Abstract No: SI-03

Tourist volume forecasting: An approach with supervised machine learning algorithms

B. R. P. M. Basnayake* and N. V. Chandrasekara

Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka
pavithramalkibasnayake@gmail.com*

The tourism industry generates almost US$4 billion of income and provides direct and indirect employment to a large number of people in the country. Expert knowledge on the travel behaviour of tourists is an important part of planning and aids decision making for all stakeholders including the government and private business organisations. There was a severe drop in tourist arrivals during the civil war and was also apparent after the more recent Easter Sunday bomb attack. The study compared the predictions of different forecasting models on tourist arrivals in Sri Lanka, in an effort to identify the most appropriate model. The supervised machine learning algorithms (MLA) applied were Time Delay Neural Network (TDNN) and Feedforward Neural Network (FFNN) with two different algorithms namely Levenberg-Marquardt (LM) and Scaled Conjugate Gradient (SCG). Recently, MLA has started playing a vital role as an effective forecasting tool. A better model in forecasting was identified using the performance criteria of the Normalized Mean Squared Error (NMSE). As an initial step, monthly data from December 2019 to January 2000 were standardized to maintain the consistency of data. The aforementioned models were trained for 100, 200 and 500 epochs separately, with different numbers of hidden layers and hidden neurons, and detected the model with minimum NMSE for further training. For the selected model from TDNN, subsequently, the transfer functions and time delays were modified. A better model was identified in 500 epochs for the network with 2 hidden layers of 4 and 3 hidden neurons with tansig transfer functions from time delay of 3 (NMSE 0.3537). For the FFNN model, input combinations were recognized using the Pearson correlation coefficient and Spearman's rank correlation coefficient. Among the trained models with the different input combinations, the model with MA3, MA6, MA9, MA12, and MA15, lag 1, lag 2, lag 3, lag 11 and lag 12 indicated the lower NMSE of 0.5244 where Moving Average (MA) indicates current and past values and depends linearly on the output variable and lags being predetermined fixed quantity of passing time. For the FFNN, a better model was identified with the adjustment of parameters. A better model was identified in 100 epochs for the network with 3 hidden layers of 3 hidden neurons in each layer with tansig transfer functions, a learning rate (η) of 0.01, a combination coefficient (μ) of 0.001 and a decreasing factor as 0.1 and increasing factor as 10 of μ (NMSE 0.2234). For the SCG algorithm, the lowest performance measurement value, NMSE was 0.3193. The model had 500 epochs with 3 hidden layers of 3, 2 and 2 hidden neurons respectively, transfer functions with tansig in all hidden layers, a sigma parameter value of 5e (-5) and a lambda of 5e (-7). The main conclusion is that all the discussed network models capture the actual behaviour of the testing set while the minimum NMSE was identified in the FFNN with the LM algorithm. The findings of the analysis are beneficial, as tourism is a global service industry and a source of foreign exchange earnings and a key employment generation sector for the country.

Keywords: Feedforward neural network, Machine learning algorithms, Normalized mean squared error, Time delay neural network, Tourism