TELKOMNIKA, Vol. 11, No. 12, December 2013, pp. 7446~7452
e-ISSN: 2087-278X

Traffic Prediction Based on SVM Training Sample Divided by Time

Lingli Li[1], Hongxia Xia[2], Lin Li[3] , Qingbo Wang[4]
School of Computer Science and Technology, Wuhan University of Technology, Wuhan, 430070, China
Corresponding author, e-mail: lilingli.wuhan@gmail.com[1], xhx@whut.edu.cn[2], cathylilin@whut.edu.cn[3]

Abstract

In recent years, the volume of traffic is rapidly increasing. When vehicles running through the tunnel are more intensive or move slowly, the tunnel environment occurs deteriorated sharply, which affects the normal operation of the vehicle in the tunnel. This paper uses the result of previous mining association rules to select feature items and to establish four training samples divided by time. Then the training samples are utilized to create the SVM classification model. Finally the trained SVM model is used to prediction the tunnel traffic situation. Through traffic situation prediction, effective decisions can be made before traffic jams, and ensure that the tunnel traffic is normal.

Keywords: *SVM, training data, test data, traffic situation prediction*

1. Introduction

A tunnel is a special part of roads. It has narrow space environment, light changes, poor visibility. Therefore, there is a potential risk of traffic accidents. Thus, the prediction of the traffic jam situation in the tunnel monitoring system is very important.

It's not difficult to find that a lot of research in this field has been done, such as Traffic State Forecasting Based on the Maximum Entropy Model [1], Short-Term Traffic Flow Forecasting Based on Multi-Dimensional Parameters [2], Grid parallel computation of online traffic status prediction, which using generalized neural network [3], and so on.

After repeated experiments, we find that the establishment of the SVM training sample is very important to forecast results, and to get the precise results four training samples based on time are established, currently, no studies have been considering to establishing the training sample based on time. However, it's very necessary to consider the time factor in traffic situation prediction city tunnel, some tunnel data changes over time. For example, light intensity will change with time, and at different time we measured the light intensity data is also different. Therefore, we establish the training sample based on time.

2. The Brief Introduction of a City Tunnel Monitoring System

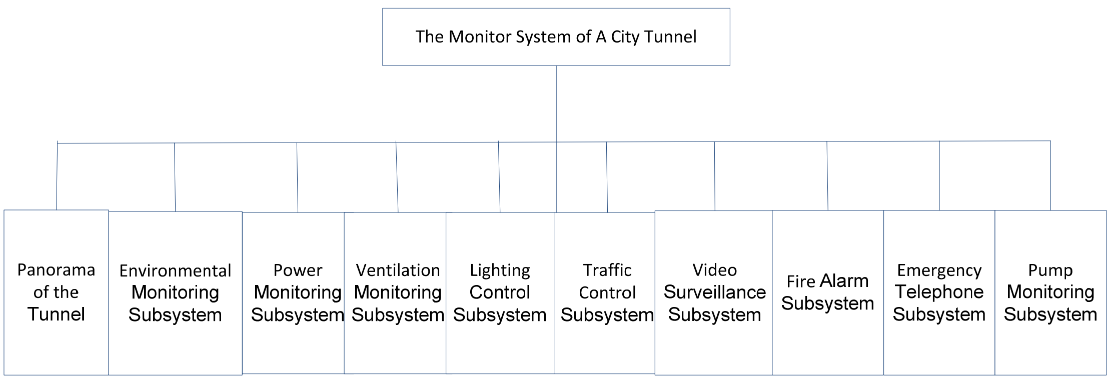

Figure 1. The Structure of our Traffic Monitor System

Received June 26, 2013; Revised August 1, 2013; Accepted August 20, 2013

The structure of our traffic monitor system is shown in Figure 1. In our experiment, CO concentration data, wind speed data, light intensity data, humidity data, temperature data, traffic volume data, and vehicle speed data is used. CO concentration data, wind speed data, light intensity data, humidity data, and temperature data is got from the environmental monitoring system, environmental monitoring system is mainly used to detect and control the tunnel environment. And traffic volume data and vehicle speed data is indirectly got from traffic control subsystem. Traffic control subsystem is mainly used to measure the total traffic volume and vehicle speed.

3. Realize the Traffic Situation Prediction based on SVM

In our previous study, we research on real-time and dynamic urban traffic [4], and association rule mining is used to mining correlated features for situation prediction. Finally, CO concentration, wind speed, light intensity, humidity, temperature, traffic volume, vehicle speed seven correlated feature items is selected and used to build training sample. Now, support vector machine (SVM) method is adopted to build SVM models based on the training samples and predict the traffic situation.

3.1. The Principle of Support Vector Machine

The principle of support vector machine theory initially comes from the processing of the data classification problems. SVM can be used for data classification and regression [5-7] and has been applied to different kinds of fields, such as traffic signs [8], image retrieval [9], gene prediction [10], component content soft-sensor of Ions color characteristics [11], and so on. The mechanism of the method can be simply described as: Looking for an optimal hyperplane that meets the classification requirements, such that the hyperplane while ensuring the classification accuracy, it is possible to maximize the hyperplane on both sides of the blank area. In theory, support vector machine can achieve the optimal classification of linearly separable data. In order to solve the nonlinear problem, Vapnik et al introduce the kernel mapping method to transform the nonlinear problems into higher dimensional linear space separable problems to solve.

3.2. Multi-class SVM Classification Method based on Binary Tree

SVM is essentially a binary classification, however, in practical applications, SVM is used to solve the multi-classification problems. For multi-classification recognition problem, SVM must be improved and generalized. There are several common multi-classification algorithms [12], in our experiment we use Multi-class SVM classification method based on the binary tree, for K-class training samples, training K-1 support vector machines.

3.2. Realize Traffic Situation Prediction

In this paper, four training samples based on time is built. Specifically, we build four training samples according to the season. Firstly, determine that the test sample data belongs to which period and select the appropriate training sample. Secondly, use the selected training sample to establish the SVM models. And finally, use SVM model to predict the test data and obtain the predicted results. The schematic diagram of traffic situation prediction based on SVM training sample divided by time is shown as Figure 2.

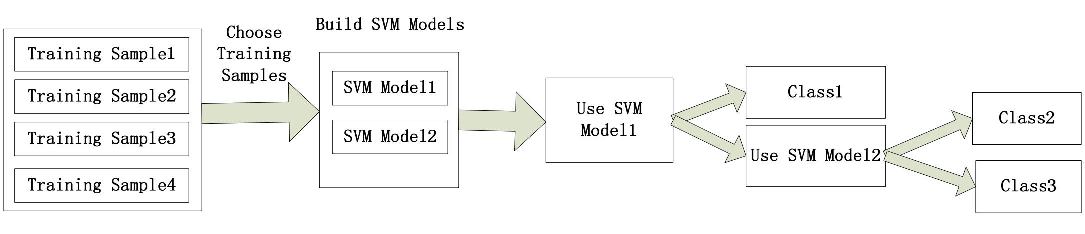

Figure 2. Traffic Situation Prediction based on SVM Training Sample Divided by Time

4. Experiment

In this experiment, libsvm-mat-2.9-1 is used in MATLAB software tool, and establishes SVM models based on binary tree multi-class SVM classification, according to the SVM models to predict the traffic jam situation in a tunnel.

The SVM algorithm in the MATLAB environment to achieve the core content is the application of optimization toolbox. Implementation of support vector machine algorithm based on MATLAB environment has the characteristics of simple, convenient and quick code transplantation, reliable performance, easy to control the program runs. Especially provide a simple, quick support vector machine algorithm research and application of technology platforms for non-computer scientists.

Through comparison and analysis of function types and the related parameters of libsvm-mat-2.9-1 software package, to find the best experimental model, finally achieve the ideal prediction.

4.1. Data Preprocessing

(1) The experimental data are measured data of a tunnel, selecting part of the data to create the SVM model, and some of the data to predict.

(2) There are seven feature items, including CO concentration, wind speed, light intensity, humidity, temperature, traffic volume, and vehicle speed.

(3) Taking into account the different time, the range of these feature items are different, the spring, summer, autumn, winter four training samples based on time and a test sample are established.

4.2. Experimental Process

Our experiment is according to CO concentration, wind speed, light intensity, humidity, temperature, traffic volume, vehicle speed seven feature items to predict whether the traffic tunnel is traffic jam, if traffic jam, subdivided into mild and serious two cases. According to the actual needs we use SVM multiclass classification method based on the binary tree to create two SVM model, and divide the tunnel traffic condition into "normal", "mild" and "serious" three classes. Firstly using SVM1 to separate the “normal” class, and then using SVM2 to separate the “mild” class and the “serious” class.

Parameters of LIBSVM involved in SVM regulation is relatively less, providing a lot of the default parameters, found after many experiments to predict experimental results on the basis of these default parameters.

The selection of SVM method, kernel function and its parameters in this experiment is described as follows:

(1) According to the Testing and comparison of experimental results, selecting C-SVR regression and RBF kernel function.

(2) The selection of model parameters “C” and “g”, according to the testing and comparison of experimental results, selecting C=1000, g=0.1.

The Concrete realization in MATLAB is described as follows:

Step 1: Giving the training sample data and the testing sample data, and each sample xi corresponds to a desired output yi, yi $\in$ {+1, -1}.

Step 2: According to the testing sample judge the sample data belongs to which quarter data, then selecting the corresponding training samples. Because different selection of training sample, the result of prediction is different, so in order to predict more accurately, we must a suitable training samples must be choosen. According to the training samples, and its desired output value, selecting SVM method, kernel function and its parameters, calling the svmtrain function to establish the SVM model.

Step 3: Using SVM model to predict the test samples data, the specific implementation is as follows:

```
[predictedY,Accuracy]= svmpredict(ClassResult,TestData,model)
```

Among them, ClassResult is our own judgment for the first, can be 1 or -1; TestData is the test data; model is the SVM model that we get from the second step.

In the program, we use “1” or “-1”to represent expected output.

Taking into account the different time, the range of feature items data is different, in order to predict well, we established four training samples based on the time. According to the test sample to choose the corresponding training samples, the final forecast results will be more accurate.

According to the multi-class SVM classification method based on binary tree, for the three categories of training samples to obtain two SVM models. For the first training time, we trained all the training data, divided the training samples into three categories, set the matched expectations values of the first and the third categories to "-1", set the second category corresponds to the expected value of all to 1, and then to call svmtrain function to establish SVM1 model. And use the SVM1 model to separate the "nomal" class. For the Second training time, we trained the first and the third categories in the training samples, reset their expectation values, all first class corresponds to the expectations set to -1, all third category corresponds to the expectations set for "1", and then call svmtrain function to establish SVM2 model. And use the SVM2 model to separate the "mild" and "serious" two classes. The binary tree model is shown in Figure 3.

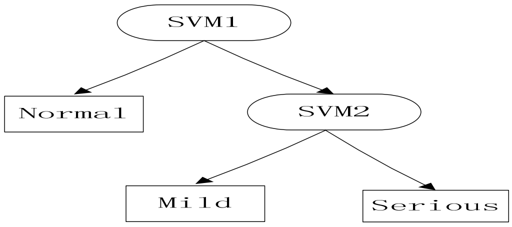

Figure 3. The Binary Tree Model

4.3. Experimental Result and Analysis

In this experiment, 100 little samples are taken for prediction. Each sample contains seven characteristics of CO concentration, wind speed, light intensity, humidity, temperature, traffic volume, and vehicle speed. Using LIBSVM in MATLAB to predict whether a traffic jam will occur, and establishing a prediction model based on SVM.

Table 1. Prediction Results of the Experiment

CO Concentration	Wind Speed	Light Intensity	Humidity	Temperature	Traffic Volume	Vehicle Speed	Prediction	Actual situation
0.72	3.31	180.14	30.51	20.87	15	55	Normal	Normal
0.72	3.61	170.13	30.51	20.87	14	54	Normal	Normal
0.72	3.21	180.14	30.51	20.87	16	60	Normal	Normal
0.72	3.29	170.13	25.3	21.75	13	46	Serious	Serious
0.72	3.31	180.14	25.18	21.75	7	46	Serious	Serious
0.72	4.07	160.12	25.3	21.87	16	39	Serious	Serious
0.72	4.25	170.13	25.3	21.87	14	47	Serious	Serious
0.72	3.74	110.08	19.13	22.87	13	46	Serious	Serious
0.72	3.73	110.08	19.93	22.87	33	25	Serious	Serious
0.72	3.1	120.09	19.25	22.87	24	20	Serious	Serious
0.72	3.43	180.14	30.59	20.75	12	53	Mild	Normal
0.72	3.68	180.14	30.59	20.87	17	50	Mild	Mild
0.72	2.79	190.15	30.47	20.75	16	60	Normal	Normal
0.72	3.13	180.14	30.59	20.87	18	55	Normal	Normal
0.72	2.42	180.14	30.51	20.87	17	54	Normal	Normal
0.72	3.31	180.14	30.51	20.87	14	55	Normal	Normal
0.72	3.61	170.13	30.51	20.87	14	54	Normal	Normal
0.72	4.32	180.14	30.59	20.87	23	50	Mild	Mild
0.72	3.21	180.14	30.51	20.87	14	60	Normal	Normal
0.72	2.96	180.14	30.63	20.87	12	54	Normal	Normal

In this experiment, four training samples: spring training data, summer training data, autumn training data, winter training data are established, and each training sample contains 200 little samples.

These feature items of the test sample can be determined that these data are measured in the summer, and then select the summer test sample to build two SVM model, according to the two SVM models to predict these test data, the precision is 97.5%. The part prediction results of the experiment are shown in Table 1.

Contrast the prediction results and the actual results we can find that use this method to predict is better. If we choose other training samples, the predicted results are not so accurate. For example, we have chosen the winter training samples to predict, the precision is 22.5%. The part prediction results of the experiment are shown in Table 2. Because in different time, the data of these feature items we measured is also different, so four training samples based on time are set up.

Table 2. Prediction Results of the Experiment

CO Concentration	Wind Speed	Light Intensity	Humidity	Temperature	Traffic Volume	Vehicle Speed	Prediction	Actual situation
0.72	3.31	180.14	30.51	20.87	15	55	Mild	Normal
0.72	3.61	170.13	30.51	20.87	14	54	Mild	Normal
0.72	3.21	180.14	30.51	20.87	16	60	Mild	Normal
0.72	3.29	170.13	25.3	21.75	13	46	Mild	Serious
0.72	3.31	180.14	25.18	21.75	7	46	Mild	Serious
0.72	4.07	160.12	25.3	21.87	16	39	Mild	Serious
0.72	4.25	170.13	25.3	21.87	14	47	Mild	Serious
0.72	3.74	110.08	19.13	22.87	13	46	Mild	Serious
0.72	3.73	110.08	19.93	22.87	33	25	Mild	Serious
0.72	3.1	120.09	19.25	22.87	24	20	Mild	Serious
0.72	3.43	180.14	30.59	20.75	12	53	Mild	Normal
0.72	3.68	180.14	30.59	20.87	17	50	Mild	Mild
0.72	2.79	190.15	30.47	20.75	16	60	Mild	Normal
0.72	3.13	180.14	30.59	20.87	18	55	Mild	Normal
0.72	2.42	180.14	30.51	20.87	17	54	Mild	Normal
0.72	3.31	180.14	30.51	20.87	14	55	Mild	Normal
0.72	3.61	170.13	30.51	20.87	14	54	Mild	Normal
0.72	4.32	180.14	30.59	20.87	23	50	Mild	Mild
0.72	3.21	180.14	30.51	20.87	14	60	Mild	Normal
0.72	2.96	180.14	30.63	20.87	12	54	Mild	Normal

Compare the two experimental results shown in Table 1 and Table 2, we found that the selection of different training samples to establish the SVM model, the final prediction results are also different.

Because in different time, the data of these feature items we measured is also different, so we set up four training samples based on time. According to the training sample data, we determine which period the test data belongs to, and then select the corresponding training samples to build the model, according to the model to predict the test data.

If the training samples based on time isn't divided and only one training sample is established, in the same test sample data, the result precision is 70%. The part prediction results of the experiment are shown in Table 3.

Compare the experimental results, we find that the results obtained by establishing the training samples based on time are more accurate. Because in different time, the data of these feature items we measured is also different, so four training samples based on time are set up. According to the training sample data, we determine which period the test data belongs to, and then select the corresponding training samples to build the model, according to the model to predict the test data, and finally get the accurate results.

Table 3. Prediction Results of the Experiment

CO Concentration	Wind Speed	Light Intensity	Humidity	Temperature	Traffic Volume	Vehicle Speed	Prediction	Actual situation
0.72	3.31	180.14	30.51	20.87	15	55	Mild	Normal
0.72	3.61	170.13	30.51	20.87	14	54	Normal	Normal
0.72	3.21	180.14	30.51	20.87	16	60	Normal	Normal
0.72	3.29	170.13	25.3	21.75	13	46	Serious	Serious
0.72	3.31	180.14	25.18	21.75	7	46	Mild	Serious
0.72	4.07	160.12	25.3	21.87	16	39	Serious	Serious
0.72	4.25	170.13	25.3	21.87	14	47	Serious	Serious
0.72	3.74	110.08	19.13	22.87	13	46	Mild	Serious
0.72	3.73	110.08	19.93	22.87	33	25	Mild	Serious
0.72	3.1	120.09	19.25	22.87	24	20	Mild	Serious
0.72	3.43	180.14	30.59	20.75	12	53	Mild	Normal
0.72	3.68	180.14	30.59	20.87	17	50	Mild	Mild
0.72	2.79	190.15	30.47	20.75	16	60	Normal	Normal
0.72	3.13	180.14	30.59	20.87	18	55	Normal	Normal
0.72	2.42	180.14	30.51	20.87	17	54	Normal	Normal
0.72	3.31	180.14	30.51	20.87	14	55	Mild	Normal
0.72	3.61	170.13	30.51	20.87	14	54	Normal	Normal
0.72	4.32	180.14	30.59	20.87	23	50	Serious	Mild
0.72	3.21	180.14	30.51	20.87	14	60	Normal	Normal
0.72	2.96	180.14	30.63	20.87	12	54	Mild	Normal

5. Optimization Problems

Use the multi-class SVM classification method based on binary tree, for K-class training samples, training K-1 support vector machine. The number of two class support vector machine is less, eliminate the situation that belongs to several classes or do not belong to any kind when decision-making. There is not unclassifiable regions, classification does not need to traverse all the classifiers, the classifier according to the hierarchy of the tree from top to bottom, the number of support vectors in classifier is reduced layer by layer, when training classifier the required training samples is less, training time is less, the classification efficiency is higher.

In addition, LIBSVM is a simple, easy to use, and efficient SVM pattern recognition and regression software package, parameters involved in the regulation of SVM is relatively small, provides default parameters, using the default parameters can solve many problems, and provides Cross Validation function. The software can solve the C-SVM classification,-SVM classification, -SVM regression and other problems, including many kinds of pattern recognition problems based on the one-to-one algorithm.

In this experiment, four training samples based on time are built and the corresponding training sample is selected to build the SVM model. This new method is more suitable for our city tunnel traffic situation.

6. Conclusion

The experimental study the method that using LIBSVM to predict traffic jams in MATLAB, establishing four training samples based on time, and according to the selected training sample to establish the prediction model based on SVM. The forecast results show that using this method to predict the traffic jam situation in the tunnel is quite accurate. The method has the advantages of generalization ability, high prediction accuracy, training speed, good stability, ease of modeling, etc, and also has a good prospect.

References

[1] DONG Hong-hui, JIA Li-min, SUN Xiao-liang, LI Chen-xi, QIN Yong, GUO Min. Traffic State Forecasting Towards Urban Freeway Based on the Maximum Entropy Model. *Journal of Transportation Systems Engineering and Information Technology*. 2010; 10(2): 112-116.

[2] LIU Xiao-ling, JIA Peng, WU Shan-hua, YU Bin. Short-Term Traffic Flow Forecasting Based on Multi-Dimensional Parameters. *Journal of Transportation Systems Engineering and Information Technology*. 2011; 11(4): 140-146.

[3] Wu Xinhong, Jin Hai. Grid parallel computation of online traffic status prediction using generalized neural network. *Journal of Huazhong University of Science and Technology*. 2009; 37(5): 49-52.

[4] Yu Hui, Zhang Weihua. Research on real-time and dynamic urban traffic: Information service system, *Telkomnika*. 2012; 10(4): 806-811.

[5] Suicheng Gu, Yuhong Guo. Learning SVM Classifiers with Indefinite Kernels. *AAAI*. 2012: 942-948.

[6] Thanes Wassantachat, Zhidong Li, Jing Chen, Yang Wang, Evan Tan. Traffic Density Estimation with On-line SVM Classifier. *AVSS*. 2009: 13-18.

[7] Rajeev Kumaraswamy, Lekhesh V Prabhu, K Suchithra, PS Sreejith Pai. SVM Based Classification of Traffic Signs for Realtime Embedded Platform. *ACC.* 2011: 193: 339-348.

[8] Zhao Lihong, Song Ying, Zhu Yushi, Zhang Cheng, Zheng Yi. Face Recognition based on multi-class SVM. *CCDC*. 2009: 5871-5873.

[9] Xiangyang Wang, Jing-Wei Chen, Hong-Ying Yang. A new integrated SVM classifiers for relevance feedback content-based image retrieval using EM parameter estimation. *ASC.* 2011; 11(2): 2787-2804.

[10] Rongxiu Lu, Hui Yang, Kunpeng Zhang. Component Content soft-sensor of SVM based on Ions color characteristics. *Telkomnika.* 2012; 10(6): 1445-1452.

[11] Xiaofeng Wang, Yuping Qin. SVM multi-class classification algorithm based on binary tree. *Journal of Hunan Institute of Engineering*. 2008; 11(3): 68-70.