

# 一种基于随机森林的OFDM系统自适应算法

王波 刘潇然\* 熊俊 辜方林 张晓瀛  
(国防科技大学电子科学学院, 湖南长沙 410073)

**摘要:** 针对动态变化的信道环境, 自适应正交频分复用(Orthogonal Frequency Division Multiplexing, OFDM)系统可以对子载波间隔和循环前缀长度进行调整, 以最大化系统的吞吐量。为了能够快速准确地找到OFDM系统在不同信道环境中的最优子载波间隔和循环前缀长度取值, 本文提出了基于随机森林的OFDM系统自适应算法。随机森林算法基于集成的思想, 能够有效处理高维度数据, 并且具有高效率、高准确率和强泛化能力等优势, 可以在复杂的数据场景下进行有效的分类。通过提取通信过程中信噪比、用户移动速度、最大多普勒频率和均方根时延扩展等信道特征与OFDM系统的子载波间隔和循环前缀长度组成训练样本, 利用随机森林算法创建了OFDM系统参数多分类模型。所提模型可以根据输入的信道特征, 实现OFDM系统子载波间隔和循环前缀长度的自适应分配。同时, 针对训练样本主要集中在少数几个系统参数类别的情况, 利用合成少数类过采样技术对较少样本数的类别进行扩充, 满足了随机森林算法对训练样本类别平衡化的需求, 进一步提高了算法的分类准确率。相比传统的自适应算法, 所提算法具有更高的分类准确率和模型泛化能力。分析和仿真结果表明, 与子载波间隔和循环前缀长度固定的OFDM系统相比, 本文所提出的自适应算法能够准确选择出最优的系统参数, 可以有效地减轻信道中符号间干扰和子载波间干扰的影响, 从而在整个信噪比范围上提供最大的平均频谱效率。基于随机森林的OFDM系统自适应算法能够动态地分配子载波间隔和循环前缀长度, 增强OFDM系统的通信质量和抗干扰能力, 实现在不同信道环境下的可靠传输。

**关键词:** 正交频分复用; 合成少数类过采样技术; 随机森林; 自适应算法

**中图分类号:** TN92 **文献标识码:** A **DOI:** 10.16798/j.issn.1003-0530.2024.06.002

**引用格式:** 王波, 刘潇然, 熊俊, 等. 一种基于随机森林的OFDM系统自适应算法[J]. 信号处理, 2024, 40(6): 1007-1018. DOI: 10.16798/j.issn.1003-0530.2024.06.002.

**Reference format:** WANG Bo, LIU Xiaoran, XIONG Jun, et al. Random forest-based adaptive algorithm for OFDM system[J]. Journal of Signal Processing, 2024, 40(6): 1007-1018. DOI: 10.16798/j.issn.1003-0530.2024.06.002.

## Random Forest-Based Adaptive Algorithm for OFDM System

WANG Bo LIU Xiaoran\* XIONG Jun GU Fanglin ZHANG Xiaoying

(College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China)

**Abstract:** In order to adapt to a dynamic channel environment, adaptive orthogonal frequency division multiplexing (OFDM) systems can adjust the subcarrier spacing and cyclic prefix length to maximize the system throughput. In order to quickly and accurately determine the optimal values for the subcarrier spacing and cyclic prefix length for an OFDM system operating in different channel environments, this study investigated a random forest-based adaptive algorithm for OFDM systems. The random forest algorithm, which is based on the principle of ensemble learning, is capable of effectively handling high-dimensional data

收稿日期: 2023-07-03; 修回日期: 2023-08-24

\*通信作者: 刘潇然 liuxiaoran10@nudt.edu.cn \*Corresponding Author: LIU Xiaoran, liuxiaoran10@nudt.edu.cn

基金项目: 国家自然科学基金(62101569); 湖南省自然科学基金(2022JJ30047)

Foundation Items: The National Natural Science Foundation of China (62101569); The Natural Science Foundation of Hunan Province (2022JJ30047)

and possesses advantages such as high efficiency, high accuracy, and strong generalization ability. It can effectively be used for classification in complex data scenarios. A dataset was created to train a multi-classification model based on a random forest for OFDM system parameters by extracting channel characteristics, including the signal-to-noise ratio (SNR), user movement speed, maximum Doppler frequency, and root mean square delay spread during the communication process, and then combining them with the subcarrier spacing and cyclic prefix length of the OFDM system. The proposed model could adaptively allocate the subcarrier spacing and cyclic prefix length in an OFDM system based on the input channel characteristics. Meanwhile, because the training samples were probably concentrated in a few categories, a synthetic minority oversampling technique was utilized to augment the samples in other categories with fewer samples. The classification accuracy of the algorithm was further improved by meeting the requirement of category balancing in the training samples for the random forest algorithm. Compared to traditional adaptive algorithms, the proposed algorithm exhibited a higher classification accuracy and model generalization capability. The analysis and simulation results indicated that compared to OFDM systems with a fixed subcarrier spacing and cyclic prefix length, the proposed adaptive algorithm could accurately select the optimal system parameters. It effectively mitigated the impact of inter-symbol interference and inter-subcarrier interference in the channel, thereby providing the maximum average spectral efficiency across the entire SNR range. In conclusion, the random forest-based adaptive algorithm for OFDM systems could dynamically allocate the subcarrier spacing and cyclic prefix length, enhance the communication quality and interference resilience of the OFDM system, and enable reliable transmission in different channel environments.

**Key words:** orthogonal frequency division multiplexing; synthetic minority oversampling technique; random forest; adaptive algorithm

## 1 引言

正交频分复用(Orthogonal Frequency Division Multiplexing, OFDM)具有抗信道频率选择性衰落能力强、频带利用率高、适合高速率数据传输等优点,广泛应用于数字广播电视系统、无线局域网和蜂窝移动通信等业务中<sup>[1]</sup>。然而,OFDM信号在复杂多变的无线信道中传输时,容易受到多径效应和多普勒效应的影响,造成符号间干扰(Inter-symbol Interference, ISI)和子载波间干扰(Inter-carrier Interference, ICI)<sup>[2]</sup>。为了减轻ISI和ICI的影响,提高OFDM系统的容量,自适应OFDM系统通过采取灵活调整子载波间隔和循环前缀(Cyclic Prefix, CP)长度等方法以适应不同的信道环境<sup>[3-8]</sup>。

传统的链路自适应技术包括自适应功率控制和自适应调制编码(Adaptive Modulation and Coding, AMC)等。自适应功率控制可以根据信道质量动态地调整发送功率,而AMC可以根据实时的信道状态信息和可靠性要求对调制编码方式进行调整。然而,传统的链路自适应技术在信道导致的ISI和ICI情况下难以发挥作用<sup>[9]</sup>。混合参数集可以看作链路自适应技术的发展和延伸。文献[10]中指出基于混合参数集的自适应OFDM将子载波间隔和CP长度定义为可变的波形参数集,在不同的信道环境下,通过动态地调整参数集为不同的用户提供最匹配的通信波形进行传输,用定制化的波形服务不同的用户和设施。

在单用户OFDM通信系统中,针对多径信道环境带来的ISI,文献[11]提出了一种具有可变CP长度的自适应OFDM系统。该系统使用不同长度CP来适应信道时延扩展的变化,能够在满足误码率性能的情况下最大限度地提高频谱效率。为了减轻信道中ICI的影响,文献[12]提出了一种子载波间隔自适应的OFDM系统。推导了系统容量的闭合表达式,并根据信道的多普勒条件和信噪比(Signal to Noise Ratio, SNR)等信息来动态地改变OFDM系统的子载波间隔,能够有效地减轻由高移动性引起的ICI的影响,从而最大化系统在时变信道中的平均频谱效率(Average Spectral Efficiency, ASE)。文献[13]分析了双选择性信道中时延扩展和多普勒扩展等对OFDM系统性能的影响,在子载波间隔和其他系统参数之间折中选择,找到了最大化信干比的参数。然而,上述研究都是基于特定场景和信道环境中的理论推导,没有将通信过程中产生的已有数据利用起来。

为了建立复杂信道特征和多个系统参数之间的映射关系,文献[7]提出了一个基于机器学习的自适应系统。文献[14]将调制编码方式和CP长度作为候选参数集,通过K近邻(K-Nearest Neighbor, KNN)算法对信道环境进行在线学习,决策出最优的参数集。然而,机器学习算法的收敛速度较慢,往往存在决策时间长的问题。文献[15]中将信道的SNR、时延扩展和多普勒扩展以及对应的系统参数组成三维的查找表,将通信过程中产

生的已有数据利用起来,通过查表的方式快速选择合适的参数。然而,查表的方式完全依赖表中的数据,无法应对未知的信道环境,泛化能力较弱。

随机森林(Random Forest, RF)<sup>[16-17]</sup>算法基于集成的思想,能够处理高维度数据,具有效率高、准确率高和泛化能力强等优势而被广泛用来解决各种分类问题。本文提出了一种基于 RF 的单用户 OFDM 系统自适应算法。通过利用通信过程中产生的数据作为训练 RF 的样本,将 SNR、均方根(Root of Mean Square, RMS)时延扩展、多普勒条件等信息作为信道特征向量,建立了基于 RF 的

OFDM 系统参数多分类模型,实现不同信道环境下的系统参数自适应分配。针对样本数据不平衡的特点,利用合成少数类过采样技术(Synthetic Minority Oversampling Technique, SMOTE)<sup>[18-20]</sup>扩充了用于 OFDM 系统参数多分类的样本数据,提高了算法的准确率和泛化能力。

## 2 系统模型

图 1 给出了联合子载波间隔和 CP 长度的自适应 OFDM 系统模型。信道特征是时变的,自适应算法可以根据当前的信道状态信息动态地调整子载波间隔和 CP 长度。

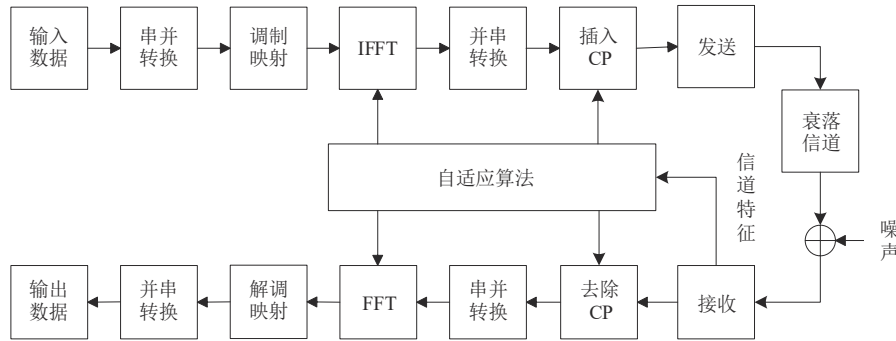


图 1 联合子载波间隔和 CP 长度的自适应 OFDM 系统

Fig. 1 Adaptive OFDM system combining subcarrier spacing and CP length

OFDM 信号在复杂的无线信道中传输时,容易受到多径效应和多普勒效应的影响,造成 ISI 和 ICI。其中,时变信道引起的 ICI 功率  $P^{ICI_1}$  表达式为<sup>[12]</sup>

$$P^{ICI_1} = \frac{\bar{P}}{6} (\pi f_{\max} T_s N)^2 \quad (1)$$

式中,  $\bar{P}$  为平均发送功率,最大多普勒频率为  $f_{\max} = f_c v/c$ ,  $f_c$  是载波频率,单位 Hz;  $v$  是用户移动速度 (km/h),  $c$  是光速。  $T_s = 1/B$  表示采样周期,  $B$  是系统带宽,  $N$  是子载波数。系统的总带宽保持不变,子载波间隔表示为  $\Delta f = B/N$ 。

CP 不足引起的 ISI 功率  $P^{ISI}$  和 ICI 功率  $P^{ICI_2}$  表达式为<sup>[4]</sup>

$$P^{ISI} = P^{ICI_2} = \begin{cases} \frac{\bar{P}}{N} \frac{L - L_{CP} + b(1 - e^{-a(L - L_{CP})})}{(1 - e^{-aL})} & L > L_{CP} \\ 0 & L \leq L_{CP} \end{cases} \quad (2)$$

式中,  $L$  表示信道路径数,  $L_{CP}$  是 CP 长度。  $b = (1 - e^{-a})^{-1}$ ,  $a = T_s/T_d$ ,  $T_d$  表示信道的最大时延扩展。

文献[5]中的分析表明,ICI 随着子载波间隔的

增大而减小。因此,在固定带宽下增加子载波间隔,即减小子载波数,可以减轻 ICI 的影响。但是,较大的子载波间隔不利于 OFDM 系统抵抗频率选择性衰落,可能会导致更多的 ISI。可以通过在 OFDM 符号中插入足够长的 CP 以消除 ISI。但是,插入的 CP 也引入了系统开销,降低了频谱利用率。因此,CP 长度需要根据时变信道 RMS 时延扩展的变化而动态地调整。

为了能够根据信道环境特征,动态地调整 OFDM 系统的子载波间隔和 CP 长度,使得 OFDM 系统可以有效地抵御信道中的 ISI 和 ICI,本文利用通信过程中产生的已有数据作为训练随机森林(RF)的样本,建立多分类模型,实现对 OFDM 系统参数的动态分配。样本中信道特征和子载波间隔、CP 长度组成的系统参数集是映射关系,因此,子载波间隔和 CP 长度可以组成用于多分类的标签:

$$\mathbf{y} = [\Delta f, L_{CP}] \quad (3)$$

假设共组成  $I$  种标签,从而标签向量可以表示为  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_I]$ 。将信道的平均信噪比  $\bar{\gamma}$ , 用户移动速度  $v$ , 最大多普勒频率  $f_{\max}$ , RMS 时延扩展  $\sigma_\tau$ ,

和信道路径数  $L$  构成信道特征:

$$\mathbf{x} = [\bar{\gamma}, v, f_{\max}, \sigma_{\tau}, L] \quad (4)$$

假设不同信道特征组成空间大小为  $n$  的信道特征向量  $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ , 则信道特征和系统参数集之间的一一映射关系可以表示为

$$\begin{aligned} & \mathbf{x}_j \rightarrow \mathbf{y}_i \\ \text{s. t. } & \begin{cases} \mathbf{x}_j \in \mathbf{X} & 1 \leq j \leq n \\ \mathbf{y}_i \in \mathbf{Y} & 1 \leq i \leq I \end{cases} \end{aligned} \quad (5)$$

RMS 时延扩展由功率时延分布 (power delay profile, PDP) 二阶中心距的平方根给出:

$$\sigma_{\tau} = \sqrt{\frac{\sum_{l=0}^{L-1} \tau_l^2 P(\tau_l)}{\sum_{l=0}^{L-1} P(\tau_l)} - \left( \frac{\sum_{l=0}^{L-1} \tau_l P(\tau_l)}{\sum_{l=0}^{L-1} P(\tau_l)} \right)^2} \quad (6)$$

其中,  $\tau_l$  和  $P(\tau_l)$  分别表示第  $l$  条路径的信道时延和功率。

考虑平均频谱效率 (ASE) 作为衡量系统性能指标, 根据文献 [12], ASE 表示为

$$C_{\text{ASE}} = \frac{N}{(N+V)\ln 2} \text{Ei} \left( 1, \frac{P^{\text{IC1}_1} + P^{\text{ISI}} + P^{\text{IC2}_2} + \sigma_w^2}{\lambda(\bar{P} - P^{\text{IC1}_1})} \right) \exp \left( \frac{P^{\text{IC1}_1} + P^{\text{ISI}} + P^{\text{IC2}_2} + \sigma_w^2}{\lambda(\bar{P} - P^{\text{IC1}_1})} \right) \quad (7)$$

其中,  $\text{Ei}(1, x) = \int_x^{\infty} (e^{-t}/t) dt$  表示指数积分,  $\sigma_w^2$  表示复高斯加性白噪声的方差。常数  $\lambda = -1.6/\ln(\varepsilon/0.2)$ ,  $\varepsilon$  表示误码率约束为  $10^{-3}$ 。通过优化子载波间隔和 CP 长度, 使得 ASE 最大, 该优化问题可以建模为

$$(\Delta f, L_{\text{CP}})^* = \arg \min_{(\Delta f, L_{\text{CP}}) \in \mathbf{Y}} (C_{\text{ASE}} | \mathbf{X}) \quad (8)$$

由于 ASE 表达式计算困难, 因此, 基于 ASE 表达式提出的自适应算法往往比较复杂, 决策时间长。针对这一问题, 本文创建了基于 RF 的多分类模型, 实现 OFDM 系统参数自适应调整, 提高决策效率。

### 3 基于随机森林的自适应算法

本文所提 OFDM 系统自适应模型如图 2 所示。

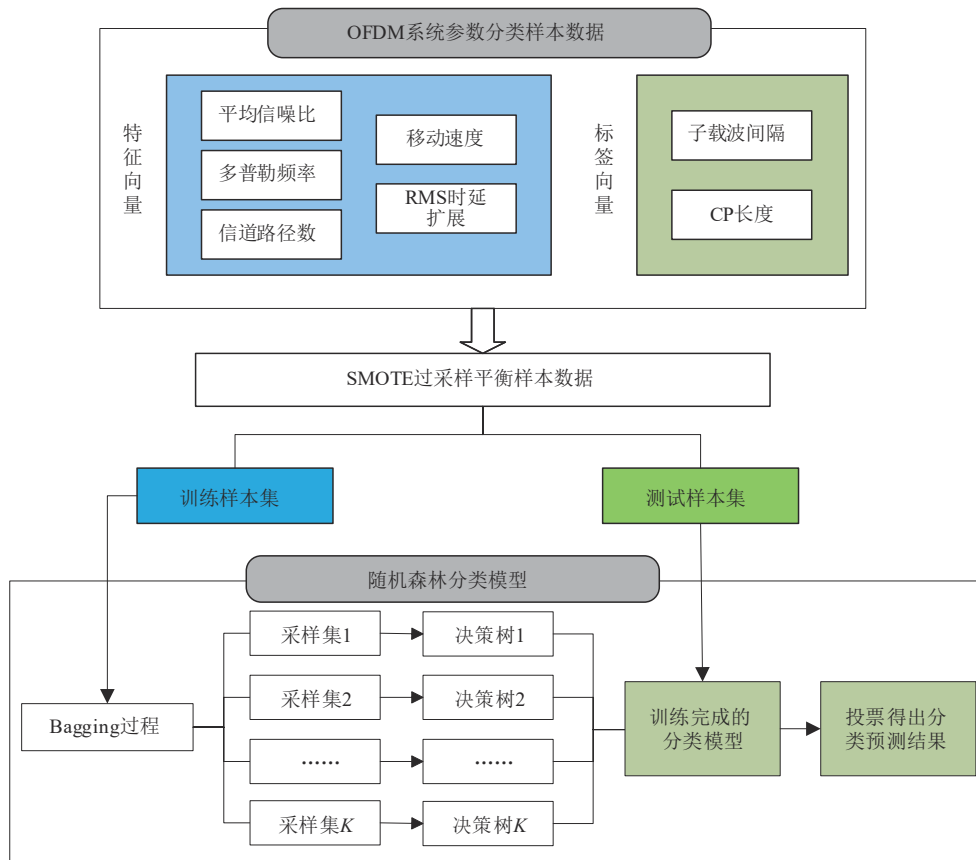


图2 基于STOME-RF的OFDM系统自适应模型

Fig. 2 STOME-RF-based adaptive model for OFDM system



由于OFDM通信过程中产生的样本数据中,部分类别的样本数量较少,因此存在数据不平衡的问题,这可能会降低随机森林(RF)算法的多分类准确率。针对这一问题,本文采取SMOTE算法对样本数据进行过采样,扩充少数类别的样本数量,以平衡样本数据,满足RF算法对样本类别平衡化的需求。进一步,利用RF算法对样本数据进行多分类,按照算法2的流程生成决策树,共生成 $K$ 棵分类决策树,形成“森林”。最后在仿真部分,利用训练数据集对建立的多分类模型进行训练,利用测试数据集测试模型的多分类准确率。本文提出的基于SMOTE-RF的OFDM系统自适应算法能够针对不同的信道特征,快速地找出一组最优的子载波间隔和CP长度,最大化系统的ASE。

### 3.1 样本预处理

清除数据中重复或样本值缺失的样本后,基于原样本特征的均值和标准差,使用Z-Score方法进行特征的标准化。信道特征向量 $\mathbf{X}$ 经过标准化后的新特征向量 $\mathbf{X}_{\text{std}}$ 为

$$\mathbf{X}_{\text{std}} = \frac{\mathbf{X} - \bar{\mathbf{X}}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})^2}} \quad (9)$$

其中 $\bar{\mathbf{X}} = 1/n \sum_{i=1}^n \mathbf{X}_i$ 表示均值。

为了消除样本特征之间的相关性,利用零相位成分分析(Zero-phase Component Analysis, ZCA)的方法对样本特征进行白化操作。ZCA白化是在主成分分析(Principal Component Analysis, PCA)白化上进行旋转,ZCA白化后的样本具有原样本的所有特征,因此,更加接近原样本。ZCA白化的具体步骤如下。

**步骤1** 计算样本特征向量的协方差矩阵为

$$\boldsymbol{\Sigma} = \frac{1}{n} \mathbf{X} \mathbf{X}^T \quad (10)$$

其中, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ 表示特征向量, $\mathbf{x}_i$ 是特征向量 $\mathbf{X}$ 的列向量, $(\cdot)^T$ 表示转置运算。

**步骤2** 利用特征分解的方法求出 $\boldsymbol{\Sigma}$ 的特征向量和特征值,将特征向量作为列向量组成特征矩阵 $\mathbf{U}$ 为

$$\mathbf{U} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m] \quad (11)$$

其中, $m$ 表示特征的数量。 $\mathbf{U}$ 中的特征向量按照对应的特征值的大小排列。

**步骤3** 将原始数据的特征向量转换到特征矩阵 $\mathbf{U}$ 构建的新空间中,得到数据集转换后的结果 $\mathbf{X}_{\text{rot}}$ 为

$$\mathbf{X}_{\text{rot}} = \mathbf{U}^T \mathbf{X} = \begin{bmatrix} \boldsymbol{\xi}_1 \mathbf{X} \\ \boldsymbol{\xi}_2 \mathbf{X} \\ \vdots \\ \boldsymbol{\xi}_m \mathbf{X} \end{bmatrix} \quad (12)$$

**步骤4** 按照式(13)对数据集进行PCA白化,使得数据的每个特征具有相同的方差。

$$\mathbf{X}_{\text{PCA},i} = \frac{\mathbf{X}_{\text{rot},i}}{\sqrt{\lambda_i}} \quad (13)$$

其中, $\lambda_i$ 是 $\mathbf{X}_{\text{rot}}$ 矩阵对角元素的值。

**步骤5** 将 $\mathbf{X}_{\text{PCA},i}$ 左乘特征矩阵 $\mathbf{U}$ ,进行ZCA白化,如式(14)所示。ZCA白化后的数据被映射回源空间,从而更加接近原始数据。

$$\mathbf{X}_{\text{ZCA},i} = \mathbf{U} \mathbf{X}_{\text{PCA},i} \quad (14)$$

### 3.2 样本平衡

SMOTE算法是在特征空间中对少数类别进行过采样,以平衡样本数据。如式(15)所示,取样本 $z_1$ 与其最近邻居 $z_2$ 之间的差值,然后将差值乘以0和1之间的随机数,产生“虚拟”样本 $z_3$ ,并将其添加到原始样本中,以平衡样本数据类别。图3(a)显示了具有三个类别的不平衡样本,图3(b)展示了使用SMOTE过采样算法对不平衡样本进行采样后的结果。可以观察到,经过采样后,原样本数量较少的类别B和类别C被扩充,样本数据达到了平衡。

$$z_3 = z_1 + \text{rand}(0, 1) \times |z_1 - z_2| \quad (15)$$

利用SMOTE对样本数据进行过采样的具体流程如算法1所示。

---

#### 算法1 SMOTE算法

---

输入:同类别样本的近邻数 $\alpha$ 。

1. 在OFDM系统参数多分类样本的特征空间中,随机选取某一少数分类样本 $z_1$ 。计算与 $z_1$ 同类型的 $M$ 个样本的欧式距离。
2. 根据计算的样本之间的距离,选取与样本 $z_1$ 最近的 $\alpha$ 个同类样本。然后在 $\alpha$ 个同类样本中,随机选择一个样本 $z_2$ ,根据式(15)合成新的样本 $z_3$ 。
3. 循环上述步骤直至所有少数类别的样本扩充完成。

输出:平衡化的样本集。

---

### 3.3 自适应算法流程

RF并联地创建多个决策树模型,每个决策树模型之间是相互独立的,使用相同训练样本集。在构建决策树模型时,每棵决策树利用Bootstrap抽样方法从训练样本集中进行有放回的随机抽样,生成新的训练样本集,同时随机选取训练样本集的部分特征组成新的特征向量进行训练。每棵决策树的训

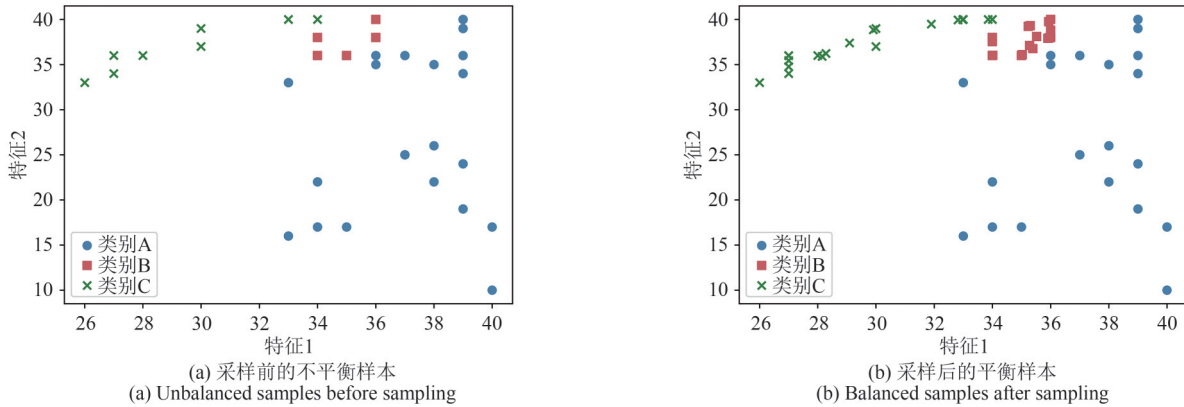


图3 SMOTE过采样结果

Fig. 3 SMOTE oversampling results

训练样本和特征各不相同,因此得出的训练结果也不相同,提高了模型的鲁棒性和泛化能力。决策树在生成过程中不进行剪枝处理。假设训练样本数据集的容量为  $n_1$ ,则RF创建决策树的具体流程如算法2所示。

**算法2** RF创建第  $k$  棵决策树算法

输入:训练样本集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , 最大深度和最大特征数。

1. 利用Bootstrap抽样方法从训练样本集中进行有放回地随机抽取第  $k$  个训练样本,抽取的样本容量和输入的原训练样本相同。
2. 随机抽取特征向量子集,选择最优的特征属性。
3. 根据当前选择的最优特征属性,将数据集划分成两个部分,建立当前节点的左、右子节点。
4. 判断决策树训练是否结束,若训练结束,则生成第  $k$  棵决策树;否则,对左、右子节点递归调用第1~3步。

输出:第  $k$  棵完整的分类决策树。

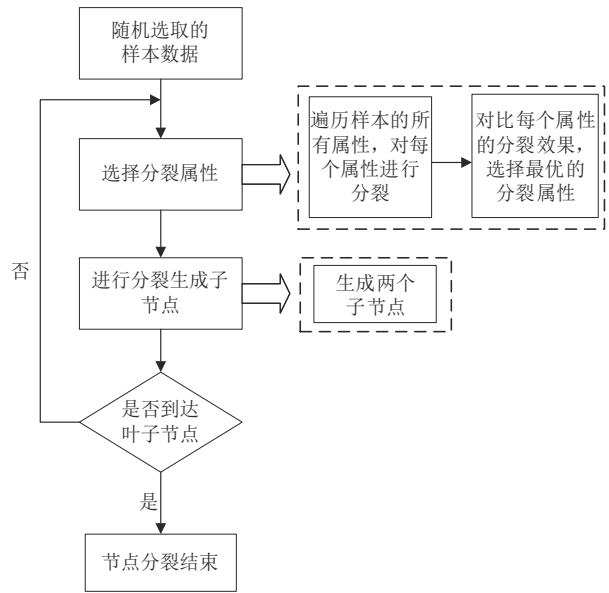


图4 决策树节点分裂流程图

Fig. 4 Node-splitting flowchart of decision tree

从算法2中可以看出,RF模型在建立过程中,为了满足森林中决策树多样性的要求,需要对训练数据进行随机抽样,包括样本随机抽样和特征随机抽样。然后选择最优的特征属性,建立节点。决策树建立节点的过程如图4所示,节点分裂依据基尼系数最小的原则。基尼系数是一个用于计算系统混乱程度的指标,计算公式为

$$Gini(D) = 1 - \sum_{i=1}^m P_i^2 \quad (16)$$

其中,  $P_i$  是类别为  $i$  的样本在样本集  $D$  中出现的概率。使用选择的属性进行分类后的基尼系数公式为

$$Gini(D) = \frac{s_1}{s_1 + s_2} Gini(D_1) + \frac{s_2}{s_1 + s_2} Gini(D_2) \quad (17)$$

其中,  $s_1, s_2$  为划分成两类的样本量,  $Gini(D_1)$  和  $Gini(D_2)$  分别为划分后的两个类别的基尼系数。

基尼系数的值越大,表示系统的混乱程度越高。决策树模型的建立过程就是不断地利用最优属性进行节点分裂,以划分样本降低训练样本集的混乱程度,最终使得相同类别的样本在同一个节点上。

利用SMOTE算法对样本数据进行平衡化处理后,进一步,利用RF算法对OFDM系统参数进行自适应分配。基于RF算法的自适应流程图和步骤分别如图5和算法3所示。

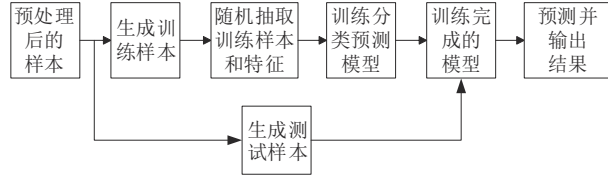


图5 基于RF的OFDM系统自适应流程图

Fig. 5 RF-based adaptive flowchart for OFDM system

**算法3** 基于RF的OFDM系统自适应算法

1. 从训练样本集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_l)\}$  中有放回的随机抽取一个样本容量也为  $n_1$  的 Bootstrap 训练样本, 然后等概率的从特征集中随机抽取特征子集组成新的特征向量, 作为 RF 中单棵决策树的训练样本。
2. 从抽取的特征向量中选择最优属性进行节点分裂, 生成一棵完整的决策树, 决策树在生成的过程中不进行减枝操作。
3. 重复步骤 1 和步骤 2, 训练  $K$  轮生成决策树集合, 构成 RF 模型。
4. 利用训练完成的模型对测试样本进行分类, 根据式(18),  $K$  棵决策树多数投票得出最终结果。

$$H(x) = \arg \max_{y \in Y} \sum_{k=1}^K Q(h_k(x) = y) \quad (18)$$

其中,  $H(x)$  表示集成的分类模型,  $h_k(x)$  表示第  $k$  棵决策树分类模型,  $Q(\cdot)$  表示示意性函数。

## 4 仿真及结果分析

### 4.1 生成样本数据

利用 MATLAB 仿真软件搭建 OFDM 通信链路平台, 仿真模拟 OFDM 在双选择性信道环境下的通信。共生成 900 个样本数据用于训练和测试 SMOTE-RF 自适应模型。其中, 用于多分类的系统参数为子载波数  $N$  和 CP 比率  $R$  各 3 种, 从而, 子载波间隔为  $\Delta f = B/N$ , CP 长度为  $L_{cp} = N * R$ , 共 9 种系统参数的组合构成标签向量。具体仿真参数如表 1 所示。

表 1 OFDM 系统参数设置

Tab. 1 Parameter settings of OFDM system

参数	值
系统带宽/MHz	5
载波频率/GHz	2
训练样本占比	0.8
测试样本占比	0.2
子载波数	128, 256, 512
CP 比率	0.25, 0.125, 0.0625

### 4.2 SMOTE-RF 模型评价

由 OFDM 系统自适应算法可知, 随机森林 (RF) 在建立的过程中, 会并行地创建多棵决策树, 算法的分类结果由所有决策树多数投票决定。因此, 决策树是 RF 模型的核心, 决策树的数量对 RF 模型分类的准确率和效率产生直接影响, 是 RF 建模过程中必须考虑的关键参数。本小节以决策树的数量为变量, 采用袋外 (Out of Bag, OOB) 数据误差和均方误差 (Mean-square Error, MSE) 两个指标, 对 SMOTE-RF 模型的泛化能力和鲁棒性进行评价。

考虑到训练样本数据集的容量为  $n_1$ , 则在一次随机采样中, 一个样本没有被采集到的概率是  $1 - 1/n_1$ , 那么  $n_1$  次采样都没有被采集到的概率是  $(1 - 1/n_1)^{n_1}$ 。当  $n_1 \rightarrow \infty$  时, 有

$$\lim_{n_1 \rightarrow \infty} \left(1 - \frac{1}{n_1}\right)^{n_1} = \frac{1}{e} \approx 0.368 \quad (19)$$

因此, 在每一轮随机采样中, 训练样本集中大约有 36.8% 的样本没有参与决策树的生成, 称之为袋外 (OOB) 数据。可以使用这些 OOB 数据检测 RF 模型的泛化能力, 得到 OOB 数据误差, OOB 数据误差越小, 模型的泛化能力越好。图 6(a) 展示了传统 RF 算法和本文所提 SMOTE-RF 算法的 OOB 数据误差曲线, 从图中可以看出, SMOTE-RF 算法的 OOB 数据误差明显低于传统 RF 算法。除此之外, SMOTE-RF 算法的 OOB 数据误差随着决策树数量的增加而降低, 但是当树的数量超过 55 时, OOB 数据误差基本保持平稳。

均方误差 (MSE) 是反映估计量与被估计量之间差异程度的一种度量, 可以采用 MSE 来衡量模型分类的有效性。假设测试样本数据集的容量为  $n_2$ , 对于测试样本集  $X = (x_1, x_2, \dots, x_{n_2})$ , 设  $Y = (y_1, y_2, \dots, y_l)$  是  $X$  的真实标签,  $\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_l)$  是测试样本集的分类结果。 $\hat{Y}$  的均方误差定义为

$$\text{MSE}(\hat{Y}) \sim E(\hat{Y} - Y)^2 \quad (20)$$

经过计算可得

$$\begin{aligned} \text{MSE}(\hat{Y}) &= E(\hat{Y} - Y)^2 = E\left[\left(\hat{Y} - E\hat{Y}\right) - \left(y - E\hat{Y}\right)\right]^2 = \\ &E(\hat{Y}^2) - (E\hat{Y})^2 + (E\hat{Y} - Y)^2 = \text{Var}(\hat{Y}) + \text{Bias}^2(\hat{Y}) \end{aligned} \quad (21)$$

式 (21) 说明, 均方误差  $\text{MSE}(\hat{Y})$  是由方差  $\text{Var}(\hat{Y})$  和偏差  $\text{Bias}(\hat{Y}) = |E\hat{Y} - Y|$  的平方两部分组



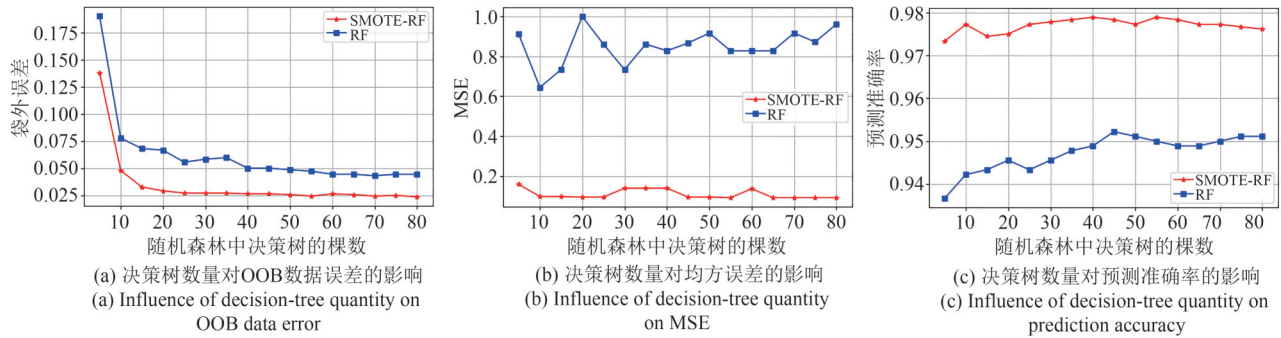


图6 决策树数量对模型性能的影响

Fig. 6 Influence of decision-tree quantity on model performance

成。偏差描述了RF算法的分类结果与真实标签的偏离程度,即算法本身的拟合能力;方差度量了训练样本集的变动导致算法性能的变化,即数据扰动的影响。为了使得模型具有较好的泛化能力,则需要减小偏差使算法能够充分拟合数据;同时,尽可能减小方差来降低数据扰动的影响。图6(b)展示了决策树的数量对MSE的影响,从图中可以观察到,所提SMOTE-RF算法的MSE明显低于传统RF算法,证明了SMOTE-RF算法是有效的,降低了传统RF算法的分类误差。进一步,可以观察到当决策树的数量为55时,SMOTE-RF算法的MSE最低。

如图6(c)所示,采用五折交叉验证的方法计算出不同决策树数量下的模型预测准确率。可以看到,SMOTE-RF算法的预测准确率明显高于传统RF算法。并且,当决策树数量为40或55时,SMOTE-RF算法的预测准确率最高。因此,综合上述分析,在实际仿真中,选择决策树的数量为55比较合适。同时,本文采用网格搜索的方法得到了SMOTE-RF中决策树的最大深度和最大特征数两个超参数的最优值。具体的超参数设置如表2所示。

表2 STOME-RF算法中的超参数设置

Tab. 2 Hyper-parameter settings for STOME-RF algorithm

超参数	参数值
决策树数量	55
最大深度	19
最大特征数	4

### 4.3 SMOTE-RF模型性能分析

#### 4.3.1 不同算法分类性能比较

为了评估比较不同分类算法的性能,观察模型

在不同类别上的分类情况,图7比较了STOME-RF算法、RF算法、决策树算法和KNN算法的混淆矩阵。矩阵中非对角线上的元素表示分类错误的类别数量,而对角线上的元素表示分类正确的类别数量。从图中可以明显地观察到,所提SMOTE-RF算法基本上在每个类别上都具有良好的分类表现,正确分类的效果明显优于其他算法。

根据混淆矩阵,分别计算出了STOME-RF算法、RF算法、决策树算法和KNN算法的准确率、精准率、召回率和F1分数,作为衡量模型分类性能的数字指标。数字评价指标的值越接近1,表示模型的性能越好,具体结果如表3所示。从表中可以看到,STOME-RF算法的准确率、精准率、召回率和F1分数都是最高的。其中,STOME-RF算法的准确率为99.17%,相较于KNN算法、决策树算法和RF算法,准确率分别提高了14.17%、4.73%和3.06%。证明了所提STOME-RF算法具有较高的可靠性和准确性。

为了进一步评估和分析STOME-RF算法的鲁棒性和泛化能力,利用MATLAB仿真软件搭建的OFDM通信链路平台生成了不同信道环境下的通信样本数据,对所提算法在不同数据集上进行测试。其中样本数据集1包含4070条数据,用于多分类的系统参数为5种子载波数和3种CP比率,共15种系统参数的组合构成标签向量。样本数据集2包含18820条数据,并进一步扩充了信道环境特征,增加了瞬时SNR特征,用于多分类的系统参数为5种子载波数、3种CP比率和9种调制方式,共135种系统参数的组合构成标签向量。

表4和表5分别计算出了STOME-RF算法、RF算法、决策树算法和KNN算法在样本数据集1和样本数据集2下的准确率、精准率、召回率和F1分数。在表4中,STOME-RF算法的准确率为98.59%,相



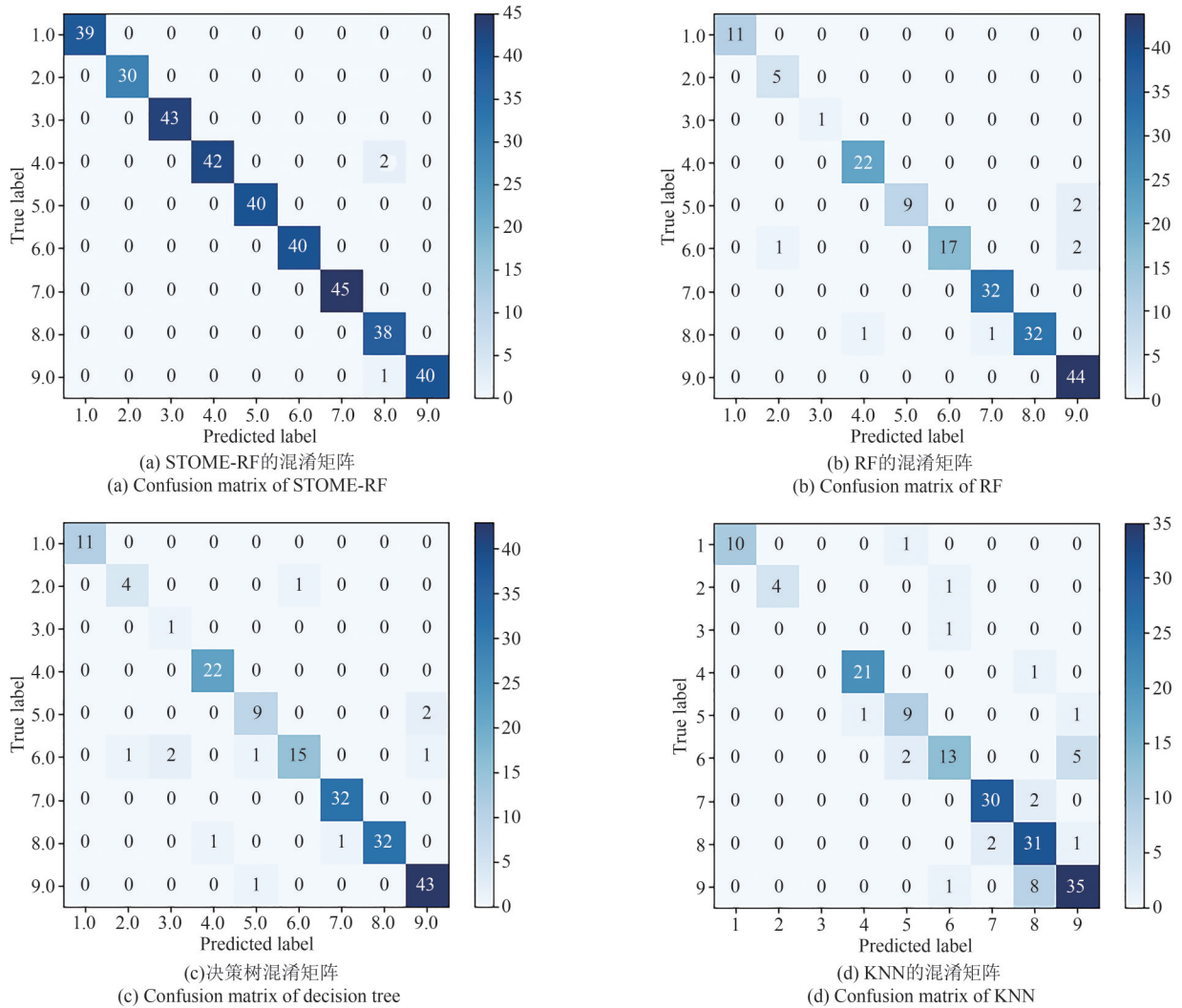


图 7 混淆矩阵对比图

Fig. 7 Comparison chart of confusion matrix

表 3 机器学习算法分类准确率比较

Tab. 3 Comparison of classification accuracies of machine-learning algorithms

使用算法	准确率	精准率	召回率	F1 分数
KNN	85.00%	78.07%	75.29%	76.31%
决策树	94.44%	86.11%	92.07%	87.10%
RF	96.11%	96.40%	95.66%	95.74%
STOME-RF	99.17%	99.22%	99.16%	99.17%

表 4 机器学习算法在数据集 1 上的测试结果比较

Tab. 4 Comparison of test results of machine-learning algorithms on dataset 1

使用算法	准确率	精准率	召回率	F1 分数
KNN	95.58%	95.33%	91.91%	93.56%
决策树	95.80%	95.58%	91.58%	93.40%
RF	96.37%	96.07%	92.83%	94.40%
STOME-RF	98.59%	98.58%	98.58%	98.58%

较于 KNN 算法、决策树算法和 RF 算法，准确率分别提高了 3.01%、2.79% 和 2.22%。从表 5 中可以看到，当数据集的维度和规模变得很高时，决策树算法和 KNN 算法的性能下降较大，而 STOME-RF 算法的准确率为 99.40%，相较于 KNN 算法和决策树

算法，准确率分别提高了 42.42% 和 32.13%。表明 STOME-RF 算法对于特征的选择有较好的鲁棒性，具有处理高维度和大规模数据的优势。和 RF 算法相比，准确率提高了 22.72%，表明 STOME-RF 算法能够有效解决数据不平衡导致的分类性能下降的

表5 机器学习算法在数据集2上的测试结果比较  
**Tab. 5 Comparison of test results of machine-learning algorithms on dataset 2**

使用算法	准确率	精准率	召回率	F1分数
KNN	56.98%	79.06%	55.29%	54.57%
决策树	67.27%	84.51%	66.40%	65.85%
RF	76.68%	89.05%	73.54%	73.67%
STOME-RF	99.40%	99.39%	99.39%	99.39%

问题。从表4和表5中还可以观察到,STOME-RF算法的精准率、召回率和F1分数都是最高的,证明了所提STOME-RF算法具有较高的鲁棒性和泛化能力。然而,由于STOME-RF算法需要预先对数据集进行过采样等处理,这增加了计算复杂度,模型训练也需要更多的计算资源和时间。在以后的工作中,将考虑进一步优化数据处理的算法,以降低算法复杂度,提高效率。

**4.3.2 基于SMOTE-RF的自适应OFDM系统容量分析**

图8显示了基于SMOTE-RF算法的自适应OFDM系统在不同信道环境下的平均频谱效率(ASE)。采用新的样本数据进行仿真,从图中可以看出,本文所提SMOTE-RF算法的ASE和理论分析结果基本吻合,证明了所提算法有效,且泛化能力强。

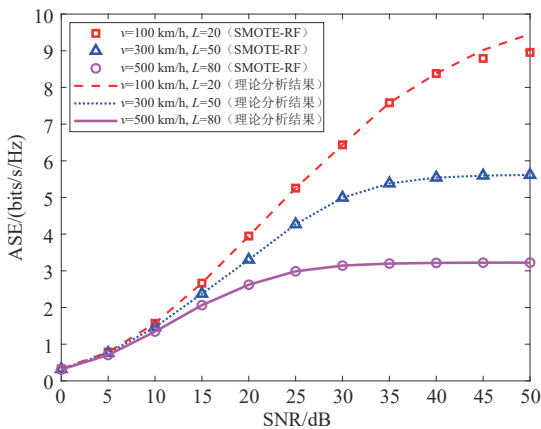
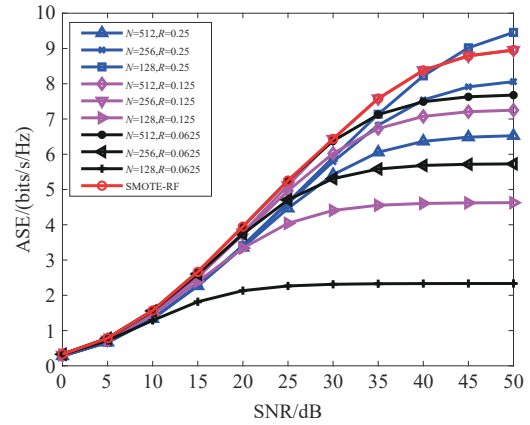
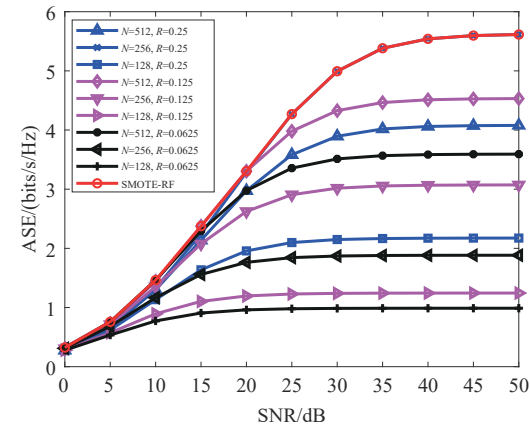


图8 SMOTE-RF在不同信道条件下的ASE  
 Fig. 8 ASE values of SMOTE-RF under different channel conditions

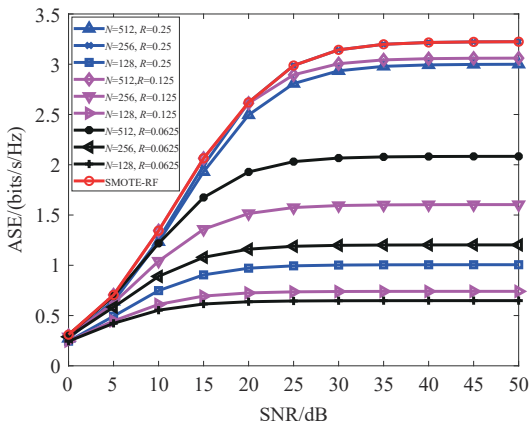
图9比较了基于SMOTE-RF算法的自适应OFDM系统和子载波数 $N$ 、CP比率 $R$ 固定的传统OFDM系统在不同信道环境下的ASE。其中,图9(a)



(a) 移动速度100 km/h, 信道路径数20下的ASE  
 (a) ASE values with a moving speed of 100 km/h and 20 channel paths



(b) 移动速度300 km/h, 信道路径数50下的ASE  
 (b) ASE values with a moving speed of 300 km/h and 50 channel paths



(c) 移动速度500 km/h, 信道路径数80下的ASE  
 (c) ASE values with a moving speed of 500 km/h and 80 channel paths

图9 SMOTE-RF算法与固定参数的OFDM系统在不同信道环境下的ASE比较

Fig. 9 Comparison of ASE values of SMOTE-RF algorithm and fixed-parameter OFDM system under different channel environments

假设用户移动速度为 100 km/h, 信道路径数为 20; 图 9(b) 假设用户移动速度为 300 km/h, 信道路径数为 50; 图 9(c) 假设用户移动速度为 500 km/h, 信道路径数为 80。从图 9 中可以看到, 无论哪一种方案, ASE 均随着 SNR 的增加而增加。但是, 当 SNR 超过一定水平之后, ASE 不会持续增加, 这是因为干扰的存在, 使得通过继续增加 SNR, 难以带来性能的持续提升。同时, 可以明显地观察到, 固定参数的传统 OFDM 系统, 只是在小的 SNR 范围内具有较高的 ASE, 而基于 SMOTE-RF 算法的自适应 OFDM 系统在整个 SNR 的取值区间上, 都具有更大的 ASE。表明所提 SMOTE-RF 算法在不同信道环境下都具有较强的适用性, 能够自适应地调整子载波间隔和 CP 长度以适应不同的信道环境, 提高系统容量。

## 5 结论

本文提出了一种基于随机森林的 OFDM 系统参数自适应分配算法。将通信过程中产生的已有数据作为样本数据, 创建了基于 RF 的多分类模型, 以实现不同信道环境下的 OFDM 系统参数的最优选择。同时利用 SMOTE 算法对少数类别的样本数量进行扩充, 解决了样本不平衡的问题, 进一步提高了 RF 算法的准确率。仿真实验验证了本文所提 SMOTE-RF 算法的有效性和可行性, 表明了所提算法具有较好的准确率和泛化能力。

## 参考文献

- [1] KANSAL L, BERRA S, MOUNIR M, et al. Performance analysis of massive MIMO-OFDM system incorporated with various transforms for image communication in 5G systems[J]. *Electronics*, 2022, 11(4): 621.
- [2] CHEN Shaoping, ZHU Cuitao. ICI and ISI analysis and mitigation for OFDM systems with insufficient cyclic prefix in time-varying channels[J]. *IEEE Transactions on Consumer Electronics*, 2004, 50(1): 78-83.
- [3] MALLICK K, ATTA R, DUTTA B, et al. Cyclic prefix length optimization in OFDM-MMWOF/PAM4-FSO integrated system for future generation smart wireless communication[J]. *Optical and Quantum Electronics*, 2022, 55(1): 64.
- [4] NTOUNI G D, KAPINAS V M, KARAGIANNIDIS G K. On the optimal tone spacing for interference mitigation in OFDM-IM systems[J]. *IEEE Communications Letters*, 2017, 21(5): 1019-1022.
- [5] DAS S S, DE CARVALHO E, PRASAD R. Performance analysis of OFDM systems with adaptive sub carrier bandwidth[J]. *IEEE Transactions on Wireless Communications*, 2008, 7(4): 1117-1122.
- [6] ZHANG Rongxin, MA Xiaoli, WANG Deqing, et al. Adaptive coding and bit-power loading algorithms for underwater acoustic transmissions[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(9): 5798-5811.
- [7] YAZAR A, ARSLAN H. A waveform parameter assignment framework for 6G with the role of machine learning[J]. *IEEE Open Journal of Vehicular Technology*, 2020, 1: 156-172.
- [8] MOHAMMADVALIEI S, SEBGHATI M, ZAREIAN H. Adaptive modulation and coding using deep recurrent neural network[J]. *Telecommunication Systems*, 2022, 81(4): 615-623.
- [9] HANZO L, HAAS H, IMRE S, et al. Wireless myths, realities, and futures: from 3G/4G to optical and quantum wireless[J]. *Proceedings of the IEEE*, 2012, 100 (Special Centennial Issue): 1853-1888.
- [10] 刘潇然. 混合参数集 OFDM 波形设计和智能适应关键技术研究[D]. 长沙: 国防科技大学, 2020.  
LIU Xiaoran. Research of OFDM waveform design and intelligent adaptation in mixed-numerology systems [D]. Changsha: National University of Defense Technology, 2020. (in Chinese)
- [11] ZHANG Zhaoyang, LAI Lifeng. A novel OFDM transmission scheme with length-adaptive Cyclic Prefix[J]. *Journal of Zhejiang University-SCIENCE A*, 2004, 5(11): 1336-1342.
- [12] DONG Zhicheng, FAN Pingzhi, HU R Q, et al. On the spectral efficiency of rate and subcarrier bandwidth adaptive OFDM systems over very fast fading channels[J]. *IEEE Transactions on Vehicular Technology*, 2016, 65(8): 6038-6050.
- [13] MATIN S A, MILSTEIN L B. OFDM system performance, variability and optimality with design imperfections and channel impediments[J]. *IEEE Transactions on Vehicular Technology*, 2021, 70(1): 381-397.
- [14] 罗海鹏, 刘玉军, 刘潇然, 等. 联合循环前缀和调制编码的链路自适应技术研究[J]. *中国电子科学研究院学报*, 2022, 17(7): 657-663.  
LUO Haipeng, LIU Yujun, LIU Xiaoran, et al. Research on link adaptation technology of joint cyclic prefix and MCS[J]. *Journal of China Academy of Electronics and Information Technology*, 2022, 17(7): 657-663. (in Chinese)
- [15] SONI T, ALI A R, GANESAN K, et al. Adaptive numerology—a solution to address the demanding QoS in 5G-V2X[C]//2018 IEEE Wireless Communications and Networking Conference (WCNC). Barcelona, Spain. IEEE, 2018: 1-6.

- [16] BREIMAN L. Random forests [J]. Machine learning, 2001, 45(1): 5-32.
- [17] AGARWAL N, SRIVASTAVA R, SRIVASTAVA P, et al. Multiclass classification of different glass types using random forest classifier[C]//2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS). Madurai, India. IEEE, 2022: 1682-1689.
- [18] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: Synthetic minority over-sampling technique [J]. Journal of Artificial Intelligence Research, 2002, 16: 321-357.
- [19] MAULIDEVI N U, SURENDRO K. SMOTE-LOF for noise identification in imbalanced data classification[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(6): 3413-3423.
- [20] CAMACHO L, DOUZAS G, BACAO F. Geometric SMOTE for regression[J]. Expert Systems with Applications, 2022, 193: 116387.

#### 作者简介



王波 男,1993年生,安徽合肥人。国防科技大学电子科学学院硕士研究生,主要研究方向为智能无线通信与网络。  
E-mail: 1451685716@qq.com



刘潇然 男,1992年生,河南洛阳人。国防科技大学电子科学学院讲师,博士,主要研究方向为无线通信信号处理、多载波波形设计和智能通信技术。  
E-mail: liuxiaoran10@nudt.edu.cn



熊俊 男,1987年生,江西丰城人。国防科技大学电子科学学院副研究员,博士,硕士生导师,主要研究方向为协同通信、物理层安全和网络资源分配等。  
E-mail: xj8765@nudt.edu.cn



辜方林 男,湖南娄底人。国防科技大学电子科学学院副研究员,博士,主要研究方向为可靠无线通信。  
E-mail: gu.fanglin@nudt.edu.cn



张晓赢 女,1980年生,湖南长沙人。国防科技大学电子科学学院副教授,博士,主要研究方向为无线信道建模、宽带无线通信。  
E-mail: zhangxiaoying@nudt.edu.cn

(责任编辑:边熙淳)