Simulation Study for Determining the Best Architecture of Multilayer Perceptron for Forecasting Nonlinear Seasonal Time Series

To cite this article: Suhartono et al 2018 J. Phys.: Conf. Ser. 1028 012214

View the article online for updates and enhancements.

Related content
- Neural networks for emulation variational method for data assimilation in nonlinear dynamics
  Hélaine Cristina Morais Furtado, Haroldo Fraga de Campos Velho and Elbert E N Macau
- Neural network prediction of solar cycle 24
  A. Ajabshirizadeh, N. Masoumzadeh Jouzdani and Shahram Abbassi
- Time Series Forecasting of Daily Reference Evapotranspiration by Neural Network Ensemble Learning for Irrigation System
  N Manikumari, A Murugappan and G Vinodhini
Simulation Study for Determining the Best Architecture of Multilayer Perceptron for Forecasting Nonlinear Seasonal Time Series

Suhartono¹, Farah Fajrina Amalia², Priyandari Dina Saputri¹, Santi Puteri Rahayu¹, Brodjol Sutijo Suprih Ulama¹,²,³,⁴,⁵

¹,²,³,⁴,⁵ Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
suhartono@statistika.its.ac.id

Abstract: Neural network is one of flexible nonlinear models that could handle various relationship patterns on data with high accuracy. The data-driven approach is one of the advantages of neural network models in solving complex problems in forecasting. The selection of the best model becomes one of the most important problems in the application of neural network for time series forecasting, which consists of determining the input, the number of neurons in the hidden layer, the activation function, and preprocessing method. This paper focuses on the simulation study to explore how to determine the best architecture of multilayer perceptron for forecasting nonlinear seasonal time series. The data that be generated from seasonal exponential smooth transition autoregressive model are used as a case study. The results show that the inputs and the number of neurons in the hidden layer are two main factors that affect significantly the forecast accuracy. In contrary, the activation function and preprocessing method do not significantly influence the forecast accuracy.

1. Introduction

Time series is a sequential series of observations recorded from time to time. There are several objectives in the study of time series, such as learning patterns of a mechanism, forecasting for future data, and performing optimal control on a system. Seasonal and trend patterns are the basic components of many time-series data, especially in economics and business [1, 2]. There are several methods that have been widely used in modeling seasonal time series data, one of them is decomposition [3]. However, analysis by decomposition method can eliminate some important information in the data. Another method commonly used is the autoregressive integrated moving average (ARIMA). This method assumes that the main component in time series is linear, so there is difficulty to apply it when the data contain nonlinearity pattern [4].

Recent researches in forecasting suggests that one of the alternative methods that can deliver promising results is neural network [5]. Neural network is one of flexible nonlinear modeling that could handle various relationship patterns on data with high accuracy. Neural network model does not require assumptions, and the pattern of relationships underlying the model is determined by performing data mining. Data driven approach is one of the advantages of neural network model in solving complex problems in forecasting. Although the neural network model is closely related to nonlinearity, the method can also use well for modeling the linear process [6].
Several studies have been conducted by comparing neural networks with conventional forecasting methods to determine the potential of neural networks in time series forecasting [7]. Some of the results show that neural networks perform better than conventional methods [8, 9, 10, 11], but there are also studies show that conventional methods provide better performance than neural networks [12, 13]. Such contradictory results can be possibly caused by the different selection of neural network models used by researchers [7].

The selection of the best model becomes one of the important things in proper application of neural network in time series forecasting [14]. The model selection includes selecting the inputs, preprocessing method, activation functions, and the number of neurons in the hidden layer [15]. This paper focuses to evaluate the impact of inputs, preprocessing method, activation function and the number of neurons in hidden layer of the neural network model for forecasting nonlinear seasonal time series by using simulation study. Moreover, the purpose of this study is to compare the forecast accuracy of the inputs, number of neurons in hidden layer, activation function, and preprocessing method of neural network for forecasting nonlinear seasonal time series. The selection of the best model is done based on the concept of cross validation [16].

2. Methods

2.1. Neural Network

Neural network is a machine learning technique that has been developed as a generalization of the mathematical model of the biological nervous system. Learning methods in neural networks can be classified into three, namely supervised learning, unsupervised learning, and reinforcement learning. The neural network model commonly used in forecasting is feed forward neural network or multilayer perceptron [17]. The following is an example of feed forward neural network architecture with three layers.

![Three Layers Feed Forward Neural Network](image)

**Figure 1.** Three Layers Feed Forward Neural Network

The relationship between output \( Y_t \) and input \((Y_{i-1}, Y_{i-2}, ..., Y_{i-p})\) is described in the mathematical equation as follows:

\[
Y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j g(\beta_{ij}) + \sum_{i=1}^{p} \beta_0 Y_{i-1} + a,
\]  

where \( \alpha_j (j = 0,1,2, ..., q) \) and \( \beta_i (i = 0,1,2, ..., p; j = 1,2, ..., q) \) are parameter or often referred to as a weight, where \( p \) is the number of input nodes and \( q \) is the number of hidden nodes. The widely used function in the hidden layer is logistic or tangent hyperbolic. Equation (1) forms a nonlinear mapping of the observed value in the past, \((Y_{i-1}, Y_{i-2}, ..., Y_{i-p})\), to future value, \(Y_t\), with the following equation

\[
Y_t = f(Y_{i-1}, Y_{i-2}, ..., Y_{i-p}, w) + a,
\]  

where $w$ is the vector of all parameters and $f$ is a function that is determined from the network structure and weight. Thus, the neural network is equivalent to the nonlinear autoregressive model [17].

2.2. Criteria for Model Evaluation

After fitting the model, it is necessary to evaluate and select the best model. One criteria of model evaluation that frequently used is Root Mean Square Error (RMSE). RMSE is the most popular model selection criteria based on the residual that be calculated by the following formula [7],

$$RMSE = \sqrt{\frac{1}{L} \sum_{l=1}^{L} e_l^2},$$  \hspace{1cm} (3)

where
- $L$ = the number of out-sample or testing dataset
- $e_l = Y_{n+l} - \hat{Y}_n(l)$
- $Y_{n+l}$ = observed value of $l$-step ahead
- $\hat{Y}_n(l)$ = forecast value of $l$-step ahead.

2.3. Data Simulation

The simulation data follows the seasonal Exponential Smooth Transition Autoregressive (ESTAR) model, i.e.

$$Y_t = 6.5Y_{t-2} \exp(-0.25Y_{t-2}^2) + a_t$$  \hspace{1cm} (4)

where $a_t \sim$ IIDN(0,0.5). Simulated data is divided into two parts, i.e. the first 360 data as training data (in-sample) and the last 40 data as testing data (out-sample). The time series plot is as follows.

Figure 2 shows the time series plot of the generated data. This graph cannot describe the nonlinearity and seasonal component clearly. These characteristics can be visualized in lags plot as follows.

Figure 3 shows the relationship between the generated data and its lags, from lag 1 to 16. This graph indicates that there is a strong nonlinear relationship between the data and its lag 7, which means the observation at a certain time depends on the observation at 7 time unit previously.
The combination of inputs, the number of neurons, activation function, and preprocessing method that will be used in selecting neural network model are as follows.

Table 1. Combination of inputs, neurons, activation function, and preprocessing method

| Component | Combination of Inputs | Count |
|-----------|-----------------------|-------|
| Input (with \(Y_7\) and without \(Y_7\)) | 1. \(X_1, X_2, X_3, X_4, X_6, X_8, X_{10}, X_{12}\) | 10 |
| | 2. \(X_1, X_3, X_5, X_7, X_9, X_{11, X_13}\) | 10 |
| | 3. \(X_1, X_3, X_5, X_7, X_9, X_{11}, X_{13, X_15}\) | 6 |
| | 4. \(X_1, X_3, X_5, X_7, X_9, X_{11, X_13}\) | 4 |
| | 5. \(X_1, X_3, X_5, X_7, X_9, X_{11}\) | 2 |
| | 6. \(X_1, X_3, X_5, X_7, X_9, X_{11}\) | 2 |
| | 7. \(X_1, X_3, X_5, X_7, X_9, X_{11, X_13}\) | 2 |
| | 8. \(X_1, X_3, X_5, X_7, X_9, X_{11, X_13}\) | 2 |
| | 9. \(X_1, X_3, X_5, X_7, X_9, X_{11, X_13}\) | 2 |
| | 10. \(X_1, X_3, X_5, X_7, X_9, X_{11, X_13}\) | 1 |
| Total | 41 |
| Number of neurons | 1, 2, 3, 4, 5, 10, and 15 | 7 |
| Activation function | Hyperbolic Tangent and Sigmoid | 2 |
| Preprocessing method | Standardized, Normalized, and Adjusted Normalized | 3 |

Based on these combinations, the number of neural network models formed is 1722 models (41 inputs \(\times\) 7 neurons \(\times\) 2 activation functions \(\times\) 3 preprocessings).

3. Results

3.1. Selection of Inputs and Number of Neurons

The data used in the simulation is generated using its lag 7. To determine whether proper selection of inputs can lead to differences of forecast accuracy, the RMSE values of 41 input combinations involving lag 7 and without involving lag 7 as inputs are compared. Furthermore, to find out whether...
the number of neurons affect the forecast accuracy, the plot will be separated based on the number of neurons used in neural network as the following graphs.

![Figure 4](image-url)

**Figure 4. The Results of RMSE Comparison of 41 Input Combinations and 7 Number of Neurons for (a) Training and (b) Testing Data**

Figure 4 indicates that the number of neurons used in the hidden layer has an important role in neural network modeling. The use of 1 neuron in the hidden layer tend to give larger RMSE compared to 2, 3, 4, 5, 10, and 15 neurons, both for training and testing data. However, the addition of neurons does not always yield smaller RMSE. Hence, in the selection of the number of neurons, trial and error need to be done to determine the optimal number of neurons that yield the smallest RMSE in testing data.

The combination of inputs without involving lag 7 yields larger RMSE than the combination of inputs involving lag 7, both for training and testing data. Moreover, the use of lag 7 alone as an input has the smallest average of RMSE in number of neurons 1, 2, and 5 for training data, and has the smallest average of RMSE in number of neurons 3, 5, and 15 in testing data, respectively.

### 3.2. Selection of The Activation Function

Figure 5 shows the RMSE comparison between the activation function that be used for forecasting simulated data. It shows that the hyperbolic tangent activation function yields a slightly lower RMSE than the sigmoid activation function, but the difference is not significant. Thus, it can be concluded that the activation function has no significant effect on the forecast accuracy.

To study further about the effect of combination between activation function and inputs, several neural network architectures are compared and the results could be seen at Figure 6. The figures indicate that the hyperbolic tangent activation function also produces a slightly lower RMSE than the sigmoid activation function only on the model with involving the $Y_{t-7}$ as inputs. Moreover, the results show that in the model without the $Y_{t-7}$ as inputs, both activation functions yield very close RMSE values.
To find out whether preprocessing method on inputs influences forecast accuracy, a comparison is done between 3 preprocessing methods, i.e. standardized, adjusted normalized, and normalized. Then, RMSE of each scenario is calculated and shown at Figure 7.
Based on Figures 7, it is known that the normalized preprocessing yields the highest RMSE value. Standardized and adjusted normalized preprocessing give lower RMSE values than normalized, but the RMSE values from both preprocessing are not significantly different. To know further about the effect of preprocessing method, several combination preprocessing method and inputs are compared and the results are shown at Figure 8.

![Figure 7](image1.png)

**Figure 7.** The Results of RMSE Comparison of Pre-processing Methods and Number of Neurons in (a) Training and (b) Testing Data

![Figure 8](image2.png)

**Figure 8.** The Results of RMSE Comparison of Pre-processing Methods and Inputs in (a) Training and (b) Testing Data

By exploring various combinations of inputs, number of neurons, activation function, and preprocessing method, five models with the smallest RMSE at testing data are obtained as shown at Table 2.
Table 2. The best five models (smallest RMSE)

| No. | Activation Function | Input | Preprocessing          | Number of neurons | RMSE  |
|-----|---------------------|-------|------------------------|-------------------|-------|
| 1.  | Tanh                | $Y_{5,1}$, $Y_{5,6}$, $Y_{5,7}$ | Adj Normalized    | 4                 | 0.502 | 0.449 |
| 2.  | Tanh                | $Y_{5,7}$                      | Standardized       | 5                 | 0.501 | 0.454 |
| 3.  | Tanh                | $Y_{5,1}$, $Y_{5,6}$, $Y_{5,7}$ | Standardized       | 3                 | 0.506 | 0.460 |
| 4.  | Tanh                | $Y_{5,1}$, $Y_{5,6}$, $Y_{5,7}$ | Standardized       | 5                 | 0.504 | 0.461 |
| 5.  | Tanh                | $Y_{5,1}$, $Y_{5,7}$          | Standardized       | 4                 | 0.502 | 0.466 |

The results on Table 2 show that the RMSE values obtained from five best models are close to 0.5 which is the actual value of generated standard deviation in simulation data. Hence, it indicates that multilayer perceptron or feedforward neural network models are capable for modeling well the simulated data.

4. Conclusions
Based on the results of this simulation study, it could be concluded that inputs and number of neurons in hidden layer are two main determinant factors that influence the forecast accuracy of multilayer perceptron or neural networks for forecasting nonlinear seasonal time series. This study shows that determination of appropriate inputs will lead to the best neural network model with high forecast accuracy. Moreover, the results also show that selection of the number of neurons in hidden layer influences the forecast accuracy, but it takes trial and error to determine the optimal number of neurons in hidden layer. Additionally, this study also shows that activation function and preprocessing method do not have a significant effect on forecast accuracy. Hence, further research is needed to validate the results of this simulation study to the real data and compare the forecast accuracy between neural network and other both classical and modern time series models.

Acknowledgments:
This research was supported by DRPM-DIKTI under scheme of “Penelitian Berbasis Kompetensi”, project No. 532/PKS/ITS/2017. The authors thank to the General Director of DIKTI for funding and to anonymous referees for useful suggestions.

References
[1] W. W. S. Wei, Time Series Analysis : Univariate and Multivariate Methods, Boston : Pearson: Addison Wesley, 2006.
[2] Suhartono, M. H. Lee and D. D. Prastyo, "Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects," in 2nd Innovation and Analytics Conference and Exhibition, IACE 2015, Kedah, Malaysia, 2015.
[3] R. Chen and R. S. Tsay, " Functional-coefficient autoregressive models," Journal of the American Statistical Association, vol. 88, no. 421, pp. 298-308, 1993.
[4] W. Ruan, Q. Z. Sheng, P. Xu, N. K. Tran, N. J. Falkner, X. Li and W. E. Zhang, "Forecasting Seasonal Time Series Using Weighted Gradient RBF Network based Autoregressive Model," in Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, Indiana, USA, 2016.
[5] R. Benzer and S. Benzer , "Application of artificial neural network into the freshwater fish caught in Turkey," International Journal of Fisheries and Aquatic Studies, vol. 2, no. 5, pp. 341-346, 2015.
[6] G. P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series," *European Journal of Operational Research*, pp. 501-514, 2005.

[7] A. Serkan and İ. D. Kocakoç, "A new model selection strategy in time series forecasting with artificial neural networks: IHTS," *Neurocomputing*, vol. 174, pp. 974-987, 2016.

[8] H. F. Zou, G. P. Xia, F. T. Yang and H. Y. Wang, "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting," *Neurocomputing*, vol. 70, pp. 2913-2923, 2007.

[9] R. Adhikari and R. K. Agrawal, "A combination of artificial neural network and random walk models for financial time series forecasting," *Neural Computing and Application*, vol. 24, no. 6, pp. 1441-1449, 2014.

[10] P. Sriraeng, G. S. Baxter and G. Wild, "Forecasting demand for low cost carriers in Australia using an artificial neural network approach," *Aviation*, vol. 9, no. 2, pp. 90-103, 2015.

[11] A. Jain and A. M. Kumar, "Hybrid neural network models for hydrologic time series forecasting," *Applied Soft Computing*, vol. 7, pp. 585–592, 2007.

[12] J. L. Callen, C. C. Y. Kwan, P. C. Y. Yip and Y. F. Yuan, "Neural network forecasting of quarterly accounting earnings," *International Journal of Forecasting*, vol. 12, pp. 475-482, 1996.

[13] W. L. Gorr, D. Nagin and J. Szczypula, "Comparative study of artificial neural network and statistical models for predicting student grade point averages," *International Journal of Forecasting*, vol. 10, pp. 17-34, 1994.

[14] S. D. Balkin and J. K. Ord, "Automatic neural network modeling for univariate time series," *International Journal of Forecasting*, vol. 16, pp. 509-515, 2000.

[15] W. Sarle, "Neural networks and statistical models," in *Proceedings Of The 19th Annual SAS Users Group International Conference*, Cary, NC, 1994.

[16] U. Anders and O. Korn, "Model selection in neural networks," *Neural networks*, vol. 12, pp. 309-323, 1999.

[17] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, no. 20, pp. 159-175, 2003.