TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 6071 - 6076

Short Term Load Forecasting Using LSTM Neural Networks

Venkates Warlu Gundu, Sishajp Simon

Article Info
Volume 82
Page Number: 6071 - 6076
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 29 January 2020*

Abstract:
Abstract— Electricity load forecast plays an important role in smart grid applications to promote the decision-making process for energy generation and consumption. Long-term forecasting is not feasible as there may be uncertainty in the forecast owing to an increase in the inclusion of renewable sources into current grids. Since the behavior of the load is highly non-linear and seasonal, Neural Networks is the best model for studying non-linear behavior within data and for forecasting purposes. Hence this paper presents an enhanced Long Short-Term Memory (LSTM) neural network model, which is used to forecastthe closing electricity load for the future interval. This paperinvestigates the performanceanalysis for optimal selectionof layers withhidden units' combination andalso discussesstatistical analysis for selection of the optimal LSTM architecture. Finally, deploy the best-adapted configuration with the lowest absolute percentage error and optimized network architecture.

Keywords: *Long Short-Term Memory,Load Forecasting, Mean Absolute Percentage Error.*

I. INTRODUCTION

Residential electricity consumption (REC) has increased by 50 times since 1971 and it is expected to grow further due to rapid electrification, increasing household incomes, and technology development [1]. The electricity demand has been increasing worldwide because of the growing population, urbanization and industrialization [2]. According to the International Energy Agency (IEA) [3-4] World energy-related CO2 emissions rise from 32.2 billion metric tons in 2012 to 35.6 billion metric tons in 2020 and to 43.2 billion metric tons in 2040.Renewable energy sources are progressively being integrated into power systems with a view to reducing CO2 emissions and relying on fossil fuels. In addition, the use of low-carbon technologies, such as electric vehicles and heat pumps, has risen in the latest years.This leads to drastic modifications in the conduct of the energy system in terms of system dynamics, generation of displacement and load profile features, which presents renewed difficulties for efficiency [5-6]. In recent days with the encouragement of incentive schemes, many residential customers installed domestic rooftop solar units. This residential generation leads to sudden surges in the distribution systemdue to its weather depended nature. Distribution system connects all consumers rich to poor, urban to suburban and then to rural [7], [8-9] stated that a feeder connected to the different behavior of customers. Based on their behavioral changes load pattern also varies, hence at any time demand may exceed its maximum demand.Hence load forecasting play an important role in the present distribution system.The literature reported in [10-12] presented day ahead price-based predictions of DR algorithms for planning home energy operations optimally, however minimizing forecasting error is a challenging issue in an efficient DR strategy for a day. Hence accurate day-ahead forecast is an important tool for economic power generation [13]. IEX (Indian Energy Exchange] follows 15-minute time block price format in day-ahead market and term ahead market for efficient use of energy

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 6071 - 6076

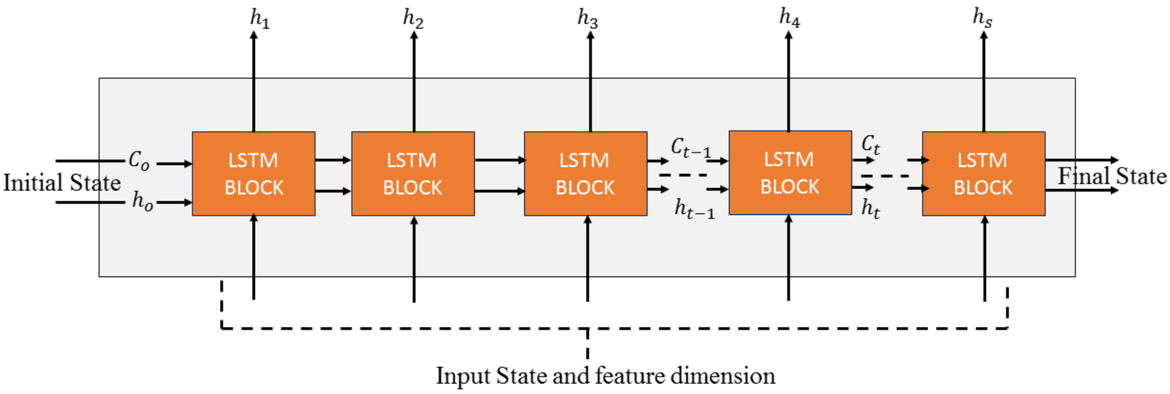


Figure.1. LSTM architecture

resources and to operate the energy market transparent, competitive and reliable, for the benefits of all stakeholders [14]. Whereas it is expected that same can be extended to the residential customers in the future smart grid environment.

Data forecasting in the time sequence is always a key technique in a variety of key issues, including weather predicting, transport planning, traffic supervision and different areas of society [15]. Forecasting and Time sequence analysis have been intensively studied for 40 years [16]. The AutoregressiveIntegrated Moving Average (ARIMA) model and Support Vector Regression (SVR) was effectively applied to forecast time sequence [17-18].However, it also has its drawbacks such as the absence of structured means to determine certain mainparameters of the model.Due to the flexible framework, in recent years, deep learning models are progressively being used in time sequence forecasting.

In particular, one of the deep learning models, Recurrent Neural Networks (RNNs), Establishes a reputation for coping with time sequence by recurring neural connections. For any conventional architecture of RNN, the effect of the input on the hiddenlayers and finally, when cycling around recurring connections, neural network output would either decrease or setback up exponentially[19].

To addressissues with RNN, LSTM have been built revolutionarily by varying the framework of the hidden neurons in traditional RNN [20]. LSTMapplications areproliferate for the forecast of time series data. To illustrate this, Wang et al. [21] used an LSTM model to forecast the subsequent moment traffic load in a particular geometric area, and Alahi et al. [22] predicted movement dynamics in crowded scenes based on LSTM. Hence this research work attempts to implement LSTM based neural network for short-term electricity load forecasting in the present smart grid networks.

II. Proposed LSTM Architecture

Long term memory-based networks, usually stated to as "LSTMs," are a superior class of RNNs adequate for long term dependency learning. Aelement called the memory block is the key component that improves the ability of LSTMs to model long-term dependencies [17]. As shown in Fig.3, the memory

block is a repeatedly connected subnet with functional segments called the memory cell and gates. The memory cell is responsible for memorizing the neural network's chronological status and multiplicative unit gates are responsible for monitoring the data flow pattern. Conversing to the relevant practical features these gates are characterized as the input gate, forget gate and output gate The input gate(I_t) regulates the amount of information that enter in to the memory cell, while the forget gate(F_t) directs the memory cell

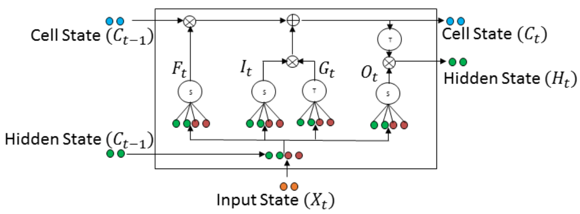


Figure.2. LSTM Cell Structure

January - February 2020
ISSN: 0193 - 4120 Page No. 6071 - 6076

how much information still remains in the present memory cell through recurring connection and the output gate(O_t) determines how much data is used to calculate the memory cell's output activation and further flows to the rest of the neural network.

$$I_t = \mathrm{Sigmoid}(W_{IH} H_{t-1} + W_{IX} X_t + b_I)$$
$$O_t = \mathrm{Sigmoid}(W_{OH} H_{t-1} + W_{OX} X_t + b_O)$$
$$F_t = \mathrm{Sigmoid}(w_{FH} H_{t-1} + W_{FX} X_t + b_F)$$
$$G_t = \mathrm{Tanh}(w_{GH} H_{t-1} + W_{GX} X_t + b_G)$$
$$C_t = (F_t C_{t-1} + I_t G_t)$$
$$H_t = O_t(\mathrm{Tanh}(C_t))(1)$$

Fig.2. highlights the functioning mechanism of LSTM cell. The association between the gates and memory cell, LSTM has a powerful capability to predict load demand with long-term dependences.

A. Optimal Architecture

The architecture of the LSTMnetwork models is given in Fig 1. This architecture consists of input and output nodes with the hidden node in the LSTM layer. Each hiddennode consists of their (LSTM) respective gates, which are discussed in section II. The input data(X) is smoothened by standardization to avoid unnecessary supremacy of certain variables using equation (2).

$$N_R(X)_i = \frac{X_i - \overline{X}}{S} (2)$$

X_i=Input data set (X_1, X_2,…X_i)
where i= number of samples
$\overline{X}$ = mean of the samples
S = standard deviation of the samples
The past historical load data is given as inputs to obtain forecasted day ahead power shown in Fig.1.The number of layers and its individual hidden nodes is designated based on statistical investigation and calculating Mean Absolute Percentage Error (MAPE) as given in equation (3).

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\text{actual load} - \text{forecast load}}{\text{actual load}}\right| (3)$$

Initially, the number of layers and hidden units are chosen on the basis of trial and error so that networks (LSTM) converge.Once the LSTM layers andhidden nodes are fixed, the number of layers(a) are changed from 1 to k and a is fixed for which MAPE is smallest. Then the number of hidden nodes(b) is varied from 1 to m for a=1 and b is fixed for which MAPE is smallest. The analogous procedure is carried out for k-1 layers. Hence the architecture ($A_{ibest}(B_{jbest})$) is selected based on the generalized equation given in (4).

$$A_{ibest}\left(N_{jbest}\right) = \mathrm{minMAPE}\{A_i(N_j)\}(4)$$
$$\forall\ a = 1,2 \ldots\ldots\ldots k \text{ and } b = 1,2,\ldots\ldots m$$

Once the number of layers and hidden nodes are fixed by using the equation (4), train the neural network for the selected input data and forecast the day ahead 15 minutes subinterval load data. The required steps to forecast load data using LSTM based neural network are briefed in the flowchart(Fig.3).

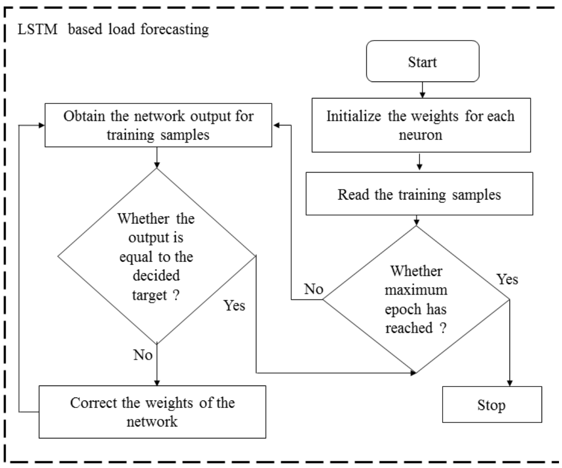


Figure.3. Flowchart of LSTM based forecast

III. Results and Analysis

The methodology proposed is implemented on GEFCom (Global Energy Forecasting Competition) data [23].The required data for training and validation is taken from 01 February 2007 to 15 March 2007 revealed in Fig.4. Here, 70% of the load samples are used for training, and 30% of the load samples are used for validation.In day ahead forecasting, n^{th} day load is mapped to $(n+1)^{th}$ day load samples[24].The finest architecture of LSTM is required foroperative load forecast in smart grid networks.The finest architecture for LSTM is therefore determined using statistical analysis as mentioned in Section II.A. The essential data for

January - February 2020
ISSN: 0193 - 4120 Page No. 6071 - 6076

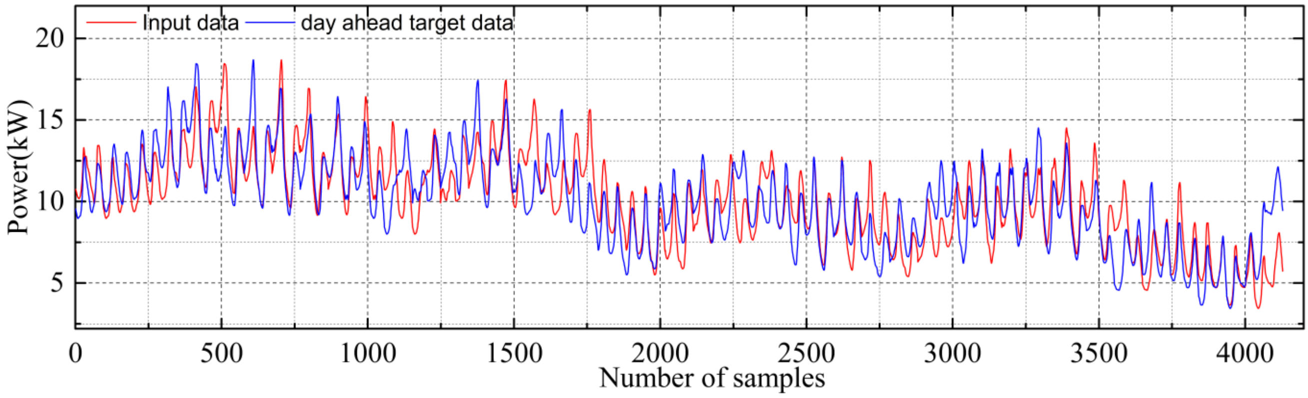


Figure.4.LSTMnetwork training dataset

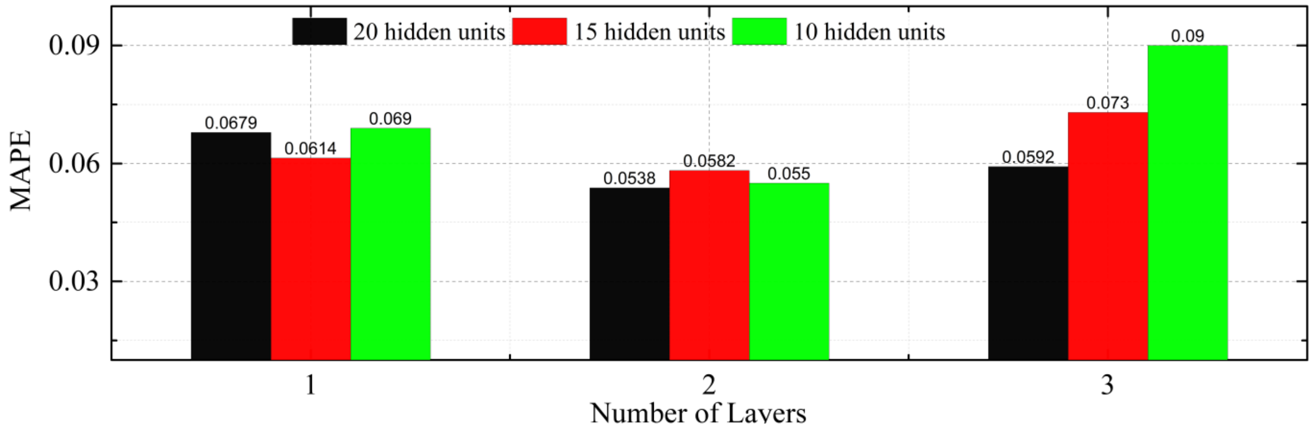


Figure.4. Optimal Layer Selection

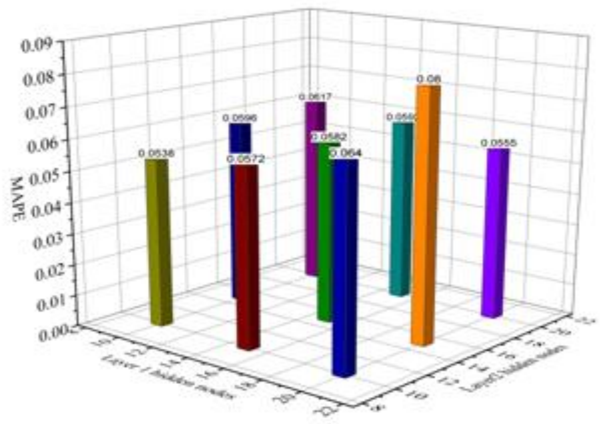
testing LSTM model is taken from 23/03/2007 to 29/03/2007(Assessment week). The parameter setting for the analysis of layer selection, hidden node selection and input data selection for proposed architecture is set as state activation function is sigmoid, gate activation function as tanh, weight learn rate factor=1, bias learn rate factor=1, weight multiplier factor=1 and bias multiplier factor=0. The number of epochs is kept the same for the entire analysis as 250.The statistical layer selection analysis is done with three layers and a combination of 10,15 and 20 hidden units is given as the following cases.

Case-I (1-Li (10)-1LSTM Architecture)

This network architecture consists of one input node, one output node,10 hidden units and the number of possible combinations for case I is given as follows.

1-1(10)-1, 1-2(10)-1, 1-3(10)-1.

Case-II (1-Li (15)-1 LSTM Architecture)



This network architecture consists of one input node, one output node,15 hidden units and the number of possible combinations for case II is given as follows.

1-1(15)-1, 1-2(15)-1, 1-3(15)-1.

Case-III (1-Li (20)-1 LSTMArchitecture)

This network architecture consists of one input node, one output node, 20 hidden units and the number of possible combinations for case III is given as follows.

1-1(20)-1, 1-2(20)-1, 1-3(20)-1

From the above-mentioned cases, the best suitable architecture is determined by using the performance metrics given in equation (4). In the first case number of hidden units are fixed as10 and the number oflayers varies from 1 to 3. Incase I layer selection,1-2(10)-1 structure is the best LSTM architecture with least MAPE (0.0538).In the second case number of hidden units are fixed as 15 and number of layers vary from 1 to 3. In case-II layer selection,1-2(15)-1 layer structure is the best LSTM architecture with least MAPE (0.0582). In the third case number of hidden units are fixed as 20 and the number of layersvaries from 1 to 3. In case-III layer selection 1-2(20)-1 layer structure is the best LSTM architecture with least MAPE (0.0555). Therefore, from the above mentioned cases it is observed that the two-layer structure is the best layer architecture for theproposed system and the same is observed from Fig.4.Hence the number of layers for the proposed LSTM architecture is fixed as 2,using two-layer architecture the number of hidden unit combinations is given as 9(referring equation(4)).They are 1-L1(10)-L2(10)-1, 1-L1(15)-L2(15)-1, 1-L1(20)-L2(20)-1, 1-L1(10)-L2(15)-1,1-L1(15)-L2(10)-1, 1-L1(10)-L2(20)-1,1-L1(20)-L2(10)-1, 1-L1(20)-L2(15)-1, 1-L1(15)-L2(20)-1.Forexample, 1-L1(20)-L2(15)-1 is one input node, one output node, layer1 with 20 hidden units and layer2 with 15 hidden units.The performance metrics Mean Absolute Percentage Error (MAPE) for the above-mentioned hidden unit combinations with two-layer architecture is shown in Fig.5 and From Fig.5 it is observed that 1-L1(10)-L2(10)-1 is the best LSTM layer and hidden unit combination with least MAPE (0.0538). Hence LSTM with two layers (layer one with 10 hidden units and layer two with 10 hidden units) is selected for forecasting the load data in advance in the smart grid environment. The forecasted load with optimal architecture (1-1(10)-2(10)-1) is shown in Fig.6 and the resulting training and validation convergence plots are shown in Fig.7. Hence this research paper suggested that LSTM based recurrent networks gives minimum forecasting error in load data forecasting in the present smart distribution system.

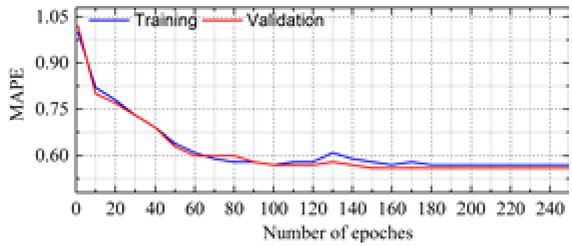


Figure.7. Training and validation Convergence plot

IV.CONCLUSION

This paper has explored the benefits of optimal selection of layers and hidden nodes for the accurate load data forecasting in the smart grid environment. Statistical analysis of LSTM architecture gives the two-layer model is the best layer structure and other than this MAPE is increasing.The accuracy also enhanced by layer 1 with 10 hidden units and that of layer 2 with 10 hidden units. Finally, it has been observed that the optimal selection of LSTM structure minimize the percentage of error and it is a good solution to forecast the load data in advanc in the smart grid environment.

V. REFERENCES

[1] AdityaChunekar, SapekshyaVarshney,Shantanu Dixit “Residential Electricity Consumption in India” Prayas Energy Group dec-2016.
[2] H.O.R.Howlader,H.Matayoshi,A.S.Noorzad,C.C.Muarapaz,andT.Senjyu, “Smart house-based optimal operation of thermal unit commitment for smart grid considering transmission constraint, ”Int. J.Sustain. Energy, Vol.37, No.5 pp.438-454, Feb.2017.
[3] International Energy Agency, “World Energy Outlook”, 2016, [Online]. Available: www.ieo.org.
[4] M.Iqbal,M.Azam,M.Naeem,A.S.Khwajaand A.Anpalagan,” Optimization Classification ,Algorithms and tools for Renewable Energy: A Review”,Vol.39,pp.640-654,July.2014.
[5] YanqingKuang, Mengqi Hu, Rui Dai, Dong Yang, “A Collaborative Decision Model for Electric Vehicle to Building Integration”, Energy Procedia,Vol. 105 , pp2077-2082,May 2017.
[6] AsifKhana, SaimMemon , Tariq Pervez Sattara, “Analysing integrated renewable energy and smart-grid systems to improve voltage quality and harmonic distortion losses at electric-vehicle charging stations”, IEEE. Translations and content mining, Vol 6,pp26404 – 26415, Apr 2018.

January - February 2020
ISSN: 0193 - 4120 Page No. 6071 - 6076

[7] MatamManjunatha, PrashantSinghb, AbhilashMandalb, Gaurav Singh Parih “Consumer Behaviour towards Electricity- a field study”, Energy Procedia,Vol.54,pp.541-548,2014 .
[8] Elisha R. Frederiks, Karen Stenner, Elizabeth V. Hobman, “Household energy use: Applying behavioural economics to understandconsumer decision-making and behavior”, Renewable and Sustainable Energy Reviews.Vol.41,pp. 1385–1394,Jan 2015.
[9] T.T. Anilkumar a, Sishaj P. Simon a, Narayana Prasad Padhy, “Residential electricity cost minimization model through open well-picoturbine pumped storage system”, AppliedEnergy ,Vol.195,pp.23-35, 2017.
[10] T. T. Kim and H. V. Poor, “Scheduling power consumption with price uncertainty,” IEEE Trans. Smart Grid, vol. 2, no. 3, pp. 519–527, Sep 2011.
[11] P. Samadi, H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, “Tackling the load uncertainty challenges for energy consumption scheduling in smart grid,” IEEE Trans. Smart Grid, vol. 4, no. 2, pp. 1007–1016, Jun 2013.
[12] Smitha Elsa Peter, I. Jacob Raglend&Sishaj P Simon, “An Architectural Frame Work of Ann Based Short Term Electricity Price Forecast Engine for Indian Energy Exchange Using Similar Day Approach” Vol. 2, no. 4, pp.111-122, Apr 2014.
[13] Tianhui Zhao, Jianxue Wang, Yao Zhang, “Day-Ahead Hierarchical Probabilistic Load Forecasting with Linear Quantile Regression and Empirical Copulas”, IEEE Access, Vol.7, pp. 80969 – 80979, June 2019.
[14] https://www.iexindia.com/
[15] R. Li et al., “Intelligent 5G: When Cellular Networks Meet Artificial Intelligence,” IEEE Wireless Commun., vol. 24, no. 5, pp. 175–83, Oct. 2017
[16] N. I. Sapankevych and R. Sankar, “Time Series Prediction Using Support Vector Machines: A Survey,” IEEE Comput. Intell. Mag., vol. 4, no. 2, pp. 24–38, May 2009.
[17] W.-C. Hong, “Application of Seasonal SVR with Chaotic Immune Algorithm in Traffic Flow Forecasting,” Neural Computing and Applications, vol. 21, no. 3, pp. 583–93, Apr 2012.
[18] I H. Witten et al., Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, Burlington, MA, USA, 4th edition, 2016, https://www.amazon.com/exec/obidos/ASIN/0128042915/departmofcompute, accessed on 11/30/2018.
[19] Yuxiu hua , Zhifeng zhao ,Rongpengli ,Xianfuchen, Zhimingliu ,Honggangzhang, “Deep Learning with Long Short-Term Memory for Time Series Prediction”, IEEECommunications Magazine,pp.1-16,mar 2019.
[20] S. Hochreiter and J. Schmidhuber, “Long Short-term Memory,” Neural Computation”, vol. 9, no. 8, pp. 1735–80, 1997.
[21] J. Wang et al., “Spatiotemporal Modeling and Prediction in Cellular Networks: A Big Data Enabled Deep Learning Approach,” Proc. 36th IEEE INFOCOM, Atlanta, GA, USA, pp. 1–9, May 2017.
[22] A. Alahi et al., “Social LSTM: Human Trajectory Prediction in Crowded Spaces,” Proc. 29th CVPR, Las Vegas, NV, USA, pp. 961–71, June 2016.
[23] http://www.drhongtao.com/gefcom/2012.
[24] Tianhui Zhao, Jianxue Wang, Yao Zhang, “Day-Ahead Hierarchical Probabilistic Load Forecasting With Linear Quantile Regression and Empirical Copulas”, IEEE Access, Vol.7, pp. 80969 – 80979, June 2019.