Abnormal Operation Diagnosis Method of Electric Energy Metering Equipment Based on Data Mining

Jianfeng Sun*, Yang Gao, Cunyu Long, Shubei Hua
* Corresponding author: sunjf0502@163.com

Yang Gao: gaoyangdaisy@outlook.com

Cunyu Long: 15597435245@163.com

Shubei Hua: a14709789294202106@163.com

Marketing Service Center of State Grid Qinghai Electric Power Company, Xining, China

Abstract: The integration of emerging technologies and power grid business in the information age has also put forward new requirements. The demand for improving the data acquisition, processing and application capabilities of power grid construction in the new era is increasing day by day. At the same time, it also puts forward higher requirements for using new technologies to improve the efficiency of production and operation. The abnormal behavior of power grid metering equipment leads to line loss, which not only causes damage to power grid facilities, but also seriously threatens the stability and safety of power grid. In view of the large range of original data of electric power equipment and the difficulties of parameter selection and low computational efficiency in the original K-means algorithm, this paper establishes an outlier detection model based on the improved K-means algorithm to preliminarily screen out the abnormal operation set of metering equipment, and further screens out the final data set through the similarity analysis of curves. Finally, the simulation and comparative experiments prove that the anomaly diagnosis detection model based on clustering analysis can achieve good results in both detection rate and the false detection rate on the data set. It provides a data basis for the operation of power equipment and also provides theoretical support for the maintenance and repair of SGCC.

Keywords: Electric energy metering equipment, Data anomaly, Diagnosis model, Data mining

1. Introduction

In the practical application scenarios of electric energy metering anomaly diagnosis, researchers have designed a series of early warning measures for emergencies with known anomalies. However, due to the uncertainty and unpredictability of unknown anomalies, the occurrence of unknown anomalies may bring more serious losses to the power grid, so it deserves more attention [1]. For small sample datasets, some categories have no actual cases or only a few cases, so it is impossible to use traditional data enhancement methods to expand the dataset. The model trained by the data set has high over-fitting, poor model accuracy, poor performance and insufficient generalization ability [2].

BDEIM 2022, December 02-03, Zhengzhou, People's Republic of China

DOI 10.4108/eai.2-12-2022.2328715

In recent years, technologies such as abnormal power consumption identification and power consumption fraud detection based on data mining theory have been proposed one after another. Reference [3] obtains the characteristic curve of each type of user load curve based on the clustering method in unsupervised learning. Reference [4] and other methods based on fuzzy support vector machines use the optimized model to screen out abnormal power consumption users. Reference [5] describes the abnormal power consumption behavior of users from multiple angles based on the abnormal power consumption identification model of multi-dimensional composite characteristics of users. Reference [6] adopts three typical intelligent detection algorithms to identify abnormal power users from the perspectives of unbalanced data distribution, data balance and data weighting so as to verify that the detection accuracy of the method is higher than that of the traditional detection method. Reference [7] analyzes the electricity consumption behavior of target user groups by means of user electricity consumption, point load curve and multi-index comprehensive score, and constructs a behavior characteristic model to identify suspected users. To sum up, a fast and accurate abnormal power consumption identification system is of far-reaching significance to power supply companies. In the face of complex power consumption data, the methods of abnormal power consumption identification are constantly updated.

For abnormal power data, the model is prone to over-fitting and lack of generalization ability and can not detect unknown anomalies. In this paper, solutions to these two problems are proposed: for the abnormal data of power equipment, the method of using small sample comparative learning is proposed, so that the model has the ability of comparative learning. Based on contrastive learning, a decision method based on confidence is proposed to make the model detect unknown anomalies. In the experimental analysis, the simulation data and the model evaluation index are compared and tested. The results indicate that the proposed method can effectively improve the generalization ability of the model.

2. Build a data model

Power data modeling is the premise of subsequent model optimization and algorithm design, and is the key to the success of deep learning algorithm. This paper introduces the characteristics of power data, data preprocessing methods, combined with professional knowledge and experience to design a reasonable power anomaly index, and gives the detailed steps to establish the data model [8]. Therefore, it is necessary to retain the information of time dimension in data analysis. The data modeling is displayed in Figure 1.

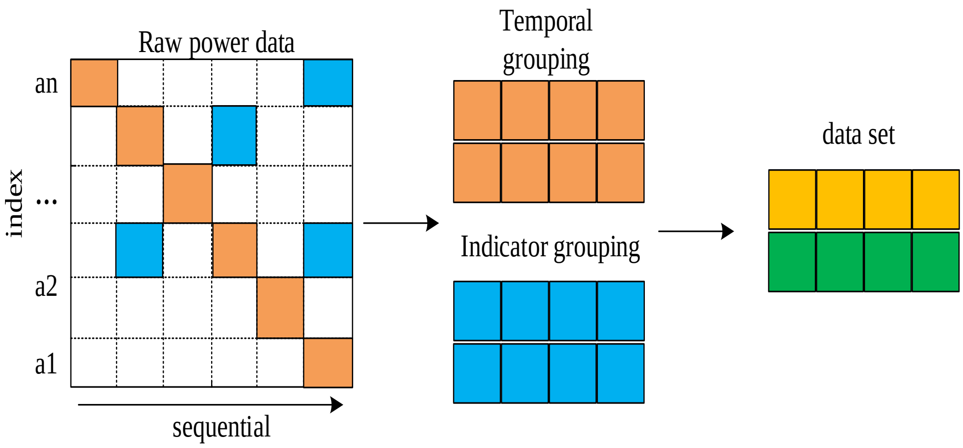

Figure 1. Ammeter data mining model

The original power data is the real data collected by the meter, with a time span of 90 days and a collection frequency of 30 minutes each time. The data contains multiple electrical parameters, including 20 electrical parameters such as three-phase voltage, current, power (including total power) and power factor (including total power factor) [9]. Since the three-phase three-wire meter does not record the electrical parameters of phase B, there are only 15 electrical parameters. The process of data modeling is displayed in Figure 2.

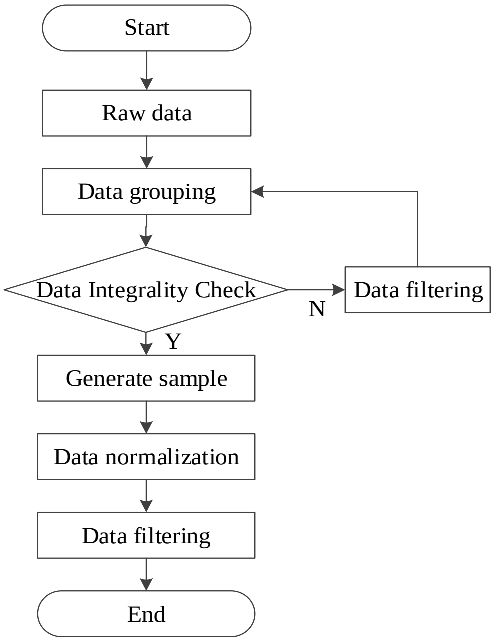

Figure 2. Data modeling process

After grouping the original data in chronological order, the power data should be denoised. For the samples with missing data, it is necessary to consider retaining or filtering directly. For those with serious missing data, it is necessary to filter directly. Then, combined with the expert rules, the abnormal indicators are added to unify and regularize the data, so that the sample dimensions of three-phase three-line and three-phase four-line are aligned. Finally, some abnormal indicators need to be normalized before generating the final power data set [10].

After the clustering is completed, a threshold needs to be set to determine the abnormal condition set S1. The model is a preliminary determination of power equipment in the overall power consumption data, and the threshold needs to be set slightly smaller, so that more suspicious objects will be identified as power equipment in the clustering process to reduce the missed detection rate. But the corresponding false detection rate will rise, that is, in the preliminary screening model of power equipment, a higher detection rate will be ensured at the expense of a certain false detection rate. The threshold issues are further explained in the case study section. After the preliminary diagnosis model of data abnormity screens the power consumption data, the cluster S1 that does not conform to the clustering result is obtained, which will be applied as the input of the analysis model of power equipment abnormity for further detection to reduce the false detection rate of the overall model [11]. The specific algorithm flow is as follows:

(1) Preprocess the original data, including the processing of large and single missing values, repeated values, and obvious error values and the standardization of the data.

(2) Determine that K value by using the double criterion of an elbow method and a profile coefficient method.

(3) Obtain a first initial clustering center v1 by calculating the median of all samples.

(4) Calculate the Manhattan distance from all sample points to the first initial cluster center v1, as shown in formula (1).

$$\lambda = \sum_{i=1}^{m} \left| x_1 - x_2 \right| \tag{1}$$

Find the maximum distance Li, and select the sample point i with the maximum distance as the second initial cluster center v2.

3. Simulation experiment

3.1 Simulation scheme

In this paper, the process of determining the important parameters in the model, including the selection of K value, the process of determining the set of outliers and the specific classification of power consumption patterns, will be described by using the real power consumption data set. In addition, the operation efficiency and detection rate of the model are tested through the comparative experiments of the model to prove the effectiveness of the proposed model in the anomaly detection of electricity consumption data sets.

In this paper, the ammeter data set is adopted as the research object, and the data structure is indicated in Table 1.

Table 1 Experimental data structure

Indicators	Parameters
Number of samples	5238
Sampling time	24h
Sampling interval	30min
Electricity consumption	152kWh
Proportion of anomalies	9%

The data set contains 5238 user data samples, including 4765 normal users and 473 abnormal users. Each node provides all the power consumption data. The data collection interval is 30 minutes, and the label of the data set can be applied to detect the model effect.

3.2 Analysis of experimental results

In this paper, K-means algorithm and K-mediods algorithm are chosen for experimental comparison with the algorithm in this paper. The results are displayed in Figure 3 from the aspects of operational efficiency and detection rate.

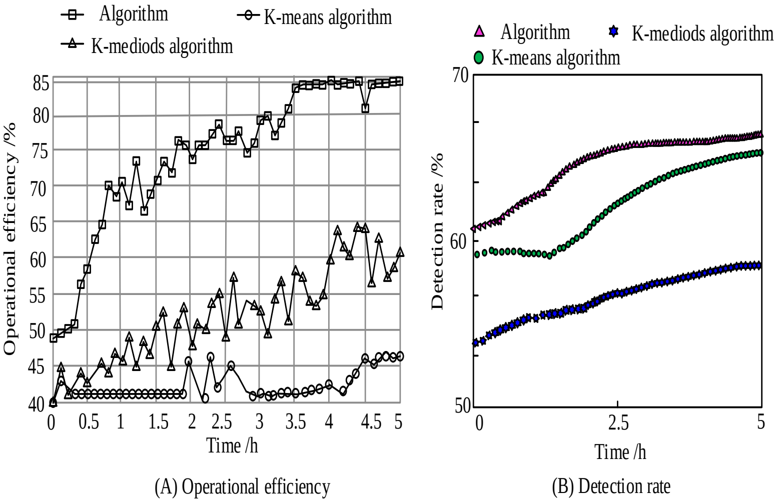

Figure 3. Comparison results of operational efficiency

In terms of detection time, the proposed algorithm is better than the traditional K-means and K-mediods algorithms in terms of operation efficiency and detection rate.

Classical machine learning algorithms and clustering algorithms alone is applied to compare the models. The steps of using machine learning algorithm to detect abnormal power consumption are roughly as follows: First, feature extraction is carried out on the preprocessed data. The way of feature extraction is based on time series, mainly including a series of statistical feature indicators such as mean and variance, rising and falling trend indicators, Fourier transform frequency domain indicators, etc. The ensemble learning algorithm is adopted as a classifier to classify normal power users and abnormal power users. The ROC curve is shown in Figure 4.

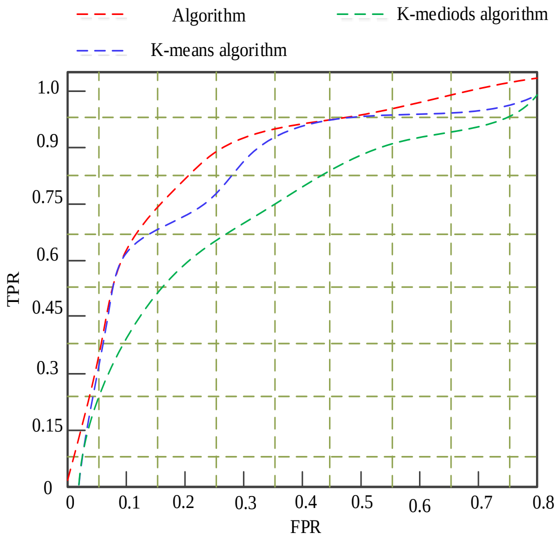

Figure 4. ROC curve test results

The AUC of the area under the ROC curve of the proposed model is higher than that of other methods, which proves the effectiveness of the proposed model.

4. Conclusion

In this paper, a preliminary diagnosis model of power equipment based on the improved K-means clustering is established, which improves the shortcomings of the traditional K-means clustering algorithm, such as difficult determination of K value and initial clustering center, low calculation efficiency, and poor performance in the case of more noise points. Secondly, according to the operation characteristics of the power equipment data, the abnormal operation modes of the equipment are classified, and the final power equipment data set is determined by the curve similarity analysis. Finally, the simulation results verify that the model can simultaneously ensure a higher detection rate and a lower false detection rate for the power equipment data set.

The samples obtained in this paper are small, among which the user sample size is scarce. The selected time is short, and the information that can be discussed and mined is limited, such as ignoring the impact of quarters and peak and valley periods. In the follow-up study, the method of the personalized confidence interval can be adopted for data sets with a small number of categories. Besides, different confidence intervals may be set for different categories, which can improve the classification accuracy of individual categories.

Acknowledgements:

This paper was supported by the Research Projects of State Grid Qinghai Electric Power company: "Research on function test and field communication technology of object-oriented protocol watt hour meter and acquisition terminal (SGYXFWJSJS2200062)".

References

[1] Oprea S V, Bra A. Machine Learning Classification Algorithms and Anomaly Detection in Conventional Meters and Tunisian Electricity Consumption Large Datasets[J]. Computers &Electrical Engineering, 2021, 94, 107329.

[2] Z. Y, H. W. Electricity Theft Detection Base on Extreme Gradient Boosting in AMI[J]. IEEE Transactions on Instrumentation and Measurement. 2021, 70: 1-9.

[3] L. C, L. G, L. G, et al. A Covert Electricity-Theft Cyber-Attack against Machine Learning-Based Detection Models[J]. IEEE Transactions on Industrial Informatics. 2021: 1.

[4] Feng Z, Huang J, Tang W H, et al. Data Mining for Abnormal Power Consumption Pattern Detection Based on Local Matrix Reconstruction[J]. International Journal of Electrical Power & Energy Systems, 2020, 123:106315.

[5] Y. G, B. F, N. Y. A Physically Inspired Data-Driven Model for Electricity Theft Detection with Smart Meter Data[J]. IEEE Transactions on Industrial Informatics. 2019, 15(9): 5076-5088.

[6] Karol S. Bagged neural networks for forecasting Polish (low) inflation[J]. International Journal of Forecasting, 2019, 35(3):1042-1059.

[7] LINGC Y, ZOU Y Z, LIN Z Q, et al. Graph embedding based API graph search and recommendation [J]. Journal of Computer Science and Technology, 2019, 34(5):993-1006.

[8] Xiangyu Kong et al. A Recursive Least Squares Method with Double-Parameter for Online Estimation of Electric Meter Errors[J]. Energies, 2019, 12(5):805.

[9] Z Zhang et al. Research on estimating method for the smart electric energy meter’s error based on parameter degradation model[J]. IOP Conference Series: Materials Science and Engineering, 2018, 366 (1).

[10] Sung F, Yang Y, et al. Learning to compare: Relation network for few-shot learning[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 1199-1208.

[11] Inoue H. Adaptive ensemble prediction for deep neural networks based on confidence level[C]. Conference on Artificial Intelligence and Statistics (AISTATS). 2019: 1284–1293.