



Short-Term Load Forecasting Based on Deep Neural Networks Using LSTM Layer

Bo-Sung Kwon¹ · Rae-Jun Park¹ · Kyung-Bin Song¹

Received: 7 May 2019 / Revised: 2 February 2020 / Accepted: 1 April 2020 / Published online: 11 May 2020
© The Korean Institute of Electrical Engineers 2020

Abstract

Short-term load forecasting (STLF) is essential for power system operation. STLF based on deep neural network using LSTM layer is proposed. In order to apply the forecasting method to STLF, the input features are separated into historical and prediction data. Historical data are input to long short-term memory (LSTM) layer to model the relationships between past observed data. The outputs of the LSTM layer are incorporated with outputs of fully-connected layer in which prediction data, for instance weather information for forecasting day, are input. The optimal parameters of the proposed forecasting method are selected following several experiment. The proposed method is expected to contribute to stable power system operation by providing a precise load forecasting.

Keywords Short-term load forecasting · Deep neural networks · Long short-term memory

1 Introduction

Load forecasting has attracted much attention in research on power systems. It enables utility suppliers to forecast loads to preserve a balance between supply and demand, reduce electricity production costs, and manage operation scheduling and future capacity planning. Mocanu et al. [1] classified load forecasting methods into three categories according to forecast range: short-term from 1 h to 1 week, mid-term from 1 week to 1 year, and long-term 1 year or more. Short-term load forecasting (STLF) is important for managing power markets and for power system operations such as unit commitment and economic dispatch. Therefore, STLF errors could interfere with reliable power system operation and cause economic losses [2].

Load forecasting methods can be classified as time series method or causal method. In time series method, the load is modeled as a function of past observed values.

Auto-regressive and exponential smoothing methods are common time-series forecasting methods, and these methods usually use time delay to forecast based on past observations. Choosing the appropriate time delay is an important step in eliminating redundant features, and time series method should select the time delay range of the previous input variables to use in model construction. This allows forecasters to increase the accuracy of the method and better understand the underlying process of time-series data [3–5].

Meanwhile, in causal method, the load is modeled as a function of some external factors, especially weather and social variables. The most common causal method is multiple linear regression (MLR). MLR is attractive because forecasters can attach physical interpretations to the MLR's components. However, MLR is in essence linear method, whereas loads are known as non-linear functions of external factors [6–8]. In addition, authors of some studies have applied other forecasting methods such as non-parametric regression [9], multiplicative auto-regression [10], Kalman Filter [11] and ARMAX [12].

Recently, many researchers have applied artificial intelligence techniques to load forecasting, including fuzzy linear regression, random forest, support vector machine [13, 14], but the method that has received the most attention is neural networks (NNs). NNs have been widely used for forecast tasks because they can model non-linearity. The basic structure of NNs, multilayer perceptron, was used to

✉ Kyung-Bin Song
kbsong@ssu.ac.kr

Bo-Sung Kwon
bosung1994@naver.com

Rae-Jun Park
rejuni@ssu.ac.kr

¹ Department of Electrical Engineering, Soongsil University, Seoul, South Korea

forecast loads using previous data [15, 16]. Various NNs structures have been applied to improve the forecasting accuracy [17–20]. Deep neural networks (DNNs) are NNs with more than one hidden layer. The multiple layers improve the feature abstraction of network, allowing for efficient learning of complex and non-linear relations [21]. DNNs are reported to produce improved performance over shallow NNs [1]. But from an intuitive point of view, it is clear that the nature of load is dynamic, not static. Changes in load are affected not only by external factors but also by past and current load conditions. In this sense, the static neural network is a lane load forecasting method, and RNNs are proposed to integrate previous load state information to current load state [22]. Additionally, LSTM is a variation of RNNs that was originally developed by Hochreiter et al. [23] to allow the preserving the weights that are forward and back-propagated through layers. Salah et al. [24] employed LSTM model for load forecasting that outperformed other machine learning approaches. Zheng et al. [25] developed a hybrid LSTM model using Xgboost and k-means. Xgboost is used to determine the feature selection and k-means is used to merge similar days into one cluster. Because load is fluctuated various non-linear factors such as social, economic, and weather factors and have time-series characteristics, various load forecasting methods have been proposed to reflect these characteristics [26]. But researchers still emphasize the need for more accurate and reliable load forecasting methods. In particular, STLF has become increasingly important in modern power systems since the rise of solar PV and wind power, through which the output is intermittent according to the weather conditions [27].

The aim of this paper is to contribute to addressing the issues related to day-ahead load forecasting. Load forecasting should reflect both the time-series characteristics of loads, and the non-linear correlations of load fluctuation factors. Day-ahead load forecasting also requires prediction data such as weather information and day of week of a given forecasting day. To address the challenge of filling this need, deep neural networks based forecasting method that can learn and extract rich features from the input is proposed. In the proposed method, long short-term memory (LSTM) layer with a fully-connected (FC) layers are combined. These receive data with different characteristics as input. LSTM layer based on recurrent neural networks (RNNs) are used to model the variability and dynamics from historical data. FC layers are used to project prediction data and form the relationship with the output of the layer.

The proposed method is evaluated respectively based on load forecasting accuracy for the total load of Korea power system in 2017 and 2018. Additionally, to objectively verify the performance of the proposed method, KPX short-term load forecasting (KSLF), which is the short-term load forecasting tool to forecast day-ahead load in Korea, is compared

with proposed method [2]. The forecast results based on the total load data of Korea power system show that the proposed forecasting method based on deep neural networks using LSTM layer outperforms alternative approaches with high accuracy.

The remainder of this paper is organized as follows: Sect. 2 provides an overview of the literature on load forecasting. Section 3 provides brief background on LSTM-based RNNs. Section 4 describes the methodology of the proposed method. Section 5 describes case study results and provides a discussion on validity of the results. Section 6 draws conclusions.

2 Long Short-Term Memory Based on Recurrent Neural Networks

In this section, the brief backgrounds on LSTM based on RNNs are provided. RNNs are special NNs that are connected in a feedback structure between units of an individual layer. They are called recurrent because they perform the same operation on all elements in the sequence. They make it possible to model data with time-series characteristics by complementing the limits of non-recurrent NNs, which independently assumes the relationship between inputs. The rolled and unrolled RNNs configurations are as follows Fig. 1.

In the Fig. 1, A is RNNs cell and the x_t is the sequence input $\{x_1, x_2, \dots, x_n\}$ at time t and h_t is the output of the cell. And h_t used as input at time $t + 1$. RNNs are able to store memory because the current output of neurons depends on the previous computations. The output of cell, h_t , is calculated to use recurrence function f , expressed as follows (1). Therefore, RNNs are algorithm that exhibit excellent performance when the input has time-series characteristics such as sequential data [28].

$$h_t = f(h_{t-1}, x_t) \quad (1)$$

However, RNNs have a gradient vanishing problem. It means that the weights propagated forward and backward

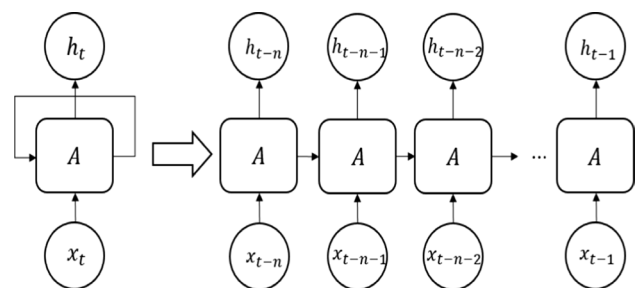


Fig. 1 The rolled and unrolled RNNs configurations [28]

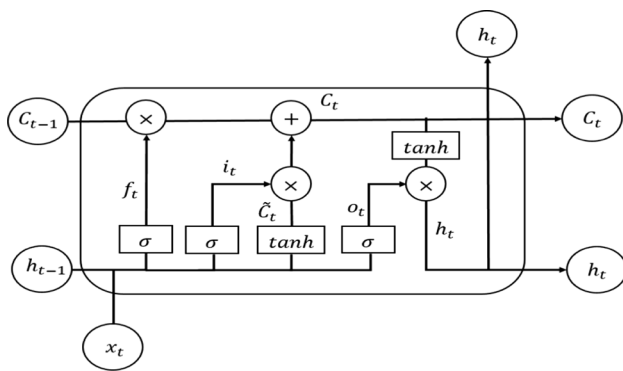


Fig. 2 The structure of LSTM [31]

through the layer decrease, and thus, the long-range dependencies cannot be preserved [29]. LSTM is proposed to overcome this gradient vanishing problem by introducing cell state and gates into RNN cells. In LSTM, the cell state and the three gates are added as follows Fig. 2. Using these gates allows preserving weights propagated through time and layers, LSTM can solve the long-range dependencies problems [30].

The output of LSTM is computed as follows (2)–(7):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{3}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{6}$$

$$h_t = o_t * \tanh(C_t) \tag{7}$$

where, f_t , i_t and o_t are the forget, input and output gates respectively. The forget gate f_t determines how much the previous state is reflected in the current state. The input gate i_t determines the new information to update the cell state. The output gate o_t determines the information to output based on the cell state. The sigmoid function, σ , is used to adjust the output to a value between 0 and 1 for the f_t , i_t and o_t following (2)–(4). The output of these gates depends on the value of the current input x_t and previous output h_{t-1} . If the gate output is 0, the values are blocked by the gate, and if it is 1, the values are stored. C_t is the cell state and \tilde{C}_t is the value for calculating the current cell state C_t . \tilde{C}_t is calculated following (5), and C_t serves as an accumulator of the state information. It is updated from previous cell state C_{t-1} following (6). h_t is the output of current LSTM following (7).

The C_t and h_t are used as input of the next time step. This process is repeated every time step. W and b are weights and biases of the LSTM cell. They are updated to a value that minimizes the differences between the output of LSTM and the target value through the training process.

3 Methodology

In order to forecast load accurately, it is necessary to identify and reflect the factors that influenced load. The factors that affect the load fluctuation include weather conditions, social events, etc. These factors have a non-linear relationship with load, and thus, it is difficult for load forecasts to reflect the effects of all such factors. Furthermore, load has strong daily, weekly, monthly time-series characteristics, and load forecasts should reflect these all characteristics. Therefore, forecasting method based on deep neural network using LSTM layer for STLF is proposed that reflects both the time-series trends of load itself and the non-linear correlation of load fluctuation factors. The output of proposed method is the hourly load for the next day.

In this section, the workflow for STLF is described in detail. The process follows six steps as shown in Fig. 3. In the following, detailed task of each steps are described.

3.1 Data Selection

3.1.1 Historical Load

The hourly loads have significant approximate periodicity. This is because people consume electricity as part of their regular production activities and lifestyles. The proposed method uses historical load as input because it can reflect trends of load [2]. Prior to using the historical loads, a time-series exploratory analysis of the load should be performed. This can be useful in identifying load trends, patterns and anomalies. For this study, the Korean nationwide load data

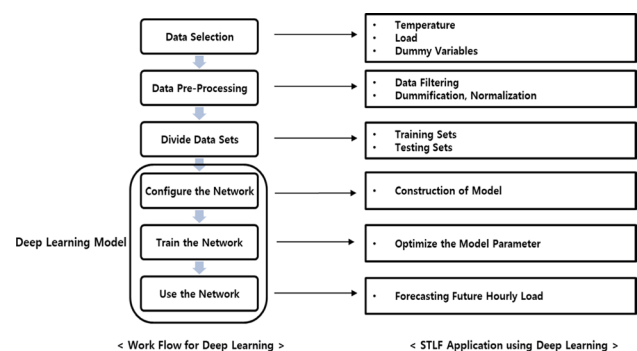


Fig. 3 The workflow for the proposed forecasting method

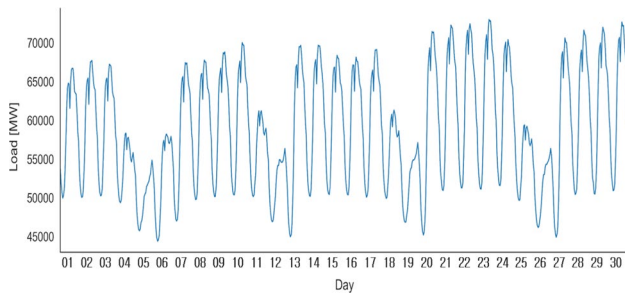


Fig. 4 Hourly load in June 2016

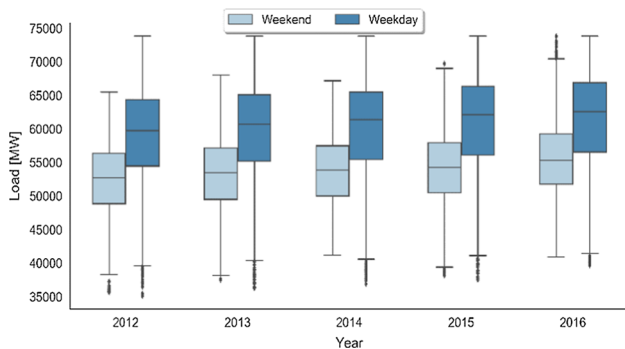


Fig. 5 Comparison between load of weekend and load of weekdays

set is used, and these data include hourly load. The Fig. 4 shows the hourly load for June of 2016.

In the Fig. 4, the loads have a similar pattern every week. This is because in general, load is closely related to the life patterns of modern society. Because electricity consumption for industrial and commercial activities is low on weekends, the weekend electricity usage patterns generally differ from the weekday patterns. And because the load is lower on weekends than on weekdays, the load on Monday morning is lower than that on other weekdays. Figure 5 shows a box plot of weekday and weekend load. In all years, the weekdays loads are higher. Because the load patterns vary by Monday, Tuesday to Friday, Saturday, and Sunday, the load patterns are classified as day type according to daily load pattern. And this features should be reflected in the STLFL.

People's behavior patterns on public holidays, especially long holidays, generally differ from their normal daily patterns, and this affects load. Therefore, the data for public holidays are filtered from inputs [13]. In addition, the average hourly load of the latest 2 days which the day type is same with the type of each input day is used as input.

3.1.2 Temperature

The weather is known to have a great influence on electricity usage because it affects people's behavior.

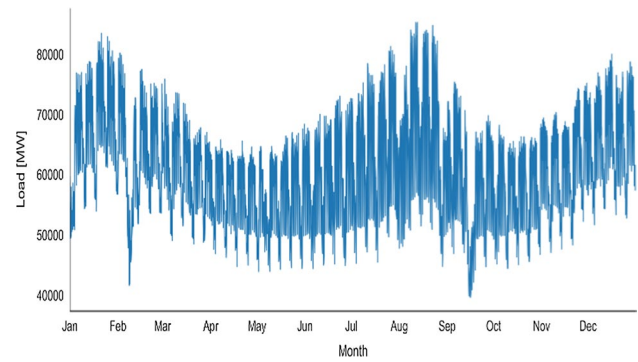


Fig. 6 The hourly load from January to December in 2016

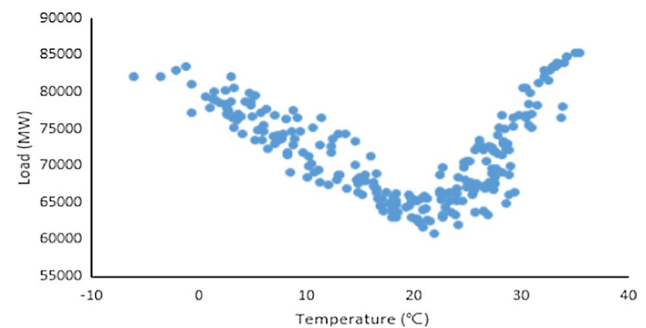


Fig. 7 The relationship of temperature and load

People tend to use more electricity when temperatures are uncomfortable, whether too hot or too cold, and to do more activities indoors. The hourly load from January to December in 2016 is shown in Fig. 6. It shows patterns related to seasonal human activities. In summer (June to August) and winter (December–February), the loads are higher due to the effect of temperature than in other month.

The effect of load by temperature in 2016 is shown non-linearly as follows Fig. 7. Here, the temperatures are weighted averages obtained by multiplying the temperature of the eight major cities in Korea by the regional weights of the eight cities [32].

The correlation coefficient between the hourly load and the hourly temperatures is 0.73, which is the yearly average of the absolute values of the correlation coefficient calculated monthly. The load correlated highly with the temperature and have non-linear relationship. Additionally, temperature is more accurately measured than are other meteorological factors. Therefore, the hourly temperatures are employed as input [33]. In the future research, we will review about the effects of other meteorological factor on load.

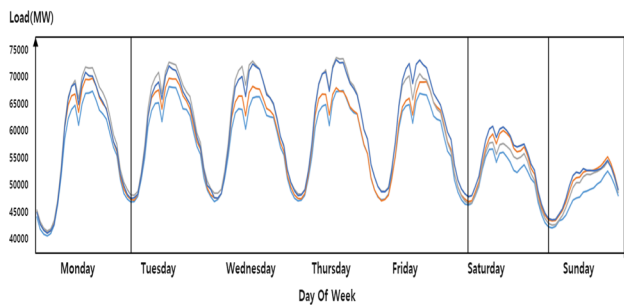


Fig. 8 Patterns of load by day of week

3.1.3 Dummy Values

As shown in Fig. 8, the load has similar load patterns for each day of week. Therefore, time and day of the week index can be useful for forecasting hourly load. So these 2 types of features are included as input in forecasting method [27, 34]. Because a day is 24 h, an incremental sequence from 1 to 24 is used as a time index. And number between 0 and 3 is specified for each day of week types to represent categorical features. Here, 0 is specified for Monday, 1 for Tuesday through Friday, 2 for Saturday, and 3 for Sunday.

3.2 Input Feature Pre-processing

Data pre-processing is an important step to obtain better performance and accuracy of neural networks. Because neural networks are sensitive to data scale, the inputs have to be normalized before they can be used in neural networks. In the load forecasting, because the difference in scale between the load and temperature is large, accurate training cannot be performed when inputs are used without being normalized. After reviewing some normalization methods, min–max normalization is selected through analysis [35]. The min–max normalization is used to scaling the data in the proposed method to range (0, 1).

3.3 Structure of Forecasting Method

The pre-processed data are divided into training and testing sets. The training set is data for training the neural networks and the testing set is input to the trained neural networks and used to forecast the load of forecasting day. Except for the presence of labels, the basic data structures of the training and testing set are similar. Here, the label is the target value to input for the training the neural networks, and the proposed method uses the load as label. The neural networks are trained by comparing the output of the final layer calculated from the input features with the target load.

Both training and testing set consist of two types of input. One is historical data and other is prediction data.

Here, historical data means recent past data used for load forecasting, consisting of past day of week, time of day, hourly temperature, hourly load, and the average hourly load of the latest 2 days. Prediction data are the data on target day, such as day of week, time of day, predicted hourly temperature, and the average hourly load of the latest 2 days. So appropriate NNs layers are used to learn the relationship and extract rich features from the data. As shown in Fig. 9, LSTM layer are used to extract features from the historical data. In addition to the historical data, prediction data should be incorporated into the networks. Therefore, the proposed method input the prediction data for the target day to FC layer to form relationship with target load. All the outputs from LSTM and FC layers are combined and used as input to next FC layer.

In the historical data, D is day index, H is time index, T is temperature, L is load, AL is the average hourly load of the latest 2 days. And the number of historical data points is determined by the LSTM sequence length. In the prediction data, PD is day index, PH is time index, PT is predicted temperature by meteorological office, and prediction data are the values for the target day. And FL is the forecasted load, which is the output of final layer.

The difference between outputs from final layer and target values is calculated using mean squared error, and each NNs parameters are optimized until the error reaches the proper value. The computationally efficient Adam optimizer [36] showed better results than other optimizer, including the steep gradient descent, Adagrad [37], Adadelata [38] and RMSProp [39]. Therefore, Adam optimizer is used to train the proposed method.

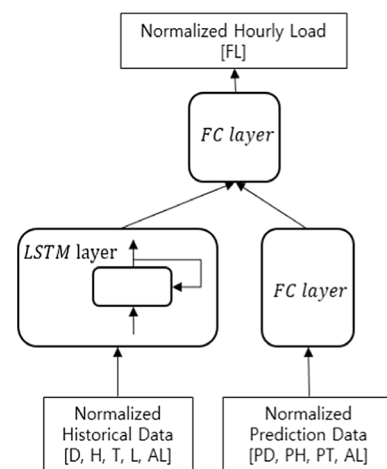


Fig. 9 The structure of proposed forecasting method

4 Case Studies

The proposed method is evaluated based on the results of load forecasting for the year 2017 and 2018. Day-ahead load forecasting is performed only the normal days except special days. The training and testing sets for the forecast are constructed using the moving window method. The Korea Power Exchange (KPX), power system operator in Korea, performs day-ahead load forecasting on previous day. Because there are no hourly data for the day on which the forecasting is performed, the data until the day before the forecast execution day is used to forecast. Therefore, the testing set for day-ahead forecasting consists of several days of data before the forecast execution day. The data from the day before the forecast execution day to 1 year ago is used as the training data. The load forecasting error rate is based on mean absolute percentage error (MAPE) as shown in (8):

$$\text{MAPE}(\%) = \frac{100}{n} \times \sum_{t=1}^n \left| \frac{L_t^{\text{Measured}} - L_t^{\text{Forecast}}}{L_t^{\text{Measured}}} \right| \quad (8)$$

where, L_t^{Measured} is load measured at time t , L_t^{Forecast} is load forecast at time t , and n is the number of time steps.

4.1 Configure the Networks

The sequence length should be determined to use LSTM layer. Here, the sequence length means the number of days of historical data used for forecast. The hyper-parameter tuning of neural networks is usually based on experience and experimentation. So, the optimal sequence length for LSTM layer is selected based experimentation. Table 1 shows the yearly average for 2016 according to LSTM sequence length. Based on the results, sequence length is selected to 3 in the proposed method.

4.2 Compare with the Results of Previous Study

In order to verify the performance of proposed method, it is necessary to compare its results with those from previous studies. To do this, the day-ahead load forecasting method, KPX Short-term Load Forecasting (KSLF), which the KPX uses is compared with proposed method. As a grid operator, KPX performs day-ahead load forecasting to operate the Korean power system. Exponential smoothing method is main scheme for day-ahead load forecasting in KSLF. It uses

Table 1 The forecasting results of yearly average MAPE for 2016 according to sequence length of LSTM layer

Sequence length	1	2	3	4
MAPE	1.58%	1.51%	1.39%	1.45%

the latest 3 days of load as input. The past load used as input is corrected to reflect the load fluctuation by temperature. The detailed operation and description of KSLF for load forecasting are described in [2]. Tables 2, 3 shows the day of week average MAPE of each method's forecasting results for 2017 and 2018.

Compared with the results of KSLF, the proposed method shows better forecasts for all days of the week. In the KSLF, Monday, Tuesday through Friday, Saturday, and Sunday are separately forecast. Also, the input data and forecasting method are different for each day of the week. This is to reflect different load characteristics by day of the week. For this reason, when forecasting Monday, Tuesday and weekend, the KSLF cannot reflect load continuity characteristics between adjacent days. Additionally, on Monday and weekends, the forecasting accuracy is low because other week data is used for forecasting. In the proposed method, forecasting accuracy is improved because forecasts use adjacent days. As shown in Fig. 10, the proposed method has a better forecasting accuracy in the Monday morning time than previous study.

The FC layer which has same factors except the presence of LSTM layer in proposed method, as follows Fig. 11, are also used for comparison.

Table 2 The forecasting results of day of week average MAPE according to each method in 2017

	KSLF (%)	Proposed algorithm (%)
Monday	2.30	1.66
Tuesday	1.99	1.57
Wednesday	1.84	1.35
Thursday	1.56	1.21
Friday	1.67	1.24
Saturday	2.12	1.69
Sunday	2.35	1.69

Table 3 The forecasting results of day of week average MAPE according to each method in 2018

	KSLF (%)	Proposed algorithm (%)
Monday	2.53	1.78
Tuesday	2.20	1.81
Wednesday	2.12	1.28
Thursday	1.71	1.29
Friday	1.97	1.43
Saturday	2.68	1.44
Sunday	2.59	1.70

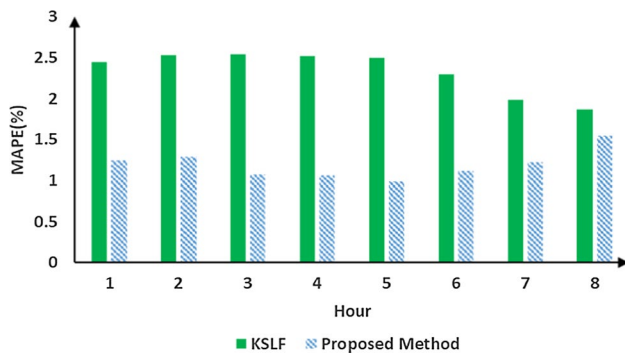


Fig. 10 Comparison the load forecasting error between KSLF and the proposed method for the Monday morning in 2018

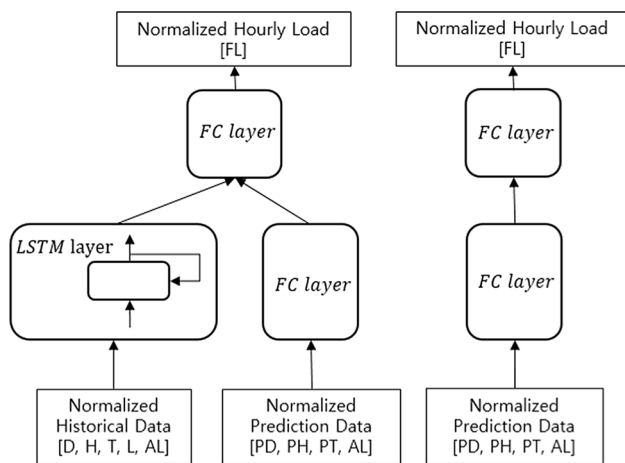


Fig. 11 The difference of structure with proposed method (left) and FC layer (right)

Table 4 The forecasting results of monthly average MAPE according to each method in 2017

	KSLF (%)	FC layer (%)	Proposed algorithm (%)
Jan	2.29	1.93	1.65
Feb	2.10	1.60	1.48
Mar	1.85	1.20	1.21
Apr	1.70	1.40	1.26
May	1.77	1.21	1.27
June	1.30	1.20	1.18
July	2.39	1.86	1.93
Aug	3.27	2.26	1.86
Sep	1.38	1.29	1.18
Oct	1.99	2.21	1.99
Nov	1.70	1.72	1.63
Dec	2.18	1.48	1.20
Average	1.99	1.61	1.49

Table 5 The forecasting results of monthly average MAPE according to each method in 2018

	KSLF (%)	FC layer (%)	Proposed algorithm (%)
Jan	2.12	2.17	1.80
Feb	2.51	1.49	1.32
Mar	2.66	1.58	1.77
Apr	1.84	1.49	1.23
May	2.00	1.43	1.18
June	1.70	1.40	1.32
July	2.87	2.03	1.99
Aug	3.10	2.64	2.34
Sep	2.59	1.68	1.34
Oct	1.55	1.52	1.26
Nov	1.64	1.44	1.19
Dec	2.67	1.76	1.54
Average	2.27	1.72	1.52

Tables 4, 5 shows the monthly average MAPE of each method’s forecasting results. It appears that adding LSTM layer to FC layer reflects the functionality between historical data for improved predictability. Because of applying hourly temperature, the accuracy of the proposed method is significantly better in summer and winter than KSLF. But the proposed method has poor forecasting performance for the period after long holidays such as Lunar New Year’s Day, Chuseok. The data for public holidays are not used as input for forecast. Therefore, the periods after long holidays are forecast using data from a day far from the forecasting day. For these reasons, there is a large error in the period after long holidays.

5 Conclusion

A forecasting method based on deep neural networks using LSTM layer is proposed to perform day-ahead load forecasting. In the proposed forecasting method, input features are processed using different types of layers according to their specific characteristics. In order to analyze the characteristics of load, time series analysis was conducted. In addition, correlation analysis with temperature is performed to select the input features. The selected input features are divided into the historical and prediction data to reflect the characteristics of the past data and the information of the forecasting day. A LSTM layer is used to extract features from past data and the FC layer is used as input for the forecasting day. In order to test the proposed forecasting method, day-ahead load forecasting is performed for normal days (excluding special days) in 2017 and 2018. Moving window method are used to select the training and testing set. The annual average

MAPE of proposed method is 1.49% for 2017, 1.52% for 2018. And the proposed method shows better performance than KSLF, and better than FC layer which have the same structure except for no LSTM layer.

The forecasting method based on deep neural networks using LSTM layer presented in this paper will contribute to stabilizing power systems and to efficient power market operation. Future research will be carried out to reflect load fluctuations due to solar radiation and wind speed changes in order to take into account fluctuations due to renewable energy.

Acknowledgements This research was supported by Korea Electric Power Corporation (Grant number: R18XA04) and “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20184010201690).

References

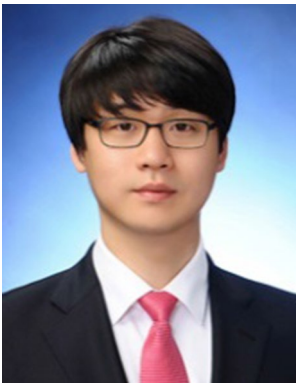
- Mocanu E, Nguyen PH, Gibescu M, Kling WL (2016) Deep learning for estimating building energy consumption. *Sustain Energy Grids Netw* 6:91–99
- Korea Power Exchange (2011) A study on short-term load forecasting technique and its application
- Ribeiro GH, Neto PSDM, Cavalcanti GD (2011) Lag selection for time series forecasting using particle swarm optimization. In: *Proceedings of the IEEE 2011 IJCNN, San Jose, CA, USA, Aug 2011*
- Box GE, Jenkins GM, Reinsel GC (2015) *Time series analysis: forecasting and control*. Wiley, Hoboken
- Chen JF, Wang WM, Huang CM (1995) Analysis of an adaptive time-series autoregressive moving-average (ARMA) model for short-term load forecasting. *Electr Power Syst Res* 34:187–196
- Haida T, Muto S (1994) Regression based peak load forecasting using a transformation technique. *IEEE Trans. Power Syst* 9:1788–1797
- Engle RF, Mustafa C, Rice J (1992) Modeling peak electricity demand. *J Forecast* 11:241–251
- Papalexopoulos AD, Hesterberg TC (1990) A regression-based approach to short-term system load forecasting. *IEEE Trans Power Syst* 5:1535–1547
- Charytoniuk W, Chen MS, Olinda PV (1998) Nonparametric regression based short-term load forecasting. *IEEE Trans Power Syst* 13:725–730
- Mbamalu GAN, Elhawary ME (1993) Load forecasting via sub-optimal seasonal autoregressive models and iteratively reweighted least squares estimation. *IEEE Trans Power Syst* 8:343–348
- Jung H, Song K, Park J, Park R (2018) Very short-term load forecasting for real-time power system operation. *J Electr Eng Technol* 13:1419–1424
- Yang HT, Huang CM (1998) A new short-term load forecasting approach using self-organizing fuzzy ARMAX models. *IEEE Trans Power Syst* 13:217–225
- Song K, Baek Y, Hong D, Jang G (2005) Short-term load forecasting for the holidays using fuzzy linear regression method. *IEEE Trans Power Syst* 20:96–101
- Hernandez L, Baladron C (2014) A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. *IEEE Commun Surveys Tut* 16:1460–1495
- Czernichow T, Piras A, Imhof K (1996) Short term electrical load forecasting with artificial neural networks. In: *Engineering Intelligent Systems for Electrical Engineering and Communications 4*. ARTICLE, pp. 85–99, Feb. 1996
- Bakirtzis AG, Petridis V, Klartzis SJ (1996) A neural network short term load forecasting model for the Greek power system. *IEEE Trans Power Syst* 11:858–863
- Papadakis SE, Theocharis JB, Kiartzis SJ (1998) A novel approach to short-term load forecasting using fuzzy neural networks. *IEEE Trans Power Syst* 13:480–492
- Bashir ZA, El-Hawary ME (2009) Applying wavelets to short-term load forecasting using PSO-based neural networks. *IEEE Trans Power Syst* 24:20–27
- Kodogiannis VS, Amina M, Petrounias I (2013) A clustering-based fuzzy wavelet neural network model for short-term load forecasting. *Int J Neural Syst* 23(05):1350024
- Fan S, Chen L (2006) Short-term load forecasting based on an adaptive hybrid method. *IEEE Trans. Power Syst* 21:395–401
- Bengio Y (2013) Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell* 35:1798–1828
- Vermaak J, Botha EC (1998) Recurrent neural networks for short-term load forecasting. *IEEE Trans Power Syst* 13:126–132
- Hochreiter S (1996) LSTM can solve hard long time lag problems. In: *Proceedings of the advances in neural information processing systems, Denver, 1996*
- Bouktif S, Fiaz A, Ouni A, Serhani MA (2018) Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches. In: *Energies*, 2018
- Zheng H, Yuan J, Chen L (2017) Long short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. In: *Energies*, 2017
- Hippert HS, Pedreira CE, Souza RC (2001) Neural networks for short-term load forecasting: a review and evaluation. *IEEE Trans Power Syst* 16:44–55
- Kong W, Dong ZY, Jia Y (2017) Short-term residential load forecasting based on LSTM Recurrent neural network. *IEEE Trans Smart Grid* 10:841–851
- Wang Y, Gan D, Sun M, Zhang N, Lu Z, Kang C (2019) Probabilistic individual load forecasting using pinball loss guided LSTM. *Appl Energy* 235:10–20
- Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. *IEEE Trans Neural Netw* 5:157–166
- Gers FA, Schmidhuber J, Cummins F (1999) Learning to forget: continual prediction with LSTM. *Neural Comput* 12:2451–2471
- Karpathy A, Fei-Fei L (2017) Deep visual-semantic alignments for generating image descriptions. *IEEE Trans Pattern Anal Mach Intell* 39:664–676
- Lim J, Kim S, Park J, Song K (2013) Representative temperature assessment for improvement of short-term load forecasting accuracy. *J KIIEIE* 27:39–43
- Song K (2014) Development of short-term load forecasting using hourly temperature. *Trans KIEE* 63:451–454
- He W (2017) Load forecasting via deep neural networks. *Proc Comput Sci* 122:308–314
- Kwon B, Park R, Song K (2018) Analysis of short-term load forecasting using artificial neural network algorithm according to normalization and selection of input data on weekdays. In: *Proceedings of the IEEE PES APPEEC, 2018*
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
- Duchi J, Hazan E, Singer Y (2011) Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* 12:2121–2159

38. Zeiler MD (2012) ADADELTA: an adaptive learning rate method. arXiv preprint [arXiv:1212.5701](https://arxiv.org/abs/1212.5701)
39. Tieleman T, Hinton G (2012) Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA Neural Netw Mach Learn 4:26–31

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Bo-Sung Kwon received his B.S. degree in Electrical Engineering from Kongju National University, Cheonan, Korea, in 2018. Currently, he is pursuing Ph.D. degree at Soongsil University, Seoul, Korea. His research interests include power system operation and control, power system economics, load forecasting of power system.



Rae-Jun Park received his B.S. and M.S. degree in Electrical Engineering from Soongsil University, Seoul, Korea, in 2011 and 2013. Currently, he is pursuing Ph.D. degree at Soongsil University, Seoul, Korea. His research interests include power system operation and control, power system economics, load forecasting of power system.



systems.

Kyung-Bin Song received his B.S. and M.S. degrees in Electrical Engineering from Yonsei University, Seoul, Korea, in 1986 and 1988, respectively. He received his Ph.D. degree in Electrical Engineering from Texas A&M University, College Station, Texas in 1995. He is currently a Professor in Electrical Engineering at Soongsil University, Seoul, Korea. His research interests include power system operation and control, power system economics, and load forecasting of power