

基于 mRMR-GA 的太赫兹信道场景识别研究

郝昕宇, 廖希, 王洋, 林峰, 罗娇, 张杰

(重庆邮电大学通信与信息工程学院, 重庆 400065)

摘要: 针对太赫兹信道场景识别中存在的特征参数冗余和场景相关性不足的问题, 结合最大相关最小冗余 (mRMR) 准则和遗传算法 (GA), 提出基于 mRMR-GA 的太赫兹信道场景识别算法。该算法通过 GA 执行信道特征的交叉和变异, 利用 mRMR 准则选择与场景相关性大的最优特征参数, 输入至反向传播神经网络模型。利用室内场景 1 745 组太赫兹信道仿真数据, 构建含 12 种信道特征的参数集进行模型训练和验证。结果表明, 所提算法识别准确率和效率分别提升 8% 和 38.8%, 收敛性和迁移泛化能力优于传统算法。

关键词: 太赫兹信道; 场景识别; 特征选择; 遗传算法; 神经网络

中图分类号: TN011

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025082

Research on scenario recognition for THz channels based on mRMR-GA

HAO Xinyu, LIAO Xi, WANG Yang, LIN Feng, LUO Jiao, ZHANG Jie

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: To address the challenges of excessive feature parameter redundancy and insufficient scene correlation in terahertz (THz) channel scenario recognition, a recognition algorithm integrating the minimal redundancy maximal relevance (mRMR) criterion with genetic algorithm (GA) optimization was constructed based on feature selection theory and evolutionary computation principles. The crossover and mutation operations of channel characteristics were executed by the genetic algorithm (GA), and the optimal feature parameters with high scenario relevance were selected using the minimum redundancy maximum relevance (mRMR) criterion. These parameters were then inputted into a backpropagation neural network model. To validate the method, a dataset containing 12 channel features was constructed with 1 745 groups of terahertz channel simulation data collected from indoor scenarios, and the model was trained and rigorously validated based on this dataset. The results demonstrate that the proposed algorithm improves accuracy and efficiency by 8% and 38.8%, respectively, and outperforms traditional algorithms in terms of convergence and transfer generalization capabilities.

Keywords: THz channel, scenario recognition, feature selection, genetic algorithm, neural network

收稿日期: 2025-02-13; 修回日期: 2025-04-24

通信作者: 廖希, liaoxi@cqupt.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62271095, No.62171071); 重庆市自然科学基金资助项目 (No.CSTB2022NSCQ-MSX1125); 重庆市教委科学技术研究资金资助项目 (No.KJZD-K202300607); 重庆市自然科学基金创新发展联合基金资助项目 (No.CSTB2022NSCQ-LZX0037)

Foundation Items: The National Natural Science Foundation of China (No.62271095, No.62171071), The Natural Science Foundation of Chongqing (No.CSTB2022NSCQ-MSX1125), The Science and Technology Research Program of Chongqing Municipal Education Commission (No.KJZD-K202300607), The Natural Science Foundation Innovation and Development Joint Fund Project of Chongqing (No.CSTB2022NSCQ-LZX0037)

0 引言

随着无线通信技术的快速发展,太赫兹通信凭借其丰富的频谱资源和高数据传输能力,已成为 6G 技术的关键组成部分。太赫兹信道特性由物理传播环境和场景中的散射体决定,不同的传播环境会导致不同的信道特性。因此,无线信道的场景识别在许多方面具有重要意义。信道特性在视距 (LoS, line of sight) 和非视距 (NLoS, non LoS) 情况下会有很大变化,在信道建模时需要进行区分^[1-2]。此外,基于到达角或到达时间的定位方法也需要以无线信号的 LoS 传播为前提^[3-4]。

典型的信道识别方法通过设置参数阈值来区分信道状态和场景,识别准确率无法保证。并且,识别过程通常需要大量的信道参数,不仅增加了计算复杂度,还可能引入冗余信息,影响识别的准确性和效率。因此,如何高效准确地识别信道场景成为一个亟待解决的问题。

无线信道十分复杂,很难通过直接建立信道参数与环境之间的映射来实时识别信道状态和场景^[5]。随着机器学习在无线信道中应用研究的深入,基于机器学习的信道场景识别也受到关注。基于机器学习的 LoS/NLoS 场景识别方法,主要包括支持向量机 (SVM, support vector machine)、反向传播神经网络 (BPNN, back propagation neural network)^[6]和随机森林 (RF, random forest) 等。信道场景识别依赖于信道特征,Cao 等^[7]提出的 SVM 方法仅需测量距离和接收信号强度即可实现 LoS/NLoS 识别,虽简化了计算,但在动态多径环境下易因特征信息不足而失效。为缓解这一问题,Nguyen 等^[8]引入相关向量机替代 SVM,通过稀疏化模型降低了计算复杂度,但仍局限于时延和功率等低维特征,未充分利用信道多维特性。Huang 等^[9]使用角度信息扩充了信道特征,提出的 SVM 场景识别算法识别错误率比传统算法降低了 6%,但 SVM 模型在工业场景中因角度信息易受多径干扰而产生性能波动。上述研究受限于对单一特征或线性组合的过度依赖,导致模型在未知环境中的泛化能力不足。

为提升模型鲁棒性,学者们将研究重点转向多维度特征融合策略。Chang 等^[10]和 Xie 等^[11]将信道脉冲响应的偏度、峰度、均方根时延扩展等信道参数作为特征,提出了基于长短时记忆 (LSTM,

long short-term memory) 网络的 LoS/NLoS 信道识别算法。Fu 等^[5]将信道统计特征 (如接收功率、时延扩展、莱斯 K 因子等) 和簇特征 (如平均簇内时延、衰减因子等) 作为输入设计了一种多输入多输出的深度神经网络 (DNN, deep neural network) 框架,实现了多场景的 LoS/NLoS 状态识别。Yang 等^[12]将功率时延谱与统计特征 (莱斯 K 因子、均方根时延扩展、阴影衰落) 融合,构建多输入 BPNN 模型,解决了传统算法在复杂场景中特征不足的问题。然而,上述算法虽通过特征融合提升了性能,却因冗余特征和人工干预引入了新的复杂度与泛化瓶颈。

随着信道特征的增多,单一算法的特征适应性出现不足。Zhou 等^[13]将 DNN 和自注意力网络 (SAN, self-attention network) 进行融合,利用 DNN 处理低维人工特征,SAN 处理高维原始信道脉冲响应 (CIR, channel impulse response) 特征的策略平衡了数据淹没与特征关联问题。Jiang 等^[14]将卷积神经网络 (CNN) 和 LSTM 网络进行结合,使用 CNN 提取特征,然后将输出馈送到 LSTM 网络得到场景分类。Li 等^[15]引入双向长短时记忆 (BiLSTM, bidirectional LSTM) 网络,使用双向时序分析捕获信号的前后文依赖,取得了比单向 LSTM 更高的识别准确率。Zhang 等^[16]在 CNN 的基础上引入了信道注意力模块,将 CIR 作为输入,利用三维卷积和信道注意力模块提升了模型特征提取能力,提升了场景识别性能。尽管这些方法在一定程度上提高了识别性能,但在处理高维度特征时仍面临计算复杂度增加的问题,且特征冗余和场景相关性不足的问题未得到根本解决。

信道场景识别的核心在于特征选择,一方面,单一信道特征无法适应多种信道环境;另一方面,冗余信道特征又会增加场景识别方法的计算复杂度,影响泛化能力。综上所述,现有算法依赖人工阈值和高维特征集,存在特征冗余和场景内关联性不足的问题,导致计算复杂度高且识别精度受限。

针对太赫兹信道场景识别面临的高维度特征冗余与场景相关性不足挑战,本文融合最大相关最小冗余 (mRMR, minimal redundancy maximal relevance) 准则和遗传算法 (GA, genetic algorithm),提出一种基于 mRMR-GA 的太赫兹信道场景识别算法。其创新性体现在 3 个方面。

1) 动态冗余控制机制: 在 GA 的交叉变异过程中嵌入 mRMR 准则, 实时剔除冗余特征, 避免无效搜索, 提升特征子集的场景相关性。与传统的多特征输入方法 (如文献[5,12]) 相比, 该机制能够有效地减少特征冗余, 提高特征子集的质量。

2) 非线性关联建模: 针对文献[7-8]线性关系假设的局限性, 通过 GA 的全局优化能力增强 mRMR 准则对非线性关系的适应性, 加快模型收敛。

3) 轻量化与泛化平衡: 优化后特征维度降至 4~6 维, 在保证识别精度的同时, 提升了计算效率。与文献[13-14]的方法相比, 该算法在降低计算复杂度方面具有显著优势, 迁移至新场景的错误率仅为 12%, 显著优于传统算法。

1 基于 mRMR-GA 的太赫兹信道场景识别算法

1.1 特征参数提取

无线信道特性可以通过不同信道参数描述, 常用的信道特征参数如下。

1) 最大接收功率 $\max(|h|^2)$: 接收天线在特定条件下能够接收的信号最大功率。

2) 接收功率峭度指数 κ : 描述了接收功率概率分布波形尖峰程度, 可以帮助分析无线信道的多径特性。

3) 接收功率偏度指数 S : 描述了接收信号功率分布的不对称性。

4) 功率上升时间 $\Delta\tau$: 发射信号从初始状态上升到指定功率水平所需的时间。

5) 时延扩展 σ_t : 描述了由于多径传播效应引起的信号到达时间扩展的程度, 反映了信号的多个路径分量在接收端到达时间的差异。

6) 莱斯 K 因子 K_r : 直射分量和散射分量的相对强度, 定义为最高功率多径分量 (MPC, multipath component) 与剩余 MPC 的功率比值。

7) 角度扩展 σ_ϕ : 描述了多径信号到达接收天线时的角度分布情况, 反映了多径信号到达角度的分

散程度。

8) 簇内时延扩展 σ_{CDS} : 描述了单个簇内多径分量到达接收端的时间分散程度, 反映了一个簇中各个多径分量的到达时间差异。

9) 簇内水平角度 ϕ_c : 描述了单个簇内多径信号在水平面上的角度分布范围, 反映一个簇中各个多径分量在接收天线或发射天线水平角度 (即水平方向角) 的分散程度。

10) 簇内水平到达角扩展 σ_{CASA} 和簇内垂直到达角扩展 σ_{CESA} : 描述了单个簇内的多径信号在水平或垂直方向上的到达角度分布范围, 反映多径簇中的多径分量在接收天线水平或垂直面上的到达角度分散程度。

1.2 特征选择算法

特征选择是特征工程中的重要环节, 特征选择算法应用特定评价标准, 从初始特征空间中挑选出性能最优的特征子集, 同时确保所选特征具有较强可解释性。常见的特征选择算法有 mRMR^[17]、RF^[18]、GA^[19]等。

表 1 为常见特征选择算法对比, 其中, n 是信道特征数量, N 是信道数据样本数量, t 是 GA 的迭代次数, M 是 GA 的种群大小。从表 1 可以看到, mRMR 计算特征与类别之间的互信息和特征之间的冗余信息, 时间复杂度较小, 最大为 $O(nN)$; RF 通过构建多个决策树来评估特征的重要性, 时间复杂度主要取决于决策树的数量、树的深度以及特征数量; GA 需要构造种群、迭代生成新一代种群以及计算对应的适应度, 其复杂度取决于迭代次数和种群大小。

1.3 mRMR-GA 特征选择算法

对输入特征的学习是信道场景识别的基础, 信道特征集中的无关信息和冗余特征会使场景识别算法过拟合或者欠拟合, 导致识别精度下降。在信道场景识别中, 主动筛选具有代表性的信道特征子集用于识别, 可以避免冗余信道特征导致的识别准确率下降。

表 1 常见特征选择算法对比

算法	时间复杂度	优点	缺点
mRMR	$O(nN)$	计算效率高, 可以有效地挖掘相关性特征	要求数据集中的特征具有一定相关性
RF	$O(nN \log(N))$	易于实现和调整	特征子集随机性大, 容易过拟合, 迁移性差
GA	$O(tM)$	全局搜索能力强, 能够在大规模特征空间中找到最优解	时间复杂度较高

传统 mRMR 算法基于互信息的单变量分析,难以捕捉特征间的非线性关联(如二次项、交互项)。例如, mRMR 通过最大化特征与目标的互信息 $I(f_i; y)$ 并最小化特征间互信息 $I(f_i; f_j)$ 选择特征,但互信息的计算依赖于概率密度估计,对高维非线性数据的敏感性较低。而 GA 因种群多样性不足可能无法充分探索特征子集的潜在最优组合,导致遗漏对场景识别更有效的特征子集。此外, GA 需要若干次迭代才能达到稳定解,计算复杂度较高。

针对以上问题,在文献[20]基础上本文提出的 mRMR-GA 特征选择算法引入 mRMR 准则引导 GA 搜索最佳特征子集,在每一代种群筛选中,优先保留高相关性、低冗余的特征子集,避免无效随机搜索。并将 GA 的全局搜索能力(通过交叉、变异生成多样性子集)与 mRMR 的局部优化(筛选最具代表性的特征)结合,提前剔除冗余特征,减少后续 BPNN 训练的特征维度,从而缩短单次迭代时间。mRMR-GA 特征选择算法的具体步骤如下。

1) 设定个体编码及初始群体: 定义信道参数集 F , F 由最大接收功率 $\max(|h|^2)$ 、接收功率峭度指数 κ 、接收功率偏度指数 S 、功率上升时间 $\Delta\tau$ 、莱斯 K 因子 K_r 、时延扩展 σ_τ 、角度扩展 σ_ϕ 、簇内时延扩展 σ_{CDS} 、簇内水平到达角扩展 σ_{CASA} 、簇内垂直到达角扩展 σ_{CESA} 、簇数 k 、簇内水平角度 φ_c 共 12 个信道特征组成。使用 0、1 编码 12 个特征,其中 0 代表特征被淘汰,1 代表特征被保留。生成 M 个个体 $\{F_1, F_2, F_3, \dots, F_M\}$, 每个个体为 LoS/NLoS 场景识别的信道特征子集, F_i 由 12 个 0、1 字符串组成, $\{F_1, F_2, F_3, \dots, F_M\}$ 组成初始群体。

2) 构造适应度函数: 利用适应度来评估候选信道特征子集的优劣。将 BPNN 输出的决定系数 R^2 作为适应度 r , 通过适应度 r 观测 GA 的优化过程。LoS/NLoS 场景识别中, 第 i 个信道特征子集为 F_i , 其适应度表示为

$$r_{F_i} = 1 - \frac{\sum_{j=1}^n (y_{\text{sim}}(F_i(j)) - y(F_i(j)))^2}{\sum_{j=1}^n (\bar{y}(F_i(j)) - y(F_i(j)))^2} \quad (1)$$

其中, n 是信道特征数量, $y_{\text{sim}}(F_i(j))$ 和 $y(F_i(j))$ 分别为 F_i 中第 j 个样本的场景标签和期望值, $\bar{y}(F_i(j))$ 是第 j 个样本的期望均值。平均适应度 r_{ave} 表示为

$$r_{\text{ave}} = \frac{1}{M} \sum_{i=1}^M r_{F_i} \quad (2)$$

3) 交叉变异信道特征参数: 保留信道特征集 $\{F_1, F_2, F_3, \dots, F_M\}$ 中适应度最高的 2 个信道特征子集, 以 F_i 和 F_j 为例, 其中特征编码分别为 $F_i = (a_1, a_2, a_3, \dots, a_{12})$ 和 $F_j = (b_1, b_2, b_3, \dots, b_{12})$ 。对 F_i 和 F_j 在 c 点进行交叉, 表示为

$$\begin{cases} F'_i = \{a_1, a_2, \dots, a_c\} + \{b_{c+1}, b_{c+2}, \dots, b_{12}\} \\ F'_j = \{b_1, b_2, \dots, b_c\} + \{a_{c+1}, a_{c+2}, \dots, a_{12}\} \end{cases} \quad (3)$$

$$c = \text{ceil}(\alpha L(F) + 1) \quad (4)$$

其中, α 为 (0,1) 内的随机数, $L(F)$ 为 F 的长度, c 为开始遗传运算的特征序号, $\text{ceil}(\cdot)$ 表示向上取整。2 个信道特征子集在 c 点进行交叉, 得到新个体 F'_i 和 F'_j 。 F'_i 和 F'_j 分别保留了 F_i 和 F_j 的部分最优特征子集。

若信道特征子集 F_i 在 d 点发生变异, 其取值范围为 $[a_{\text{min}}^d, a_{\text{max}}^d]$, 则该过程表示为

$$F'_i = \{a_1, a_2, \dots, a'_d, \dots, a_{12}\} \quad (5)$$

$$a'_d = \begin{cases} 0, \lambda \in [0, 0.5] \\ \text{ceil}(a_{\text{min}}^d + \lambda(a_{\text{max}}^d - a_{\text{min}}^d)), \lambda \in (0.5, 1] \end{cases} \quad (6)$$

其中, a_{max}^d 和 a_{min}^d 分别为原始特征 a_d 的最大值和最小值, λ 为 [0,1] 内的随机数。

4) 筛选信道特征: 步骤 3) 交叉变异后, 产生信道特征集 $\{F'_1, F'_2, F'_3, \dots, F'_M\}$ 。信道特征集 $\{F'_1, F'_2, F'_3, \dots, F'_M\}$ 中, 子集间冗余度仍然较高, 可以通过 mRMR 准则保留相关性较高的特征。计算各子集的冗余度和相关性, 以第 l 个信道特征子集 F'_l 为例, 计算为

$$\max D(F'_l, y), D = \frac{1}{|F'_l|} \sum_{f_i \in F'_l} I(f_i; y) \quad (7)$$

$$\min R(F'_l), R = \frac{1}{|F'_l|^2} \sum_{f_i, f_j \in F'_l} I(f_i; f_j) \quad (8)$$

其中, f_i 和 y 分别为第 i 个信道特征参数及其 LoS 或 NLoS 场景标签, $I(f_i, y)$ 为第 i 个信道特征参数与场景标签 y 之间的互信息, $I(f_i; f_j)$ 是第 i 个信道特征参数和第 j 个信道特征参数的互信息, $|F'_l|$ 为所选特征数量。

结合式(7)和式(8)得到最大相关最小冗余准则

$$\max \Phi(D, R), \Phi = D - R \quad (9)$$

$$\Phi(f_i) = I(f_i; y) - \frac{1}{i-1} \sum_{f_j \in F'} I(f_i; f_j) \quad (10)$$

其中, Φ 是信道特征集的 mRMR 值。在 GA 的变异操作中, 若某特征位发生变异 (0→1 或 1→0), mRMR 准则会利用式(9)立即计算新特征子集的 Φ 值, 仅保留 Φ 值高于父代的个体。这种机制迫使算法在探索新特征组合时, 始终以最大化相关性和最小化冗余性为目标, 从而有效融合非线性特征。

最后, 选择 Φ 最大的特征 $\Phi(f_i)$, 加入已选特征集 F_i''

$$F_i'' = \arg \max_{f_j \in F_i'} \{ \Phi(f_j) \} \quad (11)$$

保留相关性最大的 5 个特征, 形成新的信道特征集。

5) 迭代寻优最佳特征参数子集: 重复步骤 3) 和步骤 4), 达到适应度阈值或满足停止条件, 得到最佳的特征参数子集 F_i'' 。

1.4 信道场景识别

基于 mRMR-GA 的太赫兹信道场景识别算法将 mRMR-GA 特征选择算法和 BPNN 结合实现场景识别。mRMR-GA 特征选择算法实现最佳信道特征子集的选择, 获取的特征子集被进一步应用于 BPNN 学习, 从而提高整体场景识别性能。图 1 为基于 mRMR-GA 的太赫兹信道场景识别算法流程。

1) 将 LoS 和 NLoS 场景中的 12 个信道特征用 0 和 1 进行编码, 生成初始信道特征集 $\{F_1, F_2, F_3, \dots, F_M\}$ 。

2) 对信道数据样本添加场景标签, 按比例均匀划分为训练集和测试集, 并进行归一化处理。

3) 初始化参数, 包括 GA 的种群大小 M 、总的迭代次数 T , BPNN 的损失函数、优化器、学习率、激活函数等。

4) 利用不同特征子集训练 BPNN 并验证, 得到评价指标 R^2 作为遗传算法的适应度值 $r = \{r_1, r_2, r_3, \dots, r_M\}$, 并保留适应度值最高的两组特征子集 F_i 和 F_j 。

5) 利用 mRMR-GA 特征选择算法获得冗余度更小的信道特征集。

6) 继续步骤 4) 和步骤 5), 直到达到最大迭代次数 T , 选择此时适应度值最高的特征子集 F_i' 。

7) 将最优特征子集 F_i' 输入 BPNN 中获得最终性能指标 r 。

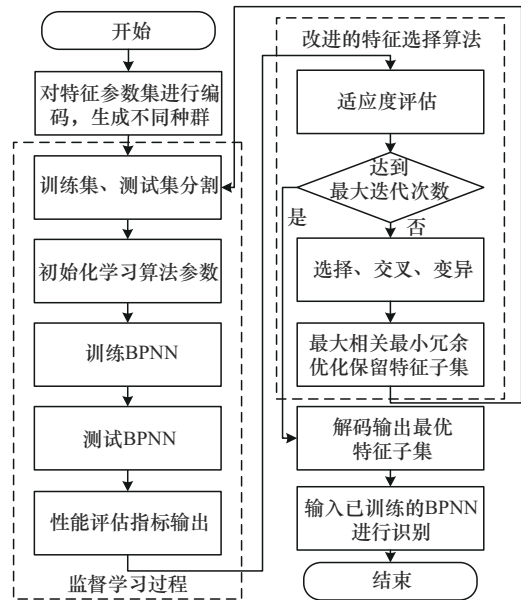


图 1 基于 mRMR-GA 的太赫兹信道场景识别算法流程

2 太赫兹室内通信场景信道仿真

BPNN 模型的训练需要大量携带标签的样本, 本文在 Wireless Insite 软件中搭建太赫兹室内通信场景并进行仿真, 获取了充足的信道参数。

根据文献[21]中描述的室内办公室通信场景构建仿真环境, 场景由走廊和办公室组成, 整体环境尺寸约为 64.21 m×25.48 m×3 m, 包含开阔走廊 64.21 m×4.52 m×3 m (①)、半开放的办公室 24.80 m×13.02 m×3 m (②) 和狭窄走廊 64.21 m×1.30 m×3 m (③) 这 3 个区域, 仿真场景的俯视图如图 2 所示。

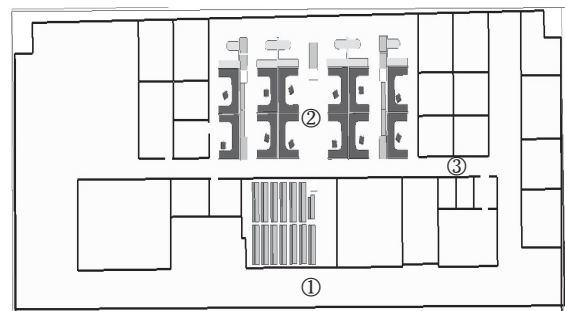


图 2 仿真场景的俯视图

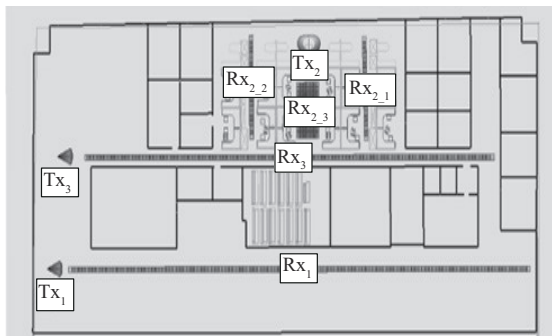
场景内布有天花板、地板、墙壁、窗户以及办公桌椅等物体。其中, 天花板为碳纤维增强塑料, 地板和墙壁为混凝土材质, 窗户为玻璃材质, 办公桌为木板和金属制成, 办公椅由木板和泡沫组成, 不同材质的电磁参数由 Wireless Insite 材料库提供, 具体材料电磁参数如表 2 所示。

表2 材料电磁参数

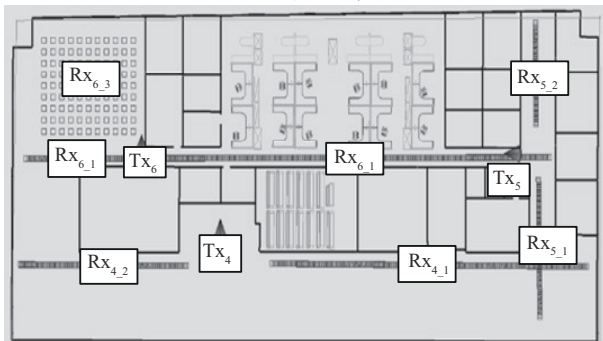
材料名称	相对介电常数	电导率/(S·m ⁻¹)	厚度/m
混凝土	5.31	1.051	0.300
金属	1.00	10 ⁷	—
玻璃	6.27	0.717	0.003
干墙	2.94	0.242	0.013
木板	1.99	0.467	0.030
碳纤维增强塑料	1.50	0.590	0.090

仿真的主要参数设置：中心频率为 220 GHz，带宽为 10 GHz。为了充分地接收多径信号，在开阔走廊场景和狭窄走廊场景采用 E 面和 H 面半波束宽度分别为 49° 和 46° 的发射天线，在半开放的办公室场景采用 E 面和 H 面半波束宽度分别为 80° 和 50° 的发射天线。发射机高度均为 2 m，采用增益为 15 dBi 的定向天线，发射信号功率为 0 dBm。接收机高度均为 1.5 m，采用增益为 25 dBi 的全向天线。系统最低信号检测功率设置为 -250 dBm，最大反射次数为 4，绕射次数为 1，不考虑透射。

为获取多样化的通信场景信息，根据场景区域特点，将场景划分为开阔走廊、半开放办公室和狭窄走廊，并采用路径 (route) 方式部署足够数量的收发机进行仿真，具体部署如图 3 所示。



(a) 视距场景



(b) 非视距场景

图 3 仿真场景收发机部署示意

在 LoS 场景中，3 个区域分别部署 3 个发射机 (Tx₁、Tx₂、Tx₃) 和 3 个接收机 (Rx₁、Rx₂、Rx₃)，设置 5 条接收机移动路径 (Rx₁、Rx_{2,1}、Rx_{2,2}、Rx_{2,3}、Rx₃)，每个测量点间距 0.2 m，共 800 个。

在 NLoS 场景中，3 个区域分别部署 3 个发射机 (Tx₄、Tx₅、Tx₆) 和 3 个接收机 (Rx₄、Rx₅、Rx₆)，设置 7 条接收机移动路径 (Rx_{4,1}、Rx_{4,2}、Rx_{5,1}、Rx_{5,2}、Rx_{6,1}、Rx_{6,2}、Rx_{6,3})，每个测量点间距 0.2 m，共 945 个。

3 性能分析

3.1 数据集

仿真获取的信道数据包含 CIR、多径功率、时延、角度等 MPC 参数，对信道数据进行参数提取，构建如表 3 所示的信道特征数据集。

表 3 信道特征数据集

场景	环境	接收机位置	样本数量/个	总计/个
LoS	开阔走廊	Rx ₁	300	800
	半开放办公室	Rx _{2,1} 、Rx _{2,2} 、Rx _{2,3}	236	
	狭窄走廊	Rx ₃	264	
NLoS	开阔走廊	Rx _{4,1} 、Rx _{4,2} 、Rx _{5,1} 、Rx _{5,2}	453	945
	狭窄走廊	Rx _{6,1} 、Rx _{6,2} 、Rx _{6,3}	492	

数据集中包含 800 个 LoS 样本和 945 个 NLoS 样本，每个样本由 12 个信道特征组成，信道特征由最大接收功率、接收功率峭度指数、接收功率偏度指数、功率上升时间、莱斯 K 因子、时延扩展、角度扩展、簇内时延扩展、簇内水平到达角扩展、簇内垂直到达角扩展、簇数、簇内水平角度组成，即 $\{\max(|h|^2), \kappa, S, \Delta\tau, K, \sigma_\tau, \sigma_\phi, \sigma_{CDS}, \sigma_{CASA}, \sigma_{CESA}, k, \phi_c\}$ 。

图 4 给出了 LoS 和 NLoS 场景下不同信道特征的概率密度函数 (PDF, probability density function)。LoS 和 NLoS 场景中，信道统计特征表现出了不同的集中趋势和分布形态，尽管部分信道特征在 LoS 和 NLoS 场景中表现出了明显的集中趋势差异，但特征分布形态存在重叠区，难以直接使用阈值对 LoS 和 NLoS 场景进行区分。

3.2 性能评估与结果分析

将信道场景识别看作二分类问题，LoS 场景和 NLoS 场景的样本数据标签分别标注为 1 和 0，

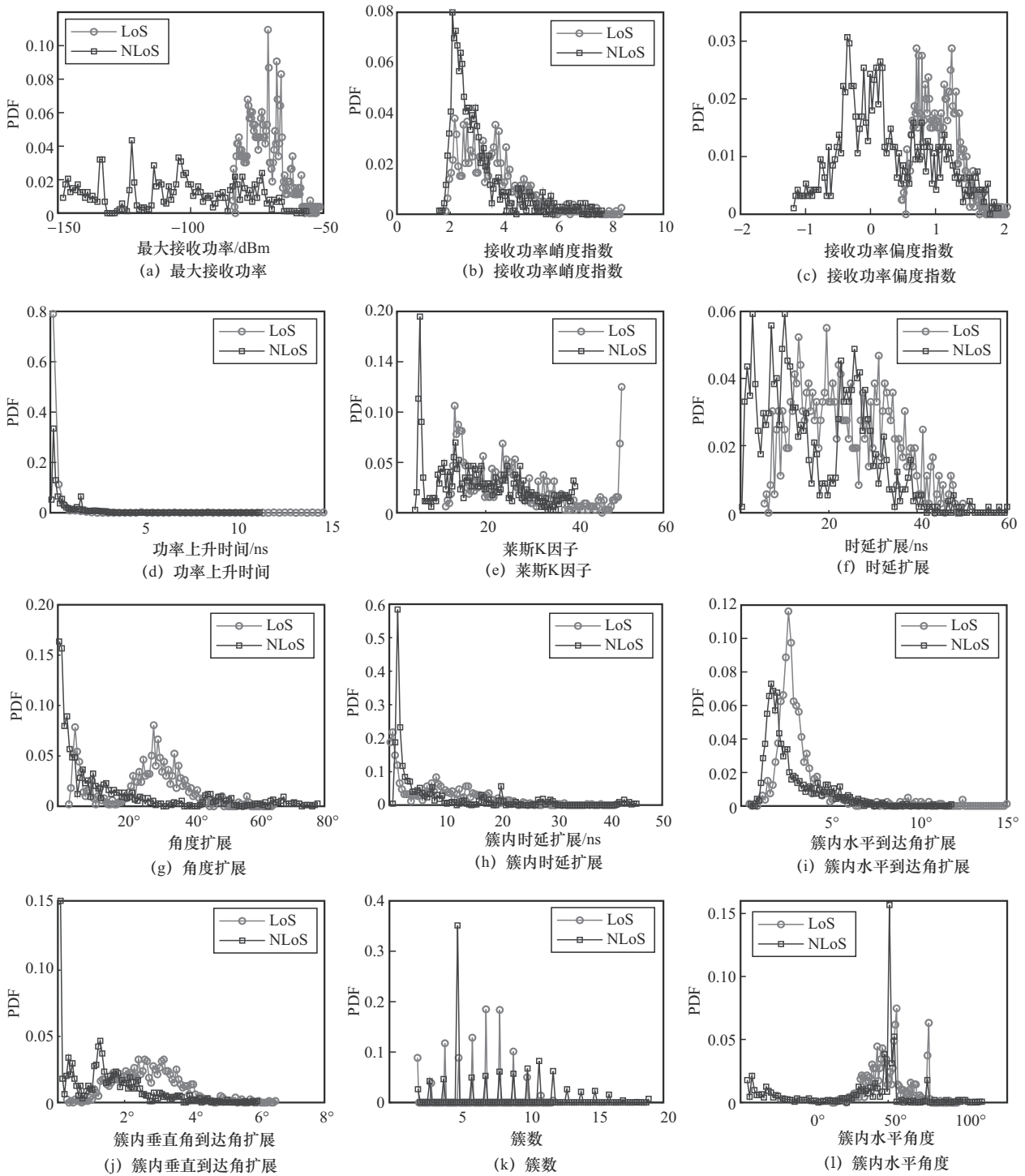


图4 不同信道特征的PDF

按照 6:4 的比例划分训练集和测试集, 保证模型有足够的训练数据进行学习同时又提供足够的测试数据来评估模型的性能。为了满足模型种群多样性的同时控制计算成本, 避免因过多迭代造成的计算浪费, 设置 GA 种群大小为 10, 最大迭代次数为 20; BPNN 的输入层神经元个数为 12, 与输

入特征维度一致, 确保所有特征被接收处理。隐藏层神经元个数为 5, 既能保证模型复杂性, 又可避免过拟合。输出层神经元个数为 1, 用于输出场景的二分类结果。采用 tansig 激活函数和 Trainlm 训练算法, 最大迭代次数为 1 000, 详细的参数设置如表 4 所示。

表 4 算法参数设置

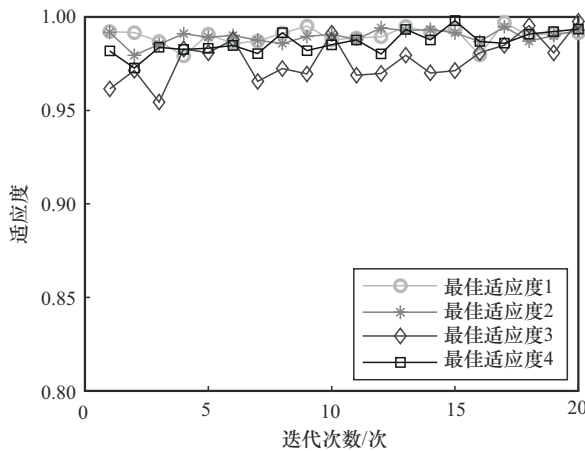
算法	参数名称	取值
遗传算法	种群大小	10
	最大迭代次数	20
BPNN	输入层神经元个数	12
	隐藏层神经元个数	5
	输出层神经元个数	1
	激活函数	tansig
	训练函数	Trainlm
	最大迭代次数	1 000
	学习率	0.01
	目标训练误差	1×10^{-6}

在信道特征数据集上对遗传算法和基于 mRMR-GA 的太赫兹信道场景识别算法进行测试，

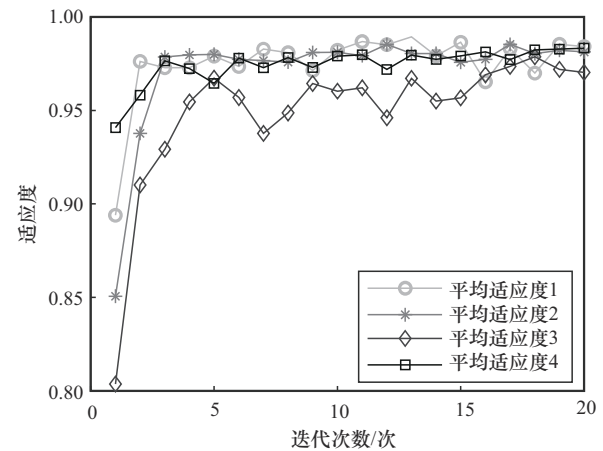
测试结果如图 5 所示。基于 mRMR-GA 的太赫兹信道场景识别算法在第 15 次迭代时达到收敛，而 GA 在 20 次时收敛，这表明本文算法收敛速度更快、稳定性也更强。此外，基于 mRMR-GA 的太赫兹信道场景识别算法的初始平均适应度约为 0.9，高于遗传算法，说明了特征工程的重要性和基于 mRMR-GA 的太赫兹信道场景识别算法的有效性。总体而言，基于 mRMR-GA 的太赫兹信道场景识别算法既保证了种群中每个个体都具有较大的适应度值，还具有较快的收敛速度和很好的寻优能力。

1) 信道特征对识别效果的影响

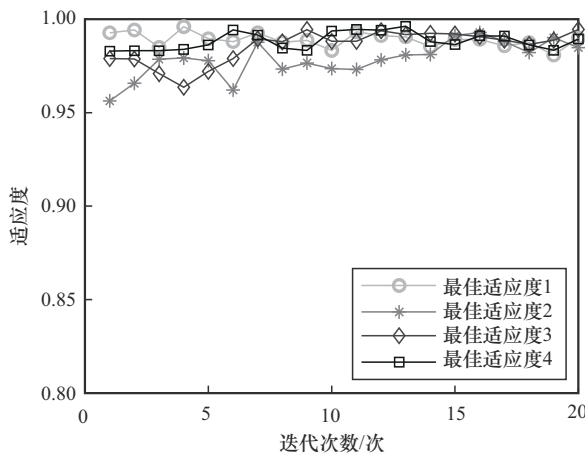
使用 mRMR、RF、GA 和基于 mRMR-GA 的太赫兹信道场景识别算法对信道特征数据集进行处理，获取信道特征子集。将不同算法获取的信道特征子集输入 BPNN 模型和 SVM 模型进行 5 折交叉验证训练。SVM 中核函数为高斯核，惩罚因子为



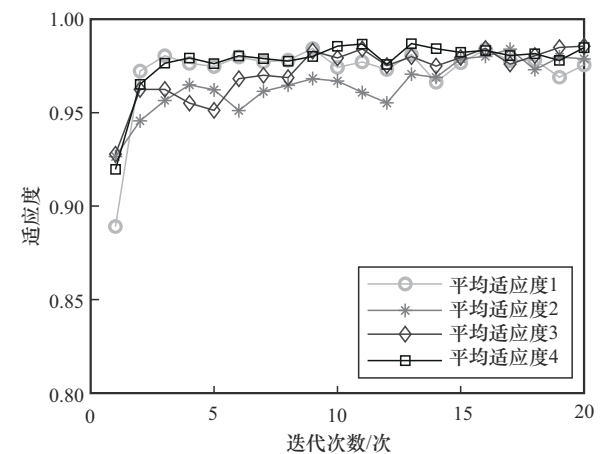
(a) 遗传算法最佳适应度



(b) 遗传算法平均适应度



(c) 基于mRMR-GA的太赫兹信道场景识别算法最佳适应度



(d) 基于mRMR-GA的太赫兹信道场景识别算法平均适应度

图 5 算法适应度比较

10, 径向基函数参数为 0.01, 最大迭代次数为 1 000。

图 6 为单一信道特征输入时 BPNN 模型的识别性能, 其中最大接收功率是场景识别最有效的特征参数, 其次是接收功率偏度指数。太赫兹频段频率较高, 在室内环境传播衰减更大, 因此 NLoS 场景下的最大接收功率和接收功率分布特性相对于 LoS 场景会呈现出明显的差异。此外, 角度扩展和基于簇的特征参数对 BPNN 模型识别性能影响较大, 这是因为室内环境散射体较多, 信号在 NLoS 场景会经过更多的反射和散射传播从而产生了更多、更分散的多径簇。图 6 中的线条是多特征参数下的 BPNN 模型识别性能, 当特征参数数量为 5 时即可取得大于 0.9 R^2 , 实现准确识别。

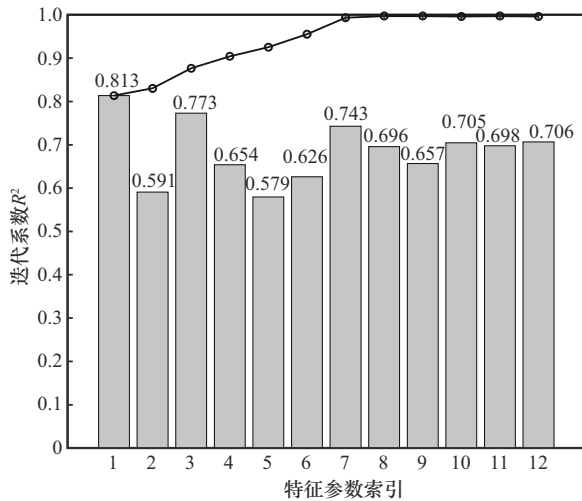


图 6 单一信道特征时 BPNN 模型的识别性能

表 5 和表 6 分别为 BPNN 和 SVM 模型在不同信道特征组合下的性能比较。将全部特征作为数据集时 BPNN 模型和 SVM 模型取得的 R^2 最高, 预测能力最强, 但模型的运行时间也更长。基于 mRMR-GA 的太赫兹信道场景识别算法选择重要性前 5 的信道特征作为数据集, 用最短的运行时间取得了较高的 R^2 指标。证明合理的特征选择可以有效提高本文算法的效率而不过分损失性能。实验表明, 相比没有使用簇特征, 使用簇特征的模型评估指标 R^2 更高, 说明簇特征在室内 LoS 和 NLoS 场景识别中具有重要作用。此外, 结合表 5 和表 6 可以看出, 算法对特征选择具有依赖性, 即不同的学习算法倾向于选择不同的特征子集。

表 5 BPNN 模型中不同信道特征参数集的性能

算法	特征组合	R^2	平均运行时间/s
BPNN	111111111111	0.996 4	3.72
BPNN	111110000000	0.925 2	3.05
BPNN+mRMR	100101001100	0.952 8	3.18
BPNN+RF	100100100011	0.977 6	2.85
BPNN+GA	101011010111	0.981 2	2.88
BPNN+mRMR-GA	101001010001	0.987 1	2.80

表 6 SVM 模型中不同信道特征参数集的性能

算法	特征组合	R^2	平均运行时间/s
SVM	111111111111	0.979 0	0.70
SVM	111110000000	0.893 0	0.68
SVM+mRMR	100101001100	0.932 2	0.62
SVM+RF	100100100011	0.965 7	0.59
SVM+GA	100011110101	0.967 6	0.61
SVM+mRMR-GA	101101000001	0.976 1	0.58

2) 算法抗噪性能分析

为了探究模型在室外环境中的适用性和鲁棒性, 在输入特征中叠加高斯噪声, 模拟室外信道衰减和干扰。实验共设置了无噪声 (1) 和 5 个噪声梯度 (2~6) 6 组, 噪声分别是对数据加入标准差为 0.1、0.2、0.3、0.4、0.5 的高斯噪声, 不同噪声梯度下的识别准确率如图 7 所示。

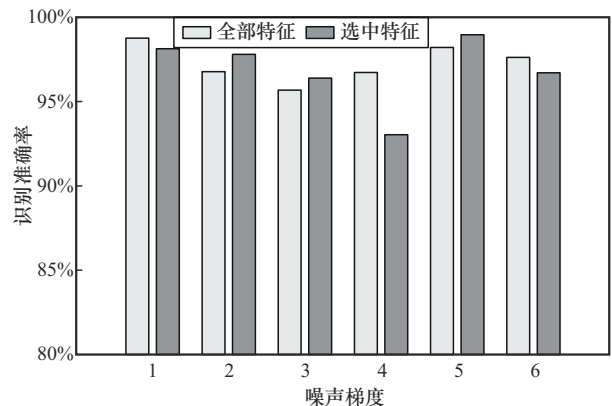


图 7 不同噪声梯度下的识别准确率

在无噪声情况下, 使用全部特征可以提高 LoS/NLoS 的识别准确率。随着噪声增大, 算法的识别准确率具有下降趋势, 冗余特征被噪声污染带来的弊端开始显现, 此时, 使用部分重要特征优于使用

全部特征。当噪声增大到一定程度后，基于 mRMR-GA 的太赫兹信道场景识别算法效果变差，但其识别率准确率仍高于 93%。

3) 机器学习算法对识别效果的影响

为进一步验证基于 mRMR-GA 的太赫兹信道场景识别算法的泛化能力，本节将其应用于多种经典机器学习模型 (BPNN、CNN、LSTM^[14]、Transformer^[22])，对比分析特征选择前后模型的识别性能差异，实验结果如表 7 所示。

表 7 mRMR-GA 特征选择算法在不同机器学习模型中的性能

模型	特征组合	识别准确率	训练时间/s
BPNN	全部特征	99.43%	2.10
	mRMR-GA 特征	90.83%	1.21
CNN	全部特征	93.12%	1.73
	mRMR-GA 特征	89.54%	1.35
LSTM	全部特征	99.28%	2.14
	mRMR-GA 特征	90.54%	1.40
Transformer	全部特征	98.71%	22.25
	mRMR-GA 特征	90.83%	14.20

使用全部特征时所有模型都可以取得较高的识别准确率，但训练时间更长。使用 mRMR-GA 对特征进行选择后模型的训练时间大幅下降，代价是降低了识别准确率。除 CNN 以外所有模型的识别准确率都超过了 90%，特别地，BPNN 在 1.21 s 的时间内完成训练并取得超过 90% 的识别准确率。

4) 算法泛化性能探究

获取室内环境中的所有信道数据并进行标注需要耗费巨大的测试成本，实际应用中算法的泛化性能至关重要。本节对具有典型特征的不同通信场景进行信道测量，并将不同场景的信道数据分别用于模型的训练和测试，以评估模型的泛化能力。

将开阔走廊和狭窄走廊的信道特征集分别作为训练集和测试集，测试算法迁移到走廊场景的识别性能。开阔走廊的 LoS 和 NLoS 场景分别有 300 条和 453 条数据，狭窄走廊的 LoS 和 NLoS 场景分别有 264 条和 492 条数据，实验使用识别错误率来衡量模型性能。

使用 mRMR、RF、GA 和 mRMR-GA 特征识别算法获取用于识别的信道特征子集，对应的信道特征组合编码及平均运行时间如表 8 所示。由于遗传

算法存在种群生成机制并且依赖于监督学习算法，每次迭代需要学习算法的评估指标 R^2 作为依据，所需的运行时间最长。mRMR-GA 特征选择算法通过 mRMR 准则在交叉变异过程中及时剔除冗余特征，相比 GA 缩短了 38.8% 的运行时间。RF 在学习算法中集成了特征选择机制，运行时间也较长。此外，mRMR、RF 获得的特征子集与 GA、本文算法获得的特征子集相差较大，这是因为 mRMR 和 RF 寻找最优特征子集的过程依赖训练数据集，而 GA 和本文算法在寻找最优特征子集的过程中将测试集上的性能指标作为指导。

表 8 不同算法特征子集

序号	特征来源	特征组合	平均运行时间/s
1	参照组	111111111111	—
2	参照组	111110000000	—
3	mRMR	101001100100	1.06
4	RF	101000100101	6.27
5	GA	100110111011	35.00
6	本文算法	100010101001	21.43

将表 8 中的信道特征子集输入 BPNN 和 SVM 中，信道场景识别结果如图 8 所示。

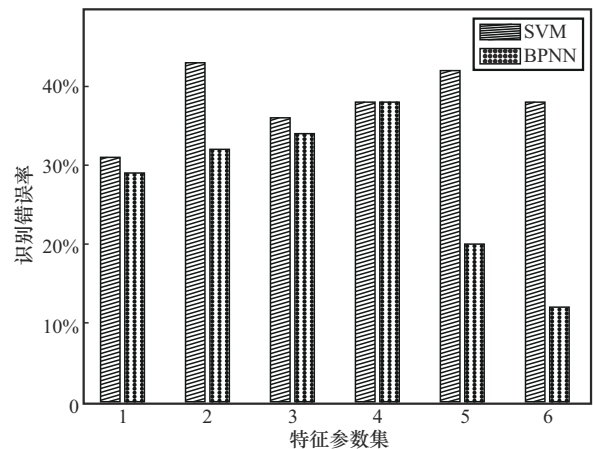


图 8 不同信道场景识别算法的识别错误率

图 8 中，采用 mRMR-GA 特征选择算法时不管是使用 BPNN 还是 SVM 都取得了优于其他特征提取算法的识别准确率。mRMR-GA 特征选择算法获得的特征子集在 BPNN 中识别错误率最低，仅为 12%，相比遗传算法^[19]降低了 8%。此外，与图 6 中算法在室内办公室场景中的识别效果比较可以看

出,在迁移到开阔走廊和狭窄走廊场景时大多数场景识别算法的识别性能有大幅度下降,但基于mRMR-GA的太赫兹信道场景识别算法依旧保持了较好的性能,表明本文算法具有较强的泛化能力。同时,也说明了在实际应用场景中,并非特征参数越多识别效果越好,特征参数的增加也可能导致信息冗余或者无关信息的出现,进而干扰正确识别。

4 结束语

本文研究了太赫兹通信系统的LoS/NLoS场景自动识别算法,针对太赫兹信道场景识别中所需特征繁多且识别效率低的问题,提出了一种基于mRMR-GA的太赫兹信道场景识别算法。该算法在遗传算法基础上嵌入了mRMR原则,通过保留与通信场景关联性更大的信道特征,减少了信道特征冗余,达到了提高算法寻优和收敛能力的目的。

仿真得到了3个典型场景下的CIR,并对本文算法性能进行了实验验证。结果表明,基于mRMR-GA的太赫兹信道场景识别算法可以提取具有代表性的信道特征,与遗传算法相比识别精度提高了8%,识别速度提高了38.8%。同时,本文算法还具有更好的泛化能力,可以迁移到不同环境的场景识别中。特征选择和场景识别的结合为复杂场景下的信号处理提供了新的思路和方法。未来的研究可以进一步优化算法流程,并将其应用于更广泛的无线通信和信号处理领域。

参考文献:

- [1] HUANG C, WANG C X, LI Z A, et al. A frequency domain predictive channel model for 6G wireless MIMO communications based on deep learning[J]. *IEEE Transactions on Communications*, 2024, 72(8): 4887-4902.
- [2] HUANG C, HE R S, AI B, et al. Artificial intelligence enabled radio propagation for communications: part II: scenario identification and channel modeling[J]. *IEEE Transactions on Antennas and Propagation*, 2022, 70(6): 3955-3969.
- [3] ZAFARI F, GKELIAS A, LEUNG K K. A survey of indoor localization systems and technologies[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(3): 2568-2599.
- [4] CHEN L C, DONG C, NIU K, et al. LSTM-embedding Transformer model for LoS/NLoS recognition in indoor localization[C]//*Proceedings of the 2023 IEEE/CIC International Conference on Communications in China (ICCC)*. Piscataway: IEEE Press, 2023: 1-5.
- [5] FU Z H, ZHANG Y, ZHAO X W, et al. A DNN-based channel state and scenario identifications for millimeter-wave communications[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(6): 8088-8093.
- [6] CUI Z C, GAO Y F, HU J, et al. LOS/NLOS identification for indoor UWB positioning based on morlet wavelet transform and convolutional neural networks[J]. *IEEE Communications Letters*, 2021, 25(3): 879-882.
- [7] CAO H J, WANG Y J, BI J X, et al. LOS compensation and trusted NLOS recognition assisted Wi-Fi RTT indoor positioning algorithm[J]. *Expert Systems with Applications*, 2024, 243: 122867.
- [8] NGUYEN T V, JEONG Y, SHIN H, et al. Machine learning for wide-band localization[J]. *IEEE Journal on Selected Areas in Communications*, 2015, 33(7): 1357-1380.
- [9] HUANG C, MOLISCH A F, WANG R, et al. Angular information-based NLOS/LOS identification for vehicle to vehicle MIMO system[C]//*Proceedings of the 2019 IEEE International Conference on Communications Workshops (ICC Workshops)*. Piscataway: IEEE Press, 2019: 1-6.
- [10] CHANG T T, JIANG S Y, SUN Y Z, et al. Multi-bandwidth NLOS identification based on deep learning method[C]//*Proceedings of the 2021 15th European Conference on Antennas and Propagation (EuCAP)*. Piscataway: IEEE Press, 2021: 1-5.
- [11] XIE J L, WANG W, XU Z L. Identification of NLOS condition in different scenarios based on wireless propagation features[C]//*Proceedings of the 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)*. Piscataway: IEEE Press, 2021: 568-571.
- [12] YANG M, AI B, HE R, et al. Machine-learning-based scenario identification using channel characteristics in intelligent vehicular communications[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(7): 3961-3974.
- [13] ZHOU T, WANG Y, LIN Y T, et al. Deep learning and hybrid fusion-based LOS/NLOS identification in substation scenarios for power Internet of Things[J]. *IEEE Internet of Things Journal*, 2024, 11(20): 33903-33914.
- [14] JIANG C H, SHEN J C, CHEN S, et al. UWB NLOS/LOS classification using deep learning method[J]. *IEEE Communications Letters*, 2020, 24(10): 2226-2230.
- [15] LI J L, LIU S L, GAO Y P, et al. UWB (N)LOS identification based on deep learning and transfer learning[J]. *IEEE Communications Letters*, 2024, 28(9): 2111-2115.
- [16] ZHANG J J, YI Q W, HUANG L, et al. Research on none-line-of-sight/line-of-sight identification method based on convolutional neural network-channel attention module[J]. *Sensors*, 2023, 23(20): 8552-8552.
- [17] PENG H C, LONG F H, DING C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, 27(8): 1226-1238.
- [18] YASSINE A, YOUNESS M, MOHAMED C, et al. A new noisy random forest based method for feature selection[J]. *Cybernetics and Information Technologies*, 2021, 21(2): 10-28.
- [19] TAN F, FU X Z, ZHANG Y Q, et al. A genetic algorithm-based method for feature subset selection[J]. *Soft Computing*, 2008, 12(2):

111-120.

- [20] SIDDIQI U F, SAIT S M, KAYNAK O. Genetic algorithm for the mutual information-based feature selection in univariate time series data[J]. IEEE Access, 2020, 8: 9597-9609.
- [21] NIE S, MACCARTNEY G R, SUN S, et al. 72 GHz millimeter wave indoor measurements for wireless and backhaul communications[C]// Proceedings of the 2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2013: 2429-2433.
- [22] SON J, KEUM I, KIM H, et al. Transformer-based environment-aware localization in the NLoS scenarios[C]//Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2024: 1-6.

[作者简介]



郝昕宇 (1996-), 男, 山东淄博人, 重庆邮电大学博士生, 主要研究方向为毫米波太赫兹信道测量与建模、信道预测、机器学习等。



廖希 (1988-), 女, 四川绵阳人, 博士, 重庆邮电大学副教授、博士生导师, 主要研究方向为 6G 毫米波太赫兹通信感知、6G 涡旋电磁波通信、散射反射通信等。



王洋 (1986-), 男, 重庆人, 博士, 重庆邮电大学教授、博士生导师, 主要研究方向为第六代移动通信技术、毫米波太赫兹信道测量与建模、涡旋电磁波、智能反射面等。



林峰 (1978-), 男, 山东烟台人, 重庆邮电大学高级工程师、硕士生导师, 主要研究方向为 5G-V2X 车联网与车路协同控制等。



罗娇 (1998-), 女, 重庆人, 重庆邮电大学硕士生, 主要研究方向为太赫兹信道测量与建模、机器学习、多径分簇等。



张杰 (1965-), 男, 山东临沂人, 博士, 重庆邮电大学特聘教授、博士生导师, 主要研究方向为室内-室外无线网络规划与优化、无线传播、小蜂窝和异构网络、自组织网络、智能建筑/电网、毫米波通信等。