Artificial Neural Networks for Seasonal Time Series Applied to Tourism Demand Forecasting

Tomás Molinet Berenguer [1], Napoleón Conde Gaxiola [2],
José Antonio Molinet Berenguer [3]

[1] Centro de Estudios Multidisciplinarios del Turismo, Universidad de Camagüey, Cuba.
E-mail: tomas.molinet@reduc.edu.cu
[2] Escuela Superior de Turismo del Instituto Politécnico Nacional, México.
E-mail: napoleon_conde@yahoo.com.mx
[3] Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, México.
E-mail: jmolinet@computacion.cs.cinvestav.mx

Abstract. In this paper, we propose a new artificial neural network (ANN) structure to predict seasonal time series and we study its application to tourism demand forecasting. Our proposal uses the seasons of tourist arrivals and values of months with similar behavior as input variables and achieves a forecast up to a year in advance. Our proposed approach is compared with respect to other forecasting techniques available in the literature, such as Autoregressive Moving Average (ARIMA) models and others ANN models. We study the validity and precision of the proposed model using two tourism demand time series. Our preliminary experimental results indicate that the proposed forecasting structure provides a longer period in advance and high forecasting accuracy than previously used ARIMA and ANN models.

Keywords: artificial neural networks, time series, tourism demand forecasting

1 Introduction

The use of artificial neural networks has received increasing attention in the modelling and forecasting of financial time series [6]. When modelling univariate time series using neural networks, the so-called time delay subset of the series is used as network input values. The values of this subset are previous to the value to be forecast, which constitutes the output of the neural network [3]. Several authors have rated that the best neural network structure is obtained when its input values correspond to the 12 months immediately prior to the month to be forecast [4,5],[9,10], which means that each time series value depends directly on the twelve previous values. Some modifications have been proposed to the structure that uses the twelve prior months. For example, in [12] the data are quarterly and use a range of 2 to 8 backward terms, *i.e.*, from 6 months to 24 months as inputs, and between 1 and 2 quarters as outputs of the network model. In [13] was proposed a new ANN methodology that uses a time index variable as the network input value, which may or may not be accompanied by the observations corresponding to the 12 previous months. They argue that thereby it is possible to capture the seasonal nature and tendency of the series in a better way. However, these ANN models only allow a better forecast a short time in advance and by using values nearby in time for the forecast the results may be affected in non-consolidated destinations in which the behaviour of tourism demand can vary sharply from one year to the next.

In recent studies were proposed neural network models which incorporate other input values to improve the forecasting accuracy in seasonal and non-linear time series. In [1] was suggested a multiplicative seasonal ANN (MSANN) to forecast time series with both trend and seasonal patterns. The input layer of MSANN is composed of two parts. The first part consists of inputs for trend component and the other one includes inputs for seasonal component, the weighted sum of each component corresponds to one of two hidden neurons. On the other hand, in [15] was proposed a neural network model which considers that time series has both linear and nonlinear components. The input layer of this model is composed by two parts, one corresponding to the linear component and the other one contain the non-linear component. The hidden layer consists of two neurons, the output value of one is equal to the weighted sum of linear component and second neuron corresponds to weighted multiplication of the non-linear component.

In this study we propose an ANN structure which, by using time series, different seasons of arrivals and values of months with a similar behaviour, can obtain a tourism demand forecast up to one year in advance, both in mature destinations and non-consolidated ones, with a quality that can be considered of high and good accuracy, respectively.

2 Formulating the proposed ANN structure

One of the most widely used neural networks in time series forecasting is the Multilayer Perceptron (MLP) [7]. The neurons in the input layer represent the values of the variable in the past and the only neuron in the output layer represents the value to be forecast. Generally, the network input values are the 12 months prior to the month to be forecast. However, this limits the forecast to only a month in advance. This is why we propose a model which will allow us to obtain the value of tourism demand with a forecast of up to a year in advance and with a similar error to the models that forecast less time in advance.

The proposed model uses the values of tourism demand in previous years as input values for the ANN, thereby we obtain a forecast of up to a year in advance and by considering several periods of time, the particular behaviour of one of them will not affect the result to a great extent. The name of the month to be forecast and the season it belongs to also make up input values, as their use enables the ANN to better adjust the specific period to be forecast. Moreover, it is considered that by using the information afforded by the months nearest in time to the month to be forecast which correspond to the same season as the network input, this helps to improve the precision of the forecast, as the behaviour of tourism demand is similar in each of them.

Fig. 1 shows the network input values, which are: the name of the month to be forecast (name of month), the season it belongs to (season), the values of this month in the previous i-th years (value in $year_{n-i}$) and the values of the two months nearest to the month to be forecast, which correspond to the same season for each of the previous i-th years ($neighbour_1$ in $year_{n-i}$ and $neighbor_2$ in $year_{n-i}$). With the aim of obtaining the forecast of the month in year n (value in $year_n$) as the output.

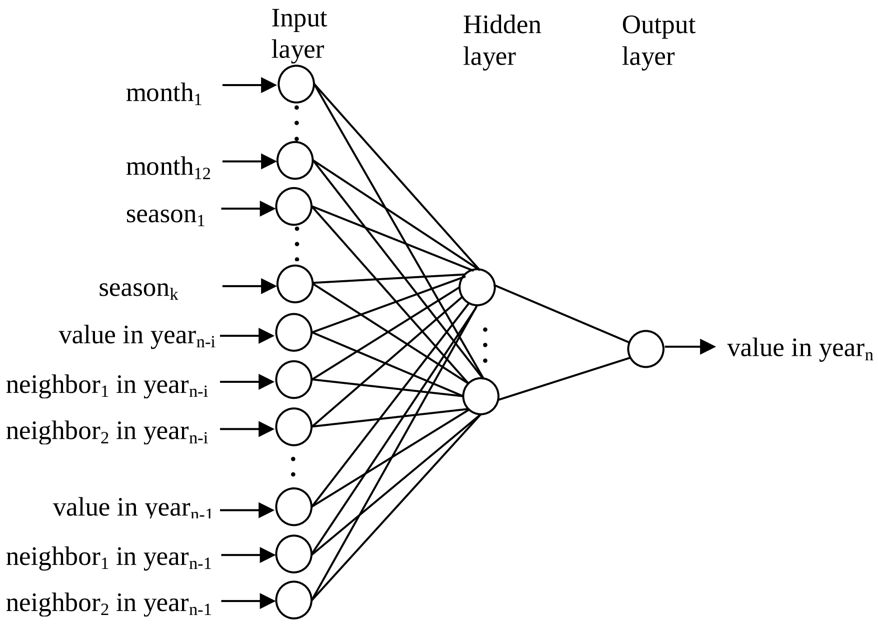

Fig. 1. Graphic representation of the proposed ANN structure.

3 Experimental results

To compare the models we forecast the tourism demand in two destinations with totally opposite characteristics, thus to know the differences that could mean the use of ANN models in forecasting demand. Mexico is a tourist destination with a very stable demand behavior while Santa Lucía de Cuba has a demand behavior complex and unstable, where the behavior of tourism demand can vary greatly from year to year. It is in destinations with these characteristics where the general accuracy of the forecasting models cannot be regarded as highly accurate.

The results obtained by the proposed model in forecasting each of the aforementioned destinations were compared with the ones found by an ANN model with two different forecasting horizons: ANN-month and ANN-year. The first used the 12 months immediately before to the month to be forecast, it means that for predict the month t is used: month $t-1$, month $t-2$, month $t-3$,…, month t-12. It is important to note that the prediction obtained with ANN-month has only a month in advance. The second variant used the 12 months that have a difference of at least one year with the month to be predicted, it means that for obtain the month t is used:

month *t-13*, month *t-14*, month *t-15*,…, month *t-24*. The ANN-year forecast with a year in advance. These results were also compared with those obtained by an ARIMA model with a forecasting horizon of one month in advance. This comparative analysis is carried out in all cases between the data of the original series and the corresponding estimated values for each model.

The criteria used in order to compare the accuracy of the different models are mean absolute percentage error (MAPE), due to its widespread use to assess the performance of the forecast [14], the Median and Huber's M-estimator, the latter two do not show the sensitivity of MAPE when faced with the existence of outlier values far from the centre of the distribution.

3.1 Definition of training, validation and test sets

The data corresponding to the tourist destinations of Santa Lucía de Cuba [11] and Mexico [2] were divided into three sets, 96 months which correspond to the training set, 12 months for the validation set and 12 months for the test set.

In the case of the tourist destination Santa Lucía de Cuba, for the proposed ANN model, ANN-month and ANN-year the observations from January 2001 to December 2008 were defined as the training set, the ones corresponding to January to December 2009 as the validation set, and from January to December 2010 as the test set. Meanwhile, for Mexico, the period ranging from January 2003 to December 2010 is considered as the training set, from January to December 2011 as the validation set, and from January to December 2012 as the test set.

Due to the fact that the proposed ANN model uses information referring to the three previous years, the training set will then have 60 patterns; on the other hand, the ANN-month model, as it only uses information from the 12 previous months, has 84 patterns in its training set, and the ANN-year, that uses information referring to the 24 previous months, has 72 patterns in its training set. Whereas, for all models both the validation set and the test set have 12 patterns.

3.2 Assessment of results

In order to obtain the best network architecture of the proposed ANN, ANN-month and ANN-year models for each destination, different training sessions were carried out, in each of which one or two hidden layers were taken, with the number of neurons varying in each of these layers between 1 and 30, avoiding a complex network structure which would impede their generalization. In the learning process the algorithms used were Back Propagation, which is widely used in different problems [7], [16] and the conjugated gradient. The former was applied independently or followed by the latter, which has the advantage of rapid convergence. In each case the number of iterations in the training varied and the network with the best performance in forecasting the validation set was chosen (see Table 1). The activation function used in the hidden neurons was the sigmoidal one and in the output neuron the linear one.

In order to choose the best ARIMA model in each destination, all the ARIMA models were constructed ARIMA(p,d,q)x(sp,sd,sq)$_{12}$ by varying the parameters *p, d, q, sp, sd* and *sq* between 0 and 2, values greater than 2 are generally not necessary [4]. The training set were used, without transforming and by applying the natural logarithm transformation; so as to calculate the coefficients of each model and the least error in the forecast of the validation set was chosen as the best model (see Table 1).

Table 1. ANN and ARIMA models obtained for forecasting each destination.

Model / Destination	proposed ANN	ANN-month	ANN-year	ARIMA
	input-hidden-output layers			
Santa Lucía	24-2-1	12-4-1	12-10-1	ARIMA (2,0,0)x(2,1,2)$_{12}$
Mexico	24-2-1	12-4-1	12-10-1	ARIMA (2,0,1)x(1,0,2)$_{12}$

In accordance with the classification given in [8] the absolute relative errors of the forecasts that are lower than 10% are considered 'highly accurate', between 10 and 20%, are considered 'good', between 20 and 50% are considered 'reasonable', and greater than 50%, 'imprecise/not very reliable'.

Below we analyse the forecasting capacity for each model, first of all in a non-consolidated destination (Santa Lucía de Cuba) and then in a consolidated one (Mexico), taking into account the aforementioned criterion in order to classify the quality of the forecast and the characteristics peculiar to each type of destination.

The neural network model applied was Multilayer Perceptron (MLP). In this study no change was made in the MLP model, but is propose a set of new inputs not previously considered and which are relevant to achieve an improvement in prediction. The input values considered in our work allow to describe more precisely the behavior of series with a marked instability. In order to test the relevance of the variables: month name and season, in the effectiveness of the forecast, we performed a sensitivity analysis to the proposed model. We constructed several subsets with the input variables and was selected the three subset that obtained the best accuracy in the forecast.

Table 2 shows the results of the sensitivity analysis. In the table, $N_{i,j}$ is the value of the *j*-th neighbor month in the year *n-i*, and V_i is the value of the month to predict but in the year *n-i*. The variable month name and season were selected in 3 networks and season was placed in the first rank. A ratio value equal to 1 or less indicates that the input variable is irrelevant and a higher value indicates more importance. The Rank indicates the order of importance of the input variables. Table 3 shows the results of the sensitivity analysis performed to all variables that are proposed, this results correspond to the tourism demand forecasting in Mexico. In the case of St. Lucia all input variables turn out to be relevant, see Table 4.

In Table 5, we can see the results obtained by the different models in the tourist destination Santa Lucía. Our propose approach achieve a prediction with a MAPE of

23.39%, based on the study of Lewis [8] this is a reasonable prediction. Since the instability of tourism demand in Santa Lucia and the results obtained by other models, we can consider that results achieved by our proposal are reasonable. Although the proposed ANN has a MAPE 5% higher than ANN-month, must be taken into account that has a forecast horizon 12 times higher than the model ANN-month, which is an important advantage for decision making.

Table 2. Results obtained for Mexico destination by creating subsets of independent variables.

	Month	Season	V_3	$N_{1,3}$	$N_{2,3}$	V_2	$N_{1,2}$	$N_{2,2}$	$N_{1,2}$
Ratio.1	1.34	1.52	1.44				1.36		
Rank.1	4	1	2				3		
Ratio.2	1.90	2.57	1.05	1.23				1.26	
Rank.2	2	1	5	4				3	
Ratio.3	1.32	2.42	1.53		1.01	1.10	1.36		0.99
Rank.3	4	1	2		6	5	3		7

Table 3. Results obtained for Mexico destination using all input variables of the model.

	Month	Season	V_3	$N_{1,3}$	$N_{2,3}$	V_2	$N_{1,2}$	$N_{2,2}$	V_1	$N_{1,1}$	$N_{1,2}$
Ratio	1.69	2.13	1.03	1.10	1.00	1.00	1.01	1.01	1.18	1.02	1.01
Rank	2	1	5	4	11	10	7	8	3	6	9

Table 4. Results obtained for Santa Lucía destination using all input variables of the model.

	Month	Season	V_3	$N_{1,3}$	$N_{2,3}$	V_2	$N_{1,2}$	$N_{2,2}$	V_1	$N_{1,1}$	$N_{1,2}$
Ratio	1.45	2.83	1.53	1.09	1.04	1.49	1.48	1.36	1.23	1.27	1.06
Rank	5	1	2	9	11	3	4	6	8	7	10

The results obtained with the ARIMA model for this destination are shown in Table 5. The accuracy achieved for this model in the validation set is lower than that obtained by the ANN models. This inadequate fit to the validation data is reflected to a greater extent in the real forecast which it is capable of achieving in the test set, as even ARIMA, which carries out the forecast a month in advance, achieves results inferior to the proposed ANN. Similar results to the ARIMA models in forecasting demand were obtained in [4].

One of the reasons for the difference between the results obtained in each of the sets is due to the fact that from 2007 to 2009 there was a continuous growth in tourism demand but in 2010 there was a sharp drop of 25% (see Fig. 2). As the ANN-month model uses real values for the months prior to the month to be forecast, it is capable of gradually identifying this decrease, however the proposed ANN model only has the information from previous years to model the behaviour of the demand. However, compared with results obtained by the ANN-year the proposed model has a better performance because forecast with an error of 4.5% lower, although both predict with the same period of time.

Table 6 shows the results obtained for all models in the validation and test sets corresponding to tourism demand in Mexico. In the validation set, the proposed ANN structure had a better fit than the others (see Fig. 3). Furthermore, in the test set it has a less error than the obtained by ANN-month, even though it forecast with a greater time in advance. Although the ANN-year structure predicts with the same period of anticipation, has a 1.1% greater error than the obtained by the proposed ANN structure. In general terms, the forecast for both models in each set can be classified as highly precise.

Table 5. Forecasts obtained with the ANN and ARIMA models in Santa Lucía de Cuba.

	Validation set			**Test set**		
	MAPE	Median	Huber's M-estimator	MAPE	Median	Huber's M-estimator
Proposed ANN	10.14	5.08	5.38	23.39	17.92	23.33
ANN-month	7.59	7.02	6.91	18.18	11.86	12.46
ANN-year	8.25	4.25	4.81	27.93	24.87	23.93
ARIMA	13.82	10.16	12.49	34.51	33.14	30.64

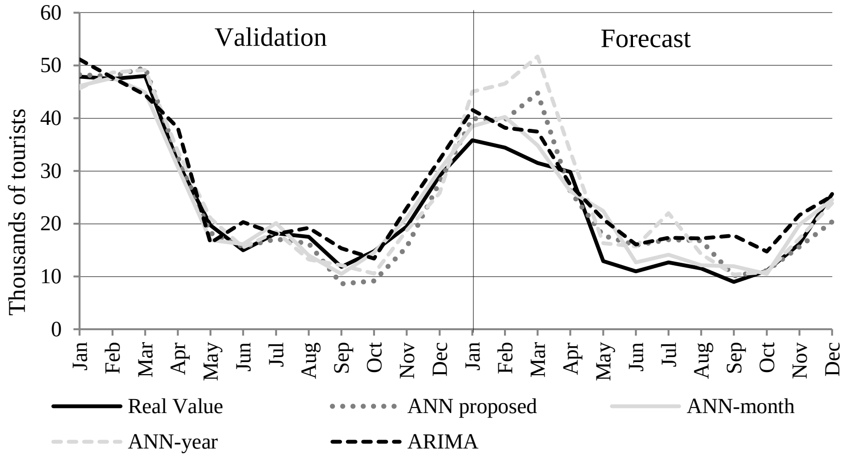

Fig. 2. Real values and forecasts obtained with the ANN and ARIMA models in Santa Lucía de Cuba.

Table 6. Forecasts obtained with the ANN and ARIMA models in Mexico.

	Validation set			**Test set**		
	MAPE	Median	Huber's M-estimator	MAPE	Median	Huber's M-estimator
Proposed ANN	1.34	0.69	0.87	3.51	3.19	2.98
ANN-month	6.98	6.6	6.85	4.07	1.77	2.35
ANN-year	6.83	4.99	5.26	4.59	3.77	3.96
ARIMA	6.52	5.56	5.67	4.22	3.04	3.27

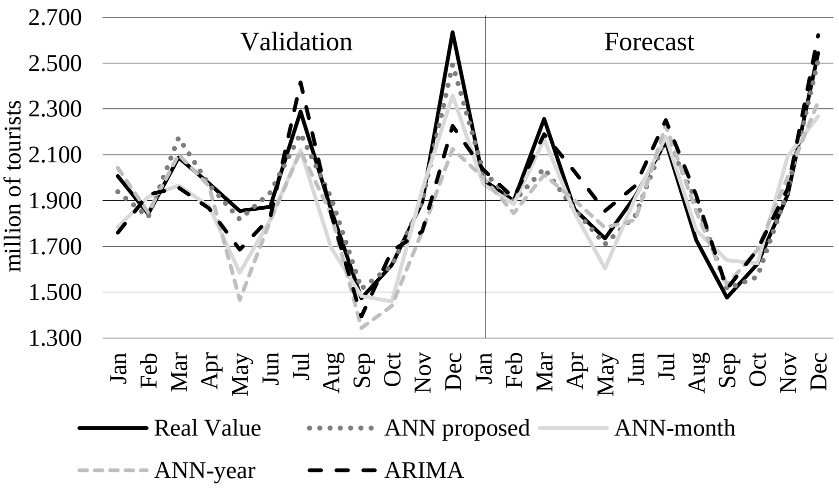

Fig. 3. Real values and forecasts obtained with the ANN and ARIMA models in Mexico.

4 Conclusions

In this study we proposed an ANN model which by using time series, different arrival seasons and values of months with similar behaviour, enabled us to obtain a demand forecast up to a year in advance. The idea of using the month and the season as input values has the aim to characterize the time series with other features that will provide more information to the neural network and improve its ability to detect patterns in time series behaviour. The input values incorporated to our network model enhance its performance mainly in unconsolidated tourist destinations, which have a high variability in its behaviour and hence, the time series have a high degree of non-stationarity. The proposed ANN was compared with an ARIMA model and two neural network models: ANN-month and ANN-year, they have a forecasting horizon of one month and one year, respectively. These last two ANN models only consider as inputs the values of the previous months. Our experimental results show that our proposal improves the performance of ANN-month and ANN-year. First, the forecasting horizon of the proposed ANN is longer than ANN-month and the forecasting accuracy is competitive. Second, the accuracy obtained by our proposal outperforms ANN-year.

The results obtained in forecasting demand in a consolidated and non-consolidated tourist destination indicate that the proposed model may be effective in forecasting demand and obtaining better results than those obtained by traditional ARIMA model and the ANN-year model. Moreover, the forecast carried out by the proposed model is similar to that obtained by the ANNs that use the 12 months prior to the one to be forecast, but with one year in advance.

5 References

1. Aladag, C. H., Yolcu, U., Egrioglu, E.: A new multiplicative seasonal neural network model based on particle swarm optimization. Neural Processing Letters, 37(3), 251-262 (2013)
2. BANXICO. Información económica. Banco de México, México (2013)
3. Bishop, C. M.: Neural networks for pattern recognition. Oxford University Press, Oxford (1995)
4. Cho, V.: A comparison of three different approaches to tourist arrival forecasting. Tourism Management, 24, 323-330 (2003)
5. Fernandes, P. O., Teixeira, J. P., Ferreira, J. M., Azevedo, S. G.: Modelling Tourism Demand: A Comparative Study between Artificial Neural Networks and the Box-Jenkins Methodology. Romanian Journal of Economic Forecasting 5, 30-50 (2008)
6. Kim, T. Y., Kyong, J. O., Kim, C., Do J. D.: Artificial neural networks for non-stationary time series, Neurocomputing 61, 439-447 (2004)
7. Law, R.: Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. Tourism Management 21, 331-340 (2000)
8. Lewis, C. D.: Industrial and Business Forecasting Method. Butterworth Scientific, London (1982)
9. Lin, C.-J., Chen, H.-F., Lee, T.-S.: Forecasting Tourism Demand Using Time Series, Artificial Neural Networks and Multivariate Adaptive Regression Splines: Evidence from Taiwan. International Journal of Business Administration 2, 14-24 (2011)
10. Montaño, J. J., Palmer, A., Muñoz, P.: Artificial neural networks applied to forecasting time series. Psicothema 23, 322-329 (2011)
11. ONE. Anuario estadístico de Cuba. Oficina Nacional de Estadística, Cuba (2011)
12. Palmer, A., Montaño, J. J., Sesé, A.: Designing an artificial neural network for forecasting tourism time series. Tourism Management, 27, 781-790 (2006)
13. Teixeira, J. P., Fernandes, P. O.: Tourism Time Series Forecast -Different ANN Architectures with Time Index Input. Procedia Technology 5, 445–454 (2012)
14. Witt, S. F., Witt, C. A.: Modelling and Forecasting Demand in Tourism. Academic Press, London (1992)
15. Yolcu, U., Egrioglu, E., Aladag, C. H.: A New Linear & Nonlinear Artificial Neural Network Model for Time Series Forecasting. Decision Support Systems, 54(3), 1340-1347 (2013)
16. Zhang, G., Patuwo, B. E., Hu, M. Y.: Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14, 35–62 (1998)