LSTM-based Short-term Electrical Load Forecasting and Anomaly Correction

Lei Zhang[1,*], *Linghui* Yang[2] , *Chengyu* Gu[1], *Da* Li[3]

[1] State Grid Shanghai Electric Power Research Institute, Department of Power Source, Shanghai 200437, China
[2] State Grid Shanghai Electric Power Research Institute, General Manager Office, Shanghai 200437, China
[3] State Grid Shanghai Electric Power Research Institute, Department of Human Resource, Shanghai 200437, China

Abstract. The Emergence of the Ubiquitous Power Internet of Things(UPIoT) facilitates data sharing and service expansion for the power system. Based on the architecture of the UPIoT and combined with deep learning technology, short-term electrical load forecasting and anomaly correction could be used to improve the overall performance. Since short-term electrical loads are non-linear and non-stationary [1] and could be easily affected by external interference, traditional load forecasting algorithms cannot recognize the correlation between the time sequence thus rendering low prediction accuracy. In this article, a Long Short-Term Memory (LSTM) based algorithm is proposed to improve the prediction accuracy by utilizing the correlation between the hourly load sequence. Then, the real-time forecasting outputs are compared to the raw data in order to detect and dynamically repair the anomaly so as to further improve the performance. Experiment results show that the proposed approach outputs low Mean Square Error (MSE) of around 0.2 and could still hold it at around 0.3 with corrected data when the anomaly is detected, which proves the accuracy and robustness of the algorithm.

1 Introduction

Electrical load forecasting plays an important role in the construction of smart grid system. Since the electric load forecasting relies on a large amount of historic data, the credibility of the time series of original load sequence has a great influence on the prediction results of any prediction method. Performing real-time anomaly correction and load forecasting simultaneously could improve the reliability of the raw data thus improving the accuracy and stability of the prediction algorithm. Therefore, the study of load prediction and anomaly repair algorithm is of great significance to improving the overall performance of the smart grid.

In the 1960s, the research into load forecasting was gradually leaning to the practical application. Load forecasting methods could be mainly categorized into two types: traditional load forecasting technology and intelligent forecasting technology [2]. Traditional power load prediction algorithms include MA, ES and ARIMA [3]. Since these algorithms are established on the assumption that the observed and the future time series are linearly correlated, they are not effective when processing time series with significant nonlinear characteristics and therefore, have no learning or self-adaptive abilities. In recent years, the load forecasting technology has entered a new era of intelligent prediction and an enormous amount of research regarding machine-learning-based load forecasting has ensued gradually matured. For instance, in [4] the authors use a support vector machines based regression model coupled with empirical mode decomposition to for long-term load forecasting. In [5] electricity demand is forecast using a kernel based multi-task learning methodologies. In [6], the authors used Artificial Neural Network (ANN) ensembles to perform the building level load forecasting. In [7, 8], Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) based algorithms are proposed and varied to achieve better overall prediction results. Though enormous number of approaches have been proposed to improve the prediction results, little research has been done to diagnose the reliability of the training data which could influence the forecasting performance to a great extent if an anomaly exists in the historical time series.

This article proposes an approach based on a four layer LSTM model for short-term load forecasting and anomaly correction. LSTM fully considered the correlation of the time series data and perfectly solved the “gradient vanishing” problem of the RNN model. In addition, by comparing the prediction results with the real data, anomalous load data in the historical time series can be detected and corrected before the next prediction iteration. In this sense, the forecasting system could not only provide accurate prediction results, but also warn the power enterprise to detect abnormal conditions of power load in time which significantly increase the practicability of the algorithm. Experiment results show that the mean square error(MSE) between the prediction and the real data is around 0.2 and has a higher prediction

[*] Corresponding author: lei.zhang8@mail.mcgill.ca

accuracy compared to RNN or other traditional algorithms. In addition, when the anomaly is induced, the MSE between the prediction results using the repaired data and the real data is still around 0.3, which proves the robustness and stability of the proposed approach.

2 An outline of the LSTM

As the simplest type of neural network, Perceptron was developed in the 1950s and 1960s by Frank Rosenblatt [9], inspired by earlier work by Warren McCulloch and Walter Pitts. Perceptron comprises only one input layer and one output layer. Following that, more complex models containing multiple hidden layers have merged. In this work, we mainly focus on the application of the back propagated neural network designed for training time series, i.e., the input of a hidden includes the output of the previous layer at this moment along with its own output at the previous moment. Typical algorithms include RNN, LSTM, etc.

2.1. The RNN architecture

RNN was first proposed by Jeffrey l. Elman in 1990 [10]. Since RNN is more capable of using the information from the past to predict the future, it is widely used in natural language processing, time series processing, voice processing and other time series related fields. A standard RNN architecture is shown in Fig. 1 where the transmission process for RNN is unfolded by time period. $x(t)$ represents the input at time t, $o(t)$ represents the output at time t, and U, V and W denotes the input, output at time t and the weight output at time $t-1$, respectively. Each arrow represents one transformation, which means that the arrows are attached with weights. As shown in the unfolded architecture of RNN, the neurons in between the hidden layer are also connected with weights in standard RNN structure which implies that the status of a hidden layer at a previous moment would affect the status of itself at a later moment as time goes on.

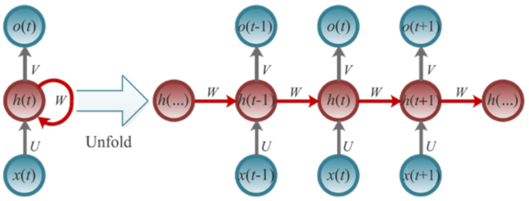


Fig. 1. Unfolded RNN architecture

The data flow of the standard RNN also possesses the characteristics of: 1) weight sharing, i.e., the U, V and W in the figure share the same value. 2) each input only establishes a weighted connection with its own route rather than other neurons. Based on the above characteristics, the output of RNN at time t can be expressed as:

$$o(t) = \mathrm{softmax}(Vh(t) + c) \tag{1}$$

$$h(t) = f\left(Ux(t) + Wh(t-1) + b\right) \tag{2}$$

where $f(\cdot)$ is the activation function of the hidden layers, generally tanh or sigmoid function.

Although RNN is able to learn the data correlation in time, it suffers from the long-term dependency problem, i.e., when the time interval between relevant information and the prediction position grows large, the historical data cannot be traced due to "gradient vanishing" or "gradient exploding", resulting in the loss of learning ability .

2.2 The LSTM architecture

In order to solve the problems of "gradient vanishing" and "gradient exploding" for RNN, Hochreiter & Schmidhuber [11] proposed a variant of RNN named as LSTM in 1997, which could learn long-term information and is widely implemented after improvement and promotion. As shown in Fig. 2, LSTM also contains a chain structure of recurrent modules. In standard RNN, the recurrent module is only composed of a simple structure such as a tanh layer. Alternatively, The LSTM module is composed of three different structures, namely, gated cell where gated means that the cell decides whether or not to store or delete information (e.g. if it opens the gates or not), based on the importance it assigns to the information. The assigning of importance happens through weights, which are also learned by the algorithm. This simply means that it learns over time which information is important and which not. A typical LSTM has three different types of gates: input, forget and output gate. These gates determine whether or not to let new input in (input gate), delete the information because it isn't important (forget gate) or to let it impact the output at the current time step (output gate).

With the evolution of neural networks, more popular variants of LSTM were invented such as peephole connections [12] and Gated Recurrent Unit (GRU) [13], which are widely used in various learning tasks related to time series. **NOTE** that the proposed approach in this article is based on the most basic LSTM structure.

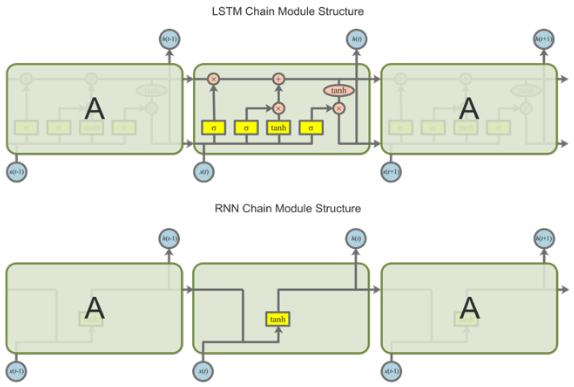


Fig. 2. LSTM and RNN Chain Module Structure

3 Using LSTM for load forecasting and anomaly correction

3.1. Load forecasting

Electrical load forecasting is a kind of prediction based on the historical load data. Normally, before conducting load forecasting, a large amount of historical data should be collected. The data should be wisely chosen since the more standard and representative the training data, the more accurate the predicted value could be obtained. After that, the original training data need to be normalized and reshaped to accelerate the convergence and the fit the model dimension. Then the model would be trained with appropriate parameters. After the model is trained, the predicted value could be obtained by passing the historical load data to the trained LSTM model. The complete training process of load forecasting using LSTM is shown in Fig. 3.

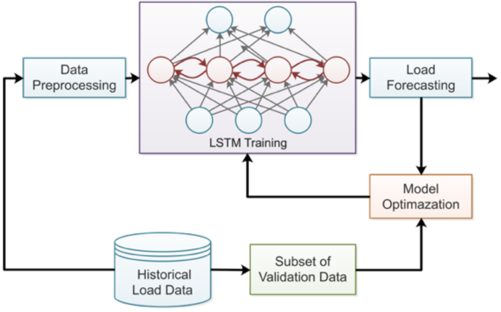


Fig. 3. The training process of load forecasting using LSTM

3.2 Anomaly correction

The anomalous data would affect the prediction results negatively since the pattern of the load sequence is destroyed and the prediction model could no longer recognize the relationship between data. Since LSTM model is distinct for its strong robustness, a small amount of anomaly will not have a serious impact on the predicted results. Therefore, the predicted results can be used to identify anomaly and correct it by the prediction results. Common anomalous data could be categorized into 3 types [14]:

- **Missing data**: Mainly due to the malfunction of the SCADA system, data in the database appears as null values.
- **Deficient data:** Mainly due to the line maintenance, equipment maintenance or electric meter damage in a certain period of time. The load curve increases and decreases drastically in a day or a period of time compared with the normal load of the adjacent day or the planned load of the day.
- **Minimax data:** The load exceeds the peak or valley value in a day dramatically in a non-peak or non-valley moment.

4 Numerical results

Experiments are conducted using Keras library in python.The dataset is the hourly power consumption in Toronto Canada from January to July 2016. After the LSTM model is trained, standard dataset is used for testing, and the accuracy of the prediction results was analyzed in terms of MAPE and MSE. In order to test the anomaly correction results for the proposed approach, corruption of the standard data (missing values, minimax values, etc.) is introduced artificially to perform prediction under anomaly situations. The prediction results are then compared with the real value at the that moment to determine whether the anomaly needs to be fixed. Finally, the prediction results before and after correction are analysed and a conclusion is drawn. The complete diagram of the proposed approach is shown in Fig. 4.

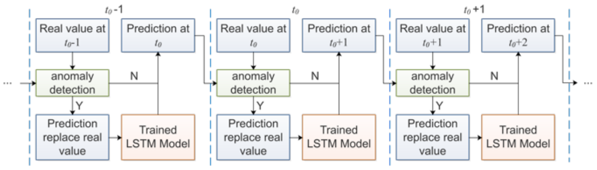


Fig. 4. Complete diagram of the proposed approach

4.1 System setup

Raw data is standardized by subtracting the mean of the original data from all the training data. Since the proposed approach uses the previous 22 load values to predict the 23rd load value, the 22 values before each moment are set to be the training features, and the load value at that moment is set to be the training label.

Since the correction results of the missing data and deficient data are similar, only the missing data case and the minimax case will be discussed in the following context. For the construction of missing data, the load value for a random period of time is assigned to be 0 in the test data. In this experiment, five consecutive random time series in the test dataset are set to be 0 while the others remained unchanged. For the construction of the minimax data, a random time point is selected from the test data. Then the previous moment of this time is set to be a random large number, and the later moment of the time point is set to be a random small number.

4.2 Prediction results for uncorrupted data

Fig. 5 shows the prediction results for uncorrupted test data (487 sets in total). Intuitively, the prediction result is very close to the real data. In order to statistically evaluate the accuracy of the prediction, Mean Absolute Percentage Error (MAPE), Mean Square Error (MSE) and Root Mean Square Error (RMSE) of the prediction and the real data are calculated and compared with RNN prediction results. The results are shown in Table 1.

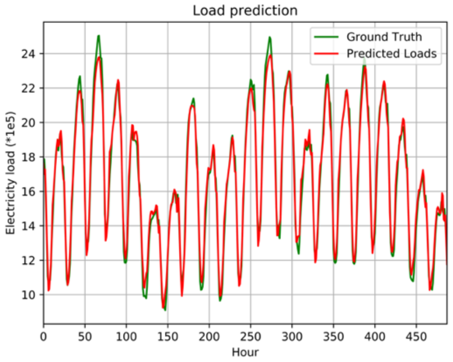


Fig. 5. Prediction results for uncorrupted data

Results shows that the overall performance of LSTM model is better than the RNN with both high prediction accuracy and short computational time.

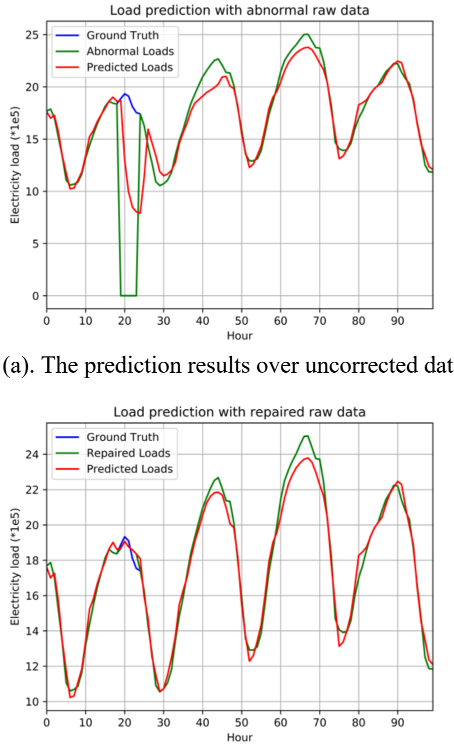
Table 1. Prediction results for LSTM and RNN

Model	MSE	RMSE	MAPE	CPU time
LSTM	0.256	0.506	2.639	0.023s
RNN	0.350	0.591	3.712	0.025s

4.2 Prediction results for corrupted data

4.2.1 The prediction results for missing data

Fig. 6(a) and 6(b) shows the prediction results over the uncorrected anomalous data and the corrected anomalous data, respectively. Obviously, the missing data has greatly affected the accuracy of prediction in period of time when corruption occurred. However the influence of corruption is getting weaker as the time sequence deviated from anomaly point which indicates that the proposed algorithm has fair anti-interference ability. Furthermore, when the dataset is fixed, it can be seen that the difference between the corrected data and the ground truth data is relatively small and the prediction results are more accurate than the results without correction, especially the results in the time period with missing data.



(a). The prediction results over uncorrected data

(b). The prediction results over corrected data

Fig. 6. The prediction results for missing data

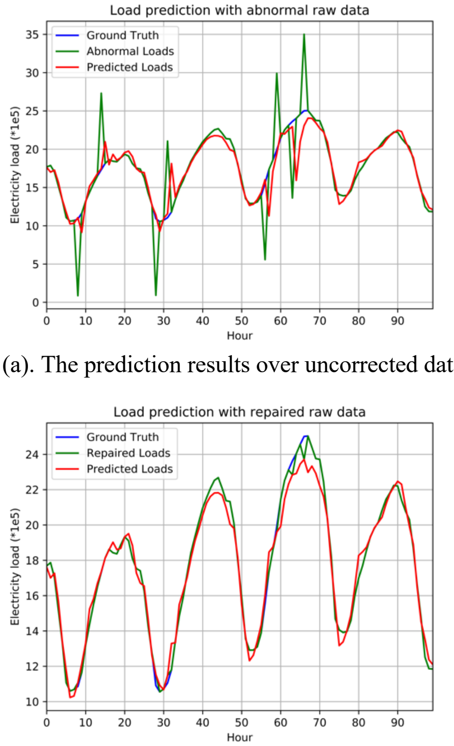
Numerical results are also presented in Table 2 to further discuss the performance. As shown in Table 2, the proposed anomaly correction algorithm results in smaller MSE and MAPE but longer CPU time due to the dynamic correction during prediction. In addition, although the prediction accuracy over the uncorrected data has decreased, it has not decreased significantly, which also proves the robustness of the system from another perspective, i.e., certain amount of interference does not cause serious damage to the stability of the system.

Table 2. Prediction results for missing data

Data	MSE	RMSE	MAPE	CPU time
Standard	0.256	0.506	2.639	0.023s
Uncorrected	1.214	1.101	3.433	0.019s
Corrected	0.269	0.519	2.678	0.073s

4.2.1 The prediction results for minimax data

Fig. 7(a) and 7(b) shows the prediction results over the uncorrected minimax data and the corrected minimax data, respectively. Similar to the missing data senecio, when the dataset is corrected, the difference between the corrected data and the ground truth data is relatively small and the prediction results are more accurate than the results without correction. However, it can be intuitively seen that the influence of minimax value on prediction is less than that of missing value. This phenomenon could possibly be explained by the sensitivity of the LSTM model to the correlation of continuous time series. Since the missing data is a series of anomaly in continuous time while the minimax data is just some discrete points over time, the damage of the data correlation is more serious for missing data which result in inferior prediction performance.



(a). The prediction results over uncorrected data

(b). The prediction results over corrected data

Fig. 7. The prediction results for minimax data

Numerical results are also presented in Table 3 to further discuss the performance. As shown in Table 3, the proposed anomaly correction algorithm results in smaller

MSE and MAPE and is even smaller than the results for missing data which further illustrates the sensitivity of the proposed model to data correlation.

Table 3. Prediction results for minimax data

Data	MSE	RMSE	MAPE	CPU time
Standard	0.256	0.506	2.639	0.023s
Uncorrected	0.550	0.741	2.912	0.025s
Corrected	0.271	0.520	2.644	0.054s

5 Conclusions

Short-term electricity load is hard to predict due to its nonlinearity and non-stationarity. However, the traditional non-learning algorithm cannot capture the correlation of load series in the time dimension rendering low prediction accuracy. In this paper, a LSTM-based algorithm is proposed and tested for short-term load prediction and anomaly correction on the basis of a large amount of existing research. The results show that the proposed approach could learn the correlation of load series and output accurate predictions with a short convergence time. In addition, when anomalous data such as missing data and the minimax data appears, the algorithm is able to detect the occurrence of anomaly and repair the anomaly adaptively. The error between the prediction result over corrected data and the standard data is small, which indicates that the proposed model has the advantage of high accuracy, stability and practical significance.

Furthermore, the algorithm of the prediction results over minimax data is better than that of the missing data in general which means that the algorithm has certain sensitivity for continuous anomaly sequence over time and cannot adapt to the data distortion over a long period of time. Therefore future research needs to be carried out into how to better improve the proposed model to make it less sensitive to long time data corruption is worth thinking. Finally, due to the existence of dynamic anomaly correction, the processing time of the algorithm is increased. How to improve the convergence speed remains to be studied.

References

1. Q. W. Ran, Y. Z. Shan, Q. Wang, et al., Wavelet neural network-PARIMA method for short-term load forecasting of power system, *Chinese journal of electrical engineering* **23(3)**, 38-42 (2003)
2. Y. Du, Z. Guo, L. Z. lu, et al., Overview of short-term power system load forecasting methods, *Technology and market* **5**, 339-340 (2015)
3. D. Alberg, M. Last, Short-Term Load Forecasting in Smart Meters with Sliding Window-Based ARIMA Algorithms, (2017)
4. L. Ghelardoni, A. Ghio, D. Anguita, Energy Load Forecasting Using Empirical Mode Decomposition and Support Vector Regression, *IEEE Transactions on Smart Grid* **4(1)**, 549-556 (2013)
5. J. B. Fiot, F. Dinuzzo, Electricity Demand Forecasting by Multi-Task Learning, *IEEE Transactions on Smart Grid*, 1-1 (2016)
6. J. G. Jetcheva, M. Majidpour, W. P. Chen, Neural network model ensembles for building-level electricity load forecasts, Energy and Buildings 84, 214-223 (2014)
7. J. Ku, R. Goomer, A. K. Singh, Long Short Term Memory Recurrent Neural Network (LSTM-RNN) Based Workload Forecasting Model For Cloud Datacenters, Procedia Computer Science 125, 676-682 (2018)
8. Y. H. Zhang, C. M. Qiu, X. He, et al., A Short-Term Load Forecasting Based on LSTM Neural Network[J], Electric Power Information & Communication Technology, (2017)
9. F. Rosenblatt, Perceptron Simulation Experiments, Proceedings of the Ire **48(3)**, 301-309 (1960)
10. J. L. Elman. *Finding structure in time*, Cognitive Science **14(2)**, 179-211 (1990)
11. S. Hochreiter, J. Schmidhuber, *Flat Minima* **9(1)**, 1 (1997).
12. G. Gelly, J. L. Gauvain, Optimization of RNN-Based Speech Activity Detection, *IEEE/ACM Transactions on Audio Speech & Language Processing* **99**, 1 (2017).
13. K. Kang, H. B. Sun, C. K. Zhang, et al., Short-term electrical load forecasting method based on stacked auto-encoding and GRU neural network, *Evolutionary Intelligence*, (2019)
14. Z Y Tian, M L Zhang, R Zhao. Identification and correction of abnormal load data in short-term power load forecasting. *Jilin electric power* **6**, 25-27 (2004