Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 158 (2019) 2922–2927

Energy Procedia

www.elsevier.com/locate/procedia

10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China

Short-Term Load Forecasts Using LSTM Networks

Shahzad Muzaffar, Afshin Afshari*

Masdar Institute, Khalifa University, Abu Dhabi, 54224, UAE

Abstract

With the increasing load requirements and the sophistication of power stations, knowing in advance about the electrical load not only at short-term periods such as hours or couple of days but also over the longer-term periods such as weeks and months is indispensable for a range of benefits such as important technical and economic impacts. Traditional methods such as ARMA, SARIMA, and ARMAX have been used for decades. In recent years, the artificial intelligence (AI) techniques such as neural networks and deep learning are emerging in the field of time series analysis. Towards this end, the artificial neural networks (ANN) and recurrent neural networks (RNN) are being explored and have shown promises in much better forecasting as compared to traditional methods. Long short-term memory (LSTM) networks are a special kind of RNN that have the capabilities to learn the long-term dependencies. In this work, we have picked up an electrical load data with exogenous variables including temperature, humidity, and wind speed. The data is used to train the LSTM network. For a fair comparison, the data is also used in traditional methods to model the load time series. The trained LSTM network and the developed models are then used to forecast over the horizons of 24 hours, 48 hours, 7 days and 30 days. The forecasts generated by the LSTM are compared with the results of traditional methods using RMSE and MAPE for all the forecast horizons. The results of a number of experiments show that the LSTM based forecast is better than other methods and have the potential to further improve the accuracies of forecasts.

Peer-review under responsibility of the scientific committee of ICAE2018 – The 10th International Conference on Applied Energy.

Keywords: LSTM for Time Series Analysis, Short Term Forecast, Long-Term Forecast, RNN for Time Series Analysis

* Corresponding author. Tel.: +971 2 810 9149.
E-mail address: afshin.afshari@ku.ac.ae

1876-6102
Peer-review under responsibility of the scientific committee of ICAE2018 – The 10th International Conference on Applied Energy.
10.1016/j.egypro.2019.01.952

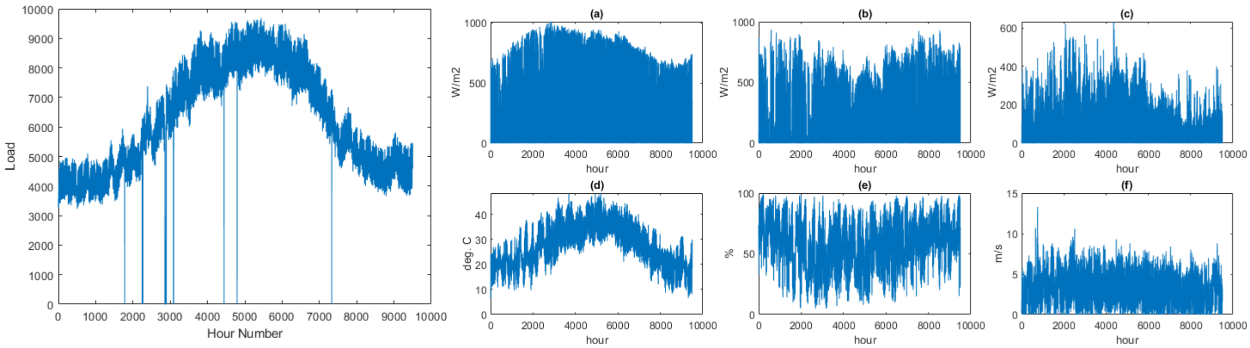

Fig. 1. Load TS (Left) and Exogenous TS (Right) (a) Global Horizontal Irradiance (b) Direct Normal Irradiance (c) Diffuse Horizontal Irradiance (d) Temperature (e) Relative Humidity (f) Wind Speed

1. Introduction

Load forecasting plays a key role in power systems operations with respect to a range of important technical and economic impacts. The analysis and management of several tasks such as market purchases/sales, day-ahead outage planning, unit commitment and economic dispatch, energy storage management, future energy contracts, power plants maintenance schedule, and portfolio structuring necessitates the fact of knowing about the upcoming demand and needs of load. Load forecasting is challenging due to several reasons such as existing daily, weekly, and annual cycles effects, and the random usage of appliances by end-users. The archives suggest that there is a significant correlation between the load data and other environmental variables, known as the exogenous/independent variables, such as temperature, humidity, irradiance, and wind speed. Such variables can be exploited for improving the forecast accuracy. The selection of the best candidate variable, toward this end, is important and may increase the undesired forecasting computational costs if selected inappropriately. Another challenge that we usually face is what model would be appropriate to incorporate both the main and exogenous time series.

To work with the aforementioned issues toward better modeling and forecast, the traditional methods such as ARMA, SARIMA, and ARMAX have been used for decades. However, there are many other methodologies which are being explored in recent years such as the artificial neural networks (ANN) and recurrent neural networks (RNN) ([5] through [11]) which have shown promises in much better forecasting as compared to traditional methods. There are also several works which have proposed the hybrid methodologies ([1] through [4]) to achieve better forecast. However, many of these are focusing only on the short-term forecast and do not provide the capability of incorporating the long-term forecasts. Towards this end, long short-term memory (LSTM) networks, a special kind of RNN, are much promising that have the capabilities to learn the long-term dependencies.

With such objectives and issues in mind, in this work, the aim is to use both the traditional and LSTM methods and concepts to analyze the electrical load data and find the best approach to achieve the better forecasts. The load data of 13 months along with exogenous variables is considered to train and validate the LSTM network. The trained network is then used to forecast at a range of forecast horizons including 25 hours, 48 hours, 7 days, and 30 days. The data is also used to model and forecast the time series with the traditional methods and are compared with the LSTM in terms of MAPE. The comparison of forecast show that the LSTM networks, given the appropriate data, outperform the traditional methods not only for short-term but also for long-term forecast.

2. Data Description

The load time series (TS) and the exogenous time series are shown in Fig. 1. The data is recorded at each hour (24 readings/day) over the time period of 13 months. Our aim is to model and forecast the main load time series. Toward this end, looking at the multicollinearity of the data, we find that there is a dependency of load over temperature and other exogenous time series. But, temperature itself is dependent on other factors as shown in Fig 1. Therefore, in the rest of the text, the temperature alone is taken as an exogenous TS. The analysis reveals that there are several components present both in the load and temperature time series. These components include the yearly, daily, and

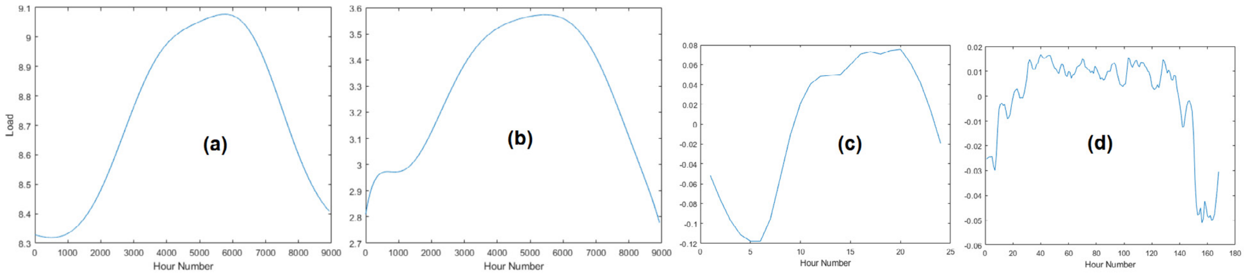

Fig. 2. Seasonality Profiles (a) Yearly Load (b) Yearly Temp. (c) Daily Load (d) Weekly Load

weekly seasonality, yearly trend, outliers, and some missing readings. It is interesting to note that temperature TS does not have outliers and weakly seasonality. Temperature does not depend on variations due to weekends but the load does. Both TS have the yearly seasonality due to change in weather and yearly trend due to several effects such as global warming. Both have daily trend due to the periodic movement of sun over the days. Before we move to the modelling part, the data needs to be prepared accordingly and TSs need to be decomposed and stationarized. Toward this end, we have adopted a step-by-step approach for analysis and decomposition of the data to enable the flexibility of picking-up and using the TS at any level of processing as per the requirement of modelling method and the analysis. The step-by-step process is given as follows.

There is variability in the number of days per month and would trouble in extracting the seasonality component of the TS. To avoid such issues and to make the data processing easy and suitable for modelling and forecasting, all the months with days less than 31 are extended by replicating the last day of the month. For example, a month with 30 days is extended to 31 days by replicating the 30th day to prepare 31st day. This conditioning is performed both on the load and temperature TSs, and at the end of this step all the months are of 31 days. Additionally, at this stage the missing data points which are represented by zeros, are replaced with infinity to facilitate outlier's management step, next subsection, to easily detect these as outliers too and take appropriate action. The outliers and missing data points present in the load TS are detected and eliminated by replacing with the new interpolated values.

Both the load and temperature TS are split into two parts, modelling data (12 months) and the validation data (13th month). Modelling data is used in all the modelling procedures, whereas, the validation data is used to verify and analyze the foretaste. From now onwards, the time series means the 12 months modelling data unless specified explicitly. The processed data at this level in this text is denoted as *DS1*. The ACF/PACF plot shows that the time series are not stationary, and we need further processing before the start of modelling process. However, further data processing is not required for LSTM except the removal of yearly seasonality because we have only one years' data to train which would lead the LSTM network learn the yearly seasonality as a continuing trend.

There is a special pattern over the year both in the load and temperature TS. This pattern is known as the yearly seasonality that is estimated using the Fourier fitting. The sum of sines and cosines is used to fit the curve and determine the model for the yearly seasonality by estimating the parameters via regression methods. The yearly seasonality components are shown in Fig. 2(a)(b). The processed data at this level in this text is denoted as *DS2*. The plots show that there is a linear trend too other than the seasonality component and is separated out using a linear equation. The processed data at this level in this text is denoted as *DS3*. There is also a special pattern over the days both in the load and temperature TS. This pattern is known as the daily seasonality. The resultant TSs of DS3 are used to estimate the daily seasonality profile for a day using the periodic averaging method. The estimated daily seasonality profiles are shown in Fig. 2(c). The daily seasonality is removed from DS3 and the processed data at this level in this text is denoted as *DS4*. As mentioned earlier, there is no weekly seasonality in temperature TS, however, the load TS does have. The resultant load TS of DS4 is used further to estimate the weekly seasonality profile for a week using the periodic averaging method. The estimated weekly seasonality profile is shown in Fig. 2(d). The daily seasonality is removed from DS4 and the processed data at this level in this text is denoted as *DS5*. The ACF and PACF of the resultant TS show that the TS are now stationary.

3. Modelling the Load and Temperature Time Series

The processed data in previous section is used to model the time series using different models including ARMA, SARIMA, ARMAX, and LSTM. ARMA and SARIMA use only the load time series to model and forecast. Whereas, the ARMAX and LSTM also use temperature TS as an exogenous variable. All the methods use 12 months data for modeling or training, and the 13th month data is used to compare and validate the forecast. The goodness of fits and the forecast uncertainties are compared. To validate the forecasts in each case, the forecasted data is converted back to the original form from the algorithmic transformed form. The logarithmic transformation at the start of data processing was used to bring the models to stability during estimation and not to remove Heteroscedasticity. Only LSTM method, as the target of this work, is explain in the following subsequent subsections. Whereas, the modeling using above mentioned methods and forecasts using these models is used to compare with the LSTM forecasts.

3.1. Training the LSTM Network

The LSTM network is supposed to learn both the long-term and short-term features of the training data. Toward this end, the type of input data has relevance to the effectiveness of learning. If the data is provided which is leading toward wrong direction or is not enough to make the features clear, the LSTM or RNN will learn accordingly and will not predict or forecast accurately. For example, the yearly seasonality in the provided data is only for one year. The LSTM will consider it a continuing trend and will predict wrongly and will lead toward zero values gradually. However, if the data of more than one year is given, LSTM can learn to predict the yearly seasonality too. Similarly, the selection of input data also contributes in deciding the accuracy of the network performance. The training process uses DS2 level of data processing that includes all kinds of seasonalities and trends except the yearly seasonality. For all the models, the data is divided into two parts. The 12 months data is used for training/modelling purpose, and the 13th month data is used for forecast validation. Whereas, for other models such as ARMA, SARIMA, and ARMAX, the DS5 level of data processing is used. The LSTM is trained with the following layer specifications and the LSTM toolbox in Matlab2018 is used to train the network.

- Input TSs = 2
- Output TSs = 2
- No. of Hidden LSTM Units = 60
- Solver = adam
- No. of Training Iterations (MaxEpochs) = 400
- Initial Learn Rate : 0.005
- Learn Rate Schedule : none
- Gradient Threshold : 1

To provide the appropriate training data, two set of vectors per load and two vectors per temperature TS are generated, *PredTrain* and *RespTrain* where PredTrain represents the predictors and RespTrain is the desired response. One PredTrain for a TS is the input TS after removing 5 last values. Whereas, one RespTrain for a TS is the input TS after removing 5 very first values. These lags of 5 are added to give more dependency on the previous values to increase the learning effectiveness. The selected optimum lag is 5 in this experiment. Each of the responses depends on previous 5 values. The effect and selection of lag, number of hidden units, and number of iterations is highlighted in the comparison section. Training data for both of the input TS is generated in the same way. Similarly, using only 13 motnh's data, the forecast test vectors are also generated. The training and forecast test data is formatted as following.

$$PredTrain = 12MonthsData\ (1 : end - lag) \tag{1}$$
$$RespTrain = 12MonthsData\ (lag + 1 : end) \tag{2}$$
$$PredTest = 13MonthData\ (1 : end - lag) \tag{3}$$
$$RespTest = 13MonthData\ (lag + 1 : end) \tag{4}$$

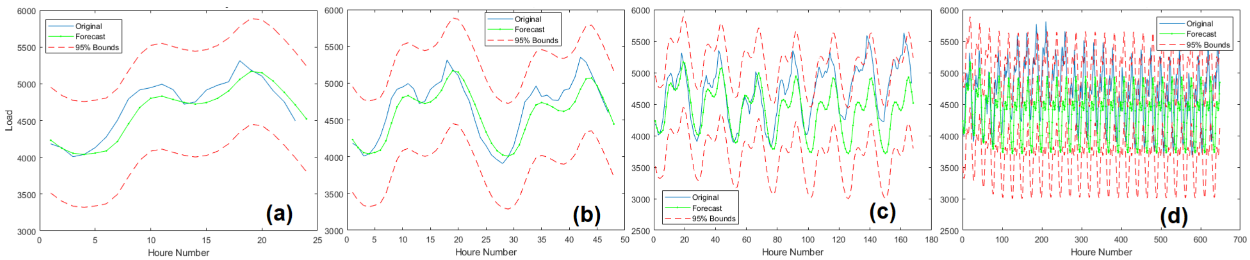

Fig. 3. LSTM Forecasts (a) 24 Hours (b) 48 Hours (c) 7 Days (d) 30 Days

3.2. Forecast

The trained LSTM network in the previous subsection is used to forecast over the horizons of 24 hours, 48 hours, 7 days, and 30 days, and the uncertainty of forecasts is measured. The forecast plots are shown in Fig. 3 and the uncertainty results are given in Table 1. 95\% bounds are also included in the plots. Looking at the RMSE and MAPE of forecasts, we can conclude that the LSTM network forecasts best as compared to other models. Contrary to other models, the pattern of MAPE continues to increase as the number of forecast hours increase.

Table 1. LSTM Forecast at Different Horizons.

LSTM Forecast Horizon	RMSE	MAPE
24 Hours	89.40	1.522
48 Hours	124.3	2.16
7 Days	374	5.97
30 Days	554.9	9.75

4. Comparison and Discussion

The forecast comparison given in Table 2 suggests that for all the forecast horizons in general and for short term especially, the LSTM outperforms others and shows a consistent behavior for the increased uncertainty with the increase in hours to forecast. As described earlier, LSTM network uses lags, reduced number of hidden units, and reduced number of training iterations. During the experiment, we have observed that increasing the number of training iterations increases the learning and forecasts keep improving. However, at certain number, the forecast accuracy maximizes for a given number of hidden units and the lag. Beyond this number, the over learning happens and the capabilities such as interpolating weakens and the network starts sticking to the input response more and more strongly, hence, errors increase. Similarly, for a given lag and number of training iterations, the forecast accuracy maximizes for a certain number of hidden units. The accuracy reduces if the number of hidden units are increased or reduced from this optimum number. Quiet similarly, if the number of hidden units and the training iterations are fixed, the dependency value of lag also has an optimum value that maximizes the forecast accuracy. The accuracy reduces if the lag is increased or reduced from its optimum number. Therefore, depending on each other, there are optimum values of these three parameters that maximize the forecast accuracy for a particular time series.

Future work includes the use of data for more than one year to achieve even better results. Moreover, this work used Matlab toolboxes and other implementations frameworks with respect to LSTM such as R and Tensor flow are of worth exploring.

Table 2. Comparison of Forecasts (MAPE) using Different Methods at Various Horizons.

Forecast Horizon	ARMA	SARIMA	ARMAX	LSTM
24 Hours	5.42	6.53	7.51	1.522
48 Hours	4.48	5.21	6.18	2.16
7 Days	3.70	4.25	4.31	5.97
30 Days	5.24	5.24	5.39	9.75

5. Conclusion

The long-term and short-term learning capabilities of the long short-term memory (LSTM) network, a special type of recurrent neural network (RNN), are suitable for the application of time series modelling and forecasting. LSTM is also simple towards handling the provided information easily instead of developing and working over complex equations as in the case of ARMAX. More and more variables can be fed to LSTM in order to improve the accuracy of forecast. Given the appropriate data with appropriate length and features, LSTM can learn all the seasonalities and trends. It has been shown in this work that LSTM outperforms the other traditional methods such as ARMA, SARIMA, and ARMAX and forecasts the load time series with reduced percentage of errors. Further improvements can be brought to LSTM forecast if data over more than one year is available from which LSTM can learn the yearly seasonality and trend as well instead of extracting these features a priori.

References

[1] A. G. Abdullah, et al. Analysis on anomalous short term load forecasting using two different approaches. International Conference on Science in Information Technology (ICSITech). Bandung, Indonesia, January 2018.
[2] I. P. Panapakidis, et al. Combining wavelet transform and support vector regression model for day-ahead peak load forecasting in the Greek power system. IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe). Milan, Italy, July 2017.
[3] P. Ray, S. Sen, A. K. Barisal. Hybrid Methodology for Short-Term Load Forecasting. IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES). Mumbai, India, December 2015.
[4] A. Abdoos, M. Hemmati, A. A. Abdoos. Short term load forecasting using a hybrid intelligent method. Knowledge-Based Systems. March 2015.
[5] X Chen, L. Wei, J. Xu. House Price Prediction Using LSTM. Computing Research Repository (CoRR). abs/1709.08432, September 2017.
[6] Malhotra, Pankaj, et al. Long short term memory networks for anomaly detection in time series. Proceedings. Presses universitaires de Louvain, 2015.
[7] L. Nikolay, J. Yosinski, L. Li, S. Smyl. Time-series Extreme Event Forecasting with Neural Networks at Uber. ICML 2017. Sydney, Australia, 2017.
[8] F. Karim, S. Majumdar, H. Darabi, S. Chen. LSTM Fully Convolutional Networks for Time Series Classification. Computing Research Repository (CoRR). abs/1709.05206, September 2017.
[9] W. Bao, J. Yue, Y. Rao. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLOS ONE. July 2017.
[10] J. C. B. Gamboa. Deep Learning for Time-Series Analysis. Computing Research Repository (CoRR). abs/1701.01887, September 2017.
[11] F. Fahiman, et al. Improving load forecasting based on deep learning and K-shape clustering. International Joint Conference on Neural Networks (IJCNN). Anchorage, AK, USA, July 2017.