# Enhanced LSTM Model for Short-Term Load Forecasting in Smart Grids

Jianing Guo[1](✉), Yuexing Peng[1], Qingguo Zhou[2], and Qingquan Lv[3]

[1] Beijing University of Posts and Telecommunication, Beijing, China
{guojianing,yxpeng}@bupt.edu.cn
[2] Lanzhou University, Lanzhou, China
zhouqg@lzu.edu.cn
[3] Wind Power Technology Center of Gansu Electric Power Company, Lanzhou, China

**Abstract.** With the rapid development of smart grids, significant research has been devoted to the methodologies for short-term load forecasting (STLF) due to its significance in forecasting demand on electric power. In this paper an enhanced LSTM model is proposed to upgrade the state-of-the-art LSTM network by exploiting the long periodic information of load, which is missed by the standard LSTM model due to its constraint on input length. In order to distill information from long load sequence and keep the input sequence short enough for LSTM, the long load sequence is reshaped into two-dimension matrix whose dimension accords to the periodicity of load. Accordingly, two LSTM networks are paralleled: one takes the rows as input to extract the temporal pattern of load in short time, while the other one takes the columns as input to distill the periodicity information. A multi-layer perception combines the two outputs for more accurate load forecasting. This model can exploit more information from much longer load sequence with only linear growth in complexity, and the experiment results verify its considerable improvement in accuracy over the standard LSTM model.

**Keywords:** Long short-term memory · Short term load forecasting · Recurrent neural network

## 1 Introduction

Load forecasting plays an important role in many departments of electric power system since it is the basis of planning generation, maintain and energy selling. Ranaweera et al. [1] has quantitatively analyzed how prediction error impacts the operation of electric power system. Douglas et al. [2] performed the assessment for system operating risk with known distribution variance of forecasting result. The analyzing of these two articles indicates electric power systems necessarily keep asking for more accurate load forecasting. With the establishment of smart grids, intelligent scheduling has become a new requirement. As the basis of electricity power scheduling, intelligent load forecasting faces new challenges of considering diversified influenced factors and being adaptive to fluctuation. Within all kinds of load forecasting, short term load forecasting (STLF)


X. Zhang et al. (Eds.): CloudComp 2019/SmartGift 2019, LNICST 322, pp. 650–662, 2020.
https://doi.org/10.1007/978-3-030-48513-9_52

forecasts the maximum or average load from one day to one week ahead. It helps the coordination between electric power system departments, the planning of generation and unit commitment scheduling. Accurate STLF is significant to electric grid for saving limited energy and asset. According to the accurate forecast information, it is possible to reasonably regulate the power generation capacity in order to avoid wasting because the power is hard to save. On the contrary, an inaccurate STLF information can lead to excess supply, or underestimation of load, resulting in costlier supplementary service. Therefore, STLF has become an important project in the field of power system and even a small percentage improve is willing to be saw.

## 2 Literature Study

Many approaches have been developed for STLF, which can be roughly categorized into statistical methods and machine learning methods. Statistical methods include regression methods and gray models. On the basis of known historical data series, a mathematical model is established to describe the relationship between load value and time in time series method. Regression methods are effective for stationary series by building equations to fit the relationship between independent and dependent variables. While for non-stationary series like load series, regression model cannot fit the fluctuation well. And the prediction is not robust with the time period which is highly influenced by factors like weather or holiday [3]. In the work of Cho et al. [4] the ARIMA model and transfer function model are applied to the short-term load forecasting by considering weather-load relationship. By comparing the effectiveness with traditional regression method and single ARIMA method, ARIMA model with transfer function achieves better accuracy. Gray models consider the electric power system as a gray system. It has small amount of computation and good prediction effect in system with uncertain factors. But it is only effective in exponential trend load [5].

Electrical load features typical randomness induced by many external factors such as temperature and special events, which facilitates the deployment of machine learning methods such as neural network and support vector machine. Due to its strong capability of well-fitting nonlinear function, machine learning-based STLF methods account for a high proportion of research. Machine learning-based methods can be classified into shallow network and deep neural network. In a neural network, hidden units at shallower location extract some simple and local information [6], while comes to the deeper layers, they can extract more complicated and global features. So with the simple structure, shallow neural network is not able to extract the complicated pattern in load series while deep neural network can fit them more accuracy.

Haqueau et al. [7] proposed a hybrid approach based on wavelet transform and fuzzy ARTMAP network and has good prediction effect in wind farm power prediction. Ghelardonil et al. [8] broke the load time series into two parts with empirical mode decomposition, respectively describing the trend and the local oscillations of the energy consumption values, and then feed them into support vector machine to train. Experiments results show the method has high prediction accuracy in load prediction. Han et al. [9] utilized the state prediction model and algorithm of weighted least squares, the main approach is to take the voltage characteristic as the basic quantity to describe the state characteristics of the system, and adopt the state estimation method to carry out load prediction, which is advantageous for dynamic load prediction. Zhang et al. [10] used extreme learning machine with ensemble structure to forecast the total load of Australian energy market. The ensemble structure helps decrease the uncertainty of prediction. Kong et al. [11] applied the long short-term memory (LSTM) recurrent neural network based framework into residential households' smart meter data. They compare the difference between the residential load forecasting and substation load and the result shows the LSTM networks achieves the best forecasting performance in residential data among other benchmark algorithm.

As a typical deep neural network for time series, LSTM has strong ability of abstracting features and learning the inner complex patterns of load series, and presents state-of-the-art performance. However, it cannot avoid gradient vanishing and exploding problems so that it has the constraints on the input length. Meanwhile, the computation complexity increases non-linearly as the input length increases, and the forecasting accuracy will not increase with the increase of input length, if not worse. In order to upgrade LSTM by overcoming the input length constraint to distill more significant features of load, in this paper an enhanced LSTM is proposed. By reshaping the long load sequences into two-dimensional matrix whose row size accords to weekly period while column size accords to monthly period, two LSTM networks are paralleled to take rows and columns as input and feed an MLP for better information merging. The model realizes better prediction performance with only linear complexity increase, and experiment result verifies its enhancement over standard LSTM model.

## 3 Enhanced LSTM Model

### 3.1 Discovery of Rules in Data

In this section we firstly statistically analyze the feature of load to lay a foundation for the design of the proposed model, and then briefly introduce standard LSTM model to mention its pros and cons. At last, an enhanced LSTM model is detailed to state its main idea to upgrade LSTM network and its structure.

**Periodicity.** Temporal correlation is the first considerable factor for time series. By analyzing load profiles, it is clear that load is periodic in weeks, months and years. By utilizing the periodicity in the forecasting, the prediction effect can be improved. Figure 1 shows the periodicity in week in the load series (see Fig. 1). Each curve describes the total used load in Toronto in three weeks of May in 2015. It can be shown in the figure that the trends of the three weeks are similar. In other words, load series is periodic on weekly scale. Monthly periodicity is shown in Fig. 2. Three figures respectively corresponding to curves in summer, winter and transitional seasons (spring and autumn) (see Fig. 2). Each curve shows the load through a month. It is clear that load of each season has similar trend. The yearly periodicity in shown in Fig. 3, which shows the load used in 12 months in 2014, 2015 and 2016 in Hangzhou. It is clear that the trend of three curves have highly similarity (see Fig. 3).

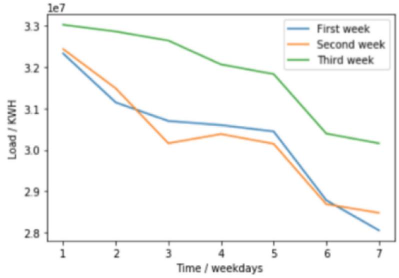



**Fig. 1.** Curves of load in a week, including three weeks of May 2015 of Toronto

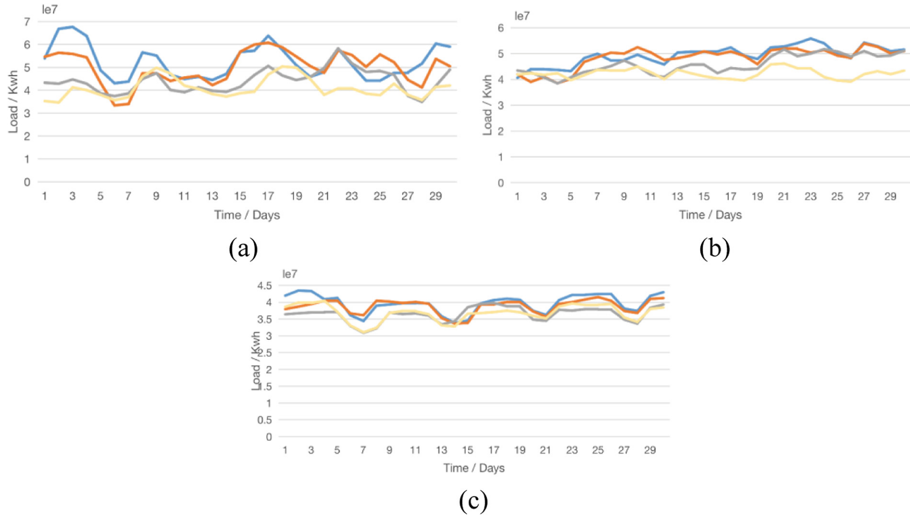



**Fig. 2.** Curves of a month in *(a)* Summer *(b)* Winter *(c)* Transitional season

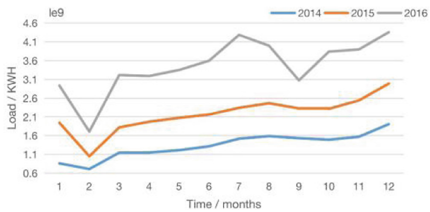



**Fig. 3.** Curves of total used load in 12 months for 2014, 2015 and 2016 in Hangzhou

**Weather.** Weather is another well-known key factor to be considered for STLF, especially the temperature factor. Paravan et al. [12] has performed an experiment showed that temperature is high positively correlated to load in summer and negatively correlate in winter. What means when temperature arises in summer, people need to use electrical product like air conditioner to cool themselves, so load arises as well. While in winter if the temperature decreases, electric blanket are used to get warm if the temperature decreases, as load arises. Figure 4 are curves of the maximum temperature and load in 2012 of Toronto in Canada (see Fig. 4). It is clear to show the negative correlation in winter and positive in summer. In spring and autumn, temperature and load are less correlated, so they are regarded as transitional seasons.

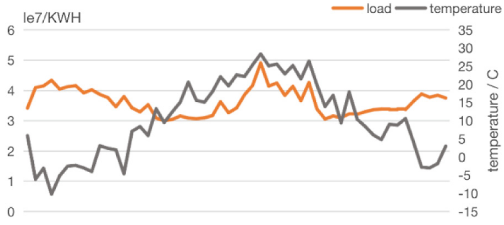



**Fig. 4.** Maximum temperature versus load in Toronto in 2012

### 3.2 Long-Short Term Model

Long short-term memory (LSTM), a variant of recurrent neural network (RNN), which is a deep network performing pretty good in sequence learning since it introduces gates to sift previous information flowing in the recurrent unit. As its name, LSTM can model data accurately with both long and short-term dependencies, and relieve the gradient vanishing and exploding problem by introducing three gates. Same as RNN, LSTM has a chain of repeating modules to process each time step's data with same flow (see Fig. 5). But LSTM adds an internal cell to process the memory of past information. Within the cell, past information is controlled by three gates: input gate, forget gate and output gate. The cell permits the network not remembering all past information, instead, the network remembers, stores and transfers only the most related information to the current value and forget the less related information. Therefore, remembering information for long periods of time is practically LSTM's default behavior, not something they struggle to learn.

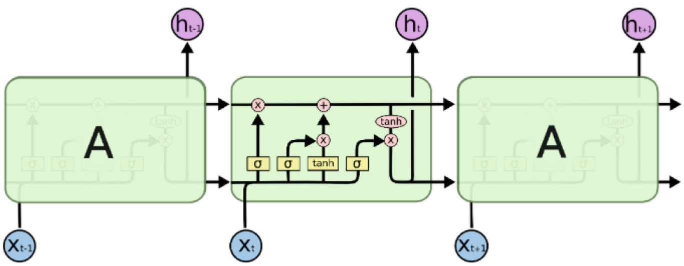



**Fig. 5.** The repeating module in an LSTM network [13]

### 3.3 The Enhanced LSTM Model

With the gates, LSTM can remember the most useful temporal information and build a relationship with current circumstance. Because of its good capacity to model sequence with long dependency, LSTM has been applied and achieved state-of-the-art results in many fields of sequence learning, such as speech recognition, machine translation and language generation However, it has not been sufficient maturely applied on time series with long relevance such as STLF which features typical large periodicity. Moreover, the training cost increases non-linearly with the length of input sequence, and the accuracy sometimes decreases when input sequence continue to increase. It is the reason only several days' data are used to predict the following days' load in the existing works [14]. However, from the rule in Sect. 3.1 it is known that load has high weekly and monthly similarity, which means we can upgrade the LSTM if we can use the data months ago instead of just few days. The proposed model utilizes much longer data to improve the prediction performance with a little payload of training complexity. The method performs better by bringing temporal dependency of load profile into full play.

The proposed model consists of two LSTM networks in parallel and a MLP, which is illustrated in Fig. 6 (see Fig. 6). The first one LSTM network of the proposed model takes advantage of daily periodicity of load series which has been shown in Sect. 3.1. Except for the proximity of the curves' trend, temperature values are always closed within several adjacent days, which is well understood. It means that the sequence of load values has high dependency within a few days. Accordingly, the first LSTM network (also termed row LSTM network which takes the rows of the reshaped data matrix as the input) takes the last seven days' load as input and external factors as well such as special event index, month index, and weather factor including maximum, minimum and average temperatures.

Besides the temporal dependency information distilled by the row LSTM network, the weekly and monthly periodicity information is exploited by the second LSTM network (namely, the column LSTM network taking the columns of the reshaped load matrix as the input sequences). In Sect. 3.1, it is proved that there is also dependency in the scale of week and month as well. For example, the load profile of one week in May is similar to the profiles of another week. What is more, the load profile of one particular day is similar to the profiles of another day with the same type of weekday in different weeks. Which means, to forecast a load value in Monday, the other LSTM network takes the data of past weeks' Monday as input, which amount to contain the temporal dependency information through two months. Therefore, the relative information useful for prediction can be extracted from the input without taking the whole months' data in the network. Except for the load value, other four influencing external factors are included as same as the first network. By utilizing the periodicity of load series, the temporal dependency can be extended from a week to two months with a little payload of training complexity.

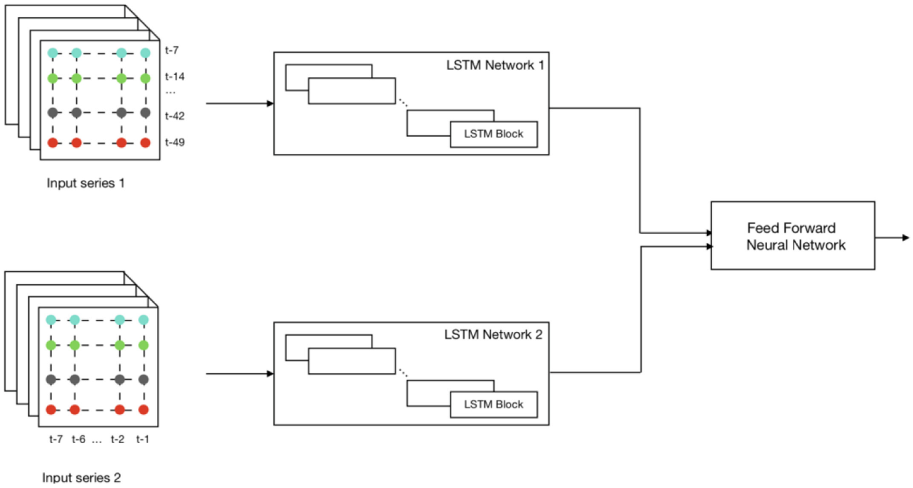



**Fig. 6.** Structure of the enhanced LSTM model

## 4 Experiments

### 4.1 Introduction of the Dataset

The experiments are performed on load and temperature dataset of Hangzhou city in East China. There are eight kinds of factors for each day, including the load in kilowatt-hour (KWH), maximum and minimum temperatures in centigrade, precipitation, holiday, day of week, month and date. Data are collected every day from 1st January 2014 to 3rd March 2017, with a total of 1185 data.

### 4.2 Data Pre-processing

Pre-processing transformations are applied in following order: (*i*) recognition and removing of abnormal data, (*ii*) trend removing, (*iii*) one-hot encoding (*iv*) standardization. Each operation is successively reversed to evaluate the forecast produced by the model.

**Data Cleaning and Missing Values Processing.** The practical dataset often contains abnormal values or may be missing. The presence of these data brings significant disturbance to normal data, hence affects the prediction accuracy. If the anomaly data is too large, it even misleads the prediction results. Therefore, the adverse effects caused by abnormal data must be eliminated. For the missing data, we fill the vacancy with the average of data before and after the lost data. As to abnormal data, in Hangzhou dataset, data from 1st January 2016 to 1st March 2017 is shown in (see Fig. 7). It can be seen that there are three obvious valleys in the load profile. The valleys near 1st February in 2016 and 2017 are regular since they are at the Spring Festival of China and people will stay at home and same valleys can be seen every year in the dataset as well. While there is an abnormal valley at 1st September 2016 and it is not a major holiday in China. By collecting news at the day, we found that G20 Financial Summit was hold in Hangzhou from 4th September to 5th September and most enterprises gave holidays for a few days before and after it. Since there is no similar rule with the data in these days in the dataset, no data can be used to fit the trend of it, what means it is unique and independent incident. As a result, there are nearly two weeks' data against the regular rule that need to be removed from the dataset. After the processing, the new dataset consists of 1171 measurements.

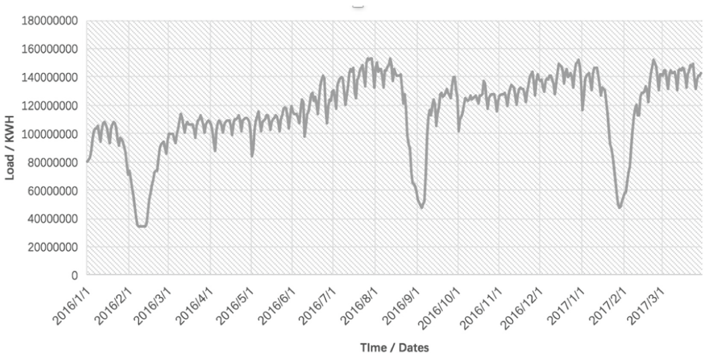
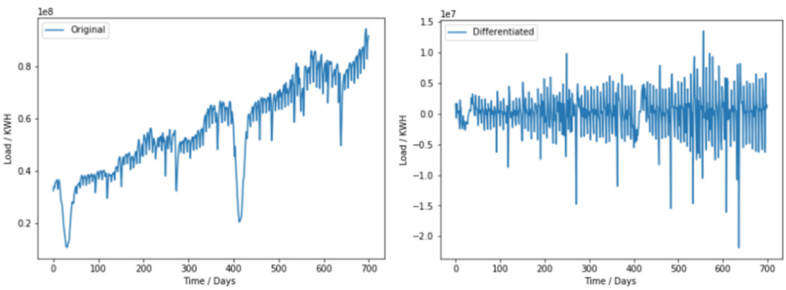



**Fig. 7.** Data from 1st Jan 2016 to 1st Mar 2017

**Processing for the Trend Factor of Load.** From the partial load profile of Hangzhou in (see Fig. 8 (a)) it can be figured out that the value of load keeps increasing through years, which is regarded as a trend item of the load dataset. Since there will be a scaling process later, the scaling effect to the proceeding years' data would be stronger than that to the later ones if the trend factor remains. This can bring change to the natural rule of the load and then reduce the forecasting accuracy. Therefore, the trend is filtered out by applying a daily differentiating. The load profile after differentiated can be seen in Fig. 8 (b). It tells the trend factor is removed from the load that can improve the forecasting accuracy of models in neural networks.




(a) (b)

**Fig. 8.** *(a)* The original load profile. *(b)* The load profile after differentiated

**One-Hot Encoding.** Some affecting factors are category variables, which are one-hot encoded for easy processing of LSTM model. One-hot encoding codes $n$- class value into $n$-tuple binary vector.

### 4.3 Experiments Set-up

The experiment is implemented in Keras library with Tensorflow as backend. After pre-processed stated in the Sect. 4.2, the dataset includes eighteen features, as shown in Table 1. For each weekday, the input series is in the form as Table 2 shows. The input for the horizontal LSTM is the data in past seven days and for the longitudinal LSTM is the past seven weeks on the same weekday. The dataset is divided into train set and test set in proportion of 80% and 20%. The input for train set is reshaped in (127, 7, 18) and for test set is (32, 7, 18).

**Table 1.** Features after pre-processing

| $m_1$ | $m_2$ | $m_3$ | $m_4$ | $m_5$–$m_{17}$ | $m_{18}$ |
|---|---|---|---|---|---|
| Max temperature | Min temperature | Rain | Holiday | Month | Power |

**Table 2.** The form of inputs

| | |
|---|---|
| Horizontal LSTM | $\{m_1(t-7), m_2(t-7), \ldots, m_1(t-6), m_2(t-6), \ldots, m_{17}(t-1), m_{18}(t-1)\}$ |
| Longitudinal LSTM | $\{m_1(t-49), m_2(t-49), \ldots, m_1(t-42), m_2(t-42), \ldots, m_{17}(t-1), m_{18}(t-1)\}$ |

In the model, both LSTM networks use two layers' structure, and the two networks are merged by Merge layer in Keras. After merging, MLP composed of two-layered fully connected layer is added to the network to adjust the weights slightly and then output a predicted value. To decide the best configuration for the model, different values of hyper parameters are evaluated. For each combination, the performance of the model is evaluated on test set, with the forecasting error defined as mean absolute percentage error (MAPE), which is defined by

$$MAPE = \frac{1}{N} + \sum\nolimits_{n=1}^{N} \frac{\lceil \widehat{y_n} - y_n \rceil}{y_n} * 100\% \tag{1}$$

where $y_n$ and $\hat{y}_n$ are the real and the forecast load value at the $\mathrm{n_{th}}$ day and N is the length of the load sequence. Several hyper-parameters need to be specified and the value of each hyper-parameter is uniformly sampled from a given interval. The process of hyper-parameters is as followed: Adam is used as gradient descent strategy, whose hyper-parameters are kept to the default value in Keras. The default step size $\varepsilon = 0.001$, the rate of exponential decay in first-order and second-order moment estimation $\rho_1 = 0.9$ and $\rho_2 = 0.999$, $\delta = 10^{-6}$. The setting of other hyper-parameters is shown in Table 3 according to repeating experiments. The simulation is performed for data with seven different weekdays (Monday, …, Sunday). To evaluate the performance of the system, the prediction is also realized in other benchmark algorithms, including single LSTM and ARIMA.

### 4.4 Results

After using different combination of hyper parameters to train the model, the best model for each type of weekday is obtained. The best configuration for the ensemble LSTM model and the comparison of mean absolute percentage error (MAPE) between ensemble LSTM and single LSTM of forecasting the load after seven days is shown in the Table 3.

**Table 3.** Configuration for ensemble LSTM and the accuracy comparison between ensemble and single model

| | Configuration | | | | | | MAPE | |
|---|---|---|---|---|---|---|---|---|
| Day | Hidden units | Epoch | $L_1$ | $L_2$ | Dropout | Dense neuron | Merged LSTM | Single LSTM |
| Mon. | 20 | 48 | 0 | 0.01 | 0.1 | 20 | 2.605% | 3.426% |
| Tue. | 20 | 52 | 0 | 0.0018 | 0.2 | 20 | 2.631% | 3.312% |
| Wed. | 10 | 49 | 0 | 0.01 | 0.2 | 20 | 2.352% | 3.287% |
| Thu. | 20 | 42 | 0.01 | 0.01 | 0.2 | 20 | 2.434% | 3.335% |
| Fri. | 20 | 68 | 0 | 0 | 0.1 | 20 | 2.509% | 3.435% |
| Sat. | 20 | 110 | 0 | 0.001 | 0.2 | 10 | 2.314% | 3.185% |
| Sun. | 20 | 43 | 0.01 | 0.01 | 0.2 | 10 | 2.625% | 3.298% |

From the results we can observe that the enhanced LSTM model achieves an MAPE performance of 2.495% for all weekdays and outperforms standard LSTM model whose MAPE is 3.325%. Except for the average accuracy, it can also be seen from the last two columns that the accuracy for weekday are all improved. The predicted and real load are shown in Fig. 9. It is clear that the enhanced LSTM's profiles are much closer to the real data, especially in the case of sharp fluctuation of the real load.

Classic ARIMA model is also evaluated, and it achieves MAPE of 3.386%, which is a little worse than standard LSTM model. The real and predicted loads are shown in Fig. 10 for the last 85 days of the test set (see Fig. 10). It can be seen that ARIMA behaves badly when a sharp fluctuation comes up, verifying ARIMA not suitable for time series with high randomness.

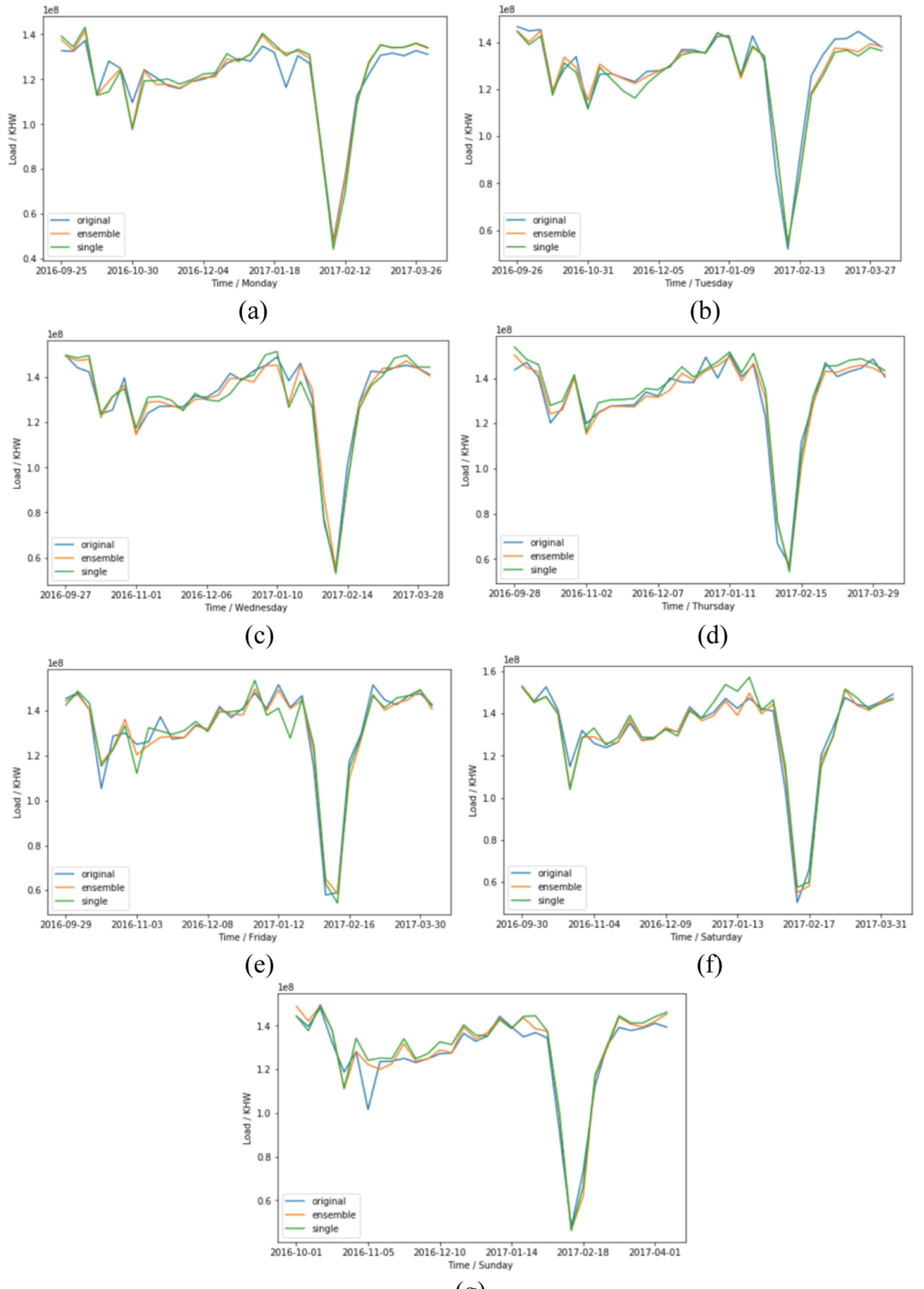



**Fig. 9.** Comparison between real value and predicted value in ensemble and single LSTM for *(a)* Mondays *(b)* Tuesdays *(c)* Wednesdays *(d)* Thursdays *(e)* Fridays *(f)* Saturdays *(g)* Sundays

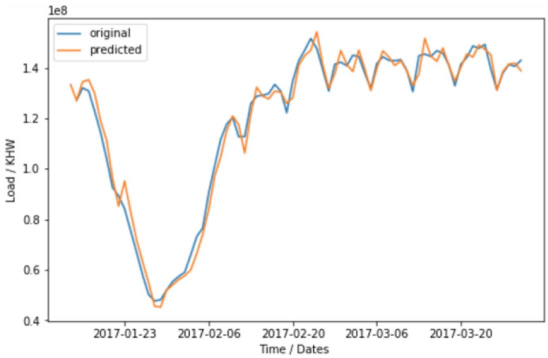



**Fig. 10.** Comparison between real value and predicted value in AIMA for the last 85 samples

## 5 Conclusion

Based on the characteristic analysis of load and the shortcoming of mainstream STLF models, an enhanced LSTM model is developed. Classic LSTM holds excellent learning capability to model the temporal inner pattern of short time sequences so that it presents the state-of-the-art performance, but it cannot avoid gradient disappear problem and thus has limit on the length of the input sequence. Since electrical load features typical weekly and monthly periodicity, the fully utility of the large time-span periodicity will upgrade the STLF performance. In this sense, the proposed model is designed by integrates two LSTM networks and an MLP. The first LSTM network takes the rows of load matrix, which is constructed by reshaping a long load sequence with row size being weekly periodicity, to exploit the similarity of the load in adjacent days, while the other LSTM network takes the columns of the load matrix as input to exploit the weekly and monthly periodicity of the load. The MLP merges the distilled information from two LSTM networks. The proposed model extends the temporal dependency from one week to two months but increases the training complexity in a linear mode. The proposed model is evaluated and compared with ARIMA and standard LSTM network. The simulation results verify its advantage over the reference models, i.e., MAPE performance is decreased from 3.736% of ARIMA and 3.325% of standard LSTM to 2.495% for seven days forecasting.

**Acknowledgment.** The work of Q. Zhou and Q. Lv is partly supported by project with Grant No. SGGSKY00FJJS1900241.

## References

1. Ranaweera, D., Karady, G., Farmer, R.: Economic impact analysis of load forecasting. IEEE Trans. Power Syst. **2**(3), 1288–1392 (1997)
2. Douglas, A.P., Breipohl, A.M., Lee, F.N.: Risk due to load forecast uncertainty in short term power system planning. IEEE Trans. Power Syst. **13**(4), 1493–1499 (1998)

3. Cho, M.Y., Hwang, J.C., Chen, C.S.: Customer short term load forecasting by using ARIMA transfer function model. In: Proceedings 1995 International Conference on Energy Management and Power Delivery, EMPD 1995. IEEE (2002)
4. Ye, G., Luo, Y., Liu, Y., Zhang, J.: Research on method of power system load forecasting based on ARMA model. Inf. Technol. **6**, 74–76 (2002)
5. Liu, B.: Research on household electricity load control strategy based on demand response. Harbin University of Science and Technology (2014)
6. Wang, G., Xiang, W., Pickering, M., Chen, C.W.: Light field multi-view video coding with two-directional parallel inter-view prediction. IEEE Trans. Image Process. **25**(11), 5104–5117 (2016)
7. Haque, A., Mandal, P., Meng, J.: A novel hybrid approach based on wavelet transform and fuzzy ARTMAP network for predicting wind farm power production. In: IEEE Industry Applications Society Annual Meeting (IAS), Las Vegas, NV, USA (2012)
8. Ghelardoni, L., Ghio, A., Anguita, D.: Energy load forecasting using empirical mode decomposition and support vector Regression. IEEE Trans. Smart Grid **4**(1), 549–556 (2013)
9. Han, L., Han, X., Lei, M.: Method for state forecasting and estimation based on nodal load forecasting and generation control. Autom. Electr. Systems **33**(4), 16–20 (2009)
10. Zhang, R., Dong, Z., Xu, Y., Meng, K.: Short-term load forecasting of Australian National Electricity Market by an ensemble model of extreme learning machine. IET Gener. Transm. Distrib. **7**, 391–397 (2013)
11. Kong, W., Dong, Z., Jia, Y., et al.: Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Trans. Smart Grid **10**(1), 841–851 (2017)
12. Xiang, W., Wang, G., Pickering, M., Zhang, Y.: Big video data for light-field-based 3D telemedicine. IEEE Netw. **30**(3), 30–38 (2016)
13. Paravan, D., Debs, A., Hansen C., Becker, D., Hirsch P., Golob, R.: Influence of temperature on short-term load forecasting using the EPRI-ANNSTLF (2003)
14. Understanding LSTM Networks. http://colah.github.io/posts/2015–08-Understanding-LSTMs
15. Bianchi, F.M., Maiorino, E., Kampffmeyer, M.C.: An overview and comparative analysis of recurrent neural networks for short term load forecasting, 11 May 2017
16. Xiang, W., Barbulescu, S.A., Pietrobon, S.S.: Unequal error protection applied to JPEG image transmission using turbo codes. In: Proceedings 2001 IEEE Information Theory Workshop (Cat. No. 01EX494), pp. 64–66 (2001)