2011 International Conference on Circuits, System and Simulation
Proc .of CSIT vol.7 (2011) © (2011) IACSIT Press, Singapore

A Comparison on Neural Network Forecasting

Hong-Choon Ong [1, a] and Shin-Yue Chan [1, b]

[1] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia.
[a] hcong@cs.usm.my, [b] shinyue_alu@yahoo.com

Abstract. This study compares the effectiveness of the Box-Jenkins model and neural networks model in making a forecast. An eighteen years bimonthly water consumption in Penang data set is analyzed. Multi-layer perceptron (MLP) in neural network with single hidden layer and double hidden layers using error-backpropagation algorithm models are used. Four MLP programs with double hidden layers which are the original data (O), the deseasonalized with the smoothing moving average method (DS), the linear detrended model (DTL) and both deseasonalized and detrended model (DSTL) and a single hidden layer MLP with both deseasonalized and detrended (DSTL) were simulated. For time series analysis, a Box-Jenkins SARIMA model is generated. The performance of the models are measured using three types of error measurement, which are mean absolute error (MAE), absolute percentage error (MAPE) and root mean squared error (RMSE). The model with smallest value of MAE, MAPE and RMSE stands out to be the best model in predicting the water consumption in Penang for the year 2008. The results showed that both SARIMA and double hidden layers MLP models perform relatively well. Furthermore, double hidden layers MLP model shows an improvement in prediction as compared with the single hidden layer MLP model.

Keywords: neural network; multi-layer perceptron; Box-Jenkins SARIMA model.

1. Introduction

Neural network has the ability to represent both linear and non-linear relationships and the network can learn these relationships directly from the raw data. Multi-layered Perceptron (MLP) model is the most common model in neural network and is a feed-forward model (Looney, 1997). Each node in one layer connects with a certain weight to every other node in the following layer. In general, more hidden layers can prevent the necessity of using unnecessary neurons to achieve highly non-linear classification. The error back-propagation algorithm used adjusts the weights to obtain more accurate forecast values. A time series is a sequence of data points (observations), collected typically at successive times, spaced at often uniform time intervals. Data collected inconsistently or only once may not be considered as a time series. Time series analysis consists of methods that attempt to understand the specific time series, identifying the underlying context of the observations and fit a suitable model to make predictions. Time series forecasting means the use of a model to predict future events based on the past data; to forecast future data points before they are measured. Time series can be applied into various fields, such as economic forecasting, sales forecasting, stock market analysis, process and quality control, inventory forecasting, production planning and many more (Chatfield, 2004).

2. A Brief Literature Review

In recent years, artificial neural networks have been widely used as a useful tool to model many areas of engineering applications. Sarıdemir (2008) presented the models in artificial neural networks (ANN) for predicting compressive strength of concretes containing metakaolin and silica fume that had been developed at the age of 1, 3, 7, 28, 56, 90 and 180 days. The multi-layered feed forward neural networks models were fed with these parameters to predict the compressive strength values of concretes containing metakaolin and silica fume. The results of this study showed that neural networks have strong potential for prediction, which proved that artificial neural networks are capable of learning and generalizing from examples and experiences. Ong, et al. (2008) gave on a functional approximation comparison between neural networks and polynomial regression. The results showed that approximation using polynomial regression generates lower fraction of variance unexplained (FVU) value as compared to approximation using neural networks with single hidden layer multi-layered perceptron (MLP) and double hidden layer MLP except for the complicated functions. Moreover, double hidden layer MLP showed better estimation results than single hidden layer MLP without considering the number of parameters being used. On the other hand, statistical models have commonly been used in time series data analysis and forecasting. Zou and Yang (2003) proposed combining time series model for forecasting for a better performance. Lam, *et al.* (2008) used autoregressive integrated moving average (ARIMA) model to measure the intervention effects and the asymptotic change in the simulation results of the business process reengineering that is based on the activity model analysis. A case study of a purchasing process of a household appliance manufacturing enterprise that involves 20 purchasing activities was used as the training data. The results indicated that the changes can be explicitly quantified and the effects of business process reengineering can be measured. In general, many time series are asymptotically unstable and intrinsically non-stationary. Box-Jenkins models solve these problems by imposing on data transformations, such as differencing and logarithm. Grillenzoni (1998) discussed on a method for modeling time series with unstable roots and changing parameters. A method of adaptive forecasting which is based on the optimization of recursive estimators was applied to well-known data sets. He had demonstrated its validity in several implementations and for different model structures.

3. Methodology

3.1. Neural network –multi-layered perceptron

Multi-layered perceptron is the most common neural network model. It's abilities in achieving highly nonlinear classification and noise tolerance make it suitable to be used in this study. Bipolar sigmoid function is applied in this study. In a single hidden layer model, the perceptron computes associated outputs from multiple real-valued inputs by forming a linear combination according to its input weights. The data set is treated as inputs and denoted as $\{ x_i \}$, where i = number of data and has an associated set of exemplar outputs target vector. A random initial synaptic weight set for the hidden and output layers are selected. Finally, the current weights are re-iterated and updated using the error back-propagation algorithm to force each of the input exemplar feature vectors to be mapped closer to the target vector that identifies a class in the input space. In the double hidden layered MLP, the perceptrons also computes associated outputs from multiple real-valued inputs by forming a linear combination according to its input weights and the data set is treated as inputs. The main difference is that the synaptic weight set has an additional second hidden layer. Error back-propagation is used to update the error computed in the network with respect to the network's modifiable weights. The error value which is the difference between output value and desired value is then computed. The errors will be propagated back into the network using a gradient descent technique so that weights adjustment can be done.

3.2. Design of Box-Jenkins model

There are three primary stages in building a Box-Jenkins time series model, which are model identification, model estimation and model diagnostics.

1) Model identification: There is a need to determine if the series is stationary, has a trend component and if there is any sign of seasonality that needs to be modeled. Stationarity, trend and seasonality can be detected by plotting the time series graph. If there is trend component present in the data, the graph will not be in a random form. Stationarity can also be checked via sample auto-correlation (ACF). In a stationary series, the auto-covariance function, γ_k and the auto-correlation function, ρ_k show the following properties:

1. $|\gamma_k| \leq \gamma_0$.
2. $\rho_0 = 1$ and $-1 < \rho_k < 1$ for $k \neq 1$.
3. $\gamma_k = \gamma_{-k}$ and $\rho_k = \rho_{-k}$ for all k.

Seasonality can be accessed from sample auto-correlation (ACF) at various lag k and also sample partial auto-correlation (PACF) at various lag k of the process. The behavior of the ACF and PACF is shown in Table 1. After stationarity and seasonality have been addressed, the next step is to identify the order (the p and q) of the autoregressive and moving average terms. The behavior of ACF and PACF will indicate the order form of the autoregressive and moving average terms.

TABLE I. BEHAVIOR OF ACF AND PACF

	AR(p)	MA(q)	ARMA(p,q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

2) Model estimation: Parameters estimation for the Box-Jenkins models is a quite complicated non-linear estimation problem. Therefore, high quality software program that fits Box-Jenkins model should be used to estimate the parameter (Anderson, 1994).

Consider Y_t is stationary time series and $\{\varepsilon_t\}$ are independently and identically distributed with $N(0,\sigma_\varepsilon^2)$, an ARMA(p,q) model with unknown parameters $\varphi=\{\varphi_1,\varphi_2,......\varphi_p\}$, $\theta=\{\theta_1,\theta_2,......\theta_q\}$ and $\sigma_\varepsilon^2=E(\varepsilon_t^2)$ is given as:

$$Y_t=\varphi_1 Y_{t-1}+\varphi_2 Y_{t-2}+......+\varphi_p Y_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-......-\theta_q\varepsilon_{t-q}$$

The main approaches to fitting Box-Jenkins models are maximum likelihood estimation and least squares estimation.

3) Model diagnostics: The ACF plot and PACF plot of the estimated residual $\{\hat{\varepsilon}_t\}$ have the ability to indicate whether the residuals are independent and normally distributed. If the ACF plot and PACF plot show the points are distributed randomly or do not show any pattern, then this indicates that the residuals are independent and normally distributed. However, if these assumptions are not satisfied, a more appropriate model needs to be fitted. This means that, the model identification step has to be redone to re-develop a better model that satisfies the diagnostic test. Although a good model may not be obtained on the first trial, however a new and better model can be easily derived (Wei, 1990). Correlogram of residuals and correlogram of residuals squared are sufficient for determining the status of white noise. Q-statistics is useful to test normality of the residuals within a range of an independent variable. In other words, it served as a tool to test whether the series is white noise. The last two columns in the correlogram represent Ljung-Box Q-statistics and their p-value respectively. The Q-statistics at lag k is a test statistics for the null hypothesis where there is no auto-correlation up to order k. It is computed as follows:

$$Q=T(T+2)\sum_{j=1}^{k}(T-j)^{-1}r_j^2$$

where r_j represents the j-th auto-correlation and T represents number of observations. If the p-value is less than 0.05, then it is significant. Therefore, the null hypothesis is rejected, and the error terms are correlated. If the p-values are greater or equal to 0.05, then the p-values are not significant at all lags. Thus, the null hypothesis is not rejected, and the residuals are not correlated (Harvey, 2001).

3.3. Criteria comparison by error measurement

Three types of error measurement are used in this study, which are the mean absolute error (MAE), the percentage error (MAPE) and the root mean squared error (RMSE).

Mean Absolute Error (MAE)

MAE = Average of the absolute values of forecasting error

$$=\frac{\sum_{i=1}^{I}\left|X_i-\hat{X}_i\right|}{I}$$

where i = 1, 2, ..., I, X_i = original data, $\hat{X}_i$ = forecast data

Mean Absolute Percentage Error (MAPE)

MAPE = Percentage of the mean ratio of the error to the original data