Performance of Modeling Time Series Using Nonlinear Autoregressive with eXogenous input (NARX) in the Network Traffic Forecasting

Haviluddin
Faculty of Mathematics and Natural Science
Dept. of Computer Science, Universitas Mulawarman, Indonesia
haviluddin@unmul.ac.id

Rayner Alfred
Faculty of Computing and Informatics,
Dept. of Computer Science, Universiti Malaysia Sabah, Malaysia
ralfred@ums.edu.my

***Abstract* — A time-series data analysis and prediction tool for learning the network traffic usage data is very important in order to ensure an acceptable and a good quality of network services can be provided to the organization (e.g., university). This paper presents the modeling using a nonlinear autoregressive with *eXogenous* input (NARX) algorithm for predicting network traffic datasets. The best performance of NARX model, based on the architecture 189:31:94 or 60%:10%:30%, with delay value of 5, is able to produce a pretty good with Mean Squared Error of 0.006717 with the value of correlation coefficient, *r*, of 0.90764 respectively. In short, the NARX technique has been proven to learn network traffic effectively with an acceptable predictive accuracy result obtained.**

Keywords—NARX; network traffic; MSE; correlation coefficient

I. INTRODUCTION

Time series analysis tools that are used for modelling and forecasting time series datasets are widely used in various fields including economic field (i.e. business, finance, foreign exchange, and stock problems), investment, engineering, energy, internet, and network traffic. Indeed, an accurate prediction ability is highly required in order to assist the process of decision making. In the literature review, numerous strategies have been established in the general framework of time series prediction. These techniques can be grouped into two main categories: statistical and machine learning (ML) methods. There are several types of methods that are derived from the statistics such as autoregressive (AR), moving average (MA), autoregressive moving-average (ARMA), autoregressive integrated moving-average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), and seasonal autoregressive integrated moving-average (SARIMA). Statistical methods are reliable enough to be used in forecasting, if the amount of data is not too much with linear data types. Meanwhile, the results of forecasting have been less accurate when using a lot of data, due to the fact that the mathematical model generated is quite complicated, and difficult to be implemented by using a nonlinear data type [1-3].

On the other hand, machine learning (ML) has been also besides these statistical models. For instance, the Artificial Neural Networks (ANN) is one of the ML methods, in which it is widely used for analyzing and forecasting time series data in the past four decades. Additionally, many researchers have been using ANN widely as a time series analysis method to solve problems due to its efficiency in solving linear and nonlinear problems [4-6]. Among the ANN extension methods include the multilayer perceptron's with back propagation (BP), recurrent neural networks (RNN), and a radial basis function (RBF) neural network, that can provide efficient and accurate forecasting, also being able to analyze especially by using nonlinear data as a representation of the real world [1, 2, 7-12]. The motivation of this paper is to present a topology and training scheme of a neural network that is able to forecast the network traffic with some degree of accuracy using a one-step ahead prediction. It is hoped that this paper can provide insights to support network engineer management in providing an efficient bandwidth traffic control management for the campus communities. This paper will study the Nonlinear Auto Regressive with *eXogenous* input neural network (NARX) model, in order to address the issue of time series data that has non-linear characteristics. Section II describes the methodology used in this work. Section III outlines the experimental setup. Section IV presents the analysis and discussion results, and Section V concludes this paper.

II. METHODOLOGY

In this section, related works on the general network traffic prediction models will be presented, including the time series analysis performed by using the NARX model.

A. Time Series

A time series dataset is a dataset that consists of observations ordered in time. In principle, time series model is used to predict the values of data $(U_{n+1}, U_{n+2}, \dots, U_{n+n})$ based on the data $(U_{n+1}, U_{n+2}, \dots, U_{n+n})$ [13]. In this study, the time series dataset is obtained from the ICT server of Universitas Mulawarman. Each network traffic data was captured by using the CACTI software from 20 – 26 June 2013 (314 samples series data). The dataset and plot dataset are shown in Table I and Fig. 1.

TABLE I. REAL NETWORK TRAFFIC DATA 20-26 JUNE 2013

Date	Time		Inb+ Out	Date	Time		Inb+ Out
6/20	11:00	1	21980000	6/24	12:00	171	4528000
	11:30	2	25700000		12:30	172	3603000
	…		…		…		…
	…		…		…		…
	11:30	26	7645000		11:30	218	5969000
6/21	12:00	27	6293000	6/25	12:00	219	7933000
	12:30	28	5185000		12:30	220	8637000
	…		…		…		…
	…		…		…		…
	11:30	74	11661000		11:30	266	13230000
6/22	12:00	75	8390000	6/26	12:00	267	11540000
	12:30	76	7307000		12:30	268	13350000
	…		…		…		…
	…		…		…		…
	11:30	122	14530000		11:30	314	9650000
6/23	12:00	123	10517000				
	12:30	124	6715000				
	…		…				
	…		…				
	11:30	170	5236000				

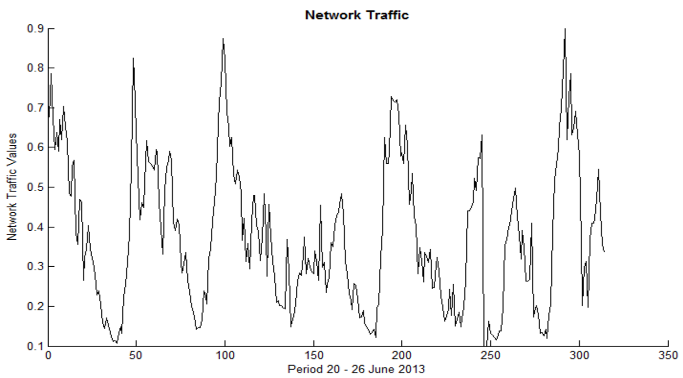


Fig. 1. Plot network traffic data 20-26 June 2013

In this study, the daily network traffic data was analyzed by using a MATLAB R2013b. The daily network traffic data needs to be normalized into the interval [0, 1] by using the statistical data normalization, which is usually expressed as:

$$\bar{X} = \frac{0.8(X - X_{min})}{(X_{max} - X_{min})} + 0.1 \tag{1}$$

where: $\bar{X}$ is the actual value of sample; X_{max} takes a large value, and X_{min} takes a samples of data is less than the minimum value. Next, inverse transform process is performed in order to get the actual value and the final dataset obtained after the normalization process is illustrated in Table II.

TABLE II. NETWORK TRAFFIC DATA AFTER NORMALIZATION

	Date	Time		Inb+ Out		Date	Time		Inb+ Out
Training	6/20	11:00	1	0.62411	Training	6/24	12:00	171	0.20797
		11:30	2	0.71282			12:30	172	0.18591
		…		…			…		…
		…		…			…		…
		11:30	26	0.28230			11:30	218	0.24233
	6/21	12:00	27	0.25006		6/25	12:00	219	0.28916
		12:30	28	0.22364			12:30	220	0.30595
		…		…	Validation		…		…
		…		…			…		…
		11:30	74	0.37806			11:30	266	0.41547
	6/22	12:00	75	0.30006		6/26	12:00	267	0.37517
		12:30	76	0.27424	Testing		12:30	268	0.41833
		…		…			…		…
		…		…			…		…
		11:30	122	0.44647			11:30	314	0.33010
	6/23	12:00	123	0.35078					
		12:30	124	0.26012					
		…		…					
		…		…					
		11:30	170	0.22485					

B. *NARX model*

The NARX is an abbreviation of nonlinear autoregressive with *exogenous* input, based on the linear ARX model which is also called as NARX recurrent neural networks that is capable of modeling efficiently a time series dataset with feedback connections enclosing several layers of the network.

Some researchers suggest that NARX is one of a great methods in the Neural Network (e.g. much faster in convergence) [8-10]. The NARX is one of models which can be used for demonstrating nonlinear systems especially in modelling the time series datasets. In other words, they are called dynamic networks categories with feedback connections enclosing several layers of the network [9, 11, 14]. In this study, NARX used a feedback dynamic neural network with the structure as shown in Fig. 2.

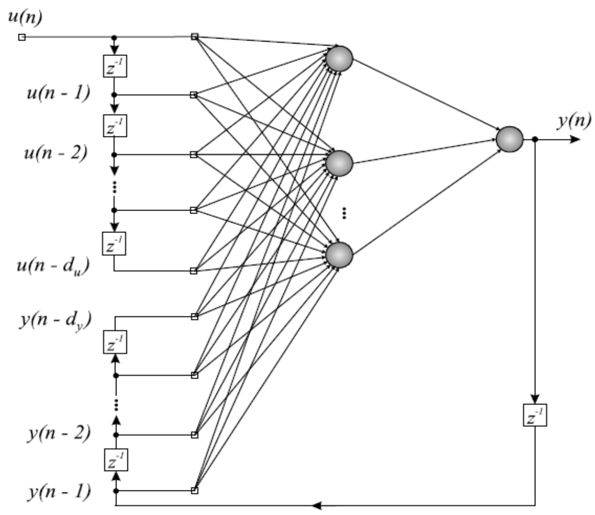


Fig. 2. The NARX with feedback dynamic neural network architecture

The inputs of NARX network consist of two categories: the external and the previous outputs to the network. The NARX model formula is defined as follows:

$$Y_{(n+1)} = f[y_{(n)}, \dots, y_{(n-d_y+1)}; u_{(n)}, u_{(n-1)}, \dots, u_{(n-d_u+1)}] \quad (1)$$

where u are the targets values; n are the past predicted values; d_y and d_u are the input and output orders; y is the exogenous variable (y_n and u_n denote the external (exogenous) output and input of network at time n); and f is a nonlinear function (generally used multilayer perceptron (MLP)); $d_y \leq 1$ and $d_u \geq 1$, $d_y \leq d_u$ are memory delays. NARX model may be written in vector form as

$$Y_{(n+1)} = f[y_{(n)}; u_{(n)}] \quad (2)$$

where $y_{(n)}$ and $u_{(n)}$ denote the external (exogenous) output and input.

In this study, NARX model can identify the exogenous y_n and u_n variables that effect the estimation of time series. The input order offers the number of past exogenous variables that are fed into the network. The steps of the NARX before the training process includes

a) Setup the datasets; the data will be divided into training, validation and testing. In this study, the data divided into three experiment: 189-31-94, 220-31-63, and 252-31-31 or 60%:10%:30%, 70%:10%:20%, and 80%:10%:10%).
b) Create the architecture of NARX configuration properly; The NARX model architecture is used 10 neurons in hidden layers with Levenberg-Marquardt (*trainlm*) learning algorithm. Levenberg-Marquardt has great computational optimization algorithm that can only be used in small networks and often characterized as more stable and efficient [9].
c) Diagnostic checking; to suggest alternative architecture(s) by changing the delay time, the datasets, and the number of neurons in hidden layer. In this experiment, diagnostic checking is used the variety numbers of delay values of each NARX model architecture.
d) Forecasting; the satisfactory model can be used for prediction purposes.

The statistical methods are used to measure the results of method NARX in predicting (actual and forecast data) [4, 9]. In this experiment, two criterion is used to evaluate the performance of the NARX prediction model. Firstly, minimizing the Mean Squared Error (MSE) is used as the first criteria and it is calculated as follow:

$$MSE \sum n\ i = 1\ (x_i - y_i)^2/n \quad (3)$$

Secondly, minimizing the correlation coefficient r is used as the second criteria as shown below, where, x_i is the observed network traffic; x is the mean of x_i; y_i is the predicted network traffic, y is the mean of y_i; n is the number of dataset. The best fit between observed and predicted values, which indicated by an increase in the MSE is 0 and r is 1 [15]. The correlation-coefficient r calculated as follow:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \quad (4)$$

III. Experimental Results

In this test, the time series data must be arranged in order of time in one period. The considered datasets are composed of 314 samples. Then, the datasets are divided into three partitions as training, validation and testing partitions. In this experiment, three architectures (e.g., 189-31-94, 220-31-63, and 252-31-31 or 60%:10%:30%, 70%:10%:20%, and 80%:10%:10%) have been setup to be performed. Then, the NARX model architecture with 10 neurons in hidden layers is also configured. Then, the variety numbers of delay values of each NARX model architecture with the Levenberg-Marquardt (*trainlm*) learning algorithm are also configured.

Furthermore, in order to determine the best NARX models, the training automatically stops indicated by the MSE have been satisfied. In other words, the MSE performance value is used to get the difference between the actual and predicted data. A comparison for all NARX model architecture by using the MSE and r values was carried out. The lower the MSE computed the better is the model in learning the time series data. The NARX model architecture of training results as shown in Table III.

IV. Results and Discussion

In this section, describes of the modeling method of NARX with feedback dynamic neural network. After successfully building and training of NARX models, some following result provides.

Based on Table III, the all values of r is almost all above 0.90 during training of network. This shows that the output produced by the network is closely similar to the target and that the model is satisfactory. After several experiments performed using the three architectures, the results show that the NARX model produces the best result when using parameters 189:31:94 or 60%:10%:30% in which performance training value produced was < 0.90.

TABLE III. The NARX training results

NARX Architecture (Training-Validation-Testing)	Memory Delay Time	Mean Square Error (MSE)	Correlation Coefficient r
	1:1	0.005472	0.90790
	1:2	0.005487	0.86392
189:31:94 (60%:10%:30%)	1:3	0.005221	0.88179
	1:4	0.006132	0.90029
	1:5	**0.006717**	**0.90764**
	1:6	0.005404	0.87530
	1:1	0.005382	0.90892
	1:2	0.005416	0.90612
220:31:63 (70%:10%:20%)	1:3	0.005351	0.90007
	1:4	0.005674	0.89996
	1:5	0.005056	0.88300
	1:6	0.006185	0.89373

252:31:31 (80%:10%:10%)	1:1	0.005455	0.91045
	1:2	0.005474	0.90896
	1:3	0.005650	0.90690
	1:4	0.005172	0.90944
	1:5	0.005416	0.91794
	1:6	0.004610	0.92144

The overall performance of NARX model above 0.90, and between 0.86 - 0.90 have been provided. In other words, the NARX architecture is possible to be used as a model to predict a time series dataset, as shown in Fig 3.

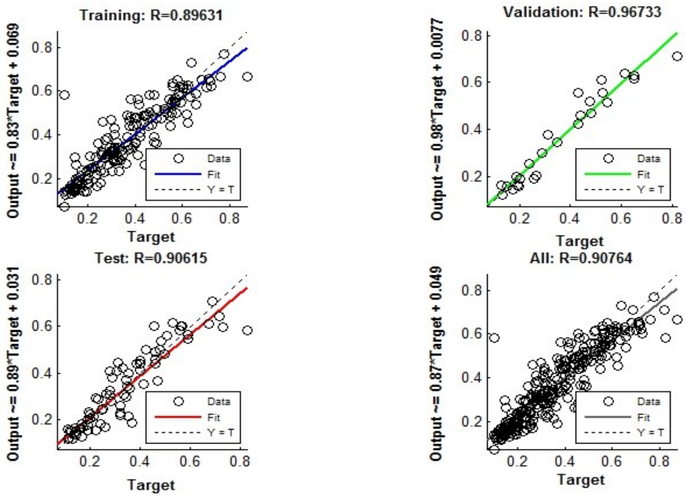
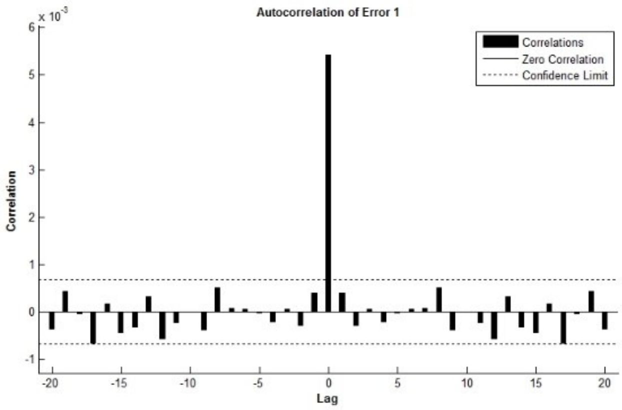
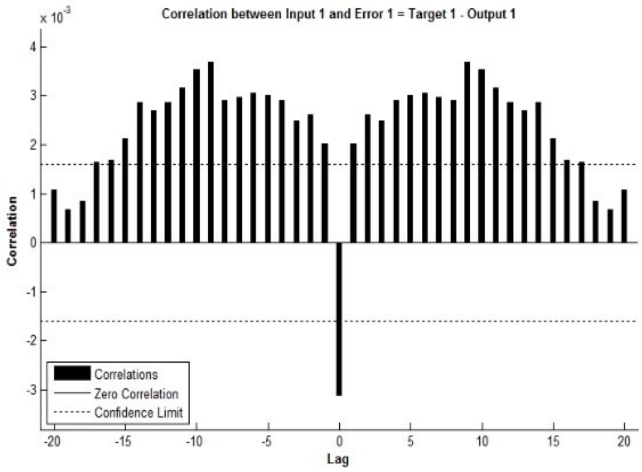


Fig. 3. The NARX of regression by using architecture 189:31:94 or 60%:10%:30%

Based on the prediction rule, there are three, namely error autocorrelation, input-error cross-correlation and time series response parameters that have been observed. For an accurate prediction model; error autocorrelation function of value should only be one non-zero, and it should hit at zero lag; this would imply that the expectation slips were totally uncorrelated with one another; Beside the one at zero slack, fall approximately within the 95% confidence limits around zero, so the model is from every angle attractive, are indicated in Fig 4 (a, b), 5, and 6.



(a)



(b)

Fig. 4. Error Autocorrelation (a) and Input-Error Cross-Corelation (b) by using architecture 189:31:94 or 60%:10%:30%

For this analysis, the best model in order to predict ahead based on performance of NARX methods by using MSE have been selected. Furthermore, the results obtained show that the best value of MSE and *r* for NARX architecture model is 0.006717 and 0.90764 respectively. Therefore, the NARX architecture model with configuration 189:31:94 or 60%:10%:30% can be used as an alternative model to predict the daily network traffic, because a pretty good MSE and *r* values has been produced. The plot can be seen in Fig 6.

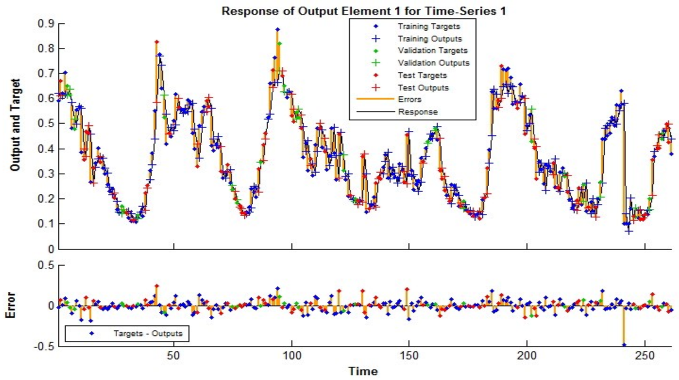


Fig. 5. The NARX of Time Series Response by using architecture 189:31:94 or 60%:10%:30%

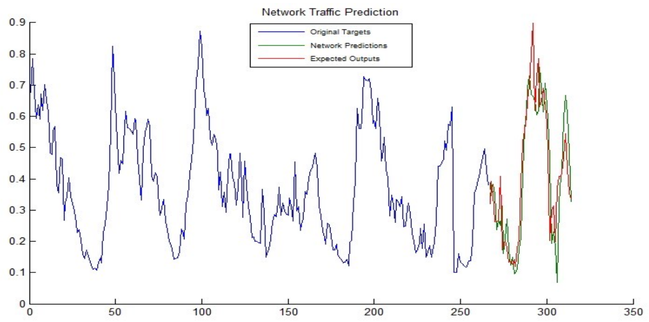


Fig. 6. Plot NARX of prediction by using architecture 189:31:94 or 60%:10%:30%

V. CONCLUSIONS

In conclusion, the daily network traffic dataset can be modelled by using the NARX. In this paper, the performance of the time series prediction model using the NARX architecture model is investigated. Based on the findings, the NARX architecture models can be used easily to model and predict time series daily network traffic dataset. The obtained results shown that NARX network can be successfully applied to complex univariate time series modelling and prediction tasks. The model's accuracy in predicting daily network traffic by MSE and *r* values have been measured. The NARX model can be considered as an alternative model for forecasting purposes of daily network traffic in the future. In future, we will evaluate the proposed approach on several other applications that require long-term time series predictions.

REFERENCES

[1] Haviluddin and A. Jawahir, "Comparing of ARIMA and RBFNN for short-term forecasting," *International Journal of Advances on Intelligent Informatics (IJAIN),* vol. 1, (1), March 2015, pp. 15-22, 2015.

[2] M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, "Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir," *Journal of Hydrology,* vol. 476, pp. 433–441, 2013.

[3] A. E. Permanasari, I. Hidayah, and I. A. Bustoni, "SARIMA (Seasonal ARIMA) Implementation on Time Series to Forecast The Number of Malaria Incidence," in *International Conference on Information Technology and Electrical Engineering (ICITEE)*, Yogyakarta, Indonesia, 2013, pp. 203 - 207.

[4] Haviluddin and R. Alfred, "Daily Network Traffic Prediction Based on Backpropagation Neural Network," *Australian Journal of Basic and Applied Sciences (AJBAS),* vol. 8 (24), Special 2014, pp. 164-169, 2014.

[5] G. Sermpinis, C. Dunis, J. Laws, and C. Stasinakis, "Forecasting and trading the EUR/USD exchange rate with stochastic Neural Network combination and time-varying leverage," *Decision Support Systems,* vol. 54, (2012), pp. 316–329, 2012.

[6] Haviluddin and I. Tahyudin, "Time Series Prediction Using Radial Basis Function Neural Network," *International Journal of Electrical and Computer Engineering (IJECE),* vol. 5, (4), August 2015, pp. xxx - xxx, 2015.

[7] Haviluddin and R. Alfred, "Forecasting Network Activities Using ARIMA Method," *Journal of Advances in Computer Networks (JACN),* vol. 2, (3) September 2014, pp. 173-179, 2014.

[8] A. Andalib and F. Atry, "Multi-step ahead forecasts for electricity prices using NARX: A new approach, a critical analysis of one-step ahead forecasts," *Energy Conversion and Management,* vol. 5, pp. 739–747, 2009.

[9] L. Xiaofeng and Y. Chunshan, "Research and Application of Data Mining and NARX Neural Networks in Load Forecasting," *International Journal of Database Theory and Application,* vol. 7, (2), 2014, pp. 13-24, 2014.

[10] J. M. P. Menezes Jr and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing,* vol. 71, (2008), pp. 3335–3343, 2008.

[11] A. Alzahrani, J. W. Kimball, and C. Dagli, "Predicting Solar Irradiance Using Time Series Neural Networks," in *Complex Adaptive Systems*, Philadelphia, PA, 2014, pp. 623 – 628.

[12] Purwanto, C. Eswaran, and R. Logeswaran, "A Comparison of ARIMA, Neural Network and Linear Regression Models for the Prediction of Infant Mortality Rate," in *2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation*, 2010.

[13] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. (2008). *Time Series Analysis Forecasting and Control 4th Edition*.

[14] H. Xie, H. Tang, and Y.-H. Liao, "Time Series Prediction Based on NARX Neural Networks: An Advanced Approach," in *The Eighth International Conference on Machine Learning and Cybernetics, 12-15 July 2009*, Boading, China, 2009, pp. 1275-1279.

[15] A. Khamis and S. N. S. B. Abdullah, "Forecasting Wheat Price Using Backpropagation And NARX Neural Network," *The International Journal Of Engineering And Science (IJES),* vol. 3, pp. 19-26, 2014.