm; ②插入螺栓螺母拧紧后通过焊接工艺牢固固定; ③持续加压测量曲线斜率得到垂向静刚度。基于试验加载工况, 在有限元仿真分析中相应地施加了 3 个分析加载步, 如图 8 所示。

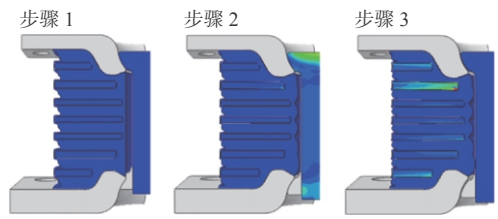


图 8　仿真分析中板簧橡胶支座的加载分析步

Fig. 8　Loading analysis steps for leaf spring rubber bearings in simulation analysis

针对试验过程中插入螺栓和螺母的这一操作, 在有限元仿真中难以在加载过程中通过装配新部件来模拟这一实际过程。因此, 需要在一开始就装配螺栓和螺母部件, 并设置相互绑定的接触关系。为了防止螺母与安装底板之间发生干涉, 在步骤 1 中无须建立二者间相互接触关系, 如图 8 的步骤 1 所示。当板簧橡胶支座预压到 5 mm 时, 到达图 8 的步骤 2 加载阶段, 此时建立了螺母和安装底板的瞬时接触关系。由于板簧橡胶支座特殊的结构设计, 螺母可以很好地固定在底板下方。这一加载分析步的设置成功地模拟了螺栓和螺母的拧紧焊死过程。最后持续下压至 7.5 mm 的位移量, 进行静刚度的分析计算, 如图 8 的步骤 3 所示。图 9 为整个有限元仿真分析过程中的板簧橡胶支座的力-位移曲线。

根据图 9 可以发现, 随着 Step-2 分析步中螺母与安装底板之间建立接触关系, 承载钢板上的反作用力瞬间降至 0 kN, 即无须外加作用力即可固定板簧橡胶支座的预压 5 mm 的装配结构。计算 60～80 kN 下的刚度曲线斜率以表示板簧橡胶支座的仿真静刚度, 得到仿真分析下的刚度为 19.5 kN/mm, 与试验刚度的平均值相比, 二者误差仅为 4.4%。

在对刚度曲线的分析中, 发现了一个有趣的现象, 即力-位移曲线存在拐点, 本文将借助有限元仿

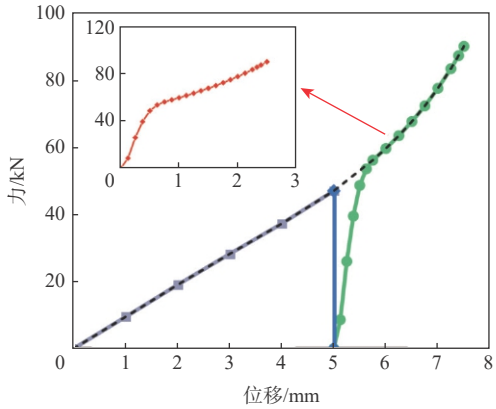


图9 板簧橡胶支座的仿真刚度曲线

Fig. 9 Simulated stiffness curve of leaf spring rubber bearings

真分析阐明其背后原因。由于装车用板簧橡胶支座经过预压和螺栓螺母拧紧定位的工序，在图8的步骤2加载阶段，本应由承载钢板承担的反作用力转移至螺栓螺母组合结构中，使得步骤2加载阶段的初始时刻，承载钢板上的外力为0 kN。随着下压位移量的增加，螺栓螺母组合结构承受的反作用力逐渐降低，承载钢板上的外力逐渐增加，两者的合力即为板簧橡胶支座在承载该下压位移量时所需的总压力值，对应图9所示的虚线部分。而当板簧橡胶支座下压到一定位移量时，螺母完全脱离安装底板，此时螺栓螺母组合结构不再受反作用力的作用，板簧橡胶支座在该下压量的压力值完全由承载钢板提供。继续下压出现如图9所示的曲线拐点现象，拐点之后的曲线斜率即可用于表示板簧橡胶支座的垂向静刚度。这也直接证明了试验测试条件下，将60～80 kN的曲线斜率定义为板簧橡胶支座垂向刚度的合理性。通过分析可知，刚度曲线的折线是由预压5 mm的工艺造成的结构受力引起的。从某种层面上来讲，这使得橡胶支座在动态服役工况下，橡胶材料始终处在非完全松弛的动态振动加载条件下。支座橡胶材料是由具有明显应变诱导结晶型的天然橡胶材料组成的，非完全松弛的动态振动加载条件下天然橡胶的疲劳耐久性表现较佳，这有利于板簧橡胶支座的疲劳耐久性提升。

4.2 疲劳寿命仿真分析与验证

本节基于哑铃型橡胶试样的伸张疲劳测试结果，结合超弹性本构方程与Thomas模型参数，计算得到橡胶材料的初始裂纹尺寸。伸张疲劳测试结果如表4所示，其中编号为1#～8#的试样为峰值应变为150%时的疲劳测试数据，疲劳寿命在36 102～52 374次；编号为9#～16#的试样为峰值应变为200%时的疲劳测试数据，疲劳寿命在11 587～22 389次。结合疲劳寿命测试结果与下式可以计算橡胶材料的裂纹前兆体尺寸c_0[20]：

表4 哑铃型橡胶试样的裂纹前兆体尺寸

Table 4 Crack precursor size of dumbbell-shaped rubber specimens

样品编号	峰值应变/%	疲劳寿命/cycle	裂纹前兆体尺寸/μm
1#	150	36 102	54.25
2#	150	44 268	39.04
3#	150	39 948	46.08
4#	150	49 951	32.13
5#	150	36 562	53.15
6#	150	52 374	29.77
7#	150	41 925	42.62
8#	150	49 075	33.03
9#	200	11 587	62.76
10#	200	16 351	36.01
11#	200	13 872	46.95
12#	200	22 389	21.69
13#	200	16 342	36.04
14#	200	17 197	33.20
15#	200	12 927	52.60
16#	200	19 248	27.68

$$c_0=\left[\frac{N_{\mathrm{f}}(F-1)r_{\mathrm{c}}(2\pi W_{\max})^F}{T_{\mathrm{c}}^F}\right]^{\frac{1}{1-F}} \tag{4}$$

式中：N_{f}为试样的伸张疲劳寿命；c_0为初始裂纹前兆体尺寸；$W_{\max}$为伸张疲劳测试峰值应变水平下的应变能密度。

结果表明，裂纹前兆体的尺寸分布在29.8～62.8 μm，可以发现裂纹前兆体的尺寸是一个范围而非确定的单一数值，本工作中橡胶材料裂纹前兆体尺寸的最大值是最小值的2.9倍。从概率学的角度对橡胶制品的初始裂纹尺寸进行推算更具有科学性。

在橡胶裂纹前兆体尺寸的分析中，Weibull分布函数[21]可以用来定量描述疲劳故障率F与裂纹尺寸c_0的关系规律，模型表达式如下：

$$F(x)=1-\mathrm{e}^{-\left(\frac{x}{\eta}\right)^{\beta}} \tag{5}$$

式中：$F(x)$为系统的失效率（failure rate function）的函数；x为随机变量；β为形状参数；η为缩放因子。基于橡胶单轴拉伸疲劳试验的测试数据，推导双参数Weibull分布模型对应的裂纹尺寸分布函数，并将其应用到板簧橡胶支座疲劳寿命的预测中。通过表4的数据对Weibull分布模型拟合，得到如图10所示的拟合曲线。其中β=3.61，η=0.043，可决系数R^2=0.978。当F=0.5时，计算得到橡胶材料的中值裂纹尺寸c_0=38.5 μm。

基于初始裂纹的中值尺寸数据，对板簧橡胶支座在260 kN垂向压缩载荷下的滞后温升和疲劳寿

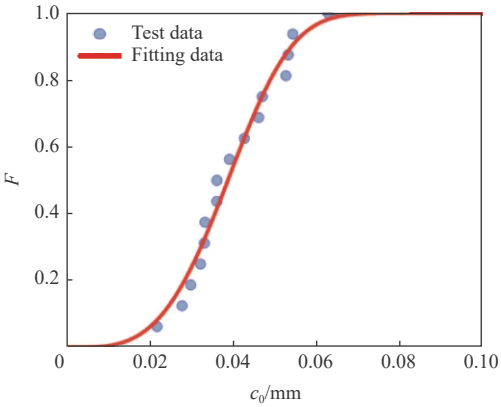


图 10　疲劳故障率试验数据与 Weibull 分布函数拟合曲线

Fig. 10　Fitting curve of fatigue failure rate test data with Weibull distribution function

命进行仿真分析, 并与台架疲劳结果对比。260 kN 压缩载荷下板簧橡胶支座的应变分布云图如图 11 所示。最大主应变为 119%, 应变集中主要出现在内钢板左右两侧边缘区域以及橡胶与安装底板和承载钢板接触的外边缘处。

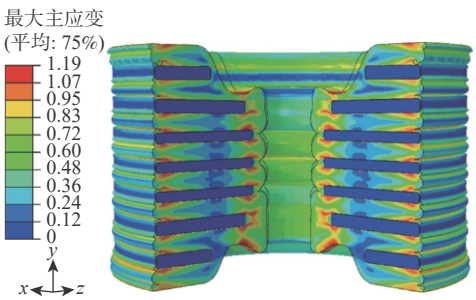


图 11　板簧橡胶支座的应变分布云图

Fig. 11　Contour plot of strain of leaf spring rubber bearings

对板簧橡胶支座的稳态温度和疲劳寿命进行仿真分析。结合试验测试时的环境温度, 将初始温度场设置为 32℃。图 12 和图 13 分别显示了板簧橡胶支座在垂直压缩 260 kN 下的稳态温度分布和疲劳寿命分布云图。

板簧橡胶支座达到稳态时的自生热温度的表面最高仿真结果为 56.7℃, 与表面温度最高值的试验结果相比, 误差为 5.0%。板簧橡胶支座的最小疲

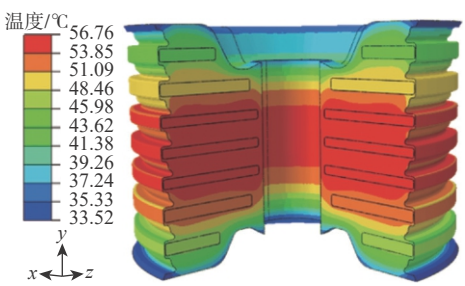


图 12　板簧橡胶支座的稳态自生热温度分布云图

Fig. 12　Contour plot of self-heating steady state temperature distribution of leaf spring rubber bearings

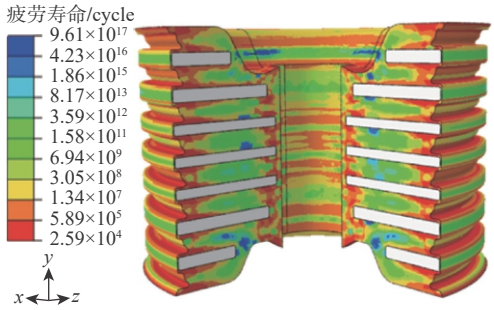


图 13　板簧橡胶支座的疲劳寿命分布云图

Fig. 13　Contour plot of fatigue life distribution of leaf spring rubber bearings

劳寿命为 2.59×10^4 次, 与试验测得的中值疲劳寿命的误差为 7.5%。试验数据与仿真数据的对比说明了本工作中的热力耦合疲劳寿命仿真分析方法是合理的。

需要进一步解释说明的是, 台架疲劳试验结果如图 14 所示, 通过统计 15 个板簧橡胶支座的台架疲劳测试结果发现, 板簧橡胶支座的疲劳寿命在 1.6×10^4～4.1×10^4 次, 波动性达±43.9%。

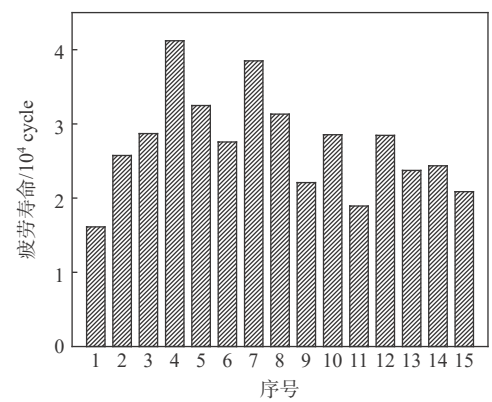


图 14　板簧橡胶支座在 260 kN 垂向压缩载荷下的台架疲劳寿命

Fig. 14　Bench fatigue life of leaf spring rubber bearings under a vertical compressive load of 260 kN

结合裂纹前兆体尺寸的概率分布结果, 计算了板簧橡胶支座在垂向 260 kN 压缩载荷下的热力耦合疲劳寿命的分布情况, 如表 5 所示。

仿真计算结果表明, 裂纹前兆体的尺寸与疲劳耐久性有直接影响, 裂纹尺寸越大, 板簧橡胶支座的疲劳寿命越短。疲劳寿命在 1.8×10^4～3.9×10^4 次之间, 仿真分析下疲劳寿命的波动性与台架疲劳试验结果具有较高的一致性。裂纹前兆体尺寸的不均一性与橡胶原材料及其在加工过程中的混炼工艺、气泡杂质等因素密切相关[22], 进而造成橡胶制品台架疲劳试验结果的波动性。裂纹前兆体尺寸与疲劳寿命的非线性拟合关系如图 15 所示, 拟合精度较高($R^2>0.998$), 这一关系也可用于板簧橡胶

表 5 板簧橡胶支座的疲劳寿命与裂纹前兆体尺寸

Table 5 Fatigue life and crack precursor size of leaf spring rubber bearings

样品编号	裂纹前兆体尺寸/μm	疲劳寿命/cycle
1#	54.25	20 080
2#	39.04	25 600
3#	46.08	22 670
4#	32.13	29 240
5#	53.15	20 420
6#	29.77	31 200
7#	42.62	24 050
8#	33.03	28 932
9#	62.76	18 019
10#	36.01	27 100
11#	46.95	22 400
12#	21.69	38 920
13#	36.04	27 100
14#	33.20	28 360
15#	52.60	20 940
16#	27.68	33 480

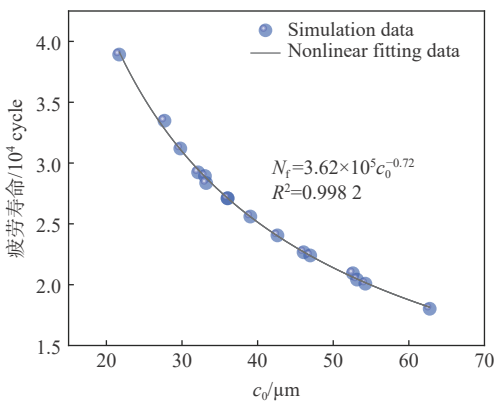


图 15 板簧橡胶支座疲劳寿命与裂纹前兆体尺寸之间的非线性关系

Fig. 15 Non-linear relationship for fatigue life and crack precursor size of leaf spring rubber bearings

支座在简单压缩工况下疲劳寿命的快速预测。

5 结 论

1）根据中国重型汽车集团有限公司提供的结构参数对板簧橡胶支座进行三维模型的构建与加载分析步的设置，通过静态垂向压缩的仿真计算，得到板簧橡胶支座的静刚度为 19.5 kN/mm，与试验测试刚度的误差仅有 4.4%，并基于不同仿真时刻支座的受力特点揭示了刚度曲线出现拐点的原因。

2）在 Abaqus 和 Endurica 联合仿真模拟的基础上，利用台架疲劳试验机对板簧橡胶支座的仿真疲劳寿命进行分析验证。在垂向 260 kN 的压缩载荷下板簧橡胶支座的仿真疲劳寿命为 2.59×10^4 次循环，稳态温度为 56.7℃，与试验测试平均结果相比，误差分别为 7.5% 和 5.0%。证明了热力耦合疲劳仿真寿命有限元分析方法的有效性及材料测试参数的准确性。

3）基于伸张疲劳的试验数据，通过两参数 Weibull 分布模型系统地研究了橡胶材料的裂纹前兆体的尺寸及其分布，并计算出不同裂纹前兆体下板簧橡胶支座的疲劳寿命结果，合理地解释了台架疲劳测试的结果出现波动性的原因。

4）本文在可靠几何模型和试验测试的基础上，探究疲劳寿命出现波动性的原因。该工作可对轮胎、球铰、空气弹簧等其他弹性元件的刚度和疲劳寿命的研究提供指导。

参考文献（References）

[1] 李腾英, 郑洪喜. 橡胶支座总成注射模的设计[J]. 特种橡胶制品, 2014, 35(5): 65-70.
LI T Y, ZHENG H X. Design of injection mold for rubber bearing assembly[J]. Special Purpose Rubber Products, 2014, 35(5): 65-70 (in Chinese).

[2] ZARRIN-GHALAMI T, FATEMI A, LEE Y L. Fatigue life prediction of an automobile cradle mount[J]. SAE International Journal of Passenger Cars - Mechanical Systems, 2013, 6(1): 351-359.

[3] ZINE A, BENSEDDIQ N, NAÏT ABDELAZIZ M. Rubber fatigue life under multiaxial loading: numerical and experimental investigations[J]. International Journal of Fatigue, 2011, 33(10): 1360-1368.

[4] 丁智平, 陈吉平, 宋传江, 等. 橡胶弹性减振元件疲劳裂纹扩展寿命分析[J]. 机械工程学报, 2010, 46(22): 58-64.
DING Z P, CHEN J P, SONG C J, et al. Analysis of fatigue crack growth life for rubber vibration damper[J]. Journal of Mechanical Engineering, 2010, 46(22): 58-64(in Chinese).

[5] KANSAKE B A, FRIMPONG S. Analytical modelling of dump truck tire dynamic response to haul road surface excitations[J]. International Journal of Mining, Reclamation and Environment, 2020, 34(1): 1-18.

[6] LI F Z, LIU J P, MARS W V, et al. Crack precursor size for natural rubber inferred from relaxing and non-relaxing fatigue experiments[J]. International Journal of Fatigue, 2015, 80: 50-57.

[7] BARBASH K P, MARS W V. Critical plane analysis of rubber bushing durability under road loads[R]. SAE Technical Paper, 2016.

[8] MARS W V, FATEMI A. Multiaxial fatigue of rubber: Part II: experimental observations and life predictions[J]. Fatigue & Fracture of Engineering Materials & Structures, 2005, 28(6): 523-538.

[9] ZINE A, BENSEDDIQ N, ABDELAZIZ M N, et al. Prediction of rubber fatigue life under multiaxial loading[J]. Fatigue & Fracture of Engineering Materials & Structures, 2006, 29(3): 267-278.

[10] MOHAMEDMEKI Z M, ESMAIL F J, AJEEL E A. Fatigue life analysis of laminated elastomeric bearing pad[J]. Materials Today: Proceedings, 2021, 42: 2361-2368.

[11] NYAABA W, BOLARINWA E O, FRIMPONG S. Durability prediction of an ultra-large mining truck tire using an enhanced finite

element method[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2019, 233(1): 161-169.

[12] MARS W V, CASTANIER M, OSTBERG D, et al. Digital twin for tank track elastomers: predicting self-heating and durability[C]// Proceedings of the 2017 Ground Vehicle Systems Engineering and Technology Symposium (GVSETS), 2017: 1-10.

[13] HE H, LIU J M, ZHANG Y R, et al. Heat build-up and rolling resistance analysis of a solid tire: experimental observation and numerical simulation with thermo-mechanical coupling method[J]. Polymers, 2022, 14(11): 2210.

[14] ROBERTSON C G, TUNNICLIFFE L B, MACIAG L, et al. Characterizing distributions of tensile strength and crack precursor size to evaluate filler dispersion effects and reliability of rubber[J]. Polymers, 2020, 12(1): 203.

[15] LIU C, GU B C, CHEN J F, et al. Thermo-mechanical coupling analysis of edge-cracked rubber specimen focusing on the crack tip: experimental observation and numerical simulation[J]. Materials Today Communications, 2022, 31: 103348.

[16] HE H, ZHANG Q, ZHANG Y R, et al. A comparative study of 85 hyperelastic constitutive models for both unfilled rubber and highly filled rubber nanocomposite material[J]. Nano Materials Science, 2022, 4(2): 64-82.

[17] WILLETT P R. Heat generation in tires due to the viscoelastic properties of elastomeric components[J]. Rubber Chemistry and Technology, 1974, 47(2): 363-375.

[18] LIU C, GU B C, WANG F, et al. Waveform impact on thermo-mechanical fatigue crack growth of a non-crystallizing rubber: experimental observation and numerical simulation[J]. Composites Part B: Engineering, 2023, 255: 110604.

[19] THOMAS A G. Rupture of rubber. V. Cut growth in natural rubber vulcanizates[J]. Journal of Polymer Science, 1958, 31(123): 467-480.

[20] GUO H, LI F Z, WEN S P, et al. Characterization and quantitative analysis of crack precursor size for rubber composites[J]. Materials, 2019, 12(20): 3442.

[21] BHATTACHARYA P, BHATTACHARJEE R. A study on weibull distribution for estimating the parameters[J]. Wind Engineering, 2009, 33(5): 469-476.

[22] HUNEAU B, MASQUELIER I, MARCO Y, et al. Fatigue crack initiation in a carbon black-filled natural rubber[J]. Rubber Chemistry and Technology, 2016, 89(1): 126-141.

Simulation analysis and experimental study on stiffness and fatigue life fluctuation of leaf spring rubber bearings

JI Na[1], LIU Juan[3], WANG Haoran[3], GAO Rui[3], LU Yonglai[1, 2], LI Fanzhu[1, 2, *]

(1. State Key Laboratory of Organic-Inorganic Composites, Beijing University of Chemical Technology, Beijing 100029, China;
2. College of Materials Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China;
3. China National Heavy Duty Truck Group Co., Ltd., Jinan 250031, China)

Abstract: The leaf spring rubber bearings are important load-bearing and vibration-absorbing components in heavy trucks and often suffer from short fatigue life and poor durability under cyclic loading conditions. This work took the leaf spring rubber bearing as the research object, calculated its vertical static stiffness, and compared it with the experimental stiffness. The reason why the experimental stiffness curve is folded was explained through bearing forces. Based on the critical plane method and fatigue crack propagation theory, the dynamic heat build-up and fatigue life of the leaf spring rubber bearing under a vertical compressive load of 260 kN were predicted by using a thermo-mechanical coupling analytical approach. The results show that the error between the simulated maximum temperature and the experimental temperature is around 5.0%, and the error between the predicted fatigue life and the mean value of the bench test is 7.5%. Based on the tensile fatigue test data of rubber specimens, a two-parameter Weibull distribution model for crack precursor size is fitted, and the initial size range of microcracks is inferred to be between 29.8 μm and 62.8 μm, with a median of 38.5 μm. The fatigue simulation life of the leaf spring rubber bearing calculated based on the distribution data of the crack precursor size ranges from 18 000 to 39 000 cycles, quantitatively explaining the fluctuation of the bench fatigue test life. This work can provide guidance for the optimal design of the fatigue life of vibration-absorbing rubber products.

Keywords: leaf spring rubber bearing; static stiffness; fatigue life; Weibull distribution; finite element analysis

Received: 2023-05-23; **Accepted**: 2023-06-30; **Published Online**: 2023-07-21 15:50
URL: link.cnki.net/urlid/11.2625.V.20230721.1439.001

* **Corresponding author**. E-mail: lifz@buct.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0271

可变拍幅扑翼飞行器模型设计与气动特性分析

梁金泽[1]，潘天宇[1,2]，郑孟宗[1,*]，彭连松[2]，曹梦达[1]

(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 能源与动力工程学院，北京 102206)

摘　　要：目前微型扑翼飞行器(MAV)在军事以及民用方面都正在普及，发挥着越来越重要的作用。扑翼生物在飞行过程中展现了其气动效率高、动作敏捷、可稳定悬停等优点。本文以扑翼昆虫为研究对象，提出了一种新型的扑翼飞行器，设计了一种可变旋转轨道半径的扑翼机械结构模型和控制系统。进行旋转半径变化时，芯片控制变轨电机转动，进而改变变位滑块距离旋转中心的位置，通过传动机构最终表现为扑翼拍动幅度的改变。针对设计的机械结构，进行坐标系建立和表达式的推导，同时利用XFlow进行流体仿真试验，通过基准试验、双翼变幅值仿真流体试验以及单翼变幅值仿真流体试验，得到相应的结论，验证了可变拍幅扑翼飞行器模型的合理性和可行性，为后续研究及样机制作提供了数据支持。

关 键 词：机械设计；可变幅值；扑翼翻转；多模式飞行；流场模拟；动力分析

中图分类号：V211.7

文献标志码：A　　**文章编号**：1001-5965(2025)05-1735-12

近年来，众多学者和工程技术人员意识到微型无人飞行器的广阔应用前景，军事科学家也越来越深刻地认识到微型无人飞行器在未来数字化、信息化战场中不可替代的作用。由于体积小，微型无人飞行器可以进入复杂的空间执行特殊任务，在救灾、勘探、侦察等领域具有明显的优势。

从工程应用的角度来看，扑翼机构与旋翼机构、固定翼机构等相比有显著灵活性、机动性和隐蔽性。同时随着机构规模的缩小，旋翼与固定翼产生升力的效率大大降低，而扑翼模式产生升力具有强的非定常性从而具有较高的效率[1]，扑翼模式的非定常气动机制有：Weis-Fogh机制、延时失速机制、尾迹捕获机制、翻转机制[2]。自然界中昆虫扑翼是拍动和翻转两种运动的耦合，扑翼运动规律的改变对产生的升力大小有较大影响。

传统的扑翼飞行器使用直流电机驱动单曲柄双摇杆机构产生升力，如孙卫等[2]采用传统的单曲柄结构双摇杆结构，通过电机直接驱动齿轮旋转，带动连杆的摆动从而实现固定轨迹的扑翼运动。为了提高运动的稳定性，Dileo等[3]在传统的单曲柄双摇杆机构基础上，增设滑槽装置，使滑块在滑槽中往复周期运动从而带动扑翼进行拍动。张焱等[4]在研究扑翼飞行器的飞行性能时，从单曲柄双摇杆机构出发，经过优化设计最终使用五杆的曲柄摇杆机构作为传动机构。胡超等[5]考虑到设计结构的难易程度，确定采用单曲柄双摇杆机构作为仿鸟扑翼飞行器的机械传动机构。这种机构虽然结构简单，但长时间拍动时，两翼容易产生相位差，两翼间拍动的相位差会使两翼气动力变化规律不同步，引起产生的升力不足进而导致扑翼飞行器无法正常飞行。

针对以上拍动幅值不可变的扑翼飞行器，近些年国内外研究重心逐步转移到可变拍动幅值的扑翼飞行器。与幅值不可变的扑翼飞行器不同，可变

收稿日期：2023-05-24；**录用日期**：2023-07-28；**网络出版时间**：2023-08-15 08:42
网络出版地址：link.cnki.net/urlid/11.2625.V.20230814.0937.002
基金项目：国家自然科学基金(52006002,51976005,52206040)
*通信作者. E-mail：zhengmengzong@buaa.edu.cn

引用格式：梁金泽，潘天宇，郑孟宗，等. 可变拍幅扑翼飞行器模型设计与气动特性分析[J]. 北京航空航天大学学报，2025，51(5)：1735-1746. LIANG J Z，PAN T Y，ZHENG M Z，et al. Model design and aerodynamic characteristic analysis of variable-amplitude flapping wing aircraft[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1735-1746 (in Chinese).

拍幅扑翼飞行器的设计难点不仅仅在机械结构和传动设计，更重要的是对控制系统的设计。如侯月阳等[6]提出的可变幅扑翼飞行器，采用辅助电机实现对扑翼幅值的改变，但该设计未进行运动姿态的实时监测反馈并且扑翼拍动幅值不能实现连续变化。Sato等[7]研制的具有变幅连接机构的三自由度扑翼机器人，在一定程度上验证了幅值改变可以实现对气动力的调节，但是未设计翻转机构，仅靠拍动获得的升力变化范围很小，且很难获得超过其自身重量的升力。Kumar等[8]研制的可变拍动幅值的仿海鸥无人机，通过舵机的牵引改变推动杆的运动行程，进而改变拍动幅值，扑翼采用串联式控制，每个翅膀均安装三个控制舵机来控制扑翼的三个自由度，机构冗杂且产生的气动力远不能达到飞行的要求。林镇炜等[9]设计的一种攻角可调的双段翼仿生扑翼飞行器采用舵机配合8051单片机进行实时攻角调节，这种设计思想可以较好地满足飞行器对灵活性和准确性的要求，但设计的控制系统模块较多，对主芯片运算性能提出较高要求。张亚锋等[10]通过实验证明扑动幅值角对仿生扑翼气动力特性的影响，进一步阐述了扑翼飞行器应对不同飞行情况时扑翼变幅拍动的重要性，为设计可变拍动幅值的扑翼飞行器提供了实验基础。综上，变幅值拍动的价值在于，扑翼飞行器通过变拍动幅值能够对气动力进行控制，从而实现多种飞行模式，提高了扑翼飞行器的灵活性和环境适应性。

本文以仿生学为切入点，首先设计幅值可变的单扑翼结构，扑翼机构可以完成拍动带动翻转的运动，可实现对拍动和翻转运动的主动控制，与真实扑翼运动规律相符，其次对扑翼机构的控制系统进行设计，包括控制原理、硬件选型和电路设计。最后对扑翼模型进行流体仿真试验，试验结果表明，通过改变拍动幅值可以实现扑翼飞行器的平稳飞行、起飞降落、转向的功能，从而验证了可变拍幅扑翼机器人设计的合理性和可行性。

1 机械传动机构

1.1 建立机器人坐标系

为了准确地描述扑翼飞行器运动，在设计模型中采取了两种机器人坐标系：惯性坐标系$Oxyz$和两个动坐标系$Oxyz$、$O'x'y'z'$。惯性坐标系描述扑翼飞行器整体的运动情况，惯性坐标系的原点为扑翼飞行器的重心，x轴方向为飞行器产生水平推力方向，z轴方向为飞行器产生竖直升力方向，由坐标系右手螺旋定则，可以确定y轴方向。两个动坐标系描述左右两翼的局部运动，两个原点O、O'分别为左右两翼运动时翼根的固定点，y轴和y'轴与扑翼上的传动杆同向，z轴和z'轴分别垂直于y轴和y'轴并且分别与左右扑翼面平行，x轴和x'轴由右手螺旋定则可以确定。扑翼飞行器坐标系如图1所示。

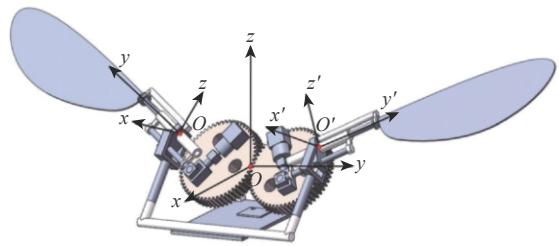


图1 飞行器坐标系示意图

Fig. 1 Coordinate system of aircraft

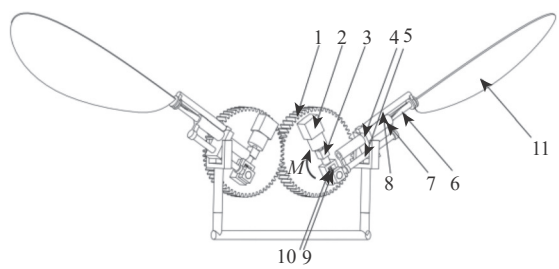
1.2 扑翼参数定义

定义拍动频率为10 Hz，用拍动角α和翻转角β来描述扑翼的运动。拍动角α定义为：扑动时Oy轴在yOz($y'O'z'$)平面的投影与Oy($O'y'$)轴的瞬时夹角($0° \leqslant |\alpha| \leqslant 90°$)。翻转角$\beta$定义为：扑动时$yOz$($y'O'z'$)平面与$yOz$平面的瞬时夹角($0° \leqslant |\beta| \leqslant 90°$)。

1.3 传动机构设计

1.3.1 整体机构设计

图2为飞行器整体机构示意图。因整体机构左右对称，图中只对一侧机构进行标注。



1. 传动齿轮；2. 变轨电机；3. 旋转装置；4. 摆动装置；5. 传动杆；6. 翻转运动轴；7. 联动架；8. 扑翼固定轴；9. 变位滑块；10. 丝杆；11. 扑翼。

图2 整体机构示意图

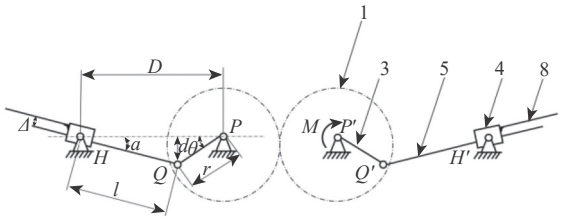
Fig. 2 Overall mechanism of aircraft

1.3.2 拍动机构设计

使用双曲柄-双摆杆机构作为幅值可变扑翼飞行器的拍动机构。图3为双曲柄-双摆杆机构的示意图。

其中P、P'、Q、Q'、H、H'为铰接点符号。图3中其他各参数的含义见表1。

两齿轮啮合运动，左右翼运动规律相同。拍动运动过程为：在主电机的带动下，传动齿轮开始以角速度ϕ开始绕P(P')转动，旋转装置一端插入齿轮中心，一端有滑槽放置变位滑块，固定在齿轮上的旋转装置带动变位滑块随之以角速度ω开始转动。传动杆一端与变位滑块进行铰接，铰接点为Q



1. 传动齿轮; 3. 旋转装置; 4. 摆动装置;
5. 传动杆; 8. 扑翼固定轴。

图 3　双曲柄-双摆杆机构示意图

Fig. 3　Double crank-double pendulum mechanism

表 1　传动机构参数表

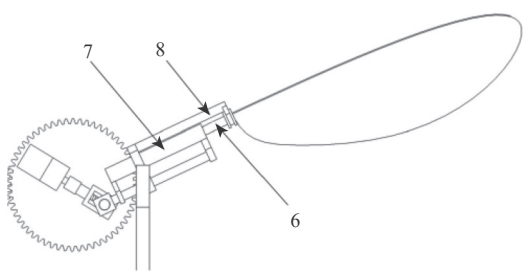
Table 1　Transmission mechanism parameters

变量符号	变量解释
θ	齿轮转角
α	拍动角
D	齿轮中心与摆动装置距离
r	旋转半径
l	连杆有效长度
Δ	扑翼与连杆的偏移距离
d	过渡参数

(Q'), 在转动运动的带动下, 传动杆另一端在摆动装置内做往复运动, 摆动装置进而绕铰接点H(H')进行摆动。扑翼固定轴与摆动装置固定连接, 二者可视为一体进行运动。扑翼与扑翼固定轴沿轴向相对位移为 0, 扑翼固定轴摆动带动扑翼实现拍动。

1.3.3　翻转机构设计

翻转机制对于真实的昆虫扑翼产生升力具有重要意义。为了实现对昆虫扑翼的仿生, 本文采用拍动带动翻转的设计, 使得扑翼能够按照已知的翻转规律实现主动翻转。翻转机构的示意图如图 4 所示。



6. 翻转运动轴; 7. 联动架; 8. 扑翼固定轴。

图 4　翻转机构

Fig. 4　Flip mechanism

联动架结构示意图如图 5 所示, 联动架一端为通孔, 与传动杆的一端紧固配合, 联动架另一端为一连杆端盖, 传动杆的另一端插入端盖固定, 联动架长轴通孔内有一固定圆柱体, 与翻转运动轴配合运动。

翻转运动轴结构示意图如图 6 所示, 上部与扑翼相连, 翻转运动轴包括 6-1 旋转卡槽和 6-2 滑槽两部分, 旋转卡槽与扑翼旋转轴的通孔一侧同心配合, 翻转运动轴可以进行绕轴线的转动, 旋转卡槽起到轴向固定和周向旋转的作用。滑槽的平面展开图形状依据翻转角的运动规律进行形状设计。

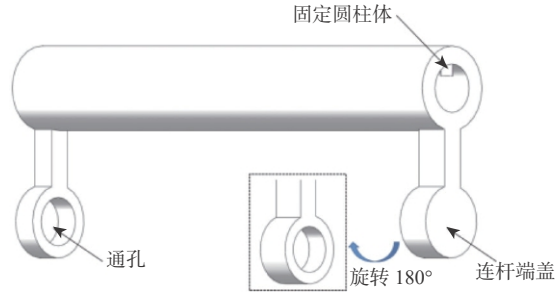


图 5　联动架示意图

Fig. 5　Linkage frame

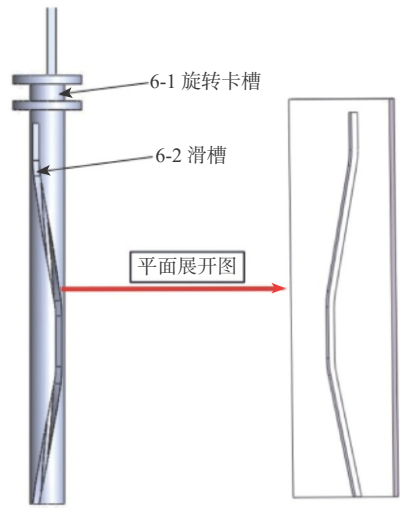


图 6　翻转运动轴示意图

Fig. 6　Inverted axis of motion

翻转运动过程为: 当传动杆在摆动装置圆孔内运动时, 带动联动架进行相同规律的运动, 联动架上的固定圆柱体在翻转运动轴的滑槽内滑动。翻转运动轴被轴向固定, 只可进行绕轴线的转动, 当联动架进行往复运动时, 固定圆柱体也进行往复运动, 因为固定圆柱体紧贴滑槽, 所以翻转运动轴在固定圆柱体接触力的作用下, 按照滑槽的轨迹进行转动, 进而带动扑翼实现满足需求的翻转运动。

本设计提出的扑翼机构, 可实现拍动带动翻转的主动控制运动规律。

1.4　运动学分析

依据机构运动简图, 对扑翼机构进行运动学分析:

$$D = r\cos\theta + l\cos\alpha \tag{1}$$

$$d = r\sin\theta = l\sin\alpha \tag{2}$$

$$\alpha = \arcsin(r\sin\theta / l) \tag{3}$$

$$l = \sqrt{r^2 + D^2 - 2Dr\cos\theta} \tag{4}$$

$$\theta = \omega t \tag{5}$$

推导出拍动角的公式：

$$\alpha = \arcsin\left(\frac{r\sin\theta}{\sqrt{r^2 + D^2 - 2Dr\cos\theta}}\right) \tag{6}$$

其中$D = 32$ mm，$\omega = 10$ rad/s。

$r = r_{\max} = 17$ mm时，$\alpha \in [-20°, 20°]$。

$r = r_{\min} = 13$ mm时，$\alpha \in [-25°, 25°]$。

翻转角的运动规律由图6设计的滑槽形状决定，依据本课题组的研究，扑翼的翻转角度曲线随时间变化呈现为梯形，因此所设计滑槽平面展开图的形状也大致为梯形。

对图2所示机构进行运动学分析，利用ADAMS中动力学模块进行仿真，将各零件导入，并施加约束与驱动，得到单周期内扑翼的仿真动画分解如图7所示，便于观察扑翼的运动，图7中扑翼机器人只保留扑翼部分。

通过运动仿真求得扑翼飞行机器人的翼尖轨迹如图8所示。

由图8的仿真结果可以看出翼尖轨迹为近似三维空间的“8”字形，符合真实扑翼的运动轨迹，

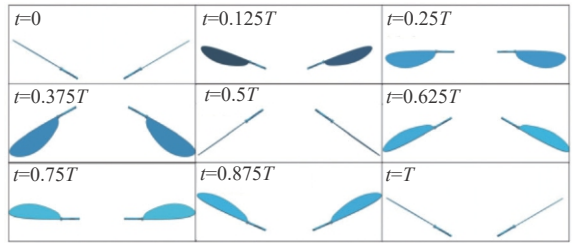


图7 单周期扑翼运动分解图

Fig. 7 Decomposition of single-period flapping wing motion

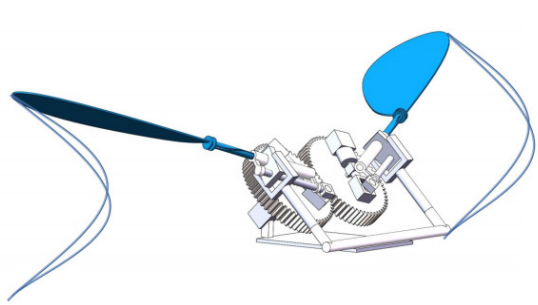


图8 ADAMS模型的翼尖轨迹曲线

Fig. 8 Wingtip trajectory curve of ADAMS model

验证了设计模型的合理性与可行性。

1.5 扑翼设计

飞行生物的飞行性能与自身质量、扑翼频率、翼展等参数有特定的关系，这种关系定义为尺度率，如表2所示，它反映了飞行生物各参数之间的函数关系[11]。

表2 尺度率参考表

Table 2 Scale rate reference

生物种类	扑翼面积/m^2	翼展/m	展弦比	翼载荷/(N·m^{-2})
昆虫及鸟类(除蜂鸟)	$\propto 0.16m^{0.72}$	$\propto 1.17m^{0.39}$	$\propto 8.56m^{0.06}$	$\propto 62.2m^{0.28}$
蜂鸟	$\propto 0.69m^{1.04}$	$\propto 2.24m^{0.53}$	$\propto 7.28m^{0.02}$	$\propto 14.3m^{-0.04}$

尺度定律表明，已知扑翼飞行器的质量，可以近似确定其他几个扑翼参数。它的发现对扑翼飞行器的设计，特别是翼型参数的确定具有重要意义。由于尺度率的表达式是基于大量统计原理得出的近似公式，不同飞行生物有其独特的飞行模式，因此相关参数无法精确计算，只能作为参考依据，需要针对具体扑翼飞行器进行适当调整[2]。由于飞行器机构的安装误差和效率的损失，气动特性会低于设计要求，因此实际参数应略大于设计值。目前已有的小型扑翼飞行器的扑翼质量集中在10～20 g，如傅雅宁等[1]设计的双扑翼重量约为17 g；刘岚等[11]提出随着设计工艺的改进以及材料的优化，可以设计的扑翼飞行器的最小质量约为16 g。基于以上数据的数量级，本设计的扑翼机构模型的质量m范围不超过20 g。

参考尺度率的定义，为了满足气动力的要求，按照扑翼机构重量为20 g进行初步计算，几个重要的扑翼参数选取以及设计取值如表3所示。

表3 扑翼参数与取值

Table 3 Parameters and values of flapping wing

扑翼面积/m^2	翼展/m	展弦比	翼载荷/(N·m^{-2})	扑翼厚度/m	材料
0.006 24	0.208 48	6.564 86	18.028 56	0.001	PET

基于牟晓蕾等[12]对鹰蛾这种大展弦比昆虫的研究，本文以展弦比这个无量纲的参数为切入点。以鹰蛾的翼型为原型，结合计算结果，利用Solidworks进行扑翼设计，得到扑翼外形简图如图9所示。

通过以上机构设计，本文提出的扑翼飞行器可以完成前飞(快速、中速、慢速)运动以及转向(左转、右转)运动，在这几类运动中左右扑翼展现运

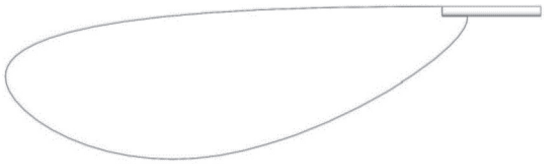


图 9　扑翼外形简图

Fig. 9　Outline of flapping wing

动规律, 不同运动状态下扑翼动作比照如表 4 所示。

表 4　飞行器不同运动状态下扑翼动作比照表

Table 4　Action comparison of flapping wing of aircraft under different motion states

运动类型	左侧扑翼	右侧扑翼
前飞(快速)	急速大幅拍动、翻转	急速大幅拍动、翻转
前飞(中速)	中速中幅拍动、翻转	中速中幅拍动、翻转
前飞(慢速)	舒缓小幅拍动、翻转	舒缓小幅拍动、翻转
转向(左转)	舒缓小幅拍动、翻转	急速大幅拍动、翻转
转向(右转)	急速大幅拍动、翻转	舒缓小幅拍动、翻转

2　控制系统设计

控制系统的硬件选用 TB6612FNG 芯片作为电机驱动芯片, DFRduino UNO R3 芯片作为主控芯片, MPU6050 六轴传感器模块作为监测反馈芯片, 选择 716 空心杯电机作为动力装置, 选用电池组作为电源, 用细铜线作为电路连接线。TB6612FNG 芯片、DFRduino UNO R3 芯片和电池采用外接的形式, 通过细导线将控制系统与扑翼飞行器进行连接, 实现外部控制。MPU6050 安装在置物板上, 实时监测飞行器姿态, 置物板与扑翼飞行器连接。主电机通过电机支撑架安装在置物板上, 输出轴驱动齿轮转动, 两个传动电机安装在齿轮上, 实现幅值调节。控制系统连接图如图 10 所示。

主电机驱动齿轮按照设定的转速转动, 进而通过机械传动实现扑翼运动。变轨电机实现图 2 中铰接点 Q、Q' 转动半径的微调, 工作过程为: 当 MPU-6050 传感器检测到偏转信号传给主控芯片, 主控芯片控制变轨电机带动丝杆运动。丝杆与变位滑块啮合转动, 从而改变变位滑块在旋转装置滑槽内的位置, 对旋转半径进行微调, 进而改变扑翼幅度。滑块变位结构细节如图 11 所示。

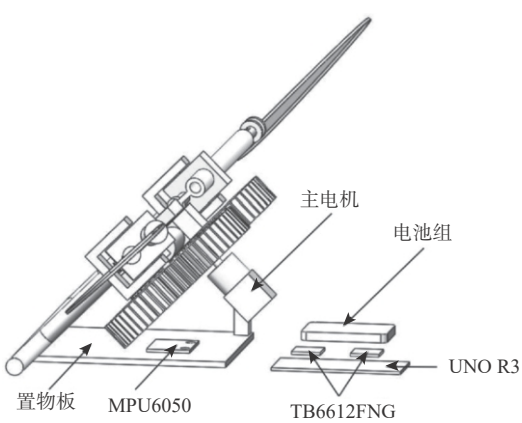


图 10　控制系统连接图

Fig. 10　Connection of control system

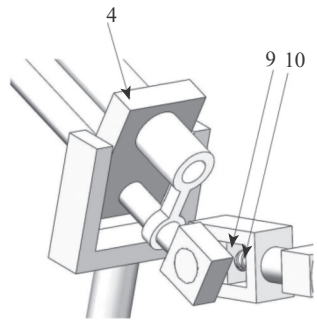


图 11　滑块变位结构细节图

Fig. 11　Details of displacement slider structure

整个控制系统采用负反馈的控制方式。如上文所述调节平衡的功能, 当拍动幅度发生改变后, 相应的飞行器的姿态也会发生改变, 此时 MPU6050 传感器实时反馈偏转角度给主控芯片, 从而实现闭环的负反馈控制系统, 如图 12 所示。依据以上所选的硬件设备和电路原理设计。

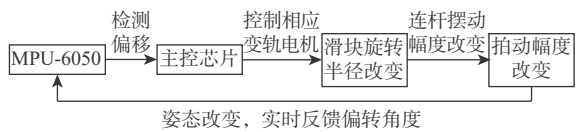


图 12　负反馈控制系统原理图

Fig. 12　Principle of negative feedback control system

3　扑翼飞行器运动试验分析

3.1　气动力分析

XFlow 采用无网格技术, 与传统流体处理软件相比, 在绘制网格时节省了大量时间。XFlow 采用大涡模拟 (large eddy simulation, LES) 作为复杂湍流流动的求解器, 并采用局部方法求解湍流问题, 对复杂流动具有较强的数值模拟能力[13-14]。

参考对于 XFlow 在非定常仿真中收敛性和精度的研究, 本文对仿真所用的格子尺寸和时间步进行调整以满足网格无关性, 最终确定格子尺寸为 0.02c, 计算域 15c×15c×10c(c 为扑翼的平均弦长)。每周期计算 600 步, 确保格子尺寸、计算域和时间步再细化, 气动力变化在 0.5% 以内。

本文采用计算流体力学仿真软件 XFlow 对扑翼飞行器气动力进行分析。以双翼最大幅值扑动作为基准算例, 进行两组流体试验, 分别为:

1）双翼变幅值仿真流体试验。双翼从最大幅值进行拍动同时改变为双翼最小幅值进行拍动。

2）单翼变幅值仿真流体试验。一侧扑翼仍为最大幅值进行拍动，另一侧扑翼变为最小幅值进行拍动。

由式(6)的计算结果可知，拍动的最大幅值为50°，拍动角α变化区间为[−25°, 25°]。变幅值后的扑动幅值为40°，拍动角α变化区间为[−20°, 20°]。

基于以上两组流体试验，可以说明产生升力的大小和气动力的控制对扑翼飞行器多模式飞行的重要性。

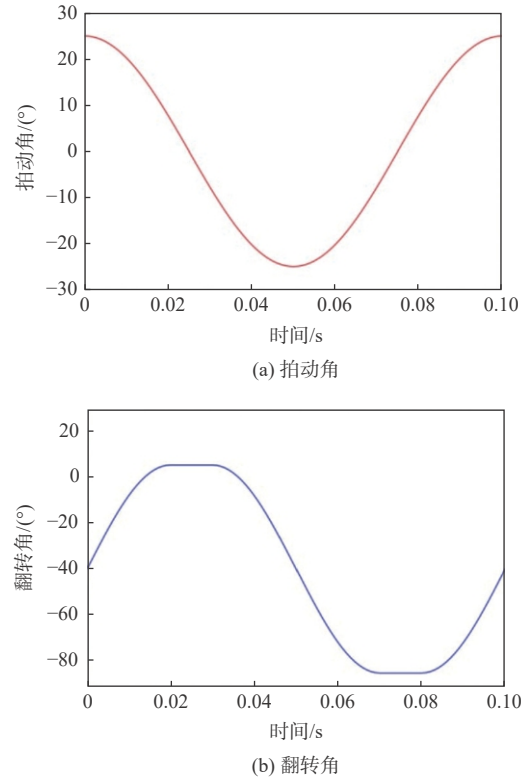
3.2 双翼变幅值仿真流体试验

首先进行两翼最大幅值扑动的流体试验并将其作为基准算例。通过两翼最大幅值扑动的流体试验，比较完整周期内的平均升力和扑翼机构重量之间的关系，验证扑翼飞行器的飞行能力。当扑翼飞行器在稳定水平面飞行时，若飞行器产生较大角度的侧翻角(阈值为5°)时，以向左侧侧翻为例，MPU-6050检测到侧翻角度，传输给控制芯片，进而控制左翼变轨电机转动一定圈数从而增大变位滑块与旋转中心的距离，进而增大左翼拍动幅度，右翼保持原有运动规律，左翼的拍动幅值将改变，进而改变左翼气动力的大小，使飞行器重新平衡。若发生右侧侧翻，控制规律类似。采用XFlow进行流场仿真需设定扑翼的拍动角和翻转角。拍动角的变化函数由公式(6)得出，拍动角α的取值范围为[−25°, 25°]。翻转角的数据通过高速相机观测活体获取[15]。

利用MATLAB对拍动角和翻转角的实验数据进行曲线拟合，获得的拍动角变化曲线和翻转角变化曲线如图13所示。

依据以上获取的实验数据，将左右扑翼的模型导入Xlow，设置两翼的坐标位置和旋转固定点，以及两翼的运动规律函数表达式。确定计算域和计算边界，选取计算域类型为虚拟风洞。本文对计算域的大小选取进行最优化选择。若计算域过小，扑翼的气动影响范围会受限，导致计算结果不准确；若计算域过大，仿真试验精度不高且计算时间较长，对计算平台有较高的硬件要求。因此，为了得到较优的仿真结果，本次设定的计算域的尺寸为500 mm×500 mm×500 mm。虚拟风洞以及扑翼模型如图14所示。

虚拟风洞的介质设置为空气，流场速度为0.1 m/s，攻角为45°，流场温度为289.15 K，流体密度为1.23 kg/m^3，动力黏性系数为1.79×10^{-5} Pa·s，外部加速度为−9.81 m·s^{-2}，风洞出口压强为0 Pa，壁面相对



(a) 拍动角

(b) 翻转角

图13 两翼最大幅值时拍动角和翻转角曲线

Fig. 13 Flapping angle and flip angle at maximum amplitude of both wings

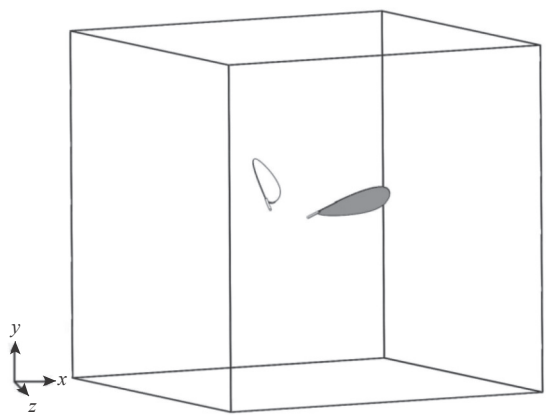


图14 计算域与扑翼模型

Fig. 14 Computing domain and flapping wing model

滑移为0。为了充分分析扑翼飞行器的气动特性，在仿真中采用了自适应轨迹改进和尾迹增强[16-17]。

对于不同状况下的升力、阻力，一般采用无量纲数升力系数C_{L}与推力系数C_{T}来进行比较。

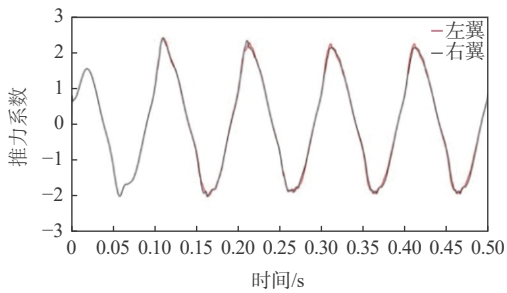
升力系数和推力系数分别定义为

$$\begin{cases} C_{\mathrm{L}} = \dfrac{F_Y}{0.5\rho S U} \\ C_{\mathrm{T}} = \dfrac{F_X}{0.5\rho S U} \end{cases} \tag{7}$$

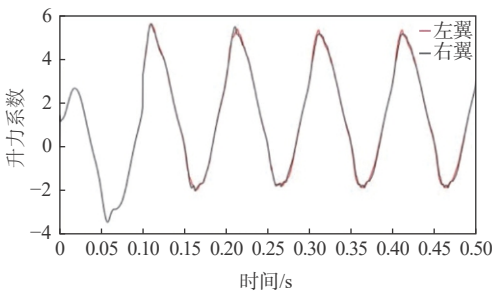
式中：ρ为流体密度；S为模型扑翼表面积；U为翼尖平均拍动速度；F_X为推力；F_Y为升力。依据数值仿

真结果, 结合式(7), 计算推力系数和升力系数。

图 15 为两翼最大幅值扑动时左翼、右翼以及整体产生的推力系数和升力系数。由于在此飞行模式下, 左右扑翼规律相同, 因此左右翼的推力曲线和升力曲线基本吻合。为了进一步探究扑翼飞行器的升力是否满足设计要求, 需进行平均升力系数的计算。由图 15(b)可知, 由于前两个周期为扑翼的起拍阶段, 没有建立起完整良好的流场, 故从第 3 个周期开始, 选取 3 个完整周期分别计算左右翼单周期升力系数以及平均升力系数, 计算结果如表 5 所示。



(a) 推力系数



(b) 升力系数

图 15　两翼最大幅值时推力系数和升力系数

Fig. 15　Thrust coefficient and lift coefficient at maximum amplitude of both wings

表 5　基准算例升力系数

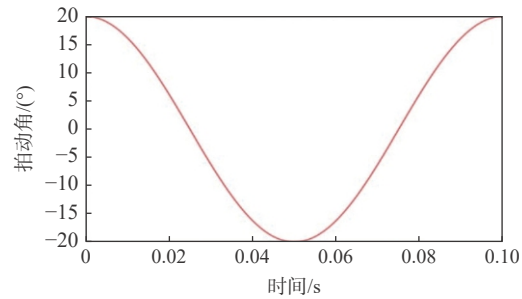
Table 5　Lift coefficients for reference example

周期序号	左翼升力系数	右翼升力系数
3	1.538 5	1.547 0
4	1.514 5	1.498 4
5	1.510 5	1.498 4

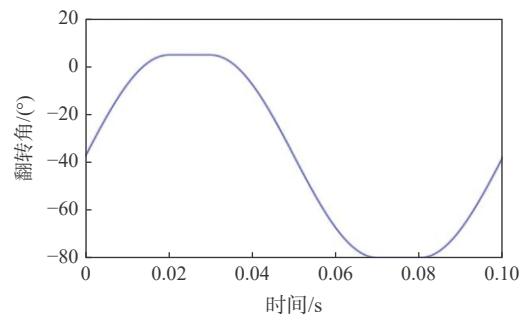
注: 左翼平均升力系数为1.521 2, 右翼平均升力系数为1.514 6。

由计算结果可知, 仿生扑翼飞行器的升力系数、推力系数保持着周期性的规律变化。左右翼完整单周期内的升力系数以及平均升力系数的大小均大于 0, 下拍阶段, 升力系数为正、推力系数为正, 证明下拍阶段产生升力和推力; 上拍阶段, 升力系数为负、推力系数为负, 证明上拍阶段主要产生阻力。扑翼的翻转在产生升力时起着重要的作用, 可以看出, 在一个周期内, 上半周期产生的平均推力系数略大于下半周期, 而上半周期产生的平均升力系数明显大于下半周期, 这就为飞行器的起飞功能提供了动力, 从而验证了设计的合理性和可行性[18]。

基于以上基准算例, 进行双翼变幅值仿真流体试验。当扑翼飞行器接收到返回指令时, 飞行器将从稳定水平面飞行转为下降飞行, 此时不再以最大拍动幅值进行飞行, 控制芯片控制变轨电机转动一定圈数从而带动变位滑块在滑槽内向心运动, 扑翼的拍动幅值将减小, 进而减小气动力的产生实现降落, 在此运动过程中, MPU-6050 仍保持运行, 保证下降过程中的飞行器平衡。由以上基准算例可知, 前两个周期未建立良好的流场关系, 并且稳定拍动时, 每个周期产生的平均升力差别不大, 考虑到结果的准确性以及计算的复杂性, 选取 5 个运动周期进行一组探究性试验, 测试扑翼飞行器最小的气动特性, 两翼的扑动幅值均为最小, 结合式 (6), 得到拍动角α的取值范围为 [−20°, 20°]。同理拍动角和翻转角的变化曲线如图 16 所示。



(a) 拍动角



(b) 翻转角

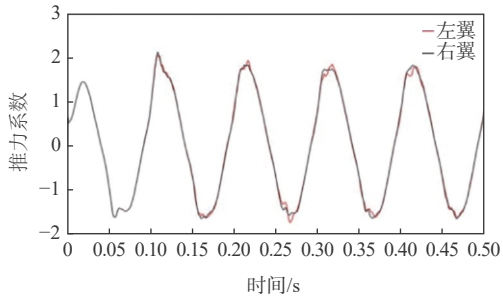
图 16　双翼变幅值时拍动角和翻转角曲线

Fig. 16　Flapping angle and flip angle at variable amplitude of both wings

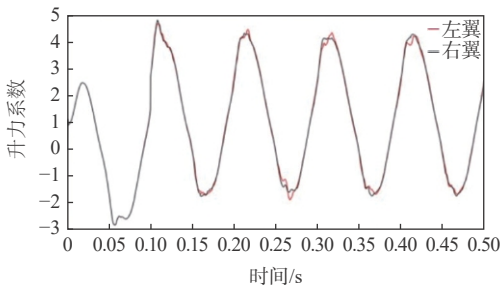
经过 XFlow 的虚拟风洞试验可以得到双翼变幅值仿真流体试验的推力系数和升力系数如图 17 所示。

参考基准算例计算平均升力方法, 从第 3 个周期开始, 选取 3 个完整周期计算双翼变幅值时产生的平均升力系数。

由表 6 可以看出, 两翼在以最小幅值扑动时,



(a) 推力系数



(b) 升力系数

图 17 双翼变幅值时推力系数和升力系数

Fig. 17 Thrust coefficient and lift coefficient at variable amplitude of both wings

产生的升力系数较上述试验有明显的减小，但仍满足升力条件。可以看出，仿生扑翼飞行器的升力系数峰值、推力系数峰值随着拍动幅值的增加而增大，平均升力系数的大小与拍动角的覆盖范围呈正相关。扑翼飞行器飞行时z方向上的加速度大小可以反映飞行器气动性能的优劣。对飞行器在基准算例和双翼变幅值模式下z方向上的加速度进行计算，结果如图 18 所示。

表 6 双翼变幅值时升力系数

Table 6 Lift coefficient at variable amplitude of both wings

周期序号	左翼升力系数	右翼升力系数
3	1.239 7	1.224 8
4	1.192 5	1.201 7
5	1.200 1	1.211 1

注：左翼平均升力系数为1.210 8，右翼平均升力系数为1.212 5。

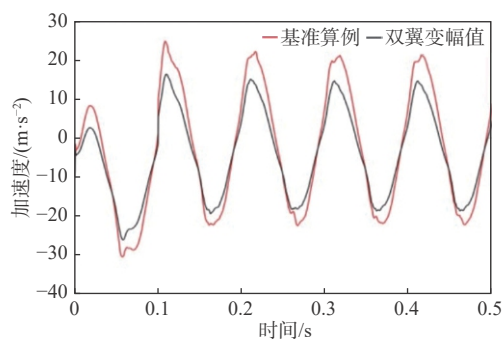


图 18 加速度变化

Fig. 18 Variation of acceleration

基准算例下，起拍阶段z方向上的加速度大小可以实现的变化区间为 [−30.16, 24.91]。在稳定飞行阶段，z方向上的加速度大小可以实现的变化区间为 [−22.13, 22.16]。双翼变幅值时，z方向上的加速度大小可以实现的变化区间为 [−18.15, 16.36]。由此可见，飞行器可产生连续的且数值变化范围较大的加速度，说明飞行器可以覆盖较大范围的气动力。

3.3 单翼变幅值仿真流体试验

当左右两个变位滑块在旋转装置滑槽内的位置不同(即旋转半径不同)时，左右两侧的扑翼幅度也不同，导致左右翼产生的气动力大小也不同。推力在转向中起着决定性的作用，由牛顿第二定律可知，左右翼推动的速度不同，产生速度差即实现转向的功能。以右转向为例，图 19 显示了转向时两翼的瞬时速度、加速度，两翼和整体的转向运动轨迹。可以看出左右翼有明显的运动差异，这就是产生转向的源动力。

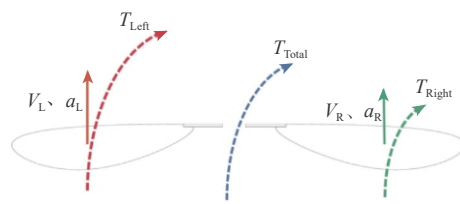


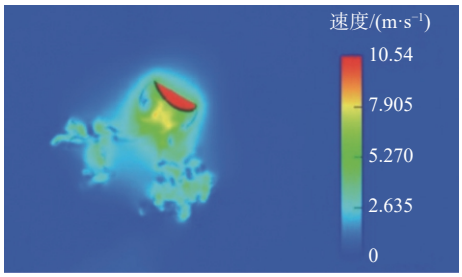
图 19 转向原理图

Fig. 19 Principle of steering

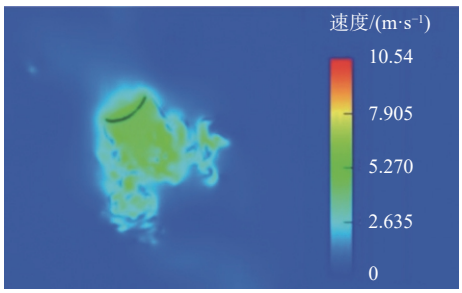
当飞行器要实现转向功能时，左右扑翼需进行不同规律的扑翼运动，进而产生左右不对称的推力，从而实现飞行器飞行姿态的改变[19]。

以扑翼飞行器右转为例，基于基准算例，进行单翼变幅值仿真流体试验。当扑翼飞行器接收到右转指令时，MPU-6050 执行转向角度控制的功能，飞行器将从稳定水平面飞行转为右转飞行，控制芯片控制右翼变轨电机转动一定圈数从而带动变位滑块在滑槽内向旋转中心运动，右翼拍动幅值减小最终以最小拍动幅值规律进行运动，左翼仍以最大拍动幅值规律进行运动，右翼产生的气动力较左翼产生的气动力明显减小，从而实现飞行器右转，当转过一定角度 (设定为 90°) 时，MPU-6050 检测到转向角度，立即向控制芯片发送指令，控制右翼变轨电机运行，在短时间内使得右翼变为最大幅值运动，然后重新在水平面稳定飞行。在 XFlow 进行数值模拟，得到左右两翼下拍阶段产生最大拍速时速度场的云图如图 20 所示。

从左右两翼下拍阶段产生最大拍速时的速度场分布来看，扑翼发生了一定程度的柔性变形，这



(a) 左翼最大拍速云图



(b) 右翼最大拍速云图

图 20　左右翼最大拍速云图

Fig. 20　Maximum flapping velocity cloud image of left and right wings

与实际扑翼的变形是一致的[20]。其次, 左翼附近速度场的整体速度明显比右翼附近速度场的大, 说明该飞行模式下左右两翼可以产生足够大的速度差完成转向。

同理从第 3 个周期开始, 选取 3 个完整周期分别计算左右翼平均推力。此处需要验证扑翼飞行器的转弯性能, 给出单翼变幅值时左翼、右翼和总体的推力如图 21 所示。

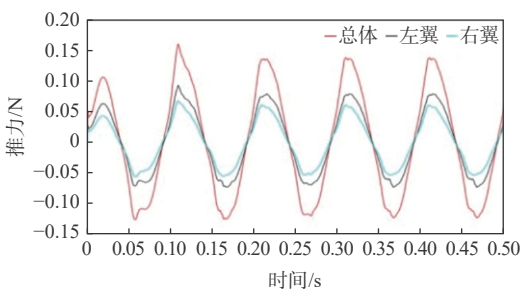


图 21　单翼变幅值时推力

Fig. 21　Thrust at variable amplitude of single wing

由表 7 可以看出, 左右扑翼以不同运动规律进行扑动时, 左翼产生的推力明显大于右翼产生的推力, 结合图 19 转向原理图, 该飞行模型可以提供足够的推力差实现右转功能。

为了进一步探究飞行器的转向性能, 在单翼变幅飞行模式下, 对旋转角速度进行求解。模型俯视图如图 22 所示, C_0、C_1、C_2分别为飞行器的重心、左翼的重心、右翼的重心, 在分析两翼推力分布时, 可将推力集中在C_1、C_2两点处。L为重心之间的距离, 用于转向力矩的计算。M为合推力矩, F_{T1}、F_{T2}分别为左右翼简化后的推力。

表 7　单翼变幅值时计算推力

Table 7　Thrust calculation at variable amplitude of single wing

周期序号	左翼单周期平均推力/N	右翼单周期平均推力/N
3	0.002 47	0.000 832
4	0.002 44	0.000 843
5	0.002 48	0.000 837

注: 左翼平均推力0.002 46 N, 右翼平均推力0.000 837 N。

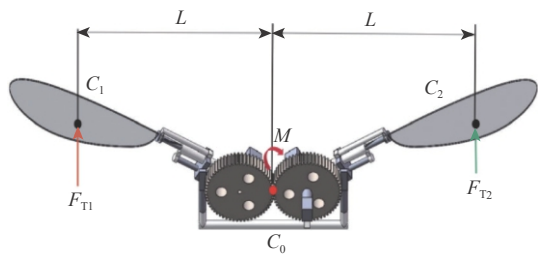


图 22　旋转力矩分析

Fig. 22　Analysis of rotational moment

为了直观看出推力差对扑翼飞行器转向性能的影响, 此处建立推力差和旋转角速度的关系式。定义ω、λ为飞行器绕C_0点的角速度、角加速度, J为飞行器绕C_0点的转动惯量。具体求解过程如下:

$$M(t) = L(F_{T1}(t) - F_{T2}(t)) \tag{8}$$

$$\omega(t) = \lambda(t)t \tag{9}$$

$$\mathrm{d}\omega = t\mathrm{d}\lambda + \lambda\mathrm{d}t \tag{10}$$

$$M\mathrm{d}t = J\omega_2 - J\omega_1 = J\mathrm{d}\omega \tag{11}$$

$$M\mathrm{d}t = J(t\mathrm{d}\lambda + \lambda\mathrm{d}t) = L(F_{T1}(t) - F_{T2}(t))\mathrm{d}t \tag{12}$$

$$\frac{L}{J}[F_{T1}(t) - F_{T2}(t)] = t\frac{\mathrm{d}^2\omega}{\mathrm{d}t^2} + \frac{\mathrm{d}\omega}{\mathrm{d}t} \tag{13}$$

其中, $J = 10\,000\ \mathrm{g\cdot mm^2}$, $L = 140\ \mathrm{mm}$。

由式 (8)～式 (13) 推出关于旋转角速度的二阶微分方程:

$$t\omega'' + \omega' = \frac{L}{J}(F_{T1}(t) - F_{T2}(t)) \tag{14}$$

求解微分方程并且绘制角速度随时间变化的曲线, 角速度单位为°/s, 如图 23 所示。

可以看出, 在转向时角速度的大小和时间呈正相关, 并且在 0.5 s 的时间内, 旋转角速度由 0 提升至 46(°)/s, 说明有较大的旋转角加速度, 转向性能较好。

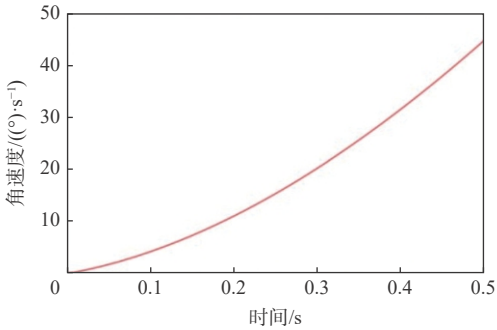


图 23 角速度变化图

Fig. 23 Variation of angular velocity

4 结 论

本文设计了一种可变拍动幅值的扑翼飞行器，对扑翼飞行器进行了机械结构设计和电路设计，并通过 XFlow 的流场仿真试验进行动力特性分析，研究结论如下。

1）本文设计的扑翼机构，实现单电机对扑翼拍动和翻转运动的同时控制。通过 MPU6050 芯片的姿态控制扑翼幅值的变化，进而实现飞行平衡的控制。通过数值模拟的方法，初步获得了新型扑翼飞行器的试验数据，为后续研究及样机制作提供了数据支持。

2）基准算例中升力系数和推力系数各周期内平均值均大于 0，说明扑翼产生的升力和推力能够满足运动条件。整个飞行阶段 y 方向上的加速度大小变化区间为 $[-30.16, 24.91]$，在双翼变幅值前后可以实现快速上升和缓慢降落的设计目标。

3）单翼变幅值试验的结果表明，在左翼变幅值后，在 0.5 s 内就可以将转弯角速度提升至 46(°)/s，表明设计的机械结构可以产生足够大的角加速度完成转向，单翼变幅值可以实现紧急避障、快速转向的功能。

参考文献（References）

[1] 傅雅宁, 赵世恒. 仿生蜻蜓扑翼飞行器模型设计与气动特性分析[J]. 机械设计, 2021, 38(10): 79-87.
FU Y N, ZHAO S H. Design of the aircraft model for bionic dragonflies' flapping wings and analysis on the aerodynamic characteristics[J]. Journal of Machine Design, 2021, 38(10): 79-87(in Chinese).

[2] 孙卫, 冯春鹏. 仿蜻蜓扑翼飞行器机构设计及气动力研究[J]. 飞行力学, 2016, 34(5): 21-25,29.
SUN W, FENG C P. Mechanism design and aerodynamic research on a dragonfly flapping-wing air vehicle[J]. Flight Dynamics, 2016, 34(5): 21-25,29(in Chinese).

[3] DILEO C, DENG X Y. Design of and experiments on a dragonfly-inspired robot[J]. Advanced Robotics, 2009, 23(7-8): 1003-1021.

[4] 张焱, 侯悦民, 王思琪. 扑翼飞行器的动力特性及其气动性能[J]. 液压与气动, 2020, 44(12): 81-88.
ZHANG Y, HOU Y M, WANG S Q. Dynamic characteristics and aerodynamic performance of flapping-wing aircraft[J]. Chinese Hydraulics & Pneumatics, 2020, 44(12): 81-88(in Chinese).

[5] 胡超, 武锦鑫. 一种仿鸟扑翼飞行器的设计研究[J]. 现代制造技术与装备, 2019, 55(7): 16-19.
HU C, WU J X. Design and research of a bird-like flapping wing vehicle[J]. Modern Manufacturing Technology and Equipment, 2019, 55(7): 16-19(in Chinese).

[6] 侯月阳, 侯宇, 孙伟, 等. 可变幅扑翼飞行器设计及多飞行模式实现[J]. 机械传动, 2021, 45(5): 50-56.
HOU Y Y, HOU Y, SUN W, et al. Design of variable amplitude flapping wing aircraft and realization of multiple flight modes[J]. Journal of Mechanical Transmission, 2021, 45(5): 50-56(in Chinese).

[7] SATO T, FUJIMURA A, TAKESUE N. Three-DoF flapping-wing robot with variable-amplitude link mechanism[J]. Journal of Robotics and Mechatronics, 2019, 31(6): 894-904.

[8] KUMAR D, GOYAL T, VEMURI S K, et al. Development and analysis of gull inspired UAV flapping wing[C]//33rd AIAA Applied Aerodynamics Conference. Reston: AIAA, 2015: 3157.

[9] 林镇炜, 王丽丽, 肖杰康, 等. 一种攻角可调的双段翼仿生扑翼飞行器设计[J]. 机械研究与应用, 2021, 34(1): 108-111,114.
LIN Z W, WANG L L, XIAO J K, et al. Design of a bionic flapping wing aircraft with adjustable angle of attack[J]. Mechanical Research & Application, 2021, 34(1): 108-111,114(in Chinese).

[10] 张亚锋, 李郁, 田卫军. 扑动幅值角对仿生扑翼气动力特性的影响[J]. 机械工程与自动化, 2019(1): 18-20.
ZHANG Y F, LI Y, TIAN W J. Effects of flapping amplitude angle on aerodynamic force of flapping wing micro air vehicle[J]. Mechanical Engineering & Automation, 2019(1): 18-20(in Chinese).

[11] 刘岚, 方宗德, 侯宇, 等. 微扑翼飞行器的尺度律研究与仿生设计[J]. 中国机械工程, 2005, 16(18): 1613-1617.
LIU L, FANG Z D, HOU Y, et al. Bionic design and scaling laws for flapping-wing MAVs[J]. China Mechanical Engineering, 2005, 16(18): 1613-1617(in Chinese).

[12] 牟晓蕾, 孙茂. 昆虫翅平面形状和展弦比对其气动性能的影响[J]. 北京航空航天大学学报, 2011, 37(11): 1359-1364,1371.
MOU X L, SUN M. Effects of wing planform on aerodynamics force production of stroking model insect wing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(11): 1359-1364,1371(in Chinese).

[13] 陈茂伟, 孙茂. 蜂蝇快速起飞过程的实验观测及力学分析[J]. 航空学报, 2014, 35(12): 3222-3231.
CHEN M W, SUN M. Experimental measurement and force analys-

is of a fast takeoff in dronefly[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(12): 3222-3231(in Chinese).

[14] 彭连松, 郑孟宗, 潘天宇, 等. 翼间干涉效应对蜻蜓悬停气动性能的影响[J]. 航空学报, 2021, 42(7): 124571.
PENG L S, ZHENG M Z, PAN T Y, et al. Effects of tandem-wing interactions on aerodynamics of hovering dragonflies[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(7): 124571(in Chinese).

[15] 高倩, 郑孟宗, 李志平, 等. 蜻蜓爬升过程飞行特征实验研究[J]. 北京航空航天大学学报, 2016, 42(6): 1271-1278.
GAO Q, ZHENG M Z, LI Z P, et al. Experimental study on flight performance of dragonfly during climbing[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(6): 1271-1278(in Chinese).

[16] QIN M C. Design and research of dragonfly robot under the background of artificial intelligence[J]. Wireless Communications and Mobile Computing, 2022, 2022(1): 1727965.

[17] HU Z, MCCAULEY R, SCHAEFFER S, et al. Aerodynamics of dragonfly flight and robotic design[C]//2009 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2009: 3061-3066.

[18] 朱建阳, 蒋林, 雷斌. 三维仿生悬停扑翼的时间非对称扑动气动特性[J]. 航空动力学报, 2017, 32(4): 858-864.
ZHU J Y, JIANG L, LEI B. Aerodynamic performance of time asymmetric flapping of a three-dimensional flapping hovering bionic wing[J]. Journal of Aerospace Power, 2017, 32(4): 858-864(in Chinese).

[19] GAMMILL M, HASSANALIAN M. Aerodynamic analysis of manta ray inspired fixed and flapping-wing drones for high altitude Venus exploration[C]//ASCEND 2021. Reston: AIAA, 2021: 4143.

[20] FERREIRA R, CAMACHO E A, NEVES F P, et al. Wing design and analysis for micro air vehicle development[C]//AIAA SCITECH 2022 Forum. Reston: AIAA, 2022: 1057.

Model design and aerodynamic characteristic analysis of variable-amplitude flapping wing aircraft

LIANG Jinze[1], PAN Tianyu[1, 2], ZHENG Mengzong[1, *], PENG Liansong[2], CAO Mengda[1]

(1. Research Institute of Aero-Engine, Beihang University, Beijing 102206, China;
2. School of Energy and Power Engineering, Beihang University, Beijing 102206, China)

Abstract: At present, micro air vehicles (MAVs) are being popularized in both military and civilian fields and are playing an increasingly important role. The flapping wing organism shows its advantages of high aerodynamic efficiency, quick action, and stable hover during flight. In this paper, by studying flapping wing insects, a new type of flapping wing aircraft was proposed, and a mechanical structure model and control system with variable rotating orbit radius were designed. When the rotating radius changed, the chip controlled the rotation of the variable rail motor and then changed the position of the displacement slider from the rotation center, which was finally represented by the change in the flapping amplitude of the flapping wing through the transmission mechanism. According to the designed mechanical structure, the coordinate system was established, and the expression was derived. At the same time, XFlow was used to carry out fluid simulation tests. Through the benchmark experiment, the fluid simulation test of the variable amplitude of the two wings, and the fluid simulation test of the amplitude of the variable single wing, the corresponding conclusions were obtained, and the rationality and feasibility of the variable-amplitude flapping wing aircraft model were verified. It provided data support for follow-up research and prototype production.

Keywords: mechanical design; variable amplitude; flip of flapping wing; multimode-flight; flow field simulation; dynamic analysis

Received: 2023-05-24; **Accepted**: 2023-07-28; **Published Online**: 2023-08-15 08:42

URL: link.cnki.net/urlid/11.2625.V.20230814.0937.002

Foundation items: National Natural Science Foundation of China (52006002,51976005,52206040)

* **Corresponding author**. E-mail: zhengmengzong@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0283

高超变体飞行器再入轨迹罚函数序列凸规划

王仰杰[1]，龙腾[1,2]，李俊志[1]，徐广通[3]，孙景亮[1,2,*]

(1. 北京理工大学 宇航学院，北京 100081；2. 北京理工大学重庆创新中心，重庆 401120；3. 浙江大学湖州研究院，湖州 313000)

摘 要：为实现高超声速飞行器由固定构型的单点最优向可变构型的全包线持续最优的跨越升级，设计了类乘波体变体飞行器布局与变后掠-变展长复合变形方案。在此基础上，针对高超声速变体飞行器再入轨迹规划问题求解难度大、规划耗时高的问题，提出了自适应信赖域更新的罚函数序列凸优化方法。采用对数凸化策略凸化路径约束，提高近似精度；引入虚拟控制，对动力学等式约束进行变量替换；定制二阶锥约束，并采用罚函数方法将其加入目标函数中，引导迭代结果向可行域逼近；设计自适应信赖域更新策略，加速序列凸优化算法收敛。仿真结果表明：相比于固定构型，高超变体飞行器增程16.63%，增程效果明显；相比于hp伪谱法，所提算法求解耗时降低了89.24%，具有更高的时效性。

关 键 词：高超声速变体飞行器；再入轨迹规划；对数凸化；虚拟控制；自适应信赖域

中图分类号：V448.1

文献标志码：A **文章编号**：1001-5965(2025)05-1747-13

高超声速飞行器具有快响应、强机动、大空域、宽速域等特点，能够执行全球快速到达、远程目标精确打击和广域监视封锁等特定作战任务，具有极高的军事战略及商业价值[1-2]。然而传统固定构型高超声速飞行器在不同速度下外表面的绕流特性差异较大，无法保证全飞行包线性能最优，飞行能力的提升和任务边界的拓展受到限制。高超声速变体飞行器能够依据飞行环境和飞行剖面自适应调整外形以改善空气动力学特性，减小阻力，优化飞行性能，满足多种任务要求，实现传统飞行器由固定构型的单一工况最优向可变构型的全包线持续最优的跨越升级[3]，已成为国内外航空航天领域的新兴热点[4-7]。

再入滑翔轨迹规划技术作为高超声速变体飞行器遂行宽速域机动、强突防打击等任务的核心技术之一，其实时性对飞行器多任务适配和多环境适应能力的提升至关重要[8]。但是再入滑翔过程中的动压、过载、热流、禁飞区约束以及高度非线性动力学导致轨迹规划问题面临计算高耗时难题。此外，变体设计引入了新的设计变量，使得再入动力学和路径约束维度大幅升高。其次，气动结构外形改变引起压心、焦点的变化，飞行器受力情况变化更为复杂；变体附加力耦合导致动力学模型非线性程度增加，进一步加剧了该阶段飞行器快时变、高动态、强耦合的特点。因此，考虑再入滑翔过程中高超变体飞行器变体附加耦合、模型维度升高及非线性增强等情况，提升轨迹规划时效性，对提高其战场生存与突防能力，获得全飞行包线最优的气动性能和操纵性能具有重要意义。

以伪谱法为代表的直接法求解轨迹规划的研究已比较成熟，且已初步应用于变体飞行器轨迹优化中，并取得了一定的成果[9-15]。杨博等[9]设计了一种满足多种飞行状态下保持气动最优的折叠翼飞行器，以折叠角为控制变量，并采用高斯伪谱法

收稿日期：2023-05-26；录用日期：2023-09-08；网络出版时间：2023-10-20 13:08
网络出版地址：link.cnki.net/urlid/11.2625.V.20231020.1023.003
基金项目：国家自然科学基金 (62003036,62203256)；航空科学基金 (2019ZC072003)；北京理工大学青年教师学术启动计划项目 (XSQD-202201005)；北京理工大学研究生科研水平和创新能力提升专项计划 (2022YCXZ017)

* 通信作者. E-mail：sunjingliangac@163.com

引用格式：王仰杰，龙腾，李俊志，等. 高超变体飞行器再入轨迹罚函数序列凸规划 [J]. 北京航空航天大学学报，2025，51（5）：1747-1759.
WANG Y J，LONG T，LI J Z，et al. Re-entry trajectory planning for hypersonic morphing vehicles using penalty sequence convex programming[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51（5）：1747-1759（in Chinese）.

进行轨迹优化，有效提高了航程。王娜等[10]以变后掠翼导弹为研究对象，以后掠角和攻角为控制变量，采用hp自适应伪谱法对变后掠翼导弹末端弹道进行优化以获得更大的终端速度。Dai等[11]采用hp自适应伪谱法对所设计的高超声速变后掠翼飞行器进行轨迹优化，通过比较终端时间自由条件下最大航程，有效地证明了高超声速变后掠翼飞行器的航程优势。Wei等[13]将打靶法与高斯伪谱法相结合，提出了一种面向变后掠导弹的离线混合弹道优化方法。岳彩红等[15]建立了以展长变形量为控制量的再入轨迹优化模型，并采用基于梯度计算方法的改进伪谱法有效地提高了再入滑翔轨迹优化求解效率。上述研究多是针对亚/跨/超声速变体对象轨迹优化问题开展，且仅将伪谱法作为求解工具，重点是通过轨迹规划结果来分析变体对航程或机动能力的影响，并未关注轨迹规划问题求解的时效性。考虑到数值优化算法的求解性能取决于问题是否非凸[16]，而伪谱法只是将连续最优控制问题转化为可供NLP算法求解的离散形式，并未改变问题非凸性，因此难以满足高超声速变体飞行器轨迹时效性、鲁棒性、收敛性的需求，无法实现多约束条件下高超声速变体飞行器再入轨迹快速生成。

近年来，凸优化方法由于具有多项式时间复杂度、计算速度快以及在可行域中必然收敛的理论优势，广泛应用于航空航天领域的轨迹规划[17-18]。Açıkmeşe等[19-21]采用凸优化技术对行星探测器的动力下降段和火箭返回着陆段进行优化。在此基础上，Lu等[22]提出了高效的序列凸优化方法，通过离散化和线性化手段将非线性约束转化为凸约束，成功求解了高超声速飞行器再入轨迹规划问题。Wang等[23]建立了以时间为自变量的再入动力学模型，使用序列凸优化技术快速求解再入轨迹，其计算耗时远低于伪谱法。通过分析可知，凸优化技术在航空航天的应用仍然存在挑战：凸化技术与模型强相关，需要针对具体的应用对象和过程采用特定的凸技术与松弛技术，以进一步提高求解过程收敛性。

凸优化方法主要分为无损凸优化和序列凸优化两类。其中，无损凸优化技术收敛性是可以得到理论证明的，但是其对应用对象要求严格，无法广泛应用于一般对象。而序列凸优化方法对应用对象要求较低，求解效率较高，已经在高超声速飞行器再入轨迹规划领域中获得了广泛应用，是实现在线快速规划的有效手段。通过文献调研可知，目前尚未有研究将凸优化方法应用于高超声速变体飞行器再入滑翔轨迹规划问题上。针对高超声速变体飞行器在变体过程中气动布局动态调整造成再入轨迹求解难度增大、规划耗时提高的问题，有必要根据模型特点开展对序列凸优化方法应用的进一步研究，进而满足再入段轨迹规划的时效性要求。

本文首先提出类乘波体变体飞行器布局与机翼大尺度复合变形方案，建立高超声速变体飞行器再入轨迹规划的非凸最优控制问题。在此基础上，通过逐次线性化和梯形积分离散方法将其转化为基于基准轨迹的凸最优控制模型。同时，设计虚拟控制策略和信赖域收缩策略，采用罚函数序列凸优化方法对凸优化值问题进行序列求解，提高算法的鲁棒性和时效性。最后通过数值仿真试验，对本文提出的算法进行了验证。

1 变体方案设计

高超声速变体飞行器飞行空域跨越电离层、中间层及平流层，飞行速度经历高超声速、超声速与跨声速多个阶段，导致中小尺度变形对飞行性能影响效果不明显[24]。因此，为有效权衡效费比和可靠性，本文采用以变展长和变后掠为代表的大尺度机翼变形方案。

1.1 机体布局方案

考虑到再入过程大气较为稀薄，本文采用由乘波体机身和变形机翼组成的类乘波体构型[25]，其中变形机翼可实现变后掠与变展长两种变体形式，变形过程中机翼弯度、厚度及扭转等不发生变化。如图1所示，机翼布置在机身尾部两侧，机翼后缘与机身底面平行。考虑机翼变后掠和变展长需求，机翼分为前缘后掠的内翼段和平直的外翼段，其安装角和上反角均为0。

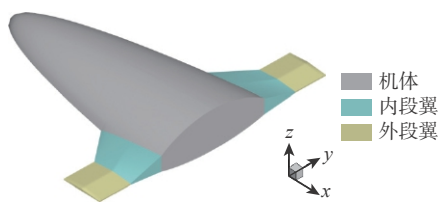


图1 机体布局方案

Fig. 1 Scheme of body layout

1.2 机翼变形方式

相较于旋转式变后掠，剪切式变后掠在升阻比和阻力方面具有更大的优势[25]，因此本文采用剪切式变形方案实现机翼变后掠。另外，采用滑轨式伸缩机翼模型实现外翼段变展长[26]，外翼段在内翼段内部伸缩。机翼变形示意如图2所示。

如图2所示，内段翼通过铰链 A、D 与机体相连，外段翼通过铰链 B、C 与内段翼相连，机翼 AB 段、BC 段、EF 段、FG 段、GH 段、HE 段长度固

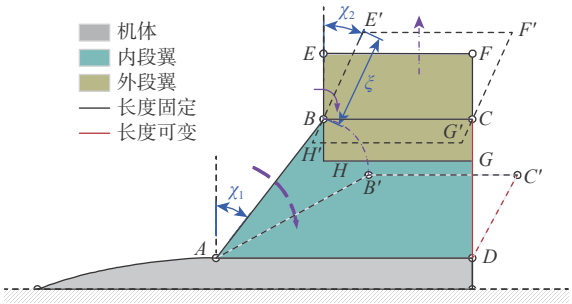


图 2　机翼变形示意图

Fig. 2　Diagram of morphing wing

定, CD 段长度可调, 外段翼的前缘与后缘平行。图 2 中实线表示变化前机翼整体的连接情况, 虚线则表示内段翼相对于机体及外段翼相对于内段翼的变化情况。紫色虚线箭头表示运动方向, 包括内外段翼的剪切式变后掠与外段翼的伸缩式展长变化。内外段翼通过旋转, 改变后掠角χ_1和χ_2, 外段翼通过伸缩改变展长ζ。

2　高超变体再入轨迹规划问题建模

2.1　高超变体气动模型

再入滑翔过程中, 高超变体飞行器构型时变, 高度与速度变化极大, 无法忽略声速及大气密度变化, 因此高超声速变体飞行器气动模型受海拔h, 马赫数M, 攻角α, 后掠角χ_1、χ_2及外段翼前缘长度ξ共同影响。

考虑外段翼后掠角χ_2变化受内段翼后掠角χ_1约束, 将外段翼后掠角的变化范围设置为归一化的区间$\chi_2\in[0,1]$, 当内外段翼后缘平齐时, $\chi_2=0$; 当外段翼的后缘与内段翼的前缘平齐时, $\chi_2=1$。外段翼展长ζ同时受外段翼后掠角χ_2和外段翼前缘长度ξ影响, 故引入外段翼前缘长度ξ表示外段翼展长变化。气动模型设计变量范围设置如下式所示:

$$\begin{cases} 30\ \text{km}\leqslant h\leqslant 65\ \text{km} \\ 7\leqslant M\leqslant 18 \\ 10^\circ\leqslant\alpha\leqslant 20^\circ \\ 8^\circ\leqslant\chi_1\leqslant 35^\circ \\ 0\leqslant\chi_2\leqslant 1 \\ 0.205\ m\leqslant\xi\leqslant 0.820\ m \end{cases} \tag{1}$$

在飞行工况和机翼变形量范围内采用拉丁超方试验设计方法[27]生成多组仿真设置条件, 采用 CFD 仿真软件求解高超声速气动力, 即可获得多组离散的升阻力气动系数数据。

变体气动模型设计变量维度较高, 且无法明确升阻力系数与各变量间的多项式表达形式。此外, 变形机翼后掠角及前缘长度的引入导致同一工况下不同构型的变体飞行器具有相同的气动参数, 即变体气动模型呈现出多峰值特点。当采用插值方法建立气动模型时, 会出现尖刺, 从而气动参数可能出现局部不可导的情况, 不利于后续动力学方程线性化处理。因此根据上述分析, 采用采样要求低、构造简单的径向基函数(radial basis function, RBF)[28]对升力系数$C_{\rm L}$、阻力系数$C_{\rm D}$和参考面积S分别构造代理模型进行拟合, 如下:

$$\begin{cases} C_{\rm L}=f_{\rm RBF_C_L}(h,\alpha,M,\chi_1,\chi_2,\xi) \\ C_{\rm D}=f_{\rm RBF_C_D}(h,\alpha,M,\chi_1,\chi_2,\xi) \\ S=f_{\rm RBF_S}(\chi_1,\chi_2,\xi) \end{cases} \tag{2}$$

采用十折交叉验证法进行模型校验, 拟合精度指标为均方根误差(root mean square error, RMSE), 验证结果如表 1 所示。

表 1　升力系数、阻力系数和参考面积拟合精度

Table 1　RMSE of $C_{\rm L}$, $C_{\rm D}$ and S

拟合精度指标	$C_{\rm L}$	$C_{\rm D}$	参考面积/m^2
RMSE	0.009 6	0.007 4	0.019 5

根据表 1 中拟合精度统计结果可知, 所构造的 RBF 气动代理模型能够有效拟合变体飞行器的气动系数随海拔、攻角、马赫数、后掠角及外段翼前缘长度的关系, RBF 参考面积代理模型能够有效拟合变体飞行器参考面积与机翼变形量的关系。

2.2　高超变体再入动力学方程

不考虑机体结构发生弹性形变, 忽略地球自转及地球曲率带来的影响, 纵向平面内高超声速变体飞行器三自由度动力学方程如下[12]:

$$\begin{cases} \dot h=v\sin\gamma \\ \dot x=v\cos\gamma \\ \dot v=\dfrac{-D}{m}-g\sin\gamma \\ \dot\gamma=\dfrac{L}{mv}-\dfrac{g\cos\gamma}{v} \end{cases} \tag{3}$$

式中: v为飞行速度; γ为弹道倾角; x为射程; m为飞行器质量; g为重力加速度; L和D分别为气动升力和阻力, 与机翼变形量有关, 计算如下:

$$\begin{cases} L=qSC_{\rm L} \\ D=qSC_{\rm D} \end{cases} \tag{4}$$

式中: $q=\rho v^2/2$为动压; $\rho=\rho_0{\rm e}^{-h/h_s}$为海拔$h$处的大气密度; S为变体飞行器的特征面积; $C_{\rm L}$和$C_{\rm D}$为气动参数。升力系数$C_{\rm L}$、阻力系数$C_{\rm D}$和参考面积S与外段翼前缘长度、内外段翼后掠角有关, 均为机翼变形量的函数, 具体计算见 2.1 节。

变体飞行器通过实时改变机翼后掠角和前缘长度实现机体构型的动态调整, 以达到飞行性能全程最优。然而, 在变体飞行器以高超声速再入的过

程中，为保证机体结构稳定及机上载荷正常工作，降低飞行控制系统的设计与实现难度，需避免机翼变形量发生突变或高频振荡。针对上述问题，本文采用后掠角和前缘长度变化率作为控制量，选取海拔h、射程x、速度v、弹道倾角γ机翼变形量内段翼后掠角χ_1、外段翼后掠角χ_2及外段翼前缘长度ξ作为状态量，如下：

$$\boldsymbol{s}=\left[h,x,v,\gamma,\chi_1,\chi_2,\xi\right]^{\mathrm{T}} \tag{5}$$

$$\boldsymbol{u}=\left[\dot{\chi}_1,\dot{\chi}_2,\dot{\xi}\right]^{\mathrm{T}} \tag{6}$$

则式(3)可改写为

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}\boldsymbol{t}}=\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t})+\boldsymbol{b}_1\cdot\boldsymbol{u} \tag{7}$$

式中：

$$\begin{cases} f(\boldsymbol{s},\boldsymbol{t})=\begin{bmatrix} v\sin\gamma \\ v\cos\gamma \\ -\dfrac{D}{m}-g\sin\gamma \\ \dfrac{L}{mv}-\dfrac{g\cos\gamma}{v} \\ \hdashline \boldsymbol{0}_{3\times1} \end{bmatrix} \\ \boldsymbol{b}_1=\begin{bmatrix} \boldsymbol{0}_{4\times3} \\ \hdashline 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{cases} \tag{8}$$

采用类升岭分段式攻角-速度剖面方案[29]，攻角模型为

$$\alpha=\begin{cases} \alpha_{\max} & v\geqslant v_2 \\ \alpha_{\mathrm{mid}}+\alpha_{\mathrm{bal}}\sin\left[\dfrac{\pi(v-v_{\mathrm{mid}})}{v_1-v_2}\right] & v_1\leqslant v\leqslant v_2 \\ \alpha_{\max(L/D)} & v<v_1 \end{cases} \tag{9}$$

式中：v_1和v_2分别为设置的两个速度分段参数；$\alpha_{\max}$为攻角上限；$\alpha_{\max(\mathrm{L/D})}$为最大升阻比对应攻角；$v_{\mathrm{mid}}=(v_1+v_2)/2$；$\alpha_{\mathrm{mid}}=\left(\alpha_{\max}+\alpha_{\max(\mathrm{L/D})}\right)/2$；$\alpha_{\mathrm{bal}}=\left(\alpha_{\max}-\alpha_{\max(\mathrm{L/D})}\right)/2$。

2.3 再入轨迹约束模型

2.3.1 路径约束

为保证高超声速变体飞行器遂行任务，再入滑翔段轨迹规划需要满足热流率、动压及过载等路径约束：

$$\dot{Q}_{\mathrm{s}}=k_Q\rho^{0.5}v^{3.15}\leqslant\dot{Q}_{\mathrm{s,max}} \tag{10}$$

$$q=\frac{1}{2}\rho v^2\leqslant q_{\max} \tag{11}$$

$$n=\frac{\sqrt{L^2+D^2}}{mg}\leqslant n_{\max} \tag{12}$$

式中：$\dot{Q}_{\mathrm{s}}$为驻点处热流率；$\dot{Q}_{\mathrm{s,max}}$为驻点处最大热流率；k_Q为热流密度常数；$q_{\max}$为飞行器能承受的最大动压值；n为过载；$n_{\max}$为飞行器能承受的最大过载值。

联立式(10)～式(12)，将路径约束写成统一形式为

$$\boldsymbol{P}(\boldsymbol{s}(t),\boldsymbol{u}(t))\leqslant0,t\in[t_0,t_{\mathrm{f}}] \tag{13}$$

式中：$\boldsymbol{P}=\left[\dot{Q}_{\mathrm{s}}-\dot{Q}_{\mathrm{s,max}},q-q_{\max},n-n_{\max}\right]^{\mathrm{T}}$；$t_0$为滑翔再入初始时刻；$t_{\mathrm{f}}$为滑翔再入终端时刻。

2.3.2 初值与终端约束

根据再入任务需求，一般会对初值及终端状态进行约束，约束形式可表示为

$$\psi(\boldsymbol{s}(t_0),\boldsymbol{s}(t_{\mathrm{f}}),t_0,t_{\mathrm{f}})=0 \tag{14}$$

即

$$\begin{cases} \boldsymbol{s}(t_0)=\boldsymbol{s}_0, \\ \boldsymbol{s}(t_{\mathrm{f}})=\boldsymbol{s}_{\mathrm{f}} \end{cases} \tag{15}$$

式中：$\boldsymbol{s}_0$和$\boldsymbol{s}_{\mathrm{f}}$分别为初始状态变量和终端状态变量。通常情况下，初值约束包括所有状态量约束，而终端约束会根据再入任务需求进行部分约束设置。

2.3.3 状态与控制变量边界约束

考虑实际飞行过程中高超声速变体飞行器飞行范围、飞行性能及变形机翼机构运动约束，有必要对状态变量与控制量的边界进行约束，避免规划结果超出飞行器最大能力范围。状态与控制变量的边界约束为

$$\varphi(\boldsymbol{s}(t),\boldsymbol{u}(t),t)\leqslant0 \tag{16}$$

即

$$\boldsymbol{s}_{\min}\leqslant\boldsymbol{s}\leqslant\boldsymbol{s}_{\max},\ \boldsymbol{u}_{\min}\leqslant\boldsymbol{u}\leqslant\boldsymbol{u}_{\max} \tag{17}$$

式中：$\boldsymbol{s}_{\max}$和$\boldsymbol{s}_{\min}$分别为再入过程中状态变量的上下边界值；$\boldsymbol{u}_{\min}$和$\boldsymbol{u}_{\max}$分别为再入过程中控制变量的最小值和最大值。

2.3.4 终端时间约束

对于终端时间自由的问题，考虑到高超声速变体飞行器再入过程受任务约束或燃料约束，还需要对终端时间进行约束：

$$t_{\mathrm{f,min}}\leqslant t_{\mathrm{f}}\leqslant t_{\mathrm{f,max}} \tag{18}$$

式中：$t_{\mathrm{f,max}}$和$t_{\mathrm{f,min}}$分别为再入过程终端时间的上下边界值。

2.4 最优控制问题模型

为充分体现变体的增程效果，本文以最大射程为性能指标，且终端时间自由，建立高超声速变体飞行器再入滑翔轨迹优化问题模型为

$$\begin{cases} P^0:\min\limits_{\boldsymbol{s},\boldsymbol{u},t_{\mathrm{f}}}-x(t_{\mathrm{f}}) \\ \quad\text{s.t.}\quad \text{式(7),式(13),式(14),式(16),式(18)} \end{cases} \tag{19}$$

3　模型凸化与离散化

3.1　时间归一化处理

引入变量τ对再入滑翔时间进行归一化处理, 将问题P^0转化为终端时间固定的最优控制问题进行求解, 变量τ定义如下:

$$\tau=\frac{t-t_0}{t_{\mathrm{f}}-t_0}\in[0,1] \tag{20}$$

则式 (7) 可转化为

$$\frac{\mathrm{d}\boldsymbol{s}}{\mathrm{d}\tau}=\boldsymbol{f}(\boldsymbol{s},\tau)+\boldsymbol{b}_2\cdot\boldsymbol{u} \tag{21}$$

式中: 当$t_0=0$时,

$$\boldsymbol{f}(\boldsymbol{s},\tau)=\boldsymbol{f}(\boldsymbol{s},t)\frac{\mathrm{d}t}{\mathrm{d}\tau}=\boldsymbol{f}(\boldsymbol{s},t)\cdot t_{\mathrm{f}} \tag{22}$$

$$\boldsymbol{b}_2=\boldsymbol{b}_1\cdot\frac{\mathrm{d}t}{\mathrm{d}\tau}=\boldsymbol{b}_1\cdot t_{\mathrm{f}} \tag{23}$$

3.2　动力学方程线性化

采用序列小扰动线性化方法, 在参考轨迹邻域内通过泰勒级数展开, 只保留一阶线性项, 将非线性动力学等式约束转化为仿射函数形式, 完成动力学模型的凸化处理。

设问题P^0的第k步迭代求解的结果为$\{\boldsymbol{s}^k,\boldsymbol{u}^k,t_{\mathrm{f}}^k\}$, 对于式 (21) 所示动力学系统, 由于状态量与控制量已经解耦, 在$\{\boldsymbol{s}^k,t_{\mathrm{f}}^k\}$处对$\boldsymbol{f}(\boldsymbol{s},\tau)$进行线性化。由于变形机翼后掠角和前缘长度的引入, $\boldsymbol{f}(\boldsymbol{s},\tau)$中气动力相关项在线性化过程中需要对机翼变形量进行求导计算, 使得线性化动力学方程维度升高, 进而导致再入轨迹优化问题维度升高。矩阵$\boldsymbol{b}_2$由于引入新的变量t_{f}导致控制量$\boldsymbol{u}$呈现非线性, 为避免因线性化导致控制量维度升高, 采用t_{f}^k近似替代t_{f}[22], 再次将矩阵$\boldsymbol{b}_2$固化为常值矩阵, 不再需要对控制量进行线性化处理。式 (21) 在$\{\boldsymbol{s}^k,\boldsymbol{u}^k,t_{\mathrm{f}}^k\}$处经过线性化处理可转化为如下形式:

$$\begin{aligned}\dot{\boldsymbol{s}}=&\boldsymbol{f}(\boldsymbol{s}^k,t_{\mathrm{f}}^k)+\frac{\partial\boldsymbol{f}(\boldsymbol{s},\tau)}{\partial\boldsymbol{s}}\Big|_{\boldsymbol{s}=\boldsymbol{s}^k,t_{\mathrm{f}}=t_{\mathrm{f}}^k}(\boldsymbol{s}-\boldsymbol{s}^k)+\\&\frac{\partial\boldsymbol{f}(\boldsymbol{s},\tau)}{\partial t_{\mathrm{f}}}\Big|_{\boldsymbol{s}=\boldsymbol{s}^k,t_{\mathrm{f}}=t_{\mathrm{f}}^k}(t_{\mathrm{f}}-t_{\mathrm{f}}^k)+\boldsymbol{b}_2\boldsymbol{u}\end{aligned} \tag{24}$$

式中: $\boldsymbol{s}^k=\left[h^k,x^k,v^k,\gamma^k,\chi_1^k,\chi_2^k,\xi^k\right]^{\mathrm{T}}$; 雅可比矩阵$\left[\partial\boldsymbol{f}(\boldsymbol{s},\tau)/\partial\boldsymbol{s}\right]$和$\left[\partial\boldsymbol{f}(\boldsymbol{s},\tau)/\partial t_{\mathrm{f}}\right]$计算如下:

$$\frac{\partial f(\boldsymbol{s},\tau)}{\partial\boldsymbol{s}}=\begin{bmatrix}0&0&a_{13}&a_{14}&0&0&0\\0&0&a_{23}&a_{24}&0&0&0\\a_{31}&0&a_{33}&a_{34}&a_{35}&a_{36}&a_{37}\\a_{41}&0&a_{43}&a_{44}&a_{45}&a_{46}&a_{47}\\\multicolumn{7}{c}{\boldsymbol{0}_{3\times7}}\end{bmatrix}t_{\mathrm{f}} \tag{25}$$

$$\frac{\partial\boldsymbol{f}(\boldsymbol{s},\tau)}{\partial t_{\mathrm{f}}}=\boldsymbol{f}(\boldsymbol{s},t) \tag{26}$$

省略上标k, 式 (25) 中各项系数如下:

$$\begin{cases}a_{13}=\sin\gamma,\ a_{14}=v\cos\gamma,\\a_{23}=\cos\gamma,a_{24}=-v\sin\gamma,\\a_{31}=-\dfrac{D_h}{m}+\dfrac{2\mu_g\sin\gamma}{r^3},\\a_{33}=-\dfrac{D_v}{m},\ a_{34}=-g\cos\gamma,\\a_{35}=-\dfrac{D_{\chi_1}}{m},\ a_{36}=-\dfrac{D_{\chi_2}}{m},\\a_{37}=-\dfrac{D_\xi}{m},a_{41}=\dfrac{L_h}{mv}+\dfrac{2\mu_g\cos\gamma}{vr^3},\\a_{43}=\dfrac{L_vv-L}{mv^2}+\dfrac{g\cos\gamma}{v^2},\ a_{44}=\dfrac{g\sin\gamma}{v},\\a_{45}=\dfrac{L_{\chi_1}}{mv},\ a_{46}=\dfrac{L_{\chi_2}}{mv},\ a_{47}=\dfrac{L_\xi}{mv},\\D_h=\dfrac{\partial D}{\partial h}=-\dfrac{D}{h_s},D_v=\dfrac{\partial D}{\partial v}=\dfrac{2D}{v}+\dfrac{1}{2}\rho v^2s\dfrac{\partial C_{\mathrm{D}}}{\partial v},\\L_h=\dfrac{\partial L}{\partial h}=-\dfrac{L}{h_s},L_v=\dfrac{\partial L}{\partial v}=\dfrac{2L}{v}+\dfrac{1}{2}\rho v^2s\dfrac{\partial C_{\mathrm{L}}}{\partial v}\end{cases} \tag{27}$$

式中: $r=h+R_0$为地心距; C_{L}和C_{D}分别为升阻力系数; μ_g为万有引力参数; $h_s=7\,254.3$为与大气高度相关常值系数; L_{χ_1}为升力对后掠角χ_1的导数。以此类推, 不再赘述。

为保证线性化的有效性, 需要在迭代过程中对状态量$\boldsymbol{s}$和终端时间t_{f}补充信赖域约束, 即

$$\begin{cases}\left|\boldsymbol{s}-\boldsymbol{s}^k\right|\leqslant\boldsymbol{\varepsilon}_{\mathrm{s}}\\\left|t_{\mathrm{f}}-t_{\mathrm{f}}^k\right|\leqslant\boldsymbol{\varepsilon}_{\mathrm{t}}\end{cases} \tag{28}$$

式中: $\boldsymbol{\varepsilon}_{\mathrm{s}}\in\mathbf{R}^7$和$\boldsymbol{\varepsilon}_{\mathrm{t}}\in\mathbf{R}$为常值向量。信赖域约束会对算法的迭代产生影响, 一般会设置较小的信赖域, 以此降低线性化带来的误差, 但是过小的信赖域导致迭代次数增加, 降低时效性。

3.3　路径约束对数凸化

针对热流、动压、过载约束的数值特性, 为进一步提高近似精度, 采用对数凸化方法对路径约束进行凸化, 将非线性不等式约束转化为仿射不等式约束。

为简化过载约束的数学表达形式。对式 (12) 左右两边同时平方, 整理可得

$$\tilde{n}=L^2+D^2\leqslant\tilde{n}_{\max} \tag{29}$$

式中: $\tilde{n}_{\max}=(n_{\max}\cdot mg)^2$。

对式 (10)、式 (11) 和式 (29) 等式两边取对数, 可得

$$\begin{cases}P_{\dot{Q}_{\mathrm{s}}}=\ln\left(k_Q\sqrt{\rho_0}\right)-\dfrac{h}{2h_{\mathrm{s}}}+3.15\ln v-\ln\dot{Q}_{\mathrm{s,max}}\leqslant0\\P_q=\ln\left(\dfrac{g_0R_0\rho_0}{2}\right)-\dfrac{h}{h_{\mathrm{s}}}+2\ln v-\ln q_{\max}\leqslant0\\P_{\tilde{n}}=\ln\left(\dfrac{\rho_0S}{2}\right)-\dfrac{h}{h_{\mathrm{s}}}+\ln\left(C_{\mathrm{L}}^2+C_{\mathrm{D}}^2\right)+2\ln v-\ln\tilde{n}_{\max}\leqslant0\end{cases} \tag{30}$$

式中，通过对数凸化策略实现了路径约束中部分高度相关项的无损线性化，完全保留了路径约束中部分高度相关项的非线性特征，在一定程度上提高了路径约束的凸化精度。考虑变形量χ_1,χ_2,ξ对气动力的影响，式(30)中除$\ln v$和$\ln\left(C_{\mathrm{L}}^2+C_{\mathrm{D}}^2\right)$两项外，其余均为线性项。由高超变体气动建模过程可知，气动力系数与机翼变形量呈现非线性关系，因此在对再入路径约束进行凸化时需要对气动力系数进行线性化处理。对$\ln v$和$\ln\left(C_{\mathrm{L}}^2+C_{\mathrm{D}}^2\right)$在$\boldsymbol{\ell}^k=\left[v^k,\chi_1^k,\chi_2^k,\xi^k\right]^{\mathrm{T}}$处进行小扰动线性化，有

$$\ln(v)\approx\ln v^k+\frac{1}{v^k}\left(v-v^k\right) \tag{31}$$

$$\begin{aligned}\ln\left(C_{\mathrm{L}}^2+C_{\mathrm{D}}^2\right)=&\ln\left(\left(C_{\mathrm{L}}^k\right)^2+\left(C_{\mathrm{D}}^k\right)^2\right)+\\&\frac{2\left(C_{\mathrm{L}}^k+C_{\mathrm{D}}^k\right)}{\left(C_{\mathrm{L}}^k\right)^2+\left(C_{\mathrm{D}}^k\right)^2}\left[\frac{\partial C_{\mathrm{L}}}{\partial\chi_1},\frac{\partial C_{\mathrm{L}}}{\partial\chi_2},\frac{\partial C_{\mathrm{L}}}{\partial\xi}\right]\begin{bmatrix}\chi_1-\chi_1^k\\\chi_2-\chi_2^k\\\xi-\xi^k\end{bmatrix}\end{aligned} \tag{32}$$

综上，联立式(30)～式(32)，记$\tilde{\boldsymbol{P}}=[P_{\dot{Q}_{\mathrm{s}}},P_q,P_{\tilde{n}}]^{\mathrm{T}}$，将对数凸化的路径约束$\tilde{\boldsymbol{P}}_{\mathrm{L}}$整理成矩阵形式如下：

$$\tilde{\boldsymbol{P}}_{\mathrm{L}}=\tilde{\boldsymbol{P}}^k+\left.\frac{\partial\tilde{\boldsymbol{P}}}{\partial\boldsymbol{\ell}}\right|_{\boldsymbol{\ell}=\boldsymbol{\ell}^k}\left(\boldsymbol{\ell}-\boldsymbol{\ell}^k\right) \tag{33}$$

式中：

$$\tilde{\boldsymbol{P}}^k=\begin{bmatrix}\ln\left(k_Q\sqrt{\rho_0}\right)-\dfrac{h}{2h_{\mathrm{s}}}+3.15\ln v^k-\ln\dot{Q}_{\mathrm{s,max}}\cdot\\\ln\left(\dfrac{g_0R_0\rho_0}{2}\right)-\dfrac{h}{h_{\mathrm{s}}}+2\ln v^k-\ln q_{\max}\cdot\\\ln\left(\dfrac{\rho_0S}{2}\right)-\dfrac{h}{h_{\mathrm{s}}}+2\ln v^k+\ln\left(\left(C_{\mathrm{L}}^k\right)^2+\left(C_{\mathrm{D}}^k\right)^2\right)-\ln\tilde{n}_{\max}\end{bmatrix} \tag{34}$$

$$\left.\frac{\partial\tilde{\boldsymbol{P}}}{\partial\boldsymbol{\ell}}\right|_{\boldsymbol{\ell}=\boldsymbol{\ell}^k}=\begin{bmatrix}\dfrac{3.15}{v^k}&0&0&0\\\dfrac{2}{v^k}&0&0&0\\\dfrac{2}{v^k}&\eta\dfrac{\partial C_{\mathrm{L}}}{\partial\chi_1}&\eta\dfrac{\partial C_{\mathrm{L}}}{\partial\chi_2}&\eta\dfrac{\partial C_{\mathrm{L}}}{\partial\xi}\end{bmatrix} \tag{35}$$

其中：$\eta=2\left(C_{\mathrm{L}}^k+C_{\mathrm{D}}^k\right)\Big/\left(\left(C_{\mathrm{L}}^k\right)^2+\left(C_{\mathrm{D}}^k\right)^2\right)$。

3.4 梯形积分离散化处理

采用配点法同时离散状态变量与控制变量，将时域[0,1]均匀拆分成N个子区间，离散步长为$\Delta t=1/N$，离散点共有$N+1$个：$\boldsymbol{s}_i=\boldsymbol{s}(\tau_i)$，$\boldsymbol{u}_i=\boldsymbol{u}(\tau_i)$，$\tau_i=i/N,i=0,1,\cdots,N$。

用梯形积分公式离散连续时间系统式(24)可得

$$\begin{aligned}\boldsymbol{s}_{i+1}=&\boldsymbol{s}_i+\frac{\Delta t}{2}\left(\dot{\boldsymbol{s}}_{i+1}+\dot{\boldsymbol{s}}_i\right)=\\&\boldsymbol{s}_i+\frac{\Delta t}{2}\left(\bar{\boldsymbol{f}}_{i+1}+\frac{\partial\boldsymbol{f}_{i+1}}{\partial\boldsymbol{s}}\Delta\boldsymbol{s}_{i+1}+\frac{\partial\boldsymbol{f}_{i+1}}{\partial t_{\mathrm{f}}}\Delta t_{\mathrm{f},i}\right)+\\&\frac{\Delta t}{2}\left(\bar{\boldsymbol{f}}_i+\frac{\partial\boldsymbol{f}_i}{\partial\boldsymbol{s}}\Delta\boldsymbol{s}_i+\frac{\partial\boldsymbol{f}_i}{\partial t_{\mathrm{f}}}\Delta t_{\mathrm{f},i}\right)+\frac{\Delta t}{2}\boldsymbol{b}_2\left(u_{i+1}+u_i\right)\end{aligned} \tag{36}$$

式中：$\boldsymbol{f}_i=\boldsymbol{f}\left(\boldsymbol{s}_i,t_{\mathrm{f},i}\right)$；$\bar{\boldsymbol{f}}_i=\boldsymbol{f}\left(\bar{\boldsymbol{s}}_i,\bar{t}_{\mathrm{f},i}\right)$；$\Delta\boldsymbol{s}_i=\boldsymbol{s}_i-\bar{\boldsymbol{s}}_i$；$\Delta t_{\mathrm{f},i}=t_{\mathrm{f}}-\bar{t}_{\mathrm{f},i}$。

将式(24)代入式(36)中，整理可得

$$\begin{aligned}&\boldsymbol{A}^{k+1}\boldsymbol{s}_i^{k+1}+\boldsymbol{A}^k\boldsymbol{s}_i^k+\boldsymbol{B}^{k+1}\boldsymbol{u}_i^{k+1}+\boldsymbol{B}^k\boldsymbol{u}_i^k+\\&\left(\boldsymbol{C}^{k+1}+\boldsymbol{C}^k\right)\Delta t+\left(\boldsymbol{D}^{k+1}+\boldsymbol{D}^k\right)=\boldsymbol{0}\end{aligned} \tag{37}$$

相关常值矩阵系数如下：

$$\begin{cases}\boldsymbol{A}^{k+1}=\dfrac{\overline{\Delta t}}{2}\left.\dfrac{\partial f}{\partial\boldsymbol{s}}\right|_{\boldsymbol{s}=\bar{\boldsymbol{s}}^{k+1}}-\boldsymbol{I}\\\boldsymbol{A}^k=\dfrac{\overline{\Delta t}}{2}\left.\dfrac{\partial f}{\partial\boldsymbol{s}}\right|_{\boldsymbol{s}=\bar{\boldsymbol{s}}^k}+\boldsymbol{I}\\\boldsymbol{B}^{k+1}=\boldsymbol{B}^k=\dfrac{\boldsymbol{b}_2}{2}\overline{\Delta t}\\\boldsymbol{C}^{k+1}=\dfrac{1}{2}\left(f\left(\bar{\boldsymbol{s}}^{k+1}\right)+\boldsymbol{b}_2\bar{\boldsymbol{u}}^{k+1}\right)\\\boldsymbol{C}^k=\dfrac{1}{2}\left(f\left(\bar{\boldsymbol{s}}^k\right)+\boldsymbol{b}_2\bar{\boldsymbol{u}}^k\right)\\\boldsymbol{D}^{k+1}=-\dfrac{\overline{\Delta t}}{2}\left(\left.\dfrac{\partial f}{\partial\boldsymbol{s}}\right|_{\boldsymbol{s}=\bar{\boldsymbol{s}}^{k+1}}\bar{\boldsymbol{s}}^{k+1}+\boldsymbol{b}_2\bar{\boldsymbol{u}}^{k+1}\right)\\\boldsymbol{D}^k=-\dfrac{\overline{\Delta t}}{2}\left(\left.\dfrac{\partial f}{\partial\boldsymbol{s}}\right|_{\boldsymbol{s}=\bar{\boldsymbol{s}}^k}\bar{\boldsymbol{s}}^k+\boldsymbol{b}_2\bar{\boldsymbol{u}}^k\right)\end{cases} \tag{38}$$

将再入滑翔过程中的约束条件离散化，得到

$$\tilde{P}_{\mathrm{c}}(i)\leqslant0 \tag{39}$$

$$\psi(\boldsymbol{s}(0),\boldsymbol{s}(N))=0 \tag{40}$$

$$\varphi(\boldsymbol{s}(i),\boldsymbol{u}(i))\leqslant0 \tag{41}$$

3.5 高超变体再入凸优化问题模型

根据上述分析，建立高超变体飞行器再入滑翔轨迹凸优化问题P^1：

$$\begin{cases}P^1:\min\limits_{\boldsymbol{s},\boldsymbol{u},t_{\mathrm{f}}}-x(N)\\\quad\text{s.t.}\quad\text{式(28),式(37),式(39)}\sim\text{式(41)}\end{cases} \tag{42}$$

4 自适应信赖域更新的罚函数序列凸优化方法

机翼后掠角与前缘长度的引入使得高超声速飞行器的再入动力学方程与再入路径约束维度大幅升高，增大了再入轨迹优化问题的求解难度。此外，通过线性化与离散化处理后带来的数值误差由于模型维度升高而进一步增大。上述原因导致原问题中的某些约束无法满足，进而导致再入轨迹凸优化问题无法求解。针对上述问题，提出一种基于自适应信赖域更新的罚函数序列凸优化方法(adaptive trust-region-based penalty sequential convex programming, ATP-SCP)，在序列凸优化方法框架下，采用虚拟控制策略，松弛动力学等式约束；引入二阶锥约束，保证问题凸性，对其进行罚函数优化，确保序列凸优化算法的迭代可行性；设计自适应信

赖域更新机制, 提高算法的收敛性。

4.1　虚拟控制与罚函数策略

变体设计的引入使得高超声速飞行器的动力学方程维度升高, 在线性化近似过程舍弃了泰勒展开的二阶及高阶项, 在一定程度上忽略了原系统的非线性特征, 特别是当维度升高后, 系统非线性特征损失更多, 因此会导致线性化动力学方程不能满足等式约束, 轨迹迭代无法收敛到可行解的情况。为提升高超声速变体飞行器再入滑翔轨迹凸优化算法的收敛性, 在序列凸优化求解框架下, 将虚拟控制策略与罚函数优化相结合, 保证序列凸优化算法迭代的收敛性。

为防止出现由于连续线性化产生的误差导致序列迭代不收敛的问题, 引入虚拟控制量$\boldsymbol{\upsilon}\in\mathbf{R}^7$, 对线性化动力学等式约束进行替换, 则式 (37) 为

$$\boldsymbol{A}_i^{k+1}\boldsymbol{s}_i^{k+1}+\boldsymbol{A}_i^{k}\boldsymbol{s}_i^{k}+\boldsymbol{B}_i^{k+1}\boldsymbol{u}_i^{k+1}+\boldsymbol{B}_i^{k}\boldsymbol{u}_i^{k}+\left(\boldsymbol{C}^{k+1}+\boldsymbol{C}^{k}\right)\Delta t+\left(\boldsymbol{D}^{k+1}+\boldsymbol{D}^{k}\right)+\boldsymbol{\upsilon}_i=\boldsymbol{0}\tag{43}$$

虚拟控制量$\boldsymbol{\upsilon}$的引入, 扩大了凸优化问题$\mathcal{P}^1$的解空间。为保证不改变问题$\mathcal{P}^1$的凸性, 在原问题基础上需增加一个二阶锥约束, 避免线性化动力学过度失真导致再入轨迹规划结果不合理:

$$\|\boldsymbol{\upsilon}_i\|_2\leqslant\eta_i,\qquad i=1,2,\cdots,N\tag{44}$$

式中: η_i为$\boldsymbol{\eta}\in\mathbf{R}^N$的第$i$个分量, 即在每个离散点处对虚拟控制量施加边界约束。

在问题$\mathcal{P}^1$的基础上, 结合式 (43)、式 (44), 采用罚函数方法将虚拟控制以线性项引入目标函数中, 则性能指标变为

$$J=\min\ \left(-x(N)+\omega\left\|\boldsymbol{\eta}\right\|_2\right)\tag{45}$$

式中: ω为罚系数。通过罚函数项引导再入轨迹解向可行域内部收敛, 逐代降低动力学违背度直到满足动力学等式约束。

4.2　信赖域更新机制

信赖域约束会对算法的迭代产生影响。信赖域过小导致可行域较小, 会出现无解或迭代求解耗时增加。信赖域过大会导致迭代求解难以收敛, 且无法保证解的准确性。

机翼前缘长度和后掠角作为变形量的引入使得高超声速变体飞行器再入轨迹规划问题的空间复杂度升高, 待求解个数的增加导致序列凸优化方法求解难度增大, 算法收敛速度较慢。针对上述问题, 提出一种信赖域序列更新策略, 更新机制如下:

$$\boldsymbol{\varepsilon}^{k+1}=\begin{cases}\beta\cdot\boldsymbol{\varepsilon}^{k} & J^{k+1}\leqslant J^{k}\\ \boldsymbol{\varepsilon}^{k} & J^{k+1}>J^{k}\end{cases}\tag{46}$$

式中: $\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}_{\mathrm{s}}$为状态变量信赖域; $\beta\in(0,1)$为信赖域缩放系数; $\boldsymbol{\varepsilon}^k$为第$k$代凸优化求解过程中的信赖域半径; J^k为第k代凸优化求解的纵向平面内高超声速变体飞行器的最大射程, $k\geqslant 2$为正整数。第$k+1$代为当前代, 当$J^{k+1}\leqslant J^k$, 则当前性能指标未得到改善, 保持当前信赖域半径, 寻找更优的可行解; 反之, 则缩减信赖域半径, 加速收敛。

4.3　算法流程

在凸优化问题P^1的基础上, 结合虚拟控制策略与信赖域更新策略, 构造新的凸优化问题P^2可构造如下:

$$\begin{cases}P^2:\ \min\limits_{\boldsymbol{s},\boldsymbol{u},t_{\mathrm{f}}}\ -x(N)+\omega\|\boldsymbol{\eta}\|_2\\ \quad\text{s.t.}\qquad \text{式(28),式(39)}\sim\text{式(41),式(43)}\sim\text{式(44)}\end{cases}\tag{47}$$

针对凸优化问题P^2, 使用 ATP-SCP 方法求解高超声速变体飞行器再入轨迹规划问题, 具体步骤如下所示。

输入: 离散点数 N, 信赖域半径$\boldsymbol{\varepsilon}_{\mathrm{s}}^0$, 信赖域收缩系数$\beta$, 罚系数$\omega$, 收敛容差$\delta$, 状态变量与控制变量边界$[s_{\min},s_{\max}]$、$[u_{\min},u_{\max}]$, 初末状态$[s_0,s_{\mathrm{f}}]$。

输出: 最优轨迹 $\boldsymbol{S}^*$, 最优控制指令 $\boldsymbol{U}^*$, 终端时间t_{f}^*。

1. Set $k=0$, 初始化积分获得参考轨迹$\boldsymbol{S}^0$和初始终端时间t_{f}^0
2. while $\Delta_{\mathrm{s}}^k>\delta$, do
3. 　　根据$\boldsymbol{S}^k$、$\boldsymbol{U}^k$和t_{f}^k构造凸优化子问题P^2
4. 　　采用凸优化求解器求解子问题P^2, 获得$\boldsymbol{S}^{k+1}$, $\boldsymbol{U}^{k+1}$, t_{f}^{k+1}
5. 　　更新状态代差$\Delta_i^{k+1}=\left|\boldsymbol{S}_i^{k+1}-\boldsymbol{S}_i^{k}\right|$
6. 　if $\Delta_{\mathrm{s}}^{k+1}\leqslant\delta$, then
7. 　　Set $\boldsymbol{S}^*=\boldsymbol{S}^{k+1}$, $\boldsymbol{U}^*=\boldsymbol{U}^{k+1}$ and $t_{\mathrm{f}}^*=t_{\mathrm{f}}^{k+1}$
8. 　　break
9. 　else
10. 　　计算相邻两代的性能指标J^{k+1}和J^k
11. 　　根据信赖域更新机制更新信赖域$\boldsymbol{\varepsilon}^{k+1}\leftarrow\boldsymbol{\varepsilon}^k$
12. 　　Set $k=k+1$
13. 　end
14. end
15. return $\boldsymbol{S}^*$, $\boldsymbol{U}^*$ and t_{f}^*.

步骤 1　(第 1 行): 初始化迭代代数$k=0$, 离散点数N, 信赖域半径$\boldsymbol{\varepsilon}^0$, 信赖域收缩系数$\beta$, 罚系数$\omega$, 收敛容差$\delta$, 高超声速变体飞行器再入过程状态变量与控制变量边界$[s_{\min},s_{\max}]$、$[u_{\min},u_{\max}]$, 初末状态$[s_0,s_{\mathrm{f}}]$, 初始控制量取$\boldsymbol{u}_i=\left[\dot{\chi}_{1,i},\dot{\chi}_{2,i},\dot{\xi}_i\right]^{\mathrm{T}}=\boldsymbol{0}_{3\times1}(i=1,2,\cdots,N)$。采用龙格库塔积分方法生成初始参考轨迹$\boldsymbol{S}^0$和初始终端时间$t_{\mathrm{f}}^0$。

步骤 2 （第 2～4 行）：根据上一代状态变量 $\boldsymbol{S}^k$、控制变量 $\boldsymbol{U}^k$ 和参考终端时间 t_{f}^k 构造凸优化子问题 P^2，采用凸优化求解器求解子问题，获得新的状态变量 $\boldsymbol{S}^{k+1}$、控制变量 $\boldsymbol{U}^{k+1}$ 和参考终端时间 t_{f}^{k+1}。

步骤 3 （第 5～9 行）：判断是否满足收敛容差 $\left|\boldsymbol{s}^{k+1}-\boldsymbol{s}^k\right| \leqslant \delta$，若满足，则退出迭代，转入步骤 5；否则，转入步骤 4。

步骤 4 （第 10～15 行）：根据式 (45) 计算相邻两代的性能指标，并根据式 (46) 所示准则更新信赖域，更新 $k=k+1$，并返回步骤 2。

步骤 5 （第 16 行）：迭代终止，输出最优状态轨迹 $\boldsymbol{S}^*$、控制变量 $\boldsymbol{U}^*$、终端时间 t_{f}^*。

5 数值仿真验证

本节开展再入滑翔轨迹规划数值仿真实验，验证所提算法的有效性和时效性。PC 机硬件环境为：AMD Ryzen 7 3700X 8-Core 处理器，16 GB 运行内存；软件环境为：64 位 Windows11 操作系统，MATLAB R2020a。程序实现过程中采用 Yalmip 工具包[30] 完成凸优化子问题建模，调用 ECOS 求解器[31] 进行优化解算。

高超声速变体飞行器参数如表 2 所示。

基于 2.1 节设置的飞行工况，现将攻角模型式 (9) 中相关参数设置如图 3 所示：攻角-速度剖面速度边界值 $v_1=3\,000$ m/s、$v_2=4\,500$ m/s，攻角边界值 $\alpha_{\max}=18°$、$\alpha_{\max(\mathrm{L/D})}=12°$。

表 2 高超声速变体飞行器参数

Table 2 Parameters of hypersonic morphing vehicle

飞行器性能参数	数值
质量 m/kg	5 500
固定外形参考翼面积 S/m^2	10.94
最大热流率 $\dot{Q}_{\mathrm{s,max}}$/(kW·m^{-2})	1 500
热流相关系数 k_Q	5×10^{-8}
最大动压 $q_{\max}$/kPa	38
最大过载 $n_{\max}$	$3g$

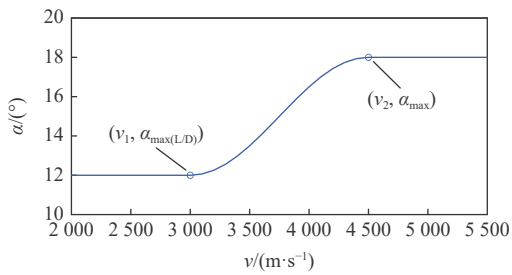


图 3 攻角-速度剖面

Fig. 3 Profile of the angle of attack-velocity

5.1 最大射程数值仿真试验

为验证变体对高超声速飞行器射程的影响，设计想定为：从指定初始状态开始再入滑翔，抵达末端置顶海拔终止。对比固定构型情况（机翼保持初始后掠角和展长不变）和变体情况下的最大射程，其中固定构型采用方案攻角加零控制量积分获得轨迹，变体构型采用所提凸优化方法在机翼变形量范围内进行轨迹优化。再入滑翔初末状态如表 3 所示。

表 3 初末状态 1

Table 3 Initial and final state 1

状态	射程 x/km	海拔 h/km	速度 v/(m·s^{-1})	弹道倾角 γ/(°)	内段翼后掠角 χ_1/(°)	内段翼后后掠角 χ_2	展长 ξ/m
初始状态	0	55	4 950	−1	8	0	0.41
终端状态		30	2 000				

ATP-SCP 算法中配点数 $N=100$，初始信赖域及收敛容差设置如下式所示，信赖域单位均为国际单位制，信赖域收缩系数 $\beta=0.7$，罚系数 $\omega=1\,000$。

$$\begin{aligned}\boldsymbol{\varepsilon}^0&=\left[3\times10^3,1.5\times10^5,\ 10^3,\frac{5\pi}{180},\ 0.1,0.1,0.1,\ 1\right]^{\mathrm{T}}\\ \boldsymbol{\delta}&=\left[50,10^3,\ 10,\frac{\pi}{180},10^{-3},10^{-3},10^{-3},\ 10^{-2}\right]^{\mathrm{T}}\end{aligned}\tag{48}$$

仿真对比结果如图 4～图 6 所示。其中，图 4 为部分状态变量对比，图 5 为再入滑翔过程中热流率、动压和过载约束随时间的变化关系，图 6 为机翼变形量随时间的变化关系，外段翼后掠角及其变化率为归一化后的结果。

从图 4 (a) 可知，由于龙格库塔数值积分时间步长的影响，固定构型再入终端高度为 29.49 km，略低于设定的终端终止高度。固定构型再入过程飞行时间为 890 s，射程为 2 994 km，变体再入过程飞行时间 1 056 s，射程为 3 492 km，飞行时间增加了 18.65%，射程增加了 16.63%，证明了变体对高超声速飞行器增程具有重要意义。根据图 5 中路径约束对比结果可知，固定构型再入过程中不满足最大动压约束，而变体构型满足所有路径约束。再入滑翔过程中，过载、动压及驻点热流率的峰值分别降低了 24.25%、30.53% 和 16.96%，对飞行器的安全稳定飞行具有重要意义。高超再入初始阶段，从初始状态到第一个滑跳点，飞行器以最大攻角 20°飞行，在开始的 0～200 s，内外段翼后掠角基本不发生变化，机翼前缘长度变大，飞行器特征面积增大，进而

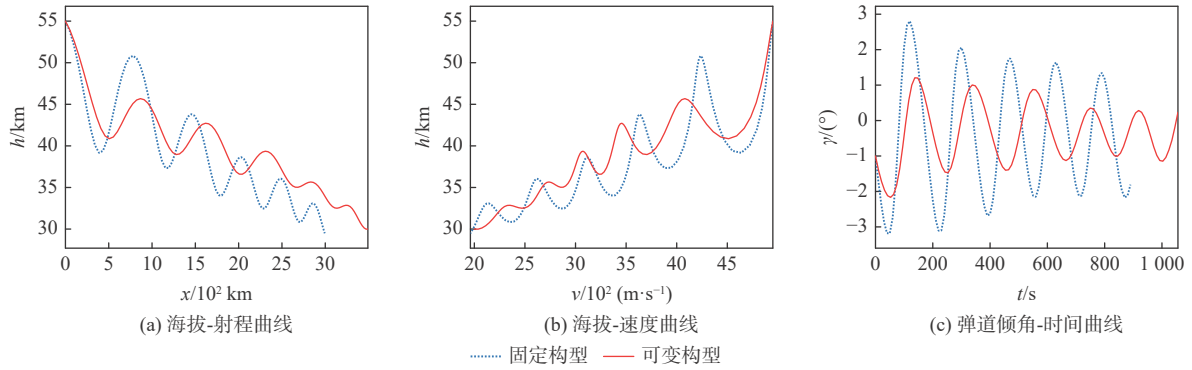


图 4　固定/可变构型状态变量对比

Fig. 4　Comparison on state variables of fixed and morphing configuration

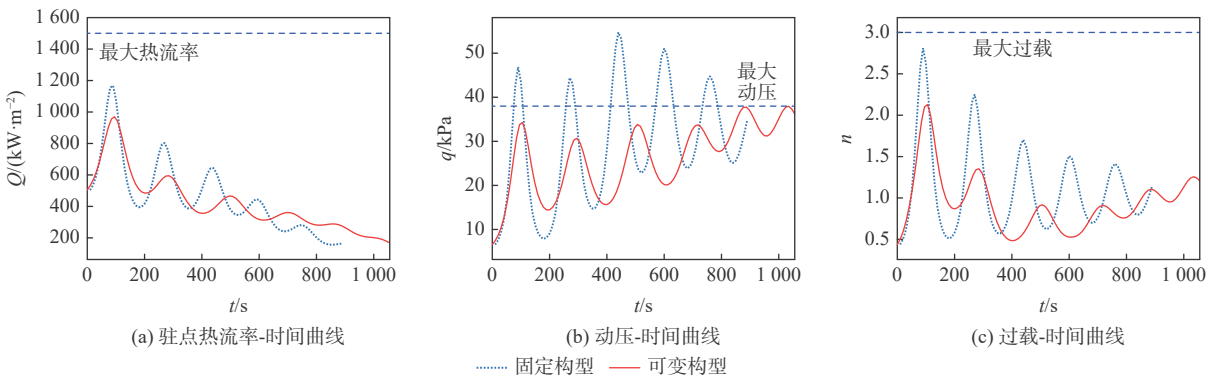


图 5　固定/可变构型路径约束对比

Fig. 5　Comparison on path constraints of fixed and morphing configuration

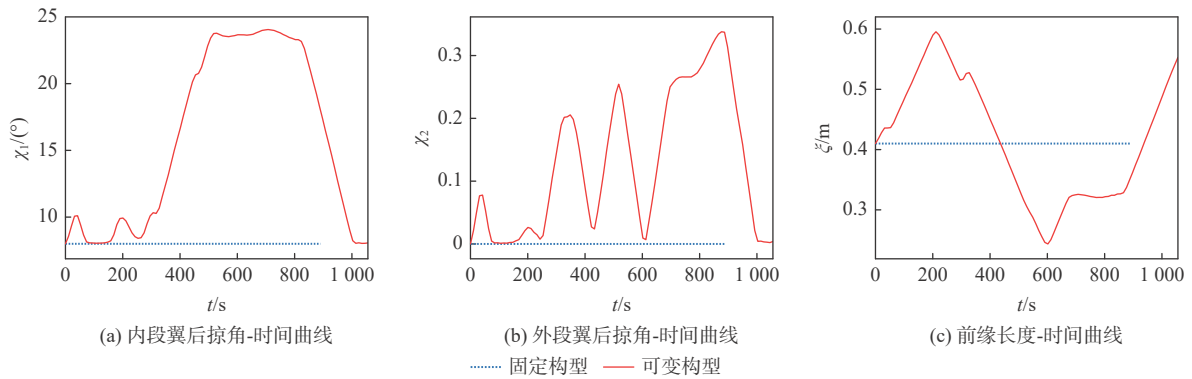


图 6　固定/可变构型机翼变形量

Fig. 6　Comparison on morphing variables of fixed and morphing configuration

增大飞行器的升力; 在 200～600 s, 前缘长度逐渐减小, 内段翼后掠角增大, 在保持较大升力的同时降低阻力; 600 s 以后, 飞行器开始保持最优升阻比对应的攻角 12°飞行。在 600～850 s, 前缘长度逐渐增大, 内段翼保持较大后掠角, 外段翼后掠角增大, 即在增大飞行器的升力的同时尽可能减小阻力; 在最后阶段, 飞行器由于速度降低, 为尽可能保持最优升阻比或逼近最优升阻比条件, 内外段翼后掠角均减小, 前缘长度变大, 增大飞行器的特征面积, 尽可能提高升力。

根据上述高超声速变体飞行器再入过程分析可知, 高超声速变体飞行器在飞行过程中根据飞行空域和速域会自适应调整后掠角与前缘长度, 通过调整升力与阻力变化, 在保持较大升阻比的同时尽可能降低阻力, 从而降低能量损失, 以期增大射程。

5.2　时效性数值仿真试验

为验证本文所提算法在时效性方面的优势, 选择目前轨迹规划领域应用最为广泛的 hp 伪谱法作为对比, 其中 hp 伪谱法基于 GPOPS-Ⅱ 工具包[32] 实现, 相关参数设置采用默认参数。

高超声速变体飞行器再入滑翔初末状态如表 4 所示, 配点数、初始信赖域、收敛容差、信赖域收缩

表 4 初末状态 2

Table 4 Initial and final state 2

变量	射程x/km	海拔h/km	速度v/(m·s^{-1})	弹道倾角γ/(°)	攻角α/(°)	内段翼后掠角χ_1/(°)	外段翼后后掠角χ_2	前缘长度ξ/m
初始状态	0	61	5 100	0	18	8	0	0.41
终端状态		30			12			

系数、罚系数等设置与 5.1 节保持一致。

统计两种算法求解耗时与性能指标结果如表 5 所示。根据结果分析可知，ATP-SCP 方法在最优性方面略低于 hp 伪谱法，相差约为 1.39%，但其求解耗时降低了 89.24%，极大地提升了高超声速变体飞行器再入滑翔轨迹规划效率。

表 5 求解耗时与最优性对比

Table 5 Comparison of solving time and optimality

算法	求解耗时/s	最大射程/km
ATP-SCP	4.10	3 695.54
hp伪谱法	38.11	3 747.67

两种算法的仿真结果如图 7～图 9 所示。对图 7 和图 8 综合分析可知，两种方法均能较好地满足再入滑翔过程的热流率、动压和过载约束，均为典型的跳跃滑翔式弹道。相较于 hp 伪谱法，所提方法获得的轨迹在再入前期速度衰减较快，滑跳幅度更低，更为平缓，而在再入后期速度衰减较慢，仍保持一定高度的滑跳，直至到达终端指定高度。两种方法的热流率和动压峰值相当，但所提方法获得的再入轨迹的过载峰值相较于 hp 伪谱法降低了约 24.89%。

图 9 给出了再入过程中机翼的变形量及控制量随再入时间的变化关系。对于内外段翼后掠角，两种方法优化出的结果趋势相当，在变化幅度上有所不同，且 hp 伪谱法优化出的后掠角幅度变化幅度大，所提算法优化出的后掠角变化幅度更小。对于前缘长度，hp 伪谱法优化出的结果在 350～850 s 有先增大后减小的趋势，变化幅度较大，且基本达到前缘长度的上限，而所提方法优化出的结果是先减小再增大，变化幅度较小，且在初始前缘长度附近变化。进一步分析可知，所提算法优化结果使得高超声速变体飞行器每次滑跳时最高点与最低点高度差更低，即能量损失更少，且控制量变化同样较为平滑。结合图 8 可知，所提方法能够使高超声速变体飞行器以较小机翼的变化量达到与 hp 伪谱法同样的效果，降低了对变形机构强度的要求，提高了变体方案的可行性。

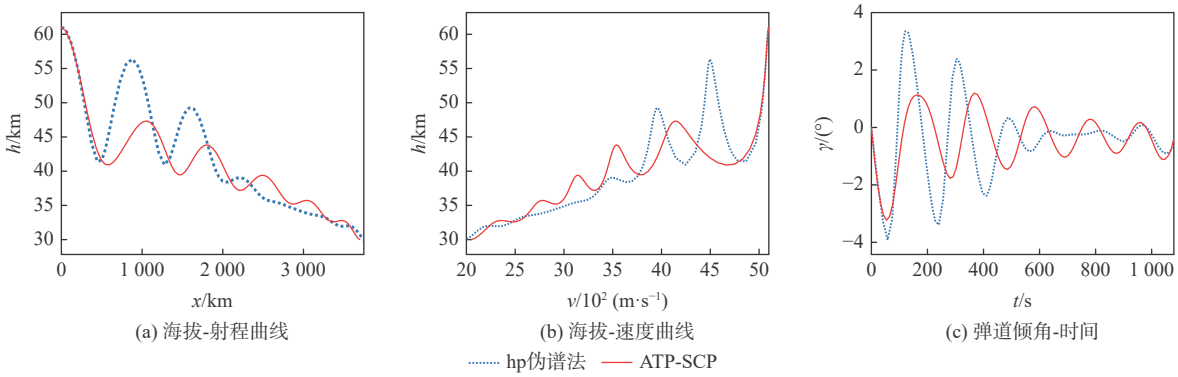


图 7 hp 伪谱法与 ATP-SCP 算法状态变量对比

Fig. 7 Comparison on state variables of hp pseudospectrum and ATP-SCP algorithm

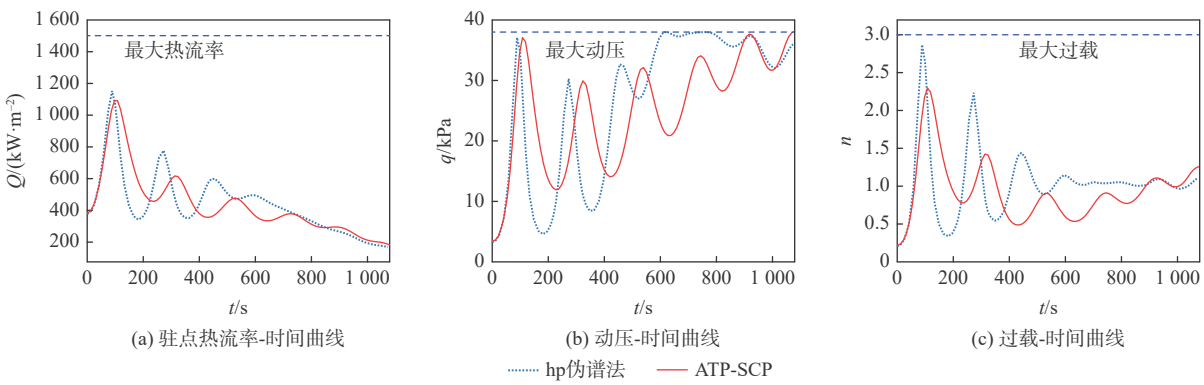
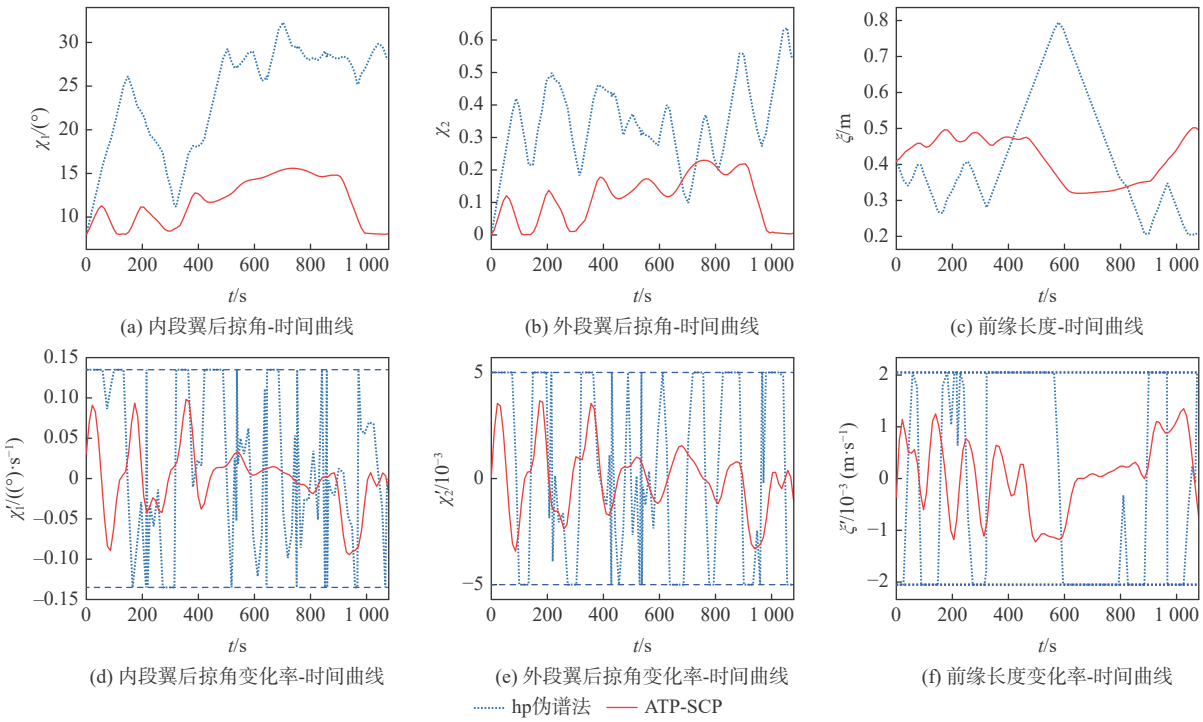


图 8 hp 伪谱法与 ATP-SCP 算法路径约束对比

Fig. 8 Comparison on path constraints of hp pseudospectrum and ATP-SCP algorithm



(a) 内段翼后掠角-时间曲线　(b) 外段翼后掠角-时间曲线　(c) 前缘长度-时间曲线

(d) 内段翼后掠角变化率-时间曲线　(e) 外段翼后掠角变化率-时间曲线　(f) 前缘长度变化率-时间曲线

图 9　hp 伪谱法与 ATP-SCP 算法变形量及对应控制变量对比

Fig. 9　Comparison on morphing and control variables of hp pseudospectrum and ATP-SCP algorithm

6　结　论

本文在类乘波体变体飞行器布局与变后掠-变展长复合变形方案基础上, 针对高超声速变体飞行器再入轨迹规划问题求解难度大、规划耗时高的问题, 开展了快速轨迹规划方法研究。

1) 提出一种基于自适应信赖域更新的罚函数序列凸优化算法 (ATP-SCP), 采用对数凸化方法减小路径约束线性化误差, 通过引入虚拟控制和自适应信赖域更新策略, 提高序列凸优化方法的鲁棒性和时效性。

2) 在最优性相当的前提下, ATP-SCP 求解高超声速变体飞行器再入轨迹规划问题的计算耗时较 hp 伪谱法降低了 89.24%, 验证了所提算法的效率优势。此外, ATP-SCP 求解结果(机翼后掠角及展长)变化幅度更小, 控制量变化更为平滑, 进一步提高了变体方案的可行性。

本文建立的高超声速变体飞行器动力学模型较为简单, 尚未考虑地球自转与飞行器姿态与变体附加力, 在后续研究中将进一步建立更为精细的变体再入动力学模型, 为再入轨迹规划提供更为准确的模型支撑。

参考文献（References）

[1] 孙聪. 高超声速飞行器强度技术的现状、挑战与发展趋势[J]. 航空学报, 2022, 43(6): 527590.
SUN C. Development status, challenges and trends of strength technology for hypersonic vehicles[J]. Acta Aeronautica et Astronautica Sinica, 2022, 43(6): 527590(in Chinese).

[2] 张灿, 王轶鹏, 叶蕾. 国外近十年高超声速飞行器技术发展综述[J]. 战术导弹技术, 2020(6): 81-86.
ZHANG C, WANG Y P, YE L. Summary of the technological development of overseas hypersonics in the past ten years[J]. Tactical Missile Technology, 2020(6): 81-86(in Chinese).

[3] 朱广生, 段毅, 姚世勇, 等. 跨流域高速飞行器气动设计研究现状及思考[J]. 宇航学报, 2023, 44(3): 358-367.
ZHU G S, DUAN Y, YAO S Y, et al. Research status and consideration on aerodynamic design of hypersonic flight vehicle covering various flow regimes[J]. Journal of Astronautics, 2023, 44(3): 358-367(in Chinese).

[4] 白鹏, 陈钱, 徐国武, 等. 智能可变形飞行器关键技术发展现状及展望[J]. 空气动力学学报, 2019, 37(3): 426-443.
BAI P, CHEN Q, XU G W, et al. Development status of key technologies and expectation about smart morphing aircraft[J]. Acta Aerodynamica Sinica, 2019, 37(3): 426-443(in Chinese).

[5] LI D C, ZHAO S W, DA RONCH A, et al. A review of modelling and analysis of morphing wings[J]. Progress in Aerospace Sciences, 2018, 100: 46-62.

[6] AJAJ R M, PARANCHEERIVILAKKATHIL M S, AMOOZGAR M, et al. Recent developments in the aeroelasticity of morphing aircraft[J]. Progress in Aerospace Sciences, 2021, 120: 100682.

[7] WEISSHAAR T A. Morphing aircraft systems: historical perspectives and future challenges[J]. Journal of Aircraft, 2013, 50(2): 337-353.

[8] LI Z H, HU C, DING C B, et al. Stochastic gradient particle swarm optimization based entry trajectory rapid planning for hypersonic glide vehicles[J]. Aerospace Science and Technology, 2018, 76: 176-186.

[9] 杨博, 朱一川, 魏延明, 等. 折叠式变体飞行器轨迹优化及控制分析[J]. 中国空间科学技术, 2020, 40(3): 64-75.
YANG B, ZHU Y C, WEI Y M, et al. Trajectory optimization and control analysis of folding wing aircraft[J]. Chinese Space Science and Technology, 2020, 40(3): 64-75(in Chinese).

[10] 王娜, 陈洁卿, 明超, 等. 基于 hp 自适应伪谱法的变后掠翼导弹弹道优化设计[J]. 弹道学报, 2016, 28(4): 24-29.
WANG N, CHEN J Q, MING C, et al. Optimization design for trajectory of morphing-wing missile based on hp-adaptive pseudospectral method[J]. Journal of Ballistics, 2016, 28(4): 24-29(in Chinese).

[11] DAI P, YAN B B, LIU R F, et al. Integrated morphing strategy and trajectory optimization of a morphing waverider and its online implementation based on the neural network[J]. IEEE Access, 2021, 9: 59383-59393.

[12] BAO C Y, WANG P, TANG G J. Integrated method of guidance, control and morphing for hypersonic morphing vehicle in glide phase[J]. Chinese Journal of Aeronautics, 2021, 34(5): 535-553.

[13] WEI Z F, CHENG Y, GUO X X, et al. Hybrid trajectory optimization method and tracking guidance for variable-sweep missiles[J]. Mathematical Problems in Engineering, 2021, 2021(1): 6669618.

[14] 陈铁彪, 龚旻, 王洪波, 等. 临近空间可变形滑翔飞行器轨迹优化与性能分析[J]. 宇航学报, 2018, 39(9): 944-952.
CHEN T B, GONG M, WANG H B, et al. Trajectory optimization and performance analysis of the near space morphing glide vehicles[J]. Journal of Astronautics, 2018, 39(9): 944-952(in Chinese).

[15] 岳彩红, 唐胜景, 郭杰, 等. 高超声速伸缩式变形飞行器再入轨迹快速优化[J]. 系统工程与电子技术, 2021, 43(8): 2232-2243.
YUE C H, TANG S J, GUO J, et al. Reentry trajectory rapid optimization for hypersonic telescopic deformable vehicle[J]. Systems Engineering and Electronics, 2021, 43(8): 2232-2243(in Chinese).

[16] ROCKAFELLAR R T. Convex analysis[M]. Princeton, NJ: Princeton University Press, 1997.

[17] LIU X F, LU P, PAN B F. Survey of convex optimization for aerospace applications[J]. Astrodynamics, 2017, 1(1): 23-40.

[18] MALYUTA D, YU Y, ELANGO P, et al. Advances in trajectory optimization for space vehicle control[J]. Annual Reviews in Control, 2021, 52: 282-315.

[19] AÇıKMEŞE B, BLACKMORE L. Lossless convexification of a class of optimal control problems with non-convex control constraints[J]. Automatica, 2011, 47(2): 341-347.

[20] HARRIS M W, AÇıKMEŞE B. Lossless convexification of non-convex optimal control problems for state constrained linear systems[J]. Automatica, 2014, 50(9): 2304-2311.

[21] MAO Y Q, SZMUK M, AÇıKMEŞE B. Successive convexification of non-convex optimal control problems and its convergence properties[C]//2016 IEEE 55th Conference on Decision and Control (CDC). Piscataway: IEEE Press, 2016: 3636-3641.

[22] LIU X F, SHEN Z J, LU P. Entry trajectory optimization by second-order cone programming[J]. Journal of Guidance, Control, and Dynamics, 2016, 39(2): 227-241.

[23] WANG Z B, GRANT M J. Constrained trajectory optimization for planetary entry *via* sequential convex programming[J]. Journal of Guidance, Control, and Dynamics, 2017, 40(10): 2603-2615.

[24] 陈铁彪, 王洪波, 康永来, 等. 变形技术在助推-滑翔飞行器中的应用前景分析[J]. 战术导弹技术, 2017(5): 1-5,11.
CHEN T B, WANG H B, KANG Y L, et al. Application prospects of morphing technology in the boost-glide vehicle[J]. Tactical Missile Technology, 2017(5): 1-5,11(in Chinese).

[25] 武宇飞. 多源模型驱动的高超声速变体飞行器总体方案设计优化研究[D]. 北京: 北京理工大学, 2023.
WU Y F. Multi-models driven conceptual design optimization of hypersonic morphing aircraft[D]. Beijing: Beijing Institute of Technology, 2023(in Chinese).

[26] 陈钱, 白鹏, 李锋. 可变形飞行器机翼两种变后掠方式及其气动特性机理[J]. 空气动力学学报, 2012, 30(5): 658-663.
CHEN Q, BAI P, LI F. Morphing aircraft wing variable-sweep: two practical methods and their aerodynamic characteristics[J]. Acta Aerodynamica Sinica, 2012, 30(5): 658-663(in Chinese).

[27] VALE J. Aero-structural optimization and performance evaluation of a morphing wing with variable span and camber[J]. Journal of Intelligent Material Systems and Structures, 2011, 22(10): 1057-1073.

[28] VIANA F A C, GOGU C, GOEL T. Surrogate modeling: tricks that endured the test of time and some recent developments[J]. Structural and Multidisciplinary Optimization, 2021, 64(5): 2881-2908.

[29] 李广华. 高超声速滑翔飞行器运动特性分析及弹道跟踪预报方法研究[D]. 长沙: 国防科学技术大学, 2016.
LI G H. Motion characteristics analysis and trajectory prediction for hypersonic glide vehicles[D]. Changsha: National University of Defense Technology, 2016(in Chinese).

[30] LOFBERG J. YALMIP: a toolbox for modeling and optimization in matlab[C]//Proceedings of the 2004 IEEE International Conference on Robotics and Automation. Piscataway: IEEE, 2004: 284-289.

[31] DOMAHIDI A, CHU E, BOYD S. ECOS: an SOCP solver for embedded systems[C]//2013 European Control Conference (ECC). Piscataway: IEEE Press, 2013: 3071-3076.

[32] PATTERSON M A, RAO A V. GPOPS-II: a MATLAB software for solving multiple-phase optimal control problems using hp-adaptive Gaussian quadrature collocation methods and sparse nonlinear programming[J]. ACM Transactions on Mathematical Software, 2014, 41(1): 1-37.

Re-entry trajectory planning for hypersonic morphing vehicles using penalty sequence convex programming

WANG Yangjie[1], LONG Teng[1, 2], LI Junzhi[1], XU Guangtong[3], SUN Jingliang[1, 2, *]

(1. School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China;
2. Chongqing Innovation Center, Beijing Institute of Technology, Chongqing 401120, China;
3. Huzhou Institute of Zhejiang University, Huzhou 313000, China)

Abstract: To realize continuous leapfrog upgrades of the hypersonic vehicle from single-point optimal fixed configuration to full envelope optimal of morphing configuration, a quasi-wave rider profile and composite deformation scheme morphing wingspan and sweep are designed. On this basis, to reduce the computational burdens of reentry trajectory planning, the adaptive trust-region-based penalty sequence convex programming method is proposed. To increase the approximate accuracy, the path restrictions are communicated using the logarithmic convexification technique. A virtual control is introduced to replace the dynamic equation constraints. Using the penalty function method, modify the second-order cone constraint and incorporate it into the objective function to direct the iterative results in order to approximate the feasible domain. An adaptive trust region updating strategy is designed to accelerate the convergence of the sequence convex optimization algorithm. As demonstrated by the simulation results, the hypersonic morphing vehicle's range extension is 16.63% when compared to the fixed configuration, and the ATP-SCP computing time is 89.24% less than when compared to the HP pseudospectral method.

Keywords: hypersonic morphing vehicle; reentry trajectory planning; logarithmic convexification; virtual control; adaptive trust region

Received: 2023-05-26; **Accepted**: 2023-09-08; **Published Online**: 2023-10-20 13:08

URL: link.cnki.net/urlid/11.2625.V.20231020.1023.003

Foundation items: National Natural Science Foundation of China (62003036,62203256); Aeronautical Science Foundation of China (2019ZC072003); Beijing Institute of Technology Research Fund Program for Young Scholars Foundation (XSQD-202201005); BIT Research and Innovation Promoting Project (2022YCXZ017)

* **Corresponding author**. E-mail: sunjingliangac@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0299

重复使用火箭垂直着陆段在线轨迹优化方法

施棋[1,2]，齐瑞云[1,2,*]，佘宇琛[3]，胡存明[3]

(1. 南京航空航天大学 航空航天结构力学及控制全国重点实验室，南京 210016；2. 南京航空航天大学 自动化学院，南京 211106；
3. 上海航天控制技术研究所，上海 201109)

摘　　要：重复使用运载火箭垂直着陆段在制导中存在初始条件不确定、过程约束复杂和终端约束严苛的问题，针对这些问题，提出了一种结合凸优化和四次多项式制导的在线轨迹优化方法。采用考虑终端姿态约束和以燃料最优为目标的四次多项式制导方法，计算得到初始轨迹和终端着陆时间；根据不同的初始条件得到不同的轨迹和时间，确保在不同初始条件下凸优化均能求解成功。同时，将初始轨迹代入同样以燃料最优为目标的凸优化算法，得到制导指令并满足过程约束和终端约束条件，实现软着陆。仿真分析表明相比于常值凸优化算法，所提出的方法对不同偏差的适应性更好，着陆成功率更高，并且省略了着陆终端时间寻优的过程，求解效率更高，使用的燃料更少。

关 键 词：重复使用运载火箭；垂直着陆段；凸优化；轨迹优化；软着陆
中图分类号：V448; TJ765
文献标志码：A　　**文章编号**：1001-5965(2025)05-1760-10

由于航天事业的成本巨大，自从20世纪60年代开始，可重复使用运载器的概念应运而生[1]。各国都积极探索重复使用运载技术在航天事业中的可能性。从最初实现部分可重复使用的航天飞机，到现在太空探索公司(SpaceX公司)发射的可成功实现地面或海上软着陆的Falcon运载火箭，重复使用运载技术在最近几十年内不断发展和进步。目前可重复使用运载器可以通过起飞和降落的方式分成三大类：垂直起飞水平降落(vertical take-off horizontal landing, VTHL)运载器、垂直起飞垂直降落(vertical take-off vertical landing, VTVL)运载器和水平起飞水平降落(horizontal take-off horizontal landing, HTHL)运载器。本文所讨论的是VTVL运载火箭。该运载火箭在一二级分离之后，二级运载火箭继续上升完成既定的任务，火箭一子级返回地面，在返回过程中需要经历6个阶段分别是：调姿段、修航段、高空下降段、动力下降段、气动减速段和垂直着陆段[2]。可重复使用运载火箭的最后一个阶段是垂直着陆段，历经此段后最终软着陆于预设位置。因此垂直着陆段是整个火箭一子级返回过程的成功与否的关键。垂直着陆段的前一个阶段是气动减速段，气动减速段对火箭有约束和干扰，使得垂直着陆段初始的条件存在一定的偏差，火箭需要根据初始条件和当前火箭状态进行在线轨迹优化。火箭的回收方式也分为两种：海上回收和陆地回收。本文选取的火箭所使用的回收方式为海上回收。海上着陆方式能够降低火箭的飞行距离，减少所需要的燃料，相比于陆地回收，不需要调整火箭的速度和进入地球大气层时的速度。但海上回收平台的面积更小，需要更精确的定点着陆技术，并且垂直着陆段对于终端约束和过程约束都有着很高的要求，海上回收存在着严重的气动扰动

收稿日期：2023-05-30；**录用日期**：2023-08-11；**网络出版时间**：2023-09-08 14:42
网络出版地址：link.cnki.net/urlid/11.2625.V.20230907.0942.004
基金项目：国家自然科学基金(62373189)；航空航天结构力学及控制全国重点实验室开放课题
* **通信作者**. E-mail：ruiyun.qi@nuaa.edu.cn

*引用格式：施棋，齐瑞云，佘宇琛，等. 重复使用火箭垂直着陆段在线轨迹优化方法[J]. 北京航空航天大学学报，2025，51(5)：1760-1769.
SHI Q，QI R Y，SHE Y C，et al. Online trajectory optimization method for vertical landing phase of reusable launch vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1760-1769 (in Chinese).*

会影响着陆的精度, 对制导控制方法的研究有更高的要求。

重复使用运载火箭着陆段根据计算的特点主要可以分成三大类: 解析制导法、计算制导法和基于学习的制导方法等[3]。解析制导法通常可以对着陆的位置进行精确控制, 如多项式制导算法, 该算法已应用于阿波罗登月计划中登月舱的着陆制导段[4]。我国的嫦娥三号着陆器也采用改进的多项式制导实现月面软着陆[5]。文献 [6] 考虑终端姿态约束基于最优控制方法设计出重复使用运载火箭的四次多项式制导律, 并用蒙特卡罗打靶验证了在大气环境各种偏差下火箭一子级能够成功实现软着陆。文献 [7] 利用线性二次型调节器对着陆段进行跟踪制导, 该算法的优点就在于能够在飞行时间的函数中包含着陆器的标称位置、速度和加速度, 并利用最优控制的理论求解系数。但是该算法不能考虑火箭在飞行过程中的约束, 并且在算法中不考虑气动阻力, 不能应用于复杂环境。

计算制导法能够实时获得制导律, 基于优化算法在线求解包含目标函数、运动学约束、过程约束和初始终端约束的问题, 求解的结果为最优轨迹和最优控制指令, 其相对于解析制导法最大的优点在于包含过程约束并且能够进行在线轨迹优化。在计算制导法中应用比较广泛的是凸优化方法, 文献 [8] 率先提出了将轨迹优化问题转化为一个二阶锥规划(second-order cone program, SOCP)问题, 该问题可以根据燃料最优的性能指标和初始终端过程约束等计算出一条最优轨迹。文献 [9] 针对 Xombie 垂直起降运载火箭利用无损凸化的方法设计了针对该火箭的二阶锥问题求解器, 并可以在有干扰和不确定性的情况下在线求解出满足可靠性和着陆精度的最优轨迹。文献 [10] 采用了凸优化方法求解了以燃料最优为目标的制导律, 导引火箭到达着陆点。文献 [11] 提出了一种轨迹优化算法, 该算法将伪谱法与改进的凸优化方法相结合, 同时提升了算法的精度和快速性。文献 [12] 根据火箭在动力减速段面临的约束, 基于分段凸优化和线性二次型调节器求解出以燃料最优为目标的轨迹跟踪制导律。文献 [13] 结合偏置比例导引法和凸优化, 求解了存在落角、落速和推力约束的火箭。文献 [14] 将序列凸优化应用于模型方程的笛卡儿表示, 并将此方法同时应用于气动控制下降段和动力着陆段。文献 [15] 协调了气动控制和推进以实现燃料最优的精确着陆, 以攻角和推力同时作为输入, 使用凸优化方法求解着陆问题。文献 [16] 设计了一种同伦迭代策略凸化火箭系统中的非线性项, 此方法提高了系统的自主性。文献 [17] 为了缓解着陆段初期的不确定性的影响, 使用最优反馈制导算法, 在每个制导周期内重新规划最优轨迹实现软着陆。而在基于学习的制导方法中, 目前研究较为广泛的是主要基于深度学习和强化学习的制导方法。文献 [18] 针对垂直起降的可重复使用运载火箭着陆段的制导问题, 基于强化学习算法得到着陆段的制导策略并验证了该策略在不同偏差下的着陆精度。该制导方法是不基于模型, 然而, 目前有效的真实训练样本比较少, 难以保证所训练出来的控制策略的优越性。

在现有的着陆段的制导方法中, 解析制导法不考虑过程约束和气动阻力, 对当前存在严重气动扰动的火箭并不适用; 凸优化方法处理终端时间不固定的着陆问题, 需要遍历寻找最优时间效率较低, 在初始条件具有偏差的情况下, 一般的凸优化方法不满足在线的需求; 而基于学习的制导方法不能保证控制策略有效, 3 种制导方法均不满足于在线轨迹优化问题的需求。因此为了提高凸优化算法的效率, 并且同时保证控制策略的有效性, 本文提出了一种在线轨迹优化算法。

本文基于凸优化方法求解考虑气动阻力可重复使用运载火箭着陆段的轨迹优化问题, 引入 4 次多项式制导方法计算初始轨迹, 所得到的初始轨迹相较线性化方法更加符合实际, 更容易使得凸优化算法收敛。通过 4 次多项式制导能够得到不同初始条件下不同的着陆终端时间, 相比于遍历寻优提高了算法的计算效率, 因此采用 4 次多项式制导和凸优化算法结合能够更精确并且更快得到满足约束条件的控制指令, 引导火箭成功实现软着陆。同时, 本文验证了该算法在各种偏差下的适应性, 实现了可重复使用运载火箭垂直着陆段的在线轨迹优化。

1　可重复使用运载火箭着陆动力学模型

1.1　可重复使用运载火箭动力着陆段动力学模型

首先以目标着陆点为参考点建立固定坐标系, x 轴方向为水平横向, z 轴方向为水平纵向, y 轴方向为垂直于着陆平台。在软着陆坐标系下建立三自由度运载火箭动力学模型, 不考虑火箭的姿态控制, 仅考虑其质心的三自由度平移, 并且作出以下假设: 1)忽略地球的旋转和曲率的变化, 认为海平面是一个平面; 2)由于着陆段的高度已位于大气的最底层, 可以忽略地球重力加速度的变化; 3)火箭的发动机推力始终沿着箭体轴的方向; 4)只考虑气

动加速度阻力,忽略其他气动力。

根据软着陆坐标系建立可重复使用运载火箭一子级的动力学模型[19]:

$$\begin{cases}\dot{\boldsymbol{r}}(t)=\boldsymbol{v}(t)\\ \dot{\boldsymbol{v}}(t)=\boldsymbol{g}+\boldsymbol{T}(t)/m(t)+\boldsymbol{a}_{\mathrm{D}}(t)\\ \dot{m}(t)=-\dfrac{\|\boldsymbol{T}(t)\|}{g_0 I_{\mathrm{sp}}}=-\alpha\|\boldsymbol{T}(t)\|\end{cases}\tag{1}$$

式中:$\boldsymbol{r}=\begin{bmatrix}r_x & r_y & r_z\end{bmatrix}^{\mathrm{T}}$和$\boldsymbol{v}=\begin{bmatrix}v_x & v_y & v_z\end{bmatrix}^{\mathrm{T}}$分别为火箭三维的位置和速度矢量;$\boldsymbol{g}$为重力加速度矢量,$\boldsymbol{g}=[0\quad -g_0\quad 0]^{\mathrm{T}}$;$\boldsymbol{T}=\begin{bmatrix}T_x & T_y & T_z\end{bmatrix}^{\mathrm{T}}$为火箭三维的推力矢量;$m$为火箭的质量,火箭的初始质量为结构质量加上初始燃料质量,$m_0=m_{\mathrm{wet}}=m_{\mathrm{dry}}+m_{\mathrm{fuel}}$;$I_{\mathrm{sp}}$为火箭子级的发动机比冲;$\boldsymbol{a}_{\mathrm{D}}$为火箭的空气阻力加速度矢量,其计算公式为

$$\boldsymbol{a}_{\mathrm{D}}(t)=-\frac{C_{\mathrm{D}}S_{\mathrm{ref}}}{2m(t)}\rho\|\boldsymbol{v}(t)\|\boldsymbol{v}(t)\tag{2}$$

式中:C_{D}为气动力轴向系数;S_{ref}为着陆时火箭子级的气动参考面积;ρ为大气密度。大气密度与高度成反比:

$$\rho=\rho_0\mathrm{e}^{-\beta h}\tag{3}$$

式中:h为高度;β为高度密度比;ρ_0为标准密度。根据火箭的质量表达式,可以计算出火箭在飞行过程中的剩余质量与初始质量的比值:

$$\begin{cases}\displaystyle\int_{t_0}^{t_{\mathrm{c}}}\frac{\dot{m}(t)}{m(t)}\mathrm{d}t=-\alpha\int_{t_0}^{t_{\mathrm{c}}}\frac{\|\boldsymbol{T}(t)\|}{m(t)}\mathrm{d}t\\ \ln\dfrac{m(t_{\mathrm{c}})}{m(t_0)}=-\alpha\displaystyle\int_{t_0}^{t_{\mathrm{c}}}\frac{\|\boldsymbol{T}(t)\|}{m(t)}\mathrm{d}t\\ \dfrac{m(t_{\mathrm{c}})}{m(t_0)}=\mathrm{e}^{-\alpha\int_{t_0}^{t_{\mathrm{c}}}\frac{\|\boldsymbol{T}(t)\|}{m(t)}\mathrm{d}t}\end{cases}\tag{4}$$

式中:t_{c}为火箭子级在着陆时从初始点到当前的状态飞行的时间;t_0为初始时刻。因此,火箭在飞行过程中的剩余质量为

$$m(t_{\mathrm{c}})=m(t_0)\mathrm{e}^{-\alpha\int_{t_0}^{t_{\mathrm{c}}}\frac{\|\boldsymbol{T}(t)\|}{m(t)}\mathrm{d}t}\tag{5}$$

1.2 可重复使用运载火箭动力着陆段约束条件

火箭一子级在垂直着陆段的初始阶段、飞行过程中和末端都需要满足各种约束条件,包括位置、速度等的精度要求,还有姿态、燃料等的约束,并且在整个飞行过程中火箭需要满足状态方程约束。所有的控制变量和状态变量都需要满足1.1节所描述的火箭一子级动力学模型中的关系表达式。

1.2.1 初始和终端约束

火箭一子级的初始约束即为气动减速段末端着陆段开始时各状态变量的初始值,终端约束即为飞行末段的位置和速度在理想状态同时降到0 m和0 m/s。由于1.1节中假设3)的存在,火箭在着陆时也要求火箭的姿态保持严格的垂直,因此火箭的加速度矢量在水平横向和侧向在理想状态下需要降到0 m/s^2,但在竖直方向不为0 m/s^2,因此定义初始和终端约束如下:

$$\begin{cases}\boldsymbol{r}(t_0)=\boldsymbol{r}_0,\boldsymbol{v}(t_0)=\boldsymbol{v}_0,m(t_0)=m_0\\ \boldsymbol{r}(t_{\mathrm{f}})=[0\quad 0\quad 0]^{\mathrm{T}},\boldsymbol{v}(t_{\mathrm{f}})=[0\quad 0\quad 0]^{\mathrm{T}}\\ \boldsymbol{T}(t_{\mathrm{f}})=\begin{bmatrix}0 & T_y(t_{\mathrm{f}}) & 0\end{bmatrix}^{\mathrm{T}}\\ T_y(t_{\mathrm{f}})\neq 0\end{cases}\tag{6}$$

式中:t_0为初始时刻;t_{f}为终端时刻;$\boldsymbol{r}_0$和$\boldsymbol{v}_0$分别为三轴初始位置和速度向量;m_0为火箭一子级初始质量;$T_y(t_{\mathrm{f}})$为y轴末端推力。

由于垂直着陆段是整个飞行的最后阶段,所剩余的燃料不会太多,因此还需要满足燃料约束,所用的燃料不能大于所剩余的燃料。最终的剩余质量需大于火箭一子级除燃料外的结构质量:

$$m_{\mathrm{dry}}\leqslant m(t_{\mathrm{f}})\tag{7}$$

为了保证火箭在着陆时不会与海上着陆平台相撞,因此设定高度约束条件:

$$r_y(t)\geqslant 0\tag{8}$$

1.2.2 推力约束

推力约束中包括最大推力和最小推力约束,推力调节速率约束。

火箭一子级着陆段所使用的发动机的性能是有限的,其调节范围和调节速率都会受到限制:

$$\begin{cases}T_{\min}\leqslant\|\boldsymbol{T}(t)\|\leqslant T_{\max}\\ \left|\dfrac{\mathrm{d}\|\boldsymbol{T}(t)\|}{\mathrm{d}t}\right|\leqslant\Delta T_{\max}\end{cases}\tag{9}$$

式中:$T_{\min}$和$T_{\max}$分别为最小推力和最大推力;$\Delta T_{\max}$为最大推力变化速率。

为了保证火箭飞行过程中的安全性和姿态的稳定,火箭的推力与y轴方向之间的夹角有推力指向约束:

$$T_y(t)\geqslant\|\boldsymbol{T}(t)\|\cos\theta_{\max}\tag{10}$$

式中:$\theta_{\max}$是火箭一子级推力与y轴方向之间的最大偏差夹角。

1.2.3 目标函数

最后阶段的火箭剩余燃料不会太多,需要以燃料最省的方式完成火箭的软着陆,得出火箭子级在过程中的推力控制矢量。

根据求解的目标得出凸优化的目标函数是使得最终的剩余质量最大,根据函数的单调性,目标函数为

$$J=\int_{t_0}^{t_f}\frac{\|\boldsymbol{T}(t)\|}{m(t)}\mathrm{d}t \tag{11}$$

上述所描述的优化问题中式 (1) 为动力学约束, 式 (6)～式 (10) 为初始、过程和终端约束, 式 (11) 为目标函数。该问题中仍然存在着非凸约束—质量随时间变化的非凸约束和推力的非凸约束, 需要将这些非凸约束进行凸化处理。

1.2.4　优化问题凸化处理

根据文献 [20] 对模型的凸化处理过程和证明过程, 对本文的优化问题进行无损凸化。引入松弛变量$\Gamma(t)$代替$\|\boldsymbol{T}(t)\|$, 并引入新的约束:

$$\begin{cases}\|\boldsymbol{T}(t)\|\leqslant\Gamma(t),\\ T_{\min}\leqslant\Gamma(t)\leqslant T_{\max}\end{cases} \tag{12}$$

根据文献 [20], 求解引入松弛变量后的最优控制问题得到的最优解仍然为原问题的最优解, 并能够将推力的非凸约束转化为凸约束。而质量出现在分母导致的非凸性通过变量替代的方式消除其影响, 令

$$\begin{cases}\sigma(t)=\dfrac{\Gamma(t)}{m(t)}\\ \boldsymbol{u}(t)=\dfrac{\boldsymbol{T}(t)}{m(t)}\\ z(t)=\ln(m(t))\end{cases} \tag{13}$$

将新的变量代入目标函数可以得到燃料最优等价于$\int_{t_0}^{t_f}\sigma(t)\mathrm{d}t$极小, 状态方程中的质量表达式$\dot{m}(t)=-\alpha\|\boldsymbol{T}(t)\|$就变成了$\dot{z}(t)=-\alpha\sigma(t)$。变量代换后状态变量为$\boldsymbol{x}=\begin{bmatrix}\boldsymbol{r}^{\mathrm{T}} & \boldsymbol{v}^{\mathrm{T}} & z\end{bmatrix}^{\mathrm{T}}$, 根据新的变量将约束中的部分不等式转换成新的形式:

$$\begin{cases}\ln m_{\mathrm{dry}}\leqslant z(t_{\mathrm{f}})\\ T_{\min}\mathrm{e}^{-z(t)}\leqslant\sigma(t)\leqslant T_{\max}\mathrm{e}^{-z(t)}\\ \left|\sigma(t_{i+1})\mathrm{e}^{z(t_{i+1})}-\sigma(t_i)\mathrm{e}^{z(t_i)}\right|\leqslant\Delta T_{\max}\\ u_y(t)\geqslant\sigma(t)\cos\theta_{\max}\end{cases} \tag{14}$$

式中: 第二项推力上下限约束的左侧不等式为凸集, 但是右侧不等式为非凸集, 两者均不符合锥规划的原则, 因此需要对其进行处理, 将$\mathrm{e}^{-z(t)}$利用泰勒级数在$z_0(t)$处展开, 用展开后的多项式近似为原来的函数, 因此展开后式 (14) 的第二个不等式变为

$$T_{\min}\mathrm{e}^{-z_0(t)}\left(1-(z(t)-z_0(t))+\frac{(z(t)-z_0(t))^2}{2}\right)\leqslant\sigma(t)\leqslant T_{\max}\mathrm{e}^{-z_0(t)}(1-(z(t)-z_0(t))) \tag{15}$$

式中: $z_0(t)=\ln(m_{\mathrm{wet}}-\alpha T_{\max}t)$表征的含义为系统在每个 t 时刻都采用最大推力导致的最小质量的对数, 因此对于质量对数产生了新的约束条件:

$$\ln(m_0-\alpha T_{\max}t)\leqslant z(t)\leqslant\ln(m_0-\alpha T_{\min}t) \tag{16}$$

原问题转化成了 SCOP 问题:

$$\begin{cases}\min J=\displaystyle\int_{t_0}^{t_f}\sigma(t)\mathrm{d}t\\ \quad\dot{\boldsymbol{x}}(t)=f(\boldsymbol{x}(t),\boldsymbol{u}(t),t)\\ \quad\boldsymbol{r}(t_0)=\boldsymbol{r}_0,\boldsymbol{v}(t_0)=\boldsymbol{v}_0,m(t_0)=m_0\\ \quad\boldsymbol{r}(t_{\mathrm{f}})=[0\ \ 0\ \ 0]^{\mathrm{T}},\boldsymbol{v}(t_{\mathrm{f}})=[0\ \ 0\ \ 0]^{\mathrm{T}},\ \|\boldsymbol{u}(t)\|\leqslant\sigma(t)\\ \quad\boldsymbol{T}(t_{\mathrm{f}})=\begin{bmatrix}0 & T_y(t_{\mathrm{f}}) & 0\end{bmatrix}^{\mathrm{T}},T_y(t_{\mathrm{f}})\neq 0\\ \quad r_y(t)\geqslant 0,u_y(t)\geqslant\sigma(t)\cos\theta_{\max}\\ \quad\ln m_{\mathrm{dry}}\leqslant z(t_{\mathrm{f}})\\ \quad\ln(m_{\mathrm{wet}}-\alpha T_{\max}t)\leqslant z(t)\leqslant\ln(m_{\mathrm{wet}}-\alpha T_{\min}t)\\ \quad T_{\min}\mathrm{e}^{-z_0(t)}\left(1-(z(t)-z_0(t))+\dfrac{(z(t)-z_0(t))^2}{2}\right)\leqslant\sigma(t)\\ \quad\sigma(t)\leqslant T_{\max}\mathrm{e}^{-z_0(t)}(1-(z(t)-z_0(t)))\\ \quad\left|\sigma(t_{i+1})\mathrm{e}^{z(t_{i+1})}-\sigma(t_i)\mathrm{e}^{z(t_i)}\right|\leqslant\Delta T_{\max}\end{cases} \tag{17}$$

2　凸优化问题离散化

2.1　利用初始轨迹进行线性化

该问题中, 状态方程的阻力项和推力调节约束中的指数函数中仍然存在着非线性项会导致求解的复杂化, 因此需要产生初始化轨迹, 利用初始化轨迹对其进行线性化, 并且离散时需要给出终端时间t_{f}, 本文采用文献 [6] 设计的四次多项式制导的方法给出初始化轨迹和终端时间, 四次多项式制导以燃料最优作为优化目标, 同时在求解制导律的过程中引入了姿态约束, 这与凸优化求解的目标函数和终端推力约束是相吻合的, 因此, 使用四次多项式制导求解出的初始轨迹与最优轨迹更接近, 并且四次多项式制导求解初始轨迹的速度快, 符合在线求解的需求。但是四次多项式制导也存在问题, 在制导律求解的过程中四次多项式制导不包含气动阻力项, 不能将其直接当作实际轨迹使用。

根据四次多项式制导得出控制加速度为

$$\boldsymbol{u}_a=\boldsymbol{a}_0-\boldsymbol{g}=\boldsymbol{a}_{\mathrm{f}}+\frac{12(\boldsymbol{r}_{\mathrm{f}}-\boldsymbol{r}_0)}{t_g^2}-\frac{6(\boldsymbol{v}_{\mathrm{f}}+\boldsymbol{v}_0)}{t_g}-\boldsymbol{g} \tag{18}$$

式中: t_g为火箭子级当前状态到着陆点的剩余飞行时间; $\boldsymbol{a}_0$为实时加速度指令; $\boldsymbol{a}_{\mathrm{f}}$为终端加速度指令。

由于最后的垂直着陆段的终端时间不固定, 根据文献 [6] 中哈密顿函数$H^*(t_g)=0$能够得出有关于t_g的四次多项式为

$$\boldsymbol{g}^2t_g^4-4\left(\boldsymbol{v}_{\mathrm{f}}^2+\boldsymbol{v}_0^2+\boldsymbol{v}_{\mathrm{f}}^{\mathrm{T}}\boldsymbol{v}_0\right)t_g^2+24(\boldsymbol{v}_{\mathrm{f}}+\boldsymbol{v}_0)^{\mathrm{T}}(\boldsymbol{r}_{\mathrm{f}}-\boldsymbol{r}_0)t_g-36(\boldsymbol{r}_{\mathrm{f}}-\boldsymbol{r}_0)^2=0 \tag{19}$$

从式(18)可以看出，在t_g等于0处会出现奇异，文献[6]中使用剩余时间保持的方法避免奇异，而本文则设定求解停止条件，当火箭子级在y轴方向的实时距离小于设定值时停止求解。常值凸优化算法在寻找最优着陆时间时，需要在外部建立循环，算法效率很差。而本文可以使用四次多项式制导记录根据初始状态求解出的剩余时间，取整后作为最终的着陆终端时间，而不需要寻找最优着陆时间，减小计算量。

用四次多项式制导，求解得到初始参考轨迹和终端时间，初始参考轨迹用于第一次凸优化求解，将相应的值代入阻力项和推力调节约束中的指数函数，而凸优化后续的求解过程中所使用的参考轨迹为上一次凸优化求解得到的轨迹，迭代得到最终的最优解。

2.2 问题离散化

凸优化问题需要求解必须是一个有限维的离散问题，目前的变量替代后的优化问题还是一个无限维的优化问题，因此，必须对其进行离散化。离散化首先需要选择时间间隔Δt，根据时间区间$[0,t_f]$可以得到时间节点个数N，利用这些参数进行连续系统的状态方程离散化。

将动力学方程写成矩阵形式：

$$\begin{bmatrix}\dot{\boldsymbol{r}}(t)\\ \dot{\boldsymbol{v}}(t)\\ \dot{z}(t)\end{bmatrix}=\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & -\dfrac{C_{\mathrm{D}}S_{\mathrm{ref}}}{2m(t)}\rho\|\boldsymbol{v}(t)\| & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{r}(t)\\ \boldsymbol{v}(t)\\ z(t)\end{bmatrix}+\begin{bmatrix}\boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}\\ \boldsymbol{I} & \boldsymbol{I} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0} & -\alpha\end{bmatrix}\begin{bmatrix}\boldsymbol{u}(t)\\ \boldsymbol{g}\\ \sigma(t)\end{bmatrix}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}\boldsymbol{\eta}(t) \tag{20}$$

式中：$\boldsymbol{0}$为维数为3×3的零矩阵，$\boldsymbol{I}$为维数为3×3的单位矩阵，$\boldsymbol{A}$矩阵中的$-(C_{\mathrm{D}}S_{\mathrm{ref}})/(2m(t))\rho\|\boldsymbol{v}(t)\|$项为非线性项，在第一次离散时需使用4次多项式求解初始轨迹代入非线性项，之后的离散都是用上一个迭代产生的参考轨迹代入该项，因此可以使用线性定常系统的离散方式：

$$\boldsymbol{x}(k+1)=\boldsymbol{G}\boldsymbol{x}(k)+\boldsymbol{h}\boldsymbol{\eta}(k) \tag{21}$$

式中：$\boldsymbol{G}(\Delta t)=\boldsymbol{\phi}(t)|_{t=\Delta t}$；$\boldsymbol{h}(\Delta t)=\left(\int_0^{\Delta t}\boldsymbol{\phi}(t)\mathrm{d}t\right)\boldsymbol{B}$，$\boldsymbol{\phi}(t)$为状态转移矩阵；$\boldsymbol{\phi}(t)=\boldsymbol{L}^{-1}\left[(s\boldsymbol{I}-\boldsymbol{A})^{-1}\right]$。

推力调节约束中的指数函数与动力学方程中的非线性项类似，均用上一个迭代产生的轨迹近似替代，并且离散后可以将目标函数近似为

$$\min J=\int_{t_0}^{t_f}\sigma(t)\mathrm{d}t\approx\min J=\sum_{k=1}^{N}\sigma(k)\Delta t \tag{22}$$

离散化后的最优控制问题变为

$$\begin{cases}\min J=\displaystyle\sum_{k=1}^{N}\sigma(k)\Delta t\\ x(t_{k+1})=G_k^{i-1}x(t_k)+h_{i_k}\eta(t_k)\\ \boldsymbol{r}(t_0)=\boldsymbol{r}_0,\boldsymbol{v}(t_0)=\boldsymbol{v}_0,m(t_0)=m_0\\ \boldsymbol{r}(t_{\mathrm{f}})=[0\quad 0\quad 0]^{\mathrm{T}},\boldsymbol{v}(t_{\mathrm{f}})=[0\quad 0\quad 0]^{\mathrm{T}},\ \|\boldsymbol{u}(t_k)\|\leqslant\sigma(t_k)\\ T_{\min}\mathrm{e}^{-z_0(t_k)}\left(1-(z(t_k)-z_0(t_k))+\dfrac{(z(t_k)-z_0(t_k))^2}{2}\right)\leqslant\sigma(t_k)\\ \sigma(t_k)\leqslant T_{\max}\mathrm{e}^{-z_0(t_k)}(1-(z(t_k)-z_0(t_k)))\\ [0\quad 1\quad 0]\boldsymbol{r}(t_k)\geqslant 0,\ [0\quad 1\quad 0]\boldsymbol{u}(t_k)\geqslant\sigma(t_k)\cos\theta_{\max}\\ \ln m_{\mathrm{dry}}\leqslant z(t_{\mathrm{f}})\\ \ln(m_{\mathrm{wet}}-\alpha T_{\max}t_k)\leqslant z(t_k)\leqslant\ln(m_{\mathrm{wet}}-\alpha T_{\min}t)\\ \left|\sigma(t_{k+1})\mathrm{e}^{z\left(t_{k+1}^{i-1}\right)}-\sigma(t_k)\mathrm{e}^{z\left(t_k^{i-1}\right)}\right|\leqslant\Delta T_{\max}\end{cases} \tag{23}$$

式中：i为凸优化迭代次数第i次迭代；k为离散化求解时的第k个离散点。

基于4次多项式和凸优化的求解方案步骤流程如表1所示。

表1 凸优化方案流程表

Table 1 Convex optimization scheme flow table

步骤	内容
STEP1	获取着陆段初始位置和速度信息
STEP2	四次多项式制导求解初始轨迹$\boldsymbol{x}_k^0(k=1,2,\cdots,N)$，并给定终端时间$t_{\mathrm{f}}$
STEP3	根据前一次的状态量$\boldsymbol{x}_k^{i-1}$凸优化求解式(23)得到控制量$\boldsymbol{u}_k^i$
STEP4	控制量$\boldsymbol{u}_k^i$代入动力学模型式(1)积分得到真实状态量$\boldsymbol{x}_k^i$，$i=i+1$
STEP5	If $\boldsymbol{x}_k^i$不满足终止约束条件 跳到STEP3 else $\boldsymbol{x}_k^i$满足终止约束条件 $\boldsymbol{x}_k^i,\boldsymbol{u}_k^i$即为可行解

凸优化的初始轨迹对于求解的成功与否非常关键，相比于常值凸优化，即对于给定的初始和终端条件直接以线性化的方式计算出初始轨迹，使用四次多项式制导计算的初始轨迹更接近于最后的最优解。并且以燃料最优为目标的火箭软着陆问题是一个终端时间自由的最优控制问题，对于一般的凸优化算法来说，需要在凸优化程序的外面建立循环遍历找到时间段内性能最优、最省燃料的终端时间，这样就会导致程序的复杂性、实时性下降，不能在实际应用中发挥作用。但是本文的凸优化算法与四次多项式制导方法结合，在四次多项式制导中设定约束条件，达到约束条件后即停止求解，记录此时的终端时间，简化算法的复杂度，更高效地计算出最优轨迹，实现在线轨迹优化。

3　在线轨迹优化算例仿真

为了验证本文所设计算法的有效性, 以 Falcon9 火箭为例, SpaceX 公司已经成功使用 Falcon9 火箭在海上回收着陆平台成功着陆。垂直着陆段的结构参数均选择 Falcon9 火箭一子级的结构参数, 推力的数值为一台 Merlin-1D 发动机的数值, 并根据 SpaceX 公司公布的火箭数据得到大致的 Falcon9 火箭垂直着陆段的位置、速度等数据, 本次仿真参数如表 2 所示[19]。本文数值仿真计算机采用 Intel Core i7-9700 3.00 GHz CPU, 基于 MATLAB2020b 环境进行仿真验证, 并使用 MATLAB 的 CVX 凸优化工具箱, 选择了工具箱中的 SeDuMi 求解器进行求解。

表 2　Falcon 火箭一子级部分参数

Table 2　Partial parameters of a sub-stage of Falcon rocket

火箭参数	取值	火箭参数	取值
$m_{\text{fuel}}/\text{kg}$	12 230	$\boldsymbol{r}_0/\text{km}$	$[3\ 4.5\ 2.6]^{\text{T}}$
m_{dry}/kg	22 200	$\boldsymbol{v}_0/(\text{m}\cdot\text{s}^{-1})$	$[-150\ -320\ -260]^{\text{T}}$
m_{wet}/kg	34 530	$S_{\text{ref}}/\text{m}^2$	10
$g_0/(\text{m}\cdot\text{s}^{-2})$	9.8	C_{D}	2.2
$T_{\max}/\text{kN}$	845	$I_{\text{sp}}/(\text{m}\cdot\text{s}^{-1})$	282
$T_{\min}/\text{kN}$	0	$\rho_0/(\text{kg}\cdot\text{m}^{-3})$	1.225
$\Delta T_{\max}/\text{kN}$	100	β/m^{-1}	$1.378\ 5\times10^{-4}$

3.1　蒙特卡罗打靶仿真验证

垂直着陆段是整个返回过程中的最后一个阶段, 垂直着陆段的初始条件受到复杂的大气环境的影响, 每次的飞行过程中各不相同。其前一阶段的气动减速段的终端位置也有一定的误差范围, 因此在最后这个阶段的制导算法需要对不同的初始条件均具有适应能力, 分别对质量、比冲、标准密度、初始位置和初始速度进行偏差, 设定的初始条件如表 3 所示, 所设定的初始质量为正态分布, 其余均为均匀分布。

表 3　蒙特卡罗打靶初始条件设定范围

Table 3　Monte Carlo targeting initial condition setting range

偏差项	初始条件范围	偏差项	初始条件范围
$\boldsymbol{r}_x/\text{km}$	[2,4]	m_{wet}/kg	均值34 530 标准差 500
$\boldsymbol{r}_y/\text{km}$	[3.5,5.5]	$I_{\text{sp}}/(\text{m}\cdot\text{s}^{-1})$	[270,290]
$\boldsymbol{r}_z/\text{km}$	[1,3]	$\rho_0/(\text{kg}\cdot\text{m}^{-3})$	[1,1.5]
$\boldsymbol{v}_x/(\text{m}\cdot\text{s}^{-1})$	[−200,−100]	$\boldsymbol{v}_z/(\text{m}\cdot\text{s}^{-1})$	[−300,−200]
$\boldsymbol{v}_y/(\text{m}\cdot\text{s}^{-1})$	[−350,−250]		

基于表 3 的偏差条件得到的两种算法的蒙特卡罗打靶的仿真结果如图 1～图 4 所示, 一共打靶 300 次, 并统计着陆终端位置、速度和消耗燃料的

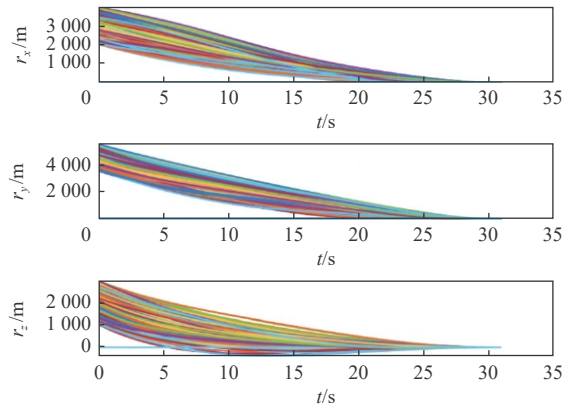


图 1　使用常值凸优化的火箭位置蒙特卡罗打靶曲线

Fig. 1　Rocket position Monte-Carlo hitting curve with constant convex optimization

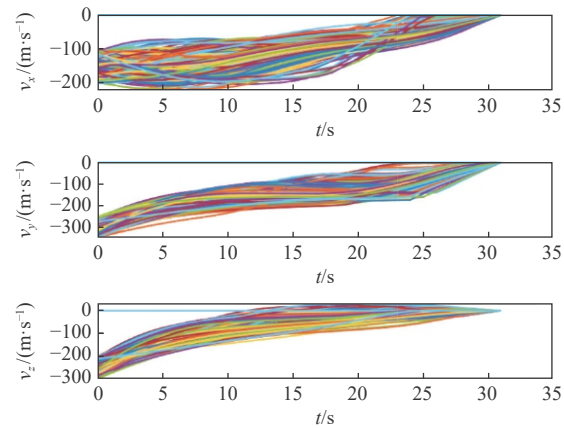


图 2　使用常值凸优化的火箭速度蒙特卡罗打靶曲线

Fig. 2　Rocket velocity Monte-Carlo hitting curve with constant convex optimization

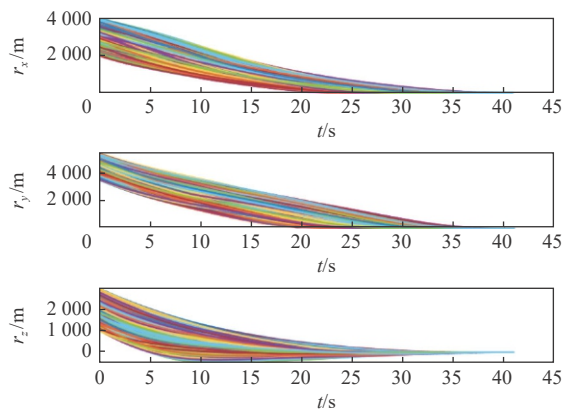


图 3　使用本文算法的火箭位置蒙特卡罗打靶曲线

Fig. 3　Rocket position Monte-Carlo hitting curve utilizing the algorithm in this paper

均值、标准差和极值。常值凸优化和本文算法的蒙特卡罗打靶统计曲线和统计结果如图 5 和图 6 所示。

本文在不同偏差下进行了蒙特卡罗打靶, 一共打靶 300 次, 图 5 和图 6 记录了火箭成功软着陆后

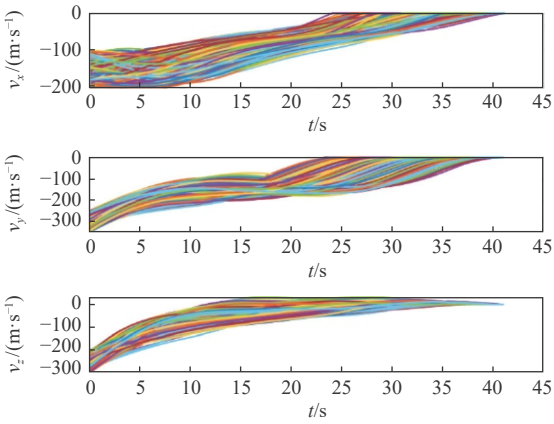


图4 使用本文算法的火箭速度蒙特卡罗打靶曲线

Fig. 4 Rocket velocity Monte-Carlo hitting curve utilizing the algorithm in this paper

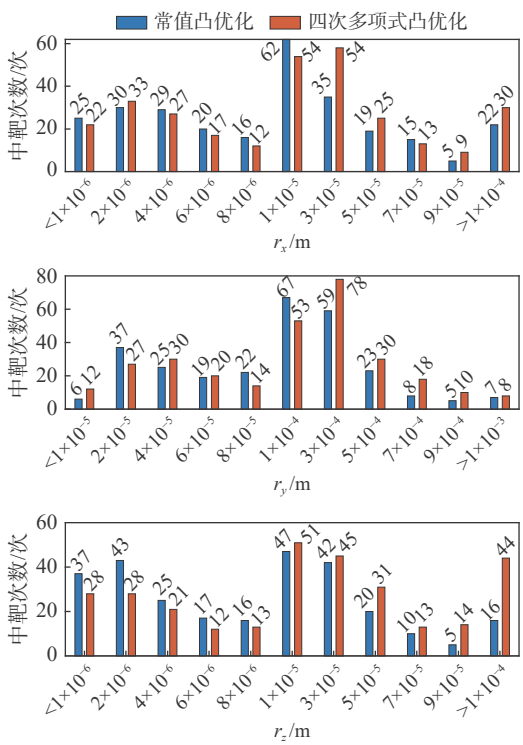


图5 火箭终端位置蒙特卡罗打靶结果统计对比

Fig. 5 Statistical comparison of rocket position Monte-Carlo hitting results

的终端位置和速度的分布图。根据图5和图6的仿真结果统计，常值凸优化方法成功软着陆278次，本文算法成功软着陆300次。因此常值凸优化方法共出现22次不能求解的情况，而本文算法没有出现不可求解的情况，并且两种方法得到的终端位置和速度的分布十分接近。从图5、图6和表4的结果可知，四次多项式制导与凸优化的结合与常值凸优化相比能够在不影响精度的同时，提高着陆的成功率，可靠性更高。本文所提出的算法可以基于不同的初始条件求解不同的初始轨迹和终端着陆时间，使得凸优化能够收敛。常值凸优化使用线性化的方式根据初始条件求解出初始轨迹，但是每次的终端着陆时间是设定好的，不能根据初始条件的不同而改变，因此在求解的成功率上会比本文算法差距大。只要垂直着陆段开始时剩余的燃料足够，火箭发动机比冲和大气标准密度偏差合理，气动减速段末端的位置和速度在一定范围内，使用该算法在线制导得到的着陆终端位置和速度能够满足软着陆的要求，并且飞行过程中也能够满足约束。因此验证了在不同的初始状态下该算法能够保证火箭子级垂直软着陆于海上回收着陆平台。

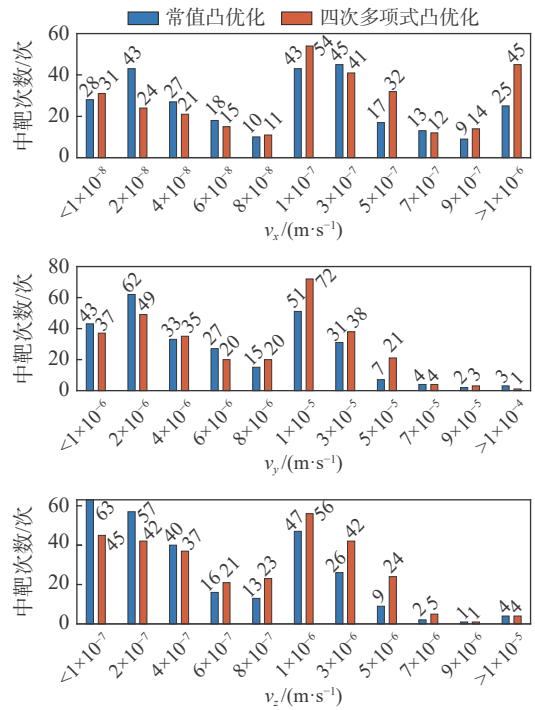


图6 火箭终端速度蒙特卡罗打靶结果统计对比

Fig. 6 Statistical comparison of rocket terminal velocity Monte-Carlo hitting results

3.2 给定终端时间下的凸优化求解

将四次多项式制导求解停止条件设定高度在

表4 火箭子级垂直着陆段蒙特卡罗打靶结果统计

Table 4 Statistical table of Monte Carlo shooting results of rocket substage vertical landing section

参数	$r_x(t_f)$/m	$r_y(t_f)$/m	$r_z(t_f)$/m	$v_x(t_f)$/(m·s^{-1})	$v_y(t_f)$/(m·s^{-1})	$v_z(t_f)$/(m·s^{-1})	Δm/kg
均值	3.42×10^{-5}	2.70×10^{-4}	4.87×10^{-5}	2.42×10^{-7}	1.39×10^{-5}	1.53×10^{-6}	4 696.2
标准差	5.93×10^{-5}	3×10^{-4}	8.82×10^{-5}	1.18×10^{-6}	1.76×10^{-5}	2.13×10^{-6}	667.01
极值	4.12×10^{-4}	0.002 7	0.001	1.14×10^{-5}	1.33×10^{-4}	1.81×10^{-5}	6 294.7

小于 0.000 1 m 后停止, 停止时求解的终端着陆时间为 31.18 s, 因此设定本文凸优化和常值凸优化终端着陆时间均设定为 31 s。本文所提出的凸优化算法与常值凸优化方法在给定终端时间下对比的结果如表 5 所示。将得到的结果与常值凸优化方法进行对比, 如图 7～图 11 所示。

表 5　四次多项式凸优化与常值凸优化对比

Table 5　Quadratic polynomial convex optimization vs. constant convex optimization

参数	四次多项式凸优化	常值凸优化
$r_x(t_f)$/m	1.27×10^{-5}	2.98×10^{-5}
$r_y(t_f)$/m	-9.75×10^{-5}	-1.85×10^{-4}
$r_z(t_f)$/m	-4.76×10^{-6}	-1.18×10^{-5}
$v_x(t_f)/(\mathrm{m\cdot s^{-1}})$	1.71×10^{-7}	8.31×10^{-7}
$v_y(t_f)/(\mathrm{m\cdot s^{-1}})$	-3.93×10^{-6}	-4.89×10^{-6}
$v_z(t_f)/(\mathrm{m\cdot s^{-1}})$	-3.09×10^{-7}	-5.69×10^{-7}
$m(t_f)$/kg	30 104.06	30 047.02
求解时间/s	12.63	16.38

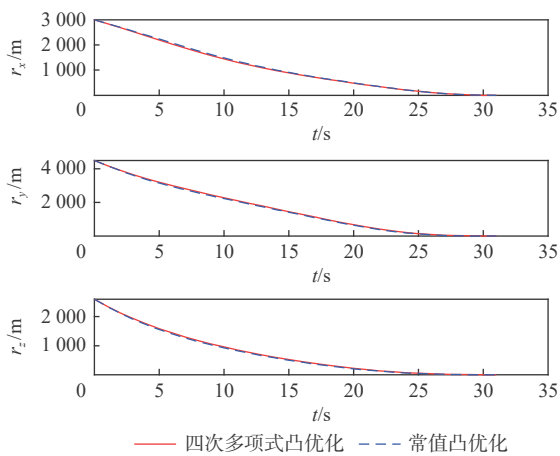


图 7　火箭一子级三轴位置曲线

Fig. 7　The three-axis position curve of the rocket

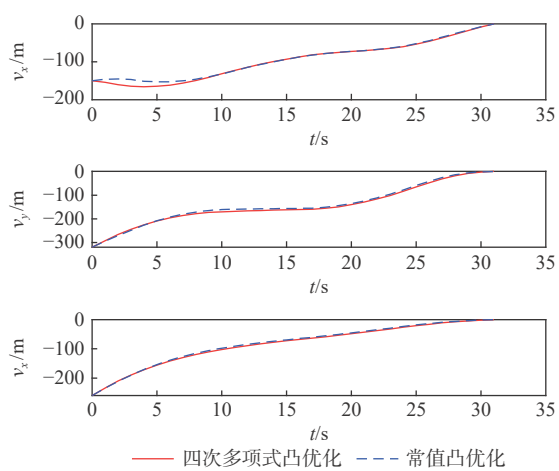


图 8　火箭一子级三轴速度曲线

Fig. 8　The three-axis velocity curve of the first sub-stage of the rocket

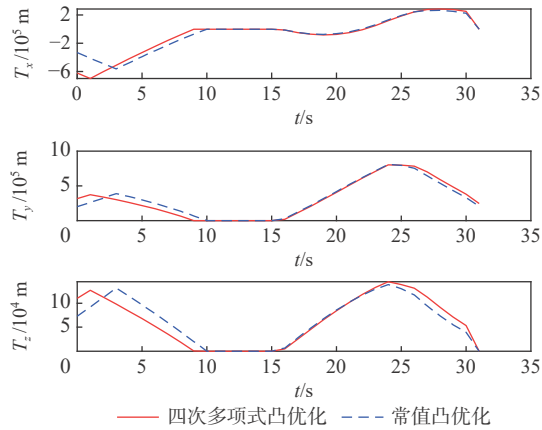


图 9　火箭一子级三轴推力曲线

Fig. 9　Three-axis thrust curve of a rocket substage

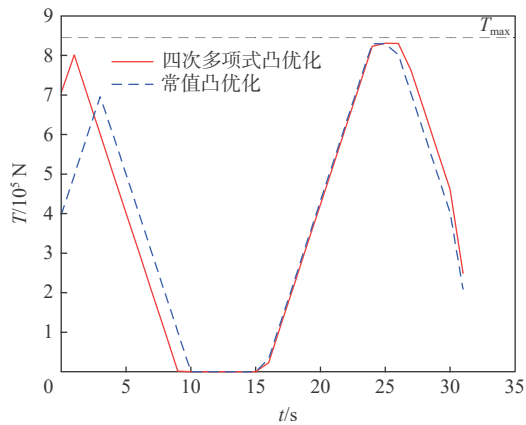


图 10　火箭一子级总推力曲线

Fig. 10　Total thrust curve of a rocket substage

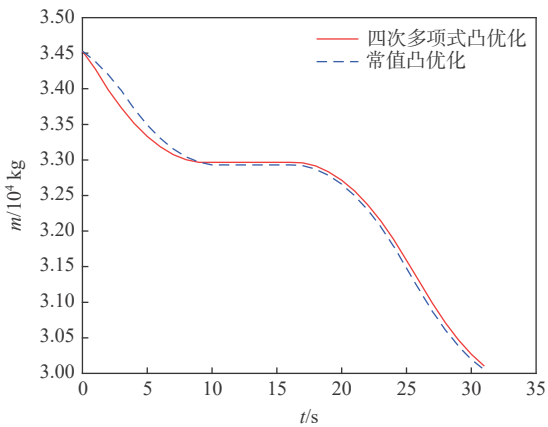


图 11　火箭一子级质量变化曲线

Fig. 11　The mass change curve of the first sub-stage of the rocket

所设计的算法和常值凸优化方法均能够计算制导律实现最终的着陆, 并满足着陆时位置、速度的约束条件, 并且推力矢量在着陆时仅有垂直方向存在推力, 水平横向和侧向均为 0, 过程中也满足推力限幅。但是四次多项式制导与凸优化的结合相比于常值凸优化来说在节省燃料和求解时间方面

具有优势，节省了约57 kg的燃料，并且本文所设计的凸优化算法在和常值凸优化方法达到相同精度时，整体运行的求解时间为12.63 s，而常值凸优化方法的求解时间为16.38 s。此时间还不包括外部最优时间的寻优，从求解的效率来看，本文所设计的凸优化算法的实时性更好，并且常值凸优化方法需要在凸优化外建立循环，找到最优的终端着陆时间，如果把这部分时间加上，本文算法将远远快于常值凸优化方法。因此对于实际情况来说，本文所提出的算法有更好的应用效果。

4 结 论

本文提出了一种可重复使用运载火箭垂直着陆段的轨迹在线优化方法，该方法保证了火箭能够在满足各种约束条件下精确软着陆于预定平台，并保证了问题求解的效率。

1）火箭一子级垂直着陆段使用四次多项式制导法，该算法与凸优化均以燃料最省为目标，并能够考虑终端姿态约束，得出的初始轨迹更符合最后的实际轨迹，更容易使得凸优化算法收敛，并且通过该制导方法能够计算出终端着陆时间，与凸优化结合降低了算法的复杂度，提高了效率。

2）通过蒙特卡罗打靶统计了本文算法和常值凸优化方法在不同初始状态偏差下的打靶结果，对比了两者的着陆成功次数，并统计了本文算法所得到的最终着陆时火箭位置、速度和质量的均值、标准差和极值，定量验证了所提算法在垂直着陆段初始条件不确定的不同偏差下均能够实现软着陆，相比于常值凸优化方法提升了着陆成功率，可靠性较高。

3）使用四次多项式制导计算出初始轨迹代入凸优化算法与常值凸优化方法相比，在达到相同着陆的精度时，所设计的算法在燃料消耗和求解时间上相比常值凸优化方法有所减少，此求解时间并不包括常值凸优化方法着陆时间寻优的部分。

综上所述，本文提供的算法能够适用于火箭子级垂直着陆段的在线轨迹优化，实现软着陆。

参考文献（References）

[1] BONO P. Rombus-an integrated systems concept for a reusable orbital module/booster and utility shuttle [C]//Summer Meeting. 1963: 271.

[2] 吴荣. 垂直起降可重复使用火箭返回制导与控制方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2019.
WU R. Research on guidance and control method of vertical takeoff and landing reusable rocket return[D]. Harbin: Harbin Institute of Technology, 2019(in Chinese).

[3] 宋征宇, 蔡巧言, 韩鹏鑫, 等. 重复使用运载器制导与控制技术综述[J]. 航空学报, 2021, 42(11): 525050.
SONG Z Y, CAI Q Y, HAN P X, et al. Review of guidance and control technologies for reusable launch vehicles[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(11): 525050(in Chinese).

[4] KLUMPP A R. Apollo lunar descent guidance[J]. Automatica, 1974, 10(2): 133-146.

[5] 张洪华, 关轶峰, 黄翔宇, 等. 嫦娥三号着陆器动力下降的制导导航与控制[J]. 中国科学: 技术科学, 2014, 44(4): 377-384.
ZHANG H H, GUAN Y F, HUANG X Y, et al. Guidance navigation and control for Chang'E-3 powered descent[J]. Scientia Sinica (Technologica), 2014, 44(4): 377-384(in Chinese).

[6] 韦常柱, 琚啸哲, 徐大富, 等. 垂直起降重复使用运载器返回制导与控制[J]. 航空学报, 2019, 40(7): 322782.
WEI C Z, JU X Z, XU D F, et al. Guidance and control for return process of vertical takeoff vertical landing reusable launching vehicle[J]. Acta Aeronautica et Astronautica Sinica, 2019, 40(7): 322782(in Chinese).

[7] 张康. 火箭子级垂直回收轨迹优化与制导研究[D]. 长沙: 国防科技大学, 2018.
ZHANG K. Research on optimization and guidance of rocket substage vertical recovery trajectory[D]. Changsha: National University of Defense Technology, 2018(in Chinese).

[8] ACIKMESE B, PLOEN S R. Convex programming approach to powered descent guidance for Mars landing[J]. Journal of Guidance, Control, and Dynamics, 2007, 30(5): 1353-1366.

[9] SCHARF D P, REGEHR M W, VAUGHAN G M, et al. ADAPT demonstrations of onboard large-divert Guidance with a VTVL rocket[C]//2014 IEEE Aerospace Conference. Piscataway: IEEE Press, 2014: 1-18.

[10] 张志国, 马英, 耿光有, 等. 火箭垂直回收着陆段在线制导凸优化方法[J]. 弹道学报, 2017, 29(1): 9-16.
ZHANG Z G, MA Y, GENG G Y, et al. Convex optimization method used in the landing-phase on-line guidance of rocket vertical recovery[J]. Journal of Ballistics, 2017, 29(1): 9-16(in Chinese).

[11] WANG J B, CUI N G, WEI C Z. Optimal rocket landing guidance using convex optimization and model predictive control[J]. Journal of Guidance, Control, and Dynamics, 2019, 42(5): 1078-1092.

[12] 吴杰, 张成, 李淼, 等. 基于凸优化和LQR的火箭返回轨迹跟踪制导[J]. 北京航空航天大学学报, 2022, 48(11): 2270-2280.
WU J, ZHANG C, LI M, et al. Rocket return trajectory tracking guidance based on convex optimization and LQR[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2270-2280(in Chinese).

[13] 安泽, 熊芬芬, 梁卓楠. 基于偏置比例导引与凸优化的火箭垂直着陆制导[J]. 航空学报, 2020, 41(5): 323606.
AN Z, XIONG F F, LIANG Z N. Landing-phase guidance of rocket using bias proportional guidance and convex optimization[J]. Acta Aeronautica et Astronautica Sinica, 2020, 41(5): 323606(in Chinese).

[14] SAGLIANO M, HEIDECKER A, MACÉS HERNÁNDEZ J, et al. Onboard guidance for reusable rockets: aerodynamic descent and powered landing[C]//AIAA Scitech 2021 Forum. Reston: AIAA, 2021: 0862.

[15] LIU X F. Fuel-optimal rocket landing with aerodynamic controls[J]. Journal of Guidance, Control, and Dynamics, 2019, 42(1): 65-77.

[16] WANG J B, LI H X, CHEN H B. An iterative convex programming method for rocket landing trajectory optimization[J]. The Journal of the Astronautical Sciences, 2020, 67(4): 1553-1574.

[17] MA L, WANG K X, XU Z H, et al. Optimal feedback guidance for powered landing of reusable rockets[C]//2019 Chinese Control and Decision Conference (CCDC). Piscataway: IEEE, 2019: 915-920.

[18] 何林坤, 张冉, 龚庆海. 基于强化学习的可回收运载火箭着陆制导[J]. 空天防御, 2021, 4(3): 33-40.
HE L K, ZHANG R, GONG Q H. Landing guidance of reusable launch vehicle based on reinforcement learning[J]. Air & Space Defense, 2021, 4(3): 33-40(in Chinese).

[19] 郭超. 基于凸优化的火箭垂直返回精确着陆轨迹在线优化[D]. 哈尔滨: 哈尔滨工业大学, 2020.
GUO C. On-line optimization of rocket vertical return accurate landing trajectory based on convex optimization[D]. Harbin: Harbin Institute of Technology, 2020(in Chinese).

[20] BLACKMORE L, ACIKMESE B, SCHARF D P. Minimum-landing-error powered-descent guidance for Mars landing using convex optimization[J]. Journal of Guidance, Control, and Dynamics, 2010, 33(4): 1161-1171.

Online trajectory optimization method for vertical landing phase of reusable launch vehicle

SHI Qi[1, 2], QI Ruiyun[1, 2, *], SHE Yuchen[3], HU Cunming[3]

(1. State Key Laboratory of Mechanics and Control for Aerospace Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;
2. School of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
3. Shanghai Institute of Space Control Technology, Shanghai 201109, China)

Abstract: In order to overcome the difficulties of ambiguous initial circumstances, intricate process constraints, and strict terminal restrictions in the guidance of reusable launch vehicles, this work suggests an online trajectory optimization technique that combines convex optimization and fourth-order polynomial guidance. The method utilizes a fourth-order polynomial guidance approach considering terminal attitude constraints and fuel optimality as the objective to compute the initial trajectory and terminal landing time. By considering different initial conditions, different trajectories and times are obtained to ensure a successful solution of convex optimization under varying circumstances. Additionally, by applying the initial trajectory to a convex optimization algorithm with the same fuel optimality objective, guidance commands are generated to satisfy process constraints and terminal constraints, enabling a soft landing. Simulation analysis demonstrates that the proposed method exhibits better adaptability and higher reliability to different deviations compared to constant convex optimization algorithms. Furthermore, by doing away with the need to optimize terminal landing time, the suggested approach increases solution efficiency and lowers fuel consumption.

Keywords: reusable launch vehicle; vertical landing section; convex optimization; trajectory optimization; soft landing

Received: 2023-05-30; **Accepted**: 2023-08-11; **Published Online**: 2023-09-08 14:42

URL: link.cnki.net/urlid/11.2625.V.20230907.0942.004

Foundation items: National Natural Science Foundation of China (62373189); National Key Laboratory of Aerospace Structural Mechanics and Control Open Project Funding

* **Corresponding author**. E-mail: ruiyun.qi@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0262

双径向洛伦兹力磁轴承动力学及解耦控制方法

李宗育[1]，王卫杰[2,*]，王丽芬[2]，任元[2]，樊亚洪[3]，李磊[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)

摘　　要：为实现高精度磁浮转台的快速响应和振动抑制，针对洛伦兹力磁悬浮万向稳定平台的高精度解耦控制问题，首先阐明径向洛伦兹力磁轴承(RLFMB)工作原理，基于等效磁路法对单个径向洛伦兹力磁轴承进行动力学建模，构建转子四自由度平转运动动力学模型。设计一种基于内模结构的解耦控制器，通过补灵敏度函数Bode图验证内模控制器强鲁棒的特点。仿真算例结果表明，跟踪性能上，偏转、平动的响应时间分别较PID控制方法降低59.3%和28.2%；抗扰性能上，偏转、平动干扰残余量较PID控制方法降低38.8%和86.2%，此方法可应用于平台对载荷系统的高精、高稳、高动态指向控制。

关 键 词：洛伦兹力磁悬浮万向稳定平台；径向洛伦兹力磁轴承；解耦控制；补灵敏度函数；内模控制

中图分类号：V448.2

文献标志码：A　　**文章编号**：1001-5965(2025)05-1770-11

随着航天技术的快速发展，航天应用对卫星载荷的高精度定向和快速机动提出了新的要求[1]，传统的机械磁轴承存在机械摩擦和机械振动[2-4]，影响轴承输出力矩和使用寿命[5-6]。磁悬浮平台具有非接触、微振动、长寿命的优势，相较于传统采用隔振机构的航天器平台，采用磁悬浮技术实现航天器平台对载荷扰动的主动振动控制和振动抑制[7]可以满足卫星超高敏捷、超高稳定、超高精度的性能要求，同时抑制微振动对航天器指向精度、指向稳定度以及敏捷性能的影响。因此，磁悬浮轴承是研究设计“三超”航天器平台的一项热点技术。

磁悬浮轴承依据受力机理可分为磁阻力和洛伦兹力磁轴承(Lorentz force magnetic bearing, LFMB)。磁阻力磁轴承具有较大的承载能力，定位精度较高，但其电磁力与控制电流呈平方关系，且存在位移负刚度，控制系统复杂[8-9]。与磁阻力磁轴承相比，LFMB线性度好，控制带宽大，能够快速实现稳定悬浮，在微重力环境下是一种理想的悬浮支撑方式，其为载荷提供的控制力矩与电流成正比，无位移负刚度，具有更高的控制精度[10-16]。LFMB在应用方面分为部件级应用和平台级应用。

部件级上主要用于飞轮、CMG等航天器姿控执行机构。2005年德国Teldix公司设计的磁悬浮框架飞轮(magnetically suspended gimballing flywheel, MSGFW)样机，采用洛伦兹力磁轴承实现转子五自由度主动悬浮以及偏转控制[17]。北京航空航天大学提出的MSGFW工程样机可以实现姿态控制和敏感一体化[10]。2017年哈尔滨工业大学提出的径向被动轴向主动的混合式磁轴承[18]，可以实现五自由度磁轴承系统的稳定悬浮。北京控制工程研究所设计的径向洛伦兹力磁轴承[19]，在磁间隙内定转子发生一定偏移时仍能实现高性能的稳定悬浮。

收稿日期：2023-05-22；**录用日期**：2023-06-30；**网络出版时间**：2023-09-01 14:33
网络出版地址：link.cnki.net/urlid/11.2625.V.20230901.1108.002
基金项目：国家自然科学基金(52075545)
*通信作者. E-mail：wangweijie@126.com

引用格式：李宗育，王卫杰，王丽芬，等. 双径向洛伦兹力磁轴承动力学及解耦控制方法[J]. 北京航空航天大学学报，2025，51(5)：1770-1780. LI Z Y，WANG W J，WANG L F，et al. Dynamics and decoupling control method of double radial Lorentz force magnetic bearings[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1770-1780 (in Chinese).

2022 年北京航空航天大学 Xiang 等提出的姿态控制磁悬浮飞轮[20], 用锥形磁轴承和洛伦兹力磁轴承输出陀螺力矩控制转子三自由度平动和两自由度偏转。

目前的平台级应用主要有在轨整星隔振平台、磁悬浮吊舱等。国际空间站机柜上应用的隔振系统 MVIS(microgravity vibration isolation system), 利用 8 个磁悬浮作动器分散布局, 实现六自由度主动隔振控制。2014 年上海卫星工程研究所(509 所)提出的“双超”卫星平台[21]具有超高指向精度、超高稳定度的技术要求, 通过非接触磁浮[22]和柔性线缆连接机构[23]实现动静隔离、主从协同控制, 该平台首次成功应用于 2021 年发射的“羲和号”太阳探测试验卫星。北京控制工程研究所(502 所)王有懿等于 2016 年提出的超静隔振平台[24], 采用 6 个独立的作动器控制 6 个支杆实现平台的主动隔振, 2022 年基于“三超”卫星平台的北京三号 B 卫星进入预定轨道。进一步, 北京控制工程研究所设计一种洛伦兹力磁悬浮旋转关节[25]用来连接平台舱和载荷舱, 实现载荷舱在轨六自由度运动。509 所设计的卫星平台没有对指向敏捷性进行约束, 502 所采用相互独立的作动器实现整星平台超静隔振。

航天工程大学研究团队提出的洛伦兹力磁悬浮万向稳定平台采用全洛伦兹力磁悬浮支承载荷系统, 其中一对双径向洛伦兹力磁轴承 (radial Lorentz force magnetic bearing, RLFMB) 为稳定平台提供两自由度平动和两自由度旋转机动。在磁悬浮平台姿态控制过程中, 控制系统受到噪声、负载扰动等影响, 采用常规的 PID 控制器难以获得满意的控制效果[26]。

现有的高精度转台控制算法包括基于状态观测器的补偿控制[27]、滑膜变结构控制[28]、模糊控制[29]等。基于状态观测器的补偿控制算法虽然对外界扰动有较好的抑制作用, 但是观测器存在滞后, 会降低系统的响应速度。滑膜变结构控制算法响应迅速, 但是在结构变换过程中可能引起系统抖振。模糊控制不需要控制对象的精确数学模型, 但是在解模糊后会存在较明显的静差。因此这些控制算法无法同时满足高精度转台快速响应和振动抑制的要求。

针对上述问题, 本文以万向稳定平台为研究对象, 阐明径向洛伦兹力磁轴承工作原理, 建立单个 RLFMB 动力学模型, 对四自由度平转运动耦合特性进行分析, 采用内模控制方法[30]设计四自由度平转运动解耦控制器, 最后通过仿真算例分析双 RLFMB 控制系统的跟踪性能和抗扰性能。

1　RLFMB 原理分析与动力学建模

1.1　RLFMB 工作原理分析

整个洛伦兹力磁悬浮万向稳定平台结构主要由平台舱、框架舱、载荷舱、七自由度磁悬浮机构构成。七自由度磁悬浮机构包括一对轴向洛伦兹力磁轴承、一对径向洛伦兹力磁轴承和两组旋转洛伦兹力磁轴承。框架舱中的一对径向洛伦兹力磁轴承对称安装在转轴两侧, 可以实现 y 轴方向和 x 轴方向两自由度平动, 以及绕 y 轴方向和绕 x 轴方向两自由度旋转机动, 径向磁轴承主动控制载荷仓绕 x 轴旋转如图 1 所示。

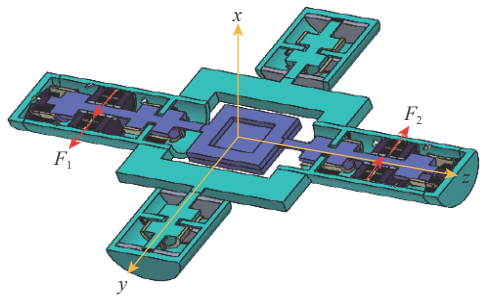


图 1　磁轴承主动控制绕 X 轴旋转示意图

Fig. 1　Active control rotation of magnetic bearing around X-axis

考虑到 LFMB 线性度好、出力精度高、控制带宽大, 但是出力较小、承载能力较弱等特点, 同时为降低磁轴承转子质量, 方便进行地面承载能力验证, 设计磁轴承磁钢部分作为定子, 线圈部分作为转子。

单个 RLFMB 主要包括上定子组件、下定子组件和转子组件。如图 2 所示, 其中, 上定子组件和下定子组件均包括导磁环、内磁钢环、外磁钢环; 转子组件包括线圈骨架、瓦形线圈。上定子组件中, 内磁钢环和外磁钢环分内外放置在导磁环下面; 下定子组件中, 内磁钢环和外磁钢环分内外放置在导磁环上面; 所述内磁钢环、外磁钢环轴向充磁, 且内磁钢环和外磁钢环充磁方向相反, 上定子组件与下定子组件的内磁钢环充磁方向相同, 上定子组件与下定子组件的外磁钢环充磁方向相同。

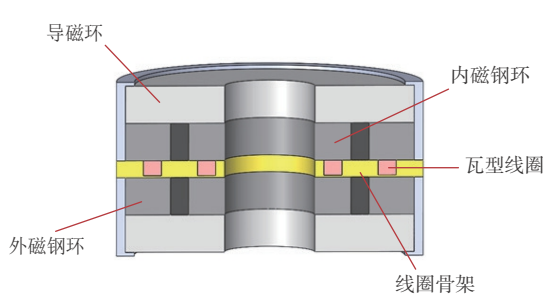
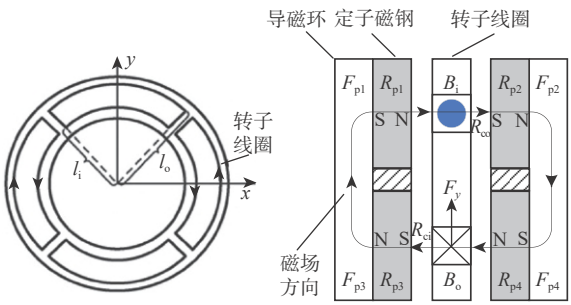


图 2　单个 RLFMB 截面结构示意图

Fig. 2　Sectional structure of single RLFMB

1.2 基于等效磁路法的单RLFMB动力学建模

忽略磁钢漏磁和线圈自感磁场以及边缘效应的影响，根据磁轴承的轴截面结构建立单个RLFMB磁路图，如图3所示，图中磁路包括4组磁钢环的磁动势$F_{p1}\sim F_{p4}$，磁动势对应的永磁磁阻$R_{p1}\sim R_{p4}$，穿过内外磁钢气隙的气隙磁阻R_{ci}和R_{co}，永磁磁路可以在气隙中产生沿轴向的恒定均匀垂直磁场B，其中磁钢环的磁动势与磁化长度线性相关，所有磁阻均为串行连接，建立的等效磁路如图4所示。



(a) 瓦型线圈示意图 (b) 磁轴承截面磁路图

图3 RLFMB磁路图

Fig. 3 Magnetic circuit of RLFMB

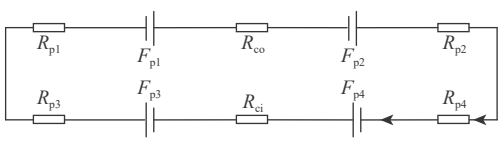


图4 等效磁路图

Fig. 4 Equivalent magnetic circuit

根据磁路欧姆定律可以得到线圈主磁通φ_m和磁感应强度B的数学表达式为

$$\phi_m=\frac{\sum_{i=1}^{4}F_{pi}}{\sum_{i=1}^{4}R_{pi}+R_{co}+R_{ci}} \tag{1}$$

$$B=\begin{bmatrix}B_i & B_o\end{bmatrix}^{\mathrm{T}}=\left[\mathrm{diag}\{A_i,A_o\}\right]^{-1}\begin{bmatrix}\phi_m & \phi_m\end{bmatrix}^{\mathrm{T}} \tag{2}$$

式中：$F_{pi}=H_c l_{pi}$；F_{pi}为磁钢的磁动势；H_c为磁钢的矫顽力；l_{pi}为磁钢的磁化长度；B_i和B_o分别为内、外磁钢气隙的磁通密度；A_i和A_o分别为内、外磁通密度对应的等效截面积。

转子组件中线圈部分由两组垂直分布的瓦型线圈构成，在恒定均匀磁场作用下产生y、x两通道的洛伦兹力。分析单个瓦型线圈受力时，因磁场与线圈平面垂直，对线圈受洛伦兹力在y、x方向进行投影并积分，可以得到单个瓦型线圈受力如下式所示，长度为dL的电流元所受电磁力为$BIl\mathrm{d}\theta$：

$$\begin{bmatrix}F_y\\F_x\end{bmatrix}=\begin{bmatrix}2\int_0^{\frac{\varphi}{2}}B_oIl_o\cos\left(\frac{\varphi}{2}-\theta\right)\mathrm{d}\theta+2\int_0^{\frac{\varphi}{2}}B_iIl_i\cos\left(\frac{\varphi}{2}-\theta\right)\mathrm{d}\theta\\2\int_0^{\frac{\varphi}{2}}B_oIl_o\sin\left(\frac{\varphi}{2}-\theta\right)\mathrm{d}\theta-2\int_0^{\frac{\varphi}{2}}B_iIl_i\sin\left(\frac{\varphi}{2}-\theta\right)\mathrm{d}\theta\end{bmatrix}=\begin{bmatrix}2B_oIl_o\sin\frac{\varphi}{2}+2B_iIl_i\sin\frac{\varphi}{2}\\0\end{bmatrix} \tag{3}$$

式中：φ为瓦型线圈张角；l_o和l_i分别为瓦型线圈长弧段和短弧段到线圈中心的距离。

根据式(3)可知，4组瓦型线圈在Y、X通道中出力不存在耦合，可以对两个方向受力进行单独分析，当没有产生干扰力矩时，单个径向磁轴承受到洛伦兹力为

$$F_1=NBIL=2NIB_oL_M+2NIB_iL_m \tag{4}$$

式中：L_M为瓦型线圈长弧段在环形磁场中的有效磁感应长度；L_m为瓦型线圈短弧段在环形磁场中的有效磁感应长度；L_M和L_m可以表示为

$$\begin{bmatrix}L_M\\L_m\end{bmatrix}=2\sin\frac{\varphi}{2}\begin{bmatrix}l_o\\l_i\end{bmatrix} \tag{5}$$

将式(2)与式(4)联立可得单个RLFMB动力学模型：

$$\begin{bmatrix}F_y\\F_x\end{bmatrix}=2NI\frac{\sum_{i=1}^{4}F_{pi}}{\sum_{i=1}^{4}R_{pi}+R_{co}+R_{ci}}\begin{bmatrix}1\\1\end{bmatrix}\left(\frac{L_M}{A_o}+\frac{L_m}{A_i}\right) \tag{6}$$

式中：F_y和F_x分别为磁轴承在y通道和x通道上的出力。

在正常工作状态下，磁化长度、等效磁阻以及内外磁通对应的等效截面积基本保持不变，可以认为输出电磁力和线圈电流线性相关，因此式(6)等效为

$$\begin{bmatrix}F_y\\F_x\end{bmatrix}=K_iI\begin{bmatrix}1\\1\end{bmatrix} \tag{7}$$

式中：K_i为线圈电流刚度系数。

1.3 双RLFMB平转运动动力学建模

当磁轴承转子受到干扰时，转子两端两通道上4个位移传感器产生偏移为α_y、β_y、α_x、β_x，经过控制器产生的控制电流分别为I_y^{L}、I_y^{R}、I_x^{L}、I_x^{R}，转子平动的运动方程可以表示为

$$\begin{cases}m\ddot{r}_y^{\mathrm{L}}=K_iI_y^{\mathrm{L}}\\m\ddot{r}_y^{\mathrm{R}}=K_iI_y^{\mathrm{R}}\\m\ddot{r}_x^{\mathrm{L}}=K_iI_x^{\mathrm{L}}\\m\ddot{r}_x^{\mathrm{R}}=K_iI_x^{\mathrm{R}}\end{cases} \tag{8}$$

在只考虑转子平动运动时，每个通道两端磁轴承中瓦型线圈的控制电流等大同向，产生的洛伦兹力等大同向，转子平动动力学方程可以简化为

$$\begin{bmatrix} m\ddot{r}_y(t) \\ m\ddot{r}_x(t) \end{bmatrix} = 2K_{\mathrm{i}} \begin{bmatrix} I_y^{\mathrm{L}} \\ I_x^{\mathrm{L}} \end{bmatrix} \tag{9}$$

考虑转子转动运动时，每个通道两端磁轴承中瓦型线圈的控制电流等大反向，产生的洛伦兹力等大反向。当转子左端磁轴承出力为正，右端磁轴承出力为负时，转子进行顺时针转动。在转子实际转动过程中转动角度较小，转动弧段可以近似为位移传感器敏感偏移：

$$\begin{cases} \theta_y L \approx \sin\theta_y L = r_y^{\mathrm{L}} \\ \theta_x L \approx \sin\theta_x L = r_x^{\mathrm{L}} \end{cases} \tag{10}$$

式中：θ_y^{L}和θ_x^{L}分别为 y 和 x 通道上转子偏移角度，L 为磁轴承到转轴中心的距离，转子转动的运动方程可以表示为

$$\begin{cases} I_y^{\mathrm{L}} = I_y^{\mathrm{R}} \\ I_x^{\mathrm{L}} = I_x^{\mathrm{R}} \\ J_x\ddot{\theta}_y(t) = K_{\mathrm{i}} I_y^{\mathrm{L}} L - K_{\mathrm{i}} I_y^{\mathrm{R}} L \\ J_y\ddot{\theta}_x(t) = K_{\mathrm{i}} I_x^{\mathrm{R}} L - K_{\mathrm{i}} I_x^{\mathrm{L}} L \end{cases} \tag{11}$$

式中：J_x和J_y分别为磁悬浮转子在 y 轴和 x 轴的转动惯量，转子转动动力学方程可以简化为

$$\begin{bmatrix} J_x\ddot{\theta}_y(t) \\ J_y\ddot{\theta}_x(t) \end{bmatrix} = 2K_{\mathrm{i}}L \begin{bmatrix} I_y^{\mathrm{L}} \\ I_x^{\mathrm{R}} \end{bmatrix} \tag{12}$$

据上述转子平动和转动的受力分析，联合式 (8) 和式 (11) 可以得到磁轴承产生的控制力和控制力矩如下：

$$\begin{aligned} \begin{bmatrix} F_y \\ M_y \end{bmatrix} &= K_{\mathrm{i}} \begin{bmatrix} 1 & 1 \\ L & -L \end{bmatrix} \begin{bmatrix} I_y^{\mathrm{L}} \\ I_y^{\mathrm{R}} \end{bmatrix} = K_{\mathrm{i}} T_y \begin{bmatrix} I_y^{\mathrm{L}} \\ I_y^{\mathrm{R}} \end{bmatrix} \\ \begin{bmatrix} F_y \\ M_y \end{bmatrix} &= K_{\mathrm{i}} \begin{bmatrix} 1 & 1 \\ -L & L \end{bmatrix} \begin{bmatrix} I_x^{\mathrm{L}} \\ I_x^{\mathrm{R}} \end{bmatrix} = K_{\mathrm{i}} T_x \begin{bmatrix} I_x^{\mathrm{L}} \\ I_x^{\mathrm{R}} \end{bmatrix} \end{aligned} \tag{13}$$

根据式 (9) 和式 (12)，并考虑控制系统存在的时延τ，RLFMB 单通道平动和转动动力学方程组可以表示为

$$\begin{cases} m\ddot{r}_y(t) = 2K_{\mathrm{i}} I_y^{\mathrm{L}}(t-\tau) \\ J_x\ddot{\theta}_y(t) = 2K_{\mathrm{i}} L I_y^{\mathrm{L}}(t-\tau) \end{cases} \tag{14}$$

对式 (14) 进行拉普拉斯变换可以得到

$$\begin{cases} ms^2 r_y(s) = 2K_{\mathrm{i}} I(s)\mathrm{e}^{-\tau s} \\ J_x s^2 \theta_y(s) = 2K_{\mathrm{i}} L I(s)\ \mathrm{e}^{-\tau s} \end{cases} \tag{15}$$

式中：s 为时域到复频域的算子，可以得到平转运动状态方程为

$$\begin{bmatrix} r_y(s) \\ \theta_y(s) \end{bmatrix} = \begin{bmatrix} \dfrac{2K_{\mathrm{i}}}{ms^2} \\ \dfrac{2K_{\mathrm{i}}L}{J_x s^2} \end{bmatrix} \mathrm{e}^{-\tau s} I_y^{\mathrm{L}}(s) \tag{16}$$

2　洛伦兹力磁浮平台四自由度运动解耦控制

2.1　双 RLFMB 四自由度平转运动耦合特性分析

根据式 (16) 可知，磁悬浮平台转子平动和转动本不存在耦合，但是当载荷舱转子同时发生平动和偏转运动时，平动位移和偏转角度都由电涡流传感器进行检测，使得测量后转子的平动量和转动量存在测量值耦合。如图 5 所示为 x 通道平动和绕 x 轴偏转检测方案示意图。

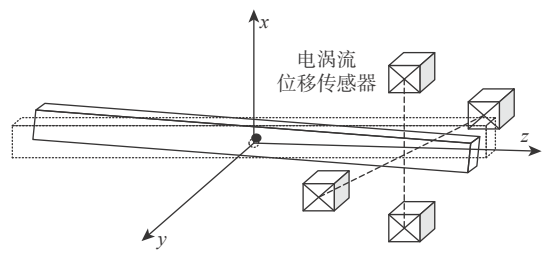


图 5　x 通道检测方案

Fig. 5　x-channel detection scheme

利用传感器测量原理和平转运动偏移值的代数关系，将位移传感器测量值近似为转子平转运动的偏移值，即

$$\begin{aligned} \begin{bmatrix} \alpha_y \\ \beta_y \end{bmatrix} &= \begin{bmatrix} 1 & L_s \\ 1 & -L_s \end{bmatrix} \begin{bmatrix} r_y \\ \theta_y \end{bmatrix} = T_{sy} \begin{bmatrix} r_y \\ \theta_y \end{bmatrix} \\ \begin{bmatrix} \alpha_x \\ \beta_x \end{bmatrix} &= \begin{bmatrix} 1 & -L_s \\ 1 & L_s \end{bmatrix} \begin{bmatrix} r_x \\ \theta_x \end{bmatrix} = T_{sx} \begin{bmatrix} r_x \\ \theta_x \end{bmatrix} \end{aligned} \tag{17}$$

在 RLFMB 闭环控制回路中，对位移传感器测量值进行负反馈控制，得到前向通路中控制电流与传感器测量值的关系为

$$\begin{aligned} \begin{bmatrix} I_y^{\mathrm{L}} \\ I_y^{\mathrm{R}} \end{bmatrix} &= -G_y(s) \begin{bmatrix} \alpha_y \\ \beta_y \end{bmatrix} \\ \begin{bmatrix} I_x^{\mathrm{L}} \\ I_x^{\mathrm{R}} \end{bmatrix} &= -G_x(s) \begin{bmatrix} \alpha_x \\ \beta_x \end{bmatrix} \end{aligned} \tag{18}$$

式中：$G_y(s)$、$G_x(s)$为磁轴承电流控制的传递函数。

联立式 (13)、式 (17)、式 (18) 可以得到磁轴承控制力和控制力矩关于转子运动偏移量的位姿关系：

$$\begin{aligned} \begin{bmatrix} F_y \\ M_y \end{bmatrix} &= -K_{\mathrm{i}} T_y G_y(s) T_{sy} \begin{bmatrix} r_y \\ \theta_y \end{bmatrix} \\ \begin{bmatrix} F_x \\ M_x \end{bmatrix} &= -K_{\mathrm{i}} T_x G_x(s) T_{sx} \begin{bmatrix} r_x \\ \theta_x \end{bmatrix} \end{aligned} \tag{19}$$

将$K_{\mathrm{i}} T_y G_y(s) T_{sy}$和$K_{\mathrm{i}} T_x G_x(s) T_{sx}$作为控制对象，定义

$$\begin{aligned} -K_{\mathrm{i}} T_y G_y(s) T_{sy} &= \begin{bmatrix} -G_{\mathrm{P}}(s) & \\ & -G_{\mathrm{Z}}(s) \end{bmatrix} \\ -K_{\mathrm{i}} T_x G_x(s) T_{sx} &= \begin{bmatrix} -G_{\mathrm{P}}(s) & \\ & -G_{\mathrm{Z}}(s) \end{bmatrix} \end{aligned} \tag{20}$$

式中：$G_{\mathrm{P}}(s)$和$G_{\mathrm{Z}}(s)$分别为转子平动和转动的控制过程。

化简得到

$$\begin{aligned} G_y(s)&=\frac{1}{K_{\mathrm{i}}}T_y\begin{bmatrix} G_{\mathrm{P}}(s) & \\ & G_{\mathrm{Z}}(s)\end{bmatrix}T_{sy}\\ G_x(s)&=\frac{1}{K_{\mathrm{i}}}T_x\begin{bmatrix} G_{\mathrm{P}}(s) & \\ & G_{\mathrm{Z}}(s)\end{bmatrix}T_{sx}\end{aligned}\tag{21}$$

将式(21)代入式(18)得到磁轴承控制方程为

$$\begin{aligned}\begin{bmatrix} I_y^{\mathrm{L}}\\ T_y^{\mathrm{R}}\end{bmatrix}&=-\frac{1}{K_{\mathrm{i}}}T_y\begin{bmatrix} G_{\mathrm{P}}(s) & \\ & G_{\mathrm{Z}}(s)\end{bmatrix}T_{sy}\begin{bmatrix}\alpha_y\\ \beta_y\end{bmatrix}\\ \begin{bmatrix} I_x^{\mathrm{L}}\\ T_x^{\mathrm{R}}\end{bmatrix}&=-\frac{1}{K_{\mathrm{i}}}T_x\begin{bmatrix} G_{\mathrm{P}}(s) & \\ & G_{\mathrm{Z}}(s)\end{bmatrix}T_{sx}\begin{bmatrix}\alpha_x\\ \beta_x\end{bmatrix}\end{aligned}\tag{22}$$

2.2 双RLFMB内模解耦控制器设计

为验证内膜控制器在高精度转台控制系统中的性能优势，与工程实际中广泛应用的PID控制器作为对比。PID控制方法以其结构简单、稳定性和适应性较强等优点在工程实践中得到广泛应用。PID控制器不仅可调参数多，而且变化多种多样，可以很好地对控制信号进行信号修正和误差补偿。

PID控制规律在控制系统中可以用如下代数式表示：

$$\frac{Y(s)}{R(s)}=K_{\mathrm{p}}+\frac{K_{\mathrm{i}}}{s}+K_{\mathrm{d}}s\tag{23}$$

式中：K_{p}、K_{i}、K_{d}分别为对系统误差信号进行比例、积分、微分量的加权，通过加权可以得出控制信号，实现对受控对象模型的控制。

根据式(22)，磁轴承控制方程中$G(s)$均采用PID控制方式，可以得到基于PID的双RLFMB解耦控制器。控制系统中有6个可调参数用于调节系统动态性能。

在实际控制系统中，大多数的控制对象使用常规PID控制即可满足实际的需要，但是对于具有高精度、高敏捷、高动态要求的磁浮平台，常规的PID控制的动态性能指标和抗干扰性能还有待提升。例如，PID控制总是不可避免产生超调，同时PID中积分反馈的引入容易产生较大震荡，闭环反应迟钝，这些都会对磁浮平台高精度控制产生影响。因此将可以同时满足快速响应和振动抑制要求的内模结构应用于磁浮平台的控制中。

内模控制(internal model control, IMC)是从化工应用发展起来的一种对参数变化不敏感的鲁棒控制方法，它不要求被控对象有准确的数学模型，并且控制器参数单一、便于调节。内模控制器框图如图6所示，$Q_1(s)$和$Q_2(s)$作为内模控制器具有两个维度，$Q_1(s)$用于调整系统的跟踪性能，$Q_2(s)$用于调整系统的抗干扰性能。$P(s)$为被控对象，$P_{\mathrm{m}}(s)$为被控对象的数学模型，$d(s)$为系统的干扰量。

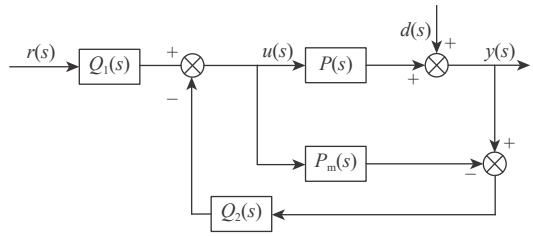


图6 IMC模型框图

Fig. 6 IMC model block diagram

其输入与输出间的传递函数为

$$\begin{aligned}Y(s)=&\frac{P(s)Q_1(s)}{1+Q_2(s)[P(s)-P_{\mathrm{m}}(s)]}R(s)+\\ &\frac{1-P_{\mathrm{m}}(s)Q_2(s)}{1+Q_2(s)[P(s)-P_{\mathrm{m}}(s)]}D(s)\end{aligned}\tag{24}$$

$$\begin{aligned}G_{\mathrm{r}}(s)&=\frac{P(s)Q(s)}{1+Q_2(s)[P(s)-P_{\mathrm{m}}(s)]}\\ G_{\mathrm{d}}(s)&=\frac{1-P_{\mathrm{m}}(s)Q_2(s)}{1+Q_2(s)[P(s)-P_{\mathrm{m}}(s)]}\end{aligned}\tag{25}$$

不考虑建立数学模型误差情况下，即$P(s)=P_{\mathrm{m}}(s)$，取$Q(s)=P_{\mathrm{m}}^{-1}(s)$，在任何干扰作用下都可以得到$G_{\mathrm{r}}=1, G_{\mathrm{d}}=0$，此时系统具有很好的抗干扰特性和跟随特性。为增加系统的鲁棒性，可以在控制器中加入低通滤波器$F(s)$，引入滤波器后的$Q_1(s)$、$Q_2(s)$满足：

$$\begin{cases}Q_1(s)=\dfrac{F_1(s)}{P(s)}\\ Q_2(s)=\dfrac{F_2(s)}{P(s)}\end{cases}\tag{26}$$

式中，

$$\begin{cases}F_1(s)=\dfrac{1}{(\lambda_1+1)^2}\\ F_2(s)=\dfrac{1}{(\lambda_2+1)^2}\end{cases}\tag{27}$$

根据IMC输入与输出的传递函数可以对控制器进行简化，得到简化后的控制器框图如图7所示。

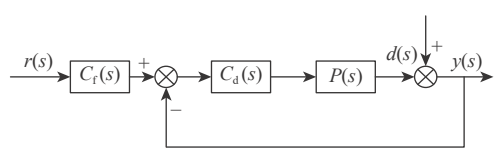


图7 简化IMC模型框图

Fig. 7 Simplified IMC model block diagram

图中，$C_{\mathrm{f}}(s)$、$C_{\mathrm{d}}(s)$与$Q_1(s)$、$Q_2(s)$的关系为

$$\begin{cases}C_{\mathrm{f}}(s)=\dfrac{Q_1(s)}{Q_2(s)}\\ C_{\mathrm{d}}(s)=\dfrac{Q_2(s)}{1-P(s)Q_2(s)}\end{cases}\tag{28}$$

根据上文，磁轴承产生y通道的控制力和控制力矩模型为

$$\begin{bmatrix} F_y \\ M_y \end{bmatrix} = K_{\mathrm{i}} \begin{bmatrix} 1 & 1 \\ L & -L \end{bmatrix} \begin{bmatrix} I_y^{\mathrm{L}} \\ I_y^{\mathrm{R}} \end{bmatrix} \tag{29}$$

以转子受同轴径向磁轴承产生洛伦兹力引起 y 通道上的平动为例, 该通道上被控对象的伪线性表达式为

$$P(s) = \frac{r_y(s)}{I(s)} = \frac{2K_{\mathrm{i}}}{ms^2} \tag{30}$$

将式 (26) 、式 (27)、式 (30) 代入式 (28) 可以得到 y 通道上转子平动 IMC 控制器表达式如下:

$$\begin{cases} C_1(s) = \dfrac{(\lambda_2 s+1)^2}{(\lambda_1 s+1)^2} \\ C_2(s) = \dfrac{ms^2}{2K_{\mathrm{i}}\lambda_2^2 s^2 + 4K_{\mathrm{i}}\lambda_2 s} \end{cases} \tag{31}$$

同理, 可以得到 y 通道上转子转动 IMC 控制器表达式为

$$\begin{cases} C_3(s) = \dfrac{(\lambda_4 s+1)^2}{(\lambda_3 s+1)^2} \\ C_4(s) = \dfrac{J_x s^2}{2K_{\mathrm{i}}L\lambda_4^2 s^2 + 4K_{\mathrm{i}}L\lambda_4 s} \end{cases} \tag{32}$$

可以看到, $C_1(s)$、$C_2(s)$、$C_3(s)$、$C_4(s)$形式上较为简单, 有 4 个可调参数λ_1、λ_2、λ_3、λ_4, 其中λ_1和λ_3用来调节系统的跟踪性能, λ_2和λ_4用来调节系统的抗干扰性能。采用同样的方法可以设计出转子通道平动和偏转运动的内模控制器, 模型框图如图 8 所示。

根据式 (22), 磁轴承控制方程中$G(s)$采用内模控制器, 控制器中输入为位移传感器测量值, 改变平动和转动控制器中 4 个可调参数可以提升控制系统动态性能。

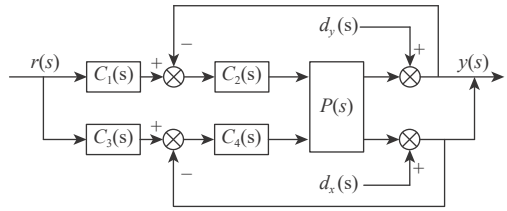


图 8　平转运动 IMC 模型框图

Fig. 8　Block diagram of IMC model for translation and deflection

根据上述磁轴承控制器可以得到洛伦兹力磁悬浮万向稳定平台两自由度平动和两自由度偏转运动的内模控制结构框图如图 9 所示, 图 10 为系统仿真模型。

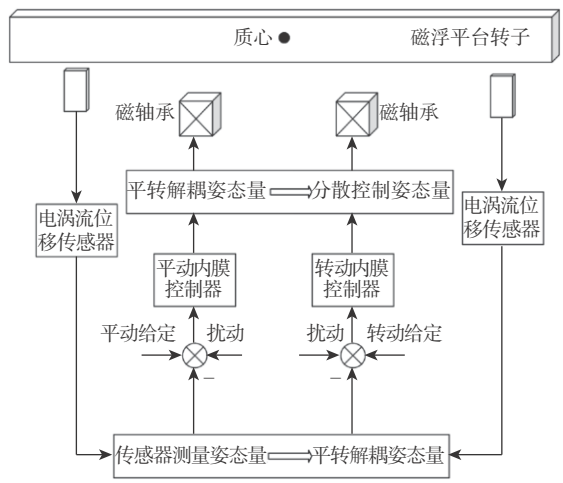


图 9　洛伦兹力磁浮平台内模控制结构框图

Fig. 9　Block diagram of IMC structure of Lorentz force magnetic levitation platform

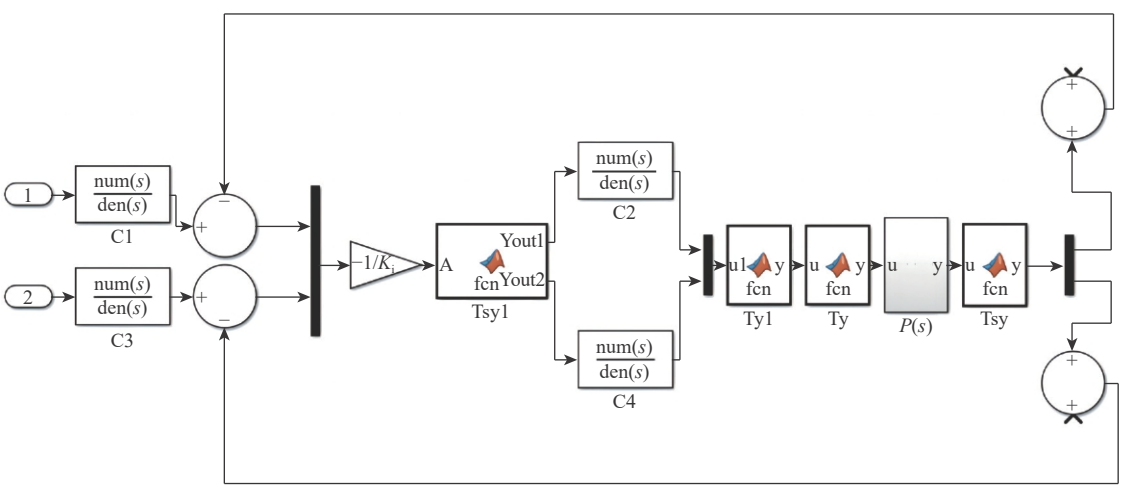


图 10　系统仿真模型

Fig. 10　System simulation model

2.3　基于补灵敏度函数的控制器鲁棒性分析

为验证控制系统强鲁棒的控制性能, 引入补灵敏度函数$T(s)$观察闭环系统的性能, 补灵敏度函数$T(s)$是对系统鲁棒稳定性的度量函数[31]。定义$T(s)$为系统输出$Y(s)$与给定输入$R(s)$之间的传递函数:

$$T(s) = \frac{C_{\mathrm{f}}(s)C_{\mathrm{d}}(s)P(s)}{1 - C_{\mathrm{d}}(s)P(s)} \tag{33}$$

其增益值越小，系统的抗扰能力越强[32]，即系统的鲁棒稳定性越好。

根据仿真及调参结果，PID 控制器总会出现不同程度的尖峰谐振，考虑是系统阻尼低、相角裕度过小导致的，因此分别设计相位补偿器对转子偏转和平动的 PID 控制器进行串联校正。对于转子偏转运动控制，设计相位补偿器在$\omega_m = 3.77$ Hz 处将系统的相角提高$\varphi_m = 44.97°$；对于转子平动控制，设计相位补偿器在$\omega_m = 14.62$ Hz 处将系统的相角提高$\varphi_m = 51.34°$，控制器得到优化之后，谐振峰值明显降低，但是控制系统带宽增加，抗扰能力减弱。将 IMC 的补灵敏度函数 Bode 图和优化后 PID 控制器的补灵敏度函数 Bode 图进行对比，主要参数如表 1 所示，IMC 的补灵敏度函数和 PID 控制器的补灵敏度函数 Bode 图如图 11 所示。

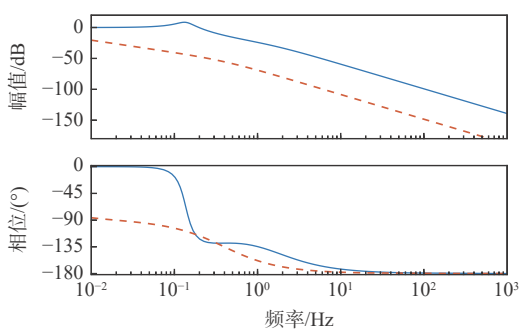
由图 11 可知，转子转动控制过程中，IMC 在 0.01～1 000 Hz 的频段范围内补灵敏度函数增益幅值随频率升高呈下降趋势，并且始终小于 0 dB，PID 控制器在 0.01～0.045 Hz 频段范围内补灵敏度函数在 1dB 以下，在 0～0.13 Hz 频段范围内增益幅值小幅度上升，在 0.13 Hz 达到峰值 8.5 dB，在之后的频段范围内增益幅值快速衰减。转子平动控制过程中，IMC 在 0.01～1 000 Hz 的频段范围内补灵敏度函数增益幅值随频率升高呈下降趋势，同样始终小于 0 dB，PID 控制器在 0.01～0.07 Hz 频段范围内补灵敏度函数在 0.2 dB 以下，在 0～0.46 Hz 频段范围内增益幅值小幅度上升，在 0.46 Hz 达到峰值 7.25 dB，在之后的频段范围内增益幅值快速衰减。

对比两种控制方法的补灵敏度函数 Bode 图可以得到，IMC 在 0.01～1 000 Hz 的频段范围内补灵敏度函数增益幅值始终小于 PID 控制器，并且随着频率的增加，IMC 的增益幅值衰减速度略大于 PID 控制器，内模控制器不仅没有出现谐振峰值，而且增益幅值低，衰减速度快，因此，IMC 的抗干扰性能优于 PID 控制器，具有强鲁棒的特点。

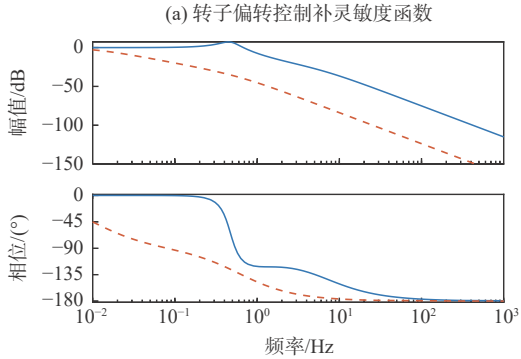
表 1　洛伦兹力平台系统参数

Table 1　Lorentz force platform system parameters

参数	数值	参数	数值
$K_i/(\mathrm{N\cdot A^{-1}})$	52.5	$J_y/(\mathrm{kg\cdot m^2})$	1.76
$J_x/(\mathrm{kg\cdot m^2})$	1.76	m/kg	14.0
L/m	0.15	L_s/m	0.20
λ_1	0.001	λ_2	0.005
λ_3	0.05	λ_4	0.05



(a) 转子偏转控制补灵敏度函数



(b) 转子平动控制补灵敏度函数

图 11　补灵敏度函数 Bode 图

Fig. 11　Bode diagram of complementary sensitivity function

3　内模解耦控制仿真分析

3.1　跟踪性能仿真分析

为了验证 IMC 控制方法的跟踪性能，设计 RLFMB 驱动洛伦兹力平台转子绕 Y 轴偏转方式如图 12 中实线所示，8 s 时，令转子产生正向偏转阶跃信号，偏转角为 0.5°，16 s 时令转子产生反向转动，偏转角恢复为 0°；24 s 时，令转子产生反向偏转阶跃信号，偏转角为 0.5°，32 s 时令转子产生反向转动，偏转角恢复为 0°。

利用 MATLAB 对 PID 控制和 IMC 控制方法的跟随性能进行仿真分析，转子偏转响应曲线如图 12 中虚线所示，偏转跟踪实验动态性能指标如表 2 所示。从仿真结果可以看出，PID 控制方法对偏转阶跃信号的跟踪过程中，超调量约为 12.7%，响应时间约为 0.81 s，保持在 5% 误差带内的调节时间约为 2.56 s；采用 IMC 控制方法对偏转阶跃信号进行跟

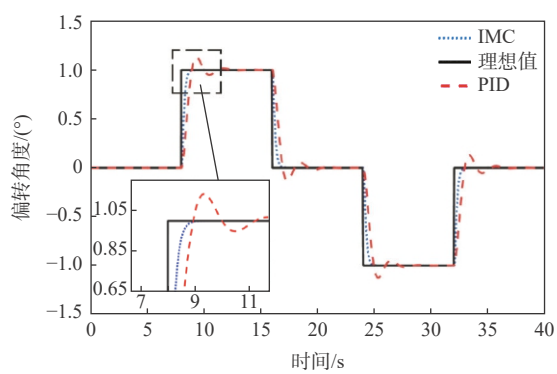


图 12　偏转响应曲线对比

Fig. 12　Comparison of deflection response curves

表 2　偏转跟踪实验动态性能指标

Table 2　Dynamic performance index of deflection tracking experiment

控制系统	调节时间/s	响应时间/s	超调量/%
PID	2.56	0.81	12.7
IMC	0.47	0.33	0

踪, 超调量为 0, 响应时间约为 0.33 s, 较 PID 控制方法下降 59.3%, 保持在 5% 误差带内的调节时间约为 0.47 s, 较 PID 控制方法下降 81.6%。仿真结果表明, 经过参数整定后, IMC 控制方法与 PID 控制方法相比具有更好的跟踪性能, 响应速度更快。

同时施加平动阶跃指令信号, 使 RLFMB 驱动洛伦兹力平台转子沿 y 轴正向运动, 阶跃指令信号如图 13 中实线所示, 15 s 时, 转子沿 y 轴正向运动 1 mm。利用 MATLAB 对 PID 控制和 IMC 控制方法的跟随性能进行仿真分析, 转子平动响应曲线如图 13 所示, 平动跟踪实验动态性能指标如表 3 所示。从仿真结果可以看出, PID 控制方法对平动阶跃信号的跟踪过程中, 超调量约为 8.3%, 响应时间约为 1.35 s, 保持在 5% 误差带内的调节时间约为 2.78 s; 采用 IMC 控制方法对偏转阶跃信号进行跟踪, 超调量为 0, 响应时间约为 0.97 s, 较 PID 控制方法下降 28.2%, 保持在 5% 误差带内的调节时间约为 1.25 s, 较 PID 控制方法下降 55.0%。仿真结果表明, 经过参数整定后, IMC 控制方法与 PID 控制方法相比具有更好的跟踪性能, 响应速度更快。

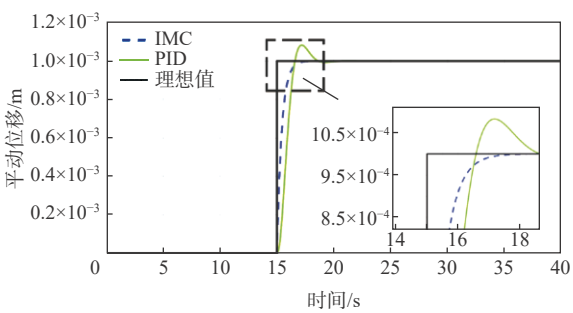


图 13　平动响应曲线对比

Fig. 13　Comparison of translation response curves

表 3　平动跟踪实验动态性能指标

Table 3　Dynamic performance index of translation tracking experiment

控制系统	调节时间/s	响应时间/s	超调量/%
PID	2.78	1.35	8.3
IMC	1.25	0.97	0

结合图 12 和图 13 可以看到, 对 RLFMB 同时输入偏转和平动阶跃信号, 控制器可以分别对洛伦兹力磁悬浮平台的转子进行平动和转动的控制。

根据表 2、表 3 的性能指标可以看到, 内模控制方法相比传统的 PID 控制方法动态性能更优且没有超调量, 跟踪效果好, 响应速度快, 平转运动的响应时间都有不同程度的下降, 其中偏转信号响应时间下降较为明显。强鲁棒、快响应能更好地满足洛伦兹力磁悬浮平台超高敏捷、超高精度的技术需求。

3.2　抗扰性仿真分析

为验证控制系统抗扰性能, 定义转子偏转角度为 0°时为平衡位置, 在 1 s 给位于平衡状态的转子施加绕 y 轴的偏转正弦干扰信号 $0.2\sin(8\pi t)°$, 频率为 4 Hz。利用 MATLAB 对 PID 控制和 IMC 控制方法的抗干扰性能进行仿真分析, 转子响应曲线如图 14 所示。可以看出, 采用 PID 控制方法可以将扰动幅值控制在 0.14°以内, 转子受正弦信号干扰后产生幅值约为 0.12°的往复偏转, 干扰残余量为原干扰信号的 61%; 采用 IMC 控制方法可以将扰动幅值控制在 0.074°以内, 转子受正弦信号干扰后产生幅值约为 0.074°的往复偏转, 干扰残余量为原干扰信号的 37.0%, 偏转幅度较 PID 控制方法减小 38.8%。

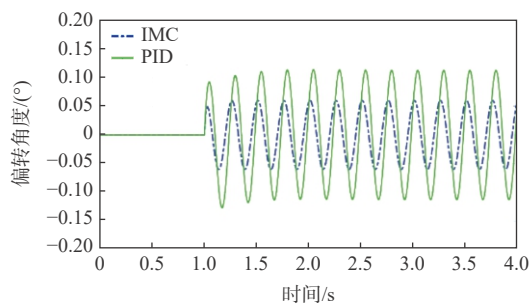


图 14　偏转干扰曲线对比

Fig. 14　Comparison of deflection interference curves

同样, 定义转子平动位移等于 0 mm 时为平衡位置, 在 1 s 给位于平衡状态的转子施加沿 y 轴正向的平动正弦干扰信号 $2\sin(4\pi t)$ mm, 频率为 2 Hz。利用 MATLAB 对 PID 控制和 2IMC 控制方法的抗干扰性能进行仿真分析, 转子响应曲线如图 15 所

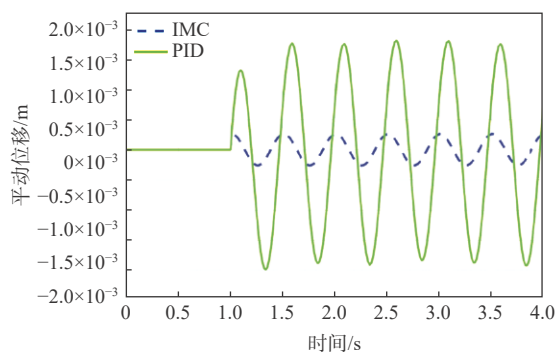


图 15　平动干扰曲线对比

Fig. 15　Comparison of translation interference curves

示。可以看出，采用PID控制方法可以将扰动幅值控制在1.9 mm左右，转子受正弦信号干扰后产生幅值约为1.9 mm的往复偏转，干扰残余量为原干扰信号的95%；采用IMC控制方法可以将扰动幅值控制在0.26 mm左右，转子受正弦信号干扰后产生幅值约为0.26 mm的往复偏转，干扰残余量为原干扰信号的13%，偏转幅度较PID控制方法减小86.2%。

4 结 论

对洛伦兹力磁悬浮万向稳定平台四自由度运动进行动力学建模，设计基于内模结构的解耦控制器分别对平动转动进行控制，对比内模控制和PID控制的系统性能，通过仿真分析得到如下结论。

1）内模控制方法在0.01～1 000 Hz的频段范围内补灵敏度函数增益幅值始终小于PID控制器，内模控制方法具有较好的鲁棒性能。

2）基于控制系统仿真模型的结果表明：内模控制方法可以提高系统的跟踪性能和抗扰性能，偏转、平动跟踪性能上的响应时间分别较PID控制方法降低59.3%和28.2%，偏转、平动抗扰性上的干扰残余量较PID控制方法降低38.8%和86.2%。

为使本文提出的控制方法对高精度磁浮转台有更好的动态性能和抗扰性能，还需优化控制结构和控制系统的各项参数。

参考文献（References）

[1] 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 precise orientation control technology of spacecraft[D]. Harbin: Harbin Institute of Technology, 2011(in Chinese).

[2] 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).

[3] 崔培玲, 盖玉欢, 房建成, 等. 主被动磁悬浮转子的不平衡振动自适应控制[J]. 光学 精密工程, 2015, 23(1): 122-131.
CUI P L, GAI Y H, FANG J C, et al. Adaptive control for unbalance vibration of active-passive hybrid magnetically suspended rotor[J]. Optics and Precision Engineering, 2015, 23(1): 122-131(in Chinese).

[4] 张激扬, 刘虎, 王虹, 等. 飞轮扰振特性及振动控制方法[J]. 空间控制技术与应用, 2014, 40(5): 1-7,18.
ZHANG J Y, LIU H, WANG H, et al. Microvibration characteristics of flywheels and its vibration control approaches[J]. Aerospace Control and Application, 2014, 40(5): 1-7,18(in Chinese).

[5] RAHMAN Z H, SPANOS J T, LASKIN R A. Multiaxis vibration isolation, suppression, and steering system for space observational applications[C]//Proc. SPIE 3351, Telescope Control Systems III. 1998: 73-81.

[6] 周伟勇. 航天器飞轮动力学建模与振动控制研究[D]. 长沙: 国防科学技术大学, 2012.
ZHOU W Y. Research on dynamic modeling and vibration control of spacecraft flywheel[D]. Changsha: National University of Defense Technology, 2012(in Chinese).

[7] LIU Q, WANG K, REN Y, et al. Novel repeatable launch locking/unlocking device for magnetically suspended momentum flywheel[J]. Mechatronics, 2018, 54: 16-25.

[8] LEE J J, AHN C B, CHOI J, et al. Development of magnetic bearing system for a new third-generation blood pump[J]. Artificial Organs, 2011, 35(11): 1082-1094.

[9] SAITÔ M, FUKUSHIMA K, SATO N, et al. Development of low disturbance magnetic bearing wheel (MBW) with inclined magnetic poles[J]. The Journal of Space Technology and Science, 2005, 25: 48-68.

[10] 刘彬, 房建成, 刘刚. 一种磁悬浮陀螺飞轮方案设计与关键技术分析[J]. 航空学报, 2011, 32(8): 1478-1487.
LIU B, FANG J C, LIU G. Design of a magnetically suspended gyrowheel and analysis of key technologies[J]. Acta Aeronautica et Astronautica Sinica, 2011, 32(8): 1478-1487(in Chinese).

[11] 李佳益. 磁悬浮微框架动量轮动力学建模与仿真研究[D]. 长沙: 国防科学技术大学, 2012.
LI J Y. Research on dynamic modeling and simulation of momentum wheel of magnetic levitation micro-frame[D]. Changsha: National University of Defense Technology, 2012(in Chinese).

[12] XIANG B, TANG J Q. Suspension and titling of vernier-gimballing magnetically suspended flywheel with conical magnetic bearing and Lorentz magnetic bearing[J]. Mechatronics, 2015, 28: 46-54.

[13] 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).

[14] 赵勇. 新型洛伦兹力磁轴承磁路设计与分析[D]. 北京: 北京石油化工学院, 2019.
ZHAO Y. Design and analysis of magnetic circuit of a new Lorentz force magnetic bearing[D]. Beijing: Beijing Institute of Petrochemical Technology, 2019(in Chinese).

[15] 熊颖, 刘强, 李衡, 等. 新型洛伦兹惯性稳定平台控制器设计与实现[J]. 火力与指挥控制, 2022, 47(1): 140-144.
XIONG Y, LIU Q, LI H, et al. Design and implementation of a new controller for Lorentz inertial stabilized platform[J]. Fire Control & Command Control, 2022, 47(1): 140-144(in Chinese).

[16] 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, 2017, 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]. Journal of Beijing University of Aeronautics and Astronautics, 2017, 43(3): 559-566(in Chinese).

[17] 熊颖. 新型洛伦兹悬浮定位台控制技术研究[D]. 北京: 北京石油化工学院, 2022.
XIONG Y. Research on control technology of new Lorenz suspension positioning table[D]. Beijing: Beijing Institute of Petrochemic-

al Technology, 2022(in Chinese).

[18] GERLACH B, EHINGER M, RAUE H K, et al. Gimballing magnetic bearing reaction wheel with digital controller[C]//The 6th International ESA Conference on Guidance, Navigation and Control Systems[S.l.]:[s.n.]. 2005: 35-40.

[19] 杨磊, 李建普, 杜金龙, 等. 一种径向磁轴承结构及包含该结构的多自由度磁悬浮机构: CN111828475A[P]. 2020-10-27.
YANG L, LI J P, DU J L, et al. Radial magnetic bearing and multi-axis levitation system: CN111828475A[P]. 2020-10-27(in Chinese).

[20] XIANG B, LIU H, YU Y J. Gimbal effect of magnetically suspended flywheel with active deflection of Lorentz-force magnetic bearing[J]. Mechanical Systems and Signal Processing, 2022, 173: 109081.

[21] 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): 7-11,30.
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): 7-11,30(in Chinese).

[22] 唐忠兴, 姚闯, 何闻, 等. 高精度非接触磁浮机构设计及其输出特性测试研究[J]. 上海航天 (中英文), 2020, 37(1): 135-141,149.
TANG Z X, YAO C, HE W, et al. Research on high-precision non-contact maglev device design and its output characteristic test[J]. Aerospace Shanghai (Chinese & English), 2020, 37(1): 135-141,149(in Chinese).

[23] 李峥, 赵艳彬, 姚闯, 等. 双超卫星平台柔性线缆构型设计与仿真[J]. 上海航天 (中英文), 2022, 39(4): 170-176.
LI Z, ZHAO Y B, YAO C, et al. Configuration design and simulation of flexible cables for dual-super satellite platform[J]. Aerospace Shanghai (Chinese & English), 2022, 39(4): 170-176(in Chinese).

[24] 王有懿, 汤亮, 何英姿. 超静平台隔振与指向一体化控制方法[J]. 航天控制, 2016, 34(6): 33-39.
WANG Y Y, TANG L, HE Y Z. Integrated control method for vibration isolation and pointing control of ultra quiet platform[J]. Aerospace Control, 2016, 34(6): 33-39(in Chinese).

[25] 赵同爽, 张激扬, 王英广, 等. 磁悬浮旋转关节建模及解耦控制方法研究[J]. 空间控制技术与应用, 2022, 48(2): 71-79.
ZHAO T S, ZHANG J Y, WANG Y G, et al. Modeling and decoupling control method of magnetically levitated rotary joints[J]. Aerospace Control and Application, 2022, 48(2): 71-79(in Chinese).

[26] 张井岗, 李临生. 一种二自由度内模控制方法[J]. 工业仪表与自动化装置, 2002(4): 63-65.
ZHANG J G, LI L S. A control method performed by a two-degree-of-freedom internal model[J]. Industrial Instrumentation & Automation, 2002(4): 63-65(in Chinese).

[27] PENG C, FANG J C, XU X B. Mismatched disturbance rejection control for voltage-controlled active magnetic bearing *via* state-space disturbance observer[J]. IEEE Transactions on Power Electronics, 2015, 30(5): 2753-2762.

[28] KANG M S, LYOU J, LEE J K. Sliding mode control for an active magnetic bearing system subject to base motion[J]. Mechatronics, 2010, 20(1): 171-178.

[29] 张玉涵, 蓝益鹏. 磁悬浮平台直线同步电动机 TS 型模糊控制的研究[J]. 电机与控制应用, 2021, 48(10): 1-6,13.
ZHANG Y H, LAN Y P. Research on TS fuzzy control of linear synchronous motor for maglev platform[J]. Electric Machines & Control Application, 2021, 48(10): 1-6,13(in Chinese).

[30] HOWZE J W, BHATTACHARYYA S P. Robust tracking, error feedback, and two-degree-of-freedom controllers[J]. IEEE Transactions on Automatic Control, 1997, 42(7): 980-983.

[31] 房建成, 孙津济, 樊亚洪. 磁悬浮惯性动量轮技术[M]. 北京: 国防工业出版社, 2012: 279-280.
FANG J C, SUN J J, FAN Y H. Magnetically suspended inertial momentum wheel technology[M]. Beijing: National Defense Industry Press, 2012: 279-280(in Chinese).

[32] 夏长峰, 蔡远文, 任元, 等. 磁悬浮控制敏感陀螺转子前馈解耦内模控制[J]. 北京航空航天大学学报, 2018, 44(3): 480-488.
XIA C F, CAI Y W, REN Y, et al. Feedforward decoupling and internal model control for rotor of magnetically suspended control and sensing gyroscope[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(3): 480-488(in Chinese).

Dynamics and decoupling control method of double radial Lorentz force magnetic bearings

LI Zongyu[1], WANG Weijie[2,*], WANG Lifen[2], REN Yuan[2], FAN Yahong[3], LI Lei[1]

(1. Graduate School, Aerospace Engineering University, Beijing 101416, China;
2. Department of Aerospace Science and Technology, Aerospace Engineering University, Beijing 101416, China;
3. Beijing Institute of Control Engineering, Beijing 100190, China)

Abstract: In order to achieve rapid response and vibration suppression of high-precision magnetic levitation turntable and ensure high-precision decoupling control of Lorentz force magnetic levitation omnidirectional stable platform, the working principle of radial Lorentz force magnetic bearings (RLFMBs) was first elucidated. Based on the equivalent magnetic circuit method, a dynamics model of a single RLFMB was established, and a four-degree-of-freedom translational motion dynamics model of the rotor was constructed. A decoupling controller based on the internal model structure was designed, and the strong robustness of the internal model controller was verified by the Bode diagram of the complementary sensitivity function. The simulation results show that in terms of tracking performance, the response time of deflection and translation is reduced by 59.3% and 28.2%, respectively, compared with the proportion-integral-differential (PID) control method. In terms of anti-interference performance, the residual interference of deflection and translation is reduced by 38.8% and 86.2% compared to the PID control method. This method can be applied to high-precision, high-stability, and high dynamic pointing control of the platform's load system.

Keywords: Lorentz force magnetic levitation omnidirectional stable platform; radial Lorentz force magnetic bearing; decoupling control; complementary sensitivity function; internal model control

Received: 2023-05-22; **Accepted**: 2023-06-30; **Published Online**: 2023-09-01 14:33
URL: link.cnki.net/urlid/11.2625.V.20230901.1108.002
Foundation item: National Natural Science Foundation of China (52075545)
* **Corresponding author**. E-mail: wangweijie@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0274

AP虚拟位置感知下的双维度分区定位算法

张阳，秦宁宁*

(江南大学 物联网技术应用教育部工程研究中心，无锡 214122)

摘 要：针对WiFi指纹定位中离线建库效率和成本以及在线匹配定位效率和精度难以平衡的问题，提出了无线接入点(AP)虚拟位置感知下的双维度分区定位算法。算法利用几何特性和传播模型，基于实际测量数据感知AP虚拟位置。离线阶段将空间权重替换为符合信号特性的信号距离权重进行指纹库插值扩充，有效降低了建库成本，同时基于信号覆盖特性对指纹库双维度划分精简，保留各区域内高价值定位信息。在线阶段引入传播相似度与距离和的差异化评判标准，高效精细搜索邻近参考点(RP)。在实测场景中，所提指纹定位算法能够兼顾质量和效率，插值效果相对其他方案提升了17.4%以上。定位算法精度对比同类算法有至少7.1%的提高，在室内场景下具备一定的应用价值。

关 键 词：室内定位；虚拟AP；指纹库扩充；分区；无线接入点；邻近参考点
中图分类号：TN92
文献标志码：A **文章编号**：1001-5965(2025)05-1781-12

伴随互联网技术的高速发展，物理空间与网络世界的边界由于数字化力量的逐渐渗透而日渐模糊，而提供目标定位的位置服务则成为现实与虚拟的重要纽带。由于全球定位系统(global positioning system, GPS)和北斗等室外卫星定位系统无法满足室内定位稳定、准确的需求，基于WiFi的位置指纹定位因其布设成本低、适用范围广和定位可靠性高的优点在诸多定位方案中脱颖而出[1]。

指纹定位法通过物理位置和信号空间的双端映射，比较在线测量值和离线数据在高维空间的指纹相似度筛选邻近参考点(reference point, RP)，进而间接确定目标位置。根据定位过程的时序变换，一般分为离线和在线两个阶段。离线阶段在场景内选取大量参考点并采集WiFi接入点(access point, AP)的接收信号强度(reccived signal strength indicator, RSSI)作为离线指纹，描绘位置特征；在线阶段利用目标位置处的测量数据匹配指纹库具有高相似指纹信息的参考点，通过已有位置完成目标定位[2]。

离线指纹建库需要在预设的多处RP点采集场景内所有可探测AP的RSSI值，而现代建筑技术的发展导致室内区域具有面积庞大、结构复杂的特点，对此类定位场景中海量定位信息进行采集会产生大量的人力物力成本。为解决这一问题，研究人员普遍选取插值方法，探索通过少量高精准RSSI指纹数据扩充生成有效实用的指纹库的方案。反距离加权(inverse distance weighted, IDW)是一种典型空间插值方法，利用与物理距离负相关的位置权重估计待插值点指纹数据，但忽略了不同AP的信号特性，拟合结果往往波动较大[3]。文献[4]则针对异源AP的传播特性建立路径损耗模型，利用小区域内信号波动的相对一致性模拟参数自适应完成指纹预测，然而环境的复杂性和已知AP位置的先决条件使得该方法普适性较低。文献[5]利用改进的克里金算法模拟待插值点的RSSI分布，插值效

收稿日期：2023-05-24；**录用日期**：2023-09-01；**网络出版时间**：2023-10-18 11:58
网络出版地址：link.cnki.net/urlid/11.2625.V.20231017.1417.002
基金项目：长三角科技创新共同体联合攻关项目(2023CSJGG1700)；国家自然科学基金(61702228)；中国高校产学研创新基金(2021ITA10003)
* **通信作者**. E-mail：qinnn@jiangnan.edu.cn

引用格式：*张阳，秦宁宁. AP虚拟位置感知下的双维度分区定位算法[J]. 北京航空航天大学学报，2025，51(5)：1781-1792.*
ZHANG Y，QIN N N. Dual-dimensional partition localization algorithm under AP virtual location perception[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(5)：1781-1792 (in Chinese).

果相对较好，但要求算法中变异函数对于空间特征的精准刻画，对参数选取具有严格限制。

高质量的离线指纹库能够明显提升系统定位性能，匹配算法则关联着定位的效率和精度。对于大型定位场景，采用聚类算法将其划分成多个子区域，通过由粗到细的逐次定位，以二次匹配方式降低邻近RP的搜索范围，加快匹配速度。常用的聚类算法有K-means[6]、FCM[7]和APC[8]等，它们均基于RP的相似度迭代完成子类归属判定，这种划分方式无法规避信号波动引起的类外RP误判到错误子类的现象，同时结果对环境特征契合效果较差。加权K近邻算法(weighted K nearest neighbor, WKNN)因其简单高效的特性被广泛运用于定位匹配阶段，算法以欧氏距离为相似度量标准筛选K个邻近参考点进行位置加权估计，然而单一的相似判别标准和固定K值降低了算法对于不同场景的灵活适用性。文献[9]提出动态K近邻算法弥补了人工选取K值的局限性。文献[10]利用信号特性，提出一种加权欧氏距离平衡信号与空间的差异。文献[11]则进一步提出了一种RP相对排序的准则自适应搜索K值，同时提升了参考点集合质量。对传统欧氏相似度的改进仍难以避免邻近RP误判带来的定位精度下降缺陷，同时缺少对信号本身特性的精准刻画，冗余的计算步骤也可能导致定位效率降低。

针对离线指纹建库代价高、在线定位匹配算法的效率和精度薄弱等问题对定位结果的影响，本文提出了一种AP虚拟位置感知下的双维度分区定位算法(dual-dimensional partition localization algorithm under AP virtual location perception, DDPL)。离线阶段以AP的虚拟位置为基点，将实采点与待插值点的传播距离差值转化为样本空间权重，通过AP加权插值对未采样RP点进行预测实现指纹库的密度扩充。并考虑AP在场景中的传播特性，分析区域内AP在信号域和空间域的相关性，完成RP和AP的双维度区域划分。在线时为避免边缘点误判情况，粗定位聚合相似区域提高搜索容错率。细粒度定位时引入传播相似度与距离和的差异化评判标准，高效精细搜索邻近RP集，提升定位性能。

1 系统模型

1.1 指纹定位算法

指纹定位算法通过目标处的RSSI向量，即指纹，作为目标空间位置的唯一标识。基于WiFi的指纹定位一般分为离线和在线阶段，离线阶段采集指纹数据并构建指纹库，定位区域为大型定位场景时通常会进行区域划分得到一系列子指纹库。在线阶段比较测量目标处的指纹信息和离线指纹库匹配得到目标位置。本文通过解决不同定位环节现存问题提升系统整体性能，定位流程如图1所示。

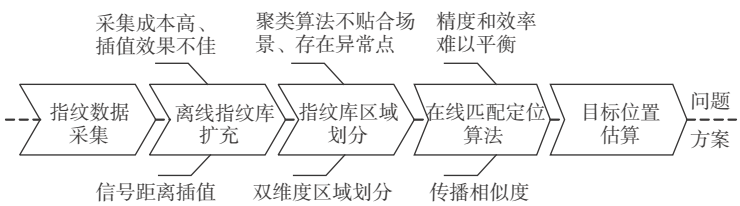


图1 指纹定位流程

Fig. 1 Fingerprint localization process

1.2 位置指纹库建立

不失一般性，将定位区域离散网格化，并于每个网格点按一定间隔均匀布置采样点。本文未作特殊说明下，场景与网络的相关变量与含义通识表述如下。

取每个区域网格的几何中心作为采集接收信号强度的参考点，RP(n)的位置记为$\mathrm{RPL}(n)=(x_n,y_n)$，其中$n\in\{1,2,\cdots,N\}$。RP位置集合为$\mathbf{RPL}=\{\mathrm{RPL}(1),\mathrm{RPL}(2),\cdots,\mathrm{RPL}(N)\}$。场景中可探测到共$M$个AP设备信号，第$m$个AP表示为$\mathrm{AP}(m)\in\mathbf{AP}$，其中$m\in\{1,2,\cdots,M\}$。

受到环境干扰，RSSI测量数据产生波动，以正态分布进行模拟表征。定义指纹集$\mathbf{FP}=\{\mathbf{FP}(1),\mathbf{FP}(2),\cdots,\mathbf{FP}(N)\}$，其中$\mathbf{FP}(n)=[\overline{r_{n,1}},\overline{r_{n,2}},\cdots,\overline{r_{n,M}}]$表示RP($n$)处的指纹**RP**数据。$\overline{r_{n,m}}=\dfrac{1}{Q}\sum_{q=1}^{Q}r_{n,m}^{q}$则为RP($n$)处对AP($m$)的$Q$次测量数据均值。样本标准差$\sigma_{n,m}=\dfrac{1}{Q-1}\sum_{q=1}^{Q}(r_{n,m}^{q}-\overline{r_{n,m}})^2$标识AP($m$)在RP($n$)处的波动程度。

在线阶段，在目标位置$\mathrm{TPL}(g)=(x_g,y_g)$测得来自$M$个AP的接收信号强度$\mathbf{TP}(g)=[\overline{r_{g,1}},\overline{r_{g,2}},\cdots,\overline{r_{g,M}}]$。RP和TP的坐标、对应采集的RSSI测量值和处理后的指纹集三者规律化收集，构建了本文的初始指纹库。

2 AP虚拟位置

WiFi设备无信号接收限制，但受收集成本和隐

私性制约, 通常服务中无法获取设备位置信息, 因此 AP 实际位置通常是未知的。已有研究通过信号传播模型来估算 AP 位置[12-13], 但受环境影响, 计算得到的位置与理想条件下的实际位置存在较大偏差。如图 2 所示, RP($n1$)和RP($n3$)在障碍物的影响下, 接收信号强度相较自由传播的RP($n1$)存在衰弱, 因此拟合得到的 AP 位置发生偏移, 定义环境阻碍下信号跳动引起的偏移位置为 AP 虚拟位置。以 AP 虚拟位置替代真实 AP 位置, 可以在定位计算时一定程度上抵消信号受环境影响的损耗, 弥补在线阶段的定位模型误差。

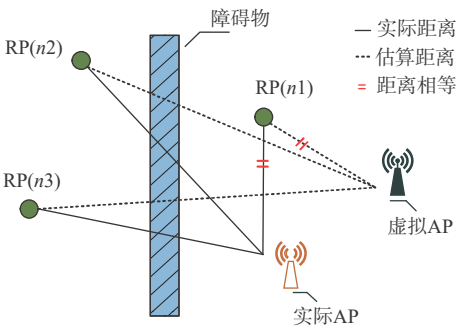


图 2　AP 位置偏差

Fig. 2　AP location deviation

2.1　AP 轨迹圆

AP 的信号覆盖区域可近似看作以节点自身位置为圆心, 最大传播范围为半径的圆形区域。信号强度随着传播距离增大而非线性衰减。对数正态阴影衰减模型适用于描述此类非视距条件下的信号传播特征, 通过转换可求得 AP_m 与 RP_n 的距离 $d_{n,m}$:

$$d_{n,m}=10^{\left(\frac{R_{0,m}-\overline{r_{n,m}}+X}{10\eta}\right)} \tag{1}$$

式中: 距 AP(m)为参考距离 $d_{n,m}$ 处的信号接收强度 $\overline{r_{n,m}}$, $R_{0,m}$ 为 AP(m)在 d_0 处的信号接收强度, d_0 一般为 1 m, η 为环境影响因子, X 为环境噪声。将定位场景中的区域大小约束至较小的范围时, 物理位置邻近的 RP 拥有相同的环境特征, 可近似看作共享 η 和 X。选取在空间中位置接近的RP($n1$)和RP($n2$), 定义二者与AP(m)的距离比 $p^m_{n1,n2}$:

$$p^m_{n1,n2}=\frac{d_{n1,m}}{d_{n2,m}}=10^{\left(\frac{\overline{r_{n2,m}}-\overline{r_{n1,m}}}{10\eta}\right)} \tag{2}$$

由上可知, $RSSI_{0,m}$ 和 X 被消除, 意味着在数据采集阶段无须考虑设备和部分环境噪声影响, 促进了定位过程降本增效。根据阿氏圆定理, 与两点的距离比为常数且不为 1 的动点轨迹是一个圆[13]。若 $RPL(n1)=(x_{n1},y_{n1})$ 和 $RPL(n2)=(x_{n2},y_{n2})$, 定义 AP(m)的虚拟位置坐标为 $APL(m)=(x^a_m,y^a_m)$, 则轨迹圆描述如下:

$$\sqrt{(x_{n1}-x^a_m)^2+(y_{n1}-y^a_m)^2}=p^m_{n1,n2}\sqrt{(x_{n2}-x^a_m)^2+(y_{n2}-y^a_m)^2} \tag{3}$$

当满足式 (3) 时, 其中:

$$\Delta=\left(\frac{2x_{n2}(p^m_{n1,n2})^2-2x_{n1}}{1-(p^m_{n1,n2})^2}\right)^2+\left(\frac{2y_{n2}(p^m_{n1,n2})^2-2y_{n1}}{1-(p^m_{n1,n2})^2}\right)^2-4\frac{(x_{n1})^2+(y_{n1})^2-(x_{n2}p^m_{n1,n2})^2-(y_{n2}p^m_{n1,n2})^2}{1-(p^m_{n1,n2})^2}>0 \tag{4}$$

则根据式 (3) 和式 (4) 可确定, 由RP($n1$)和RP($n2$)确定 AP(m)的轨迹圆心 $C^m_{n1,n2}=(x^{rm}_{n1,n2},y^{rm}_{n1,n2})$ 坐标为 $\left(\frac{x_{n1}-x_{n2}(p^m_{n1,n2})^2}{1-(p^m_{n1,n2})^2},\frac{y_{n1}-y_{n1}(p^m_{n1,n2})^2}{1-(p^m_{n1,n2})^2}\right)$, 半径为 $r^{rm}_{d,n1,n2}=\left|\frac{(p^m_{n1,n2})^2}{1-(p^m_{n1,n2})^2}\right|d_{n1,n2}$, AP($m$)的位置在圆上。

2.2　AP 虚拟位置确定

成对 RP 可以确定一个 AP 轨迹圆, 不同 RP 组合使轨迹圆拥有不同的圆心和半径。如图 3 所示, AP(m)位置等概率分布于由RP($n1$)、RP($n2$)和RP($n3$)两两确定的 $C^m_{n1,n3}$、$C^m_{n2,n3}$ 和 $C^m_{n2,n3}$ 上, 3 个轨迹圆的交点则为虚拟 AP 的最佳位置。

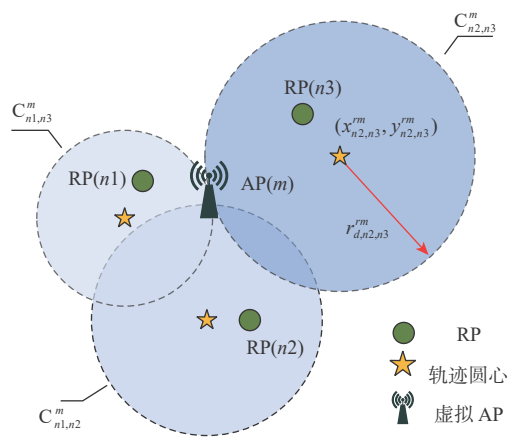


图 3　AP 虚拟位置

Fig. 3　AP virtual location

实际场景中, 信号非视距传播的路径损耗会导致轨迹圆的交点不唯一, 因此利用非线性最小二乘原理可以估算出与所有轨迹圆相距最近的坐标作为 AP 虚拟位置。

为减弱 AP 信号自身波动和远距 RP 无效信号接收强度值的影响, 提升容错裕量, 本文采取均匀采样方法以 r_t 为比例抽取 $r_t\cdot N$ 个 RP 点, 重复抽样产生 T 组。对于第 $t\in\{1,2,\cdots,T\}$ 组集合, AP(m)的位置 $APL(mt)=(x^{a(t)}_m,y^{a(t)}_m)$ 可由非线性最小二乘拟合得到:

$$APL(mt)=\underset{x^{a(t)}_m,y^{a(t)}_m}{\operatorname{argmin}}\sum_{rm=1}^{N^t}\left(y_{rm}-\sqrt{(x^{a(t)}_m-x_{rm})^2+(y^{a(t)}_m-y_{rm})^2}\right)^2 \tag{5}$$

式中：$N^t=\dfrac{r_t\cdot N(r_t\cdot N-1)}{2}$为第$t$组中成对RP确定的轨迹圆数量。对$T$组RP集合按式(5)执行可得到AP($m$)的待定虚拟位置集合{APL($m1$),APL($m2$),⋯,APL($mT$)}。

通常待定位置分布较为集聚，而非视距RP或远距RP的信号波动会造成待定位置显著偏移真实位置。而局部异常因子算法(local outlier factor, LOF)通过比较待选点周围样本点的平均密度和待选点位置的密度，可有效分辨样本总体的离群值。对待选位置集合执行LOF剔除异常位置维护待选点分布的紧密性，确定AP(m)的虚拟位置为滤除后的T'个虚拟位置集合均值$\mathrm{APL}(m)=(x_m^a,y_m^a)=\dfrac{1}{T'}\sum_{t=1}^{T'}(x_m^{a(t)},y_m^{a(t)})$。逐一计算所有AP虚拟位置最终得到AP虚拟位置集合$\boldsymbol{APL}=\{\mathrm{APL}(1),\mathrm{APL}(2),\cdots,\mathrm{APL}(M)\}$。

3 离线指纹库扩充

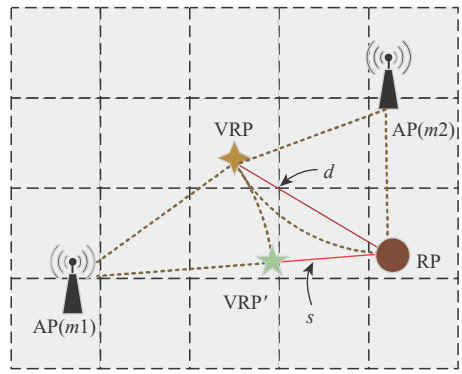
逆距离加权插值法利用RP点在空间中的近邻性，通过位置距离判定实采点的插值权重。然而在实采RP分布较稀疏的情况下无法获得丰富的数据量，进而会放大环境干扰对距离权重的影响，降低插值效果。同时忽略同一RP处不同AP在空间内信号浮动的差异，采用相同距离权重会极大削弱信号的分布特征，削弱AP异源差异对预测值的影响。

对比图4(a)和图4(b)，VRP′为VRP在以AP($m1$)为圆心向RP传播时具有相同传播距离的同位点，二者位置接收到来自AP($m1$)的RSSI应基本相同。设VRP′与RP的距离差sd为信号距离，图中RP与VRP的距离d增加，信号传播距离sd反而减小。而对AP($m2$)的信号强度进行预测时，VRP与RP传播距离相似，此时RSSI值应基本相同，d的引入反而降低了RP的权重。可见，采用信号距离差更能表述不同RP的插值权重，提高插值效果。

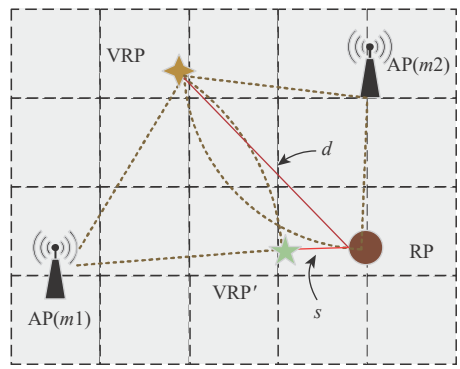
设待插值点坐标$\mathrm{VRP}(i)=(x_i^v,y_i^v)$，则对于任意实采点RP($j$)，其与VRP($i$)的关于AP($m$)的信号距离$s_{ij}^m$表示为

$$s_{ij}^m=\left|\sqrt{(x_m^a-x_i^v)^2+(y_m^a-y_i^v)^2}-\sqrt{(x_m^a-x_j)^2+(y_m^a-y_j)^2}\right| \tag{6}$$

在邻近区域内，VRP(i)和RP(j)由于环境特征基本相同，信号分布相似度较高，由图4可知$s_{ij}^m<d_{ij}$，同时二者物理距离越近则信号距离越近。随着二者物理距离逐渐增大，环境影响带来的信号波动会使得信号距离参考价值急剧降低。因此本文引入RP相对VRP的环境相似度ξ，用于衡量实采点和待插值点的物理距离，剔除远距离实采点RP对插值



(a) RP与VRP距离近



实采点 待插值点 信号传播路径

(b) RP与VRP距离远

图4 信号距离和物理距离对比

Fig. 4 Comparison between signal distance and physical distance

效果的影响：

$$\begin{cases}\mathrm{RP}(j)\in\boldsymbol{\Omega}_i^v & \dfrac{\frac{1}{M}\sum_{m=1}^{M}s_{ij}^m}{d_{ij}}\geqslant\xi\\ \mathrm{RP}(j)\notin\boldsymbol{\Omega}_i^v & \text{else}\end{cases} \tag{7}$$

式中：当RP(j)相对VRP(i)的对于所有AP的平均信号距离与物理距离比值小于ξ时，将RP(j)从VRP(i)的邻近区域$\boldsymbol{\Omega}_i^v$中剔除。最后选择$\boldsymbol{\Omega}_i^v$中RP参与插值。定义AP(m)下$\mathrm{RP}(j)\in\boldsymbol{\Omega}_i^v$在VRP($i$)的信号距离权重$u_{ij}^m$：

$$u_{ij}^m=\frac{(s_{ij}^m+\varepsilon)^{-1}}{\sum_{\mathrm{RP}(l)\in\boldsymbol{\Omega}_i^v}(s_{il}^m+\varepsilon)^{-1}} \tag{8}$$

其中，$\varepsilon\to0$，作用为防止分母为0。可进一步求得VRP(i)的预测指纹信息$\boldsymbol{FP}(vi)=\sum_{\mathrm{RP}_j\in\boldsymbol{\Omega}_i^v}\boldsymbol{u}_{ij}\boldsymbol{FP}(j)$，其中$\boldsymbol{u}_{ij}=[u_{ij}^1,u_{ij}^2,\cdots,u_{ij}^M]$。以AP虚拟位置为锚点，对所有待采样点执行信号距离权重插值获得虚拟指纹信息，对指纹库进行信息填充，至此完成离线指纹库构建。

4 离线双维度区域划分

传统聚类方式以RP之间的信号空间距离衡量

相似性, 通过不断迭代完成子指纹库的划分。这种尺度单一的划分标准容易忽视场景空间结构和信号漂移的影响, 部分物理距离远、信号距离近的 RP 被归属于同一类, 导致待定位点预测位置偏移。本文提出一种结合信号域与空间域的指纹库划分方式, 将小范围区域内 AP 作为划分锚点, 从 RP 个体挖掘与场景区域的内在联系, 以熵值评判作为划分依据, 完成基于 AP 与 RP 信号的双维度区域划分。

4.1　区域 AP 集团

信号传播具有随距离增加而非线性衰减的特征, 因此 AP 周边小范围区域拥有较高的信号强度, 同时由于减少了远距区域环境内信号折射、障碍阻挡等影响, 信号价值更加稳定真实。将这一变化特征作为区域划分尺度, 累积同一区域内邻近 AP 的信号热度差异作为不同区域间的辨别标准, 形成信号与区域强耦合的 AP 集团。

AP 聚类算法选取经典 K-means, 为兼顾 AP 间在物理空间和信号空间的相关性, 本文对相似度标准提出改进, 在物理距离的基础上引入 AP 在 RP 总体上的信号分布相似性[14]。根据皮尔逊系数定义 AP_{m1} 和 AP_{m2} 在信号域上的相似系数 $\rho_{m1,m2}$:

$$\rho_{m1,m2}=\frac{\sum_{n=1}^{N}(\overline{r_{n,m1}}-a_{m1})(\overline{r_{n,m2}}-a_{m2})}{\sqrt{\sum_{n=1}^{N}(\overline{r_{n,m1}}-a_{m1})^2}\sqrt{\sum_{n=1}^{N}(\overline{r_{n,m2}}-a_{m2})^2}} \tag{9}$$

式中: $a_{m1}=\frac{1}{N}\sum_{n=1}^{N}\overline{r_{n,m1}}$, 为 AP($m1$)的 RSSI 均值。$\rho_{m1,m2}\in[-1,1]$, 值越大, 信号空间中 AP($m1$)和 AP($m2$)越接近, 综合 AP 在空间域和信号的衡量尺度, 确立新的相似度标准 $\mathrm{Sim}(m1,m2)$:

$$\mathrm{Sim}(m1,m2)=\mathrm{Euc}(m1,m2)\cdot \mathrm{e}^{\frac{1}{1-\rho(m1,m2)+\varepsilon}} \tag{10}$$

其中, $\mathrm{Euc}(m1,m2)=[(x_{m1}-x_{m2})^2+(y_{m1}-y_{m2})^2]^{-\frac{1}{2}}$ 表示 AP($m1$)和 AP($m2$)的空间距离。$\mathrm{Sim}(m1,m2)$ 值越大, 相似程度越高。为避免分母为 0, 限制因子 $\varepsilon>0$ 且 $\varepsilon\to 0$。

根据 Sim 在给定 K 取值情况下, 对 AP 归属判定, 得到 K 个 AP 集团 $\{\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2,\cdots,\boldsymbol{\Omega}_K\}$。其中 $\mathrm{AP}(m(k))\in\boldsymbol{\Omega}_k$, $k\in\{1,2,\cdots,K\}$ 为隶属于第 k 个 AP 集团的第 m 个 AP。

4.2　熵值判定聚类

AP 维度的划分实现了具有区域特征强相关的 AP 子集团, 将 AP 集团在各自区域独有的信号特征作为锚点, 进一步发掘 RP 对单点 AP 的独有归属因子, 可以由点及面推算 RP 对不同 AP 集团的归属度。

对于 RP 接收到的多源 AP 信号而言, 接收强度较高的 AP 个体无疑与其有更紧密的联系, 因此在判定归属时应该更偏向该 AP 所属的区域集团。以 $\mathrm{AP}(m(k))$ 在指纹点 $\mathrm{RP}(n)$ 的接收信号强度表示二者归属因子 $\mu_{n,m}$:

$$\mu_{n,m}=\overline{r_{n,m}}\cdot \mathrm{e}^{\frac{1}{\sigma_{n,m}}} \tag{11}$$

式中引入了 $\mathrm{RP}(n)$ 处 $\mathrm{AP}(m(k))$ 的样本标准差 $\sigma_{n,m}$ 表征 $\mathrm{AP}(m(k))$ 的信号波动程度, 可间接提升信号稳定的 AP 权重。累计 RP 在不同区域集团内的归属因子值, 从而放大不同区域对其的吸引力, 完成 RP 区域归属判定。熵值以信息量度量事件发生的不确定性, 利用熵定权形式将 RP 归属于不同 AP 集团的概率转化为归属因子特征分布的期望:

$$i_{n(k)}=-\frac{\sum_{m=1}^{N(\boldsymbol{\Omega}_k)}\mu_{n,m(k)}\log(\mu_{n,m(k)})}{\log[N(\boldsymbol{\Omega}_k)]} \tag{12}$$

式中: $N(\cdot)$ 表示区域中的 AP 数量。$i_{n(k)}$ 为 $\mathrm{RP}(n)$ 对 $\boldsymbol{\Omega}_k$ 中所有 AP 的归属度累计值, 利用熵值形式反映了 $\mathrm{RP}(n)$ 被划分至该区域的概率, $i_{n(k)}$ 越低, 被划分至 $\boldsymbol{\Omega}_k$ 的可能性越高, 因此选取熵值权重最低的区域作为归属区域, 即 $\min(i_{n(k)}),k\in\{1,2,\cdots,K\}$。据式 (12) 计算 N 个 RP 对各个区域的熵值权重, 确定所有 RP 归属, 并更新所有 RP 的区域标签, 若 $\mathrm{RP}(n)\in\boldsymbol{\Omega}_k$ 则表示为 $\mathrm{RP}(n(k))$。

4.3　区域划分流程

本文通过离线双维度划分, 将 AP 和 RP 共同化整为零分散至各个子区域, 降低了总体指纹库体量同时保留了区域内子指纹库内的丰富定位信息, 优化了不同区域的信号覆盖整体效果。具体流程如下所示。

输入: **APL**, **FP**, $\boldsymbol{\sigma}$

输出: $\boldsymbol{\Omega}_k$

1. Initialize $\boldsymbol{\Omega}_k=\varnothing$　　\\$\varnothing$表示空集

2. For $\mathrm{AP}(m1)$, $\mathrm{AP}(m2)\in\mathrm{AP}$　　\\ $m1,m2\in[1,M]$& $m1\neq m2$

3. compute $\rho_{m1,m2}$, $m_{m1,m2}$　　\\ AP 相似度, 基于式 (9)、式 (10)

4. End for

5. For $\mathrm{AP}(m)\in\mathbf{AP}$　　\\ 区域 AP 集团

6. $\mathrm{AP}(m)=K-\mathrm{means}(\mathrm{APL})$　　\\ 执行 K-means

7. $\mathrm{AP}(m(k))\to\boldsymbol{\Omega}_k$　　\\ 更新 AP 区域标签

8. End for

9. For $\mathrm{RP}(n)\in\mathbf{RP}$

10. For $\mathrm{AP}(m(k))\in\boldsymbol{\Omega}_k$

11. compute $\mu_{n,m(k)}$　　\\ 归属度因子, 基于式 (11)

12. End for

13. compute $i_{n(k)}$　　\\ RP 熵值权重, 基于式 (12)

14. Ext[min($i_{n(k)}$)] → $\boldsymbol{\Omega}_k$ \\ Ext[·]表示将RP(n)提取至熵值最小的子区域

15. End for

16. Return $\boldsymbol{\Omega}_k$ \\ $\boldsymbol{\Omega}_k$为RP、AP双维度信息构成子指纹库

5 在线定位

5.1 归属区域聚合

分区定位算法通常分为粗定位和细粒度定位两个阶段。在粗定位阶段，确定测试点归属区域时一般通过比较聚类中心点与测试点的相关程度，该方式依赖于聚类中心点对自身聚类特征表达的准确性，但无法避免聚类中心点信息量限制带来的区域选择误判[15]。本文将AP集团的区域特征作为差异，根据式(11)、式(12)确定测试点归属区域。由TP(g)处采集的指纹数据得到熵值权重集合$\mathbf{IG}(g)=\{i^g_{(1)},i^g_{(2)},\cdots,i^g_{(k)},\cdots,i^g_{(K)}\}$。考虑到不同AP集团在区域交界处大多具有相似的覆盖强度，信号均值差异不明显，利用区域阈值γ将具有相近熵值权重系数的区域合并，提升TP在边界处适配RP信息裕量：

$$\frac{i^g_{(k)}-\min(\mathrm{IG}(g))}{\min(\mathrm{IG}(g))}<\gamma \tag{13}$$

式中，min(IG(g))表征熵值最低的初始选择区域。将满足阈值条件的适配区域合并入待选区域$\boldsymbol{\Omega}^g$，确定RP($n(g)$),AP($m(g)$)$\in\boldsymbol{\Omega}^g$，将RP($n(g)$),AP($m(g)$)对应的指纹数据存入待选指纹点集。

5.2 邻近RP匹配

粗定位完成在线区域判定后，将通过细粒度定位筛选邻近RP集最终确定测试点位置。用于测试点位置计算的RP点应尽量靠近目标位置，通常基于指纹信息向量的欧氏距离、余弦距离等相似度判别方式进行筛选，由于环境干扰和目标位置的信号波动因素，无法保证筛选点和TP之间的紧密性。在位置邻近的距离约束下，TP和邻近RP在同一AP的传播距离总是相接近，基于这一原理，本文提出以传播距离比衡量待选RP和目标位置向量的紧密程度，实现邻近RP筛选。

5.2.1 传播相似度

通过对数正态阴影模型可以一定程度上模拟信号的传播特性，而在细粒度定位阶段，TP周围分布的近邻RP环境布局非常相似，RP之间的环境损耗也非常相近。而参考点与目标位置越接近，同一AP发射信号传播至二者时的传播距离也会越相似。

由式(2)可知区域$\boldsymbol{\Omega}^g$中，RP($n(g)$)与AP($m(g)$)的距离$d_{n(g),m(g)}$，进而可得TP(g)和RP($n(g)$)与AP($m(g)$)传播距离比$\dfrac{d_{g,m(g)}}{d_{n(g),m(g)}}=10^{\left(\frac{\overline{r_{g,m(g)}}-\overline{r_{n(g),m(g)}}}{10\eta}\right)}$，比值越小二者越接近，因此可用$\boldsymbol{\Omega}^g$内AP的距离比作为传播相似度psi($n(g)$)的衡量依据：

$$\mathrm{psi}(n(g))=\frac{1}{N(\boldsymbol{\Omega}^g)}\sum_{m(g)=1}^{N(\boldsymbol{\Omega}^g)}\frac{d_{g,m(g)}}{d_{n(g),m(g)}} \tag{14}$$

式中：psi($n(g)$)为所有AP下TP和RP($n(g)$)的传播距离比之和。待选RP与TP位置越接近，二者在共有AP的传播距离也越接近，psi($n(g)$)值越小。

5.2.2 最优RP集

为自适应选取置信度最高的指纹点确定最优邻近RP集，将式(14)计算得到的传播相似度进行升序排列，初始选择前C个RP点{RP(1(g)),RP(2(g)),⋯,RP($c(g)$),⋯,RP($C(g)$)}。结合离线已知的**RPL**和**APL**，可得RP($c(g)$)与$\boldsymbol{\Omega}^g$内所有AP的位置距离和$d_{s,c(g),m(g)}=\sum_{m(g)=1}^{N(\boldsymbol{\Omega}^g)}d_{c(g),m(g)}$。邻近RP集合分布应相对集聚，因而集合中离群的异常RP距离和相对其余指纹点差异显著，可通过下列判别条件剔除：

$$\begin{cases}\mathrm{RP}(c(g))\in\mathrm{LRP}(g),\ \left|d_{s,c(g),m(g)}-d_{s,l(g),m(g)}\right|\leqslant\\ \qquad\dfrac{1}{C-1}\displaystyle\sum_{c\neq1}\left|d_{s,c(g),m(g)}-d_{s,l(g),m(g)}\right|\\ \mathrm{RP}(c(g))\notin\mathrm{LRP}(g),\ \text{其他}\end{cases} \tag{15}$$

$d_{s,1(g),m(g)}$为$p_{n(g)}$值最小，传播相似度最高的RP(1(g))，将其作为核心指纹点来判定离群RP点，确定最优邻近RP集合LRP(g)={RP(1(g)),RP(2(g)),⋯,RP($\hat{c}(g)$),⋯,RP($\hat{C}(g)$)}，包含$\hat{C}$个RP。

5.3 位置计算

$\mathbf{LRP}_g$对应的传播相似度{$p_{1(g)},\cdots,p_{\hat{c}(g)},\cdots,p_{\hat{C}(g)}$}体现了不同RP的相似程度级别，利用其作为权重配比，定义$RP(\hat{c}(g))$的权重$w_{\hat{c}(g)}=(p_{\hat{c}(g)})^{-1}\Big/\sum_{\hat{c}=1}^{\hat{C}}(p_{\hat{c}(g)})^{-1}$，最终可确定TPL($g$)的坐标$(x_g,y_g)$如下：

$$(x_g,y_g)=\sum_{\hat{c}=1}^{\hat{C}}w_{\hat{c}(g)}(x_{\hat{c}(g)},y_{\hat{c}(g)}) \tag{16}$$

6 定位流程

本文所提DDPL定位算法的整体流程如图5所示。

离线阶段，利用最小二乘拟合与离群值剔除完成AP的虚拟位置估计，并据此进一步通过信号距离替代物理距离完成离线指纹库的插值扩充，提高了建库效率。同时基于AP区域特征划分AP集团，从熵值变化角度为RP合理分配归属区域，形成高质量的小尺度指纹库。在线阶段，通过区域聚合方式扩充待选RP集合定位信息丰富度，提出传播

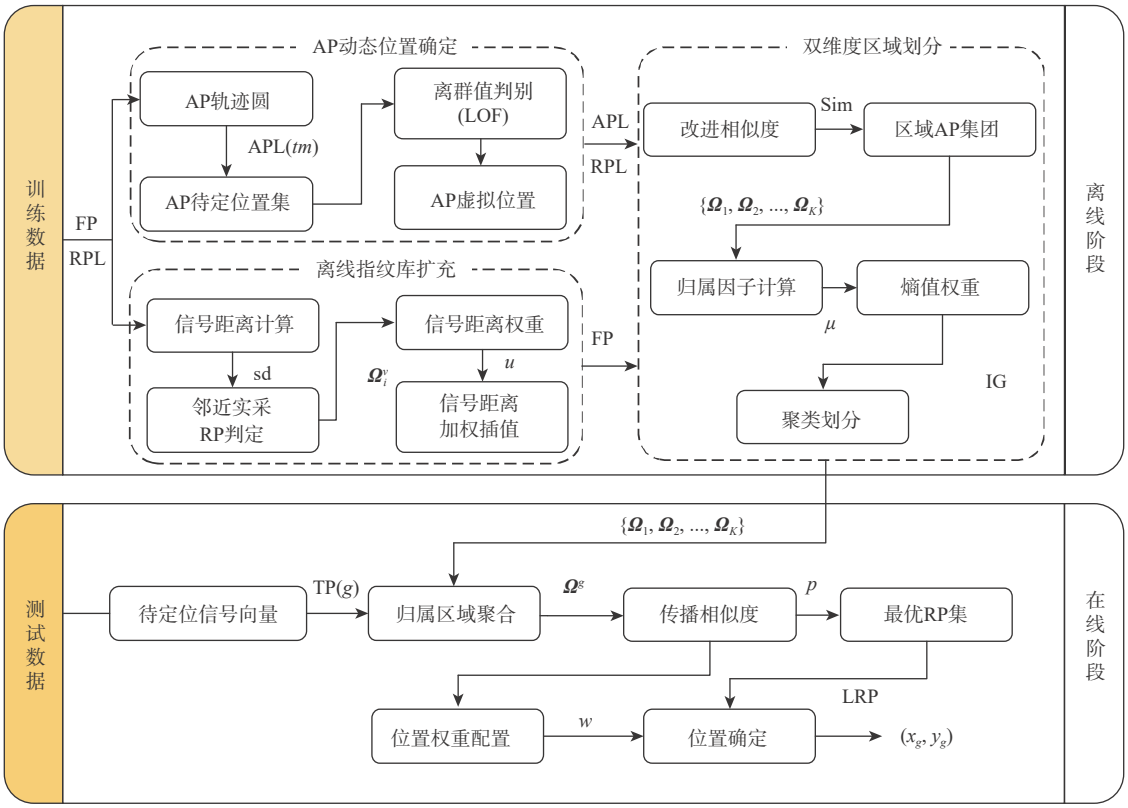


图 5　DDPL 定位流程

Fig. 5　DDPL algorithm process

相似度筛选邻近 RP 集合, 比较 RP 与 AP 距离和差异自适应剔除异常离群 RP, 获取高置信度的近邻 RP 权重, 最终确定目标位置。

7　实验分析

7.1　实测场景与数据处理

为实际测试本文所提 DDPL 算法的定位效果, 实验场景选取江南大学物联网工程学院 C 区三楼进行定位测试。将实验场景网格化, 四边走廊均匀铺设共 $N=368$ 个 RP, RP 之间间隔均为 1 m。为消除异质测量设备带来的影响, 实验人员持同一设备在每个 RP 点采集数据 $Q=50$ 次, 每次间隔 1.2 s。场景中采集的 AP 数量为 $M=30$, 均来自走廊两侧房间内部署的 WiFi 设备, 实际位置未知, 以 MAC (media access control, 媒体存取控制)地址识别并编号。测试数据由实验人员沿走廊中间环行一周, 测得 $G=185$ 个 TP 数据。已有研究表明, 走廊等室内区域中环境影响因子在 2.2～2.8 波动[16], 不失一般性, 本文选取 $\eta=2.6$。经实测, 设 $r_t=0.5$, $T=60$ 为稳定实验参数配置。所有数据以 MATLAB2016b 处理, RP 处未测得的 AP 信号强度以−100 dBm 标识。

7.2　实验参数配置

7.2.1　环境相似度

环境相似度是通过实采点和待插值点之间信号距离和物理距离的相关程度, 筛选待插值点位置邻近且具有参考价值的 RP。而参与插值的 RP 点对于扩充后的指纹库质量息息相关, 为了得到贴近场景特征的环境相似度 ξ, 不失一般性, 实验抽取比例 $h=0.2$ 的参考点用于预测剩余参考点的指纹数据, 形成完整指纹库。对比不同 ξ 时对所有 AP 的平均拟合误差, 拟合误差为预测值和实际值的差值绝对值。

图 6 展示了 ξ 对拟合误差的影响, 随着 ξ 值的增加, 平均拟合误差呈先降后升的趋势。由于距待插值点较远的 RP 参与信号值估算, 环境对信号分布的影响差异, 当 ξ 较小时, 此类 RP 的传播距离与真

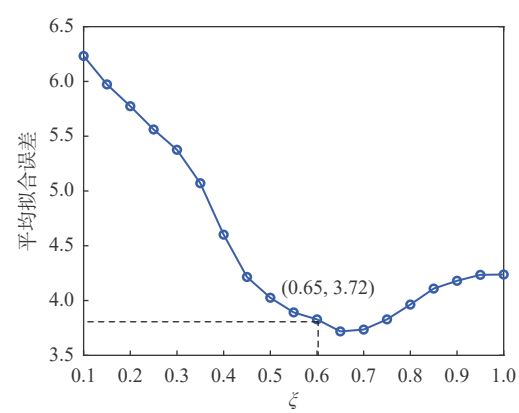


图 6　ξ 对平均拟合误差的影响

Fig. 6　Effect of ξ on average fitting error

实信号距离存在较大出入。随着ξ增加，远距RP被逐渐剔除，邻近RP的权重被提升，在$\xi=0.65$时达到最低点，此时平均拟合误差仅为3.72。而ξ的继续增加对邻近的RP要求更加严苛，使得部分预测贡献价值高的RP被剔除，降低了插值效果。为保证实验效果，后续实验均取$\xi=0.65$。

7.2.2 区域阈值

为分析在线阶段归属区域聚合弥补部分待定位点定位信息不足，避免聚类硬性划分导致的聚类边缘另一侧邻近RP点失落的问题，本文针对区域阈值展开最优值研究。实验选定区域阈值范围$\gamma\in[0.1,0.8]$，且以0.01的增量对所有TP点进行目标定位，并以不同γ下的平均定位误差作为衡量标准，相应的γ与平均定位误差关系如图7所示。

区域阈值影响着目标定位时在线即时指纹库的聚合规模，γ越大，则聚合规模越大，更多的RP参与能够提供更多的定位参考信息，然而聚合多个区域也会带来额外的计算成本，降低计算效率。由图7可以看出，随着γ值的增大，平均定位误差呈下降趋势。在初期阶段，区域边缘待定位点由于提高了信息裕量，使得定位误差快速下降，在$\gamma=0.26$时定位误差骤降至0.886 m。然而γ继续增加也会导致指纹库夹杂其他区域AP的信息特征而削弱了匹配定位的辨识力，因此定位误差上下波动。算法中γ在

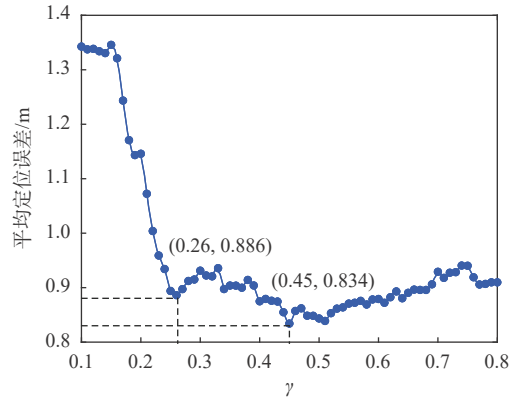


图7 γ对平均定位误差的影响

Fig. 7 Effect of γ on average localization error

取值0.45时达到最优，不失一般性，选取实验场景的区域阈值$\gamma=0.45$。

7.2.3 区域AP集团数量

为验证信号分布相似度的引入对定位性能的影响，最大化的贴合场景特征，体现AP集团对定位场景的覆盖优势，需要根据场景特征确定最优AP集团数量K。为保证对比时的参数一致性，本文选取区域阈值$\gamma=0.45$、初始邻近RP数量$C=5$，选取$K\in[4,8]$，以定位误差为评价标准，对比欧氏距离Euc相似度与本文所提相似度Sim效果，同时分析K值影响。测试结果如表1所示。

表1 2种相似度K值误差对比

Table 1 Comparison of errors of two similarity K values

K	平均值/m		最大值/m		最小值/m		误差方差/m^2	
	Euc	Sim	Euc	Sim	Euc	Sim	Euc	Sim
4	**0.958**	0.960	5.823	**4.887**	0.004	0.004	0.934	**0.823**
5	0.932	**0.910**	6.124	**5.823**	0.004	0.004	0.865	**0.976**
6	0.913	**0.878**	4.980	**3.882**	0.004	0.004	0.764	**0.664**
7	0.953	**0.948**	**7.162**	7.486	0.004	0.004	**0.967**	0.987
8	1.135	**0.988**	8.105	**7.486**	0.004	0.004	1.646	**1.334**

由表1可以看出，在不同的K值下本文所提相似度的整体误差波动水平表现均优于Euc。当K值由小到大增加时，平均定位误差呈先下降后上升的趋势，并于$K=6$处取得最小值。这是因为K值差异对子指纹库中AP和RP数量的影响，K较小时，信号区域分布特性不同的AP可能被分入同一子区域，带来的信息交杂冗余会提升RP误判的可能性。而K值较大时，区域数量增加，同时子指纹库的个体规模均过小，目标位置周围的邻近RP信息被撕裂至多个区域，降低了位置估算时信息的丰富度。因此需要均衡子指纹库缩减规模和定位信息裕量，后续实验中均取定位误差最小时参数$K=6$。

7.3 信号距离插值分析

7.3.1 信号拟合精度

为了验证本文所提信号距离加权插值的信号拟合效果，选取反距离加权插值IDW[3]、基于克里金插值的I-Krige[5]和自适应路径损耗模型的PL[4]4种算法进行对比分析。一般情况下，实采RP数量越少，扩充的指纹库定位精度越差，选取不同参考点抽取比例为$h=0.25$、$h=0.45$和$h=0.75$时扩充至完整指纹库对比4种算法的平均拟合误差，为避免抽取随机性对误差估计的影响，每次实验均重复20次，结果取均值。

图8为4种算法的信号拟合效果，随着抽取比例的增长，提供的原始指纹信息增多，4种算法的平均拟合误差都在下降。而本文所提信号距离加权插值的拟合误差始终优于其他3种算法，随抽取比例增加分别为2.42、1.96和1.32。IDW将空间距离作为权重赋予所有AP，没有考虑AP信号分布的各向差异，拟合误差始终最大。I-Krige考虑了空间整

体特征的影响, 拟合精度整体较高, 但信号在定位区域无法保证一致性, 导致变异函数拟合精度较低从而影响插值精度。PL 需要精准刻画信号在空间中的传播特性, 实际环境中障碍物和衰弱效应使估算距离存在较大误差。

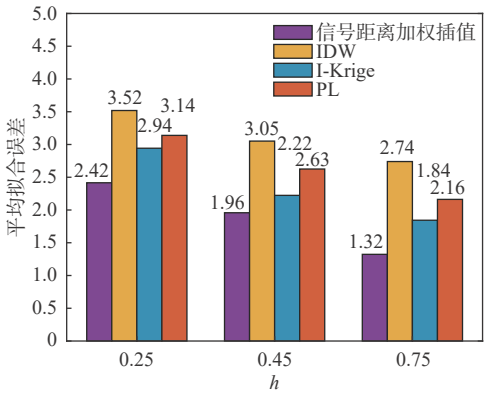


图 8　4 种插值算法的平均拟合误差

Fig. 8　Average fitting error of four interpolation algorithms

7.3.2　插值效果比较

为进一步分析本文所提信号距离加权插值算法整体对定位精度的影响, 选取比例为 $h=0.25$、$h=0.45$和$h=0.75$时信号距离加权插值所得指纹库与实采指纹库比较定位性能, 如图 9 所示。同时选取经典扩充算法 I-Krige、PL 和 IDW, 折中抽取 $h=0.45$的参考点扩充指纹库至原始规模, 定位误差对比如图 10 所示。实验均采用本文所提在线定位算法进行邻近点搜索和定位, 以定位的累积误差分布分析插值算法的性能。

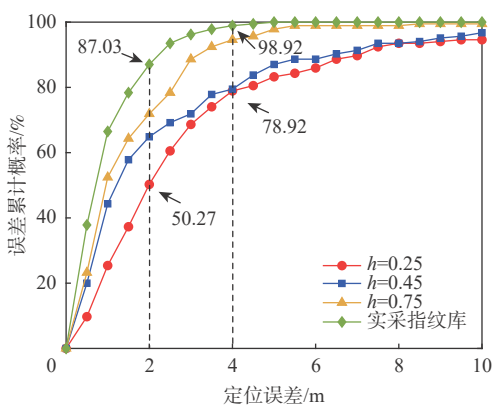


图 9　不同插值比例的累计误差分布

Fig. 9　Cumulative error distribution of different interpolation ratios

由图 9 可知, 随抽取比例不断增加, 信号距离加权插值所得指纹库的定位信息更加接近实采指纹库, 整体定位效果逐渐上升。图 10 中, 本文所提算法定位精度始终优于其他 3 种算法。在 2 m 的误差范围之内数据比例达到了 64.86%, 高出 PL 算法达 27.54%, 在 6.5 m 时突破 91.35%, 对比其他算法仍具有相当优势。

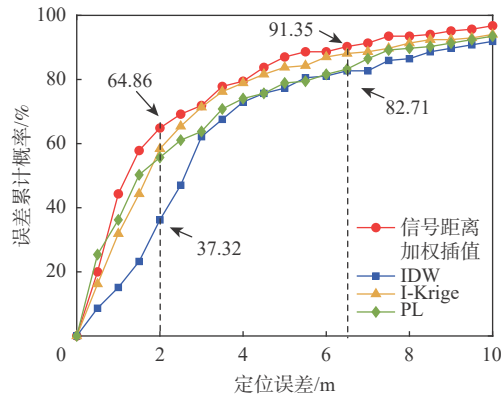


图 10　4 种插值算法的累计误差分布

Fig. 10　Cumulative error distribution of four interpolation algorithms

表 2 给出了 4 种插值算法和不同抽取比例时信号距离加权插值的定位误差对比, 信号距离加权插值的平均定位误差、最大值和误差方差均优于其他 3 种, 在插值比例$h=0.25$时与 PL 在$h=0.45$时定位精度相当, 可用更少的实采信息恢复指纹库。同时定位精度和稳定性方面表现优秀, 相对于 I-Krige 提升了 17.4%, 表明了信号距离加权插值不仅提升了指纹建库效率, 同时具备优异的定位性能。

表 2　不同插值算法的定位误差对比

Table 2　Comparison of localization errors of different interpolation algorithms

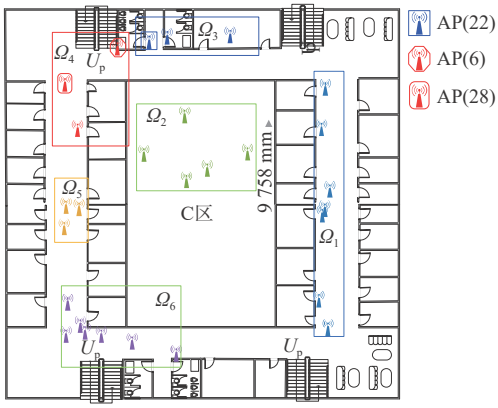
插值算法	h	平均值/m	最大值/m	最小值/m	方差/m^2
实采指纹库		**0.834**	**3.522**	**0.001**	**0.677**
	0.25	3.522 7	39.95	0.104	5.410
信号距离加权插值	0.45	2.467	18.59	0.017	2.792
	0.75	1.504	12.50	0.082	1.521
IDW	0.45	3.978	38.45	0.068	4.894
I-Krige	0.45	3.253	26.32	0.015	4.135
PL	0.45	3.477	30.77	0.023	3.958

7.4　双维度划分分析

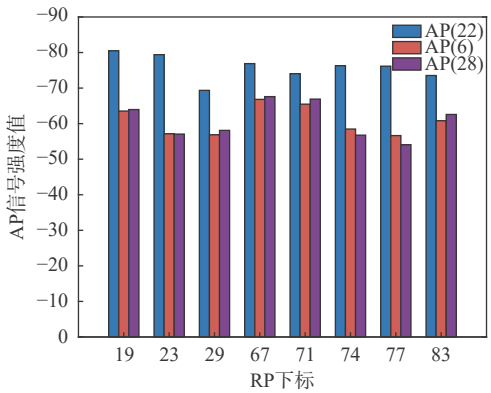
7.4.1　AP 集团划分效果

基于 7.2.3 节的实验结果, 本文将所有虚拟 AP 位置划分得到 6 个区域 AP 集团。信号分布相似度的引入使得 AP 的聚类标准不仅仅限于物理空间的邻近性, 利用皮尔逊相关系数, 信号域联系程度更高的 AP 被判定至一起的概率会得到提升, 如图 11(a)中 AP(6)与 AP(22)在空间上更邻近但却和信号分布更接近的 AP(28)分至同一区域。

而图 11(b) 中可以看到 3 个 AP 在随机抽取的 RP 集合中的 RSSI 分布, 很明显 AP(6)与 AP(28)在定位区域中的信号分布更为相似, 因此在区域归属判定时受到信号相关性的吸引同被聚类至$\boldsymbol{\Omega}_4$。



(a) 区域AP集团分布



(b) 3个AP在部分RP处的信号分布

图 11 区域 AP 集团效果

Fig. 11 Regional AP group effect

7.4.2 聚类效果分析

按 4.3 节的聚类方法，以 AP 集团作为锚点，在定位场景中形成了 6 个子区域，区域标签则分别按 AP 集团的标识作为编号。

如图 12 所示，在走廊西侧形成了 $\boldsymbol{\Omega}_4$、$\boldsymbol{\Omega}_5$和 $\boldsymbol{\Omega}_6$3 个子区域，参考图 11(a)，可知 3 个区域内的 AP 位置都相对集中，且信号热力分布差异明显，因此 RP 基本围绕 AP 集团形成了界限分明的 3 个子区域。而在走廊东侧，探测到的 AP 位置相对分散，在同一 RP 来自不同 AP 的信号辐射强度存在高低差异，在均衡考量下都被分入了 $\boldsymbol{\Omega}_1$。在东南侧角落，具有较强信号的 AP 数量稀少，同时整体覆盖强度比较弱，因此被归入相对信号较强的区域 $\boldsymbol{\Omega}_2$内。

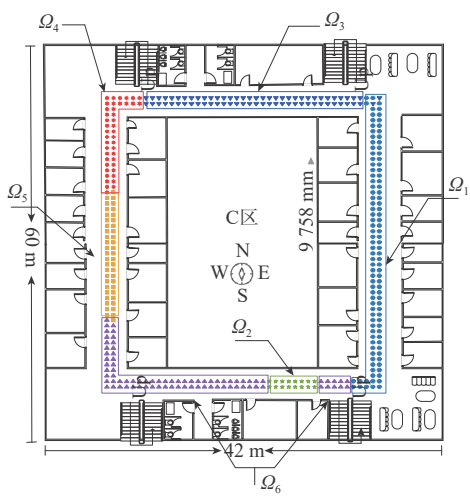


图 12 双维度聚类效果

Fig. 12 Dual-dimensional clustering effect

7.5 定位算法性能比较

7.5.1 近邻点个数

为探究近邻点数量 C 对定位效果的影响，本文选取同样以近邻点集匹配为特征定位的 AAS[11]、WKNN 和改进的加权 K 近邻算法[10]，简称 IWKNN，在离线阶段选取了 20% 的参考点作为测试集，其余作为训练集，经过 10 次实验重复求取每次的平均定位误差探究邻近点数量对定位性能的影响。

图 13 给出了 4 种算法在 $C \in [1,10]$时的平均误差波动，可以看出本文所提算法 DDPL 相较于其余算法保持了最低定位误差的优势。从整体看，WKNN 的定位误差最高，在邻近 RP 数 $C=5$时平均定位误差最低，然而仍比 DDPL 高出 0.13 m。AAS 和 IWKNN 分别在 $C=4$和 $C=6$时达到最低点，定位误差相对 DDPL 存在 0.12 m 和 0.06 m 的涨幅。

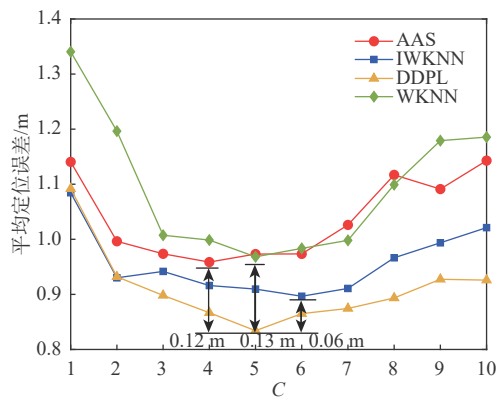


图 13 邻近点数量 C 与平均定位误差关系

Fig. 13 Relationship between number of adjacent points C and average localization error

7.5.2 整体性能比较

为了验证本文所提算法 DDPL 的整体定位性能，利用测得数据得到估算位置，将其与真实位置的物理距离差值作为定位误差，同 IWKNN、AAS 以及 WKNN 进行精度比较。为保证算法公平性，4 种算法均采取了各自的最佳参数，DDPL 中 $K=6$，$\gamma=0.45$，$C=5$；AAS 取 $C=4$；IWKNN 取 $C=6$；WKNN 取 $C=5$。

图 14 给出了 4 种算法的累积误差概率分布，本文所提算法在误差容忍 0.5 m 以内的范围中表现略次于其他 3 种算法。这是因为本文匹配算法是以 AP 虚拟位置为基础计算传播距离，由于环境的干扰估算位置并不完全准确，部分参考点的权重计算

存在偏差, 所以低定位误差数据占比较低。然而整体上本文算法均优于其他 3 种, 在 1 m 内误差占比达到了 71.9%, 在 2 m 时更是达到 93.5%, 相对于其他算法具有明显的提升。

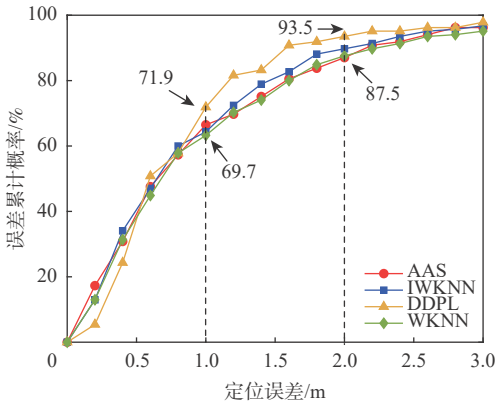


图 14　4 种定位算法的累计误差概率分析

Fig. 14　Cumulative error probability analysis of four localization algorithms

表 3 给出了 4 种算法包括平均定位误差、最小定位误差、最大定位误差和误差方差的对比。可以看出, 本文所提 DDPL 不仅在平均定位误差拥有最低值 0.834 m, 其余各个误差指标也优于另外 3 种算法。离线阶段通过区域特征划分 AP 集团降低子指纹库规模的同时也避免在线定位冗余 AP 的干扰, 而利用位置距离和的差异筛选则进一步优化了近邻点 RP 集, 多重优化措施保证了定位精度的提升。

表 3　4 种算法定位误差对比

Table 3　Comparison of localization errors of four algorithms

定位算法	平均值/m	最大值/m	最小值/m	方差/m^2
AAS	0.959	4.795	0.015	0.886
WKNN	0.968	5.411	0.064	0.921
IWKNN	0.897	4.199	0.041	0.838
DDPL	**0.834**	**3.522**	**0.001**	**0.677**

相对于全局匹配的 WKNN 算法, 本文将定位场景中的$N=368$个参考点降低至子区域 16～112 个参考点, 以子区域的局部搜索代替全局搜索提升了定位效率。而对比基于聚类分区的定位算法如 DDPL 等, 算法仅保留子区域 AP 集团信息, 降低 AP 维度至 3～ 8 个, 进一步降低了在线定位的搜索量, 同时匹配算法基于离线数据优化近邻点 RP 集引入的计算量相对可控, 因此算法的执行效率也表现优异。

8　结　论

1) 环境干扰产生的信号波动, 使得 AP 的位置估算产生偏移得到虚拟位置。通过几何原理消除小范围内环境变量, 并剔除离群估算位置, 所得虚拟 AP 更符合信号真实特性。

2) 通过分析 AP 信号特征, 完成 AP 和 RP 的双维度划分。得到小规模高质量的子指纹库同时提高了在线匹配效率。

3) 信号距离兼顾信号在空间域和信号域的传播特性, 插值结果稳定有效。实验表明, 插值精度相较于 I-Krige 提升了 17.4%。

4) 在线匹配算法通过区域阈值灵活调整待选 RP 信息裕量, 利用传播相似度高效检索邻近 RP。定位精度相较于 IWKNN 和 WKNN 分别提高了 7.1% 和 13.8%。

5) 算法中包括 AP 子集团划分方式、插值算法性能等环节的优化将作为后续研究的重点方向。

参考文献（References）

[1] KIM GEOK T, ZAR AUNG K, SANDAR AUNG M, et al. Review of indoor positioning: radio wave technology[J]. Applied Sciences, 2021, 11(1): 279.

[2] NINH D B, HE J, TRUNG V T, et al. An effective random statistical method for Indoor Positioning System using WiFi fingerprinting[J]. Future Generation Computer Systems, 2020, 109: 238-248.

[3] ISMAIL A H, MIZUSHIRI Y, TASAKI R, et al. A novel automated construction method of signal fingerprint database for mobile robot wireless positioning system[J]. International Journal of Automation Technology, 2017, 11(3): 459-471.

[4] BI J X, WANG Y J, LI Z K, et al. Fast radio map construction by using adaptive path loss model interpolation in large-scale building[J]. Sensors, 2019, 19(3): 712.

[5] 刘辉元, 马金辉, 黄琼. 基于改进克里金插值的室内定位位置指纹库构建方法[J]. 重庆邮电大学学报 (自然科学版), 2017, 29(6): 751-757.
LIU H Y, MA J H, HUANG Q. Construction method of fingerprint database based on improved Kriging interpolation for indoor location[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2017, 29(6): 751-757(in Chinese).

[6] ZHANG C B, QIN N N, XUE Y B, et al. Received signal strength-based indoor localization using hierarchical classification[J]. Sensors, 2020, 20(4): 1067.

[7] 李昂, 付敬奇, 沈华明, 等. 基于模糊聚类和猫群算法的室内定位算法[J]. 仪器仪表学报, 2020, 41(1): 185-194.
LI A, FU J Q, SHEN H M, et al. Indoor positioning algorithm based on fuzzy clustering and cat swarm optimization[J]. Chinese Journal of Scientific Instrument, 2020, 41(1): 185-194(in Chinese).

[8] CUI H M, LIU K W. Indoor positioning and fingerprint updating based on affinity propagation clustering[C]//2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC). Piscataway: IEEE Press, 2018: 226-230.

[9] SHIN B, LEE J H, LEE T, et al. Enhanced weighted K-nearest neighbor algorithm for indoor Wi-Fi positioning systems[C]//2012 8th International Conference on Computing Technology and In-

formation Management (NCM and ICNIT). Piscataway: IEEE Press, 2012: 574-577.

[10] 王博远, 刘学林, 蔚保国, 等. WiFi 指纹定位中改进的加权 k 近邻算法[J]. 西安电子科技大学学报, 2019, 46(5): 41-47.
WANG B Y, LIU X L, YU B G, et al. Improved weighted k-nearest neighbor algorithm for WiFi fingerprint positioning[J]. Journal of Xidian University, 2019, 46(5): 41-47(in Chinese).

[11] TAO Y, ZHAO L. Fingerprint localization with adaptive area search[J]. IEEE Communications Letters, 2020, 24(7): 1446-1450.

[12] 薛卫星, 李清泉, 周宝定. 虚拟 AP 支持的室内非线性约束区域定位[J]. 测绘学报, 2021, 50(5): 569-579.
XUE W X, LI Q Q, ZHOU B D. Virtual AP based indoor localization in area without linear constraints[J]. Acta Geodaetica et Cartographica Sinica, 2021, 50(5): 569-579(in Chinese).

[13] XU F, HU X K, LUO S W, et al. An efficient indoor Wi-Fi positioning method using virtual location of AP[J]. ISPRS International Journal of Geo-Information, 2020, 9(4): 261.

[14] XIAO J X, LIAO B W, LI C G. Improved indoor positioning based on fingerprint algorithm using reference point subregion and access point optimization[C]//2019 IEEE 19th International Conference on Communication Technology (ICCT). Piscataway: IEEE Press, 2019: 461-468.

[15] BI J X, HUANG L, CAO H J, et al. Improved indoor fingerprinting localization method using clustering algorithm and dynamic compensation[J]. ISPRS International Journal of Geo-Information, 2021, 10(9): 613.

[16] SOLAHUDDIN Y F, MARDENI R. Indoor empirical path loss prediction model for 2.4 GHz 802.11n network[C]//2011 IEEE International Conference on Control System, Computing and Engineering. Piscataway: IEEE Press, 2011: 12-17.

Dual-dimensional partition localization algorithm under AP virtual location perception

ZHANG Yang，QIN Ningning*

(Engineering Research Center of Internet of Things Technology Applications of Ministry of Education，Jiangnan University，Wuxi 214122，China)

Abstract：To balance the efficiency and cost of offline database construction, as well as the efficiency and accuracy of online matching and localization in WiFi fingerprint localization, this paper proposed a dual-dimensional partition localization (DDPL) algorithm based on wireless access point (AP) virtual location perception. The algorithm used geometric characteristics and propagation models to perceive AP virtual locations based on actual measurement data. In the offline phase, the spatial weight was replaced by the signal distance weight that conformed to the signal characteristics for the fingerprint database interpolation expansion, which effectively reduced the cost of database construction. At the same time, the dual-dimensional division of the fingerprint database was simplified based on the signal coverage characteristics, and the high-value localization information in each region was retained. In the online phase, the differential evaluation criteria of propagation similarity and distance sum were introduced, and the adjacent reference points (RPs) were searched efficiently and finely. In the measured scene, the proposed fingerprint localization algorithm could take into account both quality and efficiency, and the interpolation effect was improved by more than 17.4% compared with other schemes. The accuracy of the localization algorithm was at least 7.1% higher than that of similar algorithms, which had a certain application value in indoor scenes.

Keywords：indoor localization；virtual AP；fingerprint database expansion；partition；access point；reference point

Received：2023-05-24；**Accepted**：2023-09-01；**Published Online**：2023-10-18 11：58

URL：link.cnki.net/urlid/11.2625.V.20231017.1417.002

Foundation items：Yangtze River Delta Sci-Tech Innovation Community Joint Research Project (2023CSJGG1700); National Natural Science Foundation of China (61702228); China University Industry, University and Research (IUR) Innovation Fund (2021ITA10003)

* **Corresponding author**. E-mail：qinnn@jiangnan.edu.cn

《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51卷 第5期　2025年5月

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)
(Monthly, Started in 1956)
Vol.51 No.5　May 2025

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
9 771001 596250 05>

微信公众号　超星电子书　QQ群

国内定价：50.00元／期