Global Journal of Engineering and Technology Review

Journal homepage: www.gjetr.org

Global J. Eng. Tec. Review 4 (1) 1 - 8 (2019)

Forecasting East Asian Tourist Arrivals to Thailand with Adaptive Neuro-Fuzzy Inference System

*Chawalsak Phetchanchai [1], Chuthawuth Chantaramalee [1], Napatsarun Chatchawalanont [1]

[1] Computer Science Program, Suan Dusit University, Nakhon Ratchasima road, Dusit, Bangkok 10300, Thailand

ABSTRACT

Objective – This research aims to propose the approach of forecasting tourist arrivals to Thailand.

Methodology/Technique – Adaptive Neuro-Fuzzy Inference System (ANFIS) was used as our forecasting method by using fuzzy C-means clustering as a technique for partitioning training dataset

Findings – The appropriate parameter of time lag was found for each dataset of East Asian tourist arrivals to Thailand.

Novelty – The forecasting procedure with the appropriate parameter of time lag was represented our work as a novelty idea.

Type of Paper: Empirical.

Keywords: Tourist arrivals forecasting, East Asian countries, adaptive neuro fuzzy inference system, fuzzy C-means clustering, Takagi–Sugeno fuzzy inference system.

1. Introduction

Tourism is the rapid growth industry in the world due to the increasing number of tourists traveling across the countries around the world despite of negative news like terror attacks and political unrest. World tourism organization (WTO) reported that in the year 2017, there were 1.32 billion tourists which increased 7 % from the year 2016 [1]. Further, WTO also forecasts that in the year 2030 there will be 1.8 billion tourists travelling around the world [2]. On the report "2018 Global Economic Impact & Issues" illustrates that travel and tourism impact is 10.4% of global GDP and 9.9% of total employment in 2017 [3].

*Paper Info: Revised: January 11, 2019

Accepted: February 19, 2019

* Corresponding author: Chawalsak Phetchanchai

E-mail:

Affiliation: Computer Science Program, Suan Dusit University, Nakhon Ratchasima road, Dusit, Bangkok 10300, Thailand

In Thailand, the arrivals of foreign tourists also dramatically increase. In the year 2017, there were 35.38 tourists, it increased 8.8% from the year 2016, and it was estimated to increase 6.1% or at 37.55 million tourists in the year 2018. Number of tourist arrivals directly relates to the facilities that the tourism business to prepare for supporting tourists for example, hotels, events, food services, transportation etc. This is also important for government sector to conduct country planning, policy issuing, and budget providing to support fundamental facilities. Thus, the correct forecasting of the future number of tourists is very important.

Tourist arrivals to Thailand comes from different regions around the world. However, the majority of them is tourists from East Asia countries such as China, Hong Kong, Japan , Korea, and Taiwan. By 2017 there were 14.45 million visitors from this region accounted for 40.85% of total tourist arrivals and with growth from 2016 as high as 11.74%. From this information, tourists from East Asia countries illustrate the very high potential visitors which can generate a huge income for the country.

Recently, forecasting tourist number relies on the time series model such as Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) [4], Singular Spectrum Analysis (SSA) [5, 6], and time-varying parameter based structural time series model (TVP-STSM) [7]. Fortunately, since the development of artificial intelligence techniques many works have successfully applied these techniques including adaptive neuro-fuzzy inference system(ANFIS) [8-10]. However, applying of ANFIS in forecasting of East Asian tourist arrivals to Thailand is not available. In this paper, the approach of applying ANFIS to forecasting East Asian tourist arrivals to Thailand is proposed.

2. Adaptive Neuro-Fuzzy Inference System

ANFIS is a hybridization of two models; neural networks and fuzzy inference system proposed by Jang [11]. ANFIS architecture is an adaptive network that uses a supervised learning algorithm for finding the optimal membership function parameters. ANFIS has a function similar to the model of Takagi–Sugeno fuzzy inference system [12]. The structure of ANFIS comprises of 5 layers of nodes. Number of nodes in each layer depends on the input dimensions and number of desired membership functions.

2.1 ANFIS Architecture

For simplicity, assume that there are two inputs x and y, and one output $\hat{S}$. Figure 1 shows the 5 layers ANFIS architecture.

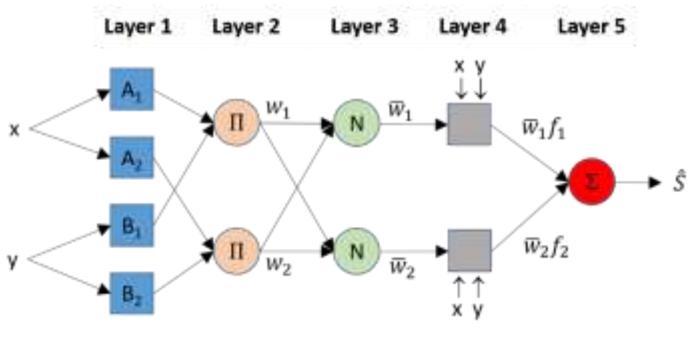

Figure 1. ANFIS Architecture

From the figure 1, two “If-Then” rules for Takagi–Sugeno model are available as follows.

Rule 1 = If x is A_1 and y is B_1 Then $f_1 = p_1x + q_1x + r_1$

Rule 2 = If x is A_2 and y is B_2 Then $f_2=p_2x+q_2x+r_2$

where A1 A2 and B1, B2 are the membership functions of each input x and y (part of the antecedents), while p1, q1, r1 and p2, q2, r2 are linear parameters in part "Then" (part of consequence) of Takagi-Sugeno fuzzy inference model.

Referring to figure 1, ANFIS architecture consists of five layers. The first and fourth layers contain adaptive nodes, while the other layers contain fixed nodes. A brief description of each layer is as follows.

Layer 1: all the nodes in this layer are adaptive nodes, which indicate that the shape of membership function can be modified through training. The output of node i in this layer can be shown as below:

$$O_i^1 = \mu_{A_i}(x) \tag{1}$$

where x is a representation of a crisp value for feeding to node i, and Ai illustrates the linguistic description for example; low, middle, and high. A symbol μ illustrates the membership function (MF). Various MFs are available such as Gaussian MF, Trapezoidal MF, Sigmoidal MF, etc. In this paper, the Gaussian MF is used and is shown as follow:

$$\mu_{A_i}(x) = e^{\frac{(x-c_i)^2}{2\sigma_i^2}} \tag{2}$$

where σ_i and c_i are the parameters of the Gaussian MF.

Layer 2: All nodes in this layer are illustrated as circle nodes which are labeled as $\prod$. Each node is performed as the firing strength of each rule and done by T-norm operators.

$$O_i^2 = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(x), \quad i = 1,2 \tag{3}$$

Layer 3: Nodes in the 3rd layer are also indicated as circle nodes which are labeled as N. The ith node is the ratio of the ith rules' firing strength to the sum of all rules' firing strengths. The outputs of the 3rd layer can be calculated as follow.

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1+w_2}, \ i = 1,2 \tag{4}$$

Layer 4: Every node in this layer is adaptive. Parameters used in this layer are represented as consequent parameters. The outputs of this layer can be shown as:

$$O_i^4 = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i y + r_i), \ i = 1,2 \tag{5}$$

where the parameters pi, qi, and ri are updated by LSE.

Layer 5: it has only one node in the last layer. The node computes the overall output calculated from all incoming information. The output of this layer is illustrated as follow.

$$O_i^5 = \hat{S} = \sum_{i=1}^{2} w_i f_i, \ \ i = 1,2 \tag{6}$$

2.2 Forecasting Performance Measurement

In order to forecast the future number of tourist arrivals, ANFIS is possibly to do this. The forecasted values are needed to compare to the actual values for evaluating the performance of forecasting.

For both training and testing processes, the error between actual values and an output or forecasted values is considered to minimize. Calculation of the error (e) for the ith iteration of training phase or the ith sample of testing phase can be shown as below:

$$e_i = S_i - \hat{S}_i \tag{7}$$

where e_i illustrates the error for the ith iteration in training phase, or the ith sample in the testing phase, and S_i and $\hat{S}_i$ represent the actual output and the forecasted output respectively. However, the error is usually used in the form of error function or cost function of the training and testing processes. Several error functions are available. Followings illustrate error functions used in this research.

2.1.1 Mean squared error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \qquad (8)$$

2.1.2 Root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \qquad (9)$$

2.1.3 Mean absolute error (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |e_i| \qquad (10)$$

3. The proposed Forecasting Approach

This section, the application of using ANFIS in forecasting the future number of tourist arrivals to Thailand from five East Asia countries; China, Hong Kong, Japan, Korea and Taiwan is conducted. Following figure shows the steps of applying ANFIS for forecasting tourist arrivals from five East Asia countries.

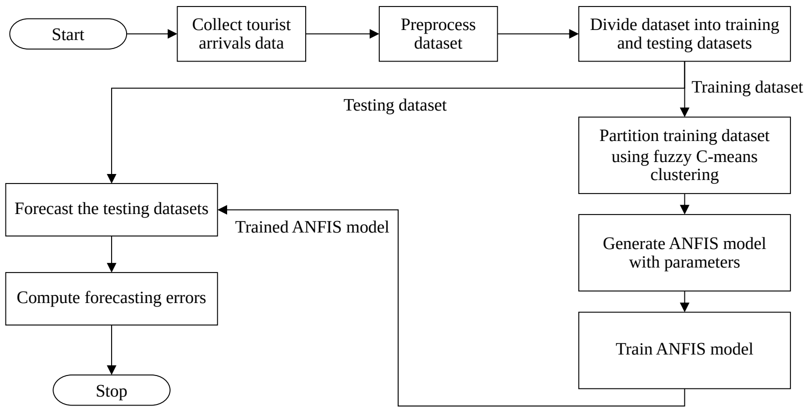

Figure 2. Procedure of forecasting approach.

3.1 Collect tourist arrivals data

In this step, tourist arrivals to Thailand from January 2008 to December 2017 are collected from the website of Ministry of Tourism [13]. The collected data comprise of number of monthly travelers from five East Asia countries; China, Hong Kong, Japan, Korea, and Taiwan totally 120 months. Following shows plots of these number of tourists.

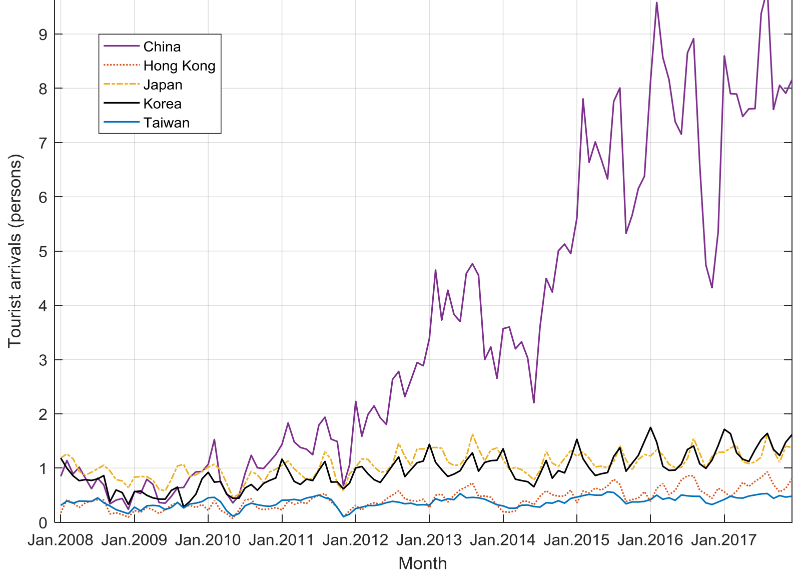

Figure 3. Monthly tourists number from China, Hong Kong, Japan, Korea, and Taiwan.

3.2 Preprocess dataset

Collected data for each country are required to preprocess before using in the forecasting model as followings.

•Data transformation. This step, the real number of tourist arrivals are transformed into the range of [0,1] by using min-max normalization. Following is the formula of min-max normalization.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} (x'_{max} - x'_{min}) + x'_{min} \tag{11}$$

where

- x and x' are the original data and the transformed data respectively
- x_{min} and x_{max} are the original maximum and minimum values respectively
- x'_{min} and x'_{max} are the normalized minimum and maximum values respectively (equal to 0 and 1 respectively for this research)

• In the forecasting concept, we aim to forecast the next month value based on the values of previous q months lag. Thus, the next month value is a function of previous q-month values.

$$x_{t+1} = f(x_{t-q}, \dots, x_{t-1}, x_t) \tag{12}$$

3.3 Dividing dataset

This step, data are aimed to divide into two sets; training set, and testing set. In this research we use the ratio of training set and testing set as 8:2.

3.4 Partitioning training dataset

We use fuzzy C-means clustering technique to partition the universe of discourse of input variables or the training dataset in to groups with their fuzzy similarity.

3.5 Generating ANFIS model

This step, number of groups specified in previous step corresponds to number of rules used in the ANFIS model. In addition, type and number of membership function are also required to specify. In this research, five Gaussian MFs are used.

3.6 Training ANFIS model

This step, the training dataset are used for training ANFIS model. We use the hybrid learning algorithm [11] which is a combination of least-squares and back-propagation gradient descent methods, to optimize the membership function parameters.

3.7 Forecasting testing dataset

The trained ANFIS model was used to conduct forecasting the corresponding testing dataset.

3.8 Computing errors

The forecasted outputs are required to calculate errors comparing to the actual transformed actual values by using MSE, RMSE, and MAE error functions.

4. Experiments and results

In this section, we aim to experiments the proposed forecasting approach to the collected datasets of China, Hong Kong, Japan, Korea, and Taiwan. Each dataset required to transform into the range [0,1] and reform to q previous months lag where q equals to 2,3,4,5, and 6 months. Further, each training set are divided into group using fuzzy C-means clustering method which number of cluster varies from 2 to 9 clusters.

The results of our experiments illustrated in the table below. The results show only the best results by determining the MSE value for each dataset in the form of error values.

Table 1. Best results of different datasets with their parameters.

Country	Lag	Number of MFs	Errors with training dataset			Errors with testing dataset		
			MSE	RMSE	MAE	MSE	RMSE	MAE
China	6	3	0.0039	0.0624	0.0384	0.0271	0.1647	0.1223
Hong Kong	3	3	0.0140	0.1183	0.0901	0.0536	0.2315	0.0903
Japan	4	7	0.0018	0.0428	0.0259	0.0134	0.1156	0.0914
Korea	6	2	0.0137	0.1037	0.0795	0.0107	0.1998	0.1762
Taiwan	2	2	0.0233	0.1149	0.0879	0.0132	0.1526	0.1220

From the table, it is clear that the errors for both training and testing of all datasets are different. Data of China and Korea give the best forecasting in testing sets with the parameters of time lag 6 and number of MFs are 3 and 2 MFs respectively. For data of Hong Kong, Japan, and Taiwan use different time lag of datasets as 3, 4, and 2 and with best number of MFs as 3, 2, and 2 MFs respectively. The first two best forecasting performance results based on MSE values found in testing dataset of Korea and Japan respectively, and the worst forecasting performance result can be found in testing dataset of Hong Kong.

The performance of forecasting in both training and testing datasets can be demonstrated as shown in figure below.

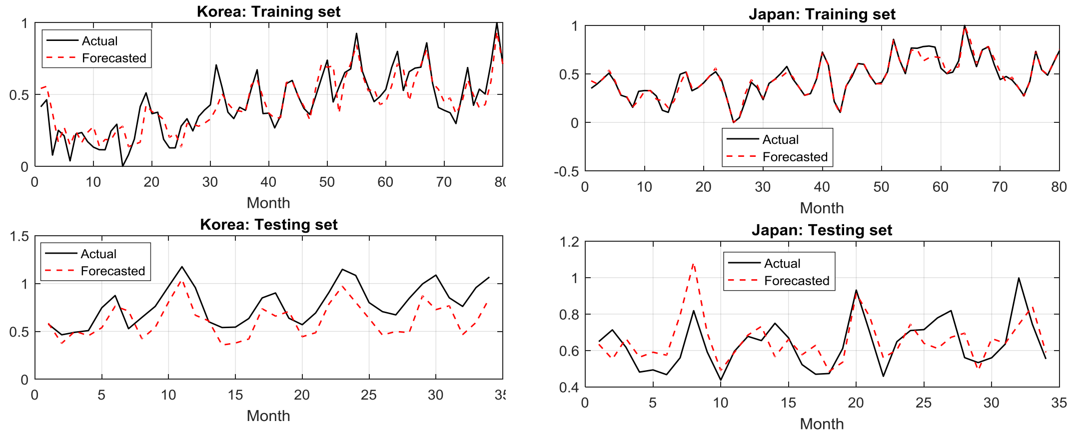

Figure 4. Forecasting performance plots of Korea and Japan in both training and testing sets.

5. Conclusion

In this study, adaptive neuro-fuzzy inference system has been used for forecasting number of tourist arrivals in future month based on difference time lags, number of membership function, with different datasets of tourist visiting Thailand from East Asia countries such as China, Hong Kong, Japan, Korea, and Taiwan. Training datasets are partitioned into clusters by using fuzzy C-means clustering method with different number of clusters varied from 2 to 9 clusters. The results show that the best performance is data of Korea, the second best performance is data of Hong Kong. Finally, the worst forecasting performance dataset is data of Japan.

Acknowledgements

This research is supported by National Research Council of Thailand and Suan Dusit University, Thailand.

References

Journal article

C. Lim, and M. McAleer, "Time series forecasts of international travel demand for Australia," Tourism Management, vol. 23, no. 4, pp. 389-396, 2002/08/01/, 2002.
H. Hassani et al., "Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis," Tourism Management, vol. 46, pp. 322-335, 2015/02/01/, 2015.
H. Hassani et al., "Forecasting accuracy evaluation of tourist arrivals," Annals of Tourism Research, vol. 63, pp. 112-127, 2017/03/01/, 2017.
H. Song et al., "Forecasting tourist arrivals using time-varying parameter structural time series models," International Journal of Forecasting, vol. 27, no. 3, pp. 855-869, 2011/07/01/, 2011.
M.-S. Chen, L.-C. Ying, and M.-C. Pan, "Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system," Expert Systems with Applications, vol. 37, no. 2, pp. 1185-1191, 2010.
R. Law, and N. Au, "A neural network model to forecast Japanese demand for travel to Hong Kong," Tourism Management, vol. 20, no. 1, pp. 89-97, 1999.
R. Law, and N. Au, "A neural network model to forecast Japanese demand for travel to Hong Kong," Tourism Management, vol. 20, no. 1, pp. 89-97, 1999.
R. Law, "Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting," Tourism Management, vol. 21, no. 4, pp. 331-340, 2000.

J.-S. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” IEEE transactions on systems, man, and cybernetics, vol. 23, no. 3, pp. 665-685, 1993.
T. Takagi, and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” IEEE transactions on systems, man, and cybernetics, no. 1, pp. 116-132, 1985.

Monographs:

G. G. Manzo, 2018 Global Economic Impact & Issues, World Travel & Tourism Council, 2018.
UNWTO, “UNWTO World Tourism Barometer and Statistical Annex, Advance Release January 2018,” UNWTO World Tourism Barometer (English version), vol. 16, no. 1, pp. 1-56, 2018/01/31, 2018.
World Tourism Organization, UNWTO Tourism Highlights: 2017 Edition, p.^pp. 16: UNWTO, 2017.
Ministry of Tourism & Sports. "Tourism statistics," 11/20/2018, 2018; https://www.mots.go.th/more_news.php?cid=411.
[1] UNWTO, “UNWTO World Tourism Barometer and Statistical Annex, Advance Release January 2018,” UNWTO World Tourism Barometer (English version), vol. 16, no. 1, pp. 1-56, 2018/01/31, 2018.
[2] World Tourism Organization, UNWTO Tourism Highlights: 2017 Edition, p.^pp. 16: UNWTO, 2017.
[3] G. G. Manzo, 2018 Global Economic Impact & Issues, World Travel & Tourism Council, 2018.
[4] C. Lim, and M. McAleer, “Time series forecasts of international travel demand for Australia,” Tourism Management, vol. 23, no. 4, pp. 389-396, 2002/08/01/, 2002.
[5] H. Hassani et al., “Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis,” Tourism Management, vol. 46, pp. 322-335, 2015/02/01/, 2015.
[6] H. Hassani et al., “Forecasting accuracy evaluation of tourist arrivals,” Annals of Tourism Research, vol. 63, pp. 112-127, 2017/03/01/, 2017.
[7] H. Song et al., “Forecasting tourist arrivals using time-varying parameter structural time series models,” International Journal of Forecasting, vol. 27, no. 3, pp. 855-869, 2011/07/01/, 2011.
[8] M.-S. Chen, L.-C. Ying, and M.-C. Pan, “Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system,” Expert Systems with Applications, vol. 37, no. 2, pp. 1185-1191, 2010.
[9] R. Law, and N. Au, “A neural network model to forecast Japanese demand for travel to Hong Kong,” Tourism Management, vol. 20, no. 1, pp. 89-97, 1999.
[10] R. Law, “Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting,” Tourism Management, vol. 21, no. 4, pp. 331-340, 2000.
[11] J.-S. Jang, “ANFIS: adaptive-network-based fuzzy inference system,” IEEE transactions on systems, man, and cybernetics, vol. 23, no. 3, pp. 665-685, 1993.
[12] T. Takagi, and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” IEEE transactions on systems, man, and cybernetics, no. 1, pp. 116-132, 1985.
[13] Ministry of Tourism & Sports. "Tourism statistics," 11/20/2018, 2018; https://www.mots.go.th/more_news.php?cid=411.