Bus Passenger Flow Prediction and Division Based on BP Neural Network

Haoyu Wang[1, *], Chengfei Yang[2] and Hui Zhou[3]

[1]School of Automotive Engineering, Wuhan University of Technology, Wuhan430070, China.

[2]School of Automation, Wuhan University of Technology, Wuhan 430070, China.

[3]School of Information Engineering, Wuhan University of Technology, Wuhan 430070, China.

*Corresponding author: why12312311@163.com

Keywords: Traffic Flow Prediction, Bus Dispatching, Bp neural network, Fisher Clustering Method.

Abstract: Bus dispatching is a crucial issue in bus operation, and a suitable dispatching method can significantly save operating costs. This paper proposes a passenger flow prediction and division method based on BP neural network and Fisher optimal dissection method for the dynamic dispatch of buses. Based on this method, the bus dispatcher can increase or decrease the number of buses before peak traffic according to the results of prediction and division to achieve economic efficiency and other purposes. This paper uses the passenger flow data of a certain line in Guangzhou for a month to train and test the network and obtain good results.

1. Introduction

Buses are an essential part of public transportation. In order to better serve the public, bus dispatching is a significant part of bus operation. A good dispatch plan can ensure that the best economic benefits can be obtained under the premise of meeting the demand for passenger flow. The passenger flow of each line of the bus changes with time, and the bus dispatch should also be dynamic. However, because it takes time for the bus from the departure station to reach the demand station, the bus dispatch is lagging. Therefore, the prediction of passenger flow plays an essential role in the dynamic dispatch of buses. This paper proposes a feasible scheme for predicting the passenger flow of bus lines.

2. Back Propagation Neutral Network

2.1 The structure of Back Propagation neural network

Back Propagation (BP) neural network is a multilayer feed-forward neural network trained by the Error Back Propagation algorithm. It has arbitrary complex pattern classification ability and excellent multidimensional function mapping ability. It is widely used in data compression, function approximation, pattern recognition, and classification[1]. In structure, the BP network has an input layer, and it has multidimensional function mapping ability. The network topology of the implied and output layers is shown in Figure.

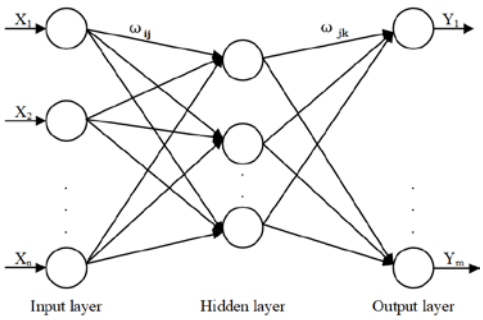

Figure 1. Structure of BP neural network.

Where, $X_1,X_2,...X_n$ are the input parameters of the BP neural network, ω_{ij} are the neural network weight parameters from the input layer to the hidden layer; ω_{jk} is the neural network weights from the hidden layer to the output layer Parameters; $Y_1,Y_2,...Y_m$ are the output parameters of the BP neural network, which express the function mapping relationship from n independent variables to m dependent variables.

2.2 Learning procedure of BP neural network

The learning process of the BP neural network is divided into two parts: forward transmission of information and reverse transmission of error. The input layer transmits the received information to the neurons in the middle layer and then performs information processing and exchange. After the information passes through all hidden layers, Finally, the output layer outputs the result. If the output value does not match the expectation, the error is reversed, and the weights and thresholds of each layer are corrected in the way of error gradient descent. The learning and training are improved until the error is acceptable or the predetermined number of learning times is completed.[2]

Assume the BP neural network has L layers, the number of neurons in each layer is n_L, and the actual output value is y_{ij}, then for a given N sample (x_k,y_k), the system error after training is

$$E_k = \frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{nL}\left(y_{ij} - \bar{y}_{ij}\right)^2 \tag{1}$$

The total error is

$$E = \frac{1}{2N}\sum_{k=1}^{N} E_k \tag{2}$$

The correction weight is

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \mu\frac{\partial E_i}{\partial \omega_{ij}} \tag{3}$$

Where t is the number of iterations, μ is the learning efficiency.

2.3 Passenger flow prediction model based on BP neural network

(1) Data preprocessing

Since the input variables of the neural network need to have the ability to extract and have a significant impact on the variables, the initial sample data must be preprocessed before the network can be trained. Besides, since the weight range is in the interval [0,1], the sample preprocessing is mainly to limit the data in the interval [0,1] for subsequent calculations. The normalization processing formula is as follow

$$y = \frac{(y_{\max}-y_{\min})(x-x_{\min})}{x_{\max}-x_{\min}} + y_{\min} \tag{4}$$

Where x is the data set which needs to be normalized, $y_{\max}$=1, $y_{\min}$=-1

(2) Selection of Factors Affecting Bus Passenger Flow

In order to obtain a better forecasting effect, it is necessary to clarify the main influencing factors of the forecasting model. According to the distribution map of bus passenger flow, time interval and working day are the main factors affecting passenger flow. Besides, weather, temperature, holidays, and road conditions will also affect citizens' travel. The above influencing factors are divided into two categories: quantifiable influencing factors and non-quantifiable influencing factors. For these two types of influencing factors, we directly use quantitative values for the former, such as the time is at 6 o'clock, 12 o'clock, the temperature is at 16°C, etc. We use the method of rating for the latter. For example, the weather can be divided into sunny, cloudy, light rain, etc, and the values are 1, 2, 3, etc, respectively. Exceptional circumstances such as sunny to cloudy can be assigned a value of 1.5. In

addition, although road conditions and seasons will also interfere with passenger flow, the degree of impact is relatively small compared to other parameters, and it is not the main factor affecting bus passenger flow. Considering the statistical difficulty of information and the characteristics of passenger flow changes, the influence factors selected in this paper are time period, holidays, weather conditions, and temperature

(3) Construction of BP neural network

1) Selection of Hidden layer

Experience has shown that for continuous functions if the independent variables are in the closed interval, they can all be approximated by a single hidden layer BP network[3]. The increase in the number of hidden layers will increase the system's ability to deal with complex problems but will also reduce the data processing speed and lengthen the training time. Therefore, this paper selects a single hidden layer multiple input single output BP network to establish a prediction model

2) Selection of Hidden layer

Hidden layer nodes are the core of data processing and storage. Too many nodes will increase the amount of network calculations, and the network is easy to overfit. When the number of nodes is too few, the learning effect of the network will decrease, and the number of training will increase accordingly. At present, there is no precise formula for determining the number of hidden layer nodes. This article refers to the following empirical formula to select the number of hidden layer nodes. After determining the approximate range, perform multiple training comparisons and then select the most appropriate number of nodes [4].

$$l = \sqrt{n + m} + a \tag{5}$$

where n and m are the numbers of nodes in the input layer and output layer, and a is a constant in the interval [0,10].

3) Selection ofTransfer function

BP network has a variety of transfer functions. The Log-sig (s-type) transfer function is most used in single hidden layer networks. The input value can be selected arbitrarily, and the output value is in the interval [0,1]; tan-sig (The input value of the hyperbolic tangent (S-type) transfer function can be arbitrarily selected, and the output value is in the interval [-1,1]; the input and output of the linear transfer function ‘purelin’ can be arbitrarily selected. In addition, there are positive linear transfer function ‘poslin’, radial basis transfer function ‘radbas’, and so on. The multi-layer network can select different transfer functions to obtain better prediction results

(4) Selection of learning rate

The selection of the initial value of the weight seriously affects the training speed of the BP neural network. If it is selected improperly, it will cause the network to fall into the Local optimal interval, and the convergence speed will be slow. The choice of learning rate affects the time and stability of neural network training. This article selects the random number with the initial value (0,1), generally selects the learning rate between 0.01-0.8, compares the test results, and then determines the appropriate parameters.

(5) Passenger flow prediction

Use the sample data to train the BP neural network, and predict the passenger flow situation in the future time period by inputting the environmental parameter values at the current moment.

3 Fisher Optimal Dissection Method

3.1 Procedures of the Fisher optimal dissection method

The Fisher optimal dissection method is one of the methods of cluster analysis. In general clustering, the classification is usually based on the similarity coefficient of the samples, which will cause the samples to lose their original arrangement order. However, in some practical problems, the original arrangement order of the original samples cannot be disrupted. The samples of the same type must be adjacent to each other. If there are n ordered samples $\boldsymbol{x_1}, \boldsymbol{x_2}, \dots, \boldsymbol{x_n}$, the form of each category

must be $\{x_i, x_{i+1}, \ldots, x_j\}, i < j$, this classification method is called the Fisher optimal dissection method (Fisher algorithm)[5]. Let n ordered samples be$x_1, x_2, \ldots, x_n$,the procedures of the ordered sample clustering method are as follows:

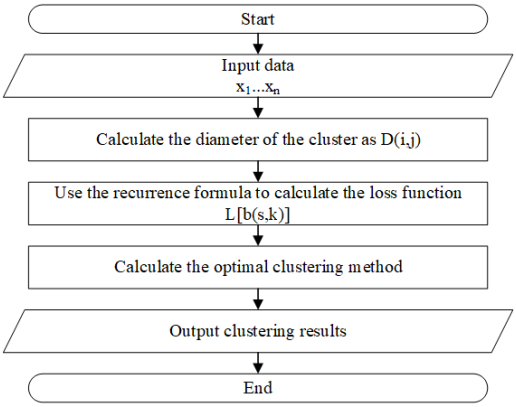

Figure 2. Flow chart of Fisher optimal dissection method

3.2 Passenger flow division model based on Fisher optimal dissection method

Due to the uncertainty of the length of each bus line and the number of stops, peaks and flat peaks cannot be divided by a fixed threshold. The problem of dividing the daily passenger flow is to classify and merge adjacent time periods with similar passenger flow according to certain rules and obtain different passenger flow intervals. Use the Fisher optimal dissection method to classify and build a model based on the passenger flow data at various times of the day according to the following steps:

1) Convert the passenger flow data of each period to its proportion in the total passenger flow of the day, as s ordered samples;

2) Select the number of clusters and calculate the diameter of all possible classes.

Use the sum of squares of the difference between the value of the class and the class mean, that is, the sum of squared deviations, to express the size of the diameter[6]

$$D(i,j) = \sum_{t=i}^{j} (x_t - \bar{x})^2 \tag{6}$$

where, if a certain type of G contains samples $\{x_i, x_{i+1}, \ldots, x_j\}, i < j$, then the mean is:

$$\bar{x} = \frac{1}{j-x+1} \sum_{t=i}^{j} x_t \tag{7}$$

3) Calculate the minimum loss function $L(b_{s,k}^*)$

Assume $b_{s,k}: \{i_1, i_1+1, i_2-1\}, \{i_2, i_2+1, i_3-1\} \ldots, \{i_k, i_k+1, s\}, (i_1 = 1 < i_2 < i_k < n)$ represents a classification method that divides s samples into k categories, then its loss function is

$$L(b_{s,k}) = \sum_{j=1}^{k} D(i_j, i_{j+1} - 1), i_{k+1} = s + 1 \tag{8}$$

The smaller the loss function value, the more reasonable the classification method. If the optimal division method for minimizing the loss, function is $b_{s,k}^*$, the recursive formula of Fisher's algorithm can be used to obtain:

$$\begin{cases} L(b_{s,2}^*) = \min\limits_{2 \le j \le s} \{D(1, j-1) + D(j, s)\} \\ L(b_{s,k}^*) = \min\limits_{k \le j \le s} \{(b_{j-1,k-1}^*) + D(j, s)\} \end{cases} \tag{9}$$

After all the procedures above, we can obtain the best solution$b_{s,k}^*$.

4) Compare the minimum loss function values corresponding to the number of different categories, comprehensively consider the convenience of scheduling, determine the number of clustering intervals, and then get the result.

4 Experiment and Analysis

4.1 Data set and parameters

(1) Data Set

This paper uses December passenger flow data of a Guangzhou bus to verify the role of the model (observation time is from December 1st to 31st all day, as shown in Figure.3). The data includes the date, weather conditions, temperature, holiday conditions, time period.

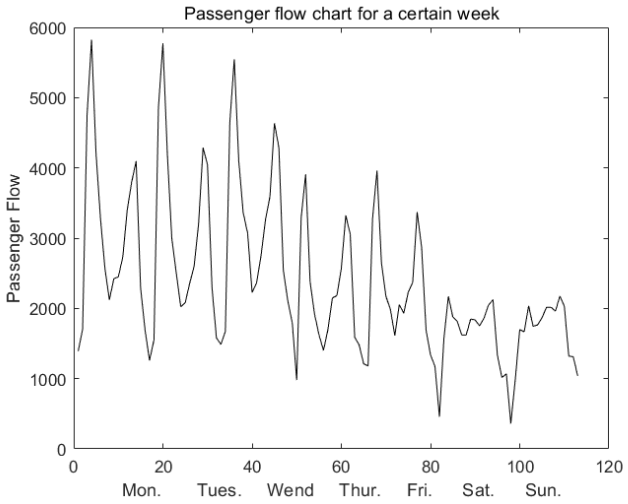

Figure 3. Passenger flow chart for a certain week

(2) Parameters

Use the ‘newff’ function in MATLAB to create a BP neural network. The number of neural network layers is 3. After multiple tests and comparisons, the numbers of neurons in each layer are set to 4, 7, and 1, and the transfer functions of the three-layer network are selected as ‘ansig’, ‘tansig’, and ‘purelin’, respectively. The training function is‘trainlm’.

After the parameters are set, randomly divide the data set into the training set and testing set. The training set is used to train the network; thus, the trained network could be used to predict the data of the test set. Use the Fisher optimal dissection method to cluster the predicted data, and compare the result with the original data.

4.2 Result

The prediction results of the neural network are shown in Figure 4:

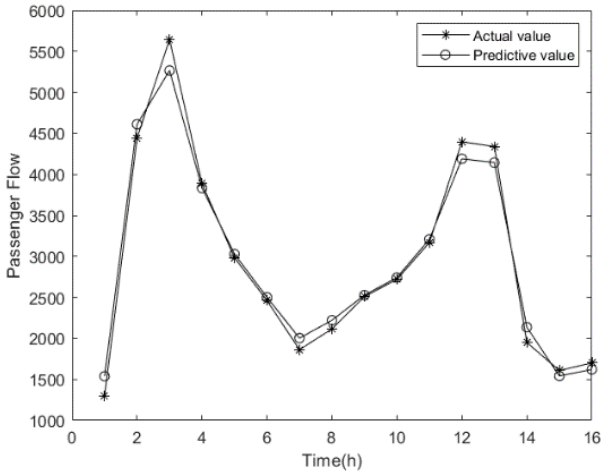

Figure 4. Result of BP neural network prediction

According to the figure above, it could be seen that the BP algorithm prediction curve trend is basically the same as the actual passenger flow curve trend, and the data difference is tiny. Based on the prediction, we can divide the passenger flow with the Fisher optimal dissection method.Following is the result of clustering:

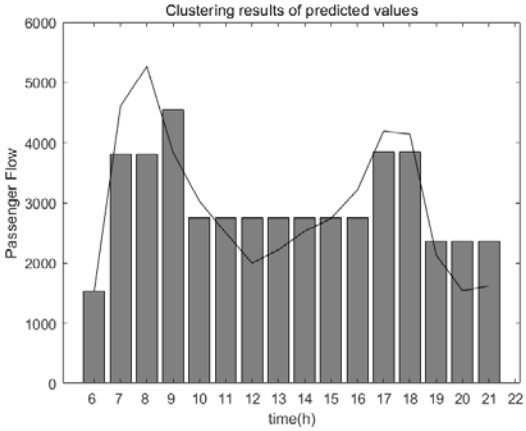

Figure 5. Clustering results of predicted values

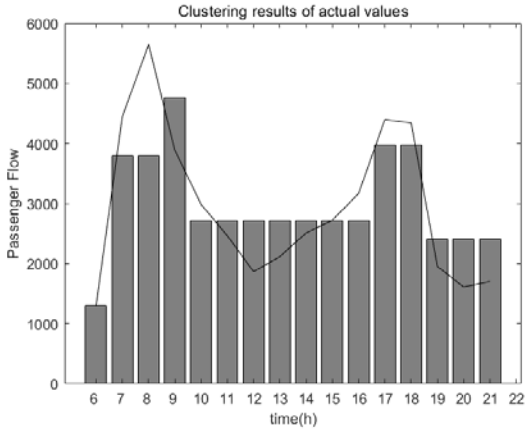

Figure 6. Clustering results of actual values

It can be seen that the predicted value and the actual value have the same trend, and the predicted result is in line with the actual one.It could be used as a basis for bus dispatch.

4.3 Analysis

The regression diagram of the prediction results of the BP neural network is as follows:

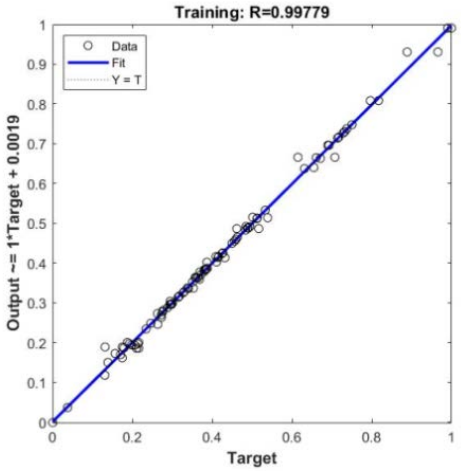

Figure 7. BP neural networkregression diagram

It can be obtained that R = 0.99764, and the deviation between the predicted data and the actual is 4.78%, which shows that the model parameters selected in this paper have a better fitting prediction effect and have achieved the stability of the prediction effect.

5 Conclusions

This paper proposes a scheme for bus passenger flow prediction and division based on BP neural network and Fisher optimal dissection method. The feasibility of the model is verified by experiments and comparisons using the data of a certain line of Guangzhou buses. Based on the results of passenger flow predicted and divided by this method, the peak time of passenger flow can be obtained, which could be used to determine the bus dispatch plan.

References

[1] Yao Qi, Wang Linshan. Forecasting Model of Road Traffic Accident Based on BP Neural Network [J]. Journal of Binzhou University, 2016, 32(06): 45-53.

[2] Zhang D, Lou S. The application research of neural network and BP algorithm in stock price pattern classification and prediction – ScienceDirect [J]. Future Generation Computer Systems, 2021, 115:872-879.

[3] Robert, Hecht-Nielsen. Theory of the backpropagation neural network [J]. Neural Networks, 1988.

[4] Yang H, Li X, Qiang W, et al. A network traffic forecasting method based on SA optimized ARIMA-BP neural network [J]. Computer Networks, 2021: 108102

[5] ZHANGShi-ying, HUANG Wei-hong.Applicability of the optimum partition method and the clustering methods forsomeorder sample [J].Journal of Applied Mathematics,1987(2):40-45

[6] HE Hongwei, ZHANG Ailing. The application of Fisher methodto dividing seismicity period in Yunnan province [J]. Journal ofSeismological Research, 1994, 17(3): 231-239.