

DOI: 10.19783/j.cnki.pspc.180814

# 基于改进随机森林的电力线通信优化算法研究

谢文旺, 孙云莲, 黄雅鑫

(武汉大学电气与自动化学院, 湖北 武汉 430072)

**摘要:** 针对传统信道估计技术会占用频谱资源的缺陷, 提出了一种基于改进随机森林的解映射优化算法。首先针对信源数据固有的不平衡性引入了改进 SMOTE 算法进行预处理, 基于电力线信道特性确定了少数类数据的合成规则, 并以解映射模块的子区间误码率作为评价指标进行性能分析。搭建了宽带电力线通信系统模型, 以实际电表数据作为信源数据, 在 500 m 的 18 径电力线信道模型下进行了仿真测试。实验结果表明, 所提算法可以很好地弥补电表数据固有的不平衡性对随机森林性能的影响, 极大地降低了子区间误码率的波动性。在各种信噪比环境下, 引入改进随机森林算法均可以很好地优化解映射模块性能, 提高宽带电力线通信质量, 降低误码率。

**关键词:** 电力线通信; OFDM; 不平衡数据集; 改进 SMOTE 算法; 随机森林

## Research on power line communication optimization algorithm based on improved random forest

XIE Wenwang, SUN Yunlian, HUANG Yaxin

(School of Electric Engineering and Automation, Wuhan University, Wuhan 430072, China)

**Abstract:** Aiming at the defect that traditional channel estimation technology will occupy spectrum resources, an improved algorithm based on improved random forest is proposed. Firstly, the improved SMOTE algorithm is preprocessed according to the inherent imbalance of the source data. Based on the characteristics of the power line channel, the synthetic rules of the minority data are determined, and the performance of the subinterval error rate of the mapped module is analyzed as the evaluation index. A broadband power line communication system model is built, and the actual meter data is used as the source data to carry out the simulation test under 500 m 18 path power line channel model. The experimental results show that the algorithm can make up for the influence of the inherent imbalance of the meter data on the performance of the random forest, and greatly reduce the fluctuation of the bit error rate. In all kinds of signal to noise environment, the improved random forest algorithm can well optimize the performance of the module, improve the quality of broadband power line communication and reduce the bit error rate.

This work is supported by Science and Technology Project of China Southern Power Grid (No. 035300KK52150007).

**Key words:** power line communication; OFDM; imbalanced dataset; improved SMOTE algorithm; random forest

## 0 引言

近年来, 电力线通信技术以其经济性好、通信效率高、可靠性好等技术优势在用电信息采集、智能化小区等领域获得了广泛的应用<sup>[1-2]</sup>。并且随着能源互联网的发展, 能量双向流动以及网络共享已经成为未来电力行业的一个发展趋势。其中, 电力线通信技术以其无需专门架设通信线路、信号保密性好和安全性高等优势在未来的电力行业发展中也具有很大的发展潜力<sup>[3]</sup>。

目前, 在我国的电力线通信技术的发展中, 宽带电力线通信技术以其良好的抗多径衰落性能、便捷的调制解调方式以及高效的传输速度展现出了比窄带电力线通信技术更加广阔的发展前景。传统的宽带电力线通信系统为了获得良好的通信质量, 普遍引入了信道估计技术。信道估计技术需要借助导频序列, 依据接收端所接收的导频信号估计出信号所经过的信道的各项特性, 从而在接收端正确地解调出信源发送的信号<sup>[4]</sup>。该技术的引入虽然可以提高宽带电力线通信系统的通信质量, 但由于导频序列的插入却会占用宝贵的频谱资源, 大大降低通信的有效性, 在频谱资源日益宝贵的今天具有很大的

基金项目: 南方电网公司科技项目资助 (035300KK52150007)

发展弊端。近年来, 为了提高电力线通信系统性能, 很多学者对信道估计技术进行了深入的研究, 文献[5-7]均是在此基础上所提出的新算法, 但其仍然没有摆脱信道估计技术占用宝贵频谱资源的弊端, 实质上仍是一种以牺牲通信系统的有效性来换取可靠性的一种方式。

因此, 本文提出了一种基于改进 SMOTE 的随机森林算法用于宽带电力线通信系统解映射模块的性能优化。首先通过改进 SMOTE 算法对原始的信源数据进行预处理, 基于电力线信道特性确定负类数据合成规则, 构建出优化数据集从而降低信源数据的不平衡性。然后在 OFDM 的解映射模块引入随机森林算法, 利用随机森林算法良好的分类预测能力优化解映射模块性能, 降低误码率。本文搭建了宽带电力线通信系统的仿真模型, 在 500 m 的 18 径电力线信道模型下进行了仿真测试, 证明了实验结果。

## 1 基于随机森林的电力线通信优化原理

### 1.1 决策树和随机森林原理

决策树(Decision Tree, DT)又称为判定树, 是在分类预测领域广泛应用的一种树结构, 也是构成随机森林的基本单位<sup>[8-9]</sup>。其简单直观且能够实现可视化, 有助于对数据的后续分析和决策, 尤其对于离散数据集的分类预测具有很好的精度。对于宽带电力线通信系统, 其映射与解映射模块的数据集类别由其映射方式决定。就目前所应用的映射方式而言, 无论是 QPSK 还是 16QAM 的映射方式, 其映射的类别均较少且属于典型的离散型样本数据, 因此在其中引入决策树理论上可以在一定程度上提高解映射模块的性能, 降低误码率。

随机森林(Random Forest, RF)则是由一系列决策树组成的集成分类器, 其学习结果由其基本单位决策树的分类结果投票表决决定。并且大量学者通过实验的手段证明了随机森林算法具有很好的预测准确率, 对噪声具有很好的抗性, 鲁棒性好<sup>[10-12]</sup>。由于构建随机森林时采用了 bagging 方法对样本进行随机有放回抽样, 以此消除不同决策树之间的内在联系, 能够有效避免决策树容易陷入过度拟合和局部最优的缺点, 因此引入随机森林算法理论上可以表现出更好的抗噪声和抗多径衰落的能力。

### 1.2 基于随机森林的电力线通信优化模型

对于宽带电力线通信系统, 误码产生的根本原因是: 在信号传输过程中, 由于电力线信道多径衰落以及各种复杂电力线噪声的影响, 使得信号在传输过程中发生衰落和变形。部分信号受到的影响严

重就会发生失真, 从而使得解映射模块接收到的部分信号游离出原始的对应区域而产生误码, 这也是误码产生的直接原因。因此本文在解映射模块引入随机森林算法提高解映射模块的分类预测性能, 利用电力线通信系统的历史数据作为训练样本集进行网络训练, 其具体步骤如下所述。

1) 取宽带电力线通信系统的编码映射和解映射模块的历史数据构成训练样本集  $S$ , 并对其通过 Bagging 方法进行有放回的均匀抽样随机生成若干个新的训练样本集  $S(i)$ , 样本集数量与随机森林中设定的决策树数量相同, 均为  $t$ 。

2) 以上述每一个训练样本集  $S(i)$  独立地构建出一棵完全分裂的决策树, 不同的决策树集合到一起即构成了随机森林模型。对于本文所构建模型, 决策树的生成采用经典的 C4.5 决策树算法, 即以信息增益率作为特征选择的依据。其中, 信息增益率  $GainRatio(S, A)$  是用信息增益  $Gain(S, A)$  和分裂信息量  $SplitInformation(S, A)$  共同定义的, 如式(1)一式(4)所示。

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (1)$$

$$Gain(S, A) = Entropy(S) - \sum_{v \in V(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

$$SplitInformation(S, A) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (3)$$

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (4)$$

式中:  $Entropy(S)$  表示集合  $S$  的熵;  $V(A)$  是属性  $A$  的值域;  $S_v$  是集合  $S$  中在属性  $A$  上值等于  $v$  的子集;  $c$  表示集合  $S$  中的状态个数;  $p_i$  为子集中第  $i$  个属性值的样本数所占的比例。

3) 统计解映射模块实际接收到的信号经过每棵决策树后的分类预测结果, 通过投票表决的方式取投票数最多的那一类即为实际信号的预测类别。

综上所述, 基于随机森林算法的电力线通信系统优化模型的原理框图如图 1 所示。

## 2 系统建模与信源数据的不平衡性

### 2.1 电力线通信系统模型的参数选取

对于宽带电力线通信系统建模过程, 很多学者已经对此进行了深入的研究<sup>[13-16]</sup>, 在此不再赘述。本文建模中, 信源数据选取我国广东省某小区用电

信息采集系统实际采集到的各项电能数据。为了分别测试电力线通信系统对各项电能数据的通信性能，本文随机选取其中的日用电量数据 200 组和瞬

时功率、瞬时电流以及瞬时电压数据各 150 组组成原始的信源数据，各项原始数据如图 2 所示。

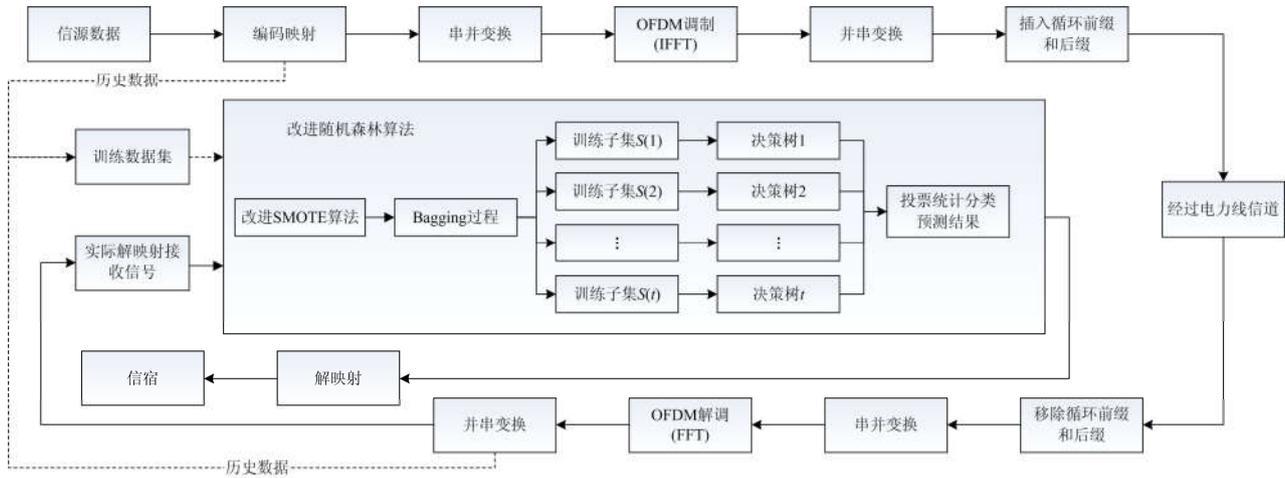


图 1 基于随机森林算法的宽带电力线通信优化原理图

Fig. 1 Optimization principle of broadband power line communication based on random forest algorithm

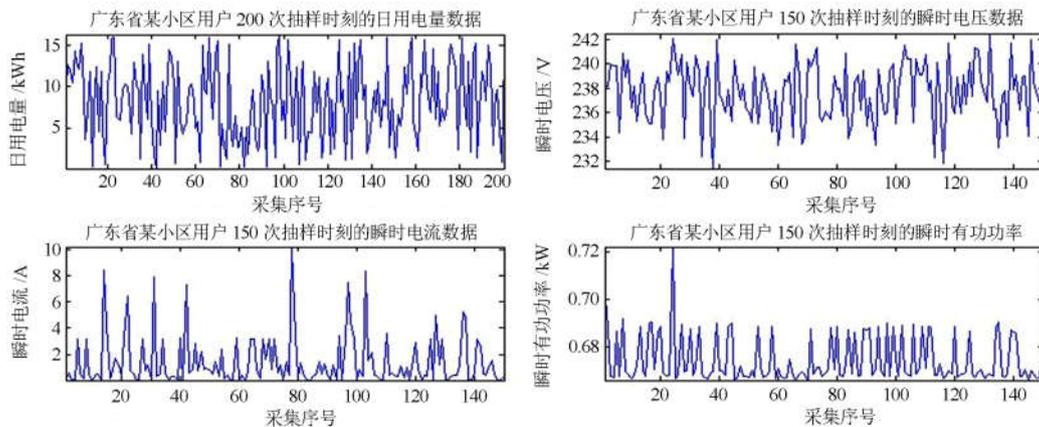


图 2 原始的信源数据

Fig. 2 Original source data

对于 PLC 系统，在数据传输前还需要对原始数据进行相应处理形成待传输数据，本文对数据的处理依据《中华人民共和国电力行业标准(DL/T 645-2007)》中对智能电表与集中器之间的数据通信时的长度规定，具体的参数规定如表 1 所示。

表 1 各项电能数据项的长度规定

Table 1 Length of the electrical energy data items

数据格式	数据长度 (字节)	单位	功能		数据项名称
			读	写	
XXXX.XX	4	kWh	*		总电能
XXX.X	2	V	*		电压
XXX.XXX	3	A	*		电流
XX.XXXX	3	kW	*		瞬时功率

由表 1 可知，不同的电能数据项在传输过程中所占的字节数和精度是不同的。按照规定，电能数据和电压数据分别占用 4 个字节和 2 个字节长度，电流和瞬时功率则均占用 3 个字节长度且具有更高的精度。因此原始的信源数据经过表 1 所示的处理转换为  $200 \times 32 + 150 \times 16 + 150 \times 2 \times 24 = 16\,000$  bit 的二进制数据用于传输。对于电力线信道模型的搭建，本文的模型参数选取文献[17]中利用改进粒子群优化算法对实际 500 m 电力线信道的 18 径参数辨识结果，具体参数如表 2 所示。信道模型搭建后，通过仿真得到其在 1~20 MHz 的频率范围内的衰减特性如图 3 所示。

表 2 500 m 的 18 径电力线信道模型的参数辨识结果

Table 2 Parameter identification results of a 500 m 18 path power line channel model

衰减参数	$k$	$\alpha_0$	$\alpha_1$
	1	0.001 9	$2.28 \times 10^{-9}$
路 径 参 数	$i$	$g_i$	$d_i / m$
	1	0.016	80
	2	0.031	91
	3	0.091	103
	4	-0.042	133
	5	-0.032	142
	6	-0.031	198
	7	0.025	248
	8	-0.025	292
	9	0.061	331
	10	-0.025	370
	11	0.051	447
	12	-0.042	510
	13	0.032	630
	14	-0.043	704
	15	0.031	850
	16	-0.017	950
	17	-0.012	1 150
18	-0.012	1 150	

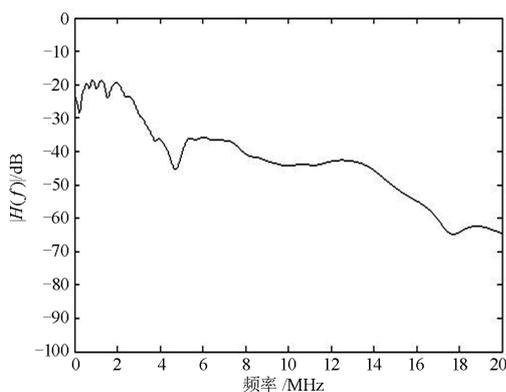


图 3 500 m 的 18 径信道模型的衰落曲线

Fig. 3 Fading curve of the 500 m 18 path channel model

由图 3 可见, 所建电力线信道模型在 1~20 MHz 的频率范围内具有典型的频率选择性衰落特性, 并且通过对比可见, 图 3 中的衰落曲线与文献[17]中的实测 500 m 的电力线信道特性吻合很好, 在此模型上进行测试具有很好的代表性。

### 2.2 数据集固有的不平衡性

将前文所述 OFDM 模型与电力线信道模型结合即为完整的宽带电力线通信系统模型。模型建立后, 在未引入信道估计和随机森林算法时, 调节  $SNR=15$  dB, 解映射模块所接收信号的星座图如图 4 所示。

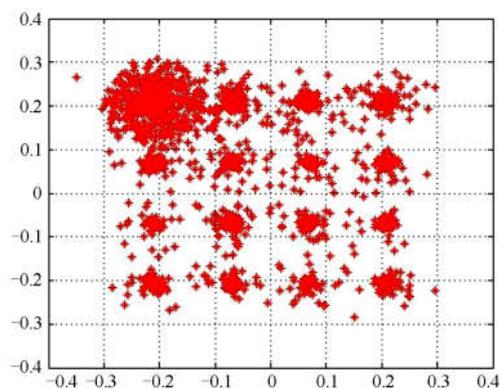


图 4 解映射模块接收信号的星座图

Fig. 4 Constellation of the receiving signal of the demap module

由图 4 可见, 原始信号经过 500 m 的电力线信道后发生了严重的信号衰落, 同时由于噪声的影响使得信号在星座图上发生了弥散现象, 从而使得不同映射区域的信号之间发生交叠, 这将增加解映射模块正确解映射的难度, 提高解映射模块的误码率。同时, 图 4 所示的星座图表现出了典型的不平衡数据特性, 即在 16QAM 解映射的 16 个映射区间(从左到右、从上到下依次将其编号 1 至 16)中, 区间 1(正类)的数据量远远多于其他 15 个映射区间(负类)的数据量, 这是由于表 1 中对各项电量数据通信时的长度规定导致待传输数据中区间 1 中数据大量增多造成的, 这也是电表数据的固有特性。对于不平衡数据集, Weiss G M 在文献[18]中详细分析了其各项特性和会带来的一系列问题, 并证明了不平衡数据是导致随机森林算法性能下降的主要原因。在本文的实验过程中, 若不对待传输的数据集进行预处理, 直接在解映射模块引入随机森林算法的仿真结果如图 5 所示。

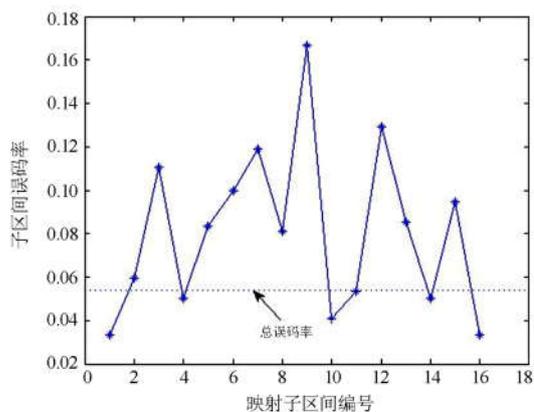


图 5 不同映射区间的数据误码率

Fig. 5 Bit error rate of data in different mapping intervals

由图 5 可见, 在  $SNR=15$  dB 的环境下, 在解映射模块引入随机森林算法后, 可使通信的总体误码率降低到 5% 左右, 通信质量得到了显著提高。但由于电表数据集转换成待传输数据后固有的不平衡性, 导致正类数据的误码率远远低于负类数据, 某些负类区间(例如区间 9)的误码率甚至达到了总体误码率的三倍以上。因此, 直接以总误码率作为通信系统的性能评价指标具有误导性。即使总误码率很低, 但某些负类区间内部的误码率仍然较高, 因此还需引入子区间误码率作为新的评价指标。当考虑各个子区间误码率时, 直接引入随机森林算法后各个子区间误码率的波动性大, 并且由于区间 1 的数据量过大, 会导致随机森林的性能具有很大的偏向性, 从而导致很多负类区间的子区间误码率相较总误码率普遍偏高。当实际传输数据包含大量的某一负类区间数据时, 实际通信系统的通信性能会变得很差, 因此还需引入新的优化算法对其进行改进。

### 3 改进 SMOTE 优化算法

#### 3.1 SMOTE 算法简介

SMOTE(Synthetic Minority Over-sampling Technique)是一种典型的不平衡数据处理算法, 在数据处理领域获得了广泛的应用<sup>[19-20]</sup>。其基本原理是通过人工合成新的负类数据的方式达到对其过采样的目的, 从而实现降低数据集不平衡性的作用。SMOTE 算法人工合成数据的规则如(5)式所示。

$$f_{j(\text{new})} = f_i + \text{rand}(0,1) \times (x_{ij} - f_i) \quad (5)$$

式中:  $f_{j(\text{new})}$  ( $j=1,2,\dots,m$ ) 是新合成的  $m$  个人工合成数据;  $f_i$  ( $i=1,2,\dots,n$ ) 为负类样本,  $n$  表示其总个数;  $x_{ij}$  ( $j=1,2,\dots,m$ ) 为负类样本  $f_i$  的近邻样本。

SMOTE 算法与传统通过单纯复制负类样本的过采样方式相比具有自身的独特性能, 由于人工合成数据不是原样本的简单复制, 而是在样本与其近邻的连线上随机产生数据, 在降低数据集不平衡性的同时, 通过对新增的合成样本的训练和学习, 理论上可以进一步提高随机森林算法的性能和精度<sup>[21-22]</sup>。SMOTE 算法的基本前提是相距较近的负类样本之间的样本仍然是负类, 但对于电力线通信而言, 在信噪比较低时, 不同映射区间边界的数据会发生一定程度交叠, 因此在区间边缘上的数据与其近邻连线上的数据很大概率会落入其他区间, 若直接引入传统的 SMOTE 算法, 则会导致新增的人造样本数据在负类区间边界会产生新的误码, 同时还会造成负类区间边界模糊。因此, 为了在降低数据集不平衡性的同时保证系统的整体性能, 还需要

对传统 SMOTE 算法进行改进。

#### 3.2 改进 SMOTE 算法原理

基于传统 SMOTE 算法在电力线通信领域的缺陷, 本文提出了一种改进的 SMOTE 算法。为了防止人工合成样本在训练学习的过程中降低随机森林的性能, 必须保证对于同一类样本的合成样本仍然属于该类别。同时为了提高随机森林对新样本数据的学习能力, 人造样本也应尽量保持与原始样本之间的差异性。对于电力线通信系统, 不同信噪比下解映射模块所接收到的信号区间交叠程度不同, 信噪比较高时, 区间交叠程度小甚至不发生交叠, 当信噪比降低时, 不同的信号区间之间的交叠程度逐渐增大。为了保证上述样本合成条件, 本文通过引入适应性阈值对样本生成区域进行调节, 改进 SMOTE 算法的数据合成规则如式(6)~式(8)所示。

$$f_{j(\text{new})} = f_{ic} + \text{rand}(\lambda_i, \gamma_i) \times (f_{ic} - f_i) \quad (6)$$

$$f_{ic} = \frac{1}{n} \sum_{i=1}^n f_i \quad (7)$$

$$\gamma_i = \alpha_i \frac{|f_l - f_{ic}|_{\min}}{|f_l - f_{ic}|_{\max}} \quad (8)$$

式中:  $f_{ic}$  表示负类数据的数据中心, 其计算式如式(7)所示;  $\gamma_i$  和  $\lambda_i$  表示适应性阈值的上下限, 用于控制合成数据的生成区域。在本文的建模过程中,  $\lambda_i$  在 0 到  $\gamma_i$  之间进行取值,  $\lambda_i$  与  $\gamma_i$  相差越大, 生成区域越大, 但当  $\lambda_i$  取值过小时, 数据生成区域过于靠近数据中心, 会降低随机森林对合成数据的学习能力。阈值上限  $\gamma_i$  由邻近区间的边缘数据距负类数据中心的距离与负类区域范围的比值决定, 如式(8)所示。式中,  $f_l$  ( $l=1,2,\dots,n_l$ ) 表示与负类区间邻近的区间中的数据,  $n_l$  表示邻近区间中的数据个数,  $\alpha_i$  为调节系数, 一般为保证合成的数据边界与邻近区间之间留有一定的隔离区域, 常取  $\alpha_i < 1$ 。引入上述改进 SMOTE 算法以后, 接收信号的星座图如图 6 所示。

由图 6 可见, 引入改进 SMOTE 算法以后, 新合成的负类数据可以降低数据集固有的不平衡性。合成数据以负类数据的中心为生成中心, 且数据的生成具有一定的方向性。适应性阈值的设置可以灵活控制合成数据的生成空间, 例如由于区间 1 的数据量大、分散性强, 与之相邻的区间 2 和区间 5 的数据生成区域相对较小, 以保证合成数据参与随机森林训练不会带来新的误码, 与之远离的负类区间的数据生成区域则相对较大, 以保证合成数据对随机森林学习能力的提升。在随机森林算法前引入改进 SMOTE 算法对待传输数据进行预处理以后, 解

映射模块接收信号的子区间误码率对比如图 7 所示。

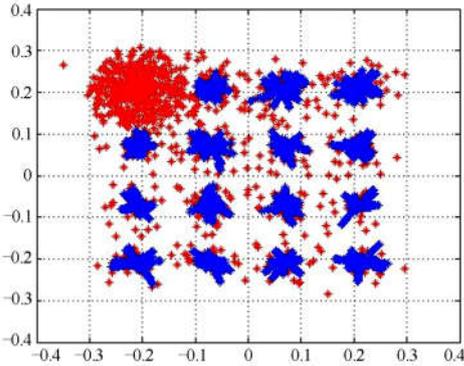


图 6 改进 SMOTE 处理后的星座图

Fig. 6 Constellation after processing of improved SMOTE algorithm

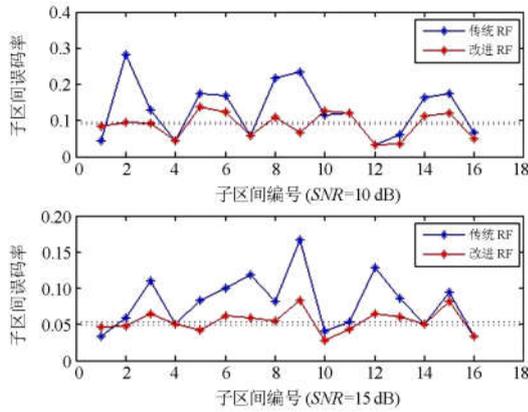


图 7 子区间误码率对比图

Fig. 7 Comparison diagram of bit error rate in subinterval

由图 7 可见, 引入改进随机森林算法以后, 与之前相比, 总误码率仅有略微降低, 这是由于数据总体基数较大, 新合成数据对随机森林算法的性能提升空间较为有限。但引入改进 SMOTE 算法以后, 各负类的子区间误码率均得到了很好的降低, 子区

间误码率与总误码率之间的差异性降到了很低的水平, 很好地解决了不平衡数据集下随机森林性能的偏向性问题, 此时以误码率作为通信系统的性能评价指标具有较好的代表性。接着, 本文在不同的信噪比下对改进随机森林算法进行了性能测试, 仿真结果如图 8 所示。

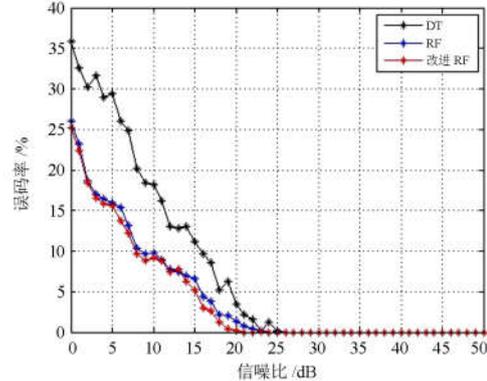


图 8 不同信噪比下的误码率对比图

Fig. 8 Comparison diagram of bit error rate under different signal to noise ratio

由图 8 可见, 引入改进随机森林算法以后, 在不同的信噪比条件下, 通信系统的误码率均得到了显著降低, 与决策树相比, 其误码率曲线下落速度更快, 尤其在低信噪比环境下的性能均更加优越。与传统随机森林算法相比, 改进后的算法可以降低电表数据集固有的不平衡特性, 在显著提高各负类子区间误码率的同时, 总误码率曲线也有一定程度降低, 极大提升了通信系统的整体性能。实验结果表明, 当  $SNR > 10$  dB 时, 引入改进随机森林算法可将电力线通信系统的误码率降低到 10% 以下, 当  $SNR > 20$  dB 时, 可将误码率降低到 1% 以下, 此时在信宿模块接收信号的对比图如图 9 所示, 信宿可以很好地恢复出经过电力线传输的信源信号。

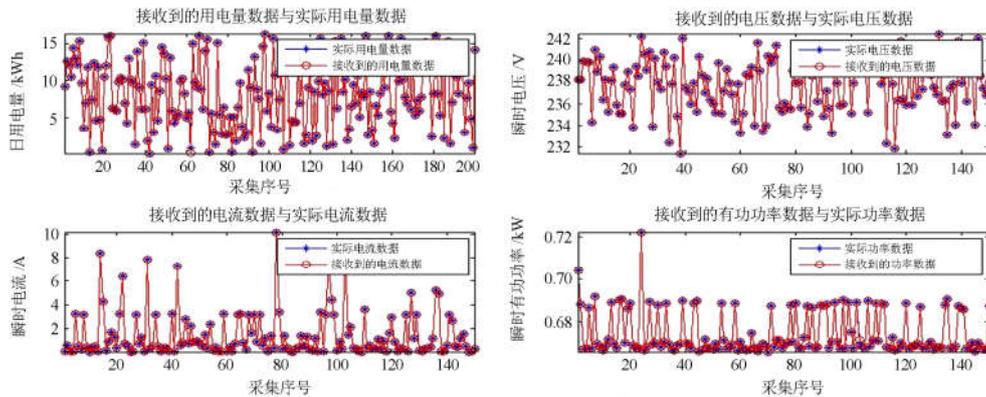


图 9 信宿接收信号的对比图(SNR=20 dB)

Fig. 9 Contrast diagram of the signal received by sink (SNR=20 dB)

## 4 结论

本文提出了一种改进的随机森林算法用于优化电力线通信系统的解映射模块性能。首先通过改进 SMOTE 算法对训练数据集中的历史数据进行预处理,以降低数据集原有的不平衡性。基于电力线信道特性确定负类数据的合成规则,引入适应性阈值调节数据生成区间,有效降低了子区间误码率的波动性,从而消除系统总误码率的误导性。随后利用改进随机森林算法对电力线通信系统解映射模块性能进行优化。实验结果表明,在各种信噪比环境下,改进随机森林算法均能有效提高电力线通信质量,且表现出比决策树更好的噪声抗性和衰落抗性。同时该算法与传统信道估计技术相比不占用宝贵的频谱资源且实现方便,对于实际电表的各项电能数据的测试结果均表现出较好的性能,对电力线通信系统的研究具有很好的借鉴意义。

### 参考文献

- [1] 何玉钧, 刘毅, 周生平. 基于物元可拓模型的电力通信网风险评估[J]. 电力系统保护与控制, 2017, 45(14): 64-69.  
HE Yujun, LIU Yi, ZHOU Shengping. Risk evaluation for electric power communication network based on matter-element extensible model[J]. Power System Protection and Control, 2017, 45(14): 64-69.
- [2] 田洪迅, 王宏刚, 万涛, 等. 基于 BP 神经网络的配电网可靠性关联因素灵敏度计算方法[J]. 电力系统保护与控制, 2017, 45(19): 71-77.  
TIAN Hongxun, WANG Honggang, WAN Tao, et al. A methodology of computing sensitivity of distribution system reliability with respect to driving factors based on a BP neural network[J]. Power System Protection and Control, 2017, 45(19): 71-77.
- [3] 彭显刚, 李壮茂, 邓小康, 等. 智能电网框架下的高级量测体系研究述评[J]. 广东电力, 2017, 30(12): 7-14.  
PENG Xiangang, LI Zhuangmao, DENG Xiaokang, et al. Research on advanced metering infrastructure under smart grid framework[J]. Guangdong Electric Power, 2017, 30(12): 7-14.
- [4] 陈可, 胡晓光. 基于低压电力线正交频分复用系统虚载波改进信道估计算法[J]. 电力自动化设备, 2013, 33(4): 135-141.  
CHEN Ke, HU Xiaoguang. Improved channel estimation algorithm based on virtual carriers of low-voltage power line OFDM system[J]. Electric Power Automation Equipment, 2013, 33(4): 135-141.
- [5] 唐英杰, 董月军, 任宏亮, 等. 基于时频域卡尔曼滤波的 CO-OFDM 系统相位噪声补偿算法[J]. 光学学报, 2017, 37(9): 47-58.  
TANG Yingjie, DONG Yuejun, REN Hongliang, et al. Phase noise compensation algorithm based on Kalman filtering in time and frequency domains for CO-OFDM system[J]. Acta Optica Sinica, 2017, 37(9): 47-58.
- [6] 谷双春, 张润生. 一种基于改进 Lloyd-Max 的信道估计算法[J]. 无线电通信技术, 2017, 43(3): 56-59.  
GU Shuangchun, ZHANG Runsheng. A new channel estimation method based on improved Lloyd-Max[J]. Radio Communications Technology, 2017, 43(3): 56-59.
- [7] 刘剑飞, 石慧敏, 曾祥焯, 等. 基于小波降噪和改进 ISFA 的 CO-OFDM 信道估计方法[J]. 中国激光, 2015, 42(12): 145-151.  
LIU Jianfei, SHI Huimin, ZENG Xiangye, et al. Channel estimation method based on wavelet and improved ISFA for CO-OFDM system[J]. Chinese Journal of Lasers, 2015, 42(12): 145-151.
- [8] 王守相, 周凯, 苏运. 基于随机森林算法的台区合理线损率估计方法[J]. 电力自动化设备, 2017, 37(11): 39-45.  
WANG Shouxiang, ZHOU Kai, SU Yun. Line loss rate estimation method of transformer district based on random forest algorithm[J]. Electric Power Automation Equipment, 2017, 37(11): 39-45.
- [9] 黄晗, 孙堃, 刘达. 基于随机森林的电力系统小时负荷预测研究[J]. 智慧电力, 2018, 46(5): 8-14.  
HUANG Han, SUN Kun, LIU Da. Hourly load forecasting of power system based on random forest[J]. Smart Power, 2018, 46(5): 8-14.
- [10] 曹正凤. 随机森林算法优化研究[D]. 北京: 首都经济贸易大学, 2014.  
CAO Zhengfeng. Study on optimization of random forests algorithm[D]. Beijing: Capital University of Economics and Business, 2014.
- [11] 李军锋, 王钦若, 李敏. 结合深度学习和随机森林的电力设备图像识别[J]. 高电压技术, 2017, 43(11): 3705-3711.  
LI Junfeng, WANG Qinruo, LI Min. Electric equipment image recognition based on deep learning and random forest[J]. High Voltage Engineering, 2017, 43(11): 3705-3711.

- [12] 刘达, 孙堃, 黄晗. 基于 EEMD 和随机森林的月度负荷预测[J]. 智慧电力, 2018, 46(6): 12-18.  
LIU Da, SUN Kun, HUANG Han. Monthly load forecasting based on EEMD and random forest[J]. Smart Power, 2018, 46(6): 12-18.
- [13] 索超男, 张慧, 赵雄文. 小波基在低压电力线信道有色背景噪声建模中的应用研究[J]. 电力系统保护与控制, 2017, 45(4): 121-125.  
SUO Chaonan, ZHANG Hui, ZHAO Xiongwen. Research on the application of wavelet basis functions on modeling of colored background noise for low-voltage power line channels[J]. Power System Protection and Control, 2017, 45(4): 121-125.
- [14] 李松浓, 胡晓锐, 郑可, 等. 低压电力线载波通信信道衰减特性测量与分析[J]. 电力系统保护与控制, 2018, 46(4): 99-106.  
LI Songnong, HU Xiaorui, ZHENG Ke, et al. Measure and research on attenuation characteristics of voltage power line communication channel[J]. Power System Protection and Control, 2018, 46(4): 99-106.
- [15] 曹旺斌, 尹成群, 谢志远, 等. 多输入多输出宽带电力线载波通信信道模型研究[J]. 中国电机工程学报, 2017, 37(4): 1136-1142.  
CAO Wangbin, YIN Chengqun, XIE Zhiyuan, et al. Research on broadband MIMO power line communications model[J]. Proceedings of the CSEE, 2017, 37(4): 1136-1142.
- [16] 易琳, 秦晓科, 王刚, 等. 基于改进蚁群算法的电力巡检路径规划[J]. 广东电力, 2018, 31(3): 115-119.  
YI Lin, QIN Xiaoke, WANG Gang, et al. Path planning for electric power patrol based on improved ant colony algorithms[J]. Guangdong Electric Power, 2018, 31(3): 115-119.
- [17] 张旭辉, 张礼勇, 梁宵. 基于改进粒子群优化算法的电力线通信多径传输模型参数辨识[J]. 电网技术, 2009, 33(1): 75-79.  
ZHANG Xuhui, ZHANG Liyong, LIANG Xiao. Parameter identification of multi-path transmission model for power line communication based on improved particle swarm optimization[J]. Power System Technology, 2009, 33(1): 75-79.
- [18] WEISS G M. Mining with rarity: a unifying framework[J]. SIGKDD Explorations, 2004, 6(1): 7-19.
- [19] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [20] 黄建明, 李晓明, 瞿合祚, 等. 考虑小波奇异信息与不平衡数据集的输电线路故障识别方法[J]. 中国电机工程学报, 2017, 37(11): 3099-3107, 3365.  
HUANG Jianming, LI Xiaoming, QU Hezuo, et al. Method for fault type identification of transmission line considering wavelet singular information and unbalanced dataset[J]. Proceedings of the CSEE, 2017, 37(11): 3099-3107, 3365.
- [21] QUINLAN J R. Induction on decision tree[J]. Machine Learning, 1986, 1(1): 81-106.
- [22] 王宇恒. 推荐系统中随机森林算法的优化与应用[D]. 杭州: 浙江大学, 2016.  
WANG Yuheng. Improvement and application of random forest algorithm in recommender systems[D]. Hangzhou: Zhejiang University, 2016.

收稿日期: 2018-07-05; 修回日期: 2018-10-13

作者简介:

谢文旺(1993—), 男, 硕士研究生, 研究方向为电力信息技术、电力线通信技术; E-mail: xiewenwang\_sz@163.com

孙云莲(1962—), 女, 教授, 博士生导师, 研究方向为电力信息技术、智能电网。E-mail: ylsun@whu.edu.cn

(编辑 周金梅)