EDUCATIONAL BANDWIDTH TRAFFIC PREDICTION USING NON-LINEAR AUTOREGRESSIVE NEURAL NETWORKS

SHWAN DYLLON

School of Engineering, London South Bank University, 103 Borough Road, London, SE1 0AA, UK

TIMOTHY HONG

School of Engineering, London South Bank University, 103 Borough Road, London, SE1 0AA, UK

OUSMANE ABDOULAYE OUMAR

School of Engineering, London South Bank University, 103 Borough Road, London, SE1 0AA, UK

PERRY XIAO

School of Engineering, London South Bank University, 103 Borough Road, London, SE1 0AA, UK

Time series network traffic analysis and forecasting are important for fundamental to many decision-making processes, also to understand network performance, reliability and security, as well as to identify potential problems. This paper provides the latest work on London South Bank University (LSBU) network data traffic analysis by adapting nonlinear autoregressive exogenous model (NARX) based on Levenberg-Marquardt backpropagation algorithm. This technique can analyse and predict data usage in its current and future states, as well as visualise the hourly, daily, weekly, monthly, and quarterly activities with less computation requirement. Results and analysis proved the accuracy of the prediction techniques.

1. Introduction

Prediction of time series network traffic analysis is important for many decision-making processes, also to understand network performance, reliability and security, as well as to identify potential problems. Many types of research have been carried out such as statistical analysis, machine learning, neural networks, Fast Fourier transforms (FFT) wavelets, Continuous wavelet transform (CWT),

fuzzy set, chaos theory for that purpose to deal with the different characteristics of time series.

Artificial neural network (ANN) is an excellent tool for machine learning in terms of performance, classification and regression problems, which is very useful for time series and forecasting. Neural network techniques used comprehensively in many organisational applications such as financial, economic, energy systems, media (weather) to predict and forecasting their time series information

With network traffic analysis, network security staff would be able to identify any malicious or suspicious packets within the traffic, whilst network administrators could monitor the download/upload speeds, throughput etc., and therefore to have a sounder understanding of web operations.

Many techniques have been used in the network data traffic analysis. The neural network (NN), likewise known as the ANN, has been applied for prediction, as easily as to identify the presence of anomalies [1 -3]. Pattern recognition has been utilized for traffic data classification [4, 5], and chaos theory has been employed for the correlation and the prediction of time series data, and to identify the nonlinear dynamical behaviour of real-time traffic data [6, 7].

Applied back-propagation algorithm and feed-forward neural net models (FNNs) parameterized for nonlinear time series [8]. ANN forecasting has been compared with an autoregressive integrated moving average with exogenous variables (ARIMAX) model to predict the future [9], this mainly limited use in ANN for capturing the first-order non-stationarity in a time series data. Recurrent neural networks, recurrent neural networks (RNN) model has been evaluating the performance to identify and forecast in order to discover the presence of anomalies [10].

Different forecasting models were able to capture different aspects of the data available for prediction. Bates and Granger showed that a linear combination of forecasts would give a smaller error variance than any of the individual methods [11]. The basic idea of the model combination in forecasting is to use each models unique feature to capture different patterns in the data. With the ensemble model, the problem of model selection can be eased with little additional exertion. In summation, the ensemble method is potential to produce stable forecasts and less potential to make catastrophic predictions than any single network used in isolation [12]. Combining several models to improve the forecasting accuracy has been extensively analysed in the traditional forecasting literature. The question is how to combine these different forecasts into a single forecast, which is assumed a more accurate forecast. A simple approach to combining network outputs is to simply average them together. The NNE provides a prediction based on the norm of the MLPNN, RNN and RBFNN outputs [13]. A model based on the simple

average of neural networks, Gaussian process regression, and linear models for the NN5 time series forecasting competition [14].

Recently, several mathematical programming approaches have been proposed to determine the weights in combined forecasts. Several types of objective concerning accuracy to be used in forecast combinations [15] and considered to minimize mean absolute percentage error and the maximum absolute percentage error independently [16]. However, most previous focus on minimizing squared errors or absolute errors, they should be considered at the same time. An ensemble-forecasting model integrating ARIMA with ANN based on combined objectives, the combined objectives are composed of sum squares errors (SSE) and MAXAE, the MAXAE achieved by an ensemble model based on minimizing SSE, and then 0.618 criteria are used in deciding the weight of two objectives. This method is examined by using the data of Canadian Lynx data series. The empirical results show that the prediction using the proposed ensemble model is generally safer than other examples introduced in this field in terms of the same evaluation measurements.

2. Methodology

Data capture: In order to capture the data, we needed permission to access the LSBU network. This was easily acquired, as the researcher was working as a senior network engineer on the infrastructure team.

Once access was acquired, it was time to capture traffic data on the network. There are many tools available to capture network traffic. At the bottom end of the market, there are tools such as Wireshark and network analyser; in the mid-range, there is the SolarWinds network-monitoring tool; and at the upper end of the market, we have For this study, LSBU network traffic raw data were first captured using the Paessler Router Traffic Grapher (PRTG) network monitoring tool, which has been selected for this study.

The PRTG network-monitoring tool is equipped with SNMP, sFlow and port utilisation in order to provide the necessary data that is then used to analyse the network.

Many vendors use sFlow protocol to collect IP traffic information. Today, sFlow has become an industry standard for traffic monitoring. This protocol exports network flow data from routing devices.   The sFlow analysis provides the results of network traffic from the two main links for LSBU for traffic modelling, traffic engineering, capacity planning and forecasting, and anomaly and attack detection. This analysis allows us to find approximate average traffic on both core switches and their main link to the outside world. Other studies have been carried out

investigating SNMP statistics and the methodology for network bandwidth expansion. The methodology investigates trends and changes that have an effect on providing good IP backbone prediction and upgrades to the backbone infrastructure.

In this research, we studied the traffic of the LSBU network to see the efficiency of the entire infrastructure and to identify and manage the traffic both inbound and outbound at the gateway.

Data analysis tools: The process of inspecting, cleaning, transforming, and modelling data with the goal to discover useful information is data analysis. Network traffic analysis is primarily done to get in-depth insight into what type of traffic/network packets or data is flowing through a network. Typically, network traffic analysis is done through a network monitoring or network bandwidth monitoring software/application. In our case, monitoring download/upload speeds, throughput, content, etc. to understand network operations.

The PRTG tool provides some functionality in order to analyse network traffic

Model design and code: The program used the latest technique for analyzing and predicting the future trends on LSBU's network data, based on nonlinear autoregressive exogenous model (NARX) based on Levenberg-Marquardt backpropagation algorithm to solve an autoregression problem with external input with NARX neural network is proposed in [17]. NARX predict time series $y(t)$ given p past values of series y and another external series $x(t)$.

The following NARX equation shows the time series prediction behavior

$$y(t) = h(x(t-1), x(t-2), ..., x(t-k), y(t-1), y(t-2), ..., y(t-p)) + \in (t) \quad (1)$$

The program will normalise the attributes to reach the target accuracy by dividing the entries into 90-10 percent ratio for training and testing also training parameters such as training function, learning rate etc. Then generate the date for ahead number of days for which predicting the output by using the previously trained neural network with the ahead number of days for 24 hours.

The results of the analysis will provide the accuracy of the prediction techniques with less computation. Also, plot the predicted data usage in its current and future states, as well as visualise the hourly, daily, weekly, monthly, and quarterly activities.

Figure 1 describes each processing step in the program for the proposed admission control algorithms. Moreover, there are different scenarios with different steps in this project, which are related to the simulation. The simulation's approach includes the simulation methodology, which demonstrates the simulation design including different algorithms.
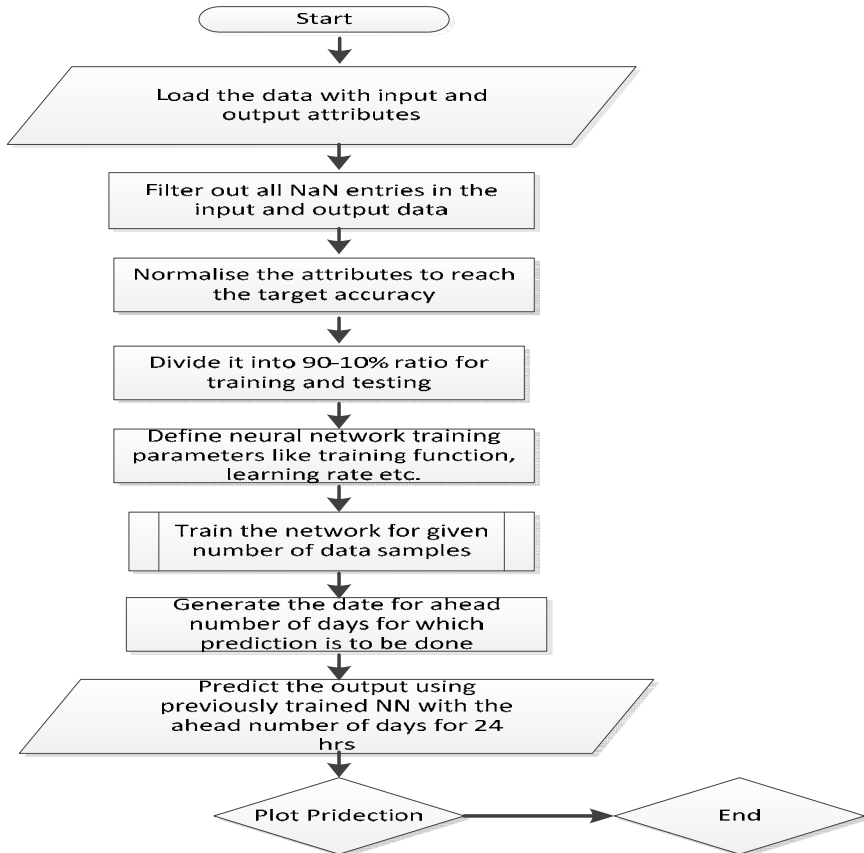
Figure 1: Flowchart of MATLAB code.

3. Analysis Results

In this study, we have recorded over two years of WWW (World Wide Web) traffic data from the 1st of January 2016 to the 1st of February 2018. Within this data, exactly two years of data, from the 1st of January 2016 to the 1st of January 2018, has been selected for training and testing the neural networks. We will predict the network traffic for the period between the 1st of January 2018 to the 1st of February 2018. The predicted traffic will be then compared with the real traffic during this period. Figure 1 shows the two year total WWW traffic data, from the 1st of January 2016 to the 1st of January 2018. The data was recorded at a rate of once per hour. Figure 2 shows the randomly selected 70% of the total WWW traffic data, used for training the neural networks. Figure 3 shows the randomly

selected 30% of the total WWW traffic data, used for testing the neural networks. This 30% includes a 15% validation ratio and another 15% testing ratio.
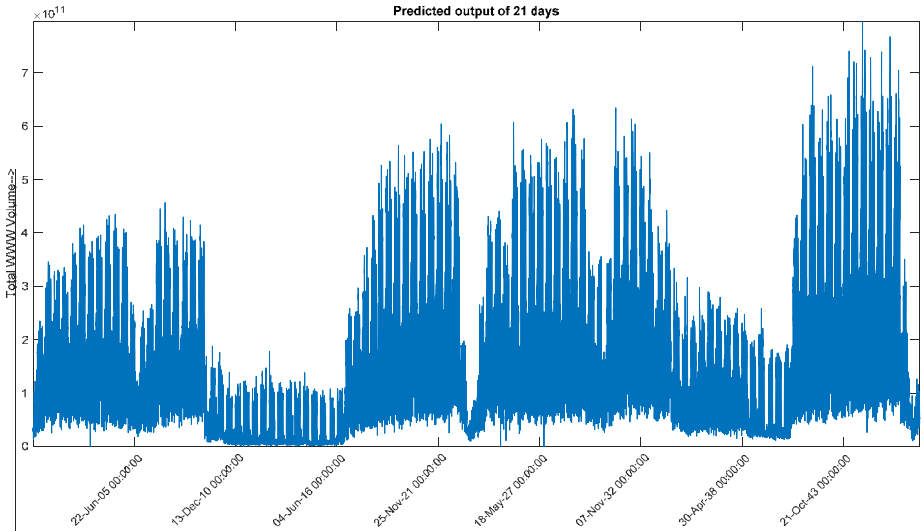

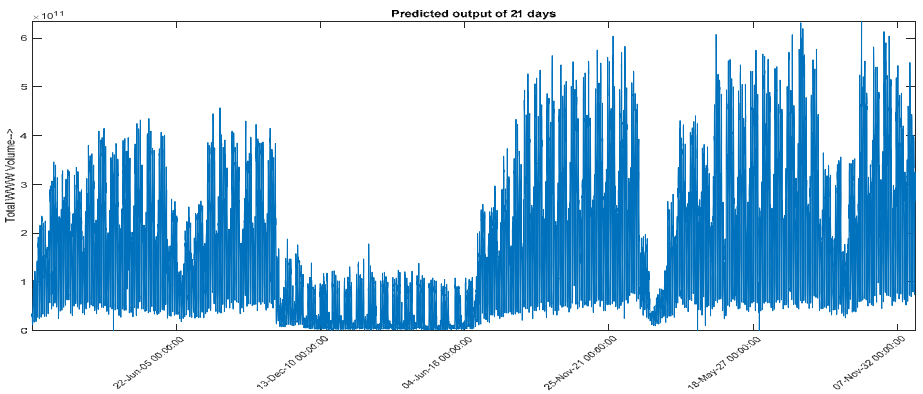Figure 2: Total WWW traffic from 1st Jan 2016 to 1st Jan 2018.

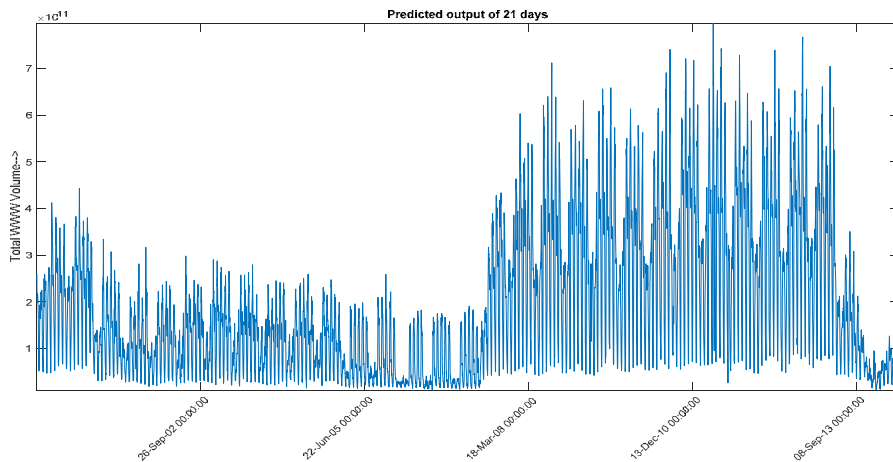Figure 3: 70% of the total WWW traffic used for training.

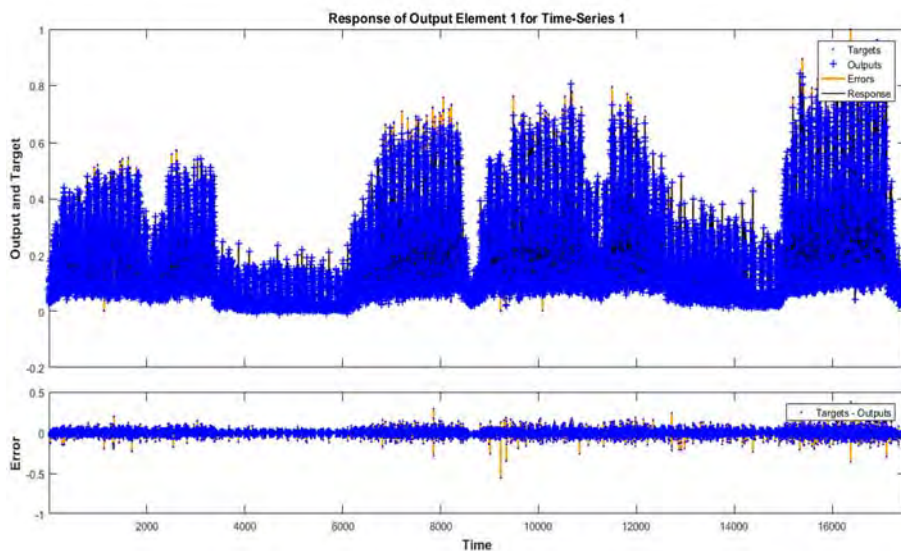Figure 4: 30% of the total WWW traffic used for testing.

Figure 5: Add the time series response plot, something similar to above.

Figure 5 shows the one week predicted WWW traffic from 1st Jan 2018 to 8th Jan 2018. The predicted traffic data can clearly show the day and night effect over the week. Figure 6 & 8 shows the regression plot of one week predicted traffic data,

258

which can prove the accuracy of the model by showing how close the output is to the actual values.
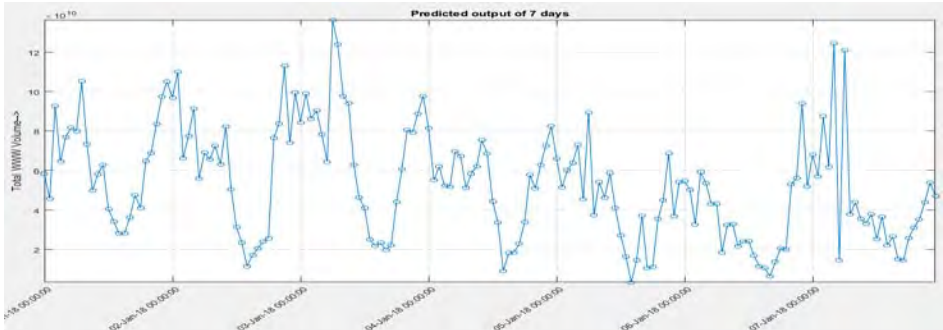

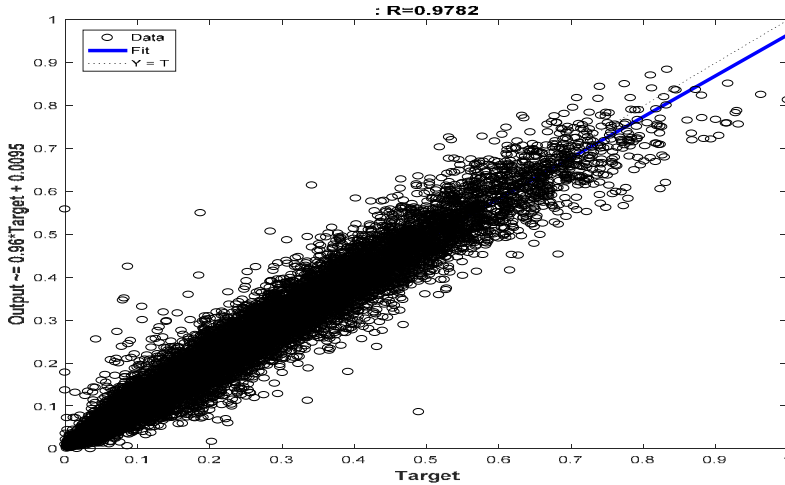Figure 6: One-week predicted WWW traffic from 1st Jan 2018 to 8th Jan 2018.

Figure 7: Regression plot of one week.

Figure 7 shows the four weeks predicted WWW traffic from 1st Jan 2018 to 29th Jan 2018. The predicted traffic data shows that in the first week, total traffic is relative low, this is likely due to the fact that university is still within the three weeks Christmas and New Year break. In the second week, traffic starts to pick, as the university re-open again. In the third and fourth week, the traffic reaches maximum, which not only show the day and night effect, but also the weekday and weekend effect.
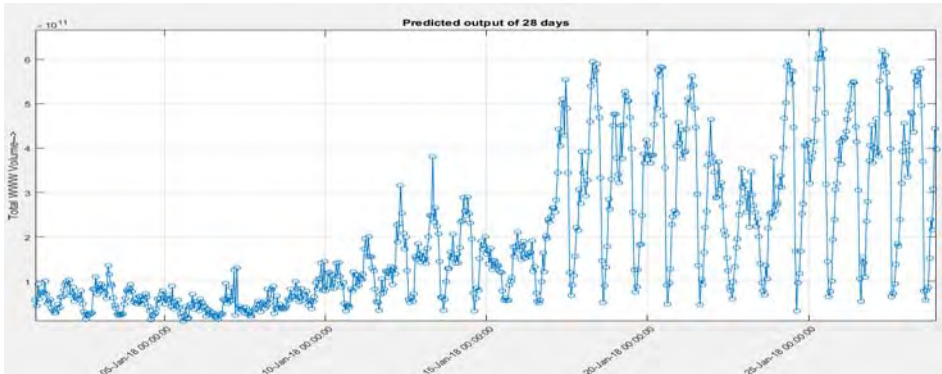
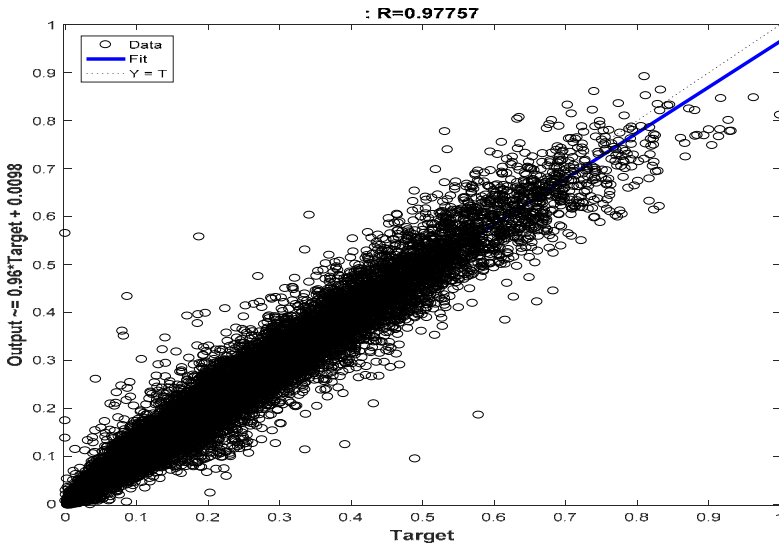Figure 8: Four weeks predicted WWW traffic from 1st Jan 2018 – 29th Jan 2018.

Figure 9: Regression plot of four weeks

4. Conclusion

We have developed a nonlinear autoregressive exogenous neural network model (NARX) based on Levenberg-Marquardt backpropagation algorithm for time series network traffic analysis. This study has implemented the methodology by developing a neural network model to predict the future trends of the LSBU bandwidth data traffic. It helped reasonably predict what traffic could be similar, given anomalies and outcomes. It has also proven that applying NARX neural

network can identify fundamental network usage and provide more accuracy on the prediction.

Furthermore, the prediction allowed us to focus on specific time in order to analyse the actual readings taken from the PRTG tool. This will allow us to predict user-driven network usage much more accurately in the future, as an efficient and effective way for decision-making process. The result shows that NARX neural network is a good method for predicting time series data.

References

1. Zhang, J., Song, C., Hu, Y., Yu, B., 2012. Improving robustness of robotic grasping by fusing multi-sensor. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp.126–131.
2. Pablo, B.J. et al., 2016. Artificial Neural Network and Monte Carlo Simulation in a hybrid method for time series forecasting with generation of L-scenarios. (1), pp.665–670.
3. Guo, T., Xu, Z., Yao, X., Chen, X., Aberer, K., and Funaya, K., 2016. Robust Online Time Series Prediction with Recurrent Neural Networks. 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp.816–825. Available at: http://ieeexplore.ieee.org/ document/7796970/.
4. Rocha, E., Salvador, P. and Nogueira, A., 2011. A real-time traffic classification approach. 2011 International Conference for Internet Technology and Secured Transactions (ICITST), (December), 620–626.
5. Han, L., Huang, L., Hu, Q., Han, X., Shi, J., 2012. Fast fourier transform based IP traffic classification system for SIPTO at H(e)NB. 7th International Conference on Communications and Networking in China, pp.430–435. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6417521.
6. Shang, P., Li, X., Kamae, S., 2005. Chaotic analysis of traffic time series, 25, 121–128. http://doi.org/10.1016/j.chaos.2004.09.104
7. Feng, H., Shu, Y. and Yang, O.W.W., 2009. Nonlinear analysis of wireless LAN traffic. Nonlinear Analysis: Real World Applications, 10(2), pp.1021–1028. Available at: http://www.sciencedirect.com/science/article/pii/S1468121807002386.
8. Zhang, J. et al., 2012. Improving robustness of robotic grasping by fusing multi sensor. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp.126–131.
9. Pablo, B.J. et al., 2016. Artificial Neural Network and Monte Carlo Simulation in a hybrid method for time series forecasting with generation of L-scenarios. (1), pp.665–670.
10. Guo, T. et al., 2016. Robust Online Time Series Prediction with Recurrent Neural Networks. 2016 IEEE International Conference on Data Science and Advanced Analytics(DSAA),pp.816–825.Availableat: http://ieeexplore.ieee.org/ document/7796970/.
11. J. M. Bates,and C. W. J. Granger, "The Combination of Forecasts," Operations Research Quarterly, vol. 20, no. 4, pp. 45l-468,1969.

12. G. P. Zhang and V. L. Berardi, "Time series forecasting with neural network nsembles: an application for exchange rate prediction," Journal of the Operational Research Society, vol. 52, pp. 652-664, 2001.
13. A. Chaouachi, R. M. Kamel, R. Ichikawa, H. Hayashi, and K. Nagasaka, "Neural Network Ensemble-based Solar Power Generation Short-Term Forecasting. International," Journal of Information and Mathematical Sciences, vol. 5, no. 4, pp. 332-337, 2009.
14. R. R. Andrawis, A. F. Atiya, H. El-Shishiny, "Forecast Combinations of Comptioal Intelligence and Linear Models for the NN5 Time Series Forecasting competition. International Journal of Forecasting," International Journal of Forecasting, vol. 27, no. 3, pp.672-688,2011.1883–1887, 2002.
15. G. R. Reeves, K. D. Lawrence, "Combining forecasts given different types of objectives," European Journal of Operational Research, vol. 51, pp.65-72,1991.
16. K. F. Lam,H. W. Mui, H. K. Yuen, "A note on minimizing absolute percentage error in combined forecasts," Computers & Operations Research, vol. 28, pp.1141-1147,2001.