Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 418303, 8 pages
http://dx.doi.org/10.1155/2013/418303

Research Article

Accurate Multisteps Traffic Flow Prediction Based on SVM

Zhang Mingheng,[1] Zhen Yaobao,[2] Hui Ganglong,[3] and Chen Gang[4]

[1] *School of Automotive Engineering, Dalian University of Technology, Dalian 116024, China*
[2] *School of Mechanical Engineering, Dalian University of Technology, Dalian 116024, China*
[3] *School of Navigation, Dalian Maritime University, Dalian 116026, China*
[4] *Department of Mechanical and Manufacturing Engineering, Aalborg University, 169220 Aalborg, Denmark*

Correspondence should be addressed to Zhang Mingheng; gloriazhang@163.com

Received 29 August 2013; Accepted 12 September 2013

Academic Editor: Rui Mu

Copyright © 2013 Zhang Mingheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate traffic flow prediction is prerequisite and important for realizing intelligent traffic control and guidance, and it is also the objective requirement for intelligent traffic management. Due to the strong nonlinear, stochastic, time-varying characteristics of urban transport system, artificial intelligence methods such as support vector machine (SVM) are now receiving more and more attentions in this research field. Compared with the traditional single-step prediction method, the multisteps prediction has the ability that can predict the traffic state trends over a certain period in the future. From the perspective of dynamic decision, it is far important than the current traffic condition obtained. Thus, in this paper, an accurate multi-steps traffic flow prediction model based on SVM was proposed. In which, the input vectors were comprised of actual traffic volume and four different types of input vectors were compared to verify their prediction performance with each other. Finally, the model was verified with actual data in the empirical analysis phase and the test results showed that the proposed SVM model had a good ability for traffic flow prediction and the SVM-HPT model outperformed the other three models for prediction.

1. Introduction

1.1. Backgrounds. Accurate traffic flow prediction is an important research content in intelligent transportation system (ITS). The traffic flow information predicted rapidly and accurately is essential for traffic control, guidance, and providing traffic information services to the public. Especially in recent years, urban traffic congestion are now becoming the major problems in the traffic management all over the world, which produce a series influence hindering the social sustainable development and people's daily work. The transit agencies also have realized that the rapid and accurate traffic information can help them efficiently make reasonable and effective traffic guidance strategy. Thus, there is a growing demands in providing accurate traffic flow duration and diffusion prediction over a period of time.

The urban transport system has the characteristics such as nonlinear, stochastic, and time-varying. In order to achieve accurate prediction, some scholars applied traffic flow model to explain the dynamic changes and evolution of traffic flow. However, the traffic flow model is relatively complex in construction and has some difficulties with practical application. Therefore, the models based undetermined predictive method encounters the problems with model construction and solving. In contrast, nonmathematical models such as nonparametric regression, neural networks, and SVM are now widely applied to the traffic flow prediction due to their characteristics of self-learning and without complicated mathematical model construction.

On the other hand, most of the traditional prediction is based on the single-step method, which can only provide the current or single-step traffic parameters with obtained traffic data. The information provided is not enough for the public or traffic agencies' making decision. Thus, multisteps traffic flow prediction is essential for obtaining the traffic state trends over a certain period in the future. From the perspective of dynamic decision, the development trends of traffic states within a certain time in the future are more important than the current traffic state. For example, the traffic congestion is considered occurred with the significant differences between the obtained traffic data and prediction trends. So we can estimate the range of possible spread of congestion and duration according to the traffic parameters of multistep prediction results.

However, accurate prediction of traffic flow is very difficult due to many stochastic variables involved (e.g., traffic conditions). Therefore, the deployment of traffic flow prediction model is a challenging task.

1.2. Literature Review. Accurate real-time traffic flow prediction is prerequisite and key to realize intelligent traffic control and guidance, and it is also the objective requirement to intelligent traffic management. Over the past decades, various sophisticated techniques and algorithms have been developed for traffic flow prediction. These methods can be roughly categorized as prediction methods based on mathematics and physics, including the historical average model, time series model, Kalman filter model, and exponential smoothing model; the other methods based on nonmathematical models such as artificial neural network (ANN), nonparametric regression (NPR), and SVM. Here, only a brief introduction about the typical method is made, more detailed information is found in their literature, respectively.

1.2.1. Kalman Filter Models. The Kalman filter, also known as linear quadratic estimation (LQE), is an efficient recursive procedure that estimates the future states of dependent variables. It is originated from the state-space representations in modern control theory. Wang et al. [1] developed a traffic state estimator model based on stochastic macroscopic traffic flow modeling and extended Kalman filtering. One major innovative feature of the traffic state estimator is the online joint estimation of important model parameters (free speed, critical density, and capacity) and traffic flow variables (flows, mean speeds, and densities). Hage et al. [2] proposed and validated an algorithm for estimating the urban links travel times based on an unscented Kalman filter (UKF). The algorithm integrates stochastically the vehicle count data from underground loop detectors at the end of every link and the travel times from probe vehicles. Their results showed that the proposed methodology can be used for estimating travel time in real time. Yuan et al. [3] proposed a state estimator based on the EKF technique, in which observation models for both Eulerian and Lagrangian sensor data (from loop detectors and vehicle trajectories, resp.) are incorporated. The results indicate that the Lagrangian estimator is significantly more accurate and offers computational and theoretical benefits over the Eulerian approach.

1.2.2. Support Vector Machine Models. Support vector machine, a novel supervised learning method used for classification and regression, has been recently proved to be a promising tool for both data classification and pattern recognition [4–6]. SVM shows very resistant to the overfitting problem, achieving high generalization performance in solving various time series forecasting problems, which has been applied in prediction of time series [7, 8]. These successful applications motivate researchers to apply SVM in the intelligent traffic system (ITS). Yang and Yu [9] proposed a freeway dynamic traffic flow forecasting model bases on SMO support vector machine. With the analysis of the freeway macroscopic dynamic traffic flow model and carrying out detail research on selecting parameter of SMO support vector machines, a test experiment was carried out and the result showed that the average relative error of forecasting was less than 3.84%. Yu et al. [6, 10] developed the SVM-based models to predict bus arrival time. In the models, travel speeds of the preceding buses of the same bus route were used to estimate traffic conditions. Their results showed that the SVM-based models outperformed the ANN and historic mean prediction models, and SVM seemed to be a powerful alternative for bus arrival time prediction. With versatile parallel distributed structures and adaptive learning processes, ANN and SVM have recently been gaining popularity in bus travel/arrival time prediction. Yu et al. [11] applied the SVM to predicting bus travel times. In which, a decay factor was introduced to adjust the weights between new and old data and the test results showed that the SVM with the decay factor and the adaptive algorithm had better prediction accuracy and dynamic performance than other existing algorithms.

In summary, due to the characteristics of urban traffic state such as uncertainty, nonlinearity and complexity, some researchers use traffic flow model to illustrate the dynamic changes of traffic state, to predict the evolution of the traffic flow, and then achieved the short-term traffic flow prediction model. However, the structure of the traffic flow model is complex relatively, which brings difficulty to the practical application. The method based on mathematical model is hard to meet the practical real-time requirements and the need for accuracy due to its difficulties in model building and solving. In contrast, the method based on nonmathematical model does not need complex model building, the prediction accuracy can meet the requirements of the intelligent transportation system. Therefore, these methods have been widely applied to the traffic flow prediction.

1.3. Contributions. Urban road traffic conditions are not only closely related to historical period conditions, but also to the upstream and downstream road state. At present, most of the accurate traffic flow prediction methods are focusing on the relation analysis between the traffic conditions and the historical traffic data from the time dimension of view. While the spatial dimension such as the upstream and downstream traffic condition and the traffic mode changes daily are ignored. In addition, most of the current traffic flow predictions are concentrated on the single-step prediction, less in the multisteps prediction. In other words, the prediction concentrates more on the coming traffic flow estimation for a specific time point, not concerns with the traffic condition changing trends in the next certain time periods. But in fact, from the perspective of dynamic decision making, the traffic condition changing trends are more important than the current traffic situation for the traffic management. Therefore, it is absolutely essential to establish a multisteps model for the traffic flow prediction, which can be used for estimating the road traffic status trends accurately and the prediction result can be applied in improving rationality of the traffic management and the travel decisionmaking.

This paper seeks to make two contributions to the literature. Firstly, it attempts to develop the models to predict multisteps traffic flow with multiple steps using real-world data. It is expected to help the transit agencies efficiently make reasonable and effective traffic guidance strategy. Secondly, in order to improve the prediction accuracy, not only the historical traffic data but also the daily space-time sequences data are taken into consideration during the input state vector constructing. The performance of the proposed model can provide valuable insight for researchers as well as practitioners.

The structure of this paper is organized as follows: Section 2 provides the accurate real-time prediction model for predicting traffic flow with multisteps; Section 3 presents a case study together with results and analysis including performance evaluation of the proposed model; and lastly, the conclusions are given in Section 4 together with suggestions for further study.

2. Methodologies

2.1. Support Vector Machine Models. SVM is a type of learning algorithm based on statistical learning theory, which can be adjusted to map the complex input-output relationship for the nonlinear system without dependent on the specific functions. Unlike other nonlinear optimization methods, the solution of SVM always can achieve the global optimal solution without limitation to a local minimum point and it shows the strong resistance to the over-fitting problem and the high generalization performance.

Given the samples $(x_1, y_1), (x_2, y_2), \ldots, (x_i, y_i)$ $(x_i \in X \subseteq R^n,\ y_i \in Y \subseteq R)$, SVM make use of a nonlinear mapping ϕ to map x into a high-dimensional feature space H in which linear approximation is conducted to find the mapping function so that we can get a better approximation for the given data samples. Based on the statistical learning theory, we can get the function as follows:

$$f(x) = \omega\phi(x) + b. \tag{1}$$

Regression can be defined as a problem that minimized the risks for a loss function. The optimal regression function is the minimum and regularized generic function Q under certain constraints:

$$Q = \frac{1}{2}\|\omega\|^2 + \frac{C}{l}\sum_{i=1}^{l} L_{\varepsilon}(y_i, f(x_i)), \tag{2}$$

where, ω is a standard vector. The first item is named as the regularized term, which makes the function flat and improve its generalization ability; the second item is named as the experience risk generic functional, which can be determined by different loss functions; C is used to balance the relationship between structure risks and experience risks $(C > 0)$. When the insensitive loss function is defined with

$$L_{\varepsilon}(y_i, f(x_i)) = \max(|y_i - f(x_i)| - \varepsilon, 0). \tag{3}$$

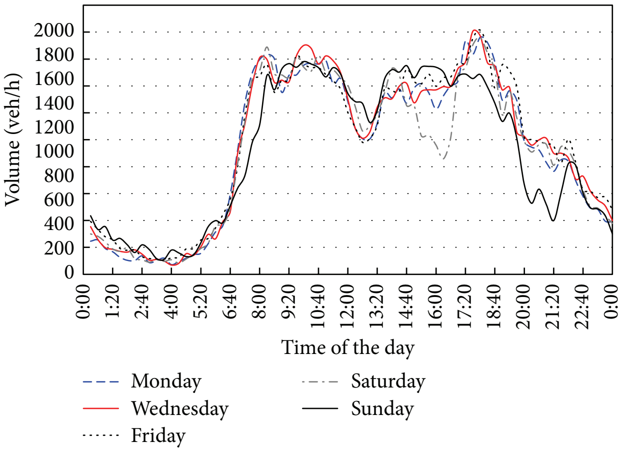


FIGURE 1: Traffic volume diurnal variation trends.

The minimization of (2) is a convex quadratic optimization problem, and then the problem can be inferred with the Lagrange multipliers a_i and a_i^*, shown as following:

$$\omega - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i = 0. \tag{4}$$

Then, the SVM decision function is

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\phi(x_i)\cdot \emptyset(x) + b\right). \tag{5}$$

The following equation can be obtained:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x_j) + b, \tag{6}$$

where $K(x_i, x_j) = \phi(x_i)$ denotes the inner product of the vector in the feature space. All of the kernel functions can be directly computed in the input space.

2.2. Multisteps Prediction Methods Based on SVM. Traffic flow data are time-series data which have the characteristics such as autocorrelation and historical changes with similarity. Figure 1 shows the traffic volume changes for a certain road section during a few days in a week. In which, during the week days, the traffic volume changes gradually. That is to say the traffic volume data have the autocorrelation properties which can be demonstrated with the relationships between the traffic flow data at time t and before. Therefore, the prediction of the traffic flow can be conducted according with this autocorrelation properties.

Furthermore, other some cues can be inferred. Firstly, the traffic volume changes with similarity each day due to the inhabitant regular daily traveling. The data waveform presents like a saddle and the peak/valley of the wave appeared in the same time. In this paper, this regular changing pattern is named as the historical change of similarity, which can be used for establishing the traffic flow changing model to predict future multisteps traffic state.

Secondly, although the daily traffic volume has the same changing trends, the models have different properties from each other. For example, the traffic volume curve are similar from Monday to Friday, but on Saturday and Sunday, the curve presents different pattern from the week-days, the peak/valley of traffic volume curve is not obvious for distinguishing.

Therefore, a historical data model for prediction should be constructed which can be used for predicting the traffic flow historical models for every day (from Monday to Sunday). When the database is not enough for predicting, we also can establish the models for week-days and week ends, respectively. In addition, the trends of traffic flow data also can be influenced by weather or holidays.

Besides the above demonstration about the traffic flow properties in time domain, the data are also correlated in space domain. The traffic flow data has some correlations between the upstream and downstream for a certain road section. The upstream traffic flow will reach the downstream after a while; therefore, we can predict the future traffic flow condition of the downstream based on this correlation.

Given the traffic volume $q_{i,j}(t)$ of road section i, time t and day j, the observation volume sequence $Q_{i,j} = \{q_{i,j}(1), q_{i,j}(2), \ldots, q_{i,j}(t), \ldots\}$, the prediction of the traffic volume $\widehat{q}_{i,j}(t)$, then the forecasting traffic volume sequence of road section in the future n steps are expressed as $\widehat{Q}_{i,j} = \{\widehat{q}_{i,j}(t), \widehat{q}_{i,j}(t+1), \ldots, \widehat{q}_{i,j}(t+n)\}$. The historical pattern data sequence is defined as $HQ_{i,k} = \{hq_{i,k}(1), hq_{i,k}(2), \ldots, hq_{i,k}(t), \ldots\}$, in which $hq_{i,k}(t) = (1/w)\sum q_{i,j}(t)$ is the historical data of section i on the kth day (from Monday to Sunday), and w is the number of weeks for the data included.

Based on the description above and the space-time characteristics of the traffic flow data, in order to predict the traffic flow for the n steps $(t+1, \ldots, t+n)$ at section i, time t, three input feature vectors are designed for the prediction model:

(1) autocorrelation time feature vector: $Q_{i,j} = \{q_{i,j}(t-m), \ldots, q_{i,j}(t)\}$;

(2) historical model feature vector: $HQ_{i,k} = \{hq_{i,k}(1), \ldots, hq_{i,k}(t-1), hq_{i,k}(t), hq_{i,k}(t\ +1), \ldots, hq_{i,k}(t+n)\}$, $k = 1, 2, \ldots, 7$, which represents from Monday to Sunday, respectively;

(3) space correlation feature vector: $Q_{i-1,j} = \{q_{i-1,j}(t-m), \ldots, q_{i-1,j}(t)\}$, $Q_{i-2,j} = \{q_{i-2,j}(t-m), \ldots, q_{i-2,j}(t)\}, \ldots$. The number of time sequence m is determined by the time consumption from the upstream to the current road section.

The relationship between each input vector and output vector is shown in Figure 2.

From the aspect of theoretical analysis, for the single-step mode, accurate results can be obtained with only consideration of the time state and space state vectors due to the gradient traffic flow characteristics. On the contrary, for the multisteps mode, the traffic flow changing trends play a more important role in prediction; therefore, more accurate results can be achieved by taking the historical data into account.

Based on the above analysis, the combination of three state vectors is used as the input data for SVM, which are as follows:

(1) the combination of one-dimensional data for the specified road section in the previous intervals (T): $Q_{i,j} = \{q_{i,j}(t-m), \ldots, q_{i,j}(t)\}$, which describes the traffic volume changing property with the time intervals, this property is similar to the single-step prediction method;

(2) the combination of multidimensional space-time data of the upstream and the specified road section (PT): $Q_{i-1,j} = \{q_{i-1,j}(t-m), \ldots, q_{i-1,j}(t)\}$, $Q_{i-2,j} = \{q_{i-2,j}(t-m), \ldots, q_{i-2,j}(t)\}$, which describes the correlation of the traffic flow changing in time domain, especially the correlations between the upstream and downstream for a certain road section;

(3) the combination of the time-series data and the historical pattern data for the specified road section in the current interval (HT): $Q_{i,j} = \{q_{i,j}(t-m), \ldots, q_{i,j}(t)\}$, $HQ_{i,k} = \{hq_{i,k}(1), \ldots, hq_{i,k}(t-1), hq_{i,k}(t), hq_{i,k}(t+1), \ldots, hq_{i,k}(t+n)\}$, which describes the historical similarity of the traffic volume changing property demonstrated in the Figure 1;

(4) the combination of the time-series data for the upstream and the specified road section in the current interval and the historical pattern data for the target road section (HPT).

Each of the combination vectors above are used as the input variable of the SVM model, and $\widehat{Q}_{i,j} = \{\widehat{q}_{i,j}(t), \widehat{q}_{i,j}(t+1), \ldots, \widehat{q}_{i,j}(t+n)\}$ as the output variable, the SVM-T, SVM-TP, SVM-TH and SVM-TPH can be achieved respectively during the SVM training process.

3. Case Study

3.1. Evaluation of Model Performance. As an indicator to reflect the accuracy and availability of the prediction model, error plays an important role in evaluating the prediction model. Common error indictors include absolute prediction error and relative prediction error, which can be calculated as follows, respectively:

$$\begin{aligned} E_a(t) &= q_i(t) - \widehat{q}_i(t), \\ E_r(t) &= \frac{q_i(t) - \widehat{q}_i(t)}{q_i(t)}, \end{aligned} \tag{7}$$

where $E_a(t)$ denotes the absolute prediction error at time t, $E_r(t)$ denotes the relative prediction error, $q_i(t)$ denotes the measured value, and $\widehat{q}_i(t)$ denotes the prediction value. $E_a(t)$ and $E_r(t)$ are the errors of the single-step prediction, while for the dynamic multisteps prediction, both the value and error can be achieved during each prediction process, which can be described with $\widehat{q}_i(t+1), \widehat{q}_i(t+2), \ldots, \widehat{q}_i(t+n)$. $\widehat{q}_i(t+1)$ is the prediction value of $(t+1)$ at the current time t, where n is the prediction step size. The greater the n, the more prediction

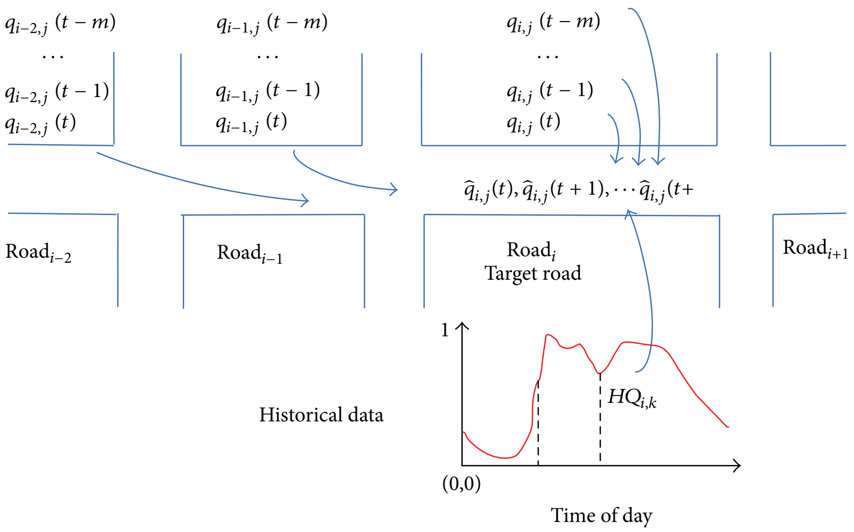


Figure 2: Prediction mechanism based on the traffic flow space-time characteristics.

error we can achieve. But a maximum theoretical limit $n_{\max}$ exists for avoiding unlimited prediction loop.

In order to obtain an objective and accurate evaluation of the prediction effect, the mean value of all the errors is used as the evaluation index for multisteps prediction, the formulas are as follows:

$$ME_a(t) = \frac{1}{n}\sum_{t=t+1}^{n}\left(q_i(t) - \widehat{q}_i(t)\right),$$
$$ME_r(t) = \frac{1}{n}\sum_{t=t+1}^{n}\left(\frac{q_i(t) - \widehat{q}_i(t)}{q_i(t)}\right), \tag{8}$$

where $ME_a(t)$ is the mean value of the absolute error for multisteps prediction at time t, $ME_r(t)$ is the mean value of the relative error. In fact, for multisteps prediction, the time-series synthetical error is more important for the prediction result. Therefore, $ME_a(t)$ and $ME_r(t)$ are the direct sum rather than the sum of absolute values of $E_a(t)$ and $E_r(t)$.

3.2. Data Collection and Processing. The presented SVM model for traffic flow prediction was tested with the actual survey data of Gaoerji Road in Dalian, China, dated from May 14 (Monday) to 20 (Sunday), 2012. Gaoerji road is a one-way lane with several imports/exports and goes from Zhongshan road to Yierjiu street through the city center with a distance of 3.8 km. The traffic state is highly congested in the morning and afternoon peaks. In this research, the actual survey started from 7:00 in the morning. The collected data was obtained with SCOOT (Split, Cycle, and Offset Optimization Technology) and consist of the traffic volume during the peak time (PT; 0700–0830 h) and off-peak time (OT; 1000–1200 h) with recording data once every 5 minutes.

There are some influence factors including missing, error and random noise in the raw datasets which is obtained with SCOOT. Therefore, the data must be identified, repaired and smoothed firstly. Then the time-series data is achieved with normalization method. The main advantage of data normalization is to accelerate convergence velocity of iteration during the SVM training. Another advantage is to facilitate subsequent data processing. Singular value sample data might cause numerical problems because the kernel values usually depend on the inner products of feature vectors such as the linear kernel or the polynomial kernel. Therefore, it is recommended that each attribute should be linearly scaled to the range $[-1, +1]$ or $[0, +1]$. In this research, the data sets were scaled to the range between 0 and 1.

Table 1: Four popular kernel functions.

Kernel function	Expression	Comment
Linear kernel	$K(x_i, x_j) = x_i^T x_j$	
Polynomial kernel	$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$	$\gamma > 0$
Radial basis function (RBF) kernel	$K(x_i, x_j) = e^{-\gamma\|x_i - x_j\|^2}$	$\gamma > 0$
Sigmoid kernel	$K\left(x_i, x_j\right) = \tanh\left(\gamma x_i^T x_j + r\right)^d$	

x_i, x_j are input vectors and γ, r, and d are kernel parameters.

3.3. Model Identification. In Section 2.2 demonstrated above, the input vectors for traffic flow prediction model have been discussed and constructed. Here, we only give an introduction to SVM briefly. More detailed descriptions can be found in [12]. Given a training set of instance-label pairs, the input vectors can be mapped implicitly and hence very efficiently into high-dimensional space (maybe infinite) by the kernel function Φ, and SVM finds a linear separating hyper plane with the maximal margin in this higher dimensional space. At present, new kernel functions are proposed by researchers, some popular kernel functions are as Table 1.

The previous researches [13, 14] suggested that radial basis function (RBF) kernel had less numerical difficulties with application and was efficient for traffic state and traffic

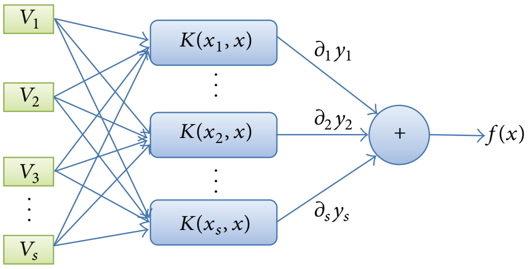


Figure 3: Structure of the SVMs model for traffic flow prediction.

flow prediction. Thus, RBF kernel function is used for the SVM model in this study. Before the RBF kernel is being used, three parameters associated should be determined: (C, ε, γ). Parameter C is a regularization constant or penalty parameter which is assigned in advance before each SVM model training process. This parameter allows striking a balance between the two competing criteria of margin maximization and error minimization, whereas the slack variables ε indicate the distance of the incorrectly classified points from the optimal hyper plane [15]. The larger the C value, the higher the penalty associated to misclassified samples. Parameter γ is the parameter of RBF function kernel. In summary, the entire SVM model training process is a process that the parameters are optimized and determined by simulation. Kavzoglu and Colkesen [16] pointed out that the parameters determination was a practical way to identify good parameters. Thus, in this research, they are calibrated by grid-search method. During the search process, all possible combinations of (C, ε, γ) are tested and the one with the best performance is chosen. After conducting the grid search on the training dataset, the optimal parameters were selected as $(2^{-2}, 2^{-5}, 1.22)$, and a cross validated on validation dataset is conducted. The various SVMs are implemented with Libsvm (National Taiwan University, available at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/) on a Microsoft Windows platform.

The structure of the SVMs model for *SVM-T*, *SVM-PT*, *SVM-HT*, and *SVM-HPT* is shown as Figure 3. In this structure, the feature vector v is composed with the different dimensions of each SVM model, respectively. For example, in *SVM-T model*, $v = (V_1, V_2, \ldots, V_s)$ denotes the traffic volume of road section i and day j at different times $\{q_{i,j}(t - s - 1), \ldots, q_{i,j}(t)\}$. The features vector definitions for each SVMs model are demonstrated in Section 2.2.

To determine the SVM's inputs vector for practical application, some comparison tests have been conducted between the models with different input variables (T, PT, HT, and HPT) which have been calibrated based on the data collected. The comparison of prediction errors about the four models on off-peak period and peak period prediction error results are shown as Figure 4.

During the whole prediction process, the data processing was divided into three steps: respectively for training, cross-validation, and testing. Firstly, about 10% of samples data were set as testing data. Then, 70% of the remaining samples data were assigned to training and the others to cross-validation. In particular, the training and testing processes were conducted with the same datasets for the four models in order to have the same basis of comparison.

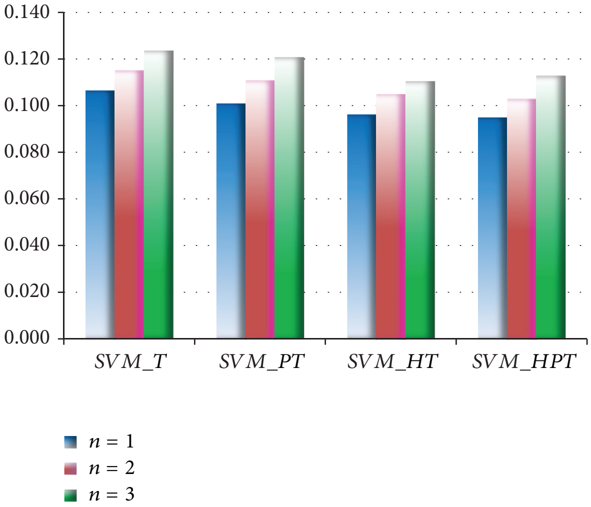


Figure 4: Prediction error results of the SVM model with different input vectors.

3.4. Results Analysis. From the comparison results of the four models (*SVM-T*, *SVM-PT*, *SVM-HT,* and *s-HPT*) in Section 3.3, some conclusions can be drawn as follows.

(1) Firstly, in the case of single-step prediction, the number of prediction steps $n = 1$. The prediction error curves in Figure 4 indicate that the results of the prediction error are similar with each other for the four models and *SVM_PT* has more accuracy than *SVM_T*. The reason of this case is that the upstream of the traffic volume can reach the current road section after a while, so a more accurate prediction result can be obtained with the introduction of upstream traffic volume into *SVM-T*.

(2) Secondly, in the case of multisteps prediction, the number of prediction steps $n > 1$. For the case of $n = 2$, the prediction error has the similar characteristics to single-step model. For the case of $n > 2$, the prediction error of *SVM_T* increases significantly with the steps n increasing. The reason for this is that in *SVM_T* model, the traffic volume data is time-series data, the distance between the predicted points and observing data has the different impact effect for the prediction accuracy, the more closely the more important for prediction result. Furthermore, for the multisteps prediction, the *SVM_HT* and *SVM_HPT* have a good prediction performance compared with the *SVM_T* which only uses the time-series data for predicting. The reason for this is mainly that the historical data can represent the traffic trends changing mode with traffic volume for the current road section, so the prediction result can be better with the introduction

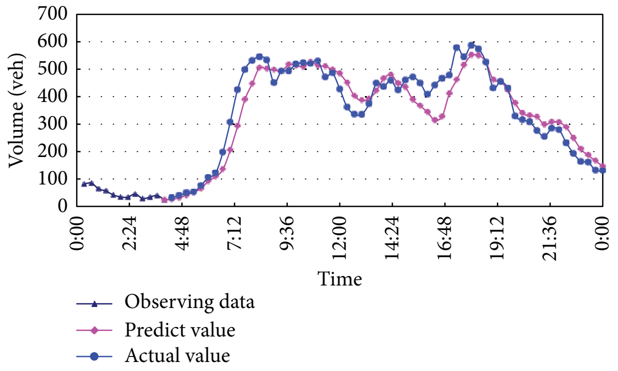


Figure 5: Example of dynamic multisteps prediction results of *SVM_PT*.

of upstream traffic volume data and historical data into *SVM-T*.

(3) It is important to note that the execution performance of *SVM_HPT* and *SVM_HT* is different due to the complicated input vector dimensions in upstream traffic volume data and historical data. Therefore, in practical application, the *SVM_HT* model is a priority selection for prediction under the condition of meeting the requirement prediction accuracy in order to reduce the time-consumption of the whole system.

In order to verify the actual performance of the prediction model presented above, a test was conducted with *SVM_HT* and the result was shown as Figure 5. In this test, a certain whole day traffic volume data were used for predicting the traffic trends and the sample interval was assigned to 20 minutes so as to simplify the computation process. From the result, it can be seen that the prediction curves agree well with the actual observing data. Test of the prediction error found that the traffic volume mean error is 16.8 vehicles and the mean relative error is 12.8%. Thus it can be seen that there seems to be a strong potential effectiveness to be used to the practical application.

4. Conclusions

The urban transport system has the characteristics such as nonlinear, stochastic, and time-varying. Therefore, artificial intelligence methods are now receiving more and more attentions in ITS. In order to predict traffic flow with multisteps accurately, this paper proposed a SVM model for the prediction. In the present research, the traffic volume data with actual observation surveys in urban area of Dalian was used to predict the traffic flow by SVM models. In order to obtain the prediction effect with different input vectors, some comparison tests are conducted with different input variables (T, PT, HT, and HPT). The test results showed that the proposed SVM model had a good ability for traffic flow prediction and the comparison of different input vectors indicated that the *SVM-HPT* model outperformed the other three models for prediction.

In this paper, only the traffic volume data was used to estimate the current traffic conditions. Further study will consider more factors analysis such as the input vectors dimensions and the prediction steps so as to enhance the performance of the proposed prediction models.

Acknowledgments

This work was jointly supported by grants from the Humanities and Social Sciences Foundation of the Ministry of Education in China (Project no. 12YJCZH280), the Ph.D. Programs Foundation of Ministry of Education of China (Project no. 20112125120004), and National Natural Science Foundation of China (Project no. 11272075).

References

[1] Y. Wang, M. Papageorgiou, and A. Messmer, “Real-time freeway traffic state estimation based on extended Kalman filter: adaptive capabilities and real data testing,” *Transportation Research A*, vol. 42, no. 10, pp. 1340–1358, 2008.

[2] R. M. Hage, D. Betaille, F. Peyret, and D. Meizel, “Unscented Kalman filter for urban network travel time estimation,” *Procedia—Social and Behavioral Sciences*, vol. 54, no. 4, pp. 1047–1057, 2012.

[3] Y. Yuan, J. W. C. van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn, “Real-time lagrangian traffic state estimator for freeways,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 1, pp. 59–70, 2012.

[4] V. N. Vapnik, “An overview of statistical learning theory,” *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.

[5] B. Yu, Z. Z. Yang, and B. Z. Yao, “Bus arrival time prediction using support vector machines,” *Journal of Intelligent Transportation Systems*, vol. 10, no. 4, pp. 151–158, 2006.

[6] B. Yu, W. H. K. Lam, and M. L. Tam, “Bus arrival time prediction at bus stop with multiple routes,” *Transportation Research C*, vol. 19, no. 6, pp. 1157–1170, 2011.

[7] L. J. Cao and F. E. H. Tay, “Support vector machine with adaptive parameters in financial time series forecasting,” *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, 2003.

[8] B. Z. Yao, C. Y. Yang, J. B. Yao, and J. Sun, “Tunnel surrounding rock displacement prediction using support vector machine,” *International Journal of Computational Intelligence Systems*, vol. 3, no. 6, pp. 843–852, 2010.

[9] J. H. Yang and X. N. Yu, “The support vector machines prediction model of freeway dynamic traffic flow,” *Journal of Xi’an Technological University*, no. 3, pp. 280–284, 2009.

[10] B. Yu, Z. Z. Yang, K. Chen, and B. Yu, “Hybrid model for prediction of bus arrival times at next station,” *Journal of Advanced Transportation*, vol. 44, no. 3, pp. 193–204, 2010.

[11] B. Yu, T. Ye, X. M. Tian, G. B. Ning, and S. Q. Zhong, “Bus travel-time prediction with forgetting factor,” *Journal of Computing in Civil Engineering*, 2012.

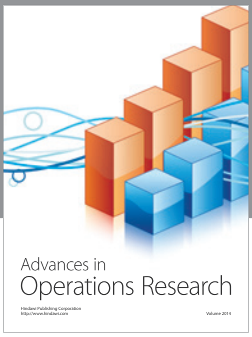
[12] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

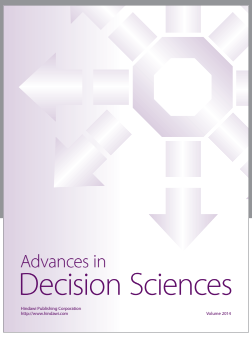
[13] V. Tyagi, S. Kalyanaraman, and R. Krishnapuram, “Vehicular traffic density state estimation based on cumulative road acoustics,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1156–1166, 2012.

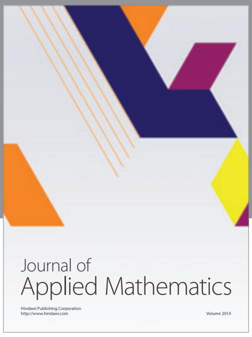
[14] Q. Li, "Short-time traffic flow volume prediction based on support vector machine with time-dependent structure," in *Proceedings of the IEEE Intrumentation and Measurement Technology Conference (I2MTC '09)*, pp. 1730–1733, May 2009.

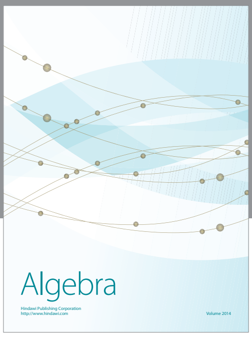
[15] T. Oommen, D. Misra, N. K. C. Twarakavi, A. Prakash, B. Sahoo, and S. Bandopadhyay, "An objective analysis of support vector machine based classification for remote sensing," *Mathematical Geosciences*, vol. 40, no. 4, pp. 409–424, 2008.

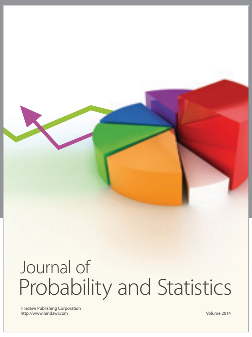
[16] T. Kavzoglu and I. Colkesen, "A kernel functions analysis for support vector machines for land cover classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 11, no. 5, pp. 352–359, 2009.











Hindawi

Submit your manuscripts at
http://www.hindawi.com