Prediction of passenger flow for north central railway region through ANN

Anoop Pratap Singh¹, Ajay Tripathi², Ravi Kumar Dwivedi¹, Anurag Garg³ and Rajan Kumar¹

¹Mechanical Engineering Department, Maulana Azad National Institute of Technology Bhopal, India
²Mechanical Engineering Department, Government Engineering College, Raipur, India
³Department of Electronics & Communication Engineering, IPS College of Technology & Management, Gwalior, India

E-mail: anoop2651@gmail.com

Abstract. A new method of prediction method is described in this paper. The passenger rate for the north central railway (NCR) region is estimated by using artificial neural networks (ANN). An ANN model is developed here that can logically estimate passenger flow rates. Which helps the decision makers to make the strategies according to the falling population. In this analysis data from the North Central Railway region from January 2009 to December 2015 have been taken, data such as passenger revenue, months and years, festival seasons and passenger numbers. When predicting passenger flows for the months of January and February 2016, an error of less than 2.9% is found. Therefore, it’s concluded that an ANN prediction method is applied in passenger flow prediction in railways.

Keywords: Passenger flow prediction, back propagation (BP), neural network, data analysis

1. Introduction

The success of any firm depends on good predictability of strategies. Indian Railways deals with large amounts of data such as passenger details, baggage details, commercial cargo details, train travel information and many more on a daily basis. Quality of service in any service providing firm depends directly on the number of customers or passengers in the case of railways. If railway data can be converted into valuable inputs of any forecasting model and predict passenger flow, it will help to strategize according to falling customers. Service providers are more efficient and reliable by the use of predictive models in the formulation of strategies [1][2].

1.1. Artificial Neural Network (ANN)

1.1.1. Neurons

The ANN works on the functioning of the human brain. It consists of several numbers of neurons. Neurons are interconnected by different values of weight, which perform the functions of the human brain [3]. It can formulate the guidelines by learning from known results and make nonlinear and complex laws for further use. Neurons of ANN have the strong ability of nonlinear mapping which
helps to work with the complex problems [4][5]. ANN is able to work with the linear and nonlinear problem functions.

1.1.2. Structure of Neural Network

BP Network is layered structure of the network. Each layer consists of several neurons. The several layer feed-forward network learning errors is based on the back-propagation algorithm [6]. Structure of neural networks provides the information about the neurons and layers counting in the ANN models. A simplest form of ANN model is having minimum two layers named input layer and output layer. Neurons in the input layer is equal to the inputs and neurons in the output layer is equal to output. More complex problems and functions deals by the multilayer ANN model. Figure 1, shows the connections of nodes in the ANN model.

Figure 1. The Artificial Neural Network Architecture

Minimum 3 layers are needed by a Back-Propagation network. The number of layers is increases as the complexity of the function increases. First layer on a structure is an input layer and one is output layer m-between a hidden layer is also present. Hidden layers varies as the complexity of the desired problem function varies [7].

Neurons become active when the input data is sampled to the BP network. Data is entered in the ANN model through the input layer neurons and transferred to the output layer neurons through hidden layers neurons [8]. The neurons are interconnected with the help of the different synaptic weights. As signal moves from input to output layer this is termed as the forward action. At this time the error occurred at the output layer transferred to back and updated the weights added with the neurons. The correction of error travels from the output layers to the input layer through the hidden layer. This may decrease the output response first but this propagation of error updates the weight and helps to reduce the final error of the model. This whole process is termed as the training of the network.

2. Modeling with Neural Networks

2.1. Research Data

Time in months with years, ticket revenue and festival season, are taken as the input parameter for the ANN. Number of passengers are the output parameter for the ANN [9]. This data is used for the training of the ANN. After the collection of the data it is seen that the data desperation is on a long range like data of revenue and number of passengers in lakhs and years and months are in ones and thousands. Therefore, there is an essential need of data processing.
2.2. Data Set and Processing

After analyzing the input data, it is seen that the months data ranges from January to December and years data ranges from 2009 to 2015 and revenues are in lakhs. In the case of ANN input parameters, the higher value shows the greater impact on the output as compare to the smaller values which is an undesirable situation. To overcome this situation all the training data should be converted in the range of [0-1]. There are multiple normalization techniques available for this, but in this case the formula mentioned below used for the normalizing the input and output data

$$\bar{x} = \frac{(x - x_{\text{min}})}{(x_{\text{max}} - x_{\text{min}})}$$  \hspace{1cm} (1)

Here,

$\bar{x}$ - Normalized data

$x_{\text{min}}$ - Minimum value for the single type of data

$x_{\text{max}}$ - Maximum value for the single type of data

2.3. Experimental Scheme

Observation of the input data shows there is increase of passengers on the festival seasons of Holi and Diwali. This is incorporated in the ANN by adding an extra input “Festive Seasons”.

Table 1. Sample Data Sets

| Set No. | 1 | 2 | 3 | . | . | 82 | 83 | 84 |
|---------|---|---|---|---|---|----|----|----|
| Month   | January | February | March | . | . | October | November | December |
| Year    | 2009 | 2009 | 2009 | . | . | 2015 | 2015 | 2015 |
| Festive Season | 0 | 0 | 1 | . | . | 0 | 1 | 0 |
| Revenue ($\times 10^5$ Rs.) | 7471 | 7920 | 8300 | . | . | 14322 | 14894 | 16790 |
| Number of Passengers ($\times 10^4$) | 1255 | 1380 | 1550 | . | . | 1387 | 1408 | 1362 |

Table 1 shows the input samples for the training of an ANN. In this case 84 sets of data are used to train the neural network. First Month names is given by the numbers from 1 to 12 for the ease of mathematical input parameter. Then processing is done on the data mentioned on Table 1 and converted all the data in the set of [0-1] by using the formula 1. This processed data is analyzed with the help of Neural Network Tool Box of MATLAB 15. And a optimizes neural network model is generated.

ANN performance is highly dependent on the architecture of the Neural network. Number of the hidden layers have a great impact on the results of the neural model [10]. If there is a greater number of hidden layers them it also increases the error as neural model is overtrained. If hidden layers are lesser as required then it unable to performs the desired function of ANN. So, the optimum number of the hidden layer is a mandatory requirement of any ANN.
Number nodes in the ANN or neurons is again an important parameter which influence the response of the neural network. If the number of neurons is more than required it will increase the value of error and if the number of neurons is less than required then it unable to performs its functionality [11]. So Optimum number of nodes is again a prime importance of an ANN model.

3. Results
This ANN model consists of four neurons in input layer and one neuron on the output layer with respect to four inputs and one output. This neural network models works on the principle of Back Propagation. So, the number of neurons in the hidden layer may varies. In this case the optimum number of neurons is checked in between 10 to 25 for the minimum value of error. Based on the analysis 20 nodes on the hidden layer is the optimum number of neurons with least error of 0.004181. Errors with respect to number of neurons on the neural network is mentioned in Table 2.

Table 2. ANN Training Error

| Number of Neurons | Network Error |
|-------------------|--------------|
| 16                | 0.009211     |
| 18                | 0.007737     |
| 20                | 0.004181     |
| 22                | 0.019486     |
| 24                | 0.018565     |

Figure 2 shows the performance graph of ANN. Total number of epochs are done on the analysis is 33. The minimum value of error 0.0041809 is the least error (Taking all three sets under consideration). Training error is continuous decreasing with the number of epochs increasing this show the over trained nature of ANN. Validation and test error show an increasing nature with epoch. So, the minimum error with respect to validation and test data is taken for the minimum value of error for ANN.
Figure 2. Performance Graph

Figure 3. Regression Graph
Figure 3 shows regression graphs, in these fit line shows the actual behaviour of ANN and circles shows the data points for different sections. First, second and third graph of figure 3 shows the regression graphs for training, validation and test data sets. Fourth graph shows the overall regression graph for ANN with the R value 0.91046 which is closer to 1 as desired. This represents a better fit for ANN model. Table 3 shows the final architecture of the ANN.

**Table 3. Final Structure of the Network**

| Structure of the Network | Training Function |
|--------------------------|-------------------|
| 4X20X1                   | trainlm           |

| Neurons of Hidden Layer | Network Least Mean Square Error |
|-------------------------|--------------------------------|
| 20                      | 0.0041809               |

Table 4 shows prediction from the designed ANN is very close to the actual values. The percentage of error is less than 2.9%. Therefore, it is said that ANN is successfully and efficiently applied to the railway passenger flow problem.

**Table 4. Predicted value**

| Year | Month | Number of Passenger Information from Railway Site | Predicted Number of Passengers | % Variation |
|------|-------|--------------------------------------------------|--------------------------------|-------------|
| 2016 | January | 12700000                                         | 12709779.6                    | -0.077004724 |
|      | February | 13620000                                         | 13225858.4                    | 2.893844347  |

4. Conclusions
In this work, an artificial neural network model has been successfully applied for forecasting. Passenger flow for the next two months are predicted with the help of trained ANN models and it is found that there is an error of less than 3% of the actual data. Which is a clear indication of the successful implementation of ANN on the passenger flow prediction problem. This model helps strategic manufacturers in a very effective way. An ANN has the ability to deal with complex problems as well as problems that cannot be derived into easy functions.

Acknowledgement
The author is thankful to Indian Railways, for providing the information regarding number and revenue generated from passengers through its website.

References
[1] Liu L and Chen R C 2017 A Mrt daily passenger flow prediction model with different combinations of influential factors *Proc. - 31st IEEE Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2017* 601–5
[2] Mullins R J 2001 Application of An Artificial Neural Network on Railway Passenger Flow Prediction *J. Trauma* 50 385–6
[3] Xu W, Huang H K and Qin Y 2004 Prediction of railway passenger flow based on temporal data mining Proc. 2004 Int. Conf. Mach. Learn. Cybern. 3 1550–4

[4] Mehrpouya M, Gisario A, Nematollahi M, Rahimzadeh A, Baghbaderani K S and Elahinia M 2021 The prediction model for additively manufacturing of NiTiHf high-temperature shape memory alloy Mater. Today Commun. 26

[5] Song J, Romero C E, Yao Z and He B 2017 A globally enhanced general regression neural network for on-line multiple emissions prediction of utility boiler Knowledge-Based Syst. 118 4–14

[6] Zhao Y, Nan J, Cui F Y and Guo L 2007 Water quality forecast through application of BP neural network at Yuqiao reservoir J. Zhejiang Univ. Sci. A 8 1482–7

[7] Agarwal S 2014 Data mining: Data mining concepts and techniques

[8] Chan H L and Pang K W 2011 Association Rule Based Approach for Improving Operation Efficiency in a Randomized Warehouse Methodology 704–10

[9] Zheng D and Wang Y 2011 Application of an artificial neural network on railway passenger flow prediction Proc. 2011 Int. Conf. Electron. Mech. Eng. Inf. Technol. EMEIT 2011 149–52

[10] Hart M and Porter G 2004 The impact of cognitive and other factors on the perceived usefulness of OLAP J. Comput. Inf. Syst. 45 47–56

[11] Jaswal K, Kumar P and Rawat S 2015 Design and development of a prototype application for intrusion detection using data mining 2015 4th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2015