CICTP 2017

TRANSPORTATION REFORM AND CHANGE—EQUITY, INCLUSIVENESS, SHARING, AND INNOVATION

PROCEEDINGS OF THE 17TH COTA INTERNATIONAL CONFERENCE OF TRANSPORTATION PROFESSIONALS

July 7–9, 2017
Shanghai, China

JOINTLY ORGANIZED BY
Chinese Overseas Transportation Association (COTA)

Tongji University

SPONSORED BY
Transportation Research Board

Institute of Transportation Engineers (ITE)

Transportation and Development Institute
of the American Society of Civil Engineers

EDITED BY
Haizhong Wang
Jian Sun
Jian Lu
Lei Zhang
Yu Zhang
ShouEn Fang

Published by the American Society of Civil Engineers

Random Forests for Freeway Short-Term Traffic Speed Prediction

Hongyu Zheng[1]; Xiaowei Chen[2]; and Xiqun (Michael) Chen[3]

[1]Institute of Transportation Engineering, College of Civil Engineering and Architecture, Zhejiang Univ., 866 Yuhangtang Rd., Hangzhou 310058, China
[2]Institute of Transportation Engineering, College of Civil Engineering and Architecture, Zhejiang Univ., 866 Yuhangtang Rd., Hangzhou 310058, China
[3]Institute of Transportation Engineering, College of Civil Engineering and Architecture, Zhejiang Univ., 866 Yuhangtang Rd., Hangzhou 310058, China (corresponding author). E-mail: chenxiqun@zju.edu.cn

ABSTRACT

Short-term traffic speed prediction plays a crucial role in intelligent transportation systems, e.g., vehicle navigation. Short-term traffic speed forecasting used to be analyzed only considering temporal factors. In this paper, we take both temporal and spatial factors into account. Features of the upstream and downstream locations are considered to establish the random forest (RF) and support vector regression (SVR) models. In order to improve the prediction accuracy, two brand new variables (i.e., absolute congestion and relative congestion) calculated from raw data are proposed. The RF is built to test the new variables' influence on the prediction accuracy. Traffic data extracted from the California Freeway Performance Measurement System (PeMS) is used to build and test the speed prediction models. It is verified that both upstream and downstream speed variables are able to improve model accuracy, and the two proposed new features are helpful to further improving the accuracy.

Keywords: Speed forecasting; temporal and spatial correlation; random forest; support vector regression

INTRODUCTION

Intelligent transportation systems (ITS) gradually play an increasingly significant role in people's daily life, e.g., traffic control systems for management authorities, and intelligent navigation systems for travelers. The ability to accurately predict traffic states (e.g., traffic speed, density, and travel time) is vital to the implementation of ITS. As one of the most important traffic state indicators, short-term speed prediction is critical for intelligent navigation systems to assist travelers on en-route decision making, e.g., route planning, advisory speed, and proactive diversion. Meanwhile, short-term forecasted speed is the most intuitive information shown on dynamic message signs or the navigation interface with travelers.

Short-term traffic speed prediction has attracted great attention for decades and all over the world. Numerous models and algorithms have been proposed to predict short-term traffic speed, e.g., time series model (McFadden et al., 2001), artificial neural networks (ANN) model (Huang and Ran, 2003), support vector regression (SVR) model (Vanajakshi and Rilett, 2004), long short-term memory neural network (LSTM NN) model (Ma et al., 2015). Most recently, the random forest (RF) and gradient boosting regression tree (GBRT) were also used in traffic flow and travel time prediction (Zhang and Haghani, 2015). From the perspective of methodology, traffic speed prediction is as mature as traffic flow and travel time prediction.

In addition, there are studies associated with traffic state prediction considering temporal and spatial correlations of neighboring traffic sensors. Kamarianakis and Prastacos (2003)

demonstrated the application of univariate and multivariate techniques in an urban network. In that research, since speeds of different locations were modeled in a way that permitted to take into account what was happening in the adjacent and other locations in the network, the model could estimate traffic speeds across the network when a traffic flow shock was introduced at some points of the network. Min and Wynter (2011) developed a highly accurate and scalable method for traffic prediction considering the spatial characteristics of a road network in a way that reflected not only the distance but also the average speed on the links. Yao et al. (2016) devised an accurate method for short-term traffic prediction using the SVR algorithm, which indicated spatial characteristics might influence the prediction accuracy, but temporal and spatial correlations were not considered in that research.

The contributions of this paper are twofold: (a) Two traffic speed prediction models based on the SVR and RF algorithms are separately established which take temporal and spatial correlation into account to improve the prediction accuracy. The influence of upstream and downstream sensors is analyzed. (b) Based on that, two brand new explanatory variables are proposed, i.e., the absolute congestion (Abs.con) that describes the congestion status of the target segment compared with its long-term average speed, e.g., for the whole previous year, and the relative congestion (Rel.con) that depicts the congestion status of predicted road segment compared with its neighbor segments, which fill the research gap that existing traffic speed prediction models seldom consider temporal and spatial effects. For demonstrative purposes, we test the influence of the two variables on the prediction accuracy using traffic data extracted from the Freeway Performance Measurement System (PeMS, 2016).

The rest of this paper is organized as follows. Section 2 presents the short-term speed forecasting models used in this paper. In Section 3, the influence of different predictors on the model accuracy is discussed, the RF and SVR models based on a one-year training set are established, validated and tested. Section 4 draws conclusions and outlooks the further research.

METHODOLOGY

In this paper, the algorithms we employed include SVR and RF, which are introduced as follows.

Support vector regression

SVR is a widely-used supervised learning method, and has been widely used in statistical classification and regression analyses (Cortes and Vapnik, 1995). SVR has sought to improve the structural risk minimization, which ensures the learning generalization ability, and realizes the minimization of the empirical risk and confidence interval.

Suppose the given training set is $\left\{\left(\mathbf{x}_1, y_1\right),\left(\mathbf{x}_2, y_2\right),\cdots,\left(\mathbf{x}_N, y_N\right)\right\}$, for which the i th D-dimensional input predictor is $\mathbf{x}_i \in \mathcal{R}^D$ and the corresponding output response is $y_i \in \mathcal{R}^1$, where $\mathcal{R}^1$ and $\mathcal{R}^D$ are the 1-dimensional and D-dimensional feasible regions, respectively. Our goal is to find a function $f(\mathbf{x})$, which will make the most ε deviation from the actually obtained targets y_i, $i=1,...,N$, in the training set, and meanwhile as flat as possible. The regression function is given by

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b, \quad \mathbf{x} \in \mathcal{R}^D, b \in \mathcal{R}^1 \tag{1}$$

where, $\mathbf{w}$ is the normal vector of the hyper plane formed by $\mathbf{w}^{\mathrm{T}}\mathbf{x} + b = 0$, the distance from a

point out of the plane to the plane is b.

One way to ensure this is to minimize the norm, equal to a convex optimization problem, where, such a function f may not exist that approximates all pairs $(\mathbf{x}_i, y_i)$ with the precision ε.

$$\begin{aligned} &\text{Min } \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ &\text{s.t. } \begin{cases} y_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_i - b \le \varepsilon \\ \mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b - y_i \le \varepsilon \end{cases} \end{aligned} \tag{2}$$

For the soft margin loss function, we can introduce the relaxation of nonnegative variables ξ_i and ξ_i^* to cope with infeasible constraints. The objective function of the SVR can be transferred into

$$\text{Min}\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\frac{1}{2}\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right) \tag{3}$$

where, *C* is the regularization parameter, which controls the degree of punishment beyond the error of the sample.

And the ε-insensitive loss function $|\xi|_\varepsilon$ is given by

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \tag{4}$$

After introducing the Lagrangian multipliers $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$, the optimization problem (3) is converted into a dual problem (5).

$$\begin{aligned} L \equiv \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\left(\zeta_i + \zeta_i^*\right) - \sum_{i=1}^{N}\left(\eta_i\zeta_i + \eta_i^*\zeta_i^*\right) \\ - \sum_{i=1}^{N}\alpha_{\mathrm{i}}\left(\xi + \zeta_i - y_i + \mathbf{w}^T\mathbf{x}_i + b\right) - \sum_{i=1}^{N}\alpha_{\mathrm{i}}^*\left(\xi + \zeta_i^* + y_i - \mathbf{w}^T\mathbf{x}_i - b\right) \end{aligned} \tag{5}$$

Let the partial derivatives of $L(\mathbf{w}, b, \eta_i, \eta_i^*)$ in terms of $\mathbf{w}, b, \eta_i, \eta_i^*$ be zero, the optimal $\mathbf{w}$, $f(\mathbf{x})$ and b can be obtained by exploiting KKT conditions. In this paper, the linear kernel function is used to realize the prediction.

Random forests

RF is an ensemble learning method for classification and regression (Ho, 1998). Based on the mean prediction of the individual trees for regression, RF works by constructing a multitude of decision trees and outputting the prediction.

The training dataset for each decision tree is drawn by sampling with replacement, which is the general technique of bootstrap aggregating or bagging. As shown in Figure 1, the dataset D is selected as the original one to generate bootstrap datasets for every individual tree model (e.g., D_1, D_2, and D_K). Then the RF algorithm selects a random subset of the features at each candidate split in the learning process, which will promise some strong variables get a great impact on the model. Finally, based on input variables, every single tree will provide a prediction. The output of RF is the mean prediction of the individual trees.

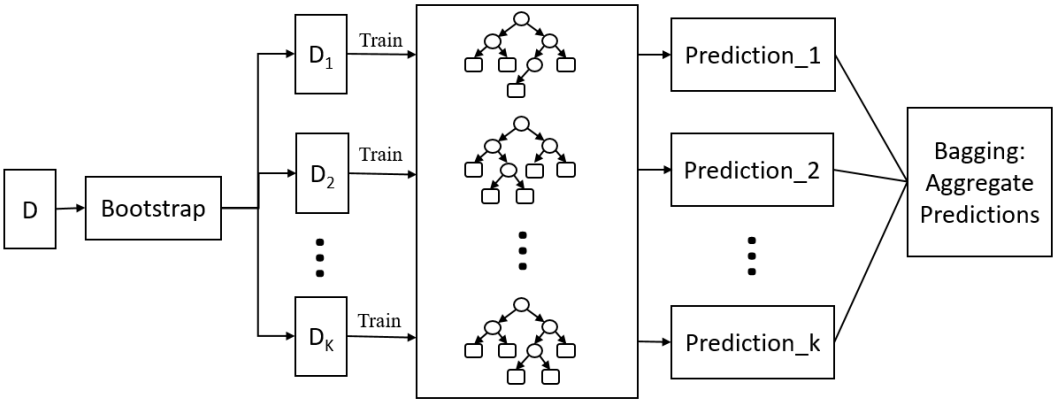

Figure 1. Flowchart of RF.

TRAFFIC SPEED PREDICTION

Data

In this study, the traffic data were extracted from the Freeway Performance Measurement System (PeMS, 2016), which provides real-time and historical data collected from over 42,000 detectors deployed on freeways throughout California. Measurements from January 1, 2015 to October 31, 2016 (data of the whole year of 2015 are used as the training set, while data in 2016 are used as the test set) are collected from five loop detectors located on US101-N from Blanken Ave. to Bayshore Blvd., San Francisco, CA. Figure 2 presents the layout of the five loop detectors (i.e., X1, X2, X3, X4 and X5) with latitude and longitude information, denoted by dots. The raw data contain the station ID, lane-by-lane vehicle counts, speed, and occupancy with the 5-min time interval for every single detector, while only speed data is selected for this paper. The lane-by-lane vehicle counts weighted average speed of each detection station is given by

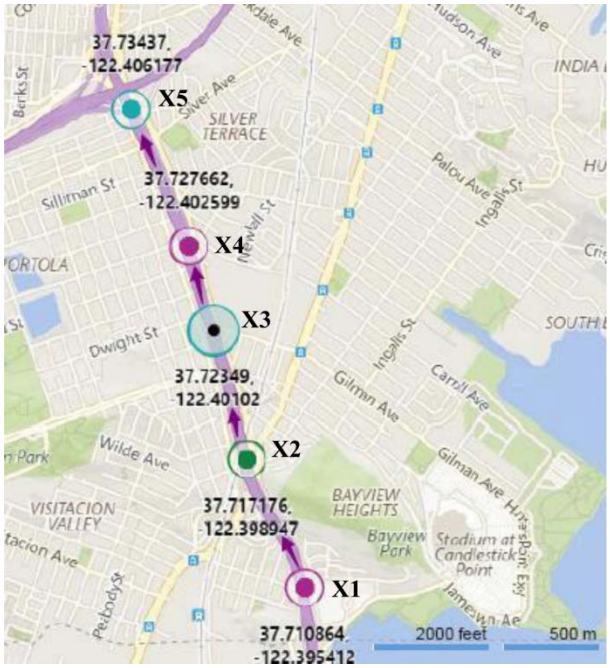

Figure 2. Layout of five loop detectors located on US101-N.

$$\bar{v}(t) = \frac{\sum_{l=1}^{L} \left(q_l(t) \cdot v_l(t) \right)}{\sum_{l=1}^{L} q_l(t)} \tag{6}$$

where $\bar{v}(t)$ is the average speed at time t, L is the number of lanes, $q_l(t)$ is the traffic count of lane l, and $v_l(t)$ is the average speed of lane l at time t.

This paper focuses on the short-term traffic speed prediction considering temporal and spatial correlations. In order to describe the congestion status of the predicted road segment, two new variables (i.e., Abs.con, Rel.con) related to the predicted detector are proposed. Abs.con describes the congestion status of the target segment compared with its long-term historical average speed, e.g., for the previous whole year. Rel.con that depicts the congestion status of predicted segment compared with its neighboring segments.

Totally, there are 27 variables considered in the prediction model, as shown in Table 1, including the two brand new variables calculated from the training set. The first column is the corresponding output of the SVR/RF models and the remaining 26 columns are the input variables. The output of the model is the 5-min speed of detector X3 at time T, denoted as V_{X3T}. Among the 26 input variables, the time of day is indexed from 0 to 287, each of which represents five minutes of one day. The day of week is indexed from 1 to 7 to represent Monday through Sunday. Month is the month information for the observation (from 1 to 12). V_{X1T1}, V_{X2T1}, V_{X3T1}, V_{X4T1} and V_{X5T1} represent the speeds of the five loop detectors (i.e., X1, X2, X3, X4 and X5) at time T1 standing for five minutes before T. Similarly, V_{X1T2}, V_{X2T2}, V_{X3T2}, V_{X4T2} and V_{X5T2} at time T2 are speeds ten minutes before T and V_{X1T3}, V_{X2T3}, V_{X3T3}, V_{X4T3} and V_{X5T3} at time T3 are speeds fifteen minutes before T. $\Delta V_{X3T12} = V_{X3T1} - V_{X3T2}$ and $\Delta V_{X3T23} = V_{X3T2} - V_{X3T3}$ are the speed changes of detector X3. Besides, Abs.con and Rel.con are defined by Eqs. (7-8) and four logical values for ΔV_{X3T12}, ΔV_{X3T23}, Abs.con and Rel.con are also listed in the training set (1 negative; 0 positive).

$$\text{Abs.con} = V_{X3T1} - \frac{1}{k}\sum_{i=1}^{k} V_{X3Ti} \quad (k = 288 \times 365) \tag{7}$$

$$\text{Rel.con} = V_{X3T1} - \frac{1}{5}\sum_{i=1}^{5} V_{XiT1} \tag{8}$$

Measures of effectiveness (MoE)

The MoE we use in this paper include the root mean square error (RMSE), mean absolute percentage error (MAPE), normalized root mean squared error (NRMSE), and symmetric mean absolute percentage error (SMAPE), defined by

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}[y_j - f(\mathbf{x}_j)]^2} \tag{9}$$

$$\text{MAPE} = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{y_j - f(\mathbf{x}_j)}{y_j}\right| \times 100\% \tag{10}$$

$$\text{NRMSE} = \sqrt{\frac{\sum_{j=1}^{M}[y_j - f(\mathbf{x}_j)]^2}{\sum_{j=1}^{M}[y_j]^2}} \times 100\% \quad (11)$$

$$\text{SMAPE1} = \frac{1}{M}\sum_{j=1}^{M}\frac{\left|y_j - f(\mathbf{x}_j)\right|}{y_j + f(\mathbf{x}_j)} \times 100\% \quad (12)$$

$$\text{SMAPE2} = \frac{\sum_{j=1}^{M}\left|y_j - f(\mathbf{x}_j)\right|}{\sum_{j=1}^{M}[y_j + f(\mathbf{x}_j)]} \times 100\% \quad (13)$$

where, M is the number of the test set.

Table 1. Sample of the Training Set

V_{X3T}	Time of day	Day of week	Month	V_{X1T1}	V_{X2T1}	V_{X3T1}	V_{X4T1}	V_{X5T1}
72.7	285	6	1	68.1	72.9	73.6	73.1	63.7
73.3	286	6	1	68.1	72.7	72.7	72.6	65.6
73	287	6	1	68.3	72.7	73.3	73	66.6
72.3	0	7	2	68.8	73.2	73	73.4	67.4
71.6	1	7	2	68.1	71.3	72.3	72.7	69.4
72.5	2	7	2	69.1	71.2	71.6	72.3	66.4

…	V_{X5T3}	ΔV_{X3T12}	0/1	ΔV_{X3T23}	0/1	Abs.con	0/1	Rel.con	0/1
…	63.2	0.2	0	0.5	0	9.31	0	3.32	0
…	63.1	-0.9	1	0.2	0	8.41	0	2.36	0
…	63.7	0.6	0	-0.9	1	9.01	0	2.52	0
…	65.6	-0.3	1	0.6	0	8.71	0	1.84	0
…	66.6	-0.7	1	-0.3	1	8.01	0	1.54	0
…	67.4	-0.7	1	-0.7	1	7.31	0	1.48	0

Results

We analyze the long-term and overall correlation of five time series that are measured by the five detection stations for the whole year of 2015. Table 2 shows that there exist significant correlations among different time series. It is interesting to find that no matter at upstream or downstream, the correlation gradually weakens with the distance between the locations. It is obvious that spatial correlations play an important role in the short-term traffic speed prediction.

First, we aim to explore the influence of different combinations of detection stations (e.g., target location, downstream and upstream locations) on the prediction accuracy. Using the 10-week traffic speed data from April 4, 2016 to Jun 12, 2016, we build four RF models of different feature combinations and then compare their prediction performance. Parameters tuning is carried out as follows: the seven-week data (between April 4, 2016 and May 22, 2016) are used as the training set for the parameter tuning purpose, the two-week data (between May 23, 2016 and June 5, 2016) are used as the validation set, and the last-week data are used as the test set (June 6-12, 2016). For a demonstrative purpose, the objective functions are set as the MAPE

error of the validation set for RF and the 5-fold cross validation MAPE error for SVR, respectively. The best parameters correspond to the minimum validation error for RF and the minimum cross validation error for SVR. As shown in Table 3, it is obvious that RF outperforms SVR in terms of those errors. After parameters tuning, the results of the four models that combine different detection locations are shown in Table 4. The RF model that considers features of the target, upstream and downstream locations outperforms other RF models.

Table 2. Correlation Matrix of Five Time Series

Detector	X1	X2	X3	X4	X5
X1	1	0.78	0.74	0.74	0.66
X2	0.78	1	0.80	0.78	0.65
X3	0.74	0.80	1	0.88	0.73
X4	0.74	0.78	0.88	1	0.85
X5	0.66	0.65	0.73	0.85	1

Table 3. Parameters Tuning with Features of the Target Location Only

MoE	RF			SVR		
	Training error	Validation error	Test error	Training error	Cross validation error	Test error
MAPE	0.84%	1.83%	1.74%	1.97%	1.99%	1.83%

Table 4. Test Error of RF for June 6-12, 2016

MoE	Features of target location only	Features of target and downstream locations	Features of target and upstream locations	Features of target, upstream and downstream locations
RMSE	1.66	1.61	1.59	1.58
NRMSE	2.50%	2.43%	2.39%	2.37%
MAPE	1.88%	1.78%	1.82%	1.76%
SMAPE1	0.93%	0.89%	0.90%	0.88%
SMAPE2	0.80%	0.77%	0.78%	0.77%

Among all of the input variables in Table 1, ΔV_{X3T12}, ΔV_{X3T23}, Abs.con and Rel.con are calculated to describe the congestion status of the target location. We also introduce four binary variables as potential features corresponding to the four continuous variables aforementioned. In this paper, we select some of those variables as input features for model training, rather than predicting the future status. Using the 10-week traffic speed data from April 4, 2016 to Jun 16, 2016, we build six RF and SVR models of different feature combinations and then compare their prediction performance. Parameters tuning is carried out as follows: the seven-week data (between April 4, 2016 and May 22, 2016) are used as the training set for the parameter tuning purpose, the two-week data (between May 23, 2016 and June 5, 2016) are used as the validation set, and the last-week data are used as the test set (June 1-5, 2016). Table 5 shows test errors of the six models. The basic features indicate the features from the time of day to V_{X5T3} as shown in

Table 1. The four continuous features represent ΔV_{X3T12}, ΔV_{X3T23}, Abs.con and Rel.con, the 0/1 values of which denote the four binary features.

As shown in Table 5, both RF and SVR models with the four continuous features perform better than the basic features only and the basic features plus four binary features. The models based on the basic features plus four binary features perform the worst in the test set. Although the improvement is not very significant, we believe that the change of prediction accuracy is due to the introduction of variables considering the data source is relatively good. Therefore, we conclude that incorporating ΔV_{X3T12}, ΔV_{X3T23}, Abs.con and Rel.con can improve the prediction accuracy. But the binary features show the congestion status of the target location as 0 or 1, which ignores the level of congestion. In the followings, we will focus on the impact of the four continuous features on the prediction accuracy.

Table 5. Feature Comparison for RF and SVR Models

MoE	RF			SVR		
	Basic features plus four continuous features	Basic features plus four binary features	Basic features only	Basic features plus four continuous features	Basic features plus four binary features	Basic features only
RMSE	1.49	1.57	1.53	1.55	1.60	1.60
NRMSE	2.25	2.36	2.31	2.34	2.41	2.41
MAPE	1.69	1.78	1.75	1.73	1.79	1.79
SMAPE1	0.84	0.88	0.87	0.86	0.89	0.89
SMAPE2	0.74	0.76	0.75	0.77	0.79	0.79

On the basis of the eighteen basic variables in Table 1, we continuously establish eight RF models by adding the eight combinations of the four continuous features to the basic features, see Table 6. Then the best model is selected that achieves the least test error, denoted by MAPE. Comparing the performance of the SVR and RF models, we find that SVR and RF models perform similarly among the eight combinations of features in Table 6. RF generates a bit smaller test error than SVR. Comparing model 1 with other models in Table 6, we find that those four features, i.e., ΔV_{X3T12}, ΔV_{X3T23}, Abs.con and Rel.con, can improve the prediction accuracy. In addition, it is worth mentioning that model 7 and model 8 have few difference in the prediction accuracy. As a result, model 8 is used in this paper.

Table 6 Eight Possible Combinations of the Four Continuous Features

Features	1	2	3	4	5	6	7	8
ΔV_{X3T12}				√		√	√	√
ΔV_{X3T23}				√		√	√	√
Abs.con			√		√		√	√
Rel.con		√			√	√		√

Furthermore, we apply RF and SVR to train the whole year data of 2015. Then we randomly select two weeks from the data of 2016 as the test sets, each of which includes traffic accidents, on February 19, 2016 and June 8, 2016, respectively. In this paper, we name February 15-21, 2016 as Week 1, June 6-12, 2016 as Week 2, February 19, 2016 as Day 1, and June 8, 2016 as

Day 2. The parameters tuning and test results based on the one-year data are shown in Table 7. For the RF model, the optimal number of trees to grow is 600, number of variables randomly sampled as candidates at each split is 7. For the SVR model, the tradeoff parameter is $C = 1$, and the linear kernel is employed for simplicity. Prediction results on the test sets of both models are presented in Figures 3-4.

Table 7. Parameters Tuning with Features of Multiple Locations

MoE	RF			SVR		
	Training Error	Validation Error	Test Error	Training Error	Cross Validation Error	Test Error
MAPE	0.58	0.57	1.57	1.02	1.14	2.91

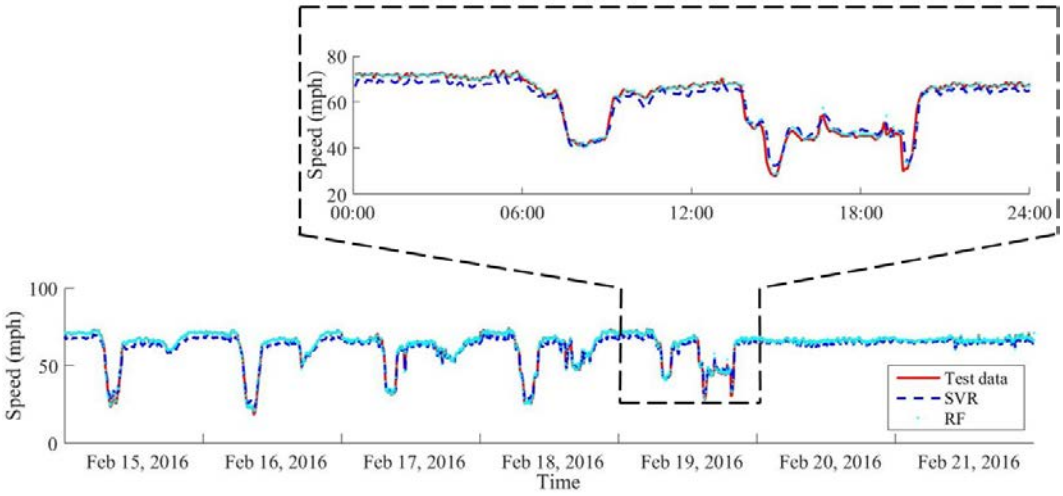

Figure 3. Prediction results for February 15-21, 2016.

Tables 8-9 show that both models fit the field measurements well, while the RF model outperforms the SVR model in terms of all of the five MoE for the two test sets and two days with traffic accidents, where the smaller errors are highlighted in bold. For the two weeks, both models can fit the trends well, whenever weekdays or weekends. The weekdays show obvious AM and PM peaks at 7:00-9:00 and 4:30-6:30, respectively, while the weekends experience no peaks. When accidents occurred, as highlighted at 15:00-19:00 in Figures 3-4, traffic breakdowns can be seen and the congestion lasted longer than the usual PM peaks. In short, the traffic speed prediction model considering features of the target, upstream and downstream locations results in more favorable predication performance, and the numerical experiments show that the RF model is able to predict short-term speed more accurately.

CONCLUSIONS AND FUTURE WORK

The goal of this paper is to devise an advanced and accurate method for short-term traffic prediction and thereby support the real-time traffic guidance and traffic control. Both the RF and SVM models that consider spatial and temporal parameters are developed. The models this paper employs are calibrated by using traffic data extracted from PeMS. More than one-year data at the 5-min time interval are used to train (10-month data), validate (two-month data), and test (two-week data) the models. The results show that the models considering features of both the target

location and its neighboring locations perform better than the model without considering the spatial information. Meanwhile, the RF model considering both the upstream and downstream information results in the highest prediction accuracy. Based on that, four continuous features (i.e., ΔV_{X3T12}, ΔV_{X3T23}, Abs.con, Rel.con) are proposed to describe the congestion status of the target location and are used as the significant prediction features. The results also show that the introduction of the four continuous features can improve the prediction accuracy.

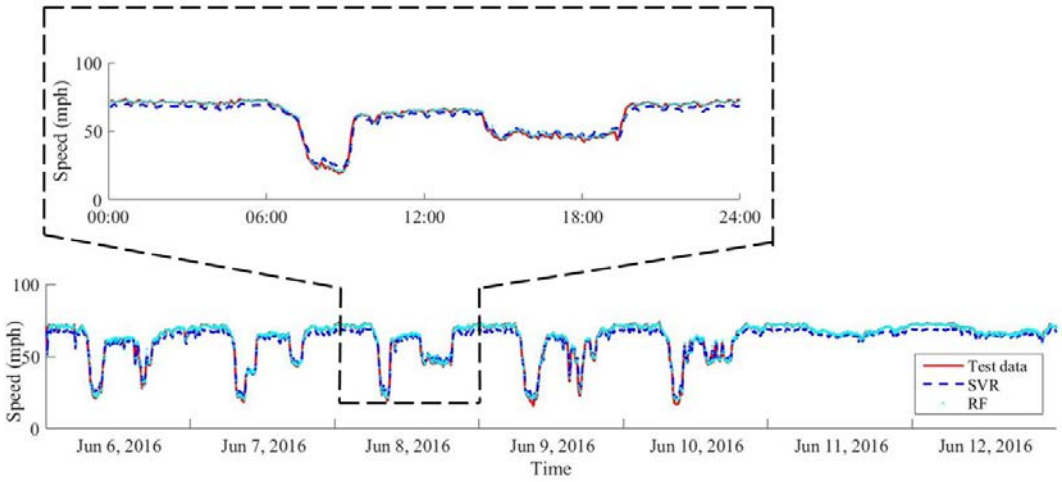

Figure 4. Prediction results for June 6-12, 2016.

Table 8. Prediction Errors of RF and SVR for the Two Test Sets

MoE	RF		SVR	
	Week 1	Week 2	Week 1	Week 2
RMSE	**1.38**	**1.73**	2.49	2.70
NRMSE	**2.16%**	**2.72%**	3.89%	4.28%
MAPE	**1.59%**	**2.25%**	3.59%	4.29%
SMAPE1	**0.79%**	**1.10%**	1.80%	2.11%
SMAPE2	**0.7%**	**0.88%**	1.70%	1.88%

Table 9. Prediction Errors of RF and SVR for Two Days of Traffic Accidents

MoE	RF		SVR	
	Day 1	Day 2	Day 1	Day 2
RMSE	**1.98**	**1.76**	2.89	2.67
NRMSE	**3.22%**	**2.86%**	4.69%	4.32%
MAPE	**2.27%**	**2.34%**	4.19%	4.23%
SMAPE1	**1.10%**	**1.17%**	2.06%	2.10%
SMAPE2	**0.96%**	**0.95%**	1.94%	1.91%

The current work only considers the influence of different combinations of the temporal and spatial influence factors. Future work will concentrate on the ranking of features mentioned in this literature, and explore the single-step and multi-step predictions considering the temporal and spatial correlations.

ACKNOWLEDGEMENTS

This research is financially supported by Zhejiang Provincial Natural Science Foundation of China (LR17E080002), the Enjoyor Corporation (2016ERCITZJ-KF03), National Natural Science Foundation of China (51508505, 51338008), the Key Science and Technology Innovation Team of Zhejiang Province (2013TD09), and the Fundamental Research Funds for the Central Universities.

REFERENCES

Cortes, C. and Vapnik, V. (1995). "Support-vector networks." *Machine Learning*, 20, 273–297.

Performance Measurement System (PeMS). (2016). [Online Access on November 1, 2016]. Accessible at http://pems.dot.ca.gov/

Ho, T. K. (1998). "The random subspace method for constructing decision forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.

Huang, S. H. and Ran, B. (2003). "An application of neural network on traffic speed prediction under adverse weather condition." *Proceedings of 82nd Annual Meeting of the Transportation Research Board* (Compendium of Papers, CD-ROM).

Jiang, X., Zhang, L., and Chen, X. (2014). "Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China." *Transportation Research Part C*, 44, 110–127.

Kamarianakis, Y. and Prastacos, P. (2003). "Forecasting traffic flow conditions in an urban network: comparison of multivariate and univariate approaches." *Transportation Research Record, 1857*, Transportation Research Board, Washington, D.C., 74–84.

Ma, X., Tao, Z., Wang, Y., Yu, H., and Wang, Y. (2015). "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data." *Transportation Research Part C*, 54, 187–97.

McFadden, J, Yang, W. T., and Durrans, S. R. (2001). "Application of artificial neural networks to predict speeds on two-lane rural highways." *Transportation Research Record, 1751*, Transportation Research Board, Washington, D.C., 9–17.

Min, W. and Wynter, L. (2011). "Real-time road traffic prediction with spatio-temporal correlations." *Transportation Research Part C*, 19(4), 606–616.

Vanajakshi, L. and Rilett, L. R. (2004). "A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed." *Proceedings of 2004 IEEE Intelligent Vehicles Symposium*, 14–17.

Yao, B., Chen, C., Cao, Q., Jin, L., Zhang, M., Zhu, H., and Yu, B. (2016). "Short-term traffic speed prediction for an urban corridor." *Computer-Aided Civil and Infrastructure Engineering*, in press.

Zhang, Y. and Haghani, A. (2011). "A gradient boosting method to improve travel time prediction." *Transportation Research Part C*, 58, 308–24.