Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 5, pp. 3342–3354, DOI 10.12694/scpe.v25i5.3088

ANALYSIS OF ABNORMAL FREEZING DATA AND UPDATING ALGORITHM FOR ELECTROMECHANICAL ENERGY METER TERMINALS

SHUZHI ZHAO*, YUE DU†, SHANSHAN HE‡, JIAO BIAN§ AND JIABO SHI¶

Abstract. Due to the rapid development of science and information technology, electricity information acquisition system has been widely used in the electricity data collection of users. With the massive electricity data collection, it is difficult to adopt traditional data processing methods to meet the abnormal data processing. In order to effectively mine abnormal information in electricity consumption data, an anomaly detection model based on the Isolation Forest (iForest) algorithm is proposed. Firstly, the daily load curve with strong regularity is used as the characteristic index of anomaly monitoring, and the users with abnormal electricity consumption data are preliminarily screened. Secondly, on the basis of electrical variables, the suspected abnormal users are further analyzed, and the anomaly identification model of electricity data is established to automatically classify the voltage at the metering point. Moreover, combined with the current data, the abnormality of the electric energy metering device is identified, and then the validity of the model is determined through on-site verification. Finally, according to the participating voltage of the fault phase, the 96-point voltage data frozen during the failure period is analyzed and the correction coefficient is adjusted. The results reveal that the electricity data detection model based on the iForest algorithm has significant advantages in computational efficiency. Through the cumulative recall and Precision-Recall (P-R) curves of the model, it is found that the majority of abnormal users can be detected only by detecting a few users with high abnormal scores, which shows that the model has high efficiency. The decision tree algorithm combined with the current data can effectively identify the anomalies of the energy metering device, which verifies the validity of the anomaly identification model of the electricity consumption data.

Key words: Energy meter, Electricity information acquisition system, Abnormal electricity consumption data, Isolated Forest algorithm, Decision Tree algorithm

1. Introduction.

As one of the most important technical bases in the power system, the electric energy metering technology is constantly improving with the improvement of technical and service levels [1,2]. As the basic equipment in the power automation system, electric energy metering plays a vital role in the power system, mainly responsible for data reading, processing, and storage. The application of big data technology promotes the integrity and accuracy of power-metering terminal data . The electric energy metering technology is mainly used in the electricity information acquisition system (hereinafter referred to as the "EIAS"), which is the basic platform for the collection of users' electric power information and can acquire and real-time monitor the power information of all power consumers. The measurement anomaly detection function is principally to monitor and analyze the collected data in real-time, and report the abnormal power, load, voltage, and current of electricity users to the EIAS [3,4].

At present, there are more and more studies on the abnormality of terminal data of energy meters in the power system. Hasan et al. (2019) proposed a power theft detection system based on the Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) architecture. An LSTM model based on CNN is implemented for data classification in a smart grid. The results denoted that the proposed scheme can better classify most ordinary users and a few users with abnormal electricity consumption data [5]. Ji et al. (2021)

*Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd., Tangshan, Hebei, 063000, China. (Corresponding author's e-mail:ShuzhiZhao7@163.com)
†Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd., Tangshan, Hebei, 063000, China.(YueDu13@126.com)
‡Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd., Tangshan, Hebei, 063000, China.(ShanshanHe9@163.com)
§Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd., Tangshan, Hebei, 063000, China.(JiaoBian19@126.com)
¶Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd., Tangshan, Hebei, 063000, China.(JiaboShi52@163.com)

Table 2.1: Data field description of EIAS

Name	Implication
ID	Meter number
DATA_DATE	Data acquisition time
DATA_TYPE	Power type
ORG_NO	Number of the power supply company
DATA_WHOLE_FLAG	Data collection success identification
R1-R96	96-point load data
CONS_ID	User ID
CONS_NAME	User name
CONS_SORT_CODE	User type
ELEC_ADDR	User address
METER_ID	Number of measuring points
ASSET_NO	Asset number of electricity meter

put forward an estimation method for real-time robust auxiliary state for power system prediction based on the Bayesian framework, deep learning, and Gaussian mixture model (GMM). By combining anomaly detection technology in machine learning (ML) with GMM, abnormal data in measurement information can be accurately determined and deleted. Numerical simulations on IEEE 118-node and IEEE 300-node test systems show that the proposed method has high accuracy and robustness [6]. Liu et al. (2021) raised a framework based on general data mining, which can extract typical power load patterns and discover the insightful information hidden in the patterns. The proposed framework was applied to analyze the time series electricity data of three practical office buildings in Chongqing, and its validity was verified [7]. Yan and Wen (2021) proposed a power theft detector using metering data based on extreme gradient lift. Preprocessing of measurement data, including recovery of missing and wrong values and normalization. Compared with 8 ML methods such as support vector machine and Decision Tree (DT), the proposed method can detect power theft with higher accuracy or a lower false alarm rate. Experimental results also manifested that the proposed method was robust when the data was unbalanced [8].

To sum up, the current anomaly detection algorithm for the energy meter's power system mainly constructs a normal data model and identifies data inconsistent with the model as abnormal data, which leads to excessive redundancy of information and low computational efficiency. At the same time, it can be found that mining abnormal electricity consumption data is helpful to improve the efficiency of power enterprises. To improve computing efficiency, an anomaly detection model of electricity data is established based on the Isolation Forest (iForest) algorithm, and users suspected to be abnormal are identified through preliminary screening combined with a daily load curve. Then, according to the electrical variable of the abnormal user, the voltage is accurately classified by the DT algorithm and determined by combining the current data, identifying the abnormal electric energy metering device. Ultimately, the calculation method of the correction coefficient of electric quantity is put forward.

2. Abnormal Detection of Electricity Consumption Data Based on the iForest Algorithm.

2.1. The data basis of the model.

2.1.1. Choice of information fields for electricity consumption data. Abnormal electricity consumption data shows high line loss, large loss of electricity sold, etc. The above phenomena are mainly caused by line and equipment quality, management loopholes, energy meter quality, and abnormal consumption behavior. Since the daily load curve is characterized by strong regularity and a more obvious shape, anomalies in electro-data can be found more easily. Therefore, the daily load curve is used to analyze electric power data [9]. The fields of the initial electricity load data set selected from the EIAS are exhibited in Table 2.1:

The data set used above retains the meter number, user type, power type, 96-point load data, and other fields [10]. The user type is three non-resident users, and the power type is positively active. Now, the main

Table 2.2: Types of dirty data

Types	Missing value	Duplicate value	Minimax	Load burr	Impact load
Presentation	The table of data has NA or blank	The data of the user's electricity load appears to be repeated at some point	Too large or too small power load data	Data increases or decreases suddenly between adjacent time periods	The meter reads down over a continuous period of time

special transformer terminals used in the system are the 96 protocol and the 05 protocol. The 96 protocol can only be used to collect the voltage, current, and electricity data of the user's energy meter at zero. The special transformer terminal of the 05 protocol collects the voltage, current, and energy data of the user's energy meter every 15 minutes, with a total of 96 points in 24 hours. To facilitate data description, the 96-point load data are reduced to 24 points.

2.1.2. Data cleaning. In the process of dirty data processing, the types of dirty data should be analyzed and summarized first, and then targeted processing should be carried out according to its manifestation. After obtaining the power load data of industrial users, dirty data is identified by data specification principles. Common types of dirty data are outlined in Table 2.2:

In the processing of dirty data, firstly, the redundant data in the data set should be deleted. In a data set, the customer name and time uniquely determine a data record. If multiple records are the same, the redundant data needs to be deleted. Secondly, it is necessary to maintain the integrity of the data set. The problem of missing data in the data set must be properly handled according to the current situation. Every user must have the electricity reading data for every hour per day. If there is a small amount of missing data, the severity of the missing should be analyzed. The missing severity is as follows:

1. The curve is missing 20% of its reading points;
2. The curve continuously misses more than 2 consecutive readings.

If the data missing reaches the above two conditions, the user will be excluded from the research range, and the remaining load curves containing missing data will be repaired by the multi-stage Lagrange interpolation method. The repair of the missing value of the load curve is written as Eqaution 2.1:

$$P_t = \frac{\sum_{k=1}^{m_1} P_{t-k} + \sum_{i=1}^{m_2} P_{t+i}}{m_1 + m_2} \tag{2.1}$$

m_1 and m_2 refer to the number of forward periods and backward periods, and t stands for the time when the load data is missing. After data cleaning, X is recorded as $(n-\lambda) \times 24$th order effective load curve matrix composed of $(n-\lambda)$ effective daily load curves.

2.1.3. Data dimension reduction. Electricity load data is easily affected by many factors such as income, price policy, and temperature. The results caused by these influences cannot be fully reflected through the distance, and the similarity of the shape or contour of the time series cannot be fully guaranteed. In order to fully reflect the similarity between loads while taking into account the operational efficiency, six commonly used daily load characteristic indexes are selected to comprehensively reflect the electricity consumption characteristics [11].

All-day load rate reflects all-day load variation:

$$a_1 = \frac{P_{av}}{P_{max}} \tag{2.2}$$

The maximum hourly utilization rate of the whole day reflects the time utilization efficiency:

$$a_2 = \frac{P_{sum}}{24P_{max}} \tag{2.3}$$

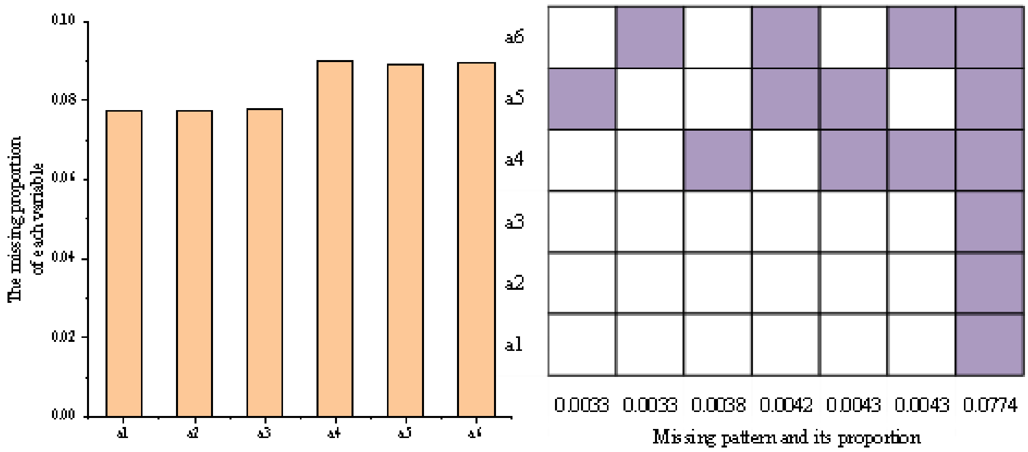


Fig. 2.1: Data missing pattern after dimensionality reduction of the daily load curve

The daily peak-valley difference throughout the day reflects the capacity of the peak regulating of the power grid:

$$a_3 = \frac{P_{max} - P_{min}}{P_{max}} \tag{2.4}$$

Peak load rate reflects peak load variation:

$$a_4 = \frac{P_{av.peak}}{P_{av}} \tag{2.5}$$

Normal load rate reflects the change of normal load:

$$a_5 = \frac{P_{av.sh}}{P_{av}} \tag{2.6}$$

The load rate in the valley period reflects the load change in the valley period:

$$a_6 = \frac{P_{av.val}}{P_{av}} \tag{2.7}$$

By using the load characteristic index to reduce the characteristic dimension of the effective load curve matrix, the $(n\lambda) \times 6$th order characteristic dimension reduction matrix is obtained, which is recorded as Y. Through visualization processing of the overall situation of data after dimensionality reduction, the result is expressed in Figure 2.1:

Figure 2.1 signifies the missing pattern of the daily load curve. After dimensionality reduction, nearly 90% of the samples do not miss any information, and the 6 features of 7.7%of the samples are Not a Number (NaN), indicating that the value does not exist, which proves that there is no electricity or the account has been canceled. These users will be deleted. The remaining six missing patterns are caused by only a small amount of load during one part of the day and no power during the other. Then divided by the daily average load, it is judged by the computer as 0 / 0 type, the value does not exist, and it is displayed as a missing value, so these six types of users are listed as suspicious users.

2.2. Data anomaly detection model based on the iForest algorithm. The daily load curves of 5972 users in S City on July 18, 2021, were taken as the research object. The daily load curves were mainly selected for small and medium-sized special transformer users and three general industrial and commercial users. The sampling interval of samples was 15 minutes, and there were 96 measurement points in total. After data dimension reduction and cleaning, a total of 4872 effective daily load curves were obtained. There were 63 abnormal users, accounting for 1.29%.

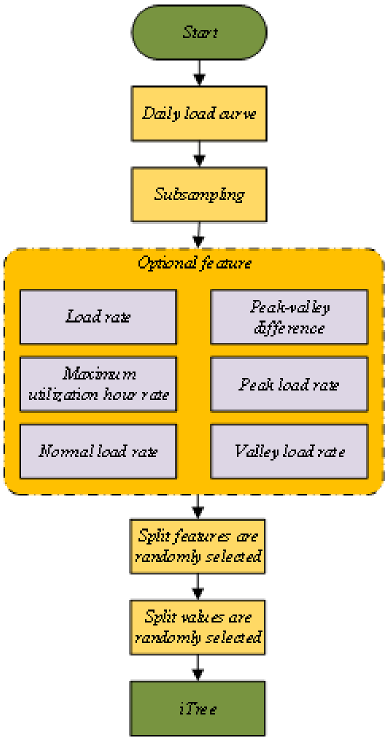


Fig. 2.2: The flow chart of the iTree construction

2.2.1. The construction of the Isolation Tree (iTree). The iForest is mainly composed of the iTree, which refers to a random binary tree in which each node contains two child nodes or leaf nodes [12]. The flow chart of the iTree construction is displayed in Figure 2.2:

Among them, the features in the data set of the daily load curve are all continuous variables. The construction steps of iTree are as follows:

Step 1: A feature is randomly selected among the 6 daily load characteristic indexes;

Step 2: A value k of the characteristics selected in step 1 is randomly selected;

Step 3: According to the characteristics, records are classified every day, and records with characteristics smaller than k are placed in the left branch, and records with characteristics greater than or equal to k are placed in the right branch;

Step 4: Then the left and right branches are constructed recursively until the following two conditions are met; There are only multiple identical records or one record in the incoming data set; The height of the tree reaches the specified height.

2.2.2. The construction of the iForest. The construction of iForest is similar to the method of random forest, both of which are carried out by random sampling. Each tree needs to be constructed through part of the data set to ensure that each tree has certain differences. The construction process of the iForest is revealed in Figure 2.3:

In the process of constructing iForest, the sampling size should be limited on the one hand, and the maximum depth should be set for each iTree on the other hand. Finally, it is necessary to calculate the power value of the tested user. In the process of evaluating the tested user, iForest can only evaluate a single customer at a time. Meanwhile, during the process of evaluation, each iTree needs to be traversed, the statistical query object is in the position of the leaf node, and then the average path length is employed to calculate the abnormal score. Finally, the user is evaluated by the value of the abnormal score, and then the user type is judged.

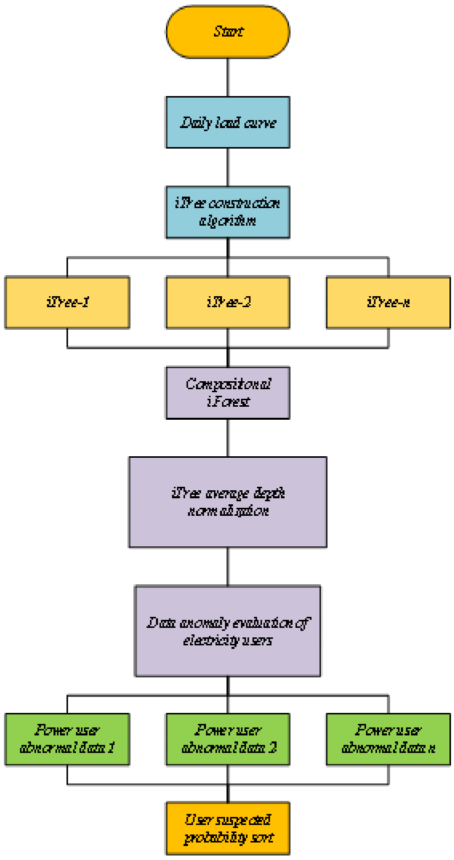


Fig. 2.3: The construction process of the iForest

2.3. Abnormal recognition of electricity consumption data based on the DT algorithm. The iForest algorithm is adopted to model the daily load curve of power users for abnormal detection, which is helpful to automatically screen users with abnormal data suspicion and realize the preliminary screening of abnormal users. Because the detection only by daily load curve is easy to cause misjudgment, the further analysis combined with other electrical variables of suspicious users can effectively improve the accuracy of detection.

The steps for the construction of the anomaly identification model of electricity power data based on the DT algorithm are demonstrated in Figure 2.4 [13].

The construction of the training set is mainly to sort out the date, meter number, and voltage data of the day in the EIAS, and sort out the transformer ratio and connection mode in the meter. Finally, the above data are combined to construct the training set of DT.

The DT algorithm is used to process the training set. Firstly, the training set is sorted, then it is divided by the threshold value of each data and the information gain is calculated. Furthermore, the threshold value is selected according to the maximum gain and the training set is divided.

The generation of DT: The root node and leaf node of DT correspond to a classification rule and synthesize all paths into a rule set, which is stored in a two-dimensional array.

Check the rationality of DT: the classification rules of DT are checked to see if there is any wrong decision. If so, the training set is adjusted until the classification is correct.

2.4. Update method for data exception. In actual work, the fault phase voltage of the data of the power acquisition system and the field measured data during the failure period is not 0, but eventually stabilizes

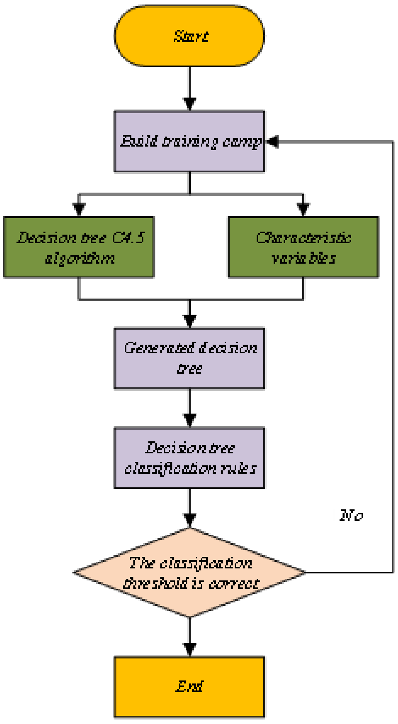


Fig. 2.4: The construction process of anomaly identification model for electricity consumption data

at a value not 0 with the change of time, which is called residual voltage [14, 15]. The residual voltage makes the amount of electricity measured by the energy meter during the voltage loss period contain the fault phase element and the amount of electricity measured by the non-fault phase element under the residual voltage. Thereupon, the calculation method of correction coefficient is determined by calculating the value of γ_{ab}, γ_{cb} according to the actual situation.

$$K = \frac{P_T}{P_F} = \frac{\sqrt{3}UIcos\varphi}{\gamma_{ab}UIcos(30^\circ + \varphi) + \gamma_{cb}UIcos(30^\circ - \varphi)} \tag{2.8}$$

The determination of γ_{ab} and γ_{cb} is related to the metering principle of the intelligent energy meter, and the measured electric energy is illustrated in Eqaution 2.9:

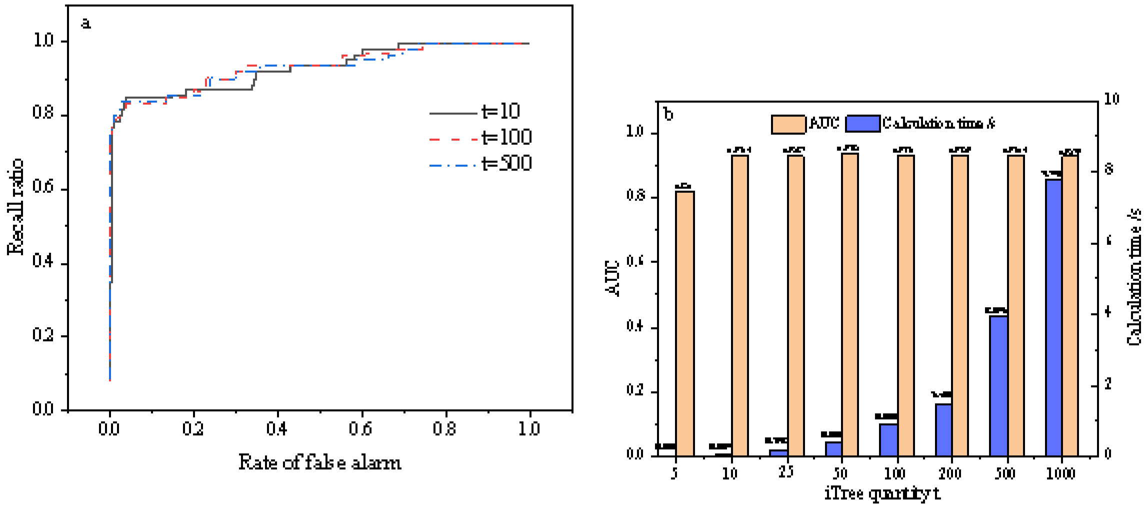
$$P = \frac{1}{T}\int_0^T u(t) \cdot i(t)dt \tag{2.9}$$

T represents the cycle of Alternating Current (AC) voltage and current.

By taking Δt as the sampling interval for voltage and current, the left discretization of Eqaution 2.9 is as follows:

$$P = \frac{1}{T}\sum_{k=1}^{N} u(k) \cdot i(k), T = N\Delta t \tag{2.10}$$

It can be seen that the determination of the correction factor is related to the current and voltage data during the failure. Through the EIAS, the frozen AC data of the Potential Transformer (PT) during the failure of voltage breakdown and loss can be known. Since the fusing on the PT side is independent of the current,



(a) ROC curve

(b) Computation time and AUC

Fig. 3.1: Different iTree quantities

the calculation of the correction factor only needs to consider the voltage. The voltage curve can be obtained by freezing the voltage data of the electricity acquisition system. Through linear regression or piecewise linear regression on the voltage change curve, the average value γU of voltage during fault is obtained.

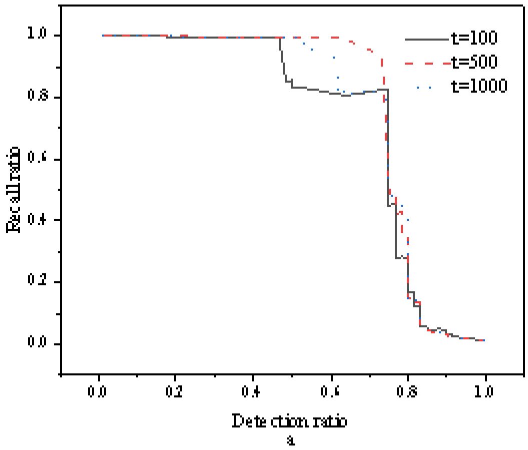
3. Results and Discussion.

3.1. Model parameter analysis. The iForest algorithm is mainly based on the thought of ensemble learning. There are two extremely important parameters in the anomaly detection model of electro-data, iTree sampling scale Ψ and the integration scale t. The simulation experiment is carried out on the system based on a Central processing unit (CPU) of dual-core 2.3GHz with 8GB memory, and the program is written through R language.

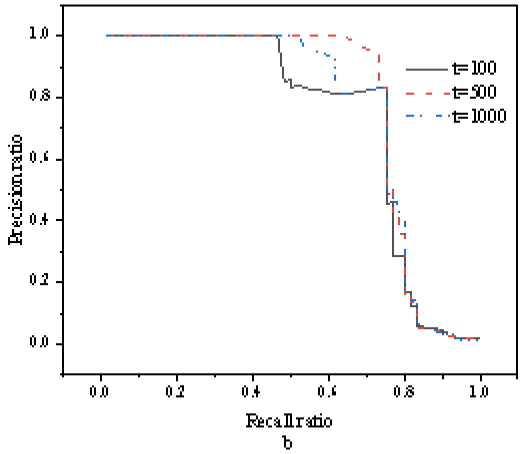
3.1.1. Quantity t of the iTrees . The iForest algorithm forms iForest by generating a certain number of itrees. It is mainly by means of random sampling to extract Ψ subset and construct iTree, and guarantee the diversity of iTree. Therefore, the number of iTree determines the size of the ensemble learning of the model. The Receiver operating characteristic curve (ROC) of the iTree with different numbers is portrayed in Figure 3.1.

Figure 3.1a portrays that the ROC curve is very close. The Area Under Curve (AUC) is obtained by calculating different numbers of iTrees respectively. According to the data in Figure 3.1b, the length of the path can be well covered when the number of iTrees reaches 100. After that, increasing the number of iTrees does not significantly improve the AUC, which is about 0.93 at this time. The cumulative recall ratio curve and Precision-Recall (P-R) curve of the iForest algorithm under different iTree numbers are shown in Figure 3.2.

In Figure 3.2a, when the number of iTrees is greater than 100, the gap between curves is very small. When the detection rate is less than 0.03, the curve has a very large upward trend, and when the detection rate is greater than 0.03, the curve tends to be flat. In the phase where the detection rate is less than 0.03, only the top 3% of users need to be detected to detect about 80% of abnormal users. At the stage where the detection rate is greater than 0.03, only 20% of abnormal users can be detected by detecting the remaining 97% of users. Thereby, the research focus of cumulative recall ratio is the stage where the detection rate is less than 0.03. As can be seen from the P-R curve in Figure 3.2b, when the iTree reaches more than 100, the precision ratio can exceed 80% when the recall ratio is 70%. Combined with Figure 3.2a, it can be found that 80% of abnormal users can be detected by detecting the top 2.5% of users for the abnormal score. When the detected users reach

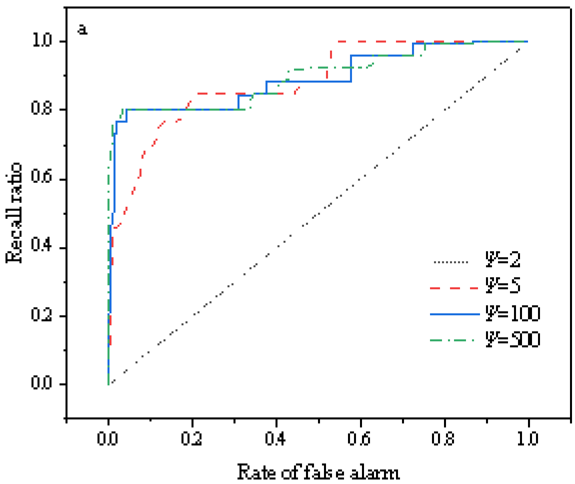


(a) Cumulative recall ratio curve

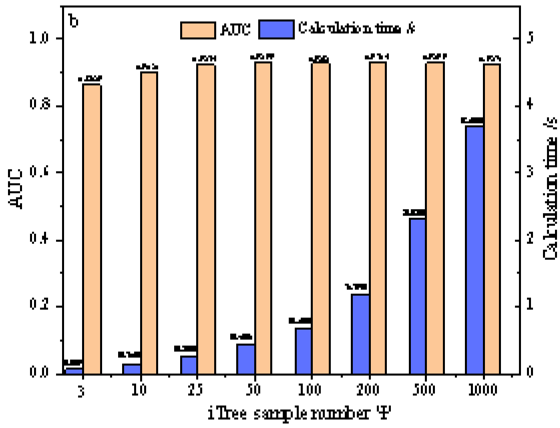


(b) P-R curve

Fig. 3.2: The iForest algorithm with a diverse number of iTrees



(a) Relationship between ROC curve and iTree samplings



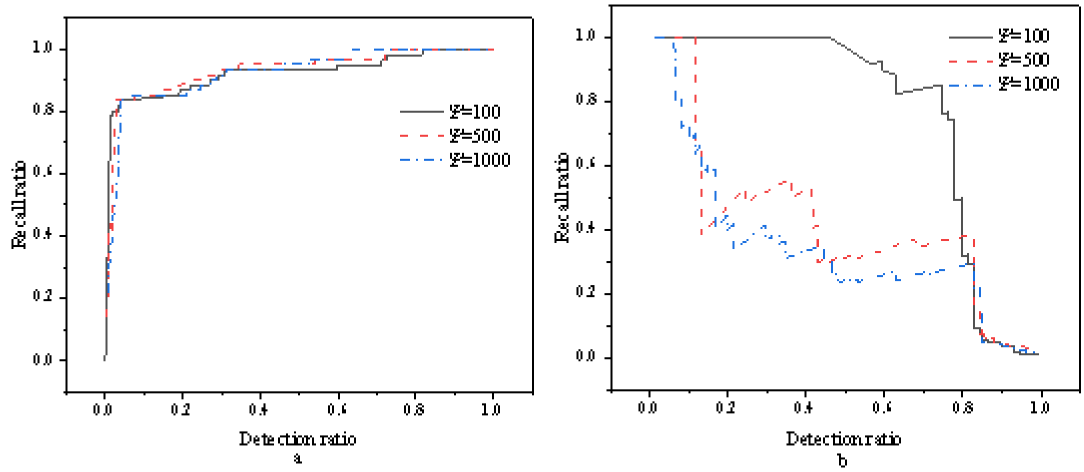
(b) Calculation time and AUC

Fig. 3.3: The iForest algorithm with different iTree sampling numbers

3.5%, the precision ratio decreases significantly, only about 40%.

3.1.2. The iTree sample numbers Ψ. For any object in the data set, different sample numbers will affect the user's abnormal score and affect the final output of the model. Thus, it is important to study the sensitivity of model parameters. The relationship between the ROC curve and the iTree sampling number is indicated in Figure 3.3.

In Figure 3.3a, when the collection number of iTree is small, the performance of the model is poor, but when the number of iTree reaches a certain value, the ROC curve will be very close. It can be seen from the data in Figure 3.3b that as the number of iTree samples increases, the calculation time continues to increase. The area AUC under the ROC curve is not normal with the change of parameters, on the contrary, it decreases a little. The P-R and cumulative recall ratio curves of the iForest algorithm with diverse iTree sampling numbers



(a) cumulative recall ratio curve (b) P-R curve

Fig. 3.4: The iForest algorithm with various iTree sampling numbers

Table 3.1: The confusion matrix of performance evaluation of the voltage classification model

Prediction	Normal voltage N	HV H	LV L
Normal voltage N	339	0	0
HV H	0	451	0
LV L	0	0	354

are implied in Figure 3.4.

The cumulative recall ratio curve in Figure 3.4a shows that when the sampling number is 100, 65% of abnormal users can be detected by detecting 2% of users. When the number of samples is much larger than 100, only 42% of abnormal users can be detected by detecting the top 2% of users. The P-R curve in Figure 3.4b can be more significantly found that when the sample numbers of iTree is too large, the performance of the model is poor. The reason for the above phenomenon is primarily that the purpose of iTree sampling is to better separate normal users from abnormal users, and the more sampled data, the worse the ability of the iForest algorithm to identify anomalies.

3.2. DT of voltage classification for power users. For the sake of explanation, the training set mainly covers 10kV three-phase measurement points. The model is constructed by using R language, and the DT of voltage classification is obtained, as plotted in Figure 3.5.

The output results in Figure 3.5 are voltage classification, where N, H, and L represent normal voltage, high voltage (HV), and low voltage (LV), respectively. The judgment condition of DT is the value between each node, and the whole DT mainly contains 10 decision points. Taking the left-most root node as an example, when the voltage value is less than 51V, the phase sequence is B and the connection mode is three-phase and three-wire, the voltage is normal; If the phase sequence is B and the connection mode is three-phase and four-wire, it is LV. If the phase sequence is A or C, it is LV. The confusion matrix of performance evaluation of the voltage classification model is expressed in Table 3.1.

Table 3.1 exhibits that there are 1166 data in the DT training set of voltage classification, and the classification accuracy is 100%. In general, the classification model is an overfitting phenomenon when the success rate of the classification reaches 100%, which is the characteristic of DT. The fitting phenomenon is more beneficial to the accurate classification of voltage between different measurement points, and the DT is more sensitive to the training set, which is conducive to the dynamic adjustment of voltage judgment rules.

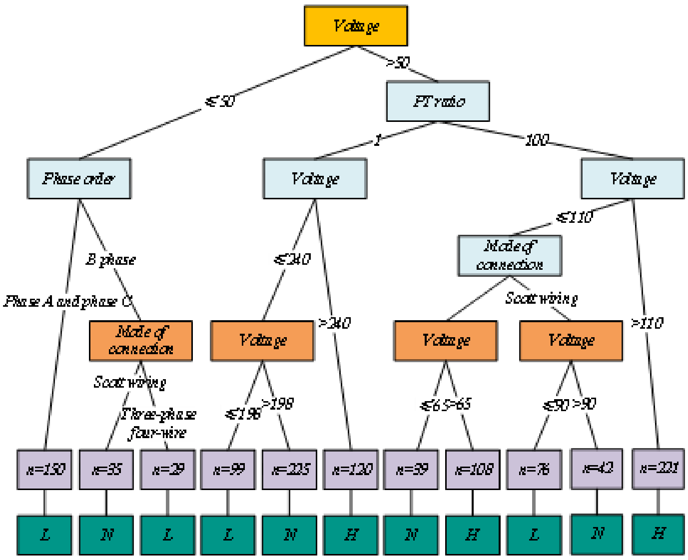


Fig. 3.5: The DT of voltage classification for users

Table 3.2: The meter data and file information of abnormal users

User ID	123456XX
Meter reading at the beginning of the failure	371.68
Meter reading at the end of the failure	392.15
Counting numbers during failure	17.32
CT	12
PT	100
Combined multiplier	1200

3.3. Data anomaly detection of power users. Through iForest algorithm, data anomalies of users of special transformers in S City are detected. After preliminary screening, 179 suspected abnormal users are detected. According to the energy meter number of the abnormal user, the date and 96-point voltage in the EIAS of the user are sorted out. The voltage transformer ratio and connection mode of the meter in SG186 are also sorted out, and the above two parts are combined. Then DT algorithm is used to identify the voltage anomaly of the metering device of the suspected abnormal users, and 65 abnormal users are output. Finally, 45 users with the highest degree of suspicion are sent out operation and maintenance work orders after manual review. After on-site verification, 37 abnormal users are identified. Figure 3.6 demonstrates the statistics of the feedback results of abnormal troubleshooting for users of special transformers:

3.4. Update of abnormal electricity consumption data for energy meters. Among the anomalies determined by the energy meter, the fuse failure of the voltage transformer is broken the most. Taking a case of fuse burn-out of a voltage transformer verified in S City as an example, it is found that there has an LV phenomenon through model recognition. The 96 points of voltage and current are obtained and sorted out from the EIAS. From A certain time period, the C-phase voltage has a large jump, the A-phase voltage is normal, and the current of the A and C phases is normal, indicating that the power supply of the user is normal, but the electric energy metering device is abnormal. The user data and file information obtained through the EIAS and SG186 are signified in Table 3.3:

According to the 96-point voltage data of the abnormal user, after the fuse on the primary side of C-phase PT is blown, the voltage between C and B phases on the secondary side quickly drops to about 14V

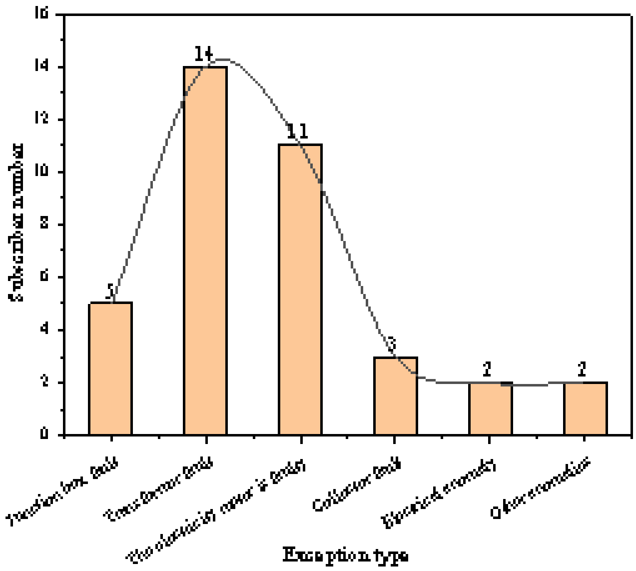


Fig. 3.6: Abnormal feedback results of electric energy metering device

and remains stable. The voltage data collected by the EIAS every 15 minutes during the fault period are statistically analyzed. It can be seen that the average voltage between C and B phases during the failure period is 14V, namely $\gamma_{cb} = 0.14$. It means that the power factor of the user is relatively stable, and the average value is $cos\varphi = 0.85$, obtained after conversion $\varphi = 31.79°$.

4. Conclusion. With the swift growth of the power industry, a variety of power equipment terminals have appeared. Based on the analysis and research status of electro-data anomalies, the data anomalies of energy meter terminals are studied. Firstly, the data of the daily load curve is preprocessed and the iForest algorithm is selected to construct the abnormal detection model of electricity data. Additionally, the accuracy and calculation time of the model are explored by analyzing the different values of the sample numbers of iTrees. Secondly, on account of the voltage and current data of suspected abnormal users initially screened, the anomaly identification model of electricity consumption data of the DT algorithm is implemented. Besides, the voltage at different metering points is classified and judged in combination with current data to identify abnormal energy metering devices. Finally, the case of an abnormal user in S City is analyzed, and the 96-point voltage data of the EIAS during the fault period are studied. Furthermore, the correction coefficient is adjusted on the basis of the participating voltage of the fault phase. The results manifest that the anomaly detection model of electricity power data constructed by the iForest algorithm has high validity and computational efficiency. The traditional recharge method does not consider the residual voltage, but the recharge method proposed here makes the recharge more accurate by determining the residual voltage and ensuring the accuracy of the charge. However, since only electrical variables such as voltage and current are considered in the analysis of anomalies in electricity data, the accuracy of identification needs to be improved. It is hoped that more in-depth exploration can be carried out in the subsequent research, and more variables can be introduced to improve the accuracy of recognition.

REFERENCES

[1] Avancini D. B., Rodrigues J. J. P. C., Rabêlo R. A. L., et al. (2021) A new IoT-based smart energy meter for smart grids[J]. International Journal of Energy Research, 45(1), 189-202.
[2] Santhosh C., Kumer S. V. A., Krishna J G., et al. (2021) IoT based smart energy meter using GSM[J]. Materials Today: Proceedings, 46, 4122-4124.

[3] Butt O. M., Zulqarnain M., Butt T. M. (2021) Recent advancement in smart grid technology: Future prospects in the electrical power network[J]. Ain Shams Engineering Journal, 12(1): 687-695.
[4] Zheng K., Chen Q., Wang Y., et al. (2018) A novel combined data-driven approach for electricity theft detection[J]. IEEE Transactions on Industrial Informatics, 15(3), 1809-1819.
[5] Hasan M. N., Toma R. N., Nahid A. A., et al. (2019) Electricity theft detection in smart grid systems: A CNN-LSTM based approach[J]. Energies, 12(17), 3310.
[6] Ji X., Yin Z., Zhang Y., et al. (2021) Real-time robust forecasting-aided state estimation of power system based on data-driven models[J]. International Journal of Electrical Power & Energy Systems, 125, 106412.
[7] Liu X., Ding Y., Tang H., et al. (2021) A data mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data[J]. Energy and Buildings, 231, 110601.
[8] Yan Z., Wen H. (2021) Electricity theft detection base on extreme gradient boosting in AMI[J]. IEEE Transactions on Instrumentation and Measurement, 70, 1-9.
[9] Wang Z., Fu Y., Song C., et al. (2019) Power system anomaly detection based on OCSVM optimized by improved particle swarm optimization[J]. IEEE Access, 7, 181580-181588.
[10] Chapaloglou S., Nesiadis A., Iliadis P, et al. (2019) Smart energy management algorithm for load smoothing and peak shaving based on load forecasting of an island's power system[J]. Applied energy, 238, 627-642.
[11] Shi Y., Yu T., Liu Q., et al. (2020) An approach of electrical load profile analysis based on time series data mining[J]. IEEE Access, 8, 209915-209925.
[12] Zhang Y., Zhang J., Yao G., et al. (2020) Method for clustering daily load curve based on SVD-KICIC[J]. Energies, 13(17), 4476.
[13] Jiang J., Li T., Chang C., et al. (2022) Fault diagnosis method for lithium-ion batteries in electric vehicles based on isolated forest algorithm[J]. Journal of Energy Storage, 50, 104177.
[14] Charbuty B., Abdulazeez A. (2021) Classification based on decision tree algorithm for machine learning[J]. Journal of Applied Science and Technology Trends, 2(01), 20-28.
[15] Abd Rahman F. A., Ab Kadir M. Z. A., Ungku Amirulddin U. A., et al. (2021) Computation of energy absorption and residual voltage in a fourth rail LRT station arresters in EMTP-RV: A comparative study[J]. Urban Rail Transit, 7(2): 71-83.

Edited by: B. Nagaraj M.E
Special issue on: Deep Learning-Based Advanced Research Trends in Scalable Computing
Received: Dec 27, 2023
Accepted: Mar 18, 2024