Comparison between Artificial Neural Network and Non-Linear Auto Regressive Moving Average (NARMA) Models for Internet Traffic Prediction

Muhammad Naqiuddin Sahrani
Faculty of Electrical Engineering
Universiti Teknologi Mara,
40450 Shah Alam, Selangor, Malaysia
naqiuddinsahrani@yahoo.com

Abstract— This paper compare the performance of Artificial Neural Network with Non-Linear Auto Regressive Moving Average (NARMA) models by predicting internet traffic. The ANN models that used in this paper was Non-Linear Auto Regressive (NAR) model where this model was train to predict the internet traffic. The NARMA model result was obtained by combine the result of NAR and Moving Average (MA) models where the result of MA model obtain by predicting the residual of NAR model. The performance of both models was evaluated by using model fitting and model validation tests. NARMA models show better results in internet forecasting.

Index Terms—internet traffic, forecasting, Nonlinear Auto-Regressive Moving Average (NARMA) model, Nonlinear Auto-Regressive (NAR) model, prediction.

I. INTRODUCTION

Networking services have been become demanded services in many areas of industry such as online shopping, trading, business and many more. With the capability of networking services, peoples can communication each other from difference place without troubling their self to meet the customers.

Since the networking services has become demand services [1, 2], there are many problems that have risen. Network traffic have becomes one of problem that need to be handled immediately. When at peak time, network congestion [3, 4] might have chance to occur. It might lead to have longer delay, high jitter and massive packet losses [3]. To let customers to get access to the server smoothly, predicting internet traffic [5] is some of the ways that can help to avoid this problem that might be occur to the user. The company might take some precaution or countermeasure before the problems become worse [5].

Because of demands from customers to access the server smoothly, internet traffic prediction has become an active research field in networking engineering area [5, 6]. There are many methodologies that have been used in some studies [7-9]. Most of the studies used to increase the prediction efficiency and accuracy [5, 8] and to get dynamic bandwidth allocation [2, 10]. Traditional traffic models are one of methods that have been used before but these models have its own weakness. The lack of selection of methodologies when choose the appropriate model can cause the model fail to used. Besides that, these models have low-level efficiency when it's come in non-linear process of network traffic time series [5, 8, 11]. As solution, nonlinear models such as artificial neural networks (ANN) are used. Besides that, traffic classification also have been proposed[12] to detect any unusual patterns that can be happens. In modern network management, these methods might have important role that can make any network traffic run smoothly.

This paper presents two types of forecasting models: NAR and NARMA for predicting the internet network traffic. Both models results are analyzed in order to compare which model predict network traffic more efficiency. The future network traffic is predicted by using NAR model using the past data of network traffic. The NARMA model is created by combined the outputs of NAR model and MA model where for MA model use residuals from NAR model in order to improve the prediction outputs.

This paper is organized by starting a literature review about this research in Section II, the experimental overview is presented in Section III and in Section IV, the results of the experiments and discussion about the results are presented. Finally, conclusion about this research are presented in Section V.

II. LITERATURE REVIEW

A. Internet Traffic Prediction

There are a lots study that related to the internet forecasting has been done before this. There are several methods that have been used by previous researchers to predict internet traffic. Even there are many methods that have been used but the main purpose of their research is to be able to predict internet traffic more precisely.

The study in [13] compared in terms of Root Mean Square Error (RMSE) criterion between Autoregressive-Generalized Autoregressive Conditional Heteroscedasticity (AR-GARCH) error model and seasonal autoregressive integrated moving average (ARIMA) models. The author evaluates both models performance for forecasting internet traffic. It was concluded that the seasonal AR-GARCH models outperformed the SARIMA models in terms of forecasting accuracy with respect to the RMSE criterion.

The authors using fuzzy interference system in approach study in [14]. The study is developed based on Wang and Mendel algorithm. To evaluate the performance of the model, authors use the parameter of RMSE by using cross-validation to handle under fitting which mean test the parameter using extrapolation of fuzzy rules. They conclude that the performance of FIS prediction is satisfied by using two to three predictor variables. However, there still have remaining issues that need to solve in the future research.

There also have a study using Two-Dimensional Correlation and Single Exponential Smoothing (TDCNTP) method[5]. The authors predict the network traffic based on the periodicity and continuity of network traffic. They predict errors of network that might be happen and take countermeasures to overcome that problem before it happen. The authors evaluate the result by compare the result to the traditional network traffic prediction model and conclude that the prediction efficiency and accuracy has improving and almost same as actual network traffic.

B. ANN & its applications to network forecasting

ANN is defined as adaptive model inspired by learning process of the human brain [10, 15, 16]. However, most of problems that face at real world are exist in nonlinear condition. This has led to the use of nonlinear estimators such as Artificial Neural Networks (ANN) [17], ARMA, NARMAX and some other estimator.

ANN is powerful estimators that can theoretically learn any relationship between the input and output provided that the structure has sufficient hidden units. The Multi-Layer Perceptron (MLP) is one of the more established ANN. In MLP, the characteristic to classify difficult and nonlinear cases can be overcome.

The study in [7] introduced a methodology to predict internet traffic using Artificial Neural Networks in the long-term period of time by collecting information based on flows to build the time series. The authors also presented a proposed four traffic forecasting models based on artificial neural networks. The authors compared the proposed methods against the Holt-Winters method. Generally, they conclude that the proposed models can be accepted as a good method for Internet traffic prediction.

In [10], the authors use multi-step prediction to predict video time series using neural networks. There are two models that used in the study which are MLP and NARX models. The authors conclude that both models results are worse than single-step prediction. The advantage of these methods is both models can predict multi outputs in single step. However, the results of these models can match with single-step prediction results.

C. NAR and NARMA model

The NAR model is a model that can be used for forecasting where this model use the past output in order to predict future values[18, 19]. The NAR model can be defined as:

$$y(t) = N[y(t-1), y(t-2), .. y(t-n+1)] + \varepsilon(t) \quad (1)$$

where:

y(t) = system output
ε(t) = residuals
y(t-1) = past system output
N = coefficient

To improve the result of NAR model, a new model which can be name as NARMA model is created by including the residual into the model[19]. The model can be defined as[20]:

$$y(t) = N[y(t-1, y(t-2), .. y(t-n+1), \varepsilon(t-1), \varepsilon(t-2), .. \varepsilon(t-n+1)] + \varepsilon(t) \quad (2)$$

where :

ε (t-1) = past residual outputs.

There not much research that focuses on networking area using NARMA model. Mostly, NARMA model used to identification, controller and forecasting.

The study in [20] use to compare between NARMA-L2 Neuro controller with nonlinear PID controller. The authors has control the speed of DC motor by develop a conventional PID and NARMA-L2 controller and train the data using MATLAB. The result shows that NARMA-L2 performance better that PID controller. The advantages of this model are there are no overshoot and the performance to track speed more efficiently.

There also have a study that compare between NAR and NARMA model[19]. The authors have forecasting tourism by using both model and analyze the results. The result shown that NARMA model provide more efficiently in the terms of Mean Square Error (MSE) and other tests.

III. METHODOLOGY

In this section, the experimental setup of the proposed models is described. Sections III-A present the hardware description and Section III-B present the experimental overview. Next, the data description present in Section III-C and parameters that needs to adjust present in Section III-D. Finally, testing methods present in Section III-E

A. Hardware Description

During this research, all of tests were done on an ASUS laptop with Intel® core™ i5 Central Processing Unit (CPU) (running at 2.40GHz) with 6.00GB of Random Access Memory (RAM), Microsoft Windows 7 Home Premium (64-bit) was installed as the operating systems. All programs were implemented using MATLAB version 2012a (7.14.0.739).

B. Experimental Overview

These experiment can be represent in the block diagram which shown in Figure 1. In this research, NARMA model was obtained by add up the output of two models which are MLP 1 and MLP 2. MLP 1 represents the ANN model where in this research is NAR model was used and for MLP 2 represents the ANN model which is MA model. The input for the MLP 1 is

used from the dataset which can represents as y(t) and the output from MLP 1 is represent as $\hat{y}(t)$. The input for the MLP 2 is used from the residual r(t) of MLP 1 which can get from the result of subtraction of $\hat{y}(t)$ and y(t) and the output of MLP 2 can be represent as $\acute{r}(t)$. Finally, for the NARMA model part can be obtained by add up for both output and can be represent as $\hat{y}_{NARMA}$.

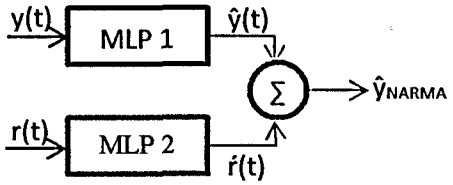

Figure 1: Block representation for both models

C. Data Description

For this research, the dataset that used was taken from open sources data from The Internet Traffic Archive website[21]. The dataset was sponsored by ACM SIGCOMM where BC-Oct89Ext is taken as dataset. The trace BC-Oct89Ext began at 23:46 on October 3, 1989, where the first 1 million external arrivals were captured. The tracing was done at the Bellcore Morristown Research and Engineering facility, where all traffic between Bellcore and the Internet was done.

The original data need to be filter before this research can be done because the original have too many data point. It will make the process to perform the experiments become complicated. After being filtered, the data that taken was only 394 data points and divided into 3 sets where 276 data point for training set, 59 data point for validation set and 59 for testing set. The dataset for the input was plotted into graph as shown in Figure 2

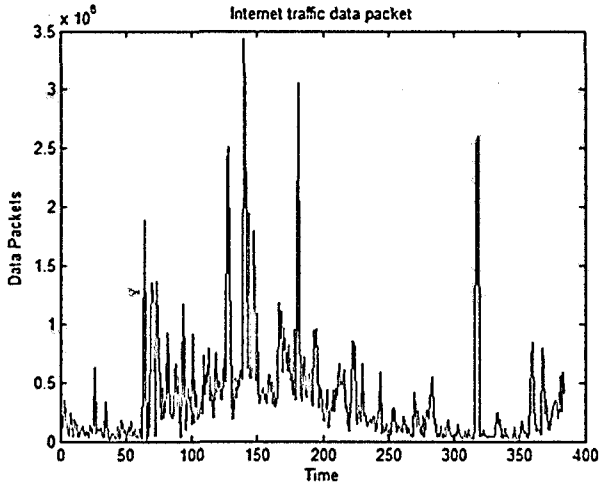

Figure 2: Internet network traffic data packet

D. Parameters Description

This section describes the parameter that used to perform the experiment for both models. The parameters that used are the number of lag and training number for the models. The suitable lag and training number need to pass the tests. The details of the test are described on Section III-E below. The parameter was choose based on try and error method. The parameters values are choose from 5 to 25. Lag is the number of past experiment values that used to predict the network traffic.

E. Testing Methods

There are two types of testing method that will used during testing the parameters which are fitting test and model validation test. For the fitting test, the tests used are Mean-Square Error (MSE) and One Step A-Head (OSA) prediction, and for model validation test, the tests used are Correlation and Histogram test. The detailed about the tests will be described in detail below.

a. Mean-Square Error

The most common method that used to measure the performance of the system is using MSE.
MSE can be defining as equation below [11, 22, 23]:

$$MSE = \frac{1}{n}\sum_{n=1}^{n}\left(\hat{y}(n) - y(n)\right)^2 \quad (3)$$

b. One Step Ahead Prediction

OSE is one of test that can be used to test the accuracy of the model's output [11, 22]. OSA[24] can be written as

$$\hat{y}(t) = f(u(t), u(t-1)\ldots u(t-n_u), y(t-1), y(t-2)\ldots y(t-n_y)) \quad (4)$$

Where f(t) is non-linear function, u is the input system and y is the output system.

c. Residual test

Residual test is test that plots all of the residuals in the dataset. Residuals are the different between the original outputs with the predicted data[19]. Residuals also can be defined as:

$$\varepsilon(t) = y(t) - \hat{y}(t) \quad (6)$$

where
ε (t) = residuals
y (t) = actual output data
ŷ (t) = predicted output data

d. Correlation Tests

Correlation test is other tests which more convince method to test the performance of model. This test is used to check the prediction error ε(t) of the pattern by measuring the similarity between two datasets. The system can be considered acceptable if the result is within the acceptable range which is 95% confidence bands [22, 25].

e. *Histogram Analysis*

Histogram analysis is an analysis that represent in the form of graphical where it is a summary of distribution of the dataset. In prediction, this method used to view the distribution of the residual. The histogram will represent the white noise as a symmetric bell-shaped distribution.

IV. Results & Discussion

This section discusses the results for both proposed models. There are several test are used to determine the efficiency and validity of the models which were discuss in Section III-E.

The comparison of the model are presented in Section IV-A. The results for NAR model are presented in Section IV-B and for the results for NARMA model are presented in Section IV-C.

A. Comparison of Models

Table 1 present the comparison of Mean Squared Error (MSE) results. For the MSE value, the smaller values indicates that the model show better result to used. Both training and testing data show improvement even just slight. The results of NARMA model show that improve and more efficiently than NAR model.

	Training	Testing
NAR	0.0161	0.1321
NARMA	0.0131	0.1288
Percentages of Improvement (%)	18.5085	2.5337

Table 1: Comparison of MSE between NAR and NARMA models

Testing Methods	NAR		NARMA	
	Training	Testing	Training	Testing
Histogram Plot	Pass	Pass	Pass	Pass
Residual Plot	Pass	Pass	Pass	Pass
OSA prediction	Pass	Pass	Pass	Pass
Autocorrelation Test	0	1	0	1
	Most of coefficients approaching zero point	Most of coefficients approaching confident boundary	Most of coefficients approaching zero point	Most of coefficients approaching zero point
Cross-Correlation Test	3	2	2	2
	Most of coefficients approaching zero point	Most of coefficients approaching confident boundary	Most of coefficients show better results	Most of coefficients show better results

Table 2 Comparison of testing method between NAR and NARMA models

Table 2 show the comparison of the tests that was done to the both models. Generally, both models were passed the tests that have done to the models. In order to compare the performance of both models, correlation test results were the suitable tests to evaluate the performance. The results of the test for NAR model is presented in Section IV-B and for NARMA model is presented in Section IV-C. The results will be presented in the detail.

B. NAR Model Results

The results of the NAR model are shown in Figure 3 and Figure 4. The autocorrelation test (Figure 3) shows that this model passes the test. There only one coefficient that pass the confident boundary for the testing dataset.

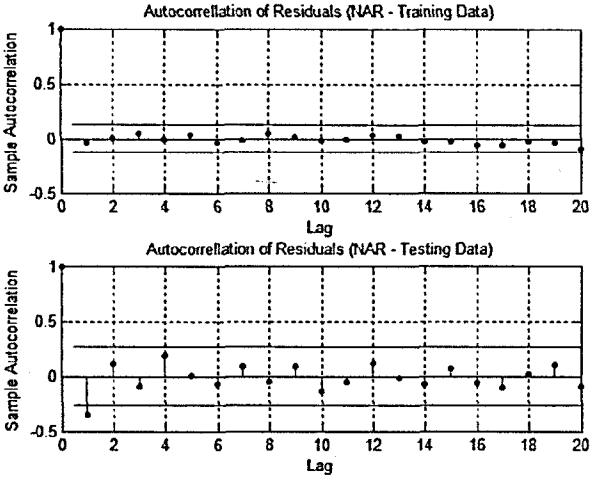

Figure 3: Autocorrelation test for NAR model

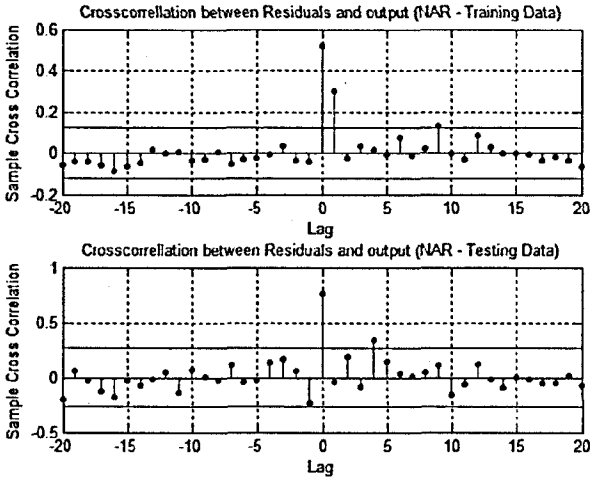

Figure 4: Cross-correlation test for NAR model

Figure 5 show the results of cross-correlation for both training dataset and testing dataset. Testing dataset show better result that training dataset. There are three coefficients that passed the confident boundary for training dataset and there are two coefficients that passed the confident boundary.

C. NARMA Models Results

The outputs from NAR and MA model were combining to in order to create NARMA model. The results for this model were presented from Figure 5 and Figure 6.

For autocorrelation test (Figure 5) show that this model were pass the test because all of coefficients were inside the confident bound. However, there is one coefficient that passed the confident boundary.

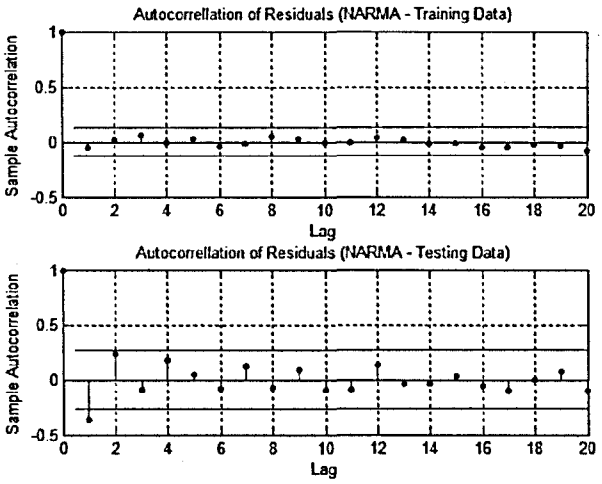

Figure 5: Autocorrelation test for NARMA model

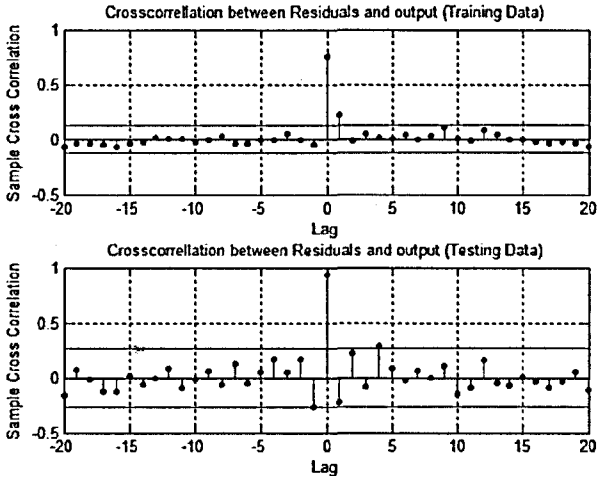

Figure 6: Cross-correlation test for NARMA model

For cross-correlation test (Figure 6) shown that several coefficients where outside the bound but still can be assume *pass the test. There are two coefficients that passed the confident boundary for both training dataset and testing dataset.* However, there are slightly improvement for the coefficients where it more approaching to the zero line.

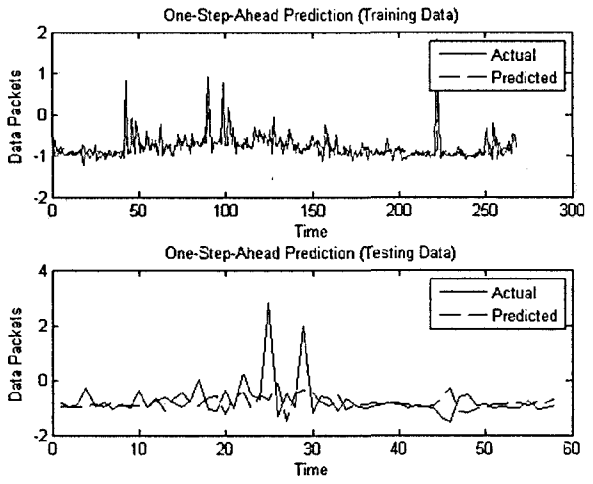

Figure 7: OSA test for NARMA model

The OSA test (Figure 7) shows the comparison between actual data with predicted result. Generally, for the training data, it shown that the prediction results generally follow the original dataset but for testing data, the prediction not clearly follow the original data. The histogram plot (Figure 8) for NARMA model almost follows the Gaussian curve for training data but testing data shown clearly Gaussian curve.

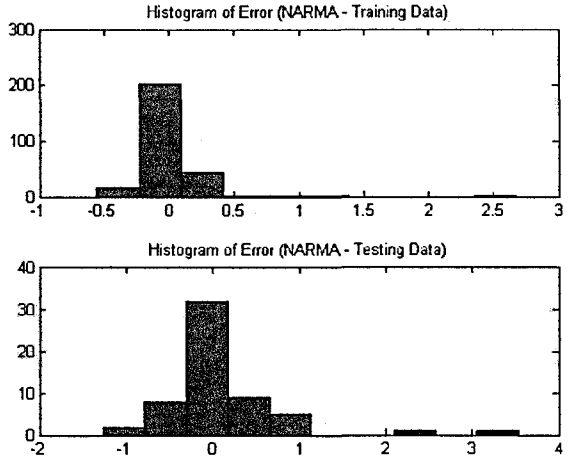

Figure 8: Histogram plot for NARMA model

V. CONCLUSION

This paper has compare between Artificially Neural Network (ANN) and Non-Linear Auto regressive Moving Average (NARMA) Models for internet traffic prediction. Both *models has tested by using fitting test and validation test. For fitting test, one step a-head prediction and mean square error*

has used and for validation test, correlation test and histogram test has been used to evaluate the performance and both model has compare. Both models show that pass the tests and can be used to predict the network traffic. From the MSE result and correlation tests, it can be concluded that NARMA model shown better performance compare to the ANN model.

REFERENCE

[1] B. Tao, G. Shanqing, Z. Zonghua, R. Ando, and Y. Kadobayashi, "Practical network traffic analysis in P2P environment," in *Wireless Communications and Mobile Computing Conference (IWCMC), 2011 7th International*, 2011, pp. 1801-1807.

[2] M. J. O.Gavade and M. P.K.Kharat, "NEURAL NETWORK BASED APPROACH FOR MPEG VIDEO TRAFFIC PREDICTION " p. 6, 2011.

[3] N. Zhenyu and G. Qing, "An improved AQM scheme with adaptive reference queue threshold," in *Communications and Networking in China (CHINACOM), 2011 6th International ICST Conference on*, 2011, pp. 589-593.

[4] G. Terdik and T. Gyires, "Internet Traffic Modeling with Levy Flights," in *Networking, 2008. ICN 2008. Seventh International Conference on*, 2008, pp. 468-473.

[5] K. Jian, Z. Dongwei, W. Xinyu, W. Yanwen, and M. Qinghua, "A network traffic prediction method using two-dimensional correlation and Single exponential smoothing," in *Communication Technology (ICCT), 2013 15th IEEE International Conference on*, 2013, pp. 403-406.

[6] S. Koyano, S. Ata, I. Oka, and K. Inoue, "A High-Grained Traffic Prediction for Microseconds Power Control in Energy-Aware Routers," in *Utility and Cloud Computing (UCC), 2012 IEEE Fifth International Conference on*, 2012, pp. 347-352.

[7] M. L. F. Miguel, M. C. Penna, J. C. Nievola, and M. E. Pellenz, "New models for long-term Internet traffic forecasting using artificial neural networks and flow based information," in *Network Operations and Management Symposium (NOMS), 2012 IEEE*, 2012, pp. 1082-1088.

[8] Z. Run, C. Yinping, and F. Xing-an, "A network traffic prediction model based on recurrent wavelet neural network," in *Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on*, 2012, pp. 1630-1633.

[9] A. Morici, D. Gajic, and L. D. Gregorio, "Access Network Traffic Models for Energy Efficient Architectures," 2013.

[10] F. Pilka and M. Oravec, "Multi-step ahead prediction using neural networks," in *ELMAR, 2011 Proceedings*, 2011, pp. 269-272.

[11] V. K. Dabhi and S. Chaudhary, "Time Series Modeling and Prediction Using Postfix Genetic Programming," in *Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on*, 2014, pp. 307-314.

[12] Z. Jun, C. Chao, X. Yang, Z. Wanlei, and X. Yong, "Internet Traffic Classification by Aggregating Correlated Naive Bayes Predictions," *Information Forensics and Security, IEEE Transactions on*, vol. 8, pp. 5-15, 2013.

[13] K. Sahm, "Forecasting internet traffic by using seasonal GARCH models," *Communications and Networks, Journal of*, vol. 13, pp. 621-624, 2011.

[14] C. K. Maurya and S. Minz, "Fuzzy inference system for Internet traffic load forecasting," in *Computing and Communication Systems (NCCCS), 2012 National Conference on*, 2012, pp. 1-4.

[15] T. Liu, "Optimizing mining association rules based on Artificial Neural Network," in *World Automation Congress (WAC), 2012*, 2012, pp. 1-4.

[16] A. RajaRajan, "Brain disorder detection using artificial neural network," in *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*, 2011, pp. 268-272.

[17] S. Su, W. Zhang, and S. Zhao, "Online fault prediction for nonlinear system based on sliding ARMA combined with online LS-SVR," in *Control Conference (CCC), 2014 33rd Chinese*, 2014, pp. 3287-3291.

[18] I. M. Yassin, A. Zabidi, M. K. M. Salleh, and N. E. A. Khalid, "Malaysian tourism interest forecasting using Nonlinear Auto-Regressive (NAR) model," in *System Engineering and Technology (ICSET), 2013 IEEE 3rd International Conference on*, 2013, pp. 32-36.

[19] I. M. Yassin, A. Zabidi, Syahril Nizam Kadir, and Nooritawati Md Tahir, "Malaysian Tourism Interest Forecasting using Nonlinear Auto-Regressive Moving Average (NARMA) Model," in *2014 IEEE Symposium on Wireless Technology and Applications (ISWTA)*, 2014, pp. 193 - 198.

[20] S. K. Valluru, M. Singh, and N. Kumar, "Implementation of NARMA-L2 Neuro controller for speed regulation of series connected DC motor," in *Power Electronics (IICPE), 2012 IEEE 5th India International Conference on*, 2012, pp. 1-7.

[21] *The Internet Traffic Archive*. Available: http://ita.ee.lbl.gov/index.html

[22] N. A. Jalil and I. Z. M. Darus, "NARMA-L2 Vibration Controller for Flexible Structure with Non-Collocated Sensor-Actuator " presented at the Fifth International Conference on Computational Intelligence, Modelling and Simulation, 2013.

[23] Z. Muhammad, Z. M. Yusoff, M. H. F. Rahiman, and M. N. Taib, "Modeling of steam distillation pot with ARX model," in *Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on*, 2012, pp. 194-198.

[24] D. Guanqun, K. Fataliyev, and W. Lipo, "One-step and multi-step ahead stock prediction using backpropagation neural networks," in *Information, Communications and Signal Processing (ICICS) 2013 9th International Conference on*, 2013, pp. 1-5.

[25] S. Nomm, K. Vassiljeva, J. Belikov, and E. Petlenkov, "Structure identification of NN-ANARX model by genetic algorithm with combined cross-correlation-test based evaluation function," in *Control and Automation (ICCA), 2011 9th IEEE International Conference on*, 2011, pp. 65-70.