A Deep Learning Based Real-time Load Forecasting Method in Electricity Spot Market

Qipei Zhang1*, Jixiang Lu1,a, Zihong Yang 1,2,b, Mengfu Tu1,c

1Nari technology co,ltd., Nanjing 211106, China
2State Key Laboratory of Intelligent power grid protection and operation control, Nanjing 211106, China
*a Corresponding author e-mail: zhangqipei@sgepri.sgcc.com.cn
aiujixiang@sgepri.sgcc.com.cn, byangzhihong@sgepri.sgcc.com.cn,
catumengfu@sgepri.sgcc.com.cn

Abstract. This paper analyzes the potential influence in Chinese electricity market due to the reform and access of the electricity spot market. On this occasion, a deep learning based model for load forecasting is proposed to improve the market operator's precise scheduling level and assist power retailers in managing bid strategies. Long-Short Term Memory (LSTM) unit is used to modeling, which is one of the most popular techniques of deep learning. In addition, historical power load data and meteorological data of Suzhou and Lianyungang in China from January 2015 to December 2017 are used for the study to training and evaluate forecasting model. As a result, this paper shows the compare results with exiting machine algorithm for load forecasting.

1. Introduction
As the fundamental work of the electricity market, load forecasting plays an increasingly important role with the electricity market develops. It not only becomes the basis for power plant quotation, but also an important premise for ensuring the safe and stable operation of the power grid. Its forecasting accuracy directly affects the benefits of power plant, market operator and other market participants.

In general, the electricity spot market can be divided into two parts: the day ahead market and the real-time market. The quotation and market-clearing price of the day ahead market can not only be used as a price signal for the entire market, but also the main basis for the formulation of following day’s schedule. The market operators will dispatch the contribution of power unit on the second day in advance based on the trading results of the day ahead market. Therefore, those units with long start-up times can be started in time to ensure timely supply of electricity [1].

In the electricity market, market operators will always monitor the real-time balance of power supply and load demand. Even if the generation schedule of the following day is formulated based on the day ahead market transactions, there still are deviation from actual load. Therefore, real-time market and intraday rolling schedule are needed to keep the power demand-supply balance. Obviously, the day ahead load forecasting and hourly forecasting cannot meet the requirements of the intraday rolling power generation schedule [2]. An efficient and reliable real-time load forecasting method with periods of 15min and 5min are urgently needed for electricity spot market.

Recently, machine learning technology has become a research hotspot in various academic areas. In particular, deep learning has been widely used in language models [3], machine translation [4], speech recognition [5] and has achieved great success. Load forecasting is a regression problem in
nature, there are some scholars use machine learning algorithms to solve this problem. For example, in [6], a least squares support vector machines (LS-SVM) method is used to forecast short-term load, in [7] authors proposal a LSTM network based model for short-term residential load forecasting. In addition, the types of power collection data and meteorological data are increasingly abundant, which provides the possibility to further improve the accuracy of short-term load forecasting.

In this paper, it is aim to make contributions to improve the short-term load forecasting accuracy of electricity spot market. First, various types of load data and meteorological data are analyzed and processed to extract corresponding features. Then these features are coupling and vectorizing to represent power load with high dimensions. This representation of power load in vector space creates a new time series data which features multi-source and heterogeneous spatial. The LSTM unit is used to construct the neural network model, and Early stop and Dropout are used for preventing overfitting while training model. Besides, Random Forest is served as benchmark. In case study, our experiments verify those effective features among numerous raw data and figure out a valid network structure.

2. Literature study and contributions

2.1. Literature study
The power load data features time series and non-linearity. According to its features, numerous researches on short-term load forecasting models are carried from past and present. Among many of these literatures, there are some classical time series forecasting model. For example, regression analysis method [8], exponential smoothing model method [9], Kalman filtering method [10], multiple linear regression method [11] and Fourier expansion method model [12]; representative model is autoregressive integral moving average model [13, 14] (Autoregressive Integrated Moving Average Model, ARIMA). Its basic idea is to treat the electrical load sequence as a random sequence, and a certain mathematical model is used to approximate the sequence. This model forecasts electrical load from the historical data and owns advantage of considering the time-series relationship of data. However, the nonlinear relationship of data still cannot be expressed by model.

Machine learning method is also be used for load forecasting. For example, BP neural network is used to forecast electrical load in [15-17], and grey projection and random forest algorithm are used in [18]. Literature [19] uses the gradient descent decision tree GBDT for forecast. In [20], authors process the feature of data by PCA, and then the support vector machine is used to figure out the regression. These algorithms are short of consideration of temporal correlation of electrical data, and need to artificially add time features to ensure the precision of forecast.

In reference [21], experimental data demonstrates the error based on LSTM power load forecasting is lower than that of the traditional model ARIMA. However, since only the load data is used as feature, other factors such as meteorological data are not considered, the performance of this model dose not reach the state of art. LSTM unit is used in short-term load forecasting in reference [22], but the training and test data are limited with half a year and the forecast period is half hour, its generalization has not been proved.

2.2. Contributions
To improve the accuracy of load forecasting by LSTM units and give full play to the advantages of deep learning neural networks, this paper combines the prior knowledge and experimental results of load forecasting to extract features from meteorological information, day types and electricity prices to enhance the performance model. Besides, a vectorization method of multi-source data is proposed to fast processing multiple data for implementing load forecasting model.
significantly improvement, like in natural language processing [23] and body and gesture perception [24].

In this paper, we summarize three main types of influencing factors used in traditional load forecasting methods: historical load information, meteorological information, and day type information, and extract features from these data for our model.

For characteristic of the load trend, there is one method choosing the load on the day of last year, plus the load accumulation from last year to this year, as the trend feature for model. Obviously, this method uses load data with large span and long period, which cannot obtain an accurate load trend and is not suitable for short-term load forecasting. We use the load of the last few days as parameters and use this data as input to the model for each forecast. In fact, the load trend chosen in this way contains the impact of weather and residential electricity habits on the load.

Meteorological factors have a crucial influence on short-term load forecasting. Among them, the common influencing factor is temperature, followed by humidity, wind force, and weather type. Extreme weathers lead to rapid changes in load. Therefore, multiple weather data are used to building the model.

The type of date is another important factor in the short-term load forecast. In fact, the current urban electricity load in China is still dominated by industrial electricity loads. And there is a significant electricity load decline between workday and holiday. This paper takes date type account into the forecasting of load. All the time series data are tagged with date labels.

In addition, the economic adjustment factors of electricity prices are also considered as load influencing factors. Since the industrial electrical load is about 90% of the total load, in order to normalize the model parameters, number 1, 2 and 3 are used to represent the price of the valley, flat and peak ratio.

4. Model structure and feature extraction

4.1. Model structure
Recurrent Neural Network (RNN) is fundamentally different from traditional feedforward neural networks. It is a sequence-based model that establishes the temporal correlation between previous information and the current situation. For the time series problem, which means the decision made by RNN at time T-1 will affect the decision of its current time T. As a result, RNN is suitable for processing and predicting the interval and delay events in the time series [25]. Based on daily experience, people's daily electricity habits are one of the important factors affecting the load of same period, which is exactly in line with the predictive characteristics of RNN.

The Long-Short Term Memory (LSTM) is an improved RNN that is first proposed by Hochreiter et al. [26] and is modified by Alex Graves to add additional gates of forgetting [27]. The improved LSTM solves the problem of “gradient disappearance” while training model. It features learning dependence information of long-term and short-term from time series. LSTM is one of the most successful RNN architecture at present, and it is widely used in many scenarios.

![Basic LSTM cell](image)

Figure 1. Basic LSTM cell
The LSTM basic cell includes the Forgetting Gate, Input Gate and Output Gate [27], as shown in Figure 1. In the forgetting gate, the input $x_t$, the status memory unit $S_{t-1}$, and the output $h_{t-1}$ jointly determine the forgotten portion of the status memory unit. The $x_t$ in the input gate determines retention vector of state memory unit after sigmoid and tanh conversion. The final output $h_t$ is determined by the updated $S_t$ and the input $x_t$. The calculation formula is shown in (1)-(6), where $W_{fs}, W_{fh}, W_{lx}, W_{ln}, W_{fx}, W_{fh}, W_{ox}$ and $W_{oh}$ are the input weight matrices of corresponding activation functions.

\[
\begin{align*}
    f_t &= \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \\
    i_t &= \sigma(W_{lx}x_t + W_{lh}h_{t-1} + b_i) \\
    g_t &= \varnothing(W_{gx}x_t + W_{gh}h_{t-1} + b_g) \\
    o_t &= \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \\
    s_t &= g_t \odot i_t + s_{t-1} \odot f_t \\
    h_t &= \varnothing(s_t) \odot o_t
\end{align*}
\]

4.2. Training skills
Deep neural networks contain many non-linear hidden layers, which allow the network to learn the extremely complex relationships between input and output. In the case of insufficient data, deep neural networks will experience over-learning and over-fitting. Solutions to overfitting usually include Early stop, which stops training as soon as the performance of the validation begins to deteriorate, and L1, L2 regularization, light weight distribution, etc. [28].

Nitish Srivastava and Geoffrey Hinton et al. proposed Dropout method achieves the performance closes to Bayesian gold standard by learning a cluster of parameter-sharing sets with exponential number of models. The term Dropout means to discard part of the neuron cells in a neural network. These discarded cells are temporarily removed from the network, and all its input and output connections while training.

The probability $P$ of retaining neurons in the same layer is same for all neurons when implementing Dropout. Besides, there are no interaction between every single neuron. Even though, the probability of each neuron participating in each training is $p$, it always presents during testing [29]. This paper adopts both Early stop and Dropout methods to avoid overfitting while training the model.

4.3. Feature extraction
The short-term load forecast of the Chinese state grid company requires to forecast the local total load on the following day. The forecast frequency is every 15 minutes, so the load at 96 different time points during the day needs to be predicted. The electricity spot market intraday trading periods are usually 5 minutes and 15 minutes. Therefore, every 5 minutes and 15 minutes load data per day are important parts of our training data. Weather conditions are usually considered to be closely related to electricity load. As a result, this paper also chooses six types of weather information for training data including wind speed, wind direction, rainfall, temperature, humidity, and air pressure.

To find out the impact of the date type on load varying, we select one week's load data with similar weather and draw the trend graph, as shown in Figure 2.
The experimental results indicate that the electricity load on weekends and holidays is significantly lower than the working day. Therefore, in the actual model training, we divide the date features into working days and holidays respectively.

After the experimental test, all features are selected from the original load data, meteorological data and peak-valley electricity price factors as the input of model training, and the working days and holidays are separately trained. The detail feature selection is shown in Table 1.

![Figure 2. Typical one-week load](image)

| Features          | Description                                          |
|-------------------|------------------------------------------------------|
| L(d-1, t)         | The load of 1 day ago, t time                        |
| L(d-2, t)         | The load of 2 days ago, t time                       |
| L(d-3, t)         | The load of 3 days ago, t time                       |
| L(d-4, t)         | The load of 4 days ago, t time                       |
| L(d-5, t)         | The load of 5 days ago, t time                       |
| L(d-6, t)         | The load of 6 days ago, t time                       |
| L(year-1, t)      | The load of 1 year ago, t time                       |
| Workday           | workday                                              |
| Holiday           | weekend and other holiday                            |
| Temperature       | Forecast of temperature, t time                      |
| Humid             | Forecast of humid, t time                            |
| Wind Speed        | Forecast of wind speed, t time                       |
| Wind Direction    | Forecast of wind direction, t time, Angle based     |
| Precipitation     | Forecast of precipitation possibility, t time        |
| Pressure Peak     | Forecast of air pressure, t time                     |
| Peak flat         | Flat electricity price                               |
| Valley            | Valley electricity price                             |

Inspired by natural language processing tasks of word to vector, we coupling all the corresponding features to construct a new time series. Every single power load data is mapping to a vector space which including above features. Finally, the input vector of our power data is present as shown in Figure 3. The Day-1 in the following figure refers to the power load of one day ahead, and Day-2, Day-3, Day-4, Day-5 and Day-6 refer to the power load of corresponding days ago.
5. Case study

5.1. Data standardization

In this experiment, min-max standardization is used to linearly transform the original data, and the data size is constrained between [0, 1]. The formula is defined as follow:

\[ x^* = \frac{x - \text{min}}{\text{max} - \text{min}} \]  \hspace{1cm} (7)

Where \( x^* \) is the transferred value, max is the maximum value in the sample data, and min is the minimum value in the sample data.

5.2. Experimental evaluation index

To assess the performance of the model, our experiment refers to the Chinese National Grid Corporation's evaluation of load forecasting indicators: mean absolute percentage error (MAPE), root mean square error (RMSE) and forecast accuracy (FA). The corresponding formulation is written as follows:

\[ \text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|x_{\text{act}}(i) - x_{\text{pred}}(i)|}{x_{\text{act}}(i)} \]  \hspace{1cm} (8)

\[ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_{\text{act}}(i) - x_{\text{pred}}(i))^2} \]  \hspace{1cm} (9)

\[ \text{FA} = \left( 1 - \frac{|x_{\text{act}}(i) - x_{\text{pred}}(i)|}{x_{\text{act}}(i)} \right) \times 100\% \]  \hspace{1cm} (10)

5.3. Results

Our experiments test the day ahead load forecasting and intraday load forecasting of 5min and 15min periods from hyperparameters and influence features of model. Besides, the input length of load time series has been tested in day ahead load forecasting. These tests adopt the control variable method to carry out stepwise tuning.

In the day ahead load forecasting, we adjust the input load time series length for error comparison test under the condition of fixed training epochs, batch size and input features. Table 2 shows the test result that the best performance comes out with input length 96 which is the same length of load needs to forecast within a day.

| Table 2 Model parameters combination test result |
|-----------------------------------------------|
| Model parameters combination test             |
| input sequence | epoch | Batch size | features | MAPE  |
| 96           | 500   | 512        | 9        | 0.0299 |
| 48           | 500   | 512        | 9        | 0.0341 |
| 24           | 500   | 512        | 9        | 0.0335 |
| 12           | 500   | 512        | 9        | 0.0338 |
| 6            | 500   | 512        | 9        | 0.0339 |
| 4            | 500   | 512        | 9        | 0.0342 |
In order to further analyze and verify the effect of the selected impact features on the forecasting, meteorological information, historical load and date type are selected for test. The test results present at Table 3 indicate both meteorological information, date type and historical load data have a positive correlation contribution to the test results. Similarly, forecasting workdays and holidays separately can significantly improve accuracy.

| Feature factor combination test result |   
|--------------------------------------|
| **Feature factor combination test**   |   
| Meteorological data load data date type epoch batch size MAPE |   
| Meteorological data workday/holiday 20 512 0.0385 |   
| history load workday/holiday 20 512 0.0349 |   
| Meteorological data history load workday/holiday 20 512 0.0369 |   
| Meteorological data history load workday and holiday 20 512 0.0337 |   

Finally, keeping other parameters fixed, and increasing the number of layers of LSTM to test the effect of increasing the depth of the model. Result shows that increasing the model depth by stacking the LSTM layer can improve the model prediction ability. However, test loss increases when the LSTM layer is increased to 5 layers in our case. The detail test result is shown in Table 4.

| Model structure combination test result |   
|----------------------------------------|
| **Model structure combination test**    |   
| LSTM Layers epoch Batch size features MAPE |   
| Two layers 500 512 16 0.0282 |   
| Three layers 500 512 16 0.0274 |   
| Four layers 500 512 16 0.0265 |   
| Five layers 500 512 16 0.0297 |   

The above experimental results determine the parameters, influence features and structure of the model. Then we compare the model with the benchmark of random forest. Figure 4 presents the forecasting result of day ahead load by our model and random forest, our proposed model gains the better performance.

![Figure 4. Forecasting curve and actual curve comparison](image)

We use the bus load of a 220kv substation at one city of Jiangsu province as test data to test the performance of the LSTM and random forest models in the ultra-short-term load forecast with periods of 5 minutes and 15 minutes. As the results shown in Figure 5, both models have achieved good performance, while our proposed model gain a higher forecast accuracy.
6. Conclusion
In this paper, we explore the relationship between electricity market and short-term load forecasting, and discuss different machine learning model for load forecasting. The influence features have been found and proved by experiment. Finally, the results show that the proposed model can accurately forecast the daily load and the real-time load in the power spot market. Such short-term load forecasting can not only extract features in the historical data, but also use the current time data to correct the load forecast at the next period. This model can apply to regional-wide node load forecasting, and in the future can assist the power retailer's quotation strategy and the power operator's node price setting. Obviously, with the reform of the electricity market, the integration of electric vehicles and renewable power generation, the uncertainty of power load will continue to increase. In the future work, these uncertain factors can be specifically analyzed and integrated into the neural network.

References
[1] Zhou JingYang, Wang Bin, Zhou JieYing, et al. Applied Model and Analysis of Dispatching Plan in Smart Grid Dispatching and Control Systems Under Market Mechanism[J]. Automation of Electric Power Systems, 2015, 39(1):124-130.
[2] Kang Chongqing, Xia Qing, Hu Zuohao, et al. New Connotation of Forecasting Issues in Electricity Market[J]. Automation of Electric Power Systems, 2004, 28(18):1-6.
[3] Sundermeyer M, Schluter R, Ney H, et al. LSTM Neural Networks for Language Modeling.[C]. conference of the international speech communication association, 2012: 194-197.
[4] Ilya Sutskever, Oriol Vinyals, Quoc V. Le.Sequence to Sequence Learning with Neural Networks. In Arxiv preprint arXiv:1409.3215v3, 2014
[5] Medennikov I, Bulusheva A. LSTM-Based Language Models for Spontaneous Speech Recognition[C]. international conference on speech and computer, 2016: 469-475.
[6] Huang YuanSheng, Deng JiaJia, Yuan ZhenZhen, SVM short-term forecasting based on ARMA error calibration and the adaptive particle swarm optimization[J]. Power System Protection and Control, 2011, 39(14):26-32.
[7] Kong W, Dong Z Y, Jia Y, et al. Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network[J]. IEEE Transactions on Smart Grid, 2017, PP(99):1-1.
[8] Mahmud M A. Isolated Area Load Forecasting using Linear Regression Analysis: Practical Approach[J]. Energy and Power Engineering, 2011, 03(04): 547-550.
[9] Ostertagova, Eva & Ostertag, Oskar. Forecasting Using Simple Exponential Smoothing Method[J]. Acta Electrotechnica et Informatica,2012, 12. 62–66. 10.2478/v10198-012-0034-2.
[10] Zheng T, Girgis A A, Makram E B, et al. A hybrid wavelet-Kalman filter method for load forecasting[J]. Electric Power Systems Research, 2000, 54(1): 11-17.
[11] Wang Yong, Huang Guoxin, Peng Daogang, Application of multiple linear regression method with feedback in power load forecasting[J]. Computer Application and Software, 2008(01):82-84.
[12] Beiraghi M, Ranjbar A M. Discrete Fourier Transform Based Approach to Forecast Monthly Peak Load[C]. Ieee pes asia-pacific power and energy engineering conference, 2011: 1-5.
[13] Cho M Y, Hwang J C, Chen C S, et al. Customer short term load forecasting by using ARIMA transfer function model[C]. International conference on energy management and power delivery, 1995: 317-322.
[14] Hor C, Watson S J, Majithia S, et al. Daily Load Forecasting and Maximum Demand Estimation using ARIMA and GARCH[C]. Ieee international conference on probabilistic methods applied to power systems, 2006: 1-6
[15] ZHOU Ying YIN Bang-de REN Ling BIAN Xue-fen, Study of Electricity Short-term Load Forecast Based on BP Neural Network[J]. Electrical Measurement & Instrumentation, 2011,48(02):68-71.
[16] SU Xueneng, LIU Tianqi, CAO Hongqian, et al. A Multiple Distributed BP Neural Networks Approach for Short-term Load Forecasting Based on Hadoop Framework[J], Proceedings of the CSEE,2017,37(17):4966-4973+5216.
[17] Zhu Haibing, Cui Yu, Xiong Hao. Load forecasting based on improved BP neural network[J]. Modern Electronics Technique, 2016,39(20):64-66.
[18] Xiaooyu W U, Jinghan H E, Pei Z, et al. Power System Short-term Load Forecasting Based on Improved Random Forest with Grey Relation Projection[J]. Automation of Electric Power Systems, 2015, 39(12):50-55.
[19] Zheng Kaiwen, Yang Cao. Short-term load forecasting research based on Gradient Boosting Decision Tree(GBDT)[J]. Gui Zhou Electronic Technology, 2017, 20(2):82-84.
[20] Bao Quanqing, Lin Qilin, Wu Guodong. Short-term load forecasting based on meteorological factors and date attributes [J]. Electrical applications, 2017,36 (11): 46-51
[21] Chen Zhuo, Sun Long-xiang. Short-Term Electrical Load Forecasting Based on Deep Learning LSTM Networks[J].Electronic Technology. 2018(1).
[22] Chen L, Wang Z, Wang G. Application of LSTM Networks in Short-Term Power Load Forecasting Under the Deep Learning Framework[J]. Electric Power Information & Communication Technology, 2017.
[23] Bei Chao, Hu Bo. Influence of language prior knowledge on Natural language processing task of neural network model [J]. Chinese Journal of Information, 2017,31 (06): 10-17.
[24] Du yu. Research on the key techniques of posture and gesture perception based on depth machine learning [D]. Zhejiang University, 2017.
[25] Kong W, Dong Z Y, Jia Y, et al. Short-Term Residential Load Forecasting based on LSTM Recurrent Neural Network[J]. IEEE Transactions on Smart Grid, 2017, PP(99):1-1
[26] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
[27] Graves A, Mohamed A, Hinton G. Speech Recognition with Deep Recurrent Neural Networks[J]. 2013, 38(2003):6645-6649.
[28] S. J. Nowlan and G. E. Hinton. Simplifying neural networks by soft weight-sharing. Neural Computation, 4(4), 1992.
[29] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1):1929-1958.