International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 8018-8030

https://doi.org/10.70517/ijhsa464687

Power User Behavior-Oriented Demand Forecasting Algorithm and Service Optimization Model

Ting Qian[1,*], Rubing Xie[1] and Qingshan Xu[1]
[1] Fuzhou Power Supply Branch Power Supply Service Center, State Grid Jiangxi Electric Power Co., Ltd., Fuzhou, Jiangxi, 344099, China
Corresponding authors: (e-mail: qianting799@163.com).

Abstract This paper takes the creation of user profiles, prediction of electricity demand, construction of an electricity service optimization model, and satisfaction of electricity user needs as its research approach. Using electricity big data technology, it obtains residential electricity consumption behavior data from aspects such as basic electricity consumption, equipment electricity consumption, advanced electricity consumption, and abnormal electricity consumption. Through quantitative analysis of the obtained user electricity consumption behavior data, it generates user behavior feature tags from aspects such as basic and behavioral characteristics. By combining the generated user behavior tags with the characteristics of changes in electricity consumption behavior data, the core content of user electricity consumption behavior profiles is derived, thereby achieving precise user profiling for residential users. Additionally, based on existing research, short-term and medium-to-long-term influencing factors are screened out, and the Attention-Bi-LSTM model is used for electricity demand forecasting. Y State Grid Power Marketing Unit was selected as the experimental subject, and the power user behavior characteristics were calculated and analyzed. The proposed model was used to predict power user demand. The proposed prediction model not only fits the original data curve well but also maintains the prediction error within the range of [-5000, 6000], demonstrating high-precision prediction performance.

Index Terms Attention-Bi-LSTM model, user profile, power user demand prediction, behavioral characteristics

I. Introduction

With the advancement of industrialization and urbanization, and the increasing reliance on electrical energy, accurate forecasting of electricity demand has become increasingly important [1], [2]. Currently, the primary methods and technologies for forecasting electricity demand include statistical methods and machine learning [3]. Statistical methods analyze historical electricity demand data from users and utilize statistical models such as time series models and regression analysis to predict future electricity demand [4]-[6]. These models can capture trends, seasonality, and periodicity in historical data and apply them to future predictions [7], [8]. Machine learning-based electricity demand forecasting is an emerging technology that analyzes large amounts of historical user data to uncover correlations between variables, enabling accurate electricity demand predictions [9]-[11]. Common machine learning algorithms, including linear regression, decision trees, support vector machines, and random forests, can be selected based on specific circumstances for modeling and training [12]-[14].

Accurate electricity demand forecasting helps balance electricity supply and demand, avoiding power shortages or resource waste caused by supply-demand imbalances [15], [16]. Additionally, by precisely predicting electricity demand, power companies can optimize power resource allocation, improve operational efficiency, and provide better services [17], [18]. Of course, excellent service cannot be achieved through a single approach; it requires power companies to address multiple aspects, including improving power supply quality, introducing intelligent power services, providing 24/7 service, enhancing interaction with users, and strengthening information security measures [19]-[21]. Only by implementing comprehensive measures can high-quality power services be achieved, user needs be met, and the sustainable development of the power industry be promoted [22], [23].

Literature [24] aims to improve the accuracy and flexibility of power demand prediction in smart grids using machine learning algorithms, verifying the effectiveness of machine learning for energy demand prediction. Its greatest advantage lies in improving operational efficiency through more intelligent energy transmission scheduling. Literature [25] evaluates the performance of optimization algorithms such as genetic algorithms (GA) and firefly algorithms (FF), particularly in electricity demand forecasting, and proposes a hybrid intelligent electricity demand forecasting algorithm that integrates wavelet transform (WT) and fuzzy ARTMAP (FA) networks, optimized using the FF algorithm. Literature [26] conducted an investigation into practical methods for predicting future electricity load demand based on the Preferred Reporting Items for System Reviews and Meta-Analyses (PRISMA) guidelines, finding the superiority of a hybrid approach combining artificial neural networks with meta-heuristic techniques and proposing improvement suggestions. Literature [27] examined important user power demand forecasting methods based on power big data and neural networks, analyzing the construction of user power demand forecasting models by elucidating power big data and neural networks, and verifying the accuracy of power demand forecasting models based on big data and long short-term memory networks. Literature [28] designed a calculation method for power grid load demand under a multi-energy coupling model, analyzed the factors influencing load under this model, and proposed a least-squares support vector machine optimized by the minimum redundancy maximum association model and adaptive fireworks algorithm for power demand load prediction. The results validated the effectiveness of the aforementioned methods. Literature [29] proposed a hybrid algorithm to improve prediction accuracy. This algorithm uses the Non-Dominated Sorting Genetic Algorithm II (NSGAII) to select input vectors, with the fitness function being a Multi-Layer Perceptron Neural Network (MLPNN). By using the results of NSGAII as input for the Adaptive Neuro-Fuzzy Inference System (ANFIS), the high prediction accuracy of the MLPNN-ANFIS system was validated. Reference [30] proposes a short-term electricity demand forecasting technique that combines two distinct methods: the Elman neural network (ELM) and the adaptive network-based fuzzy inference system (ANFIS). Research indicates that this method outperforms advanced methods such as independent ELM and ANFIS. Literature [31] discusses various techniques and methods for forecasting electricity demand in residential, industrial, and agricultural sectors and examines the role of demand response in managing peak electricity demand and maintaining grid stability by shifting usage to off-peak hours.

This paper designs a data collection process and processing methods for residential electricity consumption behavior data based on the principles of electricity big data technology. Resident electricity consumption behavior data is processed, and resident user behavior labels are proposed, with a focus on explaining the mathematical representation and generation of user behavior labels. Resident electricity consumption behavior data is integrated with multiple behavior labels to construct a resident user profiling process. The selection and meaning of electricity demand influencing factors are then described, and the Attention-Bi-LSTM model is proposed as a prediction model for electricity user demand, forming a prediction method for electricity user demand. Subsequently, select the power marketing unit of Y State Grid as the research object, collect user electricity consumption values and user behavior data from July 15, 2020, and conduct preliminary user clustering. Based on the power user load curve, construct a behavior feature set and calculate the redundancy of each feature, then perform feature clustering and analysis of power users. Finally, verify the performance of the proposed prediction model by comparing the model prediction values with the original values.

II. Analysis of residential electricity consumption behavior and profiling methods

II. A.Acquisition of residential electricity consumption data

In order to provide data support for the analysis of residential electricity consumption behavior, power big data technology is utilized to obtain residential electricity consumption behavior data from basic electricity consumption data, equipment electricity consumption data, advanced electricity consumption data, and electricity consumption anomalies. Figure 1 shows the principle of power big data technology.

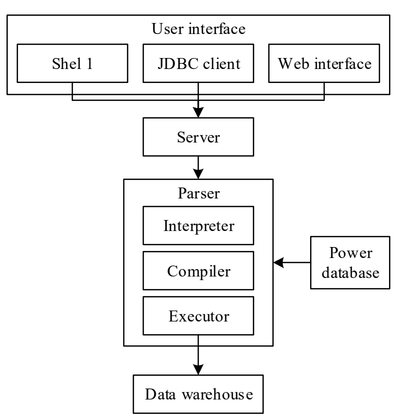

Figure 1: The principle of power big data technology

Power big data technology involves multiple steps, including data collection, data storage, and data processing. The process of collecting residential electricity consumption behavior data is shown in Figure 2.

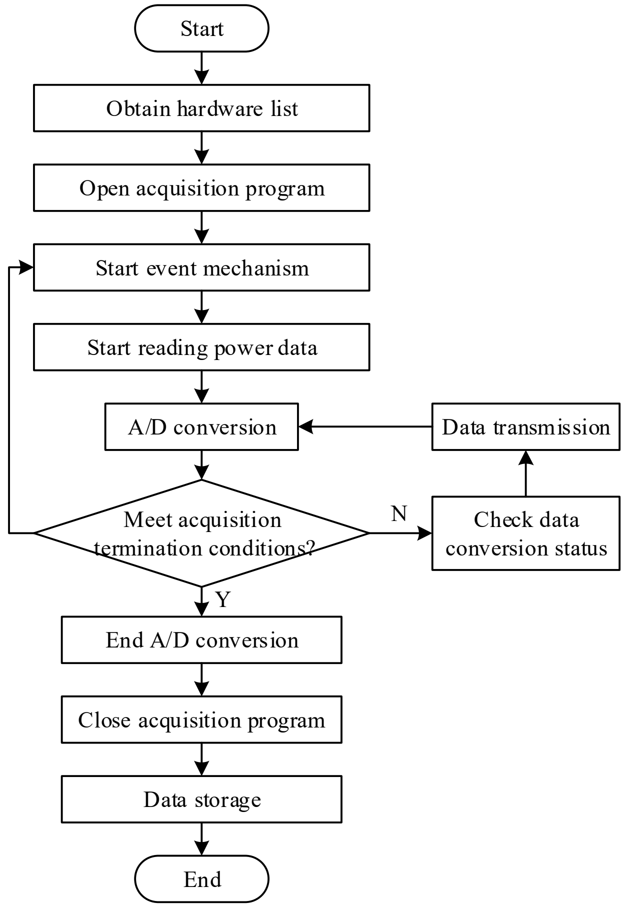

Figure 2: The data collection process of residents' electricity consumption behavior

The collected resident user behavior data includes electricity consumption data, voltage data, current data, specific device electricity consumption data, device usage duration data, switch status data, gear position, and operating conditions. Taking electricity consumption data as an example, the collection results can be expressed as in Equation (1):

$$x_E = \kappa_g \int (x_U \times x_I) dt \tag{1}$$

In the equation, κ_g is the data collection coefficient for residential electricity consumption data, whose specific value is determined by the operational status of power big data technology. x_U and x_I represent the voltage and current data collected by power big data technology, respectively, and t denotes time. Using the above method, data on other residential electricity consumption behaviors can be collected. To ensure the quality of the initial data collection and the operability between different types of residential electricity consumption behavior data, the initial collected data needs to be processed. The specific processing process is as shown in Equation (2):

$$\begin{cases} x_q = y_c(x) \\ x_g = \dfrac{x - x_{\min}}{x_{\max} - x_{\min}} \end{cases} \tag{2}$$

In the formula, $y_c(\ )$ is the data cleaning function, and $x_{\min}$ and $x_{\max}$ correspond to the minimum and maximum values of the initial behavior data collection. Finally, the processing results of residential electricity consumption behavior data are assigned to the initial values to achieve real-time acquisition of residential electricity consumption behavior data.

II. B.Generation of user behavior tags

The process of constructing a three-dimensional profile of power grid users involves an in-depth analysis of user information to extract representative behavioral feature tags. These tags directly reflect user characteristics, and when they are systematically integrated and correlated, they can collectively paint a comprehensive and multidimensional picture of power grid users. Based on the quantitative analysis of residential user behavior, feature tags are generated from basic and behavioral aspects. Basic tags refer to the collection of fundamental attribute information about grid users. These tags are determined by collecting basic user information, including key characteristics such as the user's name, age, gender, income level, and occupational position. Residential user behavioral tags specifically include electricity consumption behavior tags, payment behavior tags, and complaint behavior tags. Electricity consumption behavior tags primarily reflect users' personalized characteristics in electricity consumption, encompassing information such as electricity consumption patterns and direct perceptions of power supply quality. The generation results of peak electricity consumption periods and electricity consumption pattern stability tags in residential electricity consumption behavior tags are as shown in Equation (3):

$$\begin{cases} B_{X-p} = \arg\max\left(\frac{E_a}{24}\times t_h\right) \\ B_{X-m} = \sqrt{\frac{\sum_{t=1}^{t_h}\left(E_i(t)-\bar{E}\right)^2}{t_h}} \end{cases} \tag{3}$$

In the equation, E_a, $E_i(t)$, and $\bar{E}$ represent the total electricity consumption, electricity consumption in the i th hour, and average hourly electricity consumption, respectively, where t_h is the number of hours. In the actual label generation process, based on the calculation results of residential electricity consumption behavior analysis indicators, certain variables in equation (3) are assigned values, thereby generating the results of electricity consumption behavior label components. User payment behavior labels accurately depict the behavioral characteristics of users when settling grid fees, covering key elements such as payment tiers and selected payment methods. User demand behavior labels focus on the various service needs generated by grid users during the period of enjoying power supply services, reflecting users' feedback and expectations regarding the service quality and work efficiency of power supply companies. Using the expression in Equation (3), the generation results for all other user behavior label components can be obtained.

II. C.Building a profile of residential users

Combine the generated user behavior tags with the characteristics of changes in captured electricity usage data to form the core content of the user electricity usage behavior profile. Figure 3 shows the process of constructing a residential user profile.

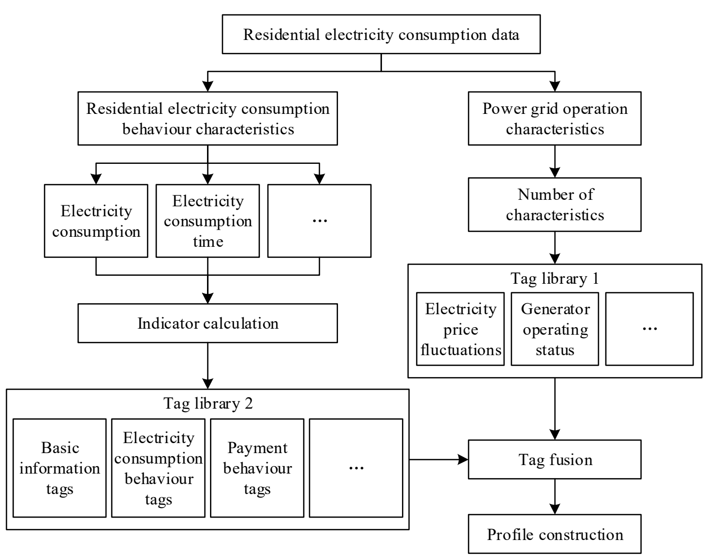

Figure 3: The process of building a resident user profile

According to the process shown in Figure 3, the results of constructing resident user profiles can be quantified as shown in Equation (4):

$$\gamma = \left\langle t_B, R, Z \right\rangle \tag{4}$$

In the formula, t_B , R , and Z represent the label generation time, label name, and user electricity consumption behavior status, respectively. Assigning the label component generation results to the variable Z yields the final user profile construction results. In the actual user profile construction process, various user behavior characteristics and labels can be classified to ensure the accuracy of the user profile construction.

III. Methods for forecasting electricity demand

III. A. Screening of factors affecting electricity demand

III. A. 1) Short-term influencing factors

The factors influencing short-term electricity demand forecasts include meteorological factors, thermal power generation, electricity policy factors, time factors, new energy vehicle sales, and other random factors. This study categorizes the factors influencing short-term electricity demand into significant factors and non-significant factors. Significant factors are those that can cause significant fluctuations in electricity demand within a short period of time. The most significant factor is meteorological factors, such as temperature, humidity, and atmospheric pressure. The next significant factor is thermal power generation. Thermal power generation, as the primary power generation method in China, has historical power generation data that largely reflect historical electricity demand. Therefore, in this study, thermal power generation is classified as a significant factor. Next are time factors, primarily the number of days during holidays and statutory holidays. Since statutory holidays can also cause short-term fluctuations in electricity demand, this is also classified as a significant factor. Non-significant factors are those that require prolonged, sustained influence to cause fluctuations in electricity demand. Non-significant factors generally include economic factors and policy factors. Since this study aims to improve the accuracy of short-term electricity demand forecasts, non-significant factors are assumed to remain stable in this study, with a focus on in-depth analysis of significant factors.

In terms of selecting temperature values, since the data source for this study is provincial power demand, the temperature indicator is taken as the average temperature to represent the provincial temperature status. The provincial average temperature is obtained by calculating the weighted average of the temperature values of each region within the province. Let the provincial average temperature be T_a , and its expression is shown in Formula (5):

$$T_a = \frac{\sum_{i=1}^{n} T_i}{n} \tag{5}$$

In the equation, $T_i\left(i=1,2,\ldots,n\right)$ represents the temperature values of each prefecture-level city, and n is the total number of prefecture-level cities in the province.

To enhance the validity and scientific rigor of the experimental data, this study collected three types of temperature data: maximum temperature, minimum temperature, and average temperature. Additionally, the Pearson correlation coefficient was used to verify the correlation between the three types of temperature data and electricity demand data obtained from historical data. The Pearson correlation coefficient is calculated as shown in Equation (6):

$$R_s = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2}} \tag{6}$$

In the formula, R_s is the correlation coefficient between X and Y, X_i is the value of the independent variable, $\bar{X}$ is the average value of the independent variable, Y_i is the value of the dependent variable, and $\bar{Y}$ is the average value of the dependent variable. The value of R_s ranges from 0 to 1, with values closer to 1 indicating a stronger correlation between the two variables. When $R_s < 0$, it indicates that there is no correlation between the two variables.

In the test results, the correlation coefficients of the three types of temperature indicators with electricity demand are all greater than 0.3, indicating that the three types of temperature indicators have a significant impact on electricity demand. Therefore, all three types of temperature indicators are considered as research factors.

The GBDT model was used to extract important features from the initially selected significant factors, which include maximum temperature, minimum temperature, average temperature, humidity, atmospheric pressure, thermal power generation, and the number of holidays and statutory holidays. Since it is currently impossible to directly determine whether new energy vehicles are significant or non-significant factors, they are also included as influencing factors in the analysis.

Among the initially selected influencing factors (maximum temperature, minimum temperature, average temperature, humidity, atmospheric pressure, thermal power generation, new energy vehicle sales, and the number of public holidays and statutory holidays), the feature importance of new energy vehicle sales and atmospheric pressure for short-term electricity demand is 0. The reason atmospheric pressure has a low impact on electricity demand may be that it indirectly influences electricity demand through changes in weather conditions, resulting in a lower contribution to electricity demand. New energy vehicle sales do not exhibit significant short-term trends, and the impact of increased sales on electricity demand is negligible. Therefore, this study excludes these two factors and further investigates the remaining factors as influencing factors.

III. A. 2) Medium- and long-term influencing factors

Ten influencing factors were selected as the initial objects of analysis, including: thermal power generation, crude oil processing volume, natural gas production, GDP, population size, industrial added value growth rate, secondary industry added value, tertiary industry added value, disposable income of residents, total retail sales of consumer goods, and electricity consumption baseline. This section further screens these eleven influencing factors using GBDT-LASSO. Since the electricity consumption baseline uses actual historical electricity demand data, the remaining ten indicators are screened. First, the GBDT model is used to rank the weights of the above factors, determining the extent to which the ten influencing factors affect medium- and long-term electricity demand. Then, factors with low influence are removed, and VIF is used to test for multicollinearity. Finally, LASSO regression is used to eliminate multicollinearity among the factors, ultimately determining the system of factors influencing medium- and long-term electricity demand (thermal power generation, crude oil processing volume, GDP, population size, industrial added value growth rate, natural gas production, secondary industry added value, tertiary industry added value, disposable income of residents, and total retail sales of consumer goods).

Calculations show that crude oil processing volume and natural gas production have a negligible impact on medium- and long-term electricity demand, with a feature contribution of 0, while the remaining eight factors have significantly higher feature contributions than these two. Therefore, this study excludes crude oil processing volume and natural gas production, making the factor system more scientifically sound and reasonable.

Therefore, gradient boosting trees were used to calculate the influence of the ten factors on electricity demand. After calculation, this study excluded natural gas production and crude oil processing volume as indicators, and retained the remaining nine indicators, including electricity consumption base, as the factor system for medium- and long-term electricity demand forecasting.

III. B. Attention-Bi-LSTM Model

To address the issues of gradient explosion and vanishing gradients in recurrent neural networks, LSTM employs a gating mechanism to control the updating or discarding of information. It introduces input gates, forget gates, and output gates to remove content that is not relevant to the current situation, thereby extending the retention time of information and enabling the preservation of information from a longer time span. The gating mechanism architecture of LSTM is shown in Figure 4.

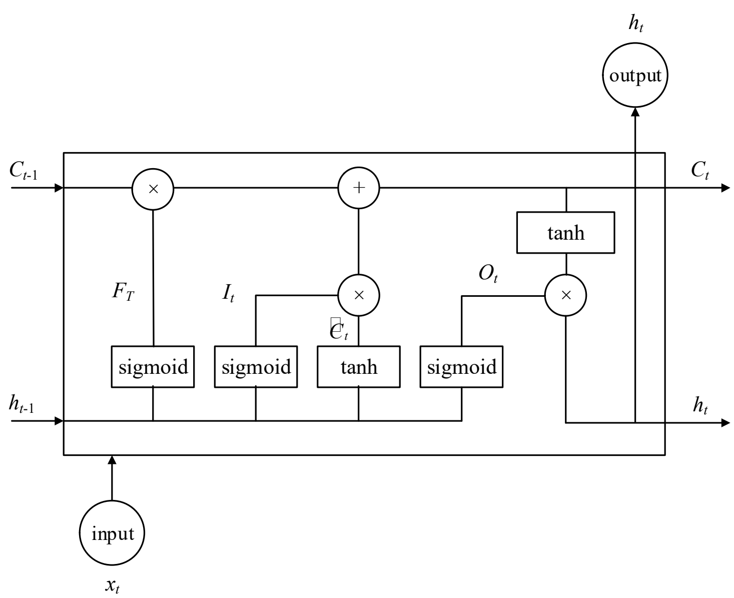

Figure 4: The gating mechanism architecture of LSTM

The inputs to the LSTM gate are the current time step input X_t and the hidden state H_{t-1} from the previous time step. The output is calculated by a fully connected layer using the sigmoid activation function (σ). The overall framework is given by equations (7)-(11):

$$\text{Input gate } I_t = \sigma\left(X_t W_{xi} + H_{t-1} W_{hi} + b_i\right) \tag{7}$$

$$\text{Forget gate } F_t = \sigma\left(X_t W_{xf} + H_{t-1} W_{hf} + b_f\right) \tag{8}$$

$$\text{Gate control unit } \tilde{C}_t = \tanh\left(X_t W_{xc} + H_{t-1} W_{hc} + b_c\right) \tag{9}$$

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \tag{10}$$

$$\text{Output gate } O_t = \sigma\left(X_t W_{xo} + H_{t-1} W_{ho} + b_o\right) \tag{11}$$

In the equation: X_t is the mini-batch input vector at time step t, W_{xi} and W_{hi} are the weight matrices of the input gates, b_i is the bias term of the input gates, W_{xf} and W_{hf} are the weight matrices of the forget gates, b_f is the bias term of the forget gates, $\tilde{C}_t$ is the candidate memory cell to be computed for short-term memory, W_{xc} and W_{hc} are the weight matrices, and b_c is the bias term of the gating unit, C_t is the current state of the gating unit, C_{t-1} denotes the previous state of the unit, W_{xo} and W_{ho} are the weight matrices for the output gate, and b_o is the bias term for the output gate. The tanh function with a range of $[-1,1]$ is used as the activation function, and element-wise multiplication $\odot$ is used to control the flow of information in the hidden state.

The output gate O_t controls the flow of information from the memory cell to the hidden state, and the final output H_t is given by equation (12):

$$H_t = O_t \odot \tanh C_t \tag{12}$$

Unlike LSTM, the Bi-LSTM model combines forward LSTM and backward LSTM. The bidirectional mode takes into account the overall information hidden in the data, performing feature extraction through both forward and backward dimensions, and combining the results extracted in both directions in a specific manner. This effectively mitigates the adverse effects caused by the order of input data on the final results in a single LSTM model, resulting in more comprehensive outcomes.

Currently, the Attention mechanism has been widely used in fields such as handwriting recognition and computer vision. When applied in deep learning, the Attention mechanism can filter out key information from input data, assign higher weights to these key information for effective decision-making, and calculate the probability distribution of attention, thereby eliminating the unreasonable influence of input data on output data. This enhances the influence of key input data, emphasizes the different influences of input data on output data, optimizes feature extraction, and improves prediction performance. Its structure is shown in Figure 5. The structure of the Attention-Bi-LSTM model obtained by incorporating the Attention mechanism into the Bi-LSTM model is shown in Figure 6.

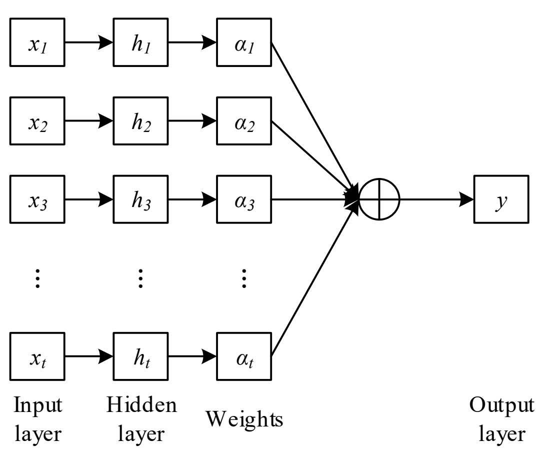

Figure 5: Attention mechanism structure

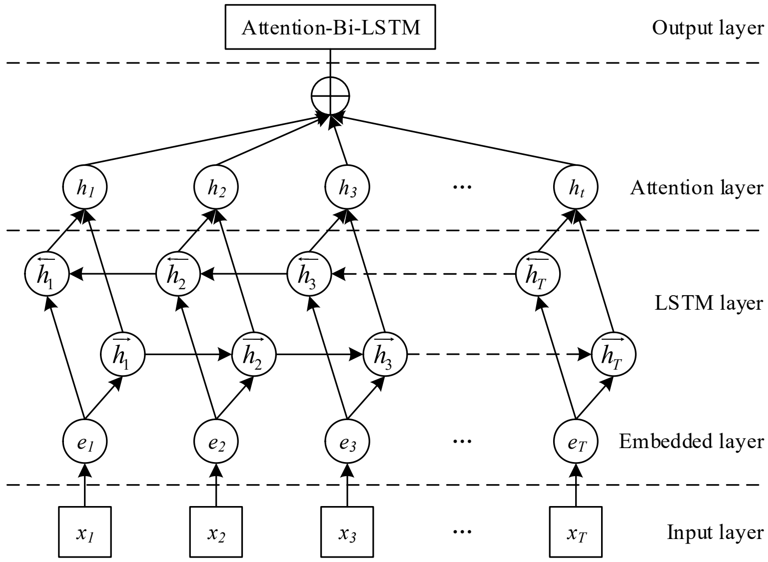

Figure 6: The structure of the Attention-Bi-LSTM model

IV. Analysis of electricity user behavior data and model performance evaluation

IV. A. Characteristics of residential electricity consumption behavior

IV. A. 1) Collection of user electricity consumption data

This experiment selected a power marketing unit of the Y Country Grid as the research object. In 2020, this unit's annual electricity sales reached 1.46x1011kWh. Based on the actual measurement data from its smart meters, the daily electricity consumption data of 6,560 users on July 15, 2020, was selected. The daily electricity consumption data of each user was collected every 30 minutes, containing 47 collection values. The original data set is shown in Figure 7.

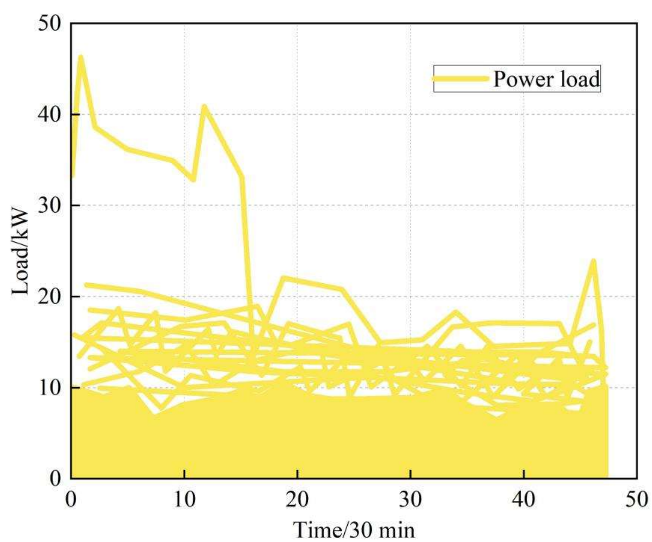

Figure 7: Unlabeled Electricity Customer Load Aggregation

IV. A. 2) Collection of user behavior data

In terms of user distribution, corporate users are primarily concentrated in the industrial and commercial sectors, where they are sensitive to electricity prices, as electricity costs constitute a significant portion of their overall expenses. Residential users are spread across both urban and rural areas and represent a key customer base for State Grid Power Company. Additionally, the company is committed to enhancing the quality and efficiency of its power supply services through measures such as streamlining electricity application processes, reducing costs, and improving transparency, all aimed at increasing customer satisfaction. Collecting electricity user behavior data as sample data, as shown in Table 1, this paper categorizes users into the following seven types: (U1) residential, (U2) corporate, (U3) commercial, (U4) industrial, (U5) residential, (U6) school, and (U7) government.

Table 1: Data collection of power user behavior samples

Serial Number	User type	Electricity consumption / kWh	Maximum power /kW	Voltage level /kV	Electricity charge/yuan
1	U1	850	8	0.35	500
2	U2	13000	60	12	6500
3	U3	8500	100	0.51	4500
4	U4	25000	250	40	13000
5	U5	700	7	0.35	425
6	U6	6500	65	0.53	3500
7	U7	4500	45	15	2500

Observing Table 1, it can be seen that the electricity consumption of resident No. 1 is abnormal, and electricity theft is initially suspected. However, after investigation, it was found that the number of people in the user's home had increased recently, and electricity consumption had increased due to changes in weather and lifestyle habits. The data mining algorithm did not take lifestyle factors into account, leading to a misjudgment of user behavior.

IV. A. 3) User Clustering

For the aforementioned power dataset, when users are divided into three categories, the aggregation evaluation metric contour coefficient reaches its maximum value, while the DBI index reaches its minimum value, indicating the best clustering effect. Therefore, 3 is preliminarily selected as the optimal number of categories for this dataset. By taking the weighted average of the load curve families for the aforementioned three categories, the average load changes for each category of users are obtained, as shown in Figure 8. From the clustering results, users are distinguished based on their overall load levels, achieving good classification performance. The peak load times for the first and second categories of users are both at 1:15 PM, while the minimum load times are 4:15 AM and 5:15 AM, respectively. For the third category of users, the minimum load occurs at 12:45 PM, and the minimum load occurs at 4:15 AM. The morning load accounts for a larger proportion of the daily load compared to the first and second categories of users.

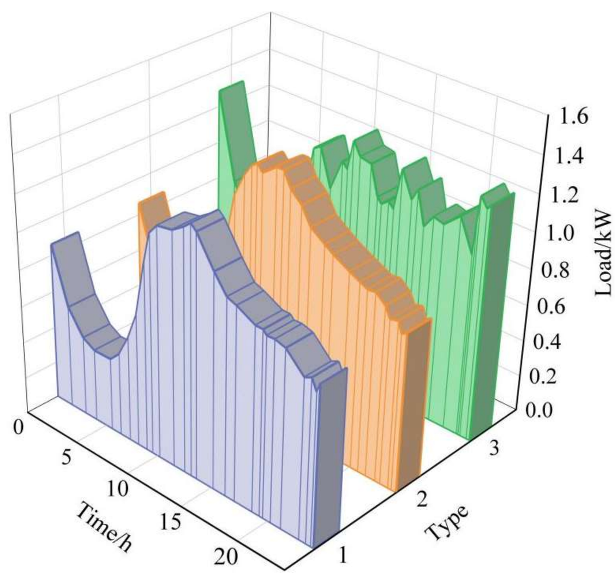

Figure 8: Average load profiles for three types of electricity consumer

IV. B. Drawing a profile of electricity users

IV. B. 1) Calculation of power user characteristics

For the construction of the original feature set T for the load curve of unlabeled power users in this paper, the feature set is defined as T = {daily maximum load, time of maximum load occurrence, peak-to-valley difference, average load, daily minimum load, time of minimum load occurrence, daily electricity consumption, peak-time electricity consumption rate, off-peak electricity consumption rate, valley-time electricity consumption rate, daily load factor, peak-to-valley difference rate}. The 12 features are sequentially numbered as T1 to T12. The peak-to-valley difference rate is defined to reflect the user's demand response capability, calculated as the ratio of the difference between the maximum and minimum loads to the maximum load. Define peak, off-peak, and valley consumption rates as the ratio of electricity consumption during peak, off-peak, and valley periods to total electricity consumption, reflecting users' electricity consumption characteristics. Define the load factor as the ratio of average load to maximum load, reflecting daily load variations, i.e., demand response regulation capability. Using the distance correlation coefficient as the evaluation metric, preliminarily calculate the redundancy of each feature in the feature set. The calculation results are shown in Table 2.

Table 2: Distance coefficient based feature set matrix for power users' electricity usage

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12
T1	1	0.243	1.103	0.91	0.651	0.206	0.91	0.275	0.361	0.484	0.472	0.474
T2	0.243	1	0.246	0.232	0.186	0.196	0.232	0.285	0.382	0.361	0.277	0.285
T3	1.103	0.246	1	0.851	0.558	0.208	0.851	0.289	0.37	0.505	0.501	0.52
T4	0.91	0.232	0.851	1	0.931	0.211	1	0.216	0.332	0.345	0.473	0.354
T5	0.651	0.186	0.558	0.931	1	0.185	0.931	0.242	0.226	0.178	0.534	0.559
T6	0.206	0.196	0.208	0.211	0.185	1	0.211	0.187	0.264	0.318	0.194	0.215
T7	0.91	0.232	0.851	1	0.931	0.211	1	0.216	0.332	0.345	0.473	0.354
T8	0.275	0.285	0.289	0.216	0.242	0.187	0.216	1	0.624	0.466	0.423	0.394
T9	0.361	0.382	0.37	0.332	0.226	0.264	0.332	0.624	1	0.659	0.301	0.34
T10	0.484	0.361	0.505	0.345	0.178	0.318	0.345	0.466	0.659	1	0.405	0.487
T 11	0.472	0.277	0.501	0.473	0.534	0.194	0.473	0.423	0.301	0.405	1	0.943
T12	0.474	0.285	0.52	0.354	0.559	0.215	0.354	0.394	0.34	0.487	0.943	1

Table 2 shows that there is a high degree of redundancy (0.206–1.103) between the peak-to-valley difference (T3) and the daily maximum load (T1). Daily electricity consumption is highly correlated with the daily minimum load and average load, and it is necessary to filter redundant features using a feature evaluation function.

IV. B. 2) Clustering of characteristics of electricity users

The average payment amount and load growth coefficient are used as two value characteristics of electricity users to assess user value. Using the user profiling method proposed in this paper, electricity users are classified into four categories: high-value users (high average payment amount and large growth coefficient), ordinary users (high average payment amount and small growth coefficient), potential users (low average payment amount and large growth coefficient), and low-value users (low average payment amount and small growth coefficient). After obtaining all value characteristics, the K-means algorithm is used to cluster power users based on their characteristics, enabling accurate and rapid classification of power users into four categories, as shown in Figure 9.

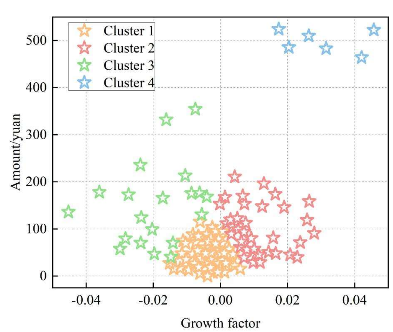

Figure 9: K-means clustering results

Figure 9 clearly shows four distinct clusters. Cluster 1 has a low average payment amount and a small growth coefficient, indicating low-value users. Cluster 2 has a low average payment amount but a large growth coefficient, indicating potential users. Cluster 3 has a high average payment amount and a small growth coefficient, indicating ordinary users. Cluster 4 has a high payment amount and a large growth coefficient, indicating high-value users.

IV. C. Performance evaluation of prediction models

Based on the complexity of the dataset and the number of data points, this paper sets up a two-layer LSTM network, with each layer having 60 memory units. To find the optimal batch_size and epochs parameters, cross-validation is performed using the Grid-SearchCV method, where the batch_size parameter represents the number of samples selected for each training run, and the epochs parameter represents the number of times the learning algorithm operates on the entire training dataset. The Grid-SearchCV method refers to an exhaustive search of specified parameter combinations. Based on the cross-validation results, the optimal batch_size was determined to be 50, and the optimal epochs were determined to be 15. These two optimal parameters were used to train the LSTM model and applied to the load data set of the power marketing unit of Y State Grid Corporation of China for calculation. The prediction performance of the LSTM model is shown in Figure 10, and the prediction error is shown in Figure 11. It can be seen that the prediction curve closely fits the original data curve, with the prediction error remaining within the range [-5000, 6000], indicating that the model has good prediction performance.

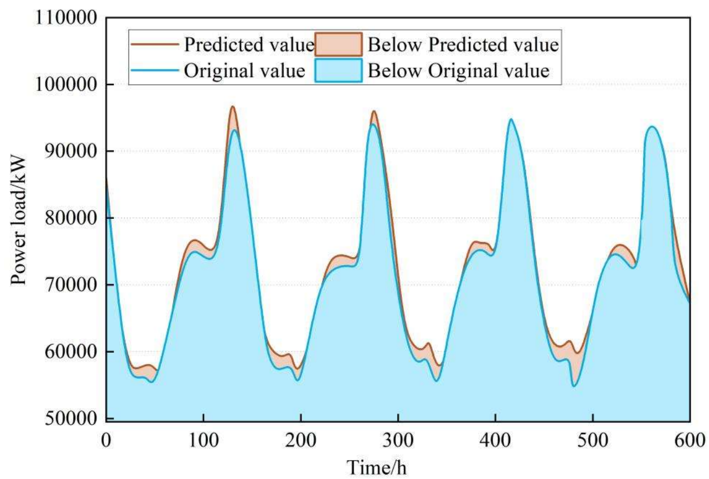

Figure 10: LSTM prediction effect

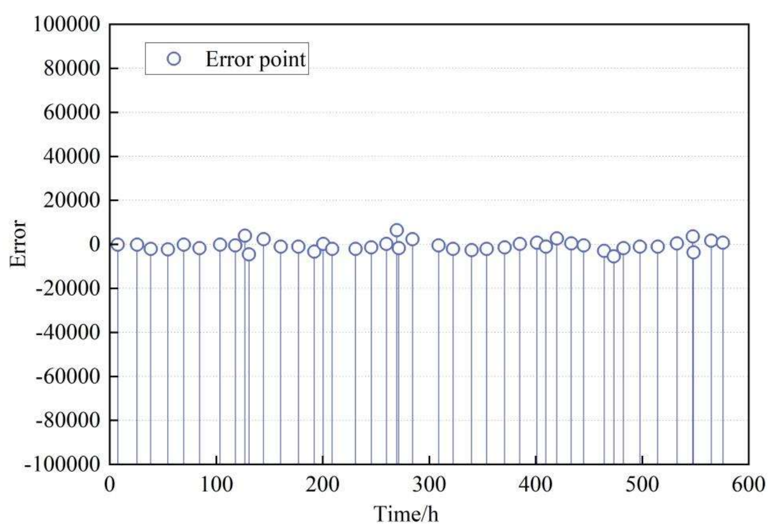

Figure 11: LSTM prediction error scatter plot

V. Conclusion

This paper employs power big data technology to collect residential electricity consumption behavior data, generate user behavior tags, and proposes a method for constructing residential user profiles. Additionally, power demand influencing factors are categorized into two types: short-term and medium-to-long-term. The Attention-Bi-LSTM model is utilized for residential power demand forecasting. This approach achieves precise forecasting of residential power demand driven by power user behavior data, providing effective data references for optimizing power services.

In an experiment using Y State Grid Power Marketing Units as a sample, the user profiling method proposed in this paper was used to classify power users into four categories: high-value users, ordinary users, potential users, and low-value users. The proposed prediction model aligns with the original data curve trends and directions for load forecasting of user electricity demand in the dataset, with prediction errors controlled within the range of [-5000, 6000].

Funding

This work was supported by the State Grid Jiangxi Electric Power Co., Ltd. Fuzhou Power Supply Branch Science and Technology (Project No.5218F0220002).

References

[1] Morales-Acevedo, A. (2014). Forecasting future energy demand: Electrical energy in Mexico as an example case. Energy Procedia, 57, 782-790.

[2] Muralitharan, K., Sakthivel, R., & Vishnuvarthan, R. (2018). Neural network based optimization approach for energy demand prediction in smart grid. Neurocomputing, 273, 199-208.

[3] He, Y., Wang, M., Guang, F., & Zhao, W. (2020). Research on the method of electricity demand analysis and forecasting: The case of China. Electric Power Systems Research, 187, 106408.

[4] Verdejo, H., Awerkin, A., Becker, C., & Olguin, G. (2017). Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile. Renewable and Sustainable Energy Reviews, 74, 512-521.

[5] Perera, S., Dissanayake, S., Fernando, D., De Silva, S., & Rankothge, W. (2019, December). Supply and demand planning of electricity power: a comprehensive solution. In 2019 IEEE Conference on Information and Communication Technology (pp. 1-6). IEEE.

[6] Pełka, P. (2023). Analysis and forecasting of monthly electricity demand time series using pattern-based statistical methods. Energies, 16(2), 827.

[7] Jiménez, J., Pertuz, A., Quintero, C., & Montaña, J. (2019). Multivariate statistical analysis based methodology for long-term demand forecasting. IEEE Latin America Transactions, 17(01), 93-101.

[8] Ghazal, T. M. (2022). Energy demand forecasting using fused machine learning approaches. Intelligent Automation & Soft Computing, 31(1), 539-553.

[9] Javanmard, M. E., & Ghaderi, S. F. (2023). Energy demand forecasting in seven sectors by an optimization model based on machine learning algorithms. Sustainable Cities and Society, 95, 104623.

[10] Hrnjica, B., & Mehr, A. D. (2020). Energy demand forecasting using deep learning. Smart cities performability, cognition, & security, 71-104.

[11] Aderibigbe, A. O., Ani, E. C., Ohenhen, P. E., Ohalete, N. C., & Daraojimba, D. O. (2023). Enhancing energy efficiency with ai: a review of machine learning models in electricity demand forecasting. Engineering Science & Technology Journal, 4(6), 341-356.

[12] Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. Applied energy, 238, 1312-1326.

[13] del Real, A. J., Dorado, F., & Durán, J. (2020). Energy demand forecasting using deep learning: applications for the French grid. Energies, 13(9), 2242.

[14] Grandón, T. G., Schwenzer, J., Steens, T., & Breuing, J. (2024). Electricity demand forecasting with hybrid classical statistical and machine learning algorithms: Case study of Ukraine. Applied Energy, 355, 122249.

[15] Wang, X., Wang, H., Bhandari, B., & Cheng, L. (2024). AI-empowered methods for smart energy consumption: A review of load forecasting, anomaly detection and demand response. International Journal of Precision Engineering and Manufacturing-Green Technology, 11(3), 963-993.

[16] Panda, S., Mohanty, S., Rout, P. K., Sahu, B. K., Bajaj, M., Zawbaa, H. M., & Kamel, S. (2022). Residential Demand Side Management model, optimization and future perspective: A review. Energy Reports, 8, 3727-3766.

[17] Mir, A. A., Alghassab, M., Ullah, K., Khan, Z. A., Lu, Y., & Imran, M. (2020). A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons. Sustainability, 12(15), 5931.

[18] Zhang, H., Du, W., Xia, X., Wang, X., & Han, Z. (2025, March). Research on Business Process Optimization Model of Customer Service Center of Power Enterprises Based on Deep Learning. In 2025 IEEE International Conference on Electronics, Energy Systems and Power Engineering (EESPE) (pp. 257-263). IEEE.

[19] Zhao, Z., & Liu, J. (2011, June). A new communication services optimization method based on services risk balancing degree for power system. In 2011 International Conference on Computer Science and Service System (CSSS) (pp. 994-997). IEEE.

[20] Kong, L., Liang, L., Xu, J., Zhang, W., & Zhu, W. (2019). The optimization of pricing strategy for the wind power equipment aftermarket service. Industrial Management & Data Systems, 119(3), 521-546.

[21] Fan, H., Yu, K., Li, Z., & Shahidehpour, M. (2019). Optimization of power supply capacity of distribution network considering the participation of power sales companies in spot power trading. IEEE Access, 7, 99651-99657.

[22] Miao, X., & Ilić, M. D. (2020). High quality of service in future electrical energy systems: a new time-domain approach. IEEE Transactions on Sustainable Energy, 12(2), 1196-1205.

[23] Saini, S. (2018). Analysis of service quality of power utilities. International Journal of Research in Engineering Application & Management, 3(11), 1-8.

[24] Hossain, A., Ridoy, M. H., Chowdhury, B. R., Hossain, M. N., Rabbi, M. N. S., Ahad, M. A., ... & Hasan, M. S. (2024). Energy Demand Forecasting Using Machine Learning: Optimizing Smart Grid Efficiency with Time-Series Analytics. Journal of Environmental and Agricultural Studies, 5(1), 26-42.

[25] Haque, A. U., Mandal, P., Meng, J., & Pineda, R. L. (2012). Performance evaluation of different optimization algorithms for power demand forecasting applications in a smart grid environment. Procedia Computer Science, 12, 320-325.
[26] Arnob, S. S., Arefin, A. I. M. S., Saber, A. Y., & Mamun, K. A. (2023). Energy demand forecasting and optimizing electric systems for developing countries. IEEE Access, 11, 39751-39775.
[27] Liu, Y., Qiao, R., Wu, H., Han, K., & Xin, Z. (2024, February). Power Demand Forecasting Method for Important users based on Power Big Data and Neural Network. In 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-5). IEEE.
[28] Liu, D., Wang, L., Qin, G., & Liu, M. (2020). Power load demand forecasting model and method based on multi-energy coupling. Applied Sciences, 10(2), 584.
[29] Jadidi, A., Menezes, R., de Souza, N., & de Castro Lima, A. C. (2019). Short-term electric power demand forecasting using NSGA II-ANFIS model. Energies, 12(10), 1891.
[30] Wu, C., Li, J., Liu, W., He, Y., & Nourmohammadi, S. (2023). Short-term electricity demand forecasting using a hybrid ANFIS–ELM network optimised by an improved parasitism–predation algorithm. Applied Energy, 345, 121316.
[31] Sharma, M., Mittal, N., Mishra, A., & Gupta, A. (2023). Survey of electricity demand forecasting and demand side management techniques in different sectors to identify scope for improvement. Smart Grids and Sustainable Energy, 8(2), 9.