Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.22, September 2024

Studying the Intelligent Power Distribution and Consumption of the Power Grid and its Cost Structure Using Big Data Analysis

Huan Xu[1], Bin Chen[1], and Qiuyong Yang[1]

[1] China Southern Power Grid Co., Ltd.
(China)
E-mail: xuhuan@csg.cn

Abstract. Based on the analysis of big data, this paper studies the impact of user behavior response on the cost structure of the microgrid system in the power grid system. The article first conducts in-depth research on distributed power generation and energy storage systems, focusing on the principles and output characteristics of smart power distribution and utilization in power grids, researches smart power distribution and utilization systems in power grids, and makes a more comprehensive discussion of the current situation. Secondly, a big data analysis platform was built, and distributed storage and computing were studied. The platform was used to perform distributed storage and regulation of electricity consumption data, and the electricity consumption information data was divided into the important load, controllable load and transferable load, constructed a microgrid system model based on electricity consumption behavior response, and analyzed a calculation example. After that, a micro-grid system was simulated based on HOMER software, and the optimal capacity configuration of the system was performed. Under this configuration, the micro-grid system has the highest economic efficiency. At the same time, a demand-side load control system was built. Introduced distributed power as a controllable load, integrated new energy access and load control technology, coordinated the contradiction between the power grid and distributed power, and completed a cost-benefit analysis. Finally, for demand-side management electricity price response, peak-valley time-of-use price, the most important implementation method, according to the cost structure theory, the peak-valley period is divided according to the membership function in fuzzy mathematics, and the user's response model to peak-valley time-of-use price is established. The experiment uses the original data to simulate, find the user response model parameters based on the load transfer rate, and complete the comparative analysis of the effect under the peak-valley time-of-use electricity price, which is of great significance to the implementation and improvement of the peak-valley time-of-use electricity price project. The analysis results of the calculation examples show that the method constructed in this paper can effectively realize the power quality analysis of the distribution network in the big data environment. The research results provide technical support for the management of the rural grid voltage deviation of the power company, and lay the foundation for improving the safe operation and management of the power grid.

Key words. Big Data Analysis, Smart Power Distribution and Consumption of Power Grid, Cost Structure, Simulation.

1. Introduction

With the development of informatization and smart grid, various electronic components and shock loads are widely used in reality, the power quality problem in the power system is getting more and more serious, and power quality problem has become one of the research difficulties home and abroad [1]. Power supply radius, power supply points, grid structure, etc. are the key factors that cause the energy quality of the distribution network. Non-linear, impulsive, fluctuating loads and intelligent power access deteriorate the power quality of the distribution network. The fact that the area is vast, few power points, too long a power supply radius, and insufficient distribution transformer capacity have led to relatively serious power quality problems in the distribution network [2]–[5]. Therefore, this article faces the actual needs of power quality analysis of regional distribution grids, and studies how to solve the problem of power quality analysis of regional distribution grids in a big data environment.

With the continuous development of mobile Internet and other technologies, people's lifestyles have undergone great changes. People can easily use mobile phones to achieve a series of activities closely related to life such as shopping, ordering meals, booking air tickets, and purchasing electricity. In the context of the energy Internet, electricity users will be both consumers and consumers of electricity. The future smart grid should meet people's various power supply and power needs, and can know the feed-in tariffs and power generation costs of various energy sources at any time, and use this information to adjust their own power consumption behaviors to achieve energy saving, emission reduction, and reduction [6]–[8]. Using big data technology to analyze users' electricity consumption behaviors, build user electricity consumption behavior models, and provide users with practical energy-saving and electricity-saving strategies. We construct a micro-grid energy management system to realize the two-way interaction between users and the grid. Analyze a specific micro-grid case and calculate the cost savings for the micro-grid system after the user's electricity consumption behavior is optimized.

The research content of this article is the energy management system in intelligent power microgrid systems. Currently, most research on microgrid energy management systems has not taken into account the user's response to the demand side. The article combines big data and artificial intelligence algorithms to integrate multiple statistical data from the multidimensional analysis interface, displaying the current power operation situation in a certain area from multiple dimensions. The displayed data includes low voltage, three-phase imbalance, defect data, hidden danger data, heavy overload data, criterion layer data, fault data, and state evaluation data. In multidimensional analysis, information on various dimensions of the power supply area can be viewed based on the selected time and power supply area. Based on the above big data technology, a comprehensive analysis of multidimensional data is conducted on the output and demand sides of distributed energy generation, fully utilizing renewable energy to meet the electricity needs of users. This article summarizes the relevant concepts of power quality, laying a theoretical foundation for power quality analysis. Firstly, based on the characteristics of big data on power quality in distribution networks, data cleaning, outlier identification, and other methods are used to preprocess power quality monitoring data, construct big data on power quality, and provide data support for subsequent power quality analysis; Secondly, based on the characteristics of big data in quality analysis, a Hadoop based big data analysis architecture for power quality is designed, and a power quality analysis method based on Hadoop big data mining is constructed; Then, based on MapReduce technology, analyze the parallel operation mechanism of power quality big data. Finally, based on the voltage monitoring data of the regional distribution network, a MapReduce design for voltage deviation analysis of the distribution network is carried out. By designing Map functions and Reduce functions separately, rapid analysis of voltage deviation in rural power grids in the regional distribution network is achieved.

2. Literature Review

In order to solve the problem of power quality in the distribution network, it is necessary to collect various distribution network power quality data, and use certain methods to conduct scientific and effective parameter detection and analysis on the collected data, classify the power quality problems, and explore the breeding of power quality in the distribution network. The inherent law of occurrence and development provides method support for predicting or dealing with power quality problems in distribution networks, and proposes targeted prevention measures or treatment plans for power quality in distribution networks. Through power quality detection and analysis, various indicators and parameters of power quality can be extracted, and the power quality status in the power grid can be grasped in real time, and then the power quality can be analyzed in depth, and various power quality disturbance phenomena can be discovered in time [9]–[11].

Hossain *et al*.[12] stipulated the rights and obligations of the country from the perspective of the confidentiality of the use of new energy sources, especially the restrictions on confidentiality conditions for the energy giants, paving the way for the development and efficient use of smart power. The development of smart power is relatively unstable and dangerous, so it is often tried in remote areas. Based on this, Stergiou *et al*. [13] provided implementation monitoring on the use of unrenewable and stable energy, and entered the risk indicators into smart power hidden dangers. In the inspection shutdown platform, a modern sensory monitoring method was used to achieve an attempt to manage and control dangerous power grids. In the popularization and use of smart grids, problems such as artificial power theft that traditional supervision platforms cannot solve have also emerged. In response to this current situation, Hadi *et al*. [14] proposed this multi-energy integration power monitoring technology, which effectively prevents the problem of artificial power theft without video monitoring. On this basis, scholars have used statistical methods to analyze the commercialization behavior of domestic customers in detail, and the results show that the power supply model represented by large-scale contracting and circulating supply needs to be vigorously promoted among production contractors.

Therefore, in order to remove the noise in the power quality signal data, Al-Ali *et al*. [15] use the discrete wavelet model and the threshold judgment method to propose a disturbance signal denoising algorithm based on the discrete wavelet threshold. The algorithm has good denoising effects, but the threshold calculation is complicated. They used wavelet transform and rule-based methods to enhance the classification accuracy of power quality problems, and proposed a combination of wavelet multi-resolution analysis and improved Prony method to realize the identification of common power quality problems, in order to meet the complex identification accuracy of the electric energy signal. In addition, the wavelet transform is combined with other mathematical tools, such as artificial neural networks and mathematical morphology, to reduce the impact of noise, thereby realizing power quality disturbance analysis. The above method effectively analyzes power quality signals in a model constructed by combining discrete wavelet models with artificial neural networks. Akhavan-Hejazi and Mohsenian-Rad [16] use real-time stream processing technology and memory computing technology in big data to construct a wind turbine gearbox fault analysis method based on naive Bayes and neural networks. In terms of power plant optimization, scholars have established an offline power station boiler combustion history case library, used big data technology to conduct in-depth exploration and analysis of the historical case library, and matched the real-time power station boiler combustion situation with the historical case library to obtain the current power station. The optimal parameter value of boiler combustion optimizes the combustion system of the power station boiler. At present, the research results of big data technology in the field of power quality analysis are relatively few. With the development of monitoring collection technology, level and capability, the amount of monitoring data that can be analyzed for power quality is getting larger and larger. In the face of massive power quality monitoring data, the requirements for power quality analysis technology are further improved [17]. The application research of big data mining technology in the field of power quality analysis is increasing, and it has gradually become a hot spot [18]–[20].

3. Grid Smart Power Distribution and Consumption Based on Big Data Analysis and its Cost Structure Model Construction

A. *Big Data Hierarchy Analysis*

There are many information systems in the distribution network, which collect massive amounts of data every day and accumulate an increasingly large amount. These data have a large volume and a wide range of sources. According to the characteristics of microgrid data hierarchical storage, it can be decomposed into various subsystems; secondly, each subsystem is abstractly designed into individual agents, and the physical nodes of each distributed storage subsystem are mapped to basic terminal agents one by one. In terms of refining the commonality of physical nodes, each functional subsystem is abstracted into corresponding functional individual agents; finally, the system storage tasks are completed through multi-agent exercises, and the individual agents are combined to complete the distributed storage function. Figure 1 is the big data hierarchy analysis framework.

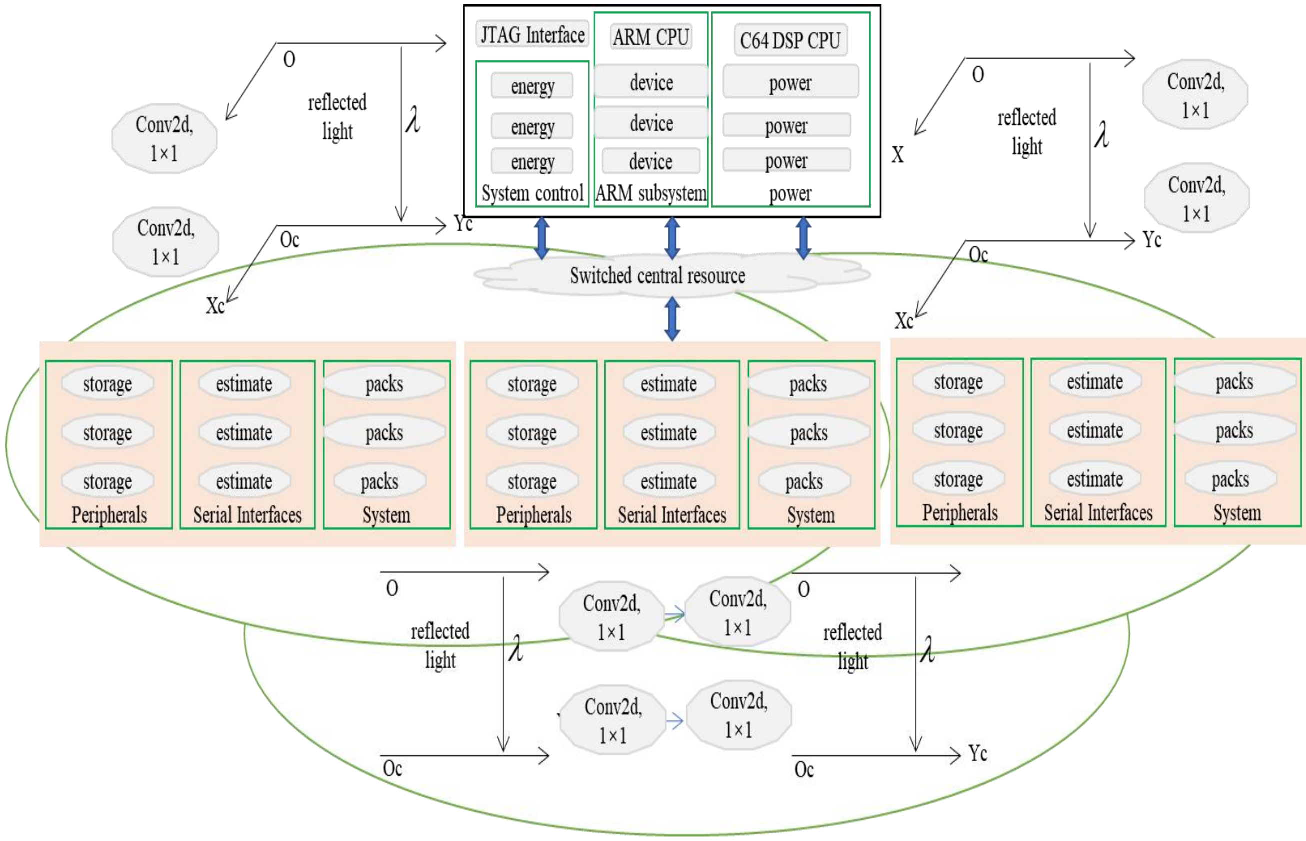

Figure 1. Big Data Hierarchy Analysis Framework

As an energy generating device for smart power storage management, BS mainly plays a role in supplying power to smart power microgrids. Therefore, it is particularly important to estimate the basic power of smart power microgrids. The battery packs are usually connected in series and in parallel to enhance the inherent driving capability of the smart power microgrid. Each branch is connected in series with the power supply. The number of independent batteries in each series is determined by the terminal voltage of the smart power microgrid. The selection of the number of branches between the smart power microgrid is determined by the total energy storage of the smart power microgrid. In addition, the security of the power grid requires necessary series redundancy between each link to ensure the continuous ability of power supply under special circumstances. The specific description is shown in formula (1):

$$\lim_{s\to\infty} U(v(m+1)) = \lim_{s\to\infty} U(v(m))U^{t}(s) = C \tag{1}$$

$$\begin{cases}[v(t-1)\,|\,v1, v2,, v_n] = [v(t)\,|\,v_n] \\ [u(t+1)\,|\,u1, u2,, u_n] = [u(t+1)\,|\,u_n]\end{cases} \tag{2}$$

The structure described in formula (2) is generally composed of one or more management agents and reaction agents with different functions, which can combine the characteristics of a distributed structure and a centralized structure. All Agent subsystems have equal status and make decisions on their own internal behaviors. The management agent only needs to uniformly deploy and manage some of the agents, solve the resource management and task planning among the agents, and exchange information with other management agents in real time. Therefore, the hybrid structure can combine the advantages of the first two structures, adapt to the complex operating environment, and become the most suitable architecture for multi-agent systems at present.

$$\begin{cases} m_1(t) = \left(U_{1,t} - I_{1,t}^2\right)/2 \\ m_1(t-1) = \left(U(m,t) - I(m,t)\right)/2 \end{cases} \tag{3}$$

$$\Delta y_t = \theta y_t - \sum_{i=1}^{p} \lambda_i \Delta y_t - \mu_i^2 \tag{4}$$

Compared with other classic load forecasting methods, the fuzzy algorithm has a great uniqueness, that is, as a non-linear system, it only controls the forecast result from the input and output angle, and ignores its calculation process. The fuzzy forecasting algorithm is obviously different from the gray theory. For example, when load forecasting is performed, the gray theory only depends on the historical load data, while the fuzzy algorithm does not depend on the historical data. The first step in using the trend extrapolation method to carry out forecasting work is to analyze the historical load trend characteristics, select the best model fit, and finally combine the differential evolution method to obtain the forecast value.

$$f(w,d) - idf(w_i) - \ln \frac{|D|}{\left|\{D_i : D_j \in w_j\} - 1\right|} = 0 \tag{5}$$

Commonly used trend models include logarithmic trend models, exponential trend models, polynomial trend models, linear trend models, power function trend models, and logistic models. The selection of the fitted curve has a great influence on the accuracy of the prediction result. The prediction result will also change with the change of the selected curve type. Therefore, the selection of the fitted curve is particularly important when applying the trend extrapolation method.

B. Correlation of the Cost Structure of Power Distribution and Consumption in the Power Grid

After designing the access unit of the microgrid, it is necessary to consider how to store a large amount of real-time and historical monitoring data: that is, how to use a limited storage space to store a large amount of analysis data. Here we use a distributed data storage method, using Google's main recommendation. The distributed file system HDFS writes data from the cloud server to the client file system in the form of master and node nodes to realize the mapping storage of large amounts of data. For microgrid data that already has distributed storage capabilities, it is necessary to consider how to use these data to study user behavior: here we use the typical K-means clustering method and key data algorithms for data mining and secondary mining; use standards pattern mapping matches user groups and user behaviors with guiding behaviors to obtain user behavior guidance. Figure 2 shows the correlation distribution of the cost structure of power distribution and consumption in the power grid.

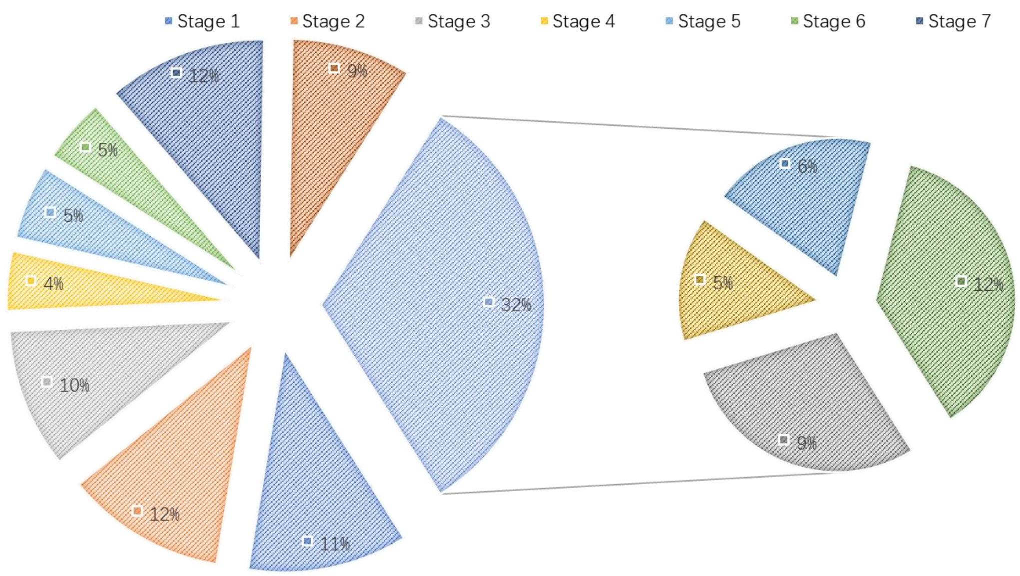

Figure 2. Correlation Distribution of the Cost Structure of Power Distribution and Consumption in the Power Grid

In terms of the overall structure, the power system is mainly composed of power plants (hydropower, wind power, thermal power, nuclear power, photovoltaics, etc.), transmission and transformation lines, substations (boost substations, load center substations, etc.) and users for consisting of electric energy production and consumption system. Because electric energy is not easy to store, the production, transmission, distribution, conversion and consumption of electric energy are almost simultaneously carried out. The producers (various generators), transporters (mainly power supply companies at all levels), and consumers (mainly including users from multiple sources) of electric energy in the power system form an organic and dynamic whole, and the components of the power system have a common influence power quality. The indicators of power quality are always dynamic changes. Power quality disturbances are potentially harmful and widely spread, and it is very difficult to evaluate their indicators. Corresponding to the voltage deviation index, the lack of reactive power of the grid system or the mismatch of the reactive power compensation capacity of the grid system is the basic cause of the deviation of the grid power supply voltage. Changes in the operation mode of the power system, changes in line impedance, active load and reactive load, etc. are the main manifestations of voltage deviation. When a positive deviation of the power supply voltage occurs, that is, when the power supply voltage is too high, it will cause equipment over-voltage, affect the insulation performance of electrical equipment and reduce equipment service life, force some reactive power compensation equipment to withdraw from operation, increase grid harmonic pollution, and pollute power system environment. For example, it causes ageing of transmission lines, insulation breakdown and overcurrent, etc.; it causes overexcitation and no-load loss of voltage transformers and other equipment to increase.

C. Grid Smart Distribution and Consumption Clustering

The design purpose of the intelligent distribution and energy storage system for the power grid is to achieve multi energy import, intelligent control, and multi data access of the intelligent power system. Among them, it is hoped that traditional electric energy and other new types of energy can be effectively introduced through energy conversion devices, and the effective energy supply of energy storage system modules can be realized through the series and parallel design of internal network battery packs. The energy storage system can be used as a power source to supply power to the load, but also it can be charged as a load by the microgrid system or the main network, and it is between the power source and the load. The purpose of energy storage system management is to calculate the amount of electricity that can be provided to the load in the energy storage system at each moment and the amount of electricity that needs to be absorbed from the microgrid or purchased from the main network, that is, to achieve real-time acquisition of customer usage data from the energy storage system need. Through the introduction of the process control algorithm, the flow of the customer side and the energy side can be controlled intelligently, and the effective feedback of the customer's power consumption data can be realized under the function of the power monitoring module to provide a platform foundation for customer behavior analysis and customer behavior guidance. Figure 3 shows the cluster distribution of smart power distribution and consumption in the grid.

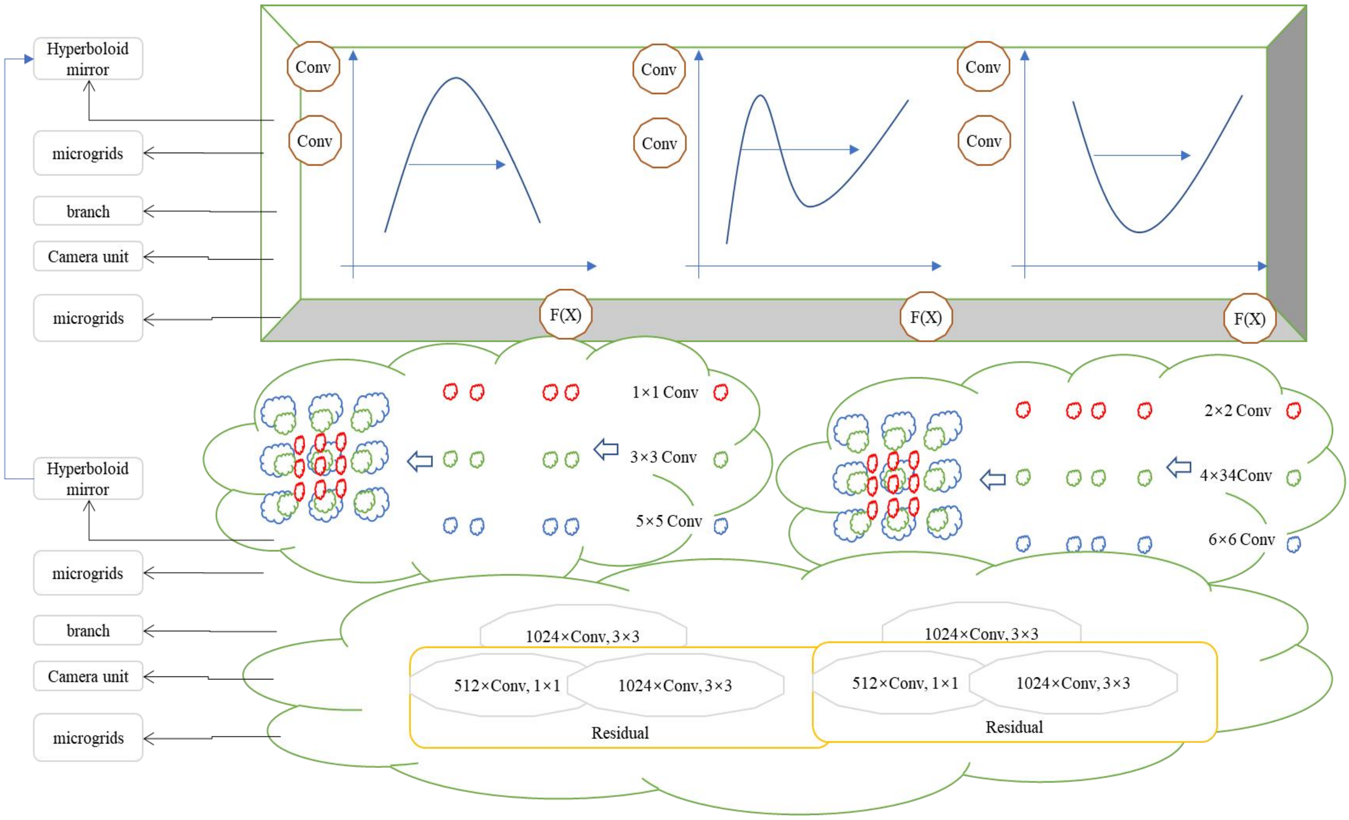

Figure 3. Cluster Distribution of Smart Power Distribution and Consumption in the Grid

In terms of data structure, power quality data mainly includes structured data, unstructured data and semi-structured data. Structured data has a standardized form and data model, which can be directly processed and analyzed by the computer. How to distribute the data collected by automation; unstructured data has no predefined model, such as video collection data of the distribution network; semi-structured data is between structured and unstructured data, such as grid geographic information system data. Non-linear equipment and loads in the power grid are the main reasons for the generation of harmonics in the power grid. For example, the quality of power generation sources is not high, the transformers and capacitors in the power transmission and distribution system, and the thyristor rectifier equipment, frequency conversion devices, electric arc furnaces, etc. in the electrical equipment. The above devices are the sources of harmonics in the power grid. When the standard sinusoidal current passes through the non-linear equipment or load, it does not show a linear relationship with the voltage, and a non-standard sinusoidal current is formed, thus forming the harmonics of the power grid. The adverse effects of grid harmonics are mainly reflected in the following aspects: for rotating electrical machines, the harmonic currents in the windings of the rotating electrical machines will increase the loss of additional power and generate noise.

D. *Structural Model Weight Rating*

In order to store multi-source distribution network power quality monitoring data, different structure model rating databases need to be built. When performing power quality analysis, it is necessary to conduct a comprehensive analysis of real-time and historical power quality monitoring data. In this way, it is necessary to centrally store the multi-modal power quality monitoring data in a distributed database or storage cluster, and then the mass power quality data is cleaned and preprocessed. The biggest challenge for the import and preprocessing of multi-modal massive power quality monitoring data is the large amount of imported data, multiple sources, and incomplete structure. User electricity behavior analysis refers to real-time evaluation of user behavior by acquiring basic household information data, electricity consumption information data, environmental data and other data. Through the accumulation of a large amount of evaluation data, the final evaluation of customer behavior and customer quality is formed. The classification of user behavior association refers to the discovery of the regularity of individual user behavior and user group behavior through the statistics of user behavior data to provide guidance for group behaviors and lay the foundation for user behavior guidance. User behavior guidance is to make correct pattern matching of user categories and regular behaviors based on associated data analysis and cluster analysis, so as to achieve the ultimate goal of accurately correcting and guiding user behaviors based on the results of data mining. Figure 4 shows the weight distribution of the smart power distribution and consumption structure model of the grid.

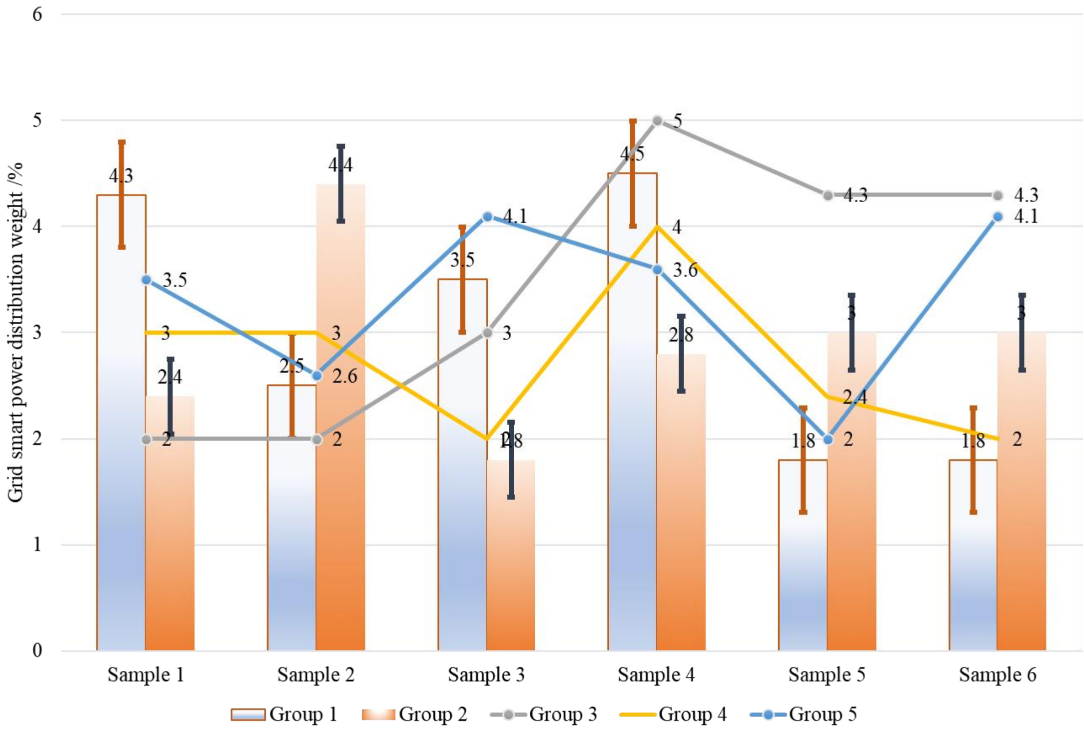

Figure 4. Weight Distribution of Smart Power Distribution Structure Model of Power Grid

Hadoop Distributed File System (HDFS) is designed as a distributed file system suitable for running on common hardware. HDFS uses a master-slave (Master/Slave) structure model to build distributed storage services. An HDFS cluster is composed of a NameNode and several DataNodes. Among them, the NameNode acts as the main server to manage the namespace of the file system and the client's file access operations, and the DataNode stores the actual data. Under normal circumstances, before the classification and analysis of the user load in the system, the structure of the system load cannot be accurately recognized, and the previous data or experience cannot be used to guide the classification, so the problem of classification load is an unsupervised classification problem. Analytical clustering is also called point group analysis and group analysis. It is a multiple statistical algorithm for classifying index problems or quantitatively studying samples. The classification objects are placed in a certain multi-dimensional space and are divided according to the degree of closeness of their spatial relationship. Generally speaking, the identification based on the different attributes of things is cluster analysis. Things with similar attributes are grouped into one category, and high similarity is reflected in the same category of things. HDFS allows users to store data in the form of files. From an internal point of view, the file is divided into several data blocks, and these several data blocks are stored on a group of DataNodes. The NameNode performs naming space operations of the file system, such as opening, closing, and renaming files or directories. It is also responsible for the mapping of data blocks to specific DataNodes.

4. Application and Analysis of Smart Power Distribution and Consumption of Power Grid and its Cost Structure Model Based on Big Data Analysis

A. Big Data Feature Extraction

Big data collection and storage are the data sources of big data processing and are the first important link in the big data processing process. Data collection is mainly to receive various types of massive data from multi-source clients through RFID radio frequency, sensor data, mobile Internet, etc., and store them in various databases or servers in a certain manner. Since there may be thousands of users accessing and operating at the same time, the degree of concurrency is high. Therefore, in the process of big data collection, the high number of concurrency is its main feature and challenge, compared to the data collected by the electricity information collection system. Based on the design of the power quality big data analysis framework and the regional hierarchical and zoning development model, the power quality big data platform adopts a hierarchical structure. The overall service-oriented SOA architecture is adopted to realize the loose coupling between the components of the platform and the collaboration of the overall platform data. The sub-platform provides various services such as power quality data integration and interface calls to the upper-level main platform. Figure 5 shows the feature extraction distribution of big data for smart power distribution.

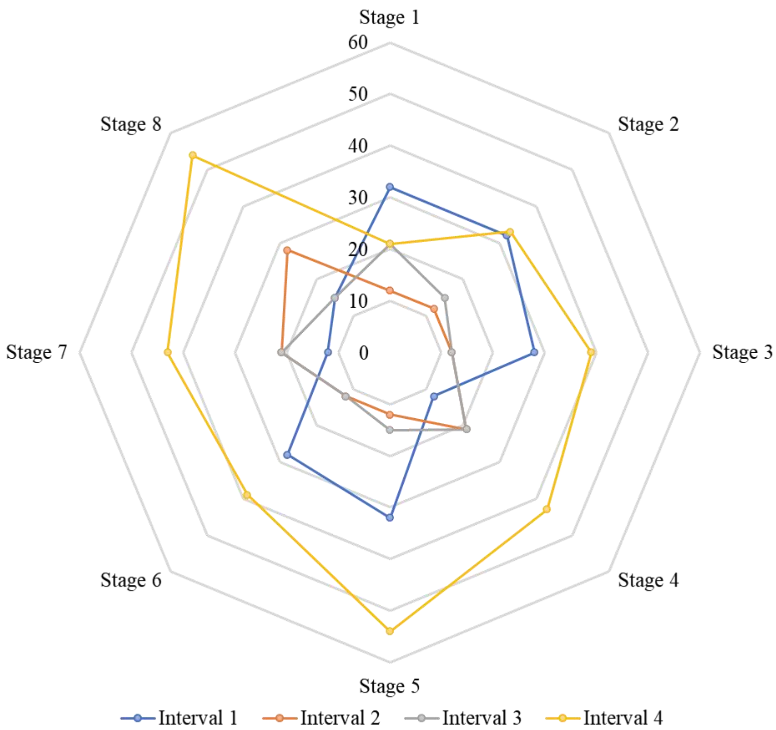

Figure 5. Feature Extraction and Distribution of Big Data for Smart Power Distribution

The calculated arithmetic mean value of each type of load curve is regarded as the typical daily load curve of such users. After cluster analysis, the 370 user loads in the original power grid can be analyzed and managed using the above 13 typical daily load curves, which greatly reduces the complexity and difficulty of implementing load management on the system. Further applications on this basis are very advantageous. At the same time, according to the typical daily load curve, the corresponding typical daily load rate, daily minimum load rate, etc. can also be determined. Since the pros and cons of the related classification results will have a certain impact on the accuracy of predictive modeling to a certain extent, even if the related results are obtained after classification, it is very necessary to analyze the degree of clustering of the curve classes. This type of curve can be modeled only when the curve has a good classification effect; if the curve is relatively discrete, it means that a certain type of classification effect is not good. In this case, in order to retain the initial information, this type of curve is required. The curve is classified twice until the final result meets the requirements.

B. Realization of Power Grid Intelligent Distribution and Utilization Simulation

The time domain simulation method is the most widely used power quality analysis method, and it can adapt to multi-source component models and system failures and operations. The time-domain simulation method mainly constructs the electronic components of the power system into a full-system model based on the topological relationship between the components, and analyzes and studies the power quality problems in combination with the time-domain simulation program. For transient power quality problems with short duration, uncertain occurrence time, and higher requirements for frequency domain analysis, such as voltage sag, voltage rise, and voltage interruption, time-domain simulation methods are often used. The power system is basically composed of generators, excitation systems, prime movers and governors, and networks and loads. Transient stability analysis mainly studies the swing characteristics of generator rotors. The actual power grid has many generators and loads. Connection and mutual influence have caused the complexity of the power system transient stability analysis. The simulation program is mainly to achieve transient stability time-domain analysis by solving a set of simultaneous differential equations and algebraic equations. The core is the initial value of the variable at t0 (usually the value of the variable under steady-state conditions calculated by the power flow). Figure 6 is a time-domain analysis of power grid smart transient stability.

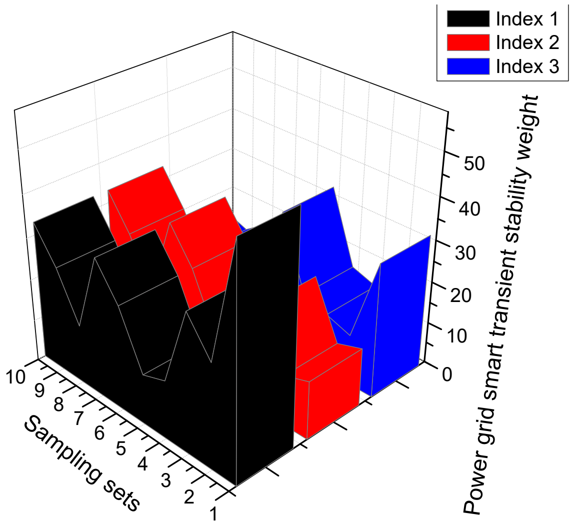

Figure 6. Time Domain Analysis of Power Grid Smart Transient Stability

Figure 6 reflects the relevant characteristic data of the daily load curve and time changes. The daily load curve can intuitively reflect the changing trend and magnitude of user load, and the curve can also more comprehensively reflect the electrical load characteristics of different users. In the trading of power contracts, the typical daily load curve is used as the basis for power futures, analysis of the peak shift benefits of interconnected systems, related distribution of electricity, and review of peak shaving capabilities. Furthermore, using the daily load curves of different types of users, the relevant substations and dispatching centers of various levels of power grids can study the impact of changes in the user structure on the power load, so as to have a more comprehensive grasp of the load changes. Compared with the fuzzy C-means clustering method, the results of the new load curve classification method are not much different, but its advantage is that it can greatly reduce the workload of classification and can be more effective and accurate. The standard deviation of various types of curves is obtained by calculation. If the standard deviation of each category can be made below 0.07, then the classification effect is very good, and the degree of curve clustering is relatively good. Table 1 is the description of the daily load peak regulation capacity of the power grid.

Table 1. Description of Daily Load Peak Regulation Capacity of Power Grid

Name	Ingredient 1	Ingredient 2	Ingredient 3
Regulation supply	80.3	53.5	44.1
Regulation demand	37.1	43.6	80.3
Power capacity	27.6	47.9	56.7
Power emergency	34.1	44.7	44.7
Regulation factors	34.4	42	86.4

In order to analyze the impact of different influencing factors on the number of emergency repair work orders, this paper introduces Pearson correlation coefficient (PCCs), respectively calculates the correlation coefficients between emergency repair work orders and influencing factors, and quantitatively analyzes the number of emergency repair work orders in different regions. The possibility that the rest of the points on the load curve lie in the valley and peak periods can be verified by the semi-trapezoidal membership function. The partial small semi-trapezoidal membership function is used to determine the possibility of each point being in the valley period, and the possibility of each point on the load curve being in the peak period is determined by the partial large membership function. The value of PCCs ranges from -1 to 1. A positive value indicates a positive correlation, and a negative value indicates a negative correlation. Under normal circumstances, the absolute value of the PCCs index below 0.3 means linear irrelevance, 0.3~0.5 means low-degree linear correlation, 0.5~0.8 means medium program correlation, and above 0.8 means highly linear correlation.

Based on the analysis of the above experimental results, it can be concluded that electricity demand involves multiple influencing factors. For electricity forecasting, it can be divided into a single indicator load forecasting model in a certain aspect, or a multi indicator load forecasting model that comprehensively considers various factors such as user structure and different user types. The latter is currently a hot topic in the research of intelligent power prediction, and its development direction can be divided into vertical in-depth research of indicators and horizontal joint research directions. This has led to various differentiated and multi indicator power load forecasting methods. In the face of different electricity demand forecasting problems, the article selects appropriate methods that are practical to establish models and conduct in-depth analysis.

C. Case Application and Analysis

In order to effectively solve the problems in power quality data storage and calculation analysis, the existing storage and calculation resources are formed into clusters through technologies such as virtualization and segmentation. The HDFS distributed file system and HBase distributed database are used to store the power quality monitoring data of this level on the cluster, provide data services for power quality analysis of the power quality monitoring points of this level, and realize the functions of power quality calculation and analysis at this level. The upper-level main platform provides multi-modal power quality monitoring data services. The frequency domain analysis method can study the stability and transient performance of the system through frequency characteristics; it can also determine the corresponding relationship between the frequency characteristics and the performance indicators of the transition process of the second-order system, and it is more convenient to analyze the influence of the parameters in the system on the system transient. The impact of the response. However, the frequency domain analysis method requires complicated calculations to obtain the frequency spectrum and restore it to a time domain signal, and the sine wave components in the frequency domain cannot reflect the moment of their occurrence. Figure 7 shows the distribution of smart grid frequency characteristics based on big data.

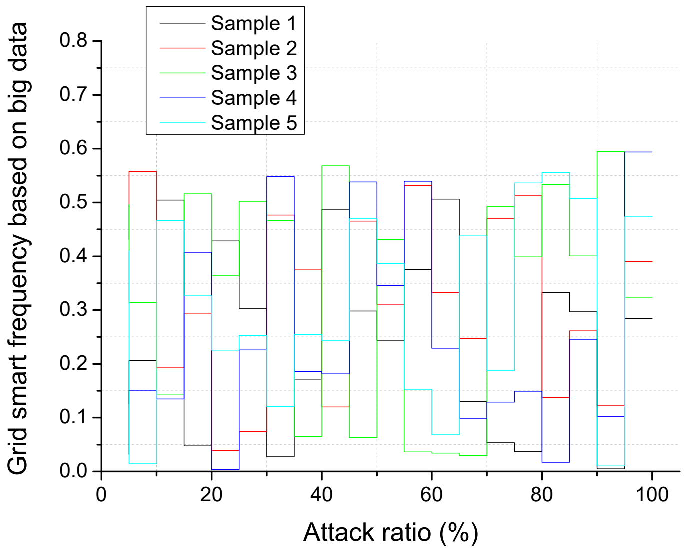

Figure 7. Distribution of Smart Frequency Characteristics of Power Grid Based on Big Data

First, the fuzzy C-means clustering (FCM) method and the load curve classification method are used to classify the user loads of the power grid. Based on the clustering algorithm theory of division, the idea of the FCM method is to ensure that the similarity between dissimilar groups is the smallest and the similarity between objects classified into the same category is the largest. At first, it is necessary to know the number of clusters and establish a fuzzy similarity matrix to start iteration, and display the clustering results until the objective function converges to the minimum. The distance clustering method is used to classify the sample data, and the optimal number of classifications is 10, and the clustering degree of these 10 load curves is obtained by analysis. Through calculation, the standard deviation of each type at each time is obtained, and the curve of the standard deviation of each type of load is drawn. After the data is preprocessed, according to the subnet division method introduced, calculate the relevance Q of emergency repair work orders between regions, and take e=0.7. According to the calculation results, the C city, D city, M city, and N city to be predicted are divided into six subnets. City C is divided into a subnet, marked as Subnet 1; City D is divided into a subnet, marked as Subnet 2; City M is divided into two subnets, marked as Subnet 3 and Subnet 4 respectively; City N is divided Are two subnets, denoted as subnet 5 and subnet 6, respectively. It can be seen from the correlation degree of each subnet repair work order curve that the Q value between different subnets after subnet division is less than e, which meets the subnet division rules. Figure 8 shows the clustering results of smart transient stability of the power grid.

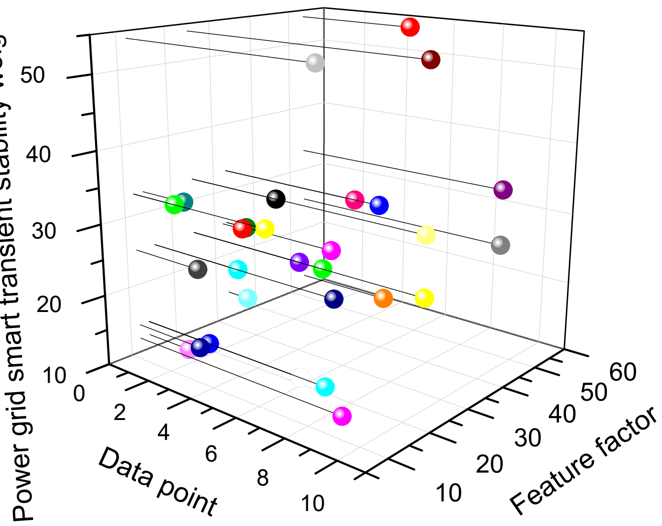

Figure 8. Power Grid Smart Transient Stability Clustering Results

According to the standard deviation effect diagram, the aggregation degree of the three curve classes (class 1, 7, and 9) is not ideal, the effect after classification is relatively poor, and the typicality of the mean curve of the load is relatively poor. This direct prediction or identification will result in large errors. Among the three categories, the better classification effect is the seventh category, and the standard deviation is controlled below 0.07. It is necessary to perform secondary classification on the curves in the three categories with unsatisfactory aggregation effects to ensure the accuracy of each category. They are sequentially classified into 6 categories, and then the 6 types of curves that have completed the secondary classification are divided into 6 categories. Analyzing the degree of clustering, and then finding the standard deviation and standard deviation curve of the six categories. For the secondary classification of the 1, 7, and 9 categories, good classification results have been achieved, and the aggregation degree of the curve classes in each category has been significantly improved. Finally, the 370 user daily load curves are divided into 13 categories, and the standard deviation of each curve is controlled below 0.07. The classification effect is good. Through the above classification process, it can be seen that the FCM method is a widely used clustering method, which can successfully complete the classification of large-scale curves, and has a good classification effect. However, because this method needs to pre-specify the number of classifications and test the effectiveness of clustering, it has the defects of low accuracy and low classification efficiency. Through the above experiments, it is verified that the distribution network operation status evaluation system implemented in this article can reflect the overall operation status of the distribution network; Through comparative analysis, weak links and deficiencies in the current power quality of the distribution network can be identified; At the same time, the necessity and feasibility of introducing big data technology in the evaluation of distribution network operation status were verified through functional and performance testing.

5. Conclusion

This paper conducts research on the prediction of the emergency repair work order of the distribution network based on big data technology, applies big data technology and artificial intelligence algorithms to the emergency repair work order prediction of the distribution network, as the algorithm basis of the fault emergency repair module in the monitoring and analysis tool in the special period. The article studies the distributed power forecasting management module. Firstly, it starts with the distributed architecture. It uses multi-agent technology, Xmesh+ZigBee technology and big data architecture to design the power supply foundation suitable for the smart power microgrid in this paper, which is designed from the three aspects of micro-grid data access, distributed storage, and intelligent use, which lays the foundation for the subsequent user analysis. The experiment transforms users' demand for purchasing smart electricity into an in-depth clustering analysis based on big data platforms. First, we introduce Google's popular big data analysis framework, and then from the perspective of the business layer, the data collection layer, data analysis layer and model building layer are separately analyzed. The model simulation based on the K-means clustering algorithm and the correlation analysis of the Apriori algorithm are made. At the same time, using the school as a case, the current situation of electricity consumption in the school was analyzed, and the electricity consumption of the student dormitory was classified and analyzed, and the electricity consumption of the laboratory was analyzed by clustering and correlation, and then the optimized electricity consumption was used. The behavior model analyzes the school's microgrid system, and the experimental results are good. Finally, the distributed prediction method for emergency repair work orders of distribution network based on big data technology was successfully applied to the fault emergency repair module in the special period monitoring and analysis tool of the power grid, mainly to realize the prediction function of emergency repair work orders of distribution network. The calculation example analysis uses the system platform to select the prediction data of emergency repair work orders from multiple cities, which verifies the accuracy and feasibility of the prediction method in this paper.

References

[1] B. P. Bhattarai *et al.*, "Big data analytics in smart grids: State-of-the-art, challenges, opportunities, and future directions," *IET Smart Grid*, vol. 2, no. 2, pp. 141-154, Mar. 2019, doi: 10.1049/iet-stg.2018.0261.

[2] K. Wang *et al.*, "Wireless big data computing in smart grid," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 58-64, Apr. 2017, doi: 10.1109/mwc.2017.1600256wc.

[3] C. Tu, X. He, Z. Shuai, and F. Jiang, "Big data issues in smart grid–A review," *Renewable Sustainable Energy Rev.*, vol. 79, pp. 1099-1107, Nov. 2017, doi: 10.1016/j.rser.2017.05.134.

[4] C. Fan, D. Yan, F. Xiao, A. Li, J. An, and X. Kang, "Advanced data analytics for enhancing building performances: From data-driven to big data-driven approaches," *Build. Simul.*, vol. 14, no. 1, pp. 3-24, Oct. 2020, doi: 10.1007/s12273-020-0723-1.

[5] W. L. Chin, W. Li, and H. H. Chen, "Energy big data security threats in IoT-based smart grid communications," *IEEE Commun. Mag.*, vol. 55, no. 10, pp. 70-75, Oct. 2017, doi: 10.1109/mcom.2017.1700154.

[6] Y. Zhang, T. Huang, and E. F. Bompard, "Big data analytics in smart grids: A review," *Energy Inf.*, vol. 1, no. 1, Aug. 2018, doi: 10.1186/s42162-018-0007-5.

[7] S. Garg et al., "Edge computing-based security framework for big data analytics in VANETs," *IEEE Network*, vol. 33, no. 2, pp. 72-81, Mar. 2019, doi: 10.1109/mnet.2019.1800239.

[8] D. Kaur, G. S. Aujla, N. Kumar, A. Y. Zomaya, C. Perera, and R. Ranjan, "Tensor-based big data management scheme for dimensionality reduction problem in smart grid systems: SDN perspective," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 10, pp. 1985-1998, Oct. 2018, doi: 10.1109/tkde.2018.2809747.

[9] S. Mujeeb, N. Javaid, M. Ilahi, Z. Wadud, F. Ishmanov, and M. K. Afzal, "Deep long short-term memory: A new price and load forecasting scheme for big data in smart cities," *Sustainability*, vol. 11, no. 4, p. 987, Feb. 2019, doi: 10.3390/su11040987.

[10] K. Wang, H. Li, Y. Feng, and G. Tian, "Big data analytics for system stability evaluation strategy in the energy Internet," *IEEE Trans. Ind. Inf.*, vol. 13, no. 4, pp. 1969-1978, Aug. 2017, doi: 10.1109/tii.2017.2692775.

[11] K. Wang, C. Xu, Y. Zhang, S. Guo, and A. Y. Zomaya, "Robust big data analytics for electricity price forecasting in the smart grid," *IEEE Trans. Big Data*, vol. 5, no. 1, pp. 34-45, Mar. 2019, doi: 10.1109/tbdata.2017.2723563.

[12] E. Hossain, I. Khan, F. Un-Noor, S. S. Sikander, and Md. S. H. Sunny, "Application of big data and machine learning in smart grid, and associated security concerns: A review," *IEEE Access*, vol. 7, pp. 13960-13988, 2019, doi: 10.1109/access.2019.2894819.

[13] C. Stergiou, K. E. Psannis, B. B. Gupta, and Y. Ishibashi, "Security, privacy & efficiency of sustainable cloud computing for big data & IoT," *Sustainable Comput. Inf. Syst.*, vol. 19, pp. 174-184, Sep. 2018, doi: 10.1016/j.suscom.2018.06.003.

[14] M. S. Hadi, A. Q. Lawey, T. E. H. El-Gorashi, and J. M. H. Elmirghani, "Big data analytics for wireless and wired network design: A survey," *Computer Networks*, vol. 132, pp. 180-199, Feb. 2018, doi: 10.1016/j.comnet.2018.01.016.

[15] A. R. Al-Ali, I. A. Zualkernan, M. Rashid, R. Gupta, and M. Alikarar, "A smart home energy management system using IoT and big data analytics approach," *IEEE Trans. Consum. Electron.*, vol. 63, no. 4, pp. 426-434, Nov. 2017, doi: 10.1109/tce.2017.015014.

[16] H. Akhavan-Hejazi and H. Mohsenian-Rad, "Power systems big data analytics: An assessment of paradigm shift barriers and prospects," *Energy Rep.*, vol. 4, pp. 91-100, Nov. 2018, doi: 10.1016/j.egyr.2017.11.002.

[17] Z. Zhou, H. Yu, C. Xu, Z. Chang, S. Mumtaz, and J. Rodriguez, "BEGIN: Big data enabled energy-efficient vehicular edge computing," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 82-89, Dec. 2018, doi: 10.1109/mcom.2018.1700910.

[18] M. Du, K. Wang, Y. Chen, X. Wang, and Y. Sun, "Big data privacy preserving in multi-access edge computing for heterogeneous Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 62-67, Aug. 2018, doi: 10.1109/mcom.2018.1701148.

[19] Y. Cao *et al.*, "Mobile edge computing for big-data-enabled electric vehicle charging," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 150-156, Mar. 2018, doi: 10.1109/mcom.2018.1700210.

[20] S. F. Rafique *et al.*, "Global power grid interconnection for sustainable growth: Concept, project and research direction," *IET Gener. Transm. Distrib.*, vol. 12, no. 13, pp. 3114-3123, May 2018, doi: 10.1049/iet-gtd.2017.1536.