Alice: A LSTM Neural Network Based Short-Term Power Load Forecasting Approach in Distributed Cloud-edge Environment

Zhengyang Ding¹, Jijun Wang², Yiqing Cheng² and Chengxun He³,*

¹State Grid Jiangsu Electric Power Co., LTD. Information & Techecommunication Branch, Nanjing, China
²Jiangsu Electric Power Information Technology Co., LTD., Nanjing, China
³School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China

*Corresponding author email: 002631@nuist.edu.cn

Abstract. Load forecasting, as the baseline for decision-making, plays a key role in the management and control of the grid. Nevertheless, the rapid evolution of the smart grid has brought a dramatic increase in the volume of user-side data, traditional load forecasting approaches have to face the challenge of ensuring the accuracy of dynamic forecasting under the circumstance of the widespread application of big data. Meanwhile, the advance of the Industrial Internet of Things (IIoT) enables smart meters to acquire more plentiful data, which improves the accuracy of short-term load forecasting with appropriate utilization, but the gradual increase in the magnitude of data brought by IIoT technique has also left the computing equipment under great pressure. To address these challenges, in this paper, a Long Short-Term Memory (LSTM) network based short-term load forecasting approach deployed in the distributed cloud-edge environment, named Alice, is devised to deliver more precise results for the smart forecasting of power load. It adopts the LSTM network to perform the forecasting tasks and extends the whole system to the cloud-edge platform to implement parallel neural computing. Moreover, the Toeplitz Inverse Covariance-Based Clustering (TICC) algorithm is invited to enhance the efficiency of LSTM. Eventually, experimental evaluations on real dataset elaborate the superiority of the proposed Alice approach when compared to other state-of-the-art approaches.

1. Introduction

Accurate short-term load forecasting is an inevitable condition for the stable running of power station and the scientific management of smart grid. It is of vital significance to improve the utility of power generation equipment and guide the dynamic planning, generation control, and long-term development of smart grids [1]. The elementary procedure of short-term load forecasting is to estimate future values by analysing past values, its research object is an uncertain and complex stochastic process [2]. In recent years, compared with the traditional short-term load forecasting approaches, the participation of artificial intelligence (AI) approaches e.g., LSTM neural network has improved the accuracy of forecasting once again. Kong et al. [3] proposed a novel approach based on LSTM network, achieved bright results in lifting the performance of forecasting.

Riding with the fast-evolving cloud computing and IIoT techniques, a large amount of IoT smart devices e.g., smart meters are deployed inside the power grid system. The continuous application of
smart meters has accelerated the evolution of traditional power grids to smart grids [4]. The smart meters obtain the user's power consumption data according to a certain frequency and offload them to the cloud computing center in real-time to perform forecasting tasks, which leads to the improvement of processing efficiency [5]. Aiming to solve the ultra-large scale power load forecasting, a data mining approach for the distributed load forecasting approach based on cloud computing and hybrid gene expression programming is presented by Deng et al. [6], which confirmed that the combination of cloud computing and smart algorithm develops the ability to improve the potency of load forecasting.

At the same time, AI is decentralizing, with the cloud and the edges are dividing gradually. The migration of AI from the cloud to the edges is an inevitable process in many real application scenarios with the rapid evolution of edge computing philosophy. Consequently, a new solution, the cloud-edge collaborative computing architecture for AI, is burgeoning and thriving [7]. Lai et al. [8] conducted a research through combining LSTM neural network with edge computing, to analysing the characteristics of industrial big data, which highly improved the recognition efficiency of smart electrical equipment. Design a novel AI computing architecture for smart short-term load forecasting in the cloud-edge parallel computing environment, further enhance the efficiency of data computing, is the right direction for the next generation of large-scale short-term power forecasting approaches.

The main contributions of this paper are summarized as:

- The overall system architecture of short-term power load forecasting model is constituted firstly in the distributed cloud-edge computing environment.
- The LSTM neural network based forecasting approach enhanced by the TICC clustering algorithm is devised.
- The experimental evaluations and comparison analysis are conducted to demonstrate the accuracy and efficiency of our designed approach Alice.

2. System Design
Currently, the exploding of IIoT and edge computing plays an irreplaceable role in the evolution of the traditional grid to the modern smart grid. Contrary to cloud computing, with the support of distributed architecture, edge computing reduces the latency caused by the centralized transmission. Better fault tolerance and scalability make edge computing more suitable for distributed load forecasting.

The smart power load forecasting system is deployed in a distributed cloud-edge collaborative environment as shown in Figure 1.

![Figure 1. System architecture for LSTM neural network based power load forecasting in cloud-edge environment.](image-url)
Load data $X = \{x_1, x_2, ..., x_N\}$ generated by different electricity users e.g., industry, education, government affairs, or families are collected by smart meters and offloading to the edges of the network in real time, where $N$ is the amount of raw data. The edge computing units analyze these data and offloading the outsize data to the powerful cloud computing center for forecasting. The smaller load data are processed inside the edge and migrated the forecasting results to the cloud once the forecasting is done. As thus, the efficiency of load forecasting, the utility of each equipment, and the energy consumption of the overall system are improved significantly.

Meanwhile, inside the cloud and edges, an improved LSTM neural network based approach is designed for load forecasting. Aiming to reduce the similarity of feature scale between sequence data, accelerate the training speed of LSTM neural network and improve the forecasting accuracy of load data, the primary business is to divide the multivariate time series of $X$.

3. Methodology

3.1. Data Clustering by Employing TICC Algorithm

Similar repeating patterns exist in the multivariate time series commonly, therefore, seemingly complex datasets could be accounted as time series with a thin amount of states by exploring these patterns. Subsequence clustering algorithm is a practical way to seek these repeated patterns [9]. If the time series with similar characteristics of electricity consumption are segmented and clustered before the load forecasting, the accuracy and efficiency of the load forecasting would be improved certainly. Therefore, the TICC algorithm [10] is employed to cluster the multivariate time series data $X$ collected by the smart meters. Aiming to cluster $N$ observations into $K$ categories, the TICC problem is defined as

$$\arg\min_{\theta \in M, \mu} \sum_{i=1}^{K} \left( \| \alpha \circ \mu_i \|_1 + \sum_{X_t \in \mathcal{P}_i} (-\ell(X_t, \theta_i) + \beta) \right)$$

where $M$ is the Toeplitz matrices set of symmetric block, $\alpha$ and $\beta$ are the regularization parameters. $\| \alpha \circ \mu_i \|_1$ is an $l_1$-norm penalty coefficient to encourage a sparse inverse covariance. $\ell$ is the log likelihood.

In order to solve this highly non-convex maximum likelihood problem, each $x_i$ are assigned to diverse clusters by employing linear programming functions firstly. Then, the Toeplitz lasso problem is calculated by employing the well-known alternating direction method of multipliers (ADMM). After that, we cluster the time sequence data to updating the cluster parameters iteratively. The clustered result set $Y = \{y_1, y_2, ..., y_K\}$ is output eventually.

3.2. Data Rectification and Normalization

In the course of load data collected by smart meters, sometimes anomalies and missing occur. If all data are directly taken as the input of the algorithm without preprocessing, not only the training time and computing pressure will be increased, but also affect the accuracy of the forecasting, even result in massive interference to the model. Therefore, in order to avoid the above problems, it runs preprocessing for these factors that affect the forecasting process after the clustering process is indispensable. The primary is the analysis and correction of the raw data. If the selected data is bad data, weight correction is required as

$$\tilde{x}_d^h = \xi x_d^{h-1} + \psi x_d^{h+1} + \zeta x_d^h$$

where $x_d^h$ is the value of $h$-hour of $d$-th day, $\xi, \psi$ and $\zeta$ are weight coefficients. In order to promote the visualization of variables and measurement of indicators, numerical data are required to be normalized after the outlier processing, thus the obtained $n$-th ($n \leq N$) load data values are unified between 0 and 1, which is defined as follows.

$$\tilde{x}_n = \frac{x_n - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$
After the rectification and normalization, load data are divided into training sets and testing sets. Next step is inputting training sets to LSTM network for model generation and results forecasting.

### 3.3. Load Forecasting by Employing LSTM Network

LSTM network is a very suitable approach for the time series forecasting in deep learning algorithms. Having overcome the dilemma of gradient disappearance or explosion in the Recurrent Neural Network (RNN) in mind, as shown in Figure 2, LSTM redesigns the memory unit while maintaining the basis structure of the RNN [11].

Through the sophisticated setting of the control gates, the input sequence is guaranteed to be continuously superimposed by the hidden layer in the new time state. The previous information can continue to transmit backward without fear disappearing, which makes LSTM have delightful long short-term memory capabilities, and provides better efficiency for the massive data processing of long-term time sequences. Therefore, LSTM grows excellent forecasting power for the data with distinct cyclical trends such as the power load data. Each LSTM cell contains 3 disparate gates, i.e., forget gate $f_t$, input gate $i_t$ and output gate $o_t$, which are defined as follows.

$$
\begin{align*}
    f_t &= \sigma(w_f[C_{t-1}, h_{t-1}, x_t] + b_f) \\
    i_t &= \sigma(w_i[C_{t-1}, h_{t-1}, x_t] + b_i) \\
    o_t &= \sigma(w_o[C_t, h_{t-1}, x_t] + b_o)
\end{align*}
$$

(4)

The main purpose the forget gate exists, is to decide which information needs to be retained and which information needs to be updated. A sigmoid activation function $\sigma$ is employed to constrain the output value between [0,1]. When the output of the forget gate equals to 0, all the information of the antecedent state is discarded. On the contrary, all the information of the previous state is retained. After passing through the forget door, both input gate and output gate are passed successively, add the obtained result to the hidden state at last.

At each moment, the LSTM unit records long-term dependencies between time series data and captures features from adjacent information. The memory state of candidate moment and the current moment are as follows:

$$
\begin{align*}
    \tilde{C}_t &= \tanh(w_c[h_{t-1}, x_t] + b_c) \\
    C_t &= f_tC_{t-1} + (1 - f_t)\tilde{C}_t
\end{align*}
$$

(5)

After clustering and normalization, each sequence can be regarded as load oscillation characteristics and trend characteristics affected by diverse factors. Divide these sequences into the training set and testing set, input to the LSTM network, and concentrate the results of each sequence to acquire the final result.
4. Experiment Evaluation

4.1. Dataset
We evaluated the performance of the presented approach in the dataset contains electricity load data of a city for 3 months, the sampling period is one time per one hour. The Back Propagation neural network (BP) and basic LSTM neural network are employed as the benchmark approaches. Benchmark approaches are conducted in cloud platform only. All the power load data are input into these neural networks and compare the forecasting results. The parameter setting of LSTM neural network training model will affect the convergence of network. The initial weights are set to a random distribution between [0,1]. Learning rate, number of hidden layers, batch size and other parameters are fine-tuned according to the specific performance.

4.2. Performance Evaluation
Figure 3 shows the results of the power load forecasting by different approaches within 24 hours of a day. Figure 4 shows the corresponding absolute error of the forecasting results by each approach. According to the relative position of each curve in the figures, it appears that in contrast with BP neural network and basic LSTM neural network, our proposed approach Alice, which combines the cloud and edge computing, enhanced by the TICC algorithm, achieves higher accuracy, and the error of different time is relatively stable. The basic LSTM network performed less well, and the BP network performed worst.

Then, we conducted experiments on forecasting the power load results for the next 5 days. The mean absolute percentage error (MAPE), the root-mean-square deviation (RMSE) are employed to evaluate the performance of each approach. The smaller MAPE and RMSE get, the higher accuracy of the approach is, and the better performance of forecasting accomplish. The details of each indicator are depicted in Figure 5 and Figure 6. It appears that the MAPE value of our proposed approach is vibrated between 2.5 and 3.5, and the RMSE value is about 30, which is a remarkable improvement in contrast with the two comparison approaches.

Figure 3. Comparison of Load value by different approaches.  
Figure 4. Comparison of absolute error by different approaches.  
Figure 5. Comparison of MAPE by different approaches.  
Figure 6. Comparison of RMSE by different approaches.
5. Conclusion and Future Work

In this paper, a generalized framework for the short-term power load forecasting problem is put forward. First, the distributed cloud-edge computing architecture is devised for the load forecasting model. Then, we explored the LSTM neural network for the training and forecasting by using power load data. Moreover, aiming to further improve the LSTM neural network, the TICC clustering algorithm is extended to our work. Eventually, experimental evaluations and comparison analyze are conducted. The forecasting results verified that our proposed approach Alice is better than the comparison approaches. For future work, more sophisticated network will be designed to further improve the accuracy of load forecasting.

Acknowledgments

This work is supported by the research on key Technologies of Situational Awareness of Big Data Information Security in complex Network environment (J2018021).

References

[1] Chen, K., Chen, K., Wang, Q., He, Z., Hu, J., He, J. (2018). Short-term load forecasting with deep residual networks. IEEE Transactions on Smart Grid, 10(4):3943-3952.
[2] Molzahn, D. K., Dörfler, F., Sandberg, H., Low, S. H., Chakrabarti, S., Baldick, R., Lavaei, J. (2017). A survey of distributed optimization and control algorithms for electric power systems. IEEE Transactions on Smart Grid, 8(6):2941-2962.
[3] Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE Transactions on Smart Grid, 10(1):841-851.
[4] Wang, Y., Chen, Q., Hong, T., Kang, C. (2018). Review of smart meter data analytics: Applications, methodologies, and challenges. IEEE Transactions on Smart Grid, 10(3):3125-3148.
[5] Bera, S., Misra, S., Rodrigues, J. (2014). Cloud computing applications for smart grid: A survey. IEEE Transactions on Parallel and Distributed Systems, 26(5):1477-1494.
[6] Deng, S., Yuan, C., Yang, L., Zhang, L. (2018). Distributed electricity load forecasting model mining based on hybrid gene expression programming and cloud computing. Pattern Recognition Letters, 109:72-80.
[7] X. Xu, C. He, Z. Xu, L. Qi, S. Wan and M. Z. A. Bhuiyan, Joint Optimization of Offloading Utility and Privacy for Edge Computing Enabled IoT, IEEE Internet of Things Journal, vol. 7, no. 4, pp. 2622-2629, April 2020, doi: 10.1109/JIOT.2019.2944007.
[8] Lai, C., Chien, W., Yang, L. T., Qiang, W. (2019). LSTM and edge computing for big data feature recognition of industrial electrical equipment. IEEE Transactions on Industrial Informatics, 15(4):2469-2477.
[9] Sun L, Wu F, Zhan T, et al. Weighted Nonlocal Low-Rank Tensor Decomposition Method for Sparse Unmixing of Hyperspectral Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020, 13: 1174-1188.
[10] Hallac D, Vare S, Boyd S, et al. Toeplitz inverse covariance-based clustering of multivariate time series data[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017: 215-223.
[11] Greff K, Srivastava R K, Koutnik J, et al. LSTM: A search space odyssey[J]. IEEE transactions on neural networks and learning systems, 2016, 28(10): 2222-2232.