

基于决策树和聚类算法的智能电表 误差估计与故障检测

史鹏博, 李蕊, 李铭凯, 赵成, 朱锦山

(国网北京电力科学研究院, 北京 100162)

摘要: 针对智能电表故障检测问题,研究了低压电能系统模型,提出了一种电能表误差估计方法。该方法采用决策树对异常数据进行过滤,并对不同损失水平的数据进行分类,然后对数据进行聚类,得到不同使用行为的数据集。此外,建立了仪表数据矩阵,采用递推算法求解矩阵方程,对仪表误差进行估计。最终通过算例验证了所提方法的有效性。

关键词: 计量学; 智能电表; 电表误差估计; 故障检测; 异常数据过滤; 损失水平

中图分类号: TB971

文献标识码: A

文章编号: 1000-1158(2022)08-1089-06

Error Estimation and Fault Detection of Smart Meter Based on Decision Tree and Clustering Algorithm

SHI Peng-bo, LI Rui, LI Ming-kai, ZHAO Cheng, ZHU Jin-shan

(State Grid Beijing Electric Power Research Institute, Beijing 100162, China)

Abstract: Aiming at the problem of intelligent watt hour meter fault detection, the low-voltage power system model is studied, and an error estimation method of watt hour meter is proposed. The method uses the decision tree to filter the abnormal data, classifies the data with different loss levels, and then clusters the data to obtain the data sets with different usage behaviors. In addition, the instrument data matrix is established, and the recursive algorithm is used to solve the matrix equation to estimate the instrument error. Finally, the effectiveness of the proposed method is verified by an example.

Key words: metrology; smart meter; meter error estimation; fault detection; abnormal data filtering; loss level

1 引言

智能电网融合了电网基础设施和网络系统,具有典型的网络物理系统(cyber physical system, CPS)^[1],其旨在将电力技术应用于能源系统,以实现双向电力流^[2]。智能电表作为整个配电网的传感器,可以测量用电量并将记录的数据发送给能源管理中心^[3],对应数据的使用有助于通过CPS为智能电网规划和运行提供及时的决策^[4]。对于智能电表数量众多、应用场合复杂的情况,如何提高智能电表故障检测能力成为电网企业关注的焦点^[5-6]。

智能电网先进计量基础设施(advanced metering infrastructure, AMI)为人们提供了丰富的测量、通信、计算和存储资源,这些信息引起了人们的关

注^[7]。用数据分析检测故障智能电表的方法很多。文献[8]提出了一种综合考虑层次结构和仪表数据的智能电表生成模型,但当面对大量的智能电表数据时,该方法效果急剧下降;文献[9]提出了一种面向概念漂移的异常检测能力定义方法,该方法采用长短期记忆网络来描述和预测消费者的行为,然而检测精度并不高;文献[10]描述了一种考虑电流和功率特征之间关系的比较方法来检测异常设备,考虑较为单一,无法处理较为复杂的情况;文献[11]介绍了一种数据挖掘技术来检测异常数据和能量窃取,该方法结合最大信息系数理论和聚类技术,发现任意形状的异常用户,对历史数据的依赖较为严重;文献[12]中提出了一种结合Bollinger带和部分可观测马尔可夫决策过程的技术来检测异常数据和能量窃取;并采用概率自适应模型提高效率,然而其操

收稿日期: 2020-08-14; 修回日期: 2022-06-23

作过于复杂,计算量大;文献[13]提出一种通过比较前后能耗来检测欺诈和异常使用的方法;文献[14]提出了一种基于随机矩阵理论的异常使用和能量盗窃检测器,研究了能量消耗与系统异常状态的关系,但考虑的因素较为单一。

在相关的工作中,虽然已经有了许多成熟的方法^[15~17],并且一些研究工作已经得到了实际应用,但在实际应用中仍然存在一些局限性。利用机器学习和强化学习相关算法检测异常电能表的思想虽然简单易行,但其结果取决于具体的系统结构和数据集。通过检查仪表误差来检测故障仪表是一种有效的方法,在线仪表误差估计是一种能够实现大规模检测的新思路。误差估计精度仅限于数据和系统信息,所采集的系统特征和测量数据越多,估计精度越高;然而,一些系统信息存在如能量损失、系统信息丢失等现象,在实际系统中很难获得,缺少这些信息将影响检测精度和方法性能。

本文主要提出了大范围误差检测模型。采用能耗数据构造线性方程,由方程的解可导出仪表误差估计量;在其基础上,结合分类和递归理论的仪表误差计算方法,对能量损失率水平相近的数据,采用决策树和聚类方法对不同的配置数据进行分类,构造线性方程,用递推最小二乘法求解,得到误差估计量;最终通过算例分析证明所提模型的有效性。

2 方案综述

2.1 方案概述

电能数据由智能电表记录并送数据中心存储,具体拓扑图如图1所示。在低压电能系统中,电能表采用拓扑结构,即在一些低精度单相电能表前安装高精度三相电能表。高精度电能表是电网的主电能表,低精度电能表是电能网的子表。其中,子表记录一个住宅用户的电压、电流和能耗值,而主表记录一个区域内所有用户的总值。仪表功能异常、内部元件退化、工作环境变化等所造成的抄表误差是智能电表检定中广泛使用的一种指标,将电表误差与官方规定的阈值进行比较得到抄表误差。若仪表误差超过阈值,仪表将被视为故障仪表。

本研究的目的是以电能数据分析的方式,藉由估测电表读数误差来侦测故障电能表。仪表误差估计方法包括4个步骤。首先,进行数据采集,作为工作的基础部分,其需要获取能源信息(电能消耗、电压、电流等)和系统信息(仪表ID、位置号等),其以

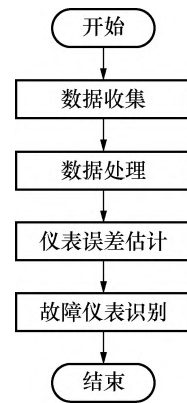


图1 故障仪表检测流程

Fig. 1 Workflow of meter detection

固定的频率传输到数据中心;然后,对数据进行处理,去除异常仪表数据和聚类代表性剖面数据;其次,根据能量平衡关系建立方程,求解得到仪表误差;最后,估计误差,若其高于调节阈值的则为故障仪表。

2.2 数据采集

仪表数据采集自AMI系统。如图2所示,仪表通过电力线载波、RS485和无线通信方式向接入点发送测量数据。通常安装在接入点的集中器,用于收集电表数据,然后上传到云数据中心进行进一步分析。数据包括电能表测量数据,如电能消耗、电压、电流等,系统信息包括电能表ID、表号、日期、时间等。

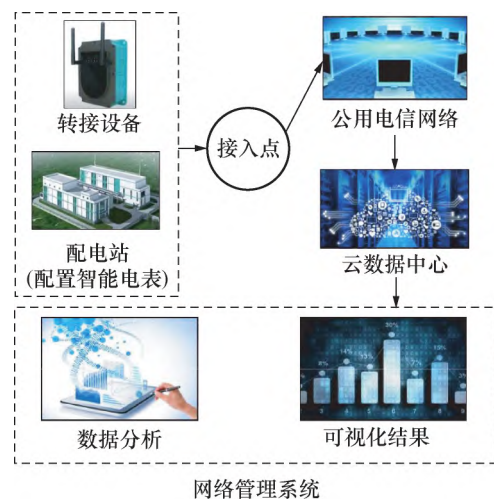


图2 AMI数据采集系统

Fig. 2 AMI data acquisition system

2.3 模型

电能首先通过主电能表,然后再通过子表。在同一时间间隔内,通过主表的能量流为通过子表能量的总和。具体关系如下:

$$\sum_{j=1}^n E_j(i) = E_0(i) \quad (1)$$

式中: $E_j(i)$ 是子表 j 支路 i 上的实际能量消耗; $E_0(i)$ 是实际总能量消耗。

在实际能源系统中,由于内部元件的退化以及环境变化、安装不正常等外部因素的影响,存在测量误差。此外,能量损失的存在使得功率损耗也较难估计。若考虑能量损失和仪表误差,式(1)可改写为:

$$\sum_{j=1}^n \frac{\varphi_j(i)}{1 + \alpha_j} + E(i) = \varphi_0(i) \quad (2)$$

式中: α_j 为子表 j 的相对误差; $\varphi_j(i)$ 为子表 j 的记录能耗值; $\varphi_0(i)$ 为主表的使用值; $E(i)$ 是系统中的能量损失。其中,主表精度高于子表误差,其误差可以忽略。

通过 n 次测量,可以建立矩阵方程如下:

$$\Phi^T \Theta = Y \quad (3)$$

其中:

$$\Phi = \begin{bmatrix} \varphi_1(1) & \varphi_2(1) & \cdots & \varphi_i(1) \\ \varphi_1(2) & \varphi_2(2) & \cdots & \varphi_i(2) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_1(n) & \varphi_2(n) & \cdots & \varphi_n(n) \end{bmatrix}$$

$$Y = \begin{bmatrix} \varphi_0(1) - E(1) \\ \varphi_0(2) - E(2) \\ \vdots \\ \varphi_0(n) - E(n) \end{bmatrix} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{bmatrix}$$

$$\Theta = \begin{bmatrix} \frac{1}{1 + \alpha_1} \\ \frac{1}{1 + \alpha_2} \\ \vdots \\ \frac{1}{1 + \alpha_n} \end{bmatrix} = \begin{bmatrix} \theta(1) \\ \theta(2) \\ \vdots \\ \theta(n) \end{bmatrix}$$

仪表误差可由式(3)算出。智能电表自耗 $e_M(i)$ 、电力线损耗 $e_N(i)$ 和漏电损耗 $e_L(i)$ 是电能损耗的主要组成部分。能量损失 $E(i)$ 的表达式如下:

$$E(i) = e_M(i) + e_N(i) + e_L(i) + \varepsilon(i) \quad (4)$$

式中 $\varepsilon(i)$ 是误差项。其余各项的表达式如下:

$$e_M(i) = \sum_{j=0}^{n-1} \bar{p}_j t(i) \quad (5)$$

$$e_N(i) = U^2 \sigma t(i) \quad (6)$$

$$e_L(i) = \sum_{j=0}^{n-1} (\bar{I}(i))_j^2 r_j t(i) \quad (7)$$

式中: \bar{p}_j 为子表 j 的额定功率; σ 为线路漏电电导; $t(i)$ 为运行时间; U 为线路电压; σ 为线路漏电电

导; $\bar{I}(i)$ 为线路电流; r_j 为从子表 j 到主仪表的线路电阻。在实际系统中, σ 和 r_j 很难得到,损失也很难准确估计。因此,在本文模型中 $e_N(i)$ 和 $e_L(i)$ 可以看作是能量损失误差项的一部分。

2.4 数据处理

数据处理是估计模型中的一项重要工作。在实际系统中,由于一些参数的缺乏,能量损失很难估计,而能量损失项对方程求解有重要影响。如果能量损失误差项大于所有子表误差引起的能量损失,则误差估计会变得准确。智能电能表的抄表误差是一个不随使用时间增长的常数;反之,在不同的能量负荷下,仪表误差会发生变化。

在模型中,主仪表的使用值应大于子仪表的总和。考虑到窃电和电源断线,应剔除子表读数之和超过主表的数据。为了减少电能损耗和电能负荷对仪表误差求解的影响,本文提出一种基于决策树和聚类的数据处理方法。

数据首先分为异常数据和正常数据,然后将正常数据分为轻载和重载负荷。轻载数据是指电表电流值小于电表额定电流,重载时电表电流大于额定电流。然后对不同能量损失率的数据进行分类。比值 γ 为:

$$\gamma = \frac{\varphi_0(i) - \sum_{j=1}^n \varphi_j(i)}{\varphi_0(i)} \quad (8)$$

当 γ 落在不同的比值范围内时,数据可以分为不同的能量损失水平。范围为从 $(0, \gamma_{\max}/k]$ 到 $((k-1)\gamma_{\max}/k, \gamma_{\max}]$, k 是节点数。选择所需的聚类中心数 R 和 K-means 过程迭代移动中心,使点与中心之间的总距离最小。对于给定的观测集 $x(i) = \{x(i)_1, x(i)_2, \dots, x(i)_n\}$ 和质心 $c(j) = \{c(j)_1, c(j)_2, \dots, c(j)_n\}$ 。 $x(i)$ 与质心 $c(j)$ 之间的距离为:

$$\text{dist}(x(i), c(j)) = \|x(i) - c(j)\|_2$$

$$= \sqrt{\sum_{m=1}^n (x(i)_m - c(j)_m)^2} \quad (9)$$

给定簇 $C = \{C^1, C^2, \dots, C^R\}$ 的质心为 $c(j)$, 聚类的目标是使每个点到质心的距离最小:

$$\min \sum_{m=1}^n \sum_{x(i) \in C^R} \|x(i) - c(j)\|_2 \quad (10)$$

中心评估为:

$$c(i) = \frac{\sum_{x(i) \in C^R} x(i)}{n_i} \quad (11)$$

式中 n_i 为中心 $c(i)$ 的观测点数量。

主要采用决策树方法对异常数据进行初步识

别,并将正常数据划分为一个类似的能量损失水平部分。然后对数据进行聚类,得到相似的特征数据用于仪表误差估计。

2.5 仪表误差估计

子表矩阵是一个 $N \times N$ 维矩阵,其解是与仪表误差有关的参数。递归模型可以同时更新参数,节省存储资源。为了重复更新估计结果,提高数据使用效率,采用递推加权最小二乘模型求解方程。

为使残差平方和最小,方程(3)的代价函数为:

$$J = \sum_{i=1}^n (y(i) - \varphi(i) \theta(i))^2 \quad (12)$$

θ 的递推加权最小二乘解为:

$$\hat{\theta} = (\Phi^T(n) W(n) \Phi(n))^{-1} \Phi^T(n) W(n) Y(n) \quad (13)$$

式中: $W = \text{diag}(w_1, w_1, \dots, w_n)$, W 是权重因子矩阵。递归方程为:

$$\hat{\theta}(n+1) = \hat{\theta}(n) + \zeta(n) (y(n+1) - \Phi^T(n+1) \hat{\theta}(n)) \quad (14)$$

式中: $\hat{\theta}(n+1)$ 表示新的估计量; $\zeta(n)$ 是更新因子。

$$\hat{\theta} = (\psi^T(n) W(n) \Phi(n))^{-1} \Phi(n-1) \quad (15)$$

在权重矩阵中使用指数遗忘因子。则估计方程为:

$$\hat{\theta}(n+1) = \hat{\theta}(n) + P(n+1) \Phi(n+1) (y(n+1) - \Phi^T(n+1) \hat{\theta}(n)) \quad (16)$$

$$\zeta(n) = P(n) \Phi(n+1) \cdot (\Phi^T(n+1) P(n) \Phi(n+1) + \lambda)^{-1} \quad (17)$$

$$P(n+1) = (I - \zeta(n) \Phi^T(n+1)) P(n) \frac{1}{\lambda} \quad (18)$$

式中 λ 是遗忘因子。

均方根误差(root mean square error, RMSE) 被广泛应用于通过计算估计值与实际值之间的距离来评价估计性能。较小的 RMSE 意味着更好的性能。其定义为:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{\alpha}_i - \alpha_i)^2} \quad (19)$$

式中: $\hat{\alpha}_i$ 为估计误差; α_i 为实际误差值; n 为子表个数。

3 算例分析

数据集采集自城市居民小区低压电能系统中。在这个能源系统中部署了一个主表和 122 个子表。数据集由 2014 年 8 月至 2016 年 8 月间每 24 h 的所

有电表的能耗值组成。此外,主表每 15 min 记录 1 次电压和电流值。大多数用户是社区中的常住用户。该子表的技术参数如下:额定功率 1 100 W,额定电压 220 V,额定电流 5 A。除了电表电量测量值外,仪表 ID、系统 ID 和社区 ID 都记录在数据集中。此外还删除了数据集中的空行和格式错误的行、冗余数据和无效数据,然后提取出可用的数据。假设所有电表数据有效,系统中无窃电和设备故障。在此期间,所有仪表误差保持不变。然后随机嵌入故障表通过将超过 2% 的主仪表误差分配给某些子仪表。大于 2% 的仪表误差估计值将被视为故障仪表。首先对数据进行筛选和分类,然后对数据进行聚类 and 构造线性方程来估计仪表误差。

3.1 仪表误差估计结果

估计的仪表误差如图 3 所示。 X 轴为测量次数, Y 轴为仪表误差。当新数据输入到模型时,仪表误差更新。

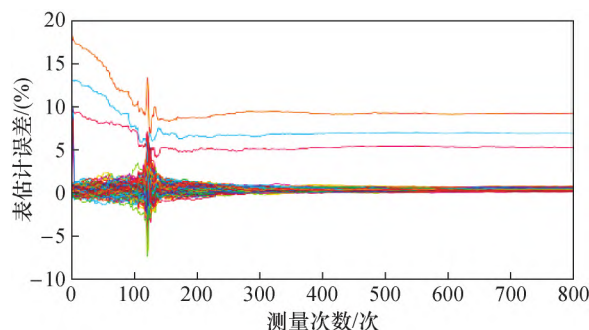


图 3 不同测量次数下的估计误差

Fig. 3 Estimated error in different measurements

在图 3 中,每一行显示了每个递归的误差估计量。随着测量值的增加,估计量收敛到稳定值。3 个仪表的误差明显高于其他仪表,有些仪表的估计误差非常接近于 0,所需测量的数量取决于数据特点。进行大约 300 次测量,可以找到误差估计值的稳定值。该递推模型能很好地估计仪表误差,且误差变化不大。

图 4 显示了当所有值保持稳定时误差估计器的详细信息。可以看出,电表 28、55 和 79 的表计误差大于 2%。79 号仪表的最大误差为 8.596%,而 55 号仪表和 28 号仪表的误差估计值分别为 6.213% 和 4.669%,其他仪表的误差估计值在 -2% 到 2% 的正常范围内。该方法的 RMSE 为 0.21%,这意味着估计误差与真实误差值的偏差很小。误差估计器能很好地反映仪表的误差特性。

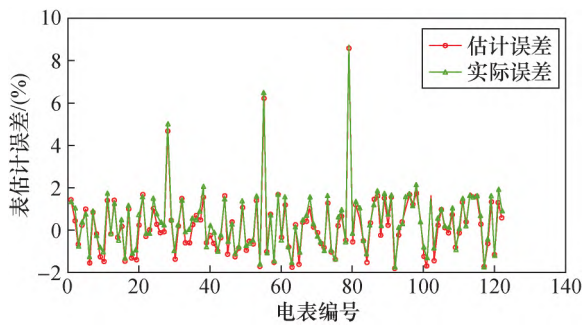


图 4 仪表误差估计

Fig. 4 Meter error estimator

3.2 方案对比

在相同的数据下,将所提方法与 LU 分解(LU factorization)、广义最小残差法(generalized minimal residual method ,GMRES) 进行了比较。图 5 显示,由于子表矩阵的病态性,LU 因子分解有一个不收敛的估计量。广义最小残差法得到的误差估计量含有奇异值。表 1 显示了 3 种方法的 RMSE,所提出的方法具有最小的 RMSE,而 LU 因子分解具有最大的 RMSE。仪表误差阈值为 2%,RMSE 不应大于 2%。LU 分解方法表现出了糟糕的性能,而且通过对比 RMSE 值可见,所提模型比 GMRES 性能更好。

图 5 中的合计误差为所提方法估计误差 RMSE 变化曲线,曲线在零轴附近小幅波动,表明所提方法整体估计误差准确度高。

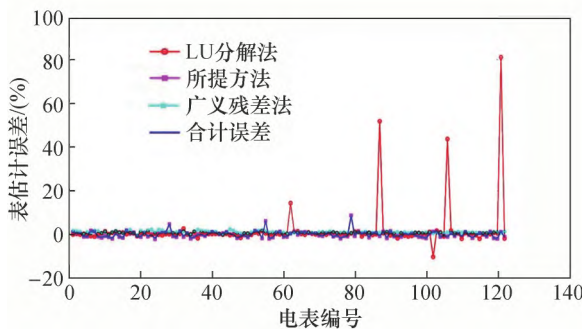


图 5 仪表误差估计对比

Fig. 5 Meter error estimator comparison

表 1 性能比较

Tab. 1 Performance comparison

	LU 分解	GMRES	所提方法
RMSE	92%	0. 9%	0. 21%

3.3 检测正确率分析

图 6 展示了系统中故障仪表数量与所提模型检测正确率之间的关系。检测正确率是指正确划分电

表状态的数量所占的总电表数量的百分比。可以看出,当故障表数不大于 23 时,所提模型运行良好。正确率在 90% 以上,如果系统中部署的故障不超过 10 个,则可以检测到所有异常数据。随着故障仪表的增多,模型性能会越来越差。

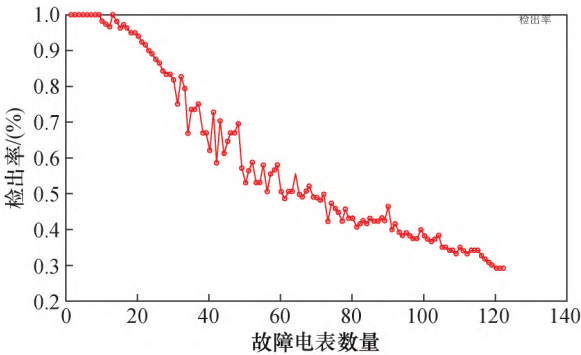


图 6 检测正确率

Fig. 6 Detection rate

4 结 论

本文介绍了一种检测智能电表故障的数据分析方法,主要提出了一种基于决策树和聚类理论的电表误差估计递推算法。实验结果表明该方法具有良好的分类性能,且在相同的数据背景下,与常规方法相比,所提模型的检出正确率更高。由于智能电表数据分析在电表检测中有着广泛的应用和较高的经济价值,所提方法能为相关人员在边缘计算环境下的图像仪表误差在线估计和智能仪表数据分析等方面提供理论支持。

[参 考 文 献]

[1] 钱雅伦. 基于智能算法的配电网故障自愈方法研究 [D]. 北京: 华北电力大学, 2018.

[2] 薛禹胜, 李满礼, 罗剑波, 等. 基于关联特性矩阵的电网信息物理系统耦合建模方法 [J]. 电力系统自动化, 2018, 42 (2): 11 - 19.

Xue Y S, Li M L, Luo J B, et al. Coupling modeling method of power grid information physical system based on correlation characteristic matrix [J]. *Automation of Electric Power Systems*, 2018, 42 (2): 11 - 19.

[3] 马爽, 徐震, 王利明. 基于集合论的电网信息物理系统模型构建方法 [J]. 电力系统自动化, 2017, 41 (6): 1 - 5.

Ma S, Xu Z, Wang L M. Modeling method of power grid information physical system based on set theory [J]. *Automation of Electric Power Systems*, 2017, 41 (6): 1 - 5.

- [4] 张淑清,任爽,陈荣飞,等. 基于大数据简约及 PCA 改进 RBF 网络的短期电力负荷预测 [J]. 计量学报, 2018, 39 (3): 392-396.
Zhang S Q, Ren S, Chen R F, et al. Short term power load forecasting based on big data reduction and PCA improved RBF network [J]. *Acta Metrologica Sinica*, 2018, 39 (3): 392-396.
- [5] Sun Q, Li H, Ma Z, et al. A comprehensive review of smart energy meters in intelligent energy networks [J]. *IEEE Internet Things J*, 2016, 3 (4): 464-479.
- [6] Alahakoon D, Yu X. Smart electricity meter data intelligence for future energy systems: A survey [J]. *IEEE Trans Ind Informat*, 2016, 12 (1): 425-436.
- [7] 郝建军,王启银,张兴忠. 基于支持向量机的电网通信入侵检测技术 [J]. 电测与仪表, 2019, 56 (22): 109-114.
Hao J J, Wang Q Y, Zhang X Z. Intrusion detection technology of power grid communication based on support vector machine [J]. *Electrical Measurement Instrumentation*, 2019, 56 (22): 109-114.
- [8] 谢鑫,王黎敏,张红雨,等. 智能电表可靠性预计标准对比研究 [J]. 计量学报, 2019, 40 (S1): 60-63.
Xie X, Wang L M, Zhang H Y, et al. Comparative study on reliability prediction standards of smart meters [J]. *Acta Metrologica Sinica*, 2019, 40 (S1): 60-63.
- [9] Fenza G, Gallo M, Loia V. Drift-aware methodology for anomaly detection in smart grid [J]. *IEEE Access*, 2019, 7: 9645-9657.
- [10] Cernazanu-Glavan C, Marcu M. Anomaly detection using power signature of consumer electrical devices [J]. *Adv Electr Comput Eng*, 2015, 15 (1): 89-94.
- [11] 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 Y, Cui C, Zhang L, et al. Research on Intelligent non intrusive load monitoring system [J]. *Journal of Electric Power Science and Technology*, 2019, 34 (2): 155-160.
- [12] Liu Y, Hu S. Cyberthreat analysis and detection for energy theft in social networking of smart homes [J]. *IEEE Trans Comput Social Syst*, 2015, 2 (4): 148-158.
- [13] 刘迪迪,马丽纳,孙浩天,等. 智能电网时变电价下基站的动态能量管理 [J]. 北京邮电大学学报, 2019, 42 (5): 54-61.
Liu D D, Ma L N, Sun H T, et al. Dynamic energy management of base station under time-varying price of smart grid [J]. *Journal of Beijing University of Posts and Telecommunications*, 2019, 42 (5): 54-61.
- [14] Xiao F, Ai Q. Electricity theft detection in smart grid using random matrix theory [J]. *IET Generation Transmission & Distribution*, 2018, 12 (2): 371-378.
- [15] 张淑清,张赞,刘海涛,等. 多重分形去趋势波动分析及改进决策树在电能质量分析中的应用 [J]. 计量学报, 2021, 42(4): 424-431.
Zhang S Q, Zhang Y, Liu H T, et al. Application in Power Quality Analysis Based on Multifractal Detrended Fluctuation Analysis and Improved Decision Tree [J]. *Acta Metrologica Sinica*, 2021, 42(4): 424-431.
- [16] 张淑清,乔永静,姜安琦,等. 基于 CEEMD 和 GG 聚类的电能质量扰动识别 [J]. 计量学报, 2019, 40 (1): 49-57.
Zhang S Q, Qiao Y J, Jiang A Q, et al. Power Quality Disturbance Identification Based on CEEMD and GG Clustering [J]. *Acta Metrologica Sinica*, 2019, 40(1): 49-57.
- [17] 席禹,陈波,袁智勇,等. 配电自动化系统主站网络化在线测试技术研究 [J]. 计量学报, 2020, 41(5): 614-619.
Xi Y, Chen B, Yuan Z Y, et al. Research on Network Online Testing Technology for Main Station of Power Distribution Automation System [J]. *Acta Metrologica Sinica*, 2020, 41(5): 614-619.



第一作者: 史鹏博 (1989-), 满族, 北京人, 硕士研究生, 工程师, 主要研究方向为电能计量技术。

Email: c3346789@163.com