

11-6-2023

Research on Fisher-SVM feature selection based load identification

Kaining LUAN

State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210028, China

Shihai YANG

State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210028, China

Yixuan HUANG

State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210028, China

Kaijie FANG

State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210028, China

Hanmiao CHENG

State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210028, China

See next page for additional authors

Follow this and additional works at: <https://jepst.researchcommons.org/journal>

Recommended Citation

LUAN, Kaining; YANG, Shihai; HUANG, Yixuan; FANG, Kaijie; CHENG, Hanmiao; and HUANG, Shi (2023) "Research on Fisher-SVM feature selection based load identification," *Journal of Electric Power Science and Technology*. Vol. 38: Iss. 4, Article 25.

DOI: 10.19781/j.issn.1673-9140.2023.04.025

Available at: <https://jepst.researchcommons.org/journal/vol38/iss4/25>

This Scientific Research is brought to you for free and open access by Journal of Electric Power Science and Technology. It has been accepted for inclusion in Journal of Electric Power Science and Technology by an authorized editor of Journal of Electric Power Science and Technology. For more information, please contact dlxb04@163.com.

Research on Fisher-SVM feature selection based load identification

Authors

Kaining LUAN, Shihai YANG, Yixuan HUANG, Kaijie FANG, Hanmiao CHENG, and Shi HUANG

基于 Fisher-SVM 特征选择的负荷辨识研究

栾开宁¹, 杨世海¹, 黄艺璇¹, 方凯杰¹, 程含渺¹, 黄 时²

(1. 国网江苏省电力有限公司, 江苏 南京 210028; 2. 江苏智臻能源科技有限公司, 江苏 南京 211111)

摘要: 针对当前非侵入式负荷辨识中不同设备特征选择上主观性、盲目性的问题, 提出基于 Fisher-SVM 特征选择的非侵入式负荷辨识算法。首先, 基于高频采样终端提取入户侧电流、电压原始数据, 使用傅里叶变换对原始信号分解得有功、无功及谐波时间序列; 其次, 将负荷波形分为 4 个阶段并计算得到负荷波形暂稳态特征; 然后, 通过 Fisher-SVM 算法在不同分类器中对特征进行选择, 得到最优分类特征子集, 并利用 Sigmoid 函数对结果进行概率校准; 最后, 根据贝叶斯理论对各分类器进行融合从而实现对不同负荷的辨识。以 3 类台区 831 户实际用户进行算法测试。结果表明, 该算法能够有效利用不同电器负荷印记的独特性, 克服特征选择上的盲目性, 提高负荷辨识能力。

关键词: 非侵入式; 负荷辨识; Fisher; 支持向量机; 特征提取

DOI: 10.19781/j.issn.1673-9140.2023.04.025 中图分类号: TM933 文章编号: 1673-9140(2023)04-0230-10

Research on Fisher-SVM feature selection based load identification

LUAN Kaining¹, YANG Shihai¹, HUANG Yixuan¹, FANG Kaijie¹,
CHENG Hanmiao¹, HUANG Shi²

(1. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210028, China; 2. Jiangsu Intever Energy Technology Co., Ltd., Nanjing 211111, China)

Abstract: Aiming at the subjectivity and blindness of different device feature selection in current non-intrusive load identification, a non-intrusive load identification algorithm based on Fisher-SVM feature selection is proposed. Firstly, the original data of household-side current and voltage are extracted based on the high frequency sampling device. Fourier transform is used to decompose the original signal into active power, reactive power and harmonic time series. Secondly, the load waveform is divided into four stages and the transient characteristics of the load waveform are calculated. Then, by utilizing the Fisher-SVM algorithm for feature selection among different classifiers, the optimal subset of classification features is obtained. Additionally, the results are calibrated using the Sigmoid function for probability calibration. Finally, different classifiers are integrated based on Bayesian theory to achieve identification of different loads. The algorithm is tested on a dataset consisting of 831 actual users from three different distribution areas. The results show that the algorithm effectively exploits the uniqueness of different electrical load imprints, overcomes the blindness in feature selection, and increases the load identification ability.

Keywords: non-intrusive; load monitoring; Fisher; SVM; feature extraction

收稿日期: 2022-03-26; 修回日期: 2022-09-05

基金项目: 国家重点研发计划“科技助力经济 2020”重点专项(SQ2020YF F0426410)

通信作者: 黄 时(1994—), 男, 硕士, 工程师, 主要从事非侵入负荷辨识、低压台区等方向的算法研究; E-mail: 15251892108@163.com

近年来,随着社会智能用电、清洁能源的不断发展,需求侧的精细化管理对电能的高效利用和调度优化起到了重要作用^[1-2]。非侵入式负荷辨识(non-intrusive load monitoring, NILM)是精细化管理的有效技术之一,通过采集入户用电信号,实时辨识家庭电器的用能情况,为电网的预测优化提供了重要依据。从用户角度出发,为用户提供用能分析管理、用能优化方案以及设备故障检测等服务,有效提升用户用能效率,明确用户电器设备用能账单,从而加强用户自身节能意识^[3]。

早期负荷辨识技术模式为侵入式,即在各类电器设备前安装传感器从而实时监控设备运行状态。此技术虽然可以做到对每种负荷的精确监控,但不可避免会被监测系统内的各种用电设备、传感器硬件成本所限制。与之相对的 NILM 方案,只需在电表处安装相应终端设备,通过负荷分解来获取各个电器的运行波形,此方案成本低,安装方便,易被客户接受,受到了大多数学者和工业部门的青睐。

国内外学者在负荷辨识算法方面做了大量的工作,该算法主要分为数据采集、数据预处理、特征提取以及负荷辨识 4 个步骤,其中大多数学者在特征提取与负荷辨识两方面进行了创新性研究。在特征提取方面,文献[4-5]分析了各种电器开启阶段的暂态特征,建立家用负荷暂态特征标准化模板,以开启特征作为负荷辨识匹配的标准;文献[6]考虑提取稳态特征值,利用动态时间弯曲(dynamic time warping, DTW)算法计算模板与识别负荷的距离从而进行负荷辨识;文献[7]基于 DB 小波变化提取小波特征作为训练的输入,从而创建各电器特异性特征。对于负荷辨识研究方面,文献[8]基于有功、无功特征建立了长短期记忆(long short-term memory, LSTM)网络深度学习模型;文献[9]通过将用电设备的相似特征转换为离散模糊数矩阵,进而实现负荷的辨识。上述文献的研究方式均为寻找特定电器的特异性特征进行模式匹配,人为对不同分类器构造特征量,该方式对特定类别电器(电热、电机类)辨识时拥有良好的性能,而对同类设备由于特征近似,效果往往难以满足实际商业环境的需求。

针对负荷辨识中的特征选择问题,本文提出一种基于 Fisher-SVM 特征选择的负荷辨识算法。首先,从设备的高频数据信号中分解出功率与谐波信号序列;然后,将设备运行时段分为开启、过渡、运行和关断阶段,根据不同阶段特点提取暂态稳态特征,再基于 Fisher-SVM 算法进行特征选择,生成每个分类器最佳特征子集;最后,基于贝叶斯理论得出最优辨识结果。

1 非侵入负荷辨识算法原理

1.1 非侵入负荷辨识流程

NILM 在居民入户处采集用户用电负荷信号,将总负荷信号分离成多个用电设备独立运行的负荷信号,从而辨识出各用户用电负荷类型与运行状态^[10],负荷辨识流程如图 1 所示。

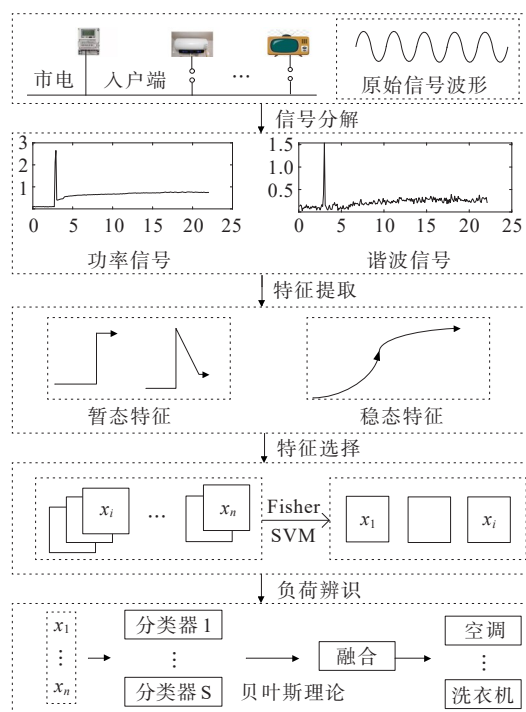


图 1 负荷辨识流程

Figure 1 Flow of load identification

1) 数据采集。负荷辨识终端数据采集模块于用户入户侧前采集高频电压、电流数据。

2) 信号分解。基于原始电压电流信号和快速傅氏变换(fast fourier transformation, FFT),将时序电压电流信号转变为时序功率与谐波有效值信号,并根据此信号分离出不同的设备子信号。

3) 特征提取。根据设备子信号的不同阶段,分别计算负荷波形的各类暂态与稳态特征。

4) 特征选择。针对不同类别电器的分类器,基于 Fisher-SVM 算法选择最优特征子集,提升辨识效果。

5) 负荷辨识。基于贝叶斯理论将不同分类器融合形成最优辨识结果。

1.2 原始信号分解

非侵入式终端设备从入户侧采集用户的高频电压、电流时域信号数据,设备采样频率为 6.4 K,每个周期电压、电流的采样点数 $N_s=128$,为了能更清晰地显示负荷特征,将 5 个周期作为一个有效值点进行输出,假定 FFT 后的电流 k 阶频域信号 $I(k)=a(k)+jb(k)$,将其转换为常见的电压、电流与电流谐波的有效值计算公式^[11-12]:

$$\begin{cases} u^{(0)}=|U(0)|/5N_s \\ i^{(0)}=|I(0)|/5N_s \\ i^{(k)}=|\sqrt{2} \cdot I(k)|/5N_s \end{cases} \quad (1)$$

式中, $u^{(0)}$ 、 $i^{(0)}$ 分别为电压、电流的基波有效值; $i^{(k)}$ 为电流 k 次谐波的有效值。由此可求得平均有功、无功功率分别为

$$\begin{cases} p = u^{(0)}i^{(0)} + \sum_{k=1}^{N_i} u^{(k)}i^{(k)} \cos(\varphi_u^k - \varphi_i^k) \\ q = \sum_{k=1}^{N_i} u^{(k)}i^{(k)} \sin(\varphi_u^k - \varphi_i^k) \end{cases} \quad (2)$$

其中,

$$\begin{cases} \varphi_u^{(k)} = \arg(U(k)) \\ \varphi_i^{(k)} = \arg(I(k)) \end{cases} \quad (3)$$

式中, $\varphi_u^{(k)}$ 、 $\varphi_i^{(k)}$ 分别为 k 次电压、电流谐波相位角。

本文以二、三次谐波分别作为偶、奇次谐波的代表,即有功功率、无功功率、二次谐波、三次谐波有效值序列分别为 P 、 Q 、 I_2 、 I_3 。对于多信号的叠加问题,考虑高频信号下短时间(秒级)窗口内有且仅有一个设备开启,在此窗口内即可分离出不同信号叠加的子信号,并计算其暂态特征以及短时稳态特征,基于此种方式实现不同设备信号的分离,获得不同设备的子信号^[6]。

1.3 特征提取

各类家用电器设备由于物理结构与运行的差

异存在其特有波形信号,分为开启、过渡、稳定运行和关断阶段^[13]。定频空调全运行周期波形如图 2 所示, $t_0 \sim t_1$ 为开启阶段,由于电机的启动电流为额定电流的 3~5 倍,因此开启阶段存在冲击特征; $t_1 \sim t_2$ 为过渡阶段,设备状态缓慢变化并进入稳态阶段; $t_2 \sim t_3$ 为稳态运行阶段,此时功率、谐波等特征量均不会改变; $t_3 \sim t_4$ 为关断阶段。

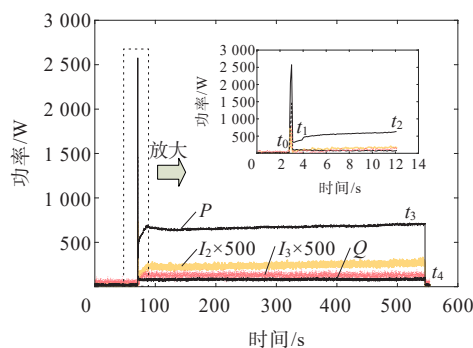


图 2 定频空调全周期运行波形

Figure 2 Full-cycle operational waveform of a fixed-frequency air conditioner

电器在整个运行阶段(开启、过渡、稳定运行和关断阶段)都蕴含着一定的统计规律(负荷特征)^[14],将这些负荷特征分为稳态和暂态特征。暂态特征一般发生在开启、关断阶段,即电器的内部结构与控制差异导致的功率或谐波的突变;稳态特征一般发生在过渡、稳定运行阶段,即随着电器控制策略的变化功率和谐波发生的平稳改变。通常将不同电器的可靠负荷特征作为负荷印记(load signature, LS),它是实现负荷辨识的基础。不同设备的总负荷印记如表 1 所示;常用电器波形如图 3 所示。

电流的总谐波失真(total harmonic distortion, THD)系数为

$$I_{\text{THD,St}} = \sqrt{\sum_{k=2}^{N_k} \left(\frac{i^{(k)}}{i^{(0)}} \right)^2} \quad (4)$$

式中, N_k 为最大非失真谐波数。

不同电器对于不同的负荷特征敏感度不同,拥有不同的负荷印记。在某种设备辨识过程中,如何在特征选择中逐步迭代消除不重要的特征或减小其在最后辨识中的权重,使得在特定空间区域内性能更好是负荷辨识中的一大难题。

表 1 暂态、稳态特征总结

Table 1 Summary of transient and steady state characteristics

特征状态	序号	符号	特征说明
稳态	1	$P_{St}/Q_{St}/I_{2,St}/I_{3,St}$	稳态运行阶段:有功功率/无功功率/二次谐波/三次谐波 ^[15]
	2	$I_{THD,St}$	电流的总谐波失真系数
	3	$P_{Var,St}/Q_{Var,St}/I_{2,Var,St}/I_{3,Var,St}$	稳态运行阶段:有功功率/无功功率/二次谐波/三次谐波方差 ^[16-17]
	4	T_{St}	稳态运行时长
	5	$P_{max,Ts}/P_{min,Ts}/P_{mean,Ts}$ $Q_{max,Ts}/Q_{min,Ts}/Q_{mean,Ts}$	过渡阶段:有功功率最大值、最小值、平均值
	6	$I_{2,max,Ts}/I_{2,min,Ts}/I_{2,mean,Ts}$ $I_{3,max,Ts}/I_{3,min,Ts}/I_{3,mean,Ts}$	过渡阶段:无功功率/二次谐波/三次谐波的最大值、最小值、平均值 ^[17]
	7	$P_{Var,Ts}/Q_{Var,Ts}/I_{2,Var,Ts}/I_{3,Var,Ts}$	过渡阶段:有功功率/无功功率/二次谐波/三次谐波方差 ^[17]
	8	T_{Ts}	过渡阶段时长
暂态	9	$P_{On}/Q_{On}/I_{2,On}/I_{3,On}$	电器开启:有功功率/无功功率/二次谐波/三次谐波 ^[17]
	10	$P_{Off}/Q_{Off}/I_{2,Off}/I_{3,Off}$	电器关断:有功功率/无功功率/二次谐波/三次谐波 ^[17]
	11	$P_{Im}/Q_{Im}/I_{2,Im}/I_{3,Im}$	电器开启阶段:有功功率/无功功率/二次谐波/三次谐波冲击
	12	T_{Im}	电器冲击时长

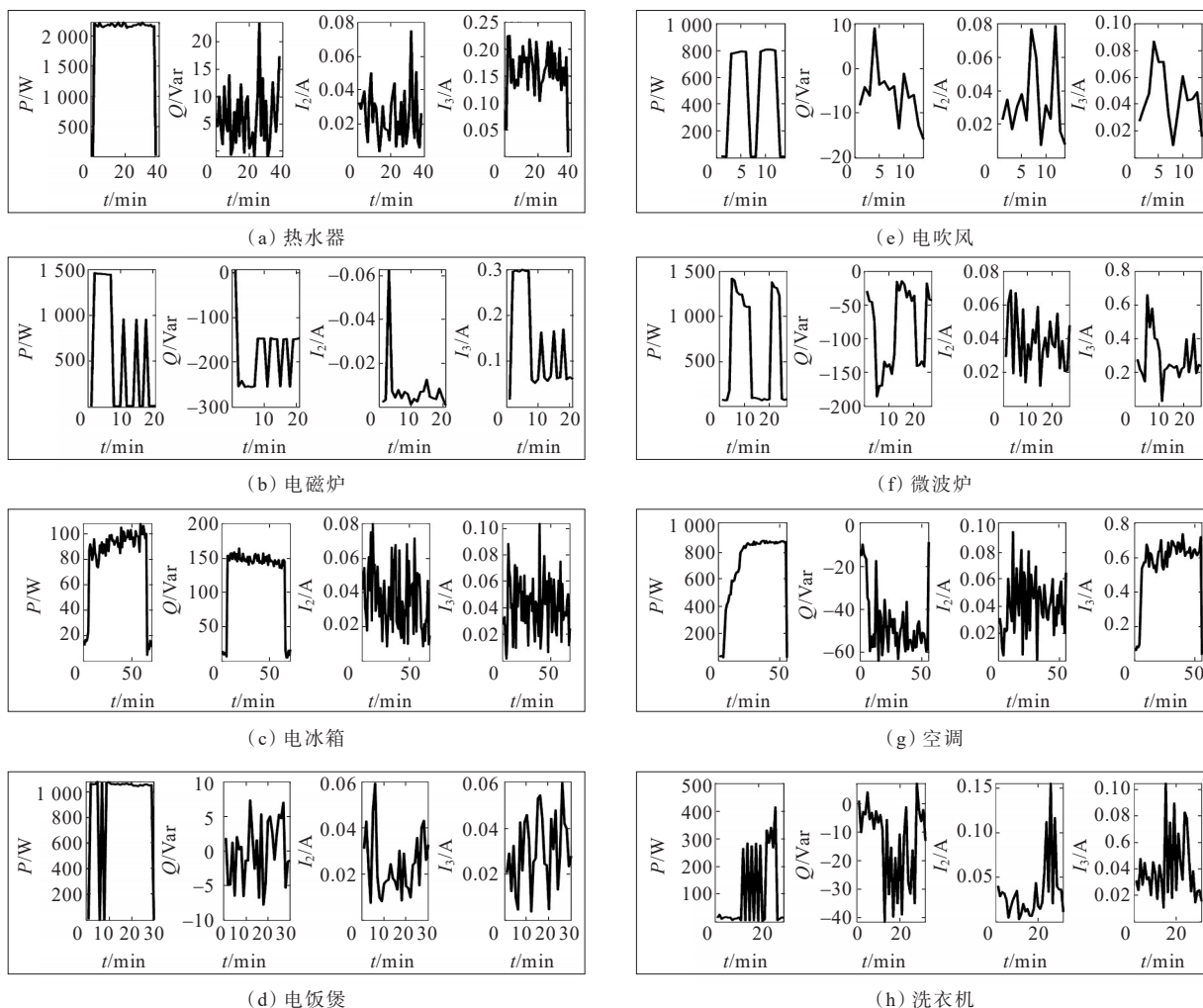


图 3 常用的电器波形

Figure 3 Commonly used electrical waveforms

2 基于 Fisher-SVM 的特征选择负荷辨识

2.1 基于 Fisher-SVM 的特征选择算法

高维特征虽然蕴含着足量的信息,但是往往存在着大量的冗余,影响辨识效果。例如,洗衣机运行时电机不断旋转,功率与谐波因此不停发生变化,对于它的辨识而言,电器开启时的有功功率、无功功率、谐波等暂态特征不仅无法起到正向辨识作用,反而会给算力以及辨识结果带来负面影响,因此在特征中有效负荷印记必不可少。

Fisher线性判别是目前常采用的特征选择方法,在模式识别领域中被广泛应用,可以有效评估某个特征在分类器中的有效性^[18]。基本原理:通过选取 Fisher线性判别函数计算得到最佳鉴别矢量,并将样本数据映射至最佳鉴别矢量上,从而通过 Fisher比率评估类间离散度 S_b 与类内离散度 S_w 的比值。计算公式为

$$S_b = (\bar{m}_1 - \bar{m})^2 + (\bar{m}_2 - \bar{m})^2 \quad (5)$$

$$S_w = \frac{1}{n_1} \sum_{x \in X_1} (x - \bar{m}_1)^2 + \frac{1}{n_2} \sum_{x \in X_2} (x - \bar{m}_2)^2 \quad (6)$$

$$F_{r,k} = \frac{S_{b,k}}{S_{w,k}} \quad (7)$$

式(5)~(7)中, X_1 为正训练样本集, $X_1 = \{x_1, x_2, \dots, x_{n_1}\}$, 共计 n_1 个样本; X_2 为负训练样本集, $X_2 = \{x_1, x_2, \dots, x_{n_2}\}$, 共计 n_2 个样本; \bar{m}_1 为正样本均值; \bar{m} 所有样本均值; \bar{m}_2 为负样本均值; $F_{r,k}$ 为第 k 个特征的类间离散度 $S_{b,k}$ 与类内离散度 $S_{w,k}$ 的比值。

特征选择是从网络数据集中筛选出对分类器分类性能影响最大的特征子集^[19]。支持向量机 (support vector machine, SVM) 是由结构风险最小化准则发展而来,学习训练过程中其将分类问题转换成二次规划问题。一个超平面 $H: \mathbf{w} \cdot \mathbf{x} + b = 0$ 使得 X_1 与 X_2 两类数据分开, \mathbf{w} 表示最优平面的法向量, b 表示分类阈值。其优化问题可转变为

$$\begin{cases} \min f(\mathbf{w}, \boldsymbol{\varepsilon}) = \frac{1}{2} \|\mathbf{w}\|^2 + \xi \sum_{i=1}^{n_1+n_2} \varepsilon_i \\ \text{s.t. } y_i(\mathbf{w} \cdot \mathbf{f}(x_i) + b) \geq 1 - \varepsilon_i \\ \varepsilon_i \geq 0 \end{cases} \quad (8)$$

式中, ε_i 为惩罚因子,因训练时无法保证每个正样

本,故在实际中均为正样本; ξ 为惩罚参数,用于惩罚错误分类。

以洗衣机辨识为例, Fisher-SVM 的特征选择算法流程如下。

1) 输入正样本为洗衣机的集合 X_1 、其余电器的集合 X_2 。

①判断正、负样本数量 n_1 、 n_2 是否均衡,若相差2倍以上,则基于 Smote 过采样算法进行样本均衡^[20]。Smote 算法流程:对于少数类中每一个样本 x ,以欧氏距离为标准计算其到少数类样本集中所有样本的距离,得到其 k 近邻;对于每一个随机选出的近邻 $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ 分别与原样本按照如下公式构建新的样本,即

$$x_{\text{new}} = x + \text{rand}(0, 1) \cdot (x^{(i)} - x) \quad (9)$$

②训练集中计算特征集每维特征的 Fisher 比值 $F_{r,k}$ 。根据计算的 $F_{r,k}$ 比值,对特征集 F 进行降序排序,删除排名前 50% 或 $F_{r,k}$ 小于 0.4 的特征子集。

③高关联度特征类群中只保留 $F_{r,k}$ 最大的特征。

④剩余特征中将排名第 1 的特征删除特征库 F ,并基于 SVM 算法进行训练,得到分类正确率。

⑤判断前次分类正确率,若正确率变大则再次重复④,并直至变小为止。

2) 输出最终特征子集 F' 。

2.2 负荷辨识算法

由上述输出可知,算法经过特征选择后生成了不同电器最佳训练模型,每个模型的目的为辨识其是否为某种电器。对于负荷辨识而言,特征序列已知而电器类型是未知的,即负荷辨识问题转换为已知特征向量 f_i 和 S 个不同电器分类器,其属于哪一类电器的问题。

贝叶斯理论的特征融合与分类器算法相辅相成,是经典模式识别算法,即用计算的方法根据样本的特征将样本划分到一定的类别^[21-22]。模型空间 Ω 包含 S 种模式类,即 $\Omega = \{\omega_1, \omega_2, \dots, \omega_S\}$, 根据最小错误率的贝叶斯决策理论,若将样本 x 分为第 j 类,则该类就是在已知样本 x 下后验概率最大的模式类,即

$$\text{if } F(\omega_j) = \max_{k=1,2,\dots,S} P(\omega_k|x), \forall k \leq S \quad (10)$$

$$x \rightarrow \omega_j$$

式中, $P(\omega_k|x)$ 为第 k 类的后验概率。

对于式(10)的决策过程,在负荷辨识中,将其看作分类器的输出结果,即属于第 j 类电器,可得到基于贝叶斯理论的分类器融合算法。已知有 S 个分类器,用 r_i 作为第 i 个分类器输出结果的向量度量,例如 r_i 为洗衣机的二分类器 $\{0,1\}$,分类器检测某样本得 $\{\text{样本为非洗衣机的概率}, \text{样本为洗衣机的概率}\}$,假设此处样本为 y ,同理有

$$\text{if } F(\omega_j) = \max_{k=1,2,\dots,S} P(\omega_k|r_1, r_2, \dots, r_M), \forall k \leq S \quad (11)$$

$$y \rightarrow \omega_j$$

其中, $P(\omega_k|r_1, r_2, \dots, r_M)$ 为 S 个分类器输出结果下第 k 类的后验概率。由于 SVM 分类输出结果仅为 0 或 1,因此需要进行概率校准,进一步获得相应标签的概率。Platt Scaling 将分类模型的输出结果通过 Sigmoid 函数转化为概率。假定分类模型对样本 x_i 的输出结果为 $f(x_i)$,使用 Sigmoid 函数转换后的概率估计如下:

$$P(y = 1|f(x_i)) = 1/(1 + e^{Af(x_i)+B}) \quad (12)$$

可以通过对数极大似然函数估计 A, B 两参数:

$$L_{\log} = - \sum_{i=1}^n y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (13)$$

式中, L_{\log} 为对数损失函数; $p_i = 1/(1 + e^{Af(x_i)+B})$, 表示概率校准后的概率值; y_i 为样本实际类别值。

2.3 基于 Fisher-SVM 特征选择的负荷辨识算法流程

基于 Fisher-SVM 特征选择的负荷辨识算法流程如图 4 所示。

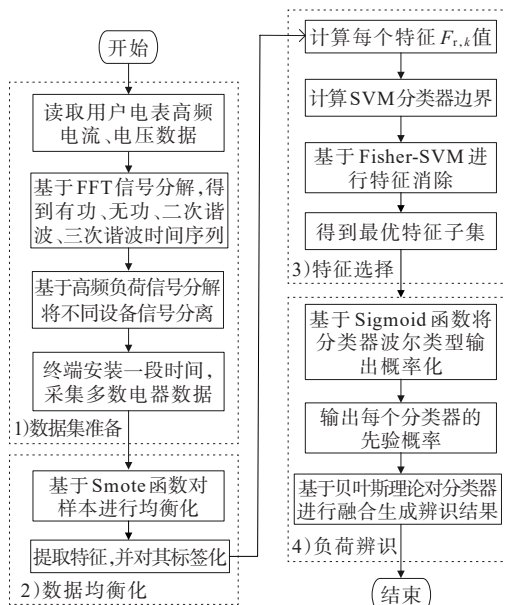


图 4 算法流程

Figure 4 Overall algorithm workflow

1) 数据集准备。现场数据采集并基于原始信号实现设备子信号分离。

2) 数据均衡化。将分离的子信号根据不同类别的数量进行均衡化处理。

3) 不同分类器的特征选择。根据上述数据集,基于 Fisher-SVM 算法对每个分类器进行特征选择,得到最优特征子集。

4) 负荷辨识。基于 Sigmoid 函数将分类器的结果进行概率校准,从而得到先验概率,并基于贝叶斯理论构造多分类器的融合,以整体最小错误率为目标得到最终辨识结果。

3 算例分析

3.1 数据集描述

分析数据集分别来源于南京某高、中、低档 3 个台区,其中现场安装如图 5 所示。3 个类型台区分别存在 51、220、560 户家庭,采集用户正常用电数据,并根据数据分解,人工提取 12 种基本常用电器有功功率、无功功率以及谐波等数据作为样本,最终得到每类电器不少于 1 000 的样本量,共计 27 个稳态特征、13 个暂态特征(表 1)。

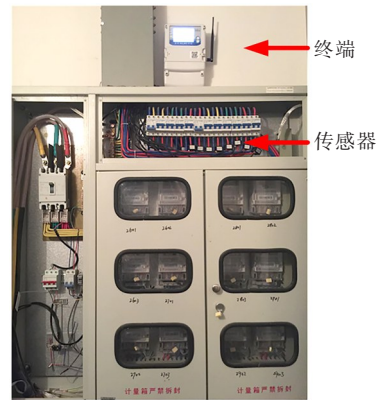


图 5 现场设备安装示意

Figure 5 On-site equipment installation

以某用户为例,其某时段原始数据波形(有效值)如图 6 所示,可以看出,该用户从凌晨至下午 3 点依次使用了定频空调、微波炉和变频空调,基于此波形分解用户真实场景下每个设备的运行波形,并提取其特征作为数据样本集。

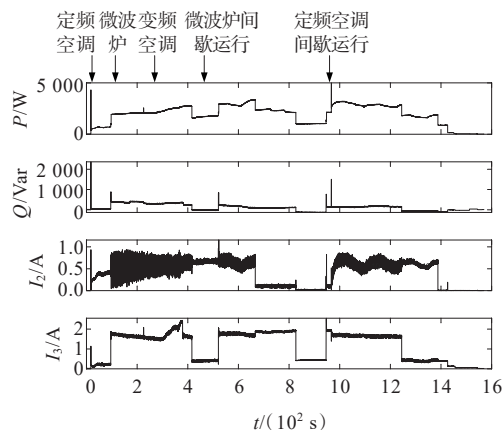


图6 基于终端采集的某用户某时段用电原始数据波形

Figure 6 Original waveform of electricity consumption data for a specific user during a certain period, collected at the terminal

3.2 数据均衡化

由于本文的数据集来源于真实场景下用户的实际用电情况,不同家用电器的使用频率存在较大的差异,例如空调开启过程中通常随着当前室内温度的变化而导致频繁启停事件,因此,事件发生的频率较高,而常用的家用设备如微波炉其开启频率通常固定,与空调的样本数存在较大的差异,如表2所示。

表2 Smote函数样本数量对比

Table 2 Smote function sample number comparison

电器	原样本数量	Smote函数后样本数量	电器	原样本数量	Smote函数后样本数量
热水器	157	1 099	电磁炉	178	1 068
空调	324	972	微波炉	193	965
电冰箱	146	1 022	电采暖	58	1 044
洗衣机	318	954	吹风机	188	940
电饭煲	448	896	电视机	553	1 106
电烤箱	512	1 024			

不平衡数据训练正确率通常不具备实际价值,数据集中的多数类主导了整体辨识率,如图7所示,其SVM决策分类器无限接近微波炉样本,显然不是最优决策边界。因此,本文基于Smote函数对样本集进行样本均衡化处理,为作图便利,仍以空调作为正样本,微波炉作为负样本,二维特征分别选择开启时有功功率突变 P_{on} 和无功功率突变 Q_{on} ,如图8所示,中间决策边界得到校验,效果得到明显改

善,如仍以原决策边界判定,微波炉辨识错误率高达24.3%。

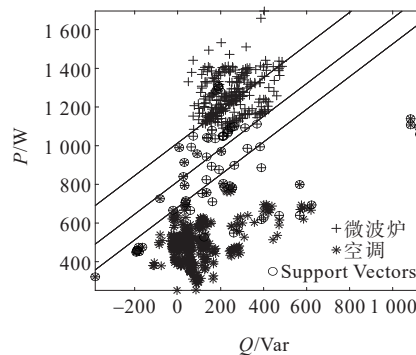


图7 数据不均衡下SVM聚类结果

Figure 7 SVM clustering result under imbalanced data

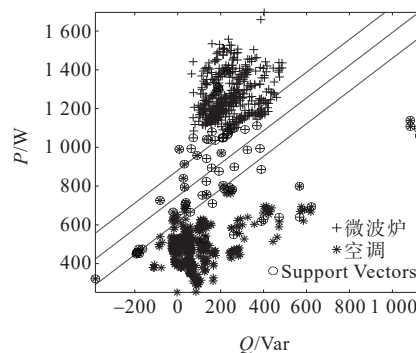


图8 基于Smote数据均衡化聚类

Figure 8 Clustering based on Smote data balancing

所有设备均衡化后样本量见表2,经过样本均衡化后部分少量电器的样本与其余样本数据达到了均衡。为证明数据均衡化的优势,本文引入 $F_{1-score}$ 、 G_{mean} 来评价不均衡分类效果,其中 $F_{1-score}$ 为精确值与召回率的组合,只有精确值与召回率均在很大的情况下, $F_{1-score}$ 才是一个较好的结果;在少数和多数类分类精度同时很高时, G_{mean} 为一个较好的结果。

$$F_{1-score} = 2 \times \frac{P_{re} R_{ec}}{P_{re} + R_{ec}} \quad (14)$$

$$G_{mean} = \sqrt{\frac{T_P}{T_P F_N} \cdot \frac{T_N}{T_N F_P}} \quad (15)$$

式(14)、(15)中, P_{re} 为精确值,即正样本预测正确数/(正样本预测正确数与负样本预测为正确数之和); R_{ec} 为召回率,即正样本预测正确数/(正样本预测正确数与负样本预测正确数之和); T_P 为正样本辨识正确的数量; T_N 为正样本辨识错误的数量; F_N 为负样本辨识正确的数量; F_P 为负样本辨识错误的数量。

通过 $F_{1-score}$ 、 G_{mean} 评估改进后的差异,保证其结果具有对比性,因此,各个设备印记特征均选择 P_{On} 、 Q_{On} ,结果如图 9 所示,各个块状数据表示经过 Smote 函数前、后 $F_{1-score}$ 与 G_{mean} 参数的变化,其中 A 区域为 $F_{1-score}$ 参数,B 区域为 G_{mean} 参数。结果显示,在样本数据不平衡度较高且 2 个设备特征较为接近时,无论是 $F_{1-score}$ 或是 G_{mean} 指标均有明显提升。一些设备之间参数有所下降,经研究发现,由一些异常数据参与 Smote 函数导致,整体辨识率得到了较大的提升。

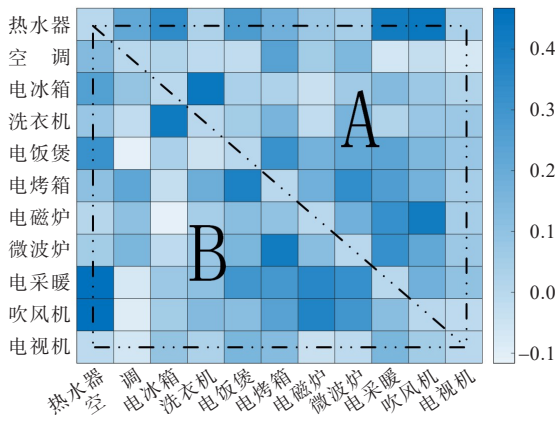


图 9 使用 Smote 均衡化前、后指标变化热力图
Figure 9 Heatmap of indicator changes before and after applying the SMOTE balancing

3.3 特征选择

数据集均衡化后再进行特征选择,本文基于 Fisher-SVM 算法进行特征消除,从而得到最优特征子空间。

数据集中以空调的负荷辨识为例,基于特征模型生成空调与微波炉的分类器。首先,计算各个特征 F_r ,将其从小到大进行排列,并删除前 50% 或 F_r 小于 0.4 的特征(特征差异不明显),如图 10 所示,共计 40 个特征,经过 F_r 筛选后最终得到 20 个特征;然后,对剩余特征进行相关度计算,对于高关联度特征仅保留 F_r 最大的特征,在此算例中,有功冲击、三次谐波冲击以及冲击时间存在高关联度,只保留有功冲击;最后,依次删除 F_r 最小的特征,直到计算 SVM 聚类正确率不再上升为止,最终得到最优特征子空间及 SVM 分类器最优参数模型。案例中最终剩下 6 个最优特征子集: P_{lm} 、 $Q_{var, Ts}$ 、 T_{St} 、 $Q_{var, St}$ 、 $I_{THD, St}$ 、 $I_{3, St}$ 。

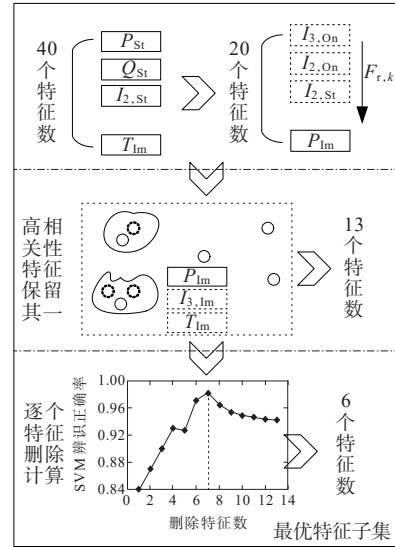


图 10 特征消除过程
Figure 10 Feature elimination process

3.4 辨识效果

分别从每个电器设备以及每户的层面分析算法性能,并与常用的负荷辨识方式^[5-6]进行比较。文献[5](方法 2)基于用户功率序列对用户的暂态特征进行辨识,其暂态过程主要为有功的开启和关断,即对应本文的 P_{On} 、 P_{Off} ,并基于最大隶属度原则进行负荷辨识;文献[6](方法 3)以稳态电流为信号分离的基础,根据原始信号进行建模,最终基于 DTW 算法进行负荷辨识,考虑本文用户数量多,将高频信号稳态特征等效为本文特征中的稳态特征。文献[5]、[6]分别建立了设备类型特征库,为了与本文算法(方法 1)进行对比,不再使用交叉验证的方法测试正确率,3 种方法分别以一半用户负荷建立设备类型特征库或相应的分类器模型,以另外一半用户负荷作为测试样本。将用户测试样本辨识结果进行加权平均,3 种方法中 11 个不同设备的测试结果如图 11 所示。

对于电饭煲、电磁炉、微波炉等电热设备,采用 3 种方法相比较,方法 2 辨识效果与方法 1、3 相比明显偏低,溯源原始结果显示:在以 120 个用户作为总样本库时,各个电热有功暂态特征相互耦合严重,电热水器、电饭煲、电烤箱等设备的有功过于接近,导致对其很难分辨,平均辨识率低于 50%,方法 1、3 平均辨识率超过 80%,均有不错的辨识率;而对于空调、冰箱等设备,其具备独特的暂态特征,稳态特征不明显,因此,方法 1、2 的辨识正确率超过 80%,

而对于方法3而言,由于其文献中数据源未考虑到空调或电机类设备,其稳态特征不明显,且空调会随着时间变化功率发生波动,因此对其辨识效果不佳;对于微功率设备电视机而言,3种方案的辨识效果均低于60%,其原因在于由于电视机属于低功率设备,其暂态特征不明显,而稳态特征中电视机会随着画面的不同进行无功功率的波动,稳态特征很难定量。具体辨识结果如表3所示。

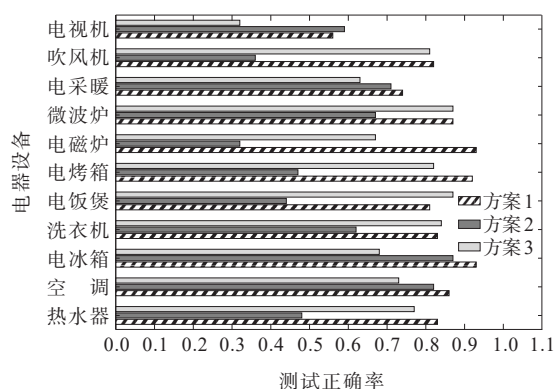


图 11 各个设备辨识结果对比

Figure 11 Comparison of identification results of each device

表 3 3类电器整体辨识结果对比

Table 3 Comparison of the overall identification results of three types of electrical appliances

设备类型	不同方法辨识率/%		
	1	2	3
电热	86.44	48.87	81.23
电机	91.17	83.34	68.42
微功率	51.44	52.17	31.23
合计	88.43	67.73	76.42

就整体效果而言,相较于基于人工经验而进行盲目的特征选择,本文对于整体的负荷辨识成功率有了显著的提升,相较于方法2,用户平均辨识正确率提升了20.7%;相较于方法3,在其平均辨识率已经超过75%的情况下,用户平均辨识正确率提升了11.01%,说明本文算法具有较高的准确率。

4 结语

针对负荷辨识中不同负荷类型间波形相似性与差异性导致现场辨识准确度低的问题,本文提出

了一种基于 Fisher-SVM 特征选择的负荷辨识算法。基于同类型设备波形的相似性提取暂态与稳态特征,基于不同类别设备波形的差异性结合其特征所具备的特异性构建辨识分类器,从而解决了特征冗余导致辨识率低的问题。

1) 对于真实场景下不同类别数据样本差异超过2倍的场景,在基于 Smote 函数算法实现负荷辨识数据样本的均衡化后,平均辨识正确率提升12%以上。

2) 对于真实场景下聚类特征选择中的主观性、盲目性问题,本文基于 Fisher-SVM 特征选择的负荷辨识算法相较于常规人工特征选择方法,综合正确率提升10%以上。

但是,电器种类和运行情况是无穷的,本文对于设备信号分解阶段依旧认识不足,即多电器均处于过渡阶段,采样得到的信号时刻变化,本文仅仅针对小窗口负荷变化分离了数据样本,未对整体分离样本进行优化。

参考文献:

- [1] 刘凯诚,钟鸣,曾平良,等.考虑分布式可再生电源和储能智能配电网可靠性评估综述[J].电测与仪表,2021,58(7):1-11.
LIU Kaicheng,ZHONG Ming,ZENG Pingliang,et al.Review on reliability assessment of smart distribution networks considering distributed renewable energy and energy storage[J]. Electrical Measurement & Instrumentation, 2021,58(7):1-11.
- [2] 李振华,陶渊,赵爽,等.智能配电网状态估计方法研究现状分析[J].电力科学与技术学报,2019,34(1):115-122.
LI Zhenhua,TAO Yuan,ZHAO Shuang,et al. Research situation analysis of state estimation in smart distribution networks[J]. Journal of Electric Power Science and Technology,2019,34(1):115-122.
- [3] 武昕,严萌,郭一凡,等.基于结构化特征图谱的组合支持向量机非侵入式负荷辨识[J].电力系统自动化,2022,46(12):210-219.
WU Xin,YAN Meng,GUO Yifan,et al.Non-intrusive load identification by combined support vector machine based on structured characteristic spectrum[J]. Automation of

- Electric Power Systems,2022,46(12):210-219.
- [4] 李如意,张鹏,刘永光,等.基于随机森林的非侵入式家庭负荷辨识方法[J].电测与仪表,2021,58(4):9-16.
LI Ruyi,ZHANG Peng,LIU Yongguang,et al.Non-intrusive household load identification method based on random forest[J].Electrical Measurement & Instrumentation,2021,58(4):9-16.
- [5] 雷怡琴,孙兆龙,叶志浩,等.电力系统负荷非侵入式监测方法研究[J].电工技术学报,2021,36(11):2288-2297.
LEI Yiqin,SUN Zhaolong,YE Zhihao,et al. Research on non-intrusive monitoring method of power system load[J]. Transactions of China Electrotechnical Society, 2021, 36 (11):2288-2297.
- [6] 祁兵,董超,武昕,等.基于DTW算法与稳态电流波形的非侵入式负荷辨识方法[J].电力系统自动化,2018,42(3):70-76.
QI Bing,DONG Chao.WU Ting,et al.Non-intrusive load identification method based on DTW algorithm and steady-state current waveform[J]. Automation of Electric Power Systems,2018,42(3):70-76.
- [7] 许春蕾,刘文颖,郭虎,等.基于小波特征提取的可调节负荷辨识方法[J].可再生能源,2019,37(6):845-851.
XU Chunlei,LIU Wenying,GUO Hu,et al.Adjustable load identification method based on wavelet feature extraction [J].Renewable Energy Resources,2019,37(6):845-851.
- [8] 刘恒勇,刘永礼,邓世聪,等.一种基于LSTM模型的电力负荷辨识方法[J].电测与仪表,2019,56(23):62-69.
LIU Hengyong,LIU Yongli,DENG Shicong,et al.A power load identification method based on LSTM model[J]. Electrical Measurement & Instrumentation, 2019, 56(23): 62-69.
- [9] 孙毅,崔灿,张璐,等.智能用电非侵入式负荷监测系统研究[J].电力科学与技术学报,2019,34(2):155-160.
SUN Yi,CUI Can,ZHANG Lu,et al. Research on nonintrusive load monitoring system for global energy internet[J]. Journal of Electric Power Science and Technology,2019,34(2):155-160.
- [10] 武昕,焦点,高宇辰.基于非侵入式用电数据分解的自适应特征库构建与负荷辨识[J].电力系统自动化,2020,44(4):101-109.
WU Ting,JIAO Dian,GAO Yuchen. Construction of adaptive feature library and load identification based on decomposition of non-intrusive power consumption data[J]. Automation of Electric Power Systems, 2020, 44(4): 101-109.
- [11] 裘星,尹红,张之涵,等.基于V-I轨迹与高次谐波特征的非侵入式负荷识别方法[J].电力工程技术,2021,40(6):34-42.
QIU Xing,YIN Hong,ZHANG Zhihan,et al.Non-intrusive load identification method based on V-I trajectory and high-order harmonic feature[J].Electric Power Engineering Technology,2021,40(6):34-42.
- [12] 吴静,赵伟.一种用于分析电网谐波的多谱线插值算法[J].中国电机工程学报,2006,26(8):55-60.
WU Jing,ZHAO Wei.An algorithm of MICA for analyzing harmonics in power system[J]. Proceedings of the CSEE, 2006,26(8):55-60.
- [13] 余登武,刘敏,汪元芹.基于GRNN与注意力机制模型的非侵入式家用负荷分解[J].智慧电力,2021,49(3):74-79.
YU Dengwu,LIU Min,WANG Yuanqin. Non-invasive household load decomposition based on GRNN and attention mechanism model[J]. Smart Power, 2021, 49(3): 74-79.
- [14] 祁兵,韩璐.基于负荷空间划分的非侵入式辨识算法[J].电测与仪表,2018,55(16):19-25.
QI Bing,HAN Lu.A non-intrusive identification algorithm based on partition of the load space[J]. Electrical Measurement & Instrumentation,2018,55(16):19-25.
- [15] 杨秀,李安,孙改平,等.基于改进GMM-CNN-GRU混合的非侵入式负荷监测方法研究[J].电力系统保护与控制,2022,50(14):65-75.
YANG Xiu,LI An,SUN Gaiping,et al.Non-invasive load monitoring based on an improved GMM-CNN-GRU combination[J].Power System Protection and Control,2022,50(14):65-75.
- [16] 苏欣,袁葆,张瑞,等.基于非侵入式负荷识别的交流电压幅值测量[J].电网与清洁能源,2022,38(12):101-106.
SU Xin,YUAN Bao,ZHANG Rui,et al.AC voltage amplitude measurement based on non-intrusive load identification[J].Power System and Clean Energy, 2022, 38 (12): 101-106.
- [17] SRINIVASAN D, NG W S, LIEW A C. Neural-network-based signature recognition for harmonic source identification[J]. IEEE Transactions on Power Delivery, 2006,21(1):398-405.
- [18] 谢小英,牛益国,于惠慧,等.基于相空间重构与原子分解的复杂电压暂降特征参数辨识[J].电力科学与技术学报,2020,35(5):103-110.

- 2007,14(3):643-648.
- [14] 葛国伟,廖敏夫,黄金强,等.双断口真空断路器配合特性仿真与试验[J].电工技术学报,2016,31(22):57-65.
GE Guowei, LIAO Minfu, HUANG Jinqiang, et al. Simulation and experimental of matching characteristics of vacuum circuit breakers with double-break[J]. Transactions of China Electrotechnical Society, 2016, 31(22):57-65.
- [15] 李博,包涌泉,彭振东,等.基于改进型直流真空断路器弧后暂态仿真及介质恢复特性分析[J].电工技术学报,2021,36(8):1752-1760.
LI Bo, BAO Yongquan, PENG Zhendong, et al. Post-arc transient simulation and dielectric recovery analysis based on improved DC vacuum circuit breaker[J]. Transactions of China Electrotechnical Society, 2021, 36(8):1752-1760.
- [16] 文化宾,宋永端,邹积岩,等.新型126 kV高压真空断路器的设计及开断能力试验研究[J].中国电机工程学报,2011,31(34):198-204.
WEN Huabin, SONG Yongduan, ZOU Jiyan, et al. Test on novel design and breaking capacity for 126 kV high voltage vacuum circuit breaker[J]. Proceedings of the the CSEE, 2011, 31(34):198-204.
- [17] 吴燕,余勇祥,韩筛根,等.真空开关容性电流开断技术[J].高压电器,2017,53(3):26-43.
WU Yan, YU Yongxiang, HAN Shaigen, et al. Technique of capacitive switching for vacuum switches[J]. High Voltage Apparatus, 2017, 53(3):26-43.
- [18] 邹丹旦,潘捍宇,钱沐杨.针板缺陷位形下局部放电击穿特性实验教学研究[J].实验技术与管理,2021,38(7):191-196.
ZOU Dandan, PAN Hanyu, QIAN Muyang. Experimental teaching research on breakdown characteristics of partial discharge under defect configuration of needle-plate[J]. Experimental Technology and Management, 2021, 38(7):191-196.
- [19] 刘素贞,饶诺歆,张闯,等.基于LabVIEW的电磁超声无损检测系统的设计[J].电工技术学报,2018,33(10):2274-2281.
LIU Suzhen, RAO Nuoxin, ZHANG Chuang, et al. Design of electromagnetic ultrasonic nondestructive testing system based on LabVIEW[J]. Transactions of China Electrotechnical Society, 2018, 33(10):2274-2281.
- [20] 修金鹏. Arduino与LabVIEW互动设计[M].北京:清华大学出版社,2014,41-47.
XIU Jinpeng. Arduino and LabVIEW interactive design [M]. Beijing: Tsinghua University Press, 2014, 41-47.
- [21] 陈树学,刘莹. LabVIEW宝典[M].北京:电子工业出版社,2011,568-575.
CHEN Shuxue, LIU Xuan. LabVIEW handbook[M]. Beijing: Electronic Industry Press, 2011, 568-575.
- [22] 于丽霞,金源,刘吉,等.强电磁环境下光纤测速系统设计[J].兵器装备工程学报,2020,41(8):234-238.
YU Lixia, JIN Yuan, LIU Ji, et al. Design of optical fiber velocity measurement system under strong electromagnetic environment[J]. Journal of Ordnance Equipment Engineering, 2020, 41(8):234-238.
- [23] 杨芳,唐小亮,尹文阔,等.基于多参量分布式光纤传感的架空输电线路风振监测分析[J].电力系统保护与控制,2022,50(11):169-177.
YANG Fang, TANG Xiaoliang, YIN Wenkuo, et al. Analysis of transmission line wind vibration monitoring based on a multi parameter distributed optical fiber sensor[J]. Power System Protection and Control, 2022, 50(11):169-177.

(上接第239页 = Continued from page 239)

- XIE Xiaoying, NIU Yiguo, YU Huihui, et al. Voltage sag characteristic parameter identification method based on phase space reconstruction and atomic decomposition[J]. Journal of Electric Power Science and Technology, 2020, 35(5):103-110.
- [19] TAPIA E, BULACIO P, ANGELONE L. Sparse and stable gene selection with consensus SVM-RFE[J]. Pattern Recognition Letters, 2012, 33(2):164-172.
- [20] CHENG K, ZHANG C, YU H, et al. Grouped SMOTE with noise filtering mechanism for classifying imbalanced data [J]. IEEE Access, 2019, 7:170668-170681.
- [21] 何小飞,童晓阳,孙明蔚.基于贝叶斯网络和D-S证据理论的分布式电网故障诊断[J].电力系统自动化,2011,35(10):42-47.
HE Xiaofei, TONG Xiaoyang, SUN Mingwei. Distributed power system fault diagnosis based on bayesian network and dempster-shafer evidence theory[J]. Automation of Electric Power System, 2011, 35(10):42-47.
- [22] 杨东升,孔亮,胡博,等.基于多特征序列融合的负荷辨识方法[J].电力系统自动化,2017,41(22):66-73.
YANG Dongsheng, KONG Liang, HU Bo, et al. Load identification method based on multi-feature sequence fusion[J]. Automation of Electric Power System, 2017, 41(22):66-73.