

Research Article

Multidimensional Intelligent Distribution Network Load Analysis and Forecasting Management System Based on Multidata Fusion Technology

Weijie Cheng ¹, Renli Cheng ¹, Jinsheng Liu ¹, Weizhe Ma ¹, Jie Li ²,
Weiling Guan ², Daolu Zhang ² and Tao Yu ²

¹Shenzhen Power Supply Bureau Co., Ltd., Shenzhen 518000, Guangdong, China

²School of Electric Power, South China University of Technology, Guangzhou 510641, Guangdong, China

Correspondence should be addressed to Renli Cheng; 1025696549@qq.com and Tao Yu; taoyu1@scut.edu.cn

Received 17 November 2020; Revised 8 December 2020; Accepted 18 December 2020; Published 13 February 2021

Academic Editor: Wei Yao

Copyright © 2021 Weijie Cheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the work efficiency of load characteristic analysis and realize lean management, scientific prediction, and reasonable planning of the distribution networks, this paper develops a multidimensional intelligent distribution network load analysis and prediction management system based on the fusion of multidimensional data for the application of multidimensional big data in the smart distribution network. First, the framework of the software system is designed, and the functional modules for multidimensional load characteristic analysis are designed. Then, the method of multidimensional user load characterization is introduced; furthermore, the application functions and the design process of some important function modules of the software system are introduced. Finally, an application example of the multidimensional user load characterization system is presented. Overall, the developed system has the features of interoperability of data links between functional modules, information support between different functions, and modular design concept, which can meet the daily application requirements of power grid enterprises and can respond quickly to the issued calculation requirements.

1. Introduction

With the continuous socioeconomic development, the maximum load of the power grid in China's Guangdong region continues to grow, the peak-to-valley differential gradually increases, the contradiction between supply and demand of the power grid at different periods is very prominent, and the difficulty of peak adjustment continues to increase, which poses a potential hazard to the stability of the power system [1], but also brings obstacles to the planning and construction of the power grid, electricity market transactions, power load forecasting, and power market management and strategy analysis [2–7]. At the same time, under the background of industrial transformation and upgrading and gradual adjustment of economic structure, backward enterprises with high energy consumption and strong pollution in Guangdong Province have

been eliminated in large numbers, which has a significant impact on the load characteristics of the power grid. Therefore, a comprehensive study and analysis on the load characteristics of Guangdong Province is of great significance to improve the planning and construction of the overall power grid and ensure its stable operation.

In recent years, the analysis and research of load characteristics have been paid more and more attention, and researchers from local power grid companies and power industry have conducted research on load characteristics from different angles [8]. In literature [9], the Southern Power Grid Company of Yunnan adopts the sampling analysis method to study the load characteristics in different industries in Yunnan Province and makes an outlook on the development trend of load. In [10], State Grid Hunan Power Grid Company used load data to complete the calculation of several load indicators including annual maximum load and

analyzed the different load rates and their related influencing factors in different seasons based on their power load structure. Literature [11] collected real distribution network load data of Tangshan central city, focused on the analysis of residential district load, proposed indicators such as occupancy rate and demand factor to supplement the traditional load indicator system, and completed the construction of different types of district load characteristic models. The work in [12] summarizes the load characteristics indicators in the Zhengzhou Power Grid, improves the grey correlation analysis by entropy weighting, and quantifies the factors that cause changes in load characteristics to identify the main influencing factors.

On one hand, the abovementioned studies generally adopt the methods of case elimination and mean value substitution to carry out simple preprocessing on load data. These implementations may cause a waste of load data information and resources, weaken the objectivity of data, and even bring wrong results. On the other hand, the traditional load characteristic analysis often focuses on the total regional load, but lacks the analysis and comparison of the load of each industry according to the industry classification standard. In addition, in the face of the diversity of load changes and the advancement of load characteristic analysis methods, there is no comprehensive load characteristic analysis system that can keep up with the times and facilitate the development of the power grid to obtain user information and carry out reasonable planning.

To support the construction of the Greater Bay Area Smart City Cluster, China Southern Power Grid has not only accelerated the planning and construction of the Greater Bay Area Smart Grid but also adjusted its strategy and proposed “Digital China Southern Power Grid” to transform and upgrade to a comprehensive energy service provider [13]. Nowadays, the measurement system of electric power enterprises has accumulated a huge amount of fine-grained data, such as historical load data, GIS data, customer information data, and public sector service data. These data have tremendous direct and indirect value. The application of big data analysis technology and visualization and interaction technology can conduct multidimensional research and analysis of load characteristics, establish a standardized and reasonable index analysis system according to different time scales, and grasp different load change laws, while load characteristics analysis can guide the load prediction work of the distribution network. Scientific and accurate load prediction is conducive to the realization of scientific planning and management of the distribution network, to ensure the stable operation of the power grid, and to improve the economic efficiency of power grid enterprises [14].

In recent years, researchers from local power grid companies and the power industry have carried out corresponding studies from different perspectives. In [13], the airplane theory of big data in smart power distribution network application is proposed, the key technology of massive data application is sorted out, the application method of big data in different scenarios is described, and the direction of transmission and interaction of data and

energy flow is clarified through the application roadmap. Researchers in [15] studied the unified historical load data samples based on the typical daily load curves on weekdays and weekends, statistically analyzed the load characteristic indicators such as daily load rate and daily peak-to-valley differential rate, and predicted the medium- and long-term load characteristic indicators of Guizhou Power Grid to provide reference for power grid planning. In [16], the spring daily load data of Harbin are taken as a sample, and the quantitative relationship between load characteristics and meteorological factors is established by multiple regressions based on correlation analysis via the SPSS (i.e., Statistical Product and Service Solutions) software. Based on the historical load characteristic curves and meteorological data of Changde City, a statistical analysis of the main influencing factors is carried out in [17] to determine the degree of influence of specific scenarios such as small hydropower grid connection and electricity price fluctuation on regional load and the influence period and to provide revised guidance for regional load forecast based on the results. The work in [18] decomposes the residential load into basic load and seasonal load at the same time for two-level analysis, analyzes the potential size of residential load regulation based on adaptive fuzzy *c*-means algorithm clustering, and realizes the differential analysis of users' electricity consumption characteristics through a new classification method to provide data support for peak fault management and tariff formulation.

In general, although a lot of research work has been carried out on load characteristics, as well as load prediction at home and abroad, most of the current research studies separate the two separately and do not correlate them well. Most of the work uses the approaches such as case elimination or mean substitution to simply preprocess load data, which causes a waste of information and resources on load data, weakens the objectivity of the data, and may even bring wrong results. In terms of the object of analysis, most of the work is focused on the system level, lacking detailed study by industry, and the load forecast does not carry out an in-depth study on the level of substation feeder in the planning direction. In terms of engineering applications, the load management system developed at present generally has the problems of insufficient data analysis and processing ability, poor human-computer interaction performance, and low visualization degree. Therefore, in the actual application of engineering, there is a set of load characteristic comprehensive analysis system that can keep pace with the times and facilitate the grid to obtain user information and make reasonable planning.

To address the abovementioned problems, the research work of this paper is implemented based on the actual operational data in the Guangdong power grid metering system, using linear interpolation to fill in and recover the missing data in the massive load characteristic data. On this basis, a flexible and extensible multidimensional load characteristic analysis system is developed using Java development language. The system takes microservices and microapplications as the core architecture, realizes the analysis and research of load characteristics of various

industries in different time scales, and provides reference for power grid load prediction and planning and construction through comparison, statistics, and summary. Firstly, this paper introduces the overall architecture and data structure of the software system, taking into account the actual requirements of the business; then, it introduces and develops the functional implementation and logical connection of each submodule of the system; and finally, it gives an application demonstration of the software system, taking into account the actual calculation examples.

The rest of this paper is organized as follows. In Section 2, the development and design of multidimensional load characterization analysis modules are introduced. In Section 3, the multidimensional user load characteristic analysis methods are elaborated. The design process of application functions for this software system is expounded in Section 4. Furthermore, the system visualization interface development and design are demonstrated in Section 5. Section 6 conducts an application case study based on the developed software system. Finally, Section 7 concludes the paper.

2. Development and Design of Multidimensional Load Characterization Analysis Modules

In order to realize the load characteristic analysis based on the industry division, extract the load characteristic indexes of each user, summarize the change law, and put forward the specific suggestions for the subsequent load prediction and power grid planning, this paper develops a set of multidimensional load characteristic analysis system based on the Java language, using microservice and microapplication architecture. The system should meet the following technical requirements: flexible expansion of all functions; data sharing among various functional modules; data collection interface that supports various collection protocols; storage requirements for various data types and good security protection; rapid response to issued demands; and the underlying algorithm that supports the development of multiple languages at the same time.

This section introduces the development methodology of the multidimensional load characterization system, including microservice architecture design, functional modular development, data acquisition interface design, data storage and security, efficient data dynamic refresh technology, and algorithm integration design.

2.1. Microservice Architecture Design. Traditional monolithic application architectures put all functions into a single WAR/JAR package with few external dependencies to achieve centralized management and reduce development difficulty. But again, this will lead to various drawbacks such as overly centralized functionality, code and data centering; high complexity; formation of technical debt accumulation; limited scalability; and the need to refactor the project for any maintenance.

Therefore, this paper chooses to use the microservice architecture [19] as the core of software development, as shown in Figure 1. Each service is highly autonomous, and

each service focuses only on a specific business function, which makes the business clear and easy to develop and maintain [20]. At the same time, multiple services can run without waiting for the completion of other services, which achieves a smooth scalability effect and greatly reduces the start-up work time of individual microservices. In addition, when expanding a single functional module, unlike traditional architectures that require refactoring the entire project, the microservice architecture only requires the deployment of specific services that need to be expanded and modified.

The introduction of microservice architecture enables flexible expansion of the system, reduces the development cycle of each functional module, and provides great convenience for real-time updates of system functions.

2.2. Modular Development of Functions. Each function in the system is developed modularly, as shown in Figure 2. In particular, the front-end HTML (Vue or React) is requested through the Nginx gateway as an ajax request to the Restful API service.

The API gateway is built using Zuul to provide cross-module HTML page access and cross-domain data access, and the introduction of Nginx facilitates horizontal load balancing.

Each functional module is injected into the registry in the form of a microservice, and the data between the modules are exchanged through the API gateway, which facilitates load balancing and ensures the long-term stability of the entire system. Different functional modules use corresponding interfaces for data transmission and data sharing.

2.3. Data Acquisition Interface Design. As for the data acquisition interface, its design principle is shown in Figure 3. Specifically, by building a data acquisition management platform, it manages and monitors the operation state of various types of interfaces and then further implements the start/stop management of interfaces and configuration management of measurement points, so as to meet the functions of supporting various data types. The data acquisition platform supports structured data sources supported by the business system database and semistructured data sources made of electronic reports, and it also supports various acquisition protocols such as custom TCP protocol and IEC101/102/103/104 statutes. In addition, by building a data warehouse, the system integrates multisource data, uses data mining to form corresponding reports, and provides real-time query, CIM analysis, and other functions.

2.4. Data Storage and Security Module Development. In order to ensure to meet the requirements of multitype data storage, with large data volume and high fault-tolerance performance, the system platform has built a variety of databases to support a variety of data types, including columnar library, NoSql library, and memory library. At the same time, the platform supports database deployment using convenient

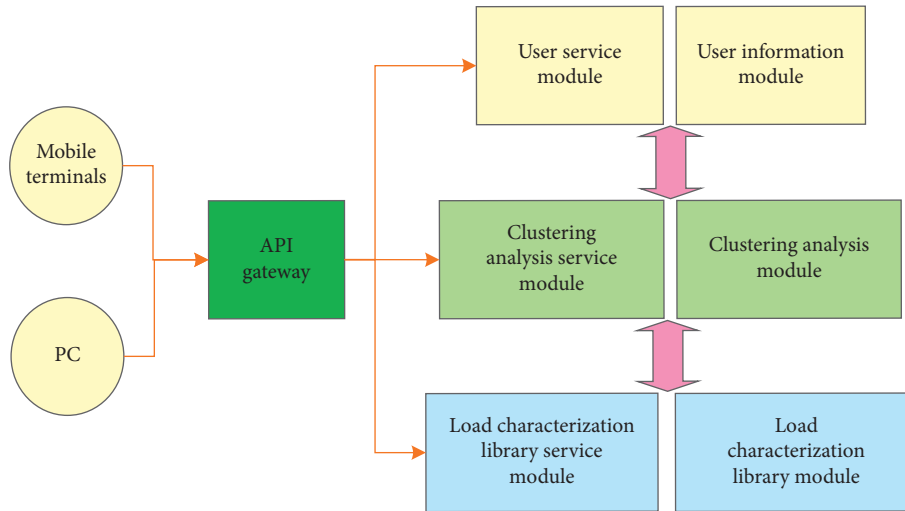


FIGURE 1: Microservice architecture design for the software system.

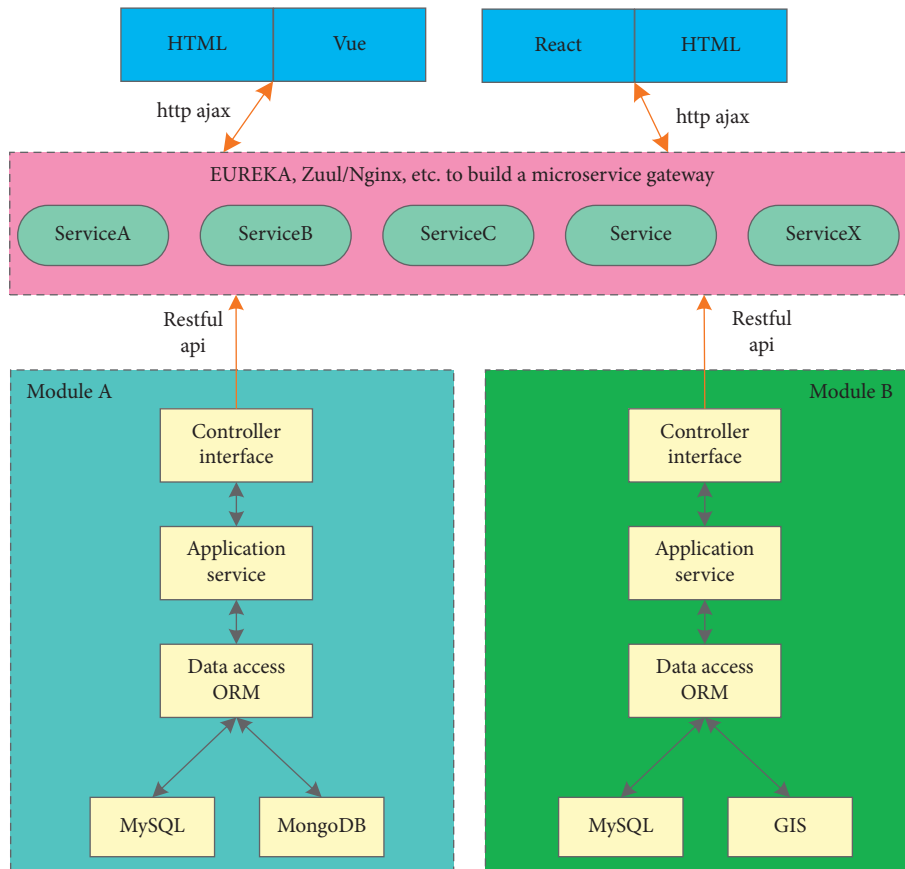


FIGURE 2: Modular development of functions for the software system.

and reliable extended multinode replica set to meet petabyte-level massive data storage, in which the master node function is set to complete the read and write of all nodes, while the rest of the replica nodes are responsible for completing the data synchronization and backup work, as shown in Figure 4. The replica nodes in the replica set use the heartbeat mechanism to monitor the state of the master node in real time and determine whether it has failed. When a

failure is detected, a master node election mechanism will be initiated within the cluster to elect a new master node to achieve automatic switching and complete the implementation of high fault tolerance.

In addition, the security part shown in Figure 4 mainly consists of two aspects: data and system. For data security, the system has established a long-term effective data backup and security management mechanism, realizing real-time

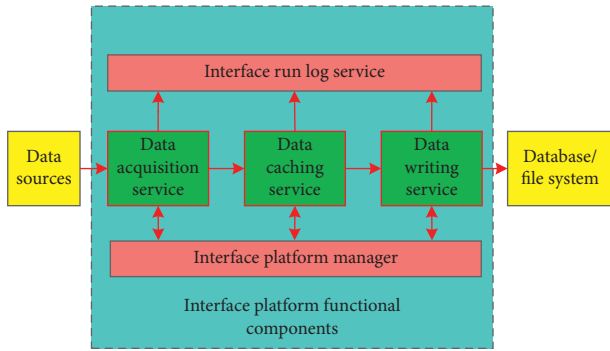


FIGURE 3: Demonstration of the design principles of the data acquisition interface.

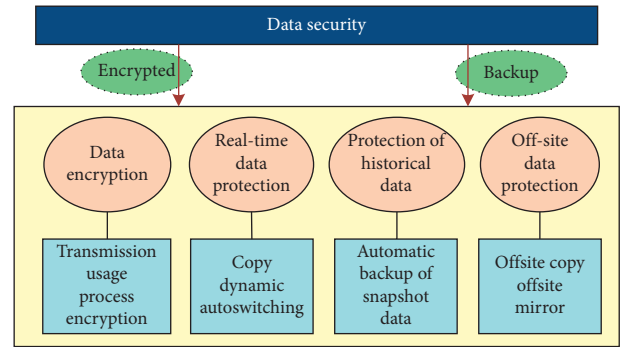


FIGURE 5: Schematic design of the data security protection module.

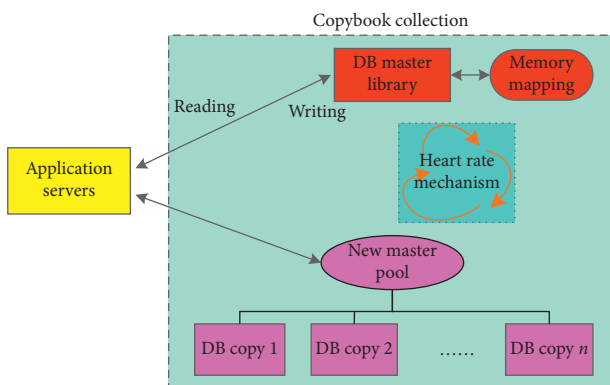


FIGURE 4: Schematic diagram of the copy set deployment model.

data protection, historical data protection, and off-site data protection, as shown in Figure 5.

In Figure 5, in terms of system security, the system ensures the stable operation of applications providing external services based on the redundant hot standby feature and adopts security protection technology to constitute Web protection against XSS cross-site script attacks and SQL script injection and other intrusions. In addition, it further strengthens protection by providing a set of authentication and authorization system for the industrial Internet platform.

2.5. Design of the Efficient Dynamic Data Refresh Technique.

In order to improve the speed of data transmission over the network when the configuration page responds to the data loading, an efficient data dynamic refresh technique is designed to improve the data loading performance, as shown in Figure 6. Specifically, we have optimized the system in the following three aspects:

- (i) Microsoft SignalR library is used based on the Web Socket duplex communication protocol to push the change data from the server to the client and improve the real-time data corresponding ability
- (ii) The configuration page supports multiple polling intervals to avoid requesting low-frequency changing data every time, such as 15-second

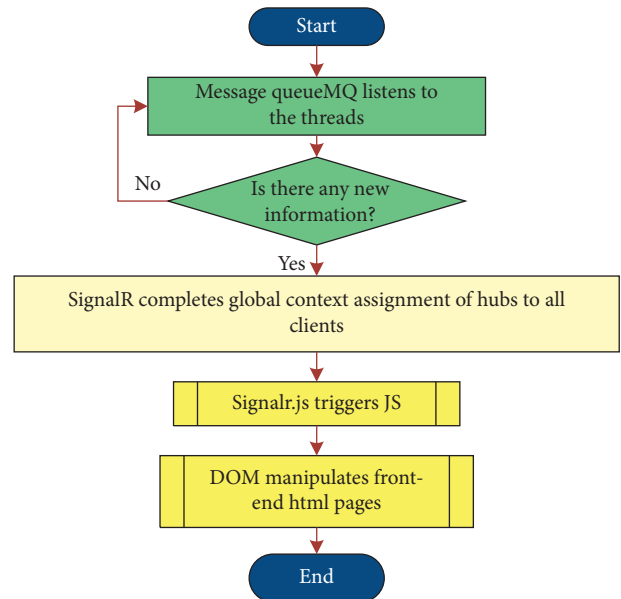


FIGURE 6: Flowchart design for dynamic data refresh.

estimates for one-hour voltage, current, and power curves

- (iii) The server side establishes correspondence with the page, and the server side compares and contrasts the real-time data of the page and only returns incremental difference data when polling the requested data

2.6. Design of Algorithm Integration. At the same time, in order to support multilingual algorithms including Java/Python/C#, the system establishes an algorithm integration platform, as shown in Figure 7. The platform architecture provides a standard unified invocation interface, http request and application/json data interaction, supports automatic identification and dynamic loading of algorithm libraries, and is integrated with microservice architecture, using Zuul gateway proxy interception to achieve algorithm access rights control. In the specific implementation of the underlying algorithm API, the system is developed in Python, taking into account the data structure specification, code simplicity, and intuition.

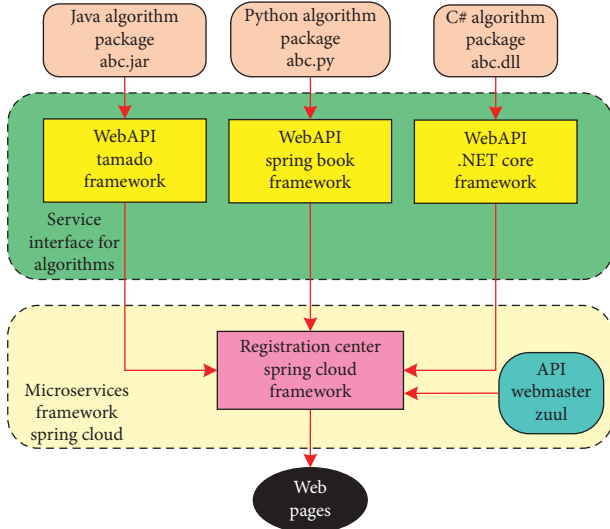


FIGURE 7: Integrated architecture design of algorithms.

3. Multidimensional User Load Characteristic Analysis Methods

In recent years, the load growth, industrial transformation, and economic structure adjustment have affected the load characteristics of power grid in many aspects. In the face of the complex situation of the load of today's power grid, to expand the load characteristics research ideas, fully exploiting the load characteristics from multiple angles can give more accurate and comprehensive guidance on employment expansion and installation, load prediction, and so on. With the help of the abovementioned simulation system, this paper establishes a single user's load characteristic database and then launches the research on the load characteristics of users in various industries for electricity consumption mode. The process is shown in Figure 8, with the following specific steps:

- (i) call the data preprocessing API to analyze the data obtained from the grid metering system, screen out abnormal data, and process them accordingly
- (ii) call the load characteristic index calculation API to extract the corresponding characteristic indexes according to the typical load curve of users in various industries
- (iii) call the daily load curve extraction method API, select the method according to the demand and application scenario, and extract the typical curve of a single user
- (iv) call the clustering algorithm API to cluster multiple users in an industry and analyze the typical power consumption patterns of users in the industry

3.1. Load Characteristic Data Preprocessing. There are various methods of data preprocessing, and considering that simple processing methods such as case exclusion and mean substitution can waste load data information and resources

and even lead to wrong conclusions, the data preprocessing API of this system adopts the linear interpolation method. The method uses a data recovery algorithm with strong continuity and autocorrelation before and after each time point on the same day, which assumes that a small number of consecutive load data points show a linear pattern of variation. The average of the complete values before and after the missing data is calculated, and the missing load is filled in. Depending on the specific location where the missing data occurs, it can be divided into two ways: first and last missing and intermediate missing. Depending on the method of missing data, there are some differences in the repair methods used. For the former, the complete data closest to the missing value will be used as the result of the repair, calculated as follows:

$$\begin{aligned} a_1 &= a_2 = \dots = a_s, \\ a_N &= a_{N-1} = \dots = a_e, \end{aligned} \quad (1)$$

where a_s is the nearest nonmissing value to the first place in the daily load curve, a_e is the nearest nonmissing value to the last place, and N is the data dimension of the single load data.

As for the latter, if it is a single missing value, then we take the average of the complete values before and after as its repair value; if it is missing at multiple consecutive points, then we find a linear expression for the nonmissing data points before and after and find out all the missing data points according to the proportion, and the formula is as follows:

$$a_x = a_m + \frac{a_n - a_m}{n - m} (x - m), \quad (2)$$

where a_x is the missing data sought and a_m and a_n are the nearest complete values before and after, respectively.

In general, the data preprocessing API of this system establishes the data preprocessing method library. The logical block diagram of the load data preprocessing method study in this method library is shown in Figure 9.

In Figure 9, the error evaluation method for data recovery uses the absolute magnitude percentage error (AMPE). In the case of multipoint loss, the mean absolute percentage error is used, which is defined as follows:

$$\text{AMPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{P'_i - P_i}{P_i} \right| \times 100\%, \quad (3)$$

where P' is the algorithm fill value, P is the true data value, and n is the number of missing points. The overall recovery effect of the algorithm can be obtained by comparing the median of the AMPE indicators of various data preprocessing methods under multiple data loss scenarios. Furthermore, the stability of the algorithm can be obtained by comparing the difference between the maximum and minimum values or the difference between the upper and lower quartiles.

3.2. Establishing Multidimensional Load Features. In general, conventional load characteristic indicators [21, 22] are selected and calculated according to three time spans: daily,

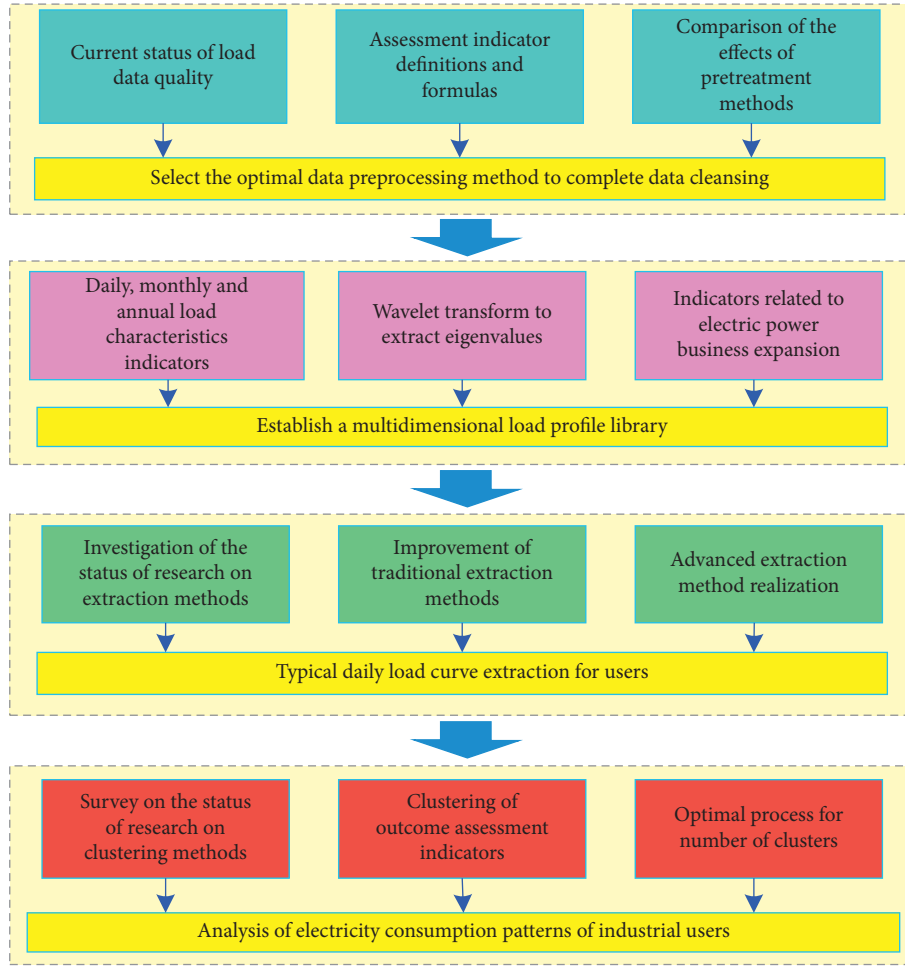


FIGURE 8: Multidimensional load characterization analysis process design.

monthly, and yearly, to describe the temporal characteristics of the load. However, the aforementioned time domain analysis results do not show the fluctuating changes of the actual load well. Therefore, the system introduces the wavelet analysis method to carry out the analysis and research of the load frequency domain. In addition, the three indicators of saturation load density, practical coefficient, and stage coefficient reflect the release law of the load after installation, which is of great significance to guide the work of industry expansion installation. The abovementioned load characteristic indexes are selected and calculated to establish the multidimensional load characteristic library, and the extraction method of each characteristic index is as follows.

3.2.1. Extraction Method of the Daily Load Characteristic Index. The daily load characteristic index can be used to calculate the daily load rate $\gamma_{d,av}$, minimum daily load rate $\gamma_{d,min}$, and peak-to-valley differential rate ϵ_d for different reporting periods (year-round, summer, and winter). The formula is as follows:

$$\begin{aligned} \gamma_{d,av} &= \frac{P_{av}}{P_{max}} \times 100\%, \\ \gamma_{d,min} &= \frac{P_{min}}{P_{max}} \times 100\%, \\ \epsilon_d &= \frac{\epsilon_{max}}{P_{max}} \times 100\%, \end{aligned} \tag{4}$$

where P_{av} , P_{max} , and P_{min} are the daily average load, daily maximum load, and daily minimum load during the reporting period, respectively, and the maximum value of daily peak-to-valley difference during the reporting period.

3.2.2. Method of Extracting the Monthly Load Characteristic Index. For the monthly load characteristic index, for different reporting periods (different months), the monthly average daily load rate $\gamma_{m,av}$, the monthly minimum load rate $\gamma_{m,min}$, and the monthly average daily peak-to-valley differential rate $\epsilon_{m,av}$ can be calculated for different reporting periods. The formula is as follows:

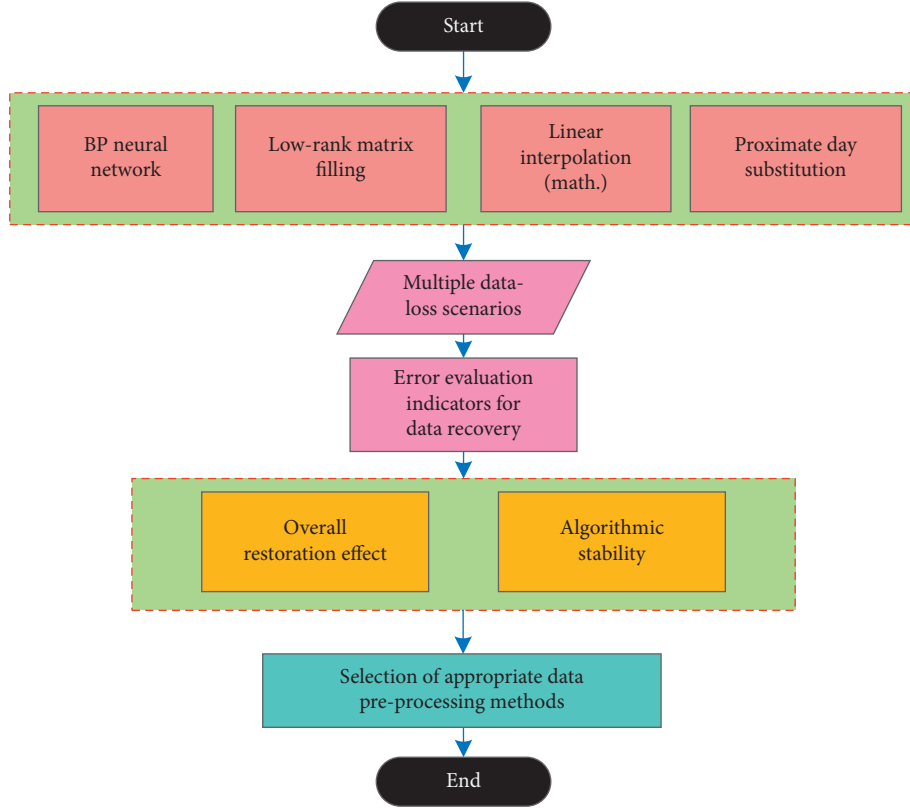


FIGURE 9: Logical block diagram of load data preprocessing method investigation.

$$\begin{aligned} \gamma_{m,av} &= \frac{\sum_{i=1}^{N_{\text{month}}} \gamma_{m,i}}{N_{\text{month}}}, \\ \gamma_{m,\min} &= \min\{\gamma_{m,1}, \gamma_{m,2}, \dots, \gamma_{m,N_{\text{month}}}\}, \\ \varepsilon_{m,av} &= \frac{\sum_{i=1}^{N_{\text{month}}} \varepsilon_{m,i}}{N_{\text{month}}}, \end{aligned} \quad (5)$$

where $\gamma_{m,i}$ is the daily loading rate on day i of the month, $\varepsilon_{m,i}$ is the daily peak-to-valley differential rate on day i of the month, and N_{month} is the number of days in the month.

3.2.3. Method of Extracting the Annual Load Characteristics Indicator. For the annual load characteristic index, the annual maximum load curve can be plotted and the annual minimum load rate, annual maximum peak-to-valley ratio, seasonal imbalance coefficient, and annual average daily peak-to-valley ratio can be calculated. The annual maximum load curve is plotted by connecting the maximum load values in each month by a polyline. The minimum annual load $\gamma_{y-\min}$, maximum annual peak-to-valley differential $\varepsilon_{y-\max}$, quarterly imbalance coefficient ρ , and average annual daily peak-to-valley differential $\varepsilon_{y-\text{av}}$ are calculated as follows:

$$\begin{aligned} \gamma_{y-\min} &= \min\{\gamma_{y-1}, \gamma_{y-2}, \dots, \gamma_{y-i}, \dots, \gamma_{y-N_{\text{year}}}\}, \\ \varepsilon_{y-\max} &= \max\{\varepsilon_{y-1}, \varepsilon_{y-2}, \dots, \varepsilon_{y-i}, \dots, \varepsilon_{y-N_{\text{year}}}\}, \\ \rho &= \frac{P_{m,av}}{P_{m,\max}} \times 100\%, \\ \varepsilon_{y-\text{av}} &= \frac{\sum_{i=1}^{N_{\text{year}}} \varepsilon_{y-i}}{N_{\text{year}}}, \end{aligned} \quad (6)$$

where, γ_{y-i} is the daily load rate on day i of the year, ε_{y-i} is the daily peak-to-valley differential rate on day i of the year, N_{year} is the number of days in the year, $P_{m,av}$ is the average value of the maximum load for each month, and $P_{m,\max}$ is the maximum value of the maximum load for each month.

3.2.4. Frequency Domain Load Characteristic Index Extraction Method. As a classical algorithm in frequency domain analysis, the wavelet transform is widely used in many research fields such as power system analysis [23]. By accurately locating the time, wavelet transform analysis can truly reflect the detailed information of power loads in the

frequency domain dimension. Therefore, it plays an outstanding role in improving the overall accuracy of load clustering and user identification, and better representing the load fluctuation. The db3 wavelet is selected to carry out a three-layer wavelet decomposition of the preprocessed user load data, with the wavelet energy, root mean square value, absolute mode mean, and standard deviation of each layer coefficient, and the calculation formula of these steady-state indicators is as follows:

$$\begin{aligned}
 Ea_i &= \sum_{j=1}^{N_1} |a_{i,j}|^2, \\
 Ed_i &= \sum_{j=1}^{N_2} |d_{i,j}|^2, \\
 rms_a_i &= \sqrt{\frac{1}{N_1} \sum_{j=1}^{N_1} |a_{i,j}|^2}, \\
 rms_d_i &= \sqrt{\frac{1}{N_2} \sum_{j=1}^{N_2} |d_{i,j}|^2}, \\
 ua_i &= \frac{1}{N_1} \sum_{j=1}^{N_1} |a_{i,j}|, \\
 ud_i &= \frac{1}{N_2} \sum_{j=1}^{N_2} |d_{i,j}|, \\
 std_a_i &= \sqrt{\frac{1}{N_1} \sum_{j=1}^{N_1} (a_{i,j} - \text{mean_}a_i)^2}, \\
 std_d_i &= \sqrt{\frac{1}{N_2} \sum_{j=1}^{N_2} (d_{i,j} - \text{mean_}d_i)^2},
 \end{aligned} \tag{7}$$

where $i = 1, 2, 3$; $a_{i,j}$ is the j th modal value of the layer i approximation coefficient; $d_{i,j}$ is the j th modal value of the layer i detail coefficient; and N_1 and N_2 are the data lengths of the approximation coefficient and detail coefficient, respectively; Ea_i represents the energy value of the layer i approximation coefficient; Ed_i represents the energy value of the layer i detail coefficient; rms_a_i represents the RMS value of the layer i approximation coefficient; rms_d_i denotes the root mean square value of the layer i coefficient of detail; ua_i denotes the absolute modal mean of the layer i coefficient of approximation; ud_i denotes the absolute modal mean of the layer i coefficient of detail; std_a_i denotes the standard deviation of the layer i coefficient of approximation; std_d_i denotes the standard deviation of the layer i coefficient of detail; $\text{mean_}a_i$ denotes the mean of the layer i coefficient of approximation; $\text{mean_}d_i$ denotes the mean of the layer i coefficient of detail; and $\text{mean_}d_i$ denotes the mean

of the layer i coefficient of approximation. The average of the layer detail coefficients.

3.2.5. Saturation Load Density, Practical Coefficients, and Phase Coefficient Extraction Methods. In order to show the real situation of electricity consumption by users, the ratio between the size of the user's load and the size of the installed capacity is defined as the utility factor, there are a total of N electricity users, then $P_{ij\max}$ is the maximum annual load of user i in year j , $P_{i\max}$ is the maximum annual load of user i in the final year, and S_i is the floor area of user i , and then, the formula for each indicator is as follows:

$$\begin{aligned}
 \rho_i &= \frac{P_{i\max}}{S_i}, \\
 \eta_i &= \frac{P_{ij\max}}{P_{ibz}}, \\
 \gamma_{ij} &= \frac{\eta_{ij}}{\eta_i},
 \end{aligned} \tag{8}$$

where ρ_i denotes the saturation load density of the user i ; η_i denotes the utility factor of the user i ; $P_{ij\max}$ denotes the annual maximum load of the user i in the final year j ; P_{ibz} denotes the reported capacity of the user i ; and γ_{ij} is the phase factor of the user i in the j th year.

3.3. Typical Daily Load Curve Extraction Method. There is no uniform regulation on the typical daily load curve extraction method, and the traditional methods mainly include the daily load rate selection method, daily maximum load selection method, and fixed day selection method [24]. According to the actual situation of power grid load data, the actual demand of power grid planning work, and the insufficiency of the abovementioned methods in terms of scientificity, the fixed daily selection method with certain limitations is removed; the daily load rate selection method is improved; the typical daily load curve is selected based on the inclusion of more daily load characteristic indexes; the improvement algorithm is called daily load characteristic index selection method; and the maximum daily load selection method is retained. In addition to the improvement of the traditional method, this paper proposes a symbolic aggregation approximation method based on time-series dimensionality reduction [25] and a typical daily load curve extraction method based on nonparametric kernel density estimation [26] in combination with cutting-edge data mining techniques to extract the typical load curves more scientifically and reasonably.

Thus, it can be seen that the typical daily load curve extraction methods are various. Based on this, the daily load extraction API of this system establishes a library of typical daily load curve extraction methods, which is described as follows.

Through investigation and summarization on the traditional typical daily load curve extraction methods, we find that there is no unified regulation on the extraction method

of typical daily load curve, and the traditional methods mainly include the daily load rate selection method, daily maximum load selection method, and fixed day selection method. Based on the actual situation of power grid load data, the actual demand of power grid planning, and the insufficiency of the abovementioned methods in terms of scientificity, the fixed day selection method is used, with certain limitations eliminated.

To this end, the daily load rate selection method is improved in the developed system. Concretely speaking, more daily load characteristics indexes are included as the basis for selecting the typical daily load curves; the improvement algorithm is developed, and it is called daily load characteristics index selection method; and the daily maximum load selection method is retained.

In addition to the improvement of the traditional methods, the software system also adopts data mining techniques to form a symbolic aggregation approximation method based on time-series dimensionality reduction and a typical daily load curve extraction method based on nonparametric kernel density estimation, so as to extract the typical daily load curve more scientifically and rationally.

Finally, the system evaluates the rationality and characterization of the typical daily load curve selection by reasonably selected evaluation indexes. Overall, the logical block diagram for the study of the typical daily load curve extraction method in this system is shown in Figure 10.

Based on Figure 10, in order to further verify the effectiveness of the abovementioned single-user daily load curve extraction algorithm, we adopt a correlation calculation method to quantify the extraction effect of the software system and further choose different numbers of sample sets of load datasets for a typical daily load curve extraction experiment. The data interval of each load curve in the sample set is 15 minutes, and the day is divided into 96 time sections for analysis. The basic idea of the abovementioned correlation calculation method is to determine the degree of correlation based on the degree of similarity between curves. In practical terms, the method is an analytical comparison of the geometry between several curves, i.e., it is believed that the closer the geometry is, the closer the development and change trend is and the greater the degree of correlation is. This method can be used to compare the degree of fit between several forecast curves corresponding to several forecasting models and one actual curve. Actually, the greater the degree of correlation, the better the corresponding forecast model and the smaller the fit error.

For example, we assume that $X_i = (x_1, x_2, \dots, x_{96})$ is the daily load curve on day i , and the typical daily load curve extracted is $Y = (y_1, y_2, \dots, y_{96})$. Correspondingly, the correlation coefficient between the two time series is calculated as r_i as follows:

$$r_i = \frac{\sum_{k=1}^{96} (x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{96} (x_k - \bar{x})^2 (y_k - \bar{y})^2}}, \quad (9)$$

where \bar{x} and \bar{y} denote the mean value of the sequences of X_i and Y and the sample correlation coefficient r of the load characteristic indicator data is a consistent estimate of the overall correlation coefficient ρ of the indicator. In fact, the closer the value of r is to 1, the stronger the correlation between the typical daily load curve, denoted by Y , and the daily load curve, denoted by X_i , is for day i . The correlation coefficient between the typical daily load curve Y and the daily load curve X_i for day n of the user's daily load curve can be calculated, namely,

$$r = \frac{1}{n} \sum_{i=1}^n r_i. \quad (10)$$

Based on this, the overall extraction effect of the algorithm can be obtained by comparing the median of the correlation r index of various typical daily load curve extraction methods; the stability of the algorithm can be obtained by comparing the difference between the maximum and minimum values or the difference between the upper and lower quartiles.

3.4. Industry-Based Load Clustering Method. According to the standard of GB/T 4754-2017-*Classification of National Economy Industries*, multiple user loads in each industry are analyzed using multiple clustering algorithms to extract the corresponding typical load curves. Among them, the system encapsulates a variety of clustering analysis methods including hierarchical clustering algorithm [27], k -means algorithm [28, 29], and Minibatch k -means algorithm in the API of the clustering analysis module. In this paper, the description of load clustering and typical daily load curve extraction module is introduced based on the k -means method.

As one of the prototype algorithms for cluster analysis, k -means has been maturely studied and applied at this stage. Specifically, these methods usually start with an initial prototype and then iterate on the results several times according to a certain rule until convergence conditions are met.

The k -means algorithm has the advantage of being simple and easy to use, and the results are generally convincing. However, the disadvantage of the k -means algorithm is that the number of classifications depends on subjective experience and it is difficult to ensure that the number of clusters selected is the optimal number. For this reason, the software system adopts an improved k -means method based on a cluster validity evaluation index to process the load data in order to extract the relevant typical daily load curves.

Concretely speaking, this paper sets the contour coefficient as an index for evaluating the effectiveness of clustering results to determine whether the selection of the number of classifications is reasonable. To evaluate the merits of the clustering results, we need to analyze both the degree of cohesion within classes and the degree of separation between classes, and the contour coefficient effectively combines the two. The overall contour coefficient can be calculated as follows:

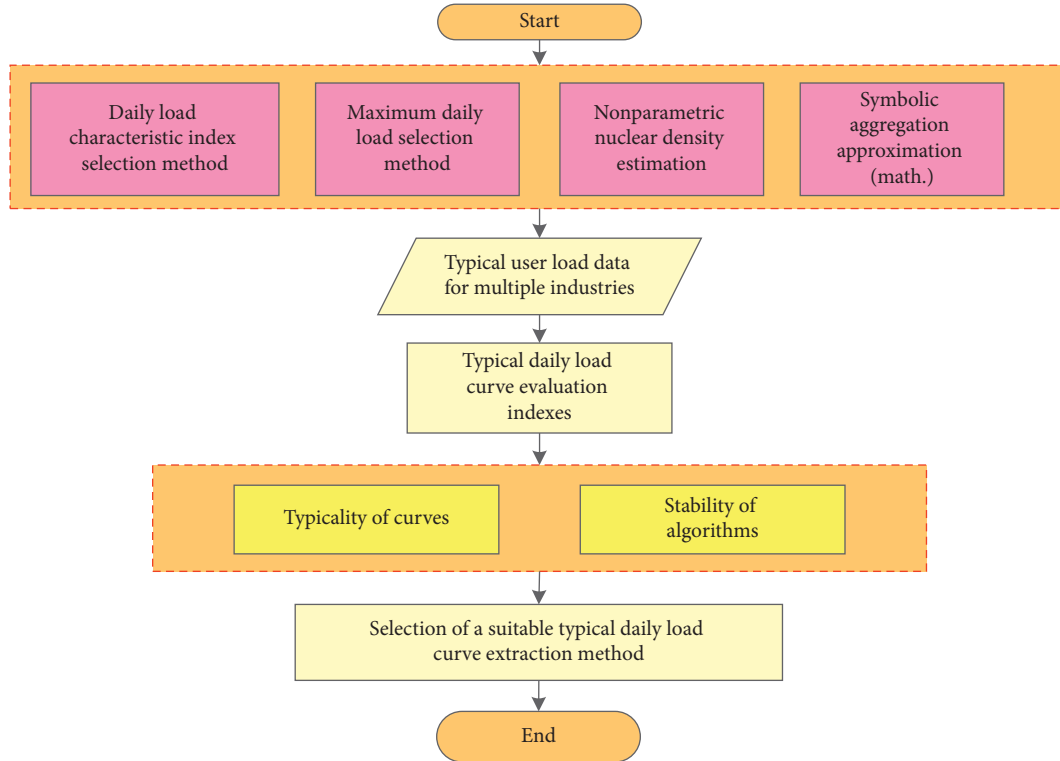


FIGURE 10: Research logic block diagram of a typical daily load curve extraction method.

$$S_{total} = \frac{\sum_{i=1}^n (b(i) - a(i)) / (\max \{a(i), b(i)\})}{n}, \quad (11)$$

where $a(i)$ is the average distance from point i to other points in the cluster it belongs to (i.e., cohesion), $b(i)$ is the average distance from point i to the nearest point in another cluster (i.e., separation), and n is the total number of clustering objects.

It can be seen that the value of the total contour coefficient ranges from -1 to 1 , and a larger value represents a higher combined cohesion and separation score, which can be considered a better clustering effect. Based on the tendency of the contour coefficient to change with the number of classifications, the most reasonable number of classifications is selected to improve the abovementioned k -means drawback. Clustering analysis is performed on the load of multiple users in an industry in order to grasp the overall electricity consumption behavior of users in that industry. The specific steps of the abovementioned content are shown in Figure 11.

According to the clustering result under the optimal number of clusters obtained in the abovementioned steps, the user load in this industry can be divided into appropriate power consumption mode types, and the clustering center is the typical daily load curve of the user load under this power consumption mode.

4. Application Functions Design

According to the application requirements of power grid companies, the smart distribution network load analysis and prediction management system developed in this paper contains several functional modules for data management,

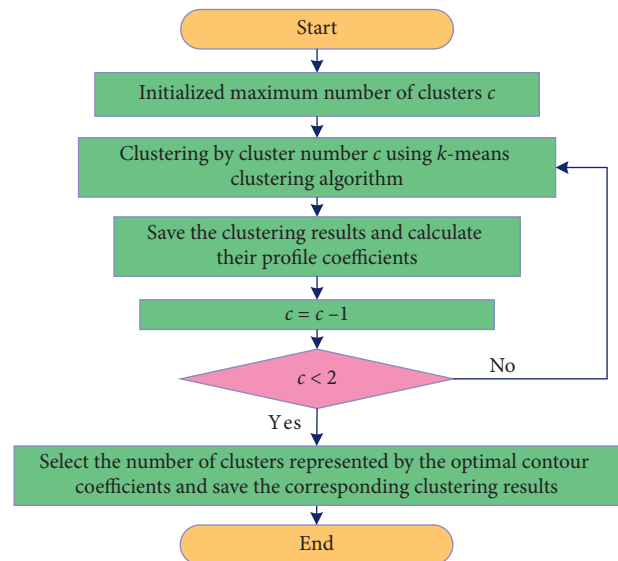


FIGURE 11: Research logic block diagram of the proposed industry-based load clustering method.

load characteristic analysis, business expansion reporting and installation, load prediction, power load optimization, distribution network planning, and system management, as demonstrated in Figure 12.

4.1. *Data Management Module.* The data management module is the basic module used for data processing, in which the user information function allows queries the

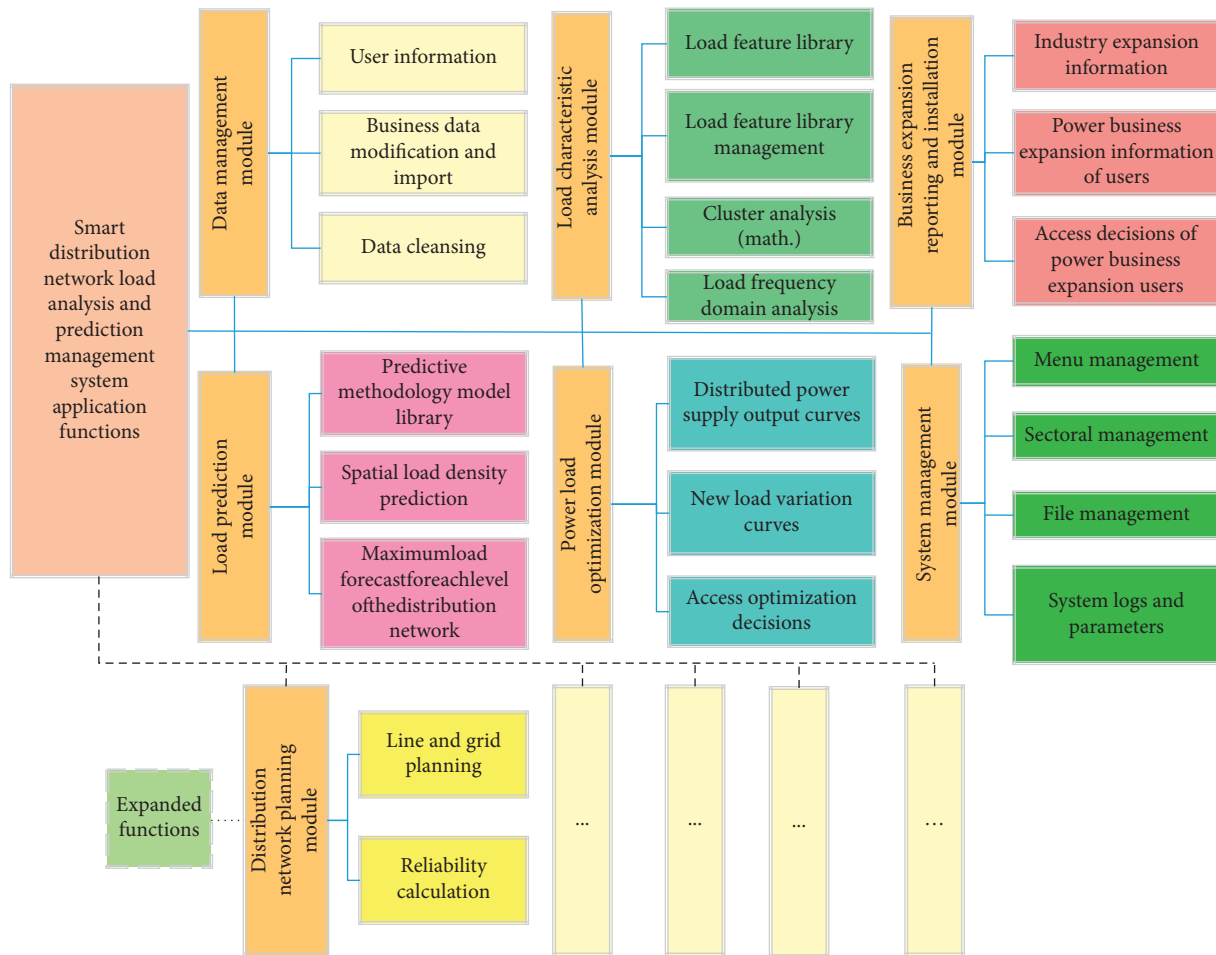


FIGURE 12: System application functional structure design.

user's name, access point, user industry type, user reported capacity, user reported installation time, floor area, scale factor, and other information, and at the same time, the data imported from the underlying database can be manually modified in this module, as well as the import of business data.

The load data are mainly data from 96 measurement points per day for distribution transformers, which requires a lot of data preprocessing operations due to the varying conditions of the measurement locations, resulting in average data quality, more bad data, and less truly usable data. Data cleaning mainly deals with missing values, zero values, sudden outliers, load curves as straight lines, and so on.

For a number of consecutive days missing a large number of data, a similar day approximation substitution method is used to repair and replace data throughout the day in order to carry out the distribution of the load addition and calculation processing functions. For single-day anomalous data, the low-rank matrix filling algorithm is used for processing, and the load curve containing the missing data and the historical load curve form a low-rank matrix, which is then calculated using the low-rank matrix filling algorithm [19], as shown in Figure 13 for the flow design of the low-rank matrix algorithm.

4.2. Load Characteristic Analysis Module. The load characteristic analysis module includes the establishment of load characteristic database and the management of the load characteristic database. The load characteristic database is calculated by a specific algorithm, and it can be divided into two categories of numerical characteristic indicators and curve-type indicators from the indicator type and three categories of daily characteristic indicators, monthly characteristic indicators, and annual characteristic indicators from the time dimension [20, 21], as shown in Tables 1–3, respectively.

In order to accurately identify the load and interpret the long and short period components contained in the load, the module sets up the load frequency domain analysis function on the basis of traditional load indicators, in order to improve the sensitivity of the load feature library to load fluctuations. The frequency domain analysis is based on the principle of wavelet transform and extracts the wavelet energy of the approximate signal, the wavelet energy of the detail signal, the root mean square value of the approximate signal, the root mean square value of the detail signal, the absolute mode mean of the approximate signal, the absolute mode mean of the detail signal, the standard deviation of the approximate signal, and the standard deviation of the detail signal as the analysis indexes.

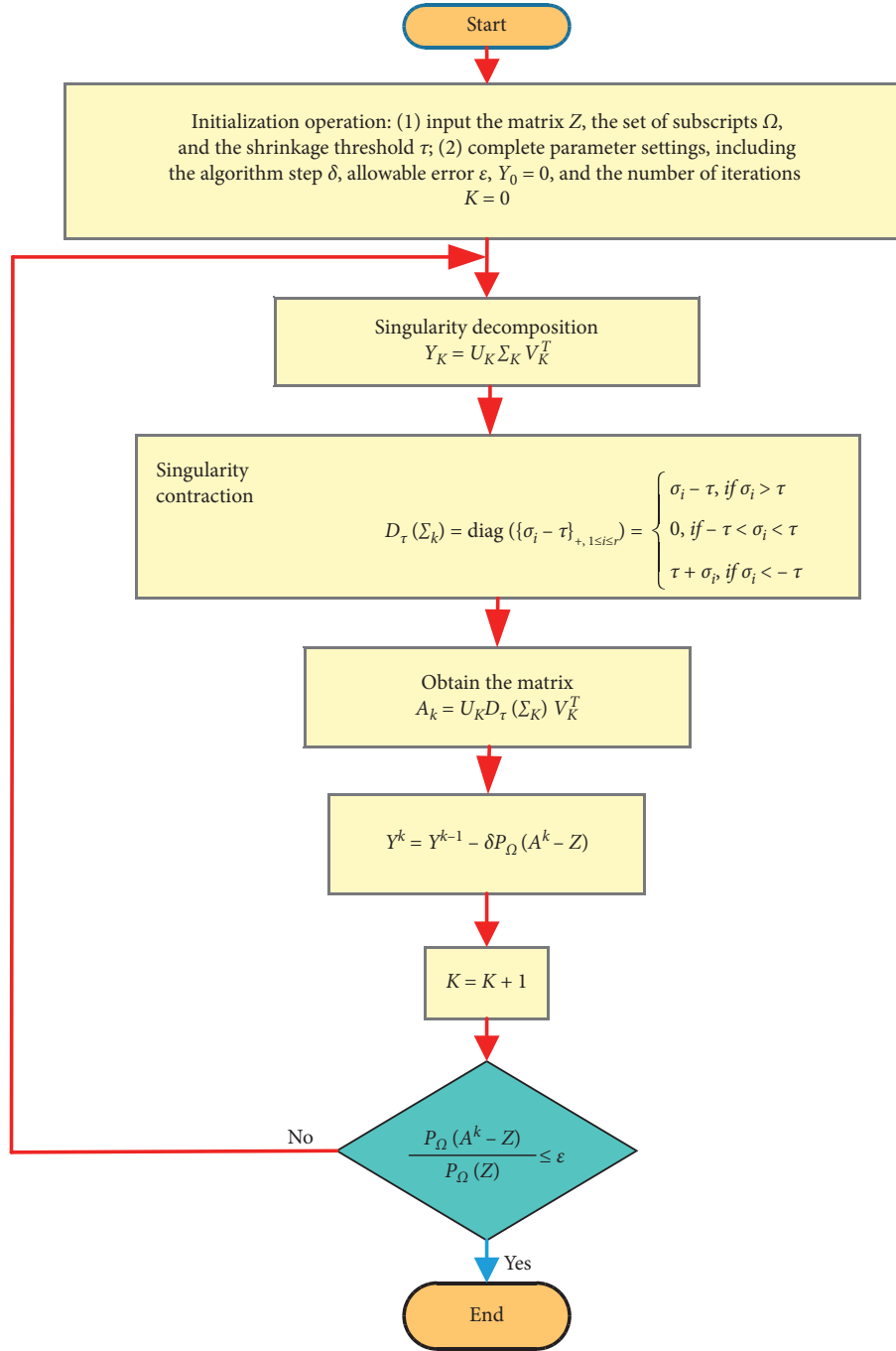


FIGURE 13: Low-rank matrix algorithm design flow.

TABLE 1: Daily load characteristics index.

Load characteristics index	Index definition
Daily maximum load	Maximum value of 96 measured point loads per day
Daily minimum load	Minimal value of 96 measured point loads per day
Average daily load	Daily average of 96 measurement point loads
Daily load rate	Ratio of average daily load to maximum daily load
Daily minimum load rate	Ratio of daily minimum load to daily maximum load
Daily peak-to-valley difference	Difference between daily maximum load and daily minimum load
Daily peak-to-valley ratio	Ratio of daily peak-to-valley difference to daily maximum load
Daily load curve	96 measurement point load connection curves per day

TABLE 2: Monthly load characteristics index.

Load characteristics index	Index definition
Monthly maximum load	Maximum load for the day of the month on maximum load day
Monthly minimum load	Minimum load for the day of the month on minimum load day
Average monthly load	Average of monthly average daily load
Monthly average daily load rate	Average of monthly daily load rates
Monthly minimum load rate	Minimum value of monthly daily load rate
Maximum monthly peak-to-valley difference	Maximum value of monthly daily peak-to-valley difference
Monthly average daily peak-to-valley ratio	Average of monthly daily peak-to-valley ratios
Monthly maximum load curve	Maximum daily load connection curve per month

TABLE 3: Annual load characteristics index.

Load characteristics index	Index definition
Annual maximum load	Maximum load for each month of the year
Annual minimum load	Minimum load for each month of the year
Annual minimum load rate	Minimum value of annual daily minimum load rate
Annual maximum peak-to-valley difference	Maximum value of annual daily peak-to-valley difference
Annual maximum peak-to-valley ratio	Maximum value of annual daily peak-to-valley rate
Quarterly imbalance coefficient	The ratio of the average of the sum of the maximum load days of the month of the year to the maximum load of the year
Annual maximum load utilization hours	Ratio of annual electricity consumption to annual maximum load
Annual maximum daily load curve	Maximum daily load connection curve per day per year

According to the industry classification, the typical daily load curves of a large number of users in the same industry are selected for clustering analysis, as shown in Figure 14, which is the typical daily load curve clustering flow chart designed in this paper.

Figure 15 demonstrates the clustering result of the summer load curves obtained from the garment industry in a certain place. Based on Figure 15, the API of the load characteristic analysis module in the developed software system encapsulates five effective clustering approaches, including fuzzy C -means clustering algorithm, fuzzy C -means clustering algorithm based on grey wolf optimization [22], hierarchy-based clustering method, k -means clustering algorithm, and Minibatch k -means clustering algorithm. Different clustering algorithms in the module will face different data samples, and they have different clustering effects; thus, the module chooses the contour coefficient as a validity indicator, which combines intraclass tightness and interclass separation as a measure of the clustering effect. The k -means algorithm, for example, has the advantage of being simple and easy to use, and the results are generally convincing. However, the disadvantage of the k -means algorithm is that the number of categories depends on subjective experience and it is difficult to guarantee that the selected number of clusters is the optimal number.

To solve the abovementioned problems, this system uses an improved k -means method based on the cluster validity evaluation index to process the load data in order to extract the relevant typical daily load curves. To solve the abovementioned problem, this paper will set the contour

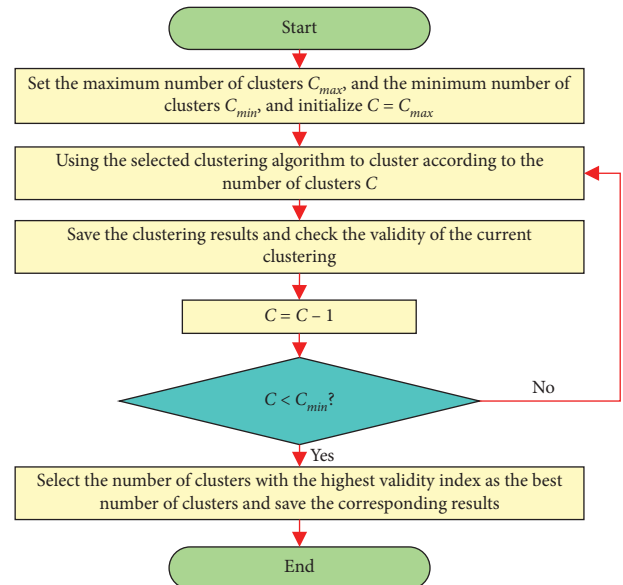


FIGURE 14: Flowchart design for clustering of typical daily load curves.

coefficient as an index for evaluating the effectiveness of clustering results to determine whether the selection of the number of classifications is reasonable. To evaluate the merits of the clustering results, we need to analyze both the degree of cohesion within classes and the degree of separation between classes, and the contour coefficient effectively

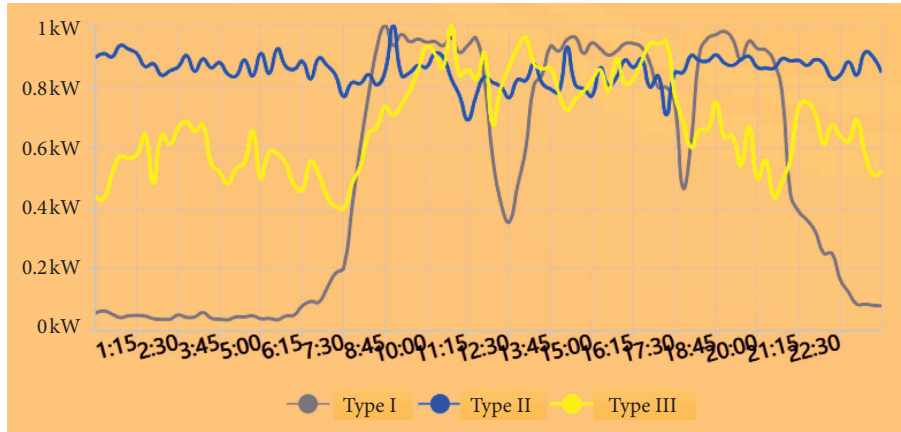


FIGURE 15: Summer daily load curve clustering results in a local garment industry.

combines the two. The overall contour coefficient [23] can be calculated by the following formula:

$$SIL_{\text{tol}} = \frac{\sum_{k=1}^{k=N} (D_1(k) - D_2(k)) / (\max\{D_1(k), D_2(k)\})}{N}, \quad (12)$$

where $D_1(k)$ is the average distance between vector k and other points in the cluster where it is located (intracluster compactness); $D_2(k)$ is the average distance between vector k and the points in the cluster closest to where it is located (inter-cluster separation); and N is the current number of clusters. It can be seen that the total profile coefficient values range from -1 to 1 , and the larger value represents higher combined cohesion and separation score, which can be considered a better clustering effect.

The abovementioned k -means drawbacks can be improved by judging the trend of the contour coefficient as the number of classifications changes and selecting the most reasonable number of classifications based on this. Clustering analysis is performed on the load of multiple users in an industry to grasp the overall electricity consumption behavior of users in that industry, as shown in Figure 14.

4.3. Business Expansion Reporting and Installation Module.

In this module, the distribution load data samples from the power grid SCADA systems are adopted and analyzed. According to the classification standard of GB/T 4754-2017 formulated in China, this module can be hierarchically divided into gates, major categories, medium categories, and minor categories. This division involves a total of 119 subcategories, which can be used to further realize differentiated and refined management and conduct information extraction and load analysis for the industry users.

According to the information of industry users' reporting time, reporting capacity, commissioning time, and load change in the year of commissioning from 1 to 3 years, the utility coefficient and stage coefficient of each user are calculated, and the confidence intervals of industry utility

coefficient and stage coefficient are determined by statistical methods to provide data support for reasonable load prediction [24]. Identifying the load patterns of large users is the key to developing a power access plan for large users. Only by understanding the daily load pattern of the newly connected large users can the system manager make better use of the complementary peaks and valleys between large users for reasonable planning. The system manager can enter the information of the user to be installed, such as the user's region, industry, installed capacity, floor area, volume ratio, the nature of electricity consumption, access time, production plan, peak hours of electricity consumption, production equipment, and nonproduction equipment, such as rated operating capacity.

Before the new users are connected, the typical load curve is selected scientifically from the load characteristic database based on the matching of installation information and industry expansion information, combined with the nature of user power consumption and production plan. Different matching methods are set up according to the information provided by the customers. For the information-rich customers, the load rate, minimum daily load rate, peak-to-valley differential rate, peak period load rate, flat period load rate, valley period load rate, and other calculation indexes are estimated by using the capacity of various types of equipment and information on usage behavior, and the load curves of different electricity consumption modes are selected by using artificial intelligence methods [30–36] such as random forest and probabilistic neural network. For users with fuzzy information, discrete simulated electricity consumption behavior curves are plotted and the Euclidean distance is used to discriminate. By establishing a scientific and reasonable matching method, the clustering curves in the load characteristic database can be selected with higher accuracy. At the same time, according to the existing information matching, the best recommended practical coefficients and stage coefficients are selected from the industry expansion information database, and the load prediction maximum value for new users in different years of the forecast period is obtained

by combining with the reported capacity. For a specific type of load, the system manager can also set special matching rules.

It is worth mentioning that, in the application process, the daily load curve of the new customer is selected by the matching algorithm, and the historical typical daily load curve is approximated as the future daily load curve for the load overlay analysis. The load curves in the feature library are standard curves, which only determine the shape, and the load level is still determined by the maximum value of the load forecast.

4.4. Load Prediction Module. Medium- and long-term load forecasting [25, 26] is a practical guide for the planning and construction of distribution networks and can initially estimate the scale of new transmission and distribution facilities to be built during the planning period. According to different forecast levels and available data sources, the model library of the medium- and long-term load forecasting method is established by integrating various forecast ideas, as shown in Figure 16.

Theoretical studies on feeder-level prediction are still scarce both at home and abroad because feeder load is smaller than system load and is subject to greater variability due to fluctuations. Feeder access to the distribution transformer load is prone to sudden changes, less stable, and difficult to seek the law of change.

In the prediction module of this paper, bottom-up distribution superposition, random forest, and top-down load allocation based on area prediction data are used as the application algorithms. Among them, the distribution overlay and random forest algorithm sink the analysis object to the level of the distribution transformer, combined with the classification criteria mentioned in the industry expansion installation module, using the data management module of the distribution transformer itself and the distribution transformer historical load data, the distribution transformer according to the public and private transformer, the nature of electricity consumption, the power supply area, the operation time, and other information to classify the distribution transformer, and different types of transformers will have different loads. For the prediction principle, the load maturity time varies for different substations, so different growth rates are set. In this case, the transformer forecasts for the same feeder are added together and multiplied by the simultaneous rate of the feeder to obtain the forecast load for the feeder.

In contrast, the top-down load allocation method first takes the 110kV substation as the object of forecast, the forecast of the 110kV main station is usually based on the current load of the 110kV main station increasing at a certain natural growth rate, and then, it adds the installed capacity of large users and the load plan transfer amount to get the result. The predicted load value is obtained for each feeder. Part of the actual feeder load data is selected for prediction [27] and then combined with the experience of experts, through manual verification and analysis of the connected distribution and other feeder state information, and the prediction results can be revised more accurately [25], as shown in Figure 17.

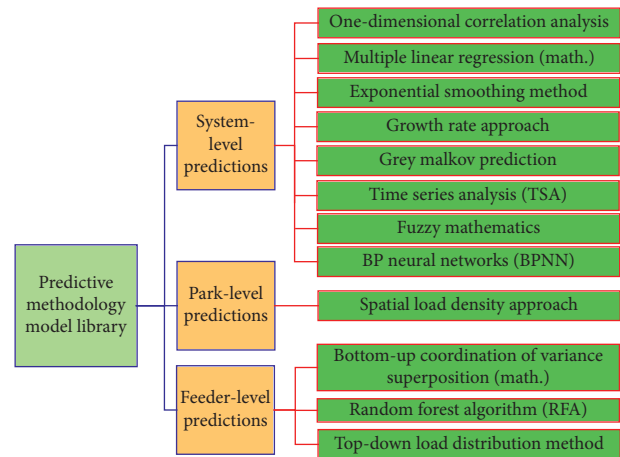


FIGURE 16: Predictive methodology model library design.

Space load density can not only calculate the maximum load value but also get the spatial distribution of the load in combination with the information of the control map of the plot, which is an important reference for the zoning of high-voltage power supply, line layout, and power supply range determination. From the calculation of historical load data, floor area, and operation time, the access load density and saturated load density of different types of users can be known, and in combination with the release coefficient of the development year, the three recommended schemes of high, medium, and low annual load value of the corresponding plot can be obtained by entering corresponding information in the function interaction column [28]. This method is used for incremental distribution network load forecasting for new parks.

The load prediction of the stock distribution network, combined with the GIS system, can clearly observe the load distribution topology and predict the maximum load of the distribution transformer in the next few years, which can roughly realize the heavy overload warning and provide data support for the capacity expansion and line modification of the existing distribution substation.

4.5. Power Load Optimization Module. The access to distributed energy sources, energy storage devices, and new types of loads such as electric vehicles changes the load characteristic curve of the traditional distribution network. Appropriate consideration of the timing characteristics, future increment, and commissioning costs of distributed energy and new types of load can effectively guide the commissioning of distributed power supply and electric vehicle charging stations in the region.

Considering the development of major new load electric private cars and space load demand, we make a forecast of the future electric vehicle ownership in the region based on the improved BASS model with reference to the data from the Shanghai Transportation Industry Development Report (2018) and the Shanghai New Energy Vehicle Industry Big Data Research Report (2018). With reference to the National Highway Traffic Safety Administration data, the modeling

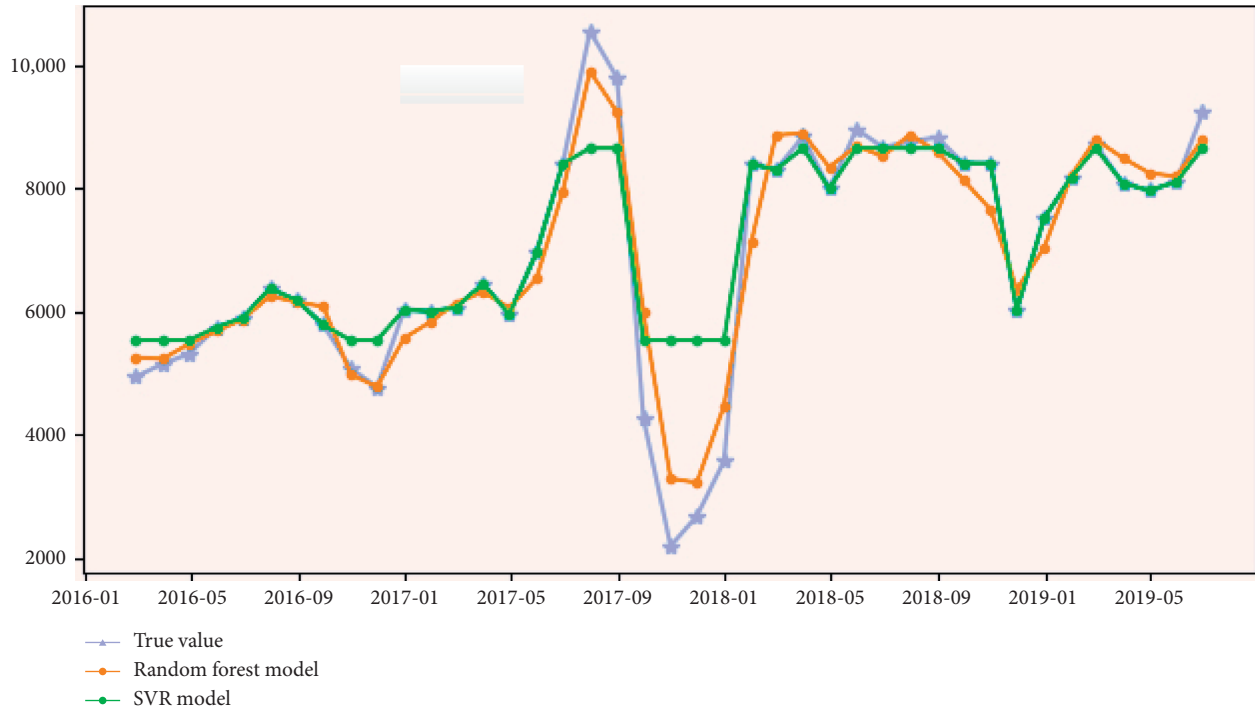


FIGURE 17: Comparative analysis of feeder-level load prediction results during Jan 2016 to May 2019.

and analysis of the temporal and spatial behavior of the vehicle based on Monte Carlo simulation is used to obtain the quantified demand of the maximum load for different charging decisions and typical EV daily load curves under different scenarios in the future forecast year [29].

The access decisions of this module are divided into load characteristic decisions, as well as economic decisions. The aforementioned industry expansion charging module can match the typical power consumption curve of new users of the industry expansion from the load characteristic analysis module through the charging information, which can be based on the load rate of the feeder near its geographic location, based on the principle of increasing the flatness of the load curve of the power supply point, to achieve the optimal decision of user access. It is also possible to make decisions on the transfer of power to existing users near the feeder to improve the load characteristic index.

At the same time, we set the initial distribution network loss, distributed power construction and operation cost, environmental benefit cost, energy storage cost simple model [37], and the access amount of PV, wind, and microgas units, combined with physical constraints to establish the solution of the multiobjective optimization function and variable weight analysis to compare the economic benefits of different conditions, for managers and investors to make a reference for decision-making.

4.6. Distribution Network Planning Module. This module is based on the load characteristics and load prediction module. After importing the basic data, the optimization algorithm is invoked to carry out park distribution network planning and quickly generate wiring schemes and

economic and technical indicators in various wiring modes to provide reference and basis for planners.

According to the order of the distribution network to carry out park planning, the first thing to do is to import the topology data and control map required for the primary distribution network frame planning from the GIS and then import the load characteristics of different industries from the load characteristics module, while supporting the addition, deletion, and modification of components. After importing the basic information, the user is supported to carry out grid drawing, and according to the cable corridor information on the control map, the path of the constructable transmission pipeline corridor is drawn on the control map, and the land parcel surrounded by the drawn path must be closed. The backend system will automatically identify the parcel area, path length, and other information based on the user's mapped line layer. At the same time, the road planning function column can display different layers, including the map, control map layer (mainly used for road drawing), load point layer (to display the load node and power node defined on the load prediction function module), path map layer (to display the manually drawn road network topology), and feeder map layer (to display the geographic alignment of the feeder once planned).

According to the plot information on the control plan, the location of each load node is marked on the drawing line layer, and the software automatically matches the user-marked load node with the plot load information in the control plan imported by the user and predicts the plot load by the spatial load density method of the load prediction module. The software also supports the user to manually complete or correct the process of matching the marked load nodes with the parcel load information in the imported

control regulations. The user can then select the wiring pattern and priority principle and customize the basic investment parameters and reliability parameters according to the actual requirements.

After deploying the abovementioned basic data, the optimization algorithm is invoked to optimize the distribution network planning model and generate the optimized primary grid, and the reliability and tidal current calculation results (including node voltage, line current, and line loss) of each feeder primary grid are also displayed for the planners to verify and adjust the planning results.

4.7. System Management Module. This module is the management operation and maintenance module of the system, though the menu management can freely adjust the system page, department management can increase the subordinate use units, and file management can download file data through the visualization of uploading, system parameters involving system copyright information and authentication code open options, and system logs to record the changes made by the user to the system operations.

Overall, as shown in Figure 18, the data full-link implementation flow of the software system is developed in this paper [38–42]. Among them, the software system selects some visual components to display the output results of the multidimensional load characteristic analysis [43]. The page of load characteristic library shows the user's typical daily load curve, day, month, year, and other regular load characteristic indexes; the page of load frequency domain analysis shows the user's load wavelet-transformed approximate signal, detailed signal, and the steady-state characteristic quantity of each layer signal; the page of the practical coefficient and stage coefficient shows the user's saturated load density, practical coefficient, and stage coefficient indexes; and the page of the practical coefficient and stage coefficient shows the user's saturated load density, practical coefficient, and stage coefficient indexes. The clustering analysis page displays the results of the electricity consumption patterns of selected industries.

Based on Figure 18, the system microservice architecture adopted in the interface of the software system developed in this paper brings flexible and expandable capabilities, and the modular development mode ensures the information interaction between each functional module, while the construction of the data acquisition management platform and the multinode replica set deployment of the database make the data interface of the system support a variety of data acquisition protocols and ensure the security of data storage. In terms of algorithm integration design, the system builds an algorithm integration platform, provides a standard unified call interface, and supports automatic identification and call of multiple languages.

5. System Visualization Interface Development and Design

In this paper, some visual components are used to display the output results of multidimensional load characteristic analysis [37–43]. The page of the load characteristic

database shows the user's typical daily load curve, day, month, year, and other general load characteristic indexes; the page of load frequency domain analysis shows the user's load wavelet-transformed approximate signal, detailed signal, and steady-state characteristic quantity of each signal layer; the page of the practical coefficient and stage coefficient shows the user's saturated load density, practical coefficient, and stage coefficient indexes; and the page of load frequency domain analysis shows the user's load wavelet-transformed approximate signal, detailed signal, and steady-state characteristic quantity of each signal layer. The clustering analysis page displays the results of the electricity usage patterns for the selected industry.

5.1. Load Feature Library Page Development. The load characteristic library analysis page developed in this system includes the typical daily load curve and characteristic index display interface, the monthly maximum load and minimum load curve interface, and the annual maximum load curve interface. As shown in Figure 19, it is the interface of the annual maximum load curve. This page adopts the nonparametric kernel density estimation method as the typical daily load curve extraction method, and the statistical time period is from 2018 to 03-01 00:00:00 to 2019-08-31 00:00:00.

This page consists of 5 sections: the typical daily load curve, daily load characteristic, maximum monthly load curve, monthly load characteristic, and annual load characteristic. The abovementioned graphical information is calculated and visualized according to the selected region, time span, user industry, user name, and typical daily load curve extraction method.

The selected region, time span, and typical daily load curve extraction method can be selected by drop-down operation, and the industry and user name provide keyword search function. The typical daily load curve section provides cursor hint function and displays specific time and data information with mouse movement.

5.2. Load Frequency Domain Characterization Page Development. The interface of user frequency domain analysis and characteristic coefficient is shown in Figure 20. The interface for load frequency domain analysis is mainly composed of two panels: figure and table. The load signal table shows the steady-state eigenvalues of approximate and detail signals obtained after 3-layer wavelet transformation of daily load data. The load signal diagram plots the approximate signals and detailed signals and displays the shape trend of each signal layer visually. The abovementioned graphical information is calculated and visualized according to the selected region, time span, and industry and user name.

5.3. Utility Factor and Stage Factor Page Development. In this system, the interface mainly consists of three panels: the practical coefficient and saturation load density, user stage coefficient, and industry stage coefficient, which are visualized

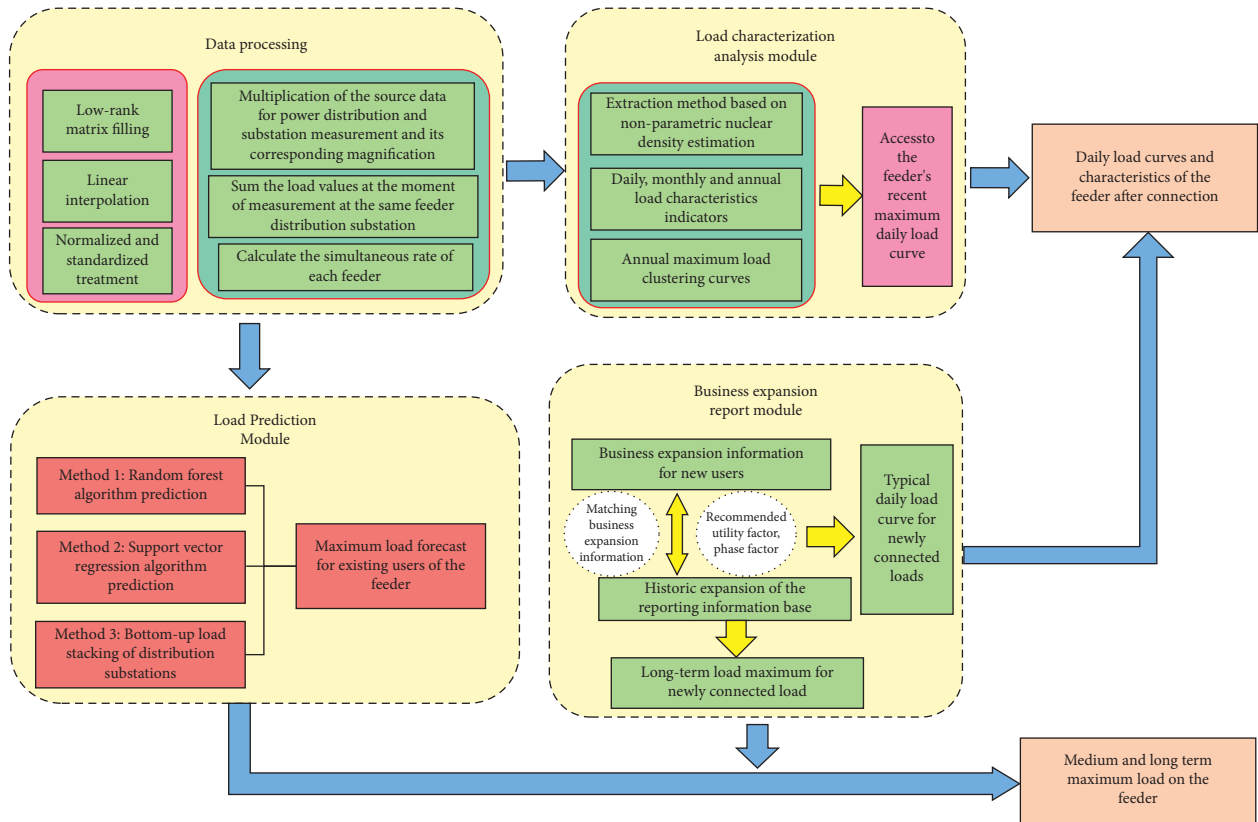


FIGURE 18: Flowchart design for the full data link implementation.

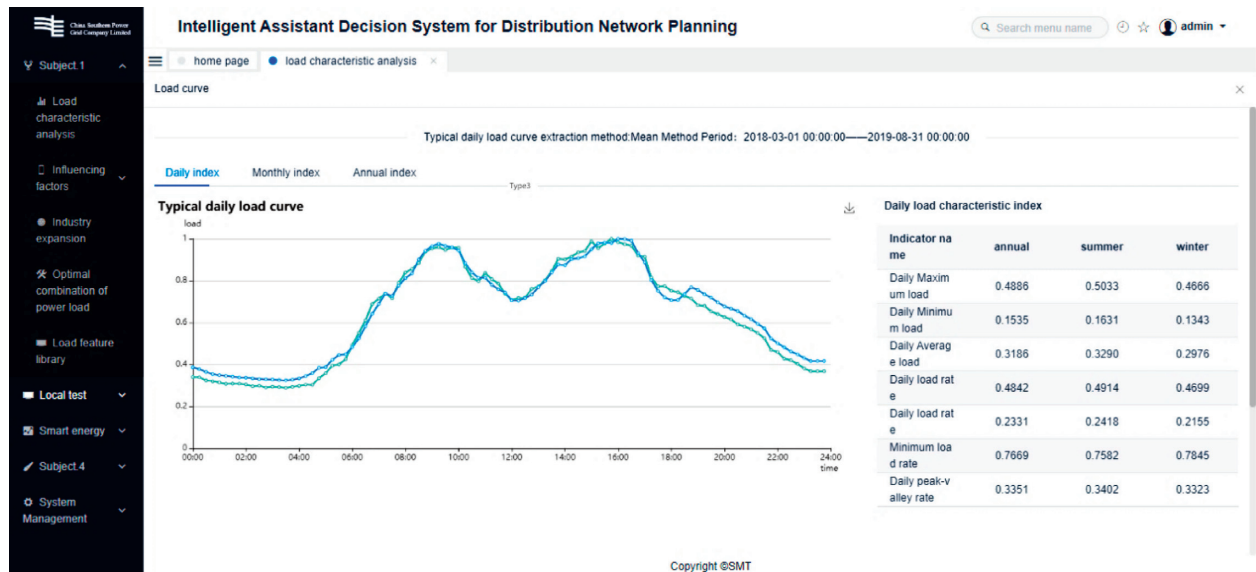


FIGURE 19: The interface display of the annual maximum load curve.

in the form of table for the specified user load and the industry to which it belongs, as demonstrated in Figure 21.

5.4. Clustering Analytics Page Development. In this system, the interface is mainly composed of two panels, namely, the industry typical daily load curve and the specified type of

load characteristic library, and the industry typical daily load curve panel plots and displays the electricity consumption patterns obtained by clustering multiple user loads in the industry. By touching the power consumption pattern legend with the mouse, you can select the power consumption pattern to be analyzed and get the characteristic database information of this type of user load. The panel

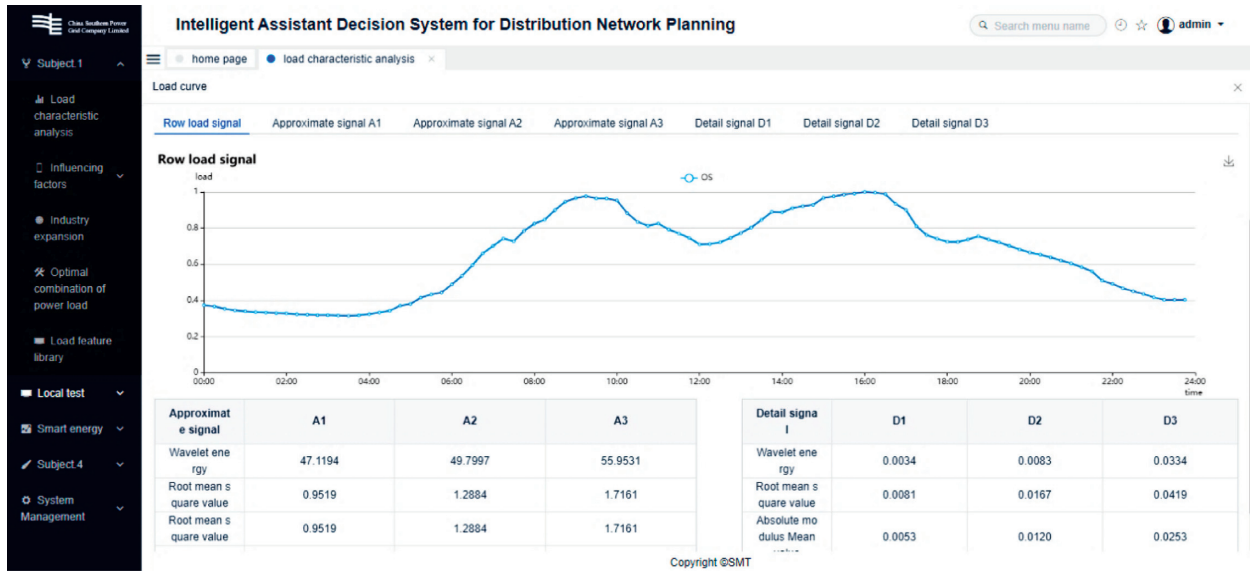


FIGURE 20: User frequency domain analysis and eigenvalue interface display.

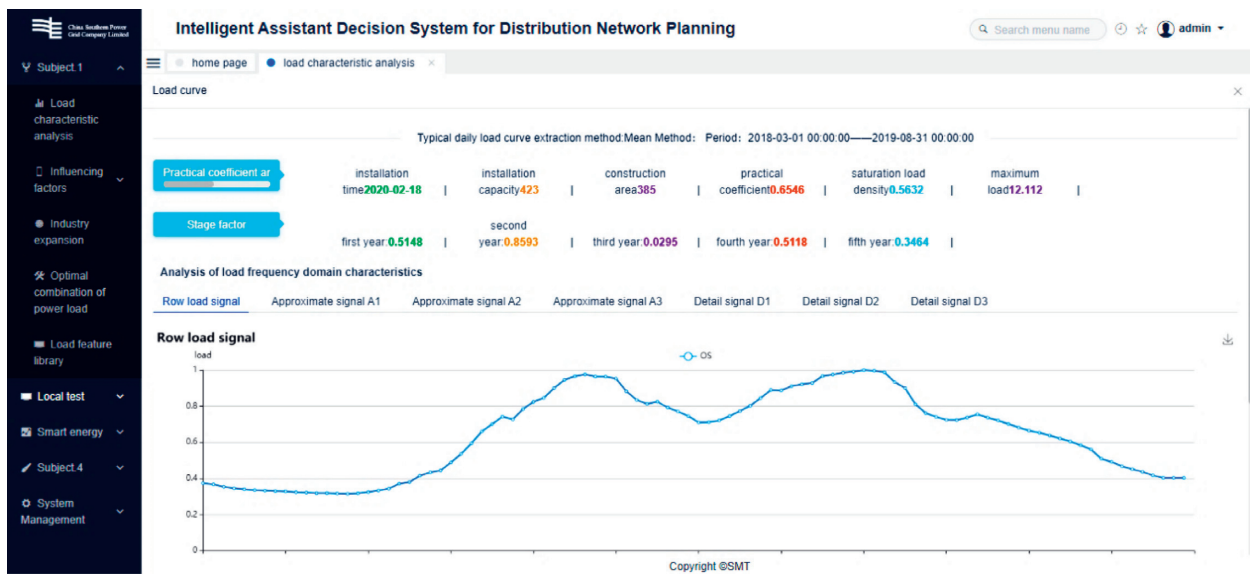


FIGURE 21: Utility factor and stage factor page display.

structure of the specified type of load characteristics database is the same as the abovementioned load characteristics database page. The abovementioned graphical information is calculated and displayed visually according to the selected region, time span, user industry, and load clustering method. Based on those mentioned above, the clustering analytics page is demonstrated in Figure 22.

6. Application Case Analysis of the Developed Software System

In order to realize the data link intercommunication between different modules of the software system, the application example selects a textile company as the business expansion user; its land area is 28,987 square meters, the

construction area is 29,886 square meters, the floor area ratio is 1.031, and the installed capacity is 630 kVA.

From the load characteristic library, two types of typical daily load curves are obtained according to the clustering curve of the textile industry, as shown in Figure 23. One of them is the typical load curve of an enterprise, where the load rises extremely fast at about 8 : 00 a.m. and falls back during lunch break at 12 : 00 noon, then re-enters the peak at about 14 : 00 a.m. and falls back briefly at about 17 : 00 a.m., and then re-enters the peak period and slowly decreases at 20 : 00 a.m. Finally, the load decreases slowly at 20 : 00 a.m. and falls back to the peak period.

The other typical curve shows the phenomenon of continuous peak without obvious peak and valley difference. Comparing and analyzing the two curves, it can be

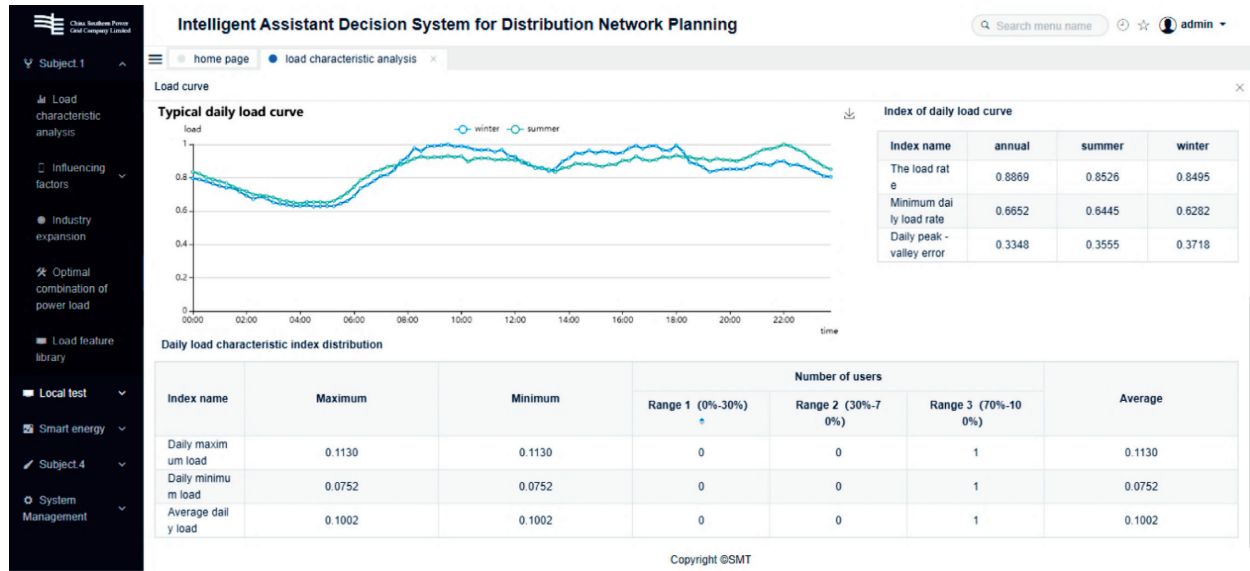


FIGURE 22: Clustering analytics page display.

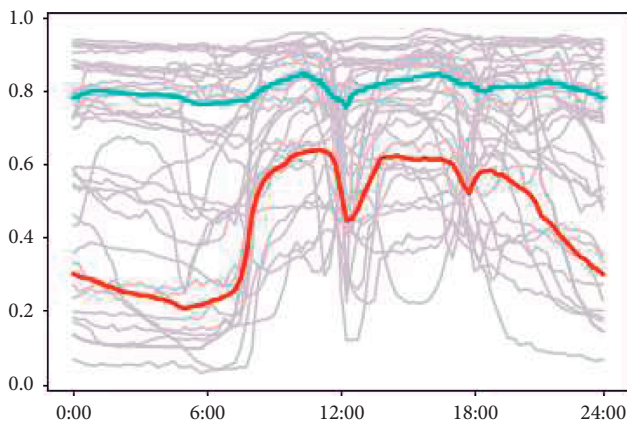


FIGURE 23: Typical daily load curve clustering results for the textile industry.

concluded that the former is a traditional hand-loom textile industry, which is not highly automated and most of the production requires human participation, so it shows a typical “three-peak” curve. But, the latter is a new type of textile industry with higher degree of automation, where most of the production work is carried out by machines instead of performing manually, and can run all day, so there is no obvious difference between time periods.

By comparing the production schedule information of the customer, this textile customer is not yet fully automated and involves other human activities; thus, a triple peak type of curve is chosen as a typical daily load curve.

The location of the user can choose two feeders 721 and 709. The load characteristic module can get the clustered typical daily load curves of these two feeders in the period where the maximum load is. By superimposing the new customer’s typical daily load curves on each of the two feeders, the new feeder load curves and characteristics can be obtained. As can be seen in Figures 24

and 25, connecting to the 721 feeder reduces the peak-to-valley differential and increases the minimum load factor, while connecting to the 709 feeder increases the peak-to-valley differential, and finally, the 721 feeder is chosen in combination with the feeder load.

The two feeders, No. 721 and No. 709, are available for this customer location. The Load Characterization module can be used to obtain clustered typical daily load curves for each of these two feeders for the time period where the maximum load is located. By superimposing the typical daily load curves for the new customer on each of the two feeders, the new feeder load curves and characteristics can be obtained. As can be seen from Figures 24 and 25, access to the 721 feeder can reduce the peak-to-valley differential ratio and increase the minimum load factor, while access to the 709 feeder will increase the peak-to-valley differential ratio, which, combined with the load of the feeder itself, is the final choice to access the 721 feeder.

The aforementioned known new user area installed capacity, industry information, and production plans, so you can match the approximate industry expansion information in the industry expansion information base to get the recommended 1~3 years of operation of the utility factor, and $\eta_1 = 0.550$, $\eta_2 = 0.734$, and $\eta_3 = 0.762$ according to the installed capacity can be the expected load value of each year. From the historical load information of feeder 721, three forecasting methods are applied to forecast its load maximum, and the weighted combination method is used to obtain the expected maximum load value in the following years, as shown in Table 4.

We use the principle of near large and far small and weigh the simultaneous rates for each year of feeder 721 to get the new simultaneous rate as the predicted simultaneous rate $Sp = 0.756$. The maximum user load is multiplied by the simultaneous rate and superimposed on the feeder load to get the predicted maximum load of the feeder in the next few years.

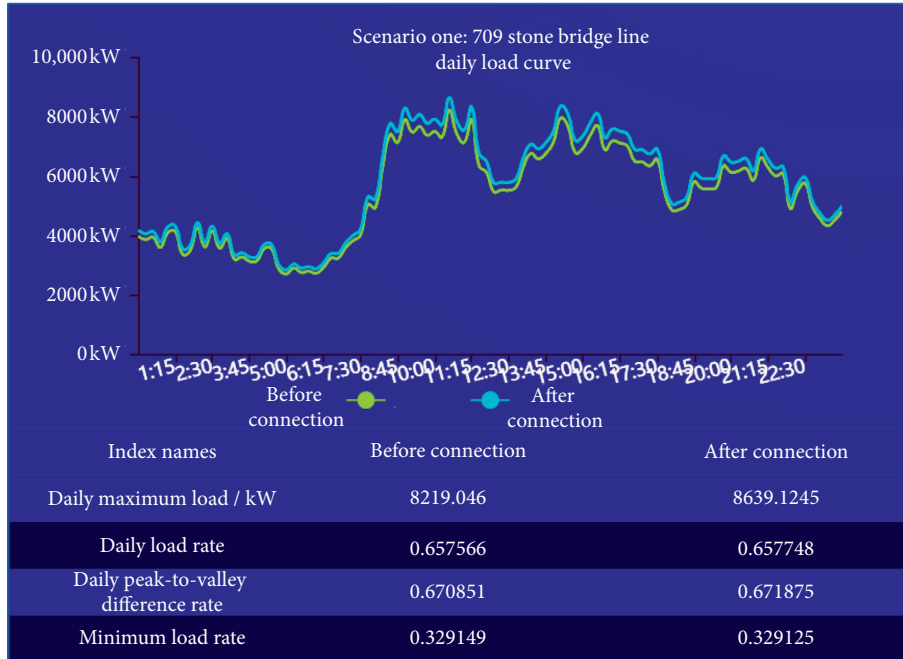


FIGURE 24: Display of the load curve No. 709 and its characteristic indicators.

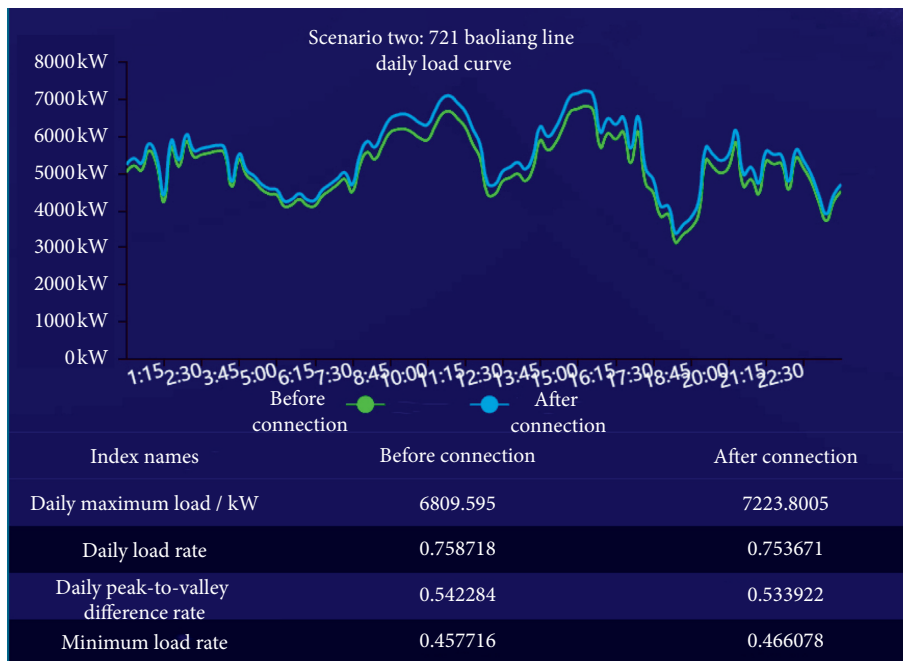


FIGURE 25: Display of the feeder load curve No. 721 and its characteristic indicators.

TABLE 4: Statistics on expected maximum load values for the year of 2020 to 2022.

Year	2020	2021	2022
Annual maximum load for new users/kW	338.4	459.35	486.22
Annual maximum load on feeder no. 721/kW	7019	7235	7604
Annual maximum load after connection/kW	7295.35	7642.25	7986.58

7. Conclusions

In this paper, a multidimensional load characteristic analysis system based on Java language with microservices and microapplications as the core architecture is built through a detailed analysis of the requirements of load characteristic software at the present stage. The microservice architecture of the system brings flexible and expandable capabilities, while the modular development mode ensures the information interaction between the functional modules. In terms of algorithm integration design, the system builds an algorithm integration platform, provides a standard unified invocation interface, and supports automatic identification and invocation of multiple languages.

Due to the interoperability of the data links of the software system described in this paper, the functional business platform is scalable and has new functional interfaces, so in the subsequent research, not only can we make use of richer data sources and conduct more in-depth theoretical research on modules that have not yet been fully developed, such as power optimization combinations, but also introduce the primary and secondary side coordination planning functions of the smart distribution network to provide interactive planning of the line network and generate components to serve the distribution network planning with the results of load characteristics and load forecast distribution. The solid foundation laid by diversified data management and control for functional expansion can further realize the functions of energy saving and environmental protection assessment, economic analysis, and comprehensive energy system planning for the distribution network.

At present, the system has completed the development of the load-related factor correlation analysis module and left the relevant interface of the load prediction module, and how to use the quantitative analysis results obtained from the load characteristic analysis and relevant factor correlation analysis module to further complete the prediction of regional load will be the focus of the next research and development work of this paper.

After the system is put into pilot operation and stabilized, it will continue to improve different application scenarios and eventually integrate the modules into a comprehensive decision-making management system integrating information control, distribution network index evaluation, transformer capacity determination, line network planning, and other functions and continuously improve the operation level and service capability of the smart distribution network.

Data Availability

The underlying data supporting the results of our study can be found from the Shenzhen Power Supply Bureau Co., Ltd.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

W. C. and T. Y. conceptualized the study; R. C. and T. Y. formulated the methodology; W. G. and D. Z. were responsible for software; validation was performed by W. C., J. L., W. G., and T. Y.; formal analysis was performed by J. L.; W. G. and J. L. were involved in investigation; R. C. and W. M. were responsible for resources; data curation was performed by W. G.; W. C., R. C., J. L., W. G., D. Z., and T. Y. prepared the original draft; W. C., R. C., J. L., W. M., and T. Y. were involved in review and editing; visualization was performed by W. G. and D. Z.; R. C. and T. Y. were involved in supervision; project administration was performed by W. C. and R. C.; and W. C., R. C., and T. Y. acquired fund. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors would like to express their gratitude to the experts for their helpful suggestions, who have enhanced this work's content. This research was funded by the Science and Technology Project of China Southern Power Grid Co., grant no. 090000KK52190072.

References

- [1] L. Tang, *Load Characteristics Analysis and Forecasting Research of Power System*, South China University of Technology, Guangzhou, China, 2010.
- [2] L. F. Cheng and T. Yu, "Typical scenario analysis of equilibrium stability of multi-group asymmetric evolutionary games in the open and ever-growing electricity market," *Proceedings of the CSEE*, vol. 38, no. 19, pp. 5687–5703, 2018.
- [3] X. Liu, Z. Xiao, R. Zhu, J. Wang, L. Liu, and M. Ma, "Edge sensing data-imaging conversion scheme of load forecasting in smart grid," *Sustainable Cities and Society*, vol. 62, Article ID 102363, 2020.
- [4] L. F. Cheng and T. Yu, "Decision-making behavior investigation for general multi-strategy evolutionary games in the spontaneous formation of long-term bidding equilibria of a power generation market," *Proceedings of the CSEE*, vol. 40, no. 21, pp. 6936–6955, 2020.
- [5] X. He, Q. Ai, and R. C. Qiu, "A big data architecture design for smart grids based on random matrix theory," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 674–686, 2017.
- [6] L. Cheng, G. Liu, H. Huang et al., "Equilibrium analysis of general N-population multi-strategy games for generation-side long-term bidding: an evolutionary game perspective," *Journal of Cleaner Production*, vol. 276, Article ID 124123, 2020.
- [7] Y. Zhang, C. Deng, R. Zhao, and S. Leto, "A novel integrated price and load forecasting method in smart grid environment based on multi-level structure," *Engineering Applications of Artificial Intelligence*, vol. 95, Article ID 103852, 2020.
- [8] L. F. Cheng, T. Yu, X. S. Zhang, and L. F. Yin, "Machine learning for energy and electric power systems: state of the art and prospects," *Automation of Electric Power Systems*, vol. 43, no. 1, pp. 15–43, 2019.
- [9] H. L. Yan and X. Z. Zhang, "Analysis on industries load characteristic of Yunnan province," *Yunnan Electric Power*, vol. 44, no. 5, pp. 107–108+114, 2016.

- [10] G. H. Li, H. Chen, Y. Y. Chen et al., "Analysis of the load rate characteristics of Hunan power grid," *Hunan Electric Power*, vol. 36, no. 5, pp. 58–60, 2016.
- [11] F. Yun, *Characteristics Analysis and Application for Power Consumer in Cities of Tangshan*, North China Electric Power University, Baoding, China, 2018.
- [12] T. Y. Kang, S. P. Yin, X. F. Wang et al., "Load characteristics of large city power grid and analysis on its influencing factors," *Electrical Measurement & Instrumentation*, vol. 53, no. 6, pp. 51–56, 2016.
- [13] L. F. Cheng, *Theoretical Investigation on the Long-Term Evolutionarily Stable Equilibrium of Multi-Population Strategic Games in Electricity Market*, South China University of Technology, Guangzhou, China, 2019.
- [14] L. Y. Wang, J. Y. Xu, and Y. Shang, "Coordinated optimal scheduling of multi-microgrid system under mechanism of time-of-use electricity price based on multi-agent improved particle swarm optimization algorithm," *Journal of Power Supply*, vol. 17, no. 4, pp. 130–139, 2019.
- [15] X. Y. Tang, H. J. Wan, H. C. Ye et al., "Analysis and prediction of typical daily load characteristics of Guizhou power grid," *Electric Power*, vol. 48, no. 9, pp. 24–30, 2015.
- [16] D. S. Xu, W. Yang, Z. Wei et al., "Short term load characteristic analysis and prediction research based on SPSS," *Power System Protection and Control*, vol. 37, no. 21, pp. 147–151, 2009.
- [17] F. Liao, C. Y. Xu, J. G. Yao et al., "Analysis of load characteristics and influencing factors in Changde area," *Power System Technology*, vol. 36, no. 7, pp. 117–125, 2012.
- [18] D. S. Luo, Q. Du, C. H. Bie et al., "Analysis of residential differentiated electricity consumption behavior based on load decomposition," *Power System Protection and Control*, vol. 44, no. 21, pp. 29–33, 2016.
- [19] J. H. Chen, M. Chen, and Y. M. Pu, "B/S system performance analysis based on microservices architecture," *Computer Systems & Applications*, vol. 29, no. 2, pp. 233–237, 2020.
- [20] X. M. Qie and J. J. Zhang, "Design of terminal service platform based on microservice architecture," *Digital Technology & Application*, vol. 38, no. 6, pp. 124–126, 2020.
- [21] S. F. Hai, S. Y. Sun, X. H. Fan et al., "Influence of induction motor parameter change considering frequency characteristics on load characteristics," *Power System Protection and Control*, vol. 47, no. 8, pp. 17–25, 2019.
- [22] Q. Z. Zhu, Z. Dong, and N. Ma, "Forecasting of short-term power based on just-in-time learning," *Power System Protection and Control*, vol. 48, no. 7, pp. 92–98, 2020.
- [23] W. J. Zhao, L. Q. Shang, and J. F. Sun, "Power quality disturbance classification based on time-frequency domain multifeature and decision tree," *Protection and Control of Modern Power Systems*, vol. 4, pp. 337–342, 2019.
- [24] B. E. Xu and H. Lan, "Selection method of typical daily load curve based on improved fuzzy clustering," *Electrical Measurement & Instrumentation*, vol. 56, no. 4, pp. 21–26, 2019.
- [25] X. D. Wang, J. Y. Liu, Y. B. Liu et al., "Shape clustering algorithm of typical load curves based on adaptive piecewise aggregate approximation," *Automation of Electric Power Systems*, vol. 43, no. 1, pp. 110–121, 2019.
- [26] Y. J. Gao, Y. J. Sun, W. H. Yang et al., "Study on load curve's classification based on nonparametric kernel density estimation and improved spectral multi-manifold clustering," *Power System Technology*, vol. 42, no. 5, pp. 1605–1612, 2018.
- [27] W. W. Tso, C. D. Demirhan, C. F. Heuberger, J. B. Powell, and E. N. Pistikopoulos, "A hierarchical clustering decomposition algorithm for optimizing renewable power systems with storage," *Applied Energy*, vol. 270, Article ID 115190, 2020.
- [28] J. A. Liu, Y. B. Liu, and M. C. Cheng, "A distributed load clustering algorithm based on quantile radius dynamic K-means," *Power System Protection and Control*, vol. 47, no. 24, pp. 15–22, 2019.
- [29] Y. X. Wu, C. Gao, H. Z. Cao et al., "Clustering analysis of daily load curves based on GWO algorithm," *Power System Protection and Control*, vol. 48, no. 6, pp. 68–76, 2020.
- [30] L. Cheng and T. Yu, "A new generation of AI: a review and perspective on machine learning technologies applied to smart energy and electric power systems," *International Journal of Energy Research*, vol. 43, no. 6, pp. 1928–1973, 2019.
- [31] O. Dumitrascu, M. Dumitrascu, and D. Dobrotă, "Performance evaluation for a sustainable supply chain management system in the automotive industry using artificial intelligence," *Processes*, vol. 8, no. 11, p. 1384, 2020.
- [32] L. F. Cheng, T. Yu, X. S. Zhang, and B. Yang, "Parallel cyber-physical-social systems based smart energy robotic dispatcher and knowledge automation: concepts, architectures and challenges," *IEEE Intelligent Systems*, vol. 34, no. 2, pp. 54–64, 2019.
- [33] X. Zhang, Z. Xu, T. Yu, B. Yang, and H. Wang, "Optimal mileage based AGC dispatch of a GenCo," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 2516–2526, 2020.
- [34] X. Zhang, T. Tan, B. Zhou, T. Yu, B. Yang, and X. Huang, "Adaptive distributed auction-based algorithm for optimal mileage based AGC dispatch with high participation of renewable energy," *International Journal of Electrical Power & Energy Systems*, vol. 124, Article ID 106371, 2021.
- [35] X. S. Zhang, Q. Li, T. Yu et al., "Consensus transfer Q-learning for decentralized generation command dispatch based on virtual generation tribe," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2152–2165, 2018.
- [36] L. Cheng and T. Yu, "Smart dispatching for energy internet with complex cyber-physical-social systems: a parallel dispatch perspective," *International Journal of Energy Research*, vol. 43, no. 8, pp. 3080–3133, 2019.
- [37] Z. Wang, H. Wang, Z. Lin et al., "Modeling of regional electrical heating load characteristics considering user response behavior difference," *International Journal of Electrical Power & Energy Systems*, vol. 123, Article ID 106297, 2020.
- [38] J. Zhao, Z. Wang, T. Yang, J. Xu, Z. Ma, and C. Wang, "Design of a novel modal space sliding mode controller for electro-hydraulic driven multi-dimensional force loading parallel mechanism," *ISA Transactions*, vol. 99, pp. 374–386, 2020.
- [39] S. Lin, X. Lin, Y. Liu et al., "Development of load modeling platform for electric power system," *Electric Power*, vol. 38, no. 11, pp. 87–91, 2005.
- [40] Y. Xie, H. Zhang, C. Li et al., "Development approach of a programmable and open software package for power system frequency response calculation," *Protection Control of Modern Power Systems*, vol. 2, no. 2, pp. 189–198, 2017.
- [41] T. Yu, L. F. Cheng, and X. S. Zhang, "The weakly-centralized Web-of-Cells based on cyber-physical-social systems integration and group machine learning: theoretical investigations and key scientific issues analysis," *Scientia Sinica Technologica*, vol. 49, no. 12, pp. 1541–1569, 2019.
- [42] X. Zhou, J. Tan, and N. Shroff, "Flexible load balancing with multi-dimensional state-space collapse: throughput and heavy-traffic delay optimality," *Performance Evaluation*, vol. 127, pp. 176–193, 2018.
- [43] Y. Ma, Z. Ling, W. Cai, Y. Cui, L. Zhou, and J. Wang, "A novel coast-down no-load characteristic test and curve conversion method for large-scale synchronous condenser," *Electric Power Systems Research*, vol. 172, pp. 77–85, 2019.