Research on Smart Electric Meter Data Mining Technology Method for Line Loss Diagnosis of Low Voltage Station Area

Yang Fuli
State Grid Chongqing Electric Power Research Institute
No. 80, Middle Section of Huangshan Avenue
Yubei District
Chongqing, China
E-mail: 18680887087@163.com

Hou Xingzhe
State Grid Chongqing Electric Power Research Institute
No. 80, Middle Section of Huangshan Avenue
Yubei District
Chongqing, China
E-mail: cqhhxz@163.com

***Abstract*—Line loss can be divided into statistical line loss, technical line loss and management line loss according to structure. It not only refers to the energy loss in the form of heat energy, but also the management line loss caused by the electricity stealing behavior [1]. The calculation of power system line loss and the realization of system lean management are of great significance in guiding the reduction of energy conservation and the promotion of line loss management. To this end, in-depth analysis of the massive user data accumulated in the marketing automation process of the electricity information system in recent years, so as to establish a reasonable and efficient mathematical model of line loss analysis. By mining the useful information behind these data in smart electric meter, the abnormal power usage behavior detection of the user is realized, so as to achieve the purpose of preventing electric larceny and leakage and thereby reducing the line loss. This paper proposes a layer-based power line electric larceny detection method based on data mining technology .This method optimizes the traditional LOF algorithm and is a weighted LOF algorithm. By performing weighted outlier analysis on massive user data, the location of abnormal power users can be more efficiently completed.**

Keywords-Component Management Line Loss, Data Mining; Layer-Based Analysis; Weighted LOF Algorithm; Outlier Analysis; Abnormal User Location

I. INTRODUCTION

Line loss is an important economic and technical indicator for power supply enterprises. The level of line loss management is a concentrated expression of the power supply enterprise's operational capability and comprehensive management capability[2], and its effect directly affects the economic benefits of power supply enterprises. As the end of the power network, the low-voltage station area provides power supply to the largest number of residential users and small and micro enterprise users. It is the most important and complicated part of line loss management.

In order to reduce the management line loss, this paper combines the outlier analysis method in data mining to study the abnormal power consumption behavior. Moreover, in order to solve the limitations of the traditional abnormal power detection method, a weighted power line abnormal power detection method based on the analytic hierarchy process is proposed.

II. ABNORMAL ELECTRICITY DETECTION BASED ON DATA MINING

As an emerging data processing method, data mining can effectively handle data analysis in high-volume and multi-complex scenarios. Therefore, it is considered to introduce data mining to effectively cope with the challenge of detecting abnormal user data in large-scale user power consumption data.

The abnormal electric energy usage detection model based on data mining is mainly divided into three parts: user electrical data acquisition and processing, construction of abnormal electric energy usage model, user detection, detection result analysis and verification. As shown in Figure 1 below:

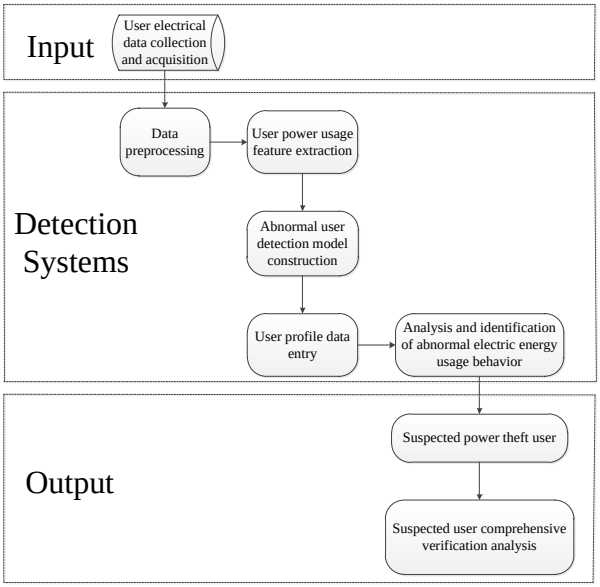

Figure 1. The abnormal electric energy usage detection model based on data mining

III. PRINCIPLE OF ANOMALY DETECTION BASED ON FEATURE OUTLIER ANALYSIS OF ELECTRICAL ENERGY USE

A. Feature extraction of electrical energy use

Extracting abnormal power usage information caused by electric larceny behavior is often not isolated, and a electric

larceny scenario may trigger multiple anomalies. If the test is based on a single indicator, it is likely that omission or misjudgment will occur. Therefore, efficient anti-tampering work should be carried out by extracting comprehensive features from a variety of abnormal power usage phenomena and around the quantifiable feature quantities caused by various abnormal power usage behaviors. Taking a single-phase user as an example, the analysis of the power consumption evaluation indicators[3-7] is as follows: daily average voltage, daily average power factor, daily average current imbalance rate, average daily frozen power of the previous seven days, and power imbalance rate.

B. Principle of electric larceny determination based on outlier detection

Outlier detection, also known as anomaly detection, is designed to find objects with different behavioral characteristics in the sample group. Common detection methods mainly include distribution-based, depth-based, distance-based, density-based, and offset-based five categories.

Considering the power user environment, the density-based detection algorithm can better adapt to the complex electrical dataset of internal structure. Therefore, the most representative LOF algorithm[8] in density detection algorithm is selected and applied to the abnormal power detection. In order to give the reader a better understanding of the algorithm, the following concepts are introduced here:

Definition 1 The k-distance neighborhood $N_K(p)$ of the object p, that is, the set of all objects that do not exceed the distance $dist_k(p)$ from the object p. Its expression is

$$N_k(p) = \{o \mid o \in D, dist(p, o) \leq dist_k(p)\} \quad (1)$$

In the formula,$dist_k(p)$ represents the k-distance of the object p.

Definition 2 The local reachable density of the object p $lrd_k(p)$ whose mathematical expression is

$$\mathrm{lrd_k(p)} = \frac{\|\mathrm{N_K(p)}\|}{\sum_{o \in \mathrm{N_K(p)}} \mathrm{reachdist_k(p,o)}} \quad (1)$$

Where $reachdist_k(p, o)$ represents the reachable distance of object o to object p, and the reachable distance is defined as follows: Given a natural number k, the reachable distance of object p with respect to object o is

$$\mathrm{reachdist_k(p, o) = \max(dist_k(p), dist(p, , o)} \quad (2)$$

Definition 3 The local outlier factor (LOF) of object p is defined as

$$\mathrm{LOF_k(p)} = \frac{\sum_{o \in \mathrm{N_K(p)}} \frac{\mathrm{lrd_k(o)}}{\mathrm{lrd_k(p)}}}{\|\mathrm{N_K(p)}\|} \quad (4)$$

It can be seen from the formula that the smaller the density of the data object is, the larger the density of the object in the k-distance neighborhood is, the larger the local outlier factor LOF value will be, and the greater the degree of outliers. Detection of group points. According to this principle, the user outlier degree is obtained by the LOF algorithm, and the degree of suspicion of the user's electricity theft is expressed, and the abnormal user detection is completed according to the degree of the suspect.

IV. WEIGHTED LOF ALGORITHM AND ITS APPLICATION IN THE DETECTION OF STOLEN ELECTRIC ENERGY

Due to the large difference in user power characteristics, this makes the power data set internally exhibit complex distribution characteristics. However, the LOF algorithm can effectively avoid the influence of uneven data distribution on the detection results, and has a good detection effect on the power dataset with unbalanced density. Because different electrical indicators have different meanings, if the outliers are obtained directly by the traditional LOF algorithm and used to indicate the degree of suspicion of users electric larceny, this is unreasonable when multiple electrical indicators are of different importance to the suspected electric larceny. If the different indicator data have the same degree of outliers, it does not mean that the two objects have the same suspicion of electric larceny.

Considering that the selected electrical indicators are of different importance for suspicion of electric larceny, it is necessary to analyze the probability that the abnormality of each electrical indicator data can represent the electric larceny [8]. At the same time, the analytic hierarchy process is used to give reasonable weight to each detection index, and the weighted LOF algorithm is used to comprehensively quantify the user's suspicion of electric larceny. The resulting comprehensive outliers are used to characterize the degree of user suspicion of electric larceny, and the efficiency of user detection of electric larceny is improved. Finally, the effectiveness of the proposed detection method is verified by experiments.

A. Index weight determination

The traditional LOF algorithm does not distinguish the meaning of different indicator data. The outliers obtained by the algorithm can only represent the abnormal degree of the user's power consumption characteristics, but can't explain the degree of user's suspicion of electric larceny. Therefore, it is necessary to assign corresponding weights to different dimensional data of the traditional LOF algorithm. The greater the probability that the electrical index anomaly can represent the electric larceny, the greater the role it should play in the overall suspected electric larceny analysis, and the greater the weight, and vice versa. Therefore, the analytic hierarchy process (AHP) [9] is introduced as a mathematical tool for weight quantification.

The basic idea of AHP is to decompose complex problems and form a hierarchical structure according to the dominance relationship. At the same time, according to a certain ratio scale, the judgment is quantified through pairwise comparison, and the relative judgment matrix is calculated to determine the relative importance of the elements of the hierarchy. The specific steps include:

1) Constructing a hierarchy of evaluation indicators for electric electric larceny.

The degree of suspicion of the user's electric larceny is taken as the object to be evaluated, and various quantifiable electrical characteristic parameters caused by the phenomenon of electric larceny are used as the evaluation index set, thereby constructing the suspicion evaluation system for electric larceny as shown in Figure 2.

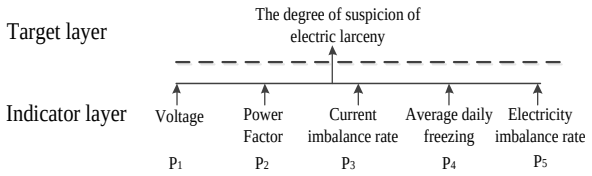

Figure 2. Assessment system for the suspicion of electric larceny

2) Electrical parameters.

According to the degree of importance relative to the suspected electric larceny, the judgment matrix P is formed. The relative weights of the indicators are compared using the scale of 9 /9 to 9 /1 [10]. The specific scale is shown in Table 1.

TABLE I. THE 9 /9 ～ 9 /1 STANDARD DEGREE METHOD

grade	Scaling	The importance of the former over the latter
K=1	9/9	
K=3	9/7	
K=5	9/5	
K=7	9/3	
K=9	9/1	
K=2,4,6,8	9/(10-K)	Intermediate value between adjacent judgments

Refer to the expert experience to build the indicator judgment matrix P based on the probability that the electrical indicator data anomaly can represent the electric larceny. As shown in Formula (5):

$$P = \begin{bmatrix} 1 & 9/8 & 8/9 & 9/4 & 9/8 \\ 8/9 & 1 & 7/9 & 1 & 1 \\ 9/8 & 9/7 & 1 & 9/3 & 9/7 \\ 4/9 & 5/9 & 3/9 & 1 & 5/9 \\ 8/9 & 1 & 7/9 & 9/5 & 1 \end{bmatrix} \quad (5)$$

After comprehensively analyzing the characteristics of each electrical parameter in the user's abnormal power consumption, it is concluded that the importance of current suspicion of electric larceny is higher than other indicators. The metering voltage, power factor, and power imbalance rate are all important to the test results. However, since the power factor depends on the load nature of the user in addition to the power grid, there are some fluctuations that are normal. The power imbalance rate is limited by the communication system communication capacity and current acquisition frequency. The data obtained will inevitably have less fluctuations with the actual power consumption. Therefore, the power factor and the power imbalance rate are slightly lower than the voltage.

3) Judgment Matrix P

The eigenvector corresponding to the maximum eigenvalue λ_{max} of the matrix is obtained, and the normalized eigenvector is the weight of the abnormal power consumption index. At the same time, the consistency check of the judgment matrix is performed. According to the above steps, the weight of the suspected evaluation index of electric larceny is as shown in the Table 2.

TABLE II. INDEX WEIGHT DISTRIBUTION OF ELECTRIC LARCENY

Index	p1	p2	p3	p4	p5
Weight wi	0.2296	0.1992	0.2690	0.1030	0.1192

B. Electric larceny analysis based on weighted LOF algorithm

Here we briefly discuss the impact of distance metrics on the results. At present, the wider distance measure is the Euclidean distance, so that $i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $j = (x_{j1}, x_{j2}, \ldots, x_{jn})$ are two objects described by n numerical attributes. The Euclidean distance between objects i and j is defined as

$$d(i,j) = \sqrt{\sum_{k=1,\ldots,n}(x_{ik} - x_{jk})^2} \quad (6)$$

In the analysis of suspected electric larceny, due to the different meanings of electrical indicators, the contribution of different indicators to the suspected power stealing is slightly different. Therefore, the corresponding weights are set for different electrical indicators in the outlier detection. The improved distance expression is:

$$d(i,j,w) = \sqrt{\sum_{k=1}^{n} w_k{}^2 (x_{ik} - x_{jk})^2} \quad (7)$$

Where $w_k \in [0, 1]$, which represents the weight of the k-th dimension attribute. The decrease in w_k, the greater the compressibility of the attribute corresponding to the coordinate axis, the smaller the effect.

When calculating the comprehensive outliers, the weighted Euclidean distance is used to weigh the comprehensive distance between any users, and then the comprehensive outliers that can represent each user's suspicion of electricity sneak are obtained. The larger the outlier factor is, the more The higher the possibility of stealing electricity.

The overall structure of the power stealing analysis system based on the weighted LOF algorithm is shown in the figure 3.

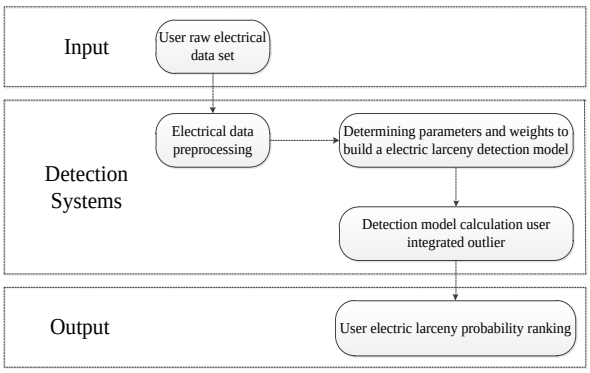

Figure 3. System structure diagram of ElectricLarceny Detect

V. EXPERIMENTAL RESULTS AND ANALYSIS

Compare and analyze the detection effect of traditional LOF algorithm and this algorithm on electric larceny record. The data set used was from a power company's anti-electricity larceny Inspection, and the data set covered 1 143 households.

A. *Results evaluation criteria*

Several classifier evaluation tools are used to measure the detection effect of the algorithm, and the records in the data set are summarized in the confusion matrix form according to the two criteria of the real category and the classification model. As shown in the following Table 3, each column in the table represents a forecast category, and each row represents the true attribution category of the data.

TABLE III. CONFUSION MATRIX

	Forecast category	
True attribution category	True Positive（TP）	False Negative（FN）
	False Negative（FN）	True Positive（TP）

Based on the confusion matrix, multiple evaluation criteria can be derived: precision = TP / (TP + FP), recall rate Recall = TP / (TP + FN), true rate TPR = TP / (TP + FN), false positive rate FPR = FP / (FP + TN).

In practice, normal users and abnormal users have imbalances in class distribution. In order to more intuitively express the final detection results, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) are introduced here. The ROC curve describes the relative relationship between the growth rate of FPR and TPR in the confusion matrix, and uses the area under the ROC curve. AUC uses a numerical value to indicate the performance of the classifier. The larger AUC represents better performance.

B. *Algorithm detection result*

Figure 4 shows the changes in the recall rate of the two types of detection algorithms with the detection rate. It can be seen from Fig. 3 that the whole tamper detection can be roughly divided into two parts. When the detection rate is low, the curve grows rapidly. When the detection rate exceeds 0.2, the rising momentum slows down and eventually stabilizes. This means that only about the top 20% of the suspects with high suspicion detection can detect about 80% of sneak users. In order to save costs, the abnormal power consumption detection can focus on the user with a large suspect coefficient of the detection algorithm output, thereby improving the efficiency of abnormal power detection.

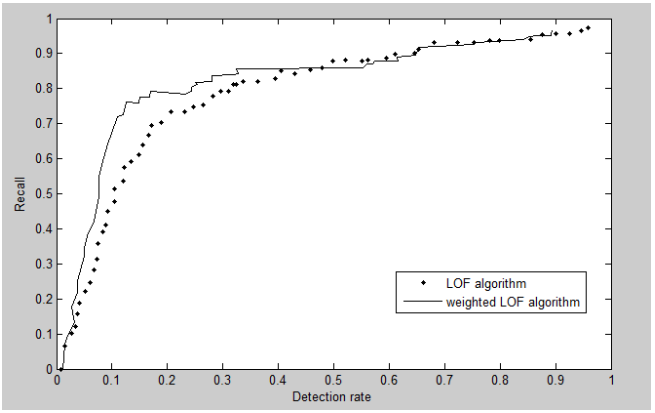

Figure 4. Cumulative recall curve for the two algorithms

Figure 5 shows the ROC curves for the two detection algorithms. According to the meaning of electrical parameters, this paper comprehensively quantifies the importance of different electrical indicators for cyber electricity analysis. The improved algorithm's detection results more reasonably explain the user's suspicion of thief, so the area under the modified ROC curve is higher than the traditional LOF algorithm, which means that the overall detection effect of the thief user is better than the traditional LOF algorithm.

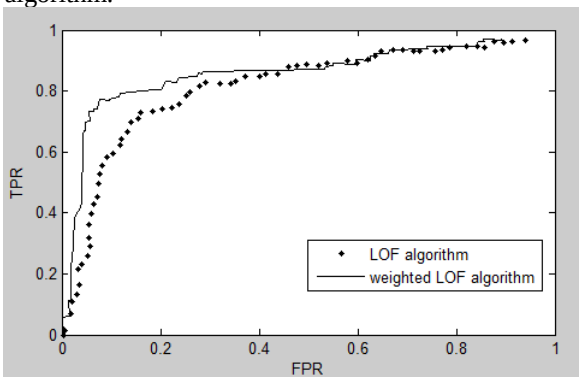

Figure 5. ROC curve for the two algorithms

VI. CONCLUSION

Through in-depth analysis of the user's electricity consumption data obtained by the electricity information collection system, the power behavior information hidden behind the data can be effectively revealed, and the effective detection of various abnormal power usage behaviors can be completed. In this paper, based on the theory of data outlier analysis, a method based on weighted LOF algorithm is

proposed. This method integrates multiple electrical indicators to evaluate the user's full-scale tampering. The method can complete the detection of a variety of abnormal powers by monitoring a plurality of quantifiable electrical parameters caused by the power stealing phenomenon. At the same time, it is only necessary to detect the user with a high suspect coefficient output from the outlier analysis to complete the detection of most power theft users.

ACKNOWLEDGMENT

This work was supported by the State Grid Corporation Science and Technology Project "Research on key technologies of risk monitoring and life evaluation of smart meters".

REFERENCES

[1] HUANG Jing, ZHANG Min, XU Wei, CHEN Xiyin, LU Xi, LI Chao, et al. A Study on the Diagnostic Model for the Line Loss in the Low-Voltage Transformer District Based on the Data Mining Technology for the Synchronous Line Loss System. Journal of Chongqing Electric Power College, 2018;23(6) : 26-30

[2] LI Jianning, MA Xiaoli, TAN Huamin, JIANG Chen, et al. A nomaly Diagnosis System for Low-Voltage Area Line Loss Based on Wireless Communication and Big Data Technology. Electricity and energy , 2019; 40(1) : 36-40

[3] Han Gujing, Yin Xiaogong, Qin Liang, et al. A novel technique of Preventing electricity-stealing in current method for electric power Measuring equipment. Electrical Measurement & Instrumentation, 2007; 44(10) : 29-32

[4] Li Dayong, Wang Yu, Li Canbing, et al. A design of an boxopening Recorder for anti power-stealing based on RFID Electrical Measurement& Instrumentation, 2008; 45(10) : 51-55

[5] Chen Chen. The design and improvement of anti-steal watt-hour Meter. Electrical Measurement & Instrumentation,2009; 46 (9A) : 113-116

[6] Wang Hui, Liu Fei. Application of wireless communication technology in electricity larceny prevention. Electrical Measure Instrumentation, 2015; 52(1) : 124-128

[7] Han Songlin, Shang Dezu. The anti-pilfering electricity abilities Distinguished of new types watt-hour meter. Electrical Measurement & Instrumentation, 2002; 39(11) : 46-49

[8] Breunig M M. LOF: identifying density-based local outliers. ACM SIGMOD International Conference on Management of Data. ACM, 2000: 93-104

[9] Tan Zhiyuan. Design and implementation of online abnormal electricity utilization and risk monitoring system based on electricity Behavior analysis. Guangzhou: South China University of Technology, 2015

[10] Chen Jinfu, Zhao Yunfei, Zhou Renjun. Analytic hierarchy process And its application in power system. Electric Power Automation Equipment,2004,24(12) : 20-23

[11] Wang Hao, Ma Da. Scale evaluation and new scale methods.Systems Engineering-Theory & Practice, 1993; 13(5) : 24-26