Image Processing, Electronics and Computers
L. Trajkovic et al. (Eds.)

doi:10.3233/ATDE240489

Quantitative Analysis and Simulation of Electricity Bill Recovery Risk Project Performance Based on Machine Learning Algorithm

Weiting LIAO[1], Xiaoyan YANG, Hongyu SU, Qiongrong YAO and Xiaohong XIE
Nanning Power Supply Bureau Guangxi Power Grid Co., LTD, 530000, Nanning, China

Abstract. Electricity bill recovery project involves a lot of data, including user electricity data, arrearage situation, recovery situation, etc., the data is easy to exist incomplete and inaccurate, resulting in the analysis of electricity bill recovery risk project is more difficult. Therefore, a simulation method based on machine learning algorithm for quantitative analysis of project performance of electricity bill recovery risk is proposed. Subspace identification method is used to create the performance distribution model of electricity bill recovery project and collect electricity bill recovery data. The classification of electricity rate recovery data is realized by using machine learning algorithm. The collected performance data of electric charge recovery risk projects are normalized and the local density of electric charge recovery data is calculated. The density peak clustering algorithm is used to determine the clustering center of electricity and electricity bill recovery data, construct the characteristic curve of electricity and electricity bill recovery data, and complete the quantitative processing of electricity and electricity bill recovery risk project performance. The simulation results show that: The F1 score of the research method is higher, and the false detection rate is always lower than 0.1, which indicates that the research method has reliable application performance.

Keywords. Machine learning algorithm, electricity bill recovery, risky projects, performance quantification, simulation experiment

1. Introduction

In recent years, with the rapid development of social economy and the acceleration of urbanization, the recovery of electricity charges has become an important part of energy management. However, there is still a risk of electricity bill recovery in many areas, that is, some residents or enterprises often default on electricity bills. This situation not only affects the economic benefits of power companies, but also has a negative impact on the entire energy supply system.

Therefore, it is of great research significance to conduct quantitative analysis on the performance of electricity bill recovery risk projects [1]. Through in-depth analysis of the risk factors of electricity bill recovery, including the ability of users to pay, electricity consumption behavior and policies and regulations, it can help power

[1] Corresponding Author: Weiting LIAO, 215531617@qq.com

companies better formulate recycling strategies and measures to improve the efficiency and success rate of recycling. At the same time, quantitative analysis of performance can also promote power companies to optimize management processes, reduce operating costs, improve service quality and enhance market competitiveness [2]. In the current intelligent and data-driven era, the use of big data analysis, artificial intelligence and other technical means to conduct quantitative analysis of the performance of electricity bill recovery risk projects can not only improve the management level of power companies, but also help establish a more accurate and robust electricity bill recovery model. The continuous deepening of this research field will help promote the modernization of the field of energy management, and realize the sustainable development of the power industry and the maximization of social benefits.

Aiming at the quantitative analysis of power risk, Huang J, et al. proposes a risk assessment method for power grid dispatching operation based on Markov chain and two-point estimation [3]. Risk theory is introduced into scheduling operation evaluation. Based on the characteristics of decomposition and memory-free scheduling operations, Markov chain is used to simulate the step-by-step execution of scheduling operations. At the same time, four indexes of voltage overlimit, power flow overload, static voltage stability and chain fault are selected to quantify the consequences of the power grid, and the influence of load fluctuation is considered in the indexes. Zhang A A, et al. takes transformer as the research object and designs a quantitative risk analysis method [4]. Firstly, the index system of transformer health evaluation is established, based on which the reliability model of transformer real-time health status is constructed. Secondly, the restoration function is used to quantify the influence of different maintenance methods on transformer reliability, and the transformer reliability before and after maintenance is determined by equivalent draft age method. Then, the operation risk economy of transformer is analyzed, and a reasonable maintenance strategy optimization model is constructed. Finally, the improved pigeon swarm algorithm is used for iterative optimization, so as to determine the optimal maintenance scheme and maintenance time.

However, the method proposed in the above literature has some disadvantages. (1) These methods have limited treatment of system complexity and uncertainty, and it is difficult to fully simulate the operation of complex power systems or the actual working environment of transformers. (2) These methods may be subjective and limited in the establishment of models and the selection of indicators, and cannot cover all potential risk factors, resulting in the accuracy and comprehensiveness of evaluation results. (3) These methods may have limitations in algorithm design and optimization strategies, and it is difficult to fully consider the optimal solutions under various conditions, so there may be local optimal solutions in practical applications, affecting the accuracy and reliability of decision-making. Taken together, these methods need to be further improved and optimized to improve their applicability and effectiveness in the field of power system risk assessment and optimization.

Therefore, this paper proposes a quantitative analysis simulation method based on machine learning algorithm for project performance of electricity bill recovery risk.

2. Research Fundamentals

2.1 Collection of Electricity Charge Recovery Data

In order to complete the reliable quantitative evaluation of the performance of electricity bill recovery projects under abnormal behaviors, subspace identification method is adopted to create the performance distribution model of electricity bill recovery projects [5]. Subspace identification is an effective data analysis technique, which can help to identify the hidden patterns and laws in the collected data, and then build the corresponding performance distribution model. Through the subspace identification method, we can understand the characteristics and distribution of electricity bill recovery project performance more comprehensively and accurately, and provide a solid foundation for the subsequent data processing and evaluation [6]. A single distributed power nodes are used to monitor the network status. By fusing fuzzy scheduling strategy, a set of performance data collection model of electricity bill recovery project is established. The model combines the characteristics of fuzzy logic, and can capture the complex relationships and patterns in the performance data of electricity bill recovery projects more effectively. Through this model, the data related to the evaluation of the performance of the electricity bill recovery project can be comprehensively collected and converted into a series of performance characteristics. This method of integrated fuzzy scheduling strategy can not only improve the efficiency and accuracy of data collection, but also help managers better understand the changing trend and characteristic law of the performance of the electricity bill recovery project, so as to realize the in-depth analysis of the performance data and optimize the decision. This sequence of performance characteristics provides an important reference for the performance evaluation of electricity bill recovery projects, and helps to promote the development of electricity bill recovery projects in a more efficient and accurate direction. Written as:

$$B=\left\{B_1,B_2,\cdots,B_N\right\} \tag{1}$$

Where, B_N is the N feature vector in the spatial distribution of performance of electricity bill recovery project.

The spatial distribution feature vectors are normally correlated. If B meets the K distribution function, then the channel model for performance transmission control of the electricity bill recovery project is:

$$x(n)=c(n)+d(n) \tag{2}$$

Where, $c(n)$ represents the vector data set of performance data of electricity bill recovery project under abnormal behavior, and $d(n)$ is the characteristic data set of electricity bill recovery situation distribution.

By adopting the channel balance adjustment strategy, the performance feature space of the electricity bill recovery project is reconstructed successfully, and the original performance feature data is adjusted and optimized according to certain rules and standards. In the adjusted feature space, a vector data set containing various performance indicators is established, which can describe the performance of the electricity bill recovery project more comprehensively and accurately. In this reconstructed vector data set, the statistical analysis of project performance data series is realized. Through the data statistics of various indicators, we can deeply understand the changing trend, correlation and abnormal situation of project performance, which provides an important reference for further performance evaluation and decision-making. This process not only improves the overall ability to control the performance data of the electricity bill recovery project, but also provides a more detailed and comprehensive perspective for the project manager to better formulate effective management strategies and improvement plans, and promote the steady development and optimization of the operation of the electricity bill recovery project. The characteristic quantity of statistical output is:

$$H = \chi(t) l^{-j2\pi} \delta(t) \tag{3}$$

Where, $\chi(t)$ is the unbalanced feature of network security debugging, t is the output delay of network frequency band, δ is the modulation frequency of network channel, and l is the characteristic vector of network fluctuation.

2.2 Classification of Electricity Bill Recovery Data Based on Machine Learning Algorithm

Data classification has become a very popular research content in the field of machine learning. Accurate classification of electricity charge recovery data is conducive to subsequent retrieval and processing [7]. The machine learning algorithm is introduced to carry out the classification and processing of electricity bill recovery data. The principle of machine learning algorithm in the classification and processing of electricity recovery data is to learn and train the existing data samples, learn the patterns and features of the data, and then classify the unknown data according to these learned rules. Typically, machine learning algorithms use labeled data samples to train models to recognize and classify different categories of data. In the classification and processing of electricity bill recovery data, the machine learning algorithm can automatically discover the potential relationships and rules in the data, realize the effective classification of the data, and can quickly and accurately classify the data according to the characteristics of the new data, providing a scientific basis for the management and decision-making of electricity bill recovery project. The general learning process of the machine learning algorithm is shown in Figure 1:

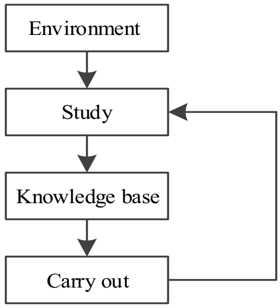

Figure 1. General process of machine learning algorithm

As the application field of machine learning is very wide, a variety of new machine learning algorithms continue to emerge, all algorithms can be applied to the classification of electricity recovery data. Therefore, the following application of machine learning to carry out the classification and processing of electricity bill recovery data is mainly divided into the following two stages:

The main purpose of a classification model in a machine learning algorithm is to select the appropriate pre-defined document for the input document. A predefined set of documents is applied to training, and it is trained by a document algorithm. The following uses the classification model in the machine learning algorithm to carry out scoring processing on the feature words of the electricity bill recovery data [8]:

(1) The classification model in the machine learning algorithm is used to classify all the electricity bill recovery data, and the electricity bill recovery with similar content is divided into a class;

(2) The association rule mining algorithm in the machine learning algorithm is used to extract the correlation of electricity bill recovery data and the final keywords corresponding to electricity bill recovery;

(3) For each type of electricity rate recovery, calculate the word frequency of the specific keywords extracted from each type of electricity rate recovery;

(4) In each class, it is necessary to carry out clustering processing on the key words of electricity bill recovery data with the same word frequency;

(5) Sorting and processing all kinds of electricity charge recovery data according to keyword word frequency;

(6) Assign an index to each class. When the index value is 0, it indicates the class with the highest word frequency, and so on, the index is assigned to other classes to complete the classification.

3. Method

In order to realize the quantitative analysis of the risk items of electricity and tariff recovery data, the relevant electricity and tariff recovery data has been collected from the power system in the above article. After data collection, to improve the accuracy and quality of the collected data, it is necessary to conduct standardized processing of the collected data [9]. The specific process of standardization is as follows:

$$\begin{cases} W_i = \dfrac{w_{i,j} - \varepsilon_i}{\sigma_i} \\ \varepsilon_i = \dfrac{\sum_{j=1}^{J} w_{i,j}}{m} \end{cases} \tag{4}$$

In the above formula, W_i represents the normalized processing of electricity and electricity bill recovery data, $w_{i,j}$ represents the original electricity and electricity bill recovery data, ε_i represents the average value of collected electricity and electricity bill recovery data, σ_i represents the standard deviation of collected electricity and electricity bill recovery data, and m represents the quantity of electricity and electricity bill recovery data.

Through the above formula, the standardized processing of electricity and electricity fee recovery data is carried out to improve the quality and accuracy of electricity and electricity fee recovery data, and to lay a foundation for the subsequent determination of the clustering center of electricity and electricity fee recovery data. In determining the clustering center of the electricity bill recovery data. In this paper, the density peak clustering algorithm is used for calculation. Density peak clustering algorithm, as a clustering algorithm based on data density distribution, determines the cluster center by looking for density peak points in the data, and assigns the data points to different clusters. The algorithm can automatically discover cluster centers and process data sets with arbitrary shapes. In the course of application, the clustering algorithm determines the local density of each data point by calculating the distance of each data point to other data points in its neighborhood. When the point with the highest local density and closest to the centroid or central point of the cluster is found, it is the clustering center of the cluster [10]. Density clustering algorithm has strong data clustering ability, and the clustering process is more efficient. In the process of clustering, it is necessary to calculate the local density of electricity charge recovery data. The specific calculation process is as follows:

$$\rho = B\left(R_{ij} - R_s\right) \tag{5}$$

In the above formula, ρ represents the local density of electricity and electricity bill recovery data, $B(\cdot)$ represents the discrete function of electricity and electricity bill recovery data, R_{ij} represents the interval of different electricity and electricity bill recovery data in the data set, and R_s represents the standard value of the interval of electricity and electricity bill recovery data [11]. Based on the above formula, the local density of electricity and electricity bill recovery data is calculated, and the weight centers of different clusters of electricity and electricity bill recovery data are determined, and the clustering centers are determined according to the calculation results to realize the clustering processing of electricity and electricity bill recovery data. In the above process, the specific process of using the density peak clustering algorithm to determine the clustering center of electricity and electricity bill recovery data is shown as follows:

$$\begin{cases} p_i = \dfrac{r_{i+N} - r_i}{N} \\ o = \arg\left[\max\left(\dfrac{U_i^1}{U_1^{i-1}}\right)\right] \end{cases} \tag{6}$$

In the above formula, p_i represents the central weight of electricity and electricity bill recovery data in different clusters, U_i represents the change trend of the central weight, r_{i+N} and r_i represent different clustering ranges and distances between different electricity and electricity bill recovery data, N represents the clustering range of electricity and electricity bill recovery data, and o represents the clustering center of electricity and electricity bill recovery data.

Through the above formula, the clustering center of the electricity and charge recovery data is determined [12]. According to the distance between adjacent data points, the collection data of electricity and electricity charges can be divided into different clusters to realize the clustering processing of electricity and electricity charges collection data [13]. At the same time, considering that under the action of density peak clustering algorithm, the clustering center of power rate recovery data is dynamic. Therefore, in the calculation, it is also necessary to sort the calculated central weight, and further determine the cluster center point according to the size of the weight. At this point, the clustering center of power rate recovery data based on density peak clustering algorithm has been determined.

Based on this, the characteristic curves of different data are drawn, and their changing trends and shapes are analyzed, so as to find the risk items in them, connect all the risk items together, and construct the characteristic curves of the risk items in the data of electric power bill recovery. The characteristic curve of the risk project thus constructed is shown in Figure 2.

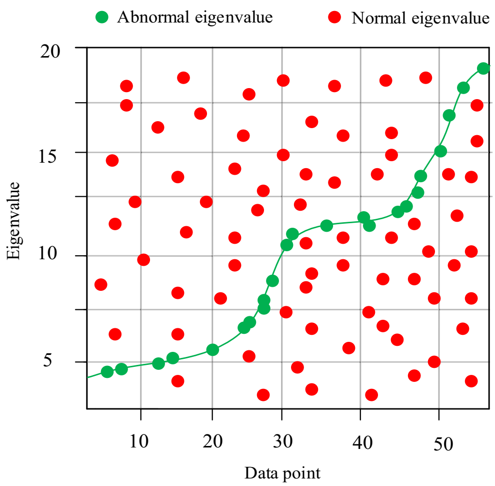

Figure 2. Characteristic value curve of risk items in electricity bill recovery data

As shown in Figure 2, in the above figure, the risk items of electricity and electricity bill recovery data are basically distributed on the same curve, and their values are all within a certain range. Normal values are more widely distributed than risk items. Taking the above risk project characteristic curve of electricity and electricity bill recovery data as the basis, it lays a foundation for the subsequent quantitative analysis of risk project of electricity and electricity bill recovery data. At this point, the construction of the risk project characteristic curve of electric power bill recovery data is completed.

On the basis of the above design, the paper makes use of the generated risk item characteristic curve to realize the quantitative analysis of the risk item of the electricity bill recovery data [14]. The specific quantitative analysis process is shown in Figure 3.

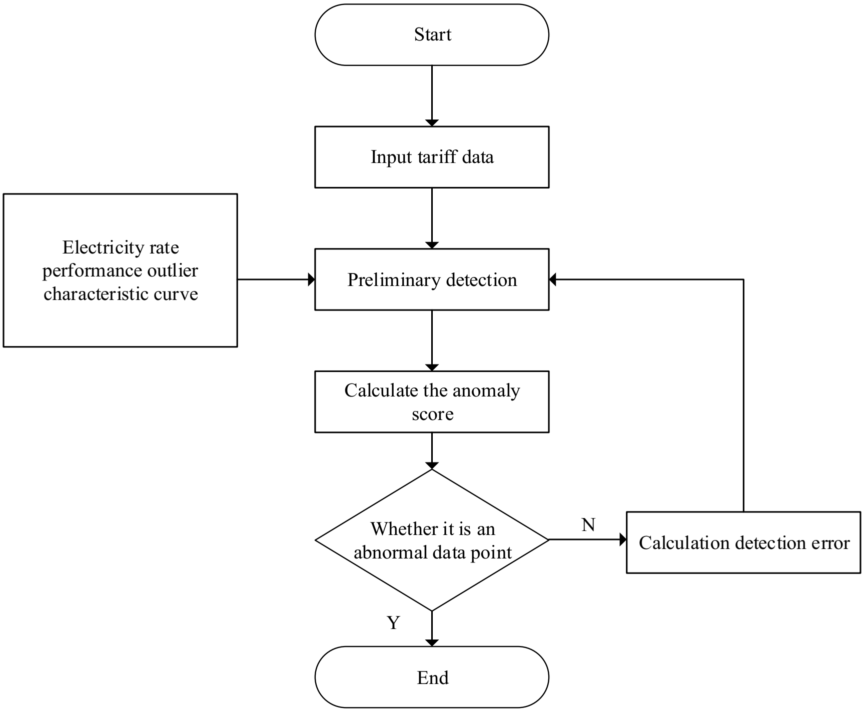

Figure 3. Process of quantitative analysis of risk items of electricity bill recovery data

As shown in Figure 3, in the process of quantitative analysis of the above risk items, the collected power and electricity bill recovery data set is first input, and the generated power and electricity bill recovery data risk project characteristic curve is reused. Conduct preliminary quantitative analysis of the risk items in the input data set, and according to the quantitative analysis results. Calculate the abnormal scores of data points on the characteristic curve of risk items, and then judge them [15]. Determine whether it is a data anomaly. If the data point is an anomaly, the quantization analysis result is directly output, if not, the quantization analysis error needs to be calculated. The data point is judged again, and the results of direct judgment meet the requirements. In the above process, the abnormal score is calculated as follows:

$$Y = \ln x + e \tag{7}$$

In the above formula, Y represents the calculated anomaly score, and e represents the set quantization analysis error value. Through the above formula, the abnormal scores of the electricity and electricity bill recovery data are calculated, and the risk items of the electricity and electricity bill recovery data are quantitatively analyzed according to the calculation results, and the corresponding quantitative analysis results are output. The above quantitative analysis results are analyzed, and the risk items are dealt with, the causes of the risk items are judged, and the corresponding treatment measures are taken to ensure the stable operation of the power system. At the same time, the data of electricity charge recovery is monitored in real time, and its changes and characteristics are updated in real time to ensure the timeliness and efficiency of the quantitative analysis method. At this point, the design of the quantitative analysis method for the risk items of the electricity rate recovery data based on the density peak clustering algorithm has been completed.

4. Validate

In order to improve the reliability of the experimental results, corresponding control experiments were set up. Among them, the design method in this paper is method 1, the method in literature is method 2, and the method in literature is method 3 [3-4]. In order to compare the performance of the above three methods in practical application, the experiments designed are as follows.

4.1 Preparation for Experiment

In order to verify the effect of the design method in practical application, experimental tests are carried out. In the experiment, a power company is taken as an example, and the power system of the power company is taken as the experimental object. The data acquisition instrument is used to collect the operation data of the power system of a power company from multiple sensors, including the electricity consumption and electricity charges of power users for subsequent experimental testing. The energy meter contains a large number of power system operation data, which can provide solid data support for this experiment. The computer can analyze the collected data and record the corresponding experimental results.

4.2 Discussion of Experimental Results

In order to verify the effect of the above three methods in practical application, the F1 score of the three methods is taken as the evaluation index, and the quantitative analysis effect of the three methods is compared. The higher the F1 score, the better the quantitative analysis. In the experiment, three methods were used to conduct multiple quantitative analysis of the experimental data set, and F1 scores of the three methods were recorded in the process of quantitative analysis. The specific statistical results are shown in Table 1.

Table 1. Quantitative analysis results of the three methods

Number of experiments	Three ways to score F1		
	Method 1	Method 2	Method 3
100	0.98	0.85	0.74
200	0.99	0.84	0.72
300	0.97	0.86	0.76
400	0.96	0.82	0.78
500	0.98	0.86	0.74
600	0.97	0.87	0.79
700	0.96	0.88	0.75
800	0.95	0.86	0.77
900	0.96	0.85	0.76
1000	0.95	0.84	0.74

As shown in Table 1, in the above experimental results, F1 scores of method 1 are higher than those of method 2 and method 3 in a single experiment. Overall, the average F1 score of Method 1 is 0.97, the average F1 score of Method 2 is 0.85, and the average F1 score of method 3 is 0.74. Thus, method 1 has the highest F1 score. Therefore, the method designed in this paper has a good quantitative analysis effect. This is because the performance quantitative analysis method combines the advantages of a variety of technologies, such as the performance distribution model created by subspace identification method, data classification realized by machine learning algorithm, and density peak clustering algorithm. The subspace identification method helps accurately capture the patterns and rules in the performance data of the electricity bill recovery project, the machine learning algorithm effectively classiifies the data, and the density peak clustering algorithm can accurately determine the clustering center of the data and build a representative characteristic curve of the risk project. These comprehensive technical means make the research method more comprehensively describe the performance of electricity bill recovery risk projects, improve the comprehensiveness and accuracy of the assessment, and thus make F1 score higher.

In order to further verify the effect of the above three methods in practical application, the false detection rate of the three methods is taken as the evaluation index, and the false detection rate of the three methods is compared under different data volumes. The specific statistical results are shown in Figure 4.

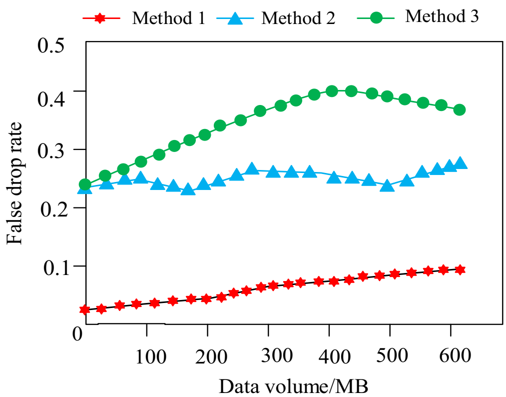

Figure 4. Quantitative analysis results of the three methods

As shown in Figure 4, in the above quantitative analysis results, it can be seen that with the increase of data volume, the false detection rate of method 1 and method 2 is gradually increasing, and the false detection rate of method 1 is always much lower than that of the other two methods. It can be seen that method 1 has the lowest false detection rate. Therefore, the method designed in this paper has the best quantitative analysis effect. This is because this research method adopts relatively strict and accurate standardized processing when dealing with electricity bill recovery data, and combines local density calculation and density peak clustering algorithm. Normalized processing helps to eliminate outliers and noise in the data, reduce the volatility and uncertainty of the data, and thus improve the quality and stability of the data. At the same time, by calculating the local density and determining the clustering center, the valuable data patterns and characteristics are successfully identified, and the false detection rate is effectively reduced to ensure the accuracy and reliability of the performance of the project. By combining these methods, the false detection rate can be maintained at a low level, which enhances the stability and reliability of the overall evaluation.

5. Conclusion

In this paper, a quantitative analysis and simulation method of electricity bill recovery risk project performance based on machine learning algorithm is designed, which combines subspace identification method and density peak clustering algorithm. The performance distribution model of electricity bill recovery data is successfully constructed, and the classification and cluster analysis of electricity bill recovery risk project data are realized. This method provides a new way of thinking and technical means for the quantitative treatment of risk project performance in the field of electric power. In the future, we can further explore the wider application of machine learning algorithms in the field of electricity bill recovery, constantly optimize methods and algorithms, improve the accuracy and reliability of performance evaluation, and provide more effective support and guidance for electricity bill recovery management decisions, so as to promote the sustainable development and improvement of the power industry.

References

[1] Gong Y, Otomo Y, Igarashi H. Sensorless metal object detection for wireless power transfer using machine learning. COMPEL: The international journal for computation and mathematics in electrical and electronic engineering, 2022, 41 (3): 807-823.
[2] Ounifi H A, Gherbi A, Kara N. Deep machine learning-based power usage effectiveness prediction for sustainable cloud infrastructures. Sustainable Energy Technologies and Assessments, 2022, 52 (Aug. Pt. A): 101967.1-101967.15.
[3] Huang J, Chen Y F, Ma X, et al. Risk Assessment Method for Power Grid Dispatching Operation Based on Markov Chain and Two-point Estimation Method. Power Capacitor & Reactive Power Compensation, 2023, 60 (3): 29-35.
[4] Zhang A A, Liu Z Q, Li Q, et al. Transformer Maintenance Strategy Considering Health Assessment and Risk Quantification. Transformer, 2023, 60 (3): 29-35.
[5] Khalkho A M, Mohanta D K. RBFNN assisted transient instability-based risk assessment of cyber-physical power system. International journal of electrical power and energy systems, 2022, 137 (May):107787.1-107787.11.

[6] Gong Y, Otomo Y, Igarashi H. Sensorless metal object detection for wireless power transfer using machine learning. COMPEL: The international journal for computation and mathematics in electrical and electronic engineering, 2022, 41 (3): 807-823.
[7] Mohammadi S F, Mohsen A S, Mengchu Z, et al. Application of machine learning methods in fault detection and classification of power transmission lines: a survey. Artificial Intelligence Review: An International Science and Engineering Journal, 2023, 56 (7): 5799-5836.
[8] Kwon S. QoS-Aware Data Center Operations Based on Chance- and Risk-Constrained Optimization. IEEE transactions on cloud computing, 2022, 10 (4): 2887-2896.
[9] Umunnakwe A, Parvania M, Nguyen H, et al. Data-driven spatio-temporal analysis of wildfire risk to power systems operation. IET generation, transmission & distribution, 2022, 16 (13): 2531-2546.
[10] XU Q, WANG S Z. Big Data Association Rule Mining Simulation Based on Improved Apriori Algorithm. Computer Simulation, 2023, 40 (7):509-513.
[11] ULLAH I, LIU K, YAMAMOTO TI, et al. Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley additive explanations. International journal of energy research, 2022, 46 (11): 15211-15230.
[12] YE T, MOU F G, DUAN Y C. Intermittent Arc Fault Detection Based on Machine Learning in Resonant Grounding Distribution Systems. CSEE Journal of Power and Energy Systems, 2023, 9 (2): 599-611.
[13] SHEHZAD F, JAVAID N, ASLAM S, et al. Electricity theft detection using big data and genetic algorithm in electric power systems. Electric Power Systems Research, 2022, 209 (Aug.): 107975.1-107975.10.
[14] SINGH M, CHAUHAN S. A hybrid-extreme learning machine based ensemble method for online dynamic security assessment of power systems. Electric Power Systems Research, 2023, 214 (Jan. Pt. B): 1-9.
[15] PENG X S, LI C, JIA S Y, et al. A short-term wind power prediction method based on deep learning and multistage ensemble algorithm. Wind energy, 2022, 25 (9): 1610-1625.