Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.22, September 2024

Research on Electric Energy Metering Anomaly Detection and Classification Algorithm Under Multi-Source Data Fusion

Fusheng Wei[1] , Zhifeng Zhou[2], Lu Yang[3], Jun Zhu[4*], Kefei Guan[5]

[1] Guandong Power Grid. Co. Ltd.
Mailing Address, Guangzhou, Guangdong, China

[2] China Southen Power Grid Co. Ltd.
Mailing Address, Guangzhou, Guangdong, China

[3] Measurement Center of Guangdong Power Grid Co. Ltd.
Mailing Address, Guangzhou, Guangdong, China

[4] Shaoguan Power Supply Bureau of Guangdong Power Grid Co. Ltd.
Mailing Address, Shaoguan, Guangdong, China
E-mail: gdzhujun@tom.com

[5] Jiangmen Power Supply Bureau of Guangdong Power Grid Co. Ltd.
Mailing Address, Jiangmen, Guangdong, China

Abstract. The renewable energy grid is a distributed power grid, which needs to be parallel to the thermal power grid to ensure the stability of its own power generation. Therefore, renewable energy has the problem of multiple sources of power generation data and complex electric energy metering, which has always limited the development of renewable energy. In order to improve the grid-connected power supply quality of renewable energy and improve the accuracy of electric energy metering, this paper proposes a binary classification algorithm to collect, amplify and standardize multi-source data. Then, the Fourier function is used to normalize the multi-source data to shorten the differences between different devices. In particular, the attribute features are classified and the input data adjustment parameters are set. Finally, renewable energy's electric energy data value is measured, and the abnormal data characteristics are obtained through the second-order derivative to realize data mining and electric energy metering anomaly detection. The simulation results show that the binary classification can complete the metering and anomaly detection of electric energy in the distributed power grid, and the classification processing can simplify the complexity of multi-source data, improve the accuracy of metering anomaly detection, and make its dispersion more reasonable, and the results are between 90~95%, which is better than the online monitoring method. The time for anomaly result detection is 25s, which is shorter than the previous algorithm, and cross-domain data search is realized, with a cross-domain rate of 50~60%. Therefore, the classification algorithm proposed in this paper can meet the requirements of electric energy metering under multi-source data fusion and achieve rapid anomaly detection.

Key words. Multi-Source Fusion Data, Electric Energy Metering, Anomaly Detection, Data Mining.

1. Introduction

Electric energy metering is affected by the change of relevant data in the power grid system, which is the guarantee of the stable operation of the power grid system, and it is necessary to accurately identify abnormal energy anomaly data [1]. However, in the actual operation of the power system, due to the collected energy data, there will be some abnormal data, which will affect the electric energy metering, and in the long run, it will have an adverse impact on the production and power generation of the power system, renewable energy belongs to the grid-connected system, and its transmission is unstable, affected by many factors such as light, wind, season and national policies [2], so the probability of abnormal electric energy metering is higher Long-term electric energy metering of renewable energy will not only increase the cost of power generation but also increase the complexity of metering Therefore, finding an effective electric energy metering and detection method is an urgent problem to be solved [3], in addition, wind power generation and photovoltaic power generation have the characteristics of distribution, and their data are mainly multi-source data, and the data span is large, which has an impact on the detection of electric energy metering anomalies and reduces their detection accuracy. Some scholars believe that multi-source data can be detected by conversion methods and qualitative methods [4], stripping away their original multivariate attributes, and reducing the differences between data by expanding and shrinking functions, but this method will affect the overall structure of the data and the enlightening effect of the data. Some scholars also believe that the anomaly detection of electric energy metering has the characteristics of large data volume and complex data, and its measurement results fluctuate greatly and the calculation time is relatively long [5].In the process of in-depth study of the detection method of abnormal data of electric energy metering, through the anomaly detection framework of electric energy metering data detection, the relevant data of electric energy metering are obtained, and the deep network is used to aggregate and compare and label the abnormal data of electric energy metering, so as to complete the detection of electric energy metering data under multi-source data fusion. Therefore, in the process of electric energy metering anomaly detection based on data analysis, some scholars analyze the likelihood value of the maximum electric energy measurement data through big data analysis [6], expand the data feature extraction of electric energy metering data under multi-source data fusion, and complete the detection and classification of electric energy metering anomalies through operations such as transition probability sequences.On this basis, this paper proposes a binary classification method to deal with the difference of multi-source data, simplify the data and process of distributed energy calculation, and improve the identification accuracy and efficiency of abnormal metering. Firstly, the electric energy metering data is detected and classified with reference to the relevant references, and the difference analysis of the electric energy metering data is focused on under the anomaly of the electric energy metering data fused with multi-source data. Then, the abnormal data samples were identified, and the false alarm parameters of the electric energy metering anomaly detection were added to ensure the stability of the electric energy metering anomaly detection. Finally, the relevant data were mined, and the binary classification calculation was carried out to find out the shortcomings of anomaly detection in different methods compared with traditional electric energy metering. To sum up, this paper proposes that the binary classification algorithm is based on data mining for electric energy metering detection, which aims to improve the accuracy of electric energy metering anomaly detection and classification, and meet the needs of power system anomaly detection.

2. Mathematical Description of Anomaly Detection in Electric Energy Metering

A. Multi-Source Data Attribute Separation

Grid anomalies in renewable energy can lead to energy metering differences, so it is important to describe the volatility and variability of energy metering data and remove other attributes. In this paper, the power data is expanded to improve the depth of data mining through the feature classification method of data mining, and the feature clustering of abnormal data is carried out. At the same time, it is necessary to expand the classification of data features and construct a collection of abnormal data of electric energy metering $A=\left(a_1,a_2,\cdots,a_m\right)$, a_j includes samples of abnormal data of electric energy metering and a $d_i(t)$ collection of abnormal data of electric energy metering. The weight of the anomaly data sample is N , and the weight of the anomaly data cluster is M . Under multi-source data fusion, an abnormal data clustering center is set for the sample $d_i'(t-1)$. Hypothesis 1: Under multi-source data fusion, the feature set of the abnormal data of electric energy metering, $y_i(x)$ and the characteristic set is described as $d_i'(1)$, then the abnormal data of electric energy metering is expressed as E_i :

$$E_i=\sum_{i=1}^{H_i}\exp\left\{-\frac{\left\|a_j-N\right\|^2}{\left(y_i(x)_a/2\right)^2}\right\}\left|d_i(t)\right. \tag{1}$$

In equation (1), $\exp\{\cdot\}$ it is the difference function of multi-source data, which can realize the unification of data. Under multi-source data fusion, the radius of the electric energy metering interval is , s_i and the maximum density in the interval is $\rho(s_i)$, then the anomaly density index is set to ind_i . E_j is an anomalous data property that describes energy metering. Under the initial multi-source data fusion, the clustering center of the abnormal energy metering data is a_i , and the calculation of the characteristics of the abnormal data is shown in equation (2).

$$E_j=E_k\cdot\exp\left\{-\frac{\left\|a_j-a_i\right\|^2}{\left(s_i/2\right)^2}\right\}\cdot ind_i \tag{2}$$

In formula (2), Under the multi-source data fusion, $\frac{E_{k+1}}{E_1}$ is the abnormal data features of electric energy metering can be obtained. The smaller the value is, the better the classification effect of the abnormal data features of electric energy metering is $\frac{E_{k+1}}{E_1}$.

B. Adjustment of the Proportion of Electricity Generated by Renewable Energy

Renewable energy generation is not stable, so it is necessary to combine it with thermal power generation to ensure the continuity of power generation. However, the adjustment of the proportion of renewable energy power generation is directly related to the stability of power generation, so it is necessary to set the adjustment threshold to reduce the impact of distributed generation instability on the computing structure, in other words, the adjustment of the proportion of power generation has an impact on the clustering quality of abnormal data characteristics of electric energy metering, so the high-order statistical model is used to compensate the electric energy metering parameters, and the abnormal data feature extraction and optimization algorithm is designed to realize the classification of abnormal data characteristics of electric energy metering.The set $R_g=\left\{r_{g1},r_{g2},\cdots,r_{gb}\right\}$ of electric energy metering abnormal data is assumed, the optimal set of individual feature classification of electric energy metering abnormal data is $N=\left\{n_1,n_2,\cdots,n_n\right\}$, and the global optimal set of feature classification of electric energy metering abnormal data is $R_j=\left\{r_{j1},r_{j2},\cdots,r_{jb}\right\}$. At this case, the feature component update scheme of electric energy metering abnormal data under multi-source data fusion is expressed as:

$$y_{jb}^{(t+1)} = y_{jb}^{t} + F_i z_i (r_{jb}^{t} - y_{jb}^{t}) + F_{i+1} z_{i+1} (r_{gb}^{t} - y_{gb}^{t}) \tag{3}$$

In formula (3), the data point in the first node z_i in the multi-source data fusion is j, the acceleration coefficient of the data feature extraction optimization algorithm is b, and the random value y_{jb}^{t} in the data feature F_i extraction optimization algorithm. Solve the distribution cluster maximum of electric energy metering abnormal data characteristics under multi-source data fusion, expressed as:

$$g_{\max} = \left| \max \left[g_{ji}(t) \right] \right| \tag{4}$$

In formula (4), the average particle size $g_{\max}$ of electric energy metering abnormal data features, $\max\left[g_{ji}(t)\right]$ is a function of the maximum, under multi-source data fusion is solved and expressed as:

$$g_{mid} = \left| \sum_{i=1}^{n} \sum_{j=1}^{b} g_{ji}(t) \right| \tag{5}$$

In formula (4) and formula (5), the distribution cluster i in the first sampling point dimension of electric energy metering abnormal data is j, the dimension of electric energy metering abnormal data is $g_{ji}(t)$, and the total number of abnormal samples of electric energy metering data under the fusion of multi-source data is b. Assuming, the characteristic clustering of energy metering anomaly data of higher-order statistics v is expressed as:

$$v = \frac{\left| g_{\max} - g_{mid} \right|}{g_{\max}} \tag{6}$$

The cycle iteration of the value can complete the optimization selection of abnormal data characteristics of electric energy measurement under multi-source data fusion a_j, obtain the abnormal data characteristics of electric energy metering data, and complete the feature extraction of abnormal electric energy metering data under multi-source data fusion. Through the cyclic iterative calculation of the power generation ratio, the instability of the distributed power grid can be greatly eliminated, and the impact of multi-source data fusion on the results can also be verified. When adjusting the proportion, it is necessary to obtain the initial value of electric energy metering data based on the previous power generation data and analyze and extract the characteristics of abnormal data.

C. Search for Electric Energy Anomaly Detection Points

There are two assumptions in the anomaly search: first, the anomaly data is on different devices q_j and is constrained by multi-source data; Second, the anomalies are located in different locations α_j, and the different points are far apart [7]. In this case, for the analysis of abnormal energy metering data under the constraint of multi-source data q_j, two calculation parameters need to be adopted, namely local anomaly coefficient e_c and distance anomaly coefficient y_j. The calculation of the local anomaly coefficient is closely related to the distance anomaly coefficient, so the conversion coefficient needs to be added to the anomaly data, which is calculated as follows:

$$q_j = \frac{\sum_i \beta \cdot (e_{ji} - e_c)}{y_j} \tag{6}$$

In equation (6), under the multi-source data fusion, the distance between the sample points of the abnormal electric energy metering data y_j can describe the number of sampling points of the abnormal electric energy metering data e_{ji} and the stage distance between the β abnormal electric energy metering data set Y and the parameter parameters α_j.

$$\alpha_j = \min_{q_i \succ q_j} e_{ji} \tag{7}$$

In Eq. (7), it means that multi-source data is comprehensively calculated, and the range is $\alpha_j = \max e_{ji}$. However, datasets have the highest frequency of anomalies. In this case, the abnormal probability of energy metering is related to the degree of distribution of renewable energy, so it is necessary to enter the point where the renewable energy is located and conduct a local search based on it. The local density of renewable energy is small, and the distance is significant to the anomaly, which is the focus of the search Y. Under the constraint of multi-source data, the sampling map of the electric energy metering data was obtained, and the two-dimensional plane search map was made to preliminarily determine the range of anomalies $e(y_j, y_i)$. Then, an iterative search is performed in the possible ranges $M_h(y_j)$ to determine the dataset's clustering centre. Judging from the detection effect of abnormal data of electric energy metering, the maximum and minimum points of abnormal points are determined, and the range search is carried out to obtain the abnormal results of electric energy metering finally. The difference between random data samples and samples is sorted and eliminated in the original energy metering data set. The formula for calculating search points is as follows.

$$M(y_j) = \left\{ i \in Y \left| e(y_j, y_i) \le e(y_j, M_h(y_i)) \right. \right\} \tag{8}$$

The abnormal points within the search range q_j^{scop} are judged as follows.

$$q_j^{scop} = \min[y_j - \left(\frac{1}{H} \cdot \sum_{y_i} e(y_i)^2 \right)] \tag{9}$$

Equation (9) represents the anomaly search of electric energy metering under multi-source data fusion $e(\cdot)$, and obtains the local anomaly comparison value according to the probabilistic method, and gradually expands the search range, as shown in Equation (10).

$$\alpha_j^{scop} = \max e_{ji} \tag{10}$$

Suppose the electric energy measurement data sample under multi-source data fusion is obtained through the calculation of formula (9) and formula (10). In that case, it is judged as the abnormal data in the electric energy measurement data set under multi-source data fusion, so that the abnormal detection and classification algorithm of electric energy measurement under multi-source data fusion is completed. Suppose the abnormal data samples of electric energy metering under the constraints of multi-source data are obtained through the calculation of equations (9) and (10), and the search range is expanded. In that case, the final multi-source data anomaly set is obtained e_{ji} to complete the classification and calculation of outliers.

3. Experimental Simulation and Analysis

A. Case Introduction

Taking a certain station area as an example, the power supply voltage is 110Kv, and the mixed power generation of photovoltaic, wind and thermal power is adopted, the power supply users are 10,000, and the electric energy meter is 560 sets, and the sub-meter nesting method is adopted. The renewable energy power generation equipment has a number and GIS signal positioning, and the power data management system monitors the whole station area. This paper's calculation method has been added to the submodule. There is no regional isolation between different stations, and there is no isolation between wind power and thermal power, and all equipment has a management system for centralized processing and data collection, and the electric energy metering range in the station area is shown in Figure 1.

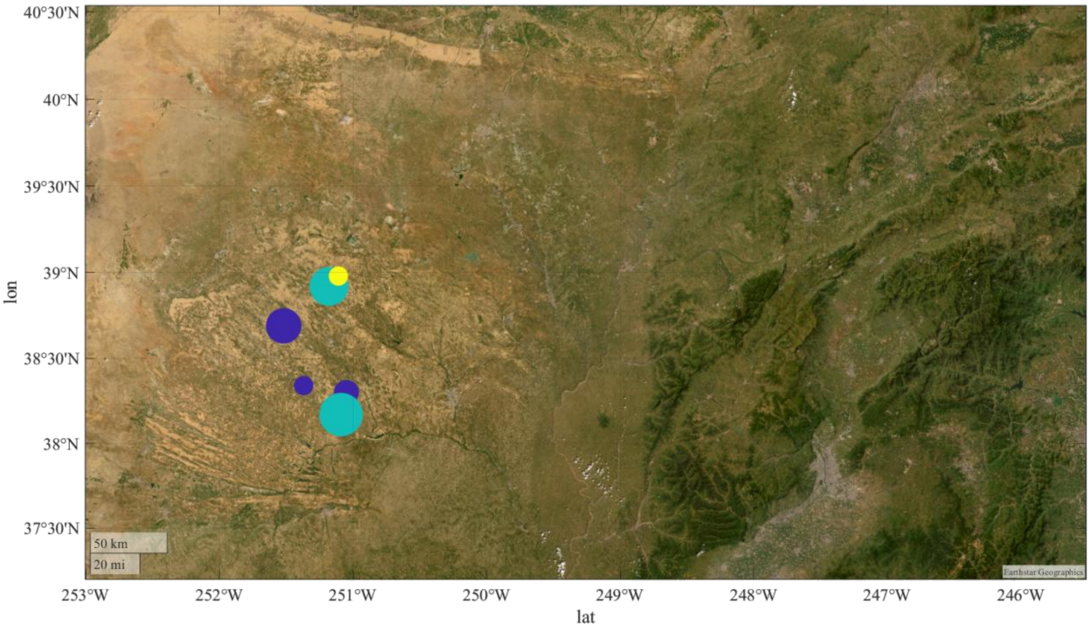


Fig. 1 The Power Supply Range of Photovoltaic, Wind and Thermal Power Mixture

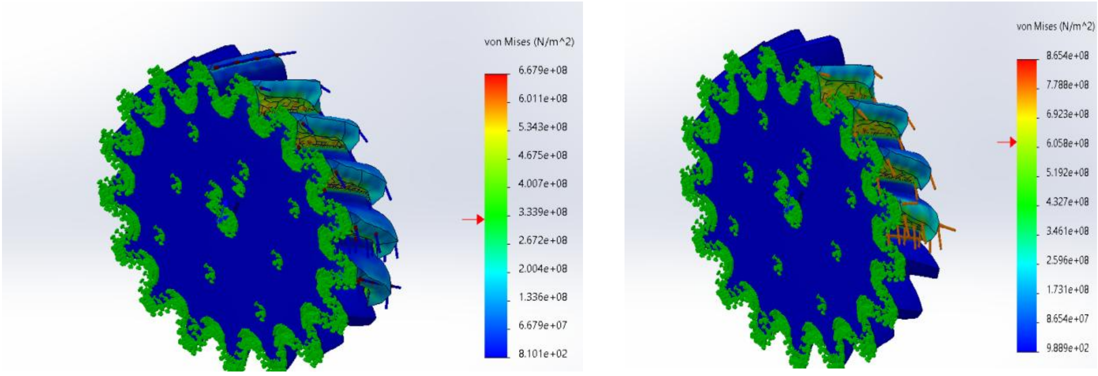
B. Abnormal Device Detection

The anomaly detection of electric energy metering under the constraint of source data is used to locate the abnormal equipment, which verifies the effectiveness of the proposed method. In the process of positioning, the abnormal data is mainly analyzed based on the GIS signal of the equipment, and the multi-source data analysis across the station area is realized. The detection results of abnormal electric energy measurement are shown in Table 1.

Table 1. Measurement Anomaly Detection Results of Multi-Source Data

Desk area	Billing	Aging failures	Other failures
1	95	95	Mutual sensor failure
2	154	154	Terminal failure
3	178	178	transformer fault
4	217	217	Table meter failure
5	320	320	Mutual sensor failure

The analysis in Table 1 shows that the proposed data mining algorithm can effectively realize the multi-source data fusion under photovoltaic, geothermal and wind power generation and detect the anomaly of electric energy metering at the user end. If the detected abnormal data is consistent with the amount of historical abnormal data, it is necessary to search for abnormal points in the later stage and find the cause of abnormal data of electric energy metering in combination with the dichotomy method to meet better the electric energy billing needs of the power system. Among them, the aging detection results of rotor equipment in wind power generation are shown in the figure below.



The Binary Classification Method Identifies the Parts of The Device

Faulty Parts Identified by Conventional Methods

Fig. 2 Faulty Parts Identified by Different Methods

As seen from Figure 3, the dichotomy method can identify the degree of alienation of electric energy metering, which is a theoretical method to reflect the anomaly detection of multi-source data, and the higher the degree of alienation, the better the detection effect of abnormal electric energy metering. For cases with a low degree of alienation and similar causes of abnormalities, this paper uses data mining in the binary classification method to achieve an in-depth comparison of homologous indicators, as shown in Figure 3.

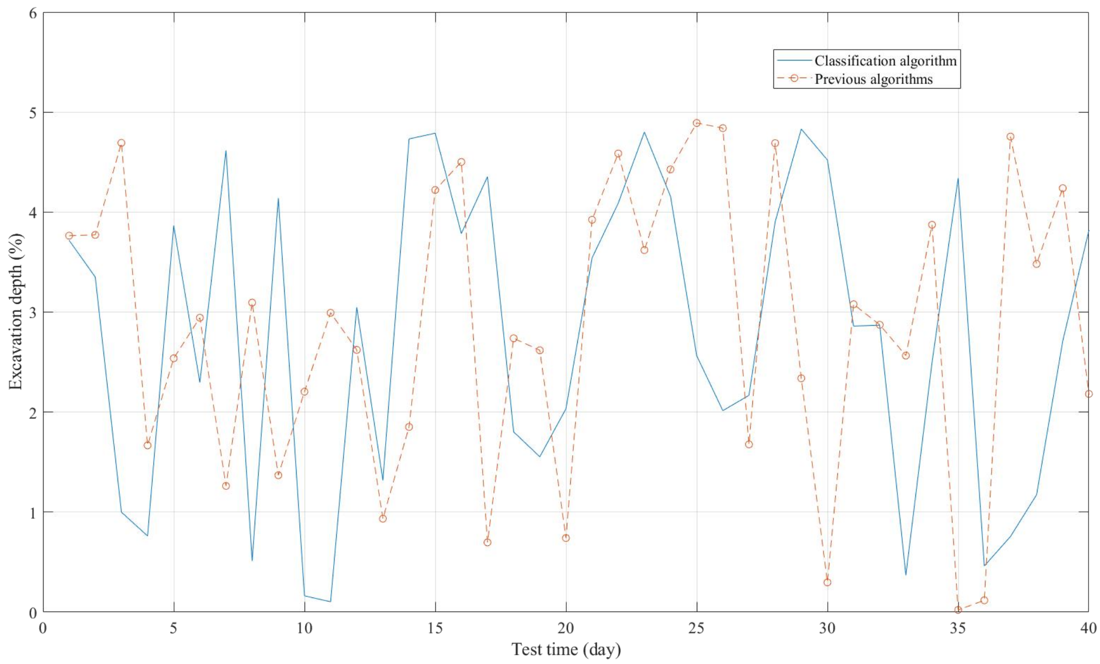


Fig. 3 Data Mining due to the Same Measurement Exception

From Figure 3, it can be seen from the analysis of the mining curve that with the gradual increase of the search for the abnormal range of electric energy metering, the mining depth of the data used shows a downward trend, but the downward trend is small, while the identification rate of abnormal electric energy metering increases. This is mainly due to the increase of uncertainty factors when mining the same index, and the anomaly detection method in this paper increases the random coefficient to reduce the interference rate between different dimensions and then improves the effectiveness of identification. In addition, in the case of a large amount of electric energy metering data and large disturbance of photovoltaic power generation equipment, the classification test algorithm can reduce the interference and improve the detection rate of electric energy metering anomalies.

C. Accuracy of Identification of Abnormal Energy Metering

The multi-source data test is faced with the characteristics of complex data, large data volume and diverse data structures, so it is necessary to judge its recognition accuracy, and compare it with the previous algorithms, and the calculation results in this paper are as follows.

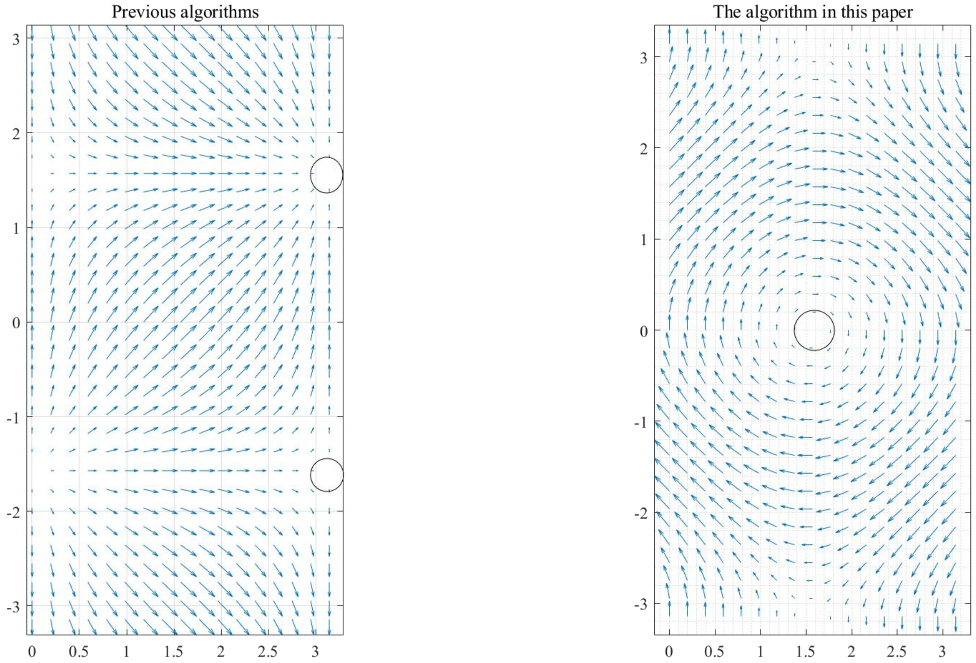


Fig. 4 Comparison of Detection Accuracy of Electric Energy Metering Anomalies Under Different Calculation Results

By comparing the accuracy of different methods in Figure 4 for the anomaly detection of electric energy metering, it can be seen that the accuracy of the data mining algorithm proposed in this paper for the anomaly detection of electric energy metering is better than that of the other two literature methods, and the accuracy of anomaly detection of electric energy metering has been maintained at more than 95%. However, the overall accuracy of energy metering anomaly detection in this paper does not change much, which remains between 80% and 90%, indicating that the accuracy and stability of the proposed method for energy metering anomaly detection are better than those of other methods. However, there are two directions to accuracy, the accuracy of different data sources and the accuracy of the same data source. Therefore, it is necessary to compare the same data source under the constraint of multi-source data. The results of the comparison are shown in Figure 5.

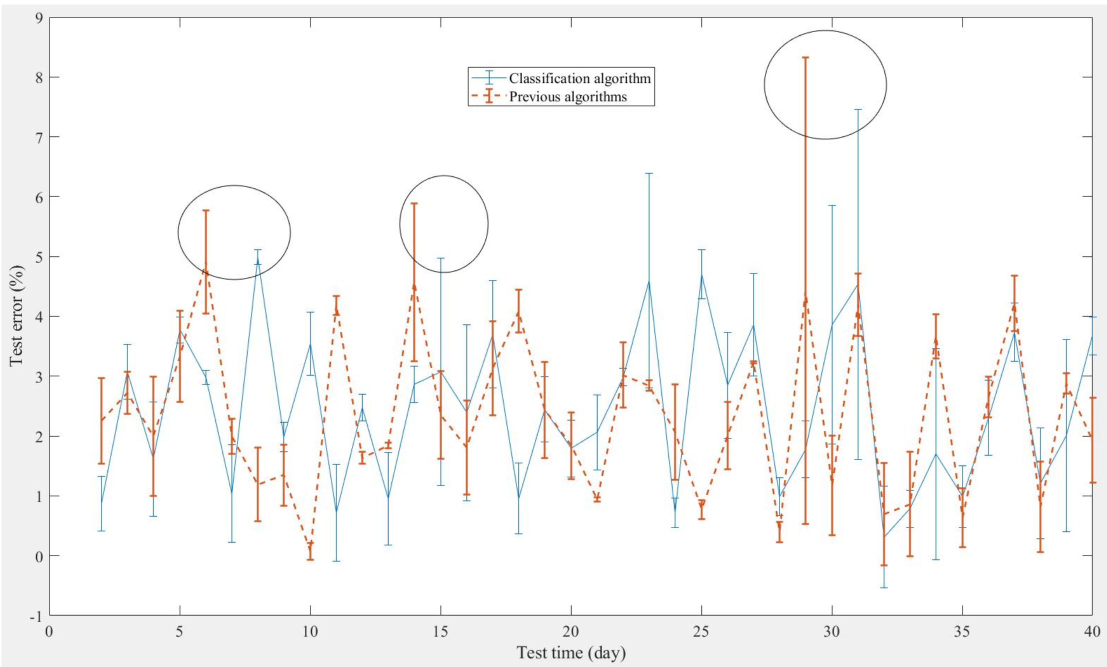


Fig. 5 Comparison of False Positive Rates of Unified Data Sources

The above analysis shows that the test results of the taxonomy in this paper for the same data source are accurate, and the false positive rate is relatively low. In 1~5 months, the overall false alarm rate is about 5.2%, and the false alarm rate of the previous algorithm for the anomaly detection of electric energy measurement has been kept at about 10%, so the false positive rate of the proposed method is better than that of other methods, therefore, the use of classification algorithm for abnormal detection of electric energy measurement meets the needs of the overall electric energy measurement anomaly detection, and realizes the accurate identification of multiple sources and homologous.

D. Detection Time of Abnormal Energy Metering

The detection time of electric energy metering anomaly is an important content, not only the occupation of system resources, but also the detection of data structure, so the detection time is an important evaluation index, but the detection time is divided into absolute detection time and relative detection time, as shown in Table 2.

Table 2. Comparison of Detection Time of Different Anomaly Detection Methods

Method	**Detect Content**	**Relative Detection Time**	**Absolute Detection Time**
Methods of this Article	Equipment	6.14	15.89
	Distance	7.58	12.24
Previous Methods	Equipment	8.92	17.54
	Distance	14.74	18.57
Average Standard	Device + Distance	9.42	14.23

As can be seen from Table 2, The method proposed in this paper is used in the previous algorithms for the abnormal detection of electric energy metering; the non-detection time and the relative detection time are used in the previous algorithms, and they are lower than the average standard, so it shows that the binary classification method can better detect the data. The reason is mainly because the binary classification method can eliminate the redundant data and realize the initial analysis of the data, and secondly, the complexity of the data in the analysis process, such as photovoltaic power generation, the data affected by night, day and weather and other factors are comprehensively judged. The binary classification method can solve the fusion between the power grid distributed data and thermal power data, maintain the stability between the two, and identify the difference between the two by zooming in and out of the function.Overall, the detection results presented in this paper are small, indicating that the abnormal detection effect of electric energy metering is better.

4. Conclusion

Photovoltaic power generation and hydropower generation are the main forms of renewable energy power generation, which are combined with traditional thermal power generation into a composite grid-connected structure, between which the user is provided with continuous current, but photovoltaic power generation and thermal power generation have the problem of unstable power generation, which is disturbed by weather and other factors, so it is necessary to identify the stability of power generation of the two, to judge the abnormal power metering, in addition, there are also problems of equipment ageing and line depreciation in the system management between renewable energy and thermal power generation, which is also another factor that leads to electric energy metering.Therefore, this paper analyzes the multi-source data of different stations by binary classification method and eliminates the interfering factors of photovoltaic power generation to improve the accuracy of electric energy metering and detection. The results show that the overall detection accuracy rate reaches 90%, and the results of electric energy detection are consistent under the constraints of homologous and multi-source data, which are higher than 90%. The binary classification method can remove the differences between different stations, shorten the complexity of the data, improve the effect of its analysis, reduce the detection time, and control it between 6~9 seconds, which has a good advantage. There are also some shortcomings in this study, mainly because the power grid belongs to the confidentiality industry, and it is difficult to obtain in-depth data, so there is a primitive deviation in the accuracy of electric energy metering anomalies, and the future will focus on data search and collection to better meet the needs of metering difference detection in the power system.

References

[1] X. Ge *et al.*, “Spatio-temporal knowledge graph based forest fire prediction with multi source heterogeneous data,” *Remote Sensing*, vol. 14, no. 14, p. 3496, 2022.

[2] Z. He, H. Shao, J. Cheng, Y. Yang, and J. Xiang, “Kernel flexible and displaceable convex hull based tensor machine for gearbox fault intelligent diagnosis with multi-source signals,” *Measurement*, vol. 163, p. 107965, 2020.

[3] Y. Huang, T. Li, C. Luo, H. Fujita, S. J. Horng, and B. Wang, “Dynamic maintenance of rough approximations in multi-source hybrid information systems,” *Information Sciences*, vol. 530, pp. 108-127, 2020.

[4] B. Li, D. Pi, Y. Lin, I. A. Khan, and L. Cui, “Multi-source information fusion based heterogeneous network embedding,” *Information Sciences*, vol. 534, pp. 53-71, 2020.

[5] S. Lin *et al.*, “MDF-SA-DDI: predicting drug–drug interaction events based on multi-source drug fusion, multi-source feature fusion and transformer self-attention mechanism,” *Briefings in Bioinformatics*, vol. 23, no. 1, p. bbab421, 2022.

[6] X. Liu, L. Jiao, L. Li, X. Tang, and Y. Guo, "Deep multi-level fusion network for multi-source image pixel-wise classification," *Knowledge-Based Systems*, vol. 221, p. 106921, 2021.
[7] W. Xu, Y. Pan, X. Chen, W. Ding, and Y. Qian, "A novel dynamic fusion approach using information entropy for interval-valued ordered datasets," *IEEE Transactions on Big Data*, vol. 9, no. 3, pp. 845-859, 2022.
[8] L. Yang, X. Zhang, W. Xu, and B. Sang, "Multi-granulation rough sets and uncertainty measurement for multi-source fuzzy information system," I*nternational Journal of Fuzzy Systems*, vol. 21, pp. 1919-1937, 2019.
[9] J. Zhang, C. Li, Z. Sun, Z. Luo, C. Zhou, and S. Li, "Towards a unified multi-source-based optimization framework for multi-label learning," *Applied Soft Computing*, vol. 76, pp.425-435, 2019.
[10] K. Zhang, T. Gao, and H. Shi, "Bearing fault diagnosis method based on multi-source heterogeneous information fusion," *Measurement Science and Technology*, vol. 33, no. 7, p. 075901, 2022.
[11] P. Zhang *et al*., "A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data," *Information Fusion*, vol. 80, pp. 87-103, 2022.
[12] Y. Zhang, Q. Li, W. Tu, K. Mai, Y. Yao, and Y. Chen, "Functional urban land use recognition integrating multi-source geospatial data and cross-correlations," *Computers, Environment and Urban Systems*, vol. 78, p. 101374, 2019.