Short-Term Traffic State Prediction Based on Support Vector Machine

Jinjie QIU[1*]
Highway College, Chang'an University
Middle Section of S. 2nd Ring Road, Xi'an, Shaanxi 710064, China
13299111675@163.com
Xiaoqing LIU[2]
Highway College, Chang'an University
Middle Section of S. 2nd Ring Road, Xi'an, Shaanxi 710064, China
254727669@qq.com
Ruyue LIU[3]
Highway College, Chang'an University
Middle Section of S. 2nd Ring Road, Xi'an, Shaanxi 710064, China
2322563774@qq.com

ABSTRACT

In practice, it is easy to cause the phenomenon that planned optimal path not be consistent with the actual best path due to the change of the roads traffic conditions. Given this situation, it is very important to predict future changes in short-term traffic flow conditions.In this paper, a mathematical model was established on the short-term traffic flow prediction based on support vector regression machine, with the speeds of the predicted roads as well as its connected sections as the input set, the kernel function was selected to train the support vector regression machine.   Then the trained vector regression machine was applied and the input parameters were used to predict the traffic speed of the next time period. Finally, the model was tested by real-time monitoring data of some sections of the Second Ring Road in Xi'an. The prediction results show that the model is effective.

Keywords: Traffic Engineering; Traffic Flow; Short Term Traffic Flow Forecasting; Machine Learning; SVR (support vector regression)

1. INTRODUCTION

Traffic condition prediction technology is an important part of the research and application of intelligent transportation system. In the past 30 years, it has attracted wide attentions of researchers in various fields of transportation. Especially in the field of intelligent transportation system, accurate and efficient real-time traffic prediction can not only provide travelers with travel information, but also assist traffic operation management center to achieve concentrated traffic control and guidance, which can improve the efficiency of traffic operation.

At present, there are many researches on short-term traffic condition prediction, and dozens of short-term traffic state prediction methods proposed by domestic and foreign scholars. These prediction methods can be summarized into two categories according to technical means: one is linear

prediction method, which mainly includes time series prediction method, parameter regression prediction, state space prediction method and Kalman filter prediction method; the other is nonlinear prediction. These methods mainly include gray system theory method, neural network prediction method and support vector machine regression prediction method. Since short-term traffic flow systems have typical nonlinear characteristics, nonlinear prediction methods usually achieve relatively good prediction accuracy.

In the nonlinear prediction methods, the support vector machine regression model can effectively solve the learning problems of small sample and high dimension by adopting the structural risk minimization principle and introducing kernel function method. By transforming the learning algorithm into the optimization problem of convex quadratic programming, the global optimization problem is effectively solved. Extreme value problem. Support Vector Machine (SVM) model is widely used in nonlinear prediction because of its advantages, complete statistical learning theory and good generalization performance.

In addition, support vector regression (SVR) also has many advantages in the research of short-term traffic condition prediction, in which the accounting method provides an important component module for SVM. The sample data is implicitly expressed as a feature space in which a nonlinear learner is trained, so the problem of computational feature mapping is also overcome effectively.

Studies have shown that there are spatial correlation characteristics of urban road traffic status. The change of road traffic state will affect other adjacent sections in a relatively short time, resulting in the changes of traffic state of other sections within a certain space. Based on this feature, it is feasible to predict the change of traffic state of this section in a short time in the future with the relevant traffic state as input variable.

Based on the above conditions, this paper proposed a traffic folw condition prediction model based on support vector machine regression(SVR). The speeds of the predicted road segments and its Connected sections were taken as input, and then the appropriate kernel function was selected to train the support vector regression machine. The support vector regression machine completed by training was applied to predict the traffic line speeds of the next period. Finally, the model was evaluated by real-time monitoring data of some roads in the second ring road of Xi'an, and the experimental results were obtained.

2. RESEARCH AREA OVERVIEW AND RESEARCH METHODS

Research Area Overview

Known as Chinana anciently, Xi'an is the capital of Shaanxi Province. It is located in the middle of the Guan Zhong Plain. It is one of the three international metropolises explicitly built by China and the ninth national central city in the country. The industrial distribution of Xi'an is mainly concentrated in four regions: high-tech development zone, economic and technological development zone, Qujiang new zone and Chanba eco-tourism zone. The four industrial concentration areas are located around the main city downtown, which has great impact on the traffic in the city. In addition, as a famous historical and cultural city, Xi'an receives more tourists every year and also puts a lot of pressure on traffic.

The urban road network in the main urban area of Xi'an is adopted as the development mode of " Square grid, Ring & radiation", gradually forming Road network structure of the city of 'three

2

verticals, three horizontals, three rings, eight radiations and one beltway' composed bytwo axes of East and West, North and South, Ring NO.1 , Ring NO.2 and Ring NO.3 as the three ring roads, and eight roads of Taibah Road, Tayyip Road, Xianning Road, Huaqing Road, Taihua Road, Zhuhong Road, Daxing Road and Kunming Road as the radiation road and Ring City Expressway. Which provides road infrastructure services for the main urban areas with dense population, active social activities and traffic demand.This paper takes the road network in the second ring road of Xi'an as the research object. The trunk road system bears the main traffic pressure of the city, its traffic status has a great impact on urban road traffic congestion. Therefore, the spatial and temporal distribution characteristics of traffic congestion in this study area are based on the trunk road system.

Data acquisition and processing

Amap is the leading provider of digital map content, navigation and location services in China. The Amapopen platform is a one-stop development service platform for developers. It can provide developers with all-round development tools from data storage, management, retrieval to map service capabilities, not only to achieve map positioning, display, labeling, Search, driving route planning, bus inquiry, navigation and other functions, can also quickly achieve group purchase, message push, map business card and other services through standardized components, thereby effectively reducing development time and reducing development costs. The data of this study was derived from the API traffic situation map of Amap open platform.A small program was written with python to obtain the traffic state data within the Xi'an Second Ring Road at intervals of 10 min in the morning peak period(8:00-10:00). (where the longer section is divided into small sections)

Building prediction models

Because the change process of traffic state is a real-time, nonlinear, high-dimensional, non-stationary random process, the randomness and uncertainty of traffic state changes become stronger with the shortening of statistical period. The short-term changes in traffic status are not only related to the traffic status of the past few periods of the road, but also affected by the upstream and downstream traffic conditions. The problem to be solved in short-term traffic state prediction is how to find out the regularity and establish the forecast from the traffic state changes with randomness and uncertainty, according to the traffic state parameters obtained by the vehicle detector, combined with other influencing factors. Model to predict traffic state changes in the next few hours.

Support Vector Machine (SVM) has many advantages in the study of short-term traffic state prediction. The kernel algorithm provides an important component module for the support vector machine. The kernel function method directly combines two steps to establish a nonlinear learner, implicitly expresses the sample data as a feature space, and trains a nonlinear learner in it, thus surpassing the problem of calculating the feature map that is originally needed. In the case of traffic prediction based on a large number of high-dimensional, random, and nonlinear data research problems, the application of the kernel function method can achieve a multiplier effect. Therefore, the process of using the support vector machine method is largely a process to choose a suitable kernel function. The problem of Support Vector Machine is the kernel function.

SVR is widely used in the transportation field because of its small structural risk and ability to avoid dimensionality disasters and over-fitting. SVR is a nonlinear kernel learning algorithm with sparse solution, which converts linear inseparable problems into high through kernel functions. The linear problem of dimensional space, the goal is to achieve the least risk of model structuring, and to

find a compromise between data fitting precision and complexity of approximation function to avoid over-fitting. According to the difference of loss function, it is often applied to traffic flow prediction. The SVR model mainly includes: ε-SVR, υ-SVR and least squares SVR, of which ε-SVR is the most widely used.

According to the algorithm of ε-SVR support vector regression machine and the basic principle of short-term prediction of traffic state, a predictive model of support vector machine regression can be established.

1 Constructing training set

Let the current time period bet, the speed of one road segment i in the period t and the speed of the upper and lower road segments of the road segment i be the input feature set, and predict the average linear velocity of the t+1 period of the road segment i. Construct the following training set

$$S=\{(x_1,y_1) \quad (x_2,y_2) \quad (x_3,y_3)\cdots(x_l,y_l)\} \qquad (1)$$

Among them, $y_1 \quad y_2 \quad y_3 \cdots y_l$ are the output sets of the training samples for the known average linear velocities $x_1 \quad x_2 \quad x_3 \cdots x_l$ of the past several periods of section i, and i is the data set of the training sets, in which x_i is the speed set of the upper and lower sections of section i in the t period.

2 Solving support vector machines

Select the appropriate parameters , C and the kernel function $K(x \quad x')$ to construct the optimization equation and solve the support vector machine:

Firstly, the error probability ε =0 in the quadratic loss function corresponds to the least squareregression (also known as ridge regression) principle of the weighted attenuation factor, and the original problem can be obtained as follows:

$$\min\|\omega\|^2 + \sum_{i=1}^l \xi_i^2 \qquad (2)$$

$$\text{s.t.} \quad y_i - <\omega \cdot x_i> = \xi_i (i=1,2,3\cdots l)$$

According to the Lagrange theory, the Lagrange function is obtained

$$\min L\,(\omega,\xi,a) = \|\omega\|^2 + \sum_{i=1}^l \xi_i^2 + \sum_{i=1}^l a_i(y_i - <\omega \cdot x_i> \xi_i) \qquad (3)$$

Find the derivative and let it be 0, get $\omega = \dfrac{a_i}{2}$

Turn it into a dual problem:

4

$$\max_{a^{(*)} \in R} W(a^{(*)}) = \sum_{i=1}^{l} y_i(a_i^* - a_i) - \frac{1}{2}\sum_{i=1}^{l}(a_i^* - a_i)\sum_{j=1}^{l}(a_j^* - a_j)K(x_i, x_j)$$

$$\text{s.t. } \begin{cases} \sum_{i=1}^{l}(a_i^* - a_i) = 0 \\ a_i^* \in \left[\frac{C}{l}\right] i = 1,2,3\cdots l \qquad (4) \\ \sum_{i=1}^{l}(a_i^* + a_i) \leq \quad \cdot C \end{cases}$$

Where: the original problem of the objective function $W(a^{(*)}))$ represents the minimum function of the error between the actual value and the predicted value, $K(x_i, x_j)$ is the kernel function, which implicitly defines the feature space directly calculates the result of the target from nonlinear data. , C is selected based on the prediction results of multiple trainings.

The optimal solution derived from the dual problem $\overline{a^{(*)}} = (\overline{a_1^*}, \overline{a_1}, \overline{a_2^*}, \overline{a_2}, \cdots \overline{a_l^*}, \overline{a_l})$, which is equivalent to selecting the support vector from the training set as the weight coefficient of each output input vector in the optimization set in the optimization function.

(3) Constructing prediction function

Using the corresponding regression equation $f(x) = <\omega \cdot x> + b$, construct the prediction function as:

$$f(x) = \sum_{i=1}^{l}(a_i^* - a_i)K(x, x_i) + b^* \quad (5)$$

In the kernel function of the prediction function, x_i is the training sample in the training set S, which is obtained by the i loop one by one, and x is the current time of the road segment i to be predicted back to the p time period (the t-p+1 time period to the first Vector combination of various traffic state influence parameters at time t)

Among them, b^* can be calculated as follows:

Select two components $\overline{a_j^*}$ $\overline{a_k}$ in the open interval $(0 \quad \frac{C}{l})$;

$$b^* = \frac{1}{2}[y_i + y_k - (\sum_{i=1}^{l}(\overline{a_i^*} - \overline{a_i})K(x_i, x_j) + \sum_{i=1}^{l}(\overline{a_i^*} - \overline{a_i})K(x_i, x_k))] \quad (6)$$

The resulting objective function value $f(x)$ is the predicted result.

With this, the short-term traffic average linear velocity prediction model based on support vector regression machine has been established. Finally, the average linear velocity of the next period can be predicted by using the prediction function $f(x)$ by inputting the combination of the traffic parameter vectors X of this section and the upper and lower sections in the current period and the past several periods.

3. EXAMPLE ANALYSIS

The data of this study was derived from the API traffic situation map on the Amap open

platform. A small program was written with python to obtain the traffic state data within the Xi'an Second Ring Road at intervals of 10 min in the morning peak period (8:00-10:00)(the longer section is divided into small sections). The purpose of this study is to build a training set through the historical data of the road network, and then solve the support vector machine to predict the future short-term traffic state changes, and compare with the actual data to evaluate the prediction effect.

Take Nanguanzheng Street of Xi'an Road Network as a research sample construction training set.The Upper and lower sections of NanguanZhengjie include Chang'an South Road, Youyi West Road, Youyi East Road, Huancheng South Road, Nanjie Street, Zhenxing Road, Gymnasium Road, etc. Considering the impact of the connected road sections traffic flow on the road section traffic, the speed of the above section including Nanguan Zheng Street section is used as input set to predict the speed of Nanguan Zheng Street section at the next time interval.

Table 1. The speeds of the study road

Roads	8:00	8:10	…
NanguanZhengjie	30	20	
West Youyi Road	25	30	
Eest Youyi Road	15	10	
SouthHuancheng Road	45	40	
NanDa Street	30	35	
Zhenxing Road	25	25	
Gymnasium Road	20	25	
Zhongnao Street	15	10	
Chang'an South Road	35	40	
Nanguo Road	20	20	

Construct the training set as follows

$$S=\{(x_1,y_1) \quad (x_2,y_2) \quad (x_3,y_3)\cdots(x_{12},y_{12})\}$$

Among them x_i is the speed set of the road segment and its upstream and downstream road segments at the starting time (8:00), and y_i is the speed of the road segment at the next time interval, and is also the output set of the training set. Based on the above data:

$$x_1 = (30 \quad 25 \quad 15 \quad 45 \quad 30 \quad 25 \quad 20 \quad 15 \quad 35 \quad 20) \quad y_1 = (20)$$

$$x_2 = (20 \quad 30 \quad 10 \quad 40 \quad 35 \quad 25 \quad 25 \quad 10 \quad 40 \quad 20) \quad y_2 = (30)$$

…

All the data in the study period are integrated through the above process, and all the training sets of the research sample can be obtained.

Python and its compiler PyCharm sklearn toolkit can be used to realize training set training

6

and support vector machine regression prediction.

In this study, the Radial Basic Function (RBP) kernel function was selected as the support function vector machine regression prediction kernel function, and predicted that the speed of the road segment is 23.6km/h at 9:10, which is close to the actual 20km/h. which shows that the method is effective in predicting short-term road network traffic conditions.

The 27 sections of the road network in Xi'an were selected as the research samples. After the training, the vector machine regression prediction was used as the test set to evaluate the validity of the model. The error conditions in all samples are as follows:
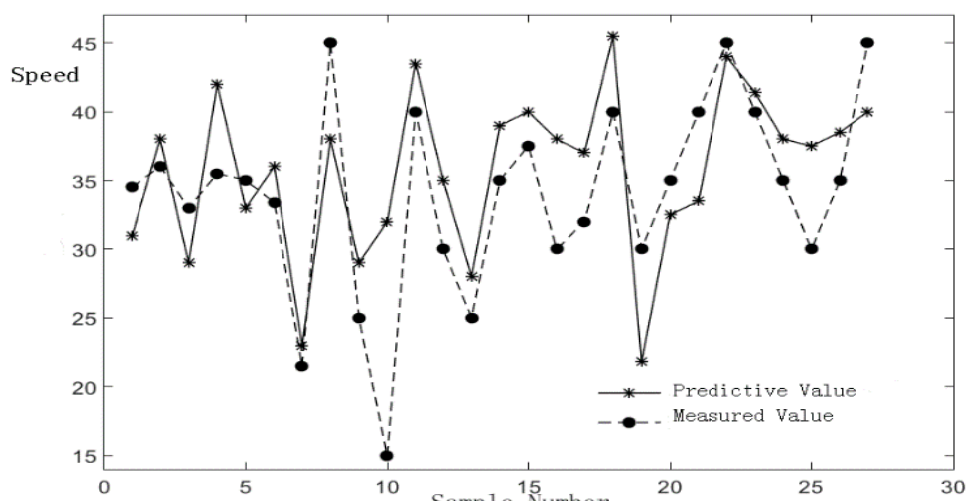

Figure 1. Comparison of forecast results

As can be seen from the above table, except for a few samples, the prediction results of the model in most sections of the highway have high accuracy, basically meet the prediction accuracy requirements. For the sample sections with large prediction error, the upper and lower link sections can be adjusted to meet the accuracy requirements.

According to the above method, the speed of each section and its upstream and downstream sections can be used to predict the traffic state changes in the short term in the future, and the traffic state changes of the whole network can be predicted by analyzing and predicting each section.

4. PROSPECT AND LACK OF RESEARCH

In this study, a short-term traffic state prediction model based on support vector machine is constructed, and an example of Xi'anwas analyzed to prove the effectiveness of the model to some extent. The traffic state short-term prediction results can provide support for real-time traffic control and traffic flow guidance systems, and also can provide path decision basis for future vehicle networking and smart car driverless mode.

Due to the difficulty of the traffic history data, the author only considers the average linear velocity in the model as the traffic state parameter, and uses the average speed of the predicted road segment and its adjacent upstream and downstream sections to average the specified time. The line speed is predicted to reflect the traffic state at the next moment of the road segment. Although the traffic impact parameters considered in the example analysis are small, the small error of the prediction results indicates the validity of the model.

The model still needs to be improved in practical applications. For example, how to extend the

7

model to the multi-point prediction method while maintaining the prediction accuracy and occupying less computing resources, so that the method can meet the practical application is the next urgent problem. In addition, the model's parameter optimization, prediction accuracy improvement, model application scale and other aspects need to be more in-depth research and discussion.

REFERENCES

Chunjiao Dong, Chunfu Shao, Chengxiang Zhuge, Meng Meng. Short Term Prediction Model Of Traffic Free Flow Based On Spatio-Temporal Characteristics[J]. China Civil Engineering Journal, 2013,8(8):111-118.

Eleni I. Vlahogianni, Matthew G. Karlaftis, John C. Golias. Short-term Traffic Forecasting Where We are and Where We're Going[J]. Transportation Research Part C Emerging Technologies 2014(43) 3-19.

Haiting Lu, Ning Zhang, Huang Wei. Research Progress of Short-term Traffic Flow PredictionMethods[J]. Journal of Transportation Engineering and Information,2009,7(4) 84-91.

Kamarianakis Y, Prastacos, P.(2003). "Forecasting Traffic Flow Conditions in an Urban Network:Comparison of Multivariate and Univariate Approaches" *Transportation Research Record, 1857*,   Transportation Research Board   Washington, D.C. ,74-84.

Lunhui Xu, Hui Fu. Traffic Information Intelligent Prediction Theory and Method[M]. Beijing. Science Press 2009.

Weiqing Xiong, Xiaobo Yan, Shouxu Jiang. Short-term Traffic Flow Prediction based on BP NeuralNetwork and Fuzzy Inference System[J]. Intelligent Computer and Applications,2015(2):43-51.

Zhaosheng Yang, Yuan Wang, Qing Guan. Short Term Traffic Flow Forecasting Method Based on Support Vector Machine[J] ,Journal of Jilin University (Engineering Edition), 2006 36(6) 881 —884

Zhao Liu, Wei Du, Dongmei Yan. Short-term traffic flow prediction technology based on K-nearest neighbor algorithm and support vector regression combination[J]. Journal of Highway and Transportation Research and Development, 2017, 34(5):122-128.

ZhishengYao,ChunfuShao,Yongliang Gao. Research on Methods of Short-term Traffic Forecasting based onSupport Vector Regression[J]. Journal of Beijing Jiaotong University,2006, 30(3):19-22.