HK.NCCP
International Journal of Intelligent Information and Management Science
ISSN: 2307-0692, Volume 8, Issue 6, December, 2019

Traffic Forecasting based on the Combined Corecasting Model

Minxin Lu*, Jingyang Sun
School of Economics and Finance, Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract: In this paper, aiming at the problem of traffic jam time prediction, in order to determine the more accurate traffic jam time under the road, four indicators of road level, traffic flow, time period and road length in Internet road condition data are selected as input variables, so as to establish BP neural network and SVM model to predict the travel time of the output variables. Considering the disadvantages of using only a single prediction method, the multi-model fusion prediction algorithm (MMFA) was selected to improve the prediction accuracy of the output variables by using the output information of multiple prediction methods. The BP neural network model and SVM model were combined to obtain a relatively accurate prediction value of the travel time.

Keywords: Clogging time prediction; Traffic flow; BP neural network; SVM model

1. Introduction

In the management and control of road traffic, travel time, as a key index, can not only reflect the current traffic status, but also measure the traffic delay and traffic efficiency. At the same time, with the continuous development of the current traffic flow induction system, the phenomenon of traffic congestion is becoming more and more common, and real-time route navigation has become a common concern of the public. Therefore, the accurate prediction of traffic jam time is of great value to the research of path navigation system, which is not only conducive to better planning the travel time, but also conducive to the detection and management of road traffic [1-2].

Xuemei Zhou, Xiaoguang Yang et al, in intelligent transportation system as the background, the statistical regression theory, combined with a large number of measured data to forecast the journey time, but they are only the section length and the number of intersections as impact factors, without considering road traffic speed, the influence of the traffic flow and time period, obviously predicted results is difficult to convince; Wenxia Zhou, Jianmin Xu et al, proposed to use kalman filter algorithm to predict the journey time of public transport by analyzing the driving characteristics of vehicles [3-4]. However, this method does not take into account the influence of passage speed, time classification and traffic flow on journey time, but only establishes a prediction model in a relatively ideal situation, and the data support is insufficient and unconvincing; Yang chao, Hu yao et al, constructed a stochastic volatility model based on the actual road segment journey time data, solved the model parameters with markov chain monte carlo method, and predicted the journey time with standard stochastic volatility (SV-N) model and thick tail stochastic volatility (SV-T) model. The results show that the thick tail stochastic volatility model is more accurate in real time prediction. However, the data of this model is the journey time data of two consecutive days for two sections with different road grades. The single control variable of the road is not studied in different time periods, and the lane variable is also studied in a specific way. Therefore, it is not of high reference value. In a word, the existing literature is more or less imperfect and needs to be improved.

2. Determine Multiple Model Weights

2.1. Source and processing of data

This paper collected the Internet road data of Shenzhen on March 25, 2018, and selected 15 different roads to collect data for empirical research. According to the tidal characteristics of urban traffic during working days, the time period is divided into two stages: flat peak and peak. The peak is from 7:00 to 9:00 and 17:00 to 19:00, while the flat peak is from 10:00 to 15:00, and is quantified as: 1 and 0.

2.2. Model input variable selection

The factors that affect the accuracy of vehicle travel time prediction are random and uncertain, complicated and diverse, such as rush hour, traffic flow, sudden traffic accidents, sudden changes in weather, unpredictable human factors, etc [5]., all of which will affect the accuracy of vehicle travel time prediction. Through the actual investigation and analysis of the running characteristics of vehicles and the road environment, the following factors are taken as the input variables of the model:

Road grade: the road grade divides the capacity of road design and is one of the evaluation index parameters of road state. The higher the road grade is, the better the road capacity is accordingly. Therefore, the influence of road grade factors in the prediction of journey time cannot be ignored. Road grade is divided into 5 levels.

Traffic flow: the size of traffic flow reflects the congestion of vehicles on the road. The more vehicles on the road, the slower the speed of buses will be, thus affecting the journey time of vehicles. The passage speed is determined by the road state. Therefore, this paper maps the size of traffic flow with the passage speed of different roads [6].

Time: the influence of peak time and peak time on vehicle travel is significantly different. The traffic flow during the peak hours is large and the running speed is slow. The traffic volume is small and the vehicle speed is relatively fast.

Road length: the different length of road segment will obviously lead to different driving time of vehicles. The driving speed of vehicles on the road segment is within a certain range. The longer the length of road segment is, the longer the corresponding journey time will be.

We recorded 5 sets of data at different stages, different road lengths and different road grades. By processing the data, we obtained a total of 150 sets of data for different roads in two periods, of which 130 sets of data were used as training data and 20 sets of data were used as test data.

2.3. Combination of multiple prediction methods

If only a single prediction method is used, even if it has more advantages and higher prediction accuracy, it will have some disadvantages in different road conditions and different time range [7-8]. If can output information comprehensive utilization many kinds of forecasting method, and the high accuracy of output data occupy larger weight, correspondingly low accuracy of the output data of low weight, finally makes the output results of several prediction for fusion, is not only can improve the accuracy of traffic data short-term prediction, and can guarantee the stability of the prediction error. Therefore, this paper selects multi-model fusion prediction algorithm (MMFA) to fuse BP neural network model and SVM model, so as to give full play to the prediction advantages of BP neural network and support vector machine [9].

In the process of multi-model fusion, weight plays a key role in the final prediction. Traditional determination methods, such as survey method, statistical analysis method and single factor analysis method, give the weight of each factor is fixed. But in the process of actual prediction, the precision of prediction is not always stable, so the MMFA weight is designed for forecasting data fusion can change as the prediction error and the calculation process of continuous adjustment, ensure high precision in the basic prediction method to predict the results of the final prediction results have a relatively large impact [10-11]. Based on this design, the concept of dynamic error is introduced. With the advance of time, the dynamic error can reflect the prediction effect of the prediction method in a previous period of time.

Dynamic error $e_{d,i}(k)$ is defined as:

$$e_{d,i}(k)=\frac{1}{k}\left[e_{ar,i}(k)+e_{ar,i}(k-1)+...+e_{ar,i}(t-n)\right]$$

In the formula, $e_{d,i}(k)$ represents the dynamic error of method I in k period, which is actually the $e_{d,i}(k)$ mean of method I in n period before k, I =1,2, corresponding to BP neural network and SVM model respectively. N represents the cumulative number of errors, which is 5 in this paper. $e_{ar,i}(k)$ represents the absolute value of the relative error obtained by the prediction analysis of method I in period k, and the calculation formula of $e_{ar,i}(k)$ is:

$$e_{ar,i}(k)=\left|\frac{Tk-\hat{T}_k^i}{Tk}\right|$$

Where, T_k represents the measured data in the k period; $\hat{T}_k^i$ represents the predicted value of method I in time period k.

On the basis of obtaining the dynamic error value of various prediction methods, the data fusion weight of the results of various prediction methods can be analyzed and obtained. The main method is the inverse proportion method, and the weight of prediction results in k period can be analyzed through the dynamic error in period $(k-n)$, which requires that ω_k^i is a function that keeps changing with the change of $e_{d,i}(k-1)$. The application criterion of the inverse proportion method is: the weight of the fusion and the dynamic error must show an inverse proportion relationship, that is, the value of the initial weight $\dot{\omega}_k^i$ should be obtained by using the inverse proportion method:

$$\dot{\omega}_k^i=\frac{1}{e_{d,i}(k-1)}$$

Then the initial weight $\omega_i^*(k)$ was normalized to obtain the fusion weight of the final prediction results of each method:

$$\omega_k^i=\frac{\dot{\omega}_k^i}{\sum_{i=1}^{2}\dot{\omega}_k^i}$$

3. Prediction of Travel Time

3.1. BP neural network

130 training samples were randomly generated from 150 processed data as the training set of BP model, and the remaining 20 test samples as the test set. The number of iterations was set as 1000, the minimum error of iteration goal was set as 0.001, the display frequency was set as 10, that is, each training was displayed once for 10 times, and the learning rate was set as 0.05.By MATLAB software programming, the predicted journey time of the test set can be obtained, and the regression diagram of the neural network can be further calculated as model R^2 =0.95563. It can be seen that the R values of the three parts of the training, validation and test are all close to 1, and the overall R value also reaches 0.97132. The fitting effect of the model is good. Finally, the predicted value is compared with the sample value of the test set as follows:

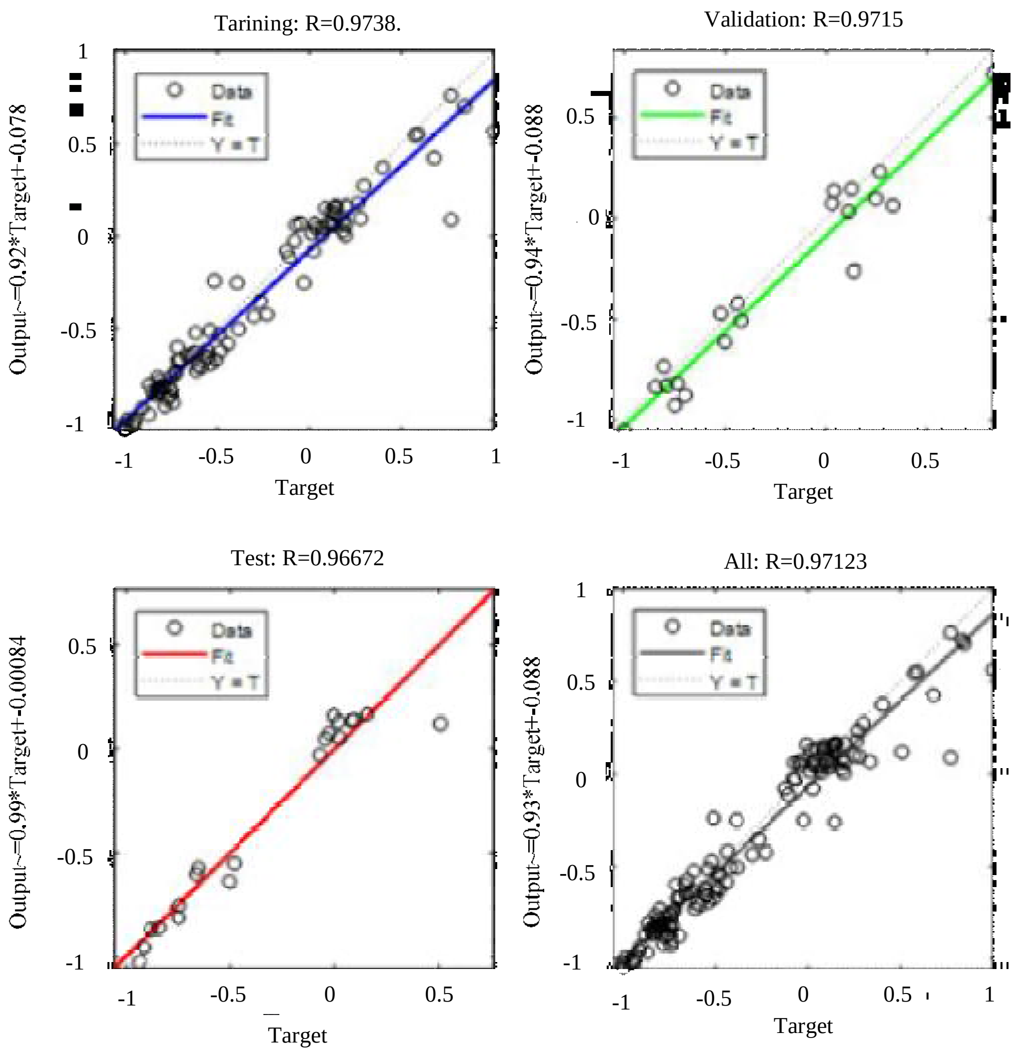


Figure 1. Regression diagram of a neural network

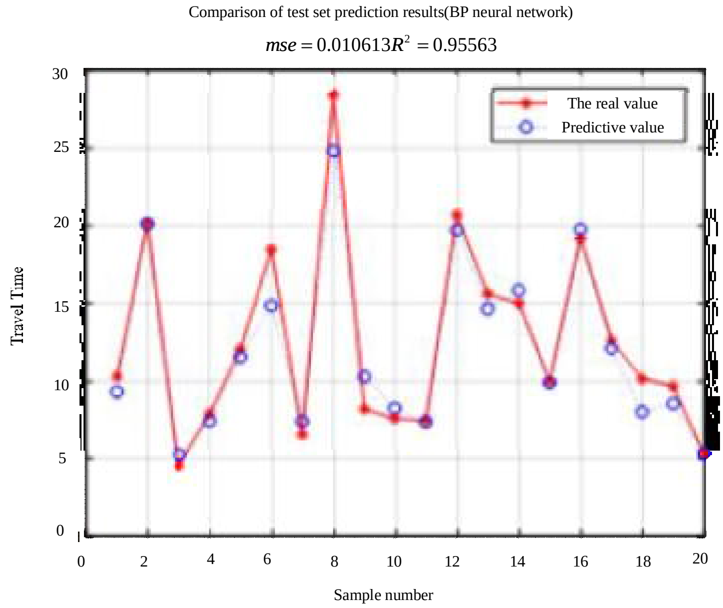
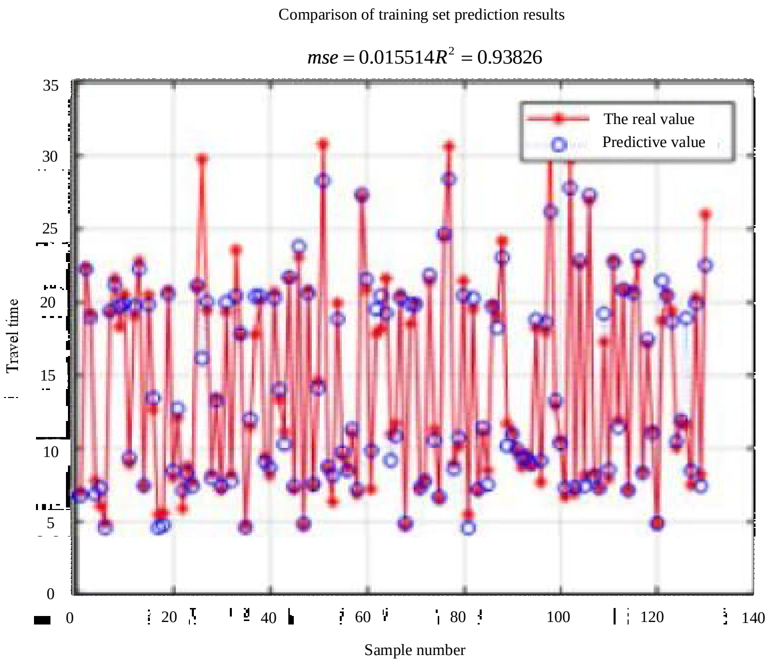


Figure 2. Comparison of prediction results

3.2. SVM model

130 training samples were randomly generated from 150 processed data as the training set of SVM model, and the remaining 20 test samples as the test set. First, the cross validation method was used to find the best c/g parameters, and then the RBF kernel function was used to train the SVM model. By MATLAB software programming, the predicted journey time of the test set could be obtained, then the model =0.9541 was further calculated, and the predicted value and sample value were compared as follows:



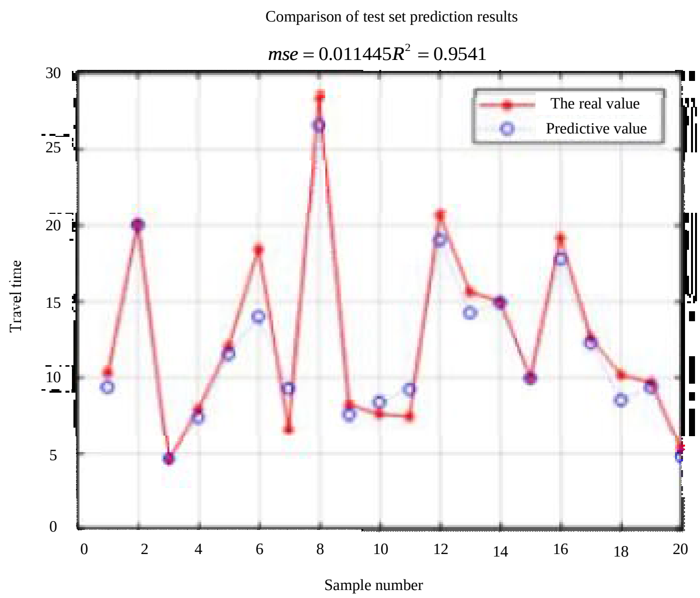


Figure 3. Comparison of prediction results

3.3. Combined prediction model

The weights of BP neural network prediction and SVM prediction methods are respectively 0.711 and 0.289 through the inverse proportion method [12]. Further, the journey time prediction value of the combined prediction method is calculated according to the weights of each method. Finally, the output results of the three methods are compared with the actual values. Figure 4 shows the comparison between the test set output predicted by BP neural network, SVM and the combination of the two methods and the actual value.

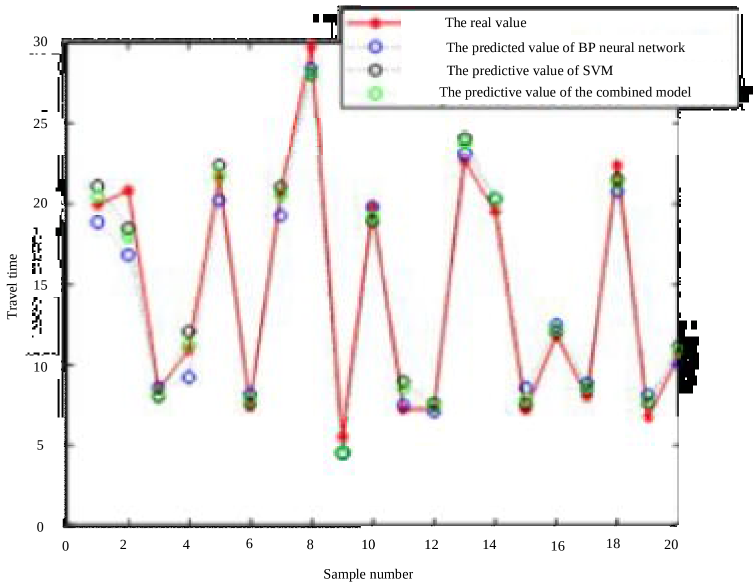


Figure 4. Comparison diagram of output results of each model

Can be seen from the figure 4, relative to the SVM and BP neural network forecasting method, combination forecast of the forecast can be well fit the change of the actual journey time, especially the seven journey time prediction and the fit of the actual value is very high, in more than one sample, combination forecasting are higher than other prediction precision, shows that the combined forecasting method is more suitable for journey time prediction.

4. Evaluation of the Effect of Travel Time Prediction

4.1. Description of model evaluation index

MRE, LSE, MAE, MSE, RMSE and other evaluation indexes are commonly adopted in traffic information prediction. For the sake of simplification, two indexes, fitting ability and correlation, are adopted in this paper to evaluate and predict the effect. Among them, average relative error (MRE) refers to the average relative error between the predicted value and the actual value, which reflects the overall difference between the predicted value and the true value. The minimum error sum of squares (LSE) is an index to evaluate the correlation between the predicted value and the true value. The smaller the value, the higher the correlation between the predicted value and the actual value. The calculation of these two indicators is shown in equations respectively.

$$MRE = \frac{1}{n}\sum_{k=1}^{n}\frac{\left|Tk - \hat{T}^{i}_{k}\right|}{T^{i}_{k}}$$

$$LSE = \frac{\sqrt{\frac{1}{n}\sum_{k=1}^{n}(T_k - \hat{T}_k)^2}}{\frac{1}{n}\sum_{k=1}^{n}T_k}$$

Where, k represents the order of samples, $k = 1, 2, \cdots, n$, n represents the number of samples. T_k represents the actual value; $\dot{T}_k$ represents the predicted value.

4.2. Model comparison evaluation

As can be seen from table 1, compared with the relative errors of SVM and BP neural network prediction, the MRE and LSE values of the combined prediction method are lower, which indicates that the accuracy of the combined prediction method is higher, and the predicted value is closer to the actual journey time, indicating that the combined prediction model has better performance for the problems in this paper.

Table 1. Evaluation index results of each prediction method

The evaluation index	MRE/%	LSE
The SVM prediction	1.1445	0.044428
The SVM prediction	1.0613	0.033537
Combination forecast	1.0272	0.030125

5. Conclusion

Through the above empirical studies, we respectively established the BP neural network model, SVM prediction model and combination prediction model for travel time. Finally, by comparing and analyzing the errors of the three models, we concluded that the combination prediction model has the highest prediction accuracy and is more suitable for the study of travel time prediction.

References

[1] Yang C., Hu Y., Shang M.J., et al. Journey time prediction based on SV model. Journal of Guizhou University (Natural Science Edition). 2019, (05), 11-17.

[2] Yang H., Wang Z.Y., Zou Y.J., et al. Prediction model of short journey time based on the second revision. Journal of Tongji University (Natural Science Edition). 2019, (10), 1454-1462.

[3] Ding H.F., Li Y.H., Liu B., et al. Combined prediction of journey time of express road based on BP neural network and SVM. Computer Application Research. 2016, (10), 2929-2931, 2936.

[4] Yao Z.H., Jiang Y.S., Han P., et al. Prediction model of traffic flow with small time granularity based on neural network. Transportation System Engineering And Information. 2017, (01), 67-73.

[5] Zhao X.W. Bayesian estimation and empirical analysis of SV model based on MCMC method. China University of Mining and Technology, 2017.

[6] Li W.F., Duan Z.Y., Yang D.Y. Analysis of journey time fluctuation of expressway. Computer Engineering and Application. 2015, 51(04), 260-265.

[7] Ming Z. Analysis and application of financial stochastic volatility expansion model. Yanshan University. 2010.

[8] Yang X.P. Study on prediction method of urban road trip time. Jilin University. 2005.

[9] Zhou X.M., Yang X.G., et al. Research on the prediction method of bus journey time. Transportation and computer. 2002, (06), 12-14.

[10] Zhou W.X., Xu J.M., et al. Prediction of bus journey time based on kalman filter algorithm. Traffic Standardization. 2007, (z1), 174-177.

[11] Taylor S.J. Modeling stochastic volatility: a review and comparative study modelin. Mathematical Finance. 1994, 4(2), 183-204.

[12] Wang L., Wang G.W., Ning X.J., Zhang J.L., et al. Application of BP neural network to the prediction of coal ash melting characteristic temperature. Fuel. 2020.