

# Electricity Demand Prediction of Smart Grid in Smart Cities Using K-means Algorithm

**Shurui Wang**

Huazhong University of Science and Technology

**Aifeng Song** (✉ [saf0217@126.com](mailto:saf0217@126.com))

North China University of Water Resources and Electric Power

**Yufeng Qian**

Hubei University of Technology

---

## Research Article

**Keywords:** K-means algorithm, smart city, smart grid, electricity demand, BP neural network

**Posted Date:** March 30th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-355422/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

The aims are to unify big data management among various departments in smart city construction, establish a centralized data collection and sharing system, and provide fast and convenient big data services for smart applications in various fields. The national grid industry is taken as the research object. A new electricity demand prediction model is proposed based on smart city big data's characteristics. This model integrates meteorological, geographic, demographic, corporate, and economic information to form an intelligent big database. The K-mean algorithm mines and analyzes the data to optimize the electricity user information. The electricity prediction model is established using the Backpropagation (BP) neural network algorithm. The electricity market is evaluated through an in-depth exploration of data relationships to verify the effectiveness of the model proposed. Results demonstrate that the K-mean algorithm can significantly improve electricity user segmentation accuracy, separate the different regional electricity consumption, and categorize different electricity users. The electricity demand network model constructed can significantly improve the prediction accuracy, and the mean error rate is 3.2671%. The model's training time improved by the additional momentum factor is significantly reduced, and the mean error rate is 2.13%. The above results can provide a theoretical and practical basis for electricity demand prediction and personalized marketing, as well as development planning for the electricity sector.

## 1. Introduction

A smart city is the effective integration of real world elements and information virtual world via new technologies and management methods, such as the Internet of Things and cloud computing, to create a visual, measurable, perceivable, interactive, and controllable intelligent production and life approaches and urban operating mechanisms. Therefore, the resource utilization can be improved, the cost and energy conservation can be strengthened, and the service efficiency can be promoted [1]. Constructing smart cities reduces the negative impact on the natural environment and supports the low-carbon economy and circular economy. The remote cloud computing center completes big data analysis, establishes reasonable and efficient control strategies, and continues to innovate based on people to provide a decision-making basis for urban management, providing various intelligent and convenient services to the public [2]. From a development content perspective, smart cities are inseparable from smart electricity consumption. With the increasingly prominent energy crisis and environmental problems, a large number of distributed energy sources and energy storage devices are connected. Moreover, the industrial society has put forward higher requirements for the reliability of power supply and power quality [3]. In China, the major advocate and construction unit of the smart grid is the State Grid Corporation of China. Under its leadership, various information technologies have been vigorously integrated, pursuing the safety performance of the power system, the stability of power supply and distribution, and improving the efficiency of energy use, thereby achieving the dual reduction of energy consumption and environmental pollution levels in the power supply and distribution process [4]. The smart grid is designed to further optimize the production, distribution, and consumption of electric energy by obtaining more customer electricity information and to highly integrate advanced sensor technology,

Internet of Things technology, information technology, communication technology, and other modern high-tech approaches to form a new type of integrated network system. Hence, information flow can be exchanged between power grid equipment and can also implement advanced functions such as real-time intelligent control of the power grid, remote debugging, online analysis and decision-making, as well as the multi-site collaborative work depending on the specific situation [5]. In smart grid-related technologies, forecasting the power demand is very important in the power industry, which has critical significance in the operation, planning, and control of the power system. Power demand forecasting is a vital foundation for ensuring the safe operation of the power system, achieving the scientific management and unified dispatching of the power grid, and formulating a reasonable development plan [6]. Electric energy consumption is affected by multiple factors. However, traditional forecasting introduces fewer factors as variables, and only considers the internal data of the power system, such as maximum load and regional average electricity consumption, failing to fully consider the impact of changes in other factors on the forecast [7]. Therefore, studying the power demand of smart grids is of great significance to the development of the national grid.

There are many studies on electricity demand prediction in the past few years. Classic and traditional electricity prediction methods have failed to adapt to the changing electricity market. Some new methods have been applied to electricity demand prediction. He et al. (2017) quantitatively analyzed the factors that affected electricity prediction and proposed a long-term electricity demand prediction model suitable for the new economic normal; compared with the traditional models, this model has high prediction accuracy [8]. Laouafi et al. (2017) proposed a new prediction combination method to generate very short-term electricity demand prediction under normal and abnormal load conditions and found that the system could achieve good prediction accuracy and avoid large prediction errors [9]. Mirjat et al. (2018) predicted that Pakistan's average electricity growth rate would be 8.35% in subsequent years using deep learning and electricity data from 2015 to 2017; this data was 19 times the existing base [10]. Khalifa et al. (2019) used electricity consumption data to model the Qatar electricity market; they found that more energy consumption would be generated around 2030 and proposed to improve electricity efficiency by reducing electricity consumption [11]. The above research shows that different scholars have adopted different algorithm models for electricity demand prediction. Compared with the traditional models, the performance has been greatly improved, and future electricity demand can be successfully predicted. However, some models are based on the demand of the electricity market. During prediction, they may fail to comprehensively consider the influencing factors of electricity demand from other aspects. Therefore, establishing a prediction model by finding the critical influencing factors from appropriate methods and massive data is crucial for electricity demand prediction.

By investigating literature about smart city and smart grid, an electricity demand prediction model is proposed according to the actual situation, which is for the new characteristics of current smart city development and the new needs of smart grids. A segmented model is created by segmenting electricity users and adopting the smart city database to analyze the differences in population, legal person, economy, and geographic information. Innovatively, a new electricity prediction algorithm is proposed,

realizing high-precision prediction via the correlation between data. The results can provide a theoretical and practical basis for constructing smart grids and research ideas for smart city construction.

The contributions are: (1) based on traditional customer segmentation methods, the differences are analyzed through the demographic, legal person, economic and geographic information provided by the smart city basic database to create a customer segmentation model. The reform of China's corporate system continues to deepen, and the power industry is now on the agenda. In this way, China's power companies are no longer the same as those in the past. The power generation, power supply, and distribution of the companies are under unified management by the national government, with a fixed source of customers and stable economic benefits. Nowadays, China's power companies have also been promoted to the market, which will face more competition. Thus, the power industry in China must respect the laws of the market like private enterprises, improve the concept of market competition, change its attitude towards customers, and strive to increase profits. (2) The characteristic relationships among the following economic indicators are extracted: the total electricity consumption and the area's population, GDP (Gross Domestic Product), GDP per capita, the number of units of industrial enterprises above designated size, and total imports and exports from the basic database of smart cities. Combined with power companies' internal data, a medium-term power demand forecasting model is established.

## **2. Methods**

### **A. Methods of electricity demand prediction**

The three most commonly used electricity prediction methods are the classic prediction method, the traditional prediction model, and the intelligent prediction model. (1) Classic prediction methods include trend extrapolation method, classified electricity demand prediction method, and load density method. These prediction methods are widely used, but most of them analyze the relationship between some simple variables and lack in-depth data analysis; therefore, the prediction accuracy is often low [12]. (2) Traditional prediction methods include the regression analysis method, which establishes the relationship between the dependent variables and known load data and predicts the electricity system's load using mathematical analysis. Applications of the time-series method include the exponential smoothing method and the Census-H Decomposition method; the random time series methods include the state space method, the Box-Jenkins method, and the Markov method [13]. According to the given data, the relationship between the variable and the dependent variable is determined, and the regression equation and various parameters are found. Based on the obtained equation, the dependent variable is obtained from the existing independent variables, and finally, the electricity prediction data are obtained. (3) When there are large random factors in historical electricity demand, the prediction effect is flawed and is greatly affected by bad data in the time series. In recent years, the electricity market has become increasingly complicated. Classic prediction methods and traditional prediction methods cannot adapt to the nonlinear, multi-variable, time-varying, and random characteristics of the electricity market. Hence, some new prediction methods are used in electricity demand prediction. The laws are extracted to

establish a knowledge base for reasoning and judgment based on real experience [14]. The detailed comparison results of the advantages and disadvantages are shown in Table 1.

TABLE I: Comparison of different electricity demand prediction methods

Recognition methods	Advantage	Disadvantage
Classical prediction method	The broadest range of applications and the most extended use time	Lack of scientific theory, low prediction accuracy
Traditional predictive model	More substantial data analysis capabilities and higher model accuracy	The algorithm runs for a long time and requires high system configuration
Intelligent prediction model	Higher prediction accuracy	Time-consuming database construction

## B. K-means clustering algorithm

**(1) Algorithm utilization:** At present, smart grid construction needs to comply with market laws and rely on its competitiveness to achieve the survival and development of enterprises. The core is to fully understand all users, improve user experience, and enhance user loyalty. For customer segmentation, the traditional segmentation method uses a single indicator and cannot effectively divide users. With the development of smart cities and the advancement of big data technology, a large amount of data can be obtained, while data mining technology can be used to extract the required indicators to segment power customers [15]. Currently, among data mining algorithms, the K-Mean clustering algorithm has attracted the attention of many scholars due to its simple implementation and high efficiency.

The K-means clustering algorithm uses the distance of the target data as an evaluation indicator to measure the similarity. When the distance between two objects is small, the similarity between these data is relatively high. This type of algorithm usually consists of a group of relatively close objects, and the final goal is to obtain a data group with a compact distance and a high degree of separation [16].

**(2) Algorithm principle:** the initial dataset is assumed as  $(x_1, x_2 \dots x_n)$ , and each data unit ring is a  $p$ -dimensional vector (the  $p$ -dimensional vector is composed of  $p$  eigenvalues). The K-means clustering algorithm's goal is to divide the original dataset into  $K$  categories  $G=\{G_1, G_2, \dots, G_k\}$  with a given number of categories  $k$  ( $k \leq n$ ). Each iteration of the K-means clustering algorithm must check whether the classification of each data unit is correct. If it is classified into the wrong category, the data must be adjusted. When the adjusted data is clustered with  $k$  points in the space as the center and the next iteration starts, the value of each cluster center is updated in turn until the cluster center does not change,

indicating that the clustering criterion function has converged, and the best clustering result is got [17]. The algorithm's workflow is shown in Figure 2.

### C. Backpropagation (BP) neural network algorithm

Artificial Neural Network (ANN) is composed of numerous neurons, in which these neurons are connected. ANN has a strong nonlinear mapping ability [18]. BP neural network is a multi-layer feedforward network trained according to the backpropagation algorithm [19]. The topological structure of the BP neural network algorithm model includes an input layer, a hidden layer, and an output layer, as shown in Figure 2.

BP neural network is usually composed of multiple layers and multiple neurons, which are mainly divided into an input layer, a hidden layer, and an output layer [20]: The input vector should be:

$$x = [x_1, x_2, x_3 \dots x_i, \dots x_m], i = 1, 2, \dots m \quad (1)$$

The output vector should be:

$$y = [y_1, y_2, y_3 \dots y_k, \dots y_n], k = 1, 2, \dots n \quad (2)$$

The neuron input of the hidden layer should be:

$$h^{(l)} = \begin{bmatrix} h^{(l)}_1, h^{(l)}_2, h^{(l)}_3 \\ \dots h^{(l)}_j, \dots h^{(l)}_{sl} \end{bmatrix}, j = 1, 2, \dots sl \quad (3)$$

In (3),  $sl$  is the number of neurons in the  $l$ -th layer. Assuming that  $w^{(l)}_{ij}$  is the connection weight between the  $j$ -th neuron in the  $l$ -th layer,  $b^{(l)}_i$  is the threshold of the  $i$ -th neuron in the  $l$ -th layer, and  $net^{(l)}_i$  is the input of the  $i$ -th neuron in the  $l$ -th layer, the following equation can be obtained:

$$h^{(l)}_i = f(net^{(l)}_i) \quad (4)$$

$$net^{(l)}_i = \sum_{j=1}^{sl-1} w^{(l)}_{ij} h^{(l-1)}_j + b^{(l)}_i \quad (5)$$

### D. Electricity user segmentation model

**(1) Overall framework:** based on the above theory, the primary database of smart cities is utilized to establish a functional structure model for electricity user segmentation (Figure 3). First, a data warehouse is established. Then, relevant user segmentation data are extracted for data analysis. Afterward, data are cleaned and converted. The association analysis method is adopted for data mining, and finally, the mining results are analyzed. The primary smart city database is the foundation of the entire model. Pre-processing of data is the guarantee for real and effective mining results. The effectiveness of user segmentation depends largely on selecting user consideration standards and establishing measurements. The adopted mining method based on actual needs is the key to the entire model.

**(2) User segmentation:** currently, the research and the adopted electricity data analysis technology use traditional data mining and statistical methods. The smart city public primary database data are added to the electricity big data analysis for improving the algorithm's accuracy and usability. Besides, the big data method is utilized for prediction, and the used data types are as many as possible. The prediction task that has been impossible previously can be completed by exploring the data association relationship, which can ensure high precision simultaneously. The electricity value of the population analyzes users' electricity value from a personal perspective and obtains the gathering area of high-potential users through personal information and social insurance information.

**(3) Data processing:** for confidentiality concerns, the Dongfangtong TI-ETL tools, and data desensitization technology are utilized to transform sensitive or confidential information through desensitization rules, thereby protecting the sensitive and private data. For some missing data, the citizen's name, gender, address, and other elements are extracted from other information. Various social security databases are integrated, and the collected information is used to roughly restore the demographic information and provide the basis for electricity user segmentation.

**(4) Algorithm realization:** it is divided into population electricity value information, enterprise commercial value distribution, and macroeconomic information. For human electricity value information, as shown in Figure 4, the electricity users are segmented, and the integrated resident data are arranged in a table from high to low in the order of individual units according to residents' social insurance information, social security information, corporate information, and the potential and influence of electricity use. Each administrative district is taken as a unit. As for the data of the corporate legal person, the legal person's registered capital is used as the analysis target, and the social insurance information is based on the insurance amount, the social security information based on the subsidy amount, and the housing provident fund based on the monthly payment amount [21].

The distribution of enterprise value is shown in Figure 5. The organization code is used as the search basis to match each commission, office, and bureau's source data. The evaluation and ranking are based on four dimensions: business category, registered capital, annual turnover, and the number of employees. Each administrative district is taken as a unit. For corporate legal person's data, K-means cluster analysis is performed based on turnover, registered capital, and the number of employees. For business categories, the conversion weight ratio of the construction industry to turnover is set to 10%, that of the manufacturing is 100%, the wholesale retail industry is 30%, and the service industry is 15% [22].

In terms of the macro-economic information, the macro value evaluation aims to evaluate the administrative districts' electricity consumption potential for the segmentation of electricity users. The data of some districts can be accurate to the administrative streets. According to previous studies, several significant data categories are selected, such as regional Gross Domestic Product (GDP), per capita disposable income, per capita GDP, total asset investment, and trade data [23].

Five districts A, B, C, D, and E in a city are chosen for empirical research. In District A, four companies are chosen from the industrial area A11-A14; in District B, four companies are chosen from the office building

area B11-B14; in District C, four neighborhoods are selected from the residential areas C11-C14; District D and E are comparison controls.

## E. Model of electricity demand prediction

**(1) Impact indicators:** electric energy consumption is affected by multiple factors. However, traditional prediction introduces fewer factors as variables. These methods only consider the internal data of electricity companies and fail to fully consider the impact of changes in other factors on prediction. Therefore, the prediction accuracy of electricity information is very limited [24]. The external information of electricity companies provided by the primary database of smart cities is fully utilized. Given the impact of changes in various relevant objects on the prediction value, the prediction of regional annual electricity consumption is researched involving multiple influencing factors.

In addition to the electricity companies' factors, the total electricity consumption is also affected by many factors such as population and corporate trends, economic conditions, energy policies, and electricity price adjustments. In particular, the development of social and economic operations needs to consume a large amount of electric energy, and there is a correlation between electricity consumption and economic indicators [25]. There are many statistical dimensions of economic data. The characteristic relationship of a district's social product, per capita GDP, price index, total import and export, and other economic indicators are combined with the internal data of electricity companies to establish a mathematical model for electricity consumption prediction [26]. The specific names of the data indicators are as follows: (1) macroeconomic data; (2) policies and other external data; (3) regional electricity consumption data in past years; (4) electricity user segmentation data.

**(2) Normalization processing:** ANN usually normalizes the data before training. The training effects of different transfer functions are different to avoid neuron oversaturation [27]. The input data value must be within [0,1], which is the characteristic requirement of the transfer activation function. Therefore, the original data of the network must be processed. The original data are normalized, and the equation is as follows:

$$x_{i0} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (6)$$

In (6),  $x_{i0}$  is the  $i$ -th feature parameter after normalization,  $x_i$  is the original  $i$ -th feature parameter,  $x_{\min}$  is the minimum value of the  $i$ -th feature parameter, and  $x_{\max}$  is the maximum value of the  $i$ -th feature parameter.

**(3) Parameter determination:** after preliminary experiments, a three-layer network model structure with a hidden layer is determined. The number of neurons is 18, and the logsig transformation function is used. The number of neurons in the second layer is the same as the number of output variable vectors, and the output layer uses a pure linear transformation function. The input feature parameters are 65; that is, the number of input layer nodes is 65, and the number of output layer nodes is 5. Generally, increasing the

number of nodes in the hidden layer can reduce the network's training error more than increasing the number of hidden layers. The BP neural network algorithm can be set as a three-layer structure to map the  $n$ -dimensional input layer to the  $m$ -dimensional output layer. Therefore, the number of hidden layers in the network is determined as 1. When applying a neural network for electricity prediction, the reference equations for selecting the number of hidden layer neurons are as follows:

$$h = 2m + 1 \quad (7)$$

$$h = \sqrt{n+m} + \alpha^* \quad (8)$$

In (7) and (8),  $h$  is the number of nodes in the hidden layer,  $m$  is the number of nodes in the input layer, and  $n$  is the number of nodes in the output layer. After a comprehensive comparison of experiments, the number of nodes in the hidden layer is selected as 18. The S-tangent tansing is selected as the activation function for hidden layer neurons, and the activation function for output layer neurons is the S-type logarithmic logsig function.

**(4) Data source:** The streaming data in the power grid come from the collection of smart meters, PMUs, and various sensors. These data are large in scale, diverse in structure, and fast. To accurately obtain the electricity consumption data of different electric equipment of users, the electric power company has installed a large number of smart meters, which will send real-time electricity consumption information to the grid every 5 minutes. The real-time collection of streaming data requires the characteristics of fast collection speed, high reliability, real-time monitoring of data changes, and simple data processing. Therefore, the collection system is a distributed, reliable, and highly available system of massive log aggregation, which can monitor and receive data from the client and send it out. When a node fails, the log file is transferred to other nodes without loss, ensuring data integrity.

## F. System improvement and verification

**(1) System improvement:** the influence of changing trends on the error surface. The BP neural network may fall into a local minimum, which can be prevented by the additional momentum. The adjustment equations for weight and threshold with additional momentum factor are:

$$\Delta w_{ij}(k+1) = (1 - mc)\eta\sigma_i p_j + mc\Delta w_{ij}(k) \quad (9)$$

$$\Delta b_{ij}(k+1) = (1 - mc)\eta\sigma_i p_j + mc\Delta b_{ij}(k) \quad (10)$$

In (9) and (10),  $w$  is the weight vector,  $k$  is the number of training,  $mc$  is the momentum factor,  $\eta$  is the learning rate,  $\Delta b_{ij}$  is the gradient of the error function, and  $\sigma_i$  and  $p_j$  are the correlation coefficients.

**(2) Clustering algorithm evaluation:** Here, the K-means clustering algorithm uses the square-error and criterion function to evaluate the clustering performance.  $X$  represents the given dataset, and each data unit is a  $p$ -dimensional feature vector. It is set to  $K$  categories. The algorithm randomly selects  $k$  data as the starting cluster center analyzes the distance from each data unit to the cluster and divides the data

into the array sink where the corresponding cluster center is located. It is supposed that  $X$  contains  $k$  data groups  $X_1, X_2 \dots X_k$  the amount of data units in each data group is  $m_1, m_2, \dots, m_k$ ; the cluster centers of each data group are  $n_1, n_2, \dots, n_k$ . The used square error equation [28] is:

$$E = \sum \sum \|p - m_i\| \quad (11)$$

**(3) BP neural network training:** the training times of the neural network is 10,000, the expected error of training is 0.02, and the learning rate is 0.01. The pre-processed data matrix is imported into MATLAB and normalized. The newff function is employed to establish a BP neural network model. After the calculation, the data go through the denormalization processing, and the MATLAB toolbox performs calculations to obtain the relative error percentage and the predicted electricity demand value.

### 3. Results

#### A. Results of electricity user segmentation

Different electricity consumption areas can be divided through the above user segmentation model, as shown in Table 2 and Figure 1. The four companies in District A are in the high electricity consumption area in their administrative district. The medium electricity consumption area is the office area of District B. The area with low electricity consumption is the residential electricity area of District C. This is consistent with the actual result. Then, all areas are divided into residential electricity and commercial electricity areas. According to the model, the number of all residential users and commercial users in the city can be subdivided into four levels from P1-P4 from more to less electricity consumption. The number of residents with the highest electricity consumption reached 25,678. The number of commercial users with the highest electricity consumption reached 298. Such results also confirm the effectiveness of the segmentation model proposed.

TABLE II: Regional electricity potential distribution

Area category	High electricity consumption area	Middle electricity area	Low electricity area
Bend over	A11	B11	C11
	A12	B12	C12
<i>Squat down</i>	A13	B13	C13
	A14	B14	C14

#### B. Performance analysis of electricity demand prediction

Figure 7 shows the results of electricity demand prediction calculated by the BP neural network model. The electricity demand prediction results in different areas are not much different from the actual value, and the largest relative error is 5.129%. The minimum value is 2.294%, and the average relative error is 3.2671%. Hence, the BP neural network has a good prediction effect. As shown in Figure 8, the model's training error is analyzed, and the results suggest that when the training time of this model is 5,000s, the model performance tends to be stable. After comparing the literature, the training speed is very long, so that it cannot satisfy the actual demand.

### C. Performance analysis results of the improved algorithm

After optimization, as shown in Figures 9 and 10, after optimizing the algorithm, the improved BP algorithm has a smaller average relative error, with an average relative error of 2.13%, and a faster training speed compared with the unimproved BP neural network. Besides, the improved algorithm has reached a stable state in the 2000s, and its training accuracy is higher, which is more advantageous in predicting electricity demand. Hence, the improved algorithm meets the basic requirements of the electricity company for electricity prediction and has particular practical application value.

### E. Performance comparison of different models

As shown in Figure 11, the proposed model is compared with the latest models. The accuracy of the BP neural network used in the literature [29] is relatively high and remains at 79.63%. Compared with methods in other literature, the proposed model has the best performance. The accuracy rate of the model is 85.25%, and the average value is also the largest, which is 83.72%. Hence, the proposed model is significantly better than the latest models in terms of algorithm performance.

TABLE III: Performance comparison results of different models

Models	Accuracy (%)	Recall (%)	F1-Measure	Mean value
Literature [29]	0.7963	0.6625	0.8253	0.761366667
Literature [30]	0.8142	0.6768	0.8649	0.7853
Literature [31]	0.7125	0.6972	0.8165	0.742066667
Literature [32]	0.8073	0.7323	0.8246	0.788066667
The proposed model	0.8525	0.7848	0.8745	0.837266667

## 4. Discussion And Analysis

The comprehensive application of big data in the smart city system can provide trend judgments and information sharing, promoting the innovative management and interconnection of cities and the healthy competition of various industries and sustainable development of the entire society. Besides, with the increasingly prominent energy crisis and environmental problems, a large number of distributed energy

sources and energy storage devices are connected, and the industrial society has put forward higher requirements for power reliability and power quality. The load forecast of the power system is of great significance to the dispatch of the power system, which is the fundamental basis for formulating power generation and transmission plans and an influential aspect of the modern development of the power market. Improving the accuracy of power system load forecasting can effectively improve the economic benefits of the power sector and promote the safe and economic operation of the power grid. Here, the demand forecasting problem of the power system is studied by consulting a large number of relevant domestic and foreign documents.

The characteristics and significance of power system demand forecasting are introduced in the context of the smart city big data. Traditional power data analysis methods are mostly based on limited sample data. New algorithms are proposed, or existing algorithms are improved based on proprietary data in this industry and field to achieve higher prediction accuracy and faster processing speed. The big data method for forecasting requires as many types of data as possible. The previously impossible forecasting task is completed by exploring the data association relationship, thereby ensuring high accuracy. During power information forecasting, the focus of work is shifted from the research of complex algorithms to the preparation of big data acquisition. On this basis, the existing methods of power demand forecasting are compared and their respective characteristics are analyzed, which is also reported in the relevant literature [33]. After data mining, the K-means clustering analysis algorithm is used to establish a power customer segmentation model for the in-depth processing of the basic data of smart cities. Through cleaning, conversion, and establishment of new data warehouses, its intrinsic data characteristics are pointed out, and the intrinsic value of data is tapped [34]. Compared with the traditional customer segmentation methods in the past, depending on the basic data of smart cities can describe customer behavior more comprehensively and accurately, which provides a reference for power companies to formulate appropriate marketing strategies and other in-depth research.

## 5. Conclusion

The establishment and mining of big data are utilized to establish a smart grid user segmentation model through the K-means algorithm for electricity demand prediction. Through cleaning, conversion, and establishing new data warehouses, the intrinsic value of data is deeply explored. Compared with the traditional user segmentation method, the proposed model can describe user behavior more comprehensively and accurately relying on the primary data of smart cities. Also, the results of electricity user segmentation are incorporated into the input samples. The BP neural network method is utilized for electricity demand prediction on the electricity system to establish a BP neural network model. The input data are normalized, proving the validity of the model. Although a suitable electricity demand prediction model has been established, several shortcomings are found. First, there is no mature multi-indicator prediction model for the relationship between time and social factors and electricity demand. The electricity demand prediction model constructed is still in the research phase based on actual cases, lacking universal applicability. Second, modeling the impact of annual weather conditions on electricity load is demanding. In electricity load prediction, the analysis of electricity consumption is accurate based

on weather changes. However, annual weather changes are very macroscopic and challenging to measure in local areas accurately. Hence, generating accurate digital reports on economic activities, people's lives, trade, and transportation in particular areas is challenging. Therefore, it is difficult to be reflected in the electricity prediction model. In the future, these aspects will be explored profoundly to improve the proposed electricity demand prediction model continuously.

## Declarations

### Compliance with Ethical Standards

**Conflict of Interest:** All Authors declare that they have no conflict of interest.

**Ethical approval:** This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent was obtained from all individual participants included in the study.

## References

- [1] M. Marjani *et al.*, "Big IoT data analytics: architecture, opportunities, and open research challenges," IEEE Access. 2017;5:5247-5261.
- [2] H. Fu *et al.*, "Intelligent decision-making of online shopping behavior based on internet of things," International Journal of Information Management. 2020;50:515-525.
- [3] K. Zhang *et al.*, "Security and privacy in smart city applications: Challenges and solutions," IEEE Communications Magazine. 2017;55(1):122-129.
- [4] L. Anthopoulos, "Smart utopia VS smart reality: Learning by experience from 10 smart city cases," Cities. 2017;63:128-148.
- [5] L. Wang *et al.*, "Frame Construction and Task Analysis of Smart Energy in the Construction of Smart City," Energy and Power Engineering. 2020;12(4):25-36.
- [6] E. Mengelkamp *et al.*, "A blockchain-based smart grid: towards sustainable local energy markets," Computer Science-Research and Development. 2018;33(1-2):207-214.
- [7] T. Ahmad, and H. Chen, "Potential of three variant machine-learning models for forecasting district level medium-term and long-term energy demand in smart grid environment," Energy. 2018;160:1008-1020.
- [8] Y. He *et al.*, "Urban long term electricity demand forecast method based on system dynamics of the new economic normal: the case of Tianjin," Energy. 2017;133:9-22.

- [9] A. Laouafi *et al.*, "Online electricity demand forecasting based on an effective forecast combination methodology," *Electric Power Systems Research*. 2017;148:35-47.
- [10] N. H. Mirjat *et al.*, "Long-term electricity demand forecast and supply side scenarios for Pakistan (2015–2050): A LEAP model application for policy analysis," *Energy*. 2018;165:512-526.
- [11] A. Khalifa, M. Caporin, and T. Di Fonzo, "Scenario-based forecast for the electricity demand in Qatar and the role of energy efficiency improvements," *Energy Policy*. 2019;127:155-164.
- [12] M. Kim *et al.*, "A hybrid neural network model for power demand forecasting," *Energies*. 2019;12(5):931.
- [13] I. Ghalekhondabi *et al.*, "An overview of energy demand forecasting methods published in 2005–2015," *Energy Systems*. 2017;8(2):411-447.
- [14] K. G. Boroojeni *et al.*, "A novel multi-time-scale modeling for electric power demand forecasting: From short-term to the medium-term horizon," *Electric Power Systems Research*. 2017;142:58-73.
- [15] M. Azaza, and F. Wallin, "Smart meter data clustering using consumption indicators: responsibility factor and consumption variability," *Energy Procedia*. 2017;142:2236-2242.
- [16] L. Bai *et al.*, "Fast density clustering strategies based on the k-means algorithm," *Pattern Recognition*. 2017;71:375-386.
- [17] G. Zhang, C. Zhang, and H. Zhang, "Improved K-means algorithm based on density Canopy," *Knowledge-based systems*. 2018;145:289-297.
- [18] J. C. Whittington, and R. Bogacz, "Theories of error back-propagation in the brain," *Trends in cognitive sciences*. 2019;23(3):235-250.
- [19] E. O. Neftci *et al.*, "Event-driven random back-propagation: Enabling neuromorphic deep learning machines," *Frontiers in neuroscience*. 2017;11:324-331.
- [20] B. Scellier, and Y. Bengio, "Equilibrium propagation: Bridging the gap between energy-based models and backpropagation," *Frontiers in computational neuroscience*. 2017;11:24-31.
- [21] H.-W. Kim, and Y.-S. Jeong, "Secure authentication-management human-centric scheme for trusting personal resource information on mobile cloud computing with blockchain," *Human-centric Computing and Information Sciences*. 2018;8(1):11-23.
- [22] F. Li *et al.*, "A clustering network-based approach to service composition in cloud manufacturing," *International Journal of Computer Integrated Manufacturing*. 2017;30(12):1331-1342.
- [23] M. Della Peruta, "Adoption of mobile money and financial inclusion: a macroeconomic approach through cluster analysis," *Economics of Innovation and New Technology*. 2018;27(2):154-173.

- [24] T. Olsson, L. F. Barcellos, and L. Alfredsson, "Interactions between genetic, lifestyle and environmental risk factors for multiple sclerosis," *Nature Reviews Neurology*. 2017;13(1):25-33.
- [25] V. Gupta, and S. Pal, "An overview of different types of load forecasting methods and the factors affecting the load forecasting," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. 2017;5(4):729-733.
- [26] F. Amara *et al.*, "Household electricity demand forecasting using adaptive conditional density estimation," *Energy and Buildings*. 2017;156:271-280.
- [27] P. Pořízka *et al.*, "Impact of laser-induced breakdown spectroscopy data normalization on multivariate classification accuracy," *Journal of Analytical Atomic Spectrometry*. 2017;32(2):277-288.
- [28] S. C. Satapathy *et al.*, "Multi-level image thresholding using Otsu and chaotic bat algorithm," *Neural Computing and Applications*. 2018;29(12):1285-1307.
- [29] Wang D *et al.* Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Applied Energy*, 2017, 190: 390-407.
- [30] Casteleiro Roca J L *et al.* Short-term energy demand forecast in hotels using hybrid intelligent modeling. *Sensors*, 2019, 19(11): 2485-2496.
- [31] Dinesh C *et al.* Residential power forecasting using load identification and graph spectral clustering. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2019, 66(11): 1900-1904.
- [32] Johansson C *et al.* Operational demand forecasting in district heating systems using ensembles of online machine learning algorithms. *Energy Procedia*, 2017, 116: 208-216.
- [33] Al Ogaili A S *et al.* Review on scheduling, clustering, and forecasting strategies for controlling electric vehicle charging: challenges and recommendations. *Ieee Access*, 2019, 7: 128353-128371.
- [34] Kazemzadeh M R *et al.* A hybrid data mining driven algorithm for long term electric peak load and energy demand forecasting. *Energy*, 2020, 204: 117948-117956.

## Figures

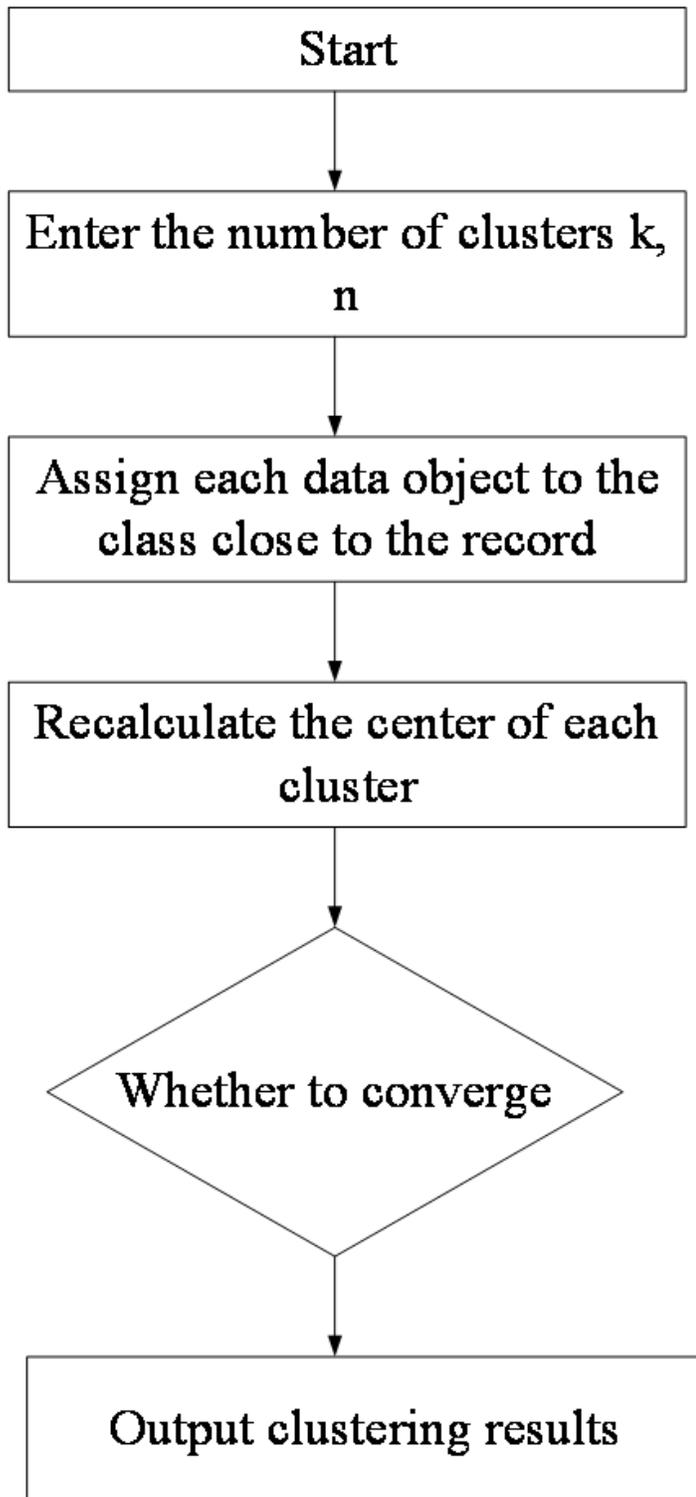


Figure 1

The workflow of the K-means algorithm

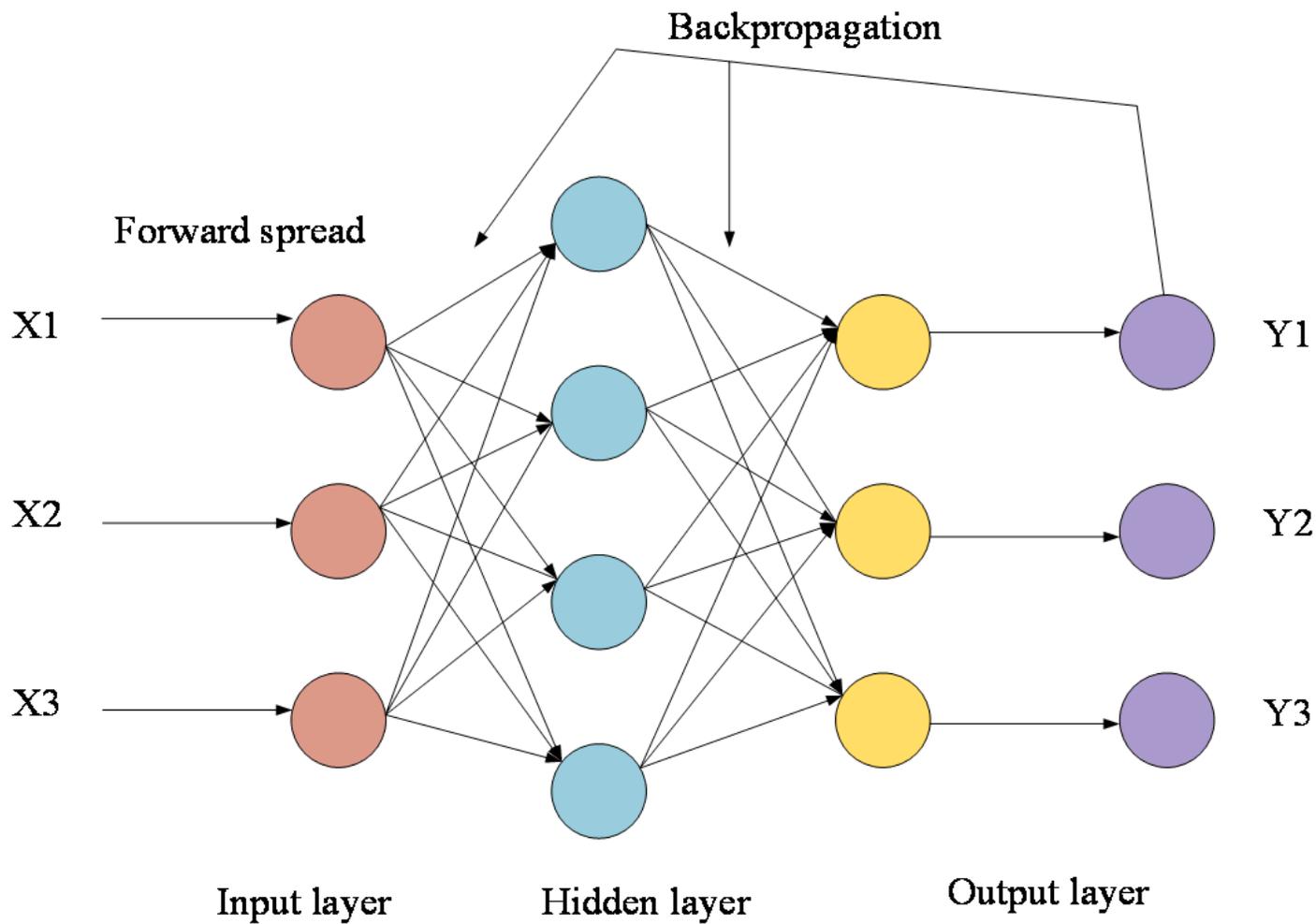


Figure 2

Structure of BP neural network

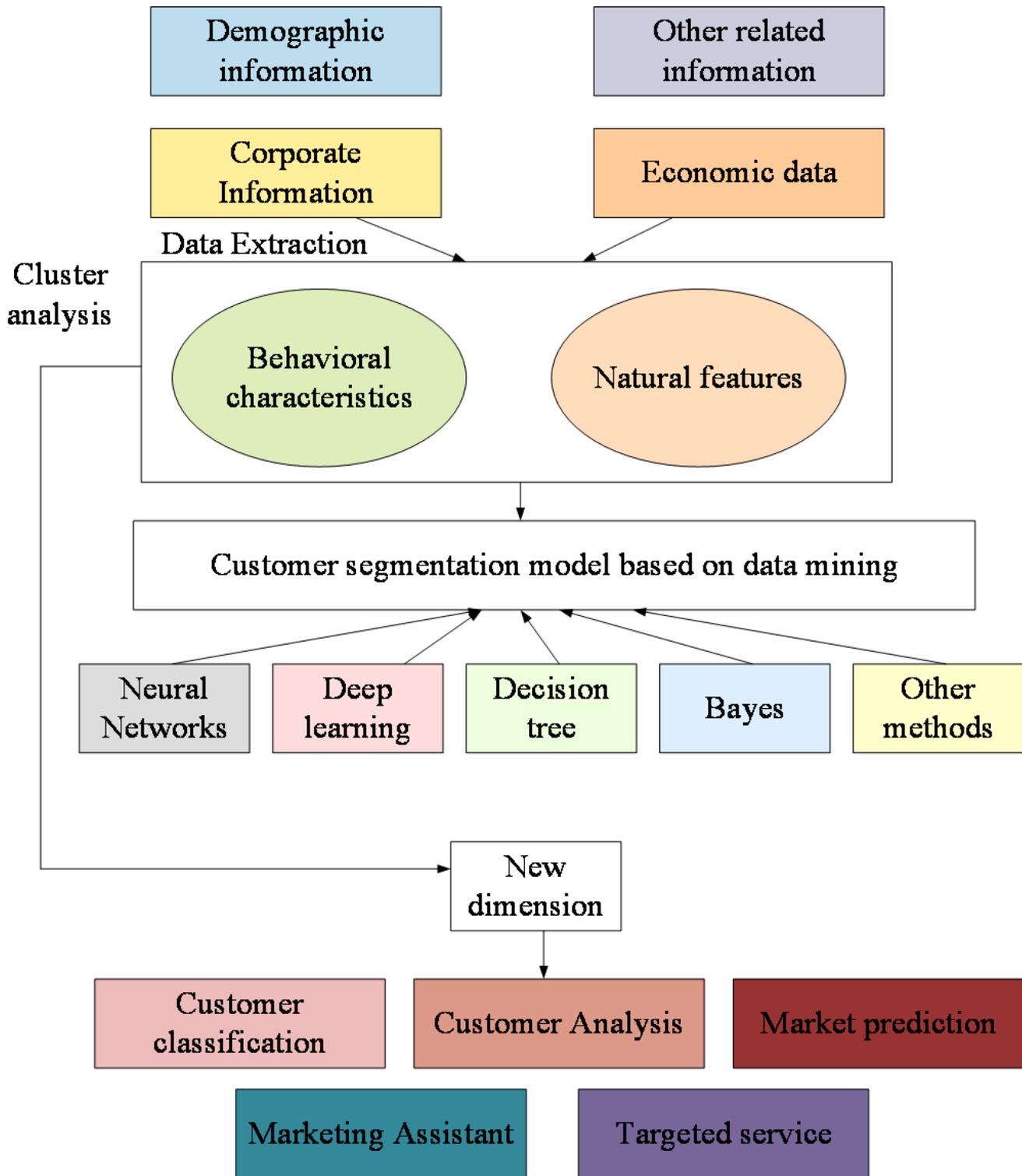
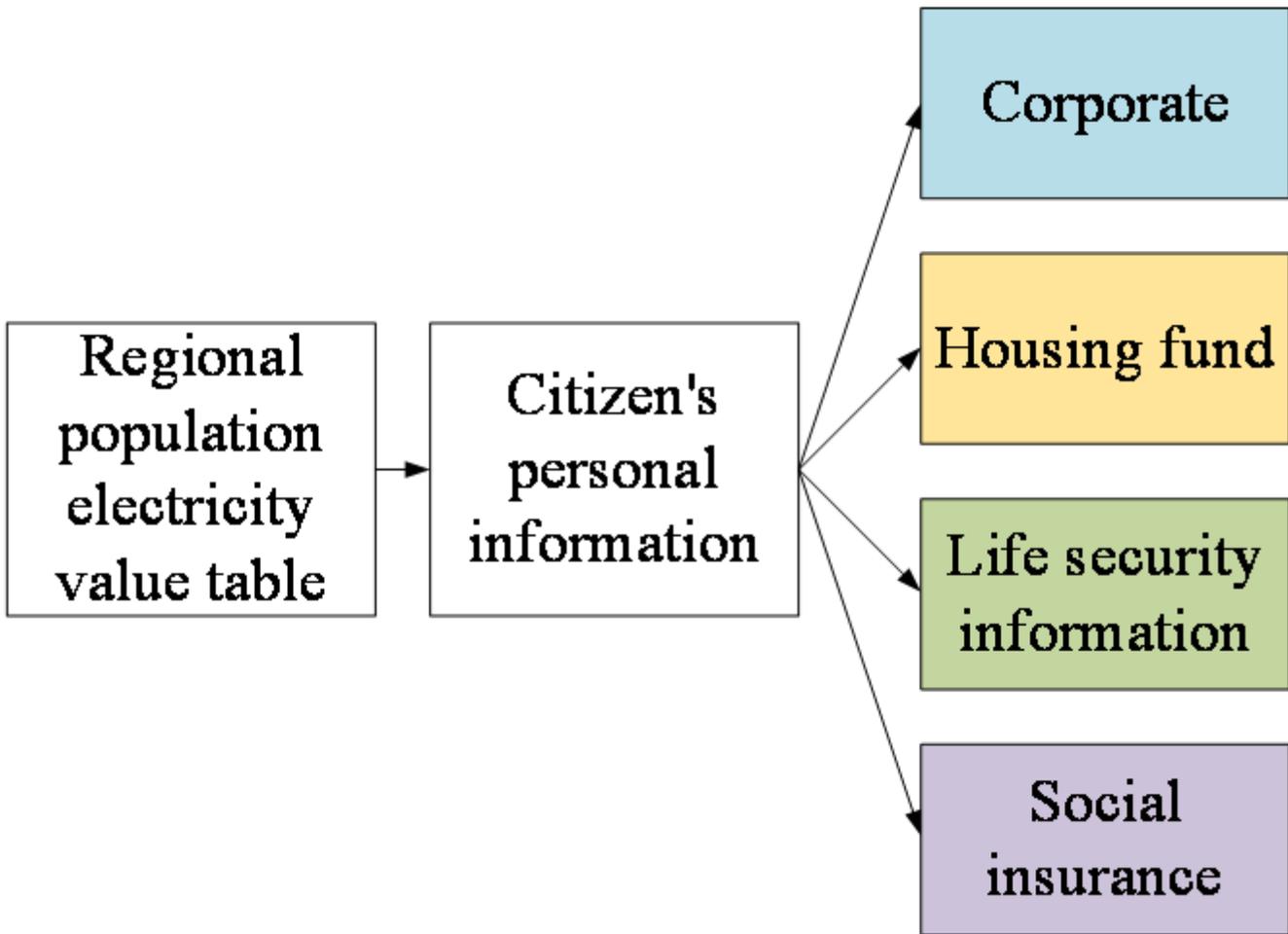


Figure 3

Functional structure of electricity user segmentation



**Figure 4**

Structure of population electricity value realization

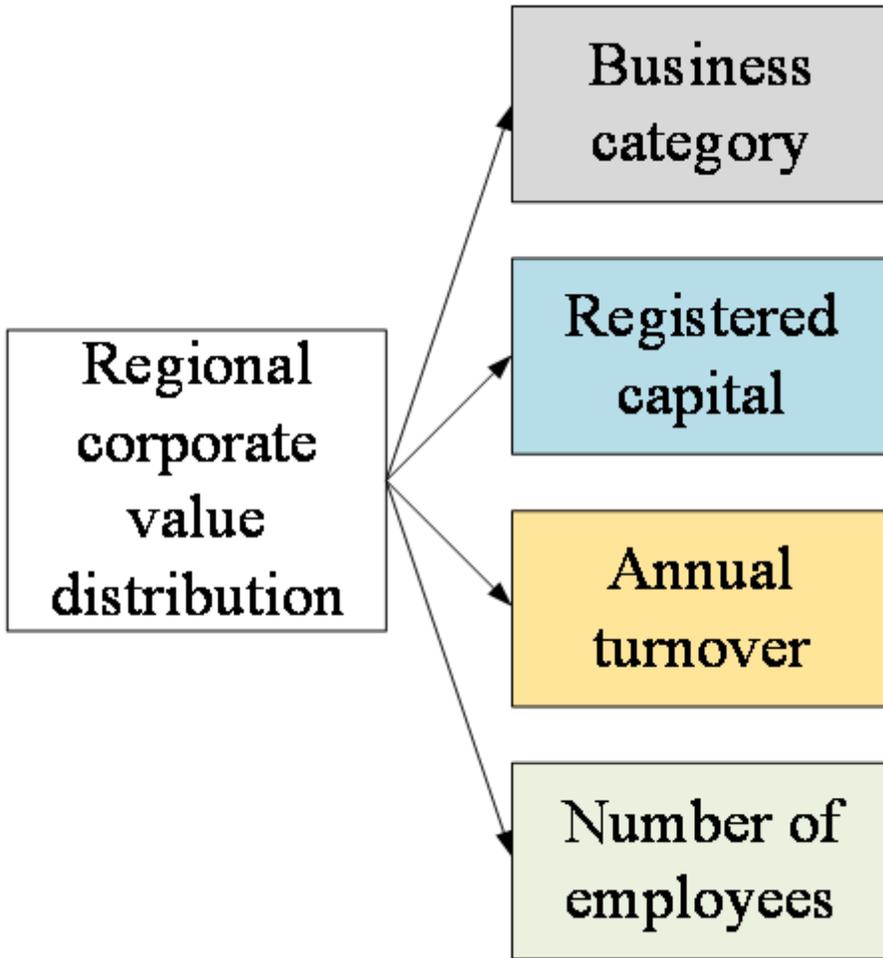
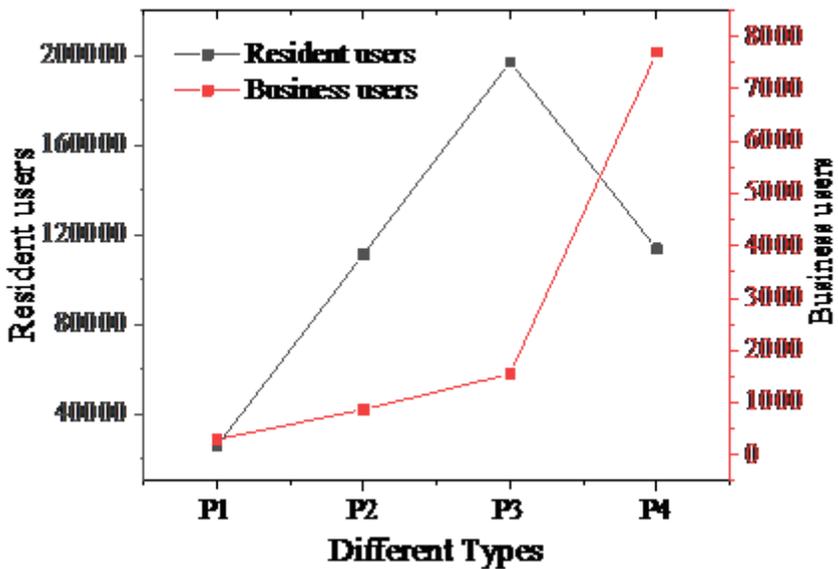


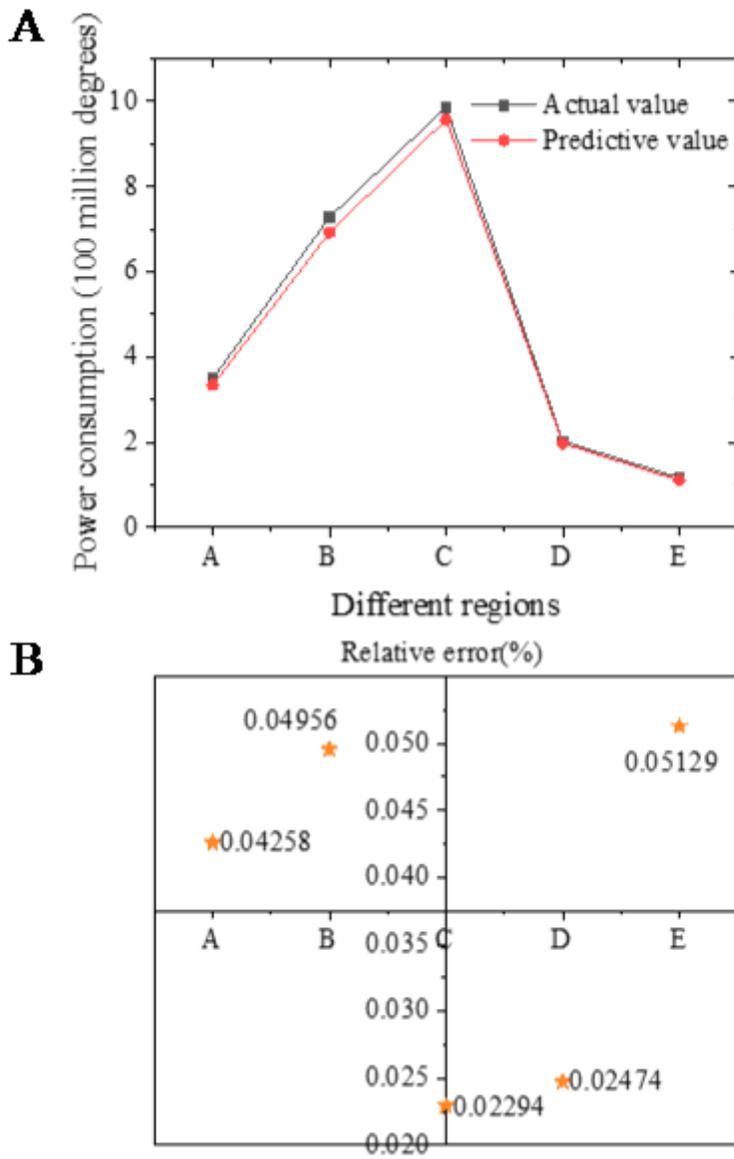
Figure 5

Structure of enterprise commercial value realization



**Figure 6**

Performance analysis of the model based on BP neural network



**Figure 7**

Electricity demand prediction results based on BP neural network

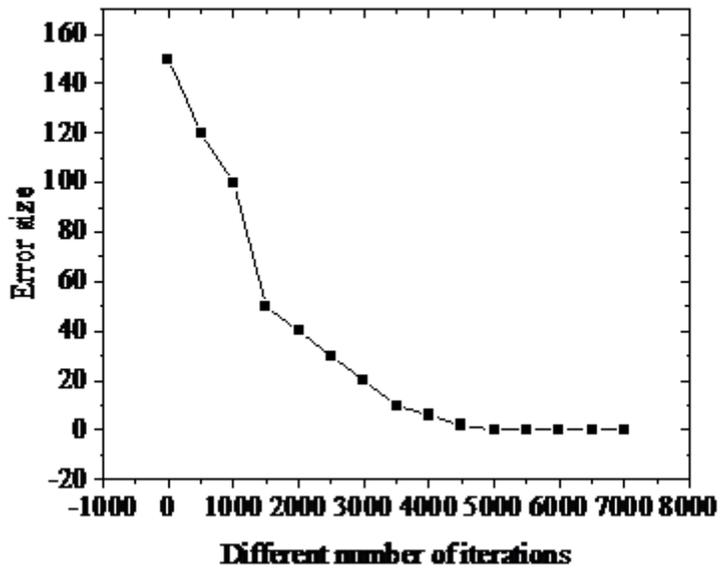
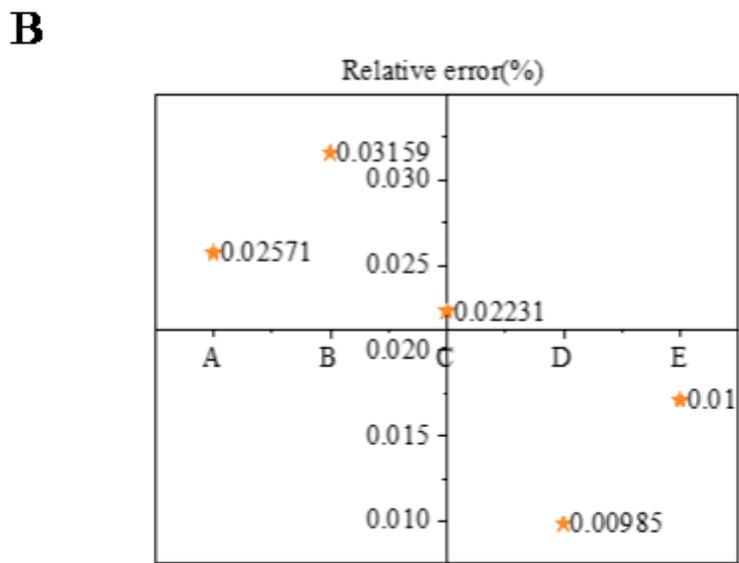
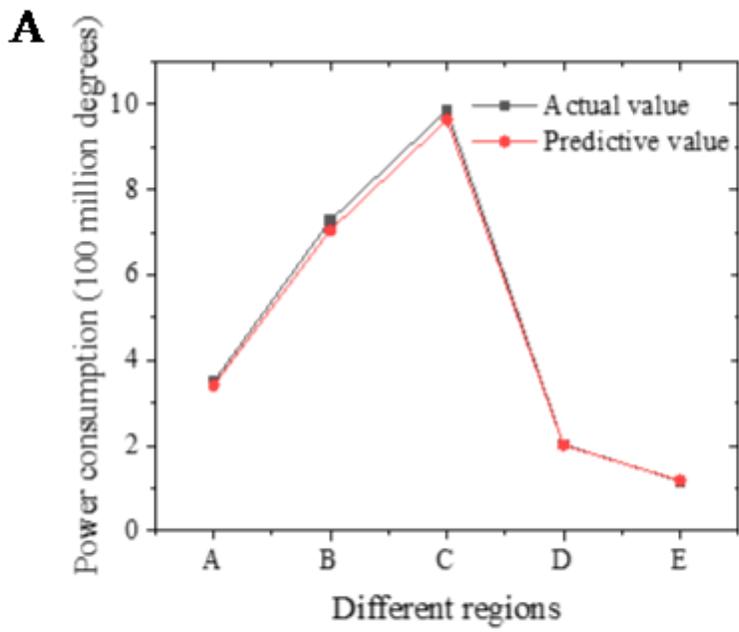


Figure 8

Performance analysis of the model based on BP neural network



**Figure 9**

Electricity demand prediction results based on improved BP neural network

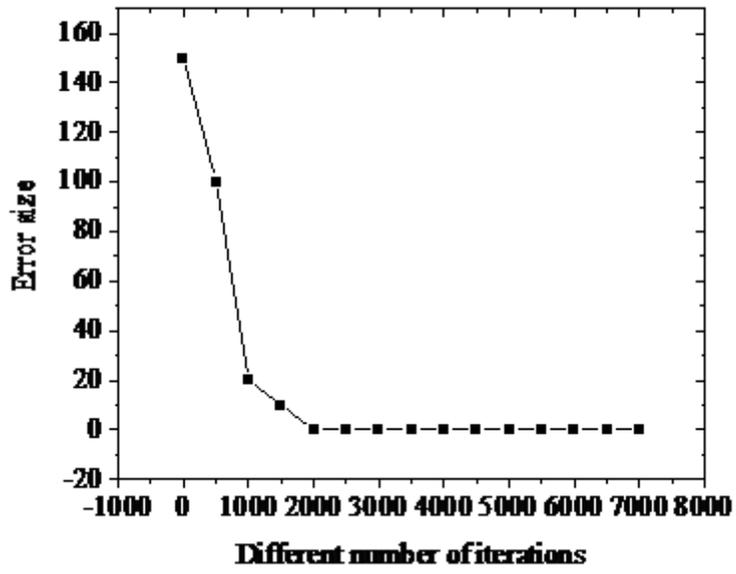


Figure 10

Model performance analysis based on improved BP neural network