WILEY | Hindawi

Retraction

Retracted: Electronic Communication Fault Signal Recognition Based on Data Mining Algorithm

Wireless Communications and Mobile Computing

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Luo, "Electronic Communication Fault Signal Recognition Based on Data Mining Algorithm," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 9641328, 8 pages, 2022.

WILEY | Hindawi

Research Article

Electronic Communication Fault Signal Recognition Based on Data Mining Algorithm

Jun Luo (ID)

Guangzhou College of Technology and Business, Guangzhou, Guangdong 510850, China

Correspondence should be addressed to Jun Luo; 31115326@njau.edu.cn

In order to solve the technical problem of fault signal recognition in the field of communication, this paper proposes an electronic communication fault signal recognition method based on data mining algorithm. Firstly, the K-means clustering algorithm is used to determine the cluster number k according to some attributes or class characteristics of the communication class samples, and the communication sample types are classified into a certain class so that the communication sample data in the cluster can be closely distributed and the data within a certain class range can be calculated by Euclidean distance formula. Then, this paper clusters the data. In the clustering data, BP neural network model is used to train and calculate the obtained clustering data again, which can map and deal with the complex nonlinear relationship between the fault information data of different clustering categories. The results show that the final error accuracy can be raised to about 20% by using the method in this paper. *Conclusion*. The algorithm designed in this paper can quickly predict the factors affecting the communication and find the communication fault information.

1. Introduction

Network communication plays a very important role in the rapid development of information technology. With the rapid development of information technology, the communication network is also developing rapidly with an unprecedented trend. The types of services provided by the communication network to people are constantly increasing and updating. It is playing a more and more important role in the whole human social economy, social activities, and daily life [1]. All these show that the communication network is no longer just a medium of mutual communication, but a distributed information processing platform that provides comprehensive services for human beings. Therefore, only scientific and effective management of communication network can ensure the accurate and uninterrupted operation of important communication services. Among them, the fault management of communication network is a very important link [2].

With the development of the Internet and high-tech industries, a large number of data are produced all the time in the manufacturing industry, service industry, and all

aspects of human life. How to extract the essence from the massive historical data and obtain valuable knowledge and laws for human society and scientific and technological progress is one of the issues favored by many scholars [3]. Datamining technology is a deep-seated data analysis method. Using data mining technology, we can obtain potentially valuable information from massive historical data and express the mined useful information in the form of rules or concepts. Applying these rules or concepts to fault diagnosis can provide support for decision-making.

Data mining is an interdisciplinary research field. The so-called data mining is the process of finding information of interest from a large number of irregular and potential data [4]. The typical application of data mining in the communication field is to analyze and discover the correlation rules of alarms based on historical alarm data and use the discovered rules to analyze and predict the possible faults of network components, thus reducing the work intensity of network management personnel and improving the work efficiency [5]. Data mining can analyze the existing alarm information (including historical alarm data and current alarm data) to obtain the association rules between alarms.

These valuable rules can be used for the location and detection of network faults and the prediction of serious faults. According to the analysis of the current alarm information, we can get the possibility of various subsequent situations, which plays a role in preventing dangerous events, so that the communication network can operate safely [6]. The advantage of data mining method is that it does not need to know the network topology relationship. When the network topology changes, it can automatically find new alarm correlation rules by analyzing the historical alarm records [7, 8]. Therefore, the alarm correlation analysis system based on data mining can quickly adjust to the dynamic communication network and solve the new problems in the communication network.

2. Literature Review

In the context of the increasing number of substations, in order to ensure the safe and stable operation of substations, higher requirements are put forward for substation duty personnel. As the most important helper of duty personnel, the importance of the monitoring background is becoming increasingly prominent. The substation monitoring background has four types of signals: telemetry, remote signaling, remote control, and remote regulation, which covers all the power equipment (primary, secondary, and automation equipment) in the substation. It can not only monitor the normal operation of the substation, but also, when the power equipment fails, the monitoring background can timely send an alarm signal to remind the personnel on duty. With the increase of the number of substations, the number of power equipment is increasing, which means that the types and number of substation monitoring background alarm signals are also increasing. If there are no new technologies and methods to improve the monitoring background, the substation duty personnel will be more and more overwhelmed [9].

Data mining technology refers to extracting interesting (nontrivial, implied, and potentially useful) information or patterns from large databases. Since the 1990s, it has been gradually studied and applied. Xie and others proposed the diagnosis function of fault information extracted from the system model by using data mining and algorithm [10]. Cui and others proposed several evaluation methods to determine the relative significance of input variables in the data mining model. Several methods are applicable to classification tasks, and the practicability and accuracy are evaluated according to the characteristics of fault diagnosis [11]. Bai and others studied the ability of using data mining to predict faults and find anomalies early [12]. Farhadi and others studied and defined the alarm information syntax model and theoretically studied the feasibility of mining correlation data from alarm data containing a lot of noise using the alarm correlation rule data mining algorithm [13]. Asami and others proposed to build a fault diagnosis platform, exchange and share diagnosis and maintenance information, and make maintenance plans according to their own experience and site conditions [14]. Zhou and others introduced the data mining model in the relay protection fault

information processing system and proposed a data mining technology based on rough set theory [15] in view of the possible distortion of information in the formation and transmission of real-time information based on diagnosis. Moayedi and others introduced data mining technology into the research of Power System SCADA alarm information noise data identification. Based on the decision tree algorithm commonly used in data mining technology classification and analysis methods, a data classifier is designed and implemented. They judge and analyze SCADA alarm information of power system, discover classification rules, and then eliminate SCADA remote signaling noise data, which is an effective attempt of data mining technology in the analysis and research of power system alarm information [16]. Li and others use association rule mining for attribute reduction, modify the threshold for interactive mining, directly extract the best combination of attribute reduction, and then use the reduction decision table formed by the best combination of attribute reduction and interactive mining of association rules to carry out diagnostic reasoning for fault information in various cases [17]. Liu and others proposed a new distribution network fault diagnosis technology based on hybrid data mining method in view of some defects of single data mining method in distribution network fault diagnosis [18].

This paper classifies the communication fault information through the clustering analysis algorithm and then quickly diagnoses the communication data fault type according to the BP neural network model, which provides a valuable technical reference for the healthy and green operation of the smart grid. At the same time, it also has good academic research significance and engineering application value.

3. Research Methods

3.1. Fault Information Analysis Framework System. In this paper, a communication fault information recognition method based on data mining algorithm is designed, which combines the clustering analysis algorithm with BP neural network to extract, process, analyze and calculate the interference information in the communication system, and obtain the fault signal. In the design of this paper, a fault information analysis architecture diagram is also proposed, as shown in Figure 1. In Figure 1, it mainly includes information collection part, information calculation part, and data information management part.

In the information acquisition part, a large number of information acquisition units are set up in the communication information system. The information acquisition unit includes a variety of sensors, such as vibration sensor, temperature sensor, humidity sensor, and magnetic field sensor. These sensors are used to collect various vibration, temperature, humidity, and other signals in the communication system. Then, users extract the features of these sensing information, input the extracted features to the computer processing system for storage, and perform data analysis, diagnosis, and display at the computer processing system. In this paper, the clustering algorithm is used to classify
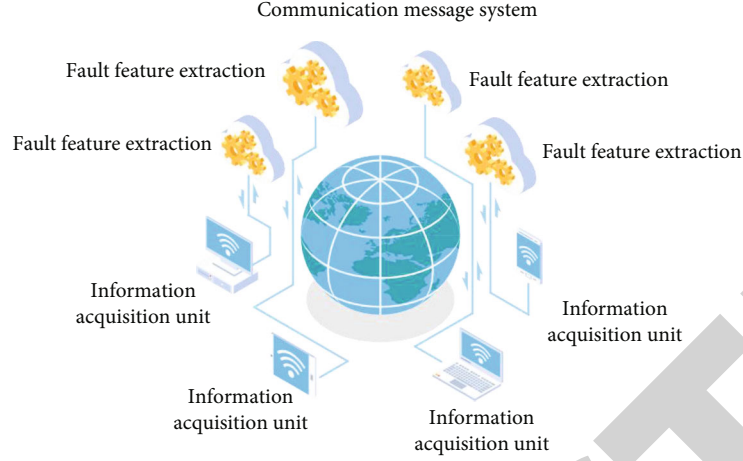
FIGURE 1: Schematic diagram of fault information analysis framework system.

the collected information data, classify and learn the interference factors that affect the communication, and obtain the range of fault information. Then, the BP neural network model is used to further analyze and calculate a small category of data. By mapping and processing the complex nonlinear relationship of a certain data type, the accurate estimation and detection of fault information in the communication system can be realized. The final data is monitored by the monitor, and the user is informed of the fault condition through the waveform display [19].

In the design of this system, the data processed by clustering algorithm and BP neural network model are available for user diagnosis and analysis in the general monitoring center. In the general monitoring center, the user can clearly see the processing data and clearly describe the fault signal according to the data processing. Fault data can also be transferred to the cloud through the Internet, and data storage can be realized in the cloud. The data in the general monitoring center is remotely transmitted to the remote upper management center through the industrial CAN bus or other communication protocols (TCP/IP) and the remote communication network for monitoring by the management personnel at a higher level. According to the monitoring data, the value of the lower management unit is used as the control guide.

3.2. Fault Information Analysis Algorithm. In this paper, the model of the K-means cluster algorithm and the BP neural network algorithm are combined to solve the communication failure signal as shown in Figure 2. A cluster is a set of statistical analysis methods that divides a research object into relative quality groups (clusters) that can group a collection of physical or abstract objects into many categories of similar objects. The BP network algorithm model is a multilayer transmission network prepared according to the error feedback algorithm and can learn and store a large number of input and output pattern mapping relationships. In this way, the accuracy of information fault signal evaluation is greatly improved. The following describes [20] in detail.

3.2.1. Clustering Analysis Algorithm. In the design of this paper, K-means algorithm of clustering analysis algorithm

is used to cluster the fault information types, and K-means algorithm can deal with the unlabeled data. As shown in Figure 3, the main steps are as follows:

(1) Extract the sample data and select the center point of the initial cluster. That is, K fault information sample data are randomly extracted from the sample data including temperature, vibration, power grid fault, load, humidity, harmonic, magnetic field, and power grid ripple, and the sample data is taken as the center of the initial value dataset (sample cluster). In this step, the steps of data preprocessing are included. In this step, a threshold for the number of iterations can be set

(2) Divide the points of the sample cluster and divide the points of each sample data cluster into sample clusters represented by centers that are relatively close to them, i.e., points that are relatively close to the center point of the original cluster divided into one class. In this step, we enter the distance formula to obtain the following equation:

$$d(x, y)^2 = \sum_{i=1}^{n} (x_i - y_i)^2 = x^- y_2^2, \tag{1}$$

where x and y, respectively, represent different samples; n represents the dimension of the fault data sample; and $d(x, y)$ is the Euclidean distance. Using the above formula, the distance between each fault signal sample data and these central sample parameters is calculated according to the center point of the cluster sample of each fault data, and the corresponding fault information data is redivided according to the minimum distance.

(3) The center point of each sample data point in a different sample cluster is used to represent the center point of the sample cluster. According to the difference of parameter data, the distance between the center point of each sample data and these cluster
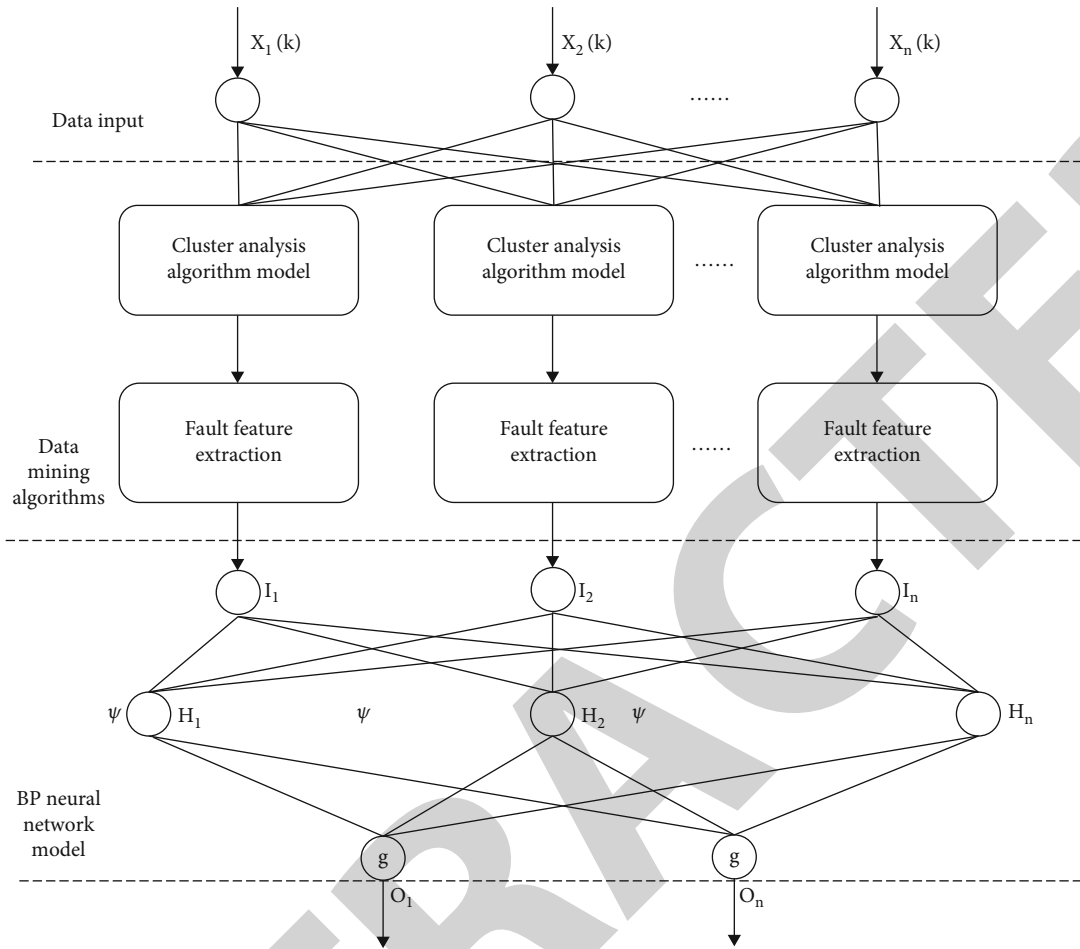
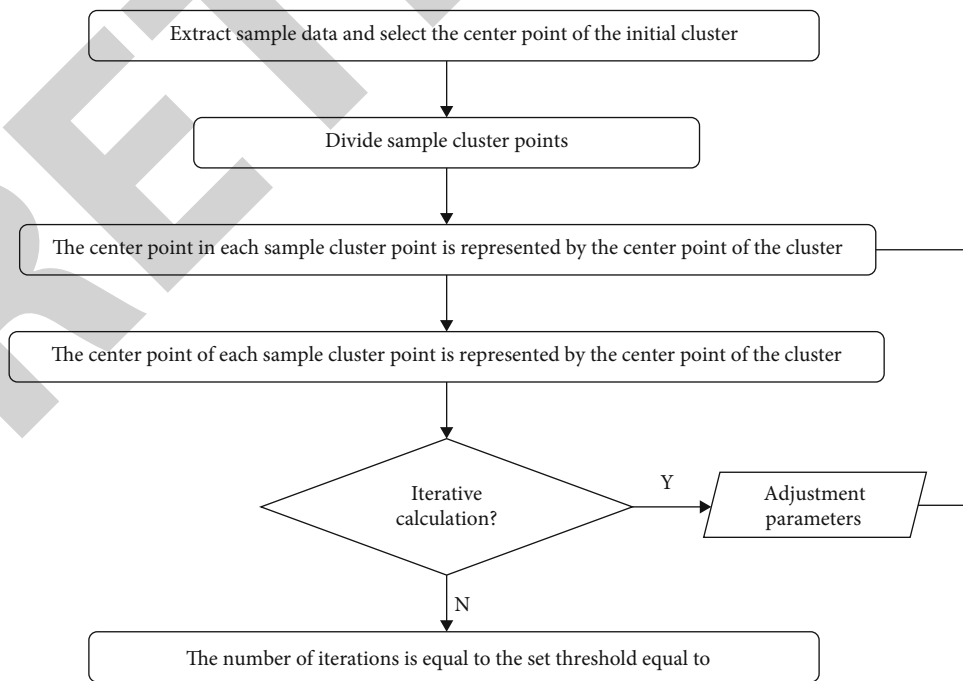Figure 2: Schematic diagram of fault information analysis algorithm.

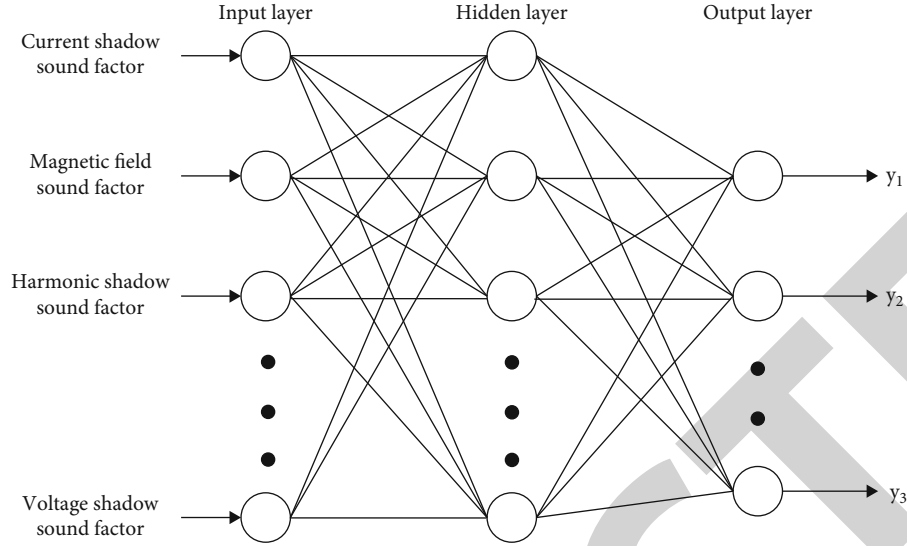Figure 3: Clustering analysis algorithm flow.

FIGURE 4: Algorithm model of BP neural network model.

TABLE 1: Calculation of test sample data.

Data type	Sample data	Test time/S	Recall rate/%	Accuracy rate/%	Model accuracy/%
Grid ripple	200	10	90.6%	92.1%	90.8%
Current	200	10	91.2%	92.5%	90.5%
Load	200	10	90.8%	92.3%	90.9%
Harmonic	200	10	91.1%	92.4%	90.6%

information data centers can be calculated again according to the center points of different cluster information sample data, and the corresponding fault information sample data can be recalculated according to the above minimum distance. Divide again. The minimum data calculated for each hour is the D matrix, followed by

$$
\boldsymbol{D} = \begin{bmatrix} x_{11}, x_{12} \cdots\cdots x_{1n} \\ x_{21}, x_{22} \cdots\cdots x_{2n} \\ \cdots\cdots \\ x_{k1}, x_{k1}, x_{k2} \cdots\cdots x_{kn} \end{bmatrix}, \tag{2}
$$

where X is the set of minimum values.

3.2.2. BP Neural Network Algorithm Model. Following the above processing method, the BP network algorithm model is used to map more complex nonlinear relationships in the trauma information sample in a timely manner. Because the BP network algorithm model has high learning efficiency, fast diagnosis speed, and high accuracy, it can quickly diagnose the types of communication data failures in the clustered data, making the information for dealing with communication failures more accurate.

When using the above method, first select the results obtained by the cluster analysis algorithm, and then conduct training and education. The BP neural network model (as shown in Figure 4) consists of three layers: the input layer, the latent layer, and the output layer. The input layer usually contains various types of data such as temperature, vibration, mains damage, load, humidity, harmonics, magnetic fields, and mains waves. By constantly adjusting the weight and threshold in the neural network, it gradually approaches the required level data. As a result, the output error is ultimately reduced. Follow these formulas to adjust your BP neural network model.

The formula for adjusting the output layer weight system is

$$
\Delta\omega_{ki} = \eta o_k^p \left(1 - o_k^p\right) \left(t_k^p - o_k^p\right) o_i^p. \tag{3}
$$

The formula for adjusting the weight coefficient of the hidden layer is

$$
\Delta\omega_{ij} = \eta O_i^p \left(1 - O_i^p\right) \sum_{k=1}^{L} O_k^p \Delta O_k^p \Delta\omega_{ki} O_j^p. \tag{4}
$$

The quadratic accurate function model of the input mode pair in each fault information sample is

$$
J_p = \frac{1}{2} \sum_{p=1}^{N} J_P = \frac{1}{2} \sum_{K=1}^{L} \left(t_k^p - o_k^p\right). \tag{5}
$$

The total accurate function expression for N fault information samples is shown in

$$
J = \sum_{p=1}^{N} J_p = \frac{1}{2} \sum_{p=1}^{N} \sum_{k=1}^{L} \left(t_k^p - o_k^p\right)^2. \tag{6}
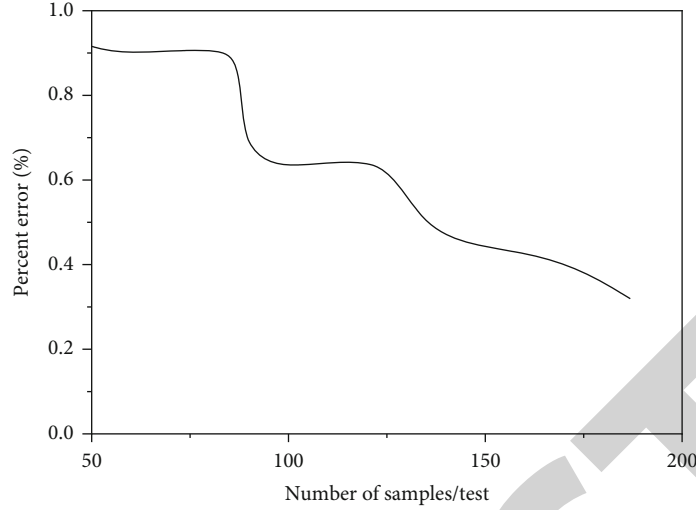$$

Figure 5: Calculation results of BP neural network model.

At the beginning of calculation, in order to improve the learning accuracy when extracting complex communication information, the sample data is first standardized. Assuming that the type of input communication fault information is m and the sample is x_{ij}, the standardization of input data x_{ij} is carried out according to the steps of the following formulas:

$$Z_{ij} = \frac{(x_{ij} - \bar{x}_j)}{\delta_j}, \tag{7}$$

$$x_j = \frac{1}{N}\sum_{i=1}^{N} x_{ij}, \tag{8}$$

$$\delta_j^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_{ij} - \bar{x}_j)^2, \tag{9}$$

where $i = 1, 2, \cdots, N$; $j = 1, 2 \cdots, m$, in the above formula and Z_{ij} is the data after standardization. The standardized formula can be

$$y_i' = \frac{q(y_i - y_{\min} + b)}{(y_{\min} - y_{\min} + b)}, \tag{10}$$

where y_i is the output fault information sample; y_i' is the standardized fault information sample data, and $y_{\max}, y_{\min}$ are the maximum and minimum values in the output fault data samples. Of which $0 < q < 2$; $0 < b < 2$, then the number of hidden layer nodes is determined to be between 7 and 9, the value from input layer to hidden layer is between 0.2 and 0.6, and the value from hidden layer to output layer can be between 0.2 and 0.3. According to the above formula, a neural network model can be established.

4. Result Analysis

When simulating the clustering analysis algorithm, the hardware environment is selected: the Pentium (R) CPU is 256

memory, the hard disk capacity is 80 g, the software environment is Windows XP operating system, JDK1.5, and the BP neural network model is simulated based on MATLAB. Four kinds of grid fault information (grid ripple, current, load, and harmonic) are selected as the test. When verifying the correctness of clustering, F-measure is selected as the evaluation standard to evaluate the accuracy of clustering classification algorithm by using the accuracy and recall in information retrieval. The calculation formula is: as follows

Accuracy calculation formula:

$$\mathrm{recall}(i, j) = \frac{n_{ij}}{n_j}, \tag{11}$$

Recall rate calculation formula:

$$\mathrm{recall}(i, j) = \frac{n_{ij}}{n_i}, \tag{12}$$

$$F(i, j) = \frac{2 * \mathrm{recall}(i, j) \times \mathrm{presision}(i, j)}{\mathrm{recall}(i, j) + \mathrm{presision}(i, j)}, \tag{13}$$

The final value of F:

$$F1\sum_{n}^{n_i} \max\{F(i, j)\}, \tag{14}$$

The calculation table of test sample data is shown in Table 1.

Then, start the BP neural network model in the MATLAB interface, and select according to the sample type, and the error data curve calculated by the BP neural network model is shown in Figure 5.

Through the above experiments, this paper uses the clustering classification algorithm to classify various data that affect the communication information types, such as temperature, vibration, power grid fault, load, humidity, harmonic, magnetic field, and power grid ripple, according to

different categories, and then adjusts the weights and thresholds in the neural network to gradually approach the required results, so as to minimize the output error. Using the method in this paper, the final error accuracy is raised to about 20%. Therefore, the algorithm designed in this paper can quickly predict the factors affecting the communication and find the communication fault information.

5. Conclusion

This paper classifies the communication fault information (temperature, vibration, power grid fault, load, humidity, harmonic, magnetic field, power grid ripple, etc.) through the clustering analysis algorithm, which is conducive to increasing the recognition degree of different communication numbers. Then, according to the BP neural network model, it further quickly diagnoses the fault type of communication data, making it more accurate to process the communication fault information. In the whole calculation process, the calculation speed is increased. The accuracy of the user training data from the clustered communication fault data is helpful for the user to quickly diagnose the fault data types in the communication system, enable the user to quickly find and solve problems, and provide valuable technical reference for the healthy and green operation of the smart grid. At the same time, it also has good academic research significance and engineering application value.

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.

Acknowledgments

References

[1] J. Rak, P. E. Heegaard, and B. E. Helvik, "Resilience of communication networks to random failures and disasters: an optimization perspective," *Networks*, vol. 75, no. 4, pp. 337–339, 2020.

[2] N. Yuvaraj, K. Srihari, G. Dhiman, K. Somasundaram, and M. Masud, "Nature-inspired-based approach for automated cyberbullying classification on multimedia social networking," *Mathematical Problems in Engineering*, vol. 2021, 12 pages, 2021.

[3] J. Jayakumar, S. Chacko, and P. Ajay, "Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment," *Wireless Communications and Mobile Computing*, vol. 2021, 8 pages, 2021.

[4] X. Liu, J. Liu, J. Chen, F. Zhong, and C. Ma, "Study on treatment of printing and dyeing waste gas in the atmosphere with Ce-Mn/GF catalyst," *Arabian Journal of Geosciences*, vol. 14, no. 8, article 737, 2021.

[5] R. Huang, P. Yan, and X. Yang, "Knowledge map visualization of technology hotspots and development trends in China's textile manufacturing industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 243–251, 2021.

[6] J. Gu, W. Wang, R. Yin, C. V. Truong, and B. P. Ganthia, "Complex circuit simulation and nonlinear characteristics analysis of GaN power switching device," *Nonlinear Engineering*, vol. 10, no. 1, pp. 555–562, 2021.

[7] P. Li, L. Yin, B. Zhao, and Y. Sun, "Virtual screening of drug proteins based on imbalance data mining," *Mathematical Problems in Engineering*, vol. 2021, no. 16, pp. 1–10, 2021.

[8] L. Huang, "Design of an IoT DDoS attack prediction system based on data mining technology," *The Journal of Supercomputing*, vol. 8, pp. 1–23, 2021.

[9] W. Tang and H. Chen, "Research on intelligent substation monitoring by image recognition method," *International Journal of Emerging Electric Power Systems*, vol. 22, no. 1, pp. 1–7, 2021.

[10] X. Xie, W. Wu, J. Xu, X. Jiang, and H. Zhang, "Abnormal network data mining model based on deep training learning," *International Journal of Internet Protocol Technology*, vol. 13, no. 4, p. 228, 2020.

[11] Y. Wang, F. Tian, and Y. Bai, "Characteristics analysis of applied mathematics in colleges and universities based on big data mining algorithm model," *Security and communication networks*, vol. 2022, Article ID 7978031, 13 pages, 2022.

[12] X. Bai, C. Liao, M. Xu, and Y. Zheng, "Correction of atmospheric model through data mining with historical data of two-line element," *IEEE Access*, vol. 8, pp. 123272–123286, 2020.

[13] H. Farhadi, A. Esmaeily, and M. Najafzadeh, "Developing a decision tree based on data mining method for detecting the influential parameters on the power of flood destruction," *Amirkabir (Journal of Science and Technology)*, vol. 53, no. 5, 2021.

[14] H. Asami, M. Golabi, and M. Albaji, "Simulation of the biochemical and chemical oxygen demand and total suspended solids in wastewater treatment plants: data-mining approach," *Journal of Cleaner Production*, vol. 296, no. 2-4, article 126533, 2021.

[15] R. Zhou, H. Chen, H. Chen, E. Liu, and S. Jiang, "Research on traffic situation analysis for urban road network through spatiotemporal data mining: a case study of Xi'an, China," *IEEE Access*, vol. 9, pp. 75553–75567, 2021.

[16] X. Xu, L. Li, and A. Sharma, "Controlling messy errors in virtual reconstruction of random sports image capture points for complex systems," *International Journal of Systems Assurance Engineering and Management*, vol. 5, 2021.

[17] M. Raj, P. Manimegalai, P. Ajay, and J. Amose, "Lipid data acquisition for devices treatment of coronary diseases health stuff on the internet of medical things," *Journal of Physics: Conference Series*, vol. 1937, article 012038, 2021.

[18] L. Xin, L. Jianqi, C. Jiayao, and Z. Fangchuan, "Catalytic conversion and DFT analysis of post DDBD-catalysis system for degradation of toluene, ethyl acetate and acetone with different metal-oxides catalysts," *Journal of Rare Earth*, vol. 3, 2022.

[19] R. Huang, S. Zhang, W. Zhang, and X. Yang, "Progress of zinc oxide-based nanocomposites in the textile industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 281–289, 2021.

[20] E. Guo, V. Jagota, M. Makhatha, and P. Kumar, "Study on fault identification of mechanical dynamic nonlinear transmission system," *Nonlinear Engineering*, vol. 10, no. 1, pp. 518–525, 2021.