



ISSN 1001-893X
CN 51-1267/TN
CODEN DIJSFP

www.teleonline.cn

TE3[®]

中文核心期刊
中国科技核心期刊
中国核心学术期刊
中国应用型权威期刊



关注本刊 更多资源



扫码收藏 阅读分享

电讯技术

Telecommunication Engineering

2025

第65卷 第3期 (Vol.65 No.3)

3

ISSN 1001-893X



9 771001 893250



中国电子科技集团公司第十研究所 主办

The Tenth Research Institute of China Electronics Technology Group Corporation



诚信花开 文明社会



中宣部宣教局 中国文明网

电 讯 技 术

Dianxun Jishu

(1958 年创刊,月刊)

第 65 卷

第 3 期(总第 436 期)

2025 年 3 月 28 日

目 次

应用基础与前沿技术

- 一种基于博弈协调策略的通信网络优化算法 陈瑞霞,海 洁,庞学民,吴青娥(339)
- 基于度量学习和子域自适应的辐射源个体识别 周 锋,杜奕航,赵 芸,乔晓强,张 涛(347)
- 基于分布式小型前端组阵的长波信号智能检测方法 陈 怀,胡贇鹏,沈智翔,李 睿(354)
- 基于 Arnold 映射的抗分选混沌调制 PRI 设计方法 刘光霞,李 琦,韩壮志,魏英珍,杨彦端(363)
- 基于 UNet 结构的大规模 MIMO 系统 CSI 反馈设计 刘 庆,李 义,李 璘,王有军,李 康,熊林麟,王 平(371)
- 一种识别和检测人工智能生成文本的算法 王雨欣,刘柯飞,李雪莲,王红军(378)

电子与信息工程

- 基于模型的复杂航天电子载荷系统工程技术方法 赵晓虎,孙 杰,吴慧伦,柴 霖(385)
- 基于有限方向随机游走的火箭发射测控机动布站方法 董 浩(392)
- 5G NR 网络中一种新颖的移动用户速度估计方法 周宝龙,陆 犇,杨洪生,唐 亮(398)
- 双 RIS 辅助 MIMO 系统混合张量信道估计算法 李双志,邢益博,雷豪杰(407)
- 一种解决移动平台下卫星接收遮挡的双天线去重拼接方法 王英杰,曾富华,王钧慧,陈文渊(415)
- 双频段高冗余度列车多普勒测速雷达设计与实现 彭泽霄,郝洪民,李 博(422)
- 基于子阵列的大规模 ISAC 系统混合波束赋形设计 郭鸿儒,郭应鸿,顾忆宵,夏 斌(429)
- 一种用于低分辨毫米波通信系统的低采样率定时恢复算法 李世宝,赵成锁,李作志(437)
- 基于亚模函数的可见光通信 MIMO-OFDM 系统天线选择算法 贾科军,贺耀民,张芳芳,蔺 莹,薛建彬,郝 莉(445)
- 非地面通信网络多普勒频偏多级高精度估计方法 李昌森,申 滨,黄晓舫(454)
- 基于传输时间的 LoRa 网络参数分配与功率控制算法 宋雨昕,杨茂恒,贾奇铭,章 辉(463)
- 基于 STAR-RIS 和中继集成系统的下行 NOMA 覆盖增强技术 陈思宇,申 滨,元文军(469)
- 一种超宽带四臂螺旋钻孔雷达天线 李佳颖,刘建霞,赵珍珍,师振盛,张安学(477)

综述与评论

- 基于深度半监督学习的目标检测技术综述 何智杰,肖 玮,刘南清,高甲博,柯学良,曲乃铸(484)

其他

- 公益广告:诚信花开 文明社会 封二
- 稿约 封三

TELECOMMUNICATION ENGINEERING

(Monthly, Started in 1958)

Vol. 65

No. 3 (Series No. 436)

March 28, 2025

CONTENTS

Application Fundamental Research and Advanced Technology

- A Communication Network Optimization Algorithm Based on Game Coordination Strategy
..... CHEN Ruixia, HAI Jie, PANG Xuemin, WU Qing'e (339)
- Specific Emitter Identification Based on Metric Learning and Subdomain Adaptation
..... ZHOU Feng, DU Yihang, ZHAO Yun, QIAO Xiaoqiang, ZHANG Tao (347)
- An Intelligent Detection Method of Long-wave Signal Based on Distributed Small Front Arrays
..... CHEN Huai, HU Yunpeng, SHEN Zhixiang, LI Rui (354)
- A Design Method of Anti-sorting Chaotic Modulation PRI Based on Arnold Mapping
..... LIU Guangxia, LI Qi, HAN Zhuangzhi, WEI Yingzhen, YANG Yanduan (363)
- Design of UNet-based CSI Feedback in Massive MIMO System
..... LIU Qing, LI Yi, LI Lin, WANG Youjun, LI Kang, XIONG Linlin, WANG Ping (371)
- An Artificial Intelligence-generated Text Recognition and Detection Method
..... WANG Yuxin, LIU Kefei, LI Xuelian, WANG Hongjun (378)

Electronics and Information Engineering

- A Model-based Complex Spacecraft Electronic Payloads System Engineering Technology Method
..... ZHAO Xiaohu, SUN Jie, WU Huilun, CHAI Lin (385)
- A Mobile Deployment Method for Rocket Launch TT&C Stations Based on Random Walk in Finite Direction DONG Hao (392)
- A Novel Speed Estimation Scheme for Mobile Users in 5G NR Networks
..... ZHOU Baolong, LU Ben, YANG Hongsheng, TANG Liang (398)
- A Hybrid Tensor Channel Estimation Algorithm for Double-RIS Assisted MIMO Systems
..... LI Shuangzhi, XING Yibo, LEI Haojie (407)
- A Dual-antenna De-duplication and Montage Method for Solving Satellite Reception Occlusion on Mobile Platforms
..... WANG Yingjie, ZENG Fuhua, WANG Junhui, CHEN Wenyuan (415)
- Design and Implementation of a Dual-frequency High Redundancy Train Doppler Speed Measurement Radar
..... PENG Zezhou, GAO Hongmin, LI Bo (422)
- Hybrid Beamforming Design for Subarray-based Integrated Sensing and Communication Systems
..... GUO Hongru, GUO Yinghong, GU Yixiao, XIA Bin (429)
- A Low Sampling Rate Timing Recovery Algorithm for Low Resolution Millimeter Wave Communication Systems
..... LI Shibao, ZHAO Chengsuo, LI Zuozhi (437)
- An Antenna Selection Algorithm for Visible Light Communication MIMO-OFDM Systems Based on Submodular Function
..... JIA Kejun, HE Yaomin, ZHANG Fangfang, LIN Ying, XUE Jianbin, HAO Li (445)
- A High-precision Doppler Frequency Offset Estimation Method for Non-terrestrial Networks
..... LI Changmiao, SHEN Bin, HUANG Xiaoge (454)
- A Parameter Allocation and Power Control Algorithm Based on Transmission Time for LoRa Networks
..... SONG Yuxin, YANG Maoheng, JIA Qiming, ZHANG Hui (463)
- Downlink NOMA Coverage Enhancement Based on Integrated STAR-RIS and Relay ... CHEN Siyu, SHEN Bin, YUAN Wenjun (469)
- An Ultra-wideband Quadrifilar Helix Borehole Radar Antenna
..... LI Jiaying, LIU Jianxia, ZHAO Zhenzhen, SHI Zhensheng, ZHANG Anxue (477)

Summarization and Review

- Survey of Object Detection Technology Based on Deep Semi-supervised Learning
..... HE Zhijie, XIAO Wei, LIU Nanqing, GAO Jiabo, KE Xueliang, QU Naizhu (484)

DOI:10.20079/j.issn.1001-893x.240422001

一种基于博弈协调策略的通信网络优化算法*

陈瑞霞¹, 海洁¹, 庞学民¹, 吴青娥²

(1. 郑州西亚斯学院 电信与智能制造学院, 郑州 451150; 2. 上海第二工业大学 计算机与信息工程学院, 上海 201209)

摘要:针对通信系统中信息流拥挤与通信阻塞问题,提出了一种基于博弈协调策略算法的通信网络优化模型。研究了一种博弈协调策略新算法,设计了请求事件、响应总队长计算模型,构建了一种通信信号区域协调新模型;对冗余请求事件和响应提出了冗余信息滤除方法。协调优化了给定区域内多路通信信号,通过仿真计算比较了在不同的协调概率情况下的通信效果,结果表明,所提算法平均通信准确度和通信速度较现有常用 Pursuit-evasion game 算法分别提高了 19.03% 和 1.37 ms,且具有更强抗干扰能力。研究结论可对急诊、火警、加密传输、军事通信、密集资源优化调度等领域提供基础方法和数据支撑。

关键词:通信网络;通信阻塞;区域协调;冗余信息滤除

开放科学(资源服务)标识码(OSID):



中图分类号:TN915 文献标志码:A 文章编号:1001-893X(2025)03-0339-08

A Communication Network Optimization Algorithm Based on Game Coordination Strategy

CHEN Ruixia¹, HAI Jie¹, PANG Xuemin¹, WU Qing'e²

(1. School of Telecommunications and Intelligent Manufacturing, Zhengzhou Sias University, Zhengzhou 451150, China;

2. School of Computer and Information Engineering, Shanghai Polytechnic University, Shanghai 201209, China)

Abstract: For the issues of information flow congestion and communication congestion in communication system, a communication network optimization model system based on game coordination strategy algorithm is proposed. A new algorithm of game coordination strategy is developed, with calculation model of request event and response total length designed, thereby constructing a new communication signal regional coordination model. A redundant information filtering method is proposed for redundant request events and responses. The multi-channel communication signals in a given area are coordinated and optimized, and the simulation calculation is completed to compare the communication effect under different coordination probabilities. From experimental comparison results, the average communication accuracy and communication speed of the proposed algorithm have been improved by 19.03% and 1.37 ms respectively compared with the commonly used Pursuit-evasion game algorithm, and it has stronger anti-interference ability. This research can provide fundamental methods and data support for emergency treatment, fire alarm, encrypted transmission, military communication, intensive resource optimization scheduling and other fields.

Key words: communication network; communication blocking; regional coordination; redundant information filtering

* 收稿日期:2024-04-22;修回日期:2024-08-02

基金项目:河南省科技攻关项目(232102210034);郑州西亚斯学院重点学科信息与通信工程(0810)

通信作者:陈瑞霞 Email:10746@sias.edu.cn

0 引言

随着电话、短信、Web 网络等发展,网络通信阻塞现象逐渐加剧,尤其是在大型网络通信系统中。为有效应对信息流拥挤和通信阻塞挑战,自 21 世纪以来,众多国家纷纷投入信息分流控制研究。但由于网络通信系统随机性、复杂性,以及当前信息分流模型局限性,这些努力在优化网络通信效果方面并未取得理想成果。

Sharma 等人^[1]提出了一种基于 Agent 的远程智能教学系统模型,可以提高动态交互性能,但对大量的信息存储使模拟视频信息的传输和分发不适应通信带宽的要求。蒲玮等人^[2]提出了另一种基于 Agent 行动图的作战建模方法,虽然解决了一般基于 Agent 作战建模过程中没有军事人员主导下的军事概念模型问题,但对于利用数字压缩算法压缩视频信息容量进行视频的数字化传输会丢失信息。郭喆^[3]提出的 Agent 建模已成为应对作战复杂适应系统的高效建模仿真手段之一,成功解决了众多军事难题,但对强干扰、频繁截获信息的能力等特点不能达到军事通信的标准。随着 ABM 方法在建模仿真领域的实践应用,毛新军等人^[4]提出的面向 Agent 程序设计逐渐崭露头角,与面向对象程序设计并行成为新兴软件范式。然而,该文献方法无法解决数据链易受到电磁干扰而导致通信中断的问题。顺应这一趋势,宋兆涵等人^[5]提出了面向 Agent 软件工程概念,包括基于知识工程、对象技术以及角色和组织模型等多种软件工程方法学的相继提出。虽然该文献分析了数据链失锁的敏感频带以及干扰信号阈值,但没有给出数据链通信中断机理。为更精确地描述 Agent 行动图,Rull 等人^[6]采用对象约束语言来约束类图,从而实现建模语法形式化描述。简玉梅等人^[7]提出了一种基于 Agent 自信度的无线传感器网络多跳路由协议,均衡了网络能量消耗,延长了网络生存时间。该文献虽然研究了基于能效最优的功率分配算法,然而没有将功率分配问题转化为求解最大系统能效的优化问题。陈东海等人^[8]分析了非均匀分簇算法,提出了将簇头节点分为管理节点、融合节点和转发节点,通过优化结构来节省网络能耗,但是没有将求解非凸优化问题转化为求解凸优化问题。Sun 等人^[9]将单跳通信方式改成多跳通信方式,由簇头负责将邻居簇首数据转发给 Sink 节点,但没有通过循环迭代给出优化的闭合解来达到降低算法的复杂度。冯进等人^[10]提出了一种混合

Agent 框架,在框架中对每一层智能给出解决方案,并应用在电网通信网络区域协调中。该文献虽然建立了正交频分复用抗干扰方案,但没有给出抗干扰性能误码率具体运算公式与方法。

Fu 等人^[11]针对车辆同步途中转移提货和配送问题,提出了一种混合整数线性规划模型和一种具有可接受解的启发式算法,但没有给出能发送精确相对位置坐标偏移量的模块化集成导航通信数据链建立方法。Li 等人^[12]给出了基于关键字感知路由规划研究,分为精确匹配和基于近似匹配路由规划,但没有解决对模拟调频数据链的频率的通用转换以达到免劫持、免攻击的目的,更未实现对模拟数据链依赖性的突破性优化。Nahum 等人^[13]提出了一种多目标模型,同时最小化控制交叉口数量的疏散路线与最优疏散路线之间的疏散时间间隔,但对多输入多输出的交叉口控制没有给出大量异构数据的冗余通信过滤方法。Bertolino 等人^[14]提出了一种针对给定程序流程图寻找路径覆盖的广义算法。该算法简单灵活,可通过对当前问题采用最合适选择策略来解决不可行路径问题,但没有具体给出有效信号的压缩和加密算法以及冗余数据的过滤方法,以解决对海量图片的存储利用率和图片安全性问题。Avola 等人^[15]针对无人机重复访问 n 个同等优先级目标持续性监测问题,给出最优解,并开发了较快地计算最优解算法。然而,对在空域、时域、频域上怎样实现抗干扰问题,没有讨论时空编码的相结合来实现空间分集、时间分集和频率分集,以及大量无用信息的通信约简。

针对现有文献没有解决的网络通信阻塞问题,本文提出了疏通网络阻塞现象优化模型,最后连接到局域网进行实验验证。

1 本文模型

本模型由阻塞状态指标计算、优化策略、冗余滤除 3 个功能模块组成,如图 1 所示。

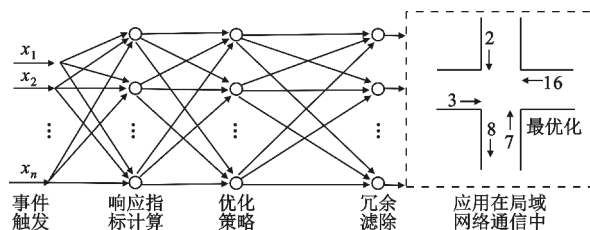


图 1 网络阻塞疏通优化模型

2 影响网络阻塞的关键指标计算

2.1 响应队长

假定 $N(t)$ 表示在时刻 t 系统中正在被服务的请求数,即 $N(t)$ 表示时刻 t 处理单元响应系统等待响应总队长, $t \geq 0$, 设间隔时间为 Δt 内有 j 个请求概率为

$$p_{ij}(\Delta t) = P\{N(t+\Delta t) = j | N(t) = i\} \quad (1)$$

式中: $i, j = 0, 1, 2, \dots$ (下同)。

则由随机过程分析计算得到

1) $p_{i,i+1}(\Delta t) = P\{\text{在 } \Delta t \text{ 内有一个请求但系统响应未完成}\} + \sum_{j=2}^{\infty} P\{\text{在 } \Delta t \text{ 内有 } j \text{ 个请求但系统响应完成 } j-1\} = \lambda \Delta t + o(\Delta t);$

2) $p_{i,i-1}(\Delta t) = P\{\text{在 } \Delta t \text{ 内没有一个请求但响应完成一个}\} + \sum_{j=1}^{\infty} P\{\text{在 } \Delta t \text{ 内有 } j \text{ 个请求但响应完成 } j+1\} = \mu \Delta t + o(\Delta t);$

3) 类似分析可得 $p_{ij}(\Delta t) = o(\Delta t), |i-j| \geq 2$ 。

综合,有

$$p_{ij}(\Delta t) = \begin{cases} \lambda \Delta t + o(\Delta t), & j = i+1, i \geq 0 \\ \mu \Delta t + o(\Delta t), & j = i-1, i \geq 1 \\ o(\Delta t), & |i-j| \geq 2 \end{cases} \quad (2)$$

于是 $\{N(t), t \geq 0\}$ 是可列无限状态 $E = \{0, 1, 2, \dots\}$ 上的请求-响应过程,其参数为

$$\begin{cases} \lambda_i = \lambda, & i \geq 0 \\ \mu_i = \mu, & i \geq 1 \end{cases} \quad (3)$$

假定 $\rho = \lambda/\mu$, 根据 λ 和 μ 意义,可以称 ρ 为系统通信强度。

于是,在统计平衡条件下 $\rho < 1$, 平均响应队长为

$$\bar{N} = E[N] = \sum_{j=0}^{\infty} j p_j = \frac{\rho}{1-\rho}, \rho < 1 \quad (4)$$

另外,根据响应队长分布可知,系统空闲的概率为 $p_0 = 1 - \rho$, 而系统繁忙概率为 ρ 。显然, ρ 越大,系统越繁忙。

2.2 响应时间

假定请求是先到先响应,根据上面讨论,可推理得到请求等待时间如下:

在统计平衡 $\rho < 1$ 下,请求等待时间分布函数 $W_q(t) = P\{W_q \leq t\}$ 为

$$W_q(t) = 1 - \rho e^{-\mu(1-\rho)t}, t \geq 0 \quad (5)$$

平均等待时间为

$$\bar{W}_q(t) = E[W_q] = \frac{\rho}{\mu(1-\rho)}, \rho < 1 \quad (6)$$

3 博弈协调优化算法

借助多 Agent 技术,并结合博弈论,本文构建了一种针对网络通信的区域优化协调模型。该模型旨在强化交换机与路由端口间协同合作,从而实现对特定区域内通信信号精准调控与优化。此模型可更加有效地解决通信阻塞和信息流拥挤问题,显著提升网络通信效率和稳定性。

3.1 模型结构

局域网通信模型由多个局域单元共同构建。在多 Agent 通信控制系统中,存在两类主要 Agent: 一种是区域 Agent (Area-Agent), 负责特定区域内通信管理与协调; 另一种则是路由端口 Agent (Cross-Agent), 主要承担信息在不同局域单元之间传输与交换任务。这两类 Agent 协同工作,共同保障整个局域网通信系统顺畅运行。每个 Agent 主要包括 3 个层次: 进行 Agent 间协调且生成最终决策的协作层、进行 Agent 间信息交互的通信层和通过通信层将控制任务信息传递给其他 Agent 的控制层。

3.2 建立协调模型

3.2.1 协调模型

在网络通信全过程中,特定时刻响应数量有限,而端口带宽宽度也固定,导致在某一时间点网络能够处理的信息量和数据传输速度都受到一定限制。每个 Agent 通过信息交换来掌握其他 Agent 效用函数、策略空间等特性,并根据这些 Agent 可能采取的策略来制定自身决策。这一过程实质上是一个基于全面信息的博弈过程,确保了各 Agent 之间有效协调。

一次博弈协调,定义为

$$G = \{A, I, S, U\} \quad (7)$$

式中: $A = \{1, 2, \dots, n\}$ 是所有 Agent 集合 (在博弈协调过程中,行为和决策 Agent 会努力选择最佳行动策略,以尽可能提升自身收益水平); 每个 Agent 都具有独特行动策略和特征信息 I ; S 表示 Agent 所有可能行动或策略的集合,即 $S_i = \{\text{水平双向, 垂直双向, 水平转换, 垂直转换}\}$ 表示第 i 个 Agent 策略空间,它涵盖了 Agent 所有可行选择范围,通过在这个策略空间中寻找最优解,Agent 做出最有利于自身的决策; U 为 $Q_i(t)$ 收益函数,表示 Agent 获得利益,在特定策略组合下,体现 Agent 得失情况 (U_i 表示第 i 个 Agent 所获得利益。 $Q_i(t)$ 表示在 t 时刻等候

在第 i 个端口的响应数向量, $Q_i(t) = \{Q_{i,hr}(t), Q_{i,hl}(t), Q_{i,vu}(t), Q_{i,vd}(t)\}$, $Q_{i,hr}(t)$ 、 $Q_{i,hl}(t)$ 、 $Q_{i,vu}(t)$ 、 $Q_{i,vd}(t)$ 分别表示 t 时刻等候在第 i 端口水平向右、水平向左、垂直向上、垂直向下 4 个方向的响应数。 Q_i 表示 i 端口响应数阈值向量, $Q_i = \{Q_{i,hr}, Q_{i,hl}, Q_{i,vu}, Q_{i,vd}\}$, $Q_{i,hr} \setminus Q_{i,hl} \setminus Q_{i,vu} \setminus Q_{i,vd}$ 分别表示不同方向上等候响应数的阈值)。

每个 Agent 基于其掌握的信息 I , 在策略集 S 中选择最适宜策略, 再通过持续协调过程, 最终实现盈利 Nash 均衡状态, 即

$$U_i(t|s_i^*, s_{-i}^*) \geq U_i(t|s_i, s_{-i}^*), \forall s_i \in S_i \quad (8)$$

式中: s_i^* 代表第 i 个 Agent 所选策略; s_{-i}^* 代表除 i 之外所有 Agent 策略所组成的向量。

整个协调过程可分为 3 个层级: 上层主要聚焦于区域 Agent 与邻近区域 Agent 之间协同合作; 中层则关注区域 Agent 与端口 Agent 之间协调与配合; 而下层则更加关注端口 Agent 与其相邻端口 Agent 之间协调与沟通。3 个层级之间协调关系共同构建了整个网络通信系统协同机制, 如图 2 所示。

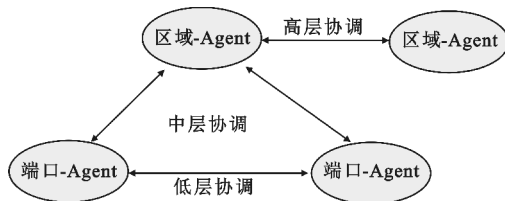


图 2 协调层次

3.2.2 协调算法

各 Agent 节点按照公式(8)算法进行博弈, 在博弈中实施协调优化, 如图 3 所示。

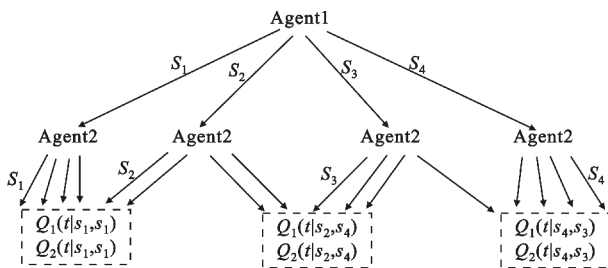


图 3 博弈树

第一步, 当端口 Agent 响应排队数量超过其设定阈值, 它会主动向相邻端口 Agent 发送请求。在通信负载较重时, 寻求相邻端口 Agent 协助, 以缓解自身处理压力。

第二步, 在接收到请求后, 相邻端口 Agent 会积极响应, 并构建出博弈树。博弈树分支上标注着不同策略选择, 框图内字母代表各种策略下盈利情况。随后, 通过遍历和搜索此博弈树, 利用公式(8)来确定 Nash 均衡点。通过策略分析和盈利比较, 找到一种能使所有参与 Agent 都获得满意收益的平衡状态。

第三步, 如果在搜索博弈树后未能找到 Nash 均衡, 端口 Agent 会转向本端口 Agent 发送请求; 否则, 各个 Agent 将采纳在 Nash 均衡状态下的策略, 并按此策略来控制端口设备, 本次协调过程随即结束。

第四步, 一旦区域 Agent 接收到请求, 它会立即启动对所辖端口 Agent 的博弈协调流程, 力求达到 Nash 均衡状态。如果在协调过程中未能实现这一目标, 区域 Agent 会主动向邻近的 Agent 发出协调请求, 以寻求更广泛合作与支持。

第五步, 收到请求的相邻区域 Agent 会迅速响应, 共同开展博弈协调, 旨在找到 Nash 均衡。若经过共同努力仍未能取得突破, 整个协调过程将宣告终止。此时, 所有参与的 Agent 将维持原有策略, 继续执行各自通信任务, 确保网络通信系统稳定运行。

4 冗余信息过滤方法

用户在实际请求网络事件时, 若得到系统的即时响应或响应时间在设定阈值内, 则请求等待进入系统接受服务; 若响应时间超过设定阈值, 就放弃本次请求事件。对请求而又没接受系统服务的事件响应或小概率响应, 可以滤除。另外, 网络对请求事件响应完, 通信系统还存留了一些已经响应的历史记录, 这些事件也是冗余信息, 都可以通过冗余信息过滤方法将其过滤掉。为使网络响应更快, 这里给出一种可变概率冗余信息过滤方法。

由第二部分对请求与响应队长、等待响应总队长计算, 对网络通信中每个节点在每单位时间内处理请求事件效率做计算, 给出区域网络节点效率值 $m_i(X)$ 。根据下一时刻有无请求行动判断当前时刻请求是否为正常事件, 如果是, 则计算在内; 如果不是, 则忽略。那么可计算 $m_i(X)$ 如下:

$$m_i(X) = \frac{n_i}{N} \quad (9)$$

式中: X 是请求事件; $m_i(X)$ 是第 i 个节点对事件 X 响应效率值; n_i 是单位时间内正常请求事件队长; N

是第 i 个节点在单位时间内响应总队长,即服务队长。

由 $m_i(X)$, 定义总效率函数为 $B^*(X) = \sum_{D \subseteq X, i} m_i(D)$ 。

根据上面效率值计算,进一步缩小有效响应范围,滤除无效响应,最终求得过滤掉冗余请求事件的约简结果,即设单位时间内总请求事件为集合 X ,总效率函数为 $B^*(X)$ 。若在集合 X 中,设 Y_1 为去掉其中一个元素后的集合, $B^*(Y_1)$ 为其总效率函数,若 $|B^*(X) - B^*(Y_1)| < \varepsilon$, 即某个元素值低于预设阈值 ε , 则认为该元素可以被剔除。这一过程会持续进行,每次迭代都会去除不满足条件的元素,直至得到最小过滤子集 Y_k 。 Y_k 就是最终过滤结果。

5 阻塞优化算法及其在局域网中的应用

以通信网络中一个特定端口为例,说明上述协调算法如何运作。在这个特定场景中, Agent 1、Agent 2 和 Agent 3 分别扮演 3 个端口 Agent 角色,并统一由一个区域 Agent 管理。为方便分析,对每个 Agent 策略集 S 进行简化处理,现在包含以下 4 种策略: s_1 代表水平双向通信; s_2 代表垂直双向通信; s_3 代表水平转换通信; s_4 代表垂直转换通信,端口间距离为 10 m。当需要协调通信时,这些端口 Agent 会开始相互作用,通过构建博弈树和搜索最优策略来达成纳什均衡。它们会考虑各自收益矩阵,根据当前网络状态、其他 Agent 策略选择以及自身通信需求来做出决策。在此过程中,每个 Agent 都会评估不同策略组合下等待响应数、通信效率以及可能的冲突情况。它们会尝试找到一种策略组合,使得在给定条件下,所有 Agent 收益都能达到最大化或至少达到一个可接受平衡点。一旦找到这种策略组合,即达到纳什均衡,每个 Agent 就会根据自己的策略来控制端口,执行相应通信操作。这样,整个通信网络性能就得到优化,减少通信阻塞和信息流拥挤问题,提高通信效率和质量。

在 t_0 时刻, Agent 2 检测到其水平向左的排队响应数 $Q_{2,h}(t_0) = 16$ 超出了预设阈值 Q_{hl} , 这里设定 $Q_{hl} = 12$ 。因此, Agent 2 向 Agent 1 发出了协调请求。Agent 1 在接收到请求后,立即与 Agent 2 进行博弈协调。这场博弈协调核心目标就是减少排队响应数,而这正是 Agent 1 和 Agent 2 通过协调所能获得

的收益。它们的策略组合将直接影响收益大小。设每个周期内到达端口每个方向响应数为 λ , 同时每个发送信息时段内通过端口响应数为 μ , 这里 $\mu = 2.5\lambda$ 。端口 1 和端口 2 在 t_1 时刻排队响应数,实际上由 Agent 1 和 Agent 2 在 t_0 时刻所采取策略决定。通过构建博弈树,不同策略组合将导致不同排队响应数变化,进而影响到 Agent 收益。有以下 4 种可能情况:

① Agent 1 和 Agent 2 都选择某一特定策略 s_1 , 那么两个端口等待响应数分别为 $11 + 2\lambda$ 和 $9 + 4.5\lambda$ 。

② Agent 1 选择某一策略 s_1 , 而 Agent 2 选择另一策略 s_2 , 两个端口等待响应数分别为 $11 + 2\lambda$ 和 $19 + 3.5\lambda$ 。

③ Agent 1 选择策略 s_2 , Agent 2 维持原先选择策略 s_1 , 那么两个端口等待响应数分别为 $10 + 3.5\lambda$ 和 $9 + 2\lambda$ 。

④ Agent 1 和 Agent 2 保持同样选择策略 s_2 , 两个端口等待响应数分别为 $10 + \lambda$ 和 $19 + \lambda$ 。

每个 Agent 决策背后都有一个收益矩阵作为支撑。通过这个矩阵,可以清晰地看到,当 $\lambda < 10$ 满足某个条件时, Agent 1 会选择策略 s_1 , 而 Agent 2 也会选择策略 s_1 , 这样它们就能达到纳什均衡状态;而在 $\lambda > 10$ 满足另一个条件时, Agent 1 会选择策略 s_2 , Agent 2 选择策略 s_2 , 同样也能实现纳什均衡。具体来说,当 $\lambda < 10$ 时, Agent 1 会选择 s_1 (水平双向) 作为行动策略,而 Agent 2 也会选择 s_1 (水平双向);而当 $\lambda > 10$ 时, Agent 1 会选择 s_2 (垂直双向), Agent 2 同样选择 s_2 (垂直双向)。这样, Agent 1 和 Agent 2 就能根据各自选择的策略来精准控制端口,实现通信优化。

上层协调与下层协调思路非常相似,只需将区域 Agent 视为端口 Agent 来处理即可。遵循同样决策逻辑和纳什均衡原则,就能实现区域间通信优化。

在整个通信过程中,本文使用博弈算法,建立 Agent 协调模型,其通信网络拓扑结构如图 4 所示。然后,建立收益博弈算法: $t = t_0 + \Delta t > T_0$ 和 $Q_{i,h}(t) - Q_{i,v}(t) > L$, L 为设定等待响应的响应队长阈值, T_0 为仿真结束时间, t_0 为赋予水平或垂直方向通信响应时间初值, Δt 为通过 $Q_{i,h}(t) - Q_{i,v}(t) > L$ 博弈后再次开通水平或垂直方向通信延长时间。在 $Q_{i,h}(t) - Q_{i,v}(t) > L$ 博弈过程中,使用模糊神经网络训练收益

$Q_{i,h}(t)$ 或 $Q_{i,v}(t)$ 满足 Nash 均衡, 确定水平或垂直方向通信延长时间, 迭代训练直至满足仿真结束条件。

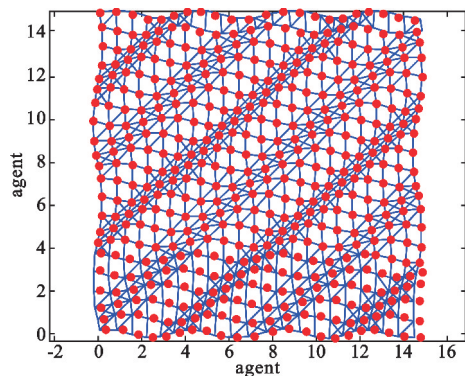


图4 建立的 Agent 协调模型通信网络拓扑结构

按协调概率分别为 $p=0.1$ 、 $p=0.2$ 、 $p=0.3$ 、 $p=0.4$ 、 $p=0.5$ 的程度进行博弈, 网络实际通信效果非常逼近预期设定的通信效果, 逼近误差效果如图 5 所示。

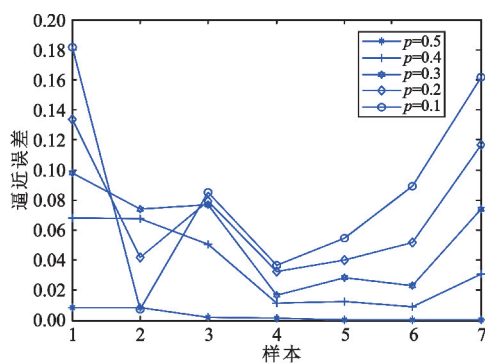


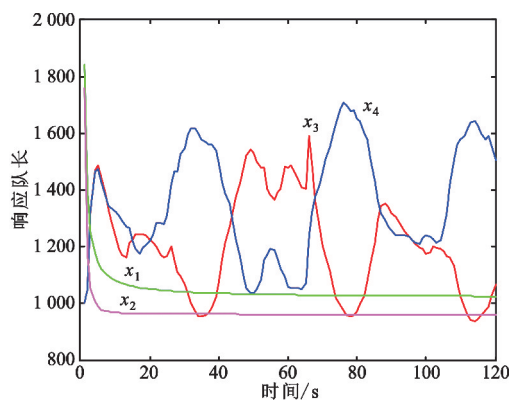
图5 实际通信与设定通信效果的逼近误差

由图 5 知, 随着协调概率增大, 逼近误差越来越小, 说明通过各级博弈进行协调, 通信效果更好。协调概率大的通信效果保真率高, 失真性小, 通信速度快, 通信量小。

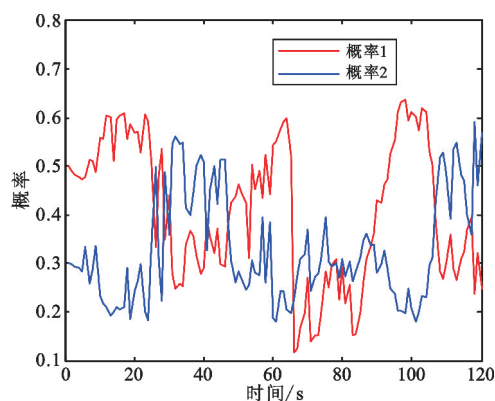
6 本文优化算法与传统通信算法比较

6.1 比较分析

为验证本文提出的优化算法对通信网络作用效果比单独传统网络通信效果好, 给出了响应队长与响应时间的比较。以变概率约简冗余信息后, 响应队长逐渐变短, 响应时间也变短。响应队长和响应时间可分别由公式(4)和(6)计算, 得到仿真结果如图 6 和图 7 所示。



(a) 响应队长与响应时间对比



(b) 接受服务的两种不同概率

图6 响应队长的新方法与传统方法比较

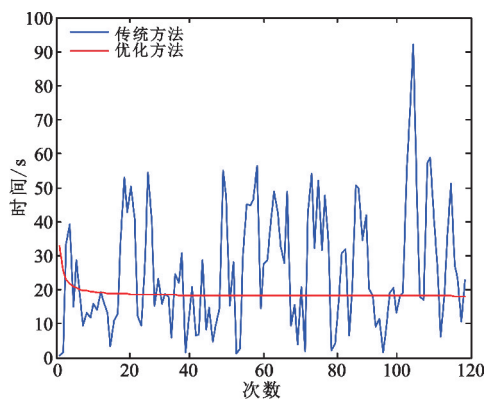


图7 响应时间比较

从图 6 可以看出, 本文算法处理请求响应快, 所以接受响应等待队长短, 等待队长也较稳定, 曲线变化平稳。而传统通信响应时间长, 等待事件队列根据请求事件接受服务的概率也不同, 等待时间长短不稳定, 表现在队长曲线幅度变化不定。图 6(a) 中 x_1 和 x_2 分别表示新算法在请求事件以两种不同概率接受服务下的等待队长, x_3 和 x_4 分别表示传统算法在请求事件以两种不同概率接受服务下的等待

队长。

由图 7 可以看出,本文算法响应时间短,随着迭代次数增加,响应时间趋于一个稳定值。而传统通信算法响应时间长,随着迭代次数增加,曲线振荡且幅度偏高。

6.2 比较优化模型

除以上比较特点外,本文算法准确率也更高。为衡量准确率,可定义如下:

编写网络系统响应次数计数程序。设第 j 时间段内响应总次数为 N_j ,通过计数程序统计出有效服务次数为 n_j ,对实验重复进行 N_0 次,那么定义有效响应平均精度 P_a 为

$$P_a = \frac{n_j}{N_j}, a = 1, 2, \dots, N_0 \tag{10}$$

定义总平均精度 P_m 为

$$P_m = \frac{\sum_{a=1}^{N_0} P_a}{N_0} \tag{11}$$

在 1 s 内 400 次请求,对每次实验重复进行 100~500 次,利用计数程序统计得到每次有效响应 n_j ,然后分别由公式(10)和(11)计算出平均精度和总平均精度值,如图 8 所示。

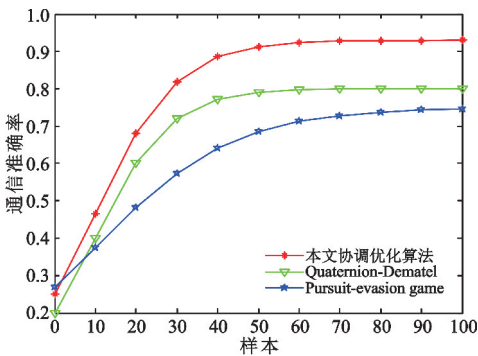


图 8 提出的网络优化模型与传统通信方法在有效响应中的比较

从图 8 可以看出,随着次数增加,优化模型精度越来越好,说明问题越复杂,优化模型解决得越好。本文算法具有更好的疏通阻塞能力,有效通信准确率可达 90% 以上。

本文对几种通信算法进行了 100 次对比实验。实验中使用 50 组电网数据进行训练,其余 50 组数据作为测试样本。在实验中,功率信号被分成帧。每帧的仿真时间约为 10 ms,正确通信速率可以通过正确传输的帧数除以总帧数得到。

上述实验过程重复 20 次,取平均值。将本文算

法与现有合作博弈算法进行比较。实验结果表明,本文算法平均特征准确率为 90.72%,文献[1]提出的四元数 Dematel 算法平均特征准确率为 77.48%,文献[15]提出的 Pursuit-evasion game 算法平均特征准确率为 71.69%,如图 8 和表 1 所示。另外,本文算法的通信速度为 2.89 ms,而四元数 Dematel 算法为 3.97 ms,Pursuit-evasion game 算法为 4.26 ms。

表 1 本文算法和现有通信算法综合比较

优化通信算法	平均通信准确度/%	通信速度/ms	抗干扰能力
本文算法	90.72	2.89	强
Quaternion-Dematel ^[1]	77.48	3.97	弱
Pursuit-evasion game ^[15]	71.69	4.26	弱

从图 8 和表 1 可以看出,在大量通信系统中,本文算法正确通信率高于现有通信算法,运行时间最短,具有较好的通信能力。

图 9 给出了本文提出的优化模型系统的稳定性仿真结果,可见该优化模型系统的稳定性误差振动幅度在开始时较大,随着时间推移逐渐收敛到一个稳定值。这说明本文提出的优化系统具有真正稳定收敛性。

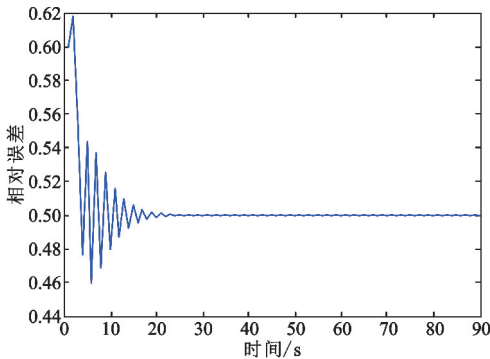


图 9 优化模型的稳定性

7 结束语

本文提出的算法通过计算阻塞状态参数来识别网络瓶颈和潜在阻塞点,利用构建区域优化协调模型,采用博弈协调优化算法对给定区域内通信信号实现了精准调控与优化,最后通过一种可变概率冗余信息滤除方法精简网络资源,提升了网络性能。通过实验室模型搭建和计算分析,给出的优化模型比传统的通信速度提高了 10%,可以有效解决信息流拥挤与通信阻塞问题,优化通信效果,同时具有较强抗干扰能力。

下一步将利用本文研究成果提出不同异构数据链的编码算法,解决视频流传输、语音传输等的失真、耗时、干扰、依赖性强的问题;建立深度卷积神经网络核函数优化的抗拥挤、抗阻塞模型,将其应用在无人机数据链的抗劫持、抗攻击、抗干扰的安全通信传输中。

参考文献:

- [1] SHARMA V, KUMAR R, KUMAR R. QUAT-DEM: quaternion-DEMATEL based neural model for mutual coordination between UAVs[J]. Information Sciences, 2017, 418: 74–90.
- [2] 蒲玮,李雄. 基于 Agent 行动图的作战建模方法[J]. 系统工程与电子技术, 2017, 39(4): 795–805.
- [3] 郭喆. 基于排队论的战术通信混合接入方法[J]. 电讯技术, 2021, 61(1): 58–62.
- [4] 毛新军,胡翠云,孙跃坤,等. 面向 Agent 程序设计的研究[J]. 软件学报, 2012, 23(11): 2885–2904.
- [5] 宋兆涵,曾贵明,梁君. 基于双频通信的空间飞行器自组网动态时隙分配[J]. 电讯技术, 2022, 62(3): 305–310.
- [6] RULL G, FARRÉ C, QUERALT A, et al. AuRUS: explaining the validation of UML/OCL conceptual schemas[J]. Software & Systems Modeling, 2015, 14(2): 953–980.
- [7] 简玉梅,张韩飞,高飞,等. 一种基于 Agent 自信度的簇间多跳路由协议[J]. 传感技术学报, 2017, 30(1): 126–132.
- [8] 陈东海,李长庚. 基于簇头功能分化的无线传感器网络成簇算法[J]. 传感技术学报, 2015, 28(2): 244–248.
- [9] SUN Y Q, PENG J, LIU T. Uneven clustering routing protocol based on dynamic partition for wireless sensor network[J]. Journal of Communications, 2014, 35(1): 198–206.
- [10] 冯进,朱江,沈寿林. 一种基于分层智能混合决策的多 Agent 框架[J]. 火力与指挥控制, 2017, 42(1): 36–39.
- [11] FU Z X, CHOW J Y J. The pickup and delivery problem with synchronized en-route transfers for microtransit planning[J]. Transportation Research Part E: Logistics and Transportation Review, 2022, 157: 1–21.
- [12] LI K, RAO X, PANG X B, et al. Route search and planning: a survey[J]. Big Data Research, 2021, 26: 1–11.
- [13] NAHUM O E, WACHTEL G, HADAS Y. Planning tourists evacuation routes with minimal navigation errors[J]. Transportation Research Procedia, 2020, 47: 235–242.
- [14] BERTOLIN A, YIN J L, CHEN B H, et al. Automatic intermediate generation with deep reinforcement learning for robust two exposure image fusion[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 21(2): 2162–2371.
- [15] AVOLA D, CINQUE L, FAGIOLI A, et al. Automatic estimation of optimal UAV flight parameters for real-time wide areas monitoring[J]. Multimedia Tools and Applications, 2021, 80(16): 25009–25031.

作者简介:

陈瑞霞 女, 1984 年生于河南开封, 2011 年获硕士学位, 现为副教授, 主要研究方向为信息传输与处理、嵌入式系统。

海洁 女, 1979 出生于河南郑州, 2007 年获硕士学位, 现为教授, 主要研究方向为电子技术。

庞学民 男, 1963 年生于河南长垣, 2006 年获硕士学位, 现为副教授, 主要研究方向为信号分析、SoC。

吴青娥 女, 1971 年生于河南新乡, 2008 年获博士学位, 现为教授, 主要研究方向为图像处理、信息融合。

DOI:10.20079/j.issn.1001-893x.240505002

基于度量学习和子域自适应的辐射源个体识别*

周 锋^{1,2}, 杜奕航², 赵 芸¹, 乔晓强², 张 涛²

(1. 南京信息工程大学 电子与信息工程学院, 南京 210044;

2. 国防科技大学第六十三研究所, 南京 210007)

摘 要:为解决辐射源个体识别中信号传输环境变化引起的数据分布不一致,导致仅接受单一分布数据集训练的网络模型识别准确率严重退化这一问题,提出结合度量学习和子域自适应的辐射源个体识别方法。该方法借鉴了领域自适应中子域自适应的思想,应用局部最大均值差异损失来缩小不同分布下相同辐射源类别之间的差异,并在其基础上加入基于欧氏距离和余弦相似度的度量学习损失,稳定迁移效果。实验表明,在同时使用了度量学习损失和子域自适应方法后,目标域识别准确率相比于未使用迁移方法提高了 38.7% 左右,并且模型具有良好的泛化能力。

关键词:辐射源个体识别;度量学习;子域自适应;余弦相似度

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN971 文献标志码:A 文章编号:1001-893X(2025)03-0347-07

Specific Emitter Identification Based on Metric Learning and Subdomain Adaptation

ZHOU Feng^{1,2}, DU Yihang², ZHAO Yun¹, QIAO Xiaoqiang², ZHANG Tao²

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. The 63rd Research Institute, National University of Defense Technology, Nanjing 210007, China)

Abstract: In order to solve the problem of inconsistent data distribution caused by changes in the transmission environment in specific emitter identification, which leads to serious degradation in identification accuracy of network models trained only with a single distribution data set, a method for specific emitter identification that combines metric learning and subdomain adaptation is proposed. This method draws on the idea of subdomain adaptation in domain adaptation, applies local maximum mean difference loss to reduce the differences between the same emitter identification categories under different distributions, and the metric learning loss based on Euclidean distance and cosine similarity is added to stabilize the migration effect. Experiments show that after using both the metric learning loss and the subdomain adaptation method, the target domain recognition accuracy is improved by about 38.7% compared with that of the non-transfer method, and the model has good generalization ability.

Key words: specific emitter identification; metric learning; subdomain adaptation; cosine similarity

0 引言

辐射源个体识别技术(Specific Emitter Identification, SEI)^[1]是通过提取辐射源信号中独特

的射频指纹,并利用特定的分类算法对辐射源个体进行识别的一种方法。该技术能够针对同一型号、批次及工作方式下的不同辐射源,综合分析并有效

* 收稿日期:2024-05-05;修回日期:2024-06-28
基金项目:国家自然科学基金资助项目(62371463)
通信作者:杜奕航 Email:dyhcs1991@163.com

挖掘出不同辐射源个体之间的固有差异。如今通信辐射源种类的数量日益增多,SEI 技术为有效管理频谱环境的安全提供了有力支持^[2]。

近年来,随着深度学习技术的迅猛发展,这些技术开始在辐射源个体识别领域中展现出巨大潜力。与传统的辐射源个体识别方法相比,深度学习方法采用端到端的设计,不再严重依赖于专家经验。研究者们已经在辐射源个体识别领域采用了多种不同的深度学习技术方法,例如,使用了卷积神经网络来处理信号的 Hilbert 谱图^[3]、Welch 功率谱图^[4]等,相较于传统方法取得了明显的性能提升。此外,还有一维卷积神经网络和长短期记忆网络被广泛应用于直接对基带 I/Q 序列数据进行特征提取和分类^[5-6]。这些方法充分利用了深度学习的特征提取和自学习能力,极大地提升了分类性能。然而,上述研究都是基于测试样本和训练样本服从同一分布这一假设,但在实际应用中,辐射源个体所处环境是在不断变化的,如长时间下的环境温度、湿度、大气条件的变化可能会对发送设备内部器件的性能造成一定影响,并且传输信道类型的不同、外界干扰信号以及发射器或接收器的任何移动,都可能使信号在捕获中产生变化^[7-8],因此,由单一分布数据训练的网络模型在面对其他分布数据时模型性能往往会发生严重下降。

领域自适应方法被引入来有效应对这一问题,其核心思想是通过学习一个共享的特征表示,使得训练数据和待测数据之间的特征分布更加一致,从而实现更好的泛化性能。领域自适应方法已在计算机视觉领域得到了长期发展,目前可以分为基于统计差异^[9-10]和基于对抗^[11-12]两种主流方法。由于领域自适应理论和方法逐渐成熟,也有学者将其应用到辐射源个体识别上,如文献[13]利用最大均值差异(Maximum Mean Discrepancy, MMD)从多个源域中学习无关于信道噪声的射频指纹特征,文献[14]和文献[15]分别将多核 MMD 和域对抗神经网络(Domain Adversarial Neural Network, DANN)结合连续小波变换实现对不同载波频率信号的高识别率。目前利用域适应解决信号传输环境变化对 SEI 模型性能影响的研究较少。

本文针对由信号传输环境变化引起待测数据与训练数据分布不一致而导致模型性能下降的问题,借鉴了深度子域自适应^[16]中的局部最大均值差异(Local Maximum Mean Discrepancy, LMMD)方法,并结合了两种距离度量学习损失,从欧氏距离和余弦

相似度的两个角度出发,缩小源域数据的类内距离,扩大源域数据的类间距离,稳定迁移源域至目标域的识别效果。实验表明,本文提出的方法在不影响源域准确率的同时,在目标域上的识别准确率能够达到 99% 左右,并且面对未在训练阶段出现的数据集具有良好的泛化能力。

1 辐射源个体识别架构及分析

1.1 模型组成

本文提出如图 1 所示的模型结构,该模型结构主要由特征提取模块和损失计算模块构成。其中,特征提取模块由残差网络结构和多层全连接层构成,即经典神经网络 18 层残差网络(Residual Network 18, ResNet18)的变体,主要作用是利用深度神经网络的强大的特征提取能力,将样本空间变换到对应特征空间;损失计算模块主要根据特征提取模块提取出来的特征,计算相应 LMMD 损失、度量学习损失,交叉熵损失根据分类器得到的预测标签计算得出。在训练阶段,以 3 种损失的组合使用指导特征提取模块学习源域数据和目标域数据的一致性特征分布;在测试阶段,由训练阶段保存的模型参数对测试数据进行模型性能验证。

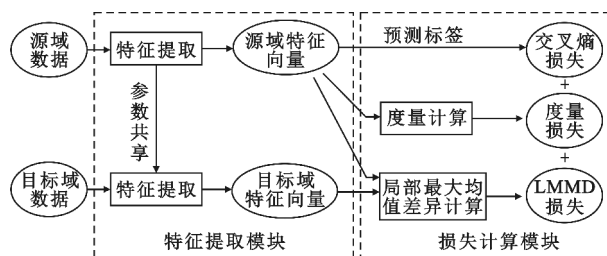


图 1 网络模型

1.2 特征提取模块

特征提取模块具体结构如图 2 所示,其为 ResNet18 的变体,改变了其中卷积核以及步长的大小以适配 IQ 基带数据,并增加多层全连接层调整最终特征向量的长度。每个 Block 结构包含了 4 层卷积层,每层卷积层都对其进行批归一化和 ReLU 激活操作。批归一化可以有效防止网络过拟合,并使其快速收敛;ReLU 激活函数则帮助网络实现对非线性数据的拟合。Block 结构左侧的连接线表示“恒等映射”,目的在于避免模型层数加深时网络出现性能退化,恒等映射每两个卷积层进行一次连接,图中在每个 Block 结构上下进行连接仅为示意。

Size 表示数据经过网络时, 在各个阶段时的尺寸。

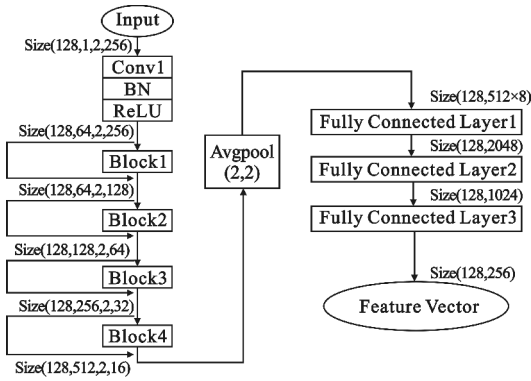


图 2 特征提取模块结构

此外, 图中显示的是将特征提取模块的最后一层全连接层的输出作为特征向量, 实际上, 每层全连接层的输出都可以作为特征向量, 用来计算相应损失。

1.3 损失计算模块

1.3.1 基于局部最大均值差异的迁移损失

在无监督的领域自适应任务中, 首先给定一个有标签的源域数据集 $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$, 其中, x_i^s 表示源域第 i 个样本, y_i^s 表示其对应的真实标签, 源域的样本总量为 N_s 。同时给出一个无标签的目标域数据集 $D_t = \{x_i^t\}_{i=1}^{N_t}$ 。 D_s 和 D_t 的样本来自不同数据分布 p 和 q (p 不等于 q)。领域自适应主要是通过将源域和目标域映射到一个共同的特征空间, 应用目标函数对齐源域与目标域的样本数据, 从而消除域之间的差异分布。领域自适应可以分为全局域自适应和子域自适应两部分, 如图 3 所示。

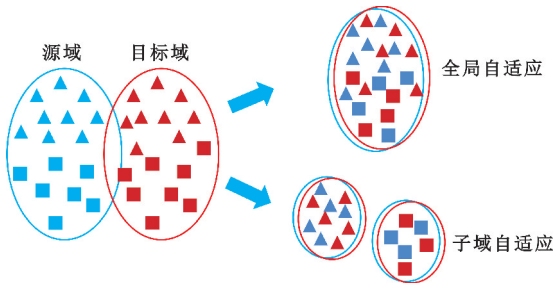


图 3 全局自适应与子域自适应

如图 3 上分支部分所示, 全局自适应从所有样本整体上对齐源域与目标域的分布, 而没有考虑不同域之间同一类别内互相之间的联系, 导致不同类别之间分布间距小, 造成不同类别样本的混淆。子

域自适应在图 3 的下分支部分进行了描述, 其挖掘源域与目标域相关的类别进行有目标的对齐。当源域和目标域的子域分布对齐时, 避免了不同类别样本的混淆, 可以提高分类精度。

MMD 因其并不需要进行参数估计, 已被广泛应用于测量源分布和目标分布之间的差异。基于深度 MMD 的方法主要关注全局分布的对齐, 忽略了同一类别两个子域之间的关系, 不同类别相互之间容易发生混淆。而 LMMD 方法能够衡量两域中相同类别之间的分布差异。LMMD 具体公式如下:

$$d_H(p, q) \triangleq E_c \| E_{p^{(c)}} [\varphi(x^s)] - E_{q^{(c)}} [\varphi(x^t)] \|_H^2 \quad (1)$$

式中: H 为具有特征核 k 的再生核希尔伯特空间 (Reproducing Kernel Hilbert Space, RKHS); x^s 和 x^t 分别是 D_s 和 D_t 中的实例; $p^{(c)}$ 和 $q^{(c)}$ 分别是 $D_s^{(c)}$ 和 $D_t^{(c)}$ 的分布, c 表示所属类别; 映射函数 $\varphi(\cdot)$ 将样本实例映射到 RKHS 中。 $d_H(p, q)$ 即表示在局部层面两域之间的分布差异大小。

为了便于使用神经网络提取的特征进行计算, 将 LMMD 的重新表示为

$$\begin{aligned} \hat{d}_H(p, q) = & \frac{1}{C} \sum_{c=1}^C \left\| \sum_{x_i^s \in D_s} \omega_i^{sc} \varphi(x_i^s) - \sum_{x_j^t \in D_t} \omega_j^{tc} \varphi(x_j^t) \right\|_H^2 = \\ & \frac{1}{C} \sum_{c=1}^C \left[\sum_{i=1}^{N_s} \sum_{j=1}^{N_t} \omega_i^{sc} \omega_j^{tc} k(z_i^s, z_j^s) + \right. \\ & \left. \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} \omega_i^{tc} \omega_j^{tc} k(z_i^t, z_j^t) - \right. \\ & \left. 2 \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} \omega_i^{sc} \omega_j^{tc} k(z_i^s, z_j^t) \right] \quad (2) \end{aligned}$$

式中: 核函数 $k(z^s, z^t) = \langle \varphi(x^s) \cdot \varphi(x^t) \rangle$, 其中 $\langle \cdot \rangle$ 表示向量内积, 这里采用高斯核函数; z_i^s 与 z_j^t 表示对应样本经过神经网络提取出的样本特征; ω_i^{sc} 和 ω_j^{tc} 分别表示属于第 c 个类别的 x_i^s 和 x_j^t 的权重, 同类别的 ω 和为 1, 公式如下:

$$\omega_i^c = \frac{y_i^c}{\sum_{(x_j, y_j) \in D} y_j^c} \quad (3)$$

式中: y_i^c 表示该样本被认为属于类别 c 的概率; ω_i^c 可以解释为该样本属于类别 c 的概率在占全部样本的比重 (在训练过程中, 实际为一个批次中的比重)。对于源域中的数据, 本文使用真实标签 y 作为一个已知向量 (对应真实标签为 1, 其余为 0) 来计算每个样本的 ω_i^c 。然而, 在目标域没有标记数据的无监督自适应中, 由于标签 y^t 不可用, 从而无法直接计算。因此, 对于没有标签的目标域 D_t , 选择使用源域分类器的输出作为概率分布分配给各个类

别。然后,可以计算每个目标样本的 ω_i^c ,并可以计算出最终 LMMD 的值。需要注意的是,访问源域的标签毫无疑问是基本准确的,而对于目标域,由于目标域的标签由源域分类器模型预测而来,因此得到标签可能是错误的,进而使用这个错误的标签会降低最终的模型性能。

1.3.2 基于欧氏距离和余弦相似度的度量学习损失

在迁移过程,源域与目标域之间的对齐效果很大程度上依赖于伪标签的准确率。而度量学习能在逐步优化源域的特征分布,即缩小类内距离,增加类间距离,以此来稳定迁移过程。

度量学习通过优化距离度量函数,引导模型在特征空间中更好地定义样本之间的距离或相似度,从而增强类内一致性和增大类间差异。本文采用了欧氏距离和余弦相似度的度量方式,如图 4 所示。

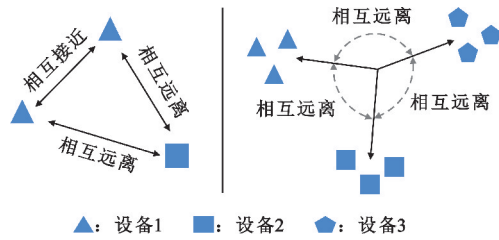


图 4 度量损失

如图 4 左图所示,欧氏距离衡量了两个样本特征间的绝对距离。利用欧氏距离使得属于相同类别的样本相互接近,而属于不同类别的样本之间相互远离。在同一批内计算不同样本间的欧氏距离,公式如下:

$$\frac{1}{N_b} \sum_{i=1}^{N_b} \max \left[\frac{1}{N_{pj}} \sum_{pj=1}^{N_{pj}} d_{\text{edu}}(x_i, x_{pj}) - \frac{1}{N_{nj}} \sum_{nj=1}^{N_{nj}} d_{\text{edu}}(x_i, x_{nj}) + \text{margin_edu}, 0 \right] \quad (4)$$

式中: $d_{\text{edu}}(\cdot)$ 表示欧氏距离(值域为0~正无穷); N_b 表示一批中样本数量; N_{pj} 表示与 x_i 为同类的样本数; N_{nj} 表示与 x_i 为异类的样本数; margin_edu 是一个可以调节的常数。正例之间的距离要比正例到负例之间的距离要小且小于 margin_edu 。在理想情况时,同类距离应为0,不同类距离应为较大的值。

如图 4 右图所示,不同于欧氏距离关注绝对距离,余弦相似度衡量不同样本特征在方向上的差异性,属于相同类别样本的特征向量在方向上会互相接近,反之则远离。在同一批内计算不同样本间的

余弦相似度,公式如下:

$$\frac{1}{N_b} \sum_{i=1}^{N_b} \max \left[-\frac{1}{N_{pj}} \sum_{pj=1}^{N_{pj}} d_{\text{cos}}(x_i, x_{pj}) + \frac{1}{N_{nj}} \sum_{nj=1}^{N_{nj}} d_{\text{cos}}(x_i, x_{nj}) + \text{margin_cos}, 0 \right] \quad (5)$$

式中: $d_{\text{cos}}(\cdot)$ 表示余弦相似度(值域为 $(-1, 1)$); N_b 、 N_{pj} 、 N_{nj} 的含义与公式(3)相同; margin_cos 同样是一个可以调节的常数(正例之间的相似度比正例与负例之间的相似度要大且大于 margin_cos 。在理想情况时,同类相似度应为1,不同类相似度应为-1,因此需要在相同类别的度量前加上负号)。

网络模型整体优化目标函数具体如下公式所示:

$$\min \sum (L_{\text{cross}} + \lambda \cdot d_{\text{H}}(p, q) + L_{\text{edu}} + L_{\text{cos}}) \quad (6)$$

式中: L_{cross} 为交叉熵损失函数; $d_{\text{H}}(p, q)$ 为子域自适应函数; λ 为动态权重参数(随迭代次数增加而变大); L_{edu} 和 L_{cos} 是两种度量损失。通过最小化损失函数,学习源域和目标域共享的分类特征,以提高分类性能。

2 实验与结果分析

2.1 数据集准备

本次实验采用的数据集为 WiSig^[17]。WiSig 数据集是一个大规模的公开 WiFi 信号数据集,其致力于实现信道和接收器不可知时的辐射源个体识别,包含了来自 174 个现成的 WiFi 发射器和 41 个 USRP 接收器的 1 000 万个数据包,跨越了一个月的 4 次捕获。发射机均遵循 IEEE802.11a/g 标准,并与 WiFi 接入点一起在中心频率 2 462 MHz 和带宽 20 MHz 的 WiFi 信道 11 上工作,接收机采样率为 25 Msample/s。

WiSig 数据集的作者表明改变接收器或使用不同日期捕获的信号会显著降低训练分类器的性能,即数据产生不同分布的表面原因也为上述两点。本文仅对捕获于不同日期的信号进行迁移分析,暂不考虑因为使用不同接收机而产生的数据分布差异问题。采集日期不同对数据的影响主要有:存在干扰信号可能影响数据质量;辐射源组件性能可能随时间或者环境因素(湿度、温度等)发生变化,从而使数据分布发生变化。在时域上,Day1 和 Day2 采集的数据样本如图 5(a)所示;在频域上,Day1 和 Day2 数据在幅值上有所差异,如图 5(b)所示。

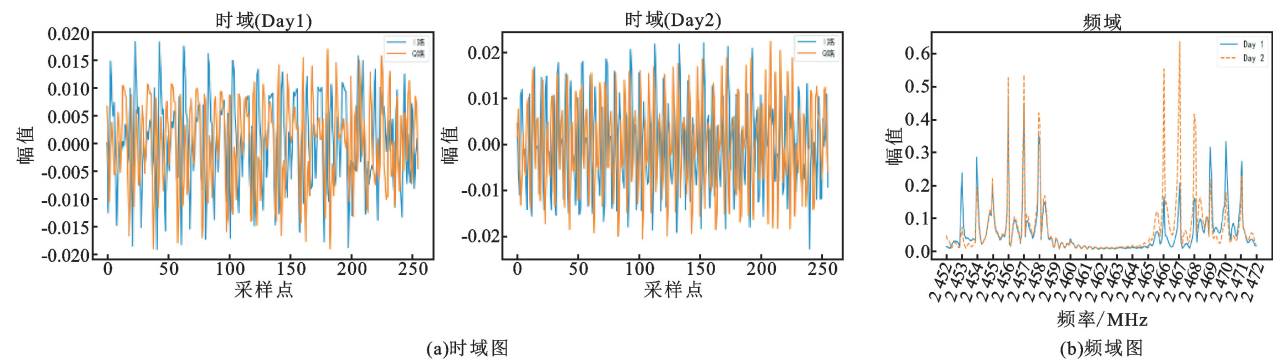


图 5 不同日期数据差异

由于 WiSig 数据集的原始数据量过于庞大,本文仅采用其中的已被打包好的数据集 ManySig 作为训练测试使用,使用捕获于不同日期的数据作为各个目标域。ManySig 数据集包含 6 台发射器和 12 台接收器,并在一个月内采集了 4 天的原始 IQ 数据,每类样本数量为 1 000 个,样本尺寸为 2×256 。

6 台发射机型号为 Atheros AR5212/AR5213,采用的接收机型号为 N210。本文将 6 台发射机作为 6 类辐射源,以 Day1 作为源域数据,Day2、Day3、Day4 作为目标域数据,接收机在 4 个不同日期下均为 1 号接收机。为模拟没有任何先验信息,使用未经过均衡操作的信号。数据的 80% 划分为训练集,20% 划分为测试集,训练集中目标域标签不可知。

实验基于 Python 下的 Pytorch 深度学习框架实现。初始学习率设置为 0.001,优化器为 Adam;源域与目标域每批输入样本数量相同,大小为 128。

2.2 横向对比实验

为了验证所提方法的有效性,使用不同的迁移学习模型进行横向对比分析。

模型 1:不使用任何迁移方法的 ResNet18。依据样本尺寸对模型进行一定修改,最终结构与图 4 所示的特征提取模块结构一致,并在最后加上一层全连接层作为分类器。仅使用该模型分别对均衡前和均衡后的数据进行训练,以此来比较使用均衡处理和迁移方法的效果,信道均衡已被文献[18]证明对去除环境干扰具有有效性。

模型 2: CORAL 方法^[10] (Deep Coral, D-CORAL)。CORAL 方法用线性变换方法将源域和目标域分布的二阶统计特征进行对齐,本质上是基于统计差异的方法,并且在深度网络中应用 CORAL 方法后,其非线性变换更加强大。

模型 3:域对抗神经网络 DANN^[11]。基于对抗的迁移学习方法,利用域判别器和特征提取器之间相互对抗来对齐源域与目标域之间的分布差异。

模型 4: 多表示领域自适应网络 (Multi-Representation Adaptation Network, MRAN)^[18],通过多表示对齐来完成跨域分类任务,并扩展了最大均值差异来计算适应损失。

模型 5: 本文提出的方法,结合度量损失和 LMMD 损失的辐射源个体识别方法。

所有模型均使用由图 4 所示的特征提取模块得到的特征向量(MRAN 添加多表示分支),然后 4 种模型对特征向量采取不同的处理,最终由计算得到的损失随迭代次数增加更新模型参数。迁移任务 A→B 表示案例 A 数据为源域,案例 B 数据为目标域。本次实验设置 3 种迁移任务:迁移任务 1 为 Day1→Day2;迁移任务 2 为 Day1→Day3;迁移任务 3 为 Day1→Day4。上述 3 种迁移任务在不同迁移学习模型下,目标域的识别准确率如表 1 所示。

表 1 不同迁移方法下各目标域识别准确率

迁移任务	目标域识别准确率/%					
	Resnet18	Resnet18 (均衡)	DeepCoral	DANN	MRAN	本文 方法
Day1→Day2	53.5	83.1	94.5	56.5	93.3	99.8
Day1→Day3	59.1	83.0	92.0	97.7	97.8	99.7
Day1→Day4	70.5	84.9	98.2	90.2	97.2	99.7

上述实验在 Day1 数据,即源域数据上的识别准确率均为 99%,但未在图中显示。由表 1 可知,在 3 个迁移任务中,结合度量学习损失和 LMMD 损失的辐射源个体识别方法在目标域上的识别准确率达到 99% 以上。在实验中发现,ResNet18 的模型性能出现了严重退化,识别准确率仅为 50%~70%,表明当训练与测试日期不相同时,两者的数据分布存在较大差异对模型性能产生了影响;模型 DANN 在迁移任务 Day1→Day2 中出现了难以收敛的现象,识别准确率仅有 55% 左右。综合 3 个迁移任务,本文提出的方法相较于未使用迁移的方法,在目标域准确率平均提高了 38.7%;相较于使用均衡处理后

的识别效果,提高约有 16.0%。

2.3 消融实验

为了验证度量学习损失和 LMMD 损失的有效性,使用消融实验对两者进行验证,即分为以下 3 种情况:设置 1,仅使用交叉熵损失;设置 2,使用交叉熵损失和 LMMD 损失;设置 3,使用交叉熵损失、LMMD 损失和度量学习损失。在 3 个迁移任务下,依据不同的设置对目标域的识别准确率进行比较。不同条件下的目标域识别准确率如表 2 所示。

表 2 消融实验下各目标域识别准确率

迁移任务	识别准确率/%		
	设置 1	设置 2	设置 3
Day1→Day2	53.5	58.1	99.7
Day1→Day3	59.7	68.8	99.3
Day1→Day4	70.6	66.7	99.6

由表 2 可以看出,当仅使用有交叉熵损失时,不同日期获取的数据分布确实存在差异,对网络的识别性能造成了较大影响。而仅采用 LMMD 损失时,并没有达到与源域识别率相近的水准(需要注意的是在设置 2 下,迁移失败的现象并不是一定发生,同样存在迁移成功的实验)。在同时使用了度量损失和 LMMD 损失下,其迁移任务中准确率最高,说明度量学习损失在源域和目标域迁移时起到了稳定的作用,增强了模型的推广能力和稳定性。3 种条件下,模型在源域数据上的识别准确率均为 99%。

本文将 Day1→Day3 迁移任务中的损失条件设置 2 的识别率可视化,如图 6 所示。在图 6 中,采用 LMMD 损失作为迁移方法时,目标域的辐射源 2 与辐射源 5 发生了混淆。我们认为发生混淆的原因是分类器给予了目标样本错误的伪标签,迫使遵从伪标签对齐的 LMMD 方法将目标域的辐射源 5 域与源域的辐射源 2 对齐,从而造成错误的对齐效果。

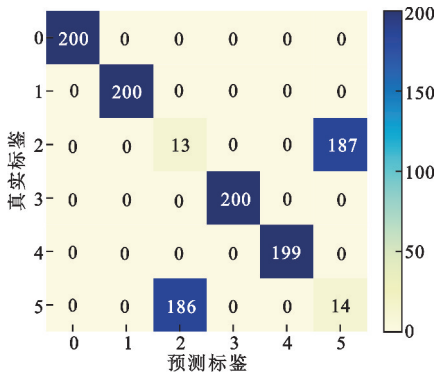


图 6 设置条件 2 的识别率混淆图

从特征降维可视化进一步分析图 6 中辐射源 2 与源域的辐射源 5 错误对齐的现象。采用 T-SNE 算法将模型提取的特征向量降至二维并可视化。图 7 中左图对应设置条件 3 的模型,右图对应设置条件 2 的模型,蓝色表示 Day1 数据,绿色表示 Day3 数据。可以看到在发生错误对齐的右图中,Day1 数据与 Day3 数据的辐射源 2 与辐射源 5 发生了混淆,符合图 6 对应结果。

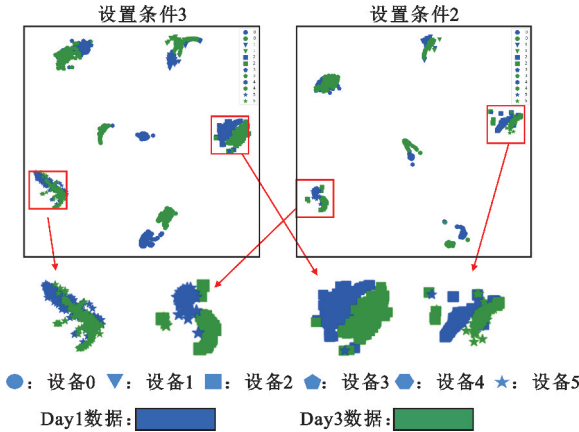


图 7 特征降维可视化

2.4 模型泛化实验

在上述实验中,每个迁移任务的训练阶段仅使用了源域(有标签)和目标域(无标签)数据,即两个不同的日期。Wisig 数据集总共提供了 4 个不同的采集数据日期。为了验证模型的泛化能力,本文将 3 个迁移任务训练出的模型保存,测试其余两个日期,识别准确率如表 3 所示。

表 3 不同日期数据集的识别准确率

迁移任务	识别准确率/%		
	Day2	Day3	Day4
Day1→Day2		98.4	94.6
Day1→Day3	99.4		99.6
Day1→Day4	94.7	99.3	

如表 3 所示,即使是模型未见过的数据集,模型最低的识别准确率也能达到 94% 以上。文献[19]同样也测试了 Wisig 不同日期的效果,该论文作者提供了两种实验设置:①训练 Day1,测试另外 3 个日期,平均识别准确率保持在 56.1%;②训练 Day2、Day3、Day4,测试 Day1,平均识别准确率保持在 77.3%。实验表明,模型在应对训练阶段未见过的数据集时依然有一个良好表现。

3 结束语

本文针对信号传输环境随时间变化而变得不可预知, 导致基于深度学习的辐射源个体识别算法性能严重下降的问题, 提出了一种基于度量学习和子域自适应的辐射源个体识别方法。该方法在子域自适应的基础上, 通过欧氏距离和余弦相似度两种度量损失, 增强类内一致性和增大类间距离, 稳定迁移效果。实验表明, 在测试不同日期下的数据时, 该方法可以有效提高深度神经网络的识别精度。

此次实验并未具体针对数据分布因何种因素导致分布不同而采用相应的解决措施, 而是将所有因素带来的影响归结于数据分布变化, 从而采用一种解决数据分布变化的方法。这在面对不同领域数据初始分布差异较大时, 模型性能可能无法达到预期要求, 因此对导致的数据分布变化的因素进行量化研究可作为进一步研究方向。

参考文献:

- [1] TALBOT K I, DULEY P R, HYATT M H. Specific emitter identification and verification [J]. *Technology Review*, 2003, 113: 113–130.
- [2] XIE L N, PENG L N, ZHANG J Q, et al. Radio frequency fingerprint identification for Internet of Things: a survey [J]. *Security and Safety*, 2024, 3: 1–32.
- [3] PAN Y W, YANG S H, PENG H, et al. Specific emitter identification based on deep residual networks [J]. *IEEE Access*, 2019, 7: 54425–54434.
- [4] 王检, 张邦宁, 魏国峰, 等. 基于 Welch 功率谱和卷积神经网络的通信辐射源个体识别 [J]. *电讯技术*, 2021, 61(10): 1197–1204.
- [5] QING G W, WANG H F, ZHANG T P. Radio frequency fingerprinting identification for Zigbee via lightweight CNN [J]. *Physical Communication*, 2021, 44: 1–8.
- [6] ZHOU X Y, HU A Q, LI G Y, et al. A robust radio-frequency fingerprint extraction scheme for practical device recognition [J]. *IEEE Internet of Things Journal*, 2021, 8(14): 11276–11289.
- [7] WONG L J, MICHAELS A J. Transfer learning for radio frequency machine learning: a taxonomy and survey [J]. *Sensors*, 2022, 22(4): 1–14.
- [8] AL-SHAWABKA A, RESTUCCIA F, D'ORO S, et al. Exposing the fingerprint: dissecting the impact of the wireless channel on radio fingerprinting [C]//*IEEE INFOCOM 2020 – IEEE Conference on Computer Communications*. Toronto: IEEE, 2020: 646–655.
- [9] LONG M S, CAO Y, CAO Z J, et al. Transferable representation learning with deep adaptation networks [J]. *IEEE Transactions on Pattern Analysis and*

Machine Intelligence, 2019, 41(12): 3071–3085.

- [10] SUN B C, SAENKO K. Deep CORAL: correlation alignment for deep domain adaptation [C]//*Computer Vision-ECCV 2016 Workshops*. Cham: Springer International Publishing, 2016: 443–450.
- [11] MURPHY K, SCHÖLKOPF B, GANIN Y, et al. Domain-adversarial training of neural networks [J]. *Journal of Machine Learning Research*, 2016, 17(1): 2096–2030.
- [12] LI X, ZHANG W, XU N X, et al. Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places [J]. *IEEE Transactions on Industrial Electronics*, 2020, 67(8): 6785–6794.
- [13] 李林, 俞璐, 蒋曾辉, 等. 基于类特征对齐的多源域适应辐射源个体识别方法 [J]. *通信技术*, 2023, 56(10): 1137–1145.
- [14] 陈浩, 杨俊安, 刘辉. 基于深度残差适配网络的通信辐射源个体识别 [J]. *系统工程与电子技术*, 2021, 43(3): 603–609.
- [15] HUANG K J, YANG J N, LIU H, et al. Deep adversarial neural network for specific emitter identification under varying frequency [J]. *Bulletin of the Polish Academy of Sciences Technical Sciences*, 2021, 69(2): 1–9.
- [16] ZHU Y C, ZHUANG F Z, WANG J D, et al. Deep subdomain adaptation network for image classification [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(4): 1713–1722.
- [17] HANNA S, KARUNARATNE S, CABRIC D. WiSig: a large-scale WiFi signal dataset for receiver and channel agnostic RF fingerprinting [J]. *IEEE Access*, 2022, 10: 22808–22818.
- [18] ZHU Y C, ZHUANG F Z, WANG J D, et al. Multi-representation adaptation network for cross-domain image classification [J]. *Neural Networks*, 2019, 119: 214–221.
- [19] CHILLET A, BOYER B, GERZAGUET R, et al. Tangled program graph for radio-frequency fingerprint identification [C]//*2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications*. Toronto: IEEE, 2023: 1–7.

作者简介:

周 锋 男, 2000 年生于江苏苏州, 2022 年获学士学位, 现为硕士研究生, 主要研究方向为通信辐射源个体识别、深度学习、迁移学习。

杜奕航 男, 1991 年生于山东济宁, 2019 年获博士学位, 现为副研究员, 主要研究方向为智能频谱认知与管控。

赵 芸 女, 1991 年生于江苏扬州, 2017 年获博士学位, 现为讲师, 主要研究方向为毫米波、亚毫米波天线及器件。

乔晓强 男, 1981 年生于江苏沐阳, 2005 年获硕士学位, 现为研究员, 主要研究方向为电磁频谱智能感知、电磁频谱安全与控制。

张 涛 男, 1988 年生于山东济宁, 2018 年获博士学位, 现为副研究员, 主要研究方向为频谱智能感知。

DOI:10.20079/j.issn.1001-893x.231219007

基于分布式小型前端组阵的长波信号智能检测方法*

陈 怀^{1,2}, 胡赞鹏¹, 沈智翔¹, 李 睿²

(1. 信息工程大学 信息工程学院, 郑州 450001; 2. 中国人民解放军 72221 部队, 济南 250000)

摘 要:面向分布式小型化磁感前端阵列接收长波信号场景,针对长波频段噪声复杂的特性,提出了一种宽带多信号智能联合检测方法。该方法基于已知参考信号样本预训练神经网络,通过神经网络在预训练阶段学习分布式接收样本矢量在各维度上的潜在复杂关联性规律,进而部署网络后可输出基于输入样本矢量联合概率的置信度量,用于判断当前样本是否存在目标信号从而得到检测结果。基于宽带信号仿真数据集进行实验,结果表明算法可直接对宽带数据进行处理,并能有效完成频谱感知,能够在低信噪比和相关噪声条件下获得接近理论处理增益的检测性能,达到 80% 以上的检测率。在此基础上,采用中科院空天信息创新研究院布设于内蒙古的超短基线电磁探测阵列 (Mini-array by Chinese Academy of Sciences, CASMA) 采集的实际长波信号数据进行性能验证,算法性能测试结果同样验证了其有效性。该方法不限于长波信号,也适用于其他具备参考信号条件下的信号盲检测的场景,实现在信道参数未知、信号微弱等盲环境条件下获得更优的目标信号检测性能。

关键词:分布式小型前端阵列;长波信号检测;相关噪声;神经网络

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN911.7 文献标志码:A 文章编号:1001-893X(2025)03-0354-09

An Intelligent Detection Method of Long-wave Signal Based on Distributed Small Front Arrays

CHEN Huai^{1,2}, HU Yunpeng¹, SHEN Zhixiang¹, LI Rui²

(1. Institute of Information Systems Engineering, Information Engineering University, Zhengzhou 450001, China; 2. Unit 72221 of PLA, Jinan 250000, China)

Abstract: In consideration of the complex noise characteristics of long-wave, an intelligent joint detection method for wideband multi-signal is proposed for the scenario in which a distributed miniaturized magnetic sensing front array is used to receive long-wave signals. The method leverages pre-trained neural networks using known reference signal samples to learn the potential correlation laws of the distributed received sample vectors in each dimension during the pre-training phase. After deployment, the network can output confidence metrics based on the joint probability of input sample vectors, enabling the determination of whether the current sample contains a target signal and thus yielding detection results. The algorithm can directly process wideband data and effectively perform spectrum sensing. It achieves near-theoretical processing gain in detection performance under low signal-to-noise ratio (SNR) and correlated noise conditions, with a detection rate exceeding 80%. Furthermore, the method is validated using actual long-wave signal data collected by the ultra-short baseline electromagnetic detection array (Mini-array by Chinese Academy of Sciences, CASMA), which is deployed in Inner Mongolia Autonomous Region by the Institute of Space Information Innovation, Chinese Academy of Sciences. The performance test results further confirm the effectiveness of the proposed method. This method is not only applicable to long-wave signals, but also to other signal blind detections with reference signal, for achieving better detection performance under such blind environmental conditions as channel parameter unknown or weak signals.

Key words: distributed miniaturized front arrays; long-wave signal detection; correlated noise; neural network

* 收稿日期:2023-12-19;修回日期:2024-02-06

基金项目:国家科学自然科学基金资助项目(62171468)

通信作者:胡赞鹏 Email:hyp1978@126.com

0 引言

长波通信频段主要包括甚低频 (Very Low Frequency, VLF)、低频 (Low Frequency, LF) 频段。长波信号频段低而波长长,辐射和接收该频段信号通常需采用较大尺寸的天线,因此传统长波通信系统需占用较大的地理空间。小型化接收天线(如磁感应机理的小型化 VLF/LF 天线)与前端的应用使得长波通信系统应用范围大大扩展。实际应用中的小型化长波磁感前端主要包括感应式磁场传感器和电容式电场传感器。中国科学院空天信息创新研究院研制了一套利用光纤方式同步的超短基线电磁探测阵列 (Mini-array by Chinese Academy of Sciences, CASMA)^[1],可用于测量长波 (20~70 kHz) 发射电台的垂直电场信号。本文在进行信号样本分析和算法验证时采用的实测数据便来源于布设于内蒙古正镶白旗的 CASMA。同时,还有其他团队的小型磁感前端如驻极体式、永磁体式、压电谐振式^[2]等运用了机械天线技术,由于实现方案的不同,天线所适用的频带范围与所能达到的辐射强度各有不同。

同时,长波信号较大的波长使得接收信号统计特性受地理位置及方向影响较大,难以准确建模刻画^[3]。采用小型化天线及前端接收长波通信频段信号时,受频段特性及感应机理影响,接收噪声也呈现出有别于更高频段和传统前端接收的统计特性,如接收信号中的噪声不仅来源于接收机热噪声,通常还包括由天线感应进来的环境噪声,使得噪声总体呈现更加复杂的统计规律。同时,在采用多个天线前端联合接收信号时,由于传播信道及接收环境的潜在客观规律,不同天线前端接收的长波通信频段的信号统计特性、噪声统计特性在一定空、时、频范围内存在显著的多源、高维相关性,这种相关性因素对分布式前端接收的信号联合检测具有十分显著的影响。通过数学模型刻画相关性,进而联合求解,是获取多天线联合检测性能增益的传统途径,但针对分布式小型化天线前端的长波信号接收样本,对多源、高维的复杂客观规律,确定模型刻画的准确性十分有限,进而大大限制了联合检测性能增益的获取。目前也有基于深度神经网络的多传感器联合信号检测方法^[3~8],但现有的研究不适用于长波信号接收中多节点间存在相关噪声、单节点非平稳等复杂背景下的盲检测。

本文提出一种基于多个分布式小型磁感前端的

长波信号分布式智能检测方法及系统,可更好地针对长波信号传播特性和接收环境相关噪声的潜在规律,提升长波信号检测性能。核心思想是,采用多个分布式小型磁感前端组成接收阵列共同接收长波信号,基于已知参考信号预训练神经网络,通过神经网络在预训练阶段学习分布式接收样本矢量在各维度上的潜在复杂关联性规律,进而部署网络后可输出基于输入样本矢量联合概率的置信度量,用于判断当前样本是否存在目标信号,得到检测结果。可在所需的长波接收宽带频段内多个目标信号频点分布信息未知、不同阵列布局的情况下,有效利用分布式接收样本在时间、空间、传播信道及接收环境等方面无法通过传统方法建模的关联性,在保证泛化能力的同时实现宽带信号盲检测,并获得较好的频段内信号检测性能。

1 小型化前端接收的长波噪声特性统计与分析

在实际的小型前端的分布式长波频段信号接收场景中,通过统计阵元之间的噪声相关系数值,对比分析不同阵元之间不同频率下的噪声相关性变化情况,发现噪声相关性与距离、频率均存在密切关系,如图 1 所示。

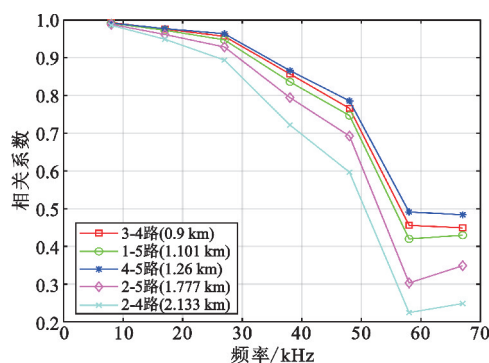
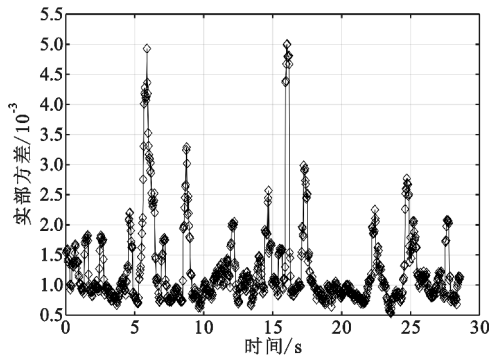


图 1 CASMA 接收噪声相关性分布

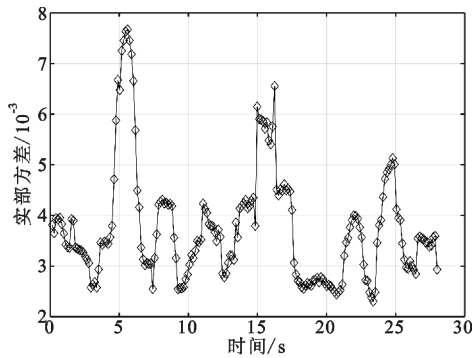
从总体趋势来看,采用小型化磁感应前端接收的长波频段噪声相关性与节点间距离、频段有关,如节点相距 2.133 km 的相关性整体较低、50 kHz 处的相关性低于 30 kHz 处的相关性等。

统计单个阵元信号的非平稳特性,分别统计 0~9 s 内不同窗长下噪声数据的均值方差。单次统计内,统计窗为时长,统计重叠长度为统计窗长的 1/10。如图 2 所示,在 300 ms 和 900 ms 的统计窗

时长下信号方差随时间的变化不同。由上述分析可见,长波频段噪声样本呈现出非平稳特性。



(a) 加窗 300 ms 时 59 kHz 噪声实部方差



(b) 加窗 900 ms 时 59 kHz 噪声实部方差

图2 不同窗长下的信号方差变化统计

2 分布式智能化盲检测算法

2.1 相关噪声条件下的分布式检测

对于经典的分布式检测问题,以 L 个阵元为例,在融合检测时,将各个传感器判定结果输出为似然函数,在融合中心进行最大后验概率融合检测^[3]。根据最大后验准则,联合后验概率可表示为

$$P(H_i | Y_1, Y_2, \dots, Y_L) = \frac{P(H_i, Y_1, Y_2, \dots, Y_L)}{P(Y_1, Y_2, \dots, Y_L)} \quad (1)$$

对于各节点噪声独立假设下有^[9]

$$\begin{aligned} P(H_i, Y_1, Y_2, \dots, Y_L) &= \prod_{l=1}^L P(Y_l | H_i) P(H_i) = \\ &= \prod_{l=1}^L P(Y_l, H_i) = \\ &= \prod_{l=1}^L P(H_i | Y_l) P(Y_l) \end{aligned} \quad (2)$$

然而若各节点噪声不独立,存在相关性,则

$$P(Y_1, Y_2, \dots, Y_L | H_i) \neq \prod_{l=1}^L P(Y_l | H_i) \quad (3)$$

即在节点间噪声相关的场景下, $P(H_i | Y_1, Y_2, \dots, Y_L)$ 的形式将更为复杂,若在处理上仍依据独立假设将导致性能的下降。

考虑一个由 L 个分布式小型磁感前端组成的阵列进行接收,则 t 时刻接收端的信号为

$$\mathbf{X}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \quad (4)$$

式中: $\mathbf{A}$ 为信道传输矩阵(在频谱感知中,传输矩阵对应信道矩阵 $\mathbf{H}$); $\mathbf{s}$ 为发射信号向量; $\mathbf{v}$ 为噪声向量。对 $\mathbf{X}$ 进行 n 次采样获得接收数据矩阵 $\mathbf{X} = [\mathbf{x}_1(n), \mathbf{x}_2(n), \dots, \mathbf{x}_L(n)]$ 。检测算法的目的即为根据 $\mathbf{X}$ 判断目标信号是否存在。

设模型中信号 $\mathbf{s}$ 为确定性信号,噪声为协方差为 $\mathbf{C}$ 的相关噪声,因而阵元之间的观测数据 $\mathbf{X}$ 也彼此相关,即在 H_0 条件下, $\mathbf{X} \sim N(0, \mathbf{C})$; 在 H_1 条件下, $\mathbf{X} \sim N(\mathbf{s}, \mathbf{C})$ 。根据 NP 检测有似然比检测^[10]如下:

$$P(\mathbf{X}; H_1) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}(\mathbf{X}-\mathbf{s})^T \mathbf{C}^{-1}(\mathbf{X}-\mathbf{s})\right] \quad (5)$$

$$P(\mathbf{X}; H_0) = \frac{1}{(2\pi)^{\frac{N}{2}} \det^{\frac{1}{2}}(\mathbf{C})} \exp\left[-\frac{1}{2}\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X}\right] \quad (6)$$

如果 $l(\mathbf{X}) = \ln \frac{P(\mathbf{X}; H_1)}{P(\mathbf{X}; H_0)} > \ln \gamma$ 判为 H_1 , 代入可

得到

$$\begin{aligned} l(\mathbf{X}) &= -\frac{1}{2}[(\mathbf{X}-\mathbf{s})^T \mathbf{C}^{-1}(\mathbf{X}-\mathbf{s}) - \mathbf{X}^T \mathbf{C}^{-1} \mathbf{X}] = \\ &= -\frac{1}{2}[\mathbf{X}^T \mathbf{C}^{-1} \mathbf{X} - 2\mathbf{X}^T \mathbf{C}^{-1} \mathbf{s} + \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} - \mathbf{X}^T \mathbf{C}^{-1} \mathbf{X}] = \\ &= \mathbf{X}^T \mathbf{C}^{-1} \mathbf{s} - \frac{1}{2} \mathbf{s}^T \mathbf{C}^{-1} \mathbf{s} \end{aligned} \quad (7)$$

将与观测数据有关的项放入门限中,即检测器可进一步修改为 $T(\mathbf{X}) = \mathbf{X}^T \mathbf{C}^{-1} \mathbf{s} > \gamma'$ 。

但是在实际应用中,获取噪声协方差需要接收端在检测前先收集一段纯噪声样本^[11]。这种方式在长波频段的噪声和频率有关这个特性上,是有一定的局限性(即实际长波信号中纯噪声的频点和信号所在的频点不同,噪声相关系数仍有差别)。此外,在进行检测前无法提前得到该频点是否存在目标信号或是纯噪声的先验信息。此外,实际场景中天线布局、天气环境等未知因素很多的情况下,无法给出确切噪声参数,如不能准确地估计噪声参数,则在实际应用时神经网络的性能无法得到保证。相较之下,设计具有较强的泛化能力的神经网络实现盲检测,更加适用于分布式小型磁感前端的长波信

号接收场景。

2.2 分布式检测最佳理论性能分析

考虑 N 副天线对信号进行接收, 假设各路信号间的时延差、频率差已获得同步补偿, 其信号模型可以写成如下形式:

$$x_i(k) = a_i s(k) + n_i(k), 1 \leq i \leq N \quad (8)$$

式中: 设 $s(k)$ 功率为 1; a_i 为第 i 路天线接收信号的幅度; $n_i(k)$ 为第 i 路天线接收信号的零均值高斯白噪声, 方差为 σ_i^2 。将上式写成矩阵形式:

$$\mathbf{x} = \mathbf{a} \mathbf{s}(k) + \mathbf{n} \quad (9)$$

式中: $\mathbf{x}$ 、 $\mathbf{a}$ 、 $\mathbf{n}$ 都为 $N \times 1$ 的列向量。定义最佳权值为 $\mathbf{W} = [w_1, w_2, \dots, w_N]^T$, 则合成信号为

$$z(k) = \mathbf{W}^T \mathbf{x} = s_c(k) + n_c(k) \quad (10)$$

式中: $s_c(k) = \sum_{i=1}^N w_i a_i s(k)$; $n_c(k) = \sum_{i=1}^N w_i n_i(k)$ 。则合成信号信噪比为

$$\begin{aligned} \text{Ratio}_c &= \frac{|s_c(k)|^2}{\text{var}[n_c(k)]} = \\ &= \frac{|\sum_{i=1}^N w_i a_i s(k)|^2}{\text{var}[\sum_{i=1}^N w_i n_i(k)]} = \\ &= \frac{|\sum_{i=1}^N w_i a_i|^2}{\text{var}[\sum_{i=1}^N w_i n_i(k)]} \quad (11) \end{aligned}$$

当各路噪声不相关时, $\text{Ratio}_c = \frac{|\sum_{i=1}^N w_i a_i|^2}{\sum_{i=1}^N w_i \sigma_i^2}$ 。根据

柯西-施瓦兹 (Cauchy-Schwarz) 不等式有 $|\sum_{i=1}^N w_i a_i|^2 =$

$$\left| \sum_{i=1}^N (w_i \sigma_i) \frac{a_i}{\sigma_i} \right|^2 \leq \left(\sum_{i=1}^N w_i^2 \sigma_i^2 \right) \left(\sum_{i=1}^N \frac{a_i^2}{\sigma_i^2} \right), \text{ 当且仅当 } w_i \sigma_i =$$

$\frac{a_i}{\sigma_i}$ 时等式成立, 即最佳合成权值为 $w_{\text{iopt}} = \frac{a_i}{\sigma_i^2}$, 对应的

最佳合成信噪比为

$$\text{SNR}_{\max} = 10 \lg(\text{Ratio}_{\max}) = 10 \lg \left(\sum_{i=1}^N \frac{a_i^2}{\sigma_i^2} \right) \quad (12)$$

可以看出, 在独立噪声条件下, 合成权值仅与信号幅度和噪声功率有关。如当信号幅度均为 1、噪声方差为 1 (即各路等信噪比) 时, 5 副天线的理想合成增益为 $\text{SNR}_{\max} = 10 \lg 5 = 6.99 \text{ dB}$ 。

接下来讨论相关噪声条件下的合成增益。由于前文设 $n_i(k)$ 为第 i 路天线接收信号的零均值高斯白噪声, 方差为 σ_i^2 , 则各路噪声间协方差为 $\text{cov}(n_i,$

$n_j)$ 。各路噪声间存在的相关性用相关系数表示, 如

$$n_i(k) \text{ 和 } n_j(k) \text{ 的互相关系数为 } r_{ij} = \frac{\text{cov}(n_i, n_j)}{\sigma_i \sigma_j}.$$

设矩阵表示的各路噪声协方差为 $\mathbf{K}$, 则有

$$\mathbf{K} = E[\mathbf{n}(k) \mathbf{n}^H(k)] = \begin{bmatrix} \sigma_1^2 & r_{12} \sigma_1 \sigma_2 & \cdots & r_{1N} \sigma_1 \sigma_N \\ r_{21} \sigma_2 \sigma_1 & \sigma_2^2 & \cdots & r_{2N} \sigma_2 \sigma_N \\ \vdots & \vdots & \ddots & \vdots \\ r_{N1} \sigma_N \sigma_1 & r_{N2} \sigma_N \sigma_2 & \cdots & \sigma_N^2 \end{bmatrix} \quad (13)$$

对应地, 合成信号信噪比为修正为

$$\text{Ratio}_c = \frac{|\sum_{i=1}^N w_i a_i|^2}{\text{var}[\sum_{i=1}^N w_i n_i(k)]} = \frac{|\mathbf{W}^H \mathbf{a}|^2}{\mathbf{W}^H \mathbf{K} \mathbf{W}} \quad (14)$$

易知 $\mathbf{K}$ 为正定的 Hermitian 矩阵, 则可存在酉变换矩阵 $\mathbf{Q}$, 使得 $\mathbf{K} = \mathbf{Q}^H \mathbf{\Lambda}^2 \mathbf{Q}$, 其中 $\mathbf{\Lambda}$ 为 $N \times N$ 的对角阵, 同时 $\mathbf{Q} \mathbf{Q}^H = \mathbf{I}$, 即 $\mathbf{Q} = (\mathbf{Q}^H)^{-1}$, 则有 $\mathbf{\Lambda}^{-1} \mathbf{Q} = \mathbf{\Lambda}^{-1} (\mathbf{Q}^H)^{-1} = (\mathbf{Q}^H \mathbf{\Lambda})^{-1}$ 。代入式 (14) 中, 则其分子绝对值符号里可写为

$$\begin{aligned} \mathbf{W}^H \mathbf{a} &= \mathbf{W}^H \mathbf{I} \mathbf{a} = \mathbf{W}^H (\mathbf{Q}^H \mathbf{\Lambda}) (\mathbf{Q}^H \mathbf{\Lambda})^{-1} \mathbf{a} = \\ &= (\mathbf{W}^H \mathbf{Q}^H \mathbf{\Lambda}) (\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a}) = \\ &= (\mathbf{\Lambda} \mathbf{Q} \mathbf{W})^H (\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a}) \end{aligned} \quad (15)$$

分母可写为

$$\begin{aligned} \mathbf{W}^H \mathbf{K} \mathbf{W} &= (\mathbf{W}^H \mathbf{Q}^H \mathbf{\Lambda}) (\mathbf{\Lambda} \mathbf{Q} \mathbf{W}) = \\ &= (\mathbf{\Lambda} \mathbf{Q} \mathbf{W})^H (\mathbf{\Lambda} \mathbf{Q} \mathbf{W}) = \|\mathbf{\Lambda} \mathbf{Q} \mathbf{W}\|^2 \end{aligned} \quad (16)$$

由柯西-施瓦兹不等式有

$$\begin{aligned} \text{Ratio}_c &= \frac{|\mathbf{W}^H \mathbf{a}|^2}{\mathbf{W}^H \mathbf{K} \mathbf{W}} = \frac{|(\mathbf{\Lambda} \mathbf{Q} \mathbf{W})^H (\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a})|^2}{\|\mathbf{\Lambda} \mathbf{Q} \mathbf{W}\|^2} \leq \\ &= \frac{\|\mathbf{\Lambda} \mathbf{Q} \mathbf{W}\|^2 \|\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a}\|^2}{\|\mathbf{\Lambda} \mathbf{Q} \mathbf{W}\|^2} = \|\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a}\|^2 \end{aligned} \quad (17)$$

当且仅当 $\mathbf{\Lambda} \mathbf{Q} \mathbf{W} = c \mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a}$ (c 为一常数, 此处设为 1) 时等式成立, 即最佳合成权值为

$$\begin{aligned} \mathbf{W}_{\text{opt}} &= \mathbf{Q}^{-1} \mathbf{\Lambda}^{-2} \mathbf{Q} \mathbf{a} = \mathbf{Q}^{-1} \mathbf{\Lambda}^{-2} (\mathbf{Q}^H)^{-1} \mathbf{a} = \\ &= (\mathbf{Q}^H \mathbf{\Lambda}^2 \mathbf{Q})^{-1} \mathbf{a} = \mathbf{K}^{-1} \mathbf{a} \end{aligned} \quad (18)$$

对应的最佳合成信噪比为

$$\begin{aligned} \text{SNR}_{\max} &= 10 \lg(\text{Ratio}_{\max}) = \\ &= 10 \lg[(\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a})^H (\mathbf{\Lambda}^{-1} \mathbf{Q} \mathbf{a})] = \\ &= 10 \lg(\mathbf{a}^H \mathbf{K}^{-1} \mathbf{a}) \end{aligned} \quad (19)$$

可以看出, 与独立噪声条件下的最佳合成结构相比, 在相关噪声条件下的最佳合成多了 $\mathbf{K}^{-1}$, 其作用在于对各路噪声去相关, 而后再进行最大比合并。

实验所用的信号样本噪声相关性与距离、频率均存在密切关系, 在不同频率、不同距离处的相关系数均不同, 即 $\mathbf{K}$ 矩阵不同。从总体趋势来看, 距离

较远的节点,噪声相关性较小,并且噪声相关性随频率增加而下降。在衡量理论合成增益时,需要结合具体的频点考虑。同时易知,相关噪声条件下和独立噪声条件下的理论合成增益差即为分布式检测最佳理论性能提升值,即同等检测概率要求下单阵元信号信噪比差值。

3 系统方案

3.1 信号预处理

针对长波信号噪声统计特性,设计了对应的信号预处理模型。考虑实际中存在单阵元信号非平稳的情况,首先将各阵元信号进行短时傅里叶变换(Short-Time Fourier Transform, STFT)(由于 FFT 无法捕捉到信号在时域分布上的不同,在非平稳或非高斯情况下对整个序列做 FFT 会丢失时间信息,所以在实现时采用短时傅里叶变换进行处理)。特别需要注意的是,进行短时傅里叶变换加窗处理时会信号进行一定的重叠保留,从而保证窗信号与窗信号之间的时间连续性^[12]。

以接收天线 1 接收的信号 $x_1(n)$ 为例,假设在 STFT 中将 $x_1(n)$ 处理为 M 个分析窗,并依次计算各个分析窗信号对应的 N_F 点 FFT 值 $y_1^1, y_2^1, \dots, y_M^1$, 其中 $y_i^1 = [Y_{i1}^1 \ Y_{i2}^1 \ \dots \ Y_{iN_F}^1]$, $i=1, 2, \dots, M$, 表示第 i 个窗内的频域分量,即接收天线 1 的信号在第 i 个窗的 N_F 点 FFT 值,其大小为 $1 \times N_F$ 。则其时频变换后输出的大小为 $N_F \times M$ 的二维时频数据块如下:

$$\mathbf{Y}_l = [\mathbf{y}_1^l \ \mathbf{y}_2^l \ \dots \ \mathbf{y}_M^l] = \begin{bmatrix} Y_{11}^l & Y_{21}^l & \dots & Y_{M1}^l \\ Y_{12}^l & Y_{22}^l & \dots & Y_{M2}^l \\ \vdots & \vdots & \ddots & \vdots \\ Y_{1N_F}^l & Y_{2N_F}^l & \dots & Y_{MN_F}^l \end{bmatrix}_{N_F \times M} \quad (20)$$

由式(20)可知,得到的二维矩阵横轴为频域纵轴为时域, $\mathbf{Y}_l$ 的第 i 列 $\mathbf{y}_i^l$ 表示第 i 个时间窗在不同频点处的 FFT 数值,第 j 行则表示第 j 个频点在不同时间窗内的采样值。同理可得接收阵元 $y_2(t), y_3(t), \dots, y_L(t)$ 的时频数据块 $\mathbf{Y}_2, \mathbf{Y}_3, \dots, \mathbf{Y}_L$ 作为多阵元预处理后的数据,并视为特征值组同步输入神经网络。

3.2 网络预训练

考虑分布式小型磁感前端接收模型和智能检测系统方案如图 3 所示,分为网络预训练和网络部署两个阶段。其中,网络预训练阶段基于仿真信号数

据或接收的实际信号数据完成神经网络训练,网络部署阶段利用预训练得到的神经网络进行信号检测。

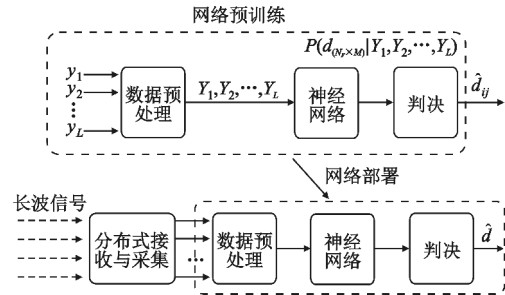


图 3 长波信号分布式智能检测方法及其系统方案

将时频数据块做一维展开后分为实数部分和复数部分同时输入作为数据集。以 L 个接收前端为例,网络输入数据集则为 $2L \times N_F \times M$ 的空、时、频三维数据。

该神经网络模块将信号检测问题简化为二元分类问题,输入训练集记为如下形式:

$$D = \{ (Y_{ij}^l, Y_{ij}^l, \dots, Y_{ij}^l, d_{ij}) \} \quad (21)$$

式中: Y_{ij}^l ($l=1, 2, \dots, L; i=1, 2, \dots, M; j=1, 2, \dots, N_F$) 是第 l 副天线的输入二维时频特征值表示; $d_{ij} \in \{0, 1\}$ 表示该样本点的真实分类标签,具体实现时编码为 One-Hot,即用向量 $[0, 1]^T$ 和 $[1, 0]^T$ 分别表示二元假设检测问题中的 H_0 和 H_1 两种假设,最后一层输出使用 Softmax 函数进行概率化输出。

神经网络映射输出归一化的预测后验概率 $\hat{p}(Y_{ij})$,表示分类标签 d_{ij} 为 1 时的预测后验概率值,用于预测目标信号在对应的时频维度上出现的概率,相当于判决是否有信号的置信度。基于神经网络的方法,其特征在于,输入信号模型覆盖的时间维、频率维、空间维度,以及输出信号的模型所描述的目标信号出现概率。该概率是以所有分布式接收前端样本为条件变量,覆盖所关注时间维、频率维度的全联合后验概率或其他数学上的等价量。

神经网络逻辑回归的损失函数采用交叉熵,旨在最大限度提高似然估计,使得在分类类别时的预测概率 $\hat{p}(Y_{ij})$ 与真实概率 $P(d_{ij} | Y_{ij}) \stackrel{\text{def}}{=} P(H_1 | Y)$ 之间的交叉熵损失最小,即训练至稳定输出两种假设的后验概率如下:

$$\begin{cases} H_1: P(H_1 | Y) = \hat{p}(Y_{ij}) \\ H_0: P(H_0 | Y) = 1 - \hat{p}(Y_{ij}) \end{cases} \quad (22)$$

本文用仿真及实采数据进行网络训练采用了长

短期记忆 (Long Short-Term Memory, LSTM) 递归神经网络、卷积神经网络 (Convolutional Neural Network, CNN)、极限学习机 (Extreme Learning Machine, ELM) 3 种网络, 网络结构如图 4 所示, 分别为由输入层、 K_1 个二维卷积层 (Convolution)、前后 2 个批量归一化层、前后 2 个 Relu 层、1 个二维最大池化层 (Pooling)、 K_2 个全连接层 (Fully Connected, FC)、1 个 Softmax 层和分类输出层连接构成的卷积神经网络; 由输入层、1 个 LSTM 层、1 个 Relu 层、 K 个全连接层、1 个 Softmax 层和分类输出层连接构成的 LSTM 网络, 梯度下降算法采用 Adam 算法, 最大迭代次数为 1 000, 初始学习率设为 0.01; 由输入层、隐藏层 (50 个神经元, 激活函数为 Sigmoid)、输出层构成的 ELM 网络。其中, CNN 网络和 LSTM 网络梯度下降算法采用 Adam 算法, 最大迭代次数为 1 000, 初始学习率设为 0.001, L2 正则化参数设为 0.000 1。

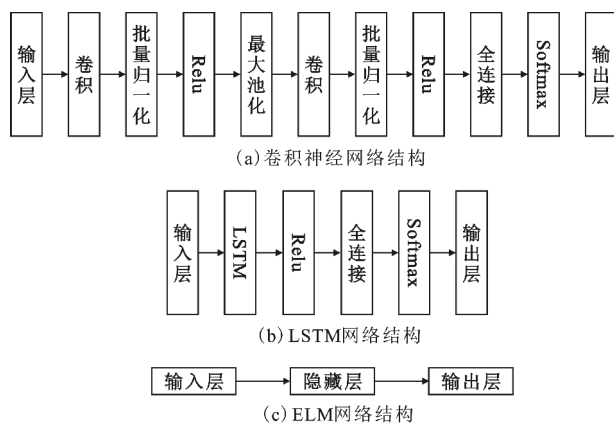


图 4 实验中采用的网络结构

L 个分布式前端输入 $2L$ 维时频数据可以视为 $2L$ 维特征, 利用神经网络有监督训练后, 判决模块根据神经网络输出的目标信号出现概率, 给出时间 m 和频点 n 处目标信号是否出现的判决结果。

3.3 网络部署

1) 分布式接收与采集

对接收的多路长波信号进行接收与数字化采集, 典型频段为 3~300 kHz, 典型调制方式为 2FSK。

2) 神经网络输入

通过时频变换, 并分实部虚部展开为 $2L$ 维矢量作为特征数据集, 和标签一起同时输入神经网络。

3) 输出信号检测结果

利用完成预训练的神经网络, 得到网络输出的目标信号出现概率, 进行判决得到信号检测结果。

4 仿真与实际信号验证

为了快速验证性能, 前期采用仿真信号对本文所提的智能检测方法进行仿真, 分为样本生成及预处理、检测网络两个步骤, 实现在已知噪声协方差的相关噪声信道环境下, 估计出目标频段内的信号分布信息。仿真 5 路并行数据, 数据集和标签生成流程如图 5 所示。首先设置接收机带宽范围内的信号数目 N , 再随机设置各个窄带信号的占用的频段范围及起止时间, 并标记作为标签。通过对原始波形数据进行预处理得到宽带时频数据作为数据集进行特征映射, 同时过瑞利信道并对宽带信号添加相关噪声, 再利用神经网络进行训练, 最终完成信号的检测。为了保证样本多样性, 生成的宽带多信号所涉及的调制方式包括 2FSK、4FSK、8FSK、BPSK、QPSK、OQPSK、8PSK、16APSK、16QAM、32APSK、64QAM (其中, MFSK 信号成形采用矩形成形, 幅相调制信号采用根升余弦成形)。

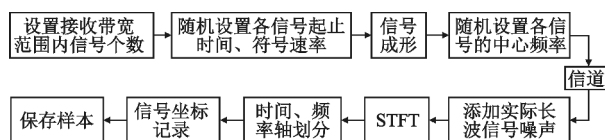
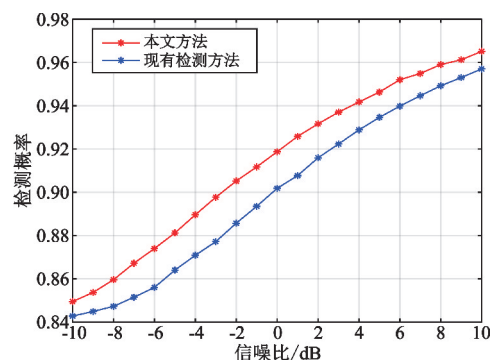


图 5 数据集和标签生成流程

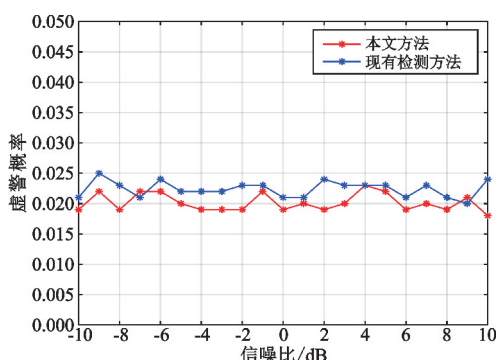
值得注意的是, 本仿真根据实际长波样本信号噪声相关这一特性, 仿真中截取实际长波信号样本中的噪声段进行添加, 并通过改变信号功率来改变信噪比。图 6 仿真所用的相关噪声的协方差矩阵 (中心频率为 55 kHz、带宽为 300 Hz 的噪声) 如下:

$$C_1 = \begin{bmatrix} 1 & 0.4515 & 0.4199 & 0.3159 & 0.4535 \\ 0.4515 & 1 & 0.4045 & 0.2645 & 0.3679 \\ 0.4199 & 0.4045 & 1 & 0.494 & 0.4965 \\ 0.3159 & 0.2645 & 0.494 & 1 & 0.49 \\ 0.4535 & 0.3679 & 0.4965 & 0.49 & 1 \end{bmatrix}$$

根据协方差矩阵易计算得理论最佳性能提升值为 2.789 dB。图 6 所示为本文方法和文献[3]所提方法仿真得到的检测概率与虚警概率随信噪比变化的曲线, 其中文献[3]所提的检测方法中融合中心判决时的似然比需要 H_1 和 H_0 的先验信息, 在仿真中通过计算真实标签的分布信息视为先验信息。从仿真结果可看出, 本文所提的智能联合检测算法具有更优的检测概率和虚警概率, 获得的性能提升最高约为 2.7 dB, 逼近于理论值。



(a) 检测概率曲线



(b) 虚警概率曲线

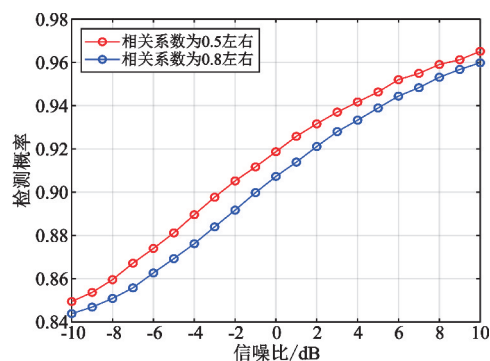
图6 不同检测方法下检测概率与虚警概率随信噪比变化曲线

再截取实际长波样本中相关系数更高的噪声段进行仿真(对应长波更低频段),结果如图7所示,具体的噪声协方差矩阵如下:

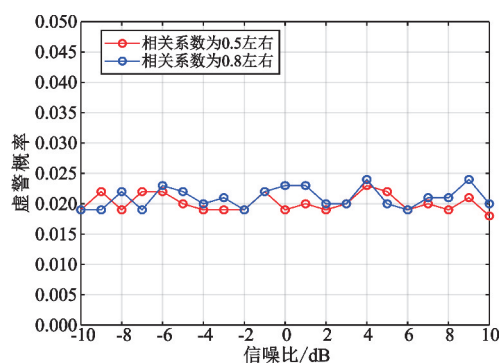
$$C_2 = \begin{bmatrix} 1 & 0.709\ 65 & 0.782\ 51 & 0.795\ 44 & 0.775\ 56 \\ 0.709\ 65 & 1 & 0.795\ 25 & 0.741\ 64 & 0.756\ 75 \\ 0.782\ 51 & 0.795\ 25 & 1 & 0.756\ 79 & 0.753\ 88 \\ 0.795\ 44 & 0.741\ 64 & 0.756\ 79 & 1 & 0.848\ 33 \\ 0.775\ 56 & 0.756\ 75 & 0.753\ 88 & 0.848\ 33 & 1 \end{bmatrix}$$

根据协方差矩阵易计算得理论最佳性能提升值为 0.887 7 dB,与相关系数为 0.5 左右的理论最佳性能提升值相差 1.901 3 dB。从图7可以看出,相同检测概率要求下,相关系数为 0.5 左右时比相关系数为 0.8 左右时性能提升约 2 dB,符合推导的理论值。

增加仿真路数仿真天线数目对检测性能的影响,分别仿真相关系数为 0.5 左右时 3 副天线和 7 副天线下的检测性能,结果如图8所示。相同检测概率要求下,天线数为 7 和天线数为 5 时比天线数为 3 时性能提升分别约 0.7 dB 和 0.2 dB,符合推导的理论值,同时天线数目越多虚警概率越低。

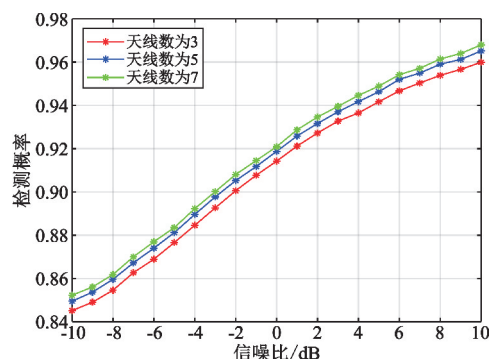


(a) 检测概率曲线

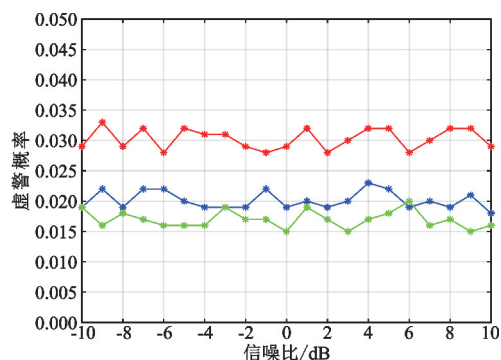


(b) 虚警概率曲线

图7 不同相关系数下检测概率与虚警概率随信噪比变化曲线



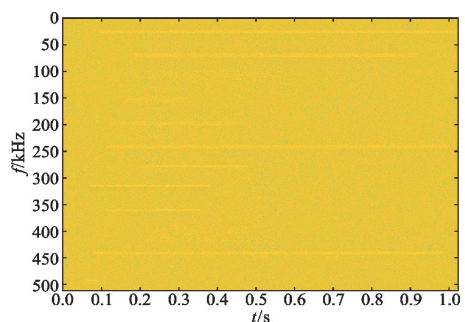
(a) 检测概率曲线



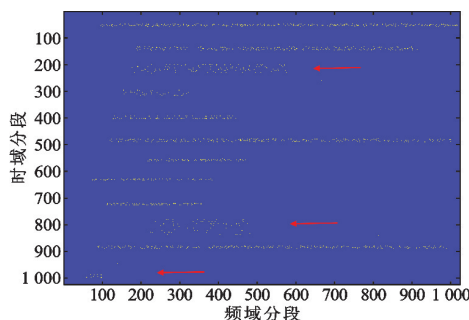
(b) 虚警概率曲线

图8 不同天线数目下检测概率与虚警概率随信噪比变化曲线

以信噪比为 -5 dB 为例,图 9(a)为单个前端接收到的信号时频图,图 9(b)为神经网络检测结果。可以看出神经网络通过分布式接收前端数据的盲融合,可以更为清晰地检测出目标带宽内微弱信号的分布(如图 9(b)中箭头所示的信号)。这表明该智能检测方法对噪声基底不敏感,在各种噪声环境下依旧能保持较好的检测鲁棒性。



(a) 单个前端接收到的信号时频图



(b) 神经网络检测结果

图 9 仿真宽带多信号检测结果

如前文所述,本文提出的智能检测方法用到的深度神经网络结构不唯一,因此,对比相关系数为 0.5 下采用的 3 种不同网络结构下的算法性能。图 10 所示为检测概率随信噪比变化的曲线,从 3 种结构对比来看,在同样仿真场景下,CNN 网络检测概率最高,ELM 检测概率最低。性能上,ELM 不需要基于梯度的反向传播来调整权重因此运算速度快,可以用于需要即时计算的场景中,但在复杂数据上 CNN、LSTM 等 DL 网络表现更优。

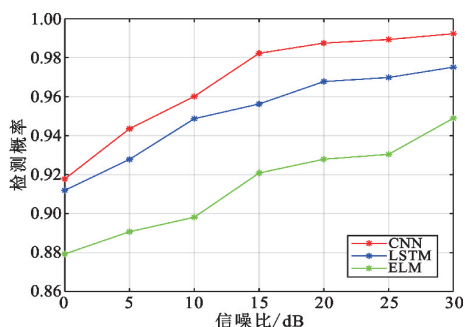
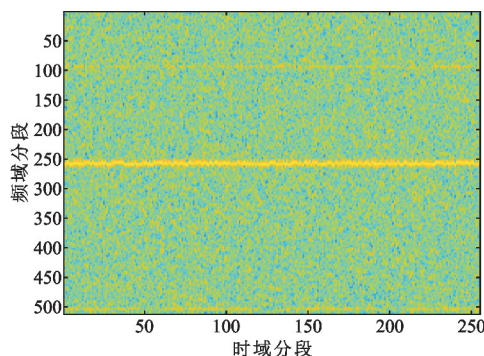


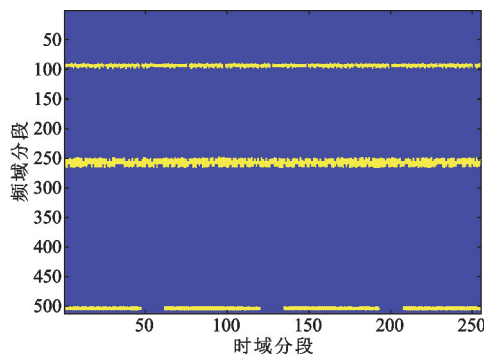
图 10 不同神经网络下的检测概率曲线

实地接收长波信号对算法进一步验证,实验所用的分布式接收前端为感应式磁场传感器,为中科院空天信息创新研究院布设于内蒙古的超短基线电磁探测阵列,有效测量频率范围为 $20 \sim 70$ kHz,包括 5 个分布在半径约 1 km 圆内的电磁探测传感器。实验采用多通道接收机将天线感应到的信号数字化并存储。

利用已知频点和带宽的长波电台信号作为参考信号进行网络训练,学习分布式小型磁感前端在特定噪声环境和系统阵列布局下的规律,如 68.5 kHz 信号(国家天文台设在河南商丘的 BPC 授时电台信号)^[13]等。这些电台的信息在网上公开可查。训练集和测试集的比例为 $4:1$,并对测试结果进行统计。对网络进行训练后求得稳定的神经网络参数并保存网络,并用于检测全频段的微弱信号。实验结果受训练集信号强度、天气等接收环境条件影响,检测概率没有仿真信号理想,但均能达到 90% 以上。以检测中心频率为 54 kHz、带宽为 6 kHz 的频段为例,图 11(a)为单个前端接收的信号时频图,图 11(b)为卷积神经网络的检测结果,验证了本文所提算法在实际应用中的有效性。



(a) 单个前端接收到的信号时频图



(b) 神经网络检测结果

图 11 实际宽带长波多信号检测结果

该方法基于参考信号进行训练,当参考信号强

度不足时,检测性能在一定程度上受限。在实际应用中,可以短暂时发参考信号,也可以利用已知的长波电台信号。同时,在实际应用中,随着观测时间增长,气候环境和天线状态发生变化,待检测信号和训练集中信号特性有明显差异时,该方法检测性能会随之下降。

5 结束语

针对小型前端组阵接收中噪声空间相关性影响检测性能的问题,本文基于联合最大后验准则,提出了一种基于神经网络的多天线长波信号智能检测算法。相比于现有传统建模方法,本文所提出的方法及系统有效利用了长波频段传播信道及分布式小型化前端接收环境所具有的潜在客观规律如分布式接收样本间的复杂统计相关性等,能够利用神经网络逼近分布式接收样本的全联合后验概率,而非依靠近似模型、独立性假设简化计算,从而在实际应用中具有更优的目标信号检测性能和系统鲁棒性。实际信号验证结果表明,相对于现有的基于独立同分布噪声背景的神经网络融合检测方法,算法可获得接近理论处理增益的检测性能,实现了基于数据的统计关联规律学习与挖掘。

后续将利用多套小型化前端进行更为机动灵活的布设,逐步开展进一步的实验系统建设与数据分析。同时,算法基于参考信号进行训练,若存在长时间观测需要进行大量数据采集时,需要不断将新接收的信号加入训练,这对算法的实时处理提出了较高的要求。后续将利用不同数据集进行训练,进一步研究所提算法对输入数据的适应性;考虑从信号合成方面对该方法进行改进,以提高算法检测性能。

参考文献:

- [1] 渠晓东,孙阳,陈冲,等.超短基线电磁脉冲阵在电磁辐射源测向中的应用[J].电子与信息学报,2019,41(4):830-836.
- [2] 崔勇,吴明,宋晓,等.小型低频发射天线的研究进展[J].物理学报,2020,69(20):165-177.
- [3] 张洪源,连政.长波通信的噪声处理研究[J].数字通信世界,2020(6):69.
- [4] 张凯,田瑶,董政.基于深度神经网络的多传感器联合信号检测方法[J].信号处理,2022,38(8):1749-1757.
- [5] 张凯,田瑶.最小错误准则下多传感器信号检测与调制识别[J].吉林大学学报(信息科学版),2023,41(3):387-395.
- [6] 苏宁远,陈小龙,关键,等.基于深度学习的海上目标一维序列信号目标检测方法[J].信号处理,2020,36(12):1987-1997.
- [7] 谷志刚,李红光,苏令华,等.基于强化学习特征选择的跳频调制方式识别[J].信号处理,2023,39(6):1089-1098.
- [8] 李杰,孙闽红,仇兆炀.时频域重叠多信号智能检测方法研究[J].信号处理,2021,37(5):878-884.
- [9] 蒲旭敏,孙致南,宋米雪,等.基于期望传播算法的多天线信号检测:架构、技术与挑战[J].电讯技术,2024,64(1):158-168.
- [10] KAY S M. Fundamentals of statistical signal processing (Volume II: detection theory) [M]. 罗鹏飞,张文明,刘忠,等译.北京:电子工业出版社,1998.
- [11] KISLIANSKY A, SHAVIT R, TABRIKIAN J. Direction of arrival estimation in the presence of noise coupling in antenna arrays [J]. IEEE Transactions on Antennas and Propagation, 2007, 55(7): 1940-1947.
- [12] 沈彩耀,于宏毅,胡赞鹏.频率选择性信道下的多天线信号分量合成算法[J].信息工程大学学报,2012,13(1):48-53.
- [13] 林自豪,刘天康,张宁.我国短波和长波授时信号的特征浅析[J].中国无线电,2023(3):48-51.

作者简介:

陈怀女,1994年生于广东汕头,2016年获学士学位,现为硕士研究生,主要研究方向为通信中的信号处理。

胡赞鹏男,1978年生于江西南昌,2006年获博士学位,现为副教授,主要研究方向为通信中的信号处理。

沈智翔男,1985年生于浙江湖州,2014年获博士学位,现为讲师,主要研究方向为通信中的信号处理。

李睿男,1990年生于山东枣庄,2016年获学士学位,现为助理工程师。

DOI:10.20079/j.issn.1001-893x.231227006

基于 Arnold 映射的抗分选混沌调制 PRI 设计方法^{*}

刘光霞¹, 李琦¹, 韩壮志², 魏英珍¹, 杨彦端¹

(1. 河北工业大学 电子信息工程学院, 天津 300401; 2. 陆军工程大学石家庄校区 电子与光学工程系, 石家庄 050003)

摘要:从主动反侦察角度出发,为了提高雷达信号在面临电子战侦察时的抗分选性能,在经典脉冲重复周期(Pulse Repetition Interval, PRI)工作形式上,提出了一种在信号脉冲序列前添加拼接干扰脉冲的方法。信号脉冲序列采用成组参差 PRI 形式,拼接干扰脉冲是基于 Arnold 映射的改进二维混沌二值化序列,该混沌序列具有良好的安全性和随机性,可提高分选的难度,破坏分选算法对雷达信号 PRI 的统计属性。通过 CDIF、SDIF 和 PRI 变换算法 3 个场景的仿真实验探究,和未添加拼接干扰脉冲对比,该方法使得 3 个场景的分选失败,不能分选出设定的 PRI 值,达到抗分选效果。

关键词:雷达信号抗分选; PRI 设计; 拼接干扰脉冲; 混沌调制; Arnold 映射

开放科学(资源服务)标识码(OSID):



中图分类号:TN974 文献标志码:A 文章编号:1001-893X(2025)03-0363-08

A Design Method of Anti-sorting Chaotic Modulation PRI Based on Arnold Mapping

LIU Guangxia¹, LI Qi¹, HAN Zhuangzhi², WEI Yingzhen¹, YANG Yanduan¹

(1. School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China;

2. Department of Electronic and Optical Engineering, Army Engineering University Shijiazhuang Campus, Shijiazhuang 050003, China)

Abstract: From the point of view of active anti-reconnaissance, in order to improve the anti-sorting performance of radar signal in the face of electronic warfare reconnaissance, a method of adding jamming pulses before signal pulse sequence is proposed in the form of pulse repetition interval (PRI). The signal pulse sequence adopts the form of group uneven PRI, and the splicing jamming pulse is an improved two-dimensional chaotic binarization sequence based on Arnold mapping. The chaotic sequence has good security and randomness, improves the difficulty of sorting, and destroys the statistical property of sorting algorithm to radar signal PRI. Through the simulation experiment of CDIF, SDIF and PRI transformation algorithms, and the comparison without adding splicing jamming pulse, the proposed method makes the separation of the three scenes fail, and the set PRI value can not be separated, and the anti-separation effect can be achieved.

Key words: radar signal anti-sorting; PRI design; splicing jamming pulse; chaotic modulation; Arnold mapping

0 引言

现代雷达在面临如干扰和反辐射攻击等电子对

抗威胁时,通常采用射频隐身技术降低雷达信号及其参数被截获和识别的概率,增强雷达的电子对抗

^{*} 收稿日期:2023-12-27;修回日期:2024-02-27

基金项目:基础加强计划技术领域基金(2020-JCJQ-JJ-376)

通信作者:李琦 Email:liqi@hebut.edu.cn

能力。一般来说,雷达告警接收机等实时宽带电子识别设备的检测过程通常包括截获、分选和识别 3 个阶段^[1-2]。为了应对电子侦察系统的威胁,主要从抗截获、抗分选和抗识别 3 个方面展开。其中,抗分选技术旨在使电子侦察系统无法从交错的脉冲流中分离出各个雷达脉冲序列,是增强雷达电子防御能力的重要途径^[3]。

为提高信号的抗分选性能,国内外学者开展了大量研究。徐梁昊等^[4]提出了一种具有规则性的重频组合干扰脉冲方法,破坏脉冲间的规律性,使得分选算法统计特性失效,达到抗分选的效果。Nan 等^[5]提出了一种利用干扰脉冲干扰到达时间序列的方法,改变雷达脉冲重复周期(Pulse Repetition Interval, PRI)的统计特性,有效地影响 PRI 分选过程,使信号处理器对雷达 PRI 方向图产生误判。戴胜波等^[6]分析了累积差值直方图(Cumulative Difference Histogram, CDIF)算法和 PRI 变换法的特点,设计了有规则的干扰脉冲。以上文献是通过添加干扰脉冲实现抗分选,但是干扰脉冲是具有一定规则的,设计过程复杂。张保群^[7]设计了满足一定约束条件的参差 PRI,使得分选过程遭到破坏,实现分选失败。余强等^[8]研究 ELINT 系统信号,针对其分选原理设计对应的干扰信号,通过仿真验证了干扰效果。汪飞等^[9]在没有添加干扰脉冲的情况下设计编队雷达的 PRI 特征,提高了编队雷达信号的抗分选性能。这些文献虽然实现了雷达的抗分选,但其分选具有局限性,不能使常见分选算法失效。

本文在经典 PRI 工作形式上,提出了一种在信号脉冲序列前添加拼接干扰脉冲的方法,选用的工作形式是成组参差 PRI。为了破坏 PRI 的规律性,利用在 Arnold 映射基础上设计的二维混沌序列调制添加的拼接干扰脉冲,从而提高拼接干扰脉冲的随机性和不确定性,实现信号的抗分选。

1 经典 PRI 工作形式

雷达信号分选,又称雷达信号去交错,是从随机交错的脉冲流中分离不同雷达脉冲序列的过程,如图 1 所示,通过适当分离各个雷达脉冲信号,使后续的信号处理工作能够高质量完成。因此,雷达信号分选技术对于电子战场态势感知至关重要,是影响后续决策的核心步骤。

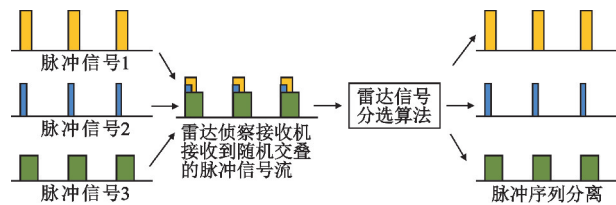


图 1 雷达辐射源分选示意

PRI 是指相邻发射脉冲前沿之间的时间间隔,对辐射源的性能具有重要影响,是进行脉冲信号分选的重要依据^[10]。早期的脉冲信号比较简单, PRI 值是固定的,即脉冲之间间隔是均匀的。现代脉冲信号为了实现抗分选和抗干扰,对 PRI 进行了多种形式改进,如图 2 所示,包括 PRI 固定、PRI 参差、PRI 抖动、PRI 滑变 4 种类型^[11],其表达式可用微信扫描本文 OSID 码,在“本文开放的科学数据与内容”中查看。

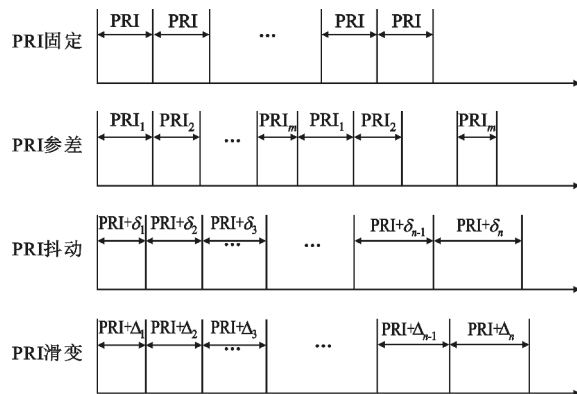


图 2 几种典型 PRI 工作形式脉冲序列

2 抗分选 PRI 设计

2.1 2D-FSBA 混沌映射

混沌系统是确定性系统中的一种不规则运动,具有随机性、遍历性以及分岔性等特征。经典一维混沌映射包括 Fuch 混沌映射、Sinusoidal 混沌映射等。Fuch 混沌映射^[12]和 Sinusoidal 混沌映射^[13]的表达式分别如(1)和(2)所示:

$$x_{n+1} = \cos\left(\frac{1}{x_n^2}\right), x_n \in (-1, 1), x_n \neq 0 \quad (1)$$

$$x_{n+1} = \lambda x_n^2 \sin(\pi x_n), x_n \in [0, 1] \quad (2)$$

式(1)中: x_n 是混沌输入, x_{n+1} 是混沌输出, $n = 1, 2, 3, \dots$ 为混沌序列中各序列值序号,在初值不为 0 的情形下均能产生混沌。式(2)中: x_n 是混沌输入, x_{n+1} 是混沌输出; λ 为 Sinusoidal 混沌映射的控制参

数,当 $\lambda = 2.3$ 时系统处于混沌状态。

Arnold 映射^[14-15]涉及元素在水平和垂直方向上的拉伸和折叠,其中拉伸是通过与拉伸矩阵相乘使位置被拉长,折叠则是通过取模过程使其回到原始区域。这是一个可逆的二维混沌映射,其迭代形式如下所示:

$$\begin{cases} x_{n+1} = (x_n + y_n) \bmod 1 \\ y_{n+1} = (x_n + 2y_n) \bmod 1 \end{cases} \quad (3)$$

Fuch 混沌和 Sinusoidal 混沌均为一维混沌映射,具有行为简单、混沌区间脆弱等弱点。受输出函数性质的影响,映射的随机数结果不是均匀分布的,即伪随机性差。为了解决这一问题,本文结合以上两种经典映射,提出一种基于 Arnold 映射的二维混沌映射 (Two-dimensional Fuch and Sinusoidal Based on Arnold, 2D-FSBA),表达式如下:

$$\begin{cases} x_{n+1} = \left(1 - ax_n^2 \sin(\pi x_n) + b \cos\left(\frac{1}{y_n^2}\right) \right) \bmod 1 \\ y_{n+1} = \left(a \cos\left(\frac{1}{x_{n+1}^2}\right) + by_n^2 \sin(\pi y_n) \right) \bmod 1 \end{cases} \quad (4)$$

式中: a 和 b 是混沌参数; $x_n \in (0, 1)$ 和 $y_n \in (0, 1)$ 是混沌输入; x_{n+1} 和 y_{n+1} 是混沌输出。混沌设计过程如图 3 所示。

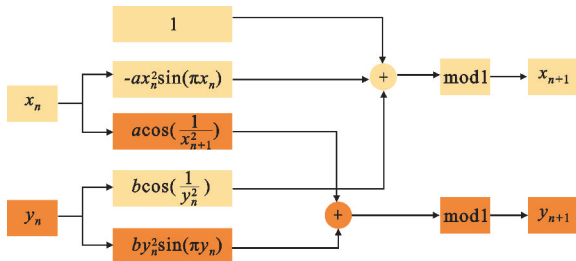


图 3 2D-FSBA 混沌系统设计流程

对 2D-FSBA 混沌映射输出序列实现二值量化,定义二值化结果为

$$z_n = \begin{cases} 0, & x_n > y_n \\ 1, & x_n \leq y_n \end{cases} \quad (5)$$

式中: z_n 为混沌序列的二值化结果,用于调制拼接干扰脉冲。

混沌调制存在安全问题,攻击者可以在不知道混沌参数甚至不知道混沌系统的情况下破解混沌信号^[16]。为了解决该问题,本文提出一种基于 Arnold 映射的混沌调制方案。该方案利用 Arnold 映射来改变混沌序列的次序,从而有效地抵御各种攻击方法,如回归映射和功耗分析等^[17]。

2.2 PRI 添加拼接干扰脉冲

雷达信号分选依赖于脉冲描述字和脉冲重复间隔。射频隐身技术中的抗分选方法可以通过调整信号参数扰乱截获接收机的分选过程,实现信号的射频隐身,主要方法包括^[18]:引入混沌技术,增强信号参数的随机性;加入扰乱脉冲,破坏脉冲间的规律性。受目前抗分选信号设计启发,本文针对成组参差 PRI 提出了一种在被保护雷达信号前拼接干扰脉冲的方法,用于干扰被保护脉冲到达时间序列,从而改变雷达信号原始 PRI 模式,破坏被保护脉冲到达时间的序列结构。信号脉冲序列结构如图 4 所示,其中被保护脉冲的脉宽固定,拼接干扰脉冲是通过 2D-FSBA 混沌映射二值量化值控制,由公式(5)可知,二值化值会随机生成 0 或 1,信号的每个被保护脉冲都会拼接随机生成的序列,所以拼接干扰脉冲会时有时无,提高了信号的随机性。被保护脉冲脉宽远大于拼接干扰脉冲脉宽,具体设定会在仿真分析中指出。

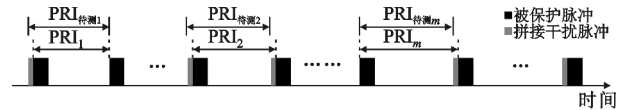


图 4 被保护脉冲添加拼接干扰脉冲时序示意

成组参差 PRI 表达式如下:

$$\text{PRI}_i = \begin{cases} \text{PRI}_1, & KM \leq i \leq KM + N_1 \\ \vdots \\ \text{PRI}_m, & KM + \sum_{j=1}^{m-1} N_j \leq i \leq KM + \sum_{j=1}^m N_j \end{cases} \quad (6)$$

式中: $K=0, 1, 2, \dots, n$; $M = \sum_{j=1}^m N_j$, M 是发出脉冲数总和,且每发出 M 个脉冲 PRI_i 重复一次;以 $\text{PRI}_1 \sim \text{PRI}_m$ 为周期分别发送 $N_1 \sim N_j$ 个脉冲。

本文采用的拼接干扰脉冲是通过 2D-FSBA 混沌映射二值化序列。不同于常见的二维混沌映射,该混沌具有更好的随机性和不确定性,使得 PRI 调制后的信号更加难以被分选出。采用此方法实现射频隐身信号的抗分选性能,添加拼接干扰脉冲公式如下:

$$\text{PRI}_i = \begin{cases} z_n + \text{PRI}_1, & KM \leq i \leq KM + N_1 \\ \vdots \\ z_n + \text{PRI}_m, & KM + \sum_{j=1}^{m-1} N_j \leq i \leq KM + \sum_{j=1}^m N_j \end{cases} \quad (7)$$

式中: $K=0, 1, 2, \dots, n$; $M = \sum_{j=1}^m N_j$; 脉冲序列中每个

PRI 值前添加的拼接干扰脉冲是随机序列 z_n 。本文主要研究雷达发射端的信号设计,对于接收方来说,由于设计的混沌序列对我方来说是已知的,可将接收到的信号脉冲反向去掉拼接干扰脉冲序列,恢复到最初的雷达信号。

3 仿真验证

3.1 2D-FSBA 混沌映射性能分析

3.1.1 平衡度

平衡度用于判断序列的随机性是否良好,通过比较序列中 0 和 1 的个数^[19]。一个理想的混沌序列应该具有数量相等的 0 和 1。如果平衡度较小,说明混沌序列的随机性较好;反之,随机性较差。设混沌序列长度为 N ,其中 0 和 1 的个数分别为 P 和 Q ,平衡度可以通过以下公式来计算:

$$E(N)=\frac{|P-Q|}{N} \tag{8}$$

图 5 展示了 2D-FSBA 混沌平衡度随序列长度变化的实验结果,生成的混沌序列长度从 0 增加到

10 000,平衡值均小于 0.002,说明序列中 0 和 1 的个数十分接近,序列有更好的随机性。

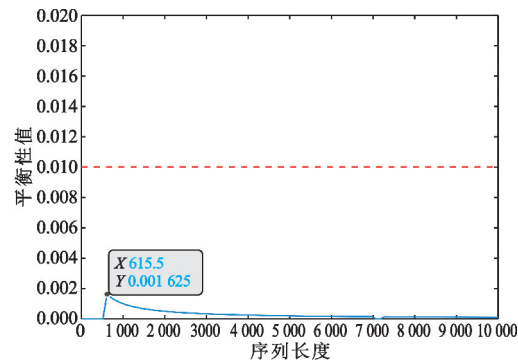


图 5 2D-FSBA 混沌平衡性

3.1.2 相空间

一个好的混沌映射应具备良好的随机性和遍历性。相空间^[20]是用状态变量表示系统运动的方法,其中每个点的位置反映了系统的状态。因此,通过观察混沌映射的相空间图中点的分布情况,可以评估混沌系统的随机性和遍历性。图 6 是 2D-FSBA 混沌和经典二维 Henon 混沌相空间图对比结果。

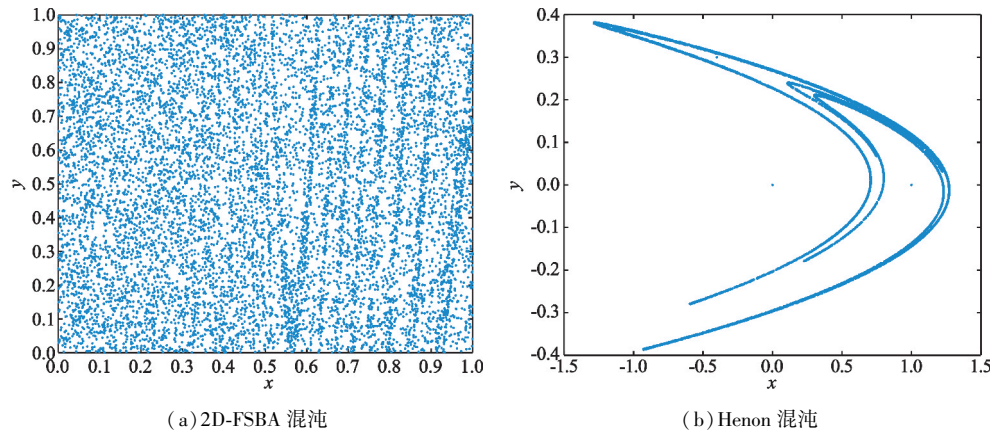


图 6 相空间图

从图 6 可看出, Henon 混沌的相图分布具有明显的形状特征且分布较为稀疏,对比还可看出 2D-FSBA 混沌的相空间图分布杂乱无章,密度较大,不具有明显的形状特征。

3.1.3 游程统计

游程是指在随机序列中连续出现 0 或 1 的子序列^[21]。游程长度是连续出现的 0 或 1 个数。在随机序列中,长度为 L 的游程为游程总数的 $1/2^L$ 。该方法验证序列中 0 和 1 的个数是否接近,当个数越接近,说明序列的平衡性越好,测试值和理论值接近一致。不平衡的序列将会出现稳定信号,不利于信号的隐身。表 1 统计了序列长度为 10 000 时,2D-

FSBA 混沌序列在 1~8 游程长度中测试值与理论值的比较。

表 1 2D-FSBA 混沌序列游程统计					
游程长度	0 的个数	1 的个数	0、1 个数比值	测试值	理论值
1	1 217	1 251	0.972 82	0.492 1	0.500 0
2	646	662	0.975 83	0.260 8	0.250 0
3	341	310	1.100 00	0.129 8	0.125 0
4	151	167	0.904 19	0.063 4	0.062 5
5	80	62	1.290 32	0.028 3	0.031 3
6	40	39	1.025 64	0.015 8	0.015 6
7	19	12	1.583 33	0.006 2	0.007 8
8	9	9	1.000 00	0.003 6	0.003 9

从表 1 可看出,不同长度的游程中,游程 1 和 0 的个数非常接近,测试值和理论值基本一致,说明 2D-FSBA 混沌产生的序列具有良好的游程特性。综上,2D-FSBA 混沌映射反映出很强的随机性和遍历性,适用于随机调制。

3.1.4 排列熵

排列熵(Permutation Entropy, PE)用于衡量离散序列的随机程度,越规则的离散序列对应的值越小,越复杂的离散序列对应的值越大。因此,PE 可以用于检测和分析混沌系统,以评估它们的复杂性和不规则性程度。2D-FSBA 混沌映射与传统一维混沌映射的 PE 值对比如图 7 所示。

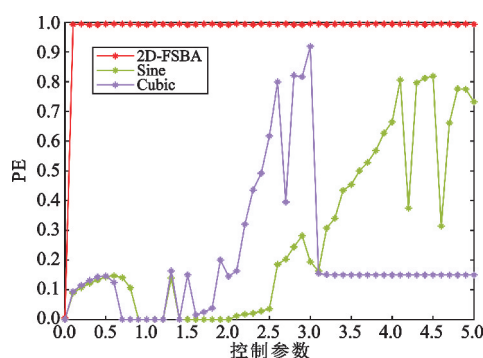


图 7 3 种混沌的 PE 值

从图 7 可看出,在参数范围内,本文所提 2D-FSBA 混沌映射的 PE 值比传统一维 Sine 和 Cubic 混沌映射 PE 值更大,且一直处于平稳状态值,说明 2D-FSBA 混沌可以产生更随机的混沌序列。

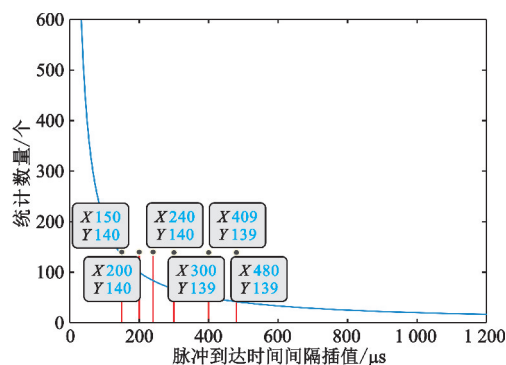
3.2 抗分选结果验证

经典的重频分选方法包括直方图法和 PRI 变换法,具有处理速度快且易于实现的优点^[22]。分选原理是统计脉冲到达时间差的频数或频谱值,当超过设定的门限值时,基于周期脉冲信号相关特性,便可以确定对应的 PRI 脉冲序列来自同一雷达。本文利用直方图法、PRI 变换法对成组参差 PRI 和本文设计的成组参差 PRI 拼接干扰脉冲进行分析,设定雷达辐射源发射的 PRI 有两组,分别为第 1 组(150 μs 、200 μs 、240 μs)和第 2 组(180 μs 、210 μs 、250 μs)。

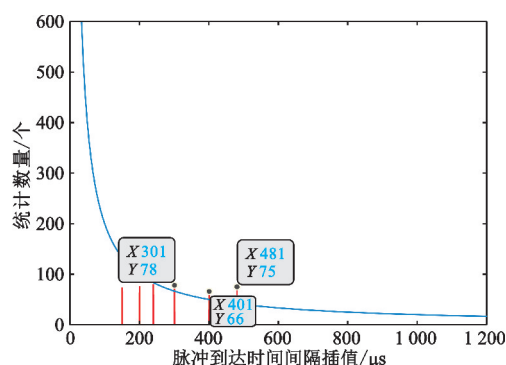
3.2.1 CDIF 分选

CDIF 算法^[23]是一种基于传统直方图统计算法的改进版,它能够进一步降低时间复杂度,提高雷达信号分选的实时性,解决传统直方图统计算法中存在的谐波问题。CDIF 算法检测门限 $T_{\text{thr}}(\tau) = \alpha T_s / \tau$, T_s 为采样时间, τ 为脉冲到达时间间隔, α 是

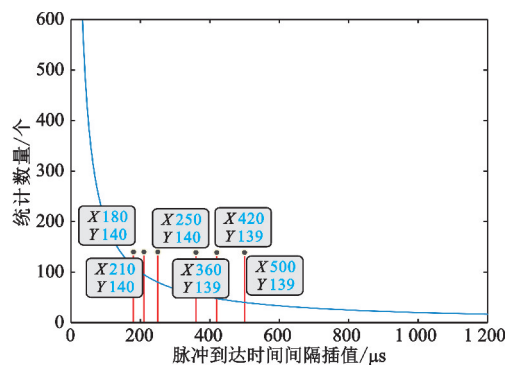
可调系数,通常设为 0.2。图 8 是两组 PRI 添加拼接干扰脉冲前后 CDIF 分选结果。



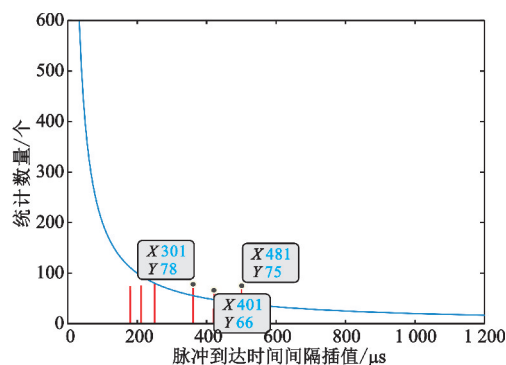
(a) 第 1 组-成组参差 PRI



(b) 第 1 组-成组参差 PRI 添加拼接脉冲



(c) 第 2 组-成组参差 PRI



(d) 第 2 组-成组参差 PRI 添加拼接脉冲

图 8 CDIF 算法分选结果

从图 8(a)、(c) 两图可看出,采用 CDIF 算法进行信号分选时,直方图中 PRI 值与其二倍值对应的峰值都超过了检测门限,成组参差 PRI 值可以全部被分选出来;但添加混沌调制的拼接脉冲后,从图 8(b)、(d) 可看出,出现真实 PRI 对应的峰值没有超过检测门限, PRI 值不能被分选出来,达成了抗分选的目的。

3.2.2 SDIF 分选

序列差值直方图 (Sequence Difference

Histogram, SDIF) 算法^[24]基本原理与 CDIF 相似,也是通过对脉冲雷达信号到达时间进行直方图统计,并根据检测门限来确定潜在的 PRI。SDIF 统计门限 $T_{thr}(\tau) = x(E - C)e^{-\tau/kN}$, E 为脉冲总数, N 为直方图内总单元数, C 是差级数, 常数 x 和 k 的最佳值由实验确定。本实验设定 $N = 2\,000$, $k = 0.15$, $x = 0.4$, 图 9 是两组 PRI 添加拼接干扰脉冲前后 SDIF 分选结果。

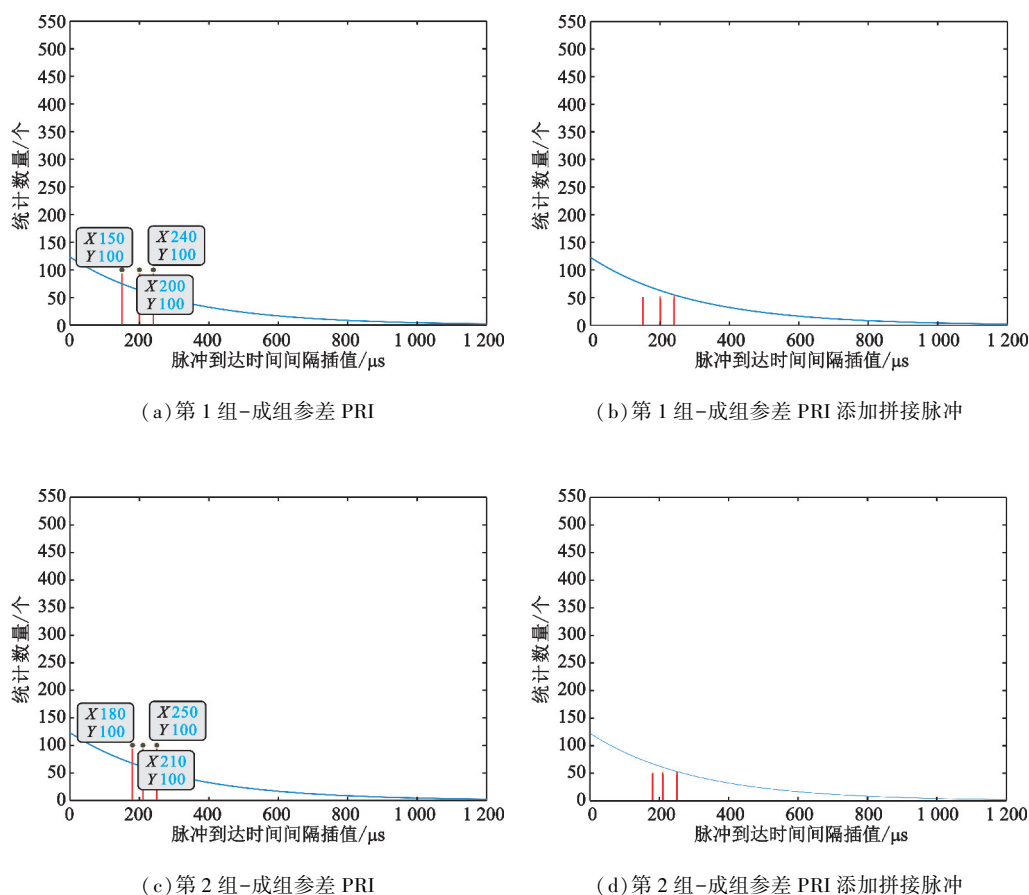


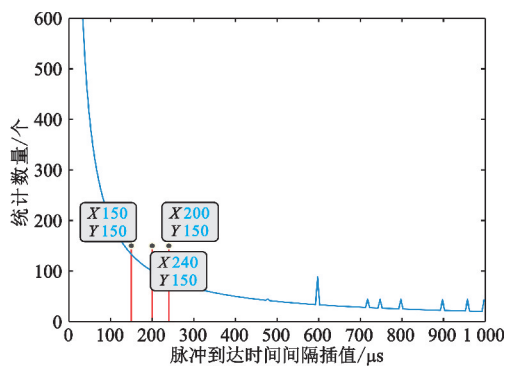
图 9 SDIF 算法分选结果

当采用 SDIF 分选算法分选时,图 9(a)、(c) 是分选出的成组参 PRI 结果,设定的两组成组参差 PRI 值完全超过检测门限的峰值,全部被分选出来;但添加混沌调制的拼接脉冲后,图 9(b)、(d) 是分选结果,说明 PRI 值不能被分选出来,实现了 PRI 的抗分选。

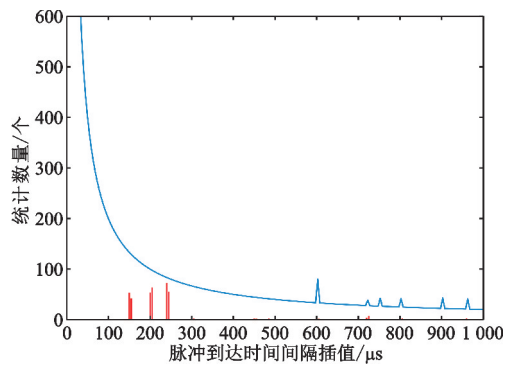
3.2.3 PRI 变换分选

PRI 变换法^[25]不同于直方图法,其核心是对

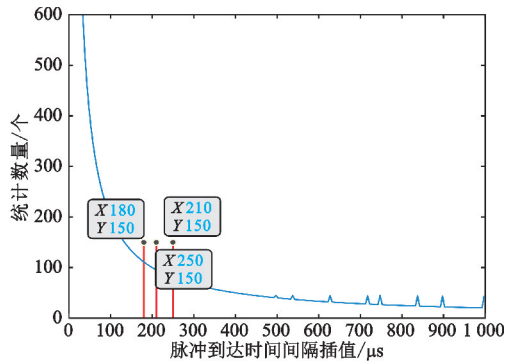
PRI 变换后的雷达脉冲序列求得 PRI 谱图,随后提取谱图中高于门限的谱峰作为参考 PRI 值,并利用这些 PRI 值进行序列检索,以获取对应的雷达信号。PRI 变换在雷达脉冲单位冲激响应自相关函数的基础上引入了一个相位因子 $e^{\frac{2\pi j t}{\tau}}$,能够在很大程度上抑制子谐波的产生并准确估计真实 PRI 值。图 10 是两组 PRI 添加拼接干扰脉冲前后 PRI 变换分选结果。



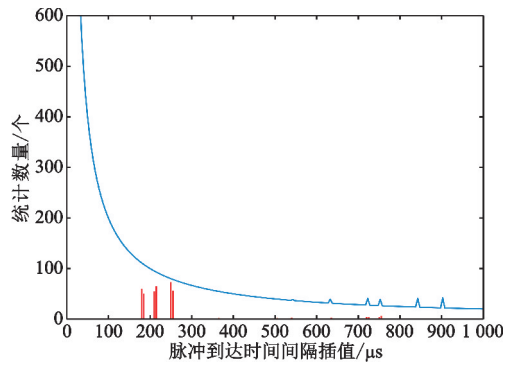
(a) 第 1 组-成组参差 PRI



(b) 第 1 组-成组参差 PRI 添加拼接脉冲



(c) 第 2 组-成组参差 PRI



(d) 第 2 组-成组参差 PRI 添加拼接脉冲

图 10 PRI 变换算法分选结果

图 10(a)、(c)说明,设定的两组成组参差 PRI 值均超过了门限函数,全部被分选出来;但添加混沌调制的拼接脉冲后,图 10(b)、(d)说明,PRI 值不能被分选出来,说明添加混沌调制拼接脉冲的 PRI 具有抗分选性能。

表 2 总结了成组参差 PRI 和成组参差 PRI 调制的分选结果。相比于添加拼接干扰脉冲,成组参差 PRI 完全被分选出来,但添加混沌调制后可以使 3 种分选算法失效,实现信号的抗分选。

表 2 3 种分选算法分选结果

算法	能否检测 PRI	
	成组参差 PRI	添加混沌调制拼接脉冲的成组参差 PRI
CDIF 算法(阶数为 2)	能	否
SDIF 算法	能	否
PRI 变换算法	能	否

4 结 论

本文提出了一种添加拼接干扰脉冲的方法,设计了一种随机性很强的 2D-FSBA 混沌映射调制拼接干扰脉冲,一方面提高了信号抗分选能力;另一方面,在己方对回波进行处理时,由于混沌调制序列对我方来说是已知的,可以快速进行信号的识别与处理。通过 CDIF、SDIF 和 PRI 变换 3 种分选算法与本文所提方法进行仿真分析,设计两组具有 3 种 PRI 值的序列,相比于原始脉冲,添加拼接干扰脉冲后的序列完全不能被分选出设定的 PRI 值。因此,本文提出的方法具有较强的抗分选能力,可以用于对抗经典的分选算法,具有工程实际意义。

参考文献:

[1] 贾金伟,刘利民,韩壮志,等. 低截获雷达抗分选信号设计技术研究综述[J]. 电讯技术,2023,63(5):748-756.

[2] 杨宇晓,汪德鑫,黄琪. 四维超混沌射频隐身跳频通信设计方法[J]. 宇航学报,2020,41(10):1341-1349.

[3] 赵子杭. MIMO 雷达射频隐身信号设计及性能分析[D]. 天津:河北工业大学,2023.

[4] 徐梁昊,姜秋喜,潘继飞,等. 一种抗重频分选的反侦察方法[J]. 四川兵工学报,2015,36(7):117-120.

[5] NAN H, PENG S R, YU J A, et al. Pulse interference method against PRI sorting [J]. The Journal of Engineering, 2019, 2019(19): 5732-5735.

[6] 戴胜波,雷武虎,程艺喆,等. 基于 TOA 分选的反电子侦察方法[J]. 电子信息对抗技术,2014,29(4):45-48.

[7] 张保群. 一种抗 SDIF 分选的脉冲重复间隔参差设计

- 方法[J]. 兵器装备工程学报, 2016, 37(9): 87-91.
- [8] 余强, 毕大平, 陈璐, 等. 对 ELINT 系统基于信号 PRI 参数分选的干扰技术[J]. 火力与指挥控制, 2016, 41(1): 143-147.
- [9] 汪飞, 刘建锋. 编队雷达的脉冲重复间隔低分选设计[J]. 信息技术, 2019, 43(3): 14-18, 23.
- [10] 石荣, 吴聪. 基于 PRI 信息的雷达脉冲信号分选技术研究综述[J]. 电讯技术, 2020, 60(1): 112-120.
- [11] 王璐璐. 雷达信号分选技术研究[D]. 哈尔滨: 哈尔滨工程大学, 2021.
- [12] 孙文捷, 张惠珍, 张健, 等. 基于 Fuch 映射的混沌蝙蝠算法[J]. 上海理工大学学报, 2014, 36(1): 26-30.
- [13] JITEURTRAGOOL N, KETTHONG P, WANNABOON C, et al. A topologically simple keyed hash function based on circular chaotic sinusoidal map network[C]//2013 15th International Conference on Advanced Communications Technology. PyeongChang: IEEE, 2013: 1089-1094.
- [14] HUANG X L, DONG Y X, JIAO K X, et al. Asymmetric pixel confusion algorithm for images based on RSA and Arnold transform [J]. Frontiers of Information Technology & Electronic Engineering, 2020, 21(12): 1783-1794.
- [15] YE H M, HUANG S M, LIU W J. Research on image scrambling method based on combination of Arnold transform and exclusive-or operation [C]//2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference. Chongqing: IEEE, 2020: 151-154.
- [16] ÁLVAREZ G, MONTROYA F, ROMERA M, et al. Breaking parameter modulated chaotic secure communication system[J]. Chaos, Solitons & Fractals, 2004, 21(4): 783-787.
- [17] LI S J, CHEN G R, ÁLVAREZ G. Return-map cryptanalysis revisited [J]. International Journal of Bifurcation and Chaos, 2006, 16(5): 1557-1568.
- [18] 贾金伟, 韩壮志, 刘利民, 等. 基于 SDIF 门限失效的雷达射频隐身信号设计原理[J]. 系统工程与电子技术, 2023, 45(6): 1693-1701.
- [19] 周艳琦. 基于人工蜂群算法的混沌参数优化及图像加密算法研究[D]. 哈尔滨: 黑龙江大学, 2022.
- [20] SHARMA M, BHARTI V. Image encryption using a new 2D bit reversed logistic map [C]//2020 11th International Conference on Computing, Communication and Networking Technologies. Kharagpur: IEEE, 2020: 1-6.
- [21] 李涵. 基于混沌系统的图像加密算法研究[D]. 淮南: 安徽理工大学, 2022.
- [22] MANICKCHAND K, STRYDOM J J, MISHRA A K. Comparative study of TOA based emitter deinterleaving and tracking algorithms [C]//2017 IEEE AFRICON. Cape Town: IEEE, 2017: 221-226.
- [23] HE Y H, SANG X, LI Y G, et al. PRI-based radar signal binning algorithm and simulation implementation [C]//2023 IEEE 7th Information Technology and Mechatronics Engineering Conference. Chongqing: IEEE, 2023: 706-710.
- [24] HAN Z Z, ZHAO Z H, GAO Z B, et al. Radar signal sorting algorithm based on pulse repetition interval [C]//2020 5th International Conference on Electromechanical Control Technology and Transportation. Nanchang: IEEE, 2020: 163-166.
- [25] 隋金坪, 刘振, 刘丽, 等. 雷达辐射源信号分选研究进展[J]. 雷达学报, 2022, 11(3): 418-433.

作者简介:

刘光霞 女, 1997 年生于河北邢台, 2021 年获学士学位, 现为硕士研究生, 主要研究方向为雷达射频隐身信号设计。

李琦 男, 1974 年生于山西晋城, 2009 年获博士学位, 现为教授, 主要研究方向为雷达信号设计。

韩壮志 男, 1972 年生于河北石家庄, 2003 年获博士学位, 现为副教授, 主要研究方向为电子对抗、数字信号处理。

魏英珍 女, 1999 年生于河北邢台, 2021 年获学士学位, 现为硕士研究生, 主要研究方向为雷达信号处理。

杨彦端 男, 2000 年生于河北邯郸, 2022 年获学士学位, 现为硕士研究生, 主要研究方向为雷达射频隐身信号设计。

DOI:10.20079/j.issn.1001-893x.240229005

基于 UNet 结构的大规模 MIMO 系统 CSI 反馈设计*

刘 庆¹, 李 义¹, 李 璘¹, 王有军¹, 李 康¹, 熊林麟¹, 王 平²

(1. 贵州电网有限责任公司毕节供电局, 贵州 毕节 551700; 2. 重庆邮电大学 通信与信息工程学院, 重庆 400065)

摘 要:从用户端获取下行信道状态信息(Channel State Information, CSI)是频分双工(Frequency Division Duplex, FDD)模式下通信系统信息高效传输的关键,然而其反馈开销随着天线规模的增加而增大,给大规模多输入多输出(Multiple-Input Multiple-Output, MIMO)系统带来了重大挑战。针对此问题,提出了一种基于连续卷积和注意力机制的 CSI 反馈网络结构 U-shaped Transformer Neural Network(UTNet)。首先,编码器和解码器分别采取编码与压缩同步、解码和重建同步的连续采样结构,实现特征提取和压缩。其次,在编码器的末端和解码器的开端分别插入 Transformer 模块,提取不同位置之间的关联信息。最后,通过调节 CSI 反馈网络参数实现对发送数据长度的控制,旨在实现 CSI 信号更加智能和高效的反馈。实验结果表明,在不同压缩率下 UTNet 的归一化均方误差(Normalized Mean Square Error, NMSE)低于-27.45 dB,相较于现有基于深度学习的方法,UTNet 能在保持更高精度的同时反馈开销更小。

关键词:大规模 MIMO; CSI 反馈; 深度学习; Transformer 模块

开放科学(资源服务)标识码(OSID):



中图分类号:TN929.5 文献标志码:A 文章编号:1001-893X(2025)03-0371-07

Design of UNet-based CSI Feedback in Massive MIMO System

LIU Qing¹, LI Yi, LI Lin, WANG Youjun, LI Kang, XIONG Linlin, WANG Ping²

(1. Bijie Power Supply Bureau, Guizhou Power Grid Corporation, Bijie 551700, China;

(2. School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: The downlink channel state information (CSI) obtained from users is the key to efficient transmission of communication system information under frequency division duplex (FDD) mode. However, the feedback overhead increases with the antenna scale, which poses significant challenges to massive multiple-input multiple-output (MIMO) systems. To address this issue, a U-shaped Transformer neural network (UTNet) structure based on continuous convolution and attention mechanism is proposed for CSI feedback. First, the encoder and decoder adopt a continuous sampling structure of encoding and compression synchronization, decoding and reconstruction synchronization, to achieve feature extraction and compression. Second, Transformer modules are inserted at the end of the encoder and the beginning of the decoder to extract correlation information between different positions. Finally, the CSI feedback network parameters are adjusted to achieve control of the length of the transmitted data, aiming to achieve more intelligent and efficient feedback of CSI signals. The experimental results show that under different compression ratios, the normalized mean square error (NMSE) of UTNet is lower than -27.45 dB, indicating that UTNet can maintain higher accuracy while having a smaller feedback overhead compared with existing deep learning-based methods.

Key words: massive MIMO; CSI feedback; deep learning; Transformer module

* 收稿日期:2024-02-29;修回日期:2024-05-11

基金项目:2022 年贵州电网有限责任公司毕节供电局电力技术开发项目(0607002022030101SC00049)

通信作者:刘庆 Email:271070837@qq.com

0 引言

为应对未来移动互联网和物联网大量智能设备的接入,学术界和工业界提出了第六代移动通信技术^[1]。此外,已有文献证明多输入多输出(Multiple-Input Multiple-Output, MIMO)系统通过在基站(Base Station, BS)端部署大规模天线阵列可获得更多的天线自由度和更高的波束增益,从而能够显著提升频谱资源利用度和通信效率^[2]。为了满足第六代移动通信技术的需求,大规模 MIMO 系统的性能应具有更高的能效比、更强的抗干扰能力、更多的天线数量和更广泛的应用场景,因而获取更高精度的下行信道状态信息(Channel State Information, CSI)的研究变得愈发迫切。

在频分双工(Frequency Division Duplex, FDD)模式下,下行链路的 CSI 需要从用户端通过下行导频获取再反馈,但天线规模的增加会导致 CSI 的复杂度和信息量呈指数级增长,在频繁更新 CSI 以实现波束赋形、自适应调制等功能时,CSI 反馈开销会急剧增大。因此,如何在降低 CSI 反馈开销的同时保持重建的 CSI 信号的准确性,已成为当前学术界研究的难点。目前,在 FDD 模式下的 CSI 反馈设计通常采用基于码本设计^[3]和压缩感知^[4-6]的方法优化。然而,基于码本设计的 CSI 反馈方案设计复杂度和反馈量随天线数目增加而显著增长,基于压缩感知的反馈方案存在信道的严格稀疏表示难以获取且不适合低时延场景的缺点。

近年来,基于深度学习的改进方案已被证实具有实时性高、CSI 重构能力强等优点^[7],受到了广泛关注。文献[8]提出了基于深度学习的 CSI 压缩反馈网络结构 CsiNet,通过在发送和接收端构建编解码网络,分别学习 CSI 反馈矩阵到码元,以及码元到 CSI 反馈的变换和逆变换,可显著提高 CSI 反馈的重建精度。在此基础上,文献[9-11]通过学习信道中的隐式相关性或引入量化模块等进一步提升了网络性能,但模型复杂度较高。文献[12]和文献[13]随后提出的 JcNet 结构和 BcsiNet 结构降低了模型的复杂度,但存在性能降低的问题。文献[14]提出的 CRNet 结构通过添加多分辨率卷积模块以及在网络训练中使用预热辅助余弦学习率调度器,在保持较好性能的同时很大程度上降低了模型复杂度,

但存在泛化能力弱的缺点,不适用于复杂环境。此外,其他方法还包括引入扩张卷积^[15]、聚合通道重构^[16]等。然而,通过增加卷积层和卷积路径提升精度的操作会大大增加网络参数和复杂度。针对上述问题,文献[17-20]尝试在编、解码器中部署注意力,文献[21]通过添加反馈延迟算法来模型网络的性能,但仍存在性能随着压缩率的增高而迅速降低的问题。

综上所述,现有方法要么以模型的复杂度为代价提高模型精度和泛化性,要么以模型的精度或泛化性为代价降低模型的复杂度。基于此,本文提出一种基于 UNet 结构和 Transformer 模块^[22]的连续卷积 CSI 反馈网络结构 UTNet (U-shaped Transformer Neural Network),以解决模型复杂度与精度、泛化性的矛盾问题。其主要的创新点与特色可概括如下:

1) 特征提取和向量编解码结合,设计出一种同步压缩模块。其特征在于通过连续下采样将 CSI 信号的特征提取和压缩同时执行,而传统模型中特征提取和压缩需分开执行。

2) 使用 Transformer 模块作为向量缩放器,代替常见模型中的全连接层,在发送端和接收端进行特征整合和缩放。通过引入位置编码、多头注意力和自注意力机制,提升模型的特征整合能力,实现对 CSI 信号更加智能和高效的处理。

3) 精简的结构。优化后的模型架构在轻量化,更易部署的同时能保持更高的精度,满足对 CSI 信号处理的高效性和实用性需求。

1 大规模 MIMO-OFDM 系统模型和 CSI 反馈结构

所述的大规模 MIMO-OFDM 系统参数设定为 BS 端有 $N_t (>> 1)$ 根发射天线,用户(User, UE)端有 $N_r = 1$ 根接收天线和 N_c 个子载波。在第 n 个子载波处的复值接收信号为

$$\mathbf{y}_n = \mathbf{h}_n^H \mathbf{v}_n \mathbf{x}_n + \mathbf{z}_n \quad (1)$$

式中: $\mathbf{y}_n, \mathbf{h}_n \in \mathbb{C}^{N_r \times 1}$ 、 $\mathbf{v}_n \in \mathbb{C}^{N_t \times 1}$ 、 $\mathbf{x}_n \in \mathbb{C}^{N_c \times N_r}$ 和 $\mathbf{z}_n \in \mathbb{C}^{N_c \times N_r}$ 分别表示接收端信号、第 n 个子载波的下行 CSI 向量、预编码向量、传输数据和信道加性高斯白噪声; $(\cdot)^H$ 表示共轭转置矩阵,所有子载波的 CSI

之和用信道矩阵 $\mathbf{H} = [h_1, h_2, h_3, \dots, h_{N_c}]^H$ 表示,其矩阵大小为 $N_c \times N_t$ 。可见,在大规模 MIMO 系统中庞大的矩阵 $\mathbf{H}$ 将产生大量的通信开销。

为解决上述问题,需要先将 CSI 矩阵 $\mathbf{H}$ 进行压缩再反馈给 BS,压缩反馈结构如图 1 所示。利用角延迟域的信道稀疏性,通过离散傅里叶变换 (Discrete Fourier Transform, DFT),将信道矩阵压缩:

$$\hat{\mathbf{H}} = \mathbf{F}_c \mathbf{H} \mathbf{F}_t^H \quad (2)$$

式中: $\mathbf{F}_c \in \mathbb{C}^{N_c \times N_c}$ 和 $\mathbf{F}_t \in \mathbb{C}^{N_t \times N_t}$ 分别是角度域和延迟域的 DFT 变换矩阵。对于角延迟域通道矩阵 $\hat{\mathbf{H}} \in \mathbb{C}^{N_t \times N_t}$,除前 N_a 行元素具有较大的值外,其余的行由接近零的元素组成,因而可删除矩阵中的近零元素,且几乎不丢失信息。然后, $\hat{\mathbf{H}}$ 的前 N_a 行用新矩阵 $\mathbf{H}_a$ 表示,达到矩阵压缩的目的。

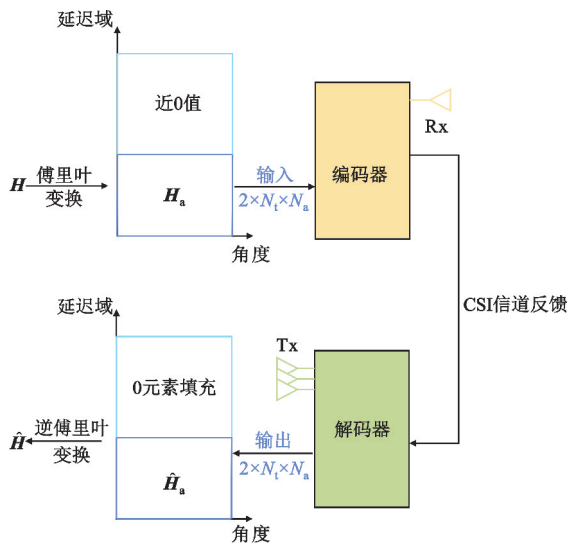


图 1 大规模 MIMO 系统中 CSI 反馈结构

2 UNet 结构

大规模 MIMO 系统中 N_a 大小由具体的应用场景、系统设计和性能要求确定,多径效应较为显著的环境中 N_a 相对较大。在实际系统中, N_a 值可能从几十到几百不等。因此,虽然 $\mathbf{H}_a$ 比 $\mathbf{H}$ 要小得多,但 $N_a \times N_t$ 的大小仍然是一个很大的矩阵。为了提高传输效率,需要进一步压缩 CSI 的 $\mathbf{H}_a$ 矩阵。

2.1 CSI 反馈设计

UTNet 将 Transformer 模块和 UNet 结构连续下

采样结合,其网络结构将提取和数据压缩步骤融合,在更大限度地保留空间信息的同时提高参数利用效率,减少过拟合,具体结构如图 2 所示。

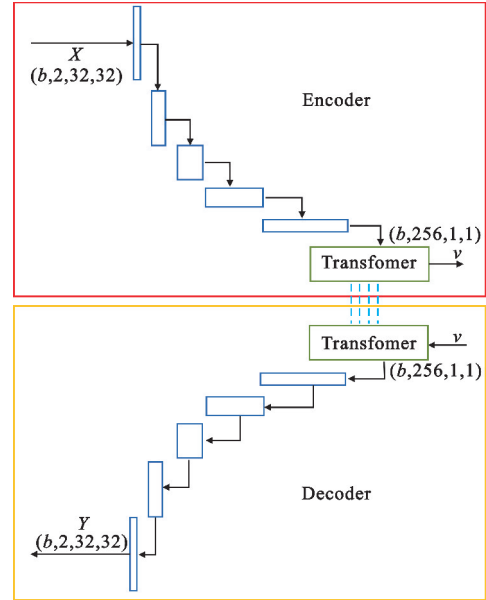


图 2 UNet 结构

CSI 反馈流程如下:转移到角域的通道矩阵 $\mathbf{H}_a$ 通过图 2 所示的 UNet 的编码器和给定的压缩率 η ,压缩成一个短的特征向量 $\vec{v}$ 并反馈给 BS。 $\vec{v} = \eta \times (N_a \times N_t \times 2)$ 为一维向量,即 $\vec{v} \in \mathbb{C}^{(\eta \times (N_a \times N_t \times 2)) \times 1}$ 。BS 端接收到 $\vec{v}$,通过 UNet 结构中的解码器将其重构为 $\hat{\mathbf{H}}_a$ 。最后,通过零填充和逆 DFT 将 $\mathbf{H}_a$ 恢复为 $\mathbf{H}$ 。因此,整个反馈网络模型可表示为

$$\mathbf{v} = E(\theta_1, \mathbf{H}_a) \quad (3)$$

$$\hat{\mathbf{H}}_a = D(\theta_2, \mathbf{v}) \quad (4)$$

式中: $E(\cdot)$ 和 θ_1 分别表示压缩操作和编码器的参数; $D(\cdot)$ 和 θ_2 分别表示恢复操作和解码器模型参数。该网络的最终目的是最小化重构误差,因此采用归一化均方误差 (Normalized Mean Square Error, NMSE) 作为最终评估指标。

2.2 编码器和解码器的结构

CsiNet 和 CRNet 结构均已证实可以有效压缩和重构 CSI 反馈,其网络结构包括用于提取特征信息的卷积块和用于特征图的压缩重构的全连接层 (Full Connection, FC)。这些网络结构的特点是,在特征提取的卷积部分,特征图大小始终保持不变,如

图 3 所示。

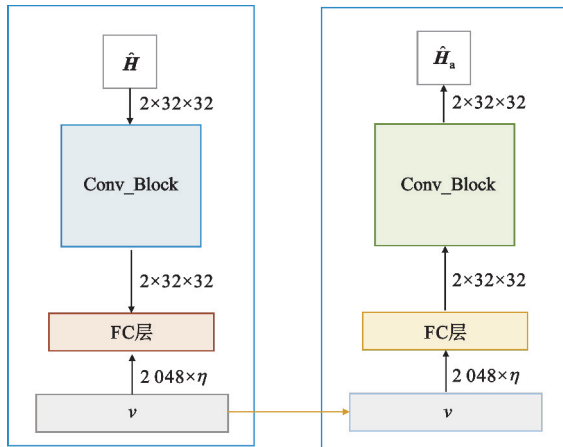


图 3 CsiNet 和 CRNet 结构

然而,FC 层参数数量取决于前一层的神经元数量。如果特征图较大,全连接层的参数数量会显著增加,增加计算开销的同时还容易引发过拟合。而且,使用 FC 层压缩特征图将丢失大部分卷积层保留的空间位置信息,难以保存完整的信号特征。针对这些问题,UTNet 采取连续卷积采样的方式处理信号,一方面通过连续卷积高效提取特征,另一方面通过连续下采样对信号进行压缩编码,以避免全连接引起的问题。与 CRNet 等网络相比,UTNet 的结构更有效地保留了信号的空间特征,以此提高信号质量。

UTNet 由 UE 端的编码器和 BS 端的解码器组成,分别设计了针对 CSI 信号输入特点连续下采样模块 Down Block 和上采样模块 Up Block,结构如图 4 所示。如图 4(a)所示,编码器采用 5 个下采样模块,在增加特征图通道的同时逐步减小特征图大小。该操作将 $2 \times 32 \times 32 (2\,048)$ 大小的特征图编码成 $256 \times 1 \times 1$ 大小的同时提取了 CSI 信号隐含的特征信息,随后将其输入 Transformer 模块,进一步缩小传输数据量的同时再次进行基于全局注意力的整合编码。其中,Transformer 输出的特征向量 $\vec{v}$ 可针对不同信道传输能力进行缩放,网络总体缩放比 η ($\vec{v}=256$ 时 $\eta=18$) 由式子 $\eta=\vec{v}/(N_a \times N_t \times 2)$ 确定。解码器结构如图 4(b)所示,首先采用 Transformer 模块对接收到的特征向量进行尺寸放大和调整,然后输入上采样模块进行特征图的解码和恢复。

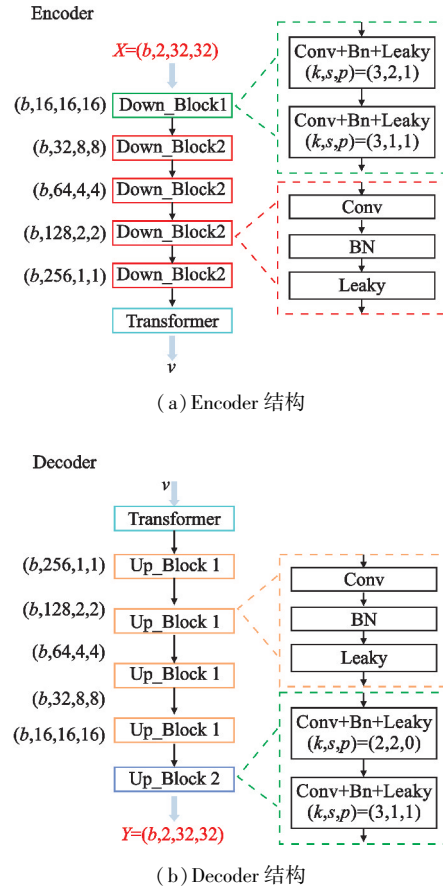


图 4 Decoder 结构和 Encoder 结构

3 仿真实验和结果分析

3.1 数据集和参数设置

本文考虑 5.3 GHz 的室内场景和 300 MHz 的室外场景两种场景。为了便于比较,采用 CsiNet 中的基本系统设置^[8]。BS 端设置 $N_t=32$ 的均匀线阵 (Uniform Line Array, ULA), UE 端设置 1 根接收天线 ($N_r=1$)。对于 FDD 系统,频域取 $N_c=1\,024$,角域取 $N_a=32$ 。数据集按照 COST2100^[23] 中的默认设置生成,15 万个独立的 CSI 反馈数据被分为训练、验证和测试数据集,分别包含 10 万个、3 万个和 2 万个信道矩阵。

整个模型基于 PyTorch 架构设计,训练过程使用 Adam 优化器,学习率设置为 0.005。网络训练 200 个 epochs,采用 L_1 Loss 和均方误差 (Mean Square Error, MSE) 衡量网络的损失,公式如下:

$$L_1 = \|v\|_1 = \sum_{i=1}^M |v_i| \quad (5)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N \|H_a - \hat{H}_a\|_2^2 \quad (6)$$

式中: H_a 和 $\hat{H}_a$ 分别表示网络的输入和输出。为了进一步衡量网络的性能,在测试阶段 NMSE 和余弦

相似度作为模型评估指标,公式如下:

$$\text{NMSE} = E \left\{ \frac{\| \mathbf{H}_a - \hat{\mathbf{H}}_a \|_2^2}{\| \mathbf{H}_a \|_2^2} \right\} \quad (7)$$

$$\rho = \frac{1}{B} \sum_{i=1}^B \frac{\| \mathbf{H}_a^H \hat{\mathbf{H}}_a \|_2}{\| \mathbf{H}_a \|_2 \| \hat{\mathbf{H}}_a \|_2} \quad (8)$$

3.2 性能评估

3.2.1 NMSE 和复杂度分析

UTNet 与 CsiNet、CsiNet+、CRNet (cosine) 采用相同损失函数和训练次数,以便更好比较各网络的精确度。室内环境中,整体性能的对比结果如表 1 所示。从表 1 可知,在相同的训练方案下,UTNet 除在 $\eta=1/4$ 的场景中和 CsiNet+ 表现相似外,在其他场景中均表现出最佳的性能;且随着压缩率的增高,与其他模型相比,精度优势也越来越明显。具体表现为,当压缩率由 $1/4$ 增高至 $1/16$ 时, CsiNet 模型误差由 -17.36 dB 增高至 -8.65 dB, CsiNet+ 模型误差由 -27.37 dB 增高至 -14.14 dB, CRNet 模型误差从 -26.9 dB 增高至 -11.35 dB,此时,UTNet 仍能保持 -27.45 dB 以下的低误差。这证明了 Transformer 在处理时序 CSI 信号时具有显著的优越性,同时也表明了 UTNet 模型能够很好地对 CSI 信号进行特征学习,提取出对于信号重建和预测有用的特征。

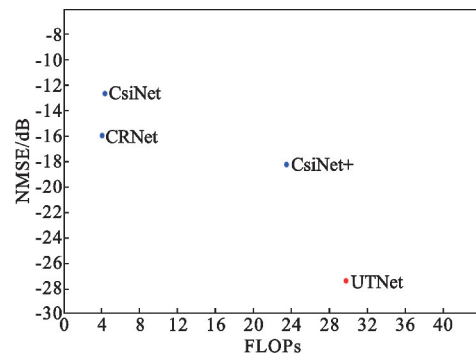
表 1 不同模型的 NMSE 和复杂度比较

η	模型	NMSE/dB		参数量/ 10^3	计算量/ 10^6
		室内	室外		
$\frac{1}{4}$	CsiNet	-17.36	-8.75	2 103	5.41
	CsiNet+	-27.37	-12.40	2 122	24.57
	CRNet (cosine)	-26.99	-12.71	2 103	5.12
	UTNet (ours)	-27.49	24.96	675	42.5
$\frac{1}{8}$	CsiNet	-12.7	-7.65	1 054	4.37
	CsiNet+	-18.29	-8.72	1 073	23.52
	CRNet (cosine)	-16.01	-8.04	1 054	4.07
	UTNet (ours)	-27.46	-24.97	675	29.8
$\frac{1}{16}$	CsiNet	-8.65	-4.51	530	3.84
	CsiNet+	-14.14	-5.73	549	23.00
	CRNet (cosine)	-11.35	-5.44	530	3.55
	UTNet (ours)	-27.50	-24.94	675	23.4

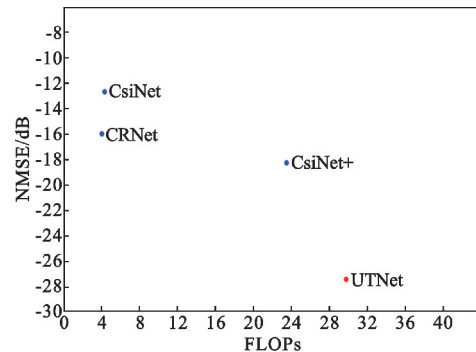
此外,表 1 也给出了各模型室外场景的性能测试结果。由于室外信号的传输通道干扰因素更加复杂,各模型在室外场景下 CSI 信号的重建效果均远差于室内场景。当 $\eta=1/16$ 时, CsiNet 在室外场景的误差达到 -4.51 dB,已经不能很好地完成信号处

理任务。在此情况下 UTNet 仍保持很好的性能,其 NMSE 最高仍为 -24.94 dB,远小于其他对比模型。UTNet 在室内外的良好表现表明,该模型具有很强的泛化能力,因此适用场景更加广泛。

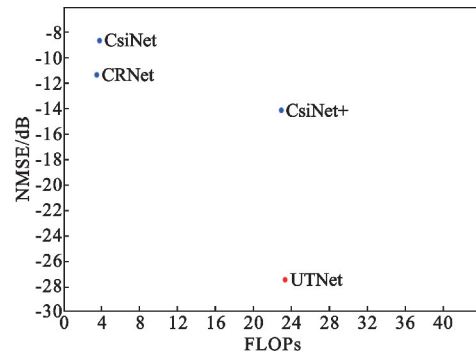
浮点运算 (Floating Point Operations, FLOPs) 常用来衡量算法的复杂度或计算量,是神经网络模型运算速度的间接衡量标准。图 5 给出了室内测试环境各 CSI 反馈结构在不同压缩率下 NMSE 和 FLOPs 之间的关系。从图 5(a)和(b)可知,UTNet 在不同压缩率下始终保持很低的重建误差,且其复杂度随着压缩率的增高迅速降低。同时,从图 5(a)、(b)和(c)可知,CRNet 在 $\eta=1/4$ 能保持较低的重建误差,而随着压缩率的增高其误差会逐渐增加至 CsiNet 附近,远高于 UTNet,未能一直保持精度优势。



(a) $\eta=1/4$



(b) $\eta=1/8$



(c) $\eta=1/16$

图 5 不同 CSI 反馈结构中 NMSE 和 FLOPs 之间的关系

不同压缩率的选择会影响模型对 CSI 的重建效果,因而需针对 UTNet 的 Transformer 模块的压缩率进行消融研究。表 2 给出了 UTNet 结构中不同压缩率对 NMSE 的影响规律,其实验原始输入信号格式为 $2 \times 32 \times 32 (2\ 048)$,当压缩率为 1/4、1/8、1/16 时,编码器输出信号长度分别为 512、256、128,从表 2 可知,对应的 NMSE 大小分别为 -27.49 dB、-27.46 dB、-27.50 dB。结果表明,在一定范围内,UTNet 结构随着压缩率的提高 NMSE 降低,但随着压缩率升高至 1/32, NMSE 反而轻微增加。此外,图 6 给出了在不同压缩率情况下 NMSE 和 FLOPs 之间的变化规律。由图 6 可知,随着压缩率增高至 1/16,模型复杂度减少到了 2.34×10^7 。上述结果综合表明,网络在 16:1 的高压缩率的情况下,不仅能保留大部分原始 CSI 信号信息,还能减少模型复杂度,更有利于模型在常用设备上的部署。

表 2 在 UTNet 模型中不同压缩率对 NMSE 的影响

压缩率	ν	Avg_NMSE/dB
1/32	64	-27.44
1/16	128	-27.50
1/8	256	-27.46
1/4	512	-27.49

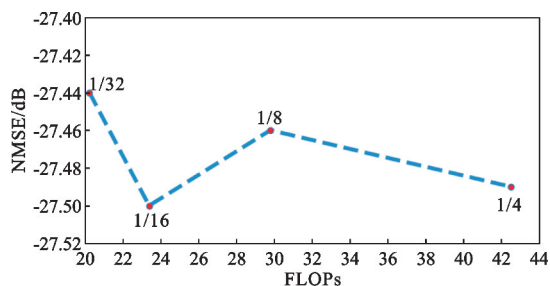


图 6 不同压缩率下 NMSE 与 FLOPs 之间的变化规律

3.2.2 余弦相似度评估

余弦相似度用于评估反馈的 CSI 特征向量与原始特征向量的相似度,其与无线通信系统的吞吐量呈正相关。

本文对 $N_a=32$ 和 $N_t=32$ 时的两组 CSI 反馈模型 CsiNet 和 UTNet 进行实验,对比结果见表 3。在 5 种压缩率下,所提出的 UTNet 的重建矩阵 $\hat{H}_a$ 的余弦相似度均高于 CsiNet。该结果表明,自注意力机制的加入使得模型对于压缩率和室外干扰因素的抗干扰性大大增加。

表 3 UTNet 和 CsiNet 模型之间的余弦相似度对比

η	模型	余弦相似度 ρ	
		室内	室外
1/4	CsiNet	0.99	0.91
	UTNet(ours)	0.99	0.95
1/8	CsiNet	0.93	0.79
	UTNet(ours)	0.99	0.95
1/16	CsiNet	0.93	0.79
	UTNet(ours)	0.99	0.95
1/32	CsiNet	0.89	0.67
	UTNet(ours)	0.98	0.94
1/64	CsiNet	0.87	0.59
	UTNet(ours)	0.96	0.94

4 结束语

本文针对常见模型泛化能力不强、高压缩率情况下精度不高的问题,提出了一种新的用于下行 CSI 反馈的神经网络架构 UTNet。模型采用连续下采样同步进行特征提取和编码,同时引入自注意力机制、位置编码和多头注意力机制来提升泛化和捕捉长距离依赖关系的能力。编解码两端的设计使得模型在轻量化的同时能在复杂环境中保持良好的性能。实验对比表明,该深度学习网络模型具有强大的泛化能力、鲁棒性和特征学习能力,有效具有较强的适用性和实用性。

然而,网络未考虑实际应用中的量化及反馈延迟等步骤,这将在未来的工作中需要进一步完善。

参考文献:

- [1] 廖勇,李雪,杨植景. 人工智能在 6G 空口物理层的潜在应用研究进展[J]. 北京邮电大学学报,2022,45(6):22-31.
- [2] 张雷,代红. 面向 5G 的大规模 MIMO 技术综述[J]. 电讯技术,2017,57(5):608-614.
- [3] HAI H, JIANG X Q, JEONG Y, et al. Low complexity precoding codebook design based on DFRST for limited feedback precoded OSTBC systems [J]. EURASIP Journal on Wireless Communications and Networking, 2017,2017(1):1-6.
- [4] 汪丽青,杨龙祥. 基于压缩感知的信道反馈重构[J]. 电讯技术,2019,59(8):880-884.
- [5] 黎明源,段红光,李振一. 基于压缩感知的大规模 MIMO 下行信道状态信息获取[J]. 电讯技术,2019,59(5):562-568.
- [6] AL-SALIH H, NAKHAI M R. Bayesian compressed

- sensing-based channel estimation for massive MIMO systems [C]//2016 European Conference on Networks and Communications. Athens;IEEE,2016:360-364.
- [7] 陈成瑞,程港,何世彪,等. 基于人工智能的大规模 MIMO 信道状态信息反馈综述[J]. 电讯技术,2021,61(9):1181-1190.
- [8] WEN C K, SHIH W T, JIN S. Deep learning for massive MIMO CSI feedback [J]. IEEE Wireless Communications Letters,2018,7(5):748-751.
- [9] WANG T Q, WEN C K, JIN S, et al. Deep learning-based CSI feedback approach for time-varying massive MIMO channels [J]. IEEE Wireless Communications Letters,2019,8(2):416-419.
- [10] CHEN T, GUO J J, JIN S, et al. A novel quantization method for deep learning-based massive MIMO CSI feedback [C]//2019 IEEE Global Conference on Signal and Information Processing. Ottawa;IEEE,2019:1-5.
- [11] GUO J J, WEN C K, JIN S, et al. Convolutional neural network-based multiple-rate compressive sensing for massive MIMO CSI feedback: design, simulation, and analysis [J]. IEEE Transactions on Wireless Communications,2020,19(4):2827-2840.
- [12] LU C, XU W, JIN S, et al. Bit-level optimized neural network for multi-antenna channel quantization [J]. IEEE Wireless Communications Letters,2020,9(1):87-90.
- [13] LU Z L, WANG J T, SONG J. Binary neural network aided CSI feedback in massive MIMO system [J]. IEEE Wireless Communications Letters,2021,10(6):1305-1308.
- [14] LU Z L, WANG J T, SONG J. Multi-resolution CSI feedback with deep learning in massive MIMO system [C]//2020 IEEE International Conference on Communications. Dublin;IEEE,2020:1-6.
- [15] TANG S P, XIA J J, FAN L S, et al. Dilated convolution based CSI feedback compression for massive MIMO systems [J]. IEEE Transactions on Vehicular Technology,2022,71(10):11216-11221.
- [16] LU Z L, ZHANG X D, HE H Y, et al. Binarized aggregated network with quantization: flexible deep learning deployment for CSI feedback in massive MIMO systems [J]. IEEE Transactions on Wireless Communications,2022,21(7):5514-5525.
- [17] 廖勇,李玉杰. 一种轻量化低复杂度的 FDD 大规模 MIMO 系统 CSI 反馈方法 [J]. 电子学报,2022,50(5):1211-1217.
- [18] 杨蓓,梁鑫,尹航,等. 基于自注意力机制的大规模 MIMO 信道状态信息特征向量反馈方法 [J]. 电信科学,2023,39(11):128-136.
- [19] JI S J, LI M. CLNet: complex input lightweight neural network designed for massive MIMO CSI feedback [J]. IEEE Wireless Communications Letters,2021,10(10):2318-2322.
- [20] 刘为波,颜彪,沈麟,等. 基于 AI 通信的大规模 MIMO 信道状态信息反馈网络 [J]. 电讯技术,2024,64(1):29-35.
- [21] 王月,段红光,郑兴林. 大规模 MIMO 系统中基于 CNN 的延迟 CSI 反馈改进算法 [J]. 电讯技术,2020,60(7):833-838.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//The 31st International Conference on Neural Information Processing Systems. Long Beach;IEEE,2017:6000-6010.
- [23] LIU L F, OESTGES C, POUTANEN J, et al. The COST 2100 MIMO channel model [J]. IEEE Wireless Communications,2012,19(6):92-99.

作者简介:

刘庆 男,1983 年生于湖南邵阳,2008 年获硕士学位,现为高级工程师,主要研究方向为无线通信系统、传感器网络、深度学习。

李义 男,1991 年生于贵州毕节,2014 年获学士学位,现为中级工程师,主要研究方向为输电技术、智能监测技术、无线传感器网络、人工智能及应用。

李璘 男,1990 年生于贵州遵义,2013 年获学士学位,现为中级工程师,主要研究方向为信号处理、输电线路防灾减灾关键技术。

王有军 男,1995 年生于贵州毕节,2018 年获学士学位,现为中级工程师,主要研究方向为输电线路数字电网技术、信号处理。

李康 男,1986 年生于河北保定,2013 年获学士学位,现为中级工程师,主要研究方向为输电线路数字化管理技术、信号处理。

熊林麟 男,1997 年生于贵州毕节,2019 年获学士学位,现为中级工程师,主要研究方向为输电智能终端通信技术、深度学习。

王平 男,1981 年生于重庆,2013 年于电子科技大学获工学博士学位,现为副教授、博士生导师,主要研究方向为新型天线、射频能量传输与收集、深度学习、无线射频感知与通信。

DOI:10.20079/j.issn.1001-893x.240727001

一种识别和检测人工智能生成文本的算法*

王雨欣¹, 刘柯飞¹, 李雪莲², 王红军³

(1. 河海大学 信息科学与工程学院, 江苏 常州 213200; 2. 扬州大学广陵学院, 江苏 扬州 225000;
3. 国防科技大学 电子对抗学院, 合肥 230031)

摘要:针对目前人工智能(Artificial Intelligence, AI)生成文本的滥用导致的学术不端、侵犯版权、隐私保护和舆情监控等问题,提出了一种基于自然语言处理的 AI 生成文本的识别和检测算法。该算法首先采用 Word2vec 方法中的连续词袋模型将文本词转换成词向量,并将词向量累加获得文本向量。随后利用 softmax 函数获取文本向量的概率分布,通过统计可视化分析 AI 生成文本的基本规律,并采用余弦相似性来判断文本类型。其次采用支持向量机递归特征消除算法判断文本是否由 AI 生成,通过 K-近邻算法对文本重生成本次数进行判断,进一步细化了文本检测的粒度。通过仿真实验验证了算法的有效性,结果显示算法识别准确率达 80% 及以上。

关键词: AI 生成文本检测; 文本向量; 余弦相似性; 支持向量机(SVM); K-近邻(KNN)算法

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号: TP181 文献标志码: A 文章编号: 1001-893X(2025)03-0378-07

An Artificial Intelligence-generated Text Recognition and Detection Method

WANG Yuxin¹, LIU Kefei¹, LI Xuelian², WANG Hongjun³

(1. College of Information Science and Engineering, Hohai University, Changzhou 213200, China;
2. Guangling College of Yangzhou University, Yangzhou 225000, China;
3. College of Electronic Engineering, National University of Defense Technology, Hefei 230031, China)

Abstract: To address such issues as academic dishonesty, copyright infringement stemming, privacy protection and public opinion monitoring from the misuse of artificial intelligence (AI)-generated texts, an recognition and detection algorithm based on natural language processing (NLP) is proposed. This algorithm initially converts words into vectors using the continuous bag-of-words (CBOW) model within Word2vec, and accumulates them into text vectors. It then applies softmax to address their probability distribution, analyze the fundamental patterns of AI-generated texts with statistical visualization, and determin the type of text by using cosine similarity. Next, a support vector machine recursive feature elimination (SVM-RFE) is used to determine whether the text is generated by AI. For AI-generated texts, the K-nearest neighbor (KNN) algorithm estimates the extent of AI involvement, further refining the granularity of text detection. Finally, simulation experiments show the algorithm's effectiveness with recognition accuracy of 80% or above.

Key words: AI-generated text detection; text vector; cosine similarity; support vector machine (SVM); K-nearest neighbor (KNN) algorithm

* 收稿日期: 2024-07-27; 修回日期: 2024-08-21
基金项目: 国家自然科学基金面上项目(61971473)
通信作者: 王红军 Email: wanghongjun17@nudt.edu.cn

0 引言

信息技术的飞速发展催生了人工智能 (Artificial Intelligence, AI) 在日常生活中的各种应用,如语音识别、文本识别、视频识别等。在这些领域内 AI 甚至超越了人类的水平,成为了引领新一轮技术革命和产业变革的战略技术。诸如 ChatGPT、文心一言等大型语言模型提供的高效率与高质量服务极大地便利了社会中的日常生活与学习场景。然而, AI 生成文本的滥用现象给人们的创造力发展、隐私安全、学术诚信和版权保护带来了不可忽视的负面影响。因此,在人工智能蓬勃发展的背景下,为避免此类负面后果,研究出可识别和检测 AI 生成文本的模型显得尤为重要。

文献[1]通过将文献摘要与 AI 重生成的摘要进行对比测试,验证了 AI 生成内容对于文字相似性检测系统的逃脱能力,表明了 AI 生成的文本能够较好地通过现有的文献相似性监测系统。中国科学院大学研究团队提出基于代理引导重采样的高效检测溯源办法^[2],在宏观 F1 分数上优于所有基线,但其依赖于选择一小部分代表性词进行重采样,因此如何准确、有效地选择这些代表性单词仍是一个挑战。文献[3]中,由斯坦福大学的研究团队研发的 DetectGPT 法通过对待检测样本生成扰动变体并比较其概率差异来进行检测,需要成百上千次地调用模型来生成文本的扰动变体进行评分,导致了极大的计算开销。文献[4]中,西湖大学团队所研发出的 Fast-DetectGPT 通过条件概率曲率来检测机器生成文本,拥有着高准确率与低成本,将检测过程加快了 340 倍,虽然有效但计算开销仍然较大,且在面对处理特定类型文本或多语言环境时可能存在局限性,无法适用于所有场景。国内学者研发的基于深度金字塔卷积神经网络 (Convolutional Neural Network, CNN) 检测 ChatGPT 生成文本的方法有着强特征提取能力和高检测准确率^[5],但其数据集类别较为局限,且对长文本处理困难。综上,目前国内外对于识别和检测 AI 生成文本的技术仍存在较大的局限性,通常准确率和效率无法兼顾,且统计量的特征获取困难、输入文本受限等问题还有待很好地解决。

自然语言处理 (Natural Language Processing,

NLP) 在大语言模型中占据着核心和基础的地位,它将人类交流沟通所用的语言转化为机器所能理解的机器语言,是一种研究语言能力的模型和算法框架,是语言学和计算机科学的交叉学科^[6]。NLP 包括自然语言处理技术和自然语言处理资源,其研究涵盖了多个领域,包括语言学、计算机科学、人工智能等^[7-8]。词嵌入是自然语言处理中一种将单词或短语映射到向量空间以捕捉其语义信息的技术,是相关文本的一种重要表示。

根据以上理论,本文提出了一种基于自然语言处理的识别和检测人工智能生成文本的算法。与现有 AIGC 识别和检测算法相比,本文算法通过引入词嵌入技术提升模型对文本的理解能力,能够在维持中高等检测准确率和效率的基础上有效处理长文本数据,并可以判断输入的文本属于哪类文本类型。此外,本文所提算法还可以有效检测目标文本经过 AI 重生成的次数,且算法泛化能力强,可满足多种应用场景的需求。

1 算法模型介绍

算法的整体流程如图 1 所示。

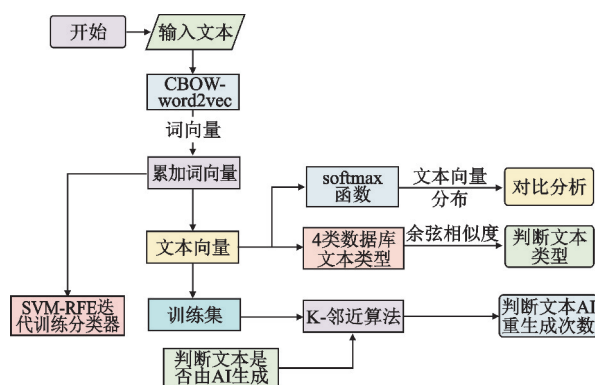


图 1 算法模型流程

针对获取的一段文本,算法运用 Word2Vec 中连续词袋模型 (Continuous Bag of Words Model, CBOW) 将文本转换成词向量,并将词向量累加以获取文本向量,且使用 softmax 函数对数据进行规范化,通过可视化分析与描述性统计获取文本基本规律,比较待测文本向量与不同类型文本向量之间的余弦相似度来判断文本类型。

获取文本向量概率分布算法流程如图 2 所示。

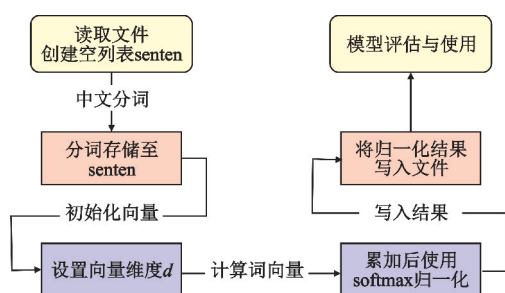


图2 获取文本向量概率分布流程

其次使用支持向量机递归特征消除算法 (Support Vector Machine-Recursive Feature Elimination, SVM-RFE) 将文本在进行特征消除的同时分为人工智能生成文本和非人工智能生成文本两类。SVM-RFE 算法流程如图3所示。

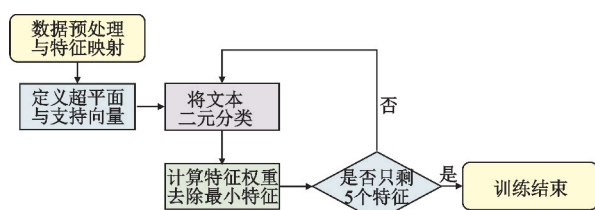


图3 SVM-RFE 算法流程

最后通过 K-近邻 (K-nearest Neighbor, KNN) 算法^[9] 来判断人工智能生成文本属于第几次重生成。KNN 算法流程如图4所示。

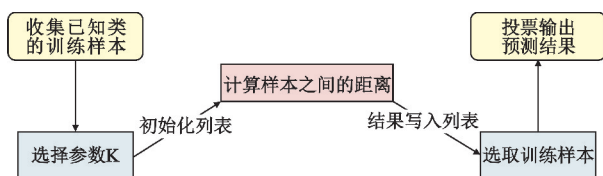


图4 KNN 算法流程

2 基于自然语言处理的 AI 生成文本识别和检测算法

2.1 文本基本规律提取

文本向量概率分布的获取主要由 Word2vec 中的 CBOW 模型来实现。Word2vec 模型是词嵌入的一种具体实现方法,目前已开源且已广泛应用于各种 NLP 任务。文献[10]描述了其核心架构为 CBOW 和 Skip-Gram 两种模型。由于 CBOW 模型计算效率高且对高频词表现更敏感,本文选用 CBOW 模型生成词向量,并对它的输入层和输出层进行了

改进,使得输入层可接收上下文单词的编码向量,将所得的词向量累加后生成文本向量,最后在输出层将数据归一化后得到最终的处理结果。

CBOW 输入层能在给定单词 W_i 和相邻窗口大小为 k 的单词的情况下可预测出单词 W_i 的概率。预测公式如公式(1)所示:

$$P(W_i | \tau(W_{t-k}, W_{t-k+1}, \dots, W_{t+k-1}, W_{t+k})) \quad (1)$$

将文本 d_i 中的词向量进行相加,得到 d_i 的文本向量表示:

$$R(d_i) = \sum_i \text{word2vec}(t) \text{ where } t \in d_i \quad (2)$$

输出层使用 softmax 函数将数据规范化到 0 和 1 之间,具体实现如下:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \text{softmax} \begin{pmatrix} W_{1,1}x_1 + W_{1,2}x_2 + W_{1,3}x_3 + b_1 \\ W_{2,1}x_1 + W_{2,2}x_2 + W_{2,3}x_3 + b_2 \\ W_{3,1}x_1 + W_{3,2}x_2 + W_{3,3}x_3 + b_3 \end{pmatrix} \quad (3)$$

$$\text{softmax}(x)_i = \frac{\exp(x)_i}{\sum_{j=1}^n \exp(x_j)} \quad (4)$$

为直观分析原文本与 AI 生成文本的差异,选择 4 类原始文本和 AI 重生成 7 遍后的文本作为分析样本,将 CBOW 模型获取的词向量累加后得到文本向量。以文本向量为输入,使用 softmax 函数进行处理并输出每一类文本向量在词嵌入 50 个维度上的概率分布。50 维的向量维度能够在不显著增加计算复杂度的情况下提供足够的容量来捕捉语言中的语义和语法特征。

其次使用折线图进行可视化以进一步对每一类原始文本和 AI 生成文本进行对比分析。为获取原文本和 AI 重生成 N 次(实验取 N 为 1~7)文本的数据特征,对原文本和 AI 多次重生成的文本的文本向量分别进行描述性统计分析,即分析方差、标准差、中位数、标准误差、最大值、最小值等统计量,以从全局视角分析并总结 AI 迭代重生成文本的数据特征间的差异,并总结出基本规律。

2.2 文本识别和检测算法实现

2.2.1 建立文本类型判断模型

对于每个需要测试的文本,计算其文本向量与数据集中每一类风格文本的余弦相似性,根据余弦值判断相似度大小。当计算出的余弦值越接近 1,代表两个文本向量越相似,即文本类型越接近,从而判断文本类型^[11]。余弦相似性的计算公式如下:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}} \quad (5)$$

使用余弦相似性计算文本相似度可以忽略文本的长度和维度差异,只关注文本之间的方向和角度差异,强调方向而非距离,即确保了本文所提算法可以处理不同规模大小的文档与数据,测量高维数据时也能保持有效性。常用的文本相似度计算方法还有编辑距离、Livingston 相似度、TF-IDF 等,但余弦相似性计算方法精度高,应用更加广泛。

2.2.2 文本二元分类

SVM 是一种用于分类和回归分析的监督学习模型^[12]。目前的机器学习领域存在众多分类技术,但研究表明支持向量机的分类性能,尤其是泛化能力好于传统的分类方法^[13-14]。文献^[15]的研究表明 SVM 的解只与支持向量有关,可以很大程度上降低计算复杂性。SVM-RFE 算法^[16]结合了 SVM 的强分类能力和 RFE 的特征消除策略,通过每次迭代消去权重最小的特征,可以帮助降低特征维度,提高模型的可解释性和泛化性,还可以处理高维数据和带有噪声与冗余特征的数据集,从而提高分类器的性能和效率。

采用 SVM-RFE 算法处理文本的步骤如下:

1) 利用上文 Word2Vec 中的 CBOW 模型将训练集转化为文本向量并存储在表格中,将 AI 生成的文本标记为 1,非 AI 生成的文本标记为 0,代入 SVM 进行训练。

2) 使用 SVM 将测试文本进行分类。

3) 使用 RFE 算法计算当第 i 个特征被移除时目标函数 L_p 的变化,这将给出每个特征的排名系数,公式如式(6)和(7)所示。设置特征数量为 5,当模型剩余 5 个最重要特征时,迭代结束。

$$\Delta L_p(i) = \frac{\partial L_p}{\partial w_i} \Delta w_i + \frac{\partial^2 L_p}{\partial w_i^2} (\Delta w_i)^2 \quad (6)$$

$$\begin{cases} \text{Rank}(i) = (w_i)^2 \\ w = \sum_{i=1}^l \alpha_i y_i x_i \end{cases} \quad (7)$$

4) 根据排名系数,得到与文本向量位置相对应的 5 个最重要特征即第 45、41、31、39、44。取 5 个特征可以避免模型过拟合,使模型可以集中在最相关的信息上。在进行分类时,从数据集中提取这些特征作为输入,并获得相应输出。

2.2.3 文本重生次数确定

KNN 算法核心思想是基于实例的学习,计算出测试文本与训练样本集的每个文本的相似度^[17]。根据某种距离度量找出训练集中与其最靠近的 K 个样本,选择出现次数最多的类别标记作为预测结果^[18]。文献^[19]说明 KNN 算法的分类误差比贝叶斯的分类误差低两倍,所以采用 KNN 算法的计算结果相对准确,同时其所具有的非参数性质使其算法易于实现。

KNN 算法实现简单,分类性能显著,在 2006 年 IEEE 数据挖掘国际会议上,KNN 算法同其他方法一起被评选为十大数据挖掘算法^[20-21]。采用 AI 重生成多次的文本作为训练数据,使用 KNN 算法对待测文本进行处理与判断。具体操作步骤如下:

1) 应用余弦相似性公式计算待测文本与已知风格类型的文本之间的余弦相似度,具有最高相似性的数据类型将决定待测文本类型。

2) 从文本库中选择与待测文本类型相同的段落,通过 AI 重生成多次作为训练样本数据,表示为文本向量。

3) 根据样本大小选择合适的 k 值,它代表最近邻居的数量,本算法在仿真中由于数据集限制选取 $k=1$ 。

4) 对于未知样本,计算其文本向量并计算其与训练集中每个已知重生成次数的样本之间的距离,采用 n 维词向量空间中的欧几里得距离。

如果有两个文本向量表示为 $(x_1, x_2, \dots, x_k)$ 和 $(y_1, y_2, \dots, y_k)$,欧几里得距离计算公式如下:

$$d = \sqrt{\left((x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2 + \dots + (x_k - y_k)^2 \right)} \quad (8)$$

5) 基于距离度量,选择与未知样本最接近的 k 个训练样本。

6) 选择 k 个选定训练样本中最频繁的类别作为未知样本的预测类别。

3 仿真结果与分析

所有仿真实验均在配备 Intel Core i9-10900F 处理器、NVIDIA Quadro P4000 GPU、64GB RAM 的硬件平台上执行,使用 python 编程语言完成实验源代码的开发。采用数据集 <https://github.com/RageLiu/AI.git>,其中包括环境与气候变化、文学与艺术、学术交流与研究、医学与健康 4 类风格的 20 篇原文章及每篇使用 AI 重生成 7 次的文章用于获

取文本生成规律,以及用于 SVM-RFE 和 KNN 模型训练与测试的段落样本共 1 000 个。

数据集选用 4 种风格类型的文本,因此算法适用的文本风格也为这 4 种,同时是测试样本的选取标准。在训练模型时,需要进行手动调参以达到目标效果。

3.1 文本生成规律获取性能

计算 AI 生成文本中的单词数量,并与样本中的单词数量进行比较,单词计数结果如图 5 所示。

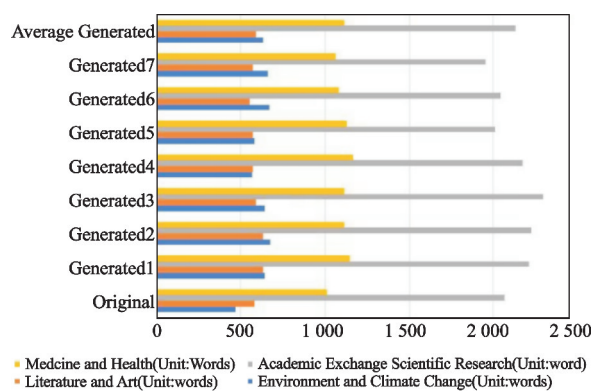


图 5 单词计数结果

不同分类的文本向量概率分布图 6 所示。

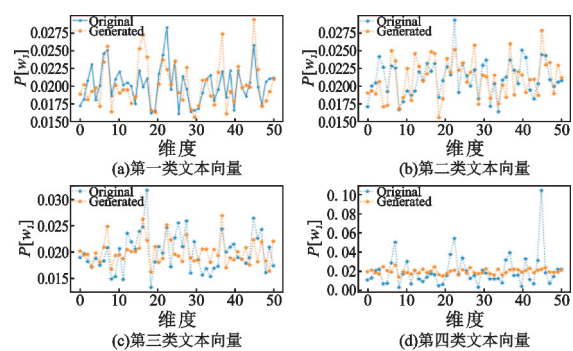


图 6 文本向量概率分布

样本和 AI 生成文本的文本向量概率折线图如图 7 所示。

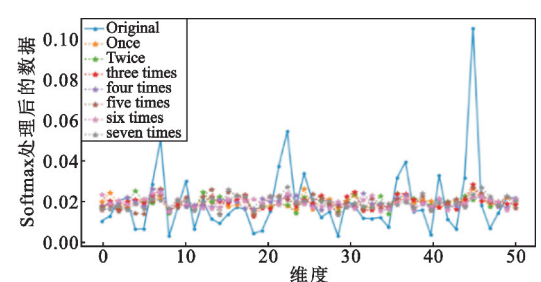


图 7 文本向量概率

统计性分析样本和 AI 生成文本向量分布方差、标准差、中位数、标准误差、最大值等统计量,结果如图 8 所示。

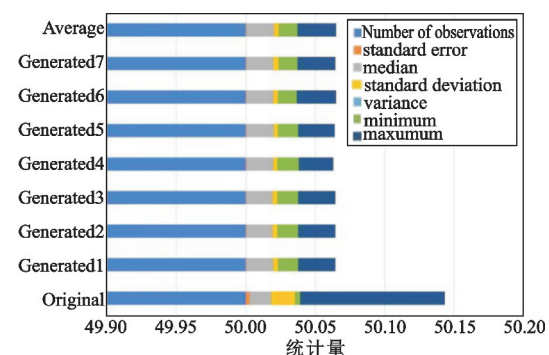


图 8 描述性统计结果

根据文本向量概率折线图和描述性统计图可以推断,样本和 AI 生成文本的文本向量分布呈现相似的变化趋势,随着 AI 重生次数的增加,文本向量分布趋于稳定。

采用文本向量可以从文本中提取突出的语义信息,使得机器可以理解和处理自然语言数据,提升文本处理和分析的效率。Word2Vec 技术能够保留词汇的上下文信息,有效提高了文本处理的准确性。

3.2 识别和检测算法性能

在模型训练中,本研究结合原始文献与 GPT-4、文心一言、讯飞星火以及 Kimi. ai 大语言模型,生成了总计 800 篇文本样本,以此作为训练集来训练 SVM-RFE 模型:将每个样本代入 Word2Vec 中的 CBOW 得到 50 维词向量分布表格后,给每个段落打上“0/1”标签,其中“0”表示为原文,“1”表示为 AI 生成。

训练好分类器后,采用 AI 随机生成的文本与文献原文共 18 个段落作为测试集,得到文本向量后代入模型即可获得相应的判断输出,结果如表 1 所示。将模型判断结果与文本本身类型相对比,准确率达 83.3%。

表 1 文本二元分类判断表

段落	是否为 AI 生成	段落	是否为 AI 生成
1	是	10	是
2	是	11	是
3	否	12	是
4	是	13	是
5	是	14	是
6	是	15	是
7	是	16	是
8	是	17	是
9	是	18	是

通过应用 KNN 算法,可以根据特定样本与已知训练样本的相似性,确定 AI 生成文本属于哪一次迭代,即可以判断文本属于第几次重生成。

采用数据集中 AI 生成重生成 5 次共 200 段文本作为训练文本,并重生成 20 个段落作为待测文本,进行 20 次实验得到每段文本的 AI 重生成次数判断,结果如表 2 所示。将 KNN 模型判断结果与文本类型相对比,准确率达 80%,可有效判断文本重生成次数。

表 2 AI 重生成次数判断表

段落	生成次数	段落	生成次数
1	1	11	2
2	2	12	4
3	1	13	2
4	2	14	3
5	3	15	1
6	3	16	3
7	2	17	3
8	4	18	4
9	3	19	1
10	5	20	3

3.3 算法性能对比分析

为了验证本算法的先进性,本文在确保本算法、文献[3-5]中所述的 DetectGPT、Fast-DetectGPT 和 CNN 等现行主流方法以及哈佛大学与麻省理工学院团队开发的 GLTR^[22]、OpenAI 推出的 OpenAIClassifier^[23] 均能有效识别和检测 AI 生成文本的前提下,对这些方法的准确率、泛用性和效率 3 种关键性能指标进行了直观的可视化对比分析,结果如图 9 所示。

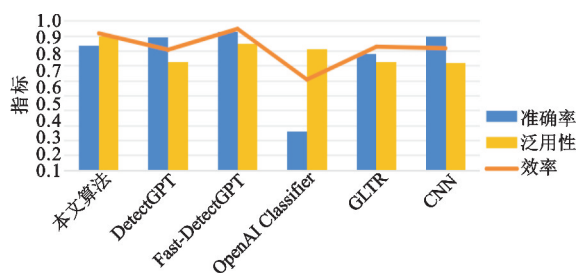


图 9 不同算法性能对比

在准确率方面,与性能相对突出的 DetectGPT、CNN 和 Fast-DetectGPT 相比,本文所提算法虽稍有逊色,但在效率和泛用性等指标上具有显著的优势。如前分析,由于采用 KNN 算法细化了文本检测的粒度,采用 SVM-RFE 算法构建了最优的子集,使得所

提算法在资源消耗等方面性能表现较为优越。同时算法能够对长文本进行有效检测,弥补了目前主流算法对文本输入长度最大值有限制的问题。

综上所述,本文所提算法具有相对优越的综合性能指标,并且在实际应用中具有显著的价值。

4 结束语

本文基于上述模型的建立与算法的实现,挖掘了 AI 生成文本的内部规则,并在此基础上通过余弦相似度判断文本分类,利用 SVM-RFE 算法判断文本是否为 AI 生成,再使用 KNN 算法来确定文本重生成的次数。但该算法仍存在进一步改进和深入研究的空间,例如,目前算法主要针对 4 种风格类型文本进行训练和测试,并且随着 AI 模型的不断演进,算法需要更新和优化以适应新的 AI 生成文本特征。此外,引入多模态融合以增强模型的可解释度与透明度;通过持续学习机制使模型能够不断从新数据中学习,适应 AI 生成文本特征的演变与多样性。总之,AI 生成文本识别和检测是一个充满挑战和机遇的领域,本文期望通过抛砖引玉促进该领域的深入研究,进而使得 AI 能够健康可持续地发展。

参考文献:

[1] 沈锡宾,王立磊. 人工智能生成学术期刊文本的检测研究[J]. 科技与出版,2023(8):56-62.

[2] SHI Y H, SHENG Q, CAO J, et al. Ten words only still help: improving black-box AI-generated text detection via proxy-guided efficient re-sampling [EB/OL]. [2024 - 07 - 20]. <https://arxiv.org/abs/2402.09199v1>.

[3] MITCHELL E, LEE Y, KHAZATSKY A, et al. DetectGPT: zero-shot machine-generated text detection using probability curvature [C]//2004 International Conference on Machine Learning. Honolulu: IEEE, 2004:24950-24962.

[4] BAO G S, ZHAO Y B, TENG Z Y, et al. Fast-DetectGPT: efficient zero-shot detection of machine-generated text via conditional probability curvature [EB/OL]. [2024 - 07 - 20]. <https://arxiv.org/abs/2310.05130v3>.

[5] 范志武,姚金良. 基于深度金字塔卷积神经网络的 ChatGPT 生成文本检测方法[J]. 数据分析与知识发现,2024,8(7):14-22.

[6] 赵京胜,宋梦雪,高祥. 自然语言处理发展及应用综述[J]. 信息技术与信息化,2019(7):142-145.

[7] 王灿辉,张敏,马少平. 自然语言处理在信息检索中的应用综述[J]. 中文信息学报,2007,21(2):35-45.

- [8] 张瀚文. 人工智能在自然语言处理中的应用[J]. 信息记录材料, 2024, 25(5) : 139-141.
- [9] COVER T, HART P. Nearest neighbor pattern classification [J]. IEEE Transactions on Information Theory, 1967, 13(1) : 21-27.
- [10] 熊富林, 邓怡豪, 唐晓晟. Word2Vec 的核心架构及其应用[J]. 南京师范大学学报(工程技术版), 2015, 15(1) : 43-48.
- [11] WANG X L, DONG X T, CHEN S X. Text duplicated-checking algorithm implementation based on natural language semantic analysis [C] // 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference. Chongqing: IEEE, 2020: 732-735.
- [12] 高斌斌, 王建军. 多分类最大间隔孪生支持向量机[J]. 西南师范大学学报(自然科学版), 2013, 38(10) : 130-135.
- [13] SCHOLKOPF B, SUNG K K, BURGESS C J C, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers[J]. IEEE Transactions on Signal Processing, 1997, 45(11) : 2758-2765.
- [14] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用, 2011, 47(3) : 123-124, 128.
- [15] 滕晓云, 徐俊, 陈德明. 基于 SVM 的信号解调算法[J]. 电讯技术, 2016, 56(10) : 1108-1111.
- [16] GUYON I, WESTON J, BARNHILL S, et al. Gene selection for cancer classification using support vector machines[J]. Machine Learning, 2002, 46(1) : 389-422.
- [17] 周庆平, 谭长庚, 王宏君, 等. 基于聚类改进的 KNN 文本分类算法[J]. 计算机应用研究, 2016, 33(11) : 3374-3377.
- [18] 王博, 钱蓉蓉, 任文平. 一种利用机器学习优化 MIMO 窃听信道天线选择的方法[J]. 电讯技术, 2020, 60(5) : 579-584.
- [19] 毋雪雁, 王水花, 张煜东. K 最近邻算法理论与应用综述[J]. 计算机工程与应用, 2017, 53(21) : 1-7.
- [20] WU X D, KUMAR V, ROSS QUINLAN J, et al. Top 10 algorithms in data mining [J]. Knowledge and Information Systems, 2008, 14(1) : 1-37.
- [21] 罗忠涛, 唐洪涛, 高天翱, 等. 基于动态自适应近邻算法的天波雷达 RD 图分类器设计[J]. 电讯技术, 2024, 64(8) : 1315-1321.
- [22] GEHRMANN S, STROBELT H, RUSH A M. GLTR: statistical detection and visualization of generated text [EB/OL]. [2024-07-20]. <https://arxiv.org/abs/1906.04043v1>.
- [23] KIRCHNER J H, AHMAD L, AARONSON S, et al. New AI classifier for indicating AI-written text [EB/OL]. [2024-07-20]. <https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/>.

作者简介:

王雨欣 女, 2003 年生于江苏扬州, 本科生, 主要研究方向为信息与通信工程。

刘柯飞 男, 2003 年生于河南项城, 本科生, 主要研究方向为物联网技术与应用。

李雪莲 女, 2004 年生于江苏南京, 本科生, 主要研究方向为计算机科学与技术。

王红军 男, 1968 年生于江苏镇江, 2011 年获博士学位, 现为教授, 主要研究方向为智能无线网络和认知无线电等。

DOI:10.20079/j.issn.1001-893x.250305005

基于模型的复杂航天电子载荷系统工程技术方法*

赵晓虎, 孙 杰, 吴慧伦, 柴 霖

(西南电子技术研究所, 成都 610036)

摘 要:针对新型复杂航天电子载荷需求多、实现功能多、新技术应用多、在轨持续演进等特点,提出了一套数字化系统工程方法,通过数字化需求分解管理、数字化系统设计、数字化协同设计、数实结合天地一体确认,贯穿系统需求分析、设计研制、在轨运行、演进升级全生命周期,实现全过程数字赋能,有效提升系统研发效率和质量。在实际工程项目中应用该方法,将系统架构设计确认时间从预计的 3 年缩减至数月,失效率下降 20%,维修时间缩短 74%。

关键词:航天电子载荷;复杂系统;系统工程;数字化协同设计

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:V57;N945.2 文献标志码:A 文章编号:1001-893X(2025)03-0385-07

A Model-based Complex Spacecraft Electronic Payloads System Engineering Technology Method

ZHAO Xiaohu, SUN Jie, WU Huilun, CHAI Lin

(Southwest China Institute of Electronic Technology, Chengdu 610036, China)

Abstract: According to the characteristics of new complex aerospace electronic payloads, such as diverse requirements, multiple functions, new technology applications, and continuous evolution in orbit, a set of digital system engineering methods is proposed. Through digital requirement decomposition and management, digital system design, digital collaborative design, and the integration of digital and physical verification in space and on the ground, this approach covers the entire life cycle of system requirements analysis, design and development, in-orbit operation, and evolution and upgrade, achieving digital empowerment throughout the process and effectively enhancing the efficiency and quality of system research and development. By applying this method in actual engineering projects, the system architecture design verification time is reduced from the expected three years to several months, the failure rate is lowered by 20%, and the maintenance time is shortened by 74%.

Key words: spacecraft electronic payloads; complex system; system engineering; digital collaborative design

0 引 言

当前航天领域的飞速发展对航天载荷的要求不断提高:承担的任务由单一变为多样,功能由固化变为可软件定义且可在轨升级,指标要求不断突破过往极限,但研制周期、成本却不断下降。新的发展趋势和需求不能再通过局部改进来满足,须从系统工

程基本的方法上有所突破。

“数字化手段”是近年来提升系统尤其复杂系统设计能力、研制效率和研制质量的重要手段之一^[1]。北京空间飞行器总体设计部建立了基于数字孪生的航天器系统工程整体模型和应用框架^[2]。中国运载火箭技术研究院先后开展了基于样机和基

* 收稿日期:2025-03-05;修回日期:2025-03-20
通信作者:孙杰 Email:sun.jie.sc@163.com

于模型的系统工程方法研究^[3-4],提出了基于数据驱动的航天产品工作方法。中国航天科技集团有限公司推行数字化设计技术,航天新研型号已实现用数字模装代替实物模装,数字化支撑“一站、两船、三箭”空间站工程建设^[5]。“北斗”系统构建了基于模型的任务保证总体架构,实现全周期风险精细化控制和智能运维^[6-7]。但是,目前各单位对数字化融入系统工程的方法大多处于探索阶段,没有形成一套较为公认行之有效的方法。

本文针对复杂航天电子载荷在论证、设计、研制、在轨运行等过程中的系统性问题,根据数字化融入系统工程的实际经验,探索出一套适合复杂航天电子载荷系统的数字化系统工程方法,取得了良好效果,为航天电子行业全面数字化转型提供了新的思路和方法。

1 方法总体框架

“载人航天大型综合化电子信息载荷”是一个典型的复杂航天电子载荷项目。该项目需要在一个通用电子平台上实现测控数传、网络通信、卫星 5G 等多种功能数十种模式应用。此类复杂航天电子产品在研制过程中面临以下难题:

- 1) 需求来源多且初期不明确,要求实现功能多且新技术应用多,系统设计难;
- 2) 空间环境复杂,工程环节繁多,实现难度大,系统验证难;
- 3) 在轨长期使用,系统持续演进,运行维护链条长,系统运营难。

面对以上难题,使用传统基于文档的系统工程方法(Document Based System Engineering, DBSE)存在以下困难:全新的系统形态使得设计需求短时间内难以明确,高的系统复杂度使得设计难以使用文本全面准确描述,众多的工作模式及状态使得故障态的关系、影响域、集成验证等分析困难,长期在轨及在轨升级使得在地面固化的系统状态必然改化。因此,亟需新的系统工程方法来满足新形态航天设备的研制需求。

本文提出的系统技术方法总体框架如图 1 所示,以标准、规范、平台、系统、工具和数据为基础,以数字化技术方法为手段,形成“需求分析、设计研制、运行监测、演进升级”闭环的系统工程方法。

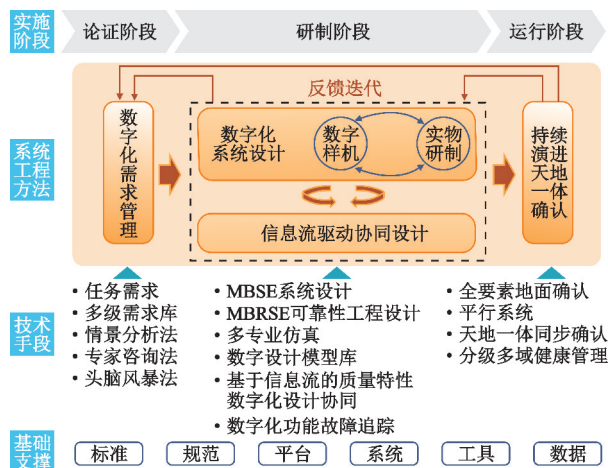


图1 总体框架

在需求分析阶段,全面收集包含使用场景、使用环境、资源要求等在内的各种需求,以数字化手段对需求进行统一管理和分析^[8],将各类需求逐步分析、分解,建立多层级需求库,避免需求的遗漏、矛盾、模糊等问题。

在设计研制阶段,首先通过基于模型的系统工程(Model-based System Engineering, MBSE)^[9]和基于模型的可靠性系统工程(Model-based Reliability System Engineering, MBRSE)^[10]实现对系统架构的综合优选,然后通过深入细致的 MBSE 建模、MBRSE 建模、各学科专业仿真,三者迭代设计、仿真、验证,实现数字样机。在完整数字样机的基础上才进行实物研制,完成测试、实物验证等。用完善充分的数字设计最大程度提升实物研制的效率和质量。

在在轨运行及演进升级阶段,通过数字样机、实物平行系统、真实设备三者间的同步、确认,实现在轨建设、任务运行、功能升级等的天地镜像运行,提高航天产品长期在轨能力。

2 基于数字化手段的需求管理与分析

传统研制方法对于研制早期大量且不够明确的需求进行管理较为困难。采用数字化的需求管理方法,首先可以规范统一需求表达方式,避免表达或理解误差;然后可以对来自不同单位、不同层级的需求进行管理,构建具有追溯关系的多层级需求库,既实现对系统需求的层层分解,又作为系统设计、分系统设计、系统验证、在轨运行与升级等的准确输入;最后,可以使得各个阶段的产品的研制成果不断与最

初的需求进行对照验证,确保所有需求被满足。

需求管理方法上,一般地,需求的收集应尽可能多,无论所收集需求是否重复甚至矛盾,尤其在系统论证早期。从离散、发散的需求到完整的设计输入由细致的需求分析完成。需求的分析大致可以按照系统、分系统、单机、模块 4 个层次递进,又进一步细分为多个需求库,如表 1 所示。需求库中的每个需求项至少包含“需求来源”和“需求分析结果”两个部分,上一级的需求分析结果可以作为下一级分析的需求来源,以实现需求追溯。不同层级需求分析一般由不同需求分析人员完成,且一般与系统设计人员相同。其次,由于没有完成实际设计,具体需求项的归属只是初步划分,最终归属关系随设计深入逐步调整。

表 1 典型需求库设置

序号	需求分析 层级	需求库
1	系统	系统任务、使用场景、使用环境、系统整体功能、工作模式、外部接口关系、系统整体性能、自主可控要求、设计建造规范要求、“六性”要求、质量体系要求、测试联试要求、交付物清单、研制计划、运营要求等
2	分系统	分系统功能、分系统性能、分系统间接口关系、分系统工作模式、分系统设计建造规范要求、分系统“六性”要求等
3	单机	单机功能、单机性能、单机间接口关系、单机工作模式、单机设计建造规范要求、单机“六性”要求等
4	模块	模块功能、模块性能、模块间接口关系、模块设计建造规范要求、模块“六性”要求等

在工具层面,应使用利用专业的需求管理工具对庞杂的需求进行管理。应选择具有协同编辑、富文本编辑、自动通知、通知评审等功能的专业需求管理工具,通过工具辅助实现对需求的分析、分解,最终实现对设计的全面准确输入和有据可查。

3 数字化系统设计及数字样机

面向复杂航天载荷系统新技术多、设备构成及工作关系复杂等系统设计难题,构建 MBSE、

MBRSE、专业仿真和实物“四位一体”的系统设计方法,如图 2 所示。首先,采用 MBSE 和 MBRSE 方法论,利用相关工具软件,迭代式开展系统架构设计。MBSE 完成对系统组成、系统互联关系、工作流程等的建模及运行逻辑仿真验证,MBRSE 完成系统故障行为、故障机理等的建模及仿真分析。后者对前者的设计提出改进建议,两者经过多次迭代实现对系统架构的优选。之后,进一步深入对分系统、单机、模块进行 MBSE 和 MBRSE 建模仿真,同时进行电路、力、热等各专业的设计和建模仿真。各专业仿真结果用于完善 MBSE 和 MBRSE 模型,主要是完善模型中相关性能指标。

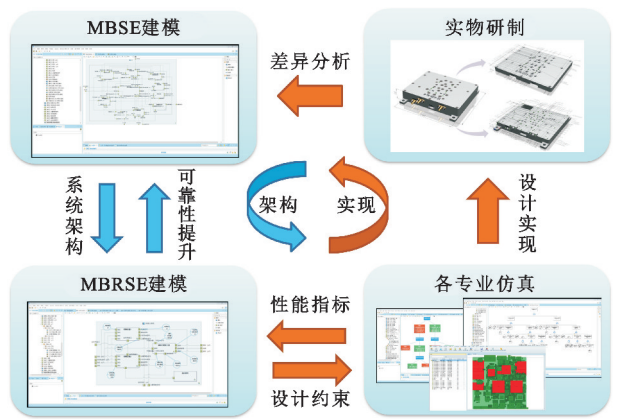


图 2 “四位一体”系统设计方法

MBSE 模型、MBRSE 模型和专业设计模型三者共同组成系统设计的数字样机。数字样机建模颗粒度依据对应实物成熟度调整,成熟度高的可以仅建模单机级模型,成熟度低的一般应建模至模块级及以下。

应在搭建完完整的数字样机之后才开始实物研制。因为即使成熟的产品模块,在不同的系统中的适用范围和约束也不同,可能对系统整体产生意料之外的影响,经过充分的数字样机验证能够最大程度避免实物研制的重复并节约联试时间。其次,在实物研制过程中,应逐模块地将实物功能性能结果从 MBSE 模型开始与数字模型进行对比。如果实物与数字模型不相符,应进行差异分析,重新实物研制或修改数字模型,始终保持实物和数字模型的一致。

MBSE 和 MBRSE 所建主要模型如表 2 和表 3 所示,表中以系统级模型为例,分系统级、单机级、模块级模型类似。专业仿真模型按常规方法进行,不再赘述。

表 2 主要 MBSE 模型

序号	模型名称	模型含义及作用
1	系统需求映射模型	在需求分析结果的基础上,建立系统组成与系统需求的映射表,使得系统设计从基本组成上就能够满足系统需求
2	系统用例模型	即系统的使用场景,或同一使用场景下的不同工作模式。不同用例下系统最基本的对外接口、工作流程等可以不同
3	系统活动模型	以系统整体为单位,与外部单位间的工作流程。此模型用于明确系统整体的功能和外部接口
4	系统组成模型	系统组成模型主要用于明确系统架构,是从系统需求和系统活动中直接设计出的模型,因此可以是逻辑上的系统组成,再从逻辑组成上拆分或组合出实际的分系统
5	系统内部互连关系模型	以系统外部接口、系统组成为基本输入,设计完成系统内部接口和互连关系
6	系统状态模型	系统在工作中的不同状态,包含不同状态、状态转移条件、时序等
7	系统指标模型	将系统级指标继承为分系统指标,或添加新的指标,并建立专门的指标计算模型,在分系统设计完成后,可自动计算系统指标

表 3 主要 MBRSE 模型

序号	模型名称	模型含义及作用
1	故障行为过程模型	基于系统组成模型,结合系统故障原理和信号传递关系,对系统的失效事件、失效状态、失效发生后的失效传递过程和逻辑等进行规范化表征,生成完整的故障传递路径
2	危险分析模型	基于故障行为过程模型,开展安全性设计分析,基于信号端口自动生成故障树模型,并自动计算事故发生概率
3	故障分析模型	基于故障行为过程模型,开展可靠性设计分析,基于任务阶段自动生成故障模式影响及危害性分析结果,基于信号端口自动生成可靠性模型,并自动计算可靠度
4	故障诊断模型	基于故障行为过程模型,开展测试性设计分析,自动生成故障诊断数字模型(包括诊断树、相关性矩阵),自动计算故障隔离率、故障检测率,定位测试性设计薄弱环节,提出改进建议
5	故障机理模型	基于数字设计模型中的热仿真模型和振动仿真模型,确定失效物理模型,进行各任务剖面下累积损伤分析,定位可靠性设计薄弱环节及其失效机理,提出改进建议
6	维修保障模型	基于数字样机模型,开展维修性综合设计定量评价,定位维修性设计薄弱环节,提出改进建议;基于系统组成模型,进行保障方案优化

在具体工具选择层面,MBSE 软件应选择支持 SysML 语言,支持 MATLAB 等语言脚本,具有用例图、块图、活动图、顺序图、状态机图等的软件。MBRSE 软件应选择能够与 MBSE 软件互通,能够使用图形化方式对故障行为过程进行建模,能够自动化进行危险分析、故障分析、故障机理等分析的软件^[11]。

以上方法可以极大提高系统架构设计能力,尤其使得新技术的应用得以“有法可依”,从而在系统架构层面就保证系统的良好性。使用该方法,使得项目组综合优选出“开放式可成长系统架构”技术方案(硬件架构采用高速信号网络传输结合通用信号模块池,软件架构采用构件化、开放式分层架构),打破了传统功能对硬件的紧耦合,赋予系统通用开放、迭代升级能力。

同时,该方法解决了传统系统研制难以在实物研制前进行验证的难题,使得在实物研制前就经过了系统级仿真验证,从而能够最大程度上减少实物研制重复,减少联试时间,降低系统研制在技术、经济、进度等各方面的风险,保证了设计与实现的一致性,避免出现系统组装后才出现系统指标不达标的情况。

4 信息流驱动的数字化工协同设计

完整的系统研制需要集合研制过程各阶段数据进行综合分析、不断迭代。采用信息流驱动的数字化工协同设计方法,以系统研制所涉及的各种信息贯穿整个研制工作,有效识别和控制研制的技术风险^[12]。信息流按设计控制、故障分析、试验验证三个方面分为设计信息流、故障信息流、试验信息流,统称系统信息流,如图 3 所示。设计信息流由功能、性能设计环节的信息组成。故障信息流由故障的起因、影响和传递等环节的信息组成。试验信息流由实验室试验、测试、联试等环节的信息组成。

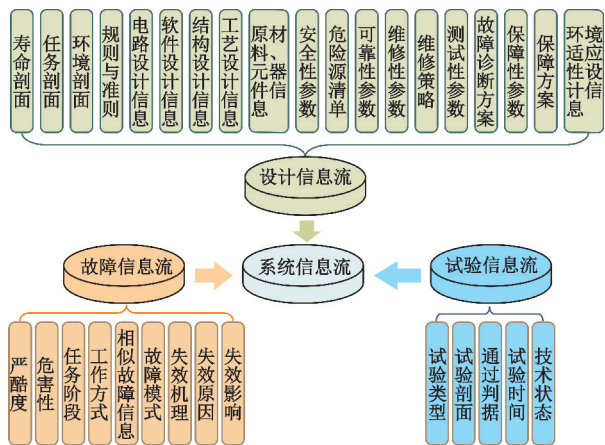


图 3 系统信息流

系统信息流设计分概要设计和详细设计两个阶段开展,具体方法如图 4 所示。每个设计阶段均包括系统分析、系统定义和系统设计三方面的工作:通过系统分析,根据系统总体和用户使用要求,确定产品功能、性能、接口和环境等输入信息流;通过系统定义,根据系统分析结果,明确输入信息流的属性;通过系统设计,综合各类信息流资源,全面验证和评估设计水平。

概要设计阶段面向任务和功能设计,主要开展需求分析、定量指标分解、定性要求确定。以可靠性工程设计为例,首先在“可靠性工程设计验证系统”中创建工程项目,然后从“产品数据管理系统”获取产品初始物料清单表,建立产品树结构,自顶向下开展系统至组件级的功能分析、任务分析、结构分析以及寿命剖面分析,确定各级产品具体的功能及特性要

求、任务时间、使用环境以及整个寿命周期经历的事件。完成系统分析和系统定义后,通过开展系统概要设计和分析,确定定量指标、安全性可靠性关键项目、在轨可更换单元等,制定试验方案和保障方案。

详细设计阶段面向物理设计和故障防控,主要开展各级产品定量设计、分析和评价,以清晰梳理出系统设计全貌。以可靠性工程设计为例,从功能和性能设计角度分析各层级设计、信号流、接口类型等,从试验角度分析试验方法和应力,从故障角度分析故障原因、故障传递关系、故障诊断措施、维修活动等。基于全面的系统定义,结合力学、热学、电磁等各学科的仿真建模,构建可靠性、维修性、测试性、保障性、安全性数字模型,定量确定故障发生概率及危害度,采用专项研制试验和系统研制试验相结合方式,充分暴露设计缺陷,持续优化设计并制定故障预案。

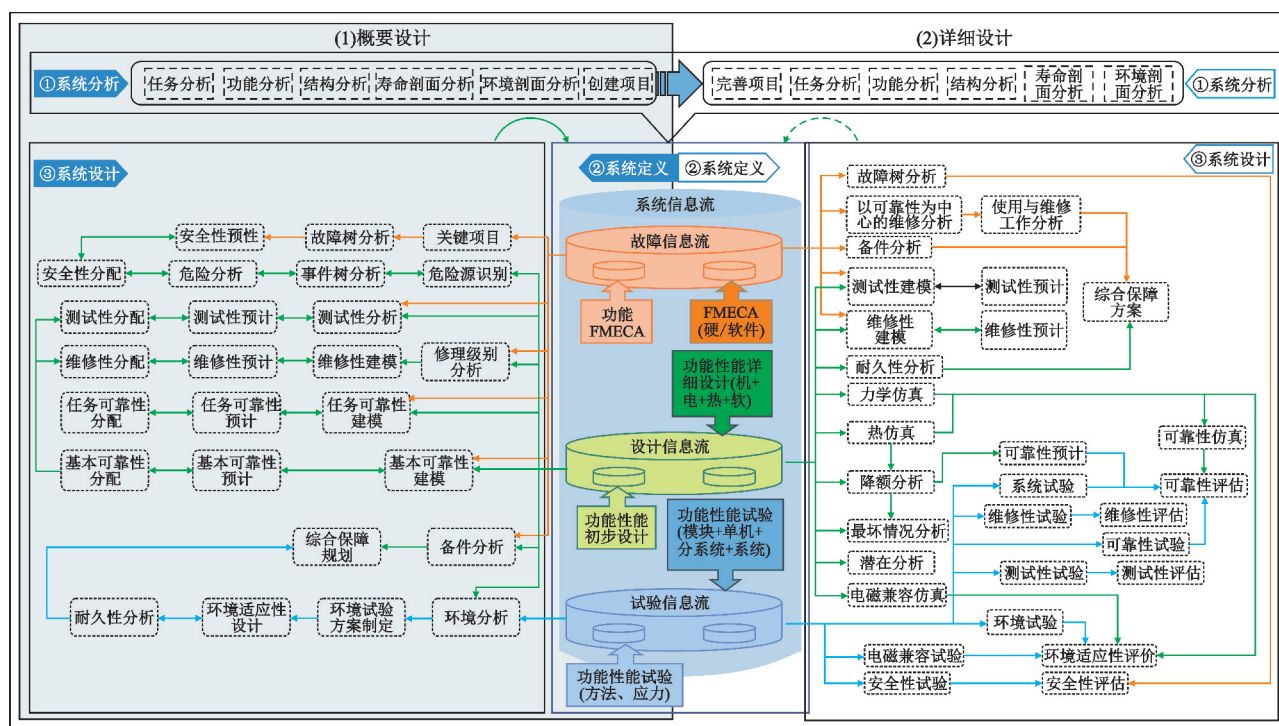


图 4 信息流驱动的数字协同设计方法

以上方法需要收集各层级系统设计结果、各种专业仿真结果、各种测试验证结果等,以在一个平台上实现系统信息流的流动。本文所选择的“可靠性工程设计验证系统”可直接调用主流 MBSE 系统设计软件、电路设计软件、力学设计软件、热学设计软件、物资信息等的设计结果,但部分结果仍需手动输入。后续需打通与更多工程设计软件的接口,增强研发数据信息自动收集功能,健全数据信息集成管理机制,从而建立更加全面、真实、有效的功能性能信息库,增强实物设计和数字设计之间的数据同源

性和设计统一性。

5 持续演进天地一体确认

传统航天器一般没有在建过程也不具备在轨升级能力,在地面完成各种测试确认技术状态后,发射升空直接使用。随着航天器功能日益多样化、任务日益复杂化,航天器在轨建设、在轨升级已经成为必然趋势,传统地面确认所有技术状态的做法不再适用。

本文采用虚实结合、模型驱动的设计理念,遵循“综合化平台+可扩展功能”的技术体制,以产品实装和数字样机为基础,建立数实结合的地面平行系统,基于天地状态同步、任务试验自动评估、故障自动分析诊断等关键技术,实现对长期在轨、在轨建设、持续演进航天电子载荷的天地一体确认。

天地一体的系统确认流程如图 5 所示。

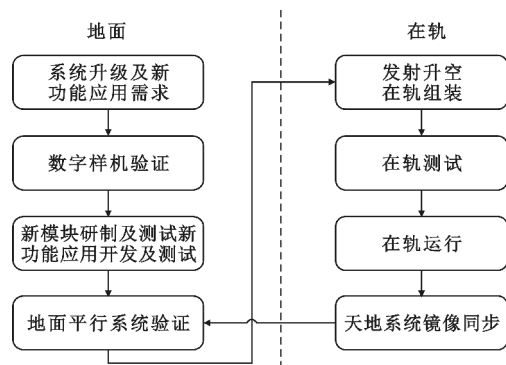


图 5 天地一体同步确认流程

1) 首先,当有了系统升级或新功能应用的需求后,对新的需求进行需求管理和分析,建立需求模型。

2) 在数字样机中新建相关模型,并进行全系统流程、功能验证。

3) 在数字样机验证无误之后,开始研制相关系

统模块或开发新的功能应用代码,并完成模块级或单机级测试确认。

4) 启动天地镜像同步,根据真实在轨系统的遥测信息,更新地面实物平行系统的状态及参数,使得地面平行系统能够真实还原在轨系统的全部状态。

5) 将新开发的系统模块或功能应用代码,放入地面平行系统中进行确认。除需重新完成全部系统级测试外,还将对新的功能应用完成全部工作模式测试确认,确保在轨后工作无误。

6) 相关模块搭载上行或功能应用代码重新上注,在轨进行组装、升级。

7) 在完成在轨测试之后,正常执行新的任务。

为提高天地状态同步的效率,减少天地数据传输的压力,提出了分级多域健康管理分析技术,实现从元器件到系统的健康管理。采用模块、单机、系统三层级的联邦管控架构,对元器件、总线、模块、单机、分系统、系统等 6 个层级的健康数据进行采集、处理和分级管理,实现 70% 健康管理数据在轨自主处理,有效提高系统天地一体健康管理效率。通过对不同层级健康数据分级提取各向量关键特征数据并进行综合分析,完成产品运行状态、环境、能源、硬件、软件等多维度的健康监视与观察。分级多域监控管理架构如图 6 所示。

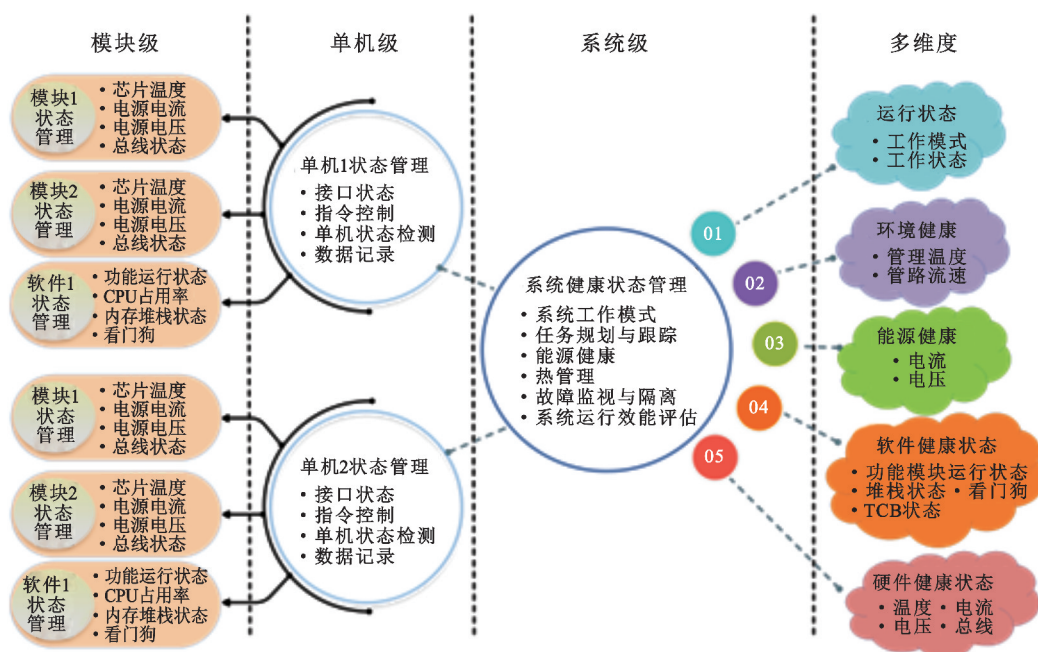


图 6 分级多域健康管理架构

以上方法可以有效解决天地环境迥异和在轨建设、在轨升级带来的航天电子产品地面验证不准确、

不充分,以及无法验证的问题,使得地面做出的在轨建设、在轨升级方案和故障诊断方案等有据可依、有

数据可依,且经过充分验证。

6 实施效果

通过开展以上系统工程技术方法研究与应用,在“载人航天大型综合化电子信息载荷”的研制过程中,成功地在系统架构设计的科学性与时效性、系统故障消减与控制、在轨运行安全可靠保障等方面实现了有效提升。

在系统架构设计的科学性与时效性方面,采用数字化需求管理和基于 MBSE 的数字化系统设计,在用户需求、技术需求等不够明确且变化的情况下,将“载人航天大型综合化电子信息载荷”这一新型载荷系统的系统架构设计确认时间从预计的 3 年缩减至数月,并从架构层面确保了后续需求修改、功能实现的可扩展性和可靠性。

在系统故障消减与控制方面,采用 MBRSE 方法,结合基于 MBSE 的数字样机、多学科设计数字模型、试验数据等进行故障行为数字化追溯,建立起了覆盖元器件、部组件、单机和系统的航天电子载荷故障追踪信息库。面向各种任务模式,利用数字化环境,深入识别在安全性、可靠性、测试性、维修性、保障性等方面的设计薄弱环节,形成优化的功能架构方案、故障诊断方案、维修保障方案辅助实物设计改进。从原理样机到正样,将本项目载人航天大型综合化电子信息载荷的严重事故发生概率由 1.6×10^{-7} 降低到 2×10^{-9} ,失效率下降 20%,维修时间缩短 74%,故障检测率提升 36%,故障隔离率提升 61%,备件量压缩了 30%。

在在轨运行安全可靠保障方面,采用覆盖载荷全生命任务周期的联邦式运行管理框架,以任务为驱动,将系统、子系统和模块三级分别由管控子系统、平台/功能子系统和模块/功能软件进行运行过程管理,实现小核心治理大平台,将系统管理的效率提高了 70%。采用自治与信息差异化处理方法,完成了全系统在升级、异常、紧急断电等不同场景下的高效运行处置。系统从发射至今,成功运行于在轨安装、在轨测试与日常运行等各种设备状态,成功支持任务上注与新功能插入试验等,并成功处置了在轨紧急告警,保障了载荷全系统的安全性和可靠性。

7 结束语

本文针对复杂航天电子载荷的设计研制问题,提出了一套面向全系统、全过程、全寿命的数字化系统工程技术方法。该方法贯穿航天电子载荷需求分析、设计研制、在轨运行、演进升级的全生命周期。通过数字化的手段,对用户需求、技术需求等进行详尽的

分解和管理,并可在研制、运行任意阶段进行追溯。构建 MBSE、MBRSE、专业仿真和实物“四位一体”的系统设计方法,实现对系统架构的优化设计,实现系统设计在实物研制前的验证,保证设计和实现的一致。采用信息流驱动的数字化协同设计方法,使用设计信息流、故障信息流、试验信息流贯穿整个研制工作,增强实物设计和数字设计之间的数据同源性和设计统一性。最后建立数实结合的地面平行系统,实现天地状态同步、任务试验自动评估、故障自动分析诊断等,实现了航天产品的天地一体确认、长期在轨运行稳定和在轨持续演进。

本文形成了针对复杂航天电子载荷的数字化系统工程的基本方法,后续将在此基础上进一步细化深入,并以此方法为路径,探索 and 开发适应其他产品系列的系统工程方法。

参考文献:

- [1] 曾照洋,彭文胜,任占勇,等. 大数据时代装备质量工作新模式[J]. 航空标准化与质量,2021(4): 16-22.
- [2] 王建军,向永清,何正文. 基于数字孪生的航天器系统工程模型与实现[J]. 计算机集成制造,2019,25(6): 1348-1360.
- [3] 王小军. 基于样机与模型的系统工程探索研究[J]. 导弹与航天运载技术,2022(5): 1-6.
- [4] 杨小龙,聂蓉梅. 运载火箭基于模型的系统工程研究[J]. 宇航总体技术,2024,8(1): 1-7.
- [5] 杨双进. 航天工程质量管理数字化创新与实践[J]. 中国质量,2023(3): 30-33.
- [6] 卿寿松,郑恒,周波,等. 基于模型的航天工程任务保证理论和技术体系研究[J]. 质量与可靠性,2023(5): 1-8.
- [7] 卿寿松,郑恒,角淑媛,等. 北斗卫星导航系统质量与可靠性工程实践[M]. 北京:中国宇航出版社,2024: 416-443.
- [8] ZHUANG C B, LIU Z W, LIU J H. Digital twin-based quality management method for the assembly process of aerospace products with the grey-Markov model and apriori algorithm[J]. Chinese Journal of Mechanical Engineering, 2022(5): 1-21.
- [9] ESTEFAN J A. Survey of model-based systems engineering (MBSE) methodologies[R]. San Diego: INCOSE, 2008: 1-70.
- [10] 任羿,王自力,杨德真,等. 基于模型的可靠性系统工程[M]. 北京:国防工业出版社,2024: 35-64.
- [11] 吴慧伦. 基于 MBSE 的微波统一测控系统测试性设计[J]. 电讯技术,2023,63(6): 811-816.
- [12] 吴慧伦,柴霖,李威. 空间站有效载荷通用质量特性协同设计[J]. 载人航天,2023,29(3): 398-406.

作者简介:

赵晓虎 男,1975 年生于四川阆中,1997 年获工学学士学位,现为高级工程师,主要从事经营管理和项目管理工作。

孙杰 男,1982 年生于四川绵阳,2012 年获博士学位,现为高级工程师,主要研究方向为测控通信网络、卫星互联网等。

吴慧伦 女,1971 年生于江苏南京,1994 年获工学学士学位,现为高级工程师,主要研究方向为航天测控通信、可靠性系统工程。

柴霖 男,1969 年生于河南洛阳,2005 年获博士学位,现为研究员,主要研究方向为航天测控通信、综合化载荷。

DOI:10.20079/j.issn.1001-893x.240422005

基于有限方向随机游走的火箭发射测控机动布站方法^{*}

董 浩

(西南电子技术研究所,成都 610036)

摘 要:针对极端对抗条件下航天应急发射机动测控能力难以量化评估的问题,提出了一种基于有限方向随机游走的火箭发射测控机动布站方法,以最大测控覆盖和最小站间重叠为目标,通过多机动站站间重叠情况限制站址随机游走方向,迭代优化生成最终机动部署方案。仿真结果表明,所提方法在给定火箭弹道下能够有效确定任意数量机动站部署位置,在测控覆盖率、站间重叠率、站址接受率等方面均优于对比方法,当机动站数量大于 5 时,对典型火箭弹道的测控覆盖率达到 98% 以上,相对于正常发射情况下全段测控要求,可有效支撑航天应急发射实施。

关键词:应急发射;测控;机动布站;随机游走

开放科学(资源服务)标识码(OSID):



中图分类号:V556;TN98

文献标志码:A

文章编号:1001-893X(2025)03-0392-06

A Mobile Deployment Method for Rocket Launch TT&C Stations Based on Random Walk in Finite Direction

DONG Hao

(Southwest China Institute of Electronic Technology, Chengdu 610036, China)

Abstract: Since it is difficult to quantitatively evaluate the mobile telemetry, tracking and command (TT&C) capability of space emergency launch under extreme confrontation conditions, a mobile deployment method for rocket launch TT&C stations based on random walk in finite direction is proposed. With the maximum coverage rate of TT&C and the minimum overlap rate between stations as optimal goals, the random walk direction of a station is limited according to the overlap between multiple mobile stations. The final deployment scheme of mobile stations is iteratively optimized. Simulation results show that the proposed method can effectively determine the positions of any number of mobile stations with a given rocket trajectory, and performs better than the comparison method in terms of coverage rate of TT&C, overlap rate between stations, acceptance rate of station, etc. When the number of mobile stations is greater than 5, the coverage rate of TT&C for typical rocket trajectory can reach more than 98%, which can effectively support space emergency launch compared with the full-section TT&C requirements under normal launch conditions.

Key words: emergency launch; TT&C; mobile deployment; random walk

0 引 言

当前,太空领域大国竞争愈加激烈,占领太空战略制高点具有重要意义。早在 2002 年,美国就提出

“快速响应空间”理念,旨在构建其快速发射补星能力,保证太空系统弹性持续可用^[1]。近年来,随着航天技术的巨大进步,进入和利用空间的门槛大幅

^{*} 收稿日期:2024-04-22;修回日期:2024-06-02

通信作者:董浩 Email:donghao@whu.edu.cn

降低,卫星发射频率及在轨卫星数量急剧增加,世界主要航天大国都在规划和建设大型低轨星座,典型代表有 OneWeb、StarLink、“鸽群”等,空间轨道和频段资源抢占态势明显,全球战略布局意义重大,并在现代局部战争或武装冲突中表现出了强大态势感知和信息支援能力^[2-4]。

航天发射测控力量是确保航天器快速进出空间的关键基础。我国已形成包括国内外陆基测控站、海基测量船、天基中继卫星在内的陆海天基一体化测控网络,为包括火箭发射在内的航天测控任务的圆满完成提供了有力支撑^[5-6],60 多年来,已成功保障 400 多次长征系列火箭发射任务^[7]。航天应急发射能力是我国航天发射能力的重要组成,是航天对抗环境下的必要举措,主要针对战时地面固定测控系统或装备易受打击而无法使用等情况,在火箭发射段进行机动测控,最大限度满足测控要求,实现航天器快速入轨或星座补网组网,是保持我国航天应急发射能力的关键举措^[8-9]。Zhao 等人^[10] 研究分析了美俄等国在快速响应运载器、快速发射基础设施等方面发展现状,强调航天应急发射对太空对抗系统的自恢复和自增强能力的决定作用。在体系建设方面,刘兴威等人^[8] 明确提出快速发射航天器已成为突发事件背景下确立空间优势的首要条件,并从体系结构、运行模式、组织结构、测试发射等方面对航天应急发射体系建设进行了分析论证;路建功等人^[11] 则在航天力量使命任务分析的基础上,提出航天发射测控力量体系化建设方法,并建议尽早推进对抗条件下按需发射、快速发射、应急发射和应急补网等能力。

在具体方法方面,陈莉丹等人^[11] 提出面向小卫星机动发射的两步测控布站优化方法;崔健等人^[12] 考虑多设备交会精度,提出基于几何精度因子(Geometric Dilution of Precision, GDOP)和设备约束的测控合理性评估方法,分析了火箭发射测控布站区域选择原则;叶君好等人^[13] 针对测控阵地勘察要素多、多因素交叉等特点,综合考虑技术和环境因素,建立了基于层次分析法的航天测控阵地勘察模型;Wu 等人^[14] 提出敏捷航天应急发射任务规划仿真验证方法,建立资源有限条件下多场域、多平台和多任务并行的航天应急发射任务规划的总体技术架构,并对发射弹道优化、多点发射碰撞计算、资源约束下发射任务规划等关键问题进行仿真分析。

目前相关文献主要从体系层面对航天应急发射

时机动测控必要性进行了论证,具体方法偏向于理论模型的构建,较难量化评估对抗条件下航天应急发射时机动测控能力限度。本文在航天应急发射机动测控需求分析的基础上,提出基于有限方向随机游走的火箭发射测控机动布站方法,以最大测控覆盖和最小站间重叠为目标,在前序机动站点位置确定的有限方向内迭代随机生成最新站点位置,并量化分析了典型火箭发射弹道的机动测控能力,为我国航天应急发射能力建设提供借鉴和支撑。

1 航天应急发射机动测控需求分析

随着我国航天事业快速发展,不同运载能力、发射方式、入轨形式的运载火箭投入使用,具备任务准备周期短、发射点位及射向变化大等特点,要求测控系统具备快速发射适应能力和大范围、全航程测控覆盖能力^[6]。虽然地面固定测控站、天基测控卫星(中继卫星、导航卫星等)覆盖范围大、测控能力强,但在对抗条件下,由于易受到攻击或干扰等原因,会部分或完全丧失测控能力,需要通过机动站保障火箭应急发射实施。航天应急发射机动测控系统组成如图 1 所示。

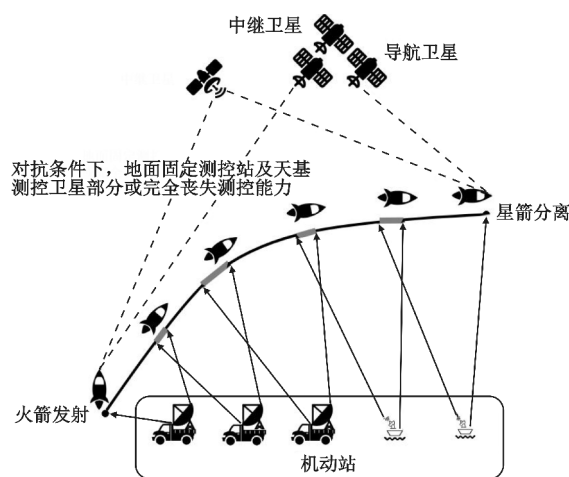


图 1 航天应急发射机动测控系统组成

机动测控是航天应急发射的关键环节,不可或缺,在满足航天应急发射测控需要的前提下,测控覆盖越高越好,通常采用多站接力跟踪的方式进行保障,其核心是布站优化目标在约束条件下的非线性规划问题。多机动站布站优化求解过程和方法,需要具备良好的收敛性和适应性,同时要减少站间重叠,避免多机动站集中分布,以提升测控装备利用能

力。因此,应以最大测控覆盖为优化目标并兼顾最小站间重叠,充分考虑多机动站间相互影响的特点,根据火箭发射理论弹道快速确定多机动站的站址位置,为实施航天应急发射和卫星快速入轨提供保障和支撑。

2 基于有限方向随机游走的火箭发射机动测控布站方法

2.1 机动测控布站优化目标

假设某次航天应急发射任务中,火箭发射和星箭分离时刻分别为 T_0 和 T_E ,火箭发射任务持续总时间为 $T=T_E-T_0$ s。进一步假设可用机动站数量共 N 个,机动站天线波束具备自动跟踪能力,最大作用距离满足面向低轨卫星的火箭发射测控需求。在满足最大工作仰角大于 5° 要求下,第 n 个机动站与运载火箭发射弹道的可见弧段为 $TW_n = \{tws_n, tve_n\}$,其由该机动站的具体部署位置 (Lat_n, Lon_n) 与火箭发射弹道决定, tws_n 与 tve_n 分别表示可见弧段 TW_n 的开始时间、结束时间。本文通过卫星工具包软件 (Satellite Toolkit, STK) 完成可见弧段计算。根据前述需求分析,航天应急发射机动测控布站的优化目标是在满足可见等条件下,机动站的部署位置可使运载火箭的测控覆盖率尽可能大,同时不同机动站间测控弧段重叠尽量少。构建航天应急发射机动测控布站优化模型如下:

1) 测控覆盖率 (Coverage Rate of Telemetry, Tracking and Command, CRT), 定义为各站点测控时间的并集与火箭发射任务持续总时间的比值。该指标越大越好, 目标函数如下:

$$\max F_1 = \frac{\bigcup_{n=1}^N \{TW_n\}}{T} \quad (1)$$

2) 站间重叠率 (Overlap Rate between Stations, ORS), 定义为各站点间测控时间的交集与火箭发射任务持续总时间的比值。该指标越小越好, 目标函数如下:

$$\min F_2 = \frac{\bigcap_{n=1}^N \{TW_n\}}{T} \quad (2)$$

2.2 随机游走有限方向范围计算

单站对火箭测控弧段的长短由其部署位置与火箭发射弹道间的空间几何关系决定,同时为了避免站间测控弧段的重叠,不同机动站间的部署位置也相互影响。航天应急发射场景下,机动测控布站优

化模型无明显解析解,本文提出一种基于有限方向随机游走的求解方法,其本质是一个搜索计算的迭代过程,关键是确定每个机动站点的搜索方向。

假设火箭发射位置为点 S ,星箭分离点位置为点 E ,第 n 个机动站点当前位置为点 O ,令 $\angle SOU = \angle SOE/2$, $\angle SOV = \angle SOE/2 + 180^\circ$ 。如果其他 $N-1$ 个机动站中有 N_{ol} ($1 \leq N_{ol} \leq N-1$) 个与当前站点存在重叠,并设重叠中心为点 P ,点 P 关于点 O 的对称点为 Q ,如图 2 所示。

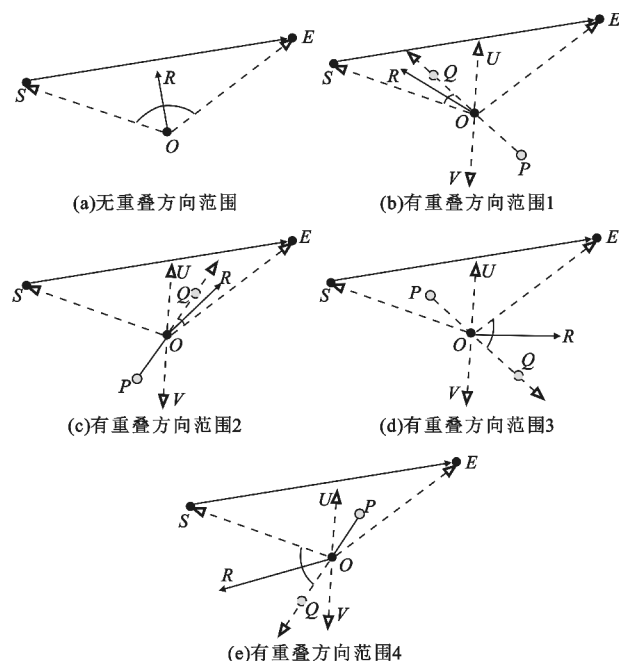


图 2 随机游走有限方向分析

$\angle SOE$ 、 $\angle SOQ$ 等角度的计算方法可以归纳为已知规定正方向下任意两个三维向量夹角计算问题。 $\angle SOU$ 、 $\angle SOV$ 与 $\angle SOE$ 存在固定几何关系,可直接计算。定义发射位置点 S 、星箭分离位置点 E 、第 n 个机动站点当前位置点 O 、重叠中心点 P 的三维坐标分别为 (x_s, y_s, z_s) 、 (x_e, y_e, z_e) 、 (x_o, y_o, z_o) 、 (x_p, y_p, z_p) , 对应的经纬高分别为 (lon_s, lat_s, alt_s) 、 (lon_e, lat_e, alt_e) 、 (lon_o, lat_o, alt_o) 、 (lon_p, lat_p, alt_p) , 由于点 Q 与点 P 关于点 O 对称,有 $\vec{OQ} = -\vec{OP} = (x_o - x_p, y_o - y_p, z_o - z_p)$, 并规定 $\vec{OS}$ 旋转至 $\vec{OE}$ 且符合右手螺旋准则的垂直方向为正方向 $\vec{w} = \vec{OS} \times \vec{OE}$, 且 $\vec{w}$ 对应归一化向量为 $\vec{u} = \frac{\vec{w}}{\sqrt{\vec{w} \cdot \vec{w}}}$ 。 $\angle SOE$ 、 $\angle SOQ$ 、 $\angle SOU$ 、 $\angle SOV$ 等角度计算结果如公式

(3)~(5)所示:

$$\begin{cases} \vec{v} = \vec{OS} \times \vec{OE} \\ \xi = \vec{v} \cdot \vec{w} \\ \alpha = \arctan\left(\frac{\|\vec{v}\|_2}{\|\vec{OS} \times \vec{OE}\|_2}\right) \\ \angle SOE = \begin{cases} \alpha, & \xi \geq 0 \\ 360 - \alpha, & \xi < 0 \end{cases} \end{cases} \quad (3)$$

$$\begin{cases} \vec{v} = \vec{OS} \times \vec{OQ} \\ \xi = \vec{v} \cdot \vec{w} \\ \alpha = \arctan\left(\frac{\|\vec{v}\|_2}{\|\vec{OS} \times \vec{OQ}\|_2}\right) \\ \angle SOQ = \begin{cases} \alpha, & \xi \geq 0 \\ 360 - \alpha, & \xi < 0 \end{cases} \end{cases} \quad (4)$$

$$\begin{cases} \angle SOU = \frac{\angle SOE}{2} \\ \angle SOV = \frac{\angle SOE}{2} + 180^\circ \end{cases} \quad (5)$$

在有站间重叠的情况下,向量 $\vec{OS}$ 、 $\vec{OE}$ 、 $\vec{OU}$ 及 $\vec{OV}$ 将球面分成了 4 个区域。为避免站间重叠,当前机动站点应向远离点 P 的方向部署,即点 S 、点 E 、点 O 及点 P 间的相对位置关系决定了当前机动站点随机行走的有限方向。如图 2 所示,根据有无重叠以及存在重叠情况下点 P 位置,当前机动站点随机行走的有限方向范围的确定分以下 5 种情况:

1) 无重叠时,当前机动站点随机行走的有限方向为 $[0^\circ, \angle SOE)$,如图 2(a) 所示;

2) 有重叠且 $0^\circ \leq \angle SOQ < \angle SOU$ 时,当前机动站点随机行走的范围为 $[0^\circ, \angle SOQ)$,如图 2(b) 所示;

3) 有重叠且 $\angle SOU \leq \angle SOQ < \angle SOE$ 时,当前机动站点随机行走的方向范围为 $[\angle SOQ, \angle SOE)$,如图 2(c) 所示;

4) 有重叠且 $\angle SOE \leq \angle SOQ < \angle SOV$ 时,当前机动站点随机行走的方向范围为 $[\angle SOE, \angle SOQ)$,如图 2(d) 所示;

5) 有重叠且 $\angle SOV \leq \angle SOQ < 360^\circ$ 时,当前机动站点随机行走的方向范围为 $[\angle SOQ, 360^\circ)$,如图 2(e) 所示。

2.3 机动站随机游走及站址确定

根据有无重叠以及存在重叠情况下点 P 位置,分前述 5 种情况确定单机动站点随机行走的方向范

围为 $[\gamma_{\min}, \gamma_{\max})$,按均匀分布随机生成角度 $\theta = \angle SOR \in [\gamma_{\min}, \gamma_{\max})$,向量 $\vec{OR}$ 即为当前机动站点随机游走的方向,相当于向量 $\vec{OS}$ 绕规定的正方向 $\vec{w}$ 旋转 $\theta = \angle SOR$ 角度后的方向。在三维空间中,按公式(6)采用罗德里格斯(Rodrigues)旋转公式计算确定旋转后矢量^[15] $\vec{OS}_{\text{rotated}}$,进而可根据公式(7)计算得当前机动站点在本次迭代后的站址位置 R 的坐标为 (x_R, y_R, z_R) ,对应的经纬高 $(\text{lon}_R, \text{lat}_R, \text{alt}_R)$ 。

$$\vec{OS}_{\text{rotated}} = \vec{OS} \cos \theta + \vec{u}(\vec{u} \cdot \vec{OS})(1 - \cos \theta) + (\vec{u} \times \vec{OS}) \sin \theta \quad (6)$$

$$\begin{cases} \vec{OR} // \vec{OS}_{\text{rotated}} \Leftrightarrow \vec{OR} \times \vec{OS}_{\text{rotated}} = 0 \\ \|\vec{OR}\|_2 = d \end{cases} \quad (7)$$

式中: d 为步进距离。其他站点位置采用同样方式在有限方向内随机生成。对于本轮迭代 k 中随机生成的 N 个新站址是否接受,根据目标函数是否得到优化来判断,即当测控覆盖率(目标函数 F_1)提升或测控覆盖率相同但站间重叠率(目标函数 F_2)下降时,则接受 N 个新站址,否则拒绝并进入下一轮迭代。站址接受准则具体如公式(8)所示:

$$\begin{aligned} \text{accepted} &= 1 \\ \text{s. t. } (F_1^k > F_1^{k-1}) \parallel [(F_1^k = F_1^{k-1}) \& (F_2^k < F_2^{k-1})] \end{aligned} \quad (8)$$

进一步定义站址接受率(Acceptance Rate of Station, ARS),以便于后续分析评估算法性能。站址接受率定义为随机生成站址被接受的迭代次数 k_{accepted} 与当前迭代总次数 k 的比值,具体如公式(9)所示:

$$\text{ARS} = \frac{k_{\text{accepted}}}{k} \times 100\% \quad (9)$$

3 仿真验证与分析

本文采用火箭弹道模拟数据对基于随机行走的火箭发射测控布站分析方法进行仿真验证和结果分析,并与无限制的随机布站方法进行对比。火箭弹道模拟数据根据典型低轨卫星轨道特点仿真生成,其发射点为 $(98.57^\circ, 39.77^\circ, 0 \text{ km})$,星箭分离点为 $(173.43^\circ, -8.10^\circ, 550 \text{ km})$,初始时所有机动站部署在发射点位置。在仿真验证时,分别设置机动站数量为 $1:1:7$ 个,最大迭代次数为 $5:5:100$,相互组合后重复执行 5 次,以分析机动站数量、最大迭代次数对布站结果的影响。

图 3 和图 4 分别显示了本文方法中测控覆盖率和站间重叠率随最大迭代次数的变化情况。总体上,测控覆盖率随着最大迭代次数的增加而升高,最大迭代次数小于 30 次时,测控覆盖率升高较快,达到 50 次以上后趋于稳定;同时随着机动站点数量的增多,测控覆盖率也明显增大但增幅逐渐收窄,数量大于 4 个时对典型火箭弹道的测控覆盖率达到 95% 以上,数量大于 5 时,达到 98% 以上,相对于正常发射情况下全段测控要求,可有效支撑航天应急发射实施。对于站间重叠率,其随最大迭代次数增加呈下降趋势但随着机动站点数量的增多而明显增加,但为了优先保障测控覆盖率优化目标,存在明显的波动。

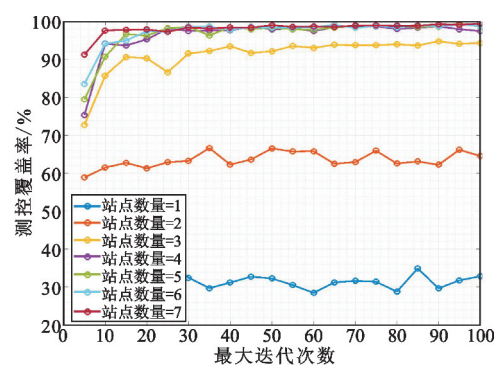


图 3 本文方法测控覆盖率随最大迭代次数变化情况

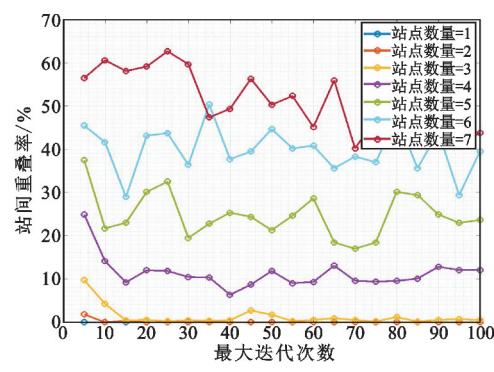


图 4 本文方法站间重叠率随最大迭代次数变化情况

通过与无限制的随机布站方法对比以分析本文方法的优势特点。对比时,设置最大迭代次数为 100,除使用方法不同外,其他参数保持相同。表 1~3 分别显示了本文方法、对比方法的测控覆盖率、站间重叠率、站址接受率随站点数量的统计结果。从表中结果中可知,本文方法与对比方法的 3 个评估指标随站点数量的增多呈相同的变化趋势,即测控覆盖率和站间重叠率均随站点数量的增多而增大,站址接受率均随站点数量的增多先升高后降低。通过对比表 1~3 中指标数据,本文方法相对于对比方

法,测控覆盖率最多提升 10.93% (站点数量为 3 时),站间重叠率越小越好,最多减少 13.1% (站点数量为 5 时),站址接受率最多提升 10% (站点数量为 3 时)。综合考虑 3 个评估指标表现,当站点数量大于等于 2 时,本文方法均明显优于对比方法。

表 1 不同方法测控覆盖率统计结果

站点数量	覆盖率/%	
	本文方法	对比方法
1	32.90	39.30
2	64.63	64.47
3	94.43	83.50
4	97.53	89.77
5	98.73	94.07
6	98.90	96.73
7	99.50	97.10

表 2 不同方法站间重叠率统计结果

站点数量	重叠率/%	
	本文方法	对比方法
1	0.00	0.00
2	0.00	1.27
3	0.53	2.83
4	12.10	16.67
5	23.67	36.77
6	39.50	49.73
7	43.83	54.63

表 3 不同方法站址接受率统计结果

站点数量	接受率/%	
	本文方法	对比方法
1	5.20	5.40
2	10.80	5.80
3	16.80	6.80
4	12.80	6.00
5	10.80	5.20
6	10.40	7.60
7	8.00	6.00

为进一步突出本文方法与对比方法的效果差异,以站点数量等于 5 为例,对测控覆盖率、站间重叠率、站址接受率随迭代次数的变化进行可视化显示,分别如图 5、图 6 和图 7 所示。从图 5 和图 6 中可得到与图 3 和图 4 相似的结论,即测控覆盖率随着迭代次数的增加而升高,站间重叠率随迭代次数增加呈下降趋势但作为次优目标函数存在明显波动情况。另外,从图 7 中可知,站址接受率随迭代次数的增加而稳定下降,主要原因是机动站址位置逐渐

收敛稳定。因篇幅所限,其他站点数量的结果和结论相似,在此不一一表述。综上,图 5~7 可视化地展示了本文方法相对于对比方法的优势。

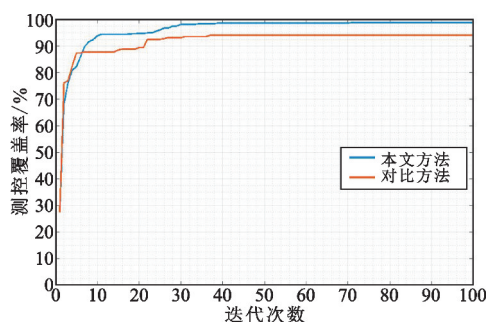


图 5 不同方法测控覆盖率随迭代次数变化曲线

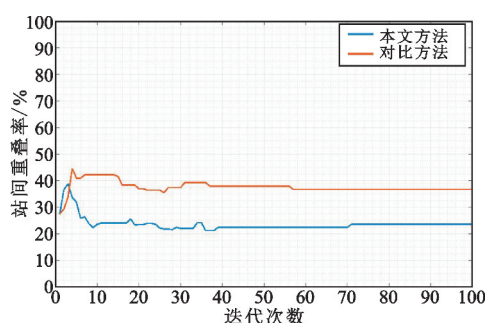


图 6 不同方法站间重叠率随迭代次数变化曲线

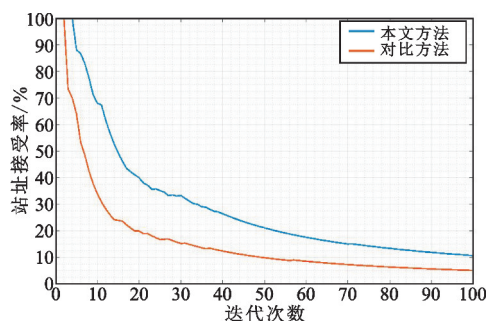


图 7 不同方法站址接受率随迭代次数变化曲线

4 结束语

本文面向航天应急发射能力构建和机动测控能力评估等需求,提出了一种基于有限方向随机游走的火箭发射机动布站方法,通过多机动站间重叠情况限制站址随机游走方向,改进了站址生成和迭代寻优效率,能够有效确定任意数量机动站的部署位置并量化评估相应的最大测控覆盖能力,在测控覆盖率、站间重叠率、站址接受率等方面均有明显优势。

本文方法后续还可继续完善,充分考虑站址的

可达性、时效性、机动距离代价等因素,优化机动站址最终部署方案。此外,还能在给定地面固定站前提下,设计实现火箭发射测控混合布站分析方法,提升方法灵活性。

参考文献:

- [1] 蔡应洲,贺绍飞,谷振丰,等. 固体火箭空间快速响应任务规划研究[J]. 现代防御技术,2021,49(1):107-115.
- [2] 马佳楠,张利萍,董浩,等. 航天测控网多模式效能评估系统设计与关键技术实现[J]. 电讯技术,2023,63(1):32-38.
- [3] 王涛,李泽西,陈学军,等. 全空域多目标测控系统研究综述[J]. 电讯技术,2023,63(10):1631-1641.
- [4] 彭中新,祁振强,钟圣,等. “星链”在俄乌冲突中的运用分析与思考启示[J]. 战术导弹技术,2022(6):121-127.
- [5] 刘保国,吴斌. 中继卫星系统在我国航天测控中的应用[J]. 飞行器测控学报,2012,31(6):1-5.
- [6] 董光亮,张国亭,韩秋龙. 运载火箭测控系统技术与发展[J]. 飞行器测控学报,2014,33(2):93-98.
- [7] 王子瑜,胡钰,彭越,等. 运载火箭地面测发控系统架构研究[J]. 测试技术学报,2023,37(3):185-193.
- [8] 刘兴威,张桂洪. 航天应急发射体系建设[J]. 国防科技,2016,37(1):14-18.
- [9] 陈莉丹,沈怀荣,邵琼玲. 小卫星机动发射测控布站优化与仿真[J]. 装备指挥技术学院学报,2002,13(2):41-44.
- [10] ZHAO L, XIN C, SHI M, et al. Current status of research on space emergency launch [M]. Bangkok: IOP Publishing, 2020.
- [11] 路建工,吕久明,李建华. 航天发射测控力量体系化建设[J]. 国防科技,2019,40(1):68-72.
- [12] 崔健,何孝港. 基于 STK 的三维布站辅助决策系统设计与实现[J]. 重庆理工大学学报(自然科学版),2013,27(6):88-92.
- [13] 叶君好,郝耀峰,李春雨,等. 基于 AHP 航天测控阵地勘察模型设计与评估[J]. 测控技术,2018,37(9):412-415.
- [14] WU F, LIU X L, WANG J, et al. Research on agile space emergency launching mission planning simulation and verification method[J]. Journal of Systems Engineering and Electronics, 2023, 34(5):1267-1284.
- [15] 韩彭威,段发阶,李杰,等. 基于双目视觉与罗德里格斯旋转的机械臂全工作空间域绝对定位精度提升方法[J]. 光学精密工程,2023,31(8):1150-1161.

作者简介:

董浩 男,1990 年生于河南周口,2018 年获博士学位,现为工程师,主要研究方向为卫星任务规划调度、遥感影像处理等。

DOI:10.20079/j.issn.1001-893x.230908002

5G NR 网络中一种新颖的移动用户速度估计方法*

周宝龙¹, 陆 犇¹, 杨洪生¹, 唐 亮²

(1. 上海瀚讯信息技术股份有限公司, 上海 201803; 2. 中国科学院上海微系统所, 上海 200050)

摘 要:在蜂窝通信系统中,移动用户速度估计是一个颇具挑战性的课题。虽然有大量文献涉及这个问题,但是由于一些不符合实际的假设,导致绝大多数方法没法应用。为此,为 5G NR (New Radio) 移动用户提出一种新颖的速度估计方法,可以克服现有方法的缺点。基本思想是通过用户对位置求导来获得移动速度。然而在非直射径环境中,用户位置估计相当困难。为了获得移动用户的精确位置估计,提出一种鲁棒性的具有自同步能力的方位角估计算法和一种改进的时延估计算法。仿真表明,提出的速度估计方法在直射径环境下,90% 的速度估计值的估计误差小于 2%;在非直射径环境下,90% 的速度估计值的估计误差小于 7%。与现有方法相比,所提速度估计方法对噪声更具鲁棒性,并且更适合实际系统。

关键词:5G; 多输入多输出(MIMO); 速度估计; 方位角估计

开放科学(资源服务)标识码(OSID):



中图分类号:TN911.23 文献标志码:A 文章编号:1001-893X(2025)03-0398-09

A Novel Speed Estimation Scheme for Mobile Users in 5G NR Networks

ZHOU Baolong¹, LU Ben¹, YANG Hongsheng¹, TANG Liang²

(1. Shanghai Jushri Technology Inc., Shanghai 201803, China;

2. Shanghai Institute of Microsystem And Information Technology, Chinese Academic of Sciences, Shanghai 200050, China)

Abstract: Mobile user speed estimation is a challenging issue for cellular communication systems. There are plenty of literatures addressed on this problem. However, most of existing methods are not applicable due to some assumptions such as Jakes model etc. In this paper, a novel speed estimation scheme is proposed for mobile users in 5G new radio systems, which can successfully conquer the defects of the existing solutions. The proposed scheme basically differentiates user position to obtain mobile speed. However, user positioning is quite difficult in non-line of sight (NLOS) channels. In order to obtain precise position estimation of a mobile user, a robust self-synchronous azimuth of departure (RSSAOD) estimation algorithm and a min-max time advanced (TA) estimation algorithm are proposed. Extensive simulations verify that with the proposed scheme, 90% of the speed estimations have above 98% correctness in the line of sight (LOS) channel and 93% correctness in the NLOS channel, respectively. Compared with existing methods, the proposed speed estimation scheme is more robust with respect to noise tolerance and is more suitable for practical systems.

Key words: 5G; multiple-input multiple-output (MIMO); speed estimation; AOD estimation

* 收稿日期:2023-09-08;修回日期:2024-05-31

通信作者:周宝龙 Email:zhoubaolong98@163.com

0 引言

在 5G NR 网络中, 导频周期和导频数量均可灵活配置。导频密度越高, 越能更好地跟踪无线信道, 从而系统具有更低的误块率 (Block Error Rate, BLER)。然而, 系统花费的开销也更多。如果基站能获得用户移动速度, 则基站能在 BLER 性能和系统开销之间取得平衡。在异构网络中, 用户移动速度是非常重要的, 因为可以用它来决定用户是接入微小区还是宏小区。另外, 小区切换、信道估计和高速移动场景下的频偏预补偿也能从移动速度获得增益。因此, 在基站端估计用户移动速度显得非常重要。

已有大量文献研究速度估计方法。根据使用的信道特征, 这些估计方法可以分成两类。

一类方法利用依赖速度的信道快衰特性来估计移动速度。例如: 文献[1]利用电平穿透率来估计移动速度, 即在给定的时间窗内统计接收信号功率穿过某个电平的次数; 文献[2]通过计算接收信号的方差来估计用户移动速度; 其他方法, 如文献[3-4]中的自相关函数法、文献[5]中的特征空间法和文献[6]中的功率谱估计法, 主要思路是估计最大多普勒频偏来获得用户移动速度。在上述方法中, 电平穿透率法和协方差法需要存储大量接收数据并且有很高的计算复杂度, 因此它们仅能应用在高速场景中; 自相关函数法、特征空间和功率谱法要求信道服从 Jakes 模型, 这限制了它们的应用。

另一类方法利用依赖速度的信道大尺度特性来估计移动速度。文献[7]使用信道阴影衰落的时变特性来估计移动速度, 文献[8]对此方法作了简化, 然而两个因素限制了这个方法的应用。第一个因素是解相关距离 D 随着传播环境和载频的变化而变化, 因此 D 只能通过经验获得。如果 D 不正确, 将会误选用户速度和度量之间的一一映射曲线, 这会导致不正确的速度估计值。第二个因素是存储用户速度与度量映射的数据库基于现场数据获得, 需要考虑相当多的信道环境用例。如果数据库包含所有用例, 数据库规模将非常巨大; 然而, 如果只包含部分用例, 速度估计误差将会很大。

本文为 5G NR 网络移动用户提出一种新颖的速度估计方法: 使用移动用户在不同时刻的空间位置来估计移动速度。其基本思想相当简单和直观,

即首先基于上行探测信号估计出用户的方位角 (Azimuth angle of Departure, AOD) 和俯仰角 (Elevation angle of Departure, EOD) 及用户与基站之间的距离 l , 然后基于 AOD、EOD 和距离 l 计算用户在不同时刻的移动距离 d , 最后移动距离 d 除以对应的时间差 Δt 得到用户移动速度。然而, 用户空间位置估计是一个颇具挑战性的任务, 特别是在非直射信道环境中。为了获得用户准确的位置, 本文设计了一种鲁棒性的具有自同步能力的方位角估计 (Robust Self-synchronous AOD algorithm, RSSAOD) 算法和一种改进的时延估计算法。

本文中 $E(\cdot)$ 、 $(\cdot)^H$ 、 $(\cdot)^T$ 、 $(\cdot)^*$ 和 $|\cdot|$ 分别表示期望、共轭转置、转置和 Frobenius 范数。

1 系统模型

考虑一个多输入多输出 - 正交频分复用 (Multiple-Input Multiple-Output Orthogonal Frequency Division Multiplexing, MIMO-OFDM) 系统, 在基站端有 $M = M_h M_v$ 根天线 (M_h 是水平面天线数, M_v 是垂直面天线数), 移动用户端有 N 根天线。不失一般性, 假定 N 等于 1。基站和用户之间的无线信道是时变、频选衰落信道, 表示为 $\mathbf{h} = [h_1, h_2, \dots, h_M]^T$, 其中 h_i 服从 $\mathcal{CN}(0, 1)$ 分布。用 s 表示发射调制符号, 那么在第 k 个子载波上的接收信号可表示如下:

$$\mathbf{y}[k] = \mathbf{h}[k]s[k] + \mathbf{n}[k] \quad (1)$$

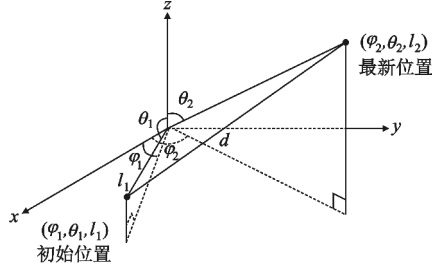
式中: $\mathbf{n}[k]$ 是噪声; $s[k]$ 满足 $E[|s[k]|^2] = 1$ 。

在 5G NR 系统, 移动用户周期性发射探测信号到基站。探测接收信号 $\mathbf{y}[k]$ 首先与导频 $s[k]$ 按式(2)进行共轭相乘, 然后把 $\mathbf{h}_{ls}[k]$ 转换成时域信号 $\mathbf{h}_{ls}[n]$, 接着对 $\mathbf{h}_{ls}[n]$ 施加一个时域窗, 如矩形窗, 以便降低噪声的影响。最后, 降噪后的 $\mathbf{h}_{ls}[n]$ 被转换成频域信号 $\hat{\mathbf{h}}[k]$, 这就是探测估计信道, 简称为 SRS (Sounding Reference Signal) 估计信道。

$$\mathbf{h}_{ls}[k] = \mathbf{y}[k](s[k])^* \quad (2)$$

2 速度估计方案设计

基于 $\hat{\mathbf{h}}[k]$, 本文第 2.1 和 2.2 小节分别推导出 AOD (用 φ 表示)、EOD (用 θ 表示) 和移动用户与基站之间的距离 l , 然后基于不同时刻的 φ 、 θ 和 l , 可计算出用户的移动距离 d , 如图 1 所示。

图1 AOD、EOD 和距离 l 几何示意

$$d = \sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} \quad (3)$$

$$\Delta x = l_2 \sin \theta_2 \cos \varphi_2 - l_1 \sin \theta_1 \cos \varphi_1$$

$$\Delta y = l_2 \sin \theta_2 \sin \varphi_2 - l_1 \sin \theta_1 \sin \varphi_1 \quad (4)$$

$$\Delta z = l_2 \cos \theta_2 - l_1 \cos \theta_1$$

式中: $(\varphi_1, \theta_1, l_1)$ 和 $(\varphi_2, \theta_2, l_2)$ 分别是移动用户在不同时刻的空间位置。故移动用户的速度可根据下式计算得到:

$$v = \frac{d}{\Delta t} \quad (5)$$

式中: Δt 是用户移动距离 d 经历的时间。

从上述公式可知,速度估计值 v 由 φ 、 θ 和 l 决定。在蜂窝移动通信系统中,移动用户到基站之间的距离 l 可以通过无线信号传播时延计算得到,这个传播时延在 3GPP 中称为 time advanced (TA)。因此,移动速度 v 的估计等效于 φ 、 θ 和 TA 的估计。接下来各小节将详细描述估计 φ 、 θ 和 TA 的算法。

2.1 鲁棒性的具有自同步能力的方位角估计算法 (RSSAOD)

由于可以采用相同的算法来估计 AOD 和 EOD,为了简化,本文以 AOD 估计为例来描述算法。

经典的 AOD 估计算法,如 MUSIC 算法^[11-12] 和 ESPRIT 算法^[13-15],需要对接收信号 $\mathbf{y}[k]$ 的 $M \times M$ 维协方差矩阵做特征矢量分解。然而,在 5G NR 系统中, M 非常大,比如 64、128,因此矩阵特征矢量分解复杂度非常高,以至于在实际系统中很难实现。而且,为了获得信号和噪声的精确统计特性,需要收集相当多的接收信号样本,这要求用户的 AOD 在信号观测期间必须保持不变,但是当用户移动速度比较高时,这个要求很难满足。文献[16]提出和差天线体制雷达空时自适应测角方法,文献[17]设计短基双天线测向阵列来增强角度估计精度,然而蜂窝系统天线阵列与雷达天线和短基双天线有区别,因此这两种方法没法应用在蜂窝系统。在蜂窝通信系

统中,当前常用的 AOD 估计算法(简记为 present AOD algorithm^[18])利用 AOD 功率谱来估计 AOD,原理如下:

$$\varphi^{\text{opt}} = \underset{\varphi}{\operatorname{argmax}} p(\varphi) \quad (6)$$

式中:覆盖扇形小区的方位角 $\varphi \in (-60^\circ, 60^\circ)$; $p(\varphi)$ 是 AOD 功率谱,

$$p(\varphi) = (\mathbf{a}(\varphi))^T \mathbf{R} \mathbf{a}(\varphi) \quad (7)$$

式中: $\mathbf{a}(\varphi) = [1, e^{j\alpha}, \dots, e^{j(M_h-1)\alpha}]^T$ 且 $\alpha = \frac{2\pi}{\lambda} d_h \sin \varphi$;

λ 是波长; d_h 是相邻天线间距; $\mathbf{R} = \frac{1}{N_{\text{sc}}} \sum_{k=1}^{N_{\text{sc}}} \hat{\mathbf{h}}_h(k) \cdot (\hat{\mathbf{h}}_h(k))^H$ 是协方差矩阵; $\hat{\mathbf{h}}_h = [h_1, h_2, \dots, h_{M_h}]^T$ 是水平面估计信道; N_{sc} 是子载波数。

上述算法在以直射径为主的环境(简称 LOS 信道)下性能良好。然而,在非直射径环境(简称 NLOS 信道)下,功率谱中与最大峰对应的 AOD 不一定是用户的真实 AOD,在绝大多数情况下,存在较大的 AOD 估计误差,这会导致不正确的速度估计值。为了改善 AOD 和速度估计精度,本文提出一种鲁棒性的具有自我同步能力的 AOD 估计算法(简称为 RSSAOD 估计算法)。RSSAOD 估计算法包括两部分:一是最小二乘 AOD 估计算法(简称 LsAOD 估计算法),它基于最小二乘(Least square)准则来估计移动用户 AOD;另外一个 AOD 同步算法,它为 LsAOD 估计算法推导初始输入。

LsAOD 估计算法的基本思想是:利用前一个 AOD 估计值 φ_{i-1} 及最新的 AOD 增量 $\Delta\varphi$,在 AOD 功率谱中搜索逼近用户真实方位角的 AOD 估计值,如式(8)所示:

$$\varphi_i^{\text{opt}} = \underset{\varphi}{\operatorname{argmin}} (p(\varphi) - p(\varphi_{\text{expect}}))^2 \quad (8)$$

s. t.

$$\text{C1: } \varphi_{\text{expect}} = \varphi_{i-1} + \Delta\varphi$$

$$\text{C2: } \dot{p}(\varphi_i^{\text{opt}}) = 0$$

$$\text{C3: } p(\varphi_i^{\text{opt}}) > p_{\text{thr}}$$

$$\text{C4: } |p(\varphi_i^{\text{opt}}) - p(\varphi_{\text{expect}})| < \varepsilon_1$$

式中: i 表示索引; ε_1 表示一个任意小的数; p_{thr} 是过滤噪声峰的阈值;一阶导数 $\dot{p}(\varphi_i^{\text{opt}}) = 0$ 和 $p(\varphi_i^{\text{opt}}) > p_{\text{thr}}$ 用来确保 $p(\varphi_i^{\text{opt}})$ 是局部极大值; $|p(\varphi_i^{\text{opt}}) - p(\varphi_{\text{expect}})| < \varepsilon_1$ 用来确保找到的 φ_i^{opt} 离 φ_{expect} 不远。基于数学模型(8),LsAOD 估计算法(算法1)具体步骤如下:

步骤 1 在当前 AOD 功率谱中,搜索 J 个最大峰对应的 AOD,表示为 $\text{AoD}_{i,j}, i=1,2,\dots,I$ 表示第 i 次 AOD 估计, $j=1,2,3,\dots,J$ 表第 i 次 AOD 估计时 AOD 功率谱中的峰索引。

步骤 2 计算 AOD 增量:

$$\Delta\text{AoD}_{i,j} = \text{AoD}_{i,j} - \text{AoD}_{i-1} \quad (9)$$

步骤 3 选择期望峰:

$$j^{\text{opt}} = \underset{j}{\text{argmin}} \left| |\Delta\text{AoD}_{i,j}| - |\Delta\text{AoD}| \right| \quad (10)$$

s. t.

$$\left| |\Delta\text{AoD}_{i,j^{\text{opt}}}| - |\Delta\text{AoD}| \right| < \varepsilon_2$$

式中: ΔAoD 是最新 AOD 增量; ε_2 是一个任意小的数。

步骤 4 输出第 i 次 AOD 估计值:

$$\text{AoD}_i = \text{AoD}_{i,j^{\text{opt}}} \quad (11)$$

在 AOD 功率谱中,表示用户真实 AOD 的峰的能量不会太小。为了降低上述算法的复杂度,步骤 1 仅需选择少数几个最大峰,比如,在 CDL (Cluster Delay Line) 信道中 J 通常取 5。另外,在 NLOS 信道中, AOD 功率变化范围比较大,优化模型 (8) 中的 p_{thr} 和 ε_1 不容易设置。不过,用户 AOD 是一个相对慢变参数,因此步骤 3 直接在角度域搜索用户 AOD。

算法输入参数 AoD_{i-1} 和 ΔAoD 的初始值,分别用 AoD_0 和 ΔAoD_0 表示,需利用当前 AOD 估计算法输出的 S 个连续样本来推导。初始值 AoD_0 和 ΔAoD_0 对 LsAOD 算法估计性能很重要,如果 AoD_0 或 ΔAoD_0 不正确,就没法获得正确的 AoD_i 估计值。下面详细描述如何根据 AOD 功率谱和 S 个 AOD 估计样本来推导初始值 ΔAoD_0 和 AoD_0 。 S 个 AOD 估计样本表示为 $\text{AoD}_s, s=1,2,\dots,S$ 。

由于 AOD 增量变化缓慢,可以根据最新 AOD 增量来推导初始值 ΔAoD_0 , 计算步骤请见 initial delaAOD 算法 (算法 2), 其中 AOD 估计样本数 $S=4$ 。

步骤 1 计算 AOD 增量 ΔAoD_s :

$$\begin{aligned} \Delta\text{AoD}_1 &= |\text{AoD}_2 - \text{AoD}_1| \\ \Delta\text{AoD}_2 &= |\text{AoD}_3 - \text{AoD}_2| \\ \Delta\text{AoD}_3 &= |\text{AoD}_4 - \text{AoD}_3| \end{aligned} \quad (12)$$

步骤 2 计算 AOD 增量差:

$$\begin{aligned} \text{diff1} &= |\Delta\text{AoD}_1 - \Delta\text{AoD}_2| \\ \text{diff2} &= |\Delta\text{AoD}_2 - \Delta\text{AoD}_3| \\ \text{diff3} &= |\Delta\text{AoD}_3 - \Delta\text{AoD}_1| \end{aligned} \quad (13)$$

步骤 3 推导初始值 ΔAoD_0 , 伪代码如下:

```
If diff1 <  $\varepsilon_3$ , then  $\Delta\text{AoD}_0 = (\Delta\text{AoD}_1 + \Delta\text{AoD}_2)/2$ 
elseif diff2 <  $\varepsilon_3$ , then  $\Delta\text{AoD}_0 = (\Delta\text{AoD}_2 + \Delta\text{AoD}_3)/2$ 
elseif diff3 <  $\varepsilon_3$ , then  $\Delta\text{AoD}_0 = (\Delta\text{AoD}_3 + \Delta\text{AoD}_1)/2$ 
end
```

在步骤 3, ε_3 是一个比较小的数。如果 diff1、diff2 和 diff3 都大于 ε_3 , 说明用户处于瞬时加速或减速状态, 将推导不出初始增量 ΔAoD_0 。此时, AOD 估计样本需滑动更新直至获得合理的初始增量, 比如, $\text{AoD}_1, \text{AoD}_2, \dots, \text{AoD}_s$ 滑动更新成 $\text{AoD}_2, \text{AoD}_3, \dots, \text{AoD}_{s+1}$ 。

获得 ΔAoD_0 后, 接下来推导 AoD_0 。直观上, 第 S 个 AOD 估计样本 AoD_s 能用作 AoD_0 , 然而 AoD_s 来自当前 AOD 估计算法, 可能正确, 也可能不正确。为了提升 AoD_0 的精度, 本文提出 initial AOD 算法 (算法 3) 来推导 AoD_0 。

initial AOD 算法使用前 $S-1$ 个样本 $\text{AoD}_s (s=1, 2, \dots, S-1)$ 和估计 AoD_s 时的 AOD 功率谱来提升 AoD_0 的正确性。由于 AOD 功率谱包含最接近用真实 AOD 的峰, 如果 ΔAoD_0 和 $S-1$ 个样本中任意一个样本是正确的, 那么一定能推导出正确的 AoD_0 。具体步骤请见算法 3。

initial delaAOD 算法和 initial AOD 算法算法用来推导 ΔAoD_0 和 AoD_0 , 统称为 AOD 同步。样本数 S 越大, 获得正确初始值 ΔAoD_0 和 AoD_0 的概率越高, 但算法复杂度也越高。在 CDL 信道中, S 通常取 4。需注意的是, 必须在 AOD 同步后才能启动 LsAOD 估计算法。

步骤 1 在 AOD 功率谱中搜索 J 个最大峰对应的 AOD, 表示为 $\text{AoD}_{s,j}, j=1,2,\dots,J$ 。

步骤 2 计算 $\text{AoD}_{s,\text{cal}}$:

$$\text{AoD}_{s,\text{cal}} = \text{AoD}_s \pm (S-s) \Delta\text{AoD}_0, s=S-1, \dots, 2, 1 \quad (14)$$

其中, 优先用 AoD_{s-1} 计算 $\text{AoD}_{s,\text{cal}}$, 然后用 $\text{AoD}_{s-2}, \dots, \text{AoD}_1$ 。如果能用 AoD_{s-1} 成功推导出 AoD_0 , 那么不必再用 $\text{AoD}_{s-2}, \dots, \text{AoD}_1$ 来计算 $\text{AoD}_{s,\text{cal}}$ 。

步骤 3 搜索最优峰:

$$j^{\text{opt}} = \underset{j}{\text{argmin}} |\text{AoD}_{s,\text{cal}} - \text{AoD}_{s,j}|, j=2,3,\dots,J \quad (15)$$

s. t.

$$|\text{AoD}_{s,\text{cal}} - \text{AoD}_{s,j^{\text{opt}}}| < \varepsilon_4$$

式中: ε_4 是一个很小的数。

步骤 4 推导 AoD_0 。如果成功获得 j^{opt} , 则 $\text{AoD}_0 = \text{AoD}_{S_j^{\text{opt}}}$, 否则 $\text{AoD}_0 = \text{AoD}_S$ 。

图 2 总结了 RSSAOD 估计算法。在图 2 中, 变量 Cal_initVal_flag 用来控制 AOD 同步和 AOD 估计的启停, 它的初始值等于 1。如果 AOD 同步获得的初始值 ΔAoD_0 或 AoD_0 不等于真实值, 则 AOD 同步是假同步, 这个假同步将在 AOD 估计过程中被检测到, 并触发一次 AOD 重同步。AOD 重同步使用新的 AOD 估计样本来推导正确的初始值 ΔAoD_0 和 AoD_0 。这意味着 RSSAOD 估计算法具有检错和纠错能力, 即自己发现错误并自己纠正错误。

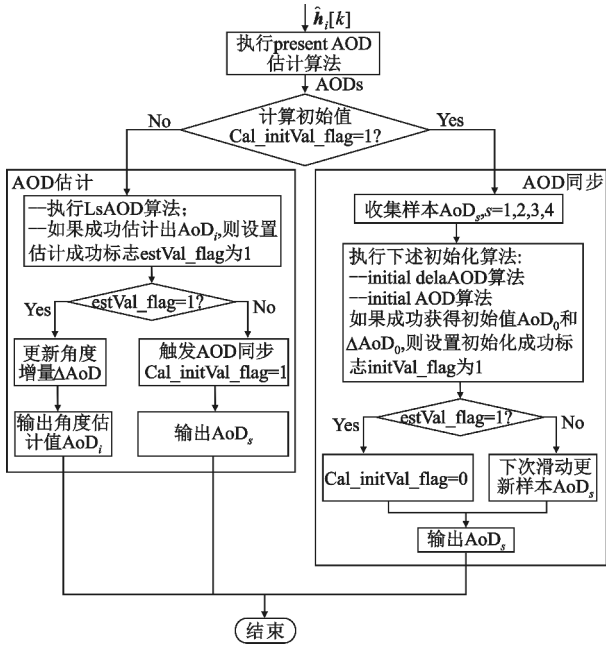


图 2 RSSAOD 估计算法

2.2 改进的 TA 估计算法

TA 由传播距离决定, 故可用 TA 来估计距离 l 。文献[19]基于接收信号自相关估计 TA, 虽然可有效应对载波频偏, 但在低信噪比 (Signal-to-Noise Ratio, SNR) 信道中会经历严重的性能下降。“当前的 TA 估计算法”利用最强径来估计 TA^[20], 然而在 NLOS 信道环境下获得的 TA 估计值远大于真实值, 导致速度估计值偏大。众所周知, 在多径信道中, 首径代表基站和用户之间传播距离对应的传输时延。受此启发, 本文提出改进的 TA 估算法 (算法 4) 基本思想是: 在信道时延功率谱中, 搜索最接近首径的

多径, 被搜索到的多径用来估计 TA。数学表达式如 (16) 所示:

$$\text{TA}^{\text{opt}} = \underset{\tau}{\operatorname{argmin}} \{ \text{local_max}(p(\tau)) \} \quad (16)$$

s. t.

$$p(\tau) > p_{\text{thr}}$$

式中: $p(\tau)$ 是信道时延功率谱; 函数 $\text{local_max}(\cdot)$ 用来寻找 $p(\tau)$ 的各极大值对应的时延 τ ; p_{thr} 是过噪声峰的阈值。基于式 (16), 改进的 TA 估计算法具体步骤如下:

步骤 1 利用 IDFT 把 SRS 估计信道变换到时域:

$$\hat{h}[n] = \text{IDFT}(\hat{h}[k]), k=1, 2, \dots, N_{\text{sc}}; n=1, 2, \dots, N_{\text{sc}} \quad (17)$$

步骤 2 计算时延功率谱 $p(n)$:

$$p(n) = \hat{h}[n] \hat{h}^*[n] \quad (18)$$

步骤 3 在 $p(n)$ 中搜索 J 个最大峰对应的时间索引, 记为 $\text{Idx}_j, j=1, 2, \dots, J$, 且 $p(\text{Idx}_j) > p_{\text{thr}}$ 。

步骤 4 从 J 个候选值选择最优索引:

$$j^{\text{opt}} = \underset{j}{\operatorname{argmin}} \{ \text{Idx}_j \} \quad (19)$$

步骤 5 计算 TA:

$$\text{TA} = T_s \times \text{Idx}_{j^{\text{opt}}} \quad (20)$$

式中: T_s 是 SRS 信号时域采样间隔。

3 仿真与分析

考虑一个 5G NR 多天线系统, 基站有 64 根天线, 终端有 1 根天线, 终端周期性发 SRS 信号给基站, 基站根据 SRS 接收信号估计 AOD 和 TA。用 CDL-C 模型模拟非直射径环境, 用 CDL-D 模型模拟直射径环境。本文比较 3 种速度估计方法的性能: 一种是本文提出的速度估计方法, 采用 RSSAOD 估计算法和改进的 TA 估计算法, 简记为“proposed method”; 一种是当前常用的速度估计方法, 采用当前的 AOD 估计算法^[18]及 TA 估计算法^[20], 简记为“present method”; 最后一种是学术界讨论比较多的基于自相关函数 (Auto Correlation Function, ACF) 的速度估计方法^[4], 简记为“ACF method”。用户以 60 km/h 的速度在小区内来回运动。表 1 列出了主要仿真参数。需要说明的是, 由于篇幅限制, 本文只给出了典型速度 60 km/h 下的仿真, 对于其他移动速度下的仿真, 可得到类似现象和结论。

表 1 仿真参数

参数名	取值	解释
仿真帧数	1×10^5	每个 SNR 运行的无线帧数
信噪比/dB	30,5	上行探测信号 SNR
子载波间隔/kHz	30	相邻子载波间隔
PRB 大小	12	每个 PRB(即资源块) 包含 12 个子载波
PRB 数	273	系统带宽分成 273 个 PRB
SRS PRB 数	256	上行探测信号占用的带宽
SRS 符号数	1	每子帧探测信号符号数
$T_{\text{Speed}}/\text{s}$	2	速度估计周期
$\varphi/(\circ)$	$(-60, 60)$	方位角 AOD 取值范围
$\theta/(\circ)$	$(-15, 15)$	俯仰角 EOD 取值范围
仿真信道	CDL-C/D	5G NR 信道模型
M	64	基站天线数
N	1	用户天线数
$v/(\text{km/h})$	60	用户移动速度

用来评估性能的速度估计误差定义如下:

$$\text{error} = |\hat{v} - v_{\text{real}}| / v_{\text{real}} \quad (21)$$

式中: v_{real} 是速度真实值; $\hat{v}$ 是速度估计值。AOD 估计误差和 TA 估计误差也采用这个定义。

3.1 CDL-C(即非直射径环境)下的仿真结果

图 3 ~ 6 分别给出了各种算法在高信噪比 (30 dB) 且 CDL-C 信道下速度估计、AOD 估计和 TA 估计的性能。如前文所述,在 NLOS 信道下,当前的 AOD 和 TA 估计算法性能不好,这从图 5 和图 6 得到了很好的验证:25% 的 AOD 相对估计误差值超过 50%,58% 的 TA 相对估计误差值超过 21%。然而,对于本文提出的 RSSAOD 估计算法和改进的 TA 估计算法,90% 的 AOD 相对估计误差值小于 5%,95% 的 TA 相对估计误差值小于 2%。这证明了本文提出的 RSSAOD 估计算法和改进的 TA 估计算法能大幅度提升 AOD 和 TA 的估计精度,进而导致本文提出的速度估计方案的性能明显优于当前的速度估计方案,图 4 很好地印证了这点:对本文所提出的速度估计方案,90% 的速度估计相对误差值小于 7%,然而对当前的速度估计方法,80% 的速度估计相对误差值在 15% ~ 35% 范围内变化。从图 3 可观察到,present method 的速度估计曲线在真实值附近波动很大,然而 proposed method 的速度估计曲线非常接近真实速度曲线。这进一步证明了 proposed method 的速度估计的正确性。另外,从图 3 也可观察到,ACF method 的速度估计值大大小于真实值,70% 的速度估计相对误差值超过 50%。原因是

ACF method 要求无线信道必须服从 Jakes 模型,然而 CDL-C 信道不满足这个条件,导致 ACF method 的速度估计误差比较大。

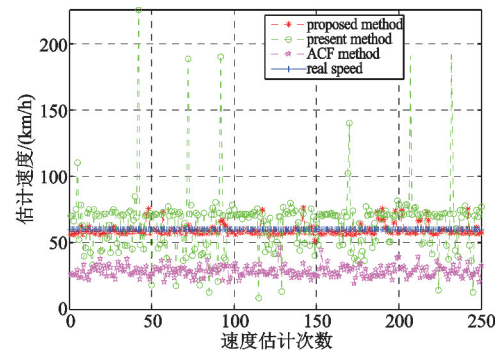


图 3 30 dB 时的速度估计曲线

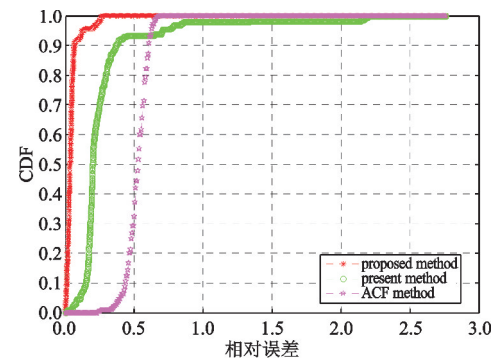


图 4 30 dB 时的速度估计相对误差 CDF 曲线

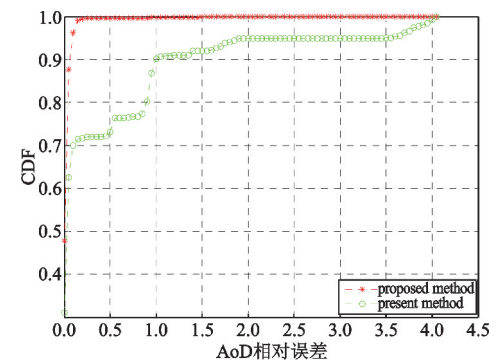


图 5 30 dB 时的 AOD 估计相对误差 CDF 曲线

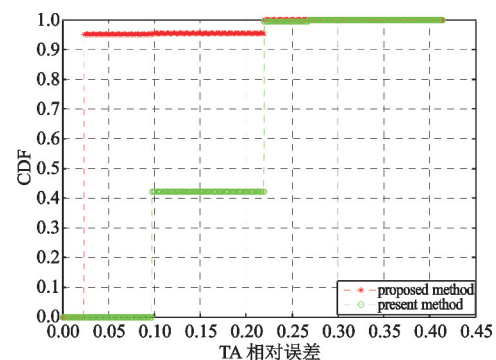


图 6 30 dB 时的 TA 估计相对误差 CDF 曲线

图 7~10 分别给出了各算法在低信噪比(5 dB)且 CDL-C 信道下速度估计、AOD 估计和 TA 估计的性能。从图 7~10 可以观察到与图 3~6 中类似的现象。另外,比较图 3~6 和图 7~10 可知,proposed method 和 present method 受噪声影响比较小,然而 ACF method 受噪声影响大。原因是 ACF method 利用信道相关性来估计速度,但是噪声会降低相关性,这导致速度估计值明显偏大;proposed method 和 present method 使用相关峰来估计 AOD 和 TA,相关峰受噪声影响小,故噪声对 AOD、TA 估计影响小,进而对速度估计影响也小。

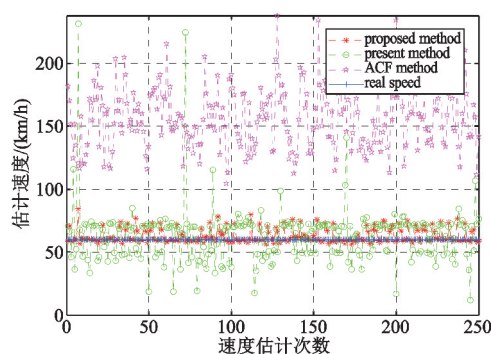


图 7 5 dB 时的速度估计曲线

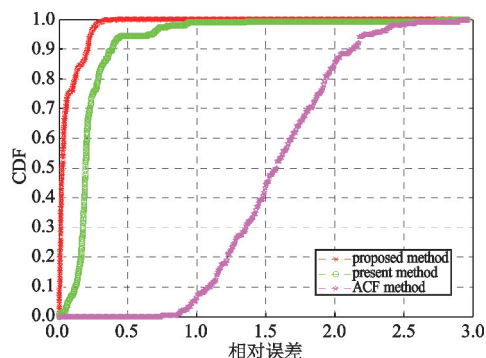


图 8 5 dB 时的速度估计相对误差 CDF 曲线

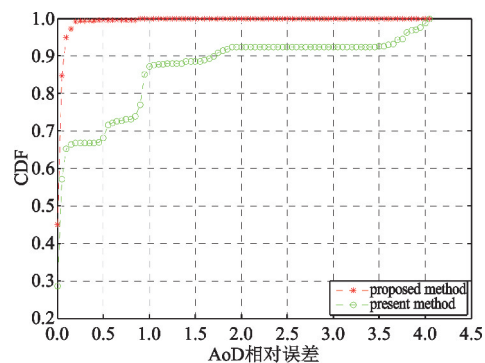


图 9 5 dB 时的 AOD 估计相对误差 CDF 曲线

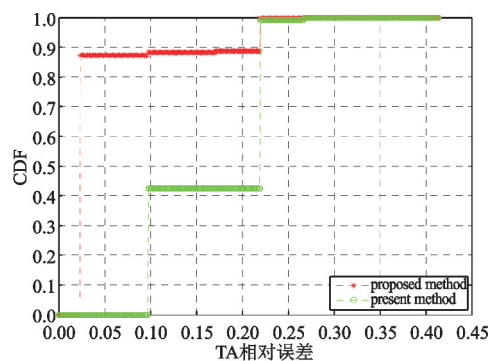


图 10 5 dB 时的 TA 估计相对误差 CDF 曲线

3.2 CDL-D(即直射径环境)下的仿真结果

图 11~14 分别给出了各种算法在高信噪比(30 dB)且 CDL-D 信道下速度估计、AOD 估计和 TA 估计的性能。如前文所述,当前 AOD 和 TA 估计算法在 LoS 信道下能很好地工作,这从图 13 和图 14 得到很好的验证:90% 的 AOD 相对估计误差值小于 4%,90% 的 TA 相对估计误差值小于 2.5%。因此,当前方案的速度估计性能几乎和本文所提出的速度估计方案的性能一致,如图 11 所示,两方案的速度估计曲线几乎重合,并且非常接近速度真实值。此外,根据图 12 可知,90% 的速度估计误差小于 2%。由于 CDL-D 信道下的信道相关性接近 1,故 ACF 方法的速度估计值严重偏小,速度估计相对误差超过 80%。

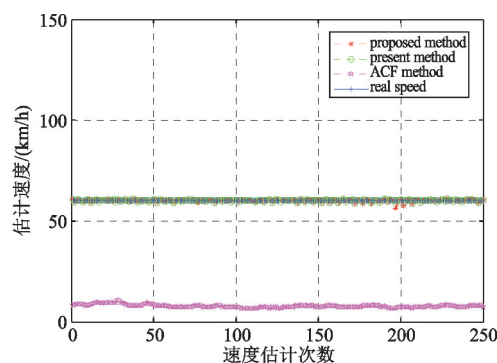


图 11 30 dB 时的速度估计曲线

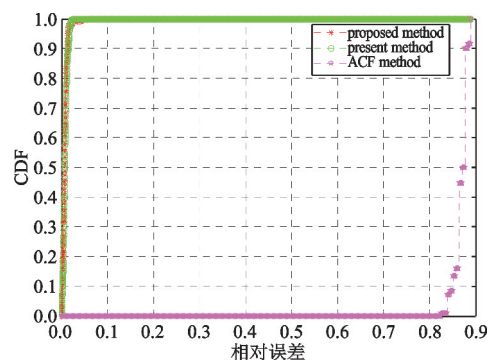


图 12 30 dB 时的速度估计相对误差 CDF 曲线

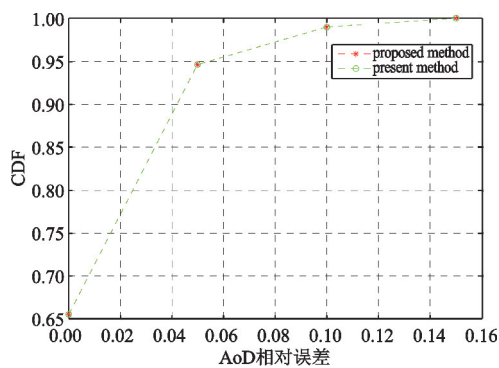


图 13 30 dB 时的 AOD 估计相对误差 CDF 曲线

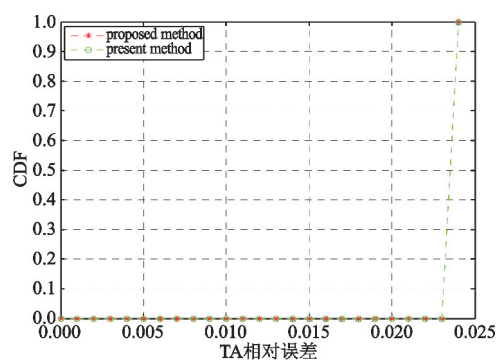


图 14 30 dB 时的 TA 估计相对误差 CDF 曲线

图 15~18 分别给出了各种算法在低信噪比 (5 dB) 且 CDL-D 信道下速度估计、AOD 估计和 TA 估计的性能。从图 15~18 可以观察到与图 11~14 中类似的现象。另外, 比较图 11~14 和图 15~18 可知, 本文所提速度估计方案和当前速度估计方案几乎不受噪声影响, 然而 ACF 速度估计方法受噪声影响非常严重, 原因与 CDL-C 信道下相同, 这里不再赘述。

比较 CDL-C 和 CDL-D 下的仿真结果可知, 本文所提速度估计方案在 CDL-D 下的性能更好。这是因为在 CDL-D 信道下出现伪峰的概率很小, 故 AOD 估计误差、TA 估计误差很小, 进而速度估计误差更小。

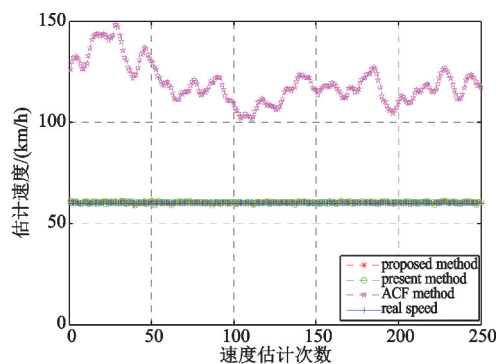


图 15 5 dB 时的速度估计曲线

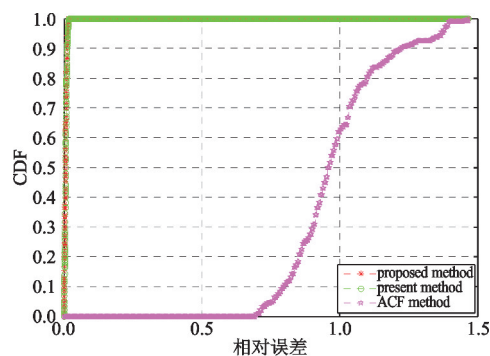


图 16 5 dB 时的速度估计相对误差 CDF 曲线

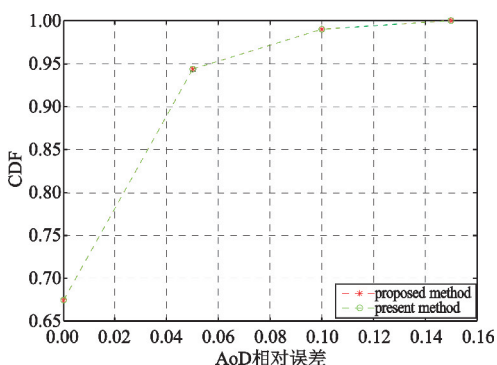


图 17 5 dB 时的 AOD 估计相对误差 CDF 曲线

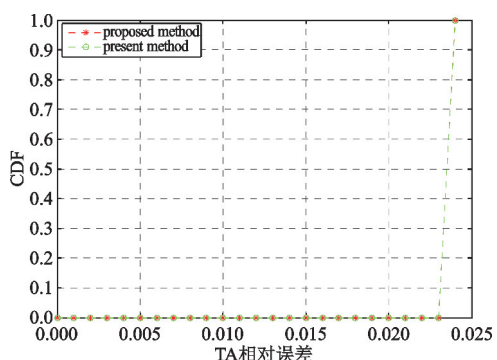


图 18 5 dB 时的 TA 估计相对误差 CDF 曲线

4 结束语

本文为 5G NR 网络中的移动用户提出了一种新颖的速度估计方案。这个方案由两部分组成: RSSAOD 估计算法和改进的 TA 估计算法。在 NLOS 信道下, 通过使用 RSSAOD 估计算法和改进的 TA 估计算法, AOD 和 TA 估计精度大幅提升, 进而速度估计精度也大幅度提升。与现有速度估计方案相比, 本文所提速度估计方案不但提升了速度估计精度, 而且对噪声更具鲁棒性, 不受限于信道模型, 比如 Jakes 模型, 可应用到任何信道场景。此外, 本文所提速度估计方案在实际通信系统中容易实现, 其性能在仿真中得到了验证。

当然,如果所在场景直射径太弱或散射径非常丰富或 SNR 非常低,AOD 估计误差和 TA 估计误差将增大,从而导致本方案的速度估计误差增大。后续将针对这种场景继续优化 AOD 估计算法和 TA 估计算法,以便减小速度估计误差。

参考文献:

- [1] ZHANG H, ABDI A. Mobile speed estimation using diversity combining in fading channels [C]//IEEE Global Telecommunications Conference. Dallas: IEEE, 2004:3685-3689.
- [2] ZHENG Y R, XIAO C S. Mobile speed estimation for broadband wireless communications over Rician fading channels [J]. IEEE Transactions on Wireless Communications, 2009, 8(1): 1-5.
- [3] DOUKAS A, KALIVAS G. Doppler spread estimation in frequency selective Rayleigh channels for OFDM systems [C]//The 5th International Symposium in Communication Systems, Networks and Digital Signal Processing. Patras: IEEE, 2006: 1-5.
- [4] MIRZA J, IKRAMY A A, BAJWAZ H. Maximum Doppler shift frequency estimation using Autocorrelation Function for MIMO OFDM systems [C]//2010 International Conference on Information and Emerging Technologies. Karachi: IEEE, 2010: 1-4.
- [5] AUSTIN M D, STUBER G L. Eigen-based Doppler estimation for differentially coherent CPM [J]. IEEE Transactions on Vehicular Technology, 1994, 43 (3): 781-785.
- [6] LIN J, PROAKIS J G. A parametric method for Doppler spectrum estimation in mobile radio channels[C]//The 27th Annual Conference on Information Science System. Baltimore: IEEE, 1993: 1-5.
- [7] HADDAD M, HERCULEA D G, ALTMAN E, et al. Mobility state estimation in LTE [C]//2016 IEEE Wireless Communications and Networking Conference. Doha: IEEE, 2016: 1-6.
- [8] HADDAD M, HERCULEA D G, CHEN C S, et al. Online mobile user speed estimation: Performance and tradeoff considerations[C]//2017 14th IEEE Annual Consumer Communications & Networking Conference. Las Vegas: IEEE, 2017: 974-979.
- [9] BANERJEE K, VAN DINH T, LEVKOVA L. Velocity estimation from monocular video for automotive applications using convolutional neural networks [C]//2017 IEEE Intelligent Vehicles Symposium. Los Angeles: IEEE, 2017: 373-378.
- [10] ILG E, MAYER N, SAIKIA T, et al. FlowNet 2. 0: evolution of optical flow estimation with deep networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1647-1655.
- [11] VARADE S W, KULAT K D. Robust algorithms for DOA estimation and adaptive beamforming for smart antenna application[C]//2009 Second International Conference on Emerging Trends in Engineering & Technology. Nagpur: IEEE, 2009: 1195-1200.
- [12] SHIRVANI-MOGHADDAM S, AKBARI F. A novel ULA-based geometry for improving AOA estimation[J]. EURASIP Journal on Advances in Signal Processing, 2011(1): 1-11.
- [13] ROY R, KAILATH T. ESPRIT-estimation of signal parameters via rotational invariance techniques [J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1989, 37(7): 984-995.
- [14] YANG W, CHEN J S, TAN Z H. An ESPRIT based DOA estimation for CDMA frequency-selective fading channels [C]//The 14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications. Beijing: IEEE, 2003: 2098-2101.
- [15] WANG Y Y, HUANG S C. An ESPRIT-based algorithm for 2D-DOA estimation [J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2011, E94-A(9): 1847-1850.
- [16] 阙庆云, 许京伟, 廖桂生. 和差天线空时自适应测角方法及性能分析[J]. 电子学报, 2023, 51(1): 42-49.
- [17] 郭丽丽. 有源天线阵列技术[D]. 西安: 西安电子科技大学, 2022.
- [18] 彭张节, 洪赟, 许威, 等. LTE 系统中基于估计波束到达角的多用户调度算法[J]. 东南大学学报(自然科学版), 2013, 43(3): 458-462.
- [19] ABDZADEH-ZIABARI H, ZHU W P, SWAMY M N S. Improved coarse timing estimation in OFDM systems using high-order statistics [J]. IEEE Transactions on Communications, 2016, 64(12): 5239-5253.
- [20] 宋宝相, 肖竹, 田红心, 等. 基于 UWB 最强路径检测的 TOA 估计算法及性能研究[J]. 测试技术学报, 2010, 24(3): 254-258.

作者简介:

周宝龙 男, 1977 年生于湖南岳阳, 博士, 教授级高级工程师, 主要研究方向为无线通信、多天线信号处理和检测。

陆 犇 男, 1973 年生于江苏南京, 博士, 研究员, 主要研究方向为宽带无线通信、大规模 MIMO。

杨洪生 男, 1967 年生于河北沧州, 博士, 教授级高级工程师, 主要研究方向为新一代移动通信。

唐 亮 男, 1983 年生于江苏南京, 博士, 研究员, 主要研究方向为宽带无线通信、自组织网络。

DOI:10.20079/j.issn.1001-893x.231202001

双 RIS 辅助 MIMO 系统混合张量信道估计算法^{*}

李双志, 邢益博, 雷豪杰

(郑州大学 电气与信息工程学院, 郑州 450001)

摘 要:为解决在双可重构智能超表面(Reconfigurable Intelligent Surface, RIS)系统中获取高维信道状态信息(Channel State Information, CSI)的挑战,提出了一种基于混合张量分解的多链路联合信道估计算法。首先,通过设计导频传输机制,将单反射链路和双反射链路的接收信号分别建模为平行因子模型和平行因子塔克(Tucker)张量模型,将信道估计问题转化为混合张量因子矩阵的拟合问题。然后,考虑到多条链路之间共享的 CSI,采用一种基于交替最小二乘迭代算法来分解混合张量,以有效估计出因子矩阵。最后,通过对该混合张量进行唯一性分析,与传统的 Khatri-Rao 分解方法相比,所提方法具备更为灵活的参数设计特点。仿真实验结果表明,该方法能够在训练块数小于 RIS 单元数的情况下有效估计反射链路 CSI。

关键词:多用户 MIMO;信道估计;双可重构智能超表面;混合张量分解

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN929.5 文献标志码:A 文章编号:1001-893X(2025)03-0407-08

A Hybrid Tensor Channel Estimation Algorithm for Double-RIS Assisted MIMO Systems

LI Shuangzhi, XING Yibo, LEI Haojie

(School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: To solve the challenge of obtaining high-dimensional channel state information (CSI) in reconfigurable intelligent surface (RIS) systems, a multi-link joint channel estimation algorithm based on hybrid tensor decomposition is proposed. Firstly, by designing a pilot transmission mechanism, the received signals of single reflection links and double reflection links are respectively modeled as parallel factor models and parallel factor and Tucker models, thereby transforming the channel estimation problem into a fitting problem of hybrid tensor factor matrices. Subsequently, considering the shared CSI among multiple links, an alternating least squares iterative algorithm is employed to decompose the hybrid tensor, effectively estimating the factor matrices. Finally, through a uniqueness analysis of the hybrid tensor, compared with the traditional Khatri-Rao decomposition method, this method possesses more flexible parameter design characteristics. Simulation results demonstrate that this method can effectively estimate the reflection link CSI when the number of training blocks is less than the number of RIS units.

Key words: multi-user MIMO; channel estimation; double-RIS; hybrid tensor decomposition

^{*} 收稿日期:2023-12-02;修回日期:2024-02-28

基金项目:国家自然科学基金青年科学基金项目(61901416);中国博士后科学基金第3批特别资助(站前(2021TQ0304));河南省青年人才托举工程(2024HYTP026)

通信作者:李双志 Email:ielsz@zzu.edu.cn

0 引言

可重构智能超表面 (Reconfigurable Intelligent Surface, RIS)^[1-3] 是 6G 系统中一种新兴且具有前景的技术。该技术具备调节无线信道的能力,通过控制入射信号的反射相移和幅度,协同地重塑无线信道,从而有助于基站 (Base Station, BS) 和用户终端 (User Terminal, UT) 之间的信息传输。由于其反射元件的无源特性,相较于传统的中继基站系统, RIS 具有低能耗低硬件成本的优势。

在 RIS 辅助的无线通信系统中,关键设计如波束成形^[4]和波束赋形^[5]、超可靠传输^[6-7]、链路自适应和资源优化^[8],在很大程度上依赖于 RIS 系统内两个子信道的信道状态信息 (Channel State Information, CSI)。为了优化 RIS 相移矩阵,需要获取两个子信道的级联信道 CSI。然而,由于 RIS 的无源和被动性质, RIS 本身缺乏收发射频链路,也无法进行信号处理,因此获得准确的信道状态信息被广泛认为是 RIS 辅助通信系统一个主要挑战。

目前大部分基于 RIS 系统的信道估计研究集中在单 RIS 系统^[9-12],仅涉及单一反射链路的 CSI 获取。然而,在城市区域、卫星通信以及室内通信等许多应用场景中,为确保多用户信号的可靠传输并减少信号盲区^[13],需要多个 RIS 的协同辅助来完成通信。文献[14-16]研究了双 RIS 设计。双 RIS 架构可以通过双反射链路实现实质性的协同功率增益提升^[5]。在这种架构中,一个 RIS 靠近 BS,另一个靠近用户,以增强容量。尽管这种协同功率增益更具吸引力,但需要更多的训练开销。同时,双 RIS 系统的反射链路更为复杂,除了传统的单反射链路之外,还需要在 RIS 之间的双反射链路中估计更多的信道系数,使得信道估计变得更加困难。文献[14]研究了具有全无源双 RIS 系统的信道估计,但未考虑单反射链路,并且仅针对单用户情况采用基于最小二乘协议进行了估计。文献[15]则针对多用户双 RIS 辅助系统提出了一种解耦的信道估计方案,依次估计两个单反射链路,然后以去耦的方式对双反射链路进行滤波估计,有效降低了导频开销,但由于不完美的信号消除引起的残余干扰,降低了信道估计精度。文献[16]则针对有源双 RIS 系统提出了一种双时间尺度的信道估计协议,以降低信道估计导频开销。

本文提出了一种新颖而高效的信道估计方案,旨在实现双 RIS 辅助的多用户上行多输入多输出

(Multiple-Input Multiple-Output, MIMO) 系统多链路的联合估计。该方案解决了双 RIS 系统中单反射链路以及双反射链路的级联 CSI 获取难题。在该方案中,为了缓解双 RIS 系统中单反射和双反射链路信号耦合的问题,首先引入了一种导频传输机制。通过将单反射链路以及双反射链路的接收信号分别建模为平行因子 (Parallel Factor, PARAFAC) 以及平行因子塔克 (PARAFAC and Tucker, PARATUCK) 张量模型,基于 3 条不同链路之间的公共信道信息,提出了两种新的张量混合方式。通过 PARAFAC-PARATUCK 混合张量模型,实现了对 3 条反射链路的联合估计,并对混合张量模型进行了唯一性分解研究。与传统张量 Khatri-Rao 分解 (Khatri-Rao Factorization, KRF) 算法^[10]对比的仿真结果显示,所提算法在各种参数限制下具有较好的估计性能和较低的导频开销。

本文使用的数学公式符号说明如下:矩阵用粗体大写字母表示;列向量用粗体小写字母表示;张量用大写花体字母表示; $\mathbf{A}^T$ 、 $\mathbf{A}^H$ 以及 $\mathbf{A}^+$ 分别表示矩阵 $\mathbf{A}$ 的转置、共轭转置以及伪逆; $\text{diag}(\mathbf{a})$ 表示 $\mathbf{a}$ 的对角化矩阵; $D_i(\mathbf{A})$ 为由 $\mathbf{A}$ 的第 i 行构成的对角矩阵; $\times_n$ 表示张量 n 模乘积; $\text{vec}(\cdot)$ 表示矩阵的向量化形式; $\|\cdot\|$ 表示矩阵或张量的 F 范数; $\mathbf{I}_M$ 表示维度为 $M \times M$ 的单位矩阵; $\mathbf{C}^{m \times n}$ 表示维度为 $m \times n$ 的复数矩阵; $\mathcal{CN}(m, n)$ 表示均值为 m 、方差为 n 的复高斯分布; $\otimes$ 和 $\odot$ 分别表示 Kronecker 乘积以及 Khatri-Rao 乘积。

1 系统模型与导频结构

本文考虑双 RIS 辅助多用户的大规模 MIMO 上行通信系统,其系统模型如图 1 所示。 U 个单天线用户通过双 RIS (称为 RIS_1 与 RIS_2) 反射与 BS 端进行通信,BS 端配置为 M 根天线,可以实现对 RIS_1 与 RIS_2 的被动反射元件的相移的实时精准控制。 RIS_1 分布在用户集群附近,由 N_1 个无源被动反射元件组成。 RIS_2 分布在基站附近,由 N_2 个无源被动反射元件组成。假设用户到基站的直接链路由于遮挡物和障碍等不利的传播因素而受阻,主要考虑经过 RIS 的反射链路,分为 3 个部分:

1) 用户端发送信号经过 RIS_1 反射到达基站的单反射链路 1:由 $\mathbf{G}_1 \in \mathbb{C}^{N_1 \times U}$ 和 $\mathbf{H}_1 \in \mathbb{C}^{M \times N_1}$ 构成,分别代表用户与 RIS_1 以及 RIS_1 与 BS 之间的信道,如图 1 中黄色链路所示。

2) 用户端发送信号经过 RIS_2 反射到达基站的单反射链路 2: 由 $\mathbf{G}_2 \in \mathbb{C}^{N_2 \times U}$ 和 $\mathbf{H}_2 \in \mathbb{C}^{M \times N_2}$ 构成, 分别代表用户与 RIS_2 以及 RIS_2 与 BS 之间的信道, 如图 1 中红色链路所示。

3) 用户端发送信号经过 RIS_1 与 RIS_2 的双反射链路反射到达基站, 形成链路 3: 由 $\mathbf{G}_1 \in \mathbb{C}^{N_1 \times U}$ 、 $\mathbf{B} \in \mathbb{C}^{N_2 \times N_1}$ 以及 $\mathbf{H}_2 \in \mathbb{C}^{M \times N_2}$ 构成, $\mathbf{B} \in \mathbb{C}^{N_2 \times N_1}$ 代表两个 RIS 之间的双反射链路信道, 如图 1 中绿色链路所示。

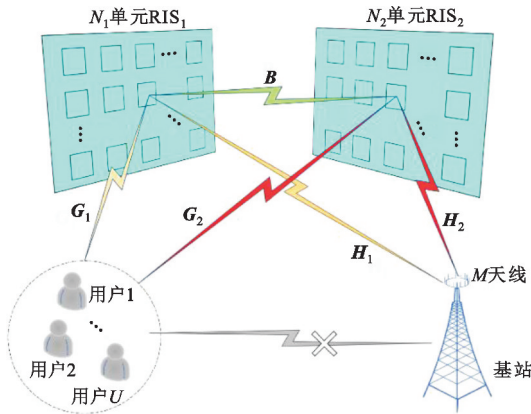


图 1 双 RIS 辅助多用户系统模型

对本文所用信道进行建模, 信道模型均采用独立同分布的瑞利衰落信道模型。具体地, $\mathbf{H}_1 \triangleq \sqrt{L_{\alpha_1}} \cdot \mathbf{A}_1$, $\mathbf{H}_2 \triangleq \sqrt{L_{\alpha_2}} \mathbf{A}_2$, $\mathbf{G}_1 \triangleq \sqrt{L_{\beta_1}} \mathbf{C}_1$, $\mathbf{G}_2 \triangleq \sqrt{L_{\beta_2}} \cdot \mathbf{C}_2$, $\mathbf{B} \triangleq \sqrt{L_{\eta}} \mathbf{D}$, L_{α} 、 L_{β} 以及 L_{η} 分别表示对应信道中的路径增益, 其服从大尺度衰落分布, 反映了不同信道之间的路径损耗差异, $\mathbf{A}_1$ 、 $\mathbf{A}_2$ 、 $\mathbf{C}_1$ 、 $\mathbf{C}_2$ 以及 $\mathbf{D}$ 为服从瑞利衰落的小尺度衰落信道模型, $[\mathbf{A}_1]_{m n_1}$, $[\mathbf{A}_2]_{m n_2} \sim \mathcal{CN}(0, 1)$, $[\mathbf{C}_1]_{n_1 u}$, $[\mathbf{C}_2]_{n_2 u} \sim \mathcal{CN}(0, 1)$, $[\mathbf{D}]_{n_1 n_2} \sim \mathcal{CN}(0, 1)$, 其中, $1 \leq m \leq M$, $1 \leq n_1 \leq N_1$, $1 \leq n_2 \leq N_2$, $1 \leq u \leq U$ 。基于双 RIS 系统分布特性, 信道 $\mathbf{H}_1$ 与 $\mathbf{G}_2$ 的路径距离相对 $\mathbf{G}_1$ 与 $\mathbf{H}_2$ 较远, 路径损耗较高, 路径增益因子存在差异, 这里引入一个缩放因子 $\omega_1 = \sqrt{L_{\alpha_1}} / \sqrt{L_{\alpha_2}}$ 与 $\omega_2 = \sqrt{L_{\beta_2}} / \sqrt{L_{\beta_1}}$ 来衡量双 RIS 系统中不同信道路径传输质量的差异。当缩放因子 ω_1 和 ω_2 较小时, 单一链路的估计结果容易导致系统整体的估计误差累积, 因此在双 RIS 系统中, 采用有效的联合多链路估计策略有利于减少这种估计误差的传播, 进而提升系统估计的性能。

在双 RIS 系统中, 由于链路较多, 分为单反射链路和双反射链路, 总的接收信号为耦合信号, 因此本文要分阶段获取耦合的接收信号模型。所提信道估

计协议分为 3 个阶段: 首先是阶段 1 关闭 RIS_2 获取链路 1 的接收信号; 然后阶段 2 关闭 RIS_1 获链路 2 的接收信号; 最后是同时打开 RIS_1 和 RIS_2 得到多条链路耦合接收信号模型。在这 3 个阶段中, 采用相同的导频时间帧结构, 如图 2 所示。

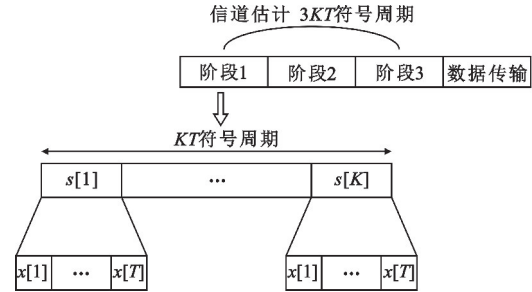


图 2 时域化导频帧结构设置

在本文的信道估计协议中, 将 RIS_1 和 RIS_2 的相移模式 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 划分为 K 块, $k=1, 2, \dots, K$, 每一块的 $s_1[k]$ 和 $s_2[k]$ 随着块数 k 的变化而变化, 对于第 k 个块上 RIS_1 和 RIS_2 相移模式为

$$\mathbf{s}_1[k] = [s_{1k} e^{j\phi_1}, \dots, s_{N_1 k} e^{j\phi_{N_1}}]^T \in \mathbb{C}^{N_1 \times 1} \quad (1a)$$

$$\mathbf{s}_2[k] = [s_{1k} e^{j\phi_1}, \dots, s_{N_2 k} e^{j\phi_{N_2}}]^T \in \mathbb{C}^{N_2 \times 1} \quad (1b)$$

式中: $s_{n_1 k}, s_{n_2 k} \in \{0, 1\}$ 和 $\phi_{n_1}, \phi_{n_2} \in (0, 2\pi]$ 分别表示 RIS_1 和 RIS_2 第 n 个元素的反射振幅以及相移, $n_1 = 1, 2, \dots, N_1$, $n_2 = 1, 2, \dots, N_2$; $\mathbf{S}_1 = [s_1[1], \dots, s_1[K]]^T \in \mathbb{C}^{K \times N_1}$ 以及 $\mathbf{S}_2 = [s_2[1], \dots, s_2[K]]^T \in \mathbb{C}^{K \times N_2}$ 分别代表 RIS_1 和 RIS_2 总的相移矩阵。在每个阶段中, 用户端发送导频为正交导频 $\mathbf{X} \in \mathbb{C}^{U \times T}$, 这些导频序列在 K 个块中反复重复, 满足 $\mathbf{X}\mathbf{X}^H = \mathbf{I}_U$ 。为了确保其正交性, 将导频信号划分为 T 时隙, $\mathbf{X} = [\mathbf{x}[1], \dots, \mathbf{x}[T]]$, 需要导频信号 $\mathbf{X}$ 满足行满秩, 这意味着 $T \geq U$, 因此在本文所用信道估计协议下总的训练序列开销为 $3KT$ 。

1.1 信号模型

在确定了双 RIS 的相移矩阵以及多链路信道所用信道模型之后, 对于基站端 3 个不同阶段在第 k 块由 U 个单天线用户发送导频的接收信号模型, 可以分别表示为

$$\mathbf{Y}_1[k] = \mathbf{H}_1 \text{diag}(\mathbf{s}_1[k]) \mathbf{G}_1 \mathbf{X} + \mathbf{N}_1[k] \quad (2a)$$

$$\mathbf{Y}_2[k] = \mathbf{H}_2 \text{diag}(\mathbf{s}_2[k]) \mathbf{G}_2 \mathbf{X} + \mathbf{N}_2[k] \quad (2b)$$

$$\mathbf{Y}_3[k] = (\mathbf{H}_1 \text{diag}(\mathbf{s}_1[k]) \mathbf{G}_1 + \mathbf{H}_2 \text{diag}(\mathbf{s}_2[k]) \mathbf{G}_2 + \mathbf{H}_2 \text{diag}(\mathbf{s}_2[k]) \mathbf{B} \text{diag}(\mathbf{s}_1[k]) \mathbf{G}_1) \mathbf{X} + \mathbf{N}_3[k] \quad (2c)$$

式中: $k=1, \dots, K; \mathbf{Y}_1[k], \mathbf{Y}_2[k] \in \mathbb{C}^{M \times U}$ 分别表示链路 1 与链路 2 的接收信号; $\mathbf{Y}_3[k] \in \mathbb{C}^{M \times U}$ 为系统中链路 1、链路 2 以及链路 3 的公共耦合接收信号; $\mathbf{N}_1[k], \mathbf{N}_2[k], \mathbf{N}_3[k] \in \mathbb{C}^{M \times U}$ 表示三阶段接收信号的加性高斯白噪声。这里可以通过一些变换来进一步简化接收信号模型的阐述, 基于 $\mathbf{X}\mathbf{X}^H = \mathbf{I}_U$, 取 $\bar{\mathbf{Y}}_1[k] = \mathbf{Y}_1[k]\mathbf{X}^H, \bar{\mathbf{Y}}_2[k] = \mathbf{Y}_2[k]\mathbf{X}^H$, 对于 $\mathbf{Y}_3[k]$, 首先要进行信号解耦, 对式 (2c) 进行处理, 将接收信号 $\mathbf{Y}_1[k]$ 和 $\mathbf{Y}_2[k]$ 的部分过滤后简化得到第三部分 $\bar{\mathbf{Y}}_3[k] = (\mathbf{Y}_3[k] - \mathbf{Y}_1[k] - \mathbf{Y}_2[k])\mathbf{X}^H$ 。简化后的 3 种张量的切片形式如下:

$$\bar{\mathbf{Y}}_1[k] = \mathbf{H}_1 \text{diag}(\mathbf{s}_1[k]) \mathbf{G}_1 + \bar{\mathbf{N}}_1[k] \quad (3a)$$

$$\bar{\mathbf{Y}}_2[k] = \mathbf{H}_2 \text{diag}(\mathbf{s}_2[k]) \mathbf{G}_2 + \bar{\mathbf{N}}_2[k] \quad (3b)$$

$$\bar{\mathbf{Y}}_3[k] = \mathbf{H}_2 \text{diag}(\mathbf{s}_2[k]) \mathbf{B} \text{diag}(\mathbf{s}_1[k]) \mathbf{G}_1 + \bar{\mathbf{N}}_3[k] \quad (3c)$$

在无噪声情况下, 将切片模型 $\bar{\mathbf{Y}}_1[k], \bar{\mathbf{Y}}_2[k]$ 以及 $\bar{\mathbf{Y}}_3[k]$ 在 K 的维度进行切片的堆叠可以得到 $\mathcal{Y}_1 \in \mathbb{C}^{M \times U \times K}, \mathcal{Y}_2 \in \mathbb{C}^{M \times U \times K}$ 以及 $\mathcal{Y}_3 \in \mathbb{C}^{M \times U \times K}$ 3 种不同的张量模型, 分别表示 3 条反射链路的接收信号模型, 其中, $\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 遵循典型的 PARAFAC 模型^[10], $\mathcal{Y}_3$ 遵循 PARATUCK 模型^[17]。3 种无噪声张量的每个 (m, u, k) 项可以写为以下标量形式:

$$\begin{cases} y_{1, muk} = \sum_{n_1=1}^{N_1} h_{1, m, n_1} g_{1, u, n_1} s_{1, k, n_1} \\ y_{2, muk} = \sum_{n_2=1}^{N_2} h_{2, m, n_2} g_{2, u, n_2} s_{2, k, n_2} \\ y_{3, muk} = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} h_{2, m, n_2} s_{2, k, n_2} b_{n_2, n_1} s_{1, k, n_1} g_{1, m, n_1} \end{cases} \quad (4)$$

式中: $\mathbf{h}_{m, n} = [\mathbf{H}]_{m, n}; \mathbf{g}_{u, n} = [\mathbf{G}]_{u, n}; \mathbf{s}_{k, n} = [\mathbf{S}]_{k, n}$ 。下面讨论 3 种接收信号张量模型对应的三线性模式展开。进行如下 3 种张量切片堆叠为 $\mathbf{Y}_i^{(1)} = [\mathbf{Y}_i[1], \dots, \mathbf{Y}_i[K]]^T, \mathbf{Y}_i^{(2)} = [\mathbf{Y}_i^T[1], \dots, \mathbf{Y}_i^T[K]]^T$ 以及 $\mathbf{Y}_i^{(3)} = [\text{vec}(\mathbf{Y}_i[1]), \dots, \text{vec}(\mathbf{Y}_i[K])]^T, i=1, 2, 3$, 得到 $\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 的 PARAFAC 模式展开为

$$\begin{cases} \mathbf{Y}_1^{(1)} = (\mathbf{S}_1 \odot \mathbf{G}_1^T) \mathbf{H}_1^T \\ \mathbf{Y}_1^{(2)} = (\mathbf{S}_1 \odot \mathbf{H}_1) \mathbf{G}_1 \\ \mathbf{Y}_1^{(3)} = (\mathbf{G}_1^T \odot \mathbf{H}_1) \mathbf{S}_1^T \end{cases} \quad (5a)$$

$$\begin{cases} \mathbf{Y}_2^{(1)} = (\mathbf{S}_2 \odot \mathbf{G}_2^T) \mathbf{H}_2^T \\ \mathbf{Y}_2^{(2)} = (\mathbf{S}_2 \odot \mathbf{H}_2) \mathbf{G}_2 \\ \mathbf{Y}_2^{(3)} = (\mathbf{G}_2^T \odot \mathbf{H}_2) \mathbf{S}_2^T \end{cases} \quad (5b)$$

以及 $\mathcal{Y}_3$ 的 PARATUCK 模型的 3 种展开为

$$\mathbf{Y}_3^{(1)} = (\mathbf{I}_K \otimes \mathbf{G}_1^T) \begin{bmatrix} D_1(\mathbf{S}_1) \mathbf{B}^T D_1(\mathbf{S}_2) \\ \vdots \\ D_K(\mathbf{S}_1) \mathbf{B}^T D_K(\mathbf{S}_2) \end{bmatrix} \mathbf{H}_2^T = (\mathbf{I}_K \otimes \mathbf{G}_1^T) \mathbf{W}_1 \mathbf{H}_2^T \in \mathbb{C}^{U \times K \times M} \quad (6a)$$

$$\mathbf{Y}_3^{(2)} = (\mathbf{I}_K \otimes \mathbf{H}_2) \begin{bmatrix} D_1(\mathbf{S}_2) \mathbf{B} D_1(\mathbf{S}_1) \\ \vdots \\ D_K(\mathbf{S}_2) \mathbf{B} D_K(\mathbf{S}_1) \end{bmatrix} \mathbf{G}_1 = (\mathbf{I}_K \otimes \mathbf{H}_2) \mathbf{W}_2 \mathbf{G}_1 \in \mathbb{C}^{M \times K \times U} \quad (6b)$$

$$\mathbf{Y}_3^{(3)} = (\mathbf{G}_1^T \otimes \mathbf{H}_2) \text{diag}(\mathbf{b}) (\mathbf{S}_1^T \odot \mathbf{S}_2^T) \in \mathbb{C}^{M \times U \times K} \quad (6c)$$

$$\text{式中: 矩阵 } \mathbf{W}_1 = \begin{bmatrix} D_1(\mathbf{S}_1) \mathbf{B}^T D_1(\mathbf{S}_2) \\ \vdots \\ D_K(\mathbf{S}_1) \mathbf{B}^T D_K(\mathbf{S}_2) \end{bmatrix} \in \mathbb{C}^{K N_1 \times N_2}; \mathbf{W}_2 = \begin{bmatrix} D_1(\mathbf{S}_2) \mathbf{B} D_1(\mathbf{S}_1) \\ \vdots \\ D_K(\mathbf{S}_2) \mathbf{B} D_K(\mathbf{S}_1) \end{bmatrix} \in \mathbb{C}^{K N_2 \times N_1}。$$

对于 $\mathbf{Y}_3^{(3)}$, 通过矩阵

性质 $\text{vec}(\mathbf{A} \text{diag}(\text{vec}(\mathbf{C})) \mathbf{B}) = (\mathbf{B}^T \odot \mathbf{A}) \text{vec}(\mathbf{C})$, 可以进一步得到

$$\text{vec}(\mathbf{Y}_3^{(3)}) = [(\mathbf{S}_1^T \odot \mathbf{S}_2^T)^T \odot (\mathbf{G}_1^T \otimes \mathbf{H}_2)] \text{vec}(\mathbf{b}) = \mathbf{W}_3 \mathbf{b} \quad (7)$$

式中: $\mathbf{W}_3 = [(\mathbf{S}_1^T \odot \mathbf{S}_2^T)^T \odot (\mathbf{G}_1^T \otimes \mathbf{H}_2)] \in \mathbb{C}^{M \times U \times K \times N_1 \times N_2}; \mathbf{b} = \text{vec}(\mathbf{B})$ 。

本文接下来将利用这 3 种张量展开的代数结构, 基于其模型间公共因子矩阵来设计双 RIS 系统的联合信道估计算法。

1.2 唯一性分析

张量信号的唯一分解特性使其在信道估计方案设计中具有独特的优势。目前, Kruskal 条件^[18]被广泛认为是平行因子张量唯一分解的充分条件。在本文中, 针对 $\mathcal{Y}_1$ 与 $\mathcal{Y}_2$, 可以建立 Kruskal 条件如下:

$$k(\mathbf{H}_1) + k(\mathbf{G}_1^T) + k(\mathbf{S}_1) \geq 2N_1 + 2 \quad (8)$$

$$k(\mathbf{H}_2) + k(\mathbf{G}_2^T) + k(\mathbf{S}_2) \geq 2N_2 + 2 \quad (9)$$

可以简化为 $\min(M, N_1) + \min(K, N_1) \geq N_1 + 2$ 和 $\min(M, N_2) + \min(K, N_2) \geq N_2 + 2$, 此时张量 $\mathcal{Y}_1$ 和 $\mathcal{Y}_2$ 存在尺度模糊的前提下, 能够实现唯一分解。 $\mathbf{S}_1$ 与 $\mathbf{S}_2$ 为已知, 尺度模糊为 $\hat{\mathbf{H}}_1 = \mathbf{H}_1 \mathbf{\Delta}_{H_1}, \hat{\mathbf{G}}_1^T = \mathbf{G}_1^T \mathbf{\Delta}_{G_1}$ 和 $\hat{\mathbf{H}}_2 = \mathbf{H}_2 \mathbf{\Delta}_{H_2}, \hat{\mathbf{G}}_2^T = \mathbf{G}_2^T \mathbf{\Delta}_{G_2}$, 其中 $\mathbf{\Delta}$ 代表包含尺度模糊的对角矩阵, 满足 $\mathbf{\Delta}_{H_1} \mathbf{\Delta}_{G_1} = \mathbf{I}_{N_1}, \mathbf{\Delta}_{H_2} \mathbf{\Delta}_{G_2} = \mathbf{I}_{N_2}$ 。对于 $\mathcal{Y}_3$ 的 PARATUCK 模型, 在满足 $\mathbf{H}_2, \mathbf{G}_1$ 以及 $\mathbf{B}$ 矩阵满秩且 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 拥有相同的列数时, 满足 $N_1 = N_2$ 时,

PARATUCK 模型满足唯一性分解, 存在列模糊以及尺度模糊为 $\bar{\mathbf{H}}_2(\mathbf{P}\mathbf{A}_H) = \mathbf{H}_2$, $\bar{\mathbf{G}}_1(\mathbf{Q}\mathbf{A}_G) = \mathbf{G}_1$ 以及 $(\mathbf{A}_H)^{-1} \mathbf{P}^T \mathbf{B} \mathbf{Q} (\mathbf{A}_G)^{-1} = \mathbf{B}$ ($\mathbf{P}, \mathbf{Q} \in \mathbb{C}^{N \times N}$) 为列交换矩阵, $\mathbf{A}_H$ 和 $\mathbf{A}_G$ 为对角模糊矩阵。对于双 RIS 系统, 单一信道的尺度模糊难以消除, 在评估不同信道时, 我们以级联信道为估计性能的标准, 并对 3 个不同的反射链路分开进行估计性能评估。

2 基于混合张量分解的信道估计方法

在 1.1 节中, 对给定的三阶段接收信号张量模型以及其三线性展开进行了分析。在本节, 将结合 3 种张量模型, 利用它们的共享信道部分进行混合张量分解, 从而提出更有效的信道估计方案, 并分析其需要的参数限制。

2.1 PARAFAC-PARATUCK 混合模型

本节提出的混合模型分为两个部分: 第一部分对接收信号 $\mathcal{Y}_1$ 和 $\mathcal{Y}_3$ 进行模式展开, 利用 $\mathcal{Y}_1$ 的单反射链路与 $\mathcal{Y}_3$ 的双反射链路中的共享信道 $\mathbf{G}_1$ 来形成一个新的混合张量信号模式展开; 第二部分对接收信号 $\mathcal{Y}_2$ 和 $\mathcal{Y}_3$ 进行模式展开, 利用 $\mathcal{Y}_2$ 和 $\mathcal{Y}_3$ 链路中共享信道 $\mathbf{H}_2$ 来得到另一个混合张量信号模式展开。具体地, 对于公式 (5a) 中的 $\mathbf{Y}_1^{(2)} \in \mathbb{C}^{MK \times U}$ 以及公式 (6b) 中的 $\mathbf{Y}_3^{(2)} \in \mathbb{C}^{MK \times U}$, 可以进行如下的堆叠:

$$\mathbf{Y}_{13} = \begin{bmatrix} \mathbf{Y}_1^{(2)} \\ \mathbf{Y}_3^{(2)} \end{bmatrix} = \begin{bmatrix} (\mathbf{S}_1 \odot \mathbf{H}_1) \\ (\mathbf{I}_K \otimes \mathbf{H}_2) \mathbf{W}_2 \end{bmatrix} \mathbf{G}_1 \quad (10)$$

该模式展开主要基于两个混合张量相同的因子矩阵 $\mathbf{G}_1$ 。

而第二种混合张量模式展开是基于公式 (5b) 中的 $\mathbf{Y}_2^{(1)} \in \mathbb{C}^{UK \times M}$ 以及 (6a) 中的 $\mathbf{Y}_3^{(1)} \in \mathbb{C}^{UK \times M}$, 可以进行如下的堆叠:

$$\mathbf{Y}_{23} = \begin{bmatrix} \mathbf{Y}_2^{(1)} \\ \mathbf{Y}_3^{(1)} \end{bmatrix} = \begin{bmatrix} (\mathbf{S}_2 \odot \mathbf{G}_2^T) \\ (\mathbf{I}_K \otimes \mathbf{G}_1) \mathbf{W}_1 \end{bmatrix} \mathbf{H}_2^T \quad (11)$$

该模式展开主要基于两个混合张量相同的因子矩阵 $\mathbf{H}_2^T$ 。

2.2 混合张量模型迭代分解算法

在提出了两种 PARAFAC-PARATUCK 张量混合模型的模式展开后, 可以通过 ALS 算法迭代联合估计双 RIS 之间信道以及公共链路信道 $\mathbf{G}_1$ 和 $\mathbf{H}_2$ 。在有噪声的情况下, 通过模式展开式 (7)、(10)、(11) 的噪声版本, 在已知双 RIS 相移矩阵 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 的情况下, 得到基于 ALS 算法的迭代解, 其代价函数

分别如下:

$$\hat{\mathbf{H}}_2 = \underset{\mathbf{H}_2}{\operatorname{argmin}} \left\| \mathbf{Y}_{23} - \begin{bmatrix} (\mathbf{S}_2 \odot \mathbf{G}_2^T) \\ (\mathbf{I}_K \otimes \mathbf{G}_1) \mathbf{W}_1 \end{bmatrix} \mathbf{H}_2^T \right\|_F^2 \quad (12a)$$

$$\hat{\mathbf{G}}_1 = \underset{\mathbf{G}_1}{\operatorname{argmin}} \left\| \mathbf{Y}_{13} - \begin{bmatrix} (\mathbf{S}_1 \odot \mathbf{H}_1) \\ (\mathbf{I}_K \otimes \mathbf{H}_2) \mathbf{W}_2 \end{bmatrix} \mathbf{G}_1 \right\|_F^2 \quad (12b)$$

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\| \operatorname{vec}(\mathbf{Y}_3^{(3)}) - \mathbf{W}_3 \mathbf{b} \right\| \quad (12c)$$

其中在迭代的过程中, 需要初始化随机矩阵 $\mathbf{B}$ 、 $\mathbf{G}_1$ 、 $\mathbf{G}_2$ 和 $\mathbf{H}_1$ 。对于每一次迭代, 首先利用混合模型 $\mathbf{Y}_{23}$ 更新 $\hat{\mathbf{H}}_2^T$:

$$(\hat{\mathbf{H}}_2^T)_i = \begin{bmatrix} (\mathbf{S}_2 \odot \mathbf{G}_2^T) \\ (\mathbf{I}_K \otimes \mathbf{G}_1) \mathbf{W}_1 \end{bmatrix}_{i-1}^\dagger \mathbf{Y}_{23} \quad (13)$$

利用混合模型 $\mathbf{Y}_{13}$ 更新 $\hat{\mathbf{G}}_1$:

$$(\hat{\mathbf{G}}_1)_i = \begin{bmatrix} (\mathbf{S}_1 \odot \mathbf{H}_1) \\ (\mathbf{I}_K \otimes \mathbf{H}_2) \mathbf{W}_2 \end{bmatrix}_i^\dagger \mathbf{Y}_{13} \quad (14)$$

并通过 $\hat{\mathbf{G}}_1$ 、 $\hat{\mathbf{H}}_2^T$ 更新矩阵 $\mathbf{W}_3$ 得到新的 $\hat{\mathbf{b}}$:

$$(\hat{\mathbf{b}})_i = (\mathbf{W}_3^i)_i \operatorname{vec}(\mathbf{Y}_3^{(3)}) \quad (15)$$

最后通过新得到的 $\hat{\mathbf{G}}_1$ 和 $\hat{\mathbf{H}}_2^T$ 去更新 $\hat{\mathbf{G}}_2$ 和 $\hat{\mathbf{H}}_1$:

$$(\hat{\mathbf{G}}_2)_i = (\mathbf{S}_2 \odot \mathbf{H}_{2,i})^{-1} \mathbf{Y}_2^{(2)} \quad (16a)$$

$$(\hat{\mathbf{H}}_1)_i = [(\mathbf{S}_1 \odot \mathbf{G}_{1,i}^T)^{-1} \mathbf{Y}_1^{(1)}]^T \quad (16b)$$

式中: i 表示迭代次数。在每次迭代完成后, 通过利用误差设置阈值判断是否达到收敛, 收敛条件为 $\|e_i - e_{i-1}\| \leq \delta$, $e_i = \|\mathcal{Y}_3 - \mathcal{Y}_{3,i}\|_F^2$, 其中, $\mathcal{Y}_{3,i}$ 被选择作为张量重构的模型, $\delta = 10^{-6}$ 为经过验证设置的跳出迭代的收敛基准。本文所提 PARAFAC-PARATUCK 张量混合迭代算法的具体算法步骤描述如下:

输入: $\mathbf{S}_1, \mathbf{S}_2, \mathbf{X}$ 以及接收信号 $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3$;

输出: $\hat{\mathbf{G}}_1, \hat{\mathbf{G}}_2, \hat{\mathbf{H}}_1, \hat{\mathbf{H}}_2, \hat{\mathbf{b}}$;

1 初始化随机矩阵 $\mathbf{B}, \mathbf{G}_1, \mathbf{G}_2, \mathbf{H}_1$;

2 对接收信号 $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3$ 进行展开以及混合得到 $\mathbf{Y}_{13}, \mathbf{Y}_{23}$

3 迭代部分:

4 $i = i + 1$;

5 while $\|e_i - e_{i-1}\| \geq \delta$ do

6 由混合模型 $\mathbf{Y}_{23}$ 根据式 (13) 更新 $\hat{\mathbf{H}}_2^T$;

7 由混合模型 $\mathbf{Y}_{13}$ 根据式 (14) 更新 $\hat{\mathbf{G}}_1$;

8 由新的 $\hat{\mathbf{H}}_2^T$ 和 $\hat{\mathbf{G}}_1$ 根据式 (15)、(16a)、(16b) 更新 $\hat{\mathbf{G}}_2, \hat{\mathbf{H}}_1, \hat{\mathbf{b}}$

9 重复步骤 6~8 直到迭代到达收敛

10 end while

2.3 算法条件分析

PARAFAC-PARATUCK 张量混合迭代算法需要

利用 PARAFAC 以及 PARATUCK 的三线性展开结构, RIS_1 以及 RIS_2 元素数 $N_1 = N_2$ (这里统一为 N), 同时需要满足 PARAFAC 模型唯一性分解条件:

$$\min(M, N) + \min(K, N) \geq N + 2 \quad (17)$$

在迭代更新过程中涉及矩阵的伪逆运算, 需要

以下两个混合部分 $\begin{bmatrix} (\mathbf{S}_2 \odot \mathbf{G}_2^T) \\ (\mathbf{I}_K \otimes \mathbf{G}_1) \mathbf{W}_1 \end{bmatrix} \in \mathbb{C}^{2U/K \times N}$ 和

$\begin{bmatrix} (\mathbf{S}_1 \odot \mathbf{H}_1) \\ (\mathbf{I}_K \otimes \mathbf{H}_2) \mathbf{W}_2 \end{bmatrix} \in \mathbb{C}^{2KM \times N}$ 列满秩, 即满足 $2UK \geq N$ 以

及 $2KM \geq N$ 。此外, 发射导频 $\mathbf{X}$ 需要满足正交性, 即 $T \geq U$, 因此以下条件也是必要的:

$$K \geq \max(N/2U, N/2M), T \geq U \quad (18)$$

传统的张量分解算法^[10]通过信号解耦基于 KRF 算法依次估计两个单反射链路, 然后以去耦的方式对双反射链路进行最小二乘求解, 所发射导频信号 $\mathbf{X}$ 需要正交, $\mathbf{S}_1$ 以及 $\mathbf{S}_2$ 列满秩以适应线性滤波, 以下条件也是必要的:

$$K \geq N, T \geq U \quad (19)$$

比较条件(18)和(19)可以注意到, 所提出的张量混合迭代算法相较于传统的 KRF 张量分解算法对信道估计所需最小的训练分块数 K 限制性要求较小。具体地, 张量混合迭代算法可以在 $K < N$ 时工作, 而 KRF 算法需要 $K \geq N$, 信道估计所需总的训练序列开销为 $3KT$, 因此, 较小的 K 可以实现较低的导频开销。

3 算法性能仿真分析

本节通过仿真来进一步验证所提出的基于 PARAFAC-PARATUCK 张量混合迭代算法的双 RIS 系统信道估计方案的估计性能, 并与传统的张量分解 KRF 算法^[9]以及最小二乘 (Least Squares, LS) 算法^[15]进行性能对比。本文针对窄带双 RIS 辅助 MIMO 系统基带信号传输过程中的信道估计研究, 导频信号是基于离散傅里叶变换 (Discrete Fourier Transform, DFT) 矩阵设计的。理论上, 本文所提出的张量混合迭代方案与 KRF 算法具有相同的导频传输模式, 但对分块数 K 的限制较小, 在接收信号解耦后, 它对 3 条不同反射链路进行了联合估计, 利用不同链路之间的共享信道信息。相比于单独链路的估计, 这提高了估计精度。同时, 在双 RIS 系统中, 当 $\mathbf{H}_1$ 和 $\mathbf{G}_2$ 的信道传输距离较远、路径损耗较高时, 缩放因子 ω_1 和 ω_2 取值较小, 联合估计能够减少

不同链路单独估计带来的误差传播, 实现更好的估计性能。

在本节中, 所提算法的估计性能由归一化均方误差 (Normalized Mean Square Error, NMSE) 衡量。3 条反射链路级联信道分别为 $\mathbf{H}_{c1} = \mathbf{H}_1 \mathbf{G}_1$, $\mathbf{H}_{c2} = \mathbf{H}_2 \mathbf{G}_2$, $\mathbf{H}_{c3} = \mathbf{H}_2 \mathbf{B} \mathbf{G}_2$, NMSE 由下式定义:

$$\text{NMSE}(\hat{\mathbf{H}}_c) = \frac{1}{R} \sum_{r=1}^R \frac{\|\mathbf{H}_c - \hat{\mathbf{H}}_c\|_F^2}{\|\mathbf{H}_c\|_F^2} \quad (20)$$

式中: R 表示蒙特卡罗仿真次数。我们在节 2.2 讨论了两种张量模型分解的尺度模糊, 在子信道相乘时, 模糊性均被消除, 在构建级联信道时相互补偿, 因此, 可以利用 3 条反射链路级联信道的 NMSE 来评估双 RIS 系统的估计性能。导频信号 $\mathbf{X}$ 以及两个 RIS 的相移矩阵 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 都可以被设计为分别满足 $\mathbf{X}\mathbf{X}^H = \mathbf{I}_U$, $\mathbf{S}_1 \mathbf{S}_1^H = \mathbf{I}_K$ 以及 $\mathbf{S}_2 \mathbf{S}_2^H = \mathbf{I}_K$ 的半酉矩阵, 一种可行的解决方案是将 $\mathbf{X}$, $\mathbf{S}_1$ 以及 $\mathbf{S}_2$ 设计为 DFT 矩阵。信噪比 (Signal-to-Noise Ratio, SNR) (单位: dB) 被定义为接收信号张量模型的功率与基站端对应链路噪声张量功率的比值。3 条链路的信噪比 SNR_l ($l = 1, 2, 3$) 设置为

$$\text{SNR}_l = 10 \lg(\|\bar{\mathcal{Y}}_l\|_F^2 / \|\mathcal{N}_l\|_F^2) \quad (21)$$

式中: $\mathcal{Y}_1 = \bar{\mathcal{Y}}_1 + \mathcal{N}_1$, $\mathcal{Y}_2 = \bar{\mathcal{Y}}_2 + \mathcal{N}_2$ 以及 $\mathcal{Y}_3 = \bar{\mathcal{Y}}_3 + \mathcal{N}_3$ 分别表示 3 条链路的接收信号张量模型。在仿真中将 3 条反射链路的信噪比统一, $\mathcal{N}_1 \in \mathbb{C}^{M \times U \times K}$, $\mathcal{N}_2 \in \mathbb{C}^{M \times U \times K}$ 和 $\mathcal{N}_3 \in \mathbb{C}^{M \times U \times K}$ 分别为 3 条路径上零均值单位方差复高斯噪声张量。通过归一化链路 1 噪声的功率进而设置链路 2 以及链路 3 的功率, 不同链路噪声之间的比值与无噪声信号功率的比值成比例。进一步对所提出算法与传统张量 KRF 算法以及 LS 算法在相同参数下的估计性能进行对比, 并且对 3 条反射链路的级联信道分开进行了评估, 用 NMSE 来衡量算法的估计性能。本文中仿真参数配置如表 1 所示。

表 1 仿真参数

参数	值
基站侧天线数 M	16
RIS_1 与 RIS_2 元素数 N_1 与 N_2	16
分块数 K	2~16
用户数 U	8
时隙数 T	8
信噪比/dB	0~30
缩放因子 ω_1 和 ω_2	0.3~1
蒙特卡罗仿真次数 R	5 000

图 3、图 4 和图 5 分别描述了基于张量混合迭代估计方案和传统 KRF 张量分解方案以及 LS 算法方案在不同信噪比下的性能对比。具体地, 研究了 $\omega_1 = \omega_2 = 1$ 以及 $\omega_1 = \omega_2 = 0.3$ 两种路径传输距离不同的应用场景 (场景 1: $\omega_1 = \omega_2 = 1$ 代表信道 H_1 以及 G_2 的路径传输距离较近的双 RIS 场景; 场景 2: $\omega_1 = \omega_2 = 0.3$ 代表信道 H_1 以及 G_2 的路径传输距离较远的场景)。从图 3 和图 4 中可以看到, 在两种场景中, 随着信噪比的增加, 本文所提方案在链路 1 以及链路 2 的归一化均方误差不断降低, 并在两种场景中相较于传统 KRF 张量分解算法以及传统 LS 算法均实现了小幅度的提升, 在相同的 NMSE 估计误差下相较于 KRF 算法在信噪比有 2 dB 的优势, 较 LS 算法有着 5 dB 的性能优势。图 4 描述了对于双反射链路即链路 3 下该方案表现出的性能优势, 基于张量混合迭代的联合估计方案在两种场景下较传统 KRF 算法以及 LS 算法都有明显的性能提升。特别地, 在场景 2 中, 由于传统算法在链路 1 和链路 2 的估计误差叠加传播, 导致性能下降, 而所提出的联合估计算法能够有效避免这种误差传播, 在高路径损耗的双 RIS 场景依然具有良好的估计性能。

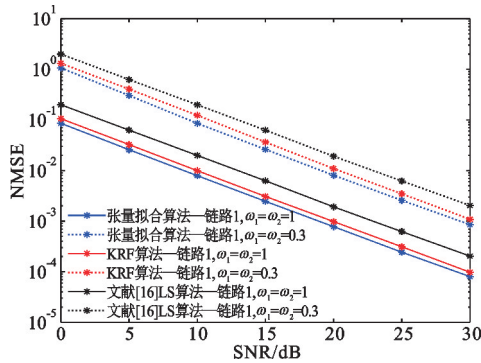


图 3 链路 1 估计性能曲线

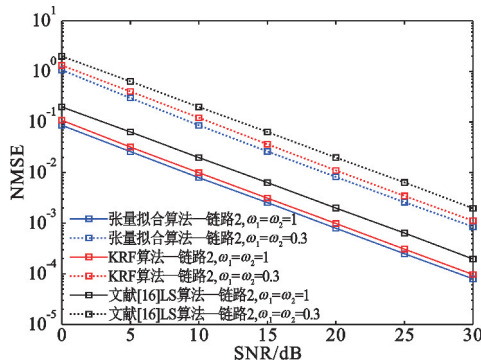


图 4 链路 2 估计性能曲线

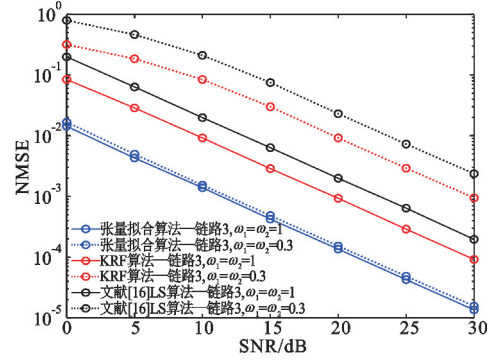
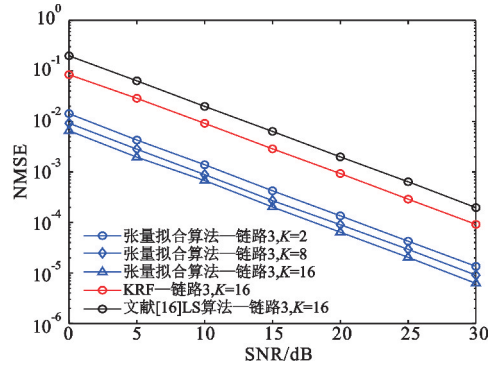


图 5 链路 3 估计性能曲线

图 6 进一步研究了 RIS 分块数 K 对于所提出的算法性能的影响。在与传统 KRF 算法的对比中, 根据在第 2.3 节的理论分析, 所提出的张量混合迭代算法在 $K < N$ 的条件下就能够进行, KRF 算法则需要 $K \geq N$ 才能估计。随着分块数的增加, 估计性能有所提升。然而, 这种提升伴随着导频开销 ($3KT$) 的增加。同时, 对比了相同开销下 ($K = 16$) 的估计性能, 仿真结果表示, 相较于传统 KRF 以及 LS 算法, 所提算法实现了更好的估计性能。这表明, 在相同估计性能下, 所提算法可以通过降低 K 来减少信道估计导频开销。

图 6 不同 K 下链路 3 性能对比曲线

4 结束语

本文提出了一种基于 PARAFAC-PARATUCK 混合张量分解的双 RIS 系统信道估计算法。该算法采用导频传输机制, 在接收端获得三阶段的接收信号; 然后对所接收的信号进行解耦处理, 并有效利用多反射链路间的共享信道信息构建混合张量模型; 最后通过交替最小二乘迭代方法, 实现了多链路联合信道估计。该方法有效避免了单一链路估计误差的扩散, 显著提升了级联信道估计性能。同时, 与传统张量方法相比, 该算法对导频训练参数的要求更为

宽松,对训练块数需求更小,有助于减少导频开销。

本文算法在后续工作中还需进一步完善,比如在用户端具有动态移动性,以及反射链路信道为时变信道或者具有更多参数信息的结构化信道时,如何利用结构化信道特性运用张量分解进行信号处理,对算法进一步补充、完善。

参考文献:

- [1] 刘凤麟,李方伟,王明月. 智能反射面辅助广义空间调制性能分析[J]. 电讯技术,2023,63(9):1383-1390.
- [2] BASAR E. Reconfigurable intelligent surface-based index modulation;a new beyond MIMO paradigm for 6G [J]. IEEE Transactions on Communications, 2020, 68(5):3187-3196.
- [3] WU Q Q, ZHANG R. Towards smart and reconfigurable environment; intelligent reflecting surface aided wireless network [J]. IEEE Communications Magazine, 2020, 58(1):106-112.
- [4] WEI X H, SHEN D C, DAI L L. Channel estimation for RIS assisted wireless communications: Part I: fundamentals, solutions, and future opportunities [J]. IEEE Communications Letters, 2021, 25(5):1398-1402.
- [5] HAN Y T, ZHANG S W, DUAN L J, et al. Cooperative double-IRS aided communication; beamforming design and power scaling [J]. IEEE Wireless Communications Letters, 2020, 9(8):1206-1210.
- [6] 张美晨,李双志,陆晶晶,等. 基于黎曼距离的大规模单入多出系统非相干星座映射方案设计[J]. 信号处理, 2021, 37(3):374-382.
- [7] 张美晨,李双志,段雪飞,等. 大规模 MIMO 系统多用户 PSK 调制方案的优化[J]. 电讯技术, 2022, 62(2):238-243.
- [8] LIU H, YUAN X J, ZHANG Y A. Matrix-calibration-based cascaded channel estimation for reconfigurable intelligent surface assisted multiuser MIMO [J]. IEEE Journal on Selected Areas in Communications, 2020, 38(11):2621-2636.
- [9] SWINDLEHURST A L, ZHOU G, LIU R, et al. Channel estimation with reconfigurable intelligent surfaces—a general framework [J]. Proceedings of the IEEE, 2022, 110(9):1312-1338.
- [10] DE ARAÚJO G T, DE ALMEIDA A L F. PARAFAC-based channel estimation for intelligent reflective surface assisted MIMO system [C]//2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop. Hangzhou:IEEE, 2020:1-5.
- [11] WEI L, HUANG C W, ALEXANDROPOULOS G C, et al. Parallel factor decomposition channel estimation in RIS-assisted multi-user MISO communication [C]//2020 IEEE 11th Sensor Array and Multichannel Signal Processing Workshop. Hangzhou:IEEE, 2020:1-5.
- [12] 刘越,李双志,穆晓敏. MIMO 系统中应用张量分解进行半盲信道估计的算法分析[J]. 信号处理, 2016, 32(1):63-69.
- [13] HAN Y T, ZHANG S W, DUAN L J, et al. Double-IRS aided MIMO communication under LoS channels; capacity maximization and scaling [J]. IEEE Transactions on Communications, 2022, 70(4):2820-2837.
- [14] YOU C S, ZHENG B X, ZHANG R. Wireless communication via double IRS; channel estimation and passive beamforming designs [J]. IEEE Wireless Communications Letters, 2021, 10(2):431-435.
- [15] ZHENG B X, YOU C S, ZHANG R. Uplink channel estimation for double-IRS assisted multi-user MIMO [C]//IEEE International Conference on Communications. Montreal:IEEE, 2021:1-6.
- [16] YANG S J, LYU W T, XIU Y, et al. Active 3D double-IRS-aided multi-user communications; two-timescale-based separate channel estimation via Bayesian learning [J]. IEEE Transactions on Communications, 2023, 71(6):3605-3620.
- [17] XIMENES L R, FAVIER G, DE ALMEIDA A L F, et al. PARAFAC-PARATUCK semi-blind receivers for two-hop cooperative MIMO relay systems [J]. IEEE Transactions on Signal Processing, 2014, 62(14):3604-3615.
- [18] STEGEMAN A, SIDIROPOULOS N D. On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition [J]. Linear Algebra and Its Applications, 2007, 420(2/3):540-552.

作者简介:

李双志 男,1990 年生于河南南阳,2018 年获博士学位,现为副教授,主要研究方向为无线通信信号处理。

邢益博 男,2000 年生于河南南阳,硕士研究生,主要研究方向为智能超表面通信和信道估计。

雷豪杰 男,2000 年生于河南周口,硕士研究生,主要研究方向为智能超表面通信和信道估计。

DOI:10.20079/j.issn.1001-893x.231219001

一种解决移动平台下卫星接收遮挡的双天线去重拼接方法^{*}

王英杰, 曾富华, 王钧慧, 陈文渊

(西南电子技术研究所, 成都 610036)

摘 要:在移动平台上(如船舶、汽车等)进行卫星通信接收的场景与固定位置平台有极大不同,需要考虑到移动平台的接收位置空间资源紧张,并且会随着平台位置的移动变化遮挡接收天线的接收效果。采用双天线接收的方式进行卫星通信可极大解决单天线遮挡问题,但会导致数据重复冗余。为解决上述问题,设计了一种基于键值对(Key, Value)结构的双天线数据去重拼接算法。根据卫星数据中的数据帧特点,以数据帧特定字段为关键字设计了数据结构。以键值对数据结构为基础对输入的双路数据进行处理,实现了数据精确去重及高效拼接输出,同时考虑了数据的实时性和可靠性。双天线去重拼接方法为移动平台中卫星持续接收数据接提供了一个新的思路,并且已经在工程中实际应用,接收效果良好。在遮挡天线 1/3 的实验情况下,采用双天线去重拼接方法的数据接收率相比于传统单天线,效果提升 15.69%。

关键词:卫星通信;移动平台;接收遮挡;双天线去重拼接

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN927;TP399 文献标志码:A 文章编号:1001-893X(2025)03-0415-07

A Dual-antenna De-duplication and Montage Method for Solving Satellite Reception Occlusion on Mobile Platforms

WANG Yingjie, ZENG Fuhua, WANG Junhui, CHEN Wenyan

(Southwest China Institute of Electronic Technology, Chengdu 610036, China)

Abstract: The scenario of satellite reception on mobile platforms such as ships, cars, etc. is greatly different from that of fixed position platforms. It is necessary to consider that the space resources of the mobile platform are tight, and the signal reception effect may be blocked with the movement of the platform. To solve these problems, a dual-antenna receiving method for satellite communication is proposed. In order to solve the redundant problem of dual-antenna data receiving repeatedly, a data de-duplication algorithm based on key-value pair structure is designed. According to the characteristics of satellite data frames, specific data fields are used as keys to process the input dual-channel data, realizing data de-duplication and efficient splicing output, and also considering the real-time and reliability of the data. The method provides a new idea for satellite to receive data continuously in mobile platform, and has been applied in engineering with good effect. In the experimental situation of blocking 1/3 of the antenna, the data reception rate using the de-duplication and montage method has improved by 15.69% compared with that of the traditional single antenna.

Key words: satellite communication; mobile platform; reception obstruction; dual-antenna de-duplication and montage

^{*} 收稿日期:2023-12-19;修回日期:2024-04-08
通信作者:王英杰 Email:wangyj_cetc@163.com

0 引 言

卫星通信具有传输距离远、覆盖范围广、不受地理环境限制等优点,在海上船舶等移动平台应用环境中,其作用更是不可或缺。卫星通信天线主要用于同步卫星信号的接收与发射,其主面辐射区域不得有遮挡,遮挡物会严重影响天线系统的增益^[1]。以海上移动平台为例,对于低轨卫星数据考虑多颗卫星从不同方位同时过境接收,为了保障接收无死角,需覆盖天线 360° 接收范围。一般海上卫星通信的接收设备都安装在船舶的甲板面的高处位置,是卫星通信设备安装时避免遮挡的首选位置。但是船舶甲板上安装空间有限,天线受到各种设备遮挡难以达到单天线 360° 全空域覆盖。如何在不影响其他船载设备的情况下,让船舶上的卫星接收天线尽可能不断接收卫星数据成为了一个难点。

文献[2]通过针对船舶某一位置点利用特定卫星通信时计算其遮挡航向来人为地规避因遮挡造成通信不畅问题并形成固定算法为相关专业人员提供帮助。文献[3]研究了障碍物边缘处的最小仰角,结合海上移动通信特点,在考虑恶劣海况带来的船舶大幅倾斜摇摆晃动的情况下,提出遮挡区域算法,并基于以上研究得出遮挡预报模型用于辅助工作人员来执行操作船体姿态变化规避遮挡。文献[4]对舰载雷达侦察装备天线被遮挡情况进行分析,然后提出了码头系泊条件下静态测试和海上动态验证相结合的试验方法,对舰载雷达侦察装备遮挡方位进行分析,可为舰艇雷达侦察装备使用及后续舰艇平台设计与装备安装架设提供参考。文献[5]综合介绍了多天线技术在空间信息传输中的应用情况及进展,展望了对应的测控技术。传统方式解决天线遮挡接收断续问题一般采用提前测算环境获取良好的天线安装位置或者根据辅助信息调整移动平台姿态等方式。良好的天线安装位置往往受限于平台空间和周边环境而有较大的局限性,而调整移动平台姿态具有操作困难繁琐、对人员和辅助信息要求高并且实时性差的缺点。

采用双天线接收方式可以弥补单天线接收区域受限问题,但会引入数据冗余重复问题。文献[6]提出了基于双数组 Trie 树的云环境下数据安全去重方案,适用于云存储环境下数据安全去重。文献[7]提出一种基于消息锁定式加密改进的块级密文去重与密钥管理的方案,支持文件级与数据块级的

两级去重方案。文献[8]提出 HsDeup 方法同时利用布隆过滤器、哈希表及各种缓存机制充分挖掘数据的时间局部性及空间局部性,提高重复指纹的查找效率及指纹预取的准确性。文献[9]提出了一种适用星地网络负载均衡的算法用于减小卫星通信中的数据负载量。当前文献中数据去重方法主要适用于云存储环境,是上传时对已存盘的数据文件的筛选去重,与双天线接收实时去重拼接数据帧的使用场景有着明显不同,不能直接适用于两路数据帧去重。

本文采用了一种在移动平台(船舶)上安装双接收天线的方案来解决遮挡导致的天线接收卫星数据断续等问题,并在保证了实时性和可靠性的前提下,设计了双路数据去重拼接算法。实验结果表明,本文所提出的方案可解决船舶卫星天线接收断续等问题,极大地提高数据的接收质量,避免卫星数据接收丢失。

1 卫星接收方案

1.1 单天线接收分析

卫星信号采用电磁波的方式进行传输,波根据其波长、功率等属性,绕过或穿透障碍物的能力有所不同。但是对于车、船载等移动平台环境,障碍物主要为金属或其他遮蔽物组成,电磁波难以穿透进行传输。理想的天线部署位置一般位于船舶等移动平台的空旷位置,四周无任何遮挡物阻断信号传播。本文旨在解决信号因遮挡导致数据接收断续问题,简化接收过程为卫星直线传播电磁波信号给接收天线,暂不关注波绕过障碍物的具体细节。

场景 1:垂直卫星接收平台面为 90° 俯仰角。方位角假设为 $0^\circ \sim 360^\circ$,其中 0° 和 360° 重合,均表示正北方向,参见图 1。天线遮挡卫星信号接收状态如图 2 所示。

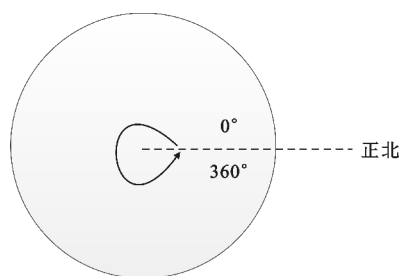


图 1 卫星天线俯视图

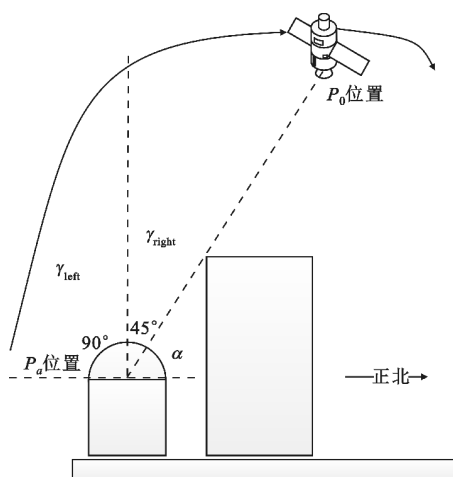


图 2 单天线接收卫星信号遮挡场景 1

卫星飞行轨迹从南往北,左边进入天线接收范围,从右边飞行出去。从图中可以分析发现,当卫星飞行到 P_0 位置的时候,开始受到障碍物的遮挡,卫星后续的广播数据因此无法被接收到。 α 为卫星接收中被遮挡的角度区域。此时卫星天线的实际有效接收范围的角度 $\Delta\gamma$ 为 $\Delta\gamma_{\text{left}} = (180^\circ, 0^\circ) \sim (180^\circ, 90^\circ)$, $\Delta\gamma_{\text{right}} = (0^\circ, 90^\circ - \angle\alpha) = (0^\circ, 90^\circ - 45^\circ) = 45^\circ$, $\Delta\gamma$ 范围为 $(180^\circ, 0^\circ) \sim (0^\circ, 45^\circ)$ 。

场景 2:当卫星飞行轨迹从北往南飞行,此时现象与场景 1 相反。实际有效接收范围的角度 $\Delta\gamma$ 为 $(0^\circ, 45^\circ) \sim (180^\circ, 0^\circ)$ 。

场景 3:当卫星从其他位置飞入接收范围的时候,如果障碍物出现在天线接收卫星信号区间中段,则可能出现卫星信号接收断开,当卫星飞过障碍物的时候通信又恢复正常,参见图 3,出现卫星接收断续情况。此时假设遮挡的接收角度范围为 α ,则接收范围为扣除 α 的角度区域,会根据障碍物的遮挡角度范围分为多段。有效接收范围的角度 $\Delta\gamma$ 为 $(180^\circ, 0^\circ) \sim (180^\circ, \angle\gamma_{\text{left}})$ 和 $(0^\circ, \angle\gamma_{\text{right}}) \sim (0^\circ, 0^\circ)$ 。

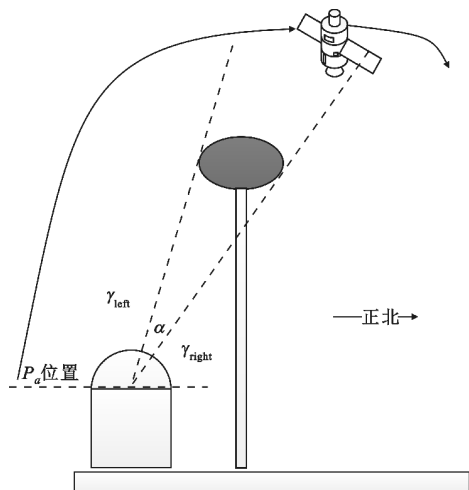


图 3 单天线接收卫星信号遮挡场景 2

1.2 双天线接收分析

1.2.1 双天线接收范围分析

采用双天线接收方案可以通过双天线安装位置互补,解决卫星通信信号被遮挡的问题。双天线安装位置尽量考虑增大与遮挡物之间的距离,避免出现双天线都无法接收到的情况。

当卫星从南向北位置飞入接收范围的时候,只要有一个接收天线能够接收到信号即可进行正常通信,参见图 4。

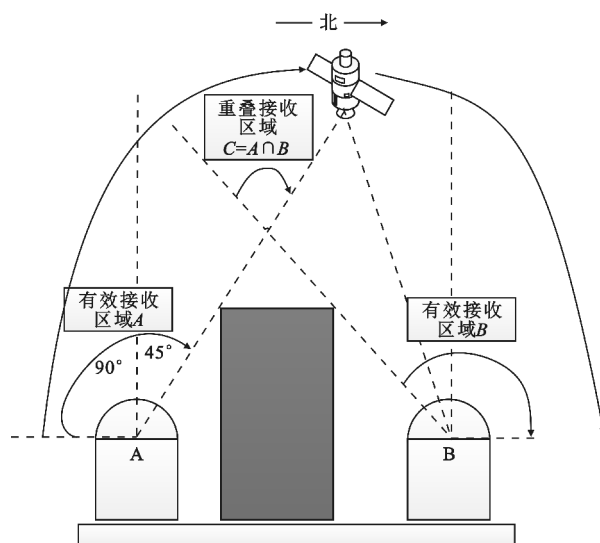


图 4 双天线接收卫星信号场景

本场景下的接收过程:

步骤 1 方位变化为由南向北。 180° 时为正南,当俯仰角跨过 90° 时,对于天线 A、B 单个来说其遮挡角度与单天线一致,此处不再赘述。卫星运动到天线 A 的有效接收角度区域,此时只有天线 A 可以正常接收,天线 B 无法接收卫星信号。

步骤 2 卫星运动到天线 B 能接收到的有效接收区域 B,但是未脱离天线 A 能接收到区域 A,此时天线 A、B 都能接收到卫星信号。重叠接收区域 C 为 A 与 B 的交集: $C=A \cap B$ 。

步骤 3 卫星运动到天线 A 的有效接收角度外,仍然在天线 B 的接收角度内,此时天线 A 无法接收到卫星信号, B 可以接收到卫星信号。

通过上面分析可以得出,当图示情况的卫星从正南向正北方向飞行时候,采用双天线方案卫星接收效果范围可以提升为 $(0^\circ, 180^\circ) \sim (0^\circ, 0^\circ)$ 。

1.2.2 双天线接收数据处理

采用双天线接收数据避免了遮挡引起的接收断

续,极大地提升了卫星接收效果,但是会出现天线接收交叠区域 C。在该区域内接收卫星信号的时候,相同的卫星数据会出现重复接收的情况。对于数据去重拼接,其重点和难点就在于实时性和可靠性,引入双天线数据去重拼接后不能增加过多的处理时延,同时要保障数据连接接收的可靠性。在卫星接收的过程中,由于在轨卫星的种类和数量存在复杂情况,还可能同时出现多颗卫星接收的情况。对于多颗卫星的情况,采用不同的一组天线进行接收。本文目前讨论的是一组天线对同一颗卫星的接收去重拼接情况。

1.3 双天线数据去重拼接算法

本文提出的双天线数据去重拼接算法,针对同一个数据源通过多个接收设备接收的场景进行了针对性的设计。分析卫星 CCSDS 广播帧数据特点发现,同一个数据源中选定数据帧的序列号作为关键标识,分析去重的时候以关键标识字段相同则认为是同一个数据帧。对于两个数据接收设备,我们认为其接收的数据内容一致。采用接收设备 1 和接收设备 2 存在重复数据,也可能因为接收链路的干扰等问题造成部分数据设备 1 接收到了而设备 2 未收到;反之亦然。

对于数据帧内容重复的部分需要根据关键标识进行去重,对于不重复部分需要根据关键标识进行拼接。

通过对数据接收端从收到数据后的第一时间进行去重拼接处理,保证后续数据接收处理系统中整个的数据流中不出现重复数据,并且补充各个数据接收设备的遗漏数据,拼接成为一个最大的数据集,尽量保障接收数据与源头数据一致。采用本算法可极大程度减少数据在后续传输过程中的吞吐量,并且保证数据的可靠性。

1.3.1 去重拼接算法原理

1) 对于输入的第 1、2 路数据用集合 A 、 B 表示,集合中的每一个元素由对应的单帧数据,以及数据帧的关键字组合形成键值对例如 $d1(k1, frame1)$: $A \{ d1(k1, frame1), d3(k3, frame3), d4(k4, frame4), d5(k5, frame5), \dots, dn(kn, frame n) \}$, $B \{ d1(k1, frame1), d2(k2, frame2), d4(k4, frame4), d5(k5, frame5), \dots, dn(kn, frame n) \}$ 。

2) 其中各自的集合中包含各自设备接收的数据帧。需要对集合 A 、 B 的数据进行合并,去重后输出。

3) 使用乒乓操作依次进行 A 、 B 数据集检索。

4) 采集 A 的第一个元素 $d1$ 。

5) 检查 $d1$ 的 $k1$ 值是否已经在输出的结果 C 集合中:如果不在 C 中则添加到 C 中,然后输出 $d1$;如果已经存在于 C 中则表示为重复元素,直接丢弃 $d1$ 元素。

6) 采集 B 的第一个元素 $d1$,重复步骤 5)。

7) 重复执行步骤 3)~6) 的到最终的去重拼接结果集合 C : $C \{ d1(k1, frame1), d1(k2, frame2), d3(k3, frame3), d4(k4, frame4), d5(k5, frame5), \dots, dn(kn, frame n) \}$ 。

图 5 给出了去重拼接的示意。输入的双路数据为输入数据 1 和输入数据 2。 $d1$ 、 $d3$ 元素为输入数据 1 的接收输出,输入数据 2 中 $d1$ 、 $d3$ 元素因为已经在输入数据 1 中输出过了,因此经过判断不再输出。 $d2$ 元素,输入数据 1 中没有接收到,输入数据 2 中进行了输出,后续元素同理。在接收过程中由于遮挡等原因可能出现输入数据 1、输入数据 2 中元素各自独有或都有的情况,通过上面步骤处理,最终在输出数据队列中输出了需要的数据帧结果。

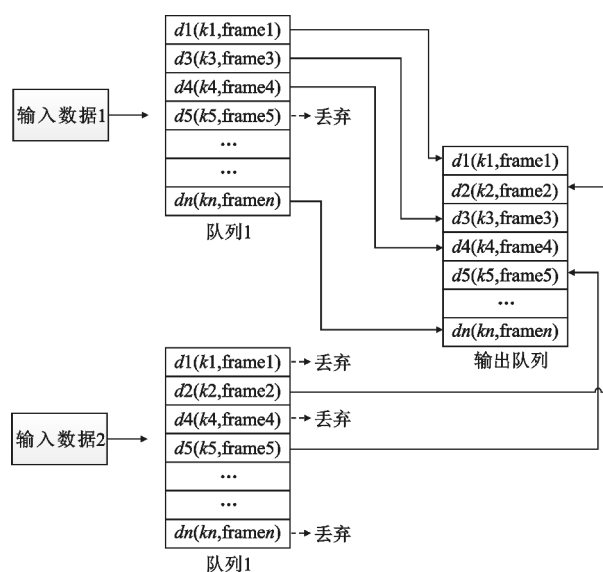


图 5 数据去重拼接算法接示意

1.3.2 卫星数据预处理

本文主要讨论卫星广播数据的去重拼接处理,在进行具体的去重拼接算法前,需要对数据进行预处理操作,找到对应的关键字段信息。

以 Landsat8 遥感卫星为例,首先需要对卫星数据进行天线接收,此过程涉及到数据帧同步、数据帧

格式解析等。通过查阅对应卫星的信息,了解到此卫星下发的广播数据采用了空间数据系统咨询委员会(Consultative Committee for Space Data Systems, CCSDS)标准^[10-12],卫星下发的数据格式为 CCSDS 帧。

卫星下发的广播数据为单帧定时下发,时刻 $T(0)$ 发送第一帧数据,此后依次周期性下发,第 n 次下发的时刻为 $T(n)$ 。卫星接收天线收到对应的卫星信号后进行对应的解调操作,获取到对应的一帧 CCSDS 帧数据,其数据关系如图 6 所示。

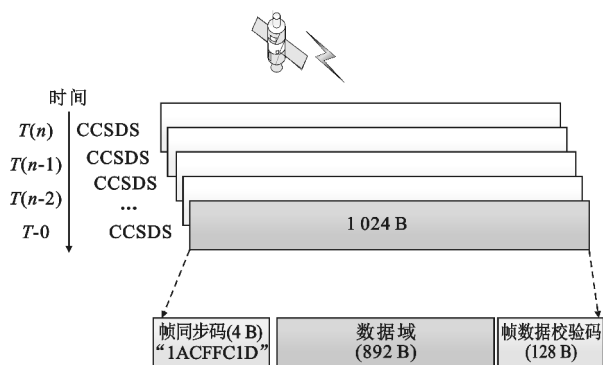


图 6 卫星广播帧数据关系

如图 7 所示,每一帧 CCSDS 帧结构的数据长度为 1 024 B,其中,帧头部分为帧同步码(0x1A, 0xCF, 0xFC, 0x1D),CCSDS 数据域部分为 892 B,结尾部分为 128 B 的校验码负责完成帧内容正确性的校验。

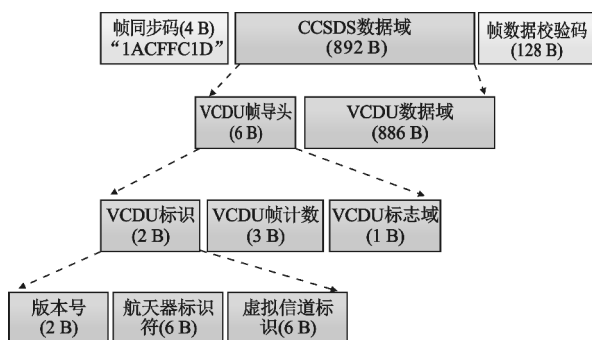


图 7 CCSDS 帧格式

掌握当前的结构还无法取到数据去重算法所需要的关键字信息,因此进行进一步的 CCSDS 帧数据域格式拆解。

VCDU 帧头中的结构详细信息如下:

1) 版本号:第 0 位和第 1 位被用作数据帧的版本号,设为 01。

2) 航天器标识符:第 2~9 位用作航天器标识,

Landsat8 是 11111010。

3) 虚拟信道标识(Virtual Channel Identifier, VCID) 第 10~15 位为虚拟信道标识,用来标明数据产生的通道。

4) 虚拟信道帧计数:第 16~39 位用作虚拟信道帧计数,计数从 0 开始,每个数据帧计数加 1,直到计数值达到 16 777 215 后重置为 0。设备重启时计数也重置为 0。

5) VCDU 标志域中含有是否启动虚拟信道帧计数采用循环计数等方式。

通过按帧结构信息进行层层解析,可以获取到对应的 VCDU 标识和 VCDU 帧计数信息。

本文目前采用 VCDU 帧计数作为去重判断的信息,设计 frameNumList (frameNum) 队列用于保存 VCDU 帧计数。

考虑到 VCDU 标志和 VCDU 帧计数编号是对应关系,不同的 VCDU 标识可能有相同的 VCDU 帧计数,因此关键字信息还需要将 VCDU 标志与 VCDU 帧计数进行组合。为后续去重算法实现,设计 map 字典,Key = VCDU 标志(key), Value = frameNumList (frameNum)。

1.3.3 去重拼接算法实现

如图 8 所示,本算法的实现流程如下:预设一个红黑树字典(Map)的数据结构,用于进行保存数据去重的属性信息。如果本帧数据已经使用过,则记录在 Map[VCDU[key],frameNumList[frameNum]]中。

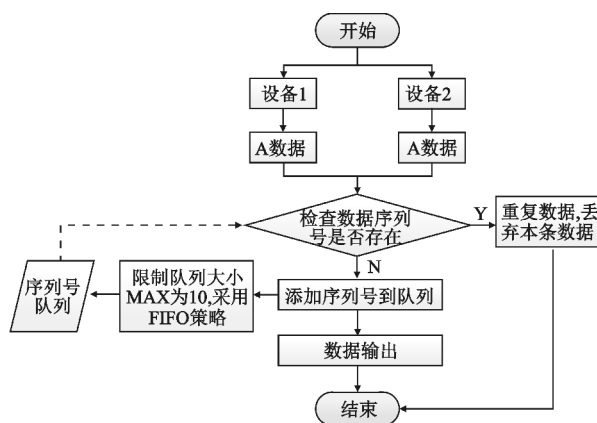


图 8 数据去重拼接流程

具体过程如下:

1) 数据预处理:将接收设备获取的数据进行预处理,形成对应的数据帧。提取帧内容中的关键字信息。对于本次的测控数据信息,关键字为 3 B 的

自增顺序数。

2)生成键值对:由接收数据驱动,插入数据预处理后的数据到队列 1 或者队列 2。

3)由循环处理线程检查队列 1 和队列 2,如果有内容,则抽取对应的一条数据元素。

4)抽取元素动作采用乒乓操作进行:首次抽取元素从队列 1 中开始,以后每一次抽取元素,都依次对两个队列进行,即上一次抽取队列 1 元素,则本次如果队列 2 中有元素就抽取队列 2 的元素。

5)数据帧去重拼接。

6)检查本次的元素($k[i]$, $frameNumList[i]$)是否存在于 Map 中:如果不存在,则输出本帧数据到输出队列中,并进行添加序列号到队列操作($k[i]$, $frameNumList[i]$)到 Map 中;如果存在,则直接丢弃本帧数据,不进行输出。

经过反复执行前述步骤,最后得到的输出队列就是去重拼接后的数据。

2 实验与结果分析

2.1 实验环境

去重数据处理计算机: Intel Core i7 处理器; 8 GB 内存和 1 TB 磁盘空间;千兆网卡×2。

其他设备:接收天线×2;发射信号模拟源;遮挡物体;信号解调处理设备。

网络拓扑:如图 9 所示,卫星接收天线通过信号线连接解调处理设备,解调处理设备通过 CAT6 网线连接到去重数据处理计算机 2 张千兆网卡上。信号发射模拟源模拟发送卫星广播信号。

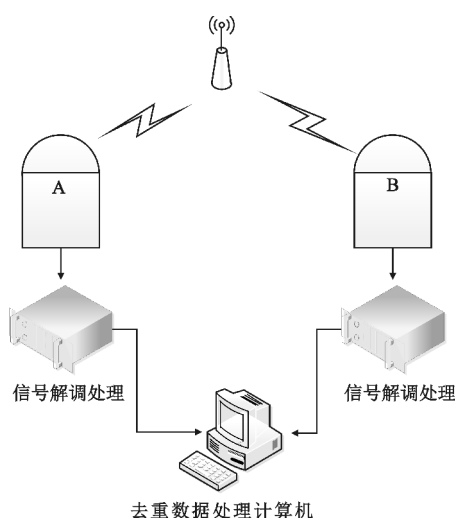


图 9 实验设备网络拓扑

2.2 实验过程

采用发射信号模拟源模拟卫星广播数据,两套信号接收天线用于接收发射信号模拟源发出的信号。双天线数据去重模块所在的计算机通过网络接收天线设备的解调数据,然后进行数据去重拼接。验证去重拼接后的效果步骤如下:

步骤 1 给实验设备通电,检查设备的连通性。

步骤 2 设置信号模拟源调制参数与接收天线的解调参数一致,确保可以完成信号解调。

步骤 3 在去重数据处理计算机上接收解调后的 CCSDS 模拟数据帧。

步骤 4 完成双天线数据去重拼接。通过源数据和接收数据进行比对,验证接收效果。

步骤 5 只开启 A 的情况下进行步骤 1~4,检查接收数据情况。在天线 A、B 与模拟信号源之间设置遮挡物同时开启 A、B 天线的情况下,检查接收数据及数据去重拼接情况。

步骤 6 更改信号模拟源的发送信号信息速率,检查不同速率下的去重数据拼接效果。

2.3 实验结果

去重拼接模块的处理时延表示为 $\Delta T = T1 - T0$, $T1$ 为模块输出数据帧的时延, $T0$ 为模块输入数据帧的时间。此两个时间在数据帧通过模块时标记上时间戳,通过帧序号来确认为同一帧数据,相减后获得去重拼接方法的数据处理时延。

实验结果表明,该算法在解决卫星接收断续问题上可以取得良好的效果,单天线和双天线接收方式下的处理时延均小于 1 ms。更多数据请用微信扫码本文 OSID 码,在“本文开放的科学数据与内容”中查看。

从图 10 可以看出,在所列出 9 种不同条件下,采用双天线接收去重拼接方法相对于传统单天线接收而言,在有天线遮挡的情况下,数据的实际接收率大幅提升。在实验环境中对单天线的接收区间遮挡 1/3 的情况下,单天线的接收率约为 68%。采用双天线的情况下只遮挡单天线,接收率可提升到 100%。对于双天线都被遮挡的情况下,需要区分遮挡的具体情况:同一时间下只要天线 A、B 没有全部被遮挡,则可以保障接收率 100%;当天线 A、B 同时被遮挡的情况下,无法有效接收数据。综合实验中,采用本方法双天线数据接收率相比于传统单天线在都能接收数据的 7 个场景下,接收数据帧效果提升 15.69%。

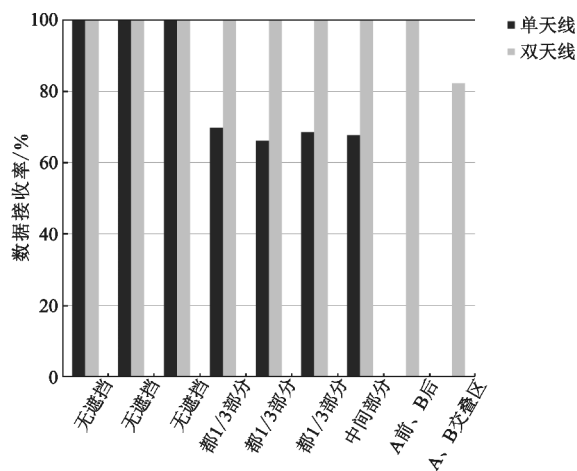


图 10 实验结果统计

如果采用两个天线独立接收,两组数据存盘后再进行事后分析去重,数据实时性上都是数分钟起步,无法保障实时性。

因此,采用双天线去重拼接接收方法可大幅度避免遮挡导致的卫星数据的接收效果差的情况。

3 结束语

本文提出了一种采用双天线接收的方法来解决移动平台卫星通信天线遮挡导致的接收断续问题,设计了双天线接收卫星数据下的去重拼接算法并进行了实验验证。实验结果表明,该方案和算法有效解决了海上卫星通信出现遮挡导致接收断续的问题,可以为海上卫星通信天线持续接收提供有效的解决方案。

虽然本文所提出的双天线接收及数据处理算法在保障连续接收卫星数据和去重、拼接方面优于常规的调整移动平台方向姿态等方法,但仍存在一些问题和需改进之处:只适用于双天线及数据输入,对三天线、四天线等多天线接收情况未进一步研究;在实现过程中只针对卫星 CCSDS 格式数据帧格式,设计适合的去重规则和策略,需要根据数据不重复的序号来进行,具有一定的局限性。因此,后续将增加天线数据,调整算法细节,提高算法的通用性和适用范围;还可以考虑信号接收门限电平来进一步提高双天线的接收的数据质量;目前已在海上移动平台

进行工程实践,对于陆地移动平台(例如车载移动平台)的不同特性也应进行拓展研究。

参考文献:

- [1] 王志波. 卫星通信天线遮挡分析[J]. 通信技术, 2020,53(7):1803-1805.
- [2] 宋汝平,蒋百灵. 船舶卫星通信遮挡航向的计算[J]. 舰船电子工程,2015,35(4):69-71.
- [3] 张海勇,刘冠邑,马迁. 海上卫星通信遮挡预报模型研究[J]. 舰船科学技术,2019,41(11):160-164.
- [4] 韩伟. 舰载雷达侦察装备遮挡范围试验方法研究[J]. 舰船电子工程,2020,40(1):173-174.
- [5] 刘燕都,王元钦,焦义文,等. 空间信息传输中的多天线技术综述[J]. 电讯技术,2020,60(3):350-357.
- [6] 吕世涛. 基于双数组 Trie 树的云数据安全去重技术的研究[D]. 广州:广东工业大学,2021.
- [7] 熊思纯. 云计算中基于客户端的存储去重与图形密码技术研究[D]. 西安:西安电子科技大学,2017.
- [8] 张攀峰. 数据去重中重复数据检测技术研究[D]. 武汉:华中科技大学,2017.
- [9] 郑健阳,李晖,周又玲,等. 一种适用于星地网络的流量负载均衡算法[J]. 电讯技术,2022,62(8):1085-1092.
- [10] HERNÁNDEZ-CABRONERO M, EVANS D, BARTIRINA-RAPESTA J, et al. Resiliency and efficiency of the CCSDS 124. 0 - B - 1 telemetry compression standard[J]. IEEE Access,2024,12:36702-36711.
- [11] TOMJA B. Enhancing versatility and efficiency of the CCSDS - 123 compression algorithm through dynamic partial reconfiguration for the HYPSON mission [D]. Norwegian: Norwegian University of Science and Technology,2023.
- [12] CHOATE M, RENGARAJAN R, STOREY J, et al. Landsat 9 geometric commissioning calibration updates and system performance assessment [J]. Remote Sensing,2011,15(14):1-26.

作者简介:

王英杰 男,1988 年生于重庆,2012 年获硕士学位,现为高级工程师,主要研究方向为测控通信系统。

曾富华 男,1983 年生于重庆,2008 年获硕士学位,现为研究员,主要从事航天测控通信系统总体技术方面的研究。

王钧慧 男,1988 年生于山西吕梁,2010 年获硕士学位,现为高级工程师,主要研究方向为飞行器测控通信。

陈文渊 男,1985 年生于河北张家口,2011 年获硕士学位,现为高级工程师,主要研究方向为飞行器测控通信。

DOI:10.20079/j.issn.1001-893x.231017003

双频段高冗余度列车多普勒测速雷达设计与实现^{*}

彭泽胄¹, 郜洪民², 李 博²

(1. 中国铁道科学研究院研究生部, 北京 100081; 2. 中国铁道科学研究院集团有限公司通信信号研究所, 北京 100081)

摘 要:为满足轨道交通高精度、宽速域、高可靠的列车测速需求,解决测速雷达速度突变问题,提出并实现了一种 24 GHz+77 GHz 双频段高冗余列车多普勒测速雷达方案。该雷达具有冗余性高、抗扰能力强、测速范围宽,以及适用交通制式广等技术优势。完成了测速雷达双频测速子系统的冗余设计,通过主备双系 DSP 芯片分别进行数字信号处理,利用嵌入式微机电系统加速度计实现通过加速度进行辅助停稳判断及速度修正的功能,并实现双系逻辑冗余功能软件设计。通过室内雷达目标模拟器及外场测试,证明该测速雷达能够实现 0~650 km/h 速度范围内的列车速度精确识别,在列车低速运行情况下测速雷达抗扰能力强,能有效解决目前列车低速状态速度突变问题。对比其他方法,在低速状态下,测速误差降低了 57.14%;在列车速度高于 50 km/h 时,测速误差降低了 30%。

关键词:列车测速;多普勒雷达;双频雷达;架构设计;微机电系统(MEMS)

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN959 文献标志码:A 文章编号:1001-893X(2025)03-0422-07

Design and Implementation of a Dual-frequency High Redundancy Train Doppler Speed Measurement Radar

PENG Zezhou¹, GAO Hongmin², LI Bo²

(1. Postgraduate Department, China Academy of Railway Sciences, Beijing 100081, China;

2. Signal and Communication Research Institute, China Academy of Railway Sciences Corporation Limited, Beijing 100081, China)

Abstract: In order to meet the high accuracy, wide speed area and high reliability train speed measurement requirements of rail transit and solve the problem of speed abrupt change of speed measurement radar, a 24 GHz+77 GHz dual-frequency highly redundant train Doppler speed measurement radar scheme is proposed and implemented. This radar has the technical advantages of high redundancy, high disturbance immunity, wide speed measurement range, and being applicable to a wide range of rail transit systems. The redundancy design of the dual-frequency speed measurement subsystem of the speed measurement radar is completed, the digital signal processing of Doppler signal is carried out by dual-series digital signal processor(DSP), and the embedded micro-electro-mechanical system(MEMS) accelerometer is used to realize the function of auxiliary stopping judgment and speed correction through acceleration, and the software design of dual-series logic redundancy function is realized. Through indoor radar target simulator and field test, it is proved that the speed measurement radar can accurately identify the train speed in 0~650 km/h, and the speed measurement radar has strong anti-interference ability under the low-speed operation of the train, which can effectively solve the problem of sudden change in the speed of the current low-speed state of the train. Compared with that of other methods, the measurement error of the proposed method is reduced by 57.14% at low speed. When the train speed is higher than 50 km/h, the measurement error of this method is reduced by 30%.

Key words: train speed measurement; Doppler radar; dual-frequency radar; architecture design; micro-electro-mechanical system(MEMS)

^{*} 收稿日期:2023-10-17;修回日期:2024-02-01

基金项目:中国铁道科学研究院集团有限公司科研项目(2020YJ043)

通信作者:郜洪民 Email:gaohm@rails.cn

0 引言

我国轨道交通事业迅猛发展,目前在轮轨交通领域,高铁客运时速已经达到 350 km/h;磁浮交通领域,多条中低速磁悬浮线路已建设运营,国家层面规划也明确指出发展时速 600 km 级高速磁浮系统。列车测速的准确性与稳定性对于提高列车运营安全性和运行效率具有重要意义^[1-2]。

近年来,列车测速技术不断发展,已有国内外研究人员做了大量研究。文献[3]详细分析了常用的轮轴速度传感器、多普勒雷达及加速度计的测速原理,并对 3 种测速方法的优缺点进行比较分析。文献[4]对国内外主要的 6 种磁浮测速方法进行综述和比较,指出了目前的测速方法均不能满足磁浮测速需求。文献[5]设计了一种异向双天线雷达,并提出了一种自适应算法校正天线波束与地面夹角,通过测试和数据分析证明了该方法能有效解决角度偏差问题。文献[6]提出了一种多普勒雷达结合轮轴传感器的联邦粒子滤波融合测速方法。文献[7]采用改进 Burg 算法设计了一种高精度的列车多普勒测速雷达。文献[8]设计了一种基于 TMS320C6745 数字信号处理器的铁路测速雷达,通过振动补偿结构消除列车车体运行振动影响,具有精确度高、稳定性好的优点。文献[9]提出了一种基于非线性纵向列车动力学模型的概率加权算法,融合算法利用轮轴传感器、雷达传感器及全球导航卫星系统进行数据融合,通过仿真实验实现了对列车速度的精确测量。文献[10]采用两个侧视调频连续波雷达对列车速度进行估计,利用两个雷达图像峰值出现的时间延迟进行速度估计。文献[11]在山梨县(Yamanashi)磁浮试验线采用了感应回线的测速定位方法,测速定位精度可达 10 cm。文献[12]将 GPS 信息与速度传感器数据进行组合应用,并利用精确数字地图提高列车定位测速精度。文献[13]基于卡尔曼滤波理论,提出了一种将轮轴速度传感器与惯性传感器融合的列车测速算法。基于多传感器融合的列车测速方法具有测速精度高的优点,但存在硬件结构复杂、计算量大且实际应用困难的问题。传统的轮轨列车主要依靠安装在轴端的光电编码器或测速电机进行测速,该方法存在车轮打滑或空转会造成测速精度低、可靠性差的缺点,且轮轴传感器无法运用于磁浮交通系统。目前国内外研究人员提出了多种磁浮测速定位方法,但大多数需基于轨旁设备实现测速功能,普遍存在结构复杂、建设维护成本高、适应性差等问题。

综上所述,目前已有的列车多普勒测速方法并不能解决列车低速状态下测速误差较大的问题,并且均为单频段单系设计方案;在目前轨道交通实际运营过程中,现有的车载测速雷达在低速工况下存在测速跳变的故障,尤其是存在列车停稳状态下受扰测速不为零的问题,现有的方法均不能很好地解决。因此,本文提出了以 24 GHz+77 GHz 双频段、宽速域、应用范围广为技术特征的车载多普勒测速雷达,利用加速度信息优化列车零速停稳判断,并通过目标仿真、现场功能测试验证了该测速雷达具有测速精度高、抗扰能力强、测速范围宽的优点。

1 雷达测速原理

车载多普勒雷达测速原理如图 1 所示。列车运行速度为 v ,在列车底部安装车载多普勒雷达,角度 θ 为雷达发射波与轨道的夹角,频率源相对轨道面的速度 $v_r = v \cdot \cos \theta$,电磁波传播速度 c 为光速,其中多普勒雷达通过发射天线向轨面发射电磁波 f_0 ,同时接收由轨面反射回来的回波 f_r 。当列车运行时,多普勒频率 f_d 是回波频率 f_r 与发射频率 f_0 的差值。对多普勒信号进行 A/D 采样并进行频域谱分析计算精确的多普勒频率值,进而可以计算出列车运行速度。由多普勒频率表达式得到式(1)的列车运行速度计算公式, λ 是发射波波长,可以看出多普勒信号 f_d 与列车速度 v 成正比。

$$v = \frac{c}{2f_0 \cos \theta} \frac{f_d}{2 \cos \theta} = \frac{f_d \lambda}{2 \cos \theta} \quad (1)$$

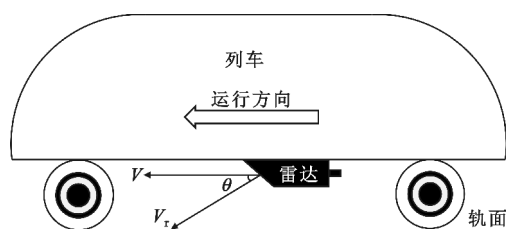


图 1 多普勒雷达测速原理

2 测速雷达硬件设计实现

列车测速雷达硬件结构如图 2 所示,硬件部分采用 24 GHz 及 77 GHz 两套结构相同、平行处理的子测速系统。子测速系统包括电源板、天线、微波模块、信号处理板。微波模块及天线实现中频信号产生及输出。通过信号处理板上的 DSP 实现中频信号频谱分析并完成列车速度求解。测速子系统通过串口通信实现主备冗余及速度融合。测速子系统分

别计算列车运行速度,若测速值异常则利用加速度进行修正。主系通过串口接收从系发送的数据后,对双系测速数据进行合理性判断,根据判断结果输出雷达工作状态及双系速度融合值。

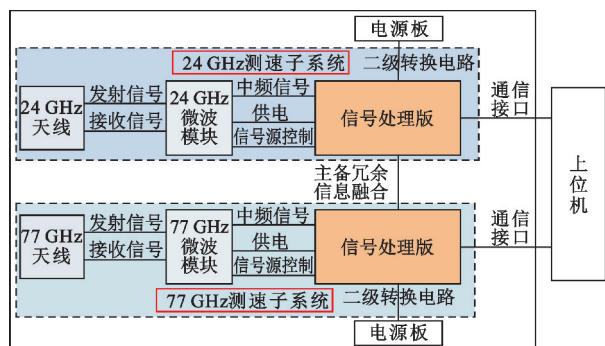


图2 列车测速雷达硬件结构

对于多普勒测速系统应选取合适的信号频段。目前在公路运输领域,24 GHz与77 GHz这两个频段应用最为广泛,在轨道交通领域列车多普勒测速则主要采用24 GHz频段^[14-17]。77 GHz的毫米波信号比24 GHz的速度分辨率和精度提高了3倍^[18],77 GHz信号频率更高波长更小,更容易受大气、雨雾衰减影响^[19-20]。考虑现有列车多普勒测速方法均为单频单系设计的特点,决定采用24 GHz与77 GHz高低双频冗余的结构,既可以有效避免同频段的相互干扰,同时能提高系统测速稳定性和精度。

车载测速雷达高冗余特性主要为硬件冗余及数据冗余。硬件异频冗余采用高低双频测速子系统的冗余设计。当主测速子系统出现子模块故障或失效时,另一个备用测速子系统继续工作,保证测速雷达的正常工作。这种高低频冗余设计可以避免同频共模干扰,提高系统的可靠性。数据冗余测速子系统分别进行速度测量,分别对原始测量值进行阈值校验判断是否进行修正,以确保雷达最终测速结果的准确性和可靠性。

3 测速雷达软件设计实现

3.1 软件总体设计

图3所示为测速系统软件流程图。24 GHz及77 GHz测速子系统工作原理及软件流程是一致的,在本文中将24 GHz测速子系统设置为主系。主、备测速子系统分别对差频信号进行A/D采样,利用数字信号滤波滤除直流分量及高频干扰噪声,采用快速傅里叶变换(Fast Fourier Transform, FFT)算法

对差频信号进行频谱分析计算列车速度,并分别由加速度信息得到修正速度。主系通过串口接收从系发送的从系状态及测速数据后,根据双系速度融合算法输出测速系统工作状态及双系融合速度值。当主系进入故障模式或从系未能在5个周期内接收到主系的反馈数据时,如图3虚线路径所示,从系通过RS485串口输出从系测速结果,实现双系冗余功能提高系统测速可靠性。

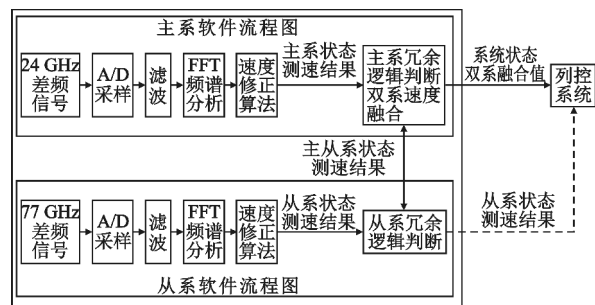


图3 测速系统软件流程

测速误差要求速度分辨率为0.5 km/h,因此雷达发射频率为24 GHz和77 GHz频率分辨率 Δf_c 分别为15.71 Hz和50.41 Hz。列车运行时速为650 km/h的工况下时,24 GHz和77 GHz对应的多普勒频率为20 427.53 Hz和65 538.34 Hz,根据采样定理和速度分辨率要求,故24 GHz和77 GHz应选取的雷达采样频率分别为50 kHz和150 kHz,采样点数为4 096点。

3.2 速度修正算法

速度修正算法的主要功能是利用加速度信息对子系统测量初始值进行速度修正,有效解决速度突变问题保证了测速平稳性。速度修正算法如图4所示。首先通过加速度进行辅助停稳判断,若判断列车处于停稳状态,则子系统测速值修正为零,避免列车在停稳状态下输出速度不为零的故障情况出现;列车在运行状态下,若子系统原始测速值 V 在阈值范围内波动,则认为正常状态不进入修正模式,直接输出原始测速值 V 。列车在运行状态下,若子系统测速值 V 超过阈值范围,则进入修正模式,利用加速度 a 进行速度修正;若修正后的速度 V_1 超过阈值则输出报错进入自检。修正速度值计算如公式(2)所示:

$$V_1 = a \cdot \Delta t + V_0 \quad (2)$$

式中: a 为列车运行加速度; Δt 为单位时间; V_0 为上一周期的子系统测速值; V_1 为修正后的测速值。速度阈值范围 ΔV_m 由列车最大制动减速度绝对值 a_m

确定, k 为调节系数, 如式(3)所示:

$$\Delta V_m = k \cdot \Delta t \cdot a_m \quad (3)$$

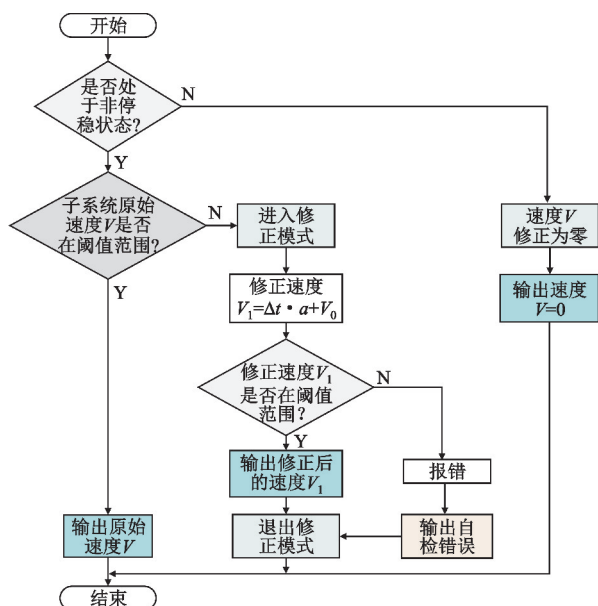


图4 速度修正算法

3.3 双系速度融合算法

双系速度融合算法如图5所示,主系通过串口接收从系发送的数据后,对双系数据进行逻辑判断,根据判断结果输出测速系统工作状态及双系速度融合值。

车载列车自动保护系统 (Automatic Train Protection, ATP) 是列车运行控制的重要车载设备,当列车测速值超过列车运行允许速度时,ATP 发出制动指令进行减速,考虑轨道交通故障安全原则,本文采用最大值选择法进行双系速度融合。因此,双系速度融合可分为以下4种情况:若两个频率的子系统原始速度都未经过修正,则输出最终速度为两者速度最大者;若两个频率的原始速度都用加速度信息进行了修正,则输出最终速度为两者速度最大者;若只有一个频率的原始速度进行了修正,则输出最终速度取非惯导修正速度值;若两个频率的原始速度都用加速度信息进行了修正且存在速度超过阈值范围,则输出速度取信号质量最佳者。

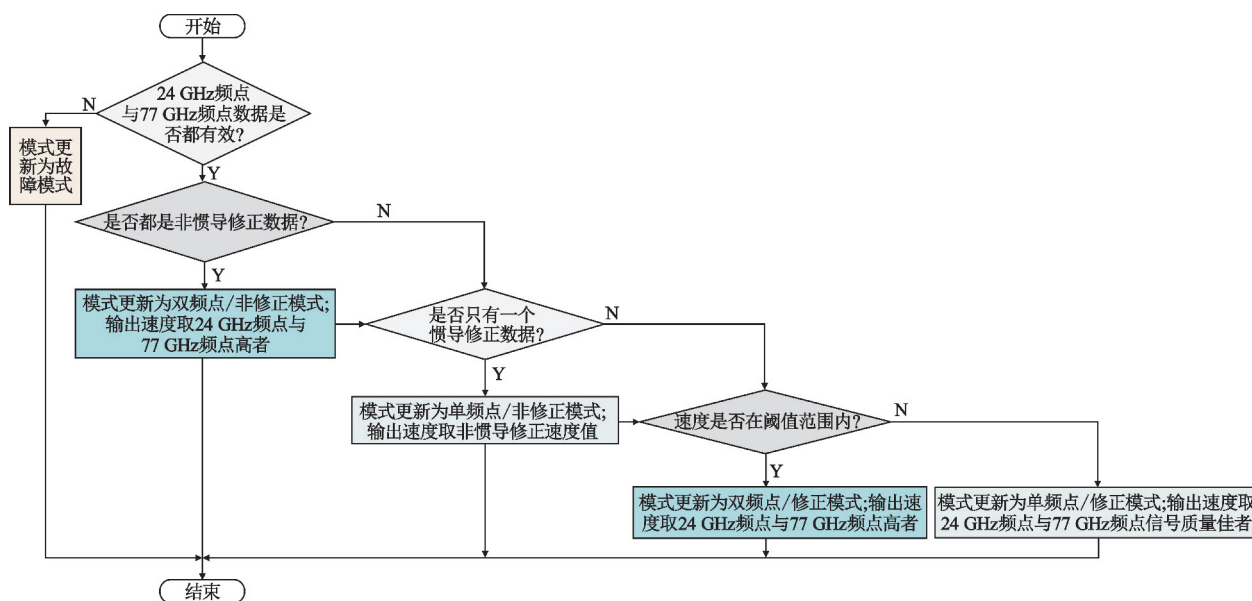


图5 双系速度融合算法

4 测试验证

4.1 模拟测试

利用雷达目标模拟器发生信号回波对列车测速雷达进行性能测试。由于雷达目标模拟器频带限制,故分别对 24 GHz 及 77 GHz 频段测速子系统进行实验室内测速性能测试。24 GHz 及 77 GHz 测速

子系统模拟测速结果如表1所示,在列车速度小于 50 km/h 时,子系统测速均误差小于 0.3 km/h;列车速度大于 50 km/h 时,子系统测速均误差小于 0.8%。对比可以发现,77 GHz 测速系统测速更精确。24 GHz 和 77 GHz 频段测速性能良好,测速量程范围为 0.5~650 km/h,满足未来高速磁浮列车的测速需求。

表 1 模拟测速结果

模拟速度/ (km/h)	测速结果/(km/h)		测速误差/%	
	77 GHz 子系统	24 GHz 子系统	77 GHz 子系统	24 GHz 子系统
650	653.90	646.88	0.60	0.48
600	603.62	597.11	0.60	0.48
500	502.79	497.59	0.56	0.48
400	402.41	398.07	0.60	0.48
300	301.81	298.92	0.60	0.36
200	201.2	199.04	0.60	0.48
150	151.1	149.28	0.73	0.48
100	100.6	99.52	0.60	0.48
90	90.5	89.71	0.56	0.32
80	80.4	79.9	0.50	0.12
70	70.3	69.73	0.43	0.39
60	60.21	59.93	0.35	0.12
50	50.11	50.12	0.11	0.12
40	40.2	39.95	0.20	0.05
30	30.3	29.78	0.30	0.22
20	20.2	19.98	0.20	0.02
10	9.71	9.81	0.29	0.19
5	5.05	5.08	0.05	0.08
1	0.78	1.09	0.22	0.09
0.5	0.38	0.48	0.12	0.02

将本文方法与文献[5-8]的测速性能指标进行对比。文献[5-8]中方法分别用方法 1、方法 2、方法 3、方法 4 表示。如表 2 和表 3 所示,在低速运行状态下,本文方法相较于其他 3 种方法测速精度更高,测速误差最少降低了 57.14%。在列车速度高于 50 km/h 时,本文方法相较于方法 3 和方法 4 更加精确,测速误差最少降低了 30%。通过对比,本文方法测速精度更高,并且能够有效解决速度突变问题。

表 2 低速运行状态下测速性能指标对比 ($V<50\text{ km/h}$)

方法	测速误差/(km/h)	是否解决速度 突变问题
本文方法	0.30	是
方法 1 ^[5]	1.00	否
方法 2 ^[6]	1.37	否
方法 3 ^[7]	0.70	否

表 3 高速运行状态下测速性能指标对比 ($V>50\text{ km/h}$)

方法	测速误差/(km/h)	是否解决速度 突变问题
本文方法	0.7	是
方法 3 ^[7]	1.7	否
方法 4 ^[8]	1.0	否

4.2 现场测试

4.2.1 低速状态功能验证测试

低速运行状态雷达测速结果如图 6 所示。在运行速度小于 5 km/h 超低速运行状态下,雷达测速结果为红色曲线双系融合速度值,测速误差约为 0.3 km/h,符合测速精度要求。雷达单系故障状态测速结果如图 7 所示。在一系故障状态下,另一系测速子系统能正常工作实现双系冗余功能,测速效果良好测速误差约为 0.3 km/h,符合测速精度要求。列车静止状态施加外部干扰雷达测速结果如图 8 所示。在列车静止状态下,对雷达天线施加外部干扰,由图可知测速子系统均出现速度突变异常,但加速度基本都为零,即判定列车处于停稳状态,将速度修正输出为零。低速运行状态施加外部干扰雷达测速结果如图 9 所示。在低速运行状态下施加外部干扰,测速子系统均出现速度突变异常,进入修正模式利用加速度信息进行速度修正,双系速度融合输出测速曲线光滑完整,符合测速误差要求。

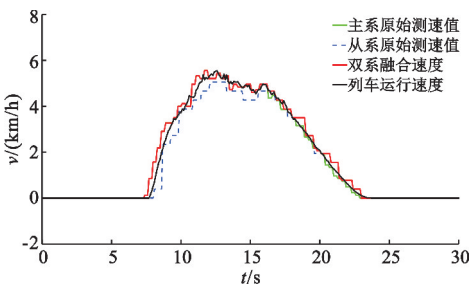


图 6 低速运行状态测速结果

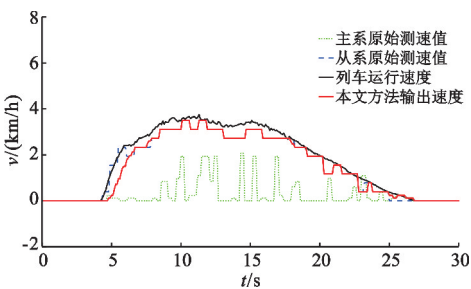


图 7 雷达单系故障状态测速结果

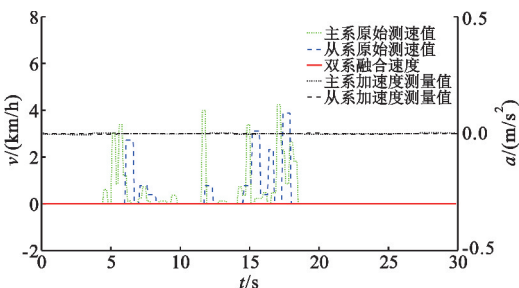


图 8 列车静止状态施加外部干扰雷达测速结果

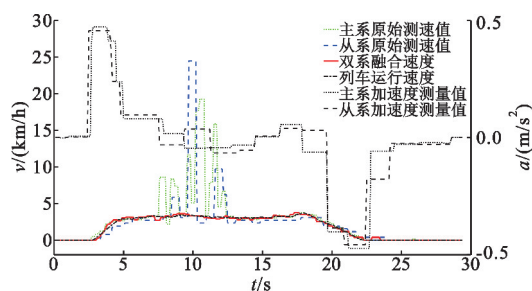


图 9 低速运行状态施加外部干扰雷达测速结果

4.2.2 现场测速性能验证测试

如图 10 所示,在国家铁道试验中心城轨试验线上开展试验,列车以 80 km/h 运行时雷达测速结果如图 11 所示,在平稳运行状态下,列车速度无突变,不启用加速度修正功能,选取输出双系速度最大值。可以得出以下结论:雷达测速效果较好,在列车约以 80 km/h 速度运行时,雷达测速值能与列车实际运行速度较好拟合,本文方法测速平均绝对误差 (Mean Absolute Error, MAE) 为 0.53 km/h,平均绝对百分比误差 (Mean Absolute Percentage Error, MAPE) 为 0.70%。雷达测量速度存在一定的测速误差,分析原因可能是雷达固定支架振动或列车固有运行振动频率导致的测量结果误差,后续需要进一步研究信号滤波去噪算法,从算法角度消除列车振动引起的测速误差值。



图 10 现场测试示意

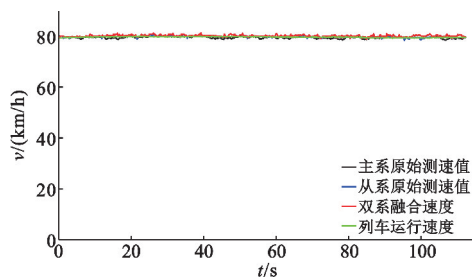


图 11 列车 80 km/h 运行速度时雷达测速结果

5 结束语

本文在轨道交通领域提出并实现了双频段高冗余列车多普勒测速雷达方案。基于 24 GHz 和 77 GHz 双系冗余测速模式,提高了列车测速可靠性和精度,以及测速雷达的环境抗扰能力。利用内嵌式加速度计进行雷达测速修正,增加了车载测速雷达的环境感知能力,提高了系统安全性和可靠性。通过对比测试验证了该测速雷达能够实现 0 ~ 650 km/h 速度范围强抗扰、高精度的速度测量,满足高速铁路、磁浮列车、市域、城际铁路等不同制式轨道交通应用需求。

为提高雷达测速精度,在后续工作中需要进一步研究列车振动对车载雷达测速误差的影响。

参考文献:

- [1] 张森,杨志杰,莫小凡,等. 列控车载设备测速测距同步表计技术研究[J]. 铁道标准设计, 2022, 66(1): 156-160.
- [2] 刘红良. 基于复杂网络理论的高速铁路列控系统设备安全风险研究[J]. 铁道运输与经济, 2022, 44(7): 82-89.
- [3] 宋剑伟. 城市轨道交通信号系统的典型测速定位方案比较[J]. 城市轨道交通研究, 2021, 24(4): 26-29.
- [4] 张世聪. 适用于磁浮列车的测速定位方法研究综述[J]. 铁道标准设计, 2018, 62(10): 186-191.
- [5] 李萌,曹林,王东峰. 用于机车测速的雷达传感器算法研究[J]. 传感器与微系统, 2016, 35(12): 69-71.
- [6] 李泽鑫,张亚东,魏维伟,等. 双雷达与轮轴传感器融合的列车高精度测速方法[J]. 铁道标准设计, 2024, 68(6): 221-228.
- [7] 黄颖. 列车多普勒测速信号处理方法研究与系统设计实现[D]. 北京: 北京交通大学, 2019.
- [8] ZHANG J, LI H W, ZHANG L H, et al. The research of radar speed measurement system based on TMS320C6745 [C]//2012 IEEE 11th International Conference on Signal Processing. Beijing: IEEE, 2012: 1843-1846.
- [9] MUNIANDI G, DEENADAYALAN E. Train distance and speed estimation using multi sensor data fusion[J]. IET Radar, Sonar & Navigation, 2019, 13(4): 664-671.
- [10] REISSLAND T, LENHART B, LICHTBLAU J, et al. Evaluation of a robust correlation-based true-speed-over-ground measurement system employing a FMCW radar[J]. International Journal of Microwave and Wireless Technologies, 2019, 11(7): 686-693.

- [11] MAEDA T, WATANABE K, ONO M. Train position detecting system using radio millimeter-waves [C]//Computers in Railways X: Computer System Design and Operation in the Railway and Other Transmit. Prague: Central Japan Railway Company, 2006: 469-476.
- [12] FILIP A. Efficient use of multi-constellation EGNOS for the European train control system [C]//2016 European Navigation Conference. Helsinki: Nordic Institute of Navigation, 2016: 1-9.
- [13] MALVEZZI M, VETTORI G, ALLOTTA B, et al. A localization algorithm for railway vehicles based on sensor fusion between tachometers and inertial measurement units [J]. Journal of Rail and Rapid Transit, 2014, 228(4): 431-448.
- [14] 丁左武, 徐杰, 周龙, 等. 车载毫米波雷达目标检测综述[J]. 电讯技术, 2024, 64(4): 655-662.
- [15] 黄富传. 基于 DSP 的毫米波雷达信号处理系统设计[J]. 中国集成电路, 2022, 31(11): 43-48.
- [16] 高辉. 77GHz 调频连续波雷达场景感知关键技术研究[D]. 南京: 南京理工大学, 2019.
- [17] 魏豪. 77G 毫米波汽车防撞雷达设计与实现[D]. 西安: 西安电子科技大学, 2022.
- [18] XU T Q, YU D Y, DU L. A bi-objective simulation facility for speed and range calibration of 24 GHz and 77 GHz automotive millimeter-wave radars for environmental perception[J]. Electronics, 2023, 12(13): 1-20.
- [19] DARLIS A R, IBRAHIM N, KUSUMOPUTRO B. Performance analysis of 77 GHz mmWave radar based object behavior[J]. Journal of Communications, 2021, 16(12): 576-582.
- [20] 徐天琪, 杜磊, 白杰, 等. 24 GHz 和 77 GHz 毫米波雷达速度测量不确定度评估比较[J]. 计量科学与技术, 2021, 65(9): 3-7.

作者简介:

彭泽青 男, 1999 年生于湖北十堰, 硕士研究生, 主要从事列车车载测速雷达研发工作。

郜洪民 男, 1974 年生于河北石家庄, 2003 年获硕士学位, 现为研究员, 主要从事铁路通信信号研究工作。

李 博 男, 1985 年生于湖北咸宁, 2009 年获硕士学位, 现为副研究员, 主要从事城市轨道交通列控系统的研发工作。

DOI:10.20079/j.issn.1001-893x.231112001

基于子阵列的大规模 ISAC 系统混合波束赋形设计*

郭鸿儒¹, 郭应鸿¹, 顾忆宵², 夏 斌¹

(1. 上海交通大学 电子信息与电气工程学院, 上海 200240; 2. 上海大学 通信与信息工程学院, 上海 200444)

摘要: 通信感知一体化(Integrated Sensing and Communication, ISAC)系统可以将通信感知功能有机融合,以取得更高的频谱效率和硬件利用率,但传统的大规模集中式天线阵列在平面波假设下无法提供距离维增益,且其混合波束赋形设计为非凸优化问题,仍是具有挑战性的难题。为此,提出了一种基于子阵列的混合波束赋形设计方案,在较低的硬件复杂度下通过扩展球面波区域范围提供距离维增益,以在满足感知性能约束和发射功率预算的前提下最大化通信速率。首先提出了一种基于分式规划和最优化最小化方法的算法,将非凸优化问题转化为凸问题后迭代求解得到一个联合波束赋形矩阵;进而提出一种基于流形优化和最小二乘法的算法,迭代求解后将其分解为数字/模拟波束赋形矩阵。仿真结果表明,基于子阵列的算法相较于集中式阵列能够获得更多的距离维信息和感知自由度,通信性能提升 40%,且流形优化后混合波束赋形方案能够很好地逼近联合优化的数字波束赋形方案的性能。

关键词: 通信感知一体化(ISAC);大规模天线;混合波束赋形;分式规划;流形优化

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN929.5 文献标志码:A 文章编号:1001-893X(2025)03-0429-08

Hybrid Beamforming Design for Subarray-based Integrated Sensing and Communication Systems

GUO Hongru¹, GUO Yinghong¹, GU Yixiao², XIA Bin¹

(1. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;
2. School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China)

Abstract: Integrated sensing and communication (ISAC) systems aim to organically merge communication and sensing functionalities to achieve higher spectral efficiency and hardware utilization. However, traditional large-scale centralized antenna arrays, under the assumption of planar waves, cannot provide range-dimensional gain, and their hybrid beamforming design presents a non-convex optimization problem, which remains a challenging issue. To address this, a hybrid beamforming design scheme based on sub-arrays is proposed. It offers range-dimensional gain by extending the spherical wave region with lower hardware complexity. The objective is to maximize communication rates under the constraints of sensing performance and transmission power budget. Initially, an algorithm based on fractional programming and optimization minimization methods is introduced, which transforms the non-convex optimization problem into a convex one and iteratively solves to obtain a joint beamforming matrix. Subsequently, a manifold optimization and least squares-based algorithm is proposed, which iteratively solves and decomposes the solution into digital/analog beamforming matrices. Simulation results indicate that the sub-array-based algorithm, compared with centralized arrays, can obtain more range-dimensional information and sensing degrees of freedom, improving communication performance by 40%. Moreover, after manifold optimization, the hybrid beamforming scheme closely approximates the performance of the jointly optimized digital beamforming scheme.

Key words: Integrated sensing and communication (ISAC); massive MIMO; hybrid beamforming; fractional programming; manifold optimization

* 收稿日期:2023-11-12;修回日期:2024-07-07
通信作者:夏斌 Email:bxia@sjtu.edu.cn

0 引言

通信感知一体化 (Integrated Sensing and Communication, ISAC) 是指将通信和感知两个功能融合在一起的一种技术。在传统的无线通信系统中,通信和感知被视为两个独立的任务,它们通常使用不同的频段和硬件设备来实现。但是,在 ISAC 中,通信和感知可以通过共享相同的频段和硬件设备来实现,可以得到更高的频谱效率和更好的空间利用率,同时也可以提高通信和感知的性能和可靠性。ISAC 技术的应用领域包括智能交通、工业物联网、智能城市等^[1-2]。在这些领域中,ISAC 可以实现更高效、更智能的通信和感知,从而提高整个系统的性能和效率。

然而,由于采用共波形发送通信和感知信号,ISAC 系统中的发射波束形成设计不仅影响用户的通信速率,还会影响传感定位的精度。现有相关文献从波束形成的设计上,对如何权衡通信和传感的性能进行了研究^[3-4]。进一步地,从单一通信用户和单一目标的简化情况,扩展到具有多个接收器和多个感测目标的情况^[5-7];文献[5]提出了一种基于波束图匹配的方法,以在保证感知性能的基础上,最小化通信用户间的多用户干扰;文献[6]引入矩阵补偿技术来解决感知与通信系统之间的相互干扰问题,实现了联合传输时高效能量利用及准确的目标检测;文献[7]采用逐步优化的方法,在优化多个目标的同时考虑了通信和雷达传感器的限制,使得系统的性能得到了显著提高。

然而,对于上述系统所采用的传统数字波束形成方案,随着天线数量的快速增加,系统的硬件成本和功耗将迅速上升,而混合波束赋形技术可以有效满足大规模天线低能耗的需求。混合波束赋形将信号处理分为数字基带域和模拟射频域,可以在保持较低硬件复杂度的同时实现较高的频谱效率^[8]。全连接 (fully-connected, FC) 结构和子阵列 (Array-of-Subarrays, AoSA) 结构是两种典型的混合波束赋形结构。基于 AoSA 结构,文献[9]提出并分析了在雷达性能与通信性能加权下的双层预编码器设计。文献[10]提出了 GoSA 模型,并进一步讨论了在保证雷达检测性能的前提下,最大化通信系统的传输速率的混合波束成形算法的设计。而且,当天线规模较大时,传统无线通信系统中使用的集中式阵列的性能将受到很大限制,且距离维不同的设计需求无法被满足^[11]。为了解决这一问题,与多径所提供

的路径内增益不同,基于子阵列的系统将信道用平面波和球面波混合建模,可以利用阵列上的波曲率提供额外的距离维自由度,使信道矩阵从低维变为高维,并基于不同子阵列的相对位置,在单一路径上充分利用空间的路径间增益,已被证明有效地提高了系统的通信性能^[12-13]。然而基于子阵列的天线技术对 ISAC 系统设计自由度的提升还鲜有研究。大规模天线通信场景下子阵列系统虽然能够提升系统能效,但是由于通信速率和感知性能的影响相互耦合,给如何有效地利用距离维度额外的信息进行通感联合设计,以在保证感知性能的同时充分提升通信速率带来了挑战。

针对上述文献的不足,本文研究了基于子阵列的 ISAC 系统的联合波束形成设计,其中多天线的基站同时为多个单天线用户服务并检测多个点目标。通过优化发射波束形成算法,在雷达感测波束相似度约束、反射系数约束和总发射功率预算的条件下,最大限度地提高可达速率。为了处理由此形成的非凸优化问题,首先采用分式规划 (Fractional Programming, FP) 方法将其目标函数转换为更易于处理的形式,然后基于最优化最小化 (Majorization Minimization, MM) 的方法迭代求解得到一个联合波束赋形矩阵,再使用基于流形优化 (Manifold Optimization, MO) 的算法迭代求解得到混合波束赋形的两个矩阵。数值结果表明,基于子阵列的系统可以弥补集中式阵列无法区分同一方向上不同距离目标的不足,获得更高的感知自由度,通信性能也有显著提升。同时,基于 MO 算法设计的分步逼近算法可以很好地逼近联合优化的感知性能,两条曲线趋势相似且波峰都几乎重合到理想目标处。所提算法得到的混合波束赋形方案可以降低非感知目标角度上的波束强度,更逼近理想的感知性能,同时联合优化解相较于混合波束赋形方案具有略高的通信性能。

1 系统模型

如图 1 所示,考虑一个下行混合预编码的通感一体化系统,基站 (Base Station, BS) 采用多个子阵列组成的天线。将路径内和路径间空间多路复用模式合并到组合信道模型中,并将其与混合波束赋形架构连接,其中大规模天线阵列被分为 N 个子阵列,每个子阵列包含 P 根天线。

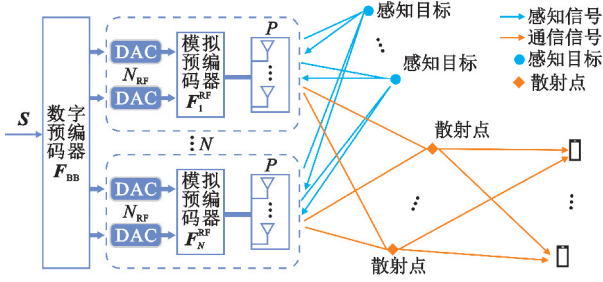


图 1 基于子阵列的大规模天线 ISAC 系统模型

基站端传输 N_s 个数据流, 服务于 U 个单天线通信用户。考虑多用户波束形成的情形, 即基站只通过一个流与每个用户通信, 即 $U=N_s$ 。此外, 基站可以同时服务的最大用户数量等于基站端 RF 链的数量, 即 $U \leq N_{RF}$ 。对于混合波束赋形, 基站端用一个基带预编码器 $\mathbf{F}_{BB} \in \mathbb{C}^{N_{RF} \times N_s}$ 连接一个射频波束形成器 $\mathbf{F}_{RF} \in \mathbb{C}^{NP \times N_{RF}}$ 。基站端的传输信号可以被建模为

$$\mathbf{x} = \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} \quad (1)$$

式中: $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$ 表示符号向量, 满足 $E[\mathbf{s}\mathbf{s}^H] = \frac{P_t}{N_s} \mathbf{I}_{N_s}$, P_t 表示总的传输能量。基于子阵列模型, 基带预编码器和射频波束形成器可以被表示为

$$\mathbf{F}_{BB} = [\mathbf{f}_1^{BB}, \mathbf{f}_2^{BB}, \dots, \mathbf{f}_U^{BB}] \quad (2)$$

$$\mathbf{F}_{RF} = \begin{bmatrix} \mathbf{F}_1^{RF} & 0 & \dots & 0 \\ 0 & \mathbf{F}_2^{RF} & \dots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & \mathbf{F}_N^{RF} \end{bmatrix} \quad (3)$$

式中: $\mathbf{F}_i^{RF} \in \mathbb{C}^{P \times N_{RF}}$ 表示第 i 个子阵列的 RF 波束形成器。

1.1 通信信号模型

考虑一个服务于多个点状目标的下行通信通道, 在传感环境中, 基站将波束引导到这些目标。第 u 个用户接收到的输出信号在信道中传输并经过组合处理后可表示为

$$\mathbf{y}_u = \mathbf{h}_u^H \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} + n_u \quad (4)$$

式中: n_u 表示加性高斯白噪声 (Additive White Gaussian Noise, AWGN), 满足 $n_u \sim \mathcal{CN}(0, \sigma_u^2)$ 。

在有限散射环境下, 式 (1) 中的天线发射和接收的波束可以看作是一个子阵列的平面波。考虑第 n 个子阵列, $\mathbf{h}_{u,n}$ 是对应的面向用户 u 有 M 条散射多路径的平坦衰落窄带信道, 可以表示为数组传播向量加权外积的和:

$$\mathbf{h}_{u,n} = \sum_{m=1}^M \alpha_m e^{-jk_\lambda D_m^{(n,u)}} \mathbf{a}_T(\theta_{m,n}) \quad (5)$$

$$\mathbf{a}_T(\theta_{m,n}) = [1, e^{-jk_\lambda d \cos \theta_{m,n}}, \dots, e^{-jk_\lambda (P-1) d \cos \theta_{m,n}}]^T \quad (6)$$

$$D_m^{(n,u)} = |\mathbf{p}(n) - \mathbf{p}(m)| + |\mathbf{p}(m) - \mathbf{p}(u)| \quad (7)$$

式中: α_m 是包括路径损失的第 m 条路径的复增益; 发射子阵列是以 d 为间隔的均匀线性阵列; $\mathbf{a}_T(\theta_{m,n})$ 表示归一化的发射数组响应向量, 其中 $\theta_{m,n}$ 表示第 u 个用户到第 n 个子阵列的第 m 条路径的离开角; $D_m^{(n,u)}$ 表示对应的距离; $\mathbf{p}(n)$ 表示第 n 个子阵列的位置。

在不同子阵列之间, 特殊排列且间距较大的单元提供了路径内多路复用, 从而提供距离维增益^[12], 整个信道应考虑在球面波模型下, 可表示为

$$\mathbf{h}_u = [\mathbf{h}_{u,1}^T, \mathbf{h}_{u,2}^T, \dots, \mathbf{h}_{u,N}^T]^T \quad (8)$$

1.2 雷达波束图

已有文献表明, MIMO 雷达的波束图设计相当于探测信号的协方差矩阵。在文献[5]中, 作者提出了一个最小二乘问题来逼近理想波束方向图。假设空间中有离散化采样的 X 个目标检测区域, 有 K 个感兴趣的检测目标区域, 对于子阵列 MIMO 雷达,

将理想的雷达波束图 $\tilde{\mathbf{P}}_d(\theta_x, D_x)$ 定义为

$$\tilde{\mathbf{P}}_d(\theta_x, D_x) = \begin{cases} 1, \theta_x \in \left[\bar{\theta}_k - \frac{\Delta_\theta}{2}, \bar{\theta}_k + \frac{\Delta_\theta}{2} \right], \\ D_x \in \left[\bar{D}_k - \frac{\Delta_D}{2}, \bar{D}_k + \frac{\Delta_D}{2} \right], k=1, 2, \dots, K \\ 0, \text{其他} \end{cases} \quad (9)$$

式中: $\bar{\theta}_k$ 表示第 k 个目标的方向; $\bar{D}_k$ 表示第 k 个目标的距离; 波束宽度为 Δ_θ, Δ_D 。

而第 x 个目标区域的实际发射波束图可表示为

$$\mathbf{B}(\theta_x, D_x) = [\mathbf{a}_T(\theta_x, D_x)]^H \mathbf{R}_T [\mathbf{a}_T(\theta_x, D_x)] \quad (10)$$

其中, 预编码波束的自相关矩阵 $\mathbf{R}_T \in \mathbb{C}^{NP \times NP}$ 记为

$$\mathbf{R}_T = E(\mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{s} \mathbf{s}^H \mathbf{F}_{BB}^H \mathbf{F}_{RF}^H) = \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{F}_{BB}^H \mathbf{F}_{RF}^H \quad (11)$$

导向向量可以表示为

$$\mathbf{a}_T(\theta_x, D_x) = [\mathbf{a}_T(\theta_x^{(1)}, D_x^{(1)}), \mathbf{a}_T(\theta_x^{(2)}, D_x^{(2)}), \dots, \mathbf{a}_T(\theta_x^{(N)}, D_x^{(N)})] \quad (12)$$

式中: $\mathbf{a}_T(\theta_x^{(n)}, D_x^{(n)})$ 表示第 n 个子阵列对应的导向向量。

用理想波束方向图 and 实际波束方向图之间的均方误差来作为衡量波束方向图相似性的指标:

$$\varepsilon(\alpha, \mathbf{F}_{RF}, \mathbf{F}_{BB}) = \sum_{x=1}^X |\alpha \tilde{\mathbf{P}}_d(\theta_x, D_x) - [\mathbf{a}_T(\theta_x, D_x)]^H \cdot \mathbf{F}_{RF} \mathbf{F}_{BB} \mathbf{F}_{BB}^H \mathbf{F}_{RF}^H [\mathbf{a}_T(\theta_x, D_x)]|^2 \quad (13)$$

式中: α 为缩放因子; $[\mathbf{a}_T(\theta_x, D_x)]^H \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} \mathbf{F}_{\text{BB}}^H \cdot \mathbf{F}_{\text{RF}}^H [\mathbf{a}_T(\theta_x, D_x)]$ 为实际波束方向图。

1.3 问题建模

由于同时求解矩阵 $\mathbf{F}_{\text{RF}}$ 和 $\mathbf{F}_{\text{BB}}$ 复杂度较高, 所以先联合优化两个矩阵, 令 $\mathbf{F}_{\text{BS}} = \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$, $\mathbf{F}_{\text{BS}} = [\mathbf{f}_1^{\text{BS}}, \mathbf{f}_2^{\text{BS}}, \dots, \mathbf{f}_U^{\text{BS}}]$ 。第 u 个用户将其他用户的信号视为干扰, 则第 u 个用户的 SINR 为

$$\gamma_u = \frac{|\mathbf{h}_u^H \mathbf{f}_u^{\text{BS}}|^2}{\sum_{i=1, i \neq u}^U |\mathbf{h}_u^H \mathbf{f}_i^{\text{BS}}|^2 + \sigma_0^2} \quad (14)$$

第 u 个用户的可达速率为 $\text{lb}(1 + \gamma_u)$ 。本文在发射功率约束和雷达性能指标约束的条件下, 通过在基站处设计混合波束赋形向量, 使所有 U 个用户的和通信速率最大化。在数学上, 和速率最大化问题可以表述为

$$\begin{aligned} \max_{\mathbf{F}_{\text{BS}}, \alpha} f_1(\mathbf{F}_{\text{BS}}, \alpha) &= \sum_{u=1}^U \text{lb}(1 + \gamma_u) \\ \text{s. t.} \quad \|\mathbf{F}_{\text{BS}}\|_F^2 &\leq P_T \\ \sum_{x=1}^X |\alpha \tilde{\mathbf{P}}_d(\theta_x, D_x) - \mathbf{B}(\theta_x, D_x)|^2 &\leq \epsilon \end{aligned} \quad (15)$$

式中: P_T 表示可用发射功率; ϵ 表示波束图相似度。

2 联合波束赋形设计

由于非凸目标函数 $f_1(\mathbf{F}_{\text{BS}}, \alpha)$, 通常很难得到上述优化问题的最优解。在本文中, 我们使用一种低复杂度的算法来解决这个问题, 该算法基于分式规划技术^[14]。FP 是一类涉及分数项(或比率)的优化问题, 可以直接采用 FP 方法来解决。优化问题可以等价转化为

$$f_{1a}(\mathbf{F}_{\text{BS}}, \boldsymbol{\beta}) = \sum_{u=1}^U \text{lb}(1 + 2\text{Re}\{\boldsymbol{\beta}_u^* \mathbf{h}_u^H \mathbf{f}_u^{\text{BS}}\} - |\boldsymbol{\beta}_u|^2 (\sum_{i=1, i \neq u}^U |\mathbf{h}_u^H \mathbf{f}_i^{\text{BS}}|^2 + \sigma_0^2)) \quad (16)$$

式中: $\boldsymbol{\beta}$ 表示辅助变量 $[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_U]$ 的集合。经多维二次变换解耦后, SINR 项可以转换为 $\mathbf{f}_u^{\text{BS}}$ 的凹函数。由于外部对数函数是非递减的凹函数, 当辅助变量 $\boldsymbol{\beta}_u$ 保持固定时, 优化问题(16)是一个 $\mathbf{f}_u^{\text{BS}}$ 的凸问题。

当给定 $\mathbf{f}_u^{\text{BS}}$ 时, 最优的 $\boldsymbol{\beta}_u$ 为

$$\boldsymbol{\beta}_u^\circ = \frac{\mathbf{h}_u^H \mathbf{f}_u^{\text{BS}}}{\sum_{i=1, i \neq u}^U |\mathbf{h}_u^H \mathbf{f}_i^{\text{BS}}|^2 + \sigma_0^2} \quad (17)$$

当给定 $\boldsymbol{\beta}_u^\circ$ 时, $\mathbf{F}_{\text{BS}}$ 的优化问题可以转化为

$$\begin{aligned} \max_{\mathbf{F}_{\text{BS}}, \alpha} f_{1a}(\mathbf{F}_{\text{BS}}, \boldsymbol{\beta}_u^\circ) \\ \text{s. t.} \quad \|\mathbf{F}_{\text{BS}}\|_F^2 &\leq P_T \end{aligned} \quad (18)$$

$$\sum_{x=1}^X |\alpha \tilde{\mathbf{P}}_d(\theta_x, D_x) - \mathbf{B}(\theta_x, D_x)|^2 \leq \epsilon$$

很容易发现, 变量 α 仅出现在第二个优化约束中, 而该约束条件是关于 α 的二次凸函数。因此, 我们可以直接使用典型的一阶最优条件即 $\frac{\partial E(\alpha, \mathbf{F}_{\text{BS}})}{\partial \alpha} = 0$ 来求解最优的 α 。定义 $\mathbf{A}_x = \mathbf{a}_T(\theta_x, D_x)[\mathbf{a}_T(\theta_x, D_x)]^H$, 则最优的 α 可以计算为

$$\alpha^\circ = \frac{\sum_{x=1}^X \tilde{\mathbf{P}}_d(\theta_x, D_x) \text{vec}^H(\mathbf{A}_x) \text{vec}(\mathbf{F}_{\text{BS}} \mathbf{F}_{\text{BS}}^H)}{\sum_{x=1}^X \tilde{\mathbf{P}}_d^2(\theta_x, D_x)} \quad (19)$$

然而, 式(18)中第二个约束对 $\mathbf{F}_{\text{BS}}$ 不是凸的。基于 MM 方法, 文献[15]提供了一种方法, 通过构造一系列易于优化的函数来解决问题, 可以获得局部最优解。优化问题被转化为

$$\begin{aligned} \max_{\mathbf{F}_{\text{BS}}} f_{1a}(\mathbf{F}_{\text{BS}}, \boldsymbol{\beta}_u^\circ) \\ \text{s. t.} \quad \|\mathbf{F}_{\text{BS}}\|_F^2 &\leq P_T \\ \sum_{u=1}^U \text{Re}\{(\mathbf{f}_u^{\text{BS}})^H \mathbf{B}_1^t \mathbf{f}_u^{\text{BS}} + 2(\mathbf{f}_u^{\text{BS}})^H \mathbf{v}_u\} + c_2^t &\leq \epsilon \\ \mathbf{b}_x &\triangleq \frac{\tilde{\mathbf{P}}_d(\theta_x, D_x) \sum_{x_1=1}^X \tilde{\mathbf{P}}_d(\theta_{x_1}, D_{x_1}) \text{vec}(\mathbf{A}_{x_1})}{\sum_{x_1=1}^X \tilde{\mathbf{P}}_d^2(\theta_{x_1}, D_{x_1})} - \text{vec}(\mathbf{A}_x) \end{aligned} \quad (20)$$

$$\mathbf{F}^t \triangleq \mathbf{F}_{\text{BS}}^t (\mathbf{F}_{\text{BS}}^t)^H \quad (22)$$

$$\begin{cases} c_1^t \triangleq \text{vec}^H(\mathbf{F}^t) (\lambda_m \mathbf{I} - \mathbf{B}) \text{vec}(\mathbf{F}^t) \\ \mathbf{B} = \frac{1}{X} \sum_{x=1}^X \mathbf{b}_x \mathbf{b}_x^H \end{cases} \quad (23)$$

$$\begin{aligned} \mathbf{B}_1^t &\triangleq \frac{2}{X} \sum_{x_1=1}^X \frac{\tilde{\mathbf{P}}_d^2(\theta_{x_1}, D_{x_1})}{\eta^2} \sum_{x_2=1}^X \tilde{\mathbf{P}}_d(\theta_{x_2}, D_{x_2}) \cdot \\ &\quad \text{vec}^H(\mathbf{A}_{x_2}) \text{vec}(\mathbf{F}^t) \sum_{x_3=1}^X \tilde{\mathbf{P}}_d(\theta_{x_3}, D_{x_3}) \mathbf{A}_{x_3} + \\ &\quad \frac{2}{X} \sum_{x_1=1}^X \text{vec}^H(\mathbf{A}_{x_1}) \text{vec}(\mathbf{F}^t) \mathbf{A}_{x_1} \end{aligned} \quad (24)$$

$$\begin{aligned} \mathbf{B}_2^t &\triangleq -\frac{4}{X} \text{Re}\left\{ \sum_{x_1=1}^X \frac{\tilde{\mathbf{P}}_d(\theta_{x_1}, D_{x_1})}{\eta} \text{vec}^H(\mathbf{A}_{x_1}) \text{vec}(\mathbf{F}^t) \cdot \right. \\ &\quad \left. \sum_{x_2=1}^X \tilde{\mathbf{P}}_d(\theta_{x_2}, D_{x_2}) \mathbf{A}_{x_2} \right\} - 2\lambda_m (\mathbf{F}^t)^H \end{aligned} \quad (25)$$

$$\begin{aligned} c_2^t &\triangleq -\sum_{u=1}^U \text{Re}\{((\mathbf{F}^t)^t)^H (\mathbf{B}_2^t)^H (\mathbf{f}_u^{\text{BS}})^t\} + \lambda_m P_T^2 + \\ &\quad \text{vec}^H(\mathbf{F}^t) (\lambda_m \mathbf{I} - \frac{1}{X} \sum_{x=1}^X \mathbf{b}_x \mathbf{b}_x^H) \text{vec}(\mathbf{F}^t) \end{aligned} \quad (26)$$

$$\mathbf{v}_u \triangleq (\mathbf{B}_2^t)^H (\mathbf{f}_u^{\text{BS}})^t \quad (27)$$

经过上述转化, 问题(20)中的优化目标和优化约束对于 $\mathbf{F}_{\text{BS}}$ 是凸问题, 并且可以被例如 MATLAB 中的凸问题工具箱(Convex Toolbox, CVX)等凸问题

求解器求解, 迭代求解联合波束赋形矩阵 $\mathbf{F}_{\text{BS}}$ 的过程可以用算法 1 (联合优化算法) 来表示, 其算法流程如下:

- 1 初始化: 满足功率约束的随机 $\mathbf{F}_{\text{BS}}^{(0)}$, 最优可达速率和 $s^{(0)}$, 迭代次数 $k=0$, 收敛阈值 ϵ_1 。
- 2 重复;
- 3 给定 $\mathbf{F}_{\text{BS}}^{(k)}$, 根据公式 (17) 求得 β_u° ;
- 4 给定 β_u° , 通过 CVX 求解优化问题 (20), 得到 $\mathbf{F}_{\text{BS}}^{(k+1)}$ 及 $s^{(k+1)}$;
- 5 令 $k=k+1$;
- 6 直到两次循环的可达速率和之差 $s^{(k+1)} - s^{(k)}$ 小于预先设定的阈值 ϵ_1 , 结束循环。

混合波束赋形设计问题接下来可以被转化为最小化 $\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}$ 与 $\mathbf{F}_{\text{BS}}$ 之间的欧氏距离, 从而将联合优化矩阵 $\mathbf{F}_{\text{BS}}$ 分解为数字波束赋形矩阵 $\mathbf{F}_{\text{BB}}$ 和模拟波束赋形矩阵 $\mathbf{F}_{\text{RF}}$, 即

$$\begin{aligned} \min_{\mathbf{F}_{\text{RF}}, \mathbf{F}_{\text{BB}}} & \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}} - \mathbf{F}_{\text{BS}}\|_F \\ \text{s. t.} & \|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F = P_T \\ & |[\mathbf{F}_{\text{RF}}]_{i,j}| = \frac{1}{\sqrt{P}}, \forall i, j \in S \\ & |[\mathbf{F}_{\text{RF}}]_{i,j}| = 0, \forall i, j \in \bar{S} \end{aligned} \quad (28)$$

当 $\mathbf{F}_{\text{RF}}$ 固定时, 可以用最小二乘法求解该问题, 得到结果为

$$\mathbf{F}_{\text{BB}} = \mathbf{F}_{\text{RF}}^\dagger \mathbf{F}_{\text{BS}} \quad (29)$$

值得注意的是, 在公式 (20) 中的功率约束已被临时移除, 这将在后续的算法 2 中处理。尽管如此, 公式 (28) 中的解已经为接收端的对应设计问题提供了全局最优解。该解可以进一步被归一化为

$$\mathbf{F}_{\text{BB}} = \frac{\sqrt{N_s}}{\|\mathbf{F}_{\text{RF}} \mathbf{F}_{\text{BB}}\|_F} \mathbf{F}_{\text{BS}}^\circ$$

由于单位模约束条件, $\mathbf{F}_{\text{RF}}$ 的优化更加复杂。在全连接情况下, 可以通过基于流行优化的方法找到 $\mathbf{F}_{\text{RF}}^{[16]}$ 。然而, 由于优化问题 (28) 中零约束条件的存在, $\mathbf{F}_{\text{RF}}$ 的设计并不是直观的。因此, 基于 MO 算法, 我们提出了以下解决方案。

当 $\mathbf{F}_{\text{BB}}$ 固定时, 优化问题 (28) 可以被转化为向量形式:

$$\begin{aligned} \min_{\mathbf{f}_{\text{RF}}} & \|\mathbf{G} \mathbf{f}_{\text{RF}} - \mathbf{f}_{\text{BS}}\|_F \\ \text{s. t.} & |[\mathbf{f}_{\text{RF}}]_i| = \frac{1}{\sqrt{P}}, \forall i \in V \\ & |[\mathbf{f}_{\text{RF}}]_i| = 0, \forall i \in \bar{V} \end{aligned} \quad (30)$$

式中: $\mathbf{G} = (\mathbf{F}_{\text{BB}}^\dagger) \otimes \mathbf{I}_{N_P} \in \mathbb{C}^{NPU \times N^2 P N_{\text{RF}}}$; $\mathbf{f}_{\text{RF}} = \text{vec} \{ \mathbf{F}_{\text{RF}} \} \in \mathbb{C}^{N^2 P N_{\text{RF}} \times 1}$; $\mathbf{f}_{\text{BS}} = \text{vec} \{ \mathbf{F}_{\text{BS}} \} \in \mathbb{C}^{NPU \times 1}$; V 和 $\bar{V}$ 分别表示 $\mathbf{f}_{\text{RF}}$ 中非零元素和零元素对应的集合。现在的目标

是排除与 $\bar{V}$ 对应的 $\mathbf{G}$ 和 $\mathbf{f}_{\text{RF}}$ 的元素, 以及找到 V 对应的 $\mathbf{f}_{\text{RF}}$ 中的元素, 以便使用 MO 算法, 并且未知向量的所有元素都符合单位模约束条件。MO 算法的思路如下:

假设去零操作后, 优化问题为

$$\begin{aligned} \min_{\mathbf{f}_{\text{RF}}} & \|\mathbf{G}' \mathbf{f}_{\text{RF}} - \mathbf{f}_{\text{BS}}\|_F \\ \text{s. t.} & |[\mathbf{f}_{\text{RF}}]_i| = \frac{1}{\sqrt{P}}, \forall i \in V' \end{aligned} \quad (31)$$

式中: $\mathbf{f}_{\text{RF}}$ 为 $\mathbf{f}_{\text{RF}}$ 去掉零元素后得到的向量; $\mathbf{G}'$ 为 $\mathbf{G}$ 去掉对应的列后得到的矩阵。由于单位模约束本质上是凸的, 因此将引入流形空间来求解该问题。如图 2 所示, 流形 $\mathbf{M}$ 是一个拓扑空间, 在每个点附近类似于欧几里得空间。换句话说, 流形上的每个点都有一个与欧几里得空间同胚的邻域。流形 $\mathbf{M}$ 上给定点 x 处的切空间 $T_x \mathbf{M}$ 由经过点 x 的曲线 $R_x(\xi)$ 的切向量 ξ 组成。

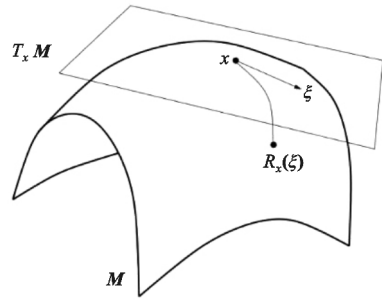


图 2 黎曼流形的切空间和切向量

首先用欧氏距离来定义复平面 $\mathbb{C}$:

$$\langle x_1, x_2 \rangle = \text{Re} \{ x_1^* x_2 \} \quad (32)$$

其等价于把 $\mathbb{C}$ 当作 $\mathbb{R}^2$ 的正则内积。进一步可以将复圆表示为

$$\mathbf{M}_{\text{cc}} = \{ x \in \mathbb{C} : x^* x = 1 \} \quad (33)$$

对于给定流形 $\mathbf{M}_{\text{cc}}$ 上的一个点 x 而言, 其可以移动的方向由切向量表示。该点处的切空间可以定义为

$$T_x \mathbf{M}_{\text{cc}} = \{ z \in \mathbb{C} : z^* x + x^* z = 2 \langle x, z \rangle = 0 \} \quad (34)$$

向量 $\mathbf{x} = \mathbf{f}_{\text{RF}}$ 组成了一个复圆流形 $\mathbf{M}_{\text{cc}}^m =$

$\left\{ \mathbf{f}_{\text{RF}} \in \mathbb{C}^m : |\mathbf{f}_{\text{RF},1}| = |\mathbf{f}_{\text{RF},2}| = \dots = |\mathbf{f}_{\text{RF},m}| = \frac{1}{\sqrt{P}} \right\}$, m 为 V' 中元素的个数, 因此, 优化问题 (31) 的搜索空间是复平面 m 个圆的积, 即 $\mathbb{C}^m$ 的具有积几何的黎曼子流形。因此, 给定点 $\mathbf{x} \in \mathbf{M}_{\text{cc}}^m$ 处的切空间可以表示为

$$T_x \mathbf{M}_{\text{cc}}^m = \{ z \in \mathbb{C}^m : \text{Re} \{ z \circ \mathbf{x}^* \} = \mathbf{0}_m \} \quad (35)$$

在所有切向量中, 与欧几里得空间类似, 其中一

个与负黎曼梯度有关的切向量表示函数减小最大的方向。因为复圆流形 M_{cc}^m 是 $\mathbb{C}^m$ 的黎曼子流形, 所以 $\mathbf{x}$ 处的黎曼梯度是由欧几里德梯度 $\nabla f(\mathbf{x})$ 在切空间 $T_x M_{cc}^m$ 上的正交投影给出的切向量梯度 $f(\mathbf{x})$:

$$\text{grad } f(\mathbf{x}) = \text{Proj}_x \nabla f(\mathbf{x}) = \nabla f(\mathbf{x}) - \Re \{ \nabla f(\mathbf{x}) \circ \mathbf{x}^* \} \circ \mathbf{x} \quad (36)$$

式(31)中代价函数的欧几里得梯度为

$$\nabla f(\mathbf{x}) = -2\mathbf{G}'^H [\mathbf{f}_{BS} - \mathbf{G}'\mathbf{x}] \quad (37)$$

接着使用回缩操作, 将向量从切线空间映射到流形本身, 当沿着切向量移动时, 确定流形上的目标位置。当搜索步长为 γ 时, 切向量 $\gamma\mathbf{d}$ 在点 $\mathbf{x} \in M_{cc}^m$ 处的回缩可以表示为

$$\text{Retr}_x: T_x M_{cc}^m \rightarrow M_{cc}^m:$$

$$\gamma\mathbf{d} \mapsto \text{Retr}_x(\gamma\mathbf{d}) = \text{vec} \left[\frac{(\mathbf{x} + \gamma\mathbf{d})_i}{|(\mathbf{x} + \gamma\mathbf{d})_i|} \right] \quad (38)$$

利用复圆流形 M_{cc}^m 的切空间、黎曼梯度和回缩, 可以开发一种基于线搜索的共轭梯度法。该方法是欧氏空间中的经典算法, 如算法 2(流形优化算法)所示:

- 1 初始化: 迭代次数 $k=0$, 设置收敛阈值 ϵ_2 。
- 2 计算共轭方向 $\mathbf{d}_0 = -\text{grad } f(\mathbf{x}_0)$
- 3 重复;
- 4 选择 Armijo 回溯线搜索步长 γ_k ;
- 5 使用回溯公式(38)找到下一个点 $\mathbf{x}_{k+1} = \text{Retr}_{x_k}(\gamma_k \mathbf{d}_k)$;
- 6 根据公式(36)和(37), 计算黎曼梯度 $\mathbf{g}_{k+1} = \text{grad } f(\mathbf{x}_{k+1})$;
- 7 计算从 $\mathbf{x}_k$ 到 $\mathbf{x}_{k+1}$ 梯度 $\mathbf{g}_k$ 和共轭方向 $\mathbf{d}_k$ 的移动向量 $\mathbf{g}_k^+$ 和 $\mathbf{d}_k^+$;
- 8 选择 Polak-Ribiere 参数 ρ_{k+1} ;
- 9 计算共轭方向 $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \rho_{k+1} \mathbf{d}_k^+$;
- 10 令 $k=k+1$;
- 11 直到黎曼梯度 $\|\mathbf{g}_k\|$ 小于预先设定的值 ϵ_2

其中, 与第 6 步中相对应的切向量 $\mathbf{d}$ 从到 $\mathbf{x}_{k+1}$ 的移动可以表示为

$$\text{Transp}_{x_k \rightarrow x_{k+1}}: T_{x_k} M_{cc}^m \rightarrow T_{x_{k+1}} M_{cc}^m: \mathbf{d} \mapsto \mathbf{d} - \Re \{ \{ \mathbf{d} \circ \mathbf{x}_{k+1}^* \} \circ \mathbf{x}_{k+1} \} \quad (39)$$

迭代求解混合波束赋形的两个矩阵的算法 3(混合波束赋形算法)流程如下:

- 1 初始化: 算法 1 得到的联合波束赋形矩阵 $\mathbf{F}_{BS}$, 相位随机 $\mathbf{F}_{RF}^{(0)}$, 迭代次数 $k=0$, 收敛阈值 ϵ_3 。
- 2 重复;
- 3 给定 $\mathbf{F}_{RF}^{(k)}$, 根据公式(29)求解 $\mathbf{F}_{BB}^{(k+1)}$, 并将其归一化;
- 4 给定 $\mathbf{F}_{BB}^{(k+1)}$, 通过算法 2 求解优化问题(30), 得到 $\mathbf{F}_{RF}^{(k+1)}$;
- 5 令 $k=k+1$;
- 6 直到两次迭代之间的目标函数值之差 $\|\mathbf{F}_{RF}^{(k)} \mathbf{F}_{BB}^{(k)} - \mathbf{F}_{BS}\|_F$ 小于 ϵ_3 , 结束循环。

综上, 为了得到混合波束赋形方案, 首先使用算法 1 得到联合波束赋形矩阵 $\mathbf{F}_{BS}$, 之后再采用算法 3

迭代求解得到两个子矩阵。其中, 算法 3 的初始值 $\mathbf{F}_{BS}^{(0)}$ 和 $\mathbf{F}_{RF}^{(0)}$ 都为随机产生, 最优通信速率 $s^{(0)}$ 则为一个设定的较大值; 算法 1 为求解联合波束赋形矩阵 $\mathbf{F}_{BS}$, 算法 3 为求解 $\mathbf{F}_{RF}$ 和 $\mathbf{F}_{BB}$ 。假设经过 $N_{\max}$ 次迭代算法 1 收敛, 其算法复杂度为 $O(N_{\max} P N^{3.5})^{[17]}$ 。当单位模约束有 m 个时, 采用流形优化的算法 3 的复杂度为 $O(m^{1.5})^{[18]}$ 。因此, 总的复杂度为 $O(N_{\max} P N^{3.5} + m^{1.5})$ 。求解联合波束赋形矩阵的算法可以收敛到局部最优解, 将两个矩阵分开求解的算法可以收敛到全局最优解, 因此算法 3 可以收敛到局部最优解。

3 仿真分析

本节对所提出的 ISAC 系统中基于子阵列和集中式阵列之间的联合波束设计的性能进行数值评估。仿真参数如下: 通信频率为 5 GHz; 考虑阵列由两个并排的子阵列组成, 每个子阵列为 8 个阵子组成的均匀线性阵列, 天线间距为半波长, 阵列中心的坐标为 $[0 \text{ m}, 0 \text{ m}]$ 。在不妨碍系统性能的前提下, 考虑系统服务于一个用户, 同时, 服务于 4 个感知目标, 感知目标的位置为 $([0.024 \text{ m}, 0.032 \text{ m}]$ 、 $[0.024 \text{ m}, -0.032 \text{ m}]$ 、 $[0.3 \text{ m}, 0.4 \text{ m}]$ 、 $[0.3 \text{ m}, -0.4 \text{ m}])$ 。本文考虑在距离与方位角二维域的理想方向图, 在感兴趣区域内的方向图增益为 0 dBm, 在感兴趣区域外的方向图增益为 $-\infty$ dBm。该理想情况下, 方向图的波束宽度为 5° , 最大增益为 16(天线数目), 无旁瓣。

图 3 为集中式阵列的感知性能图, 其中红色圈出的区域为感知目标所在的区域。由图可知, 集中式阵列能够很好地对不同角度上的感知目标进行区分, 且感知目标对应角度上的感知波束强度明显强于其他感知波束, 但对于同一角度不同距离上的两个目标, 感知波束无法对距离做出区分。这是因为集中式阵列信道矩阵维度低, 无法进一步提升性能。

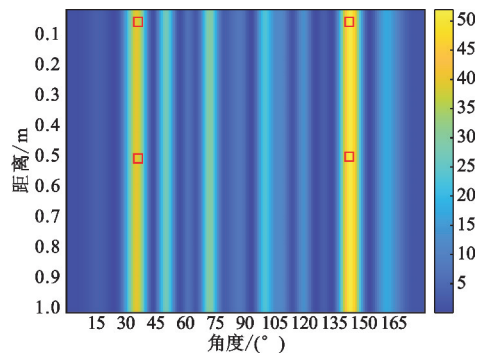


图 3 集中式阵列感知性能

图 4 为基于子阵列的感知性能图,可以看到,当把感知目标设在在 37° 和 143° 两个方向上,在 0.04 m 和 0.5 m 距离处,以及在 90° 方向上 0.9 m 处设置 5 个感知目标时,对应位置的波束强度明显较高,而在其他区域的感知波束强度则较低。结合图 5 中同一角度上子阵列和集中式阵列的对比,基于子阵列的 ISAC 系统在同一角度不同距离处的感知强度有明显变化,在感兴趣区域得到很明显的波峰,而集中式阵列的强度则没有变化,说明 ISAC 系统中采用子阵列建模可以有效补足集中式阵列无法分辨同一方向上不同距离的感知目标这一不足,获得更高的感知自由度。在感知目标的相邻区域,也会出现一条亮带,主要由于两个原因导致:一是为了保证算法有解,理想的感知目标区域在仿真时设置为感知目标的附近区域,而非一个点;二是由于优化得到的联合波束赋形矩阵的维度小于感知区域的维度,几个感知波束相互干扰,造成波束逸散。

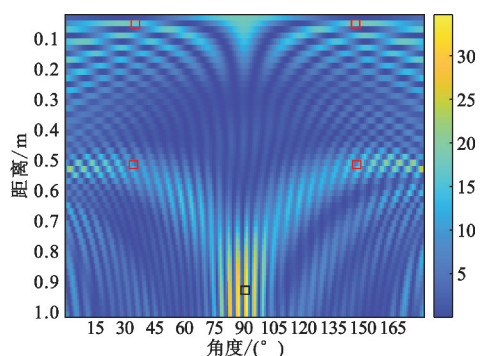


图 4 基于子阵列的感知性能

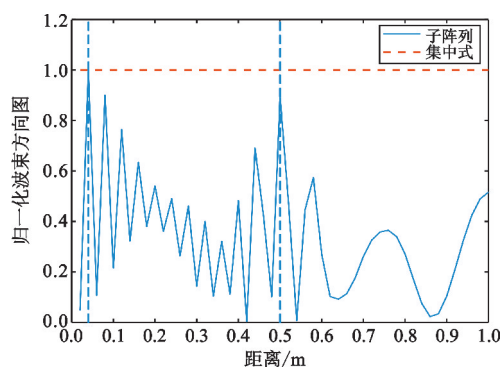
图 5 37° 方向上的感知性能

图 6 中,除了本方案提出的联合数字波束赋形和混合波束赋形,还包括了理想波束图和纯雷达场景中^[19]的波束图。可以看到,基于 MO 算法设计的分步逼近算法可以很好地逼近联合优化的感知性能。两条曲线整体趋势相似,且基于 MO 算法迭代求解得到的混合波束赋形矩阵的仿真曲线的波峰与

联合优化波束赋形矩阵的仿真曲线的波峰几乎重合在一起,都收敛到理想的波束角度上,同时基于 MO 算法得到的混合波束赋形方案可以把非感知目标角度上的波束强度降低,能够更精确地拟合理想的感知性能图。

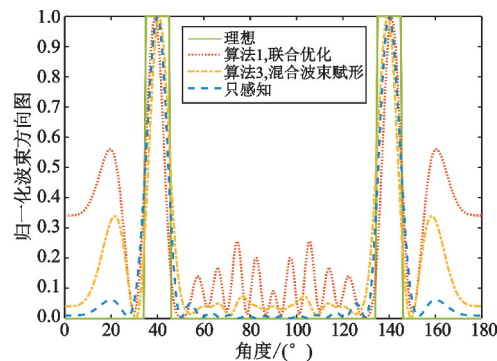


图 6 混合波束赋形算法感知性能

为了探究所提出算法的通信性能,图 7 分别比较了服务于两个用户,总天线数目为 16,集中式阵列和 $N=2$ 个分布式子阵列时,算法 1 得到的联合优化矩阵 $\mathbf{F}_{\text{BS}}$ 、算法 3 得到的混合波束赋形矩阵 $\mathbf{F}_{\text{BB}}$ 及 $\mathbf{F}_{\text{RF}}$,以及只考虑通信性能得到的联合优化矩阵 3 种情况下,计算的通信速率随信噪比变化图。从图中可以看到,随着信噪比的增大,通信速率也逐渐增大。在只考虑通信的情景中,两个子阵列相比于集中式阵列有 20% 左右的通信性能提升。虽然相比于只服务于通信的场景,保证感知性能会产生一定的通信性能损失,但在保证感知性能最大化通信速率的场景中,两个子阵列相比于集中式阵列的可达速率的提升达到了接近 40%,因此相比于集中式阵列,分布式子阵列能够有效提升系统的通信性能。同时,本文算法 1 得到的联合优化方案与算法 3 混合波束赋形方案的通信性能相差不超过 5%,说明基于流行优化设计的算法 3 具有很高的有效性,即使采用了混合波束赋形方案也基本没有性能损失,可以指导低能耗低成本波束赋形系统的设计。

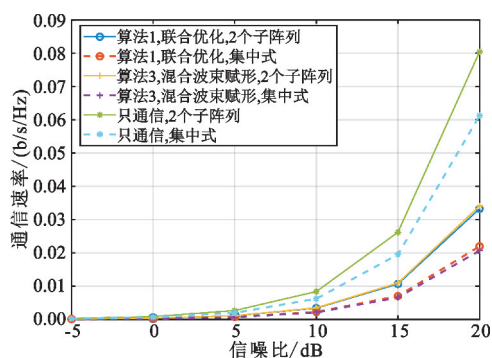


图 7 混合波束赋形算法通信速率随信噪比的变化曲线

4 结束语

本文提出了一种基于子阵列的大规模天线 ISAC 混合波束赋形系统,其中基站为多个单天线用户设备服务,同时主动检测多个目标。在满足雷达感知波束图相似度约束和发射功率预算的前提下,对模拟波束和数字波束形成器进行联合优化,最大限度地提高通信用户的可达和速率。提出了一种基于分式规划、最优化最小化和流形优化方法的高效交替算法,将得到的非凸优化问题转化为两个可解的子问题并进行迭代求解。仿真结果表明了子阵列排在 ISAC 系统中的优越性以及所提算法的有效性。

在接下来的工作中,将以低复杂度、高鲁棒性、高自适应性为目标,进一步研究基于子阵列的大规模天线 ISAC 混合波束赋形设计,进而继续研究在追踪感知目标的场景下,能够自适应调整感知目标区域的基于子阵列的大规模天线 ISAC 混合波束赋形设计。

参考文献:

- [1] CHOI J, VA V, GONZALEZ-PRELCIC N, et al. Millimeter-wave vehicular communication to support massive automotive sensing[J]. IEEE Communications Magazine, 2016, 54(12): 160-167.
- [2] 张若愚,袁伟杰,崔原豪,等. 面向 6G 的大规模 MIMO 通信感知一体化: 现状与展望[J]. 移动通信, 2022, 46(6): 17-23.
- [3] LIU F, ZHOU L F, MASOUIROS C, et al. Toward dual-functional radar-communication systems: optimal waveform design[J]. IEEE Transactions on Signal Processing, 2018, 66(16): 4264-4279.
- [4] 王佳欢,范平志,时巧,等. 一种具有多普勒容忍性的通感一体化波形设计[J]. 雷达学报, 2023, 12(2): 275-286.
- [5] LIU F, MASOUIROS C, LI A, et al. MU-MIMO communications with MIMO radar: from co-existence to joint transmission[J]. IEEE Transactions on Wireless Communications, 2018, 17(4): 2755-2770.
- [6] LIU X, HUANG T Y, SHLEZINGER N, et al. Joint transmit beamforming for multiuser MIMO communications and MIMO radar[J]. IEEE Transactions on Signal Processing, 2020, 68: 3929-3944.
- [7] WANG M L, CHEN J N, TAO J Y, et al. Over-the-air antenna array calibration for mmWave hybrid beamforming systems based on Monte Carlo Markov chain method[J]. IEEE Transactions on Vehicular Technology, 2023, 72(5): 6068-6079.
- [8] 张雷,高俊枫,向博. 最大化子阵列增益的多用户大规模 MIMO 混合预编码[J]. 电讯技术, 2021, 61(9): 1073-1079.
- [9] LIU F, MASOUIROS C. Hybrid beamforming with subarrayed MIMO radar: enabling joint sensing and communication at mmWave band[EB/OL]. [2023-11-05]. <https://arxiv.org/abs/1810.09812v2>.
- [10] ELBIR A M, MISHRA K V, CHATZINOTAS S. Hybrid beamforming for terahertz joint ultra-massive MIMO radar-communications[C]//2021 17th International Symposium on Wireless Communication Systems. Berlin: IEEE, 2021: 1-6.
- [11] GUIDI F, DARDARI D. Radio positioning with EM processing of the spherical wavefront[J]. IEEE Transactions on Wireless Communications, 2021, 20(6): 3571-3586.
- [12] SONG X H, RAVE W, BABU N, et al. Two-level spatial multiplexing using hybrid beamforming for millimeter-wave backhaul[J]. IEEE Transactions on Wireless Communications, 2018, 17(7): 4830-4844.
- [13] YAN L F, CHEN Y H, HAN C, et al. Joint inter-path and intra-path multiplexing for terahertz widely-spaced multi-subarray hybrid beamforming systems[J]. IEEE Transactions on Communications, 2022, 70(2): 1391-1406.
- [14] SHEN K M, YU W. Fractional programming for communication systems: Part I: power control and beamforming[J]. IEEE Transactions on Signal Processing, 2018, 66(10): 2616-2630.
- [15] LUO H H, LIU R, LI M, et al. Joint beamforming design for RIS-assisted integrated sensing and communication systems[J]. IEEE Transactions on Vehicular Technology, 2022, 71(12): 13393-13397.
- [16] YU X H, SHEN J C, ZHANG J, et al. Alternating minimization algorithms for hybrid precoding in millimeter wave MIMO systems[J]. IEEE Journal of Selected Topics in Signal Processing, 2016, 10(3): 485-500.
- [17] HUA M, WU Q Q, YANG L X, et al. A novel wireless communication paradigm for intelligent reflecting surface based symbiotic radio systems[J]. IEEE Transactions on Signal Processing, 2021, 70: 550-565.
- [18] LIU R, LI M, LIU Q, et al. Joint symbol-level precoding and reflecting designs for IRS-enhanced MU-MISO systems[J]. IEEE Transactions on Wireless Communications, 2021, 20(2): 798-811.
- [19] LI J, STOICA P. MIMO radar with colocated antennas[J]. IEEE Signal Processing Magazine, 2007, 24(5): 106-114.

作者简介:

郭鸿儒 男,1999 年生于山西大同,2021 年获工学学士学位,现为硕士研究生,主要研究方向为通信感知一体化。

顾忆宵 男,1995 年生于江苏海门,2023 年获工学博士学位,现为助理教授,主要研究方向为通信计算融合网络、卫星通信。

郭应鸿 男,1998 年生于四川成都,2016 年获工学学士学位,现为博士研究生,主要研究方向为通信感知一体化。

夏斌 男,1973 年生于湖南长沙,2004 年获工学博士学位,现为教授、博士生导师,主要研究方向为无线通信与网络。

DOI:10.20079/j.issn.1001-893x.231226003

一种用于低分辨毫米波通信系统的低采样率定时恢复算法^{*}

李世宝¹, 赵成锁¹, 李作志²

(1. 中国石油大学(华东)海洋与空间信息学院, 山东 青岛 266580; 2. 青岛港应急救援有限公司, 山东 青岛 266001)

摘要:针对低分辨毫米波定时恢复算法在低采样率下无法通过循环相关获取定时误差估计值的问题,提出了一种循环相关保持的低分辨低采样率定时恢复算法,并且设计了独特的预处理和后处理模块以保证循环相关估计器所必需的数据平稳特性和非混叠性。预处理模块通过均匀相位抖动以及非有理过采样获得循环相关要求的平稳基带数据。后处理模块对平稳基带数据进行复调制和低通滤波,解决了低采样率下的循环相关混叠消失问题。定时估计器模块利用数据的循环相关特性,对数据执行相关、滞后相关运算,通过相位求解方法导出定时误差估计结果值。仿真结果表明,所提算法与现有最优的 Martin 定时恢复算法相比定时误差估计精度提升了 75%。

关键词:毫米波通信;定时误差估计;循环相关;定时恢复;低采样率

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN929.5 文献标志码:A 文章编号:1001-893X(2025)03-0437-08

A Low Sampling Rate Timing Recovery Algorithm for Low Resolution Millimeter Wave Communication Systems

LI Shibao¹, ZHAO Chengsuo¹, LI Zuozhi²

(1. College of Oceanography and Space Informatics, China University of Petroleum(East China), Qingdao 266580, China;
2. Qingdao Port Emergency Rescue Co., Ltd., Qingdao 266001, China)

Abstract: For the problem that the low resolution millimeter wave timing recovery algorithm is unable to obtain the timing error estimate by cyclic correlation at low sampling rate, a low resolution and low sampling rate timing recovery algorithm for cyclic correlation holdover is proposed, and unique pre-processing and post-processing modules are designed to ensure the data smoothing characteristics and non-mixing required for cyclic correlation estimators. The preprocessing module obtains the smooth baseband data with cyclic correlation requirement by uniform phase jitter and irrational oversampling. The post-processing module performs complex modulation and low-pass filtering on the smooth baseband data to solve the problem of vanishing cyclic correlation aliasing at low sampling rates. The timing estimator module utilizes the cyclic correlation property of the data, performs correlation and hysteresis correlation operations on the data, and derives the timing error estimation result value through the phase solving method. Simulation results show that the proposed algorithm improves the timing error estimation accuracy by 75% compared with the existing Martin's algorithm, which is optimal for timing recovery.

Key words: millimeter wave communication; timing error estimation; cyclic correlation; timing recovery; low sample rate

^{*} 收稿日期:2023-12-26;修回日期:2024-04-23

基金项目:山东省自然科学基金项目(ZR2023LZH010, ZR2020MF005)

通信作者:李世宝 Email:lishibao@upc.edu.cn

0 引言

在无线通信领域,毫米波通信^[1-2]是一个提升数据传输速率^[3-4]的重要手段,已经引起了研究人员的广泛关注。但是毫米波通信高带宽的特性不可避免地导致了射频端较大的采样压力与高昂的硬件成本^[5]。为了应对这种挑战,现有的研究在射频端采用低分辨率模数转换器^[6-7] (Analog-to-Digital Converter, ADC),通过较低的量化精度来降低 ADC 的采样压力^[8]。较低量化精度的 ADC^[9]不需要自动增益控制且实现较为简单,已经得到了广泛研究^[10-11]。

ADC 以低量化精度进行采样是一种非线性行为^[12-13],这种情况下的定时同步往往考虑使用时间过采样来获取同步的参数估计^[9,12,14-18]。在完成毫米波通信系统定时恢复工作时,定时估计器的采样率要尽可能小。更低的采样率可以降低 ADC 的采样压力,节省硬件成本,以最小的代价保证后续的信道估计与解调。而在低采样率下完成低量化精度数据的定时恢复工作时,循环相关要求非混叠性以及数据平稳性无法满足,导致利用循环相关提取定时误差失败,所以十分有必要研究低采样率下低量化精度信号的定时恢复问题。

目前低采样率下低量化精度信号的定时恢复相关研究包含两类:低量化精度信号的定时恢复研究以及低采样率的定时恢复研究。低量化精度信号的定时恢复已经引起广泛的关注。文献[15]针对游程长度有限序列在快于奈奎斯特速率传输条件下的定时同步展开了研究,提出了一种考虑数据过采样的解决方案,但是该算法需要数据辅助,额外增加了系统开销。文献[19-20]指明了非数据辅助情况下一位量化信号参数估计的下界,但是并没有解决非数据辅助定时误差估计问题。文献[9,12,21]中低量化位宽的毫米波通信系统定时同步的相关算法,当每个符号的样本数量小于两个或过采样因子为 2 以下时,依赖的循环相关出现混叠,而存在混叠的循环相关无法有效利用其傅里叶变换的第一谱获取定时误差值。低采样率定时恢复也有一定的研究基础。文献[22]提出了每个符号的样本数量少于两个的定时同步算法,可以有效完成低采样率下的定时恢复,但是该算法是在无限分辨率下使用的。在低量化精度的情况下,基带信号并不是一个平稳的随机信号,因此经过量化后的信号无法满足文献

[22]中使用循环相关估计定时误差的数据平稳性要求,因此该算法无法有效获得定时误差估计值。上述提及的所有算法虽然可以分别解决低量化精度信号的定时恢复、低采样率定时恢复问题,但是在以较低的采样率获取低量化精度信号的定时误差时,传统的低分辨率毫米波定时恢复算法所依赖的循环相关信息混叠消失,无法通过傅里叶变换的第一谱求解与定时误差有关的相位值,定时误差估计失效。传统的低采样率定时恢复算法利用循环相关估计定时误差的前提条件是基带信号必须是平稳信号,而较低量化精度下该条件无法满足,因此该算法也同样无法使用循环相关提取定时误差。

针对上述问题,本文提出了一种用于低分辨率毫米波通信系统的低采样率定时恢复算法,以应对低采样率下低量化精度信号无法利用循环相关解决定时恢复估计问题。该算法基于循环相关保持原理,主要包含了预处理、后处理、定时误差估计器 3 个部分。在预处理模块采用了均匀采样与相位抖动保持低量化数据的平稳遍历特性,这将保证后续可以使用循环相关信息进行定时误差估计。后处理模块使用了低通滤波器与复调制对数据进行处理,以解决低采样率下循环相关混叠消失的问题。定时估计器利用数据的循环相关与滞后相关特性导出了定时误差估计结果值。最后通过仿真实验验证了所提算法的有效性,解决了传统算法在低采样率下无法利用低量化精度信号获取定时误差的难题。

1 信号模型

首先考虑发送端的系统模型:

$$x(t) = \sum_{n=-\infty}^{+\infty} a_n g_T(t-nT-\varepsilon T) \quad (1)$$

式中: a_n 是线性调制的复值信号,在本文中使用了 QPSK 信号作为基带符号; g_T 是发送的根升余弦滤波器,是一个实值函数且带宽限制在 $1/T$ 以内,其滚降因子为 ρ ; T 是符号持续时间; ε 为未知的定时误差,其限制在 $[-0.5T, 0.5T]$ 内。

发送端的信号经过信道,受到了零均值高斯白噪声以及未知的相位误差 ϕ 的影响,在接收端的信号表现形式为

$$r(t) = \sum_{n=-\infty}^{+\infty} a_n g_T(t-nT-\varepsilon T) e^{j\phi} + n(t) \quad (2)$$

在接收端进行信号采样:

$$r(kT_s) = \sum_{n=-\infty}^{+\infty} a_n g_T(kT_s-nT-\varepsilon T) e^{j\phi} + n(kT_s) \quad (3)$$

式中: T_s 为接收端的采样时间, 本文选择 $T/T_s = M$, 并考虑低采样率的 M 小于 2 大于 1 的情况。随后在接收端对上述信号量化:

$$y(kT_s) = \text{Csign}(r(kT_s)) = \text{Csign}\{R(r(kT_s))\} + j\text{Csign}\{I(r(kT_s))\} \quad (4)$$

式中: R 与 I 代表取 $r(kT_s)$ 的实部与虚部; Csign 函数是量化的过程, 以一位量化为例, 实际上的 Csign 函数可以表示为

$$\text{Csign}\{R(r(kT_s))\} = \begin{cases} 1, & R(r(kT_s)) \geq 0 \\ -1, & R(r(kT_s)) < 0 \end{cases} \quad (5)$$

量化后的信号 y_k 要经过匹配滤波器 g_{MF} 处理, 输出信号用来获取定时误差估计的结果值:

$$r(kT_s) = \sum_{k=-\infty}^{+\infty} y_k g_{MF}(-kT_s + nT + \varepsilon T) \quad (6)$$

2 基于循环相关保持的定时恢复算法

针对低量化精度的基带信号在低采样率的情况下循环相关信息丢失导致无法完成定时误差估计的问题, 本文提出了如图 1 所示的一种循环相关保持的定时恢复算法结构。该算法包含了预处理、后处理、定时误差估计器 3 个部分。预处理模块对中频接收信号添加了均匀采样抖动与相位抖动, 其中均匀的采样抖动通过非有理过采样实现, 而均匀的相位抖动则通过 $e^{j\varphi kT_s}$ 的复乘使得每个信号拥有不同相位。通过预处理模块获得了平稳的低量化位宽信号, 这是利用循环相关提取定时误差的必要条件。在后处理模块中, 为了解决低采样率下循环相关混叠消失进而无法获取定时误差估计值的问题, 对预处理的基带数据进行复调制与低通滤波, 保持了低采样率下的循环相关特性。在定时误差估计器模块推导了基带信号循环相关保持条件下的估计公式。

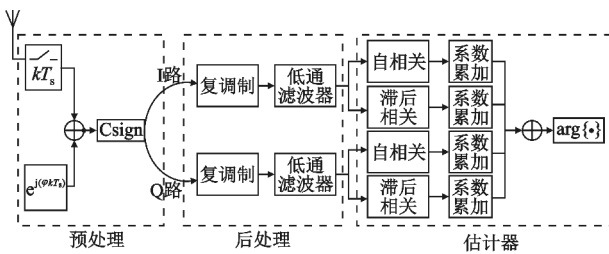


图 1 循环相关保持定时恢复示意

2.1 预处理

本节讨论均匀相位抖动与采样抖动对于低量化信号使用循环相关特性获取定时误差的重要性。随

后的部分将分析预处理的信号经过匹配滤波器仍然无法获取定时误差的原因, 这将在后处理部分解决。

为了使用循环相关提取定时误差, 要求用于估计的信号是一个平稳的随机过程。为了获得一位量化的平稳随机信号, 在中频采样的过程中引入了均匀的采样与相位抖动, 其中均匀的采样抖动可以通过非有理过采样实现, 均匀相位抖动通过符号复乘 $e^{j\varphi kT_s}$ 获得。经过预处理后的信号模型为

$$r(kT_s) = \sum_{n=-\infty}^{+\infty} a_n g_T(kT_s - nT - \varepsilon T) e^{j\varphi} e^{j\varphi kT_s} + n(kT_s) \quad (7)$$

均匀的采样相位通过非有理过采样 M 获得, 因此每个采样点会有不同的采样相位。均匀的相位抖动 $e^{j\varphi kT_s}$ 使用了抖动 φ 使每个样本 $r(kT_s)$ 拥有不同的相位。使用了均匀的采样抖动后获得的接收信号效果为

$$s(kT_s) = \sum_{n=-\infty}^{+\infty} a_n g_T(kT_s - nT - \varepsilon T - \tau_k) e^{j\varphi} e^{j\varphi kT_s} \quad (8)$$

非有理过采样因子 M 使得每一个采样点获得了 $\tau_k \in [-0.5T, 0.5T]$ 的随机相位, 消除了 $s(kT_s)$ 时间平均值对 ε 的依赖性, 并通过采样相位的随机化平均了集合平均值 ε , 获得了最终的平稳的随机过程。与上述的过程类似, $e^{j\varphi kT_s}$ 使每个样本 $s(kT_s)$ 拥有不同的相位, 消除了时间平均与集合平均中 φ 的依赖性。而 τ_k 与 φ 相互独立, 因此获得了平稳遍历过程 $s(kT_s)$ 。由于 $n(kT_s)$ 是独立的平稳随机过程, 且 $s(kT_s)$ 与 $n(kT_s)$ 是相互独立的, 得到的 $r(kT_s)$ 同样是平稳随机过程。经过量化后的信号表示形式为

$$y(kT_s) = \text{Csign}(r(kT_s)) \quad (9)$$

而 $y(kT_s) = f\{r(kT_s), r((k-1)T_s), \dots, r(T_s)\}$ 是公式(9)的量化函数 Csign 更具体的表现形式, 如果量化函数 $f(\cdot)$ 可测量, $r(kT_s)$ 是平稳的, 则 y_k 是平稳的。这个结论对于后处理部分使用循环相关提取定时误差十分重要。

经过预处理的基带信号经过匹配滤波器的输出信号为

$$z(nT + \varepsilon T) = \sum_{k=-\infty}^{+\infty} g_{MF}(nT + \varepsilon T - kT_s) e^{j\varphi kT_s} y_k \quad (10)$$

根据最小二乘准则, 定时误差的估计公式为

$$\hat{\varepsilon} = \arg\max_{\varepsilon} \sum_{n=-\infty}^{+\infty} |z(nT + \varepsilon T)|^2 = \arg\max_{\varepsilon} \sum_{n=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} |g_{MF}(nT + \varepsilon T - kT_s) e^{j\varphi kT_s} y_k|^2 \quad (11)$$

由于使用了较小的 $e^{j\varphi k T_s}$ 来对抗非线性失真,同时为了保持 y_k 信号的循环相关特性,因此在导出闭合解的时候考虑了 $\varphi=0$:

$$\hat{\varepsilon} = \arg\max \sum_{n=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} |g_{MF}(nT+\varepsilon T-kT_s)y_k|^2 \quad (12)$$

上述的过程可以理解为输出匹配滤波器的数据平方和,更换积分次序后上述公式可以表示为

$$\hat{\varepsilon} = \arg\max \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} |g_{MF}(nT+\varepsilon T-kT_s)y_k|^2 \quad (13)$$

事实上的估计可以理解为 y_k^2 与 $g_{MF}^2(nT+\varepsilon T-kT_s)$ 累乘求和的过程,进一步化简上述公式为

$$\hat{\varepsilon} = \arg\max \{E(\sum_{n=-\infty}^{+\infty} y_k^2 g_{MF}^2(nT+\varepsilon T-kT_s))\} \quad (14)$$

由于 y_k 是遍历平稳的随机过程,随机变量 y_k^2 是一个常数,上述公式只剩下一个匹配滤波器的平方项 $g_{MF}^2(nT+\varepsilon T-kT_s)$ 。上述最小二乘法的定时误差推导的形式等效于匹配滤波后的信号序列的时变相关 $r_z = E(z(k)z(k+\tau))$ 在滞后 $\tau=0$ 的情况下的定义。一般情况下,由于时变相关的周期性,在 $\tau=0$ 的情况下可以利用其傅里叶变换,也就是循环相关系数 $R(l, \tau)$ 的第一谱 $l=1$ 获取定时误差估计值:

$$\hat{\varepsilon} = \arg \left\{ \sum_{n=-\infty}^{+\infty} g_{MF}^2(nT+\varepsilon T-kT_s) e^{-j2\pi k/M} \right\} = \arg \left\{ \frac{1}{M} e^{-j2\pi \varepsilon} \frac{M}{T} \int_{-\frac{M}{2T}}^{\frac{M}{2T}} G(F+\frac{1}{2T}) G(F-\frac{1}{2T}) dF \right\} \quad (15)$$

式中的 $G(F)$ 是 $g_{MF}(t)$ 的傅里叶变换,由于 $g_{MF}(t)$ 被限制在 $1/T$ 中且是一个实偶函数,因此 $\int_{-\frac{M}{2T}}^{\frac{M}{2T}} G(F+\frac{1}{2T}) G(F-\frac{1}{2T}) dF$ 也是一个实数值。但是上述的公式仅仅在 $M \geq 2$ 时有效,当过采样 $T/T_s = M < 2$ 时无法从式(15)获取定时误差信息,因为上述的循环相关的假设不再有效。

从上述分析得知,即使经过预处理的信号仍然无法获得定时误差估计结果,这是因为低采样率下循环相关混叠消失。这个问题将在后处理模块处理。

2.2 后处理

为了获取低采样率下的定时误差,需要在后处理模块使用复调制与低通滤波器。通过上述操作消除了信号在低采样率下的混叠问题,保持了信号的循环相关特性,可以进行定时误差的估计。

首先需要对匹配滤波器的输出结果 $z(k)$ 实虚部分别进行半符号速率的复调制以获取余弦滤波器的滚降部分。假设升余弦滤波器的函数为

$$g(t) = \frac{T \sin(\frac{\pi t}{T}) \cos(\frac{\pi \rho t}{T})}{\pi t \left[1 - (\frac{2\rho t}{T})^2 \right]} \quad (16)$$

经过复调制和低通滤波信号可以表示为

$$\Re\{c(k)\} = [e^{-j\pi k/M} \text{real}\{z(k)\}] * h(k) \quad (17)$$

式中 $h(k)$ 为带宽为 $\rho/2T$ 的低通滤波器, ρ 为匹配滤波器的滚降因子系数。将经过匹配滤波器的接收信号代入到上述公式中可以获得

$$\Re\{c(k)\} = \frac{T}{2} e^{-j\pi \varepsilon/T} \sum_{n=-\infty}^{+\infty} h(nT+\varepsilon T-kT_s) \text{real}(y_k) e^{jn\pi} \quad (18)$$

其中 $h(t)$ 为经过半符号速率频移的低通滤波器与匹配滤波器的卷积结果:

$$h(t) = h_1(t) + jh_2(t) \quad (19)$$

式中的 $h_1(t) = \frac{\sin(\rho t/2T)}{\pi t}$, 而 $h_2(t)$ 可以表示为

$$h_2(t) = \frac{1}{2} \{h_1(t-T/2\rho) - h_1(t+T/2\rho)\}。$$

2.3 定时误差估计器

经过预处理与后处理模块的处理,有效解决了在低采样率下的低量化位宽数据的循环相关保持问题,因此在本节中导出了基于循环相关保持信号的定时误差估计器。经过循环相关保持处理的输出信号为式(18),利用输出的复信号的相关特性以获得定时误差估计值:

$$E\{\Re\{c(k)\} \Re\{c(k)\}\} = E\left\{ \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \text{real}(y_k)^2 h^2(nT+\varepsilon T-kT_s) \frac{T^2}{4} e^{-j2\pi \varepsilon/T} \right\} \quad (20)$$

式中的随机变量为 y_k 。在使用了均匀的相位抖动与均匀的采样抖动后, y_k 是一个平稳的随机变量。需要注意的是,只有在上述条件下才满足 y_k 是一个平稳的随机变量,因此 $E\{\text{real}(y_k)^2\}$ 是一个常数 C 。化简上述公式为

$$E\{\Re\{c(k)\} \Re\{c(k)\}\} = K \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} h_1^2(nT+\varepsilon T-kT_s) - h_2^2(nT+\varepsilon T-kT_s) C \quad (21)$$

式中: $K = \frac{T^2}{4} e^{-j2\pi \varepsilon/T}$ 。公式(21)中包含的积分项

$\sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} h^2(nT+\varepsilon T-kT_s)$ 结果可以展开为 $h_1(t)$ 的变体,其最终结果是一个实数值,因此可以直接从输出复值的平方中获取定时误差估计值:

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg \left\{ \frac{T^2}{4} e^{-j2\pi \varepsilon/T} \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} h^2(nT+\varepsilon T-kT_s) C \right\} \quad (22)$$

同样可以利用 $z(k)$ 的虚部使用相同的操作获取 $\Im\{c(k)\}^2$ 来获取定时误差值, 在有限长的数据序列中最终可以获得估计器如下:

$$\hat{\varepsilon} = -\frac{1}{2\pi} \arg\left\{ \sum_{k=0}^{N-1} \Im\{c(k)\}^2 + \Re\{c(k)\}^2 \right\} \quad (23)$$

为了进一步提升精度, 可以使用复信号的滞后相关, 二次获取定时误差估计值, 对复信号实虚部执行以下操作:

$$\Re\{c(k-1)\} = [e^{-j\pi k/M} \text{real}\{z(k-1)\}] * h(k) \quad (24)$$

将 $z(k-1)$ 代入上述的公式, 并且对输出的复信号求数学期望:

$$E\{\Re\{c(k)\} \Re\{c(k-1)\}\} = \frac{T^2}{4} e^{-\frac{j2\pi\varepsilon}{T}} e^{\frac{j\pi}{M}} \times$$

$$E\left\{ \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \text{real}(y_k) \text{real}(y_{k-1}) h^2(nT + \varepsilon T - kT_s) \right\} \quad (25)$$

如公式(20)后的文字所述, 因此 $E\{\text{real}(y_k) \cdot \text{real}(y_{k-1})\}$ 也是常数 C :

$$\hat{\varepsilon} = -\frac{1}{2\pi} e^{-\frac{j\pi}{M}} \times \arg\left\{ \frac{T^2}{4} e^{-\frac{j2\pi\varepsilon}{T}} e^{\frac{j\pi}{M}} \sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} h^2(nT + \varepsilon T - kT_s) C \right\} \quad (26)$$

由于 $\sum_{k=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} h^2(nT + \varepsilon T - kT_s)$ 最终结果是一个实数值, 因此可以从输出复值获取存在相偏 $e^{\frac{j\pi}{M}}$ 的定时误差估计值。

类似地, 以从 $z(k)$ 的虚部中获取定时误差值, 在有限长的数据序列中最终可以获得估计器如下:

$$\begin{aligned} \hat{\varepsilon}_{\text{lag}} = & -\frac{1}{2\pi} e^{-\frac{j\pi}{M}} \times \\ & \arg\left\{ \sum_{k=0}^{N-1} \Im\{c(k)\} \Im\{c(k-1)\} + \right. \\ & \left. \Re\{c(k)\} \Re\{c(k-1)\} \right\} \end{aligned} \quad (27)$$

组合上述的两个估计公式以获得更好的估计结果, 最终的估计结果如下:

$$\begin{aligned} \hat{\varepsilon} = & -\frac{1}{2\pi} \arg\left\{ \sum_{k=0}^{N-1} \Im\{c(k)\}^2 + \Re\{c(k)\}^2 + \right. \\ & e^{\frac{j\pi}{M}} [\Im\{c(k)\} \Im\{c(k-1)\} + \\ & \left. \Re\{c(k)\} \Re\{c(k-1)\}] \right\} \end{aligned} \quad (28)$$

3 算法仿真与结果分析

本节设置的对比算法为文献[9]提供的低分辨毫米波定时恢复算法以及文献[22]提供的传统低采样率定时恢复算法, 后文分别简称 Matin 算法和

Kim 算法。

图 2 给出了观察长度 $L=500$ 时不同量化位宽下的 Matin 算法、Kim 算法以及所提出循环相关保持算法的信噪比曲线。仿真设置了实验次数为 300 次以减少实验误差。在实验中参考了文献[17]设置了经典滚降因子为 0.25, 以及经典过采样倍数 M 为 1.5, 所提出循环相关提取算法应用了 0.02 的过采样抖动(过采样抖动设置值只需要满足非有理过采样使得信号具有不同的采样相位 τ_k 即可), 其中均匀的相位抖动设置为 $\varphi=0.02\pi$, 该相位抖动值参考了文献[9]要求的较低相位抖动以保证循环相关性。

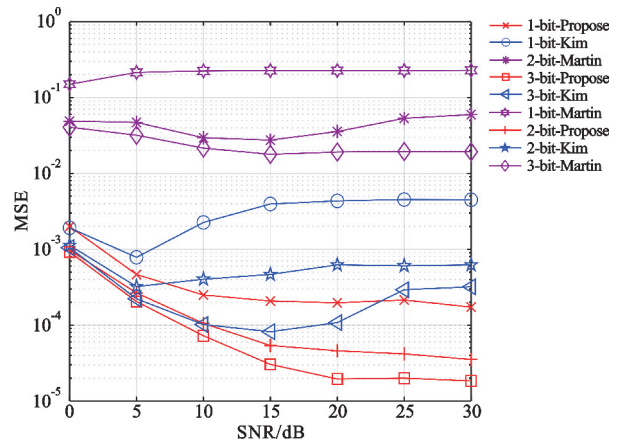


图 2 不同算法在低量化精度下的信噪比曲线

在实验设置的参数环境下, 原先的 Matin 算法在低采样率的情况下, 由于循环相关混叠的缘故已经无法获取定时误差估计值。而 Kim 算法在低采样率的情况下, 无法保证用于循环相关的数据统计特性, 信噪比曲线异常。在本文中, 综合考虑了定时误差估计过程中所必需的循环相关性以及循环相关所依赖的数据平稳特性, 成功解决了低量化位宽信号在低采样率的情况下定时误差的估计问题。随着量化位宽的提升, 所提算法的性能逐渐提升, 这是因为更多的量化比特导致的信号非线性影响在逐渐减少。

图 3 在观察长度 $L=500$ 的情况下展示了所提算法在不同信噪比下的系统误码率性能曲线。仿真设置了实验次数为 300 次以减少实验误差。在实验中设置了过采样为 1.5。所提出循环相关保持算法以及应用了 0.02 的过采样抖动, 其中均匀的采样抖动设置为 $\varphi=0.02\pi$ 。

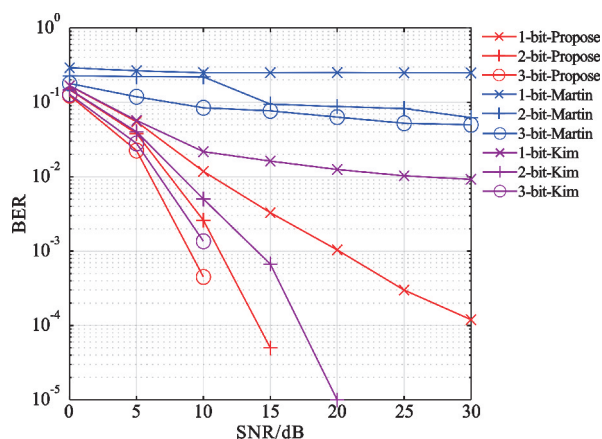


图3 不同估计算法的系统误码率对比曲线

在实验设置的参数环境下,随着信噪比的增大所提算法的系统误码性能得到了显著提升。在信噪比提高到一定情况下系统误码为0,并且随着量化比特逐渐上升,获得误码率为0时的信噪比越来越低。这是由于信号受到的非线性失真影响在逐渐降低。而 Martin 算法由于受到采样率的影响,即使在 30 dB 也存在误码。Kim 算法受到非线性失真的影响,在极低量化比特情况下,在 30 dB 仍存在较大误码,但是当量化比特逐渐上升,整体误码呈现下降趋势,但是仍不如所提算法的性能。

图4在观察长度 $L=500$ 的情况下展示了所提算法在不同滚降因子下的 MSE 性能。仿真设置了实验次数为 300 次以减少实验误差。在实验中设置了滚降因子范围为 0.1~0.9,信噪比为 30 dB,过采样倍数 M 为 1.5,所提出循环相关保持算法应用了 0.02 的过采样抖动,即设置了过采样 1.5 附近的无理数采样,其中均匀的采样抖动设置为 $\varphi=0.02\pi$ 。实验中为了充分考量算法在不同滚降因子下的性能,选取了较大滚降因子范围。

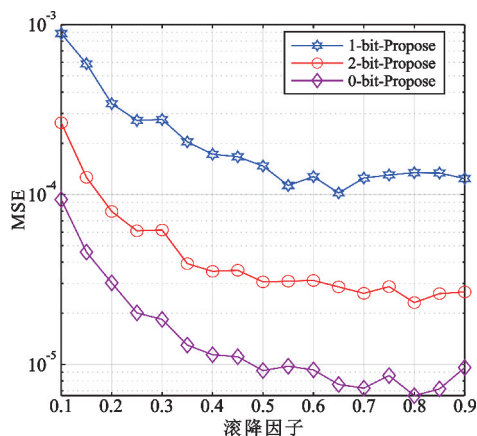


图4 所提算法在不同滚降因子下的 MSE 曲线

在实验设置的参数环境下,随着滚降因子的增大算法的性能得到了显著的提升,这是因为使用了更大的带宽获取有效的定时误差信息。但是当滚降因子增大到一定程度时算法的性能基本维持不变,这是由于带内噪声的增加制约了算法的性能进一步提升。

图5在滚降因子为 0.25 的情况下展示了所提算法在不同的观察长度的 MSE 性能。仿真设置了实验次数为 300 次以减少实验误差。在实验中设置了观察长度范围为 200~1 000,信噪比为 30 dB,过采样倍数 M 为 1.5,所提出循环相关保持算法应用了 0.02 的过采样抖动,即设置了 1.5 附近的无理数采样,均匀的采样抖动设置为 $\varphi=0.02\pi$ 。实验中为了充分考量算法在不同观察长度下的性能,选取了较大观察长度范围。

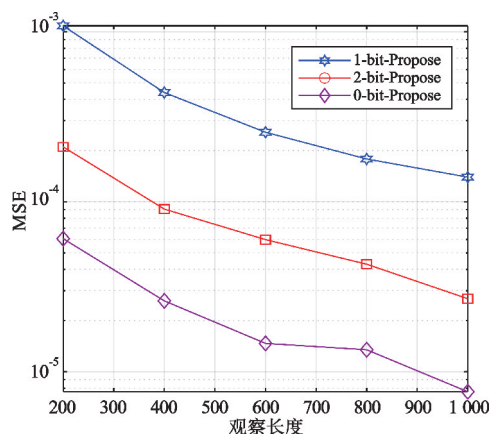


图5 所提算法在不同观察长度下的 MSE 曲线

在实验设置的参数环境下,从图5可以看出,随着观察长度的增大算法的性能得到了显著的提升,因为使用了更大的观察长度有利于获取更加精确的定时误差信息。

图6在观察长度 $L=200$ 的情况下展示了所提算法与 Martin、Kim 算法在 1.5~2.5 过采样情况下的 MSE 曲线。仿真设置了实验次数为 300 次以减少实验误差。实验设置过采样范围为 1.5~2.5,实验信噪比为 30 dB。所提出循环相关保持算法以及 Martin 算法应用了 0.02 的过采样抖动,其中均匀的采样抖动设置为 $\varphi=0.02\pi$ 。过采样区间参考了文献[17]的设置。

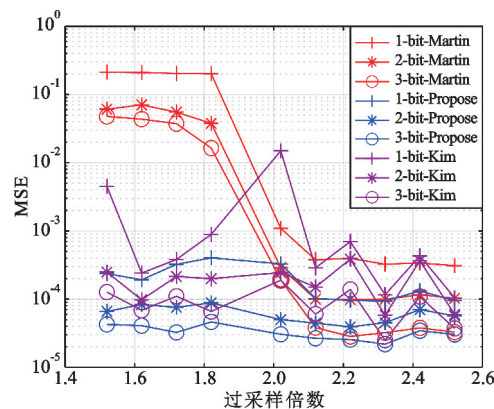


图 6 不同过采样因子下的算法性能对比

在实验设置的参数环境下,图 6 显示随着过采样倍数的增大, Martin 与所提算法出现明显分界。在过采样因子小于 2 时 Martin 失效,在过采样因子大于 2 时所提算法性能接近于 Martin。此外,随着量化比特数的提高,所提算法与 Martin 算法在有效区间内性能逐渐提升,这是由于信号受到的非线性失真影响在逐渐降低。Kim 算法由于未考虑信号非线性失真对数据统计学特性的影响,其信噪比曲线存在异常升高现象,最终导致在过采样实验区间内整体性能不如本文所提算法。

表 1 分析了 3 种算法计算复杂度,分别计算了 3 种算法实现所需要使用的实数乘法与实数加法次数。其中, L 是观察长度, M 是过采样的倍数, Q 是过采样抖动值。本文所提算法相比于 Martin 算法有较大的复杂度,但是能够在较低采样率情况下完成估计。相较于 Kim 算法,复杂度差异主要在于相位抖动的复乘以及过采样抖动,但是本文算法不会出现类似 Kim 算法估计异常的情况。

表 1 算法复杂度

算法	复杂度	
	实数乘法	实数加法
Kim ^[22]	$24LM+4$	$14LM+2$
Martin ^[9]	$12L(Q+M)$	$7L(Q+M)-1$
本文	$28L(Q+M)$	$16L(Q+M)+2$

4 结束语

针对低分辨毫米波通信系统要求低采样率以降低 ADC 的采样压力时无法通过循环相关获取定时误差信息的问题,本文提出了一种循环相关保持的定时误差估计算法。在该算法中应用循环相关保持的概念,并且设计了独特的预处理和后处理模块以

保证循环相关估计器完成定时误差估计所必需的依赖数据平稳特性和非混叠性。经过仿真验证,所提出的基于循环相关保持的定时恢复算法优于传统算法的估计结果。

本文充实了低分辨率毫米波技术在低采样率下完成定时误差估计的相关工作,文中所提算法能够以较低射频端硬件成本完成定时估计工作,可以适应要求低硬件成本与低功耗的室内场景,如 IEEE802.11ad 协议中定义的室内无线通信^[23],因此具有较强的实际应用价值。

但是,目前针对低分辨毫米波通信技术在插值恢复、载波同步等领域仍处于探索阶段,希望未来可以拓展丰富上述领域的相关研究。

参考文献:

[1] ZHANG J Y, DAI L L, LI X, et al. On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems[J]. IEEE Communications Magazine, 2018, 56(7):205-211.

[2] 李贵勇,于晓娜,高馨雨. IRS 辅助毫米波多用户系统的级联信道估计[J]. 电讯技术, 2025, 65(1):89-95.

[3] WANG H Q, SHIH W T, WEN C K, et al. Reliable OFDM receiver with ultra-low resolution ADC [J]. IEEE Transactions on Communications, 2019, 67(5):3566-3579.

[4] 张仲楷,崔高峰,和梦敏,等. 毫米波星间通信测距一体化技术研究[J]. 电讯技术, 2023, 63(1):24-31.

[5] WANG H Q, LIU T, WEN C K, et al. Optimal data detection for OFDM system with low-resolution quantization[C]//2016 IEEE International Conference on Communication Systems. Shenzhen: IEEE, 2016:1-6.

[6] CARNÌ D L, GRIMALDI D. Static and dynamic test of high resolution DAC based on over sampling and low resolution ADC[J]. Measurement, 2010, 43(2):262-273.

[7] AZIZZADEH A, MOHAMMADKHANI R, MAKKI S V A, et al. BER performance analysis of coarsely quantized uplink massive MIMO[J]. Signal Processing, 2019, 161:259-267.

[8] ZHANG R Y, ZHANG J Y, GAO Y L, et al. Bussgang decomposition-based sparse channel estimation in wideband hybrid millimeter wave MIMO systems with finite-bit ADCs[J]. Digital Signal Processing, 2019, 85:29-40.

[9] SCHLÜTER M, DÖRPINGHAUS M, FETTWEIS G P. Joint phase and timing estimation with 1-bit quantization and oversampling [J]. IEEE Transactions on Communications, 2022, 70(1):71-86.

[10] BENDER S, DÖRPINGHAUS M, FETTWEIS G. On the achievable rate of bandlimited continuous-time 1-bit quantized AWGN channels [C]//2017 IEEE

- International Symposium on Information Theory. Aachen: IEEE, 2017: 2083–2087.
- [11] CHAPEAU-BLONDEAU F, BLANCHARD S, ROUSSEAU D. Fisher information and noise-aided power estimation from one-bit quantizers [J]. Digital Signal Processing, 2008, 18(3): 434–443.
- [12] SCHLUTER M, DORPINGHAUS M, FETTWEIS G P. NDA timing estimation with 1-bit quantization and oversampling at the receiver [C]//2020 IEEE Global Communications Conference. Taipei, China: IEEE, 2020: 1–6.
- [13] KASHER M, SPASOJEVIC P, TINSTON M. Online memory-constrained frequency estimation for low-resolution non-linear ADCs [C]//2022 IEEE Wireless Communications and Networking Conference. Austin: IEEE, 2022: 956–961.
- [14] ZHU D L, BENDLIN R, AKOUM S, et al. Double-sequence frequency synchronization for wideband millimeter-wave systems with few-bit ADCs [J]. IEEE Transactions on Wireless Communications, 2020, 19(2): 1357–1372.
- [15] SCHLÜTER M, DÖRPINGHAUS M, FETTWEIS G P. On the timing synchronization under 1-bit quantization and oversampling [C]//2018 IEEE Statistical Signal Processing Workshop. Freiburg im Breisgau: IEEE, 2018: 198–202.
- [16] CHENG X T, XIA B Y, XU K, et al. Bayesian channel estimation and data detection in oversampled OFDM receiver with low-resolution ADC [J]. IEEE Transactions on Wireless Communications, 2021, 20(9): 5558–5571.
- [17] ZHU D L, BENDLIN R, AKOUM S, et al. Directional frame timing synchronization in wideband millimeter-wave systems with low-resolution ADCs [J]. IEEE Transactions on Wireless Communications, 2019, 18(11): 5350–5366.
- [18] SCHLUTER M, DORPINGHAUS M, FETTWEIS G P. Least squares phase estimation of 1-bit quantized signals with phase dithering [C]//2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications. Cannes: IEEE, 2019: 1–5.
- [19] SCHLUTER M, DORPINGHAUS M, FETTWEIS G P. Bounds on phase and frequency estimation from 1-bit quantized signals with phase dithering [C]//2019 IEEE International Conference on Communications. Shanghai: IEEE, 2019: 1–6.
- [20] SCHLÜTER M, DÖRPINGHAUS M, FETTWEIS G P. Bounds on phase, frequency, and timing synchronization in fully digital receivers with 1-bit quantization and oversampling [J]. IEEE Transactions on Communications, 2020, 68(10): 6499–6513.
- [21] OERDER M, MEYR H. Digital filter and square timing recovery [J]. IEEE Transactions on Communications, 1988, 36(5): 605–612.
- [22] KIM T K, MIN M. Improved non-data-aided feedforward symbol timing estimator for low-rate sampling systems [J]. IEEE Communications Letters, 2018, 22(5): 1010–1013.
- [23] 郭迎. 60 GHz 毫米波通信系统非理想接收机关键算法研究 [D]. 南京: 东南大学, 2024.

作者简介:

李世宝 男, 1978 年生于山东高密, 2002 年获硕士学位, 现为教授, 主要研究方向为移动计算、无线传感器网络、干扰对齐等。

赵成锁 男, 1999 年生于山西孝义, 2017 年获学士学位, 现为硕士研究生, 主要研究方向为无线通信、同步等。

李作志 男, 1978 年生于山东青岛, 2017 年获硕士学位, 现为高级政工师, 主要研究方向为陆海双向应急救援。

DOI:10.20079/j.issn.1001-893x.231228002

基于亚模函数的可见光通信 MIMO-OFDM 系统天线选择算法*

贾科军¹, 贺耀民¹, 张芳芳¹, 蔺莹¹, 薛建彬¹, 郝莉²

(1. 兰州理工大学 计算机与通信学院, 兰州 730050; 2. 西南交通大学 信息科学与技术学院, 成都 610031)

摘要:在可见光通信多输入多输出系统中,针对天线选择理论建模不足和穷举算法复杂度过高的问题,提出了基于亚模函数的天线选择方案。首先,以下行链路的信道容量最大化为目标,建立了基于亚模函数的天线选择理论优化模型,并证明了目标函数满足的单调亚模性。其次,根据亚模函数的收益递减效应,设计了基于容量最大化的天线选择算法。最后,仿真分析了非对称限幅光正交频分复用(Asymmetrically Clipped Optical Orthogonal Frequency Division Multiplexing, ACO-OFDM)和直流偏置光 OFDM(DC-biased Optical OFDM, DCO-OFDM)系统的信道容量和误码率性能。在 6 选 4 的情况下,当信噪比为 30 dB 时,所提算法与穷举最优算法的信道容量差异仅为 0.51 b/s/Hz 和 1.2 b/s/Hz,复杂度则降低了约 46.3%。另外,随着选择天线数的增多和调制阶数的增大,系统的误码率性能逐渐变差。

关键词:可见光通信(VLC);多输入多输出(MIMO);天线选择;亚模函数;收益递减效应

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN929.1 文献标志码:A 文章编号:1001-893X(2025)03-0445-09

An Antenna Selection Algorithm for Visible Light Communication MIMO-OFDM Systems Based on Submodular Function

JIA Kejun¹, HE Yaomin¹, ZHANG Fangfang¹, LIN Ying¹, XUE Jianbin¹, HAO Li²

(1. School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China;
2. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: In the visible light communication multiple-input multiple-output system, an antenna selection scheme based on submodular function is proposed to address the problems of insufficient theoretical modeling of antenna selection and high complexity of the exhaustive algorithm. Firstly, the theoretical optimization model of antenna selection based on submodular function is established with the objective of maximizing the channel capacity of the downlink, and the monotonic submodularity satisfied by the objective function is proved. Secondly, the antenna selection algorithm based on capacity maximization is also designed based on the diminishing returns effect of the submodular function. Finally, the channel capacity and bit error rate (BER) performance of asymmetrically clipped optical orthogonal frequency division multiplexing (ACO-OFDM) and DC-biased optical OFDM (DCO-OFDM) systems is analyzed through simulation. In the case of 6-select-4, when the signal-to-noise ratio is 30 dB, the difference in channel capacity between the proposed algorithm and the exhaustive optimal algorithm is only 0.51 b/s/Hz and 1.2 b/s/Hz, while the complexity is reduced by about 46.3%. In addition, the BER performance of the system gradually deteriorates with the increase of the number of selected antennas and the modulation order.

Key words: visible light communication (VLC); multiple-input multiple-output (MIMO); antenna selection; submodular function; diminishing returns effect

* 收稿日期:2023-12-28;修回日期:2024-05-06

基金项目:国家自然科学基金资助项目(61875080);甘肃省科技计划项目(22JR5RA276, 22JR5RA274, 23YFGA0062, 2022A-215);兰州理工大学博士科研启动经费项目(061903)

通信作者:贾科军 Email:kjia@lut.edu.cn

0 引言

随着发光二极管 (Light Emitting Diode, LED) 的快速普及, 高速无线接入需求的急剧增加, 以及未来 6G 布局的带动, 可见光通信 (Visible Light Communication, VLC) 技术近年来获得了迅猛发展^[1]。VLC 作为一种频谱无需授权的通信方式, 可以解决传统射频 (Radio Frequency, RF) 通信频谱资源短缺的问题, 在室内环境中实现高保密、人体无害、无电磁辐射的高速通信^[2]。VLC 也被视为未来室内无线移动通信网络的补充技术之一。

目前广泛使用的荧光粉 LED 的调制带宽一般小于 10 MHz, 限制了 VLC 系统传输速率的提高, 频带利用率较低^[3]。多输入多输出 (Multiple-Input Multiple-Output, MIMO) 技术则可以在不增加系统带宽的条件下提高频谱利用率, 改善数据传输速率。实际的室内环境中, 为满足照明需求通常会安装多个 LED, 这也为室内可见光通信 MIMO 系统 (VLC-MIMO) 的实现提供了天然条件。相比于单光源的通信方式, VLC-MIMO 系统不仅可以充分利用空间资源提高系统的信道容量和频谱效率, 还能克服单 LED 信号容易被遮挡的缺点^[4]。另外, 由于正交频分复用技术 (Orthogonal Frequency Division Multiplexing, OFDM) 可以将宽带信道转化成若干个平坦的窄带子信道, 使传播特性线性化, 从而更好地实施 MIMO 技术。因此, 将 MIMO 与 OFDM 结合的可见光通信 MIMO-OFDM 系统可以通过空间复用 (Spatial Multiplexing, SMP) 技术提供更高的数据传输速率和频谱利用率^[5]。然而, 在实际照明的多 LED 场景下, 如果所有的 LED 同时传输多个数据流, 容易导致系统的能耗和复杂度过大。同时 MIMO-OFDM 系统也会受到信道间干扰 (Inter Channel Interference, ICI) 增加、天线同步困难等问题的严重影响, 导致系统的性能变差^[6]。

天线选择作为一种低成本、低复杂度的有效方法, 可以通过特定的准则选择出具有最佳信道条件的天线子集, 实现保留 MIMO 技术优势的同时, 解决多天线能效较低和系统复杂度高的问题。在 RF 中天线选择技术已受到了广泛的研究和应用^[7-12]。

在室内 VLC-MIMO 系统的天线选择研究方面, 文献 [13] 提出了一种新的基于最小均方误差 (Minimum Mean Square Error, MMSE) 的空间调光

(Space Dimming, SD) 方案, 通过在不同的调光水平下选择最佳工作的 LED 子集, 实现调光控制的同时提高误码率 (Bit Error Ratio, BER) 性能。文献 [14] 提出了一种基于天线选择的新型调光方案, 通过递增和递减算法来选择 LED 的最优子集, 相比于传统调光方案具有更好的 BER 性能。但是, 上述方案主要还是以调光目标为前提考虑的 LED 选择。文献 [15] 提出了一种基于有效比阈值的 LED 选择方案, 通过该阈值选择通信 LED, 以提高信噪比 (Signal-to-Noise Ratio, SNR), 但也只适用于单用户的情况。文献 [16] 提出了欧氏距离天线选择 (Euclidean Distance Antenna Selection, EDAS) 算法, 通过遍历计算天线间的欧氏距离, 选取相关度较低的天线, 但遍历的复杂度较高, 限制了算法的应用。文献 [17] 利用调制符号本身的旋转对称性, 降低符号搜索空间的大小, 从而降低 EDAS 算法复杂度, 但是无法从本质上摆脱遍历的特点, 导致实际的选择效率并未有太大的提升。总之, 现有的针对 VLC-MIMO 系统的天线选择方案仍存在复杂度较高、适用场景受限和理论建模不足的问题。

本文利用集合函数亚模函数在最大化问题求解时较好的性能, 将 VLC-MIMO 系统下行链路的天线选择问题与亚模函数相结合, 建立了基于亚模函数的天线选择优化模型; 然后以信道容量最大化为目标, 证明了目标函数的单调亚模性, 使得 VLC-MIMO 系统的天线选择这一组合优化问题的求解有了理论基础; 其次, 利用亚模函数的收益递减性质设计了通信 LED 的选择算法; 最后, 在 VLC-MIMO 系统多径信道下, 对采用亚模函数的天线选择算法进行了仿真分析, 验证了其在信道容量和算法复杂度以及误码率方面的性能。

1 系统模型

采用强度调制/直接检测 (Intensity Modulation/Direct Detection, IM/DD) 的室内 VLC-MIMO 系统的通信场景如图 1 所示。建立室内空间坐标系, 坐标系原点 O 与房间左后下角重合, xOy 平面与地面重合。安装在屋顶的 N_T 个 LED 都可用于照明和通信, 但只将选中的 LED 作为发射天线, N_R 个光电探测器 (Photo Detector, PD) 全部作为接收天线。

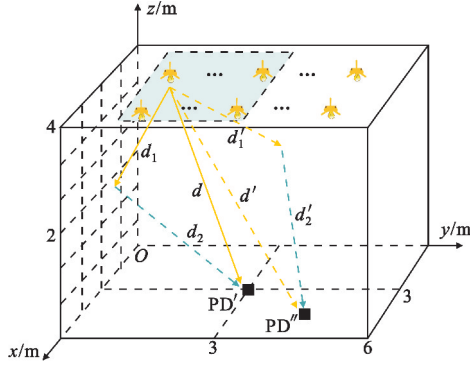


图 1 室内 VLC-MIMO 几何场景

考虑下行信道空间复用系统, LED 同步发射不同的光信号, 并考虑两种典型的光 OFDM (Optical OFDM, O-OFDM) 系统: 非对称限幅光 OFDM (Asymmetrically Clipped Optical OFDM, ACO-OFDM) 和直流偏置光 OFDM (DC-biased Optical OFDM, DCO-OFDM), 建立室内可见光通信 MIMO-OFDM 天线选择系统, 原理如图 2 所示。

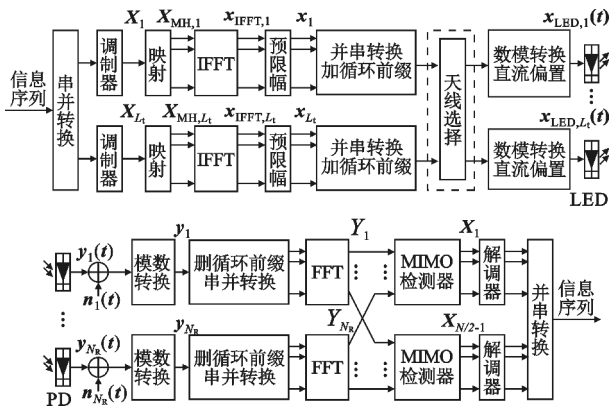


图 2 可见光通信 MIMO-OFDM 天线选择系统原理

二进制信源信息经串并转换后分成 L_1 路并行信号, 每一路信号经过 O-OFDM 调制后驱动 LED 发光, 故以第 n_1 路信号为例, 对发端原理进行说明。

首先将二进制信号进行 M 维正交幅度调制 (Quadrature Amplitude Modulation, QAM), X_{n_1} 表示功率归一化的调制符号。然后对 X_{n_1} 进行映射操作, 使输出信号 X_{MH,n_1} 满足厄米特 (Hermitian) 对称性, 即

$$X_{MH,n_1}^{(ACO)} = \begin{bmatrix} 0 & X(0) & 0 & X(1) & \cdots & X(N/4-1) \\ 0 & X^*(N/4-1) & 0 & \cdots & X^*(0) \end{bmatrix}^T \quad (1)$$

$$X_{MH,n_1}^{(DCO)} = \begin{bmatrix} 0 & X(1) & X(2) & \cdots & X(N/2-1) \\ 0 & X^*(N/2-1) & \cdots & X^*(2) & X^*(1) \end{bmatrix}^T \quad (2)$$

式中: $(\cdot)^*$ 表示共轭运算; $(\cdot)^T$ 表示矩阵转置; N 是逆快速傅里叶变换 (Inverse Fast Fourier Transform, IFFT) 的长度。可以看出, ACO-OFDM 系统所有的奇数子载波包含信息, 偶数子载波为零。DCO-OFDM 系统第 0 个和第 $N/2$ 个子载波为零, 其他子载波都携带信息。

与传统复数 OFDM 信号相同, IFFT 后输出的时域信号 x_{IFFT,n_1} 也具有较高的峰均比。为了减少 LED 非线性影响, 对 x_{IFFT,n_1} 预限幅以保证 LED 工作在线性区内。限幅信号经并串转换后, 添加循环前缀 (Cyclic Prefix, CP) 以消除多径效应的影响。然后通过天线选择算法, 从 N_T 个 LED 中选择 L_1 个作为发射天线。最后进行数模转换, 并且为了得到单极性实数信号以及提供足够的照明亮度, 添加直流偏置 B_{DC} , 驱动选出的 LED 同时发送光信号。

通常情况下, LED 的驱动信号需大于开启电压, 同时小于最大允许电压, 否则 LED 无法开启发光, 或者可能过热烧毁。LED 的驱动信号等于直流偏置与 O-OFDM 信号之和。假设 LED 的线性工作区范围为 $V_{\min} \sim V_{\max}$, 预限幅的上下限幅门限则由 LED 的工作区范围和直流偏置共同决定。ACO-OFDM 上下限幅门限分别为 $\varepsilon_{\text{top}} = V_{\max} - B_{DC}$ 和 $\varepsilon_{\text{bottom}} = \max(V_{\min} - B_{DC}, 0)$, DCO-OFDM 的限幅门限分别为 $\varepsilon_{\text{top}} = V_{\max} - B_{DC}$ 和 $\varepsilon_{\text{bottom}} = V_{\min} - B_{DC}$ [18]。

光信号经过室内 VLC-MIMO 多径信道传输后到达接收端, $\mathbf{H}$ 为 $N_R \times N_T$ 维的原始信道矩阵, 可以表示为

$$\mathbf{H} = \begin{bmatrix} h_{1,1} & h_{1,2} & \cdots & h_{1,N_T} \\ h_{2,1} & h_{2,2} & \cdots & h_{2,N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1} & h_{N_R,2} & \cdots & h_{N_R,N_T} \end{bmatrix} \quad (3)$$

式中: h_{n_r,n_t} 表示从第 n_t 个 LED 到第 n_r 个 PD 的信道增益。天线选择后所对应的传输矩阵为 $\mathbf{H}_s$, 即由原信道矩阵 $\mathbf{H}$ 中选出的部分列向量构成的 $N_R \times L_1$ 维的新矩阵。

接收端 N_R 个 PD 分别检测光信号并转换为电信号, 其中第 n_r 个 PD 的电信号为

$$y_{n_r}(t) = \gamma \sum_{n_t=1}^{L_1} h_{n_r,n_t}(t) \otimes x_{\text{LED},n_t}(t) + n_{n_r}(t) \quad (4)$$

式中: $x_{\text{LED},n_t}(t)$ 表示第 n_t 个 LED 的发射信号; γ 为光电转换系数; $n_{n_r}(t)$ 表示背景光噪声和电路热噪声之和, 是独立于信号的加性高斯白噪声 (Additive White Gaussian Noise, AWGN) [19]。

信号 $y_{n_t}(t)$ 经过模数转换和串并转换后去除 CP, 然后输入到 N 点快速傅里叶变换 (Fast Fourier Transform, FFT) 模块, 输出用于解调的频域信号 Y_{n_t} 。与发送端相对应, 提取 Y_{n_t} 中的有效子载波作为解调信号输入到线性 MIMO 检测器。为了消除或最小化 LED 发送信号间的干扰和信道噪声, 给接收信号乘以一个加权矩阵, 采用了迫零 (Zero Forcing, ZF) 和 MMSE 两种线性检测算法。ZF 检测加权矩阵为 $\mathbf{W}_{ZF} = (\mathbf{H}_s^H \mathbf{H}_s)^{-1} \mathbf{H}_s^H$, 其中, $(\cdot)^{-1}$ 表示矩阵求逆, $(\cdot)^H$ 表示矩阵的共轭转置, $\mathbf{W}_{ZF}$ 是 $L_t \times N_R$ 维的加权矩阵, $\mathbf{H}_s$ 是传输矩阵。MMSE 检测加权矩阵为 $\mathbf{W}_{MMSE} = (\mathbf{H}_s^H \mathbf{H}_s + N_0 \mathbf{I})^{-1} \mathbf{H}_s^H$, 其中, $\mathbf{I}$ 表示 $L_t \times N_R$ 维的单位矩阵, N_0 是噪声的单边功率谱密度。最后将 MIMO 检测器输出的符号输入到 QAM 解调器, 解调恢复出原始二进制信息。

2 基于亚模函数的天线选择

2.1 VLC-MIMO 信道容量

由于 VLC 系统同时实现照明和通信功能, IM/DD 系统中信息被调制为瞬时光强度, 输入信号不仅要满足非负性, 还需满足平均光功率相对稳定, 因此传统 RF 通信的信道容量计算结果不能直接用于 VLC。考虑信号非负性、峰值功率受限和平均功率不变的约束, 使输入信号满足 $0 \leq x_{LED, n_t} \leq A_m$ 。同时, 由于照明需求, 平均光强为一个常数, 即 x_{LED, n_t} 同时满足约束

$$E(x_{LED, n_t}) = \int_0^{A_m} x_{LED, n_t} f_X(x) dx = \xi P \quad (5)$$

式中: $f_X(x)$ 是 x_{LED, n_t} 的概率密度函数; ξ 表示光照目标, 满足 $0 \leq \xi \leq 1$; A_m 表示每个 LED 的峰值光功率; P 是 LED 的额定光功率, 满足 $P \leq A_m$ 。

文献[20]对 VLC-MIMO 系统信道容量进行了分析, 用奇异值分解将 MIMO 信道矩阵 $\mathbf{H}$ 分解成 L 个独立的并行子信道, $L \leq \min(N_T, N_R)$ 为信道矩阵 $\mathbf{H}$ 的秩。考虑光照条件不变的情况下, 定义第 n 个子信道的平均光强与峰值光强的比为 $\alpha_n = \xi P / A_m$, 当 $\alpha_n = 0.5$ 时, 第 n 个子信道的容量下限表示为

$$C_n \geq \frac{1}{2} \ln \left[1 + \frac{(\lambda_n A_m \sum_{j=1}^{N_T} |v_{j,n}|)^2}{2\pi e \sigma^2} \right] \quad (6)$$

式中: λ_n 为奇异值分解后对角矩阵的第 n 个元素; $v_{j,n}$ 是 $N_T \times N_T$ 维酉矩阵的第 j 行; 第 n 列元素; σ^2 是噪声方差。

奇异值分解前后系统的信息熵不变, 所以可见

光通信 MIMO 系统的信道容量可转变成 L 个子信道的信道容量之和, 即

$$C_{MIMO} = \frac{1}{2} \sum_{n=1}^L \ln \left[1 + \frac{(\lambda_n A_m \sum_{j=1}^{N_T} |v_{j,n}|)^2}{2\pi e \sigma^2} \right] \quad (7)$$

2.2 基于亚模函数的天线选择优化模型

定义: 对于给定集合 $Z = \{1, 2, \dots, z\}$, 若集合函数 $F: 2^Z \mapsto \mathbb{R}$, 或伪布尔函数 $F: \{0, 1\}^z \mapsto \mathbb{R}$, 对于 $\forall A, B \subseteq Z$, 函数 F 都满足

$$F(A) + F(B) \geq F(A \cap B) + F(A \cup B) \quad (8)$$

则称集合函数 F 为亚模函数^[21]。

性质: 如果集合函数 F 为亚模函数, 则对于 $\forall A \subseteq B \subseteq Z, \forall i \in Z \setminus B$, 满足

$$F(i|A) \geq F(i|B) \quad (9)$$

式中: $F(i|A) \triangleq F(A+i) - F(A)$ 为元素 i 在集合 A 上的边际增益, $A+i$ 表示集合 A 与单元 i 的并集。同样地, $Z \setminus B$ 表示 $Z - B$, 为集合 Z 和集合 B 的差集。亚模性是刻画边际收益递减性质的直观体现。对于集合函数而言, 亚模性表现为将单个元素添加至大集合所带来的函数值增益小于该元素添加至小集合的增益。

天线选择问题实际上是一个离散集合选择问题, 所选天线可看作是全部天线的一个子集, 而每一个子集的组合优化问题可以被描述为一个单调亚模函数最大化问题, 那么对目标函数的最大化求解过程也就是天线子集的寻优过程。因此, 本文以系统的容量表达式为目标亚模函数, 考虑基数约束下的目标函数的最大化问题, 然后构建一个以亚模函数为基础的天线选择优化模型:

$$F = C_{\text{select}} = \frac{1}{2} \ln \left[1 + \frac{(\lambda_n A_m \sum_{j=1}^{N_T} |v_{j,n}|)^2}{2\pi e \sigma^2} \right] \quad (10)$$

$$\max_{S \in \mathbf{Z}^{1 \times N_T}} C_{\text{select}}(\mathbf{S}), \mathbf{S} = \{s_i = 1\}_{i=1, \dots, N_T} \quad (11)$$

$$\text{s. t. } \mathbf{I}^T \mathbf{S} \leq L_t \quad (12)$$

式中: $\mathbf{S} = \{s_i = 1\}_{i=1, \dots, N_T}$ 表示选择向量 $\mathbf{S}$ 中被选中天线的索引; $\mathbf{I}^T \mathbf{S} \leq L_t$ 作为基数约束条件, 表示选择向量 $\mathbf{S}$ 中“1”元素的个数, 即选择的天线个数为 L_t , $L_t > 0$ 是正整数。例如 LED 的总数为 6, 选择天线数 $L_t = 4$, 则向量 $\mathbf{S}$ 中为 1 的元素个数等于 4; 若选中的为 LED₁、LED₃、LED₄ 和 LED₅, 则 $\mathbf{S} = \{1 \ 0 \ 1 \ 1 \ 1 \ 0\}$ 。

天线选择优化模型中, 目标函数的收益递减特性表现为随着选择子集中天线数目的增加, 后续选择的天线所提供的容量增益逐渐减少。下面则对天线选择优化模型中目标函数的亚模性进行证明。由

于 $\frac{A_m^2}{2\pi e \sigma^2}$ 是一个常数, 对目标函数的性质没有影响, 因此可以省略。

不失一般性, 令天线子集 n_1 对应集合 A , n_2 对应单个天线元素, 天线子集 n_3 对应由多个元素 i 组成的集合 $\{i\}$, 即 $n_1 + n_3$ 对应为集合 B , 因此 $F(i|A) = F(A+i) - F(A)$, $F(i|B) = F(B+i) - F(B)$ 可分别表示为

$$F(i|A) = C_{\text{select}}(n_1 + n_2) - C_{\text{select}}(n_1) = \frac{1}{2} \ln[1 + (\lambda_{n_1} \sum_{j=1}^{N_T} |v_{j,n_1}|)^2 + (\lambda_{n_2} \sum_{j=1}^{N_T} |v_{j,n_2}|)^2] - \frac{1}{2} \ln[1 + (\lambda_{n_1} \sum_{j=1}^{N_T} |v_{j,n_1}|)^2] \quad (13)$$

$$F(i|B) = C_{\text{select}}(n_1 + n_2 + n_3) - C_{\text{select}}(n_1 + n_3) = \frac{1}{2} \ln[1 + (\lambda_{n_1} \sum_{j=1}^{N_T} |v_{j,n_1}|)^2 + (\lambda_{n_2} \sum_{j=1}^{N_T} |v_{j,n_2}|)^2 + (\lambda_{n_3} \sum_{j=1}^{N_T} |v_{j,n_3}|)^2] - \frac{1}{2} \ln[1 + (\lambda_{n_1} \sum_{j=1}^{N_T} |v_{j,n_1}|)^2 + (\lambda_{n_3} \sum_{j=1}^{N_T} |v_{j,n_3}|)^2] \quad (14)$$

为了表达简便, 令 $K = (\lambda_{n_1} \sum_{j=1}^{N_T} |v_{j,n_1}|)^2$, $Q = (\lambda_{n_2} \sum_{j=1}^{N_T} |v_{j,n_2}|)^2$, $T = (\lambda_{n_3} \sum_{j=1}^{N_T} |v_{j,n_3}|)^2$, 进行替换后

$$F(i|A) = \frac{1}{2} [\ln(1+K+Q) - \ln(1+K)] = \frac{1}{2} \ln\left(1 + \frac{Q}{1+K}\right) \quad (15)$$

$$F(i|B) = \frac{1}{2} [\ln(1+K+Q+T) - \ln(1+K+T)] = \frac{1}{2} \ln\left(1 + \frac{Q}{1+K+T}\right) \quad (16)$$

$$\ln\left(1 + \frac{Q}{1+K}\right) \geq \ln\left(1 + \frac{Q}{1+K+T}\right) \quad (17)$$

通过上述的证明过程可得, 目标亚模函数满足 $F(i|A) \geq F(i|B)$, 具备亚模函数的单调亚模性。

2.3 基于容量最大化的天线选择算法

通过对以容量表达式为目标亚模函数的证明可得, 目标亚模函数满足收益递减性质。本文利用此性质设计了以容量最大化为目标的

天线选择算法, 算法的具体步骤如下:

步骤 1 输入原始信道矩阵 $\mathbf{H}$, 选择后信道矩阵 $\mathbf{H}_s$, 选择向量 $\mathbf{S}$, 选择天线数目 L_t , 循环次数 t 。

步骤 2 首先依次对原始信道矩阵 $\mathbf{H}$ 中的每一列 c 进行遍历; 将 t 时刻对应列 c_t 的增量表示为 $G_m(c_t) \triangleq \max_{c \in H^{(t)}} C_{\text{select}}(c|\mathbf{H}^{(t)})$ 。

步骤 3 如果 $G_m(c_t) > 0$ 且 $G_m(c_t) > G_m(c_{t-1})$, 则列 c_t 符合选择标准, 将其添加到上一次循环中的 $\mathbf{H}_s^{(t-1)}$, 即 $\mathbf{H}_s^{(t-1)} \cup [c_t] = \mathbf{H}_s^{(t)}$, 并从原始信道矩阵 $\mathbf{H}$ 中删除 c_t 列, 即 $\mathbf{H}^{(t-1)} \setminus [c_t] = \mathbf{H}^{(t)}$, 使 $\mathbf{H}_s$ 和 $\mathbf{H}$ 同时得到更新; 并将 c_t 列对应的 LED 作为被选天线的索引 i , 根据 i 的位置更新选择向量 $\mathbf{S}$, 即将 $\mathbf{S}_{t-1}$ 中第 i 个位置处元素置为 1 并赋值给 $\mathbf{S}$ 。否则, 将 $\mathbf{H}_s^{(t-1)}$ 赋值给 $\mathbf{H}_s$, 将 $\mathbf{S}_{t-1}$ 赋值给 $\mathbf{S}$, 继续寻找下一列, 目的是记录 c_t 列在 $\mathbf{H}$ 中的索引位置并在选择向量的相应位置进行更新。

步骤 4 当 $t < L_t$ 时, 说明当前选择的天线数量小于 L_t , 继续执行上一步, 且在执行完成时更新循环次数 $t = t + 1$; 否则, 此时 $t \geq L_t$, 表示已将满足条件的天线全部选出, 即 $\mathbf{H}_s^{(t-1)} = \mathbf{H}_s$, $\mathbf{S} = \mathbf{S}_{t-1}$, 结束循环并输出最终结果 $\mathbf{H}_s$ 以及 $\mathbf{S}$ 。

3 数值仿真与分析

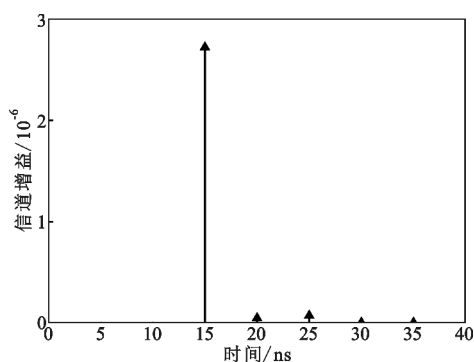
在长、宽和高分别为 6 m、6 m 和 4 m 的房间内, 安装 6 个垂直指向地面、高度为 3.5 m 的 LED。由 6 个垂直向上、距地面高度为 0.85 m 的 PD 形成矩形接收机阵列, 相邻 PD 的间隔为 0.1 m。将墙面在三维坐标方向上按间隔 0.1 m 划分为矩形微反射单元, 其他仿真参数如表 1 所示。IFFT/FFT 长度为 $N = 256$, LED 的线性工作区为 $V_{\min} = 0.1$ V, $V_{\max} = 1$ V, 直流偏置 $B_{\text{DC}} = 0.56$ V。

表 1 仿真参数

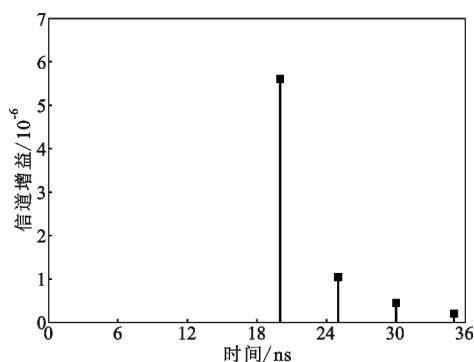
参数	取值
LED 的位置坐标 $(x, y, z)/\text{m}$	LED ₁ (2.0, 1.5, 3.5)
	LED ₂ (2.0, 3.0, 3.5)
	LED ₃ (2.0, 4.5, 3.5)
	LED ₄ (4.0, 1.5, 3.5)
	LED ₅ (4.0, 3.0, 3.5)
	LED ₆ (4.0, 4.5, 3.5)
光电转换系数 $\gamma/(\text{A/W})$	1
LED 调制带宽 $B_{\text{LED}}/\text{MHz}$	50
光电检测器视场角 $\psi_{\text{FOV}}/(\text{°})$	80
LED 的半功率角 $\theta_{1/2}/(\text{°})$	60
接收器 PD 的面积 A_{R}/cm^2	1
墙面反射率 ρ_i	0.8
每一个反射单元的面积 $\Delta A/\text{m}^2$	0.01

采用文献[19]提出的多径信道模型, 由于接收

到的光功率大约 90% 来自 LOS 分量和一次反射分量,所以本文仅考虑 LOS 信道和一次反射信道。图 3 所示为 PD 阵列的中心坐标位于室内两个典型位置 $(3, 3, 0.85)$ m 和 $(0.5, 0.5, 0.85)$ m 时, LED_4 和 PD_i 之间的多径信道增益。可以看出,当 PD 在房间中心时,LOS 路径延迟小,信道增益较强,路径矢量衰减快;当 PD 在房间墙角时,LOS 路径延迟大,信道冲激响应变小,多径信道路径矢量衰减较慢,多径分量相对较大。



(a) PD'坐标为 $(3, 3, 0.85)$ m



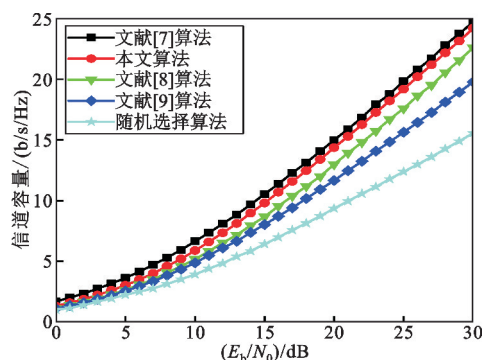
(b) PD''坐标为 $(0.5, 0.5, 0.85)$ m

图 3 LED_4 和 PD_i 之间的多径信道增益

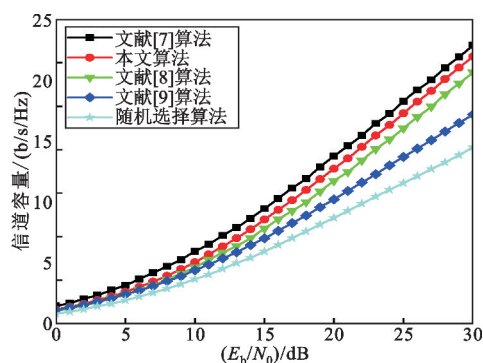
3.1 容量性能

图 4 所示为发射端选择天线数 $L_t = 4$ 时,在 4QAM 调制下,本文算法与文献[7]的穷举算法、文献[8]的范数选择算法、文献[9]的快速选择算法以及随机选择算法的信道容量比较。可以看出,无论是 MIMO-ACO-OFDM 系统还是 MIMO-DCO-OFDM 系统,信道容量与信噪比的增减趋势是一致的,而且本文算法与穷举最优选择算法的性能十分接近,且优于其他几种算法。在信噪比为 30 dB 时,在 MIMO-ACO-OFDM 系统中本文算法与穷举算法相差约 0.51 b/s/Hz,但与范数选择法、快速选择法以及

随机选择法相比则分别提升了 1.6 b/s/Hz、4.4 b/s/Hz、9.6 b/s/Hz;对于 MIMO-DCO-OFDM 系统而言,本文算法与穷举算法相差约 1.2 b/s/Hz,与范数选择法、快速选择法以及随机选择法相比则分别提升了 1.9 b/s/Hz、6.7 b/s/Hz、10.5 b/s/Hz。另外,MIMO-DCO-OFDM 系统由于自身频谱效率高,在信道容量方面具有更好的性能表现。



(a) MIMO-ACO-OFDM 系统

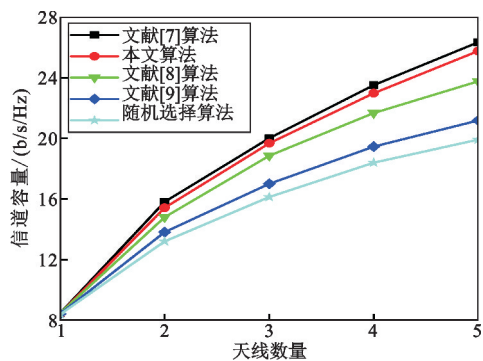


(b) MIMO-DCO-OFDM 系统

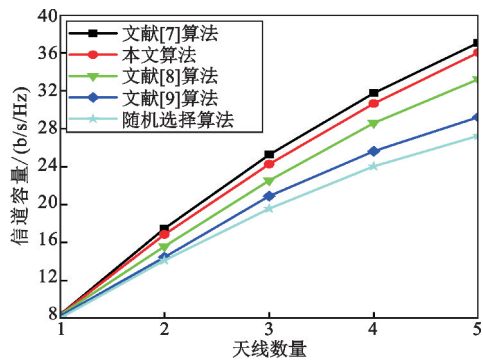
图 4 不同选择算法的容量性能比较

图 5 所示为本文算法与文献[7]的穷举算法、文献[8]的范数选择算法、文献[9]的快速选择算法以及随机选择算法在发射端选择不同天线数量时的信道容量比较。可以看出,随着发射端 L_t 数量的增多,系统的整体信道容量不断递增。然而,随着天线子集中被选天线数量的增加,后续选择的天线对于系统信道容量的增益呈现出逐渐递减的趋势。例如,对于本文选择算法的 MIMO-ACO-OFDM 系统,当选择的天线数 L_t 由 1 到 5 递增时,信道容量的增量分别为 6.98 b/s/Hz、4.26 b/s/Hz、3.30 b/s/Hz、2.68 b/s/Hz;MIMO-DCO-OFDM 系统中的增量变化分别为 8.48 b/s/Hz、7.43 b/s/Hz、5.38 b/s/Hz、4.28 b/s/Hz,明显在不断减小。这也与亚模函数所

具备的“收益递减效应”相符合,即对于优先选中的天线能给信道容量带来更大的增量表现,由此验证了基于亚模函数的天线选择模型和算法的正确性。



(a) MIMO-ACO-OFDM 系统



(b) MIMO-DCO-OFDM 系统

图 5 信道容量随天线数的变化曲线

3.2 复杂度分析

对于本文算法的复杂度, 设定 N_T 是发射端 LED 总数, L_t 是选择的天线个数, L 是将矩阵 $\mathbf{H}$ 分解后的子信道个数。整个选择过程需要调用一次遍历排序的过程和至多 L 次亚模函数的计算, 而遍历排序的过程又需要进行 $(N_T \times L_t)$ 次的乘法运算, 因此, 总的运算复杂度为 $O(N_T \times L_t \times L)$ 。本文算法与其他选择算法的复杂度比较如表 2 所示。相比于文献[7]的穷举最优算法, 本文算法避免了遍历全部组合时的冗余计算, 因此复杂度降低效果明显。

表 2 算法复杂度

选择算法	计算复杂度
文献[7]算法	$O(C_{N_T}^{L_t} \times N_T \times L_t^2)$
文献[8]算法	$O(N_T \times L_t)$
文献[9]算法	$O(N_T^2 \times L_t)$
随机选择算法	$O(1)$
本文算法	$O(N_T \times L_t \times L)$

图 6 所示为发射端 LED 数目 $N_T=6$ 时, 随着发射端选择天线数 L_t 的增加各选择算法复杂度的变化曲线。可以观察到随着 L_t 的增加, 文献[7]算法的复杂度增加幅度最大, 随机选择算法的复杂度最小。另外, 当 $L_t=4$ 时, 与文献[7]的穷举最优算法相比, 本文算法的复杂度降低了约 46.3%, 实现了选择效率的提高。

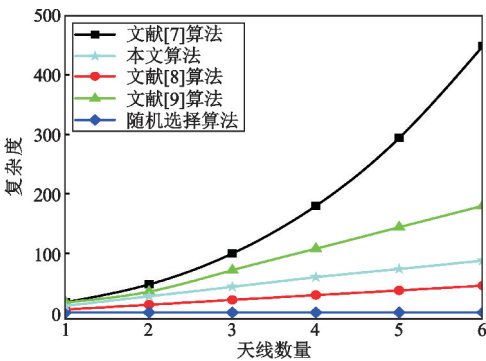
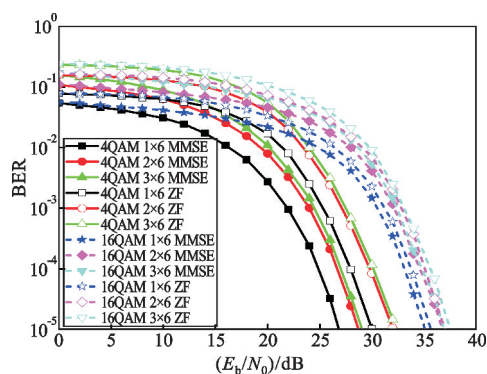


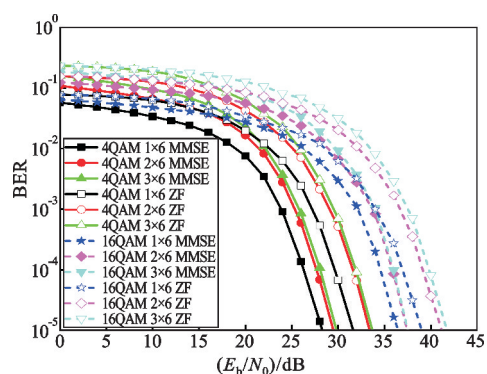
图 6 不同 L_t 时的算法复杂度对比

3.3 误码率性能

采用 ZF 和 MMSE 检测算法, 在 $PD'(3, 3, 0.85)m$ 和 $PD''(0.5, 0.5, 0.85)m$ 两个位置, 对 MIMO-ACO-OFDM 和 MIMO-DCO-OFDM 系统在 4QAM 和 16QAM 调制下的误码率性能进行了仿真分析, 结果如图 7 和图 8 所示。随着信噪比的增加, 误码率性能逐渐变好; 但随着所选天线数量的增加, 天线间的相互干扰及相关性增强, 误码率性能变差。例如, 4QAM 调制下采用 MMSE 和 ZF 检测算法时, $L_t=1$ 时 BER 性能表现最好; 当 $L_t \geq 2$ 时, BER 性能逐渐变差且差异越来越小。对于接收端采用两种不同检测算法的情况, MMSE 检测算法比 ZF 检测算法具有更好的 BER 性能, 且调制阶数越小, MMSE 比 ZF 越有优势。然而, 随着调制阶数的增大, BER 性能逐渐变差。另外, 当 PD 位于中心位置时, 接收信号功率大, 反射引起的多径信号较小, 因此性能较好。PD 位于室内边缘位置时, 与中心位置相比, 光源到探测器的直射信道增益减少, 接收信号变弱且多径增益相对较强, BER 性能变差。

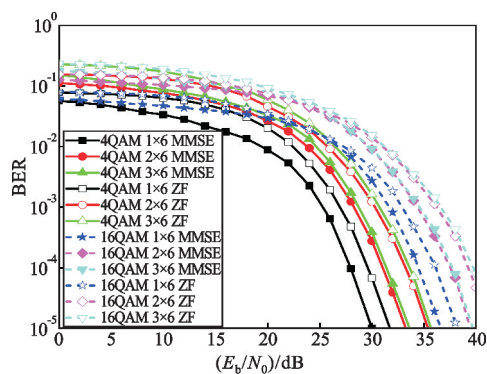


(a) MIMO-ACO-OFDM 系统

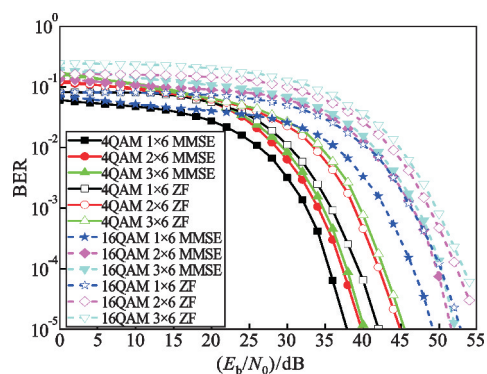


(b) MIMO-DCO-OFDM 系统

图7 PD 位于中心位置时系统的误码率性能



(a) MIMO-ACO-OFDM 系统



(b) MIMO-DCO-OFDM 系统

图8 PD 位于边缘位置时系统的误码率性能

4 结束语

本文通过将天线选择问题与亚模函数相结合,建立了基于亚模函数的天线选择优化模型。以下行信道的容量最大化为目标,证明了目标亚模函数的单调亚模性,为 VLC-MIMO 系统的天线选择问题求解提供了理论基础。又根据亚模函数的收益递减效应设计了发射端通信 LED 的选择算法,实现了天线子集的有效寻优。仿真结果表明,本文算法的容量性能与最优选择算法十分接近,同时还降低了算法的复杂度。因此,本文所提算法适用于解决室内 LED 布局完成的 VLC-MIMO 系统天线选择问题。

在后续的研究中,将针对室内 VLC 系统中多用户信道相关性较强的情形,对所提算法进行改进和优化,以进一步提升算法在该情形下选择时的性能。

参考文献:

- [1] 迟楠,王哲.可见光通信复用技术研究进展[J]. 光通信技术,2020,44(4):1-7.
- [2] DWIVEDY P, DIXIT V, KUMAR A. A survey on visible light communication for 6G: architecture, application and challenges [C]//2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications. Srinagar Garhwal: IEEE, 2023: 1-6.
- [3] FATH T, HAAS H. Performance comparison of MIMO techniques for optical wireless communications in indoor environments [J]. IEEE Transactions on Communications, 2013, 61(2): 733-742.
- [4] AKANDE K O, POPOOLA W O. MIMO techniques for carrierless amplitude and phase modulation in visible light communication [J]. IEEE Communications Letters, 2018, 22(5): 974-977.
- [5] LE N P, SAFAEI F, TRAN L C. Antenna selection strategies for MIMO-OFDM wireless systems: an energy efficiency perspective [J]. IEEE Transactions on Vehicular Technology, 2016, 65(4): 2048-2062.
- [6] YANG Y, YANG Y J, CHEN M Z, et al. Joint LED selection and precoding optimization for multiple-user multiple-cell VLC systems [J]. IEEE Internet of Things Journal, 2022, 9(8): 6003-6017.
- [7] GAO Y, VINCK H, KAISER T. Massive MIMO antenna selection: switching architectures, capacity bounds, and optimal antenna selection algorithms [J]. IEEE Transactions on Signal Processing, 2018, 66(5): 1346-1360.
- [8] KIM S. Efficient transmit antenna selection for receive spatial modulation-based massive MIMO [J]. IEEE Access, 2020, 8: 152034-152044.
- [9] GHARAVI-ALKHANSARI M, GERSHMAN A B. Fast antenna subset selection in MIMO systems [J]. IEEE

- Transactions on Signal Processing, 2004, 52(2): 339–347.
- [10] 冀笑伟, 李莉, 魏爽, 等. 大规模 MIMO 相关信道下的联合天线分组和天线选择[J]. 电讯技术, 2022, 62(5): 637–643.
- [11] ZHANG J, WANG J T, WANG Y C. Antenna selection in massive MIMO systems utilizing the submodular function [C]//2017 9th International Conference on Wireless Communications and Signal Processing. Nanjing: IEEE, 2017: 1–6.
- [12] VAZE R, GANAPATHY H. Sub-modularity and antenna selection in MIMO systems [J]. IEEE Communications Letters, 2012, 16(9): 1446–1449.
- [13] FENG Z, GUO C L, GHASSEMLOOY Z, et al. The spatial dimming scheme for the MU-MIMO-OFDM VLC system [J]. IEEE Photonics Journal, 2018, 10(5): 1–13.
- [14] WANG Z P, GUO C L, YANG Y, et al. Antenna selection based dimming scheme for indoor MIMO visible light communication systems utilizing multiple lamps [C]//2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications. Valencia: IEEE, 2016: 1–7.
- [15] WANG L, WANG C Y, CHI X F, et al. Optimizing SNR for indoor visible light communication via selecting communicating LEDs [J]. Optics Communications, 2017, 387: 174–181.
- [16] RAJASHEKAR R, HARI K V S, HANZO L. Antenna selection in spatial modulation systems [J]. IEEE Communications Letters, 2013, 17(3): 521–524.
- [17] 门宏志, 刘文龙, 王楠, 等. 空间调制系统低复杂度的天线选择算法[J]. 电子学报, 2016, 44(6): 1322–1327.
- [18] JIA K J, HAO L. Modeling of multipath channel and performance analysis of MIMO-DCO-OFDM system in visible light communications [J]. Chinese Journal of Electronics, 2019, 28(3): 630–639.
- [19] 贾科军, 郝莉, 余彩虹. 室内可见光通信多径信道建模及 MIMO-ACO-OFDM 系统性能分析[J]. 光学学报, 2016, 36(7): 49–60.
- [20] WANG J Y, DAI J X, GUAN R, et al. Channel capacity and receiver deployment optimization for multi-input multi-output visible light communications [J]. Optics Express, 2016, 24(12): 13060–13074.
- [21] KONAR A, SIDIROPOULOS N D. A simple and effective approach for transmit antenna selection in multiuser massive MIMO leveraging submodularity [J]. IEEE Transactions on Signal Processing, 2018, 66(18): 4869–4883.

作者简介:

贾科军 男, 1978 年生于陕西兴平, 博士, 教授, 主要研究方向为可见光通信、多载波调制技术。

贺耀民 男, 1998 年生于甘肃镇原, 硕士研究生, 主要研究方向为可见光通信 MIMO 系统的天线选择。

张芳芳 女, 1998 年生于甘肃庆城, 硕士研究生, 主要研究方向为可见光通信空间调制系统的天线选择。

蔺莹 女, 1981 年生于甘肃甘谷, 博士, 副教授, 主要研究方向为 NOMA 接入系统中的关键技术。

薛建彬 男, 1973 年生于甘肃会宁, 博士, 教授, 主要研究方向为无线通信理论与技术、边缘计算。

郝莉 女, 1971 年生于四川成都, 博士, 教授, 主要研究方向为高移动性无线通信系统、多址接入技术。

DOI:10.20079/j.issn.1001-893x.231208003

非地面通信网络多普勒频偏多级高精度估计方法*

李昌淼^{1,2}, 申 滨^{1,2}, 黄晓炯^{1,2}

(1. 重庆邮电大学 通信与信息工程学院, 重庆 400065; 2. 移动通信技术重庆市重点实验室, 重庆 400065)

摘 要:与地面蜂窝网络相比,非地面通信网络如低轨卫星通信场景下往往存在更高的多普勒频偏和较低的接收信噪比等问题,严重影响系统性能。为了提高同步性能,提出了一种多级频偏估计方法,以补偿由星地之间相对运动引起的多普勒频偏。首先,星历频偏预估计利用终端位置和星历信息进行粗略估计。其次,对于剩余频偏估计问题,将其等效地转化为离散傅里叶变换(Discrete Fourier Transform, DFT)插值迭代频率估计问题,并通过残余频偏粗估计和精估计分别完成。仿真结果表明,与传统的互相关方法及其他几种典型改进方法相比,所提方法具有更高的多普勒频偏估计精度,在归一化频偏为 4.487、信噪比为 0 dB 时,频偏估计的均方根误差可达到 10^{-2} ,有助于提高接入成功率和通信质量。

关键词:非地面通信网络;同步性能;多普勒频偏估计;插值迭代

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN927 文献标志码:A 文章编号:1001-893X(2025)03-0454-09

A High-precision Doppler Frequency Offset Estimation Method for Non-terrestrial Networks

LI Changmiao^{1,2}, SHEN Bin^{1,2}, HUANG Xiaojie^{1,2}

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Chongqing Key Laboratory of Mobile Communications Technology, Chongqing 400065, China)

Abstract: Compared with terrestrial cellular networks, non-terrestrial network (NTN) communication systems, for instance in the low Earth orbit satellite communication scenario, often have higher Doppler frequency shifts and lower signal-to-noise ratios, which seriously affect the system performance. To improve the synchronization performance, a frequency shift estimation method is proposed to compensate for the Doppler frequency induced by the relative motion between the satellite and the ground. Firstly, the pre-estimation of ephemeris shift uses the terminal position and ephemeris information for rough frequency compensation. Then, the residual frequency shift estimation problem is equivalently transformed into the discrete Fourier transform (DFT) iterative interpolation-based frequency estimation problem, which is completed by the coarse and fine estimation of the residual frequency shift, respectively. Simulation results show that compared with traditional cross-correlation methods and other typical methods, the proposed method has higher Doppler frequency offset estimation accuracy. When the normalized frequency offset is 4.487 and the signal-to-noise ratio is 0 dB, the root mean square error of frequency offset estimation can reach 10^{-2} , which helps to improve the user access success rate and the communication quality.

Key words: non-terrestrial networks; synchronization performance; Doppler frequency offset estimation; iterative interpolation

* 收稿日期:2023-12-08;修回日期:2024-03-11

基金项目:国家自然科学基金资助项目(62371082)

通信作者:李昌淼 Email:lichangmiao0725@163.com

0 引言

非地面网络(Non-terrestrial Network, NTN)具有覆盖范围广、传输距离远和通信质量稳定等优势,可以作为地面蜂窝网络的有效补充。典型的 NTN 包括卫星或无人机高空平台等,卫星又根据轨道高度分为低轨卫星、中轨卫星和高轨卫星,而低轨卫星相比于其他轨道卫星,其传输时延低,信道质量好,因此受到越来越多的重视^[1-3]。

当前 5G 及未来 6G 仍将使用正交频分复用(Orthogonal Frequency Division Multiplexing, OFDM)技术,它也成为星地综合网络的主要调制方案^[4-5]。然而基于 OFDM 的无线通信系统对载波频偏(Carrier Frequency Offset, CFO)非常敏感,可能导致频率同步误差,从而使得通信性能降低^[6-7]。

在 NTN 系统中, CFO 主要由发射机和接收机两端的相对运动或两端的晶振误差引起。过去 20 年里,业界和学术界对 OFDM 系统的频率同步问题进行了大量研究,然而这些方法主要针对地面蜂窝网络,在 NTN 场景中受限于高频偏和低信噪比等问题^[4],这些同步方法的性能会受到不可避免的限制。

目前,基于地面蜂窝网络的同步方法根据是否需要额外数据参与,大体可以分为两类,包括基于非数据辅助的同步方法^[7-11]和数据辅助^[12-15]的同步方法。基于非数据辅助的同步方法通常利用同步序列或 OFDM 符号的循环前缀(Cyclic Prefix, CP)进行自相关或互相关进行时频估计,其中最简单的方法是利用 OFDM 中的 CP 和数据间的相关性进行同步。这类方法具有高数据传输效率的特点,但估计精度较差,容易受到噪声的影响,频偏估计范围也较小。文献[8]通过利用同步信号序列的周期移位版本之间的自相关特性实现了较为精确的同步。为了减小延迟或同步信号的数据长度,文献[9]提出了一种基于过采样最大似然估计器的 CFO 估计方法,但该方法要求发射信号使用虚拟子载波,同时接收机的过采样会增加计算复杂度。文献[10]使用两个不同的 m 序列生成辅助同步序列,减小了同步的复杂度,但是同步准确性有所降低。

基于数据辅助的同步方法具有同步精度高和数据传输效率较低等特点。文献[13]提出利用两个完全相同的 PN 序列构成同步序列,接收端利用重复序列间的自相关性完成同步过程。文献[15]通过使用所提出的一种由 Zadoff-Chu(ZC)序列及其修

正序列中心连接对称的主同步信号序列实现了较为准确的同步。然而现有的大部分数据辅助方法要求发射信号具有某些特定特征,即需要修改发射信号的结构,具有一定程度上的实现困难。

针对低轨卫星系统的频率同步误差,在加性高斯白噪声信道中要求误差小于子载波间隔的 4%^[4]。当频偏误差处于该范围时,因频率偏移导致的系统性能损失才可以忽略,而在 700 km 圆形轨道上的低轨卫星系统中,最大的 CFO 可达到 70 kHz,这远远超过地面蜂窝网络中的 CFO。目前,还缺乏充分的研究结论以确定与无线空口技术(New Radio, NR)兼容的频率同步方法是否适用于低轨卫星系统。此外, NTN 通信链路的信噪比一般明显低于地面网络,可能超过 5 dB 以上的程度,而目前大多数 NR 兼容的同步方法在如此低的信噪比情况下无法提供精确的频偏估计。

基于上述分析,本文提出一种多级频偏同步方法,充分利用了同步序列的特征以及卫星系统信道的特点。实验结果证明本文所提频偏估计方法具备更精确的同步性能。

1 同步信号和系统模型

1.1 3GPP 同步信号

根据第三代合作伙伴计划(The 3rd Generation Partnership Project, 3GPP)标准, OFDM 系统中一帧持续时间为 10 ms, 包含 10 个子帧, 每个子帧的持续时间为 1 ms。为了辅助终端设备进行帧定时同步和频率同步, 定义了两种同步参考信号, 分别是主同步信号(Primary Synchronization Signal, PSS)和次同步信号(Secondary Synchronization Signal, SSS), 其中 PSS 分别在时隙 0 的符号 2 和符号 8 中传输。

本文假设所使用 PSS 序列由频域 ZC 序列生成, 可以表示为

$$S(n) = \begin{cases} e^{-j\frac{\pi u n(n+1)}{N_{zc}}}, & n=0, 1, \dots, \frac{N_{zc}-3}{2} \\ e^{-j\frac{\pi u (n+1)(n+2)}{N_{zc}}}, & n=\frac{N_{zc}-3}{2}+1, \dots, N_{zc}-2 \end{cases} \quad (1)$$

式中: u 表示 ZC 序列的根索引, 其值为 25、29 和 34; N_{zc} 表示使用的 ZC 序列长度为 128。根据上式可以得到 ZC 序列的频域序列特性为

$$S(k) = S(N_{zc}-k), k=0, 1, \dots, \frac{N_{zc}-3}{2} \quad (2)$$

式中: $S(k)$ 表示在第 k 个子载波上调制的频域 PSS

信号。ZC 序列的时域特性如下:

特性 1: 时域对称特性

$$s(n) = s(N_{zc} - n), n = 0, 1, \dots, \frac{N_{zc} - 3}{2} \quad (3)$$

特性 2: 恒包络特性

$$|s(n)|^2 = 1, n = 0, 1, \dots, N_{zc} - 1 \quad (4)$$

为了实现低轨卫星系统与蜂窝网络系统的融合, NTN 系统将采用与蜂窝网络通信相同的同步信号^[2]。本文采用上述 PSS 序列作为 NTN 系统的同步信号序列, 所提方法主要是利用 ZC 序列的两个特性和卫星信道的特征来估计 NTN 系统的多普勒频偏。

1.2 NTN 系统模型

根据 3GPP R17 协议, 在 NTN 通信中, 地面设置信关站作为网关, 连接到 5G 核心网络, 从而实现卫星直接与终端相连。如图 1 所示, NTN 通信系统通常包括以下组成部分: 网关, 即 NTN 系统与公共数据网络之间的关键接口; 馈电链路, 连接网关与卫星或其他空中载体平台之间的通信链路; 服务链路, 连接 NTN 终端与卫星或其他空中载体平台之间的通信链路。在服务链路上, 多普勒频偏补偿由终端用户执行, 而在馈电链路上的频偏管理则由网络实现^[3]。

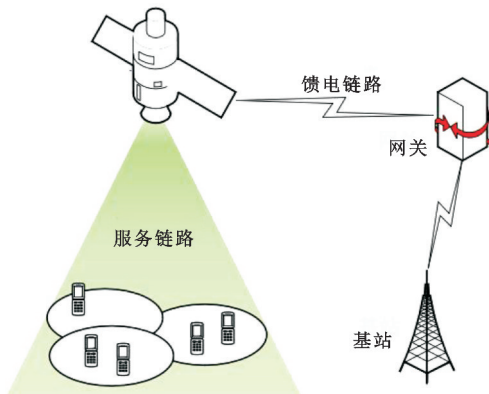


图 1 NTN 通信系统架构

设 $S(k)$ 为发射的同步信号, 经过快速傅里叶逆变换 (Inverse Fast Fourier Transform, IFFT) 之后, PSS 可以表示为

$$s(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S(k) e^{j2\pi kn/N}, 0 \leq n \leq N-1 \quad (5)$$

式中: N 表示 FFT 的序列长度。则接收信号为

$$y(n) = s(n)h(n)e^{j2\pi\epsilon n/N} + \omega(n) \quad (6)$$

式中: $h(n)$ 表示信道响应; $\omega(n)$ 为零均值方差为 δ^2

的加性高斯白噪声; ϵ 是关于子载波间隔的归一化频偏。

2 多级频偏估计方法

如图 2 所示, 本文提出一种多级频偏估计方法, 将频偏估计分为星历频偏预估计、残余频偏粗估计和残余频偏精估计三级处理模块分别完成, 星历频偏预估计利用星历数据和终端位置进行估计, 残余频偏粗估计和精估计通过对接收同步信号与本地参考信号的相关函数的离散傅里叶变换系数进行插值并以迭代方式获取, 最终实现星地之间服务链路中多普勒频偏的高精度估计。

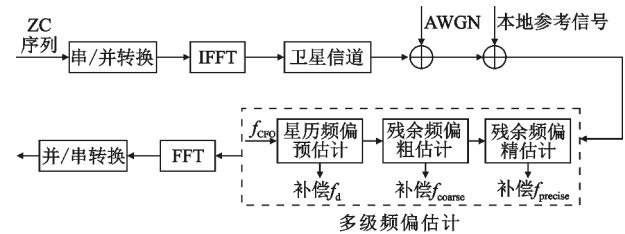


图 2 多级频偏估计模型

因此, 本文提出的多级频偏同步方法可以表示为

$$f_{CFO} = f_d + f_{coarse} + f_{precise} \quad (7)$$

式中: f_{CFO} 为星地之间服务链路的多普勒频偏; f_d 、 f_{coarse} 和 $f_{precise}$ 分别表示星历频偏预估计值、残余频偏粗估计值和残余频偏精估计值。

2.1 星历频偏预估计

针对超过一个子载波间隔的归一化多普勒频偏, 可以利用星历数据和终端地理位置进行频偏预估计。为满足终端设备在接入网络前补偿服务链路的多普勒频偏, 卫星会广播与其位置和速度相对应的星历表, 终端设备需要配备全球导航卫星系统模块, 用于在接入网络之前确定自己的位置^[1]。

为了求解该多普勒预估计值, 需要计算卫星与地面终端之间的相对速度 V_T 。由于低轨卫星与终端之间的相对速度在一段时间内不断变化, 因此会产生多普勒频偏变化率, 进而导致多普勒频偏发生快速变化。本文假设星地之间时间同步是准确的, 如果时间同步不准确或者星历信息存在误差, 可能会引入一定的误差, 这将反映在多普勒频偏变化率中。

图 3 为频偏预估计模型, 坐标系中心为地球中心 O 点, S 点表示卫星的位置, Y 点是在任意时刻 t

卫星与地心连接在地表上的投影, d 表示卫星和终端之间的距离, 仰角定义为 φ , α 表示地心角, A 表示地面终端, 假设地球半径 $r_E = 6\,371\text{ km}$ 。

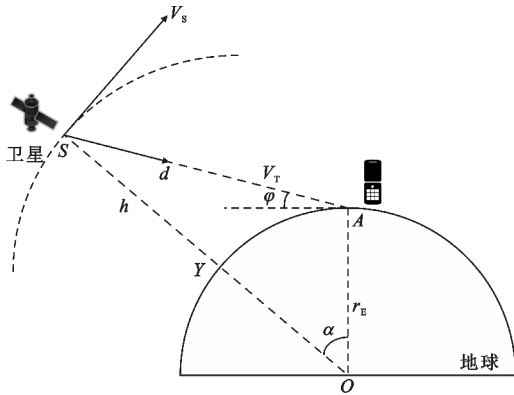


图 3 频偏预估计模型

根据余弦定律, 可以计算卫星与终端设备之间的距离为

$$d = \sqrt{r_E^2 + r^2 - 2r_E r \cos \alpha} \quad (8)$$

式中: $r = r_E + h$, h 为低轨卫星的轨道高度。为了得到卫星与地面终端之间的相对速度, 函数 d 对变量 α 求偏导得

$$V_T = \frac{r_E \cdot r \cdot \sin \alpha}{\sqrt{r_E^2 + r^2 - 2r_E \cdot r \cdot \cos \alpha}} \frac{d\alpha}{dt} \quad (9)$$

式中: V_T 为卫星与地面终端之间的相对速度; $d\alpha/dt$ 表示卫星的角速度, 用 ω_s 来代替。根据牛顿万有引力定律和牛顿第三定律可得, 卫星的角速度为

$$\omega_s = \frac{d\alpha}{dt} = \sqrt{\frac{GM}{r^3}} \quad (10)$$

式中: G 和 M 分别表示地球的引力常数和地球的质量。根据如下多普勒频偏公式:

$$f_d = \frac{V_T}{\lambda} \cos \varphi = f_c \frac{V_T}{c} \cos \varphi = f_m \cos \varphi \quad (11)$$

式中: f_c 为载波频率; c 为光速; f_m 表示最大多普勒频偏。式(9)和式(10)通过代入式(11)后进行代数运算可得多普勒频偏 f_d 与仰角 φ 的关系为

$$f_d = \frac{r_E r \sin \left\{ \arccos \left[\frac{r_E}{r} \cos \varphi \right] - \varphi \right\} \cos \varphi}{\sqrt{r_E^2 + r^2 - 2r_E r \cos \left\{ \arccos \left[\frac{r_E}{r} \cos \varphi \right] - \varphi \right\}}} \frac{\omega_s}{c} f_c \quad (12)$$

对上述方程进行求导, 可以得到多普勒频偏的一阶变化率为

$$f_d' = \frac{f_c}{c} \cdot \frac{dV_T}{dt} = \frac{r_E r \cos \left[\arccos \left(\frac{r_E}{r} \cos \varphi \right) - \varphi \right] \frac{\omega_s^2}{c} f_c}{\sqrt{r_E^2 + r^2 - 2r_E r \cos \left[\arccos \left(\frac{r_E}{r} \cos \varphi \right) - \varphi \right]}} - \frac{\sin \left[\arccos \left(\frac{r_E}{r} \cos \varphi \right) - \varphi \right] \frac{\omega_s^2}{c} f_c}{2 \sqrt{r_E^2 + r^2 - 2r_E r \cos \left[\arccos \left(\frac{r_E}{r} \cos \varphi \right) - \varphi \right]}} \quad (13)$$

根据上面分析, 通过利用星历数据和终端的地理位置, 可以计算出多普勒频偏以及多普勒频偏的变化率, 从而实现对多普勒频偏预估计值的求解与补偿。

2.2 残余频偏估计

本文提出了一种针对残余频偏的新型精细估计方法, 将频偏估计问题等效转化为 DFT 插值迭代等效频率估计问题。通过对移位的 DFT 系数进行插值, 并通过迭代处理方式以获得残余频偏的估计。

由于主同步信号 $S(n)$ 在接收机端是已知的, 因此可以获得接收信号与本地参考信号的互相关函数, 如下所示:

$$c(n) = y(n) s^*(n) = h(n) |s(n)|^2 e^{j2\pi \varepsilon_r n/N} + \omega(n) s^*(n) \quad (14)$$

式中: ε_r 表示补偿星历频偏 f_d 后的剩余归一化频偏, 即 $\varepsilon_r = \varepsilon - f_d$ 。

根据文献[16]可知, 在草原、湿地、海洋或沙漠等空地上的信道, 多径功率与视距功率之比可低至 -12 dB, 因此卫星与终端设备之间的信道可以建模为单径信道。所以这里定义卫星信道响应为 $h(n) = A e^{j\theta}$, $f_0 = \varepsilon_r/N$, 其中, A 表示信号的幅度变化, θ 表示信号的相位偏移。由于本文只考虑多普勒频偏的估计, 因此幅度变化与相位偏移忽略不计。根据式(6)可以化简得

$$c(n) = A e^{j(2\pi f_0 n + \theta)} + \bar{\omega}(n) \quad (15)$$

显然 $\bar{\omega}(n)$ 可以看作与 $\omega(n)$ 具有相同方差的加性高斯白噪声。基于该式, 可以将载波频偏的估计问题等效转化为频率为 f_0 的等效频率估计问题。

在频率估计领域, 已经有部分研究文献探讨了通过对离散傅里叶变换系数进行插值来计算信号频率的方法, 但是目前关于 NTN 系统中频偏估计的大部分研究文献使用地面通信网络中基于互相关及其他改进的方法, 该类方法暂未应用到 NTN 通信中的

多普勒频偏估计中。目前存在的非迭代方法通常依赖于峰值的离散傅里叶变换系数及其邻域的幅度值:文献[17]中提出的 A&M 算法分别使用 DFT 系数偏移 $\pm \frac{1}{2}$ 的两个系数进行插值;文献[18]中研究了最优偏移插值点 q 的选择。这些方法具有渐近无偏性和高精度的特点,在信噪比高于特定阈值时具有较好的估计性能。然而,这些方法的局限性在于它们的频率估计范围限定在 $[-f_s, f_s]$ 之间,无法提供足够的范围来估计大于 1 个子载波间隔的频偏。假设子载波间隔 f_s 为 15 kHz,根据上面分析,NTN 低轨卫星对应的最大多普勒频偏可高达近 70 kHz,则对于 NTN 通信系统,需要估计大于 5 个子载波间隔的多普勒频偏,因此此类方法刚好适用于进行星历频偏预补偿后的残余频偏估计。

首先,通过寻找离散傅里叶变换系数的频谱峰值索引 k_p 来进行初始化:

$$k_p = \arg\max_k P(k) \quad (16)$$

$$P(k) = \frac{1}{N} \left| \sum_{n=0}^{N-1} c(n) \exp \left[-j \frac{2\pi}{N} nk \right] \right|^2 \quad (17)$$

式中: $P(k)$ 为 DFT 的频谱图。

然后可以得到残余频偏粗估计的估计值为 $f_{\text{coarse}} = k_p \Delta f N$, $\Delta f = f_s / N$ 。

本文残余频偏精估计采用 DFT 移位插值迭代方法。该方法利用 A&M 算法作为残余频偏精估计的第一次迭代值,分别使用 DFT 系数偏移 $\pm 1/2$ 的两个系数进行插值,即对离散傅里叶变换系数的频谱图进行两次插值,从而提高估计性能。

因此,第一次迭代可以通过离散傅里叶变换系数偏移 $\pm 1/2$ 的两个系数进行插值获得一个初始的精细频偏估计值为

$$\hat{\delta}_1 = \frac{N}{2\pi} \arcsin \left(\sin \left(\frac{\pi}{N} \right) \operatorname{Re} \left\{ \frac{C_{0.5} + C_{-0.5}}{C_{0.5} - C_{-0.5}} \right\} \right) \quad (18)$$

$$C_{\pm 0.5} = \sum_{n=0}^{N-1} c(n) \exp \left[-j \frac{2\pi}{N} n (k_p \pm 0.5) \right] \quad (19)$$

式中: $\operatorname{Re}(\cdot)$ 表示实部; $C_{\pm 0.5}$ 表示相关信号的 DFT 系数。

然后使用初始的精细估计值 $\hat{\delta}_1$ 进行第二次迭代获取更加精细的残余频率估计值:

$$\delta_2 = \frac{1}{x(q)} \times \operatorname{Re} \left\{ \frac{C_{+q} + C_{-q}}{C_{+q} - C_{-q}} \right\} + \hat{\delta}_1 \quad (20)$$

其中,设相关信号的 DFT 的系数 $C_{\pm q}$ 为 $C_{\pm q} = \sum_{n=0}^{N-1} c(n) \exp \left[-j \frac{2\pi}{N} n (k_p + \hat{\delta}_1 \pm q) \right]$ 。

关于 $x(q)$ 的推导如下:首先定义 $\Delta_i = \delta - \delta_{i-1}$,假设 Δ_i 为第 $i-1$ 次迭代时的偏差,则第 i 次移位为 q 的 DFT 系数为

$$C_{\pm q} = A e^{j\theta} \frac{1 - \exp[j2\pi(\Delta_i \pm q)]}{1 - \exp[j\frac{2\pi}{N}(\Delta_i \pm q)]} = \frac{A N e^{j\theta}}{j2\pi} \times \frac{1 - \exp[j2\pi(\Delta_i \pm q)]}{(\Delta_i \pm q)} \quad (21)$$

由此可推出

$$C_{+q} - C_{-q} = \frac{A N e^{j\theta}}{j2\pi} \left[\frac{1 - e^{j2\pi(\Delta_i - q)}}{(\Delta_i - q)} - \frac{1 - e^{j2\pi(\Delta_i + q)}}{(\Delta_i + q)} \right] = \frac{A N e^{j\theta}}{j\pi} \times \frac{q + e^{j2\pi\Delta_i} (j\Delta_i \sin(2\pi q) - q \cos(2\pi q))}{(\Delta_i^2 - q^2)} \quad (22)$$

$$C_{+q} + C_{-q} = \frac{A N e^{j\theta}}{j\pi} \frac{\Delta_i - e^{j2\pi\Delta_i} (\Delta_i \cos(2\pi q) - j q \sin(2\pi q))}{(\Delta_i^2 - q^2)} \quad (23)$$

定义 C_{+q} 和 C_{-q} 和与差的比值为 β_i ,化简得

$$\beta_i = \frac{C_{+q} - C_{-q}}{C_{+q} + C_{-q}} = \frac{q + e^{j2\pi\Delta_i} [j\Delta_i \sin(2\pi q) - q \cos(2\pi q)]}{\Delta_i - e^{j2\pi\Delta_i} [\Delta_i \cos(2\pi q) - j q \sin(2\pi q)]} \quad (24)$$

其中,该估计只涉及 β_i 的实部,且 $\operatorname{Re}(\beta_i)$ 相当于 Δ_i 的线性函数,将 $\operatorname{Re}(\beta_i)$ 展开为泰勒级数得

$$\operatorname{Re}\{\beta_i\} = x(q) \Delta_i + o(\Delta_i^3) \quad (25)$$

$$x(q) = \frac{1 - \pi q \cot(\pi q)}{q \cos^2(\pi)} \quad (26)$$

因此,可以计算出估计的残余频偏精估计如下:

$$f_{\text{precise}} = \delta_2 \cdot \Delta f \cdot N \quad (27)$$

则残余频偏粗估计与残余频偏精估计的和,即残余频偏估计值为

$$f_0 = (k_p + \delta_2) \Delta f \cdot N \quad (28)$$

用 m 表示观测次数,结合公式(7)、(12)和(28)可将残余频偏估计的均方根误差表示为

$$E_{\text{RMS}} = \sqrt{\frac{1}{m} \sum_{i=1}^m (f_{\text{CFO}} - f_d - f_0)^2} \quad (29)$$

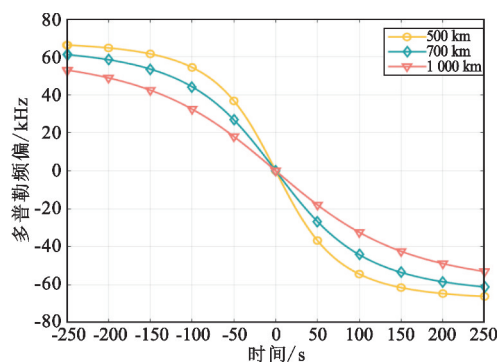
3 仿真结果

在本节中,对于星历频偏预估计,根据表 1 中提供的低轨卫星移动通信系统场景参数,得到了不同轨道高度下的多普勒频偏和多普勒频偏变化率的数值。

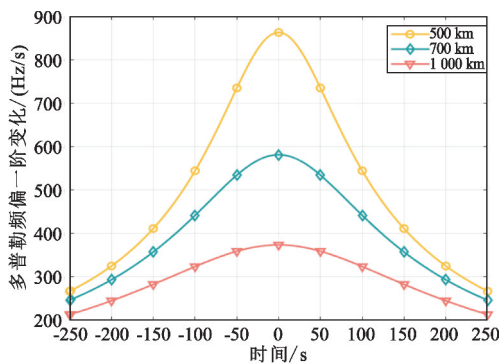
表 1 低轨卫星通信系统场景参数

参数	数值
卫星轨道高度 h/km	500, 700, 1 000
载波频率 f_c/GHz	3
卫星轨道离心率 $e/(\circ)$	0
地球半径 r_E/km	6 731
卫星轨道倾角 $i/(\circ)$	53
仰角 $\varphi/(\circ)$	60

仿真模拟了几种不同终端仰角的情况,并将其用作多普勒频偏的预估计。图 4 展示了不同轨道高度下卫星移动过程中多普勒频偏值和多普勒频偏变化率的仿真结果,假设地面用户端为近似静止,低轨卫星在过顶的过程中先靠近地面接收端,而后逐渐远离地面接收端。在图 4 中,多普勒频偏的值先减小再增大,当低轨卫星位于地面终端的正上方时,由于相对速度最小,多普勒频偏达到最小值。此外,多普勒频偏变化率的值在这个过程中由小逐渐变大再减小,当低轨卫星位于地面终端的正上方时,多普勒频偏变化率的值最大。



(a) 多普勒频偏的范围



(b) 多普勒频偏变化率的范围

图 4 不同轨道高度下的仿真结果

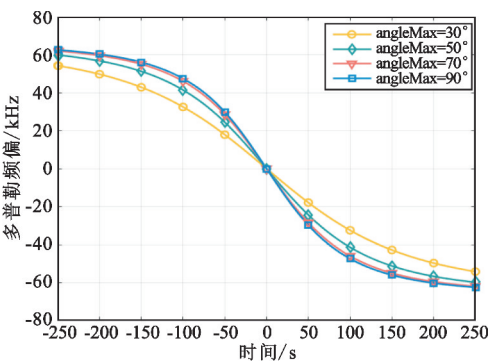
仿真结果表明,在低轨卫星通信系统中,最大多

普勒频偏可高达近 70 kHz,最大多普勒频偏一阶变化率高达近 900 Hz/s。

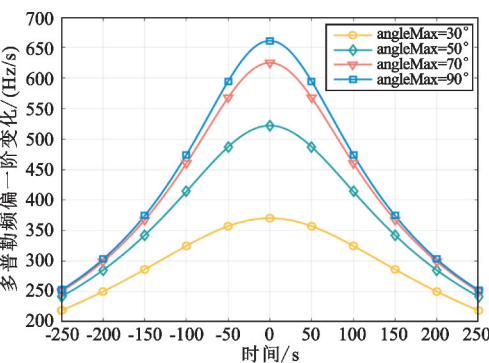
本文假设星地之间时间同步是准确的,可以利用终端位置和星历信息对由相对运动引起的多普勒频偏中大于一个子载波间隔的多普勒频偏进行初步估计补偿。

此外,不同终端和低轨卫星的相对位置对多普勒频偏曲线的影响非常敏感。为了探究这种影响,本文对处在不同位置上的用户进行了测试,分别测量了它们的多普勒频偏及其多普勒频偏变化率。在仿真中,卫星观测到的用户的最大仰角代表不同的用户位置。

如图 5 所示,较小的观测仰角对应着较小的多普勒频偏和多普勒频偏变化率。这是因为观测仰角越小,径向速度和加速度分量越小。由于不同终端所处的位置不同,其多普勒频偏和多普勒频偏变化率也会有所差异。此外,图 5 中曲线的变化趋势与图 4 相似,均呈现出低轨卫星在过顶过程中的频偏变化情况。



(a) 不同用户多普勒频偏的范围



(b) 不同用户多普勒频偏变化率的范围

图 5 卫星轨道高度为 700 km 下的仿真结果

由于终端设备的移动,星历频偏预估计所估计的多普勒频偏并不完全准确,由表 2 的星历补偿效

果可知,星历频偏预估计能够补偿大部分的频率偏移,但误差范围很小,而剩余的频偏则需要通过残余频偏估计来进行补偿。

表 2 星历估计效果表

SNR/dB	$f_{\text{CF0}}/\text{kHz}$	f_{d}/kHz
-10	67.3	66.9
-5	67.3	67.3
0	67.3	67.1
5	67.3	67.7
10	67.3	67.6
15	67.3	66.8
20	67.3	67.2

针对残余频偏估计的性能仿真中,在不同信噪比下与互相关法^[8]、对称相关法^[6]、Moose 法^[11]、A&M 算法^[17]等多个现有典型方案进行了比较。本文采用了长度为 128 的 ZC 序列,并设置 FFT/IFFT 大小为 1 024,最优偏移插值点 $q = 1/\sqrt[3]{N}$ 。所提残余频偏粗精估计法命名为迭代 A&M 法。此外,假设子载波间隔 f_s 为 15 kHz,CP 类型为正常 CP。设置归一化的残余频偏范围为 $(0, 1]$,高于低轨卫星通信系统的最大多普勒频偏值。

图 6 中给出了剩余归一化频偏为 0.1 时不同方法的多普勒频偏估计均方根误差。仿真结果表明,当信噪比大于 -5 dB 时,本文提出的方法相比其他对比方法具有更好的性能。本文所提方法具有较高的频偏估计精度,因此能够解决目前大多数 NR 兼容的同步方法在低信噪比情况下无法提供精确频偏估计的问题。

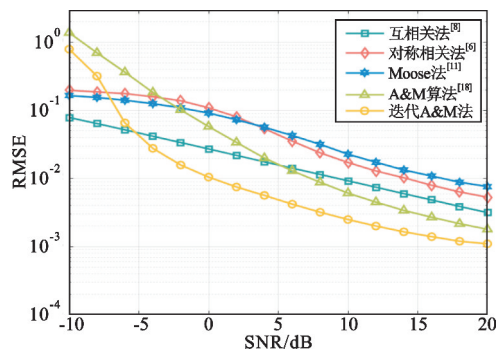


图 6 不同信噪比下的剩余频偏估计均方根误差

当信噪比小于 -5 dB 时,本文所提频偏估计方法的性能下降。这是由于所提方法基于 FFT 的寻峰,在信噪比较低的情况下会出现更多的假峰,从而

导致估计误差。在信噪比较低时,如信噪比低于 -6 dB 时,互相关法的均方根误差最低。这是因为该方法使用了整个 PSS 作为有效序列长度,是所有方法中最长的。然而,该方法未充分利用 ZC 序列的特性,因此在高信噪比条件下性能较差。当信噪比高于 4 dB 时,对称相关法的均方误差低于 Moose 方法,表现出更好的性能。这是因为对称相关法使用的序列长度比 Moose 法使用的序列长度更长。对称相关法的有效长度为 PSS 序列的一半,而 Moose 方法的有效序列长度仅为 CP 的长度。然而,由于对称相关法未考虑信道的影响,在低信噪比条件下性能低于 Moose 方法。因此,从图中可以看出,权衡 ZC 序列长度和信道响应可能进一步提高同步性能。此外,A&M 算法并未第二次迭代,与本文所用方法相比精度较低。

图 7 为剩余归一化频偏等于 1.0 时的性能估计仿真图,结果显示所提方法在剩余归一化频偏为 1.0 时与图 6 中剩余归一化频偏为 0.1 时的性能表现相近,即当信噪比大于 -5 dB 时,本文提出的方法相比其他对比方法具有更好的性能,表现最优,而互相关法、对称相关法和 Moose 法在剩余归一化频偏为一个子载波间隔时失效,即地面网络使用的传统方法在 NTN 系统具有大频偏的场景下无法进行有效的频偏估计。这验证了本文针对剩余频偏粗、精估计提出的方法在剩余频偏范围为 $(0, 1]$ 内具有相同的性能,能够有效解决 NTN 系统频偏较大及信噪比较低时传统地面频偏同步方法性能不佳的问题。

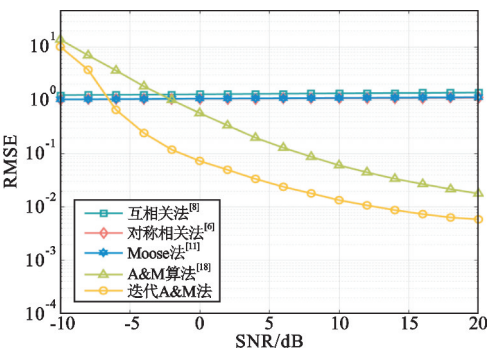


图 7 不同信噪比下的剩余频偏估计均方根误差

图 8 为这些方法在不同残余频偏下的均方根误差,设置 SNR 为 5 dB,符合 NTN 通信链路下的一般信噪比。从图 8 可以看出,本文方法在归一化频偏为 $(0, 1]$ 的范围内具有最优性能。此外,由于其他方法各自的频偏估计范围有限,无法准确估计多普

勒频偏,因此不适用于 NTN 通信系统的频偏估计。

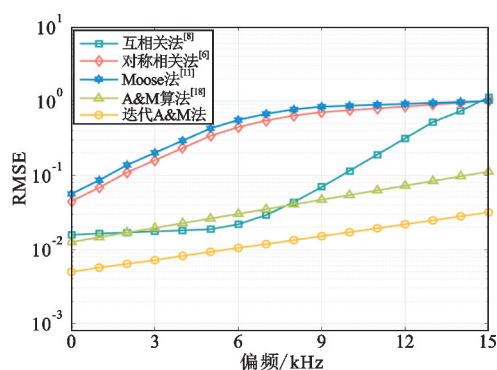


图 8 固定信噪比下的剩余频偏估计均方根误差

图 9 展示了在不同信噪比下进行的多级联合频偏估计的均方根误差。在仿真中,设频偏为表 2 所示的 67.3 kHz,即归一化频偏为 4.487,这符合低轨卫星通信系统中常见的频偏水平。此外,文献[16]中提到低轨卫星通信系统中的信噪比通常约为 5 dB,在雨天及阴天环境下系统信噪比可降低至 0 dB。为体现所提方法具有优越特性,多级联合频偏估计仿真选择将信噪比范围设置为 0~10 dB。仿真结果表明,在归一化频偏为 4.487 时,本文所提联合估计方法有更精确的性能,而互相关法、对称相关法、Moose 法及 A&M 算法由于难以有效估计超过一个子载波间隔的多普勒频偏,因此均无法进行有效的频偏估计。此外,文献[19]所提的联合 CP 的估计算法在各种信道条件下表现均一,因此随着信噪比增加,其性能保持稳定,但由于该方法基于 CP 相关的改进,其精度较低。文献[16]所提出的补零插值相关法将频偏估计分为整数频偏与小数频偏两部分分别处理。其整数频偏处理基于互相关,小数频偏处理基于补零且一次插值,其精度与本文所提方法相比略低。

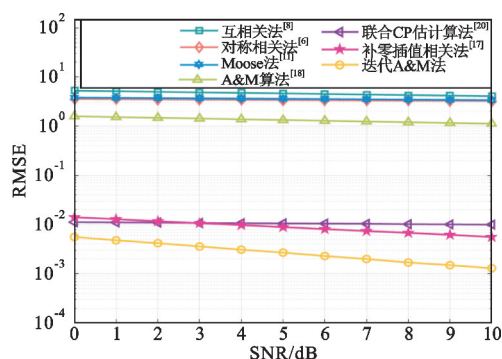


图 9 不同信噪比下的多级频偏估计均方根误差

总体而言,本文所提的联合估计方法在频偏估计方面具有准确性能,这为低轨卫星通信系统中的频偏估计提供了一种有效的解决方案。

4 结束语

针对 NTN 通信系统具有频偏大和信噪比较低等特点,本文提出了一种多级频偏估计方法。该方法首先基于星历信息和终端位置进行预估计,然后对于残余频偏通过对移位的 DFT 系数进行插值并进行迭代,获得残余频偏的估计值。性能仿真表明,该方法具有较高的估计精度且适用于实现卫星系统与地面蜂窝网络的进一步集成,有助于提高接入成功率。但是,本文未纳入同步过程中的时延估计。在未来的研究中,可以考虑加入时延估计,以实现同步过程的完整性。

参考文献:

- [1] 苏昭阳,刘留,艾渤,等.面向低轨卫星的星地信道模型综述[J].电子与信息学报,2024,46(5):1684-1702.
- [2] 徐常志,靳一,李立,等.面向 6G 的星地融合无线传输技术[J].电子与信息学报,2021,43(1):28-36.
- [3] 陈山枝.关于低轨卫星通信的分析及我国的发展建议[J].电信科学,2020,36(6):1-13.
- [4] WANG W J, TONG Y S, LI L X, et al. Near optimal timing and frequency offset estimation for 5G integrated LEO satellite communication system[J]. IEEE Access, 2019, 7: 113298-113310.
- [5] 缪德山,柴丽,孙建成,等.5G NTN 关键技术研究及演进展望[J].电信科学,2022,38(3):10-21.
- [6] WANG D, MEI Z Q, ZHANG H Q, et al. A novel PSS timing synchronization algorithm for cell search in 5G NR system[J]. IEEE Access, 2021, 9: 5870-5880.
- [7] CHEN F T, LI X, ZHANG Y, et al. Design and implementation of initial cell search in 5G NR systems[J]. China Communications, 2020, 17(5):38-49.
- [8] YOU Y H, SONG H K. Efficient sequential detection of carrier frequency offset and primary synchronization signal for 5G NR systems[J]. IEEE Transactions on Vehicular Technology, 2020, 69(8):9212-9216.
- [9] CHEN B, WANG H. Blind estimation of OFDM carrier frequency offset via oversampling[J]. IEEE Transactions on Signal Processing, 2004, 52(7):2047-2057.
- [10] MILYUTIN V S, ROGOZHNIKOV E V, PETROVSKIY K V, et al. Methods for improving the accuracy of frequency shift estimation in 5G NR [C]//2021 International Conference Engineering and Telecommunication. Dolgoprudny: IEEE, 2021: 1-5.

- [11] YOU Y H, PARK J H, AHN I Y. Complexity effective sequential detection of secondary synchronization signal for 5G new radio communication systems [J]. IEEE Systems Journal, 2021, 15(3): 3382–3390.
- [12] 孟凡军, 邓炳光, 秦启航, 等. 5G NR 小区搜索中一种频域相关快速同步算法[J]. 电讯技术, 2023, 63(4): 563–568.
- [13] MOOSE P H. A technique for orthogonal frequency division multiplexing frequency offset correction [J]. IEEE Transactions on Communications, 1994, 42(10): 2908–2914.
- [14] YIN J L, LEE M C, HSIAO W H, et al. A novel network resolved and mobile assisted cell search method for 5G cellular communication systems[J]. IEEE Access, 2022, 10: 75331–75342.
- [15] CHANG K, LEE S. Robust OFDM-based synchronization against very high fractional CFO and time-varying fading [J]. IEEE Systems Journal, 2020, 14(3): 4047–4058.
- [16] HUANG M N, CHEN J, FENG S L. Synchronization for OFDM-based satellite communication system[J]. IEEE Transactions on Vehicular Technology, 2021, 70(6): 5693–5702.
- [17] ABOUTANIOS E, MULGREW B. Iterative frequency estimation by interpolation on Fourier coefficients [J]. IEEE Transactions on Signal Processing, 2005, 53(4): 1237–1242.
- [18] SERBES A. Fast and efficient sinusoidal frequency estimation by using the DFT coefficients [J]. IEEE Transactions on Communications, 2019, 67(3): 2333–2342.
- [19] LIN J N, HOU Z W, ZHOU Y Q, et al. Map estimation based on Doppler characterization in broadband and mobile LEO satellite communications [C]//2016 IEEE 83rd Vehicular Technology Conference. Nanjing: IEEE, 2016: 1–5.

作者简介:

李昌森 男, 1998 年生于山西大同, 2020 年获学士学位, 现为硕士研究生, 主要研究方向为非地面通信网络下的同步与接入。

申 滨 男, 1978 年生于贵州兴义, 2011 年获博士学位, 现为教授、博士生导师, 主要研究方向为下一代移动通信系统、LTE/LTE-A 系统、认知无线电系统等领域的信号处理理论与技术等。

黄晓舸 女, 1982 年生于重庆, 2013 年获博士学位, 现为教授、博士生导师, 主要研究方向为移动通信技术、认知无线电动态频谱分配等。

DOI:10.20079/j.issn.1001-893x.230728001

基于传输时间的 LoRa 网络参数分配与功率控制算法^{*}

宋雨昕, 杨茂恒, 贾奇铭, 章 辉

(南开大学 电子信息与光学工程学院 天津市光电传感器与传感网络技术重点实验室, 天津 300350)

摘 要: 远距离无线电 (Long Range Radio, LoRa) 网络具有连接数量多、传输距离远、功耗低的特点, 然而海量 LoRa 终端的接入将使得 LoRa 网络的数据包碰撞率激增, 制约网络传输性能。为优化 LoRa 网络参数分配和功率控制, 以 A 类终端设备建立了 LoRa 网络冲突率的计算公式, 提出了一种基于传输时间的 LoRa 网络参数分配与功率控制算法, 其目标是 minimized 网络数据包传输平均冲突率时最小化网络能耗, 通过优化网络整体节点的参数和发送功率来实现。仿真结果表明, 与传统的 AAPA 等参数分配算法相比, 基于传输时间的 LoRa 网络参数分配与功率控制算法能显著降低冲突率 24.8% 以上, 降低网络能耗 30.7% 以上。

关键词: 物联网; LoRa 网络; 网络参数分配; 功率控制

开放科学 (资源服务) 标识码 (OSID):



中图分类号: TN914 文献标志码: A 文章编号: 1001-893X (2025) 03-0463-06

A Parameter Allocation and Power Control Algorithm Based on Transmission Time for LoRa Networks

SONG Yuxin, YANG Maoheng, JIA Qiming, ZHANG Hui

(Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, School of Electronic Information and Optical Engineering, Nankai University, Tianjing 300350, China)

Abstract: The Long Range Radio (LoRa) network has the characteristics of a large number of connections, a long transmission distance, and low power consumption. However, the access of a large number of LoRa terminals will cause a sharp increase in the data packet collision rate of the LoRa networks, restricting network transmission performance. To optimize the allocation of LoRa network parameters and power control for Class A terminal devices, a calculation formula for the LoRa network collision rate is developed. A LoRa network parameter allocation and power control algorithm based on transmission time is proposed, which aims to minimize network energy consumption when the average collision rate of network packet transmission is minimized by optimizing the parameters and transmission power of the overall nodes of the network. The simulation results show that compared with the traditional AAPA and other parameter allocation algorithms, the LoRa network parameter allocation and power control algorithm based on transmission time can significantly reduce the collision rate by more than 24.8% and reduce the network energy consumption by more than 30.7%.

Key words: Internet of Things; LoRa network; network parameter assignment; power control

0 引 言

低功耗广域网支持海量物联网终端连接^[1], 有

望解决低成本、远距离和低功耗的通信, 而上述应用通常具有低数据传输速率且没有特定的延迟要求。

^{*} 收稿日期: 2023-07-28; 修回日期: 2024-11-18
基金项目: 国家自然科学基金资助项目 (61871239, 61671254)
通信作者: 章辉 Email: zhangh@nankai.edu.cn

传统蜂窝网络虽然提供了广域范围的覆盖,但未能满足低成本和低功耗的因素^[2]。远距离无线电(Long Range Radio, LoRa)是一种典型的低功耗广域网技术,主要针对具有能源限制和远距离要求的物联网应用,在智慧城市、智能家居、工业和企业应用、智能农业等领域具有广泛的应用前景^[3-4]。

针对 LoRa 网络中的参数分配和功率控制问题,Reynders 等人^[5]计算了扩频因子之间终端的最佳分布比率,以实现每个扩频因子的相等的内部扩频因子碰撞概率,提出了一种将扩频因子和离散功率设置分配给节点的方案。该方案的目标是提高远离基站的节点由于捕获效应导致不公平的包错误率。Su 等人^[6]和 Qin 等人^[7]联合探索用户调度、扩频因子分配和功率分配,以目标信噪比要求和功率范围作为每个 LoRa 用户的约束来制定最大化系统能效的非凸优化问题,并将低功耗广域网中的资源分配被表述为信道分配和功率分配的联合优化问题,提出了一种低复杂度的匹配信道分配算法,并进一步发展了一种最优功率分配算法,以最大化低功耗广域网中达到的最小传输速率。Sandoval 等人^[8]通过计算每个节点的最佳配置设置来最大化网络吞吐量,提出了一种将最佳传输参数配置到各终端的更新策略。Ta 等人^[9]提出了在节点端利用机器学习自主决定最有效的传输设置,在节点上采用基于强化学习的算法。Li 等人^[10]通过检测网络的信道状态和链路质量,基于反馈机制,动态地优化传输功率,最大程度提高数据包的传输效率。Al-Gumaei 等人^[11]对 LoRa 网络建立数学模型,考虑能量效率、信号质量和网络容量等因素,利用数值方法实现最优传输功率控制算法。

现有文献表明,优化设计参数分配和功率控制是提高 LoRa 传输性能的重要途径。但现有优化算法所考虑的网络参数不全面或缺少对功率的控制。基于上述问题,本文针对降低 LoRa 网络能耗并降低冲突率减少数据包丢失这一问题提出解决方案。根据低功耗广域网协议(Long Range Wide Area Network, LoRaWAN)中的 A 类设备定义理论冲突模型,计算理论的冲突率,并确立 LoRa 网络的优化目标,由此提出了基于传输时间的解决方案,并通过仿真验证了其优越性能。

1 LoRa 传输理论

LoRa 主要工作于 1 GHz 以下的频段,通过调整

特定的无线电相关参数实现不同的传输特性,如载波频率(Carrier Frequency, CF)、扩频因子(Spreading Factor, SF)、带宽(Bandwidth, BW)、发送功率(Transmission Power, TP)和编码率(Coding Rate, CR)等参数。LoRa 实际传输速率主要受传输带宽、扩频因子和编码率控制制约,计算方法如下^[12]:

$$R_b = SF \times \frac{BW}{2^{SF}} \times \frac{4}{4+CR} \quad (1)$$

上述参数可在 LoRa 设备端调整,以提高整体网络性能。LoRa 数据包传输时间 T_{packet} 由前导码的传输时间 T_{preamble} 和有效负载的传输时间 T_{payload} 组成^[12]:

$$T_{\text{packet}} = (n_{\text{preamble}} + 4.25) \times \frac{2^{SF}}{BW} + \text{payloadSymNb} \times \frac{2^{SF}}{BW} \quad (2)$$

式中: n_{preamble} 是前导符号的长度; $2^{SF}/BW$ 是符号持续时间;payloadSymNb 是组成数据包有效载荷和报头的符号数,

$$\text{payloadSymNb} = 8 + \max\left(\text{ceil}\left(\frac{8PL - 4SF + 44 - 20H}{4(SF - 2DE)}\right)(CR + 4), 0\right) \quad (3)$$

式中:PL 为有效负载的长度; H 表示选择的报头类型;DE 表示数据传输过程中是否采用低速率优化;ceil 为取整函数^[12]。

如图 1 所示,LoRa 网络通常采用单网关星形拓扑结构^[2],其小区内可部署海量 LoRa 设备,因此整个 LoRa 网络的参数选择至关重要,分配方案决定了吞吐量和网络能耗的大小,可有效避免包丢失和降低能耗。由于 LoRa 设备采用电池供能,高能耗将使得设备的使用寿命降低。

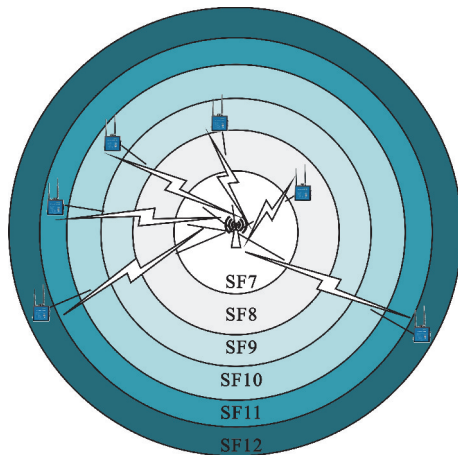


图 1 LoRa 系统设置-扩频因子 SF 距离分配

接收到的信号功率 P_{rx} 取决于发射功率 P_{tx} 和通信路径上的所有增益和损失^[13]:

$$P_{rx} = P_{tx} + GL - LPL \quad (4)$$

式中:GL 结合了所有一般的得失,包括发射机天线增益、发射机损失、接收天线增益、接收损耗、杂项损失;LPL 表示由通信环境性质决定的路径损耗,为对数距离路径损失模型。该模型通常用于对建筑密集地区的部署进行建模。利用该模型,通信距离为 d 的路径损耗可描述为^[13]

$$LPL(d) = \overline{LPL}(d_0) + 10\gamma \lg\left(\frac{d}{d_0}\right) + X_\sigma \quad (5)$$

式中: d_0 是参考距离; $\overline{LPL}(d_0)$ 是距离为 d_0 的路径损耗; γ 是路径损耗指数; $X_\sigma \sim N(0, \sigma^2)$ 。在发射端,信号覆盖范围可通过改变发射功率来调节。在接收端,覆盖范围受到灵敏度阈值 S_{rx} 的限制,而灵敏度阈值 S_{rx} 又受到扩频因子 SF 和带宽 BW 的影响。根据 Semtech 公司的 LoRa 调制解调计算工具得到表 1 的灵敏度值。

表 1 不同扩频因子和带宽下灵敏度

SF	灵敏度/dBm		
	BW = 125 kHz	BW = 250 kHz	BW = 500 kHz
7	-124.5	-121.5	-118.5
8	-127.0	-124.0	-121.0
9	-129.5	-126.5	-123.5
10	-132.0	-129.0	-126.0
11	-134.5	-131.5	-128.5
12	-137.0	-134.0	-131.0

LoRa 广域网协议 LoRaWAN 定义了基于 LoRa 的广域网通信协议,包括 A、B、C 3 种不同的类型。A 类适用于有能量限制和最低下行链路要求的设备,设备通过 Aloha 发送上行帧,并且只接收下行帧作为对上行帧的响应。A 类设备最省电,设备上行数据后会打开两个短暂的下行接收窗口。B 类扩展了 A 类,包括以预定间隔接收信标帧,并增加了以预定间隔接收服务器发起的下行链路帧。B 类设备还会根据从网关接收的时间同步信标,打开指定时间的窗口。C 类设备在不发送时保持接收模式,因此该类不适合有能量限制的设备。C 类设备的接收窗口一直处于打开的状态,只会在发送数据的时候短暂的打开关闭,因此 C 类设备更加耗电^[4]。

2 LoRa 网络参数分配和功率控制

2.1 问题描述

当 LoRa 网络的节点数量增加,网络参数的选择将严重影响网络性能,有多种 LoRa 和 LoRaWAN 设置的组合可导致不同级别的性能^[14]。影响网络容量的参数与影响数据传输时间的参数密切相关,涉及扩频因子 SF、带宽 BW 和编码率 CR。在 LoRa 网络中采用高扩频因子 SF 值可实现更远的传输距离,更高的接收灵敏度。然而对于具有相同有效负载的数据包,式(1)和式(2)表明高扩频因子 SF 值会导致低传输速率和长传输时间^[3]。通过使用更宽的带宽信道,可在更小的距离内实现更高的传输速率。另外,可调整发送功率设置以提高信号可实现的覆盖范围,并针对可能的信道损失提高成功接收终端节点发送数据包并解码的概率^[15]。参数以及信道的不同组合可用于确保成功的并发上行链路传输。多通道 LoRa 网关可以解码多达 8 个并发准正交上行链路传输^[16],因此,基于 Aloha 协议的随机接入特性,考虑冲突发生的条件及 LoRa 的传输设置,定义平均冲突率如下:

$$P_{col} = 1 - \frac{1}{6} \sum_{SF=7}^{12} \exp(-2N \cdot T_{packet, SF} \cdot \lambda \cdot p_{SF}) \quad (6)$$

式中: λ 是数据包平均发送速率; $T_{packet, SF}$ 是使用 SF 的数据包传输时间; N 是网络总节点数目; p_{SF} 是使用扩频因子 SF 的比例; $N \times p_{SF}$ 表示使用 SF 的节点数目。

在海量 LoRa 终端被部署在同一小区时,在减少数据包丢失的同时,如何在保证网络的低能耗或者降低网络能耗是亟待解决的问题。当冲突发生时,接收功率的差值会决定数据包是否被解码。因此,定义网络能耗(Network Energy Consumption, NEC)如下:

$$NEC = \sum_{i=1}^N T_{packet, i} \cdot P_{tx, i} \quad (7)$$

式中: $P_{tx, i}$ 是节点 i 使用的发送功率; $T_{packet, i}$ 是节点 i 的数据包传输时间。

则优化目标可定义为

$$\min_{P_{col}} NEC \quad (8)$$

该优化目标属于多目标优化。从上述两式可知,优化的关键在于数据包传输时间。当使用更高的扩频因子、更窄的信道带宽和更低的编码率时,传输时间会增加。采用基于 Aloha 的 MAC 层,由节点正确选择扩频因子和编码率对于可实现的网络总容量至关重要。然而,大多数时候会选择最低的扩频

因子以实现最低的能耗。此外,信道带宽的选择对整体容量也起着重要作用。LoRa 传输在更宽的信道上比在更窄的信道上传输得更快。发送功率的选择需考虑扩频因子、带宽、编码率和链路预算,以及网关灵敏度。

2.2 基于传输时间的参数分配与功率控制算法

本文的优化目标为式(8),即最小化网络数据包传输平均冲突率时最小化网络能耗,属于多目标优化问题。节点选取低功率不会影响冲突率,但会提高捕获效应带来的影响,增加数据包的丢失。为此,基于式(7)和式(8)的共同影响因子传输时间,提出一种基于传输时间的参数分配与功率控制算法,主要有五大步骤,具体如下:

输入:已确定位置信息的节点集合 U_0 。

输出:初始化设置完成的节点集合 U_1 。

步骤 1 将所有节点按照节点到网关的距离排序,得到节点集 U_0 。

步骤 2 载频列表 CF;扩频因子列表 SF;带宽列表 BW;初始化三维传输时间矩阵 T ,维度大小为 $(\text{len}(\text{CF}), \text{len}(\text{SF}), \text{len}(\text{BW}))$,值为 0;对节点集 U_0 中所有节点执行:找出 T 中最小值的位置信息 X, Y, Z ,根据 X, Y, Z 得到对应的 CF、SF、BW 值并赋给当前节点: $T[X, Y, Z] = T[X, Y, Z] + \text{在空时间 airtime}$ (在空时间 airtime 由式(2)计算得出)。

步骤 3 再次将所有节点按照节点到网关的距离排序,得到节点集 U'_0 。

步骤 4 执行以下伪代码:

For U_i in U'_0 do:

For U_j in $[U_n - U_i]$ do:

If $\text{SF}_{U_i} > \text{SF}_{U_j}$:

交换 U_i 与 U_j 的参数

步骤 5 对所有节点执行:

步骤 5.1 计算链路预算;

步骤 5.2 根据表 1 的灵敏度分配合适发送功率。

算法综合考虑节点之间的距离、传输时间、扩频因子等因素,以及参数的优化和功率分配,来有效地配置 LoRa 网络中节点的参数,以提高网络性能和可靠性。步骤 2 帮助平衡网络中节点的传输负载,避免某些节点传输时间过长,降低网络整体效率。根据扩频因子的使用条件,距离网关较近的节点需要较低的扩频因子以获得较高的数据传输速率,距离网络较远的节点需要较高的扩频因子以覆盖更远的传输范围。如果距离网关较远的节点扩频因子小,可能会导致该节点的传输范围不足以覆盖通信

距离,因此步骤 4 是一个交换扩频因子的过程,从距离更远的节点开始搜索。

3 仿真分析

本文仿真基于 LoRaSim 平台^[17],通过模拟在二维空间中放置网关和终端来部署 LoRa 网络,支持上行链路,终端定期生成数据包发送到网关。LoRaSim 使用对数距离路径损耗模型,该模型基于经验测量的 $d_0 = 40$ m 距离处的参考路径损耗值 $\overline{\text{LPL}}(d_0) = 127.47$ dB,路径损耗指数 γ 为 2.08 和 σ 为 3.57。LoRaSim 可模拟 LoRa 网络数据包的碰撞过程,分析接收灵敏度和数据包捕获效果。仿真参数如表 2 所示。

表 2 仿真参数

参数	值
载频 CF/MHz	486.3~487.7
扩频因子 SF	7~12
带宽 BW/kHz	125,250,500
编码率 CR	1~4
发送功率 TP/dBm	2,6,8,14
平均发送速率 $\lambda/(\text{packet/s})$	1×10^{-3}

本文使用了冲突率和网络能耗两个评估指标来评估 LoRa 网络的性能。冲突率包含理论冲突率与实际冲突率。理论冲突率根据式(6)计算。实际冲突率计算如下式:

$$R_{\text{col}} = \frac{C}{N_s} \quad (9)$$

式中: C 是丢失数据包数量; N_s 是发送数据包数量。

网络能耗 NEC 由式(7)定义。每个数据包的传输能量消耗取决于发送功率 TP 和传输时间,并受其他参数的影响。该指标越低,网络的生命周期越长。在许多场景中,LoRa 节点使用电池供电,因此必须将传输的能量消耗保持在最低水平。本文仿真选取节点数为 500~2 800,采取单网关组网模式。

本文选取 3 种传统算法,即文献[5]的算法 (Power, Channel and Spreading Factor Selection Algorithm, PCSF)、文献[14]的算法 (Adaptive Allocation and Parameter Adjustment Algorithm, AAPA) 和文献[18]的算法 (Proximal Algorithm, PA) 作为对比。图 2~5 展示了各分配方案在 2 800 个节点下的扩频因子与发送功率分布。等距分配方案发送功率比其他方案最低功率分布更广泛,但并不

表示网络功耗最低。从图 6 中可以得出,该分配方案网络能耗最高。文献[14]的 AAPA 分配算法以传输时间为分配基础,图 2 展示了其杂乱的扩频因子分配,这导致较多的节点需使用更高的发送功率。文献[5]的 PCSF 分配算法同时控制扩频因子和发送功率,以获得两者平衡,但未达到分配最优。文献[18]的 PA 分配算法分配载波频率和扩频因子,但时间复杂度较高。本文的分配方案解决了 AAPA 算法杂乱的扩频因子分配问题,更符合实际部署情况,且比 PCSF 算法更优化。

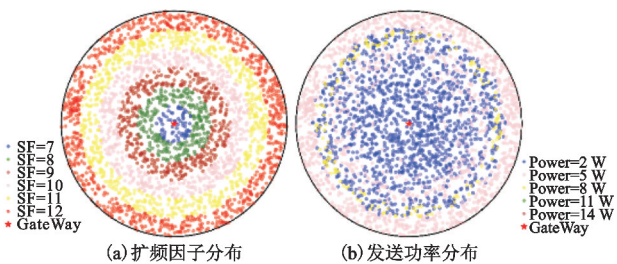


图 2 PA 算法扩频因子分布与功率分布

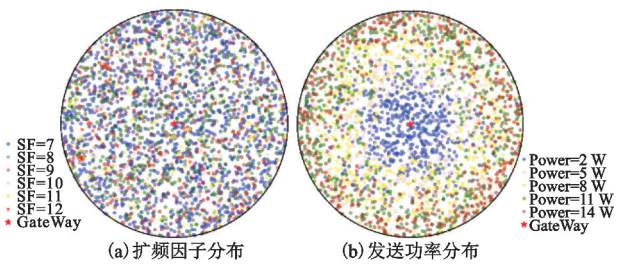


图 3 AAPA 算法扩频因子与发送功率分布

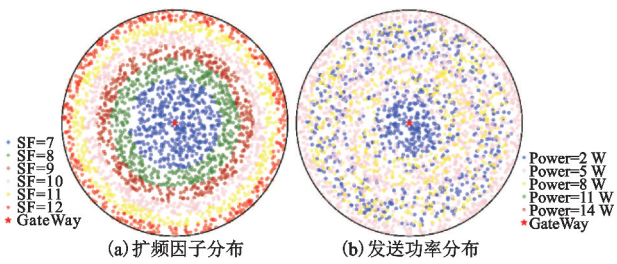


图 4 PCSF 算法扩频因子与发送功率分布

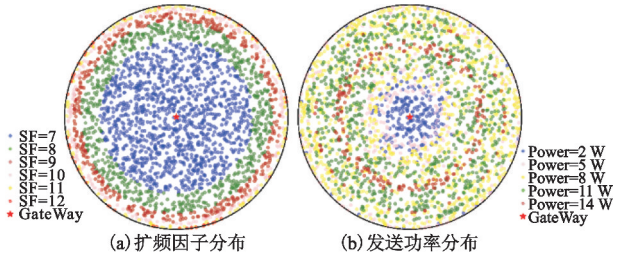


图 5 本文方案扩频因子与发送功率分布

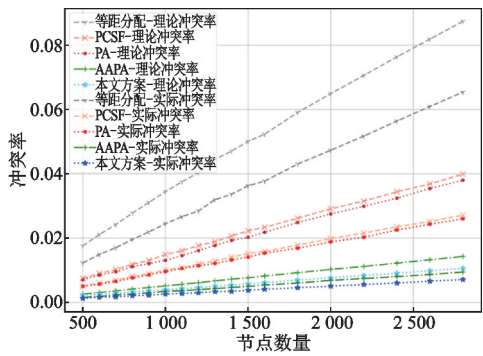


图 6 冲突率与节点数关系

从理论冲突概率的分析比较而言,本文方案在理论指标上远优于等距分配、PCSF 算法和 PA 算法,略优于 AAPA 算法。实际冲突率符合理论冲突概率的趋势,但略低于理论冲突概率。当然,理论冲突率值越低,说明网络冲突越少,网络吞吐量越高。在 2 800 个网络终端下,本文方案的网络能耗与其他方案的对比如图 7 所示。在此基础上,表 3 给出了本文各指标在其他对比方案上的占比;实际冲突率分别比 AAPA 算法低 24. 8%, PA 算法低 50. 5%, PCSF 算法低 73. 9%;网络能耗比 AAPA 算法低 30. 7%, PA 算法低 42. 1%, PCSF 算法低 73. 8%。算法复杂度如表 4 所示。

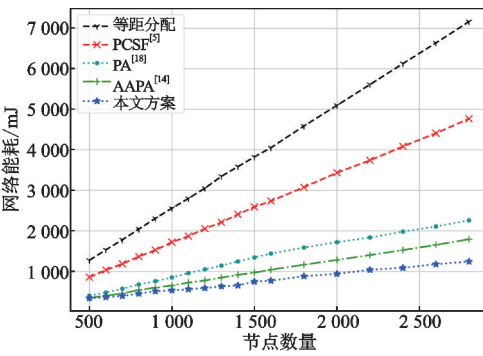


图 7 网络能耗与节点数量关系

表 3 本文方案各指标在其他对比方案上的占比

方案	理论冲突率/%	实际冲突率/%	网络能耗/%
AAPA ^[14]	74. 0	75. 2	69. 3
PA ^[18]	28. 3	28. 0	51. 3
PCSF ^[5]	26. 6	26. 1	26. 1
等距分配	12. 1	10. 8	17. 4

表 4 算法复杂度比较

算法	复杂度
本文方案	$O(N \lg N)$
AAPA 算法 ^[14]	$O(N)$
PA 算法 ^[18]	$O(2^N(N/2)^M)$
PCSF 算法 ^[5]	$O(N^2)$
等距分配	$O(N)$

4 结束语

低功耗广域网的特点是数据速率要求低、覆盖单元大,导致在同一小区中共存的终端数量非常多。LoRa 提供了不同的参数组合规范数据包的传输并提供无碰撞通信。针对现目前存在的节点参数分配不合理导致高冲突率与高能耗的问题,本文考虑优化控制冲突发生的 3 个主要参数的分布,以 A 类终端设备提出了一种基于传输时间的参数分配与功率控制算法,目标在于最小化冲突率时最小化网络能耗。在此基础上,提出了均匀每组参数组合传输时间分布的方案,计算了数据包的冲突率。仿真结果表明,在 LoRa 终端数量较高的网络,本文方案的实际冲突率只有文献[14]的 75.2%,文献[18]的 28%,文献[5]的 26.1%;网络能耗是文献[14]的 69.3%,文献[18]的 51.3%,文献[5]的 26.1%。在终端数量多的场景下,本文算法能有效提升 LoRa 网络整体性能。在实际应用中,通过本文方案可以降低冲突率和网络能耗。

未来将扩展到 LoRa 网络海量终端设备移动场景,在移动场景中需要考虑终端设备与网关的稳定通信,这涉及到选择传输参数、切换最优网关、延长电池寿命等要求。

参考文献:

- [1] GKOTSIPOULOS P, ZORBAS D, DOULIGERIS C. Performance determinants in LoRa networks: a literature review[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1721-1758.
- [2] RAZA U, KULKARNI P, SOORIYABANDARA M. Low power wide area networks: an overview[J]. IEEE Communications Surveys & Tutorials, 2017, 19(2): 855-873.
- [3] OSORIO A, CALLE M, SOTO J D, et al. Routing in LoRaWAN: overview and challenges [J]. IEEE Communications Magazine, 2020, 58(6): 72-76.
- [4] JOUHARI M, SAEED N, ALOUINI M S, et al. A survey on scalable LoRaWAN for massive IoT: recent advances, potentials, and challenges [J]. IEEE Communications Surveys & Tutorials, 2023, 25(3): 1841-1876.
- [5] REYNDERS B, MEERT W, POLLIN S. Power and spreading factor control in low power wide area networks [C]//2017 IEEE International Conference on Communications. Paris: IEEE, 2017: 1-6.
- [6] SU B B, QIN Z J, NI Q. Energy efficient resource allocation for uplink LoRa networks [C]//2018 IEEE Global Communications Conference. Abu Dhabi: IEEE, 2018: 1-7.
- [7] QIN Z J, MCCANN J A. Resource efficiency in low-power wide-area networks for IoT applications [C]//2017 IEEE Global Communications Conference. Singapore: IEEE, 2017: 1-7.

- [8] SANDOVAL R M, GARCIA-SANCHEZ A J, GARCIA-HARO J. Optimizing and updating LoRa communication parameters: a machine learning approach [J]. IEEE Transactions on Network and Service Management, 2019, 16(3): 884-895.
- [9] TA D T, KHAWAM K, LAHOUD S, et al. LoRa-MAB: a flexible simulator for decentralized learning resource allocation in IoT networks [C]//2019 12th IFIP Wireless and Mobile Networking Conference. Paris: IEEE, 2019: 55-62.
- [10] LI Y H, YANG J, WANG J L. DyLoRa: towards energy efficient dynamic LoRa transmission control [C]//2020 IEEE Conference on Computer Communications. Toronto: IEEE, 2020: 2312-2320.
- [11] AL-GUMAEI Y A, ASLAM N, CHEN X M, et al. Optimizing power allocation in LoRaWAN IoT applications [J]. IEEE Internet of Things Journal, 2022, 9(5): 3429-3442.
- [12] 王硕禾, 刘旭, 李苏晨, 等. 基于多目标遗传算法的 LoRa 参数匹配优化 [J]. 计算机工程与科学, 2020, 42(3): 434-440.
- [13] ZHANG H, LI M K, YU H D, et al. Dynamic parameter selection of LoRa edge nodes using reinforcement learning with link prior knowledge [J]. IEEE Internet of Things Journal, 2024, 11(21): 34420-34433.
- [14] BOR M C, ROEDIG U, VOIGT T, et al. Do LoRa low-power wide-area networks scale? [C]//The 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems. Malta: IEEE, 2016: 59-67.
- [15] 周超, 章辉, 杨茂恒, 等. 基于烟花爆炸式混合蛙跳算法的 LoRa 网络参数分配策略 [J]. 电讯技术, 2022, 62(6): 795-801.
- [16] BOR M, ROEDIG U. LoRa transmission parameter selection [C]//2017 13th International Conference on Distributed Computing in Sensor Systems. Ottawa: IEEE, 2017: 27-34.
- [17] ABDELFADEEL K Q, CIONCA V, PESCH D. Fair adaptive data rate allocation and power control in LoRaWAN [C]//2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks". Chania: IEEE, 2018: 14-15.
- [18] SALLUM E, PEREIRA N, ALVES M, et al. Performance optimization on LoRa networks through assigning radio parameters [C]//2020 IEEE International Conference on Industrial Technology. Buenos Aires: IEEE, 2020: 304-309.

作者简介:

宋雨昕 女, 1999 年生于山西晋城, 硕士研究生, 主要研究方向为无线通信与传感网络。

杨茂恒 男, 1998 年生于四川资阳, 硕士研究生, 主要研究方向为无线通信与传感网络。

贾奇铭 男, 1998 年生于山西忻州, 硕士研究生, 主要研究方向为智能反射面。

章辉 男, 1982 年生于湖北鄂州, 2010 年于北京邮电大学获博士学位, 现为南开大学副研究员, 主要研究方向为 6G 无线通信系统、工业物联网等。

DOI:10.20079/j.issn.1001-893x.231226001

基于 STAR-RIS 和中继集成系统的下行 NOMA 覆盖增强技术*

陈思宇^{1,2}, 申 滨^{1,2}, 元文军^{1,2}

(1. 重庆邮电大学 通信与信息工程学院, 重庆 400065; 2. 移动通信技术重庆市重点实验室, 重庆 400065)

摘 要:可重构智能表面(Reconfigurable Intelligent Surface, RIS)与中继技术相结合是近年来兴起的新型增强型无线通信技术。为实现下行覆盖增强,提出了一种基于同时反射和透射的可重构智能表面(Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface, STAR-RIS)与解码转发(Decode-and-Forward, DF)中继协同工作的下行非正交多址(Non-orthogonal Multiple Access, NOMA)双用户系统。在 Nakagami-m 信道假设下,通过联合优化基站和中继处的 NOMA 功率分配以及 STAR-RIS 的能量分裂系数,实现最大化该系统中两个用户的可实现速率之和。在满足功率约束和理想连续干扰消除的条件下,提出了一种基于两阶段中继系统流量平衡的双层迭代算法,通过内层迭代功率分配和外层迭代能量分裂系数来得到最优值。仿真结果表明,该集成系统覆盖性能在系统功率同为 1 W 的情况下要优于纯中继系统,且优势随 STAR-RIS 元件数量增多而变大;当 STAR-RIS 元件个数小于 200 或发射信噪比小于 80 dB 时,该系统相较于纯 STAR-RIS 系统亦可获得 0.5~2 b/s/Hz 的系统和速率增益。

关键词:STAR-RIS; 非正交多址(NOMA); 解码转发中继; 集成系统; 功率分配

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN929.5 文献标志码:A 文章编号:1001-893X(2025)03-0469-08

Downlink NOMA Coverage Enhancement Based on Integrated STAR-RIS and Relay

CHEN Siyu^{1,2}, SHEN Bin^{1,2}, YUAN Wenjun^{1,2}

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
2. Chongqing Key Laboratory of Mobile Communications Technology, Chongqing 400065, China)

Abstract: The combination of reconfigurable intelligent surface (RIS) and relaying technology is a new enhanced wireless communication technology emerging in recent years. To achieve downlink coverage enhancement, a dual-user system for downlink non-orthogonal multiple access (NOMA) based on simultaneously transmitting and RIS (STAR-RIS) working in concert with decode-and-forward (DF) relay is proposed. Under the Nakagami-m channel assumption, the sum of the achievable rates of the two users in this system is maximized by jointly optimizing the NOMA power allocation at the base station and the relay as well as the energy splitting coefficient of STAR-RIS. Subject to satisfying the total power constraint and ideal successive interference cancellation, a two-layer iterative algorithm based on flow balancing in two-phase relay systems is proposed to obtain the optimal value by iterating the power allocation in the inner layer and the energy splitting coefficient in the outer layer. The simulation results show that the coverage performance of the integrated system is better than that of the single relay system under the same system power of 1 W, and the advantage increases with the increase of the number of STAR-RIS components; when the number of STAR-RIS components is less than 200 or the transmission signal-to-noise ratio is less than 80 dB, the system can also achieve sum-rate gain of 0.5~2 b/s/Hz compared with single STAR-RIS systems.

Key words: STAR-RIS; non-orthogonal multiple access (NOMA); DF relay; integrated system; power allocation

* 收稿日期:2023-12-26;修回日期:2024-07-27
基金项目:国家自然科学基金资助项目(62371082)
通信作者:陈思宇 Email:1586954246@qq.com

0 引言

未来 10 年通信网络容量预计将千倍增长,无处不在的无线连接将成为现实。高度复杂的网络、高成本的硬件和日益增加的能源消耗成为未来无线通信面临的关键问题^[1-2],研究高效、频谱及资源友好的未来无线网络解决方案势在必行。

在众多新涌现出的候选新技术中,可重构智能表面(Reconfigurable Intelligent Surface, RIS)以其独特的低成本、低能耗、可编程、易部署等特点脱颖而出。RIS 对于未来无线网络的覆盖增强和容量提升有很大潜力,可提供虚拟视距链路、消除局部覆盖空洞、服务小区边缘用户等,进而实现智能可重构的无线环境^[3]。近年来,已有较多文献研究 RIS 辅助的通信系统。文献[4]通过优化 RIS 的部署位置,包括 RIS 与基站间的距离及角度来最大化小区覆盖。

然而,传统的单一反射或透射 RIS 必须满足用户位于 RIS 的同一侧的部署要求,只能实现半空间覆盖,严重限制了 RIS 部署的灵活性。在此基础上,一种新颖的同时反射和透射的可重构智能表面(Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface, STAR-RIS)的概念得以提出,其独特之处在于可以同时透射和反射入射信号,由此可以实现全空间覆盖^[5]。由于 STAR-RIS 将入射信号分割为两部分,因此需要多址方案来区分,以便位于 STAR-RIS 不同侧的用户设备(User Equipment, UE)可以成功解调。与传统的正交多址接入(Orthogonal Multiple Access, OMA)相比,非正交多址接入(Non-orthogonal Multiple Access, NOMA)允许多个用户通过在发射机处采用叠加编码(Superposition Coding, SC)和在接收机处采用连续干扰消除(Successive Interference Cancellation, SIC)来共享相同的资源块^[6],可以显著提高 STAR-RIS 辅助网络的频谱效率以及保证 UE 公平性^[7],同时 STAR-RIS 也能为 NOMA 系统提供全空间覆盖和解码灵活性。文献[5]研究了 STAR-RIS 辅助的双用户通信网络覆盖范围,针对 NOMA 和 OMA 分别提出了总覆盖范围最大化问题并使用凸优化方法求解。

在无线通信中,有源协作中继在提供可靠和节能通信方面发挥了关键作用^[8]。但中继有着信号处理复杂度高、增强噪声干扰以及高实现成本等缺点,而使用简单、低成本且无源的 RIS 可以减轻以上

缺点^[9-10]。现有研究大多将 RIS 与中继视作互相竞争的技术,但也有文献已经表明有源中继和 RIS 可以和谐共存,以提高系统的速率性能。文献[11]介绍了两种将 RIS 与解码转发(Decode-and-Forward, DF)中继相结合的混合传输方案,结果表明这两种技术可以在适当的传输场景下相互协调工作,中继可以视为 RIS 辅助传输场景的额外性能增强组件,尤其是在信道快速恶化的情况下。

综上所述,RIS 和 NOMA 技术都将是未来无线通信中的关键技术,两者与现有成熟的中继技术相结合可以发挥更大的作用。然而,现有大多数文献提出的 RIS 与中继混合系统只考虑了单用户场景,对多用户系统关注较少,也并未将 STAR-RIS 与中继联系起来。鉴于现有研究的局限性,本文将 STAR-RIS 和 DF 中继集成在一起,并与 NOMA 技术相结合,提出了一种新的双用户集成系统。集成系统的优势在于可以减小 RIS 和中继混合网络的信道估计开销,且通过控制是否激活中继可以实现灵活高效的覆盖效果。通过联合优化基站和中继处的 NOMA 功率分配以及 STAR-RIS 的能量分裂系数最大化该系统中两个用户的可实现速率之和,以增强覆盖。为解决所述优化问题,本文利用两阶段中继系统流量平衡的思想,提出了一个有效的双层迭代算法。理论分析与仿真结果表明,在功率和 STAR-RIS 元件数量受限的条件下,所提系统的性能是优于纯 STAR-RIS 系统的。另外,本文还分析了该集成设备的位置部署对于系统性能的影响,结果也验证了所提集成系统及迭代算法的有效性与性能优势。

1 系统模型

本文设计并提出了一种基于集成 STAR-RIS 和 DF 中继辅助的下行链路通信系统,如图 1 所示,基站(Base Station, BS)利用 NOMA 技术为两个用户 U_r 和 U_t 提供服务。BS 和两个用户以及中继都配备了全向天线,并且中继在 HD 模式下工作。设由于信号传播阻塞,BS 与两个用户之间皆不存在直接链路。STAR-RIS 由 N 个反射元件组成,并且智能相位调整是在理想的信道状态信息(Channel State Information, CSI)完全已知的条件下执行的。用户 U_r 位于反射区域,用户 U_t 位于透射区域,且 U_r 与 U_t 相比更接近 BS。

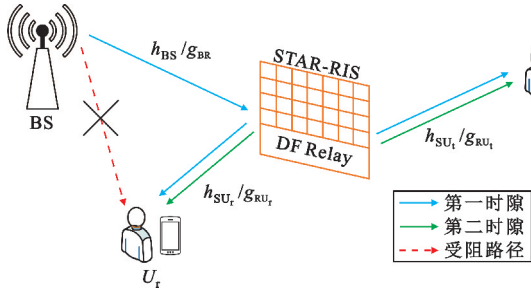


图 1 系统模型

1.1 信道模型

为了更好地描述实际传播场景中的小规模衰落,假设所有节点之间的通信链路均为 Nakagami-m 衰落信道。将第一个时隙和第二个时隙内 BS 到用户 $U_i (i=r, t)$ 、BS 到 STAR-RIS、STAR-RIS 到用户 U_i 的信道分别表示为 h_{BU_i} 、 $h_{BS} \in \mathbb{C}^{N \times 1}$ 和 $h_{SU_i} \in \mathbb{C}^{N \times 1}$, 并将第一个时隙 BS 到 DF 中继以及第二个时隙 DF 中继到用户 U_i 的信道分别表示为 g_{BR} 和 g_{RU_i} , 以上信道中的每个元素 $h \in \{g_{BR}, g_{RU_i}, h_{BU_i}, h_{BS}, h_{SU_i}\}$ 都可以表示为

$$h = \sqrt{vd^{-\alpha}} ge^{j\theta} \quad (1)$$

式中: v 表示 1 m 参考距离的路径损耗; α 和 d 分别表示两个节点间的路径损耗指数以及距离; θ 均匀分布在 $[0, 2\pi)$; g 服从 Nakagami-m 分布, 其概率密度函数为

$$f_g(x) = \frac{2m^m}{\Gamma(m)\Omega^m} x^{2m-1} e^{-\frac{m}{\Omega}x^2} \quad (2)$$

式中: $\Gamma(\cdot)$ 是伽马函数; $\Omega = E[g^2]$ 是平均功率; $m = \frac{\Omega^2}{E[(g^2 - \Omega)^2]}$ 是形状参数。

1.2 STAR-RIS 以及 DF 中继信号传输

文献[12]中提出了 STAR-RIS 的 3 种实用协议, 即能量分裂 (Energy Splitting, ES)、模式切换 (Mode Switching, MS) 和时间切换 (Time Switching, TS)。本文采用 ES 协议, 即所有 RIS 元件都可以同时进行透射和反射, 其中入射到每个元件上的信号能量分解为透射和反射两部分的信号能量。假设两个 $N \times N$ 的对角矩阵 $\mathbf{\Theta}_T = \text{diag}(t_1, t_2, \dots, t_N)$ 以及 $\mathbf{\Theta}_R = \text{diag}(r_1, r_2, \dots, r_N)$ 分别为 STAR-RIS 的透射系数和反射系数矩阵, 其中 $t_n = \sqrt{\beta_n^t} e^{j\theta_n^t}$ 和 $r_n = \sqrt{\beta_n^r} e^{j\theta_n^r}$ 分别表示 STAR-RIS 第 n 个元素的透射系数和反射系数。考虑到基于无源和无损元表面的 STAR-RIS,

反射信号和透射信号的能量之和必须等于入射信号的能量, 因此, 需要满足 $\beta_n^t + \beta_n^r = 1$, $\theta_n^t, \theta_n^r \in [0, 2\pi)$ 。

解码和转发中继接收的信号经历完整的解调和解码过程, 以及重新编码和重新调制步骤。DF 策略能够将整个通信链路拆分成发端到中继、中继到收端两个独立的子链路。中继系统的工作原理是通过接收和转发子链路之间的数据包来提高网络的性能, 如果一个子链路的吞吐量较低, 那么中继系统就需要等待这个子链路处理完数据包后才能继续接收和转发其他数据包, 从而限制了整个中继系统的吞吐量, 使得总吞吐量对应于两个子链路中的最低吞吐量。

1.3 信号传输模型

一个完整的信号传输过程分为两个时隙, 从 BS 发出的信号到达两个用户以及 STAR-RIS 和 DF 中继集成设备处, 信号通过 STAR-RIS 的反射或透射和中继的解码转发过程传输给两个用户。假设两个时隙中 CSI 保持不变, 整个信号传输过程如图 2 所示。

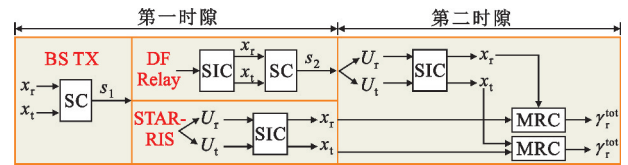


图 2 STAR-RIS 和 DF 中继集成系统的信号传输流程

具体地, 在第一个时隙, BS 发送叠加编码信号 $s_1 = (\sqrt{P_1}x_r + \sqrt{P_2}x_t)$, 其中, P_1 和 P_2 分别是 BS 分配给用户 U_r 和 U_t 的发射功率, x_i 为发送给 U_i 的信息, 满足 $E[|x_i|^2] = 1 (i=r, t)$ 。由于假设 U_r 和 U_t 相比更接近于 BS, 则从中继到用户 U_r 的信道增益 g_{RU_r} 大于从中继到 U_t 的信道增益 g_{RU_t} , 即 $|g_{RU_r}|^2 \geq |g_{RU_t}|^2$ 。在下行链路 NOMA 中, 高信道增益用户的消息以低功率电平发送, 而低信道增益用户消息以高功率电平发送^[13], 即 $P_2 > P_1$ 。因此, 中继处的接收信号可以表示为

$$y_R = g_{BR}s_1 + n_R \quad (3)$$

式中: $n_R \sim \mathcal{CN}(0, \sigma^2)$ 表示中继处的加性高斯白噪声, 方差为 σ^2 。

中继接收到信号 y_R 以后对其进行解码并重新进行叠加编码, 以便在第二个时隙转发信息。DF 中继首先将 x_r 视为噪声来解码 x_t , 然后通过 SIC 将 x_t

从叠加信号中剔除,再对 x_r 进行解码。因此,中继解码 x_t 和 x_r 时的 SINR 分别为

$$\gamma_t^R = \frac{p_2 |g_{BR}|^2}{p_1 |g_{BR}|^2 + \sigma^2} \quad (4)$$

$$\gamma_r^R = \frac{p_1 |g_{BR}|^2}{\sigma^2} \quad (5)$$

令 $g_1 = |g_{BR}|^2$, 由于 g_{BR} 服从 Nakagami-m 分布, 其模值的平方服从 Gamma 分布, 即 $g_1 \sim \text{Gamma}(m, \Omega/m)$ 。

同时,在第一个时隙信号通过 STAR-RIS 反射或者透射给两个用户,因此第一个时隙用户 U_r 和 U_t 处的接收信号可以表示为

$$y_r^I = \mathbf{h}_{SU_r}^T \mathbf{\Theta}_R \mathbf{h}_{BS} s_1 + n_r^I \quad (6)$$

$$y_t^I = \mathbf{h}_{SU_t}^T \mathbf{\Theta}_I \mathbf{h}_{BS} s_1 + n_t^I \quad (7)$$

式中: $n_r^I, n_t^I \sim \mathcal{CN}(0, \sigma^2)$ 分别为第一个时隙 U_r 和 U_t 处的加性高斯白噪声。 U_t 只需解码 x_t 并将 x_r 视为噪声,相应的 SINR 为

$$\gamma_t^I = \frac{P_2 |\mathbf{h}_{SU_t}^T \mathbf{\Theta}_I \mathbf{h}_{BS}|^2}{P_1 |\mathbf{h}_{SU_t}^T \mathbf{\Theta}_I \mathbf{h}_{BS}|^2 + \sigma^2} \quad (8)$$

令 $G_t = |\mathbf{h}_{SU_t}^T \mathbf{\Theta}_I \mathbf{h}_{BS}|^2$, 为实现收端所有信号的相位一致以实现最大的 SINR, 假设 STAR-RIS 最优透射相移为 $\theta_n^I = \angle(h_{BS,n}, h_{SU_t,n})$, STAR-RIS 所有元件的能量分裂系数保持一致, 即 $\beta_n^I = \beta_r, \beta_n^I = \beta_t, n = 1, 2, \dots, N$, 则 G_t 可以重新表示为

$$G_t = |\mathbf{h}_{SU_t}^T \mathbf{\Theta}_I \mathbf{h}_{BS}|^2 = \beta_t \left| \sum_{n=1}^N |h_{BS,n}| |h_{SU_t,n}| \right|^2 \quad (9)$$

对于复合信道功率增益 G_t 的分布, 文献[14]中提供了一种易于处理的拟合方法, 当 N 够大时, G_t 同样服从 Gamma 分布, 即 $G_t \sim \text{Gamma}(\frac{u_r^2}{4\sigma_r^2} N, 4\sigma_r^2 N)$, 其中, $\mu_r = \frac{\Gamma(m+0.5)}{\Gamma(m)} (\frac{\Omega}{m})^{0.5}$, $\sigma_r^2 = \frac{\Omega}{m} \left(\frac{\Gamma(m+0.5)}{\Gamma(m)} \right)^2$ 。

根据 NOMA 技术原理, 在第一个时隙 U_r 需要依次解码 x_t 和 x_r , SINR 分别为

$$\gamma_{r \rightarrow t}^I = \frac{P_2 |\mathbf{h}_{SU_r}^T \mathbf{\Theta}_R \mathbf{h}_{BS}|^2}{P_1 |\mathbf{h}_{SU_r}^T \mathbf{\Theta}_R \mathbf{h}_{BS}|^2 + \sigma^2} \quad (10)$$

$$\gamma_r^I = \frac{P_1 |\mathbf{h}_{SU_r}^T \mathbf{\Theta}_R \mathbf{h}_{BS}|^2}{\sigma^2} \quad (11)$$

令 $G_r = |\mathbf{h}_{SU_r}^T \mathbf{\Theta}_R \mathbf{h}_{BS}|^2$, 同样地, G_t 可以重新表示为

$$G_r = |\mathbf{h}_{SU_r}^T \mathbf{\Theta}_R \mathbf{h}_{BS}|^2 = \beta_r \left| \sum_{n=1}^N |h_{BS,n}| |h_{SU_r,n}| \right|^2 \quad (12)$$

第二个时隙 DF 中继发送重新编码后的叠加信号 $s_2 = (\sqrt{P_3} x_r + \sqrt{P_4} x_t)$ 给两个用户, P_3 和 P_4 分别是中继对于 U_r 和 U_t 的发射功率。为确保接收节点之间的公平性, 假设 $P_4 > P_3$ 。通过理想 DF 中继, 第二个时隙 U_r 和 U_t 处的接收信号可以分别表示为

$$y_r^II = g_{RU_r} s_2 + n_r^II \quad (13)$$

$$y_t^II = g_{RU_t} s_2 + n_t^II \quad (14)$$

式中: $n_r^II, n_t^II \sim \mathcal{CN}(0, \sigma^2)$ 分别为第二个时隙 U_r 和 U_t 处的加性高斯白噪声。与第一个时隙类似, 第二个时隙 U_t 只需解码 x_t , 相应的 SINR 为

$$\gamma_t^II = \frac{P_4 |g_{RU_t}|^2}{P_3 |g_{RU_t}|^2 + \sigma^2} \quad (15)$$

令 $g_2 = |g_{RU_t}|^2$, 可知 $g_2 \sim \text{Gamma}(m, \Omega/m)$ 。

U_r 需要依次解码 x_t 和 x_r , 并且相应的 SINR 被定义为

$$\gamma_{r \rightarrow t}^II = \frac{P_4 |g_{RU_r}|^2}{P_3 |g_{RU_r}|^2 + \sigma^2} \quad (16)$$

$$\gamma_r^II = \frac{P_3 |g_{RU_r}|^2}{\sigma^2} \quad (17)$$

令 $g_3 = |g_{RU_r}|^2$, 可知 $g_3 \sim \text{Gamma}(m, \Omega/m)$ 。

2 系统性能优化

2.1 优化问题提出

为了两个用户 U_r 和 U_t 获得更好信息质量, 将在两个时隙内各自接收并解码出来的信息 x_r 和 x_t 进行最大比合并 (Maximal Ratio Combining, MRC), 得到在 U_r 和 U_t 处总的 SNR 分别为 $\gamma_r^{\text{tot}} = \gamma_r^I + \gamma_r^II$ 和 $\gamma_t^{\text{tot}} = \gamma_t^I + \gamma_t^II$ 。本文的目标是通过联合优化 $P_1, P_2, P_3, P_4, \beta_r$ 以及 β_t 的值来最大化该系统中两个用户的系统和速率, 即

$$R = \frac{1}{2} \log(1 + \min\{\gamma_r^R, \gamma_t^{\text{tot}}\}) + \frac{1}{2} \log(1 + \min\{\gamma_t^R, \gamma_r^{\text{tot}}\}) \quad (18)$$

除此之外, 需要满足以下 3 个功率约束: 式

(19) 为总功率约束;式(20)为 BS 最大发射功率约束;式(21)为 DF 中继最大发射功率约束。

$$P_1 + P_2 + P_3 + P_4 \leq P_t \quad (19)$$

$$P_1 + P_2 \leq P_B \quad (20)$$

$$P_3 + P_4 \leq P_R \quad (21)$$

为了保证 NOMA 场景下 SIC 的成功执行, γ_t^R 、 $\gamma_{r \rightarrow t}^I$ 、 $\gamma_{r \rightarrow t}^{II}$ 需要大于某个特定的门限值 γ_{th} , 即

$$\gamma_t^R = \frac{P_2 g_1}{P_1 g_1 + \sigma^2} \geq \gamma_{th} \quad (22)$$

$$\gamma_{r \rightarrow t}^I = \frac{P_2 G_r}{P_1 G_r + \sigma^2} \geq \gamma_{th} \quad (23)$$

$$\gamma_{r \rightarrow t}^{II} = \frac{P_4 g_3}{P_3 g_3 + \sigma^2} \geq \gamma_{th} \quad (24)$$

因此, 优化问题可描述为

$$\begin{aligned} & \max_{P_1, P_2, P_3, P_4, \beta_r, \beta_t} R \\ \text{s. t. } & P_j \geq 0, j=1, 2, 3, 4 \\ & \beta_r + \beta_t = 1 \\ & \text{式(19) ~ (24)} \end{aligned} \quad (25)$$

由于式(25)中的目标函数是非凸的, 且优化参数是高度耦合的, 这使得优化难以处理。

2.2 联合优化算法

针对优化问题(25), 本文提出一种以两阶段流量平衡的优化算法来逐渐逼近问题(25)的最优解。根据观察, 最大化公式(18)中的任意一个最小值都会使得其中包含的变量趋于一种平衡的结果, 否则会造成不必要的资源浪费。因此, 对于本文提出的 STAR-RIS 与中继集成的两阶段传输系统, 将两者看作是一个整体, 当该集成设备的“流入”和“流出”速率大致相同时可以实现更高的可实现速率。这种目标称为流量平衡。换言之, 当两个时隙的 SINR 相等时, 即中继处和用户处的 SINR 相等时, 系统和速率最大。利用流量平衡理论, 令 $\gamma_i^R = \gamma_i^{\text{tot}}, i \in \{r, t\}$, 根据公式推导可以得到

$$P_3 = \frac{g_1 - G_r}{g_3} P_1 \quad (26)$$

$$P_4 = \frac{(A-B)(CP_3+1)}{ABCP_1^2 + (A+B)CP_1 + C} P_2 \quad (27)$$

式中: $A = g_1/\sigma^2$; $B = G_r/\sigma^2$; $C = g_2/\sigma^2$ 。从上述两个式子中, 在给定 β_r 的情况下, 4 个功率值的关系得以确定, 但并不能计算出最优的值。为解决该问题, 本节利用功率值之间的相互关系设计了一个双层迭代

算法。

首先, 外层迭代的是能量分裂系数 β_r 。因为 $\beta_r \in [0, 1]$, 根据实际需求找到一个合适的能量分裂的步进值 β_δ , 每次循环 β_r 都会加上这个固定值, 然后迭代搜索最优值。另外, 为了满足公式(23)的 SIC 约束, β_r 显然不能等于 0, 因此需要从 β_δ 开始迭代。对于内层迭代来讲, 在给定 β_r 的前提下, 事先将 P_t 分为两部分, $P_1 + P_3$ 和 $P_2 + P_4$ 。为保证下行链路 NOMA 中的 SIC 过程平稳执行, 需要满足 $P_2 > P_1$ 以及 $P_4 > P_3$, 因此令功率初始值为 $P_1 + P_3 = 0, P_2 + P_4 = P_t$ 。也就是说, 只有较低的分支在一开始对整体可实现速率有贡献。此外, 设置 $P_{13} = P_1 + P_3, P_{23} = P_2 + P_4$, 以及一个小值 P_δ , 用于调整功率电平以逐渐满足式(26)和(27)。通过将 P_δ 与 P_{13} 相加, 并从 P_{23} 中减去 P_δ , 可以计算满足式(26)和(27)的功率值。得到 4 个功率值以后判断是否满足基站以及中继处的功率约束, 如果满足则计算得到 R 。每当发现较大的 R 时, 相应的功率值就会更新。最后, 由于式(22)、(23)和(24)中的 SIC 约束保证了 NOMA 的成功, 因此需要检查所得到的功率是否符合这些约束, 如果不满足, 就意味着内层迭代的结束。

基于两阶段通信流量平衡的双层迭代优化算法伪代码如下:

```

Input:  $P_B, P_R, P_\delta, \beta_\delta, \gamma_{th}, \sigma^2, g_1, g_2, g_3, G_r, G_t$ 
Output: optimal  $R^*, \beta_r^*, P_j^* (j=1, 2, 3, 4)$ 
1 Initialize  $\beta_r = \beta_\delta, R^* = R_{\text{new}} = R_{\text{old}} = 0$ 
2 repeat
3   Initialize  $P_1 = P_3 = 0, P_2 + P_4 = P_B + P_R$ 
4   repeat
5      $P_{13} = P_1 + P_3 + P_\delta$ 
6      $P_{24} = P_2 + P_4 - P_\delta$ 
7     Calculate  $P_1, P_2, P_3, P_4$  via (26), (27)
8     if  $P_1 + P_2 \leq P_B$  and  $P_3 + P_4 \leq P_R$ 
9       Update  $R_{\text{new}}$  via (18)
10      if  $R_{\text{new}} > R_{\text{old}}$  then
11         $R^* = R_{\text{new}}$ 
12         $R_j^* = R_j (j=1, 2, 3, 4), \beta_r^* = \beta_r$ 
13         $R_{\text{old}} = R_{\text{new}}$ 
14      end if
15    end if
16  until  $\gamma_t^R < \gamma_{th}$  of  $\gamma_{r \rightarrow t}^I < \gamma_{th}$  or  $\gamma_{r \rightarrow t}^{II} < \gamma_{th}$ 
17   $\beta_r = \beta_r + \beta_\delta$ 
18 until

```

对于所提的双层迭代联合优化算法, 其算法复

杂度与内层迭代的功率步长 P_δ 以及外层迭代的能量分裂系数步长 β_δ 有关,步长越大则复杂度越低。此外,由于内层迭代的结束标志为 SIC 约束,因此也会影响复杂度,SIC 过程的 SINR 门限值越大复杂度越低。

3 仿真结果与分析

本节提供数值结果来验证所提出的集成 STAR-RIS 和中继辅助通信系统的有效性。除非另有说明,基站位于二维平面的原点 $(0,0)$,而集成设备、用户 U_r 和 U_l 分别位于 $(30,0)\text{m}$ 、 $(30,-40)\text{m}$ 、 $(30,50)\text{m}$ 。假设系统中加性高斯白噪声的功率为 $\sigma^2 = -90\text{ dBm}$, $v = 0.001$, $\gamma_{\text{th}} = 0.01$, $P_\delta = 10^{-3}$, $\alpha = 2$, $P_t = 1\text{ W}$, $P_B = 0.8P_t$, $P_R = 0.6P_t$, $m = 2$, $\Omega = 2$ 。以下仿真结果均是在 NOMA 下开展的。

图 3 展示了本文提出的集成系统分别与纯中继和纯 STAR-RIS 系统在 N 变化下的性能比较。为了公平起见,考虑纯中继以及纯 STAR-RIS 系统都采用 NOMA 技术。对于集成系统和纯中继系统, $P_1 = P_3 = 0.2\text{ W}$, $P_2 = P_4 = 0.3\text{ W}$; 对于集成系统和纯 STAR-RIS 系统, $\beta_r = 0.5$, 且对于纯 STAR-RIS 系统中 BS 对于用户 U_r 和 U_l 分配的功率分别是 0.4 W 和 0.6 W 。从左图中可以看出所提出的集成系统是优于纯中继的;从右图中可以看出来,受限于 DF 中继的功耗以及半双工模式,集成系统的可实现速率之和相较于纯 STAR-RIS 系统随 N 的增大提升并不明显,但在 N 较小的情况下集成系统是优于纯 STAR-RIS 的,因为中继可以大大改善 STAR-RIS 带来的双路径损耗。也就是说,在元件个数受限的情况下,部署中继可以提升纯 STAR-RIS 系统的性能。

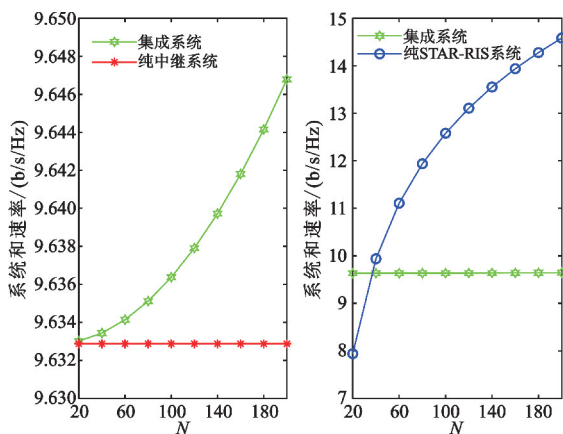


图 3 集成系统与纯中继系统和纯 STAR-RIS 系统对比

图 4 显示了本文提出的联合优化算法在不同 N 的条件下能量分裂系数 β_r 对于可实现速率的影响。从图中可以看出, β_r 越小,所提出的集成系统的性能越好。这是因为用户 U_r 比 U_l 离基站更近,因此 U_r 在执行 SIC 之后并不会收到来自用户 U_l 的干扰,且 U_r 的可实现速率应当是占两者之和的较大部分,而 U_l 则相反。根据对数函数的性质, U_l 的速率增加对总速率的增加贡献较大,因此 U_l 的速率越高则总的速率也会越大,故分裂给用户 U_l 的能量比例越大,系统性能会越好。需要说明的是,该结论只适用于功率不受限的情况下,因为在系统功率受限时 STAR-RIS 作用不大,分裂系数的值对系统的影响也就变得很小。另外,从图中可以看到随着 N 的增大,性能也会相应变好。

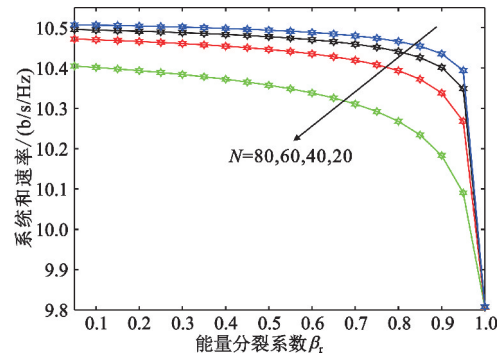


图 4 集成系统和速率随 STAR-RIS 能量分裂系数变化

图 5 对比了在不同 N 的情况下,本文所提的集成系统与纯 STAR-RIS 系统的可实现速率与 SNR 之间的关系, SNR 定义为 P_t/σ^2 , $\beta_\delta = 0.1$, $P_\delta = 10^{-2} \times P_t$ 。对于纯 STAR-RIS 系统,设定能量分裂系数为 0.5 , 用户 U_r 和 U_l 的功率分配比例为 $2:3$ 。从图中可以看出,对于两种方案, N 的增大都会提升性能,但是在低 SNR 或者 N 较小的情况下所提出的集成系统会优于纯 STAR-RIS 系统;相反,在高 SNR 和 N 较大的情况下,纯 STAR-RIS 的优势则突显出来,一方面是因为 STAR-RIS 的双路径损耗得以改善,另一方面是因为采用 HD 中继会导致带宽利用率降低。由此得出结论:在 P_t 或 N 比较小时, HD 中继可以为 STAR-RIS 辅助网络提供大的性能增益。这是因为此时网络处于功率受限状态,部署中继可以克服双路径损耗效应,从而获得更高的可实现速率。利用这一点,可以将本文提出的集成系统设置成两个模

式: 当 P_i 较低时, 可以激活中继, 而当 P_i 高的时候关闭中继, 通过模式匹配以实现灵活的覆盖增强效果以及高效率的能耗。

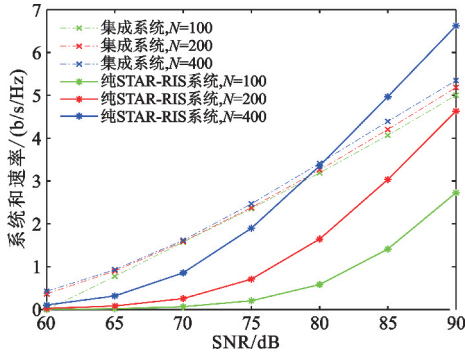


图 5 集成系统与纯 STAR-RIS 系统和速率随发射信噪比变化

图 6 对比了本文提出的集成系统与纯 FD 中继的性能, 可以看出 FD 中继的残余自干扰较大以及系统功率较低时, 集成系统有着较大的优势; 反之则纯 FD 中继性能更好, 主要是 FD 模式相较于 HD 模式还是有较大的性能优势, 但实现复杂度也会相应变大。在未来的工作中, 可以将集成系统中的 HD 中继换用 FD 中继以改善性能, 但也需考虑到 FD 中继的残余自干扰的影响。

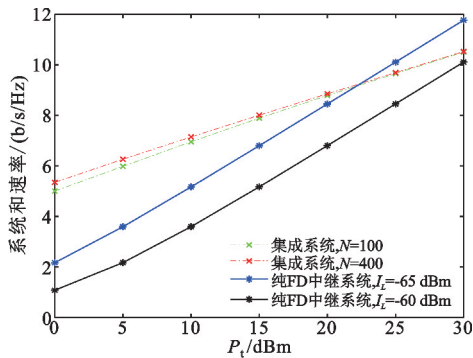


图 6 集成系统与纯 FD 中继系统和速率对比

在图 7 中通过改变集成设备的横坐标 x 来研究集成设备不同的部署位置对于性能的影响, 并考虑了不同 P_i 以及 N 场景, 集成设备、用户 U_r 和 U_l 的坐标分别为 $(x, 0)$ 、 $(50, -30)$ 、 $(50, 40)$, $v = 0.0001$ 。从图 7 中的上下两张图可以发现, 不同的 P_i 和 N 会导致系统性能随集成设备的部署变化趋势不一样。具体而言, 当 P_i 较低或者 N 较小时, 系

统总的可实现速率先升后降, 此时系统中中继影响较大, 类似于中继辅助系统, 当中继位于发端与收端中间位置时性能最优。而当 P_i 较高或者 N 较大时, 离基站越近性能越好, 此时系统中 STAR-RIS 发挥较大作用, 与 STAR-RIS 辅助系统靠近发端或收端性能最优的结论一致, 这也验证了本文所提集成系统以及算法的可行性。由于本文考虑到了信道估计的开销将 STAR-RIS 和中继进行集成, 未来考虑到用于 RIS 的信道估计技术会不断发展, 可以将两者进行独立部署来研究系统性能。

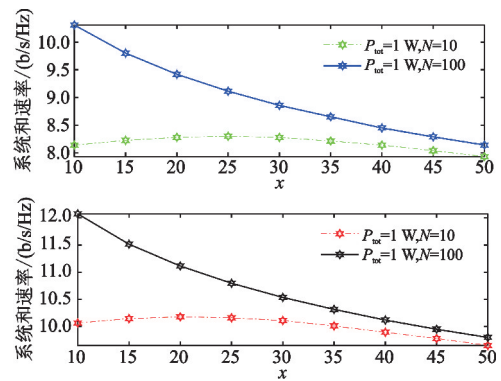


图 7 集成系统和速率随集成设备位置变化

4 结束语

针对传统 RIS 半空间覆盖以及双路径损耗的不足, 本文提出了一种用于下行 NOMA 的 STAR-RIS 和 DF 中继集成系统, 以实现覆盖增强的目标。仿真结果表明在系统功率受限或 STAR-RIS 元件数量受限的情况下, 该系统的可实现速率优于纯 STAR-RIS 辅助的系统, 即中继可以帮助 STAR-RIS 实现更好的性能。在功率不受限的情况下, STAR-RIS 对于分裂给反射用户的能量应尽可能小以保证系统总的可实现速率更大。由于本文将 STAR-RIS 与中继进行集成, 可以灵活地控制是否激活中继以实现在不同场景下的覆盖性能, 例如在功率不受限时可以关闭中继, 只使用 STAR-RIS 来辅助通信。最后仿真分析了集成设备在不同部署位置时的系统速率, 为将来实际部署提供了参考价值。

由于本文将 STAR-RIS 和 DF 中继进行集成, 未来需要考虑分开部署时的系统性能, 同时需要评估额外带来的信道估计开销。

参考文献:

- [1] ZHAO Y J, YU G H, XU H Q. 6G mobile communication networks: vision, challenges, and key technologies [J]. Scientia Sinica Informationis, 2019, 49(8): 963–987.
- [2] TARIQ F, KHANDAKER M R A, WONG K K, et al. A speculative study on 6G [J]. IEEE Wireless Communications, 2020, 27(4): 118–125.
- [3] RENZO M, ZAPPONE A, DEBBAH M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: how it works, state of research, and the road ahead [J]. IEEE Journal on Selected Areas in Communications, 2020, 38(11): 2450–2525.
- [4] ZENG S H, ZHANG H L, DI B Y, et al. Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: ris orientation and location optimization [J]. IEEE Communications Letters, 2021, 25(1): 269–273.
- [5] WU C Y, LIU Y W, MU X D, et al. Coverage characterization of STAR-RIS networks: NOMA and OMA [J]. IEEE Communications Letters, 2021, 25(9): 3036–3040.
- [6] IBRAHEEM S M, BEDAWY W, SAAD W, et al. Outage performance of NOMA-based DF relay sharing networks over nakagami-m fading channels [C] // 2018 13th International Conference on Computer Engineering and Systems. Cairo: IEEE, 2018: 512–517.
- [7] 季薇, 赵亚楠, 刘子卿, 等. 面向服务质量的 RIS 辅助的多用户 NOMA 系统功率分配方案 [J]. 电子与信息学报, 2023, 45(10): 3603–3611.
- [8] ABDULLAH Z, KISSELEFF S, MARTINS W A, et al. Cooperative hybrid networks with active relays and RISs for B5G: applications, challenges, and research directions [J]. IEEE Wireless Communications, 2024, 31(1): 126–132.
- [9] NEMATI M, PARK J, CHOI J. RIS-assisted coverage enhancement in millimeter-wave cellular networks [J]. IEEE Access, 2020, 8: 188171–188185.
- [10] HUANG C W, ZAPPONE A, ALEXANDROPOULOS G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication [J]. IEEE Transactions on Wireless Communications, 2019, 18(8): 4157–4170.
- [11] YILDIRIM I, KILINC F, BASAR E, et al. Hybrid RIS-empowered reflection and decode-and-forward relaying for coverage extension [J]. IEEE Communications Letters, 2021, 25(5): 1692–1696.
- [12] MU X D, LIU Y W, GUO L, et al. Simultaneously transmitting and reflecting (STAR) RIS aided wireless communications [J]. IEEE Transactions on Wireless Communications, 2022, 21(5): 3083–3098.
- [13] 吕铄, 嵇建波. 基于解码转发 CR-NOMA 协同系统物理层安全性能分析 [J]. 电讯技术, 2024, 64(1): 139–148.
- [14] XIE Z Y, YI W Q, WU X L, et al. STAR-RIS aided NOMA in multicell networks: a general analytical framework with gamma distributed channel modeling [J]. IEEE Transactions on Communications, 2022, 70(8): 5629–5644.

作者简介:

陈思宇 男, 1999 年生于河南焦作, 2021 年获学士学位, 现为硕士研究生, 主要研究方向为无线智能超表面。

申 滨 男, 1978 年生于贵州兴义, 2011 年获博士学位, 现为教授、博士生导师, 主要研究方向为下一代移动通信系统、LTE/LTE-A 系统、认知无线电系统等领域的信号处理理论与技术等。

元文军 女, 1997 年生于山西朔州, 2020 年获学士学位, 现为硕士研究生, 主要研究方向为图论、无线资源分配。

DOI:10.20079/j.issn.1001-893x.240304001

一种超宽带四臂螺旋钻孔雷达天线*

李佳颖¹, 刘建霞¹, 赵珍珍¹, 师振盛², 张安学²

(1. 太原理工大学 电子信息与光学工程学院, 太原 030600;

2. 西安交通大学 信息与通信工程学院, 西安 710049)

摘要:提出了一种基于电阻加载的超宽带四臂螺旋钻孔雷达天线。该天线在偶极子天线的基础上,采用 Wu-King 电阻加载方法,即天线沿纵向的电阻值依次增加,减少了天线的末端传输电流,使得天线的输入阻抗更加稳定,有效地提升了天线的带宽。同时,通过在天线臂末端设计螺旋结构,增加了天线臂上电流的流经路径,减小了天线的有效尺寸,并拓宽了低频带宽。该天线的外径为 40 mm,总长度为 574 mm。天线在 90.2~998 MHz 内驻波比小于 2.5,相对带宽高达 166.8%,接收脉冲波形拖尾极小。将天线放在洞穴里进行实际测量,测试结果与仿真结果较吻合。天线整体结构简单,加工方便,重量仅为 220.5 g,易于携带测试。因此,该天线可以较好地应用于钻孔雷达系统。

关键词:钻孔雷达;超宽带天线;电阻加载;四臂螺旋

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号:TN957.2 文献标志码:A 文章编号:1001-893X(2025)03-0477-07

An Ultra-wideband Quadrifilar Helix Borehole Radar Antenna

LI Jiaying¹, LIU Jianxia¹, ZHAO Zhenzhen¹, SHI Zhensheng², ZHANG Anxue²

(1. College of Electronic Information and Optical Engineering, Taiyuan University of Technology, Taiyuan 030600, China;

2. School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: An ultra-wideband quadrifilar helix borehole radar antenna based on resistive load is proposed. Developed from the dipole antenna, the antenna adopts the Wu-King resistance loading method to increase the antenna resistance sequentially along the longitudinal direction, which reduces the end transmission current of the antenna and effectively improves the bandwidth of the antenna. Meanwhile, the current flow path on the antenna arm is increased by designing a helical structure at the end of the antenna arm, resulting in the reduction of the antenna size and increase of the bandwidth. The outer diameter of the antenna is 40 mm, and the total length is 574 mm. The voltage wave ratio (VSWR) of the antenna is less than 2.5 in 90.2~998 MHz, the fractional bandwidth is up to 166.8%, and the received pulse waveform has very little trailing. Through the measurement of the fabricated antenna in the cave, there is a fair degree of agreement between the measured and simulated results. The overall structure of the antenna is simple, easy to process and weighs only 220.5 g, making it easy to carry around for testing. Therefore, the designed antenna can be better applied in the borehole radar system.

Key words: borehole radar; ultra-wideband (UWB) antenna; resistive load; quadrifilar helix

0 引言

雷达技术可用于检测地下的天然或人造异物,包括所有种类的埋地金属和非金属目标、高速公路

下的空洞以及地球的非匀质性。钻孔雷达(Borehole Radar)属于探地雷达的一种,由地面探地雷达发展而来,主要工作在地面以下的钻孔中。钻

* 收稿日期:2024-03-04;修回日期:2024-05-14

基金项目:国家重点研发计划(2021YFB3900102);山西省基础研究计划青年科学研究项目(20210302124171)

通信作者:刘建霞 Email:tyljx@163.com

孔雷达系统结构的工作原理与地面用探地雷达类似,都是通过电磁波进行地下介质探测^[1-10]。它通过在钻孔中放置天线,发射高频宽带电磁波并接收来自地下岩土介质的反射波,从而获取地下介质的结构特征。

天线作为雷达系统的关键部件,其性能的好坏对整个雷达探测系统有着极大的影响。伴随着越来越多的工程测试,超宽带、高辐射效率、低振铃效应、便携等诸多需求应运而生。较为经典的 TEM 喇叭天线、螺旋天线、碟形天线和双脊喇叭天线等具有宽带特性的天线被大量应用于探地雷达系统、车载雷达系统以及机载雷达系统^[11-17]。但是,钻孔雷达系统有限的尺寸限制了这些宽带天线的应用。目前,钻孔雷达系统中应用较多的天线有偶极子天线和 Vivaldi 等具有低剖面、体积小等特性的天线^[18-22]。文献[23]提出了一个工作在 0.6~4 GHz 的对偶 Vivaldi 天线,并将天线嵌入在含有金属反射器的圆柱形纤维塑料结构里,改善了低频时天线的方向性。文献[24]提出了一个半椭圆偶极子天线,通过在 V 形缝隙结构之间加载电阻,改善了时域特性,但是其带宽不是很宽(工作频段在 170~247.6 MHz)。文献[25]提出了一个锥形头的偶极子天线,通过 Wu-King 电阻加载技术使得天线频带得到拓宽,但是该天线太长,在实际应用时较为困难。

针对以上问题,本文提出了一种工作在 200 MHz 频率附近的超宽带四臂螺旋偶极子天线。采用 Wu-King 电阻加载技术,即沿天线纵向方向电流减弱,电流末端反射得到有效减少,拓展了带宽。通过在天线末端采取螺旋结构,延长电流路径,使得在天线整体尺寸保持不变的情况下低频带宽得到了拓宽。

1 四臂螺旋天线设计

1.1 四臂螺旋天线结构

本文所提出的钻孔雷达天线主要应用于煤岩界面的探测,工作环境大多是洞穴,因此所设计的四臂螺旋天线采取了偶极子天线类型。天线整体结构如图 1 所示,其中偶极子部分主要由金属偶极子部分、金属螺旋部分以及金属螺旋底部的回路部分组成,偶极子部分在仿真图里用灰色表示。图中,紫色的空心圆柱代表 PVC 固定管,加载电阻在仿真软件里是由蓝色的小圆锥体所代表,红色的圆锥体代表仿真时所用的离散馈电端口。经过 CST Studio Suite 电磁仿真软件优化得到的具体参数为 $L1=150\text{ mm}$, $L2=30\text{ mm}$, $L3=100\text{ mm}$, $L4=5\text{ mm}$, $w=6\text{ mm}$, $\text{gap}=2\text{ mm}$, $R=40\text{ mm}$ 。四臂螺旋天线总体长度 $L=574\text{ mm}$ 。

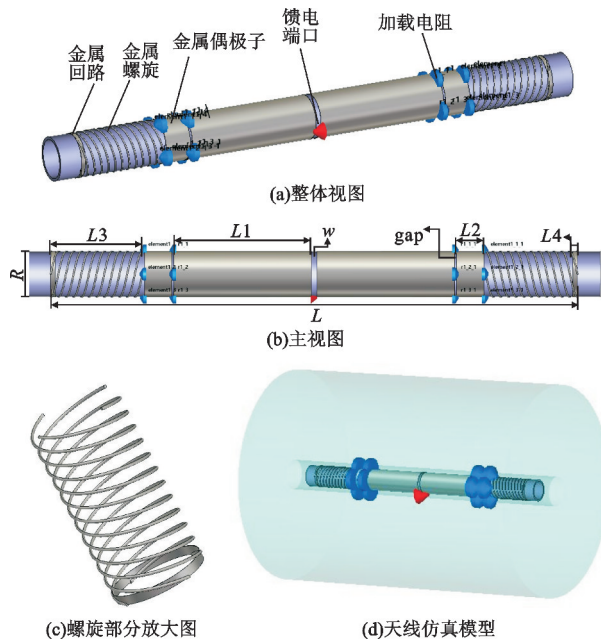


图 1 四臂螺旋钻孔雷达天线结构及仿真

为了更好地模拟验证天线在洞穴里实际工作环境,仿真时设置一个足够大的圆柱形介质块(圆柱形介质介电常数设置为 8),并在圆柱中心挖掉一个比天线略大一些的洞,模拟洞穴口,将天线放入模拟圆柱介质里进行仿真,最终仿真模型如图 1(d)所示。通过模拟真实环境进行仿真可以使天线与真实洞穴环境有一个良好的阻抗匹配,消除一些不必要的影响。

1.2 四臂螺旋天线设计原理

一般而言,偶极子天线归类于窄带天线,因此,为了符合钻孔雷达系统对高分辨率的工作需求,需要扩展偶极子天线的带宽。本文采用了 Wu-King 加载技术对天线带宽特性进行改善。

Wu-King 电阻加载理论是一种经典的无反射连续电阻加载方式。这种方法要求加载电阻按照一定的规律沿着天线的长度分布,需要的电阻数量多、阻值大,难以实际应用。为了简化实现难度, Sagnard 和 Fauchard 采用离散加载方式代替连续分布加载,不仅能扩展带宽还可以抑制拖尾波形。天线加载电阻的阻值由以下修正后的 Wu-King 公式确定:

$$R(z) = \frac{R_0}{\left(1 - \left|\frac{z}{h}\right|\right) + \left(1 - \left|\frac{z}{h}\right|\right)^2} \quad (1)$$

式中: $R_0 = \frac{\zeta_0 \Phi_0}{2\pi h}$, ζ_0 是自由空间的波阻抗, Φ_0 是频率和天线尺寸大小的函数,决定着所加载的电阻对电流的吸收效率,进一步影响天线的辐射效率,一般取

值在 3~11 之间。在实际应用中, 一些学者定义 Φ_0 为如下式子:

$$\Phi_0 = 2 \left[\ln \left(\frac{2h}{r} \right) - 1 \right] \quad (2)$$

最终由学者推导出来的阻抗加载公式为

$$Z'(z) = \frac{60\Phi}{h-|z|} \quad (3)$$

式中: h 为天线的臂长; Φ 为加载常数; z 为中心点到加载电阻位置的场长度。

更进一步, 为了满足工程对低频且小型化的需求, 在天线末端进行螺旋结构设计。螺旋结构的设计可以使天线偶极子臂上的电流路径得到延长, 在天线单元本身长度不变的情况下增加了电长度, 使得低频带宽得到拓宽。

螺旋部分结构参数由下式确定:

$$L_{ax} = N \sqrt{\frac{1}{N^2} (L_{ele} - A_{r_0})^2 - 4\pi^2 r_0^2} \quad (4)$$

式中: L_{ax} 为轴向长度; r_0 为半径; N 为圈数; L_{ele} 为螺旋的长度。通过公式计算和仿真优化最终确定 L_{ax} 为 100 mm, 螺旋圈数 N 为 3 圈。

2 天线仿真结果分析

最终, 通过仿真优化将天线分成 3 段进行电阻

加载, 被分开的两段天线之间各自通过加载电阻进行连接; 加载电阻会对天线表面上的电流产生影响, 进而使方向图发生畸变, 因此选择采用 4 个电阻并联的方式进行加载, 4 个电阻之间以 90° 为间隔进行排布。经上述公式计算以及 CST 软件仿真优化之后, 天线两端加载的电阻大小分别为 10Ω 和 19Ω (加载电阻阻值大小距离中心馈电点处依次增大)。由于 4 个电阻采取并联加载方式, 因此电阻对应的值为 40Ω 和 76Ω 。天线的回波损耗与驻波比结果如图 2(a) 所示, 可以看出天线 $|S_{11}|$ 小于 -8 dB 的频带为 160~675 MHz, 相对带宽约为 123.5%, 属于超宽带天线。

为了进一步验证所提出的四臂螺旋天线的物理特性, 图 2(b) 给出了 200 MHz 处四臂螺旋天线的电流分布, 可见天线末端螺旋臂段部分基本无反射电流, 因此, 天线近似呈行波分布。图 2(c) 是采用 CST 中的 200 MHz 的理想高斯信号对天线进行激励之后的辐射波形, 可以看出天线的时域波形拖尾几乎为 0, 且波形保真度较高, “W” 形较明显, 符合预期要求。图 2(d) 给出了在洞穴环境下天线的辐射效率仿真结果, 可以看出天线的辐射效率在 70%~90% 之间。

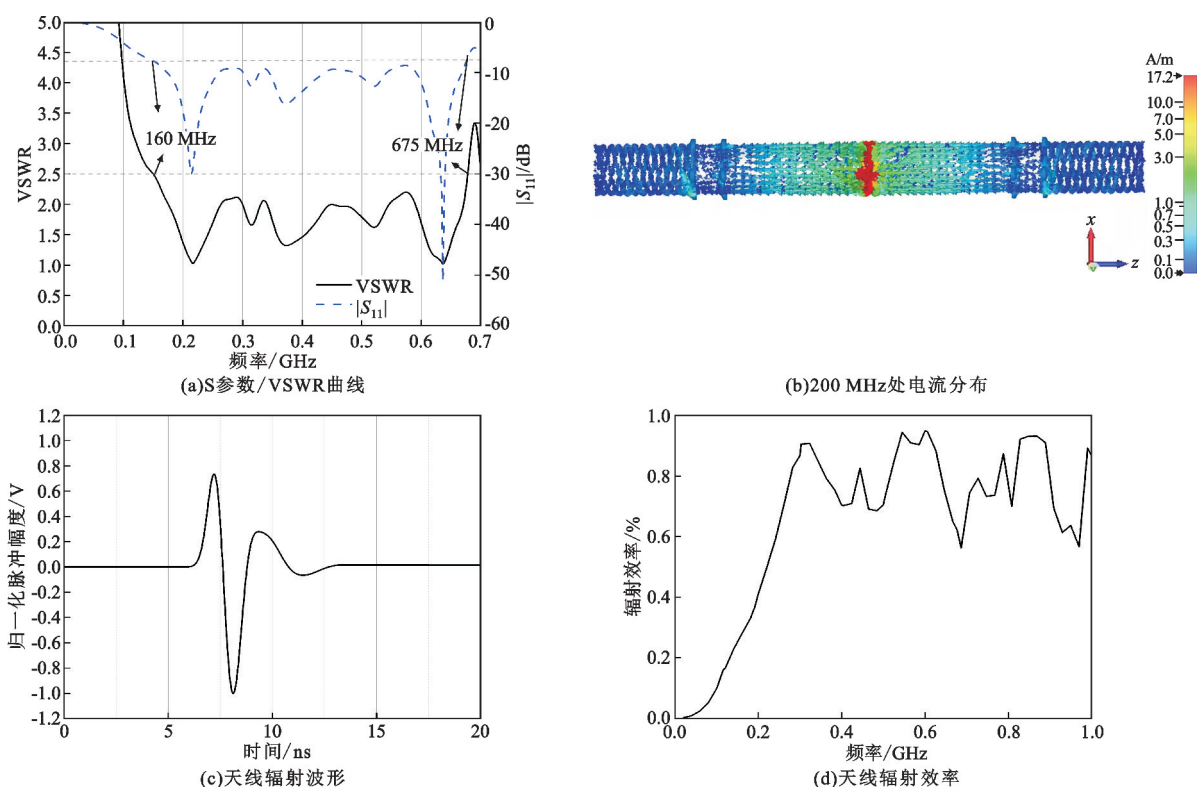


图 2 天线仿真结果

图 3 给出了天线在 200 MHz 处的方向图,可以看出经过电阻加载后方向图与偶极子天线类似,在 E 面呈“8”字形分布,在 H 面呈全向分布。因此,钻孔雷达全向发射天线的需求可以得到满足。

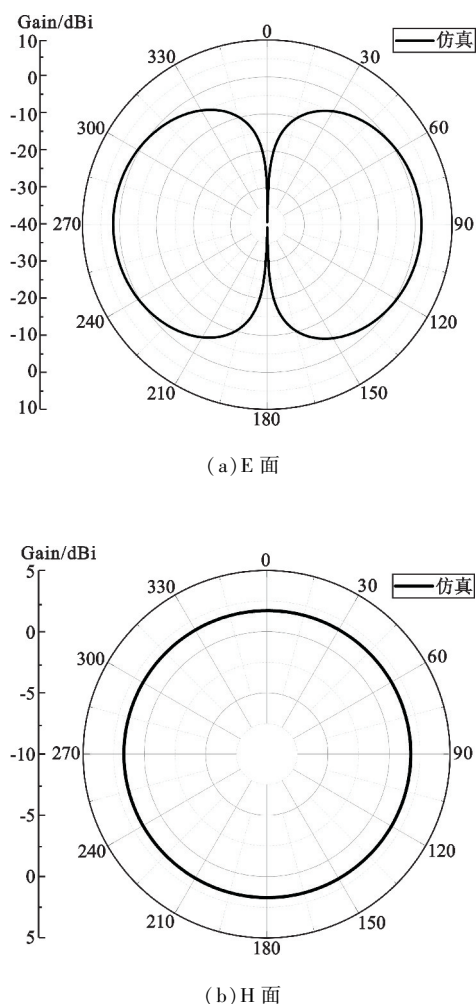


图 3 天线方向图

3 天线加工与测试

通过电磁仿真软件 CST Studio Suite 对天线的仿真模型优化后,将提出的超宽带四臂螺旋天线进行加工。考虑到目前市面上的钻孔雷达系统都存在一个普遍的问题就是重量过重,在场地测试时不方便使用。针对以上现实问题,该天线选择将铜箔缠绕在 PVC 固定管上,不仅大幅度降低了天线的整体重量,方便携带测试,而且降低了加工难度。加工出来的天线实物如图 4 所示,图片左上角放大图为电阻加载示意图。

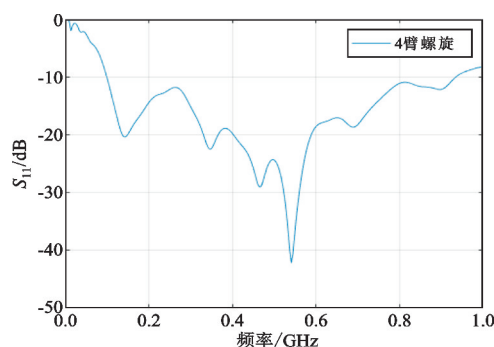


图 4 四臂螺旋天线实物

将天线放入提前钻好洞穴的土坡实验场地中进行测试,如图 5(a) 所示。该洞穴直径 1 m,深度 1.5 m。通过同轴线将天线馈电端口与便携式矢量网络分析仪 207235AE3750 连接起来。在测试之前,首先对矢量网络分析仪进行 SOLT 校准,校准之后再进行测试,测得的天线回波损耗曲线如图 5(b) 所示。可以看出在 90.2~998 MHz 频带内,天线的 $|S_{11}|$ 小于 -8 dB,相对带宽高达 166.8%,属于超宽带天线。从测试曲线可以看出,在洞穴环境下 $|S_{11}|$ 大部分频段都低于 -15 dB,说明该天线具有较宽的阻抗带宽,可保证天线辐射脉冲具有良好的波形特征。



(a) 测试场景



(b) 天线的回波损耗实测曲线

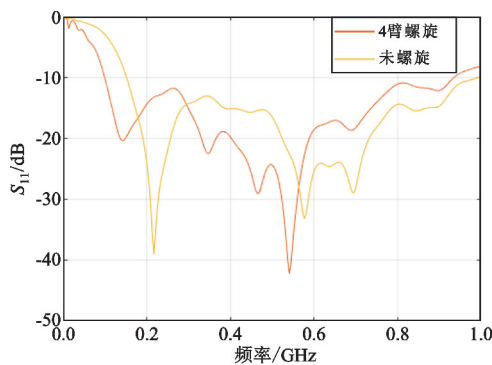
图 5 天线测试场景及响应结果

此外,为了对比螺旋结构对天线低频带宽的影响,将相同长度的只采取电阻加载的偶极子天线加工出实物进行对比,实物如图 6(a) 所示。将此天线也放入洞穴中与便携式矢量网络分析仪连接进行测试,测试出来的回波损耗如图 6(b) 所示,与四臂螺旋天线的回波损耗进行对比,可以看出天线带宽整

体向低频端移动,使得低频处的起始频率得到有效降低。因此,此螺旋结构在天线尺寸保持不变的情况下,通过延长电流路径实现了天线的小型化。



(a) 电阻加载天线实物



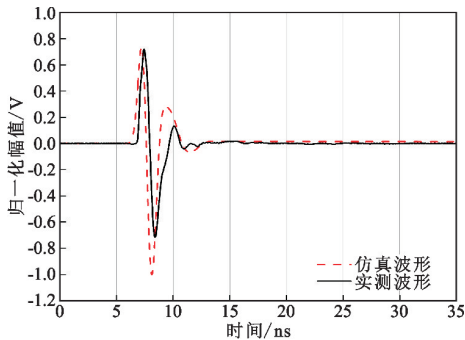
(b) 两个天线的回波损耗实测曲线

图 6 天线电阻加载实物以及测试曲线对比

将收发天线沿径向放置,采用 200 MHz 的脉冲信号进行激励,在图 7(a)所示的实验室环境下,测得的辐射波形如图 7(b)所示,可见实测波形与仿真波形吻合度良好,而且波形保真度较高,拖尾幅度小,波形的“W”形特征明显,因此满足钻孔雷达天线对时域波形的要求。



(a) 实验室测试环境



(b) 200 MHz 实测波形

图 7 天线时域波形测试

除此之外,对天线的方向图也进行了测试。如图 8 所示,将天线放在微波暗室里,并将中心馈电端口位置与激光中心对准。测试出来的二维方向图如图 9 所示,可以看出,在 3 个不同频点处 H 面为圆形,展示出了全向性,E 面为“8”字形,与仿真结果较吻合。

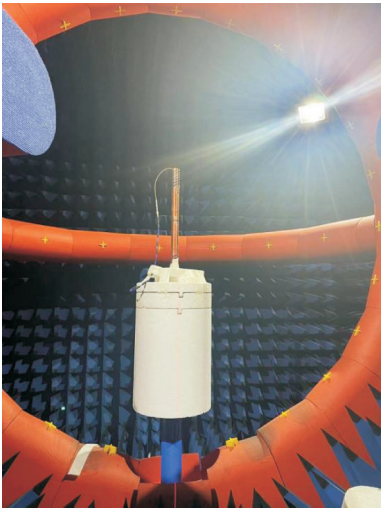


图 8 微波暗室测试

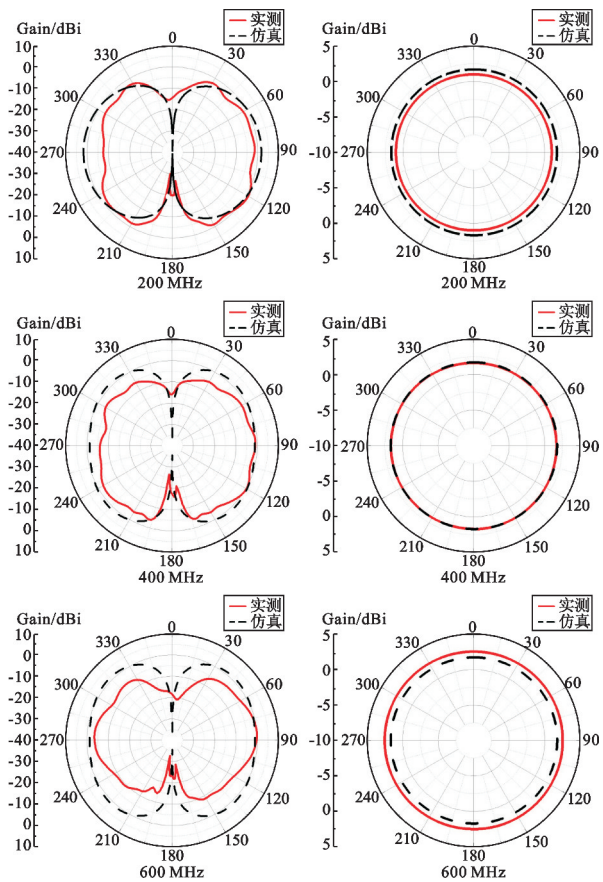


图 9 天线实测方向图

表 1 列出了其他文献用于探地雷达系统和钻孔
雷达系统所设计的天线相关特性对比,可以得知本文

设计的天线在尺寸、带宽和增益方面具有一定优势,
因此,在钻孔雷达系统应用中具有广阔的应用前景。

表 1 天线特性对比						
参考文献	天线类型	尺寸/ λ	起止频率/GHz	相对带宽/%	辐射效率/%	增益/dB
[17]	Vivaldi	2. 300 0 $\times$ 1. 730 0	0. 3000 0~2. 000 0	147. 8	未提供	11. 50
[19]	Vivaldi	1. 100 0 $\times$ 1. 390 0	0. 700 0~1. 200 0	100. 0	未提供	2. 00
[24]	偶极子	0. 350 0 $\times$ 0. 280 0	0. 170 0~0. 247 6	37. 2	85	0. 97
[25]	偶极子	0. 710 0 $\times$ 0. 022 0	0. 070 0~0. 260 0	115. 0	未提供	未提供
本文	偶极子	0. 380 0 $\times$ 0. 026 7	0. 098 8~0. 936 0	162. 0	38	6. 00

4 结束语

本文提出了一个处于洞穴环境下工作在 90. 2~998 MHz 频带范围内的基于电阻加载的超宽带四臂螺旋偶极子天线。通过 Wu-King 分布电阻加载以及末端螺旋结构的设计,天线不仅带宽得到拓宽,而且使得实现了小型化。天线较宽的阻抗带宽保证了时域波形具有较小的拖尾,波形保真度较高,符合预期结果。除此之外,电阻加载呈 90° 均匀分布,使得天线的方向图没有发生畸变,依然保持良好的全向特性,而且采取的加工方式简单且天线重量较轻,仅为 220. 5 g。因此,该天线可以较好地满足钻孔雷达系统的工程需求。

下一步,准备将该天线采样得到的信号经过窄脉冲等效采样电路系统处理后,观察雷达图像显示情况,然后再做进一步提升。

参考文献:

[1] SIEVER K, ELSER R, UCHTMANN S, et al. A decade of directional borehole radar services: from past to future [C]//2012 14th International Conference on Ground Penetrating Radar. Shanghai: IEEE, 2012: 906-910.

[2] OLSSON O, FALK L, FORSLUND O, et al. Borehole radar applied to the characterization of hydraulically conductive fracture zones in crystalline rock [J]. Geophysical Prospecting, 1992, 40(2): 109-142.

[3] 刘澜波, 钱荣毅. 探地雷达: 浅表地球物理科学技术中的重要工具 [J]. 地球物理学报, 2015, 58(8): 2606-2617.

[4] 朱兆荣, 赵守全, 秦欣, 等. 探地雷达在铁路隧道衬砌质量无损检测中的应用研究 [J]. 工程地球物理学报, 2021, 18(5): 703-708.

[5] ZHAO W K, FORTE E, FONTANA F, et al. GPR imaging and characterization of ancient Roman Ruins in the Aquileia Archaeological Park, NE Italy [J].

Measurement, 2018, 113: 161-171.

[6] YI L, ZOU L, SATO M. Practical approach for high-resolution airport pavement inspection with the Yakumo multistatic array ground-penetrating radar system [J]. Sensors, 2018, 18(8): 1-14.

[7] TEOH Y J, BRUKA M A, IDRIS N M, et al. Introduction of a ground penetrating radar system for subsurface investigation in Balik Pulau, Penang Island [J]. Journal of Physics: Conference Series, 2018, 995: 1-8.

[8] GUO L Y, YANG H L, ZHANG Q S, et al. A compact antipodal tapered slot antenna with artificial material lens and reflector for GPR applications [J]. IEEE Access, 2018, 6: 44244-44251.

[9] KITTITIANPUNYA C, KRAIRIKSH M. Object detection of ground-penetrating radar in the frequency domain using three-antenna system [J]. IEEE Transactions on Antennas and Propagation, 2022, 70(5): 3672-3679.

[10] YANG G Y, YE S B, JI Y C, et al. Radiation enhancement of an ultrawideband unidirectional folded bowtie antenna for GPR applications [J]. IEEE Access, 2020, 8: 182218-182228.

[11] SERHIR M, LESSELIER D. Wideband reflector-backed folded bowtie antenna for ground penetrating radar [J]. IEEE Transactions on Antennas and Propagation, 2018, 66(3): 1056-1063.

[12] DARVISH A, ZAKERI B, FALLAH P. 50-100 Ohms, 130 MHz-5 GHz resistively loaded bow-tie antenna for efficient UWB mono-pulse radiation [J]. AEU-International Journal of Electronics and Communications, 2021, 138: 1-11.

[13] NELLA A, SHARMA M, SRINIVASA RAO V, et al. A wideband high directional bow-tie antenna for automatic dependent surveillance-broadcast [C]//2022 IEEE Microwaves, Antennas, and Propagation Conference. Bangalore: IEEE, 2022: 33-37.

[14] KUMAR G, MEVADA P, CHAKRABARTY S, et al. Ultrawide band cage dipole antenna for ultra high

- frequency band ground penetrating radar system [J]. International Journal of RF and Microwave Computer-Aided Engineering, 2022, 32(6): 1-11.
- [15] ROJHANI N, PIERACCINI M, GOLAZARI S S. A compact TEM horn antenna for ground penetrating radar [C]//2018 International Conference on Advances in Computing, Communications and Informatics. Bangalore: IEEE, 2018: 1641-1645.
- [16] SUO Y, QI F X, LI W. Design of exponential gradient TEM horn antenna for ground penetrating radar [C]//2021 International Symposium on Antennas and Propagation. Taipei, China: IEEE, 2021: 1-2.
- [17] GUO J Y, TONG J S, ZHAO Q, et al. An ultrawide band antipodal Vivaldi antenna for airborne GPR application [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(10): 1560-1564.
- [18] NASSAR I T, WELLER T M. A novel method for improving antipodal Vivaldi antenna performance [J]. IEEE Transactions on Antennas and Propagation, 2015, 63(7): 3321-3324.
- [19] CHENG H Y, YANG H L, LI Y J, et al. A compact Vivaldi antenna with artificial material lens and sidelobe suppressor for GPR applications [J]. IEEE Access, 2020, 8: 64056-64063.
- [20] 杜荷. Vivaldi 天线及其工程设计 [D]. 西安: 西安电子科技大学, 2020.
- [21] LI X X, ZHOU H, GAO Z M, et al. Metamaterial slabs covered UWB antipodal Vivaldi antenna [J]. IEEE Antennas and Wireless Propagation Letters, 2017, 16: 2943-2946.
- [22] 王琼, 史春辉, 夏鑫淋, 等. 带引向器的非对称平衡对拓 Vivaldi 天线设计 [J]. 电讯技术, 2018, 58(1): 96-100.
- [23] GUO J Y, ZHAO Q, JIAO J, et al. An ultrawideband antipodal Vivaldi antenna for borehole radar application [C]//2017 9th International Workshop on Advanced Ground Penetrating Radar. Edinburgh: IEEE, 2017: 1-5.
- [24] TONG J S, GUO Y X, JI C G, et al. A compact wideband semi-elliptical dipole antenna for boat-borne GPR [J]. IEEE Transactions on Antennas and Propagation, 2022, 70(12): 11496-11504.
- [25] 赵泽哲, 葛良全, 刘成林. 钻孔雷达全向天线的设计与分析 [J]. 强激光与粒子束, 2013, 25(7): 1759-1762.

作者简介:

李佳颖 女, 1998 年生于河南安阳, 2020 年获学士学位, 现为硕士研究生, 主要研究方向为超宽带天线、频率选择表面。

刘建霞 女, 1970 年生于山西平遥, 2009 年获博士学位, 现为教授, 主要研究方向为微波电路与天线设计。

赵珍珍 女, 1991 年生于山西朔州, 2019 年获博士学位, 现为讲师, 主要研究方向为超表面在天线的应用、基于频率选择表面的电磁隐身。

师振盛 男, 1966 年生于山西平遥, 1989 年获硕士学位, 现为副教授, 主要研究方向为新型天线分析与设计、雷达信号处理及成像。

张安学 男, 1972 年生于河南安阳, 2003 年获博士学位, 现为教授, 主要研究方向为天线与微波技术、超宽带雷达。

DOI:10.20079/j.issn.1001-893x.240605008

基于深度半监督学习的目标检测技术综述*

何智杰^{1,2}, 肖 玮¹, 刘南清³, 高甲博¹, 柯学良¹, 曲乃铸¹

(1. 中国人民解放军陆军勤务学院, 重庆 401311; 2. 中国人民解放军 31680 部队, 四川 崇州 611233;
3. 西南交通大学 信息科学与技术学院, 成都 610031)

摘 要: 基于深度半监督学习的目标检测技术利用少量带标注信息的样本和大量无标注信息的样本进行模型训练, 可减少对标注样本的依赖, 提高准确性和效率。首先介绍了基于深度半监督学习的目标检测理论, 依据损失函数和模式设计方式的不同对其方法进行了分类, 然后基于 MS-COCO 和 Pascal VOC 数据集对典型方法进行了性能对比, 最后分析了其挑战和发展趋势, 旨在为相关研究提供参考。

关键词: 目标检测; 深度半监督学习; 半监督学习; 深度学习

开放科学(资源服务)标识码(OSID):



微信扫描二维码
听独家语音释文
与作者在线交流
享本刊专属服务

中图分类号: TP391.41 文献标志码: A 文章编号: 1001-893X(2025)03-0484-11

Survey of Object Detection Technology Based on Deep Semi-supervised Learning

HE Zhijie^{1,2}, XIAO Wei¹, LIU Nanqing³, GAO Jiabo¹, KE Xueliang¹, QU Naizhu¹

(1. PLA Army Logistics Academy, Chongqing 401311, China; 2. Unit 31680 of PLA, Chongzhou 611233, China;
3. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: The object detection technology based on deep semi-supervised learning uses a small number of samples with labeled information and a large number of samples without labeled information for model training, which reduces the dependence on labeled samples and improves the accuracy and efficiency. In this paper, the theory of object detection based on deep semi-supervised learning is introduced, and the methods are classified according to the different loss functions and pattern designs. Then, the performance of typical methods is compared based on MS-COCO and Pascal VOC data sets. Finally, their challenges and development trends are analyzed in order to provide a reference for related researches.

Key words: object detection; deep semi-supervised learning; semi-supervised learning; deep learning

0 引言

随着深度学习技术的发展, 卷积神经网络在精度上显著优于传统方法, 成为了最新的研究热点^[1], 但其检测精度依赖大量带标注信息的样本, 存在标注成本高和大量无标注信息的样本利用率低

的问题^[2]。受半监督学习思想的启发, 基于深度半监督学习的目标检测应运而生。其核心思想是使用深度学习模型, 利用少量带标注信息的样本和大量无标注信息的样本对模型进行训练, 在减少对标注样本依赖的同时具有更强的学习能力, 能够捕捉到

* 收稿日期: 2024-06-05; 修回日期: 2024-08-25

基金项目: 2023 年重庆市教委科学技术研究重点项目 (KJZD-K202312903); 2023 年度重庆市研究生科研创新项目 (CYS23778); 2024 年度重庆市研究生科研创新项目 (CYS240832)

通信作者: 肖玮 Email: 747386797@qq.com

更复杂的数据特征,适用于带标注信息的样本较少、无标注信息的样本较多的场景,为目标检测提供了新思路和方法,目前在医学影像检测^[3]和自动驾驶^[4]方面已经有了广泛应用。

当前,单独介绍目标检测和深度半监督学习的综述文献较多^[5-6],但鲜见系统而深入介绍基于深度半监督学习的目标检测综述性文献,也没有对各类方法进行分类,更没有对各类方法的基本思想、优缺点、典型算法和研究重点进行系统梳理总结。为此,本文在介绍基于深度半监督学习的目标检测理论的基础上,依据损失函数和模式设计方式的不同,将基于深度半监督学习的目标检测方法分为生成式方法、一致性正则化方法、基于图的方法、伪标记方法和混合方法,然后基于常用数据集对典型方法进行了性能对比,最后分析了其挑战和发展趋势,旨在为相关研究提供参考。

1 基于深度半监督学习的目标检测理论基础

基于深度半监督学习的目标检测是将深度学习、半监督学习和目标检测相结合,通过深度学习模型强大的特征提取能力,结合半监督学习对大量无标注信息的样本的利用,增强了模型的泛化能力,实现了对目标的准确识别和定位。深度学习通过模拟人脑的工作机制,构建深度神经网络结构,利用反向传播算法进行训练,并结合各种技术来改进模型的性能,实现对复杂数据的自动表示和学习。基于深度学习的目标检测是通过构建深度卷积神经网络并利用带标注信息的样本,模型可以自动学习从原始图像到高级抽象特征的映射,从而捕捉到目标的形状、纹理和上下文等关键信息。

基于深度半监督学习的目标检测则是利用半监督的方法来解决前者中带标注信息的样本少、大量无标注信息的样本利用率低的问题,与基于深度学习的目标检测的主要区别是输入中是否有无标注信息的样本和专门针对该样本的检测网络,其流程如图 1 所示。

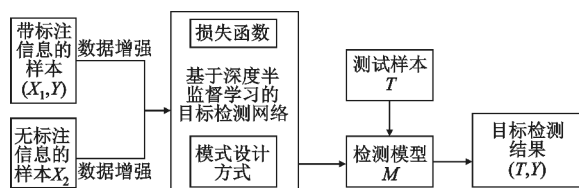


图 1 基于深度半监督学习的目标检测流程

2 基于深度半监督学习的目标检测方法

从半监督学习的角度出发,根据所采用的损失函数和模式设计方式^[7],可以将基于深度半监督学习的目标检测方法分为生成式方法、一致性正则化方法、基于图的方法、伪标记方法和混合方法。

2.1 生成式方法

生成式方法的基本思想是假设所有样本均来自同一潜在模型且分布一致,通过探索带标注信息的样本和无标注信息的样本中目标的真实分布,使用生成式对抗网络 (Generative Adversarial Network, GAN) 或自编码器生成新的样本或提取无标注信息的样本中的有用信息,最后将所有样本进行训练,得到最终检测模型,如图 2 所示。该类方法的优点是方法简单、易于实现,在带标注信息样本极少的情况下效果较好,但模型假设必须准确,生成的数据必须与真实数据分布吻合。典型的算法主要有基于 GAN 和变分自动编码器的方法。

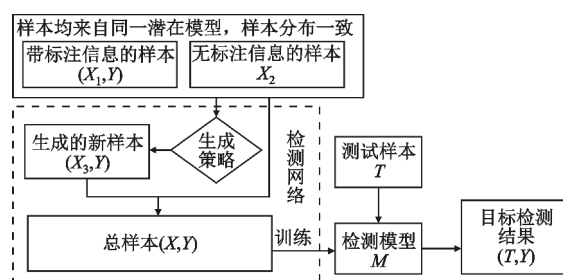


图 2 生成式方法示意

生成式方法目前的研究重点是如何得到最佳的生成策略,以产生更为精准的样本。文献[8]针对遥感领域数据量大、标注难和标注样本少容易出现拟合的问题,提出了一种基于 GAN 的深度半监督学习的目标检测方法,先后通过粗粒度的 GAN 和细粒度的 GAN 判别器设计了目标检测的网络模型。文献[9]提出了一种新的基于半监督学习主动显著性目标检测方法,通过训练变分自动编码器,选择最具有代表性的样本添加到标记池中。

2.2 一致性正则化方法

一致性正则化方法的基本思想是对带标注信息的样本训练后得到检测模型,而后再将无标注信息的样本中添加不同噪声后输入至检测模型,通过比较不同预测结果之间的一致性损失来更新迭代检测模型参数,如图 3 所示。该类方法的优点是对噪声和异常值鲁棒性较好,能够较好地减少过拟合,提高模

型的泛化能力,但计算复杂度增加,可能在加噪过程中引入误差。典型算法为 Mean Teacher^[10]。

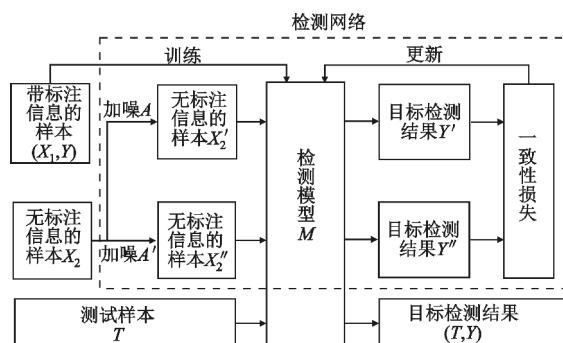


图3 一致性正则化方法示意

一致性正则化方法目前的研究重点是如何选择更为合适的加噪方式,以便训练出鲁棒性更强、泛化性更好的模型。文献[10]针对时序集成在学习大型数据集时会变得非常笨拙的问题,提出一种平均模型权重的方法,在训练时使用比时序集成更少的标签却达到了更快的学习速度和分类精度。文献[11]提出了一个改进的损失函数,即在 RPN (Region Proposal Network) 阶段生成的候选区域输入到教师模型和学生模型的 ROI (Region of Interest) head 中,通过比较生成的硬标签和软标签得到损失函数来更新模型的参数。

文献[12]针对数据、类别不平衡、噪声和误标注的问题,将算法的训练损失函数分为两部分,可以同时单阶段和双阶段检测器上工作。文献[13]针对分类任务的插值正则化直接应用于目标检测所带来的问题,提出不同类型基于插值的损失函数。文献[14]针对航空图像多方位物体较多的情况,利用自适应权值对每个伪标签预测对的方向进行一致性正则化和利用相似度进行正则化,在伪标签集和预测集之间建立了多对多的关系。文献[15]通过将图像块混合来增强模型的泛化能力,将特征块分离保持图像的局部结构信息,获得了更为丰富的图像特征。

2.3 基于图的方法

基于图的方法的基本思想是利用少量带标注信息的样本和大量无标注信息的样本构建相似性图,形成亲和矩阵,推断出无标注信息样本的标签信息,而后将带标注信息的样本和利用亲和矩阵进行标注的样本联合训练得到检测模型,如图4所示。该类方法的优点是直观且易于理解,灵活性强,但计算复杂,对所构建的图质量要求较高,在新加入数据时,

亲和矩阵需要重新构建。典型算法为基于图注意力网络(Graph Attention Network, GAT)。

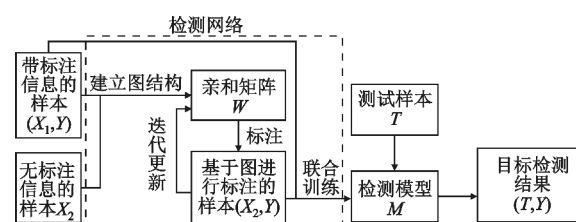


图4 基于图的方法示意

基于图的方法目前的研究重点是如何准确快速地构建出相似性图,得到亲和矩阵,以便获取更为精准的标注信息。文献[16]针对合成孔径雷达舰船目标标注样本不足的问题,提出基于 GAT 的方法,实现对无标记节点的分类,将超像素块定位到图像中并实现目标检测。文献[17]针对单一图模型存在的不够系统全面的问题,通过从同一张图像中的颜色、纹理、形状等方面来获得多视角特征数据并拓展到多图的情况,提出了基于多图模型的显著目标检测通用框架。

2.4 伪标记方法

伪标记方法的基本思想是首先通过带标注信息的样本生成一个初始检测模型,而后通过该模型对无标注信息的样本进行预测并经筛选后生成伪标签,最后利用带伪标签的样本与带标注信息的样本联合作为新的训练样本进行模型训练,如图5所示。该类方法的优点是能够较好地解决模型过拟合和泛化能力差的问题,应用最为广泛,灵活性强,但训练中的噪声对伪标签影响较大,置信度阈值的选择较为困难。根据伪标签生成时解决问题的方法侧重点,可进一步将该方法细分为改进自训练、改善类不平衡、基于稠密指导和优化伪标签4种方法。

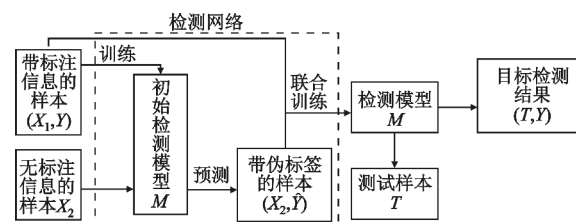


图5 伪标记方法示意

2.4.1 改进自训练方法

自训练^[18]是通过训练一个初始模型,为无标注信息的样本创建伪标签,最后将带标注信息的样本

和带伪标签的样本进行联合训练并得到目标检测模型。改进自训练方法的基本思想是对自训练的网络结构进行改进,通过提高伪标签质量来提升检测效果。该类方法的优点是能够充分发掘自训练方法的优势,提高检测精度,降低过拟合风险,但迭代过程相对复杂,对异常值比较敏感,对初始模型的依赖性强。典型算法有 STAC^[19]。

改进自训练方法目前的研究重点是如何在自训练框架的基础上提出更为简单有效的训练策略,提高训练效率和质量。STAC 在自训练基础上,对未标注信息的样本进行了强增强,而后引入了置信度阈值 τ 和非监督损失权重 loss 两个新的超参数,将训练结果经过非极大值抑制 (Non-maximum Suppression, NMS) 后用 τ 进行滤除,得到高质量的伪标签后输入模型进行训练。文献[20]提出了通过 Mean Teacher 方法增强的交互式自训练方法,使用 NMS 将最新检测结果与历史伪标签融合来提高伪标签的质量。文献[21]针对“教师-学生”模型中紧耦合效应导致的性能瓶颈,提出了循环自训练框架,用噪声伪标签进行稳健训练,避免累积确认误差。

2.4.2 改善类不平衡方法

改善类不平衡方法的基本思想是利用无标注信息中的大量背景类样本,通过调整模型的学习目标,使得模型更加关注于区分前景类样本和背景类样本,从而提升目标检测的精度。该类方法的优点是更关注于数据样本分布,类间的平衡利于更高质量伪标签的生成,但对参数比较敏感,计算复杂度较高。典型算法有 ACRST^[22]。

改善类不平衡方法目前的研究重点是如何缓解因类不平衡所带来的各类问题,从而生成更为优质的伪标签。文献[22]针对前景-背景和前景-前景不平衡的问题,提出一种新的记忆模块来缓解该不平衡问题并生成无偏伪标签。文献[23]为解决类不平衡引起的伪标签偏差问题,训练出逐渐改进的教师模型。文献[24]针对不同场景的图像存在输入数据分布偏移和类分布偏移的问题,通过估计相似性来处理样本分布差异和无标注的类别分布变化来校准有偏差的伪标签,有效处理域不一致并得到准确无偏的伪标签。文献[25]针对标签错配会导致严重的确认偏差问题,从分布级和实例级两个不同而又互补的角度提出一个有效的 LabelMatch 框架。文献[26]针对单阶段检测器因正样本较少而加剧类不平衡的问题,设计了一种动态分类的自适

应阈值策略,可以自动选择伪标签,实现质量和数量的最佳权衡。文献[27]针对类不平衡问题限制了深度半监督学习目标检测性能的问题,引入了一种自适应阈值机制和抖动模块,以过滤和细化边界框。

2.4.3 基于稠密指导方法

基于稠密指导方法的基本思想是利用数据增强后的图像作为不同的“通道”,模型对每个通道的输出结果进行预测,并根据预测结果生成稠密的伪标签,最后再利用一个阈值对伪标签进行挑选得到最终的伪标签。该类方法的优点是可以充分利用数据增强后的信息,更加全面地考虑图像的各种特征和变化,但生成的伪标签比较稠密,因此需要进行筛选和过滤,增加了计算的复杂度和时间成本。典型算法有 Dense Teacher^[28] 和 DTG^[29]。

基于稠密指导方法目前的研究重点是如何选择阈值对各个通道的输出结果进行挑选来生成更高质量的伪标签。文献[28]提出用稠密伪标签代替稀疏伪标签免除后处理,引入区域选择技术抑制稠密伪标签携带的噪声。文献[29]针对 Mean Teacher 方案中采用的从“稀疏到稠密”的范式使目标检测流程复杂化且容易累积噪声的问题,提出了一种新的“稠密到稠密”的范式,使模型中的“学生”可以得到“老师”充分、信息丰富和稠密的指导。文献[30]利用自适应滤波策略、稠密像素伪标签和不确定性一致性正则化项提高模型的泛化能力。

2.4.4 优化伪标签方法

优化伪标签方法的基本思想是通过优化策略来纠正伪标签的偏差,提高伪标签的质量。该类方法的优点是可以更好地处理噪声和异常值,但需要额外的计算成本,训练复杂度会增加。该类型常使用优化伪标签框架、软标签和抗噪学习等策略,典型算法主要有 Efficient Teacher^[31] 和 SIOD^[32]。

优化伪标签方法是当前研究的热点。该类型方法多样,效果较好,研究重点是如何设计或选择更好的优化策略来得到更为精准的伪标签。文献[31]针对基于锚的目标检测器缺乏生成高质量或灵活的伪标签结构而导致的严重不一致问题,引入了伪标签分配器,有效利用了可靠和不确定的伪标签。文献[32]基于相似度的伪标签生产模块和像素级组对比学习模块,提出了 Dual-mining 框架用于伪标签的生成。文献[33]针对检测过程中的自错误,通过同时训练两个具有不同初始参数的检测器,利用检测器间的差异自错误,并提出交叉校正机制来获得准确的伪标签。文献[34]针对单阶段检测器存在的严

重的筛选歧义性和样本歧义性的问题,提出了 JCE (Joint-confidence Estimation) 和 TSA (Task-separation Assignment) 两种通用的单阶段半监督检测模块,从而筛选出更高质量的伪标签。文献[35]提出双伪标签修正框架,充分利用上下文知识推断出更精准的伪标签。

文献[36]从难度、信息量和多样性方面主动采样信息最丰富、最多样化的无标注信息样本,并将其添加到带标签的示例集中,提高了伪标签的质量。文献[37]针对当前方法严重依赖从网络预测中过滤高质量伪标签的情况,提出了自适应融合跨尺度特征的混合尺度特征金字塔来挖掘伪标签并作出更好的预测。文献[38]针对基于深度半监督学习的图像分类方法不能很好地迁移到基于深度半监督学习的目标检测的实际和 STAC 方法在硬伪标签上的不足,使用 EMA (Exponential Moving Average) 动态更新教师模型并在提议区域上使用软标签来获得更可靠的伪标签。文献[39]将基于深度半监督学习的目标检测在无锚检测器上进行推广,并开发了一种新的相对不确定性的伪标签选择机制。

文献[40]针对长尾数据集上的标签噪声分布不一致和尾部难以分离出干净样本的特点,提出量化区域不确定性来进行抗噪声的目标检测。文献[41]针对标签中存在噪声的问题,构建两个分类头和一个蒸馏头进行集成学习,避免了对噪声标签的过度拟合,提高了对噪声的容忍度和鲁棒性。文献[42]针对自训练中存在的噪声,提出了清洁数据和噪声数据解耦、分类和定位任务解耦的双解耦训练框架。

文献[43]针对置信值不能准确过滤伪标签的问题,引入了边界框定位的分类任务,证明了无监督的边界框回归和类别分类对模型的训练有一定的贡献。文献[44]针对多阶段的训练策略严重依赖教师模型生成伪标签的质量的问题,提出了一个端到端的架构进行伪标签的生成和无标签数据的学习。文献[45]针对伪标签生成质量比较低的问题,首先将学生模型生成的候选框发送至教师模型进行“双重检查”,而后教师模型输出概率软标签用于学生模型的进一步训练。文献[46]针对 STAC 实现过程中复杂且效率低下的问题,提出了一种端到端的检测框架。该框架使用具有扩展“弱-强”数据增强的即时伪标签和协同校正方案,缓解了确认偏差问题并提高了伪标签的质量。

文献[47]为了充分利用带标注信息的样本,提

出了一种基于全局类原型的多实例对齐模型,增强了预测结果之间的一致性。文献[48]针对“教师-学生”模型过滤伪标签的不足,对分类头采用对象感知对比学习来改善伪标签,对回归头采取回归不确定性引导的伪标签来进行标签过滤。文献[49]用旋转框来替换水平框,针对航空图像尺度偏差问题引起的伪标签、标签分配和负学习不平衡问题,提出了尺度感应自适应阈值、尺度再平衡标签分配和教师引导的负性学习来保证模型无偏的学习。

2.5 混合方法

混合方法的基本思路是采用集成学习的方式,将多种基于深度半监督学习的目标检测方法进行组合应用,通过模型间相互的取长补短,获得更好的性能。该类方法的优点是可以同时利用一种或多种方法的优势,降低模型对特定数据的依赖性,但需训练多个模型,计算成本较高,需更多的时间来进行实验和调整参数,常见的组合是一致性正则化方法和伪标记方法的组合。典型算法有 PseCo^[50] 和 Consistent-Teacher^[51]。

混合方法目前的研究重点是如何选择在不增加模型复杂度的情况下,选择算法进行合理的组合集成。文献[50]针对伪标记中定位信息不准确的问题,提出了有噪伪标签边界框学习,设计了标签级和特征级一致性机制的多视图尺度不变学习策略,提升了模型的尺度不变性学习能力。文献[51]针对筛选伪标签存在的分配不一致、任务不一致和时序不一致等偏移与不稳定性问题,提出了自适应的标签分配、3D 特征对齐和基于高斯混合模型的自适应阈值 3 种模块。文献[52]针对伪标签的漏检问题,提出了混合伪标签和标签重采样方法,消除了假阴性样本的负面影响,检测性能提升明显。

文献[53]针对伪标签不准确带来的错误匹配和基于 DETR^[54] 检测器输入输出缺乏确定性对应关系的问题,提出了阶段混合匹配策略、跨视图查询一致性方法和一种新的伪标签挖掘模块,自适应地挖掘更可靠的伪标签以提高训练效率。文献[55]针对定位精度、多尺度图像和类不平衡问题,提出一种简单而有效地端到端的知识提取方法,引入了一致性正则化方法来处理大规模方差和定位精度差的问题。

除上述方法外,基于深度半监督学习的目标检测的研究热点还集中在 3D 领域、开放集条件下、与弱监督学习和语言模型相结合、基于迁移学习和基

于知识蒸馏等相关的目标检测领域^[56-69]。

文献[56]受自采样技术在深度半监督学习的分类任务的启发,提出了一个自采样半监督三维对象检测框架,通过设计扰动方案来增强网络的泛化能力。文献[57]针对在三维场景中存在的利用传统教师网络预测的伪标签不准确的问题,提出一种新的两阶段筛选过程。文献[58]针对单目 3D 目标检测中低召回伪标签的问题,提出了一种基于不确定性的过滤器来过滤低质量的噪声伪标签。文献[59]提出了一个半监督三维目标检测器,主要包含稠密伪标签生成策略和动态阈值的方法。

文献[60]针对开放集条件下未经整理的样本未有效利用的情况,提出了动态更新的类智能特征库并设计了自适应阈值。文献[61]针对现有基于深度半监督学习的目标检测由于语义扩展原因在开放集条件下性能增益较低的问题,提出了在线和离线 OOD(Out-of-Distribution)检测模块,提高了模型的鲁棒性和检测效果。文献[62]针对开放场景下目标不确定性而导致伪标签精度较低的问题,提出了一个端到端的框架,防止不确定性对模型的误导。

文献[63]针对对图像标签利用率不高的问题,

基于“教师-学生”框架和基于端到端的对象检测技术,提出了统一的架构,将无标注信息的样本、带标注信息的样本的和弱标注的样本都利用了起来。文献[64]和文献[65]针对当前算法没有充分利用点注释的问题,有效地使用深度半监督学习的方法和点标注的方法,大幅提升了检测器性能。

文献[66]针对大规模数据集标注的难点问题,提出了用视觉和语言模型中可用的丰富语义来定位和分类未标记图像中的对象并生成伪标签的方法。文献[67]根据检测新类别需要进行手动注释的问题,使用预训练的语言模型生成提示来指导模型的训练,使得模型能够在不进行重新训练的情况下适应新类别。

文献[68]运用了迁移学习的思想,通过视觉和语义知识转移来对伪标签进行校正和优化,提高了检测的准确性。文献[69]将原始的训练数据视为正样本,基于提示学习和基于真实边界框的知识蒸馏,提出了一种自我蒸馏的算法,显著提升了检测性能。

表 1 对以上方法进行了总结梳理,并提出了不同方法的适用场景,以便于开展工程应用。

表 1 深度半监督学习的目标检测方法

方法类型	优点	不足	适用场景	
生成式方法	方法简单,易于实现	模型假设必须准确,生成样本必须与真实样本分布吻合	适用于具有较大数据集,但带标注信息样本极少的情况	
一致性正则化方法	对噪声和异常值鲁棒性较好	计算复杂度增加,可能在加噪过程中引入误差	通常与数据增强结合使用,适用模型过拟合能力差和鲁棒性不强的情况	
基于图的方法	直观且易于理解,灵活性强	计算复杂,对所构建图质量要求较高,新加入数据时亲和矩阵需要重新构建	适用于对模型的可解释性有要求、数据之间存在复杂关系或结构、需要考虑全局一致性和利用样本间相似性的情况	
伪标记方法	自训练	可有效利用未标记数据,通过迭代过程逐步改进检测精度	迭代过程复杂,对异常值敏感,对初始模型的依赖性强	适用于初始模型性能较好但仍有提升空间,需持续更新和优化模型以及样本分布复杂或存在噪声的情况
	基于类不平衡	关注于数据样本分布,类间的平衡利于更高质量伪标签的生成	对网络参数比较敏感,计算复杂度较高	适用于数据集中类别分布不均匀、需要高精度检测少数类的情况
	基于稠密指导	充分利用数据增强后的信息,更全面地考虑图像的各种特征和变化	生成的伪标签比较稠密,因此需要进行筛选和过滤,增加了计算的复杂度和时间成本	适用于需要准确识别和定位图像中的稠密目标对象
	优化伪标签	更好地处理噪声和异常值	需要额外的计算成本,训练复杂度增加	适用于跨域目标检测,长尾效应较为严重的数据集和检测显著性较低的目标等情况
混合方法	同时利用一种或多种方法的优势	需训练多个模型,计算成本高	适用于算力比较充足且对模型复杂度要求不高的情况	

3 典型方法性能对比

为进一步客观评价,本节总结了在 MS-COCO^[70]和 Pascal VOC^[71]两个数据集上展开的基于深度半监督学习的目标检测算法,摘录了典型算法的性能数据进行对比。因生成式方法和基于图的方法在以上两个数据集应用时复杂度高,不易实现,故主要梳理了一致性正则化方法、伪标记方法和混合方法的性能参数。

为确保能够在同一基线上进行比较,方法性能对比中通常会使用一些基准算法及框架,主要有基于锚框的两阶段算法“Faster RCNN^[72]+FPN^[73]+ResNet-50^[74]”、基于锚框的单阶段算法“RetinaNet^[75]”“SSD^[76]”“YOLO 系列^[77-79]”、基于无锚框的算法“FCOS^[80]”和基于 Transformer 的算法“DETR”“DINO^[81]”。

3.1 MS-COCO 数据集的性能对比

MS-COCO 数据集重点关注目标在复杂背景下的检测定位能力、物体之间的上下文关系和精细化的定位,涉及的任务较难。为了能够全面反映复杂场景下算法的性能,常用平均精度模型(mean Average Precision, mAP)来衡量检测精度,表示为数据集中各类别目标的平均精度值(Average Precision, AP)的均值,其值越大,代表算法性能越好。通常随机抽取数据集中的 1%、2%、5% 和 10% 的数据作为带标注信息的样本,其余作为无标注信息的样本。图 6 中分别选取了各类方法中 mAP 值较大的部分算法进行比较,详细数据请用微信扫一扫本文的 OSID 码,在“本文开放的科学数据与内容”中查看。

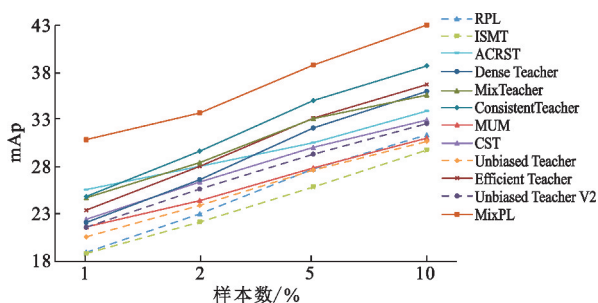


图 6 MS-COCO 数据集下部分方法性能结果

由图 6 可以看出,采用 MS-COCO 数据集,带标注信息的样本为 5% 和 10% 时,检测性能要明显好于 1% 和 2%,表明训练样本的数量越多,检测性能就越高。从性能指标看,混合方法最好,伪标记方法

次之,其中基于 DINO 框架的 MixPL 算法,引入了自注意力机制,整合了基于 Transformer 的算法的框架优势,已经成为了当前最先进的算法,但其网络参数量大,复杂度还相对较高,轻量化是未来的趋势之一。

3.2 Psacal VOC 数据集的性能对比

Psacal VOC 数据集重点关注不同目标类别的识别和定位,更注重分类的准确性和相对简单场景下的性能表现,常用 AP_{50} 和 $AP_{50:95}$ 来衡量检测精度,分别代表矩形框交并比(Intersection over Union, IoU) 阈值等于 0.50 和 0.50~0.95 时的 mAP 值。通常将 Psacal VOC07 训练集作为带标注信息的样本,Psacal VOC12 作为无标注信息的样本,相关性能指标请用微信扫一扫本文的 OSID 码,在“本文开放的科学数据与内容”中查看。

2022 年前的各类方法均采用了两阶段的 Faster RCNN 作为基础检测框架,近两年来逐渐向单阶段、无锚框和基于 Transformer 方向发展并取得了较好的效果,这与目标检测框架的发展趋势是相吻合的,说明将新的框架运用于基于深度半监督学习的目标检测是有效的,也为该方向的发展提供了思路。从复杂度上看,采用同一类检测框架的复杂度相差不大,采用不同的数据集所得到的复杂度差距也不大,整体复杂度受基础检测框架的影响较大。

采用 Psacal VOC 数据集,基于 FCOS 框架的算法,检测性能相比其他的数据集要稍好,说明数据集不同,各类算法的表现也不一。从性能指标上看,混合方法表现最好,伪标记方法次之,基于 YOLOv5 的 Efficient Teacher 算法复杂度最低,在 $AP_{50:95}$ 的表现也比较好,算法的实时性强。其中,优化伪标签中的 MixTeacher 算法的在 AP_{50} 表现也很好,从数量来看,也说明优化伪标签的方法成为了当前算法改进的主要方向,性能也得到了较大提高。

综上所述可以看出,由于混合方法能够集成各类方法的优点,在实际检测中性能较好;伪标记方法能够充分利用伪标签的特点,在检测中灵活性强,应用较多,得到了充分发展。这两种方法也是未来发展的主要方向。

4 面临挑战及发展趋势

4.1 当前面临的挑战及解决策略

近年来,虽然基于深度半监督学习的目标检测得到了快速发展,提出了很多新方法,但该领域仍面

临诸多挑战,主要表现在以下几个方面:一是在样本层面,主要在于如何处理无标注信息的样本中的噪声和不确定性。因这部分样本未经人工筛选,可能包含大量噪声和错误的信息,会对模型的训练产生干扰,降低模型的准确性。针对该问题,可以从数据增强、合适的数据清洗和采用弱监督等策略来降低噪声干扰,以此来提高数据质量。二是在模型层面,主要在于如何同时优化模型和提高模型的泛化能力。在训练过程中,模型的准确性和泛化能力不排除但存在一定的制衡因素,模型需要平衡两种类型的样本对模型的影响,在确保准确性的同时具备良好的泛化能力。针对该问题,可以从优化模型结构、改进训练策略、超参数调优和集成学习等方面综合入手,不断提高模型的准确度和泛化能力。三是在算法层面,主要在于如何在训练过程中有效利用无标注信息的样本来提升模型的性能,需要综合考虑样本的复杂性、分布特征来设计算法进行优化。针对该问题,可以采用优化伪标签、算法优化和即时教学等方法来提升模型的性能。四是在应用层面,主要在于如何选择合适的检测方法针对不同的应用场景和任务,以及如何针对特定的任务进行模型优化。针对该问题,可以从领域自适应、模型压缩和优化等方面来提高模型对场景和任务的适应性。

4.2 发展趋势

综上,基于深度半监督学习的目标检测技术在未来将会展现出更为广阔的前景。在算法创新方面,未来可在伪标注技术上进一步精细化,如引入更复杂的检测技术和无监督学习算法,结合丰富的上下文信息,提高伪标签的准确性和可靠性,减少噪声对模型训练的影响;混合方法将得到更加多样化的探索,以充分利用数据集的价值,促进模型性能的持续提升。在技术融合方面,跨任务联合优化将成为提升目标检测模型性能的重要途径。通过将目标检测与其他相关任务,如图像分割和姿态估计进行联合训练,可以充分利用不同任务之间的互补信息,增强模型的泛化能力和适应性。此外,云计算与大数据技术的深度融合,将为处理海量数据提供强有力的支持。利用云计算平台的高性能计算能力和大数据技术的数据挖掘能力,可以实现数据的快速处理、分析和模型优化,进一步推动目标检测技术的创新与发展。在实际应用方面,在军事装备、智慧医疗、自动驾驶、智能安防、智慧城市和工业自动化等领域,有诸多符合基于深度半监督学习的目标检测的相关场景,其应用范围也将得到极大拓展。

参考文献:

- [1] 李翠锦,瞿中. 复杂交通环境下多层交叉融合多目标检测[J]. 电讯技术, 2023, 63(9): 1291-1299.
- [2] 刘春磊,陈天恩,王聪,等. 小样本目标检测研究综述[J]. 计算机科学与探索, 2023, 17(1): 53-73.
- [3] 叶超. 基于半监督学习的特定场景目标检测[D]. 广州:华南理工大学, 2022.
- [4] 文海名,童孟军. 基于半监督学习的自动驾驶场景下的目标检测[J]. 微电子学与计算机, 2023, 40(2): 22-36.
- [5] 李柯泉,陈燕,刘佳晨,等. 基于深度学习的目标检测算法综述[J]. 计算机工程, 2022, 48(7): 1-12.
- [6] 韩嵩,韩秋弘. 半监督学习研究的述评[J]. 计算机工程与应用, 2020, 56(6): 19-27.
- [7] YANG X L, SONG Z X, KING I, et al. A survey on deep semi-supervised learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(9): 8934-8954.
- [8] 杜泽星,殷进勇,杨建. 基于半监督学习的遥感飞机图像检测方法[J]. 激光与光电子学进展, 2020, 57(6): 123-131.
- [9] LV Y Q, LIU B W, ZHANG J, et al. Semi-supervised active salient object detection[J]. Pattern Recognition, 2022, 123: 1-13.
- [10] TARVAINEN A, VALPOLA H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results [C]//The 31th Annual Conference on Neural Information Processing Systems. Long Beach: MIT Press, 2017: 1196-1205.
- [11] LI H D, WU Z X, SHRIVASTAVA A, et al. Rethinking pseudo labels for semi-supervised object detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2): 1314-1322.
- [12] JEONG J, LEE S, KIM J N, et al. Consistency-based semi-supervised learning for object detection [C]//The 33th Conference on Neural Information Processing Systems. Vancouver: MIT Press, 2019: 10759-10768.
- [13] JEONG J, VERMA V, HYUN M, et al. Interpolation-based semi-supervised learning for object detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 11597-11606.
- [14] HUA W, LIANG D K, LI J Y, et al. SOOD: towards semi-supervised oriented object detection [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 15558-15567.
- [15] KIM J, JANG J, SEO S, et al. MUM: mix image tiles and UnMix feature tiles for semi-supervised object detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 14492-14501.
- [16] 吕进东,王彤,唐晓斌. 基于图注意力网络的半监督

- SAR 舰船目标检测[J]. 电子与信息学报, 2023, 45(5): 1541-1549.
- [17] 肖云. 基于图的半监督学习与排序模型的视觉显著目标检测研究[D]. 合肥: 安徽大学, 2019.
- [18] SCUDDER H. Probability of error of some adaptive pattern-recognition machines[J]. IEEE Transactions on Information Theory, 1965, 11(3): 363-371.
- [19] SOHN K, ZHANG Z Z, LI C L, et al. A simple semi-supervised learning framework for object detection[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2005.04757v2>.
- [20] YANG Q Z, WEI X H, WANG B, et al. Interactive self-training with mean teachers for semi-supervised object detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 5937-5946.
- [21] LIU H, CHEN B, WANG B, et al. Cycle self-training for semi-supervised object detection with distribution consistency reweighting [C]//The 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022: 6569-6578.
- [22] ZHANG F Y, PAN T X, WANG B. Semi-supervised object detection with adaptive class-rebalancing self-training [C]//The 36th AAAI Conference on Artificial Intelligence. Vancouver: AIAA, 2022: 3252-3261.
- [23] LIU Y C, MA C Y, HE Z, et al. Unbiased teacher for semi-supervised object detection [C]//The 9th International Conference on Learning Representations. Vienna: IEEE, 2021: 1-17.
- [24] YU L H, ZHANG Y F, HONG L Q, et al. Dual-curriculum teacher for domain-inconsistent object detection in autonomous driving[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2210.08748v1>.
- [25] CHEN B B, CHEN W J, YANG S C, et al. Label matching semi-supervised object detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 14361-14370.
- [26] ZHANG Y M, YAO X X, LIU C, et al. S4OD: semi-supervised learning for single-stage object detection[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2204.04492v1>.
- [27] KAR P, CHUDASAMA V, ONOE N, et al. Revisiting class imbalance for end-to-end semi-supervised object detection [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver: IEEE, 2023: 4570-4579.
- [28] ZHOU H Y, GE Z, LIU S T, et al. Dense teacher: dense pseudo-labels for semi-supervised object detection[C]//Computer Vision - ECCV 2022. Cham: Springer, 2022: 35-50.
- [29] GANG L, XIANG L, WANG Y J, et al. DTG-SSOD: dense teacher guidance for semi-supervised object detection [C]//The 36th Conference on Neural Information Processing Systems. New Orleans: MIT Press, 2022: 1-15.
- [30] CHEN B H, LI P Y, CHEN X, et al. Dense learning based semi-supervised object detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 4805-4814.
- [31] XU B W, CHEN M T, GUAN W L, et al. Efficient teacher: semi-supervised object detection for YOLOv5 [EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2302.07577v3>.
- [32] LI H J, PAN X J, YAN K, et al. SIOD: single instance annotated per category per image for object detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 14177-14186.
- [33] MA C C, PAN X J, YE Q X, et al. CrossRectify: leveraging disagreement for semi-supervised object detection[J]. Pattern Recognition, 2023, 137: 1-5.
- [34] LIU C, ZHANG W M, LIN X R, et al. Ambiguity-resistant semi-supervised learning for dense object detection [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 15579-15588.
- [35] ZHANG L, SUN Y X, WEI W. Mind the gap: polishing pseudo labels for accurate semi-supervised object detection [C]//The AAAI Conference on Artificial Intelligence. Vienna: AIAA, 2023: 3463-3471.
- [36] MI P, LIN J H, ZHOU Y Y, et al. Active teacher for semi-supervised object detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 14462-14471.
- [37] LIU L, ZHANG B S, ZHANG J N, et al. MixTeacher: mining promising labels with mixed scale teacher for semi-supervised object detection [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7370-7379.
- [38] TANG Y H, CHEN W F, LUO Y J, et al. Humble teachers teach better students for semi-supervised object detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 3131-3140.
- [39] LIU Y C, MA C Y, KIRA Z. Unbiased teacher v2: semi-supervised object detection for anchor-free and anchor-based detectors [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 9809-9818.
- [40] WANG Z Y, LI Y L, GUO Y, et al. Combating noise:

- semi-supervised learning by region uncertainty quantification [C]//The 35th Conference on Neural Information Processing Systems. [S. l.]: MIT Press, 2021:9534–9545.
- [41] GAO J Y, WANG J, DAI S Y, et al. NOTE-RCNN: noise tolerant ensemble RCNN for semi-supervised object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul:IEEE,2019:9507–9516.
- [42] ZHENG S D, CHEN C S, CAI X W, et al. Dual decoupling training for semi-supervised object detection with noise-bypass head [C]//The AAAI Conference on Artificial Intelligence. Vancouver:AAAI,2022:3526–3534.
- [43] ROSSI L, KARIMI A, PRATI A. Improving localization for semi-supervised object detection [C]//Image Analysis and Processing- ICIAP 2022. Cham: Springer International Publishing,2022:516–527.
- [44] XU M D, ZHANG Z, HU H, et al. End-to-end semi-supervised object detection with soft teacher[C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal:IEEE,2021:3040–3049.
- [45] WANG K, NIE Y X, FANG C W, et al. Double-check soft teacher for semi-supervised object detection [C]//The Thirty-First International Joint Conference on Artificial Intelligence. Vienna:IEEE,2022:1430–1436.
- [46] ZHOU Q, YU C H, WANG Z B, et al. Instant-teaching: an end-to-end semi-supervised object detection framework [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE,2021:4079–4088.
- [47] LI A X, YUAN P, LI Z G. Semi-supervised object detection via multi-instance alignment with global class prototypes [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE,2022:9799–9808.
- [48] CHOI H, CHEN Z X, SHI X P, et al. Semi-supervised object detection with object-wise contrastive learning and regression uncertainty [EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2212.02747v1>.
- [49] ZHANG R X, XU C, XU F, et al. Rethinking scale imbalance in semi-supervised object detection for aerial images[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2310.14718v1>.
- [50] LI G, LI X, WANG Y J, et al. PseCo: pseudo labeling and consistency training for semi-supervised object detection [M]//Computer Vision-ECCV 2022. Cham: Springer,2022:457–472.
- [51] WANG X J, YANG X Y, ZHANG S L, et al. Consistent-teacher: towards reducing inconsistent pseudo-targets in semi-supervised object detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver:IEEE,2023:3240–3249.
- [52] CHEN Z M, ZHANG W W, WANG X J, et al. Mixed pseudo labels for semi-supervised object detection[EB/OL]. [2024-06-01]. <https://arxiv.org/abs/2312.07006v1>.
- [53] ZHANG J C, LIN X R, ZHANG W, et al. Semi-DETR: semi-supervised object detection with detection transformers [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE,2023:23809–23818.
- [54] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [M]//Computer Vision-ECCV 2020. Cham: Springer International Publishing,2020:213–229.
- [55] GUO Q S, MU Y, CHEN J Y, et al. Scale-equivalent distillation for semi-supervised object detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans:IEEE,2022:14502–14511.
- [56] ZHAO N, CHUA T S, LEE G H. SESS: self-ensembling semi-supervised 3D object detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle:IEEE,2020:11076–11084.
- [57] WANG H, CONG Y Z, LITANY O, et al. 3DIoUMatch: leveraging IoU prediction for semi-supervised 3D object detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE,2021:14610–14619.
- [58] YANG L, ZHANG X Y, LI J, et al. Mix-teaching: a simple, unified and effective semi-supervised learning framework for monocular 3D object detection[J]. IEEE Transactions on Circuits and Systems for Video Technology,2023,33(11):6832–6844.
- [59] LI J Y, LIU Z, HOU J H, et al. DDS3D: dense pseudo-labels with dynamic threshold for semi-supervised 3D object detection [C]//2023 IEEE International Conference on Robotics and Automation. London:IEEE, 2023:9245–9252.
- [60] LIU N Q, XU X, GAO Y J, et al. Semi-supervised object detection with uncurated unlabeled data for remote sensing images [J]. International Journal of Applied Earth Observation and Geoinformation,2024,129:1–10.
- [61] LIU Y C, MA C Y, DAI X L, et al. Open-set semi-supervised object detection [C]//Computer Vision-ECCV 2022. Cham:Springer,2022:143–159.
- [62] ZHUANG J Y, WANG K, LIN L, et al. Credible teacher for semi-supervised object detection in open scene[C]//2024 IEEE International Conference on Acoustics, Speech and Signal Processing. Seoul:IEEE,2024:5135–5139.
- [63] WANG P, CAI Z W, YANG H, et al. Omni-DETR: omni-supervised object detection with transformers[C]//2022

- IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans;IEEE,2022:9357-9366.
- [64] CHEN L Y, YANG T, ZHANG X Y, et al. Points as queries: weakly semi-supervised object detection by points [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville; IEEE, 2021: 8819-8828.
- [65] ZHANG S L, YU Z R, LIU L Y, et al. Group R-CNN for weakly semi-supervised object detection with points [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans;IEEE,2022:9407-9416.
- [66] ZHAO S Y, ZHANG Z X, SCHULTER S, et al. Exploiting unlabeled data with vision and language models for object detection [C]//Computer Vision-ECCV 2022. Cham;Springer,2022:159-175.
- [67] FENG C J, ZHONG Y J, JIE Z Q, et al. PromptDet: towards open-vocabulary detection using uncurated images [C]//Computer Vision-ECCV 2022. Cham; Springer,2022:701-717.
- [68] GUO J Y, HAN K, WU H, et al. Positive-unlabeled data purification in the wild for object detection [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville;IEEE,2021:2652-2661.
- [69] TANG Y X, WANG J, GAO B Y, et al. Large scale semi-supervised object detection using visual and semantic knowledge transfer [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas; IEEE,2016:2119-2128.
- [70] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//Computer Vision-ECCV 2014. Cham; Springer International Publishing,2014:740-755.
- [71] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88 (2):303-338.
- [72] REN S Q, HE K M, 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, 39(6):1137-1149.
- [73] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu;IEEE,2017:936-944.
- [74] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas;IEEE,2016:770-778.
- [75] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2):318-327.
- [76] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]//Computer Vision-ECCV 2016. Cham;Springer International Publishing,2016:21-37.
- [77] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas;IEEE,2016:779-788.
- [78] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu; IEEE, 2017: 6517-6525.
- [79] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. [2024-06-01]. <https://arxiv.org/abs/1804.02767v1>.
- [80] TIAN Z, SHEN C H, CHEN H, et al. FCOS: fully convolutional one-stage object detection [C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul;IEEE,2019:9626-9635.
- [81] CARON M, TOUVRON H, MISRA I, et al. Emerging properties in self-supervised vision transformers [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal;IEEE,2021:9630-9640.

作者简介:

何智杰 男,1987 年生于山西清徐,2010 年获学士学位,现为硕士研究生、工程师,主要研究方向是半监督学习、目标检测。

肖 玮 女,1982 年生于重庆,2012 年获博士学位,现为副教授、硕士生导师,主要研究方向为物联网工程与系统、计算机视觉。

刘南清 男,1994 年生于湖北荆州,2019 年获硕士学位,现为博士研究生,主要研究方向为遥感图像的目标检测。

高甲博 男,1995 年生于陕西大荔,2023 年获硕士学位,现为助理工程师,主要研究方向为深度强化学习。

柯学良 男,1997 年生于安徽黄山,2020 年获学士学位,现为硕士研究生,主要研究方向为深度学习。

曲乃铸 男,1992 年生于黑龙江讷河,2015 年获学士学位,现为硕士研究生、讲师,主要研究方向为深度学习与集群控制。

稿 约

《电讯技术》(月刊)创刊于1958年,由中国电子科技集团有限公司主管、中国电子科技集团公司第十研究所主办,系国内外公开发行的、理论与应用相结合的中文核心期刊、中国科技核心期刊、中国核心学术期刊和中国应用型权威期刊,也是国家新闻出版广电总局第一批认定的学术期刊,以及信息通信领域和电子技术、通信技术领域高质量科技期刊。

本刊主要栏目:应用基础与前沿技术;电子与信息工程;综述与评论。

本刊征稿内容:

(1)应用基础与前沿技术(主要指国家自然科学基金、国家重大科技专项、国家重点研发计划等项目以及各级重点实验室产出的电子与信息领域的应用基础研究(为了实现某一特定的或具体的应用目的或目标而获取应用原理、规律的新知识所进行的独创性研究)和前沿技术研究(高技术领域中具有前瞻性、先导性和探索性的重大技术研究));

(2)电子与信息工程(主要指网络信息体系、航空电子、通信与数据链、飞行器测控通信、导航与定位、卫星应用、情报与侦察监视、目标识别、电磁频谱战、大数据与云计算、自主化与智能化、微系统以及毫米波、太赫兹、光电子、微波光子学、量子等领域的研究与开发);

(3)综述与评论(主要指电子与信息工程领域技术发展方面的综述、评论和展望)。

欢迎学者、专家及科技人员踊跃投稿,请务必遵守《发表学术论文“五不准”》。

来稿要求及注意事项

(1)来稿务必脱密处理并经所在单位审查,应主题明确,论述合理,逻辑严谨,数据可靠,叙述清楚,文字精炼。作者及来稿不得存在学术不端情况,否则,后果自负。

(2)来稿一般不应超过6 000字,综述稿不宜超过8 000字。稿件应附英文题名、作者名、单位名、摘要和关键词,基金项目应注明项目编号。中文题名一般不超过20个汉字,必要时可加副标题。为便于文献检索,应尽可能根据《中国图书馆分类法(第五版)》提供中图分类号。

(3)摘要应包括目的、方法、结果和结论四要素,即用简洁的语言说明论文要解决的问题,主要工作过程及所采用的技术手段和方法,研究所获得的实验数据、结果及其意义。篇幅以200~300字为宜。

(4)关键词应选取能反映论文主题概念的词或词组,以3~8个为宜。中英文关键词须对应。第一个关键词应体现论文学科分类。

(5)文中涉及的物理量和计量单位及其符号的选用和书写应符合最新发布的有关国家标准。应注意区分各物理量符号的文种、大小写、正斜体(矢量和矩阵用黑斜体)、上下角标等。

(6)插图和表格只给出必要的,具体要求请分别参见《CY/T 171—2019 学术出版规范 插图》和《CY/T 170—2019 学术出版规范 表格》。

(7)来稿中引用他人的成果,应在参考文献中列出,并按顺序在正文相应位置进行标示。参考文献只择主要的,未公开发表的文献请勿列入参考文献。书写格式请参见《GB/T 7714—2015 信息与文献 参考文献著录规则》。

(8)来稿需提供全部作者简介,包括作者的姓名、性别、出生年、籍贯、已有学位获得时间、现攻读学位、职称、研究方向等内容。

(9)请务必在线投稿,并按要求和提示真实、准确填写有关信息。

(10)本刊编辑部将在两个月内给出来稿审核结论。如逾期未收到处理意见或刊用通知,作者有权对稿件另行处理。稿件一经刊出,本刊将按规定酌情从优支付稿酬并赠送当期样刊。本刊支付的稿酬中已包含作者著作权使用费。

请务必在本刊网站投稿,不接收邮箱投稿。www.teleonline.cn是本刊唯一投稿入口。

特别提示:本刊从2022年7月20日起发表论文不收取任何费用!

本刊从未委托任何组织、机构、个人代理征稿,切勿上当受骗!

电话:(028)87555632

地址:四川省成都市金牛区营康西路85号

Email:dxjs@cetc.com.cn

传真:(028)87538378

邮编:610036

网址:www.teleonline.cn

本刊被下列数据库列为来源期刊：

- 中国学术期刊综合评价数据库
- 中国科技论文与引文数据库
- 中国科技论文统计源
- 《中国学术期刊影响因子年报》统计源
- 中国核心期刊（遴选）数据库
- 中国学术期刊（光盘版）
- 中文科技期刊数据库
- 中国期刊全文数据库
- 《科技期刊世界影响力指数（WJCI）报告》
- 超星期刊出版系统
- 台湾华艺线上图书馆（Airiti Library）
- 英国IET《科学文摘》（SA）INSPEC
- 美国《剑桥科学文摘》（CSA）
- 波兰《哥白尼索引》（IC）
- 美国《乌利希国际期刊指南》（Ulrich's Periodicals Directory）
- 美国《史蒂芬斯数据库》（EBSCOhost）

本刊已荣获以下奖项：

- ★首届《CAJ-CD规范》执行优秀期刊奖
- ★工业和信息化部2007—2008年度电子科技期刊学术技术水平优秀奖
- ★工业和信息化部2009—2010年度优秀期刊奖

电讯技术

Dianxun Jishu

（月刊，1958年创刊）

第65卷 第3期(总第436期)2025年3月28日

TELECOMMUNICATION ENGINEERING

(Monthly, Started in 1958)

Vol.65 No.3(Series No.436) March 28,2025

主管单位：CETC 中国电子科技集团有限公司

主办单位：中国电子科技集团公司第十研究所

出版单位：《电讯技术》期刊社

四川省成都市金牛区营康西路85号，610036

电话：（028）87555632

传真：（028）87538378

业务联系：dxjs@cetc.com.cn

网址（投稿、审稿等）：www.teleonline.cn

社长：赵晓虎

副社长：黄金元

总编辑：乔文昇

编辑部主任：郑理

副总编：赵勇

责任编辑：李桂花

印刷：成都市优艺图文设计有限公司

国内发行：成都市报刊发行局

海外发行：中国国际图书贸易集团有限公司

（北京市车公庄西路35号第399信箱，100048）

国内订阅：全国各地邮局

出版日期：2025年3月28日

Superintendent Institution China Electronics Technology Group Corporation

Sponsor The Tenth Research Institute of China Electronics Technology Group Corporation

Publishing Institution Periodical Publishing House of Telecommunication Engineering

85 Yingkang Xilu, Chengdu 610036, Sichuan, China

Tel. 86-28-87555632

Fax. 86-28-87538378

Email dxjs@cetc.com.cn

Web Site(manuscript submission&review) www.teleonline.cn

President ZHAO Xiaohu

Vice President HUANG Jinyuan

Editor-in-Chief QIAO Wensheng

Editorial Director ZHENG Li

Deputy Editor-in-Chief ZHAO Yong

Responsible Editor LI Guihua

Printed by Chengdu Youyi Graphic Design Co., Ltd.

Distributed by

Domestic: Chengdu Post Office

Overseas: China International Book Trading Corporation(Guoji Shudian)

(35 Chegongzhuang Xilu, P.O. Box 399, Beijing 100048, China)

Subscription Post Office in China

Published on March 28,2025

中国标准连续出版物号：ISSN 1001-893X
CN 51-1267/TN

邮发代号:62-39 海外代号:M2924 国内定价:30.00元/册 国外定价: 30.00美元/册