Comparison of linear regression and neural network models forecasting tourist arrivals to Turkey

Selcuk Cankurt, Abdulhamit Subasi

International Burch University, Faculty of Engineering and Information Technologies, Francuske Revolucije bb. Ilidza, Sarajevo, 71000, Bosnia and Herzegovina.

E-mail:*asubasi@ibu.edu.ba*

Abstract

This paper develops statistical and machine learning methods for estimating tourist arrivals which is one of the donnée for planning the sustainable tourism development. Tourism is arguably one of the world's largest and fastest growing industries. Sustainable tourism development is one of the most promising generators of the sustainable economic development. Realistic tourism projections based on accurate tourism forecasting contribute much for the sustainable tourism development. The challenge of the planning and developing sustainable tourism is to see as the complex paradigm but one of the starting points is the accurate forecasting tourist arrivals. In this study, linear regression and neural network multilayer perceptron (MLP) implementations are considered to make multivariate tourism forecasting for Turkey. Comparison of forecasting performances in terms of correlation coefficient (R), relative absolute error (RAE) and root relative squared error (RRSE) measurements shows that MLP model for regression gives a better performance.

Keywords: Tourism forecasting; Tourism demand modelling; Time series; Linear regression; Neural networks; Multilayer perceptron; Multivariate tourism forecasting.

1.INTRODUCTION

Tourism demand forecasts are of great economic value both for the public and private sector. Tourism products, such as unfilled airline seats, unoccupied hotel rooms, and unused facilities, cannot be stocked because of their perishable nature (Archer, 1987). Therefore, accurately forecasting tourism demand has great importance to the sectors concerned with tourism, in order to accurate and efficient plans (Petropoulos, Nikolopoulos, & V., 2005; Pai & Hong, 2005).

According to the World Travel & Tourism Council (WTTC), travel and tourism is the biggest industry in the world. Since 1992 tourism sector is the largest industry and has the largest employer in the world (Aslan, Alper, Kaplan, Muhittin, Kula, & Ferit, 2008).

Turkey's economy grew an average of 6.0% per year in last decade. Currently Turkey is in 16th place on the list of the largest economies of the world and the fastest growing economy among members of the Organization for Economic Cooperation and Development (OECD).

The new goals of Turkish tourism were to establish an efficient tourism sector with high international competitiveness while preserving and enhancing of the country’s natural and historical environment and cultural heritage in a sustainable manner (Ministry of Culture, 2007).

The statistical methods such as linear regression are suitable for data having seasonal or trend patterns, while artificial neural techniques are also efficient for data which are influenced by the special case, like promotion or extreme crisis (Efendigil, Önüt, & Kahraman, 2009).

One major application area of ANNs is forecasting (Gooijer & J., 2006); see (Zhang, Patuwo, & Hu, 1998) and (Hippert, Pedreira, & Souza, 2001). Generally the ANNs are increasingly used to forecast demands for tourism (Law & Au, 1999; Law R. , 2000). (Pattie & Snyder, 1996) used a back-propagation neural network model with two hidden layers to forecast monthly overnight stays in US national park systems. (Law & Au, 1999) presented a feed-forward neural network with six input and one output nodes to forecast arrivals in Hong Kong. For more application area of ANN, see (Al-Saba & El-Amin, 1999), (Beccali, Cellura, Lo Brano, & Marvuglia, 2004), (Hobbs, Helman, Jitprapaikulsarn, Konda, & Maratukulam, 1998), (Sozen, Arcaklioglu, & Ozkaymak, 2005), (Sabuncuoglu, 1998), (Vellido, Lisboa, & Vaughan, 1999), (Wong, Lai, & Lam, 2000), (Ayata, Cam, & Yıldız, 2007), (Efendigil, Önüt, & Kahraman, 2009).

According to the brief review of literature especially related to tourism demands approaches, this study attempts to develop a multivariate linear regression model and a general regression neural network model for forecasting the number of the tourists coming to Turkey.

2.THEORETICAL BACKGROUND

2.1.Linear regression

Multiple linear regression (MLR) attempts to model the linear relationship called the regression function between a dependent variable and more than one independent variables as different from simple linear models with one independent variable. The dependent variable is sometimes also called the predictand, and the independent variables is called the predictors.

The model for multiple linear regression, given n observations, is

$$y_i = \beta_o + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} + \varepsilon_i \qquad (1)$$

for i = 1,2, ... n.

$x_{i,p}$ value of p^{th} predictor, β_o the intercept, also known as the bias in machine learning, β_p coefficient on the p^{th} predictor, p total number of predictors, y_i predictand, ε_i error.

2.2 MLP Approach

Artificial neural networks (ANNs) (also usually preferred Neural Networks NNs) are computing structures inspired from the biological neural networks. A neural network is made of the interconnected processing units (usually called neurons). They have the ability of learning by adjusting the strength of the interconnections which can be achieved by altering the values called weights through the input data (Haykin S. , 1999). Neuron sums the weighted inputs and conveys the net input through an activation function in order to normalize and produce a result (Jones, 2008).

The multilayer network architecture consists of an input layer, two or more hidden layers, and one output layer. Activation function is used for both the hidden and output nodes. While the sigmoid function can be used to squash the output of the neuron to 0.0 to 1.0 in the hidden layer in order to introduce the non-linearity to NN, linear activation function must use in output layer to predict the numerical values in the regression problems. MLP is trained with supervised learning include the Perceptron learning algorithm, Least-Mean-Squares learning, and Backpropagation. Backpropagation is one of the most popular approximation approaches for training the multilayer feedforward neural networks based on the Widrow–Hoff training rule (Bishop, 1995; Haykin S. , 1999; Aslanargun, Mammadov, Yazici, & Yolacan, 2007).

3.EXPERIMENTAL RESULTS

A total of 31 models were obtained on the basis of two regression models and their corresponding parameter selection which are three of them belong to linear regression models and remaining 28 ones belong to MLP models. Those models were evaluated with the validation data through three forecasting accuracy measures: correlation coefficient (R), relative absolute error (RAE), root relative squared error (RRSE).

Three linear regression models were examined on the basis of attribute selection parameter: none, M5 and greedy methods. It has been shown that the linear regression model with greedy attribute selection parameter has the best accuracy when you compare with the other linear regression models but also the worst when you compare with MLP regression models.

According to result of our linear regression model: 25 attributes don’t affect the results — WEKA builds the regression function by considering the attributes which only statistically contribute to the accuracy of the model (measured in R^2). It will not consider the attributes that don't contribute the regression equation. So this regression model is telling us that whole sale price of Turkey, consumer prize index of Canada, Denmark, Spain, Russia, number of German, France, Syrian, Poland, Romanian, Norwegian, Switzerlandian visitors, Exchange rate of Russia, Canada, Switzerland don’t affect the arrivals to Turkey. Estimated positive values (coefficients) tell us as value of those attributes increase number of the total visitors. Estimated negative values (coefficients) reduce the result — linear regression model is telling us that the bigger negative value is, the lower the total coming tourist. This can be seen by the negative coefficient in front of the variables.

Table 1 Overall performance of linear regression and MLP methods

Model	Correlation coefficient	Relative absolute error	Root relative squared error
Linear Regression	0.978	18.73%	20.70%

MLP Regression	0.9874	14.17%	15.86%

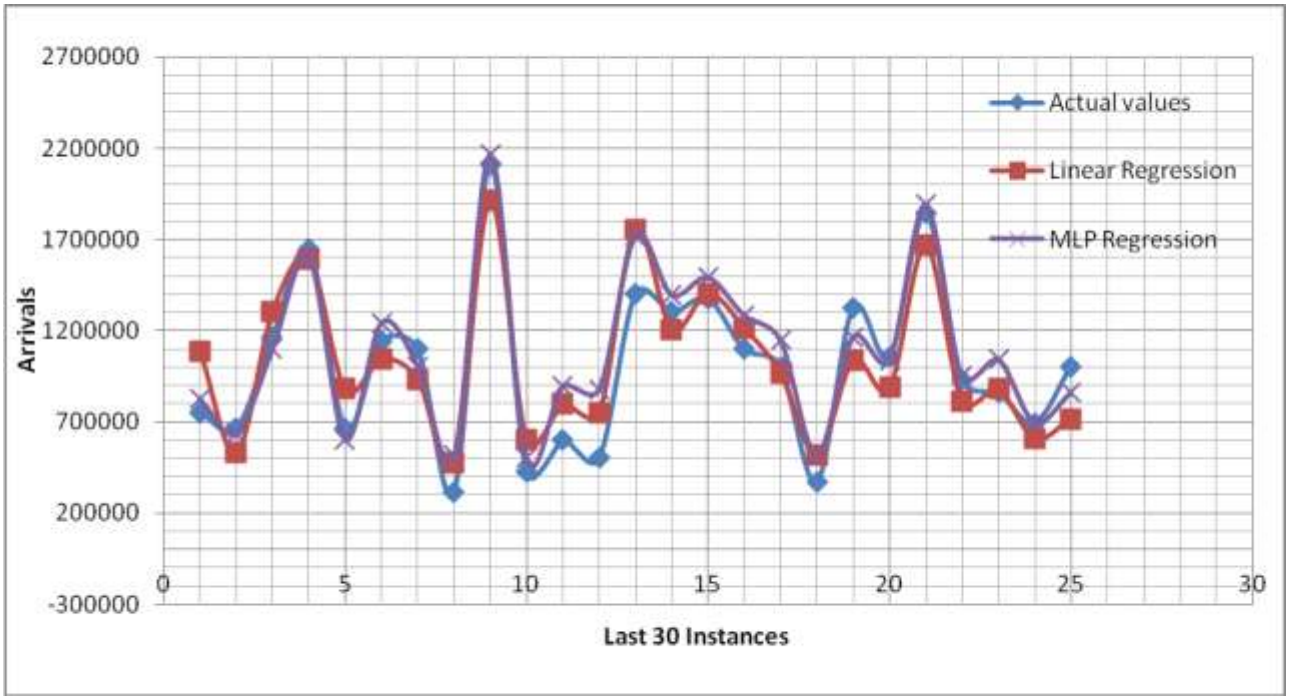

Figure 1 Comparison of MLP and linear regression methods

Among the MLP regression models presented, the best forecasting accuracy was the MLP model composed of three hidden layers with the neuron numbers of 30, 15 and 10 (abbreviated as 30-15-10). In this model the learning rate 0.03, momentum 0.8, epoch 500 values are used and backpropagation training algorithm, sigmoid activation function for hidden nodes and unthresholded linear activation function for output node are employed. It showed R 0.9874, RAE 14.17% and RRSE 15.86% accuracy results.

Results obtained from the experiments in this study, support the discussions in the literature reviews topic of this paper. As seen in the table (1) apparently, machine learning MLP regression model have better performance than statistical linear regression model.

4.CONCLUSIONS

This study presents a multivariate time-series forecasting to predict the tourism demand to Turkey by employing linear regression and multilayer perceptron methods. The real data sets respect to Turkey and its top ranked 24 tourism clients of the countries are used to compare the performance of the those methods and to find out the achievement of them on forecasting tourism demand to Turkey. Comparison of the experimental results among linear regression and MLP demonstrated that the MLP method had better forecasting accuracy. Experimental results showed that the MLP model can produce lower prediction error and higher prediction accuracy and outperformed the linear regression model. According to the experiments, it can be concluded that the tuned MLP method with the multivariate time series has enough satisfactory to forecast the tourism demand to Turkey.

In this study, linear regression model with greedy attributes selection method and MLP (30:15:10) models have shown better performance when compared with other corresponding models in forecasting the number of monthly tourist arrivals to Turkey owing to the RAE and the RRSE measures.

Unfortunately, there is no certain or systematic method to select the appropriate model. Our studies showed that among the methods mentioned above MLP regression has better performance but still we need numerous experiments to evaluate and find out the most suitable MLP regression model which can be employed on the multivariate time series forecasting.

REFERENCES

Al-Saba, T., & El-Amin, I. (1999). Artificial neural networks as applied to long-term demand forecasting. Artificial Intelligence in Engineering.

Archer, B. (1987). Demand Forecasting and Estimation. Travel, tourism, and hospitality research. A handbook for managers and researchers pp. 77-85 .

Aslan, Alper, Kaplan, Muhittin, Kula, & Ferit. (2008). Approach, International Tourism Demand for Turkey: A Dynamic Panel Data. Munich Personal RePEc Archive MPRA Paper No. 10601.

Aslanargun, A., Mammadov, M., Yazici, B., & Yolacan, S. (2007). Comparison of ARIMA, neural networks and hybrid models in time series: tourist arrival forecasting. Journal of Statistical Computation and Simulation Vol. 77, No. 1, January , 29–53.

Ayata, T., Cam, E., & Yıldız, O. (2007). Adaptive neuro-fuzzy inference systems (ANFIS) application to investigate potential use of natural ventilation in new building designs in Turkey. Energy Conversion and Management, 48, 1472–1479.

Beccali, M., Cellura, M., Lo Brano, V., & Marvuglia, A. (2004). Forecasting daily urban electric load profiles using artificial neural networks. Energy Conversion and Management.

Bishop, C. (1995). Neural Networks for Pattern Recognition.

Choy, K. L., Lee, W. B., & Lo, V. (2003). Design of an intelligent supplier relationship management system: A hybrid case based neural network approach. Expert Systems with Applications, 24, 225–237.

Efendigil, T., Önüt, S., & Kahraman, C. (2009). A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: A comparative analysis. Expert Systems with Applications 36 6697–6707.

Gooijer, J. G., & J., R. (2006). 25 years of time series forecasting. Hyndman International Journal of Forecasting 22 (2006) 443– 473.

H.Witten, I., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques (Second b.). New York: Elsevier.

Haykin, S. (1999). Neural Networks: a comprehensive foundation. (Second Edition b.). Prentice Hall.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: A review and evaluation. IEEE Transactions on Power Systems, 16, 44–55.

Hobbs, B. F., Helman, U., Jitprapaikulsarn, S., Konda, S., & Maratukulam, D. (1998). Artificial neural networks for short-term energy forecasting: Accuracy and economic value. Neurocomputing, 23, 71–84.

Jones, M. T. (2008). Artificial Intelligence: A Systems Approach. INFINITY SCIENCE PRESS LLC.

Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting.

Law, R., & Au, N. (1999). A Neural Network Model to Forecast Japanese Demand for Travel to Hong Kong. Tourism Management.

Mark Hall, E. F., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA Data Mining Software: An Update; SIGKDD Explorations. 11(1).

Ministry of Culture, T. (2007). Tourism strategy of Turkey – 2023. Ankara: Republic of Turkey of Ministry of Culture & Tourism.

Pai, P.-F., & Hong, W.-C. (2005). An Improved Neural Network Model in Forecasting Arrivals. Annals of Tourism Research, Vol. 32, No. 4, pp. 1138–1141, Elsevier.

Palmer, A., Montano, J. J., & Sese, A. (2006). Designing an artificial neural network for forecasting tourism time series. Tourism Management 27 781–790.

Pattie, D., & Snyder, J. (1996). Using a Neural Network to Forecast Visitor Behavior. Annals of Tourism Research.

Petropoulos, C., Nikolopoulos, K., & V., A. P. (2005). A technical analysis approach to tourism demand forecasting. Applied Economics Letters 12, 327–333.

Reinsel, G. C. (2003). Elements of multivariate time series analysis.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. Parallel distributed processing (pp. 318–362). Cambridge, MA: MIT Press.

Sabuncuoglu, I. (1998). Scheduling with neural networks: A review of the literature and new research directions. Production Planning and Control, 9(1), 2–12.

Sozen, A., Arcaklioglu, E., & Ozkaymak, M. (2005). Turkey's net energy consumption. Applied Energy, 81(2), 209–221. .

Vellido, A., Lisboa, P. J., & Vaughan, J. (1999). Neural networks in business: A survey of applications (1992–1998). Expert Systems with Applications, 17, 51–70.

Witt, S. F., & Witt, C. A. (1995, 3). Forecasting tourism demand: a review of empirical research, International Journal of Forecasting.

Wong, B. K., Lai, S. V., & Lam, J. (2000). A bibliography of neural network business applications research: 1994–1998. Computers & Operations Research 27,1045–1076.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial networks: The state of the art. International Journal of Forecasting, 14, 35– 62.