Artificial neural networks in forecasting tourists’ flow, an intelligent technique to help the economic development of tourism in Albania.

Dezdemona Gjylapi, MSc, PhD Candidate
University “Pavaresia” Vlore, Albania
Veronika Durmishi, PhD
University “Pavaresia” Vlore, Albania

Abstract

Tourism plays an important role in many economies and contributes greatly to the Gross Domestic Product.

In the past eight years, the number of tourist arrivals in Albania has increased rapidly, which resulted in increasing the number of tourist nights and revenue from tourism.

Tourism also provides new sources of income for the country, without having that local citizen to pay more taxes. This can be achieved by income from parking, tourist taxes, leased apartments, sales information, etc.

Early prediction on the tourist inflow mainly focuses on econometric models that have as a main feature the tourism demand being predicted by analysing factors that affect the tourists’ inflow. This approach results in being difficult, time-consuming and also expensive to determine econometric models. Traditional time series methods, such as exponential smoothing method, grey prediction method, linear regression method, ARIMA method etc., are more appropriate for the prediction of the tourist inflow. However, since they don’t apply a learning process on sample data, it is difficult for them to realize complicated and non-linear prediction on tourist inflow [1].

The aim of this paper is to present the neural network usage in the tourists’ number forecasting and to determine the trends of the future tourist inflow, thus helping tourism management agencies in making scientific based financial decisions.

Keywords: tourist inflow; tourism economy; neural networks; neuro-genetic; BPNN

The role of tourism in economic restructuring

Travel & Tourism is one of the world’s largest industries accounting 9% of global GDP of 2011. This is more than the automotive industry which accounts for 8.5%, and only slightly less than the banking sector which accounts for 11% [2].

Tourism is under an ongoing process and is one of the sectors with the fastest economic growth in the world, which has vastly expanded in the recent years. Tourism
is not just an activity, but is a collection of very special and similar activities, which include: transportation, accommodation, food and drinks, different services, culture and entertainment, conventions and trade shows, sports, etc. In addition, there are some activities that are critical success factors of the tourism, which include: financial services, telecommunication, health, and other services, such as electricity, water, and life security.

All these extensive activities come together in the production and consumption of tourism, starting with accommodation in hotels as a major tourism activity and continuing further with suppliers of inputs that are necessary for immediate consumption, such as meat and fish, dairy products, vegetables, and beverages. Tourism also establishes relationships between the construction companies and the equipment and uniforms manufacturers \(^3\). There are other important links to tourism hotels apart, and production of goods or services used by tourists, such as handicrafts, shopping, musical performances, health treatments and the employment of people working as guides. Most developed countries have estimated that the tourism economy generates jobs and income almost two times more than other similar sectors in developing countries \(^4\).

**Relevance of tourists inflow forecasting**

Tourism is seen as an economic activity that generates both positive and negative impacts, but sustainable tourism aims to achieve the best balance between economic benefits and environmental and social costs \(^4\). A sustainable tourism should make an optimum use of resources, minimize the environmental, cultural and social impacts, as well as maximize the benefits of preserving the communities \(^5\).

Forecasting the tourist inflow is of great importance for the tourism operators and managers, because by knowing the tendency of tourist inflow, is done the planning of this sector regarding the accommodation and the transport. Even though in economic terms, it is desirable to have as many tourists as possible, the cities are not always ready and prepared to accommodate ten times more visitors per day as local residents during the summer \(^4\).

Forecasting of tourist inflow is of great importance in the forecasting the expected revenues to be realized from this activity, and the impact that these revenues have on the GDP.

With the tourists inflow forecast is related also the loan guarantee that banks can predict for providing short-and long-term investments in this sector, as well as helping the tourism agencies in the preparation of joint tourist guides.
Also, the prediction of tourist inflow empowers the local governments to organize various events with the tourists. Tourists’ activities tend to make tourists live in a more interesting and exciting place.

The more developed countries have estimated that the tourism economy generates jobs and income almost two times more compared to other similar sectors in developing countries.

Different statistical methods are generally used for forecasting the tourists’ inflow, but their use encounters difficulties as they are time-consuming, this because they use the factors affecting the number of tourists, are expensive, and they do not learn from past data. In this article we present an intelligent technique such as neural networks for forecasting the number of tourists.

**Artificial neural networks overview**

Artificial Neural Networks (ANNs) provide powerful models for statistical data analysis. Their most prominent feature is their ability to “learn” dependencies based on a finite number of observations [6].

ANN is a mathematical model that performs a computational simulation of the behavior of its biologic counter-part. Neural networks have been used in many business applications for pattern recognition, forecasting, prediction, and classification, due to their ability to “learn” from the data, their nonparametric nature and their ability to generalize [7]. The most typical application areas of neural networks include finance, marketing, manufacturing, operations, information systems, and so on.

The first step in developing an ANN comprises the definition of the network architecture, which is defined by the basic processing elements (i.e. neurons) and by the way in which they are interconnected (i.e. layers).

The second step deals with the definition of the NN Learning, which implies that a processing unit is capable of changing its input or output behavior as a result of changes in the environment, i.e. to adjust the weights based on input vector values.

The third step finalizes the definition of the data used for training, testing and validating the neural network. These topics will be described in more details in the following sections.

**Anns’ architecture**

An ANN may have one or more layers of neurons, which may be fully or partially connected. Each link between two nodes has a weight \( w_{ij} \) which summarizes the knowledge of the system. The processing of the existing cases with the inputs: \( x_1, x_2, ... \)
$x_j$ and the expected results, will adjust these weights based on the difference between the actual and the expected results.

The input layer nodes are passive, doing nothing but relaying the values from their single input to multiple outputs, while the hidden layer nodes and output layer nodes are active nodes, as depicted in Figure 1 \cite{8}.

\[ \sum_{i=1}^{j} W_{ij} x_j \]

\[ f(x) = \frac{2}{1 + e^{-\alpha x}} - 1 \]

We tested the model with different $\alpha$ values, and found that the best value in this case is $\alpha=1$.

The most common structure for a neural network is three layers with full interconnection.

In this study we use the programming language C# for the design and testing of ANN models. The ANN architecture used in this paper is shown in Figure 2.
The number of neurons in input layers depends on the independent variables, whereas the number in the output layers on the dependent variable. In this paper, the number of inputs is 6 (six neurons in input layer), meaning the number of tourists in six months, the hidden layer also has six neurons in our model, and the output is 1, meaning the number of tourists of the consecutive month.

The structure of the ANN is an important factor for correct training. A correct amount of hidden layers and nodes should be used. If more are used than those are required to learn the input-output relationship, there will be more weights than necessary, which can lead to over fitting of the data and bad performance when approximating unknown cases.

**ANNs learning**

Learning is the major issue for most of the neural networks’ architectures. This makes the selection of a learning algorithm the major issue in the development of the network. Learning implies that a processing unit is capable of changing its input or output behavior because of the changes in the environment, i.e. to adjust the weights corresponding to the input vector \[^9\]. The whole process of the ANN learning has three continuous steps:

1. Calculate the interim results.
2. Compare the results with the desired objectives.
3. Adjustment of the weights and then start over the process till the required accuracy is achieved or another termination criterion is reached.

In this paper two neural networks have been implemented: the multilayer perceptron (MLP) neural networks trained by genetic algorithm (GA), and the MLP neural network trained by back-propagation (BP), both using bipolar sigmoid activation function.

Although BP - a gradient descending method - achieves very good performance in ANN training, we often use GA algorithm in the learning process of the ANN instead of it, in order to avoid wasting time and falling in a trap such as local minima. In the ANN training, the GA starts with a large population of vectors, called individuals, which are generated randomly. Each individual of the population represents a possible solution of the problem that is being analyzed. The individuals are selected to create new individuals (offspring) based on some selection criteria such as their fitness, the more adequate (fitting) they are, the more the chances they have for reproduction, and through the genetic operators it searches for the improved solutions.

**ANNs’ data sets**

In order to verify the rationality of the model and the algorithm, this paper analyzes data about Albania. The sample interval involves monthly data about tourists’ inflow
from January 2005 till June 2013, presented in Table 1. In accordance with the needs of the model, the dataset is divided into two parts, i.e., training and test set. The monthly data (99 months) about tourists’ inflow from January 2005 to March 2013 belongs to the training set, whilst monthly data about tourists inflow in the interval April – June 2013 (3 months) constitute the test set. There are no official public data regarding the number of tourists from July 2013 till now, this can be due to the political changes and the restructuring of the responsible institutions. These data was stored in a .csv file and served as input to the neural network model.

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Month</th>
<th>Tourists number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>January</td>
<td>32.321 37.630 49.437 55.635 58.640 68.997 90.712 113.107 124.506</td>
</tr>
<tr>
<td>2</td>
<td>February</td>
<td>25.214 33.107 41.173 52.284 52.787 55.181 79.189 67.083 103.089</td>
</tr>
<tr>
<td>3</td>
<td>March</td>
<td>33.471 41.345 53.532 68.942 62.933 72.864 106.649 12.921 145.711</td>
</tr>
<tr>
<td>4</td>
<td>April</td>
<td>41.891 58.723 70.040 76.867 98.735 90.139 145.171 17.387 211.761</td>
</tr>
<tr>
<td>5</td>
<td>May</td>
<td>49.148 66.696 77.577 97.995 113.362 117.261 146.114 188.174 219.468</td>
</tr>
<tr>
<td>6</td>
<td>June</td>
<td>62.201 73.402 93.533 123.249 135.740 144.663 196.704 249.004 254.921</td>
</tr>
<tr>
<td>7</td>
<td>July</td>
<td>139.446 173.124 201.428 251.926 469.109 565.262 670.713 767.246</td>
</tr>
<tr>
<td>8</td>
<td>August</td>
<td>163.681 197.976 215.120 257.185 411.252 711.546 684.886 957.427</td>
</tr>
<tr>
<td>9</td>
<td>September</td>
<td>70.805 82.731 95.148 101.461 119.358 158.545 200.486 305.125</td>
</tr>
<tr>
<td>10</td>
<td>October</td>
<td>49.286 57.181 75.109 91.319 100.783 128.452 149.917 180.786</td>
</tr>
<tr>
<td>11</td>
<td>November</td>
<td>39.840 45.227 54.994 65.691 73.165 95.288 122.986 176.807</td>
</tr>
<tr>
<td>12</td>
<td>December</td>
<td>44.533 58.914 71.674 87.584 90.181 94.701 140.123 128.397</td>
</tr>
</tbody>
</table>

Table 1. Monthly statistical data about tourist inflow. Source: MTRS Albania

Results

In this paper two neural networks have been implemented: the multilayer perceptron (MLP) neural networks trained by a genetic algorithm (GA) – referred as neuro-genetic model - and the MLP neural network trained by back-propagation (BP), both using bipolar sigmoid activation function.

The main parameters of the GA used in neuro-genetic model were set as:

- Genetic population Size = 100
- Crossover probability in genetic population = 0.7
- Mutation probability in genetic population= 0.3
- Probability to add newly generated chromosome to population = 0.25

The results of this model are shown in Figure 3.
The main parameters of the BP used in the BPNN model were set as:
- Learning rate=0.3
- Momentum=0.2

The results of this model are shown in Figure 4.

We compare BPNN and the neuro-genetic model (NN trained by GA) with the Exponential Smoothing model to verify their accuracy and effectiveness.

Details about the comparison are shown in Figure 5. Standards for evaluation on each model’s advantages and disadvantages is mainly realized by $R$ the correlation coefficient of the training set and the test set, MAD and average error (especially MAD and the correlation coefficient $R$ of the test set). The $R$ stands for degree of relevance between the predicted value and the actual value, and MAD means the degree of deviation between predicted value and actual value. The closer the correlation coefficient $R$ is to 1, the higher the degree of relevance between predicted value and actual value will be; the lower the value of MAD is, the smaller the deviation between predicted value and actual value will be. These measures of forecasting accuracy are shown in Table 2. The predicted value obtained by the neuro-genetic model is the most accurate.
Figure 4. Predicted Values of neuro-Genetic Model

Figure 5. Comparison of forecasted values of each month by different methods

<table>
<thead>
<tr>
<th></th>
<th>BPNN</th>
<th>Exponential Smooth</th>
<th>Neuro-Genetic</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.998</td>
<td>0.614</td>
<td>0.999</td>
</tr>
<tr>
<td>TS</td>
<td>-3.7</td>
<td>2.72</td>
<td>-0.73</td>
</tr>
<tr>
<td>MAD</td>
<td>6935</td>
<td>71622</td>
<td>5512</td>
</tr>
</tbody>
</table>

Table 2. Measures of forecasting accuracy
Conclusion

- Tourism is not a simple activity; it comprises many similar and distinctive activities which include: transportation, accommodation, food, services, beverages, cultural entertainment, conventions, and trade shows.
- Using statistical models for the analysis of tourists inflow, we conclude that the trend of tourists’ number in the future will be increasing and that the tourism phenomenon in Albania represents a constant and stable seasonality, both in numerical terms, as well as from the time aspect.
- The ANN models forecasted tourist flow in this study resulted very accurate, as indicated by the correlation coefficient $R_{BPNN} = 0.998$ and $R_{neuro-genetic} = 0.999$.
- The models tend to slightly over-forecast, with an average absolute error between 5512 units for neuro-genetic model and 6935 units for BPNN.
- As long as the tracking signal (TS) is between – 4 and 4, in our models TS is equal to -3.7 for BPNN and -0.73 for neuro-genetic model, we can say that the model is working correctly. In general, ANN models require large amounts of data.
- Taking into consideration that features like seasonality, nonlinear nature, and small sample characterize the data about tourists’ inflow, NN provides an effective and new method for prediction of tourists’ inflow. The NN model doesn’t need to know anything else about the tourism industry other than the time series of tourists’ inflow.
- Forecasting the number of tourists is quite important for planning the development of the whole tourism industry.

Bibliography

2. The Authority on World Travel & Tourism, “Travel & Tourism 2011”
6. Antanasijević, Davor; Pocajt, Viktor; Popović, Ivanka; Redžić, Nebojša; Ristić, Mirjana: The forecasting of municipal waste generation using artificial neural networks and sustainability indicators, Sustainability Science; Jan 2013, Vol. 8 Issue 1, p37

7. Priyanka Gaur, Neural Networks in Data Mining, International Journal of Electronics and Computer Science Engineering, ISSN- 2277-1956, Volume 1, Number 3 (p 1449- p1453)


9. Alok Bhaskar Nakate, Neural Networks Non-Linear Scaling, California State University, 2011


