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

2022-11
Vol.48 No.11

北京航空航天大学学报

第 48 卷　第 11 期　(总第 357 期)　2022 年 11 月

目　　次

抵御对抗样本攻击的指纹室内定位方法 … 张学军，鲍俊达，何福存，盖继扬，田丰，黄海燕 (2087)
微小卫星隐身构型设计及优化分析 …… 秦远田，孙汗青，岳鑫 (2102)
高速气膜镶装式浮环密封的开启特性 …… 郑娆，陈潇竹，李双喜，赵祥，施任杰，宋子锋 (2111)
基于径向基函数代理模型的 M 形杆刚度优化 …… 杨慧，范硕硕，王岩，刘荣强，肖洪 (2121)
高超声速飞行器命中精度高准确度评估 …… 彭维仕 (2130)
基于 PPO 的移动平台自主导航 …… 徐国艳，熊绎维，周彬，陈冠宏 (2138)
NO 快速检测光腔结构设计与光路仿真 …… 李文，蔡永青，陈梦凡，刘鹏 (2146)
冰雪天气下基于 MFOA-KELM 残差修正的跑道温度混合预测 ……
…… 陈斌，刘悦，李庆真，丁宇，王立文 (2153)
基于燃料电池的复合电源式装载机分层控制 …… 张志文，杜文杰，梁君飞，张艳岗，武雅文 (2165)
基于 CC1310 芯片的多通道高速率低功耗无线传感系统 …… 段瑞枫，吕燕洁，杜文基，周游 (2177)
基于注意机制的轻量化稠密连接网络单幅图像去雨 …… 柴国强，王大为，芦宾，李竹 (2186)
GPS P(Y)码功率增强监测分析 …… 李文玄，焦文海，王凯，邱瑞瑾，孙淑贤 (2193)
基于 ESO 的超局部无模型 PMSM 转速预测控制 …… 许令亮，陈桂明，李乔扬 (2204)
星载原子钟在轨剩余寿命预测方法研究 …… 凤建广，郑紫霞，龙东腾，周波，陆明泉，郑恒 (2215)
高精度球面洛伦兹力磁轴承设计与分析 … 傅百恒，王卫杰，王元钦，樊亚洪，聂辰，贾海鹏 (2222)
旋转抛物-锥形蚀坑模型及其应用 …… 刘德俊，田干，金国锋，杨正伟，任碧云，魏花丽 (2230)
基于群体意志统一的无人机协同围捕策略 …… 刘峰，魏瑞轩，周凯，丁超 (2241)
基于无监督学习和粒子滤波的非视距信号检测 …… 侯宁宁，李灯熬，赵菊敏 (2250)
飞机旋转表面结冰数值模拟 …… 郭琦，申晓斌，林贵平，张世娟 (2259)
基于凸优化和 LQR 的火箭返回轨迹跟踪制导 …… 吴杰，张成，李淼，熊芬芬 (2270)
一种基于激光位移测量的机械臂参数标定方法 …… 邵鑫，季力，邹怀武，解杨敏 (2281)
基于障碍物代价势场的移动机器人避障算法 …… 迟胜凯，谢永芳，陈晓方，彭帆 (2289)
数据立方体关联挖掘在红外抗干扰中的应用 … 吴鑫，伍友利，牛得清，蔡宇轩，徐洋，陈鞭 (2304)
基于混合稀疏表示的二维压缩感知 SAR 成像 …… 熊世超，倪嘉成，张群，罗迎，王岩松 (2314)
基于 SOM 和关联规则的民机运行风险 …… 熊明兰，王华伟，倪晓梅，蔺瑞管 (2325)
基于优化 CenterNet 的低空无人机检测方法 …… 张瑞鑫，黎宁，张夏夏，ZHOU Huiyu (2335)

期刊基本参数：CN 11-2625/V * 1956 * m * A4 * 258 * zh * P * ￥50.00 * 400 * 26 * 2022-11

(**编　辑**　张　嵘　孙　芳　卞欢欢　王　茜　李亚泰　李艳霞　苏　磊)

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

Vol. 48 No. 11 (Sum 357) November 2022

CONTENTS

A fingerprint indoor localization method against adversarial sample attacks
…… ZHANG Xuejun, BAO Junda, HE Fucun, GAI Jiyang, TIAN Feng, HUANG Haiyan (2087)
Stealthy configuration design and optimization analysis of microsatellite
…… QIN Yuantian, SUN Hanqing, YUE Xin (2102)
Opening characteristics of inlaid floating ring seal with high-speed gas film
…… ZHENG Rao, CHEN Xiaozhu, LI Shuangxi, ZHAO Xiang, SHI Renjie, SONG Zifeng (2111)
Stiffness optimization of M-shaped boom based on radial basis function surrogate model
…… YANG Hui, FAN Shuoshuo, WANG Yan, LIU Rongqiang, XIAO Hong (2121)
Evaluation of high hitting accuracy performance of hypersonic vehicle
…… PENG Weishi (2130)
Autonomous navigation based on PPO for mobile platform
…… XU Guoyan, XIONG Yiwei, ZHOU Bin, CHEN Guanhong (2138)
Optical path simulation and design of NO rapid detection optical cavity structure
…… LI Wen, CAI Yongqing, CHEN Mengfan, LIU Peng (2146)
Runway temperature hybrid prediction based on MFOA-KELM residual correction under ice and snow
…… CHEN Bin, LIU Yue, LI Qingzhen, DING Yu, WANG Liwen (2153)
Layered control of hybrid power loader based on fuel cell
…… ZHANG Zhiwen, DU Wenjie, LIANG Junfei, ZHANG Yangang, WU Yawen (2165)
Multi-channel wireless sensor system based on CC1310 chip with high speed and low power consumption
…… DUAN Ruifeng, LYU Yanjie, DU Wenji, ZHOU You (2177)
Lightweight densely connected network based on attention mechanism for single-image deraining
…… CHAI Guoqiang, WANG Dawei, LU Bin, LI Zhu (2186)
Monitoring and analysis on GPS P(Y) code power enhancement
…… LI Wenxuan, JIAO Wenhai, WANG Kai, QIU Ruijin, SUN Shuxian (2193)
Ultra-local model-free speed predictive control based on ESO for PMSM
…… XU Lingliang, CHEN Guiming, LI Qiaoyang (2204)
Method for predicting on-orbit residual life of satellite atomic clock
…… FENG Jianguang, ZHENG Zixia, LONG Dongteng, ZHOU Bo, LU Mingquan, ZHENG Heng (2215)
Design and analysis of high precision for spherical Lorentz force magnetic bearing
…… FU Baiheng, WANG Weijie, WANG Yuanqin, FAN Yahong, NIE Chen, JIA Haipeng (2222)
Rotating parabolic-conical corrosion pit model establishment and its application
…… LIU Dejun, TIAN Gan, JIN Guofeng, YANG Zhengwei, REN Biyun, WEI Huali (2230)
Multi-UAV round up strategy based on unity of group will
…… LIU Feng, WEI Ruixuan, ZHOU Kai, DING Chao (2241)
Non-line-of-sight signal detection based on unsupervised learning and particle filtering
…… HOU Ningning, LI Deng'ao, ZHAO Jumin (2250)
Numerical simulation of icing on aircraft rotating surfaces
…… GUO Qi, SHEN Xiaobin, LIN Guiping, ZHANG Shijuan (2259)
Rocket return trajectory tracking guidance based on convex optimization and LQR
…… WU Jie, ZHANG Cheng, LI Miao, XIONG Fenfen (2270)
A parameter calibration method for manipulators based on laser displacement measurement
…… SHAO Xin, JI Li, ZOU Huaiwu, XIE Yangmin (2281)
Obstacle avoidance method of mobile robot based on obstacle cost potential field
…… CHI Shengkai, XIE Yongfang, CHEN Xiaofang, PENG Fan (2289)
Application of data cube association mining in infrared anti-jamming
…… WU Xin, WU Youli, NIU Deqing, CAI Yuxuan, XU Yang, CHEN Bian (2304)
2-D compressed sensing SAR imaging based on mixed sparse representation
…… XIONG Shichao, NI Jiacheng, ZHANG Qun, LUO Ying, WANG Yansong (2314)
Operation risk of civil aircraft based on SOM and association rules
…… XIONG Minglan, WANG Huawei, NI Xiaomei, LIN Ruiguan (2325)
Low-altitude UAV detection method based on optimized CenterNet
…… ZHANG Ruixin, LI Ning, ZHANG Xiaxia, ZHOU Huiyu (2335)

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0789

抵御对抗样本攻击的指纹室内定位方法

张学军[1,*], 鲍俊达[1], 何福存[1], 盖继扬[1], 田丰[2], 黄海燕[1]

(1. 兰州交通大学 电子与信息工程学院, 兰州 730070; 2. 陕西师范大学 计算机科学学院, 西安 710062)

摘 要: 随着城市智能化的发展,基于 WiFi 接收信号强度(RSS)的指纹室内定位服务受到社会的广泛关注。深度学习技术是利用 RSS 信号获得高室内定位性能的一种重要手段,但其易遭受对抗样本攻击,给定位系统带来严重安全隐患。为此,提出了一种抵御对抗样本攻击的基于深度学习的 RSS 指纹室内定位方法(AdvILoc)。该方法基于图像识别领域对抗样本防御方法的研究和分析,结合室内 RSS 指纹数据特征单一且高维的特点,通过在 RSS 指纹室内定位深度学习模型中添加池化层、全连接层,以及满足差分隐私的噪声层来抵御对抗样本攻击,解决了基于深度学习的室内定位模型易过拟合且泛化能力不高的问题。通过添加 Dropout 层,以及设计模型参数正则化方法,提高模型抵御对抗样本攻击的鲁棒性。在 2 个真实 RSS 指纹室内定位数据集上的实验结果表明:与已有基于多层感知机(MLP)、卷积神经网络(CNN)的 RSS 指纹室内定位方法相比,所提方法在保证时间开销和基本不影响定位模型性能的情况下,提高了模型抵御对抗样本攻击的鲁棒性;在满足 l_2 范式规范的 C&W 攻击下,随着攻击大小不断增大,模型的定位准确率下降也更平稳。

关 键 词: 室内定位; 对抗样本; 深度学习; C&W 攻击; 差分隐私

中图分类号: TP309.2

文献标志码: A **文章编号**: 1001-5965(2022)11-2087-15

随着物联网和无线通信技术的蓬勃发展,室内位置服务已成为大众生活中不可或缺的一部分,如室内导航、健康感知、活动识别、灾害管理与救援等[1]。目前,众多室内定位技术利用无线局域网的相关特征来进行定位,常用的技术有 WiFi[2]、蓝牙[3]、红外线[4]和超声设备[5]等。Bahl 和 Padmanabhan[6]提出了一种基于无线射频信号的用户定位与跟踪系统 RADAR,能够为用户提供高精度的室内定位服务。然而,室内环境的复杂性(如物体遮挡、人员流动和电子设备干扰等)导致射频信号无法覆盖全区域,很难向所有用户提供同等质量的定位服务。得益于基础设施成熟、易部署、低成本和低功耗等优势,基于接收信号强度(received signal strength, RSS)的指纹室内定位技术成为主流[1-2],其核心是发现指纹信号强度和位置之间的映射关系。但是,基于 RSS 的指纹室内定位技术易受室内多径效应、衍射等影响,仅能提供 6~8 m 的定位精度。深度学习(deep learning, DL)利用其强大的特征抽取和分类能力有效获取指纹信号强度和位置之间的映射关系,克服指纹定位受室内环境影响而性能下降的缺陷,在指纹室内定位技术中得到了广泛应用[7-8]。

基于深度学习的 RSS 指纹室内定位方法使

收稿日期: 2021-12-28; **录用日期**: 2022-03-11; **网络出版时间**: 2022-04-11 08:31

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220408.1225.001.html

基金项目: 国家自然科学基金(61762058,61901201); 兰州交通大学百人青年人才培养计划; 甘肃省自然科学基金(21JR7RA282); 甘肃省高等学校产业支撑计划(2022CYZC-38); 中央高校基本科研业务费专项资金(GK202103090); 陕西省自然科学基础研究计划(2022JM-329)

* **通信作者**. E-mail: xuejunzhang@mail.lzjtu.cn

引用格式: *张学军, 鲍俊达, 何福存, 等. 抵御对抗样本攻击的指纹室内定位方法[J]. 北京航空航天大学学报, 2022, 48(11): 2087-2101. ZHANG X J, BAO J D, HE F C, et al. A fingerprint indoor localization method against adversarial sample attacks[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2087-2101(in Chinese).*

用包含定位终端和定位云服务器的云架构来实现室内定位[9]。Hsieh 等[7]提出了基于 RSS 指纹数据和信道状态信息(channel state information, CSI)的深度学习室内定位方法,将一个矩形室内房间平面划分成若干区域,每个子区域都被视为一个类,从而将室内定位服务形式化为一个深度学习分类问题,并使用多层感知机(multi-layer perception, MLP)和卷积神经网络(convolutional neural network, CNN)估计用户在房间中的位置,该方法在保证网络规模轻量化的同时取得了较好的定位性能。Wang 等[8]提出了一个基于深度卷积神经网络(deep convolutional neural networks, DCNN)的商用 5 GHz WiFi 室内定位方法 CiFi,先利用 5G WiFi 的 CSI 相位差数据估计到达角,并用到达角构建了 16 个大小为 60 × 60 的图像,再将这 16 个图像输入 DCNN 来训练离线阶段的权重,大大提高了室内定位模型的性能。但是,云作为一个不可信的实体,可能会获知深度学习模型的相关参数,生成最优的对抗性噪声,并将其添加至用户提交的 RSS 指纹数据中,导致深度学习分类模型返回错误的定位服务,使得定位系统服务质量下降,甚至失效。例如,Patil 等[10]分别通过快速梯度符号攻击、投影梯度下降攻击和动量迭代攻击对构建的室内定位深度学习模型进行主动攻击,验证了对抗样本攻击极大地降低了楼层分类和室内定位预测的性能,但没有给出防御方法。

目前,在计算机视觉、网络流量分析等领域,学者们针对对抗样本攻击提出了对抗训练、防御蒸馏、对抗样本检测及强化网络等防御方法[11],但缺乏专门针对室内定位领域的防御方法。刘西蒙等[12]对深度学习中的代表性对抗攻击及典型的防御策略进行了详细分析和总结,指出目前该领域仍没有一个能有效抵御对抗样本攻击的防御策略。王滨等[13]针对 CNN 网络流量分类器遭受对抗样本攻击的问题,提出了一种批次对抗训练的方法,利用训练过程反向传播误差的特点,在一次反向传播过程中同时完成样本梯度和参数梯度的计算,抵御白盒攻击的同时提高了训练效率。Xie 等[14]提出在模型前向传播时利用随机化来抵御对抗样本攻击,如采用随机调整和随机填充等,取得了一定的防御效果。Liu 等[15]结合模型融合与随机化提出了随机自组装(random self-ensemble, RSE)的防御方法,在神经网络模型中加入随机噪声层,并将多个随机噪声的预测结果融合在一起来增强模型鲁棒性。基于 RSE 防御方法,Lecuyer 等[16]提出了基于差分隐私(differential privacy, DP)的隐藏式安全防御机制 PixelDP,该机制在深度神经网络模型中加入满足 DP 的噪声层,随机化网络的计算结果,使对抗样本中满足 l 范式的扰动对模型输出的影响保持在 DP 稳定输出期望界限范围内。虽然上述方法能在一定程度上防御对抗样本攻击,但是其在面对数据特征单一、高维的 RSS 指纹室内定位数据时,容易出现过拟合、学习能力低等问题,导致定位模型在实时定位阶段的泛化能力和鲁棒性差,难以向用户提供可靠、稳定的定位服务。

针对以上问题,本文提出了一种抵御对抗样本攻击的基于深度学习的 RSS 指纹室内定位方法(deep learning based fingerprint indoor localization method against adversarial samples attack, AdvILoc)。AdvILoc 通过 DP 技术,在室内定位深度学习网络中加入满足 ε-DP 的噪声层,随机化模型泛化的计算结果,为用户提供满足 ε-DP 且经过鲁棒性认证的室内定位服务。同时,通过在全连接层后添加 Dropout 层,随机失活全连接层中部分神经元,削弱神经元节点间的联合适应性,在训练过程中正则化模型损失参数,避免模型过拟合、学习能力低等问题,提高了模型鲁棒性。实验验证了在实时定位阶段,AdvILoc 在抵御对抗样本攻击的同时,仍能为用户提供高效的室内定位服务。

1 预备知识

1.1 对抗样本攻击

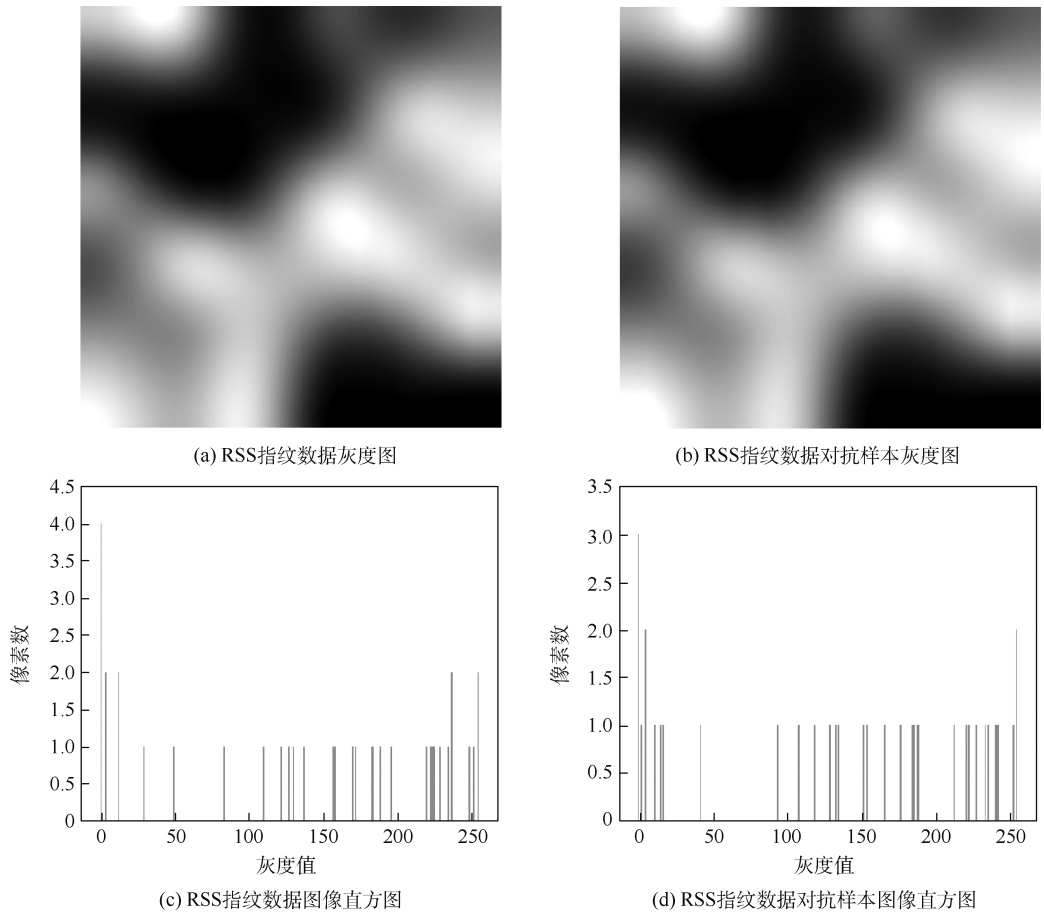
对抗样本攻击[17]是针对机器学习模型的一类主动攻击,通过故意在输入数据中添加人类肉眼无法察觉的扰动生成输入数据对抗样本,使模型以高置信度输出一个错误的分类结果。

C&W 攻击[18]是对抗样本攻击中最常用的对抗样本生成算法。针对室内定位深度学习模型,该算法通过求解最优化算式(1)生成对抗样本攻击。

$$\begin{aligned}&\min \|\boldsymbol{x}-\boldsymbol{x}'\|_2, \boldsymbol{x}' \in [0,1]^n\\&\text{s.t.}\quad C(\boldsymbol{x})=\boldsymbol{y}, C(\boldsymbol{x}')=\boldsymbol{y}' \text{ and } \boldsymbol{y}' \neq \boldsymbol{y}\end{aligned} \tag{1}$$

式中:$\boldsymbol{x}$ 为原始 RSS 指纹数据(灰度图见图 1(a));$\boldsymbol{x}'$为对应 $\boldsymbol{x}$ 生成的 RSS 指纹数据对抗样本(灰度图见图 1(b));$\boldsymbol{y}$ 为正确标签;$\boldsymbol{y}'$为目标标签;C 为训练好的室内定位 CNN 分类器。

式(1)表明,恶意攻击者依据 C 的相关参数,利用原始 RSS 指纹数据 $\boldsymbol{x}$ 生成其对抗样本 $\boldsymbol{x}'$,同



(a) RSS指纹数据灰度图　(b) RSS指纹数据对抗样本灰度图

(c) RSS指纹数据图像直方图　(d) RSS指纹数据对抗样本图像直方图

图 1　RSS 指纹数据 $\boldsymbol{x}$ 与其经过 C&W 攻击生成的对抗样本 $\boldsymbol{x}'$ 的灰度图及其直方图

Fig. 1　Grayscale and histogram of RSS fingerprint data $\boldsymbol{x}$ and its' samples $\boldsymbol{x}'$ generated by C&W attack

时能确保 $\boldsymbol{x}'$ 与 $\boldsymbol{x}$ 之间存在难以察觉的细微扰动,但经过 C 的泛化可得到不同的定位结果 $\boldsymbol{y}'$ 与 $\boldsymbol{y}$。显然,在图 1(a)、(b)中,无法用肉眼看出 $\boldsymbol{x}'$ 与 $\boldsymbol{x}$ 之间的差别。为此,图 1(c)、(d)给出了 $\boldsymbol{x}$ 与 $\boldsymbol{x}'$ 的图像直方图来观察其差异性,可以看出,$\boldsymbol{x}$ 和经过 C&W 生成的对抗样本 $\boldsymbol{x}'$ 的数据分布明显不同,将 $\boldsymbol{x}'$ 与 $\boldsymbol{x}$ 输入 C 中会得到不同的分类结果,后续实验结果也验证了 2 张 RSS 指纹数据灰度图像经模型泛化后得到了不同的定位结果。

1.2　鲁 棒 性

如果一个多分类深度学习模型 $f:\mathbf{R}^n\to\kappa$ 的预测结果对部署中可能遇到的一些看似合理输入的细微变化均不敏感,则认为该模型对对抗样本攻击具有鲁棒性[16]。

模型 $f:\mathbf{R}^n\to\kappa$ 将 n 维输入映射到所有可能标签的集合 $\kappa=\{1,2,\cdots,K\}$ 中。通常,这类模型会将一个输入 $\boldsymbol{x}\in\mathbf{R}^n$ 映射为一个分数向量 $\boldsymbol{y}(\boldsymbol{x})=(y_1(\boldsymbol{x}),y_2(\boldsymbol{x}),\cdots,y_K(\boldsymbol{x}))$,满足:$y_k(\boldsymbol{x})\in[0,1]$ 且 $\sum_{k=1}^{K}y_k(\boldsymbol{x})=1$。分数向量一般是标签上的概率分布,而且模型会返回概率最高的标签,如 $\arg\max_{k\in K}y_k(\boldsymbol{x})$。

按照文献[16],将输入 $\boldsymbol{x}$ 映射到 $\boldsymbol{y}$ 的函数表示为 Q,称其为评分函数;将给出输入 $\boldsymbol{x}$ 最终预测的函数表示为 f,称其为预测过程。

定义 1　鲁棒性。给定输入 $\boldsymbol{x}$,假设对 $\forall\boldsymbol{\alpha}\in B_p(L)$,如果 $f(\boldsymbol{x})=f(\boldsymbol{x}+\boldsymbol{\alpha})$,则称多分类模型 f 对输入 $\boldsymbol{x}$ 上大小为 L 的 p-范数攻击不敏感或鲁棒,这等价于:

$$\forall\boldsymbol{\alpha}\in B_p(L),\ y_k(\boldsymbol{x}+\boldsymbol{\alpha})>\max_{i;i\neq k}(y_i(\boldsymbol{x}+\boldsymbol{\alpha})) \tag{2}$$

式中:$B_p(L):=\{\boldsymbol{\alpha}\in\mathbf{R}^n:\|\boldsymbol{\alpha}\|_p\leqslant L\}$ 为半径为 L 的 p-范数球;$y_k(\boldsymbol{x}+\boldsymbol{\alpha})$ 为针对对抗样本的预测标签的概率评分;$\boldsymbol{\alpha}$ 为 C&W 攻击生成对抗样本时向数据添加的细微扰动。

1.3　差分隐私

DP[19]是 Dwork 和 Roth 针对统计数据库的隐私泄露问题提出的一种隐私保护技术。DP 是指为了防止数据库的计算输出中单个敏感记录的信息泄露,将随机性引入数据库计算中,确保 2 个相邻数据库的查询输出之间相差不大,从而隐藏个别记录的敏感信息。

定义 2　ε-DP。设有一个随机算法 A,O 是 A

的所有可能输出构成的集合，D 和 D' 为任意 2 个相邻数据集，且 $\rho(D,D')\leqslant 1$。对 $S\in O$，如果随机算法 A 满足 ε-DP，则

$$P(A(D)\in S)\leqslant \mathrm{e}^{\varepsilon}P(A(D')\in S) \tag{3}$$

式中：ε 为隐私预算；ρ 为海明度量，即计算 2 个数据库中不同条目的个数；$P(\cdot)$ 为概率分布。

ε-DP 有 2 个关键属性：

1）后处理属性。满足 ε-DP 的计算输出，其后续的任何计算同样满足 ε-DP。

2）输出期望稳定属性。满足 ε-DP 算法的有界输出期望对输入的微小变化不敏感。

定义 3 输出期望稳定界限[16]。假设一个输出界限为 $A(\boldsymbol{x})\in[0,b]$，$b\in\mathbf{R}^{+}$ 的随机算法 A 满足 ε-DP，则算法 A 输出的期望值满足：

$$\forall\boldsymbol{\alpha}\in B_p(1),\ E(A(\boldsymbol{x}))\leqslant \mathrm{e}^{\varepsilon}E(A(\boldsymbol{x}+\boldsymbol{\alpha})) \tag{4}$$

式中：$E(A(\boldsymbol{x}))=\int_0^b P(A(\boldsymbol{x})>t)\,\mathrm{d}t$。

2 AdvILoc 方法设计

2.1 系统架构

针对恶意攻击者利用指纹室内定位深度学习模型参数生成的对抗样本对定位模型进行干扰，极大程度改变了模型的分类结果，导致室内定位的服务质量下降甚至失效。AdvILoc 综合考虑云服务器的不可信及室内定位环境的复杂性等因素，结合 RSS 指纹数据单一且高维的特征，运用隐藏式安全保护机制，通过在离线训练阶段向指纹室内定位深度学习模型中添加满足 ε-DP 噪声层，以确保模型的输出处于输出期望稳定界限范围内，使模型具有高鲁棒性，实现抵御对抗样本攻击。系统架构如图 2 所示。

系统架构分为终端设备层和云服务器层。基于深度学习的 RSS 指纹室内定位方法分为离线训练和在线定位 2 个阶段。在离线训练阶段，终端用户为了使用室内定位服务，自愿将个人的 RSS 指纹数据上传至云服务器。云服务器接收到 RSS 指纹数据后，进行定位模型训练。在线定位阶段，云服务器使用训练好的模型为用户提供高效、可信的定位服务。

1）终端设备层。用户拥有智能终端设备（如智能手机、平板、智能监控设备等）。每个终端设备都具有计算和通信模块，用于从室内区域（如大型购物中心、地下停车场、展览厅等）的多个无线传感器信标（如 WiFi、蓝牙等）收集 RSS 指纹数据。在线定位和离线训练阶段，用户将收集的 RSS 指纹数据集用皮尔逊相关系数[20]（Pearson correlation coefficient，PCC）剔除弱相关的访问点（access point，AP），并将其转换为灰度图后发送给云服务器。

2）云服务器层。作为控制中心，其具有强大的存储和计算能力。离线训练阶段，其负责接收用户终端设备层上传的 RSS 指纹灰度图像，训练并保存抵御对抗样本攻击的指纹室内定位深度学习模型；在线定位阶段，云服务器为用户提供可抵御对抗样本攻击的高质量室内定位服务。

2.2 定位模型

文献[16]提出了一种抵御对抗样本攻击的图像分类 CNN 网络（见图 3），但其在面对特征单一、高维的 RSS 指纹数据时，存在过拟合和学习能力低等问题。

为此，AdvILoc 在借鉴图 3 网络结构的基础上，对网络架构和模型进行了优化，设计了一个能抵御对抗样本攻击的指纹室内定位 CNN 模型 RssDP，如图 4 所示。

相较于图 3 中的 CNN 网络，RssDP 通过添加 1 个池化层和 1 个全连接层，并将噪声层放置在第 1 个全连接层之后来优化网络结构。其中，池化层主要对卷积层提取的特征进行压缩；全连接

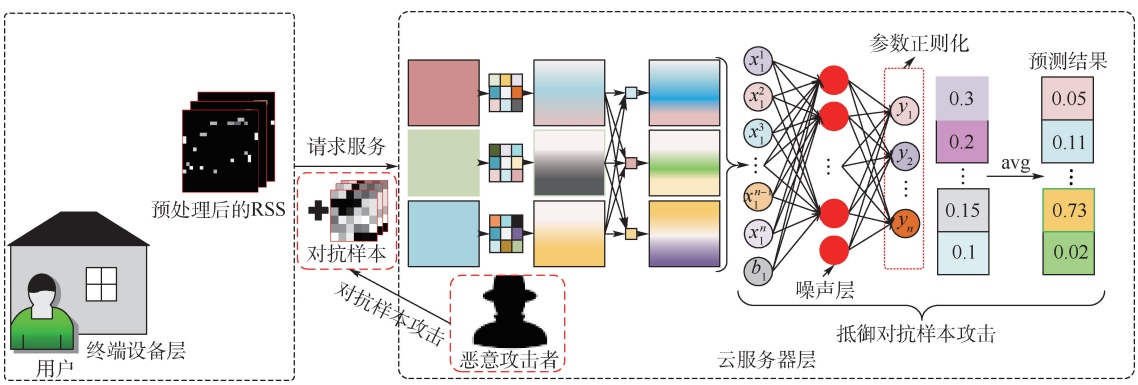


图 2 抵御对抗样本攻击室内定位系统架构

Fig. 2 Architecture of indoor localization system against adversarial sample attack

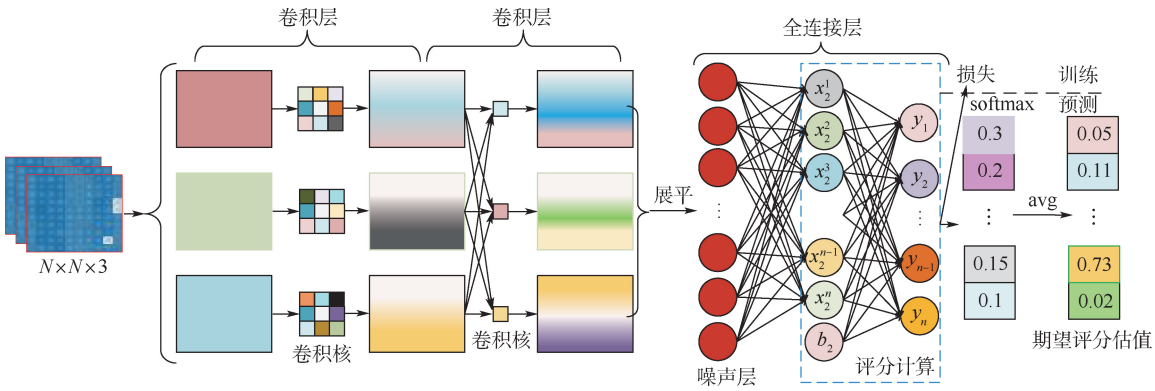


图 3　抵御对抗样本攻击的卷积神经网络结构[16]

Fig. 3　Framework of CNN against adversarial samples attack[16]

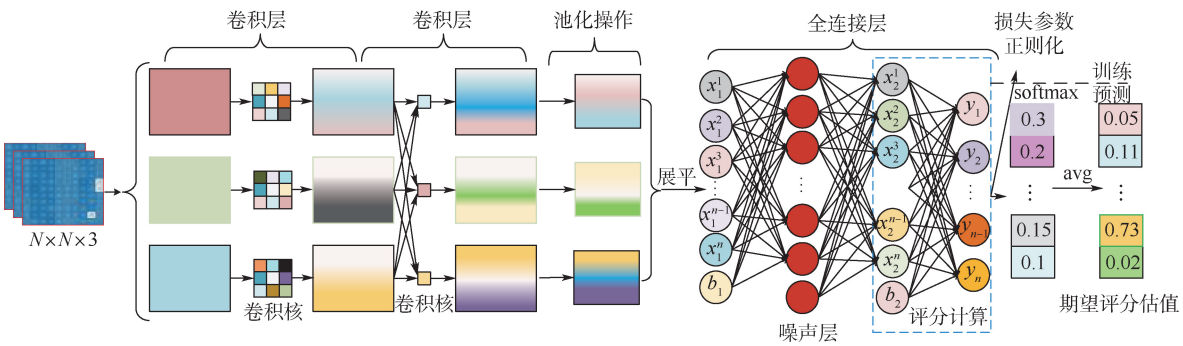


图 4　抵御对抗样本攻击的指纹室内定位深度学习模型

Fig. 4　Indoor localization DL model against adversarial sample attacks

层则将压缩后的特征映射到目标结果;由于对抗样本攻击主要是在深度学习过程中对模型进行干扰,将噪声层调至第 1 个全连接层之后,定位模型抵御对抗样本攻击的效果会更好。在模型优化方面,为了提高 RssDP 模型的泛化能力,AdvILoc 通过添加 Dropout 层与模型参数正则化措施对 RssDP 进一步优化。其中,Dropout 层可随机将一定比例的神经元失活,降低模型对部分神经元的依赖,再通过正则化模型损失函数,不断优化模型参数,提升模型鲁棒性。因此,RssDP 不仅提升了模型鲁棒性,而且解决了模型过拟合和泛化能力低的问题。

2.2.1　噪 声 层

噪声层是 RssDP 抵御对抗样本攻击的关键部分,通过添加 Laplace 噪声使其满足 ε-DP。在线定位阶段,每次调用模型时,噪声层都会在其前层输出结果 $\boldsymbol{g}(\boldsymbol{x})$ 上添加服从 Laplace 分布的噪声 $\boldsymbol{Z}=(Z_1,Z_2,\cdots,Z_m)$,得到扰动输出 $\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{Z}$。其中,噪声 $\boldsymbol{Z}$ 由函数 noise(Δ,L,ε)决定,Δ 为 p-范数输入变化的敏感度,L 为 p-范数的约束攻击界限(即 p-范数球的半径),ε 为隐私预算。在 AdvILoc 中,噪声层服从均值为 0、标准差为 $\sigma=\sqrt{2}\Delta_{p,1}L/\varepsilon$ 的 Laplace 分布,其中,$\Delta_{p,1}$ 为 $\boldsymbol{g}(\cdot)$ 的敏感度。

推论 1　$\boldsymbol{g}(\cdot)$ 的敏感度。给定输入和输出的一些距离指标(分别为 p-范数、1-范数),$\boldsymbol{g}(\cdot)$ 的敏感度 $\Delta_{p,1}$ 定义为输入变化产生的输出的最大变化,表示为

$$\Delta_{p,1}=\max_{\boldsymbol{x},\boldsymbol{x}':\boldsymbol{x}\neq\boldsymbol{x}'}\frac{\|\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{g}(\boldsymbol{x}')\|_1}{\|\boldsymbol{x}-\boldsymbol{x}'\|_p}\tag{5}$$

式中:$\boldsymbol{x}$ 为原始 RSS 指纹数据;$\boldsymbol{x}'$ 为 $\boldsymbol{x}$ 生成的对抗样本。

因为对抗样本攻击源于对 CNN 的概率评分函数 $Q(\boldsymbol{x})=(y_1(\boldsymbol{x}),y_2(\boldsymbol{x}),\cdots,y_K(\boldsymbol{x}))$ 的干扰,而对抗样本的脆弱性源于 $Q(\boldsymbol{x})$ 对输入 p-范数变化的无限敏感性,所以通过在第 1 个全连接层的计算结果上添加满足 ε-DP 的噪声,随机化计算结果。根据 ε-DP 的后处理属性,通过添加噪声将 $Q(\boldsymbol{x})$ 转换为满足 ε-DP 随机概率评分函数,使模型输出期望针对输入中的细微扰动满足有界不敏感性,则基于 RssDP 构建的指纹室内定位深度学习模型可抵御对抗样本攻击,本文定义为 ε-RssDP。

2.2.2　模型优化

为提高 ε-RssDP 的鲁棒性和泛化能力,防止模型发生过拟合,AdvILoc 从 2 个方面对 RssDP 网络进行了优化,即添加 Dropout 层和使用模型参数正则化。

1) Dropout 层优化。AdvILoc 通过在模型网

络的第1个全连接层和第2个全连接层后添加Dropout层,使25%的神经元失活,以降低模型对部分神经元的依赖,防止模型发生过拟合,提高模型鲁棒性。

2) 模型参数正则化优化。AdvILoc通过正则化CNN网络的交叉熵损失函数[16],不断优化模型参数,提高模型泛化能力,即

$$\boldsymbol{\theta}_s,\boldsymbol{\theta}_n \leftarrow \arg\min_{\hat{\boldsymbol{\theta}}_s,\hat{\boldsymbol{\theta}}_n}(\ell(\boldsymbol{y}_n,\hat{\boldsymbol{y}}_n)+R(\hat{\boldsymbol{\theta}}_s,\hat{\boldsymbol{\theta}}_n)) \tag{6}$$

式中:$\ell(\boldsymbol{y}_n,\hat{\boldsymbol{y}}_n)$为CNN网络的交叉熵损失函数;$R(\hat{\boldsymbol{\theta}}_s,\hat{\boldsymbol{\theta}}_n)$为参数正则化;$\hat{\boldsymbol{\theta}}_s$、$\hat{\boldsymbol{\theta}}_n$为正则化前的参数;$\boldsymbol{\theta}_s$、$\boldsymbol{\theta}_n$为经过正则化反向传播调整后的参数;$\boldsymbol{y}_n$为真实标签;$\hat{\boldsymbol{y}}_n$为模型预测得到的标签。

通过网络优化,AdvILoc提高了ε-RssDP的鲁棒性和泛化能力,防止模型发生过拟合现象。

2.2.3 在线定位算法

1) 终端设备层。用户U在请求定位服务时,先将自己实时获取的RSS指纹数据$\boldsymbol{x}$根据PCC值剔除弱相关的AP点,并转换为图像数据集,再发送给云服务器。

2) 云服务器层。将接收到用户U的服务请求输入到已训练好的model(·)中,实时为用户提供定位服务。其中,model(·)为存储到云服务器上抵御对抗样本攻击的室内定位深度学习模型,$\boldsymbol{x}_{ij}^{c'}$为用户提交的服务请求,S_{type}为提供给用户U的定位服务反馈。具体描述如下。

算法1 抵御对抗样本攻击的室内定位方法。

输入:$\boldsymbol{x}$、Δ、L、ε。

输出:S_{type}。

步骤1 根据PCC值剔除$\boldsymbol{x}$中弱相关的AP点,并将其转换为图像数据集$\boldsymbol{x}_{ij}^{c'}$。

步骤2 将$\boldsymbol{x}_{ij}^{c'}$上传至云服务器。

步骤3 云服务器接收$\boldsymbol{x}_{ij}^{c'}$。

步骤4 $S_{\text{type}}\leftarrow\text{model}(\boldsymbol{x}_{ij}^{c'})$ //模型预测。模型噪声层的前一隐藏层的输出结果$\boldsymbol{g}(\boldsymbol{x})$经过满足$\varepsilon$-DP且大小为noise($\Delta$,$L$,$\varepsilon$)的噪声层后,输出期望稳定界限内的定位结果$S_{\text{type}}$。

步骤5 发送S_{type}至终端设备层。

3 安全性分析

本节使用DP技术与鲁棒性条件对AdvILoc进行安全性分析。

ε-DP为指纹室内定位深度学习模型提供鲁棒性认证的前提是建立满足ε-DP的概率评分函数。首先,在给定输入样本的情况下,输入特征值(RSS指纹数据)的模型预测同样满足ε-DP。然后,结合定义3与定义1,为模型提供严格的鲁棒性证明。

定理1 AdvILoc能够提供经过严格鲁棒性认证的ε-DP保护。

证明 形式上,将$\boldsymbol{x}$的特征值(RSS指纹数据)视为数据库中的一条记录,并假设一个随机概率评分函数$Q(\boldsymbol{x})=(y_1(\boldsymbol{x}),y_2(\boldsymbol{x}),\cdots,y_K(\boldsymbol{x}))$,$y_k(\boldsymbol{x})\in[0,1]$且$\sum_{k=1}^{K}y_k(\boldsymbol{x})=1$。如果$Q(\boldsymbol{x})$满足$\varepsilon$-DP,则

$$\forall k,\forall\boldsymbol{\alpha}\in B_p(1),E(y_k(\boldsymbol{x}))\leqslant \mathrm{e}^{\varepsilon}E(y_k(\boldsymbol{x}+\boldsymbol{\alpha})) \tag{7}$$

$\forall k,b=1$,结合定义3得出,AdvILoc将模型的概率评分函数转换为满足ε-DP随机算法$A(\boldsymbol{x})$,即模型$f(\boldsymbol{x})$经过满足ε-DP的噪声层,由于ε-DP的后处理属性,$A(\boldsymbol{x})$的期望$E(A(\boldsymbol{x}))$同样满足ε-DP,模型$f(\boldsymbol{x})$从$E(A(\boldsymbol{x}))$中选取最大置信度标签$\arg\max_{k\in K}E(A(\boldsymbol{x}))$。

结合定义1,针对RSS指纹数据的对抗样本,ε-RssDP满足认证鲁棒性。

假设遵循p-范数指标的随机算法$A(\boldsymbol{x})$满足ε-DP,则

$$\forall\boldsymbol{x},\exists k\in K,E(A_k(\boldsymbol{x}))>\mathrm{e}^{2\varepsilon}\max_{i:i\neq k}E(A_i(\boldsymbol{x})) \tag{8}$$

定义$\boldsymbol{x}':=\boldsymbol{x}+\boldsymbol{\alpha}$,且根据式(7)得

$$E(A_k(\boldsymbol{x}))\leqslant\mathrm{e}^{\varepsilon}E(A_k(\boldsymbol{x}')) \tag{9a}$$

$$E(A_i(\boldsymbol{x}'))\leqslant\mathrm{e}^{\varepsilon}E(A_i(\boldsymbol{x}))\quad i\neq k \tag{9b}$$

式(9a)给出了$E(A_k(\boldsymbol{x}'))$的下界;式(9b)给出了$\max_{i\neq k}E(A_i(\boldsymbol{x}'))$的上界。则

$$E(A_k(\boldsymbol{x}'))\geqslant\frac{E(A_k(\boldsymbol{x}))}{\mathrm{e}^{\varepsilon}}$$

结合式(7)得

$$E(A_k(\boldsymbol{x}'))\geqslant\frac{\mathrm{e}^{2\varepsilon}\max_{i:i\neq k}E(A_i(\boldsymbol{x}))}{\mathrm{e}^{\varepsilon}}=\mathrm{e}^{\varepsilon}\max_{i:i\neq k}E(A_i(\boldsymbol{x}))\geqslant\max_{i:i\neq k}E(A_i(\boldsymbol{x}'))$$

故得出

$$E(A_k(\boldsymbol{x}+\boldsymbol{\alpha}))\geqslant\max_{i:i\neq k}E(A_i(\boldsymbol{x}+\boldsymbol{\alpha}))$$

$\forall\boldsymbol{\alpha}\in B_p(1)$

因此,AdvILoc能够提供经过严格鲁棒性认证的ε-DP保护,即攻击者无法通过对抗样本对模型的输出结果产生影响,保证了模型的有效性和高质量的室内定位服务。 **证毕**

4 实验结果与分析

4.1 实验配置

实验采用TensorFlow库来构建室内定位深度

学习模型,并选择 SGDM 优化器作为模型的优化方法。SGDM 算法解决了梯度下降不稳定、易陷入鞍点的缺点,并可以加快训练速度,减小震荡。迭代次数为 5 000 批次,每批次大小为 128。硬件环境为:Inter(R) Core(TM) i7-10750H CPU @ 2.20 GHz, NVIDIA GeForce GTX1660Ti 显卡, 16 GB RAM,6 GB 显存。

4.2　数据集及预处理

实验采用 Mall[21] 和 UJIIndoorLoc[22] 2 个 RSS 指纹数据集来验证 AdvILoc 的有效性。

1) Mall 数据集。采自一个 31.8 m×14.95 m 的购物中心区域,整个区域被划分为 10 个子区域。在所有位置采集点附近共部署了 22 个 BLE 信标,可以稳定检测到来自 35 个 WiFi AP 的信号。Mall 数据集包含 7 000 个有效样本,每个样本都包含一个区域 ID 和 57 维 RSS 特征(22 维蓝牙特征和 35 维 WiFi 特征)。

2) UJIIndoorLoc 数据集。采自 University Jaume I 的一个 390 m×270 m 区域的 4 层建筑,该数据集包含 19 937 个训练样本和 1 111 个测试样本,每个样本包含一个楼层 ID 和 520 维的 WiFi RSS 指纹数据。

基于深度学习的 RSS 指纹室内定位方法的核心是发现指纹数据与位置点之间的映射关系。实际中,靠近用户终端设备位置采集的 AP 信号较强,相反,远离终端设备位置采集的 AP 信号较弱,甚至不可用[23],因此需要对 RSS 数据进行预处理。针对 2 个 RSS 指纹数据集,通过计算位置点与 RSS 信号强度的 PCC 值来剔除弱相关甚至不可用的 AP,再将其转换成灰度图像。具体数据预处理步骤如下:

步骤 1　将室内区域分为 N 个均匀小区域,即 $\boldsymbol{L}_N=[l_1,l_2,\cdots,l_N]$;用户在室内区域通过终端设备接收的 M 条 RSS 指纹信息表示为 $\boldsymbol{R}_M=[\boldsymbol{r}_1,\boldsymbol{r}_2,\cdots,\boldsymbol{r}_M]$,$\boldsymbol{r}_j=[r_1,r_2,\cdots,r_m,l_i]$ $(1\leqslant j\leqslant M)$ 表示用户在区域 l_i 中 m 个 AP 收集到的第 j 条 RSS 指纹数据;室内区域中 K 个 AP 的 RSS 指纹信息表示为 $\boldsymbol{W}_K$,即 $\boldsymbol{W}_K=[w_1,w_2,\cdots,w_K]$。

步骤 2　计算 $\boldsymbol{L}_N$ 与 $\boldsymbol{W}_K$ 的 PCC 值 c。依据文献[20],当 $|c|<0.3$ 时,AP 的信号强度不可用。因此,实验剔除 $|c|<0.3$ 的 AP。

步骤 3　计算每一个 $\boldsymbol{R}$ 和 c 的哈达玛乘积(HP)[24],并将每一个 $1\times K$ 的 HP 矩阵 $\mathbf{HP}_i$ 重塑为 $m\times m$ 的 RSS 指纹灰度图像。

图 5 给出了 Mall 数据集重构为灰度图像的具体过程。

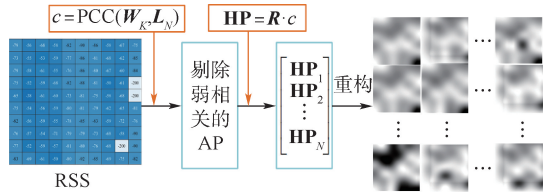


图 5　Mall 数据集预处理

Fig. 5　Preprocessing of Mall datasets

从图 5 中可以看出,Mall 数据集转化为灰度图像最主要的一个步骤是剔除弱相关的 AP,提高 RSS 指纹数据的可用性。利用 HP[24] 将 RSS 指纹数据重构为灰度图像,可以使数据更加适用于深度学习训练,提高模型对 RSS 指纹数据特征的学习能力,进而提高模型的泛化能力。

4.3　评价指标

采用模型性能、定位准确率和认证准确率 3 个指标来评估 ε-RssDP 的性能。准确率表示定位模型预测正确的比例,用于评估 CNN 的性能;认证准确率表示经过鲁棒性认证且预测正确的样本个数占总样本个数的比率,用于评估认证防御模型的性能[25-27]。

1) 模型性能。室内定位深度学习模型的训练精度 Precision 与训练损失 Loss。

2) 定位准确率。

$$\mathrm{ACC}=\frac{\sum_{i=1}^{n}\mathrm{iscorrect}(x_i)}{n} \tag{10}$$

式中:n 为测试集大小;$\mathrm{iscorrect}(x_i)$ 表示对测试样本 x_i 的预测的判断函数,如果 x_i 的预测判断为正确标签,则返回 1,否则返回 0。

3) 认证准确率。

$$\mathrm{CA}=\frac{\sum_{i=1}^{n}(\mathrm{iscorrect}(x_i)\&(\mathrm{robustsize}(\mathrm{scores},\varepsilon,L)\geqslant T))}{n} \tag{11}$$

式中:$\mathrm{robustsize}(\mathrm{scores},\varepsilon,L)$ 函数返回经过认证的鲁棒性大小;scores 为 x_i 经过模型后得到的概率预测分数;鲁棒性阈值 T 为模型在 C&W 攻击下取得最小精度时的攻击大小。将返回值与鲁棒性阈值 T 进行比较,大于 T 返回 1,小于 T 返回 0,再与 $\mathrm{iscorrect}(x_i)$ 返回值取交集,若均为 1,则返回 1,若不相同,则返回 0。

4.4　实验及结果分析

4.4.1　隐私预算 ε 对定位模型的影响

为了测试隐私预算 ε 对定位模型的影响,在 UJIIndoorLoc 数据集及 Mall 数据集上分别进行对

比实验。训练集与测试集按 7:3 划分。实验中，设定$L=1.0$，ε 取值为[0.1,1.0]，实验结果如图 6所示。

图 6(a)、(b)给出了 2 个数据集上不同 ε 对定位模型性能的影响。可以看出，在 UJIIndoorLoc 数据集和 Mall 数据集上，随着模型训练迭代次数的不断增加，模型训练精度均逐步上升，并且不同的 ε 导致模型训练精度曲线上升幅度不一致。这是因为 $\sigma=\sqrt{2}\Delta_{p,1}L/\varepsilon$，当 L 一定时，ε 越大 σ 的值越小，即添加噪声越少，对模型性能的影响越小，同理 ε 越小，对模型性能的影响越大。另外，在 Mall 数据集上，当 $\varepsilon\geqslant 0.5$ 时，随着迭代次数的增加，模型训练精度曲线有明显的上升幅度，且当 $\varepsilon=1.0$ 时，模型训练精度达到最大值 65%。在 UJIIndoorLoc 数据集上存在近似的结果，且当 $\varepsilon=1.0$ 时，模型训练精度达到最大值 99.2%，因此当 $\varepsilon=1.0$ 时，在 2 个数据集上模型性能均可达到最佳。另外，图 6(a)与图 6(b)中的模型训练精度曲线存在明显的差异，这是因为 UJIIndoorLoc 数据集在规模与特征维数上均远大于 Mall 数据集，更多的样本特征为模型训练提供了从输入到输出的更优确定性映射关系，进而获得了性能更优的定位模型。

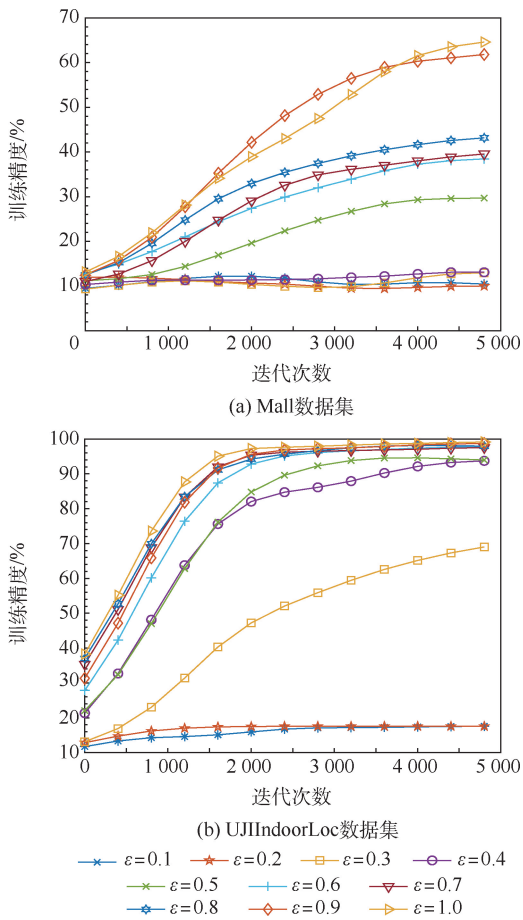


图 6 不同隐私预算 ε 对模型性能的影响

Fig. 6 Effects of different privacy budgets ε on model performance

表 1 给出了 2 个数据集下，随着 ε 的不断增大，噪声尺度 σ 及模型认证准确率 CA 的变化情况。可以看出，随着 ε 的不断增大，σ 不断减小，模型认证准确率 CA 逐步增大。在 Mall 数据集上，随着 ε 的不断增大，模型的认证准确率 CA 呈逐步上升的趋势，$\varepsilon=1.0$ 时模型的认证准确率 CA 达到最大。在 UJIIndoorLoc 数据集上，当 $\varepsilon\geqslant 0.4$ 时，模型的认证准确率 CA 均处于 73% 左右，故不同 ε 构建的模型鲁棒性相近。结合图 6，当 $\varepsilon=1.0$ 时模型训练精度达到最大，故 $\varepsilon=1.0$ 时构建的定位模型鲁棒性最优。

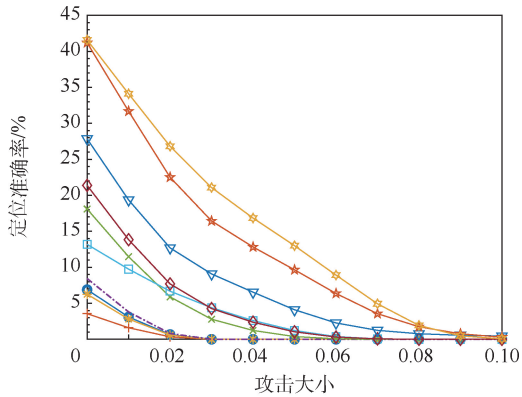
图 7(a)、(b)给出了在 2 个数据集上，C&W 攻击下不同隐私预算 ε 对模型定位准确率 ACC 的影响。可以看出，$\varepsilon=0.1$ 和 $\varepsilon=0.2$ 时的模型定位准确率 ACC 没有明显下降，即模型抵御对抗样本的能力较强，但是模型的定位准确率 ACC 仅有 10%，存在模型失效的问题。同时，在 UJIIndoorLoc 数据集上明显表现出，当 $\varepsilon\geqslant 0.4$ 时，C&W 攻击下不同 ε 对应的模型定位准确率 ACC 下降趋势近似且都比较平缓，在 Mall 数据集上，当 $\varepsilon=1.0$ 时，模型定位准确率 ACC 下降最为平缓。这说明模型抵御对抗样本攻击的效果较好。

$\varepsilon=1.0$ 时，模型认证准确率 CA 及定位准确率 ACC 均可达到最优，且为了保证 L 与攻击大小 L_{attack} 之间的关联性，确保 ε 对 L 的影响最小，故选取 $\varepsilon=1.0$ 为后续实验的隐私预算。

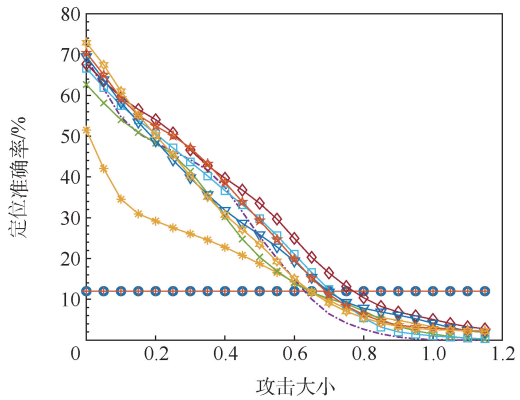
表 1 不同隐私预算 ε 下室内定位模型的认证准确率

Table 1 Certified accuracy of indoor localization model under different privacy budgets ε

数据集	ε	σ	CA/%
Mall	0.1	25.373	9.5
	0.2	12.686	10.4
	0.3	8.458	10.9
	0.4	6.343	16.6
	0.5	5.075	29.6
	0.6	4.229	38.6
	0.7	3.625	32.6
	0.8	3.172	36.8
	0.9	2.819	49.8
	1.0	2.537	50.8
UJIIndoorLoc	0.1	25.373	10.0
	0.2	12.686	11.5
	0.3	8.458	52.1
	0.4	6.343	75.2
	0.5	5.075	72.8
	0.6	4.229	73.6
	0.7	3.625	72.3
	0.8	3.172	77.8
	0.9	2.819	75.7
	1.0	2.537	73.7



(a) Mall数据集



(b) UJIIndoorLoc数据集

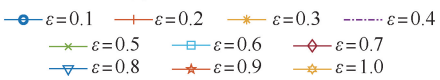


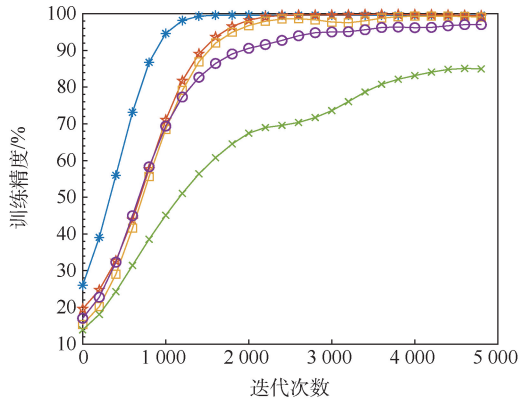
图 7　C&W 攻击下不同隐私预算 ε 对模型定位准确率的影响

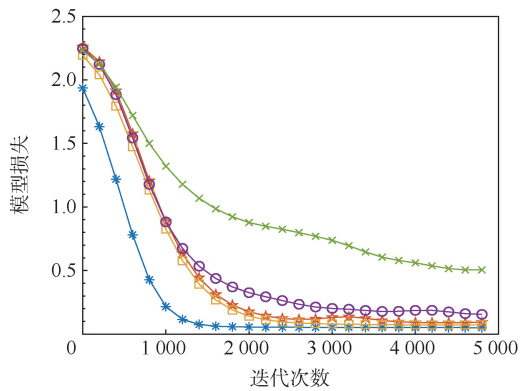
Fig. 7　Effection of different privacy budgets ε on model localization accuracy under C&W attack

4.4.2　不同约束攻击界限 L 构建的模型性能比较

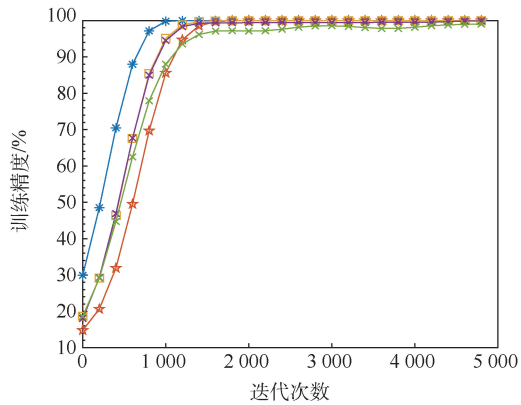
本节实验分别在 UJIIndoorLoc 数据集和 Mall 数据集上测试定位模型在不同约束攻击界限 L 的性能对比,训练集与测试集同样按照 7:3 的比例划分,$\varepsilon=1.0$。参考文献[16],约束攻击界限 L 取值为{0, 0.03, 0.1, 0.3, 1.0}。经过多次实验,C&W 攻击下模型均在鲁棒性阈值 $T=0.02$ 范围之内达到最小精度,故设定 T 的最大值为0.02。其中,$L=0$ 表示模型没有添加噪声层,即模型没有采用防御措施,本文将没有采用防御措施的室内定位深度学习模型定义为 Baseline。实验结果如图 8 所示。

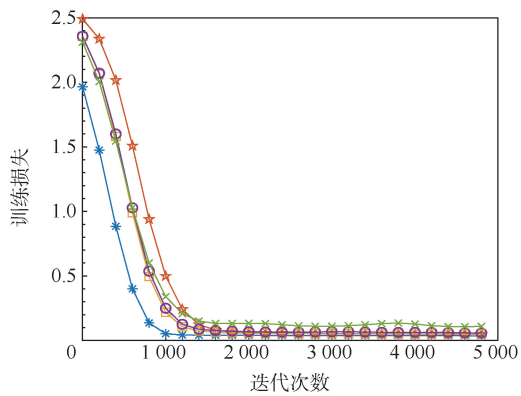
从图 8(a)、(b)中可以看出,随着模型训练迭代次数的增加,不同约束攻击界限 L 下模型的精度都在逐步上升,损失逐步下降。这是因为模型添加噪声层的影响,噪声层添加噪声的大小取决于约束攻击界限 L,即 L 越大,添加噪声越多,模型性能越差;L 越小,添加噪声越少,模型性能越好。但是,当 L 取最大值时,Mall 数据集下的模型精度依旧保持在 85% 以上,UJIIndoorLoc 数据集下模型精度几乎没有下降,因此噪声层对模型





(a) Mall数据集





(b) UJIIndoorLoc数据集

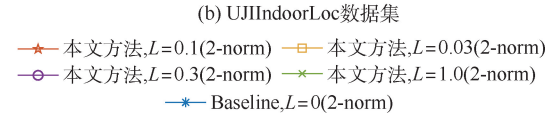


图 8　不同 p-范数约束攻击界限 L 对模型性能的影响

Fig. 8　Impact of constraint attack bounds L of different p-norm on model performance

性能并没有产生明显的不利影响。另外,由于 UJIIndoorLoc 数据集在规模与特征维数上均远大于 Mall 数据集,更多的样本特征为模型训练提供了从输入到输出的更优确定性映射关系,进而获得了性能更优的定位模型。因此,在 Mall、UJIIndoorLoc 两个数据集上模型性能变化程度存在明显差异。

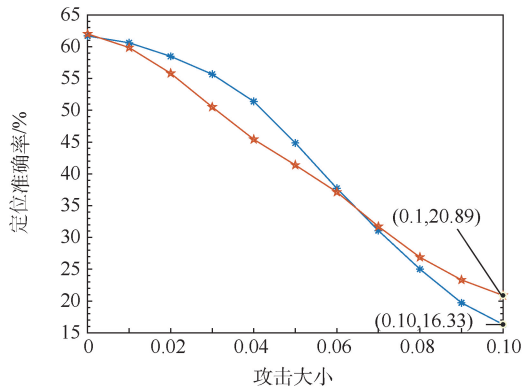
表 2 给出了室内定位深度学习模型在不同约束攻击界限 L 下的认证准确率 CA 的对比数据。可以看出,当 $L=0.03$ 时,模型在 2 个数据集中都取得了最佳的认证准确率 CA,且均大于 60.0%。同时,结合图 8 中 $L=0.03$ 时的室内定位深度学习模型训练精度与 $L=0$ 时相比没有明显的下降,故当 $L=0.03$ 时,可构建最优的室内定位深度学习模型。

图 9(a)、(b)给出了 2 个数据集下 $L=0,0.03$ 时构建的室内定位深度学习模型在 C&W 攻击下定位准确率 ACC 的实验对比结果。其中,Baseline 表示基于 $L=0$ 构建的室内定位深度学习模型。从图 9 中可以得出,在 2 个数据集上,随着约束攻击界限 L 的增大,Baseline 与本文方法的定位准确率均呈下降趋势,且本文方法均比 Baseline 呈现的曲线下降趋势更加平稳,即斜率更小。因此,AdvILoc 能够提供经过严格鲁棒性认证的室内定位深度学习模型。

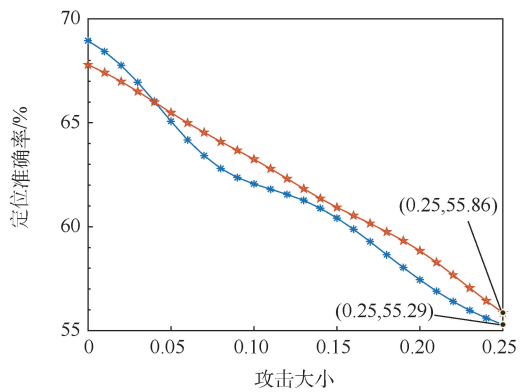
表 2 室内定位深度学习模型在不同约束攻击界限 L 下的认证准确率对比

Table 2 Certified accuracy of indoor localization DL model under different constraint attack boundaries L

数据集	认证准确率/%			
	$L=0.03$	$L=0.1$	$L=0.3$	$L=1.0$
Mall	60.0	57.2	56.4	54.6
UJIIndoorLoc	77.9	76.6	76.4	70.1



(a) Mall数据集



(b) UJIIndoorLoc数据集

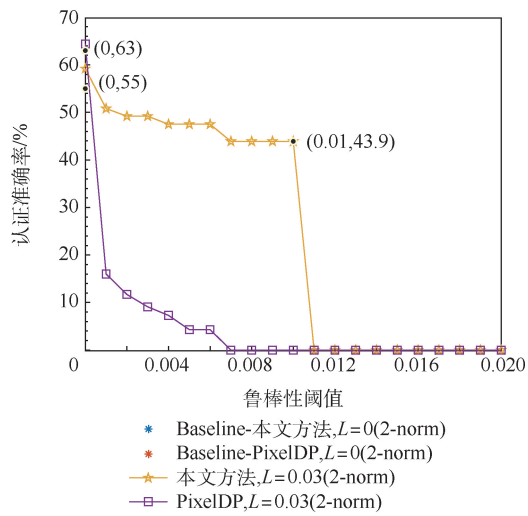
Baseline,L=0(2-norm) 本文方法,L=0.03(2-norm)

图 9 C&W 攻击下不同约束攻击界限 L 对模型定位准确率的影响

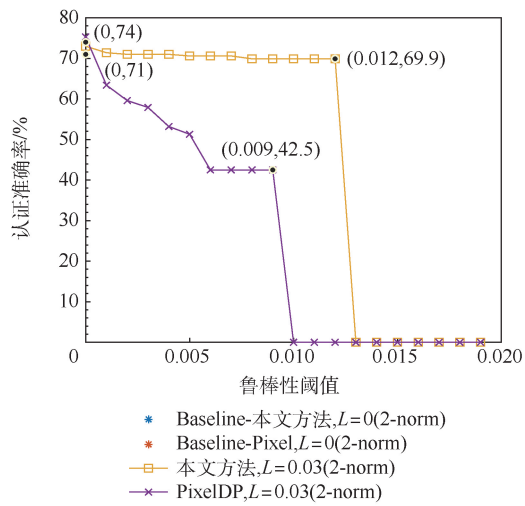
Fig. 9 Effection of different constraint attack boundary L on localization accuracy of model under C&W attack

4.4.3 与基于 PixelDP 的室内定位模型鲁棒性比较

为了验证在 PixelDP[16] 网络基础上,经过 Dropout 层及模型参数正则化优化的模型在 Mall 数据集和 UJIIndoorLoc 数据集上的表现情况,实验分别进行了认证准确率 CA 和 C&W 攻击下定位准确率 ACC 的比较评价。图 10 和图 11 给出了 2 个数据集上,当 $L=0.03$ 时,AdvILoc 与基于 PixelDP 的室内定位模型在认证准确率 CA 及 C&W



(a) Mall数据集



(b) UJIIndoorLoc数据集

图 10 AdvILoc 与基于 PixelDP 的室内定位模型认证准确率对比

Fig. 10 Certified accuracy between AdvILoc and indoor localization model based on PixelDP

攻击下定位准确率 ACC 的对比结果。其中,Baseline-本文方法表示当 $L=0$ 时优化的室内定位模型,Baseline-Pixel 和 Baseline-PixelDP 表示当 $L=0$ 时基于 PixelDP 的室内定位模型,本文方法表示优化后基于 ε-RssDP 的抵御对抗样本攻击的室内定位模型,Pixel和 PixelDP 表示基于 PixelDP 构建的抵御对抗样本攻击的室内定位模型。

从图 10(a)、(b)中可以看出,在 Mall 数据集上,鲁棒性阈值 $T=0.01$ 时 Baseline-本文方法的认证准确率 CA 比 Baseline-PixelDP 低 8%;而在 UJIIndoorLoc 数据集上,Baseline-本文方法的认证准确率 CA 比 Baseline-Pixel 高 3%。但是,随着鲁棒性阈值 T 的不断增大,由于 Dropout 层及模型参数正则化模型优化措施,在 2 个数据集上,本文方法的认证准确率 CA 较 PixelDP 和 Pixel 下降均更加平稳,且在 Mall 数据集上,当 $T=0.01$ 时本文方法的认证准确率 CA 仍可达到 43.9%,即本文方法在攻击大小为 0.01 的 C&W 攻击下最小精度仍可取得 43.9%。在 UJIIndoorLoc 数据集上,当鲁棒性阈值 $T=0.013$ 时本文方法的认证准确率 CA 仍可达到 69.9%,即本文方法在攻击大小为 0.013 的 C&W 攻击下最小精度仍可取得 69.9%。然而,当鲁棒性阈值 $T=0.007$ 及 $T=0.01$ 时,PixelDP 和 Pixel 的认证准确率 CA 已降为 0。因此,在面向对抗样本攻击时,本文方法鲁棒性更强,防御效果更佳。

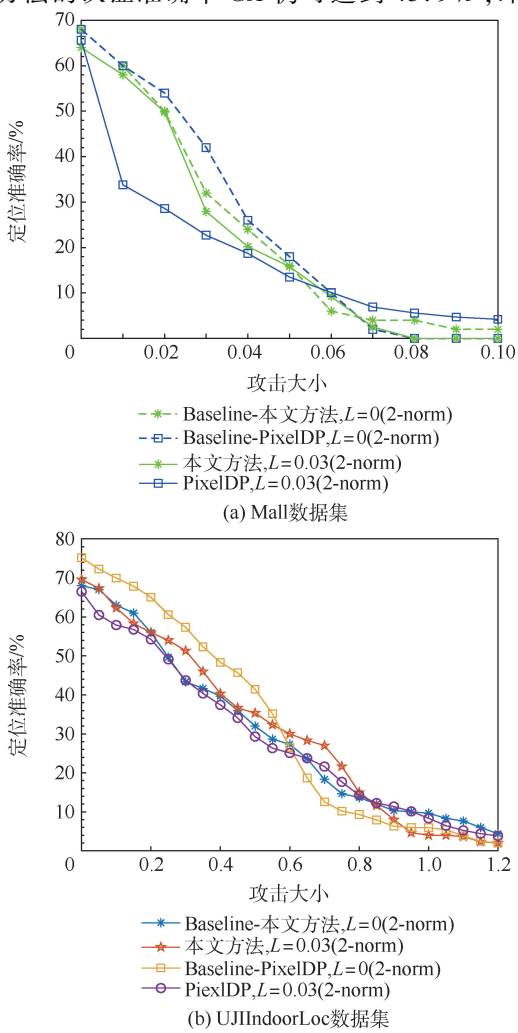


图 11　C&W 攻击下 AdvILoc 与基于 PixelDP 的室内定位模型定位准确率的对比

Fig. 11　Comparison of localization accuracy between AdvILoc under C&W attack with indoor localization model based on PixelDP

从图 11(a)、(b)中可以看出,C&W 攻击下,在 2 个数据集上,随着攻击大小 L_{attack} 不断增大,Baseline-本文方法定位准确率比 Baseline-PixelDP 和 Baseline-Pixel 下降更加平稳。其中,在 Mall 数据集上,当攻击大小 $L_{attack}<0.01$ 时,本文方法比 PixelDP 下降更平稳。虽然,当攻击大小 $L_{attack}>0.055$ 时,PixelDP 定位准确率高于本文方法,但是最终定位准确率相差小于 1%;在 UJIIndoorLoc 数据集上呈现了相近的实验结果,当攻击大小 $L_{attack}=0.5$ 时,Baseline-本文方法比 Baseline-Pixel 下降更平稳,且当攻击大小 $L_{attack}<0.9$ 时,相较于 Pixel,本文方法取得更加良好的表现。虽然当 $0.9<L_{attack}<1.2$ 时,Pixel 准确率高于本文方法,但最终定位准确率相差小于 5%。

表 3 给出了 AdvILoc 与基于 PixelDP 的室内定位模型计算效率的对比。可以看出,在 2 个数据集上,Pixel 与本文方法的训练时间和测试时间相近,在 Mall 数据集上本文方法的模型训练时间比 Pixel 仅高 0.970 2 s,这是因为相较 Pixel,本文方法添加了 Dropout 层及模型参数正则化对模型进行了优化,所以训练时间较长。在 UJIIndoorLoc 数据集上,本文方法的训练时间比 Pixel 仅低 0.180 0 s,这是因为相较 Mall 数据集,UJIIndoorLoc 数据集规模更大,且 Pixel 没有池化层对特征进行压缩,所以相较于本文方法全连接层进行线性计算时,时间更长。但在 2 个数据集上,2 种方法的测试时间基本相近。

表 3　AdvILoc 与基于 PixelDP 的室内定位模型计算效率对比

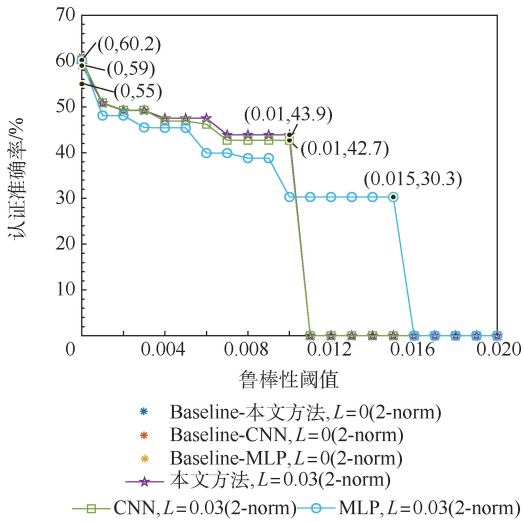
Table 3　Comparative evaluation of calculation efficiency between AdvILoc and indoor localization model based on PixelDP

数据集	模型	训练时间/s	测试时间/ms
Mall	Pixel	14.717 9	0.761 2
	本文方法	15.688 1	0.761 5
UJIIndoorLoc	Pixel	30.120 2	0.761 2
	本文方法	29.940 2	0.761 2

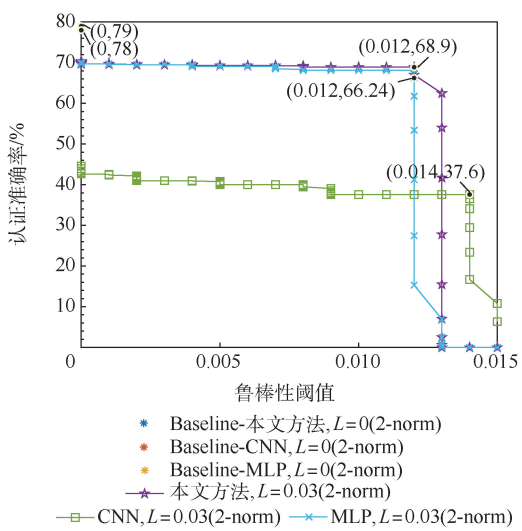
综上,相较于基于 PixelDP 的室内定位模型,AdvILoc 在保证时间开销的同时,泛化能力、鲁棒性更优。

4.4.4 不同室内定位深度学习模型鲁棒性比较

为了验证 AdvILoc 网络 ε-RssDP 的有效性,与基于 CNN[8]、MLP[7] 室内定位模型进行了比较。图 12 和图 13 给出了 2 个数据集上,在 $L=0.03$ 的条件下,ε-RssDP 与 CNN[8]、MLP[7] 室内定位深度学习模型的认证准确率和 C&W 攻击下模型定位准确率的对比实验结果。其中,Baseline-本文方法表示当 $L=0$ 时的 ε-RssDP;Baseline-MLP 表示当 $L=0$ 时基于文献[7]的室内定位模型;Baseline-CNN 表示基于当 $L=0$ 时基于文献[8]中的深度学习网络构建的室内定位模型,本文方法表示 ε-RssDP,MLP 表示基于文献[7]深度学习网络构建的抵御对抗样本攻击的室内定位模型,CNN 表示基于文献[8]深度学习网络构建的抵御对抗样本攻击的室内定位模型。



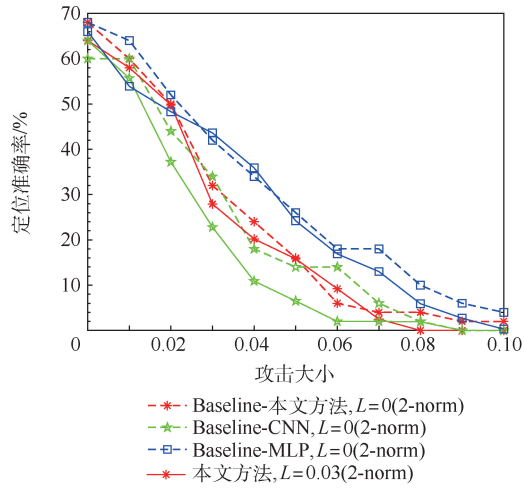
(a) Mall数据集



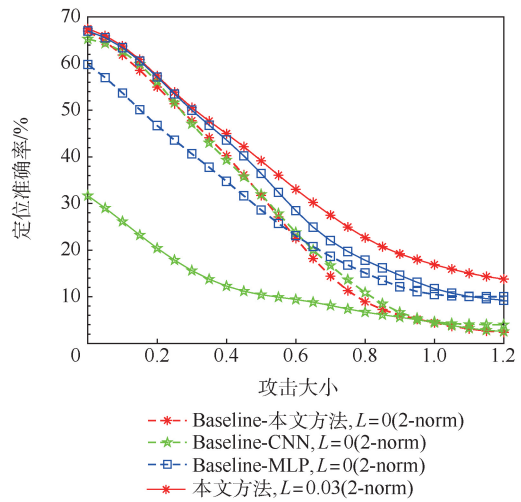
(b) UJIIndoorLoc数据集

图 12 ε-RssDP 与基于 CNN、MLP 的室内定位深度学习模型认证准确率的对比

Fig. 12 Certified accuracy between ε-RssDP and indoor localization DL model based on CNN and MLP



(a) Mall数据集



(b) UJIIndoorLoc数据集

图 13 C&W 攻击下 ε-RssDP 与基于 CNN、MLP 的室内定位深度学习模型定位准确率的对比

Fig. 13 Comparison of positioning accuracy between ε-RssDP and indoor localization DL model based on CNN and MLP network under C&W attack

从图 12(a)中可以看出,在 Mall 数据集上,Baseline-本文方法、Baseline-CNN 与 Baseline-MLP 的认证准确率 CA 相近,均达到了 55% 以上。随着鲁棒性阈值 T 的不断增大,当 $T=0.011$ 时,本文方法与 CNN 的认证准确率 CA 已降为 0,而当 $T=0.016$ 时,MLP 的认证准确率 CA 才降为 0。这是因为:虽然在网络结构上相较于 MLP,本文方法、CNN 添加了卷积层、池化层来提升模型提取特征的能力,但是 Mall 数据集的规模较小、特征较少,本文方法、CNN 中学习到卷积层提取出的特征相较 MLP 学习的少,使本文方法、CNN 模型的泛化能力较差。然而,当 $T<0.011$ 时,本文方法、CNN 模型的认证准确率 CA 均远大于 MLP,

达到 40% 以上,甚至本文方法的认证准确率 CA 达到了最大。从图 12(b)中可以看出,在 UJIIndoorLoc 数据集上,由于数据集特征多样且丰富,Baseline-本文方法与 Baseline-CNN 的认证准确率相近,均达到了 78% 以上,但 Baseline-MLP 的认证准确率不足 50%。随着鲁棒性阈值 T 不断增大,本文方法的认证准确率始终比 CNN、MLP 方法下降更平稳,且当 $T=0.012$ 时,本文方法的认证准确率仍可达到 68.9%。综上,相较于 CNN、MLP,本文方法鲁棒性更强,抵御对抗样本攻击的效果更佳。

从图 13(a)可以看出,在 Mall 数据集上,C&W 攻击下,随着攻击大小 L_{attack} 不断增大,Baseline-本文方法曲线和 Baseline-CNN 曲线均低于 Baseline-MLP 曲线,且本文方法、MLP 下降趋势相近且比 CNN 平稳,但一直低于 MLP。这同样是因为本文方法、CNN 添加了卷积层、池化层来提升模型提取特征的能力,但是 Mall 数据集的规模较小、特征较少,本文方法、CNN 中学习到卷积层提取出的特征相较 MLP 少,使本文方法、CNN 模型的鲁棒性较差。但是同时有卷积层的本文方法和 CNN,随着攻击大小 L_{attack} 不断增大,本文方法的定位准确率一直大于 CNN 的定位准确率。从图 13(b)可以看出,在 UJIIndoorLoc 数据集上,C&W 攻击下,随着攻击大小 L_{attack} 不断增大,由于 CNN 在网络结构上比 ε-RssDP 多添加了一个卷积层,使得特征提取能力更优,Baseline-本文方法曲线比 Baseline-CNN 下降更快。但是,由于 ε-RssDP 中添加 Dropout 层及模型参数正则化措施来提升模型的性能和鲁棒性,本文方法比 CNN、MLP 模型定位准确率下降平缓。

表 4 给出了在 2 个数据集上,在 $L=0.03$ 的条件下,ε-RssDP 与 CNN[8]、MLP[7] 室内定位深度学习模型的计算效率的对比实验结果。可以看出,在 2 个数据集上,Pixel 与本文方法的模型训练时间和模型测试时间相近,在 Mall 数据集上,本文方法的模型训练时间比 CNN 高 2.150 5 s,但是比 MLP 低 0.346 4 s,这是因为相较于本文方法、CNN,MLP 的网络架构没有卷积层、池化层对特征进行提取与压缩,且 Mall 数据集的规模较小、特征较少,所以全连接层进行线性计算时,MLP 的时间开销更多。然而,在 UJIIndoorLoc 数据集上,本文方法的模型训练时间比 CNN 仅高 4.669 8 s,比 MLP 仅高 4.069 4 s,这是因为较 Mall 数据集,UJIIndoorLoc 数据集规模更大,本文方法与 CNN 经过卷积层和池化层提取与压缩后的特征数据较多,且本文方法进行了 Dropout 与模型参数正则化优化,所以本文方法的时间开销较多,但是在2 个数据集上,3 种方法的测试时间开销相近。

表 4　ε-RssDP 与基于 CNN、MLP 网络结构构建的室内定位深度学习模型计算效率的对比

Table 4　Comparative evaluation of calculation efficiency between ε-RssDP and indoor localization DL model based on CNN and MLP network structure

数据集	方法	训练时间/s	测试时间/ms
Mall	CNN	13.303 7	0.761 2
	MLP	15.800 6	0.761 1
	本文方法	15.454 2	0.761 5
UJIIndoorLoc	CNN	24.043 1	0.761 2
	MLP	24.643 5	0.761 1
	本文方法	28.712 9	0.761 3

综上,ε-RssDP 在认证准确率与 C&W 攻击下的模型定位准确率方面均有较好表现,且相较于 CNN、MLP,ε-RssDP 在保证时间开销的同时,模型鲁棒性更强。

5　结　论

1)针对基于深度学习的 RSS 指纹室内定位方法易遭受对抗样本攻击导致模型鲁棒性不高甚至失效的问题,本文设计了一种基于隐藏式安全保护的 RSS 指纹室内定位深度学习模型,结合 DP、正则化和 Dropout 防御技术,提高了模型的鲁棒性,能有效抵御对抗样本攻击。

2)给出了对所提方法的安全性理论分析,证明了所提方法能够提供严格鲁棒性认证的 ε-DP 保护,使攻击者无法通过对抗样本对模型的输出产生影响,确保了模型的有效性。

3)实验结果表明,与基于 MLP、CNN 的室内定位方法相比,所提方法在保证时间开销和模型性能的情况下,提高了模型的鲁棒性;在满足 l_2 范式规范的 C&W 攻击下,随着攻击大小不断增大,模型的定位准确率下降更平稳。

后续工作将深入研究隐私预算 ε 与约束攻击界限 L 之间的关联性及其对模型性能的影响,在模型噪声层添加更为合理的随机噪声,在保证模型鲁棒性的同时提高模型的定位性能。

参考文献(References)

[1] 钱堃,张新宇,杨铮,等. 基于离开角的商用毫米波设备定位方法研究[J]. 中国科学:信息科学,2021,51(1):122-138.
QIAN K,ZHANF X Y,YANG Z,et al. AoD-based localization with cots millimeter-wave device[J]. Scientia Sinica Informationis,2021,51(1):122-138(in Chinese).

[2] 石柯,宋小妹,王信达,等. 多传感器辅助的WiFi信号指纹室内定位技术[J]. 软件学报,2019,30(11):3457-3468.
SHI K,SONG X M,WANG X D,et al. Multi-sensor assisted WiFi signal fingerprint based indoor positioning technology[J]. Journal of Software,2019,30(11):3457-3468(in Chinese).

[3] ZHANG L,LIU X,SONG J. A comprehensive study of bluetooth fingerprinting-based algorithms for localization[C]// Proceedings of the 27th International Conference on Advanced Information Networking and Applications Workshops. Piscataway:IEEE Press,2013:300-305.

[4] LEE S,HA K N,LEE K C. A pyroelectric infrared sensor-based indoor location-aware system for the smart home[J]. IEEE Transactions on Consumer Electronics,2006,52(4):1311-1317.

[5] GONZÁLEZ E,PRADOS L,RUBIO A J. ATLINTIDA:A robust indoor ultrasound location system:Design and evaluation[C]// Proceeding of the 3rd Symposium of Ubiquitous Computing and Ambient Intelligence. Berlin:Springer,2009,51:180-190.

[6] BAHL P,PADMANABHAN V N. RADAR:An in-building RF-based user location and tracking system[C]// Proceedings of the IEEE Conference on Computer Communications. Piscataway:IEEE Press,2000:775-784.

[7] HSIEH C H,CHEN J Y,NIEN B H. Deep learning-based indoor localization using received signal strength and channel state information[J]. IEEE Access,2019,7:33256-33267.

[8] WANG X Y,WANG X Y,MAO S W. Deep convolution neural networks for indoor localization with CSI images[J]. IEEE Transactions on Network Science and Engineering,2020,7(1):316-327.

[9] 张学军,何福存,盖继扬,等. 边缘计算下指纹室内定位差分私有联邦学习模型[J/OL]. 计算机研究与发展,2022(2022-02-11). https://kns.cnki.net/kcms/detail/11.1777.TP.20220211.1039.004.html.
ZHANG X J,HE F C,GAI J Y,et al. Differential federated learning model for indoor fingerprint location under edge calculation[J/OL]. Journal of Computer Research and Development,2022(2022-02-11). http://kns.cnki.net/kcms/detail/11.1777.TP.20220211.1039.004.html(in Chinese).

[10] PATIL M,WANG X,WANG X,et al. Adversarial attacks on deep learning-based floor classification and indoor localization[C]// Proceeding of the 3rd ACM Workshop on Wireless Security and Machine Learning. New York:ACM,2021:7-12.

[11] 钱亚冠,张锡敏,王滨,等. 基于二节对抗样本的对抗训练防御[J]. 电子与信息学报,2021,43(11):3367-3373.
QIAN Y G,ZHANG X M,WANG B,et al. Adversarial training defense based on second-order adversarial examples[J]. Journal of Electronics & Information Technology,2021,43(11):3367-3373(in Chinese).

[12] 刘西蒙,谢乐辉,王耀鹏,等. 深度学习中的对抗攻击与防御[J]. 网络与信息安全学报,2020,6(5):36-53.
LIU X M,XIE L H,WANG Y P,et al. Adversarial attacks and defenses in deep learning[J]. Chinese Journal of Network and Information Security,2020,6(5):36-53(in Chinese).

[13] 王滨,郭艳凯,钱亚冠,等. 针对卷积神经网络流量分类器的对抗样本攻击防御[J]. 信息安全学报,2022,7(1):145-156.
WANG B,GUO Y K,QIAN Y G,et al. Defense against sample attack against convolutional neural network traffic classifier[J]. Journal of Information Security,2022,7(1):145-156(in Chinese).

[14] XIE C,WANG J,ZHANG Z,et al. Mitigating adversarial effects through randomization[C]// Proceeding of 6th International Conference on Learning Representations,2018:1-16.

[15] LIU X,LI Y,WU C,et al. Adv-BNN:Improved adversarial defense through robust Bayesian neural network[EB/OL]. (2019-05-04)[2021-12-25]. https://arxiv.org/abs/1810.01279.

[16] LECUYER M,ATLIDAKIS V,GEAMBASU R,et al. Certified robustness to adversarial examples with differential privacy[C]// Proceedings of the 2019 IEEE Symposium on Security and Privacy. Piscataway:IEEE Press,2019:656-672.

[17] CARLINI N,WAGNER D. Towards evaluating the robustness of neural networks[C]// Proceedings of the 2017 IEEE Symposium on Security and Privacy. Piscataway:IEEE Press,2017:39-57.

[18] SHI Y,WANG S,HAN Y. Curls & Whey:Boosting black-box adversarial attacks[C]// 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2019:6519-6527.

[19] DWORK C,ROTH A. The algorithmic foundations of differential privacy[J]. Foundations and Trends in Theoretical Computer Science,2014,9(3-4):211-407.

[20] BENESTY J,CHEN J,HUANG Y,et al. Pearson correlation coefficient[M]// COHEN I,HUANG Y,CHEN J,et al. Noise reduction in speech processing. Berlin:Springer,2009:1-4.

[21] JIANG X,CHEN Y,LIU J. FSELM:Fusion semi-supervised extreme learning machine for indoor localization with Wi-Fi and bluetooth fingerprints[J]. Soft Computing,2018,22(6):3621-3635.

[22] TORRES-SOSPEDRA J. UJIIndoorLoc:A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems[C]//Proceeding of 2014 International Conference on Indoor Positioning and Indoor Navigation. Piscataway:IEEE Press,2014:261-270.

[23] JIANG X,LIU J,CHEN Y. Feature adaptive online sequential extreme learning machine for lifelong indoor localization[J]. Neural Computing and Applications,2016,27(1):215-225.

[24] STYAN G P H. Hadamard products and multivariate statistical analysis[J]. Linear Algebra and Its Applications,1973,6:217-240.

[25] WONG E,KOLTER Z. Provable defenses against adversarial examples via the convex outer adversarial polytope[C]// 35th International Conference on Machine Learning,2018:5286-5295.

[26] RAGHUNATHAN A,STEINHARDT J,LIANG P. Certified defenses against adversarial examples[EB/OL]. (2020-10-31)[2021-12-25]. https://arxiv.org/abs/1801.09344v1.

[27] DVIJOTHAM K,GOWAL S,STANFORTH R,et al. Training verified learners with learned verifiers[EB/OL]. (2018-05-29)[2021-12-25]. https://arxiv.org/abs/1805.10265v1.

A fingerprint indoor localization method against adversarial sample attacks

ZHANG Xuejun[1,*], BAO Junda[1], HE Fucun[1], GAI Jiyang[1], TIAN Feng[2], HUANG Haiyan[1]

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. School of Computer Science, Shaanxi Normal University, Xi'an 710062, China)

Abstract: With the development of urban intelligence, the indoor positioning services based on WiFi received signal strength (RSS) have attracted extensive attention of society. The deep learning technology is a powerful method to achieve high indoor positioning performance using RSS signal. However, it is vulnerable to adversarial sample attack, which brings serious security risks to the indoor positioning system. In this paper, we propose a deep learning based fingerprint indoor localization method using WiFi RSS against adversarial samples attack (AdvILoc), leveraging the research and analysis of anti-sample defense methods in the field of image recognition. The AdvILoc defend against adversarial samples attack through adding a polling layer, a full connection layer, and a noise layer with differential privacy to the fingerprint indoor positioning deep learning model, which contemplates the characteristics of single and dimension of RSS signals. It also solves the problem of overfitting and weak generalization of deep learning based fingerprint indoor localization model. Meanwhile, the robustness of the model against adversarial samples attack is improved by adding a Dropout layer and designing the parameters regularization of model. The experimental results on two real indoor RSS fingerprint datasets show that, compared with the existing indoor localization methods based on multi-layer perception (MLP) and convolution neural network (CNN), the AdvILoc improves the robustness of the localization model against adversarial samples attack without compromising the localization performance. Additionally, under the C&W attack that meets the l_2-normal form specification, the localization accuracy of the model also decreases more smoothly with the increment of the attack size.

Keywords: indoor localization; adversarial samples; deep learning; C&W attacks; differential privacy

Received: 2021-12-28; **Accepted**: 2022-03-11; **Published online**: 2022-04-11 08:31

URL: kns.cnki.net/kcms/detail/11.2625.V.20220408.1225.001.html

Foundation items: National Natural Science Foundation of China (61762058,61901201); Foundation of A Hundred Youth Talents Training Program of Lanzhou Jiaotong University; Natural Science Foundation of Gansu Province, China (21JR7RA282); Industry Support Program for College and University of Gansu Province (2022CYZC-38); The Fundamental Research Funds for the Central Universities (GK202103090); Natural Science Basic Research Program of Shaanxi (2022JM-329)

* **Corresponding author**. E-mail: xuejunzhang@mail.lzjtu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0392

微小卫星隐身构型设计及优化分析

秦远田[1,*], 孙汗青[1], 岳鑫[2]

(1. 南京航空航天大学 航天学院, 南京 210093; 2. 南京航空航天大学 电子信息工程学院, 南京 210093)

摘 要: 为研究天巡一号微小卫星的电磁散射特性,建立了具有隐身外形设计的微小卫星电磁计算模型。采用物理光学法(PO)对不同状态下的雷达散射截面(RCS)进行数值计算,并与微波暗室的试验结果对比,验证了 PO 的准确性。在此基础上,着重分析了卫星 RCS 入射角、极化、频率、电尺寸响应特性和全姿态角空间 RCS 响应特性。参考天巡一号的隐身构型设计,将天巡一号优化为对称的尖锥构型,通过不断增加尖锥棱边数来优化构型,得到具有更低 RCS 构型的橄榄体卫星。结果表明:天巡一号的隐身姿态可有效应对单站雷达威胁,最佳隐身姿态下的空间 RCS 均值低于非隐身姿态 4.89 dBsm;在 S 波段(3 GHz)下,橄榄体卫星 RCS 算术均值和 RCS 幅值分别低于天巡一号 4.77 dBsm 和 31.66 dBsm;在 X 波段(10 GHz)下,橄榄体卫星 RCS 算术均值和 RCS 幅值分别低于天巡一号 3.65 dBsm 和 43.97 dBsm。

关 键 词: 微小卫星; 雷达散射截面; 构型优化; 物理光学法; 隐身

中图分类号: V423.4; TN973

文献标志码: A **文章编号**: 1001-5965(2022)11-2102-09

微小卫星普遍具有研制周期短、发射难度小、成本低等特点[1],其与微纳卫星可以从质量上进行区分。微小卫星指质量在 1 000 kg 以下的人造卫星,包括小卫星、微卫星、纳卫星、皮卫星和飞卫星等,而微纳卫星的质量往往不足 10 kg,即微纳卫星包含于微小卫星。它们都为空间科学的蓬勃发展作出了巨大贡献[2]。

隐身卫星对未来空间攻防中起着至关重要的作用。针对卫星在不同探测系统下所面临的不同威胁,卫星隐身技术可以分为雷达隐身、红外隐身、可见光隐身和射频隐身[3]。其中,卫星的主要威胁之一来自雷达,设计和分析具有低雷达散射截面(radar cross section, RCS)外形设计的微小卫星就显得尤为重要。国内外关于隐身卫星研究进展的资料并不多见。1990 年,美国朦胧计划的首星(Misty-1)成功发射,圆柱形的卫星本体下方有一锥形充气罩,地面发射的雷达波照射至锥形充气罩会折射至其他方向,使回波能量得到极大抑制[4],同时这也是世界上首颗在轨运行的隐身卫星。郑侃[5]依据多棱面设计思路,提出了 3 种隐身微小卫星构型,并通过 RCS 对比仿真,得到了多棱面锥体的隐身效果较优这一结论。朱冬骏等[6]提出了一种兼顾雷达隐身和光学隐身的卫星构型,其光学横截面积(optical cross section, OCS)峰值仅为 0.082 m^2,不易被光学探测系统监视识别。

针对国内外学者对卫星隐身这一领域的研究还相对较少这一现状,本文通过对天巡一号进行电磁散射特性的数值模拟和分析,得到了 RCS 入射角、极化、频率、电尺寸等响应特性的分析报告,并对该卫星的构型进行优化使其能够实现 RCS 在某些角域乃至全向角域算术均值的进一步缩减,期望对今后微小卫星的隐身设计提供技术参考。

收稿日期: 2021-07-13; **录用日期**: 2021-11-05; **网络出版时间**: 2021-11-30 11:39

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20211129.1538.001.html

* **通信作者**. E-mail: qinyt@nuaa.edu.cn

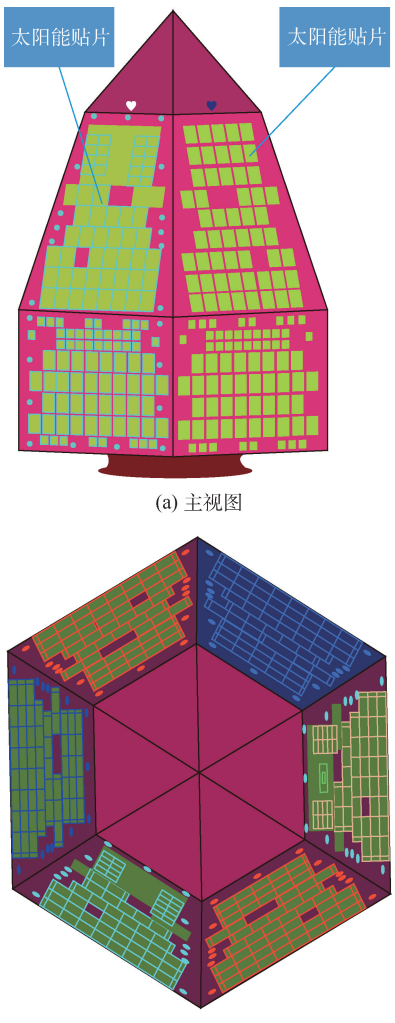
引用格式: *秦远田, 孙汗青, 岳鑫. 微小卫星隐身构型设计及优化分析[J]. 北京航空航天大学学报, 2022, 48(11): 2102-2110.*

QIN Y T, SUN H Q, YUE X. Stealthy configuration design and optimization analysis of microsatellite[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2102-2110(in Chinese).

1　RCS 数值计算流程

1.1　隐身微小卫星电磁计算模型

天巡一号星体为拥有交错斜置外形设计的多面体,从底部到顶部依次为 1 个正六边形底面(用于安置对接环,实现星箭分离)、6 个矩形面、6 个倾斜梯形面和 6 个倾斜三角形面,如图 1 所示。其中,矩形面和梯形面均安装有太阳能贴片,可随时为卫星提供能源支持。当星体尖顶向下时可实现对地雷达隐身,尖顶向上时可实现对天雷达隐身。



(a) 主视图

(b) 俯视图

图 1　天巡一号的电磁计算模型

Fig. 1　Electromagnetic calculation model of TX-1

1.2　RCS 数值计算方法

从计算电磁学的角度出发,RCS 计算方法可分为微分方程法、积分方程法、有限元法及适用于目标在光学区的高频近似方法。在本文所研究的频率范围内,卫星目标主要处于光学区。矩量法(method of moment,MOM)就是典型的积分方法,通过对矩阵采用直接计算,不存在收敛性的问题,而且由于采用面网格进行划分,未知量的总数大大降低,但是因为“源”之间存在耦合,所以矩阵属于稠密矩阵,从而导致 MOM 计算速度慢[7-11],大多数用于一些精细电小或电中尺寸的 RCS 计算。多层快速多极子(multi level fast multipole method,MLFMM)在 MOM 的基础上采用了加速算法,既降低了内存需求又提高了计算速度,可用于电大尺寸的 RCS 求解,但迭代求解时遇到病态矩阵容易出现不收敛的情况,如飞机进气道、尾喷之类的腔体结构。物理光学法(physical optics,PO)主要适用于目标在光学区的散射特性求解,从斯特拉顿-朱兰成(Stratton-Chu)散射场积分方程出发,通过切平面近似方法假设目标表面电流值等于积分面元处目标为理想光滑平面时的电流值,并据此来对物理光学电流进行积分得到目标远区散射场[12-16]。根据切平面近似所得到面元 RCS 的平方根为

$$\sqrt{\sigma} = -\frac{k}{\sqrt{\pi}}\mathrm{j}\int_S \boldsymbol{n}\cdot(\boldsymbol{e}_\mathrm{r}\times\boldsymbol{h}_\mathrm{i})\exp[k\boldsymbol{r}'(\boldsymbol{i}-\boldsymbol{s})\mathrm{j}]\mathrm{d}S \tag{1}$$

式中:$\boldsymbol{n}$ 为积分面元处的法向矢量;k 为入射电磁波数量;S 为小面元的面积;$\boldsymbol{e}_\mathrm{r}$ 和 $\boldsymbol{h}_\mathrm{i}$ 分别为接收天线的电场单位矢量和发射天线的磁场单位矢量;$\boldsymbol{i}$ 和 $\boldsymbol{s}$ 分别为电磁波入射方向的单位矢量和散射方向的单位矢量;$\boldsymbol{r}'$为“源”的位置矢量。

由于式(1)是基于 PO 的单个面元 RCS 平方根的表达式,有着有限个网格的目标按相位叠加得到的总 RCS[16]为

$$\sigma_\mathrm{t} = \left|\sum_{i=1}^{n}\sqrt{\sigma_i}\right|^2 \tag{2}$$

式中:σ_i 为第 i 个网格面元的 RCS。

1.3　目标隐身性能分析方法

为使卫星具备较好的雷达隐身能力,卫星全姿态下的 RCS 需处于一个较低的范围内,这是因为雷达探测距离与目标 RCS 密切相关。雷达作用距离方程式为

$$R_{\max} = \sqrt[4]{\left[\frac{P_\mathrm{r}C_\mathrm{t}^2\lambda^2}{(4\pi)^3}\cdot\frac{1}{S_{\min}L}\right]\sigma(F_1^2F_2^2)} \tag{3}$$

式中:P_r 为接收信号的功率;C_t 为发射/接收天线功率增益;λ 为波长;$S_{\min}$为最小可探测信号;L 为发射机的内部损耗;σ 为目标雷达散射截面;F_1^2 为从发射天线到目标的方向图传播因子;F_2^2 为从目标到接收天线的方向图传播因子。

由式(3)可以看出,雷达性能和大气传播系数无法人为改变,要想降低雷达作用距离,最好的方法就是降低目标 RCS。由 $R_{\max}\propto\sigma^{1/4}$ 可知,当

RCS 每缩减 10 dBsm，就可以使雷达作用距离降低 44%；当 RCS 缩减 40 dBsm 时，雷达探测距离可降低 90%。因此，可通过缩减目标 RCS 来降低雷达探测距离，从而使目标躲避雷达探测。

除此之外，还可以选用 RCS 算术均值和 RCS 幅值来更加具体直观地分析目标的隐身性能。

$$\sigma_{\mathrm{m}} = \max\{\sigma_i\} \tag{4}$$

$$\overline{\sigma}_{[a,b]} = \frac{\sum_{i=1}^{N}\sigma_i}{N} \tag{5}$$

式中：σ_{m} 为 RCS 幅值；σ_i 为第 i 个入射角下的 RCS 值；$\overline{\sigma}_{[a,b]}$ 为 $[a,b]$ 角域内的 RCS 算术均值。

1.4 卫星散射特性计算流程

图 2 给出了本文建立的卫星模型 RCS 求解流程，主要由前处理模块、求解计算模块及后处理模块 3 部分组成。

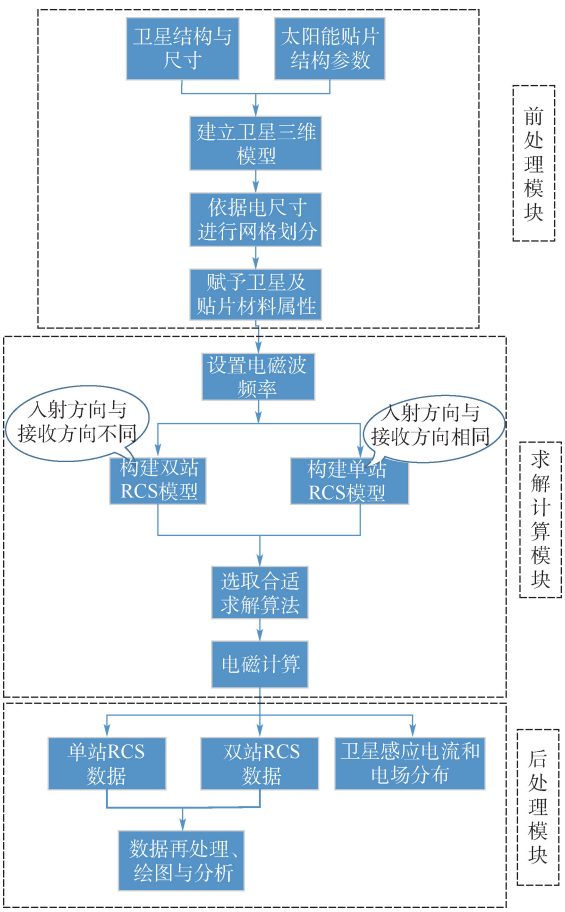


图 2 卫星电磁散射计算流程

Fig. 2 Calculation process of satellite electromagnetic scattering

2 样星微波暗室试验验证

在 C 波段（5.45 GHz）电磁波激励下的天巡一号处于光学区，具有高频散射特性，分别使用 PO 和微波暗室测试来求解天巡一号的后向雷达散射截面（单站 RCS）。由于 PO 为近似方法，在计算结果上可能会存在较大误差，需要验证 PO 数值模拟出的 RCS 的准确性。图 3 给出在上述条件下 PO 与微波暗室测试所得到的 RCS 对比结果。

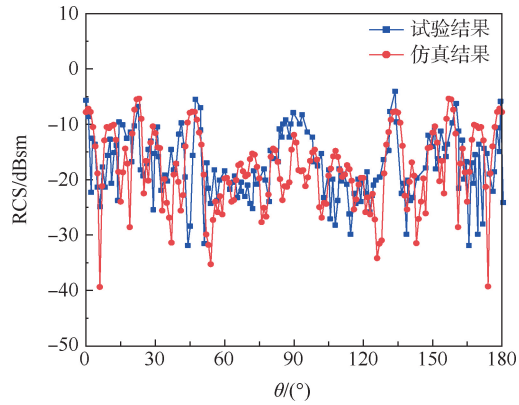


图 3 数值仿真与试验结果对比

Fig. 3 Comparison of numerical simulation and experimental results

由图 3 不难看出，采用 PO 求解出的 RCS 值与在微波暗室试验所得到的卫星 RCS 实测值吻合度较高。具体来看，在 90°为中心的（即正对卫星尖顶处）84°～96°这一对称角域内，试验结果有一明显波峰，正好对应了天巡一号尖端处。此角域内试验结果明显高于仿真结果，这是因为 PO 是采用切平面近似的方法，简单地假定被积目标表面是完全光滑平坦的，直接将表面场代入积分方程，所以 PO 并不能很好地求解尖顶散射问题。除去此误差较大的角域后，仿真和试验所得到的 RCS 均值分别为 -17.86 dBsm 和 -17.96 dBsm，两者仅相差 0.1 dBsm，证明了在光学区时采用高频近似方法来求解卫星 RCS 的准确性和可行性。

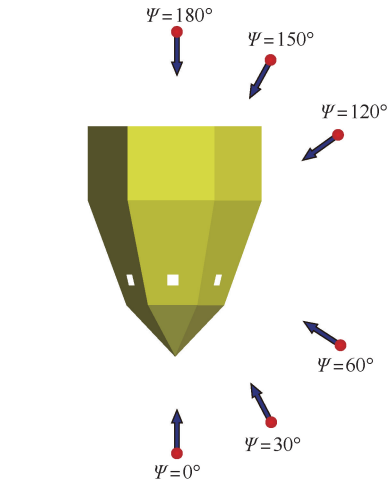
少数角域内数据还存在误差，主要是测试时卫星是用铝箔贴敷表面，难免存在缝隙、贴合不平整等问题。此外，卫星模型在测试时的姿态摆放也会存在误差，这些都是导致测试结果与仿真结果不能完全相符的主要原因。

3 天巡一号电磁散射特性分析

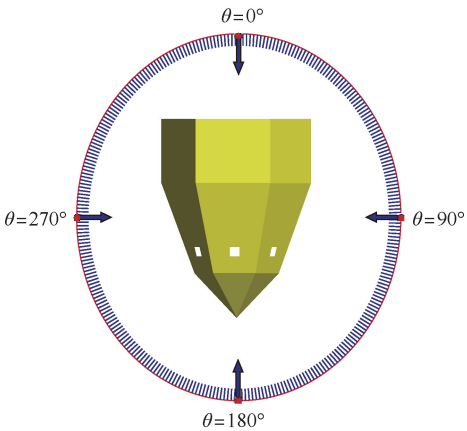
卫星隐身状态下，双站 RCS 计算模型的入射角为 0°、30°、60°，而非隐身状态下的双站 RCS 计算模型的入射角为 120°、150°、180°。单站计算模型较双站则简化许多，本文采用周向 360°入射，周向 360°接收（发射天线和接收天线处于同一位置）。具体雷达波入射示意如图 4 所示。图中：Ψ 为散射角。

3.1 RCS 入射角响应特性

图 5 给出了电磁波从地面和太空方向入射，步长均为 30°时天巡一号的双站 RCS 入射角响应



(a) 双站RCS



(b) 单站RCS

图 4　双站和单站 RCS 入射示意图

Fig. 4　Schematic of bistatic and monostatic RCS incidence

曲线分布。图中:Φ 为入射角,$0° \leqslant \Phi \leqslant 60°$代表卫星隐身姿态下电磁波入射,$120° \leqslant \Phi \leqslant 180°$代表卫星非隐身姿态下电磁波入射。

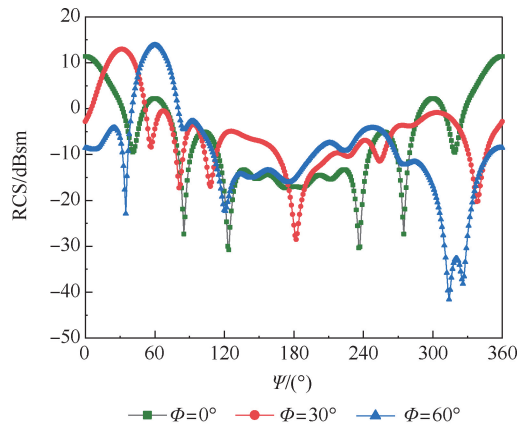
由图 5 和表 1 不难看出:

1) 当$0° \leqslant \Phi \leqslant 60°$时,双站 RCS 分布曲线的幅值出现在 0°、30°、60°处,即散射角 $\Psi = \Phi$ 时;当$120° \leqslant \Phi \leqslant 180°$时,双站 RCS 分布曲线的幅值出现在 240°、150°、180°处,即散射角 $\Psi = \Phi$ 或 $\Psi = 360° - \Phi$。

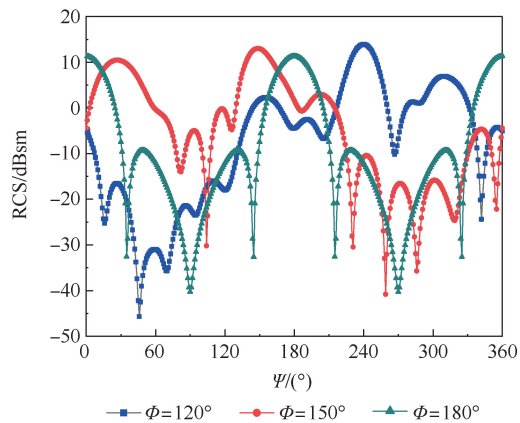
2) 在$0° \leqslant \Phi \leqslant 60°$和$120° \leqslant \Phi \leqslant 180°$散射角域内,RCS 幅值和均值相差不大,0°均值约等于 120°处均值,30°均值约等于 150°处均值,60°均值约等于 180°处均值。由此可知,天巡一号的隐身姿态对比非隐身姿态的双站 RCS 并无明显缩减,即隐身姿态不能有效应对来自双站雷达的探测和威胁。

3.2　RCS 极化响应特性

为探究极化形式对天巡一号 RCS 的影响,图 6给出了天巡一号在 4 种不同极化形式下的 RCS 极化响应特性曲线。



(a) 隐身姿态



(b) 非隐身姿态

图 5　不同入射角下双站 RCS 分布特性

Fig. 5　Characteristics of RCS distribution at bistatic with different incidence angles

表 1　不同入射角下双站 RCS 算术均值和幅值

Table 1　Average value and amplitude of RCS at bistatic with different incidence angles

入射角/(°)	RCS 算术均值/dBsm	RCS 幅值/dBsm
0	7.03	11.4
30	-5.82	12.9
60	-7.84	13.9
120	-7.84	13.9
150	-5.19	13.0
180	-9.85	11.3

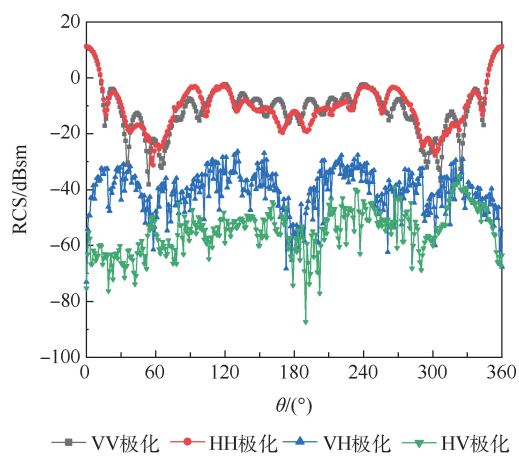


图 6　天巡一号 RCS 极化响应特性

Fig. 6　RCS polarization response characteristics of TX-1

由图 6 可以看出,同极化的单站 RCS 值要高于交叉极化 30 ~ 40 dBsm,且曲线分布更为平缓。2 种同极化(HH、VV)的单站 RCS 吻合度也较高,但交叉极化(VH、HV)之间 RCS 值差异明显。此外,同极化下的单站 RCS 在 0°(360°)处有一主瓣,但交叉极化却无明显主瓣。由以上分析可得,天巡一号的 RCS 与入射场和接收天线的极化形式有关,同极化下的 RCS 要显著高于交叉极化。

3.3 RCS 频率响应特性

图 7 和图 8 分别给出了天巡一号随频率变化下的单站 RCS 响应分布和 RCS 算术均值统计。选用的电磁波频率 f 为 1 ~ 16 GHz(涵盖了 L、S、C、X 和 Ku 波段),步长为 3 GHz,可以看出:

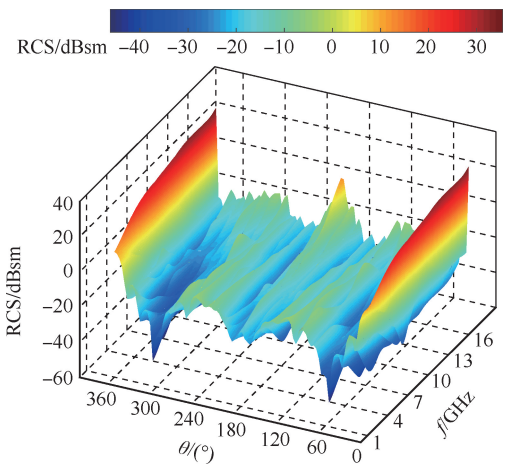


图 7 天巡一号 RCS 频率响应特性

Fig. 7 RCS frequency response characteristics of TX-1

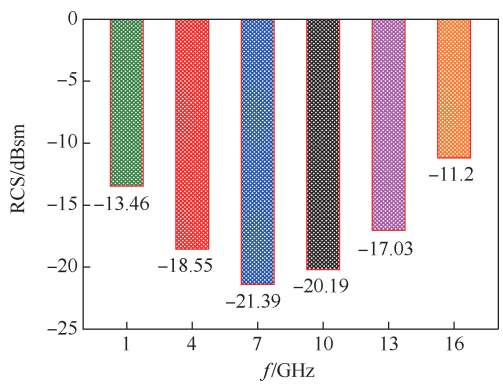


图 8 不同频率下 RCS 算术均值对比

Fig. 8 Comparison of RCS average values at different frequencies

1) 当电磁波频率从 L 波段增大至 C 波段时,RCS 算术均值逐渐降低;当电磁波频率从 C 波段增大至 Ku 波段时,RCS 均值逐渐升高。即 RCS 算术均值随频率增大会出现先减小再增大这一特点。

2) 天巡一号在全波段(指 L ~ Ku 波段)下 RCS 峰值为 35.4 dBsm(16 GHz 下散射角为 0°时),最小值低至 −45.57 dBsm(1 GHz 下散射角为 306°时),RCS 算术均值在 −11.2 ~ 21.39 dBsm,具有较低的电磁散射特征和较好的雷达隐身性能。

3) 随着入射雷达波频率的升高,在入射角为 0°(360°)和 180°处的 RCS 极大值呈现逐渐升高的趋势,但散射的波束宽度变小,响应曲面也更加尖锐,表明目标散射特性更加集中。RCS 频率特性分布曲面主要有 2 处极值,分别分布在 0°附近角域和 180°附近角域,正好对应卫星正六边形底面的镜面反射和头部尖端处的尖顶散射。

3.4 RCS 电尺寸响应特性

为探究模型电尺寸 l(实际尺寸比上工作波长)改变对目标 RCS 的影响,图 9 给出了电尺寸 l 从 3 ~ 30 变化时天巡一号 RCS 电尺寸响应分布曲线。可以看出:

1) 无论电尺寸如何变化,RCS 曲线的总体分布特性无明显改变,RCS 幅值均出现在 0°(360°)处。电尺寸较小时,RCS 曲线波动平缓;电尺寸较大时,RCS 曲线波动则明显加剧。

2) 直观来看,随着电尺寸不断增加,曲线振荡程度加剧,天巡一号 RCS 幅值、算术均值均有上升。具体来看,当目标处于谐振区,即电尺寸由 3 增大至 6 时,RCS 曲线整体略有升高但增速缓慢,波峰数量下降且散射的波束宽度也开始变小,呈现字母“W”形状;当目标处于光学区,即电尺寸由 15 增大至 30 时,RCS 分布曲线整体呈明显上升趋势且增速变快,但各波峰位置无明显改变。

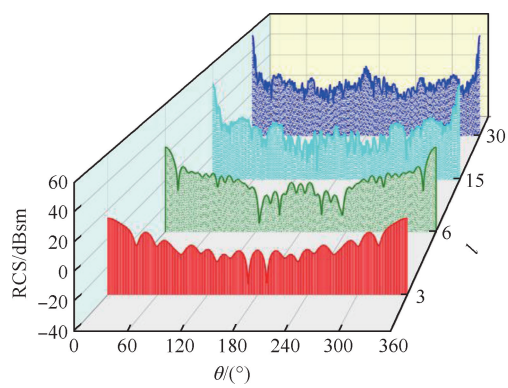


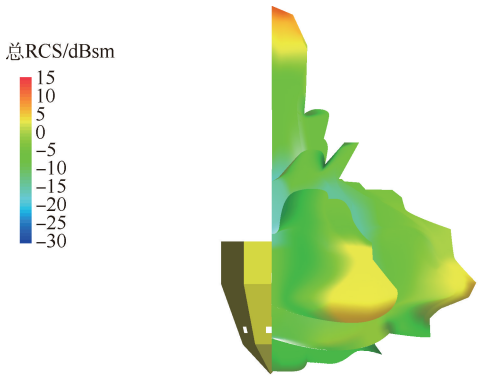
图 9 天巡一号 RCS 电尺寸响应特性

Fig. 9 RCS electric size response characteristics of TX-1

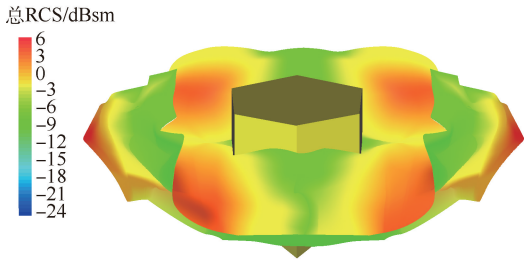
3.5 全姿态角下空间 RCS 响应特性

为进一步研究和分析天巡一号的隐身能力和强弱散射源分布情况,采用多角度雷达波组成的 3 种全姿态入射球面来照射天巡一号,3 种入射姿态分别为隐身姿态、最佳隐身姿态和非隐身姿态,入射步长均为 1°。

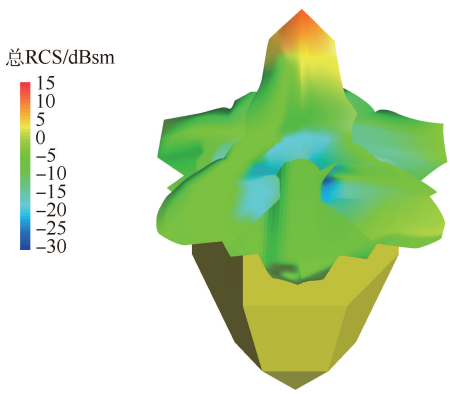
图 10 和图 11 分别给出了 3 个不同姿态角下天巡一号的空间 RCS 分布云图、RCS 算术均值和幅值对比,分析可以发现:



(a) 隐身姿态



(b) 最佳隐身姿态



(c) 非隐身姿态

图 10　天巡一号全姿态角空间 RCS 响应云图

Fig. 10　TX-1 full attitude angle RCS response cloud

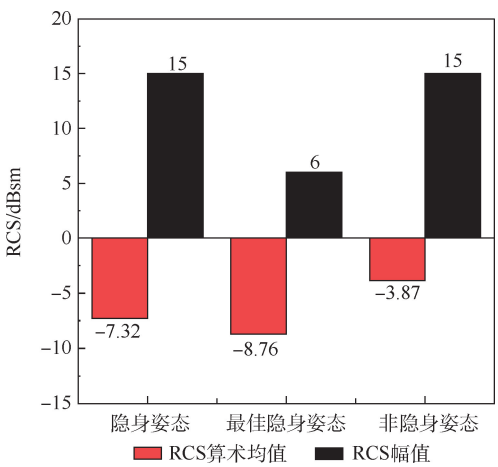


图 11　天巡一号全姿态角空间 RCS 对比

Fig. 11　TX-1 full attitude angle RCS comparison

1) 由于设计卫星构型时并未打算实现全姿态隐身,而是使卫星拥有隐身姿态和非隐身姿态,如图 10 所示,下半球面即为最佳隐身姿态,右边球面为隐身姿态,上半球面为非隐身姿态。由 RCS 云图对比可知,隐身与非隐身姿态下单站 RCS 差异显著。

2) 天巡一号的正六边形底面及与底面有公共边的 6 个矩形面是卫星的强散射源,而采用斜置外形的三角形面和梯形面为弱散射源。这主要是为了方便在底面安装对接环实现星箭分离,但光滑的正六边形底面会带来强烈的镜面反射,使卫星 RCS 升高,其他角域下天巡一号的 RCS 均处在一个比较理想的范围内。

3) 以下半球面这一最佳隐身姿态作为参考,与隐身姿态下 RCS 最大值相差 9 dBsm,但 RCS 算术均值仅相差 1.44 dBsm,即隐身姿态的隐身性能与最佳隐身姿态相差不大;与非隐身姿态下 RCS 最大值相差 9 dBsm,RCS 算术均值相差 4.89 dBsm,即非隐身姿态的隐身性能要明显低于卫星最佳隐身姿态。

4　天巡一号隐身构型优化

4.1　卫星构型优化设计方法

通过第 3 节对天巡一号隐身特性和散射源强弱的分析可知,天巡一号具有较好的隐身性能,但可通过进一步优化其隐身构型来实现更低的 RCS 和更宽的隐身角域。

图 12 为优化隐身微小卫星天巡一号构型的具体思路和方法。首先,参考天巡一号隐身构型,得到呈对称结构的四棱边尖锥构型;然后,继续增加对称尖锥的边数,至八棱边(即优化构型 A)和十棱边(即优化构型 B),如图 13 所示;最后,利用极限的思想将棱边数取到足够大,即得到最终的优化构型(橄榄体)。

4.2　构型优化方法验证

为验证图 12 中优化方法的可行性,分别选取 S 波段和 X 波段的中心频点来分析天巡一号、优化构型 A、优化构型 B 和橄榄体的隐身性能。下面先给出 S 波段(3 GHz)下 4 种隐身构型的 RCS 数值计算结果对比。

由图 14 和表 2 不难发现,在 S 波段下,最终隐身构型橄榄体在算术均值和幅值上相比较于天巡一号分别缩减了 4.77 dBsm 和 31.66 dBsm。从 RCS 算术均值出发,判断 4 种构型的隐身性能为:橄榄体 > 构型 B > 构型 A > 天巡一号,这也印证了图 12 中的构型优化方法。下面给出 4 种隐身构型在 X 波段下 RCS 分布情况。

由图 15 和表 3 不难发现,在 X 波段下,规律更加明显,即无论是从 RCS 算术均值还是 RCS 幅

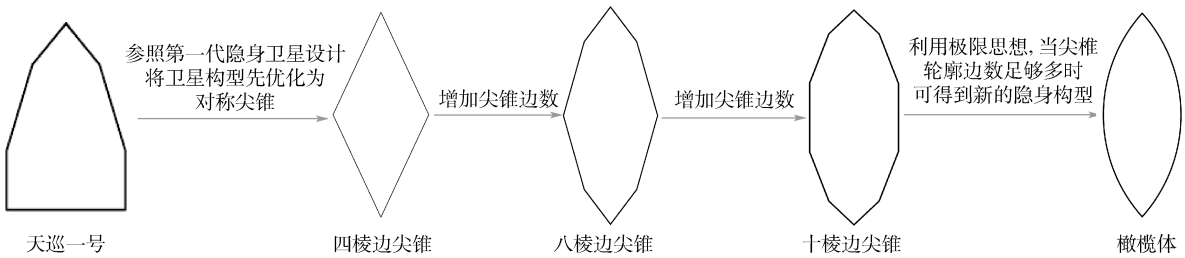
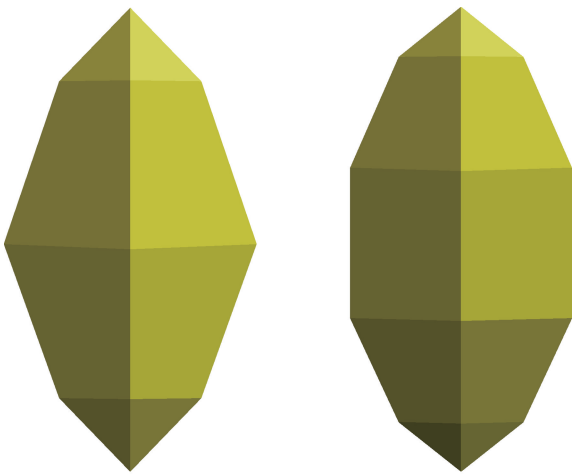


图 12 一种卫星构型优化方法

Fig. 12 A satellite configuration optimization method



(a) 优化构型A (b) 优化构型B

图 13 两种卫星优化构型

Fig. 13 Two satellite optimized configurations

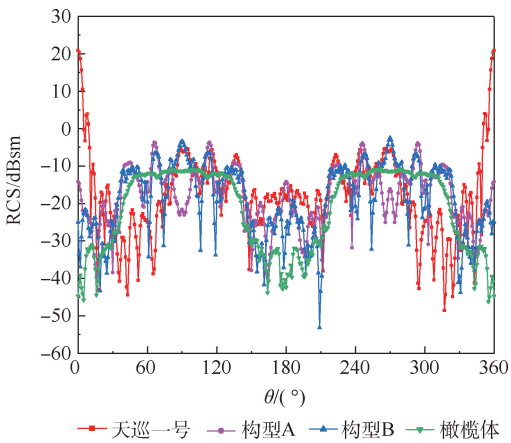


图 14 四种卫星构型 3 GHz 下 RCS 分布

Fig. 14 RCS distribution of 4 kinds of satellite configurations at 3 GHz

表 2 四种卫星构型 3 GHz 下 RCS 算术均值和幅值对比

Table 2 Comparison of average and amplitude values of RCS for 4 kinds of satellite configuration at 3 GHz

卫星构型	RCS 算术均值/dBsm	RCS 幅值/dBsm
天巡一号	-18.55	20.84
构型 A	-18.67	-3.66
构型 B	-19.82	-2.45
橄榄体	-23.32	-10.82

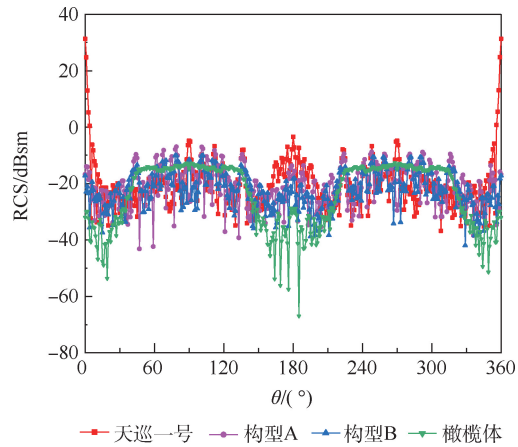


图 15 四种卫星构型 10 GHz 下 RCS 分布

Fig. 15 RCS distribution of 4 kinds of satellite configurations at 10 GHz

表 3 四种卫星构型 10 GHz 下 RCS 算术均值和幅值对比

Table 3 Comparison of average and amplitude values of RCS for 4 kinds of satellite configuration at 10 GHz

卫星构型	RCS 算术均值/dBsm	RCS 幅值/dBsm
天巡一号	-20.19	31.32
构型 A	-21.05	-7.01
构型 B	-22.30	-10.06
橄榄体	-23.84	-12.65

值来看，4 种构型的隐身性能均为：橄榄体 > 构型 B > 构型 A > 天巡一号。橄榄体构型 RCS 算术均值和幅值分别低于天巡一号 3.65 dBsm 和 43.97 dBsm，构型 A 和构型 B 也有不同程度上的 RCS 缩减，并且构型 B 的缩减效果优于构型 A。由此也再次验证了图 12 中优化设计方法的正确性和可行性。此外，橄榄体构型在以 90° 和 270° 为中心的对称角域内 RCS 分布十分平缓，反观另外 3 种构型的 RCS 曲线则振荡剧烈，存在较多散射波峰，这对于隐身构型设计来说是极为不利的。

5 结 论

为了探究天巡一号及其优化构型的电磁散射特性，本文分析了多种响应下的 RCS 分布曲线。同时，将 PO 计算出的天巡一号 RCS 和微波暗室

实测值对比,证明了采用 PO 求解是准确且合适的。在参考天巡一号隐身构型设计的基础上,构建一对称尖锥作为初始模型,在验证优化方法的可行性后,利用极限思想得到了最终的橄榄体隐身构型,并通过 RCS 对比仿真论证了橄榄体在隐身性能上的提升。

1) 对于双站雷达,天巡一号隐身姿态和非隐身姿态下的 RCS 算术均值与 RCS 幅值均相差不大,表明天巡一号的隐身姿态并不能有效应对双站雷达探测与威胁。

2) 对于单站雷达,由于正六边形底面无任何隐身设计,具有较强的镜面反射,导致其后向雷达散射截面较高。但就其隐身姿态而言,天巡一号交错斜置的多棱面外形发挥了极佳的隐身效果,可将雷达波反射至入射波所在方向以外的其他方向,从而大幅降低了单站 RCS。2 种隐身姿态下单站 RCS 算术均值分别为 -8.76 dBsm 和 -7.32 dBsm,相比较非隐身姿态 RCS 均值分别下降了 4.89 dBsm 和 3.45 dBsm。证明了天巡一号的隐身姿态在应对单站雷达探测时可发挥更好的隐身性能。

3) 对称尖锥是一种在雷达隐身上优于单侧多棱面锥体的构型设计,是本文在优化得到橄榄体卫星过程中的"桥梁"。在 10 GHz 下,4 种隐身构型的 RCS 算术均值和幅值均符合"橄榄体 < 构型B < 构型 A < 天巡一号"这一特点,表明从增加尖锥棱边数这一角度来优化隐身构型是可行的。

4) 寻求低电磁散射特性的构型是卫星隐身设计中至关重要的一环。在参考原设计(天巡一号)的基础上,通过不断优化对称锥体得到了橄榄体隐身卫星。橄榄体构型无论是在 S 波段还是 X 波段,隐身性能均有了显著提升,RCS 幅值缩减超过 30 dBsm,RCS 算术均值缩减也超过 3 dBsm。除此之外,与另外 3 种构型相比可知,橄榄体构型 RCS 曲线平缓度较好,散射波峰数量也有明显下降,这也会在一定程度上降低卫星被雷达探测系统监视识别的可能性。

本文工作为今后设计和优化微小卫星隐身构型提供了一种新的思路和参考。

参考文献 (References)

[1] 仇恒抗,杨琴,张智芳,等. 微小卫星可展式刚挠结合板太阳电池阵设计[J]. 航天器工程,2019,28(3):46-51.
QIU H K,YANG Q,ZHANG Z F,et al. Design of expandable rigid flexible combined plate solar array for microsatellite[J]. Spacecraft Engineering,2019,28(3):46-51(in Chinese).

[2] 尤政,李冠华. 多学科设计优化方法在微纳卫星总体设计中的应用[J]. 中国航天,2010(4):36-38.
YOU Z,LI G H. Application of multidisciplinary design optimization methods in the overall design of micro and nanosatellites[J]. China Aerospace,2010(4):36-38(in Chinese).

[3] HUANG H. Concept study on satellite stealth[J]. Aerospace Electronic Warfare,2010(6):009.

[4] 胡豪斌,张翔,廖文和,等. 卫星隐身技术研究进展及发展趋势[J]. 国防科技大学学报,2021,43(3):107-127.
HU H B,ZHANG X,LIAO W H,et al. Research progress and development trend of satellite stealth technology[J]. Journal of National University of Defense Technology,2021,43(3):107-127(in Chinese).

[5] 郑侃. 隐身微小卫星结构设计关键技术研究[D]. 南京:南京航空航天大学,2011.
ZHENG K. Research on the key technology of stealthy microsatellite structure design[D]. Nanjing:Nanjing University of Aeronautics and Astronautics,2011(in Chinese).

[6] 朱冬骏,张占月,赵程亮,等. 一种实现光学隐身的卫星构型设计[J]. 空间控制技术与应用,2017,43(1):61-66.
ZHU D J,ZHANG Z Y,ZHAO C L,et al. A satellite configuration design to achieve optical stealth[J]. Space Control Technology and Applications,2017,43(1):61-66(in Chinese).

[7] ZENG H T,ZHAO X W,SU Q,et al. Fast coating analysis and modeling for RCS reduction of aircraft[J]. Chinese Journal of Aeronautics,2019,32(6):1481-1487.

[8] 刘战合,苗楠,王菁,等. 不同气动布局轰炸机电磁隐身性能对比研究[J]. 科学技术与工程,2020,20(23):9640-9646.
LIU Z H,MIAO N,WANG J,et al. Comparative study on electromagnetic stealth performance of bombers with different aerodynamic layouts[J]. Science,Technology and Engineering,2020,20(23):9640-9646(in Chinese).

[9] YI M X,WANG L F,HUANG J. Active cancellation analysis based on the radar detection probability[J]. Aerospace Science and Technology,2015,46:273-281.

[10] ZHOU Z Y,HUANG J,YI M X. Comprehensive optimization of aerodynamic noise and radar stealth for helicopter rotor based on Pareto solution[J]. Aerospace Science and Technology,2018,82-83:607-619.

[11] 桑建华. 飞行器隐身技术[M]. 北京:航空工业出版社,2013.
SANG J H. Stealth technology of flying vehicles[M]. Beijing:Aviation Industry Press,2013(in Chinese).

[12] 雷浩. 物理光学算法在电大尺寸目标电磁散射中的应用[D]. 西安:西安电子科技大学,2018.
LEI H. Application of physical optics algorithms in electromagnetic scattering of electrically large targets[D]. Xi'an:Xidian University,2018(in Chinese).

[13] 吴安雯,吴语茂,杨杨,等. 矩量法-物理光学混合算法计算多尺度复合目标电磁散射场[J]. 电波科学学报,2019,34(1):83-90.
WU A W,WU Y M,YANG Y,et al. The method of moments physical optics hybrid algorithm to calculate the electromagnetic scattering field of multiscale compound targets[J]. Chinese Journal of Radio Science,2019,34(1):83-90(in Chinese).

[14] 张策. 物理光学法及其迭代加速算法在电磁散射中的应用

[D]. 西安:西安电子科技大学,2019.
ZHANG C. Application of physical optics method and its iterative acceleration algorithm in electromagnetic scattering[D]. Xi'an: Xidian University, 2019(in Chinese).

[15] 岳奎志,孙聪,姬金祖. 双立尾对战斗机隐身特性的数值模拟[J]. 北京航空航天大学学报,2014,40(2):160-165.
YUE K Z, SUN C, JI J Z. Numerical simulation of the stealth characteristics of fighter jets with twin vertical tails[J]. Journal of Beijing University of Aeronautics and Astronautics, 2014, 40(2): 160-165(in Chinese).

[16] 肖厚地,刘龙斌,吕明云. X型尾翼临近空间飞艇隐身特性仿真[J]. 北京航空航天大学学报,2015,41(1):181-186.
XIAO H D, LIU L B, LV M Y. Simulation of stealth characteristics of X-tail airship near space[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41(1): 181-186(in Chinese).

Stealthy configuration design and optimization analysis of microsatellite

QIN Yuantian[1,*], SUN Hanqing[1], YUE Xin[2]

(1. College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210093, China;
2. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210093, China)

Abstract: To study the electromagnetic scattering characteristics of the TX-1 microsatellite, an electromagnetic calculation model with a stealthy shape is established. The physical optics (PO) is used to numerically calculate the radar cross section (RCS) under different conditions to verify PO accuracy through comparing with anechoic chamber experimental results. This serves as a foundation for the development of the satellite RCS incidence angle, polarization, frequency, electric size response characteristics and full attitude angle spatial RCS response characteristics. Finally, the configuration is optimized to a symmetric pointed cone configuration referring to the TX-1's stealthy design. By increasing the quantity of sharp cone edges to optimize the configuration, the olive-shaped satellite with a lower RCS configuration is obtained. The TX-1's stealthy trait can effectively dispel the monostatic radar threat. The spatial RCS average value in the best stealth attitude is 4.89 dBsm less than that in the non-stealth attitude. Under S-band (3 GHz), the RCS average value and RCS amplitude of the olive-shaped satellite are 4.77 dBsm and 31.66 dBsm lower than those of TX-1. Under X-band (10 GHz), the RCS average value and RCS amplitude of the olive-shaped satellite are 3.65 dBsm and 43.97 dBsm less than those of TX-1.

Keywords: microsatellite; radar cross section; configuration optimization; physical optics; stealth

Received: 2021-07-13; **Accepted**: 2021-11-05; **Published online**: 2021-11-30 11:39
URL: kns.cnki.net/kcms/detail/11.2625.V.20211129.1538.001.html

* **Corresponding author**. E-mail: qinyt@nuaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0083

高速气膜镶装式浮环密封的开启特性

郑娆, 陈潇竹, 李双喜*, 赵祥, 施任杰, 宋子锋
(北京化工大学 机电工程学院, 北京 100029)

摘 要: 针对应用于航空发动机的高速气膜镶装式浮环密封,探究密封在不同结构参数、启动方式、材料结合等多因素下的开启性能。建立镶装环-石墨环-跑道的固体域模型和气膜流体域模型,得到了工作气膜厚度、气膜流场压力分布;计算密封受力,得到了密封上浮力、闭合力和开启转速等密封开启性能参数。分析镶装环与石墨环的厚度比和宽度比、镶装环-石墨环配对材料、镶装环-跑道配对材料等因素对密封开启转速的影响。搭建浮环密封试验台和浮环位移监测系统,通过试验验证了数值模拟结果。研究结果表明:浮环的镶装结构能有效改善升温时石墨环与跑道间隙减小而导致密封失效的问题;镶装环材料是影响密封开启性能的敏感参数,密封开启性能随材料线膨胀系数的升高而快速降低;镶装环与跑道的材料配对情况是影响密封开启性能的重要因素,镶装环与跑道材料相同时,石墨环与跑道处于"恒间隙"状态,在复杂温度工况下密封的开启性能更加稳定;不同的工作机组启动方式对密封开启性能影响较大,发动机采用将转速增加至工况转速再增压的启动方式时密封开启性能最好;浮环密封在高转速、高压力工况下因转速、压力变化产生的密封扰动值更大,浮环密封应避免在较高压力和转速下长时间调节工况参数。研究结果为航空发动机镶装式浮环密封的结构设计、材料选用、系统设计的研究提供了参考。

关 键 词: 轴承腔密封; 浮环密封; 开启特性; 恒间隙; 运行稳定性
中图分类号: V219; TB42
文献标志码: A **文章编号**: 1001-5965(2022)11-2111-10

高速气膜镶装式浮环密封作为一种非接触式流体动密封,因其结构简单、安全可靠而被运用于直升机发动机轴承腔密封。镶装式浮环密封的开启性能包括密封开启转速的高低和开启时浮环运行稳定性的优良。由于气体黏度小,相比于液膜来说动压效果较弱,气膜浮环密封的开启性能往往较差。当密封的结构、材料及发动机启动方式在设计选择不合理时,密封的开启难度加大,导致磨损加剧,进而影响航空发动机的安全性。

为研究浮环密封的开启性能,Childs 等[1-2]假定浮环锁死偏心率,研究了周向槽型、深度、宽度等因素对浮环动力学系数的影响。Xia 等[3]研究发现浮动环密封件的弹性变形对密封件的泄漏流量和转子动力学系数有很大影响。San Andres 和 Semanate[4]为得到不同工况的动力学系数,利用改进了浮环密封的数值模型,得到了精度更高的结果。Berger[5]研究了工程模型领域的各种摩擦模型,讨论了用于动态系统仿真的摩擦建模方式。马也等[6]针对航空发动机主轴承箱的气膜浮环密封系统提出了一种开启转速计算方法,分析了先增压后增速和先增速后增压 2 种情况下工况参数对浮环开启转速的影响。马利军等[7-8]利用

收稿日期: 2021-02-19; **录用日期**: 2021-07-03; **网络出版时间**: 2021-08-25 20:13
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210825.1244.003.html
基金项目: 国家重点研发计划 (2018YFB2000800)
* **通信作者**. E-mail: buctlsx@126.com

引用格式: 郑娆, 陈潇竹, 李双喜, 等. 高速气膜镶装式浮环密封的开启特性[J]. 北京航空航天大学学报, 2022, 48(11): 2111-2120. ZHENG R, CHEN X Z, LI S X, et al. Opening characteristics of inlaid floating ring seal with high-speed gas film[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2111-2120(in Chinese).

ANSYS Workbench 分析软件,建立了镶装式浮环密封数值计算模型和气膜数值分析模型,分析了工况参数、结构和材料对径向间隙的变化规律,以及不同条件下气膜浮环密封的上浮力、泄漏量和实际偏心距等特性参数的变化规律。夏鹏和刘占生[9]基于 Bulk flow 模型和位移-位移耦合模式建立了密封间隙流体激励和浮环结构耦合模型,研究证明了外部高压作用下,浮环间隙高度减小,入口度降低,交叉刚度相对增大。刘占生等[10]基于有限单元法求解了瞬态流体动压力,利用摄动法求解了锁死条件下浮环密封动力学系数。马纲等[11]提出了一种浮环密封系统的准动态特性计算分析方法,采用有限元法实现了密封气膜力学特性和密封稳态特性的计算仿真,从而获得了反转轴间大柔性支撑浮环柱面气膜密封系统的主要密封特性变化曲线。Arghir 等[12]建立了一种较为简单的浮环力学分析模型,并对力学模型中的上浮力提出近似的求解方法。在 Shapiro 等[13]对气膜润滑、扇形浮环密封的研究基础上,王飞等[14]采用有限差分与牛顿超松弛迭代求解雷诺方程的方法,对浅槽分瓣环型浮动环密封的密封间隙内流场压力的分布状况进行了数值模拟,研究间隙对泄漏量、静态刚度的影响。许庆余等[15]依据短轴承动力润滑理论推导出浮环油膜力分析表达式,并应用 Treasor 数值积分法,解决了浮环运动微分方程严重“病态”问题。

已有工作对于浮环密封的研究主要集中于理论分析及工况参数对浮环密封开启性能的影响,试验方面研究较少。在此基础上,本文从镶装式浮环密封的结构和材料等参数出发,采用数值模拟和试验结合的方法研究其密封开启性能,揭示镶装式浮环密封在各种结构、材料配对及发动机开启方式下密封开启性能的变化规律。

1 结构与工作原理

1.1 密封结构

镶装式浮环密封主要由石墨环、镶装环、跑道、波形弹簧和密封腔体组成。密封结构如图 1 所示。密封主体为具有自润滑效应且耐热耐腐蚀的石墨环,但石墨环在受到载荷后容易发生脆性断裂,需要在石墨环外镶装金属环以增加密封整体结构的强度。镶装式浮环结构控制间隙原理如图 2 所示。在航空发动机的高温环境下,由于石墨环的线膨胀系数低于金属跑道线膨胀系数,导致跑道直径变化值 c_2 大于石墨环直径变化值 c_1,

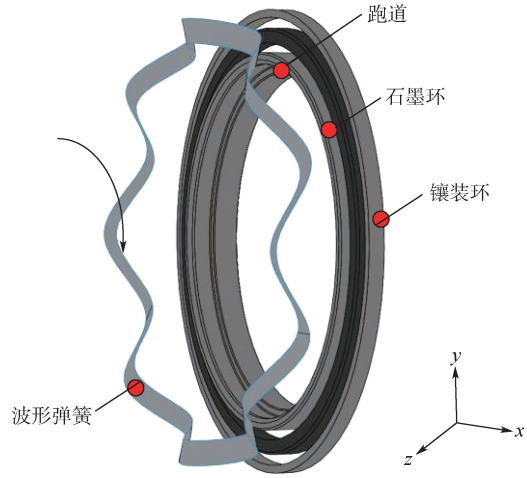


图 1 镶装式浮环密封结构

Fig. 1 Structure of inlaid floating ring seal

石墨环与跑道间隙 δ_R 减小,使其偏离设计值进而发生碰摩。使用金属镶装环可以补偿石墨环的间隙变化,其原理为:在石墨环镶装的过程中会发生压缩变形 t,原本的石墨环厚度由 h 变为 $h-t$。受热膨胀时,石墨环会膨胀,金属镶装环膨胀,同时会释放一部压缩变形 k。因此,当带有镶装环的石墨环在受热时所产生的变形 c_1 大于纯石墨环受热所产生的变形,改善石墨环与跑道之间的间隙随温度增大而大幅度变小的情况。同时控制镶装石墨环与跑道的初始间隙和波形弹簧弹力[8],实现在一定温度范围内的恒间隙并保证密封运行安全性。因此,镶装式浮环同时具有石墨环和金属环的优点,使其在工作状态下既不会脆性断裂又不会与跑道发生剧烈碰摩。波形弹簧使镶装式浮环与压盖紧密接触,从而阻止密封的径向泄漏,并防止镶装式浮环的轴向窜动。

1.2 工作原理

镶装式浮环密封在静止状态下,由于初始间隙和重力的原因使石墨环与跑道相互接触,密封处于关闭状态,当密封开始运转后,跑道旋转所产生的气膜动压力和气体流动对浮环间隙所产生的静压力,使得镶装式浮环能克服其重力与端面摩擦力,最终使密封处于开启状态。

密封的工作间隙大小影响密封开启性能[6],镶装式浮环密封仍需考虑密封的初始间隙。当初始间隙过大时,密封不会发生碰摩,但密封泄漏量较大;当初始间隙过小时,密封泄漏量较小,但密封可能会因温度的升高导致石墨环与跑道发生剧烈碰摩,密封失效。

1.3 开启性能参数

1.3.1 密封上浮力

在竖直方向上,密封上浮力 F_o 由跑道旋转产

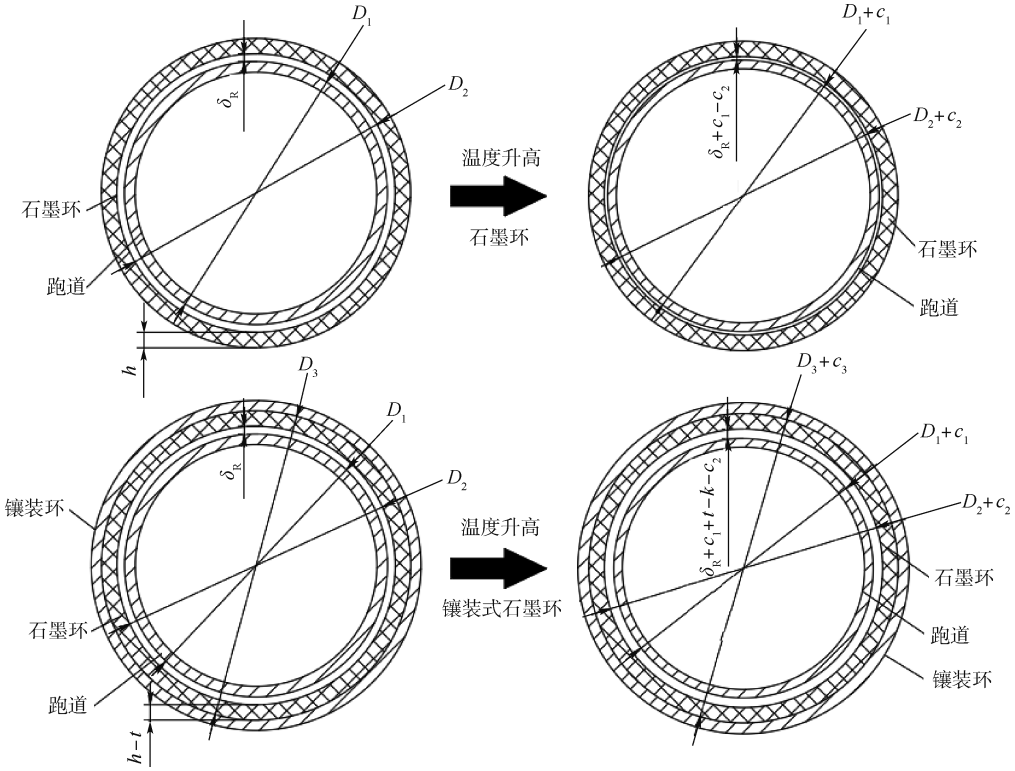


图 2　镶装式浮环结构间隙控制原理

Fig. 2　Clearance control principle of inlaid floating ring

生的气膜动压力所产生,如式(1)所示。浮环受力情况如图 3 所示。

$$\begin{cases} F_1 = \int_L^0 \int_0^{2\pi} (P_1 - P_2) R\cos\theta \mathrm{d}\theta \mathrm{d}z \\ F_2 = \int_L^0 \int_0^{2\pi} (P_1 - P_2) R\sin\theta \mathrm{d}\theta \mathrm{d}z \\ F_o = \sqrt{F_1^2 + F_2^2} \end{cases} \tag{1}$$

式中:F_1、F_2为密封上浮力分力,N;P_1为高压侧压力,MPa;P_2为低压侧压力,MPa;R 为石墨环内半径,mm;θ 为力的方向角度,(°)。

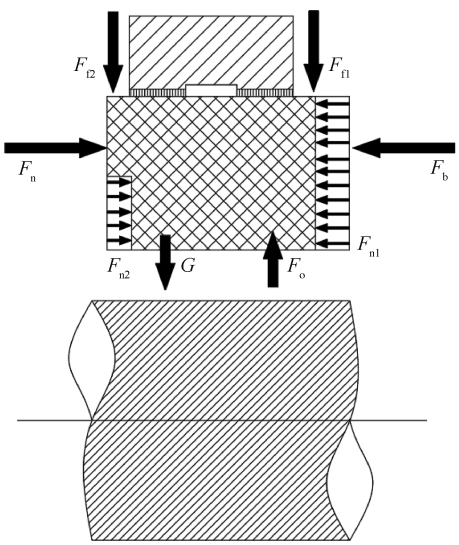


图 3　镶装式浮环受力分析

Fig. 3　Force analysis of inlaid floating ring

1.3.2　密封闭合力

在镶装式浮环密封中,密封闭合力 F_c主要由石墨环与腔体的摩擦力 F_f和镶装式浮环重力 G 组成,如下:

$$\boldsymbol{F}_c = \boldsymbol{F}_f + \boldsymbol{G} \tag{2}$$

轴向方向上,端面压紧力 F_n与波形弹簧的平衡情况如图 3 所示,可以得出端面压紧力计算式为

$$F_n = F_{n1} + F_b - F_{n2} \tag{3}$$

式中:F_{n1}为高压侧接触压力;F_{n2}为低压侧接触压力;F_b为波形弹簧弹力。

镶装式石墨环两侧的摩擦力 F_f由石墨环两边的接触力端面压紧力 F_n和波形弹簧的弹力 F_b计算得出,如下:

$$\begin{cases} F_{f1} = \mu F_b \\ F_{f2} = \mu F_n \\ F_f = F_{f1} + F_{f2} \end{cases} \tag{4}$$

式中:μ 为石墨与合金钢的摩擦系数,本文取 0.07。

1.3.3　开启转速

开启转速为密封开启性能的指标之一,密封的开启力随转速的增加而增加,当密封的上浮力与闭合力相等时,此时的转速即为开启转速。

当密封上浮力 F_o等于密封闭合力 F_c时,密封开启,如下:

$$\boldsymbol{F}_o = \boldsymbol{F}_c = \boldsymbol{F}_f + \boldsymbol{G} \tag{5}$$

考虑密封闭合力最大即摩擦力和重力在同一

直线上时,将式(3)、式(4)代入,即可得

$$F_o = (F_{n1} + 2F_b - F_{n2})\mu + G \tag{6}$$

反之,当密封上浮力 F_o 小于密封闭合力 F_c 时,密封无法开启,如下:

$$F_o < (F_{n1} + 2F_b - F_{n2})\mu + G \tag{7}$$

式中:$F_{n1} = P_1A_1$;$F_{n2} = P_2A_2$;A_1 为高压侧密封端面面积,m^2;A_2 为低压侧密封端面面积,m^2。

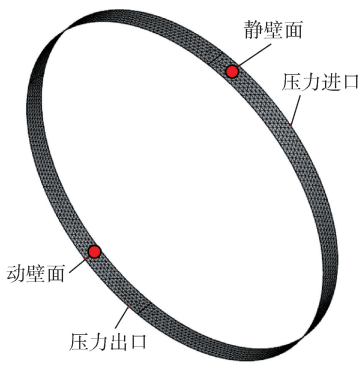
1.3.4 浮环运行稳定性

浮环运行稳定性为密封开启性能的指标之一,由浮环运行状态监测系统测得。浮环在开启时的位移量 s 的大小代表浮环运行稳定性,浮环位移量大,则表示浮环运行稳定性差,反之,表示浮环运行稳定性好。

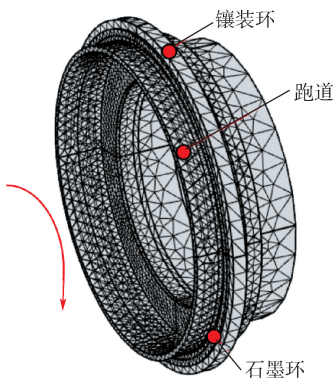
2 数值模拟

2.1 几何模型

利用 ANSYS Workbench ICEM 模块建立镶装环-石墨环-跑道固体域模型和气膜流体域模型,同时利用 ANSYS Workbench 中的 Mesh 模块进行网格划分。固体域模型和气膜流体域模型及其网格分布情况如图 4(a)、(b)所示。气膜流体域模型经网格无关化验证,当网格节点数为 196 700,



(a) 气膜流体域模型网格划分



(b) 固体域模型网格划分

图 4 固体域模型和气膜流体域模型及其网格分布

Fig. 4 Solid domain model and gas film fluid domain model and their grid distribution

网格单元数为 178 900,网格平均边长取 0.05 mm 时,网格精度与计算时间相平衡,此时计算结果较为精确。

固体域模型和气膜流体域模型的结构参数与工况参数如表 1 和表 2 所示。

表 1 结构参数

Table 1 Structural parameters

参数	数值
镶装环外径/mm	91
镶装环内径/mm	$86.5_0^{+0.035}$
石墨环外径/mm	$86.5_{+0.545}^{+0.58}$
石墨环内径/mm	$80.6_{+0.15}^{+0.172}$
镶装环宽度/mm	3.5
石墨环宽度/mm	5.5
跑道直径/mm	$80.6_{-0.156}^{-0.14}$

表 2 工况参数

Table 2 Operating parameters

参数	数值
进口压力/MPa	0 ~ 0.06
出口压力/MPa	0
气体温度/K	532
气体黏度/(Pa · s)	1.82×10^{-5}
转速/(r · min^{-1})	0 ~ 26 000

2.2 计算流程

高速气膜镶装式浮环密封开启性能数值计算流程为:镶装环-石墨环-跑道固体域计算;实际密封间隙计算;气膜流场的开启转速计算。

建立镶装环-石墨环-跑道的固体域模型,对其施加温度场和转速,分别计算镶装式浮环与跑道之间的内外径变化,通过式(8) ~ 式(13)计算流体域气膜厚度。根据固体域计算结果建立气膜流体域模型,通过施加压力、转速、温度等边界条件计算镶装式浮环的上浮力,并与密封闭合力进行比较。若密封上浮力等于密封闭合力,此时的转速则为密封开启转速;反之,则重新设置转速计算,直至二者相等。通过计算得到的固体域模型的石墨环内径变化与跑道外径变化情况如图 5 所示。数值模拟流程如图 6 所示。

$$\delta_{Rmax} = (c_{max} - e_{min})/2 \tag{8}$$

$$\delta_{Rmin} = (c_{min} - e_{max})/2 \tag{9}$$

$$S_{max} = \delta_{Rmax} + \delta_{Rhuan} - \delta_{Rpao} \tag{10}$$

$$S_{min} = \delta_{Rmin} + \delta_{Rhuan} - \delta_{Rpao} \tag{11}$$

$$D_{qm} = D_1 + \delta_{Rmax} + \delta_{Rmin} + 2\delta_{Rhuan} \tag{12}$$

$$d_{qm} = D_1 + 2\delta_{Rpao} \tag{13}$$

式中:c_{max} 为石墨环内径上偏差;c_{min} 为石墨环内径下偏差;e_{max} 为跑道外径上偏差;e_{min} 为跑道外径下

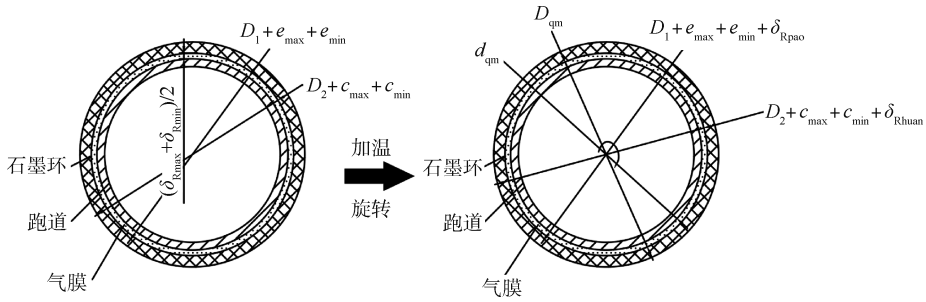


图 5　密封间隙变化

Fig. 5　Variation of sealing clearance

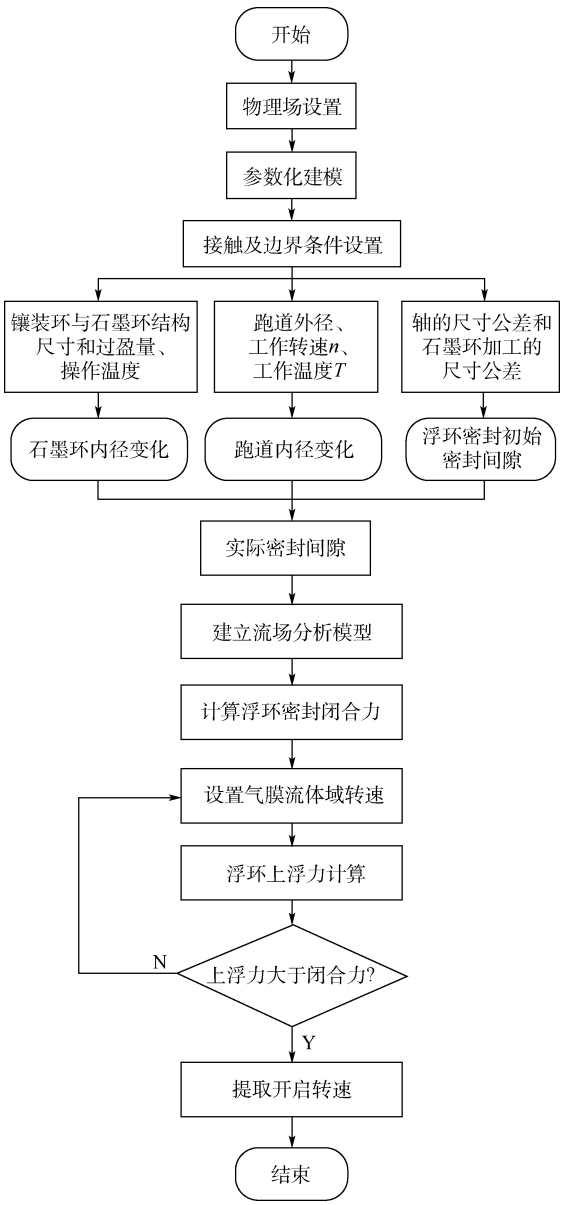


图 6　数值模拟流程

Fig. 6　Flow chart of numerical simulation

偏差;δ_{Rmax}为初始最大半径间隙;δ_{Rmin}为初始最小半径间隙;δ_{Rhuan}为石墨环内径变化值;δ_{Rpao}为跑道外径变化值;S_{max}为动态最大半径间隙;S_{min}为动态最小半径间隙;D_{qm}为气膜外径值;d_{qm}为气膜内径值。

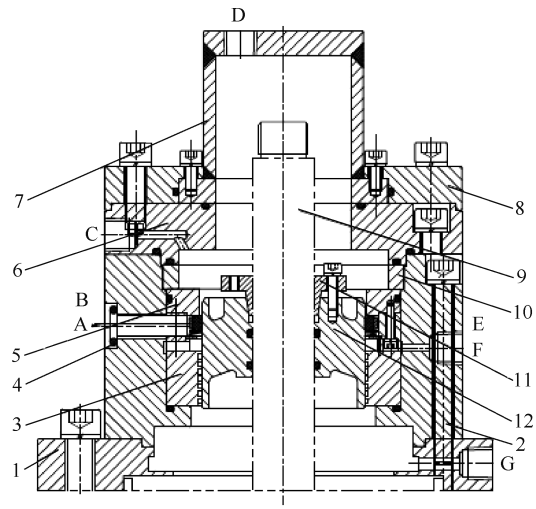
3　密封开启特性试验

3.1　试验装置与测量仪器

镶装式浮环密封开启特性试验在浮环密封立式试验台上进行,如图 7 所示。该试验台兼具机械系统、控制系统及数据采集系统。试验台的机械系统由试验台架、供气系统、油气润滑冷却系统、气体加热系统等组成。

试验台的控制系统由基于 LabView 的自制软件、数据采集卡及转速传感器、压力传感器、温度传感器制作而成,用于控制试验台的转速及设定密封的进气压力和温度。试验台进气压力的控制范围为 0 ~ 6 MPa,转速的控制范围为 0 ~ 20 000 r/min,温度的控制范围为 20 ~ 300℃。

数据采集系统由高温气体流量计、浮环位移监测系统及光学三维影像仪组成。同时,在密封腔和密封腔盖上开设了不同功能作用的管孔用于进行测量和检测。管孔排布情况如表 3 所示,表中字母与图 7 对应。浮环位移监测系统由前置器、延长线、数据线、数字显示器、显示器连接线及



1—过渡板;2—密封腔体;3—迷宫环;4—O形圈;
5—浮环密封组件;6—密封腔法兰;7—密封腔盖;
8—腔盖;9—轴;10—压紧环;11—胀紧套;12—跑道。

图 7　试验装置结构

Fig. 7　Structure of test device

表 3 试验装置管口释义

Table 3 Interpretation of nozzles of test device

管口代号	管口名称	管口作用
A	位移监测口 1	监测浮环受扰动的位移
B	进气口	提供高压高温气源
C	压力检测口 1	检测浮环低压侧压力
D	空气出口	检测浮环泄漏率
E	压力检测口 2	检测浮环高压侧压力
F	空气入口	提供气源
G	油雾润滑口	通入油雾

VIBROTEST80 电涡流传感器搭建,用于对浮环在运行状态下的位移情况进行监测测量。

3.2 试件制备与试验参数

试验所用的镶装式浮环为笔者团队自制,制备过程如下:首先,将镶装所需金属环放入工装的凹槽中,并放置于恒温加热台加热;然后,加热到给定温度,将石墨环装进工装的固定套中;最后,将石墨环压进金属环,待温度降低后镶装石墨环制作完毕。

试验采用数套自制的镶装式浮环密封,试验密封介质为压缩空气,试验参数如表 4 所示。

表 4 试验参数

Table 4 Test parameters

材料			结构参数		工况参数	
镶装环	石墨环	跑道	宽度比	厚度比	压力/MPa	温度/℃
S30408	CDJ-83	S30408	0 ~ 1.0	0 ~ 1.0	0 ~ 0.06	20 ~ 300
S30408	MAT679	S30408	0 ~ 1.0	0 ~ 1.0	0 ~ 0.06	20 ~ 300
TC4	MAT679	TC4	0 ~ 1.0	0 ~ 1.0	0 ~ 0.06	20 ~ 300
GH4169	EY308	GH4169	0 ~ 1.0	0 ~ 1.0	0 ~ 0.06	20 ~ 300

3.3 试验步骤

步骤 1 选择不同的镶装环和石墨环材料配对进行镶装。

步骤 2 将镶装式浮环密封装入试验台中,根据试验前预设的工况参数调节密封供气压力和温度,并启动油气润滑系统。

步骤 3 启动电机,增加电机转速,观察浮环位移监测系统显示数值并实时记录。

步骤 4 待浮环位移监测系统显示数值稳定后,记录此时的电机转速为密封开启转速。

步骤 5 给予密封转速和压力的扰动,记录浮环位移的变化值。

步骤 6 根据试验研究对象重复步骤 1 ~ 步骤 5。

4 开启转速影响因素分析

结合数值模拟与试验结果,分析镶装环与石墨环厚度比、镶装环与石墨环宽度比、镶装环-石墨环配对材料、镶装环-跑道配对材料、发动机启动方式对石墨环开启性能的变化规律研究。

4.1 镶装环与石墨环厚度比

选用模型的镶装环与石墨环宽度比为 0.5,材料选用 S30408-MAT679-S30408 的镶装环-石墨环-跑道材料配对形式,同时选取 0 ~ 1.0 镶装环与石墨环厚度比进行数值计算和试验,得到镶装环与石墨环厚度比-开启转速关系,如图 8 所示。

试验与数值模拟分析的开启转速随镶装环与石墨环厚度比变化趋势相同,都随镶装环与石墨环厚度比的增加先快速增加后增长速度逐渐降低,最大误差为 7.63%,在工况温度较低时,厚度比对密封开启影响较小,随着工况温度的升高,厚度比对密封开启的影响逐渐增大。研究表明,镶装环与石墨环的厚度比主要影响密封环的接触应力,接触应力随厚度比的增加而增加,两环接触应力的增大使密封难以开启。由于较小的厚度比会使金属镶装环过薄,从而影响密封环的结构强度,镶装环与石墨环的厚度比应选择 0.4 ~ 0.6 为宜。同时,根据开启转速的变化值随厚度比的变化最大为 4 000 r/min,以及受实际加工难度因素影响,镶装环与石墨环厚度比对密封开启性能的影响较小。

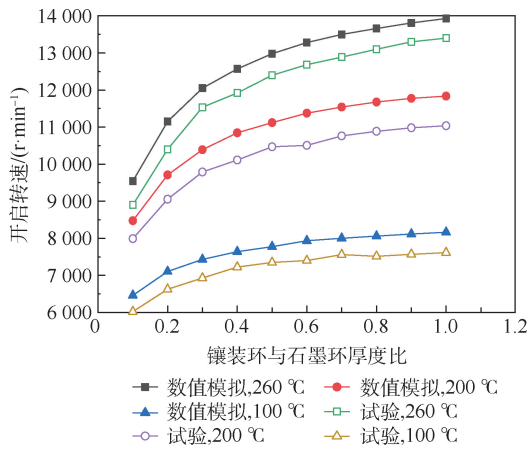


图 8 镶装环与石墨环厚度比和开启转速关系

Fig. 8 Relationship between inlay-graphite ring thickness ratio and opening speed

4.2 镶装环与石墨环宽度比

选用模型的镶装环与石墨环厚度比为 0.5,材料选用 S30408-MAT679-S30408 的镶装环-石墨环-跑道材料配对形式,同时选取 0 ~ 1.0 镶装环

与石墨环宽度比进行数值计算和试验。在工况温度较低(50 ~ 100℃)和镶装环与石墨环宽度比较大(0.5 ~ 0.9)时,宽度比对密封开启转速不敏感,随宽度比的增大而增加,但增加幅度较小;在工况温度较高(200 ~ 260℃)和镶装环与石墨环宽度比较小(0.1 ~ 0.4)时,宽度比对密封开启转速敏感,随宽度比的增大而快速增加,如图 9 所示。同样的,镶装环与石墨环的宽度比主要影响密封环的接触应力,两环的接触应力随厚度比的增加而增加,接触应力增大使密封难以开启。且由于过大的宽度比会导致镶装环与石墨环接触应力过大,石墨环破碎,过小的宽度比会导致镶装环过薄,影响其结构强度。因此,镶装环与石墨环宽度比应选择 0.4 ~ 0.6 为宜。试验值与数值模拟变化规律相同,最大误差为 8.7%,在可接受范围内。同样的,镶装环与石墨环宽度比对密封开启性能的影响也较小。

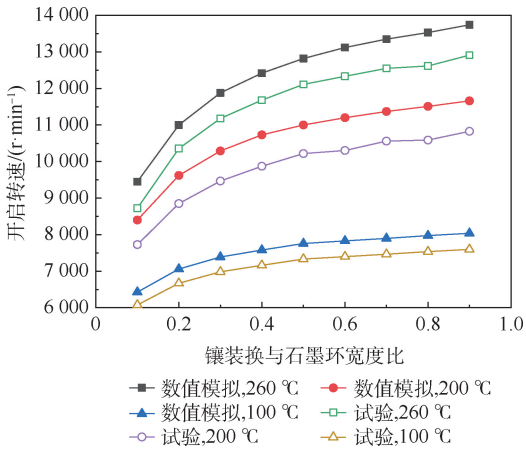


图 9　镶装环与石墨环宽度比和开启转速关系

Fig. 9　Relationship between inlay-graphite ring width ratio and opening speed

4.3　镶装环-石墨环配对材料

选取 3 种石墨环材料 CDJ-83、MAT679、EY308 与 3 种镶装环材料 TC4、GH4169、S30408 进行配对,镶装环-石墨环配对材料对密封开启转速的影响如图 10 所示。

由图 10 可知,不同配对材料对开启转速的影响较大,且使用 3 种不同石墨环材料的密封环的开启转速大小为:CDJ-83 < MAT679 < EY308;3 种不同镶装环材料的密封环开启转速大小为:TC4 < GH4169 < S30408,且镶装环材料对开启转速的影响比石墨环材料影响更大。这是因为:3 种石墨环材料的弹性模量 CDJ-83 > MAT6779 > EY308,石墨环材料的弹性模型影响了密封间隙,间接影响了密封开启转速,密封的开启转速随石墨环材料的弹性模量降低而增大。同样的,3 种镶装环材料的线膨胀系数 TC4 < GH4169 < S30408,镶装环的线膨胀系数影响了密封间隙,间接影响了密

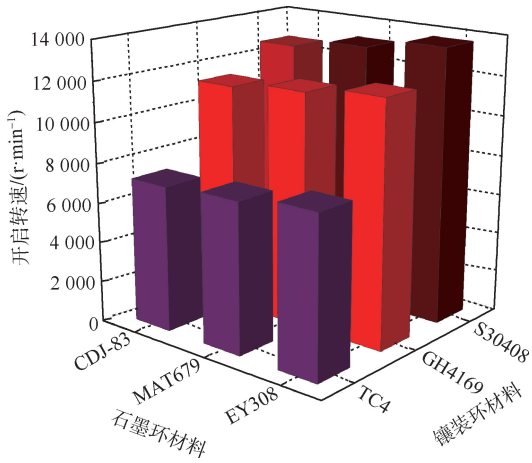


图 10　镶装环-石墨环配对材料和开启转速关系

Fig. 10　Relationship between matching material of graphite-inlay ring and opening speed

封开启转速,密封的开启转速随镶装环材料的线膨胀系数的增加而增大。因此,在实际运用中应选择线膨胀系数较小的镶装环材料和弹性模量较大的石墨环材料进行配对组合。

4.4　镶装环-跑道配对材料

选取 TC4、S30408、GH4169 三种金属材料作为镶装环和跑道材料进行配对。镶装环-跑道配对材料对密封开启转速的影响如图 11 所示。

镶装环-跑道配对材料对密封开启转速的影响主要来源为工况温度对密封开启转速的影响。通过影响镶装式浮环与跑道间的密封间隙,密封开启转速随着密封间隙的增大而升高。工况温度对密封开启转速的次要影响因素为工况温度的升高使空气黏度增大,但空气黏度对密封开启转速影响较小。因此,当镶装环与跑道材料不同时,由于 2 种不同材料的线膨胀系数不同,镶装式浮环与跑道的间隙随温度的变化而变化,密封的开启转速受温度的影响较大;但当镶装环与跑道材料相同时,由于二者的线膨胀系数相同,镶装式浮环

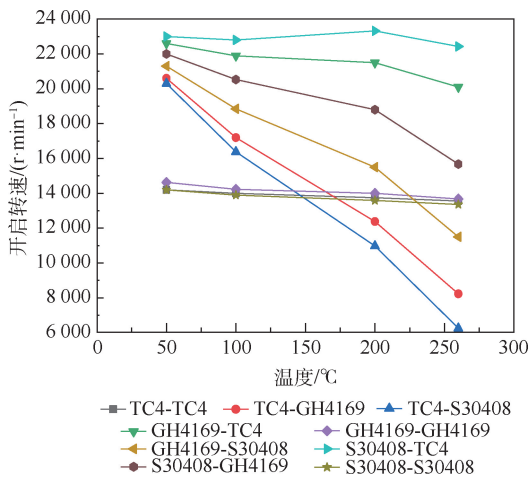


图 11　镶装环-跑道配对材料和开启转速关系

Fig. 11　Relationship between matching material of inlay ring and runway and opening speed

与跑道的间隙不会受到影响,密封处于“恒间隙”状态,此时密封的开启转速只受空气黏度变化的影响,密封的开启转速受温度影响较小。

4.5 发动机启动方式

根据发动机类型和实际工作条件,发动机的启动方式有 3 种:先增速至预定工作转速再进行增压,先增压至预定工作压力再进行增速,发动机启动时增压增速同时进行。试验分为试验台无压力启动、试验台同时增压增速启动、试验台先增压启动 3 种情况,分别对应发动机的 3 种启动方式。由图 12 可知,密封的开启转速随密封腔内压力的升高而升高,这是因为随着腔内压力的提高,密封端面压紧力 F_n 也随之增加,进而使密封的密封闭合力增加,最终使密封的开启难度增大。因此,发动机在较高工况压力时先增压启动不利于密封的开启,且不同的启动方式对密封开启转速的影响变化高达 12 000 r/min,发动机启动方式对密封开启性能影响较大。

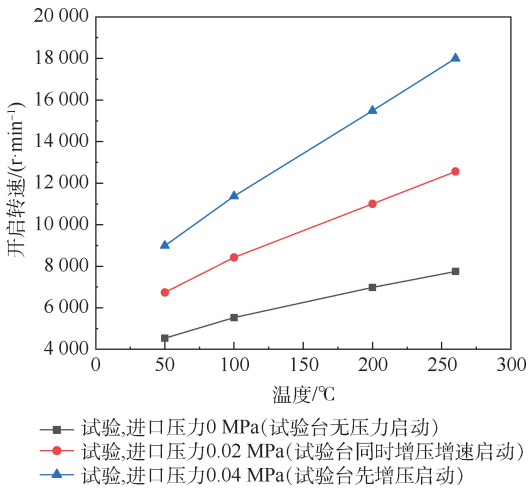


图 12 启动方式和开启转速关系

Fig. 12 Relationship between driving mode and opening speed

5 密封开启运行稳定性

利用浮环密封试验台和浮环位移监测系统对密封开启过程中压力变化和转速变化对运行稳定性的影响进行研究。

固定进口压力为 0.04 MPa,在不同转速下待密封运行稳定,再分别给其 2 000 r/min 的转速正扰动量,监测浮环的径向位移值。图 13 为稳定转速为 1 000,2 000,3 000 r/min 的扰动记录图,图 14为稳定转速为 4 000,6 000,8 000 r/min 的扰动记录图。

气膜浮环密封所处的工作转速越高,其受扰动的波动愈剧烈,故其稳定性相对于低转速工况下的浮环密封稳定性差,在转速为 8 000 r/min 时,扰动后的位移波动值最大达到了约 12 μm。对比图 13和图 14,浮环工况转速越大,其受相同扰动后重新恢复到新稳定状态的时间也会变长。

固定浮环工况压力,改变工况转速分别为 1 000,3 000,5 000 r/min,分别在相应工况转速下稳定后给予整体式气膜浮环 0.01 MPa 的压力扰动,记录并整理浮环扰动位移图像。图 15 和图 16分别为保持工况压力 0.04 MPa 和 0.06 MPa 时,密封在转速变化下的扰动值变化曲线。

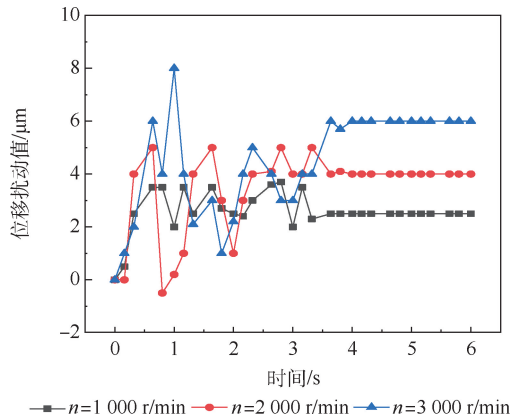


图 13 转速为 1 000 r/min,2 000 r/min,3 000 r/min 下扰动值

Fig. 13 Disturbance values at speed of 1 000 r/min, 2 000 r/min and 3 000 r/min

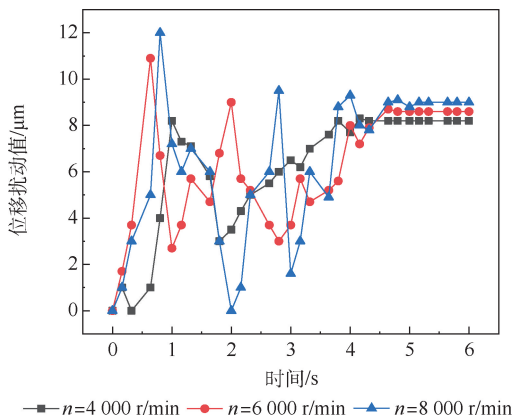


图 14 转速为 4 000 r/min,6 000 r/min,8 000 r/min 下扰动值

Fig. 14 Disturbance values at speed of 4 000 r/min, 6 000 r/min and 8 000 r/min

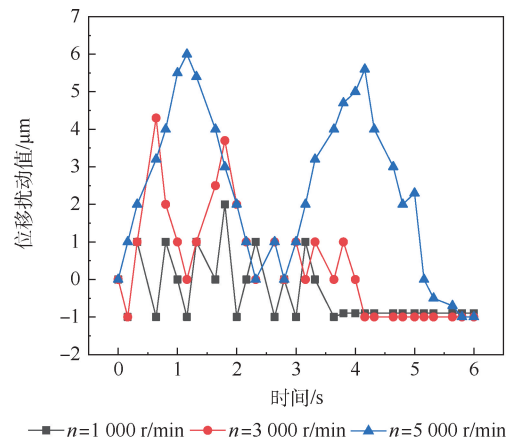


图 15 压力为 0.04 MPa 下扰动值

Fig. 15 Disturbance values under pressure of 0.04 MPa

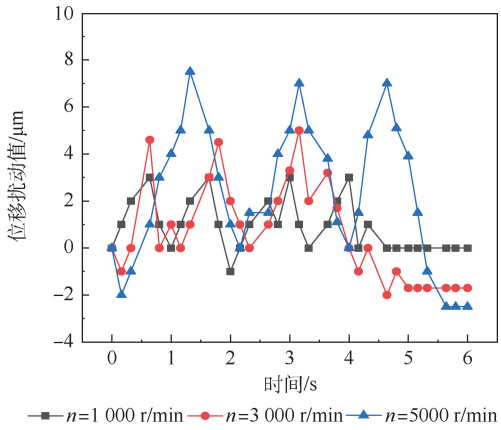


图 16　压力为 0.06 MPa 下扰动值

Fig. 16　Disturbance values under pressure of 0.06 MPa

气膜浮环密封的工况转速和压力越大,在运行过程中受到相同压力扰动后的波动程度就越剧烈。

6　结　论

针对镶装式浮环密封的开启性能,进行了数值模拟研究和试验研究,得到如下结论:

1)不同的参数对密封开启性的影响差异性很大。镶装环材料与跑道材料配对、镶装环材料与石墨环材料配对、发动机启动方式对密封开启性能影响较大;镶装环与石墨环厚度比、宽度比对密封开启性能影响较小。

2)浮环的镶装结构可以有效改善升温时石墨环与跑道的间隙较小问题,镶装环与跑道材料相同时,密封处于“恒间隙”状态,开启转速受温度变化影响较小,在工况温度变化剧烈的情况下使用可以保证密封开启性能的稳定性。

3)浮环密封在高转速高压力工况下运行时受转速、压力变化时密封扰动值较低工况下更大,因此,浮环密封应避免在较高压力和转速下长时间调节工况参数。

参考文献(References)

[1] CHILDS D W, RODRIGUEZ L E, CULLOTTA V, et al. Rotordynamic-coefficients and static (equilibrium loci and leakage) characteristics for short, laminar-flow annular seals[J]. Journal of Tribology, 2006, 128(2): 378-387.

[2] CHILDS D W, GRAVISS M, RODRIGUEZ L E. Influence of groove size on the static and rotordynamic characteristics of short, laminar-flow annular seals[J]. Journal of Tribology, 2007, 129(2): 398-406.

[3] XIA P, ZHANG G, ZHAO J, et al. Investigations on rotordynamic characteristics of a floating ring seal considering structural elasticity[C]// ASME Turbo Expo: Turbomachinery Technical Conference & Exposition. New York: ASME, 2017, 1: 1-10.

[4] SEMANATE J, SAN ANDRES L A. Analysis of multi-land high pressure oil seals[J]. ASLE Transactions, 1993, 36(4): 661-669.

[5] BERGER E. Friction modeling for dynamic system simulation[J]. Applied Mechanics Reviews, 2002, 55(6): 25-32.

[6] 马也,王庆锋,施任杰,等. 航空发动机气膜浮环密封上浮性能研究[J]. 润滑与密封, 2021, 46(1): 38-44.
MA Y, WANG Q F, SHI R J, et al. Research on the floating performance of aeroengine gas film floating ring seal[J]. Lubrication and Seal, 2021, 46(1): 38-44(in Chinese).

[7] 马利军,李双喜,马也,等. 气膜浮环密封特性参数的影响因素分析[J]. 流体机械, 2018, 46(10): 6-10.
MA L J, LI S X, MA Y, et al. Analysis of influencing factors of gas film floating ring seal characteristic parameters[J]. Fluid Machinery, 2018, 46(10): 6-10(in Chinese).

[8] 马利军,李双喜,蔡纪宁,等. 高温镶装式浮环密封的径向间隙研究[J]. 润滑与密封, 2018, 43(6): 90-95.
MA L J, LI S X, CAI J N, et al. Study on radial clearance of high temperature inlaid floating ring seal[J]. Lubrication and Seal, 2018, 43(6): 90-95(in Chinese).

[9] 夏鹏,刘占生. 浮环密封结构弹性对间隙泄漏量和动力学系数的影响[J]. 推进技术, 2017, 38(12): 2815-2821.
XIA P, LIU Z S. Effect of structural elasticity of floating ring seal on clearance leakage and dynamic coefficient[J]. Propulsion Technology, 2017, 38(12): 2815-2821(in Chinese).

[10] 刘占生,夏鹏,张广辉,等. 浮环密封运动机理及对轴系稳定性的影响[J]. 振动与冲击, 2016, 35(9): 110-116.
LIU Z S, XIA P, ZHANG G H, et al. Motion mechanism of floating ring seal and its influence on shafting stability[J]. Vibration and Impact, 2016, 35(9): 110-116(in Chinese).

[11] 马纲,席平,沈心敏,等. 柔性支承浮环柱面气膜密封准动态特性分析[J]. 航空动力学报, 2010, 25(5): 1190-1196.
MA G, XI P, SHEN X M, et al. Quasi dynamic characteristics analysis of cylindrical gas film seal with flexible support floating ring[J]. Journal of Aeronautical Dynamics, 2010, 25(5): 1190-1196(in Chinese).

[12] ARGHIR M, NGUYEN M H, TONON D. Analytic modeling of floating ring annular seals[J]. Journal of Engineering for Gas Turbines and Power, 2012, 134(5): 153-161.

[13] SHAPIRO W, LEE C, JONES H. Analysis and design of a gas-lubricated, sectored, floating ring seal[J]. Transactions of the ASME, 1988, 110(7): 525-532.

[14] 王飞,刘向锋,刘莹. 浅槽环瓣型浮动环密封的性能分析[J]. 润滑与密封, 2005(6): 74-76.
WANG F, LIU X F, LIU Y. Performance analysis of shallow groove floating ring seal[J]. Lubrication and Seal, 2005(6): 74-76(in Chinese).

[15] 许庆余,张义忠,朱因远. 密封浮环非线性振动及环心轨迹的研究[J]. 西安交通大学学报, 1982(1): 70-79.
XU Q Y, ZHANG Y Z, ZHU Y Y. Study on nonlinear vibration and center locus of floating ring[J]. Journal of Xi'an Jiaotong University, 1982(1): 70-79(in Chinese).

Opening characteristics of inlaid floating ring seal with high-speed gas film

ZHENG Rao, CHEN Xiaozhu, LI Shuangxi*, ZHAO Xiang, SHI Renjie, SONG Zifeng

(College of Mechanical and Electrical Engineering, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: This study investigates the inlaid floating ring seal with high-speed gas films in aeroengines, focusing on the opening performance of the seal with different structural parameters, starting modes, and material combinations. The solid domain model and the gas film fluid domain numerical analysis model of the inlay ring-graphite ring-runway are established, and the working gas film thickness and the pressure distribution of the gas film flow field are obtained. By calculating the seal force, the seal opening performance parameters, such as buoyancy, closing force and opening speed, are obtained. This study analyzes the effects of many factors on seal opening speed, such as the inlay-graphite ring thickness ratio and width ratio, and material matchings between the inlay and graphite rings and between the inlay ring and runway. A floating ring seal test bench and a floating ring displacement monitoring system are built, and the numerical simulation results are verified by experiments. Results show that the inlaid structure of the floating ring can effectively improve the sealing failure caused by the decrease of the gap between the graphite ring and runway when temperature rises. The material of the insert ring is a sensitive parameter, which affects the opening performance of seals. This performance decreases rapidly with the increase of the linear expansion coefficient of the material. The material matching between the inlay ring and runway is an important factor for the opening performance. When the inlay ring and runway have the same material, the graphite ring and runway are in a state with a "constant gap", and the opening performance of the seal is more stable at complex temperatures. Results also show that start-up mode has great influence on the opening performance of seals, and that this performance is best when the engine is started by increasing the speed to the working speed and then pressurized. Therefore, floating ring seals should avoid adjusting operating parameters for a long time with higher pressure and speed. The results provide a basis for structural design, material selection and system design of the inlaid floating ring seal for aircraft engines.

Keywords: bearing cavity seal; floating ring seal; opening characteristics; constant gap; operation stability

Received: 2021-02-19; **Accepted**: 2021-07-03; **Published online**: 2021-08-25 20:13

URL: kns.cnki.net/kcms/detail/11.2625.V.20210825.1244.003.html

Foundation item: National Key R & D Program of China (2018YFB2000800)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0091

基于径向基函数代理模型的M形杆刚度优化

杨慧[1,*], 范硕硕[2], 王岩[1], 刘荣强[3], 肖洪[3]

(1. 燕山大学 机械工程学院, 秦皇岛 066044; 2. 安徽大学 电气工程与自动化学院, 合肥 230601;
3. 哈尔滨工业大学 机器人技术与系统国家重点实验室, 哈尔滨 150001)

摘 要: 在航天任务的执行中,超弹性杆主要用于大型空间可展天线和太阳帆等展开和支撑。为了提高超弹性杆在展开状态下的刚度,提出了一种新型M形超弹性杆,并对M形超弹性杆的刚度进行了研究。采用ABAQUS建立M形超弹性杆的弯曲、压缩和扭转的有限元模型,利用显示动力学法对屈曲过程进行非线性数值模拟。采用全因子法进行实验设计,利用径向基函数(RBF)建立M形超弹性杆屈曲过程性能参数的代理模型。以弯曲刚度、扭转刚度和压缩刚度为优化目标,以质量为约束,选取黏结段长度和内侧带簧片圆心角为自变量建立优化模型。采用粒子群(PSO)算法进行M形超弹性杆参数优化,得到最优刚度下,黏结段长度为7.894 5 mm,圆心角为26°,并且得到刚度随其变化规律。

关 键 词: 复合材料; 可展机构; M形杆; 径向基函数; 优化
中图分类号: V214.2
文献标志码: A **文章编号**: 1001-5965(2022)11-2121-09

可展开薄壁超弹性杆件能够利用自身弯曲存储的弹性势能实现展开,在薄膜天线、太阳帆和阻力帆中广泛使用,常见的几种不同截面的薄壁杆件有豆荚杆[1-3]、TRAC[4-5]杆和圆形储能伸缩杆(STEM)[6-7]。

Mallikarachchi等研究了带弹簧和复合材料薄壁杆的缠绕[8-9]及2层织层板在小应变下的微观力学行为[10]。楚中毅等[11-12]提出了一种用于航天探测器的主动-被动复合驱动可展开豆荚杆,在模态和缠绕分析的基础上对豆荚杆参数进行了优化设计。文献[13-15]分析了可展开复合材料薄壁杆的压扁和缠绕过程的实验和数值。Bessa等[16-17]采用数据驱动法对TRAC杆进行了优化设计,但忽略了2个带弹簧之间的黏结段影响。Roybal等[18]发现在相同的压扁截面高度时,TRAC杆的截面惯性量是豆荚杆的10倍,是STEM的34倍。杨慧等[19]提出了一种N形超弹性杆,并对其后屈曲性能进行了优化。目前,对超弹性杆压扁和缠绕过程的力学性能进行的研究较多,对完全展开状态下的承载能力却很少研究。

本文以M形超弹性杆(简称M形杆)为研究对象,对展开状态下的M形杆进行了刚度分析和几何参数优化,从而提高M形杆展开状态的刚度。首先,利用有限元软件ABAQUS建立M形杆绕x、y和z轴弯曲、扭转,以及沿z轴压缩的有限

收稿日期: 2021-02-24; **录用日期**: 2021-06-06; **网络出版时间**: 2021-06-21 15:02
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210621.1255.002.html
基金项目: 安徽省自然科学基金(2108085QE221); 国家自然科学基金(51975001,51835002,52105013); 机器人技术与系统国家重点实验室开放基金(SKLRS-2020-KF-16)
* **通信作者**. E-mail: huiyang_0431@163.com

引用格式: *杨慧, 范硕硕, 王岩, 等. 基于径向基函数代理模型的M形杆刚度优化[J]. 北京航空航天大学学报, 2022, 48(11): 2121-2129. YANG H, FAN S S, WANG Y, et al. Stiffness optimization of M-shaped boom based on radial basis function surrogate model[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2121-2129(in Chinese).*

元模型;然后,提取绕 x 轴弯曲刚度 $EI_x(h,\theta_2)$、绕 y 轴弯曲刚度 $EI_y(h,\theta_2)$、绕 z 轴扭转刚度 $GJ_z(h,\theta_2)$、沿 z 轴压缩刚度 $EA_z(h,\theta_2)$,采用径向基函数(radial basis function,RBF)法建立代理模型;最后,以上述 4 种刚度为目标,以质量为约束,以黏结段长度 h 和内侧带簧片圆心角 θ_2 为变量,采用粒子群(particle swarm optimization,PSO)算法进行优化。

1 M 形杆几何模型

M 形杆截面几何尺寸如图 1 所示。M 形杆横截面被压扁之后可以绕轮毂缠绕,缠绕弯曲过程中存储应变能,释放约束之后,依靠存储的弹性势能,能够展开到初始状态。M 形杆是由 4 个带簧片组成,2 个外带簧片只有 1 个黏结段,2 个内带簧片有中心对称的 2 个黏结段,黏结段的长度为 h。内侧带簧片由 2 个圆角外切弧组成,M 形杆 4 个带簧片有 2 个独立的参数,即 h 和 θ_2。由压扁后内外侧宽度相同可以推导出如下关系:

$$\theta_1 = \frac{2r_2\theta_2 + h}{r_1} \tag{1}$$

由图 1 可以得到 M 形杆质量 M_0 为

$$M_0(h,\theta_2) = 8\rho L_m t_m (r_2\theta_2 + h) + 3t_G h L_m \rho_G \tag{2}$$

式中:杆件长度 $L_m = 0.4$ m;厚度 $t_m = 0.15$ mm 不变;ρ 为复合材料 T800 的密度;t_G 为黏结胶厚度;ρ_G 为黏结胶密度。

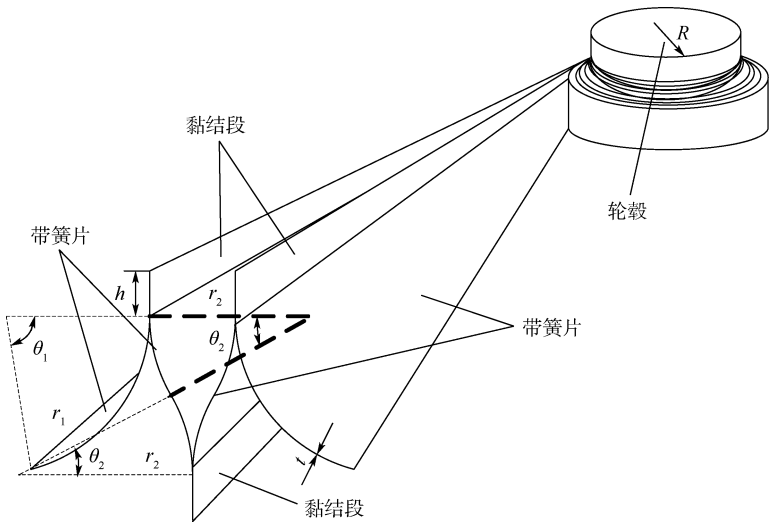


图 1 M 形杆缠绕几何示意图

Fig. 1 Geometric schematic diagram of M boom coiling

2 有限元建模分析

M 形杆的仿真模型以实际工况为依据,采用 4 层复合材料的铺设方式,复合材料叠加顺序为[45°/-45°/-45°/45°],每层的厚度为 $t = 3.75\times10^{-5}$ m,复合材料属性如表 1 所示,铺层方式如图 2 所示。

表 1 T800 材料属性

Table 1 Material properties of T800

材料属性	T800	黏结胶
纵向弹性模量 E_1/MPa	150 000	
横向弹性模量 $E_2 = E_3$/MPa	7 000	60 000
剪切模量 $G_{12} = G_{13}$/MPa	7 000	
面内剪切模量 G_{23}/MPa	4 500	
泊松比 υ	0.3	0.3
密度 ρ/(kg·m^{-3})	2 500	1 600

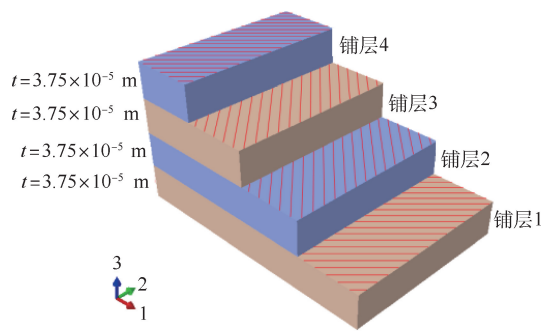


图 2 M 形杆各层材料铺设布置图

Fig. 2 Material laying layout of each layer of M boom

创建一个 Cohesive 部件模拟 M 形杆黏结段,M 形杆和黏结胶分别采用壳单元 S4R 和表面单元 C3D8R 模拟,由于 M 形杆的横截面关于 y 轴对称,只需对相互对称的一侧参数进行研究。为了得到黏结段长度 h 和内侧带簧片圆心角 θ_2 对 M

形杆临界载荷的影响，在 M 形杆两端的横截面处设置 2 个参考点 RP_1 和 RP_2，并以 2 个参考点为控制点与杆件两端建立 MPC 约束，如图 3 所示。固定参考点 RP_1，且限制参考点 RP_2 的 5 个自由度，给 RP_2 施加一定的旋转或平移自由度，UX、UY、UZ 分别为沿 x、y、z 轴水平位移，RX、RY、RZ 分别为绕 x、y、z 轴旋转位移，具体施加载荷情况如表 2 所示。

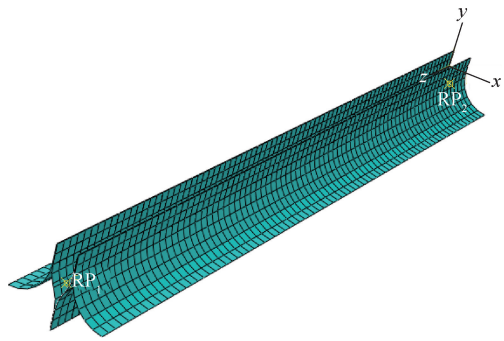


图 3　M 形杆有限元模型

Fig. 3　Finite element model of M boom

结合表 2 和图 3，M 形杆绕 x 轴弯曲刚度 $EI_x(h,\theta_2)$、绕 y 轴弯曲刚度 $EI_y(h,\theta_2)$、绕 z 轴扭转刚度 $GJ_z(h,\theta_2)$、沿 z 轴压缩刚度 $EA_z(h,\theta_2)$ 可表示为

$$EI_x(h,\theta_2) = \left(\frac{dM_x}{d\theta_x}\right)L_m \tag{3}$$

$$EI_y(h,\theta_2) = \left(\frac{dM_y}{d\theta_y}\right)L_m \tag{4}$$

$$GJ_z(h,\theta_2) = \left(\frac{dM_z}{d\theta_z}\right)L_m \tag{5}$$

$$EA_z(h,\theta_2) = \left(\frac{dF_z}{dS_z}\right)L_m \tag{6}$$

式中：M_x、M_y、M_z 分别为绕 x、y、z 轴的力矩；θ_x、θ_y、θ_z 分别为绕 x、y、z 轴旋转的角度；F_z 为轴向力；S_z 为压缩位移。

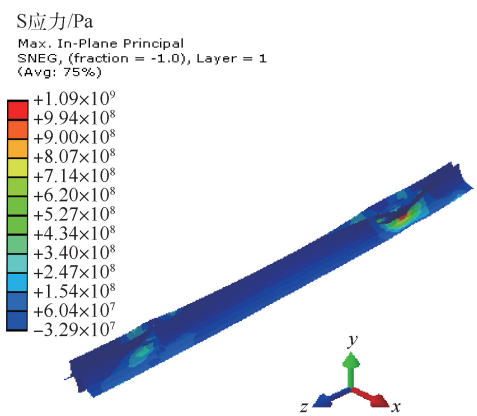
表 2　参考点 RP_2 在不同情况下的边界条件

Table 2　Reference point RP_2 boundary conditions under different conditions

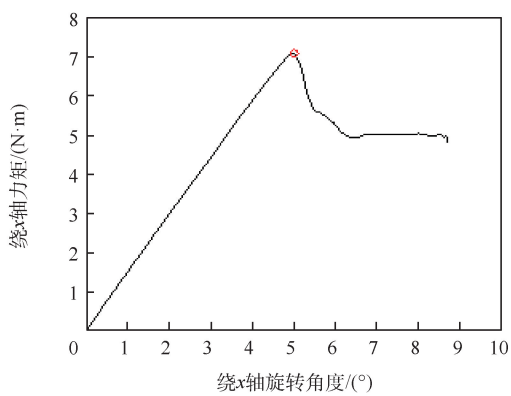
约束	绕 x 轴弯曲	绕 y 轴弯曲	绕 z 轴扭转	沿 z 轴正向压缩
UX	0	释放约束	释放约束	释放约束
UY	释放约束	0	释放约束	释放约束
UZ	释放约束	释放约束	0	施加位移载荷
RX	施加旋转载荷	0	0	0
RY	0	施加旋转载荷	0	0
RZ	0	0	施加旋转载荷	0

当 $h=7$ mm，$\theta_2=24°$，$\theta_1=68.06°$ 时，M 形杆绕 x、y 轴弯曲的应力云图和力矩曲线如图 4 和图 5所示，绕 z 轴扭转的应力云图和力矩曲线如图 6 所示，沿 z 轴压缩的应力云图和反作用力曲线如图 7 所示，图中的圆圈标记点为 M 形杆的临界载荷点。

为了得到最优刚度下的 h 和 θ_2 组合，需要在样本空间中进行大量的有限元数值模拟计算。共在样本空间中生成 3 000 个样本点，每个样本点在 Intel(R) Core(TM) i5-9400F CPU@2.90 GHz 6 核配置计算机上仿真计算得到 4 种刚度需要约 40 min，为了节约时间和资源成本，需要构建代理模型，代理模型完成后利用 MATLAB 可以在 1 s



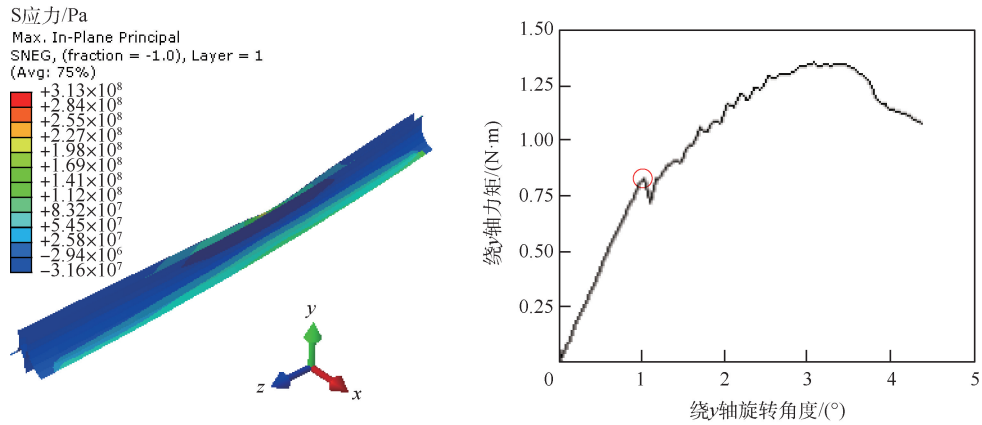
(a) 绕x轴弯曲10°应力云图



(b) 绕x轴弯曲过程力矩曲线

图 4　M 形杆绕 x 轴弯曲的应力云图和力矩曲线

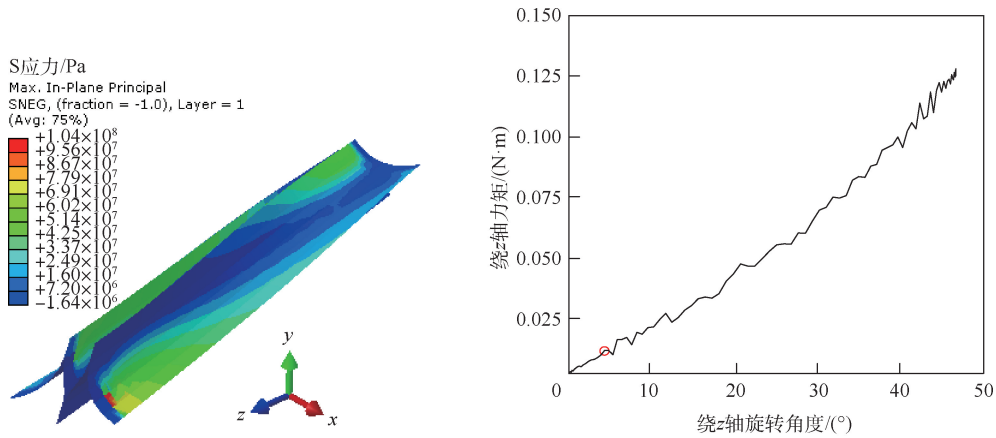
Fig. 4　Cloud chart of stress and bending moment curve of M boom around x-axis



(a) 绕y轴弯曲5°应力云图　　(b) 绕y轴弯曲过程力矩曲线

图5　M形杆绕y轴弯曲的应力云图和力矩曲线

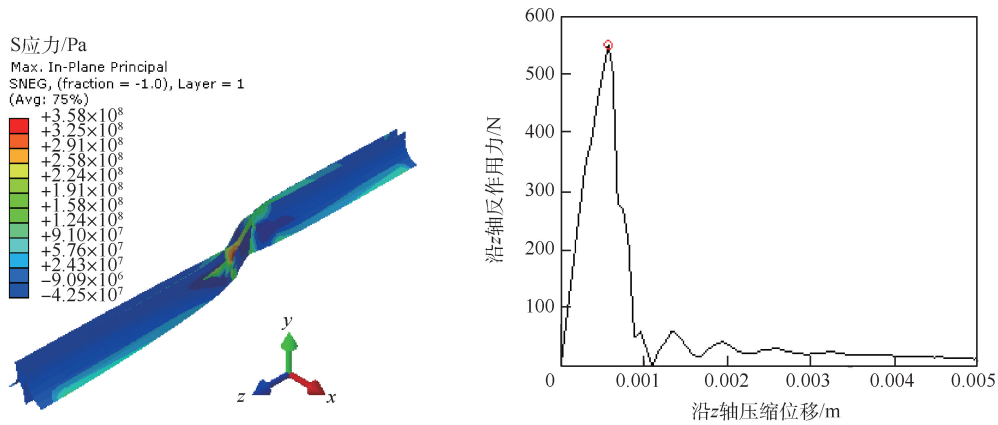
Fig. 5　Cloud chart of stress and bending moment curve of M boom around y-axis



(a) 绕z轴扭转50 °应力云图　　(b) 绕z轴扭转过程力矩曲线

图6　M形杆绕z轴扭转的应力云图和力矩曲线

Fig. 6　Cloud chart of stress and torsion moment curve of M boom around z-axis



(a) 沿z轴压缩5 mm后应力云图　　(b) 沿z轴压缩过程反作用力曲线

图7　M形杆沿z轴压缩的应力云图和反作用力曲线

Fig. 7　Stress cloud diagram and reaction force curves of M boom compressed along z-axis

左右计算出4种刚度输出数值。代理模型简单来说是通过已有样本点构建输入变量与输出目标之间的映射关系,从而利用映射关系来近似预测样本空间中其他未知样本点的输出值。为了构建关于4种刚度EI_x、EI_y、GJ_z、EA_z的代理模型,采用全因子法进行实验设计。h取值范围为5~10 mm,以1 mm等间隔,共有5,6,7,8,9,10六水平,θ_2取值范围为22°~26°,以1°为等间隔,共有22,23,24,25,26五水平,共在样本空间中生成30个样本点,其有限元分析结果如表3所示,用30个已知样本点建立代理模型预测整个样本空间中3 000个样本点输出值。

表 3　样本点

Table 3　Sample points

序号	h/mm	θ_2/(°)	$EI_x(h,\theta_2)$/(N·m²·rad⁻¹)	$EI_y(h,\theta_2)$/(N·m²·rad⁻¹)	$GJ_z(h,\theta_2)$/(N·m²·rad⁻¹)	$EA_z(h,\theta_2)$/(N·mm⁻¹)
1	5	22	16.525	8.800	0.022	318.836
2	5	23	17.764	10.718	0.026	322.861
3	5	24	19.018	12.579	0.031	314.631
4	5	25	20.672	14.756	0.037	322.566
5	5	26	21.779	17.296	0.044	341.485
6	6	22	20.299	10.467	0.028	342.641
7	6	23	21.683	12.250	0.031	352.375
8	6	24	23.548	13.527	0.036	359.440
9	6	25	24.680	16.322	0.045	354.424
10	6	26	26.529	19.273	0.051	349.309
11	7	22	24.826	11.736	0.030	362.863
12	7	23	26.121	14.097	0.038	336.068
13	7	24	28.263	15.904	0.043	386.556
14	7	25	30.448	18.814	0.049	387.839
15	7	26	31.960	21.919	0.059	417.498
16	8	22	31.014	13.979	0.035	390.809
17	8	23	24.585	16.280	0.039	391.156
18	8	24	25.897	19.031	0.050	389.042
19	8	25	35.451	22.181	0.059	412.499
20	8	26	36.913	25.155	0.068	398.496
21	9	22	34.444	16.144	0.042	418.356
22	9	23	28.561	18.814	0.050	396.581
23	9	24	30.862	21.860	0.057	421.409
24	9	25	39.797	24.364	0.072	404.325
25	9	26	43.121	24.267	0.079	430.254
26	10	22	42.957	18.604	0.051	417.984
27	10	23	42.714	21.545	0.057	426.818
28	10	24	46.798	24.402	0.064	440.339
29	10	25	47.374	28.437	0.081	454.123
30	10	26	51.486	32.035	0.094	470.183

3　径向基函数代理模型

3.1　建立代理模型

径向基函数代理模型基本的建模思想是:以设计空间中初步确定的实验样本点作为中心,选用不同的径向基函数,利用径向基函数的线性叠加拟合函数 $f(x)$,$x\in\mathbf{R}^n$,通过拟合函数估算未知点处的函数值。径向基函数 $\varphi(g)$ 作为 $f(x)$ 的基函数,待测样本点与中心点的欧氏距离作为自变量,径向基函数的取值仅由离中心点 c 的距离唯一确定,即 $\varphi(x,c)=\varphi(\|x-c\|)$。

径向基函数代理模型表达式如下:

$$f(x)\approx\sum_{i=1}^{m}\boldsymbol{\omega}\,\boldsymbol{\varphi}(\|x-x_i\|) \tag{7}$$

式中:$\boldsymbol{\omega}=\begin{bmatrix}\omega_1\\ \vdots\\ \omega_m\end{bmatrix}$;$\boldsymbol{\varphi}=\begin{bmatrix}\varphi(g_1,c)\\ \vdots\\ \varphi(g_m,c)\end{bmatrix}$;$\boldsymbol{\omega}$ 为径向基函数 $\varphi(\cdot)$ 的加权系数;$g_i=\|x-x_i\|(i=1,2,\cdots,m)$ 为未知点 x 与样本点 x_i 之间的欧氏距离。选取逆二次函数 $\phi(g)=(g^2+c^2)^{1/2}$ 作为径向基函数的核函数,形参数 c 的选取通过交叉验证法确定,由文献[20]可知,对于 Inverse quadratic 径向基核函数,c 取值范围划分为 0.1 ~ 5,所选取的 c 值为0.1、0.3、1、3、4、4.5、5。使用上述形参数 c 值,利用交叉验证法进行径向基函数代理模型的

建模,通过对比不同形参数 c 下的代理模型误差,最终选取代理模型误差最小的一组形参数 c 作为最终的形参数,建立的关于 EI_x、EI_y、GJ_z、EA_z 的4个代理模型对应核函数中的形参数 c 的取值如表4所示。使用式(7)作为近似代理模型时,应满足如下插值条件:

$$f(x_j) = Y^j \quad j = 1,2,\cdots,m \tag{8}$$

将式(8)代入式(7),可以得到

$$\boldsymbol{T}\cdot\boldsymbol{\omega} = \boldsymbol{Y} \tag{9}$$

式中:$\boldsymbol{T} = \begin{bmatrix} \varphi(g_{11}) & \cdots & \varphi(g_{1m}) \\ \vdots & & \vdots \\ \varphi(g_{m1}) & \cdots & \varphi(g_{mm}) \end{bmatrix}$;$\boldsymbol{Y}\begin{bmatrix} Y^1 \\ \vdots \\ Y^m \end{bmatrix}$,$\boldsymbol{Y}$ 为已知样本点输出响应值。$\boldsymbol{T}$ 为非奇异矩阵,且在样本点不重合条件下,$\boldsymbol{\omega}$ 存在唯一解,即

$$\boldsymbol{\omega} = \boldsymbol{T}^{-1}\cdot\boldsymbol{Y} \tag{10}$$

由式(10)可求出加权系数 $\boldsymbol{\omega}$,代入式(7)可得到代理模型表达式,对未知点的输出值进行估算。

表4 径向基函数代理模型核函数中常数 c 的取值

Table 4 Values of constant c in kernel function of RBFs surrogate model

目标量	c 值
EI_x	1
EI_y	0.3
GJ_z	0.3
EA_z	0.3

3.2 模型误差分析

用于建立代理模型的样本点处,其实际响应结果与代理模型的近似值相等,无法通过这些样本点衡量代理模型的精度,因此评估代理模型精度是否满足要求需要选取额外的样本点。可以定义相对误差的表达式,即

$$E = \frac{\tilde{f}(x_i) - f(x_i)}{f(x_i)} \quad i = 1,2,\cdots,m \tag{11}$$

式中:$f(x_i)$ 为第 i 个样本点处的有限元结果;$\tilde{f}(x_i)$ 为第 i 个样本点处的径向基函数代理模型近似结果;m 为样本点个数。由检测样本点处误差大小来评估代理模型的精度和拟合效果。建立代理模型共选取样本点30个,为保证代理模型拟合精度,避免误差测试样本点的随机性导致偏差,采用交叉验证法建立代理模型,30个样本点共分为6组,每组5个样本点,每次选择其中5组建立代理模型,余下1组做为误差检测样本点,选择检验误差最小的1组模型作为最终代理模型。最终得到4个关于 $EI_x(h,\theta_2)$、$EI_y(h,\theta_2)$、$GJ_z(h,\theta_2)$、$EA_z(h,\theta_2)$ 的代理模型,测试误差结果如表5所示。模型最大相对误差的绝对值分别为6.155%、1.666%、1.518%和5.538%,代理模型精度均满足要求。

3.3 代理模型响应面

关于 $EI_x(h,\theta_2)$、$EI_y(h,\theta_2)$、$GJ_z(h,\theta_2)$、$EA_z(h,\theta_2)$ 的响应面如图8~图11所示。由响应面图形可知,黏结段长度 h 的变化对 $EI_x(h,\theta_2)$ 和 $EA_z(h,\theta_2)$ 有显著影响,h 和 θ_2 的变化对 $GJ_z(h,\theta_2)$ 和 $EI_y(h,\theta_2)$ 产生变化影响程度相当。

表5 误差测试样本点

Table 5 Sample points of relative errors

序号	h/mm	θ_2/(°)	有限元结果		代理模型结果		相对误差 E/%	
			$EI_x(h,\theta_2)$/(N·m²·rad⁻¹)	$EI_y(h,\theta_2)$/(N·m²·rad⁻¹)	$EI_x(h,\theta_2)$/(N·m²·rad⁻¹)	$EI_y(h,\theta_2)$/(N·m²·rad⁻¹)	$EI_x(h,\theta_2)$	$EI_y(h,\theta_2)$
1	6	26	26.529	19.273	26.120	19.259	-1.542	-0.073
2	7	24	28.263	15.904	27.329	16.169	-3.305	1.666
3	7	26	31.960	21.919	32.612	22.091	2.040	0.785
4	8	22	31.014	13.979	29.105	14.097	-6.155	0.844
5	8	26	36.913	25.155	38.292	25.342	3.736	0.743

序号	h/mm	θ_2/(°)	有限元结果		代理模型结果		相对误差 E/%	
			$GJ_z(h,\theta_2)$/(N·m²·rad⁻¹)	$EA_z(h,\theta_2)$/(N·mm⁻¹)	$GJ_z(h,\theta_2)$/(N·m²·rad⁻¹)	$EA_z(h,\theta_2)$/(N·mm⁻¹)	$GJ_z(h,\theta_2)$	$EA_z(h,\theta_2)$
1	6	26	0.051 4	349.309	0.051 0	367.851	-0.778	5.308
2	7	24	0.043 4	386.556	0.042 9	366.866	-1.152	-5.094
3	7	26	0.059 3	417.498	0.058 4	400.586	-1.518	-4.051
4	8	22	0.034 8	390.809	0.035 2	395.578	1.149	1.220
5	8	26	0.068 2	398.496	0.067 7	420.565	-0.733	5.538

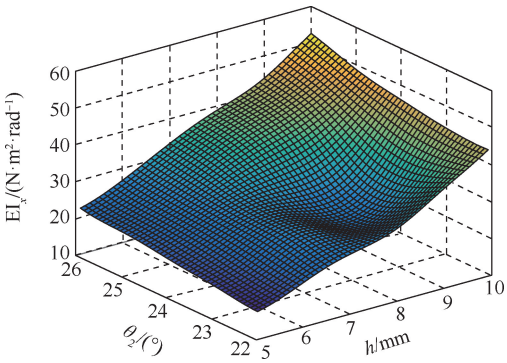


图 8　EI$_x$(h,θ_2)响应面

Fig. 8　Response surface of EI$_x$(h,θ_2)

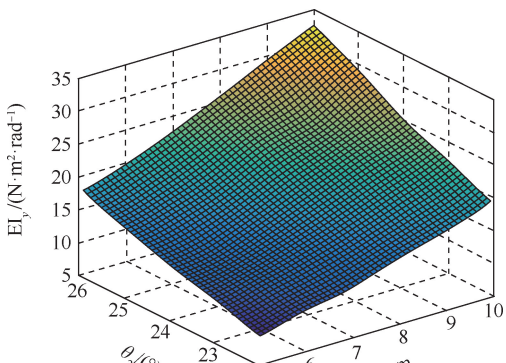


图 9　EI$_y$(h,θ_2)响应面

Fig. 9　Response surface of EI$_y$(h,θ_2)

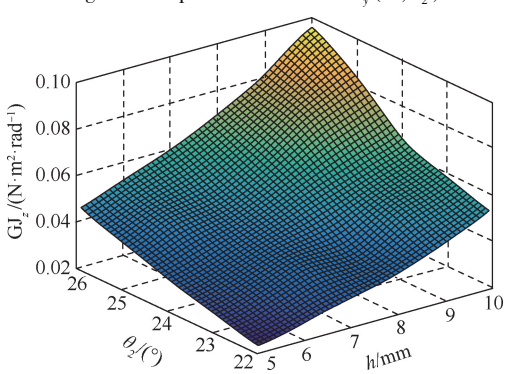


图 10　GJ$_z$(h,θ_2)响应面

Fig. 10　Response surface of GJ$_z$(h,θ_2)

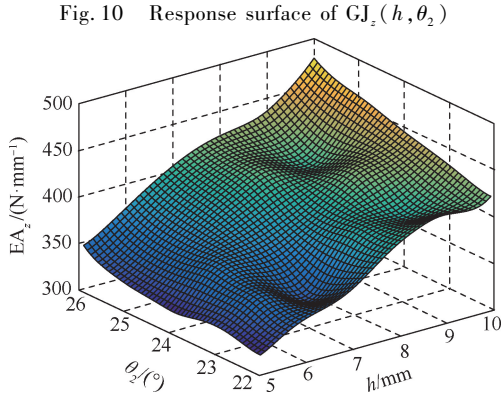


图 11　EA$_z$(h,θ_2)响应面

Fig. 11　Response surface of EA$_z$(h,θ_2)

4　优化设计

以绕 x 轴弯曲刚度 EI$_x$(h,θ_2)、绕 y 轴弯曲刚度 EI$_y$(h,θ_2)、绕 z 轴扭转刚度 GJ$_z$(h,θ_2)和沿 z 轴压缩刚度 EA$_z$(h,θ_2)为 4 个目标量,目标量取最大值,以黏结段长度 h 和内侧带簧片圆心角 θ_2 为设计变量,以质量为约束,优化设计如式(12)所示。用粒子群算法进行多目标优化,最大迭代次数设为 100,粒子群数量设为 30,惯性权重、群体学习因子及自我学习因子均设为 0.9。如式(13)所示,目标函数 O 值的计算为所有优化目标分量 X_i 和对应第 i 个目标分量的权重因子 W_i、比例因子 S_i 加权之和。如果目标分量的比例因子和权重相等,目标分量数量级较大的对优化结果影响较大,比例因子和权重的设置是为了调整在优化过程中各子目标的重要程度。为方便计算,比例因子统一设置为 1。由于 M 形杆为开口型超弹性杆,扭转刚度 GJ$_z$ 较小,对杆件整体刚度影响较大,在优化中赋予较大权重,设为 1 000;EI$_x$、EI$_y$、EA$_z$、M_0 根据自身数量级的大小,分别设置为 1.2、2、0.1、1.5。4 个目标量及约束量 M_0 的比例因子和权重如表 6 所示。在多目标优化问题中,多个目标在约束条件下相互制约,找到的一般不是所有子目标的最优解,而是 Pareto 解,可行结果即为 Pareto 最优解,一般不止一个,多目标优化问题的最终解是从所有 Pareto 最优解中挑一个最优折中解,由于 M 形杆扭转刚度对杆件整体刚度影响较大,选择扭转刚度较大的一个解做为最终解。优化后的 Pareto 结果如表 7 所示。最优设计点为第 2 个,h = 7.894 5 mm,θ_2 = 26°,θ_1 = 74.616°,结果如表 8 所示。

$$\text{Opt.}\begin{cases}\text{EI}_x(h,\theta_2)\mid_{\max}\\ \text{EI}_y(h,\theta_2)\mid_{\max}\\ \text{GJ}_z(h,\theta_2)\mid_{\max}\\ \text{EA}_z(h,\theta_2)\mid_{\max}\\ \text{s.t. } M_0 \leqslant 25\ \text{g}\\ 5\ \text{mm} \leqslant h \leqslant 10\ \text{mm}\\ 22° \leqslant \theta_2 \leqslant 26°\end{cases} \tag{12}$$

$$O = \sum \frac{W_i X_i}{S_i} \tag{13}$$

表 6　目标量和约束量的比例因子和权重

Table 6　Scale factors and weights of objective and constraint quantities

目标量	S_i	W_i
EI$_x$	1	1.2
EI$_y$	1	2
GJ$_z$	1	1 000
EA$_z$	1	0.1
M_0	1	1.5

表7 Pareto设计点

Table 7 Pareto points of design

序号	h/mm	θ_2/(°)	M_0/g	$EI_x(h,\theta_2)$/(N·m²·rad⁻¹)	$EI_y(h,\theta_2)$/(N·m²·rad⁻¹)	$GJ_z(h,\theta_2)$/(N·m²·rad⁻¹)	$EA_z(h,\theta_2)$/(N·mm⁻¹)
1	7.970 7	25.727 7	24.932 7	37.814 4	24.341 7	0.065 0	415.960 8
2	7.894 5	26	24.911 4	37.794 9	24.994 1	0.066 6	419.362 2

表8 最优设计点

Table 8 Optimal point of design

结果	对应刚度	数值
有限元	$EI_x(h,\theta_2)$/(N·m²·rad⁻¹)	36.286 2
	$EI_y(h,\theta_2)$/(N·m²·rad⁻¹)	24.828 4
	$GJ_z(h,\theta_2)$/(N·m²·rad⁻¹)	0.065 1
	$EA_z(h,\theta_2)$/(N·mm⁻¹)	430.808 6
代理模型	$EI_x(h,\theta_2)$/(N·m²·rad⁻¹)	37.794 9
	$EI_y(h,\theta_2)$/(N·m²·rad⁻¹)	24.994 1
	$GJ_z(h,\theta_2)$/(N·m²·rad⁻¹)	0.066 6
	$EA_z(h,\theta_2)$/(N·mm⁻¹)	419.362 2
相对误差 E/%	$EI_x(h,\theta_2)$	4.16
	$EI_y(h,\theta_2)$	0.67
	$GJ_z(h,\theta_2)$	2.30
	$EA_z(h,\theta_2)$	-2.66

5 结 论

1）为了节省仿真分析时间，建立了径向基函数代理模型，得到4种刚度相关的4个代理模型，最大相对误差的绝对值不大于6.155%，精度均满足要求。

2）采用粒子群算法进行参数优化，得到黏结段长度 h = 7.894 5 mm，内侧带簧片圆心角 θ_2 = 26°时，有最优的弯曲刚度、扭转刚度及压缩刚度。最优点的有限元仿真结果与代理模型预测结果之间最大相对误差仅为4.16%。

3）同内侧带簧片圆心角 θ_2 相比，黏结段长度 h 的变化对绕 x 轴弯曲刚度及沿 z 轴压缩刚度的变化影响显著；而 h 和 θ_2 的变化对绕 y 轴弯曲刚度及绕 z 轴扭转刚度的变化影响程度较为接近。

参考文献（References）

[1] SICKINGER C, HERBECK L. Deployment strategies, analyses and tests for the CFRP booms of a solar sail[C]// European Conference on Spacecraft Structures, Materials and Mechanical Testing, 2002.

[2] HU Y, CHEN W J, GAO J F, et al. A study of flattening process of deployable composite thin-walled lenticular tubes under compress and tension[J]. Composite Structure, 2017, 168: 164-177.

[3] BAI J B, CHEN D, XIONG J J, et al. Folding analysis for thin-walled deployable composite[J]. Acta Astronautica, 2019, 159: 622-636.

[4] BANIK J, MUTPHEY T. Performance validation of the triangular rollable and collapsible mast[C]// 24th Annual AIAA/USU Conferenceon Small Satellites, 2010: 1-8.

[5] YANG H, LIU R Q, WANG Y, et al. Experiment and multiobjective optimization design of tape-spring hinges[J]. Structural and Multidisciplinary Optimization, 2015, 51(6): 1373-1384.

[6] STABILE A, LAURENZI S. Coiling dynamic analysis of thin-walled composite deployable boom[J]. Composite Structures, 2014, 113: 429-436.

[7] HOSKIN A, VIQUERAT A, AGLIETTI G S. Tip force during blossoming of coiled deployable booms[J]. International Journal of Solids and Structures, 2017, 118-119: 58-69.

[8] MALLIKARACHCHI H M Y C, PELLEGRINO S. Deployment dynamics of ultrathin composite booms with tape-spring hinges[J]. Journal of Spacecraft and Rockets, 2014, 51(2): 604-613.

[9] MALLIKARACHCHI H M Y C, PELLEGRINO S. Design of ultrathin composite self-deployable booms[J]. Journal of Spacecraft and Rockets, 2014, 51(6): 1811-1821.

[10] MALLIKARACHCHI H M Y C. Predicting mechanical properties of thin woven carbon fiber reinforced laminates[J]. Thin-Walled Structures, 2019, 135: 297-305.

[11] CHU Z Y, LEI Y A. Design theory and dynamic analysis of a deployable boom[J]. Mechanism and Machine Theory, 2014, 71: 126-141.

[12] CHU Z Y, LEI Y A, LI D. Dynamics and robust adaptive control of a deployable boom for a space probe[J]. Acta Astronautica, 2014, 97: 138-150.

[13] CAI J G, ZHOU Y, WANG X Y, et al. Dynamic analysis of a cylindrical boom based on Miura Origami[J]. Steel and Composite Structures, 2018, 28(5): 607-615.

[14] ANGELETTI F, GASBARRI P, SABATINI M. Optimal design and robust analysis of a net of active devices for microvibration control of an on-orbit large space antenna[J]. Acta Astronautica, 2019, 164: 241-253.

[15] CHEN W J, FANG G Q, HU Y. An experimental and numerical study of flattening and wrapping process of deployable composite thin-walled lenticular tubes[J]. Thin-Walled Structures, 2017, 111: 38-47.

[16] BESSA M A, PELLEGRINO S. Design of ultra-thin shell structures in the stochastic post-buckling range using Bayesian machine learning and optimization[J]. International Journal of Solids and Structures, 2018, 139-140: 174-188.

[17] BESSA M A, BOSTANABAD R, LIU Z, et al. A framework for data-driven analysis of materials under uncertainty: Countering the curse of dimensionality[J]. Computer Methods in Applied Mechanics and Engineering, 2017, 320: 633-667.

[18] ROYBAL F A, BANIK J A, MURPHEY T W. Development of an elastically deployable boom for tensioned planar structures: AIAA 2007-1838[R]. Reston: AIAA, 2007.

[19] YANG H,LU F S,GUO H W,et al. Design of a new N-shape composite ultra-thin deployable boom in the post-buckling range using response surface method and optimization[J]. IEEE Access,2019,7:129659-129665.

[20] 谢进德. 代理模型技术的研究及其在汽车抗撞性中的应用[D]. 长沙:湖南大学,2013.
XIE J D. Research on surrogate model technology and its application in vehicle crashworthiness[D]. Changsha: Hunan University,2013(in Chinese).

Stiffness optimization of M-shaped boom based on radial basis function surrogate model

YANG Hui[1,*], FAN Shuoshuo[2], WANG Yan[1], LIU Rongqiang[3], XIAO Hong[3]

(1. School of Mechanical Engineering, Yanshan University, Qinhuangdao 066044, China;
2. School of Electrical Engineering and Automation, Anhui University, Hefei 230601, China;
3. State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin 150001, China)

Abstract: In the execution of space missions, the hyper-elastic boom is mainly used in the deployment and support of large space deployable antennas and solar sails. In order to improve the supporting effect of the hyper-elastic boom in the state of deployment, the stiffness of the M-shaped hyper-elastic boom (M boom) with different sizes was studied. By using ABAQUS, the M boom's bending, compression, and torsion finite element models were created, and an explicit dynamic method was used to carry out a nonlinear numerical simulation of the M boom's buckling process. The bending stiffness, torsional stiffness and compression stiffness were taken as optimization objectives, the mass as constraint, and the cross-section arc length and radius as independent variables to establish the optimization model. The full factor method was used to design the experiment, the radial basis function (RBF) was used to establish the buckling surrogate model of the M boom, using the particle swarm optimization (PSO) algorithm to perform parameter optimization of the M boom. The ideal values for the length of bonding segment and the central angle were 7.894 5 mm and 26°, and the variation rule of stiffness with the length of bonding segment and the central angle was obtained.

Keywords: composite material; deployable structure; M-shaped boom; radial basis function; optimization

Received: 2021-02-24; **Accepted**: 2021-06-06; **Published online**: 2021-06-21 15:02

URL: kns.cnki.net/kcms/detail/11.2625.V.20210621.1255.002.html

Foundation items: Anhui Provincial Natural Science Foundation (2108085QE221); National Natural Science Foundation of China (51975001,51835002,52105013); Open Fundings of State Key Laboratory of Robotics and System (SKLRS-2020-KF-16)

* **Corresponding author**. E-mail: huiyang_0431@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0094

高超声速飞行器命中精度高准确度评估

彭维仕*
(武警工程大学 装备管理与保障学院, 西安 710086)

摘 要: 为科学合理地评估高超声速飞行器的命中精度,提出了一种基于误差谱的高超声速飞行器命中精度高准确度评估方法。对获得的高超声速飞行器命中精度数据进行预处理,得到该数据的落点偏差。构建算术平均误差、均方根误差、调和平均误差、误差中位数、误差众数、几何平均误差和迭代中距误差评估指标来反映高超声速飞行器命中精度的性能。分别建立高超声速飞行器命中精度评估模型及其可信度模型。通过实例验证所提评估方法的正确性和合理性。仿真结果表明:所提方法不仅能给出高超声速飞行器命中精度的评估结果,还能给出该结果的可信度。研究工作为高超声速飞行器命中精度的高准确度评估提供了新的科学评估方法。

关 键 词: 高超声速飞行器; 误差谱; 系统评论理论与方法; 命中精度评估方法; 可信度模型
中图分类号: E924.3
文献标志码: A **文章编号**: 1001-5965(2022)11-2130-08

高超声速飞行器具备快速打击、攻击广泛、突防力强和毁伤较高的特点,在未来信息化条件下的局部战争中具有非常广阔的应用前景[1]。但是,高超声速飞行器的研制过程具备技术先进、耗资巨大、周期长等特点,导致其试验和验证评估非常困难。因为常规武器的评估方法可能无法满足高超声速飞行器高准确度的评估,所以亟需研究新理论和新方法以解决高超声速飞行器中的评估难题。

目前,关于高超声速飞行器的评估研究主要集中在某一个关键部件的性能评估,并且大多只给出了评估结果,并没有分析评估结果的可信度。例如,为了保证高超声速飞行器研制方案的科学性和合理性,文献[2]提出了通过性能数据库评价空对空飞行器的性能。为了分析高超声速拦截飞行器的气动设计特点,文献[3]提出了一种性能分析方法。为分析高超声速拦截飞行器的气动设计特点,文献[4]出了一种性能分析方法。为了选择最优的高超声速飞行器设计方案,文献[5]提出了基于 TOPSIS 和综合效用函数法的高超声速飞行器多学科设计优选方法,但在计算权重时采用了主观赋权法,大致评估结果存在一定的误差。进一步,为了选取最优的高超声速飞行器弹道优化方法,文献[6]提出了一种基于灰色关联度分析理论的高超声速飞行器弹道优化算法的综合评估方法,通过层次分析法(analytic hierarchy process, AHP)确定指标权重,并利用灰色关联度法确定被估系统与理想系统的“距离”,最终得到各方案的优劣排序。为了分析高超声速飞行器控制性能指标中底层指标能否反映高超声速飞行器的飞行性能,文献[7]提出了一种基于统计实验分析的检验方法,将可行性分析问题转化为特定映射的双映射检验问题,通过可行域的实验点数判定方案的可行性。为了解决高超声速

收稿日期: 2021-03-01; **录用日期**: 2021-06-04; **网络出版时间**: 2021-07-12 09:16
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210709.1705.001.html
基金项目: 国家自然科学基金(71801222); 陕西省自然科学基金(2018JQ6019); 博士后创新人才支持计划(BX201700104)
* **通信作者**. E-mail: peng_weishi@163.com

引用格式: 彭维仕. 高超声速飞行器命中精度高准确度评估[J]. 北京航空航天大学学报, 2022, 48(11): 2130-2137.
PENG W S. Evaluation of high hitting accuracy performance of hypersonic vehicle[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2130-2137(in Chinese).

飞行器制导控制系统性能评估中评估效率低的问题,文献[8]提出了基于优化拉丁超立方试验设计方法的高超声速飞行器制导控制系统的性能评估方法,利用优化拉丁超立方试验设计方法采用较少的采样次数以充满空间,进而提高评估的效率。针对评估中使用确定数值来表征评估数据时,无法准确反映评估数据的不确定性和模糊性的问题,文献[9]提出了一种基于区间数排序的目标威胁评估方法。为了验证高超声速飞行器的性能,文献[10]提出了基于模型迁移的方法,通过仿射相似度模型,建立了高超声速飞行器的模型。文献[11]通过构建高超声速飞行器再入控制方法的性能指标体系,结合 AHP 法和熵值法的各自优点,定量评估了高超声速飞行器再入控制系统的性能。在高超声速飞行器可信度评估方面,文献[12]提出了一种高超声速飞行器制导控制模型置信度评价方法,对高超声速飞行器系统模型进行校核,给出了该模型的置信度。

综上所述,现有高超声速飞行器性能评估方法在评估时仅给出了评估结果,未能给出该评估结果的可信度,无法准确评估高超声速飞行器的效能。因此,本文提出了一种基于误差谱的高超声速飞行器命中精度高准确度评估方法。该方法不仅能给出高超声速飞行器命中精度的评估结果,还能给出该结果的可信度。

1　高超声速飞行器命中精度数据预处理

由于高超声速飞行器属于小样本试验,经典的性能评估方法难以应用。针对小样本问题,通常利用重采样的方法对样本进行扩容,或者基于半实物仿真的方法对样本进行扩容等[13-15]。

如图 1 所示,以目标点为圆心 O,构建坐标系 XOZ。令 $(\boldsymbol{X},\boldsymbol{Z})=[(x_1,z_1),(x_2,z_2),\cdots,(x_n,z_n)]$ 为高超声速飞行器的命中精度试验数据(已通过相容性检验),其中,n 为试验样本容量,并且试验数据满足独立同分布的条件。令目标位置为 $(\boldsymbol{X}_0,\boldsymbol{Z}_0)=(x_0,z_0)$,则横向和纵向偏差为

$$\tilde{\boldsymbol{X}}=\boldsymbol{X}-\boldsymbol{X}_0=[(x_1-x_0),(x_2-x_0),\cdots,(x_n-x_0)]=(\tilde{x}_1,\tilde{x}_2,\cdots,\tilde{x}_n) \tag{1}$$

$$\tilde{\boldsymbol{Z}}=\boldsymbol{Z}-\boldsymbol{Z}_0=[(z_1-z_0),(z_2-z_0),\cdots,(z_n-z_0)]=(\tilde{z}_1,\tilde{z}_2,\cdots,\tilde{z}_n) \tag{2}$$

如图 1 所示,可得落点偏差为

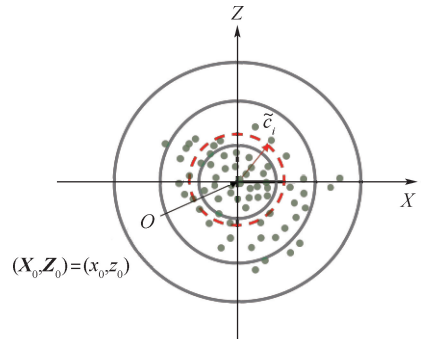


图 1　落点偏差示意图

Fig. 1　Diagram of relative distance error

$$\tilde{c}_i=\sqrt{\tilde{x}_i^2+\tilde{z}_i^2}\qquad i=1,2,\cdots,n \tag{3}$$

进一步,根据文献[14]中的扩容方法,对上述高超声速飞行器的原始数据进行扩容,其基本步骤如下:

步骤 1　随机产生一个服从均匀分布 $U(0,M)$ 的正整数 R,M 满足 $M\gg n$。

步骤 2　令 $p=\mathrm{mod}(R,n)$,其中,$\mathrm{mod}(\cdot)$ 表示求余函数,且当 $\mathrm{mod}(R,n)=0$ 时,令 $p=1$;当 $\mathrm{mod}(R,n)>n$ 时,令 $p=n$。

步骤 3　得到第 1 次复制样本中的第 1 个数据 $(x_1^*,z_1^*)=(x_p,z_p)$,其中,x_p 表示原始样本中的第 p 个元素,重复步骤 1 和步骤 2,n 次得到第 1 次复制的样本 $(x_1^*,z_1^*)=[(x_1^*,z_1^*),\cdots,(x_i^*,z_i^*),\cdots,(x_n^*,z_n^*)]$;进一步计算第 1 次复制样本的均值 $(\bar{x}_1^*,\bar{z}_1^*)=\left(\sum_{i=1}^{n}x_i^*/n,\sum_{i=1}^{n}z_i^*/n\right)$。

步骤 4　重复步骤 1 ~ 步骤 3,B 次得到重采样样本 $(\boldsymbol{X}^*,\boldsymbol{Z}^*)=[(x_1^*,z_1^*),\cdots,(x_B^*,z_B^*)]$。

步骤 5　计算重采样样本与原始样本的相似度 $\rho(f(x);f(x^*))$,如果 $\rho(f(x);f(x^*))\geqslant\rho_\varepsilon$,则输出重采样样本 $(\boldsymbol{X}^*,\boldsymbol{Z}^*)=[(x_1^*,z_1^*),\cdots,(x_B^*,z_B^*)]$,否则重复步骤 1 ~ 步骤 4 直到满足相似度条件。

扩容后样本的落点偏差为

$$\tilde{c}_j^*=\sqrt{(\tilde{x}_j^*)^2+(\tilde{z}_j^*)^2}\qquad j=1,2,\cdots,B \tag{4}$$

为了方便,将扩容样本的落点偏差记为 $\tilde{\boldsymbol{c}}^*=\{\tilde{c}_j^*\}_{j=1}^{B}$,$B$ 为扩容后的样本容量。

2　高超声速飞行器命中精度评估指标模型

2.1　算术平均误差和均方根误差评估指标模型

得到扩容样本 $\tilde{\boldsymbol{c}}^*=\{\tilde{c}_j^*\}_{j=1}^{B}$ 后,常用的命中精度评估指标为准确度和密集度。二者分别利用样本均值和方差表征弹着点与目标瞄准点的偏离

程度和离散程度,即[14]

$$\begin{cases} \mathrm{AEE}(\tilde{\boldsymbol{c}}^*) = \bar{c}^* = \dfrac{1}{B}\sum_{j=1}^{B}\tilde{c}_j^* \\ \tilde{\sigma}_{\tilde{c}^*}^2 = \dfrac{1}{B-1}\sum_{j=1}^{B}(\tilde{c}_j^* - \bar{c}^*)^2 \end{cases} \tag{5}$$

均方根是标准差 $E(\sqrt{\tilde{\boldsymbol{c}}^*\tilde{\boldsymbol{c}}^{*\mathrm{T}}})$ 的有限样本近似,因此常用均方根误差替换标准差[14]:

$$\mathrm{RMSE}(\tilde{\boldsymbol{c}}^*) = \left[\frac{1}{B}\sum_{j=1}^{B}(\tilde{c}_j^*)^2\right]^{1/2} \tag{6}$$

但是实际使用中,有时需关注极小值影响,也就是最高精度能达到多高水平,为此定义调和平均误差评估指标。

2.2 调和平均误差评估指标模型

调和平均误差评估指标的定义为[14]

$$\mathrm{HAE}(\tilde{\boldsymbol{c}}^*) = \left[\frac{1}{B}\sum_{j=1}^{B}(\tilde{c}_j^*)^{-1}\right]^{-1} \tag{7}$$

显然,调和平均误差评估指标对误差取倒数后,关注的是极小误差,因此常用于回答命中与否的问题。特别地,当 $\tilde{c}^{*k}=0$ 时,$1/c_j^*$ 无意义,使得 $\mathrm{HAE}(\tilde{\boldsymbol{c}}^*)$ 无法使用,为此令 $\tilde{c}^{*k}=0.000\ 1$ 或者换成饱和函数,如 Tanh 函数、Softplus 函数等。

在工程实际中,为了快速得到评估结果,并且遵从原始样本的信息,通常采用误差中位数和误差众数来衡量系统的命中精度。

2.3 误差中位数和误差众数评估指标模型

误差中位数是指在样本 $\tilde{\boldsymbol{c}}^*$ 中利用中间序号对应的样本点作为系统的命中精度,即[14]

$$\mathrm{ME}(\tilde{\boldsymbol{c}}^*) = \begin{cases} \tilde{c}_{B/2}^* & B\text{ 为奇数} \\ \dfrac{\tilde{c}_{B/2}^* + \tilde{c}_{(B+1)/2}^*}{2} & B\text{ 为偶数} \end{cases} \tag{8}$$

可见,中位数未考虑其他样本点的信息,仅考虑了中间点样本的信息。为了克服该问题,工程实际中通常采用误差众数来衡量系统的精度。而误差众数 $\mathrm{EMM}(\tilde{\boldsymbol{c}}^*)$ 是指样本 $\tilde{\boldsymbol{c}}^*$ 中出现频率最大的值。

进一步,在工程实际中,为了评估系统命中精度的平稳性,定义几何平均误差评估指标。

2.4 几何平均误差评估指标模型

几何平均误差评估指标的定义为[14]

$$\mathrm{GAE}(\tilde{\boldsymbol{c}}^*) = \exp\left[\frac{1}{B}\sum_{j=1}^{B}(\ln\tilde{c}_j^*)\right] \tag{9}$$

几何平均误差评估指标既不受粗大值影响,也不受极小值影响,因此其能够反映系统命中精度的平稳性。同样的,当 $\tilde{c}^{*k}=0$ 时,使得 $\mathrm{GAE}(\tilde{\boldsymbol{c}}^*)$ 也无法使用,令 $\tilde{c}^{*k}=0.000\ 1$ 或者换成饱和函数。

2.5 迭代中距误差评估指标模型

为了克服算术平均误差和均方根误差评估指标受大误差的影响,文献[16]又给出了迭代中距误差评估指标,即

$$\mathrm{IMRE}(\tilde{\boldsymbol{c}}^*) = \left\{\frac{\min\{\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B}\} + \max\{\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B}\}}{2}\right\}_{k'=1}^{B-1} \tag{10}$$

显然,迭代中距误差评估指标首先对样本数据进行排序 $\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B}$,然后取排序后样本中的最大值 $\max\{\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B}\}$ 和最小值 $\min\{\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B}\}$ 的平均值作为新的样本点 $\tilde{c}_{j(\mathrm{new})}^{*k'}$,最后用新的样本点 $\tilde{c}_{j(\mathrm{new})}^{*k'}$ 代替上一次样本中的最大值 $\max\{\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B-1}\}$ 和最小值 $\min\{\mathrm{sort}\{\tilde{c}_j^{*k'}\}_{j=1}^{B-1}\}$,如此循环迭代 $B-1$ 次后输出迭代结果。

下面举例说明迭代中距误差的实施过程。假设获得的误差数据为{1,3,8,9,15},第1次 $k'=1$ 迭代中距误差为 $\mathrm{IMRE}(\tilde{\boldsymbol{c}}^*)_1=(1+15)/2=8$;则新的误差数据为{3,8,8,9},第2次 $k'=2$ 迭代中距误差为 $\mathrm{IMRE}(\tilde{\boldsymbol{c}}^*)_2=(3+9)/2=6$;同理新的误差数据为{6,8,8},第3次 $k'=3$ 迭代中距误差为 $\mathrm{IMRE}(\tilde{\boldsymbol{c}}^*)_3=(6+8)/2=7$;新的误差数据为{7,8},最终 $k'=4$ 的迭代中距误差为 $\mathrm{IMRE}(\tilde{\boldsymbol{c}}^*)_4=(7+8)/2=7.5$。

3 基于误差谱的高超声速飞行器命中精度高准确度评估

基于误差谱的高超声速飞行器命中精度高准确度评估是指不仅要给出高超声速飞行器命中精度的结果,还要给出高超声速飞行器命中精度评估结果的可信度。

3.1 高超声速飞行器命中精度评估模型

因为定义算术平均误差、均方根误差、调和平均误差、误差中位数、误差众数、几何平均误差和迭代中距误差评估指标在评估高超声速飞行器命中精度时仅能从某一个方面反映系统的性能,所以为了准确、综合和全面地评估高超声速飞行器的命中精度,本文提出误差谱度量(error spectrum, ES)来评估高超声速飞行器的命中精度。

误差谱度量是一个综合性度量[13,17],其包含了很多常用的绝对误差度量,综合考虑了大误差和小误差的影响,因此能够给出一个更加公正的评估结果。下面介绍误差谱度量的定义。

令 $e_j=\|\tilde{c}_j^*\|$ 表示绝对误差范数,r 为[$-\infty$,

$+\infty$]上的实数,则误差谱度量的定义为[13,17]

$$S_e(r) = \{E[(e^r)]\}^{1/r} = \left\{\int e^r \mathrm{d}F(e)\right\}^{1/r} = \begin{cases} \left(\int e^r f(e)\,\mathrm{d}e\right)^{1/r} & e \text{ 为连续变量} \\ \left(\sum e_i^r p_i\right)^{1/r} & e \text{ 为离散变量} \end{cases} \tag{11}$$

式中:$F(e)$、$f(e)$和 p_i 分别表示 e 的累积分布函数、概率密度函数和分布律。

由式(11)可得,对于给定的 e,$S(r)$是关于 r 变化的一条曲线,并且误差谱曲线越低,系统的性能越好[17-18]。特别地,当 $r=-1,0,1,2$ 时,该段误差谱曲线包含了常用的 4 个绝对误差度量的指标,即 RMSE、AEE、GAE 和 HAE。

3.2　高超声速飞行器命中精度可信度模型

根据高超声速飞行器的特点,给出命中精度的评估结果后,还需给定对该结果的可信度,最终给出高超声速飞行器高准确度的评估结果。下面定义高超声速飞行器命中精度的可信度模型。

一般情况下,可信度评估可分为 2 种。第 1 种是将评估结果与真值比较,通过比较评估结果与正值的接近程度,给出该评估结果的可信度。但是,该方法在工程实际中使用比较困难,因为真值一般情况无法获得,通常采用一个与真值比较接近的值作为标准。可见,该方法的难点是寻找一个标准,作为可信度评估的参考。

如图 2 所示,第 2 种方法是专家投票法,基于少数服从多数的原理,计算评估结果的可信度。例如,若选择 $N_{ZJ}=10$ 位专家对系统进行打分,其中有 6 位专家一致认为实测值接近真实值,则该测量结果的一致程度为 $6/10=0.6$。通常情况下,将这种一致程度称为测量结果的可信度。因为第 2 种方法实现起来比较容易,所以得到了广泛应用[13-14]。下面给出高超声速飞行器命中精度的可信度。

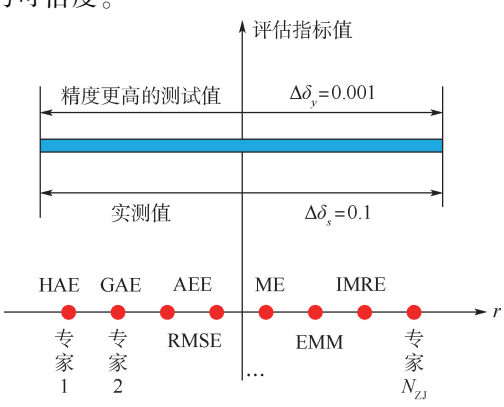


图 2　两种可信度评估方法示意图

Fig. 2　Schematic of two credibility assessment methods

高超声速飞行器命中精度可信度是指利用不同的评估指标,评价相同条件下高超声速飞行器的命中精度时评价结果的一致性程度。

比较两型高超声速飞行器的命中精度,得到 2 组高超声速飞行器的命中精度的扩容数据 $\tilde{c}^{*1}$ 和 $\tilde{c}^{*2}$,代入本文提出的算术平均误差、均方根误差、调和平均误差、误差中位数、误差众数(EMM)、几何平均误差和迭代中距误差评估指标中得到

$$\begin{cases} \mathrm{HAE}_1 > \mathrm{HAE}_2 \\ \mathrm{GAE}_1 > \mathrm{GAE}_2 \\ \mathrm{ME}_1 > \mathrm{ME}_2 \\ \mathrm{EMM}_1 > \mathrm{EMM}_2 \\ \mathrm{IMRE}_1 > \mathrm{IMRE}_2 \\ \mathrm{AEE}_1 < \mathrm{AEE}_2 \\ \mathrm{RMSE}_1 < \mathrm{RMSE}_2 \end{cases} \tag{12}$$

可以看出,针对这两型高超声速飞行器的命中精度,利用不同的指标得到不同的评估结果。上述评估结果中,HAE、GAE、ME、EMM 和 IMRE 一致认为系统 2 的精度要高于系统 1,因为系统 2 的误差小于系统 1 的误差。只有 AEE 和 RMSE 认为系统 1 的精度要高于系统 2。该评估结果的一致程度为 $M_c=5/7=0.71$。因此可得,系统 2 的精度要高于系统 1,该评估结果的可信度为 $M_c=5/7=0.71$。

下面根据 3.1 节中的命中精度度量指标,给出高超声速飞行器命中精度可信度的计算模型。

3.2.1　建立高超声速飞行器命中精度评估矩阵

将 $\tilde{\boldsymbol{c}}^* = \{\tilde{c}_j^*\}_{j=1}^{B}$ 代入 3.1 节中定义的 7 项高超声速飞行器命中精度评估指标,从而得到高超声速飞行器命中精度评估矩阵:

$$\boldsymbol{H} = \begin{bmatrix} \mathrm{RMSE}_1 & \mathrm{RMSE}_2 & \cdots & \mathrm{RMSE}_N \\ \mathrm{AEE}_1 & \mathrm{AEE}_2 & \cdots & \mathrm{AEE}_N \\ \mathrm{GAE}_1 & \mathrm{GAE}_2 & \cdots & \mathrm{GAE}_N \\ \mathrm{HAE}_1 & \mathrm{HAE}_2 & \cdots & \mathrm{HAE}_N \\ \mathrm{ME}_1 & \mathrm{ME}_2 & \cdots & \mathrm{ME}_N \\ \mathrm{EMM}_1 & \mathrm{EMM}_2 & \cdots & \mathrm{EMM}_N \\ \mathrm{IMRE}_1 & \mathrm{IMRE}_2 & \cdots & \mathrm{IMRE}_N \end{bmatrix}_{(1,\cdots,7)\times N} \tag{13}$$

式中:N 表示 N 个被评的高超声速飞行器。

3.2.2　建立高超声速飞行器命中精度可信度评估矩阵

基于上述高超声速飞行器命中精度评估矩阵 $\boldsymbol{H}$,运用皮氏接近度原则(Pitman's closeness measure)建立高超声速飞行器命中精度可信度评估

矩阵。

记 $M_c(j,k;\boldsymbol{H}_{i(j,k)})$ 为打分矩阵 $\boldsymbol{H}$ 中的第 j 列与第 k 列元素的比较结果,得分的依据为

$$R(j,k;\boldsymbol{H}_{i(j,k)}) = \begin{cases} 1 & H_{ij} < H_{ik} \\ 0.5 & H_{ij} = H_{ik} \\ 0 & H_{ij} > H_{ik} \end{cases} \tag{14}$$

式中:$H_{ij} < H_{ik}$ 表示用第 i 个评估指标评估时,第 j 个高超声速飞行器的命中精度优于第 k 个高超声速飞行器的命中精度,因为 $\boldsymbol{H}$ 中的元素是误差值;$H_{ij} = H_{ik}$ 表示用第 i 个评估指标评估时,第 j 个高超声速飞行器的命中精度与第 k 个高超声速飞行器的命中精度相当。

进一步,可得高超声速飞行器命中精度的可信度为

$$R(j,k;\boldsymbol{H}_{(1,\cdots,7)\times N}) = \frac{1}{7}\sum_{i=1}^{7} R(j,k;\boldsymbol{H}_{i(j,k)}) \tag{15}$$

进一步得到反映系统可信度的矩阵,即

$$\boldsymbol{R} = \begin{bmatrix} R(1,1;\boldsymbol{H}_{(1,\cdots,7)\times N}) & \cdots & R(1,N;\boldsymbol{H}_{(1,\cdots,7)\times N}) \\ \vdots & & \vdots \\ R(7,1;\boldsymbol{H}_{(1,\cdots,7)\times N}) & \cdots & R(7,N;\boldsymbol{H}_{(1,\cdots,7)\times N}) \end{bmatrix} \tag{16}$$

当 $R(j,k;\boldsymbol{H}_{(1,\cdots,7)\times N}) > 0.5$ 时,则系统 j 的命中精度比系统 k 的命中精度高,且该评估结果的可信度为 $R(j,k;\boldsymbol{H}_{(1,\cdots,7)\times N})$。

综上所述,根据式(11)和式(16)可得最终的评估结果如下。

命中精度:

$$S_e(r)_p > S_e(r)_q \tag{17a}$$

可信度:

$$R(p,q;\boldsymbol{H}_{(1,\cdots,7)\times N}) \tag{17b}$$

式中:$S_e(r)_p > S_e(r)_q$ 表示第 p 个高超声速飞行器的命中精度优于第 q 个高超声速飞行器的命中精度,$p,q = 1,2,\cdots,N$。

特别地,当仅评估一个高超声速飞行器的命中精度时,将期望的高超声速飞行器的命中精度作为参考对象(或为预先设计的高超声速飞行器的命中精度)$S_e(r)_o$。

命中精度:

$$S_e(r)_p > S_e(r)_o \tag{18a}$$

可信度:

$$R(p,o;\boldsymbol{H}_{(1,\cdots,7)\times N}) \tag{18b}$$

认为被评估的高超声速飞行器命中精度满足要求,该结果的可信度为 $M_c(p,o;\boldsymbol{H}_{(1,\cdots,7)\times N})$,否则该系统的命中精度不满足要求,而不满足要求的可信度为 $M_c(o,p;\boldsymbol{H}_{(1,\cdots,7)\times N})$。

4 仿真验证

为了验证本文方法的正确性,评估某三型高超声速飞行器命中精度。假设数据服从二维正态分布,分布参数如表 1 所示。

表 1 三次试验服从的二维正态分布参数

Table 1 Two-dimensional normal distribution which parameters of three tests follow

试验数据	次数	扩容样本	均值	方差
Test data1	50	1 000	[4,3]	$\begin{bmatrix}6 & 1\\ 1 & 7\end{bmatrix}$
Test data2	30	1 000	[2.5,5]	$\begin{bmatrix}4 & 3\\ 3 & 9\end{bmatrix}$
Test data3	45	1 000	[3,3]	$\begin{bmatrix}3 & 1\\ 1 & 8\end{bmatrix}$

4.1 单次仿真验证

用文献[14]中的扩容方法后得到的数据如图 3所示。由式(14)可得到误差数据如图 4 所示。

进一步,将上述 3 组数据代入式(12)可得数据的误差谱曲线,如图 5 所示。

由图 5 可得,可以直观看出 3 组数据对应的高超声速飞行器命中精度的优劣,即

$$\text{Test data1} < \text{Test data2} < \text{Test data3} \tag{19}$$

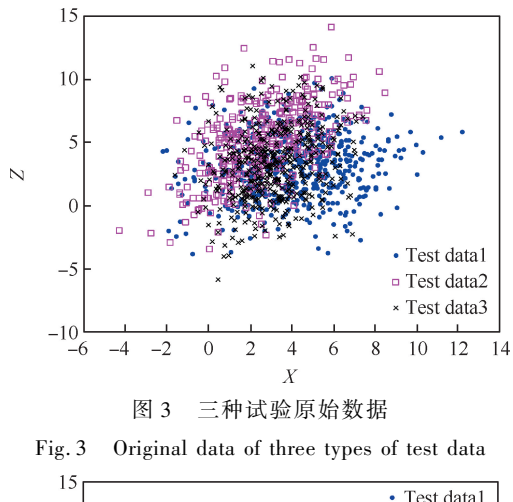


图 3 三种试验原始数据

Fig. 3 Original data of three types of test data

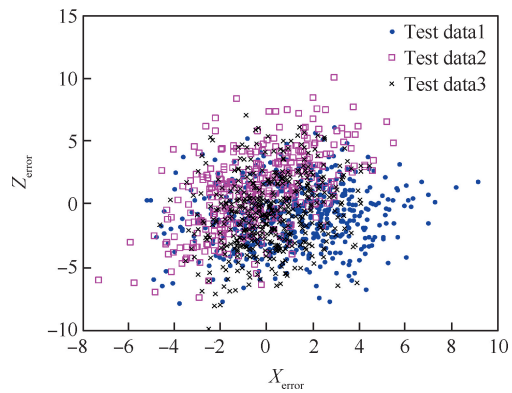


图 4 三种试验误差数据

Fig. 4 Error data of three types of test data

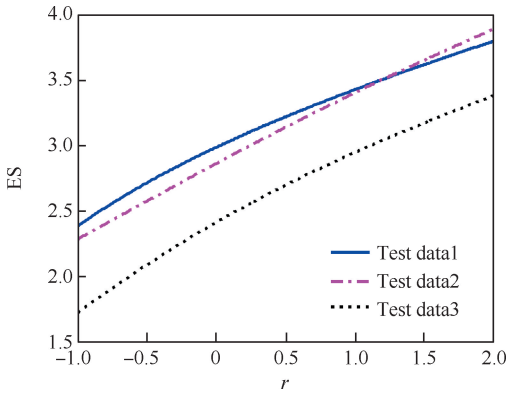


图 5　三种试验数据误差谱曲线

Fig. 5　Error spectrum curves of three types of test data

式中:$X \prec Y$ 表示系统 Y 的命中精度优于系统 X 的命中精度。

将数据代入式(13)可得

$$\boldsymbol{H} = \begin{bmatrix} 3.800\ 7 & 3.892\ 7 & 3.383\ 5 \\ 3.431\ 0 & 3.406\ 8 & 2.951\ 1 \\ 2.990\ 8 & 2.866\ 9 & 2.417\ 7 \\ 2.389\ 8 & 2.288\ 7 & 1.726\ 4 \\ 3.231\ 6 & 3.126\ 6 & 2.743\ 7 \\ 3.373\ 3 & 2.913\ 6 & 2.650\ 3 \\ 3.386\ 1 & 3.324\ 2 & 2.900\ 3 \end{bmatrix} \tag{20}$$

进一步,根据式(16)可得高超声速飞行器命中精度的可信度矩阵为

$$\boldsymbol{R} = \begin{bmatrix} 0.5 & 0.142\ 9 & 0 \\ 0.857\ 1 & 0.5 & 0 \\ 1 & 1 & 0.5 \end{bmatrix} \tag{21}$$

可见,$\boldsymbol{R}(3,1) = 1 > 0.5$,则说明系统 3 的命中精度优于系统 1 的命中精度,且可信度为 $\boldsymbol{R}(3,1) = 1$。又根据误差谱曲线的特征可知,曲线越低,性能越好,显然系统 3 对应的虚线误差谱曲线低于系统 1 对应的实线误差谱曲线。结果证明上述可信度评估的正确性。

同理,$\boldsymbol{R}(2,1) = 0.857\ 1 > 0.5$,则说明系统 2 的命中精度优于系统 1 的命中精度,且可信度为 $\boldsymbol{R}(2,1) = 0.857\ 1$。对应的误差谱曲线可得,点画线误差谱曲线与实线误差谱曲线有交叉,也进一步说明有“专家认为”系统 1 的命中精度优于系统 2 的命中精度。

4.2　多次仿真验证

根据 4.1 节中的仿真,运用 MATLAB 软件仿真 $M = 1\ 000$ 次,结果如图 6 ~ 图 8 所示。

进一步得到平均可信度矩阵为

$$\overline{\boldsymbol{R}} = \sum_{k=1}^{M} M_{ck} = \begin{bmatrix} 0.5 & 0.149\ 9 & 0.032\ 6 \\ 0.850\ 1 & 0.5 & 0.097\ 3 \\ 0.967\ 4 & 0.902\ 7 & 0.5 \end{bmatrix} \tag{22}$$

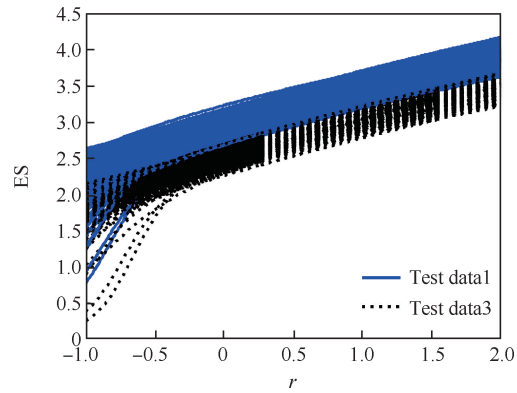


图 6　试验数据 1 和试验数据 3 的误差谱曲线

Fig. 6　Error spectrum curves of Test data1 and Test data3

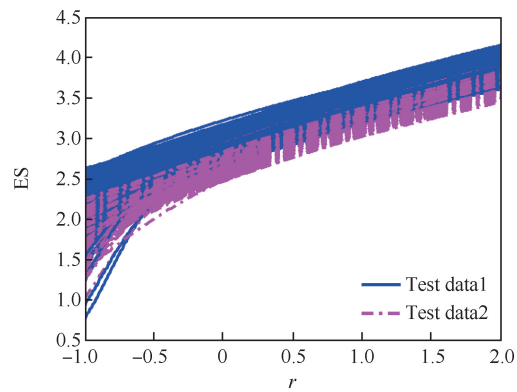


图 7　试验数据 1 和试验数据 2 的误差谱曲线

Fig. 7　Error spectrum curves of Test data1 and Test data2

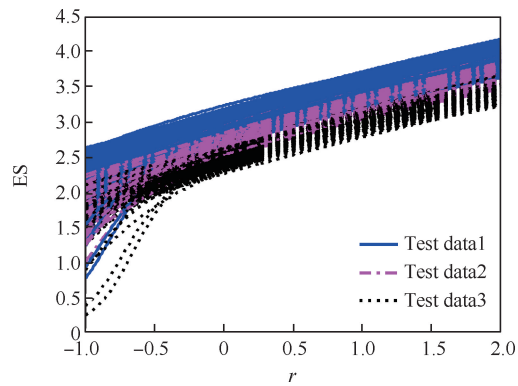


图 8　三种试验数据的误差谱曲线

Fig. 8　Error spectrum curves of three types of test data

可得 $\overline{\boldsymbol{R}}(3,1) = 0.967\ 4 > 0.5$,则说明系统 3 的命中精度优于系统 1 的命中精度,且可信度为 $\overline{\boldsymbol{R}}(3,1) = 0.967\ 4$。

同理,如图 6 所示,系统 3 对应的点线误差谱曲线低于系统 1 对应的实线误差谱曲线,因此,系统 3 的命中精度优于系统 1 的命中精度。该结果与单次仿真的结果一致。但是,从图 6 中可以看出,有少部分结果显示,系统 3 对应的点线误差谱曲线高于系统 1 对应的实线误差谱曲线。换句话说,就是系统 3 的命中精度不总是优于系统 1 的命中精度,即系统 3 的命中精度优于系统 1 的命

中精度的可信度不为1。利用本文方法得到该可信度为 $\overline{\boldsymbol{R}}(3,1)=0.967\ 4$,可以准确反映该结果,显然再次证明上述可信度评估的结果是正确的。

同理,$\overline{\boldsymbol{R}}(2,1)=0.850\ 1>0.5$,则说明系统2的命中精度优于系统1的命中精度,且可信度为 $\overline{\boldsymbol{R}}(2,1)=0.850\ 1$。同样的,图7有部分结果显示,系统2对应的点画线误差谱曲线高于系统1对应的实线误差谱曲线,此时说明系统1的命中精度优于系统2的命中精度。引入可信度能够为决策者提供更加可靠、准确的评估结果。通过可信度矩阵,给出相应的评估结果,更加准确地评估系统的优劣,而高超声速飞行器亟须这样的评估方法。

进一步由图8可得,在多次仿真的条件下,利用不同的评估指标评估高超声速飞行器的命中精度时,可能得到不一致的评估结果。此时,应该考虑用多种评估指标评估高超声速飞行器的命中精度,给出最终评估结果的一致性,即该评估结果的可信度,这样才能准确地给出高超声速飞行器的命中精度。

综上所述,可得三型高超声速飞行器命中精度评估结果为

$$\begin{cases}\text{Test data1}<\text{Test data2}\\ \overline{\boldsymbol{R}}(2,1)=0.850\ 1\end{cases}\tag{23}$$

$$\begin{cases}\text{Test data1}<\text{Test data3}\\ \overline{\boldsymbol{R}}(3,1)=0.967\ 4\end{cases}\tag{24}$$

$$\begin{cases}\text{Test data2}<\text{Test data3}\\ \overline{\boldsymbol{R}}(3,2)=0.902\ 7\end{cases}\tag{25}$$

5 结 论

本文提出了一种高超声速飞行器命中精度高准确度评估方法,不仅能给出高超声速飞行器命中精度评估结果,还能给出该结果的可信度。

1)建立了算术平均误差、均方根误差、调和平均误差、误差中位数、误差众数、几何平均误差和迭代中距误差的评估模型,从不同的方面反映了系统的命中精度,为后续高超声速飞行器命中精度及其可信度评估提供指标支撑。

2)为了综合反映高超声速飞行器命中精度的性能,提出了用误差谱曲线评估高超声速飞行器的命中精度,克服了单一评估指标仅反映系统一方面性能的缺陷。

3)将上述高超声速飞行器命中精度评估指标定义为可信度评估中的“打分专家”,进而构建高超声速飞行器命中精度评估矩阵,并根据皮氏接近度原则建立高超声速飞行器命中精度可信度评估矩阵,并给出了可信度评估结果。

基于高超声速飞行器命中精度高准确度评估方法还需经过更多的实际应用才能验证该方法的可行性。

参考文献(References)

[1] WANG W J. Hypersonic vehicle aerodynamic design using modified sequential approximate optimization[J]. Advances in Space Research,2019,63(11):3706-3720.

[2] LAUGHREY J A. Performance evaluation of an air vehicle utilizing nonaxisymrnetric nozzles[J]. Journal of Aircraft, 1981, 18(2):89-95.

[3] GONG C L,GOU J J,HU J X,et al. A novel TE-material based thermal protection structure and its performance evaluation for hypersonic flight vehicles[J]. Aerospace Science and Technology,2018,77:458-470.

[4] JAIMIE T,RAMANA G,WILBUR H,et al. Performance analysis method for evaluating design features of tactical hypersonic intercepts[C]// AIAA of Navigation and Control Conference in New Orleans. Reston:AIAA,1991:1557-1565.

[5] 邢永刚,周小勇,李懿,等. 高超声速飞行器MDO设计综合评估研究[J]. 陕西理工大学学报(自然科学版),2019,35(5):72-79.
XING Y G,ZHOU X Y,LI Y,et al. Research on the comprehensive assessment of multidisciplinary design optimization for hypersonic vehicle[J]. Journal of Shaanxi University of Technology (Natural Science Edition),2019,35(5):72-79(in Chinese).

[6] 胡怀中,郭茹霜,宫厚娟,等. 高超声速飞行器的弹道优化评估方法[J]. 国防科技大学学报,2015,37(2):117-120.
HU H Z,GUO R S,GONG H J,et al. Evaluation method for hypersonic vehicle trajectory optimization[J]. Journal of National University of Defense Technology,2015,37(2):117-120(in Chinese).

[7] 朱鸿绪,刘燕斌,曹瑞,等. 高超声速飞行器底层性能评价指标的可行性分析[J]. 航空学报,2020,41(3):323259.
ZHU H X,LIU Y B,CAO R,et al. Feasibility analysis for underlying indictors in control performance evaluation of hypersonic vehicles[J]. Acta Aeronautica et Astronautica Sinica,2020,41(3):323259(in Chinese).

[8] 张登辉,马萍,晁涛,等. 高超声速飞行器制导控制系统性能评估[J]. 系统工程与电子技术,2018,40(8):148-153.
ZHANG D H,MA P,CHAO T,et al. Performance evaluation of guidance and control system for hypersonic vehicle[J]. Systems Engineering and Electrinics,2018,40(8):148-153(in Chinese).

[9] 王楚燕,战晓苏. 基于区间数排序的高超声速飞行器目标威胁评估方法[J]. 军事运筹与系统工程,2017,31(1):17-21.
WANG C Y,ZHAN X S. Hypersonic vehicle target threat assessment method based on interval number ranking[J]. Military Operations Research and Systems Engineering,2017,31(1):17-21(in Chinese).

[10] 王浩亮,李清东,任章,等. 基于模型迁移方法的高超声速飞行器建模[J]. 北京航空航天大学学报,2016,42(12):2540-2647.
WANG H L, LI Q D, REN Z, et al. Modeling of hypersonic vehicle via model migration method[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(12): 2540-2647(in Chinese).

[11] 杨琪琛,宗群,董琦. 高超声速飞行器再入控制与性能评估方法[J]. 信息与控制,2017,46(1):33-40.
YANG Q C, ZONG Q, DONG Q. Reentry control and performance evaluation method for hypersonic vehicle[J]. Information and Control, 2017, 46(1): 33-40(in Chinese).

[12] 卢青. 高超声速飞行器制导控制与评估方法研究[D]. 西安:西北工业大学,2018:75-77.
LU Q. Research of guidance control and evaluation for hypersonic vehicle[D]. Xi'an: Northwestern Polytechnical University, 2018: 75-77(in Chinese).

[13] 彭维仕. 误差谱理论及其在导弹系统性能评估中的应用研究[D]. 西安:空军工程大学,2016:59-61.
PENG W S. Study on error spectrum theory and its application in missile system performance evaluation[D]. Xi'an: Air Force Engineering University, 2016: 59-61(in Chinese).

[14] PENG W S, FANG Y W, ZHAN R J, et al. Weapon systems accuracy evaluation using the error spectrum[J]. Aerospace Science and Technology, 2016, 58: 369-379.

[15] PENG W S. Model selection for Gaussian mixture model based on desirability level criterion[J]. Optik, 2017, 130: 797-805.

[16] YIN H L, LI X R, LAN J. Iterative mid-range with application to estimation performance evaluation[J]. IEEE Signal Process Letters, 2015, 22(11): 2044-2048.

[17] LI X R, ZHAO Z L, DUAN Z S. Error spectrum and desirability level for estimation performance evaluation[C]// Proceedings of Workshop on Estimation, Tracking and Fusion, 2007: 1-7.

[18] PENG W S, FANG Y W, DUAN Z S, et al. Enhanced error spectrum for estimation performance evaluation[J]. Optik, 2016, 127: 5084-5091.

Evaluation of high hitting accuracy performance of hypersonic vehicle

PENG Weishi*

(School of Equipment Management and Support, People Armed Police Engineering University, Xi'an 710086, China)

Abstract: To evaluate the hitting accuracy of hypersonic vehicles scientifically and reasonably, this paper proposes a high accuracy evaluation method using the error spectrum. First, the hit accuracy data of hypersonic vehicles are preprocessed to obtain the relative distance error. Secondly, the root mean square, average Euclidean error, harmonic average error, median error, error mode metric, geometric average error and iterative mid-range error are constructed to reflect the hitting accuracy of hypersonic vehicles. Then, we build the high accuracy evaluation model and its credibility model for hypersonic vehicles. Finally, two examples are given to verify the correctness and rationality of the proposed method. The simulation results show that this method can obtain not only the result of the hit accuracy evaluation, but also the credibility of the result. This research provides a new scientific evaluation method for evaluating high hitting accuracy of hypersonic vehicles.

Keywords: hypersonic vehicle; error spectrum; systematic evaluation theory and method; hit accuracy evaluation method; credibility model

Received: 2021-03-01; **Accepted**: 2021-06-04; **Published online**: 2021-07-12 09:16

URL: kns.cnki.net/kcms/detail/11.2625.V.20210709.1705.001.html

Foundation items: National Natural Science Foundation of China (71801222); National Science Foundation of Shaanxi Province of China (2018JQ6019); National Postdoctoral Program for Innovative Talents (BX201700104)

* **Corresponding author**. E-mail: peng_weishi@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0100

基于 PPO 的移动平台自主导航

徐国艳*, 熊绎维, 周彬, 陈冠宏
(北京航空航天大学 交通科学与工程学院 特种车辆无人运输技术工业和信息化部重点实验室, 北京 100191)

摘 要: 为解决强化学习算法在自主导航任务中动作输出不连续、训练收敛困难等问题,提出了一种基于近似策略优化(PPO)算法的移动平台自主导航方法。在 PPO 算法的基础上设计了基于正态分布的动作策略函数,解决了移动平台整车线速度和横摆角速度的输出动作连续性问题。设计了一种改进的人工势场算法作为自身位置评价,有效提高强化学习模型在自主导航场景中的收敛速度。针对导航场景设计了模型的网络框架和奖励函数,并在 Gazebo 仿真环境中进行模型训练,结果表明,引入自身位置评价的模型收敛速度明显提高。将收敛模型移植入真实环境中,验证了所提方法的有效性。

关 键 词: 近似策略优化算法; 移动平台; 自主导航; 强化学习; 人工势场
中图分类号: TP242.6
文献标志码: A **文章编号**: 1001-5965(2022)11-2138-08

自主导航作为无人驾驶技术中的关键,可通过结合感知、决策和控制等功能[1],来解决“我在哪”“要去哪”及“怎么去”等主要问题[2]。预期实现的功能为:用户指定目的地后,智能体可在非碰撞条件下自行去往目的地。

传统实现自主导航功能的算法多为基于搜索的路径规划算法,如 A*[3]、D*[4-5]、快速扩展随机树(rapidly exploring random tree, RRT)[6]等。此类算法的基本思想是:通过定位自身位置,根据目的点位置和障碍物位置,搜索出一条最优路径。此类算法的稳定性高,应用于实际场景时可以得到很好的效果,但存在无法进行自我学习、在复杂环境下适用性较低等问题。

强化学习的使用可以有效解决上述问题。强化学习的基本思想是:通过与环境交互得到奖励反馈,根据反馈自行更新模型参数。由于其独特的学习机制,模型可以在训练完成后迁移至无先验知识的场景中使用,且学习的场景越多,模型表现越好。

因此,近年来国内外众多专家学者开展了基于强化学习的自主导航研究。2017 年,付雪建[7]以超声波和视觉作为环境感知器,构造了基于 Q-learning的强化学习模型,将其应用在自主导航场景中,提高了移动机器人对未知环境的适应性。2019 年,为了解决强化学习的输入维度灾难问题,杨宁博[8]使用即时定位与地图构建(simultaneous localization and mapping, SLAM)完成了室内自身定位,并使用基于深度 Q 学习网络(deep Q network, DQN)的强化学习算法,在无先验地图的情况下实现自主导航移动机器人。2020 年,陶睿[9]研究了基于 DQN 及其改进 Double DQN、Dueling DQN 的强化学习算法,并分析和总结了强化学习算法在自主导航实际应用中各参数的选择方法。2021 年,何聪[10]针对无地图环境下的导航任务,提出了基于智能体自我定位和深度预测辅助任务的深度强化学习模型,实现了无地图环境

收稿日期: 2021-03-02; **录用日期**: 2021-04-11; **网络出版时间**: 2021-05-07 08:55
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210506.1639.002.html
基金项目: 国家自然科学基金 (51775016)
* **通信作者**. E-mail: xuguoyan@buaa.edu.cn

引用格式: 徐国艳, 熊绎维, 周彬, 等. 基于 PPO 的移动平台自主导航[J]. 北京航空航天大学学报, 2022, 48(11): 2138-2145.
XU G Y, XIONG Y W, ZHOU B, et al. Autonomous navigation based on PPO for mobile platform[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2138-2145(in Chinese).

下的自主导航。

以上研究虽然都实现了强化学习算法在自主导航中的应用,但依然存在着一个主要问题,即模型的输出动作离散问题,这将导致底盘的运动不流畅,使得部件损坏加速。为了解决上述问题,Tai 等[11]将改进型的深度确定性策略梯度算法(deep deterministic policy gradient, DDPG)应用于自主导航场景中,该方法使用激光雷达作为传感器,完成了端到端的强化学习模型训练,并实现了机器人的连续性控制。但输出动作连续性的解决将引发动作采样空间增大的问题,这意味着模型与环境交互的可能性增多,进而导致模型训练次数增加,收敛困难。

基于以上研究背景,本文提出了一种基于近似策略优化(proximal policy optimization,PPO)算法的移动平台自主导航方法,可有效解决强化学习算法的输出动作连续性问题及模型收敛困难问题。首先,利用全球导航卫星系统(global navigation satellite system, GNSS)和激光雷达传感器获取环境信息,并根据环境数据设计了一种基于改进人工势场算法的自身位置评价方法,作为状态输入。其次,在 PPO 算法的基础上,设计了基于正态分布的动作策略函数,移动平台的整车线速度和横摆角速度在正态分布中采样,通过该方法解决了强化学习模型输出动作连续性的问题;并根据导航场景特点,设计了强化学习模型的网络框架和奖励函数。然后,搭建了基于 Gazebo 模拟器的仿真环境对模型进行训练,结果表明,在经过自主学习后模型可有效收敛,且引入自身位置评价后的模型收敛速度明显增快。最后,将该模型移植入真实环境中进行实车实验,实验结果说明,移动平台在真实环境中可有效完成自主导航任务,验证了本文方法的有效性。

1　PPO 算法

1.1　强化学习算法

随着人工智能的兴起,强化学习算法也越来越多的应用在无人驾驶领域,其基本思想是:通过与环境交互来获取奖励,并以此进行自我学习。

强化学习算法包含 5 个主要部分:智能体、环境、状态、动作及奖励。智能体若在 t 时刻处于状态 S_t,可根据当前的策略函数 π 选择动作 A_t,该动作会影响环境,并在下一时刻返回奖励 R_{t+1},与环境交互的轨迹序列可表示为:$\{S_0,A_0,R_1,S_1,A_1,R_2,\cdots,S_t,A_t,R_{t+1}\}$。其中,智能体训练的目的是最大化获取的回合奖励,即期望回报 G_t。回报表示回合结束后的累计奖励加权值。G_t 的定义如下:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots + \gamma^{T-1} R_T = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{1}$$

式中:γ 为值小于 1 的折扣因子。

为了评价某一个状态或者动作的好坏,引入状态价值函数 $V_\pi(s)$ 和动作价值函数 $Q_\pi(s,a)$,方程分别如下:

$$V_\pi(s) = E_\pi[G_t \mid S_t = s] \tag{2}$$

$$Q_\pi(s,a) = E_\pi[G_t \mid S_t = s, A_t = a] \tag{3}$$

状态价值函数和动作价值函数反映了当前状态或动作可获取的回合平均期望回报值,因此可作为强化学习的决策指标。

1.2　基于策略的算法

现有许多强化学习算法都基于 Q 值,模型参数用于更新动作价值函数 $Q_\pi(s,a)$,策略即选择最大的 $Q_\pi(s,a)$。

不同于基于 Q 值的强化学习算法,PPO 算法将模型参数定义在策略函数 $\pi(a \mid s,\boldsymbol{\theta})$ 中:

$$\pi(a \mid s,\boldsymbol{\theta}) = \Pr\{A_t = a \mid S_t = s, \boldsymbol{\theta}_t = \boldsymbol{\theta}\} \tag{4}$$

式中:$\Pr\{A_t = a \mid S_t = s, \boldsymbol{\theta}_t = \boldsymbol{\theta}\}$ 表示在 t 时刻,状态 S_t 为 s、参数向量 $\boldsymbol{\theta}_t$ 为 $\boldsymbol{\theta}$ 时,该模型选取动作 A_t 为 a 的概率。

通过更新参数向量 $\boldsymbol{\theta}_t$,可得到更优的策略函数。参数向量更新方程如下:

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha \delta_t \frac{\nabla \pi(a \mid s,\boldsymbol{\theta}_t)}{\pi(a \mid s,\boldsymbol{\theta}_t)} \tag{5}$$

式中:α 为学习率;δ_t 为根据训练过程中求解出来的基准值。

参数向量更新的目的是使回报函数最大,回报函数定义如下:

$$\eta(\pi) = E_{s_0,a_0,\cdots}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t)\right] \tag{6}$$

式中:$r(s_t)$ 表示在 s_t 状态下的奖励值。

但在实际训练过程中,可能会由于学习率 α 设置的不合适导致策略模型表现越来越差。为解决此问题,PPO 算法定义了优势函数 $A_\pi(s_t,a_t)$[12]:

$$A_\pi(s_t,a_t) = Q_\pi(s,a) - V_\pi(s) \tag{7}$$

函数(7)表示动作 a 所获得的回合奖励与动作平均可获得的回合奖励差值。若 $A_\pi(s,a) > 0$,则说明动作 a 优于平均表现。

根据优势函数,构建了近似回报函数 $L_\pi(\tilde{\pi})$:

$$L_\pi(\tilde{\pi}) = \eta(\pi) + E_{s\sim\rho_{\theta_o},a\sim\pi_{\theta_o}}\left[\frac{\tilde{\pi}_{\boldsymbol{\theta}}(a \mid s)}{\pi_{\boldsymbol{\theta}}(a \mid s)} A_{\boldsymbol{\theta}_o}(s,a)\right] \tag{8}$$

式中:$L_{\pi}(\tilde{\pi})$表示更新后的策略 $\tilde{\pi}$ 所对应的近似回报函数;$\eta(\pi)$表示更新前策略 π 对应的回报函数;ρ_{θ_o}表示更新前策略参数为 $\boldsymbol{\theta}_o$ 所对应的状态分布。

若式(8)中的优势函数项 $A_{\theta_o}(s,a)\geqslant 0$,则策略更新后的回报函数单调不减,即模型策略更优或不变,通过该方法可提高训练效率。

2 导航状态定义

在自主导航场景中,为定义强化学习算法中的状态 S_t,需要获取自身位姿、速度和障碍物等多个数据。本文预设的场景为室外非结构化道路,因此,使用 GNSS 和激光雷达多传感器来获取环境信息。

2.1 环境感知方法

2.1.1 自身位置和速度信息感知

考虑到普通工业化的 GNSS 精度仅能达到米级别,无法满足小范围内的自主导航要求,因此,使用 RTK 技术对 GNSS 定位精度进行了提升,并将坐标从大地坐标系转换为站心坐标系[13-14],可获取坐标值(N,E,U),表示相对于站心点的东向、北向、天向距离。同时,通过在移动平台车头和车尾搭载 2 个 GNSS 天线,可根据其接收信号计算出移动平台底盘的航向角 φ。

除此之外,还可根据底盘控制器域网(controller area network, CAN)返回的消息得到移动平台的整车线速度 v 及整车横摆角速度 ω。

2.1.2 障碍物信息感知

在障碍物感知中,选用 VLP-16 线激光雷达作为障碍物感知传感器,但该激光雷达扫描出的点云数据量太大,无法直接输入给决策模块进行处理。因此,需要对激光雷达点云进行聚类分析。本文使用基于 K-D 树的欧氏距离聚类方法,通过对激光雷达点云进行聚类处理后即可得到障碍物相对自身的坐标信息(X_i,Y_i,Z_i)和长宽高信息(L_i,W_i,H_i)。

2.2 状态空间定义

强化学习是根据某一时刻状态 S_t 进行动作选择的,在本文的自主导航场景中,状态 S_t 的定义如图 1 所示。

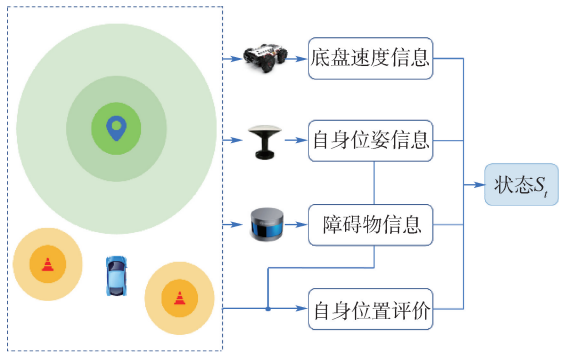


图 1 状态空间构造示意图

Fig. 1 Schematic diagram of state space construction

由图 1 可知,状态 S_t 中包含自身位姿信息(N,E,U)、航向角 φ、底盘速度信息(v,ω)及障碍物信息$(X_i,Y_i,Z_i,L_i,W_i,H_i)$。

除此之外,本文还设计了一种改进的人工势场算法,来评价自身完成任务的情况,并将其作为状态值。

2.2.1 人工势场算法

人工势场[15]作为移动平台的导航算法之一,其基本思想是:障碍物对自身产生排斥力,目的地对自身产生吸引力,势能为引力斥力之和,通过该方法在环境中构建人工势场。传统的引力势能和斥力势能如下:

$$U_{\mathrm{a}}=\frac{1}{2}k_p d_{\mathrm{g}} \tag{9}$$

$$U_{\mathrm{r}}=\begin{cases}\dfrac{1}{2}\varepsilon\left(\dfrac{1}{\tau}-\dfrac{1}{\tau_0}\right)^2 & \tau\leqslant\tau_0\\ 0 & \tau>\tau_0\end{cases} \tag{10}$$

式中:U_{a} 为引力势能;k_p 为引力因子;d_{g} 为自身与目的地的距离;U_{r} 为斥力势能;ε 为斥力因子;τ 为自身与最近障碍物间的距离;τ_0 为自身与障碍物间的距离阈值。

2.2.2 自身位置评价

本文设计了一种改进的人工势场算法作为移动平台自身位置评价,可有效评估移动平台完成任务的情况。

传统人工势场的斥力势能取最近障碍物与自身的距离 τ,但很多时候距离远的障碍物可能因为其体积大更有碰撞风险。因此,本文设计了一种危险因子用于判断障碍物的危险程度,危险因子定义如下:

$$t=d_{\mathrm{o}}-\max(L,W) \tag{11}$$

式中:d_{o} 为自身与障碍物的距离;L 和 W 分别为障碍物的长度和宽度。

计算各障碍物的危险因子并取其最小值 t^*,将斥力势能替换如下:

$$U'_{\mathrm{r}}=\begin{cases}\dfrac{1}{2}\varepsilon\left(\dfrac{1}{t^*}-\dfrac{1}{t_0}\right)^2 & t^*\leqslant t_0\\ 0 & t^*>t_0\end{cases} \tag{12}$$

式中:t_0 为危险阈值。

除此之外,由于本文中的目的地并非某个点,

而是一块区域,为更准确定义目的地区域的势能,将引力势能定义如下:

$$U'_{\mathrm{a}}=\begin{cases}\dfrac{1}{2}k_p d_{\mathrm{g}} & d_{\mathrm{g}}>d_{\mathrm{g}}^0\\ 0 & d_{\mathrm{g}}\leqslant d_{\mathrm{g}}^0\end{cases}\tag{13}$$

式中:d_{g}^0 为移动平台与目的地间的距离阈值。若移动平台与目的地距离小于阈值时,则引力势能为 0。

最终,取引力和斥力的势能和作为移动平台的自身位置评价指标:

$$U=U'_{\mathrm{r}}+U'_{\mathrm{a}}\tag{14}$$

U 值越大,说明完成任务情况越差,U 值越小,说明任务完成情况越好。

3　决策模型设计

3.1　动作策略函数

为解决在导航过程中移动平台输出动作的连续性问题,本文在 PPO 算法的基础上,根据式(4)的 PPO 动作策略函数,利用正态分布函数将策略函数重新定义如下:

$$\pi(x\mid s,\boldsymbol{\theta})=\frac{1}{\sigma(s,\boldsymbol{\theta})\sqrt{2\pi}}\exp\left[-\frac{(a-\mu(s,\boldsymbol{\theta}))^2}{2\sigma(s,\boldsymbol{\theta})^2}\right]\tag{15}$$

式中:$\sigma(s,\boldsymbol{\theta})$ 和 $\mu(s,\boldsymbol{\theta})$ 分布表示带有参数向量 $\boldsymbol{\theta}$ 的策略分布方差和期望。

基于以上动作策略函数定义,整车线速度 v 和横摆角速度 ω 将从正态分布中进行采样。相比于离散性采样,在正态分布中进行采样可以使输出动作数值连续,由此可解决输出动作连续性问题,且通过式(5)更新参数 $\boldsymbol{\theta}$,可改变正态分布的期望和方差,从而改变动作的采样概率,使好的动作采样概率更高,坏的动作采样概率更低,进而得到更优的决策模型。

但考虑到移动平台的整车线速度和横摆角速度有上限值,若不对正态分布期望进行限制,大多采样动作会超出其上限值,这意味着许多采样动作将对模型的更新起不到作用,进而导致模型收敛速度慢。因此,使用 tanh 激活函数对正态分布期望进行限制。由于 tanh 激活函数方程取值为(-1,1),将其乘以期望因子 δ_μ 可限制策略分布的期望,策略分布期望定义如下:

$$\mu(s,\boldsymbol{\theta})=\delta_\mu\tanh(x)\tag{16}$$

本文中的期望因子 δ_μ 取值为移动平台整车的线速度和横摆角速度上限值。

3.2　网络框架

根据上述策略定义,本文设计了决策模型的网络框架,如图 2 所示,包括 1 个输入层、演员网络和评论家网络。

输入层为状态空间序列(S_t, S_{t+1},…,S_T),分别输入进演员网络和评论家网络的隐藏层中。演员网络和评论家网络的隐藏层都为全连接层,且第 1 层都有 256 个节点,激活函数为 Relu;第 2 层都有 128 个节点,激活函数为 tanh。其中,演员网络用于选择策略分布 $\pi_i(a\mid s,\boldsymbol{\theta})$ 的期望 $\mu(s,\boldsymbol{\theta})$ 和方差 $\sigma_i(s,\boldsymbol{\theta})$,以构造整车线速度 v 的正态分布及整车横摆角速度 ω 的正态分布,决策动作将根据其分布函数进行采样;评论家网络用于估计评价函数 $V_\pi(s)$,根据式(7)和式(8)可知,该评价函数可计算出优势函数 $A_\pi(s,a)$,进而得到回报函数 L_π 的值,并参与演员网络策略分布的参数更新。

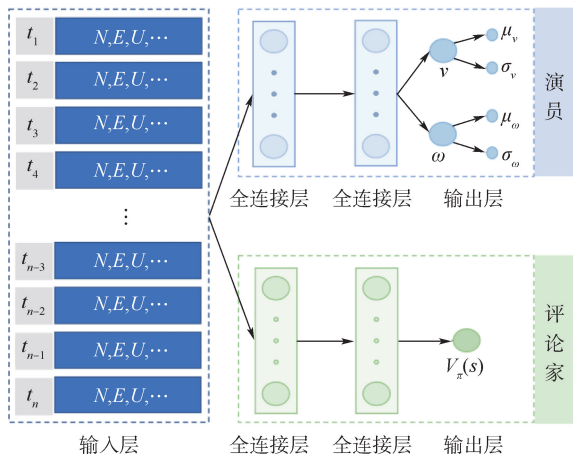


图 2　模型网络框架示意图

Fig. 2　Schematic diagram of model network framework

3.3　奖励函数

3.3.1　离散奖励

在本文的自主导航场景中,设置离散奖励如下:碰撞障碍物会得到负奖励,到达目的地会得到正奖励。除此之外,还定义了可行驶区域,超出该区域也会得到负奖励。具体的离散奖励值 r_e 设置如下:

$$r_e=\begin{cases}-10 & d_{\mathrm{a}}\geqslant d_{\mathrm{a}}^0\\ -10 & c=1\\ 20 & d_{\mathrm{g}}\leqslant d_{\mathrm{g}}^0\end{cases}\tag{17}$$

式中:d_{a} 为移动平台中心到目的地中心的距离;d_{a}^0 为可行驶区域阈值;c 为碰撞检测因子,若发生碰撞则为 1,不发生碰撞为 0。

3.3.2　基于自身位置评价的过程奖励

自主导航场景可采样的行为序列过多,可产生到达目的地的回合次数很少,因此只采用离散奖励训练效率较低。为提高模型收敛速度,本文

使用自身位置评价作为过程奖励对该模型进行引导。

引导方法为:在每一步计算自身位置评价奖励,而非仅在回合结束时计算离散奖励。根据改进的人工势场(式(14)),计算出当前时刻的人工势能,并取前后 2 步的人工势能差值 ΔU_t 作为奖励值,如下:

$$\Delta U_t = U_{t-1} - U_t \tag{18}$$

式中:U_{t-1} 和 U_t 分别为上一时刻和这一时刻移动平台的人工势能值。

考虑到移动平台有可能为赚取更多离散奖励值而不结束回合,因此,为了突出结果对移动平台的指导性,将过程奖励的值设置为比离散奖励小一个量级。上文中已对离散奖励设值,取值分别为 -10 和 20,因此将过程奖励归一化至(-1,1)之间。过程奖励具体的计算步骤如下:

步骤 1 初始化列表 $l_U=[0,1]$,计算在初始位置时移动平台的人工势能 U_0。

步骤 2 回合未结束时做如下循环:

步骤 2.1 移动平台采取动作 A_t,计算移动平台人工势能值 U_t。

步骤 2.2 计算移动平台前后 2 步的人工势能差值:$\Delta U_t = U_{t-1} - U_t$。

步骤 2.3 将移动平台人工势能值的绝对值加入列表中:$l_U = [0,1,|\Delta U_{t+1}|]$。

步骤 2.4 计算过程奖励值:$r_u = (\Delta U_{t+1} - \min(\text{list}))/(\max(\text{list}) - \min(\text{list}))$。

某个时刻的单步整体奖励值 r 为

$$r = r_e + r_u \tag{19}$$

4 仿真环境实验和实车实验

4.1 仿真环境实验

考虑到直接在真实环境中训练模型的成本太高,因此,搭建仿真环境对模型进行训练。本文的仿真环境基于 Gazebo 模拟器,并在该环境中搭建了移动平台物理模型和激光雷达传感器模型。模型搭建完毕后,将点云聚类算法应用在仿真环境中可得障碍物感知效果如图 3 所示。

在上述仿真环境的基础上,本文设定了 2 类训练场景,包含不同的障碍物分布和目的地区域,如图 4 所示。图中:红色物体表示障碍物,绿色区域表示目的地区域,紫色区域表示可行驶区域。

根据上述训练场景,设定以下训练策略:①训练回合步长设定为 160 步,每 32 步进行模型参数更新,若回合结束,则更新模型并初始环境,开始下一回合训练;②若移动平台发生碰撞、超出行驶范围及到达目的地,则返回奖励,更新模型参数并初始环境,开始下一回合训练。

为对比自身位置评价对导航模型的影响,本文分别对引入自身位置评价的模型和未引入自身位置评价的模型进行训练。其中,未引入自身位置评价的模型奖励值仅为离散奖励,训练结果如图 5 所示。

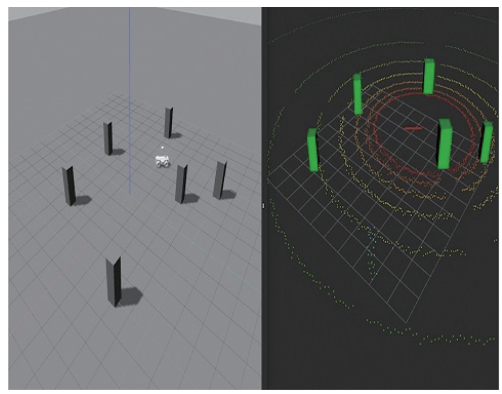


图 3 Gazebo 环境中的障碍物感知效果

Fig. 3 Obstacle detection in Gazebo environment

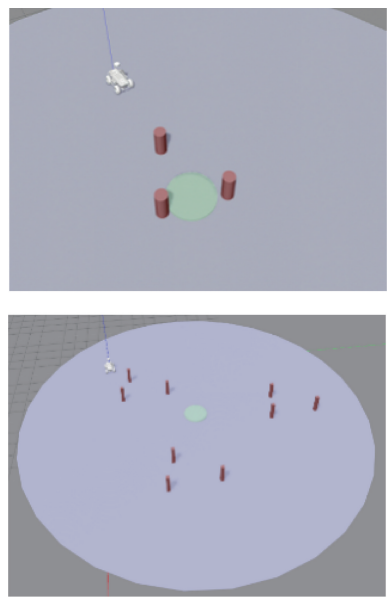


图 4 仿真环境训练场景

Fig. 4 Training scenario in simulation environment

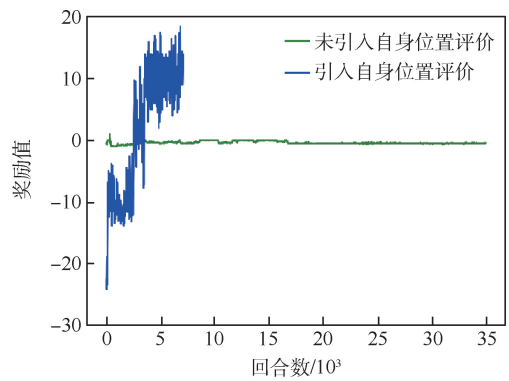


图 5 导航模型训练结果

Fig. 5 Training results of navigation model

由图 5 可以看出,未引入自身位置评价的模型训练 35 000 个回合依然无法收敛,而引入自身位置评价的模型在训练近 5 000 个回合时就可以达到收敛。训练结果表明,引入自身位置评价后模型收敛速度明显提高。

需要说明的是,由于引入自身位置评价,奖励会根据步数产生累计,在模型收敛后奖励依然波动较大,但整体不影响其完成到达目的地的任务。

模型经过自主学习后,可搜索到去往目的地的有效路径,并完成导航任务。其中,模型在训练场景中所习得路径如图 6 所示。

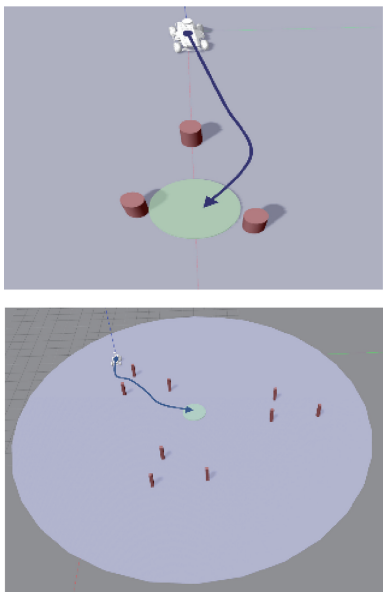


图 6　仿真环境中导航模型所习得路径

Fig. 6　Learned path of navigation model in simulation environment

4.2　实车实验

为验证所提出的自主导航模型在真实环境中是否有效,本文搭建了一套移动平台硬件如图 7 所示。基于以上硬件设施,完成了真实环境中的感知、决策和驱动 3 个功能的测试。



图 7　真实环境下搭建的移动平台

Fig. 7　Mobile platform built in real environment

在自身位姿感知测试中,在室外环境下使用 RTK 技术获取移动平台的经纬高度。随后,将其转换至站心坐标系下获得移动平台的自身位姿状态。

在障碍物感知测试中,使用 VLP-16 线激光雷达获取环境点云,并对其进行去除噪点处理和感知区域设置,最终使用聚类算法对点云数据进行聚类,获取障碍物的位置和大小数据。聚类效果以包围框方式显示,如图 8 所示。

将仿真环境中训练好的强化学习模型移植入真实平台中进行测试。测试环境中使用 3 个圆锥桶作为障碍物,并设定指定目的地进行移动平台自主导航的真实环境测试,测试场景布置如图 9 所示。

测试过程中,以工控机为计算硬件平台的条件下,从感知到驱动平均花费时长为 0.025 s,即平均每秒可以进行 40 次决策。

在测试场景中,本文提出的导航模型根据感知输入的障碍物信息、自身位置信息及人为指定的目的地信息,生成了如图 10 所示的导航轨迹曲线,证明了该模型可自主绕过障碍物并完成导航任务,验证了其有效性。

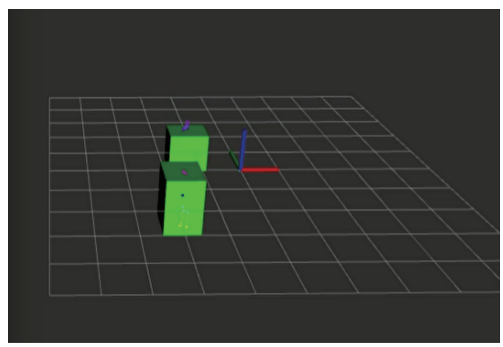


图 8　真实环境下激光雷达点云聚类效果

Fig. 8　Clustering of LIDAR point cloud in real environment



图 9　真实环境下自主导航场景

Fig. 9　Autonomous navigation scene in real environment

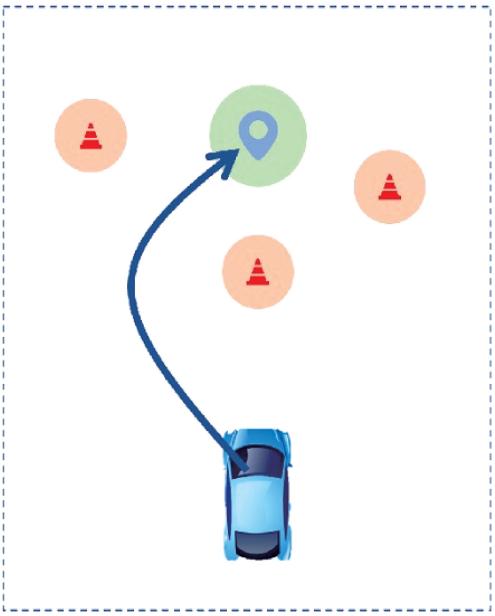


图 10 真实环境下生成的导航轨迹曲线

Fig. 10 Navigation trajectory curve in real environment

5 结 论

针对移动平台的自主导航场景,本文提出了一种基于 PPO 算法的导航模型,并完成了模型训练和实景测试,主要结论如下:

1) 在 PPO 算法的基础上,设计了基于正态分布的动作策略函数,移动平台整车线速度和横摆角速度在策略函数中采样,该方法可解决强化学习模型输出动作连续性问题。

2) 设计了基于改进人工势场的自身位置评价方法,并将其作为过程奖励函数加入模型训练,该方法可大幅度加快导航模型的训练速度。训练结果表明,引入自身位置评价指标的模型训练 5 000次即可收敛,未引入自身位置评价指标的模型训练 35 000 次以上还未完成收敛。

3) 在仿真环境和真实环境中均完成了模型测试,测试结果说明,该模型可有效绕过障碍物并到达目的地,验证了其有效性。

参考文献 (References)

[1] 王义林. 地面无人平台自主导航避障系统的研究与实现[D]. 哈尔滨:哈尔滨工业大学, 2020.
WANG Y L. Research and implementation of autonomous navigation and obstacle avoidance system for ground unmanned platform[D]. Harbin:Harbin Institute of Technology,2020(in Chinese).

[2] 秦圣然. 基于激光传感器的移动机器人导航系统研究[D]. 沈阳:沈阳工业大学,2020.
QIN S R. Research on laser sensor-based navigation system for mobile robots[D]. Shenyang:Shenyang University of Technology,2020(in Chinese).

[3] HART P E, NILSSON N J, RAPHAEL B. A formal basis for the heuristic determination of minimum cost paths in graphs[J]. IEEE Transactions on Systems Science and Cybernetics,1968,4(2):100-107.

[4] STENT A. Optimal and efficient path planning for partially-known environments[C]// Proceedings of IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press,1994,4:3310-3317.

[5] STENT A. The focussed D* algorithm for real-time replanning[C]// Proceedings of the 14th International Joint Conference on Artificial Intelligence. New York:ACM,1995:1652-1659.

[6] LAVALLE S M, KUFFNER J J. Randomized kinodynamic planning[C]// Proceedings of IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 1999, 1: 473-479.

[7] 付雪建. 基于强化学习的移动机器人自主导航研究[D]. 重庆:重庆大学,2017.
FU X J. Research on autonomous navigation of mobile robots based on reinforcement learning[D]. Chongqing: Chongqing University,2017(in Chinese).

[8] 杨宁博. 面向环境探测的全向模式移动机器人自主导航研究[D]. 哈尔滨:哈尔滨工业大学,2019.
YANG N B. Research on autonomous navigation of omnidirectional mode mobile robots for environmental detection[D]. Harbin:Harbin Institute of Technology,2019(in Chinese).

[9] 陶睿. 基于深度强化学习的移动机器人导航[D]. 济南:山东大学,2020.
TAO R. Deep reinforcement learning-based navigation for mobile robots[D]. Jinan:Shandong University,2020(in Chinese).

[10] 何聪. 基于深度强化学习的机器人视觉导航算法[D]. 南京:东南大学,2021.
HE C. Robot visual navigation algorithm based on deep reinforcement learning[D]. Nanjing:Southeast University,2021(in Chinese).

[11] TAI L, PAOLO G, LIU M. Virtual-to-real deep reinforcement learning:Continuous control of mobile robots for mapless navigation[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press,2017: 31-36.

[12] SCHULMAN J, WOLSKI F, DHARIWAL P, et al. Proximal policy optimization algorithms[EB/OL]. (2017-08-28)[2021-03-01]. https://arxiv.org/abs/1707.06347.

[13] 刘志平,余前勇,查剑锋. 空间直角坐标至两类常用坐标的快速变换[J]. 测绘科学,2015,40(3):8-11.
LIU Z P, YU Q Y, ZHA J F. Fast coordinate transformations for both XYZ-BLH and XYZ-RhA[J]. Science of Surveying and Mapping, 2015,40(3):8-11(in Chinese).

[14] 刘山洪,邓彩群. 坐标转换与坐标变换研究[J]. 吉林建筑大学学报,2016,33(1):43-47.
LIU S H, DENG C Q. Transformation of coordinate system[J]. Journal of Jilin Jianzhu University,2016,33(1):43-47(in Chinese).

[15] KHATIB O. Real-time obstacle avoidance system for manipulators and mobile robots[J]. International Journal of Robotics Research,1986,5(1):90-98.

Autonomous navigation based on PPO for mobile platform

XU Guoyan*, XIONG Yiwei, ZHOU Bin, CHEN Guanhong

(Key Laboratory of Autonomous Transportation Technology for Special Vehicles, Ministry of Industry and Information Technology, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: This paper presents an autonomous navigation method based on proximal policy optimization (PPO) algorithm for mobile platform. In this method, GNSS and LADAR are used for sensing environment information. To define the state of reinforcement learning model, an ego position evaluation method is introduced based on improved artificial potential field algorithm. After that, on the basis of PPO algorithm, a kind of action policy function is designed based on Gaussian distribution, which solves the continuity problem of the vehicle linear velocity and yaw velocity. Furthermore, the network framework and reward function of the model are also designed for navigation scenarios. In order to train the navigation model, a virtual environment based on Gazebo is built. The training results show that the ego position evaluation method obviously helps to improve the speed of model convergence. Finally, the navigation model is transplanted to a real environment, which verifies the effectiveness of the proposed method.

Keywords: proximal policy optimization algorithm; mobile platform; autonomous navigation; reinforcement learning; artificial potential field

Received: 2021-03-02; **Accepted**: 2021-04-11; **Published online**: 2021-05-07 08:55

URL: kns.cnki.net/kcms/detail/11.2625.V.20210506.1639.002.html

Foundation item: National Natural Science Foundation of China (51775016)

* **Corresponding author**. E-mail: xuguoyan@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0105

NO 快速检测光腔结构设计与光路仿真

李文*, 蔡永青, 陈梦凡, 刘鹏
(北方工业大学 机械与材料工程学院, 北京 100144)

摘 要: 根据 NO 与臭氧的化学发光反应机理,提出了一种基于化学发光反应检测原理的用于 NO 气体快速检测的新型圆柱全反射 S 型光腔结构。该结构以圆柱内壁全反射 S 型结构为光腔模型,采用圆柱体光源作为光散射元件,将化学发光光线最大效率地汇聚到探测器的感光面,增强了 NO 浓度测量信号,提高了检测精度。采用光学软件 ZEMAX 对所提模型进行光路仿真,通过对比分析表明,S 型光腔采集光路对化学发光的采集效率为 36.6%。通过实验验证,结果表明,在一定浓度范围内 NO 气体浓度与化学反应发光强度线性相关度为 0.999 2,在量程 0‰~2.5‰的范围内,检出限为 0.001‰。所提模型结构简单,满足国家标准,为尾气在线检测提供了一种实用的设计方案。

关 键 词: S 型光腔结构; 光路仿真; NO 浓度; 化学发光; ZEMAX
中图分类号: O433.1
文献标志码: A **文章编号**: 1001-5965(2022)11-2146-07

在机动车尾气排放中占比最大的是 NO 气体,因此,检测机动车尾气中 NO 浓度成为重中之重[1]。近年来,发展了如非分散红外光谱法(non-dispersive infrared, NDIR)、可调谐吸收光谱法(tunable diode laser absorption spectroscopy, TDLAS)等[2-4]一系列光谱法直接检测 NO 浓度。目前,中国主要采用电化学原理和分光光度法检测机动车尾气中的 NO。电化学检测存在受温、湿度等干扰因素影响大,标定时间长等问题,已无法满足国标 GB18352.6—2016[5]中的尾气排放相关标准,并要求在用车检测 NO_X 方法中停止使用电化学法[6]。分光光度法检测由于分辨率低、灵敏度差,无法满足相关规定。化学发光法(chemiluminescent detector,CLD)测定 NO 具有灵敏度高、线性测量量程大等优点,被广泛应用于气体的检测分析中[7]。

为提高 NO 检测精度和响应时间,基于化学发光反应检测原理,本文设计了一种新型圆柱全反射 S 型光腔结构,采用 ZEMAX 对结构模型进行光路仿真,应用 FLUENT 对 S 型光腔结构模型进行模拟验证。对比分析表明,圆柱内壁全反射 S 型光腔结构在气体检测领域具有广泛的应用前景。

1 原理分析

化学发光法检测 NO_X 的工作原理是臭氧与 NO 反应[8],生成激发态的 NO_2^*,在激发态的 NO_2^* 返回基态 NO_2 时,发射出波长为 600~2 500 nm 的光,发射出的光峰值约在1 200 nm 处。反应方程式如下:

$$NO + O_3 \rightarrow NO_2^* + O_2 \tag{1}$$

$$NO_2^* \rightarrow NO_2 + hv \tag{2}$$

式中:h 为普朗克常量;v 为发射光频率。

化学反应的发光强度 I 的计算如下:

$$I = Nhv/(At) \tag{3}$$

收稿日期: 2021-03-03; **录用日期**: 2021-06-06; **网络出版时间**: 2021-07-30 11:10
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210730.1047.002.html
基金项目: 国家自然科学基金 (51205005); 北京市科技创新服务能力建设资助项目 (PXM2017-014212-000013)
* **通信作者**. E-mail: lw@ncut.edu.cn

引用格式: 李文, 蔡永青, 陈梦凡, 等. NO 快速检测光腔结构设计与光路仿真[J]. 北京航空航天大学学报, 2022, 48(11): 2146-2152. LI W, CAI Y Q, CHEN M F, et al. Optical path simulation and design of NO rapid detection optical cavity structure[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2146-2152(in Chinese).

式中:A 为照射区域面积;N 为时间间隔 t 上的光子总数。

由式(3)可知,光学中的光强与众多因素有关,因此需要对以上因素进行分析。

2　微型采集光路的设计与优化

化学发光法检测 NO 的过程中[8-10],激发态的 NO_2 跃迁回基态 NO_2 时,会自发辐射产生波峰为 1 200 nm 的光,产生的光散射方向为发散性的,因此,提高 NO 气体检测精度的关键是如何收集更多的光[11-13],以此提高光电二极管采集信号的强度,降低背景噪声,同时缩短整体采集光路,有利于分析仪小型化发展趋势[14]。

2.1　光子检测装置

目前,美国 GE 公司研制的氮氧化物监测仪化学发光采集装置采用的是光电倍增管(PMT)[15-16]。光电倍增管响应时间快,噪声低,但响应波段在 300 ~ 500 nm 范围内具有较高的光谱响应,且对紫外光具有较高的探测性能,不适合中心波长为 1 200 nm 的氮氧化物化学发光检测,而且需要 400 ~ 600 V 的驱动电压,体积很大,不利于监测仪小型化改进。因此,本文选取体积较小、暗电流较低的 PIN 光电二极管作为化学发光探测器,对信号进行高增益放大,以实现弱光信号的检测[17]。

经过分析,最终选定滨松公司的 S12023 系列 Si-APD 参数光敏面为 2.5 mm × 2.5 mm 的光电探测器。

2.2　光腔结构的设计与比较

由于传统的化学发光法检测 NO 产生的化学发光光线发散性大、随机性强,化学发光采集效率低、光路长。因此,本文设计了一种内壁反射式光腔结构模型,能够改善化学发光光线发散角大等问题,汇聚更多的光线,以此增强检测信号。为了仿真更接近真实情况,采用 ZEMAX 非序列体光源仿真软件[18-20],对圆柱广口散射式、12 mm 透镜聚焦式、圆柱全反射式模型进行仿真与分析。通过仿真软件模拟探测器接收到光子的相对能量 a。根据不同的光腔模型,寻找最大的光子相对能量 a[21-24]。为方便计算,设定单位为 mm^3,体积内的光学强度为 1。

由 ZEMAX 非序列模块可知,整个光腔结构的光源是无规则的体光源[25],因此,需要在结构中增加光学镜面以此来得到更多的光子。下面对 3 种光腔模型进行比较。

1) 圆柱广口散射式模型

对于广口的光场,模型中选择光场为广口区域直径为 6 mm、高度为 8 mm 的圆柱形状。直径为 6 mm 圆柱形状是结合边长为 2.5 mm 的探测器而确定。为了更好地提高光子照射到探测器表面的效率,在发光体的顶端正对探测器放置一个直径为 5 mm 的光学反射镜,如图 1 所示。为了更真实地进行仿真,本文选取化学发光模型为圆柱体光源[26],得到如图 2 所示的光学简化仿真模型。经光线追迹得到的结果如图 3 所示。图中:尺寸为宽 3 mm,高 3 mm,像素 1 024 × 1 024,总点数 194 312,峰值辐照度 6.849 5 W/cm^2,总功率 3.5×10^{-2} W。分析光线条数为 500 000,由图 3 可知总撞击数为 194 312,光子的相对能量 a 的计算为

$$a = M/T \tag{4}$$

式中:M 为总撞击数;T 为分析光线条数。

通过计算式(4)可得,探测器接收到光子的相对能量 a 为 0.39;若不加光学反射镜,探测器接收到的光子相对能量为 0.37。

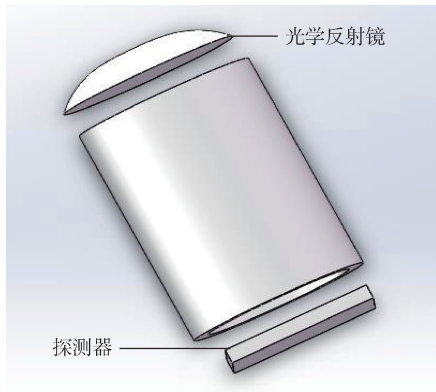
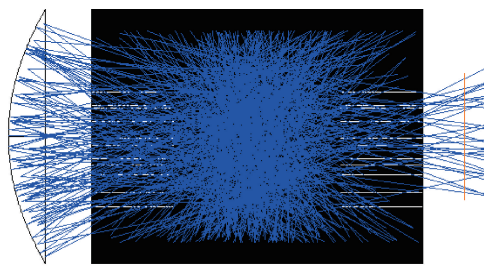
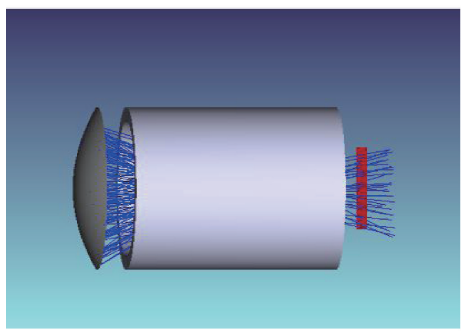


图 1　广口圆柱模型

Fig. 1　Cylindrical wide-mouth model



(a) 非序列模式3D模型



(b) 非序列模式阴影模型

图 2　光学简化仿真模型

Fig. 2　Optical simplified simulation model

2）12 mm 透镜聚焦式模型

对于选择透镜光学结构，市面上大多以 12 mm 的透镜为主，因此本文模型选择 12 mm 的透镜，将透镜放置在探测器上方，同时发光体顶端仍放有直径为 5 mm 的光学反射镜，如图 4 所示，图示中透镜的作用是使光线汇聚在一起[27]。图中尺寸为宽 3 mm，高 3 mm，像素 1 024 × 1 024，总点改为 184 811，峰值辐照度 7.851 5 W/cm^2，总功率 3.962 × 10^{-2} W。通过图 5 可知，总撞击数为 184 811，因此由式(4)计算可得 a 为 0.36，在不加反射镜的情况下 a 为 0.26。

根据探测器接收到的对光子能量比较可知，12 mm 透镜聚焦式模型不能起到汇聚光线的效果。

3）圆柱全反射式模型

先选择简单的圆柱体模型进行分析，将圆柱体上方放置反射镜的方法更换成圆柱筒内壁粘贴反射镜。模型中圆柱筒的内表面相当于全反射面，如图 6 所示，圆柱筒的直径不需要太大。

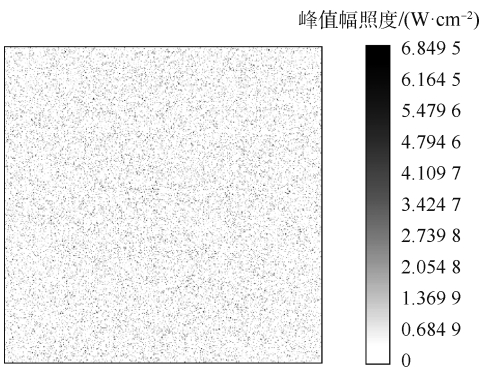
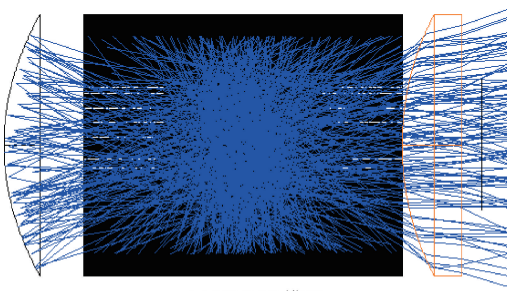
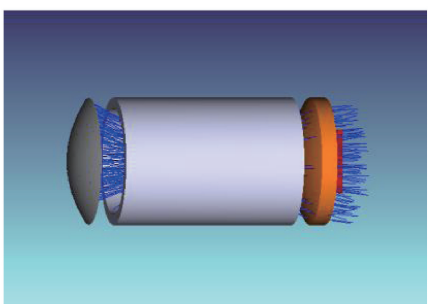


图 3 矩形探测器成像图

Fig. 3 Rectangular detector imaging



(a) NSC 3D模型



(b) NSC 阴影模型

图 4 透镜聚焦式模型

Fig. 4 Lens focusing model

以圆筒长度为 8 mm 为例，圆筒内壁为全反射的条件下，如图 7 可知，总撞击数为 738 707，因此探测器接收到的相对光子能量 a 为 1.48。图中：尺寸为宽 3 mm，高 3 mm，像素 1 024 × 1 024，

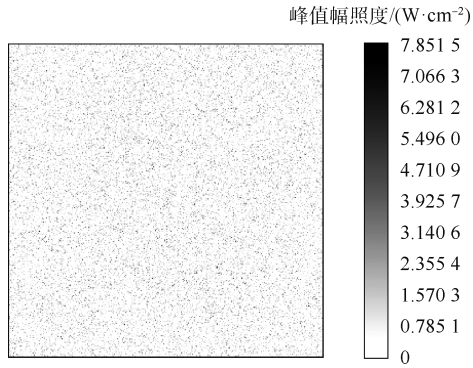
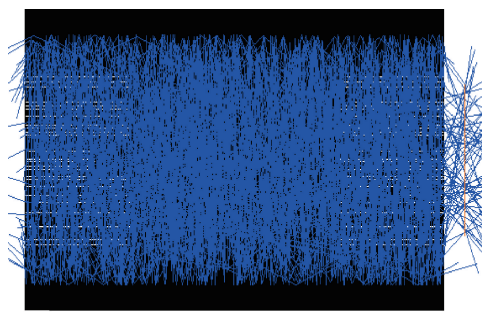
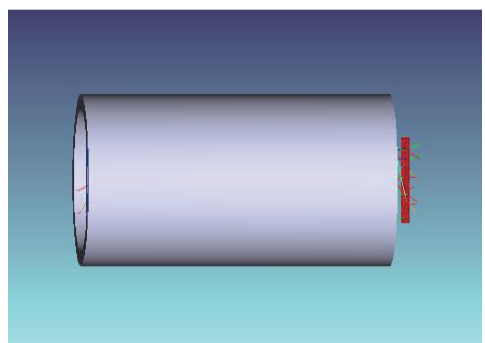


图 5 加透镜的矩形探测器成像图

Fig. 5 Rectangular detector imaging with Lens



(a) 内壁全反射NSC 3D模型



(b) 内壁全反射NSC阴影模型

图 6 圆柱全反射式模型

Fig. 6 Cylindrical total reflection model

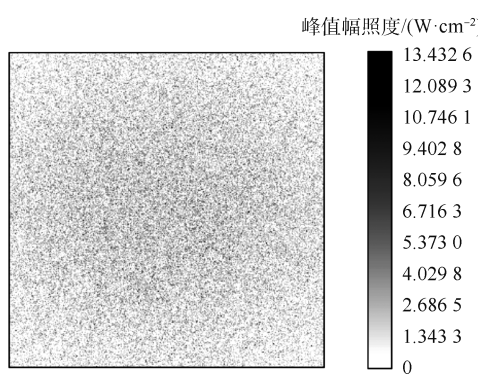


图 7 内壁全反射的矩形探测器成像图

Fig. 7 Rectangular detector imaging with total reflection of inner wall

总点数 738 707,峰值辐照度 3.433 W/cm^2,总功率 1.477 4 ×10^{-1} W。

可以看出,提高腔体内表面反射率是增加探测器信号强度的关键。由 3 种腔体模型接收相对光子能量的多少可以判定圆柱全反射式模型更有助于光子的高效收集。因此,本文设计了内壁全反射式光腔结构。

2.3　光腔结构设计的优化与仿真

化学发光法检测 NO 选取合适的光腔结构极其重要,这直接关系到 NO 与臭氧在腔体内的反应效率。因此,光腔的设计需要考虑以下条件:快速稳定的气流交换,气体反应混合充分,腔体小型化,尽可能多地将化学发光光线汇聚到探测器的感光面。在 2.2 节方案的基础上进一步优化,将选定的圆柱全反射式模型构成直径为10 mm的 S 型光腔结构模型,如图 8 所示,采用 ZEMAX 非序列模型仿真,由图 9 可知总撞击数为 1 365 467,因此探测器接收到的相对光子能量 a 较大为 2.73。图中:尺寸为宽 3 mm,高 3 mm,像素 1 024 ×1 024,总点数 1 365 467,峰值辐照度 17.784 W/cm^2,总功率 2.730 9 ×10^{-1} W。

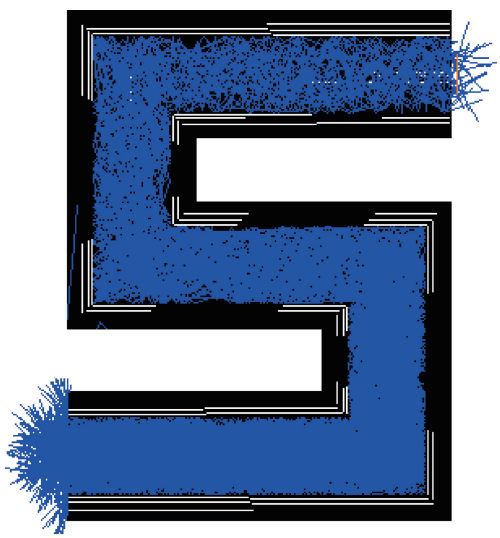


图 8　S 型光腔结构 3D 模型

Fig. 8　3D model of S-shaped optical cavity structure

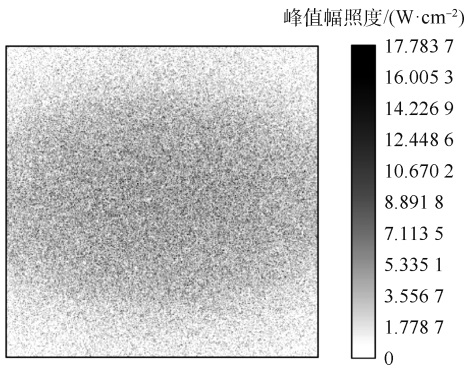


图 9　S 型光腔结构成像图

Fig. 9　S-shaped optical cavity structure imaging

为了充分证明圆柱内壁全反射 S 型光腔结构更好,本文选取了不同截面形状的全反射模型和不同形状的光腔全反射结构模型与设计的 S 型光腔结构进行对比,采用 ZEMAX 进行仿真,仿真结果表明,内壁全反射 S 型光腔结构探测器接收到的相对光子能量最大。综上所述,最终选择圆柱内壁全反射 S 型结构作为光腔模型。

通过选定的模型,利用 ZEMAX 仿真光场的模型,得到如图 10 所示的光场模型与仿真结果。

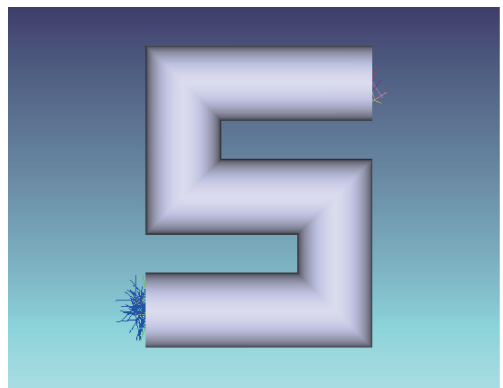


图 10　最终仿真结果

Fig. 10　Final simulation result

3　实验与分析

为了满足国家标准对腔体性能参数的要求,需要搭建标准的气体检测平台对选定的圆柱内壁全反射 S 型光腔内置反射镜模型进行性能测试。

3.1　化学发光检测设备

根据 NO 和臭氧化学反应发光波长,选定的光电探测器对光的敏感范围为 400 ~ 1 000 nm。同时设计了 APD 器件的恒温制冷功能,以此维持 APD 的增益稳定及压低器件的噪声水平。

3.2　标准气体检测的平台搭建

设计的双通道氮氧化物分析气路平台由气体预处理部分、气体前期反应制备装置、化学发光反应及其检测部分、流路控制部分等 4 部分组成,搭建流程如图 11 所示。

样气通过气泵抽取气体进入到仪器,采用设计的双通道方式测定氮氧化物浓度,部分样气进入钼转化炉将 NO_2 转化为 NO,并单独进入一个反应室,测出总的氮氧化物浓度,即 NO 和 NO_2 浓度之和。剩下的样气直接进入另一个反应室,测出其中 NO 浓度。气体混合反应均在设计的新型圆柱全反射 S 型反应室中进行。

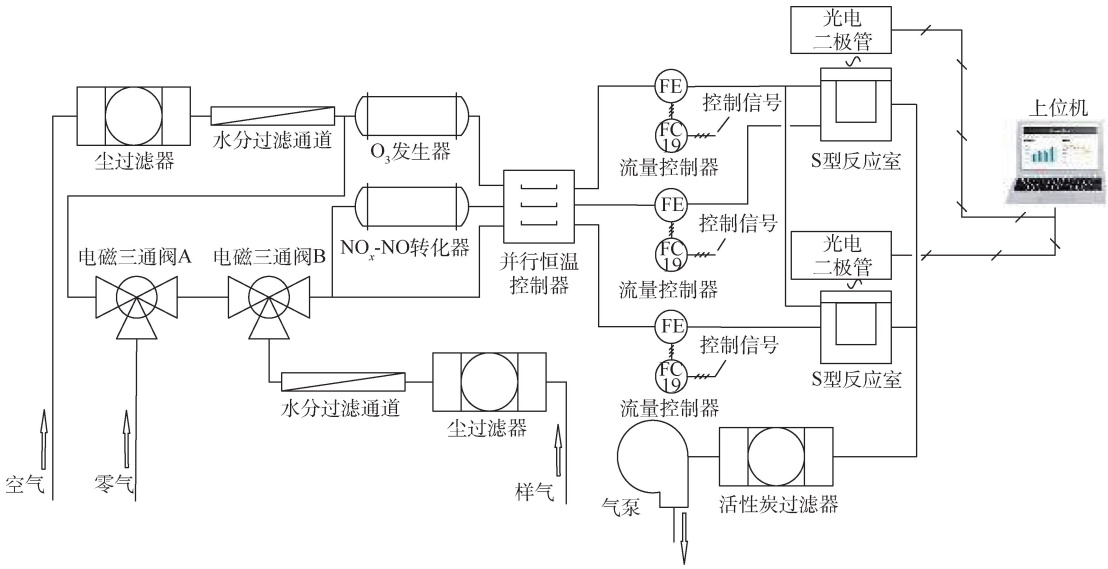


图 11 双通道氮氧化物分析气路流程

Fig. 11 Flow chart of two-channel NO_x analysis gas path

3.3 腔体实验性能

3.3.1 建立 NO 标准工作曲线

将仪器预热 15 min,达到最佳稳定状态。分别通入 0.30‰、0.50‰、0.80‰、1.0‰、1.5‰、1.8‰、2.0‰的标准气体,对其进行检测,每个标准气体测定 6 次,以室内空气的测定作为空白参照,扣除空白参照的检测电压值后,以 NO 浓度为横坐标,检测电压值为纵坐标,对得到的数据进行曲线拟合,工作曲线如图 12 所示,工作曲线对应的相关参数如表 1 所示。

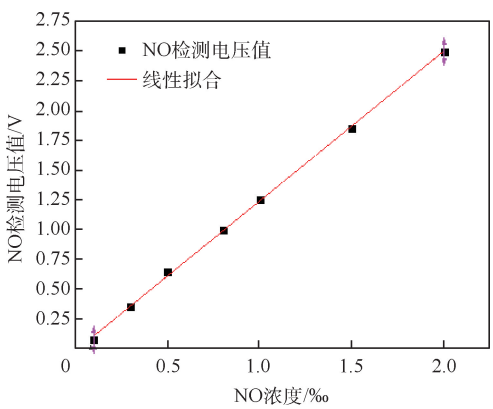


图 12 NO 检测工作曲线

Fig. 12 Working curve of nitric oxide detection

表 1 工作曲线相关参数

Table 1 Work curve related parameters

参数	方法或数值
检测方法	化学发光法
线性范围/‰	0 ~ 2.5
回归方程	$y = 0.001\ 26x - 0.019\ 78$
相关系数	0.999 2
最低检出限/‰	0.001

3.3.2 重复性测试

将空气作为零点标定,0.8‰的标准气体作为量程标定。通入 1.0‰的样气进行 6 次测定,作为腔体的重复性测试,测试结果如表 2 所示。

表 2 重复性测试结果

Table 2 Repeatability test results

测试序号	NO 浓度/‰
1	1.010
2	1.007
3	0.985
4	1.021
5	1.009
6	1.005
平均值/‰	1.006
标准偏差/%	1.18

4 结 论

本文通过分析 NO 与臭氧的化学反应产生的化学发光信号特点,采用 ZEMAX 光学仿真对其微型采集光路进行仿真,选取了最佳的光腔结构进行光子检测,S 型光腔结构模型追踪到的总撞击数为 1 365 467,与分析光线条数 500 000 进行对比,结果表明优化后的 S 型采集光路对化学发光的采集效率为 36.6%。通过实验对圆柱内壁全反射 S 型光腔结构进行测试,经过实验验证,低浓度的 NO 气体,其浓度与发光强度线性相关。在系统 0‰ ~ 2.5‰的量程范围内,系统检测线性度为 0.999 2,标准偏差为 1.18%。由此表明圆柱内壁全反射 S 型光腔结构稳定性更好,同时也实现了检测 NO 浓度的要求。综上所述,该结构具有工艺简单、光路简洁、可小型化等特点,圆柱

内壁全反射 S 型光腔结构在气体检测领域具有广泛的应用前景。

致谢　感谢徐明刚老师给予的指导和帮助,同时也感谢金工实习老师给予的技术支持。

参考文献(References)

[1] 刘萍,简家文,陈志芸. 基于集成神经网络的汽车尾气检测系统设计[J]. 环境工程学报,2016,10(4):1883-1887.
LIU P,JIAN J W,CHEN Z Y. Design of automobile exhaust gas detection system based on integrated neural network[J]. Journal of Environmental Engineering,2016,10(4):1883-1887 (in Chinese).

[2] 李世武,王琳虹,郭栋,等. 基于机动车排放的自适应信号控制模型[J]. 华南理工大学学报(自然科学版),2011,39(3):101-106.
LI S W,WANG L H,GUO D,et al. Adaptive signal control model based on vehicle emission[J]. Journal of South China University of Technology (Natural Science Edition),2011,39(3):101-106(in Chinese).

[3] 罗学科,刘鹏,李文,等. 基于 CLD 机动车尾气 NO_x 排放快速在线检测研究[J]. 环境科学与技术,2020,43(9):119-124.
LUO X K,LIU P,LI W,et al. Research on rapid on-line detection of NO_x emission from motor vehicle exhaust based on CLD[J]. Environmental Science and Technology,2020,43(9):119-124(in Chinese).

[4] 马清芝,车胜新. 我国汽车污染物排放标准制、修订发展一瞥[J]. 机械工业标准化与质量,2006(3):14-16.
MA Q Z,CHE S X. A glance at the development and revision of my country's automobile pollutant emission standards[J]. Standardization and Quality of Machinery Industry,2006(3):14-16(in Chinese).

[5] 环境保护部,国家质量监督检验检疫总局. 轻型汽车污染物排放限值及测量方法(中国第六阶段):GB18352.6—2016[S]. 北京:中国环境科学出版社,2020.
Ministry of Environmental Protection,Gneral Administration of Quality Supervision,Inspection and Quarantine. Limits and measurement methods for emissions from light-duty vehicles (CHINA 6):GB18352.6—2016[S]. Beijing:China Environmental Science Press,2020(in Chinese).

[6] 吴盛阳,胡仁志,谢品华,等. 基于腔衰荡光谱技术(CRDS)对大气总活性氮氧化物(NO_y)的实时测量[J]. 光谱学与光谱分析,2020,40(6):1661-1667.
WU S Y,HU R Z,XIE P H,et al. Real-time measurement of atmospheric total active nitrogen oxides (NO_y) based on cavity ring-down spectroscopy (CRDS)[J]. Spectroscopy and Spectral Analysis,2020,40(6):1661-1667(in Chinese).

[7] 李杭. 微型 NO 化学发光分析仪的研制与应用[D]. 合肥:中国科学技术大学,2019.
LI H. Development and application of miniature NO chemiluminescence analyzer[D]. Hefei:University of Science and Technology of China,2019(in Chinese).

[8] 孟晓辰,郝群,朱秋东,等. 基于 Zemax 的部分补偿透镜的优化设计[J]. 光学学报,2011,31(6):222-228.
MENG X C,HAO Q,ZHU X D,et al. Optimal design of partial compensation lens based on Zemax[J]. Acta Optica Sinica,2011,31(6):222-228(in Chinese).

[9] 苏杭,王桂梅,杨立洁. 大气痕量硫荧光采集光路设计与仿真[J]. 激光与光电子学进展,2020,57(5):144-149.
SU H,WANG G M,YANG L J. Design and simulation of light path for chemiluminescence collection of atmospheric trace sulfur[J]. Progress in Laser and Optoelectronics,2020,57(5):144-149(in Chinese).

[10] 李俊纬,倪屹,郭瑜,等. 基于凹面光栅的特定蛋白分析仪结构设计及性能分析[J]. 光学学报,2019,39(8):380-388.
LI J W,NI Y,GUO Y,et al. Structure design and performance analysis of specific protein analyzer based on concave grating[J]. Acta Optica Sinica,2019,39(8):380-388(in Chinese).

[11] 扶昭富,李攻科,胡玉斐. 催化化学发光分析系统研究进展[J]. 光谱学与光谱分析,2015,35(9):2450-2456.
FU Z F,LI G K,HU Y F. Research progress of catalytic chemiluminescence analysis system[J]. Spectroscopy and Spectral Analysis,2015,35(9):2450-2456(in Chinese).

[12] 金星焕,陈和,董家宁,等. 凹面光栅拉曼光谱仪的光学设计[J]. 光学学报,2015,35(9):359-365.
JIN X H,CHEN H,DONG J N,et al. Optical design of concave grating Raman spectrometer[J]. Acta Optica Sinica,2015,35(9):359-365(in Chinese).

[13] 国爱燕,白廷柱,胡海鹤,等. 固体火箭发动机羽烟紫外辐射特性分析[J]. 光学学报,2012,32(10):176-183.
GUO A Y,BAI T Z,HU H H,et al. Analysis of ultraviolet radiation characteristics of solid rocket motor plume[J]. Acta Optica Sinica,2012,32(10):176-183(in Chinese).

[14] 杨兰,蔡晓梅,周雄图,等. 圆孔液晶透镜的 ZEMAX 设计与优化[J]. 发光学报,2017,38(12):1688-1694.
YANG L,CAI X M,ZHOU X T,et al. ZEMAX design and optimization of round hole liquid crystal lens[J]. Chinese Journal of Luminescence,2017,38(12):1688-1694(in Chinese).

[15] 李杭,刘文清,姚路. 基于 PIN 光电二极管的微型一氧化氮化学发光传感器[J]. 光学学报,2019,39(5):291-295.
LI H,LIU W Q,YAO L. Miniature nitric oxide chemiluminescence sensor based on PIN photodiode[J]. Acta Optica Sinica,2019,39(5):291-295(in Chinese).

[16] 陈洪芳,汤亮,石照耀,等. 基于双波长法补偿空气折射率的激光追踪系统 ZEMAX 仿真方法[J]. 中国激光,2019,46(1):232-239.
CHEN H F,TANG L,SHI Z Y,et al. ZEMAX simulation method of laser tracking system based on dual-wavelength method to compensate air refractive index of air[J]. Chinese Journal of Lasers,2019,46(1):232-239(in Chinese).

[17] 陈洪芳,汤亮,孙衍强,等. 基于 Zemax 仿真的激光追踪测量光学系统能量分析[J]. 中国激光,2018,45(7):183-190.
CHEN H F,TANG L,SUN Y Q,et al. Energy analysis of laser tracking measurement optical system based on Zemax simulation[J]. Chinese Journal of Lasers,2018,45(7):183-190(in Chinese).

[18] HAN X Y,HUAN K W,SHENG S J. Performance evaluation and optimization design of photoelectric pyrometer detection op-

tical system[J]. Defence Technology,2020,16(2):401-407.
[19] GAN L,ZHANG H,LIU B,et al. Influence of high overload on the collimating lenses of laser ranging systems[J]. Defence Technology,2020,16(2):354-361.
[20] WU Q Y,TANG Y H,CHEN X Y,et al. Method for evaluating ophthalmic lens based on eye-lens-object optical system[J]. Optics Express,2019,27(26):37274-37285.
[21] HUANG P,HE C W,FAN B,et al. Analysis and optimization of the Glass pattern realized by a microlens array[J]. Applied Optics,2019,58(36):9976-9982.
[22] HANG T Y,WANG X L,LIU F,et al. Zemax-based optimum structural design of probe of an optical-fiber sensor[J]. Transactions of Nanjing University of Aeronautics and Astronautics,2018,35(2):309-317.
[23] LIU P,CHEN B,ZHANG Y C,et al. Simulation method of Wolter type grazing incidence telescope imaging of an X-ray region[J]. Research in Astronomy and Astrophysics,2019,19(12):13-20.
[24] LEE J,LEE J,WON Y H. Nonmechanical three-dimensional beam steering using electrowetting-based liquid lens and liquid prism[J]. Optics Express,2019,27(25):36757-36766.
[25] ZHANG J K,WANG X,CHEN X B,et al. Paraxial lens design of double-telecentric anamorphic zoom lenses with variable magnifications or fixed conjugate[J]. Journal of the Optical Society of America A-Optics,Image Science,and Vision,2019,36(12):1977-1990.
[26] BISESTO F G,CASTELLANO M,CHIADRONI E,et al. Zemax simulations describing collective effects in transition and diffraction radiation[J]. Optics Express,2018,26(4):5075-5082.
[27] WEI L M,WANG C C,DUAN W R. Simulation of detecting piston error between segmented mirrors by Fizaeu interference technique on ZEMAX[J]. Optik,2019,183:828-834.

Optical path simulation and design of NO rapid detection optical cavity structure

LI Wen*, CAI Yongqing, CHEN Mengfan, LIU Peng

(School of Mechanical and Materials Engineering, North China University of Technology, Beijing 100144, China)

Abstract: According to the chemiluminescence reaction mechanism of nitric oxide and ozone, a new cylindrical total reflection S-type optical cavity design for rapid detection of nitric oxide gas is proposed. The optical cavity model is based on the total reflection S-type structure on the inner wall of the cylinder. Cylindrical light source is used as the light scattering element, so the chemiluminescence light is concentrated on the photosensitive surface of the detector with maximum efficiency, which strengthens the measurement signal of nitric oxide concentration and improves the detection accuracy. Optical software ZEMAX was used to simulate the optical path of the model. The comparative analysis showed that the chemiluminescence collection efficiency of the S-type optical cavity collection path was 36.6%, and is verified by experiments. The experimental results showed that in a certain range of concentration, there is a good linear relationship between the luminescence intensity of the reaction and the concentration of NO gas. The linear correlation was 0.999 2. In the range of 0‰~2.5‰, the detection limit was 0.001‰. The model is simple in structure and meets the national standards. Moreover, it provides a practical idea of design for the on-line exhaust gas detection.

Keywords: S-type optical cavity structure; optical path simulation; NO concentration; chemiluminescence; ZEMAX

Received: 2021-03-03; **Accepted**: 2021-06-06; **Published online**: 2021-07-30 11:10

URL: kns.cnki.net/kcms/detail/11.2625.V.20210730.1047.002.html

Foundation items: National Natural Science Foundation of China (51205005); Beijing Science and Technology Innovation Service Ability Building Supporting Project (PXM2017-014212-000013)

* **Corresponding author**. E-mail: lw@ncut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0646

冰雪天气下基于 MFOA-KELM 残差修正的跑道温度混合预测

陈斌[1,2,*], 刘悦[1,2], 李庆真[1,2], 丁宇[1,2], 王立文[2]

(1. 中国民航大学 电子信息与自动化学院, 天津 300300; 2. 中国民航大学 航空地面特种设备研究基地, 天津 300300)

摘 要: 道面温度短时精准预测是跑道积冰预警的关键因素之一,为了解决单一机理预测模型随预测时间延长而造成误差累积的问题,提出了一种冰雪天气下跑道温度混合预测方法。将跑道温度机理预测模型与核极限学习机(KELM)相结合,建立一种数据驱动修正残差的跑道温度机理预测模型。针对果蝇优化算法(FOA)收敛速度慢、易陷入局部最小值的问题,引入权值更新函数和距离扩充因子,调整果蝇的全局寻优效果,避免陷入局部极小值。利用改进的果蝇优化算法(MFOA)对 KELM 的正则化参数与核参数联合优化,以冰雪天气下跑道温度实际数据为例,建立基于改进果蝇优化核极限学习机(MFOA-KELM)的跑道温度混合预测模型,并在不同时间尺度下对该混合预测模型进行仿真测试。实验结果表明:与单一机理预测模型相比,当预测时长为 120 min 时,MFOA-KELM 混合预测模型的平均绝对误差至少减小了 61.43%,在残差阈值为 ±0.5℃ 时,平均预测准确率为 91.25%。可见,MFOA-KELM 混合预测模型具有更高的预测准确性,研究结论显示该混合预测方法能够为机场跑道温度短时精准预测提供新思路。

关 键 词: 混合建模; 核极限学习机; 改进果蝇算法; 跑道温度; 预测
中图分类号: V219; TP311
文献标志码: A **文章编号**: 1001-5965(2022)11-2153-12

冬季机场跑道面临积冰、积雪的现象,由此常常会引发诸如航班延误、旅客滞留等问题。温度是导致跑道结冰的重要因素,对跑道温度进行短时精准预测会为跑道结冰预警提供有效信息,不仅可以提高航班运行安全裕度,还有助于机场及时采取除冰雪等应急措施,减少航班延误。

多年来,国内外对与跑道结构相似的路面温度开展了大量研究,常用的路面温度预测方法主要有统计回归法和理论分析法。统计回归法指对大量路面温度观测数据进行统计回归建模,如线性回归法[1]、时间序列法[2]等。统计回归模型简单,但需要大量实测数据,模型普适性低[3]。理论分析法是根据气象学和传热学基本原理采用数值分析方法建立的路面温度场预测模型[3]。Barbere[4]于 1957 年最先使用该方法,利用路面介质温度周期变化热传导方程来描述路面温度演变规律。此后,国内外学者对该方法不断进行改进。Christison[5]、严作人[6]等以气温和太阳辐射为边界条件,基于层状路面热传导方程,通过有限差分等数值方法求解,建立了沥青路面温度预测模型。Hu 等[7]根据空气温度、湿度、风速等环境因素建立了基于能量平衡和热传导方程的路面温度预测

收稿日期: 2021-10-29; **录用日期**: 2022-02-28; **网络出版时间**: 2022-04-26 09:03
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220425.1227.004.html
基金项目: 国家自然科学基金委员会-中国民航局民航联合研究基金 (U1933107)
* **通信作者**. E-mail: chenbindavid@163.com

引用格式: *陈斌, 刘悦, 李庆真, 等. 冰雪天气下基于 MFOA-KELM 残差修正的跑道温度混合预测[J]. 北京航空航天大学学报, 2022, 48(11): 2153-2164. CHEN B, LIU Y, LI Q Z, et al. Runway temperature hybrid prediction based on MFOA-KELM residual correction under ice and snow[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2153-2164(in Chinese).*

模型。Nuijten[8]基于能量守恒方法,综合考虑辐射通量、风速等和跑道材料之间的能量平衡关系,建立了跑道温度预测模型。机理模型能够从本质上反映气象、环境和路面结构材料对路面温度场的影响机理,不需要大量实测数据,普适性较强[3],被广泛应用于路面温度预报中[9]。然而在实际运行的机场中,使用机理模型预测时,起降飞机的尾流[8]、除冰液稀释前后释放的热量会使道面温度受到瞬时冲击[10]。除冰雪前后的道面覆盖物状况、除冰雪作业方式等多种因素都会影响机理模型热通量参数的确定,使导热微分方程在求解时需要多种假设简化解析表达式和边界条件[11],这将影响机理模型对道面温度的预测准确度。另外,机理模型的有限差分求解与离散化,当预测时长增加时导致的误差累积问题也不容忽视。

为了解决机理模型的上述缺陷,学者们分别从机理和数据分析角度进行误差补偿。基于数据分析的误差补偿方法在电力、化工、软测量领域应用广泛。姚源朝等[12]利用广义回归神经网络(generalized regression neural network, GRNN)补偿了气流床汽化炉机理模型的预测残差,建立了结合机理与 GRNN 的混合预测模型,但 GRNN 模型需要通过选择合适的目标函数才能平衡参数平滑因子。成庶等[13]使用支持向量机完成了对电容器寿命预测的补偿修正,然而支持向量机在处理规模大的数据样本时,运行速度慢,需要大量的计算时间。核极限学习机(kernel extreme learning machine, KELM)是一种具有强大泛化能力和非线性拟合能力的数据驱动方法,结构简单,运行速度快[14],但是其预测性能受到正则化参数和核函数参数的影响[15]。果蝇优化算法(fruit fly optimization algorithm, FOA)计算简单、调整方便,能够解决 KELM 参数自动优选的问题[16]。目前,FOA 和 KELM 相结合的方法已经应用于医疗、金融等领域。Mythili 和 Kumar[17]将 FOA-KELM 用于解决高维基因特征选择问题;王东[18]提出一种 FOA-KELM 核参数的时间序列预测方法。但 FOA 易陷入局部极小值,出现早熟收敛等情况[19]。

因此,本文提出了一种基于改进 FOA(modified fruit fly optimization algorithm, MFOA)优化 KELM(MFOA-KELM)的跑道温度混合预测模型。首先,通过简化边界条件建立跑道温度机理预测模型;然后,利用 KELM 强大的非线性拟合能力使机理预测残差最小化,并利用 MFOA 寻找 KELM 的最佳参数组合;最后,实现跑道温度短时预测。

1 机理预测模型

1.1 跑道边界条件

跑道结构处于外界自然环境中,受到周围多种环境因素的综合作用,如太阳辐射、降雨、降雪或起降航班等[20]。跑道温度的变化往往以影响因素热量流动的形式在跑道温度场中传导。跑道热量传导过程如图 1 所示。综合考虑各影响因素热通量,基于能量守恒定律,可得跑道温度机理预测模型的边界条件[21]如下:

$$Q_0 + Q_{l2} + H_f + L_f + Q_r + Q_w = Q_s + Q_{l1} + Q_d \tag{1}$$

式中:Q_0 为道面热通量;Q_s 为太阳短波辐射通量;Q_{l1}为大气长波辐射通量;Q_{l2}为道面长波辐射通量;H_f 为感热通量;L_f 为潜热通量;Q_r 为降雨热通量;Q_w 为降雪热通量;Q_d 为航班密度热通量。

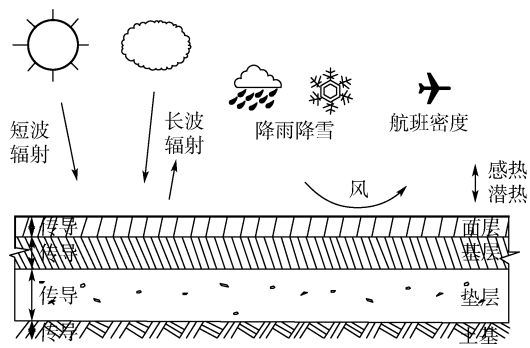


图 1 跑道热量传导过程示意图

Fig. 1 Schematic diagram of runway heat transfer process

1)太阳短波辐射通量 Q_s。其计算式为

$$Q_s = \frac{I_0 r}{1 + f_w} \tag{2}$$

$$f_w = [0.174 + 0.056\ln(1 + E(T_a))]\frac{P}{P_0} \tag{3}$$

$$E(T_a) = 6.11 H_a e^{17.269\frac{T_a-273}{T_a-35.86}} \tag{4}$$

式中:I_0 为太阳常数;r 为跑道对太阳辐射的吸收率;f_w 为大气层对太阳辐射的削弱系数;T_a 为大气温度;P、P_0 分别为大气压强、标准大气压强;H_a 为大气相对湿度。

2)长波辐射通量 Q_{l1}、Q_{l2}。其计算式为

$$Q_{l1} = \sigma T_a^4 \varepsilon_a \tag{5}$$

$$Q_{l2} = \sigma T_0^4 \varepsilon_r \tag{6}$$

式中:σ 为玻尔兹曼常数;T_0 为跑道温度;ε_a、ε_r 分别为大气相对辐射率、道面相对辐射率。

3)感热、潜热通量 H_f、L_f。其计算式为

$$H_f = c_s \rho_a c_a V(T_a - T_0) \tag{7}$$

$$L_f = L \rho_a c_w V(H_a - q_h) \tag{8}$$

$$q_h = 0.622\frac{W_a}{W_0}\cdot\frac{E(T_0)}{P-0.378E(T_0)}+\left(1-\frac{W_a}{W_0}\right)H_a \tag{9}$$

$$E(T_0) = 6.11H_a e^{17.269\frac{T_0-273}{T_0-35.86}} \tag{10}$$

式中:c_s 为感热传输系数;ρ_a 为跑道附近空气密度;c_a 为空气比热容;V 为风速;L 为凝结潜热;c_w 为水汽传输系数;q_h 为道面比湿;W_a 为机场道面吸水量;W_0 为机场道面饱和吸水量,取值范围为 $0<W_a/W_0<0.5$。

4) 降雨、降雪热通量 Q_r、Q_w。其计算式为

$$Q_r = [P_r(T_0-T_w)/3\ 600]c_L \tag{11}$$

$$Q_w = m_s c_{snow}(T_a-T_s) \tag{12}$$

式中:P_r 为降雨速率;T_w 为水膜温度;c_L 为水的比热容;m_s 为雪的质量通量;c_{snow} 为雪的比热容;T_s 为雪的温度。

5) 航班密度热通量 Q_d。航班密度热通量指飞机尾流与跑道之间的热通量,取决于航班密度 N、飞机起降时间 t_a 和飞机尾流在跑道上方产生的温度幅值 ΔT,其计算式为

$$Q_d = -Nt_a\lambda\Delta T \tag{13}$$

式中:λ 为跑道的导热系数。

飞机在起降时,飞机尾流在跑道上方升高的温度,取决于飞机的型号和推力。本文以波音 737-800 机型为例,通过计算飞机在最佳推力下起飞时的燃油消耗量,获得飞机尾流对跑道上方空气产生的温度幅值 ΔT,其计算式为

$$\Delta T = \frac{Q_f}{V_a\rho_a c_a} = \frac{Q_f}{l_a w_a h_a \rho_a c_a} \tag{14}$$

$$Q_f = m_f e_f \eta_f \tag{15}$$

式中:Q_f 为燃油燃烧释放能量;l_a 和 w_a、h_a 分别为起飞距离和受影响空气的宽度、高度;η_f 为燃料燃烧效率;e_f 为燃料能量密度;m_f 为燃料使用量;V_a 为受到飞机尾流影响的空气体积。

波音 737-800 起降时的参数如表 1 所示[8]。

表 1　波音 737-800 起降参数[8]

Table 1　Taking-off and landing parameters of Boeing 737-800[8]

参数	数值
l_a/m	1 431
w_a/m	25
h_a/m	10
m_f/kg	56
η_f	0.95
e_f/(J·kg^{-1})	43×10^6

1.2　跑道温度机理预测模型

跑道温度预测的本质是求解复杂边界条件下跑道温度的导热微分方程。由于跑道为层状结构[22],本文建立一维跑道导热方程,其结构为

$$\frac{\partial T(x,t)}{\partial t} = \frac{\lambda}{\rho c}\cdot\frac{\partial T^2(x,t)}{\partial x^2} \tag{16}$$

式中:t 为时间;x 为空间坐标;ρ 为跑道材料密度;c 为跑道材料比热容;$T(x,t)$ 为当时间为 t 时,地基深度为 x 的跑道温度。

联立式(1)、式(16),通过有限差分法网格划分,由差商近似表示微商[23],可得跑道温度机理预测模型为

$$T_m(i+1) = \frac{2\tau\lambda}{\rho c h^2}T_1(i)+\left(1-\frac{2\tau\lambda}{\rho c h^2}\right)T_0(i)+\frac{2\tau}{\rho c h}Q_0(i) \tag{17}$$

式中:i 为当前时刻,$i+1$ 为下一时刻;T_m 为跑道温度机理预测模型预测值;T_1 为跑道下方深度为 h 处的温度值;τ 为网格划分时间步长;$Q_0(i)$ 为 i 时刻跑道与外界环境交换的热通量,$Q_0(i)=Q_s(i)+Q_{l1}(i)+Q_d(i)-Q_{l2}(i)-H_f(i)-L_f(i)-Q_r(i)-Q_w(i)$。

2　基于 MFOA-KELM 残差修正的混合预测模型

2.1　核极限学习机基本原理

KELM 引入核函数代替 ELM 隐含层节点的特征映射,不需要设置隐含层节点数目、初始权值和偏置,综合考虑经验风险和结构风险[24],根据 KKT(Karush-Kuhn-Tucker)最优化条件,确定网络输出权值 φ 为

$$\varphi = \boldsymbol{H}^{\mathrm{T}}(\boldsymbol{I}/C+\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{U} \tag{18}$$

式中:$\boldsymbol{I}$ 为单位对角矩阵;C 为正则化参数,用于平衡经验风险和结构风险的分配;$\boldsymbol{H}$ 为将输入映射到隐含层的特征向量;$\boldsymbol{U}$ 为期望输出向量。

由此,获得 KELM 输出函数表达式为

$$f(x) = h_1(x)\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{I}/C+\boldsymbol{H}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{U} \tag{19}$$

定义核函数表达式为

$$\begin{cases}\boldsymbol{\Omega}_{\mathrm{ELM}} = \boldsymbol{H}\boldsymbol{H}^{\mathrm{T}}\\ \Omega_{i,j} = h_1(x_i)h_1(x_j) = K(x_i,x_j)\end{cases} \tag{20}$$

式中:$K(x_i,x_j)$ 为核函数。

则式(19)变换为

$$f(x) = \begin{bmatrix}K(x,x_1)\\ \vdots\\ K(x,x_N)\end{bmatrix}^{\mathrm{T}}(\boldsymbol{I}/C+\boldsymbol{\Omega}_{\mathrm{ELM}})^{-1}\boldsymbol{U} \tag{21}$$

核函数的选择有很多,考虑到径向基(radial basis function,RBF)核函数参数设置较少,故本文选用 RBF 核函数,定义为

$$K(x_i,x_j) = \exp(-\gamma\|x_i-x_j\|^2)\qquad \gamma>0 \tag{22}$$

2.2 改进的果蝇优化算法

2.2.1 果蝇优化算法的改进方法

FOA是一种基于果蝇觅食行为寻找全局最优解的群体新智能算法[16],但其在迭代寻优过程中会出现以下问题:果蝇的搜索步长RandomValue为固定在区间[0,1]之间的随机值,搜索步长过小会使果蝇个体收敛速度变慢,不能全局充分搜索,易陷入局部最优解。因此,为避免FOA陷入局部极值,提高果蝇的全局寻优能力和计算效率,本文采用以下方法对FOA进行改进:①采用权值更新函数w_1更新果蝇个体的搜索步长,使搜索步长不再是局限于一定范围内的随机值,平衡了果蝇的局部寻优能力和全局寻优能力。②引入距离扩充因子d计算味道浓度判定值S_i,扩充了S_i的解空间范围。

MFOA迭代寻优步骤如下:

步骤1 设置果蝇种群规模sizepop,最大迭代代数maxiteration,果蝇个体的随机初始位置X_axis、Y_axis。

步骤2 更新果蝇个体的移动位置。利用权值更新函数更新果蝇个体利用嗅觉寻找食物的距离。

$$w_1 = w_0\beta^g \tag{23}$$

$$X_i = X_axis + w_1 \cdot RandomValue \tag{24}$$

$$Y_i = Y_axis + w_1 \cdot RandomValue \tag{25}$$

式中:w_0为初始权值;β为权值系数;g为果蝇当前迭代代数;X_i、Y_i为果蝇个体的更新位置;RandomValue为区间[0,1]之间的随机数。

步骤3 计算果蝇味道浓度判定值S_i。此时果蝇的味道浓度判定值S_i为果蝇个体到原点距离的倒数与距离扩充因子d之和。

$$D_{ist} = \sqrt{X_i^2 + Y_i^2} \tag{26}$$

$$S_i = 1/D_{ist} + d \tag{27}$$

$$d = (\alpha - RandomValue)D_{ist} \tag{28}$$

式中:D_{ist}为果蝇个体到原点的距离;α为常数。

步骤4 更新果蝇群体位置,获得最佳味道浓度Smell(i)。将味道浓度判定值S_i代入味道浓度判定函数(适应度函数)中,可得到每只果蝇个体位置的味道浓度值,通过对比获得果蝇群中味道浓度最佳的果蝇个体。

$$Smell(i) = Fitness(S_i) \tag{29}$$

$$[bestSmell \quad bestIndex] = \min(Smell(i)) \tag{30}$$

式中:Fitness为味道浓度判定函数(适应度函数);[bestSmell bestIndex]为果蝇群中味道浓度最佳的果蝇个体。

步骤5 保留最佳味道浓度值和x、y坐标,果蝇群将以最佳浓度个体为目的地前往飞行。

$$Smellbest = bestSmell \tag{31}$$

$$X_axis = X[bestIndex] \tag{32}$$

$$Y_axis = Y[bestIndex] \tag{33}$$

步骤6 进入迭代寻优,重复执行步骤2~步骤5,判断味道浓度是否优于前一迭代,若是则迭代停止。

2.2.2 改进的果蝇优化算法性能评估

为验证MFOA的有效性与优越性,选用4个标准测试函数对MFOA进行仿真测试,并与FOA、粒子群优化算法(PSO)和差分进化算法(DE)对比。测试函数的表达式、自变量的搜索区间和理论最优值如表2所示[25]。其中,F_1、F_2为单峰函数,F_3、F_4为多峰函数,使用单峰函数来验证算法的收敛速度,使用多峰函数来验证算法跳出局部最小值的能力[26]。

为了运算的公平性,实验选取相同的共性参数,设置种群规模为80,迭代次数为300,对每一种测试函数独立运行30次,测试函数的维数为30。在MFOA中,设置$w_0=0.8$,$\beta=1.1$,$\alpha=0.5$。PSO算法的参数设置详见文献[27]。DE算法的参数设置详见文献[28]。选用各算法对测试函数最佳求解结果的平均值和标准差作为评价优化算法的性能指标,如表3所示。图2为MFOA和FOA对4种测试函数的适应度曲线。

表2 标准测试函数[25]

Table 2 Standard test functions[25]

函数	表达式	搜索区间	理论最优值
F_1	$F_1 = \sum_{i=1}^{d} x_i^2$	[-100,100]	0
F_2	$F_2 = \sum_{i=1}^{d} ix_i^4 + rand[0,1)$	[-1.28,1.28]	0
F_3	$F_3 = \frac{1}{4\,000}\sum_{i=1}^{d} x_i^2 - \prod_{i=1}^{d}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	[-600,600]	0
F_4	$F_4 = \sum_{i=1}^{d}[x_i^2 - 10\cos(2\pi x_i) + 10]$	[-10,10]	0

表 3　优化结果比较

Table 3　Optimization result comparison

函数	MFOA		FOA		PSO		DE	
	平均值	标准差	平均值	标准差	平均值	标准差	平均值	标准差
F_1	$8.195\ 2\times10^{-26}$	$7.752\ 3\times10^{-28}$	0.000 190 81	$7.532\ 3\times10^{-7}$	3.609 5	2.145 4	23.354 1	5.936 5
F_2	$5.769\ 5\times10^{-5}$	$4.638\ 2\times10^{-5}$	0.001 278 8	0.000 277 66	0.185 83	0.065 835	0.329 56	0.085 208
F_3	0	0	1.275×10^{-5}	$5.276\ 7\times10^{-8}$	0.016 197	0.020 383	1.211 3	0.084 048
F_4	0	0	81.204 6	95.036 1	53.307 8	12.502 6	212.245 5	9.562 8

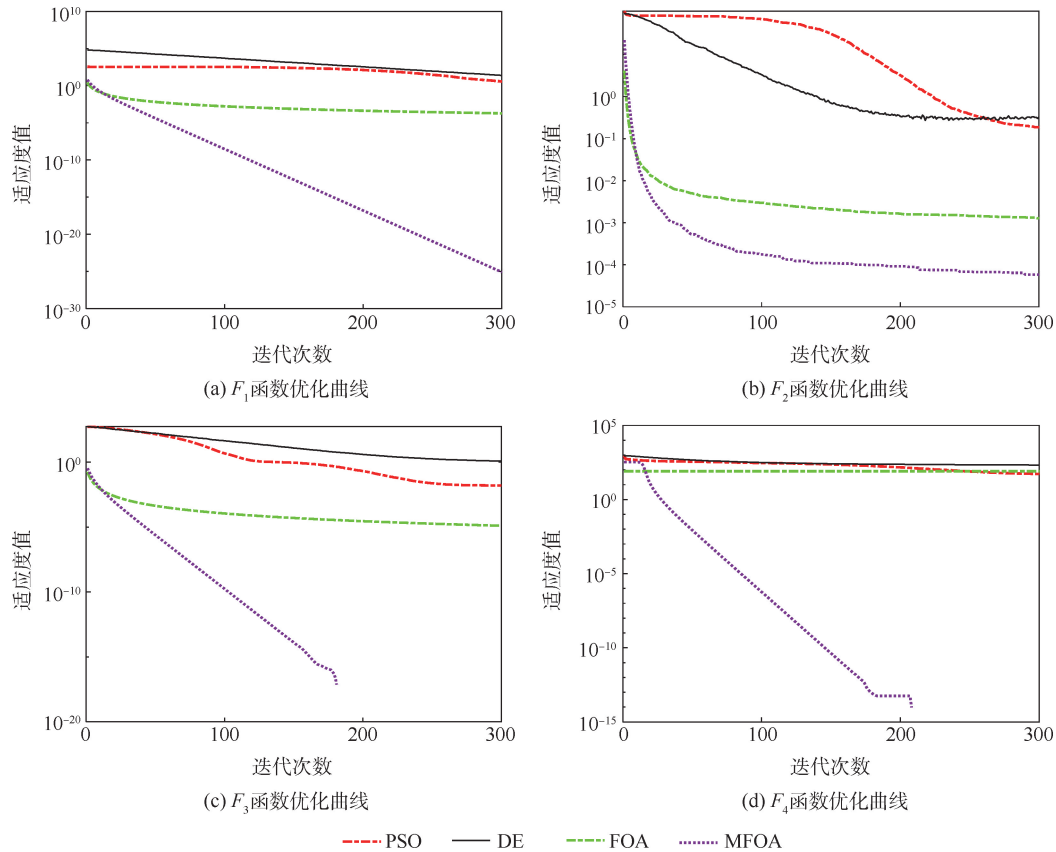


图 2　函数优化曲线

Fig. 2　Function optimization curves

由表 3 可见,在 4 个测试函数的寻优结果中,与 FOA、PSO、DE 算法相比,MFOA 均能求得最小的平均值和标准差,且在 F_3、F_4 测试函数中可以稳定求出全局最优解。由图 2 可见,在 4 个测试函数中,与 FOA、PSO、DE 算法相比,MFOA 收敛速度更快,寻优精度更高。对于 F_3、F_4 多峰测试函数,MFOA 分别在迭代次数为 182 次和 209 次时就达到了最优适应度值 0。可见,MFOA 具有更强的局部搜索能力和全局寻优能力。

2.3　基于 MFOA-KELM 残差修正的混合预测模型建模

2.3.1　混合预测模型的构建

在跑道温度混合预测模型中,以机理预测模型为基础预测模型,综合考虑跑道与其周围太阳短波辐射通量、潜热通量、航班密度等热通量之间的能量流动,基于传热分析从跑道结构特性和交通状况角度预测跑道温度。利用 MFOA-KELM 预测机理预测模型的残差,对机理预测模型预测值进行补偿修正。

基于 MFOA-KELM 残差修正的混合预测模型结构如图 3 所示。具体预测步骤如下:

步骤 1　使用机理预测模型预测跑道温度,获得机理预测模型预测值 T_m 与机理预测模型残差值 e_m。

步骤 2　建立 MFOA-KELM 模型,用于预测机理预测模型的残差,获得机理预测模型残差预测值 e_r。

步骤 3　将步骤 2 预测的残差预测值 e_r 与机

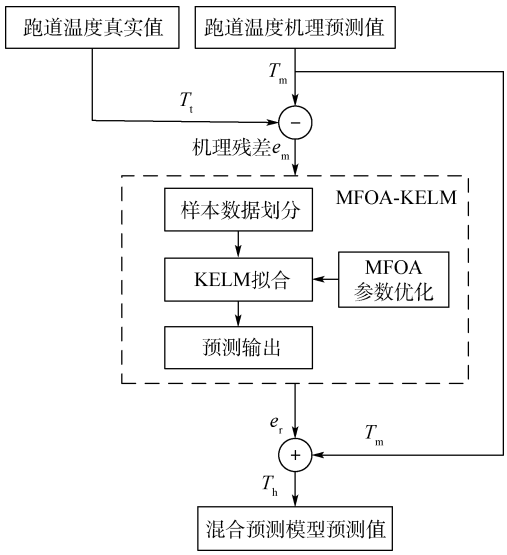


图 3 混合预测模型结构

Fig. 3 Structure of hybrid prediction model

理预测模型预测值 T_m 进行叠加,获得混合预测模型预测值 T_h。

$$T_h = T_m + e_r \tag{34}$$

2.3.2 基于 MFOA-KELM 的机理残差预测

KELM 具有强大的非线性拟合能力,能够根据跑道周围多影响因素与机理预测模型残差之间的数据关系建立非线性映射,预测下一个时刻机理预测模型的残差值。但 KELM 的预测性能与其内部正则化参数 C 与核函数参数 γ 密切相关。因此,本文采用 MFOA 对 KELM 进行参数寻优,建立 MFOA-KELM 修正模型。具体预测流程如图 4所示。

将机理残差值 e_m 作为 MFOA-KELM 模型的输出部分,将跑道周围多影响因素——气温、大气湿度、风速、大气压强和道面下方深度为 h 处的温度作为 MFOA-KELM 模型的输入部分。设置果蝇种群的初始参数,以训练样本的均方误差作为果蝇参数寻优的适应度函数。适应度函数的计算表达式为

$$\text{Fitness}(S_i) = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2 \tag{35}$$

式中:x_i 为第 i 个预测点的实际值;x_i'为第 i 个预测点的预测值;n 为预测点个数。均方误差越小,表示寻优个体的适应度值越好,MFOA-KELM 模型预测的机理残差值越准确。当 KELM 模型获得最优正则化参数 C 和核函数参数 γ 时,迭代停止,输出机理预测模型残差预测值 e_r。

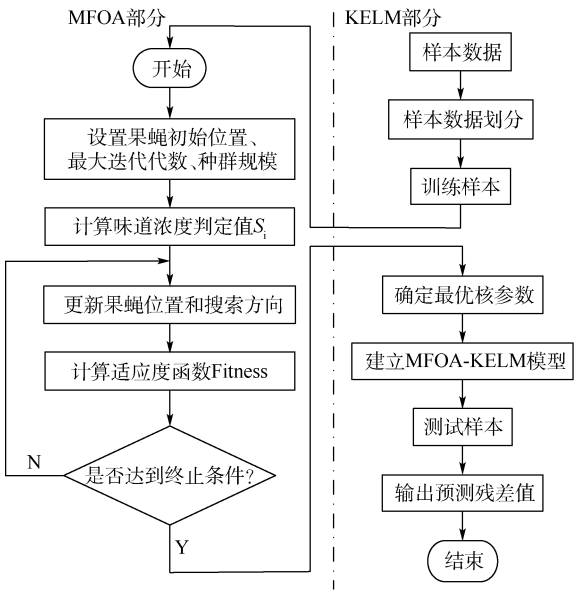


图 4 MFOA-KELM 模型流程

Fig. 4 Flow chart of MFOA-KELM model

3 实例验证

3.1 数据来源

为验证本文所提的混合预测模型对冰雪天气下跑道温度预测的可行性,笔者于 2019 年 12 月 23 日至 2020 年 1 月 7 日在模拟跑道温度实验平台中采集实验数据,实验平台如图 5 所示。模拟跑道道面为根据 MH/T 5004—2010[29] 设计的水泥混凝土道面。采用 ARS31Pro 道面传感器、Vaisala



(a) 数据采集端

(b) 数据接收端

图 5 模拟跑道温度实验系统采集平台

Fig. 5 Simulation of runway temperature experimental system collecting platform

wtx520 气象传感器、湿度传感器、PT100 温度传感器依次测量风速、气温、大气湿度、大气压强和道面下 10 cm 深度处温度。选取表 4 中4 组实验数据样本建模预测。在每组样本中选取前 10 个小时数据为输入样本集,以每分钟样本输入数据的未来 15 min,30 min,60 min,120 min温度为样本输出集。将每组样本的前 80% 数据作为训练集用于构建模型,后 20% 数据作为测试集用于模型验证。其中,样本 1 为持续性降雪天气,最低气温 -0.38℃,最高气温0.6℃。样本 2 ~ 样本 4 为阴天天气,最低气温依次为 -2.31℃、-1.64℃、-1.3℃,最高气温依次为 3.87℃、1.4℃、3.6℃。

表 4　实验数据样本

Table 4　Experimental data samples

样本	日期	时间	采样间隔/min
1	2020-01-05	20:00—08:00	1
2	2019-12-23	10:00—22:00	1
3	2019-12-28	19:00—07:00	1
4	2020-01-07	08:00—20:00	1

3.2　数据归一化

在 MFOA-KELM 预测机理预测模型残差时,为了避免各跑道温度影响因子由于量纲不同对模型预测精度产生影响,将样本数据归一化到[0,1]区间。

$$x_1 = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{36}$$

式中:$x_{\min}$为样本数据最小值;$x_{\max}$为样本数据最大值;x_i、x_1 分别为样本数据归一化前、后的数值。

3.3　模型评价指标

选取平均绝对误差(MAE)和平均预测准确率(ACC)对混合预测模型预测准确性进行评价分析[30-31]。MAE 数值越小,ACC 数值越大,表示模型预测越准确。计算公式如下:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |x_i - x_i'| \tag{37}$$

$$\mathrm{ACC} = \frac{m}{n} \times 100\% \tag{38}$$

式中:m 为预测残差在规定残差阈值内点的个数。

3.4　混合预测模型预测结果及分析

3.4.1　模型参数设置

机理预测模型相关参数设置如表 5 所示。表中:c_r 为跑道结构材料比热容。设置 FOA 种群规模为 10,迭代代数为 100,优化参数个数 $d_1 = 2$,正则化参数 C 与核参数 γ 的取值范围分别为[0,10]和[0,100],其他参数设置与 2.2.2 节相同。对训练样本以均方误差最小为优化目标,得到 MFOA-KELM 预测残差值,将其与跑道温度机理预测模型预测值叠加获得 MFOA-KELM 混合预测模型预测值。为充分验证 MFOA-KELM 混合预测模型的有效性,分别与 FOA-KELM、KELM 混合预测模型预测值对比,选取相同的数据进行预测。

表 5　机理预测模型参数

Table 5　Mechanistic prediction model parameters

参数	数值
$I_0/(\mathrm{W \cdot m^{-2}})$	1 367
P_0/kPa	101.325
$c_L/(\mathrm{J \cdot kg^{-1} \cdot K^{-1}})$	4 200
$\rho_a/(\mathrm{kg \cdot m^{-3}})$	1.29
$c_{snow}/(\mathrm{J \cdot kg^{-1} \cdot K^{-1}})$	2 100
$c_a/(\mathrm{J \cdot kg^{-1} \cdot K^{-1}})$	1 005.46
$\sigma/(\mathrm{W \cdot m^{-2} \cdot K^{-4}})$	5.67×10^{-8}
$c_r/(\mathrm{J \cdot kg^{-1} \cdot K^{-1}})$	970
$\rho/(\mathrm{kg \cdot m^{-3}})$	1 930
$\lambda/(\mathrm{W \cdot m^{-1} \cdot K^{-1}})$	1.7
$W_0/(\mathrm{kg \cdot m^{-2}})$	0.5
ε_r	0.9
r	0.7
ε_a	0.929
τ_1/s	900
τ_2/s	1 800
τ_3/s	3 600
τ_4/s	7 200

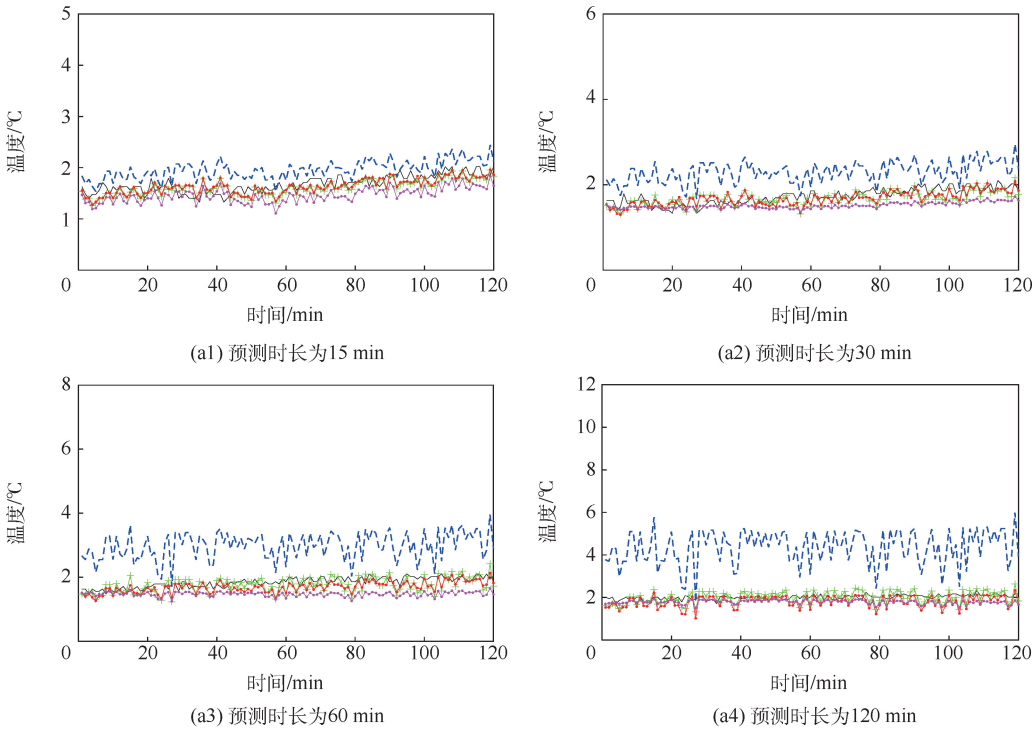
3.4.2　预测结果及模型误差对比分析

上述 3 种混合预测模型与单一机理预测模型的预测结果如图 6 所示,平均绝对误差如表 6 所示。图 6 中,每组样本的机理预测模型预测值表明,预测时长越短,机理残差越小,随着预测时间延长,预测残差不断增大。机理预测模型会随着预测时长的增加不断累积误差。在表 6 中,当预测同一时长时,机理预测模型的平均绝对误差最大。随着预测时长增加,MFOA-KELM 混合预测模型均具有最小的平均绝对误差值。当预测时长为 120 min 时,与机理预测模型、FOA-KELM 混合、KELM 混合预测模型相比,MFOA-KELM 混合预测模型的平均绝对误差分别至少减小了 61.43%、20.16% 和 26.80%。可见,引

入权值更新函数和距离扩充因子的MFOA，增强了果蝇种群对正则化参数 C 与核函数参数 γ 的全局寻优能力，能够获得更小的适应度函数值，提升了MFOA-KELM模型对机理残差的预测水平，进而使MFOA-KELM混合预测模型具有更好的预测准确性。

3.4.3 模型准确率对比及分析

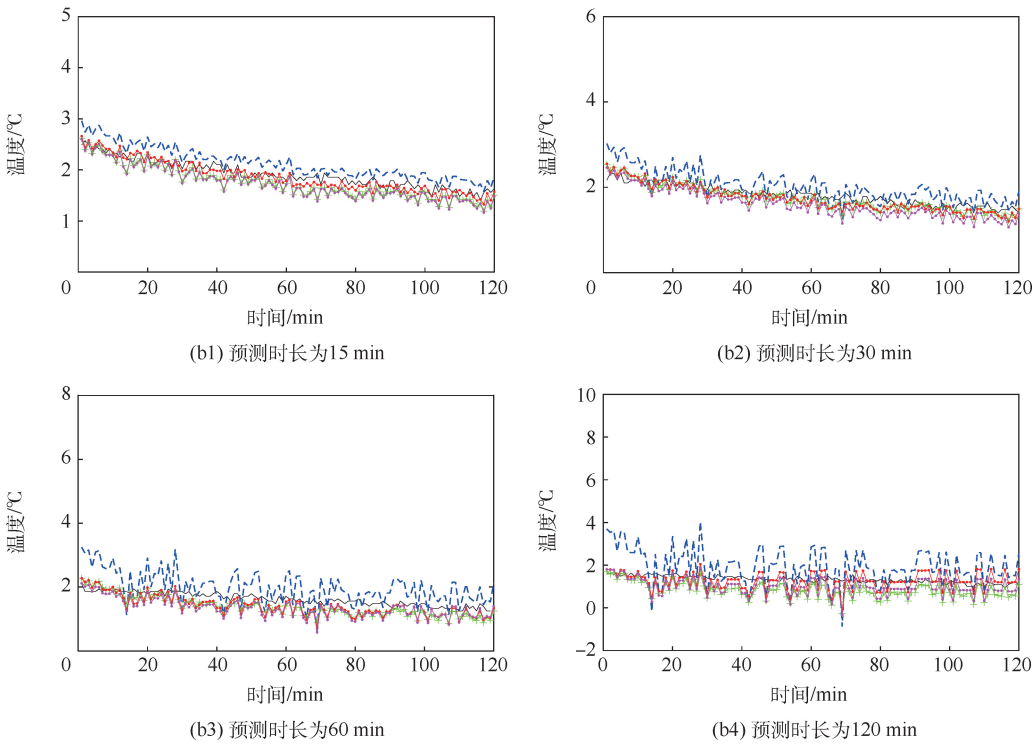
表7为4组样本数据在不同残差阈值内（±0.5℃，±1℃，±2℃）的平均预测准确率。可见，当残差阈值为±0.5℃时，在同一预测时长下，MFOA-KELM混合预测模型的预测准确率明显高于机理预测模型。在不同预测时长下，MFOA-KELM混合预测模型的平均预测准确率同为最高。当预测时长为120 min时，在3种残差阈值下，MFOA-KELM混合预测模型的平均预测准确率至少达到91.25%。可见，与单一的机理预测模



(a1) 预测时长为15 min　(a2) 预测时长为30 min

(a3) 预测时长为60 min　(a4) 预测时长为120 min

(a) 样本1



(b1) 预测时长为15 min　(b2) 预测时长为30 min

(b3) 预测时长为60 min　(b4) 预测时长为120 min

(b) 样本2

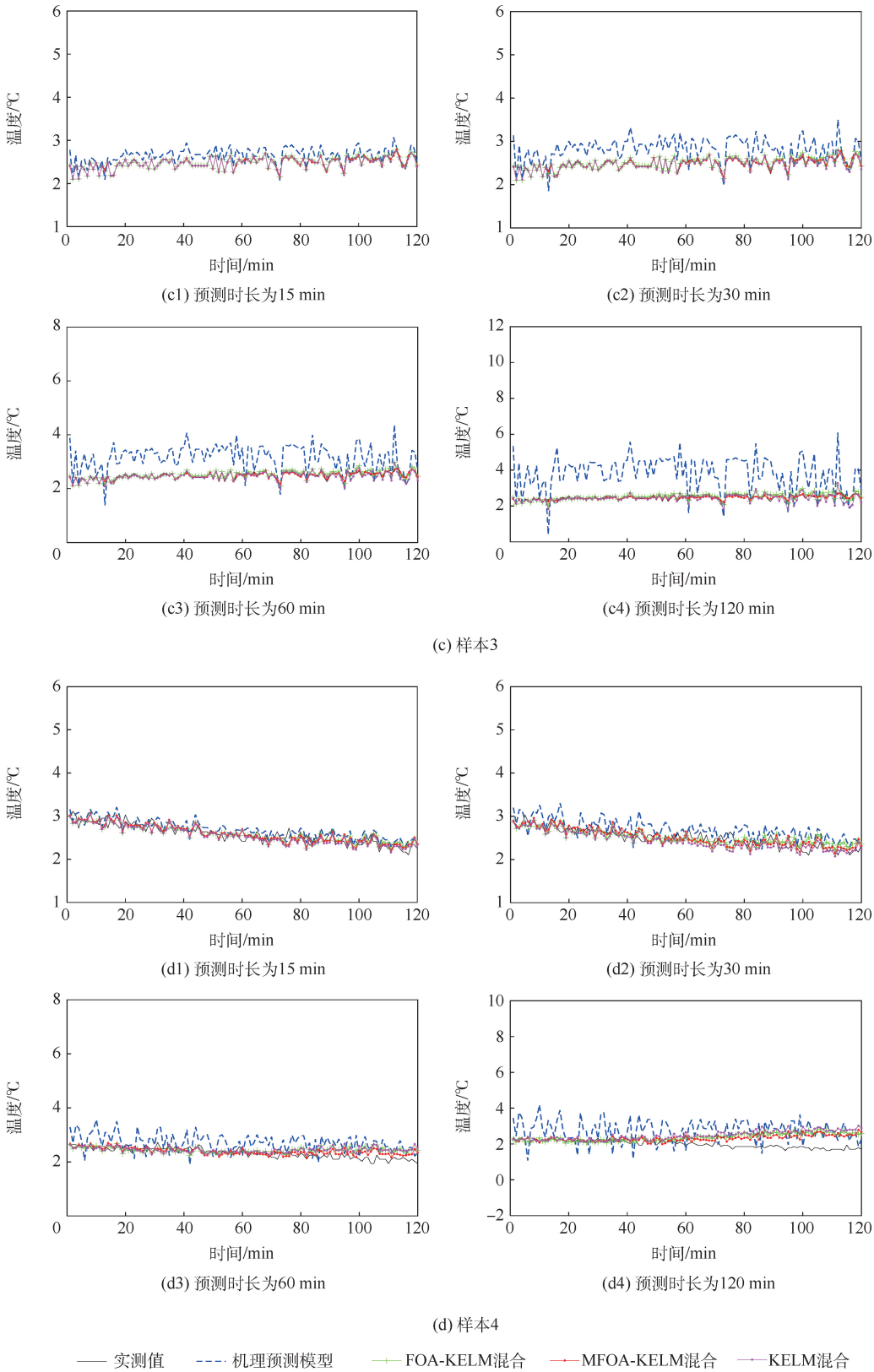


图 6　样本 1 ~ 样本 4 的预测结果

Fig. 6　Prediction results of samples 1, 2, 3 and 4

型相比，MFOA-KELM 混合预测模型利用 KELM 强大的非线性拟合能力能够充分挖掘道面温度影响因素，如气温、大气湿度、风速、地基温度等与道面温度机理预测残差之间的内部信息，由 MFOA-KELM 混合预测模型预测的残差值补偿了机理预测模型预测值与跑道温度真实值之间的偏差，使 MFOA-KELM 混合预测模型比单一机理预测模型具有更高的预测准确率。

表6 样本1~样本4的平均绝对误差

Table 6 Mean absolute error of samples 1, 2, 3 and 4

样本	预测时长/min	平均绝对误差			
		机理预测模型	MFOA-KELM混合	FOA-KELM混合	KELM混合
1	15	0.285	0.111	0.129	0.232
	30	0.575	0.130	0.147	0.222
	60	1.144	0.149	0.207	0.369
	120	2.336	0.194	0.243	0.265
2	15	0.219	0.087	0.170	0.191
	30	0.268	0.124	0.128	0.218
	60	0.479	0.191	0.269	0.296
	120	0.879	0.339	0.459	0.516
3	15	0.184	0.007	0.014	0.017
	30	0.368	0.014	0.031	0.034
	60	0.736	0.028	0.067	0.069
	120	1.472	0.055	0.106	0.138
4	15	0.144	0.093	0.098	0.099
	30	0.230	0.095	0.107	0.111
	60	0.428	0.120	0.177	0.182
	120	0.849	0.186	0.415	0.503

表7 不同残差阈值内平均预测准确率

Table 7 Average prediction accuracy under different residual threshold values

预测时长/min	预测模型	不同允许误差范围内平均预测准确率/%		
		±0.5℃	±1℃	±2℃
15	机理预测模型	97.29	100.00	100.00
	MFOA-KELM混合	100.00	100.00	100.00
30	机理预测模型	58.75	81.04	100.00
	MFOA-KELM混合	100.00	100.00	100.00
60	机理预测模型	39.58	73.96	99.58
	MFOA-KELM混合	100.00	100.00	100.00
120	机理预测模型	18.75	40.42	74.38
	MFOA-KELM混合	91.25	99.38	100.00

4 结 论

1）为提高FOA的收敛精度，避免算法出现早熟收敛的现象，本文通过引入权值更新函数和距离扩充因子来改进算法，避免原始算法陷入局部极小值。对4个标准测试函数进行仿真实验，将MFOA与FOA、其他群体智能算法进行比较，实验结果表明，本文提出的MFOA具有更高的收敛性能和全局寻优能力。

2）建立基于MFOA-KELM残差修正的跑道温度混合预测模型，以实测数据验证了该混合方法对跑道温度预测的有效性和泛化性。与单一机理预测模型相比，当预测时长为120 min时，MFOA-KELM混合预测模型的平均绝对误差至少减小61.43%，预测准确性更高。

本文混合预测模型能够为跑道积冰预警提供有效信息。后续研究中将把跑道温度混合预测值作为跑道结冰预测的目标温度值，进一步展开对跑道结冰预测的研究。

参考文献（References）

[1] KRŠMANC R, SLAK A Š, DEMŠAR J. Statistical approach for forecasting road surface temperature[J]. Meteorological Applications, 2013, 20(4): 439-446.

[2] 汤筠筠，郭忠印. 基于自回归求和移动平均的冬季路温短临预测[J]. 同济大学学报(自然科学版)，2017，45(12)：1824-1829.
TANG J J, GUO Z Y. Winter road temperature short prognosis based on autoregressive summation moving average[J]. Journal of Tongji University (Natural Science Edition), 2017, 45(12): 1824-1829(in Chinese).

[3] 秦健，孙立军. 国外沥青路面温度预估方法综述[J]. 中外公路，2005，25(6)：19-23.
QIN J, SUN L J. A review of foreign asphalt pavement temperature prediction methods[J]. Chinese and Foreign Highways, 2005, 25(6): 19-23(in Chinese).

[4] BARBERE S. Calculation of maximum pavement temperatures from weather reports[R]. Washington, D. C.: National Research Council, 1957: 1-8.

[5] CHRISTISON J T, ANDERSON K O. The response of asphalt pavements to low temperature climatic environments[C]// International Conference on the Structural Design of Asphalt Pavements, 1972: 44-56.

[6] 严作人. 层状路面体系的温度场分析[J]. 同济大学学报，1984(3)：76-85.
YAN Z R. Temperature field analysis of laminated pavement systems[J]. Journal of Tongji University, 1984(3): 76-85(in Chinese).

[7] HU Y, ALMKVIST E, LINDBERG F, et al. The use of screening effects in modelling route-based daytime road surface temperature[J]. Theoretical and Applied Climatology, 2016, 125(1): 303-319.

[8] NUIJTEN A D W. Runway temperature prediction, a case study for Oslo Airport, Norway[J]. Cold Regions Science and Technology, 2016, 125: 72-84.

[9] 邱金晶，陈锋，张珏，等. 浙江省高速公路路面温度预报模型性能检验及应用[J]. 气象科技，2020，48(4)：518-528.
QIU J J, CHEN F, ZHANG J, et al. Performance test and application of highway pavement temperature forecasting model in Zhejiang province[J]. Meteorological Science and Technology, 2020, 48(4): 518-528(in Chinese).

[10] 赵鸿铎，姚祖康，张长安，等. 飞机除冰液对停机坪水泥混凝土的影响[J]. 交通运输工程学报，2004，4(2)：1-5.
ZHAO H D, YAO Z K, ZHANG C A, et al. Effects of aircraft de-icing fluid on apron cement concrete[J]. Journal of Transportation Engineering, 2004, 4(2): 1-5(in Chinese).

[11] 孙立军，秦健. 沥青路面温度场的预估模型[J]. 同济大学

学报(自然科学版),2006,34(4):480-483.
SUN L J,QIN J. Prediction model for temperature field of asphalt pavement[J]. Journal of Tongji University (Natural Science Edition),2006,34(4):480-483(in Chinese).

[12] 姚源朝,仇鹏,许建良,等. 基于混合模型的气流床气化炉建模[J]. 化工学报,2021,72(5):2727-2734.
YAO Y Z,QIU P,XU J L,et al. Modeling of a gasifier in a fluid bed based on a hybrid model[J]. Journal of Chemical Engineering,2021,72(5):2727-2734(in Chinese).

[13] 成庶,刘嘉文,伍珣,等. 基于特征提取与误差补偿的金属化薄膜电容器剩余寿命预测[J]. 中国电机工程学报,2022,42(7):2672-2680.
CHENG S,LIU J W,WU X,et al. Residual life prediction of metalized film capacitors based on feature extraction and error compensation[J]. Proceedings of the CSEE,2022,42(7):2672-2680(in Chinese).

[14] HUANG G B. An insight into extreme learning machines:Random neurons,random features and kernels[J]. Cognitive Computation,2014,6(3):376-390.

[15] 亓孝武,李可军,于小晏,等. 基于核极限学习机和 Bootstrap 方法的变压器顶层油温区间预测[J]. 中国电机工程学报,2017,37(19):5821-5828.
QI X W,LI K J,YU X Y,et al. Transformer top oil temperature interval prediction based on nuclear limit learning machine and Bootstrap method[J]. Proceedings of the CSEE,2017,37(19):5821-5828(in Chinese).

[16] PAN W T. A new fruit fly optimization algorithm:Taking the financial distress model as an example[J]. Knowledge-Based Systems,2012,26:69-74.

[17] MYTHILI S,KUMAR A V. Hybrid feature selection and heterogeneous clustering ensemble framework for detection of circulating tumor cells[J]. Journal of Medical Imaging and Health Informatics,2016,6(5):1160-1166.

[18] 王东. 在线核极限学习机的改进与应用研究[D]. 蚌埠:安徽财经大学,2015.
WANG D. Research on the improvement and application of online kernel limit learning machine[D]. Bengbu:Anhui University of Finance and Economics,2015(in Chinese).

[19] 皮骏,马圣,张奇奇,等. 基于改进果蝇算法优化的 GRNN 航空发动机排气温度预测模型[J]. 航空动力学报,2019,34(1):8-17.
PI J,MA S,ZHANG Q Q,et al. Optimized GRNN aero-engine exhaust temperature prediction model based on improved fruit fly algorithm[J]. Journal of Aerodynamics,2019,34(1):8-17(in Chinese).

[20] 周淑贞. 气象学与气候学[M]. 北京:高等教育出版社,2004.
ZHOU S Z. Meteorology and climatology[M]. Beijing:Higher Education Press,2004(in Chinese).

[21] 翁笃鸣,李炬,高歌. 晴天太阳总辐射的参数化及气候计算[J]. 气象科学,1997,17(1):1-9.
WENG D M,LI J,GAO G. Parameterization of total solar radiation on sunny days and climate calculation[J]. Meteorological Science,1997,17(1):1-9(in Chinese).

[22] 戚春香,李瑶,杨简,等. 青藏高原机场跑道多年冻土地基温度场特征[J]. 交通运输工程学报,2019,19(1):33-47.
QI C X,LI Y,YANG J,et al. Characteristics of multi-year permafrost foundation temperature field of airport runway on Qinghai-Tibet Plateau[J]. Journal of Transportation Engineering,2019,19(1):33-47(in Chinese).

[23] 顾尔祚. 流体力学有限差分法基础[M]. 上海:上海交通大学出版社,1988.
GU E Z. Fundamentals of finite difference method for fluid mechanics[M]. Shanghai:Shanghai Jiaotong University Press,1988(in Chinese).

[24] 杨锡运,邢国通,马雪,等. 一种核极限学习机分位数回归模型及风电功率区间预测[J]. 太阳能学报,2020,41(11):300-306.
YANG X Y,XING G T,MA X,et al. A kernel limit learning machine quantile regression model and wind power interval prediction[J]. Journal of Solar Energy,2020,41(11):300-306(in Chinese).

[25] SURJANOVIC S,BINGHAM D. Virtual library of simulation experiments:Test functions and datasets[EB/OL]. [2021-10-01]. http://www.sfu.ca/~ssurjano.

[26] 蔡赛男,宋卫星,班利明,等. 基于鲸鱼算法优化 LSSVM 的滚动轴承故障诊断[J]. 控制与决策,2022,37(1):230-236.
CAI S N,SONG W X,BAN L M,et al. Fault diagnosis of rolling bearings based on whale algorithm optimized LSSVM[J]. Control and Decision Making,2022,37(1):230-236(in Chinese).

[27] 谷远利,张源,芮小平,等. 基于免疫算法优化 LSSVM 的短时交通流预测[J]. 吉林大学学报(工学版),2019,49(6):1852-1857.
GU Y L,ZHANG Y,RUI X P,et al. Short-time traffic flow prediction based on immune algorithm optimized LSSVM[J]. Journal of Jilin University (Engineering Edition),2019,49(6):1852-1857(in Chinese).

[28] 吴桂芳,崔勇,刘宏,等. 基于差分进化算法的三维电场传感器解耦标定方法[J]. 电工技术学报,2021,36(19):3993-4001.
WU G F,CUI Y,LIU H,et al. A decoupled calibration method for three-dimensional electric field sensors based on differential evolutionary algorithm[J]. Journal of Electrical Engineering Technology,2021,36(19):3993-4001(in Chinese).

[29] 中国民用航空局. 民用机场水泥混凝土道面设计规范:MH/T 5004—2010[S]. 北京:中国民用航空局,2010.
Civil Aviation Administration of China. Specification for airport-cement concrete pavement design:MH/T 5004—2010[S]. Beijing:Civil Aviation Administration of China,2010(in Chinese).

[30] 周哲韬,刘路,宋晓,等. 基于 Transformer 模型的滚动轴承剩余使用寿命预测方法[J/OL]. 北京航空航天大学学报,2021(2021-07-12)[2021-10-26]. https://doi.org/10.13700/j.bh.1001-5965.2021.0247.
ZHOU Z T,LIU L,SONG X,et al. A method for predicting the remaining service life of rolling bearings based on Transformer model[J/OL]. Journal of Beijing University of Aeronautics and Astronautics,2021(2021-07-12)[2021-10-26]. https://doi.org/10.13700/j.bh.1001-5965.2021.0247(in Chinese).

[31] 陈景龙,王日新,李玉庆,等. 一种基于 SD-ICA 的卫星电池健康状态估计方法[J]. 北京航空航天大学学报,2021,47(10):2058-2067.

CHEN J L, WANG R X, LI Y Q, et al. A SD-ICA-based method for satellite battery health state estimation[J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(10): 2058-2067(in Chinese).

Runway temperature hybrid prediction based on MFOA-KELM residual correction under ice and snow

CHEN Bin[1,2,*], LIU Yue[1,2], LI Qingzhen[1,2], DING Yu[1,2], WANG Liwen[2]

(1. School of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China;
2. Department of Aviation Ground Special Equipment, Civil Aviation University of China, Tianjin 300300, China)

Abstract: The runway surface temperature short-term accurate prediction is one of the key factors for runway icing warning. In order to solve the problem of error accumulation caused by a single mechanistic model with increasing prediction time, a hybrid runway temperature prediction method under ice and snow is proposed. The runway temperature mechanism model is combined with the kernel extreme learning machine (KELM) to develop a data-driven model for correcting the mechanism residuals. To address the problem that the fruit fly optimization algorithm (FOA) is slow to converge (converges slowly) and easily falls into local minima. By introducing a distance expansion factor and a weight update function, it is possible to modify the effect of the FOA's search for the global optimal solution and prevent falling into local minima. The modified fruit fly optimization algorithm (MFOA) is used to jointly optimize the KELM regularization parameter and the kernel parameter. A hybrid runway temperature prediction model is developed based on the modified fruit fly optimized kernel extreme learning machine (MFOA-KELM) with the actual data of runway temperature under ice and snow. The hybrid model is simulated and tested under different time lengths. The experimental results show that compared with the single mechanism prediction model, the mean absolute error of the MFOA-KELM hybrid model is reducedby at least 61.43% when the prediction length is 120 minutes, and the average prediction accuracy is 91.25% when the residual threshold is ±0.5℃. It can be seen that the MFOA-KELM hybrid model has higher prediction accuracy. The research findings show that this hybrid prediction method can provide a new idea for short time accurate prediction of airport runway temperatures.

Keywords: hybrid modeling; kernel extreme learning machine; modified fruit fly optimization algorithm; runway temperature; predict

Received: 2021-10-29; **Accepted**: 2022-02-28; **Published online**: 2022-04-26 09:03

URL: kns.cnki.net/kcms/detail/11.2625.V.20220425.1227.004.html

Foundation item: Joint Fund of Civil Aviation Research of National Natural Science Foundation of China and Civil Aviation Administration of China (U1933107)

* **Corresponding author**. E-mail: chenbindavid@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0099

基于燃料电池的复合电源式装载机分层控制

张志文[1,2,*], 杜文杰[1], 梁君飞[1], 张艳岗[1], 武雅文[1]

(1. 中北大学 能源与动力工程学院, 太原 030051; 2. 燕山大学 车辆与能源学院, 秦皇岛 066004)

摘 要: 装载机能耗高、排放差,研究装载机新能源技术具有重要意义。结合装载机工况特性提出了燃料电池与超级电容联合驱动的电源系统,围绕复杂工况下燃料电池和超级电容系统动态模型的实时工况数据进行自适应能量管理策略研究。设计了复合电源拓扑结构与动力传动方案,建立装载机复杂工况下系统多状态模型,基于 Haar 小波理论对整车系统进行功率分流,提出模糊逻辑能量管理策略动态平衡需求功率中的低频分量,采用粒子群算法对控制系统进行优化。仿真结果显示:载荷功率经过最优阈值 3 层 Haar 小波处理后,功率变化大幅度减缓,有效提升燃料电池系统的寿命;模糊逻辑控制器输出的燃料电池功率曲线变化光滑,超级电容 SOC 值处于设定区域内,提高复合电源系统的综合效率;经过粒子群算法优化控制器后,燃料电池输出平均功率同比下降约 5%,超级电容 SOC 值在约 0.6 达到动态平衡状态,改善了装载机的动态响应和稳定性。

关 键 词: 复合电源; 装载机; 多状态模型; Haar 小波; 模糊逻辑; 粒子群算法

中图分类号: U469.72+2

文献标志码: A **文章编号**: 1001-5965(2022)11-2165-12

当今社会通过科技变革与新能源技术创新推动着汽车的节能减排[1]。在工程机械领域,传统的工程机械排放大,能量利用率低,造成能源浪费与环境污染等问题[2-3]。近年来,国内外通过混合动力、纯电动驱动等技术来提高工程机械能源利用率,国外工程机械生产厂商如卡特彼勒、沃尔沃、约翰迪尔、日立建机、川崎等新能源工程机械产品早已商业化[1]。

文献[4]以某 5 t 轮式装载机为研究对象,探索工程机械新能源汽车新技术。装载机需求功率变化频繁、整车冲击振动剧烈,因此,新能源装载机能量部件需要具备高功率密度与高能量密度的特性,储能元件应具有快速充放电特性与高循环寿命。

本文设计燃料电池复合电源式装载机。以燃料电池作为主动力源、超级电容作为辅助能源构建复合电源系统,通过电力驱动满足负载液压系统与行走系统的需求。在不同负载工况下,控制系统通过调节功率分配因子动态调节燃料电池与超级电容功率输出以满足装载机工况需求。要发挥复合电源的优势需充分利用超级电容功率密度高、循环寿命长、充放电速度快的优点及燃料电池高效率、零排放、寿命长、能量密度高的优点。

韩国于 2016 年着手燃料电池挖掘机的商用化,2018 年氢能经济路线图提出要“培育以燃料电池叉车和挖掘机为中心的氢能建筑机械产业”。2020 年 3 月,现代工程机械、现代摩比斯与现代汽车联合开发全球首台氢动力工程机械设备,计划于 2023 年进行批量生产和分销。2020 年 7 月,英国工程机械生产商 JCB 开发出建筑行

收稿日期: 2021-03-02; **录用日期**: 2021-08-23; **网络出版时间**: 2021-09-28 08:55

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210924.2317.002.html

基金项目: 国家自然科学基金 (51605447); 山西省基础研究计划 (201901D211208); 山西省高等学校科技创新项目 (2019L0605)

* **通信作者**. E-mail: zhzhw666@nuc.edu.cn

引用格式: 张志文, 杜文杰, 梁君飞, 等. 基于燃料电池的复合电源式装载机分层控制[J]. 北京航空航天大学学报, 2022, 48(11): 2165-2176. ZHANG Z W, DU W J, LIANG J F, et al. Layered control of hybrid power loader based on fuel cell[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2165-2176(in Chinese).

业第一台氢动力挖掘机,已在试验场对一台由氢燃料电池驱动的 20 t 220X 挖掘机进行了为期 12 个月的测试。2020 年 3 月,潍柴动力股份有限公司氢燃料电池发动机工厂正式投产使用,预计年产量约20 000 台,这是其氢燃料电池商业化的关键一步。

高效合理的能量管理策略是解决复合电源系统功率分配的关键。2017 年,法国贝松实验室 Munoz 等[5]提出了基于电池和超级电容总功率共享的实时控制策略,建立非线性自回归神经网络预测模型,设计了离散小波变换的时域滤波器,延长了动力电池的寿命,控制策略具有良好的鲁棒性。2018 年, Ben Elghali 等[6]研究了以质子交换膜燃料电池为主体的应用于汽车上的复合电源系统,提出了一种基于自适应滤波的能量管理策略,采用模糊逻辑动态调整分离频率,驱动循环仿真表明,燃料电池与超级电容都取得了良好的性能。S. G. R. E 机构的 Ibrahim 等[7]基于模糊逻辑、向量机控制、频率解耦等策略研究燃料电池汽车能量管理,改善了电动汽车动态响应速度和稳定性,根据动力电池充放电状态、燃料电池氢气消耗量、整车效率等参数进行参数匹配。浙江大学 Zhang 等[8]在 2019 年通过多层感知与神经网络方法提取复合电源系统的特征参数,通过驱动模式识别实现燃料电池和超级电容实时功率优化分配。

文献对燃料电池复合电源系统的研究集中在汽车上,在工程机械领域的研究甚少,由于工况的差异性导致工程机械复合电源系统控制策略更加复杂。为了满足装载机工况需求,动力系统需实现大功率和高能量的双重目标,本文将燃料电池和超级电容组成复合电源系统,在频繁变化的工况下,由于超级电容的存在,可以瞬间动态平衡负载功率需求,同时燃料电池能够提供稳定的功率输出。因此,超级电容与燃料电池组成的复合电源系统是解决电动装载机电源问题的重要途径[9],但复合储能使得系统的功率与能量流变得更加错综复杂,目前亟须新技术解决多模式能量分流关系,制定高效合理的能量管理与优化策略,才能充分发挥复合储能系统的性能和优势,也是打破工程机械车载储能技术瓶颈的关键。

本文通过分析传统装载机工况特征,采集各动力系统的性能参数,设计复合电源式装载机动力传动方案,对燃料电池、超级电容、动力驱动等系统进行建模;考虑到系统能量管理过程中一些混沌和暂态需求功率分量对复合电源系统耦合特性和燃料电池耐久性的影响,基于三阶 Haar 小波变换分解需求功率;设计模糊逻辑控制策略,将小波处理后的低频负载分量、超级电容 SOC 值作为输入,输出功率分配因子,以系统的经济性为控制目标对复合电源系统进行功率分配;基于粒子群(particle swarm optimization, PSO)算法优化模糊逻辑隶属度函数,提高复合电源系统的效率与寿命,并提高动力系统的响应速度。

1 系统传动方案及模型建立

如图 1 所示,燃料电池和超级电容组成复合电源系统作为纯电动轮式装载机的能量系统,通过电动机为动力传动系统与液压系统提供动力,控制器根据实时采集工况数据进行能量管理。

复合电源系统主要包括燃料电池、超级电容、DC/DC 变换器等元件,通过电力驱动将动力分配到工作液压系统和行走系统[4]。超级电容作为辅

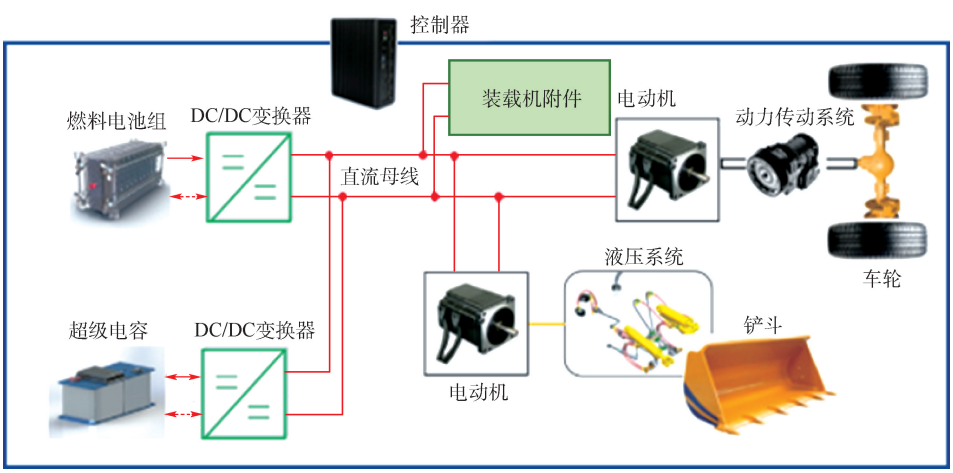


图 1 复合电源装载机动力传动方案

Fig. 1 Power transmission scheme of hybrid power loader

助电源配合燃料电池系统完成装载机各个工况需求,在装载机起步、重载、能量回收、负载变化剧烈等工况时,通过超级电容的“削峰填谷”保证电动装载机系统的高效稳定工作,并延长燃料电池系统的寿命。因此,针对不同工况需求,需要探索耦合系统的能量管理方案,首先对装载机系统进行模型建立。

1.1　燃料电池模型

燃料电池标准理论可逆电动势 E_{ref} 为

$$E_{ref} = \Delta G/(nF) \tag{1}$$

式中:ΔG 为燃料电池系统的吉布斯自由能;n 为氢气的电子数;F 为法拉第常数。

燃料电池输出电压 E_{FC} 为

$$E_{FC} = E_{ref} - E_{act} - E_{ohm} - E_{con} \tag{2}$$

式中:E_{act} 为活化过电势;E_{ohm} 为欧姆过电势;E_{con} 为浓差过电势。

燃料电池电压与电流密度变化有关,电流密度较低时,由活化极化和欧姆电压损耗引起的电压损失起主要作用,而在高电流密度时,由浓度极化引起的电压损失占主导作用[10]。图 2 为燃料电池的极化曲线,曲线分别代表阴(RHca)、阳(RHan)两极不同相对湿度下的曲线。

通过 MATLAB/Simulink 平台建立燃料电池系统模型 Simulink 模型,其结构如图 3 所示。

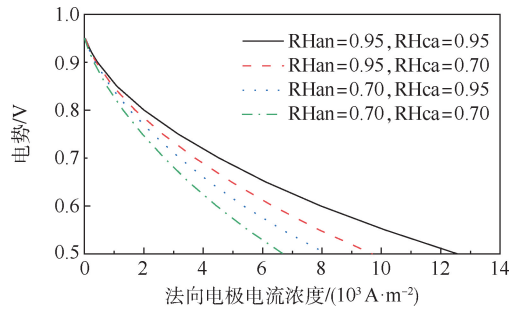


图 2　燃料电池极化曲线

Fig. 2　Polarization curves of fuel cell

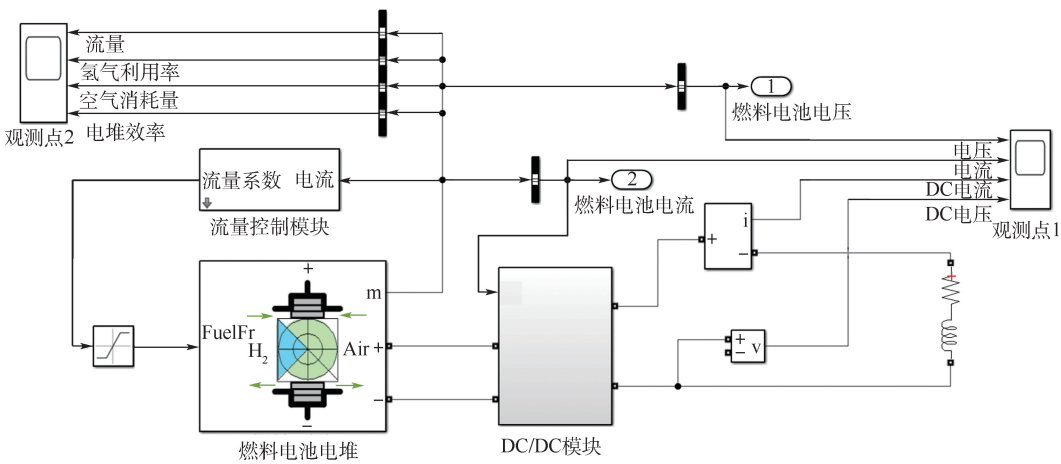


图 3　燃料电池模型结构

Fig. 3　Structure of fuel cell model

1.2　超级电容模型

电容的容量取决于两极板间存储的电荷量。图 4 为超级电容等效电路模型,主要由并联电容器 C、串联充放电电阻 R_1、并联自放电损耗电阻 R_2 组成[10]。

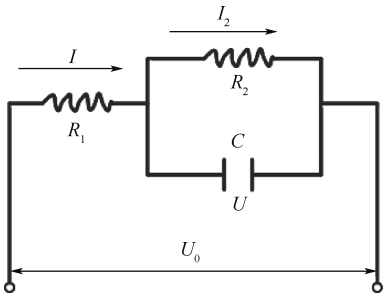


图 4　超级电容等效电路模型

Fig. 4　Equivalent circuit model of super capacitor

超级电容输出端电压 U_0、等效电容电压 U 与放电电流 I 的关系为

$$U_0 = U - IR_1 \tag{3}$$

超级电容的荷电状态(SOC)取决于输出端电压,其值为

$$\mathrm{SOC} = \frac{U_{rt} - U_{min}}{U_{max} - U_{min}} \tag{4}$$

式中:U_{rt} 为超级电容实时电压;U_{max} 和 U_{min} 分别为超级电容最大和最小工作电压。

根据以上超级电容等效电路数学公式建立如图 5 所示的 MATLAB/Simulink 模型。

1.3　其他模型

1)整车动力传动模型

装载机动力平衡方程表示为[10]

$$F_t = F_f + F_w + F_i + F_j \tag{5}$$

式中：F_t 为驱动力；F_f 为滚动阻力；F_w 为空气阻力；F_i 为坡度阻力；F_j 为惯性阻力。

2）电动机模型

建立永磁同步电机力矩平衡方程[10]：

$$T_m - T_L = J\frac{d\omega_m}{dt} + B\omega_m \tag{6}$$

式中：T_m 为电动机的电磁转矩；T_L 为电动机负载转矩；J 为电动机的转子及负载惯量；ω_m 为电动机角速度；B 为黏滞摩擦系数。

基于以上数学模型在 MATLAB/Simulink 平台建立如图 6 所示电动机模型。

3）液压系统模型

液压泵出口压力转换成泵驱动转矩，液压系统总消耗转矩 T_P 为[10]

$$T_P = \sum T_i = \sum \frac{pq}{2\pi\eta_p} \tag{7}$$

式中：T_i 为各液压泵驱动转矩；p 为各液压泵的出口压力；q 为各液压泵排量；η_p 为各液压泵机械效率。

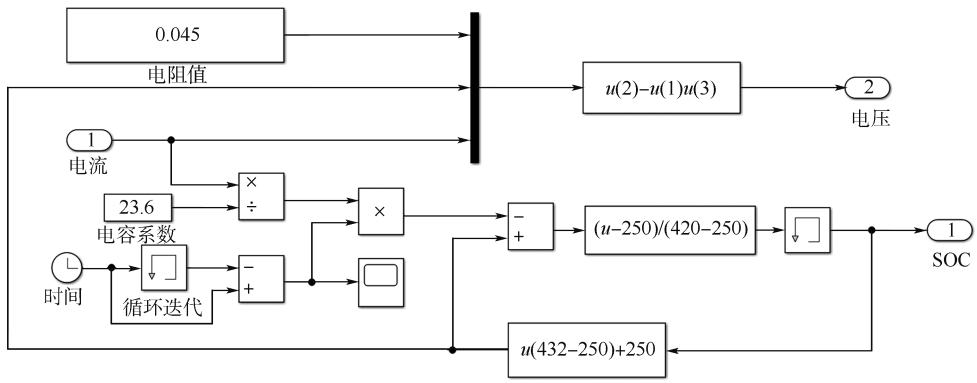


图 5 超级电容模型结构

Fig. 5 Structure of super capacitor model

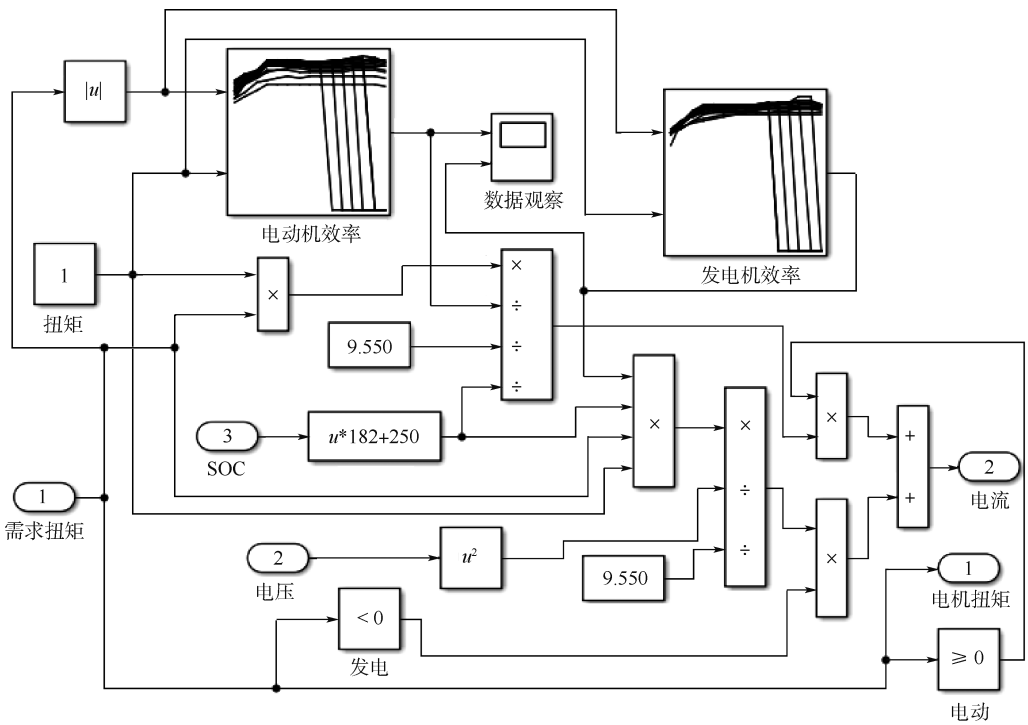


图 6 65 kW/125 kW 电动机模型结构

Fig. 6 Structure of 65 kW/125 kW motor model

2 燃料电池复合电源能量控制策略

2.1 装载机工况特性分析

图 7 为装载机作业循环过程，其工况特点是：作业循环周期短，整车频繁起停，功率变化频率与幅度大，驾驶员需要频繁加减油门、变换挡位、操控转向与工作液压系统。因此，对于燃料电池复合电源装载机循环工作过程中，输出功率、车速、载荷等相关参数频繁变化，控制系统需根据装载机的负载工况特性进行分析，并制定对应的控制策略。

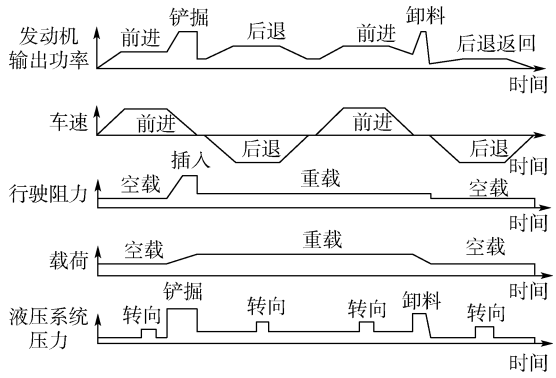


图 7　装载机工作示意图

Fig. 7　Working diagram of loader

2.2　燃料电池复合电源系统控制策略

为了满足复合电源式装载机对功率和能量的双重需求,控制系统需要将燃料电池和超级电容输出功率进行实时分配。但复合储能使得动力系统的功率与能量流变得更加错综复杂,需制定高效合理的能量管理与优化策略,解决多模式能量分流关系,才能发挥复合电源系统的性能和优势。

图 8 为控制系统工作结构。装载机复合电源能量控制系统主要工作过程如下:首先,根据复杂工况下系统多状态模型,通过模型仿真探索复杂工况对系统耦合特性。然后,基于 Haar 小波理论进行功率分配,用超级电容平衡需求功率中混沌和暂态分量,将低频分量输入到模糊逻辑控制器中进行功率分配。最后,通过 PSO 算法优化模糊逻辑隶属度函数,优化功率分配因子,实现多能源系统的最优能量控制策略。

装载机在复杂工况下,通过调节功率分配因子对燃料电池和超级电容进行功率分配与分层优化控制,分析多系统耦合特征,探索分层优化控制的机理,以满足装载机的工作需求,提高装载机能源利用率。挖掘复合电源的性能潜力,实现系统综合性能最优。

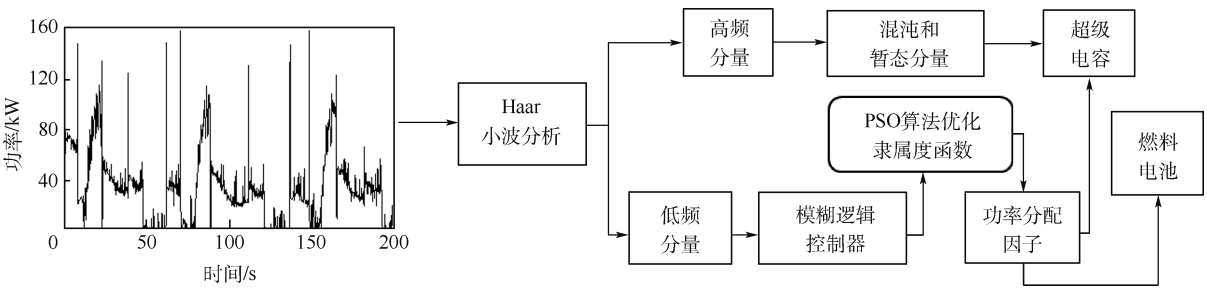


图 8　整车控制系统结构框图

Fig. 8　Structure block diagram of vehicle control system

2.3　小波功率分配

小波理论用于将给定的函数或连续时间信号分成不同尺度的分量,该技术已经成功应用于多数字信号处理和压缩领域。装载机工作过程中频繁处于功率波动、启停变化、怠速和大功率运行状态。剧烈的功率波动和频繁的启停变化是燃料电池性能衰退的关键影响因素[11-12]。因此,在复合电源系统中将需求功率进行小波分配,避免燃料电池应对一些混沌和暂态功率分量,提高燃料电池的耐久性和复合电源系统的可靠性。

Haar 小波与其他小波相比在时域上具有最短的滤波长度,非常适合装载机负载功率频繁变化的特点。本文使用 Haar 小波变换策略提取系统需求功率暂态信号。Haar 函数 $\psi(x)$ 为

$$\psi(x)=\begin{cases}1 & x\in[0,0.5)\\ -1 & x\in[0.5,1)\\ 0 & \text{其他}\end{cases}\tag{8}$$

式中:x 为时间。

本文将装载机实际载荷谱曲线加载到 Haar 小波变换策略中,将负载功率曲线分解为高频和低频部分,燃料电池系统主要负责低频功率部分,高频分量及制动能量回收部分由超级电容负责。考虑计算效率和频率的限制,设置 3 层小波分解,解压缩和重构滤波器分别为 $[H_1(z),H_0(z)]^{\mathrm{T}}$ 和 $[G_1(z),G_0(z)]^{\mathrm{T}}$。输入信号 $x(t)$ 采用 3 级 Haar 小波分解和重构。如图 9 所示,分解和重构的过程分别采用下采样和上采样的方法,$H_0(z)$ 与 $G_1(z)$ 为低通滤波器,$H_1(z)$ 与 $G_0(z)$ 为高通滤波器。原始信号 $x(n)$ 等于装载机负载功率谱的预定功率需求。

将装载机实车功率载荷谱作为原始信号 $x(n)$。经过 3 层 Haar 小波变换处理,原始信号被划分为 $x_0(n)$、$x_1(n)$、$x_2(n)$、$x_3(n)$来描述频率的特征。低频分量 P_{LC} 和高频分量 P_{HC} 分配表达式为

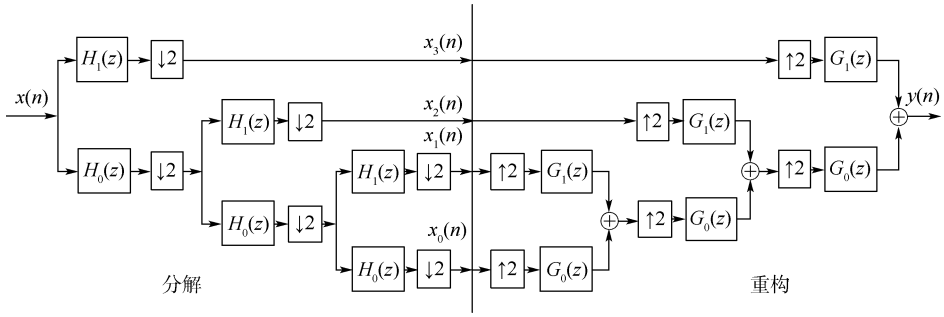


图 9 Haar 小波信号处理

Fig. 9 Haar wavelet signal processing

$$P_{\mathrm{LC}}=\begin{cases}x_0(n) & x_i(n)>0\\ 0 & \text{其他}\end{cases} \tag{9}$$

$$P_{\mathrm{HC}}=\begin{cases}x_1(n)+x_2(n)+x_3(n) & x_i(n)>0\\ x_0(n)+x_1(n)+x_2(n)+x_3(n)\end{cases} \tag{10}$$

低频分量为需求功率中稳态分量,将其作为模糊逻辑控制器的输入变量,而高频分量是需求功率中混沌和暂态分量,由超级电容平衡。

2.4 模糊逻辑控制策略

本文的控制策略目标是:通过调节功率分配因子满足装载机工况需求,使复合电源与负载更好地匹配,系统各元件高效工作,提高整机效率和燃料电池系统的寿命。复合电源装载机控制系统的特点是:动力驱动系统输出功率非线性突出,工作环境的复杂多变导致控制参数突变与信号失真,负载载荷具有很强的鲁棒性,控制单元复杂繁琐,因此很难建立精确的数学模型。模糊逻辑控制策略把人的经验控制用计算机实现,基于人的控制经验建立控制系统的知识模型,定义模糊语言变量,确定模糊集合,设计模糊逻辑推理程序,实现高鲁棒性自适应的非线性控制[13-14]。本文基于模糊逻辑理论研究复合电源装载机系统控制策略,通过实时采集负载载荷数据与超级电容 SOC 值,控制燃料电池与超级电容系统工作点。

根据装载机工况特征对其负载特性及能量流向进行分析,实现工况划分并制定相应的工作模式[5]。根据装载机、燃料电池和超级电容的特征及工况需求,划分 4 种工作模式来适配混合动力能量源[15]。控制器工作模式分为以下 4 种模式[10]。

1)模式 1。起步,超级电容独立驱动车辆。

$$P_{\mathrm{m}}=\frac{P_{\mathrm{dem}}}{\eta_{\mathrm{m}}} \tag{11}$$

$$P_{\mathrm{FC}}=0 \tag{12}$$

$$P_{\mathrm{SC}}=P_{\mathrm{m}} \tag{13}$$

式中:P_{dem}为装载机需求功率;P_{FC}为燃料电池系统的输出功率;P_{m}为电动机的输出功率;P_{SC}为超级电容输出功率;η_{m}为电动机输出效率。

2)模式 2。燃料电池系统独立驱动车辆,燃料电池系统在自己约束条件内满足需求功率并同时平衡超级电容 SOC 值。

$$P_{\mathrm{SC}}=\frac{\mathrm{SOC}-\mathrm{SOC_e}}{0.5(\mathrm{SOC_h}-\mathrm{SOC_l})}P_{\mathrm{C}} \tag{14}$$

$$P_{\mathrm{FC}}\eta_{\mathrm{DC}}=P_{\mathrm{m}}-P_{\mathrm{SC}} \tag{15}$$

$$\mathrm{SOC_e}=0.5(\mathrm{SOC_h}+\mathrm{SOC_l}) \tag{16}$$

式中:SOC、$\mathrm{SOC_e}$、$\mathrm{SOC_h}$、$\mathrm{SOC_l}$分别为超级电容的荷电状态、荷电状态期望值、设置的超级电容荷电状态的最大、最小值;P_{C}为超级电容 SOC 的调节功率;η_{DC}为燃料电池 DC/DC 变换器效率。

3)模式 3。超级电容和燃料电池联合驱动,若超级电容的 SOC 值大于期望值,超级电容参与工作并提供较大功率,剩余需求功率由燃料电池系统提供。若需求功率较大,单纯燃料电池无法满足要求,需要超级电容放电以平衡负载需求。

$$P_{\mathrm{FC}}=P_{\mathrm{FC\text{-}max}} \tag{17}$$

$$P_{\mathrm{SC}}=P_{\mathrm{m}}-P_{\mathrm{FC}}\eta_{\mathrm{DC}} \tag{18}$$

式中:$P_{\mathrm{FC\text{-}max}}$为燃料电池最大输出功率。

4)模式 4。制动回收模式,在车辆减速刹车或降臂时,电动机以发电模式工作,超级电容吸收能量而燃料电池系统不工作。

复合电源系统模糊逻辑控制结构如图 10 所示,采用功率控制,控制器的输入变量是系统需求

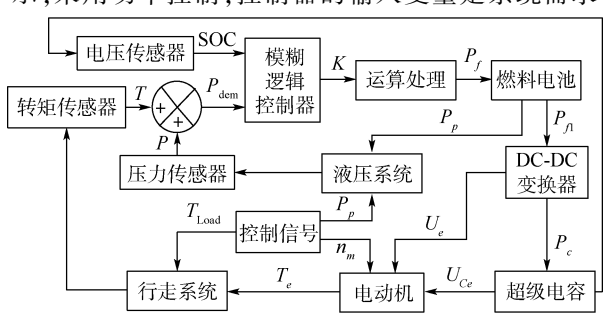


图 10 复合电源系统模糊逻辑控制结构

Fig. 10 Fuzzy logic control structure of composite power supply system

功率 P_{dem} 与超级电容 SOC 值,输出为功率分配因子 K。需求功率为工作液压系统和行走系统驱动功率之和,SOC 值为超级电容的荷电状态[10]。

在模糊逻辑控制系统中,首先设定输入与输出参数的基本论域,然后设计模糊逻辑控制器并制定控制规则,模糊逻辑控制器如图 11 所示。

在该模糊逻辑控制器中,输入变量 1 为小波处理后整车需求功率 P_{dem},设置模糊论域为[0,160],P_{dem} 的模糊子集划分为{VL,L,A,H,VH};输入变量 2 为超级电容 SOC 值,设置模糊论域为[0,1],超级电容 SOC 的模糊子集划分为{vlow,low,mid,high,vhigh};输出变量燃料电池系统功率 P_{FC} 设置模糊论域为[10,100],燃料电池 P_{FC} 模糊子集为{L1,L2,L3,L4,L5}。根据不同工况制定模糊规则集,建立隶属度函数,根据专家库制定控制规则,图 12 为对应的调节规则。

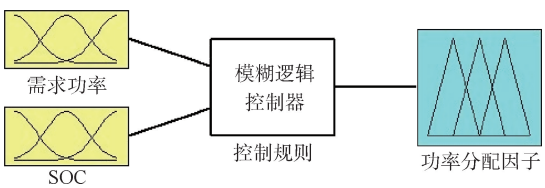


图 11　模糊逻辑控制器

Fig. 11　Fuzzy logic controller

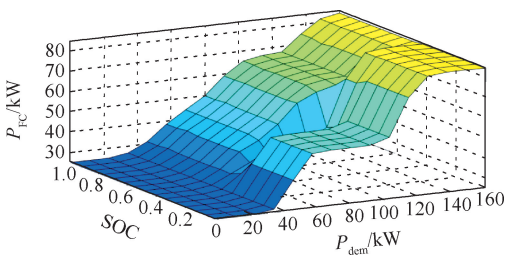


图 12　P_{dem}、SOC、P_{FC} 对应的调节规则

Fig. 12　Regulation rules of P_{dem}, SOC and P_{FC}

2.5　小波变换-模糊逻辑控制器

基于小波变换算法分离高低频需求功率,基于 MATLAB 代码来确定最优的 3 层分解阈值[16],对应 d_1、d_2、d_3 最优值为 10.664 5、14.474 8、6.813 0;3 层 Haar 小波处理后对应的负载功率曲线如图 13 所示,$s/ds = a_3 + d_3 + d_2 + d_1$,其曲线为原功率,$a_3$ 代表的曲线为分解后的低频功率曲线,d_1、d_2、d_3 代表超级电容负责的高频功率部分[10]。

经过最优阈值 3 层 Haar 小波处理后,将需求功率进行分配,一些混沌和暂态分量用超级电容平衡,有效提升燃料电池系统的寿命与整车经济性。图 14 和图 15 为 3 层 Haar 小波处理后的装载机典型工况下功率分流结构。该工况是以某 5 t轮式装载机(ZL50)为例,分析其对原生土进行作业时的载荷谱数据,控制器通过各传感器实时采集车速、动力传动轴转矩、工作液压泵出口压力及流量、转向液压泵出口压力及流量等数据。由图 14 和图 15 可见,分解后的燃料电池功率需求

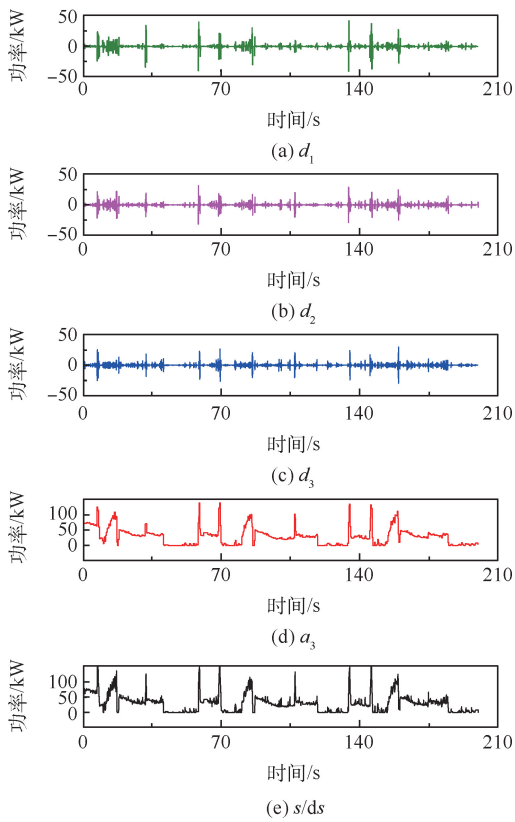


图 13　三层 Haar 小波处理载荷谱曲线

Fig. 13　Three-layer wavelet processing of load spectrum curves

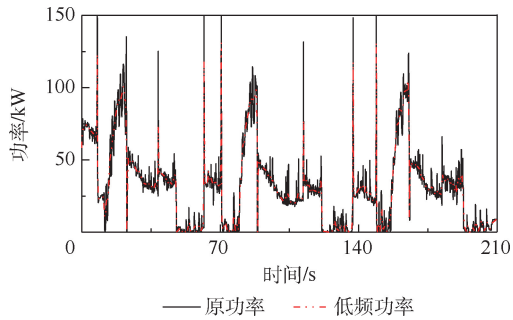


图 14　动力系统负载功率曲线

Fig. 14　Load power curves of power system

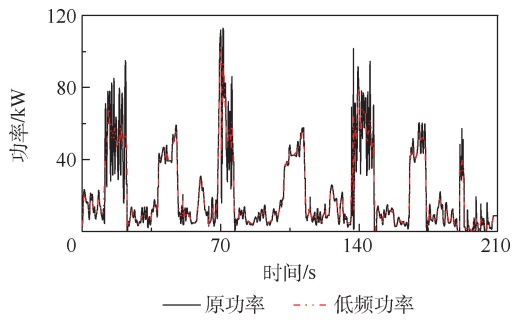


图 15　液压系统负载功率曲线

Fig. 15　Load power curves of hydraulic system

曲线具有全局低频性,将低频分量输入到模糊逻辑控制器中。

将实际载荷谱数据低频分量加载到 Simulink 模型中,模型主要包括燃料电池系统、超级电容系统、发电机、电动机、工作液压系统、控制模块等。建立复合电源装载机系统仿真模型,研究系统动力耦合特性,将复合电源控制策略集成到整车控制层中,挖掘复合电源的性能潜力。其仿真模型如图 16 所示。

将载荷谱功率数据作为模型输入,通过一个周期的对比仿真分析。图 17 为超级电容 SOC 值变化曲线,图 18 为燃料电池输出功率。对比可知,载荷功率经过 3 层 Haar 小波变换后,超级电容 SOC 值变化更加频繁,燃料电池功率变化幅度减缓,兼顾经济性的同时提高燃料电池寿命。因为超级电容在控制过程中平衡负载功率的暂态与混沌分量,导致 SOC 值变化幅度相对剧烈,但在 0.6 左右处于动态平衡状态,Haar 小波变换后燃料电池输出功率更加平滑,图中局部放大处波峰功率下降了 4.14 kW,因此,小波变换提高了系统综合效率,控制策略更合理。

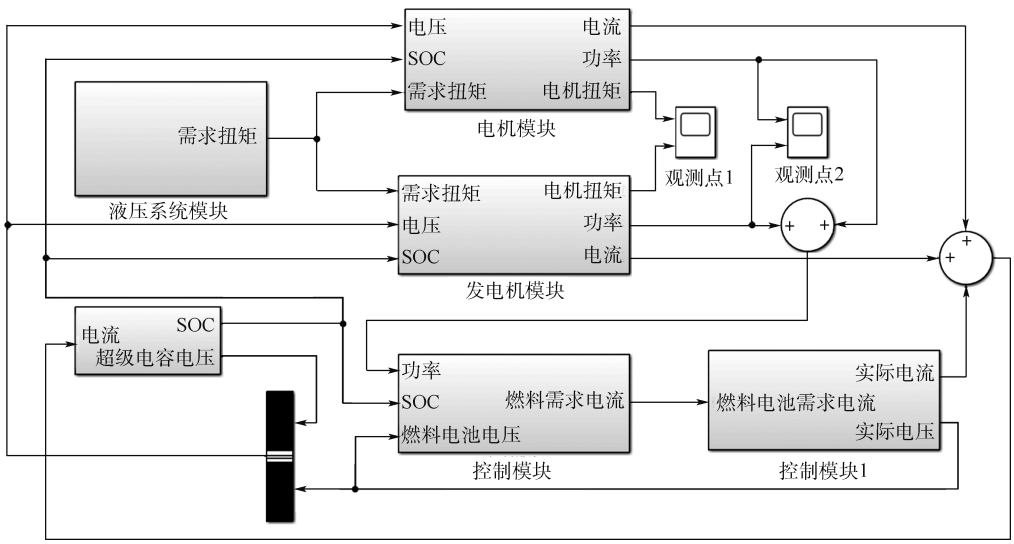


图 16 复合电源系统仿真模型

Fig. 16 Simulation model of composite power supply system

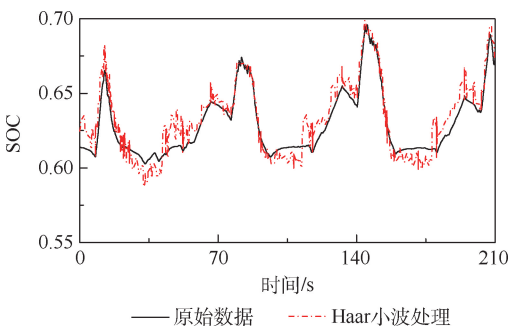


图 17 超级电容 SOC 变化曲线

Fig. 17 SOC curves of super capacitor

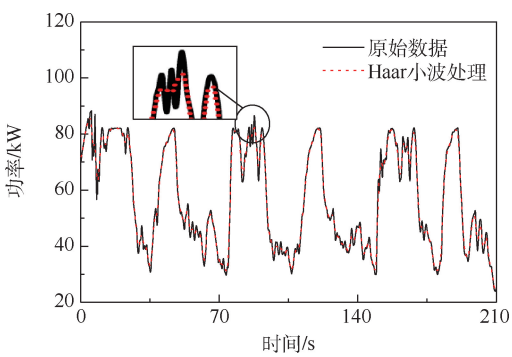


图 18 燃料电池输出功率曲线

Fig. 18 Output power curves of fuel cell

3 PSO 算法优化模糊逻辑隶属度函数

由于模糊逻辑推理过程依靠丰富的专家经验,需要反复训练才得到最佳控制模式,PSO 算法采用简单速度-位置更新公式迭代,粒子通过交换自己的成功经验信息来寻求最优解[17-18]。为适合复合电源装载机系统工况特性,利用 PSO 算法优化模糊逻辑隶属度函数,实现复杂工况下系统功率分配及优化控制策略,提高系统综合性能。

3.1 PSO 算法原理

首先,在一个设定的 D 维空间中进行目标搜索,搜索粒子群落由 N 个粒子构成,其中第 i 个粒子表示为 D 维空间的向量。第 i 个粒子的“飞行”速度也是一个 D 维向量,第 i 个粒子迄今为止搜索到的最优位置称为个体极值,最优位置是由整个粒子群到此为止搜索到的,记为全局极值,找到全局最优位置和个体极值最优位置时,粒子在这时会根据如下公式更新速度和位置[19]。PSO 算法的具体流程如图 19 所示。

$$\begin{cases} v_i(t+1) = \omega v_i(t) + c_1 r_1 (p_i(t) - x_i(t)) + \\ \qquad c_2 r_2(t)(p_g(t) - x_i(t)) \\ x_i(t+1) = x_i(t) + v_i(t+1) \end{cases} \tag{19}$$

式中:t 为当前迭代次数;r_1、r_2 表示(0,1)之间的随机数;c_1、c_2 为加速因子,一般在(0,4)之间取值;ω 为惯性权重;x_i 为粒子位置;v_i 为粒子速度函数;p_i 为个体极值;p_g 为全体极值。

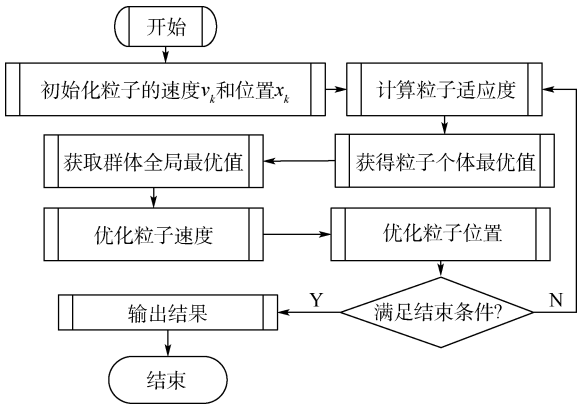


图 19　PSO 算法优化流程

Fig. 19　Flow chart of particle swarm optimization algorithm

3.2　PSO 算法优化隶属度函数

模型以装载机典型的工作周期实际载荷功率作为模糊逻辑控制器的输入之一。本文使用已知控制规则优化隶属函数的优化方法[20],在仿真系统中代入实际的模糊控制器。超级电容的 SOC 值应该保持在一定的范围内,约束条件如下:

$$\mathrm{SOC}_{\max} \geqslant \mathrm{SOC}(t) \geqslant \mathrm{SOC}_{\min} \tag{20}$$

复合电源中放电功率约束为:在装载机行驶过程中,复合电源系统的放电功率要受到其自身功率的限制,约束条件如下:

$$P_{\mathrm{FC\text{-}max}} \geqslant P_{\mathrm{FC}}(t) \geqslant P_{\mathrm{FC\text{-}min}} \tag{21}$$

$$P_{\mathrm{SC\text{-}max}} \geqslant P_{\mathrm{SC}}(t) \geqslant P_{\mathrm{SC\text{-}min}} \tag{22}$$

粒子群优化目标是降低燃料电池汽车在整个运行工况的总等效氢耗,即

$$M = \min(M_{\mathrm{FC}} + M_{\mathrm{SC}}) \tag{23}$$

式中:M_{FC}为燃料电池耗氢量;M_{SC}当超级电容等效耗氢量。

将输入输出中所有的隶属度函数进行优化,隶属函数通过 PSO 优化算法迭代了 474 次后保持不变,并达到最大适应度 1。将输入输出中所有的隶属函数优化后代入到新的模糊逻辑控制器中进行观测。优化前后模糊逻辑规则对比观察器如图 20 所示,选则观测器片段,提取 P_{dem} = 106 kW,超级电容 SOC = 0.659 时刻数据,通过优化后燃料电池输出功率为 P_{FC} = 58.3 kW。与原观察器相比,燃料电池输出功率下降 0.7 kW。

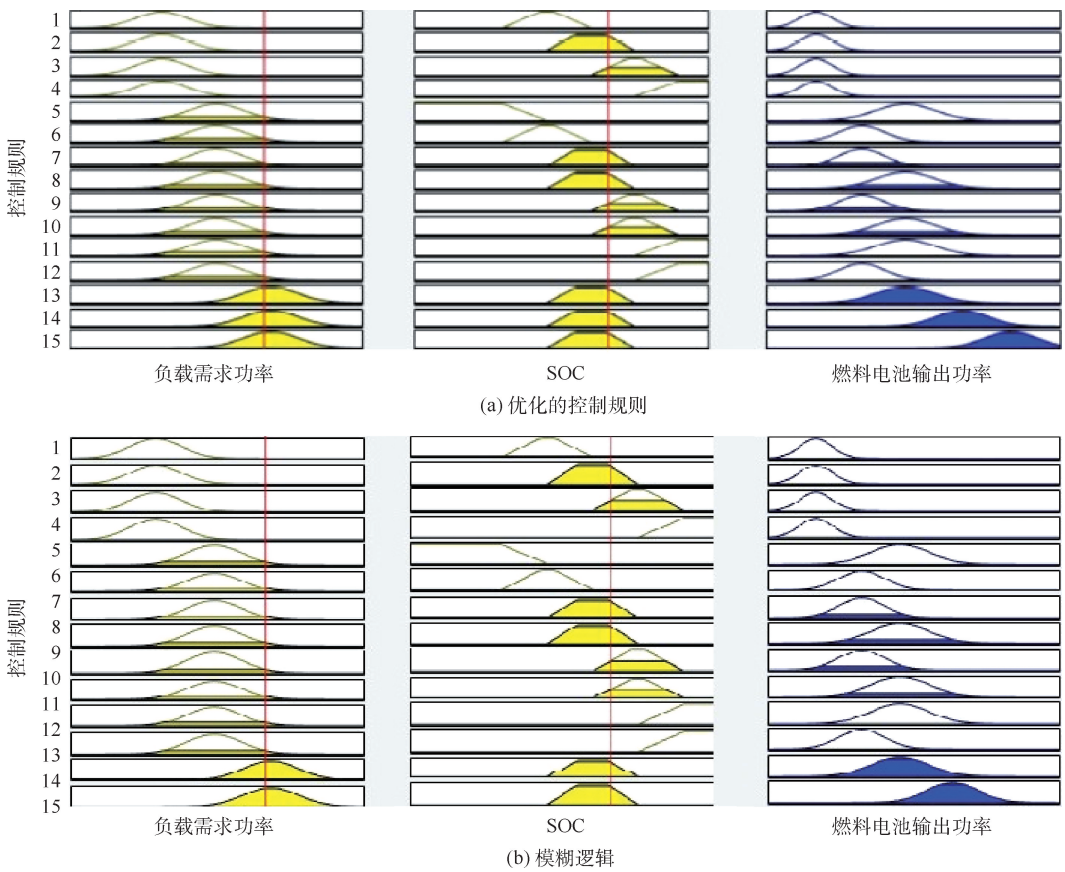


图 20　优化前后的模糊逻辑规则观察器(部分)

Fig. 20　Rule observer before and after fuzzy logic controller optimization (part)

4 PSO 算法优化-模糊逻辑控制器仿真

以装载机同样工况周期的实际载荷功率曲线作为模糊逻辑控制器的输入,按照 PSO 算法优化的隶属度函数建立控制规则,通过控制模式对比仿真分析。

经过 PSO 算法优化控制器后,燃料电池输出功率对比曲线如图 21 所示。Haar 小波处理代表通过 3 层 Haar 小波处理需求功率后进行模糊逻辑控制结果,PSO 算法优化代表通过小波处理后并对模糊逻辑隶属度函数进行粒子群优化后的控制结果。在“波峰波谷”处清晰看到燃料电池输出功率波动较少,变化平缓,在满足装载机系统动力性要求的前提下,燃料电池输出平均功率同比下降 5% 左右,有效提高燃料电池的效率。因此,通过 PSO 算法优化模糊逻辑控制器对复合电源系统功率分配因子进行修正,有效平衡燃料电池系统在快速变载、频繁启停、过载等恶劣工况下的功率输出,提高燃料电池的使用寿命与系统的综合性能。满足装载机系统的动态工况需求,保证燃料电池工作状态变化更加平稳,兼顾系统的耐久性和经济性。

如图 22 所示,经过优化后,选择在某单循环

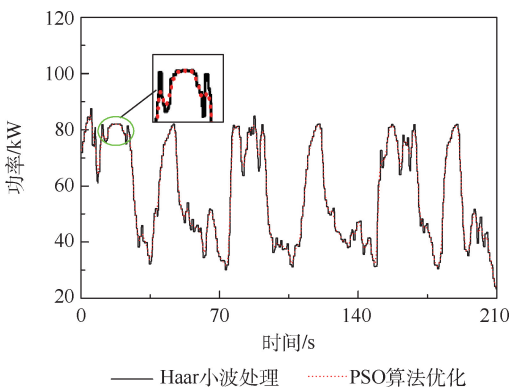


图 21 优化后燃料电池输出功率变化曲线

Fig. 21 Output power curves of optimized fuel cell

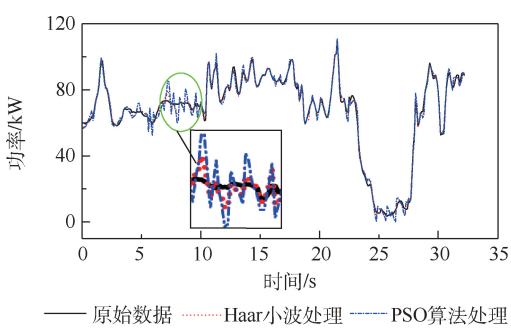


图 22 优化后超级电容输出功率变化曲线

Fig. 22 Output power curves of optimized super capacitor

工况中重载频繁工况(35 s)下研究超级电容平衡功率数据。通过 3 种控制模式进行对比仿真分析,图中原始数据代表未处理载荷谱数据进行模糊逻辑控制,Haar 小波处理代表通过 3 层 Haar 小波处理需求功率后进行模糊逻辑控制,PSO 算法优化代表通过小波处理并对模糊逻辑隶属度函数进行优化后的控制。超级电容输出功率中,原始数据曲线变化相对平滑,但 Haar 小波处理后,曲线变化幅度较大,PSO 算法优化后,曲线变化频率与变化幅度最大。因此,在复合电源系统中,为了提高燃料电池的综合效率,利用超级电容吸收负载功率中混沌和暂态分量。

图 23 为超级电容在全周期工况下 Haar 小波处理与 PSO 算法优化后的 SOC 值变化曲线对比。优化后曲线变化幅度更剧烈,但超级电容 SOC 值保持在 0.6 左右,说明超级电容能够满足工况需求,优化后的控制器效果明显。

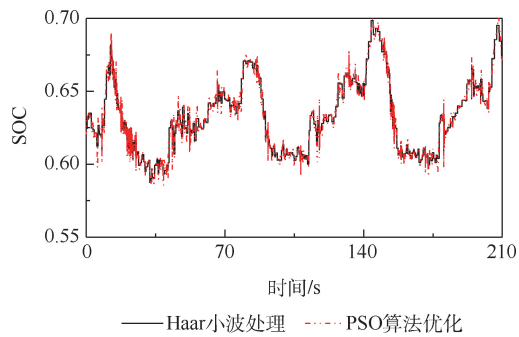


图 23 优化后超级电容 SOC 变化曲线

Fig. 23 SOC curves of optimized supercapacitor

5 结 论

本文围绕复杂工况下燃料电池和超级电容复合电源式装载机进行控制策略研究,分析多系统耦合特征,探索分层优化控制的机理。首先,建立燃料电池复合电源装载机复杂工况下系统多状态模型;其次,基于 Haar 小波理论分层处理载荷功率;然后,基于模糊逻辑控制策略探索复杂工况对系统耦合特性的影响规律;最后,基于 PSO 算法优化隶属度函数,探索复杂工况下系统功率分配及优化控制策略。通过仿真数据验证结论如下:

1) 经过最优阈值 3 层 Haar 小波处理后,将需求功率进行分配,一些混沌和暂态分量用超级电容平衡,避免燃料电池处理急速变化的功率需求,有效提升燃料电池系统寿命与整车经济性。

2) 提出模糊逻辑能量管理策略动态平衡需求功率,设计模糊逻辑控制器,制定控制规则,仿真结果显示,燃料电池功率变化平滑,提高了电源

系统的效率,超级电容工作在设定区域内,工作功率变化幅度相对剧烈,控制策略合理。

3) 经过 PSO 算法优化控制器后,燃料电池输出功率变化幅度变得更平缓,燃料电池平均功率同比下降 5% 左右。有效提高燃料电池的效率与耐久性。

参考文献 (References)

[1] LI T Y, HUANG L T, LIU H Y. Energy management and economic analysis for a fuel cell supercapacitor excavator[J]. Energy, 2019, 172: 840-851.

[2] DE MIRANDA P E V, CARREIRA E S, ICARDI U A, et al. Brazilian hybrid electric-hydrogen fuel cell bus improved on-board energy management system[J]. International Journal of Hydrogen Energy, 2017, 42(19): 13949-13959.

[3] ZHANG X, MI C, 等. 车辆能量管理:建模、控制与优化[M]. 张希, 米春亭, 译. 北京:机械工业出版社, 2013.
ZHANG X, MI C, et al. Vehicle energy management: Modeling, control and optimization[M]. ZHANG X, MI C T, translated. Beijing: China Machine Press, 2013 (in Chinese).

[4] LI T Y, LIU H Y, WANG H, et al. Hierarchical predictive control-based economic energy management for fuel cell hybrid construction vehicles[J]. Energy, 2020, 198: 117327.

[5] MUNOZ P M, CORREA G, GAUDIANO M E, et al. Energy management control design for fuel cell hybrid electric vehicles using neural networks[J]. International Journal of Hydrogen Energy, 2017, 42(48): 28932-28944.

[6] SNOUSSI J, BEN ELGHALI S, BENBOUZID M, et al. Auto-adaptive filtering-based energy management strategy for fuel cell hybrid electric vehicles[J]. Energies, 2018, 11(8): 1-20.

[7] IBRAHIM M, JEMEI S, WIMMER G, et al. Non cinear autoregressive neural network in an energy management strategy for battery lultra-capacitor hybrid electrical vehicles[J]. Electric Power Systems Research, 2016, 136: 262-269.

[8] ZHANG R D, TAO J L, ZHOU H Y. Fuzzy optimal energy management for fuel cell and super capacitor systems using neural network based driving pattern recognition[J]. IEEE Transactions on Fuzzy Systems, 2019, 27(1): 45-57.

[9] HE D, SHI Y, SONG X. Weight-free multi-objective predictive cruise control of autonomous vehicles in integrated perturbation analysis and sequential quadratic programming optimization framework[J]. Dynamic System Measurement Control, 141, 9: 91015.

[10] 杜文杰. 基于燃料电池复合双电源装载机系统功率控制研究[D]. 太原:中北大学, 2020.
DU W J. Research on power control of fuel cell compound dual energy source system[D]. Taiyuan: North University of China, 2020 (in Chinese).

[11] 吕沁阳, 滕腾, 张宝迪, 等. 增程式燃料电池车经济性与耐久性优化控制策略[J]. 哈尔滨工业大学学报, 2021, 53(7): 126-133.
LV Q Y, TENG T, ZHANG B D, et al. Optimal control strategy for economy and durability of extended range fuel cell vehicle[J]. Journal of Harbin Institute of Technology, 2021, 53(7): 126-133 (in Chinese).

[12] 胡尊严. 车用燃料电池系统耐久性建模与状态估计研究[D]. 北京:清华大学, 2019: 3-7.
HU Z Y. Durability modeling and state estimation for vehicular fuel cell system[D]. Beijing: Tsinghua University, 2019: 3-7 (in Chinese).

[13] SULAIMAN N, HANNAN M A, MOHAMED A, et al. Optimization of energy management system for fuel-cell hybrid electric vehicles: Issues and recommendations[J]. Applied Energy, 2018, 228: 2061-2079.

[14] 崔宁, 秦四成, 赵丁选. 液压挖掘机并联混合节能动力系统多目标优化控制策略[J]. 西安交通大学学报, 2016, 50(6): 116-121.
CUI N, QIN S C, ZHAO D X. A multi object optimal control strategy for a parallel hybrid power system in hydraulic excavators[J]. Journal of Xi'an Jiaotong University, 2016, 50(6): 116-121 (in Chinese).

[15] SONG K, DING Y H, HU X, et al. Degradation adaptive energy management strategy using fuel cell state-of-health for fuel economy improvement of hybrid electric vehicle[J]. Applied Energy, 2021, 285: 1-12.

[16] LI T Y, LIU H Y, WANG H, et al. Multiobjective optimal predictive energy management for fuel cell/battery hybrid construction vehicles[J]. IEEE Access, 2020, 8: 25927-25937.

[17] FAISAL M, HANNAN M A, KER P J, et al. Particle swarm optimised fuzzy controller for charging-discharging and scheduling of battery energy storage system in MG applications[J]. Energy Reports, 2020, 6: 215-228.

[18] AFZAL A, RAMIS M K. Multi-objective optimization of thermal performance in battery system using genetic and particle swarm algorithm combined with fuzzy logics[J]. Journal of Energy Storage, 2020, 32: 101815.

[19] LI J, ZHOU Q, WILLIAMS H, et al. Back-to-back competitive learning mechanism for fuzzy logic based supervisory control system of hybrid electric vehicles[J]. IEEE Transactions on Industrial Electronics, 2020, 67(10): 8900-8909.

[20] 吕柏权, 张静静, 李占培, 等. 基于变换函数与填充函数的模糊粒子群优化算法[J]. 自动化学报, 2018, 44(1): 74-86.
LV B Q, ZHANG J J, LI Z P, et al. Fuzzy partical swarm optimization based on filled function and transformation function[J]. Acta Automatica Sinica, 2018, 44(1): 74-86 (in Chinese).

Layered control of hybrid power loader based on fuel cell

ZHANG Zhiwen[1,2,*], DU Wenjie[1], LIANG Junfei[1], ZHANG Yangang[1], WU Yawen[1]

(1. School of Energy and Power Engineering, North University of China, Taiyuan 030051, China;
2. School of Vehicle and Energy, Yanshan University, Qinhuangdao 066004, China)

Abstract: It is of great significance to study new energy technologies for loaders, which have high energy consumption and poor emissions. Combined with the operating characteristics of the loader, this paper proposes a power supply system driven by fuel cell and super capacitor. Our research focused on adaptive energy management strategy resulting from dynamic model and real-time data of fuel cell and supercapacitor system under complex working conditions. Firstly, we designed a composite power supply topology and power transmission scheme. Then, a multi-state model of the system was established under complex working conditions of the loader. And, based on the Haar wavelet theory, the power of the vehicle system was split. Subsequently, a fuzzy logic energy management strategy was proposed to dynamically balance the low-frequency components of the demand power. Lastly, the particle swarm optimization algorithm was used to optimize the control system. The simulation results showed that the power change was greatly slowed down because of the optimal threshold three-layer Haar wavelet on the load power, which effectively improved the life of the fuel cell system. The fuel cell power curve output by the fuzzy logic controller also changed smoothly. Meanwhile, the SOC value of the super capacitor was within the set area. Therefore, the overall efficiency of the composite power system was improved. After optimizing the controller by the particle swarm algorithm, the average output power of the fuel cell decreased by about 5% year-on-year, and the SOC value of the super capacitor reaches a dynamic equilibrium state of about 0.6, which improves the dynamic response and stability of the loader.

Keywords: composite power supply; loader; multi state model; Haar wavelet; fuzzy logic; particle swarm optimization algorithm

Received: 2021-03-02; **Accepted**: 2021-08-23; **Published online**: 2021-09-28 08:55

URL: kns.cnki.net/kcms/detail/11.2625.V.20210924.2317.002.html

Foundation items: National Natural Science Foundation of China (51605447); Applied Basic Research Programs of Shanxi (201901D211208); Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2019L0605)

* **Corresponding author**. E-mail: zhzhw666@nuc.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0682

基于 CC1310 芯片的多通道高速率低功耗无线传感系统

段瑞枫[1,2,*], 吕燕洁[3], 杜文基[4,5], 周游[6]

(1. 北京林业大学 信息学院, 北京 100083; 2. 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083;
3. 北京航空航天大学 电子信息工程学院, 北京 100191; 4. 中国科学院计算机网络信息中心, 北京 100190;
5. 中国科学院大学 计算机科学与技术学院, 北京 100049; 6. 芯擎科技有限公司, 北京 100012)

摘 要: 为有效减轻新一代运载火箭传感器数据采集与传输系统的质量,设计并实现了基于 CC1310 芯片的无线传感器系统。采用频分复用(FDM)结合时分复用(TDM)的方式完成多节点组网并实施分组管理,组间频分复用既实现了节点数的扩增,又提升了传输速率复用倍数,分组数为 4 时子节点数量可达 100 个以上。提出主节点授时法结合多节点分时传输协议的优化设计方法,保证多节点高精准同步,避免节点间碰撞,获得了最优的组内可达速率;设计节点唤醒/休眠模式切换策略,有效降低了系统功耗。实测结果表明:2 个主节点带 5 个子节点并行工作时,传输速率可达 400 Kbps,且主节点数量增加时,系统的传输速率成比例增大;单个子节点忙时功耗不超过 60 mW,闲时功耗不超过 12 mW,平均功耗为 15.2 mW,符合低功耗要求;同时,所设计的无线传感系统具备良好的可靠性和鲁棒性。

关 键 词: CC1310; 无线组网通信; 时间同步; 高速率; 低功耗
中图分类号: TN914.5; TN924
文献标志码: A **文章编号**: 1001-5965(2022)11-2177-09

20 世纪 90 年代开始,无线传感网络技术得到快速发展,微小传感器技术和节点间的无线通信能力为传感器网络赋予了广阔的应用前景,主要应用在军事、环境、健康、家庭及其他一些感知监测区域[1-3]。在航空航天领域,当运载火箭飞行时,需要实时获取温度、湿度、振动等相关参数,传统的电缆数据传输中,数据采集区域宽,占用空间大,运载火箭工作环境严峻,在发射过程中线路极易损坏,导致测量数据丢失,因此电缆设计和保护非常复杂。采用无线传感网络时,节点间通信不会占用箭体的内部空间,节点间依赖减小,减轻了运载火箭电气系统的质量,降低了发射成本。因此无线传感网络技术在航空航天领域的应用正逐渐成为研究热点[4-5],在保证信息可靠采集的情况下,如何提高无线系统的传输速率是关键。

无线传感网络是一种分布式系统,协议的运行、调度、数据传输等均建立在精确的时间同步基础上[6-7],节点数量也是衡量系统性能的指标之一[8]。节点过少时,不能满足运载火箭多监测量的数据获取;节点过多时,相互独立的各节点协同工作的难度大幅增加,时间同步极具挑战性。传统的参考广播同步算法(reference broadcast synchronization,RBS)利用自身的本地时间与参考节点的时钟差值进行自我调节,但难以满足大数量

收稿日期: 2021-11-11; **录用日期**: 2022-02-13; **网络出版时间**: 2022-03-09 09:30
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220307.1346.004.html
基金项目: 中央高校基本科研业务费专项资金(BLX201623); 北京市自然科学基金(L202003); 国家自然科学基金(31700479)
* **通信作者**. E-mail: drffighting2008@163.com

引用格式: 段瑞枫, 吕燕洁, 杜文基, 等. 基于 CC1310 芯片的多通道高速率低功耗无线传感系统[J]. 北京航空航天大学学报, 2022, 48(11): 2177-2185. DUAN R F, LYU Y J, DU W J, et al. Multi-channel wireless sensor system based on CC1310 chip with high speed and low power consumption[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2177-2185(in Chinese).

节点的同步;能量有效参考广播同步算法(energy-efficient RBS,ERBS)对 RBS 进行了改进,但当无线传感网络环境发生改变时会产生较大的计时误差[9-10]。因此本文将研究多节点时间同步的问题。

无线传感网络可靠工作依赖于高效的传输协议,目前主流协议有 ZigBee 协议、LoRa 协议等。ZigBee 协议由于发射功率较小,因此具备低功耗的优点,但是也降低了传输速率,其最大传输速率为 250 Kbps,不满足本系统高速率的需求[11-12]。LoRa 协议在相同功耗下比其他无线传输方式传播距离更远,但是其成本偏高,占用带宽大,抗干扰能力弱,无法满足航空系统中的低成本可靠性的传输需求[13]。此外,这些协议均为通用协议,不具有保密性,不适合航空航天领域的应用,本文将研究适用于运载火箭的无线传输协议。

部署于箭体上的传感器子节点无法外接电源,需采用电池供电,因此低功耗对于本文研究的无线传感系统尤为重要,可以从硬件平台及软件传输协议的设计来实现低功耗[14]。文献[15]即从微带电路与软件协议两方面进行设计,通过轮询协议和防碰撞算法实现了功耗的降低,但电路尺寸较大且不适用于大量子节点的无线通信系统。文献[16]设计了基于星型网络拓扑结构的无线组网方式,提出了基于节点模式切换的低功耗重传方案,结果表明 3 个子节点工作时平均功耗为 3.2 mW,达到了低功耗,但并未优化传输速率。本文基于低功耗芯片 CC1310,设计节点模式切换方法来降低系统功耗。

综上所述,本文设计了一个适用于航天运载火箭的无线传感系统。基于 CC1310 芯片,提出了组间频分复用(frequency division multiplexing, FDM)结合组内时分复用(time division multiplexing, TDM)的组网方式,极大增加了系统可支持的节点数,即通道数;多个频段同时传输也提升了系统的速率复用倍数;设计主节点授时的同步方案,实现了多节点高精准同步,保证系统可靠性;设计一种多节点分时传输协议,避免节点间碰撞,提高了组内传输速率;采用节点唤醒/休眠模式切换策略,极大地降低了系统功耗。系统基于硬件编程实现,具有灵活的重配置能力,便于升级和扩展。

1 无线传感系统总体设计

本文所设计的无线传感系统由中心节点板卡和传感器子节点组成,系统整体结构如图 1 所示。中心节点板卡上包含若干个主节点和一个 FPGA

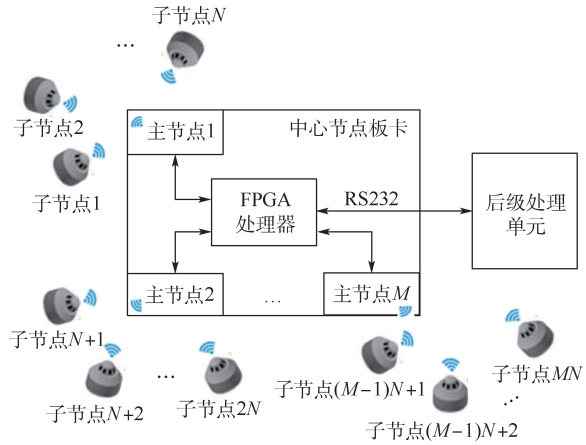


图 1 系统整体结构

Fig. 1 Overall structure of system

处理器。对所有传感器子节点进行分组,每组包含一个主节点和多个子节点。子节点将采集到的数据发送给同组内的主节点,主节点将数据汇总后发送给 FPGA,FPGA 将数据再次汇总后传送给后级处理单元。

因子节点体积与功率受限,且无线系统需要实现多通道数据高速率传输,选用 CC1310 芯片,主要有以下 4 个优势:①芯片高度集成化,集传感器数据采集、ARM 及射频处理器于一体,且体积小;②具有超低功耗的特点,支持电池供电,接收电流仅为 8 mA,休眠电流最小值仅为 2 μA;③数据传输速率最快可达 4 Mbps,支持多频段通信;④采用 GFSK 调制,抗干扰能力强,且支持自动 CRC 校验和多次重传,保证了数据传输的可靠性。

为了简化系统开发流程,节约设计成本,无线传感网络的子节点与主节点均基于 CC1310 芯片搭建。芯片内置的 ARM Cortex-M3 可用于编程,实现数据传输协议及工作模式切换。子节点核心电路单元包含一个 CC1310 芯片和若干个传感器,采用电池供电;主节点主要使用无线通信功能。核心处理器 FPGA 因没有操作系统,具有功耗低和可编程重配置特点,管脚功能也可以灵活定义,便于后续扩展。本系统选用一款中型 FPGA 处理器 EP1C12Q240I7 实现无线节点的管理及数据组帧等逻辑运算。FPGA 可以接收后级模块下发的命令来改变无线节点的工作模式等。

2 系统高速率数据传输设计

2.1 频分复用结合时分复用组网设计

CC1310 芯片支持多个频段(287 ~ 351 MHz, 359 ~ 527 MHz,718 ~ 1 054 MHz)的无线通信[17],基于此提出了组间频分复用结合组内时分复用的

组网方式,对多个子节点进行分组管理。

无线传感器系统组网示意图如图 2 所示,中心节点包含 M 个主节点,不同的主节点被配置到不同的频段,组间频分复用成倍地增加了子节点的数量及系统传输速率。一个主节点只与同频段内的 N 个子节点通信,构成一个分组。在分组内部,为避免干扰和碰撞,同一时刻最多仅有一个子节点处于发送状态,因此,采用时分复用的方式,将一个传输周期划分为多个时隙,每个子节点仅占据一个时隙进行数据传输。

另外,各节点独立工作,保持一个由自身时钟测量的本地时间。因生产加工过程会导致同型号器件存在差异,使得各节点出现计时偏差,进而影响系统时间同步,给系统的正常工作带来挑战[18]。且节点数越多,系统时间同步的复杂度也越高。采用组间频分复用方法可减少组内子节点个数,降低无线通信系统的同步复杂度。当分组数为 M 时,在同样的计时精度约束下,可容忍的计时偏差变为原来的 M 倍,该优势可用于延长同步校准周期,降低系统开销,提高传输效率。

2.2　无线传感系统组内节点时间同步方法

为避免节点碰撞,应实现高精准同步,本文设计了主节点定时授时方法。主节点均位于中心节点板卡上,具有时钟一致性,因此,仅需达到组内的子节点与主节点时钟一致,即可实现整个系统的时间同步。主节点定时授时法即每隔一个授时周期 T,主节点广播授时帧,各子节点根据监听到的授时信号调整本地计时器,消除累计误差。周期性授时同步示意图如图 3 所示。图中:每一行的命令时隙和传输时隙合称为一个传输周期 T_s,在此时间内,所有子节点完成一次数据上传。若干次传输周期之和即为一个授时周期 T,每个节点的传输时隙长度可以根据需要分别设置。

理论分析表明,在相邻节点传输时隙间留有

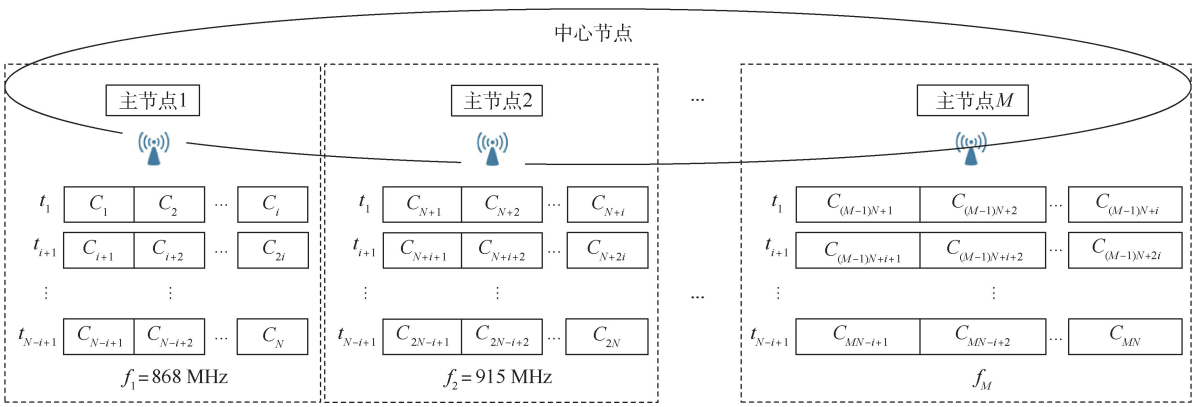


图 2　无线网络组网示意图

Fig. 2　Schematic of wireless network

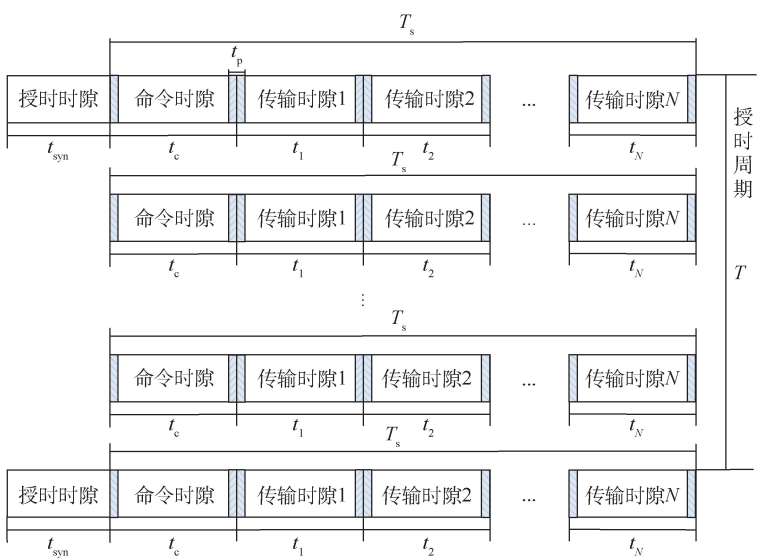


图 3　系统周期性授时同步示意图

Fig. 3　Schematic of system periodic timing synchronization

一定保护间隔 t_p，只要在授时周期 T 内，相邻节点的最大计时偏差累积量小于保护间隔，就可避免节点碰撞。因此，时间同步的关键在于：如何设计保护间隔 t_p 和授时周期 T。

计时偏差取决于节点本地时钟的频率稳定度，各无线节点采用 ARM CPU 作为核心处理器，其集成一个系统计时器，时钟频率可选，最高可达 48 MHz。本系统采用 24 MHz 频率进行计时，晶振稳定度约每 100 MHz 时偏差 ±10 Hz，即 1 s 过后子节点的计时偏差 t_b 最大可达到 2×10^{-5} s，即间隔 T 后，各节点计时误差最多相差 Tt_b。因此只要令 $t_p \geqslant Tt_b$，就可避免碰撞。如图 4 所示，最小保护间隔 t_{pmin} 与授时周期 T 的关系是典型的线性关系。

另外，传感器能量有限也是进行时间同步时首要考虑的问题。太过频繁的授时会导致节点能量消耗增加，并且授时时隙 t_{syn} 不能用于数据传输，系统同步开销增加，从而降低传输效率。授时间隔太长则要求保护间隔足够大，传输时隙内真正用于数据传输的时间减少，造成传输速率降低；保护间隔太小则会使得各子节点与主节点失去同步，产生节点碰撞，降低系统传输速率和可靠性，因此，应合理的设定授时周期。授时周期的设计方法将在 2.3 节传输协议的优化设计中给出。

在授时时隙内，主节点在组内广播 3 帧授时帧，发送 3 帧是为了确保子节点可监听到授时信息。子节点根据接收到的第 1 个授时帧编号计算得出本地时间的置 0 时刻，如图 5 所示，t 为中心节点发送一帧授时帧所需时间，空中传输时间忽略不计。如果子节点收到了编号为 $k(1\leqslant k\leqslant 3)$ 的授时帧，则再经过 $3-k$ 个 t 时刻之后将本地时钟置 0。

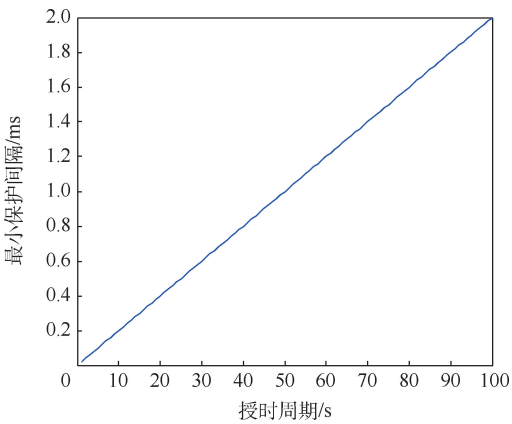


图 4 最小保护间隔与授时周期的关系

Fig. 4 Relationship between minimum protection interval and timing period

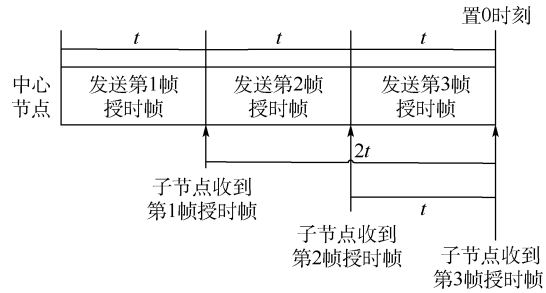


图 5 系统各节点时钟同步示意图

Fig. 5 Schematic of system nodes clock synchronization

2.3 无线传感系统多节点传输协议设计

在多通道无线传感器系统中，系统时间同步是可靠通信的基础，如何实现各子节点与主节点间的无碰撞通信，并尽可能提高可达的传输速率是系统的最终目标，这就需要合理安排每个子节点的数据发送时刻与发送时隙，并优化设计授时周期及保护间隔，即多节点传输协议的设计。

一个授时周期 T 内，包含有 r 轮传输周期 T_s，为了保证系统实时采集与传输，在一个传输周期内需要将所有子节点采集的数据均发送给主节点。以无线通信系统的一个频段为例，假设此频段内现有 N 个正常工作的子节点，且每个子节点的采样率可能不同，各子节点编号为 $i(1\leqslant i\leqslant N)$，其对应的采样率为 $f_i(1\leqslant i\leqslant N)$。为避免产生碰撞，需在相邻子节点传输时隙之间留出保护间隔 t_p。

由图 3 所示，一个传输周期 T_s 包括了命令时隙 t_c（用于完成系统的模式配置）、N 个子节点的传输时隙，协议中所有时间变量均以秒为单位。在每个传输周期开始时，都可以重新配置各子节点的采样速率和传输时隙长度。CC1310 芯片对采集到的数据进行 Q 比特的量化，其数据传输速率可利用程序测试得出，各子节点根据自己的采样率计算传输数据所需传输时隙长度。

无线通信系统中，在满足所有子节点时间同步的前提下，要保证可靠的数据传输，需严格计算各子节点的数据发送时刻。根据各子节点的发送时隙，可以推算出第 $k(1\leqslant k\leqslant r)$ 个传输周期时第 i 个子节点的发送时刻 t_{start_ki} 为

$$t_{start_ki} = T_s(k-1) + t_c + \sum_{n=1}^{i-1} t_n + \frac{1}{2}t_p \tag{1}$$

式中：t_n 为第 n 个子节点的传输时隙长度。

根据传输周期的时隙组成，得到子节点数 N 与授时周期 T 的约束关系：

$$T = rt_c + \sum_{k=1}^{r}\sum_{i=1}^{N} t_{ki} \tag{2}$$

式中：t_{ki} 为第 k 个传输周期时第 i 个节点的传输时隙。不失一般性，设一个授时周期 T 内，每个传

输周期的时间 T_s 相同,系统实际传输时所需的净荷传输速率为 R_d,表示为各个子节点的采样率乘以量化比特数再求和:

$$R_d = \sum_{i=1}^{N} f_i Q \tag{3}$$

由式(3)可知,当一个传输周期的节点数越多,量化值越精确,所需的传输速率 R_d 就越大。系统设计时,应保证最高可达的传输速率 $R_a \geqslant R_d$,才能满足航天运载火箭应用的实际要求。

单频段内最大可达的传输速率,其主要由授时周期 T、子节点个数 N 及传输时隙 t_{ki} 所决定。为了简化分析,假定各子节点采样速率相同,即所需的传输时隙 $t_{ki} \equiv t_1$。不失一般性,设命令时隙与子节点的传输时隙长度也相同。则分组内可达传输速率计算如下:

$$R_a = \frac{\frac{t_1 - t_p}{t_1} R_{t1} r N}{t_{syn} + T} = \frac{\frac{t_1 - r(N+1) t_1 t_b}{t_1} R_{t1} r N}{t_{syn} + r(N+1) t_1} \tag{4}$$

式中:R_{t1} 为 t_1 的函数,表示在 t_1 内 CC1310 全部用于传输数据时可达到的传输比特数,该数值可以通过程序测试得出。经编程测试后发现,当 $t_1 > 10$ ms 时,R_{t1} 与 t_1 呈正比例关系,表明此时 CC1310 发送程序中用于数据发送的时间远远大于程序初始化处理时间,程序设计具有高效性。实际应用中,为保证 CC1310 传输效率高,取 $t_1 > 10$ ms。结合图 3,可以得出 $T = r(N+1) t_1$,显然当 t_p 取得最小值 Tt_b 时,有利于 R_a 取到最大值,即式(4)中结果。当节点数 $N \gg 1$ 时,$(N+1) \approx N$。记 $r(N+1) = v$,保护间隔不会超过一个传输时隙的总长度,即 $t_p = v t_1 t_b \leqslant t_1$,单频段内可达传输速率最大化可建模为如下优化问题:

$$\max_v R_a = \frac{(1 - v t_b) R_{t1} v}{t_{syn} + v t_1} \qquad 2 \leqslant v \leqslant \frac{1}{t_b} \tag{5}$$

式中:子节点在 1 s 内的计时误差 t_b 取最大值 2×10^{-5} s;v 对应的为一个授时周期内所有传输时隙的个数,应取整数,其取值范围为 $2 \leqslant v \leqslant 50\ 000$。单频段内最大可达传输速率 R_{aopt} 由参数 t_1、t_{syn} 和变量 v 决定,t_{syn} 和 t_1 可预先设定。本文系统设 $t_1 = 15$ ms,测得 R_{t1} 为 3.2 Kbits。为了便于系统设计,授时时隙 t_{syn} 也取 15 ms,根据 v 的取值范围采用遍历搜索的方法对上述优化问题进行求解。授时周期 $T = v t_1$,遍历搜索后,单频段可达传输速率 R_a 随授时周期 T 的变化规律如图 6 所示。

由图 6 可知,授时周期 T 为 3.35 s 时,对应的最优值 $v_{opt} = r(N+1) = 223$,可得达到单频段传输的最大速率 $R_{aopt} = 213$ Kbps。在保证最高可达传输速率时,N 和 r 可以互换,可以根据节点数 N 来调整传输轮数 r,当 $N = 30$ 时,求得 $r = 7$,当系统有 4 个频段同时工作时,可支持的节点数为 120。实际传输时,需要在功耗和速率之间进行折中,得到一个工程上的近最佳授时周期值。当每个节点的采样速率不同时,可以将高采样率节点等效为若干个低采样率节点,再进行求解。

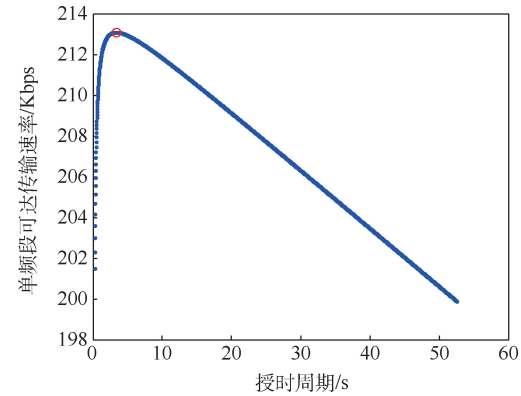


图 6　单频段可达的传输速率

Fig. 6　Achievable transmission rate on single frequency

以一个频段为例,当 M 个频段同时传输时,可达的合路传输速率 R_{asum} 为

$$R_{asum} = \sum_{m=1}^{M} R_{am} = M \overline{R}_a \tag{6}$$

可见,随着复用频段数增加,系统合路传输速率会成倍扩增,满足航天运载火箭模拟信号监测的高速率传输需求。

3　系统低功耗实现方法

CC1310 芯片具备多个模式,其中包括射频模式和低功耗模式,通过编程配置即可切换。本系统采用唤醒/休眠的模式切换策略在实现系统功能的前提下进一步降低系统功耗。

如图 7 所示,系统上电后,主节点获取当前频段内所有活跃子节点的编号和采样率。进入授时期,主节点为发送模式,向各子节点广播 3 次授时帧进行系统时钟同步;同步完成后,向各个子节点广播 2 次命令帧,用于完成系统的模式配置;进入数据传输阶段主节点切换为接收模式,接收来自各子节点的数据。在一个传输周期 T_s 末,各个主节点读取后级模块发来的配置命令。如果本次传输周期 T_s 结束时刻,正好也是一轮授时周期 T 结束时刻,则系统进入授时同步状态,主节点转换为发送模式,继而发送授时帧,否则主节点发送命令帧。

如图 8 所示,系统上电后,子节点处于接收模

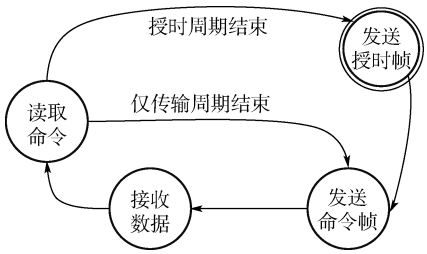


图 7 主节点状态转移图

Fig. 7 Master node state transition diagram

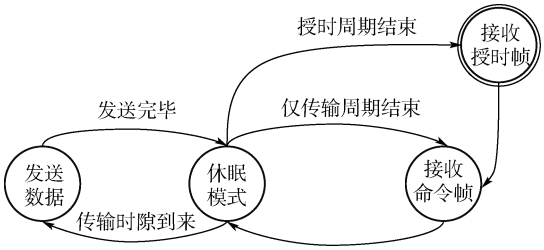


图 8 子节点状态转移图

Fig. 8 Sub-node state transition diagram

式,接收来自主节点的授时帧,根据帧编号计算本地时钟置 0 时刻,从而与主节点保持同步;之后子节点仍处于接收状态,接收主节点发送的命令帧,调整其数据发送时刻与发送时隙,以达到系统配置的要求;进入数据传输阶段后,子节点先切换为休眠模式,等待直至自己的传输时刻到来,子节点切换为发送模式,向主节点发送采集到的数据,发送完毕后再次切换为休眠模式;等待命令时隙或授时时隙到来,子节点被唤醒,切换为接收模式。

通过唤醒/休眠的模式切换策略,各子节点仅在命令时隙和自身传输时隙的有效时间内处于射频模式,其他时间均处于休眠模式。假设节点处于射频模式时功耗为 P_{RF},处于休眠模式时功耗为 P_{idle}。在唤醒/休眠模式切换策略下,单个子节点在一个完整周期 $T+t_{syn}$ 内的平均功耗 $\overline{P}_{low}$ 为

$$\overline{P}_{low}=\frac{t_{syn}+(t_c-t_p)r+(t_1-t_p)r}{T+t_{syn}}P_{RF}+\left[1-\frac{t_{syn}+(t_c-t_p)r+(t_1-t_p)r}{T+t_{syn}}\right]P_{idle} \quad (7)$$

式中:为简化系统设计,令 $t_c=t_1=15$ ms。在未使用模式切换策略时,即每个子节点全程工作于射频模式,此时单个节点在一个完整周期 $T+t_{syn}$ 内的平均功耗 $\overline{P}_{high}$ 即为 P_{RF}。由 2.3 节分析可知,为达到最高传输速率,理论上最佳的授时周期为 3.35 s,但是太过频繁授时会增加系统功耗,工程应用中,在满足所需的净荷传输速率的情况下,可适当增大授时周期来达到低功耗。设 $R_d=200$ Kbps,参见图 6,选择 $T=50$ s,$R_a=201$ Kbps,传输速率满足需求,且授时周期增长约 14 倍,对比单个子节点平均功耗,如图 9 所示。

由图 9 可知,采用模式切换并且 T 为最佳值 3.35 s 时,与未采用模式切换方式相比,单节点平均功耗显著降低,降低比例为 0.4% ~78%。当子节点数较少且传输时隙较短时,因同步时隙和命令时隙占比重较多,而这 2 个时隙节点处于射频模式,即使采用模式切换,子节点平均功耗相对较高。随着子节点数增多,模式切换的平均功耗越逼近休眠模式功耗。当工程应用中选择 $T=50$ s 时,与理论最优的 $T=3.35$ s 相比,功耗进一步降低,降低比例达 2% ~5%。可见,延长授时周期可进一步降低系统功耗,实际应用中,可根据传输速率要求,结合图 6 确定满足条件的最大授时周期 T,使得系统功耗尽可能低。

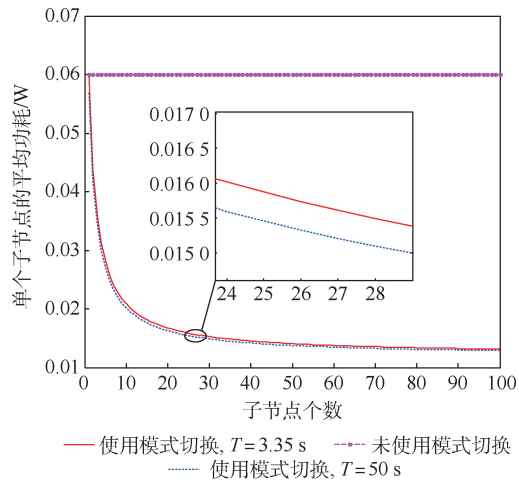


图 9 单个子节点平均功耗

Fig. 9 The average power for a sub-node

4 实验结果与分析

采用 5 个传感器子节点和 2 个主节点进行系统测试,将其分为 2 组,第 1 组的子节点编号为 1、2、4,无线收发频段为 868 MHz,第 2 组的子节点编号为 6、8,无线收发频段为 915 MHz。各子节点所采集模拟信号及采样率设置如表 1 所示。为了更好地显示系统功能和性能,本文系统也开发了上位机软件用于监视各节点工作状态。

表 1 传感器子节点采集信号与采样率

Table 1 Acquisition signals and sampling rates of sensor sub-nodes

组别(频段/MHz)	子节点编号	模拟信号源	采样率/Hz
第 1 组(868)	1	电池电压	320
	2	1 kHz 正弦波	5 120
	4	湿度传感器	320
第 2 组(915)	6	温度传感器	320
	8	电池电压	320

4.1　系统速率测试

各传感器子节点与中心节点距离 10 m 以内,进行短距离无障碍通信。上位机软件界面显示了系统各子节点所采集的模拟信号,如图 10 所示,此时波形图为子节点 2 采集的 1 kHz 正弦波。

在本文系统测试中,授时周期 $T=50$ s,$t_{syn}=15$ ms,CC1310 对模拟信号采用 8 bit 量化,为了在节点数很少时仍能测试系统的最高实际传输速率,令每个分组内每个节点在 T_s 内多次等概率出现,即用 2 个或 3 个真实节点模拟出 $N=30$ 个子节点,保证传输时隙不空闲,传输时隙长度仍为 15 ms。系统中实测的各子节点的传输速率统计如表 2 所示,每个频段内所有子节点速率之和约 200 Kbps,系统总速率达到 400 Kbps,与理论分析结果一致。因组内各子节点等概率出现,测得的各子节点传输速率相同,但其按需传输时的净荷传输速率取决于真实采样速率。

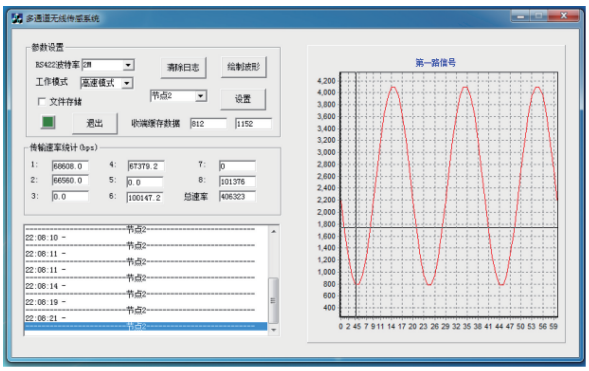


图 10　上位机软件运行工作图

Fig. 10　Computer software operation diagram

表 2　子节点传输速率

Table 2　The transmission rate of sub-nodes

组别(频段/MHz)	子节点编号	传输速率/bps	净荷传输速率/bps
第 1 组(868)	1	68 608	2 560
	2	66 560	40 960
	4	67 379	2 560
第 2 组(915)	6	100 147	2 560
	8	101 376	2 560
系统总速率		406 323	51 200

4.2　子节点功耗测试

本文系统中所设计的传感器子节点实测体积为 3 cm × 3 cm × 1 cm,其质量约 10 g。

子节点使用 3 V 干电池进行供电,测得子节点处于工作模式时的电流为 20 mA,$P_{RF}=60$ mW,处于休眠模式的电流为 4 mA,$P_{idle}=12$ mW。采用模式切换后,根据式(7),计算得出单个子节点的平均功耗为 $\overline{P}_{low}=14.9$ mW(等效于 $N=30$),在电池电量为 10 Ah 时,能连续工作 83 天,而在未采用模式切换时,仅能连续工作 20 天。

4.3　系统可靠性及鲁棒性测试

改变子节点与中心节点的距离测试系统的可靠性和鲁棒性,以子节点 6 为例,测试结果如图 11所示。由图 11 可知,子节点距离中心节点 20 m 内,可正常通信,实际传输速率在 100 Kbps 的小范围内波动;但两者距离从 20 m 增加至50 m 时,子节点的传输速率持续下降,最终在接近 50 m时,传输速率为 0;此时当子节点再次靠近主节点时,通信可自主恢复,子节点的传输速率提升直至稳定于 100 Kbps。本文系统也测试了子节点与主节点在有墙障碍时的无线通信,结果显示,当墙体为单层时,两者能正常通信,墙体大于单层时通信中断。

在所有测试场景下,均未发生节点碰撞的情况,说明系统具有可靠性。当某个子节点发生故障时,并不影响整个系统的运行,说明系统具有鲁棒性。子节点数目及采样率均可通过程序配置动态调整,系统具有较好的灵活性和可扩展性。

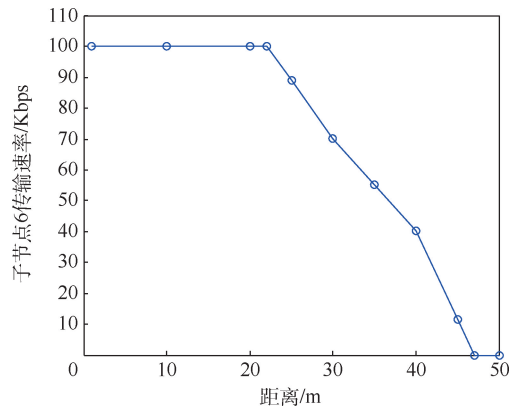


图 11　子节点 6 传输速率随距离变化的测试结果

Fig. 11　Test result of transmission rate for single sub-node 6 with changing distance

5　结　论

1) 本文基于 CC1310 芯片设计并实现了一个适用于运载火箭的多通道无线传感系统,提出频分复用结合时分复用的组网方式,增加系统可支持的子节点数,提高了系统的速率复用倍数。

2) 设计了主节点周期性授时方案,保证系统的高精准同步;并以可达传输速率最大化为目标,优化设计了分时传输协议,提升了分组内传输速率。结合工程实际应用,对最佳授时周期进行修正。

3) 提出唤醒/休眠模式切换策略,极大降低了子节点的平均功耗,充分满足航天运载火箭对

于无线传感系统的低功耗需求。

4）本文设计的系统具有较好的可靠性和鲁棒性，硬件采用CC1310芯片及FPGA平台，可编程配置，便于后期扩展升级。

参考文献（References）

[1] PATEL N R, KUMAR S. Wireless sensor networks' challenges and future prospects[C]// International Conference on System Modeling & Advancement in Research Trends. Piscataway: IEEE Press, 2018: 60-65.

[2] CARMINATI M, KANOUN O, ULLO S L, et al. Prospects of distributed wireless sensor networks for urban environmental monitoring[J]. IEEE Aerospace and Electronic Systems Magazine, 2019, 34(6): 44-52.

[3] WU J, JIANG W, MEI Y, et al. A survey on the progress of testing techniques and methods for wireless sensor networks[J]. IEEE Access, 2019, 7: 4302-4316.

[4] 罗煜缤，李洪，周广铭，等. 新一代箭载无线传感器网络系统架构综述[J]. 宇航计测技术，2020，40(4)：1-6.
LUO Y B, LI H, ZHOU G M, et al. Review on the system architecture of new generation wireless sensor network for rocket [J]. Journal of Astronautical Metrology and Measurement, 2020, 40(4): 1-6(in Chinese).

[5] WANG Y, LU J, HAO C, et al. Applicable to launch vehicle data integrated design method of wireless sensor network research [C]//2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC). Piscataway: IEEE Press, 2018: 221-225.

[6] WANG H, ZENG H, PING W. Linear estimation of clock frequency offset for time synchronization based on overhearing in wireless sensor networks[J]. IEEE Communications Letters, 2016, 20(2): 288-291.

[7] LEANDRO T B, TALES H, EDISON P. Enhancing time synchronization support in wireless sensor networks[J]. Sensors, 2017, 17(12): 2956.

[8] ABDULKAREM M, SAMSUDIN K, ROKHANI F Z, et al. Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction [J]. Structural Health Monitoring, 2020, 193: 693-735.

[9] SUN X, SU Y, HUANG Y, et al. Design and development of a wireless sensor network time synchronization system for photovoltaic module monitoring[J]. International Journal of Distributed Sensor Networks, 2020, 16(5): 1-8.

[10] 庄祎梦，贺光辉，陈洪涛，等. 无线传感器网络高精度时间同步机制[J]. 信息技术，2017(10)：1-4.
ZHUANG Y M, HE G H, CHEN H T, et al. Method of highly accurate time synchronization for wireless sensor networks[J]. Information Technology, 2017(10): 1-4(in Chinese).

[11] 麦军，邓巧茵，万智萍. 基于CC2530的ZigBee无线组网温度监测系统的设计[J]. 电子设计工程，2015，23(22)：117-121.
MAI J, DENG Q Y, WAN Z P. A ZigBee wireless network temperature monitoring system based on CC2530[J]. Electronic Design Engineering, 2015, 23(22): 117-121(in Chinese).

[12] SAHITYA G, BALAJI N, NAIDU C D, et al. Designing a wireless sensor network for precision agriculture using Zigbee[C]// IEEE International Advance Computing Conference. Piscataway: IEEE Press, 2017: 287-291.

[13] NURGALIYEV M, SAYMBETOV A, YASHCHYSHYN Y, et al. Prediction of energy consumption for LoRa based wireless sensors network[J]. Wireless Networks, 2020, 26(5): 3507-3520.

[14] SOMMER P, MARET Y, DZUNG D. Low-power wide-area networks for industrial sensing applications[C]// IEEE International Conference on Industrial Internet. Piscataway: IEEE Press, 2018: 23-32.

[15] 周洋. 无线传感网节点低功耗控制技术的研究[D]. 苏州：苏州大学，2019：33-39.
ZHOU Y. Research on low power control technology of wireless sensor network nodes[D]. Suzhou: Soochow University, 2019: 33-39(in Chinese).

[16] 朱志斌. 基于CC1310的无线传感网络记录仪设计[D]. 太原：中北大学，2021：30-36.
ZHU Z B. Design of wireless sensor network recorder based on CC1310[D]. Taiyuan: North University of China, 2021: 30-36 (in Chinese).

[17] MAHBUBUR R, DALI I, VENKATA P. et al. Implementation of LPWAN over white spaces for practical deployment[C]// Conference on Internet-of-Things Design and Implementation. New York: ACM, 2019: 178-189.

[18] JIAN L, MECHITOV K A, KIM R E, et al. Efficient time synchronization for structural health monitoring using wireless smart sensor networks[J]. Structural Control and Health Monitoring, 2016, 23(3): 470-486.

Multi-channel wireless sensor system based on CC1310 chip wtih high speed and low power consumption

DUAN Ruifeng[1,2,*], LYU Yanjie[3], DU Wenji[4,5], ZHOU You[6]

(1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China;
2. Engineering Research Center for Forestry-oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China;
3. School of Electronics and Information Engineering, Beihang University, Beijng 100191, China;
4. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
5. School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China;
6. SiEngine Technology Co., Ltd., Beijing 100012, China)

Abstract: In order to effectively reduce the weight of sensor data acquisition and transmission system for the new-generation launch vehicle, a wireless sensor system based on chip CC1310 was designed and implemented. Multi-node networking and group management are accomplished using frequency division multiplexing (FDM) and time division multiplexing (TDM). The FDM scheme among groups not only increases the number of nodes, but also improves the transmission rate multiplexing multiple. When there are four groups, the system is capable of supporting more than 100 sub-nodes. In order to ensure high precision synchronization of multi-nodes, avoid node collisions, and obtain the optimal intra-group achievable rate, we also suggested an optimal method of master node timing in combination with a multi-node time-sharing transfer protocol. Moreover, a node wake-up/idle mode switching strategy is designed to reduce system power consumption effectively. The experimental results show that the transmission rate can reach 400 Kbps when two master nodes work in parallel with five sub-nodes, and the transmission rate will increase proportionally when the number of master nodes rises. The power consumption of the single sub-node is less than 60 mW during work hours and less than 12 mW during idle hours. The average power consumption of the single sub-node is 15.2 mW, which meets the requirements of low power consumption. At the same time, the wireless sensor system designed in this paper has good reliability and robustness.

Keywords: CC1310; wireless network communication; time synchronization; high rate; low power consumption

Received: 2021-11-11; **Accepted**: 2022-02-13; **Published online**: 2022-03-09 09:30

URL: kns.cnki.net/kcms/detail/11.2625.V.20220307.1346.004.html

Foundation items: The Fundamental Research Funds for the Central Universities (BLX201623); Beijing Natural Science Foundation (L202003); National Natural Science Foundation of China (31700479)

* **Corresponding author**. E-mail: drffighting2008@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0294

基于注意机制的轻量化稠密连接网络单幅图像去雨

柴国强, 王大为, 芦宾*, 李竹
(山西师范大学 物理与信息工程学院, 临汾 041000)

摘 要: 图像中附着的雨条纹对背景造成的破坏严重影响了对图像信息的分析和后续研究。为了恢复被雨条纹破坏的背景纹理特征,提出一种基于注意机制的轻量化稠密连接网络针对单幅图像进行去雨。注意机制有利于网络准确定位降雨区域,稠密连接网络的使用增强了特征的复用,缓解了梯度消失和模型退化问题。利用多尺度通道混洗深度可分离卷积实现网络轻量化设计,降低了网络参数规模,提升了网络运行效率。在合成数据集和真实数据集上的去雨结果表明,所提算法在定量指标和定性分析上都优于现有算法。

关 键 词: 注意机制; 稠密连接网络; 轻量化设计; 图像去雨; 深度学习
中图分类号: TP391
文献标志码: A **文章编号**: 1001-5965(2022)11-2186-07

随着科技的发展,目标检测与跟踪、场景语义分割、深度信息估计[1-3]等计算机视觉算法已经广泛应用于无人机拍摄、人脸识别系统、智慧城市预警等现代化科技中。在阴雨天气下,雨条纹会对相机等传感器获取的图像造成严重的噪声污染,从而降低后续计算机视觉算法的可靠度和准确性,限制了其应用范围。对降雨时获取的图像先进行去雨处理,重构出背景纹理具有重要研究意义。现有图像去雨算法根据所研究的对象不同主要可以分为基于视频的去雨算法和基于单幅图像的去雨算法。基于视频的去雨算法通过分析连续帧图像之间的差异,根据丰富的时空关系检测出雨条纹并将其去除。相比而言,由于缺乏时空信息,单幅图像去雨难度更大。

根据是否需要数据集进行训练,单幅图像去雨算法可以分为基于传统的优化算法和基于深度学习的算法。传统算法中,Li 等[4]利用基于高斯混合模型的先验概率对雨条纹进行检测和去除,通过对背景层和雨条纹层施加预先定义的约束实现图像分解,利用结构残基将雨条纹去除。Kim 等[5]采用自适应的非局部邻域滤波对雨条纹进行均值滤波将其去除。Luo 等[6]提出了一种基于字典学习和稀疏编码的单幅图像去雨算法,其中强互斥性的字典学习逼近雨条纹层和背景层,判别稀疏编码将雨条纹去除。Wang 等[7]将字典学习和图像分解相结合实现了单幅图像去雨。

近年来,随着深度学习框架的不断优化和计算机硬件水平的飞速发展,计算机视觉算法性能和运行效率都得到了大幅度提升。Yang 等[8]提出了一种基于空洞卷积的多通道单幅图像去雨网络,通过空洞卷积扩大网络感受野,通过闭环反馈实现雨条纹的逐步去除。Wang 等[9]先利用空间感知网络得到雨条纹注意图,再利用局部到全局的方式将雨条纹去除。Ren 等[10]利用深度特征在不同阶段之间的依赖关系,提出了渐进循环去雨网络。Zhang 等[11]提出了一种多任务的去雨网络,在完成去雨的同时实现高分辨率图像的生成。Xia 等[12]提出了一种基于图像分解的去雨网络,先将雨图像分解为高频部分和低频部分,再利用简化的稠密连接网络在高频部分将雨条纹去

收稿日期: 2021-06-03; **录用日期**: 2021-08-27; **网络出版时间**: 2021-09-15 13:06
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210914.1727.012.html
基金项目: 国家自然科学基金 (62201333,62004119); 山西省基础研究计划 (20210302124647)
* **通信作者**. E-mail: lubinsxnu@sina.cn

引用格式: *柴国强, 王大为, 芦宾, 等. 基于注意机制的轻量化稠密连接网络单幅图像去雨[J]. 北京航空航天大学学报, 2022, 48(11): 2186-2192. CHAI G Q, WANG D W, LU B, et al. Lightweight densely connected network based on attention mechanism for single-image deraining[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2186-2192(in Chinese).*

除。王聪等[13-14]分别采用网络与内部尺度融合操作和自校准网络[15]实现了单幅图像去雨。

虽然现有基于深度学习的图像去雨算法取得不错进展,但是就图像去雨质量和网络参数规模方面仍有较大提升空间。本文提出一种基于注意机制的轻量化稠密连接网络,在去除雨条纹的同时降低网络参数规模,提升模型运行效率。

1　网络结构

本文所提出的基于注意机制的改进稠密连接去雨网络结构如图 1 所示。为了提升雨条纹区域的权重,利用多尺度聚集自适应雨条纹检测模块(multi-scale aggregation self-adapting rain streaks detection module, MASM)检测雨条纹,生成注意图用于引导后续模块有意识地关注降雨区域。雨条纹具有不同的大小和形状,与背景图像相互影响,单一深层网络可能无法从图像中完全去除雨条纹。因此,本文采用循环策略对雨条纹进行分步去除。针对去雨任务,对稠密连接网络进行修改,包括利用空洞卷积增加网络感受野,去除批归一化层(batchnorm layer, BN)增强网络鲁棒性,使用挤压-激励提升网络表征能力。

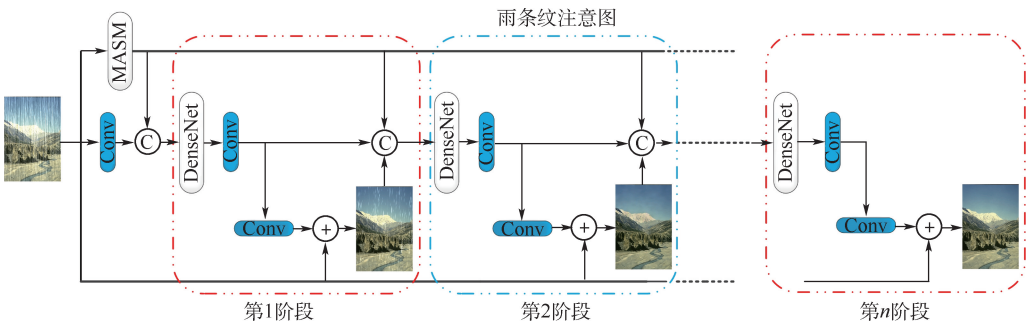


图 1　基于注意机制的改进稠密连接去雨网络结构

Fig. 1　Structure of modified densely connected rain removal network based on attention mechanism

1.1　多尺度聚集自适应雨条纹检测模块

视觉注意机制有助于计算机视觉算法准确定位于特定区域。为使去雨网络给予雨条纹区域有意识的关注,本文设计了 MASM 模块检测雨条纹,MASM 网络结构如图 2所示。

在图 2 中,输入特征图 X 经过 3 个具有不同大小卷积核的并行卷积通道,将生成的特征图沿通道维连接并且经过 1×1 的卷积核进行特征融合,输出 Y 表示特征权重,并行卷积通道操作可表示为

$$\begin{cases} X_i = F^{(k)}(X) \qquad i \in \{1,2,3\}, k \in \{3,5,7\} \\ Y = F^{(1)}([X_1, X_2, X_3]) \end{cases} \tag{1}$$

式中:X 表示输入;$X_i(i=1,2,3)$ 表示 3 个通道的输出,每个通道特征图通道维度为 32;$F^{(k)}(\cdot)$ 表示卷积核的尺寸为 $k\times k$;$[X_1, X_2, X_3]$ 表示不同特征图在通道维度上的拼接。

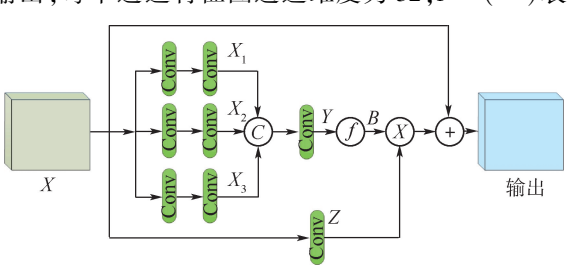


图 2　多尺度聚集自适应雨条纹检测模块结构

Fig. 2　Structure of multi-scale aggregation self-adapting rain streaks detection module

为了检测出小权重雨条纹,首先利用 softmax 函数对 Y 进行处理得到归一化的权重 B,然后将另一个分支生成的特征图 Z 与 B 相乘,最后通过自适应的方式将乘积结果与输入特征 X 相加:

$$S = \lambda(BZ) + X \tag{2}$$

式中:λ 为可学习参数;S 为雨条纹注意图。

图 3 为雨条纹检测模块生成的注意图,可以看出几乎所有雨条纹都得到了标记。

1.2　改进的稠密连接网络

在硬件条件允许的情况下加深网络深度可以有效提升网络性能,但是过深的网络会增加梯度弥散的风险,随着网络深度的增加,前面的信息就会逐渐减弱和消散。为了确保网络中最大化的信息传递,稠密连接网络[16](densely connected networks, DenseNet)使每层都与该层之前的所有层在特征维度上相连,增强了特征的复用,使得信息(前向传播时)或梯度(后向传播时)在整个网络中可以更好的保持,从而训练更深的网络模型。

原始稠密连接网络模块如图 4(a)所示,设输入特征图具有 c 个通道,每个卷积层产生 c 个通道的新特征图,在网络的前向传递过程中,每一个卷积层都可以接收之前所有卷积层的输出,第 n

个卷积层的输出可以表示为

$$X_n = \sigma(W_n^{(k)}[X_1, X_2, \cdots, X_{n-1}]) \quad (3)$$

式中:σ 表示激活函数;$W_n^{(k)}$ 表示核尺寸为 $k \times k$ 的卷积操作;$[X_1, X_2, \cdots, X_{n-1}]$ 表示特征图沿通道维度的连接。具有 n 个卷积层的稠密连接网络模块最终会产生 $n \times G$ 个新的特征图,其中 G 表示生长率。

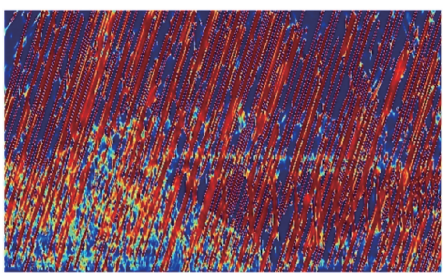
在图 4(a)所示原始稠密连接网络模块中,每个卷积层后都紧跟一个 BN 层,在训练过程中 BN 使用每一批次(mini-batch)的统计量归一化该批次的特征,在测试过程中,BN 使用所有训练数据的统计量归一化测试集。在测试集的统计量与训练集差别较大的情况下,BN 会影响网络性能,降低网络泛化能力[17]。此外,BN 层需要在 GPU 中存储标准化的数据副本,占用大量计算空间,因此本文将 BN 层从稠密连接网络模块中去除。



(a) 雨图像



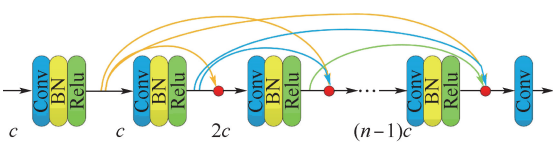
(b) 背景图



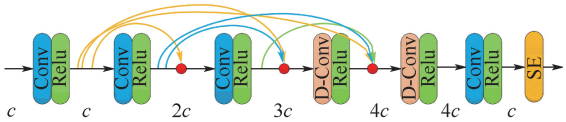
(c) 注意图

图 3 注意图可视化

Fig. 3 Visualization of attention map



(a) 原始稠密连接网络



(b) 改进的稠密连接网络

图 4 稠密连接网络

Fig. 4 Densely connected network

将部分稠密连接块中的常规卷积用空洞卷积(dilated convolution)替代,可以有效增加网络的感受野,在稠密连接网络的最后加入挤压-激励(squeeze-and-excitation,SE)模块,将通道特征和空间特征相结合,进一步增加网络特征提取能力。改进的稠密连接网络如图 4(b)所示,其中,输入特征图的通道为 64,生长率为 12。

1.3 多尺度通道混洗深度可分离卷积

稠密连接网络由于特征复用导致网络后端的卷积核通道维度较大,参数量急剧增加,带来了严重的计算负荷。深度可分离卷积将常规卷积分为深度卷积和逐点卷积,实现了特征图通道维度和空间维度的关联性解耦,可以有效减少网络参数[18-19]。在深度卷积中,每个输出特征层只对单一输入特征层负责,输入特征之间没有信息交换,在逐点卷积中,使用 1 × 1 的卷积核实现特征融合。

在上述深度可分离卷积中,不同通道之间的信息交换是通过最后一步卷积核为 1 × 1 的卷积实现的,较小的感受野导致了深度可分离卷积不利于特征通道之间的信息融合,限制了网络的表征能力。为了增强不同特征通道之间信息交流,同时获取不同尺度特征,首先将输入特征通道分组,使用不同尺寸卷积核对不同通道进行卷积,然后利用通道混洗技术[20]对不同组的卷积进行特征融合,达到特征通信目的。多尺度通道混洗的深度可分离卷积实现流程如图 5 所示。

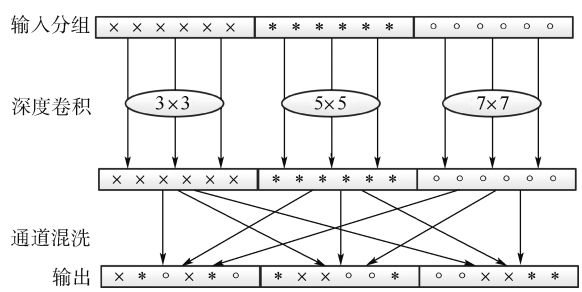


图 5 多尺度通道混洗的深度可分离卷积

Fig. 5 Depthwise separable convolution with multi-scale channel shuffle

2 损失函数

为了确保雨条纹去除的同时恢复出清晰可靠的背景纹理,结合像素目标函数、负结构相似目标

函数、感知目标函数、边缘目标函数及注意图目标函数对提出的去雨网络进行训练。

像素目标函数为去雨图像和真实背景 RGB 幅值的均方差异,其定义为

$$L_c = L_{mse}(G(x), y) \tag{4}$$

式中:$L_{mse}(\cdot)$为均方误差函数;x 为雨图像;$G(x)$为去雨结果;y 为无雨的干净背景。

负结构相似目标函数从亮度、对比度和结构 3 个方面度量 2 幅图像的区别,其定义为

$$L_{ssim} = 1 - \mathrm{SSIM}(G(x), y) \tag{5}$$

式中:SSIM 表示结构相似度指数[21],SSIM 越大,2 幅图像越相近。

感知目标函数[22]从高维度特征差异衡量 2 幅图像,定义如下:

$$L_p = \sum_{i=3,8,15,25} L_{mse}(V_i(G(x)), V_i(y)) \tag{6}$$

式中:$V_i(\cdot)$表示在 ImageNet 数据集上预训练的 VGG-16 网络提取的第 i 层特征图。

采用均方误差训练的网络一般会对生成图像的边缘纹理产生破坏[23],通过增加边缘目标函数来约束网络对边缘纹理的重构,其定义为

$$L_e = L_{mes}(W(G(x)), W(y)) \tag{7}$$

式中:$W(\cdot)$为拉普拉斯边缘检测算子卷积。

注意图目标函数可以有效保证 MASM 模块能够准确检测出降雨区域,定义为

$$L_{att} = L_{mes}(A(x), M(x)) \tag{8}$$

式中:$A(x)$为生成的雨条纹注意图;$M(x)$为雨条纹二值图(0 为无雨,1 为有雨),$M(x)$是通过从雨图中减去无雨图像,再通过与预设的阈值比较获取的。

网络总体目标函数为

$$L = \lambda_1 L_c + \lambda_2 L_{ssim} + \lambda_3 L_p + \lambda_4 L_e + \lambda_5 L_{att} \tag{9}$$

式中:$\lambda_1 \sim \lambda_5$ 为不同目标函数项的权重系数。

3　实验结果和分析

3.1　训练细节

由于无法在实际中获取大量降雨图像与其对应的干净背景图像,采用文献[24]提供的 Rain800 数据集和文献[25]提供的 Rain100H、Rain100L 数据集训练和测试相关网络。利用 Pytorch 框架搭建所提出的神经网络,使用 Adam 优化器训练模型,初始化学习率为 0.01,每 30 个 epoch 后学习率降低 10 倍,共训练 100 个 epoch。在训练过程中,式(9)的权重系数设置为 $\lambda_1 = \lambda_2 = \lambda_5 = 1$,$\lambda_3 = \lambda_4 = 0.1$。采用常用的图像评价指标峰值信噪比(PSNR)和结构相似度(SSIM)对降雨图像进行评价。将本文算法去雨结果与 2 种基于传统优化的去雨算法 DSC[6]、GMM[4]及4 种基于深度学习的去雨算法 RESCAN[26]、ReHEN[27]、SPANet[9]、UMRL[28]进行比较。

3.2　实验分析

表 1 为不同算法在测试集上去雨效果定量比较结果。可以看出,本文算法在 PSNR 和 SSIM 都取得了最好的表现,并且优势较大,说明了本文算法的有效性。

表 2 列出了不同算法的参数量及在分辨率为 320 × 640 图像上的测试时间。传统优化算法 GMM 和 DSC 由于需要学习高斯混合模型或模型字典,非常耗时。基于深度学习的算法由于有预先训练的模型可供使用,运行效率远远高于传统优化算法。得益于多尺度通道混洗深度可分离卷积的利用,本文算法在保证去雨质量的同时大大降低了网络参数,提升了模型推理效率。

表 1　不同去雨算法 PSNR 和 SSIM 对比

Table 1　PSNR and SSIM comparison of different rain removal algorithms

数据集	算法	PSNR	SSIM
Rain100H	GMM	14.262	0.544
	DSC	15.661	0.423
	RESCAN	26.451	0.846
	ReHEN	27.972	0.864
	SPANet	26.762	0.836
	UMRL	26.840	0.854
	本文	29.362	0.898
Rain100L	GMM	29.111	0.880
	DSC	24.162	0.870
	RESCAN	35.672	0.967
	ReHEN	37.412	0.980
	SPANet	35.473	0.967
	UMRL	32.305	0.966
	本文	37.524	0.985
Rain800	GMM	20.462	0.730
	DSC	18.561	0.599
	RESCAN	24.091	0.841
	ReHEN	26.961	0.854
	SPANet	25.960	0.866
	UMRL	25.155	0.879
	本文	27.155	0.884

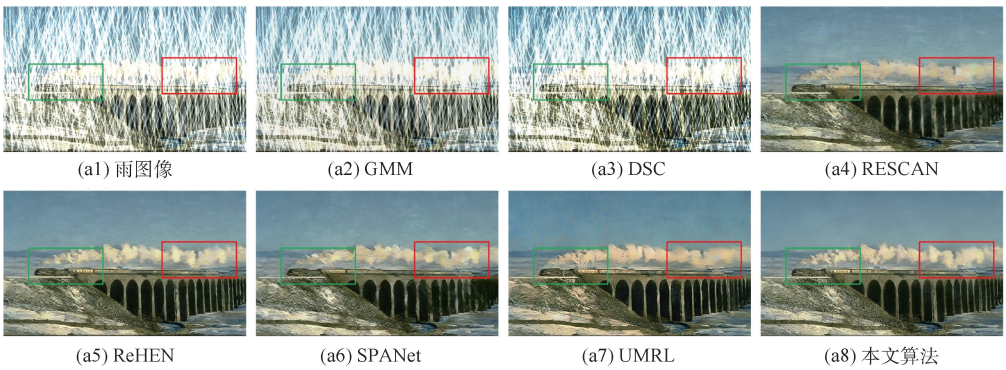
表 2　不同方法参数量与测试时间对比

Table 2　Comparison of parameters number and average running time

算法	参数量/10^6	测试时间/s
GMM		423.511
DSC		146.761
RESCAN	0.298	0.443
ReHEN	0.150	0.522
SPANet	0.219	0.145
UMRL	0.984	0.162
本文	0.112	0.122

为了直观地说明本文算法的改进效果，图6给出了2幅合成图像的去雨结果。传统优化算法DSC和GMM很难去除雨纹，背景仍然难以区分。RESCAN和SPANet中产生了较大伪影，ReHEN和UMRL丢失了边缘纹理，细节保持较差。与之对比，本文算法去掉了最多的雨条纹，同时保留了丰富的场景细节，在视觉上更令人满意。

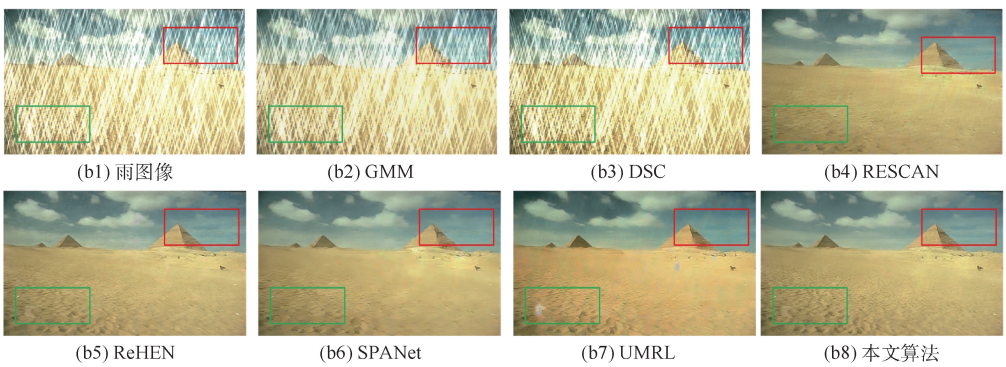
为了验证本文算法的泛化能力，图7展示了



(a1) 雨图像 (a2) GMM (a3) DSC (a4) RESCAN

(a5) ReHEN (a6) SPANet (a7) UMRL (a8) 本文算法

(a) 图像1去雨结果



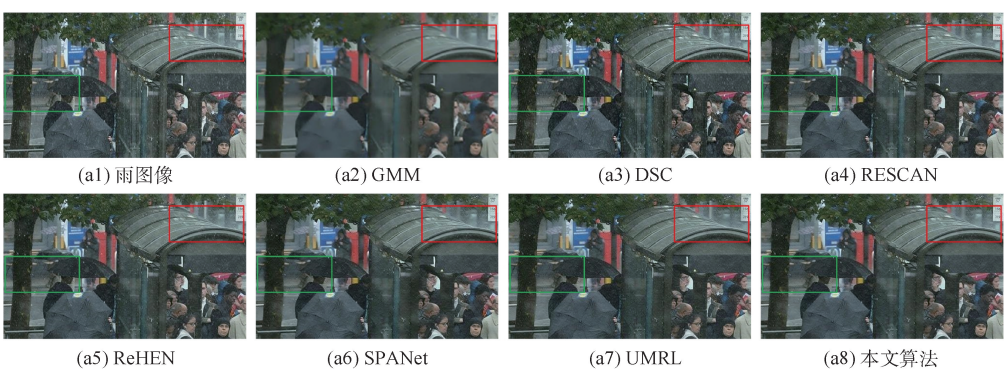
(b1) 雨图像 (b2) GMM (b3) DSC (b4) RESCAN

(b5) ReHEN (b6) SPANet (b7) UMRL (b8) 本文算法

(b) 图像2去雨结果

图6 不同算法合成数据集去雨结果可视化对比

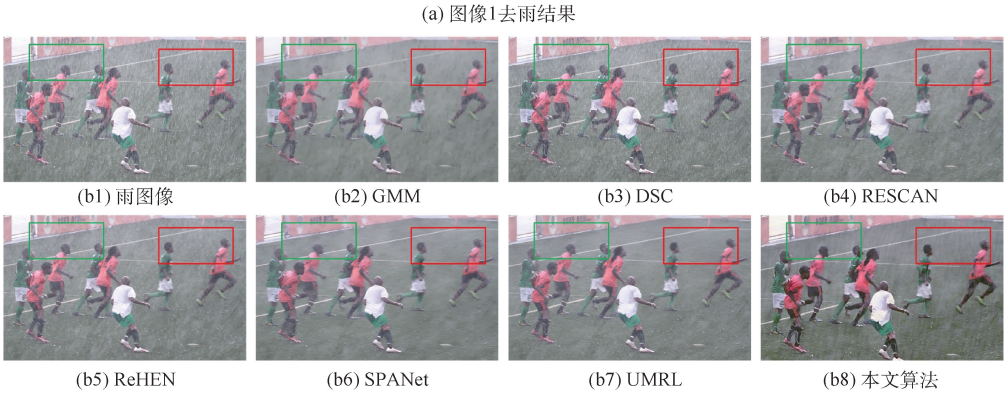
Fig. 6 Visualized comparison of different rain removal agorithms on synthetic dataset



(a1) 雨图像 (a2) GMM (a3) DSC (a4) RESCAN

(a5) ReHEN (a6) SPANet (a7) UMRL (a8) 本文算法

(a) 图像1去雨结果



(b1) 雨图像 (b2) GMM (b3) DSC (b4) RESCAN

(b5) ReHEN (b6) SPANet (b7) UMRL (b8) 本文算法

(b) 图像2去雨结果

图7 不同算法真实数据集去雨结果可视化对比

Fig. 7 Visualized comparison of different rain removal algorithms on real dataset

不同去雨算法处理真实降雨图像的结果。图像 1 中雨条纹较小,基于传统算法中的 GMM 结果使图像产生大范围的模糊,边缘细节保留较差,DSC 结果中有残留雨条纹,基于深度学习的算法取得了相近的去雨效果。图像 2 中雨条纹较大,与其他算法相比,本文算法在去除了几乎所有雨条纹的同时保留了较好的边缘细节。

为了探究阶段数量对网络去雨效果的影响,在 Rain100H 数据集上采用相同策略训练了包含 1 ~7 个阶段的网络。表 3 列出了不同网络的去雨效果。可以看出,随着去雨阶段从 1 增加到 6,去雨效果改善明显,从 6 个阶段增加到 7 个阶段,只有 SSIM 增加 0.1%,达到了网络的瓶颈,为了降低过拟合风险,兼顾去雨质量和运行效率,设定默认包含 6 个阶段。

表 3　阶段数量对去雨效果的影响

Table 3　Effect of stage number on rain removal

阶段数量	PSNR	SSIM
1	26.156	0.756
2	27.256	0.812
3	28.116	0.866
4	28.596	0.878
5	28.702	0.895
6	29.362	0.898
7	29.362	0.899

4　结　论

本文提出一种基于轻量化设计的稠密连接网络针对单幅图像进行去雨,从雨条纹检测、雨条纹去除及网络轻量化 3 方面依次展开,具体贡献如下:

1) 利用基于多尺度聚集自适应雨条纹检测模块对雨条纹进行精准定位,生成的注意图用以引导后续模块对雨条纹区域给予有意识的关注;

2) 针对图像去雨任务将稠密连接网络进行改进,通过去除归一化层增强网络泛化能力、使用空洞卷积增加网络感受野、利用挤压-激励网络增强表征能力,然后使用循环策略实现雨条纹的去除;

3) 提出多尺度通道混洗的深度可分离卷积,替代稠密连接网络中的常规卷积,在保证去雨性能的前提下大幅度降低了网络参数。与现有算法相比,本文算法在合成数据集和真实数据集上都取得了较好的去雨效果,并且运行效率高,验证了算法的效性。

参考文献 (References)

[1] LIU S, HUANG D, WANG Y. Receptive field block net for accurate and fast object detection[C] // European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 385-400.

[2] SALEH F, ALIAKBARIAN M, SALZMANN M, et al. Effective use of synthetic data for urban scene semantic segmentation[C] // European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 86-103.

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

[4] LI Y, TAN R T, GUO X, et al. Rain streak removal using layer priors[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 2736-2744.

[5] KIM J, LEE C, SIM J, et al. Single-image deraining using an adaptive nonlocal means filter[C] // IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2013: 914-917.

[6] LUO Y, XU Y, JI H. Removing rain from a single image via discriminative sparse coding[C] // International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 3397-3405.

[7] WANG Y, LIU S, CHEN C, et al. A hierarchical approach for rain or snow removing in a single color image[J]. IEEE Transactions on Image Processing, 2017, 26(8): 3936-3950.

[8] YANG W, TAN R, FENG J, et al. Joint rain detection and removal from a single image with contextualized deep networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(6): 1377-1393.

[9] WANG T Y, YANG X, XU K, et al. Spatial attentive single-image deraining with a high quality real rain dataset[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 12262-12271.

[10] REN D, ZUO W, HU Q, et al. Progressive image deraining networks: A better and simpler baseline[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 3932-3941.

[11] ZHANG D, HE J, ZHAO Y. Learning to perform joint image super-resolution and rain removal via a single-convolutional neural network[J]. Journal of Electronic Imaging, 2018, 27(6): 063024.1-063024.11.

[12] XIA H, ZHUGE R, LI H, et al. Single image rain removal via a simplified residual dense network[J]. IEEE Access, 2018, 6: 66522-66535.

[13] 王聪. 图像去雨的深度模型[D]. 大连: 大连理工大学, 2020: 5-8.
WANG C. Deep models for single image deraining[D]. Dalian: Dalian University of Technology, 2020: 5-8 (in Chinese).

[14] WANG C, XING X, WU Y, et al. DCSFN: Deep cross-scale fusion network for single image rain removal[C] // Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 1643-1651.

[15] WANG C, WU Y, SU Z, et al. Joint self-attention and scale-aggregation for self-calibrated deraining network[C] // Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 2517-2525.

[16] HUANG G, LIU Z, MAATEN L, et al. Densely connected con-

volutional networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 2261-2269.

[17] CHAI G Q, WANG Z B, GUO G D, et al. Recursive modified dense network for single image deraining[J]. Journal of Electronic Imaging, 2020, 29(3): 33006.

[18] LIU R, JIANG D, ZHANG L, et al. Deep depthwise separable convolutional network for change detection in optical aerial images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 1109-1118.

[19] 庞帆. 激光雷达惯导耦合的里程计与建图方法研究[D]. 北京: 北京建筑大学, 2020: 2-6.
PANG F. Odometer and mapping method of lidar inertial navigation coupling[D]. Beijing: Beijing University of Civil Engineering and Architecture, 2020: 2-6 (in Chinese).

[20] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2018: 6848-6856.

[21] HUANG H, HE R, SUN Z, et al. Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution[C]// International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1698-1706.

[22] JOHNSON J, ALAHI A, LI F. Perceptual losses for real-time style transfer and super-resolution[C]// European Conference on Computer Vision (ECCV). Berlin: Springer, 2016: 694-711.

[23] YANG C, MA C, YANG M. Single-image super-resolution: A benchmark[C]// European Conference on Computer Vision (ECCV). Berlin: Springer, 2014: 372-386.

[24] ZHANG H, SINDAGI V, PATEL V. Image deraining using a conditional generative adversarial network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(11): 3943-3956.

[25] YANG W, TAN R, FENG J, et al. Deep joint rain detection and removal from a singled image[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 1685-1694.

[26] LI X, WU J, LIN Z, et al. Recurrent squeeze-and-excitation context aggregation net for single image deraining[C]// European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 262-277.

[27] YANG Y, LU H. Single image deraining via recurrent hierarchy enhancement network[C]// Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1814-1822.

[28] YASARLA R, PATEL V M. Uncertainty guided multi-scale residual learning-using a cycle spinning CNN for single image deraining[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 8397-8406.

Lightweight densely connected network based on attention mechanism for single-image deraining

CHAI Guoqiang, WANG Dawei, LU Bin*, LI Zhu

(College of Physics and Information Engineering, Shanxi Normal University, Linfen 041000, China)

Abstract: The rain streaks attached to the image seriously affect the analysis and follow-up research of the image information. To restore the background texture damaged by rain streaks, this paper proposes a lightweight densely connected network based on attention mechanism to remove rain from a single image. The attention mechanism makes the network locate in the rain streaks area accurately, and the densely connected network enhances the feature reuse, alleviates the gradient disappearance and model degradation problems. The utilization of multi-scale mix channel depthwise separable convolutions realizes lightweight design by reducing the scale of network parameters and enhancing the efficiency of network operation. Both qualitative and quantitative validations on synthetic and real-world datasets demonstrate that the proposed approach can achieve competitive performance in comparison with the state-of-the-art methods.

Keywords: attention mechanism; densely connected network; lightweight design; image deraining; deep learning

Received: 2021-06-03; **Accepted**: 2021-08-27; **Published online**: 2021-09-15 13:06

URL: kns.cnki.net/kcms/detail/11.2625.V.20210914.1727.012.html

Foundation items: National Natural Science Foundation of China (62201333, 62004119); Basic Research Program of Shanxi Province (20210302124647)

* **Corresponding author**. E-mail: lubinsxnu@sina.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0676

GPS P(Y)码功率增强监测分析

李文玄[1,2,3], 焦文海[3,*], 王凯[3], 邱瑞瑾[1], 孙淑贤[2]

(1. 信息工程大学 地理空间信息学院, 郑州 450001; 2. 32021 部队, 北京 100094;
3. 北京跟踪与通信技术研究所, 北京 100094)

摘 要: 全球定位系统(GPS)的 Block IIR-M 和 Block IIF 卫星具备可编程功率输出能力,可以灵活增强单个信号分量的发射功率。为了系统评估 GPS P(Y)码的功率增强能力,对弹性功率的原理进行了理论分析,提出了 GPS 信号功率增强的监测分析方法,利用国际 GNSS 监测评估系统(iGMAS)和国际 GNSS 服务(IGS)监测站数据、高增益天线监测数据、事后精密星历对 GPS 增强 P(Y)码的覆盖性及星座性能、空间信号和用户端性能进行了分析。结果表明:在保持发射总功率和民用信号功率不变的情况下,Block IIF 和 Block IIR-M 卫星的 L1 P(Y)码和 L2 P(Y)码功率相比正常水平分别增强约 6 dB 和 5 dB;在功率增强信号覆盖区内仅利用 19 颗增强卫星进行双频单点定位,位置误差不大于 15 m(95%);当可见增强卫星数为 6,增强后的 P(Y)码载噪比为 55 dB·Hz 时,P(Y)码之间的多址干扰引起的等效载噪比下降量为 0.4 dB。

关 键 词: 全球定位系统; 弹性功率; 位置精度因子; 载噪比; 相干自适应副载波调制
中图分类号: P228; V249.3; TN967.1
文献标志码: A **文章编号**: 1001-5965(2022)11-2193-11

面对日益频发的全球导航卫星系统(global navigation satellite system, GNSS)干扰现象,卫星导航信号功率增强成为系统层面克服 GNSS 先天脆弱性、提升下行信号抗干扰能力最直接的措施,也是导航战体系的重要组成部分。全球定位系统(global postioning system, GPS) Block IIR-M 和 Block IIF 卫星现代化的重要内容之一就是增加可编程功率输出的功能,也称为弹性功率(flex power),以增加卫星播发的军用 P(Y)和 M 码的信号强度。2010 年 9 月 7 日至 12 日,在美国空军太空司令部(AFSPC)组织的 GPS 弹性功率演示训练活动中,第二太空操作中队进行了 GPS 在轨卫星的星上功率可调技术试验,Jiménez-Baños 等[1]利用伽利略实验传感器(GESS)网络的监测站数据和英国奇尔波顿天文台 25 m 抛物面天线的观测数据对试验结果进行了分析。2017 年 2 月 7 日和8 日,7 颗在轨运行的 Block IIR-M 卫星 L1 频点各个信号分量的功率发生了重新分配[2]。2017 年1 月27 日开始,第二太空操作中队对 10 颗正常运行的 Block IIF 卫星在 L1 频点上实施了区域覆盖的弹性功率操作,C/A 码和 P(Y)码的功率比在标称模式下增强约2.5 dB[3-4]。2018 年以来,GPS 功率增强行动开始进入实战,2018 年 4 月13 日至 17 日,美联合英法空袭叙利亚大马士革地区期间,GPS 信号受到人为干扰,美军对 Block IIR-M 和 Block IIF 卫星播发的 P(Y)码实施了功率增强操作,使得武器装备能够正常使用 GPS 提供的服务[3,5-6]。2019 年6 月20 日和 21 日,美军对伊朗进行空袭威胁期间,同样监测到 P(Y)码的功率增强现象[3,7]。可见,GPS P(Y)码

收稿日期: 2021-11-08; **录用日期**: 2022-02-13; **网络出版时间**: 2022-03-10 13:11
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20220309.1121.001.html
* **通信作者**. E-mail: jiaowh0927@163.com

引用格式: *李文玄, 焦文海, 王凯, 等. GPS P(Y)码功率增强监测分析[J]. 北京航空航天大学学报, 2022, 48(11): 2193-2203.*
LI W X, JIAO W H, WANG K, et al. Monitoring and analysis on GPS P(Y) code power enhancement[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2193-2203(in Chinese).

功率增强行动已成为美国军事行动的先兆和常态化操作。GPS 在轨运行的 7 颗 Block IIR-M 卫星和 12 颗 Block IIF 卫星播发的增强功率P(Y)信号,在战争和军事行动中有助于各型搭载 GPS 设备的武器装备提高抗干扰能力,在复杂电磁环境下有效确保军事行动效果。有关学者针对历次 GPS 信号功率调整和 P(Y)码功率增强现象,从功率增强强度、调制方式和差分码偏差(differential code bias, DCB)的变化等方面进行了分析[1-7]。

2021 年 6 月,随着第 5 颗 GPS III 卫星的发射入轨,目前在轨播发新一代军用 M 码信号的 GPS 卫星已达到 24 颗,但由于支持 M 码应用的地面运控系统和接收机尚没有形成完整能力[8-9],目前 P(Y)码仍是 GPS 的主用授权信号。2020 年 2 月14 日以来,GPS 对 Block IIR-M 和 Block IIF 卫星进行了常态化固定区域的弹性功率操作。

本文对 GPS 弹性功率的原理进行推导分析,综合多种数据分析 2020 年 2 月 14 日开始的 GPS P(Y)码功率增强行动,提炼 GPS 功率增强能力指标,客观评价 GPS P(Y)码功率增强能力。相关结论可为 GPS P(Y)码功率增强的实时监测分析提供技术支撑,为下一代北斗系统的信号功率设计提供相关参考,也可为干扰条件下北斗和 GPS 的功率增强对抗提供有益借鉴。

1 GPS 弹性功率原理分析

提高 GPS 的抗干扰能力是 Block IIR-M 卫星设计研发的主要目标之一,通过控制段(CS)上注指令,该型卫星搭载的新型发射机可以调整射频输出功率[10],后续的 Block IIF 卫星同样具备这种能力。在卫星发射总功率不增大的情况下,不同信号分量的功率可以实现重新分配,这种功能被称为弹性功率。

GPS 卫星在 3 个频点播发民用和军用导航信号,即 1 575.42 MHz 的 L1 频点、1 227.60 MHz 的 L2 频点和 1 176.45 MHz 的 L5 频点。所有类型的卫星均在 L1 频点上播发民用 C/A 码,在 L1 和 L2 频点上播发军用P(Y)码。2005 年 9 月首发的 Block IIR-M 卫星在 L2 频点上增加了民用 L2C 码,在 L1 和 L2 频点上增加了军用 M 码。Block IIF 卫星又增加了 L5 信号。通常情况下,GPS 卫星发射信号的总功率和功率比是恒定的。对于民用信号和 P(Y)码,发射功率满足接口规范[11-12]中给出的接收最小信号强度的要求。Barker 等[13]给出了军用 M 码的近似值。

Block IIR-M 和 Block IIF 卫星具备可编程功率输出能力,可以在 L1 和 L2 频点的各个信号之间进行发射功率的重新分配[14-15],使得单个信号分量的功率可能会超过规定的最大值,但不会超过 -150 dBW[11]。下面以 Block IIF 卫星 L1 频点分析弹性功率原理。

Block IIF 卫星 L1 频点共播发民用 C/A 码、军用 P(Y)码和 M 码 3 种信号,为了保证合路信号包络恒定,采用了相干自适应副载波调制(coherent adaptive subcarrier modulation, CASM)来解决播发新的军用 M 码的难题。L1 频点合路基带信号理论解析式[16]可由式(1)表示:

$$\begin{cases} I(t) = \sqrt{2P_{\mathrm{I}}}C_{\mathrm{P(Y)}}(t)d(t)\cos m - \\ \quad \sqrt{2P_{\mathrm{Q}}}C_{\mathrm{M}}(t)\mathrm{sqr}(t)d_{\mathrm{M}}(t)\sin m \\ Q(t) = \sqrt{2P_{\mathrm{Q}}}C_{\mathrm{C/A}}(t)d(t)\cos m + \\ \quad \sqrt{2P_{\mathrm{I}}}C_{\mathrm{P(Y)}}(t)C_{\mathrm{M}}(t)C_{\mathrm{C/A}}(t)\mathrm{sqr}(t)\sin m \end{cases} \tag{1}$$

式中:$I(t)$、$Q(t)$分别为同相和正交分量;$d(t)$和$d_{\mathrm{M}}(t)$分别为 NAV 电文数据码和新的 MNAV 电文数据码;$C_{\mathrm{P(Y)}}(t)$、$C_{\mathrm{M}}(t)$、$C_{\mathrm{C/A}}(t)$分别为 P(Y)码、M 码、C/A 码的电平值;P_{I}、P_{Q} 分别为 I、Q 支路的功率;$\mathrm{sqr}(t)$为 M 码信号扩频调制中采用的 10.23 MHz的方波;m 为调制系数。

各信号分量的功率及与信号总功率的关系可表示为

$$\begin{cases} P_{\mathrm{C/A}} = P_{\mathrm{Q}}\cos^2 m \\ P_{\mathrm{M}} = P_{\mathrm{Q}}\sin^2 m \\ P_{\mathrm{P(Y)}} = P_{\mathrm{I}}\cos^2 m \\ P_{\mathrm{IM}} = P_{\mathrm{I}}\sin^2 m \\ P_{\mathrm{T}} = P_{\mathrm{C/A}} + P_{\mathrm{M}} + P_{\mathrm{P(Y)}} + P_{\mathrm{IM}} = P_{\mathrm{I}} + P_{\mathrm{Q}} \end{cases} \tag{2}$$

式中:P_{T} 为信号发射总功率;P_{IM}为互调分量的功率。

由式(2)可知,在总功率不变的条件下,增大 P(Y)码功率可以通过减小 m、增大 P_{I}、减小 m 同时增大 P_{I} 3 种方法实现,下面分别进行分析。正常情况下,L1 频点的信号功率比和调制系数采用德国宇航中心(DLR)的高增益天线测量结果[3]:$P_{\mathrm{Q}}/P_{\mathrm{I}} \approx 65/35$,$m \approx 0.83$,$P_{\mathrm{C/A}}/P_{\mathrm{P(Y)}} \approx 2.9$ dB。

1) 减小 m。对于 L1 频点,当 m 减小为 0 时,即调制方式由 CASM 转换为四相相移键控(QPSK),此时 M 码和互调项 IM 的功率为 0,$P_{\mathrm{Q}}/P_{\mathrm{I}} = P_{\mathrm{C/A}}/P_{\mathrm{P(Y)}} \approx 2.9$ dB。根据 I 和 Q 支路及各信号分量的功率比,可知这种情况下 P(Y)码的功率增强量最大为 $\Delta P_{\mathrm{P(Y)}} = 3.4$ dB,同时 $\Delta P_{\mathrm{C/A}} = 3.4$ dB,即 P(Y)码和 C/A 码同步增强,P(Y)码和 C/A 码的功率比不变。方法 1 已在 2018 年 1 月的 Block IIF 卫星 L1 频点的各信号分量功率调整中得到验证,但由于 L1 频点总功率下降了约 0.6 ~ 1.0 dB,P(Y)码和 C/A 码的功率实际增强

约 2.5 dB[3]。

2) 增大 P_I。该情况下调制方式不变,互调分量 IM 的功率同时增加,复用效率将会下降,同时 P_Q 减小,$P_{C/A}$码的功率也减小。根据高增益天线测量结果,正常情况下,$P_I/P_T = 0.35$,$P_Q/P_T = 0.65$。极限情况下,当 $P_I = P_T$ 时,P 码功率最大只能增强 4.6 dB,但此时 P_Q 为 0,C/A 码和 M 码不能正常播发。方法 2 会影响民用用户使用 C/A 码信号。

3) 减小 m 同时增大 P_I。在方法 1 减小 m 为 0 的基础上,继续调整 I 和 Q 支路的功率比,当 P_Q/P_I 减小时,$\Delta P_{P(Y)}$将进一步增大,而 $\Delta P_{C/A}$将减小,最终可以实现 P(Y)码增强而 C/A 码功率不变的结果,此时 $\Delta P_{P(Y)} = 6.3$ dB,$P_Q/P_I \approx 31/69$,$P_{C/A}/P_{P(Y)} \approx -3.5$ dB。2020 年 2 月 14 日开始的 GPS 信号功率调整操作就是属于方法 3。

分别分析得到 Block IIF 卫星 L2 频点、Block IIR-M 卫星 L1 频点和 L2 频点在方法 3 下的功率变化情况:

1) Block IIF 卫星 L2 频点 Q 和 I 支路的功率比 $P_Q/P_I \approx 55/45$,调制系数 $m \approx 0.75$[3],按照方法 3 实现 P(Y)码增强而 C 码功率不变时,$\Delta P_{P(Y)} = 4.6$ dB,$P_Q/P_I \approx 30/70$,$P_C/P_{P(Y)} \approx -3.7$ dB。

2) Block IIR-M 卫星 L1 频点 Q 和 I 支路的功率比 $P_Q/P_I \approx 61/39$,调制系数 $m \approx 0.83$[16],按照方法 3 实现 P(Y)码增强而 C/A 码功率不变时,$\Delta P_{P(Y)} = 6.0$ dB,$P_Q/P_I \approx 28/72$,$P_{C/A}/P_{P(Y)} \approx -4.1$ dB。

3) Block IIR-M 卫星 L2 频点 Q 和 I 支路的功率比 $P_Q/P_I \approx 48/52$,调制系数 $m \approx 0.795$[16],按照方法 3 实现 P(Y)码增强而 C 码功率不变时,$\Delta P_{P(Y)} = 4.6$ dB,$P_Q/P_I \approx 23.9/76.1$,$P_C/P_{P(Y)} \approx -5.0$ dB。

可见,通过调节 m、P_I 和 P_Q 的比,可以灵活实现各信号分量功率的不同组合。方法 1 可以实现 P(Y)码和 C/A 码的同步增强,减小 M 码和互调项 IM 的功率,直至为 0;方法 2 中 P 码功率最大只能增强 4.6 dB,同时需要减小 C/A 码和 M 码的功率;方法 3 在 C/A 码功率不变的情况下,Block IIF 卫星可以实现 L1 P(Y)码功率增强 6.5 dB。3 种方法进行比较,可以看出方法 3 更符合导航战的内涵要求,无干扰时,民用用户不受弹性功率模式的影响,干扰环境下,民用用户较授权用户更容易受到干扰,弹性功率主要提高了授权用户的抗干扰能力。

GPS 具备弹性功率功能的 7 颗 Block IIR-M 卫星和 12 颗 Block IIF 卫星构成了一个功率增强星座,精确控制星座中每颗卫星弹性功率开启和终止的时刻,可以灵活产生不同的增强星座覆盖情况,达到不同的星座性能,如 2018 年 4 月、2019 年 6 月的全球覆盖和 2020 年 2 月 14 日以来的区域覆盖。

2　监测分析方法

利用全球分布的国际 GNSS 监测评估系统(iGMAS)和国际 GNSS 服务(IGS)监测站网络,实现了对 GPS P(Y)码功率增强现象的连续监测,利用事后精密星历分析 GPS 功率增强卫星的覆盖性及星座性能。使用高增益天线对功率增强前后的 GPS 射频信号进行分析,得出了信号调制方式、各信号分量功率比的变化情况。增强后P(Y)码的抗干扰能力提升程度主要通过信号分析的方法得出。

2.1　数据来源

2.1.1　监测站观测数据

iGMAS 目前已建成包括全球 27 个监测站的 GNSS 监测网络,均具备监测 BDS/GPS/GLONASS/Galileo 四系统信号的能力。IGS 及其多 GNSS 试点项目(MGEX)目前已建成全球 500 余个监测站。以上 2 个监测网络提供了分析 GPS 功率增强能力所需的观测数据。选取分布于全球的多个监测站,主要利用 2020 年 2 月 13 日至 17 日 GPS 卫星各信号分量的信号强度观测量分析功率增强能力。

2.1.2　高增益天线监测数据

中国科学院国家授时中心建设的 GNSS 空间信号质量评估系统能够实现 GNSS 空间导航信号的连续监测和信号功率的测量,系统的核心设备是位于洛南昊平站的 40 m 口径高增益天线[17]。2020 年 1 月至 3 月,该系统对 Block IIF 卫星的 L1 频点信号进行了监测,监测数据可以用来分析信号调制方式及各信号分量功率比的变化情况。

2.1.3　精密星历和精确站坐标数据

利用 2020 年 2 月 17 日的 IGS 事后精密星历进行功率增强星座的性能分析。欧洲定轨中心(CODE)第 2 092 GPS 周的站坐标用于评价增强 P(Y)码伪距单点定位的精度。IGS 事后精密星历的精度为 5 cm(RMS)、数据间隔为 15 min,站坐标的平面精度为 3 mm、高程精度为 6 mm,满足星座性能和定位精度分析的要求。

2.2　分析方法

2.2.1　全球分布的监测站数据分析

目前,iGMAS 和 MGEX 均提供 RINEX 3 格式的观测文件,其中包含了接收机输出的各个信号

分量的信号强度观测量,即载噪比密度(C/N_0)。分析 C/N_0 随时间的变化情况,若多个监测站同时观测到同一类卫星发射的相同信号超过一定阈值的 C/N_0 阶跃变化,则认为对该信号进行了功率增强,进一步可分析得到每颗卫星的功率增强操作开始和结束时刻,然后计算分析得到功率增强星座性能、信号功率增强强度、定位精度等指标参数。

1) 覆盖区域范围

利用精密星历和信号功率增强的开始、结束时刻,设置高度截止角为 5°、格网大小为 2.5°、时间间隔为 1 min,计算一个回归周期内每个空间格网点可见增强卫星数(NSAT)的 95% 分位数,然后按式(3)和式(4)计算功率增强覆盖区域面积 S 和空间覆盖率 S_p:

$$S = \sum_{i=1}^{n} S_i \cdot \mathrm{bool}\{\mathrm{NSAT}_i(95\%) \geqslant 4\} \tag{3}$$

$$S_p = S/S_g \tag{4}$$

式中:n 为空间格网点总数;i 为空间格网点序号;S_i 为第 i 个空间格网的面积;S_g 为全球表面积。

2) 可见增强卫星数

按照覆盖区域范围中的设置,计算一个回归周期内每个时空格网点的可见增强卫星数。分析全球性能时,分别计算 5°、10°、15°高度截止角下的可见增强卫星数,统计最小值、最大值和平均值。

3) 增强星座位置精度因子

格网点位置精度因子(position dilution of precision, PDOP)可用性和星座 PDOP 可用性[18]可以分别表示为

$$\begin{cases} A_j = \dfrac{N_j^{\mathrm{T}}(\mathrm{PDOP}_n)}{N_j^{\mathrm{T}}(\mathrm{PDOP})} \times 100\% \\ A_{\mathrm{R}} = \dfrac{N_{\mathrm{R}}^{\mathrm{T}}(\mathrm{PDOP}_n)}{N_{\mathrm{R}}^{\mathrm{T}}(\mathrm{PDOP})} \times 100\% \end{cases} \tag{5}$$

式中:A_j 表示特定格网点 j 的 PDOP≤n 的可用性;$N_j^{\mathrm{T}}(\mathrm{PDOP})$ 表示格网点 j 在一个回归周期内采样计算所有 PDOP 的统计数;$N_j^{\mathrm{T}}(\mathrm{PDOP}_n)$ 表示格网点 j 在一个回归周期内采样计算 PDOP≤n 的统计数;A_{R} 表示特定覆盖区域 PDOP≤n 的星座可用性;$N_{\mathrm{R}}^{\mathrm{T}}(\mathrm{PDOP})$ 表示特定覆盖区域 R 所有格网点在一个回归周期内采样计算所有 PDOP 的统计数;$N_{\mathrm{R}}^{T}(\mathrm{PDOP}_n)$ 表示特定覆盖区域 R 所有格网点在一个回归周期内采样计算 PDOP≤n 的统计数;n 为星座可用性取值。

4) 信号增强能力

C/N_0 描述了 GNSS 信号的载波功率与 1 Hz 带宽中的噪声功率之比。接收机通常将 C/N_0 作为测量值输出并作为衡量信号质量的重要参数。假设噪声功率保持恒定,则 C/N_0 也可以作为接收信号强度的指标。由于天线增益、低噪声放大器和接收机特定的 C/N_0 估计方法[19]的不同,不同监测站或者不同型号接收机的 C/N_0 测量值无法直接比较,而同一接收机的 C/N_0 变化可用于监测和评估卫星信号发射功率的变化量。

当 AS(anti-spoofing)激活时,未授权的民用接收机一般通过无码或半无码技术恢复 L2 载波,间接获取 L2 频点的载波相位测量值。较直接 P(Y)码相关法,由此产生的平方损耗会导致接收机输出的 P(Y)码载噪比下降,因此,信号功率变化前后时刻 S2W 观测值的差值不能直接代表信号功率的变化量。几种半无码技术的平方损耗曲线[20]如图 1 所示。

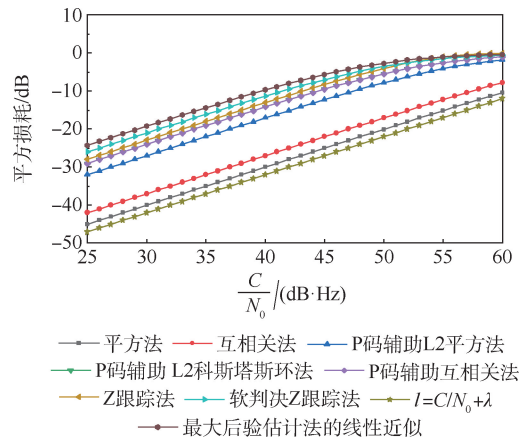


图 1 几种半无码技术的平方损耗

Fig. 1 Square loss of several semi-codeless techniques

由图 1 可知,在 L2 P(Y)码的 C/N_0 小于一定值时,所有技术的理论平方损耗曲线的斜率都近似为 1,区别只在于纵轴截距不同。不同型号的接收机可能存在差异,但只要在近似区间内,一般可以用式(6)表示的直线束来近似平方损耗曲线。

$$l = C/N_0 + \lambda \tag{6}$$

式中:l 为平方损耗;C/N_0 为直接相关法得到的 L2 P(Y)码的载噪比;λ 为常数。

由于

$$(C/N_0)_0 = C/N_0 + l \tag{7}$$

式中:$(C/N_0)_0$ 为接收机输出的 L2 P(Y)码载噪比(S2W)。

结合式(6)、式(7),可得

$$C/N_0 = [(C/N_0)_0 - \lambda]/2 \tag{8}$$

则 L2 P(Y)载噪比的变化量 $\Delta(C/N_0)$ 与 S2W 值变化量 $\Delta(C/N_0)_0$ 的关系可表示为

$$\Delta(C/N_0) = \Delta(C/N_0)_0/2 \tag{9}$$

可见,载噪比变化量与 λ 的取值无关。在平方损耗的线性区间内按式(9)计算的载噪比变化量即代表了 L2 P(Y)码的功率变化量。

5) 定位误差

利用监测站观测的增强卫星伪距进行单点定位,并与精确站坐标进行比较,得到站心坐标系中的水平、高程和位置误差。

2.2.2 高增益天线测量

民用接收机可以利用半无码技术跟踪 GPS L2 P(Y)码,一般不输出 L1 P(Y)码的观测值。具有较大有效接收面积的抛物面天线可以实现在射频端将处于噪底之下的微弱 GPS 信号提升到噪底之上。利用星座图和极大似然法[21]分析得到各频点信号分量功率比和信号调制方式的变化情况。

2.2.3 增强信号分析

利用分析得到的可见增强卫星数和信号增强能力,通过计算等效载噪比,可以对用户端的抗干扰能力提升程度和对民用信号的影响进行针对性的分析。

1) 用户端抗干扰能力提升程度

接收机同时接收多颗卫星播发的功率增强信号时,对其中某颗增强卫星信号而言,其他增强卫星信号对其构成多址干扰。功率增强情况下,各颗卫星增强 P(Y)码间的多址干扰视为匹配谱干扰,用户端 P(Y)码的等效载噪比可按式(10)计算:

$$(C/N_0)_{\text{eff}} = -10\lg\left[10^{-[C/N_0+\Delta P]/10} + \frac{N-1}{QR_{\text{C}}}\right] \tag{10}$$

式中:$(C/N_0)_{\text{eff}}$为 P(Y)码的等效载噪比;C/N_0为未增强 P(Y)码的载噪比;ΔP为功率增强量;N为可见增强卫星数;Q为抗干扰品质因数,此处取为 1.5[22];R_{C}为 P(Y)码的扩频码速率。

2) 对其他信号的影响程度

由于 L1 频点上的 C/A 码和 P(Y)码的频谱中心重叠,增强 P(Y)码信号功率将会对 C/A 码信号的接收造成影响。增强 P(Y)码对 C/A 码的影响类似于宽带干扰,用户端的 C/A 码等效载噪比可表示为

$$(C/N_0)_{\text{eff}} = -10\lg\left[10^{-(C/N_0)/10} + \frac{\sum_{i=1}^{N} 10^{(J/S)_i/10}}{QR_{\text{C}}}\right] \tag{11}$$

式中:$(C/N_0)_{\text{eff}}$为 C/A 码的等效载噪比;C/N_0为 C/A 码无干扰时的载噪比;N为可见增强卫星数;J/S为增强 P(Y)码和 C/A 码的功率比;Q为抗干扰品质因数,此处取为 2.22[22];R_{C}为 C/A 码的扩频码速率。

受 P 码干扰后 C/A 码的载噪比下降量 $\Delta(C/N_0)$可按式(12)计算:

$$\Delta(C/N_0) = C/N_0 - (C/N_0)_{\text{eff}} \tag{12}$$

3 GPS P(Y)码功率增强能力

3.1 覆盖性及星座性能

3.1.1 覆盖区域范围

分析全球分布的监测站数据中每颗卫星 P(Y)码的 C/N_0 随时间的变化情况,得到 7 颗 Block IIR-M 卫星和 12 颗 Block IIF 卫星在 2020 年 2 月 17 日的功率增强操作开始和结束时刻,如表 1 所示。Esenbuğa等[23]对星下点轨迹和覆盖区

表 1　2020 年 2 月 17 日弹性功率开启和终止时刻(GPST)

Table 1　Flex power on and off moments on Feb. 17th, 2020 (GPST)

卫星编号	卫星类型	开始时刻	结束时刻	开始时刻	结束时刻	开始时刻	结束时刻
G01	Block IIF	00:00	05:41	09:23	13:22	23:22	00:00
G03	Block IIF	00:16	07:53	13:09	14:45		
G05	Block IIR-M	06:59	11:44	14:44	20:48		
G06	Block IIF	03:56	08:08	11:44	17:54		
G07	Block IIR-M	04:13	12:47				
G08	Block IIF	00:00	02:55	06:04	11:09	21:20	00:00
G09	Block IIF	02:36	10:47				
G10	Block IIF	00:00	03:32	19:07	23:56		
G12	Block IIR-M	11:41	19:34				
G15	Block IIR-M	09:36	15:17	17:57	23:17		
G17	Block IIR-M	01:54	04:43	09:17	16:35		
G24	Block IIF	00:00	1:10	11:10	17:30	21:04	00:00
G25	Block IIF	01:07	03:56	13:33	20:31		
G26	Block IIF	02:15	07:33	17:46	23:07		
G27	Block IIF	00:00	02:02	05:32	09:51	19:55	00:00
G29	Block IIR-M	14:18	22:37				
G30	Block IIF	05:31	13:59				
G31	Block IIR-M	00:01	06:13	16:02	20:16		
G32	Block IIF	00:00	04:39	14:08	16:25	21:14	00:00

域中心坐标进行了分析。利用这些卫星的精密星历和功率增强操作时刻进行轨道仿真,以可见增强卫星数 NSAT(95%)≥4 作为判断条件,按式(3)和式(4)计算得到 2020 年 2 月 17 日一个回归周期内的 7 颗 Block IIR-M 卫星和 12 颗 Block IIF 卫星的覆盖区域面积为 66 378 228.25 km^2,覆盖率为 13%。覆盖区域范围如图 2 所示。

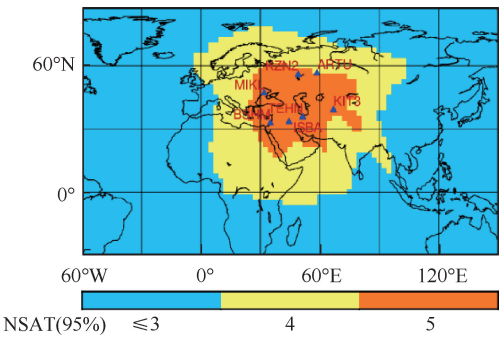


图 2 功率增强星座覆盖区域范围和可见增强卫星数

Fig. 2 Coverage area range and NSAT of power enhancement constellation

3.1.2 可见增强卫星数

高度截止角设为 5°,计算得到 2020 年 2 月 17 日的可见增强卫星数(95%),如图 2 所示。在覆盖区域范围内,可见增强卫星数为 5 颗(95%)的区域和可见增强卫星数为 4 颗(95%)的区域的面积占比分别为 30% 和 70%。

为分析 GPS 功率增强星座的全球覆盖性,假设 7 颗 Block IIR-M 卫星和 12 颗 Block IIF 卫星在一个回归周期内不间断播发功率增强信号,得到全球可见增强卫星数,如图 3 所示。分别将高度截止角设为 5°、15°、30°,统计出全球可见增强卫星数的最小值、最大值、平均值,如表 2 所示。

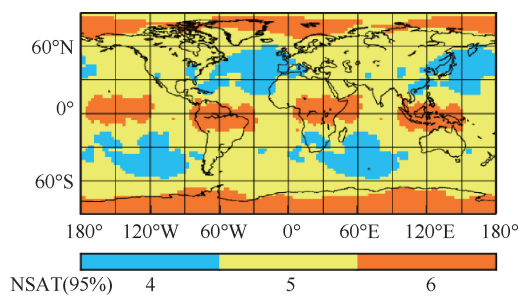


图 3 功率增强星座全球可见增强卫星数

Fig. 3 Global NSAT of power enhancement constellation

3.1.3 增强星座位置精度因子

高度截止角设为 5°,计算得到 2020 年 2 月 17 日增强星座 PDOP(95%),如图 4 所示。

假设 19 颗卫星在一个回归周期内不间断播发功率增强信号,进一步分析得到增强星座全球位置精度因子 PDOP(95%),如图 5 所示。GPS 星座不是 Walker 型星座,并且功率增强卫星数量仅为 19 颗,因此 PDOP 值在同纬度地区的分布并不相同,PDOP(95%)最大不超过 7。分别将高度截止角设为 5°、15°、30°,统计出全球 PDOP≤6 星座可用性,如表 3 所示,5°高度截止角下的星座可用性可达 98.6%。

表 2 不同高度截止角下的全球可见增强卫星数

Table 2 Global NSAT at different elevation mask angle

高度截止角/(°)	NSAT	
5	最小	3
	最大	10
	平均	6.4
15	最小	2
	最大	9
	平均	5
30	最小	0
	最大	7
	平均	3.1

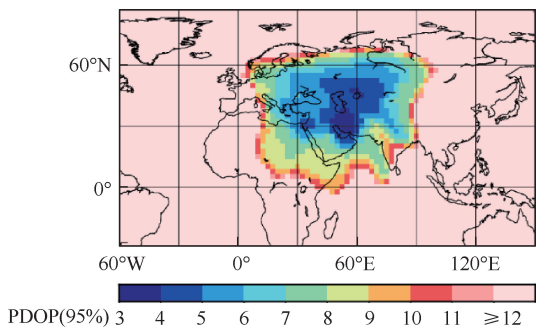


图 4 增强星座区域位置精度因子

Fig. 4 PDOP of enhanced constellation area

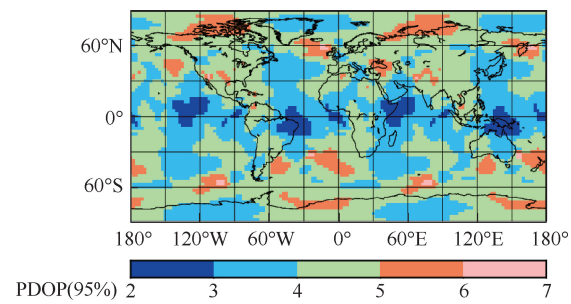


图 5 增强星座全球位置精度因子

Fig. 5 Global PDOP of enhanced constellation

表 3 PDOP≤6 星座可用性

Table 3 PDOP≤6 constellation availability

高度截止角/(°)	星座可用性/%
5	98.6
15	81.4
30	14.8

3.2　空间信号

3.2.1　信号增强能力

图 6 绘出了 IGS ABPO 站观测的 2020 年 2 月 13 日(年积日 044)和 14 日(年积日 045)Block IIR-M G17 卫星 L1 C/A、L2 P(Y)、L2 C 的信号强度 S1C、S2W、S2L。2020 年 2 月 13 日 Block IIR-M G17 卫星各信号分量的强度都保持连续状态,2 月 14 日 09:34,L2 P(Y)的信号强度出现了瞬时增大的现象,表明从该时刻起,Block IIR-M G17 卫星开启了 L2 P(Y)信号功率的增强。L2 C 在 2 月13 日和 2 月 14 日两天的信号强度基本保持一致。L1 C/A 在 2 月 14 日的信号强度相比 2 月 13 日有所减小,平均减小约 2.4 dB,进一步分析发现,从 2 月 14 日 00:00 起,Block IIR-M 卫星 L1 C/A 的信号强度都出现了减小,这表明自 2017 年 2 月 7 日开始的 7 颗 Block IIR-M 卫星 L1 频点各信号分量重新调整的操作[2]于 2 月 14 日 00:00 终止,Block IIR-M 卫星的 C/A 码功率恢复到了正常状态。分析发现,S2W 和 S1C、S2L 观测值的曲线不平行,其差距和 S2W 值的大小相关,这是由于普通民用接收机使用半无码技术恢复 L2 P(Y)码产生的平方损耗引起的,S2W 值要低于采用直接相关法得到的 L2 P(Y)码载噪比,且其差值是非线性的。

图 7 绘出了 IGS MCIL 站观测的 2020 年 2 月 13 日(年积日 044)和 14 日(年积日 045)Block IIF G24 卫星 L1 C/A、L2 P(Y)、L2 C 的信号强度 S1C、S2W、S2X。S2X 在 2 月 13 日和 2 月 14 日两天基本保持一致。2 月 13 日的 S1C 和 S2W 值在 00:34 均出现约 2.5 dB 的减小,在 23:11 均出现约2.5 dB的增大,这是由于 2017 年 1 月 27 日开始的对 Block IIF 卫星 L1 频点各信号分量功率的调整造成的,由于半无码技术,L1 C/A 码功率的增强也会造成接收机输出的 S2W 增大,这种调整于 2020 年 2 月 14 日终止,被新的功率调整方式代替。2 月 14 日的 S1C 保持连续状态,表明 L1 C/A 的信号强度恢复到了正常水平。2020 年 2 月14 日 00:00 起,S2W 瞬时增大,01:26 恢复,21:20 又增大,表明从 2 月 14 日起,Block IIF G24 卫星开启了区域覆盖的 L2 P(Y)信号功率的增强。

考虑卫星低仰角时的多路径效应和平方损耗曲线的近似区间问题,以 S2W 属于[38,45]dB·Hz 区间确定计算各信号分量功率变化量的时段,在此时段内对 2 月 13 日和 2 月 14 日两天 L1 C/A、L2 C 的载噪比作差,L2 P(Y)的载噪比变化量按式(9)计算,Block IIR-M G17 卫星和 Block IIF G24 卫星的结果分别如图 8 和图 9 所示。图中,各个历元的载噪比差值基本保持恒定。

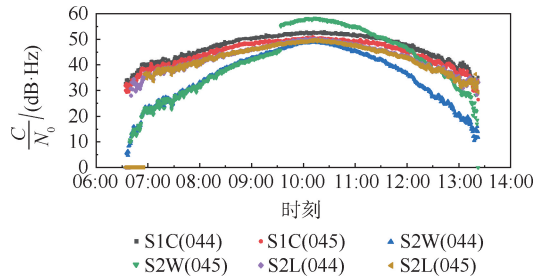


图 6　Block IIR-M G17 卫星各信号分量强度(IGS ABPO 站)

Fig. 6　Strength of each signal component of Block IIR-M G17 satellite (IGS ABPO)

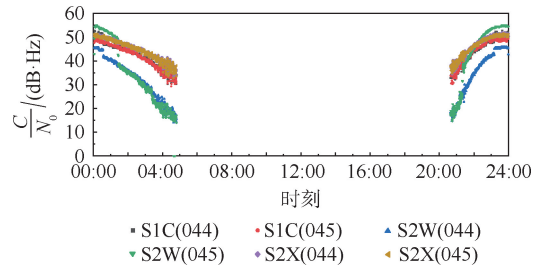


图 7　Block IIF G24 卫星各信号分量强度(IGS MCIL 站)

Fig. 7　Strength of each signal component of Block IIF G24 satellite (IGS MCIL)

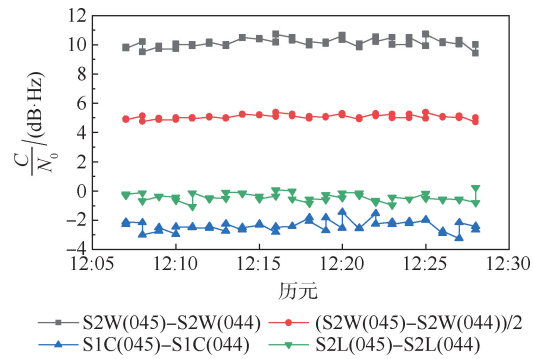


图 8　Block IIR-M G17 卫星各信号分量强度的变化

Fig. 8　Variation of strength of each signal component of Block IIR-M G17 satellite

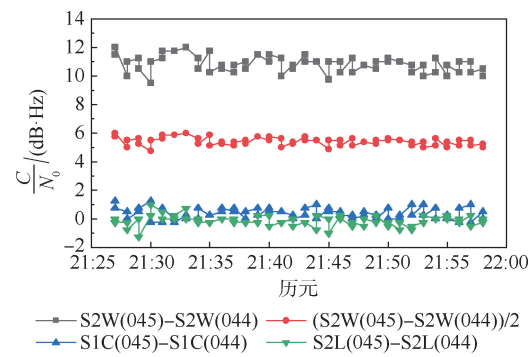


图 9　Block IIF G24 卫星各信号分量强度的变化

Fig. 9　Variation of strength of each signal component of Block IIF G24 satellite

对各历元的载噪比差值取中数,结果如表 4 所示。结果表明,Block IIR-M G17 卫星的 L2 P(Y)的功率增强了约 5.1 dB。L1 C/A 码的功率减小了约 2.4 dB,考虑到 2017 年 2 月 7 日开始的 Block IIR-M 卫星 L1 C/A 码功率增强2.4 dB的现象,经这次调整,L1 C/A 码的功率恢复到了正常水平。Block IIF G24 卫星的 L2 P(Y)码功率增强了约 5.4 dB。

3.2.2 信号调制方式及信号分量功率比

常用测量型接收机一般只输出 L2 P(Y)码观测量。2020 年 1 月至 3 月使用高增益天线对8 颗 Block IIF 卫星的下行信号进行了监测,分析得到了 L1 频点信号调制方式、L1 P(Y)码和 L1 C/A 码功率比的变化情况,如表 5 所示。

表 4 Block IIR-M G17 和 Block IIF G24 卫星各信号分量功率变化量

Table 4 Power variation of each signal component of Block IIR-M G17 and Block IIF G24 satellite

PRN	信号分量功率变化量/dB		
	L2 P(Y)	L1 C/A	L2 C
Block IIR-M G17	5.1	-2.4	-0.4
Block IIF G24	5.4	0.4	-0.1

表 5 Block IIF 卫星 L1 频点信号调制方式和功率比变化

Table 5 Variation of signal modulation mode and power ratio at L1 frequency band of Block IIF satellites

PRN	调整前调制方式	调整后调制方式	调整前 $P_{P(Y)}:P_{C/A}$/dB	调整后 $P_P:P_{C/A}$/dB	$\Delta P_P:P_{C/A}$/dB
G03	CASM	QPSK	-2.99	3.54	6.53
G09	QPSK	QPSK	-3.05	3.66	6.71
G24	QPSK	QPSK	-3.02	3.46	6.48
G25	CASM	QPSK	-2.74	3.94	6.68
G26	QPSK	QPSK	-3.09	3.42	6.51
G27	CASM	QPSK	-2.88	3.81	6.69
G30	QPSK	QPSK	-3.07	3.48	6.55
G32	CASM	QPSK	-3.31	3.63	6.94
平均值			-3.02	3.62	6.64

2020 年 1 月 16 日至 21 日的监测数据表明,8 颗 Block IIF 卫星中有 4 颗的 L1 频点是 CASM 调制,另外 4 颗为 QPSK 调制。结合监测站的信号强度观测量,可以得出采用 CASM 调制的 4 颗卫星的 L1 频点各信号分量功率比已经恢复正常状态。而采用 QPSK 调制的 G24、G30、G09、G26 等 4 颗卫星仍然处于弹性功率模式,信号分量功率配比与 2017 年 1 月 27 日相同。分析 2020 年 2 月 29 至 3 月 16 日的监测数据,发现当 8 颗 Block IIF 卫星进入特定空域后开启弹性功率时,L1 频点的调制方式均调整为 QPSK,所有卫星的 L1 P(Y)和 L1 C/A 的功率比的变化情况基本相同。由于 C/A 码功率基本未变,相当于 P(Y)码功率平均增强了 6.64 dB,这与第一节中的理论分析结果基本一致。

3.2.3 增强频点数量

综合 IGS ABPO 站、MCIL 站观测的 L2 P(Y)信号强度分析结果(见表 4)和由高增益天线监测数据分析得出的 L1 频点信号分量功率比(见表 5),可知 L1 和 L2 频点的 P(Y)码均发生了功率增强。

3.3 用户端性能

3.3.1 抗干扰能力提升程度

结合 3.1.2 节中的结论,在式(10)中将可见增强卫星数 NSAT 设为 3 ~ 6,计算得到接收机端的增强 P(Y)码等效载噪比 $(C/N_0)_{P(Y),\mathrm{eff}}$,如图 10所示。图 10 表明,功率增强情况下,由于多址干扰,接收机端的增强 P(Y)码等效载噪比 $(C/N_0)_{P(Y),\mathrm{eff}}$小于未增强 P(Y)码载噪比与功率增强量之和:

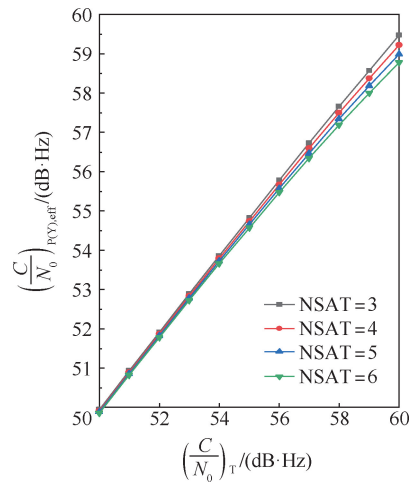


图 10 P(Y)码的等效载噪比

Fig. 10 Equivalent carrier-to-noise ratio of P(Y) code

$$(C/N_0)_T = (C/N_0)_{P(Y)} + \Delta P$$

其差值随 NSAT 和 $(C/N_0)_T$ 的增大而增大。当 NSAT 为 6,$(C/N_0)_T$ 为 55 dB · Hz 时,等效载噪比下降量为 0.4 dB;$(C/N_0)_T$ 为 60 dB · Hz 时,等效载噪比下降量可达 1.2 dB。

3.3.2 对系统内其他信号的影响程度

在特定的弹性功率模式下,增强 P(Y)码和 C/A 码的功率比是恒定的,以 Block IIF 卫星 L2

频点为例,在式(11)中将 J/S 设为 3.7 dB,计算得到增强 P(Y)码对 C/A 码干扰引起的载噪比下降量 $\Delta(C/N_0)$,如图 11 所示。可见,C/A 码载噪比下降量 $\Delta(C/N_0)$ 随可见增强卫星数 NSAT 的增大而增大,当 NSAT 一定时,$\Delta(C/N_0)$ 随 C/A 码载噪比 $(C/N_0)_{C/A}$ 的增大而增大。当 NSAT = 6,$(C/N_0)_{C/A}$ 为 50 dB·Hz 时,载噪比下降量可达 2.1 dB;$(C/N_0)_{C/A}$ 在 46 dB·Hz 以下时,载噪比下降量小于 1 dB。

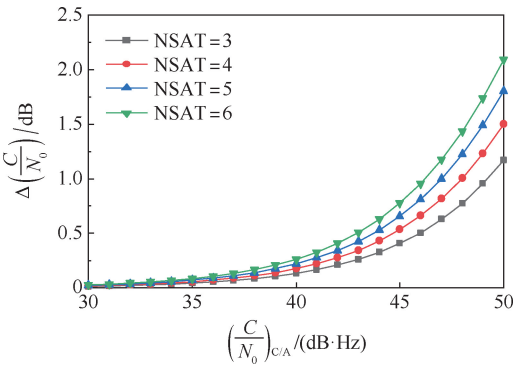


图 11　增强 P(Y)码引起的 C/A 码载噪比下降量

Fig. 11　Degradation of C/A code carrier-to-noise ratio caused by enhanced P(Y) code

3.3.3　定位误差

选择 GPS 增强 P(Y)码覆盖区域内的 7 个 IGS 监测站(点位分布见图 2),仅利用 19 颗功率增强卫星 2020 年 2 月 14 日的 L1 C/A 和 L2 P(Y)伪距观测量进行双频单点定位解算,高度截止角设为5°,采样间隔为 30 s,电离层延迟采用消电离层组合改正,对流层延迟采用萨斯塔莫宁(Saastamoinen)模型改正。进行单历元解,每个历元的解与 CODE 的周解站坐标作差,而后统计站点在站心坐标系下的水平、高程和位置误差(95%),各站点的定位误差如图 12 所示。

从图 12 可以看出,所有站点的水平误差均小于高程误差,7 个站点中的最大水平误差不大于 8 m,最大位置误差不大于 15 m。

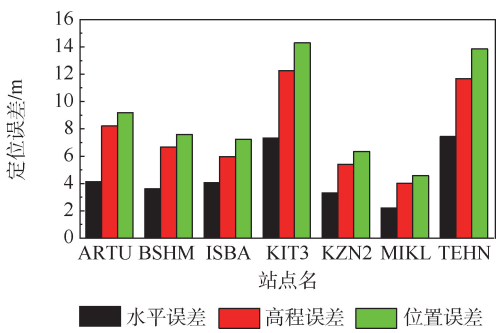


图 12　各站点定位误差(95%)

Fig. 12　Positioning error of each station(95%)

4　结　论

1) GPS 的 P(Y)码功率增强操作实现了操作时刻的精确计算与上注,增强信号覆盖区域的精确控制,验证了 GPS 对各频点信号分量功率的灵活控制能力,表明了地面段和空间段的同步协调能力达到了较高水平。

2) 在保持发射总功率和民用信号功率不变的情况下,Block IIF、Block IIR-M 卫星的 L1 P(Y)码和 L2 P(Y)码功率相比正常水平分别增强约 6 dB和 5 dB。

3) 高度截止角为 5°时,7 颗 Block IIR-M 卫星和 12 颗 Block IIF 卫星构成的功率增强星座的全球位置精度因子 PDOP≤6 星座可用性可达 98.6%。在功率增强覆盖区内选择 7 个站点,仅利用 19 颗增强卫星的观测量进行双频单点定位,位置误差不大于 15 m(95%)。

4) 当可见增强卫星数为 6,增强后的 P(Y)码载噪比为 60 dB·Hz 时,P(Y)码之间的多址干扰引起的等效载噪比下降量可达 1.2 dB。当增强 P(Y)码功率高于 C/A 码 3.7 dB,C/A 码载噪比低于 46 dB·Hz 时,增强 P(Y)码引起的 C/A 码载噪比下降量在 1 dB 以内。

致谢　感谢 iGMAS 和 IGS 提供监测站数据、精密星历和精确站坐标,感谢中国科学院国家授时中心洛南昊平站提供高增益天线监测数据。

参考文献(References)

[1] JIMÉNEZ-BAÑOS D, PERELLÓ-GISBERT J V, CRISCI M. The measured effects of GPS flex power capability collected on sensor station data[C]//2010 5th ESA Workshop on Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC). Piscataway: IEEE Press,2010:1-6.

[2] THOELERT S, HAUSCHILD A, STEIGENBERGER P, et al. GPS IIR-M L1 transmit power redistribution: Analysis of GNSS receiver and high-gain antenna data[J]. Navigation-Journal of the Institute of Navigation,2018,65(3):423-430.

[3] STEIGENBERGER P, THÖLERT S, MONTENBRUCK O. Flex power on GPS Block IIR-M and IIF[J]. GPS Solutions,2019,23(8):1-12.

[4] ESENBUĞA Ö G, HAUSCHILD A. Impact of flex power on GPS Block IIF differential code biases[J]. GPS Solutions,2020,24(4):1-9.

[5] 刘苗苗,焦文海,贾小林. 战时叙利亚地区 GPS 定位结果分析[J]. 测绘科学与工程,2018,38(4):19-25.
LIU M M, JIAO W H, JIA X L. Analysis of GPS positioning results in wartime Syria[J]. Geomatics Science and Engineering,

2018,38(4):19-25(in Chinese).

[6] 韩奇,朱克家,付钰,等.美国打击叙利亚期间GPS信号监测评估[J].导航定位学报,2019,7(3):7-10.
HAN Q,ZHU K J,FU Y,et al. Monitoring and assessment of GPS signals during US attacking on Syria[J]. Journal of Navigation and Positioning,2019,7(3):7-10(in Chinese).

[7] 刘苗苗,焦文海,贾小林.美伊冲突中的GPS信号增强分析[J].全球定位系统,2020,45(1):31-36.
LIU M M,JIAO W H,JIA X L. GPS signal enhancement analysis in the US-Iranian conflict[J]. GNSS World of China,2020,45(1):31-36(in Chinese).

[8] US Government Accountability Office. GPS modernization:DOD continuing to develop new jam-resistant capability,but widespread use remains years away:GAO-21-145[R]. Washington,D. C.:US Government Accountability Office,2021:35.

[9] CAMERON A. Benefits coming from GPS III constellation[J/OL]. GPS World,2019(2019-04-01)[2021-11-01]. https://www.gpsworld.com/benefits-coming-from-gps-iii-constellation/.

[10] RAJAN J A,TRACY J A. GPS IIR-M:Modernizing the signal-in-space[C]// Proceedings of the 2003 National Technical Meeting of the Institute of Navigation,2003:484-493.

[11] GPS Joint Program Office. Navstar GPS space segment/navigation user segment interfaces:IS-GPS-200J[S]. EI Segundo:GPS Joint Program Office,2018.

[12] GPS Joint Program Office. Navstar GPS space segment/user segment L5 interfaces:IS-GPS-705E[S]. EI Segundo:GPS Joint Program Office,2018.

[13] BARKER B C,BETZ J W,CLARK J E,et al. Overview of the GPS M code signal[C]// Proceedings of the 2000 National Technical Meeting of the Institute of Navigation,2000:542-549.

[14] RAJAN J A,IRVINE J. GPS IIR-M and IIF:Payload modernization[C]// Proceedings of the 2005 National Technical Meeting of the Institute of Navigation,2005:508-514.

[15] US Defense Science Board. The future of the global positioning system technical report[R]. US Defense Science Board,2005:32.

[16] PARTRIDGE M D,DAFESH P A. Code power measurement methodology for GPS Block IIR-M and IIF on-orbit test procedures[C]// Proccedings of the 14th International Technical Meeting of the Satellite Division of the Institute of Navigation,2001:2764-2772.

[17] 饶永南,王萌,康立,等.GPS Ⅲ首星空间信号质量监测评估[J].电子学报,2020,48(2):407-411.
RAO Y N,WANG M,KANG L,et al. Signal-in-space quality monitoring and assessment for the first GPS III satellite[J]. Acta Electronica Sinica,2020,48(2):407-411(in Chinese).

[18] 帅平,曲广吉.基于半分析式方法的全球导航星座设计[J].中国空间科学技术,2006(4):11-19.
SHUAI P,QU G J. Global navigation constellation design based on semi-analysis method[J]. Chinese Space Science and Technology,2006(4):11-19(in Chinese).

[19] FALLETTI E,PINI M,PRESTI L. Low complexity carrier-to-noise ratio estimators for GNSS digital receivers[J]. IEEE Transactions on Aerospace and Electronic Systems,2011,47(1):420-437.

[20] WOO K. Optimum semicodeless carrier-phase tracking of L2[J]. Navigation-Journal of the Institute of Navigation,2000,47(2):82-99.

[21] 康立,饶永南,王雪,等.GPSBIIF-1卫星L1频点QPSK VS CASM信号质量评估[J].宇航学报,2019,40(1):102-108.
KANG L,RAO Y N,WANG X,et al. QPSK VS CASM signals quality assessment on GPSBIIF-1 satellite L1[J]. Journal of Astronautics,2019,40(1):102-108(in Chinese).

[22] KAPLAN E D,HEGARTY C J. GPS原理与应用(第2版)[M].寇艳红,译.北京:电子工业出版社,2007:194.
KAPLAN E D,HEGARTY C J. Understanding GPS principles and applications(2nd ed)[M]. KOU Y H,translated. Beijing:Publishing House of Electronics Industry,2007:194(in Chinese).

[23] ESENBUGA Ö G,HAUSCHILD A,STEIGENBERGER P. Impact of GPS flex power on differential code bias estimation for Block IIR-M and IIF satellites[C]// Proceedings of the 2020 National Technical Meeting of the Institute of Navigation,2020:2922-2930.

Monitoring and analysis on GPS P(Y) code power enhancement

LI Wenxuan[1,2,3], JIAO Wenhai[3,*], WANG Kai[3], QIU Ruijin[1], SUN Shuxian[2]

(1. Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China;
2. 32021 Troops, Beijing 100094, China;
3. Beijing Institute of Tracking and Telecommunication Technology, Beijing 100094, China)

Abstract: With programmable power output capabilities, global positioning system (GPS) Block IIR-M and Block IIF satellites can flexibly enhance transmit power of individual signal components. In order to systematically evaluate the power enhancement capability of GPS P(Y) codes, the theoretical analysis of the flex power principle was conducted, and a monitoring and analysis method of GPS signal power enhancement was proposed. Furthermore, the coverage, constellation performance, signal-in-space and user-side performance of power-enhanced P(Y) codes were analyzed based on the data of the International GNSS monitoring and assessment system (iGMAS) and international GNSS service (IGS) monitoring station, high-gain antenna monitoring data, and precision ephemeris. According to the findings, it is possible to increase the L1 P(Y) code and L2 P(Y) code power of Block IIF and Block IIR-M satellites by about 6 dB and 5 dB, respectively, compared with the normal level, while keeping the total transmit power and civil signal power unchanged. With respect to the dual-frequency single-point positioning test only using 19 enhanced satellites, the positioning accuracy is no more than 15 m (95%) in the power-enhanced signal coverage area. While the equivalent carrier-to-noise ratio reduction caused by multi-access interference between P(Y) codes is 0.4 dB, when there are 6 visible enhanced satellites and the enhanced P(Y) code carrier-to-noise ratio is 55 dB · Hz.

Keywords: global positioning system; flex power; position dilution of precision; carrier-to-noise ratio; coherent adaptive subcarrier modulation

Received: 2021-11-08; **Accepted**: 2022-02-13; **Published online**: 2022-03-10 13:11
URL: kns.cnki.net/kcms/detail/11.2625.V.20220309.1121.001.html
* **Corresponding author**. E-mail: jiaowh0927@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0085

基于 ESO 的超局部无模型 PMSM 转速预测控制

许令亮, 陈桂明*, 李乔扬
(火箭军工程大学 作战保障学院, 西安 710025)

摘 要: 针对永磁同步电机(PMSM)由于参数变化和外界干扰导致控制性能下降的问题,提出了一种基于扩展状态观测器(ESO)的超局部无模型转速预测控制(MFSPC)方法,并解决了数字系统中的一拍延迟问题。该方法仅利用速度环的输入和输出,不考虑电机参数,避免了因电机参数变化导致的模型失配。针对传统 MFSPC 方法参数调整多、工作量大、输出脉动大、抖振明显、抗干扰性和鲁棒性差的问题,建立了 ESO,实时监测系统的总扰动,并进行前馈补偿。利用 ESO 产生的转速预测,解决了数字控制系统中的一拍延时问题,并基于频域分析,整定控制参数,提高控制性能。实验结果分析表明:所提方法有较强的抗干扰性和鲁棒性,能够稳定跟踪额定转速,有较快的动态响应。

关 键 词: 永磁同步电机; 超局部; 无模型转速预测控制; 模型失配; 扩张状态观测器; 鲁棒性
中图分类号: TM341
文献标志码: A **文章编号**: 1001-5965(2022)11-2204-11

永磁同步电机(permanent magnet synchronous motor, PMSM) 因其结构简单、体积小和功率密度高等优点,被广泛应用于运动系统中[1-2]。同时,PMSM 也是一个多变量、强耦合、非线性的复杂系统,因此,需采用合理的控制方法,提高 PMSM 的控制性能[3-4]。

传统的控制方法包括矢量控制和直接转矩控制(direct torque control, DTC)。矢量控制借鉴直流电机控制方法,建立旋转坐标变换器,实现交轴和直轴电流的解耦,进而实现磁场和转矩的解耦。但是随着电机的运行,电机参数会受到温升、磁场饱和运行状态等影响,产生参数漂移,导致模型失配,降低电机的控制性能,此时矢量控制技术难以满足实际工作需要[5-6]。为了解决这个问题,马玉梅等[7]将滑模控制和比例积分(proportional integral, PI) 控制结合,有效利用滑模控制对参数变化的不敏感性和 PI 控制的稳定性,结果表明,该方法能够有效跟踪额定转速,启动时间短,超调量较小。DTC 控制摒弃了矢量控制解耦的思想,取消了旋转坐标变换器,减弱了控制模型对参数的依赖,建立磁链和转矩调节器,输出所需的的空间电压。但是,DTC 控制存在逆变器开关不恒定、输出转矩磁链脉动较大及因转矩脉动带来的高频噪声等[8-9]。Xia 等[10]通过构造增强型切换表,建立了转矩、磁通变化率和矩阵变换器之间的解析表达式,无须坐标转换,并且具有固定的逆变器开关频率,结果表明,该方法能够有效降低转矩脉动。

相比于传统控制方法,模型预测控制方法因其结构简单,易于实现,控制效果好,受到国内外学者广泛关注。模型预测控制方法利用控制对象的数学模型,离散化得到下一时刻预测值,再通过

收稿日期: 2021-02-22; **录用日期**: 2021-07-03; **网络出版时间**: 2021-07-13 17:02
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210713.1419.002.html
基金项目: 军内科研项目 (20191C050239)
* **通信作者**. E-mail: 792757066@qq.com

引用格式: 许令亮, 陈桂明, 李乔扬. 基于 ESO 的超局部无模型 PMSM 转速预测控制[J]. 北京航空航天大学学报, 2022, 48(11): 2204-2214. XU L L, CHEN G M, LI Q Y. Ultra-local model-free speed predictive control based on ESO for PMSM[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2204-2214(in Chinese).

优化成本函数,使得预测值沿着参考轨迹趋近于期望值。但是模型预测控制方法依赖于控制对象的数学模型,在电机运转过程中,会受到内部参数漂移和外部未知干扰,导致控制性能下降,控制精度降低[11-12]。为了解决这些问题,刘珅和高琳[13]提出了改进型模型预测直接转矩控制方法,通过无差拍控制理论,计算出期望电压矢量,根据引入的虚拟电压矢量,确定期望电压所在扇区,进而求出实际电压矢量,结果表明,该方法能够有效抑制转矩和磁链脉动,提高模型鲁棒性。贾成禹等[14]提出了一种线性变参数模型预测控制算法,设计增量形式的预测模型及代价函数实现了转速的最优调节,从而克服因模型参数不匹配及外部扰动引起的稳态误差,实验结果表明,该方法具有良好的转速动态稳定性。

为了降低参数漂移对控制模型的影响,Fliess 和 Join[15]提出了无模型控制方法,该方法只利用系统的输入和输出,不考虑任何参数,将已知项和未知项干扰统称为系统总扰动,规避了系统参数对控制性能的影响,降低了模型对参数的依赖性,提高了控制性能。Zhou 等[16]提出了超局部无模型和无差拍预测控制方法,利用速度环的输入和输出变量建立了超局部控制模型,结果表明,该方法能有效提高模型鲁棒性,有较强的抗干扰能力。Zhou 等[17]建立了超局部无模型控制模型,针对系统总干扰项,利用参数辨识的方法在线进行估计,并进行前馈补偿,结果表明,该方法在一定程度上降低了电流纹波,提高了系统动态响应性能。

为了提高 PMSM 系统对参数变化、内部干扰和外部干扰的鲁棒性,获得更好的转速跟踪效果,本文提出超局部无模型转速预测控制(model-free speed prediction control, MFSPC)。利用速度环的输入和输出,不涉及任何系统参数,针对系统总扰动和数字控制系统存在的一拍延迟,建立扩张状态观测器(extended state observer, ESO),实时监测系统总扰动,并进行前馈补偿,利用 ESO 的预测转速弥补数字控制系统存在的一拍延迟。同时,采用频域分析方法,对 ESO 参数进行整定。实验结果表明,本文方法能够实现快速的动态响应,降低转速脉动,减小电流纹波,有较强的鲁棒性和抗干扰性。

1　PMSM 参数敏感性分析

根据磁场定向理论,PMSM 数学模型为

$$\begin{cases} u_d = R_0 i_d + \dfrac{\mathrm{d}\psi_d}{\mathrm{d}t} - p\omega\psi_q \\ u_q = R_0 i_q + \dfrac{\mathrm{d}\psi_q}{\mathrm{d}t} + p\omega\psi_d \\ J\dfrac{\mathrm{d}\omega}{\mathrm{d}t} = T_\mathrm{e} - T_\mathrm{L} - B\omega \end{cases} \tag{1}$$

$$\begin{cases} \psi_d = L_d i_d + \psi_\mathrm{f} \\ \psi_q = L_q i_q \end{cases} \tag{2}$$

式中:u_d 和 u_q 分别为 d-q 轴电压;i_d 和 i_q 分别为 d-q 轴电流;ψ_d 和 ψ_q 分别为 d-q 轴磁链;L_d 和 L_q 分别为 d-q 轴电感;R_0 为定子电阻;ψ_f 为永磁体磁链;p 为电机极对数;ω 为机械角速度;B 为摩擦系数;J 为转动惯量;T_e 为电磁转矩;T_L 为负载转。

将式(2)代入式(1)得到定子电压方程为

$$\begin{cases} u_d = R_0 i_d + L_d\dfrac{\mathrm{d}i_d}{\mathrm{d}t} - p\omega L_q i_q \\ u_q = R_0 i_q + L_q\dfrac{\mathrm{d}i_q}{\mathrm{d}t} + p\omega L_d i_d + p\omega\psi_\mathrm{f} \end{cases} \tag{3}$$

从式(3)可以看出,PMSM 数学模型实现了解耦。此时电磁转矩方程可写为

$$T_\mathrm{e} = 1.5pi_q[i_d(L_d - L_q) + \psi_\mathrm{f}] \tag{4}$$

对于表贴式 PMSM,$L_d = L_q = L_0$,此时,运动方程为

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = \frac{1.5p\psi_\mathrm{f}}{J}i_q - \frac{T_\mathrm{L}}{J} - \frac{B}{J}\omega \tag{5}$$

根据前向欧拉离散法,对式(5)进行离散化,得到 $k+1$ 时刻预测值:

$$\omega(k+1) = \left(1 - \frac{T_\mathrm{s}B}{J}\right)\omega(k) + T_\mathrm{s}\left(Ai_q - \frac{T_\mathrm{L}}{J}\right) \tag{6}$$

此时,根据式(6)得到速度环输出电流为

$$i_q = (AT_\mathrm{s})^{-1}\left[\omega(k+1) - \left(1 - \frac{T_\mathrm{s}B}{J}\right)\omega(k) + \frac{T_\mathrm{s}}{J}T_\mathrm{L}\right] \tag{7}$$

式中:$A = (1.5p\psi_{\mathrm{f}0})/J$;$T_\mathrm{s}$ 为采样时间。

随着电机运转,电机参数会发生变化,产生参数漂移,此时,电机参数实际值和电机铭牌上的标称值不相等。当电机参数为实际值时,此时理想的速度环输出电流为

$$i_{q1} = (A_1T_\mathrm{s})^{-1}\left[\omega_1(k+1) - \left(1 - \frac{T_\mathrm{s}B_1}{J_1}\right)\omega(k) + \frac{T_\mathrm{s}}{J_1}T_\mathrm{L}\right] \tag{8}$$

式中:$A_1 = (1.5p\psi_{\mathrm{f}1})/J_1$;$\psi_{\mathrm{f}1}$、$J_1$ 和 B_1 为电机参数的实际值;ω_1 为理想机械角速度。

但是在实际计算过程中,无法精确测量随时间变化的电机参数真实值,因此采用电机铭牌上的标称值。此时实际的速度环输出电流为

$$i_{q0}=(A_0T_s)^{-1}\left[\omega_0(k+1)-\left(1-\frac{T_sB_0}{J_0}\right)\omega(k)+\frac{T_s}{J_0}T_L\right] \tag{9}$$

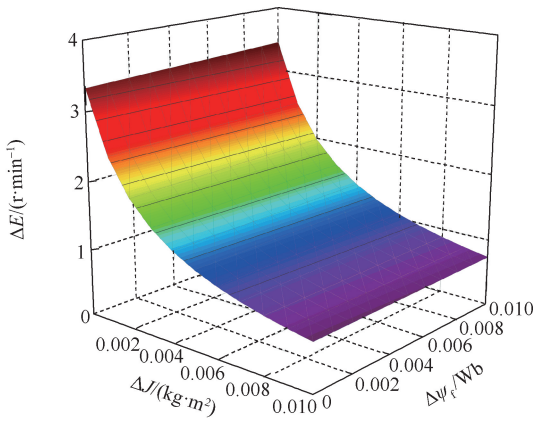
式中：$A_0=(1.5p\psi_{f0})/J_0$；ψ_{f0}、J_0 和 B_0 为电机铭牌上的标称值；ω_0 为实际机械角速度。

令速度环输出电流相等，即式(8)和式(9)相等，得到两者预测转速之间的关系为

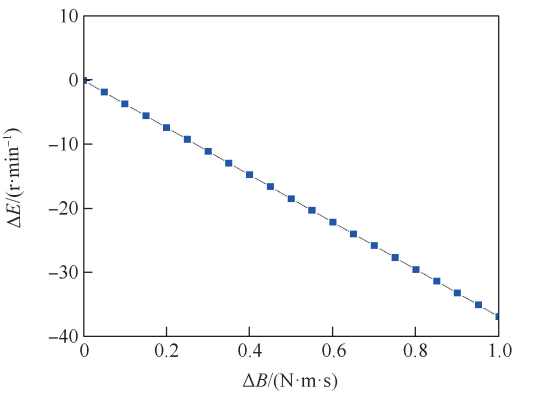
$$\omega_1(k+1)=\frac{A_1}{A_0}\omega_0(k+1)+\omega(k)\left(1-\frac{B_1T_s}{J_1}-\frac{A_1}{A_0}+\frac{A_1}{A_0}\cdot\frac{B_0T_s}{J_0}\right)+T_L\left(\frac{A_1}{A_0}\cdot\frac{T_s}{J_0}-\frac{T_s}{J_1}\right) \tag{10}$$

由式(10)可知，摩擦系数、转动惯量和永磁体磁链失配对预测转速均有影响。为了解决参数漂移导致的模型失配问题，提出超局部 MFSPC 方法。

图 1 表示在转速 1 000 r/min 和负载转矩 10 N·m 条件下，转动惯量 J、摩擦系数 B 和磁链 ψ_f 失配对速度环造成的误差 ΔE 影响。可知，3 种参数失配均影响着速度环的稳定性，转动惯量和摩擦系数失配对速度环的影响较大，而磁链参数失配引起的误差影响较小。



(a) 转动惯量和磁链失配的影响



(b) 摩擦系数失配的影响

图 1 1 000 r/min 和 10 N·m 条件下参数失配造成的误差影响

Fig. 1 Effect of errors caused by mismatch parameters at 1 000 r/min and 10 N·m

2 超局部无模型转速预测控制

2.1 超局部模型

针对一阶系统输入和输出，传统的超局部模型为[15]

$$\dot{y}=W+au \tag{11}$$

式中：y 为系统输出；u 为系统输入；W 为系统已知干扰和未知干扰总和；a 为模型结构的非物理比例因子。

$$u=\frac{\dot{y}^*-\hat{W}+\xi}{a} \tag{12}$$

式中：y^* 为系统的期望输出；$\hat{W}$ 为 W 的估计值；ξ 为所设计控制器的输出。

$$\dot{e}+\xi=0 \tag{13}$$

式中：$e=y^*-y$ 为跟踪误差。

当采用 P 控制器时，则式(12)可表达为

$$u=\frac{\dot{y}^*-\hat{W}+K_pe}{a} \tag{14}$$

$$\dot{e}+K_pe=0 \tag{15}$$

式中：K_p 为比例增益。

2.2 转速预测控制模型

考虑参数变化和未知扰动，式(5)变为

$$\frac{\mathrm{d}\omega}{\mathrm{d}t}=\frac{1.5p\psi_{f0}}{J_0}i_q-\frac{T_L}{J_0}-\frac{B_0}{J_0}\omega-\frac{1}{J_0}\cdot\left(\Delta J\frac{\mathrm{d}\omega}{\mathrm{d}t}-1.5p\Delta\psi_f+\Delta B\omega\right)+f_\omega \tag{16}$$

式中：$\Delta J=J_1-J_0$；$\Delta B=B_1-B_0$；$\Delta\psi_f=\psi_{f1}-\psi_{f0}$；$f_\omega$ 为未知扰动。

根据式(16)将 i_q 作为系统输入、ω 作为系统输出，建立超局部无模型控制结构：

$$\frac{\mathrm{d}\omega}{\mathrm{d}t}=\alpha i_q+F \tag{17}$$

式中：a 为控制器增益；F 为系统总扰动，包括系统参数漂移扰动项和未知扰动项。

采用前向欧拉离散法，对式(17)进行离散化，将连续时间模型转化为离散时间模型，得到

$$\omega(k+1)=\omega(k)+T_s(\alpha i_q(k)+\hat{F}(k)) \tag{18}$$

再求电流 i_q：

$$i_q(k)=\frac{\omega(k+1)-\omega(k)}{\alpha T_s}-\frac{\hat{F}(k)}{\alpha} \tag{19}$$

式中：$\hat{F}$ 为 F 的估计值。

为使转速更好地跟踪额定转速，令

$$\omega(k+1) = \omega^* \tag{20}$$

最终求出参考电流为

$$i_q(k) = \frac{\omega^* - \omega(k)}{\alpha T_s} - \frac{\hat{F}(k)}{\alpha} \tag{21}$$

另外,由于数字控制系统存在一拍延迟现象,如果在控制器中不加以考虑,将会降低控制系统性能[18]。为了解决数字控制系统一拍延迟问题,建立 ESO 模型结构,弥补一拍延迟。

3　建立 ESO 结构模型

基于式(17),选择 i_q 和 F 作为状态变量,建立 ESO,其状态空间方程为[19-20]

$$\begin{cases} e_\omega = z_1 - \omega \\ \dot{z}_1 = z_2 + \alpha i_q - \beta_1 e \\ \dot{z}_2 = -\beta_2 e \end{cases} \tag{22}$$

式中:e_ω 为速度环跟踪误差;β_1 和 β_2 为 ESO 增益系数;z_1 为 ω 的估计值,即 $z_1 = \hat{\omega}$;z_2 为 F 的估计值,即 $z_2 = \hat{F}$。

将式(22)进行 S 变换,得到

$$z_1(s) = \frac{(\beta_1 s + \beta_2)\omega + s\alpha i_q}{s^2 + \beta_1 s + \beta_2} \tag{23}$$

根据式(23)可以得到特征方程为

$$s^2 + \beta_1 s + \beta_2 = 0 \tag{24}$$

为使系统处于稳定状态,特征根为 $-\omega_0$,ω_0 为 ESO 的带宽,因此,ESO 增益系数 β_1 和 β_2 为

$$\begin{cases} \beta_1 = 2\omega_0 \\ \beta_2 = \omega_0^2 \end{cases} \tag{25}$$

采用前向欧拉离散法,对式(22)进行离散化,得到离散化状态空间方程为

$$\begin{cases} e(k) = \hat{\omega}(k) - \omega(k) \\ \hat{\omega}(k+1) = \hat{\omega}(k) + T_s[\hat{F}(k) + \alpha i_q(k) - \beta_1 e(k)] \\ \hat{F}(k+1) = \hat{F}(k) - T_s\beta_2 e(k) \end{cases} \tag{26}$$

为了验证离散化状态空间的稳定性,对式(26)进行 Z 变换,得到

$$\begin{cases} e(z) = \hat{\omega}(z) - \omega(z) \\ z\hat{\omega}(z) = \hat{\omega}(z) + T_s[\hat{F}(z) + \alpha i_q(z) - \beta_1 e(z)] \\ z\hat{F}(z) = \hat{F}(z) - T_s\beta_2 e(z) \end{cases} \tag{27}$$

考虑到采样时间足够小,则离散系统的传递函数为

$$G(z) = \frac{\hat{\omega}(z)}{\omega(z)} = \frac{\beta_{02}T_s + \beta_{01}z - \beta_{01}}{(z-1)^2 + \beta_{02}T_s + \beta_{01}z - \beta_{01}} \tag{28}$$

式中:$\beta_{01} = \beta_1 T_s$;$\beta_{02} = \beta_2 T_s$。

要想保证离散系统的稳定性,则离散系统的极点必须处于单位圆内,因此离散系统的特征方程为

$$(z-1)^2 + \beta_{02}T_s + \beta_{01}z - \beta_{01} = 0 \tag{29}$$

应用劳斯-赫尔维茨判据准则[21],令 $z = \frac{\lambda+1}{\lambda-1}$,得到 λ 域特征方程为

$$\beta_{02}T_s\lambda^2 + 2(\beta_{01} - \beta_{02}T_s)\lambda + (4 + \beta_{02}T_s - \beta_{01}) = 0 \tag{30}$$

根据劳斯·赫尔维茨判据准则,λ 域特征方程的各系数均大于 0,则离散系统稳定,进而求得 ESO 增益系数的取值范围,即

$$\begin{cases} \beta_{02}T_s < \beta_{01} < 4 + \beta_{02}T_s \\ \beta_{02} > 0 \end{cases} \tag{31}$$

在 PMSM 控制系统中,为了获得良好的控制效果,速度外环的带宽小于电流内环[22]。图 2 为系统控制结构。

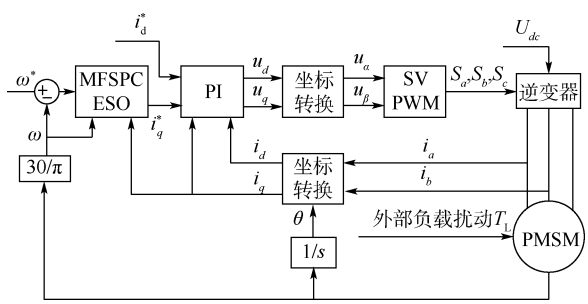


图 2　基于 MFSPC + ESO 方法的 PMSM 驱动系统控制

Fig. 2　PMSM drive system control based on MFSPC + ESO method

4　延迟补偿

为了弥补数字控制系统中的一拍延迟问题,将式(21)中的采样转速 $\omega(k)$ 和 $\hat{F}(k)$ 替代为式(26)中的和 $\hat{\omega}(k+1)$ 和 $\hat{F}(k+1)$。

$$i_q^*(k) = \frac{\omega^* - \hat{\omega}(k+1)}{\alpha T_s} - \frac{\hat{F}(k+1)}{\alpha} \tag{32}$$

通过 ESO 的转速预测值,求得参考电流 $i_q^*(k)$,弥补了传统 MFSPC 方法的一拍延迟。同时,针对系统总扰动,可以在线进行估计,并进行前馈补偿,提高控制系统的调节能力。MFSPC + ESO 控制结构如图 3 所示。

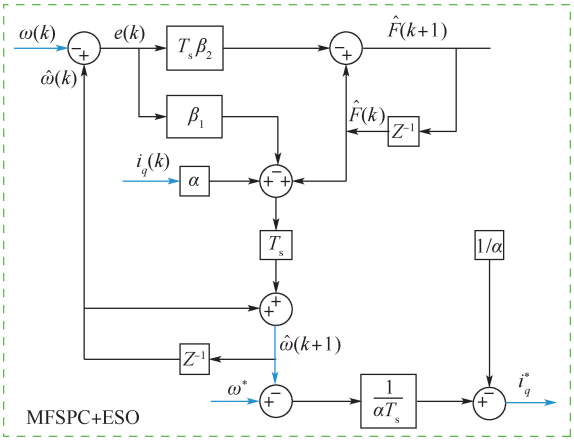


图 3 基于 ESO 的 MFSPC 转速预测控制

Fig. 3 Speed predictive control of MFSPC based on ESO

5 实验结果分析

PMSM 驱动控制系统的实验配置如图 4 所示，实验平台如图 5 所示。该实验平台基于 MATLAB/Simulink 进行伺服控制系统算法模型设计，并能够自动生成代码，通过开发软件下载到目标机上运行，便于 PMSM 控制方法的研究和验证。PMSM 模型的参数如表 1 所示。

为体现所搭建模型的控制性能，将 MFSPC + ESO 方法与传统 PI 方法、传统 MFSPC 方法、基于终端滑模控制（terminal sliding mode control，TSMC）的 MFSPC 方法（MFSPC + TSMC）[23] 结果相比较，并进行分析。MFSPC + ESO 方法控制参数 $\alpha = 4\ 000$，$\omega_0 = 100$ rad/s；PI 控制器增益参数 $K_p = 0.14$，$K = 14$；MFSPC + TSMC 方法控制参数 $\mu_1 = 1$，$\mu_2 = 5$，$\sigma_1 = 0.9$，$\sigma_2 = 0.5$。

5.1 速度环的反转性能测试

图 6 为在额定转速 1 000 r/min 下电机反转的响应波形。

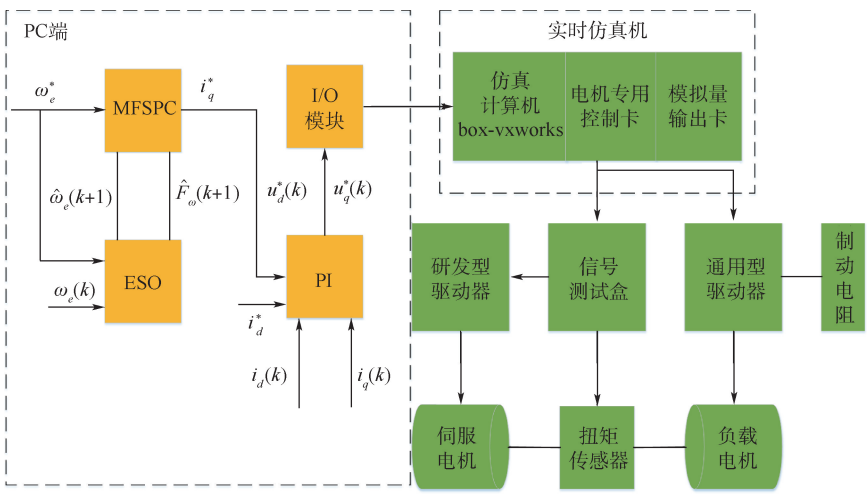


图 4 PMSM 实验平台配置

Fig. 4 PMSM experimental platform configuration

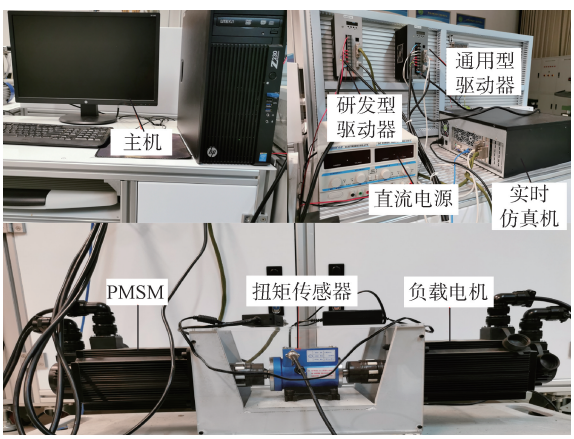


图 5 PMSM 实验平台

Fig. 5 PMSM experimental platform

由图 6(a) MFSPC + ESO 方法的电机转速响应可以看出，在电机启动阶段，几乎未产生超调，平稳启动，能够快速地达到给定转速，具有较快的动态响应，在正常运行阶段，没有产生明显的转速脉动，运行平稳。同时，图 6(b)、(c)在电机启动阶段，转速抖振现象较为严重，图 6(d)传统 PI 方法的转速产生了较大超调，不利于电机的正常启动。MFSPC + ESO 方法和传统 PI 方法的 d-q 轴电流运行较为平稳，虽然在启动阶段产生了过冲，

表 1 PMSM 模型参数

Table 1 Parameters of PMSM model

参数	数值
极对数 p	4
定子电阻 R_0/Ω	1.84
电感 L_0/mH	6.65
磁链 Ψ_{f0}/Wb	0.32
摩擦系数 $B_0/(\mathrm{N\cdot m\cdot s})$	0.008
转动惯量 $J_0/(\mathrm{kg\cdot m^2})$	0.002 7

但是能够快速回归正常，其中 MFSPC + ESO 方法产生的电流过冲较小，表现出良好的动态特性。而传统 MFSPC 方法的 q 轴电流出现了失真，脉动较为明显，MFSPC + TSMC 方法的 q 轴电流产生较大的电流纹波。当电机反转时，MFSPC + ESO 方法能够快速跟踪转速，有较快的动态响应特性，产生稳定的 d-q 轴电流，另外 3 种控制方法速度反转时，波动明显，d-q 轴电流失真，不能平稳运行。

表 2 为速度反转特性。可知，MFSPC + ESO 方法相比另外 3 种控制方法，转速超调量较小，在整个电机运行中，转速的波动方差较小，当电机反转时，回归正常转速的时间最小，表现出较好的稳定性和快速响应能力。

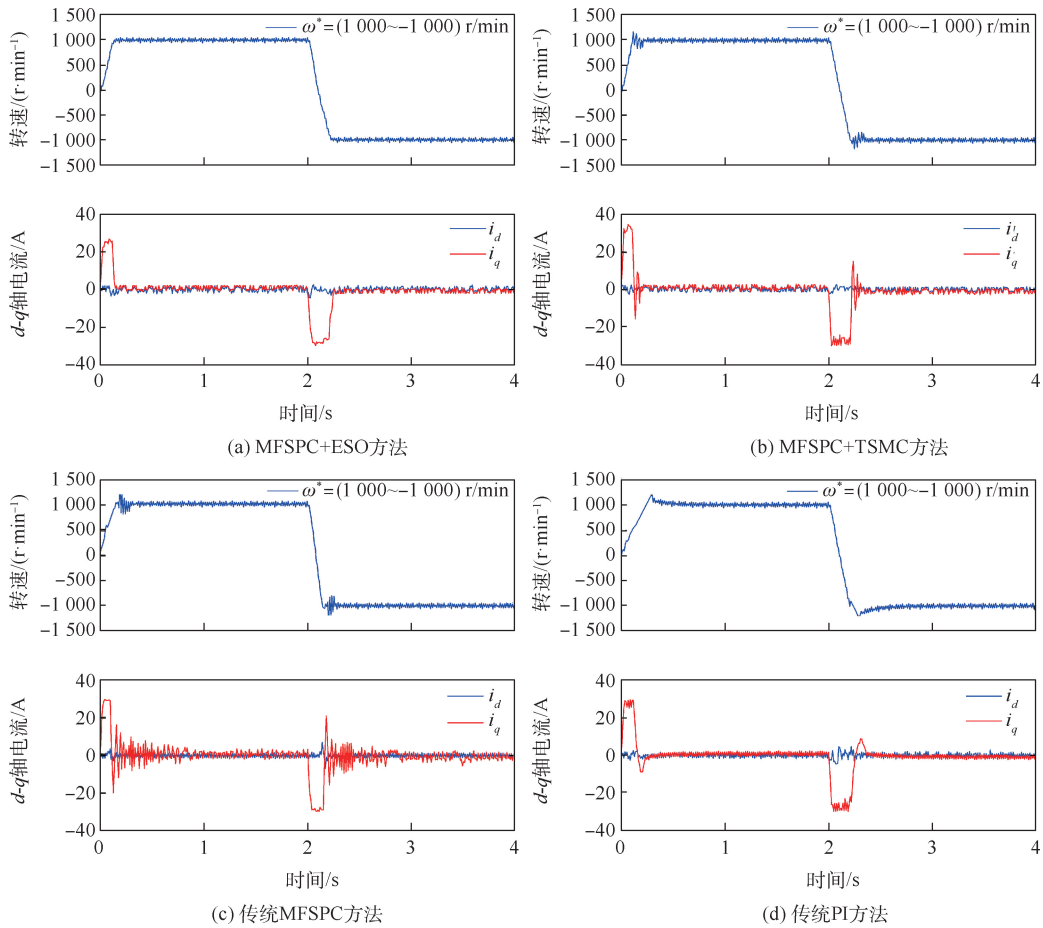


图 6　四种控制方法的反转性能对比

Fig. 6　Comparison of reversal performance of four control methods

表 2　速度反转特性

Table 2　Speed reversal characteristics

方法	转速超调/%	波动方差	回归时间/s
MFSPC + ESO	5.18	10.49	2.23
MFSPC + TSMC	17.5	12.51	2.36
MFSPC	36.05	13.01	2.29
PI	20.36	15.15	2.41

5.2　速度环的负载扰动性能测试

在宽速范围，加入 10 N · m 的外部负载扰动，电机转速的响应波形如图 7 所示。分别设置额定转速的 10%、80%、150%。由图 7 可知，在宽速范围内，MFSPC + ESO 方法能够平稳运行，在受到负载扰动时，转速下降较小，并且能够快速回归正常转速，表现出良好的跟踪和抗干扰性能。而 MFSPC + TSMC 方法，在低速为 100 r/min 时，转速失真，无法稳定精确跟踪给定转速，但是图 7(b)、(c)在受到负载扰动时，转速所受影响较小。传统 PI 方法在低速 100 r/min 时，抖振较为明显，波动较大，当加入负载扰动时，转速下降较大，抗干扰性能较低。

表 3 为在负载扰动下 4 种控制方法的速度超调和失调特性。可知，传统 PI 方法在电机启动时产生超调较大，当受到负载扰动时，失调也较明显；传统 MFSPC 方法和 MFSPC + TSMC 方法，在启动过程中产生了较大抖动，当受到负载扰动时，表现出一定的抗干扰能力；MFSPC + ESO 方法表

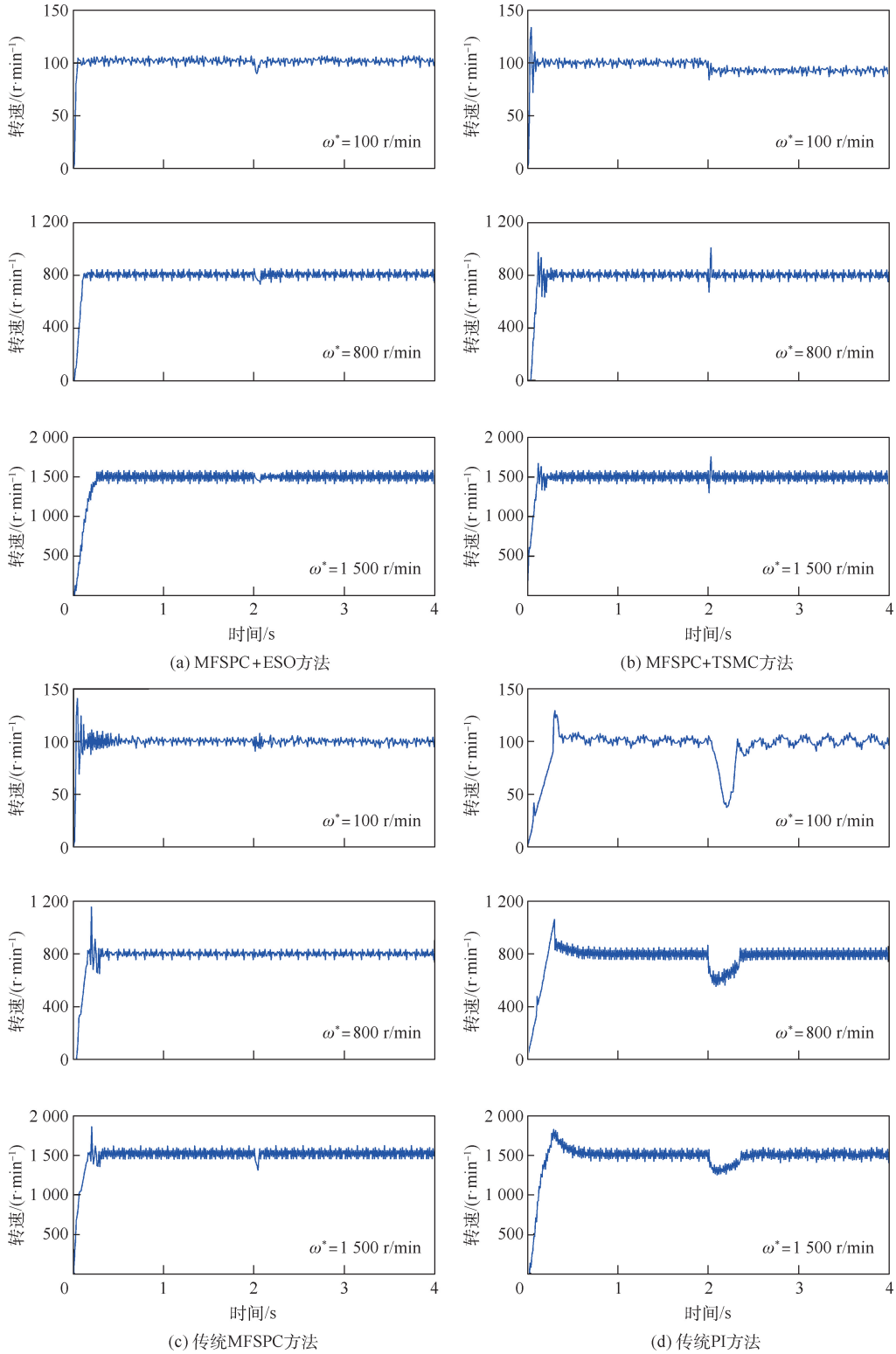


图7 四种控制方法的负载扰动性能对比

Fig. 7 Comparison of load disturbance performance of four control methods

现出较好的抗干扰性能。

图8为4种控制方法在额定转速1 000 r/min下加入额定负载15 N·m的转速响应曲线。可知,在电机启动阶段,相比于另外3种控制方法,MFSPC+ESO方法启动平稳,几乎未产生超调。当t=2 s时,加入额定负载,MFSPC+TSMC方法和传统MFSPC方法震荡较为明显,脉动比较大;传统PI方法受负载影响较大,失调较大;而MFSPC+ESO方法,转速下降较小,并且能够较快恢复正常,有较好的抗干扰性和较快的动态响应。

表 3　速度响应特性

Table 3　Speed response characteristics

方法	超调/%			失调/%		
	给定转速 100 r/min	给定转速 800 r/min	给定转速 1 500 r/min	给定转速 100 r/min	给定转速 800 r/min	给定转速 1 500 r/min
MFSPC + ESO	7.11	6.75	5.87	42.01	8.50	5.60
MFSPC + TSMC	34.01	26.38	17.40	16.11	15.88	13.21
MFSPC	41.15	45.07	24.04	9.09	5.19	12.83
PI	29.39	32.54	21.62	62.70	31.24	16.17

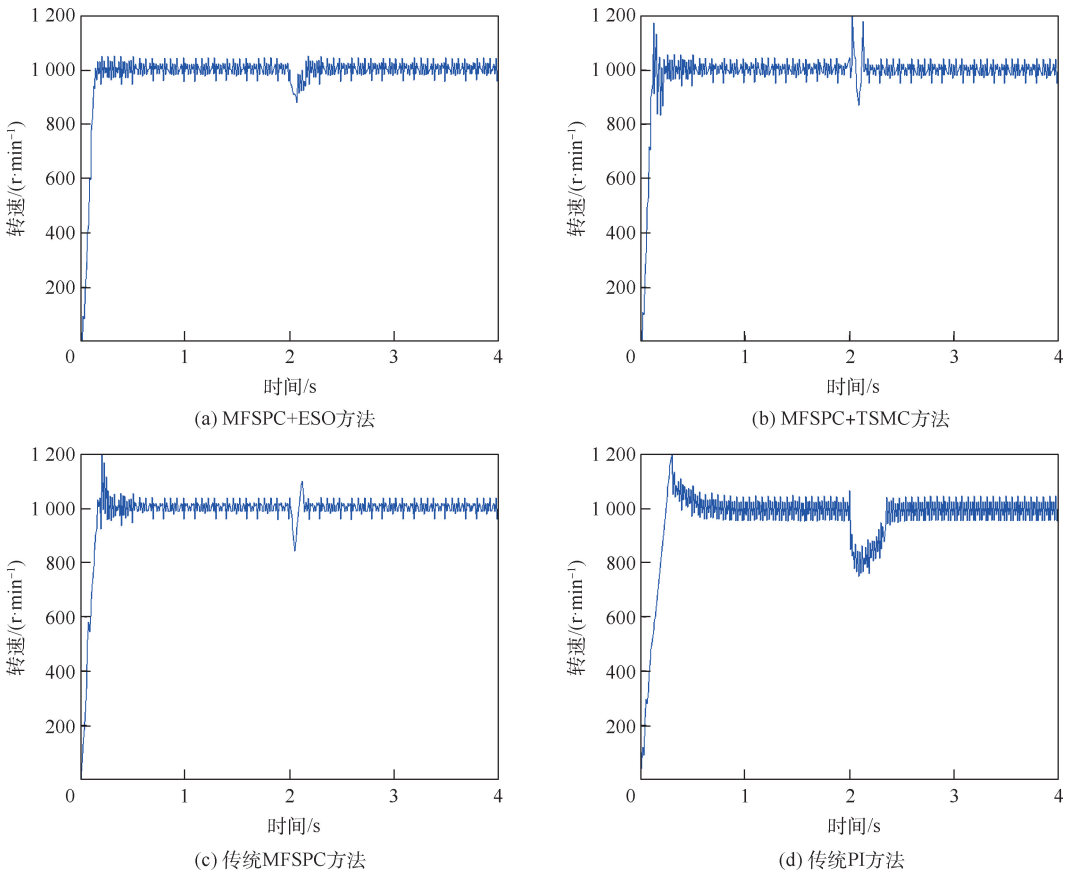


图 8　额定负载下的转速响应对比

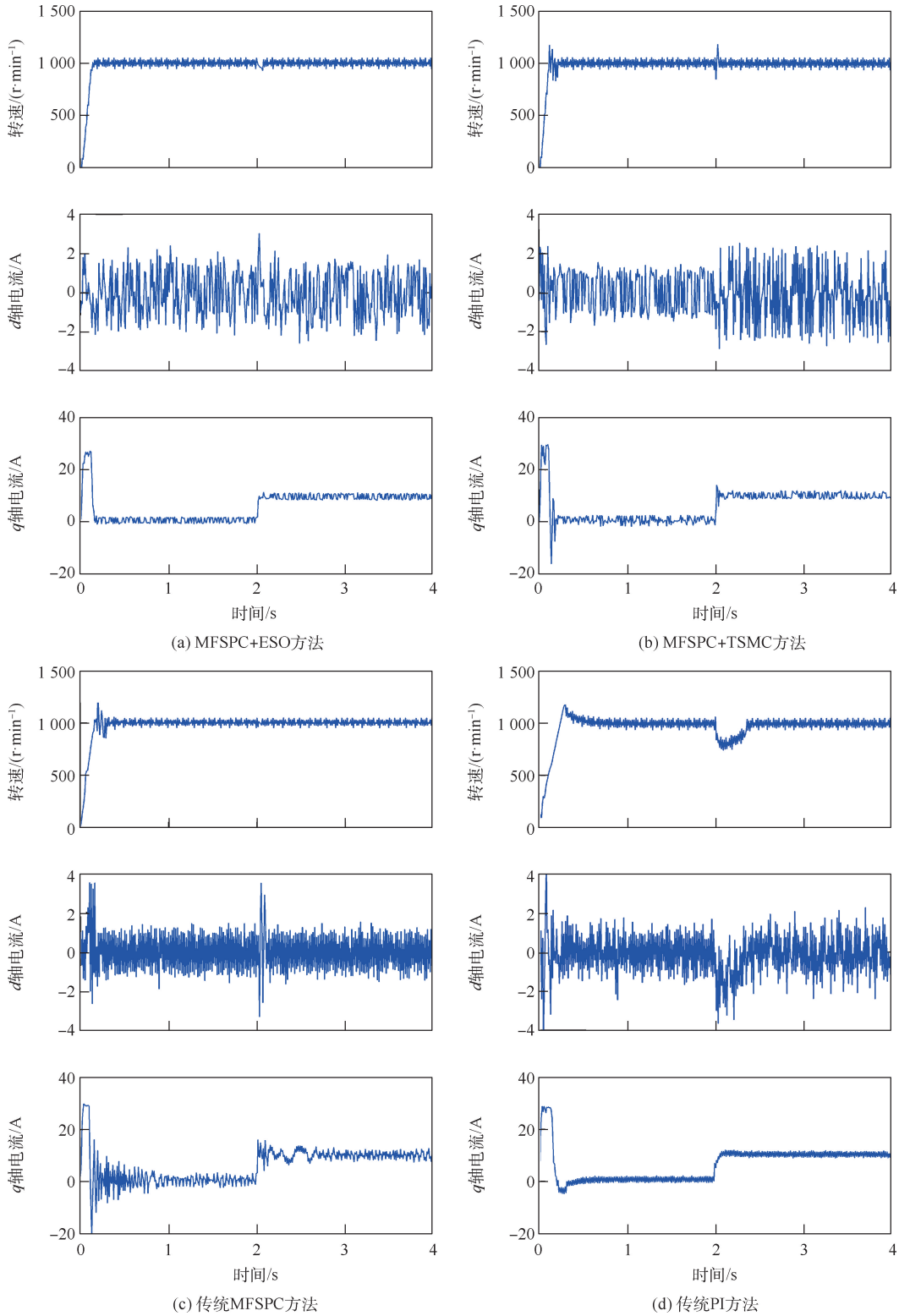
Fig. 8　Comparison of speed response under rated load

5.3　速度环的参数变化性能测试

当 PMSM 控制模型受到参数变化影响时，系统的控制性能下降。为了验证搭建模型的鲁棒性和抗干扰性能，设置 PMSM 参数磁链 $\psi_f = \psi_{f0} \rightarrow 0.8\psi_{f0}$，摩擦系数 $B = B_0 \rightarrow 2B_0$，转动惯量 $J = J_0 \rightarrow 2J_0$ 并且加入负载扰动 $T_L = 10$ N · m。

图 9 为在变参数和外部负载扰动条件下4 种控制方法的响应波形。MFSPC + ESO 方法始终表现出良好的鲁棒性和抗干扰性能。在启动阶段，传统 PI 方法出现较大超调，MFSPC + TSMC 方法和 MFSPC 方法出现较大脉动。当受到参数变化和外部负载扰动时，传统 PI 方法出现了较大失调，转速下降明显，而传统 MFSPC 方法出现了较大的转速脉动，运行不平稳，MFSPC + TSMC 方法无法准确跟踪额定转速，运转速度低于额定转速。*d-q* 轴电流在受到参数变化和负载扰动时，MFSPC + ESO 方法能够迅速跟踪，没有产生较大的电流纹波，动态特性较好，另外 3 种控制方法的电流响应波动较大，脉动较明显。

表 4 为 4 种控制方法的实现复杂度比较。MFSPC + ESO 方法与传统 PI 方法相比，不需要电机参数，各方面性能提高较大。与传统 MFSPC 方法和 MFSPC + TSMC 方法相比，MFSPC + ESO 方法仅需调节 2 个控制参数，工作量较低，表现出较强的鲁棒性和抗干扰性，转速和电流脉动较小，跟踪精度较高，表现出较优的性能。



(a) MFSPC+ESO方法　(b) MFSPC+TSMC方法

(c) 传统MFSPC方法　(d) 传统PI方法

图 9　四种控制方法的参数变化性能对比

Fig. 9　Comparison of parameter change of four control methods

表 4　四种控制方法实现复杂度比较

Table 4　Comparison of implementation complexity of four control methods

方法	电机参数	控制参数
PI	R_0, L_0, B, J, ψ_{f0}等	K_p, K_i
MFSPC		K_p, α, n_F
MFSPC + TSMC		α, σ_1, σ_2, μ_1, μ_2
MFSPC + ESO		α, ω_0

6　结　论

1）针对因电机参数漂移和外部扰动导致的电机控制性能下降，本文提出了超局部 MFSPC 方法，该方法不涉及任何电机参数，能有效降低模型对参数变化的敏感性。

2) 对于系统总扰动和数字系统中的一拍延迟,建立了 ESO,估测系统总扰动,进行前馈补偿,并利用无差拍控制原理,解决了一拍延迟问题,以提高控制精度

3) 将结合 ESO 的 MFSPC 方法与其他 3 种方法进行对比实验,结果表明,MFSPC + ESO 方法工作量较小,能够平稳启动,当受到扰动时,有较快的动态响应速度,能够准确跟踪额定转速,有较强的鲁棒性和抗干扰能力。

参考文献 (References)

[1] 陆骏,杨建国. 永磁同步电机滑模直接转速观测器[J]. 电机与控制学报,2018,22(1):86-92.
LU J,YANG J G. Direct sliding mode speed observer of permanent magnetic synchronous motor[J]. Electric Machines and Control,2018,22(1):86-92(in Chinese).

[2] 黄宴委,汤邵建,黄文超,等. 基于期望电压矢量的永磁同步电机快速速度预测控制[J]. 电机与控制学报,2020,24(4):87-95.
HUANG Y W,TANG S J,HUANG W C,et al. Fast speed predictive control of permanent magnet synchronous motor based on expected voltage vector[J]. Electric Machines and Control,2020,24(4):87-95(in Chinese).

[3] 周湛清,夏长亮,陈炜,等. 具有参数鲁棒性的永磁同步电机改进型预测转矩控制[J]. 电工技术学报,2018,33(5):965-972.
ZHOU Z Q,XIA C L,CHEN W,et al. Modified predictive torque control for PMSM drives with parameter robustness[J]. Transactions of China Electrotechnical Society,2018,33(5):965-972(in Chinese).

[4] XU L L,CHEN G M,LI Q Y. Ultra-local model-free predictive current control based on nonlinear disturbance compensation for permanent magnet synchronous motor[J]. IEEE Access,2020,8:127690-127699.

[5] 刘宁,夏长亮,周湛清,等. 基于比例增益补偿的永磁同步电机转速平滑控制[J]. 电工技术学报,2018,33(17):4007-4015.
LIU N,XIA C L,ZHOU Z Q,et al. Smooth speed control for permanent magnet synchronous motor using proportional gain compensation[J]. Transactions of China Electrotechnical Society,2018,33(17):4007-4015(in Chinese).

[6] JUNG J W,LEU V Q,DO T D,et al. Adaptive PID speed control design for permanent magnet synchronous motor drives [J]. IEEE Transactions on Power Electronics,2014,30(2):900-908.

[7] 马玉梅,谢洋,白环,等. 基于 DTC 的 PMSM 中 SMC 与纯 PI 的协同控制器研究[J]. 电工技术,2019,496(10):19-21.
MA Y M,XIE Y,BAI H,et al. Research on cooperative controller with SMC and only PI in PMSM based on DTC[J]. Electric Engineering,2019,496(10):19-21(in Chinese).

[8] 刘晓楠. 矩阵变换器-永磁同步电机系统直接转矩控制转矩波动抑制策略[J]. 中国电机工程学报,2020,40(1):300-308.
LIU X N. Torque ripple reduction in matrix converter-fed permanent magnet synchronous motor driven by direct torque control[J]. Proceedings of the CSEE,2020,40(1):300-308(in Chinese).

[9] VAFAIE M H,DEHKORDI B M,MOALLEM P,et al. A new predictive direct torque control method for improving both steady-state and transient-state operations of the PMSM[J]. IEEE Transactions on Power Electronics,2016,31(5):3738-3753.

[10] XIA C,ZHAO J,YAN Y,et al. A novel direct torque control of matrix converter-fed PMSM drives using duty cycle control for torque ripple reduction[J]. IEEE Transactions on Industrial Electronics,2014,61(6):2700-2713.

[11] 张虎,张永昌,刘家利,等. 基于单次电流采样的永磁同步电机无模型预测电流控制[J]. 电工技术学报,2017,32(2):180-187.
ZHANG H,ZHANG Y C,LIU J L,et al. Model-free predictive current control of permanent magnet synchronous motor based on single current sampling[J]. Transactions of China Electrotechnical Society,2017,32(2):180-187(in Chinese).

[12] 刘旭东,李珂,张奇,等. 基于非线性扰动观测器的永磁同步电机单环预测控制[J]. 中国电机工程学报,2018,38(7):2153-2162.
LIU X D,LI K,ZHANG Q,et al. Single-loop predictive control of PMSM based on nonlinear disturbance observers[J]. Proceedings of the CSEE,2018,38(7):2153-2162(in Chinese).

[13] 刘珅,高琳. 永磁同步电机的改进模型预测直接转矩控制[J]. 电机与控制学报,2020,24(1):10-17.
LIU K,GAO L. Improved model of predictive direct torque control for permanent magnet synchronous motor[J]. Electric Machines and Control,2020,24(1):10-17(in Chinese).

[14] 贾成禹,王旭东,周凯. 基于线性变参数模型预测控制的内置式永磁同步电机转速控制器设计[J]. 电工技术学报,2020,35(22):50-61.
JIA C Y,WANG X D,ZHOU K. Design of built-in permanent magnet synchronous motor speed controller based on linear variable parameter model predictive control[J]. Transactions of China Electrotechnical Society,2020,35(22):50-61(in Chinese).

[15] FLIESS M, JOIN C. Model-free control[J]. International Journal of Control,2013,86(12):2228-2252.

[16] ZHOU Y,LI H,ZHANG H. Model-free deadbeat predictive current control of a surface-mounted permanent magnet synchronous motor drive system[J]. Journal of Power Electronics,2018,18(1):103-115.

[17] ZHOU Y,LI H,YAO H. Model-free control of surface mounted PMSM drive system[C]// IEEE International Conference on Industrial Technology. Piscataway:IEEE Press,2016:175-180.

[18] ZHANG X,HOU B,MEI Y. Deadbeat predictive current control of permanent-magnet synchronous motors with stator current and disturbance observer[J]. IEEE Transactions on Power Electronics,2017,32(5):3818-3834.

[19] ZHANG Y, JIN J, HUANG L. Model-free predictive current control of PMSM drives based on extended state observer using

ultra-local model[J]. IEEE Transactions on Industrial Electronics, 2021,68(2):993-1003.

[20] 杨淑英,王玉柱,储昭晗,等. 基于增益连续扩张状态观测器的永磁同步电机电流解耦控制[J]. 中国电机工程学报, 2020,40(6):1985-1997.
YANG S Y, WANG Y Z, CHU Z H, et al. Current decoupling control of PMSM based on an extended state observer with continuous gains[J]. Proceedings of the CSEE, 2020, 40(6): 1985-1997(in Chinese).

[21] XU L L, CHEN G M, LI G Y, et al. Model predictive control based on parametric disturbance compensation[J]. Mathematical Problems in Engineering, 2020, 2020: 9543928.

[22] 毛海杰,李炜,蒋栋年,等. 基于线性扩张状态观测器的永磁同步电机状态估计与性能分析[J]. 电工技术学报, 2019,34(10):2155-2165.
MAO H J, LI W, JIANG D N, et al. State estimation and performance analysis based on linear extended state observer for permanent magnet synchronous motor[J]. Transactions of China Electrotechnical Society, 2019, 34(10): 2155-2165(in Chinese).

[23] WANG F X, HE L. FPGA based predictive speed control for PMSM system using integral sliding mode disturbance observer [J]. IEEE Transactions on Industrial Electronics, 2021, 68(2):972-981.

Ultra-local model-free speed predictive control based on ESO for PMSM

XU Lingliang, CHEN Guiming*, LI Qiaoyang

(College of Combat Support, Rocket Force University of Engineering, Xi'an 710025, China)

Abstract: An ultra-local model-free speed predictive control (MFSPC) method based on an extended state observer (ESO) is proposed to address the control performance decrease of permanent magnet synchronous motor (PMSM) due to parameter changes and external disturbances, and to solve the one-step delay problem in digital systems. This method reduces the model mismatch caused by motor parameter changes, using only the input and output of speed loops without motor parameters. The ESO is established to overcome the problems of the traditional MFSPC method which adjusts many parameters with large workload and output pulsation, obvious chattering, and low anti-interference and robustness. The ESO monitors the total disturbance of the system in real time, and performs feedforward compensation. It also addresses the beat delay problem in the digital control system, using the generated speed prediction, and sets the control parameters to improve control performance based on frequency domain analysis. Experimental results show that the proposed method has strong anti-interference and robustness, can track the rated speed steadily, and has faster dynamic response.

Keywords: permanent magnet synchronous motor; ultra-local; model-free speed prediction control; model mismatch; extended state observer; robustness

Received: 2021-02-22; **Accepted**: 2021-07-03; **Published online**: 2021-07-13 17:02

URL: kns.cnki.net/kcms/detail/11.2625.V.20210713.1419.002.html

Foundation item: Military Scientific Research Projects (20191C050239)

* **Corresponding author**. E-mail: 792757066@qq.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0087

星载原子钟在轨剩余寿命预测方法研究

凤建广[1,2], 郑紫霞[3], 龙东腾[3], 周波[3], 陆明泉[1,*], 郑恒[3]

(1. 清华大学 电子工程系, 北京 100084; 2. 中国卫星导航系统管理办公室, 北京 100032;
3. 中国航天标准化研究所, 北京 100071)

摘 要: 原子钟是导航卫星最关键载荷之一, 开展原子钟健康状态评估与寿命预测相关工作,对于及时掌握原子钟性能和运行状况,确保卫星导航系统提供高可用性连续性完好性的服务,具有重要意义。结合星载原子钟在轨失效特性分析,提出了一种综合考虑原子钟随机失效与耗损失效的剩余寿命预测方法。该方法融合利用原子钟寿命周期内产生的多源多维数据,通过概率统计和机器学习的途径实现原子钟的剩余寿命预测。经案例验证,所提方法合理有效,具有一定的工程应用价值。

关 键 词: 原子钟; 随机失效; 耗损失效; 机器学习; 剩余寿命预测
中图分类号: V19
文献标志码: A **文章编号**: 1001-5965(2022)11-2215-07

星载原子钟是导航卫星最关键的载荷之一,其决定导航系统导航定位、测速及授时精度,是一个国家能否具备独立发展卫星导航系统能力的核心标志之一[1-4]。开展星载原子钟在轨剩余寿命预测,对于及时掌握原子钟性能和运行状况、及时制定备份钟切换等有效在轨管理措施,以及确保卫星导航系统提供高可用性连续性完好性的服务,具有重要意义。

目前,针对航天器及单机产品寿命预测常用方法主要包括三大类:基于概率统计的方法、失效物理方法及数据驱动方法。其中,基于概率统计的方法是一种传统的寿命预测方法,需要大量样本以获得失效经验概率模型,在航天领域的应用具有一定的局限性;失效物理方法适应了航天领域样本量少的情况,主要依靠地面试验建立部件的失效机理模型,一般应用于航天器部件的设计阶段;数据驱动方法是一种较为普遍的现代故障预测方法,主要依靠关键数据和数据处理方法,完全不用考虑对象的模型特征,避免了前 2 种故障预测方法在模型建立和知识获取方面的困难,在航天器全寿命周期,特别是总装、测试、试验和在轨故障预测中具有广阔的应用前景。

针对星载原子钟的健康评估与寿命预测,目前常见的方法有概率统计方法、故障树分析方法、安全查表方法等[5]。其中,概率统计方法主要是假设星载原子钟的健康状态或寿命服从某一种分布,通过收集大量的样本数据,进而推测出当前星载原子钟的健康状态或寿命水平,该方法较为成熟,在工程中应用最广,但极其依赖样本数据,对于航天小子样产品适用性不强,另外对于耗损类的失效并不适用;故障树分析方法属于失效物理方法中的一种,应用较为广泛,主要通过计算底事件的失效概率,推导出顶事件发生故障的概率,该方法对经验数据依赖较强,且不能对顶事件的故障概率实现动态变化预测,预测精确度不高;安全查表方法类似于常用的阈值判别法,主要是结合在线监测数据和阈值,给出定性的判别结果,该方法直观、易操作,但不能做定量的判断。

收稿日期: 2021-02-22; **录用日期**: 2021-06-06; **网络出版时间**: 2021-07-13 09:37
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210712.1525.007.html
* **通信作者**. E-mail: lumq@tsinghua.edu.cn

引用格式: *凤建广, 郑紫霞, 龙东腾, 等. 星载原子钟在轨剩余寿命预测方法研究[J]. 北京航空航天大学学报, 2022, 48(11): 2215-2221. FENG J G, ZHENG Z X, LONG D T, et al. Method for predicting on-orbit residual life of satellite atomic clock[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2215-2221(in Chinese).*

本文结合星载原子钟在轨失效特性分析,通过概率统计与数据驱动相结合的方式,融合利用星载原子钟质量问题数据、可靠性/寿命试验数据及在轨遥测数据等多源数据,研究提出了一种星载原子钟在轨剩余寿命预测方法。该方法综合考虑了星载原子钟电路部分的随机失效及物理部分的耗损失效,针对不同失效类型用不同方法开展预测。其中,针对随机失效的概率统计方法,利用贝叶斯定理融合了研制阶段、地面试验阶段、在轨运行阶段及同类产品的相似数据,扩大了样本量,提高了预测准确度;而针对耗损失效的数据驱动方法,可以更好地捕捉退化数据特征,提高预测精度。

1 星载原子钟在轨失效特性分析

星载原子钟一般由物理部分和电路部分2部分组成。星载原子钟在运行过程中,不可避免受到太空复杂环境、元器件老化和供电设备故障等因素影响而发生异常[6-7]。通过以往的工程经验,星载原子钟的故障主要是物理器件故障[8],是随着运行时间延长由于耗损而导致的部件失效和性能下降,如氢源损耗、真空度下降、铷泡故障[9-10]等。而电路部分主要受空间环境影响较大,主要故障模式为无功率输出或电压波动[11],故障表现为随机。

基于上述分析,星载原子钟故障可分为随机型和耗损型2类。针对随机型故障,基于星载原子钟质量问题数据、地面可靠性试验失效数据及同类产品数据,利用概率统计方法评估其可靠寿命;针对耗损型故障,提取反映星载原子钟损耗特性的关键参数(如电离泡光强、二次谐波幅度),利用机器学习的方法建立退化寿命评估模型,并根据监测到的特征参数值,预测其在轨退化寿命;最终,综合退化寿命与可靠寿命,评估得出星载原子钟在轨剩余寿命。基于多源数据融合的星载原子钟在轨剩余寿命预测思路如图1所示。

铷原子钟、氢原子钟、铯原子钟虽然工作机理不同,但基本结构都包括物理和电路2部分,对应的失效类型均可以归纳为耗损失效和随机失效2类,故本文方法对上述3类星载原子钟具有普适性。

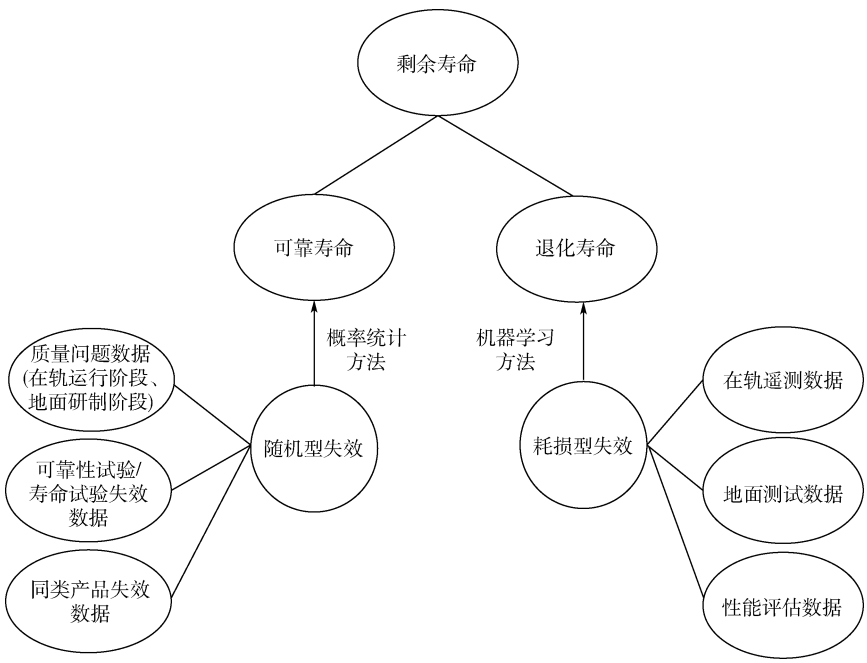


图1 星载原子钟在轨剩余寿命预测思路

Fig. 1 On-orbit residual life prediction of satellite atomic clock

2 星载原子钟在轨剩余寿命预测

2.1 星载原子钟在轨剩余寿命预测数据采集

采集星载原子钟寿命周期内多源多维数据,作为在轨剩余寿命预测的数据输入,包括质量问题数据(本产品及同类产品地面研制/在轨运行阶段质量问题)、可靠性/寿命试验数据、在轨遥测数据及性能评估数据。以氢原子钟为例,数据采集类型及示例如表1所示。

2.2 基于多源数据融合的星载原子钟在轨剩余寿命预测

2.2.1 考虑随机失效的星载原子钟可靠寿命预测

利用概率统计的方法,开展基于随机失效的星载原子钟可靠寿命预测。为扩大样本量,本文

表 1　星载原子钟在轨剩余寿命预测数据类型及示例

Table 1　Data types and examples for on-orbit residual life prediction of satellite atomic clock

数据类型	数据说明	数据示例	用途
质量问题数据	本产品在轨运行阶段质量问题发生次数、发生时间	XXXX-XX-XX,1 次	原子钟在轨可靠寿命预测
质量问题数据	本产品地面研制阶段质量问题发生次数、发生时间	XXXX-XX-XX,1 次	原子钟在轨可靠寿命预测
质量问题数据	同类产品在轨运行阶段/地面研制阶段质量问题发生次数、发生时间(同类产品指与评估产品技术状态一致的产品)	XXXX-XX-XX,1 次	原子钟在轨可靠寿命预测
地面试验数据	同批次同类产品地面可靠性/寿命试验样本个数、试验时间、失效个数	2 个样本,试验 4 000 h,失效 0	原子钟在轨可靠寿命预测
在轨遥测数据	电离泡光强	2.5 V	原子钟在轨退化寿命预测
在轨遥测数据	二次谐波幅度	1.28 V	原子钟在轨退化寿命预测
在轨遥测数据	二极管电压	1.5 V	原子钟在轨退化寿命预测
在轨遥测数据	晶振压控电压	1.36 V	原子钟在轨退化寿命预测
在轨遥测数据	微波腔中部温度	56℃	原子钟在轨退化寿命预测
在轨遥测数据	高压源电流	0 V	原子钟在轨退化寿命预测
性能评估数据	频率稳定度	10×10^{-5}	原子钟在轨退化寿命预测

融合利用了产品质量问题(本产品及同类产品地面研制/在轨运行阶段)数据及地面试验数据,具体过程如下。

1) 质量问题数据与地面可靠性/寿命试验数据融合。利用贝叶斯方法实现质量问题数据与地面可靠性/寿命试验数据的融合,具体流程如下[12]:

① 可靠性/寿命试验数据融合。

a) 数据。试验样本数、累计试验时间 T、失效数 r、失效时间。

b) 验前信息转换。将验前信息转换为先验信息,选定 $r_0=1$,等效试验时间按照式(1)计算:

$$t_0=\frac{2.022\ 32}{\lambda_0}\times10^9 \tag{1}$$

式中:$(t_0,1)$为验前信息;λ_0 为经验失效率。

式(1)将验前信息按置信度 0.6 进行压缩。

c) 相容性分析。对验前信息(t_0,r_0)与试验信息(T,r)进行相容性分析,可根据如下双边区间估计计算公式进行统计显著性检验:$\frac{\chi^2_{\alpha/2}(2r+1)}{2T}$,$\frac{\chi^2_{1-\alpha/2}(2r+1)}{2T}$。其中,$\alpha$ 为显著性水平,取值为 0.01 ~ 0.1。

若验前信息的比值 r_0/t_0 在统计显著性检验计算的区间内,则接受相容性假设,可以用验后信息进行计算;若不在该区间内,则拒绝接受相容性假设,舍弃验前信息,仅应用现场试验信息进行计算。

d) 验后信息计算。综合试验信息(T,r)和验前信息(t_0,r_0),验后信息按照式(2)和式(3)计算:

$$T_1=T_0+T \tag{2}$$

$$r_1=r_0+r \tag{3}$$

② 质量问题数据融合。

a) 数据。质量问题数 r'、累计在轨运行时间 T'、失效时间。

b) 融合过程。以上述融合可靠性/寿命试验数据后的(T_1,r_1)为先验信息,重复上述相容性分析,若相容性假设合理,则验后信息按照式(4)和式(5)计算:

$$T_2=T_1+T \tag{4}$$

$$r_2=r_1+r \tag{5}$$

2) 可靠寿命预测。利用威布尔分布开展氢原子钟在轨可靠寿命预测。假设试验产品数 n,失效数 r,相应的失效时间:$t_1,t_2,\cdots,t_r$,无失效产品的试验或在轨运行时间 $t_{r+1},t_{r+2},\cdots,t_n$。基于威布尔的可靠度及寿命计算式如下:

$$R_M(t)=\exp\left[-\left(\frac{t}{\eta}\right)^m\right] \tag{6}$$

式中:尺度参数 η 计算为

$$\eta=\left(\frac{t^*}{r}\right)^{\frac{1}{m}} \tag{7}$$

$$t^*=\sum_{i=1}^{n}t_i^m \tag{8}$$

形状参数 m 计算如下:

1) 对于失效数 $r\geqslant3$ 的情况,按照求解式(9)计算:

$$\frac{1}{m}+\frac{\sum_{i=1}^{r}\ln t_i}{r}-\frac{t_i^m\ln t_i}{\sum_{i=1}^{r}\ln t_i^m}=0 \tag{9}$$

2) 对于失效数 $r<3$ 的情况,根据工程经验确定形状参数 m,通常取 2 ~ 3。

在给定可靠度后,利用式(7)反算得到当前可靠度下的可靠寿命。

2.2.2　考虑耗损失效的星载原子钟退化寿命预测

利用机器学习方法,开展基于耗损失效的星载原子钟退化寿命预测,具体过程如图 2 所示。

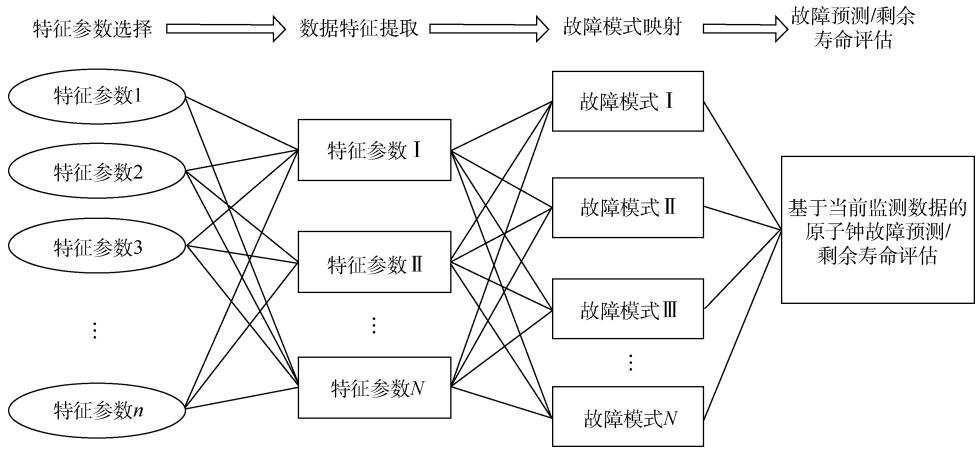


图 2 星载原子钟退化寿命预测思路

Fig. 2 Degradation life prediction of satellite atomic clock

1）特征参数选择。依据工程经验，选取表征星载原子钟性能的特征参数，以氢原子钟为例，主要的特征参数有二次谐波幅度、二极管电压、电离泡光强、高压源电流、晶振压控电压及微波腔部温度。

2）数据特征提取。利用深度置信网络（deep belief network，DBN）对数据特征进行提取，数据特征提炼了原始数据中的信息，比原始数据更具有价值。DBN 模型是由多层受限制玻尔兹曼机（restricted Boltzmann machine，RBM）与顶层 Softmax 分类器[13]构成。DBN 的结构示意如图 3[14]所示。

DBN 是一种经典的机器学习[15]方法，由多个 RBM 堆叠而成，RBM 由可见层和隐含层构成，可见层神经元和隐含层神经元之间相互连接，同一层的神经元之间相互独立。可见层和隐含层神经元状态由{0，1}代表激活和未激活状态。RBM 通过可见层和隐含层之间各神经元的连接权值来实现对隐含层的构建。数据由最底层 RBM 的可见层输入，依照上述 RBM 训练的过程自下而上完成 DBN 中所有 RBM 的无监督学习，最顶层的 RBM 输出即输入数据的提取特征。

3）故障模式映射。通过神经网络模型训练感知数据特征变化与氢原子钟故障模式之间的关联关系，将数据特征信息映射到故障模式空间。

4）剩余寿命预测。基于实时传感器获得的实时数据特征，通过数据特征与故障模式之间映射关系，计算获得当前状态下氢原子钟发生各类故障模式的置信水平，从而预测氢原子钟在轨退化寿命。

2.2.3 综合寿命预测

运用蒙特卡罗仿真综合退化寿命与可靠寿命，评估得出含置信度的产品在轨剩余寿命。流程如图 4 所示。

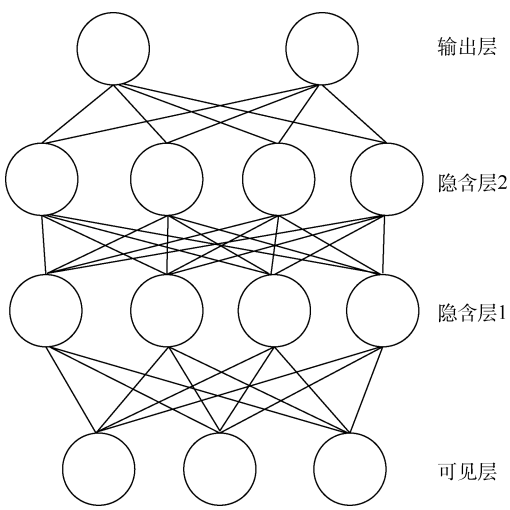


图 3 DBN 结构[14]

Fig. 3 DBN structure[14]

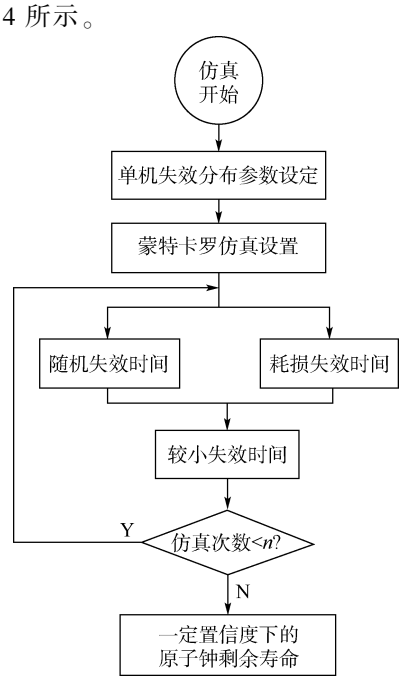


图 4 星载原子钟剩余寿命评估流程

Fig. 4 Process of assessing residual life of satellite atomic clock

利用仿真的方法综合考虑随机失效和耗损失效，对星载原子钟的剩余寿命进行评估。基本思想是：针对随机失效模型获得的可靠度函数，利用蒙特卡罗仿真产生的随机数，作为单机的可靠度，通过反算获得单机随机失效时间，与计算获得的耗损失效时间进行比对，取其最小值作为星载原子钟的在轨剩余寿命时间。

3　示例验证

本节以北斗系统某星某氢原子钟的一次实际故障为例，基于故障前 2 个月时间内的在轨遥测数据，利用 DBN 开展故障预测与退化寿命评估，从而与实际故障情况相比对，以验证本文方法的有效性。由于本产品之前未发生过随机失效，在此不对随机失效进行验证。

1）数据采集及预处理

采集故障前一段时间内氢原子钟特征参数遥测值，包括二次谐波幅度、二极管电压、电离泡光强、高压源电流、晶振压控电压、微波腔中部温度。各参数变化趋势如图 5 所示。

在故障发生前，遥测参数发生 2 次明显跳变，2 次跳变情况如表 2 所示。可知，在 2 次数据异常表现中，二次谐波幅度、二极管电压、电离泡光强、高压源电流、晶振压控电压、微波腔中部温度都发生了跳变。

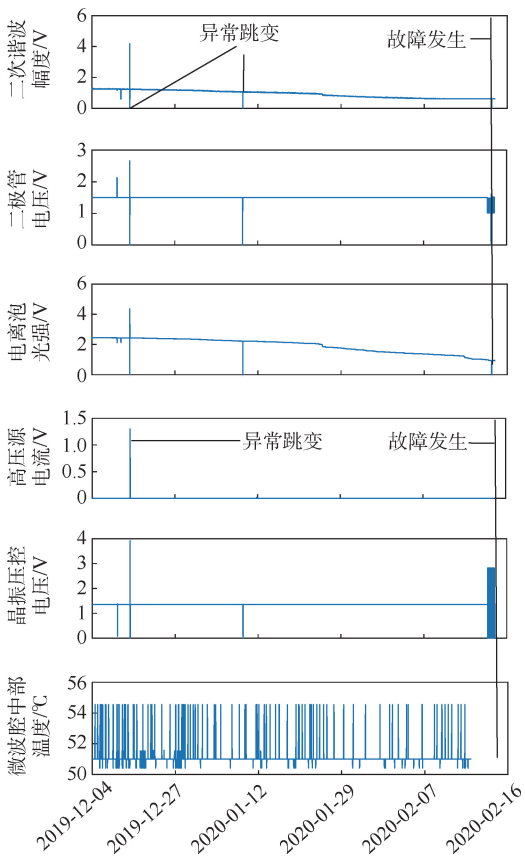


图 5　特征参数趋势分析

Fig. 5　Trend analysis of characteristic parameters

通过分析得知，2 次异常值出现的时间点上多个遥测参数同时发生跳变，不可认为是单个遥测参数的遥测误差导致而将数据剔除，因此在数据预处理过程中予以保留。

2）数据特征提取

利用 DBN 进行遥测参数的数据特征提取。通过分析 6 个遥测参数与时间序列的相关性（见图 5）发现，二次谐波幅度和电离泡光强 2 个参数随时间退化趋势较为明显，其他 4 类参数遥测随时间的变化趋势基本平稳，未呈现出退化规律性。

故本文选取二次谐波幅度和电离泡光强 2 个参数开展数据特征提取。

① 搭建数据特征提取网络结构。基于 DBN 的氢原子钟遥测数据特征提取网络结构如图 6 所示。

② 基于故障数据进行网络权重优化。最顶层 RBM 输出经过 Softmax 分类器，依据故障期间

表 2　特征参数变化情况

Table 2　Changes of characteristic parameters

参数	第 1 次跳变值	第 2 次跳变值	正常值
二次谐波幅度/V	0	0	1.28
二极管电压/V	0	0	1.5
电离泡光强/V	0	0	2.5
高压源电流/V	1.3		0
晶振压控电压/V	0/3.92	0	1.36
微波腔中部温度/℃			50.998

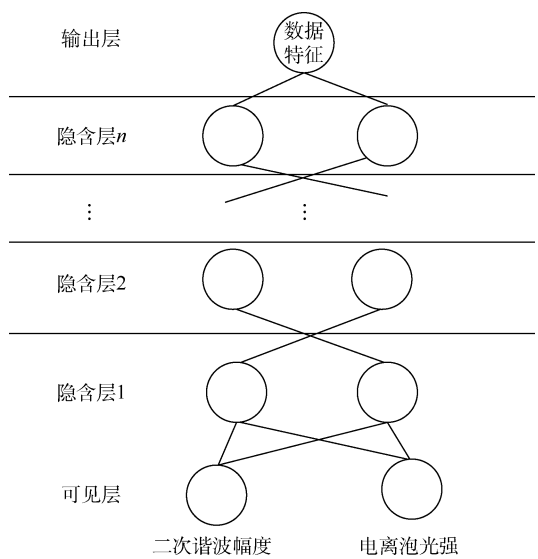


图 6　氢原子钟遥测数据特征提取网络结构

Fig. 6　Network structure for feature extraction of hydrogen clock telemetry data

氢原子钟的遥测数据,形成数据特征分类标签,该标签与原始数据标签对比并计算形成建模误差,通过误差反向传播算法对 DBN 中的 RBM 连接权值进行有监督的优化。

③ 故障预测/寿命评估模型。通过利用 DBN 获得基于二次谐波幅度和电离泡光强的数据特征后,在该数据特征的基础上进行决策分析和故障预测,利用神经网络算法建立数据特征与本次某卫星发生的故障模式之间的关联模型。

④ 故障分析。针对提取的数据特征,开展故障前一段时间内数据特征变化趋势分析,如图 7 所示。设定数据特征“0”表示健康,“1”表示故障,由数据特征变化趋势可以看出,通过二次谐波幅度和电离泡光强提取的数据特征 2019-02-07 前一直稳定在“0”左右,表示此时氢原子钟状态健康,到 2019-02-08 前后开始产生波动,并于 2020-02-09 稳定处于 0.53,并一直延续至故障实际发生之前,当故障发生后,数据特征在 0.94 ~ 0.99 范围内波动。

通过分析可看出,基于二次谐波幅度和电离泡光强提取的数据特征能够正确反映氢原子钟健康状态及退化趋势,提前发现隐含在遥测参数内的故障征兆。

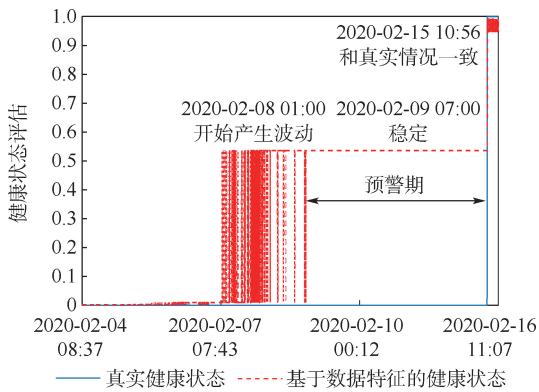


图 7 数据特征分析结果

Fig. 7 Data feature analysis results

4 结 论

本文提出了一种星载原子钟在轨剩余寿命预测方法,经上述方法分析、案例验证,得出如下结论:

1) 区别考虑了星载原子钟在轨随机失效及耗损失效的特性,针对不同失效特性,利用不同数据,采取不同途径实现相关的寿命预测,较传统概率统计等方法更具有针对性。

2) 融合利用了星载原子钟寿命周期内产生的多源多维数据,不仅考虑了当前在轨表现(在轨数据),同时结合了历史故障信息(质量问题数据、试验数据),预测结果更为可信。

3) 传统的阈值判别方法只能对参数进行最直观的观测,而本文中参数直到故障发生前都没有超出阈值,而是在故障发生时刻突然跳变,因此阈值判别并不能捕捉参数的退化特点,大多时候都不能实现提前的故障预测。本文利用 DBN 开展了星载原子钟特征参数的数据特征提取,并对故障前 2 个月的数据特征进行了变化趋势分析,预测结果与实际情况一致,一定程度上验证了本文方法可以实现故障的提前预测,为工程风险防控提供有效的决策支持。

参考文献(References)

[1] 李锡瑞,漆溢. 提高氢原子钟监控系统性能的设计[J]. 时间频率学报,2016,39(4):266-273.
LI X R, QI Y. A design for improving performance of hydrogen maser monitor system[J]. Journal of Time and Frequency, 2016,39(4):266-273(in Chinese).

[2] 翟造成. 我国氢钟研制现状与用于卫星导航的可行性[J]. 宇航计测技术,2003,23(5):1-9.
ZHAI Z C. The research status of domestic hydrogen masers and their application prospects for satellite positioning system[J]. Journal of Astronautic Metrology and Measurement, 2003, 23(5):1-9(in Chinese).

[3] 王文明. 国内外氢钟最新发展及我国氢钟未来发展趋势[J]. 导航定位与授时,2015,2(6):48-54.
WANG W M. The current development of atomic hydrogen clock and the future technology development trend in China [J]. Navigation Positioning and Timing,2015,2(6):48-54(in Chinese).

[4] 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. Performance analysis and evaluation of hydrogen atomic clock for Beidou navigation satellite[J]. Chinese Science:Physics, Mechanics & Astronomy, 2021,51(1):125-134(in Chinese).

[5] 孙海燕,谢彦民,荆涛,等. 基于模糊故障树分析法实现氢原子钟风险评估[J]. 现代导航,2010(5):64-68.
SUN H Y, XIE Y M, JING T, et al. Realization of hydrogen atomic clock risk assessment based on fuzzy fault tree analysis [J]. Modern Navigation,2010(5):64-68(in Chinese).

[6] 黄观文. GNSS 星载原子钟质量评价及精密钟差算法研究[D]. 西安:长安大学,2012.
HUANG G W. Research on algorithms of precise clock offset and quality evaluation of GNSS satellite clock[D]. Xi'an: Chang'an University,2012(in Chinese).

[7] 杨友超,李海伟. 空间电源系统寿命预测技术发展综述[J]. 宇航计测技术,2019,39(4):1-5.
YANG Y C, LI H W. Development survey of space power system life prediction[J]. Journal of Astronautic Metrology and

Measurement,2019,39(4):1-5(in Chinese).

[8] 邓小波,赵斌,王铮. 原子钟在线监测评估方法设计[J]. 无线电工程,2018,48(1):46-49.
DENG X B,ZHAO B,WANG Z. Design of online monitor and evaluation method for atomic clock[J]. Radio Engineering,2018,48(1):46-49(in Chinese).

[9] 邓爱民,陈循,张春华,等. 星载铷钟的故障模式及影响分析[J]. 国防科技大学学报,2005,27(2):102-106.
DENG A M,CHEN X,ZHANG C H,et al. The failure modes and effects analysis of the space-born e rubidium clock[J]. Journal of National University of Defense Technology,2005,27(2):102-106(in Chinese).

[10] 李锡瑞,蔡勇. 氢原子钟故障模式及影响分析[C]//第七届中国卫星导航学术年会,2016:1-5.
LI X R,CAI Y. The failure modes and effects analysis of hydrogen maser[C]//The 7th China Satellite Navigation Academic Conference,2016:1-5(in Chinese).

[11] 李荣凯,李锡瑞,蔡勇. 基于机器学习方法的氢原子钟故障诊断研究[J]. 天文研究与技术,2020,17(3):349-356.
LI R K,LI X R,CAI Y. Research on fault diagnosis of hydrogen maser based on machine learning[J]. Astronomical Research and Technology,2020,17(3):349-356(in Chinese).

[12] 中国航天标准化研究所. 宇航单机产品可靠性评估要求:QJA 307—2014[S]. 北京:中国航天标准化研究所,2014.
China Aerospace Standardization Institue. Reliability evaluation requirements of space equipments:QJA 307—2014[S]. Beijing:China Aerospace Standardization Institue,2014(in Chinese).

[13] 冉鹏,王灵,李昕. 改进 Softmax 分类器的深度卷积神经网络及其在人脸识别中的应用[J]. 上海大学学报(自然科学版),2018,24(3):75-90.
RAN P,WANG L,LI X. Improved Softmax classifier for deep convolution neural networks and its application in face recognition[J]. Journal of Shanghai University (Natural Science),2018,24(3):75-90(in Chinese).

[14] 姜宗辰,马毅,江涛,等. 基于深度置信网络(DBN)的赤潮高光谱遥感提取研究[J]. 海洋技术学报,2019,38(2):1-7.
JIANG Z C,MA Y,JIANG T,et al. Research on the extraction of red tide hyperspectral remote sensing based on the deep belief network (DBN)[J]. Journal of Ocean Technology,2019,38(2):1-7(in Chinese).

[15] 秦文健. 基于机器学习的医学影像分割关键问题研究及其在肿瘤诊疗中的应用[D]. 深圳:中国科学院深圳先进技术研究院,2019.
QIN W J. A study of medical image segmentation and its application in tumor diagnosis and therapy based on machine learning[D]. Shenzhen:Shenzhen Institutes of Advanced Technology,Chinese Academy of Sciences,2019(in Chinese).

Method for predicting on-orbit residual life of satellite atomic clock

FENG Jianguang[1,2], ZHENG Zixia[3], LONG Dongteng[3], ZHOU Bo[3], LU Mingquan[1,*], ZHENG Heng[3]

(1. Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;
2. China Satellite Navigation Office, Beijing 100032, China;
3. China Astronautics Standard Institute, Beijing 100071, China)

Abstract: An atomic clock is one of the most critical loads of a navigation satellite. It is thus of great significance to carry out the work related to the health assessment and residual life prediction of atomic clocks to understand the performance and the health status of the atomic clock in time and to ensure that the satellite navigation system provides services with high availability and continuity. Based on the analysis of on-orbit failure characteristics of the atomic clock, a method for residual life prediction of the satellite atomic clock is presented in this paper. The method considers the random failure and the loss failure of the atomic clock comprehensively based on the multi-dimensional data generated in the life cycle of the atomic clock, and predicts residual life by means of probability statistics and machine learning. Case studies show that the method proposed is reasonable, effective, and applicable to engineering practices.

Keywords: atomic clock; random failure; loss failure; machine learning; residual life prediction

Received: 2021-02-22; **Accepted**: 2021-06-06; **Published online**: 2021-07-13 09:37

URL: kns.cnki.net/kcms/detail/11.2625.V.20210712.1525.007.html

* **Corresponding author**. E-mail: lumq@tsinghua.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0103

高精度球面洛伦兹力磁轴承设计与分析

傅百恒[1], 王卫杰[2,*], 王元钦[2], 樊亚洪[3], 聂辰[4], 贾海鹏[4]

(1. 航天工程大学 研究生院, 北京 101416; 2. 航天工程大学 宇航科学与技术系, 北京 101416;
3. 北京控制工程研究所, 北京 100190; 4. 中国人民解放军 63961 部队, 北京 100020)

摘 要: 针对柱面洛伦兹力磁轴承(LFMB)偏角有限导致磁悬浮控制敏感陀螺(MSCSG)力矩输出持续时间短和气隙磁密均匀度低影响控制敏感精度的突出问题,提出了一种高精度球面 LFMB 设计与分析方法。所设计的 LFMB 转子球面导磁套和定子球面绕组均与双球面陀螺转子同球心,气隙呈球壳状,保证转子偏转时定子绕组两侧气隙宽度不变,相较于柱面 LFMB,转子可偏转角度由 ±0.6°扩大到 ±2°。利用等效磁路法推导了柱面与球面 LFMB 气隙磁密的数学解析模型,并基于 ANSYS 命令流构建了柱面与球面 LFMB 的有限元仿真模型。仿真结果表明:在转子可偏转范围内,沿偏转中心线,球面 LFMB 最大磁密较柱面下降了 34.1%;当转子不偏转时,球面 LFMB 绕组截面内的磁密均匀度较柱面提高了 11.6%;当转子偏转时,球面 LFMB 绕组截面内的磁密均匀度较柱面提高了 17.7%。所提方法为磁悬浮控制敏感陀螺控制与敏感性能的提升奠定了基础。

关 键 词: 磁悬浮控制敏感陀螺; 球面洛伦兹力磁轴承; 磁密均匀度; 等效磁路法; 有限元分析

中图分类号: V221+.3; TB553

文献标志码: A **文章编号**: 1001-5965(2022)11-2222-08

高分卫星对地观测时,卫星的姿态稳定度及超静性能是相机清晰成像的关键[1-6]。作为航天器姿态控制的一种磁悬浮解决方案,磁悬浮控制敏感陀螺(magnetically suspended control and sensing gyroscope,MSCSG)融合了高精度控制力矩输出和载体姿态敏感测量的功能,可输出大于 1 N·m、带宽大于 100 Hz 的偏转力矩[7-10]。洛伦兹力磁轴承(Lorentz force magnetic bearing,LFMB)作为 MSCSG 的功能部件,其性能好坏决定了 MSCSG 能否实现高精度敏感控制[11-15]。

LFMB 主要由转子磁钢环和定子绕组构成,磁钢环提供稳定的气隙磁场,绕组置于其中。绕组对置通电后产生安培力形成力偶以控制转子偏转,以陀螺实现高精度敏感控制。文献[16]通过仿真指出柱面 LFMB 的气隙磁密分布不均是产生干扰力矩、引起测量误差的主要因素,该干扰力矩会随着气隙磁密的不均匀度增大而增大,最终影响 MSCSG 的控制敏感精度。文献[17-18]通过 MSCSG 的控制敏感公式指出,由于气隙磁密已给定,当输入陀螺的扰动越小时,LFMB 所需的敏感电流越小,导致在信噪中难以分辨该电流信号,若适当减小气隙磁密,可增大敏感电流。文献[19-20]在柱面 LFMB 的内外永磁体靠近气隙一侧增加了铁镍软磁合金(1J50)导磁薄片,由于增加了磁阻,气隙磁密降低,但其均匀度提升并不明显,不能提高 LFMB 的控制敏感精度。

收稿日期: 2021-03-02; **录用日期**: 2021-05-05; **网络出版时间**: 2021-05-18 13:23
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210518.1013.003.html
基金项目: 国家自然科学基金 (52075545)
* **通信作者**. E-mail: wangwjie@126.com

引用格式: 傅百恒, 王卫杰, 王元钦, 等. 高精度球面洛伦兹力磁轴承设计与分析[J]. 北京航空航天大学学报, 2022, 48(11): 2222-2229. FU B H, WANG W J, WANG Y Q, et al. Design and analysis of high precision for spherical Lorentz force magnetic bearing [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2222-2229(in Chinese).

此外,研究人员还将 LFMB 应用于磁悬浮隔振平台。文献[21]利用 LFMB 高精度控制的优点,设计了一种主被动结合磁悬浮隔振平台,该平台针对 100 Hz 以下的振动,其六自由度的加速度峰值可削弱 90% 以上,但由于 LFMB 的控制行程有限,难以补偿 100 Hz 以上的振动。

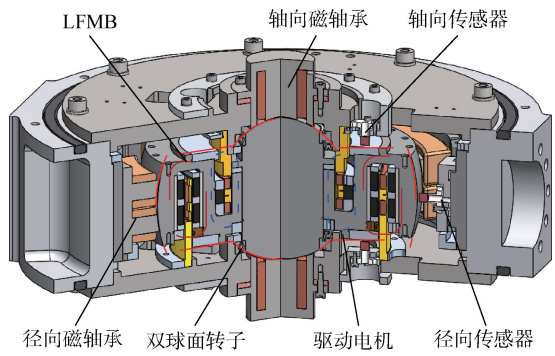
综上可知,提高 LFMB 气隙磁密的均匀度、并适当降低磁密是 MSCSG 实现高精度控制及敏感测量的关键方法,而扩大转子偏角又可提高 MSCSG 应对复杂工况时持续输出力矩的性能。因此,根据设计要求,本文提出了一种气隙磁密均匀度较高、转子偏角较大的球面 LFMB 设计方法。通过仿真结果可知,球面 LFMB 不仅能够有效提高气隙磁密的各向均匀度,增大了转子偏角,拓宽了球面 LFMB 高精度控制的区间,为 MSCSG 提高控制敏感精度、扩大适用工况奠定了基础。

1　MSCSG 高精度测控机理分析

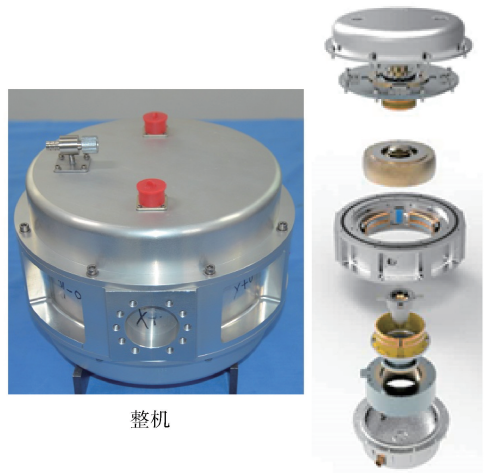
1.1　基于微框架的 MSCSG 测控机理

MSCSG 主要由无刷直流电机、轴向磁轴承、径向磁轴承、LFMB、位移传感器等部件组成,如图 1所示。轴向、径向磁轴承协同控制转子的三自由度平动悬浮,LFMB 控制转子的二自由度径向偏转,转子可实现五自由度全主动稳定悬浮,消除了定转子之间的摩擦,使得电机驱动转子实现可控的高速旋转,提高了陀螺的运行寿命及可靠性。此外,轴向及径向磁轴承的定转子磁极面均设计为球面,在转子偏转前后,其磁极面之间的气隙宽度不会发生变化,电磁力始终垂直于球形转子表面,以此消除转子平动对偏转控制的影响。

MSCSG 利用 LFMB 控制高速旋转的转子偏转以跟踪载体位姿变化实现敏感测量,及产生动量差形成控制力矩,利用 LFMB 实现一体化测控方案实现过程如下:当平台扰动传递至陀螺时,MSCSG 对姿态变化进行敏感测量,转子由于高速旋转保持定轴性,不会跟随定子偏转,利用位移传感器测量定转子的相对位移,解算位移可得转子相对定子的偏角,再往 LFMB 通入敏感电流以控制转子偏转至扰动前,此时检测定子对置绕组的敏感电流差,可得陀螺的位姿变化角速度;根据敏感测量所得角速度,再往 LFMB 通入控制电流,以对转子沿径向二自由度施加力偶控制转子瞬时偏转,则在微框架效应下,由转子偏转产生的动量差将转换为陀螺力矩,来抵消或补偿扰动,该过程即为 MSCSG 控制力矩的输出过程。



(a) 结构图



(b) 实物图

图 1　MSCSG 结构与实物图

Fig. 1　Structure diagram and material object of MSCSG

LFMB 通过控制转子的二自由度万向偏转使 MSCSG 实现控制敏感功能,因此,对 LFMB 控制精度的要求较高。

1.2　MSCSG 高精度测控的影响因素分析

根据 1.1 节中 MSCSG 的测控机理可知,衡量 LFMB 性能优劣主要有以下两方面:①力偶的控制精度;②对微小扰动的敏感测量精度。

1)影响敏感测量精度的因素

MSCSG 姿态敏感是通过测量转子相对于惯性空间的姿态角速率得到,该角速率正比于 LFMB 定子对置绕组的电流差,敏感式为

$$\begin{cases} \omega_{Bx} = \dfrac{BL_r(I_4 - I_2)}{J_z\Omega} \\ \omega_{By} = \dfrac{BL_r(I_3 - I_1)}{J_z\Omega} \end{cases} \tag{1}$$

式中:B 为气隙磁密强度;L_r 为 LFMB 定子骨架半径;Ω 为陀螺转速;J_z 为转子的极转动惯量;ω_{Bx}、ω_{By} 分别为姿态角速率沿 x、y 轴的分量;$I_4 - I_2$、$I_3 - I_1$ 分别为在 x、y 轴上 2 组对置绕组的电流差。其中,陀螺转速及极转动惯量由实际情况给定。

通过检测 LFMB 对置绕组电流差,解算得到陀螺的姿态角速率。然而当气隙磁密强度较大而

由外部输入陀螺的扰动较小时,LFMB 绕组内的敏感电流将相应减小,导致其信号在信噪中的分辨率过低,难以检测,使得 MSCSG 无法实现对微小扰动的敏感测量,且气隙磁密的不均匀分布也将进一步影响 LFMB 的偏转控制精度。

2) 影响力矩控制精度的因素

MSCSG 二自由度(绕 x 或 y 轴)的控制力矩为

$$T = 8nBL_r^2 \sin\phi \begin{bmatrix} I_2 \\ I_3 \end{bmatrix} \tag{2}$$

式中:n 为单个绕组匝数;ϕ 为单个绕组有效周向长度所对应的圆周角;I_2 和 I_3 分别为对置绕组的电流大小。其中,I_i 是控制量,由敏感所得位姿信息决定,n、L_r、ϕ、B 是已知量,由陀螺设计给出。

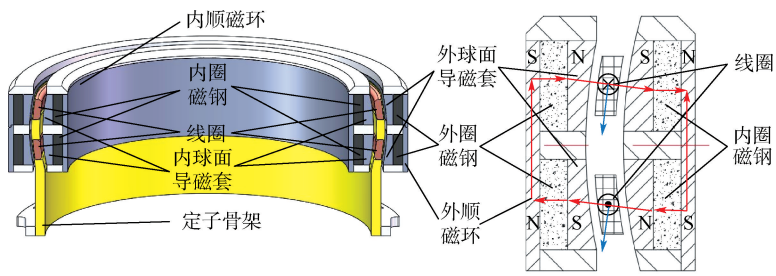
由于洛伦兹力的大小、方向、磁场强度、绕组与磁场夹角有关,若气隙磁密的均匀性较差,则在轴截面范围内,安培力分布并不均匀,LFMB 无法控制转子进行高精度偏转,使得 MSCSG 不能输出高精度力矩来补偿载体振动,且当转子偏转后,磁场与绕组的夹角小于 90°,加剧了洛伦兹力分布的不均匀性。此外,传统 LFMB 定转子均为柱面结构,其控制转子偏转的持续输出时间短,导致转子瞬时偏转产生的动量包络范围受限,MSCSG 难以持续输出力矩来补偿载体的高频振动。

2 高精度球面 LFMB 设计

为提高 MSCSG 的控制敏感精度,针对柱面 LFMB,将围绕以下两方面进行设计:①提高绕组所处气隙磁密的均匀度;②扩大转子的偏转角度。

研究发现,在磁钢靠近气隙一侧增加导磁材料,一方面可提高气隙磁密的均匀度,但会削弱磁密,另一方面导磁结构无需充磁,不影响磁路。气隙宽度保持不变,为扩大转子的偏角,将气隙设计为球面转子的同心球壳,本文提出了一种球面 LFMB 的设计方案:在转子柱面磁钢靠近气隙一侧,设计与转子球心重合的铁钴钒软磁合金(1J22)球面导磁结构,则气隙形状由圆柱环变为球壳状,定子绕组及骨架的形状相应地由柱面改为与转子球心重合的球面。

由于增加球面导磁结构不影响球面 LFMB 整体磁路的闭合,为保证转子偏转后绕组边缘仍处于均匀磁场中,磁钢的轴向高度也做调整。根据设计要求,球面 LFMB 的结构设计如图 2 所示,其主要参数设计和计算结果如表 1 所示,球面 LFMB 设计偏转角度由柱面 LFMB 的 ±0.6°扩大至 ±2°,磁钢的轴向高度相应增加 2 mm。



(a) 整体结构 (b) 轴截面及磁路

图 2 球面 LFMB 整体结构示意图

Fig. 2 Overall structure diagram of spherical LFMB

表 1 两种 LFMB 主要参数

Table 1 Main design parameters of two LFMBs

参数	柱面 LFMB	球面 LFMB
内永磁体内径/mm	119.4	119.4
内永磁体外径/mm	127.4	127.4
内球面导磁套外径/mm	138.2	136
外永磁体内径/mm		152
外永磁体外径/mm	146.2	160
外球面导磁套内径/mm		146.8
磁钢轴向高度/mm	11	13
上下永磁体间轴向距离/mm	4	4
绕组轴向高度/mm	8.5	8.5
设计偏转角度/(°)	±0.6	±2
极限偏转角度/(°)	0.8	2.5
单个绕组沿周向张角/(°)	84	84

结合实际加工、装配环节,由于气隙形状的改变,使得球面 LFMB 的装配方式较柱面有了很大变化。柱面 LFMB 在转子装配完成后,定子可竖直滑配进入气隙中。但对于球面 LFMB 的装配,由于球面结构,定子上沿直径小于内隔磁环直径、定子中心直径大于外导磁套下沿直径。因此,若将球面 LFMB 的转子部分一次性装配完成,则无法将定子竖直滑入气隙中;若扩大气隙宽度,将会大幅削弱气隙磁密,影响气隙磁路闭合,导致球面 LFMB 无法正常工作。

因此,考虑到球面 LFMB 的装配难度及可行性,本文还提出了一种球面 LFMB 装配方法,如图 3所示。首先,将定子绕组内置于骨架以增大

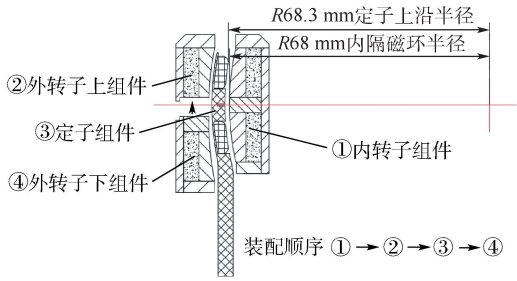


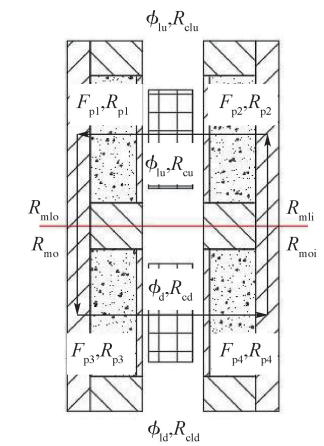
图 3　球面 LFMB 装配方案

Fig. 3　Assembly plan of spherical LFMB

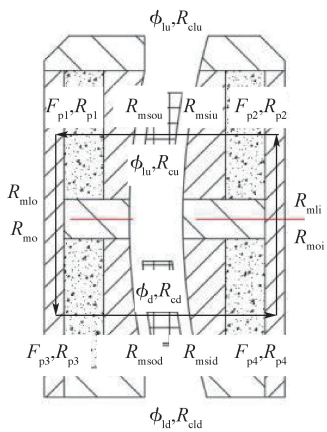
定子上沿直径,保证定子上沿能够竖直通过内隔磁环;其次,将球面 LFMB 的外转子分为上、下组件,其上组件和内转子先装配组成上半部分气隙,在定子竖直滑配进入气隙后,再将外转子下组件滑配至气隙中,从而保证了定转子各自的装配空间不发生冲突。

3　球面 LFMB 等效磁路分析

根据柱面及球面 LFMB 的轴截面结构,建立磁路图,如图 4 所示。F_{p1} ~ F_{p4}分别为 4 组磁钢环的磁动势,R_{p1} ~ R_{p4}分别为磁动势对应的永磁磁



(a) 柱面LFMB



(b) 球面LFMB

图 4　磁路图

Fig. 4　Magnetic circuit diagram

阻,R_{cu}和 R_{cd}分别为处于上、下磁钢环间穿过定子绕组的气隙磁阻,R_{clu}和 R_{cld}分别为上、下磁钢环间未穿过定子绕组的气隙磁阻,R_{mi}和 R_{mo}分别为内、外导磁环的磁阻,R_{msiu}和 R_{msou}分别为上球形内、外导磁套的磁阻,R_{msid}和 R_{msod}分别为下球形内、外导磁套的磁阻,R_{mli}和 R_{mlo}分别为内、外导磁环未经过导磁环的漏磁磁路气隙磁阻,ϕ_u和 ϕ_{lu}分别为上气隙出穿过和未穿过绕组的磁通,ϕ_d和 ϕ_{ld}分别为下气隙出穿过和未穿过绕组的磁通。

忽略定子绕组磁阻及磁钢漏磁及边缘效应可得等效磁路,如图 5 所示,柱面与球面磁路的区别在于是否有球面导磁结构,如虚线框标示。

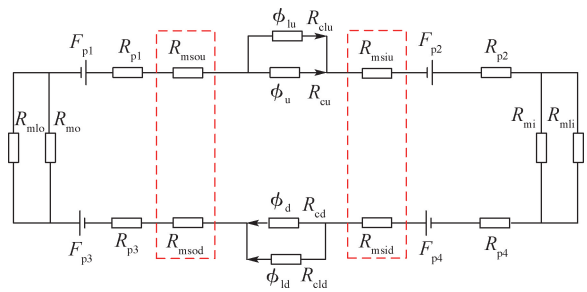


图 5　等效磁路图

Fig. 5　Equivalent magnetic circuit

根据图 5 可知,由于外导磁环磁阻 R_{mlo}与外磁钢磁阻 R_{mo}为并联磁阻,内导磁环磁阻 R_{mli}与内磁钢磁阻 R_{mli}为并联磁阻,而导磁环磁阻可近似为 0,并联后电阻也为 0,最终 LFMB 气隙内的主磁通 $\boldsymbol{\phi}_s$为

$$\boldsymbol{\phi}_s = \begin{bmatrix} \phi_{lu} \\ \phi_u \\ \phi_{ld} \\ \phi_d \end{bmatrix} = \begin{bmatrix} R_{cu}/(R_{clu}+R_{cu}) \\ R_{clu}/(R_{clu}+R_{cu}) \\ R_{cd}/(R_{clu}+R_{cu}) \\ R_{cld}/(R_{clu}+R_{cu}) \end{bmatrix} \phi_{ms} \tag{3}$$

式中:ϕ_{ms}为通过 LFMB 绕组的气隙主磁通。

$$\phi_{ms} = \frac{\sum_{i=1}^{4} F_{pi}}{\sum_{i=1}^{4} R_{pi} + R_{ms} + \dfrac{R_{clu}R_{cu}}{R_{clu}+R_{cu}} + \dfrac{R_{cld}R_{cd}}{R_{cld}+R_{cd}}} \tag{4}$$

$$F_{pi} = H_c l_{pi} \qquad i = 1,2,3,4 \tag{5}$$

$$R_{ms} = R_{msiu} + R_{msou} + R_{msid} + R_{msod} \tag{6}$$

式中:F_{p1} ~ F_{p4}为磁钢的磁动势;l_{pi}为磁钢磁化长度。由式(4)可知,2 种 LFMB 主磁通差异在于靠近气隙处的导磁结构磁阻 R_{ms}。

根据主磁通可得到对应位置的气隙磁密 $\boldsymbol{B}_z$为

$$\boldsymbol{B}_z = [B_{lu} \quad B_u \quad B_{lu} \quad B_d]^{\mathrm{T}} = [\mathrm{diag}(A_{lu},A_u,A_{ld},A_d)]^{-1} \tag{7}$$

式中：B_{lu} 和 B_u 分别为上气隙出垂直为穿过绕组的磁密和穿过绕组的磁密；B_{ld} 和 B_d 分别为下气隙出垂直为穿过绕组的磁密和穿过绕组的磁密；A_{lu}、A_u、A_{ld} 和 A_d 分别为气隙磁密强度 B 对应的 4 个等效截面积。

由于磁钢所处径向尺寸远大于其磁化宽度，则可假设 $F_{p1}=F_{p2}=F_{p3}=F_{p4}=F_p$，$l_{p1}=l_{p2}=l_{p3}=l_{p4}=l_p$，且由于 2 种 LFMB 均为上下对称的双圈结构，则 $R_{cu}=R_{cd}=R_c$，$R_{clu}=R_{cld}=R_{cl}$，$A_u=A_d=A$，$A_{lu}=A_{ld}=A_1$，$R_{msi}=R_{msiu}=R_{msid}$，$R_{mso}=R_{msou}=R_{msod}$，垂直穿过绕组的磁密 B_u 和 B_d 可写为

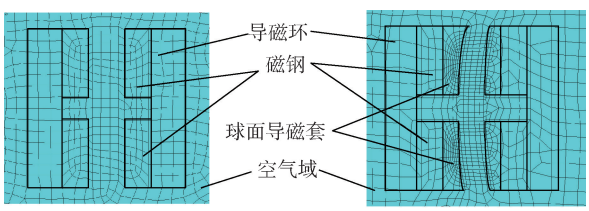
$$B_u=B_d=4H_cl_p\frac{R_{cl}}{A(4R_p+R_{msi}+R_{mso})+2AR_{cl}R_c} \tag{8}$$

由式(8)可知，气隙磁密 $\boldsymbol{B}_z$ 是关于磁钢磁化长度的增函数、关于磁阻的减函数。其中，磁钢长度相同的情况下，磁阻的大小与气隙磁阻 R_c 及 R_{cl}、球面导磁结构磁阻 R_{ms}、磁通等效面积 A 有关。球面 LFMB 比柱面多了一项由球面导磁结构带来的磁阻 R_{ms}，由此可分析球面导磁结构对于 2 种LFMB 的气隙磁密强度、各向均匀性的影响。

4 有限元仿真分析

4.1 有限元建模

为对比分析柱面及球面 LFMB 的气隙磁密分布特征，利用 ANSYS APDL 构建柱面与球面 LFMB 的电磁仿真模型。图 6 为 2 种 LFMB 的有限元计算模型。设定空气域的磁渗透系数为 1，磁钢的磁力系数为 79 600、磁渗透为1.05，二维面域网格采用自由划分。



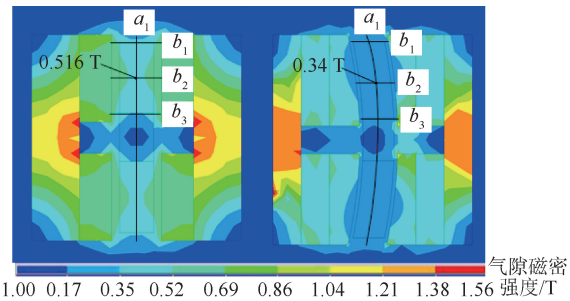
(a) 柱面LFMB (b) 球面LFMB

图 6 两种 LFMB 的有限元建模

Fig. 6 Finite element modeling of two LFMBs

4.2 仿真结果分析

柱面及球面 LFMB 的气隙磁密云图分布如图 7所示。对比可知，球面导磁结构的设计使得球面 LFMB 的气隙磁密分布比柱面更均匀，说明球面导磁结构具备平顺气隙磁场的作用。球面导磁结构延长了磁路闭合路径，导致磁路磁阻增大，当 MSCSG 进行控制敏感时，使得球面 LFMB 需要更大的电流来控制转子偏转。大电流提高球面



(a) 柱面LFMB (b) 球面LFMB

图 7 两种 LFMB 磁密分布云图及路径

Fig. 7 Magnetic density distribution nephogram and paths of two LFMBs

LFMB 功耗的同时，也提升了电流信号信噪分辨率，间接提高了球面 LFMB 的控制敏感精度。

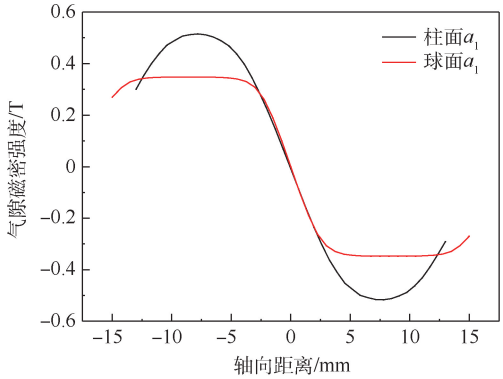
进一步分析 LFMB 的球面设计对气隙磁密分布特征的影响，分别选取了球面及柱面气隙的轴向路径 a_1 及径向路径 $b_1\sim b_3$，如图 7 所示，路径上的磁密分布如图 8 所示。对比图 8(a)轴向路径 a_1，柱面及球面 LFMB 的气隙中心的轴向磁密分布关于气隙中点对称、大小相等、方向相反，在绕组截面的轴向范围内，球面 LFMB 的轴向气隙磁密分布的平顶区域比柱面宽，其中，球面气隙轴向磁密分布在 0.33 ~ 0.34 T 之间，柱面气隙轴向磁密分布在 0.35 ~ 0.51 T 之间。对比图 8(b)径向路径 $b_1\sim b_3$，在绕组截面径向范围内，球面 LFMB 的径向气隙磁密分布的波动比柱面小，其中，球面气隙的 3 条径向磁密分布在 0.32 ~ 0.35 T 之间，柱面气隙的 3 条径向磁密分布在 0.35 ~ 0.55 T 之间。通过图 8 对比可知，当磁钢径向充磁长度、气隙宽度保持不变的情况下，球面 LFMB 的轴向、径向气隙磁密分布波动均比柱面小，球面气隙中心磁密较柱面下降了 34.1%。

由于式(1)和式(2)中的气隙磁密强度 B 为设计给定的常值，因此绕组范围内的气隙磁密分布越均匀，则 LFMB 的控制敏感精度就越高。现引入磁密均匀度以表征气隙磁密的分布，具体评估 2 种 LFMB 在转子偏转前后的磁密分布情况，其定义如下：

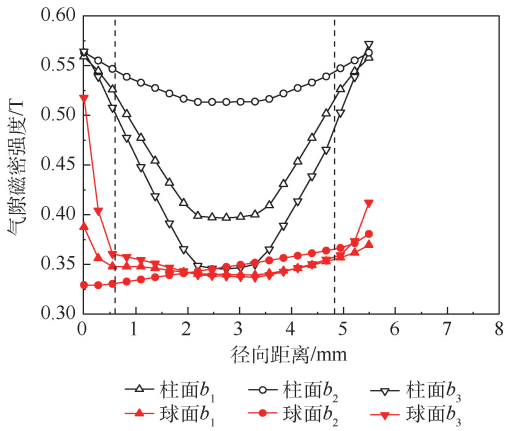
$$\xi=1-\frac{\sqrt{\dfrac{1}{n}\sum_{i=1}^{n}(\overline{B}-B_i)^2}}{\overline{B}} \tag{9}$$

式中：B_i 和 $\overline{B}$ 分别为绕组范围内选取点的磁密和磁密均值。

图 9 为 2 种 LFMB 绕组相对转子偏转前后的位置示意图，各自绕组范围内的磁密均匀度计算结果如表 2 所示。由于 2 种气隙的磁密关于气隙中点对称分布，所以只需分析转子朝一个方向偏



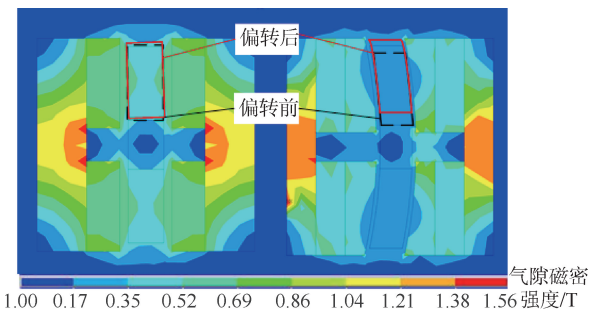
(a) 轴向路径气隙磁密强度



(b) 径向路径气隙磁密强度

图 8　两种 LFMB 气隙磁密强度分布

Fig. 8　Magnetic density distribution of two LFMBs



(a) 柱面LFMB　　(b) 球面LFMB

图 9　两种 LFMB 绕组截面偏转示意图

Fig. 9　Schematic diagram of coil-section deflection of two LFMBs

表 2　两种 LFMB 绕组截面范围内磁密均匀度的对比

Table 2　Comparison of magnetic density uniformity within coil section of two LFMBs

偏转角度/(°)	磁密均匀度	
	柱面 LFMB	球面 LFMB
0	0.882	0.984
0.6	0.834	0.982
2		0.912

转，绕组范围内气隙磁密强度变化。

在绕组截面范围内，当转子未发生偏转时，球面 LFMB 的磁密均匀度比柱面提升了 11.6%；当转子偏转 0.6°时，球面 LFMB 的磁密均匀度较柱面提升了 17.7%，此时对比各自偏转前的气隙磁密均匀度，球面 LFMB 仅降低了 0.2%，优于柱面 LFMB 降低的 5.4%；当转子偏转 2°时，球面 LFMB 的磁密均匀度较偏转前虽然降低了 7.3%，但整体磁密均匀度仍优于转子偏转前的柱面气隙。

通过对比 2 种 LFMB 气隙磁密的分布特点可知，球面 LFMB 气隙磁密的各向均匀度较柱面有了明显提升，虽然球面 LFMB 增大了磁路磁阻，导致气隙磁密下降、控制敏感电流增大，增加了整机功耗，但大电流可提高 LFMB 的控制精度。另外，球面 LFMB 的转子偏角比柱面扩大了 3 倍，使得控制力矩得以持续输出，增大了 MSCSG 的动量包络范围，同时在转子偏转前后，球面气隙磁密强度及均匀度的变化均比柱面小，从而提高了转子处于大偏角时球面 LFMB 的控制敏感精度，使得 MSCSG 能够实现高精度控制力矩持续输出、转子姿态敏感等性能要求。

5　结　论

根据柱面 LFMB 在运行过程中存在的控制精度低、持续输出力矩时间短等问题，提出了一种球面 LFMB 的设计方法，结论如下：

1）球面 LFMB 工作气隙磁密分布比柱面均匀，中心磁密较柱面下降了 34.1%。

2）绕组截面范围内，在转子偏转前后，球面 LFMB 的气隙磁密均匀度相较于柱面 LFMB 分别提升了 11.6%、17.7%，说明球面导磁结构具备平顺球面气隙磁密分布的作用。当球面 LFMB 进行控制敏感时，均匀分布的气隙磁密使得绕组范围内产生的电磁力大小均匀，提高了球面 LFMB 的控制精度。

3）由于球面 LFMB 允许转子进行 ±2°的偏转，控制力矩得以持续输出，增大了 MSCSG 的动量包络范围。且在转子偏转前后，球面气隙磁密大小及均匀度的变化较小，提高了转子处于大偏角时球面 LFMB 的控制敏感精度。

参考文献（References）

[1] BITTERLY J G. Flywheel technology: Past, present, and 21st century projections[J]. IEEE Aerospace and Electronic Systems Magazine, 1998, 13(8): 13-16.

[2] CHRISTOPHER D A, BEACH R. Flywheel technology development program for aerospace applications[J]. IEEE Aerospace and Electronic Systems Magazine, 1998, 13(6): 9-14.

[3] 林士谔. 动力调谐陀螺仪[M]. 北京：国防工业出版社，1983：25-32.

LIN S E. Dynamically tuned gyroscopes[M]. Beijing: National Defense Industry Press, 1983: 25-32 (in Chinese).

[4] PENNÉ B, TOBEHN C, KASSEBOM M, et al. A high agile satellite platform for Earth observation-performance description using new generation missions[C]// 57th International Astronautical Congress. Reston: AIAA, 2006.

[5] KURITA N, ISHIKAWA T, MATSUNAMI M. Basic design and dynamic analysis of the small-sized flywheel energy storage system-application of Lorentz force type magnetic bearing[C]// 2009 International Conference on Electrical Machines and Systems. Piscataway: IEEE Press, 2010: 1-6.

[6] TANG J Q, XIANG B, WANG C E. Rotor's suspension for vernier-gimballing magnetically suspended flywheel with conical magnetic bearing[J]. ISA Transactions, 2015, 58: 509-519.

[7] 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. Magnetically suspended control and sensing gyroscope: CN104613950B[P]. 2017-06-27 (in Chinese).

[8] 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.

[9] 房建成, 任元. 磁悬浮控制力矩陀螺技术[M]. 北京: 国防工业出版社, 2014: 25-30.
FANG J C, REN Y. Magnetically suspended control moment gyroscope technology[M]. Beijing: National Defense Industry Press, 2014: 25-30 (in Chinese).

[10] 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: 58-65.
FANG J C, SUN J J, FAN Y H. Magnetically suspended inertial momentum wheel technology[M]. Beijing: National Defense Industry Press, 2012: 58-65 (in Chinese).

[11] DOWNER J, JOHNSON B. Modelling and control of an annular momentum control device[R]. [S. l.]: NASA Con-tractor Report, 1988: 23-29.

[12] GERLACH B, EHINGER M, RAUE H K, et al. Gimballing magnetic bearing reaction wheel with digital controller[C]// 11th European Space Mechanisms and Tribology Symposium, 2005: 21-23.

[13] 夏长峰, 蔡远文, 任元, 等. MSCSG 转子不平衡振动原理分析与建模[J]. 北京航空航天大学学报, 2018, 44(11): 2321-2328.
XIA C F, CAI Y W, REN Y, et al. Principle analysis and modeling of rotor imbalance vibration in magnetically suspended control and sensing gyroscope[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(11): 2321-2328 (in Chinese).

[14] CHEN X C, CAI Y W, REN Y, et al. Spacecraft angular rates and angular acceleration estimation using single-gimbal magnetically suspended control moment gyros[J]. IEEE Transactions on Industrial Electronics, 2019, 66(1): 440-450.

[15] LIU Q, WANG K, REN Y, et al. Optimization design of launch locking protective device (LLPD) based on carbon fiber bracket for magnetically suspended flywheel (MSFW)[J]. Acta Astronautica, 2019, 154: 9-17.

[16] 辛朝军, 蔡远文, 任元, 等. 磁悬浮敏感陀螺动力学建模与关键误差源分析[J]. 北京航空航天大学学报, 2016, 42(10): 2048-2058.
XIN C J, CAI Y W, REN Y, et al. Dynamic modeling and key error sources analysis of magnetically suspended sensitive gyroscopes[J]. Journal of Beijing University of Aeronautics and Astronautics, 2016, 42(10): 2048-2058 (in Chinese).

[17] CHEN M, KNOSPE C R. Feedback linearization of active magnetic bearings: Current-mode implementation[J]. IEEE/ASME Transactions on Mechatronics, 2005, 10(6): 632-639.

[18] XU G F, CAI Y W, REN Y, et al. Application of a new Lorentz force-type magnetic bearing in tilting control for magnetically suspended control & sensitive gyro-scope with cross sliding mode control[J]. Journal of Aerospace Technology and Management, 2017, 61(1): 107-109.

[19] ZHAO Y, LIU Q, MA L M, et al. Novel Lorentz force-type magnetic bearing with flux congregating rings for magnetically suspended gyrowheel[J]. IEEE Transactions on Magnetics, 2019, 55(12): 1-8.

[20] 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).

[21] 周一恒. 主被动结合磁悬浮隔振平台的基础研究[D]. 哈尔滨: 哈尔滨工业大学, 2019: 105-110.
ZHOU Y H. Fundamental research on active and passive combined magnetic levitation vibration isolation platform[D]. Harbin: Harbin Institute of Technology, 2019: 105-110 (in Chinese).

Design and analysis of high precision for spherical Lorentz force magnetic bearing

FU Baiheng[1], WANG Weijie[2,*], WANG Yuanqin[2], FAN Yahong[3], NIE Chen[4], JIA Haipeng[4]

(1. Company of Postgraduate Management, Space Engineering University, Beijing 101416, China;
2. Department of Aerospace Science and Technology, Space Engineering University, Beijing 101416, China;
3. Beijing Institute of Control Engineering, Beijing 100190, China;
4. The 63961th Unit of the PLA, Beijing 100020, China)

Abstract: Aim at solving following problems. Firstly, the limited deflection angle of the cylindrical Lorentz force magnetic bearing (LFMB) can leads to the short duration of torque output of magnetically suspended control and sensing gyroscope (MSCSG). Secondly, the low uniformity of magnetic density between the LFMB's air gap will also affect the control sensitivity of the MSCSG. A high-precision spherical design and analysis of LFMB are proposed. The spherical LFMB's rotor spherical magnetic sleeve and stator spherical winding are both concentric with the double spherical gyroscope rotor, and the air gap is spherical shell shape, which ensures that the width of the air gap on both sides of the winding coil remains unchanged during deflection. Compared to the cylindrical LFMB, the deflection angle of the spherical LFMB is increased from $\pm 0.6°$ to $\pm 2°$. The mathematical analysis model of the air gap magnetic density of the cylindrical and spherical LFMB is derived by the equivalent magnetic circuit method, and the finite element simulation model of the cylindrical and spherical LFMB is constructed based on the ANSYS command stream. The simulation analysis shows that the maximum magnetic density of the spherical LFMB decreases by 34.1% compared to the cylindrical LFMB within the deflection range of the rotor along the deflection centerline. When the rotor is not deflected, the uniformity of the magnetic density at the cross-section of the spherical LFMB coil increases by 11.6% compared to the cylindrical. When the rotor is deflected, the uniformity of the magnetic density at the cross-section of the spherical LFMB coil increases by 17.7% compared to the cylindrical. The proposed method lays the foundation for the improvement of the control and sensitivity of the magnetically suspended control and sensing gyroscope.

Keywords: magnetically suspended control and sensing gyroscope; spherical Lorentz force magnetic bearing; magnetic density uniformity; equivalent magnetic circuit method; finite element analysis

Received: 2021-03-02; **Accepted**: 2021-05-05; **Published online**: 2021-05-18 13:23
URL: kns.cnki.net/kcms/detail/11.2625.V.20210518.1013.003.html
Foundation item: National Natural Science Foundation of China (52075545)
* **Corresponding author**. E-mail: wangwjie@126.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0106

旋转抛物-锥形蚀坑模型及其应用

刘德俊[1], 田干[1,*], 金国锋[1], 杨正伟[1], 任碧云[1], 魏花丽[2]

(1. 火箭军工程大学 导弹工程学院, 西安 710025; 2. 西安航天复合材料研究所, 西安 710025)

摘 要: 点蚀是腐蚀介质环境中金属性能劣化的主要形式之一。蚀坑导致的应力集中会使结构整体强度退化,降低结构的安全性和可靠性。因此,准确的蚀坑模型对结构应力场的分析具有重要影响。为此,在典型的蚀坑形貌基础上,结合蚀坑张开角的定义,提出旋转抛物-锥形蚀坑模型,并通过有限元仿真与拉伸破坏试验验证了模型的有效性。结果表明:蚀坑导致的应力集中极大值处于接近坑底或坑口区域,旋转抛物-锥形蚀坑模型在蚀坑坑底、坑肩及坑口处的应力集中分布带比半椭球形蚀坑模型更准确,应力敏感性更强。

关 键 词: 蚀坑; 旋转抛物-锥形模型; 半椭球形模型; 应力集中; 结构强度

中图分类号: V252

文献标志码: A **文章编号**: 1001-5965(2022)11-2230-11

火箭推进剂贮箱采用高强度铝合金材料,海底管道、陆地长输管道及压力容器采用不锈钢或低合金钢等材料。在推进剂、海水等输送介质或盛装介质腐蚀作用下,这些金属材料表面极易发生点蚀。由点蚀导致的应力集中会引起裂纹的萌生、扩展等断裂失效问题,从而降低设备的使用寿命,因此,关于蚀坑及其表征模型强度问题的研究一直广受学者关注[1-7]。

为了准确表征蚀坑产生的影响,1954 年,Eubanks[8]采用半球描述点蚀坑,实现了蚀坑应力集中的近似分析。但是半球模型无法体现蚀坑形状比对应力集中的影响,为此,1978 年,Fujita 和 Sadavasu[9]将半球形模型发展为半椭球模型,为研究蚀坑的影响提供了更为有效的方法。在此基础上,Cerit[10]通过对半椭球蚀坑应力分布的有限元分析,发现随着蚀坑形状比的改变,最大应力出现在蚀坑底部或坑口下侧位置,拟合出蚀坑与应力集中系数之间的经验公式,并将该成果应用于压力容器的强度分析[11-12],为含蚀坑状态下的压力容器安全分析提供了有效途径。类似地,Hou 和 Song[13]则基于半椭球形蚀坑模型研究了受拉伸载荷下的蚀坑应力分布。结果发现应力集中系数与蚀坑形状比线性相关,但与蚀坑相对深度成非线性关系。2016 年,Li[14]采用半椭球形模型探究了主、次蚀坑的宽度、深度和距离等因素对金属结构强度的影响,结果发现蚀坑的萌生会导致应力集中系数发生乘数效应,使结构的剩余强度大幅度地降低;2019 年,Liu 等[15]讨论了蚀坑与材料内部孔洞对应力集中的影响,提出了可表征蚀坑形状参数、缺陷距离比等因素的应力集中系数经验公式,丰富了蚀坑模型的研究范围。

但是在工程实际中,蚀坑的几何形状十分复杂,半椭球模型与真实的蚀坑具有非常大的差别。文献[16-18]也认为半椭球形蚀坑演化出的半椭圆裂纹会极大地高估蚀坑引起的应力奇异性,从而会高估结构的剩余强度,造成不必要的资源浪费。为此,Turnbull 等[19]在试验观察的基础上提出了更能反映蚀坑真实形貌的 U 形蚀坑模型,其

收稿日期: 2021-03-03; **录用日期**: 2021-06-04; **网络出版时间**: 2021-06-24 17:20
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210624.1613.001.html
基金项目: 国家自然科学基金(52075541,52272446); 陕西省自然科学基金(2020JM-354)
* **通信作者**. E-mail: tiangan_2012@163.com

引用格式: *刘德俊, 田干, 金国锋, 等. 旋转抛物-锥形蚀坑模型及其应用[J]. 北京航空航天大学学报, 2022, 48(11): 2230-2240.*
LIU D J, TIAN G, JIN G F, et al. Rotating parabolic-conical corrosion pit model establishment and its application[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2230-2240(in Chinese).

应力集中现象比半椭球蚀坑更为突出。在此基础上,张川和姚卫星[20]利用 U 形蚀坑模型研究了危险蚀坑的应力状态,并给定了危险蚀坑的判别准则。文献[21]认为具有尖深边缘且能反映蚀坑底部形貌的蚀坑模型能够准确地表征蚀坑造成的应力集中问题,但是并未建立相应的模型。因此,本文为了准确地表征点蚀坑形貌,描述点蚀产生的应力集中效应,基于蚀坑张开角的重新定义,提出了旋转抛物-锥形蚀坑模型,有效地反映出不同形状比的蚀坑形貌。同时,与半椭球形蚀坑模型进行了有限元对比分析,得到了不同形状比下的蚀坑应力分布规律。最后,采用扫描电镜(scanning electron microscope, SEM)观察拉伸破坏试验件的蚀坑截面,验证了模型的正确性。

1　旋转抛物-锥形蚀坑模型建立

腐蚀环境下,结构表面蚀坑形貌不尽相同。面积、深度、长度等形状特征突出的蚀坑不但影响结构整体强度,还能导致裂纹萌生致使材料断裂破坏。如图 1 所示,暴露于腐蚀介质中的结构蚀坑形貌,通过观察呈现 2 种典型情形[20]:图 1(a)为浅平蚀坑,特点是坑底狭长,材料表面衔接处应力相对集中;图 1(b)为尖深蚀坑,特点是坑底狭窄,应力集中区域多聚集在坑底,且比浅平蚀坑更具破坏性。

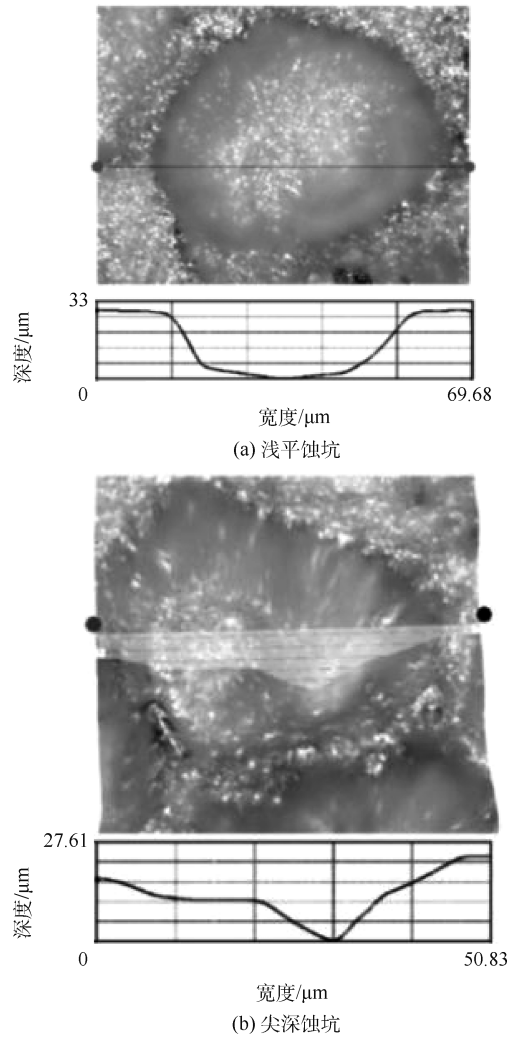


图 1　典型蚀坑图

Fig. 1　Typical corrosion pit patterns

在蚀坑形貌观察中,坑底的轮廓可近似简化为光滑过渡。因此,坑底附近区域截面可通过控制曲率半径 r 的大小来确定蚀坑坑底尖锐的程度。基于上述假设,提出了描述尖深蚀坑及浅平蚀坑的旋转抛物-锥形蚀坑模型。

1.1　尖深蚀坑截面

对于尖深蚀坑,坑底曲率半径 r 较小且小于蚀坑深度,因此,在蚀坑轮廓线底部做以 r 为半径的曲率圆。该圆作为旋转抛物-锥形模型的基础起到确定坑底位置的作用。

如图 2 所示,蚀坑截面轮廓线为 $GDICH$。以坑底 I 为原点,x、y 轴如图 2 所示,作以 O 为圆心、r 为坑底曲率半径的曲率圆,由于旋转抛物体母线的系数与曲率半径相关,因此过 I 点做抛物线 L,图 2 中 m 为蚀坑半长,h 为蚀坑深度。

设抛物线 L 方程为

$$y = px^2 \tag{1}$$

式中:p 为旋转物体母线系数。

因为 I 为抛物线顶点,则

$$p = 1/(2r) \tag{2}$$

过曲率圆分别作平行于 y 轴且与圆分别相切于点 A、B 的直线 AE 和 BF,与抛物线 L 交于点 $D(r,r/2)$ 和 $C(r,r/2)$,定义 $\angle EOF$ 为蚀坑张开角,分别将 G、D 和 C、H 线形连接,从而形成尖深蚀坑轮廓线。

根据直线解析式及 GD、CH 坐标得到解析式:

$$\begin{cases} y_{GD} = \dfrac{r/2 + h}{m - r}x + \dfrac{rm/2 - hr}{m - r} \\ y_{CH} = \dfrac{r/2 + h}{r - m}x + \dfrac{rm/2 - hr}{m - r} \end{cases} \tag{3}$$

由相应几何关系得到尖深蚀坑张开角 θ 与坑底曲率半径和蚀坑深度关系:

$$\theta = 2\arctan\left(\frac{r}{h - r}\right) \tag{4}$$

1.2　浅平蚀坑截面

对于浅平蚀坑,由于坑底曲率半径 r 较大且常大于坑深 h,如图 3 所示,因此,浅平蚀坑轮廓线的构造与尖深蚀坑轮廓线构造方法有所不同。

如图 3 所示,蚀坑截面轮廓为 $GDICH$。参照

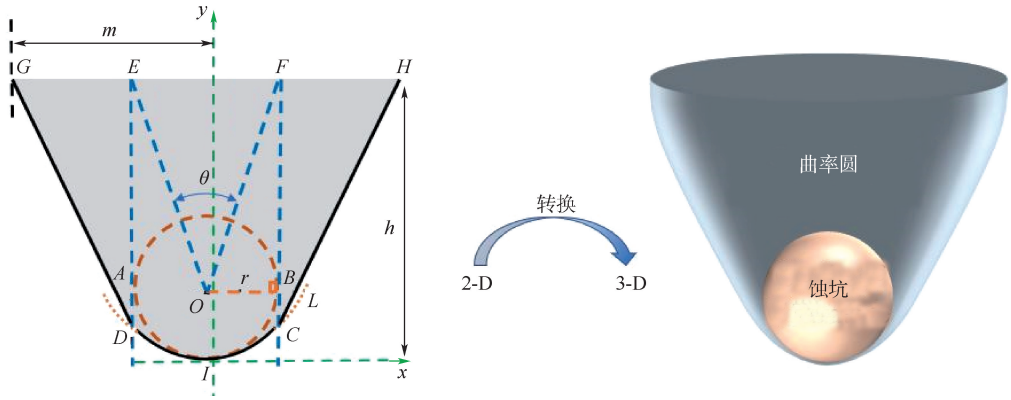


图 2 尖深蚀坑示意图

Fig. 2 Sharp-deep corrosion pit

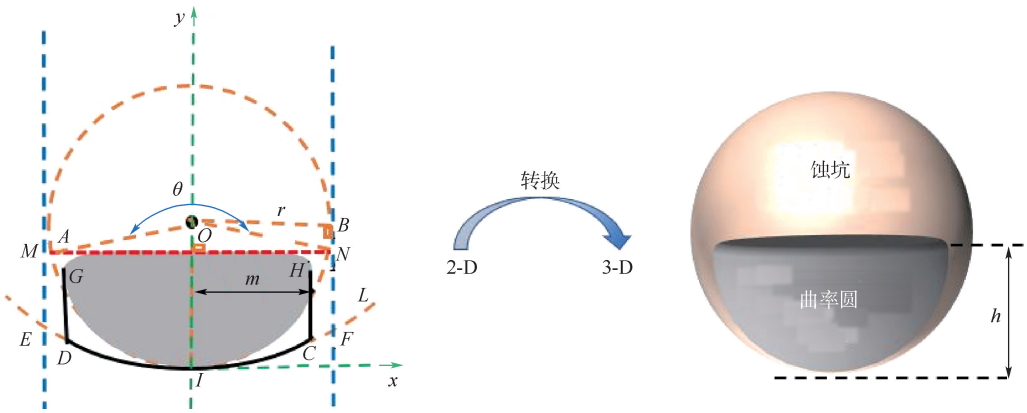


图 3 浅平蚀坑示意图

Fig. 3 Shallow-flat corrosion pit

1.1 节,I 为原点,作以 O 为圆心,r 为半径的曲率圆,根据式(1)和式(2)做过点 I 的抛物线 L。AE 和 BF 是平行于 y 轴与圆相切的 2 条切线,A、B 为切点,E、F 是与抛物线 L 的交点;延长蚀坑上边 GH 与曲率圆相交于 M、N,即 MN 为曲率圆的割线。过 G、F 作垂线与抛物线 L 交于 D、C,从而形成浅平蚀坑轮廓线。进而得到 GD、CH 两个垂线方程为

$$y_{GD/CH} = \pm m \tag{5}$$

定义圆周角 $\angle EOF$(钝角)为蚀坑张开角 θ,由几何关系得到浅平蚀坑张开角 θ 与坑底曲率半径和蚀坑深度的关系:

$$\theta = 2\pi - 2\arccos\left(\frac{r-h}{r}\right) \tag{6}$$

通过对尖深蚀坑与浅平蚀坑建模分析,能够得到张开角是关于蚀坑深度及蚀坑底部曲率半径的函数。当 $r<h$ 时,蚀坑为尖深蚀坑,此时张开角在(0,90°)内;当 $r\geqslant h$ 时,蚀坑为浅平蚀坑,此时张开角范围为[90°,180°)。为简化蚀坑模型表达,定义形状比 $\lambda = m/h$,坑底平整度 $\gamma = r/h$,得到抛物线-线型蚀坑模型如下:

1)尖深蚀坑,$\theta\in(0,90°)$,$r<h$

$$\begin{cases} y_{GD} = \dfrac{\gamma/2+1}{\lambda-\gamma}x + \dfrac{\lambda\gamma h/2 - h\gamma}{\lambda-\gamma} & -m \leqslant x < -r/2 \\ y_{CD} = \dfrac{1}{2\gamma h}x^2 & -r/2 \leqslant x \leqslant r/2 \\ y_{CH} = \dfrac{\gamma/2+1}{\gamma-\lambda}x + \dfrac{\lambda\gamma h/2 - h\gamma}{\lambda-\gamma} & r/2 < x \leqslant m \\ \theta = 2\arctan(\gamma - \gamma h) \end{cases} \tag{7}$$

2)浅平蚀坑,$\theta\in[90°,180°)$,$r\geqslant h$

$$\begin{cases} x = -\lambda h \\ y_{CD} = \dfrac{1}{2\gamma h}x^2 & -m \leqslant x \leqslant m \\ x = \lambda h \\ \theta = 2\pi - 2\arccos(\gamma - 1/\gamma) \end{cases} \tag{8}$$

1.3 蚀坑截面比较

半椭球形蚀坑是根据蚀坑近似长度及深度简化成椭圆长短径而得到的,该蚀坑模型虽然比较简便,但是其关于应力集中敏感性仍有待商榷。因此,该部分将通过改变形状比 $\lambda = m/h$、坑底

平整度 $\gamma = r/h$，将不同蚀坑张开角下的旋转抛物-锥形蚀坑与等深等长的半椭球蚀坑截面进行比较，其中 h 为 0.5 mm。如图 4～图 6 所示，当蚀坑长度和深度一定，半椭球蚀坑对蚀坑形态的变化不敏感，而旋转抛物-锥形蚀坑则能通过坑底曲率半径的改变调节坑底的尖锐或平整程度，即随着曲率半径的增大，蚀坑可由尖深坑过渡到浅平坑。

对于 $\theta < 90°$的尖深蚀坑，半椭球形蚀坑所围

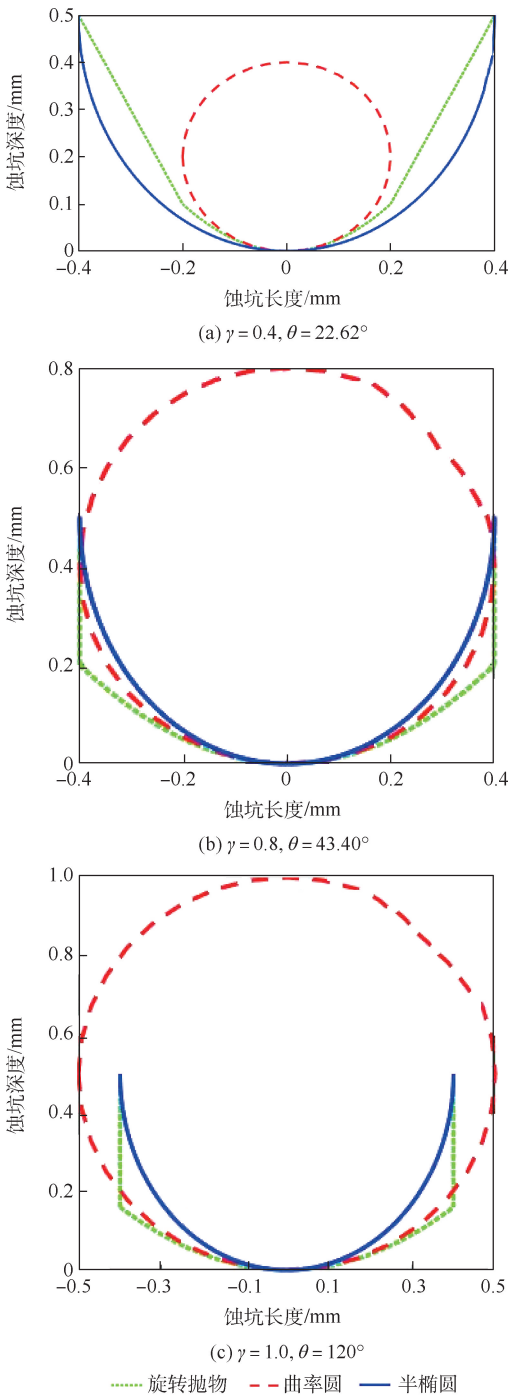


图 4　$\lambda = 0.8$ 的蚀坑形貌比较

Fig. 4　Comparison of corrosion pit morphology under $\lambda = 0.8$

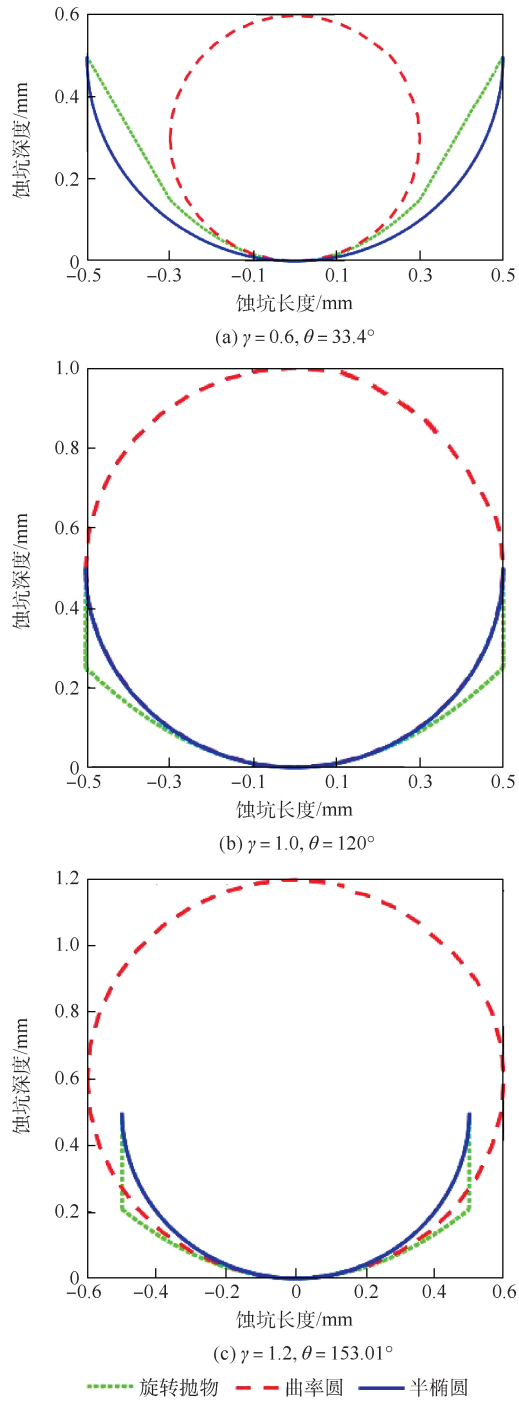


图 5　$\lambda = 1.0$ 的蚀坑形貌比较

Fig. 5　Comparison of corrosion pit morphology under $\lambda = 1.0$

成的截面面积大于旋转抛物-锥形蚀坑的截面面积，因此后者更能反映应力集中的现象；而对于 $\theta > 90°$的浅平蚀坑，旋转抛物-锥形模型截面面积更大，所多出的面积主要集中在蚀坑侧面边缘处，因此旋转抛物-锥形蚀坑从形貌上更能符合该类蚀坑的“浅、平”的特点。综上，与常规的半椭球蚀坑相比，旋转抛物-锥形蚀坑能够在蚀坑深度、长度一定的条件下针对蚀坑尖锐平整程度较为准确的反映蚀坑形貌。

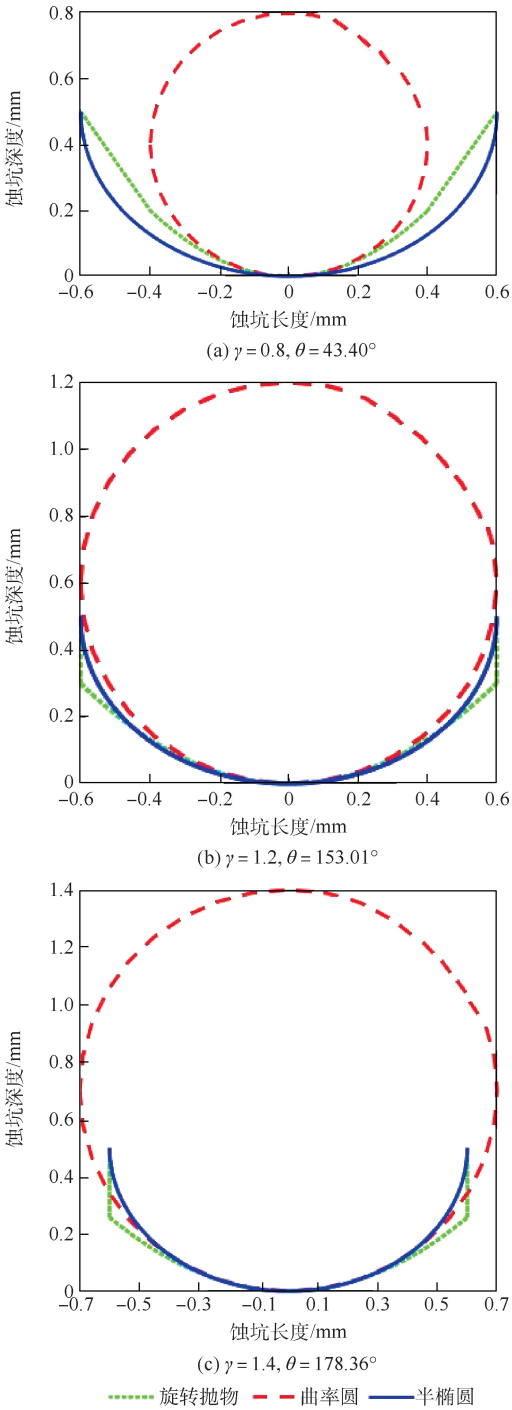


图 6 $\lambda=1.2$ 的蚀坑形貌比较

Fig. 6 Comparison of corrosion pit morphology under $\lambda=1.2$

2 三维蚀坑数值分析

2.1 有限元模型建立

分别将相同 λ 的旋转抛物-锥形及半椭球形蚀坑建立在长为 10 mm、宽为 10 mm、高为 5 mm 的长方体平板上，材料为铝合金，弹性模量为 70 GPa，泊松比 0.28，屈服应力 290 MPa。如图 7 所示，平板一端固定，另一端施加 100 MPa 的均匀压力，进而得到不同 λ 的蚀坑周边应力大小及分布情况；网格划分中，蚀坑周围采用 10 节点的

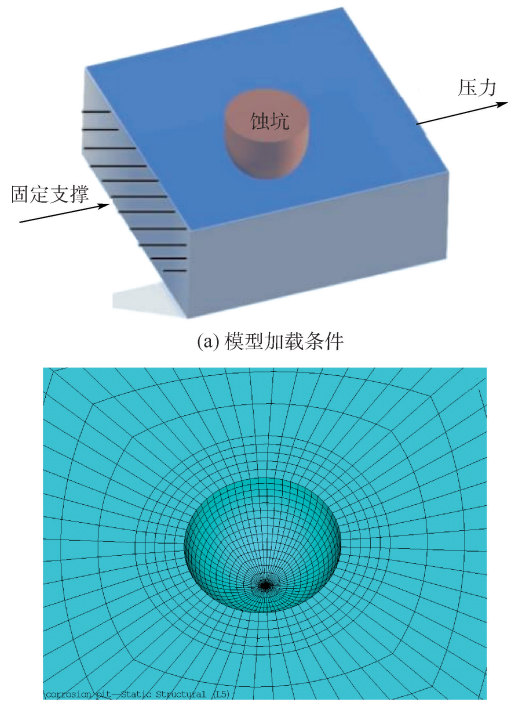


图 7 模型加载条件与蚀坑网格模型

Fig. 7 Boundary condition and mesh model

Solid187四面体单元，蚀坑其余部分通过 20 节点 Solid186 六面体单元划分，采用局部控制的方式对蚀坑周边进行局部加密。

为证实数值建模的准确性，文献[13]已对有限元法在不同蚀坑应力分析中得到了数值上的应用，并将蚀坑数值应力求解与理论求解进行了比对。因为弹性力学中带圆孔平板的理论应力分布已知，所以由圆孔中心处向外不同位置处的 Mises 应力能够被估计。

表 1 比较的是平板上半径 b 为 0.4 mm 的孔，在拉伸荷载 S 为 100 MPa 的作用下，沿与拉伸方向夹角 δ 为 90°的孔边应力数值结果与理论解析解进行比较，带孔平板如图 8 所示，可以看出在相

表 1 理论与数值 Mises 应力比较

Table 1 Comparison of theoretical and numerical Mises stress

R(垂直拉伸方向上的节点距孔中心距离/mm)	理论 Mises 应力/MPa	数值 Mises 应力/MPa	误差/%
0.4	300	301.97	0.66
0.464	220.02	232.21	5.54
0.5	193.44	200.11	3.45

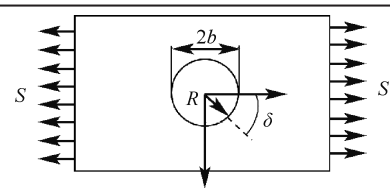


图 8 带孔平板示意图

Fig. 8 Schematic diagram of plate with hole

同位置,数值求解结果与理论 Mises 应力值的主要误差低于 6%,符合精度要求。由此可说明含蚀坑平板在拉伸载荷下的数值分析结果具有一定的准确性。

2.2　结果分析与讨论

为探究旋转抛物-锥形蚀坑适用性以及蚀坑应力变化规律,因此设定蚀坑深度 h 为 0.5 mm,λ 分别为 0.8、1.0 与 1.2,分析不同 γ 下的蚀坑应力分布。此外,建立长短径分别为 m、h 的半椭球蚀坑模型作为数值结果对比,进而得到不同蚀坑模型应力数值分析结果。

图 9 为 λ =0.8 时的蚀坑局部应力分布。当 γ 从 0.4 增长到 1.0,蚀坑张开角 θ 随之增大,即蚀坑从尖深型到浅平型转变过程中,蚀坑局部应力集中带从蚀坑坑底转移至坑口附近。θ 较小时(见图 9(a)),应力集中带在坑肩以下,且带宽较窄呈长条状分布,应力最大值靠近坑底;随着 θ 增大,应力集中带分散为两部分分别位于坑肩处,而坑底呈现出应力分布均匀无集中的结果(见图 9(b)、(c))。应力数值上,随着 θ 增大,Mises 应力最大值 σ_{max} 从 236 MPa 降低为 213 MPa,但是 θ =43.60°时,σ_{max} 为 215 MPa 与浅平蚀坑应力数值相近,说明蚀坑应力集中带接近时,蚀坑局部应力大小对 γ 的改变不再敏感。对半椭球形蚀坑,应力集中带带宽较宽且贯穿坑底、坑肩;应力最大值位于坑肩处,数值为 202 MPa(见图 9(d))。由此可知,对于给定 m、h 的半椭球形蚀坑,无法准确反映不同类型蚀坑的应力集中带,其应力数值结果没有旋转抛物-锥形蚀坑保守。

图 10 为 λ =1.0 时的蚀坑局部应力分布。当蚀坑为尖深型时(见图 10(a)),应力集中带较宽,聚集在坑肩下方的位置且呈轴对称分布;当蚀坑转变为浅平蚀坑时(见图 10(b)、(c)),应力集中带分成 3 部分,一部分在坑底,另两部分对称分布在坑肩两侧;而且 γ 越大,坑底应力分布带越向坑两侧蔓延。应力数值上,当 θ <90°时,σ_{max} 为 203 MPa,当 θ ≥90°时,由于浅平蚀坑应力带分布相近,σ_{max} 为 190 MPa 左右。由此看出,随着 λ 增大,蚀坑局部应力逐渐降低。对于半椭球蚀坑(见图 10(d)),应力集中带仍贯穿坑底与坑口,应力数值仍旧低于旋转抛物-锥形蚀坑的大小。

图 11 为 λ =1.2 时的蚀坑局部应力分布。无论是尖深或是浅平蚀坑,其应力分布带与 λ =1.0 时所对应的蚀坑应力分布大致相同,不同的是应力分布带带宽在 θ >90°时更宽更靠近蚀坑两侧;

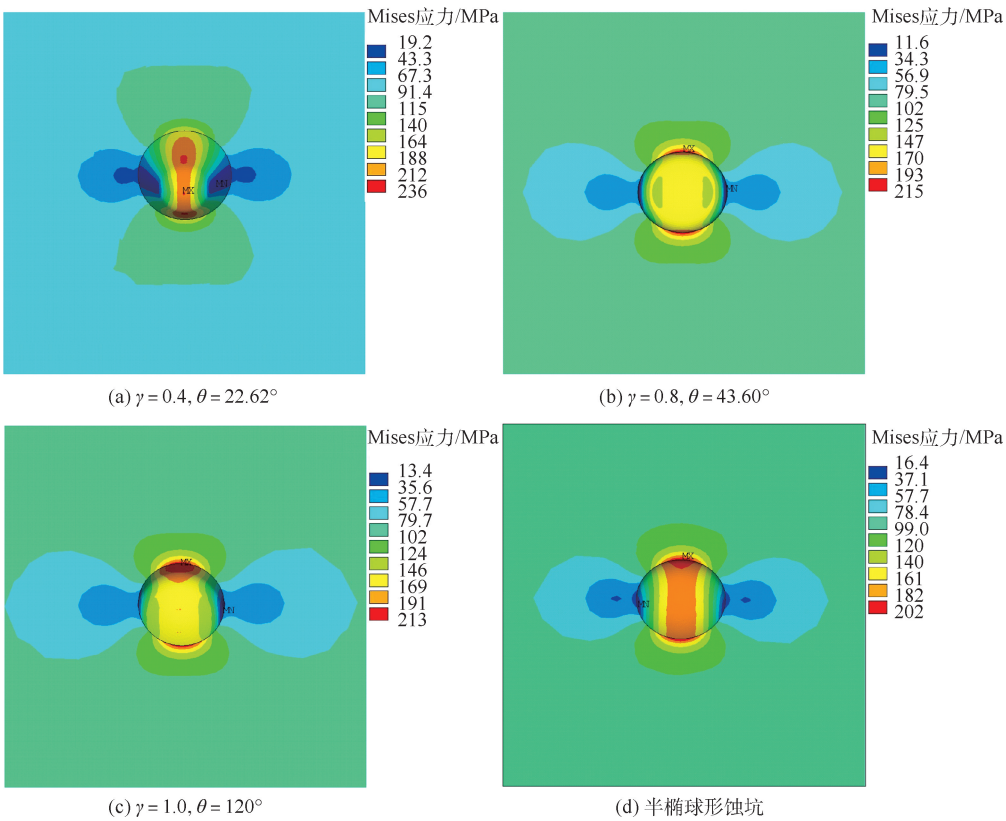
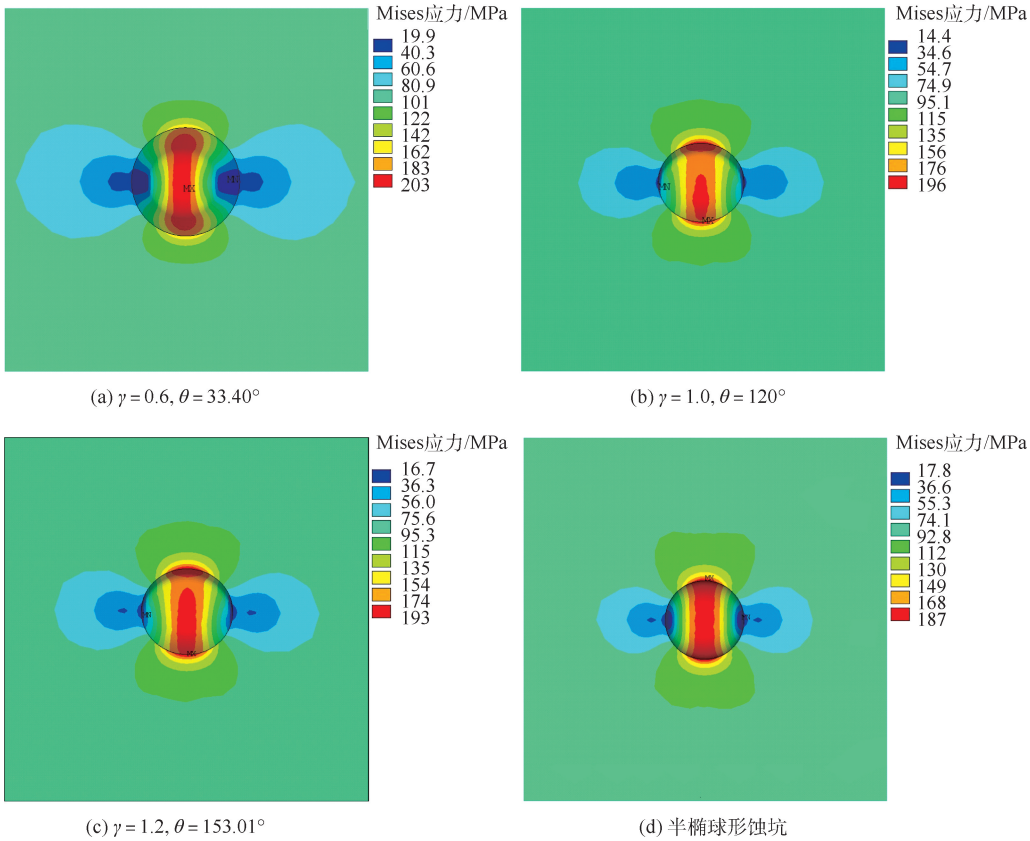


图 9　λ =0.8 下的蚀坑局部应力分布

Fig. 9　Local stress contribution of corrosion pits under λ = 0.8

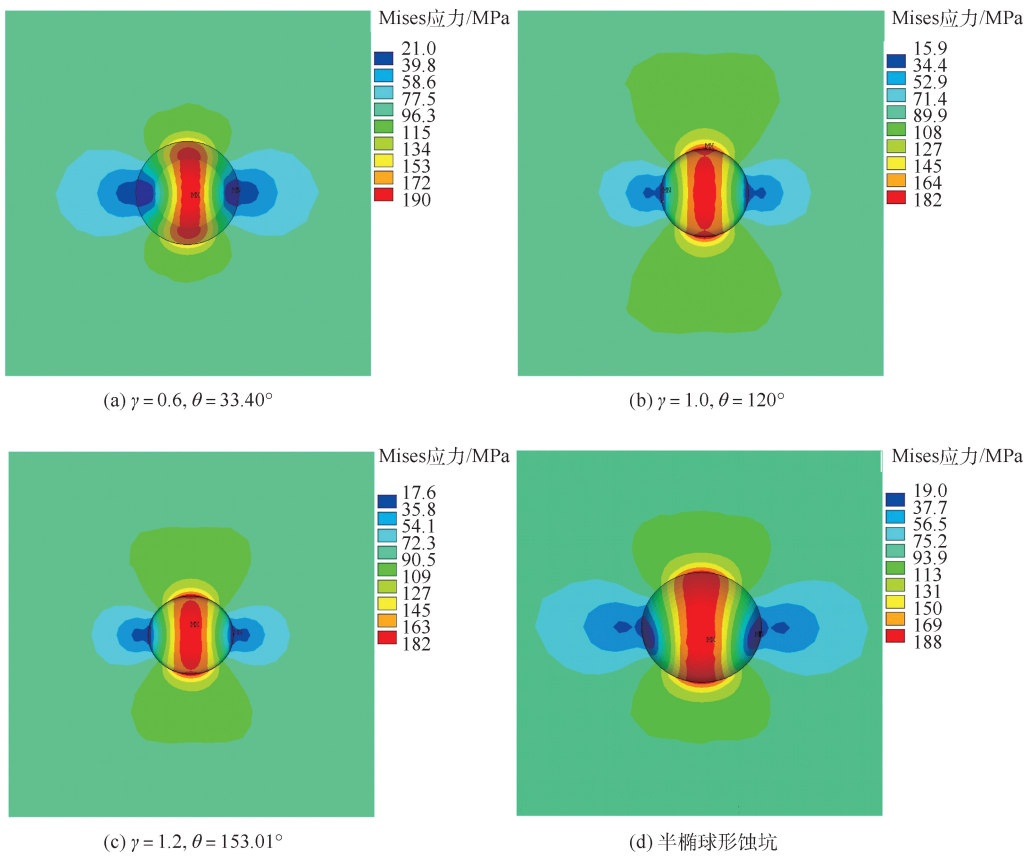


(a) $\gamma=0.6$, $\theta=33.40°$ (b) $\gamma=1.0$, $\theta=120°$

(c) $\gamma=1.2$, $\theta=153.01°$ (d) 半椭球形蚀坑

图 10 $\lambda=1.0$ 下的蚀坑局部应力分布

Fig. 10 Local stress contribution of corrosion pits under $\lambda=1.0$



(a) $\gamma=0.6$, $\theta=33.40°$ (b) $\gamma=1.0$, $\theta=120°$

(c) $\gamma=1.2$, $\theta=153.01°$ (d) 半椭球形蚀坑

图 11 $\lambda=1.2$ 下的蚀坑局部应力分布

Fig. 11 Local stress contribution of corrosion pits under $\lambda=1.2$

蚀坑 σ_{max} 数值小，表明 $\lambda>1$ 时所引起的应力集中程度低。对于半椭球蚀坑（见图 11(d)），应力分布与图 10 相同，σ_{max} 数值基本相同。

综合图 9～图 11，对于半椭球形蚀坑，其应力分布带都是垂直于拉伸载荷方向且呈左右对称分布。当 $\lambda<1$ 时，σ_{max} 位置在坑肩，而 $\lambda\geqslant1$ 时，应力分布带贯穿坑底至坑口，而且应力大小相近，因此无法判断由蚀坑导致的破坏区域的具体位置，无法衡量坑底、坑肩及坑口的应力集中敏感性，所以半椭球形蚀坑模型在结构安全强度评估上具有一定缺陷。对于旋转抛物-锥形蚀坑模型，λ 一定时，随着 γ 及 θ 的改变，蚀坑的应力集中分布位置会随之变化。当 $\lambda<1.0$ 时，θ 较小时，应力集中带在坑肩下方且带宽较窄呈长条状分布，σ_{max} 出现在坑底附近；随着 θ 增大，应力集中带则分散在蚀坑两侧。当 $\lambda\geqslant1$ 且 $\theta\leqslant90°$，应力集中带在坑肩下方且沿中心轴对称；当 $\theta>90°$，应力集中带分为 3 部分，一部分在坑底附近，其余对称分布在蚀坑两侧，而且随着 θ 的增大，应力集中带带宽增大，并沿中心轴两侧蔓延。应力分布带的存在说明该处应力集中水平相对其他位置更高，尖深蚀坑应力集中带的连片表明其受载状态下应力集中效应显著，呈现单一局部的特点；而浅平蚀坑应力集中带的分散反映了应力集中的多源性，具有整体性特征，导致裂纹萌生由于应力强度间的竞争关系而被削弱，因此尖深蚀坑会对结构安全产生更为不利的影响。相对于半椭球形蚀坑所呈现的仅仅是贯穿坑底至坑口的应力集中带，旋转抛物-锥形蚀坑能够对不同类型蚀坑引起的应力集中更为准确的判断与评估。

图 12 反映了在不同 λ 取值下，应力沿旋转抛物-锥形和半椭球形蚀坑截面曲线的数值分布。对旋转抛物-锥形蚀坑，在图 12(a)中，当 $\lambda<1.0$ 时，尖深蚀坑截面应力数值（$\gamma=0.4$）从坑底到坑口呈现先上升再下降的趋势，其 σ_{max} 在坑底附近，且坑底处整体应力大小数值较高；而浅平蚀坑截面应力数值分布则截然相反，σ_{max} 在坑口附近且坑底附近应力大小较低。图 12(b)、(c)中，尖深蚀坑截面应力数值（$\gamma<1$）从坑底到坑口呈现先下降再上升然后再下降的趋势，σ_{max} 在坑底附近；浅平蚀坑截面应力数值（$\gamma\geqslant1$）由坑口附近先下降再上升，但是坑口与坑底附近的应力数值相近，且随着 γ 的增大，数值上越来越接近相等。而对半椭球形蚀坑，当 $\lambda<1$ 时，应力数值由坑底到坑口逐渐增大，σ_{max} 在坑口处；当 $\lambda\geqslant1$ 时，应力数值由坑底到坑口先增大再减小，σ_{max} 在坑口附近。

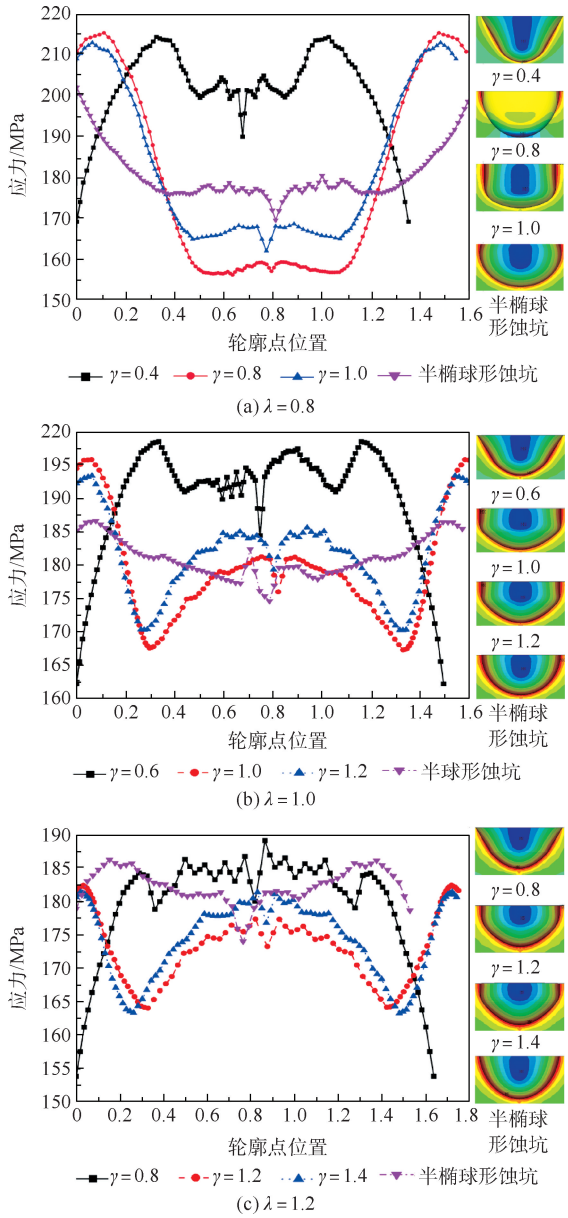


图 12　应力沿不同 λ 值的蚀坑截面曲线分布

Fig. 12　Stress along corrosion pit sections under various λ value

尤为注意的是，2 种类型蚀坑坑底处的应力总是小于其附近应力大小，而且应力大小有波动。此外，在蚀坑底部和坑口邻近范围内，应力大小分别沿着蚀坑半长方向单调递减，从而导致应力最大的位置并非在坑口或坑底，而是位于其附近位置。

为衡量旋转抛物-锥形蚀坑的应力集中效应，通过应力集中系数 K 表征，定量表达式为

$$K=\frac{\sigma_{max}}{\sigma_{norm}} \tag{8}$$

式中：σ_{max} 和 σ_{norm} 分别为表示几何形状不连续区域的局部最大弹性应力和远场名义应力。

如表 2 所示，旋转抛物-锥形蚀坑模型相对半椭球蚀坑的应力集中系数高，得到的结果更为保守，尤其对于尖深蚀坑的应力计算，更能反映尖深

表 2 旋转抛物-锥形蚀坑与半椭球形蚀坑应力集中系数

Table 2 Stress concentration factors between rotating parabolic-conial corrosion pit and semi-ellipsoid corrosion pits

γ	K		
	$\lambda=0.8$	$\lambda=1.0$	$\lambda=1.2$
0.4	2.142		
0.6		2.028	
0.8	2.128		1.904
1.0	2.152	1.959	
1.2		1.934	1.825
1.4			1.818
半椭形蚀坑	2.022	1.866	1.876

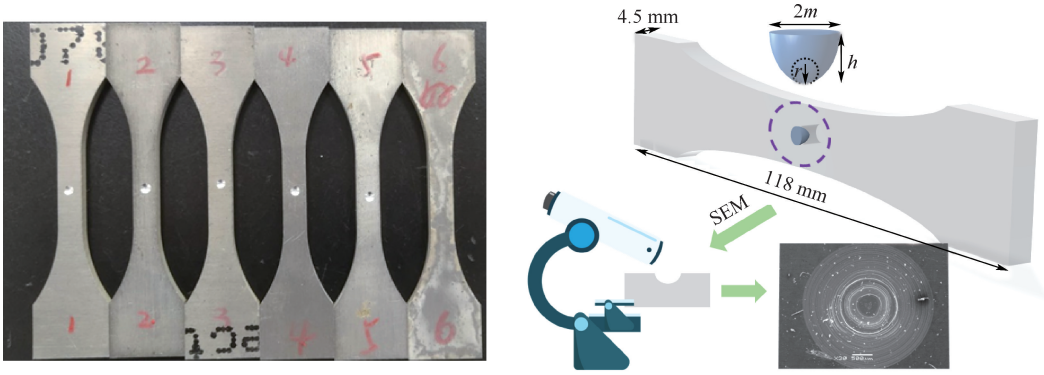
蚀坑应力集中的严重程度,有利于对点蚀结构安全强度评价。

3 试验及验证

为进一步探究蚀坑应力集中带在拉伸载荷下的分布情况,因此,对含蚀坑 2195 Al-Li 合金试样进行拉伸试验,并利用 SEM 对蚀坑拉伸断面进行观测分析。

如图 13 所示,采用长为 118 mm、宽为 25 mm、高为 4.5 mm 的 6 个铝合金拉伸试样,并在其中 5 个试样中间位置人工构造不同 λ、γ 类型的蚀坑,同时蚀坑表面进行了光滑处理。根据 GB/T 228.1—2021 金属拉伸试验标准[22],在 INSTRON 8801 拉伸机上以1 mm/min速率对试样进行拉伸试验,之后取断裂试样,在 Hitachi Su8010 扫描电镜上进行断面观察。

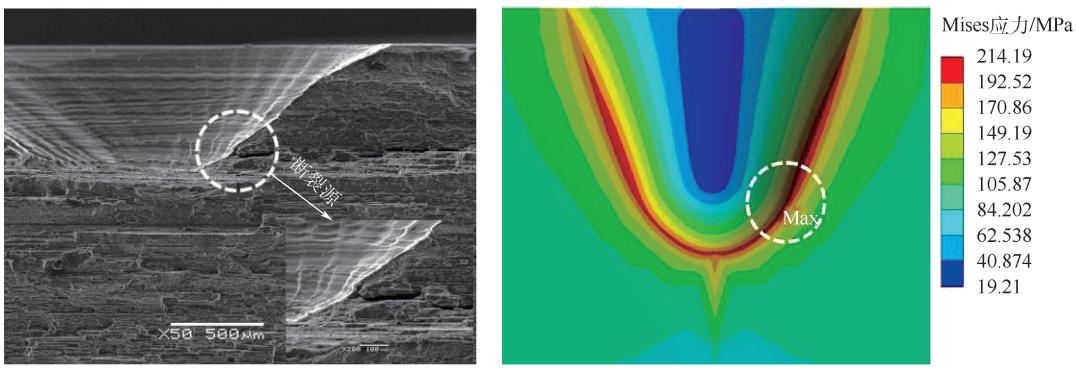
图 14 为不同 λ、γ 参数下的蚀坑断面图和对应的有限元应力云图。从断面 SEM 图中可以看出,2195 Al-Li 合金以沿晶断裂为主,且内部伴有粗大析出相和细小孔洞,由于这些细观缺陷及层状晶粒分布较为均匀,因此,对不同类型蚀坑断裂源分析及模型验证较为有利。图 14(a)、(b)为 $\lambda=0.8$、$\gamma=0.4$ 时尖深程度较高的尖深蚀坑断面和应力云图,可看出在蚀坑坑底邻近区域由于应力集中导致了断裂产生并伴随着沿晶开裂的现象,蚀坑侧边缘其余位置未发生裂纹萌生,其邻近形貌呈现穿晶断裂的特征,说明蚀坑应力集中位置对材料的使用有较大影响。图 14(c)、(d)为 $\lambda=1.0$、$\gamma=0.6$ 时尖深蚀坑断面和应力云图,表明其应力分布位置与图 14(b)近似。从断面上看,起裂位置在蚀坑坑底附近,蚀坑侧边缘没有出现与起裂位置相似的沿晶断裂形貌,其现象与图 14(a)一致。但是坑底处留有棱锥状晶粒残留,说明由于坑底应力大小比邻近区域应力数值低,使得坑底处呈现最终瞬时拉断的断裂特征。



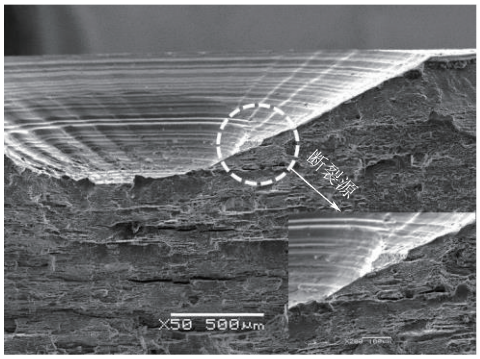
(a) 拉伸试样 (b) 试验尺寸及蚀坑SEM图

图 13 拉伸试样及蚀坑形貌 SEM 图

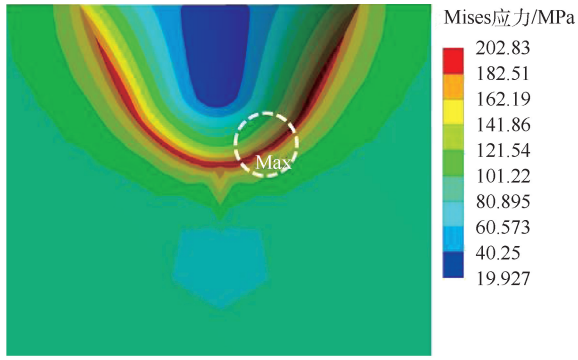
Fig. 13 Tensile samples and corrosion pit morphology of SEM



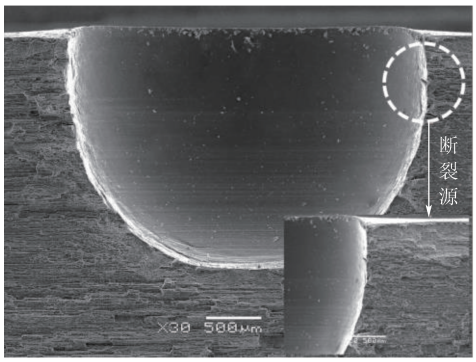
(a) $\lambda=0.8$, $\gamma=0.4$蚀坑断口形貌 (b) $\lambda=0.8$, $\gamma=0.4$蚀坑应力云图



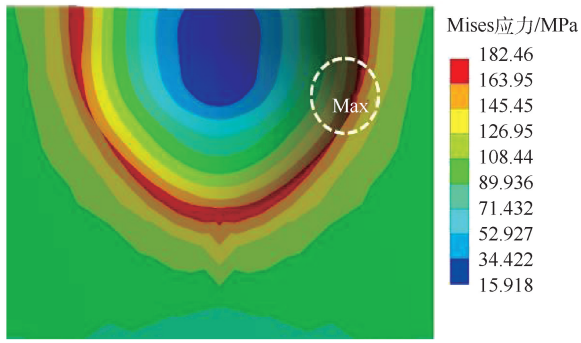
(c) $\lambda=1.0,\gamma=0.6$蚀坑断口形貌



(d) $\lambda=1.0,\gamma=0.6$蚀坑应力云图



(e) $\lambda=1.2,\gamma=1.2$蚀坑断口形貌



(f) $\lambda=1.2,\gamma=1.2$蚀坑应力云图

图 14　蚀坑断面图与应力云图

Fig. 14　Cross section and stress contour of corrosion pits

图 14(e)、(f)为 $\lambda=1.2$、$\gamma=1.2$ 时浅平蚀坑断面和应力云图,不同于尖深蚀坑,其断裂源在坑口附近区域,而位于断裂源以下的蚀坑轮廓断裂特征与尖深蚀坑边缘的断裂形貌基本一致,虽然部分位置毗连着晶间 2 次裂纹和材料内部微观孔洞等缺陷,但是蚀坑起裂位置仍然位于蚀坑应力最大处。

根据不同蚀坑类型的应力分布和断裂形貌,可发现所提出的旋转抛物-锥形蚀坑模型能够应用于不同类型的蚀坑强度分析当中,对起裂位置的判断及适应性上比半椭球形蚀坑模型更为优异。

4　结　论

点蚀是腐蚀环境中金属材料的常见缺陷形式之一。由点蚀引起的应力集中对裂纹的萌生与扩展具有重要影响。因此,准确的蚀坑模型不仅是研究裂纹产生的基础,对结构剩余强度的评价也具有十分重要的意义。为弥补典型蚀坑模型的不足,重新定义了张开角,提出了旋转抛物-锥形蚀坑模型。基于该模型研究了形状比和蚀坑平整度对蚀坑应力集中分布的影响,最后通过含蚀坑铝合金试样拉伸破坏试验验证了模型的正确性。主要的结论如下:

1) 无论是尖深型还是浅平型蚀坑,应力集中的极大值点均位于接近坑口或坑底的区域,该区域也是裂纹萌生点。

2) 与半椭球形蚀坑模型相比,旋转抛物-锥形蚀坑模型能够准确地表征坑底、坑肩及坑口区域的应力分布及相互之间应力集中的转换规律。

3) 拉伸荷载下,尖深蚀坑应力集中的特性大于浅平蚀坑,对材料损伤程度高。

参考文献(References)

[1] 刘治国,李旭东,陈川. 单个边缘微观损伤对铝合金材料蚀坑应力集中效应的影响研究[J]. 机械科学与技术,2021,40(3):456-462.
LIU Z G,LI X D,CHEN C. Influence of single marginal micro-cosmic damage on concentration effect of corrosion pit stress for aluminum alloy[J]. Mechanical Science and Technology for Aerospace Engineering,2021,40(3):456-462(in Chinese).

[2] LIU D Z,LI Y,XIE X D,et al. Effect of pre-corrosion pits on residual fatigue life for 42CrMo steel[J]. Materials (Basel,Switzerland),2019,12(13):2130-2137.

[3] ISHIHARA S,SAKA S,NAN Z Y,et al. Prediction of corrosion fatigue lives of aluminium alloy on the basis of corrosion pit growth law[J]. Fracture of Engineering Materials and Structures,2006,29(6):472-480.

[4] FARHAD F,ZHANG X,SMYTH-BOYLE D. Fatigue behaviour of corrosion pits in X65 steel pipelines[J]. Proceedings of the Institution of Mechanical Engineers,Part C:Journal of Mechanical Engineering Science,2019,233(5):1771-1782.

[5] XIANG L H,PAN J Y,CHEN S Y. Analysis on the stress corrosion crack inception based on pit shape and size of the FV520B tensile specimen[J]. Results in Physics,2018,9:463-470.

[6] HASHIM M,FARHAD F,SMYTH-BOYLE D,et al. Behavior of 316L stainless steel containing corrosion pits under cyclic loading[J]. Materials and Corrosion,2019,70(11):2009-2019.

[7] LI X W, LIANG J S, SHI T, et al. Tribological behaviors of vacuum hot-pressed ceramic composites with enhanced cyclic oxidation and corrosion resistance[J]. Ceramics International, 2020, 46(9): 12911-12920.

[8] EUBANKS R A. Stress concentration due to a hemispherical pit at a free surface[J]. Journal of Applied Mechanics, 1954, 21(3): 57-62.

[9] FUJITA T, SADAVASU T. Stress concentration due to hemispherical pit at a free surface of a thick plate under all around tension[J]. Bulletin of the Japan Society of Mechanical Engineers, 1978, 21(154): 561-565.

[10] CERIT M. Corrosion pit-induced stress concentration in spherical pressure vessel[J]. Thin-Walled Structures, 2019, 136: 106-112.

[11] CERIT M. Numerical investigation on torsional stress concentration factor at the semi elliptical corrosion pit[J]. Corrosion Science, 2013, 67: 225-232.

[12] CERIT M, GENEL K, EKSI S. Numerical investigation on stress concentration of corrosion pit[J]. Engineering Failure Analysis, 2009, 16(7): 2467-2472.

[13] HOU J, SONG L. Numerical investigation on stress concentration of tension steel bars with one or two corrosion pits[J]. Advances in Materials Science and Engineering, 2015, 2015: 413737.

[14] LI J Q. Modelling on the evolution of corrosion pit during stress corrosion[J]. Advances in Petroleum Exploration and Development, 2016, 11(2): 64-69.

[15] LIU L, HOU N, DING N, et al. Interacting effects of internal defects and corrosion pits on the stress concentration of hourglass shaped specimens[J]. Journal of Failure Analysis and Prevention, 2019, 19(4): 967-975.

[16] LI C Q, YANG S T. Stress intensity factors for high aspect ratio semi-elliptical internal surface cracks in pipes[J]. International Journal of Pressure Vessels and Piping, 2012, 96-97(1): 13-23.

[17] ATLURI S N, KATHIRESAN K. Influence of flaw shapes on stress intensity factors for pressure vessel surface flaws and nozzle corner cracks[J]. Journal of Pressure Vessel Technology, 1980, 102(3): 278-286.

[18] ACUÑA N, GONZÁLEZ-SÁNCHEZ J, KÚ-BASULTO G, et al. Analysis of the stress intensity factor around corrosion pits developed on structures subjected to mixed loading[J]. Scripta Materialia, 2006, 55(4): 363-366.

[19] TURNBULL A, WRIGHT L, CROCKER L. New insight into the pit-to-crack transition from finite element analysis of the stress and strain distribution around a corrosion pit[J]. Corrosion Science, 2010, 52(4): 1492-1498.

[20] 张川, 姚卫星. 危险蚀坑评判及疲劳寿命计算[J]. 机械强度, 2013, 35(2): 207-213.
ZHANG C, YAO W X. Judgment of critical pits and calculation of fatigue life[J]. Journal of Mechanical Strength, 2013, 35(2): 207-213(in Chinese).

[21] WANG W G, ZHOU A N, FU G Y, et al. Evaluation of stress intensity factor for cast iron pipes with sharp corrosion pits[J]. Engineering Failure Analysis, 2017, 81: 254-269.

[22] 国家市场监督管理总局, 国家标准化管理委员会. 金属材料 拉伸试验 第1部分: 室温试验方法: GB/T 228.1—2021[S]. 北京: 中国标准出版社, 2011: 11-14.
State Administration for Market Regulation, Standardization Administtration of the People's Republi of China. Metallic materials-Tensile testing-Part 1: Method of test at room temperature: GB/T 228.1—2021[S]. Beijing: Standards Press of China, 2011: 11-14(in Chinese).

Rotating parabolic-conical corrosion pit model establishment and its application

LIU Dejun[1], TIAN Gan[1,*], JIN Guofeng[1], YANG Zhengwei[1], REN Biyun[1], WEI Huali[2]

(1. The Missile Institute of Engineering, Rocket Force Engineering University, Xi'an 710025, China;
2. Xi'an Institute of Aerospace Composites, Xi'an 710025, China)

Abstract: As one of the common degradation forms of metal structure exposed to the corrosive medium, pitting may cause the local stress concentration and decrease the strength, reliability and safety of the structure of equipment. Thus, an exact pit model is useful for the stress distribution analysis of the metal structure exposed to the corrosive medium. In this instance, by surveying the pattern of typical corrosive pitting, the concept of pit open angle is redefined, and a novel model called rotating parabolic-conial model is developed. Both FEM simulations and tensile experiments are performed to validate the accuracy and efficiency of the proposed model. It is shown that the maximum of the stress concentration caused by the pitting is generally located around the bottom or mouth area of the pit. Compared with the semi-ellipsoid model, the former is more accurate and sensitive on the description of the stress distribution around the bottom shoulder and mouth of a pit.

Keywords: corrosion pit; rotating parabolic-conical model; semi-ellipsoid model; stress concentration; structural strength

Received: 2021-03-03; **Accepted**: 2021-06-04; **Published online**: 2021-06-24 17:20

URL: kns.cnki.net/kcms/detail/11.2625.V.20210624.1613.001.html

Foundation items: National Natural Science Foundation of China (52075541, 52272446); Natural Science Foundation of Shaanxi Province (2020JM-354)

* **Corresponding author**. E-mail: tiangan_2012@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0109

基于群体意志统一的无人机协同围捕策略

刘峰[1], 魏瑞轩[1,*], 周凯[1,2], 丁超[1]

(1. 空军工程大学 航空工程学院, 西安 710051; 2. 空军航空大学, 长春 130000)

摘 要: 针对无人机(UAV)协同围捕问题,提出一种基于群体意志统一的围捕策略。受人类在协作任务中的认知机理启发,引入“群体意志”定义无人机的协作认知,并构建双回路认知模型,借助图卷积网络对围捕无人机获取的局部态势进行融合认知,有效减轻无人机系统的计算负载。依靠变分推断原理和生成式自动编码器对围捕无人机进行群体意志趋同学习,依据 Apollonius 圆实现协同围捕,使无人机集群涌现出更加智能化的围捕效果。通过对比仿真验证了所提策略的有效性和智能性。

关 键 词: 协同围捕; 群体意志统一; 融合认知; 趋同学习; Apollonius 圆
中图分类号: TP391.9
文献标志码: A **文章编号**: 1001-5965(2022)11-2241-09

多无人机协同围捕是研究如何指导一群具有自主决策能力的无人机通过相互合作对单一个体或群体实现围捕。由于这类问题有着重要的理论价值和广泛的应用前景,如应用于军事领域,可大幅度提高军队无人机的自动化程度,降低军事任务成本,提高战斗效率等,因此受到很多学者的关注[1-6]。

就协同围捕策略设计而言,蔡云飞等[7]设计出 Cross-EKF 定位算法,通过交叉计算目标位置的后验估计协方差,将对动态围捕点的收敛扩展到对动态围捕面的收敛,提高了协同围捕的收敛速度及稳定性。黄天云等[8]通过对围捕行为的分解,设计了每个个体自组织运动控制器,提出了基于松散偏好规则(loose-preference rule, LP-Rule)的自组织协作围捕策略,并从理论上分析了其稳定性。李瑞珍等[9]为使机器人有效地快速围捕移动目标,提出一种基于动态围捕点的多机器人协同围捕策略,采用协商法为围捕机器人分配最佳围捕点,建立目标函数并优化,实现了围捕机器人在线路径规划。裴惠琴等[10]主要针对围捕环境受限及双方的速度比率受限问题,设计基于动态虚拟势点的“切换式”围捕策略,在提高围捕效率的同时解决了由“夹角最小”原则带来的围捕者“死锁”问题。

以上对于协同围捕策略的研究仍是建立在获取目标位置后,预先依据围捕成功的条件进行任务分配与协同环节的基础上,并基于特定控制模式完成围捕。同时,需要指出的是,围捕策略对于环境的自适应性、策略体现出的智能性有待提高。

近年来,也有部分学者通过观察总结和强化学习来探索集群协同问题[11-12]。Muro 等[11]通过观察、计算模拟北灰狼群围捕猎物行为的规则和主要特征,从狼群和猎物的 4 种简单行为中提出一个计算多智能体模型,总结出狼群的围捕策略。Lowe 等[12]从多智能体深度强化学习的角度,构建智能体合作与竞争的场景,提出多智能体深度确定性策略梯度(multi-agent deep deterministic policy gradient, MADDPG)算法,采用集中训练和

收稿日期: 2021-03-05; **录用日期**: 2021-04-09; **网络出版时间**: 2021-04-20 09:16
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210419.1608.005.html
基金项目: 科技部 2030“新一代人工智能”重大项目 (2018AAA0102403)
* **通信作者**. E-mail: ruixuanWei123@163.com

引用格式: *刘峰, 魏瑞轩, 周凯, 等. 基于群体意志统一的无人机协同围捕策略[J]. 北京航空航天大学学报, 2022, 48(11): 2241-2249. LIU F, WEI R X, ZHOU K, et al. Multi-UAV round up strategy based on unity of group will[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2241-2249(in Chinese).*

分散执行的框架,实现了群体之间的协同策略,其鲁棒性远优于传统的强化学习算法,但是,随着智能体数量的增加,系统的稳定性降低,收敛速度变慢。

基于以上研究的不足,本文探索出一种智能化的多无人机协同围捕策略。在该策略中,借助图卷积网络模型对围捕无人机进行信息融合认知设计,实现参数共享,减少模型复杂度,有效减轻计算负载;更具魅力之处在于借鉴人类在协作任务中达成认知统一的规律[13],设计群体意志趋同学习的生成式模型使围捕无人机通过学习更倾向于产生协作策略,使无人机集群涌现出更加智能化的围捕行为。仿真结果验证了基于群体意志统一的围捕策略能够有效完成多无人机协同围捕任务。

1 问题描述

假设在一个无限大且无障碍的二维空间中,有 $N(N\geqslant 3)$ 个围捕无人机对一个目标无人机进行围捕,如图1所示。

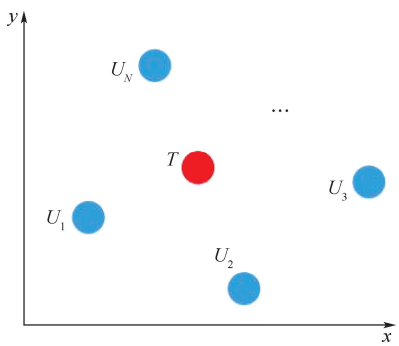


图1 N 个无人机围捕单个目标

Fig. 1 N drones rounded up a single target

其中,$U=\{U_1,U_2,\cdots,U_N\}$ 为 N 架围捕无人机的集合,$V_U=\{V_{U_1},V_{U_2},\cdots,V_{U_N}\}$ 为 N 架围捕无人机速度的集合。T 为单架目标无人机,V_T 为目标无人机的速度。为更好的体现所设计围捕策略的智能性,$\forall V_{U_i}\in V_U$,满足 $V_{U_i}<V_T$,其中,$i=1,2,\cdots,N$。所有无人机均视为质点,且不具备攻击能力,围捕无人机仅通过协作形成封闭包围圈完成围捕。

围捕开始后,围捕无人机通过协作对目标无人机实施围捕,目标无人机则按照一定的逃逸策略进行逃离。当围捕无人机在目标无人机周围形成封闭的 Apollonius 圆域时,视为围捕成功。考虑到围捕无人机固定物理防守半径 R_{U_i},当目标无人机 T 与任意一架围捕无人机 U_i 距离小于相应的 R_{U_i} 时,也视为围捕成功。

2 模型构建

2.1 无人机运动模型

假设无人机当前时刻位置向量为 $[x,y]^{\mathrm{T}}$,将无人机的运动模型表述如下:

$$\begin{cases}\dot{x}=v\cos\theta\\ \dot{y}=v\sin\theta\end{cases}\tag{1}$$

式中:θ 为航向角;v 为速度;$\dot{x}$ 为 v 的水平分量;$\dot{y}$ 为 v 的竖直分量。

可以看出,忽略姿态变化后,无人机的横向、纵向速度都可以通过 v 和 θ 进行相应配置。为拓展无人机机动能力,航向角 θ 可变,但存在角速度上限,变化示意图如图2所示。

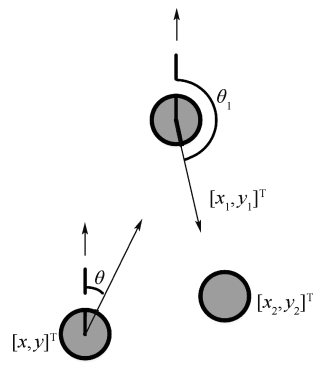


图2 无人机航向变化示意图

Fig. 2 Diagram of course change of UAV

2.2 基于 Apollonius 圆的围捕模型

Apollonius 圆的定义为平面内到2个定点的距离之比为常数 $k(k\neq 1)$ 的点的集合。针对围捕无人机与目标无人机而言,若二者速度之比为 k,则双方同时到达 Apollonius 圆上,从而实现单一方向上的拦截。

围捕无人机借助与目标无人机到达 Apollonius 圆[14]上点的时间相等这一特点进行决策,其决策原则如图3所示。

图中绿色、蓝色、红色虚线圆分别为围捕无人机 U_1、U_2、U_3 与目标无人机 T 之间构建的 Apollonius 圆,点 O_1、O_2、O_3 分别为3个圆的圆心,点 C_{11}、

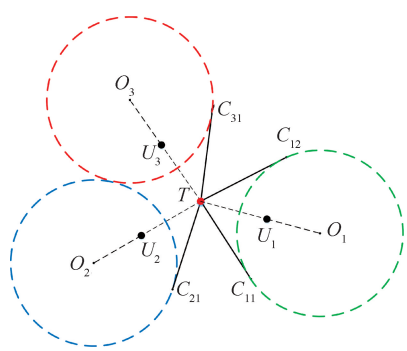


图3 围捕无人机决策静态示意图

Fig. 3 Static schematic of a roundup UAV decision making

C_{12}、C_{21}、C_{31}分别为围捕无人机当前位置与3个圆的切点。由于围捕无人机速度小于目标无人机速度，故目标无人机未被围捕前，可利用速度优势从 C_{11} 以左、C_{21} 以右或 C_{12} 以上、C_{31} 以下区域，进行逃离。若形成封闭 Apollonius 圆域，则围捕成功。

3　基于群体意志统一的协同围捕策略

3.1　双回路认知模型

当人类参与协作任务时，个体除获取自身信息外，通过观察、交流获得环境及其他个体信息，基于融合后的信息，一方面进行自我认知，形成指导自身的策略；另一方面能够进行协作认知，形成对于群体目标的认知，且群体目标会反过来影响个体的自我认知过程。本文将围捕无人机根据其获得信息形成关于协作任务的认知定义为群体意志。对于每个个体而言，群体意志是个体根据自身能获得的所有信息形成的对协作任务的认识。根据这种现象，构建了人类在协作任务中的双回路认知模型如图 4 所示。

在协作任务中，人类的认知活动分为两部分。其中，第 1 部分如图 4 红色箭头表示的实践学习认知回路，个体根据自我认知形成决策，并通过观察环境反馈的信息不断学习优化，以保证个体策略的最优性；第 2 部分如图 4 蓝色箭头表示的回路，个体将自身得到的群体意志与群体中其他个体获得的群体意志进行对比，存在差异时，试图与其他个体实现统一，此过程反过来会改变信息融合时的侧重点和对信息的认知理解，从而改变自身对能够获得的信息的融合过程，将这一过程称为群体意志趋同认知回路。

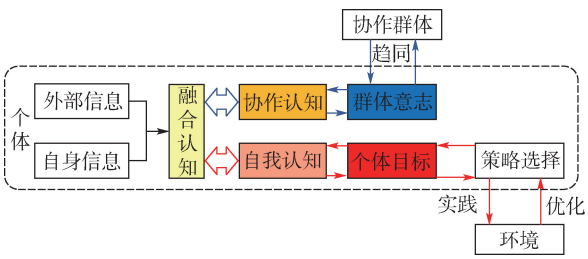


图 4　双回路认知模型示意图

Fig. 4　Schematic diagram of double-loop cognitive model

3.2　无人机融合认知方法设计

在协同围捕过程中，围捕无人机之间存在不完全联系，为有效形成个体对于整体的认识，需要构造一个能够实现信息交联，并进行融合认知的模型。

3.2.1　基于图论的信息交联关系

采用图 $\boldsymbol{G}=(V,E,\boldsymbol{H})$ 描述 N 架围捕无人机的信息交联关系，其中，$V=\{1,2,\cdots,N\}$ 为各围捕无人机节点集合，E 为边集。当围捕无人机 j 能获得围捕无人机 i 的信息，则有 $(i,j)\in E$，且将围捕无人机 i 称为围捕无人机 j 的邻居(Neighbor)。$\boldsymbol{H}$ 为 $N\times N$ 的邻接矩阵，建立式(2)模型描述围捕无人机之间的连通性。

$$h_{ji}=\begin{cases}1 & (i,j)\in E\\ 0 & \text{其他}\end{cases} \tag{2}$$

假设邻居之间的通信是对称的，即 $\forall i,j\in V$，有 $h_{ij}=h_{ji}$；定义矩阵 $\boldsymbol{D}=[d_{ij}]_{N\times N}$ 为图 $\boldsymbol{G}$ 的度矩阵，$\boldsymbol{d}_{ii}=\sum_{j=0}^{N}h_{ij}$ 为对角矩阵。

3.2.2　基于图卷积网络的融合认知模型

借助空间域和谱域融合的图卷积网络[15](graph convolutional network，GCN)在拓扑结构数据上强大的特征提取能力和分类能力，设计了基于 GCN 的融合认知模型，如图 5 所示。

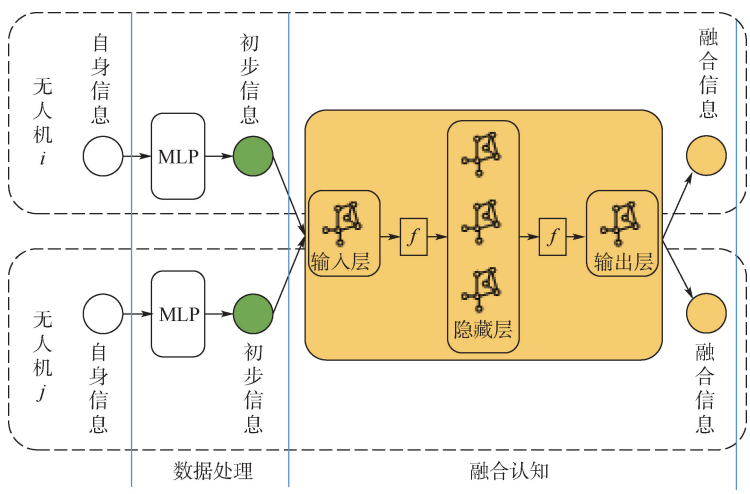


图 5　融合认知模型结构示意图

Fig. 5　Schematic diagram of fusion cognitive model

以围捕无人机 i 为例，将状态信息 $\boldsymbol{s}_i=[x_i,y_i]^{\mathrm{T}}$ 和动作信息 $a_i=\theta_i$，记为 $x_i=\{\boldsymbol{s}_i,a_i\}$，利用一个隐藏层为 2 层，每层 64 个节点，激活函数为 tanh 函数的多层感知器（multi-layer perceptron，MLP）对 x_i 进行池归一化[16]等预处理，记处理后的信息为 $x_i'=\{\boldsymbol{s}_i',a_i'\}$。将 x_i'输入图卷积网络，利用 GCN 的特征提取能力获得融合信息。融合认知过程中信息的传递过程可以表示为

$$\boldsymbol{X}^{l+1}=f(\boldsymbol{W}^l(\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{H}}\hat{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{X}^l)) \tag{3}$$

式中：$\hat{\boldsymbol{H}}=\boldsymbol{H}+\boldsymbol{I}$，$\boldsymbol{I}$ 为单位矩阵，由于邻接矩阵 $\boldsymbol{H}$ 的对角线上为 0，围捕无人机在提取特征时不能将自身的信息计算在内，因此，加上一个单位矩阵，使围捕无人机对态势实现更加全面的认识；$\hat{\boldsymbol{D}}$ 为 $\hat{\boldsymbol{H}}$ 的度矩阵，即 $\hat{\boldsymbol{d}}_{ii}=\sum_{j=0}^{N}\hat{h}_{ij}$；$\boldsymbol{W}^l$ 为第 l 层的网络权值；$\boldsymbol{X}^l$、$\boldsymbol{X}^{l+1}$分别为第 l 层的输入和输出，对于第 1 层而言，$\boldsymbol{X}^1$ 即为各围捕无人机的初始信息集合$\{x_1',x_2',\cdots,x_N'\}$，最后一层的 $\boldsymbol{X}^{l+1}$即融合后信息，增加 $\hat{\boldsymbol{D}}^{-\frac{1}{2}}\hat{\boldsymbol{H}}\hat{\boldsymbol{D}}^{-\frac{1}{2}}$项可以对不同连接数的邻居节点对当前节点的影响权重进行平衡。式(3)表示 GCN 中第 l 层向 $l+1$ 层的传递计算式。

相较于已有结果，融合认知模型的优势在于：①鲁棒性强。由于将各围捕无人机之间存在差异的连接关系作为 GCN 网络的输入，因此，不需要针对每架围捕无人机设计特定的信息提取模型。②模型复杂度低。使用单层卷积即能够实现所有相邻围捕无人机信息的融合，若增加一层则可实现与当前围捕无人机有二跳连接关系的围捕无人机信息融合。

同时，图卷积网络的层数也不宜过多，连接关系较远的围捕无人机对当前个体认知的影响较小，层数增加会出现过度平滑的问题[17]，失去了利用局部信息实现群体意志统一的意义。

3.3 群体意志的趋同学习原理

理想状态下，当围捕无人机达成群体意志统一时，存在由信息 x_i 到统一群体意志 C 之间的条件概率分布 $P(C\mid x_i)$，且每架围捕无人机根据自身的信息得到的统一群体意志 C 的分布是高度相似的（见图 6），即分布 $P(C\mid x_i)$ 与 $P(C\mid x_j)$ 高度相似。达成统一的群体意志的后验分布 $P(C\mid x_i)$ 的计算方法为

$$P(C\mid x_i)=\frac{P(x_i,C)}{P(x_i)} \tag{4}$$

式中：$P(x_i,C)$ 为 x_i 与 C 的联合分布；$P(x_i)$ 为无人机 i 获得的信息的分布。对于围捕无人机协作问题而言，式(4)很难执行，因为围捕无人机获得

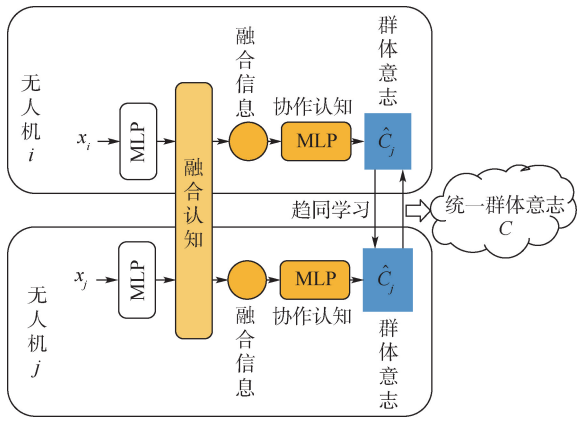


图 6 统一群体意志统一构造示意图

Fig. 6 Schematic diagram of unified structure of group will

的信息相对复杂，联合分布 $P(x_i,C)$ 难以获得，并且，达成统一后的群体意志 C 的分布 $P(C)$ 也是后续求解得到的。

因此，借鉴变分推断[18]的原理，构建一种生成式的模型，使得围捕无人机能够在任务过程中通过学习自主地逼近 $P(C\mid x_i)$。将图 6 中构建的由信息 x_i 到群体意志 $\hat{C}_i$ 的映射关系抽象地描述为条件概率分布 $Q(\hat{C}_i\mid x_i;W)$，W 为图 6 中 x_i 与 $\hat{C}_i$ 之间所有网络参数的集合。趋同学习的目标可以表示为

$$W^*=\arg\min_W D_{\mathrm{KL}}(Q(\hat{C}_i\mid x_i;W),P(C\mid x_i)) \tag{5}$$

式中：$D_{\mathrm{KL}}(\cdot)$ 为 2 个分布之间的 KL 散度（也称为交叉熵或相对熵），表示 2 个分布之间的距离。根据变分推断理论[18]，式(5)等价于

$$\begin{aligned}W^*&=\arg\max_W E_{\hat{C}_i\sim Q}[\ln(P(x_i\mid\hat{C}_i))]-\\&E_{\hat{C}_i\sim Q}\left[\ln\left(\frac{P(C)}{Q(\hat{C}_i\mid x_i;W)}\right)\right]=\\&\arg\max_W E_{\hat{C}_i\sim Q}[\ln(P(x_i\mid\hat{C}_i))]-\\&D_{\mathrm{KL}}(Q(\hat{C}_i\mid x_i;W),P(C))\end{aligned} \tag{6}$$

式中：$P(x_i\mid\hat{C}_i)$ 为根据群体意志 $\hat{C}_i$ 反向推测个体获得的信息 x_i 的概率分布；$P(C)$ 为达成统一后的群体意志 C 的先验分布，是隐式存在的，虽然不能直接获得，但是其分布与 $P(x_i\mid\hat{C}_i)$、$P(x_i)$、$Q(\hat{C}_i\mid x_i;W)$都无关；$E_{\hat{C}_i\sim Q}$为对应项的期望分布。采用神经网络对式(6)中 $P(x_i\mid\hat{C}_i)$项参数化，可以使用梯度下降的方法进行优化，逐渐实现围捕无人机群体意志趋向统一。借鉴文献[19]构造一个生成式自动编码器模型，其结构如图 7 所示。

图 7 中，黄色部分为编码器，根据信息 x_i 构造群体意志 $\hat{C}_i$，即图 6 中的群体意志构造模块 $Q(\hat{C}_i\mid x_i;W)$。绿色部分为解码器，根据编码器生

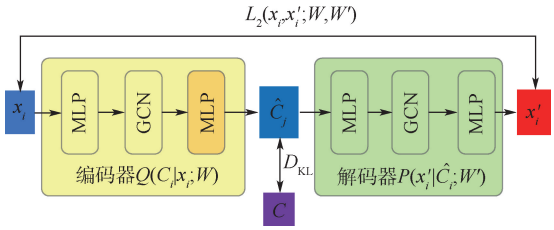


图 7　用于群体意志趋同学习的生成式自动编码器结构

Fig. 7　Generative autoencoder for group will convergence learning

成的 $\hat{C}_i$ 试图恢复信息 x_i，其输出为 x_i 的近似值 x_i'，将其写成条件概率分布形式即为 $P(x_i' \mid \hat{C}_i; W')$，其中 W' 表示解码器部分神经网络的参数。图中 $\hat{C}_i$ 与 C 之间的 KL 散度对应式(6)中的第 2 项 $D_{KL}(Q(\hat{C}_i \mid x_i; W), P(C))$，能够使围捕无人机学习到的群体意志分布尽可能接近统一的群体意志的先验分布 $P(C)$。$L_2(x_i, x_i'; W, W')$ 为原始信息 x_i 与重构信息 x_i'之间的重构误差，是式(6)中第 1 项 $E_{\hat{C}_i \sim Q}[\ln(P(x_i \mid \hat{C}_i))]$ 的最小二乘距离表达形式。通过两方面的误差传播，上述生成式自动编码器能够不断优化神经网络参数，从而实现群体意志的趋同逼近。需要指出的是，$D_{KL}(Q(\hat{C}_i \mid x_i; W), P(C))$ 计算的是 2 个分布之间的距离，生成式自动编码器不像传统的自动编码器那样输出实数向量，而是输出一个分布模型。

为便于结构计算，借助深度神经网络极强的非线性逼近能力，对编码器中负责协作认知的 MLP 模块(图 6、图 7 中橙色模块)进行如图 8 所示的调整(隐藏层为 1 层，每层节点数为 128，激活函数为 tanh 函数)。为能够获得更大收益，构建从状态到动作策略的映射 $S \rightarrow \boldsymbol{\pi}: \boldsymbol{\pi}_{\boldsymbol{\omega}}(s)$，其中 $\boldsymbol{\omega}$ 是能够通过学习进行调整的参数矩阵。最终，协作认知 MLP 的输入为融合认知环节 GCN 网络的输出，输出为 $\hat{C}_i^{\mu}$、$\hat{C}_i^{\sigma}$ 分别表示 $\hat{C}_i$ 的均值和标准差，且 $\hat{C}_i^{\mu}$、$\hat{C}_i^{\sigma}$ 均服从与 MLP 结构相关的截断高斯分布。

综上，群体意志趋同学习的训练过程表示为

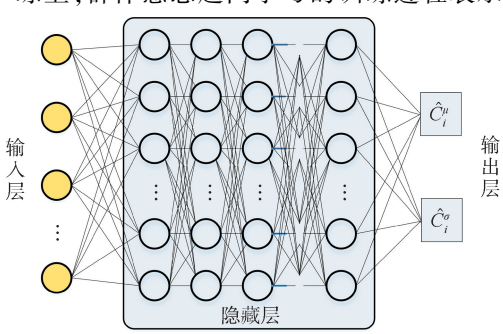


图 8　协作认知模块的神经网络结构

Fig. 8　Neural network structure of cooperative cognition module

$$W^* = \arg\min[L_2(x_i, x_i'; W, W') + D_{KL}(Q(\hat{C}_i \mid x_i; W), P(C))] \tag{7}$$

4　仿真及分析

4.1　训练集参数设定

为验证所设计策略的有效性，取围捕无人机数目 $N=4$ 进行仿真实验，初始位置随机生成，大致散布在目标无人机周围，最远不超过 150 km，具体参数如表 1 所示。

表 1　无人机参数设定

Table 1　UAV parameter setting

无人机类型	速度/(km·s^{-1})	固定物理防守半径/km
围捕无人机 U_1	0.15	3
围捕无人机 U_2	0.12	2
围捕无人机 U_3	0.12	2
围捕无人机 U_4	0.12	2
目标无人机 T	0.18	

设定 t 时刻奖励函数如下：

$$r_t = \underbrace{10\sum_{i=1}^{N}[D_i^{ID}(t-1) - D_i^{ID}(t)]}_{①} + \underbrace{10\text{flag}}_{②} \tag{8}$$

式中：$D_i^{ID}(t)$ 表示 t 时刻第 i 个围捕无人机与目标无人机的距离，为便于处理数据及突出奖励，对距离和乘以 10。训练开始时，记初始时刻奖励 $r_0 = 0$。flag 定义为

$$\text{flag} = \begin{cases} 1 & \text{围捕成功} \\ 0 & \text{其他} \end{cases} \tag{9}$$

式(8)第①部分促使围捕无人机接近目标无人机进行围捕，第②部分为围捕成功后反馈奖励值 10。当 $\forall i \in \{1,2,\cdots,N\}, D_i^{ID}(t-1) - D_i^{ID}(t) < 0$ 成立时，判定围捕任务失败。

仿真流程如图 9 所示。其中，M 为所设置的

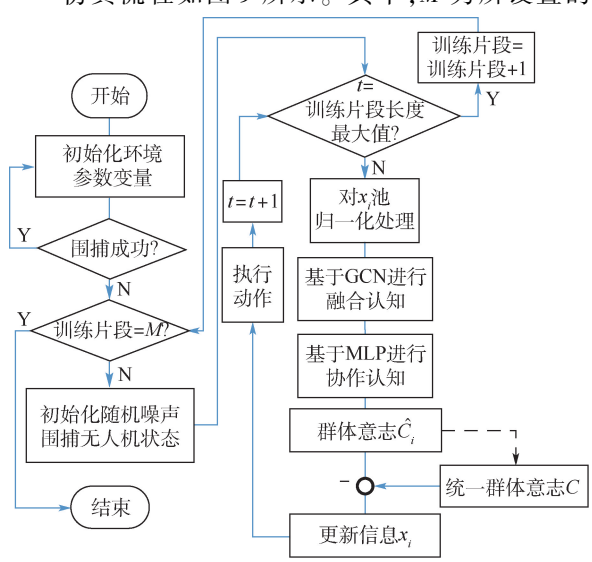


图 9　仿真流程

Fig. 9　Simulation flow chart

训练片段数量上限,t 为常数。

4.2 仿真结果

通过对比实验分析群体意志统一对无人机协作围捕学习的影响,实验组采用基于群体意志统一的策略训练围捕无人机,对照组采用与实验组相同的网络结构,但采用传统的分布式优化策略,只进行基于实践的学习,而不进行趋同学习。

记录下实验后期某片段围捕无人机获得的实时奖励,如图 10 所示。

从图 10 中可以看出,实验组与对照组的学习效果大约在第 6 500 回合出现差别。进行趋同学习的实验组获得的初始奖励较低,但围捕无人机系统经过基于群体意志统一的学习,能够在较短回合内提升奖励值,并经过大约 45 000 回合的训练后,通过协作实现了对目标无人机的围捕。而采用传统的分布式优化策略,只进行实践学习的对照组能够更快的获得较高的初始奖励,随后奖励值的增长速度较慢,远小于实验组。同时,由于对照组个体更倾向于接近目标无人机,扩大自身收益,在局部回合较实验组能够获得更高的实时奖励,但从任务整体的收益来看,最终并未完成围捕任务。

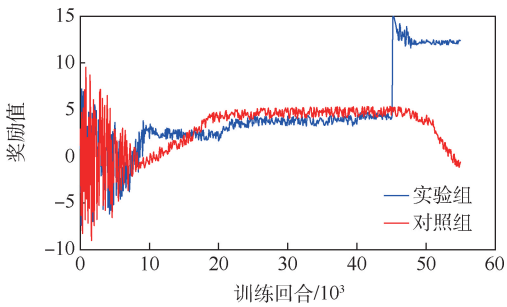
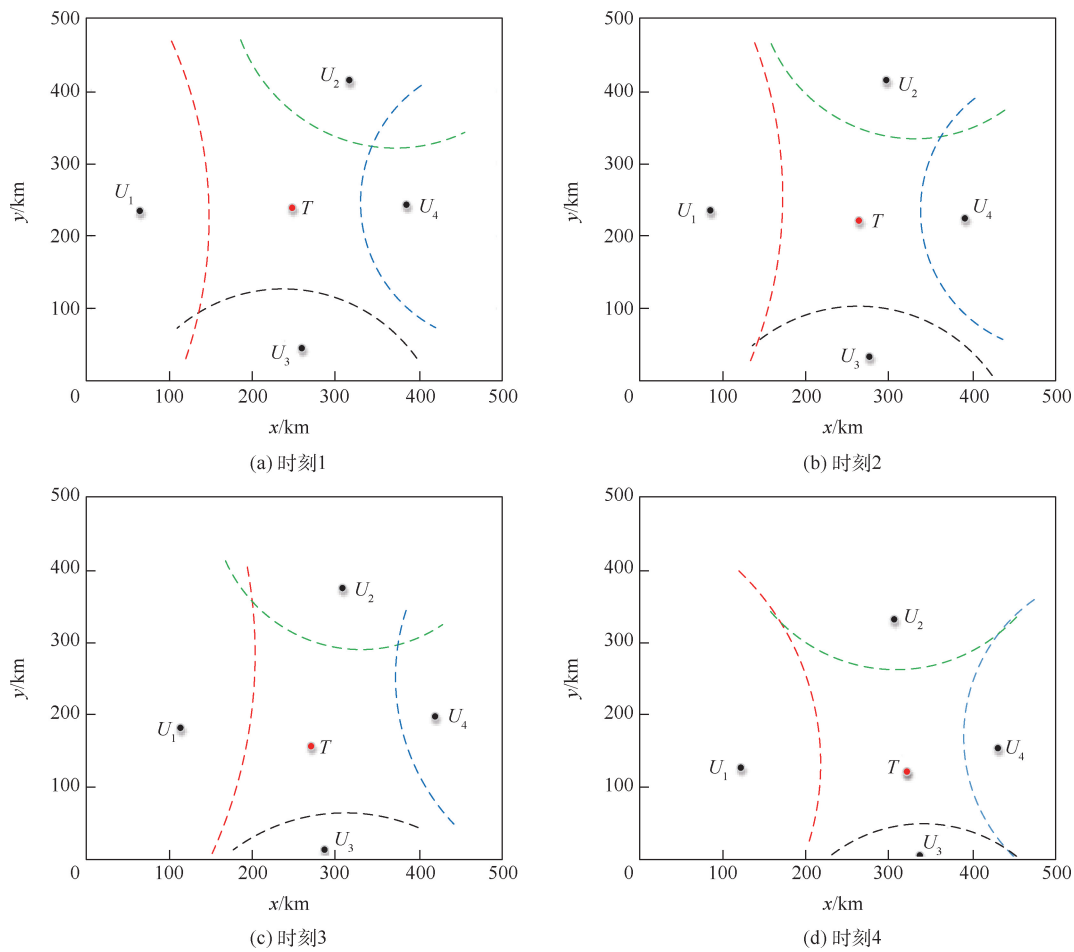


图 10 训练实时奖励

Fig. 10 Training real-time rewards

为了更直观地说明问题,将训练 48 000 回合的围捕无人机系统接入训练环境,获取实验组和对照组的拦截过程截图。

如图 11 和图 12 所示,红色实心点为目标无人机的当前位置,红、绿、蓝、黑 4 条虚线弧分别为围捕无人机 U_1、U_2、U_3、U_4 与目标无人机形成的 Apollonius圆弧。从实验组和对照组围捕过程中可以看出,采用基于统一群体意志的协作围捕策略,围捕无人机更倾向于互相配合,尽快形成封闭



(a) 时刻1 (b) 时刻2

(c) 时刻3 (d) 时刻4

图 11 实验组围捕过程

Fig. 11 Experimental group round up process

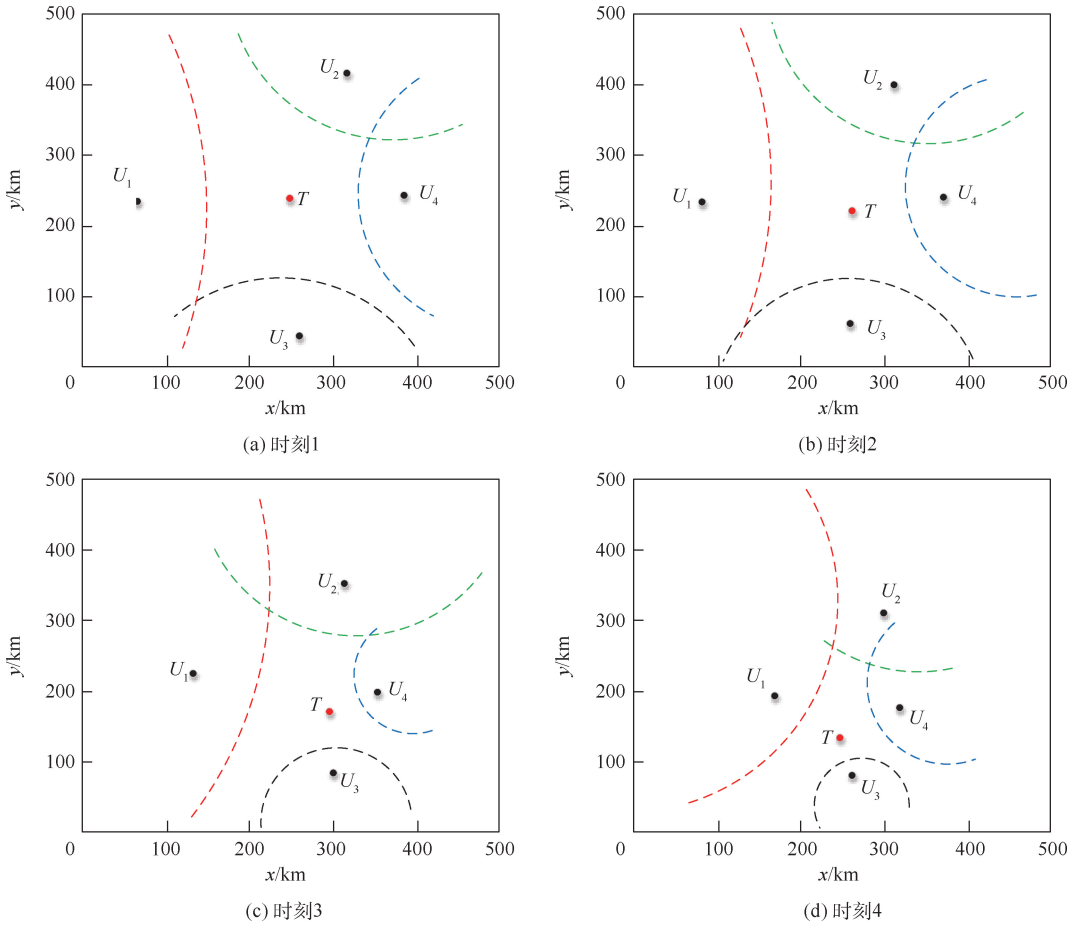


图 12　对照组围捕过程

Fig. 12　Control group round up process

Apollonius 圆域完成围捕，实验组相比于对照组而言，尽管获得的奖励函数式(8)第①部分值较低，但是获得奖励函数整体值较高。对照组中围捕无人机更倾向于扩大自身收益，利用其他个体的 Apollonius 圆，尽快接近目标无人机，以固定物理防守半径完成围捕，却被目标无人机利用速度优势获得更好的局部态势。正如图 11 和图 12 所示，对照组与实验组拥有相同的初始条件，如图 12(a)和图 11(a)所示；在 2 组相应的围捕策略指导下，经过一段时间的运动后，对照组较实验组有更好的围捕态势，如图 12(b)所示；在目标无人机向下运动压缩围捕无人机 U_3 的 Apollonius 圆时，对照组围捕无人机 U_1 和 U_4 并没有像实验组及时保护 U_3 的弱侧，却径直接近目标无人机，反被目标无人机利用速度冲击 U_1 和 U_3 空档，没有及时完成合围，如图 12(c)和图 12(d)所示。

5　结　论

本文所设计的基于群体意志统一的多无人机协同围捕策略能够有效解决不同数量的无人机集群在固定环境下的围捕问题。设定围捕无人机集群速度小于目标无人机速度，同时摆脱复杂、精确的模型控制，提出群体意志趋同学习原理，在保证全局有解的情况下，使整个围捕无人机系统涌现出的智能化得到充分发挥。较采用传统的分布式优化策略，仅进行实践学习的围捕无人机能够更快的完成围捕，提高围捕成功率。

同时，本文只考虑了无约束的二维空间环境，仅能将所设计围捕策略简单适用于在指定高度层下，定速无人机协同围捕单架定速目标无人机的工程应用中，具有一定的局限性。下一步将针对有障碍物约束的二维场景及无障碍约束的三维空间环境，逐步进行深入扩展研究，并结合三维空间下的多无人机协同围捕工程应用，以增强该围捕策略在复杂环境下的鲁棒性。

参考文献（References）

[1] 段海滨，申燕凯，赵彦杰，等. 2019 年无人机热点回眸[J]. 科技导报，2020，38(1)：170-187.
DUAN H B，SHEN Y K，ZHAO Y J，et al. Review of technological hotspots of unmanned aerial vehicle in 2019[J]. Science Technology Review，2020，38(1)：170-187(in Chinese).

[2] 段海滨，张岱峰，范彦铭，等. 从狼群智能到无人机集群协

同决策[J]. 中国科学:信息科学,2019,49(1):112-118.
DUAN H B,ZHANG D F,FAN Y M,et al. From wolf pack intelligence to UAV swarm cooperative decisionmaking[J]. Scientia Sinica: Informationis, 2019, 49(1): 112-118 (in Chinese).

[3] ALEXOPOULOS A, SCHMIDT T, BADREDDIN E. A pursuit-evasion game between unmanned aerial vehicles[C]// 2014 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO). Piscataway: IEEE Press, 2014:74-81.

[4] 吴子沉,胡斌. 基于态势认知的无人机集群围捕方法[J]. 北京航空航天大学学报,2021,47(2):424-430.
WU Z C,HU B. Swarm rounding up method of UAV based on situation cognition[J]. Journal of Beijing University of Aeronautics and Astronautics,2021,47(2):424-430(in Chinese).

[5] YASUYUKI S,HIROFUMI O,TADASHI M,et al. Cooperative capture by multi-agent using reinforcement learning application for security patrol systems[C]//2015 10th Asian Control Conference (ASCC). Piscataway:IEEE Press,2015:1-6.

[6] 陈志鹏,李健. 基于动物集群行为的无人机群目标围捕策略[J]. 现代计算机,2018(2):11-14.
CHEN Z P,LI J. Cluster UAVs hunting strategy based on biological collective behaviors[J]. Modern Computer,2018(2):11-14(in Chinese).

[7] 蔡云飞,唐振民,张浩峰. 基于Cross-EKF定位的多机器人协作围捕策略研究[J]. 控制与决策,2010,25(9):1313-1317.
CAI Y F,TANG Z M,ZHANG H F. Multi-robots cooperative hunting strategy based on Cross-EKF localization[J]. Control and Decision,2010,25(9):1313-1317(in Chinese).

[8] 黄天云,陈雪波,徐望宝,等. 基于松散偏好规则的群体机器人系统自组织协作围捕[J]. 自动化学报,2013,39(1):57-68.
HUANG T Y,CHEN X B,XU W B,et al. A self-organizing cooperative hunting by swarm robotic systems based on loose-preference rule[J]. Acta Automatica Sinica,2013,39(1):57-68(in Chinese).

[9] 李瑞珍,杨惠珍,萧丛杉. 基于动态围捕点的多机器人协同策略[J]. 控制工程,2019,26(3):510-514.
LI R Z,YANG H Z,XIAO C S. Cooperative hunting strategy for multi-mobile robot systems based on dynamic hunting points[J]. Control Engineering of China,2019,26(3):510-514(in Chinese).

[10] 裴惠琴,陈世明,孙红伟. 动态环境下可扩展移动机器人群体的围捕控制[J]. 信息与控制,2009,38(4):437-443.
PEI H Q,CHEN S M,SUN H W. Pursuit control of scalable swarm system of mobile robots in dynamic environment[J]. Information and Control,2009,38(4):437-443(in Chinese).

[11] MURO C,ESCOBEDO R,SPECTOR L,et al. Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations[J]. Behavioural Processes,2011,88(3):192-197.

[12] LOWE R,WU Y,TAMAR A,et al. Multi-agent actor-critic for mixed cooperative-competitive environments[EB/OL]. (2017-06-07)[2021-02-14]. https://arxiv.org/abs/1706.02275v1.

[13] MAO H Y,LIU W L,HAO J Y,et al. Neighborhood cognition consistent multi-agent reinforcement learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence,2020,34(5):7219-7226.

[14] LIANG L,DENG F,PENG Z H,et al. A differential game for cooperative target defense[J]. Automatica,2019,102:58-71.

[15] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[C]// International Conference on Learning Representations (ICLR 2017),2017:1-14.

[16] IOFFE S,SZEGEDY C. Batch normalization:Accelerating deep network training by reducing internal covariate shift[C]// Proceedings of the 32nd International Conference on International Conference on Machine Learning,2015,37:448-456.

[17] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems,2021,32(1):4-24.

[18] BLEI D M,KUCUKELBIR A,MCAULIFFE J D. Variational inference:A review for statisticians[J]. Journal of the American Statistical Association,2017,112(518):859-877.

[19] NAIR A, PONG V, DALAL M, et al. Visual reinforcement learning with imagined goals[EB/OL]. (2018-07-12)[2020-12-04]. https://arxiv.org/abs/1807.04742.

Multi-UAV round up strategy based on unity of group will

LIU Feng[1], WEI Ruixuan[1,*], ZHOU Kai[1,2], DING Chao[1]

(1. College of Aeronautical Engineering, Air Force Engineering University, Xi'an 710051, China;
2. Air Force Aviation University, Changchun 130000, China)

Abstract: To solve the problem of unmanned aerial vehicle (UAV) coordinated rounding up, a strategy was proposed based on the unity of group will. Inspired by the cognitive mechanism of human beings in collaborative tasks, this paper introduces "group will" to define the collaborative cognition of UAVs, builds a double-loop cognitive model, and integrates the cognition of the local situation acquired by the rounded up UAVs with the help of the graph convolutional network, so as to effectively reduce the computing load of UAVs. On the basis of the variational inference principle and generative autoencoder, the group will convergence learning is carried out on the UAV, and the coordinated rounding up is realized on the basis of the Apollonius circle so that the UAV cluster emerges a more intelligent rounding up effect. The simulation results show the effectiveness and intelligence of the designed rounding up strategy.

Keywords: cooperative rounding up; unity of group will; fusion cognition; convergent learning; Apollonius round

Received: 2021-03-05; **Accepted**: 2021-04-09; **Published online**: 2021-04-20 09:16

URL: kns.cnki.net/kcms/detail/11.2625.V.20210419.1608.005.html

Foundation item: Science and Technology Innovation 2030-Key Project of "New Generation Artificial Intelligence" (2018AAA0102403)

* **Corresponding author**. E-mail: ruixuanWei123@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0077

基于无监督学习和粒子滤波的非视距信号检测

侯宁宁[1,2], 李灯熬[2,3,*], 赵菊敏[1,2]

(1. 太原理工大学 信息与计算机学院, 太原 030024; 2. 山西省空间信息网络工程技术研究中心, 太原 030024;
3. 太原理工大学 大数据学院, 太原 030024)

摘 要: 全球卫星导航系统(GNSS)是目前应用最广泛的定位技术,研究城市峡谷中的定位问题时,由于高楼大厦的阻塞,仍存在非视距传播导致的性能退化问题。为此,提出了无监督学习粒子滤波(UL-PF)算法。在卫星信号分类阶段,使用核 k-means 聚类的无监督学习分类方法,在定位阶段,使用通过聚类算法优化的粒子滤波方法。所提算法考虑了采样粒子在状态空间分布中的内在相似性,探索在每个聚类中选择一个粒子作为重要粒子,利用时间序列相关技术提高重采样粒子集的多样性。实验表明:在城市场景中,所提算法的平均定位精度从传统算法的15 m提高到约 5 m,收敛时间从 500 s 缩短到 200 s 左右。

关 键 词: 全球卫星导航系统; 非视距分类; 核 k-means; 粒子滤波; 无监督学习
中图分类号: TN967.1
文献标志码: A **文章编号**: 1001-5965(2022)11-2250-09

非视距和多径信号是影响全球卫星导航系统(global navigation satellite system, GNSS)定位精度困难大的误差源之一。反射或非视距信号与直接视距信号结合时,会产生不必要的多径效应,仍然是误差的主要来源[1-2]。多径不能通过差分技术去除,从而限制了多径易发区域的定位精度[3-4]。检测和抑制非视距(non line of sight, NLOS)方法主要分为 3 类:基于天线方面的双极化天线[5],基于接收机方面的先进接收机设计和传感器集成[6],基于导航处理器方面的三维建筑模型[7]。由于现实环境中非视距信号的复杂性,他们都需要外部硬件或最新的地图,这些昂贵的或不实际的大众市场应用程序,仅使用单一的检测方法无法达到满意的效果。

近年来,一些学者提出利用机器学习方法区分非视距信号。文献[8]提出使用支持向量机识别视距/非视距(line of sight/non line of sight, LOS/NLOS)场景。文献[9]提出了一种新的基于有监督机器学习算法的 LOS/NLOS 分类方法。上述研究都是基于监督学习的 GNSS 信号分类方法,需事先对学习样本进行标记,因此,需要借助额外的硬件或软件。无监督学习是一种对不含标记的数据建立模型的机器学习方法。文献[10]提出一种无监督机器学习的数据聚类方法,将测量数据分为 LOS 信号和 NLOS 信号。文献[11]提出采用“高斯混合模型的期望最大化”的无监督机器学习方法,对 LOS 和 NLOS 分量进行分类,但 2 种方法均应用于室内无线信号,借鉴该思路可以将无监督学习对信号的分类应用于 GNSS 信号,同时为了解决传统 k-means 算法对簇中心敏感的问题,使用了核 k-means 聚类算法,旨在提高 GNSS 设备在复杂环境下的定位精度,可以避免额外软硬件的使用,提高运算效率。对于定位优化部分,本文提出无监督学习-粒子滤波(unsuper-

收稿日期: 2021-02-08; **录用日期**: 2021-04-23; **网络出版时间**: 2021-05-11 11:39
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210511.1005.001.html
基金项目: 国家自然科学基金(61772358); 国家重点研发计划(2018YFB2200900)
* **通信作者**. E-mail: lidengao@tyut.edu.cn

引用格式: 侯宁宁, 李灯熬, 赵菊敏. 基于无监督学习和粒子滤波的非视距信号检测[J]. 北京航空航天大学学报, 2022, 48(11): 2250-2258. HOU N N, LI D A, ZHAO J M. Non-line-of-sight signal detection based on unsupervised learning and particle filtering [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2250-2258(in Chinese).

vised learning-particle filtering,UL-PF)算法为提高定位精度做 2 次保障。

聚类是一种无监督的学习分组技术,其提供了一种自动形成相似事物聚类的方法,这就像自动分类[12]。粒子滤波是一种视觉跟踪工具,依赖于蒙特卡罗链框架和贝叶斯估计,适用于非线性和非高斯系统[13]。文献[14]提出采用粒子滤波方法结合行人航位推算算法和磁性指纹图谱有效提高室内定位的精度和效率。文献[15]提出了一种基于遗传算法的简化粒子滤波方法。文献[16]提出了一种改进蒙特卡罗定位算法的概率激光定位方法,通过加入卡尔曼滤波后的 GNSS 信息来增强粒子权值,从而提高自主车辆定位精度。文献[17]根据粒子的状态值将采样粒子集合的状态空间划分为可变大小的可数测度空间,实验结果表明,在空间域对采样粒子进行简单的启发式划分后,性能有显著的提高。因而可以进一步研究一种挖掘采样粒子空间分布相关信息的有效方法。

对于定位优化部分,本文使用通过聚类算法优化的粒子滤波方法,为提高定位精度做二次保障。首先,提出了一种基于核 k-means 聚类算法的采样粒子状态空间分割方法,该方法考虑了采样粒子在状态空间分布中的内在相似性。其次,探索在每个聚类中选择一个粒子作为重要粒子,利用时间序列相关技术提高重采样粒子集的多样性。最后,通过伪距残差似然函数有效性检验,证明了该算法重采样的粒子集与原始粒子集来自同一分布。实验结果表明,与传统粒子滤波算法相比,所提算法能获得更准确的估计。

1　GNSS 信号核 k-means 聚类算法

1.1　GNSS 信号特征提取分析

使用无监督学习对 GNSS 信号进行分类的关键是从原始观测数据中提取主要特征。原始数据一般包含伪距、载波相位、载噪比和多普勒频移等,他们都与 GNSS 信号类型有着密切的联系。而任何单一特征都不可能对 GNSS 信号进行有效的分类,因此,需要通过不同特征的组合来提高分类的准确性。

1)载噪比。一般来说,LOS 载噪比较高,而 NLOS 载噪比较低。在城市峡谷环境中,经过光滑的建筑物表面反射后,NLOS 信号的强度可能与 LOS 强度相当,这种情况下,仅利用载噪比已经难以将二者区分开来。

2)伪距残差。当观测环境较好时,伪距残差偏大的对应 NLOS,但当观测环境不好时,可见卫星少于 4 颗时,此时的观测方程不再超定,伪距残差为 0,不足以作为区分指标。

3)卫星高度角。通常情况下,LOS 卫星的高度角较大,而城市峡谷环境中,可见卫星能见度差,LOS 卫星被高楼遮挡,卫星高度角也比较大,因此使用卫星高度角可能出现误判,还会破坏卫星的几何分布。

1.2　k-means 算法

对于复杂环境下 GNSS 接收机接收到的卫星信号一般分为 LOS、多径和 NLOS 信号 3 类,每一类信号都与载噪比、伪距残差和卫星高度角有一定的内在联系。根据这一特性,本文采用核 k-means 算法进行信号的聚类,当其中样本离聚类中心最近时,即归为该类。对于给定的一组样本集 $X=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\}$,$\boldsymbol{x}_N=\{r_n,v_n,e_n\}$ 为标准化载噪比、伪距残差和卫星高度角组成的特征向量,N 为样本数,n 为样本的特征向量下标,k-means 算法将其划分为 M 个簇 $C_1,C_2,\cdots,C_M$,使得每个数据点与其最近的聚类中心的满足平方欧几里得距离和最小,即

$$E(\boldsymbol{m}_1,\cdots,\boldsymbol{m}_M)=\sum_{i=1}^{N}\sum_{k=1}^{M}I(\boldsymbol{x}_i\in C_k)\|\boldsymbol{x}_i-\boldsymbol{m}_k\|^2 \tag{1}$$

式中:$\boldsymbol{x}_i$ 为样本,i 为样本编号;I 为 M 个簇中选取的样本数据集;k 为簇的编号;$\boldsymbol{m}_k$ 为均值向量,表达式为

$$\boldsymbol{m}_k=\frac{1}{|C_k|}\sum_{\boldsymbol{x}_i\in C_k}\boldsymbol{x}_i \tag{2}$$

k-means 算法的缺点是:最终的分类结果取决于集群中心的起始位置,且星团是线性分开的。为了克服这个问题,选择使用核 k-means 聚类算法[18]。核 k-means 聚类算法通过使用 k-means 算法中包含 N 个数据大小的初始位置的确定性全局,一次动态添加一个聚类中心[19]。

1.3　核 k-means 聚类算法

1)初始化。在数据集中随机选取 K 作为初始质心 c_p,$1\leqslant p\leqslant K$,p 为样本数量,K 为数据集的数量。

2)聚类。计算每条记录与质心之间的相似度 $\mathrm{sim}(r_i,c_p)$,并将记录 r_i 分配给相似的 c_p 质心。

3)质心重新计算。记录 r_i 将平均嵌入向量分配给 c_p 质心,重新计算每个群集的新质心,即 $c'_p=\frac{1}{nc_p}\sum_{i=1}^{nc_p}r_i$。$nc_p$ 为质心 c_p 指示的集群中的记录数。

4) 终止条件。所有旧质心的矩阵为 $\boldsymbol{\Theta} = [c_1, c_2, \cdots, c_K]^{\mathrm{T}}, 1 \leqslant p \leqslant K$，所有新质心的矩阵为 $\boldsymbol{\Theta}' = [c_1', c_2', \cdots, c_K']^{\mathrm{T}}$，下一步，欧几里得距离为

$$d(\boldsymbol{\Theta}, \boldsymbol{\Theta}') = \sqrt{\sum_{p=1}^{K} (c_p - c_p')^2} = \sqrt{\sum_{p=1}^{K} \sum_{i=1}^{\lambda} (v_{p,i} - v_{p,i}')^2} \tag{3}$$

式中：$v_{p,i}$ 和 $v_{p,i}'$ 分别为 c_p 和 c_p' 的质心元素。

核 k-means 聚类算法是 k-means 算法的推广，通过非线性变换 $\boldsymbol{\Phi}$ 数据点从空间输入映射到高维特征空间。在 k-means 算法核中，由于重叠数据在新的维度空间中可以是线性的，因此，期望数据能够很好地分开[18]。

用核 k-means 聚类算法约简的目标函数等同于特征中的分组误差，如下[19]：

$$E(\boldsymbol{m}_1, \cdots, \boldsymbol{m}_M) = \sum_{i=1}^{N} \sum_{k=1}^{M} I(\boldsymbol{x}_i \in C_K) \| \boldsymbol{\Phi}(\boldsymbol{x}_i - \boldsymbol{m}_k) \|^2 \tag{4}$$

$$\boldsymbol{m}_k = \frac{\sum_{i=1}^{N} I(\boldsymbol{x}_j \in C_K) \boldsymbol{\Phi}(\boldsymbol{x}_i)}{\sum_{i=1}^{N} I(\boldsymbol{x}_j \in C_K)} \tag{5}$$

$$\| \boldsymbol{\Phi}(\boldsymbol{x}_i - \boldsymbol{m}_k) \|^2 = K_{ii} - \frac{2 \sum_{j=1}^{N} I(\boldsymbol{x}_i \in C_K) K_{ij}}{\sum_{i=1}^{N} I(\boldsymbol{x}_i \in C_K)} + \frac{\sum_{j=1}^{N} \sum_{i=1}^{N} I(\boldsymbol{x}_j \in C_K) I(\boldsymbol{x}_l \in C_K) K_{jl}}{\sum_{j=1}^{N} \sum_{i=1}^{N} I(\boldsymbol{x}_j \in C_K) I(\boldsymbol{x}_l \in C_K)} \tag{6}$$

式中：K_{ii}、K_{ij} 和 K_{jl} 分别为对应样本 ii、ij、jl 的径向基函数的核函数，本文选择径向基函数(RBF)，B-Spline核定义在区间[-1,1]上。由递归公式给出：

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = B_{2p+1}(\boldsymbol{x}_i - \boldsymbol{x}_j) \tag{7}$$

式中：$p \in X$ 且 $B_{i+1} = B \otimes B_0$，X 为样本集，$\otimes$ 指圈乘运算。得到核函数：

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{p=1}^{d} B_{2n+1}(\boldsymbol{x}_{ip} - \boldsymbol{x}_{jp}) \tag{8}$$

使用径向基函数核 k-means 聚类算法，通过式(3)～式(8)，所有的 GNSS 信号都被归类到不同的集群，k-fold 将样本数据划分为 m 个部分，复杂环境中 GNSS 信号通常分为 LOS、多径和 NLOS 共 3 类，因此选择 k-fold = 3。由于卫星与终端的距离保持在一定的范围内，LOS 环境中同步的峰值信噪比(PSNR)也会相应保持在一个区间内。采用无监督机器学习的数据聚类方法，将实测数据分为 LOS、NLOS 和多径信号。图 1 为基于核 k-means 聚类算法的 GNSS 信号的仿真分类结果。图 1(a)为分类前只能已知目前存在 3 种信号，但无法判别和区分，图 1(b)为加入特征向量提取后的聚类方法可以识别 3 种信号分别对应的分类，蓝色为多径，红色为 LOS，绿色为 NLOS，横轴为伪距残差 ρ，纵轴为 PSNR。NLOS 的 PSNR 为负值，多径与 LOS 的 PSNR 为正值，而且多径的伪距残差大于 LOS。结果表明，在二维状态空间中，按一定的度量被分成 3 个簇。不同的分类结果以不同的颜色显示，可以看到具有特定属性的信号聚类在一起，而不是简单的分割。最后，利用本节所述的信号分类方法，便可见判定卫星信号是 LOS 还是 NLOS 传播，需从伪距观测量中剔除该卫星的信号。将受 NLOS 效应干扰严重的卫星剔除，获得精度更高的伪距观测量，为后续优化定位算法做准备。

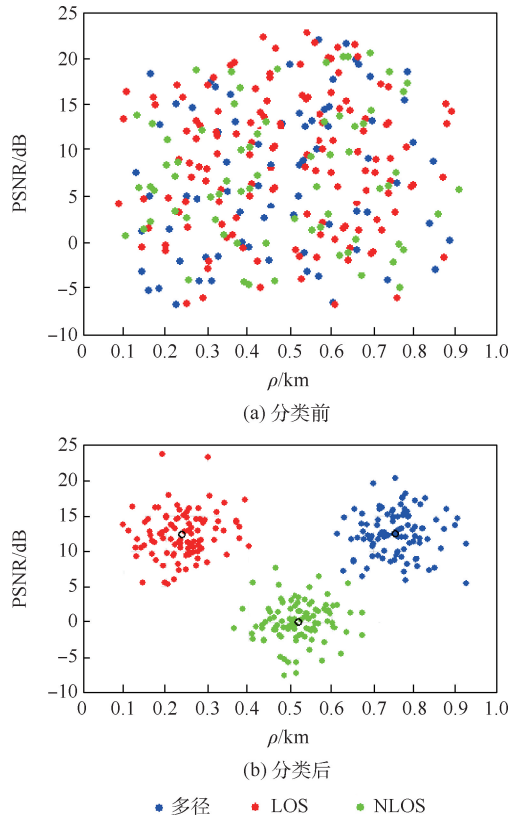


图 1 基于核 k-means 算法的 GNSS 信号仿真分类结果

Fig. 1 GNSS signal simulation classification results based on kernel k-means algorithm

2 粒子滤波定位算法

步骤 1 粒子初始化处理。对 GPS 接收机输出的原始测量值进行初始化处理，接收机输出的原始伪距集 $\boldsymbol{\rho} = [\rho_1, \rho_2, \cdots, \rho_N]^{\mathrm{T}}$ 和经接收机计算得到的参考位置 y，用来初始化位置候选点。使

用该方法实现位置估计,需要 4 颗及以上 GNSS 卫星提供的原始伪距测量值。首先分配位置候选点 $P^{(i)}=(x^{(i)},y^{(i)})$,$i$ 为候选点的编号。用随机分布在参考位置的二维高斯分布粒子来表示候选点。在先前的估计位置 $x(t-\Delta t)$ 也要放置粒子,t 为选取的位置候选点对应的采样时间。在重采样期间,选择正态分布生成粒子,一半粒子的分布为 $N(y(t),\boldsymbol{\Sigma}_0)$,$\boldsymbol{\Sigma}_0$ 表示正态分布粒子的均值,另一半粒子的分布为 $N(x(t-\Delta t),\boldsymbol{\Sigma}_{\Delta t})$。在城市环境定位场景中,行人在 1 s 内发生剧烈位置变化的可能性极低,因此,可以利用历元$x(t-\Delta t)$的位置信息。

步骤 2　测量伪距集与模拟伪距集差计算。假设 $P^{(i)}$ 作为地面参考位置,候选点 $P^{(i)}$ 和第 n 时刻的卫星位置 x_n^{SV} 之间的模拟伪距 $\hat{\boldsymbol{\rho}}_n^{(i)}$,如下:

$$\hat{\boldsymbol{\rho}}_n^{(i)}=R_n^{(i)}+c(\delta t^{r(i)}-\delta t_n^{\mathrm{SV}})+I_n+T_n+\varepsilon_n^{\mathrm{refl}(i)} \tag{9}$$

式中:通过伪距几何距离的计算,$c\delta t_n^{\mathrm{SV}}$ 为卫星钟差等效距离;$c\delta t^{r(i)}$ 为接收机钟差等效距离;I_n 为电离层延迟;T_n 为对流层延迟;$\varepsilon_n^{\mathrm{refl}(i)}$ 为反射延迟距离;δt_n^{SV}、I_n 和 T_n 分别为利用广播星历模型、Klobuchar 模型和 Saastamoinen 模型校正后的值。估计接收机钟差 $\delta t^{r(i)}$ 可以最小化模拟伪距集和测量伪距集之间的差异。如果模拟伪距集与测量伪距集受信号影响的方式相同,那么模拟伪距集应该与测量伪距集相似。通过比较 2 组伪距集,可以计算模拟伪距集 $\hat{\boldsymbol{\rho}}_n^{(i)}$ 和测量伪距集 $\boldsymbol{\rho}_n$。伪距差 $d_{\mathrm{pr}}^{(i)}$ 可以定义为

$$d_{\mathrm{pr}}^{(i)}=\min_{\delta t^{r(i)}}\sum_n^{N^{\mathrm{sim}}}\frac{|\boldsymbol{\rho}_n-\hat{\boldsymbol{\rho}}_n^{(i)}(\delta t^{r(i)})|}{N^{\mathrm{sim}}} \tag{10}$$

式中:N^{sim}为经过非视距排除后重新选择的伪距观测量个数。

步骤 3　评估候选点位置。伪距相似性决定每个候选点的似然函数$\alpha^{(i)}(t)(i=1,2,\cdots,N_P)$,似然函数计算如下:

$$\alpha^{(i)}(t)=\begin{cases}\exp\left(-\dfrac{d_{\mathrm{pr}}^{(i)2}}{\sigma_0^2}\right)\\ 0\end{cases} \tag{11}$$

$d_{\mathrm{pr}}^{(i)}<C_{\mathrm{pr}}$情况下为有效粒子,其他情况下为无效粒子。似然函数根据伪距集的相似性确定,利用 NLOS 信号检测结果对样本点进行有效性检验。根据状态转移概率分布函数,将采样的粒子转移到状态空间中的任意位置。因此,采样粒子的多样性可以用粒子在状态空间中的分布来表示。为了避免重采样过程中随机丢弃一些权重较小的粒子,保持重采样粒子集的多样性,采用核 k-means 聚类算法对采样的粒子进行分组,使用一个合理的度量。然后,在每组中选择一个称为重要粒子的粒子,传播到下一个。根据分类结果评估位置候选点,聚类算法更准确的确认采样点,识别出 NLOS 卫星,则排除该采样点,从而提高粒子滤波算法定位精度。在本实验所选的小区环境中,狭窄的街道两侧建筑物距离分布密集,而且行人不易发生剧烈位置变化,根据行人与本实验环境中建筑物之间的距离选择合适的阈值,选择一个阈值 $C_{\mathrm{pr}}=10$ m 来排除异常的 NLOS 候选点。σ_0 是一个用来稳定估计值的常数值,城市环境中建筑物密集,而郊区建筑物分散对应该常数值会偏大,本文的城市环境选择设置为 20 m。

步骤 4　输出定位结果。对所有有效样本不同时刻位置 $P^{(i)}(t)$ 加权平均,为最终粒子滤波校正位置提出的方法的定位结果为

$$x(t)=\frac{\sum_{i=1}^{N_p}\alpha^{(i)}(t)P^{(i)}(t)}{\sum_{i=1}^{N_p}\alpha^{(i)}(t)} \tag{12}$$

3　基于 UL-PF 算法

本节工作是利用核 k-means 聚类算法进一步研究采样粒子的空间分布特征,然后在每个聚类中通过时间和序列分析选择一些被称为重要粒子的粒子,以增强重采样粒子的多样性。核 k-means 聚类算法通过研究采样粒子的空间分布特征,通过时间和序列分析选择一些重要粒子,以增强粒子滤波重采样粒子的多样性,能够缓解粒子退化,将优化后粒子滤波算法用来后续评估候选点位置,可以提高卫星导航系统的定位精度。

1)复制所有粒子候选点。接收机输出的原始伪距构成粒子候选点,复制选取的位置粒子候选点,将所有候选点$(\rho_t^h,\alpha_t^h)^{N_t}$权重 α_t^h 乘以采样大小 N_t,将 $N_t\alpha_t^h$ 的整数部分分配给 n_t^h:

$$n_t^h=\lfloor N_t\alpha_t^h\rfloor \tag{13}$$

将第 h 个时刻粒子(ρ_t^h,α_t^h)复制 n_t^h 次,并分配相等的新权重,计算粒子残差权重:

$$\alpha_t^h=\frac{N_t\alpha_t^h-\lfloor N_t\alpha_t^h\rfloor}{N_t} \tag{14}$$

2)使用核 k-means 聚类算法将候选粒子集划分为带有 N_p 个粒子的聚类 g_k:

$$g_k\propto\{(\rho_{t,k}^h,\alpha_{t,k}^h)\mid i=1,2,\cdots,N_p\} \tag{15}$$

3）选择每个类中下一次筛选的重要粒子：

$$g_k \rightarrow (\hat{\rho}_{t,k},\hat{\alpha}_{t,k})\begin{cases}\hat{\rho}_{t,k} = \rho_{t,k}^{h} \\ \qquad i \in N_p, j \in N_p, S(i) > S(j) \\ \hat{\alpha}_{t,k} = \sum_{i=1}^{N_p} \alpha_{t,k}^{h}\end{cases} \tag{16}$$

式中：$S(i)$代表时间序列分析和核 k-means 选择结果。

4）将第 1 步中复制的粒子设置为相同的权重：

$$\alpha_t = \frac{1 - \sum_{j=1}^{k} \hat{\alpha}_{t,j}}{\sum_{i=1}^{k} n_i} \tag{17}$$

综上所述，本文的优化定位过程如下：首先，使用北斗卫星导航系统（beidou navigation satellite system，BDS）卫星导航系统的伪距观测量，选择位置候选点和粒子初始化处理，将第 2 节介绍的 GNSS 信号分类结果作为 NLOS 信号检测结果，用来参与后续伪距差的计算；然后，将优化后的粒子滤波器用来优化接收机定位结果，同时，核 k-means 聚类算法更新粒子权重用来评估位置候选点；最后，将优化后卫星导航系统的伪距观测量，参与最终的定位结果解算。

4 实验结果

4.1 实验环境

实验过程使用 u-blox NEO-M8N GNSS 接收器和 SIS600B1L1 中频采样器采集数据，实验设备如图 2（a）、（b）、（d）所示，测试数据的采集地点选择山西太原小店区太航小区，这一区域是城市环境中常见的双侧街道环境，实验环境如图 2（c）所示。

4.2 定位性能对比

仿真基于 BDS 卫星的星历数据，行人动态数据集的轨迹图及按照本文算法的信号分类结果如图 3 所示。图 3（a）表示分类前的候选点分布图，图 3（b）根据分类结果评估位置候选点，在 A、B 区域没有受到跨街街道的影响，LOS 卫星可见数目较多，由于受到四周高楼遮挡，在 C、D、E 横街区域接收到 NLOS 卫星较多。核 k-means 聚类算法更准确的确认采样点，识别出 NLOS 卫星，则排除该采样点，从而提高粒子滤波算法定位精度。

图 4 为动态场景下定位误差对比。A、B 段 LOS 卫星较多，定位效果较好，A 段误差比 B 段大，是因为 A 段建筑物周围还有树木遮挡。C、D、

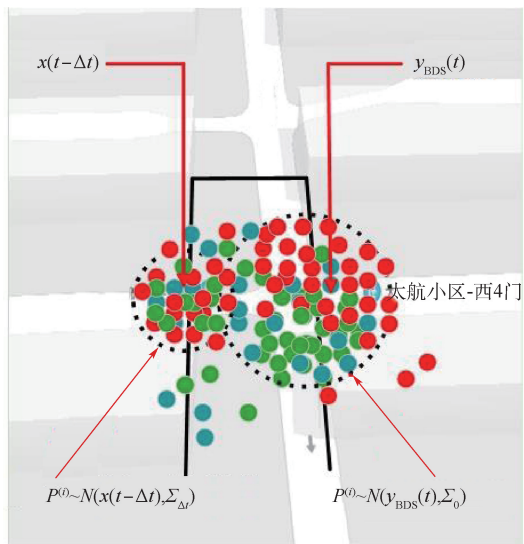


(a) GNSS接收器　(b) 中频采样器

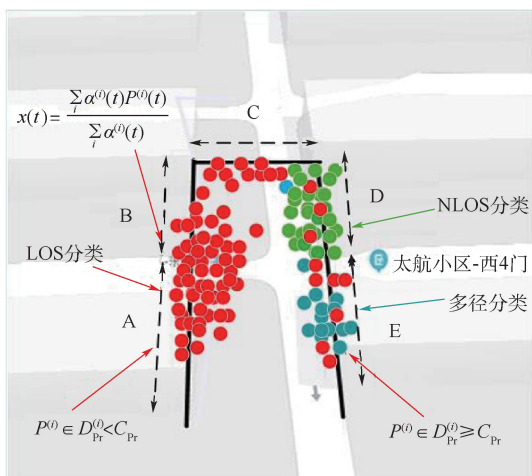
(c) 实验环境　(d) 数据采集

图 2　实验条件与数据采集

Fig. 2　Experimental conditions and data acquisition



(a) 位置候选点分布

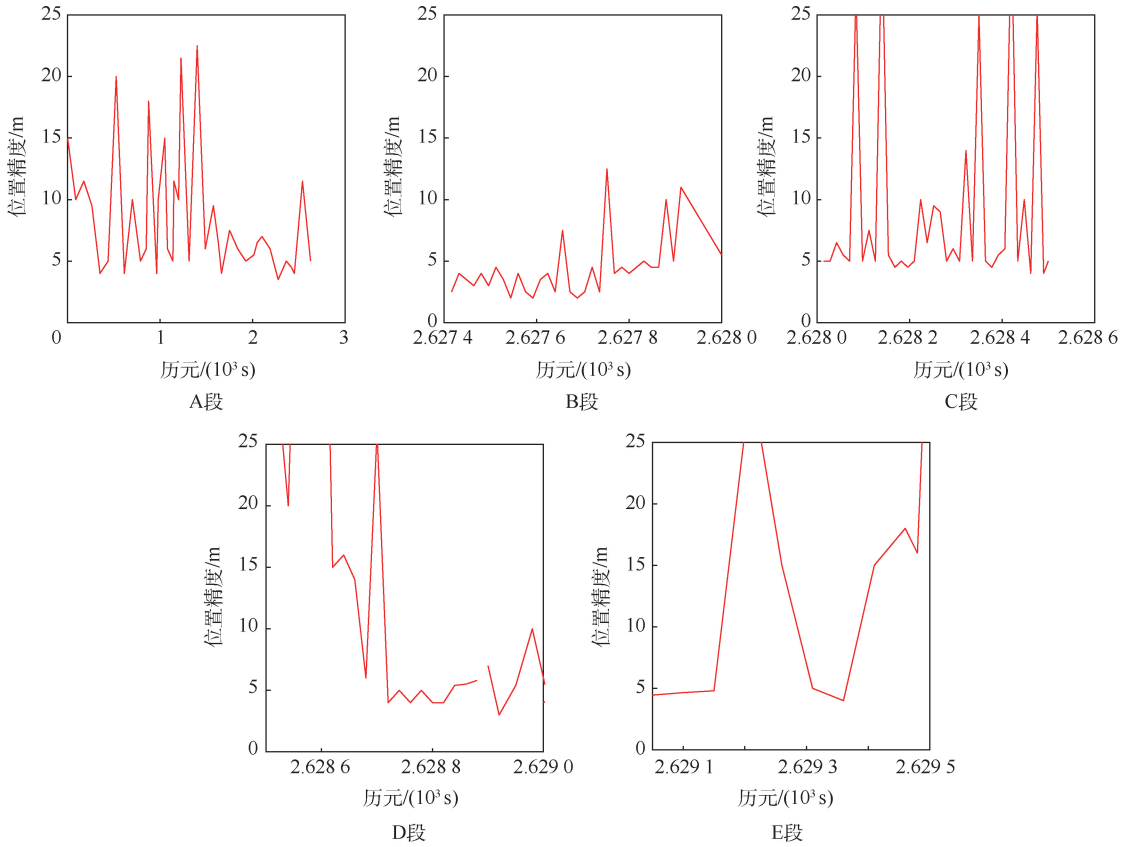


(b) 分类结果

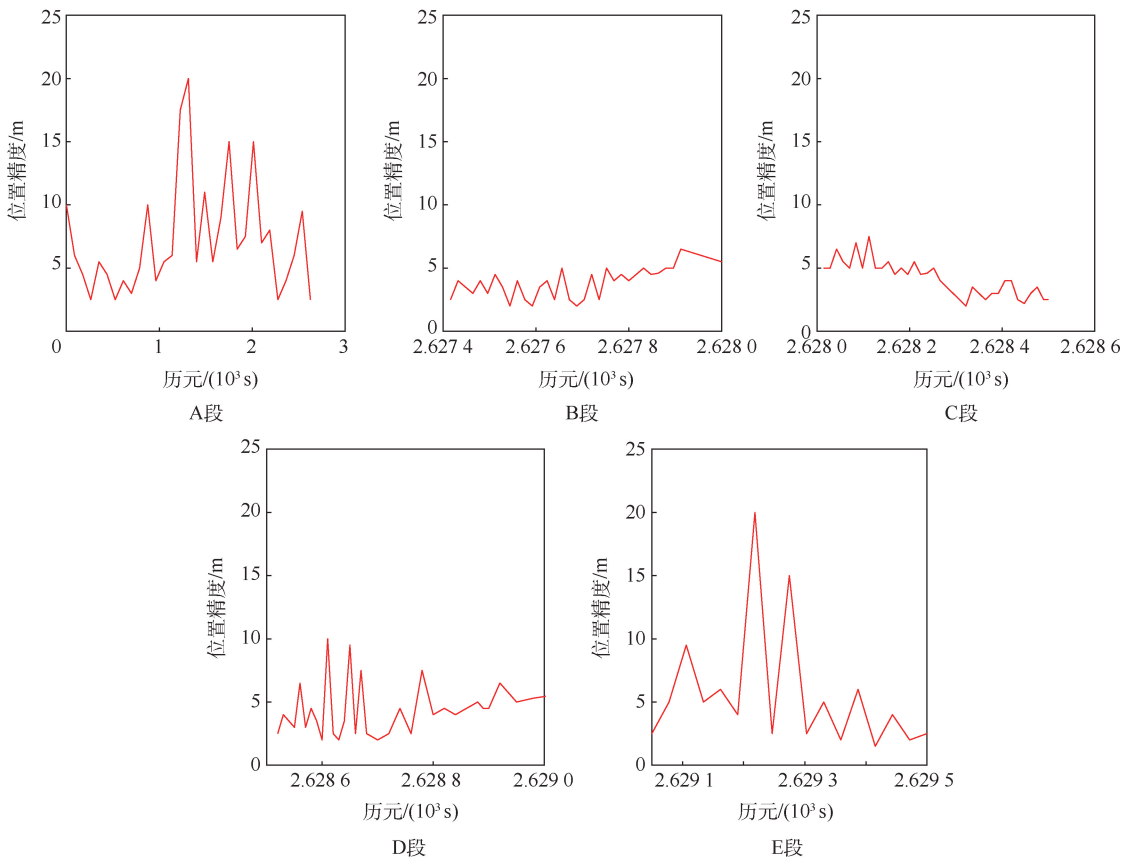
图 3　位置候选点分布及分类结果

Fig. 3　Distribution of location candidate points and classification results

E 段 NLOS 卫星较多,未剔除 NLOS 干扰前误差　　甚至高达 25 m,而且误差曲线不规则不光滑,严



(a) 未剔除NLOS



(b) 剔除NLOS

图 4　动态场景中剔除 NLOS 前后定位误差

Fig. 4　Positioning error before and after NLOS elimination in dynamic scene

重影响定位精度。确定 C、D、E 段信号分类结果后,将剔除 NLOS 卫星再进行计算,而剔除 NLOS 后的误差在 5 m 左右,表明 NLOS 信号会给定位带来一定程度的系统误差,正确地去除 NLOS 能够有效地缓解这种系统误差。

4.3 算法性能对比

表 1 给出了核 k-means 聚类算法与现有的监督机器学习算法、k-最近邻(KNN)、朴素贝叶斯(NB)、决策树(DT)和最小二乘支持向量机(LS-SVM)的计算复杂度和性能比较。为简单起见,使用运行时间来度量计算复杂度。仿真中,分别生成500 个LOS 波形和 500 个 NLOS 波形来训练有监督机器学习算法,不同算法的性能对比如表 1所示。

表 1 不同算法对比

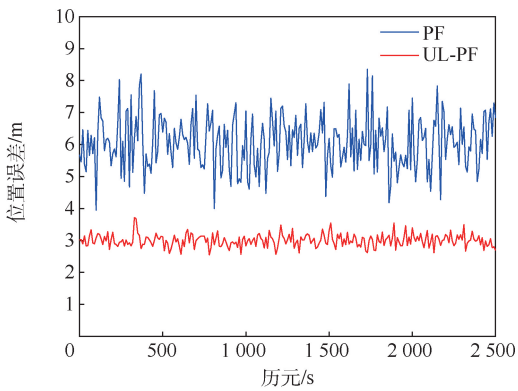
Table 1 Comparison of different algorithms

算法	运行时间/s	LOS 检测准确度/%	NLOS 检测准确度/%	正确率/%	错误率/%
KNN	0.036 9	89.2	95.2	92.2	7.8
NB	0.039 1	85.4	88.2	86.8	13.2
DT	0.055 1	97.2	94.3	95.75	2.25
LS-SVM	0.128 5	98.4	99.4	98.9	1.1
核 k-means	0.060 2	97.9	98.9	98.4	1.6

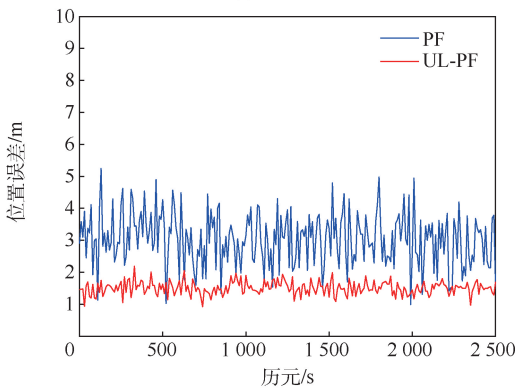
从表 1 中可以看出,无监督机器学习核 k-means 聚类算法与有监督机器学习算法相比,在性能上有一定的差异,但由于没有训练阶段,所以不需要训练数据。核 k-means 聚类算法只需要最先进的监督机器学习 LS-SVM 所需的 44% 的运行时间,但其实现了 LS-SVM 算法几乎相同的 NLOS 检测准确度。

为进一步证明本文提出的 UL-PF 算法性能,图 5 为复杂环境和空旷环境中定位误差。复杂环境中,传统粒子滤波(particle filter, PF)算法定位误差平均值约为 6 m,UL-PF 算法平均值在3 m左右,并且波动幅度较小,定位性能较为稳定。空旷环境中,传统粒子滤波算法定位误差平均 3 m 左右,UL-PF 算法平均在 1.5 m 左右。

图 6 将本文算法与传统算法、PF 和扩展卡尔曼滤波(extended Kalman filter, EKF)算法进行



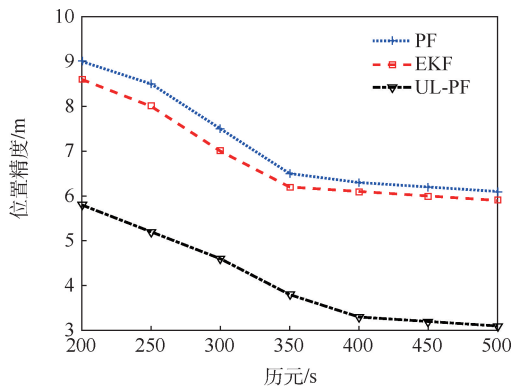
(a) 复杂环境



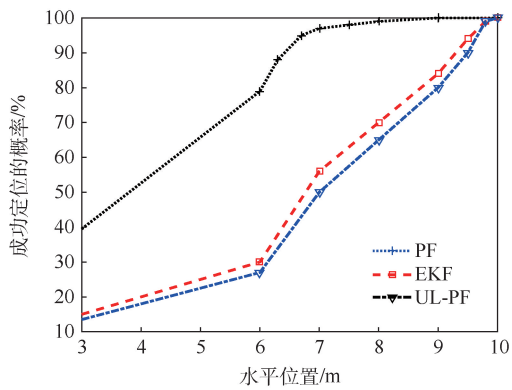
(b) 空旷环境

图 5 不同环境下本文算法与传统算法定位误差对比

Fig. 5 Comparison of positioning errors between the proposed algorithm and traditional algorithms in different environments



(a) 收敛时间对比



(b) 区位概率对比

图 6 不同算法的收敛时间和区位概率对比

Fig. 6 Comparison of convergence time and location probability of different algorithms

对比。图 6(a)显示了使用单个 LEO 卫星的平均定位精度的仿真结果。纵轴表示平均定位精度。测量数据中含有大量的 NLOS 数据,导致传统算法定位精度较差,收敛时间较长。当测量时间达到 350 s 时,传统方法的精度逐渐收敛到 6 m 左右。与传统算法相比,基于非监督学习 NLOS 排除的定位方法大大降低了 NLOS 误差的影响,使定位精度显著提高,收敛时间缩短当测量时间达到 200 s 时,定位精度约为 6 m;当测量时间达到 400 s时,定位精度收敛到约 3 m。

图 6(b)显示了城市场景中的位置概率。由于卫星定位精度和定位成功率与测量误差密切相关,本文算法对测量数据中的非视距信号进行分类排除,可以有效降低非视距信号误差的影响,提高定位成功率。如图 6(b)所示,本文算法的成功定位概率明显高于其他 2 种算法,在定位误差为 7 m 时,定位成功率可达 99%,而传统算法在定位误差为 10 m 时,定位成功率可分别达到 97% 和 99%。

5　结　论

针对城市环境中非视距效应问题,本文介绍了一种基于无监督学习和粒子滤波的非视距信号检测方法,进而通过优化的粒子滤波器来优化定位结果,为提高精度提供 2 次保障。实验结果表明,剔除 NLOS 信号影响后,能有效提高平均定位精度和缩短算法收敛时间。

1) 与传统方法相比,基于非监督学习 NLOS 排除的终端定位方法定位精度显著提高,收敛时间缩短,当测量时间达到 200 s 时,定位精度约为 6 m;当测量时间达到 400 s 时,定位精度收敛到约 3 m。

2) 所提方法的成功定位概率明显高于粒子滤波和扩展卡尔曼滤波 2 种方法,在定位误差为 7 m时,定位成功率可达 99%,而传统方法在定位误差为 10 m 时,定位成功率可分别达到 97% 和 99%。

尽管本文针对缓解 NLOS 效应已经有一定研究成果,但仍有需要深化和改进的方面。机器学习方法对于不同的数据集场景有着不同的分类效果,本文只采用低速动态的行人数据集进行实验,还需进一步研究接收机处于车辆实时高速动态环境下场景变化对分类效果的影响,此外,针对 GNSS 信号的载噪比,伪距残差和卫星高度角的样本数据集,还需选择更加合理的特征向量归一化方法。本文所开展的研究具有一定的社会价值,可以应用于驾考系统、智慧城市及无人驾驶等方面,加快城市现代化进程建设。

参考文献(References)

[1] 曹建波,王宇飞,董军. 码相关参考波形技术在卫星导航信号中的多径抑制性能研究[J]. 现代导航,2019,10(2):79-83.
CAO J B,WANG Y F,DONG J. Research of multipath mitigation performance for CCRW technique in navigation signal[J]. Modern Navigation,2019,10(2):79-83(in Chinese).

[2] 孙旭升. GPS 多径抑制算法研究报告[J]. 信息通信,2017,30(2):13-14.
SUN X S. Research report on GPS multipath suppression algorithm[J]. Information & Communications,2017,30(2):13-14(in Chinese).

[3] 李惟,朱云龙,王峰,等. GNSS 多径信号模型及测高方法[J]. 北京航空航天大学学报,2018,44(6):1239-1245.
LI W,ZHU Y L,WANG F,et al. GNSS multipath signal model and altimetry method[J]. Journal of Beijing University of Aeronautics and Astronautics,2018,44(6):1239-1245(in Chinese).

[4] 王蓉,陈思,胡伟锋,等. 卫星导航定位误差仿真分析[J]. 计算机测量与控制,2018,26(10):264-268.
WANG R,CHEN S,HU W F,et al. Simulation and analysis of navigation satellite positioning error[J]. Computer Measurement & Control,2018,26(10):264-268(in Chinese).

[5] 王冠宇,孙蕊,程琦. 城市峡谷中基于 GPS 双极化天线的定位方法[J]. 导航定位学报,2020,8(4):13-19.
WANG G Y,SUN R,CHENG Q. Positioning method based on GPS dual-polarized antenna in urban canyon environment[J]. Journal of Navigation and Positioning,2020,8(4):13-19(in Chinese).

[6] 何通,刘怡俊. 基于地图拟合的 GNSS 接收机抗多径干扰技术[J]. 电脑知识与技术,2019,15(6):261-262.
HE T,LIU Y J. Anti-multipath interference technology of GNSS receiver based on map fitting[J]. Computer Knowledge and Technology,2019,15(6):261-262(in Chinese).

[7] YUAN Y,SHEN F,LI X D. GPS multipath and NLOS mitigation for relative positioning in urban environments[J]. Aerospace Science and Technology,2020,107(12):106-128.

[8] MARANÒ S,GIFFORD W M,WYMEERSCH H,et al. NLOS identification and mitigation for localization based on UWB experimental data[J]. IEEE Journal on Selected Areas in Communications,2010,28(7):1026-1035.

[9] YOZEVITCH R,MOSHE B B,WEISSMAN A. A robust GNSS LOS/NLOS signal classifier[J]. Navigation,2016,63(4):429-442.

[10] ZHU F,BA T E,ZHANG Y,et al. Terminal location method with NLOS exclusion based on unsupervised learning in 5G-LEO satellite communication systems[J]. International Journal of Satellite Communications and Networking,2020,38(5):425-436.

[11] FAN J C,AWAN A S. Non-line-of-sight identification based on unsupervised machine learning in ultra wideband systems[J].

IEEE Access,2019,7:32464-32471.

[12] 陈唯实,闫军,李敬. 基于 Rao-Blackwellized 蒙特卡罗数据关联的检测跟踪联合优化[J]. 北京航空航天大学学报,2018,44(4):700-708.
CHEN W S, YAN J, LI J. Joint optimization of detection and tracking with Rao-Blackwellized Monte Carlo data association [J]. Journal of Beijing University of Aeronautics and Astronautics,2018,44(4):700-708(in Chinese).

[13] 李凯锋,王保良,黄志尧,等. K-均值聚类在 CCERT 系统流型辨识中的应用[J]. 北京航空航天大学学报,2017,43(11):2280-2285.
LI K F, WANG B L, HUANG Z Y, et al. Application of K-means clustering in flow pattern identification of CCERT system [J]. Journal of Beijing University of Aeronautics and Astronautics,2017,43(11):2280-2285(in Chinese).

[14] NING F S, CHEN Y C. Combining a modified particle filter method and indoor magnetic fingerprint map to assist pedestrian dead reckoning for indoor positioning and navigation[J]. Sensors (Basel, Switzerland),2019,20(1):185-193.

[15] DE MIGUEL M Á, GARCÍA F, ARMINGOL J M. Improved LiDAR probabilistic localization for autonomous vehicles using GNSS[J]. Sensors (Basel, Switzerland), 2020, 20(11): 139-145.

[16] YANG W M, SONG L, TEE C A T, et al. Unsupervised learning grouping-based resampling for particle filters[J]. IEEE Access,2019,7:127265-127275.

[17] LI T C, SATTAR T P, SUN S D. Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters[J]. Signal Processing,2012,92(7):1637-1645.

[18] TZORTZIS G, LIKAS A. The global kernel k-means clustering algorithm[C]// 2008 IEEE International Joint Conference on Neural Networks. Piscataway: IEEE Press,2008:10365371.

[19] AGRAWAL A, GUPTA H. Global K-means (GKM) clustering algorithm: A survey[J]. International Journal of Computer Applications,2013,79(2):20-24.

Non-line-of-sight signal detection based on unsupervised learning and particle filtering

HOU Ningning[1,2], LI Deng'ao[2,3,*], ZHAO Jumin[1,2]

(1. College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China;
2. Shanxi Province Spatial Information Network Engineering Technology Research Center, Taiyuan 030024, China;
3. College of Data Science, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: Global navigation satellite system (GNSS) is the most widely used positioning technology at present. Due to high-rise structures blocking the signal, the performance deterioration caused on by non-line-of-sight propagation still remains while studying the positioning problem in urban canyons. In order to solve this problem, the unsupervised learning-partinle filter (UL-PF) algorithm is proposed. In the satellite signal classification stage, the unsupervised learning classification method using kernel k-means clustering is used. In the positioning stage, the particle filter method optimized by clustering algorithm is used. The method first considers the inherent similarity of the sampled particles in the state space distribution. Secondly, it explores how to select one particle as the key particle in each cluster and increase the diversity of resampled particle sets by using time series correlation techniques. Experiments show that the average positioning accuracy of the proposed algorithm in urban is improved from 15 m to about 5 m, and the convergence time is reduced from 500 s to about 200 s.

Keywords: global navigation satellite system; non-line-of-sight classification; kernel k-means; particle filtering; unsupervised learning

Received: 2021-02-08; **Accepted**: 2021-04-23; **Published online**: 2021-05-11 11:39

URL: kns.cnki.net/kcms/detail/11.2625.V.20210511.1005.001.html

Foundation items: National Natural Science Foundation of China (61772358); National Key R & D Program of China (2018YFB2200900)

* **Corresponding author**. E-mail: lidengao@tyut.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0081

飞机旋转表面结冰数值模拟

郭琦[1], 申晓斌[1,2,*], 林贵平[1], 张世娟[3]

(1. 北京航空航天大学 航空科学与工程学院, 北京 100191;
2. 中国空气动力研究与发展中心 结冰与防除冰重点实验室, 绵阳 621000; 3. 武汉航空仪表有限责任公司, 武汉 430074)

摘 要: 在 Shallow-Water 结冰热力学模型的基础上,搭建了一套适用于三维旋转表面的非稳态结冰模型,采用 Jacobi 迭代法和 Gauss-Seidel 迭代法对表面结冰进行非稳态迭代求解。用所提模型对简化后的旋转桨叶模型进行计算,并与 FENSAP 软件结果进行对比,验证了所提模型的准确性,分析了转速、水滴直径和液态水含量等因素对旋转表面结冰冰形和水膜流动的影响。结果表明:随着转速的增加,结冰范围和水膜覆盖范围偏移愈加明显;结冰范围和水膜覆盖范围随水滴直径增长逐渐增加,水膜厚度也逐渐增大;结冰厚度和水膜厚度随液态水含量增长而相应增加,水膜覆盖范围也明显变大。

关 键 词: 飞机结冰; 旋转; 水膜流动; 非稳态; Shallow-Water 模型
中图分类号: V211.3
文献标志码: A **文章编号**: 1001-5965(2022)11-2259-11

飞机在结冰气象条件下飞行时,其迎风面上会积聚过冷水滴,发生冰的聚积,该现象为飞机结冰[1]。飞机积冰对飞机性能有重要影响,会造成升力和失速迎角减小,阻力增加[2],甚至导致严重的飞行事故[3]。随着航空科学技术的快速发展,航空飞行器得到广泛运用,由于飞机结冰引发的飞行安全问题愈发凸显。目前对飞机结冰的研究有结冰气象条件的探测、过冷水滴的撞击特性、结冰热力学模型与冰形预测、结冰对飞行性能的影响等,本文针对飞机旋转表面结冰进行数值模拟研究。

目前主要通过实验研究与数值仿真分析,对飞机结冰展开研究,其中数值方法具有经济性好、执行周期短等优点,并且由于计算流体力学技术和计算机技术的不断发展与成熟,数值仿真在飞机结冰研究中的作用越发重要。在国内外学者的共同努力下,飞机结冰数值模拟已取得了相当大的进展,结冰数值研究的内容由霜冰预测发展到了明冰的模拟,由二维计算逐步向三维仿真推进,由固定部件表面向旋转部件表面发展,三维旋转表面复杂流动结冰机理与模拟成为了目前研究的焦点。

传统的结冰计算一般采用 Messinger[4] 结冰热力学模型,该模型假设某个控制体中的未凝结液态水会全部流出到其他控制体中,没有考虑到控制体自身存在水膜的情况。美国国家航空航天局(NASA)于 20 世纪 80 年代研发成功的 LEWICE 软件[5]就是基于该结冰模型开发的。之后在 Messinger 模型的基础上,发展出了 Myers[6] 模型和 Shallow-Water[7] 结冰热力学模型,两者都考虑了控制体表面水膜存在的情况,不同的是,Myers 模型还考虑了冰层和水膜中的温度梯度分布。加拿大的 FENSAP-ICE 软件[7],是利用 Shallow-Water 模型对结冰表面的冰形进行求解。基于这些结冰模型,国内外学者对飞机结冰展开了

收稿日期: 2021-02-19; **录用日期**: 2021-05-07; **网络出版时间**: 2021-05-12 10:19
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210512.0828.001.html
基金项目: 国家自然科学基金(51806008); 结冰与防除冰重点实验室开放课题(IADL20200307)
* **通信作者**. E-mail: shenxiaobin@buaa.edu.cn

引用格式: 郭琦, 申晓斌, 林贵平, 等. 飞机旋转表面结冰数值模拟[J]. 北京航空航天大学学报, 2022, 48(11): 2259-2269.
GUO Q, SHEN X B, LIN G P, et al. Numerical simulation of icing on aircraft rotating surfaces[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2259-2269(in Chinese).

一系列模拟仿真研究。易贤等[8]发展了一种采用迭代求解的三维结冰热力学模型,该模型基于传统的 Messinger 二维模型,引入了结冰表面溢流水的流动。申晓斌等[9]基于 Messinger 结冰模型,建立了一种适用于三维结冰表面的液态水溢流模型,并完成了对三维发动机结冰冰形的计算。曹广州等[10]采用 Myers 结冰模型,对三维机翼模型的结冰增长和表面液态水流动进行了仿真求解,且能够在计算中自动辨别过冷水滴在结冰表面发生碰撞后的结冰冰形。Cao 和 Huang[11]在 Myers 模型的基础上,建立了三维结冰模拟模型,计算了后掠机翼的冰形。雷梦龙等[12]基于 Myers 结冰热力学模型,对其结冰类型判别方法进行改进,采用离散算法对表面液态水溢流和冰层生长过程进行计算,模拟了机翼表面的结冰过程。

同时,随着飞机结冰模拟技术的发展,各国学者针对飞机旋转部件结冰仿真计算也进行了多项研究。Bain 等[13]用 LEWICE3D 软件进行旋转桨叶的结冰计算,将旋转结冰过程分为 2 个步长进行计算,没有考虑旋转作用对表面水膜的作用力及对流动特性的影响。Reid 等[14-15]基于 Shallow-Water 溢流模型,利用旋转坐标系的方法对旋转桨叶的结冰进行了模拟。Switchenko 等[16]针对旋转桨叶采用二维截面的方法简化计算,通过二维多步法模拟特征截面处的结冰情况。Wang 与 Zhu[17]在旋转坐标系下计算了流场与水滴撞击特性,考虑离心力对溢流水膜分配的影响,进行了结冰分析。Dong 等[18]采用欧拉法求解了发动机处的水滴撞击特性,根据结冰热力学平衡模型得到了旋转帽罩的结冰冰形。赵秋月等[19]基于拉格朗日法,对旋转部件表面的水收集系数计算方法进行研究,将其用于发动机旋转整流帽罩的水滴撞击特性计算,发现整流罩表面的局部水收集系数会随着转速和锥体锥角的增大,沿表面外形线下降加快,撞击区域减小。吴孟龙等[20]基于欧拉法对发动机旋转帽罩三维表面的水滴撞击特性进行模拟,研究帽罩表面的水滴撞击特性受旋转速度的影响,发现旋转速度对帽罩表面的水滴撞击特性影响较小。

现在使用的结冰热力学模型几乎都是在固定部件表面建立的,一般认为空气剪切力为水膜的驱动力,用其来确定水膜的流动特性,而对于旋转部件表面的受力分析及旋转作用对水膜运动的影响分析较少,很少考虑离心力和科里奥利力的方向与大小对结冰效果的作用,旋转结冰过程中的运动力学与传热学机理研究尚不全面。本文采用 Shallow-Water 结冰模型,对飞机旋转表面的水膜流动和结冰冰形进行模拟,并分析转速、水滴直径和液态水含量等因素对旋转表面结冰的影响,为飞机旋转部件的结冰研究提供参考。

1 数学模型

为了研究旋转表面的水膜流动和结冰冰形,基于 Shallow-Water 结冰热力学模型[7],分析结冰表面水膜的流动与传热过程,建立动量守恒方程、质量守恒方程和能量守恒方程,进而计算旋转坐标系下的非稳态结冰过程与冰形。

1.1 动量守恒方程

旋转表面的水膜主要受到空气剪切力、压力梯度、重力、旋转导致的惯性力等因素的影响,选取表面微元控制体进行受力分析,如图 1 所示。用 S 表示曲面的表面,S'表示 S 在法向量 $\boldsymbol{n}$ 方向的投影,则 S 和 S'之间的距离就是水膜的厚度 h。

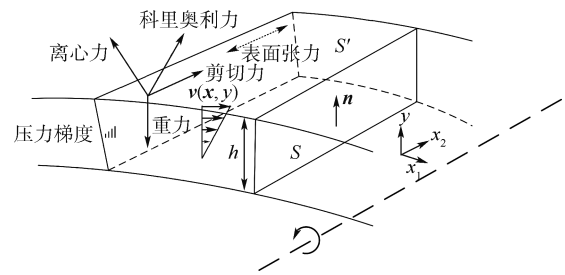


图 1 旋转表面水膜受力分析示意图

Fig. 1 Schematic diagram of force analysis of water film on rotating surface

由于水膜在结冰表面流动速度较小,且不考虑水滴撞击的影响,对水膜作常物性处理,可假设结冰表面的水膜流动为不可压层流流动[10],基于不可压流体的雷诺平均 Navier-Stokes 方程对水膜溢流过程进行建模,在旋转坐标系下其动量方程表述为

$$\frac{\partial}{\partial t}(\rho_{\mathrm{w}}\boldsymbol{v}) + \nabla\cdot(\rho_{\mathrm{w}}\boldsymbol{v}\boldsymbol{v} + p\boldsymbol{I}) = \frac{\partial}{\partial y}\left(\mu_{\mathrm{w}}\frac{\partial \boldsymbol{v}}{\partial y}\right) + \rho_{\mathrm{w}}[\boldsymbol{g} - 2\boldsymbol{\omega}\cdot\boldsymbol{v} - \boldsymbol{\omega}\cdot(\boldsymbol{\omega}\cdot\boldsymbol{r})] \tag{1}$$

式中:ρ_{w} 为水的密度;$\boldsymbol{v}$ 为水膜相对速度;p 为水膜内部压力;$\boldsymbol{I}$ 为单位张量;μ_{w} 为水的动力黏度;y 为表面法向坐标;$\boldsymbol{g}$ 为重力加速度;$\boldsymbol{\omega}$ 为转速;$\boldsymbol{r}$ 为当地矢径。其中,方程右边第 2 项为惯性力项,包括了科里奥利惯性力与牵连惯性力。

考虑结冰表面的水膜厚度很小,其主要受空气剪切力和旋转导致的惯性力的作用,忽略非稳态项、压力梯度项和重力项的影响[21],则动量方程可简化为

$$\rho_w[2\boldsymbol{\omega}\cdot\boldsymbol{v}+\boldsymbol{\omega}\cdot(\boldsymbol{\omega}\cdot\boldsymbol{r})]=\frac{\partial}{\partial y}\left(\mu_w\frac{\partial \boldsymbol{v}}{\partial y}\right) \tag{2}$$

水膜沿表面的流动方向为 $\boldsymbol{x}=(x_1,x_2)$。垂直方向水膜流动的边界分别是水膜-固体表面界面和空气-水膜界面。

水膜-表面界面在旋转参考系下的条件为

$$\begin{cases}\boldsymbol{v}=\boldsymbol{0}\\ y=0\end{cases} \tag{3}$$

空气-水膜界面处,由于表面水膜很薄,水膜的速度梯度由空气剪切力占主导作用,边界条件为

$$\begin{cases}\mu_w\dfrac{\partial \boldsymbol{v}}{\partial y}=\boldsymbol{\tau}\\ y=h\end{cases} \tag{4}$$

式中:$\boldsymbol{\tau}$ 为水膜所受的空气剪切力。

根据动量方程(2)和边界条件(3)、(4),得到旋转坐标系下,水膜速度沿 y 方向的分布:

$$\boldsymbol{v}(y)=\frac{\rho_w y^2}{2\mu_w}(2\boldsymbol{\omega}\cdot\boldsymbol{v}+\boldsymbol{\omega}\cdot(\boldsymbol{\omega}\cdot\boldsymbol{r}))+\frac{y}{\mu_w}[\boldsymbol{\tau}-\rho_w h(2\boldsymbol{\omega}\cdot\boldsymbol{v}+\boldsymbol{\omega}\cdot(\boldsymbol{\omega}\cdot\boldsymbol{r}))] \tag{5}$$

水膜和壁面交界处,速度为 0;水膜顶端,速度达到最大值。在法向进行积分平均化,得到水膜的相对速度为

$$\boldsymbol{v}=\frac{h}{2\mu_w}\boldsymbol{\tau}-\frac{\rho_w h^2}{3\mu_w}(2\boldsymbol{\omega}\cdot\boldsymbol{v}+\boldsymbol{\omega}\cdot(\boldsymbol{\omega}\cdot\boldsymbol{r})) \tag{6}$$

可见,水膜速度是由惯性力加速度和剪切应力共同决定的,而惯性力加速度与该速度本身有关,存在耦合关系,在求解过程中需要通过耦合迭代求解的方式予以处理。

1.2　质量守恒方程

图 2 为所有流入与流出控制体的质量,包括:控制体上表面水滴的撞击和蒸发或升华,下表面冰的生成,同时还有水的流入流出,以及控制体自身水膜厚度的变化[10]。

根据图 2 可得质量守恒方程为

$$\rho_w\left[\frac{\partial h}{\partial t}+\mathrm{div}(\boldsymbol{v}h)\right]=\dot{m}_{imp}-\dot{m}_{evap}-\dot{m}_{ice} \tag{7}$$

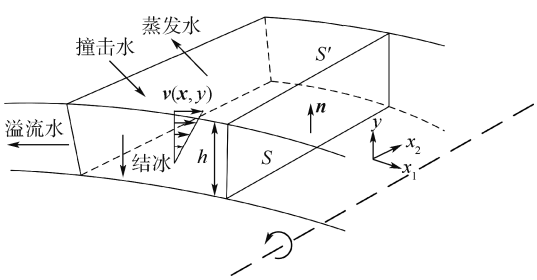


图 2　控制体质量守恒示意图

Fig. 2　Schematic diagram of mass conservation of control volume

式中:$\dot{m}_{imp}$ 为撞击到控制体表面的液态水收集率;$\dot{m}_{evap}$ 为控制体内液态水的蒸发速率;$\dot{m}_{ice}$ 为控制体内的积冰冻结速率;t 为结冰时间。

式(7)可表示为

$$\rho_w\left[\frac{\partial h}{\partial t}+\frac{\partial}{\partial x_1}(v_1 h)+\frac{\partial}{\partial x_2}(v_2 h)\right]=\dot{m}_{imp}-\dot{m}_{evap}-\dot{m}_{ice} \tag{8}$$

式中:v_1 和 v_2 分别为水膜平均速度 $\boldsymbol{v}$ 在 x_1 和 x_2 方向分量的大小。

由于非稳态计算中,是对每一个时间步长进行分别计算,需要将结冰时间分成多个时间步长,每一个时间步长作为一个稳态。因此,对质量守恒方程进行离散求解,可表示为

$$\rho_w\frac{h-h^{old}}{\Delta t}A+\sum\rho_w h(v_1\cdot\boldsymbol{n}_n)\Delta x_2+\sum\rho_w h(v_2\cdot\boldsymbol{n}_n)\Delta x_1=\dot{m}_{imp}-\dot{m}_{evap}-\dot{m}_{ice} \tag{9}$$

式中:h^{old} 为上一时间步长控制体内的水膜厚度;A 为控制体底面面积;Δt 为时间步长;$\boldsymbol{n}_n$ 为流入流出面的单位外法向量;Δx_1 和 Δx_2 分别为当前控制体的底面边长。

式(9)中,等号左边第 1 项对应的是控制体内水膜质量随时间的变化,第 2 和第 3 项表示控制体内液态水在 x_1 和 x_2 方向所有的流入流出。从而,可将每个时间步长结冰计算中的质量守恒方程表示为

$$(m_w-m_w^{old})+(m_{out}-m_{in})=m_{imp}-m_{evap}-(m_{ice}-m_{ice}^{old}) \tag{10}$$

式中:m_w^{old} 和 m_{ice}^{old} 分别为上一时间步长控制体内存在的水膜和冰层的质量;m_w 和 m_{ice} 分别为当前时间步长控制体内存在的水膜和冰层的质量;m_{in} 和 m_{out} 分别为当前时间步长流入和流出该控制体的液态水质量;m_{imp} 为当前时间步长撞击到控制体表面的液态水质量;m_{evap} 为当前时间步长控制体内液态水蒸发的质量。

当前时间步长控制体内液态水膜质量为

$$m_w=\rho_w hA \tag{11}$$

假设水滴均匀撞击在控制体表面,当前时间步长撞击到控制体表面的液态水质量为

$$m_{imp}=\mathrm{LWC}\cdot\beta v_\infty A\Delta t \tag{12}$$

式中:v_∞ 为远场来流速度;LWC 为液态水含量;β 为局部水收集系数。

当前时间步长流出该控制体的液态水质量为

$$m_{out}=\rho_w l_n h(\boldsymbol{v}\cdot\boldsymbol{n}_n)\Delta t \tag{13}$$

式中:l_n 为流出面的底边长度。

计算当前时间步长控制体内蒸发水的质量,

其中,蒸发速率可通过 Chilton-Colburn 比拟理论确定,利用对流换热系数 h_c 确定相应位置的对流传质系数 k_c,如下:

$$\frac{h_c}{c_p}Pr^{2/3} = k_c Sc^{2/3}$$

式中:Sc 为 Schimidt 数;h_c 为对流换热系数;k_c 为对流传质系数;c_p 为定压比热;Pr 为普朗特数,从而可推导蒸发速率为

$$\dot{m}_{evap} = \frac{h_c}{c_{p,air}} \cdot \frac{MW_{water}}{MW_{air}}\left(\frac{Pr}{Sc}\right)^{2/3} \cdot \left(\frac{p_{v,sat}}{p_T} \cdot \frac{T_T}{T}\left(\frac{p_T}{p}\right)^{1/\gamma} - \frac{p_{v,\infty,sat} r_h}{p_\infty}\right) \tag{14}$$

式中:$c_{p,air}$ 为空气的定压比热;MW_{water} 和 MW_{air} 分别为水和空气的分子量;$p_{v,sat}$ 为饱和蒸气压;p_T 和 T_T 分别为总压和总温;T 为当前时间步长控制体温度;p 为壁面控制体绝对压力;r_h 为相对湿度;在水膜表面该值为 1;$p_{v,\infty,sat}$ 为入口饱和蒸气压;p_∞ 为入口环境压力。

同时可定义冻结系数 f,是指在微元表面上实际结冰量和控制体内所有水之比:

$$f = \frac{m_{ice}}{m_w^{old} + m_{ice}^{old} + m_{imp} + m_{in} - m_{evap}} \tag{15}$$

1.3 能量守恒方程

图 3 为控制体内的能量守恒示意图,设流入能量为正。其中,流入控制体的能量包括水滴撞击带来的能量,从邻近控制体流入的水带来的能量和控制体内结冰释放的能量,流出控制体的能量包括对流换热、导热、热辐射带走的能量,蒸发或升华带走的能量和流出控制体的水带走的能量[7]。由于导热和热辐射对能量传递的影响较小,一般可忽略。

根据图 3 可得能量守恒方程为

$$\rho_w\left[\frac{\partial h c_{p,w} T}{\partial t} + \mathrm{div}(\boldsymbol{v} h c_{p,w} T)\right] = \dot{Q}_{imp} - \dot{Q}_{evap} - \dot{Q}_{ice} - \dot{Q}_c \tag{16}$$

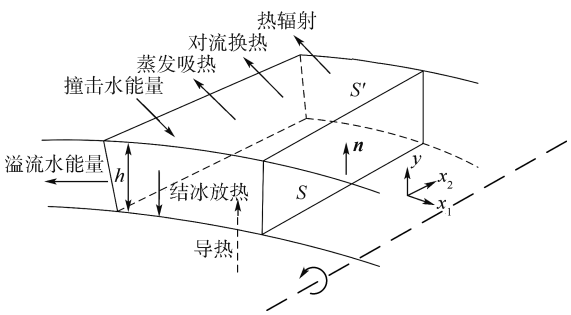


图 3 控制体能量守恒示意图

Fig. 3 Schematic diagram of energy conservation of control volume

式中:$c_{p,w}$ 为水的定压比热容;T 为当前时间步长控制体的温度;$\dot{Q}_{imp}$ 为单位时间撞击水携带的能量;$\dot{Q}_{evap}$ 为单位时间蒸发水带走的能量;$\dot{Q}_{ice}$ 为单位时间所结冰层携带的能量;$\dot{Q}_c$ 为单位时间对流换热带走的能量。

式(16)可表示为

$$\rho_w\left[\frac{\partial h c_{p,w} T}{\partial t} + \frac{\partial}{\partial x_1}(v_1 h c_{p,w} T) + \frac{\partial}{\partial x_2}(v_2 h c_{p,w} T)\right] = \dot{Q}_{imp} - \dot{Q}_{evap} - \dot{Q}_{ice} - \dot{Q}_c \tag{17}$$

参考质量守恒方程,对式(17)进行离散求解,可得

$$\rho_w c_{p,w}\frac{hT - h^{old}T^{old}}{\Delta t}A + \left[\sum \rho_w h(v_1 \cdot \boldsymbol{n}_n)\Delta x_2 T + \sum \rho_w h(v_2 \cdot \boldsymbol{n}_n)\Delta x_1 T\right] c_{p,w} = \dot{Q}_{imp} - \dot{Q}_{evap} - \dot{Q}_{ice} - \dot{Q}_c \tag{18}$$

式中:T^{old} 为上一时间步长控制体的温度;等号左边第 1 项对应的是控制体内水膜自身携带的能量随时间的变化;第 2 项表示在 x_1 和 x_2 方向控制体内所有的流入流出的液态水携带的能量。从而,可将每个时间步长结冰计算中的能量守恒方程表示为

$$(Q_w - Q_w^{old}) + (Q_{out} - Q_{in}) = Q_{imp} - Q_{evap} - (Q_{ice} - Q_{ice}^{old}) - Q_c \tag{19}$$

将式(19)各项整理,可写为

$$Q_w^{old} + Q_{ice}^{old} + Q_{imp} + Q_{in} - Q_{out} - Q_{evap} - Q_c = Q_w + Q_{ice} \tag{20}$$

式中:Q_w^{old} 和 Q_{ice}^{old} 分别为上一时间步长计算结束后控制体内存在的水膜和冰层自身携带的能量;Q_w 和 Q_{ice} 分别为当前时间步长控制体内存在的水膜和冰层自身携带的能量;Q_{imp} 为撞击水携带的能量;Q_{in} 和 Q_{out} 分别为流入和流出该控制体的水携带的能量;Q_{evap} 为水膜蒸发带走能量;Q_c 为对流换热带走的能量。

上一时间步长计算结束后控制体内存在的水膜和冰层自身携带的能量为

$$Q_w^{old} = m_w^{old} c_{p,w} T^{old} \tag{21}$$

$$Q_{ice}^{old} = m_{ice}^{old}[c_{p,ice}(T^{old} - T_0) - L_{fus} + c_{p,w} T_0] \tag{22}$$

式中:$c_{p,ice}$ 为冰的定压比热容;L_{fus} 为结冰时的固化潜热;$T_0 = 273.15$ K。

当前时间步长控制体内存在的水膜和冰层自身携带的能量为

$$Q_w = m_w c_{p,w} T \tag{23}$$

$$Q_{ice} = m_{ice}[c_{p,ice}(T - T_0) - L_{fus} + c_{p,w} T_0] \tag{24}$$

撞击水携带的能量为

$$Q_{imp} = m_{imp}\left[c_{p,w}(T_\infty - T_0) + \frac{1}{2}v_\infty^2\right] \tag{25}$$

式中：T_∞ 为远场来流温度。

流出该控制体的水携带的能量为

$$Q_{out} = m_{out} c_{p,w} T \tag{26}$$

水膜蒸发带走的能量 Q_{evap}，在不同结冰情况下采用不同的计算公式。

若没有发生结冰，即控制体内是液态水膜时：

$$Q_{evap} = [L_{evap} + c_{p,w}(T - T_0)] m_{evap} \tag{27}$$

若发生部分结冰，即结明冰情况下，控制体内温度处于相变平衡温度时：

$$Q_{evap} = L_{evap} m_{evap} \tag{28}$$

若全部结冰，即结霜冰情况下：

$$Q_{evap} = [L_{sub} + c_{p,ice}(T - T_0) - L_{fus}] m_{evap} \tag{29}$$

式中：L_{evap} 和 L_{sub} 为凝固点温度时的蒸发潜热和升华潜热。

对流换热带走的能量：

$$Q_c = h_c A (T - T_\infty) \Delta t \tag{30}$$

令 $Q = Q_{in} + Q_{imp} - Q_{evap} - Q_{out} - Q_c$，联立式(20)～式(30)，对能量守恒方程进行求解，可得到结冰冰层质量为

$$m_{ice} = \frac{m_{ice}^{old}[c_{p,ice}(T^{old} - T_0) - L_{fus} + c_{p,w} T_0]}{c_{p,ice}(T - T_0) - L_{fus} + c_{p,w} T_0} + \frac{c_{p,w} m_w^{old} T^{old} + Q}{c_{p,ice}(T - T_0) - L_{fus} + c_{p,w} T_0} \tag{31}$$

2　求解方法

通过动量方程、质量和能量守恒方程及引入的冻结系数，联立方程式(6)、式(10)、式(15)、式(31)可以对结冰数学模型进行求解。

基于 Fluent 计算得到的外部空气-水滴流场数据[22]，利用 Fluent 提供的用户自定义函数(user defined function，UDF)，求解旋转结冰数学模型，发展了一种循环遍历控制体的非稳态迭代求解方法。具体步骤如下：

步骤 1　在上一时间步水膜与冰厚的基础上(初始时刻都为 0)，对动量方程式(6)迭代求解，得到所有微元控制体内水膜的平均速度和相应的流动方向。

步骤 2　根据水膜的流动速度和方向，循环全部控制体，找到边界及驻点位置作为计算的开端。

步骤 3　从驻点处开始对控制体展开计算，假设控制体内初始温度为 0℃，并全部结冰，计算控制体内的质量和能量守恒方程，从而得到冻结系数 f，对冻结系数进行判断，得到其结冰类型(明冰、霜冰、不结冰)，根据结冰类型更新冻结系数、温度、结冰量、溢流水量、水膜厚度等。利用更新的结果迭代计算动量方程、质量和能量守恒方程，直至收敛，该控制体计算完成，将当前控制体溢流水流出量作为输入值，赋值给临近控制体，邻近控制体流入的溢流水质量可知。

步骤 4　继续计算邻近控制体，直至全部控制体计算完成，并利用更新后的水膜厚度，对全部控制体内水膜速度进行更新，返回步骤 1 开始计算，直至完成收敛判断，则当前时间步长计算完成，输出当前时间步长冰形。

步骤 5　进入下一时间步长结冰计算，直至所设定结冰时间全部计算完成。

上述步骤的具体流程如图 4 所示。

其中步骤 3 和 4 为结冰计算关键，对其进行进一步详细说明：

1) 假设控制体初始结冰状态。假设初始控制体温度为 0℃，控制体内结明冰，且水全部发生冻结，可知 $m_{out}=0$，$m_w=0$，$Q_{out}=0$，$Q_w=0$，且 m_w^{old} 和 m_{ice}^{old} 对应上一时间步长的液态水量和结冰量，为已知量。

通过上述假设可以对式(31)求解，得到结冰量，从而可根据式(15)得到冻结系数 f。

2) 对冻结系数 f 进行判断。

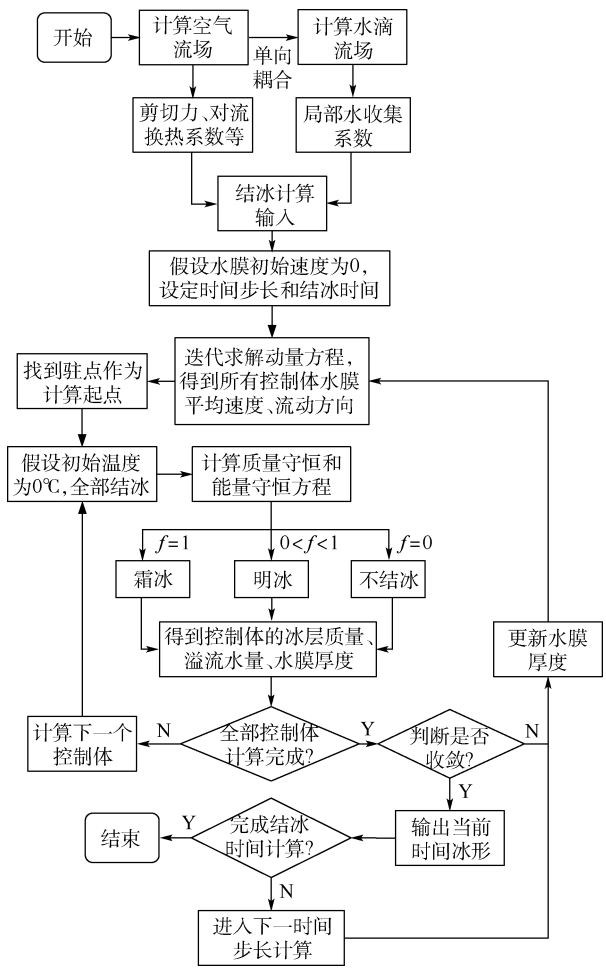


图 4　旋转表面结冰求解流程

Fig. 4　Flow chart for solving icing on rotating surfaces

① 当 $f \geqslant 1$ 时,说明此时控制体结霜冰,控制体内水全部发生冻结,没有溢流水流出,控制体内不存在液态水,令 $f=1$,则可得

$$\begin{cases} m_{\text{out}} = 0 \\ Q_{\text{out}} = 0 \\ m_{\text{w}} = 0 \\ Q_{\text{w}} = 0 \\ m_{\text{ice}} = m_{\text{w}}^{\text{old}} + m_{\text{ice}}^{\text{old}} + m_{\text{imp}} + m_{\text{in}} - m_{\text{evap}} \\ Q_{\text{ice}} = m_{\text{ice}}[c_{\text{p,ice}}(T - T_0) - L_{\text{fus}} + c_{\text{p,w}} T_0] \end{cases} \tag{32}$$

② 当 $0 < f < 1$ 时,说明此时控制体结明冰,部分水发生冻结,控制体温度处于相变平衡温度。控制体内结冰量由式(31)可知,且联立式(10)、式(11)、式(13)可得

$$m_{\text{out}} + m_{\text{w}} = m_{\text{w}}^{\text{old}} + m_{\text{ice}}^{\text{old}} + m_{\text{imp}} + m_{\text{in}} - m_{\text{evap}} - m_{\text{ice}} \tag{33}$$

$$m_{\text{out}} + m_{\text{w}} = \rho_{\text{w}} l_n h(\boldsymbol{v} \cdot \boldsymbol{n}_n) \Delta t + \rho_{\text{w}} h A \tag{34}$$

联立式(33)和式(34)可得水膜厚度为

$$h = \frac{m_{\text{w}}^{\text{old}} + m_{\text{ice}}^{\text{old}} + m_{\text{imp}} + m_{\text{in}} - m_{\text{evap}} - m_{\text{ice}}}{\rho_{\text{w}} l_n (\boldsymbol{v} \cdot \boldsymbol{n}_n) \Delta t + \rho_{\text{w}} A} \tag{35}$$

观察式(6)、式(35)可知,水膜流动速度和水膜厚度相互影响,相互耦合,无法单独求解,需要联立求解。通过 Jacobi 迭代法和 Gauss-Seidel 迭代法,在对单一控制体迭代计算时,将该控制体上一迭代步骤计算结果中的水膜厚度和水膜速度带入当前迭代步骤,循环迭代求解式(6)得到水膜速度用于判定流入流出水量和方向,再通过式(35)求出当前迭代步骤的水膜厚度,与水膜速度一起用于下一迭代过程,直至水膜流动速度和水膜厚度达到收敛,当前控制体计算完成,可用于结冰计算。

在全部控制体结冰计算完成后,利用更新后的水膜厚度和水膜速度,从步骤 1 开始重新判定流动速度和流动方向,再次对所有控制体进行结冰计算,直至前后 2 次计算误差达到一定程度,完成收敛。通过计算单一控制体内收敛和全部控制体收敛,在个体-整体双重收敛的作用下,完成对水膜流动速度和水膜厚度的计算。

通过迭代计算得到水膜厚度和水膜流动速度后,即可得明冰状态下的结冰参数:

$$\begin{cases} m_{\text{out}} = \rho_{\text{w}} l_n h(\boldsymbol{v} \cdot \boldsymbol{n}_n) \Delta t \\ Q_{\text{out}} = m_{\text{out}} c_{\text{p,w}} T \\ m_{\text{w}} = \rho_{\text{w}} h A \\ Q_{\text{w}} = m_{\text{w}} c_{\text{p,w}} T \\ m_{\text{ice}} = m_{\text{ice}} \\ Q_{\text{ice}} = m_{\text{ice}}[c_{\text{p,ice}}(T - T_0) - L_{\text{fus}} + c_{\text{p,w}} T_0] \end{cases} \tag{36}$$

③ 当 $f=0$ 时,说明此时控制体内不结冰,只存在液态水膜,即 $m_{\text{ice}}=0$,且由式(35)可知水膜厚度为

$$h = \frac{m_{\text{w}}^{\text{old}} + m_{\text{ice}}^{\text{old}} + m_{\text{imp}} + m_{\text{in}} - m_{\text{evap}}}{\rho_{\text{w}} l_n (\boldsymbol{v} \cdot \boldsymbol{n}_n) \Delta t + \rho_{\text{w}} A} \tag{37}$$

联立式(6)和式(37),采用明冰计算时的迭代求解方法,可得水膜厚度和流动速度,从而得到液态水状态下相应的结冰参数:

$$\begin{cases} m_{\text{out}} = \rho_{\text{w}} l_n h(\boldsymbol{v} \cdot \boldsymbol{n}_n) \Delta t \\ Q_{\text{out}} = m_{\text{out}} c_{\text{p,w}} T \\ m_{\text{w}} = \rho_{\text{w}} h A \\ Q_{\text{w}} = m_{\text{w}} c_{\text{p,w}} T \\ m_{\text{ice}} = 0 \\ Q_{\text{ice}} = 0 \end{cases} \tag{38}$$

3) 通过对 f 的判断可以得到控制体内的结冰状态,从而计算得到当前时间步长控制体的溢流水流出量,水膜厚度和结冰量,然后将当前控制体溢流水流出量作为输入值,赋值给临近控制体。

4) 进行下一控制体计算,直至全部控制体计算完成,并利用更新后的水膜厚度,对控制体内水膜速度进行更新,返回步骤 1 重新计算,直至完成收敛,则当前时间步长计算完成。

3 仿真计算分析

针对飞机上的所有旋转部件结冰进行研究,分析旋转部件的结冰特性,为便于仿真模型扩展到所有旋转部件,参考文献[17]的处理方式,将简化后的等截面圆柱作为结冰壁面,其直径为 60 mm,长度为 200 mm,对称安置于直径 600 mm 的转轴上,来流空气沿转轴轴向方向,如图 5 所示。

将入口来流边界设为速度进口,出口边界设为压力出口,结冰表面和转轴均设为旋转壁面,计算条件为:压力 101 325 Pa,迎角 0°,结冰温度 263.15 K,来流速度 60 m/s,转速 30 rad/s,液态水含量(LWC)0.8 g/m^3,水滴直径 20 μm,结冰时间 420 s。取时间步长 0.02 s 与 0.05 s 进行验证

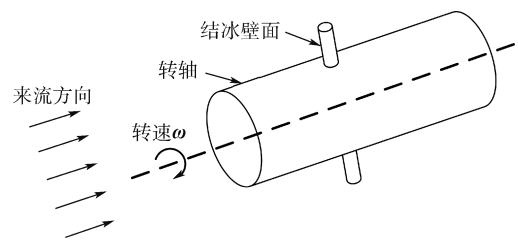


图 5 旋转圆柱模型示意图

Fig. 5 Schematic diagram of rotating cylinder model

计算，对比结果发现最终结冰冰形和水膜相近，本文取时间步长 0.05 s 进行计算，所得圆柱三维结冰冰形如图 6 所示，圆柱中间截面上的二维冰形和水膜厚度随时间变化如图 7 和图 8 所示。

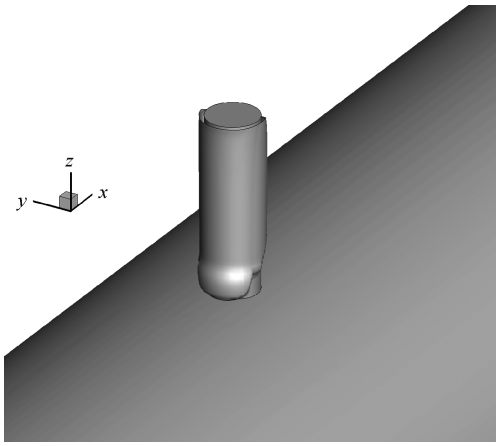


图 6　旋转圆柱三维冰形

Fig. 6　Three-dimensional ice shape of rotating cylinder

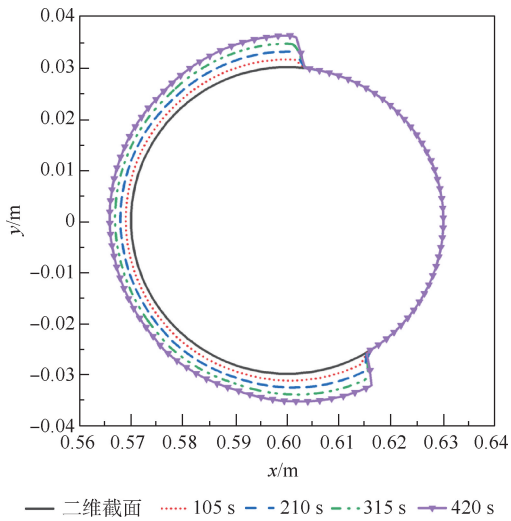


图 7　结冰冰形随时间变化

Fig. 7　Ice shape changing over time

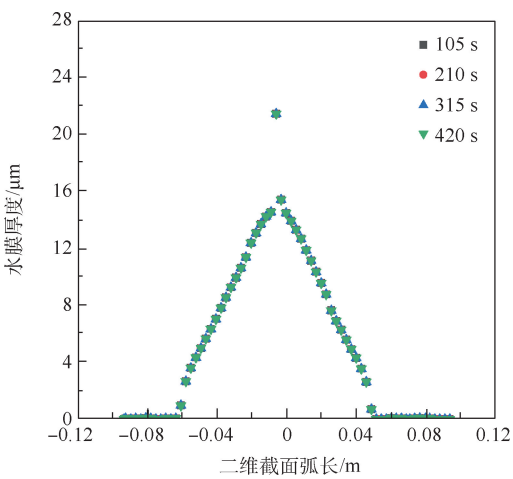


图 8　水膜厚度随时间变化

Fig. 8　Water film thickness changing over time

由图 7 和图 8 可知，随着结冰时间的增长，不断有撞击水撞击到结冰壁面上，结冰表面水收集量逐渐增多，导致圆柱表面结冰厚度逐渐增加，当结冰环境温度不变时，结冰能力基本不变，只有厚度随时间发生明显变化；且随着时间的增长，水膜厚度和覆盖范围变化很小，水膜达到一定结冰时间后基本不变。同时由于旋转的作用，圆柱表面上的结冰区域和水膜覆盖区域均向圆柱下表面（$-y$ 方向）发生偏移，下表面的结冰区域和水膜覆盖范围大于上表面。

水膜厚度虽然整体变化趋势符合预期，但是在驻点处数值偏大，通过对数据进行分析，推测由于驻点处速度趋近于 0，水膜在驻点处基本不发生溢流，液态水在该点产生积聚，导致驻点处水膜厚度偏大。可在之后的研究中加入压力的影响，当剪切力趋于 0 时，将压差作为驻点处水膜流动的主要驱动力，对模型算法进一步改进。

由于旋转表面结冰实验研究匮乏，缺乏相关实验数据用以验证数值模拟的准确性，因此为了验证所提模型的可靠性，将该模型算法的结冰计算结果与 FENSAP 软件获得的结冰冰形进行对比，来验证所提模型的合理性。在转速为 30 rad/s，结冰时间 420 s 的工况下开展计算，将最终冰形结果进行对比，如图 9 和图 10 所示。

观察图 9 和图 10 的对比可知，两者的结冰厚度和结冰范围基本相同，其中本文所搭建的结冰模型的结冰范围略微偏小；两者的冰形变化趋势也基本一致，在旋转作用下同样向圆柱下表面发生了偏移；水膜厚度变化曲线基本吻合，只有在驻点处由于速度趋近于 0，使得结冰模型计算精度不够，水膜厚度数值偏大，其余点处吻合良好。本

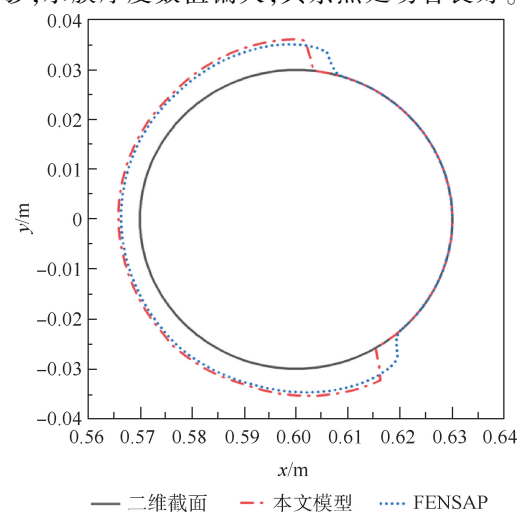


图 9　本文模型和 FENSAP 冰形结果对比

Fig. 9　Comparison of ice shape between the proposed model and FENSAP

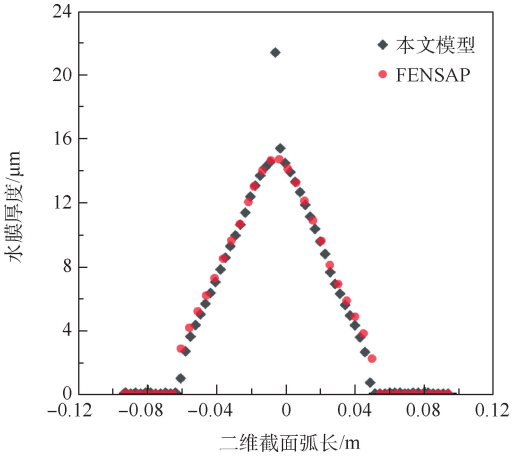


图 10 本文模型和 FENSAP 水膜厚度对比

Fig. 10 Comparison of water film thickness between the proposed model and FENSAP

文模型与 FENSAP 计算结果对比分析,证明本文模型的结冰冰形和水膜溢流计算结果合理,所提结冰模型求解方法正确。

通过改变转速,研究转速对旋转圆柱表面结冰和水膜流动的影响,将转速分别设为 0,30,60 rad/s,其二维中间截面的结冰冰形和水膜厚度如图 11 和图 12 所示。

根据图 11 和图 12 比较 3 种不同转速结果可知,随着转速的增加,水膜受到的惯性力和空气剪切力等逐渐增加,圆柱表面液态水的溢流能力逐渐增强,结冰范围和水膜覆盖范围向圆柱下表面(-y 方向)出现明显偏移,转速越大,偏移现象愈加明显。且随着转速的增加,冰层厚度基本不变,水膜厚度略有减小。

通过改变撞击水滴的直径,研究其对旋转圆柱表面结冰和水膜流动的影响,撞击水滴直径分别设为 20,25,30 μm,其二维中间截面的结冰冰形和水膜厚度如图 13 和图 14 所示。

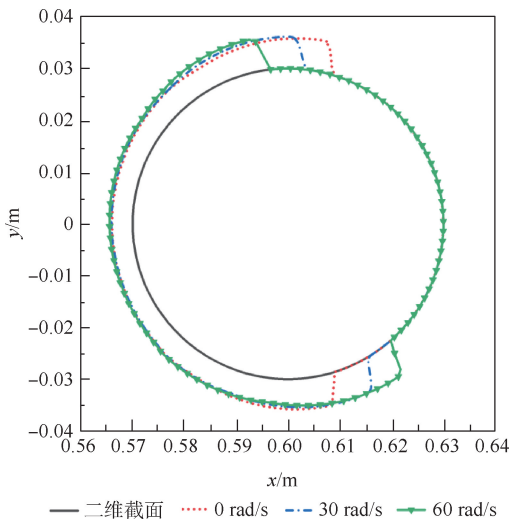


图 11 结冰冰形随转速变化

Fig. 11 Ice shape changing with speed

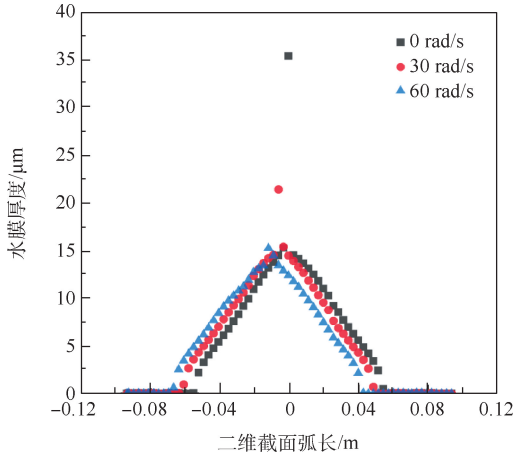


图 12 水膜厚度随转速变化

Fig. 12 Water film thickness changing with speed

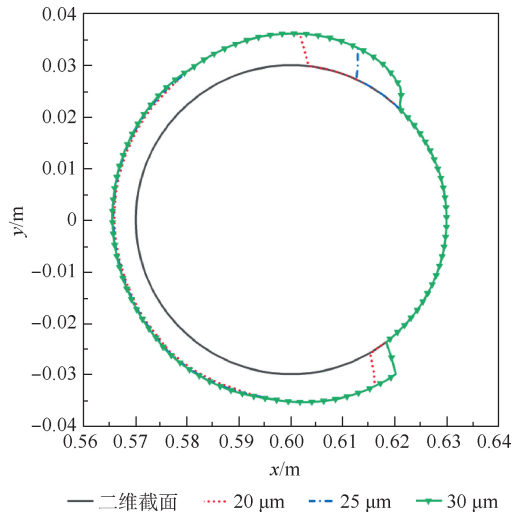


图 13 结冰冰形随水滴直径变化

Fig. 13 Ice shape changing with droplet diameter

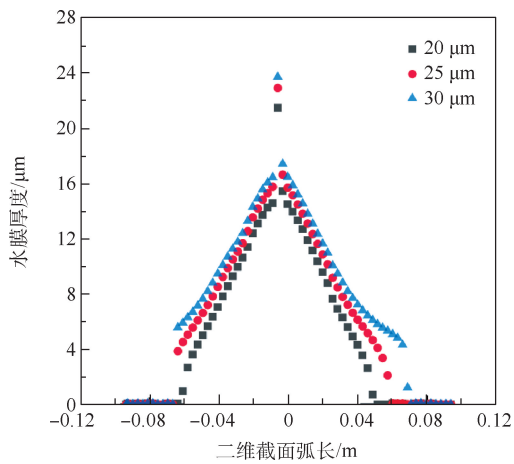


图 14 水膜厚度随水滴直径变化

Fig. 14 Water film thickness changing with droplet diameter

通过图 13 和图 14 比较 3 种不同撞击水滴直径的计算结果可知,随着水滴直径的增加,水滴惯性更大,不易受到气流的作用而绕过迎风表面,更容易和圆柱表面发生碰撞,水滴撞击范围和撞击量相应增加,同时由于当前环境结冰能力恒定,冰

层厚度随水滴直径变化较小,未凝结的液态水相应增多,会在空气剪切力和离心力等的作用下发生溢流,因此结冰范围和水膜覆盖范围随水滴直径增长而逐渐增加,水膜厚度也逐渐增大。

通过改变液态水含量,研究其对旋转圆柱表面结冰和水膜流动的影响,液态水含量分别设为 0.6,0.8,1.0 g/m^3,其二维中间截面的结冰冰形和水膜厚度如图 15 和图 16 所示。

由图 15 和图 16 比较 3 种不同液态水含量的仿真计算结果可知,随着液态水含量的增长,相同时间撞击到迎风表面的水滴增多,导致圆柱表面的液态水逐渐增多,同时由于当前结冰环境的结冰能力恒定,结冰厚度只发生了小幅度增长,可溢流的未凝结液态水逐渐增多,水膜厚度也相应增加,在空气剪切力和离心力等的作用下,未凝结水膜发生溢流,其覆盖范围明显变大,结冰范围也明显增加。

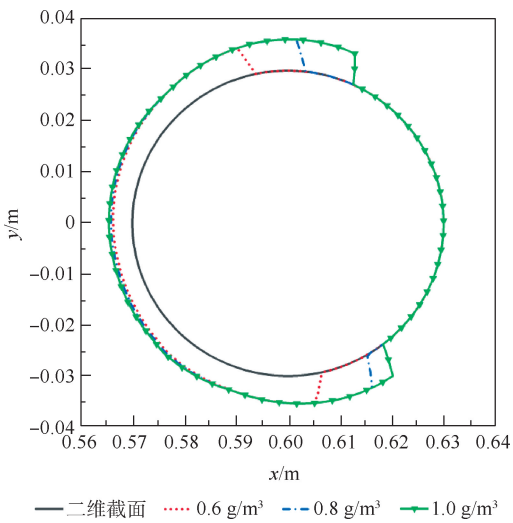


图 15　结冰冰形随液态水含量变化

Fig. 15　Ice shape changing with liquid water content

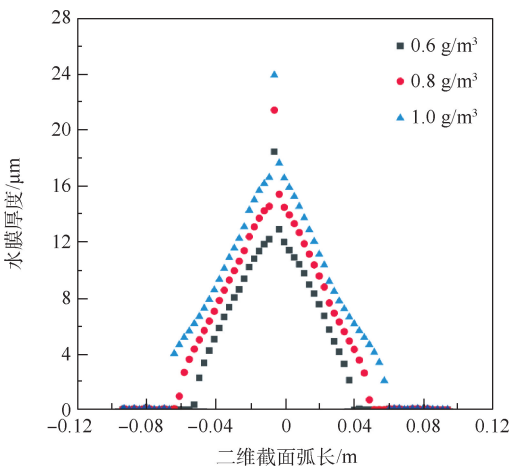


图 16　水膜厚度随液态水含量变化

Fig. 16　Water film thickness changing with liquid water content

4　结　论

考虑旋转部件表面的水膜流动受到空气剪切力与惯性力影响,建立了旋转表面的结冰热力学模型,通过个体-整体双重收敛的方式进行非稳态迭代求解。以旋转圆柱表面为研究对象,分析各影响因素对结冰和水膜流动的作用,得出以下结论。

1) 所提模型的非稳态求解方法经过验证,可以有效对三维旋转部件表面的结冰冰形和水膜流动进行模拟,在驻点处水膜计算尚存在不足,可在之后的研究中对模型算法进一步改进。

2) 由于旋转的作用,圆柱表面的结冰区域和水膜覆盖范围会发生偏移,且随着结冰时间的增长,旋转圆柱表面的结冰厚度也随之增加,其结冰区域和水膜溢流受结冰时间影响较小。

3) 随着转速的增加,旋转圆柱表面结冰区域和水膜覆盖范围的偏移现象愈加明显,冰层厚度基本不变,水膜厚度略微减小。

4) 随着水滴直径的增加,旋转圆柱表面的冰层厚度变化较小,其结冰区域和水膜覆盖范围随水滴直径增长逐渐增加,水膜厚度也逐渐增大。

5) 随着液态水含量的增长,旋转圆柱表面的结冰厚度出现小幅增长,水膜厚度相应增加,结冰范围和水膜覆盖范围也明显变大。

参考文献 (References)

[1] CAO Y H, TAN W Y, WU Z L. Aircraft icing: An ongoing threat to aviation safety[J]. Aerospace Science and Technology, 2018, 75(4): 353-385.

[2] SHEN X B, LIN G P, YU J, et al. Three-dimensional numerical simulation of ice accretion at the engine inlet[J]. Journal of Aircraft, 2013, 50(2): 635-642.

[3] DU Y X, GUI Y W, XIAO C H, et al. Investigation on heat transfer characteristics of aircraft icing including runback water [J]. International Journal of Heat and Mass Transfer, 2010, 53(19-20): 3702-3707.

[4] MESSINGER B L. Equilibrium temperature of an unheated icing surface as a function of air speed[J]. Journal of the Aeronautical Sciences, 1953, 20(1): 29-42.

[5] GHENAI C, KULKARNI S, LIN C X. Validation of LEWICE 2.2 icing software code: Comparison with LEWICE 2.0 and experimental data[C]//43rd AIAA Aerospace Sciences Meeting and Exhibit. Reston: AIAA, 2005: 1249.

[6] MYERS T G. Extension to the messinger model for aircraft icing [J]. AIAA Journal, 2001, 39(2): 211-218.

[7] BOURGAULT Y, BEAUGENDRE H, HABASHI W G. Development of a Shallow-Water icing model in FENSAP-ICE[J]. Journal of Aircraft, 2000, 37(4): 640-646.

[8] 易贤,桂业伟,朱国林. 飞机三维结冰模型及其数值求解方法[J]. 航空学报,2010,31(11):2152-2158.
YI X,GUI Y W,ZHU G L. Numerical method of a three-dimensional ice accretion model of aircraft[J]. Acta Aeronautica et Astronautica Sinica,2010,31(11):2152-2158(in Chinese).

[9] 申晓斌,林贵平,卜雪琴,等. 发动机进气道短舱前缘结冰三维模拟研究[J]. 航空学报,2013,34(3):517-524.
SHEN X B,LIN G P,BU X Q,et al. Three-dimensional simulation research on ice shape at engine inlet nacelle front[J]. Acta Aeronautica et Astronautica Sinica,2013,34(3):517-524(in Chinese).

[10] 曹广州,吉洪湖,胡娅萍,等. 模拟飞机迎风面三维积冰的数学模型[J]. 航空动力学报,2011,26(9):1953-1963.
CAO G Z,JI H H,HU Y P,et al. An icing model for simulating three dimensional ice accretion on the upwind surfaces of a plane[J]. Journal of Aerospace Power,2011,26(9):1953-1963(in Chinese).

[11] CAO Y H,HUANG J S. New method for direct numerical simulation of three-dimensional ice accretion[J]. Journal of Aircraft,2014,52(2):650-659.

[12] 雷梦龙,常士楠,杨波. 基于Myers模型的三维结冰数值仿真[J]. 航空学报,2018,39(9):121952-121962.
LEI M L,CHANG S N,YANG B. Three-dimensional numerical simulation of icing using Myers model[J]. Acta Aeronautica et Astronautica Sinica,2018,39(9):121952-121962(in Chinese).

[13] BAIN J,CAJIGAS J,SANKAR L,et al. Prediction of rotor blade ice shedding using empirical methods[C]// AIAA Atmospheric and Space Environments Conference. Reston:AIAA,2010:7985.

[14] REID T,BARUZZI G,OZCER I,et al. FENSAP-ICE simulation of icing on wind turbine blades,Part 1:Performance degradation[C]// 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. Reston:AIAA,2013:750.

[15] REID T,BARUZZI G,OZCER I,et al. FENSAP-ICE simulation of icing on wind turbine blades,Part 2:Ice protection system design[C]// 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition. Reston:AIAA,2013:751.

[16] SWITCHENKO D,HABASHI W,REID T,et al. FENSAP-ICE simulation of complex wind turbine icing events,and comparison to observed performance data[C]// 32nd ASME Wind Energy Symposium,2014:1399.

[17] WANG Z Z,ZHU C L. Numerical simulation of three-dimensional rotor icing in hovering flight[J]. Proceedings of the Institution of Mechanical Engineers,Part G:Journal of Aerospace Engineering,2018,232(3):545-555.

[18] DONG W,ZHU J J,ZHENG M,et al. Numerical study of ice accretion on rotating aero-engine cone[C]// 29th Congress of the International Council of the Aeronautical Sciences,2014.

[19] 赵秋月,董威,朱剑鋆. 发动机旋转整流帽罩的水滴撞击特性分析[J]. 燃气涡轮试验与研究,2011,24(4):32-35.
ZHAO Q Y,DONG W,ZHU J J. Droplets impinging characteristic analysis of the rotating fairing of aero-engine[J]. Gas Turbine Experiment and Research,2011,24(4):32-35(in Chinese).

[20] 吴孟龙,常士楠,冷梦尧,等. 基于欧拉法模拟旋转帽罩水滴撞击特性[J]. 北京航空航天大学学报,2014,40(9):1263-1267.
WU M L,CHANG S N,LENG M Y,et al. Simulation of droplet impingement characteristics of spinner based on Eulerian method[J]. Journal of Beijing University of Aeronautics and Astronautics,2014,40(9):1263-1267(in Chinese).

[21] ROTHMAYER A,TSAO J. Water film runback on an airfoil surface[C]// 38th Aerospace Sciences Meeting and Exhibit. Reston:AIAA,2000:237.

[22] 申晓斌,张志强,林贵平,等. 旋转部件复杂表面水滴撞击计算[J]. 空气动力学学报,2016,34(6):709-713.
SHEN X B,ZHANG Z Q,LIN G P,et al. Droplet impingement calculation on complex suface of rotating part[J]. Acta Aerodynamica Sinica,2016,34(6):709-713(in Chinese).

Numerical simulation of icing on aircraft rotating surfaces

GUO Qi[1], SHEN Xiaobin[1,2,*], LIN Guiping[1], ZHANG Shijuan[3]

(1. School of Aeronautic Science and Engineering, Beihang University, Beijing 100191, China;
2. Key Laboratory of Icing and Anti/De-icing, China Aerodynamics Research and Development Center, Mianyang 621000, China;
3. Wuhan Aviation Instrument Corporation, Wuhan 430074, China)

Abstract: Based on Shallow-Water icing thermodynamic theory, an unsteady icing model is developed to simulate icing on three-dimensional rotating surfaces. Jacobi and Gauss-Seidel iterative algorithms are adopted to numerically solve the governing equations of unsteady processes of surface icing. This method is used to calculate the simplified rotating blade model, and the results are compared with those of FENSAP, verifying the accuracy of the model. Then the effect of the rotation speed, droplet diameter and liquid water content on the rotating surface water film flow and icing shape is investigated. Results show that as the rotation speed increases, the deviation of the icing range and the water film coverage becomes more obvious. With the increase of the droplet diameter, the icing range and the water film coverage and thickness gradually increase. The icing and water film thicknesses increase as a result of the increase of the liquid water content, and the water film coverage also increases significantly.

Keywords: aircraft icing; rotation; water film flow; unsteady-state; Shallow-Water model

Received: 2021-02-19; **Accepted**: 2021-05-07; **Published online**: 2021-05-12 10:19

URL: kns.cnki.net/kcms/detail/11.2625.V.20210512.0828.001.html

Foundation items: National Natural Science Foundation of China (51806008); Funding for the Open Project of the Key Laboratory of Icing and Anti/De-icing (IADL20200307)

* **Corresponding author**. E-mail: shenxiaobin@buaa.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0084

基于凸优化和 LQR 的火箭返回轨迹跟踪制导

吴杰[1], 张成[1], 李淼[2], 熊芬芬[1,*]

(1. 北京理工大学 宇航学院, 北京 100081; 2. 北京特种机电研究所, 北京 100012)

摘 要: 可重复使用运载火箭动力减速段制导,面临各种苛刻的过程约束、终端约束及燃料最省的迫切需求,给制导带来巨大挑战。因此,提出一种基于分段凸优化和线性二次调节器(LQR)的轨迹跟踪制导律。采用分段凸优化方法对火箭基准速度进行跟踪,大幅简化了优化模型从而降低凸优化求解的计算量,同时确保火箭在各种初始误差和模型误差的情况下燃料最省。采用 LQR 方法实现对火箭飞行位置轨迹的高精度跟踪,抵抗各种误差和干扰的影响。仿真结果表明:相对于传统的 LQR 跟踪制导方法,所提方法能大幅减少燃料消耗,且在各种误差和干扰下具有较高的轨迹跟踪精度和较强的抗干扰能力;相比于现有的滚动凸优化方法,所提方法能显著降低求解计算量,且方法可靠性更高。

关 键 词: 可重复使用运载火箭; 动力减速段; 轨迹跟踪制导; 线性二次调节器; 凸优化
中图分类号: V448; TJ765
文献标志码: A **文章编号**: 1001-5965(2022)11-2270-11

可重复使用运载火箭可大幅降低单次发射成本、提高火箭发射密度、降低残骸的危害[1-2],是未来运载技术发展的必然趋势。近年来,随着以美国猎鹰 9(Falcon-9)为代表的可重复使用运载火箭的多次成功发射、回收和重复使用,掀起了世界各国对可重复使用运载器的研究热潮。

火箭返回主要包括调姿段、修航段、高空下降段、动力减速段、气动减速段和垂直着陆段[1]。对于调姿段和高空下降段,火箭处于稠密大气层外且发动机关机,通常火箭无制导飞行能力;对于修航段,火箭主要任务是调整航程,飞行轨迹也较为固定[3]。而本文所研究的动力减速段,是指运载器进入大气层后,利用发动机反向点火进行制动减速的飞行阶段。动力减速飞行过程中,要求运载器以一定的速度飞行到预定区域,运载器面临严峻的热流约束、过载约束和动压约束等过程约束;同时运载器动力学具有非线性和高动态等技术特征;此外,为提高可重复使用火箭的运载能力,还要求其再入过程中使用燃料尽可能少。这些都对动力减速段轨迹规划与制导提出了巨大挑战。

标称轨迹跟踪制导方法广泛应用于航天器再入制导,其根据返回器在飞行过程中的端点约束和过程约束形成一条标称轨迹,然后进行在线轨迹跟踪制导。标称轨迹跟踪方法早期主要采用 PID 控制方法,Harpold 和 Graves[4] 以航天飞机再入返回为应用背景,纵向制导采用 PID 控制,根据实际飞行状态与标准阻力加速度-速度剖面之间的偏差生成控制指令,但该方法依赖于标称轨迹,不能适应较大的初始误差。近年来,基于现代控制理论的标称轨迹跟踪制导设计方法得到广泛研究。Roenneke 和 Cornwell[5] 针对低升阻比再入飞行器使用基于线性反馈的方法来跟踪轨迹。Dukeman[6] 针对航天飞机再入段,采用线性二次

收稿日期: 2021-02-19; **录用日期**: 2021-04-18; **网络出版时间**: 2021-05-07 09:22
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210506.1717.003.html
基金项目: 十三五装发预研 (202020201320)
* **通信作者**. E-mail: fenfenx@bit.edu.cn

引用格式: 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).

调节器(linear quadratic regulator,LQR)方法进行闭环反馈控制,实现对标称轨迹的跟踪,取得了良好的跟踪性能。张大元等[7]针对防空导弹弹道跟踪问题,利用 LQR 理论设计弹道跟踪制导律并验证存在初始误差和风干扰下的跟踪效果。刘晓东等[8]、宋晨等[9]研究了自适应动态面标称轨迹跟踪方法,该方法同样具有良好的跟踪性能。但是,这些方法一般都是以对轨迹进行高精度跟踪,即位置跟踪偏差最小作为最优控制目标,无法考虑燃料最省。而且,即便在基准轨迹的设计过程中考虑了燃料最省,但是在火箭返回过程中面临着各种复杂干扰和模型的强不确定性,轨迹跟踪中必然需要消耗额外的燃料。因此,需要在轨迹跟踪中对燃料消耗进行重规划,以期达到尽可能减少燃料消耗的目的。此外,基于 LQR 的轨迹跟踪方法需要提供精确的基准控制输入才能保证较好的跟踪效果,而这在实际中很多情况下难以保证。

近年来,Acikmese 和 Blackmore 等[10-12]成功采用凸优化方法求解火星软着陆制导问题,引发了基于优化的制导算法的技术创新。由于凸优化具有计算效率高的特点,且能有效处理过程、终端等各类约束,在飞行器轨迹优化和制导中得到越来越多的应用。Liu 和 Lu[13]提出序列凸化方法将非线性最优控制问题转化为二阶锥规划问题,并成功应用于航天器交会对接[14]等问题。Wang 和 Grant[15]同样采用序列凸优化算法解决高超声速飞行器再入问题,该方法采取离线轨迹优化,无法实现在线制导。路钊[16]将序列凸优化与滚动时域控制结合,形成了所谓的滚动序列凸优化方法,实现了在线轨迹跟踪制导。序列凸优化方法由于不需要对最优控制问题进行前期繁杂的模型变换、凸化处理,通用性较强,是目前求解非凸最优控制问题的主要方法。但是,该方法的最优解较为依赖模型的具体形式,当动力学系统涉及复杂非线性气动力、状态量变量维数较高时,存在迭代次数多甚至难以收敛的问题,由于计算量较大、可靠性较差,难以适应当前箭载计算机的计算能力,无法满足在线制导的要求[17]。

综上可知,基于滚动序列凸优化方法的轨迹跟踪方法由于集成了凸优化方法,显然能方便地实现燃料最省的要求,但是极有可能存在计算量过大、可靠性无法保证,而无法满足在线制导要求的问题。为了提高滚动凸优化方法的计算效率和求解可靠性,同时追求燃料最省,本文提出一种分段序列凸优化和 LQR 的轨迹跟踪方法;相比于传统的 LQR 跟踪制导方法,能有效减少燃料消耗,且可保证具有较高的轨迹跟踪精度和较强的抗干扰能力;相比于现有的滚动凸优化方法,可显著降低计算量、提高方法可靠性。本文提出的轨迹跟踪方法充分利用分段凸优化方法能进行在线优化的优势以实现燃料最省,同时利用 LQR 具有良好轨迹跟踪性能的优势实现高精度的轨迹跟踪,且无需提供精确的基准推力控制输入,实际使用更加灵活。通过将火箭的运动分解为包含速度和位置相关的两部分,在此基础上利用分段凸优化方法实现火箭的速度跟踪,同时达到燃料最省的目的;利用 LQR 实现火箭的飞行轨迹跟踪。对于速度跟踪,由于仅包含速度相关的状态量,动力学模型维度降低,凸优化方法求解的效率可大为提升。同时,为进一步降低凸优化方法计算量,在每一个轨迹分段仅求解当前轨迹分段内的最优控制问题,并非从当前分段到飞行终端的整个时域。

1　火箭运动方程

考虑到动力减速段火箭始终保持尾部向前飞行,速度方向不发生改变,将火箭箭体坐标系定义为 x 轴指向箭体尾部,火箭所受的空气动力均建立在此坐标系中。火箭纵向平面几何关系如图 1 所示,其中 O' 为火箭质心位置。

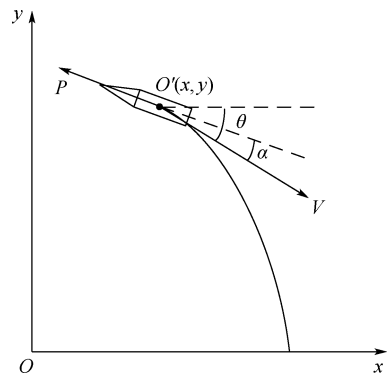


图 1　火箭纵向平面几何关系

Fig. 1　Geometric diagram of rockets in longitudinal plane

运载火箭动力减速段飞行航程一般在 100 km 左右,因此可以忽略地球自转;火箭飞行高度一般在 20 ~ 70 km 之间,重力加速度可简化为常值。将火箭质心运动方程建立在发射坐标系内,纵向平面内的火箭再入运动方程如下:

$$\begin{cases}\dfrac{\mathrm{d}V}{\mathrm{d}t} = -\dfrac{P\cos\alpha + D}{m} - g\sin\theta \\ \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \dfrac{-P\sin\alpha + Y}{mV} - \dfrac{g\cos\theta}{V} \\ \dfrac{\mathrm{d}x}{\mathrm{d}t} = V\cos\theta \\ \dfrac{\mathrm{d}y}{\mathrm{d}t} = V\sin\theta \\ \dfrac{\mathrm{d}m}{\mathrm{d}t} = -\dfrac{P}{I_{\mathrm{sp}}}\end{cases} \tag{1}$$

式中:V为火箭飞行速度;θ为速度倾角;x、y分别为火箭飞行位置的横、纵坐标;m为火箭质量;D为火箭所受阻力;Y为升力;I_{sp}为发动机比冲;α为攻角;P为火箭的推力;g为重力加速度。D和Y分别表示为

$$\begin{cases} D = \dfrac{1}{2}\rho V^2 S_{ref} C_a \\ Y = \dfrac{1}{2}\rho V^2 S_{ref} C_n^{\alpha} \alpha \end{cases} \tag{2}$$

式中:S_{ref}为参考面积;C_a为阻力系数;C_n^{α}为升力系数关于攻角α的导数,本文将气动升力视作攻角α的线性函数;ρ为大气密度,可以表示为

$$\rho = \rho_0 \exp\left(\frac{-y}{h}\right) \tag{3}$$

式中:$\rho_0 = 1.225\ \mathrm{kg/m^3}$为海拔为0时的空气密度;$h = 7\ 254.3\ \mathrm{m}$。

轨迹跟踪的核心是对位置的精确跟踪,而对实际飞行时间和标称轨迹对应的飞行时间之间的差异不作要求,因此无需一定采用时间为自变量。如果取火箭的航程或高度为自变量,不仅可保证对位置的精确跟踪,而且在凸优化方法的应用中无需优化飞行时间(火箭飞行至预定点的时间无法提前确定),从而可降低系统状态变量的维数,提高计算效率和可靠性。因此,在式(1)的基础上,选择火箭位置x坐标为自变量,得到新的火箭运动方程如下:

$$\begin{cases} \dfrac{dV}{dx} = -\dfrac{P\cos\alpha + D}{mV\cos\theta} - \dfrac{g\sin\theta}{V\cos\theta} \\ \dfrac{d\theta}{dx} = \dfrac{-P\sin\alpha + Y}{mV^2\cos\theta} - \dfrac{g}{V^2} \\ \dfrac{dy}{dx} = \tan\theta \\ \dfrac{dm}{dx} = -\dfrac{P}{I_{sp} V\cos\theta} \end{cases} \tag{4}$$

式中:关于速度倾角θ的有些项出现在分母中,当$\theta = 90°$时会出现奇异的情况。然而,可重复使用运载火箭返回过程在动力减速段之后还有气动减速段和垂直着陆段,因此在动力减速段速度倾角一般不会大于70°,该模型可以应用于火箭动力减速段。

2 轨迹跟踪制导律设计

2.1 轨迹跟踪制导策略

获取一条基准轨迹,获取基准轨迹的方法不限,但要保证基准轨迹满足过程约束、终端约束,这样才能让火箭顺利完成动力减速段的制导任务。为了保证火箭返回过程中满足各种约束及使用燃料最少的要求,基准轨迹的获取可采用基于凸优化或高斯伪谱法[18-19]的轨迹规划方法。

基准轨迹确定后,轨迹跟踪制导采用速度跟踪和位置跟踪独立设计的制导方案,基于LQR的位置跟踪制导确保对轨迹位置的高精度跟踪,通过不断修正过载控制量,减少轨迹位置偏差;基于分段凸优化方法的速度跟踪采用分段跟踪策略,每个轨迹分段采用凸优方法化求解最优控制问题,通过对始/末端速度施加约束以实现跟踪,确保火箭在存在初始速度误差和模型误差的情况下,实现对速度的高精度跟踪,同时追求燃料最省。

2.2 基于分段凸优化的速度跟踪

对于速度跟踪制导律设计,为减少凸优化方法的计算量,将速度跟踪按基准轨迹进行分段,滚动时域控制中控制变量的优化只需保证对当前跟踪段上的速度跟踪,而不需要保证从当前分段到飞行结束的整个全程的速度跟踪,从而大幅度降低凸优化的计算时间。具体来说,首先将基准轨迹按坐标x分成M段,可获取M段分段轨迹和$(M+1)$个分段点,所有的分段点将作为速度跟踪的基准点,在每一分段上的基准点处施加速度相关的初始约束与终端约束。可对x采用均匀分段,也可根据实际情况进行不均匀分段,比如采用类似伪谱法的节点配置方式[20]。本文通过大量测试发现,均匀分段与非均匀分段方法精度相差不大,而前者实现起来更加简单直接,因此本文采用均匀分段的方法。

采用序列凸优化方法求解第$j(1 \leqslant j \leqslant M)$个轨迹分段内的最优控制问题,考虑燃料最省的性能指标,及该段内基准点处初始与终端约束以保证速度跟踪的精度,得到该段的最优控制量即推力P。同时,将推力迅速作用到该分段上,预测该段终端位置处的状态,并根据实际环境考虑各种干扰和不确定性,进行状态更新,将其作为下一个轨迹分段$j+1$的初始状态,进行下一个轨迹分段上的凸优化求解,直至达到整个飞行轨迹的终点。通过这种分段凸优化的形式,在系统受到各种外界干扰和存在模型误差情况下,也能及时进行修正,不会产生误差积累,确保速度的高精度跟踪。

当考虑对速度进行跟踪时,仅考虑火箭速度和质量的变化如下:

$$\begin{cases} \dfrac{dV}{dx} = \dfrac{-P\cos\alpha - D}{mV\cos\theta} - \dfrac{g\sin\theta}{V\cos\theta} = f_1 \\ \dfrac{dm}{dx} = \dfrac{-P}{I_{sp} V\cos\theta} = f_2 \end{cases} \tag{5}$$

式中:f_1和f_2分别为速度V和质量m对x的微

分;方程左端的变量 θ、阻力 D 取标称轨迹值。由于 α 较小,可取 $\cos\alpha=1$。

对非线性动力学等式约束式(5)进行凸优化,为书写方便将式(5)简写为

$$\dot{\boldsymbol{X}}(x)=\boldsymbol{f}(\boldsymbol{X}(x),\boldsymbol{U}(x))=[f_1,f_2]^{\mathrm{T}} \tag{6}$$

式中:$\dot{\boldsymbol{X}}(x)$为状态变量,$\boldsymbol{X}=[V,m]^{\mathrm{T}}$;控制量 $\boldsymbol{U}$ 为推力。设第 k 次迭代的参考轨迹上状态和控制量分别为 $\boldsymbol{X}_{\mathrm{ref}}(k)=[V_k,m_k]^{\mathrm{T}}$ 和 $P(k)=P_k$。以下统一将 $\boldsymbol{X}_{\mathrm{ref}}(k)$ 记为 $\boldsymbol{X}_k$,在参考轨迹 $\{V_k,m_k,P_k\}$ 处对非线性状态方程式(6)采用逐次线性化方法进行凸优化处理,进行一阶泰勒展开,得

$$\dot{\boldsymbol{X}}(x)=\boldsymbol{A}_k\boldsymbol{X}(x)+\boldsymbol{B}_k\boldsymbol{U}(x)+\boldsymbol{C}_k \tag{7}$$

式中:各项系数矩阵中的下角标 k 表示在第 k 次迭代对应的参考轨迹处取值,各项系数矩阵可分别表示为

$$\begin{cases}\boldsymbol{A}_k=\left[\begin{array}{cc}\dfrac{\partial f_1}{\partial V} & \dfrac{\partial f_1}{\partial m}\\ \dfrac{\partial f_2}{\partial V} & \dfrac{\partial f_2}{\partial m}\end{array}\right]\Bigg|_{\boldsymbol{X}_k,\boldsymbol{U}_k}=\\ \qquad\left[\begin{array}{cc}\dfrac{-P_k-D}{m_kV_k^2\cos\theta}+\dfrac{g\sin\theta}{V_k^2\cos\theta} & \dfrac{P_k+D}{m_k^2V_k\cos\theta}\\ \dfrac{P_k}{I_{\mathrm{sp}}V_k^2\cos\theta} & 0\end{array}\right]\\ \boldsymbol{B}_k=\left[\begin{array}{c}\dfrac{\partial f_1}{\partial P}\\ \dfrac{\partial f_2}{\partial P}\end{array}\right]\Bigg|_{\boldsymbol{X}_k,\boldsymbol{U}_k}=\left[\begin{array}{c}-\dfrac{1}{V_km_k\cos\theta}\\ -\dfrac{1}{I_{\mathrm{sp}}V_k\cos\theta}\end{array}\right]\\ \boldsymbol{C}_k=\boldsymbol{f}-\boldsymbol{A}_k\boldsymbol{X}_k-\boldsymbol{B}_k\boldsymbol{U}_k=\\ \qquad\left[\begin{array}{c}f_1-\dfrac{\partial f_1}{\partial V}V_k-\dfrac{\partial f_1}{\partial m}m_k-\dfrac{\partial f_1}{\partial P}P_k\\ f_2-\dfrac{\partial f_2}{\partial V}V_k-\dfrac{\partial f_2}{\partial m}m_k-\dfrac{\partial f_2}{\partial P}P_k\end{array}\right]\end{cases} \tag{8}$$

在式(7)的基础上,进一步进行离散化。欧拉离散法[16-17]形式简单、精度足够,在最优控制问题求解中广为采用,本文选用欧拉法进行离散。若需进一步提高精度,也可考虑梯形[21]或伪谱离散法[22]。设离散周期为 T,离散点数为 N,则状态方程离散为

$$\boldsymbol{X}(i+1)=\boldsymbol{A}(i)\boldsymbol{X}(i)+\boldsymbol{B}(i)\boldsymbol{U}(i)+\boldsymbol{C}(i) \tag{9}$$

式中:$\boldsymbol{A}(i)=T\boldsymbol{A}_k+\boldsymbol{I}_2$,$\boldsymbol{I}_2$ 为 2×2 的单位矩阵;$\boldsymbol{B}(i)=T\boldsymbol{B}_k$;$\boldsymbol{C}(i)=T\boldsymbol{C}_k$;$i$ 为第 i 个离散点。

通过加入该段基准轨迹点对应的速度相关的初始约束与终端约束,实现对该分段内的速度跟踪。对于第 j 个分段,加入始端约束为

$$V_0^j=V_{\mathrm{f}}^{j-1} \tag{10}$$

式中:下标“0”表示该段的始端约束;下标“f”表示上一分段的速度终值,即始端约束的状态取上一段的终值,而不是基准弹道对应的值,从而实现状态的更新。终端约束为

$$V_{\mathrm{f}}^j=V_{\mathrm{Bf}}^j\pm e_1 \tag{11}$$

式中:下标“Bf”为该分段末端基准轨迹对应的速度终值;e_1 为允许的终端误差,引入 e_1 将终端速度约束设定到一个区域,而非一个点,可提高凸优化求解的可靠性。另外,控制量推力 P_i 存在幅值约束:

$$P_{\min}\leqslant P_i\leqslant P_{\max} \tag{12}$$

优化目标为

$$J=\sum_{i=1}^{N}P_i^2 \tag{13}$$

式中:N 为每个分段的离散点数。式(13)的含义是使离散控制量的平方和最小,即最小化每个分段燃料消耗。

经过凸化和离散化处理后,得到第 j 个分段的二阶锥规划问题如下:

$$\begin{cases}\text{Find}\quad P_i(x)\qquad i=1,2,\cdots,N\\ \text{Min}\quad J=\sum\limits_{i=1}^{N}P_i^2\\ \text{s.t.}\begin{cases}\dot{\boldsymbol{X}}(x)=\boldsymbol{A}_k\boldsymbol{X}(x)+\boldsymbol{B}_k\boldsymbol{U}(x)+\boldsymbol{C}_k\\ P_{\min}\leqslant P_i\leqslant P_{\max}\\ V(1)=V_0^j\\ V(N)=V_{\mathrm{f}}^j\pm e_1\end{cases}\end{cases} \tag{14}$$

式(14)中的凸优化问题由于仅涉及 2 个状态量和一个控制量,无需对模型进行繁琐的变换和凸化剪裁,且优化时域较短,可采用序列凸优化[23]进行快速求解,提高了凸优化求解的速度与可靠性。

2.3　基于 LQR 的轨迹跟踪

轨迹跟踪时考虑如下运动方程:

$$\begin{cases}\dfrac{\mathrm{d}\theta}{\mathrm{d}x}=\dfrac{-P\sin\alpha+Y}{mV^2\cos\theta}-\dfrac{g}{V^2}\\ \dfrac{\mathrm{d}y}{\mathrm{d}x}=\tan\theta\end{cases} \tag{15}$$

运载火箭等飞行器一般会安装惯性测量装置,可以对火箭受到的加速度进行直接测量,因此选用侧向加速度作为控制变量。火箭侧向加速度可表示为

$$a_y=\frac{-P\sin\alpha+Y}{m} \tag{16}$$

由式(16)可得火箭运动方程为

$$\begin{cases}\dfrac{\mathrm{d}\theta}{\mathrm{d}x}=\dfrac{a_y}{V^2\cos\theta}-\dfrac{g}{V^2}=f_3\\ \dfrac{\mathrm{d}y}{\mathrm{d}x}=\tan\theta=f_4\end{cases} \tag{17}$$

式中:f_3 和 f_4 分别为速度倾角 θ 和高度 y 对 x 的微分。

取扰动系统状态变量 $\boldsymbol{Y}=[\Delta\theta \quad \Delta y]^{\mathrm{T}}$，其中：

$$\begin{cases}\Delta\theta = \theta - \theta_{\mathrm{B}} \\ \Delta y = y - y_{\mathrm{B}}\end{cases} \tag{18}$$

式中：下标“B”表示基准轨迹参数。采用小扰动法对动力学方程进行线性化，可得到状态方程为

$$\frac{\mathrm{d}\boldsymbol{Y}}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}\begin{bmatrix}\Delta\theta \\ \Delta y\end{bmatrix} = \boldsymbol{A}\begin{bmatrix}\Delta\theta \\ \Delta y\end{bmatrix} + \boldsymbol{B}\Delta a_y \tag{19}$$

式中：$\Delta a_y = a_y - a_{y\mathrm{B}}$ 为控制量。矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 的表达式为

$$\begin{cases}\boldsymbol{A} = \begin{bmatrix}\dfrac{\partial f_3}{\partial\theta} & \dfrac{\partial f_3}{\partial y} \\ \dfrac{\partial f_4}{\partial\theta} & \dfrac{\partial f_4}{\partial y}\end{bmatrix} = \begin{bmatrix}\dfrac{a_y\sin\theta}{V^2\cos^2\theta} & 0 \\ \dfrac{1}{\cos^2\theta} & 0\end{bmatrix} \\ \boldsymbol{B} = \begin{bmatrix}\dfrac{\partial f_3}{\partial a_y} \\ \dfrac{\partial f_4}{\partial a_y}\end{bmatrix} = \begin{bmatrix}\dfrac{1}{V^2\cos\theta} \\ 0\end{bmatrix}\end{cases} \tag{20}$$

根据 LQR 理论，对线性系统选定拉格朗日型性能指标如下：

$$J = \frac{1}{2}\int_0^{\infty}(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Y} + \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}\boldsymbol{U})\,\mathrm{d}t \tag{21}$$

式中：$\boldsymbol{Q}$ 和 $\boldsymbol{R}$ 为需要设计的权重矩阵，且满足 $\boldsymbol{Q}$ 半正定，$\boldsymbol{R}$ 正定；$\boldsymbol{Y}$ 为系统状态变量。根据 LQR 理论，如果该系统受到外界干扰而偏离基准状态，则应施加控制量 $U=-\boldsymbol{K}^*\boldsymbol{Y}$ 使系统回到基准状态附近并同时满足 J 最小，且反馈系数 $\boldsymbol{K}^*$ 为

$$\boldsymbol{K}^* = \boldsymbol{R}^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{P} \tag{22}$$

式中：$\boldsymbol{P}$ 为黎卡提方程的解。黎卡提方程可以表示为

$$-\boldsymbol{P}\boldsymbol{A} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{B}\boldsymbol{R}^{-1}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{P} - \boldsymbol{Q} = 0 \tag{23}$$

当黎卡提方程无解时，一般可取上一步制导周期的解，这样既可以保证算法的工程实用性，同时也不会影响算法的使用性能。

基于 LQR 的位置跟踪制导律工作流程，首先，在基准轨迹上按坐标 x 等间隔取样，按照式(19)获取火箭侧向扰动运动的线性化模型；然后，选择权重指标，计算最优反馈系数矩阵 $\boldsymbol{K}^*$；最后，根据实际飞行的系统状态偏差计算控制量 $U=-\boldsymbol{K}^*\boldsymbol{Y}$，通过数值积分手段对位置状态方程进行预测，积分步长为 Δx。

本文方法的信息流程如图 2 所示。

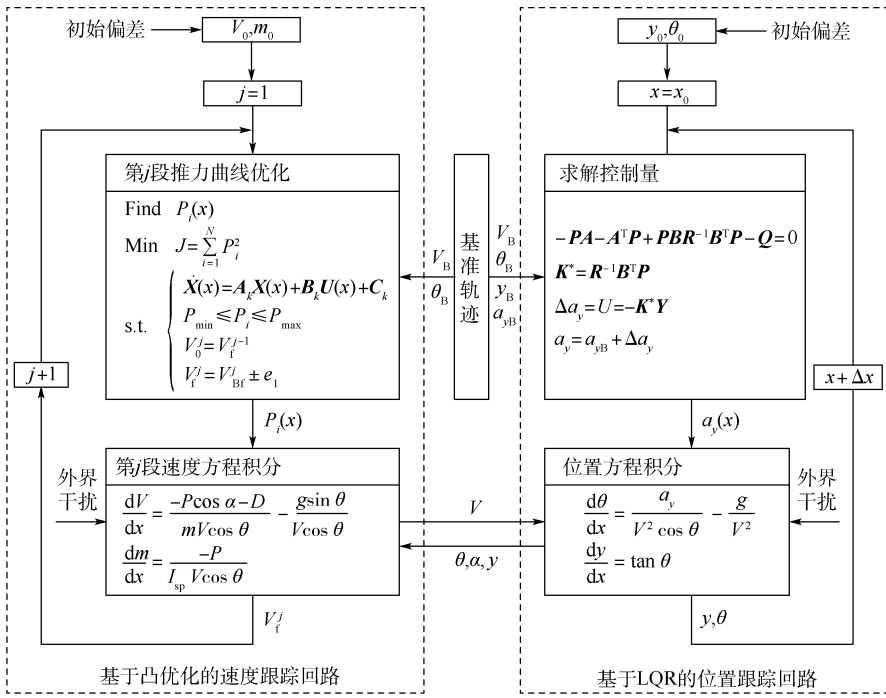


图 2 本文方法的轨迹跟踪信息流程

Fig. 2 Information flow chart of trajectory tracking with the proposed method

3 仿真验证

3.1 本文方法与滚动凸优化方法的仿真对比分析

以某运载火箭动力减速段为例进行仿真，火箭初始状态和终端状态如表 1 所示。基于分段凸优化的速度跟踪由于计算量相较于解析式的 LQR 大，其制导周期一般更长。本文方法仿真中速度跟踪分段数取 10，按 x 坐标均匀分段，则制导周期为 1/10 的总横向航程；在每个速度跟踪周期内，分段凸优化先规划出该段内的推力曲线，基于此对式(5)中的速度进行预测；LQR 跟踪器计算出控制量 a_y，然后对位置方程(17)中的位置

(y,θ)进行预测。由于 LQR 解析计算速度非常快,仿真中其制导周期取为每个凸优化速度跟踪制导周期的 1/200(按 x 坐标进行等分)。为了比较公平,基于直接滚动凸优化方法全程同样取 10 个制导周期。仿真所使用电脑配置为 i5 8400 CPU、8 GB 内存的普通台式机上。

2 种方法的轨迹曲线如图 3 所示。本文方法的基准轨迹是一条离线优化轨迹,在仿真过程中,由于没有考虑外界的各种干扰因素,采用本文方法的飞行轨迹应与滚动凸优化方法的轨迹基本重合,图 3 的仿真结果也验证了这一点。

表 1　火箭初始状态和终端状态

Table 1　Initial and terminal states of rocket

状态	速度/(m·s^{-1})	速度倾角/(°)	航程 x/m	高度 y/m
初始	2 762.9	−28.82	0	71 410
终端	1 211.9	−37.64	85 129	17 342

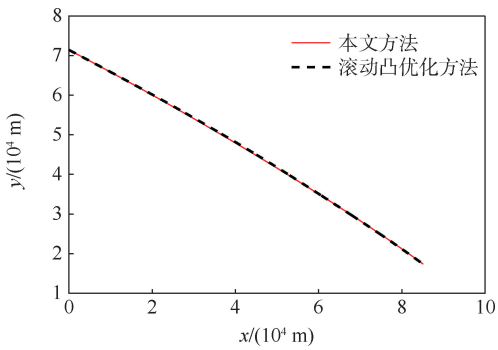


图 3　轨迹曲线

Fig. 3　Trajectory curves

2 种方法的计算时间如表 2 所示。从中可以看出,本文方法的轨迹跟踪制导方法凸优化求解时间为 22.66 s,远小于滚动凸优化方法的计算时间 103.95 s。本文方法由于将轨迹跟踪问题简化为速度跟踪和位置跟踪 2 部分,位置跟踪采用高效率的 LQR 跟踪方法,速度跟踪形成的最优控制问题由于只涉及到 2 个状态变量和 1 个控制量,并采用分段控制策略,计算时间大幅减少。由仿真验证可见,本文方法相比于滚动凸优化方法可大幅度减低计算量。

在仿真过程中发现,直接采用滚动凸优化方法在求解时容易出现无解的情况。这主要是火箭返回问题涉及的模型状态量和控制量多,过程约束较多且较为苛刻,动力学高度非线性,现有的滚动凸优化方法利用逐次逼近的方法对整个非线性动力学方程线性化,极有可能导致凸化后的问题可行域非常小,实际使用过程中会出现无解的情况,可靠性难以保证。通过引入变量变换、松弛等方法对模型进行凸化和裁剪,可以明显改善凸优化求解的性能,提高可靠性,但凸优化方法针对不同的问题需要专门研究,这对工程人员数学基础和凸优化经验要求较高,通用性与工程适用性较差。而采用本文所提方法,速度跟踪的最优控制问题中系统状态量和控制量减少,且采用分段控制策略可以减少每次凸优化求解中离散点个数,因此该方法在仿真过程中未出现无解的情况,具有更高的可靠性。

表 2　两种方法的计算时间

Table 2　Computation time of two methods

方法	计算时间/s
LQR + 凸优化	22.66
滚动凸优化	103.95

综上所述,本文方法在计算速度、求解可靠性等方面明显优于滚动凸优化方法。

3.2　本文方法与 LQR 方法的仿真对比分析

本节主要从考虑初始误差和同时考虑各种干扰误差 2 种情况对 LQR 方法和本文方法进行仿真对比分析,重点分析 2 种方法的跟踪精度和燃料耗损性能。针对火箭返回运动的强非线性特性,为了减少 LQR 线性化引起的模型误差,在以下仿真过程中,LQR 方法跟踪的基准弹道离散点数取 2 000 个点。

3.2.1　考虑初始误差的轨迹跟踪仿真

火箭初始误差主要包括初始速度误差(ΔV_0)、速度倾角误差($\Delta\theta_0$)及高度误差(Δy_0)等,考虑初始速度、速度倾角和高度分别存在 +100 m/s、+2°和 +1 000 m 的初始误差。在相同的场景及参数设置下,分别采用传统的 LQR 方法和本文方法进行轨迹跟踪制导仿真。

考虑 ΔV_0 下的仿真结果如图 4 ~ 图 6 所示。可以看出,2 种跟踪方法的跟踪精度基本相当,速度误差在最初始段均得以迅速消除;本文方法速度倾角跟踪误差全程小于 0.05°,高度跟踪误差全程小于 17 m。

从图 5 和图 6 中可见,本文方法高度跟踪误差和速度倾角跟踪误差出现小幅周期性振荡,这是由于对速度采用分段跟踪,每段只跟踪该段的速度终值,而对中间过程的速度不进行精确的跟踪,因此在端点之间会产生一定的速度偏差,而位置回路会使用速度回路传递的速度量信息,因此会造成速度倾角和高度跟踪误差小幅振荡,但振荡幅度很小,能保证较高的跟踪精度。

考虑 $\Delta\theta_0$ 下的仿真结果如图 7 ~ 图 9 所示。

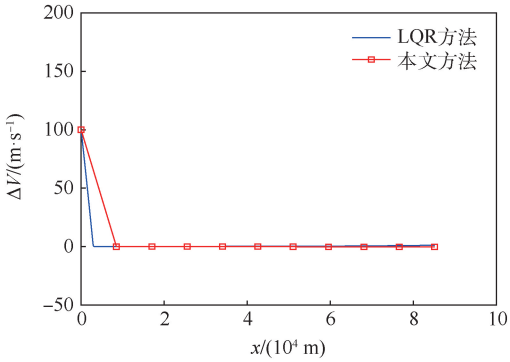


图 4 速度误差曲线(考虑 ΔV_0)

Fig. 4 Velocity error curves (with ΔV_0)

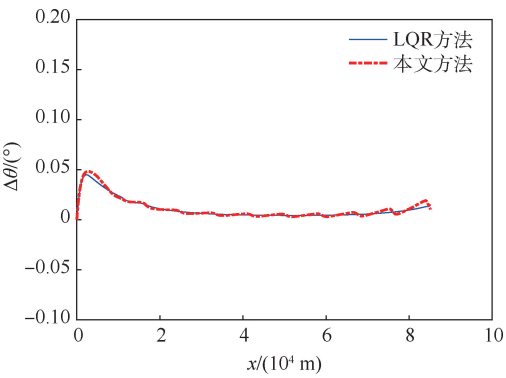


图 5 速度倾角误差曲线(考虑 ΔV_0)

Fig. 5 Flight path angle error curves (with ΔV_0)

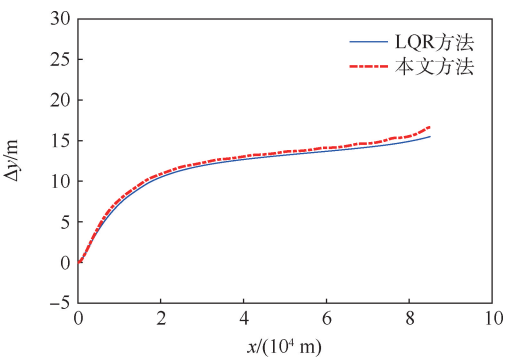


图 6 高度误差曲线(考虑 ΔV_0)

Fig. 6 Altitude error curves (with ΔV_0)

可以看出,2 种跟踪方法的跟踪精度基本相当,速度倾角误差在飞行过程中逐渐得以消除并很快收敛到 0;本文方法在速度跟踪点上误差均小于 1 m/s,高度误差由于受到初始速度倾角误差的影响逐渐增大,但随着速度倾角偏差的修正,高度跟踪误差也得到进一步修正,初始速度倾角误差消除之后全程高度误差不超过 18 m。

考虑初始高度误差下的仿真结果如图 10 ~ 图 12 所示。可以看出,2 种跟踪方法的跟踪精度基本相当,本文方法在速度跟踪点上最大误差不超过 1 m/s,由于初始高度误差的影响,速度倾角跟踪误差先增大后逐渐收敛于 0,高度跟踪误差迅速减小,之后全程高度误差不超过 17 m。

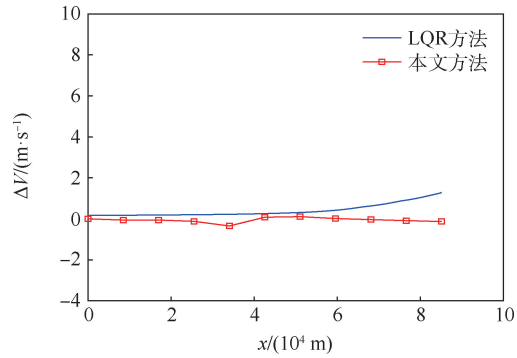


图 7 速度误差曲线(考虑 $\Delta\theta_0$)

Fig. 7 Velocity error curves (with $\Delta\theta_0$)

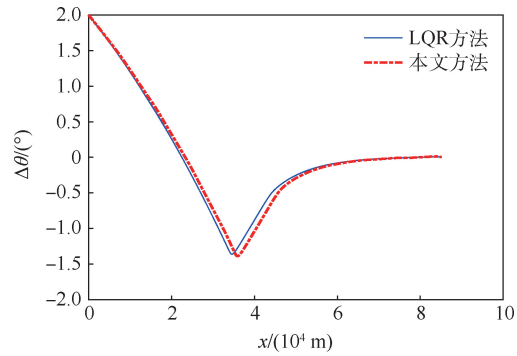


图 8 速度倾角误差曲线(考虑 $\Delta\theta_0$)

Fig. 8 Flight path angle error curves (with $\Delta\theta_0$)

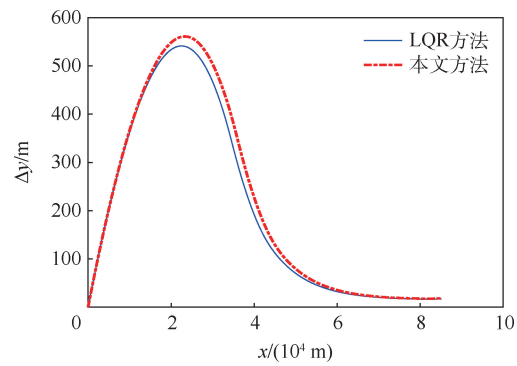


图 9 高度误差曲线(考虑 $\Delta\theta_0$)

Fig. 9 Altitude error curves (with $\Delta\theta_0$)

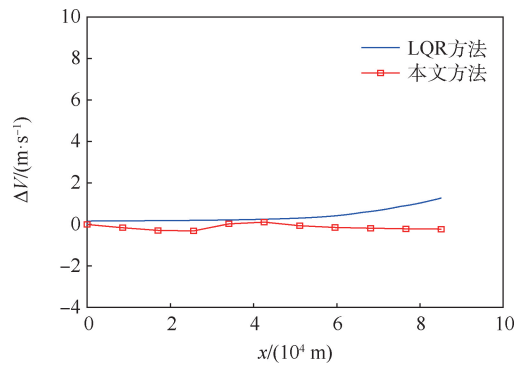


图 10 速度误差曲线(考虑 Δy_0)

Fig. 10 Velocity error curves (with Δy_0)

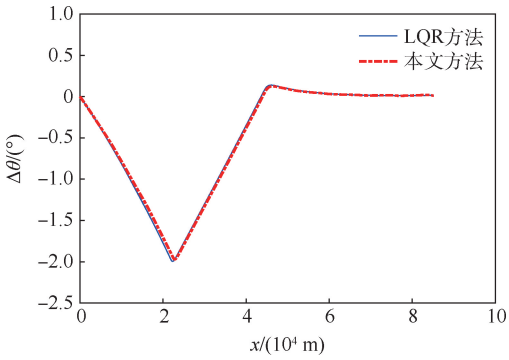


图 11　速度倾角误差曲线(考虑 Δy_0)

Fig. 11　Flight path angle error curves (with Δy_0)

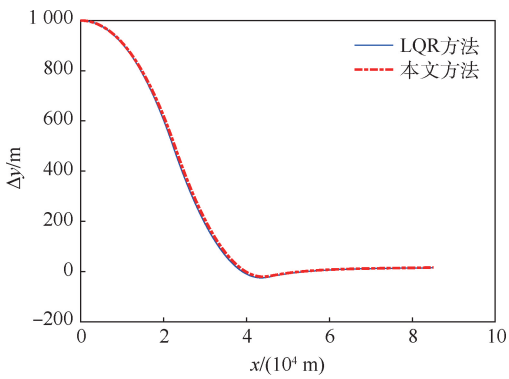


图 12　高度误差曲线(考虑 Δy_0)

Fig. 12　Altitude error curves (with Δy_0)

各种初始误差下采用 2 种方法的火箭终端剩余质量如表 3 所示。可以看出,本文方法火箭的终端质量高于传统的 LQR 方法,终端质量提升 268 ~ 272 kg,燃料更省。

表 3　不同初始条件下火箭终端质量

Table 3　Terminal mass of the rocket under different intial conditions

方法	火箭终端质量/kg		
	$\Delta V_0 = 100$ m/s	$\Delta\theta_0 = 2°$	$\Delta y_0 = 1\,000$ m
本文方法	65 877	67 611	67 563
LQR 方法	65 609	67 339	67 292

考虑到动力减速段火箭飞行高度在 20 ~ 70 km之间,空气相对比较稀薄,不能充分发挥气动减速效应达到降低使用燃料的目的,对于更稠密大气层内的轨迹跟踪其节省的燃料将更为可观。

从本节仿真结果可以看出,2 种方法的跟踪精度基本相当,而本文方法的轨迹跟踪制导方法可以在保证较好跟踪的情况下降低燃料消耗,且无需提供精确的推力控制输入,使用更加灵活。

3.2.2　考虑各种干扰误差的蒙特卡罗仿真

3.2.1 节仿真都是考虑单一的偏差因素,有必要综合考虑各种误差因素同时作用下的轨迹跟踪性能。同时考虑火箭初始状态误差、气动模型误差、风等因素的干扰,各种干扰因素、大小和分布形式如表 4 所示,采用蒙特卡罗方法进行仿真。

采用本文方法对火箭动力减速段进行跟踪,500 次蒙特卡罗仿真均成功求解,仿真结果如图 13 ~ 图 18 所示。其中,图 13 ~ 图 15 为速度误差、速度倾角误差和高度误差的变化过程,可见整个制导过程这些误差都快速减小,最终都收敛到较小值,说明轨迹跟踪中偏差修正的有效性。图 16 ~ 图 18 为终端速度误差、速度倾角误差和高度误差的分布。可以看到,500 次仿真中,3 种终端误差分别位于[- 1, 0. 6] m/s、[0. 006, 0. 018]°、[15,18] m 的区间内,处于可接受的范围。

表 4　蒙特卡罗仿真考虑的误差项

Table 4　Disturbances in Monte-Carlo simulation

误差项	标准差	分布形式
初始速度/(m · s^{-1})	100(3σ)	正态
初始速度倾角/(°)	2(3σ)	正态
初始高度/m	1 000(3σ)	正态
气动阻力/%	10	均匀
气动升力/%	10	均匀
质量偏差/kg	50(3σ)	正态
比冲偏差/%	3(3σ)	正态
风速/(m · s^{-1})	30(3σ)	正态

注:σ 为方差。

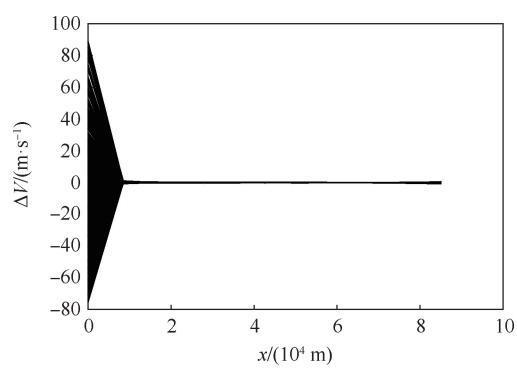


图 13　速度误差(蒙特卡罗仿真)

Fig. 13　Velocity error (Monte-Carlo simulation)

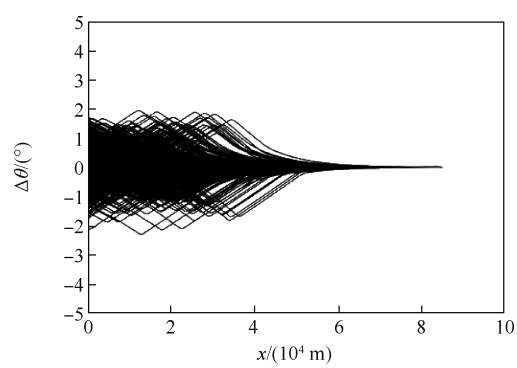


图 14　速度倾角误差(蒙特卡罗仿真)

Fig. 14　Flight path angle error (Monte-Carlo simulation)

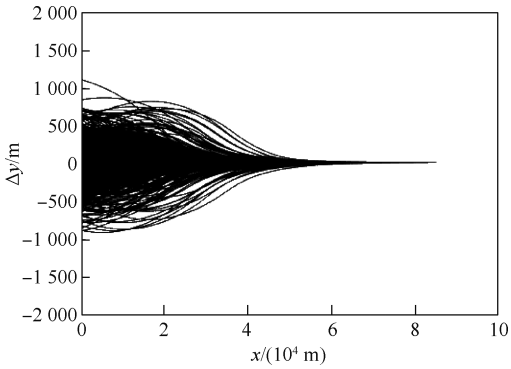


图15 高度误差(蒙特卡罗仿真)

Fig. 15 Altitude error (Monte-Carlo simulation)

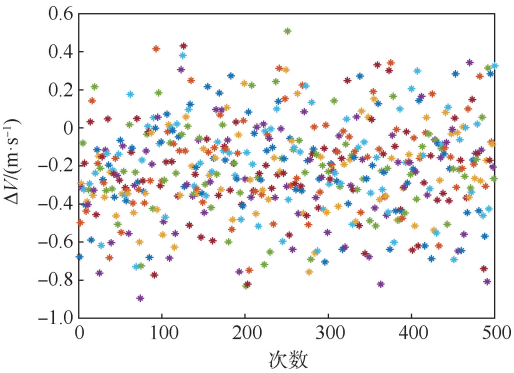


图16 终端速度误差(蒙特卡罗仿真)

Fig. 16 Terminal velocity error (Monte-Carlo simulation)

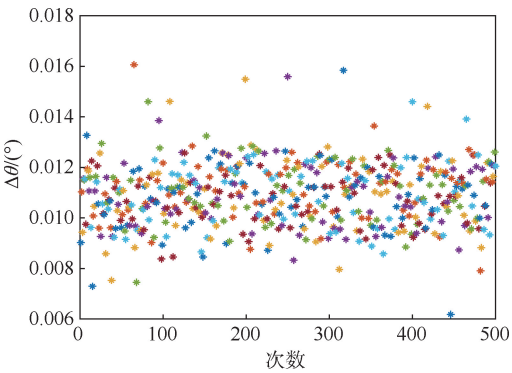


图17 终端速度倾角误差(蒙特卡罗仿真)

Fig. 17 Terminal flight path angle error (Monte-Carlo simulation)

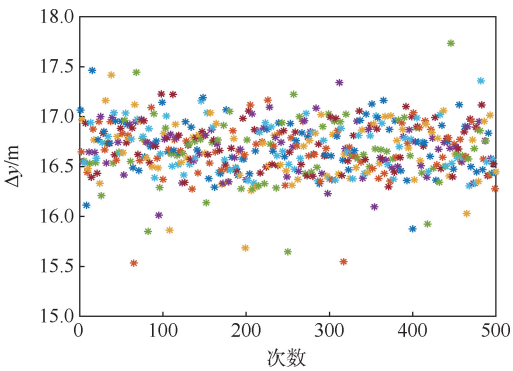


图18 终端高度误差(蒙特卡罗仿真)

Fig. 18 Terminal altitude error (Monte-Carlo simulation)

整体而言,当考虑初始干扰、风干扰和其他各种干扰因素下,采用本文方法轨迹跟踪性能良好,可以抵抗多种干扰因素可能导致的较大跟踪误差,全程跟踪精度良好,说明本文方法具有较强的抵抗模型不确定性和外界干扰的能力,具有良好的跟踪性能和鲁棒性能。

4 结 论

1) 本文针对可重复使用运载火箭动力减速段,提出了一种基于LQR和分段凸优化的标称轨迹跟踪制导方法。将动力学模型分解为速度和位置相关的2部分,采用分段凸优化方法进行速度跟踪,同时最小化燃料消耗;采用LQR方法实现对火箭飞行轨迹的高精度跟踪,同时抵抗各种误差和干扰的影响。

2) 采用数值仿真对本文方法进行验证,仿真结果表明,相对于传统的LQR跟踪制导方法,本文方法轨迹跟踪精度与LQR基本相当,对不确定性干扰具有较强的鲁棒性,但本文方法能一定程度降低燃料消耗,且无需基准轨迹对应的推力控制量信息,实际应用更加灵活。

3) 本文方法采用分段控制策略,无需求解全段道优化控制问题,相较于直接滚动凸优化方法计算效率大幅提升,且具有较高的可靠性。本文方法也可应用于其他推力可控的飞行器轨迹跟踪制导,尤其是大气较为稠密的情况,其燃料节省效果将更加可观。

参考文献 (References)

[1] 崔乃刚,吴荣,韦常柱,等. 垂直起降可重复使用运载器发展现状与关键技术分析[J]. 宇航总体技术,2018,2(2):27-42.
CUI N G,WU R,WEI C Z,et al. Development and key technologies of vertical takeoff vertical landing reusable launch vehicle [J]. Astronautical Systems Engineering Technology, 2018, 2(2):27-42(in Chinese).

[2] 王振国,罗世彬,吴建军. 可重复使用运载器研究进展[M]. 长沙:国防科技大学出版社,2004:1-20.
WANG Z G,LUO S B,WU J J. Development of reusable launch vehicle technology[M]. Changsha: National University of Defense Technology Press,2004:1-20(in Chinese).

[3] 高朝辉,张普卓,刘宇,等. 垂直返回重复使用运载火箭技术分析[J]. 宇航学报,2016,37(2):145-152.
GAO Z H,ZHANG P Z,LIU Y,et al. Analysis of vertical landing technique in reusable launch vehicle[J]. Journal of Astronautics,2016,37(2):145-152(in Chinese).

[4] HARPOLD J C, GRAVES C A. Shuttle entry guidance[J]. Journal of Astronautical Sciences,1978,27(3):239-268.

[5] ROENNEKE A J,CORNWELL P J. Trajectory control for a low-

lift re-entry vehicle[J]. Journal of Guidance, Control, and Dynamics, 1993, 16(5): 927-933.

[6] DUKEMAN G. Profile-following entry guidance using linear quadratic regulator theory[C]// AIAA Guidance, Navigation, and Control Conference and Exhibit. Reston: AIAA, 2002: 4457.

[7] 张大元,雷虎民,王君,等. 一种反馈线性化弹道跟踪制导律设计[J]. 弹道学报,2014,26(4):7-12.
ZHANG D Y, LEI H M, WANG J, et al. A trajectory tracking guidance law based on feedback linearization method[J]. Journal of Ballistics, 2014, 26(4): 7-12(in Chinese).

[8] 刘晓东,黄万伟,禹春梅. 含扩张状态观测器的高超声速飞行器动态面姿态控制[J]. 宇航学报,2015,36(8):916-922.
LIU X D, HUANG W W, YU C M. Dynamic surface attitude control for hypersonic vehicle containing extended state observer [J]. Journal of Astronautics, 2015, 36(8): 916-922(in Chinese).

[9] 宋晨,周军,郭建国,等. 高超声速飞行器基于路径跟踪的制导方法[J]. 宇航学报,2016,37(4):435-441.
SONG C, ZHOU J, GUO J G, et al. Hypersonic vehicle guidance based on path-following [J]. Journal of Astronautics, 2016, 37(4): 435-441(in Chinese).

[10] 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.

[11] ACIKMESE B, CARSON J M, BLACKMORE L. Lossless convexification of nonconvex control bound and pointing constraints of the soft landing optimal control problem[J]. IEEE Transactions on Control Systems Technology, 2013, 21(6): 2104-2113.

[12] 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.

[13] LIU X F, LU P. Solving nonconvex optimal control problems by convex optimization[J]. Journal of Guidance, Control, and Dynamics, 2014, 37(3): 750-765.

[14] LU P, LIU X F. Autonomous trajectory planning for rendezvous and proximity operations by conic optimization[J]. Journal of Guidance, Control, and Dynamics, 2013, 36(2): 375-389.

[15] 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.

[16] 路钊. 高超声速飞行器再入末段轨迹在线优化[D]. 哈尔滨:哈尔滨工业大学,2014:27-74.
LU Z. Online trajectory optimization for the terminal stage of re-entry hypersonic vehicles [D]. Harbin: Harbin Institute of Technology, 2014: 27-74(in Chinese).

[17] 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).

[18] 雍恩米,唐国金,陈磊. 基于 Gauss 伪谱方法的高超声速飞行器再入轨迹快速优化[J]. 宇航学报,2008,29(6):1766-1772.
YONG E M, TANG G J, CHEN L. Rapid trajectory optimization for hypersonic reentry vehicle via Gauss pseudospectral method [J]. Journal of Astronautics, 2008, 29(6): 1766-1772(in Chinese).

[19] 王少奇,马东立,杨穆清,等. 高空太阳能无人机三维航迹优化[J]. 北京航空航天大学学报,2019,45(5):936-943.
WANG S Q, MA D L, YANG M Q, et al. Three-dimensional optimal path planning for high-altitude solar-powered UAV[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(5): 936-943(in Chinese).

[20] DARBY C L, HAGER W W, RAO A V. An hp-adaptive pseudospectral method for solving optimal control problems[J]. Optimal Control Applications and Methods, 2011, 32(4): 476-502.

[21] 李樾,韩维,陈清阳,等. 凸优化算法在有人/无人机协同系统航迹规划中的应用[J]. 宇航学报,2020,41(3):276-286.
LI Y, HAN W, CHEN Q Y, et al. Application of convex optimization algorithm in trajectory planning of manned/unmanned cooperative system[J]. Journal of Astronautics, 2020, 41(3): 276-286(in Chinese).

[22] 唐胜景,王肖,郭杰. 基于 hp 伪谱凸优化的高超声速滑翔飞行器轨迹优化与制导[J]. 战术导弹技术,2020(5):66-75.
TANG S J, WANG X, GUO J. Trajectory optimization and guidance for hypersonic gliding vehicles based on hp pseudospectral convex programming [J]. Tactical Missile Technology, 2020(5): 66-75(in Chinese).

[23] QUOC T D, SAVORGNAN C, DIEHL M. Real-time sequential convex programming for optimal control applications [C]// Modeling, Simulation and Optimization of Complex Processes, 2012: 91-102.

Rocket return trajectory tracking guidance based on convex optimization and LQR

WU Jie[1], ZHANG Cheng[1], LI Miao[2], XIONG Fenfen[1,*]

(1. School of Astronautics, Beijing Institute of Technology, Beijing 100081, China;
2. Beijing Institute of Special Mechanic-Electric, Beijing 100012, China)

Abstract: For the powered descent phase of the reusable launch vehicle, various strict process constraints, terminal constraints and requirements on fuel saving exist, which bring great challenges to the guidance. This paper proposes a trajectory tracking guidance method based on convex optimization and linear quadratic regulator (LQR). The improved receding horizon convex optimization method is used to track the reference velocity of the rocket without requiring accurate thrust control input, which greatly simplifies the optimization model, and thus saves the computational cost of solving the optimal control problem. Meanwhile, the fuel is reduced as far as possible under various initial errors and model errors. On the other hand, the LQR technique is used to track the position of the rocket with high precision. The simulation results show that compared with the traditional LQR tracking guidance method, the proposed method can obtain comparable tracking accuracy, while greatly reducing fuel consumption. And compared to the existing receding horizon convex optimization, the proposed method can evidently reduce the computational cost and improve reliability.

Keywords: reusable launch vehicle; powered decent phase; trajectory tracking guidance; linear quadratic regnlator; convex optimization

Received: 2021-02-19; **Accepted**: 2021-04-18; **Published online**: 2021-05-07 09:22

URL: kns.cnki.net/kcms/detail/11.2625.V.20210506.1717.003.html

Foundation items: Science Challenge Project (TZ2019001); Pre-research of Equipment Development During the "13th Five-Year Plan" (202020201320)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0093

一种基于激光位移测量的机械臂参数标定方法

邵鑫[1], 季力[1], 邹怀武[2], 解杨敏[1,*]

(1. 上海大学 机电工程与自动化学院, 上海 200444; 2. 上海市空间飞行器机构重点实验室, 上海 201108)

摘 要: 针对传统机器人参数标定方法,通常依赖昂贵设备,技术实现成本高的问题,提出一种新型低成本的基于激光位移测量的机器人标定系统,利用安装于机械臂末端的激光位移计进行与外部参照物立方体的相对位置测量,并设计实现了六自由度机械臂的运动学参数标定方法。基于单维度位移测量值,采用平面度约束与平面角度约束共同构建误差函数,并利用非线性优化方法求得最终的标定结果。在实验系统上进行标定方法的验证,并与利用激光追踪仪标定的实验结果进行对比。实验数据表明:所提方法在操作简便、成本低廉的同时,能够获得与使用昂贵的外部测量仪器近似的标定效果。

关 键 词: 机械臂; 运动学参数标定; 激光位移测量; 组合约束; 非线性优化

中图分类号: TP241.2

文献标志码: A **文章编号**: 1001-5965(2022)11-2281-08

机器人的定位精度是衡量其性能最重要的指标之一。其重复定位精度主要受关节控制误差的影响,现国内外应用的机器人一般能够达到亚毫米级。但其绝对定位精度一般在毫米级,通常为2~3 mm[1],这在很大程度上降低了其在精密作业中的应用效果。研究表明,机器人绝对定位精度的最大影响因素为其运动学参数的误差,其影响约占到系统总误差的65%~95%[2]。机器人的运动学参数标定是解决这一问题的基本方法。其原理为构造及优化某种与运动约束相关的误差函数,以矫正运动学模型中的机器人名义结构参数,使其更接近于真实加工及安装状态。

一般来讲机器人的参数标定按照约束建立的方法可分为三大类[3]:外部直接测量法[4-6]、外约束标定法[7-8]、自约束标定法[9-11]。外部直接测量法通常利用高精度仪器对机器人末端位置进行三维测量,并以此作为标定约束,这类方法数学建模过程简单,但通常需要昂贵的测量仪器和实施环境支持;外约束标定法利用安装于机器人末端的传感器在不同的空间位置测量与外部参照物的相对位姿关系,从而形成空间约束,这类方法使用的测量仪器相对简单廉价,但需要设计合理的外部参照物;自约束标定法则是通过构建封闭运动链以利用自身的运动学建立几何约束,虽其误差函数构建不依赖参照物,但通常需要外部参考建立封闭运动链,且其数学模型相对复杂。

本文的工作属于外约束标定的范畴。外约束标定通常具有易于计算和实践的优势,已获得广泛的应用。在已有的研究中使用最常见的测量装置有视觉传感器[12-13]和接触式传感器[14-15]两大类。接触式传感器操作较为复杂,且存在碰撞和仪器损伤风险,因此在近年的研究中越来越多地

收稿日期: 2021-02-24; **录用日期**: 2021-06-13; **网络出版时间**: 2021-07-12 15:13

网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210712.1337.003.html

基金项目: 上海市自然科学基金(19ZR1455500); 上海市科学技术委员会启明星计划(19QA1403500); 上海市空间飞行器机构重点实验室开放课题(YYHT-F805201808013)

* **通信作者**. E-mail: xieym@shu.edu.cn

引用格式: *邵鑫, 季力, 邹怀武, 等. 一种基于激光位移测量的机械臂参数标定方法[J]. 北京航空航天大学学报, 2022, 48(11): 2281-2288. SHAO X, JI L, ZOU H W, et al. A parameter calibration method for manipulators based on laser displacement measurement[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2281-2288(in Chinese).*

使用视觉进行标定[16-17]。视觉测量的优势是操作相对简单,但需要进行传感器自标定,且测量精度易受参照物和环境(如标定板和光照)等因素影响。为此本文提出一种利用末端安装的激光位移计测量与外部参照物相对位置关系的方法,同时具有使用便捷、成本低廉、测量稳定及数据获取简单的特点。

为了合理约束多自由度机械臂空间运动,本文设计了相互近似垂直的 3 个平面参照物。研究表明,单纯的平面方程不足以对多自由度机器人形成足够的空间约束[18],因此本文提出了平面度与平面夹角相结合的约束构建方式,形成多目标误差函数。相关的标定方法在实验系统上进行了验证,且与基于高精度激光追踪仪的参数标定进行对比实验,并对两者误差进行对比分析,从而对本文方法的参数标定效果进行了全面性的评价。

1 运动学建模

机器人的运动学模型是参数标定的基础。考虑到串联机器人建模的标准化和连续性,本文使用坐标系后置 DH 参数法[19]建立其运动学模型。机器人基坐标系位于安装基座的中心处,根据 DH 参数法建立每个连杆坐标系,再用齐次变换矩阵描述相邻连杆之间的空间关系[20]。六自由度机械臂 DH 模型坐标系如图 1 所示,机械臂 DH 参数的初始名义值如表 1 所示。

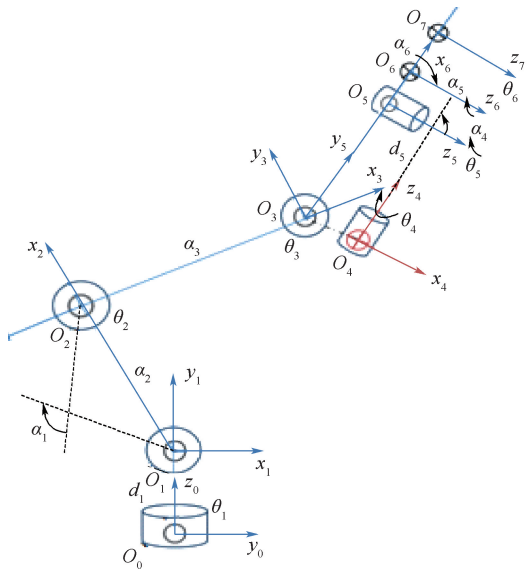


图 1 机械臂 DH 模型坐标系

Fig. 1 DH model coordinate system of manipulators

表 1 DH 参数

Table 1 DH parameters

关节 i	连杆长度 a_i/mm	连杆偏角 α_i/(°)	连杆距离 d_i/mm	连杆夹角 θ_i/(°)
1	0	90	184	θ_1
2	372	0	0	θ_2
3	382	0	0	θ_3
4	0	-90	0	θ_4
5	0	-90	280.4	θ_5
6	36	0	0	θ_6

由连杆坐标系$\{i\}$和相应的连杆参数建立机械臂运动学方程,得到末端执行器相对于基坐标系$\{B\}$的齐次变换矩阵:

$$ {}_6^0\boldsymbol{T} = {}_1^0\boldsymbol{T}\,{}_2^1\boldsymbol{T}\,{}_3^2\boldsymbol{T}\,{}_4^3\boldsymbol{T}\,{}_5^4\boldsymbol{T}\,{}_6^5\boldsymbol{T} \tag{1} $$

其中相邻关节对应的位姿变换矩阵可以写为

$$ {}_i^{i-1}\boldsymbol{T} = \mathbf{Rot}_z(\theta_i)\cdot\mathbf{Trans}_z(d_i)\cdot\mathbf{Trans}_x(a_i)\cdot\mathbf{Rot}_x(\alpha_i) = \begin{bmatrix} \cos\theta_i & -\sin\theta_i\cos\alpha_i & \sin\theta_i\sin\alpha_i & a_i\cos\theta_i \\ \sin\theta_i & \cos\theta_i\cos\alpha_i & -\cos\theta_i\sin\alpha_i & a_i\sin\theta_i \\ 0 & \sin\alpha_i & \cos\alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{2} $$

式中:$\mathbf{Rot}_z(\theta_i)$和$\mathbf{Rot}_x(\alpha_i)$分别为绕 z 轴和 x 轴旋转 θ_i 和 α_i 对应的旋转矩阵;$\mathbf{Trans}_z(d_i)$和$\mathbf{Trans}_x(a_i)$分别为在 z_i 和 x_i 方向上平移 d_i 和 a_i 对应的平移矩阵。

由于编码器回零误差、加工装配误差等因素,每个连杆的 4 个 DH 参数都存在误差,记为 $\Delta\theta_i$、Δa_i、$\Delta\alpha_i$、Δd_i,则六自由度机械臂共有 24 个误差参数。为建立机器人末端误差与各误差参数之间的定量关系模型,修正 DH 参数列表如表 2所示。

表 2 DH 误差参数

Table 2 DH parameters with kinematic errors

关节 i	连杆长度 a_i/mm	连杆偏角 α_i/(°)	连杆距离 d_i/mm	连杆夹角 θ_i/(°)
1	Δa_1	$90+\Delta\alpha_1$	$184+\Delta d_1$	$\theta_1+\Delta\theta_1$
2	$372+\Delta a_2$	$\Delta\alpha_2$	Δd_2	$\theta_2+\Delta\theta_2$
3	$382+\Delta a_3$	$\Delta\alpha_3$	Δd_3	$\theta_3+\Delta\theta_3$
4	Δa_4	$-90+\Delta\alpha_4$	Δd_4	$\theta_4+\Delta\theta_4$
5	Δa_5	$-90+\Delta\alpha_5$	$280.4+\Delta d_5$	$\theta_5+\Delta\theta_5$
6	$36+\Delta a_6$	$\Delta\alpha_6$	Δd_6	$\theta_6+\Delta\theta_6$

故包含机器人误差参数后的变换矩阵可表示为

$$ \begin{aligned} {}_i^{i-1}\boldsymbol{T}^* = {} & \mathbf{Rot}_z(\theta_i+\Delta\theta_i)\cdot\mathbf{Trans}_z(d_i+\Delta d_i)\cdot \\ & \mathbf{Trans}_x(a_i+\Delta a_i)\cdot\mathbf{Rot}_x(\alpha_i+\Delta\alpha_i) \\ & i=1,2\cdots,6 \end{aligned} \tag{3} $$

2　参数标定方法

2.1　标定原理

传统的机器人参数标定通常需要借助激光跟踪仪、三坐标测量仪等昂贵且操作复杂的工业级设备完成。本文提出了一种创新性的简易标定方法,利用价格低廉且操作简易的激光位移传感器对立方体标定物外部信息进行间接测量,利用多个测量点进行多平面拟合和多平面角度约束,通过已建立的参数误差模型优化求解关节参数误差。

其实验方案如图 2 所示。激光位移传感器安装于机械臂末端,机械臂操作空间放置两两垂直的立方标定板。激光位移传感器发射出激光束照射在标定板上,通过变换机械臂位姿,采集 3 块标定板平面上多点测量数据。

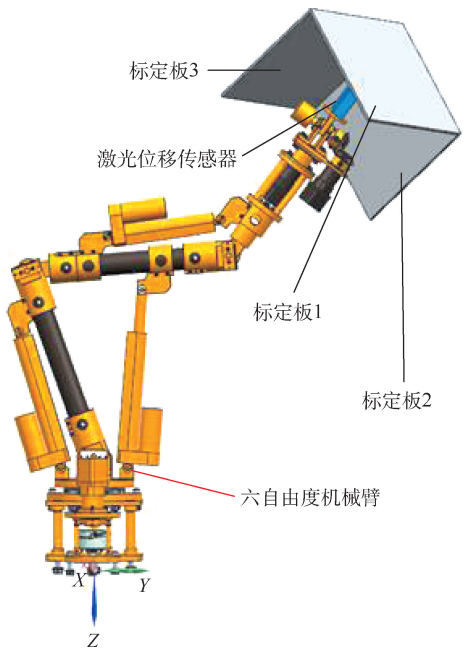


图 2　实验方案

Fig. 2　Experimental scheme

具体测量方案原理如图 3 所示,激光位移传感器的末端参考点到标定板平面测量点之间的距离用 L 描述,即为激光位移传感器读数。机器人末端坐标系为 $O_f x_f y_f z_f$,机械臂基坐标系至末端坐标系的位姿变换矩阵为 ${}^w_t\boldsymbol{T}$,激光位移传感器末端参考点在 $O_f x_f y_f z_f$ 下坐标向量为 $\boldsymbol{p}(x_0, y_0, z_0)$,单位方向向量为 $\boldsymbol{n}(n_x, n_y, n_z)$。

根据激光位移传感器的读数,可以得到 3 个平面上的测量点相对于激光位移传感器坐标系的位置信息。通过最小化测量点在机械臂基坐标系下 3 个平面约束和 3 个平面之间角度约束的误差,可求得机器人 DH 优化参数。

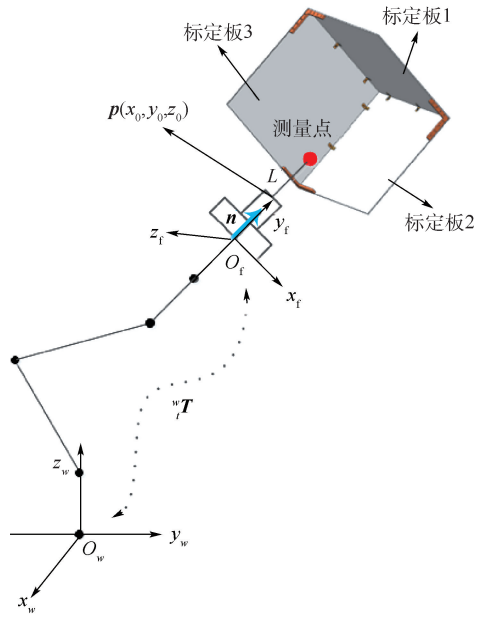


图 3　参数标定原理

Fig. 3　Schematic diagram of parameter calibration

2.2　标定方法

由图 3 可知,标定板上的测量点在末端坐标系 $O_f x_f y_f z_f$ 下的坐标 $\boldsymbol{q}$ 可描述为

$$\boldsymbol{q} = \boldsymbol{p} + L \cdot \boldsymbol{n} \tag{4}$$

式中:$\boldsymbol{q} = (x, y, z)^T$,则根据机器人末端坐标系与基坐标系之间的齐次变换关系可得

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = {}^w_t\boldsymbol{T} \begin{bmatrix} x_0 + Ln_x \\ y_0 + Ln_y \\ z_0 + Ln_z \\ 1 \end{bmatrix} \tag{5}$$

由于在实际的测量过程当中,激光位移传感器实际安装也存在误差,即 L 与 $\boldsymbol{n}$ 值并不完全准确,为此定义实际安装参数为 $L + \Delta L$ 和 $\boldsymbol{n} + \Delta\boldsymbol{n}$。综合考虑表 2 中 24 个 DH 误差参数及 ΔL、$\Delta\boldsymbol{n}$,系统需要修正的总误差参数个数为 28。实际测量点 $\boldsymbol{q}^*$ 的齐次坐标可表示为

$$\begin{bmatrix} x^* \\ y^* \\ z^* \\ 1 \end{bmatrix} = {}^w_t\boldsymbol{T}^* \begin{bmatrix} x_0 + (L + \Delta L)(n_x + \Delta n_x) \\ y_0 + (L + \Delta L)(n_y + \Delta n_y) \\ z_0 + (L + \Delta L)(n_z + \Delta n_z) \\ 1 \end{bmatrix} \tag{6}$$

记标定板 1、2、3 上的测量点集为 Q_1^*、Q_2^* 和 Q_3^*,分别包含 N_1、N_2、N_3 个测量点。记第 i 块标定板上第 l 个测量点坐标为 $\boldsymbol{q}_{li}^* = (x_{li}^*, y_{li}^*, z_{li}^*)^T$。定义 3 块标定板所在平面在机器人基坐标下平面方程分别描述为

$$k_i x + b_i y + c_i z + 1 = 0 \qquad i = 1,2,3 \tag{7}$$

式中:k_i、b_i、c_i 为平面方程系数。

对于 3 个平面的 N_i 个点分别采用最小二乘

法进行平面方程拟合,即优化平面方程系数使式(8)中 S 为最小:

$$S(A,B,C) = \sum_{l=1}^{N_l} (Ax_{li}^* + By_{li}^* + C - z_{li}^*)^2 \quad (8)$$

记所得最优系数为 A^*、B^*、C^*,则可得式(7)中的平面方程系数为

$$\begin{cases} k_i = -\dfrac{A}{C} \\ b_i = -\dfrac{B}{C} \\ c_i = \dfrac{1}{C} \end{cases} \quad (9)$$

计算测量点到拟合平面误差为

$$\Delta d_i = \frac{1}{N_i}\sum_{l=1}^{N_i} \frac{|k_i x_{li}^* + b_i y_{li}^* + c_i z_{li}^* + 1|}{\sqrt{k_i^2 + b_i^2 + c_i^2}} \quad (10)$$

为了保证平面拟合度准确性标定点平面约束给出的误差可定义为

$$\Delta e_1 = \sum_{i=1}^{3} \Delta d_i \quad (11)$$

另外,3 个拟合平面两两之间的夹角余弦值为

$$\cos\beta_{ij} = \frac{|k_i a_j + b_i b_j + c_i c_j|}{\sqrt{k_i^2 + b_i^2 + c_i^2}\sqrt{k_j^2 + b_j^2 + c_j^2}} \quad (12)$$

式中:i 和 j 为 2 个不同平面的序号。设标定板的实际安装角度为 β_{12}^*、β_{23}^*、β_{13}^*,β_{12}、β_{23}、p_{13} 为两平面夹角的拟合值,则标定板之间的角度约束误差定义为

$$\Delta e_2 = (\arccos\beta_{12} - \beta_{12}^*) + (\arccos\beta_{13} - \beta_{13}^*) + (\arccos\beta_{23} - \beta_{23}^*) \quad (13)$$

同时考虑标定板平面及标定板间角度约束,建立关于激光位移传感器参数标定的误差函数,如式(14)所示,其变量包括 24 个 DH 误差参数及 4 个激光位移器安装误差参数。

$$e = (1-\mu)\Delta e_1 + \mu\Delta e_2 \quad (14)$$

式中:μ 为平面和角度影响因素的权重参数;误差 e 为标定参数的非线性函数,采用能够解决对多约束和多变量结合的非线性问题的全局最优解求解的算法 OQNLP[21] 进行数值优化并调整最优值,即可得到参数标定结果。

3 标定实验及结果分析

3.1 基于激光位移传感器的标定实验

标定实验场景如图 4 所示,标定实验系统主要由六自由度机械臂、基恩士 LK-H080 型号激光位移传感器和标定板组成。该型号激光传感器测量量程为 62 ~ 98 mm,绝对定位精度为 0.01 mm,重复定位精度可达 0.3 μm。

为确定其安装夹角,利用了光学动作捕捉系统对 3 块标定板上放置的反光标志点进行三维空间精确捕捉,如图 5 所示。该系统镜头像素为 400 万,三维定位精度为 ±0.1 mm。通过拟合每个平面求出标定板间夹角测量值 β_{12}^* 为 89.94°,β_{23}^* 为 89.97°,β_{13}^* 为 90.2°。已知 3 块标定板平面加工精度为 0.1 mm。则根据当前标定板的尺寸其角度误差为 0.15°。

基于激光位移传感器标定实验流程主要如下:

1) 将激光位移传感器固定在机械臂末端,将标定板固定于末端运动范围内并保证其平面位置处于激光位移传感器测量量程内。

2) 开启激光位移传感器,将激光点照射于标定板 1 上,记录激光位移传感器读数值 L_i 和机械臂两关节连杆夹角值 θ_i。调整机械臂位姿,记录标定板 1 不同位置的测量值,尽量使测量激光点在标定板上均匀分布。对标定板 2 与 3 重复此步骤。

3) 将数据值导入标定误差模型中,利用非线性优化算法最小化目标函数式(14),求解误差参数最优值。

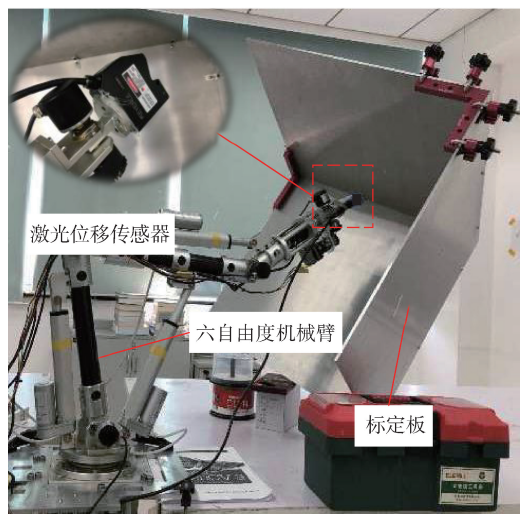


图 4 激光位移传感器标定实验场景

Fig. 4 Experimental scene of laser displacement sensor calibration

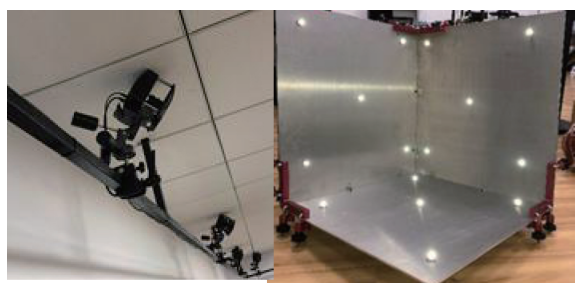


图 5 标定板夹角求解实验场景

Fig. 5 Experimental scene of calibration plate angle solution

3.2 基于激光跟踪仪的对照标定实验

作为对比验证,本文进行了激光跟踪仪的参

数标定实验。实验中使用的设备为美国 API 公司生产的 T3 型激光跟踪仪,激光跟踪距离为 0 ~ 60 m,静态测量精度为 5 μm/m,动态测量精度为 110 μm/m。使用激光跟踪仪可测得机械臂末端相对于基坐标系 $\{B\}$ 的三维坐标,详细过程请参照文献[22]。建立目标函数 F 为末端三维坐标的计算值与测量值误差和,如式(15)所示,其变量为 24 个 DH 误差参数。同样利用非线性数值优化得出使目标函数模型误差最小的 DH 误差参数,即可得到标定结果。

$$F = \sum_{i=1}^{M_i} [(x_i - x_i')^2 + (y_i - y_i')^2 + (z_i - z_i')^2]$$
$$i = 1,2,\cdots,M_i \tag{15}$$

式中:x_i、y_i、z_i 为第 i 个点的计算坐标;x_i'、y_i'、z_i'为其测量坐标;M_i 为测量总点数。

基于激光跟踪仪的机器人参数标定实验场景如图 6 所示。标定实验流程主要如下:

1) 建立机器人基座坐标系,将激光跟踪仪测量坐标系更改为机器人坐标系。

2) 转动机器人各关节,记录每个时刻的机械臂两关节连杆夹角值 θ_i,并记录每个时刻激光跟踪仪读数x_i'、z_i'。

3) 将数据值导入标定误差模型中,利用非线性优化算法最小化目标函数求解参数标定最优值。



图 6　激光跟踪仪标定实验场景

Fig. 6　Experimental scene of laser tracker calibration

3.3　实验结果

完全未校核的机器人其名义参数值常大大偏离实际值,因此,通常需要对其进行多次迭代以减少初始偏差对标定优化精度的影响[23]。在本文实验工作中,对标定过程进行了二次迭代。本节主要讨论后一次的精准校正过程和结果。

由 2.2 节可知,基于激光位移传感器的参数标定实验,评估目标函数误差 e 求解精度的指标有2 个:①激光位移传感器测量点到拟合平面的平均距离误差 Δe_1;②拟合所得 3 个平面与实际标定板平面平均角度误差 Δe_2。权重 μ 取值一定程度上影响参数优化精度,在不同的 μ 值下,优化后平面误差及角度误差如表 3 所示。μ 值越小,越倾向于优化角度误差。优选结果应考虑实验条件下作为合理的误差分配参考。如 3.1 节所述,本实验条件标定板平面度误差为 0.1 mm,角度测量误差为 0.15°。据此应选取 μ 为 0.07 时的标定结果,得到 $e_1 = 1.08$ mm,$e_2 = 0.67°$,分别为标定物自身误差的 10 倍和 4.5 倍,为有效标定误差。相应的 DH 参数标定结果如表 4 所示。

绘制相应的拟合平面,如图 7 所示,红点为根据优化结果进行 DH 参数补偿后的激光位移传感器测量点。可以看出,标定补偿后的三维点基本位于标定平面上,且拟合所得 3 个平面基本两两垂直,符合实验预期。

基于激光跟踪仪的参数标定对照实验结果如图 8 所示。使用经过标定的参数后,末端点三维位置基本与跟踪仪测量量吻合,其平均误差为 1.01 mm。

表 3　权重 μ 取值

Table 3　Values of weight μ

μ	e_1/mm	e_2/(°)
0.01	1.04×10^{-5}	67.04
0.02	0.009 9	13.41
0.07	1.08	0.67
0.71	1.2	0.034

表 4　DH 参数变化值

Table 4　Values of DH parameter change

关节 i	Δa_i/mm		$\Delta\alpha_i$/(°)		Δd_i/mm		$\Delta\theta_i$/(°)	
	粗标定	精标定	粗标定	精标定	粗标定	精标定	粗标定	精标定
1	-0.19	0.92	-4.52	0.45	-4.40	0.94	-3.35	-0.89
2	3.09	0.29	5.92	0.04	-2.41	0.92	-1.99	0
3	0.16	0.19	3.59	0.89	-2.46	0.59	-0.79	0.04
4	-10.24	0.07	1.78	0.62	-2.57	0.4	3.142 3	0.78
5	5.84	0.51	-2.56	0.06	-19.35	0	-2.11	0.4
6	0.95	0.18	5.13	0.33	1.35	0.05	-1.73	0.26

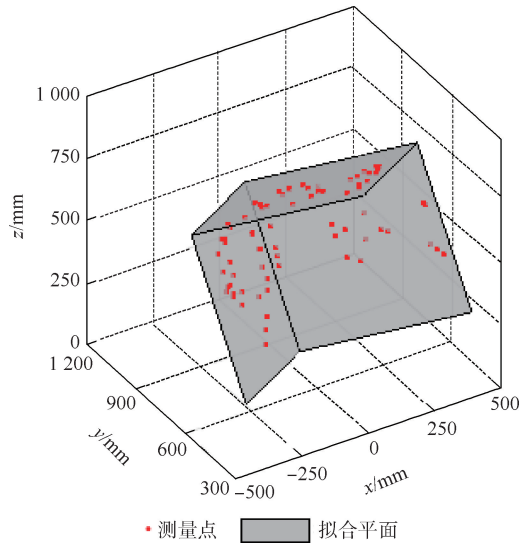


图 7 拟合平面结果

Fig. 7 Graph of fitted plane

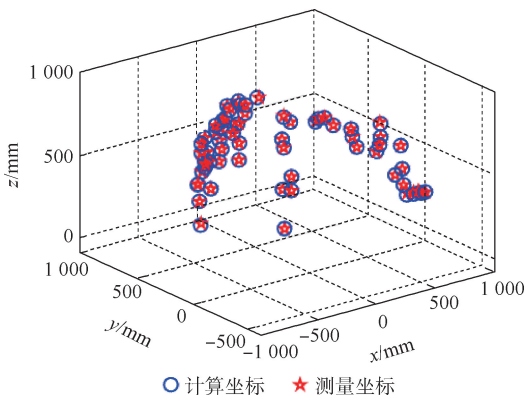


图 8 基于跟踪仪标定后计算坐标和测量坐标对比

Fig. 8 Comparison of calculated coordinates and measurement coordinates after tracker-based calibration

3.4 误差分析与讨论

本文从 2 个方面对标定误差进行讨论:①基于激光位移传感器的标定与激光跟踪仪标定精度的比较;②基于激光位移传感器的标定误差与系统重复定位精度的比较。

激光位移传感器与激光跟踪仪在各自 53 组标定实验数据中的距离误差分布情况如图 9 所示,其误差数据对比如表 5 所示。从图 9 和表 5 可知,使用激光跟踪仪平均误差为 1.01 mm,标定后定位误差减小了 72.48%;激光位移传感器平均误差为 1.08 mm,标定后误差减小了 71.72%。两者达到了相似的标定结果。若需进一步提升使用激光位移传感器的标定精度,可提高标定板的加工平面度与平面夹角测量精度以增加标准参照物的尺寸精度。

另外,为了衡量最终标定误差范围相对于实验系统自身精度的合理性,对作为实验对象的六自由度机械臂系统的重复定位误差进行了测量,

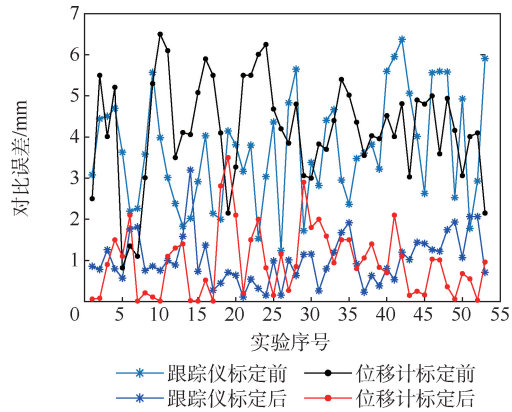


图 9 标定误差对比

Fig. 9 Comparison of calibration errors

表 5 标定实验误差值对比

Table 5 Comparison of error values of calibration experiment

实验类别	误差平均值/mm	最大值/mm	最小值/mm	方差/mm^2
跟踪仪标定前	3.67	6.37	1.24	1.34
跟踪仪标定后	1.01	3.19	0.11	0.36
位移计标定前	3.82	6.52	0.82	1.27
位移计标定后	1.08	3.45	0.01	0.39

得到平均重复定位误差为 0.74 mm。机器人最终标定误差主要由其固有重复定位误差和标定后残余的运动学参数误差两部分组成。由表 5 标定结果可知,由残余参数误差所致的定位误差仅占系统绝对定位总误差的 31.5%,为系统重复定位误差的 50%。

4 结 论

本文针对传统机械臂参数标定方法,设备昂贵、技术实现成本高的问题,设计了一种低成本、易操作的利用机械臂末端激光位移计作为测量手段的机械臂参数标定方法。通过实验验证,得到如下结论:

1)设计了立方体面作为外参标定物,利用非接触式激光测量手段,相对接触式测量更为便捷安全,相对视觉测量更为稳定简易。

2)通过构建立方体平面约束及平面间角度约束设计误差函数,利用非线性优化算法最小化目标函数求解可求得误差参数最优值。

3)实验结果表明,本文提出的方法能够达到与使用昂贵的激光跟踪仪相比拟的标定精度,达到良好的标定效果。

参考文献(References)

[1] 高涵,张明路,张小俊,等. 基于最优测量结构选择的机械

臂参数标定[J]. 仪器仪表学报,2018,39(9):1-9.
GAO H,ZHANG M L,ZHANG X J,et al. Manipulator parameter calibration based on optimal measurement structure selection [J]. Chinese Journal of Scientific Instrument,2018,39(9):1-9 (in Chinese).

[2] GATTI G,DANIELI G. A practical approach to compensate for geometric errors in measuring arms:Application to a six-degree-of-freedom kinematic struct-ure[J]. Measurement Science and Technology,2007,19(1):015107.

[3] 陈启军,周自强,刘成菊,等. 基于视觉的六自由度机械臂运动学参数辨识[J]. 同济大学学报(自然科学版),2020,48(7):1023-1029.
CHEN Q J,ZHOU Z Q,LIU C J,et al. Kinematics parameter identification of six-degree-of-freedom manipulator based on vision[J]. Journal of Tongji University (Natural Science Edition),2020,48(7):1023-1029(in Chinese).

[4] 谭月胜,孙汉旭,贾庆轩,等. 基于旋量理论及距离误差的机械臂标定新方法[J]. 北京航空航天大学学报,2006,32(9):1104-1108.
TAN Y S,SUN H X,JIA Q X,et al. New manipulator calibration method based on screw theory and distance error[J]. Journal of Beijing University of Aeronautics and Astronautics,2006,32(9):1104-1108(in Chinese).

[5] HE J,DING Q,GAO F,et al. Kinematic calibration methodology of hybrid manipulator containing parallel topology with main limb[J]. Measurement,2019,152:107334.

[6] 张铁,戴孝亮,杜亮. 基于距离测量的机器人误差标定及参数选定[J]. 北京航空航天大学学报,2014,40(5):585-590.
ZHANG T,DAI X L,DU L. Robot error calibration based on distance measurement with parameter selection[J]. Journal of Beijing University of Aeronautics and Astronautics, 2014,40(5):585-590(in Chinese).

[7] HE S,MA L,YAN C,et al. Multiple location constraints based industrial robot kinematic parameter calibration and accuracy assessment[J]. International Journal of Advanced Manufacturing Technology,2019,102(5):1037-1050.

[8] DU G L,ZHANG P. IMU-based online kinematic calibration of robot manipulator[J]. The Scientific World Journal, 2013,2013:139738.

[9] 陈宵燕,张秋菊,孙沂琳. 串联机器人多模式标定与刚柔耦合误差补偿方法研究[J]. 农业机械学报,2019,50(3):396-403.
CHEN X Y,ZHANG Q J,SUN Y L. Multi-mode calibration and rigid-flexible coupling error compensation method of serial robot [J]. Transactions of the Chinese Society for Agricultural Machinery,2019,50(3):396-403(in Chinese).

[10] 朱齐丹,谢心如,夏桂华,等. 基于光轴约束的机械臂运动学标定方法[J]. 哈尔滨工程大学学报,2019,40(3):433-439.
ZHU Q D,XIE X R,XIA G H,et al. Kinematics calibration method of mechanical arm based on optical axis constraint[J]. Journal of Harbin Engineering University,2019,40(3):433-439(in Chinese).

[11] YANG P,GUO Z,KONG Y. Plane kinematic calibration method for industrial robot based on dynamic measurement of double ball bar[J]. Precision Engineering,2020,62:265-272.

[12] 李永泉,王皓辰,张阳,等. 一种基于手眼视觉的并联机器人标定方法[J]. 中国机械工程,2020,31(6):722-730.
LI Y Q,WANG H C,ZHANG Y,et al. A calibration method of parallel robot based on hand eye vision[J]. China Mechanical Engineering,2020,31(6):722-730(in Chinese).

[13] PARK I W,LEE B J,CHO S H,et al. Laser-based kinematic calibration of robot manipulator using differential kinematics [J]. IEEE/ASME Transactions on Mechatronics, 2012, 17(6):1059-1067.

[14] JOUBAIR A,BONEV I A. Kinematic calibration of a six-axis serial robot using distance and sphere constraints[J]. International Journal of Advanced Manufacturing Technology,2015, 77(1-4):515-523.

[15] JOUBAIR A,BONEV I A. Non-kinematic calibration of a six-axis serial robot using planar constraints[J]. Precision Engineering,2015,40(1):325-333.

[16] 吴孟桦,胡晓兵,赵清祥,等. 基于双结构光视觉的机器人运动学参数标定方法[J]. 组合机床与自动化加工技术,2020(6):37-40.
WU M H,HU X B,ZHAO Q X,et al. Robot kinematics parameter calibration method based on dual structured light vision [J]. Modular Machine Tool and Automatic Manufacturing Technology,2020(6):37-40(in Chinese).

[17] SLOCUM K,HAYES M J D,LEITNER M,et al. Kinematic calibration of six-axis robots[C]// Proceedings of the CSME Forum 2002,2002:149-152.

[18] HADGE H,BIDAUD P,JARDIN N. Practical consideration on the identification of the kinematic parameters of the Staubli TX90 robot[C]//13th World Congress in Mechanism and Machine Science,2011:43.

[19] 杨源,曹彤,刘达. 一种利用标定板的机械臂 DH 参数标定方法[J]. 机械传动,2017,41(6):187-190.
YANG Y,CAO T,LIU D. A DH parameter calibration method of a manipulator using a calibration plate [J]. Mechanical Transmission,2017,41(6):187-190(in Chinese).

[20] COLSON J C,PERREIRA N D. Kinematic arrangements used in industrial robots[C]// Proceedings of the 13th International Symposium on Industrial Robots,1983:540-544.

[21] UGRAY Z,LASDON L,PLUMMER J,et al. Scatter search and local NLP solvers:A multistart framework for global optimization [J]. Informs Journal on Computing,2007,19(3):328-340.

[22] 任永杰,邾继贵,杨学友,等. 利用激光跟踪仪对机器人进行标定的方法[J]. 机械工程学报,2007,43(9):195-200.
REN Y J,CHU J G,YANG X Y,et al. Method of calibrating robot using laser tracker[J]. Chinese Journal of Mechanical Engineering,2007,43(9):195-200(in Chinese).

[23] 董慧颖,李文广. 一种基于平面精度的机器人标定方法及仿真[J]. 中国机械工程,2011,22(17):2039-2042.
DONG H Y,LI W G. A robot calibration method and simulation based on plane accuracy[J]. China Mechanical Engineering, 2011,22(17):2039-2042(in Chinese).

A parameter calibration method for manipulators based on laser displacement measurement

SHAO Xin[1], JI Li[1], ZOU Huaiwu[2], XIE Yangmin[1,*]

(1. School of Mechatronics Engineering and Automation, Shanghai University, Shanghai 200444, China;
2. Shanghai Key Laboratory of Spacecraft Mechanism, Shanghai 201108, China)

Abstract: The traditional method of robot parameter target usually requires an expensive equipment, and the technical realization cost is high, a novel low-cost calibration method for manipulators is proposed, which utilizes a laser displacement sensor installed on the end of the manipulator to measure the relative position of the manipulator end with respect to an external cubic reference. A kinematic parameter calibration is designed for a 6 degree of freedom manipulator. Based on single-dimensional measurements, the error function is constructed with combined consideration of plane flatness and plane angles, and further minimized with a nonlinear optimization method to obtain the final calibration results. The proposed method is testified on an experimental system and compared with the calibration by a laser tracker. Results show that this method has similar calibration accuracy to the methods using expensive external measurement devices and provides a handy and cheap solution to manipulator calibration.

Keywords: manipulators; kinematic parameter calibration; laser displacement measurement; combined constraints; nonlinear optimization

Received: 2021-02-24; **Accepted**: 2021-06-13; **Published online**: 2021-07-12 15:13

URL: kns.cnki.net/kcms/detail/11.2625.V.20210712.1337.003.html

Foundation items: Shanghai Natural Science Foundation (19ZR1455500); Shanghai Science and Technology Committee Rising-Star Program (19QA1403500); Open Project of Shanghai Key Laboratory of Spacecraft Mechanism (YYHT-F805201808013)

* **Corresponding author**. E-mail: xieym@shu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0095

基于障碍物代价势场的移动机器人避障算法

迟胜凯, 谢永芳, 陈晓方*, 彭帆
(中南大学 自动化学院, 长沙 410083)

摘 要: 移动机器人所处的环境通常是动态的,机器人需要及时做出响应,同时保证路径的平滑度及与障碍物间的安全距离。针对此问题,提出了一种基于障碍物代价势场的移动机器人动态避障算法。通过建立静态栅格地图及障碍物的代价势场,获得动态场景下的等势线及经过起点、终点的切线,求解最小生成树获得初始候选路径,针对路径的长度、障碍物距离及平滑度对候选路径锚点进行调整。通过引入障碍物速度对代价势场的影响,使得机器人能对移动中的障碍物做出及时的响应。为验证所提算法的有效性,在分辨率为 1 200×1 000 m 的栅格场景下分别对静态场景和动态场景进行仿真,结果表明:所提算法能够在保证路径具有较高的平滑度且与障碍物间保持安全距离的条件下使路径尽可能得短;同时在动态障碍物场景下依然能保持路径的平滑和避障的安全性,满足动态场景下移动机器人路径规划的要求。

关 键 词: 移动机器人; 路径规划; 代价势场; 动态障碍物; 避障
中图分类号: TP242.6
文献标志码: A **文章编号**: 1001-5965(2022)11-2289-15

随着机器人技术的快速发展,机器人在工业、服务业、安防等领域的应用越来越广泛,机器人也越来越多的出现在生活中。移动机器人最重要的功能是路径规划和避障,导航的目的是找到从起点到终点具有避障能力的最优或次优路径[1]。现实世界的环境是动态变化的,需要不断更新地图中的障碍物信息,使机器人在现实环境中实现灵活的自主移动。因此,在保证路径规划最优或次优的前提下,如何让机器人对场景中的静态和动态障碍物快速的做出响应,同时保证路径的平滑度和安全性,对于机器人的避障来说非常重要。

机器人所处的环境通常是三维的,三维环境地图[2-3]是描述真实三维世界最直接的方法,在三维地图中可以获得障碍物更多的形状、轮廓、姿态、语义等信息,对于辅助机器人的导航和避障来说意义重大。而基于三维点云地图的三维路径规划[4]具有很高的时间复杂度和空间复杂度,不利于实时避障的实现。因此在二维地图场景下利用三维障碍物信息,实现基于二维地图的导航和路径规划,对于提高机器人避障的灵活性是很有必要的。

栅格法环境建模是移动机器人导航的经典方法,该方法将机器人所处的实际环境分割为大小相同的栅格,每个栅格的不同值对应地图的不同障碍物占用状态[5]。因此在栅格地图下利用三维场景信息及三维传感器数据,可以大大提升移动机器人导航与避障的灵活性与安全性。在栅格地图场景下,常用的路径规划算法有 A* 算法、人工势场法、蚁群算法、快速扩展随机树(rapidly-exploring random tree star, RRT*)算法、遗传算法、D* 算法等[6]。A* 算法是求解最优路径非常有效的算法,但是当路径规划的场景较大时,并且存在障碍物遮挡时,A* 算法的计算量就会非常大,内存占用严重,并且运算时间长,路径规划的效率较

收稿日期: 2021-03-01; **录用日期**: 2021-05-14; **网络出版时间**: 2021-05-26 10:21
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210525.1309.003.html
基金项目: 广东省重点领域研发计划(2021B0101200005); 国家自然科学基金(62133016); 国家杰出青年科学基金(61725306)
* **通信作者**. E-mail: xiaofangchen@csu.edu.cn

引用格式: 迟胜凯, 谢永芳, 陈晓方, 等. 基于障碍物代价势场的移动机器人避障算法[J]. 北京航空航天大学学报, 2022, 48(11): 2289-2303. CHI S K, XIE Y F, CHEN X F, et al. Obstacle avoidance method of mobile robot based on obstacle cost potential field[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2289-2303(in Chinese).

低,赵晓等[7]针对此问题提出了基于跳点搜索的改进算法,通过筛选跳点进行扩展,直到生成最终路径,扩展过程中使用跳点代替 A* 算法中大量可能被添加到有效点集合和无效点集合的不必要节点,从而减少计算量。王洪斌等[5]针对多目标复杂场景,提出了 A* 算法与动态窗口法相结合的混合算法,通过设置多个临时目标点,利用目标成本函数对所有目标进行优先级判定,进而利用改进的 A* 算法规划出一条经过多个目标点的最优路径,一定程度上改善了算法的效率问题,但是对于导航路径的安全性及路径平滑性的问题,依然没有很好的解决。Liu 等[8]采用了改进蚁群算法进行全局路径规划,在搜索全局路径的过程中,将信息素扩散与势场力相结合,使信息素沿障碍物势场力的方向扩散,提高了蚁群算法的收敛速度,但蚁群算法为找到全局最优路径,搜索空间必须足够大,会消耗大量的内存空间,依然存在最优路径搜索效率低的问题。林依凡等[9]提出了一种无碰撞检测改进 RRT* 运动规划算法,针对 RRT* 算法碰撞检测过程中耗时过长的问题,提出使用在代价函数中增加碰撞风险评估函数的方法取代所有的碰撞检测计算,提高了算法的寻路效率,但是并没有改善导航的路径平滑性不好、算法不稳定的特点,且没有考虑与障碍物的安全距离。薛学华等[10]提出了一种改进人工势场法,针对局部最优值问题提出了一种自适应参数调整回溯法,解决了传统人工势场的路径不可达与局部最优值问题,通过人工势场法与可视图法切换的策略克服了人工势场目标不可达问题,但是该算法同样没有考虑导航路径的安全性及平滑性,且得到的路径长度通常不是最优的。

Zafar 和 Mohanta[6]对目前现有的移动机器人的路径规划和避障算法及改进方案进行了研究和对比,现有研究在解决同一场景下机器人避障问题时只考虑了单一的路径评价指标,如只考虑路径最短、平滑度最优、耗能最少或算法的效率最高,在单一指标上能够达到较优的效果,因此只能适用一些特定的场景。以往研究的另一个局限是,在进行静态或动态环境下局部路径规划过程中,没有考虑机器人的大小,也没有考虑移动中的障碍物对于移动机器人避障的影响。而在一些工业应用场景下,不仅要求路径尽可能地短,出于安全考虑需要机器人和某些障碍物保持安全距离,并且要求路径具有较高的平滑度。

基于以上原因,本文提出了一种基于障碍物代价势场的动态路径生成及路径调整算法。根据激光雷达、深度相机等传感器数据获取动态障碍物运动状态信息。然后由栅格地图构建障碍物的代价势场,根据机器人的形状、大小及其他避障参数选择某一合适高度的等势曲线,通过求解由机器人当前位置、目标点、等势线及其切线得到的最小生成子树,获得初始候选路径。对初始候选路径采样获得初始候选路径锚点。并在代价势场下对初始候选路径锚点的障碍物代价和路径长度代价进行调整,得到最终的路径。算法在保证路径和障碍物的安全距离以及较好的路径平滑度的同时,能够使规划得到的路径尽可能地短。对于动态移动的障碍物,通过引入障碍物速度对代价势场的影响,使得在移动障碍物场景下也能获得较好的结果。

1 机器人避障问题建模

1.1 机器人模型

在平面环境中,差动转向移动机器人的机械结构简单、运动灵活、稳定性高,是目前较为流行的地面移动机器人。本文假设机器人装有稳定的底层控制系统,可以通过陀螺仪、激光雷达、里程计等传感器数据完成对机器人速度、姿态的跟踪。设 E-N 为机器人所处的世界坐标系,X-Y 为机器人本体的坐标系,如图 1 所示,机器人的当前状态可以用一个向量 $[e,n,\theta_w]$ 表示,其中 $[e,n]$ 表示机器人在世界坐标系中的坐标,θ_w 表示机器人的航向角。设机器人左右两侧驱动轮距离为 W,驱动轮直径为 L,右轮和左轮角速度分别为 $\dot{\omega}_R$ 和 $\dot{\omega}_L$,机器人的运动状态可以表示为

$$\begin{cases}\dot{x} = \dfrac{L}{2}(\dot{\omega}_L + \dot{\omega}_R)\\ \dot{y} = 0\\ \dot{\theta} = \dfrac{L}{W}(\dot{\omega}_L - \dot{\omega}_R)\end{cases} \tag{1}$$

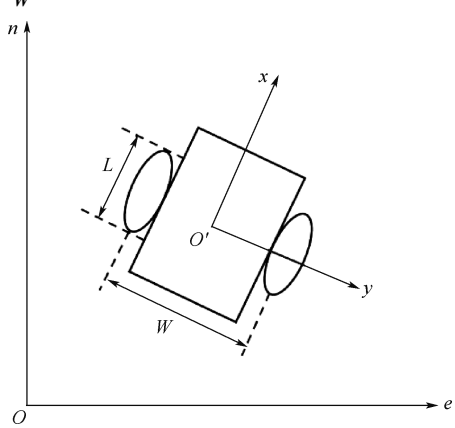


图 1 差动转向机器人的俯视示意图

Fig. 1 Top view of differential steering robot

式中:$\dot{x}$、$\dot{y}$ 分别为在本体坐标系下机器人沿 x 轴、y 轴的线速度;$\dot{\theta}$ 为机器人转向的角速度。假设 $\boldsymbol{T}_0$ 为机器人坐标系到世界坐标系的坐标转换矩阵,则机器人在世界坐标系下的状态可以由式(2)坐标变换矩阵得到

$$\begin{bmatrix} \dot{e} \\ \dot{n} \\ \dot{\theta}_w \end{bmatrix} = \boldsymbol{T}_0 \cdot \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} \tag{2}$$

1.2　动态地图模型

虽然机器人只会在一个二维的平面内运动,但机器人所处环境是三维的,而直接在三维地图下进行路径规划和避障,会使计算的复杂度大大增加,不利于实时的实现和大尺度环境下的应用。为了在导航过程中充分利用三维数据,实现本文的避障算法,通常基于多线激光雷达数据实现动态地图的更新,通过建立由三维点云数据到二维占用栅格地图的动态映射来更新动态障碍物地图信息。由本地机器人在动态二维地图的基础上完成避障任务。

机器人在进行导航和避障之前会建立一个全局场景的静态栅格地图,假设这个初始场景的栅格地图不会发生变化。机器人在运动过程中会对激光雷达产生的点云数据进行实时处理,根据具体机器人的大小和尺寸进行滤波去除无效的点云信息,利用可能会与机器人发生碰撞的障碍物有效点云信息更新对应的障碍物占用栅格地图[11](occupancy grid map),得到针对不同动态障碍物的动态概率栅格地图,即不同栅格中存在障碍物的概率。根据占用栅格地图中的障碍物表示方法,对于地图中的任意栅格,定义栅格状态算子 s,$p(s=0)$ 表示该栅格被障碍物占用的概率,使用 $p(s=1)$ 表示该栅格不被障碍物占用的概率,二者的和为 1,使用二者的比值 $\mathrm{Odd}(s)=p(s=1)/p(s=0)$ 描述当前栅格的障碍物占用状态,初始状态下默认 $p(s=1)=p(s=0)$。

在地图更新过程中假设不同栅格间的状态变化互不影响,对于地图中的任一栅格单元,假设在 t 时刻之前的障碍物占用状态为 $\mathrm{Odd}_{t-1}(s)$,在 t 时刻有新的测量值 z_t,其在 t 时刻的障碍物占用状态按式(3)进行计算,$\mathrm{Odd}(s\mid z_t)$ 表示在 z_t 发生的条件下 s 的发生概率。

$$\mathrm{Odd}(s\mid z_t)=\frac{p(s=1\mid z_t)}{p(s=0\mid z_t)} \tag{3}$$

由贝叶斯公式可知

$$\begin{cases} p(s=1\mid z_t)=\dfrac{p(z_t\mid s=1)p(s=1)}{p(z_t)} \\ p(s=0\mid z_t)=\dfrac{p(z_t\mid s=0)p(s=0)}{p(z_t)} \end{cases} \tag{4}$$

由式(4)代入式(3)后得

$$\mathrm{Odd}(s\mid z_t)=\frac{p(z_t\mid s=1)}{p(z_t\mid s=0)}\mathrm{Odd}(s) \tag{5}$$

同时对式(5)两边取对数可得

$$\ln \mathrm{Odd}(s\mid z_t)=\ln\left(\frac{p(z_t\mid s=1)}{p(z_t\mid s=0)}\right)+\ln \mathrm{Odd}(s) \tag{6}$$

因此,只需根据当前传感器测量值对 $\ln(p(z_t\mid s=1)/p(z_t\mid s=0))$ 进行更新,即可实现障碍物地图数据的更新,记 l_t 为 t 时刻栅格单元的对数占用状态式,l_{t-1} 为 $t-1$ 时刻栅格单元的对数占用状态式,l_{meas} 为传感器测量值的对数描述形式,则可以简化为 $l_t=l_{t-1}+l_{\mathrm{meas}}$。测量值的模型有 2 种状态,分别为 free 和 occupied,带迭代更新过程中都是定值。这样,即可基于占用栅格地图对动态障碍物进行描述,l_t 的值越大,则表示对应栅格存在障碍物的概率越大。在此基础设 ρ 为更新后得栅格障碍物概率描述式,$C_{\max}$ 表示地图中允许的栅格代价最大值,为一个常数。则当栅格的概率描述值 $l_t<0$ 时,$\rho=0$。当 $0\leqslant l_t<C_{\max}$ 时,$\rho=l_t/C_{\max}$,当 $l_t>C_{\max}$ 时,$\rho=1$。

假设本地机器人的全局静态栅格地图是由 m_1 行 m_2 列的栅格组成,则全局静态栅格地图可以用一个二维矩阵 $\boldsymbol{M}(x,y)$ 表示,栅格障碍物概率为

$$\boldsymbol{M}(x,y)=\begin{cases} 0 & (x,y)\text{ 处没有障碍物} \\ \rho & 0<\rho<1,\text{存在障碍物概率为 }\rho \\ 1 & (x,y)\text{ 处存在障碍物} \end{cases} \tag{7}$$

则由激光雷达数据得到的局部动态障碍物地图可以表示为 $M'(x,y)$(其中 x 与 y 满足条件 $0<x<m_1$,$0<y<m_2$)。在执行动态避障任务时局部动态障碍物地图会和全局静态地图进行叠加。

1.3　静态障碍物的表示

本文基于三维传感器数据得到了机器人对应的二维概率栅格地图,提出了基于障碍物代价势场的机器人障碍物表示算法。静态障碍物栅格的占用情况由栅格代价值 C 和栅格存在障碍物概率 ρ 这 2 个参数描述。先需要建立全局静态栅格地图的代价势场。设 $\rho_{op}=\rho$ 为静态障碍物代价势场中的栅格障碍物概率,其中 ρ 为由占用栅格地图更新得到的障碍物概率,下标 p 为距离当前栅

格最近且障碍物概率不为 0 的栅格、o 为当前栅格。对于障碍物概率 $\rho\neq 0$ 的栅格,其障碍物代价 $U_p = C_{\max}\rho_{op}$。对于障碍物概率 $\rho = 0$ 的栅格,障碍物代价为

$$U_p = \begin{cases} C_{\max}\rho_{op}\dfrac{1}{(1+d_{op})^2} & d_{op} \leqslant d_{\max},\rho = 0 \\ 0 & d_{op} > d_{\max},\rho = 0 \end{cases} \tag{8}$$

式中:若 $p(x_p,y_p)$ 为距离当前栅格最近且障碍物概率不为 0 的栅格坐标,ρ_{op} 为栅格 p 的障碍物概率;d_{op} 为 p 到当前栅格的欧氏距离;$d_{\max}$ 为障碍物栅格所能影响的最远距离,为一个常数。这样可以得到静态栅格地图的代价势场,如图 2 所示,栅格的边长为单位 1,黑色栅格代表障碍物概率 $\rho\neq 0$ 的栅格,带有灰度的栅格为由式(12)计算得到的代价势场。

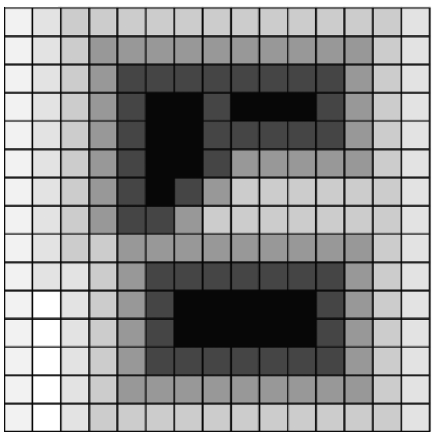


图 2 障碍物附近的代价势场栅格示意图

Fig. 2 Schematic diagram of cost potential field grid near obstacle

1.4 动态障碍物的表示

在实际生产过程中机器人面对的不仅有静态的障碍物,还需要对场景中的动态障碍物做出灵活的响应。由三维场景下的激光数据可以获得动态障碍物相对于机器人的速度。根据机器人在世界坐标系下的姿态 $[e,n,\theta_w]$,可以得到当前时刻动态障碍物坐标系相对于世界坐标系下的速度。定义 $\boldsymbol{O}_{\text{obs}}[x,y]$ 和 $\boldsymbol{v}$ 分别为动态障碍物当前时刻的位置和速度,其中 $[x,y]$ 为动态障碍物对象位置,向量 $\boldsymbol{v}$ 为动态障碍物的移动速度,动态障碍物的移动速度为

$$v = \frac{1}{t}\sqrt{(x-x_a)^2+(y-y_a)^2} \tag{9}$$

式中:$A(x_a,y_a)$ 为上一时刻动态障碍物的坐标;t 为 2 次坐标更新的时间差。设 $B(x_b,y_b)$ 为地图中障碍物概率 $\rho = 0$ 的任一栅格坐标,$p(x_p,y_p)$ 为动态障碍物 $\boldsymbol{O}_{\text{obs}}$ 中距离点 B 最近且 $\rho\neq 0$ 的栅格坐标。令向量 $\boldsymbol{d} = (x_b - x_p, y_b - y_p)$,障碍物速度对代价势场的影响为

$$U_v = \begin{cases} C_{\max}\rho_{op}\dfrac{\lambda\|\boldsymbol{v}\|\varphi}{\|\boldsymbol{d}\|^2} & d_{op} \leqslant d_{\max},\varphi \geqslant 0 \\ 0 & d_{op} > d_{\max} \text{ 或 } \varphi < 0 \end{cases} \tag{10}$$

式中:λ 为障碍物速度对代价势场的影响系数,值越大表示障碍物速度对障碍物代价势场的影响越大;ρ_{op} 为栅格 p 的障碍物概率;φ 为向量 $\boldsymbol{v}$ 和向量 $\boldsymbol{d}$ 夹角的余弦值,即

$$\varphi = \frac{\boldsymbol{v}\cdot\boldsymbol{d}}{\|\boldsymbol{v}\|\times\|\boldsymbol{d}\|} \tag{11}$$

由式(8)和式(10)可得,动态障碍物场景下的障碍物代价为

$$U = U_p + U_v \tag{12}$$

2 路径生成与避障

2.1 切线的计算

为了计算避障过程中的初始候选路径,根据第 1 节本文算法计算栅格地图中的障碍物代价势场。如图 3 所示,在栅格地图中,设点 R 为机器人的当前位置,黑色栅格表示障碍物概率不为 0 的区域,白色栅格表示障碍物概率为 0 的区域,由式(12)更新障碍物代价势场,然后根据具体机器人的形状、大小、避障规则等选择合适高度的等势曲线。图 3 中红色曲线为障碍物的一条等势曲线,由于在栅格地图中长度的计算是不连续的,栅格地图中的等势曲线可以用图 3 中的绿色栅格表示。为方便计算,设点 A 为等势曲线的起点,按顺时针方向,等势曲线可以由一系列离散的栅格 $a_1,a_2,\cdots,a_{n_e}$ 表示,其中 n_e 为等势曲线包含栅格总个数。其长度 l_{n_e} 按如下方法计算:

$$l_{n_e} = \sum_{i=1}^{n_e} d_i \tag{13}$$

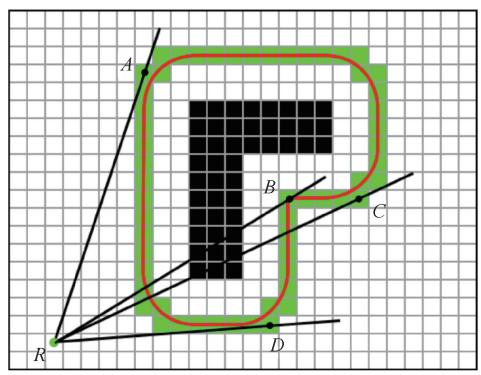


图 3 基于代价势场计算切线及等势曲线长度

Fig. 3 Calculate tangent and equipotential curve length based on cost potential field

式中:$d_i(i=1,2,\cdots,n_e-1)$为栅格点 a_i 到 a_{i+1}的曼哈顿距离,d_{n_e}为栅格点 a_{n_e}到 a_1 的曼哈顿距离。

为计算等势曲线的切线,需要计算等势曲线中栅格点与机器人当前时刻位置 $P_0(x_0,y_0)$间的斜率。设按顺时针方向栅格点 $a_i(i=1,2,\cdots,n_e)$坐标表示为(x_i,y_i),则经过栅格点 a_i的斜率 s_i 为

$$s_i=\frac{y_i-y_0}{x_i-x_0}\qquad x_i-x_0\neq 0 \tag{14}$$

沿顺时针方向依次计算栅格斜率,在顺时针方向上斜率出现极值的栅格点即为等势曲线上的切点,如图 3 中的点 A、B、C、D。对于出现斜率的绝对值无穷大的栅格点,需要进行单独判断,即判断等势曲线中其所在邻域内的栅格点是否出现在过该点与机器人当前位置 R 间连线的同一侧,如果是,则可判断为切点。

为了保证切点的有效性,需要保证在按顺时针方向寻找切点的过程中,相邻切点间的等势曲线长度大于常数 d_m,且相邻的有效切点不能出现相同的斜率,d_m 表示等势曲线上相邻切点间允许的最短等势曲线长度,一般设置为一个较小的数,其值越大则等势曲线上所允许的切点越稀疏。如图 3 中如果切点 B 与切点 C 间的长度小于 d_m,则可以认为切点 B 或切点 C 是无效的。对于求出的任意有效切点,如图 3 中的切点 A,其到点 R 的连线长度表示为切点 A 到机器人当前位置 R 间的欧氏距离。

2.2　候选路径生成

根据得到的障碍物代价势场和切线的计算方法,可以得到不同障碍物的等势曲线及相应的切线。为了更好地描述在避障过程中候选路径的生成方式,如图 4 所示,图中没有画出具体的网格,其中黑色实线为障碍物的边界,包含实时检测到的动态障碍物,灰色曲线为其中一条障碍物代价值为 h_c 的等势曲线。假设机器人当前位于点 R,机器人当前的目标点为点 G,为了得到最初的备选路径,分别计算经过点 R 和点 G 的障碍物等势曲线的切线,如图 4 的红色直线和绿色直线所示,为分别经过点 R 和点 G 的若干个切点。由于图中存在 3 个不连通的障碍物,等势曲线分为独立的 3 段,根据图 4 中切点所在的等势曲线可以将所有的切点分为 3 组,分别为 set1 = {1,6,10,12}、set2 = {2,3,8,11}、set3 = {4,5,7,9},其中点 R 和点 G 自成一个分组。

在实际求解过程中,每个分组内的切点数目是不确定的,假设包含起始点和目标点在内共有 n_t 个节点,每个节点可以和剩下的 n_t-1 个节点产生连接,而每 2 个节点间的连接是无向的,无需计算 2 次,因此可以得到连接的边数为 $n_t(n_t-1)/2$,实际上在同一个分组内的点是不需要进行连接的;所以最终得到的有效连接数会少于 $n_t(n_t-1)/2$。如图 5 所示,图中绘出的点 10 和点 12 的部分有效连接。

对 2 点之间的代价做如下定义:对于同一个分组内的任意 2 点,连接代价为其所在等势曲线上 2 点之间较短的一条曲线长度,即栅格代价较小的一侧。对于不在同一个分组内的两点,如果其间存在直线连接,并且没有穿越等势曲线,即没有穿过障碍物代价值为大于 h_c 的栅格,其连接代价为 2 点之间的欧式距离,如图 5 中点 R 到点 1 之间的连接;如果其间的直线连接穿越了等势曲线,其连接代价为相应的等势曲线部分与 2 点间直线部分的长度之和,如图 5 中点 12 到点 9 之间的连接,其连接代价包含了两部分,一部分为点 12和图中点 p 之间曲线连接中较短的一段等势曲线的长度,另一部分为点 p 到点 9 之间直线连接的长度。

应用切线计算方法和代价定义,可以得到包括机器人当前位置和目标点在内的所有切点之间的有效连接及相应的连接代价,所有的节点及其间的连接代价构成了一个包含起始点和目标点的无向图。为了找到一条可以从起始点到目标点的有效且连接代价最小的候选路径,需要对无向图

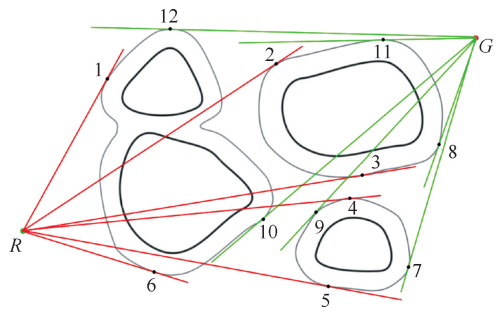


图 4　障碍物等势曲线的切点

Fig. 4　Tangent point of obstacle equipotential curve

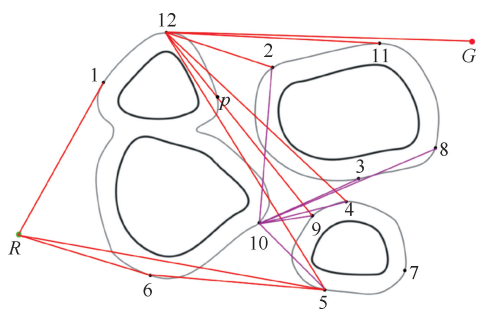


图 5　基于等势曲线切点生成的部分连接代价

Fig. 5　Partial connection cost generated based on equipotential curve tangent points

进行遍历,可以采用深度优先或广度优先算法,或基于最小生成树的求解算法,求解包含起始点和目标点的最小子树,即为代价最小的有效路径。如图 6 所示,图中展示了 3 条连接代价较小的路径。

通过路径生成规则,最终可以得到 1 条无碰撞的有效候选路径,记为 P^*,如图 7 所示,其中红色曲线为在 U 型障碍物中的初始候选路径,R 为机器人的位置,G 为当前目标点。此时得到的路径已经相对较为平滑,但对于机器人来说并不是一条代价最优的路径,需要对路径进行进一步的调整。

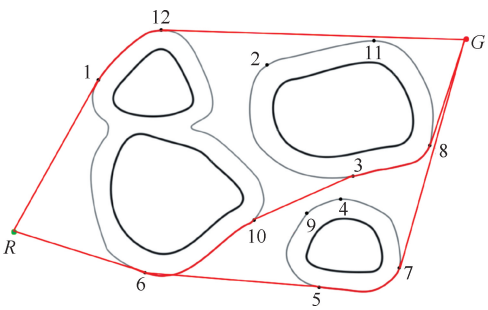


图 6 由最小生成树得到的 3 条代价较小的初始路径

Fig. 6 3 initial paths with lower cost obtained from the minimum spanning tree

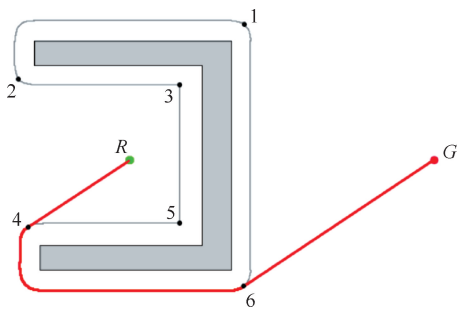


图 7 U 型障碍物中候选路径生成示意图

Fig. 7 Schematic diagram of candidate path generation in U-shaped obstacle

2.3 机器人不同初始姿态下的候选路径

2.2 节所提初始候选路径生成方案没有考虑机器人的初始姿态,默认机器人是可以进行原地转向的差动转向机器人,但是在一些特定的场景下,机器人无法进行原地转向或转向具有较高的成本时,就需要考虑机器人的初始姿态了。设机器人的最小转弯半径为 R_0,机器人的当前姿态如图 8 所示,根据最小转弯半径 R_0 定义 2 个最小转向圆[12],2 个圆在机器人的当前位置 $P_0(e,n,\theta_w)$ 与机器人的航向 θ_w 相切,其中点 P 为 2.1 节得到的候选路径中节点 Q 与初始转向圆的一个切点。由于初始候选路径中 P_0Q 穿过了最小转向圆的内部,存在最小转弯半径约束的机器人无法进行路径跟随,因此在生成候选路径时,取最小转弯半径上的圆弧$\overset{\frown}{P_0P}$及切线 PQ 代替初始候选路径中的线段 P_0Q,最终得到考虑初始最小转弯半径的场景下的初始候选路径,如图 8 中红色路径所示。

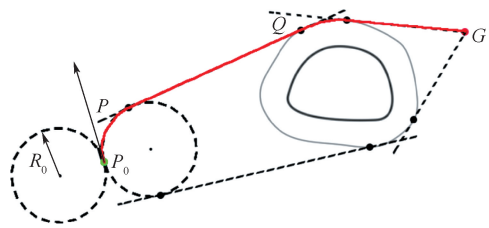


图 8 初试姿态及最小转向半径约束下的初始路径

Fig. 8 Initial path under constraints of initial attitude and the minimum turning radius

2.4 候选路径调整

在栅格地图中,初始候选路径是由紧密相邻的栅格表示的,不利于后续路径的调整,因此,需要对初始候选栅格路径按照一定的规则进行采样,得到一系列离散的点,称之为锚点。机器人的路径可以用一系列的锚点 $p_s,p_1,p_2,\cdots,p_e$ 表示,其中,p_s 表示路径起始点,p_e 表示路径终点。在采样过程中使用曼哈顿距离来表示栅格地图中 2 个栅格之间的距离。在调整过程中,相邻锚点间的距离过小会使锚点所受合力因相互抵消而变小,需要更多的调整轮数,同时也会增大计算量,距离过大则会使锚点过于稀疏,路径的平滑度变差,因此,为保证整条路径中锚点稀疏程度的一致性,设 D_{max} 为路径调整过程中的采样间距,约定在对候选路径进行采样时,满足采样后相邻 2 个锚点之间的距离 D_i 在小于常量 D_{max} 的前提下取最大值。此时采样后的候选路径 P^*,可以由锚点集合 $P^*=\{p_s,p_1,p_2,\cdots,p_e\}$ 表示,如图 9 所示。

下面定义一种路径调整算法,将候选路径 P 调整为近似最短的安全路径。设锚点 $p_s,p_1,p_2,\cdots,p_e$ 所在栅格的障碍物代价值为 $c_s,c_1,c_2,\cdots,c_e$。则候选路径的障碍物代价 C_{obs} 为

$$C_{obs}=\frac{1}{n_a}(c_s+c_1+c_2+\cdots+c_e) \tag{15}$$

式中:c_i 为锚点所在栅格的代价值,c_s 为机器人当前位置的栅格代价值;c_e 为目标点的代价值;n_a 为采样锚点的个数。

设锚点 $p_s,p_1,p_2,\cdots,p_e$ 中相邻锚点间的距离代价值为 $h_1,h_2,\cdots,h_e$。则候选路径的长度代价 H_{obs} 为

$$H_{obs}=h_1+h_2+\cdots+h_e \tag{16}$$

式中:$h_i(i=1,2,3,\cdots,e)$ 为相邻锚点之间的曼哈

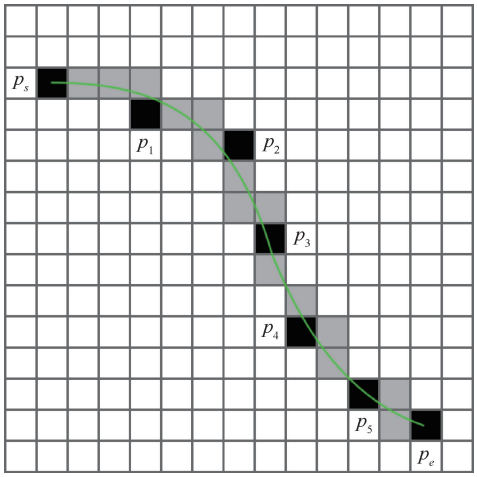


图 9　栅格地图中初始路径采样表示

Fig. 9　Sampling representation of initial path in raster map

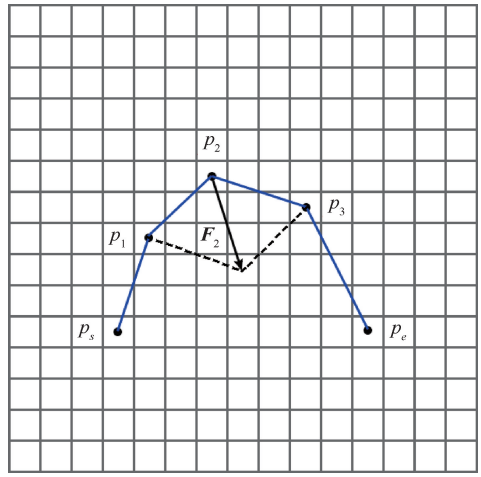


图 10　候选路径中锚点 p_2 所受合力 $\boldsymbol{F}_2$

Fig. 10　Resultant force $\boldsymbol{F}_2$ on anchor point p_2 in candidate path

顿距离。此时机器人由当前位置到目标点加权代价 C_{cost} 为

$$C_{\text{cost}} = \mu C_{\text{obs}} + (1-\mu) H_{\text{obs}} \tag{17}$$

式中:μ 为障碍物代价和距离代价对应的权值系数。

为了进一步对路径进行调整,定义相邻锚点之间存在吸引力。通过锚点之间的引力可以进一步调整候选路径的长度。约定每个锚点只会受到相邻锚点的影响而不会受到其他锚点的影响。设 p_1 的坐标为 (i_1, j_1),p_2 的坐标为 (i_2, j_2),p_3 的坐标为 (i_3, j_3),$\boldsymbol{f}_1$ 表示由点 p_2 到 p_1 的引力向量,$\boldsymbol{f}_2$ 表示由点 p_2 到 p_3 的引力向量,如图 10 所示,可以计算锚点 p_2 受到的合力 $\boldsymbol{F}_2 = \boldsymbol{f}_1 + \boldsymbol{f}_2$。此时锚点 p_2 由于受到合力 $\boldsymbol{F}_2$ 的作用沿 $\boldsymbol{F}_2$ 的方向移动。若合力 $\boldsymbol{F}_2$ 表示为 (F_x, F_y),则锚点 p_2 调整之后的坐标为 $p_2(i_2 + \sigma F_x, j_2 + \sigma F_y)$,其中 σ 为调整增益,满足 $0 < \sigma < 1$。

为保证调整过程中整条路径中锚点稀疏程度的一致性,约定在路径调整过程中相邻 2 个锚点的距离 D_i 满足约束条件:$d_{\min} < D_i < 2d_{\min}$,其中 $d_{\min}$ 为相邻 2 个锚点的最小距离,一般取 $d_{\min} = 0.75 D_{\max}$,并且初始锚点 p_s 和目标点锚点 p_e 为固定点,在调整过程中不会发生变化。当出现 2 个不相邻锚点 p_{i-1} 和 p_{i+1} 之间的距离 $D_i < d_{\min}$ 时,可以通过删除锚点 p_i 来满足约束条件。当2 个相邻锚点 p_{i-1} 和 p_i 之间的距离 $D_i > 2d_{\min}$ 时,可以通过插入锚点 p_{new} 来满足约束条件,p_{new} 的坐标为

$$p_{\text{new}} = \left(\frac{i_{x-1} + i_x}{2}, \frac{j_{x-1} + j_x}{2} \right) \tag{18}$$

本节给出了在满足距离代价最优的条件下候选路径的调整方案,下面考虑在满足障碍物代价最优的条件下候选路径的调整方案。

在建立障碍物的二维栅格地图时,建立了从障碍物边界向外梯度下降的代价势场,用以表示机器人经过此处的障碍物代价。候选路径 P^* 是由障碍物等势曲线得到的,因此可以基于障碍物代价势场对路径 P^* 进行调整,使路径的总代价 C_{cost} 最小。

对于栅格地图中候选路径的每一个锚点,按照当前位置梯度下降的方向进行调整。需要计算每一个锚点处的梯度大小和方向,对于候选路径中的任意锚点 $p_{(i,j)}$,以 p 点为中心的 3×3 栅格进行卷积计算。由梯度的计算式:

$$\boldsymbol{f}_G(x,y) = \Delta \boldsymbol{f}(x,y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right)^{\mathrm{T}} \tag{19}$$

定义 $\boldsymbol{E}(x,y)$ 为代价势场中以锚点 p 为中心的 3×3 矩阵,则锚点 p 在 x 方向上的梯度为

$$\boldsymbol{G}_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \cdot \boldsymbol{E} \tag{20}$$

在 y 方向上的梯度为

$$\boldsymbol{G}_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \cdot \boldsymbol{E} \tag{21}$$

可以得到锚点 p 的梯度大小为

$$G = \sqrt{|\boldsymbol{G}_x|^2 + |\boldsymbol{G}_y|^2} \tag{22}$$

由式(23)计算锚点 p 的梯度方向:

$$\boldsymbol{\Theta} = \arctan\left(\frac{|\boldsymbol{G}_y|}{|\boldsymbol{G}_x|} \right) \tag{23}$$

为了避免候选路径在调整的过程中因梯度方向的改变导致调整后的路径不平滑,定义单位向量 $\boldsymbol{g} = (\boldsymbol{G}_x/G, \boldsymbol{G}_y/G)$ 为梯度的方向向量。锚点 p 在 x 方向上的梯度调整量为

$$g_x = \omega\left(\frac{|\boldsymbol{G}_x|}{G}\right)^3 G = \frac{\omega|\boldsymbol{G}_x|}{1+(|\boldsymbol{G}_y|/|\boldsymbol{G}_x|)^2} \tag{24}$$

在 y 方向上的梯度调整量为

$$g_y = \frac{\omega|\boldsymbol{G}_y|}{1+(|\boldsymbol{G}_x|/|\boldsymbol{G}_y|)^2} \tag{25}$$

式中：ω 为梯度调整量的增益系数。则锚点 p 在梯度调整后的坐标为 $p(i-g_x, j-g_y)$。梯度下降的方向为当前待调整锚点在障碍物代价势场作用下的调整方向。候选路径的总代价 C_{cost} 为

$$\begin{aligned} C_{\text{cost}} &= \mu C_{\text{obs}} + (1-\mu)H_{\text{obs}} = \\ &\mu(c_s + c_1 + c_2 + \cdots + c_e) + \\ &(1-\mu)(h_1 + h_2 + \cdots + h_e) \end{aligned} \tag{26}$$

为了保证路径调整的可靠性，路径在每一步调整的过程中需要满足如下约束条件：

1）若 h_c 为初始候选路径中所选等势曲线对应的障碍物代价值，在路径调整过程中候选路径中的任何一个锚点都不会从当前所在栅格移动到障碍物代价值高于 h_c 的栅格。如果针对当前锚点的调整会使得调整之后的锚点所在栅格的障碍物代价大于 h_c，则跳过对当前锚点的位置调整继续调整下一个锚点。

2）如果考虑机器人的初始姿态及最小转弯半径 R_0，则应当保证调整之后的路径锚点不会出现在最小转向圆内部。如果针对当前锚点的调整会使得调整之后的锚点与最近的转向圆的圆心距离小于 R_0，则跳过对当前锚点的位置调整继续调整下一个锚点。

3）为了保证路径在调整过程中的平滑度，设 $R_{\min}(R_{\min} > R_0)$ 为路径调整过程中的最小曲率半径，对于任意相邻的 3 个锚点，其确定的圆半径 r 不小于最小曲率半径 $R_{\min}$。在路径调整过程中，对锚点沿梯度方向调整和沿相邻锚点合力方向上的调整均会影响该锚点处的曲率圆半径，而在调整过程中不存在 2 种调整方式都使锚点处的曲率圆半径变小的情况，在对锚点沿梯度方向调整过程或沿相邻锚点合力方向上的调整过程中，针对已满足最小转弯半径条件的锚点，如果调整后的结果会破坏最小转弯半径的条件，则跳过对当前锚点的位置调整继续调整下一个锚点。如果初始路径中存在不满足最小转弯半径条件的锚点，针对这样的锚点，如果经过路径长度或势场代价的调整会使调整后的结果对应的曲率圆半径减小，则跳过对当前锚点的相应调整继续调整下一个锚点。因此，通过本文算法可以保证路径的平滑度，使路径满足最小转弯半径的约束条件。路径中锚点的转弯半径按式(27)进行计算。

图 11 为锚点 p_1、p_2、p_3 所确定的曲率圆，p_1 的坐标为 (i_1, j_1)，p_2 的坐标为 (i_2, j_2)，p_3 的坐标为 (i_3, j_3)，圆心坐标 $O(x, y)$ 为 p_1、p_2 与 p_2、p_3 所决定的 2 条中垂线交点决定，即解下面二元一次方程组：

$$\begin{cases} y - \dfrac{j_1 + j_2}{2} = -\dfrac{i_1 - i_2}{j_1 - j_2}\left(x - \dfrac{i_1 + i_2}{2}\right) \\ y - \dfrac{j_2 + j_3}{2} = -\dfrac{i_2 - i_3}{j_2 - j_3}\left(x - \dfrac{i_2 + i_3}{2}\right) \end{cases} \tag{27}$$

则 p_1、p_2、p_3 所确定的曲率圆半径为

$$r = \sqrt{(x-i_1)^2 + (y-j_1)^2} \tag{28}$$

本文提出如下路径规划算法来调整候选路径 P^*，具体过程如下：

步骤 1 以 $D_{\max}$ 为采样间距对初始候选路径进行采样，使得采样后相邻 2 个锚点之间的曼哈顿距离 D_i 在小于常量 $D_{\max}$ 的前提下取最大值，得到候选路径锚点集合 $P = \{p_s, p_1, p_2, \cdots, p_e\}$。在调整过程中任意相邻的 2 个锚点的曼哈顿距离 D_i 满足约束条件 $d_{\min} < D_i < 2d_{\min}$，其中 $d_{\min}$ 为任意 2 个锚点的最小距离。

步骤 2 执行以下循环：

1）计算锚点 $p_k(k = 1, 2, 3, \cdots, e-1)$ 所受到的合力 $\boldsymbol{F}(p_k)$。

2）尝试沿 $\boldsymbol{F}(p_k)$ 方向移动锚点 p_k，判断从起始点到当前锚点的 C_{obs} 是否减小，如果是，则更新当前锚点。

3）判断锚点是否满足距离约束条件，如果 $D_k < d_{\min}$ 则删除锚点，如果 $D_i < 2d_{\min}$ 则插入新锚点。

4）循环执行第 2 步直至完成集合中所有锚点的调整。

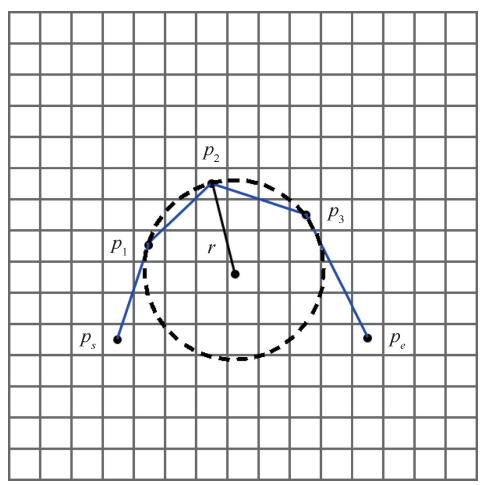


图 11 锚点 p_2 对应的曲率圆半径

Fig. 11 Radius of circle of curvature corresponding to anchor point p_2

步骤 3　执行以下步骤：

1）计算锚点 $p_k(k=1,2,3,\cdots,e-1)$ 处的梯度大小和方向，并沿着负梯度的方向根据梯度的绝对值大小移动锚点 p_k。

2）判断锚点是否满足距离约束条件，如果 $D_k < d_{\min}$，则删除锚点，如果 $D_i < 2d_{\min}$ 则插入新锚点。

3）重复第 1 步和第 2 步，直至完成集合中所有锚点的调整。

4）比较第 3 步调整之后的路径总代价 C_{cost} 是否比调整之前小，如果是则第 3 步调整生效，否则不生效。

步骤 4　重复执行步骤 2 和步骤 3 并迭代更新锚点，直至 C_{cost} 的变化量 $\Delta C_{\text{cost}} < C_{\text{con}}$ 或循环达到一定次数，则退出循环。其中 $C_{\text{con}} > 0$ 为一个给定的阈值。

根据本文提出的局部避障算法，可以得到一条由采样锚点构成，相对较短且安全的路径。机器人在实际路径跟随的过程中可以采用分段 2 次或 3 次样条插值法对调整后包含有限锚点的路径进行跟随。

3　实验与仿真

为了验证本文算法的特性，在 MATLAB 中进行了仿真。设定栅格地图的分辨率为 1 200 × 1 000 m。在本节中，分别验证了本文避障算法在静态障碍物环境中及包含移动障碍物的动态场景下的性能。在场景中，静态障碍物可以是墙、机器和设备。动态障碍物可以是行走的人，移动的设备或其他机器人。如图 12 所示，机器人可以通过静态栅格地图及基于传感器的障碍物信息生成包含障碍物速度信息的代价势场，并基于本文算法实现动态避障到达目的地。本文重点针对动态障碍物场景下单个机器人的避障进行仿真测试实验。

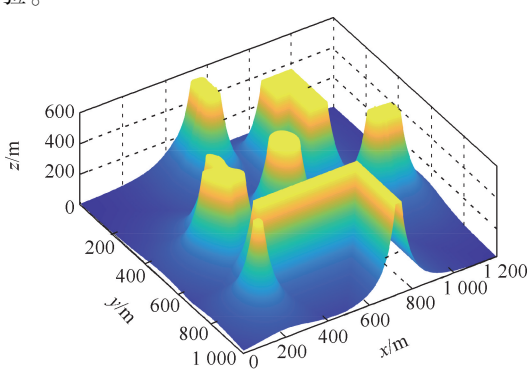


图 12　静态栅格地图的代价势场

Fig. 12　Cost potential field of static grid map

3.1　静态场景下的避障分析

在计算机仿真中，通过计算栅格地图中障碍物栅格的代价，可以得到如图 12 所示的代价势场，高度越高的区域表示机器人由此处经过的代价越大。在更新障碍物代价势场时可以得到代价势场的等势曲线，图 13 为 5 组对应不同障碍物代价值的等势曲线。由图 12 与图 13 可知，等势曲线对应的代价值 h_c 取值越大，则等势曲线越靠近障碍物。因此，为保证移动机器人导航过程的安全性，避免与障碍物发生碰撞，机器人导航过程中障碍物代价值 h_c 的选择应使得等势曲线与障碍物的边界间的最短距离 d_c 略大于 0.5 倍机器人的宽度或机器人外接圆半径，由图 13 可知 h_c 的取值过小会使得路径过于远离障碍物进而无法通过一些狭窄的通道，因此实际应用中应根据机器人的具体外形参数选择合适的 h_c。其中令式(8)中 d_{op} 的值取机器人的外接圆半径，此时计算出的 U_p 的值即为 h_c 对应的障碍物代价临界值。

假设机器人当前位于点(550，900) m，向目标点栅格(950，100) m 行进，如图 14 所示，其中黑色区域代表障碍物。取障碍物代价势场的最大值 $C_{\max}=600$，初始候选路径中所选等势曲线对应的障碍物代价值 $h_c=200$，可以在图 12 所示的代价势场中得到对应代价值 $h_c=200$ 的若干段平滑的等势线。

然后分别寻找经过起始点和目标点的等高线的 n_t 个切点，不同分组内的点，其间的连接代价包含栅格地图中的障碍物代价 C_{obs} 和距离代价 H_{obs}。使用经典的 Prim 算法可以得到包含 n_t-1 条边的最小生成树，最优候选路径即为最小生成树中包含起始点和目标点的最小子树，如图 14 中的蓝色曲线所示，即为最优候选路径由采样之后的锚点组成的路径。此时的路径虽然可以用于机器人的导航，但路径的代价和平滑度都不是最优

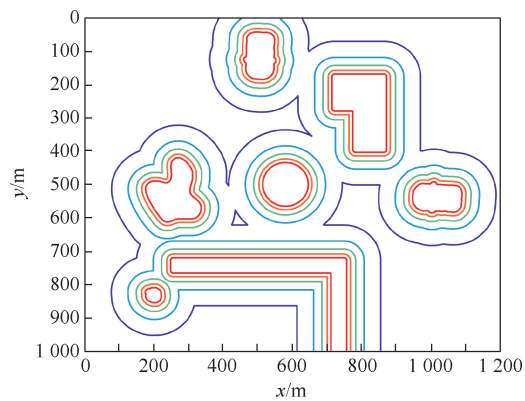


图 13　障碍物代价势场的等势曲线

Fig. 13　Contour plot in cost potential field of obstacles

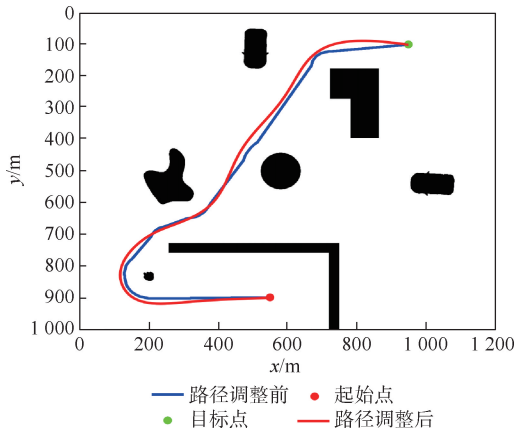


图 14 静态障碍物场景中本文算法避障效果

Fig. 14 Obstacle avoidance effect of the proposed method in static obstacle scene

的,需要进行调整。根据 2.4 节中的路径调整算法,对候选路径进行调整,这里障碍物代价和距离代价对应的权值系数 μ 取 0.65,最终得到图 14 中红色曲线所示的调整后的最优路径。

3.2 动态场景下的避障分析

为了测试本文算法在动态场景下的性能,设定了如图 15 所示的动态障碍物场景。假设机器人从 R 点出发,目标点为 G,其中 A 为移动的行人障碍物,B 和 C 为 2 个移动的小车障碍物,在机器人避障的过程中其位置会发生变化。黑色的区域为静态障碍物。场景中的障碍物小车 B 和 C 及行人障碍物 A 运动轨迹为随机的曲线。在避障过程中动态障碍物的位置和速度会发生变化,因此,在生成动态障碍物的代价势场时需要考虑障碍物的速度大小和方向。按照 2.4 节提到的算法生成动态障碍物的速度代价势场,和静态地图的代价势场进行叠加,得到包含动态障碍物信息的代价势场。图 16 为按照第 2 节提出的路径生成和调整方案在实现动态避障过程中机器人与动态障碍物交互的仿真图。当机器人移动到图 16(a)中位置时,动态障碍物代价势场及实时路径如图 16(b)所示,当机器人移动到图 16(c)中位置时,动态障碍物代价势场及实时路径如图 16(d)所示。通过对比发现,机器人可以很好地根据障碍物的运动状态对当前局部路径进行调整,获得较好的避障效果。在动态障碍物场景下机器人避障路径及与动态障碍物的交互结果如图 17所示,R 为机器人的起始位置,G 为机器人的当前目标点。图 17 展示了机器人及动态障碍物在运动过程中的运动轨迹,其中蓝色的路线为障碍物小车 B 的移动轨迹,黄色的路线为障碍物小车 C 的移动轨迹,橙色路径为行人障碍物 A 的移动轨迹。

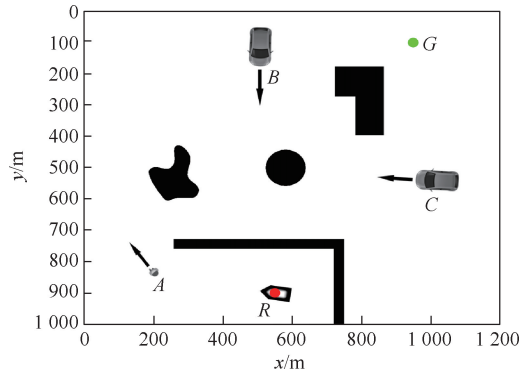


图 15 初始动态场景中障碍物运动状态

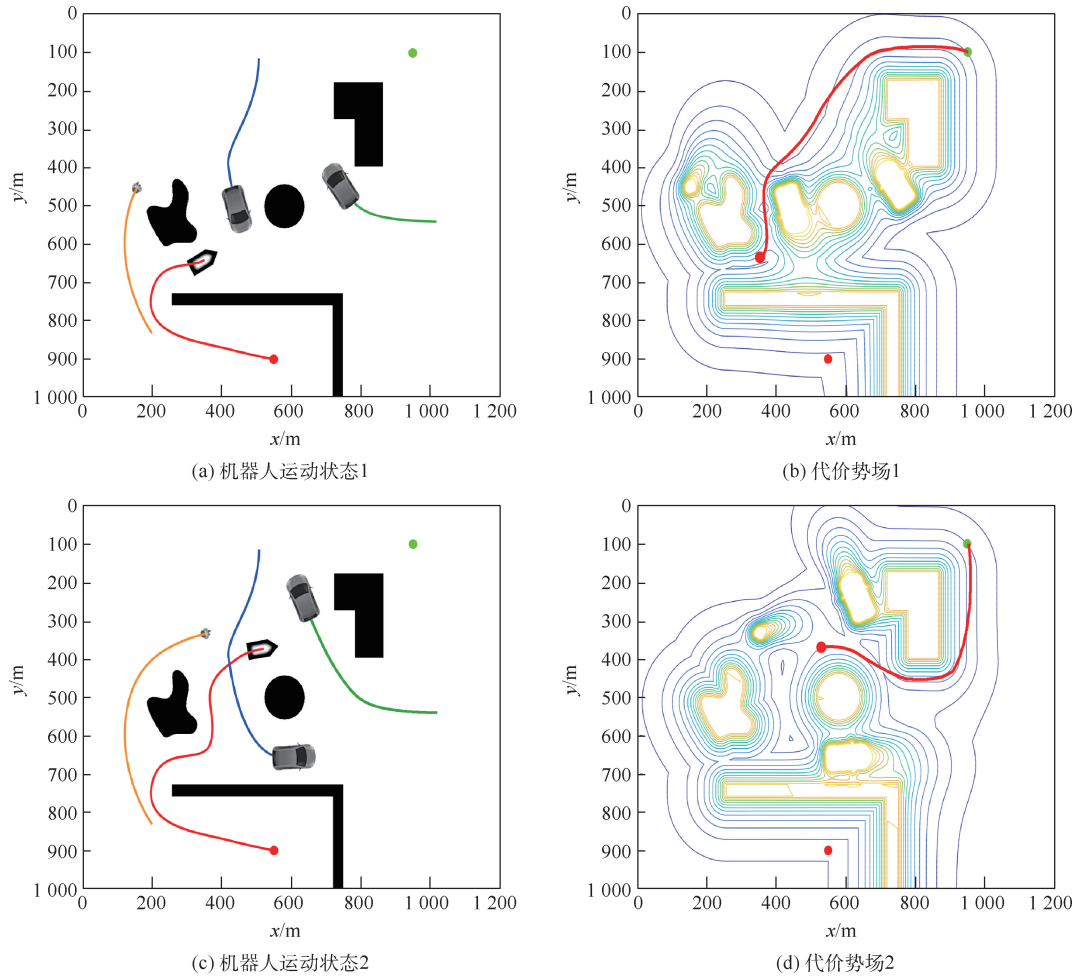
Fig. 15 Obstacle motion state in initial dynamic scene

红色路径为机器人在避障过程中的运动轨迹,可以看出,在动态移动障碍物场景下机器人可以保持较为平滑的运动轨迹。图 18 展示了机器人在移动的过程中距离动态障碍物 A、B、C 及当前目标点 G 的实时距离,其中横坐标是移动机器人从起始点出发运行的时间,纵坐标是和动态障碍物之间的距离。由于在避障过程中引入了障碍物速度对代价势场的影响,在障碍物前进方向上的区域会产生对应的障碍物速度代价。可以看出,由于机器人在运动过程中通过障碍物速度可以考虑动态障碍物未来一段时间的位置对避障过程的影响,在移动障碍物场景下具有较高的灵活性。

3.3 不同算法的结果比较

本文提出了一种基于等势线和最小生成树的路径生成及代价势场下的路径调整算法。本节中,为了比较不同路径规划算法的特性和性能,构建了如图 19 所示的障碍物场景,对比了模拟人工势场(APF)算法、RRT* 算法、A* 算法和本文算法。为了方便不同算法搜索时间及搜索空间的对比,定义了地图遍历率的评价指标,其值定义为在完成 1 次路径规划过程中,栅格地图中被访问过的栅格数量与地图总栅格数量的比值,反映了算法的空间搜索效率。同时基于式(17)的路径评价指标,本文对比了 4 种算法在时间消耗及路径代价上的量化指标,如表 1 和表 2 所示,其中障碍物权值系数 μ 取 0.40。

经典的 APF 算法[13]定义了 2 个主要的势场:引力势场和斥力势场。引力势场会吸引机器人向着目标点移动,障碍物产生的斥力势场会排斥机器人。最终机器人由受到的虚拟引力和虚拟斥力的合力牵引着向目标点移动,但是他也存在着一些缺点,如局部最优值问题和目标不可达问题。如图 19 中的绿色曲线所示,由于存在局部最优



(a) 机器人运动状态1　(b) 代价势场1

(c) 机器人运动状态2　(d) 代价势场2

图 16　动态场景中机器人实时避障路径

Fig. 16　Real-time obstacle avoidance path of robots in dynamic scenarios

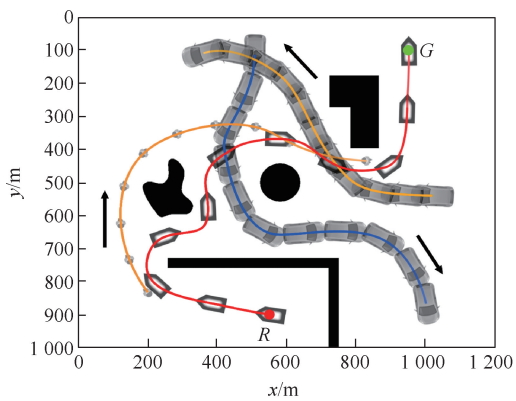


图 17　机器人与动态障碍物交互过程中的运动轨迹

Fig. 17　Movement trajectory during interaction between robot and dynamic obstacles

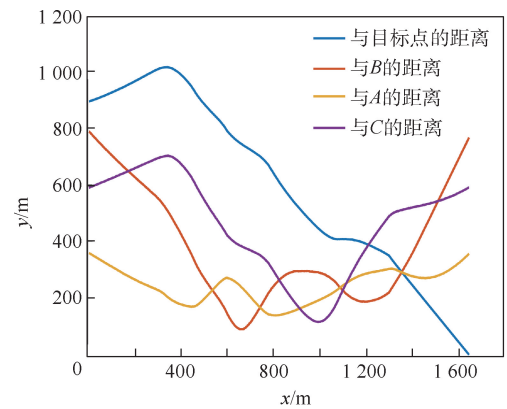


图 18　机器人距离障碍物 A、B、C 及目标点的实时距离

Fig. 18　Real-time distance of robot from obstacles A, B, C and target point

值,APF 算法无法完成路径规划。当不存在局部最优值时,如图 20 所示障碍物场景,APF 算法可以在人工势场中根据梯度下降法或邻域搜索法快速完成路径搜索,具有较高的时间效率及较低的地图遍历率,然而由表 1 和表 2 中的数据可知,由于没有考虑路径和障碍物间的距离及路径平滑度等因素,其路径具有较高的长度代价和加权代价,并不是最优的选择。RRT* 算法[14]是以起始点为根节点,通过随机采样的算法增加叶子节点,生成一棵随机扩展树,当存在叶子节点进入目标区域,就得到了从起始点到目标点的路径。他的优点是实现较为简单,速度快,无论自由度多少,多复杂的约束都能适用。但缺点也非常明显,如图 19 和图 20 中的橙色曲线所示,由于其路径的随机生长

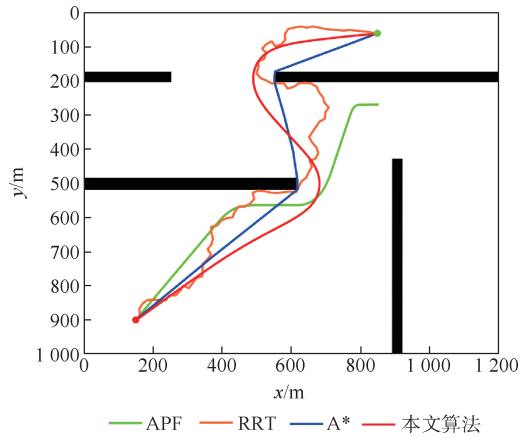


图 19 场景 1 中不同算法的路径规划结果对比

Fig. 19 Comparison of path planning results of different methods in Scenario 1

表 1 场景 1 中不同算法路径代价指标的 10 次平均值对比

Table 1 Ten mean comparison of path cost indexes of different methods in Scenario 1

评价指标	APF	RRT*	A*	本文算法
算法耗时/s	0.114	1.794	4.475	1.645
长度代价	17.98	23.02	16.72	19.24
障碍物代价	14.03	18.66	19.43	11.06
加权代价	15.61	20.41	18.35	14.33
地图遍历率/%	0.18	1.64	9.34	2.64

表 2 场景 2 中不同算法路径代价指标的 10 次平均值对比

Table 2 Ten mean comparison of path cost indexes of different methods in Scenario 2

评价指标	APF	RRT*	A*	本文算法
算法耗时/s	0.110	1.125	1.210	1.636
长度代价	18.63	17.93	14.40	15.73
障碍物代价	15.35	21.61	20.49	13.86
加权代价	16.67	20.14	18.05	14.61
地图遍历率/%	0.21	1.03	2.12	2.53

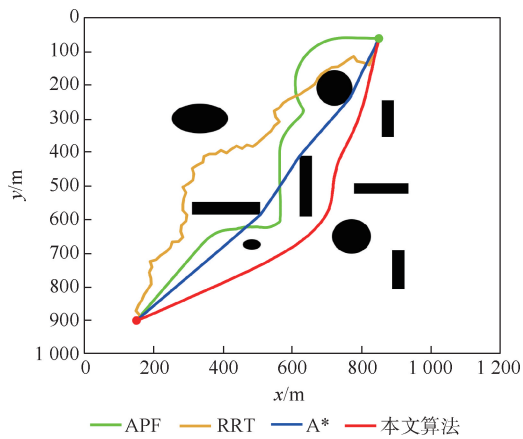


图 20 场景 2 中不同算法的路径规划结果对比

Fig. 20 Comparison of path planning results of different methods in Scenario 1

方式,其生成的路径通常包含棱角,不够光滑,并且靠近障碍物,具有较高的时间代价和障碍物代价。根据 RRT* 算法中扩展树的随机生长方式,算法在执行过程中需要对整个地图空间内进行随机采样,会产生很多无效的栅格采样点,因此其地图遍历率较高。其路径规划参数对比如表 1 和表 2所示,其路径长度代价和障碍物代价都相对较高,并且不是最优的。

A* 算法[15]的显著特点是基于栅格地图完成路径的搜索,由起点开始,每次优先选择距离起点代价和距离终点的预计代价之和最小的节点作为下一个要遍历的节点依次遍历,直到到达终点。A* 算法没有考虑和障碍物的安全距离,并且生成的路径不够平滑,如图 19 和图 20 中的蓝色曲线所示。在实际应用过程中需要进行进一步的优化。在目标点被障碍物遮挡且地图分辨率较高的栅格地图场景中,A* 算法会有大量的无效节点栅格被加入有效点集合,因此搜索过程中花较多的时间查找有效点集合,仿真中耗时高于另外3 种算法,并且具有较高的地图遍历率,同时路径的障碍物代价值也相对较高,在实际过程中很容易与障碍物发生碰撞。当目标点不被障碍物遮挡或存在较少的障碍物遮挡时,如图 20 所示仿真场景,算法的地图遍历率和耗时相比图 19 仿真场景有所降低。APF 的时间复杂度相对较低,并且具有较小的障碍物代价值,但由于其存在局部最优问题,在存在局部最优值的场景中无法完成避障任务。RRT* 算法采用概率随机树的方式生成避障路径,也存在障碍物代价高的问题,并且由于树的生长方式是随机的,在相同的场景下每次的时间消耗和障碍物代价都可能发生变化。由表 3 的数据可知,RRT* 算法在相同场景下进行多次的路径规划仿真时,算法的长度代价方差与障碍物代价方差都比较大,得到的路径具有随机性,这对于移动机器人的动态避障是不利的,相比之下 APF、A* 及本文算法在多次重复实验时差较小,具有较高的稳定性。本文算法同时考虑了路径长度,障碍物代价,路径平滑度等多种因素,通过求解代价势场等势线切点和求解最小生成子树的方式获得

表 3 场景 2 中不同算法路径代价方差对比

Table 3 Variance comparison of path costs of different methods in Scenario 2

评价指标	APF	RRT*	A*	本文算法
耗时方差	2.16×10^{-5}	5.28×10^{-3}	4.47×10^{-5}	2.246×10^{-5}
长度方差	0	1.01	0	0
代价方差	0	3.66	0	0

初始路径,假设锚点间存在引力,并且会受到代价势场的影响,通过调整路径锚点获得有效路径,由表 1 和表 2 的数据可知,本文算法在时间复杂度和障碍物代价值上都具有较好的指标,同时也具有良好的地图遍历率指标。

3.4　复杂场景下的自主避障

本节测试了在相对较为复杂的场景下本文算法的避障效果,并对比了 APF 算法、A* 算法和本文算法的不同表现,栅格地图的大小同样是 1 200 × 1 000 m,如图 21 所示,黑色栅格区域为障碍物,红色点(300,900)为起始点,绿色点(900,60)为目标点。选择高度为 $h_c = 120$ 的等势线可以在代价势场中得到图 22 红色曲线所示的初始候选路径。为了方便比较,地图中没有设置局部最优值,保证了 APF 算法也能够完成避障。

复杂场景下的避障结果如图 21 所示。3 种算法都能成功避障,可以看出,相比本文算法,APF 算法的轨迹不够平滑,并且选择了一条较远的路径,在路径长度上没做到最优。A* 算法的轨迹相对较为平滑,但是没有考虑与障碍物间的距离,显然不利于实现安全避障。不同路径的量化指标如表 4 所示,根据 Yao 等[16]提出的指标量化方法,全局平滑指标 $GS = \frac{1}{T-2}\sum_{t=2}^{t-1}\langle p_{t-1}p_t, p_tp_{t+1}\rangle$,局部平滑指标 $LS = \max_{t=2,\cdots,t-1}\langle p_{t-1}p_t, p_tp_{t+1}\rangle$,其中 p_t 满足 $\forall t \in \{1,2,\cdots,T\}$ 表示算法生成路径上离散的栅格坐标。与障碍物的最小距离 d_{obs}^{min} 反映了路径的安全性,越大表示得到的路径越安全。全局平滑指标 GS 反映了避障过程中机器人航向的平均变化量,越小表示路径越平滑。LS 反映了轨迹的局部平滑度,越小表示越容易实现路径的跟踪。

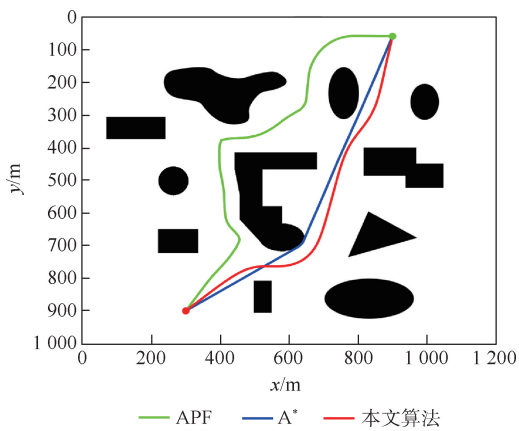


图 21　复杂场景下的避障结果对比

Fig. 21　Comparison of obstacle avoidance results in complex scenarios

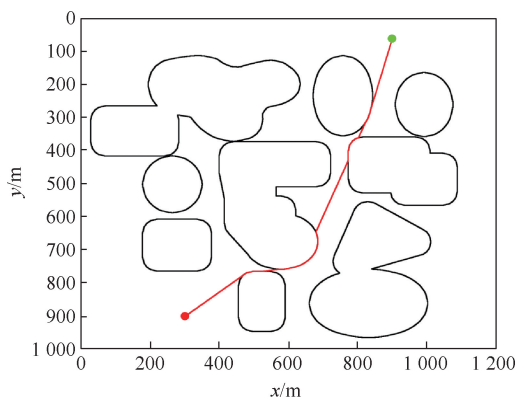


图 22　基于等势曲线生成的初始候选路径

Fig. 22　Initial candidate path generated based on equipotential curve

表 4　不同算法评价指标的 10 次平均值对比

Table 4　Ten average comparison of cost indexes of different methods

评价指标	APF	A*	本文算法
加权路径代价	15.71	17.36	12.64
全局平滑指标	2.13	1.26	1.72
局部平滑指标	8.48	1.47	1.51
障碍物最小距离	36.22	1.41	31.51

实验结果表明,本文算法通过对候选路径在路径长度和障碍物代价上的调整,得到的路径具有较好的平滑性和安全性,同时加权路径代价也相对较小。

4　结　论

针对动态障碍物场景下的移动机器人安全避障问题,本文提出了基于障碍物代价势场的避障路径生成及调整算法,并进行了仿真验证。

1)本文所提算法可以得到较为平滑的路径,能够在和障碍物保持安全距离的前提下做到路径最短。在静态场景下,如果障碍物本身的位置不会发生变化,通过所提算法动态的生成可行路径,并在机器人不断运行的过程中对路径进行调整,可以获得可靠的导航和避障结果。

2)对于移动的障碍物,所提算法加入了其移动速度对障碍物代价势场的影响,在障碍物前进方向的区域范围内产生较强的代价势场,为机器人规避移动中的障碍物预留了充分的裕量,使得机器人可以对障碍物未来的运动态势进行一定程度的预测,对于移动中的障碍物也能够达到较好的响应效果。

3)所提算法综合考虑了路径的长度和障碍物代价,相比其他算法,在路径平滑指标、长度代价等单一指标上不是最优的,但是通过所提算法

对路径的调整使路径具有较好的安全性及较小的加权代价,有利于实际应用中实现良好的路径跟踪和避障。

参考文献(References)

[1] PANDEY A,PANDEY S,PARHI D R. Mobile robot navigation and obstacle avoidance techniques: A review[J]. International Robotics & Automation Journal,2017,2(3):96-105.

[2] RUBLEE E,RABAUD V,KONOLIGE K,et al. ORB:An efficient alternative to SIFT or SURF[C]// 2011 International Conference on Computer Vision. Piscataway:IEEE Press,2011:6-13.

[3] HORNUNG A,WURM K M,BENNEWITZ M,et al. OctoMap:An efficient probabilistic 3D mapping framework based on octrees[J]. Autonomous Robots,2013,34(3):189-206.

[4] YANG L,QI J,SONG D,et al. Survey of robot 3D path planning algorithms[J]. Journal of Control Science and Engineering,2016,2016:1-22.

[5] 王洪斌,尹鹏衡,郑维,等. 基于改进的A*算法与动态窗口法的移动机器人路径规划[J]. 机器人,2020,42(3):346-353.
WANG H B,YIN P H,ZHENG W,et al. Mobile robot path planning based on improved A* algorithm and dynamic window method[J]. Robot,2020,42(3):346-353(in Chinese).

[6] ZAFAR M N,MOHANTA J C. Methodology for path planning and optimization of mobile robots:A review[J]. Procedia Computer Science,2018,133:141-152.

[7] 赵晓,王铮,黄程侃,等. 基于改进A*算法的移动机器人路径规划[J]. 机器人,2018,40(6):903-910.
ZHAO X,WANG Z,HUANG C K,et al. Mobile robot path planning based on an improved A* algorithm[J]. Robot,2018,40(6):903-910(in Chinese).

[8] LIU J,YANG J,LIU H,et al. An improved ant colony algorithm for robot path planning[J]. Soft Computing,2017,21(19):5829-5839.

[9] 林依凡,陈彦杰,何炳蔚,等. 无碰撞检测RRT*的移动机器人运动规划方法[J]. 仪器仪表学报,2020,41(10):257-267.
LIN Y F,CHEN Y J,HE B W,et al. Non-collision checking RRT* algorithm for mobile robot motion planning[J]. Chinese Journal of Scientific Instrument,2020,41(10):257-267(in Chinese).

[10] 薛学华,冯辉,徐海祥,等. 基于改进人工势场法的USV全局路径规划[J]. 武汉理工大学学报(交通科学与工程版),2020,44(6):978-983.
XUE X H,FENG H,XU H X,et al. Global path planning of USV base on improved artifical potential field method[J]. Journal of Wuhan University of Technology (Transportation Science & Engineering),2020,44(6):978-983(in Chinese).

[11] THRUN S. Learning occupancy grid maps with forward sensor models[J]. Autonomous Robots,2003,15(2):111-127.

[12] SAVKIN A V,LI H. A safe area search and map building algorithm for a wheeled mobile robot in complex unknown cluttered environments[J]. Robotica,2018,36(1):96.

[13] KOREN Y,BORENSTEIN J. Potential field methods and their inherent limitations for mobile robot navigation[C]// IEEE International Conference on Robotics and Automation. Piscataway:IEEE Press,1991:1398-1404.

[14] KARAMAN S,FRAZZOLI E. Sampling-based algorithms for optimal motion planning[J]. The International Journal of Robotics Research,2011,30(7):846-894.

[15] HART P E,NILSSON N J,RAPHAEL B. A formal basis for the heuristic determination of minimum cost paths[J]. IEEE Transactions on Systems Science and Cybernetics,1968,4(2):100-107.

[16] YAO P,WANG H,SU Z. Real-time path planning of unmanned aerial vehicle for target tracking and obstacle avoidance in complex dynamic environment[J]. Aerospace Science and Technology,2015,47:269-279.

Obstacle avoidance method of mobile robot based on obstacle cost potential field

CHI Shengkai, XIE Yongfang, CHEN Xiaofang*, PENG Fan

(School of Automation, Central South University, Changsha 410083, China)

Abstract: A mobile robot operates in a dynamic environment, so it must react quickly, maintain a clear path, and keep a safe distance from obstacles. To solve this problem, this paper proposes a dynamic obstacle avoidance method for mobile robots based on the obstacle cost potential field. By establishing a static grid map and the obstacle cost potential field, the equipotential lines in the dynamic scene and the tangents passing through the start to end points are obtained. Then the initial candidate path is obtained by the minimum spanning tree. The candidate path anchor points for the length of the path, the distance from the obstacle and the smoothness are optimized. He candidate path anchor points are then optimized for the path's length, distance from the obstruction, and smoothness. By introducing the influence of obstacle speed on the cost potential field, the robot can respond to the moving obstacle in time. In order to verify the effectiveness of the algorithm, the static scene and the dynamic scene are simulated separately in a grid scene with a resolution of 1 200 × 1 000 m. The results show that the algorithm in this paper can ensure that the path has a high degree of smoothness and maintain safety between obstacles. Moreover, it makes the path as short as possible under the condition of distance. At the same time, it can still maintain the smoothness of the path and the safety of obstacle avoidance in the dynamic obstacle scene, which can meet the requirements of mobile robot path planning in the dynamic scene.

Keywords: mobile robot; path planning; cost potential field; dynamic obstacle; obstacle avoidance

Received: 2021-03-01; **Accepted**: 2021-05-14; **Published online**: 2021-05-26 10:21

URL: kns.cnki.net/kcms/detail/11.2625.V.20210525.1309.003.html

Foundation items: Key-Area Research and Development Program of Guangdong (2021B0101200005); National Natural Science Foundation of China (62133016); National Science Fund for Distinguished Young Scholars (61725306)

* **Corresponding author**. E-mail: xiaofangchen@csu.edu.cn

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0096

数据立方体关联挖掘在红外抗干扰中的应用

吴鑫[1], 伍友利[1,*], 牛得清[1], 蔡宇轩[1], 徐洋[2], 陈鞭[1]

(1. 空军工程大学 航空工程学院,西安 710038; 2. 中国空气动力研究与发展中心, 绵阳 621000)

摘 要: 为了分析空空导弹在红外抗干扰过程中,各干扰量与脱靶量之间的耦合关系,利用红外抗干扰试验仿真平台,通过设置干扰量获得海量数据,构建数据立方体对海量数据进行数据预处理,同时统计出数据集中各事务的支持度计数,将处理后的数据利用 FP-Growth 关联规则算法挖掘干扰量与脱靶量之间的关联规则。通过对关联规则的分析得到:导弹进入角与弹目距离是影响导弹脱靶量大小的主要因素,各个干扰量之间不能随意组合,应该根据导弹来袭方向、目标机动方式选择诱饵投掷策略,否则干扰效果会大大下降。同时设置一组对比试验,验证基于数据立方体的 FP-Growth 算法处理高维数的脱靶量数据,拥有较高的挖掘效率。

关 键 词: 数据立方体; 数据处理; FP-Growth 算法; 关联规则; 红外抗干扰
中图分类号: TN216
文献标志码: A **文章编号**: 1001-5965(2022)11-2304-10

在现代空战中,随着战场环境复杂化、对战目标智能化、作战空间多元化的特点,对红外空空导弹性能提出了更高、更严的要求。其中包括高隐蔽性、高效费比、灵活机动,尤其是需要拥有很强的抗干扰能力[1]。自 20 世纪 90 年代红外诱饵技术的广泛应用,红外导弹的抗干扰能力开始备受各个军事大国关注,红外导弹抗干扰能力设计和评估一直是领域内公认的一个难题[2]。

经过 20 几年的发展,红外抗干扰能力设计及评估方法层出不穷。胡朝晖和闫杰[3]为解决红外空空导弹抗干扰性能的定量评估,提出一种贝叶斯估计算法,但该算法依赖于先验概率,且假设的属性之间相互独立,而抗干扰性能各属性之间相关性较大,因此该算法的局限性较大;王涛和王祥[4]从作战的空、时、频、能量、信息 5 个维度提出一套红外抗干扰评估方法;李慎波等[5]则从导弹作战参数出发研究对红外诱饵干扰效能的影响。随着红外抗干扰能力设计的深入,红外对抗环境的量化建模也成为了研究热点,通过对场景构成和红外物理特性构建红外抗干扰要素模型,对对抗中的环境复杂度进行量化[6]。同时各种机器学习算法如随机森林、支持向量机、贝叶斯分类器等[7-9]与红外抗干扰评估的融合,产生各个算法下的评估方法。但上述方法中均未提到抗干扰指标体系中各指标与红外导弹脱靶量之间的关系,各干扰因素与脱靶量之间的强、弱耦合关系为抗干扰能力的设计及评估提供了着重点。

本文改变诱饵弹投掷参数、弹目相对态势及目标机动方式,这些参量在本文中统称为干扰量,再利用红外抗干扰仿真平台获取海量数据,建立数据立方体将数据进行预处理,最后通过数据挖掘算法得出干扰量与脱靶量之间的关联规则,根据关联规则分析出各干扰量与脱靶量之间的耦合强弱;同时为飞行员在相应态势下如何投掷诱饵,

收稿日期: 2021-03-01; **录用日期**: 2021-06-11; **网络出版时间**: 2021-07-06 15:40
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210706.1304.002.html
基金项目: 国家自然科学基金 (61973253)
* **通信作者**. E-mail: 18359943663@163.com

引用格式: *吴鑫, 伍友利, 牛得清, 等. 数据立方体关联挖掘在红外抗干扰中的应用[J]. 北京航空航天大学学报, 2022, 48(11): 2304-2313. WU X, WU Y L, NIU D Q, et al. Application of data cube association mining in infrared anti-jamming[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2304-2313(in Chinese).*

如何进行战术机动规避导弹提供辅助决策。

1　红外抗干扰数据获取

1.1　红外对抗模型构建

由于本文试验需基于大量的样本试验数据,为此需要构建红外对抗模型,其仿真流程如图 1 所示。从整个仿真流程来看,红外对抗模型主要包括诱饵弹、目标、导弹 3 个部分,其中除了解算诱饵弹、目标、导弹的运动模型外,还需要解算红外辐射模型。图中的初始参数包括:①导弹-目标初始态势信息;②诱饵投掷器信息;③目标机动信息;④各对抗实体自身属性信息,如质量、时间常数等。

导弹、目标、诱饵弹的运动模型依据各自的质心运动方程与质心动力学方程建立[10],对于红外辐射模型的构建,本文采用单线组谱带模型[11]及计算流体力学的方法分别离线计算目标与诱饵的红外辐射强度。

1.2　干扰量设定

本文试验中干扰量包括诱饵投掷参数(parameters of infrared bait, PIB)、目标机动方式(target maneuvering mode, TMM)及弹目相对态势(missile and target relative position, MTRP)[12-13]。其中 PIB 包括投掷总数、每次投掷个数、每组投掷个数、投掷时机、投掷组内间隔、投掷组间间隔、投掷速度、水平投掷角度、垂直投掷角度;TMM 包括无机动、左转弯、右转弯、跃升;MTRP 包括弹目距离和进入角。

根据实际作战态势及红外对抗模型边界条件的约束,本文制定了如表 1 所示的干扰量参数设定范围。结合 4 种目标机动方式,采用控制变量的方法,将所有的干扰量组合所形成的干扰态势遍历一遍,作为输入量导入红外对抗模型中,最终得出 50 000 条不同态势下的脱靶量数据作为后续试验的样本集。

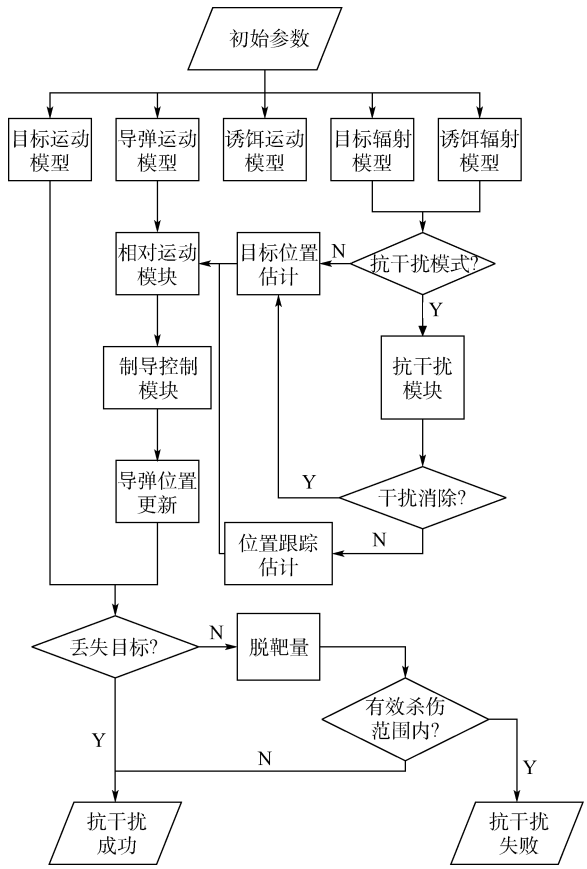


图 1　红外对抗仿真流程

Fig. 1　Simulation process of infrared countermeasure model

表 1　干扰量参数设定范围

Table 1　Range of disturbance variable parameters

干扰量	参数	取值设定
诱饵投掷参数	投掷总数 N_s/个	12,24,36,48
	每次投掷个数 N_e/个	1,2
	每组投掷个数 N_f/个	4,6,12
	投掷时机 T_l/s	0.5,1,1.5,2,2.5
	投掷组内间隔 T_i/s	0.1,0.2,0.3,0.4,0.5
	投掷组间间隔 T_e/s	0.2,0.4,0.6,0.8,1.0
	投掷速度 V_f/(m·s^{-1})	10,20,30,40
	水平投掷角度 θ_l/(°)	0,30,60,90,120,150
	垂直投掷角度 θ_v/(°)	-90, -60,0,30,60,90
弹目相对态势	进入角 θ_m/(°)	10,20,30,70,80,90,130,140,150
	弹目距离 D_r/km	(0,3],(3,4],(4,5],(5,6],(6,7],(7,8],(8,9],(9,10]

2　基于数据立方体的数据预处理

2.1　构建数据立方体

定义 1　多维数据模型(multi-attribute data model, MDM)。将不同对抗态势下多维数据进行记录,可将其表达为 MDM = (F_0;F_1;…;F_i;M),其中 F_i 为数据的维度,M 为维度度量。以红外导弹抗干扰过程为例,红外对抗的多维数据模型可

以表示为MDM=(诱饵投掷参数;弹目相对态势;目标机动方式;脱靶量;4)。其中,PIB包含9个参量,可表示为:PIB=(N_s,N_e,N_f,T_l,T_i,T_e,V_f,θ_l,θ_v);MTRP包含2个参量,可表示为:MTRP=(θ_m,D_r);TMM为1-无机动、2-左转弯、3-右转弯、4-跃升;将脱靶量(miss distance,MD)分为3个等级:MD1<12 m为摧毁、严重毁伤(目标解体或丧失飞行能力);12 m≤MD2<16 m为中度毁伤(目标失去作战能力但还能飞行);MD3≥16 m为轻微毁伤、逃逸成功。

例如,MDM=(24,2,6,1,0.2,0.8,20,30,60;70,3.5;4;MD1)可解析为:目标携带红外诱饵弹24枚,投掷分为3组每组6枚,每次投掷2枚,当仿真经过1 s后开始投掷,1个大组中组内投掷间隔为0.2 s,组与组之间的投掷间隔为0.8 s,诱饵的投掷速度为20 m/s,投掷方向与机体横轴(x轴)夹角为30°、机体纵轴(y轴)夹角为60°;弹目相对态势的初始距离为3.5 km,导弹进入角为70°,此时目标采取跃升机动对导弹进行规避,最后脱靶量为MD1级,说明在此红外对抗仿真作战态势下,目标逃逸失败。

定义2 数据立方体(data cube,DC)。数据立方体是由数据的维和事实定义的,允许对多维数据进行建模和观察。对于不同红外对抗态势下的仿真结果,按照不同维度中参量的组合构建一个数据集合。按照多维数据模型的集合,所构成的数据立方体有4个维度:PIB、MTRP、TMM、MD。

通过4个维度构成的分组总数,可能的分组是{(PIB,MTRP,TMM,MD),(PIB,MTRP,TMM),(MTRP,TMM,MD),(PIB,MTRP,MD),(PIB,TMM,MD),(PIB,MTRP),(MTRP,TMM),(TMM,MD),(PIB,TMM),(PIB,MD),(MTRP,MD),(PIB),(MTRP),(TMM),(MD),(all)},其中(all)指不对任何数据分组,即不进行分维的原始数据。这些分组形成了数据立方体的方体格,如图2所示。

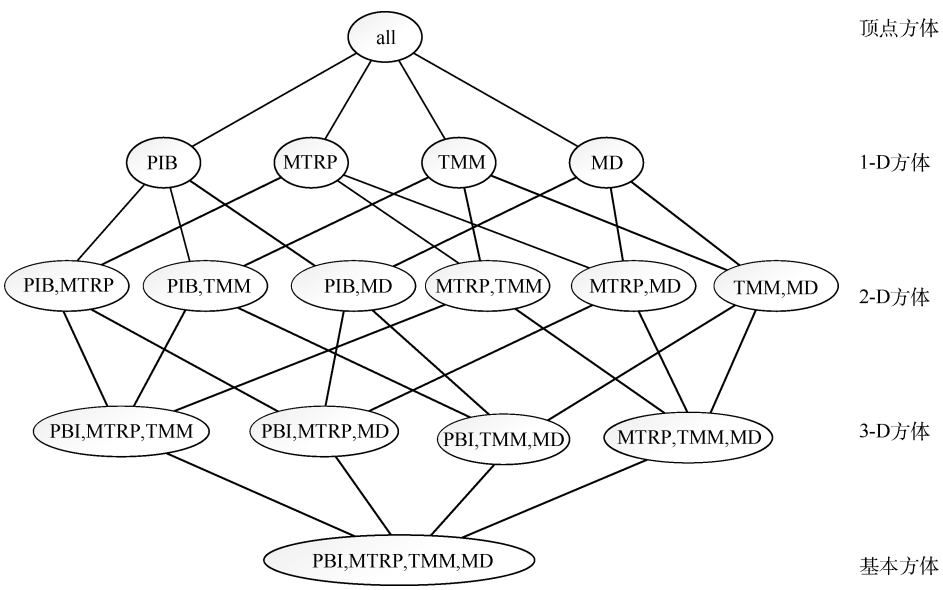


图2 四维数据立方体

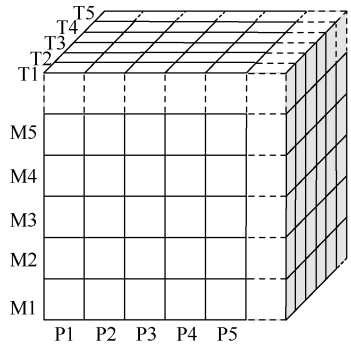
Fig. 2 Four-dimensional data cube

其中基本方体包含PIB、MTRP、TMM、MD这4个维,是低泛化(最特殊化)的方体,该方体中每一条由此四维组成的数据就是红外对抗模型输入量与输出量。顶点方体是最高泛化(最不特殊化)的方体,包含红外对抗模型运行后的所有数据。如果从顶点开始,沿方体的格向下探查,这等价于在数据立方体中下钻;如果从基本方体向上探查,则类似于上卷。通过对数据立方体下钻、上卷等方式,对方体中的数据类型及数据模式进行预处理。

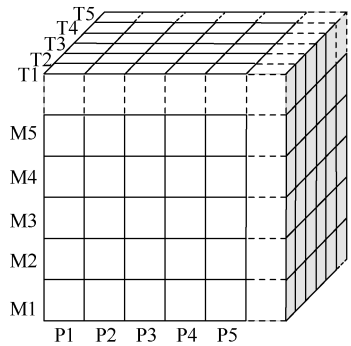
通过图2观察四维数据立方体的数据变得有点麻烦,因此通过降维的手段,把基本方体转换成3-D方体序列。

2.2 数据立方体的降维

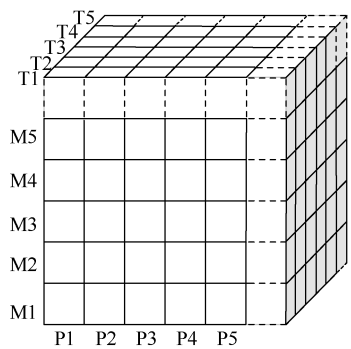
由于四维的立方体数据不好进行观察和处理,本文选用MD作为固定维,将MD按3个等级(MD1、MD2、MD3)把数据分为三大类,这样就可将基本方体转化为3个3-D方体,3-D方体的3个维度:诱饵投掷参量(P)、弹目相对态势(M)、目标机动方式(T)。降维后数据立方体示意图如图3所示。



(a) MD1<12 m



(b) 12 m≤MD2<16 m



(c) MD3≥16 m

图 3　按 MD 降维后数据立方体示意图

Fig. 3　Diagram of data cube after dimension reduction by MD

2.3　数据立方体的查询操作

在 OLAP 软件中关于对数据立方体的操作包括 Cube、Roll-up 及 Drill-down 等操作符号[14-15],通过 SqlServer 的操作步骤,得出降维后数据立方体各参量之间的映射关系如表 2 ~ 表 4 所示。以 MD1 立方体为例,在 SqlServer 中关于数据立方体的具体查询步骤为

```
select PIB,MTRP,TMM,SUM
from MD1
group by PIB by PIB
group by MTRP by MTRP
group by TMM by TMM
```

该步骤可以解释为从 MD1 数据立方体中分别以 PIB、MTRP、TMM 进行分组,并且统计每一个干扰量及其组合的数目。

表 2　MD1 立方体的 1-D 关系

Table 2　1-D relation table of MD1 cube

PIB	MTRP	TMM	数目
N_{s1}	*	*	10 656
⋮	⋮	⋮	⋮
*	θ_{m1}	*	5 556
⋮	⋮	⋮	⋮
*	*	1	9 878

注:* 代表可取任意参数。

表 3　MD1 立方体的 2-D 关系

Table 3　2-D relation table of MD1 cube

PIB	MTRP	TMM	数目
N_{s1}	θ_{m1}	*	3 078
⋮	⋮	⋮	⋮
*	θ_{m1}	1	2 043
⋮	⋮	⋮	⋮
N_{s1}	*	1	4 028
⋮	⋮	⋮	⋮

注:* 代表可取任意参数。

表 4　MD1 立方体的 3-D 关系

Table 4　3-D relation table of MD1 cube

PIB	MTRP	TMM	数目
N_{s1}	θ_{m1}	1	1 059
⋮	⋮	⋮	⋮

注:* 代表可取任意参数。

3　关联规则生成

3.1　关联规则算法相关概念

Apriori 算法与 FP-Growth 算法是关联规则算法中的经典算法。由于 Apriori 算法在产生完整的频繁模式集之前需要对数据库进行多次扫描,同时产生大量的候选频繁集,这就使 Apriori 算法在时间和空间复杂度较大,但 FP-Growth 算法只需对数据库进行 2 次扫描,使用事物数据集的压缩来有效地产生频繁项集,因此 FP-Growth 算法比标准的 Apriori 算法要快几个数量级[16]。故选择 FP-Growth 算法作为本试验关联规则生成算法。

定义 3　关联规则(Rules)。关联规则就是形如 $X \rightarrow Y$ 的表达形式,其中 $X, Y \subseteq I$ 且 $X \cap Y = \phi$,I 是所有项的集合。关联规则就是要找出支持度、置信度都大于等于支持度阈值与置信度阈值的所有规则。

定义 4　支持度(Sup)和置信度(Conf)。假设 X 和 Y 都是数据集中的事务,支持度就是确定规则可以用于给定数据集的频繁程度,而置信度确定 Y 在包含 X 的事务中出现的频繁程度。这 2 种度量的表示形式如下:

$$\mathrm{Sup}(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \tag{1}$$

$$\mathrm{Conf}(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \tag{2}$$

式中:$\sigma(\cdot)$表示项集的支持度计数;N 为数据集事务总数。

定义 5 Kulczynski 度量指标(Kulc)。仅用支持度-置信度框架来衡量关联规则的好坏存在许多局限性,支持度的缺陷在于许多潜在意义的模式由于包含支持度小的项而被删去;置信度的缺陷在于该度量忽略了规则后件中项集的支持度。为了解决这一问题,引入 Kulc 度量指标来进行模式评估。

$$\mathrm{Kulc}(X,Y) = \frac{1}{2}(P(X \mid Y) + P(Y \mid X)) = \frac{1}{2}\left(\frac{\sigma(X \cup Y)}{\sigma(X)} + \frac{\sigma(X \cup Y)}{\sigma(Y)}\right) \tag{3}$$

由式(3)可以看出,Kulc 度量值仅受 X、Y 和 $X \cup Y$ 的支持度影响,而不受事务总个数的影响,当 $\mathrm{Kulc}(X,Y) > 0.5$,则事务 X 与 Y 之间呈正相关;当 $\mathrm{Kulc}(X,Y) < 0.5$,则事务 X 与 Y 之间呈负相关;当 $\mathrm{Kulc}(X,Y) = 0.5$,则事务 X 与 Y 之间无明显关联,且 $\mathrm{Kulc}(X,Y)$ 的值越靠近 0.5,则 X 与 Y 之间的关联程度越弱。

定义 6 不平衡比度量指标(IR)。不平衡比度量指标是评估规则蕴含式中项集 X 与 Y 之间的不平衡程度,是对 Kulc 度量指标的补充。虽然当 $\mathrm{Kulc}(X,Y)$ 的值接近 0.5 时,事务 X 与 Y 之间无明显关联,但是出现这样的关联规则有可能是因为事务中出现了 0 事务,即不包含任何考察项集的事务,所以得对 $\mathrm{Kulc}(X,Y)$ 接近 0.5 的值用 IR 指标进行二次考察,若 $\mathrm{IR}(X,Y)$ 值接近于 0,那么事务 X 与 Y 之间的确毫无关联;若 $\mathrm{IR}(X,Y)$ 接近于 1,则事务 X 与 Y 之间关联较强,即分析出的关联规则可用。不平衡比度量指标定义为

$$\mathrm{IR}(X,Y) = \frac{|\mathrm{Sup}(X) - \mathrm{Sup}(Y)|}{\mathrm{Sup}(X) + \mathrm{Sup}(Y) - \mathrm{Sup}(X \cup Y)} \tag{4}$$

定义 7 FP 树。将数据集中的各个事务数据项按照支持度排序后,把每个事务中的数据项按降序依次插入到一棵以 NULL 为根节点的树中,同时在每个节点处记录该节点出现的支持度,最终构造出 FP 树。FP-Growth 算法采用分治策略将一个问题分解为较小的子问题,每一次递归,通过更新前缀路径中的支持度计数和删除非频繁项来构建条件 FP 树。举一具体实例展示 FP 树及条件 FP 树的生成过程。

算法第 1 次扫描数据集,计算出各参数频数,并设定最小支持度,将频数小于最小支持度的参数项进行删除,依据频数将每条数据的参数从大到小重新进行排列,过程如图 4 所示。

紧接着算法第 2 次扫描数据库,从根节点“NULL”出发,依次将每条数据读入,出现相同的节点进行累加,如所给例子,先读入 TID1 生成第 1 条分支;之后读入第 2 条数据形成第 2 条分支;再读入第 3 条数据,由于与第 2 条数据有共同的参量 I2,所以在 I2 节点上计数加 1,以此类推将数据集遍历一遍生成 FP 树,过程如图 5所示。

依据生成的 FP 树依次考察末节点,删除末节点所包含的前缀路径,该路径剩余节点构成的树即为该考察节点的条件 FP 树。各末节点对应的条件 FP 树如图 6 所示。

事务	项集
1	I1, I3, I5
2	I1, I2
3	I2, I5
4	I2, I4
5	I1, I3
6	I1, I2, I3
7	I2, I4, I5
8	I1, I2, I5

项集	支持度计数
I1	5
I2	6
I3	3
I4	2
I5	4

事务	项集
1	I1, I5, I3
2	I2, I1
3	I2, I5
4	I2, I4
5	I1, I3
6	I2, I1, I3
7	I2, I5, I4
8	I2, I1, I5

图 4 参数频数统计及数据中参数重新排序

Fig. 4 Parameter frequency statistics and parameter reordering in each piece of data

3.2 干扰量数据格式处理

FP-Growth 算法需要将实在意义的干扰量参量转化为数据格式标签,即将不同数据类型的干扰量参量进行统一编号;将带有编号的数据导入 FP 算法进行挖掘分析;最后根据其编号参量具体的物理意义,得出相应的关联规则。将干扰量参量编号后的结果如表 5 所示。

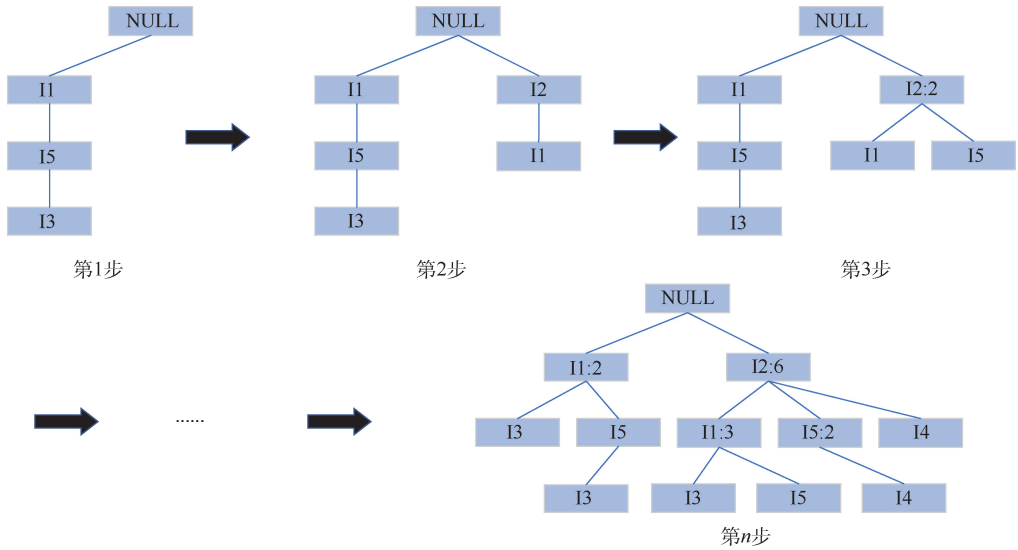
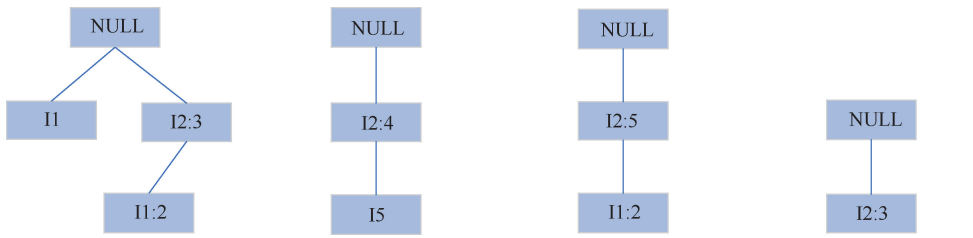


图 5　读入数据生成 FP 树

Fig. 5　Data reading to generate FP tree



(a) 末节点I5的条件FP树　(b) 末节点I4的条件FP树　(c) 末节点I3的条件FP树　(d) 末节点I1的条件FP树

图 6　各末节点对应的条件 FP 树

Fig. 6　Conditional FP tree for each final node

表 5　干扰量参量编号

Table 5 Disturbance variable numbering

干扰量	参数	取值设定
诱饵投掷参量	N_s/个	12(4),24(5),36(6),48(7)
	N_e/个	1(8),2(9)
	N_f/个	4(10),6(11),12(12)
	T_l/s	0.5(13),1(14),1.5(15),2(16),2.5(17)
	T_i/s	0.1(18),0.2(19),0.3(20),0.4(21),0.5(28.2)
	T_e/s	0.2(23),0.4(24),0.6(25),0.8(26),1(27)
	V_f/(m·s^{-1})	10(28),20(29),30(30),40(31)
	θ_l/(°)	0(32),30(33),60(34),90(35),120(36),150(37)
	θ_v/(°)	-90(38),-60(39),0(40),30(41),60(42),90(43)
弹目相对态势	θ_m/(°)	10(44),20(45),30(46),70(47),80(48),90(49),130(50),140(51),150(52)
	D_r/km	(0,3](53),(3,4](54),(4,5](55),(5,6](56),(6,7](57),(7,8](58),(8,9](59),(9,10](60)
目标机动	TMM	无机动(61),左转弯(62),右转弯(63),跃升(64)
脱靶量	MD/m	MD1(1),MD1(2),MD1(3)

注:括号内数字为编号。

4　试验分析

通过第 2 节构建的数据立方体统计出数据集中各事务的支持度计数,来减少 FP-Growth 算法遍历数据集的次数,从而提高该算法生成关联规则的收敛速度,收敛速度的比较将在 4.3 节中论述。本文所构建三维立方体,将最后挖掘算法得出的关联规则进行划分:①各方体内部的关联规则挖掘,旨在得出各干扰量之间应该如何组合使用才能对导弹的干扰效果影响最大,从而保证载机的安全;②方体间外联关联规则的挖掘,通过对各个方体之间的关联规则进行比较,得出对脱靶量影响较大的干扰量。

4.1　各方体内部关联规则

由于试验中感兴趣的低命中率结果总数相对于整个数据集合为少数,采取低支持度度量指标,

防止将感兴趣的规则剔除掉,故将最小支持度指标(minSup)设置为 0.5%,最小置信度指标(minConf)设置为 0.7;Kulc 度量指标的边界条件设定为 Kulc < 0.4 或 Kulc > 0.6;若 0.4 ≤ Kulc ≤ 0.6 时,则判断不平衡比度量指标(IR),并且最小不平衡比度量指标设置为 minIR > 0.9。

1) MD1 < 12 m 内部关联规则

对 MD1 子立方体中的数据,按照所设置的各指标条件,进行关联规则的挖掘,对挖掘出的关联规则进行筛选。其部分的关联关系如表 6 所示。

表 6 MD1 部分内部关联规则(MD1 < 12 m)

Table 6 Inner association rules of MD1 (MD1 < 12 m)

序号	Conf	Sup	Kulc	IR	规则
1	0.989	0.011	0.571	0.968	13,52,59→23,1
2	0.976	0.006	0.538	0.925	37,47,53→17,1
3	0.972	0.007	0.549	0.901	21,37,53→17,1
4	0.966	0.005	0.533	0.935	5,43,53→17,1
5	0.961	0.009	0.559	0.933	9,36,53→17,1

通过分析表 6 可得到以下结论:

① 序号 1 规则。该规则的置信度达到了 0.989,说明了导弹做迎头攻击,在弹目距离较远(8 ~ 9 km)的情况下,若载机诱饵起投时间早,则载机不应该再采取小间隔投掷策略,否则很快就会将诱饵投掷完毕,在拦截末期目标将无诱饵可用,载机未将导弹完全诱骗,导弹又重新将载机识别,最终脱靶量较低。

② 序号 2 ~ 5 规则。表 6 规则都有较高的置信度,说明当弹目距离非常近的时候(小于 3 km),红外诱饵弹的起投时机不应该过晚,否则有可能诱饵还未投掷结束,导弹就已经命中载机,干扰效果极差。

通过表 6 分析可知,根据弹目距离选择诱饵的投掷方式对干扰的效果起到决定性作用,弹目距离远应选择“晚起投 + 大间隔”的投掷方式;弹目距离近应选择“早起投 + 小间隔”的投掷方式。

2) 12 m ≤ MD2 < 16 m 内部关联规则

针对该立方体中的数据,进行关联规则的挖掘,挑选部分感兴趣的关联规则进行分析,规则在表 7 呈现。

表 7 MD2 部分内部关联规则(12 m ≤ MD2 < 16 m)

Table 7 Inner association rules of MD2 (12 m ≤ MD2 < 16 m)

序号	Conf	Sup	Kulc	IR	规则
1	0.996	0.005	0.513	0.975	34,53→61,2
2	0.988	0.006	0.512	0.975	44,56→8,2
3	0.975	0.008	0.517	0.967	6,32→8,2

通过表 7 规则可以得到如下结论:

① 序号 1 规则。在迎头攻击的情况下,载机无需采取机动就可以获得较好的干扰效果。

② 序号 2、3 规则。若载机的诱饵总数固定,导弹采取尾后攻击方式,当弹目距离较远时,载机可以采取每次单投 1 个的投掷策略,目的是为了延长红外诱饵的干扰时间,使其持续干扰。

通过表 7 分析可知,导弹要尽量避免采用迎头攻击方式,若导弹采取尾后攻击,载机要想办法调整干扰策略,使其干扰时间延长,从而增强干扰效果。

3) MD3 ≥ 16 m 内部关联规则

该组数据的挖掘结果如表 8 所示。

表 8 MD3 部分内部关联规则(MD3 ≥ 16 m)

Table 8 Inner association rules of MD3 (MD3 ≥ 16 m)

序号	Conf	Sup	Kulc	IR	规则
1	0.986	0.006	0.555	0.918	17,60,62→27,3
2	0.977	0.006	0.706	0.588	17,60,62→22,3
3	0.984	0.006	0.511	0.978	10,13,53→9,3
4	0.965	0.005	0.543	0.988	13,24,53→10,3

对表 8 数据加以分析可得如下结论:

① 序号 1、2 规则。当弹目距离较远(9 ~ 10 km)时,若载机采取“晚起投 + 大间隔”的投掷策略,导弹脱靶的概率达到 0.986。

② 序号 3、4 规则。当弹目距离较近(小于 3 km)时,载机想要保证对导弹的干扰效果,应采取“早起投 + 小间隔”的投掷策略。

由表 8 分析可得,弹目距离在 9 ~ 10 km 之间时,采取“晚起投 + 大组间”的投掷策略;在弹目距离小于 3 km 时,采取“早起投 + 小间隔”的投掷策略,可以获得良好的干扰效果。

通过对各立方体内部关联规则的分析,可以得到以下结论:

1) 导弹要尽量避免采用迎头攻击的攻击方式。

2) 根据弹目距离的远近,选择合适的起投时机以及投掷间隔。

3) 干扰参量的随意组合会造成干扰效果的下降。

4.2 方体间外联关联规则

按照 4.1 节设置同样的度量指标边界条件,对方体间外联关联规则进行筛选,得到如表 9 所示部分方体间外联关联规则。

选取表 9 中最具代表性的部分方体间外联关联规则进行分析。

1) 序号 1 规则。17,26,53→1,过了 2.5 s 后

表 9　部分方体间外联关联规则

Table 9　Part of outer association rules

序号	Conf	Sup	Kulc	IR	规则
1	0.986	0.010	0.508	0.984	17,26,53→1
2	0.983	0.006	0.505	0.990	52,53,63→1
3	0.980	0.006	0.505	0.991	32,44,64→1
4	0.973	0.006	0.505	0.990	11,52,64→1
5	0.972	0.007	0.506	0.988	37,49,59,64→1
6	0.992	0.005	0.510	0.979	17,34,63→2
7	0.987	0.005	0.511	0.978	52,55,61→2
8	0.977	0.006	0.512	0.977	35,52,61→2
9	0.975	0.007	0.514	0.973	44,58,64→2
10	0.971	0.006	0.513	0.974	33,48,63→2
11	0.996	0.009	0.537	0.927	50,58,61→3
12	0.987	0.008	0.531	0.937	9,52,58,62→3
13	0.978	0.009	0.537	0.927	8,51,58→3
14	0.974	0.009	0.534	0.932	35,51,58→3
15	0.972	0.007	0.527	0.946	52,60,63→3
16	0.967	0.009	0.537	0.927	49,54,64→3

目标以每组组间间隔为 0.8 s 开始投掷红外诱饵弹,弹目距离小于 3 km,在这种干扰态势下导弹脱靶量小于 12 m 的置信度达到 0.986。分析原因为弹目距离过近,若目标投掷诱饵的起投时机晚,并且采用大间距时间间隔投掷策略,导弹可能已经进入了盲飞区,此时诱饵已经不能对导弹造成干扰,故干扰效果极差,在这种情况应采取“快速、短频”的投饵方式。

2) 序号 6 规则。17,34,63→2,过了 2.5 s 后目标以 60°水平投掷夹角投掷诱饵弹,同时采取右机动规避导弹,在这种干扰态势下导弹脱靶量介于 12 ~ 16 m,置信度为 0.992。分析原因为目标诱饵弹起投时机晚,目标采取右机动规避导弹的同时投掷诱饵,并且投掷方向与机动方向一致,这样就使得诱饵尽可能的覆盖了目标,同时增加了覆盖时长,这样对导弹的干扰效果大大加强,最终造成脱靶量上升。

3) 序号 3 与序号 9 规则。规则 3 与规则 9 相比,若导弹采用尾后攻击方式,同时目标采取跃升机动方式进行规避,此时采取的诱饵投掷方式应该是垂直方向投掷而不是水平方向,原因是要让诱饵尽可能的将发动机尾焰进行遮挡,使得诱饵在导引头视场中停留时间足够长,从而增加导引头的识别误差,降低命中精度。

4) 序号 11 ~ 15 规则。通过这 5 条规则可以发现都有一个共同的干扰因素,即导弹以大进入角方式对目标进行追踪,进入角大于 130°的范围属于前半球面,攻击方式属于迎头攻击,在这种情况下目标蒙皮对发动机尾焰遮挡严重,目标释放的红外诱饵弹为高辐射能量区域,极易诱骗导弹朝诱饵弹方向飞去,以致于目标花费极低的代价,仍可造成较大脱靶量,如规则 11,目标无需采取任何机动方式,导弹依旧未跟踪到目标。

5) 序号 4 规则。由序号 11 ~ 15 规则分析可知,当导弹做迎头攻击时,目标只需花费少许代价便可使干扰效果显著。但如果导弹做迎头攻击时,同时目标采取跃升机动方式进行规避,这使得被蒙皮遮蔽的发动机尾焰暴露在导引头视场中,从而使得干扰效果大大下降,故导弹为迎头攻击时,载机最好不要采取跃升机动对其进行规避。

6) 序号 2、7、11 规则。虽然序号 7 规则进入角大于序号 11 规则的进入角,但是由于弹目距离的不同造成干扰效果不同。在序号 11 规则中,由于弹目距离远,虽然导引头有足够时间识别目标并做弹体姿态调整,但由于目标蒙皮对尾焰的遮蔽,使得目标与诱饵的红外辐射能量差异过于悬殊,导引头识别误差越来越大,最终跟丢目标;序号 7 规则,由于弹目距离较近,诱饵还未将导弹完全诱骗,此时导弹已经进入盲飞区,将不再受诱饵干扰,故干扰效果下降,脱靶量下降;而序号 2 规则,最后的结果更能说明这一结论,弹目距离太近(小于 3 km),即使导弹以大进入角进入,相对态势较差,但仍能保证较高命中率。

7) 序号 5 与序号 16 规则。当导弹的进入角为 90°时,目标采取跃升机动进行规避,此时导弹速度与目标速度垂直,两者相对速度达到最大,目标采取的机动方式虽然导引头能够识别,但是在这种情况下导弹没有足够的过载对弹体姿态进行快速调整,如序号 16 规则,又因初始弹目距离较近,就更没有足够的时间给弹体进行姿态的调整,同时加上红外诱饵弹的干扰,在“双重干扰”的条件下,造成了最后脱靶量较大。但序号 5 规则,红外诱饵弹的投掷方向与导弹来袭方向同向且异侧,干扰效果大大降低,与此同时弹目距离较远,导弹有足够的时间进行弹体姿态的调整,故脱靶量大大降低。

通过对上述关联规则的分析可以得到以下结论:导弹的进入角与弹目距离是影响干扰效果的主要因素;诱饵的投掷方向要与目标机动及导弹来袭方向配合使用,使得诱饵弹对发动机尾焰的遮蔽时间尽量长,将会提升干扰效果。

4.3　规则挖掘效率验证

为了说明在不同数据量下的传统 FP-Growth 算法与基于数据立方体处理数据后的 FP-Growth

算法(DC_FP)的挖掘效率,本试验分别在数据量为1 000,5 000,10 000,15 000,20 000,50 000进行2种算法的对比,置信度设置为0.7、支持度设置为0.5,对比结果如表10所示。

该对比试验是在同一电脑硬件下完成的,由对比的结果可知,传统的FP-Growth算法与DC_FP算法在小数据规模下,算法效率相当,但随着数据规模的增大,传统的FP-Growth算法与DC_FP算法之间的计算效率差距也越来越大,尤其当数据量超过10 000条后,传统的FP-Growth算法的挖掘效率下降明显,原因是由于数据维数较多(13维),造成每条数据划分为的子问题多,同时又要遍历数据集统计各项集的支持度计数,故造成了算法性能的下降。

表10 不同数据量下算法效率对比

Table 10 Comparison of algorithm efficiency with different data volumes

数据量	T_{FP}/s	T_{DC_FP}/s
1 000	56.4	55.9
5 000	534.6	108.7
10 000	5 063.7	305.3
15 000	9 886.2	502.9
20 000	20 043.2	745.8
50 000	*	1 034.9

注:*代表数据量过大,结果未获得。

5 结 论

基于数据立方体对FP-Growth算法进行改进,并对红外抗干扰数据进行关联规则挖掘,最终得到以下结论:

1)导弹的进入角与弹目距离是影响脱靶量大小的主要因素,在作战过程中导弹应避免使用迎头攻击的作战方式,尾后攻击命中率比迎头攻击大。

2)根据弹目距离远近选择合适的掷饵策略,长距离选择"晚起投、大间隔"掷饵方式;短距离选择"早起投,短间隔"掷饵方式。

3)不同的干扰参量不能随意组合,否则会造成干扰性能的下降,应该保证干扰时长,通过干扰量的配合使用尽量隐藏发动机尾焰,减少其在导引头视场中的暴露面积。

4)通过对比试验,传统的FP-Growth算法与DC_FP算法随着数据规模的增大,两者的关联规则挖掘效率差距也越来越大。通过统计出数据集中各事务的支持度计数,来减少FP-Growth算法遍历数据集的次数,可以较好的提高挖掘效率。

参考文献(References)

[1] 李立坤.精确制导技术现状及发展方向[J].航空兵器,2004(1):1-4.
LI L K. Present situation and development direction of precision guidance technology[J]. Aero Weaponry, 2004(1):1-4(in Chinese).

[2] 王晓铭,王玫.防空导弹武器抗干扰试验技术[J].上海航天,2013,30(2):34-38.
WANG X M, WANG M. ECCM test technology for antiaircraft missile weapon system[J]. Aerospace Shanghai, 2013, 30(2): 34-38(in Chinese).

[3] 胡朝晖,闫杰.红外空空导弹抗干扰性能评估方法研究[J].空军工程大学学报(自然科学版),2008,25(3):45-52.
HU Z H, YAN J. An evaluation method of infrared missile countermeasures performance[J]. Journal of Air Force Engineering University(Natural Science Edition), 2008, 25(3):45-52(in Chinese).

[4] 王涛,王祥.红外导弹抗干扰能力多维度评估方法[J].红外技术,2014,36(7):573-576.
WANG T, WANG X. An evaluation method with five dimensions for infrared missile anti-jamming[J]. Infrared Technology, 2014, 36(7):573-576(in Chinese).

[5] 李慎波,李韬锐,童中翔,等.导弹战术参数对面源红外诱饵干扰效能影响[J].红外与激光工程,2018,47(7):1-10.
LI S B, LI T R, TONG Z X, et al. Influence of tactical parameters of missile on jamming effectiveness of surface source infrared decoy[J]. Infrared and Laser Engineering, 2018, 47(7): 1-10(in Chinese).

[6] 牛得清,伍友利,徐洋,等.点源红外诱饵干扰下环境复杂度量化建模[J].红外与激光工程,2020,49(2):211-219.
NIU D Q, WU Y L, XU Y, et al. Quantification modeling for environmental complexity under point source infrared decoy interference[J]. Infrared and Laser Engineering, 2020, 49(2):211-219(in Chinese).

[7] 马潮,陆志洋,余海鸣,等.红外成像导引头抗干扰性能评估方法研究[J].空天防御,2018,1(4):44-47.
MA C, LU Z F, YU H M, et al. Research on evaluation method of anti-interference performance of infrared seeker[J]. Air & Space Defense, 2018, 1(4):44-47(in Chinese).

[8] 葛辰杰,陆志洋,洪泽华,等.基于支持向量回归与多核集成的红外成像导引头抗干扰性能评估方法[J].上海航天,2019,36(5):94-98.
GE C J, LU Z F, HONG Z H, et al. Anti-interference capability evaluation based on SVM regression and multi-kernel boosting[J]. Aerospace Shanghai, 2019, 36(5):94-98(in Chinese).

[9] 杨开,李少毅,张凯,等.基于朴素贝叶斯分类器的空中红外目标抗干扰识别方法研究[J].飞控与探测,2019,2(4):62-70.
YANG K, LI S Y, ZHANG K, et al. Research on anti-jamming recognition method of aerial infrared target based on naive Bayes classifier[J]. Flight Control & Detection, 2019, 2(4): 62-70(in Chinese).

[10] 周卫文,康美玲,周泽强. 红外诱饵对成像制导导弹影响规律研究[J]. 红外与激光工程,2019,48(12):118-124.
ZHOU W W, KANG M L, ZHOU Z Q. Research of infrared flares influence mechanism on the imaging guidance missile [J]. Infrared and Laser Engineering, 2019, 48(12): 118-124 (in Chinese).

[11] 张志波,童中翔,王超哲,等. 视线方向飞机红外辐射特性建模与仿真[J]. 激光与红外,2013,43(8):890-895.
ZHANG Z B, TONG Z X, WANG C Z, et al. Modeling and simulation of airplane infrared characteristic along the sightline [J]. Laser & Infrared, 2013, 43(8): 890-895 (in Chinese).

[12] LI S, ZHANG K, YIN J, et al. A study on IR target recognition approach in aerial jamming environment based on Bayesian probabilistic model[J]. IEEE Access, 2019, 7: 50300-50316.

[13] XU Y, FANG Y W, WU Y L, et al. Association rule mining for the infrared countermeasure by the PF-Growth algorithm[C]// Proceedings of the 37th China Control Conference, 2018: 8043-8048.

[14] FREDERIC G, LAKSHMANAN L V S. nD-SQL: A multi-dimensional language for interoperability and OLAP[C]// International Conference on VLDB, 1998: 134-145.

[15] 冯玉才,向隆刚. 维上带层次数据立方的自底向上计算[J]. 小型微型计算机系统,2004(8):1477-1481.
FENG Y C, XIANG L G. Bottom-up computation of data cube with hierarchy along dimension[J]. Journal of Chinese Computer Systems, 2004(8): 1477-1481 (in Chinese).

[16] TAN P N, STEINBACH M, KUMAR V. Association analysis: Basic concepts and algorithms[M]. 3rd ed. Beijing: Posts and Telecom Press, 2011: 200-240.

Application of data cube association mining in infrared anti-jamming

WU Xin[1], WU Youli[1,*], NIU Deqing[1], CAI Yuxuan[1], XU Yang[2], CHEN Bian[1]

(1. School of Aeronautical Engineering, Air Force Engineering University, Xi'an 710038, China;
2. China Aerodynamics Research and Development Center, Mianyang 621000, China)

Abstract: This study aims at analyzing the coupling relationship between disturbance variables and miss distance of air-to-air missiles in infrared anti-jamming. Massive data are obtained by setting disturbance variables through an experimental simulation platform. The data cube is then constructed to preprocess the massive data and compute the support count of each transaction. The processed data are used to mine the association rules between disturbance variables and miss distance by FP-Growth algorithm. Analysis of the association rules show that the entry angle of the missile and the distance between the missile and the target are the main factors for miss distance. The disturbance variables cannot be combined freely, but should be selected according to the attack direction of the missile and the target maneuver mode. Otherwise, the jamming effect will be greatly reduced. A group of comparative experiments is also set up to verify the high efficiency of FP-Growth algorithm based on the data cube in processing the miss distance data with a high dimension.

Keywords: data cube; data processing; FP-Growth algorithm; association rule; infrared anti-jamming

Received: 2021-03-01; **Accepted**: 2021-06-11; **Published online**: 2021-07-06 15:40
URL: kns.cnki.net/kcms/detail/11.2625.V.20210706.1304.002.html
Foundation item: National Natural Science Foundation of China (61973253)
* **Corresponding author**. E-mail: 18359943663@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0101

基于混合稀疏表示的二维压缩感知 SAR 成像

熊世超[1], 倪嘉成[1,*], 张群[1,2], 罗迎[1,2], 王岩松[3]

(1. 空军工程大学 信息与导航学院, 西安 710077; 2. 复旦大学 波散射与遥感信息国家教育部重点实验室, 上海 200433;
3. 中国人民解放军 93303 部队, 沈阳 110000)

摘 要: 压缩感知(CS)理论在合成孔径雷达(SAR)成像中应用广泛。针对包含城市、河流等区域的非稀疏场景压缩感知 SAR 成像,提出基于近似观测模型的混合稀疏表示(MSR)压缩感知 SAR 成像方法。该方法将复杂的 SAR 图像分解成点、线、面,并将线、面分别通过离散余弦变换和曲波变换转换到稀疏域,使压缩感知的稀疏性条件得以满足,通过求解基于近似观测模型的二维压缩感知优化问题重建非稀疏场景的 SAR 图像。所提方法能够实现降采样率条件下对包含城市、河流等非稀疏场景区域的成像,仿真场景和实测场景成像结果表明了所提方法的有效性。

关 键 词: 合成孔径雷达; 压缩感知; 稀疏表示; 近似观测; 曲波变换
中图分类号: V243.2; TN957.5
文献标志码: A **文章编号**: 1001-5965(2022)11-2314-11

合成孔径雷达(synthetic aperture radar,SAR)是一种主动式、全天时、全天候、高分辨的对地观测雷达系统,具有广泛的应用场景。然而受奈奎斯特采样定律的限制,距离向分辨率的不断提高、距离幅宽的不断扩大,给雷达采样系统、数据存储、数据传输和处理带来严峻的挑战[1-2]。

近年来迅速发展的压缩感知(compressed sensing,CS)理论通过研究稀疏采样信号的精确重建问题,突破了奈奎斯特采样定律,CS 理论指出如果信号本身或在某个变换域是稀疏的,那么就可以实现利用少量观测数据高概率重建原信号[3-5]。为避免传统压缩感知 SAR 成像将回波数据从二维转为一维,Fang 等[6]提出近似观测稀疏 SAR 成像模型,直接对二维回波数据进行处理,不仅降低了存储负担、提高了重构速度,且保留了 SAR 图像的二维结构特征,具有更好的成像效果。在近似观测模型框架下,适用于大斜视的基于非线性调频变标算子的 CS 成像[7]、具有与匹配滤波方法相近计算复杂度的实时稀疏 SAR 成像[8]等方法相继被提出。

一般情况下 SAR 场景散射系数本身并不是稀疏的,在较低的采样率下直接运用压缩感知理论进行成像很难得到高质量 SAR 图像,因此需要进一步寻找 SAR 场景的稀疏表示方法。离散余弦变换(discrete cosine transform, DCT)、小波变换(wavelet transform, WT)、曲波变换(curvelet transform, CT)等都是常用的信号稀疏表示方法[9-11]。Cetin 等[12]指出很难给包含复杂特征的 SAR 场景找到一个通用的稀疏表示方法,为解决此问题,联合字典稀疏表示[13]、结构化稀疏表示[14]、字典学习方法[15]相继被提出,胡长雨等[16]利用字典学习方法解决了逆合成孔经雷达(inverse synthetic aperture radar, ISAR)中的稀疏成像问题,而字典学习方法具有依赖训练数据集、适应未知场景的

收稿日期: 2021-03-02; **录用日期**: 2021-06-18; **网络出版时间**: 2021-07-07 14:47
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210707.1349.002.html
基金项目: 国家自然科学基金 (62001508,61871396); 陕西省自然科学基础研究计划 (2020JQ-480,2020JM-348)
* **通信作者**. E-mail: littlenjc@sina.com

引用格式: 熊世超, 倪嘉成, 张群, 等. 基于混合稀疏表示的二维压缩感知 SAR成像[J]. 北京航空航天大学学报, 2022, 48(11): 2314-2324. XIONG S C, NI J C, ZHANG Q, et al. 2-D compressed sensing SAR imaging based on mixed sparse representation [J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2314-2324(in Chinese).

泛化能力未知等缺点。为了克服传统字典学习方法和固定字典方法的不足,Li 等[17]提出混合稀疏表示(mixed sparse representation, MSR)方法,其基本思路是:将图像分解为点、线、面的组合;对点直接在空域进行重建;对线和面部分则先进行稀疏表示,变换到稀疏域后再进行重建。Li 等[18]将其思想引入到 CS-SAR 成像中来,且考虑到 SAR 图像是复图像的特点,提出了基于近似观测模型的幅度-相位分离的 CS-SAR 成像方法,用 DCT 字典来对面目标进行稀疏表示,而这种方法忽略了 SAR 图像线目标的稀疏表示,导致城市、河流等线条分明的区域成像质量受限。

针对存在的问题,本文提出一种在降采样条件下的非稀疏场景 CS-SAR 成像方法——混合稀疏表示压缩感知 SAR 成像(mixed sparse representation CS-SAR, MSR-CS-SAR),MSR-CS-SAR 方法首先建立 CS-SAR 近似观测模型,其次对将要重建的 SAR 复图像的相位和幅度分离,将幅度分解为点、线、面,对面进行 DCT,对线进行 CT,最后通过迭代阈值算法(iterative thresholding algorithm, ITA)[19]求解基于近似观测模型的 MSR 优化问题,得到重建的非稀疏场景 SAR 图像。除了面的稀疏表示外,MSR-CS-SAR 方法还用 CT 对 SAR 图像中线进行稀疏表示,CT 在表示图像边缘方向性特征上具有优势,使得重建的 SAR 图像中线条的表现力增强。综合点、线、面的分解与重建,实现了较低降采样率条件下非稀疏 SAR 场景的高质量成像。实验对仿真场景和 RADARSAT-1 实测场景进行 CS-SAR 成像,比较了传统基于近似观测的压缩感知算法、文献[18]方法和本文所提 MSR-CS-SAR 方法的成像效果,证明了 MSR-CS-SAR 方法的有效性。

1　CS-SAR 成像模型

图 1 为条带式 SAR 成像几何。SAR 通过发射线性调频(linear frequency modulated, LFM)信号来实现距离向的高分辨,而通过方位向的合成孔径来实现方位高分辨。

在传统 CS-SAR 成像模型中,利用降采样矩阵,实现对回波数据的降采样,得到如下欠定方程[8]:

$$\boldsymbol{y}_s = \boldsymbol{\Theta}\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}_0 \tag{1}$$

式中:$\boldsymbol{y}_s \in \mathbf{C}^{n\times 1}$ 为向量化的回波矩阵 $\boldsymbol{Y}$;$\boldsymbol{x} \in \mathbf{C}^{l\times 1}$ 为向量化的场景散射系数矩阵 $\boldsymbol{X}$;$\boldsymbol{n}_0$ 为加性噪声;$\boldsymbol{\Phi} \in \mathbf{C}^{n\times l}$ 为根据先验知识构造出来的观测矩阵;$\boldsymbol{\Theta}$ 为降采样矩阵。

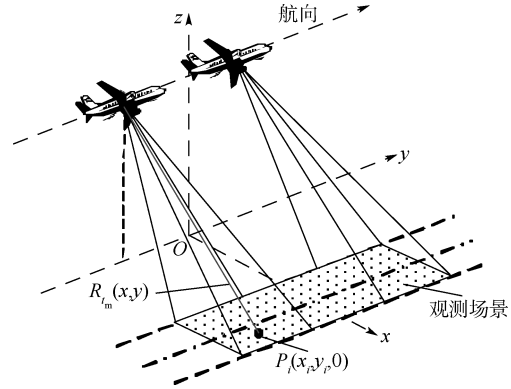


图 1　条带式 SAR 成像几何示意图

Fig. 1　Geometry of strip mode SAR

根据 CS 理论,当 $\boldsymbol{x}$ 具有稀疏性且 $\boldsymbol{\Theta\Phi}$ 满足约束等距性质时,可以将欠定方程的求解转化为一个优化问题,用拉普拉斯乘子法将式(1)转化成无约束的优化问题:

$$\min_{\boldsymbol{x}}\{\|\boldsymbol{y}_s - \boldsymbol{\Theta}\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_q^q\} \tag{2}$$

式中:λ 为正则化参数。

式(2)所描述的 CS-SAR 成像模型对回波和场景散射系数进行向量化,导致观测矩阵 $\boldsymbol{\Phi}$ 需要占据很大的存储空间。近似观测模型将精确观测矩阵 $\boldsymbol{\Phi}$ 用匹配滤波成像算法的一系列解耦的算子来近似表示,不仅能大幅减少观测矩阵的存储空间,还能实现与匹配滤波方法相近的成像速度[6]。线性调频变标算法(chrip scaling algorith, CSA)是一个典型的匹配滤波成像方法[20],成像过程用近似观测模型中的成像算子可表示为

$$M(\boldsymbol{Y}) = \boldsymbol{F}_a^{\mathrm{H}}(\boldsymbol{F}_a\boldsymbol{Y}\circ\boldsymbol{H}_1\boldsymbol{F}_r\circ\boldsymbol{H}_2\boldsymbol{F}_r^{\mathrm{H}}\circ\boldsymbol{H}_3) \tag{3}$$

式中:$\boldsymbol{Y}$ 为回波数据;"∘"为 Hadamard 积;$\boldsymbol{F}$ 和 $\boldsymbol{F}^{\mathrm{H}}$ 分别为离散傅里叶变换(discrete fourier transform, DFT)矩阵和逆 DFT 矩阵;下标 a 和 r 分别表示方位向和距离向;$\boldsymbol{H}_1$、$\boldsymbol{H}_2$、$\boldsymbol{H}_3$ 分别为二次相位函数、距离压缩和徙动校正相位函数、方位压缩和剩余相位补偿相位函数。则 CSA 的近似观测逆成像算子为

$$I(\boldsymbol{X}) = \boldsymbol{F}_a^{\mathrm{H}}(\boldsymbol{F}_a\boldsymbol{X}\circ\boldsymbol{H}_3^*\boldsymbol{F}_r\circ\boldsymbol{H}_2^*\boldsymbol{F}_r^{\mathrm{H}}\circ\boldsymbol{H}_1^*) \tag{4}$$

式中:$\boldsymbol{X}$ 为场景散射系数矩阵;$(\cdot)^*$ 为对矩阵进行共轭转置操作。可以得到如下 CS-SAR 成像优化问题:

$$\min_{\boldsymbol{X}}\{\|\boldsymbol{Y}_s - \boldsymbol{\Theta}_a I(\boldsymbol{X})\boldsymbol{\Theta}_r\|_{\mathrm{F}}^2 + \lambda\|\boldsymbol{X}\|_q^q\} \tag{5}$$

式中:$\|\cdot\|_{\mathrm{F}}$ 为 Frobenius 范数;$\boldsymbol{\Theta}_a \in \mathbf{R}^{N_a\times n_a}$($N_a < n_a$)和 $\boldsymbol{\Theta}_r \in \mathbf{R}^{N_r\times n_r}$($N_r < n_r$)分别为方位向和距离向的采样矩阵;$\lambda$ 为正则化参数,控制着数据保真度和稀疏度之间的平衡;$\boldsymbol{Y}_s$ 为降采样收集的回波。

此优化问题可以方便地用 ITA 求解,本文选

取 q 为1时的迭代软阈值算法进行求解。在ITA求解过程的每轮迭代中，按照式(6)进行更新为

$$\boldsymbol{X}^{(i+1)} = E_{1,\lambda\mu}(\boldsymbol{X}^{(i)} + \mu M(\boldsymbol{\Theta}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{X}^{(i)})\boldsymbol{\Theta}_{\mathrm{r}})\boldsymbol{\Theta}_{\mathrm{r}}^{\mathrm{T}})) \tag{6}$$

式中：μ 为迭代的步进大小，与迭代过程是否收敛密切相关，λ 和 μ 参数的具体选择用文献[6]的方法；$E_{1,\lambda\mu}$ 为软阈值算子，其式为

$$E_{1,\lambda\mu}(x) = \begin{cases} \mathrm{sign}(x)(|x| - \lambda\mu) & |x| > \lambda\mu \\ 0 & \text{其他} \end{cases} \tag{7}$$

2 基于近似观测模型的MSR-CS-SAR成像

非稀疏场景的压缩感知SAR成像存在2个问题：①非稀疏场景SAR图像包含丰富的地理特征，要找到一个通用的稀疏表示字典对所有特征进行稀疏表示是很困难的；②SAR图像本身是复图像，在运用压缩感知理论进行重建SAR图像的过程中需要把SAR图像的相位和幅度分开求解。针对这2个问题，可以借鉴混合稀疏表示模型[17]和幅度-相位联合求解[18]的思想，首先将SAR复图像的相位和幅度分开，之后将SAR图像的幅度分解为点、线、面，对线进行CT，对面进行DCT，建立关于幅度-相位的优化问题，联合求解SAR图像的幅度和相位。

SAR图像利用点、线、面3种分量来表示不同的特性是合理的，原因如下：根据雷达目标散射机理，SAR图像本质上反应了地面散射点的散射系数，强散射点一定是存在的；线分量和面分量从光学图像中借用而来，虽然SAR复图像与光学图像本身有区别，但是对SAR复图像进行去相位、幅度归一化操作之后生成的SAR图像相当于光学图像的灰度图，因此，从SAR图像中分解出线分量、面分量也是合理的。3.1节实验验证了这种分解方式的合理性。

由于本文只需要SAR图像的幅度，将SAR复图像分解为幅度-相位分离、点线面幅度相加的形式：

$$\boldsymbol{S} = \boldsymbol{A} \cdot (\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3) \tag{8}$$

式中：$\boldsymbol{S}$ 为SAR复图像；$\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$ 分别为点、线、面的幅度；$\boldsymbol{A}$ 为SAR图像相位因子。

结合近似观测模型和混合稀疏表示模型，CS-SAR成像问题可以写为

$$\min_{\boldsymbol{X}_1,\boldsymbol{X}_2,\boldsymbol{X}_3} \{\|\boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A} \cdot (\boldsymbol{X}_1 + \boldsymbol{X}_2 + \boldsymbol{X}_3))\boldsymbol{\Theta}_{\mathrm{r}}\|_{\mathrm{F}}^2 + \lambda_1\|\boldsymbol{X}_1\|_1 + \lambda_2\|C(\boldsymbol{X}_2)\|_1 + \lambda_3\|D(\boldsymbol{X}_3)\|_1\} \tag{9}$$

式中：$D(\cdot)$ 为DCT；$C(\cdot)$ 为CT；λ_1、λ_2、λ_3 分别为对点、线、面的正则化约束参数；式(9)后3项分别为对点分量、DCT后的面分量、CT后的线分量进行正则化稀疏约束。

通过求解式(9)，点、线、面都可以被重建出来，最后将三者相加即为重建的SAR图像。求解过程按照 $\boldsymbol{X}_1$、$\boldsymbol{X}_2$、$\boldsymbol{X}_3$ 的顺序分3步进行，在每一步的求解中又细分为残差更新和幅度估计，在点部分求解过程中，不需要相位估计，在线、面部分求解之前进行相位估计，具体步骤详述如下。

首先初始化点、线、面幅度 $\boldsymbol{X}_1^{(0)}$、$\boldsymbol{X}_2^{(0)}$、$\boldsymbol{X}_3^{(0)}$ 和相位因子 $\boldsymbol{A}^{(0)}$。第 $i+1$ 次迭代，在求解点部分过程中，第1步先更新残差：

$$\boldsymbol{Y}_0 = \boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A}^{(i)} \cdot (\boldsymbol{X}_2^{(i)} + \boldsymbol{X}_3^{(i)}))\boldsymbol{\Theta}_{\mathrm{r}} \tag{10}$$

式中：$\boldsymbol{X}_2^{(i)}$、$\boldsymbol{X}_3^{(i)}$ 分别为第 i 次迭代 $\boldsymbol{X}_2$、$\boldsymbol{X}_3$ 的值；$\boldsymbol{A}^{(i)}$ 为第 i 次迭代的相位因子。第2步把线、面幅度 $\boldsymbol{X}_2^{(i)}$、$\boldsymbol{X}_3^{(i)}$ 固定，求解点的幅度 $\boldsymbol{X}_1^{(i+1)}$，即令 λ_2、λ_3 为0，则式(9)转化为

$$\min_{\boldsymbol{X}_1}\|\boldsymbol{Y}_0 - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{X}_1^{(i)})\boldsymbol{\Theta}_{\mathrm{r}}\|_{\mathrm{F}}^2 + \lambda_1\|\boldsymbol{X}_1^{(i)}\|_1 \tag{11}$$

然后利用ITA对式(11)进行求解，可以得到 $\boldsymbol{X}_1$ 的迭代式为

$$\boldsymbol{X}_1^{(i+1)} = E_{1,\lambda_1\mu_1}(\boldsymbol{X}_1^{(i)} + \mu_1(\boldsymbol{A}^{(i)*} \cdot M(\boldsymbol{\Theta}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{Y}_0 - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{X}_1^{(i)})\boldsymbol{\Theta}_{\mathrm{r}})\boldsymbol{\Theta}_{\mathrm{r}}^{\mathrm{T}})) \tag{12}$$

式中：$\boldsymbol{A}^*$ 为 $\boldsymbol{A}$ 的共轭。令

$$\boldsymbol{Y}_1 = \boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{X}_1^{(i+1)})\boldsymbol{\Theta}_{\mathrm{r}} \tag{13}$$

在求解线、面之前，通过求解优化问题式(14)来得到相位。

$$\min_{\boldsymbol{A}}\|\boldsymbol{Y}_1 - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A}^{(i)} \cdot (\boldsymbol{X}_2^{(i)} + \boldsymbol{X}_3^{(i)}))\boldsymbol{\Theta}_{\mathrm{r}}\|_{\mathrm{F}}^2 \tag{14}$$

可以通过匹配滤波方法估计相位信息[18]：

$$\boldsymbol{A}^{(i+1)} = \exp(\mathrm{angle}(M(\boldsymbol{\Theta}_{\mathrm{a}}^{\mathrm{T}}\boldsymbol{Y}_1\boldsymbol{\Theta}_{\mathrm{r}}^{\mathrm{T}}))) \tag{15}$$

在求解面部分过程中，第1步更新残差为

$$\boldsymbol{Y}_2 = \boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A}^{(i+1)} \cdot (\boldsymbol{X}_1^{(i+1)} + \boldsymbol{X}_2^{(i)}))\boldsymbol{\Theta}_{\mathrm{r}} \tag{16}$$

式中：angle(·)为获取相位角。

第2步把求出来的相位 $\boldsymbol{A}^{(i+1)}$ 和之前的点、线幅度 $\boldsymbol{X}_1^{(i+1)}$、$\boldsymbol{X}_2^{(i)}$ 固定，通过求解式(17)的优化问题得到面的幅度 $\boldsymbol{X}_3^{(i+1)}$：

$$\min_{\boldsymbol{X}_3}\|\boldsymbol{Y}_2 - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A}^{(i+1)} \cdot \boldsymbol{X}_3^{(i)})\boldsymbol{\Theta}_{\mathrm{r}}\|_{\mathrm{F}}^2 + \lambda_3\|D(\boldsymbol{X}_3^{(i)})\|_1 \tag{17}$$

用ITA求解 $\boldsymbol{X}_3^{(i+1)}$ 的迭代表达式为

$$D(\boldsymbol{X}_3^{(i+1)}) = E_{1,\lambda_3\mu_3}(D(\boldsymbol{X}_3^{(i)}) + \mu_3(\boldsymbol{A}^{(i+1)*} \cdot M(\boldsymbol{\Theta}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{Y}_2 - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A}^{(i+1)} \cdot \boldsymbol{X}_3^{(i)})\boldsymbol{\Theta}_{\mathrm{r}})\boldsymbol{\Theta}_{\mathrm{r}}^{\mathrm{T}}))) \tag{18}$$

在求解线部分幅度的过程中，方法跟求解点、面部分相似，先按式(19)更新残差：

$$\boldsymbol{Y}_3 = \boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} I(\boldsymbol{A}^{(i+1)} \cdot (\boldsymbol{X}_1^{(i+1)} + \boldsymbol{X}_3^{(i+1)}))\boldsymbol{\Theta}_{\mathrm{r}} \tag{19}$$

再构建优化问题式(20),通过迭代式(21)对线部分幅度 $\boldsymbol{X}_2^{(i+1)}$ 进行求解。

$$\min_{\boldsymbol{X}_2}\left\|\boldsymbol{Y}_3-\boldsymbol{\Theta}_{\mathrm{a}}I(\boldsymbol{A}^{(i+1)}\cdot\boldsymbol{X}_2^{(i)})\boldsymbol{\Theta}_{\mathrm{r}}\right\|_{\mathrm{F}}^2+\lambda_2\left\|C(\boldsymbol{X}_2^{(i)})\right\|_1 \tag{20}$$

$$C(\boldsymbol{X}_2^{(i+1)})=E_{1,\lambda_2\mu_2}(C(\boldsymbol{X}_2^{(i)})+\mu_2(\boldsymbol{A}^{(i+1)^*}\cdot M(\boldsymbol{\Theta}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{Y}_3-\boldsymbol{\Theta}_{\mathrm{a}}I(\boldsymbol{A}^{(i+1)}\cdot\boldsymbol{X}_2^{(i)})\boldsymbol{\Theta}_{\mathrm{r}})\boldsymbol{\Theta}_{\mathrm{r}}^{\mathrm{T}}))) \tag{21}$$

令 $\boldsymbol{X}^{(i+1)}=\boldsymbol{X}_1^{i+1}+\boldsymbol{X}_2^{i+1}+\boldsymbol{X}_3^{i+1}$。当经过一定的迭代次数,或 $(\|\boldsymbol{X}^{(i+1)}-\boldsymbol{X}^{(i)}\|_2^2/\|\boldsymbol{X}^{(i)}\|_2^2)\leqslant\delta$ 时(δ 为一个很小的值),则完成了整个求解过程。本节给出了基于近似观测模型的 CS-SAR 算法流程,图 2 为算法流程。

1) 输入降采样回波数据 $\boldsymbol{Y}_{\mathrm{s}}$,降采样矩阵 $\boldsymbol{\Theta}_{\mathrm{a}}$、$\boldsymbol{\Theta}_{\mathrm{r}}$,最大迭代次数 $i_{\max}$,CSA 近似观测成像算子 $M(\cdot)$,CSA 近似观测逆成像算子 $I(\cdot)$。

2) 迭代次数 $i=0$,初始化点、线、面幅度,$\boldsymbol{X}_1^{(0)}=\boldsymbol{X}_2^{(0)}=\boldsymbol{X}_3^{(0)}=\mathrm{zeros}(l_{\mathrm{a}},l_{\mathrm{r}})$,初始化相位因子 $\boldsymbol{A}^{(0)}=\mathrm{zeros}(l_{\mathrm{a}},l_{\mathrm{r}})$。

3) 点部分重建:根据式(10)和式(12),计算并更新第 $i+1$ 次的残差 $\boldsymbol{Y}_0$ 和点部分幅度 $\boldsymbol{X}_1^{(i+1)}$。

4) 面部分重建:根据式(15)、式(16)和式(18),计算并更新第 $i+1$ 次的相位 $\boldsymbol{A}^{(i+1)}$、残差 $\boldsymbol{Y}_2$、和面部分幅度 $\boldsymbol{X}_3^{(i+1)}$。

5) 线部分重建:根据式(19)和式(21),计算并更新第 $i+1$ 次的残差 $\boldsymbol{Y}_3$ 和线部分幅度 $\boldsymbol{X}_2^{(i+1)}$。

6) 点、线、面组合:$\boldsymbol{X}^{(i+1)}=\boldsymbol{X}_1^{(i+1)}+\boldsymbol{X}_2^{(i+1)}+\boldsymbol{X}_3^{(i+1)}$。

7) 令 $i=i+1$。

8) 当 $i=i_{\max}$,或满足迭代结束条件时,结束迭代,输出 $\boldsymbol{X}^{i_{\max}}$,否则跳转到第 3 步。

分析 MSR-CS-SAR 算法的计算复杂度,设 I_{n} 为

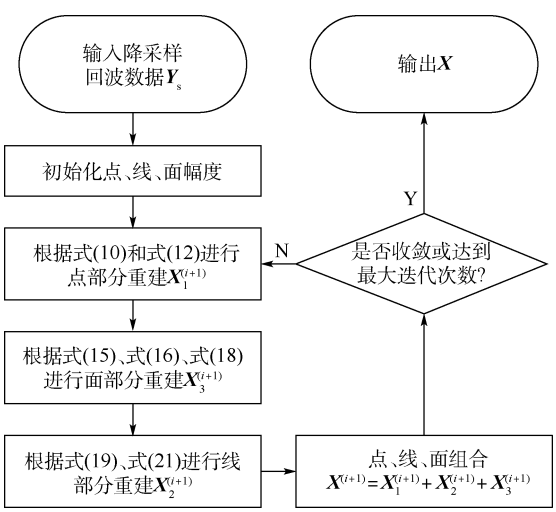


图 2　MSR-CS-SAR 算法的流程

Fig. 2　Flow block of MSR-CS-SAR algorithm

算法整体迭代次数,$M_{\mathrm{n}}=n_{\mathrm{a}}\times n_{\mathrm{r}}$ 为回波采样点数,n_{a} 为回波方位采样点数,n_{r} 为回波距离采样点数。计算量主要在于成像算子和逆成像算子,匹配滤波成像算法 CSA 的算法复杂度为 $O(M_{\mathrm{n}}\log_2 M_{\mathrm{n}})$,而阈值算子 $E_{1,\lambda u}$ 的复杂度为 $O(M)$,CT 的复杂度[21]为 $O(M_{\mathrm{n}}\log_2\sqrt{M_{\mathrm{n}}})$,因此,经过 I 次迭代 MSR-CS-SAR 方法的计算复杂度为

$$O(I_{\mathrm{n}}(M_{\mathrm{n}}\log_2 M_{\mathrm{n}}+M_{\mathrm{n}}+M_{\mathrm{n}}\log_2\sqrt{M_{\mathrm{n}}})) \tag{22}$$

与参考文献[18]相比,MSR-CS-SAR 算法增加了对线部分进行稀疏表示的曲波变换,因此,在优化求解中主要增加了曲波变换所带来的时间损失。

分析 MSR-CS-SAR 算法的存储空间。传统 CS-SAR 成像把 $n_{\mathrm{a}}\times n_{\mathrm{r}}$ 回波矩阵和 $l_{\mathrm{a}}\times l_{\mathrm{r}}$ 场景散射系数矩阵转化成一维向量的形式,导致观测矩阵维度为 $n_{\mathrm{a}}n_{\mathrm{r}}\times l_{\mathrm{a}}l_{\mathrm{r}}$,$l_{\mathrm{a}}$ 为场景方传采样点数,l_{r} 为场景距离采样点数。MSR-CS-SAR 采用基于近似观测模型的 CS-SAR 成像,成像模型中的观测矩阵由一系列的相位函数矩阵和傅里叶变换矩阵组成,这些矩阵的维度远远小于 $n_{\mathrm{a}}n_{\mathrm{r}}\times l_{\mathrm{a}}l_{\mathrm{r}}$,节省了存储空间。

进一步分析算法的收敛性。MSR-CS-SAR 算法在分别更新点、线、面时,求解的优化问题是经典的以 $L1$ 范数为正则项的 LASSO 问题,迭代软阈值算法是解决此类凸优化问题的方法之一,其收敛性在文献[20]中已经得到证明。MSR-CS-SAR 算法在迭代软阈值算法的基础上,利用了交替迭代的思想,分别对点、线、面及相位分别进行优化。每次迭代中,在更新某一部分的解时固定其他部分,如在更新点部分的解时,固定线、面、相位的部分,通过匹配滤波算子提取残差中的信息,经过软阈值算子的阈值收缩之后,大的值被保留,而小的值被置 0。在求解相位时,使用近似解。交替迭代中每一个子步骤分别用迭代阈值算法更新迭代解,残差都逐渐减小,这样在不断的迭代中,总的迭代残差会越来越小,最终几乎不变,算法收敛。

3　实验结果

3.1　可行性实验

首先进行包含多种特征的 SAR 图像的分解实验来验证 MSR-CS-SAR 算法的可行性,如图 3 所示,以图 3(a)CSA 成像结果图作为参照图,在全采样的情况下用 MSR-CS-SAR 算法将参照图分解为点部分图 3(b)、线部分图 3(c)和面部分图 3(d),然后 3 个部分相加可以得到整体合成

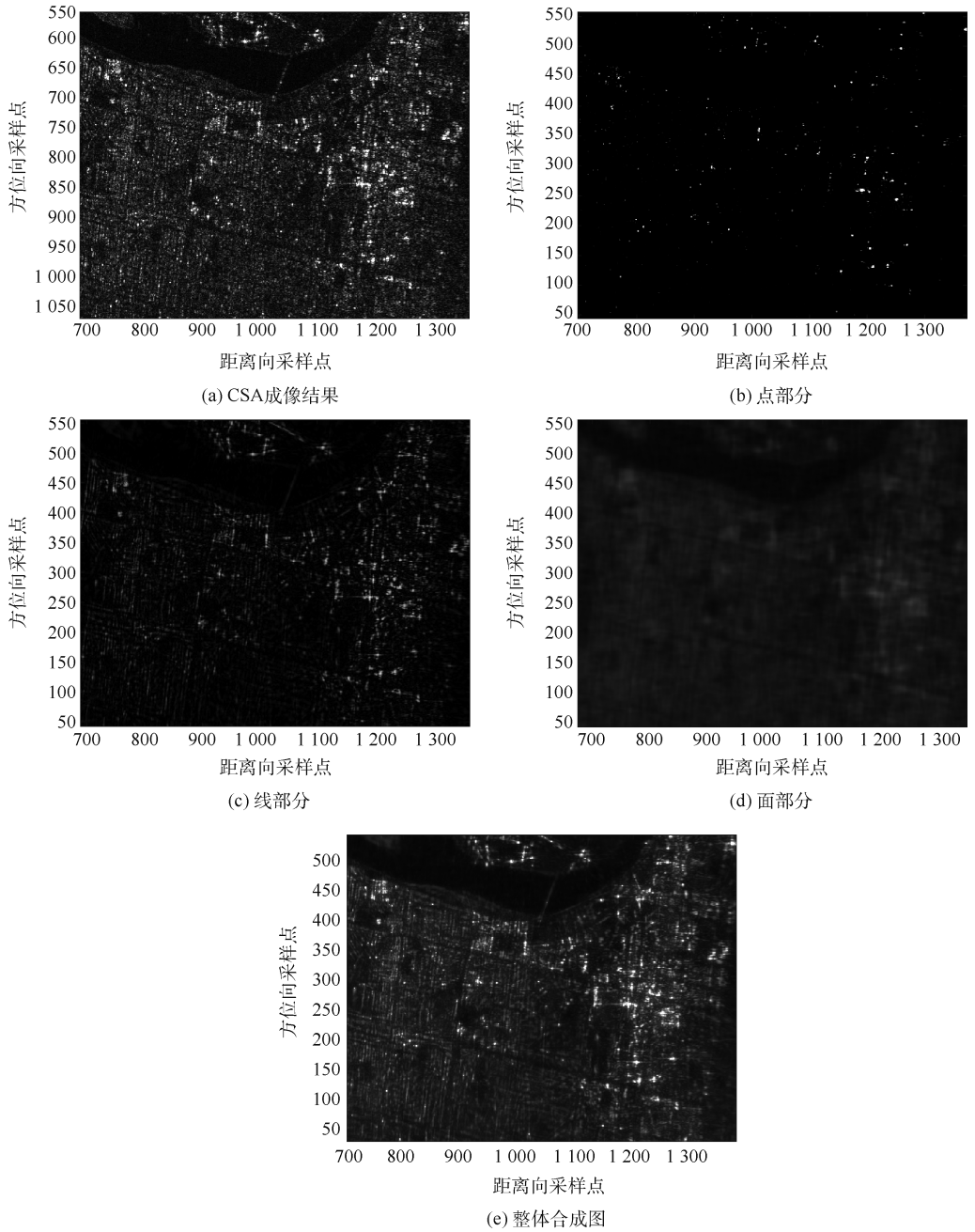


图3 SAR图像分解实验

Fig. 3 Decomposition of SAR

图3(e)。整体合成图3(e)与参照图3(a)相似度很高,说明可以直接将SAR图像表示为点、线和面相加的形式,由此验证了将复杂的SAR图像分解成点、线、面分量是合理的。

3.2 有效性实验

为了验证MSR-CS-SAR方法的有效性,在本节进行了仿真场景和实测场景的实验。回波的降采样率定义为方位向降采样率乘以距离向降采样率,例如,方位向降采样率为80%,距离向降采样率为80%,则回波降采样率为80% × 80% = 64%。实验设置36%、64%、81%这3种回波降采样率,对比了文献[6]的近似观测模型基本方法、文献[18]方法、MSR-CS-SAR方法的成像效果。基本方法和MSR-CS-SAR方法所采用的近似观测模型成像算子均为CSA算子。实验中的曲波变换使用CurveLab Toolbox[20]。

3.2.1 仿真场景实验

在仿真场景实验中,设置了场景中心距雷达10 000 m,场景范围为100 m × 100 m的区域,发射的线性调频信号带宽为75 MHz,载波频率为10 GHz,雷达平台速度为110 m/s。仿真场景如图4所示,包含5个点目标、1个长方形面目标,1个圆形面目标和由2条直线、2条曲线组成的线目标。线目标中包含2种不同散射系数的线条,

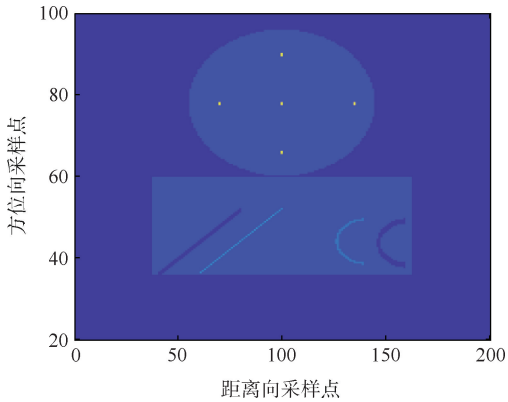


图 4　仿真场景示意图

Fig. 4　Simulation scene

一种比面部分的散射系数大,称之为蓝线,而另一种比面部分的散射系数小,称之为黑线。直线和曲线各包含 1 条蓝线和 1 条黑线。设置蓝黑 2 种不同散射系数的线条,目的在于对比 3 种方法的优劣。

第 i 次迭代中,分别使用式(23)、式(24)来计算超参数 μ、λ:

$$\mu^{(i)} = \frac{\left\| \boldsymbol{\Theta}_{\mathrm{a}} \boldsymbol{I}(\Delta \boldsymbol{X}^{(i)}) \boldsymbol{\Theta}_{\mathrm{r}} \right\|_{\mathrm{F}}^{2}}{\left\| \Delta \boldsymbol{X}^{(i)} \right\|_{\mathrm{F}}^{2}} \tag{23}$$

$$\lambda^{(i)} = \frac{\left| \boldsymbol{X}^{(i)} \right|_{K_0+1}}{\mu_i} \tag{24}$$

式中:$\Delta \boldsymbol{X}^{(i)} = M(\boldsymbol{\Theta}_{\mathrm{a}}^{\mathrm{T}}(\boldsymbol{Y}_{\mathrm{s}} - \boldsymbol{\Theta}_{\mathrm{a}} \boldsymbol{I}(\boldsymbol{X}^{(i)}) \boldsymbol{\Theta}_{\mathrm{r}}) \boldsymbol{\Theta}_{\mathrm{r}}^{\mathrm{T}})$,$\left| \boldsymbol{X}^{(i)} \right|_{K_0+1}$ 为 $|\boldsymbol{X}^{(i)}|$ 的第 K_0+1 个最大值[6];K_0 与 $|\boldsymbol{X}^{(i)}|$ 的稀疏度有关,$|\boldsymbol{X}^{(i)}|$ 越稀疏,K_0 的值就越小,因此点、线、面 3 种不同分量的求解中,首先需要分析场景中点、线、面 3 种不同分量在各自稀疏域的稀疏度,然后根据稀疏度来选择各自的 K_0。依据人工调参得到的经验可知,点、线、面分量的稀疏度分别设置为总数据量的 0.5% ~1%、0.5% ~3%、2.5% ~30%。实验选取的点、线、面稀疏度分别为 0.4%、2.56%、32.1%。

图 5 ~图 7 分别为 36%、64%、81% 降采样率条件下 3 种成像方法的对比结果。由图 5 ~图 7 的结果可以看出,在不同的降采样率条件下,3 种方法都能重建出点目标和蓝线目标。从图 5 ~图 7,降采样率越高,重建出的点目标与蓝线目标越清晰。这是因为 3 种成像方法都对图像空域的稀疏性进行了约束,而点目标和蓝线目标都具有空域稀疏的特性。在 36% 降采样率条件下,3 种方法重建的面目标都存在不同程度的条纹现象。在 64%、81% 降采样率条件下,文献[18]和 MSR-CS-SAR 方法重建的面目标条纹现象消失,符合仿真场景面目标光滑的特点,但基本方法仍存在

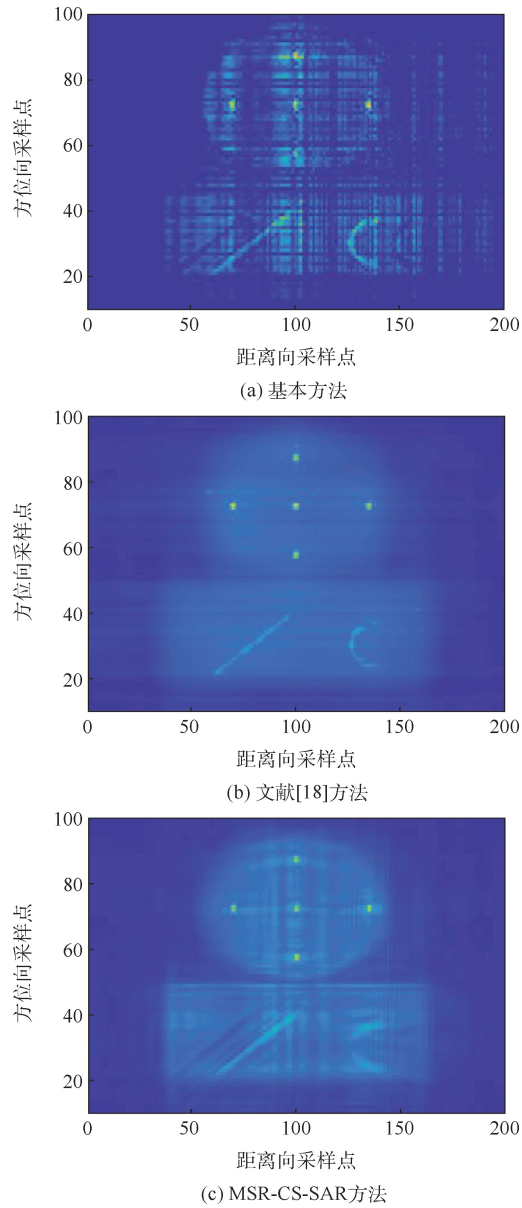
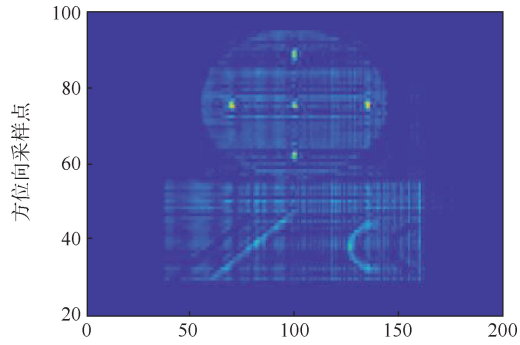


图 5　36% 降采样率仿真场景成像效果对比

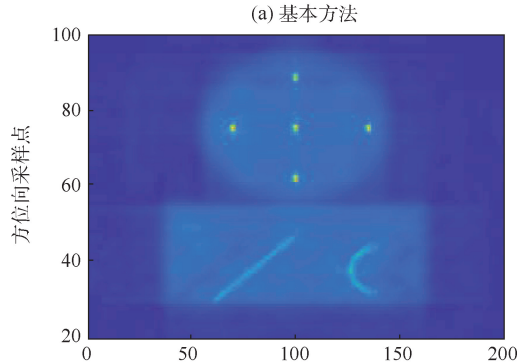
Fig. 5　Comparison of simulation scene SAR images under 36% downsample

条纹。除了点目标、蓝线目标之外,MSR-CS-SAR 方法还能较好重建出黑直线,且面目标的边缘比文献[18]的结果更锐利。这是由于 MSR-CS-SAR 方法除了对面部分进行 DCT 之外,还利用 CT 对线条进行稀疏表示,增强了 SAR 图像中线部分的表现力。蓝线可以看成由多个相邻的强散射点组成,其本身在图像中是稀疏的,可以通过算法中的点部分求解,但是黑线并不是由强散射点组成的,只能通过面部分求解,而面部分的 DCT 对线条特征的表示能力有限,所以文献[18]方法能够高质量重建蓝线,而对黑线的表现力较弱。由于 MSR-CS-SAR 方法使用了 CT 对线部分进行稀疏表示,蓝线和黑线都能被较好的重建。

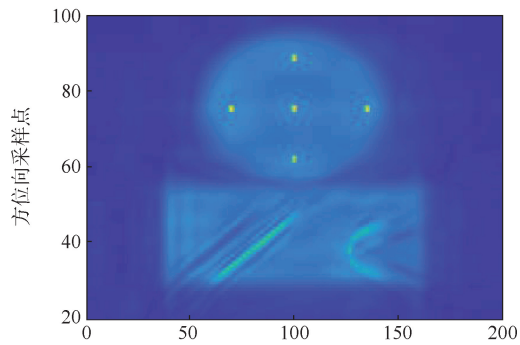
为进一步评估 MSR-CS-SAR 方法的成像质量,



(a) 基本方法



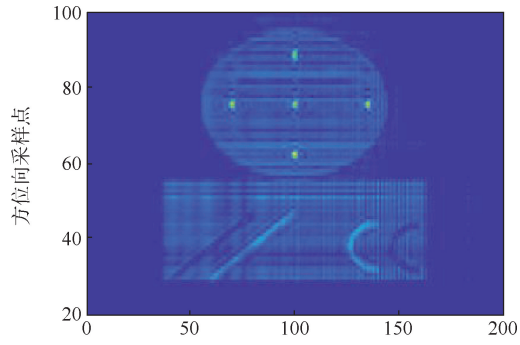
(b) 文献[18]方法



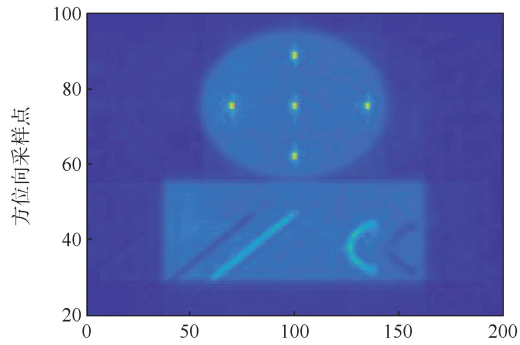
(c) MSR-CS-SAR方法

图 6 64% 降采样率仿真场景成像效果对比

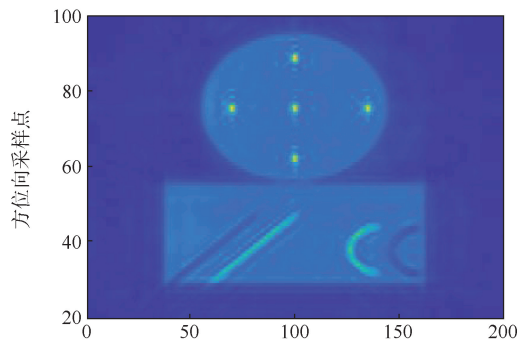
Fig. 6 Comparison of simulation scene SAR images under 64% downsample



(a) 基本方法



(b) 文献[18]方法



(c) MSR-CS-SAR方法

图 7 81% 降采样率仿真场景成像效果对比

Fig. 7 Comparison of simulation scene SAR images under 81% downsample

以全采样数据 CSA 成像结果作为参考图像,基本方法、文献[18]方法、MSR-CS-SAR 方法的成像结果作为待评图像,采用式(25)的相对均方误差(relative mean square error, RMSE),式(26)的峰值信噪比(peak signal noise ratio,PSNR)和平均结构相似性(mean structural similarity index matrix, MSSIM)作为衡量图像质量的指标,MSSIM 通过对图像进行分区,按式(27)计算各分区的结构相似性(structural similarity index matrix,SSIM),然后再求取平均值得到。对于 RMSE,其值越小,成像质量越好,对于 PSNR 和 MSSIM,其值越大,成像质量越好。

$$\mathrm{RMSE} = \frac{\|\boldsymbol{X} - \hat{\boldsymbol{X}}\|}{\|\boldsymbol{X}\|} \tag{25}$$

$$\mathrm{PSNR} = 20\lg\left(\frac{\mathrm{MAX}_I}{\sqrt{\|\boldsymbol{X} - \hat{\boldsymbol{X}}\|}}\right) \tag{26}$$

$$\mathrm{SSIM} = \frac{(2\mu_X\mu_{\hat{X}} + C_1)(2\sigma_{X\hat{X}} + C_2)}{(\mu_X^2 + \mu_{\hat{X}}^2 + C_1)(\sigma_X^2 + \sigma_{\hat{X}}^2 + C_1)} \tag{27}$$

式中:μ_X、$\mu_{\hat{X}}$分别为 $\boldsymbol{X}$、$\hat{\boldsymbol{X}}$ 的均值;σ_X、$\sigma_{\hat{X}}$分别为 $\boldsymbol{X}$、$\hat{\boldsymbol{X}}$ 的标准差;$\sigma_{X\hat{X}}$为 $\boldsymbol{X}$ 和 $\hat{\boldsymbol{X}}$ 的协方差;C_1、C_2 为常数。

3 种降采样率条件下不同方法仿真场景成像结果的 RMSE、PSNR 和 MSSIM 如表 1 所示。最好的指标结果用加粗字体强调。结果显示,MSR-CS-SAR 方法的 RMSE 更低,而 PSNR 和 MSSIM 更大,说明具有更好的成像质量。

3.2.2 实测场景实验

实测场景实验采用 RADARSAT-1 卫星实测

数据。实验选取了河流入海口附近包含丰富点、线、面特征的一部分非稀疏场景,验证 MSR-CS-SAR 方法在较低降采样条件下对非稀疏场景的成像能力。与 3.2.1 节实验方法一样,分别在 36%、64%、81% 降采样率条件下,对 3 种方法的成像效果进行对比。为了得到关于 3 种成像方法的 RMSE、PSNR、MSSIM 量化比较结果,选择的参考图像为全采样条件下经过特征增强[17]之后的图像,如图 8 所示,相比 CSA 成像结果,特征增强的图像有更少的噪声。表 2 为 3 种成像方法效果的量化比较结果。最好的指标结果用加粗字体强调。

表 1　不同采样率下仿真场景成像结果的 RMSE、PSNR、MSSIM

Table 1　RMSE, PSNR, MSSIM of simulation scene SAR images under different downsampling ratio

方法	36% 降采样率			64% 降采样率			81% 降采样率		
	RMSE	PSNR	MSSIM	RMSE	PSNR	MSSIM	RMSE	PSNR	MSSIM
基本方法	0.786	−9.268	0.053	0.735	−8.695	0.066	0.503	−5.390	0.109
文献[18]方法	0.316	−1.357	0.181	0.259	0.385	0.219	0.138	5.835	0.365
MSR-CS-SAR 方法	**0.309**	**−1.176**	**0.234**	**0.206**	**2.338**	**0.276**	**0.111**	**7.679**	**0.385**

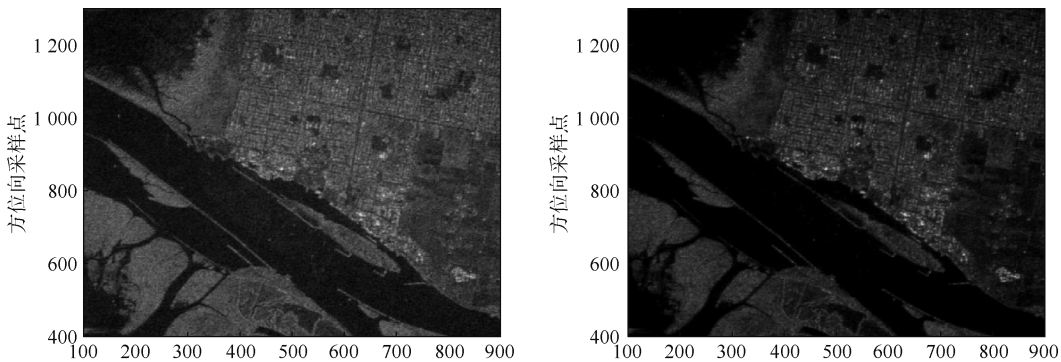


图 8　全采样率实测场景成像效果对比

Fig. 8　Comparison of real scene SAR images under 100% downsample

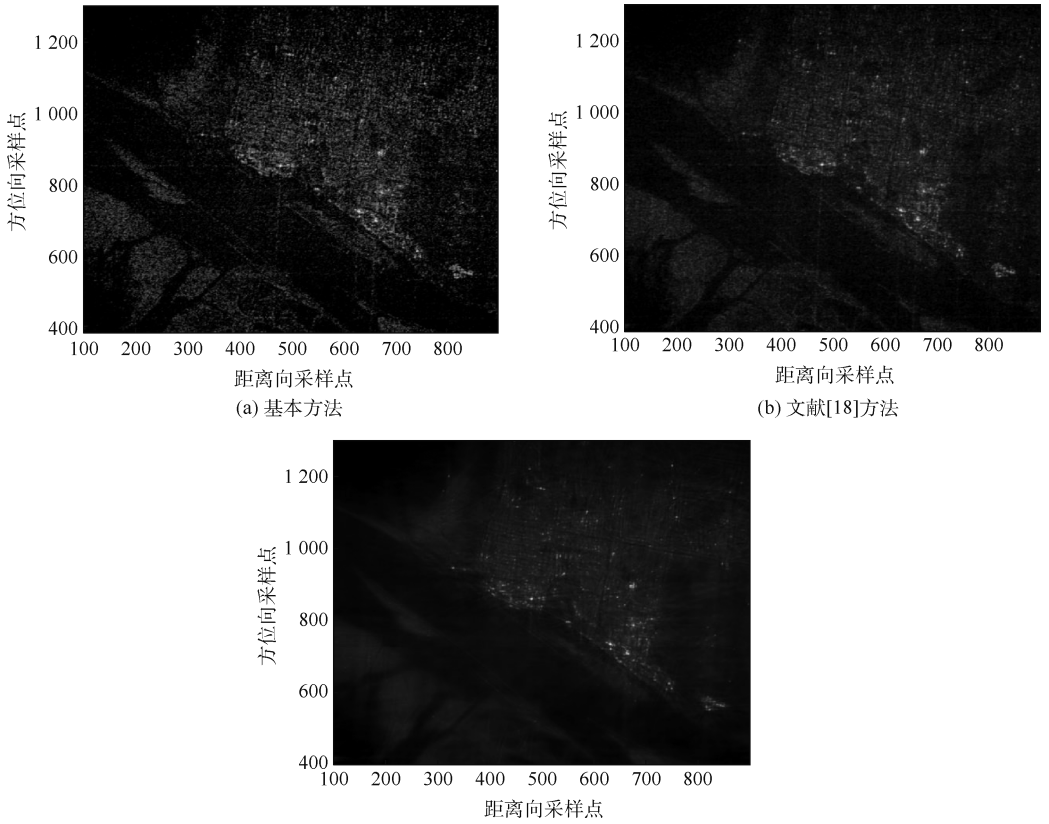
表 2　不同采样率下实测场景成像结果的 RMSE、PSNR、MSSIM

Table 2　RMSE, PSNR, MSSIM of real scene SAR images under different downsampling ratio

方法	36% 降采样率			64% 降采样率			81% 降采样率		
	RMSE	PSNR	MSSIM	RMSE	PSNR	MSSIM	RMSE	PSNR	MSSIM
基本方法	3.387	7.771	0.142	4.157	5.993	0.173	4.831	4.688	0.174
文献[18]方法	0.180	33.28	**0.951**	0.099	38.44	0.975	0.093	39.01	0.979
MSR-CS-SAR 方法	**0.152**	**34.71**	0.950	**0.094**	**38.83**	**0.978**	**0.077**	**40.65**	**0.985**

如图 9 ~ 图 11 所示,在 36%、64%、81% 降采样率条件下,基本方法能较好重建强散射点,但是重建的大陆边缘轮廓和线条不连续,陆地光滑区域存在较多噪点。文献[18]方法和 MSR-CS-SAR 方法都能重建出陆地、河流的轮廓和陆地中多个强散射点目标,而 MSR-CS-SAR 方法重建的大陆边缘和河流轮廓更锐利,同时也很好地重建了城市区域中的道路部分。如图 9 所示,在 36% 降采样率条件下,由于数据缺失严重,3 种方法的成像质量都不高,基本方法和文献[18]方法重建的陆地轮廓模糊、噪点较多,MSR-CS-SAR 方法的陆地区域出现条纹现象,但 MSR-CS-SAR 方法重建出了较为清晰的线条。如图 10、图 11 所示,在 64%、81% 降采样率条件下,相比基本方法和文献[18]方法,MSR-CS-SAR 方法能重建出清晰且连续的线条,包括图像下半部分散射稀疏较大的白色线条及上半部分陆地中散射稀疏较小的城市道路线条。3 种降采样率条件下不同方法仿真场景成像结果的 RMSE、PSNR 和 MSSIM 如表 2 所示。最好的指标结果用加粗字体强调,在 36% 降采样率下,MSR-CS-SAR 方法具有最低的 RMSE 及最高的 PSNR,但 MSSIM 比文献[18]方法略低。在 64%、81% 降采样率下,MSR-CS-SAR 方法的 RMSE 最低,PSNR 和 MSSIM 最大,说明具有更好的成像质量。

综上所述,相比基本方法和文献[18]方法,MSR-CS-SAR 方法能够更好地重建图像中的线部分,包括陆地轮廓、河流轮廓、城市道路重建图像的整体质量较高。这是由于基本方法只对空域的稀疏性进行了约束,不适用于降采样条件下的非稀

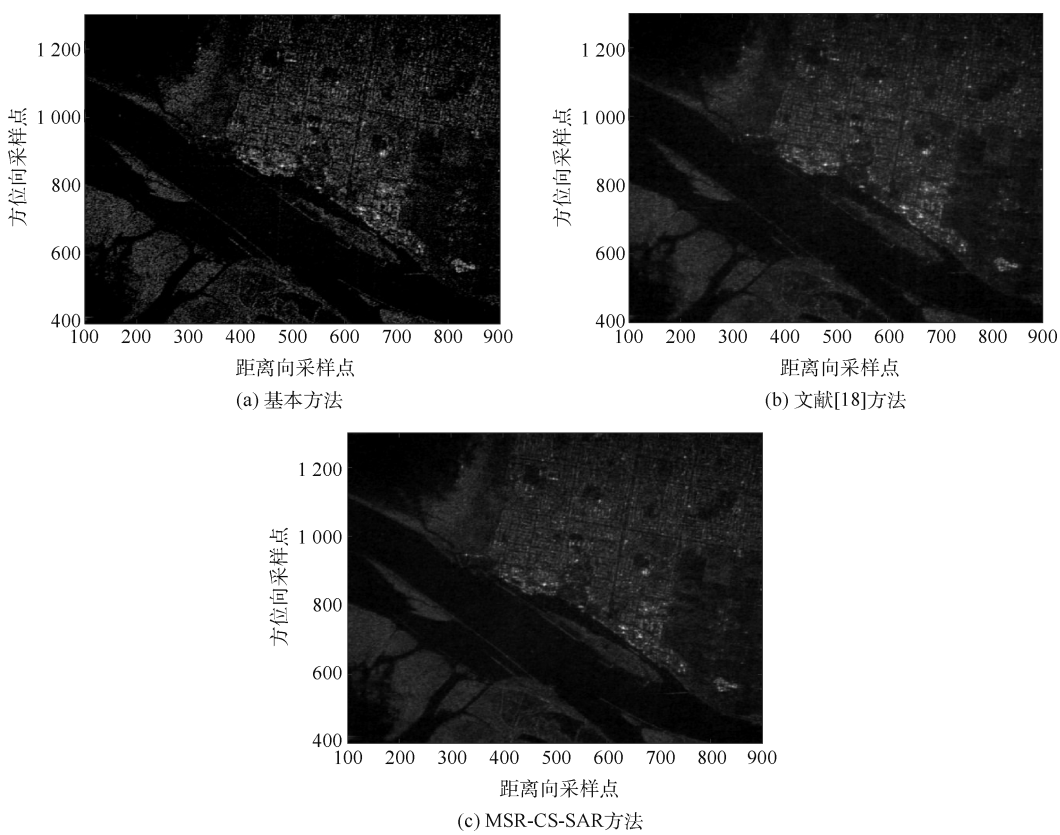


(a) 基本方法 (b) 文献[18]方法

(c) MSR-CS-SAR方法

图9 36%降采样率实测场景成像效果对比

Fig. 9 Comparison of real scene SAR images under 36% downsample



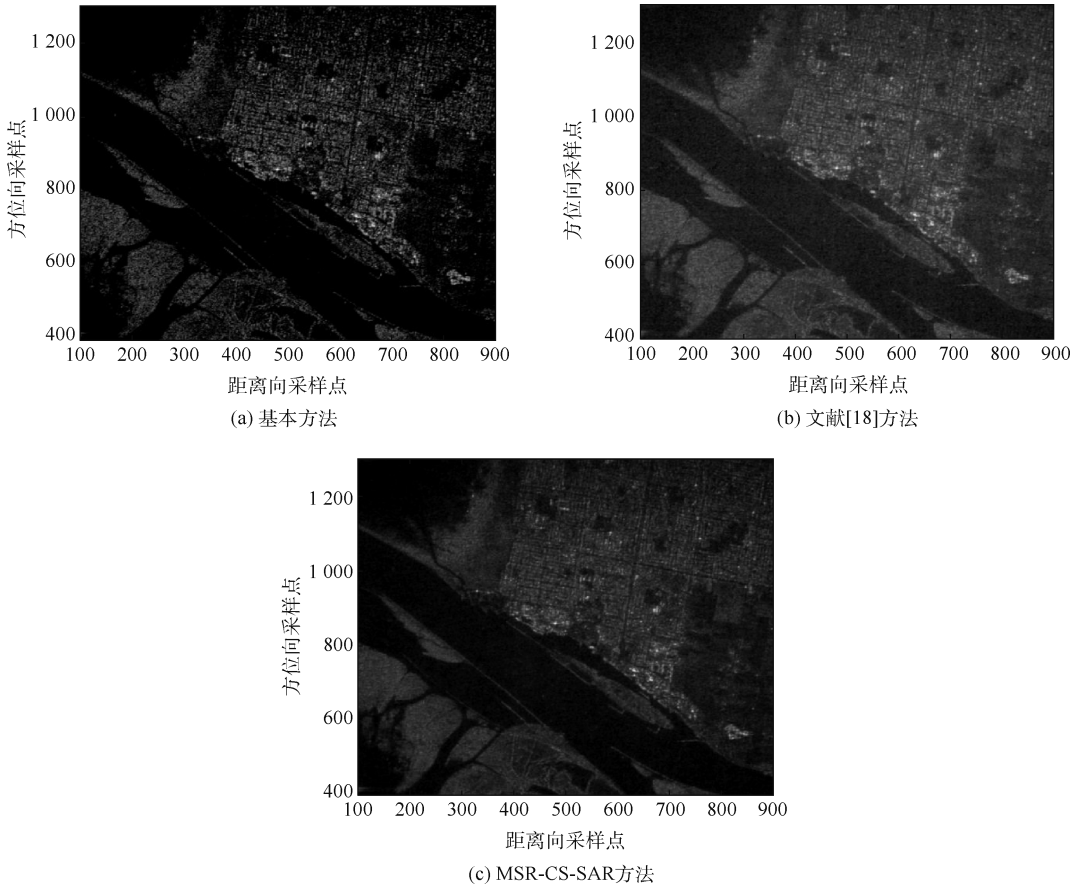
(a) 基本方法 (b) 文献[18]方法

(c) MSR-CS-SAR方法

图10 64%降采样率实测场景成像效果对比

Fig. 10 Comparison of real scene SAR images under 64% downsample

疏场景成像;文献[18]方法对空域稀疏性进行了约束,同时对面目标进行了DCT,但没有对线部分进行稀疏表示,导致成像结果中线部分质量不高;而MSR-CS-SAR方法用CT对线部分也进行了稀疏表



(a) 基本方法　(b) 文献[18]方法

(c) MSR-CS-SAR方法

图 11　81% 降采样率实测场景成像效果对比

Fig. 11　Comparison of real scene SAR images under 81% downsample

示,成像结果中的线部分效果有明显提升。

4　结　论

针对非稀疏场景 CS-SAR 成像,提出了一种基于近似观测模型的 MSR-CS-SAR 成像方法。先将 SAR 复图像的幅度和相位分离,然后将幅度分解成点、线、面,将线、面分别转换到曲波域和 DCT 域,能够增强 SAR 图像中的点、线、面特征。通过实验验证了以下结论:

1) 相比传统 CS-SAR,MSR-CS-SAR 方法运用近似观测模型进行 CS-SAR 成像,求解 SAR 图像的幅度,具有更低的计算复杂度和较小的存储空间占用。

2) MSR-CS-SAR 方法结合了混合稀疏表示模型和近似观测模型,通过求解 CS 优化问题得到高质量的 SAR 图像,实现了降采样条件下的非稀疏场景 SAR 成像。

3) 仿真场景和实测场景实验都表明在降采样条件下,MSR-CS-SAR 方法仍能对包含城市道路、河流、光滑的陆地等成分的非稀疏场景高质量成像。

所提方法考虑的是正侧视条件下的压缩感知 SAR 成像问题,斜视 SAR 或者其他更复杂的情况将是下一步的研究内容。

参考文献 (References)

[1] 吴一戎,洪文,张冰尘,等. 稀疏微波成像研究进展[J]. 雷达学报,2014,3(4):383-395.
WU Y R,HONG W,ZHANG B C,et al. Current developments of sparse microwave imaging[J]. Journal of Radars,2014,3(4):383-395(in Chinese).

[2] 倪嘉成,张群,顾福飞,等. 基于马尔科夫链的单站 SAR 海面场景宽幅高分成像算法[J]. 航空学报,2016,37(12):3793-3802.
NI J C,ZHANG Q,GU F F,et al. Mono-static SAR HRWS imaging algorithm of sea surface based Markov chain[J]. Acta Aeronautica et Astronautica Sinica,2016,37(12):3793-3802(in Chinese).

[3] DONOHO D. Compressed sensing[J]. IEEE Transactions on Information Theory,2006,52(4):1289-1306.

[4] CANDES E J,WAKIN M B. An introduction to compressive sampling[J]. IEEE Signal Processing Magazine,2008,25(2):21-30.

[5] SAMADI S,CETIN M,MASNADI-SHIRAZI M A. Sparse representation-based synthetic aperture radar imaging[J]. IET Radar Sonar and Navigation,2009,5(2):182-193.

[6] FANG J,XU Z B,ZHANG B C,et al. Fast compressed sensing SAR imaging based on approximated observation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2014,7(1):352-363.

[7] 顾福飞,张群,杨秋,等. 基于 NCS 算子的大斜视 SAR 压缩感知成像方法[J]. 雷达学报,2016,5(1):16-24.
GU F F,ZHANG Q,YANG Q,et al. Compressed sensing ima-

ging algorithm for high-squint SAR based on NCS operator[J]. Journal of Radars,2016,5(1):16-24(in Chinese).

[8] BI H,BI G A,ZHANG B C,et al. From theory to application: Realtime sparse SAR imaging[J]. IEEE Transactions on Geoscience and Remote Sensing,2020,58(4):2928-2936.

[9] SHKVARKO Y V,Y ANEZ J I,AMAO J A. Radar/SAR image resolution enhancement via unifying descriptive experiment design regularization and wavelet-domain processing[J]. IEEE Geoscience and Remote Sensing Letters,2016,13(2):152-156.

[10] ZHANG C L,LIU Y L,WAN F Y,et al. Multi-faults diagnosis of rolling bearings via adaptive customization of flexible analytical wavelet bases[J]. Chinese Journal of Aeronautics,2020,33(2):407-417.

[11] 彭才,常智,朱仕军. 基于曲波变换的地震数据去噪方法[J]. 石油勘探,2008,47(5):461-464.
PENG C,CHANG Z,ZHU S J. Noise elimination method based on curvelet trans-form[J]. Geophysical Prospecting for Petroleum,2008,47(5):461-464(in Chinese).

[12] CETIN M,STOJANOVIC I,ONHON O,et al. Sparsity-driven synthetic aperture radar imaging: Reconstruction,autofocusing,moving targets and compressed sensing[J]. IEEE Signal Processing Magazine,2014,31(4):27-40.

[13] SADEGH S,MUJDAT C,MOHAMMAD A M. Multiple feature enhanced SAR imaging using sparsity in combined dictionaries[J]. IEEE Geoscience and Remote Sensing Letters,2013,10(4):821-825.

[14] SHEN F F,ZHAO G H,LIU Z C,et al. SAR imaging with structural sparse representation[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2015,8(8):3902-3910.

[15] MICHAL A,MICHAEL E,ALFRED B. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation[J]. IEEE Transactions on Signal Processing,2006,54(11):4311-4322.

[16] 胡长雨,汪玲,朱栋强. 结合字典学习技术的ISAR稀疏成像方法[J]. 电子与信息学报,2019,41(7):1735-1742.
HU C Y,WANG L,ZHU D Q. Sparse ISAR imaging exploiting dictionary learning[J]. Journal of Electronics & Information Technology,2019,41(7):1735-1742(in Chinese).

[17] LI F,XIN L,GUO Y,et al. A framework of mixed sparse representations for remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing,2017,55(2):1210-1221.

[18] LI B,LIU F L,ZHOU C B,et al. Mixed sparse representation for approximated observation-based compressed sensing radar imaging[J]. Journal of Applied Remote Sensing,2018,12(3):1-21.

[19] DABECHIES I,DEFRISE M,DEMOL C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on Pure and Applied Mathmatics,2004,57(11):1413-1457.

[20] RANEY R K,RUNGE H,BAMLER R,et al. Precision SAR processing using chirp scaling[J]. IEEE Transactions on Geoscience and Remote Sensing,1994,32(4):786-799.

[21] CANDES E,DEMANET L,DONOHO D. Fast discrete curvelet transforms[J]. Multiscale Modeling & Simulation,2006,5(3):861-899.

2-D compressed sensing SAR imaging based on mixed sparse representation

XIONG Shichao[1], NI Jiacheng[1,*], ZHANG Qun[1,2], LUO Ying[1,2], WANG Yansong[3]

(1. Information and Navigation College, Air Force Engineering University, Xi'an 710077, China;
2. Key Laboratory of Wave Scattering and Remote Sensing Information, Fudan University, Shanghai 200433, China;
3. 93303 Unit of PLA, Shenyang 110000, China)

Abstract: Compressed sensing (CS) theory has been applied in synthetic aperture radar (SAR) in the recent years. A 2-D CS SAR imaging method is proposed using mixed sparse representation (MSR) based on approximate observation model in non-sparse scene compressed sensing SAR imaging. Firstly, non-sparse scene with complicated ground features is decomposed into point-like, edges and smooth components. Then, edges and smooth components are transformed into sparse domain by discrete cosine transform and curvelet transform respectively. And based on approximate observation model, SAR images are derived from 2-D CS optimization problem. Owing to the sparse representation method of non-sparse scene, the proposed method can realize high quality SAR imaging for non-sparse scene. Compared to the existing method, the proposed method has better reconstruction quality for region containing distinct edges and lines, such as city and river. Both the simulation scene and real scene experiments demonstrate the effectiveness of the proposed method.

Keywords: synthetic aperture radar; compressed sensing; sparse representation; approximate observation; curvelet transform

Received: 2021-03-02; **Accepted**: 2021-06-18; **Published online**: 2021-07-07 14:47

URL: kns.cnki.net/kcms/detail/11.2625.V.20210707.1349.002.html

Foundation items: National Natural Science Foundation of China (62001508,61871396); Natural Science Basic Research Program of Shaanxi (2020JQ-480,2020JM-348)

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

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0102

基于 SOM 和关联规则的民机运行风险

熊明兰, 王华伟*, 倪晓梅, 蔺瑞管
(南京航空航天大学 民航学院, 南京 211106)

摘 要: 为充分认知民机风险,实现从事故中学习,以重大民机事故(MCAA)为研究对象挖掘出事故深层次的致因特征。针对 MCAA 信息具有可读性差,系统行为具有非线性导致的无法直接获取运行风险信息,难以直接建立事故致因的关联与映射关系,提出一种从 MCAA 中学习民机运行风险特征的方法。针对民机运行特点,结合事故信息及认知可靠性和失误分析方法(CREAM),设计出 MCAA-CREAM 模型,并构建民机多属性技术重大事故数据集。采用自组织映射(SOM)模型,完成对事故的聚类分析和抽象特征映射,以 2D 地图形式增强风险因素可读性,利用关联规则有效挖掘风险因素间的强关联关系。

关 键 词: 民机安全; 风险特征; 事故分析; 自组织映射模型; 关联规则
中图分类号: V328.2
文献标志码: A **文章编号**: 1001-5965(2022)11-2325-10

随着中国国产大飞机项目的不断推进,C919 已取得中国民航局适航认证,ARJ-21 已进入持续载客运营阶段,对民机系统的安全保障能力提出了更高的要求。认知运行风险的基础是数据,但民机作为特高安全性和可靠性系统,难以采集到直接的事故信息,两者之间的矛盾更增加了运行风险分析的难度。因此,为及时识别安全风险,实现安全关口前移,就有必要从事故数据中充分挖掘出有价值的风险信息。

事故分析是实施危险识别和安全管理的基本前提。重大民机事故(major civil aircraft accidents,MCAA)的发生往往不是单一风险因素导致,而是多种风险因素的耦合。通过对风险因素的有效识别,可以反映出民机在运行过程中存在的安全问题,实现事故预防及风险管控[1]。相关学者从不同角度研究了如何准确认知安全风险。Imai 等[2]分析了传感器数据错误对飞行安全的影响,并提出使用容错数据流处理可以减少传感器软件错误的风险。Kelly 和 Efthymiou[3]对 50 次飞机受控飞行进入地形进行分析,强调人为因素是受控飞行进入地形事故的重要原因。Pan 等[4]提出一种多分类信息融合方法对结构性风险进行评估,为结构的健康状况及关键风险因素的深入分析提供了方案。Chen 等[5]在概率基础上对设备故障进行风险评估,所提出的方法可在紧急情况下为设备提供早期预警。董雷霆等[6]提出可以结合数字孪生技术,对飞机的结构疲劳风险进行管控。

随着数据可用性和计算能力的提高,越来越多的学者采用可视化数据挖掘的方法[7]对事故进行研究。Moura 等[8]将人工神经网络方法用于重大事故数据集分析,提出从事故中学习、增强事故沟通风险的方法。余冠华[9]基于人工智能中的自组织映射方法分析了事故的规律及关键致因,并提出相应的改进措施。Zaranezhad 等[10]通过集成人工神经网络、模糊系统和元启发式算法,提出了用于早期事故预测的最佳模型。Asgary 等[11]开发了一种简单新颖的飞机风险地图,为深

收稿日期: 2021-03-02; **录用日期**: 2021-07-16; **网络出版时间**: 2021-07-29 16:19
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210729.1409.001.html
基金项目: 国家自然科学基金 (U1833110)
* **通信作者**. E-mail: wang_hw66@163.com

引用格式: *熊明兰, 王华伟, 倪晓梅, 等. 基于 SOM 和关联规则的民机运行风险[J]. 北京航空航天大学学报, 2022, 48(11): 2325-2334. XIONG M L, WANG H W, NI X M, et al. Operation risk of civil aircraft based on SOM and association rules[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2325-2334(in Chinese).*

度调查飞机坠毁原因提供基础。李哲等[12]通过对飞行安全参数进行风险预测,借助仿真手段为事故演化提供了可视化的分析方法。Zhang 等[13]利用深度学习的方法对事故进行分析和提取特征,进一步提高了事故检测的准确性。

重大民机事故的风险信息分散在不同来源的文本、视频、音频等非结构化数据资料中,不利于风险信息的准确获取。以上方法为重大民机事故的研究提供了一定的方法和技术,但不利于增强对民机运行风险信息挖掘的研究,也不利于提高其可读性和解释性,不能快速有效的传输相关风险信息。

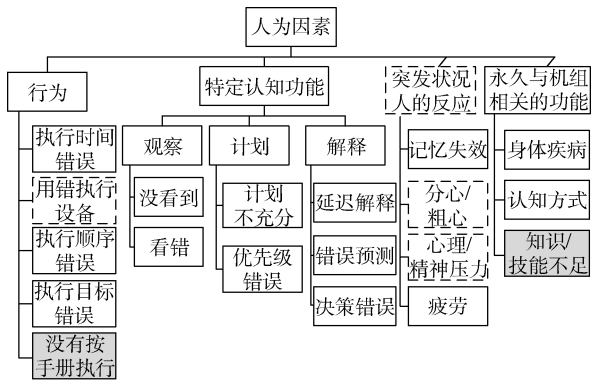
本文通过分析收集到的 MCAA 数据,针对民机运行特点,改进认知可靠性和失误分析方法(cognitive reliability and error analysis method,CREAM),得出 MCAA 的 40 个风险因素。提取的 40 个风险因素系统、全面地表征了民机运行过程中的风险致因,通过逆向挖掘运行风险的深层次特征,探知导致事故的根本原因,实现事故的预防。构建的 MCAA 多属性技术数据集作为模型输入空间样本,实现非结构化数据与结构化数据的转换,结合自组织映射(self-organizing maps,SOM)模型输出民机运行事故 2D 聚类地图,使得风险因素有效集成,以更直观的方式对民机运行风险进行分析与控制,同时利用关联规则(association rule,AR)识别出各风险因素间的影响及关联关系。研究结果表明,本文所提出的方法可以直观、快速的传递民机运行风险信息,深入解析风险要素间的关联关系,挖掘风险致因间的演化规律,为民机运行风险研究提供一种新的技术和手段。

1 基本方法

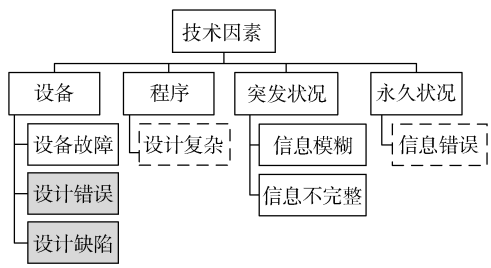
1.1 MCAA-CREAM 模型

美国学者 Hollnagel 提出了 CREAM 的概念[14],用于事故案例原因调查和预测可能发生的不安全事件,为大多数工业领域事故风险因素的调查提供了有效途径。然而,现有的 CREAM 是一种通用模型框架,考虑到民机运行具有区别于一般领域的自身特性,直接将 CREAM 运用到民机运行风险因素分析的效果并不理想,因此需针对民机运行特点对其进行改进。

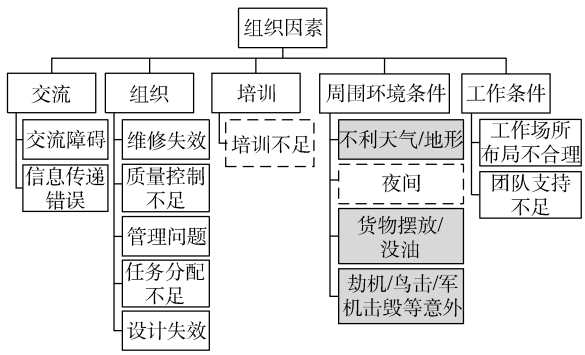
通过分析收集到的 MCAA 数据,并结合民机运行的风险理论及特殊性,对原有的 CREAM 进行事故因素的删除、增加等修改,得到适合民机运行风险研究的 MCAA 认知可靠性和失误分析方法(major civil aircraft accidents-cognitive reliability and error analysis method, MCAA-CREAM)模型。设计出的模型保留了原有 CREAM 的 3 个经典组别:人、技术、组织,如图 1(a)~(c)所示。图中白色框图表示原有 CREAM 存在的因素,灰色填充框图表示增加的因素,虚线框图表示与原有因素功能相近的类别,删除的因素没有在模型中体现。



(a) MCAA-CREAM“人为因素”组别



(b) MCAA-CREAM“技术因素”组别



(c) MCAA-CREAM“组织因素”组别

图 1 MCAA-CREAM 模型

Fig. 1 MCAA-CREAM model

1.2 SOM 模型数据挖掘流程

数据挖掘的目的是揭示事故的通用结构及重要特征。SOM 模型是一种包含无监督学习的过程[15],可通过训练学习[16]将高维数据转化到低维空间,并保留数据的拓扑结构[17]。将 SOM 模型用于事故分析[18],可以更加直观地展示事故数据信息。其算法原理[19]由以下 4 个步骤组成:

1)初始化和归一化。对所有权向量进行随机初始化,设 $\eta(0)$ 为初始学习率,$r(0)$ 为初始邻域半径。

2) 竞争。通过欧氏距离式(1)可以寻找获胜的输出神经元节点 j^*,其权向量对应记为 $\boldsymbol{w}_{j^*}$。

$$d(t) = \min_{j\in\{1,2,\cdots,n\}} \|\boldsymbol{x}_i(t) - \boldsymbol{w}_j(t)\| \tag{1}$$

式中:$\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{im})$ 为输入向量,设有 m 个样本;$\boldsymbol{w}_j = (w_{j1}, w_{j2}, \cdots, w_{jn})$ 为各神经元对应的权向量;$d(t)$ 为计算出的距离。

3) 合作。得到获胜神经元后,使用领域函数 $N_{j^*}(t)$ 定义相邻神经元的最佳匹配单元,通常是高斯函数方程,如下:

$$N_{j^*}(t) = \eta(t)\exp\left(-\frac{\|\boldsymbol{w}_{j^*} - \boldsymbol{w}_j\|}{2r^2(t)}\right)$$
$$j \in m_{j^*}(t) \tag{2}$$

式中:$\eta(t)$ 为学习率;$r(t)$ 为邻域半径。

4) 适应。由于 $\eta(0)$、$r(0)$ 的初始值较大,输出的结果并不好,而在 SOM 模型适应过程中,所有相邻神经元的权重都将按顺序更新,$\eta(t)$ 与 $r(t)$ 不断衰减。通过不断迭代,地图会逐渐收敛,同时神经元根据权向量 $\boldsymbol{w}_j$ 逐渐趋于聚类中心,规则如下:

$$\boldsymbol{w}_j(t+1) = \boldsymbol{w}_j(t) + \eta(t)N_{j^*}(t)[\boldsymbol{x}_i(t) - \boldsymbol{w}_j(t)] \tag{3}$$

其中,学习率 $\eta(t)$ 随时间的推移逐渐减小,如下:

$$0 \leqslant \eta(t) \leqslant \eta(t-1) \leqslant 1 \tag{4}$$

1.3　民机运行风险关联规则

关联规则可以挖掘事故风险之间的耦合关系[20],识别数据中一些关联性较强的规则。目前的关联规则算法主要采用支持度和置信度作为评价指标[21],而由于在 MCAA 事故数据集中特别重大事故数据占比较少,且大多为军机击毁、劫机等意外情况,事故致因单一,因此,将此关联规则方法运用到 MCAA 数据集时发现往往会得到一些无效甚至错误的强关联关系。为解决这一问题,采用考虑提升度及事故等级间权重差异的关联规则方法[9]。

需要指出的是,民机运行相较于其他运输方式更容易受到天气等不可控环境的影响,风险机制更加复杂,产生的事故影响及逃生可能完全不同。尤其是人为因素,往往受到组织因素的影响;同时又与其他因素发生耦合机制及动力学机制。将 SOM 方法和关联规则应用到民机运行安全研究中,能从纷繁复杂的多因素致因机制中,挖掘出潜在的深层次特征,有效识别风险,提升安全风险控制的精准性。充分考虑了民机所特有的风险致因因素,关注在民航安全风险表征的基础上挖掘深层次风险特征,改进后的模型对民机运行安全的研究更具适用性,为民机运行风险研究提供一种新的思路。

设 $I = \{i_1, i_2, \cdots, i_m\}$ 为项目集合,$D = \{t_1, t_2, \cdots, t_n\}$ 为聚类后的 MCAA 事故数据集,由多个事故组成,每个事故 $t_i(i = 1, 2, \cdots, n)$ 均包含项集 I 中的一项或者若干项,满足 $t_i \in I$,风险因素 A 与风险因素 B 之间的关联规则表达形式为 $A \rightarrow B$,其中 $A, B \in I$ 且 $A \cap B \neq \varnothing$。将重大事故等级权重设为 1,特别重大事故等级权重设为 2,则事故数据集 D 中的事故对应等级权值为 $W = \{w_1, w_2, \cdots, w_k, \cdots, w_n\}$,$w_k \in [1, 2]$ 且 w_k 为整数。A 的加权支持度为所有包含 A 的事故 t_i 对应的等级权值 w_i 之和除以事故数据库中所有事故对应的等级权值之和。相关等式如下:

$$W_s(A) = \frac{\sum_{k=1, A=t_k}^{n} w_k}{\sum_{k=1}^{n} w_k} \tag{5}$$

$$W_s(A \rightarrow B) = \frac{\sum_{k=1, A\cap B=t_k}^{n} w_k}{\sum_{k=1}^{n} w_k} \tag{6}$$

$$W_c(A \rightarrow B) = \frac{W_s(A \rightarrow B)}{W_s(A)} \tag{7}$$

$$W_l(A \rightarrow B) = \frac{W_c(A \rightarrow B)}{W_s(B)} = \frac{W_s(A \rightarrow B)}{W_s(A)W_s(B)} \tag{8}$$

式中:W_s 为加权支持度;W_c 为加权置信度;W_l 为加权提升度。设 W_{mins} 为最小支持度,W_{minc} 为最小置信度,若在事故数据库 D 中,$W_s(A \rightarrow B) \geqslant W_{mins}$,$W_c(A \rightarrow B) \geqslant W_{minc}$,则 $A \rightarrow B$ 为强关联规则。

2　基于 SOM 和关联规则的民机运行风险研究算法设计

2.1　重大民机事故多属性技术的数据集构建

根据改进的 MCAA-CREAM 模型,对收集到的 MCAA 进行数据处理和信息提取。在 1.1 节中,将民机运行风险因素分为 7 个技术风险因素,19 个人为风险因素,14 个组织风险因素,共计 40 个民机运行风险因素特征。利用上述 40 个风险因素对 241 个案例进行分析,形成 241 个案例样本、40 个风险因素的特征矩阵,将这一矩阵称为重大民机事故的多属性技术事故(major civil aircraft accidents-multi-attribute technological accident, MCAA-MATA)数据集,记为 $\boldsymbol{M}$,其表达式为

$$\boldsymbol{M}=\begin{bmatrix} x_{1,1} & \cdots & x_{1,40} \\ \vdots & \ddots & \vdots \\ x_{241,1} & \cdots & x_{241,40} \end{bmatrix} \tag{9}$$

该数据集能清晰的描述所有收集到的 MCAA 案例及其风险因素特征,为后续的可视化数据挖掘提供了基础。

2.2 基于 SOM 和关联规则的民机运行风险研究算法流程

基于 SOM 和关联规则的民机运行风险研究算法流程如图 2 所示。其算法流程如下:

1)结合 MCAA-CREAM 模型将收集到的 MCAA 非结构化数据转化为结构化数据,形成 MCAA-MATA 数据集 $\boldsymbol{M}$。

2)根据 1.2 节中的 SOM 模型数据挖掘流程,将数据集作为样本输入,经过多次迭代计算最终得到民机运行事故聚类结果。

3)针对得到的每一聚类结果,根据 1.3 节中计算各聚类类别中风险要素的关联规则,通过计算候选项集、频繁项集、加权支持度、加权置信度以及加权提升度得到有效的强关联规则。

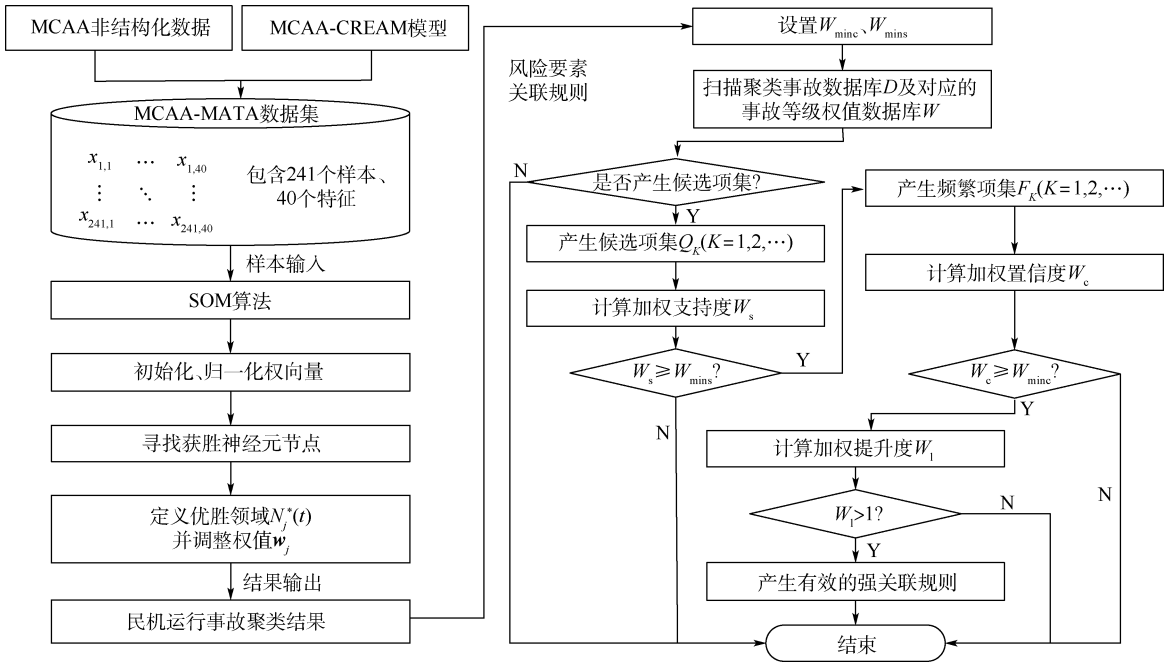


图 2 基于 SOM 和关联规则的民机运行风险研究算法流程

Fig. 2 Operational risk algorithm of civil aircraft based on SOM and AR

3 实例分析

数据主要来源于:美国联邦航空管理局(FAA)、美国航空安全网官方网站(ASN)、航空安全报告体系(ASRS)。根据国际民航组织(ICAO)对事故的定义,选取的 MCAA 均为民机重大事故及以上的案例,发生时间为 1972 年 12 月—2020 年 1 月,共计 241 个 MCAA 案例。

3.1 聚类结果

民机运行风险的 SOM 模型是通过相似性将 MCAA 进行聚类,事故越相似,在输出空间中的位置就越接近。利用 Viscovery SOMine® 软件,为了保证算法的收敛性,通过多次训练,当高斯领域半径设为 0.5,学习率为 0.5 对样本进行训练时,学习率和优胜邻域值都随着迭代次数增加而不断减小,算法逐渐收敛,映射结果良好。可将 241 个 MCAA 事故案例分为 3 类(C1,C2,C3),以可视化的图形形式增强风险因素的可读性。如图 3 所示,显示了输入空间[数据集 $\boldsymbol{M}(241\times40)$]在输出空间 2D 地图中的可视化表示。

为验证聚类的可靠性,Viscovery SOMine® 软



图 3 2D 聚类地图

Fig. 3 2D cluster map

件提供了聚类质量指标直方图,通过为每种可能的聚类安排一个索引来对可能的分组进行分类。如图 4 所示,当群集聚类结果为 3 时,其质量指标最高为 73,表示其解释更具有可靠性,因此选取 3 个聚类对民机运行风险因素进行聚类分析。

在此聚类条件下,数据频率和量化误差会均匀分布在整个地图上,表明网络训练映射效果良好。对 MCAA 的聚类基本信息进一步分析,可得表 1 所示的聚类信息结果。

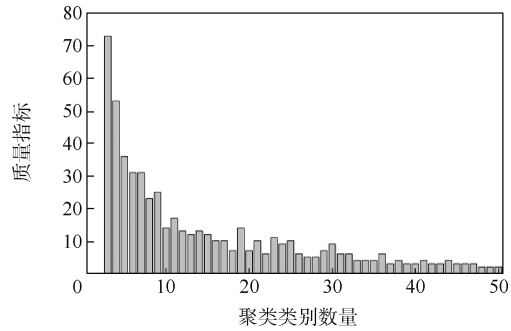


图 4　聚类质量指标

Fig. 4　Cluster quality indicator

表 1　聚类特征

Table 1　Clusters features

聚类族编号	包含案例数/个	频率/%	风险因素				
			最小值/个	最大值/个	平均值/个	中位数	众数
C1	111	46.06	1	12	5.16	5	4
C2	93	38.59	1	12	3.67	3	2
C3	37	15.35	3	15	8.5	8	7,8

类别 C1 包含 111 个 MCAA,占比最大,为 46.06%。事故风险因素在 1 ~ 12 个之间波动,平均风险因素为 5.16 个,中位数为 5,在该类别中多数案例包含风险因素为 4 个。

类别 C2 包含 93 个 MCAA,占比 38.59%。风险因素在 1 ~ 12 个之间波动,与类别 C1 不同的是,该类别中事故平均风险因素为 3.67 个,中位数为 3,多数事故包含风险因素为 2 个,表明该类别中的事故风险因素较为分散。

类别 C3 包含 37 个 MCAA,占比 15.35%,是最小的类别。事故风险因素在 3 ~ 15 个之间波动,案例平均风险因素为 8.5 个,中位数为 8,在该类别中多数事故包含风险因素为 7 个与 8 个,是 3 个类别中包含风险因素最多也是最集中的类别。

3.2　聚类解释

图 3 显示了具有相似特征的 MCAA 在地图中的分布结果,表 2 为数据集 $\boldsymbol{M}$ 中各个风险特征聚类的统计结果。经过分析,3 个类别具有鲜明的风险水平特征,主要体现在风险致因因素的明显区别上,这 3 种风险致因因素对事故的作用机制、致因规律与表现形式上均有所不同:

1) 在类别 C1 中,主要以组织因素为主,其中由于不利天气/地形(60.40%)引起的 MCAA 占主导因素。此外,在 C1 类中,还严重受到运营人管理问题(35.10%)及飞行员技能/知识不足(32.40%)的影响。从技术角度看,设备故障(18.90%)是主要原因;人为因素中飞行员决策错误(29.70%)、分心/粗心(28.80%)也是重要的风险原因。

2) 类别 C2 中,技术因素中的设备故障(40.09%)及设计缺陷(39.80%)是该类别的首要风险因素。管理问题(38.70%)及质量控制不足(35.50%)并列为第二大风险影响因素;该分类主要考量技术原因包括鸟击/军机击毁/劫机等意外情况(31.20%)。

3) 类别 C3 中最突出的是人为组织原因,如管理问题(73.00%)、团队支持不足(67.60%)、任务分配不足(62.20%)。此外,人为因素在该类别中也占有很大影响,如认知方式(56.80%)、没有按手册执行(48.60%)、心理/精神压力(43.20%)。

3.3　主要风险要素分析

对 MCAA 进行重新组织以地图形式呈现时,主要以事故案例中的风险因素相似性进行聚类,可以识别民机运行风险并显示数据集中事故特征的重要联系。将影响民机运行的风险因素划分为 40 个因素(具体因素见 MCAA-CREAM 模型),由于篇幅限制无法一一呈现。现展示 241 个 MCAA 案例中影响比较大的 8 个风险因素地图(见图 5),包括管理问题、不利天气/地形、没有按手册执行、设计缺陷、分心/粗心、决策错误、技能/知识不足、设备故障,其他风险因素以文字进行说明。图 5 中:深灰色表示不存在该特征,灰色表示存在该特征,浅灰色是多个类别交集的边界。风险因素地图在于通过事故表征和基本描述,挖掘事故致因与事故之间的深层次关联,在此基础上

表2 风险特征聚类统计结果

Table 2 Characteristic clustering statistical results of each risk factor

风险类别	风险子类别	C1 频率/%	C2 频率/%	C3 频率/%
设备	设备故障	18.90	40.09	27.00
	设计错误	0	12.90	0
	设计缺陷	15.30	39.80	21.60
程序	设计复杂	0.90	12.90	2.70
突发状况	信息模糊	14.40	6.50	8.10
	信息不完整	5.40	4.30	0
永久状况	信息错误	8.10	5.40	0
行为	执行时间错误	25.20	3.20	16.20
	用错执行设备	19.80	3.20	24.30
	执行顺序错误	17.10	2.20	10.80
	执行目标错误	9.90	0	8.10
	没有按手册执行	27.00	4.30	48.60
观察	没看到	11.70	2.20	16.20
	看错	4.50	1.10	5.40
计划	计划不充分	10.80	1.10	21.60
	优先级错误	10.80	1.10	21.60
解释	延迟措施	15.30	2.20	18.90
	错误预测	17.10	2.20	13.50
	决策错误	29.70	11.80	35.10
突发状况人的反应	记忆失效	5.40	0	10.80
	分心/粗心	28.80	1.10	40.50
	心理/精神压力	10.80	0	43.20
	疲劳	11.70	3.20	27.00
永久与机组相关的功能	身体疾病	0	0	16.20
	认知方式	11.70	8.60	56.80
	技能/知识不足	32.40	9.70	40.50
交流	交流障碍	5.40	1.10	18.90
	信息传递错误	24.30	5.40	2.80
组织	维修失效	1.80	19.40	8.10
	质量控制不足	9.90	35.50	10.80
	管理问题	35.10	38.70	73.00
	任务分配不足	0.90	0	62.20
	设计失效	2.70	21.50	13.50
培训	培训不足	3.60	5.40	18.90
周围环境条件	不利天气/地形	60.40	7.50	37.80
	夜间	6.30	6.50	10.80
	货物摆放/没油	0.90	9.70	0
	劫机/鸟击/军机击毁等意外	0	31.20	0
工作条件	工作场所布局不合理	0.90	4.30	2.70
	团队支持不足	7.20	2.20	67.60

为识别潜在风险源,控制事故提供依据。通过对风险因素地图分析,可将实际事故与图像直接关联起来,从而快速传递民机运行风险,增强事故信息可读性、全面性和系统性。

由图5(a)可以看出,组织因素中管理问题,在每一类中都占有很大的比例,因此不是事故聚类的依据。

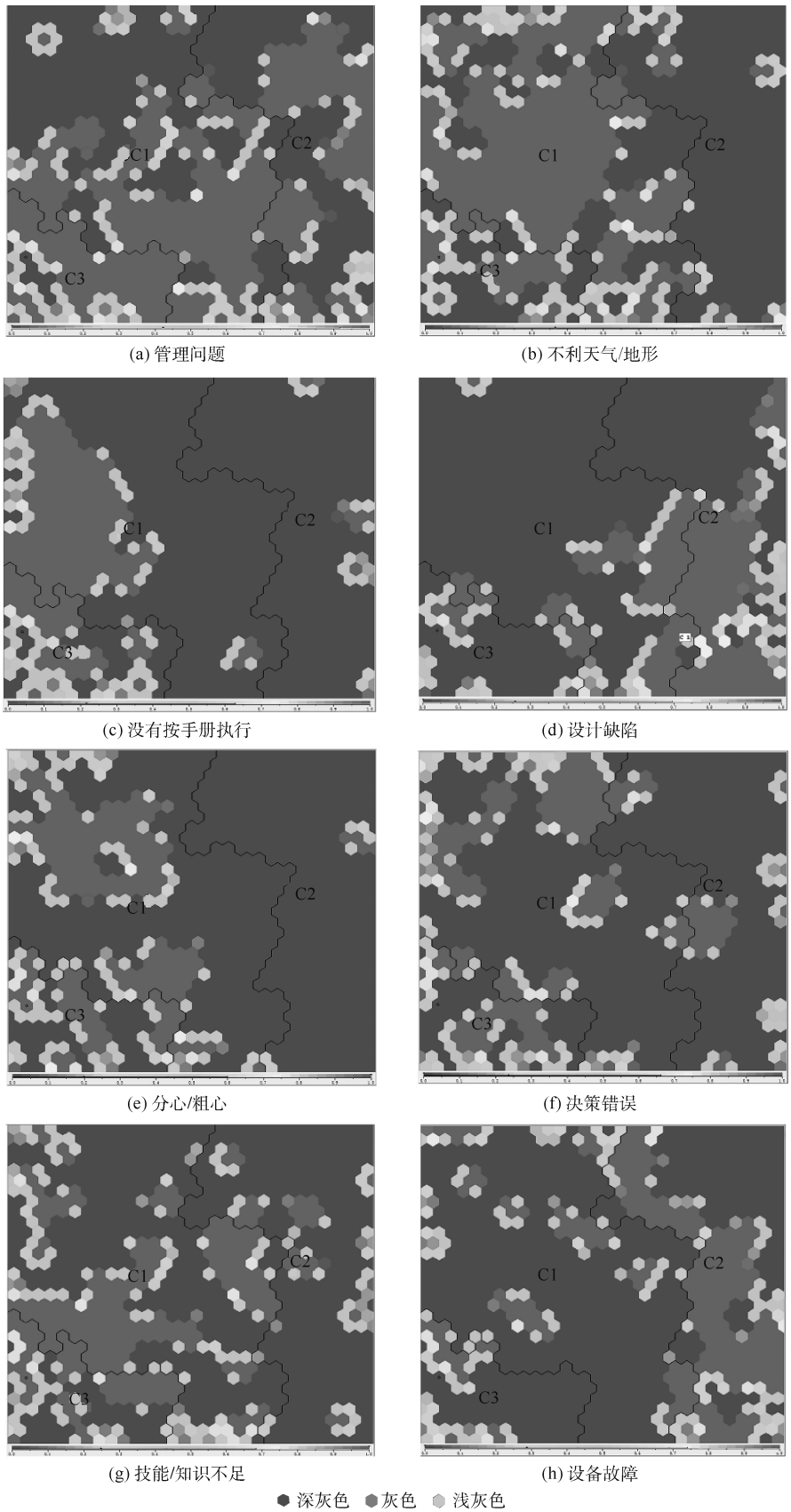
1个组织因素(不利天气/地形)及3个人为因素(分心/粗心、没有按手册执行、技能/知识不足)占据了C1的大部分区域,意味着这些是导致分组C1聚类的主要因素。此外,传递信息错误、错误预测、用错方式覆盖了C1面积的1/3左右;人为因素中错误预测、没看到、执行时间错误、执行顺序错误、信息传递错误,及技术因素中设备信息模糊、设备信息不完整、设备信息显示错误都主要分布在C1,可见技术设备的缺陷与人为错误的联系十分紧密。大量的人为错误划分到这一类,表明有效的培训策略不应只是书面程序的形式,还应适当培养在紧急情况下需要的决策技能和重点关注对象,采用适当的操作流程等。

技术因素中设备故障、设计缺陷及组织因素中质量控制不足占据了C2的大部分区域,可以看出设备故障往往与设计缺陷/组织能力有关,且设备问题常伴随质量控制问题和设计缺陷。维修失效虽然完全被划分在C2,但可以通过图形看出,不能将加强维修能力看作是减少设备故障可能性的唯一解决方案。此外,设备设计复杂、设备设计错误主要分布在C2,可见设备的风险管控与技术因素的关联性较高。

组织因素中任务分配不足、团队支持不足及人为因素中认知方式是C3聚类的主要事故原因,占据大部分面积。人为因素中的延迟措施、用错执行设备、没看到、认知方式、计划不充分、优先级错误、看错、记忆失效、疲劳、心理/精神压力、交流障碍、身体疾病、技能/知识不足、决策错误主要分布在C3区域。这表明人的组织管理对民机运行安全有着重要的影响。

3.4 关联规则分析

根据2.2节中关联规则的算法流程,经过多次试验,将MCAA中各事故类别最小支持度设为0.10,最小置信度设为0.60,利用关联规则对各聚类群进行分析,得到关联规则集。图6中数字1/数字2代表该关联规则的支持度/置信度。由图6可知,在事故聚类群1中不利天气/地形条件下最容易暴露民机运行中存在的安全问题,通过关联规则有效挖掘出了不利天气/地形情况下关联的7个强致因,其中人为因素是造成C1类MCAA发生的主要因素,包括:执行时间错误、设备信息模糊、传递信息错误、分心/粗心、技能/知识不足及管理问题;同时有效挖掘出了其他3个强关联规则:执行时间错误导致延迟措施、设备信息模糊导致信息传递错误、分心/粗心造成的执行顺序错误,技能/知识不足往往是由于培训不足导致。



(a) 管理问题　(b) 不利天气/地形

(c) 没有按手册执行　(d) 设计缺陷

(e) 分心/粗心　(f) 决策错误

(g) 技能/知识不足　(h) 设备故障

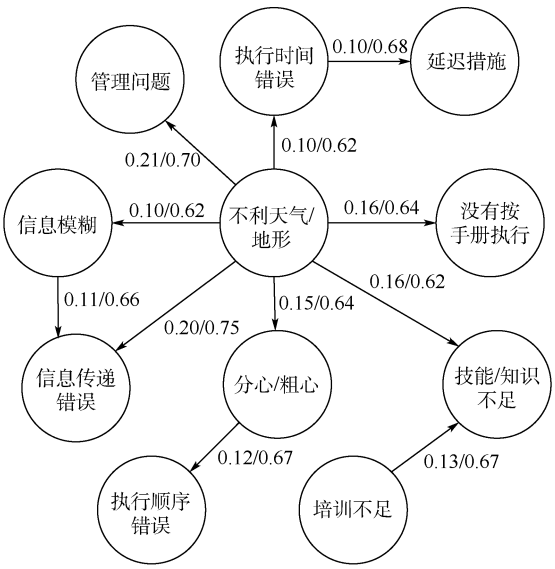
图 5　MCAA 中 8 个风险因素可视化地图

Fig. 5　Visual maps of 8 risk factors in MCAA

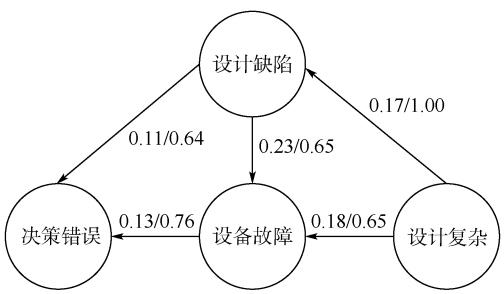
事故聚类群 C2 中主要项集为技术因素中的设备故障,挖掘出了 3 个与设备故障具有强关联规则的事故致因项:设计缺陷、设计复杂及决策错误。这证明在飞机设计过程中,飞机上的使用设备的缺陷会在运行过程中暴露出来,设计缺陷及设计于复杂会导致设备故障或致使机组人员做出

错误决策。

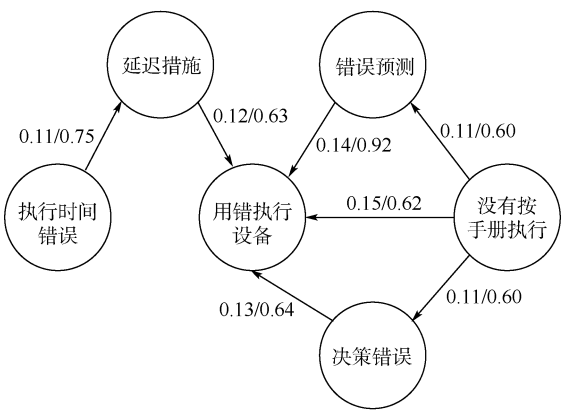
如图6(c)所示,事故聚类群C3中存在4条人为因素致因链:①执行时间错误—延迟措施—用错执行设备;②没有按手册执行—错误预测—用错执行设备;③没有按手册执行—用错执行设备;④没有按手册执行—决策错误—用错执行设备。以此有效挖掘出了用错执行设备的4个主要致因,并阐明没有按手册执行将会带来的不利影响,因此需加强对机组人员的管理与培训,实际操作时必须根据飞机机型结构按照手册严格执行。



(a) 事故聚类群C1强关联规则可视化



(b) 事故聚类群C2强关联规则可视化



(c) 事故聚类C3强关联规则可视化

图6 事故聚类群强关联规则可视化

Fig. 6 Visualization of AR in accident clustering group

4 结 论

结合SOM模型,将重大民机事故的多属性、复杂结构化高维数据转化为2D可视化地图,为民机运行风险信息的获取提供了新的手段。通过对重大民机事故的研究,可以得出以下结论:

1)利用关联规则在SOM模型可以从重大事故中进行学习,可以快速传递民机运行风险,增强事故信息可读性,充分理解事故现象挖掘出深层次的致因特征。

2)风险地图为研究者提供直观、清晰的视觉理解,在输出空间中完全保留输入数据,并可以检索数据集记录,帮助研究者将事故与图像进行联系。

3)提出的方法有助于理解重大事故原因之间的关系。如需加强飞行员在不利天气环境下的操作技能和培训;先进的维护手段也不能保证设备不出现故障,需进一步采取措施;人的组织管理对民机运行安全有着重要的影响。

4)加权关联规则更有助于挖掘风险致因间的演化规律,为民机的安全运行保障提供科学借鉴。

通过关联分析与可视化技术为解析事故发生的深层次特征进行了探索,随着数据提取挖掘及主动学习技术的深入,还可以通过集成、增强学习技术等,提升信息挖掘的深度和广度,为制定安全管理对策提供支持。

参考文献(References)

[1] 张凡,魏法杰,李权葆. 复杂装备研制项目的风险源识别[J]. 北京航空航天大学学报,2017,43(5):975-980.
ZHANG F,WEI F J,LI Q B. Risk source identification of complex equipment development project[J]. Journal of Beijing University of Aeronautics and Astronautics,2017,43(5):975-980 (in Chinese).

[2] IMAI S, BLASCH E, GALLI A, et al. Airplane flight safety using error-tolerant data stream processing[J]. IEEE Aerospace and Electronic Systems Magazine,2017,32(4):4-17.

[3] KELLY D, EFTHYMIOU M. An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017[J]. Journal of Safety Research,2019,69:155-165.

[4] PAN Y,ZHANG L M,WU X G,et al. Multi-classifier information fusion in risk analysis[J]. Information Fusion,2020,60:121-136.

[5] CHEN Z Y,HUANG X P,YU S,et al. Risk analysis for clustered check dams due to heavy rainfall[J]. International Journal of Sediment Research,2021,36(2):291-305.

[6] 董雷霆,周轩,赵福斌,等. 飞机结构数字孪生关键建模仿真技术[J]. 航空学报,2021,42(3):113-141.

DONG L T, ZHOU X, ZHAO F B, et al. Key technologies for modeling and simulation of airframe digital twin[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(3): 113-141 (in Chinese).

[7] GOERLANDT F, GOITEAB H, VALDEZ B O, et al. An analysis of wintertime navigational accidents in the Northern Baltic Sea[J]. Safety Science, 2017, 92: 66-84.

[8] MOURA R, BEER M, PATELLI E, et al. Learning from major accidents: Graphical representation and analysis of multi-attribute events to enhance risk communication[J]. Safety Science, 2017, 99: 58-70.

[9] 余冠华. 基于多属性铁路事故数据集的聚类和关联规则分析方法研究[D]. 北京:北京交通大学, 2019: 15-19.
YU G H. Research on clustering and association rule analysis method based on multi-attribute railway accident dataset[D]. Beijing: Beijing Jiaotong University, 2019: 15-19 (in Chinese).

[10] ZARANEZHAD A, MAHABADI H A, DEHGHANI R M. Development of prediction models for repair and maintenance-related accidents at oil refineries using artificial neural network, fuzzy system, genetic algorithm, and ant colony optimization algorithm[J]. Process Safety and Environmental Protection, 2019, 131: 331-348.

[11] ASGARY A, ANSARI S, DUNCAN R, et al. Mapping potential airplane hazards and risks using airline traffic data[J]. International Journal of Disaster Risk Reduction, 2015, 13: 276-280.

[12] 李哲, 徐浩军, 薛源, 等. 基于风险预测的飞行安全操纵空间构建方法[J]. 北京航空航天大学学报, 2018, 44(9): 1839-1849.
LI Z, XU H J, XUE Y, et al. Construction method of flight safety manipulation space based on risk prediction[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(9): 1839-1849 (in Chinese).

[13] ZHANG Z H, HE Q, GAO J, et al. A deep learning approach for detecting traffic accidents from social media data[J]. Transportation Research Part C, 2018, 86: 580-596.

[14] HOLLNAGEL E. Cognitive reliability and error analysis method (CREAM)[M]. Oxford: Elsevier Science Ltd., 1998: 151-190.

[15] HU R J, RATNER K, RATNER E, et al. ELM-SOM+: A continuous mapping for visualization[J]. Neurocomputing, 2019, 365: 147-156.

[16] AVSAR B, ALIABADI D E. Parallelized neural network system for solving euclidean traveling salesman problem[J]. Applied Soft Computing, 2015, 34: 862-873.

[17] MERLIN P, SORJAMAA A, MAILLET B, et al. X-SOM and L-SOM: A double classification approach for missing value imputation[J]. Neurocomputing, 2009, 73(7): 1103-1108.

[18] COMBERTII L, DEMICHELA M, BALDISSONE G. A combined approach for the analysis of large occupational accident databases to support accident-prevention decision making[J]. Safety Science, 2018, 106: 191-202.

[19] 安璐. 基于自组织映射的期刊主题研究[D]. 武汉:武汉大学, 2009: 26-29.
AN L. Studying journal subjects with self-organizing map[D]. Wuhan: Wuhan University, 2009: 26-29 (in Chinese).

[20] ZHOU Y, LI C S, DING L Y, et al. Combining association rules mining with complex networks to monitor coupled risks[J]. Reliability Engineering and System Safety, 2019, 186: 194-208.

[21] TEW C, GIRAUD C C, TANNER K, et al. Behavior-based clustering and analysis of interestingness measures for association rule mining[J]. Data Mining and Knowledge Discovery, 2014, 28(4): 1004-1045.

Operation risk of civil aircraft based on SOM and association rules

XIONG Minglan, WANG Huawei*, NI Xiaomei, LIN Ruiguan

(School of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: To fully understand the risks of civil aircraft and learning from accidents, the major civil aircraft accidents (MCAA) are used as the research object to dig out its deep-level causal characteristics. Due to the poor readability information and nonlinear system behavior of MCAA, it is difficult to directly obtain the operational risk information or establish association with mapping relationship of accident factors, a method of learning the operational risk characteristics of civil aircraft from major accidents is proposed. According to the operation characteristics of civil aircraft, and drawing on MCAA information and cognitive reliability, and error analysis method (CREAM), the MCAA-CREAM model is designed. Furthermore, the civil aircraft multi-attribute technology major accident dataset was constructed. To complete the cluster analysis and abstract feature mapping, we take the dataset as a sample, input it into the self-organizing maps (SOM) model, and enhance the readability of risk factors in the form of a 2D map. The strong association between risk factors can be mined by association rules.

Keywords: civil aircraft safety; risk characteristics; accident analysis; self-organizing maps model; association rules

Received: 2021-03-02; **Accepted**: 2021-07-16; **Published online**: 2021-07-29 16:19

URL: kns.cnki.net/kcms/detail/11.2625.V.20210729.1409.001.html

Foundation item: National Natural Science Foundation of China (U1833110)

* **Corresponding author**. E-mail: wang_hw66@163.com

http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2021.0108

基于优化 CenterNet 的低空无人机检测方法

张瑞鑫[1], 黎宁[1,2,*], 张夏夏[1], ZHOU Huiyu[3]

(1. 南京航空航天大学 电子信息与工程学院, 南京 211106;
2. 光电控制技术重点实验室, 洛阳 471000; 3. 莱斯特大学, 莱斯特 LE1 7RH)

摘 要: 为实现对“低慢小”无人机(UAV)的有效探测,提升检测精度和定位质量,提出一种基于联合注意力和 CenterNet 的低空无人机检测方法。针对通用目标检测算法小目标漏检率高的问题,引入解耦的非局部算子,捕捉光学图像目标区域的关联性。利用无人机群个体间的相似性,将离散的无人机特征相互关联,降低漏检率。为获得更加精准的检测框,对 CenterNet 的标签编码策略和边界框回归方式进行优化,引入定位质量损失,提升检测框定位质量。实验结果表明:优化后的 S-CenterNet 算法相比原始 CenterNet 算法平均准确率提升了 8.9%,检测框定位质量有明显改善。

关 键 词: 目标检测; 深度学习; 联合注意力; CenterNet; 无人机
中图分类号: TP391
文献标志码: A **文章编号**: 1001-5965(2022)11-2335-10

随着无人机技术的快速发展,低空小型无人机(unmanned aerial vehicle, UAV)在航拍、农业、军事等领域得到广泛应用,但同时也为空中航路安全、公共安全及军事安全带来威胁。民用领域,近年来消费型无人机在中国快速普及,全国多个机场出现了民用无人机扰航事件[1],无人机的合规合法化使用成为了新的焦点问题。军用领域,利用蜂群无人机进行战场侦察、信息干扰甚至发动攻击逐渐成为了一种有效的战争手段。因此,作为无人机反制与管控的重要环节,对“低慢小”无人机的有效探测成为亟待解决的问题。

目前对无人机的探测手段按信号形式主要分为 3 类:雷达探测、无线电探测、光电探测。雷达探测作用距离远,几乎不受天气影响,但无法识别悬停的无人机,无法区分慢速移动的无人机与树木等其他物体;无线电探测通过对无人机与操控者的通信信号进行分析,来对无人机进行识别,这种方式可以实现对无人机详细特征的精准识别,但无法识别频谱特征不存在于频谱库或处于电磁静默状态的无人机;光电探测使用可见光和红外探测器对无人机进行识别与跟踪,精度高且成本较低,相比其他 2 种具有更广阔的应用场景和更高的实用性。

随着卷积神经网络在图像处理领域的广泛应用,目标检测算法已经从传统的人工设计特征算子转变为利用卷积神经网络自动提取特征。目前基于深度学习的目标检测算法主要分为一阶段和二阶段网络。二阶段网络首先产生候选区域,包含目标大概的位置信息,然后对候选区域进行分类和目标位置回归。Girshick 等[2]将深度学习技术应用于目标检测领域,提出基于区域提取的 RCNN 算法,利用选择性搜索算法生成1 000 ~ 2 000个候选区域,并将这些区域送入卷积神经网络中进行特征提取,利用支持向量机进行分类并

收稿日期: 2021-03-05; **录用日期**: 2021-05-05; **网络出版时间**: 2021-06-02 08:50
网络出版地址: kns.cnki.net/kcms/detail/11.2625.V.20210601.1134.002.html
基金项目: 航空科学基金 (ASFC-20175152036); 人工智能重点项目 (1004-56XZA19008)
* **通信作者**. E-mail: lnee@nuaa.edu.cn

引用格式: 张瑞鑫, 黎宁, 张夏夏, 等. 基于优化 CenterNet 的低空无人机检测方法[J]. 北京航空航天大学学报, 2022, 48(11): 2335-2344. ZHANG R X, LI N, ZHANG X X, et al. Low-altitude UAV detection method based on optimized CenterNet[J]. Journal of Beijing University of Aeronautics and Astronautics, 2022, 48(11): 2335-2344(in Chinese).

进行位置精修。RCNN 相比传统算法精度大大提升,但候选区域生成阶段包含过多冗余计算,运行速度较低。为提高二阶段网络的速度和精度,学者们陆续提出了一系列 RCNN 的变体:Fast RCNN[3]、Faster RCNN[4]、Cascade RCNN[5]、Mask RCNN[6]等。虽然二阶段网络在精度上有优异的表现,但其缓慢的速度限制了在实时性任务中的应用。一阶段网络直接生成目标的类别概率和位置坐标,准确度一般低于二阶段网络,但在速度上有较大优势,代表算法有 YOLOV3[7]、RetinaNet[8]、CornerNet[9]等。

无人机检测任务的首要难点为机体在光学图像中所占像素尺寸较小。一般认为所占像素尺寸小于整张图像 1% 的目标为小目标,目标检测通用数据集 COCO 将小目标定义为像素尺寸小于 32 × 32 的目标。小型无人机成像后所占像素往往更小,在像素尺寸为 1 960 × 1 080 的光学图像数据集中,大部分无人机所占像素尺寸在 15 × 15 ~ 35 × 35 之间,包含特征信息较少,因此,针对低空无人机小目标检测任务需要设计专用算法。现有的目标检测算法在特征提取阶段普遍包含过多下采样,小目标的特征信息随着下采样的次数增多逐渐丢失,检测器在小目标检测任务中普遍存在漏检率高、定位质量低的问题。故而本文对目标检测算法的特征提取模块进行重新设计,采用先下采样再上采样的沙漏结构,获得深层语义特征的同时维持较大的特征图分辨率,并使用旁路连接融合小目标及定位信息清晰的浅层特征与深层语义特征。此外,现有的目标检测算法主要依据图像局部范围内的特征信息判定该区域是否存在目标,孤立于其他局部区域的检测。而蜂群无人机个体之间相似度较高,可以利用特征的相似性对群体中的个体进行特征增强。因此采用改进的非局部算子捕捉空间长距离特征之间的关联性,增强无人机特征信息,减少漏检情况的发生。最后引入定位质量监督,优化标签编码策略和边界框回归方式以提升定位质量,实现了对低空无人机的有效检测和精准定位,为后续无人机跟踪、轨迹预测任务奠定基础。

1 CenterNet 目标检测算法

CenterNet[10]是一种端到端可微的无锚框(anchor free)单阶段目标检测算法。大多数成功的目标检测器列举所有潜在目标的位置,并对这些区域进行交并比(intersection over union, IOU)计算、分类,这样的做法是低效的,并且需要非极大值抑制(non maximum suppression, NMS)等额外的后处理。CenterNet 将目标检测问题转换为关键点估计问题,输入图片经过全卷积网络产生热图,热图的局部峰值对应物体中心,并对物体中心回归边界框的高度和宽度。相比基于锚框(anchor based)的算法更加简洁、快速和精确。

1.1 检测原理

输入图像为宽度 W 和高度 H 的 RGB 图像,经过编解码网络(encoder-decoder)输出到 3 个分支,分别为物体中心概率的关键点热图、物体尺寸、下采样取整带来的中心点偏移。在中心点与分类分支,对关键点热图执行简单的 3 × 3 最大池化获得局部峰值点,取代 NMS 后处理。用中心点偏移补偿目标中心后,结合物体尺寸即可获得检测框,网络结构如图 1 所示。

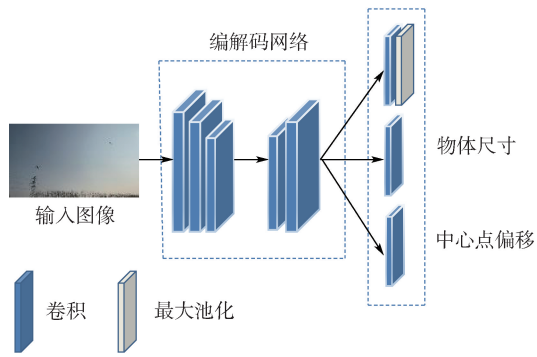


图 1 CenterNet 结构简图

Fig. 1 CenterNet structural sketch

1.2 标签编码

基于锚框的目标检测器使用标签的方式为在训练过程中将真值框(ground truth box, GT)与大量的锚框进行交并比计算,大于某一阈值则认为锚框内存在目标,这一方法包含大量冗余计算,减慢了算法运行速度。

CenterNet 将真值框编码为目标中心坐标和目标尺寸,并用高斯散射核将下采样后的目标中心投射到热图上:

$$Y_{xyc} = \exp\left[-\frac{(x-\tilde{p}_x)^2+(y-\tilde{p}_y)^2}{2\sigma_p^2}\right] \tag{1}$$

式中:σ_p 为尺寸自适应标准差,与目标面积大小相关;$Y_{xyc} \in [0,1]^{\frac{W}{r}\times\frac{H}{r}\times C}$ 为热图标签,xyc 分别为热图的横纵坐标及通道号,r 为下采样倍率;p 为目标中心,$\tilde{p} = p/r$ 为下采样后的目标中心;C 为类别数,用于在训练过程中与预测热图进行损失计算。

1.3 损失函数

CenterNet 的损失函数由三部分组成,中心点损失、中心点偏移损失、物体尺寸损失。中心点损

失选择连续标签的 Focal Loss[8]：

$$L_k = \frac{-1}{N}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\ln(\hat{Y}_{xyc}) & Y_{xyc}=1\\ (1-Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha}\ln(1-\hat{Y}_{xyc}) & \text{其他}\end{cases} \tag{2}$$

式中：N 为关键点个数，下标 k 为第 k 张图像，超参数 α 和 β 遵循文献[9]选用 2 和 4。物体尺寸和中心点偏移均采用均绝对误差(MAE)损失：

$$L_{\text{size}} = \frac{1}{N}\sum_{k=1}^{N}|\hat{S}-s| \tag{3}$$

式中：$\hat{S}$ 为预测的物体尺寸；s 为标签真实尺寸。

$$L_{\text{off}} = \frac{1}{N}\sum_{k=1}^{N}\left|\hat{O}-\left(\frac{p}{r}-\hat{P}\right)\right| \tag{4}$$

式中：$\hat{O}$ 为预测的中心点偏移；$p/r-\hat{P}$ 为实际中心点偏移；p 为标签真实中心点坐标；$\hat{P}$ 为预测的中心点坐标。

完整损失函数即为三者相加：

$$L_{\text{det}} = L_k + \lambda_{\text{size}}L_{\text{size}} + \lambda_{\text{off}}L_{\text{off}} \tag{5}$$

式中：λ_{size} 和 λ_{off} 为权重系数。

2　联合注意力的 CenterNet 优化

无人机反制系统中无人机探测的准确性、实时性至关重要。主流目标检测器针对无人机小目标的检测普遍存在漏检率偏高、检测框质量较差的问题，对后续的跟踪和预测有较大影响。为此，对 CenterNet 进行算法优化，引入联合注意力及检测框质量估计机制，以提高平均准确率。

2.1　非局部信息统计

蜂群无人机是近些年备受关注的新型作战概念之一，未来空域中无人机很可能以集群的方式出现。目前大多数目标检测器利用小卷积核提取局部信息，并通过不断堆叠卷积层和下采样层来扩大感受野判断某一区域内是否存在目标，浅层特征信息在下采样过程中不断损失，最终也无法获得表征 2 个目标之间关系的全局信息。因此引入一种非局部信息统计机制，以捕捉长距离特征之间的依赖关系，使得多个无人机小目标之间相互关联、相互激活，减少漏检现象。

Non-local 于 2018 年由 Wang 等[11]提出，其思想来源于数字图像滤波领域的非局部均值(non-local mean)操作及自然语言处理领域中的自注意力(self-attention)机制，最初用于行为识别领域，捕捉视频序列的长范围依赖。Non-local 的一般表达为

$$y_i = \frac{1}{C(\boldsymbol{x})}\sum_{\forall j}f(x_i,x_j)g(x_j) \tag{6}$$

$$f(x_i,x_j) = \sigma(\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{k}_j) \tag{7}$$

式中：$\boldsymbol{x}$ 为输入特征图矩阵；i 为输出位置，输出 y_i 与任意位置 j 均相关，使用函数 $f(x_i,x_j)$ 度量位置 i 和位置 j 的相似度；函数 g 为输入特征在位置 j 的响应函数；最后通过响应因子 $C(\boldsymbol{x})$ 进行标准化处理。式(7)中，$\boldsymbol{q}_i = W_{\boldsymbol{q}}x_i$，$\boldsymbol{k}_j = W_{\boldsymbol{k}}x_j$ 为自注意力机制 KQV 模型中的查询向量和键向量。

Non-local 算子度量特征图任意 2 个位置的关联程度，并利用 2 个位置之间的相似性对特征进行加权处理，自适应增强神经网络需要关注的位置，抑制无关信息。为了让 Non-local 操作适应神经网络结构，将整个 Non-local 操作转换为矩阵乘法运算和矩阵加法运算，如图 2 所示。

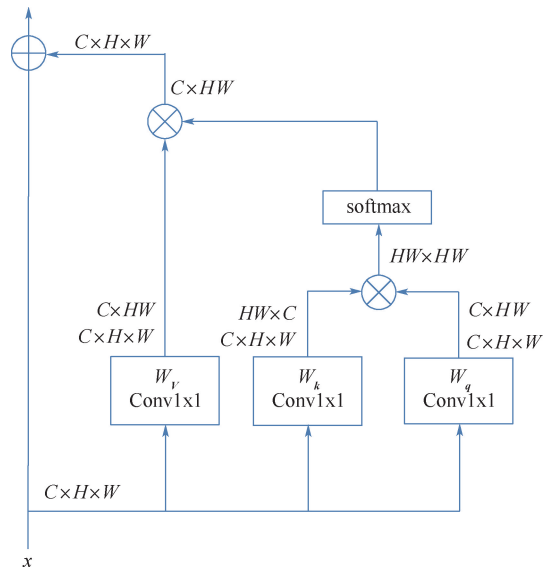


图 2　Non-local 结构

Fig. 2　Non-local structure

然而，文献[12]对 Non-local 提取的注意力图进行可视化，发现 Non-local 操作确实能够关注具有目标附近的区域，但对于不同查询位置的注意力图几乎是一致的。文献[12]认为 Non-local 操作学习到的注意力图是不依赖于位置的，换而言之，Non-local 操作不能表征任意两特征之间的关系，只能在空间域上关注某些重要特征。文献[13]对 Non-local 进行了进一步研究，认为 Non-local 将两项独立的任务联合学习，导致了文献[12]中的问题，文献[13]将 Non-local 公式中的相似度计算函数 $f(x_i,x_j)$ 解耦为

$$\sigma(\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{k}_j) = \sigma((\boldsymbol{q}_i-\boldsymbol{\mu}_{\boldsymbol{q}})^{\mathrm{T}}(\boldsymbol{k}_j-\boldsymbol{\mu}_{\boldsymbol{k}})+\boldsymbol{\mu}_{\boldsymbol{q}}^{\mathrm{T}}\boldsymbol{k}_j) \tag{8}$$

式中：$\boldsymbol{\mu}_{\boldsymbol{q}}$ 为 $\boldsymbol{q}$ 的均值；$\boldsymbol{\mu}_{\boldsymbol{k}}$ 为 $\boldsymbol{k}$ 的均值。

经过白化(whiten)的成对项解释了 2 个特征之间的关系，一元项代表了每个特征的显著性。经实验发现，成对项倾向于学习相同类别区域内关系，一元项学习显著边界，但当两者在 Non-local

中联合学习时,两者的作用相互影响,这可能对学习两类视觉线索有害。基于此项研究,将 Non-local 拆分为 PNL(pairwise NL)和 UNL(unary NL)。图 3 为 PNL 结构。为了保持 PNL 学习到的区域之间关系,将 PNL 放置于浅层支路连接,与编解码网络输出的特征进行通道拼接。

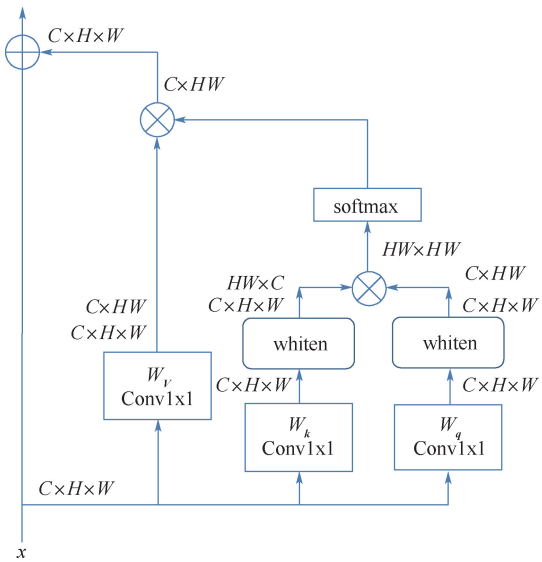


图 3 PNL 结构

Fig. 3 PNL structure

然而,Non-local 仅关注空间域特征间关系,而不同特征通道内的信息对最终目标正确检测的重要程度也有所不同。所以,在 UNL 模块的基础中添加通道注意力机制,组成 CNL(channel NL),同时捕获空间与通道的显著性,突出有效信息,抑制冗余信息,如图 4 所示。大多数通道注意力在计算各通道重要性时采用先降维再升维的方式来降低计算量,并捕获所有通道间的关系。文献[14]的实验表明降维对通道注意预测带来副作用,捕获所有通道之间的依赖也是低效且不必要的。因此在 CNL 中使用不降维的局部跨通道交互策略,使用一个自适应的 1(1 卷积代替进行降维的 2 个全连接层,卷积核大小 k 代表参与局部跨通道交互的通道数,k 与通道维数成正比,通过这样的方式降低计算量。

2.2 网络结构设计

基于 ResNet、DLA、Hourglass,CetnerNet 被设计了 4 种编解码网络[15-17]。输入图像经过下采样获得原图的 1/32 或者 1/128 大小的特征图后,经转置卷积上采样到原图的 1/4 大小的输出特征图。经初步实验表明,DLA 获得最佳精度和速度权衡,因此选择 DLA 为基准编解码网络。考虑到过多的下采样对浅层定位信息的破坏性及浅层特征对小目标检测的重要性,减少下采样模块,移除

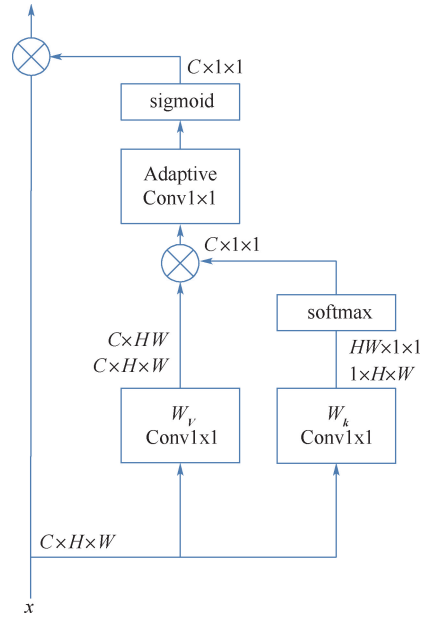


图 4 CNL 结构

Fig. 4 CNL structure

32 倍下采样模块,特征图大小到达原图的 1/16 时进行上采样。Non-local 文章中建议在浅层使用 Non-local 模块,因为深层特征在不断的下采样过程中逐渐损失空间信息及特征点之间的关联性,所以在 DLA 中 2 倍下采样处额外添加浅层分支,并使用 PNL 模块捕捉特征图空间关联性,与输出特征图聚合,最后在编解码网络中使用 CNL 模块捕获通道间依赖关系,突出重要信息,抑制冗余信息。综上,设计 DLAX28 作为最终的编解码网络。

编解码网络输出的特征图送入目标中心点分支和目标尺寸分支。相比 CenterNet 移除了中心点偏移分支,目标尺寸分支改为回归中心点到 4 条边的距离,让中心点偏移包含在尺寸回归中,计算复杂度不变但并行度提高。再者,边界框中心点可能落在背景上,可能导致较大中心点回归误差,回归中心点到 4 条边距离的边界框编码方式更加灵活,适应复杂情况的能力更强。原始 CenterNet 和 S-CenterNet-DLAX28 网络整体架构如图 5所示。

为了提升计算资源受限平台下的性能表现,同时设计轻量化版本 DLAS28。用深度可分离卷积代替普通的卷积操作,降低参数量和浮点运算次数。受 CSPNet[18] 启发,将 2 倍下采样浅层分支处的通道进行分流,一半的通道信息通过完整的编解码网络进行特征提取,一半的通道信息直接通过分支进行全局信息提取,并与输出特征图进行通道拼接。CSPNet 认为这样可以不同的累积层重复学习复制的梯度信息,并减少计算量。经实验发现,通道分流再聚合在获得可忍受的精度

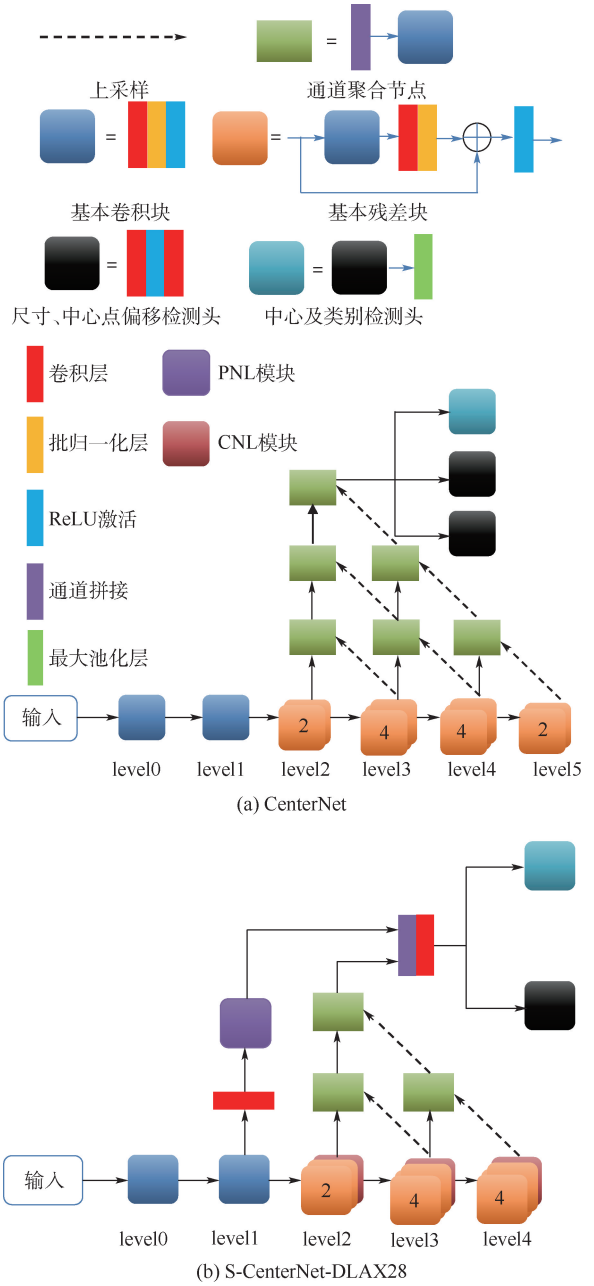


图 5　CenterNet 与 S-CenterNet-DLAX28 结构

Fig. 5　CenterNet and S-CenterNet-DLAX28 structure

损失的同时，推理时间减少约 30%。

2.3　标签编码策略优化

CenterNet 与 CornerNet 都使用圆形高斯散射核将真值框目标中心投射到标签热图上，高斯散射核如图 6 所示。但物体的长和宽往往是不一致的，对于长宽比较大的物体，预测中心点在 x 轴方向和 y 轴方向偏移同样的距离应当获得不同的误差。因此使用椭圆高斯散射核进行标签编码，以获得更精准的目标中心点定位：

$$Y_{xyc}=\exp\left(-\frac{(x-\tilde{p}_x)^2}{2\sigma_x^2}-\frac{(y-\tilde{p}_y)^2}{2\sigma_y^2}\right) \tag{9}$$

式中：$\sigma_x=\frac{2\sqrt{2}}{3}\left(1-\sqrt{\frac{2\text{min_iou}}{1+\text{min_iou}}}\right)w$，$w$ 为物体的宽；$\sigma_y=\frac{2\sqrt{2}}{3}\left(1-\sqrt{\frac{2\text{min_iou}}{1+\text{min_iou}}}\right)h$，$h$ 为物体的高，min_iou 为超参数，用于限制圆高斯散射核投射区域的面积，实验中取 0.7。

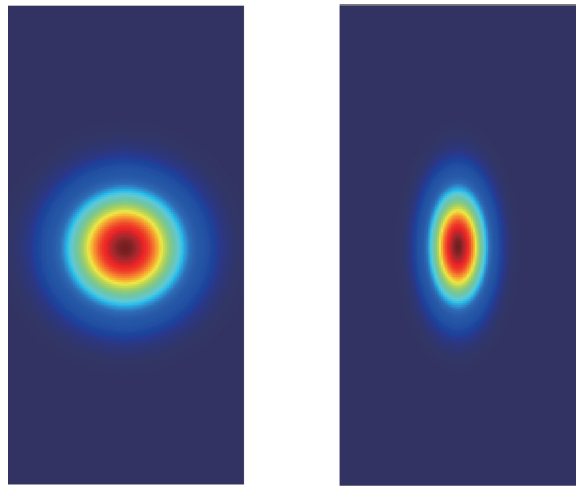


图 6　高斯散射核示意图

Fig. 6　Schematic diagram of Gaussian scattering kernel

2.4　损失函数设计

S-CenterNet 的损失函数由 3 项组成：中心点损失、物体尺寸损失、定位质量损失。

$$L_{\text{det}}=L_k+\lambda_{\text{size}}L_{\text{size}}+\lambda_{\text{MIOU}}L_{\text{MIOU}} \tag{10}$$

CenterNet 只将 $Y_{xyc}=1$ 的点作为正样本点，其余所有点视为负样本点，造成严重的正负样本不平衡问题，导致训练过程的损失由负样本的损失主导。因此，设置比例因子，增强正样本的损失，削弱负样本的损失。此外，$Y_{xyc}=1$ 附近的高质量样本同样能对训练起到监督作用，所以，将 $Y_{xyc}\geqslant 0.95$ 的样本视为正样本点。中心点损失：

$$L_k=\frac{-1}{N}\begin{cases}\lambda(1-\hat{Y}_{xyc})^{\alpha}\ln(\hat{Y}_{xyc}) & Y_{xyc}\geqslant 0.95\\ (\hat{Y}_{xyc})^{\alpha}\ln(1-\hat{Y}_{xyc}) & \text{其他}\end{cases} \tag{11}$$

式中：λ 为超参数，实验中取 0.75；N 为正样本点数量。

物体尺寸损失沿用 MSE 损失，计算公式同式(3)，其中 $\hat{S}$ 为中心点到上下或两侧距离之和，仍然为物体的长或宽。

CenterNet 对物体的中心点及尺寸的训练是独立的，这是检测框定位质量不佳的重要原因。为了解决小目标定位质量普遍较差的问题，提出马氏交并比(MIOU)损失对定位质量进行度量，在计算损失的过程中将预测的中心点位置及目标尺寸解码为预测框，将预测框和真值框的重合程

度作为训练的监督项,使得 S-CenterNet 可以获得更加精准的检测框。

$$L_{\mathrm{MIOU}} = 1 - \left(\mathrm{IOU} - \frac{m_1 + m_2 + m_3 + m_4}{d^2}\right) \tag{12}$$

式中:$m_1 \sim m_4$ 分别为预测框和真值框 4 条边界中心点之间的曼哈顿距离;d 为能够同时包含预测框和真值框的最小闭包区域的对角线的欧式距离,具体计算方式如图 7 所示。

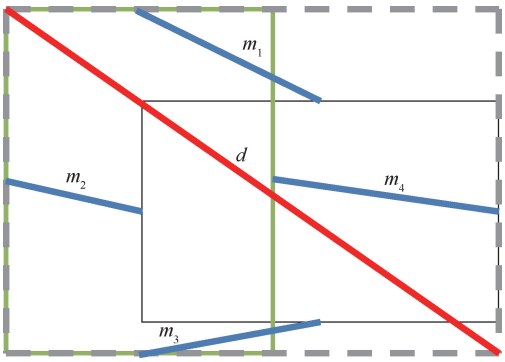


图 7 MIOU 计算方法

Fig. 7 MIOU calculation method

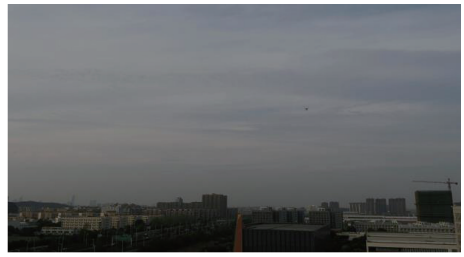
3 实 验

为验证 S-CenterNet 模型在低空无人机检测任务中的有效性,在相同运行环境及无人机数据集的前提下,将 S-CenterNet 和 CenterNet 进行对比实验。

3.1 数据集构建

由于目前没有公开的通用无人机数据集,首先进行数据集构建。为保证数据的多样性,数据采集在不同光照、不同天气的条件下进行,总共采用 4 种型号的无人机:大疆精灵 4、大疆 M200 V2、大疆 M600、自制无人机。拍摄距离由 20 ~ 3 000 m,近程使用无人机载光学摄像头,远程拍摄使用雷达扫描到可疑目标后引导光电设备对准无人机。

数据采集总共获得 6 682 张图像,经过翻转、旋转、裁剪、亮度调整扩充为 40 092 张图像。经人工标注后,按照 6:2:2 的比例划分训练集、验证集和测试集,标签中的目标尺寸经 kmeans 算法聚类后获得的聚类中心为(14,16)、(22,25)、(30,35)。图 8 为数据集部分图像。



(a) 单机



(b) 双机



(c) 多机

图 8 数据集部分图像

Fig. 8 Part of dataset image

3.2 训练细节

实验算法的运行环境为:深度学习框架 Pytorch1.2;编程语言 Python3.6;操作系统 Ubuntu 16.04;处理器 Intel(R) Xeon(R) Silver 4114;内存 128 GB;GPU 为 2 张 Nvidia GeForce RTX 2080Ti 12 GB;CUDA 10.0。

训练使用 Adam 优化器,初始学习率为 1.5×10^{-4},训练 100 轮,在第 60 轮和第 80 轮学习率衰减为原来的 0.1 倍。

3.3 实验结果与分析

3.3.1 评估标准

目标检测算法的一般评估标准为平均准确率(mAP),mAP 由查准率(Precision)和召回率(Recall)计算而来,查准率计算式为

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\% \tag{13}$$

式中:TP 为与真值框的交并比大于某一阈值的检测框数量;FP 为与真值框的交并比小于某一阈值的检测框数量,或检测到同一个真值框的多余检测框的数量。召回率计算式为

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\% \tag{14}$$

式中:FN 为没有被检测到的真值框的数量。以查准率为纵坐标,召回率为横坐标绘制 P-R(Precision-Recall)曲线,曲线下的面积即为某一类平均准确率 AP,mAP 为所有类别 AP 值的平均。AP50 表示 IOU 阈值为 0.5,AP75 表示 IOU 阈值为 0.75,COCO 中的 AP 一般表示 IOU 阈值为[0.5:0.05:0.95]共

10 个值下的平均结果,综合反映了检测准确率和定位质量。

实验还使用前向推理时间衡量图片经过网络的 GPU 耗时;使用网络参数量衡量模型的复杂度。

3.3.2　消融实验

首先对主干网络结构进行调整,移除 DLA34 的 32 倍下采样模块,并在 2 倍下采样处添加支路连接,支路下采样后与编解码网络输出特征图通道拼接。表 1 ~ 表 6 为消融实验结果,黑体代表该组实验中表现最好结果。表 1 为对网络结构调整前后性能对比,准确率略微上升,前向推理时间减少1 ms。表 2 展示了成对项非局部算子对无人机检测的效果提升。由于 PNL 计算复杂度较高,在支路增加一个 PNL 模块推理时间上升 4 ms,不考虑嵌入多个 PNL 模块。表 3 展示了在不同阶段集成 CNL 模块的结果。实验结果表明,在编解码网络中添加 CNL 模块能普遍提升平均准确率,综合考虑速度与准确率,最终选定在 L_2、L_3 和 L_4 层添加 CNL,并在支路添加一个 PNL。

表 4 对比了使用圆形高斯散射核与椭圆高斯散射核进行标签编码对检测性能的影响。实验表明,区分长和宽的标签编码方式可以更好地引导中心点分支的训练,能够提升中心点预测和边界框回归的准确率。

表 1　网络结构调整

Table 1　Network structure adjustment

网络	mAP/%	AP50/%	AP75/%	推理时间/ms
DLA34	41.9	89.8	31.8	11
DLA28	42.1	90.0	32.1	10

表 2　PNL 性能对比

Table 2　PNL performance comparison

PNL	mAP/%	AP50/%	AP75/%	推理时间/ms
×	42.1	90.0	32.1	10
√	44.8	93.2	34.7	14

注:“√”表示采用相关改进策略,“×”表示不采用相关改进策略。

表 3　CNL 性能对比

Table 3　CNL performance comparison

阶段	mAP/%	AP50/%	AP75/%	推理时间/ms
L_2	42.3	90.5	32.0	11
L_3	42.6	91.1	32.2	11
L_4	42.5	90.8	32.3	11
$L_2 + L_3 + L_4$	43.3	92.0	33.1	13

表 4　标签编码方式

Table 4　Label encoding method

高斯核	mAP/%	AP50/%	AP75/%	推理时间/ms
圆形	42.1	90.0	32.1	10
椭圆	42.5	90.1	32.8	10

表 5　定位质量损失和比例因子

Table 5　Positioning quality loss and scale factor

损失函数	mAP/%	AP50/%	AP75/%	推理时间/ms
Focal Loss	42.1	90.0	32.1	10
FL + MIOU	46.1	92.5	40.2	10
FL + λ	42.8	90.7	33.4	10
FL + MIOU + λ	46.6	92.7	41.1	10

表 6　消融实验结果

Table 6　Results of ablation experiments

网络结构	椭圆高斯核	损失函数	mAP(0.5:0.95)/%	AP50/%	AP75/%	mAR(0.9:0.95)/%	推理时间/ms	参数量/MB
×	×	×	41.9	89.8	31.8	51.8	11	74.0
√	×	×	45.6	92.2	38.2	55.3	17	30.0
×	√	×	42.5	90.1	32.8	52.4	11	74.0
×	×	√	46.6	92.7	41.1	54.9	11	74.0
√	√	×	46.3	93.7	39.0	55.7	17	30.0
×	√	√	47.3	93.4	43.7	57.1	11	74.0
√	×	√	50.2	94.9	47.2	59.1	17	30.0
√	√	√	50.5	95.5	47.8	59.5	17	30.0

注:“√”表示采用相关改进策略,“×”表示不采用相关改进策略。

对于损失函数,增加定位质量损失和比例因子的结果在表 5 中展示。定位质量损失引导预测框与真值框尽可能重合,比例因子缓解正负样本不平衡问题,避免训练被负样本损失主导,并且在预测时不增加额外的计算开销,大幅提升定位准确率。

最后,对各模块的改进进行消融实验。表 6 展示了包含引入浅层信息、加入解耦的非局部算子和通道注意力在内的网络结构设计、标签编码方式及损失函数的改进对算法性能提升的贡献。结果显示,在可容忍的时间开销增加下,算法平均准确率和平均召回率(mAR)均有较大提升。

3.3.3　算法性能评估

为验证优化算法的有效性,采用相同超参数,选用多个主干网络对 CenterNet 和 S-Centernet 进行评估,性能对比如图 9 所示。更加详细的实验结果如表 7 和表 8 所示,DCN(deformable ConvNets)表示在上采样过程中使用可变形卷积[19]。

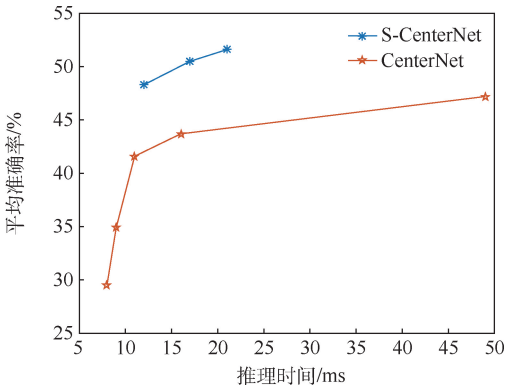


图 9 改进前后算法性能对比

Fig. 9 Algorithm performance comparison before and after improvement

表 7 算法准确率对比

Table 7 Algorithm accuracy comparison

算法框架	主干网络	mAP/%	AP50/%	AP75/%
CenterNet	ResNet-18	29.5	77.8	13.3
	ResNet-101	34.9	85.7	18.2
	DLA34	41.9	89.8	31.8
	DLA34-DCN	43.7	90.2	34.2
	Hourglass	47.2	93.5	42.3
S-CenterNet	DLAS28	48.3	94.1	44.0
	DLAX28	50.5	95.5	47.8
	DLAX28-DCN	51.6	95.3	50.1

表 8 算法速度及模型复杂度对比

Table 8 Algorithm speed and model complexity comparison

算法框架	主干网络	推理时间/ms	参数量/MB
CenterNet	ResNet-18	8	63.3
	ResNet-101	19	214.3
	DLA34	11	74.0
	DLA34-DCN	16	78.8
	Hourglass	49	765.7
S-CenterNet	DLAS28	12	8.3
	DLAX28	17	30.0
	DLAX28-DCN	21	32.1

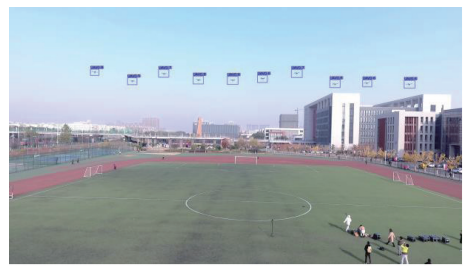
实验表明,使用 DLA34 和 Hourglass 的 CenterNet 网络在无人机数据集有不错的准确率表现,但是 Hourglass 网络参数量大、推理速度慢,不适合实时性要求强的低空无人机检测任务,使用 DLA34 和 DLA34-DCN 可以达到较好的准确率-速度权衡,AP50 达到 90% 左右,但 AP75 较低,表明存在回归框定位质量不高的问题。

S-CenterNet 相比 CenterNet 有普遍的性能提升,如表 7 所示。AP50 有 5%~6% 的提升,漏检率降低,检测出更多难样本。AP75 提升较大,普遍在 15% 以上,表明检测框定位质量大幅提升,为后续的稳定跟踪、预测等任务奠定良好的基础。

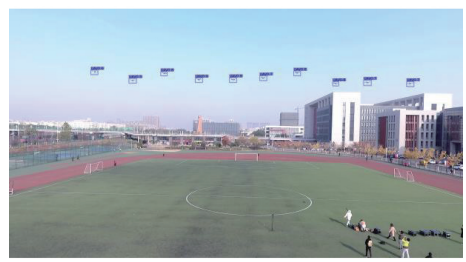
DLAX28 由 DLA34 改进而来,联合注意力机制与网络结构调整带来的额外时间开销使得推理时间增加了 6 ms,但 mAP 获得 8.9% 的提升。轻量版主干网络 DLAS28 采用类似 CSPNet 的通道分流,mAP 损失约 2%,但仍超过使用 Hourglass 的 CenterNet 网络,推理时间降为 12 ms,获得与 DLA34 相近的时间开销,但准确率优于 DLA34 约 7%。值得一提的是,通道分流后网络宽度变小,整体网络参数量仅 8.3 MB,是原始 DLA34 的 1/8 左右,适用于计算资源和存储资源受限的平台。

3.4 可视化结果

图 10 为算法改进前后不同场景的低空无人机检测结果对比。图 10(a)中所有无人机被检测出来,但检测框中包含过多背景,为后续的跟踪等



(a) CenterNet场景1



(b) S-CenterNet场景1



(c) CenterNet场景2



(d) S-CenterNet场景2

图 10 可视化结果

Fig. 10 Visualized results

任务带来更多的噪声;得益于 S-CenterNet 的定位质量损失及更好地利用边界信息的边界框编码机制,图 10(b)中的检测框更加准确、紧凑。在远距离光电设备拍摄的场景 2,图 10(c)漏检现象较为严重,图 10(d)则能够将所有无人机检测出来,漏检现象得到缓解。

4　结　论

本文提出了一个基于 CenterNet 与联合注意力的目标检测算法,使用非局部算子提取全局特征,突出需要关注的区域,并优化标签编码与边界框回归方式。

1) 融合非局部特征能够提升编解码网络性能、降低网络对深度的依赖,用更少的层数达到相同的效果,并降低群目标检测的漏检率。

2) 定位质量损失可以大幅提升边界框定位质量,获得更加紧凑的边界框。

3) 通道分流策略通过减少网络宽度与重复梯度信息的学习,降低计算复杂度和内存成本,在可接受的准确率损失下带来速度提升。

参考文献 (References)

[1] 于飞,刘东华,贺飞扬. 无人机“黑飞”对电磁空间安全的挑战[J]. 中国无线电,2018(8):43-44.
YU F,LIU D H,HE F Y. UAV “black flying” challenges the safety of electromagnetic space[J]. China Radio,2018(8):43-44(in Chinese).

[2] GIRSHICK R,DONAHUE J,DARRELL T,et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2014:580-587.

[3] GIRSHICK R. Fast-RCNN[C]// Proceedings of the IEEE Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1440-1448.

[4] REN S,HE K,GIRSHICK R,et al. Faster R-CNN:Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2017(6):1137-1149.

[5] CAI Z,VASCONCELOS N. Cascade R-CNN:Delving into high quality object detection[C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2018:6154-6162.

[6] HE K,GKIOXARI G,DOLLÁR P,et al. Mask R-CNN[C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Piscataway:IEEE Press,2017:2980-2988.

[7] REDMON J,FARHADI A. YOLOv3:An incremental improvement[EB/OL]. (2018-04-08)[2021-01-12]. https://arxiv.org/abs/1804.02767.

[8] LIN T,GOYAL P,GIRSHICK R,et al. Focal loss for dense object detection[C]// Proceedings of 2017 IEEE International Conference on Computer Vision. Piscataway:IEEE Press,2017:2999-3007.

[9] LAW H,DENG J. CornerNet:Detecting objects as paired keypoints[J]. International Journal of Computer Vision,2018,128:642-656.

[10] ZHOU X,WANG D,KRÄHENBÜHL P. Objects as points[EB/OL]. (2019-04-16)[2021-01-12]. https://arxiv.org/abs/1603.06937.

[11] WANG X,GIRSHICK R,GUPTA A,et al. Non-local neural networks[C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2018:7794-7803.

[12] CAO Y,XU J,LIN S,et al. Global context networks[J/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020(2020-12-24)[2021-01-12]. https://doi.org/10.1109/TPAMI.2020-3047209.

[13] YIN M,YAO Z,CAO Y,et al. Disentangled non-local neural networks[EB/OL]. (2020-11-08)[2021-01-12]. https://arxiv.org/abs/2006.06668.

[14] WANG Q,WU B,ZHU P,et al. ECA-Net:Efficient channel attention for deep convolutional neural networks[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2020:11531-11539.

[15] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press,2016:770-778.

[16] YU F,WANG D,SHELHAMER E,et al. Deep layer aggregation[C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2403-2412.

[17] NEWWLL A,YANG K,JIA D. Stacked hourglass networks for human pose estimation[EB/OL]. (2016-03-22)[2020-12-28]. https://arxiv.org/abs/1603.06937.

[18] WANG C,LIAO H Y,WU Y,et al. CSPNet:A new backbone that can enhance learning capability of CNN[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2020:1571-1580.

[19] ZHU X Z,HU H,LIN S,et al. Deformable ConvNets v2:More deformable,better results[C]// Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2019:9308-9316.

Low-altitude UAV detection method based on optimized CenterNet

ZHANG Ruixin[1], LI Ning[1,2,*], ZHANG Xiaxia[1], ZHOU Huiyu[3]

(1. College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;
2. Science and Technology on Electro-optic Control Laboratory, Luoyang 471000, China;
3. University of Leicester, Leicester LE1 7RH, UK)

Abstract: To achieve effective detection of "low, slow and small" unmanned aerial vehicle (UAV) and improve detection accuracy and positioning quality, we propose a low-altitude UAV detection method based on joint attention and CenterNet. Aiming at the problem of high miss-detection rate of small targets in general target detection algorithms, a decoupled non-local operator is introduced to capture the relevance of target regions in optical images. Utilizing the similarity between individuals of the UAV group, the discrete UAV features are correlated to each other to reduce the missed detection rate. Moreover, to obtain more accurate detection boxes, we optimized the label coding strategy and bounding box regression method of CenterNet, and the positioning quality loss is introduced to improve the positioning quality of the detection boxes. Experimental results show that the optimized S-CenterNet algorithm has an average accuracy increase of 8.9% compared with the original CenterNet, and the detection boxer positioning quality has been significantly improved.

Keywords: object detection; deep learning; joint attention; CenterNet; unmanned aerial vehicle

Received: 2021-03-05; **Accepted**: 2021-05-05; **Published online**: 2021-06-02 08:50

URL: kns.cnki.net/kcms/detail/11.2625.V.20210601.1134.002.html

Foundation items: Aeronautical Science Foundation of China (ASFC-20175152036); Key Project on Artificial Intelligence (1004-56XZA19008)

* **Corresponding author**. E-mail: lnee@nuaa.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 405,407 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

《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

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48卷 第11期　2022年11月

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

JOURNAL OF BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS (JBUAA)

(Monthly, Started in 1956)

Vol.48 No.11　November 2022

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

ISSN 1001-5965

中国标准连续出版物号：ISSN 1001-5965 / CN 11-2625/V　　国内定价：50.00元／期

