A Method for Anti-electric Theft Based on User Electricity Load Forecasting and Historical Load Data

Zhi Zhang¹, Xianguang Dong¹, Bo Liang², Yanjie Dai¹, Zhiru Chen¹, Pingxin Wang¹

¹ State Grid Shandong Electric Power Research Institute, Jinan, 250003, China
² State Grid Linyi Power Supply Company, Linyi, 276000, China
E-mail: 270882626@qq.com

Abstract. With the development of socioeconomic and the increasing demand for electricity, the electricity safety requirements are getting higher and higher and the importance of anti-electric theft has become increasingly prominent. At present, the phenomenon of electric theft is frequent, but the quality of relevant inspectors and related technologies are inadequate in identifying electricity stealing. In view of the above problems, this paper proposes a new method for anti-electric theft, which can predict the load of users. User load forecasting curve is formed, and combined with user history data, user forecasting load curve and industry load characteristic data, logistic regression algorithm is used to establish user electric theft judgment model and judge the possibility of user electric theft. The experimental results on the relevant datasets show that the proposed algorithm has a good effect.

1. Introduction
In recent years, with the development of the means of electric theft the scope is also expanding, and the related litigation cases are also on the rise. This not only affects the normal operation of the power system, but also causes significant losses to the relevant power enterprises. According to incomplete statistics, the annual loss caused by electric theft amounts to 20 billion yuan. In order to promote the healthy and stable development of electric power enterprises, it is imminent to carry out relevant anti-electric theft and electricity detection work. Common anti-electric theft measures mainly include technical means and management means. Technical means mainly use related hardware equipment to fight against electric theft, but the shortcomings are unclear objectives, waste of resources and time-consuming. The management means is to check the point of electric theft by mobilizing the enthusiasm of grass-roots personnel. The shortcoming is that it can not effectively obtain evidence, quantify and timely feedback information.

2. Relevant analysis of anti-electric theft
At present, the anti-electric theft has received the attention of power enterprises at all levels, has also received widespread attention from all walks of life, the corresponding means of electric theft are also constantly developing. By using the data collected by the telecommunication information acquisition system, the parameters of users’ power consumption state can be monitored in real time. We can also use large data technology to select voltage, current, power and electricity as the electrical reference in the criterion of electric theft. Through the combination of current collected data and historical data to analyze electricity consumption behavior, suspicious users of electric theft are found. For example,
Tang dengping\cite{4} et al. designed a high-performance standard digital watt-hour meter for complex working conditions, which can meet the requirements of accuracy and calibration. Ye Peng\cite{5} designed an on-line calculation method for compensating the loss of voltage fault of three-phase watt-hour meter, which reduces the measurement error and improves the accuracy.

Among the above anti-electric theft measures, more and more scholars focus on the research of anti-electric theft technology\cite{6,7}. The main problems of the current anti-electric theft technology are the inaccuracy of the judgment and the inaccuracy of the user's characteristic information mining. In view of the above problems, this paper proposes a new method for anti-electric theft on the basis of user electricity load forecasting and historical load data. The advantages of the method mainly includes:

Based on the logistic regression algorithm, combining the most relevant data of users, the short-term load forecasting data of users and the historical load data of users, an electric theft judgment model is established to improve the effect of electric theft judgment.

3. Algorithm principle

The core content of the algorithm includes three modules: data processing module, load forecasting module which has been established in our other paper\cite{1} and electric theft judgment module. The load forecasting module is used to get the user's electricity characteristic curve. In the electric theft judgment module, it is compared with the industry characteristic curve to judge the possibility of electric theft.

This part is the core content of this paper, in which the logistic regression algorithm is used as the data processing principle. The used data contain two parts: user forecasting data, associated factor data and industry electricity characteristic sequence data, with the flow chart judged shown in Figure 1.

![Fig. 1 Judgment of User Electric Theft Behavior](image)

The electric theft is judged as per the user's historical data on the basis of the logistic regression models, which mainly includes the model training and the model test, with the steps as follows:

**Step 1:** The model sample data selected contain two parts, in which the first one is the comparison data between user electricity forecasting data and industry electricity characteristics sequence; the second one is the associated factor data of the user electricity. The equal proportion electric theft sample data and the normal electricity sample data are selected. The corresponding proportion of data is drawn from the sample set at random to constitute the training set data and test set data;
The comparison data of electricity characteristic sequence data between different users and industries are calculated with the correlation coefficient:

\[ r(X,Y) = \sum_{p=1}^{m} \frac{\text{Cov}(X_p,Y_p)}{\sqrt{\text{Var}[X] \text{Var}[Y]}} \]  

(1)

In the above formula, \( X \) indicates the predicted user electricity sequence, \( Y \) indicates the industry electricity characteristics sequence data, and \( X_p \) and \( Y_p \) are the electricity data corresponding to the same time point. \( \text{Cov}() \) indicates the covariance, while \( \text{Var}[\cdot] \) indicates the variance.

**Step II:** The collective training sample is defined as \( X=\{x_1, x_2, \ldots\} \), and the characteristic weight vector \( \theta=(\theta_1, \theta_2, \ldots) \), then the objective function \( f(\theta)=\theta^\top X \), the category set \( C \in \{C_1, C_2\} \), the initialized iterations number \( k=0 \), the permissible error \( \varepsilon>0 \), and the random assignment is conducted to \( \theta \) on the basis of certain scope.

**Step III:** Iterative solution \( k+1 \): The quasi-newton method is adopted for the optimal solution of objective function, and the gradient of objective function:

\[ \nabla f = \sum_{\substack{x \in X, C=C_2}} \frac{p_{1i}p_{i2}x_i}{p_{ni} + N_i} - \frac{\sum_{\substack{x \in X, C=C_1}} p_{ni} \sum_{\substack{x \in X, C=C_1}} p_{1i}p_{i2}x_i}{(\sum_{\substack{x \in X, C=C_1}} p_{ni} + N_i)^2} \]

(2)

Calculating the Hessian matrix \( H_{k+1} \):

\[ H_{k+1} = H_k + \left(1 + \frac{q^{(k)}}{p^{(k)}q^{(k)}} H_k \frac{q^{(k)}}{p^{(k)}q^{(k)}}\right) \frac{p^{(k)}q^{(k)}}{p^{(k)}q^{(k)}} \]

(3)

Wherein, \( p^{(k)}=\theta^{(k)}），q^{(k)}=\nabla f(\theta^{(k)})-\nabla f(\theta^{(k)})）, p_{i2} \) indicates the number of examples which are correctly classified as category \( C_2 \), and Hessian matrix \( H_1 \) is the unit matrix.

Gradient descent direction

\[ d^{(k)} = -H_k \nabla f \]

(4)

It starts from \( x(k) \), searches along the direction of \( d(k) \) and works out the step size \( \lambda_k \), the equation is as follows;

\[ f(\theta^{(k)} + \lambda_k d^{(k)}) = \min_{\lambda \geq 0} f(\theta^{(k)} + \lambda \lambda_k d^{(k)}) \]

(5)

Update the characteristics weight vector \( \theta=(\theta_1, \theta_2, \ldots) \),

\[ \theta^{(k+1)} = \theta^{(k)} + \lambda_k d^{(k)} \]

(6)

**Step IV:** Substitute characteristic weight vector into objective function \( f(\theta) \) to judge whether the formula (14) is true.

\[ \| \nabla f(\theta^{(k+1)}) \| > \varepsilon \]

(7)

If true, it returns to step III and continues entering the iteration solution process; if not true, it obtains the optimization objective function calculated, and enter Step V.

**Step V:** Construct the diagnostic model of user electric theft behavior based on the optimization objective function,

\[ P(C = C_1 | x_j) = \frac{1}{1 + e^{-\theta^{(k)}}} \]

(8)

and test the test set, if the results meet the electric theft judgment requirement, complete constructing the model, if not, return to the step II and conduct the random assignment to \( \theta \).
4. Experiment and Analysis

For the purpose of verifying the effectiveness of method in this paper, the experiment takes the electricity information collection system of certain Electric Power Company of State Grid as the background implementation platform and the basic data source.

4.1 Data acquisition

50,000 samples in 2010–2017 withdrawn from the SG186 marketing business application system are corresponding to the user measuring index data obtained by the relevant users from the electricity information collection system, the user archive data obtained from SG186 marketing business application system, the abnormal matter data obtained from the acquisition terminal and the short-term electricity characteristic data in different industries. Besides, the meteorological data obtained from meteorological system are collected and summarized in combination with the holidays and festivals data, network crawling social data and policy information. The characteristics database selected initially contain 8 types of information: user's basic information, user's electricity information, information at measuring point, information in electric energy meter, electricity information, load information, meteorological information and social status information (40 associated factor data in total). The obtained data are preprocessed, including the data cleaning, data complementation, data definition and storage.

4.2 Assessment index

This paper selects the following assessment indexes in the experiment: Precision, Accuracy, Recall and $F_{\text{measure}}$, with the formula as follows:

\[
\text{Accuracy} = \frac{(T / P + F / N)}{N} \quad (9)
\]

\[
\text{Precision} = \frac{T / P}{T / P + F / P} \quad (10)
\]

\[
\text{Recall} = \frac{T / P}{(T / P + T / N)} \quad (11)
\]

\[
F_{\text{measure}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Recall} + \text{Precision}} \quad (12)
\]

The above formula, $TIF$ indicates whether the electric theft exists actually in the sample, and $P/N$ indicates whether the model forecasting has the electric theft.

4.3 Experimental results analysis

Based on the daily load forecasting data obtained and the user load data in the sample, the diagnostic model of user electric theft behavior is constructed. This paper optimizes the logistic regression forecasting by repeating training process for 4 times. The number of training samples reaches 20000, 6 characteristics weight vectors are obtained accumulatively for the deassign, the times of iterative optimization are limited to 200, and its successive accuracy rate and the change of iterations are shown in Figure 2. In Figure 2, three models with higher accuracy are selected to test the samples on the basis of accuracy and recall rate, with results shown in Table 1.

The curve a in figure 3 indicates that the electric theft judgment model contains the electricity load forecasting data, while the curve b indicates the electric theft judgment model doesn't contain the electricity load forecasting data. We could find that the integration of correlation data of electricity load forecasting and industry characteristics in the electric theft judgment model could improve the electric theft judgment effect of electric theft judgment model to some extent.
Based on the data in Table 1, the model 2 owns good diagnostic performance of user electric theft, hence, the model 2 is adopted as the final diagnostic performance of user electric theft behavior.

In order to describe the reasonability of the method in this paper, based on the above analysis, two kinds of methods in the electric theft prevention research hotspot field are selected as the comparison solution of electric theft behavior recognition, and the scheme in this paper is used as the third scheme as follows:

1) Option 1. Establish the identification method of electric theft behavior based on the gathered data and the normal distribution outlier algorithm\(^8\).

2) Option 2. Propose a kind of recognition method of electric theft behavior based on the electricity characteristics analysis by analyzing and handling with the obtained user electric theft behavior data reasonably\(^9\).

Based on the data in Table 1, the model 2 owns good diagnostic performance of user electric theft, hence, the model 2 is adopted as the final diagnostic performance of user electric theft behavior.

In order to describe the reasonability of the method in this paper, based on the above analysis, two kinds of methods in the electric theft prevention research hotspot field are selected as the comparison solution of electric theft behavior recognition, and the scheme in this paper is used as the third scheme as follows:

1) Option 1. Establish the identification method of electric theft behavior based on the gathered data and the normal distribution outlier algorithm\(^8\).

2) Option 2. Propose a kind of recognition method of electric theft behavior based on the electricity characteristics analysis by analyzing and handling with the obtained user electric theft behavior data reasonably\(^9\).

Based on the data in Table 1, the model 2 owns good diagnostic performance of user electric theft, hence, the model 2 is adopted as the final diagnostic performance of user electric theft behavior.

In order to describe the reasonability of the method in this paper, based on the above analysis, two kinds of methods in the electric theft prevention research hotspot field are selected as the comparison solution of electric theft behavior recognition, and the scheme in this paper is used as the third scheme as follows:

1) Option 1. Establish the identification method of electric theft behavior based on the gathered data and the normal distribution outlier algorithm\(^8\).

2) Option 2. Propose a kind of recognition method of electric theft behavior based on the electricity characteristics analysis by analyzing and handling with the obtained user electric theft behavior data reasonably\(^9\).

Based on the data in Table 1, the model 2 owns good diagnostic performance of user electric theft, hence, the model 2 is adopted as the final diagnostic performance of user electric theft behavior.

In order to describe the reasonability of the method in this paper, based on the above analysis, two kinds of methods in the electric theft prevention research hotspot field are selected as the comparison solution of electric theft behavior recognition, and the scheme in this paper is used as the third scheme as follows:

1) Option 1. Establish the identification method of electric theft behavior based on the gathered data and the normal distribution outlier algorithm\(^8\).

2) Option 2. Propose a kind of recognition method of electric theft behavior based on the electricity characteristics analysis by analyzing and handling with the obtained user electric theft behavior data reasonably\(^9\).

Based on the data in Table 1, the model 2 owns good diagnostic performance of user electric theft, hence, the model 2 is adopted as the final diagnostic performance of user electric theft behavior.

In order to describe the reasonability of the method in this paper, based on the above analysis, two kinds of methods in the electric theft prevention research hotspot field are selected as the comparison solution of electric theft behavior recognition, and the scheme in this paper is used as the third scheme as follows:

1) Option 1. Establish the identification method of electric theft behavior based on the gathered data and the normal distribution outlier algorithm\(^8\).

2) Option 2. Propose a kind of recognition method of electric theft behavior based on the electricity characteristics analysis by analyzing and handling with the obtained user electric theft behavior data reasonably\(^9\).
model training complexity is weaker than that of the first two kinds of models. Hence, the time spent in the model training is higher than that of the comparison algorithm, but its results in $F_{\text{measure}}$ and \textit{Accuracy} are superior to that of the comparison algorithm, which indicates the analysis method of associated factors of user electricity load adopted in this paper could excavate the associated factors of the user electricity load better and integrates the forecasting data and true data in the judgment model of electric theft, further improving the accuracy of model forecasting.

In the current research, the logistic regression algorithm is applied quite commonly\cite{10}, which indicates such method has significant advantage. Based on the data in Table 2, the electric theft recognition effect of method in this paper is superior to that of the comparison method as a whole, and also owns certain practicability in the model training and adaptability. The electric theft prevention work could be supplemented in practice.

5. Summary
In this paper, a method of identifying electric theft based on user short-term load forecasting and feature analysis is constructed on the background of the implementation platform of a power company's electricity information acquisition system. Firstly, the daily load forecasting data of users are obtained according to the historical data of users. And it is compared with the industry load characteristic data in the system to obtain the comparative data. Combined with various factors, the most relevant factors of different user groups are obtained through analysis. The above data are used as input of electric theft model, and a judgment model of electric theft is constructed based on logistic regression algorithm. The quasi-Newton algorithm is used to solve the optimal objective function, which can effectively reduce the number of iterations to obtain the diagnosis model of electric theft behavior and realize the judgment of electric theft.

This method is based on user's prediction data and real historical data. In the establishment of electric theft model, the factors of pre-judgment are integrated to improve the accuracy of identifying suspected users of electric theft. The algorithm analysis data show that this method is feasible and helpful to improve the detection of electric theft, which is of vital significance for maintaining normal work of electric theft.

References
[1] Dong Xianguang, Zhang Zhi, et al. The Research on User Short-term Electricity Load Forecasting for Judging Electric Theft[J]. International Conference on Communications, Information System and Computer Engineering, 2019.
[2] Liu Z M. Research on the Anti Power theft methods and strategies in Power Supply Company[D]. Beijing: North China Electric Power University, 2013
[3] Chen W Y, et al. Analysis of anti-stealing electric power based on big data technology[J]. Journal of electronic measurement and instrumentation, 2016,30(10):1558-1567.
[4] Tang D P, et al. A high performance reference digital energy meter for complex operating conditions[J]. Electrical Measurement & Instrumentation, 2018,55(22):140-146.
[5] Ye P, et al. On-line Calculating of Recovery Volume for Voltage Loss Fault of Three-phase Electric Energy Meter[J]. Electronic Technology & Software Engineering, 2018(13):229.
[6] Tian L, et al. Abnormal Power Consumption Analysis Based on Density-based Spatial Clustering of Applications with Noise in Power Systems[J]. Automation of Electric Power Systems, 2017(5) :64-70.
[7] Hu W, et al. Fault identification based on LOF and SVM for smart distribution network[J]. Electric Power Automation Equipment, 2016(6):7-12.
[8] Wang Xinxia, Wang ke, Jiao Dongxing, et al. Research on anti-stealing based on normal distribution outlier algorithm[J]. Electrotechnical Application, 2017, 36(7) : 60-65(in Chinese).
[9] Shi Y L, et al. Stealing Behavior Recognition Method Based on Electricity Characteristics Analysis[J]. Journal of Computer Research and Development, 2018,55(08):1599-1608.
[10] Li Z J. Small enterprise credit evaluation model based on hierarchical logistic regression[J]. Statistics & Decision, 2016(7):178-182(in Chinese)