Journal of Critical Reviews

ISSN- 2394-5125 Vol 7, Issue 8, 2020

FORECASTING INDONESIA TOURIST ARRIVALS TO MALAYSIA BASED ON NONLINEAR AND LINEAR MODEL

A. Rafidah[1*], Ani Shabri[2], Y.Suhaila[3], Erni Mazuin[4]

[*1,3,4] Technical Foundation, University Kuala Lumpur
[2]Mathematics, Universiti Teknologi Malaysia
Email: *[1]rafidahali@unikl.edu.my

Received: 25.03.2020 ***Revised: 23.04.2020*** ***Accepted: 01.06.2020***

Abstract
The development of economic and industry tourism depend upon how well the accuracy of number tourist arrivals forecasting is managed. The study aims to reduce computation complexity and enhance forecasting accuracy of decomposition ensemble model and wavelet method by incorporating intrinsic mode functions (IMFs) reconstruction. The empirical results indicated that the proposed model statistically outperformed all the considered benchmark models including the most popular wavelet with support vector machine (WSVM) model, decomposition ensemble model (Benchmark EMD-SARIMA and EMD_WSVM). To determine the performance, four statistical measures were applied, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Then, the best ranked model is measured using Mean of Forecasting Error (MFE) to determine its under and over-predicted forecast rate. The results show that EMD_WSVM ranked first based on four measures for Thailand tourist arrivals. The MFE results also indicates a small value of over-predicted values compared to the observed tourist arrivals values for Indonesia. The MAPE of the proposed EMD_WSVM data of Indonesia is <10% that indicate as excellent fit. In conclusion, the proposed method of pre-processing data using EMD and wavelet method enhanced the forecasting accuracy of the SVM model.

Index Terms-- Forecasting, tourist arrivals, SVM model, WSVM model and EMD_WSVM model.

DOI: http://dx.doi.org/10.31838/jcr.07.08.19

INTRODUCTION

Tourism has become one of the largest and fastest growing industries in the modern world (Werthner and Ricci, 2004). The global economic development is primarily linked to the tourism industry particularly in the number of tourist arrivals (Song and Li, 2008 and Song et al., 2011). Malaysia is one of the most popular tourist destinations. Strong exchange rate and political stability have made Malaysia an affordable destination in Southeast Asia. Since the first launching of "Visit Malaysia Year" in 1990, Malaysian government has launched several tourism promotions programs to attract foreign tourists especially from Asia, Middle East, and Europe. (Nanthakumar, Han and Kogid, 2013).

The tourism industry is an increasingly important industry for Malaysia; therefore, the policymakers and industry players have paid a close attention to the development of the tourism industry (Liang, 2014). Therefore, continuous tourism data, such as the tourist arrival or tourism demand data are necessary. With the help of the data, the pattern of the trend can be determined, thus the design and planning can be done accordingly. For instance, tourism research in various fields like promoting, marketing, forecasting and planning are indeed imperative and important. Because of this, the ability in forecasting tourist arrivals, it will aid the government and other tourism related organizations in their future planning.

However, most tourism time series data of practical relevance are complex nonlinear characteristics and chaotic in nature and strong seasonality of most tourism series. Because of their complicated nonlinearity and seasonality, currently existing methods cannot exactly deal with both issues. Accurate forecasting of tourist arrivals remains a difficult task attracted attentions in the literatures, and it is greatly necessary to develop new forecasting techniques to obtain satisfied accurate level. There is a need to enhance the prediction of future tourist arrivals.

To overcome the problem, several hybrid models have been developed to improve the accuracy of the forecasting value. One of the new forecasting hybrid models is wavelet with support vector machine (WSVM) model. However, wavelet-based hybrid models still have limitations in nonlinear and non- stationary time series forecasting because of the linearity restrictions of the wavelet transform. Thus, the accuracy of this hybrid model is still lacking in presenting the lowest error in forecasting result. Therefore, to overcome this drawback, this study will be concentrating on the modified the existing hybrid model WSVM combined with the empirical mode decomposition (EMD) to improve its prediction accuracy.

METHODOLOGY

A. Hybrid Wavelet Support Vector Machine (WSVM) Model

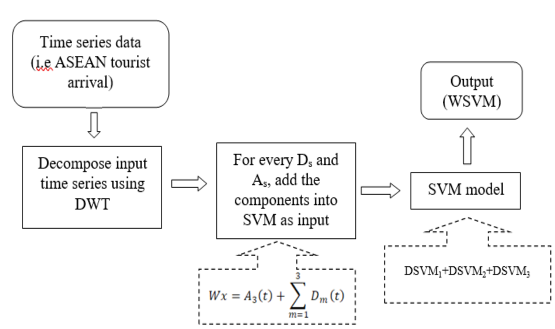

Figure 1. WSVM model structure

In the wavelet analysis, the original signal time series is broken into the low frequency which presented by a scale of approximate mode and the five resolution levels of high frequency. The signal is then represented by its features, which are becoming the wavelet coefficients. The sets of wavelet coefficients that generated from both high and low frequencies were sum up and tally for each input variables which were for seven ASEAN countries tourist arrivals. The total wavelet coefficients from every single input variable were divided into training and testing data. The input variables that carry through the wavelet coefficients then iterated into SVM model. Figure 3.4 shows the flowchart of implementing the algorithm.

B. Hybrid EMD-SARIMA Model

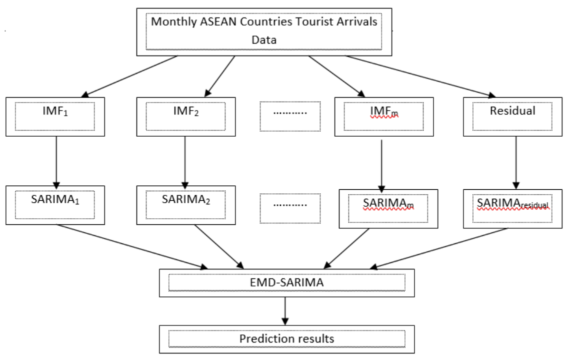

Figure 2. EMD-SARIMA model structure

The framework of hybrid consists of EMD-SARIMA model is demonstrated in Figure 2. The steps for achieving the final forecast are outlined below:

1. Original time series data is partitioned into a number of sub-series EMD methods.
2. Extract in the form of IMF and residue from the original data.
3. After the decomposition process, each of the obtained IMFs and residue components were analyzed using autocorrelation value in order to search for the stochastics and deterministic components.
4. Develop a suitable SARIMA model for each IMF and residue.
5. In the final step, the forecast value of each of the input models for all IMFs and residue components in the previous step will be reconstructed as a sum of all components and will be used as results accordingly.

C. Hybrid EMD_WSVM Model

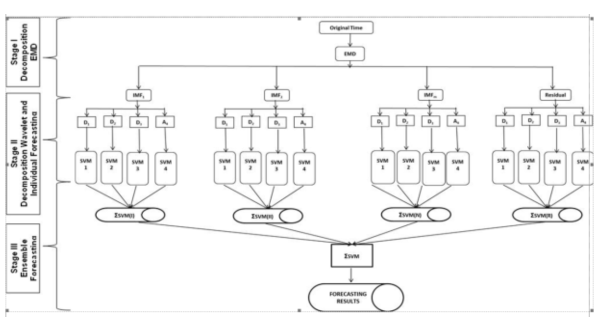

Figure 3. EMD-WSVM model structure

The following Figure 3 describes the above procedure, EMD_WSVM forecasting method that involves three stages:

(a) Stage I – Decomposition EMD. Monthly Indonesia tourist arrivals data time series were decomposed into n-IMF components and one residue, using EMD technique. The extracted IMF components represent a range of high to low frequencies and reveal various periodic patterns of tourist arrivals where each IMF component represents the local characteristic time scale by itself (Shabri, 2015a). Then, the IMF and residual are identified as stochastics or deterministic component based on is deterministic components otherwise stochastics (Aamir and Shabri, 2018).

(b) Stage II – Decomposition Wavelet. In stage two, the IMF stochastics and deterministic component decomposed into Discrete Wavelet Transformation (DWT). DWT was chosen to be used due to its simplicity and ability to compute results faster (Pandhiani and Shabri, 2015). In the wavelet analysis two the original signal time series is broken into the low frequency which presented by a scale of approximate mode and the two resolution levels of high frequency.

(c) Stage III – Forecasting Stage. In stage three, the SVM forecasting technique was employed on the chosen signals. The new dataset is further input into the SVM forecasting technique model accordingly and compared it with single model SVM using several performance measurements.

DATA

Monthly data of tourist arrivals for Indonesia countries consist of 204 observations which ranged from January 1999 till December 2015.

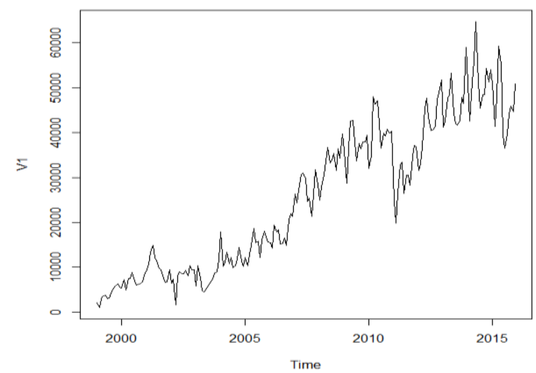

Figure 4. Monthly Indonesia tourist arrivals (Jan 1999 to Dec 2015

RESULTS AND DISCUSSION

In this experiment were carried out to evaluate the performance of Thailand tourist arrivals model forecasting using the root means square errors (RMSE), mean absolute errors (MAE), mean absolute percentage (MAPE) and mean forecast error (MFE).

The table 1 indicates that EMD_WSVM reported to have improvements over Benchmark EMD-SARIMA with about 77.67%, 74.25% and 3.84% reductions in MAE, RMSE and MAPE respectively and improvement for r. EMD_WSVM also produces significant improvements over WSVM where the reductions of MAE, RMSE and MAPE are by 36.56%, 27.06%, and 2.13% respectively.

Table 1. Comparative Performance of All Models

Method	RMSE	MAE	MAPE (%)
WSVM	40413.9741	25959.7653	10.96
EMD-SARIMA	99579.55701	84801.9700	12.67
EMD_WSVM	**25639.04**	**18935.06**	**8.83**

For further analysis MFE value for WSVM and EMD_WSVM had under predicted the forecasts by 10365.97 and 18541.85 respectively. Meanwhile, EMD-SARIMA has over predicted the number tourist arrivals value by -211563.6. Figure 5 compares the predicted and observed monthly tourist arrivals for Indonesia for EMD_WSVM models using 1:1 line to examine the agreement level between the values. Meanwhile, Figures 6 shows the comparison of the observed and the predicted values of monthly tourist arrivals for Indonesia. Most of the points in Figure 5 scattered close to the observation data indicating good prediction capability.

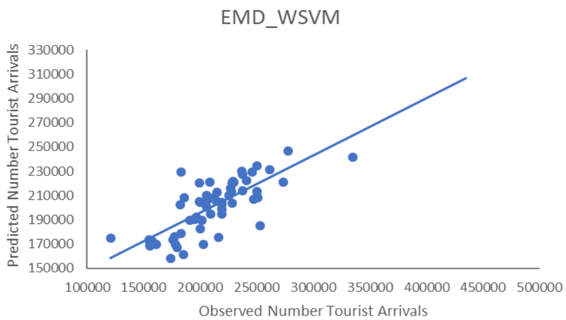

Figure 5. Plot 1:1 Graph

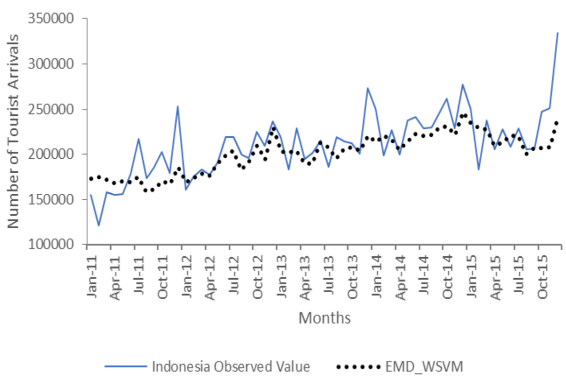

Figure 6. Time series Graph for EMD_WSVM

CONCLUSION

A hybrid forecasting framework of Empirical Mode ecomposition and wavelet with linear and nonlinear model has been study in this paper. The forecasting framework is tested using the data from the year 1999 to 2015 from the Ministry of Tourism, Malaysia and is compared with other widely used forecasting models (EMD-SARIMA and WSVM model) and the results show that the hybrid nonlinear model EMD–WSVM framework can improve the forecasting accuracy to a great level and that the forecasting result is consistent as well as for different forecasting steps. In the future, the robustness of the EMD–WSVM framework will be discussed by testing it with data where there is more seasonality (or less seasonality); also, the data sensitivity of the framework will also be discussed by testing it with more data (or less data).

ACKNOWLEDGMENT

This work has been supported by Research & Innovation, Universiti Kuala Lumpur, Malaysian Institute of Industrial Technology (MITEC).

REFERENCES

1. Werthner, H. and Ricci, F. (2004) 'E-commerce and tourism', Communications of the ACM, 47(12), pp. 101–105. doi: 10.1145/1035134.1035141.
2. Song, H. et al. (2011) 'Forecasting tourist arrivals using time-varying parameter structural time series models', International Journal of Forecasting, 27(3), pp. 855–869. doi: 10.1016/j.ijforecast.2010.06.001.
3. Song, H. and Li, G. (2008) 'Tourism Demand Modelling and Forecasting A Review of Recent Research', Tourism Management, 29(2), (2005), pp. 1–28.
4. Nanthakumar, L., Han, A. S. and Kogid, M. (2013) 'Demand for Indonesia, Singapore and Thailand Tourist to Malaysia: Seasonal Unit Root and Multivariate Analysis', International Journal of Economics and Empirical Research, 1(2), pp. 15–23.
5. Liang, Y. H. (2014) 'Forecasting models for Taiwanese tourism demand after allowance for Mainland China tourists visiting Taiwan', Computers and Industrial Engineering, 74(1), pp. 111–119. doi: 10.1016/j.cie.2014.04.005.
6. Shabri, A. (2015a) 'A novel hybrid ensemble learning paradigm for tourism forecasting', 192, pp. 192–200. doi: 10.1063/1.4907444.
7. Aamir, M. and Shabri, A. (2018) 'Improving Crude Oil Price Forecasting Accuracy via Decomposition and Ensemble Model by Reconstructing the Stochastic and Deterministic Influences', Advanced Science Letters, 24(6), pp. 4337–4342. doi: 10.1166/asl.2018.11601.
8. Pandhiani, S. M. and Shabri, A. Bin (2015) 'Time Series Forecasting by Using Hybrid Models for Monthly Streamflow Data', Applied Mathematical Sciences, 9(57), pp. 2809–2829. doi: 10.12988/ams.2015.52164.

AUTHORS

First & Correspondence Author: A. Rafidah, Lecturer: Technical Foundation Department, Universiti Kuala Lumpur, Bandar Seri Alam, johor Bahru Malaysia. rafidahali@unikl.edu.my.

Second Author: Ani Shabri – Assoc Prof, Mathematics Department, Universiti Technology Malaysia, Skudai Johor Bahru, Malaysia.