Saturated Load Forecast of Hebei Province Based on LSTM

Shiyao Hu¹, Yang Zhao¹, Xiaolin Tan¹, Shiping Geng²,*, Chunguang He¹ and Xiaopeng Guo²

¹State Grid HeBei Economic Research Institute, Hebei, China
²School of Economics and Management, North China Electric Power University, Beijing, China

*Corresponding author e-mail: Gengsp@ncepu.edu.cn

Abstract. Aiming at the characteristics of saturation load forecasting with strong uncertainty and greater time correlation, and comprehensively considering the relevant factors that affect the saturation load level, this paper establishes a multi-factor saturation load forecasting model based on long-short term memory neural networks. Firstly, the judgment basis of saturated load state is determined according to the related factors of foreign developed regions; then, based on the Logistic curve model, the socio-economic factors affecting the saturated load level are predicted as the input of the network model; finally, based on the historical social and economic development data and electricity consumption data of Hebei Province, the future social and economic development of Hebei Province is divided into four different scenarios, and the long-short term memory neural network model is used to predict, respectively discuss the time and scale of Hebei Province entering the saturation load level under different scenarios, and provide reference and suggestions for the future power planning of Hebei Province.

Keywords: Long-Short Term Memory, Artificial Neural Network, Saturated Load Forecasting, Saturation Judgment

1. Introduction

With the continuous development of China's social economy, the city's power demand is also increasing [1,2]. Generally, the growth of electric power will make corresponding trend changes according to the change of economic development. However, when the social and economic development reaches a certain stage, it will be affected by the city's population, urban planning and other factors. The annual growth rate of power load will gradually decrease due to the maturity and relative stability of urban economic development, and the load level will tend to a saturated state, that is, the saturated load of the city [3]. The research on the scale of urban saturated load has important guiding significance for the planning and development of urban power industry, especially the construction and transformation of urban power grid [4,5].

Different from the traditional load forecasting method for specific years, the time expansion of urban saturated load forecasting is often larger, and the factors to be considered are extensive [6]. It includes the city's policy planning, economic development, energy and resource conditions, social
population and living standards. At present, the methods and models for saturated load forecasting at home and abroad mainly include economic trend analysis method, load trend analysis method and intelligent algorithm model \[7,8\]. The method of economic trend analysis is based on the analysis of urban economic development, population and functional area planning, including macro prediction method and micro prediction method. In reference \[9\], the population of a city in East China is predicted, and the saturation scale and saturation time of the whole social power consumption of the city are predicted by using the method of per capita power consumption prediction. In reference \[10\], according to the long-term planning of a city, based on the spatial saturated load density load forecasting method, the saturated load value of each functional area of the city is predicted. The load trend analysis method is to fit the load curve according to the historical data to predict the scale of saturated load and the time of outgoing line. The commonly used method is Logistic curve model, which focuses on analyzing the scale of saturated load and the occurrence time of saturated load. Literature \[11\] summarizes the economic operation and power consumption over the years, and forecasts the long-term development trend of power consumption and maximum load in Hubei Province by using Logistic model, and obtains the saturation scale and time point of future power demand. In reference \[12\], the growth process of power demand is divided into three stages, and a modified adaptive Logistic model is proposed. Based on the actual power demand data in East China, the model is verified to be practical and accurate. However, traditional economic and load trend forecasting methods often only make statistical forecasts for a single dependent variable, ignoring the influence of economy and society on power development in the process of load development. More and more machine learning and intelligent algorithms are gradually applied to load level forecasting. In reference \[13\], BP neural network was applied to predict the power consumption of consumers from the single time factor and considering the influence of multiple factors such as temperature factor. The results show that BP neural network has good self-learning processing ability for multiple factors, and the prediction effect based on multiple factors is better. In reference \[14\], considering that direct use of the maximum value of measured load data for spatial load forecasting will reduce the prediction accuracy, a spatial forecasting method based on chance fuzzy information granulation and support vector machine is proposed. An engineering example shows the effectiveness and practicability of this method.

To sum up, the traditional saturated load forecasting model does not consider the various correlation factors affecting the saturation load level, which makes its prediction accuracy not high and has limitations. However, most intelligent algorithms, such as BP neural network prediction model, take into account various correlation factors, but its delayed effect on influencing factors and load persistence cannot be expressed. Therefore, this paper constructs a long-short term memory (LSTM) neural network model for saturated load forecasting. The memory unit in the network stores historical load and related factor data to meet the requirements of time sequence and influence delay. Finally, based on the historical social and economic development data and electricity consumption data of Hebei Province, the future electricity consumption of Hebei Province is predicted, and according to the law of electricity consumption growth, the time and scale of Hebei Province entering saturation stage under different scenarios are divided.

2. Judgment Basis of Saturated Load Stage
Based on the analysis of typical countries and regions, such as the United States, the United Kingdom, Japan and so on, the criteria for judging the saturation stage of power demand are summarized, as shown in Table 1. The standard is used to test whether a city has entered the stage of saturated load. The judgment indicators can be roughly divided into two categories: the main indicators: the growth rate of electricity consumption and the annual maximum load growth rate; and the auxiliary indicators: industrial structure, growth rate of permanent residents, urbanization rate, per capita GDP, etc.
Table 1. Index system for determining saturated load level

| Country / region | Growth rate of electricity consumption | Maximum load growth rate | Growth rate of resident population | Industrial structure (first, second and third industries) | Urbanization rate | GDP per capita (U.S. dollar / person) |
|------------------|----------------------------------------|--------------------------|------------------------------------|----------------------------------------------------------|------------------|-------------------------------------|
| The United States| 1.94%                                  | 1.77%                    | 0.59%                              | 1.1:25.4:73.5                                            | 76.1%            | 32847.1                             |
| The United Kingdom| 0.89%                                  | 1.43%                    | 0.29%                              | 1.2:28.5:70.3                                            | 78.9%            | 23326.3                             |
| Japan            | 1.98%                                  | 1.40%                    | 0.24%                              | 1.8:29.9:68.3                                            | 78.3%            | 36304.9                             |

Criteria: <2% <2% <0.65% >70% > 15000 U.S. dollar / person

From Table 1 and related literature, it can be considered that when a country or region enters the saturated load settlement, its electricity consumption and annual load growth rate are less than 2% [15,16]. In our country, due to different factors such as the geographical location, resource distribution, economic development degree, and policies of different cities, it is possible that only part of the auxiliary indicators of a certain area meet the criterion of the saturation load stage, while the other part cannot. Therefore, it is believed that the growth rate of electricity consumption or maximum load in this area for 5 consecutive years is less than 2%, and other auxiliary standards can meet 1-2 of them [17].

3. LSTM Prediction Model

3.1. Model Principle

The long-short memory neural network was proposed by HOCHREITER and SCHMIDUBER (1997) [18], it has been improved and popularized by GRAVES, it belongs to a kind of recurrent neural network (RNN). Its advantage is that RNN only has the function of memory temporary storage, while LSTM has the function of long-short term memory. It can effectively eliminate the gradient disappearance of RNN in dealing with long-term dependence. In many problems, LSTM has achieved considerable success and has been widely used.

As a derivative of RNN, the hidden layer of LSTM has two main lines, and its calculation unit structure is shown in Figure 1. The upper main line is the main memory line, in which \( C_t \) is the state of memory unit at time \( t \); the next main line mainly controls the input and output, where \( X_t \) is the input value outside the network at time \( t \), and \( O_t \) is the output value of LSTM at time \( t \).

![LSTM computing unit](image)

Figure 1. LSTM computing unit
An LSTM computing unit consists of four parts: input gate, forgetting gate, output gate and memory unit. Firstly, the forgetting gate mainly helps the network to forget the past input data and reset the memory unit. The forgetting state of the memory unit is determined by the value between 0-1 output by the Sigmoid excitation function. 1 stands for "completely reserved", and 0 represents "completely forgotten". The calculation process is as follows:

$$f_t = \sigma(W_f \cdot [O_{t-1}, x_t] + b_f)$$  

(1)

Next is the input gate: to decide what information to store in the memory unit, including two parts: one is the Sigmoid layer of the "input gate layer" to determine the value we want to update; the other part is the tanh layer to generate a new candidate value \(C_{\text{c-l}}\), which makes the memory unit generate an updated state value by combining the two parts. The calculation process is as follows:

$$i_t = \sigma(W_i \cdot [O_{t-1}, x_t] + b_i)$$  

(2)

$$C_{\text{c-l}} = \tanh(W_C \cdot [O_{t-1}, x_t] + b_C)$$  

(3)

After the forgetting gate forgets the information, the input gate adds new information, so that the final output of the memory unit is as follows:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot C_{\text{c-l}}$$  

(4)

The output gate mainly determines what information to output. The output is based on the memory unit. First, it passes through the tanh layer (making the output value within [-1,1]), and then through the Sigmoid layer (which part of the memory unit needs to be output), and finally outputs the useful information. The calculation process is as follows:

$$h_t = \sigma(W_h \cdot [O_{t-1}, x_t] + b_h)$$  

(5)

The final output is:

$$O_t = h_t \cdot \tanh(C_t)$$  

(6)

In the above formula, \(f_t\) is the forgetting weight; \(i_t\) is the input weight; \(h_t\) is the output weight; \(W_f, W_i, W_C, W_h\) are the weight matrix of forgetting gate, input gate, memory unit and output gate respectively; \(b_f, b_i, b_C, b_h\) are the bias vector of forgetting gate, input gate, memory unit and output gate respectively; \(\sigma\) is Sigmoid function.

3.2. Prediction Steps

In this paper, the steps of saturated load forecasting model based on LSTM are as follows, and the prediction flow chart is shown in Figure 2.

Step 1: selection of related factors. By constructing grey correlation analysis, quantify the correlation between social, economic, energy and other correlation factors related to load development to regional electricity consumption, and screen key factors with larger correlation coefficients and bring them into the forecast model.

Step 2: prediction of correlation factors for social and economic influencing factors, such as population, regional GDP and other indicators, the Logistic curve, that is, the S-shaped curve, is used to predict the trend. The logistic curve formula is:

$$y = \frac{1}{a + be^x}$$  

(7)
The proportion of the structure of the secondary industry and the tertiary industry, and the consumption index of residents are extrapolated by comparing the predicted areas with the similar regions abroad.

Step 3: data processing. Because the units of the sample data of the selected Association variables are not uniform and the order of magnitude is quite different, therefore, the selected indicators are processed without dimension to eliminate the errors caused by the difference of units and dimensions. The result after processing is between [0,1]. The specific formula is as follows:

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$  \hspace{1cm} (8)

Where $X^*$ is the normalized result; $X$ is the original value of the sample; $X_{\max}$ and $X_{\min}$ are the maximum and minimum values of the variable samples respectively.

Step 4: construct LSTM prediction model to train prediction. Select 10% of the input data of the prediction area as the test set to test the accuracy of the LSTM prediction model in terms of prediction. After the model test is completed, the prediction results of related variables obtained in step 2 are input into the LSTM prediction model, and the extrapolation prediction results of future electricity consumption are output.

Step 5: enter saturation time point prediction, and saturation scale prediction. Calculate the growth rate of electricity consumption and the changes of other relevant judgment indicators according to the data of the future regional electricity consumption predicted and processed by the LSTM forecasting model. According to the judgment basis in Table 1, determine the load scale and entry time of the predicted area entering the saturation stage.

Where $a$ is the saturation height of the curve; $b$ and $k$ are the coefficients related to the event respectively; the upper limit of the curve is $a^{-1}$; $a > 0$, $b > 0$, $k < 0$.

Figure 2. LSTM prediction process

4. Example Analysis
Select Hebei Province from 1995 to 2019 GDP, Permanent population, per capita GDP, proportion of secondary industry, proportion of tertiary industry, urbanization rate and consumer price index of Hebei Province are the influencing factors of electricity consumption. The LSTM network prediction model is constructed. The variable sample data from 1995 to 2014 are selected as the training set, and the sample data from 2015 to 2019 are selected as the test set to verify the validity of the model, and the future saturated electricity scale of Hebei Province is predicted.

4.1 Validation of LSTM Prediction Model

Firstly, the sample data is normalized. After processing, the grey correlation degree is calculated for the seven selected correlation variables, and the key factors affecting the power consumption level are screened. The results of correlation degree are shown in Table 2. Selecting 0.7 as the screening standard of grey correlation degree, it can be seen that the grey relational grade of the proportion of secondary industry and consumption index is lower than others, and the grey relational grade of other variables is greater than 0.7. Therefore, the remaining five variables, namely population, urbanization rate, GDP, per capita GDP and the proportion of tertiary industry, are used as input variables of LSTM model to participate in the training and testing of the model.

| Variable     | Population | Urbanization rate | GDP    | Per capita GDP | Proportion of secondary industry | Proportion of tertiary industry | Consumer index |
|--------------|------------|-------------------|--------|---------------|---------------------------------|--------------------------------|----------------|
| Grey relational grade | 0.70395    | 0.7727            | 0.8223 | 0.8620        | 0.6234                          | 0.7080                         | 0.6393         |

To verify the effectiveness of the LSTM proposed in this paper, the LSTM model prediction, particle swarm optimization BP neural network prediction and support vector machine (SVM) model prediction are carried out based on the historical data to forecast the power consumption of Hebei Province from 2015 to 2019. The average absolute error MAE, the average relative percentage error absolute value MAPE, and the mean square error MSE are used to test and evaluate the prediction results.

The training set data are brought into the three models to train the model, and three prediction models are obtained. The test set prediction results are shown in Figure 3, and the error is shown in Table 3:

![Figure 3. Power consumption forecast results of Hebei Province in 2015-2019](image-url)
Table 3. Prediction errors of different models

| Prediction model                                      | Mae / 100 million kwh | MAPE/% | MSE / 100 million hours |
|-------------------------------------------------------|------------------------|--------|-------------------------|
| PSO-BP model considering five key factors             | 212.648                | 6.165  | 217.800                 |
| Support vector machine (SVM) model considering five key factors | 271.0069               | 7.310  | 391.201                 |
| LSTM prediction model considering seven related factors | 268.244                | 7.950  | 300.374                 |
| LSTM prediction model considering five key factors     | 161.616                | 3.657  | 164.958                 |

It can be seen from Figure 3 and table 3 that The LSTM prediction model established in this paper has the best prediction effect based on five key factors after removing two related variables with lower correlation degree, and its overall error is the smallest, indicating that the prediction accuracy is high. However, the LSTM prediction model with seven correlation variables has larger prediction error and lower prediction accuracy, which indicates that if the influence variables with small correlation degree are input, the prediction accuracy of the model will be affected and the accuracy of the model will be reduced. Compared with the PSO BP neural network model and support vector machine model used in this paper, it can be seen from Figure 3 that the forecast trend of power consumption in 2015-2019 is close to the actual power consumption, but it can be seen from table 3 that the prediction results of these two models considering five key factors are less effective than LSTM prediction model, and the accuracy is not as high as that of LSTM prediction.

4.2. Forecast of Saturated Electricity in Hebei Province

According to the steps of saturated power load forecasting, five key variables, namely permanent population, urbanization rate, GDP, per capita GDP and the proportion of tertiary industry, are extrapolated and predicted. Based on the historical data of Hebei Province from 1995 to 2019, the logistic curve model is used to extrapolate the resident population, urban population and GDP. At the same time, the fluctuation value of historical data relative to the model fitting value is obtained, which can provide reference for future scenarios. For the calculation of the proportion of the three industries, the evolution of the industrial structure of Hebei Province in the future is calculated by comparing the industrial structure of Hebei Province over the years with the historical industrial structure of Britain, the United States and Japan.

The settings are as follows:

Scenario 1: rapid economic and social development (GDP growth rate is greater than 2%, the permanent population growth rate is greater than 2%), and the share of tertiary industry is 2% higher than the judgment standard.

Scenario 2: rapid economic and social development (GDP growth rate is greater than 2%, the permanent population growth rate is greater than 2%), and the share of tertiary industry is 2% lower than the judgment standard.

Scenario 3: population, GDP and proportion of tertiary industry are all extrapolated predicted equilibrium values;

Scenario 4: Economic and social development is slow (population is less than 2% of the equilibrium value, GDP is less than 2% of the equilibrium value), and the share of tertiary industry is 2% lower than the judgment standard.

Through the LSTM prediction model proposed in this paper, the power consumption of Hebei Province under the above four scenarios are predicted. The prediction results are shown in Figure 4:
According to the extrapolation results of the five key factors and the future electricity consumption of Hebei Province obtained by LSTM prediction model, the judgment in Table 1 is based on the time point of Hebei Province entering the saturation load stage. The judgment results of saturation time point under four scenarios are shown in Table 4.

### Table 4. Saturation time points of different scenes

| Judgment index | Scenario 1 | Scenario 2 | Scenario 3 | Scenario 4 |
|----------------|------------|------------|------------|------------|
| The growth rate of electricity consumption for 5 consecutive years is less than 2% | 2031 | 2032 | 2032 | 2033 |
| The growth rate of resident population is less than 0.65% | 2028 | 2028 | 2029 | 2028 |
| Urbanization rate > 70% | 2029 | 2029 | 2030 | 2031 |
| Tertiary industry accounts for more than 65% | 2031 | 2032 | 2032 | 2032 |

It can be seen from Table 4 that if the power growth rate is less than 2% for five consecutive years, scenario 1 will enter the saturated load stage in 2031 due to the rapid economic development and the rapid development of the tertiary industry in the industrial structure. Scenario 2 and scenario 3 will enter the saturated load stage in 2032, and scenario 4 will enter the saturated load stage later than other scenarios due to the slow economic development. By meeting 1-2 auxiliary judgment indicators, scenario 1 and scenario 2 will enter the saturated load stage in 2029, scenario 3 will enter the saturated load settlement in 2030, and scenario 4 will enter the saturated load stage in 2031. Considering the main index and auxiliary index, the time point of each scenario entering saturation load is obtained. The saturated year is brought into the prediction result of LSTM model, and the package and scale of four scenarios are obtained respectively, as shown in Table 5:

### Table 5. Saturation time and saturation scale of different scenarios

| Scenario | Time point / year of final saturation stage | Final saturated load scale / 100 million kwh |
|----------|------------------------------------------|-------------------------------------------|
| 1        | 2031                                     | 56428                                     |
| 2        | 2032                                     | 52036                                     |
| 3        | 2032                                     | 51615                                     |
| 4        | 2033                                     | 49103                                     |

It can be seen from Table 5 that under different scenarios, due to the economic growth, population growth and the proportion of tertiary industry in industrial institutions, the final predicted saturated load scale is also different. Among them, scenario 1 has the largest saturated load scale of 5642.8 billion kwh when the economic and social development is fast and the proportion of tertiary industry is relatively high; the final saturated load scale obtained by scenario 2 and scenario 3 is relatively consistent, which is about 5200 billion kwh. The reason may be that although the social and economic development of scenario 2 is better, the development of the tertiary industry is relatively slow, and the
industrial power consumption is relatively low, so the final power consumption scale is lower than that of scenario 1. In scenario 4, due to the slow economic development, relatively small population and low development level of tertiary industry, the final saturated load scale is the smallest, which is 4910.3 billion kwh. Based on the above analysis results, it can be concluded that the time for Hebei Province to enter the saturation load stage in the future is 2031-2033, and the saturated load scale range is 4910.3 billion kwh-5642.8 billion kwh.

5. Conclusion

Based on the historical data of Hebei Province from 1995 to 2019, this paper forecasts the time point of saturated load and the scale of saturated load by constructing LSTM forecasting model. The conclusions are as follows:

(1) Compared with the conventional forecasting model, the LSTM prediction model proposed in this paper has a long-term memory unit which is updated with time. Through the memory and storage of historical information, the LSTM prediction model can meet the requirements of load sequence continuity and the influence of influence factors on load delay. Through the calculation of seven related variables related to the load level, five key factors affecting the load level were selected by grey correlation degree, and these factors were used as model input variables to participate in the model training and prediction. Compared with the prediction results of the BP neural network model and support vector machine model optimized by particle swarm optimization, the superiority and accuracy of the model prediction were verified.

(2) Based on the analysis of the countries and regions that have entered the saturation stage, this paper summarizes the judgment basis of the indicators and auxiliary indicators that have entered the saturation stage, and based on the revised rules, it determines that Hebei Province has entered the saturation stage.

(3) Based on LSTM prediction model, this paper sets up four different economic and social development scenarios to forecast the electricity consumption of Hebei Province from 2020 to 2050. Through comprehensive analysis, it is concluded that the province will enter the saturation stage from 2031 to 2033, with the scale ranging from 4910.3 billion kwh to 5642.8 billion kwh.

References

[1] Guan Meng. Research on saturated load forecasting of Xiamen Power Grid [D]. Fuzhou University, 2017. (in Chinese).
[2] He Miao. Research on multi factor saturated load forecasting of power system [D]. Shenyang Institute of engineering, 2018. (in Chinese).
[3] He Miao, Ye Peng, He Ningning. Review of saturated load forecasting methods [J]. Shandong industrial technology, 2017 (23): 217. (in Chinese).
[4] Huang Bihe, Guo Weijiong, Liu Tao. Development goals and strategic measures of DSM in Hebei Province during the 13th five years plan [J]. DSM, 2017, 19 (03): 1-4 + 20.
[5] Ma Huiping. Load forecasting of Shijiazhuang urban power grid based on spatial load forecasting method [D]. North China Electric Power University, 2018. (in Chinese).
[6] Bastian Dietrich, Jessica Walther, Matthias Weigold, Eberhard Abele. Machine learning based very short term load forecasting of machine tools[J]. Applied Energy, 2020, 276.
[7] Li Qian. Discussion on urban saturated load forecasting method [J]. Electrical technology, 2020 (05): 6-7. (in Chinese).
[8] He Miao, Ye Peng, he Ningning, Li Shan. A review on saturated load forecasting of power system [J]. Journal of Shenyang Institute of Engineering (NATURAL SCIENCE EDITION), 2017, 13 (04): 340-346. (in Chinese).
[9] Zhang Jianping, Liu Jiefeng, Chen Yidong, Yang Zonglin, Ma zeliang, Cheng Haozhong. Saturated load forecasting based on per capita electricity consumption and per capita power load [J]. East China electric power, 2014, 42 (04): 661-664. (in Chinese).
[10] Shen Hao, Lu Junyan, Yu Jie, Shi Jinxiao, Liu Jianqing, He Yangmu. Load forecasting method
based on spatial saturated load density [J]. Electrical automation, 2018,40 (03): 57-59 + 104. (in Chinese).

[11] Ding Heng, Tong Yafang, Guo Sen, Zhou Xiaobing. Prediction of saturated power demand in Hubei based on logistic model [J]. Hubei electric power, 2015,39 (08): 31-34. (in Chinese).

[12] Yang Tialhua, Chen Yin. Power demand saturation prediction based on modified adaptive logistic model [J]. China electric power, 2017,50 (05): 114-120. (in Chinese).

[13] Shi Xiuqian, Yu Xiaoming. Power load forecasting based on BP neural network [J]. Industrial control computer, 2020,33 (03): 21-22. (in Chinese).

[14] Xiao Bai, Zhao Xiaoning, Jiang Zhuo, Shi Yonggang, Jiao Mingxi, Wang Fu. Spatial load forecasting method based on fuzzy information granulation and support vector machine [J]. Power grid technology: 1-10 [2020-05-12].https://doi.org/10.13335/j.1000-3673.pst.2019.1463. (in Chinese).

[15] Shang Fangyi, Li Jie. Application of combined forecasting in saturated load forecasting [J]. Power and energy, 2017,38 (02): 120-125. (in Chinese).

[16] Tang Xiao. Analysis of influencing factors of saturated load and Research on prediction method [D]. Hunan University, 2016. (in Chinese).

[17] Mao Xuejiao, Tan Jian, Yao Yingbei, Li Bingjie, Wu Chen. Saturated load forecasting method and its application based on long-short term memory neural network [J]. Hydropower energy science, 2019,37 (06): 192-195 + 168. (in Chinese).

[18] Jitendra Kumar,Rimsha Goomer,Ashutosh Kumar Singh. Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters[J]. Procedia Computer Science,2018,125.