International Journal of Engineering & Technology, 7 (2.2) (2018) 104-107

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Modeling of time series data for forecasting the number of foreign tourists in east Kalimantan using fuzzy inference system based on ARX model

Said Keliwar[1]*, Arief Bramanto Wicaksono Putra[2]*, Jehad Hammad[3], Haviluddin[4]

[1]*Department of Tourism, State Polytechnic of Samarinda, East Kalimantan, Indonesia*
[2]*Department of Information Technology, State Polytechnic of Samarinda, East Kalimantan, Indonesia*
[3]*Alquds Open University*
[4]*Faculty of Computer Science and Information Technology, Mulawarman University, Samarinda, East Kalimantan, Indonesia*
**Corresponding author E-mail: ariefbram@gmail.com, saidkeliwar@polnes.ac.id*

Abstract

The government agencies require accurate tourism demand forecasts to plan the required tourism infrastructure, such as accommodation location planning and transportation development. Tourism demand forecasts can be viewed from various factors, one of which is the number of tourists per period. Without ignoring the number of domestic tourists, the increasing number of foreign tourists is prioritized by the government to increase the country's foreign exchange. Usually, a tourist destination is to visit some tour objects and need some accommodation and hotel sites to rest. With this consideration, the forecasting number of foreign tourists can be done by using data on the number of tour objects, accommodation and hotel sites, foreign and domestic tourists from the previous period. Data on the number of domestic tourists used to measure the tendency of foreign tourists compared with domestic tourists to all existing tour objects. All data history can be viewed as time series data. Conventionally, many researchers have employed traditional methods of time series analysis, modeling, and forecasting such as ARX (Autoregressive with exogenous input) model. FIS (Fuzzy Inference System) is a system that processes the input mapping formulation provided to produce output using Fuzzy Logic. The aim of this study is to forecast the number of foreign tourists by using FIS through a training process conducted by adapting the number of linguistic variables. All the training data are modeled by using the ARX model.

Keywords: *foreign tourist, domestic tourist, ARX model, FIS.*

1. Introduction

The tourism industry has become one of the most important sectors in many countries in the world. This industry has contributed a major economic resource to these countries. There are two types of tourists are domestic and foreign tourists. Domestic tourists are every visitor who visiting an area within the country of residence. Foreign tourists are every visitor who visiting an area outside the country of residence that lives for at least 24 hours and not exceeding one year [1].

Government agencies require accurate tourism demand forecasts to plan the required tourism infrastructure, such as accommodation location planning and transportation development. Forecasting is very important in many types of organizations because the prediction of future events is one of the important factors in the decision-making process [2]. Forecasting performance evaluation of artificial neural network modeling compared with the time series method, the estimated arrival of tourists is more than the estimated overnight stay. Seasonality and volatility are important features of tourism data, therefore the context is a favorable feature for comparing the performance of linear model forecasting with non-linear alternative approach [3], [4].

A research has been conducted to forecast demand travel models involving stages of trip generation and distribution with different types of traffic density models[5]. Other studies have compared between modern and classical methods, in which classical methods are represented by Box Jenkins, ARIMA, SARIMA, Holt-Winters and time series regression methods [6]–[8]. While the modern method is represented by Fuzzy time series [9].

Generally, in the time series area, we can divide the forecasting method into a classical or traditional method and modern methods. Several models of fuzzy time series have been developed to construct fuzzy time series models for long-term predictions. The objective study of time series fuzzy models is how to improve forecasting accuracy by controlling uncertainty and involving fuzzy number support [10]–[12].

Conventionally, researchers use traditional analysis, modeling and forecasting methods such as AR (Autoregressive with exogenous input) model [13]. Approaches Autoregressive with exogenous inputs (ARX) and artificial neural network (ANN) models for the detection and imputation of anomalies in time series data are used to extract the characteristics of time series [14], In addition there is a proposed a two-stage weighted-least-squares regression approach, in which the prediction method includes a combination of two separate time-indexed ARX models to improve the prediction accuracy of the cooling load over different forecasting periods [15], [15].

However, they only provide reasonable accuracy and still have problems with stationary and linearity assumptions. Because of these constraints, came the idea of an alternative solution that is Fuzzy Inference System that processes the input mapping formulation provided to produce output using Fuzzy Logic. The aim of this study is to forecast the number of foreign tourists by using

FIS through a training process conducted by adapting the number of linguistic variables. All the training data are modeled by using the ARX model.

2. Experimental Details

2.1. The ARX Model

The general structure of the ARX model can be written as follows:

$$y(t)+a_1y(t-1)+...+a_ny(t-n)= b_1u(t-1)+...+b_mu(t-m)+e(t) \quad (1)$$

Equation (1) can be described as follows [4]:

$$\begin{aligned} y(t) &= \left(a_1q^{-1}+...+a_nq^{-n}\right)y(t) \\ &\quad +\left(b_1q^{-1}+...+b_mq^{-m}\right)u(t)+e(t) \\ &= \left(\sum_{i=1}^{n} a_iq^{-i}\right)y(t)+\left(\sum_{j=1}^{m} b_jq^{-j}\right)u(t)+e(t) \end{aligned} \quad (2)$$

The variables $u(t)$ and $y(t)$ are the inputs and outputs of each system, $e(t)$ is the white noise signal, t is the time unit and q^{-1} represents the delay operator. The variables a_i and b_j are model parameters to be estimated.

2.2. Fuzzy Logic

Fuzzy logic variables have truth values that range between 0 and 1. Unlike traditional binary sets that have true or false values. A fuzzy logic is represented as a form or logical system based on the set theory of fuzzy set. The usefulness of the fuzzy set lies in its ability to model uncertain or ambiguous data. The fuzzy set has been defined as a set of fuzzy boundaries whose elements have a degree of membership. It has an infinite range of truth values which can be anywhere in the range from 0 to 1, normally it is called the degree of membership [16] [16]. The ambiguity in the fuzzy set is expressed by the Membership Function [17]. One of the various Membership Functions (MF) is Triangular MF as shown in Figure 1. An example of a Fuzzy set with three linguistic variables (Low, Medium, High) using triangular MF as shown in Figure 2.

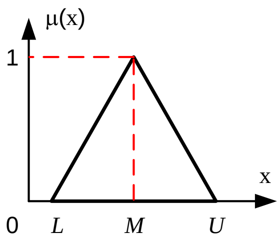


Fig. 1: Triangular MF.

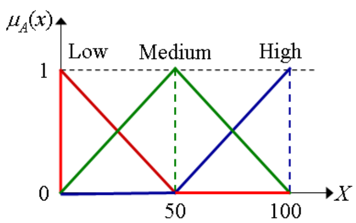


Fig. 2: Fuzzy Set (Triangular MF)

.

Mathematically, Triangular MF is expressed by:

$$\mu(x)=trimf(x,L,M,U)=\begin{cases} 0 & x\le L, x\ge U \\ \dfrac{x-L}{M-L} & L\le x\le M \\ \dfrac{U-x}{U-M} & M\le x\le U \end{cases} \quad (3)$$

FIS provides a means of mapping the available input space to the expected output space. Mapping is done on the basis of IF-THEN rules that utilize the "AND" or "OR" connectors to make the necessary decisions. The input to the FIS may be crisp or fuzzy, but the result is always the fuzzy set of FIS. The two most important processes in the FIS are fuzzification and defuzzification. FIS as shown in Figure 3 consist of four functional blocks, they are:

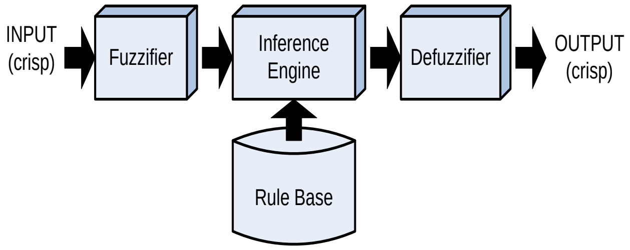


Fig. 3: Fuzzy Inference System.

1. Fuzzifier, convert the crisp value of inputs into fuzzy value through a fuzzy set which defined using certain membership function for each input. This process called fuzzification.
2. Rule Base contains several *IF-THEN* rules.
3. Inference engine applies every *IF-THEN* rule by using *AND* (min) operator for antecedent part to consequent part and applies the implication method using *AND* (min) operator to the consequent part. The result of the implication method is a fuzzy set. All the results of applying the implication method of all the rules are then aggregated using *OR* (max) operator, which the result is a fuzzy set as well.
4. Defuzzifier, convert the fuzzy set as the result aggregation method into a crisp value.

Mamdani's fuzzy inference system is the most commonly used fuzzy method. The Mamdani method uses the aggregation process of all the implication processes of each rule in the form of a fuzzy set to produce a single output. This process is called defuzzification. One of the most popular defuzzification methods is the centroid method that returns the center of the area under the curve. For Fuzzy Set X with a finite universe of discourse, the centroid method is expressed by [6]:

$$x_c=\left(\sum_{i=1}^{n} x_i.\mu(x_i)\right)\Big/\sum_{i=1}^{n}\mu(x_i) \quad (4)$$

Where x_c is the centroid value, x_i is x value of ith point corresponds to $\mu(x_i)$ from the aggregation result, $\mu(x_i)$ is the degree of membership (fuzzy number) of x_i, n is the number of $(x_i,\mu(x_i))$ pairs in the curve.

2.3. Time Series Data Modelling

Usually, a tourist destination is to visit some tour objects and need some accommodation and hotel sites to rest. With these considerations, the data used in this study is the number of tour objects, accommodation and hotel sites, foreign and domestic tourists as shown in Table 1. Foreign tourist ratio expressed by:

$$FT_{ratio} = \frac{(no.\ of\ \ foreign\ tourist)}{(no.\ of\ \ domestic\ tourist)} \tag{5}$$

In this study, 1st order ARX model is used to model the time series data which expressed by:

$$y_1(t)+y_2(t)=-a_1.y_1(t-1)-a_2.y_2(t-1)$$
$$+b_1.u_1(t-1)+b_2.u_2(t-1)+e_1(t)+e_2(t) \tag{6}$$

Where, $y_1(t)$ and $y_1(t\text{-}1)$ are the numbers of domestic tourist in current and previous year, respectively, $y_2(t)$ and $y_2(t\text{-}1)$ are foreign tourist ratio in current and previous year, respectively, $u_1(t\text{-}1)$ and $u_2(t\text{-}1)$ are the number of tour object and accommodation & hotel sites in previous year, respectively. For example, from the year 2001-2003 are obtained in Table 2.

Table 1: The number of tour object, accommodation & hotel site, foreign and domestic tourist

Year	2001	2002	2003	2004	2005	2006	2007	2008
Tour Objects	180	180	325	316	316	393	393	393
Accom. & Hotels	313	314	314	311	321	352	353	384
Foreign Tourist	12.599	14.094	11.781	16.930	17.299	19.169	19.746	20.142
Domestic Tourist	535.782	560.283	841.781	765.122	776.598	782.037	793.000	808.860
Foreign Tourist Ratio	0,0235	0,0252	0,0140	0,0221	0,0223	0,0245	0,0249	0,0249
Year	2009	2010	2011	2012	2013	2014	2015	2016
Tour Objects	393	413	423	430	435	435	516	527
Accom. & Hotels	392	393	420	432	432	491	491	591
Foreign Tourist	23.768	24.410	29.768	28.273	32.973	53.257	59.285	70.976
Domestic Tourist	1.131.906	1.174.626	1.564.013	1.667.467	1.926.769	3.914.769	4.270.740	5.030.586
Foreign Tourist Ratio	0,0210	0,0208	0,0190	0,0170	0,0171	0,0136	0,0139	0,0141

Table 2: Example of time series data using 1st order ARX model from Table 1

Year (t-1)	Input				Year (t)	Output	
	$u_1(t\text{-}1)$	$u_2(t\text{-}1)$	$y_1(t\text{-}1)$	$y_2(t\text{-}1)$		$y_1(t)$	$y_2(t)$
2001	180	313	535.782	0,0235	2002	560.283	0,0252
2002	180	314	560.283	0,0252	2003	841.781	0,0140
2003	325	314	841.781	0,0140	2004	765.122	0,0221

Each variable has the universe of discourse (U) in the range [min, max] as shown in Table 3. If all inputs and outputs have the same fuzzy set with 3 linguistic variables, then every universe of discourse of input and output is partitioned into 3 partitions of equal length intervals as shown in Table 4. Each data is then labeled in order of partition as shown in Table 5.

Table 3: The universe of discourse of the time series data from Table 1

	Tour Objects ($\boldsymbol{u}_1$)	Accom. & Hotels ($\boldsymbol{u}_2$)	Domestic Tourist ($\boldsymbol{y}_1$)	Foreign Tourist Ratio ($\boldsymbol{y}_2$)
Min	180	311	535.782	0.01360
Max	527	591	5.030.586	0.02516

Table 4: Partition result from Table 3

No. Partition	u_1		u_2		y_1		y_2	
1	180	296	311	404	535.782	2.034.050	0,0136	0,0175
2	296	411	404	498	2.034.050	3.532.318	0,0175	0,0213
3	411	527	498	591	3.532.318	5.030.586	0,0213	0,0252

Table 5: Labelling result from Table 2

Year (t-1)	Input				Year (t)	Output	
	A	B	C	D		E	F
2001	1	1	1	3	2002	1	3
2002	1	1	1	3	2003	1	1
2003	2	1	1	1	2004	1	3

From Table 5 then obtained the rules as follows:

No	antecedent	consequent
1	*IF **A** is **Low** AND **B** is **Low** AND **C** is **Low** AND **D** is **High***	*THEN **E** is **Low**, **F** is **High***
2	*IF **A** is **Low** AND **B** is **Low** AND **C** is **Low** AND **D** is **High***	*THEN **E** is **Low**, **F** is **Low***
3	*IF **A** is **Medium** AND **B** is **Low** AND **C** is **Low** AND **D** is **High***	*THEN **E** is **Low**, **F** is **High***

Using FIS, the 1st order ARX model of time series data as above can be declared as:

$$y_1(t)+y_2(t)=Fuzzy\begin{pmatrix} y_1(t-1), y_2(t-1), \\ u_1(t-1), u_2(t-1) \end{pmatrix}+e_1(t)+e_2(t) \tag{7}$$

By training $Fuzzy(\bullet)$ such that $e_1(t)+e_2(t)\to 0$ then obtained $Fuzzy(\bullet)\to y_1(t)+y_2(t)$. Training data is generated from Table 1 after modeling using 1st order AR model. The rule base is built on training data using data labeling technique as above. The performance function of the training results using MAPE (Mean Absolute Percentage Error) expressed as:

$$MAPE=\frac{1}{N}\sum_{i=1}^{N}\frac{\left|y(t_i)-y_{fuzzy}(t_i)\right|}{y(t_i)} \tag{8}$$

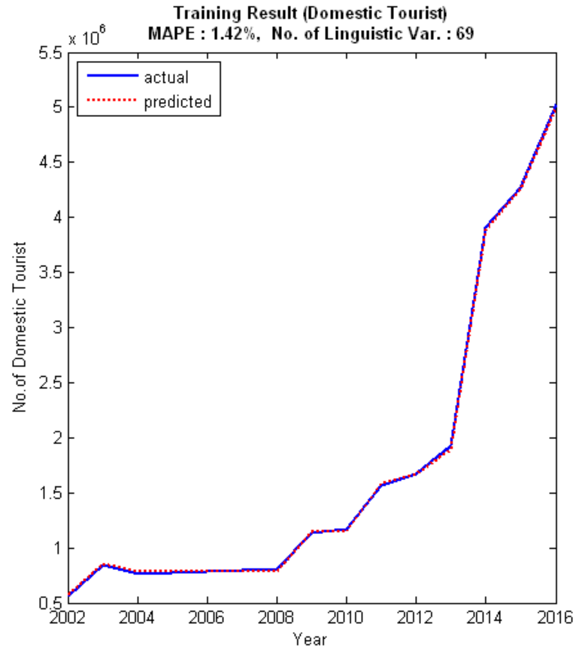
where N is the number of training data; $y(t_i)$ is the ith training target; $y_{fuzzy}(t_i)$ is the ith output of FIS. Implementation of FIS is done by using MATLAB programming tool. In this study, the FIS uses 5 linguistic variables for initialization and MAPE target ($e(t)$) is 1%.

3. Result and Discussion

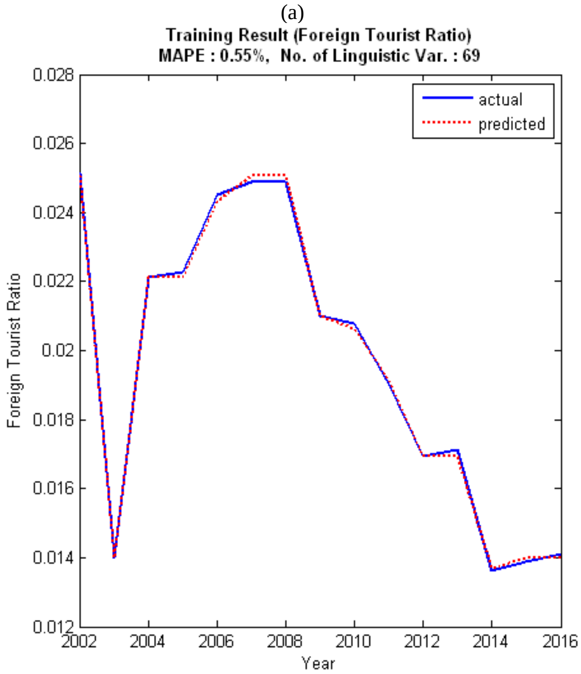
The training result is shown in Figure 4. Final MAPE is 1.42% for y_1 (no. of domestic tourist) and 0.55% for y_2 (foreign tourist ratio) with the number of linguistic variables of 69. The predicted no. of foreign tourist calculated by using the formula:

$$no.\ of\ \ foreign\ tourist=(FT_{ratio})*(no.\ of\ \ domestic\ tourist)$$

This predicted data are then used to validate the model already trained. The result of trained model validation is shown in Figure 5 with MAPE of 1.46%.



(a)



(b)

Fig. 4: Training result

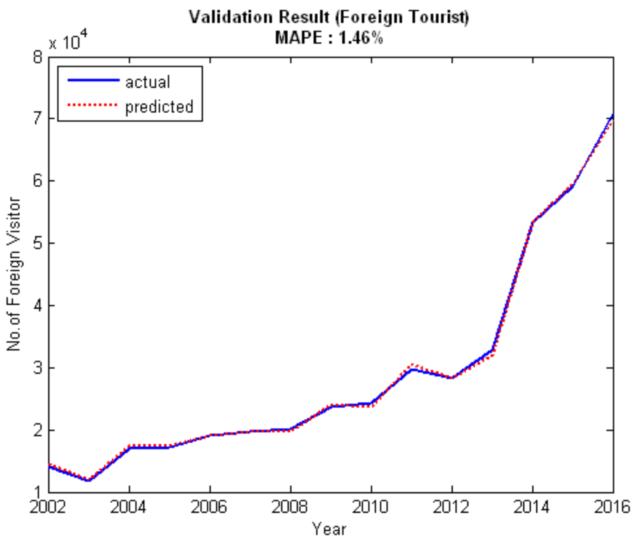


Fig. 5: Validation result

4. Conclusion

The results of this study prove that with certain MAPE, FIS model with input in the form of time series data based on ARX model can predict more than one output by adapting the number of linguistic variables during the training process. In this study did not consider the existence of the same rules based on training data. Future work, the development of this proposed method can be done taking into account the ignorance of the same rules.

References

[1] B. Statistics, *East Kalimantan Province in Figures 2006-2017*. 2017.

[2] Y. Wang, "The Tourism Demand of Nonlinear Combination Forecasting based on Time Series Method and WNN," *Int. J. u-e-Service, Sci. Technol.*, vol. 8, no. 3, pp. 301–310, 2015.

[3] O. Claveria and S. Torra, "Forecasting tourism demand to Catalonia: Neural networks vs. time series models," *Econ. Model.*, vol. 36, no. September 2013, pp. 220–228, 2014.

[4] M. Gan, Y. Cheng, K. Liu, and G. L. Zhang, "Seasonal and trend time series forecasting based on a quasi-linear autoregressive model," *Appl. Soft Comput. J.*, vol. 24, no. November, pp. 13–18, 2014.

[5] B. Prasad and K. Molugaram, "Development of mode choice models of a trip maker for Hyderabad metropolitan city," *Int. J. Eng. Technol.*, vol. 7, pp. 1–7, 2018.

[6] A. Rahman and A. S. Ahmar, "Forecasting of primary energy consumption data in the United States: A comparison between ARIMA and Holter-Winters models," in *AIP Conference Proceedings*, 2017, vol. 1885.

[7] A. S. Ahmar *et al.*, "Modeling Data Containing Outliers using ARIMA Additive Outlier (ARIMA-AO)," *J. Phys. Conf. Ser.*, vol. 954, 2018.

[8] A. S. Ahmar, "A Comparison of α-Sutte Indicator and ARIMA Methods in Renewable Energy Forecasting in Indonesia," *Int. J. Eng. Technol.*, vol. 7, no. 1.6, pp. 9–11, 2018.

[9] M. H. Lee, M. E. Nor, H. J. Sadaei, N. H. A. Rahman, and N. A. B. Kamisan, "Fuzzy Time Series: An Application to Tourism Demand ForecastingNo Title," *Am. J. Appl. Sci.*, vol. 9, no. 1, pp. 132–140, 2012.

[10] A. Tale, A. S. Gusain, J. Baguli, R. Sheikh, and A. Badar, "Study of Load Forecasting Techniques using," *Int. J. Advenced Res. Electr. Electron. Instrum. Eng.*, vol. 6, no. 2, pp. 510–518, 2017.

[11] A. Cankurt, Selcuk Subasi, "Comparison of Linear Regression and Neural Network Models Forecasting Tourist Arrivals to Turkey," *Eurasian J. Sci. Eng.*, vol. 1, no. 1, pp. 21–25, 2015.

[12] A. E. Abow Mohammed, "Using analysis of time series to forecast the number of patients with tuberculosis: a case study in Khartoum state from 2007 to 2016," *Int. J. Adv. Stat. Probab.*, vol. 6, no. 1, p. 24, 2018.

[13] S. Rachad, B. Nsiri, and B. Bensassi, "System identification of inventory system using ARX and ARMAX models," *Int. J. Control Autom.*, vol. 8, no. 12, pp. 283–294, 2015.

[14] H. N. Akouemo and R. J. Povinelli, "Data Improving in Time Series Using ARX and ANN Models," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3352–3359, 2017.

[15] Y. Guo, E. Nazarian, J. Ko, and K. Rajurkar, "Hourly cooling load forecasting using time-indexed ARX models with two-stage weighted least squares regression," *Energy Convers. Manag.*, vol. 80, no. April, pp. 46–53, 2014.

[16] S. S. M. Khalifa, K. Saadan, and N. M. Norwawi, "Risk Assessment of Mined Areas using Fuzzy Inference," *Int. J. Artif. Intell. Appl.*, vol. 6, no. 118, pp. 37–51, 2015.

[17] T. M. Inc, "Fuzzy Logic Toolbox™ User's Guide," *The Mathworks, Inc*, 2014. .