Research Article

Hazard Trend Identification Model Based on Statistical Analysis of Abnormal Power Generation Behavior Data

Gaojun Xu 1, Xusheng Qian 1, Xiaodong Li 2, and Weijiang Wu 1

1State Grid Jiangsu Electric Power Company Limited Marketing Service Center, Nanjing 210019, Jiangsu, China
2Nanjing Power Supply Company, State Grid Jiangsu Electric Power Company Limited, Nanjing 210000, Jiangsu, China

Correspondence should be addressed to Gaojun Xu; 2010412211@hbut.edu.cn

Received 13 July 2022; Revised 28 July 2022; Accepted 3 August 2022; Published 31 August 2022

Copyright © 2022 Gaojun Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to solve the problem of abnormal data identification for key indicators with the deepening and development of power enterprise reform, this paper proposes a method of dangerous trend identification model based on a statistical analysis of abnormal power generation behavior data. The method includes a data access scheme, feature extraction scheme, and anomaly detection algorithm. The experimental results show that the proportion of users whose electricity consumption behavior conforms to the peak period electricity consumption $>$ normal period electricity consumption $>$ valley period electricity consumption exceeds 90%. More than 85% of users’ electricity consumption behavior is in line with the proportion of electricity consumption that is less than 0.25 in millet hours. The proportion of users whose fluctuation coefficient of electricity consumption in the valley period is less than 1 exceeds 85%, and 99.9% of users’ fluctuation coefficient of electricity consumption in the valley period is less than 5, which proves that abnormal power generation behavior data and abnormal power consumption data can bring early warning to some dangerous power consumption behaviors. The statistical analysis model of abnormal power generation behavior data can play a positive role in the identification of dangerous trends.

1. Introduction

Power statistical data are an important basis for the government to implement macrocontrol, strengthen professional supervision, formulate national economic planning, energy, power, and other special planning, and an important reference for studying and judging the economic and energy situation and formulating industrial policies [1]. We should attach great importance to the statistical work of power regulation. We should not only systematically collect, sort out, and share power statistical information but also use statistical information and data to carry out in-depth comprehensive analysis and special research, so as to better serve the energy and economic decision-making, industry management and supervision, and “deregulation service” reform. With the continuous deepening and development of the reform of power enterprises, energy conservation and consumption reduction in high energy-consuming industries is very important to solve energy problems and is also the fundamental guarantee to achieve the energy conservation and consumption reduction goal of the 13th five-year plan [2]. The national energy administration and government departments attach great importance to the authenticity and effectiveness of energy conservation and consumption reduction-related indicators when conducting power supervision and statistics. As the main bodies of China’s power market are relatively concentrated, the legal awareness of market subjects is relatively weak, and the market supervision system is not perfect; the market power risk has become the main risk of China’s power market, showing the characteristics of many forms, difficult prevention, difficult supervision, and great influence, which greatly increases the difficulty of power market construction. The current power generation statistical model lacks an effective data quality control system, and even some power plants report false values to meet the requirements of energy conservation, consumption reduction, and carbon emission, which may form abnormal data, affecting the cognition of relevant departments and enterprises on power generation. Therefore, it is necessary to carry out abnormal data
identification for key indicators to monitor the actual operation level of the power plant. For example, the influence coefficient of a 1% change in the auxiliary power rate on coal consumption for power supply is 3.499%, and each 1% decrease in the load rate will affect the increase of the auxiliary power rate by 0.06%. Therefore, it is of great significance to take the power consumption rate of the power plant as the key index, analyze the potential law of the power consumption rate of the power plant by using the historical data, and identify whether the reported value is abnormal, so as to identify the dangerous trend in all aspects.

2. Literature Review

Chinese and foreign scholars have put forward some identification methods and preventive measures against violations in the power market. Ji and Bai and others built a regulatory index system to judge the use of market power on the generation side [3]. Kulikov and others built a framework for identifying violations in the whole process of electricity market transactions [4]. Li et al., respectively, used game theory, genetic algorithm, and joint evolutionary optimization to determine whether a single power producer participated in collusion [5]. Yang et al. used econometric methods to analyze the collusive bidding behavior between the two major power generation companies in the UK [6]. Phanab et al. proposed that anomaly detection methods for time series include frequency-based methods and machine learning-based methods [7]. The former has the advantage of high processing efficiency and can detect novel patterns in linear time and space. However, due to the use of time series discretization technology, meaningful data may be lost in the process of conversion. The basic idea of the latter is to predict the sequence and determine the abnormal value through the difference between the predicted value and the actual value. It mainly includes two categories: the support vector machine method and the neural network method. As a kind of dynamic neural network, the nonlinear active autoregressive model has the functions of memory and association of historical information and strong adaptability to the time-varying characteristics of time series. At present, it has been used in the fields of wind speed and photovoltaic power prediction. However, these studies have the following problems and shortcomings: first, they only pay attention to the establishment of market supervision indicators but do not specifically explain how to use indicators to identify abnormal trading behavior. Second, the market model has been greatly simplified, and only a small number of market members are concerned, but there is little research on dealing with large-scale datasets generated by actual market operations. At present, there is a lack of research on abnormal power consumption behavior detection based on the power consumption data of users (residents, communities, buildings, enterprises, etc.). However, the existing anomaly detection methods for equipment operation status mainly design data mining analysis algorithms and match them based on the known characteristics of equipment operation status reflected in the data. They do not use the method of automatically learning and summarizing universal laws from massive data, so they cannot automatically identify the abnormal behavior characteristics different from universal laws or unknown [8].

Facing the big data early warning demand of social security risk, this paper studies the problem of abnormal trend identification and early warning based on the data mining and analysis of people’s daily behavior (such as power consumption behavior) and puts forward the characteristic identification method and the early warning model of abnormal power consumption behavior. Combined with the structure of the power system and the characteristics of power consumption data, the schemes of data security access, distributed storage, parallel computing analysis, and visual display are designed, and the construction of power consumption anomaly analysis and early warning system is completed [9]. The technical implementation scheme of electricity behavior anomaly detection and the public security risk early warning method based on power big data analysis is as follows.

We read the user profile, electric energy representation value, electricity consumption in each period, and other data required for power consumption behavior analysis from the relational database of the power grid data acquisition system and use the security access technology to store the data in the HBase distributed database of the Hadoop platform in real time. Then, the MapReduce algorithm is designed to preprocess the data, extract the characteristics of power consumption behavior, cluster based on multi features and time series, and summarize the general laws of residential power consumption behavior [10].

3. Method

3.1. Data Item Extraction. The user profile database and power operation database are the core of the relational database in the power grid data acquisition system. In the data access stage, first, we select the data items to be read in the analysis and early warning stage, including the file data such as the meter number, user identity, and power consumption location, as well as the current positive indication value, electricity consumption in each period, and other data.

3.2. Construction of Intermediate Warehouse. We set up an interface server and a relational database in the power intranet. The data are synchronized from the database of the power grid data acquisition system to the intermediate database through the goldengate data synchronization technology. Among them, the archive data in the user archive database and the historical operation data in the power operation database are synchronized to the intermediate database at one time by using the data pump technology, and the newly collected power consumption data of the power grid data acquisition system are synchronously updated to the intermediate database every day [11].

3.3. Data Transmission Method. The data transmission channel adopts strong isolation technology [12]. This
technology is technology that allows the external network to access the intranet database, but the database seen by the external network is only a virtual database. At the same time, access to the intranet database, but the database seen by the technology is technology that allows the external network to realize interaccess. If the social security risk precontrol system based on abnormal electricity use behavior detection is deployed on the private network of political, legal, and public security departments, a special data transmission line is required. One end is connected to the intermediate library deployed in the power intranet and the firewall outside the IPS and other security systems through the above steps, and the other end is connected to the corresponding private network through the security access platform (dedicated to the public security network) [13]. The intermediate database is configured as the data source, and only one-way transmission of data points (intermediate database) to points (database of power abnormal analysis system) is allowed.

### 3.4. Distributed Data Storage

After associating the file data read from the intermediate database with the operation data according to the user ID, the historical data are stored in the HBase distributed database in the key value mode. Since the user ID of the user meter is unique, the digital string spliced between the user ID and the acquisition time (date) is used as the key, and the acquisition terminal number and the cumulative value of electricity consumption in each period are used as the column value. When new data are entered, we first generate data records in key-value format and then execute the query and insert each into the data block where the same key is located [14].

### 3.5. Feature Extraction Scheme

During equipment condition monitoring, clear fault characteristics can be used for anomaly identification. In contrast, there is no clear definition of abnormal power consumption behavior before. The residential electricity data collected in a city for one month (30 days) are selected as the training set, and the feature extraction scheme is as follows:

**Step 1.** Data preprocessing to eliminate the data with class B abnormal characteristics.

**Step 2.** Calculate the power consumption in each period of the day.

The power department divides a day into three periods according to the electricity consumption, including the peak period, normal period, and valley period. For residential power consumption, the peak period is the period with the highest power demand, generally in the early morning and at night (taking the area in Dongba district as an example, the peak period is from 6:00 to 10:00 and from 18:00 to 22:00). The valley period is the period with the lowest power demand, generally from midnight to early morning (22:00 to 6:00 in the East 8th District). Electricity demand in the normal period is between the peak period and the valley period. For each meter that is not excluded, the electricity consumption in three periods of each day is calculated as follows:

For each meter, the positive electric energy indication value of each day in a month is $\sum_{\text{cur totali}}$, where $i \in [1, 2, 30]$, and then, the daily power consumption is

$A_i = \sum_{\text{cur totali}} \sum_{i=1}^{i+1} \sum_{\text{cur peaki}}$, If $A_i < 0$, or the sum of the indicated electricity consumption in three periods of a day is not equal to $\sum_{\text{cur totali}}$, the data of the corresponding electricity meter shall be deleted.

**Step 3.** Calculate the proportion of electricity consumption in each period of the day.

The proportion of peak, average, and valley electricity consumption in the total electricity consumption of the day is calculated as follows:

- Power consumption during peak hours: $A_{i,\text{peak}} = \sum_{\text{cur peaki+1}} \sum_{\text{cur peaki}}$;
- Electricity consumption in the normal period: $A_{i,\text{normal}} = \sum_{\text{cur normali+1}} \sum_{\text{cur normali}}$;
- Electricity consumption during the valley period: $A_{i,\text{valley}} = \sum_{\text{cur valleyi+1}} \sum_{\text{cur valleyi}}$.

**Step 4.** Calculate the fluctuation coefficient of daily power consumption, power consumption in each period, and the proportion of power consumption in each period.

The average value of each parameter is calculated. Taking the daily power consumption as an example, the monthly average value is
The variance of each parameter is calculated, and taking the power consumption in the peak period as an example, the variance is

\[
\text{var}(A_{\text{peak}}) = \sum_{i=1}^{30} (A_{\text{peak}}^i - \text{mean}(A_{\text{peak}}))^2.
\]  

(2)

The fluctuation coefficient of each parameter is calculated. Taking the proportion of electricity consumption in the valley period as an example, the fluctuation coefficient is as follows:

\[
\text{cv}_{\text{ratio_valley}} = \frac{\sqrt{\text{var}(\text{ratio}_{\text{valley}})}}{\text{mean}(\text{ratio}_{\text{valley}})}.
\]

(3)

3.6. Anomaly Detection Algorithm. The abnormal detection process of electric behavior is divided into two parts: abnormal behavior detection and abnormal feature analysis. The former stage is used to identify abnormal power consumption behavior, and the latter stage is used to match the characteristics of abnormal power consumption behavior with those of criminals in actual cases to generate accurate early warning information.

Step 5. The MapReduce computing model is used in the abnormal behavior discovery process.

(1) We perform two types of map tasks, respectively; that is, we calculate the daily power consumption, power consumption in each period of the day, and the proportion of power consumption in each period of the day of each meter for the latest week and the latest month, as well as the mean value, variance, and fluctuation coefficient of these parameters in the selected period.

(2) The reduce process performs clustering. The selected clustering features are shown in Table 2. Clustering shows that the data characteristics are different from those of most meters [15].

(3) Using the LOF (local outlier factor) algorithm for further anomaly detection, the highest abnormal level of power consumption behavior can be found according to the score value distribution.

Step 6. In the abnormal feature analysis process, the abnormal behavior data extracted from the abnormal behavior discovery process and the existing feature database are calculated to confirm whether the abnormal power consumption behavior meets the alarm conditions and outputs the accurate early warning type [16].

(1) We verify whether the fluctuation coefficient of night power consumption and its proportion increases sharply at the same time. As shown in Figure 2, the gang gathered at the home of a criminal for two consecutive days before the crime, and its midnight (valley time) power consumption increased sharply, while there was no significant change in the peak, period and peacetime power consumption, resulting in a sudden increase in the valley time power consumption fluctuation coefficient and the valley time power consumption proportion fluctuation coefficient of the meter. The warning type generated by such an abnormal behavior is abnormal power consumption.
consumption at night. The judgment conditions are as follows: for users whose fluctuation coefficient of electricity consumption in the valley period and the proportion of electricity consumption in the valley period are greater than 50%, we calculate the proportional relationship between the electricity consumption in the valley period on the day of the sudden change and the previous day and the electricity consumption in the other two periods, \( A_{i,\text{valley}}/(A_{i,\text{peak}}+A_{i,\text{normal}}) \). If the value on the day is greater than twice the value on the previous day, this kind of alarm information will be generated.

(2) The fluctuation coefficient of electricity consumption in each period continues to increase. As shown in Figure 3, a criminal used high-power electrical appliances to make crime tools at home during the preparation stage. The electricity consumption in each period has increased significantly and fluctuated sharply within the time window. The warning type generated by such an abnormal behavior is the abnormal fluctuation of power consumption throughout the day [17]. The specific judgment conditions are as follows: the whole day power consumption fluctuation coefficient \( CV_{\text{cur_total}} \) is calculated. If the value increases for three consecutive days and the cumulative increase exceeds 100%, this kind of alarm information will be generated.

(3) The electricity consumption decreases sharply (even approaches zero) and continues to rise. As shown in Figure 4, on the eve of the crime, the perpetrator left his home and joined the crime gang. Its power consumption is close to 0, and there is no fluctuation in power consumption in each period [18]. The alert type generated by such an abnormal behavior is suspected of leaving home.

3.7. Alarm Visualization Scheme. Visualization of early warning information: the visualization interface is mainly based on the application of GIS Geographic Information System. Based on the power consumption address information (place name, longitude, and latitude) in the file data, the location of the electricity meter that generates the abnormal power consumption early warning is displayed on the GIS map, and all the power consumption early warning information within the jurisdiction is displayed on the page in the form of a list. The contents displayed in the list include alarm level, user name, and alarm time. In addition, the location of the alarm is visually displayed on the map. You can click the alarm location to view the user details, alarm type, power consumption curve, and other specific information.

Visualization of regional early warning: we calculate regional abnormal electricity consumption indicators from the two dimensions of judging whether the location of abnormal electricity consumption meters is concentrated in a certain area and whether the electricity consumption of suspected personnel in each key monitoring area is

| Serial number | Feature name       | Feature definition                        |
|---------------|--------------------|-------------------------------------------|
| 1             | \( CV_{\text{cur_total}} \) | Fluctuation coefficient of power consumption throughout the day |
| 2             | \( CV_{\text{cur_peak}} \) | Fluctuation coefficient of power consumption in the peak period |
| 3             | \( CV_{\text{cur_normal}} \) | Fluctuation coefficient of electricity consumption in the normal period |
| 4             | \( CV_{\text{cur_valley}} \) | Fluctuation coefficient of electricity consumption in the valley period |
| 5             | \( CV_{\text{ratio_peak}} \) | Fluctuation coefficient of electricity consumption in the peak period |
| 6             | \( CV_{\text{ratio_normal}} \) | Fluctuation coefficient of electricity consumption in the normal period |
| 7             | \( CV_{\text{ratio_valley}} \) | Fluctuation coefficient of electricity consumption in the valley period |

Figure 2: Abnormal characteristics of the sudden increase of power consumption at night.

Figure 3: Abnormal characteristics of sudden increase and sharp fluctuation of electricity consumption throughout the day.

Figure 4: Abnormal characteristic of the sudden decrease of power consumption at night.
abnormal [19]. The level of indicators mainly depends on the level of abnormal electricity consumption in the region and the number of abnormal electricity meters. Different colors are used to represent different regional early warning levels.

4. Results and Discussion

The distribution of electricity consumption in each period and its fluctuation coefficient were analyzed. Through the distribution of each parameter, the general law of normal electricity consumption behavior can be obtained.

The proportion of users whose electricity consumption behavior is in line with peak period electricity consumption > normal period electricity consumption > valley period electricity consumption exceeds 90%. More than 85% of the users whose electricity consumption behavior is in line with the valley period have electricity consumption proportion <0.25.

The proportion of users whose fluctuation coefficient of electricity consumption in the valley period is less than 1 exceeds 85%, and the fluctuation coefficient of 99.9% of users’ electricity consumption in the valley period is less than 5.

Obviously, if the electricity consumption and its proportion in the valley period increase significantly and the electricity consumption and its proportion in each period fluctuate violently (the value of the fluctuation coefficient is significantly greater than the value in the normal range), it belongs to abnormal electricity consumption behavior [20].

5. Conclusion

In this paper, a method of hazard trend identification model based on a statistical analysis of abnormal power generation behavior data is proposed. This method also analyzes the power consumption in each period, the proportion of power consumption in each period, and the distribution of its fluctuation coefficient, calculates the proportion of power consumption in each period of the day, and reverses the early warning of abnormal power generation behavior data on the identification of dangerous trends to prove that the power consumption rate of the power plant is taken as the key index. It uses historical data to analyze the potential law of the power consumption rate of power plants and identify whether the reported value is abnormal. It is of great significance to identify dangerous trends in all aspects.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Liao, L. Wei, T. Kim, and W. Su, Modeling and Analysis of Residential Electricity Consumption Statistics: A Tracy-Widom Mixture Density Approximation, IEEE Access, Piscataway, NJ, USA, 2020.
[2] K. Giwa and O. D. Osahon, “Article average annual committed effective dose and threshold consumption rate of bovine samples collected from a government owned abattoir, ikpoba slope, Benin city, Nigeria,” Canadian Journal of Physics, vol. 98, no. 8, pp. 742–751, 2020.
[3] M. Ji and X. Bai, “Construction of the sponge city regulatory detailed planning index system based on the swmm model,” Environmental Technology and Innovation, vol. 23, no. 19, 2021.
[4] A. A. Kulikov, “Application of biometric systems in face identification technologies,” Russian Technical Journal, vol. 9, no. 3, pp. 7–14, 2021.
[5] X. Li, J. Liu, M. S. Obaidat, P. Vijayakumar, Q. Jiang, and R. Amin, “An unlinkable authenticated key agreement with collusion resistant for vanets,” IEEE Transactions on Vehicular Technology, vol. 70, pp. 7992–8006, 2021.
[6] C. Yang, W. Zhang, F. Zhang et al., “Research and dynamic analysis based on nonlinear identification of sports goods econometrics,” Journal of Physics: Conference Series, vol. 1802, no. 2, 2021.
[7] T. T. H. Phanab, É. P. Caillaultac, L. Alain, and A. Biganda, “Dynamic time warping-based imputation for univariate time series data—sciencedirect,” Pattern Recognition Letters, vol. 139, pp. 139–147, 2020.
[8] A. Hemeida, S. Alkalaf, A. Mady, E. Mahmoud, M. Hussein, and A. M. Baha Eldin, “Implementation of nature-inspired optimization algorithms in some data mining tasks, Aim Shams Engineering Journal, vol. 11, no. 2, pp. 309–318, 2020.
[9] Y. Zhou and P. Zhang, “Reachable eigenanalysis,” IEEE Transactions on Power Systems, vol. 35, pp. 4936–4939, 2020.
[10] S. Graham, D. Huffer, and J. Blackadar, “Towards a digital sensorial archaeology as an experiment in distant viewing of the trade in human remains on instagram,” Heritage, vol. 3, no. 2, pp. 208–227, 2020.
[11] K. Yuan, X. Guo, and J. Tian, “Research and implementation of clock synchronization technology based on ptp,” Journal of Physics: Conference Series, vol. 1757, no. 1, 2021.
[12] Y. Zhang, C. Zhang, Y. Wang, K. Yu, G. Xue, and J. Crowcroft, “Kylinx: simplified virtualization architecture for specialized virtual appliances with strong isolation,” ACM Transactions on Computer Systems, vol. 37, pp. 1–27, 2019.
[13] M. Sayad Haghighi, F. Farivar, and A. Jolfaei, "A machine learning-based approach to build zero false-positive.ipss for industrial iot and cps with a case study on power grids security," *IEEE Transactions on Industry Applications*, vol. 7, 2020.

[14] M. F. Ginting, K. Otsu, J. A. Edlund, J. Gao, and A. A. Agha-Mohammadi, "Chord: distributed data-sharing via hybrid ros 1 and 2 for multi-robot exploration of large-scale complex environments," *IEEE Robotics and Automation Letters*, vol. 6, pp. 5064–5071, 2021.

[15] S. Chattopadhyay, A. Dey, P. K. Singh, Z. W. Geem, and R. Sarkar, "Covid-19 detection by optimizing deep residual features with improved clustering-based golden ratio optimizer," *Diagnostics*, vol. 11, no. 2, p. 315, 2021.

[16] L. Yan, K. Cengiz, and A. Sharma, "An improved image processing algorithm for automatic defect inspection in TFT-LCD TCON," *Nonlinear Engineering*, vol. 10, no. 1, pp. 293–303, 2021.

[17] R. Huang and X. Yang, "The application of TiO\textsubscript{2} and noble metal nanomaterials in tele materials," *Journal of Ceramic Processing Research*, vol. 23, no. 2, pp. 213–220, 2022.

[18] R. Huang and X. Yang, "Analysis and research hotspots of ceramic materials in textile application," *Journal of Ceramic Processing Research*, vol. 23, no. 3, pp. 312–319, 2022.

[19] Q. Liu, X. Liu, T. Liu et al., "Seasonal variation in particle contribution and aerosol types in shanghai based on satellite data from modis and caliop," *Particuology*, vol. 51, pp. 18–25, 2020.

[20] P. Ajay and J. Jaya, "Bi-level energy optimization model in smart integrated engineering systems using WSN," *Energy Reports*, vol. 8, pp. 2490–2495, 2022.