COMPARISION OF TIME SERIES FORECASTING METHODS USING NEURAL NETWORKS AND BOX-JENKINS MODEL

Ani bin Shabri
Jabatan Matematik, Fakulti Sains UTM
Skudai Johor Malaysia
e-mail : shabri99@hotmail.com

Abstract The performance of the Box-Jenkins methods is compared with that of the neural networks in forecasting time series. Five time series of different complexities are built using back propagation neural networks were compared with the standard Box-Jenkins model. It is found that for time series with seasonal pattern, both methods produced comparable results. However, for series with irregular pattern, the Box-Jenkins outperformed the neural networks model. Results also show that neural networks are robust, provide good long-term forecasting, and represent a promising alternative method for forecasting.

Keywords Neural Networks, Back Propagation, Forecasting, Robust.

Abstrak Kajian ini membincangkan keupayaan pengkaedahan Box-Jenkins bila dibandingkan dengan kaedah Rangkaian Neural dalam peramalan siri masa. Lima siri masa yang kompleks dibangunkan menggunakan kaedah rambatan balik Rangkaian Neural dan dibandingkan dengan model Box-Jenkins yang piawai. Analisis kajian menunjukkan bahawa bagi data siri masa bermusim, kedua-dua kaedah menghasilkan keputusan yang setanding. Walau bagaimana pun, untuk siri masa yang berbentuk tidak menentu, kaedah Box-Jenkins menghasilkan keputusan yang kurang baik berbanding Rangkaian Neural. Hasil ini juga menunjukkan bahawa Rangkaian Neural adalah teguh, menghasilkan peramalan yang baik untuk jangka panjang, dan boleh menjadi kaedah alternatif untuk peramalan.

Katakunci Rangkaian Neural, Rambatan Balik, Peramalan, Teguh.

1   Introduction

Time series forecasting is highly utilised in predicting economic and business trends. Many forecasting methods have been developed in the last few decades. The Box-Jenkins method is one of the most widely used time series forecasting methods in practice [1]. It is also one of the most popular models in traditional time series forecasting and is often used as a benchmark model for comparison with neural networks (Holger and Graeme [2], Tang and Fishwick [4], and Zhang, Patuwo and Hu [7]).

Recently, artificial neural networks (NN) that serve, as a powerful computational framework, have gained much popularity in business applications. Neural networks have been successfully applied to loan evaluation, signature recognition, time series forecasting, classification analysis and many other difficult pattern recognition problems (Tang and

Fishwick [3]). However, the use of neural networks is rapidly increasing, and in recent years they have been successfully used in prediction in economic, business and hydrology.

Concerning the application of neural networks to time series forecasting, there have been many reviews. Sharda and Patil [3] conducted a comparison between neural networks model and the Box-Jenkins method. They concluded that the simple neural networks could forecast about as well as the Box-Jenkins system. Tang and Fishwick [4] reported that for time series with long memory, the Box-Jenkins and the neural networks methods produced comparable results. However for series without memory, the neural networks outperformed the Box-Jenkins model.

In this paper, a comparative study is carried out to investigate the forecasting capability of neural networks and Box-Jenkins model, which are among those forecasting models most successfully, applied in practice.

2   The Box-Jenkins Method

The Box-Jenkins method is one of the most popular time series forecasting methods in business and economics. The method uses a systematic procedure to select an appropriate model from a rich family of models, namely, Integrated Autoregressive Moving Average (ARIMA) models. A general ARIMA model has the following form (Bowerman and O'Connell [1]).

$$\Phi_p(B)\Phi_P(B^L)(1-B^L)^D(1-B)^d\, y_t = a + \Theta_q(B)\Theta_Q(B^L)\varepsilon_t$$

where $\Phi(B)$ and $\Theta(B)$ are autoregressive and moving average operators respectively; B is the back shift operator; ε_t is random error with normal distribution $N(0,\sigma^2)$; a is a constant, and y_t is the time series data, transformed if necessary.

The Box-Jenkins method performs forecasting through the following process :
1. Model identification: Historical data is used to tentatively identify an appropriate Box-Jenkins model.
2. Estimation: Estimate the parameters of the tentatively identified model.
3. Diagnostic checking: Various diagnostics are used to check the adequacy of the estimated model and if need be, alternative models may be considered.
4. Forecasting: The best model chosen is used for forecasting.

3   Artificial Neural Networks

Artificial Neural Networks (ANN) are forms of computing inspired by the functioning of the brain and nervous system. The structure and operation of ANNs have been discussed by numerous authors (Holger and Graeme [2], Toth, Barth and Montanari [5]).

Structure. The processing elements (PE) are usually arranged in layers: an input layer, an output layer, and one or more layers in between, called hidden layers
(Figure 1). The connections between the PEs are weighted. The strength of each connection weight can be adjusted; a zero weight represents the absence of a connection.
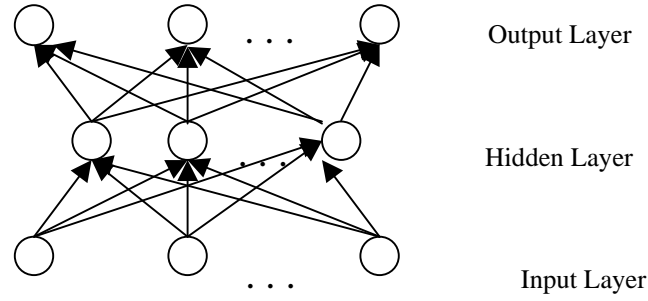
Figure 1: Typical Artificial Neural Network

Operation. The propagation of data through the networks starts with the introduction of an input stimulus at the input layer. The data then flow through, and are operated on by, the networks until an output stimulus is produced at the output layer.

The architecture of a typical node is shown in Figure 2. Each PE receives the weighted outputs $(w_{ji}x_i)$ from PEs in the previous layer, which are summed and added to a threshold value (θ_j) to produce the node input. The node input of this process is given by

$$I_j = \sum_i w_{ij}x_i + \theta_j$$
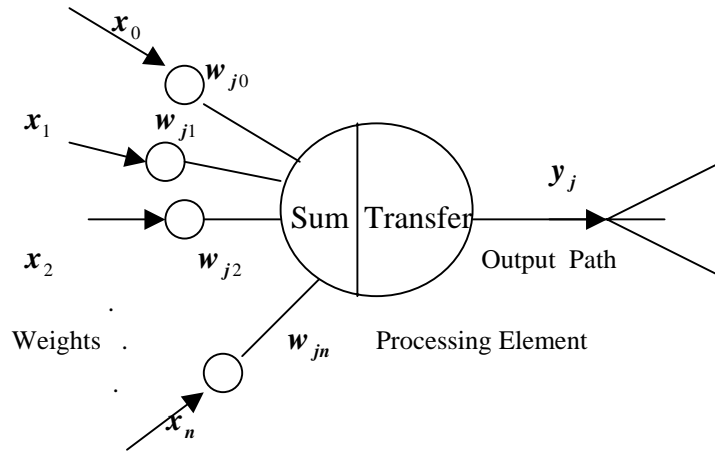
Figure 2 : Operation of a Typical Processing Element

The purpose of the threshold is to scale the input to a useful range. The node input (I_j) is then passed through a nonlinear transfer function $(f(I_j))$, such as hyperbolic tangent function (Figure 2) to produce the node output (y_j), which is passed to the weighted input paths of many other PEs.

Learning. Learning is the process in which the weights are adjusted in response to training data provided at the input layer and depending on the learning rule, at the output layer.

During training, the output predicted by the networks $(y_j(t))$ is compared with the actual output $(d_j(t))$ and the mean squared error (MSE) between the two is calculated. The error function at time t, $E(t)$ is given by

$$E(t) = (1/2)\sum \left(y_j(t) - d_j(t)\right)^2$$

The aim of the training is to find a set of weights that will minimize the error function. Initially, the weights are assigned as small, arbitrary values. The weights are updated systematically using a learning rule during learning progress. There are several methods of finding the weight increment of which the gradient descent method is most common.

4   Data

The five time series are selected. The data series are labeled and they are shown in figure 3. The monthly total International Airline Passengers for 1949-1960, monthly Hotel Room Average for 1973-1986 (Bowerman and O'Connell [1]) ,and Water Demand in Utopia for 1963-1982 data show an increasing trend and seasonal pattern. The Standard Malaysia Rubber 20 (SMR20) for 1972-1991 and Palm Oil Production in Malaysia for 1978-1991 also show seasonal patterns though not very clear.

5   Forecasting Results

The data series are normalized in the range [-1, 1] before feeding into the neural networks. Forecasts from the neural networks output were transformed to the original data scale before the mean error was computed.

Although there has been some research on the design of optimal neural networks structure, it is still largely an art to determine the number of hidden layers and number of units in each hidden layer. Neural networks structure is denoted as I x H x O, where I, H and O represent number of input units, hidden units, and output units respectively.

In this study, the number of hidden units equals the number of input units, which is set to be 12 corresponding to twelve months. The number of output units, also set to be 12, corresponds to 12 periods forecasting because the series data had seasonal pattern with 12 period seasonal. A back propagation neural networks, the most popular networks in business application (Tang and Fishwick [4]), is used in this study. A sigmoid hyperbolic tangent function is used as the activation function. The generalized delta learning rule and the quadratic function were used. The initial weights were randomly distributed between +0.1 and –0.1. The performance of each model was measured by computing the root mean square error (RMSE). The model which has the lowest RMSE was then used for forecasting.

Of the original time series, the last 12 data series are saved and they are not used in model building process. It uses to compare with 12 step period forecasted value.

Box-Jenkins model forecasting was carried out with the time series analysis package STATGRAFIC.

The forecasting results are summarized in Table 1. Note that all values represent the mean square error for the 12 period forecast.

Table 1: Mean Square Error (MSE) for 12 Period Forecasting using The Neural Networks and The Box-Jenkins Model

Data									
Airline		SMR20		Water Demand		Palm Oil		Hotel Room	
BJ[1]	NN[2]	BJ	NN	BJ	NN	BJ	NN	BJ	NN
0.852	1.217	0.923	0.749	0.088	0.106	0.037	0.036	0.704	0.860

Here **BJ**[1] – Box-Jenkins Model.

Here **NN**[2] – Neural Networks Model.

Table 1 shows that the performances of neural networks compared with Box-Jenkins model for multiple-step-period forecasting are different for the all series. For SMR20 and Palm Oil data, with for a 12-step-period the neural networks is better. For another data series, the neural networks seem worse. A neural networks and Box-Jenkins model are shown in figure 2, where the last series value is compared with 12 period ahead forecast. Figure 2 shows clearly that when stepwise forecasting is used for 12-step-period forecast, the neural networks models is a good performer compared to the Box-Jenkins method in all series data except SMR series data. This suggests that the neural networks is a better choice for long term forecasting. Examining the forecast error resulting from the Box-Jenkins model and the neural networks, the neural networks provides very nice forecasts for all time series.

6   Conclusion

Neural networks provide a promising alternative approach to time series forecasting. For time series with trend and seasonal pattern, both Box-Jenkins model and neural networks model perform well. The Box-Jenkins model can nicely forecast when the data series have trend and seasonal pattern. The neural networks found nicely forecast any series data. There are no systematic procedures for building neural networks forecasting models. The performance of neural networks model depends on many factors such as the nature of data series, the networks structure, and the training procedure.

Reference

[1]   B. L. Bowerman and R. T. O'Connell, *Time Series Forecasting : Unified Concepts and Computer Implementation,* 2nd Edition, Duxbury Press, USA., 1987.

[2]   R. Holger and C. D. Graeme, *The Use of Artificial Neural Networks For the Prediction of Water Quality Parameters,* Water Resource Research, Vol.32, 1013-1022, 1996.

[3]   R. Sharda and R. Patil, *Neural Network as Forecasting Experts : An Empirical Test.* International Joint Conference on Neural Networks, Vol. 1, 491-494, 1990.

[4]   Z. Tang and P. A. Fishwick, *Feedforward Neural Networks as Models For Time Series Forecasting,* ORSA Journal on Computing 5, 374-85, 1993.

[5]   E. Toth, A. Barth and A. Montanari, *Comparison of Short-Term Rainfall Prediction Models For Real-Time Flood Forecasting*, Journal of Hydrology, Vol.239,   132-147, 2000.

[6]   J. Yao, Y. Li, Cl Tan, *Option Price Forecasting Using Neural Networks.* The International Journal of Management Science, Vol. 28, 455-466, 2000.

[7]   G. P. Zhang, B. E. Patuwo and M. Y. Hu, *A Simulation Study of Artificial Neural Networks For Nonlinear Time-Series Forecasting,* Computer & Operations Research, Vol. 28, 381-396, 2001.