Spatio-temporal Variable Selection Based Support Vector Regression for Urban Traffic Flow Prediction

Yanyan Xu*
Department of Automation, Shanghai Jiao Tong University
& Key Laboratory of System Control and Information Processing,
Ministry of Education of China, Shanghai, 200240, China
Phone: (86-21)3420-4328
Fax: (86-21)3420-4340
Email: xustone1987@gmail.com

Bin Wang
Department of Automation, Shanghai Jiao Tong University
& Key Laboratory of System Control and Information Processing,
Ministry of Education of China, Shanghai, 200240, China
Email: binwang@sjtu.edu.cn

Qing-Jie Kong
State Key Laboratory for Management and Control of Complex Systems
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
Email: qjkong.cas@gmail.com

Yuncai Liu
Department of Automation, Shanghai Jiao Tong University
& Key Laboratory of System Control and Information Processing,
Ministry of Education of China, Shanghai, 200240, China
Email: whomliu@sjtu.edu.cn

Fei-Yue Wang
State Key Laboratory for Management and Control of Complex Systems
Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
Email: feiyue.trans@gmail.com

* Corresponding author

5395 text words + 6 figures + 2 tables = 7395 words

Submission date: November 15, 2013

ABSTRACT

Short-term urban traffic flow prediction remains a difficult yet important problem in the intelligent transportation systems (ITS). Most previous spatio-temporal based urban traffic flow prediction techniques just pay attention to building the relationship between the adjacent upstream and downstream road segments using various models. While in this paper, we take advantage of the spatial and temporal information from all available road segments in the road network to predict the short-term traffic volume accurately. However, the available traffic states can be high-dimensional for high-density or large scale road networks. Therefore, we present a spatio-temporal variable selection based support vector regression (VS-SVR) model fed with the high-dimensional traffic data collected from all available road segments. Our prediction framework can be presented as a two-stage model. In the first stage, we employ the multivariate adaptive regression splines (MARS) model to select a set of predictors most related to the target one from the high-dimensional spatio-temporal variables, and reasonable weights are assigned to the selected predictors. In the second stage, the kernel learning method, support vector regression (SVR), is trained on the weighted variables in the second stage for prediction. In the experiments, we employ the actual traffic volume collected from a subarea of Shanghai, China, every 10 minutes. The experimental results indicate that the proposed spatio-temporal variable selection based support vector regression model can generate preferable results in contrast with the time series based autoregression (AR) method, the separate MARS model, and the SVR model.

INTRODUCTION

Along with the frequently occurring traffic congests in urban road network, reliable short-term urban traffic flow prediction is becoming an extremely crucial but tough task in the intelligent transportation systems (ITS), especially in the advanced traveler information systems (ATIS) and the advanced transportation management systems (ATMS). Accurate traffic prediction assists the traffic managers in carrying out reasonable policies to relieve the congest or giving optimal routes to the travelers. In the past few decades, a great number of models have been proposed to predict the urban traffic flow based on historical data or spatio-temporal correlation. In general, the previous methods can be classified into two categories: time series methods and spatio-temporal based methods.

The time series based traffic prediction methods employing the parametric or non-parametric models build the relationship between the historical traffic states and the desired value. Among those methods, the autoregressive integrated moving average (ARIMA) may be the most frequently used parametric prediction method (*1, 2, 3*). In addition, the Kalman filtering was also applied to the prediction of traffic flow (*4*). However, the non-parametric models always perform better than the parametric models as they are more flexible to the nonlinear processes. Some non-parametric regression methods were applied to the traffic prediction successfully, such as k-nearest neighbor (k-NN) approach (*5*), artificial neural network (ANN) approach (*6*), and regression trees (*7*). Although the time series based approaches have limited information, some researchers still work on such methods due to its low cost computation. For example, Tchrakian *et al.* developed a spectral analysis approach for real-time traffic flow prediction (*8*). In summary, owing to considering the traffic flow as time series, most of these approaches could perform well in the freeway or a location on an artery in urban, instead of the complicated urban road network.

Subsequently, researchers introduced the spatial information in the traffic prediction models, especially in urban area. A large variety of multivariate spatio-temporal correlation approaches have been developed to predict traffic flow in urban arteries. For instance, Hobeika *et al.* (*9*) addressed the short-term traffic flow prediction based on the traffic states from current and upstream roads as well as the average average states. Stathopoulos *et al.* (*10*) developed a state space approach which fed with the data from upstream roads to improve on the downstream locations. Djuric *et al.* used the continuous conditional random fields (CCRF) to predict the travel speed based on the spatio-temporal correlations (*11*). Min and Wynter (*12*) developed a real-time traffic prediction method with spatio-temporal correlation based on the vector autoregressive model. Besides, some advanced statistical learning approaches have also been extensively utilized to build the relationship between the upstream and downstream traffic states, such as the support vector regression (SVR) (*13, 14*), Bayesian graphical model (*15*), Gaussian processes (*16*) and so on.

Although a lot of spatio-temporal correlation methods have been proposed for urban traffic flow prediction, most of the previous methods just pay attention to building the relationship between the adjacent upstream and downstream traffic states using various models. While in this paper, we consider the traffic states from all available road segments in the road network to predict the short-term traffic volume on the target road accurately. However, towards the high-density or large road network, the number of the available variables fed into the prediction model can be high-dimensional. Consequently, we propose a spatio-temporal variable selection based support vector regression (VS-SVR) model fed with the high-dimensional traffic data collected from all available road segments in the network. The proposed traffic prediction framework can be presented as a two-stage model. In the first stage, we employ the multivariate adaptive regression splines (MARS) model to select a set of predictors most related to the target one from the high-dimensional spatio-temporal variables, and reasonable weights are assigned to the selected predictors. Afterwards, the SVR model is trained on the weighted variables in the second stage for prediction. In the experiments, the actual traffic volume collected from a subarea of Shanghai, China, every 10 minutes is employed to evaluate the proposed model. The experimental results indicate that the proposed spatio-temporal variable selection based support vector regression model can generate preferable results in contrast with the time series based autoregression (AR), MARS, and SVR model.

The remainder of this paper is organized as follows: section 2 describes the details of the proposed spatio-temporal variable selection based SVR traffic prediction model; section 3 briefly introduces the real traffic data used in our work; the spatio-temporal correlation and the short-term traffic volume prediction results are illustrated and analyzed in Section 4. Moreover, the AR, MARS, and SVR prediction methods are implemented for comparison with the proposed model in this section; finally, some concluding remarks and directions for the future work are given in Section 5.

METHODOLOGY

Variable selection plays an important role in many applications for which datasets with tens or hundreds of variables are available in classification and regression problems. According to Guyon *et. al.*'s suggestion (*17*), the contribution of variable selection is summarized as follows: improving the prediction performance of the predictor, saving computing resources of the prediction methods, and providing a better understanding of the relationship between the predictor and the response.

Towards the urban traffic prediction, the spatio-temporal correlation of the road segments in a road network is more complicated. Although the road segments are not direct adjacent, the traffic states on these segments may interact. Hence in our research, the current and historical traffic states from all available road segments are collected as the input variables of the prediction model. Thus, the independent variables can be high-dimensional, especially when the road network is large and crowded. The prediction of the traffic flow at a target road segment can be treated as a regression problem with high-dimensional input. Therefore, we desire to obtain better prediction results through variable selection on such high-dimensional regression problem.

In this paper, we implement the variable selection using a nonlinear and adaptive model, multivariate adaptive regression splines (MARS). MARS is a kind of embedded variable selection method as it performs variable selection via a built-in mechanism (*18*). The prediction ability of

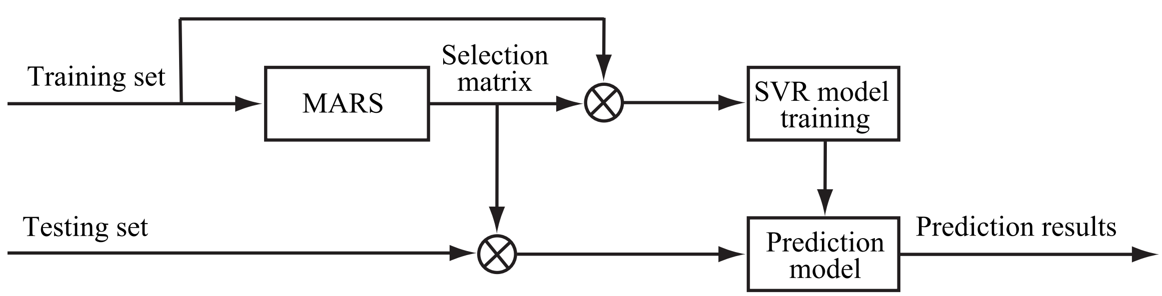


FIGURE 1 The framework of the proposed urban traffic prediction model

MARS model is uncompetitive in contrast with some advanced statistical learning models. Therefore, in our work, MARS is used as a filter which performs variable selection before a learning based prediction method, SVR. The framework of our proposed urban traffic prediction method is illustrated in Figure 1. The details of the methodology are described in the following content.

Spatio-temporal Variable Selection

As we feed the traffic states from all of the road segments at different time intervals into the prediction model and desire to build the relationship between all the predictors and the target, the set of independent variables $\mathbf{X}$ fed into the prediction model is high-dimensional. Suppose that we have N observations, $\mathbf{X} = \{\mathbf{x}_i\}_{i=i}^{N}$, where $\mathbf{x}_i \in \Re^p$ contains p predictors. Then suppose the corresponding response variables are $\mathbf{y} = \{y_i\}_{i=1}^{N}$. To select a subset $\mathbf{X}'$ from $\mathbf{X}$ that can obtain best prediction performance, assume that the response is generated by

$$y_i = g(\mathbf{x}_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \delta^2) \tag{1}$$

where ε_i is the normal error with a mean of zero and a variance of δ^2. If we define $e(\mathbf{X})$ as the estimation error function expressing the error between $\mathbf{y}$ and $g(\mathbf{X})$, the best predictor set $\mathbf{X}'$ should obey the following condition:

$$e(\mathbf{X}') \leq e(\mathbf{X}'') \tag{2}$$

where $\mathbf{X}''$ is any other possible subset of $\mathbf{X}$ except $\mathbf{X}'$. Hence, the critical issues in the variable selection are two fold: the fitting method and the definition of the error function. Considering that MARS is able to remove those variables with negative effect and keep informative variables, we employ MARS to select the significative spatio-temporal variables.

A Brief Introduction to MARS

MARS is a flexible regression model for multivariate regression proposed by Friedman (*18*). As Friedman stated, the objective function $g(\mathbf{x})$ in MARS is assumed to be composed of a series of basis functions, each of which has its support on a distinct region (*19*). The objective function is as shown in the following equation:

$$g(\mathbf{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\mathbf{x}) \tag{3}$$

where β_0 is a constant bias; β_m is the regression coefficient estimated to yield the best fit to the desired relationship between predictor and response; $B_m(\mathbf{x})$ is the basis function; M is the number of basis functions. Generally, $B_m(\mathbf{x})$ can be expressed as the product of spline functions:

$$B_m(\mathbf{x}) = \prod_{l=1}^{L_m} \phi_{m,l}(x_{v(m,l)}) \tag{4}$$

where L_m is the degree of the interaction of basis B_m and $v(m,l)$ is the index of the predictor variable depending upon the mth basis function and the lth spline function. For each m, $B_m(\mathbf{x})$ can consist of a single or a product of two or more spline functions, and no input variable can appear more than once in the product. These spline functions often take the form as follows

$$\phi_{m,l}(x_{v(m,l)}) \in \left\{ b_q^+(x_{v(m,l)} - t_{m,l}), b_q^-(x_{v(m,l)} - t_{m,l}) \right\} \tag{5}$$

with

$$b_q^+(x-t) = [+(x-t)]_+^q = \begin{cases} (x-t)^q, & if\, x > t \\ 0 & otherwise \end{cases} \tag{6}$$

$$b_q^-(x-t) = [-(x-t)]_+^q = \begin{cases} (t-x)^q, & if\, x < t \\ 0 & otherwise \end{cases} \tag{7}$$

where $[.]_+$ denotes the positive part of the argument; x is the predictor split and t is the threshold on the predictor, named knot; q is the power of the spline function.

Variable Selection in MARS Model

The "optimal" variable selection from $\mathbf{X}$ in the MARS model is figured out with a two-stage process: forward growing and backward pruning. In the forward growing stage, MARS initially constructs an overfitted model by adding an excessively large number of basis functions. These basis functions are allowed to interact with one another or be restricted to entry as additive components only. During this stage, MARS uses a greedy algorithm to consider whether to add a basis function to the MARS model by searching all combinations of all values of all variables. The selection of basis functions from the initial set is achieved by determining a constant function. Meanwhile given a configuration for the $B_m(\mathbf{x}_i)$, the coefficients β_m are estimated by minimizing the residual sum of squares (RSS) criterion with the following form

$$RSS(M) = \sum_{i=1}^{N} (y_i - \hat{g}_M(\mathbf{x}_i))^2 \tag{8}$$

where $\hat{g}_M(\mathbf{x}_i)$ is the estimation of y_i with M basis functions; N is the total number of the predictors. Finally, new pairs of functions are considered at each phase until the change of residual error is too small to continue or until the maximum number of basis functions specified at the beginning is reached (*18*).

The second stage in the MARS strategy is a pruning procedure. In this backward stepwise stage, basis functions are deleted according to their contribution to the model, that is, the least contribution, the eariler it is deleted, until an optimal balance of bias and variance is found. The backward removal is performed by suppressing the basis functions that contribute to a minimal residual error. This stage consists of reducing the complexity of the model by increasing its generalizability. This process can be conducted by means of generalized cross validation (GCV):

$$GCV(M) = \frac{\sum_{i=1}^{N} (y_i - g_M(\mathbf{x}_i))^2}{(1 - C(M)/N)^2} \tag{9}$$

where M is the number of linearly independent basis functions in Equation 3, being proportional to the number of basis function parameters; $C(M)$ is a complexity penalty function to avoid overfitting and indicates the effective number of parameters in the model. Usually, it is defined as $C(M) = d \cdot M$, where d is the penalizing parameter. According to the suggestion from (*18*), d is chosen as 4 in this paper. Finally, the basic choice of MARS model has $\hat{M} = \arg\min_M GCV(M)$ additive terms.

After creating a MARS model through the above two stages, we can track the GCV changes during the building of the model for each predictor. The importance of the variables can be estimated via accumulating the reduction in the statistic when each predictor's feature is added to the model. If a predictor (including spatial and temporal traffic volume) was rarely or never used in any MARS basis function, it has little or no influence on the target road segment. In our study, the GCV changes are normalized to 0 to 1, where 1 denotes that the predictor is the most important one among all of the predictors, while 0 denotes that the predictor is useless to the response. Finally, the GCV changes of all predictors constitutes the variable selection weight $\boldsymbol{\omega}$.

Support Vector Regression Model

In the second stage of our prediction framework, we employ the SVR model to learn the relationship between the selected predictors and the response in our work. Support vector machine (SVM) is a widely used classification technique developed by Vapnik and co-workers in 1995 (*20*). A version of SVM for regression named SVR has been proposed in 1997 by Vapnik, Steven Golowich, and Alex Smola (*21*) for solving nonlinear regression problems. In recent years, SVR has shown remarkable generalization abilities in the prediction of traffic flow (*13, 14*).

Given the training data set $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$, q variables are selected from $\mathbf{X}$ by:

$$\mathbf{x}'_i = \boldsymbol{\omega} \cdot \mathbf{x}_i \tag{10}$$

where $\boldsymbol{\omega}$ is the variable selection matrix generated using MARS and $\mathbf{x}'_i \in \Re^q$. Then a kernel function $\phi(\mathbf{x}'_i) = \{\phi_1(\mathbf{x}'_i), \ldots, \phi_D(\mathbf{x}'_i)\}^T$ is employed to transform the q-dimensional input into a higher D-dimension Hilbert space. Based on the nonlinear mapping, SVR model can be expressed as follows

$$f(\mathbf{x}'_i) = \mathbf{w}^T \phi(\mathbf{x}'_i) + b \tag{11}$$

where $\mathbf{w} = \{w_1, w_2, \ldots, w_D\}^T$ is a weight vector and b is bias.

In SVR, Vapnik (*20, 21*) introduced a ε-insensitivity loss function ignoring errors of size less than ε. The loss function is defined using the following equation:

$$L_\varepsilon(f(\mathbf{x}'_i) - y_i) = \max\{0, |f(\mathbf{x}'_i) - y_i| - \varepsilon\} \tag{12}$$

where ε is the region of ε-insensitivity; when the predicted value falls outside the band area, the loss is equal to the difference between the predicted value and the margin. Otherwise, The function gives zero loss. Such error function is therefore more tolerant to noise and is thus more robust. As in the hinge loss, there is a region of no error, which causes sparseness.

After introducing the loss function, the $\mathbf{w}$ and b are obtained by minimizing the following equation:

$$E(\mathbf{w}) = \frac{1}{2}\mathbf{w} \cdot \mathbf{w} + C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon(f(\mathbf{x}'_i) - y_i) \tag{13}$$

where C is a parameter controlling the tradeoff between the penalty and margin. The issue of SVR is to find the acceptable function $f(\cdot)$ corresponding to a minimum $E(\mathbf{w})$. Analogous to the soft margin hyperplane, Vapnik (*20*) introduced slack variables to account for deviations out of the ε-zone. Thus Equation (13) can be transformed to:

$$\begin{cases} \min & \frac{1}{2}(\mathbf{w} \cdot \mathbf{w}) + C\frac{1}{N}\sum_{i=1}^{N}(\xi_i + \xi'_i) \\ \text{s.t.} & y_i - [\mathbf{w} \cdot \phi \mathbf{x}'_i + b] \le \varepsilon + \xi'_i \\ & [\mathbf{w} \cdot \phi \mathbf{x}'_i + b] - y_i \le \varepsilon + \xi_i \\ & \xi_i, \xi'_i \ge 0, \quad i = 1, 2, \ldots, N \end{cases} \tag{14}$$

where we use two types of slack variables ξ_i and ξ_i' to keep them positive. Actually, we can see these as two hinges added back to back, one for positive and one for negative slacks. In practice, Equation (14) is often solved through its dual problem:

$$\begin{cases} \max & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\beta_i-\beta_i')(\beta_j-\beta_j')K(\mathbf{x}_i',\mathbf{x}_j') - \varepsilon\sum_{i=1}^{N}(\beta_i+\beta_i') - \sum_{i=1}^{N}(\beta_i-\beta_i') \\ \text{s.t.} & 0 \le \beta_i \le C/N,\ 0 \le \beta_i' \le C/N,\ \sum_{i=1}^{N}(\beta_i-\beta_i') = 0,\ \ i = 1,2,\ldots,N \end{cases} \tag{15}$$

where β_i and β_i' are the Lagrange multipliers of two constraints in Equation (14); $K(\mathbf{x}_i',\mathbf{x}_j') = \phi(\mathbf{x}_i')\cdot\phi(\mathbf{x}_j')$ is the kernel function satisfying Mercer condition. If $\overline{\beta}$ and $\overline{\beta'}$ are the solutions of Equation (15), then we have

$$b = \begin{cases} y_i - \varepsilon - \sum_j(\overline{\beta} - \overline{\beta}_i)K(\mathbf{x}_i',\mathbf{x}_j') & \beta_i \ge 0 \\ y_i + \varepsilon - \sum_j(\overline{\beta} - \overline{\beta}_i)K(\mathbf{x}_i',\mathbf{x}_j') & \beta_i' \ge 0 \end{cases} \tag{16}$$

After the above transformations, the solution to Equation (13) is transformed into the estimation of β_i, β_i', and b. Finally, in the testing set, we can predict the response of $\mathbf{x}'$ using:

$$f(\mathbf{x}') = \sum_{i=1}^{N}(\beta_i-\beta_i')K(\mathbf{x}',\mathbf{x}_i') + b \tag{17}$$

Moreover, before training the SVR model, the predictors $\mathbf{X}$ and response $\mathbf{y}$ are scaled internally to zero mean and unit variance. The center and scale values are returned for later predictions.

DATA SET DESCRIPTION

The work in this paper focuses on the short-term prediction of the traffic volume of the target road segment in Shanghai urban road network area. The raw data set is collected from the Sydney Co-ordinated Adaptive Traffic System (SCATS) of Shanghai road network provided by Shanghai Traffic Information Center along with the precise map of Shanghai road network. The selected field data are the total numbers of vehicles passing certain loop detectors during every interval of 10 minutes along the road links whose unit is vehicles per hour (*vehs/h*). The duration of the data is from February 1 to February 29, 2012, except February 24 as there was no data on this day. Since the urban traffic states on weekdays are prone to be congested and are more complicated, we only consider the traffic volume on weekdays in this paper.

The traffic volumes in a subarea of Shanghai road network are investigated in our study. The map and the location of the subarea is shown in Figure 2. The subarea consists of 12 bidirectional road segments labeled by circled capital letters. The average length of the segments is about 400m. Furthermore, the footnotes 1 and 2 are used to denote the direction of the vehicle stream. Footnote 1 denotes the traffic streams from north to south and west to east, and footnoot 2 denotes the reverses. For example, D_1 denotes the link between its upstream roads, A_1, C_1, F_2 and downstream roads, B_2, E_1, G_1 and flows from west to east; G_2 denotes the link between its upstream roads, I_1, L_2, J_2 and downstream roads, B_2, D_2, E_1 and flows from south to north.

As shown in the map, there are 24 road segments in the selected road network (we consider the bidirectional roads as two individual road segments). However, in practice, some loop detectors don't work or there are no detectors on some road segments. In our data set, traffic data from road B_1, B_2, D_1, F_1, G_1, K_1, and L_1 are unavailable. Therefore, we fed the traffic volume from the

FIGURE 2 A subarea road network of Shanghai and the location of the subarea.

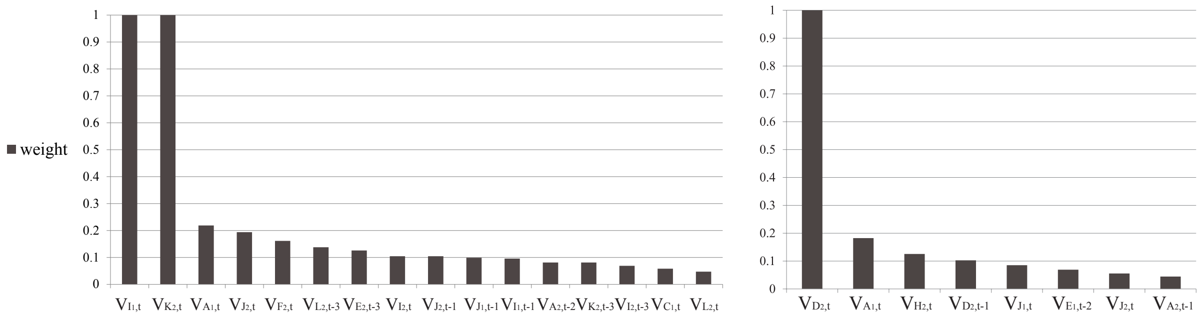
17 available road segments into the proposed prediction model. Additionally, two road segments with different directions, I_1 and D_2, are selected as the target roads to predict. I_1 is a 3 lanes road segment directing to downtown from suburb. D_2 is a 2 lanes road segment directing to suburb from downtown.

In addition, there are still missing data existing in the traffic data set of the remaining road segments. Therefore, we eliminate the missing data when selecting the significant predictors and training the prediction model. When we perform the prediction model on testing date set, the missing data is replaced using its prediction value.

EXPERIMENTS AND DISCUSSIONS

In order to build the proposed prediction model, the available traffic volume is divided into two subsets: training set and testing set. The training set contains the early 17 weekdays in our data set, from February 1 to February 23, 2012. The remaining 3 weekdays, February 27, 28, and 29, are used for model evaluation. The training set is used to obtain the contribution weight ω via the MARS model in the first step, and then the weighted spatio-temporal variables in the training set are reused to train the SVR model. After the SVR prediction model is built, the testing set is employed to evaluate the performance of our traffic prediction model.

The predictor from each road segment contains the current traffic volume V_t and the traffic volume in the earlier d time intervals, $\{V_{t-1}, \ldots, V_{t-d}\}$. To determine the number of earlier time intervals d, Xie *et. al.* employed an autocorrelation function (ACF) method (*16*). However, previous spatio-temporal models on urban traffic prediction always construct the independent variable $\mathbf{X}$ with smaller time intervals (*15*). Therefore in our experiments, we tested the effects of several time intervals, e.g. 2, 3, 4, 5. After these attempts, we found that the proposed model can achieve



(a) I_1 (b) D_2

FIGURE 3 The selected spatio-temporal variables for road I_1 and D_2

better prediction results with 3 time intervals in terms of MAPE. Consequently, four predictors, $\{V_t, V_{t-1}, V_{t-2}, V_{t-3}\}$, are drawn from each road segments and then compose the independent variable $\mathbf{X}$. Therefore, the independent variable $\mathbf{X}$ in the training and testing set contains a total of 68 predictors. The response is the observation average traffic volume at the object road in the later 10 minutes V_{t+1}.

Moreover, in the SVR model, the kernel function used for the traffic prediction is $K(x_i, x_j) = exp(-\sigma \left\|x_i - x_j\right\|^2)$ where σ is the kernel parameter. In the SVR training process, the parameters (C, σ, ε) are selected by a 5-fold cross-validation on the training set. The searching grid over the (C, σ, ε) is $[2^{-2}, 2^{-1}, \ldots, 2^6] \times [2^{-6}, 2^{-5}, \ldots, 2^2] \times [2^{-5}, 2^{-4}, \ldots, 2^2]$.

Results of Spatio-Temporal Variable Selection

As mentioned in Section 2.1, the descent of GCV importance in MARS model can be used to measure the contribution of the spatio-temporal variables to the object road segment. After the two-step optimization of the MARS model, the contribution weight ω can be drawn from the descent of GCV importance. Finally, the weights of the selected predictors for I_1 and D_2 are illustrated in Figure 3(a) and 3(b), respectively. The predictors which are not in the figure are regarded to have little impact on the traffic state of the object road segment and will not be used in the following SVR model building.

In Figure 3, we can find that the two predictors most related to the response $V_{I_1,t+1}$ are its current state $V_{I_1,t}$ and the state from one of its upstream road segment, $V_{K_2,t}$. For road D_2, the current traffic state $V_{D_2,t}$ is conceivably the most related predictor. Meanwhile the figures also indicate that some road segments would influence on the future traffic states of the target road although they don't belong to the adjacent upstream or downstream roads.

Prediction Results Analyses

Following the spatio-temporal variable selection, we train the SVR model using the selected predictors with different weights. To verify the proposed prediction model, the 3-order AR model is built to predict the testing set using the temporal traffic states. Moreover, the MARS model implemented in the first step is also carried out on the testing set. On the other hand, the SVR model is also directly trained and tested on all of the predictors without variable selection. So we can evaluate the performance of the spatio-temporal variable selection strategy proposed in this paper more intuitively.

As evaluating indicators, two measures for forecasting error analysis, root mean square error (RMSE) and mean absolute percentage error (MAPE), are employed to evaluate the performance of the proposed model. RMSE and MAPE are defined as follows:

$$RMSE = \sqrt{\left[\frac{1}{K}\sum_{k=1}^{K}\left(V_k - \hat{V}_k\right)^2\right]} \tag{18}$$

$$MAPE = \frac{1}{K}\sum_{k=1}^{K}\frac{\left|V_k - \hat{V}_k\right|}{V_k} \times 100\% \tag{19}$$

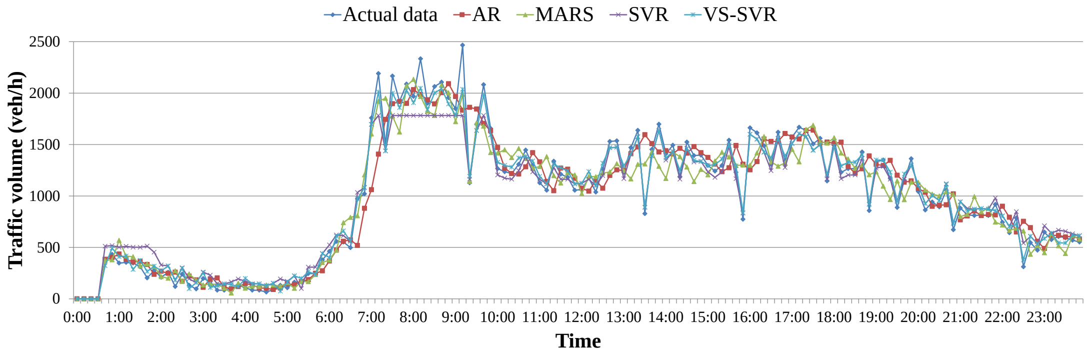
where K is the total number of intervals during the testing stage; V_k denotes the actual traffic volume; $\hat{V}_k$ is the prediction value produced by the proposed or comparison models.

In the model testing phase, the four models were carried out on the traffic volumes from road I_1 and D_2 during 3 weekdays, from February 27 to 29. Their prediction results on road I_1 on the three days are plotted with the actual data in Figures 4(a), 4(b), and 4(c), respectively. Figure 5 plots the prediction results of road D_2 on February 27. As can be observed in Figure 4 and 5, the daily traffic phenomenological trends on road I_1 and D_2 are different in the morning and evening peak. On road I_1, the morning peak is much stronger than the evening peak. The performance of road D_2 is just the opposite: the evening peak is stronger than the morning peak. Such differences are due to the origin-destination (OD) pattern existing between the suburb and downtown. Most people work in downtown, therefore, there is a heavy trend from west to east in the morning and from east to west in the evening in the subarea. However, the proposed spatio-temporal VS-SVR provides reliable prediction results, even though there are different OD trends between different roads.

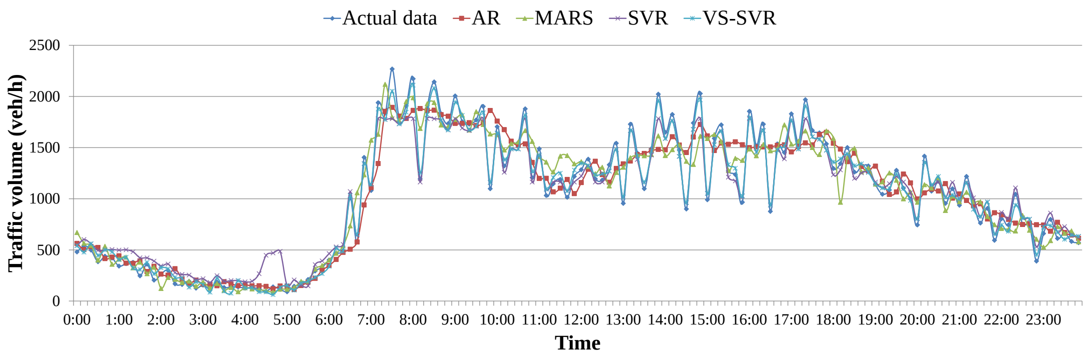
Moreover, from these figures, we can discover that the AR model generates an obvious delay during the prediction process, especially when the traffic states change suddenly, e.g. the climbing and declining phase during the morning and evening peak. MARS model also can not follow the frequent change of the actual volume timely. By contrast, the SVR and the proposed VS-SVR, as non-parametric methods, perform much better than the two parametric prediction methods.

For a further comparison of SVR and the proposed VS-SVR, we plot the prediction results on road I_1 by SVR and VS-SVR during the morning peak on February 29 in Figure 6. Obviously, the prediction results of VS-SVR is more close to the actual value, especially when the traffic volume is higher than 1500 veh/h. Therefore, we can conclude that the VS-SVR turned in a better performance than the SVR during the peak hour.

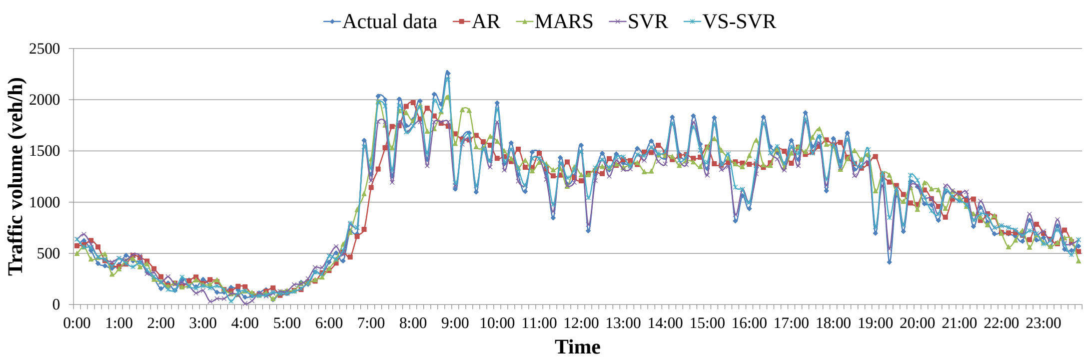
Furthermore, the RMSE and the MAPE of the proposed and comparison prediction models are summarized in Table 1. According to the FHWA quality standards, the maximum acceptable prediction error is 20%; 10% should be an ideal error. From this perspective, the AR model shows the worst performace on our real urban traffic volume and is slightly better than the maximum acceptable error. As a parametric nonlinear model, MARS performs better than AR, but its total MAPE is still up to 18.10% (for I_1). Even though the SVR model is a kernel learning method which always preforms excellently in most cases, there is still a way to the ideal standard. In comparison with SVR, the total MAPE of VS-SVR is much closer to the ideal error (10%) for road I_1 and D_2. Moreover, SVR is easily overfitting on high-dimensional predictors. As shown in Figure 4(b) and Table 1, SVR yields unexpected prediction errors during 4:00AM-5:00AM, which



(a) Feb. 27, Monday



(b) Feb. 28, Tuesday



(c) Feb. 29, Wednesday

FIGURE 4 Prediction results comparison of road I_1

leads to a high MAPE on Feb. 28. This is caused by the overfitting yield during the testing stage. In contrast, the proposed VS-SVR reduces the dimensionality of the predictors so as to reduce the risk of overfitting.

Besides the comparison of the accuracy of the prediction, we also discussed the computation demand of the prediction models. The average times of 10 executions for the model training and testing on road I_1 are listed in Table 2. The time is measured on a PC with 2.8 GHz Intel CPU, 4GB RAM and 64-bit operating system. As we can see from the table, VS-SVR consumed much more time than SVR in model training stage, but in practice, the model training can be completed

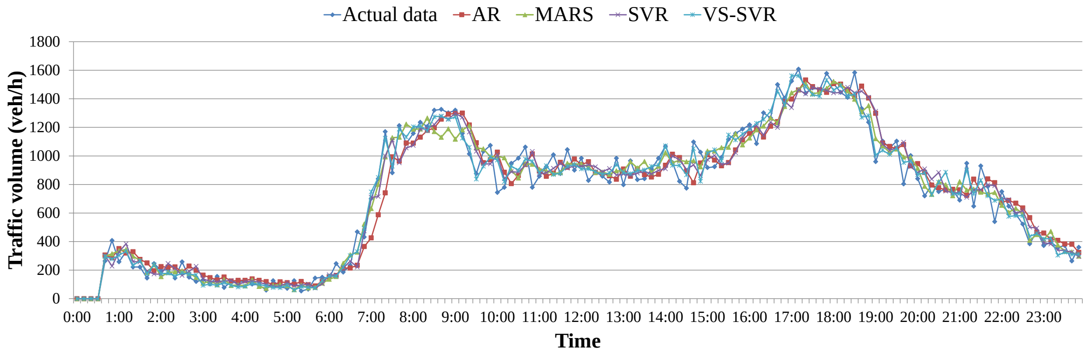


FIGURE 5 Prediction results comparison of road D_2 on Feb. 27, Monday

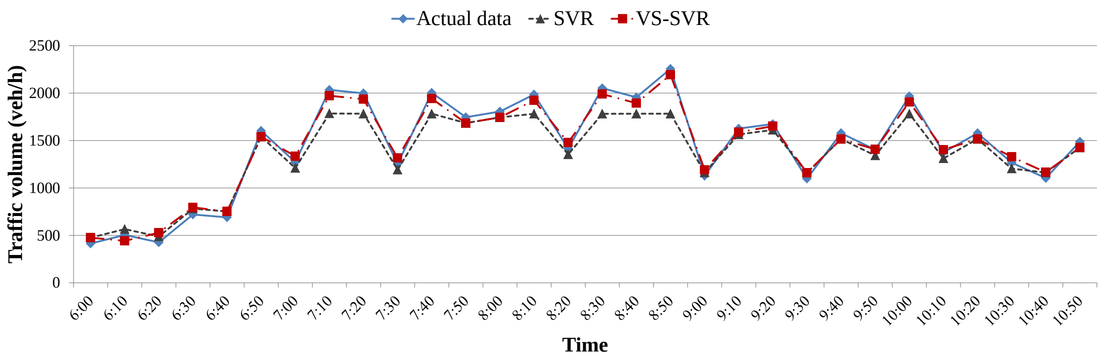


FIGURE 6 Prediction results comparison of the morning peak of road I_1 on Feb. 29

TABLE 1 Prediction error comparison of the four models

Road	Models	RMSE				MAPE(%)			
		Feb.27	Feb.28	Feb.29	Total	Feb.27	Feb.28	Feb.29	Total
I_1	AR	212.55	221.59	242.97	226.19	19.04	18.01	21.91	19.66
	MARS	185.03	206.24	205.19	199.18	19.63	16.13	18.60	18.10
	SVR	128.18	111.65	96.85	110.03	17.61	20.50	12.62	16.90
	VS-SVR	**75.74**	**59.85**	**77.91**	**71.58**	**12.98**	**9.39**	**10.80**	**11.04**
D_2	AR	121.50	130.04	126.48	126.10	18.48	18.18	20.21	18.96
	MARS	95.82	112.36	100.57	103.32	14.27	15.32	16.21	15.28
	SVR	108.68	120.55	107.72	112.50	15.75	15.21	16.64	15.86
	VS-SVR	**67.95**	**80.62**	**69.85**	**73.07**	**11.49**	**12.65**	**12.86**	**12.34**

off-line. The prediction stage of VS-SVR took much less time and could meet the real-time on-line application.

Based on above discussions, we can draw a conclusion that the proposed two-step VS-SVR model can exploit the spatio-temporal correlations reasonably, and achieves reliable prediction results in a short time. Meanwhile, the difference of OD trend has very little impact on the reliability

TABLE 2 Running time comparison of the four models

Models	No. of predictors	Training time(s)	Prediction time(s)
AR	4	0.13	0.006
MARS	68	522.47	0.03
SVR	68	3.67	0.19
VS-SVR	68	534.94	0.09

of the proposed model. Moreover, although we test the model on a small network, the model also can work on large networks. For extremely large road network (e.g. Shanghai road network with thousands roads), although the model could run directly, we suggest that the road network should better be divided into small subareas first (e.g. subareas with less than one hundred roads).

CONCLUSIONS

This paper proposed a novel urban traffic volume prediction methods based on the spatio-temporal variable selection strategy. The proposed prediction framework contains two critical modules, the variable selection module, and the prediction module. In the first stage, the MARS model is employed as a filter variable selection method. In the second stage, the SVR model is employed to train the relationship between the selected predictors and the response. On the other hand, different from other urban traffic prediction models, the traffic data from all of the road segments are fed into the proposed model other than the upstream roads of the target. Consequently, the proposed model can exploit the spatio-temporal correlations of the road segments using MARS model. In the experiment, in order to evaluate the performance of the proposed prediction model, the linear AR model, the MARS model, the SVR models are employed for comparison. The experiments are carried out on a subarea of the real road network in Shanghai. The experimental results indicate that the spatio-temporal variable selection based SVR model is an effective approach for short-term traffic volume prediction in complex urban road network.

For future work, the proposed method can be extended in several ways, e.g., considering the spatio-temporal correlations under different traffic states, such as morning peak, evening peak, and the stable traffic stage.

ACKNOWLEDGMENTS

This work was partly supported by China National 863 Key Program under Grant 2012AA112307, China NSFC Program under Grant 61104160 and 61375019.

REFERENCES

[1] Voort, M. V. D., M. Dougherty, and S. Watson, Combining kohonen maps with arima time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, Vol. 4, No. 5, 1996, pp. 307–318.

[2] Williams, B. M., P. K. Durvasula, and D. E. Brown, Urban freeway travel prediction: application of seasonal ARIMA and exponential smoothing models. *Transportation Research Record: Journal of the Transportation Research Board, No. 1644*, 1998, pp. 132–141.

[3] Williams, B. M., Multivariate vehicular traffic flow prediction: evaluation of ARIMAX mod-

eling. *Transportation Research Record: Journal of the Transportation Research Board, No. 1776*, 2001, pp. 194–200.

[4] Okutani, I. and Y. J. Stephanedes, Dynamic prediction of traffic volume through Kalman filtering theory. *Transportation Research Part B: Methodological*, Vol. 18, No. 1, 1984, pp. 1–11.

[5] Smith, B. L., B. M. Williams, and R. K. Oswald, Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 4, 2002, pp. 303–321.

[6] Dougherty, M. S. and M. R. Cobbett, Short-term inter-urban traffic forecasts using neural networks. *International Journal of Forecasting*, Vol. 13, No. 1, 1997, pp. 21–31.

[7] Xu, Y., Q.-J. Kong, and Y. Liu, Short-Term Traffic Volume Prediction Using Classification and Regression Trees. In *The 2013 IEEE Intelligent Vehicles Symposium*, Gold Coast, Australia, 2013, pp. 493–498.

[8] Tchrakian, T. T., B. Basu, and M. O'Mahony, Real-Time Traffic Flow Forecasting Using Spectral Analysis. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 2, 2012, pp. 519–526.

[9] Hobeika, A. and C. K. Kim, Traffic-flow-prediction systems based on upstream traffic. In *Proc. Vehicle Navigation and Information Systems Conference*, Yokohama, Japan, 1994, pp. 345–350.

[10] Stathopoulos, A. and G. M. Karlaftis, A multivariate state-space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 2, 2003, pp. 121–135.

[11] Djuric, N., V. Radosavljevic, V. Coric, and S. Vucetic, Travel Speed Forecasting by Means of Continuous Conditional Random Fields. *Transportation Research Record: Journal of the Transportation Research Board, No. 2263*, 2011, pp. 131–139.

[12] Min, W. and L. Wynter, Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 4, 2011, pp. 606–616.

[13] Zhang, Y. and Y. Xie, Forecasting of Short-Term Freeway Volume with v-Support Vector Machines. *Transportation Research Record: Journal of the Transportation Research Board, No. 2024*, 2007, pp. 92–99.

[14] Castro-Neto, M., Y.-S. Jeong, M.-K. Jeong, and L. D. Han, Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications*, Vol. 36, No. 3, Part 2, 2009, pp. 6164 – 6173.

[15] Sun, S., C. Zhang, and G. Yu, A Bayesian network approach to traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, No. 1, 2006, pp. 124–132.

[16] Xie, Y., K. Zhao, Y. Sun, and D. Chen, Gaussian Processes for Short-Term Traffic Volume Forecasting. *Transportation Research Record: Journal of the Transportation Research Board, No. 2165*, 2010, pp. 69–78.

[17] Guyon, I. and A. Elisseeff, An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, Vol. 3, 2003, pp. 1157–1182.

[18] Friedman, J. H., Multivariate adaptive regression splines. *The Annual of Statistics*, Vol. 19, No. 1, 1991, pp. 1–67.

[19] Clarke, B., E. Fokoué, and H. H. Zhang, *Principles and Theory for Data Mining and Machine Learning*. Springer Series in Statistics, Springer-Verlag, Berlin, Germany, 2009.

[20] Vapnik, V. N., *The Nature of Statistical Learning Theory*. Springer, New York, 1995.

[21] Vapnik, V., S. Golowich, and A. Smola, Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Neural Information Processing Systems*, Vol. 9, 1997.