SCIREA Journal of Mathematics

http://www.scirea.org/journal/Mathematics

January 11, 2017

Volume 1, Issue 2, December 2016

Modelling Singapore Tourist Arrivals to Malaysia by Using SVM and ANN

Rafidah Ali[1,a*] and Ani Shabri[2,b]

[1] University Kuala Lumpur Malaysia Institute of Industrial Technology, Persiaran Sinaran Ilmu, Bandar Seri Alam,81750 Johor Bharu Johor Malaysia

[2]Departments of Mathematics, Science Faculty, University of Technology Malaysia, Skudai, Johor Malaysia

Email: [a]rafidahali@mitec.unikl.edu.my, [b] ani@utm.my,

ABSTRACT.

The tourism industry is an increasingly important national industry for Malaysia. Government policymakers and business managers pay close attention to the development of the tourism industry. In this study, two machine learning methods, artificial neural network (ANN) and support vector machine (SVM) were utilized to predict the Singapore tourist arrival to Malaysia. This country is neighbouring country for Malaysia and tourists are more flexible to visit Malaysia using rails and road transportations. This paper aims at finding an accurate forecasting model in order to make the tourism industry grow stably. This study uses monthly time series data from 2010 (January) until 2014 (December). The experiment shows that ANN model outperform SVM base on the criteria Root Mean Absolute Error (RMSE).

Keywords: Support Vector Machine (SVM), Artificial Intelligence (AI), Artificial Neural Network (ANN)

1.INTRODUCTION.

Refer [1] in the past few decades, the tourism industry has emerged as the fastest growing sector, and has spread widely around the world. Tourism expenditure has become an important source of economic activity, employment, tax revenue, income and foreign exchange. Therefore, every country needs to understand its international visitors and tourism receipts, to help formulate responsive policies on tourism quickly. The accurate forecasting tourism demand would facilitate for assisting managerial, operational and tactical decision making of both the private and the public sector. Therefore the selection of forecasting model is the important criteria that will influence to the forecasting accuracy [10]. An accurate estimation tourist arrival requires expert systems such as Support Vector Machine (SVM) and Artificial Neural Networks (ANN) [2].

This paper focuses on the application of new approach to solve problem nonlinear patterns. According to the literature, there were two methods machine learning which were popular for forecasting time series data. Among these techniques were artificial neural network (ANN) and support vector machine (SVM). Since the performance of these approaches was still questionable, empirical study was always utilized as a basis to benchmark these techniques [8].

There has been a growing interest in tourism demand research over the past decades base on Claveria and Torra. [3] The objectives of this paper are to predict Singapore Tourist Arrivals to Malaysia by using SVM model and ANN model, to determine applicability of machine learning SVM and ANN in modelling the prediction of Singapore Tourist Arrivals time series and to compare the prediction results by SVM model with ANN model. The research will be focused on the forecasting Singapore Tourist Arrivals by using SVM and ARIMA model. In this paper, the data gather from monthly Singapore Tourist Arrivals Malaysia.. The data was collected from January 1999 to December 2015. The monthly Singapore Tourist Arrivals were selected for this study.

This paper is organized as follows. Section 2 presents model input and describes various kinds of forecasting models in detail. In Section 3, these forecasting models are compared and evaluated. Finally, Section 4 presents the main conclusion of the paper.

2. FORECASTING MODELS

2.1 SVM Model

The SVM is a new technique for regression. The basic concept of the SVM is to map nonlinearly the original data *x* into higher dimensional feature space. The SVM predictor is trained using a set of time series history values as inputs and a single output as the target value.[9]

Consider a given training set of *n* data points $\{x_i, y_i\}_{i=1}^{n}$ with input data $x_i \in R^n$, *p* is the total number of data patterns and output $y_i \in R$. SVM approximate the function in the following form :

$$y(x) = w^T \phi(x) + b \tag{2.1}$$

where $\phi(x)$ represent the higher dimensional feature space, which is nonlinearly mapped the input space *x*.

In SVM for function estimation, the estimation by minimizing regularized risk function:

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} L_\varepsilon(y_i) \tag{2.2}$$

is an arbitrary penalty parameter called the regularization constant.

Basically, SVM penalize $f(x_i)$ when it departures from y_i by means of an ε-insensitive loss function :

$$L_\varepsilon(y_i) = \begin{cases} 0 & \text{if} \left| f(x_i) - y_i \right| < \varepsilon \\ \left| f(x_i) - y_i \right| - \varepsilon & \text{otherwise} \end{cases} \tag{2.3}$$

The minimization of expression (2.2) is implemented by introducing the slack variable

ξ_i^- and ξ_i^+. Specifically, ε-Support Vector Regression (ε-SVR) solves the following

quadratic programming problem [1]:

$$\min_{\omega, b, \xi_i^-, \xi_i^+} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}\left(\xi_i^- + \xi_i^+\right) \tag{2.4}$$

Subject to;

$$y_i - (\omega'\phi(x_i) + b) \le \varepsilon + \xi_i^-$$
$$(\omega'\phi(x_i) + b) - y_i \le \varepsilon + \xi_i^+$$
$$\nabla i, \xi_i^- \text{ and } \xi_i^+ \ge 0$$

The solution to this minimization problem is of the form

$$f(x) = \sum_{i=1}^{m} (\lambda_i - \lambda_i^*) K(x_i, x) + b \qquad (2.5)$$

Where λ_i and λ_i^* are the Langrage multipliers associated with the constrains $y_i - (\omega'\phi(x_i) + b) \le \varepsilon + \xi_i^-$ and $(\omega'\phi(x_i) + b) - y_i \le \varepsilon + \xi_i^+$ respectively.

The kernel function can be defined as:

$$K(x_i, x_j) = \phi(x_i)'\phi(x_j) \qquad (2.6)$$

The value of the kernel is equal to the inner product of two vectors x_i and x_j in the feature space $\phi(x_i)$ and $\phi(x_j)$

Below are the Kernal types:

- Polynomial (homogeneous) : $k(x_i, x_j) = (x_i, x_j)^d$
- Polynomial (inhomogeneous) : $k(x_i, x_j) = (x_i \cdot x_j + 1)^d$
- Gaussian Radial Basis Function : $k(x_i, x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$, for $\gamma > 0$ (or $\gamma = 1/2\sigma^2$
- Hyperbolic Tangent : $k(x_i, x_j) = \tanh\left(\kappa \mathrm{x_i} \cdot x_j + C\right)$ for some $\kappa > 0$ and $C < 0$

The radial basis kernel is a popular choice in the SVM literature. Therefore our computations are based on such a kernel.[4,5]

2.2 ANN Model

The development of ANN models was based on [6] studying the relationship between input variables and output variables. Basically, the neural architecture consisted of three or more layers, i.e., input layer, output layer and hidden layer as shown in Fig. 1. The function of this network was described as follows:

$$\mathrm{Y_j} = \mathrm{f}\left(\sum \mathrm{w_{ij}} \mathrm{X_{ij}}\right) \qquad (2.7)$$

where Yj is the output of node j, f (.) is the transfer function, wij the connection weight between node j and node i in the lower layer and Xij is the input signal from the node i in the lower layer to node j.

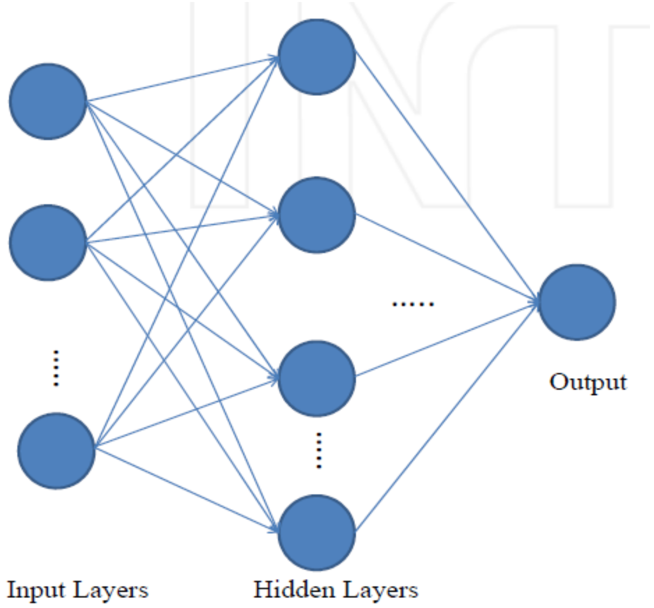

Fig 1: The architecture of a neural network

3. RESULT AND DISCUSSION

For analysis, the minimum RMSE from the SVM and ANN models are compared. Comparison performances of ANN and SVM models are given in Table 1. Figure 2 and Figure 3 the time series plot of actual and predicted testing data monthly Singapore Tourist Arrivals to Malaysia using SVM model and ANN model. The value of RMSE of ANN model is lower than those of SVM model. The results suggest that ANN model is superior to the ANN model in the modelling and Singapore Tourist Arrivals to Malaysia.

	SVM	ANN
RMSE	0.050	0.003

Table 1: The Comparison of RMSE between SVM Model and ANN Model of Singapore Tourist Arrivals to Malaysia.

As a result, SVM model is superior to the ANN model in the modelling and forecasting of Singapore Tourist Arrivals to Malaysia.

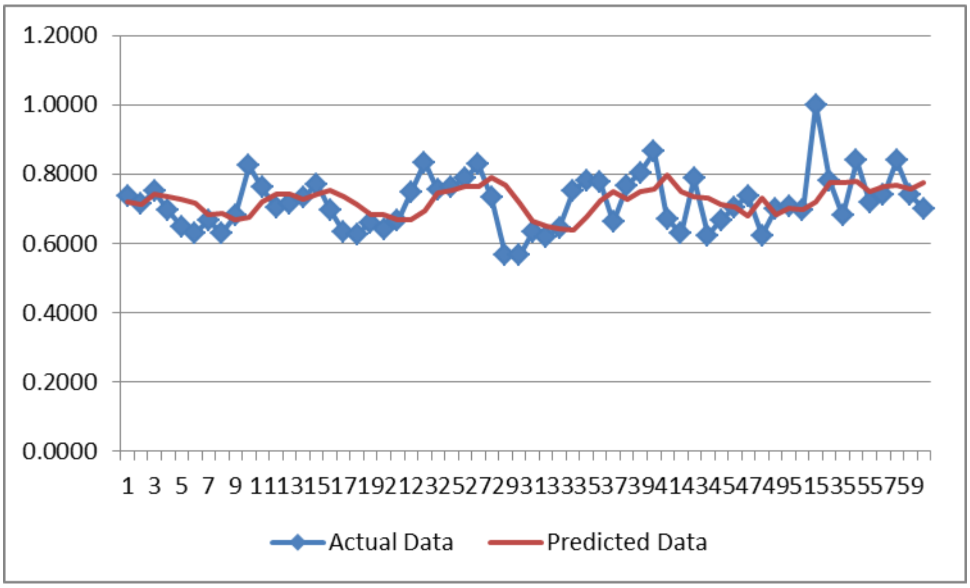

Figure 2: The Graph of Actual Testing Data and Predicted for Testing Data for Model p= 4,5 and 6 Singapore Tourist Arrivals to Malaysia Using SVM Model

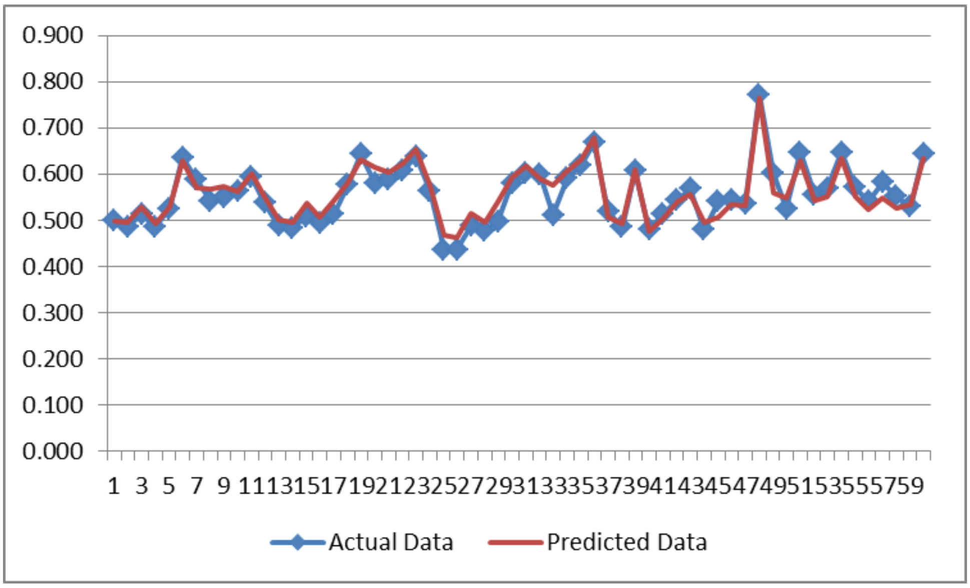

Figure 3: The Graph of Actual Testing Data and Predicted for Testing Data Singapore Tourist Arrivals to Malaysia Using ANN Model

4. CONCLUSION

From this paper, Artificial Neural Network (ANN) and Support Vector Machine (SVM) were used for monthly Singapore Tourist Arrivals to Malaysia. The modelling process data had been done by using computer software called MINITAB 14 and STATISTICA. The best results from SVM model were compared with the best fit of ANN model. From the analysis using both methods, it was concluded that the suitable model to predict monthly Singapore Tourist Arrivals to Malaysia is ANN model. It is because the ANN model has the smallest RMSE. This study has fulfilled and achieved its objective and scope which is to study ANN model and SVM model and also to compare the output from both of these methods.

REFERENCE

[1] O.Claveria, E.Monte and S.Torra, Tourism demand forecasting with different neural networks models (2013)

[2] S. Ismail, R. Samsudin and A. Shabri, A Comparison of Time Series Forecasting Using Support Vector Machine and Artificial Neural Network Model (2010)

[3] O.Claveria, E.Monte and S.Torra, Forecasting tourism demand to Catalonia: Neural networks vs. time series models (2014)

[4] J.Shawe-Taylor, N.Cristianini, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University (2000)

[5] K.Kandananond, A Comparison of Various Forecasting Methods for Autocorrelated Time Series (2012)

[6] A.Suliman, N.Nazri, M.Othman, M.A.Malek, and K.R.K.Mahamud, Artificial Neural Network And Support Vector Machine In Flood Forecasting: A Review (2013)

[7] P.F. Pai and C.S. Lin Using Support Vector Machine In Forecasting Production Values Of Machinery Industry In Taiwan. International Journal of Manufacturing Technology 27 (12) (2004) 205-210.

[8] A.Muhannad and M.M.Shotar. The Application Of Time Series Modelling To Some Major Economic Variables. Ph.D Qatar University.

[9] X.G.Hua, Y.Q.Ni ,J.M. Ko and K.Y. Wong, Modeling Of Temperature–Frequency Correlation Using Combined Principal Component Analysis And Support Vector Regression Technique (2007)

[10] A.Rafidah and Y.Suhaila, Modeling River Stream Flow Using Support Vector Machine Vol. 315 (2013) pp 602-605