

文章编号: 2095-4980(2024)07-0800-07

基于数据挖掘算法的电网调度信号异常数据提取方法

张洪略^a, 万毅^b, 王家军^c, 石家德^d, 金贵红^d

(贵州电网有限责任公司 a. 电力调度控制中心, 贵州 贵阳 550000; b. 都匀供电局, 贵州 都匀 558000;
c. 兴义供电局, 贵州 兴义 562400; d. 凯里供电局, 贵州 凯里 556000)

摘要: 在获取电网调度信号后, 大多采用传统深度置信型辨识制度提取异常数据, 只能获取低维数据包含的异常信息参量, 使得最终数据提取结果曲线下面积(AUC)值较低。因此, 为了提高电网调度信号异常数据提取结果的 AUC 值, 提出基于数据挖掘算法的电网调度信号异常数据提取方法。应用独立成分分析算法处理电网调度信号, 去除信号中的噪声信息。并对去噪后的信号进行小波分解, 得到多个子信号数据集。运用数据挖掘算法中的聚类算法分析子信号数据集, 得到数据样本特征, 并在考虑属性特征密度指标的情况下完成数据特征分类, 获取异常数据特征。最后, 在支持向量数据描述的辅助下, 检测出电网调度信号异常数据, 汇总这部分数据即可完成异常数据提取。实验结果表明, 所提方法应用后得到的异常数据提取结果 AUC 值总是大于 0.85, 证明了其具优越的应用效果。

关键词: 数据挖掘算法; 电网调度信号; 异常数据; 特征提取; 去噪; 小波分解

中图分类号: TP391

文献标志码: A

doi: 10.11805/TKYDA2023381

Abnormal data extracting from power grid dispatching signals based on data mining algorithms

ZHANG Honglue^a, WAN Yi^b, WANG Jiajun^c, SHI Jiade^d, JIN Guihong^d

(a. Power Dispatching Control Center, Guizhou Power Grid Co., LTD., Guiyang Guizhou 550000, China;
b. Duyun Power Supply Bureau, Guizhou Power Grid Co., LTD., Duyun Guizhou 558000, China;
c. Xingyi Power Supply Bureau, Guizhou Power Grid Co., LTD., Xingyi Guizhou 562400, China;
d. Kaili Power Supply Bureau, Guizhou Power Grid Co., LTD., Kaili Guizhou 556000, China)

Abstract: After acquiring the power grid dispatching signals, traditional deep confidence identification systems are mostly used for anomaly data extraction, which can only obtain the anomaly information parameters contained in low-dimensional data, resulting in a lower Area Under the Curve (AUC) value of the final data extraction result. Therefore, in order to improve the AUC value of the anomaly data extraction results of the power grid dispatching signals, an anomaly data extraction method for power grid dispatching signals based on data mining algorithms is proposed. The power grid dispatching signals are processed using the Independent Component Analysis(ICA) algorithm to remove noise from the signals. The denoised signals are then subjected to wavelet decomposition to obtain multiple sub-signal datasets. Clustering algorithms in data mining algorithms are employed to analyze the sub-signal datasets to obtain the characteristics of the data samples, and data feature classification is completed considering the attribute feature density index to obtain the anomaly data characteristics. Finally, with the assistance of the Support Vector Data Description(SVDD), the abnormal data in the power grid dispatching signals are detected, and summarizing this part of the data can complete the anomaly data extraction. The experimental results show that the AUC value of the anomaly data extraction results obtained after applying the proposed method is always greater than 0.85, proving its superior application effect.

Keywords: data mining algorithms; power grid dispatch signal; abnormal data; feature extraction;

denoising; wavelet decomposition

在电能使用量不断增长的社会背景下，电网调度问题备受关注^[1]。尤其在智能电网广泛发展后，其可以在复杂的供电环境下，实现电网系统的合理调度，从而保证电力系统的供电稳定性。但在电网调度规模逐渐增大后，智能调度工作会受到多种因素干扰，频繁出现异常情况，对供电产生负面影响^[2]。因此，如何准确提取电网调度信号中的异常数据，指导电网调度工作，成为学者们探讨的热门话题。

文献[3]考虑不同维度上电力调度数据的分布存在明显的差异性，设计一种基于马氏距离度量方法的异常数据提取方法。分别计算数据分布中心与每个数据点之间的距离，并应用对数区间隔策略构建异常检测子树，进一步分析马氏距离度量值，即可提取出所有信号异常数据。但是，该方法未去除信号中的噪声信息，导致异常数据提取中受到噪声干扰，效果较差。文献[4]应用独立成分分析算法处理电网调度噪声信号，并对去噪信号进行小波分解，观察信号特征后可以推算出单个信号数据样本的异常值。结合异常判断阈值，即可找出异常数据。但是该方法未考虑属性特征密度指标无法准确对数据特征分类，导致异常数据整体提取精确度较低。文献[5]结合孤立森林算法和动态选择集成原理，建立多个电力调度数据异常检测器，并引入元学习策略，确保所有异常检测器可以通过多种指标互补，实现电网调度信号异常数据的动态选择，依据选择结果即可实现异常数据准确提取。该方法进行异常数据提取时，未对信号进行分解，难以准确找到异常数据。

为了快速、有效地解决电网调度异常问题，本研究提出一种以数据挖掘算法为核心的电网调度信号异常数据提取方法，在调度信号去噪和分解处理后，运用数据挖掘领域的聚类算法，准确提取数据特征，基于此检测出电网调度信号中包含的所有异常数据，从而完成高质量数据提取。

1 设计基于数据挖掘算法的电网调度信号异常数据提取方法

1.1 电网调度信号去噪处理

针对电网调度信号进行异常分析时，为了避免脉冲噪声带来的负面影响，需要先利用独立成分分析算法进行信号去噪处理^[6]。去噪过程中，假设当前电网调度信号中包含数个随机变量，将其与混合矩阵、分离矩阵结合起来，即可构建图 1 所示的信号分离模型。

利用图 1 所示的模型处理电网调度信号，本质上是对高斯数据进行线性变换处理。先分离所有信号独立条件获得分离矩阵，并将其应用到观测信号上即可得出源信号估计值，如式(1)所示。

$$y = E[(Wx) \times (Wx)^T] \tag{1}$$

式中： x 表示观测信号； y 表示源信号估计值； E 表示数学期望函数； W 表示正交分离矩阵； T 表示转置。

考虑到分离矩阵、源信号中拥有未知参数会影响信号去噪结果，在处理源信号时，以最大矩阵与分离信号独立性为目标，求出贴合实际情况的源信号。随后考虑到异常数据主要处在观测信号中，针对其构建式(2)所示的噪声观测信号模型。

$$\varepsilon(t) = WA[y(t) + N] \tag{2}$$

式中： t 表示信号采集时刻； A 表示混合矩阵； ε 表示噪声信号； N 表示随机噪声异常值：

$$N = \sum_{i=1}^x \beta_i \times y \tag{3}$$

式中： i 表示电网调度信号内包含的第 i 个脉冲噪声； β 表示脉冲噪声振幅。

通过上述分析，提取出电网调度信号中包含的所有噪声干扰信息，将其从原始信号中剔除，完成信号去噪处理。

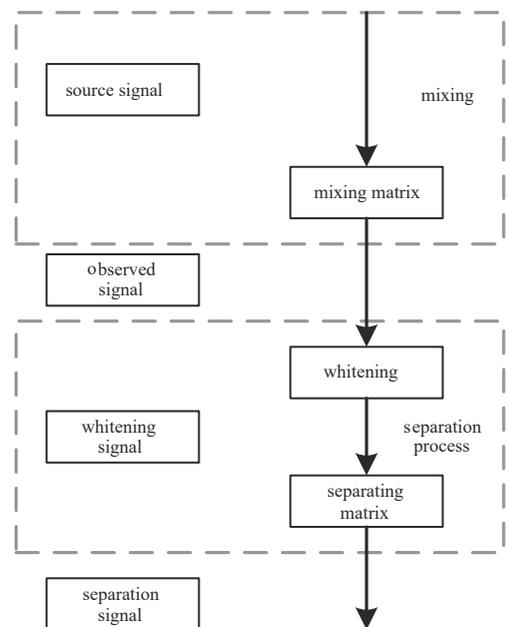


Fig.1 Structural diagram of the signal separation model
图 1 信号分离模型结构图

1.2 建立调度信号小波分解方案

为了便于分析电网调度信号中的异常问题, 依托于傅里叶变换思想对去噪后的信号进行小波分解, 得到一组不同频率的本构分量信号^[7], 针对每个本构分量分别观察, 有利于准确找到异常数据。实际操作过程中, 主要采用了离散小波变换原理完成信号分解, 如式(4)所示。

$$\rho(m, p) = 2^{-(m/2)} \sum_{q=0}^Q y'(t) \psi\left(\frac{q-p^{2^m}}{2^m}\right) \quad (4)$$

式中: ρ 表示离散小波变换结果; m 表示尺度因子; p 表示平移参数; q 表示离散点; Q 表示离散点数量; y' 表示去噪处理后的电网调度信号; ψ 表示母波信号。

现有的调度信号经过一次离散小波变换结束后, 转变为低频信号和低频信号, 小波分解过程可以定义为式(5)。

$$y'(t) = \chi_1(t) + D_1(t) = \chi_2(t) + D_2(t) + D_1(t) \quad (5)$$

式中: D_1 表示高频分量信号; χ_1 表示低频分量信号; χ_2 、 D_2 分别表示低频信号 χ_1 再次分解形成的低频、高频近似分量。

在两级小波分解完成后, 得到的低频近似分量信号曲线较为光滑, 描述了调度信号的总体变化趋势^[8], 另外两个高频细节分量则详细描述了局部信号变化特征, 基于此可以展开后续异常数据提取。

1.3 设计基于数据挖掘的异常特征获取方法

从小波分解后形成的 3 个子信号数据集入手, 运用数据挖掘算法对其深入分析, 了解电网调度信号数据的波动性与变异性, 从而获取数据异常特征。考虑到电网调度信号异常数据提取涉及的数据量较多, 且其中包含大量高维数据, 为了全面获取异常特征, 应用数据挖掘算法中的聚类算法进行特征提取。

针对电力调度信号数据样本设置多个聚类中心, 并计算出每个数据样本与聚类中心之间的模糊隶属度^[9], 假设每个聚类中电网调度数据存在数个特征, 则每个特征的密度指标可以表示为:

$$\theta_j = \sum_{l=1}^L \exp\left\{-\frac{\|c_j - c_l\|^2}{(r_c/2)^2}\right\} \quad (6)$$

式中: θ 表示密度指标; j 、 l 表示电力调度信号数据; L 表示电力调度信号数据样本集; c 表示数据特征; r 表示特征邻域区间半径。

经过多次数据聚类处理后, 根据已知的聚类中心更新结果, 将密度指标变换为式(7)。

$$\theta_j = \theta_a \exp\left\{-\frac{\|c_j - c_l\|^2}{(r'_c/2)^2}\right\} \quad (7)$$

式中: a 表示电网调度信号数据聚类次数; r' 表示数据密度指标的邻域区间半径。

基于特征密度指标, 可以求取出调度信号数据特征, 如式(8)所示。

$$\varphi = \frac{\theta_{a+1}}{\theta_1} \quad (8)$$

考虑到阈值的选择, 会影响最终特征提取质量, 本研究在数据特征聚类分析过程中, 引入高阶统计模型, 计算出可以应用到特征提取优化的补偿计量参数^[10]。再对每个聚类中提取的数据特征进行聚类分析, 使其分类为个体最优解集合和全局最优解集合, 并设置数据特征分类的更新模式为:

$$\delta_{gb}^{(a+1)} = \delta_{gb}^{(a)} + \tau_1 z_1 (d_{gb}^{(a)} - \delta_{gb}^{(a)}) + \tau_2 z_2 (d_{gb}^{(a)} - \delta_{gb}^{(a)}) \quad (9)$$

式中: δ 表示数据点; g 表示节点; b 表示特征维度; τ_1 、 τ_2 表示优化加速系数; z_1 、 z_2 表示随机值; d 表示特征个体; $\delta_{gb}^{(a)}$ 、 $d_{gb}^{(a)}$ 分别表示个体最优解集、全局最优解集中的特征。

随后, 电网调度信号数据特征分布聚类的最大值和平均粒度进行计算, 得到:

$$\zeta' = \left| \max [c_{gb}(t)] \right| \quad (10)$$

$$\bar{\zeta} = \left| \sum_{b=1}^B \zeta_{gb}(t) \right| \quad (11)$$

式中： ζ' 表示特征分布聚类最大值； $\bar{\zeta}$ 表示平均粒度； B 表示维度数量； ζ 表示特征分布聚类值。

结合式(10)和式(11)，可以定义高阶统计量的特征聚类度为：

$$v = \frac{|\zeta' - \bar{\zeta}|}{\zeta'} \quad (12)$$

在基于数据挖掘算法进行信号数据特征提取时，不断更新特征聚类度取值，使得数据特征聚类参数得到优化，选择出有效的特征聚类，得到电网调度信号异常特征提取结果。

1.4 实现调度信号异常数据提取

在异常特征获取完成后，进行简单的归一化处理，使得所有特征值的量纲保持统一。再运用支持向量机模型进一步分析异常数据特征，完成正常特征和异常特征的分类，检测出调度信号异常数据，从而提取出所有异常数据。对经典的支持向量机模式进行改进，得到支持向量数据描述(SVDD)算法，建立一个超球用来识别异常数据，如图2所示。

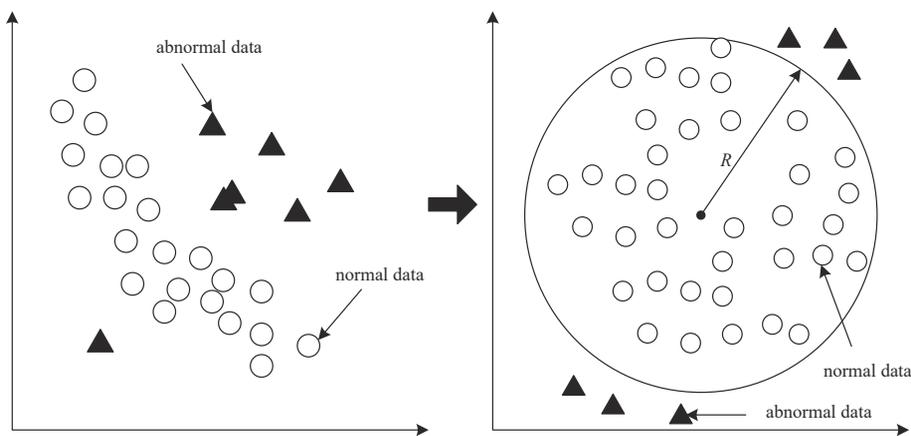


Fig.2 SVDD model structure diagram
图2 SVDD模型结构图

运用图2所示的SVDD模型进行异常数据提取时，将所有提取的数据特征依据一个非线性映射函数，映射到高维特征空间，再分析高维特征空间中异常特征的像与超球中心之间的间距，以此来判断当前数据是否属于异常数据。

$$\omega(\Phi) = \|\mu(\Phi) - e\|^2 \quad (13)$$

式中： Φ 表示异常特征样本； ω 表示特征像与超球中心的距离； μ 表示非线性映射函数； e 表示超球中心。

当式(13)计算结果大于超球半径，即可判断当前电网调度信号数据属于异常数据，需要将其提取出来，遍历所有获取的异常特征后，即可完成异常数据提取。

2 实验

2.1 实验数据

应用数据挖掘算法设计新的电网调度信号异常数据提取方法后，为了检验该方法的应用效果，从某电力系统管理系统中采集部分电网调度信号，组成6个实验数据集，每个数据集的具体信息如表1所示。

设置异常数据阈值指标为密度间距0.35。在实验开始之前，需要对所有电网调度信号进行去噪处理，以任意一段电网调度信号为例，去噪前后的对比结果如图3所示。

根据图3可知，去噪前信号的功率集中在420~500 W，无法看出信号变化，去噪处理后电网调度信号分离出干扰信号，变化规律更加明显，有利于准确提取信号异常数据。按照同样的操作方式处理所有电网调度信号，并根据去噪后的信号更新数据集数据样本。

表1 数据集详细信息

Table1 Dataset details

nomenclature	sample capacity	dimension	normal sample size	abnormal sample size	ratio/%
S1	6 435	36	4 399	2 036	31.64
S2	36 752	9	34 108	2 644	7.19
S3	2 665	5	1 693	972	36.47
S4	4 898	11	4 710	188	3.84
S5	3 505	21	1 848	1 657	47.28
S6	768	8	500	268	34.90

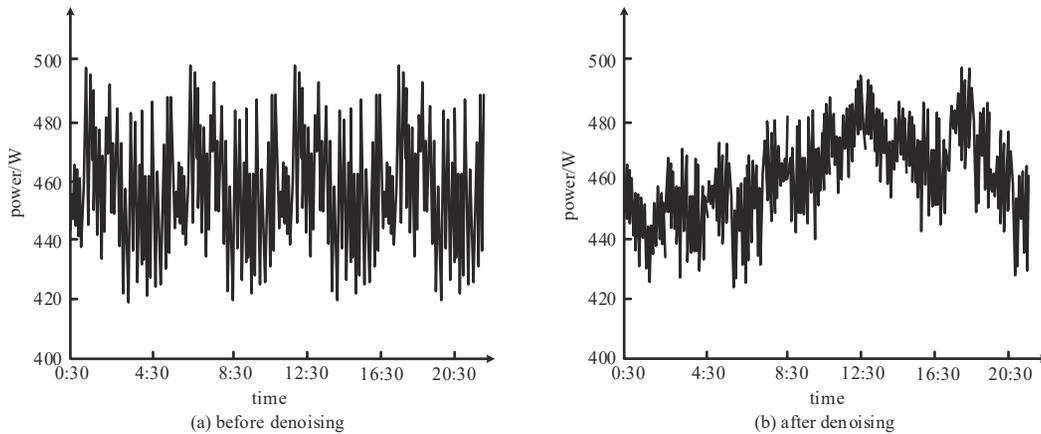


Fig.3 Power grid dispatching signal before and after denoising

图3 去噪前后电网调度信号

2.2 异常数据提取结果

基于聚类算法进行电网调度信号数据特征提取，是异常数据提取的前提。考虑到不同聚类中心选择方法应用后，所得到的聚类结果存在差异，本次实验过程中，选择3种典型的聚类中心选择方法进行测试，得到的聚类结果如图4所示。

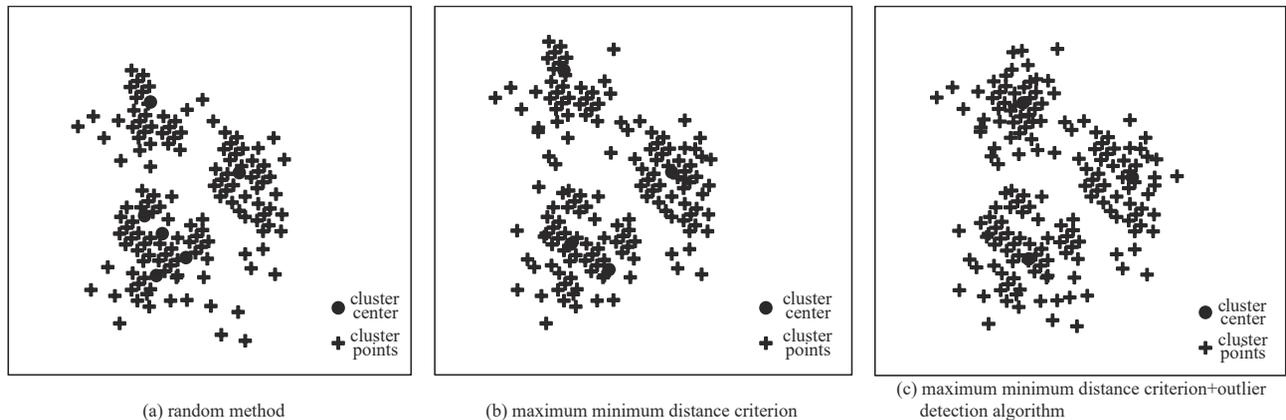


Fig.4 Cluster results for different cluster center selection methods

图4 不同聚类中心选取方法的聚类结果

从图4可以看出，结合最大最小距离准则和离群检测算法的聚类中心选择方法，所形成的聚类中心数量更少，且聚类中心分布更均匀，证明该方法应用后聚类效果最优。

以表1所示的6个数据集为例，在上述聚类中心选择方法的帮助下进行聚类分析，提取出电网调度信号异常特征，并通过支持向量数据描述算法进一步分析，最终提取出异常样本数据数量分别为2 036、2 644、972、188、1 657和268。这一提取结果与数据集详细信息基本相符，证明了所提方法是可行的。

2.3 方法性能对比

考虑到最终提取出的异常数据中可能包含部分识别错误的正常数据，而提取数据中正常数据比例最小，是异常数据提取方法的设计目标。为了进一步观察方法应用性能，选择文献[3]、[4]和[5]方法同步开展实验，对于

数据集 S1 来说，统计不同方法数据提取结果后，可以绘制图 5 所示的接收者操作特征(Receiver Operating Characteristic, ROC)曲线(以假阳率和真阳率为横纵坐标。假阳率表示在二分类问题中，将负样本错误地预测为正样本的比例；真阳率指正确地将正样本判定为正样本的比例)。

统计 ROC 曲线下面积，即可得到不同方法异常数据提取结果 AUC 值，按照这种计算方式，计算出不同方法对不同数据集的异常数据提取结果 AUC 值，生成图 6 所示的对比结果。

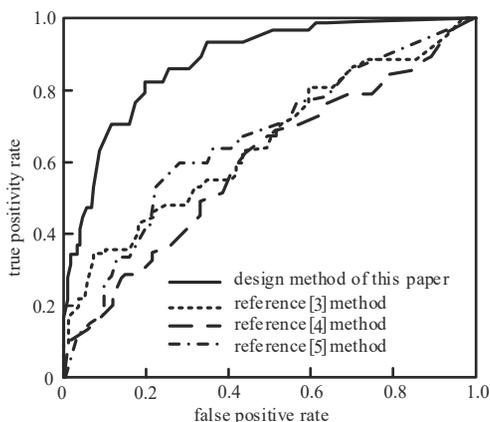


Fig.5 ROC curves of data extraction results using different methods in dataset S1
图 5 数据集 S1 不同方法数据提取结果 ROC 曲线

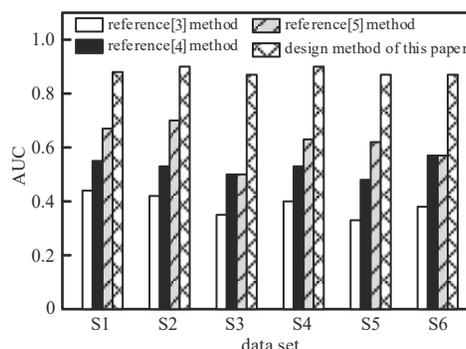


Fig.6 Comparison of AUC values of abnormal data extraction results of different methods
图 6 不同方法异常数据提取结果 AUC 值对比

根据图 6 可知，新提出的方法应用后，可以得到更加准确的电网调度信号异常数据提取结果，该提取结果的 AUC 值始终保持在 0.85 以上，而另外 3 篇文献提出的数据提取方法 AUC 值明显更低，这就验证了基于数据挖掘算法的新方法的优越性。

基于数据集 S1 和 AUC 指标，设计消融实验，测试本文方法进行电网调度信号去噪处理、调度信号小波分解、特征数据挖掘、实现异常数据提取 4 个步骤对异常数据提取结果的影响。实验结果如表 2 所示。由表 2 可知，电网调度信号去噪处理、调度信号小波分解、特征数据挖掘、实现异常数据提取这 4 个步骤中任意步骤缺失均会降低异常数据提取精确度。

表 2 消融实验结果

Table 2 Results of the ablation experiments

measure	AUC
denoising processing+signal wavelet decomposition+feature data mining+anomaly data extraction	0.85
denoising processing+signal wavelet decomposition+anomaly data extraction	0.81
signal wavelet decomposition+anomaly data extraction	0.78
denoising processing+anomaly data extraction	0.75
abnormal data extraction	0.66

3 结论

以智能电网调度过程中的异常检测为研究目标，提出一种结合数据挖掘算法的电网调度信号异常数据提取方法，在聚类算法的辅助下，准确找到信号异常数据并提取出来，从而满足人们越来越高的异常检测要求。

参考文献：

[1] 刘震宇. 基于支持向量机的电力系统调度数据异常检测方法[J]. 自动化技术与应用, 2023,42(6):24-27,37. (LIU Zhenyu. Anomaly detection method of power system dispatching data based on support vector machine[J]. Techniques of Automation and Applications, 2023,42(6):24-27,37.) doi:10.20033/j.1003-7241.(2023)06-0024-05.

[2] 李强,张立梅,白牧可. 基于多元数据特征和改进随机森林的智能配电网异常数据辨识[J]. 科学技术与工程, 2023,23(5):2007-2015. (LI Qiang, ZHANG Limei, BAI Muke. Identification of abnormal data in smart distribution network based on multivariate data features and improved random forest[J]. Science Technology and Engineering, 2023,23(5):2007-2015.) doi:10.3969/j.issn.1671-1815.2023.05.027.

[3] 王锋,高欣,贾欣,等. 一种基于对数区间隔离森林的电力调度数据异常检测集成算法[J]. 电网技术, 2021,45(12):4818-4827. (WANG Feng, GAO Xin, JIA Xin, et al. An anomaly detection ensemble algorithm for power dispatching data based on log-

- interval isolation[J]. Power System Technology, 2021,45(12):4818–4827.) doi:10.13335/j.1000-3673.pst.2020.2112.
- [4] 刘峰,朱颀,张凯,等. 数据挖掘下电网调度信号异常数据提取方法[J]. 湘潭大学学报(自然科学版), 2021,43(6):74–80. (LIU Feng, ZHU Jie, ZHANG Kai, et al. Abnormal data extraction method of power grid dispatching signal under data mining[J]. Journal of Xiangtan University(Natural Science Edition), 2021,43(6):74–80.) doi:10.13715/j.cnki.nsjxu.2021.06.011.
- [5] 傅世元,高欣,张浩,等. 基于元学习动态选择集成的电力调度数据异常检测方法[J]. 电网技术, 2022,46(8):3248–3256. (FU Shiyuan, GAO Xin, ZHANG Hao, et al. Anomaly detection for power dispatching data based on meta-learning dynamic ensemble selection[J]. Power System Technology, 2022,46(8):3248–3256.) doi:10.13335/j.1000-3673.pst.2022.0017.
- [6] 管萸,谢小川,胡琳,等. 基于数据标签的智能电网监控与异常检测[J]. 工程科学与技术, 2023,55(3):243–254. (GUAN Ti, XIE Xiaochuan, HU Lin, et al. Outlier data detection and monitoring of smart electric grid based on data-tag[J]. Advanced Engineering Sciences, 2023,55(3):243–254.) doi:10.15961/j.jsuese.202101248.
- [7] 田银磊,刘书伦. 基于神经网络的船舶通信网络异常数据识别[J]. 舰船科学技术, 2022,44(17):148–151. (TIAN Yinlei, LIU Shulun. Recognition of abnormal data in ship communication network based on neural network[J]. Ship Science and Technology, 2022,44(17):148–151.) doi:10.3404/j.issn.1672-7649.2022.17.030.
- [8] 况华,何鑫,何觅,等. 基于双向长短期记忆神经网络的配网电压异常数据检测[J]. 科学技术与工程, 2021,21(24):10291–10297. (KUANG Hua, HE Xin, HE Mi, et al. Abnormal voltage data detection of distribution network based on bidirectional long short-term memory neural network[J]. Science Technology and Engineering, 2021,21(24):10291–10297.)
- [9] 孙滢涛,张锋明,陈水标,等. 基于多域特征提取的电力数据异常检测方法[J]. 电力系统及其自动化学报, 2022,34(6):105–113. (SUN Yingtao, ZHANG Fengming, CHEN Shuibiao, et al. Power data anomaly detection algorithm based on multi-domain feature extraction[J]. Proceedings of the CSU-EPSCA, 2022,34(6):105–113.) doi:10.19635/j.cnki.csu-epsa.000805.
- [10] 张继丹,肖东,侯燕曦. 基于MES频谱数据异常贡献度估计与后验分析[J]. 太赫兹科学与电子信息学报, 2022,20(12):1277–1284. (ZHANG Jidan, XIAO Dong, HOU Yanxi. Estimation of anomaly contribution in spectrum data based on MES and backtesting[J]. Journal of Terahertz Science and Electronic Information Technology, 2022,20(12):1277–1284.) doi:10.11805/TKYDA2021178.)

作者简介:

张洪略(1994–), 男, 硕士, 高级工程师, 主要研究方向为电力调度自动化 .email:burehonglue@163.com.

万毅(1988–), 男, 本科, 高级工程师, 主要研究方向为电力调度自动化.

王家军(1977–), 男, 本科, 高级工程师, 主要研究方向为电力调度自动化.

石家德(1990–), 男, 本科, 高级工程师, 主要研究方向为电力调度自动化.

金贵红(1987–), 男, 本科, 高级工程师, 主要研究方向为电力调度自动化.