A Comprehensive Method of Recovering Electricity Stealing Based on Data Mining

Jin Tingchao¹ᵃ, Wang Jianbo¹ᵇ, Cheng Shuya²ᶜ, Cai Hui²ᵈ, Xie Yue²ᵉ, Wang Ying²ᶠ

¹State Grid Zhejiang Electric Power Co., Ltd. Taizhou Power Supply Company, Taizhou 318000, China
²College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China

ᵃemail: jintingchao@163.com,ᵇemail: jianbosgcc_lq@163.com,
ᶜemail: 1820760593@qq.com,ᵈemail: caihui@cjlu.edu.cn,
ᵉemail: xieyue@cjlu.edu.cn,ᶠemail: sara.wy@163.com.

*Corresponding author’s e-mail: ᵇemail: caihui@cjlu.edu.cn

Abstract: Aiming at the current situation that the current methods of recovering the stolen electricity often use manual analysis and estimation, which leads to unsatisfactory results, this paper proposes a comprehensive electricity recovery method based on data mining. First of all, it is determined for the stealing time and other relevant data. Then, for users who can calculate the correction coefficients, the correction coefficient method is used to analyze the electricity theft electricity, and the users who meet the requirements of the prediction algorithm are analyzed using the corresponding forecast algorithm time series or neural network. Finally, the remaining users use the calibrated capacity of the electric energy meter instead of the actual load to analyze the stolen electricity. Time series forecasting uses an improved ARIMA algorithm based on white noise to predict the stolen electricity. The neural network uses an improved BP neural network based on independent variables to predict the stolen electricity. The research results show that this method takes into account the characteristics of a single power-stealing user while taking into account the user’s periodicity. It can reasonably recover the electricity of power-stealing users and provides a new idea for power recovery.

1. Introduction

When the electric energy metering device is operating under a fault condition, the calculated electricity data is incorrect. These incorrect data are likely to cause huge economic losses in the power supply department. At this time, how to correctly measure the stolen electricity is the problem the electricity department and the measurement department must face [1]. The existing electricity theft work is mainly based on the "Electricity Law of the People's Republic of China", "Electricity Inspection and Management Measures" and "Power Supply Business Rules", using manual analysis and estimation [2]. However, when manual analysis is used to estimate the stolen electricity, due to lack of solid data support and too flexible, the estimated stolen electricity may be insufficiently persuasive, difficult for users to accept, and serious user disputes may occur. Therefore, a reasonable
and comprehensive method for estimating the amount of electricity theft is proposed, and the electricity theft of users is analyzed from the user's historical electricity consumption data, so as to provide power companies with a scientific basis for the stolen electricity in time.

Literature [3-4] describes that the failure of the electric energy metering device will affect the accurate measurement of the user's electric energy. For this reason, a method for calculating and estimating the amount of electricity theft by using the multi-function electronic electric energy meter's split-phase electric energy recording function after a failure is proposed. But this method is too dependent on the function of the electric energy meter. Literature [5], based on the analysis and comparison of the data characteristics of the electric energy acquisition system and the energy management system, combined with calculation examples, and proposed two calculation methods for estimating the amount of electricity theft: line loss method and power method. This method is not comprehensive enough, and cannot be applied to dedicated transformer users without line loss. Literature [6-7] has studied a set of measurement failure event monitoring and alarm system for the frequent occurrence of provincial gateway metering failures in recent years, and gave specific software and hardware implementations. However, it is not possible to estimate the amount of electricity theft for a specific user. Reference [8] deeply analyzes the reasons for the calculation errors in the phasor analysis method of the electricity correction coefficient based on examples, and proposes a correction calculation method for estimating electricity theft based on the data of the metering automation system under the condition of obvious load fluctuations. It is applicable to cases with correction coefficients, but not applicable to cases without correction coefficients. Literature [9-10] proposed a high-accuracy measurement fault interval detection algorithm, which uses a robust random forest method to predict the electricity consumption in the fault interval. Due to the rolling forecasting method, cumulative errors may occur, causing the model to have a slightly poorer effect on data restoration in a long failure interval, and it cannot predict the daily electricity consumption for a long period of time. Literature [11] starts with the wiring of the electric energy meter, first analyzes the types of three-phase electric energy meter loss of voltage, and then proposes a new method of using the phase voltage without loss of voltage to track the electricity under the loss of voltage fault. Literature [12] uses a combination of cat group algorithm CSO and BP neural network for short-term load forecasting. Specifically, it uses four-dimensional data and adopts the cat group algorithm to optimize the weights and thresholds of the BP neural network to predict the electric quantity at an hourly time point. In [13], aiming at the difficulty of improving the accuracy of load forecasting models with multiple input data characteristics, the article proposes a parallel multi-model fusion hybrid neural network ultra-short-term load forecasting method. However, the one-dimensional data used in practice, that is, the user’s daily electricity consumption data, only one-dimensional data is used because the relevant power supply department system can be directly called without barriers, and other data such as average temperature cannot be imported into the relevant department data system in time. In addition, information security cannot be guaranteed when connected to the Internet, so this article uses the daily electricity consumption data of each user over the years.

The methods mentioned in the above literature mostly rely on hardware facilities and cannot be widely applied to all users. Based on the analysis of historical electricity consumption data, this paper adopts methods such as correction coefficient method, improved time series ARIMA model, improved BP neural network algorithm, and "Power Supply Business Rules" basis. After investigating the electricity theft situation and historical electricity consumption data of electricity users, a method for calculating electricity theft from multiple angles is proposed, the overall calculation process is designed, and the calculation model for estimating electricity theft is established. Finally, the user's historical data collected by a power supply company in a certain place is used to verify the rationality, comprehensiveness, effectiveness and practicability of the proposed method.

2. Research Ideas

After analyzing the power theft situation of users and the user’s historical power consumption, this paper adopts a method suitable for users to calculate the amount of power theft, and based on this,
designs a set of total procedures for calculating the amount of power that is applicable to all users. As shown in Figure 1. Proceed as follows:

First, if the electricity stealing user uses an illegal electricity recharge card to steal electricity, it will be calculated based on the difference between the electricity read on the spot and the electricity final settlement of the power supply company.

Second, if the number of days of stealing electricity cannot be judged on the spot, according to the "Power Supply Business Rules", "the stealing time cannot be checked, at least 180 days", the stealing time is calculated as 180 days.

Third, analyze the electricity theft situation of the electricity theft user. If the electricity theft user has a correction coefficient, the actual electricity consumption is determined by the correction coefficient, and the actual electricity consumption is added to calculate the electricity theft on the day of the deadline. The traditional correction coefficient method is adopted, and this article will not introduce it in detail.

Fourth, if the user’s historical electricity consumption meets the forecasting requirements of the forecasting algorithm, the improved time series ARIMA model or the improved BP neural network algorithm is used to predict electricity consumption, and the electricity theft is calculated based on the predicted electricity consumption and the actual electricity consumption.

Fifth, if the user who steals electricity does not meet the judgment conditions of the first four steps, the standard capacity indicated by the calibrated current value of the billing electric energy meter will replace the actual load. According to the "Power Supply Business Rules", "power consumption for production and operation is 12 hours or 24 hours a day, and electricity consumption for life is 6 hours a day", and finally the amount of electricity stealing is calculated based on user load, power stealing time and days of stealing electricity.

![Fig.1 General idea of electricity compensation](image)

3. Principle analysis

3.1. BP neural network algorithm

BP neural network [14-15] also known as back-propagation neural network, is a method of continuously modifying the network weights and thresholds through the training of sample data to make the error function drop along the negative gradient direction and approach the desired output. BP neural network is mainly composed of input layer, hidden layer and output layer.

The BP neural network model is generally used for classification. The change required for regression prediction is to remove the third layer of nonlinear conversion, or to replace the nonlinear
activation function $\sigma(x)$ with a function of $f(x) = x$. The main reason for this is that the output range of the general non-linear activation function is between 0 and 1, which is approximately equal to normalization, and the model prediction requires unprocessed values.

For the hidden layer, as long as there are enough nodes in the hidden layer, a nonlinear function can be approximated with arbitrary precision. However, too many neurons in the hidden layer will increase the amount of network calculation and easily cause overfitting. Therefore, this article refers to the following empirical formula in selecting the number of neurons in the hidden layer:

$$l = \sqrt{n} + m + a$$ (1)

Among them, $n$ is the number of neurons in the input layer, $m$ is the number of neurons in the output layer, and $a$ is a constant between $[1,10]$. According to formula (1), the interval range of the number of neurons can be calculated, and the number of hidden layer neurons selected in this article is 5.

3.2. Time series ARIMA algorithm

If a sequence $\{X_t\}$ becomes a stationary sequence after d-order difference, and the ARMA model can be used to model the differentiated stationary sequence, then the model structure of the sequence $\{X_t\}$ is called a single-product autoregressive moving average model [16]. The model is also called ARIMA($p,d,q$), where $d$ is the difference order, $p$ is the autoregressive order, and $q$ is the moving average order. The steps of the time series ARIMA model are:

- The white noise test is used to determine whether the time series ARIMA model can be used.
- Use ADF to test the stationarity of the sequence.
- Choose the combination of AIC criterion and BIC criterion to determine the order.
- Model fit and prediction.
- Use white noise to check whether all the sequence information is extracted.

4. Calculation process and algorithm for estimating the amount of electricity theft

Based on correction coefficient method, time series ARIMA model, BP neural network algorithm, and "Power Supply Business Rules" basis and other methods, establish a calculation model for estimating the amount of electricity theft. Complete the calculation of the entire model based on data analysis software.

The input data of the model are: the voltage and current data of the stealing user, the stealing time, the power factor $\phi$ and the historical electricity consumption data. If there is no data on voltage, current and power-stealing time, it will be indicated by ‘-1’ for software identification.

4.1. Preliminary judgment on estimating electricity theft

First of all, when judging whether the user is stealing electricity, it can directly determine whether the electricity stealing depends on the length of time, and determine the number of days of stealing electricity. If you do not rely on the duration of the electricity theft, the difference between the electricity read on the spot and the final settlement of the power supply company is determined. If the time for stealing electricity cannot be determined, it is calculated as 180 days.

Then, the correction coefficient is calculated based on the voltage, current, and power factor data in the electricity consumption data of the electricity theft user to determine the amount of electricity that needs to be supplemented.

If the electric energy meter at the power-stealing user is a single-phase meter with no voltage and current data, it is necessary to determine whether the historical power consumption of the user meets the requirements of the prediction algorithm.

If it does not meet the requirements of the forecasting algorithm, the actual load is replaced by the capacity indicated by the calibrated current value of the meter. The rated power of the meter for low-voltage users is 8kW.
4.2. Algorithm predicts electricity consumption

4.2.1. Improved BP neural network algorithm prediction

If the BP neural network prediction is regarded as the mapping relationship between the independent variable and the dependent variable, when the data is one-dimensional data, the independent variable is generally an increasing number sequence: ‘1,2,3,...’. However, considering the characteristics of the electricity consumption sequence itself, one year is regarded as a cycle, and each year has an increasing increment. The independent variable of the electricity consumption data sequence from 2017 to 2018 is defined as: "1, 2, 3... 364, 365, 1.1, 1.2...364.1, 365.1". At this time, the BP neural network algorithm can be used to extract the characteristic curve of the data sequence.

For the daily historical electricity consumption data 'extract characteristic curve', if the historical electricity consumption data does not have obvious periodic characteristics, the prediction result will inevitably have a large error; if the electricity consumption of the user is 0, the algorithm analysis will also be affected. Therefore, users with a daily electricity consumption of 0, which account for more than 20%, and users with a maximum value of less than 1200 in the autocorrelation function are both filtered out. Both thresholds are determined by hundreds of users.

Due to the small number of power-stealing users, and there is no definite and accurate information about the power-stealing time and the electricity of power-stealing users, this section uses normal users in a certain area of Zhejiang Province for analysis in the modeling and analysis process. The advantage is that the power stealing time can be customized and determined, and the power stealing has an accurate true value, which greatly improves the accuracy and feasibility of modeling and prediction.

The method for determining the threshold of the proportion of days when the daily electricity consumption is 0: First, establish a BP neural network model for 200 normal users and predict the results, and calculate the fitting error rate and prediction error rate of 200 users. Next, if the threshold is 10%, the users whose mark exceeds the threshold of 10% are 0, and the probability of the prediction error rate of the users whose mark is 0 is too large (specific error rate>30%) is calculated. The same method calculates the probability that the prediction error rate of 200 normal users is too large when the threshold is 20% to 50%. Finally, the selected threshold is 20%. At this time, the probability that the prediction error rate of the user marked as 0 is too large is 90.2%.

The method for determining that the maximum value of the autocorrelation function is less than the threshold 1200: First, establish a BP neural network model for 200 normal users and predict the results, and calculate the fitting error rate, prediction error rate and autocorrelation function value. Next, arrange the maximum value of the autocorrelation function among all users from small to large, and calculate the probability that the prediction error rate of users below the threshold is too large when the threshold is 500 to 5000. Finally, the selected threshold is 1200. At this time, the probability that the prediction error rate of users who are less than the threshold is too large is 85.2%.

The moving average filter is used to preprocess the data to remove fluctuations in a small range and improve the accuracy of the algorithm. Since the cumulative power is predicted at the end, the prediction result is not affected.

For the establishment of the BP neural network prediction model: First, determine the number of nodes. Since the input data is a defined one-dimensional independent variable, the number of nodes in the input layer is 1. The number of neurons in the hidden layer is 5 (derived from formula (1)). Since the output data is historical daily electricity consumption, the number of nodes in the output layer is 1. Second, determine that the number of network iterations Epochs is 3000, the expected error Goal is 0.00001, and the learning rate γ is 0.05. Use MATLAB’s neural network toolbox for network training to complete the modeling; finally, use the trained network to predict the future electricity consumption of users who steal electricity.

The BP neural network prediction algorithm is verified by the historical daily power consumption of normal users. First, model the daily electricity consumption for 730 days from January 1, 2017 to December 31, 2018, and obtain a characteristic curve. As shown in the fitted values in Figure 2, the
fitted values for 2017 and 2018 are shown in the figure for the same characteristic curve because the cycle is the year. The forecast value is the daily electricity consumption for a total of 31 days in January 2019 (as shown in Figure 2. The cumulative electricity consumption fitting error rate is 0.45%, and the forecast error rate is 0.93%.

But not every user can use the BP neural network to extract representative characteristic curves and make predictions. 150 normal users are used for BP neural network prediction, and the cumulative power consumption fitting error rate and prediction error rate are obtained. The probability distribution diagram of the fitting error rate is shown in Figure 3(a). It can be seen from Figure 3(a) that the fitting error rate of a total of 35 users in the fourth and fifth categories is relatively large, and most of the corresponding prediction error rates are relatively large (Figure 3(b)). In Figure 3(b), the error rate greater than 100% is calculated as 100%. Looking at the prediction error rates of the first three categories in Figure 3(a), half of the users have a large error rate.

In summary, the use of BP neural network to extract representative characteristic curves for prediction requires greater user requirements, and this method can be used as an auxiliary method. But if it meets the BP neural network prediction, it can predict the electricity consumption for a long time or even a year.
4.2.2. Improved time series ARIMA model forecasting

Time series ARIMA model prediction also considers two judgment conditions: screening out users who steal electricity with a daily electricity consumption of 0, which accounts for more than 20%, and users whose maximum value in the autocorrelation function is less than 1200. It is also found that the fitting data is too low during the analysis process and the prediction result error is too large. The judgment condition is added: if the daily power consumption in the fitting data is less than the maximum value * 0.1, it exceeds 50% of the total days of the fitting data. It does not apply to time series ARIMA model forecasting.

The method of determining the threshold in the judgment conditions: For normal users using the time series ARIMA model to predict the 28 users whose fitting values are too large, draw the maximum value *0.1~0.4 threshold line in each graph, and finally determine the most common point Threshold, the threshold is the maximum value*0.1. Establish a time series ARIMA model for 120 normal users and predict the results, and calculate the prediction error rate for 120 users; if the number of days with daily electricity consumption less than the maximum *0.1 exceeds 50% of the total number of days in the fitted data, the mark exceeds the threshold 50% Calculate the probability that the prediction error rate of users marked as 0 is too large (specific error rate>30%). The same method calculates the probability that the prediction error rate of 120 normal users is too large when the threshold is 20%-50%; finally, the threshold is selected as 50%.

Using the BP neural network to predict the historical daily power consumption of the same user, a total of 730 days of daily power consumption from January 1, 2017 to December 31, 2018, was modeled according to the time series ARIMA model steps. Make sure the model is ARIMA(2,0,1). The model predicts in January 2019 (Figure 4(a)), the prediction error rate is 9.97%, which is greater than the BP neural network prediction error value. However, the same 150 normal users are modeled and predicted by the time series ARIMA model. After passing the judgment conditions, 98 users meet the time series modeling, and 81 users have a fitting error rate of less than 0.3 (Figure 4(b)) , the probability is 82.6%, which is more stable than the BP neural network. But the forecast error rate will increase with the forecast time, so it is suitable for short-term forecasting.
4.2.3. Comprehensive algorithm prediction and analysis

Through the analysis of the two algorithms, it is finally determined that when short-term forecasting, the combined algorithm of time series ARIMA model as the main and BP neural network algorithm is used to predict electricity consumption. The main overall process is shown in Figure 5.

The model uses the historical daily electricity consumption of 150 normal users, and predicts the electricity consumption through the overall algorithm flow of Figure 5, and finally obtains the prediction error rate of 150 users (the point represented by the comprehensive algorithm in Figure 6). The BP neural network and the time series ARIMA model were used for 150 users respectively, and a total of 730 days of daily electricity consumption modeling from January 1, 2017 to December 31, 2018 was used to predict January 2019, and the forecast fitting was obtained Rate (as represented by the BP neural network and the time series ARIMA model in Figure 6). It can be seen from Figure 6 that the comprehensive algorithm greatly reduces the prediction error rate, and basically does not exceed the threshold line. Among 150 uses, 37 uses calculate and forecast electricity consumption based on capacity, which is represented by a -10% error rate in the figure. 98 uses use time series ARIMA model to predict electricity consumption, and 15 uses use BP neural network forecast energy.
used. For the 113 users who use algorithms to predict electricity consumption, 97 users have a prediction error rate of less than 30%, with a probability of 85.8%, as shown in Figure 7.

Fig.5 Total algorithm forecast total electricity consumption process

Fig.6 Prediction fitting rate of algorithm for electricity consumption prediction
5. Case Analysis and Discussion

In order to verify the accuracy of the algorithm in the actual estimation of electricity theft, seven actual electricity theft users with complete and relatively correct estimated electricity theft information in a certain area of Zhejiang Province are used to predict electricity consumption and calculate the electricity theft using the algorithm in this paper.

Tab.1 Comparison between the forecast and the actual power demand of the electricity stealing users

| User ID | Electricity stealing time period/day | Specific algorithm   | Predicted electricity stolen/kW.h | Actual electricity stolen/kW.h |
|---------|-------------------------------------|----------------------|----------------------------------|-------------------------------|
| 6249    | 20180625-20180919                   | BP neural network    | 814.04                           | 2kW*6h*87day=1044            |
| 7180    | 20181201-20181205                   | BP neural network    | 43.71                            | 3.92kW*6h*5day=118           |
| 8193    | 20181210-20181218                   | ARIMA model          | 120.40                           | 2.68kW*6h*9day=145           |
| 2730    | 20181028-20190125                   | ARIMA model          | 720.82                           | 1.5kW*6h*90day=810           |
| 1505    | 20181010-20181218                   | Capacity instead of load | 3250.26                      | 1.5kW*6h*70day=630          |
| 9113    | 20181220-20190114                   | BP neural network    | 588.53                           | 17.22kW.h/天*26day=447       |
| 2453    | 20181122-20190101                   | Capacity instead of load | 1309.45                       | 3.8kW*6h*41day=934.8         |

Although the actual electricity theft verified on the spot cannot be absolutely correct, it has reference value. As shown in Table 1, there is a discrepancy between the predicted value of a few users and the actual value verified on the spot, and the predicted value of most users is similar to the actual value verified on the spot, and the error rate is less than 25%. This algorithm can be considered to be scientific. The reason for the excessive error of the user 1505 is that the capacity of the electric energy meter is 8kW, and the power of the stealing appliance is determined to be 1.5kW in the actual stealing, and the defined capacity is different. The comprehensive algorithm can directly use the data in the relevant department system, use fewer types of data, establish a certain correctness and scientific model, and can be applied to a variety of situations to achieve effective practical applications. If it is possible to extract the electricity consumption data at a time point of 24 hours on the same date in the past years, it is also possible to make a point-in-time electricity consumption forecast. Because the relevant departments have not yet needed time-point forecasts, the comprehensive algorithm only makes daily electricity forecasts.

6. Conclusions

This article proposes a new comprehensive method for estimating the amount of electricity theft from users, establishing a model and verification based on real data. The rationality of the actual electricity theft verification method combined with field investigation. The research results show that the method breaks through the limitations of a single algorithm through multiple analyses such as comprehensive correction coefficients, improved time series ARIMA model, improved BP neural network, and "Power Supply Business Rules" basis, and can estimate the amount of electricity theft for all users who steal electricity Reasonable analysis. In practice, the method of on-site inquiry is adopted. The on-site inquiry method is to analyze the electricity of each household after determining the line or
station area with abnormal line loss based on the line loss. The method in this paper effectively eases the workload of the on-site query method, and accurately locates users with abnormal power consumption, takes full advantage of the huge data of the power supply department, and reduces the workload. Moreover, it is possible to directly use fewer data types in the relevant departmental system to establish a certain correctness and scientific model. And the model can be applied to a variety of situations to achieve effective practical applications. In addition, future work needs to consider more methods at the same time, discover more suitable algorithms to integrate into the overall method, and improve the effectiveness and stability of the method. At the same time, in terms of stealing electricity, relevant departments should also formulate relevant enterprise standards and corresponding legislative measures.

Acknowledgments
Supported by the Youth Science Foundation of Natural Science Foundation of Zhejiang Province (LQ17E070003)

References
[1] Huang Hongqiao, Yang Maotao, Wang Haiyuan, et al. Calculation method of power recovery after power metering device failure [J]. China Metrology, 2017, 5:118-120.
[2] Huang Yulong, Wu Bin, Liu Jing. Calculation of retroactive power in electricity theft by single-phase energy measurement [J]. Journal of Shijiazhuang University of Applied Technology, 2018, 30 (2): 31-33.
[3] Ding Pengcheng, Deng Yuehua, Deng Zhihai, et al. Discussion on calculation method of power recovery after power metering device failure [J]. Electrical technology, 2013, 4: 25-27.
[4] Chen Jedong. New Method Of Supplementing Electric Volume after Voltage—Loss of Electric Energy Measuring Unit [J]. Electrical technology, 2005, 10:64-65.
[5] SHEN Li, LEI Ming, LI Fan, et al. Research on the Calculation Method of Reclaiming the Missing Energy in Power Grid System [J]. Electrical Measurement & Instrumentation, 2013, 50(11A):42-46.
[6] Cao Hongmeng, pan Zhihong. Method of power recovery at power grid gateway [J]. China power, 2003, 36 (5): 68
[7] YANG Shuai, CHEN Xiang qun, ZHAO Dan, et al. Research on Alarming Instrument and Master Station for Gateway Energy Measuring Failures [J]. Electrical Measurement & Instrumentation , 2014, 51(13):21-28.
[8] LIU Fen. Correction Calculation Method of Electric Energy Compensation Based on Data of Metering Automation System [J]. Electrical technology, 2019, 16:57-58.
[9] CHEN Guangkai, CHEN Shuhong, PAN Wei, et al. Electricity Consumption Rectification Algorithm Based on Random Forest and Rolling Prediction[J]. Smart Power, 2018, 46(12):45-50.
[10] CHEN Lvpeng, YIN Linfei, Yu Tao, et al. Short-term Power Load Forecasting Based on Deep Forest Algorithm[J]. Electric Power Construction, 2018, 39(11):42-50.
[11] CHEN Jinyou, Peng Zhaohuang, Cai Chunhuang. Online Calculation of Make-up Electrical Power During Voltage-loss of Three-phase Watt-hour Meter [J]. Automation of Electric Power Systems, 2013, 37(19):100-104.
[12] Wang Kejie, Zhang Rui. Research on short-term power load forecasting method based on improved BP neural network [J]. Electrical Measurement & Instrumentation, 2019, 56(24):1-7.
[13] ZHUANG Jiayi, YANG Guohua, ZHENG Haofeng, et al. Ultra-short-term Load Forecasting Using Hybrid Neural Network Based on Parallel Multi-model Combination[J]. Electric Power Construction, 2020, 41(10):1-8.
[14] LI Ya, LIU Liping, LI Baiqing, et al. Calculation of Line Loss Rate in Transformer District Based on Improved K-Means Clustering Algorithm and BP Neural Network [J], 2016, 36(17):4543-4551.

[15] SONG Jian, SHU Hongchun, DONG Jun, et al. Comprehensive Load Forecast Based on GM(1,1) and BP Neural Network[J], Electric Power Construction, 2020, 41(5):75-80.

[16] Wu Yuxia, Wen Xin. Short term stock price prediction based on ARIMA model [J]. Statistics and decision, 2016, 000 (023): 83-86.