EUROPEAN UNION'S HISTORY, CULTURE AND CITIZENSHIP

UNIVERSITY OF PITEŞTI

FACULTY OF LAW AND ADMINISTRATIVE SCIENCES

Prediction Tourism Demand Using Artificial Neural Networks

Fernandes, Paula (pof@ipb.pt)
Department of Economics and Management

Teixeira, João (joaopt@ipb.pt)
Department of Electrical Engineering
School of Technology and Management (ESTiG)
Polytechnic Institute of Bragança (IPB)
Apartado 134, 5301-857 Bragança, Portugal
Tel: +351273303103, Fax: +351273313051

ABSTRACT

The aim of this research is to quantify the tourism demand using an Artificial Neural Network (ANN) model. The methodology was focused in the treatment, analysis and modulation of the tourism time series: "Monthly Guest Nights in Hotels" in Northern Portugal recorded between January 1987 and December 2006, since it is one of the variables that better explain the effective tourism demand. The model used 4 neurons in the hidden layer with the logistic activation function and was trained using the *Resilient Backpropagation* algorithm. Each time series forecast depended on 12 preceding values. The developed model yielded acceptable goodness of fit and statistical properties and therefore it is adequate for the modulation and prediction of the reference time series.

Key Words: Time Series, Tourism Demand, Artificial Neural Networks, Prediction.

1. INTRODUCTION

Tourism demand is usually measured in terms of the number of tourist visits from an origin country to a destination country. Alternative measures used to quantify tourism demand are tourist expenditures by visitors from an origin country in the destination country and the number of nights spent in the destination country (Witt & Witt, 1995). According with same authors accurate forecasts of tourism demand are prerequisite to the decision-making process in many organizations of the private or public sector. Any information concerning the future evolution of tourism flows is of great importance to hoteliers, tour operators and other industries concerned with tourism or transportation, in order to adjust their policy and corporate finance.

In this context, this paper examines the accuracy of a forecasting model in predicting tourism demand, as represented by the number of Monthly Guest Nights in Hotels in the North of Portugal. The Artificial Neural Networks (ANN) methodology is employed to forecast the volume of tourist from January 1987 to December 2006. This methodology was inspired in the biologic theories of human brain function. The human brain is composed of several non-linear processors densely interconnected operating in parallel, these being the principal advantages compared with other forecast techniques. This was achieved through a study of the reference time series whose past values were known and whose objective was to obtain a model that better predicts the behaviour of the time series under study.

This paper is organized in the following structure: first, there is an overview section that examines the summary theoretical foundation of neural networks. Based on the theoretical analysis, a neural network is developed for forecasting the "Monthly Guest Nights in Hotels" in Northern Portugal. Real data from official publications in Portugal is used for the neural network development. The analysis results of forecasting are described in the next section. Some concluding remarks are given in the final section.

2. NONLINEAR MODELLING APPROACH

A neural network is a collection of interconnected simple processing elements. Every connection in a neural network has a weight attached to it. There are countless learning methods for neural networks. However, they can be classified into two groups, namely supervised and unsupervised method. Supervised learning requires historical data with examples of both dependent and independent variables to train the network. The known answers are worked as a teacher to correct the behaviour of the training network. Unsupervised learning method creates its own model to interpret the data without known answers (Basheer & Hajmeer, 2000; Fernandes, 2005).

Neural networks are the most versatile nonlinear models that can represent both nonseasonal and seasonal time series. The most important capability of neural networks compared to other nonlinear models is their flexibility in modelling any type of nonlinear pattern without the prior assumption of the underlying data generating process (Haykin, 1999; Zhang, 2003).

A feedforward neural network learns from a supervised training data to discover patterns connecting input and output variables. Feedforward recall is a one-directional information processing neural network in which the signal flows from the input units to the output units in a forward direction (Kuan & White, 1994; Nam & Schaefer, 1995; Yao *et al*., 2000).

Backpropagation is the most popular neural network training algorithm that has been used to perform learning on feedforward neural networks. It is a method for assigning responsibility for mismatches to each of the processing units in the network, which is achieved by propagating the gradient of the activation function back through the network to each hidden layer, down to the first hidden layer. The weights are then modified so as to minimize the mean squared error between the network's prediction and the actual target (Thawornwong & Enke; 2004). The Backpropagation neural network consists of an input layer, an output layer and one or more intervening layers also referred to as hidden layers. The hidden layers can capture the nonlinear relationship between variables. Each layer consists of multiple neurons that are connected to neurons in adjacent layers. Since these networks contain many interacting nonlinear neurons in multiple layers, the networks can capture relatively complex phenomena (Hill *et al*., 1996; Chiang *et al*., 1996; Basheer & Hajmeer; 2000). Many variant were developed of Backpropagation training algorithm. In our case we adopted the *Resilient Backpropagation* [RP] (Reidmiller & Braun, 1993), because it can combine fast convergence, stability and generally good results.

The most popular feedforward three-layer network for forecasting problems can be written as (Fernandes, 2005):

$$Y_t = b_{2,1} + \sum_{j=1}^{n} w_j f\left(\sum_{i=1}^{m} W_{ij} y_{t-i} + b_{1,j}\right) \qquad [1]$$

where,

m, is the number of input nodes; n, is the number of hidden nodes; f, is a sigmoid transfer function; $\{w_j, j = 0,1,\ldots,n\}$, is a vector of weights from the hidden to output nodes; $\{W_{ij}, i = 0,1,\ldots,m; j = 1,2,\ldots,n\}$, are weights from the input to hidden nodes; $b_{2,1}$ and $b_{1,j}$, are the bias associated with the nodes in output and hidden layers, respectively.

The equation shows a linear transfer function used in the output node. The input variables are the lagged past observations if the time series data are used in model building.

3. PREDICTION TOURISM DEMAND IN THE NORTH OF PORTUGAL

3.1. DESIGNING AN NEURAL NETWORK MODEL

The data used in this research are Monthly Guest Nights in Hotels in the North of Portugal, published in the Portuguese National Statistical Institute. The total sampling period examined is from January 1987 to December 2006 (see Table A.1, in Appendix).

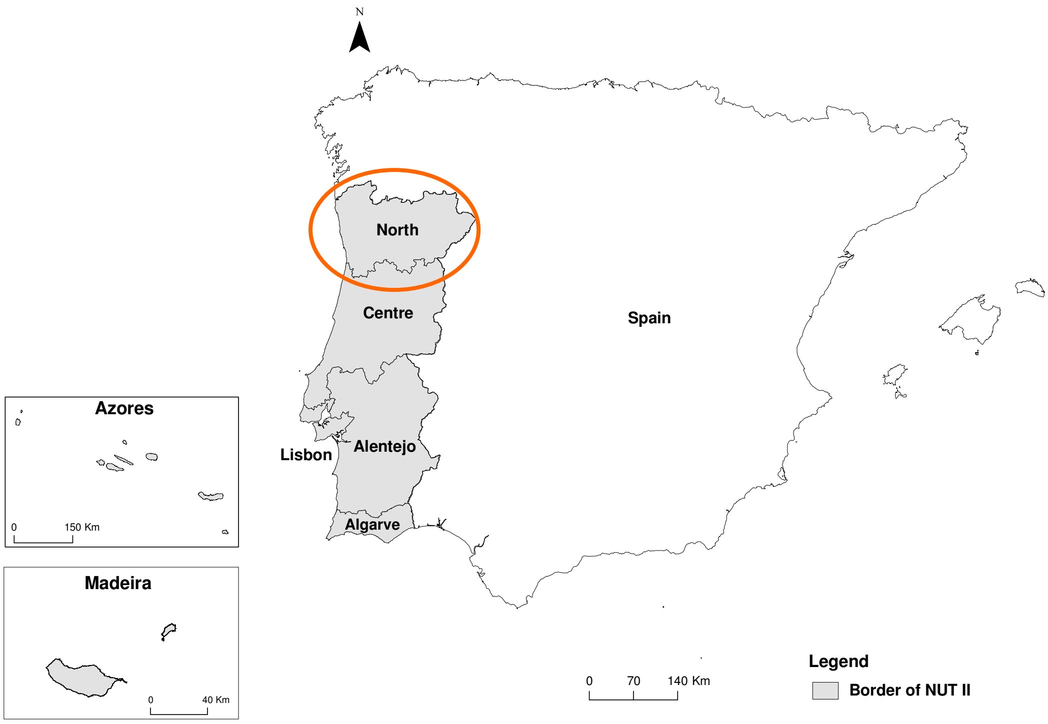

Fig. 1: Regions of Portugal.

The Northern region of Portugal is delimited in Fig 1. The time series "Monthly Guest Nights in Hotels" presents a clear increasing seasonality (Fig. 2a), the natural log-transformation is performed to stabilize the seasonal variations and variance, and we have another time series. During this study we call this time series, the Transformed Original Data [TOD] (Fig. 2b).

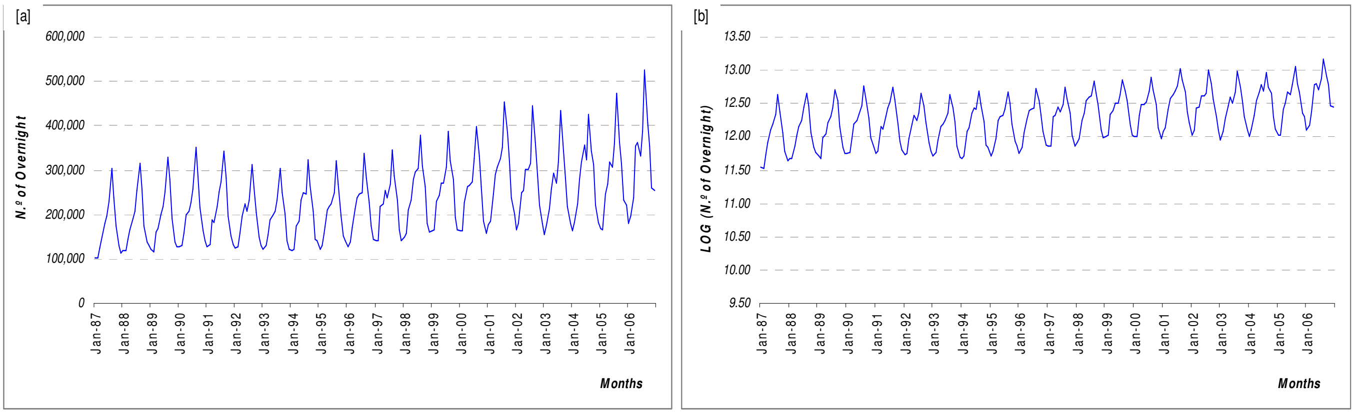

Fig. 2: Overnights in the North of Portugal from 1987:01 to 2006:12: [a] Original Data; [b] Transformed Original Data.

The ANN model used in this study is the standard three-layer feedforward network. Since the one-step-ahead forecasting is considered, only one output node is employed. The activation function for hidden nodes is the logistic function [*Logsig*]: $f(x)=\frac{1}{1+e^{(-x)}}$; and for the output node the identity function [*Lin*]: $f(x)=x$. Bias terms are used in both hidden and output layer's nodes. The fast *Resilient* Backpropagation algorithm[1] is employed in training process. The ANN is randomly initialised with weights and bias values. The selection of the architecture is supported in the author's work Fernandes (2005), Fernandes and Teixeira (2007). For selecting the architecture several experiments with different architectures was carried out (train and test) and selected the better architectures according to the results in a validation set using hundreds of training session. The elected architecture consists of 12 input nodes in the entrance layer, 4 hidden nodes in the second layer and one node in the output layer (1-12;4;1). The input of the model consists of the 12 previous numbers, corresponding to the last 12 months overnights. The output is the predicted overnights for the next month.

Several training sessions for each identified situation have been performed with different initial weights. From this number of training sessions we retain the ANN

[1] Provide by the MATLAB 7.1 neural network toolbox.

(concerning its weights) that obtain better forecast results in each situation under the validation set. In this particular situation we performed 250 training sessions.

In order to compare the performance, the root mean squared error (RMSE[2]) between the observed and predicted values are used as the agreement index and we adopted this index to select the best ANN model.

Also in the training process, for each session we need to establish the number of iterations and the goal. In the present study we defined our goal as an error (RMSE between target and predicted values) of the order of $1x10^{-4}$.

The data sample was divided in a sub-set for training, a sub-set for validation and a sub-set for test. The data set between January 1988 until December 2003 was used for training. It must be notice that the data between January and December 1987 was used as the input data for predicting January 1988 till December. The data between January and December 2004 was used for the validation set. This set is used for early stop training if the RMSE does not decrease in a number (5 in this case) of training iterations. This early stop training condition avoids the ANN to over fit the training data without improvements in a data not used in the training phase. Finally the data between January 2005 and December 2006 was used as data never seen in the training and selection process and used just to present the results of the model with never seen data.

4. RESULTS AND DISCUSSION

In this section we will examine the results of ANN under the test set. For this purpose we will compare the predicted data of ANN with the target values (original data) for the years 2005 and 2006 (the test set). The selection process of the better ANN is controlled by the minimum RMSE in the training set.

We should look at the performance in the test set. The predicted values for the years of 2005 and 2006 with ANN model and its APE and MAPE are presented in Table 1. APE is the absolute percentage error given by the expression [2]. MAPE is the Mean absolute percentage error given by the expression [3].

$$APE = \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100. \qquad [2]$$

2

$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - P_t)^2}{n}}$; *where A is the target value, P denotes the value of prediction and n the total number of observations.*

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100. \qquad [3]$$

Table 1: Prediction Tourism Demand with ANN model, APE and MAPE in the period 2005:01 to 2006:12.

2005			Months	2006		
TOD		Target Data		Target Data	TOD	
Values	APE				Values	APE
182.389	8.5%	168100	January	180.700	189.349	4.8%
181.870	9.0%	166800	February	195.100	183.731	5.8%
220.635	10.7%	247000	March	237.200	234.591	1.1%
284.692	6.0%	268500	April	352.600	285.916	18.9%
301.171	5.0%	316900	May	361.200	316.248	12.4%
333.732	8.5%	307700	June	331.500	312.298	5.8%
340.731	5.0%	358500	July	388.400	376.036	3.2%
416.740	11.8%	472400	August	524.500	414.580	21.0%
357.019	1.4%	362200	September	406.500	363.306	10.6%
328.557	4.0%	315900	October	353.300	337.129	4.6%
237.594	1.8%	233400	November	258.800	264.057	2.0%
196.989	11.0%	221300	December	254.700	218.612	14.2%
----	***6.9%***	**----**	***MAPE***	**----**	**----**	***8.7%***

Analysing the presented results in Table 1 and according to the Criteria of MAPE for Model Evaluation in Lewis (1982), the predicted data with the selected model has a highly accurate forecast, because the results are lower than 10%. We can say the predicted data for each year follow the behaviour of the target data. Therefore, in Table 1 we can observe an additional difficulty for the model imposed by the fact those years 2005 and 2006 have and increasing number of overnights, and this increasing phenomenon was present in the training set only in 2001. This phenomenon was due to the fact that the city of Guimarães and the Douro Region were considered World Cultural Heritage, and the city of Porto was the European Capital of Culture in 2001 (Fernandes & Teixeira, 2007).

5. CONCLUSIONS

This paper describes the process of modelling tourism demand for the north of Portugal, using an artificial neural network model. Data used in the time series was obtained from official publications - Portuguese National Statistics Institute. The time series was considered in the logarithmic transformed data and divided in a sub-set for training, in a validation set, to stop the training process earlier and a test data set to examine the level of forecasting accuracy.

The model has 4 neurons in the hidden layer with the logistic activation function and was trained using the *Resilient Backpropagation* algorithm. The ANN model has the 12 preceding values as the input. The analysis of the output forecast data of the selected ANN model showed a relatively close result compared to the target data.

Finally and considering the results, the developed model yielded acceptable goodness of fit and statistical properties and therefore it is adequate for the modulation and prediction of the reference time series; and the ANN methodology becomes interesting to forecast because it allows the use of a non linear model for seasonal time series

REFERENCES

Basheer, I.A. and **Hajmeer**, M.; "Artificial Neural Networks: fundamentals, computing, design and application"; *Journal of Microbiological Methods*; N.º 43, (2000); pp.3/31.

Chiang, W.C.; **Urban**, T.L. and **Baldridge**, G.W.; "A neural network approach to mutual fund net asset value forecasting"; *Omega, The International Journal of Management Science*; Vol. 24, N.º 2, (1996); pp. 205/215.

Fernandes, Paula Odete; "*Modelling, Prediction and Behaviour Analysis of Tourism Demand in the North of Portugal*"; Ph.D. Thesis in Applied Economy and Regional Analysis; Valladolid University - Spain; (2005).

Fernandes, Paula O. and **Teixeira**, João Paulo; "A new approach to modelling and forecasting monthly overnights in the Northern Region of Portugal"; *Proceedings of the 4th International Finance Conference;* Université de Cergy; Hammamet, Medina, Tunísia; (2007).

Haykin, Simon; "*Neural Networks. A comprehensive foundation*"; New Jersey, Prentice Hall; (1999).

Hill, Tim; **O'Connor**, Marcus and **Remus**, William; "Neural network models for time series forecasts"; *Management Science*; Vol. 42, N.º 7; (1996); pp.1082/1092.

Kuan, Chung-Ming and **White**, Halbert; "Artificial Neural Network: An econometric perspective"; *Econometric Reviews*; N.º 13; (1994); pp.1/91.

Lewis, C.D.; "*Industrial and Business Forecasting Method*"; Butterworth Scientific; London; (1982).

Nam, Kyungdoo and **Schaefer**, Thomas; "Forecasting International Airline Passenger Traffic Using Neural Networks"; *Logistics and Transportation Review*; Vol. 31, N.º 3; (1995); pp.239/251.

Reidmiller, M and **Braun**, H.; "A direct adaptive method for faster backpropagation learning: The RPRO algorithm"; *Proceedings of the IEEE International Conference on Neural Networks*; (1993).

Thawornwong, Suraphan and **Enke**, David; "The adaptive selection of financial and economic variables for use with artificial neural networks"; *Neurocomputing*; N.º 56; (2004); pp.205/232.

Witt, S. F. and **Witt**, C. A; "Forecasting Tourism Demand: a review of empirical research"; *International Journal of Forecasting*; N.º 11; (1995); pp. 447/475.

Yao, Jingtao; **Li**, Yili and **Tan**, Chem Lim; "Option price forecasting using neural networks"; *Omega, The International Journal of Management Science*; N.º 28; (2000); pp.455/466.

Zhang, G. Peter; "Time series forecasting using a hybrid ARIMA and neural network model"; *Neurocomputing*; N.º 50; (2003); pp.159/174.

APPENDIX A

Table A.1: Overnights in the North of Portugal from 1987:01 to 1995:12 (cont.).

Years / Months	1987	1988	1989	1990	1991	1992	1993	1994	1995
January	102.447	118.011	122.217	126.671	126.826	124.194	121.469	118.606	122.480
February	102.123	117.547	116.837	129.802	131.653	127.474	129.284	122.988	130.393
March	125.401	142.687	160.658	158.701	188.999	157.536	154.734	175.261	156.645
April	150.042	167.118	169.326	197.757	182.290	196.087	189.142	185.525	209.263
May	180.430	189.823	199.158	207.876	219.187	223.918	198.402	232.075	218.666
June	197.113	207.729	218.595	227.159	251.295	207.907	207.216	248.237	222.720
July	229.293	254.523	252.634	257.633	273.927	231.801	231.453	246.274	247.589
August	304.847	315.113	329.014	351.500	341.490	312.026	304.576	322.366	320.750
September	238.542	258.287	278.074	284.867	283.378	259.023	249.583	266.094	269.433
October	173.503	174.359	189.664	216.286	197.241	205.400	202.792	206.256	196.466
November	130.187	137.933	138.683	162.062	152.554	149.289	141.976	144.803	152.340
December	114.229	128.774	127.730	139.683	132.802	130.963	120.748	139.706	140.643
TOTAL	**2.048.157**	**2.211.904**	**2.302.590**	**2.459.997**	**2.481.642**	**2.325.618**	**2.251.375**	**2.408.191**	**2.387.388**

Table A.1: Overnights in the North of Portugal from 1996:01 to 2004:12.

Years / Months	1996	1997	1998	1999	2000	2001	2002	2003	2004
January	126.910	140.430	148.218	163.696	162.389	176.690	165.653	155.527	162.900
February	139.403	141.183	157.415	165.988	162.637	186.586	181.005	177.818	181.900
March	172.393	219.465	209.929	228.149	226.010	245.261	249.214	214.106	224.600
April	213.973	224.382	232.767	242.744	262.865	291.395	253.274	258.519	279.800
May	239.142	253.833	280.326	269.854	264.497	306.743	302.028	293.531	317.300
June	245.264	238.334	296.612	270.126	273.881	325.568	301.465	271.454	355.300
July	248.398	266.993	303.866	306.031	324.962	351.955	314.560	318.706	324.400
August	336.086	345.672	377.645	385.868	397.405	452.581	444.991	433.211	426.900
September	280.769	288.409	309.700	321.248	331.155	383.793	361.181	343.534	342.100
October	225.734	232.052	263.522	280.597	263.217	319.417	287.383	281.472	311.500
November	175.438	166.835	180.796	193.062	186.445	238.925	221.910	219.463	221.200
December	143.163	141.349	161.273	166.990	157.210	202.351	179.766	178.439	182.800
TOTAL	**2.546.673**	**2.658.937**	**2.922.069**	**2.994.353**	**3.012.673**	**3.481.265**	**3.262.430**	**3.145.780**	**3.330.700**