International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:3, No:3, 2009

Comparison of Parametric and Nonparametric Techniques for Non-peak Traffic Forecasting

Yang Zhang, Yuncai Liu

***Abstract*—**Accurately predicting non-peak traffic is crucial to daily traffic for all forecasting models. In the paper, least squares support vector machines (LS-SVMs) are investigated to solve such a practical problem. It is the first time to apply the approach and analyze the forecast performance in the domain. For comparison purpose, two parametric and two non-parametric techniques are selected because of their effectiveness proved in past research. Having good generalization ability and guaranteeing global minima, LS-SVMs perform better than the others. Providing sufficient improvement in stability and robustness reveals that the approach is practically promising.

***Keywords*—**Parametric and Nonparametric Techniques, Non-peak Traffic Forecasting

I. Introduction

THE success of intelligent transportation system (ITS) much depends on the provision of accurate real-time information and predictions of traffic status. Due to the importance of traffic forecasts, more research attention has been focused on this subject in recent decades. Some papers such as extensive review appear [1]-[3], which attract significant scientific interest in more flexible approaches. After summarizing former works, traffic forecasting is classified quite differently on the basis of diverse classification standards. The type of forecasting depends on the following groups of factors: single road link or transportation network, freeways or urban streets, physical models or mathematical methodologies, univariate or multivariate method, etc. As illustrated in the statement, traffic prediction "can be separated into two paradigms: the empirical based, incorporating fairly standard statistical methodology on one hand, and that based on traffic process theory, either of demand or of supply, on the other" [1]. By applying statistical methodology or heuristic methods in traffic forecasting, the empirical approaches can be classified into two categories: parametric and nonparametric techniques.

To obtain accurate forecasts, since the early 1980s, extensive variety of parametric approaches has been employed ranging from historical average algorithms, smoothing techniques, linear and nonlinear regression, filtering techniques, to autoregressive linear processes including autoregressive moving average (ARMA) family that is regarded as a milestone in forecasting field. Lately extraordinary development of distinct nonparametric techniques, including nonparametric regression, neural networks, etc., has shown their great potential alternative to their parametric counterparts. In essence, nonparametric statistical regression can be regarded as a dynamic clustering model that relies on the relationship between dependent and independent traffic variables. It attempts to identify past information that are similar to the state at prediction time, which leads to easily implemented nature. Some researchers demonstrated that nonparametric techniques generally perform well due to their strong ability to capture the nondeterministic and complex nonlinearity of traffic time series [4]-[11]. Many computational intelligence (CI) techniques including fuzzy systems, machine learning, and evolutionary computation have been successfully adopted in the field. Particularly, artificial neural networks (ANNs) such as the radial basis function neural network (RBF-NN) [12], [13], have been successfully applied. The theory of support vector regression (SVR) has also been introduced by several researchers to model traffic characteristics and predict traffic states [7], [14], [15]. The recent applications of various CI techniques and hybrid intelligent systems have shown their good potential on traffic forecasting.

Least squares support vector machines (LS-SVMs) proposed by J. A. K Suykens [16], [17] are closely related to regularization networks and Gaussian processes, but emphasize and exploit primal-dual interpretations additionally. The early application of the method to financial time series forecasting has obtained breakthroughs and plausible performance [18]. In studies of *k*-nearest neighbor (*k*-NN) nonparametric approaches to traffic forecasting, state vectors are found to be essential to ensure more accurate prediction [19]. The traffic in non-peak period is the essential part of daily traffic. Unpredictable congestion occurred in the relatively stable traffic inevitably brings some difficulty for all predictors. Thus, the quality of forecasts in non-peak hours is fundamental to traffic prediction. And the LS-SVMs method is proposed to predict and analyze traffic status in non-peak period. Meanwhile, a hybrid state space method is applied to determine the appropriate input and output dimensions.

Due to the simplicity of the parametric techniques and the effectiveness of the nonparametric ones, both types of techniques are chosen for comparison: historical-mean (HM), autoregressive moving average (ARMA), radial basis functions (RBF) networks, and support vector regression (SVR) models.

This work was supported by the National Science & Technology Supporting Program during the 11th Five-year Plan Period (No. 2006BAJ18B02), China.

Y. Zhang and Y. Liu are with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, 200240, China. (phone: +86-21-3420-4028; fax: +86-21-3420-4340; e-mail: zhang-yang@sjtu.edu.cn, whomliu@sjtu.edu.cn).

To validate our method, the forecast performance is measured by different indices of forecast accuracy. The results show that the proposed approach is generally better than the other models in both effectiveness and robustness through the comparative case analysis.

II. Proposed Scheme

A. Least Squares Support Vector Machines (LS-SVMs)

Compared with standard support vector machines (SVMs) [20], LS-SVMs apply linear least squares criteria to the loss function instead of traditional Quadratic Programming (QP) method, which leads to the advantages of fast convergence, high accuracy and low computational efforts [16], [17]. The algorithm is introduced simply and mathematically. Suppose a training data set is

$$D = \{(\boldsymbol{x}_i, y_i) : i = 1, 2, \cdots, l\} \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^n$ is the input variable set; $y_i \in \mathbb{R}$ is the output variable set; l corresponds to the size of the training data. In the weight space (primal space), the LS-SVMs formulation can be described as:

$$\begin{cases} \min\limits_{\boldsymbol{w},b,\boldsymbol{e}} J(\boldsymbol{w},\boldsymbol{e}) = \dfrac{1}{2}\|\boldsymbol{w}\|^2 + \dfrac{1}{2}\zeta\sum\limits_{i=1}^{l} e_i^2, \\ s.t.\ \ y_i = \boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{x}_i) + b + e_i, \quad i = 1, 2, \cdots, l \end{cases} \tag{2}$$

where $\boldsymbol{\Phi}$: $\mathcal{L} = \mathbb{R}^n \rightarrow \mathcal{H}$ is a nonlinear mapping function that maps the input vector $\boldsymbol{x}$ into a higher (possibly infinite) dimensional feature space $\mathcal{H}$; $\boldsymbol{w} \in \mathcal{H}$ is the weight vector; $e_i \in \mathbb{R}$ is the error variable; $\boldsymbol{e} = [e_1, \ldots, e_l]^{\mathrm{T}} \in \mathbb{R}^l$ is the error vector and b a bias term. In addition, J is the loss function, and ζ is the adjustable error term corresponding to a weighted least squares cost function. To solve the above minimization problem, the Lagrangian function is defined by:

$$L(\boldsymbol{w}, b, \boldsymbol{e}, \boldsymbol{a}) = J(\boldsymbol{w}, \boldsymbol{e}) - \sum_{i=1}^{l} \alpha_i [\boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{x}_i) + b + e_i - y_i] \tag{3}$$

where $\alpha_i \in \mathbb{R}$ are named Lagrange multipliers or support values that constitute the support vector $\boldsymbol{\alpha}$ ($\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_l]^{\mathrm{T}} \in \mathbb{R}^l$). The conditions for optimality are given by:

$$\begin{cases} \dfrac{\partial L}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum\limits_{i=1}^{l} \alpha_i \boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{x}_i), \\ \dfrac{\partial L}{\partial b} = 0 \Rightarrow \sum\limits_{i=1}^{l} \alpha_i = 0, \\ \dfrac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \zeta e_i, \quad i = 1, 2, \cdots, l, \\ \dfrac{\partial L}{\partial \alpha_i} = 0 \Rightarrow \boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{x}_i) + b + e_i - y_i = 0, \quad i = 1, 2, \cdots, l. \end{cases} \tag{4}$$

In standard SVMs, $\boldsymbol{w}$ and $\boldsymbol{\Phi}(\boldsymbol{x}_i)$ are never calculated. After variables $\boldsymbol{w}$ and $\boldsymbol{e}$ are eliminated, the following linear system can be easily obtained:

$$\begin{bmatrix} 0 & \mathbf{1}^{\mathrm{T}} \\ \mathbf{1} & \boldsymbol{\Omega} + \dfrac{1}{\zeta}\mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \boldsymbol{y} \end{bmatrix} \tag{5}$$

where $\boldsymbol{y} = [y_1, \ldots, y_l]^{\mathrm{T}}$, $\mathbf{1} = [1, \ldots, 1] \in \mathbb{R}^l$, and $\boldsymbol{\Omega} = \{\boldsymbol{\Omega}_{ij}\}_{l \times l}$. Here Mercer's condition is applied within the matrix $\boldsymbol{\Omega}$

$$\boldsymbol{\Omega}_{ij} = \boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{x}_i)\boldsymbol{\Phi}(\boldsymbol{x}_j) = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{6}$$

In the optimum the weight vector can be denoted by $\boldsymbol{w} = \sum_{i=1}^{l} \alpha_i \boldsymbol{\Phi}(\boldsymbol{x}_i)$, and the LS-SVMs regressor is obtained by applying the Mercer's condition:

$$f(\boldsymbol{x}) = \boldsymbol{w}^T\boldsymbol{\Phi}(\boldsymbol{x}) + b = \sum_{i=1}^{l} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{7}$$

where $\boldsymbol{\alpha}$ and b can be obtained by solving the above matrix equation. For positive definite kernel function $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$, three typical kernels are commonly used. These are linear kernel with formula $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$, polynomial kernel of degree d with formula $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j + 1)^{\mathrm{d}}$ and radial basis function (RBF) kernel with formula

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\|\boldsymbol{x}_j - \boldsymbol{x}_i\|^2 / 2\sigma^2\right) \tag{8}$$

where σ is a tuning parameter. The paper focuses on the use of the RBF kernel for its good performance and advantages in time series forecasting problem proved in past research. The precision and convergence of LS-SVMs are both affected by ζ and σ.

B. State Space Method

The state space methodology has a long historical background [19]. There the state contains vectors that include all the information about a certain system that carries over into the future. Specifically, the measurements during each time interval t, t-1,…, t-d compose a state vector where d is an appropriate number of lags ($d \in \mathbb{N}$). Suppose there are N weeks chosen for training in the traffic data obtained. For a particular day in a week (Monday, for example), the state vector $\boldsymbol{X}_k(t, d)$ of the traffic parameters measured every 1 hour can be described by:

$$\boldsymbol{X}_k(t, d) = \left[V_k(t), \boldsymbol{x}_k(t, d)\right] \tag{9}$$

where

$$\boldsymbol{x}_k(t, d) = \left[V_k(t-1), \cdots, V_k(t-d)\right], \quad k \in [2, \mathrm{N}], \ t \in [1, 24] \tag{10}$$

can be chosen as the input variables in training the RBF-NN, SVR and LS-SVMs models; $V_k(t)$ denotes the traffic parameter

during the current time interval t in the week k; $V_k(t\text{-}1)$ represents the one during the previous 1-hour interval, etc. All indices in $\boldsymbol{x}_k(t, d)$ are closely related to $V_k(t)$. Particularly, when $t \le d$, $\boldsymbol{x}_k(t, d)$ contains the last $d\text{-}t\text{+}1$ parameters measured in the day before the chosen particular day. After d is appropriately determined, the vectors $\{\boldsymbol{x}_k(t, d), V_k(t)\}$, $k \in [2, N]$ can be used as input-output pairs in the training process for each t, $t \in [1, 24]$. And the number of each group of training samples is N-1 (24 groups). Then for each time interval t in the week N+1, the vector $\boldsymbol{x}_{N+1}(t, d)$ is used to obtain the final forecasting result $\bar{V}_{N+1}(t)$. It is usually compared with the original measurement $V_{N+1}(t)$ that means the parameter measured during the corresponding time interval t in the week N+1. In practice, $\boldsymbol{x}_k(t, d)$ may not only contain the d lagged values but also be supplemented with the historical information $V_{hist,k\text{-}1}(t)$ that represents the historical index at the time of day and day of the week associated with the time interval t in the week k-1 along the cyclical curve. The combination forms a hybrid state space. The comparison of different nonparametric forecasting results is investigated after specifying the variable d.

C. Measures of Forecast Accuracy

Two statistics are used to assess the quality of forecasting. The mean absolute percentage error (MAPE) and the variance of absolute percentage error (VAPE) are irrelevant to the unit of the measures and insensitive to the changes in the magnitude of forecasts:

$$MAPE = \left(\frac{1}{T} \sum_{t=1}^{T} \left| V_O(t) - \bar{V}_P(t) \right| / V_O(t) \right) \times 100\% \tag{11}$$

$$VAPE = \sqrt{\frac{T \sum_{t=1}^{T} \left(\frac{\left| V_O(t) - \bar{V}_P(t) \right|}{V_O(t)} \right)^2 - \left(\sum_{t=1}^{T} \frac{\left| V_O(t) - \bar{V}_P(t) \right|}{V_O(t)} \right)^2}{T(T-1)}} \times 100\% \tag{12}$$

where $V_O(t)$ is the observed value of the measure during the time interval t ($V_O(t) \neq 0$); $\bar{V}_P(t)$ is the predicted value of the measure; and T is the number of forecasting periods (T=102 in the experiment). Specifically, the MAPE calculates the average relative error between the forecast output and actual observed data, which reflects the accuracy of the forecasting. The VAPE calculates the sum of the deviations from the average performance during the forecasting in all periods, which represents the stability of a forecasting model. Meanwhile, the percentage error (PE) between our prediction and the original data is also applied:

$$PE(t) = \left[\left(V_O(t) - \bar{V}_P(t) \right) / V_O(t) \right] \times 100\% \tag{13}$$

III. Other Forecasting Methods

Two parametric and two nonparametric methods are used as representative techniques to supply overall comparisons.

A. Historical-mean (HM) Model

HM model, a simple conventional parametric technique, is described as below:

$$\bar{V}_{Aver,K}(t) = \frac{1}{K} \sum_{i=1}^{K} V_{hist,w+1-i}(t) \tag{14}$$

where $\bar{V}_{Aver,K}(t)$ is the average which represents the forecasting result $\bar{V}_{w+1}(t)$ and K is the number of the historical weeks before the week w+1. Namely, the forecasting result is obtained from the average of the historical traffic data at the same time of day and day of the week. This method is used as a representative of parametric techniques to supply a comparison.

B. Autoregressive Moving Average (ARMA) Models

Autoregressive moving average (ARMA) models, including purely autoregressive (AR) and purely moving-average (MA) models as special cases, are one of the most popular classes of linear time series models [21]. The models are frequently applied to model linear dynamic structures, to depict linear relationships among lagged variables, and to provide effective linear forecasting. The AR and MA classes can be further extended to modeling more complicated dynamics of time series. Combining AR and MA forms together yields the ARMA model defined as

$$\begin{aligned} V_{k+1}(t) = {} & b_1 V_{k+1}(t-1) + \cdots + b_p V_{k+1}(t-p) \\ & + \varepsilon_t + a_1 \varepsilon_{t\text{-}1} + \cdots + a_q \varepsilon_{t\text{-}q} \end{aligned} \tag{15}$$

where $\{\varepsilon_t\}$ is a simple type of stochastic process, denoted as $\{\varepsilon_t\} \sqcup \mathrm{WN}(0, \sigma^2)$; $p, q \geq 0$ are integers, and (p, q) represents the order of the model; $V_{k+1}(t)$ denotes the traffic parameter during the current time interval t in the week k+1; $V_{k+1}(t\text{-}1)$ represents the one during the previous 1-hour interval, etc. The white noise $\{\varepsilon_t\}$ serves as a building block in defining more complex linear time series processes and reflects information that is not directly observable. ARMA model is one of the most frequently used families of parametric models in the time series analysis. This is due to their flexibility in approximating many stationary processes.

C. Radial Basis Function Neural Network (RBF-NN)

As a competitive nonparametric regression approach, the RBF-NN model for predicting traffic data series is applied in this paper. The method is different from the conventional multilayer perceptrons (MLPs) approach in which the nonlinearity of the model is only embedded in the hidden layer of the network [22]. There are many additional advantages in RBF-NN model, which have been proved recently. Compared with MLPs, the network can be well developed to become more adaptive to universal approximations with more accuracy and less time. Moreover, the RBF network outputs become linear

functions of the output layer weights when the basis functions are appropriately fixed. The above two advantages make us select the highly developed tool for comparison.

D. Support Vector Regression (SVR)

SVR has been introduced within the context of statistical learning theory and structural risk minimization (SRM) principle. Researchers regard it as a powerful methodology for linear and nonlinear regression. Benefiting from the SRM principle, the SVR can gain a much better ability on generalization which is especially important for machining learning algorithms. In brief, the SVR maps the inputs into a higher dimensional feature space with an appropriate kernel inner product, then in the mapped space minimizing the loss value with Quadratic Programming (QP) techniques can determine some parameters which exclusively denotes a regression function [20], [23]. Together with the determined parameters and the fixed regression formulation, the regression function can be ascertained. N.B. before the SVR processing, an appropriate kernel function and loss function must be chosen to get a better solution for our problem. For the proved efficiency of SVR in the early research, and both LS-SVMs and SVR predictors belonging to the same SVMs family, SVR predictor is also compared with ours.

IV. Experimental Results

Data for this study come from the Performance Measurement System (PeMS), which can be accessed through the Internet [24]. The travel time index (TTI) is commonly used in the analysis of traffic status. It expresses the average amount of extra time it takes to travel in the peak relative to free-flow travel. And it can present congestion levels in a format that is easy to understand and communicate to the general public. Considering it from a practical perspective, the PeMS supplies the TTI directly on its website to the public for reference and evaluates the traffic situation in the whole freeway network. The traffic data of 24 weeks from May 1 to Oct. 15, 2006 are used in the paper. The 1-hour lane-aggregated average TTI data are downloaded because of our access to limited traffic data. The data for a particular day start every 1 hour between 00:00 am and 23:00 pm. Our attention is focused on predicting the TTI of the last week based on the former.

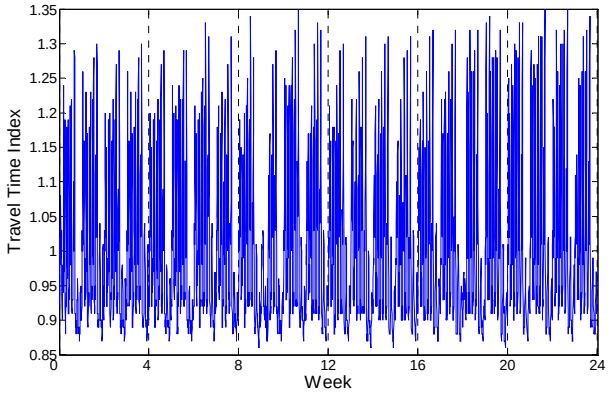


Fig. 1 Plot of all traffic data obtained from the transportation network (24 weeks)

Fig. 1 shows the total 24 weeks traffic data hourly. In this section, the out-of-sample forecasting ability of the models is evaluated. Fig. 2 illustrates the above hourly average TTI data of 1,680 time points. The data are continuously recorded over a period of the selected 70 days (the first 10 weeks). With the plan axes of 24 hours and 70 days, the three-dimensional graph shows the periodical pattern of 24 hours with two peak periods.

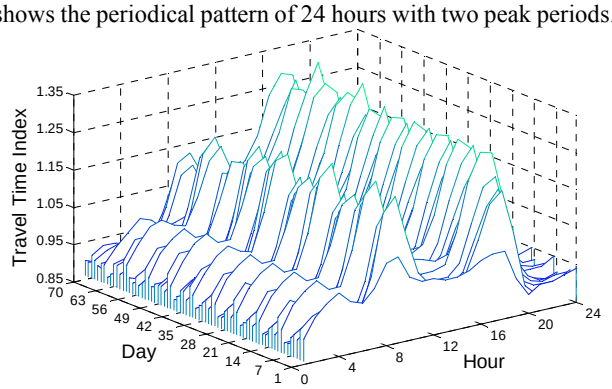
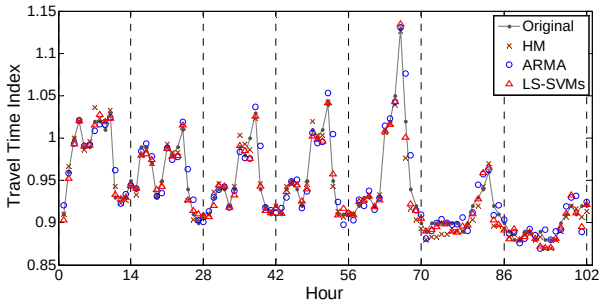


Fig. 2 Three dimensional graph of the first 10 weeks

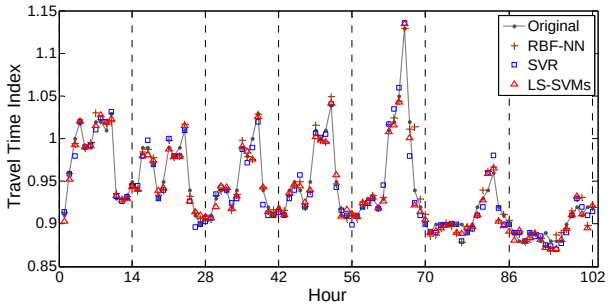
Based on the analysis of the TTI of each day, the non-peak periods from the data are selected. Specifically, the time intervals (6:00 am – 10:00 am, 15:00 pm – 19:00 pm on weekdays; 13:00 – 20:00 pm on weekends) contain the morning and evening peak periods. Therefore, there are 14 and 16 time points in the non-peak period for each weekday and each weekend day respectively. Namely, there are totally 102 time points lying in the non-peak hours of one week. And the aim of the paper is to predict these points in the 24th week.

Different parameters (K, d, etc.) are applied for the predictors in the experiments. Due to their general performance in our other studies, the most representative models with better performance measured in MAPE are chosen for simplicity: HM (K=3), ARMA (1,1), RBF-NN (d=1), SVR (d=3), and LS-SVR (d=4). For the three nonparametric techniques, the training process is based on the information in the non-peak periods of the former 23 weeks. Specifically, the former 12 weeks are selected for training, while the latter 11 are chosen for validation in the proposed model (N=23). Parameters adjusted appropriately help to obtain better performance. Due to distinct differences among the 102 time points of the week, 102 (ζ , σ) pairs with ζ ranging from 0.05 to 20000 and σ from 0.5 to 100 help LS-SVMs producing the final forecast. An LS-SVMs software kit is applied in the experiments [25].

Fig. 3 compares the forecast performance of the HM, ARMA, RBF-NN, and SVR models with that of the LS-SVMs model in two groups. Fig. 3(a) presents the original data and the forecasts from the parametric techniques at 102 time points separately, while Fig. 3(b) displays the results from the nonparametric techniques for comparison. Correspondingly, Fig. 4 compares the PEs of these predictors point by point.

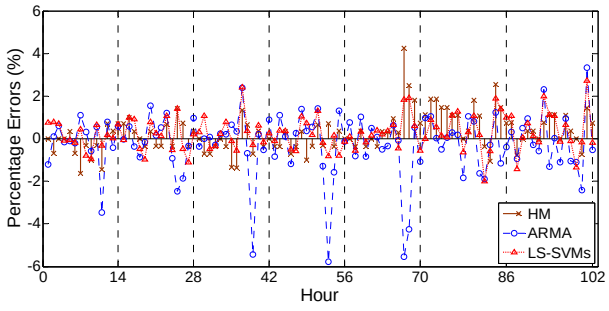


(a) HM, ARMA and LS-SVMs

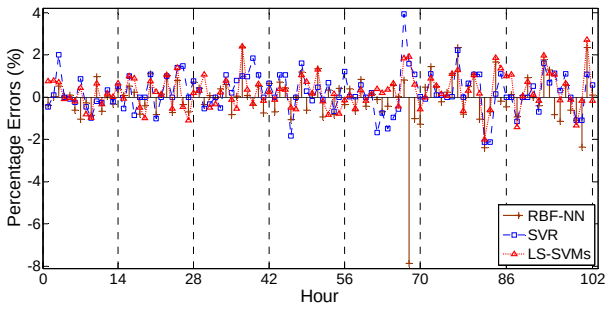


(b) RBF-NN, SVR and LS-SVMs

Fig. 3 Comparison of forecasts from the parametric and nonparametric models



(a) HM, ARMA and LS-SVMs



(b) RBF-NN, SVR and LS-SVMs

Fig. 4 Comparison of PEs from the parametric and nonparametric models

From the figures, it is obvious that most PEs of the nonparametric predictors are less than 4%. This indicates that the nonparametric techniques outperform the parametric ones. It also can be found that the simple HM model can provide relatively stable and accurate forecasts. Meanwhile, the two SVMs family members generally perform better, and nearly all the absolute values of PEs of our model are less than 2%. A predicted value from the RBF-NN model is visibly inaccurate (21:00 pm, Friday), which makes its PE up to about -8%. And the SVR model also produces such an inaccurate value with the PE up to 4% (20:00 pm, Friday). Moreover, it seems that more difficulties exist for all models to provide accurate forecasts on weekends.

Showing the prediction results from different models, Table I lays strong emphasis on the study using the measures of forecast accuracy: MAPE and VAPE. It can be easily noticed that the performance of the two SVMs family members is better than that of the other models. Moreover, compared with the HM, ARMA, RBF-NN, and SVR predictors, our model reduces 8.31%, 34.18%, 13.85%, and 6.61% in MAPE respectively. Meanwhile, it also reduces 15.28%, 49.73%, 38.31%, and 17.93% in VAPE respectively. This can fully demonstrate that our model is more accurate and robust than the other four.

TABLE I
COMPARISON OF PREDICTION PERFORMANCE IN MAPE & VAPE USING DIFFERENT PREDICTORS FOR NON-PEAK HOURS (%)

Predictor	MAPE	VAPE
HM (K=3)	0.7030	0.6553
ARMA (1, 1)	0.9793	1.1045
RBF-NN (d=1)	0.7482	0.9000
SVR (d=3)	0.6902	0.6765
LS-SVMs (d=4)	0.6446	0.5552

In order to compare the models from a holistic perspective, the analysis using another measure of forecast accuracy, the PE, is further produced. After the PEs are calculated for each model, the numbers of the predicted time points lying in different ranges of |PEs| (the absolute values of PEs) are statistically analyzed. The range boundaries are set as 1%, 2%, and 4%. For the five groups of non-peak period forecasts, Table II clearly shows the comparisons of the numbers computed from each model. It can be seen that only the two SVMs family members have no PEs above 4%. Simple comparisons show that our model has more predicted points lying in the range less than 1% and fewer points in the range more than 2%.

TABLE II
THE NUMBERS OF PREDICTED POINTS LYING IN DIFFERENT RANGES OF |PEs|

Predictor	[0, 1%]	(1%, 2%]	(2%,4%]	Above 4%
HM (K=3)	76	23	2	1
ARMA (1, 1)	67	25	6	4
RBF-NN (d=1)	79	17	5	1
SVR (d=3)	67	30	5	0
LS-SVMs (d=4)	78	22	2	0

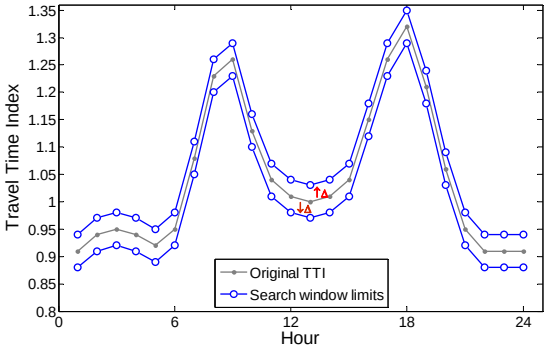


Fig. 5 Original TTI and search window

As shown in Figure 5, the observed TTI $V_{TTI}(t)$ ($t \in N$) and a variant Δ ($\Delta>0$, $\Delta \in R$) can produce a search window that is two Δ wide. At each time point t, the upper boundary of the window is $V_{TTI}(t)+\Delta$ and the lower one is $V_{TTI}(t)-\Delta$. The use of the window can determine the distribution of the predicted TTI and evaluate the accuracy of each model in another way. Obviously, when the search window expands with Δ increasing, more predicted points lie inside it.

Table III lists the specific numbers of the predicted TTI lying inside different ranges (Δ_1, Δ_2] using the models. In the table, Δ_1 and Δ_2 determine the search window W_1 with width=$2\Delta_1$ and W_2 with width=$2\Delta_2$ ($\Delta_1<\Delta_2$). The predicted points lying in the range (Δ_1, Δ_2] means that these points lie inside window W_2 and outside W_1. Examining the presented numbers, it can be seen that 51.96% forecasts from the LS-SVMs model lie inside the window with width=0.010 (Δ_2=0.005). Particularly, there are no predicted points lying outside the search window with Δ=0.040. This proves the robustness and accuracy of the proposed model from another point of view.

TABLE III
THE NUMBERS OF PREDICTED POINTS LYING IN DIFFERENT RANGES (Δ_1, Δ_2]

(Δ_1, Δ_2]	HM	ARMA	RBF-NN	SVR	LS-SVR
(0, 0.005]	52	41	46	48	53
(0.005, 0.010]	32	32	36	38	34
(0.010, 0.025]	17	23	19	15	14
(0.025, 0.040]	0	3	0	0	1
(0.040, 0.060]	1	3	0	1	0
(0.060, 0.075]	0	0	1	0	0

Measuring the forecast accuracy in four ways, the analyses show that the proposed model generally outperforms the others in non-peak traffic forecasting. It demonstrates the effectiveness and robustness of the LS-SVMs model applied in the special situation.

V. CONCLUSION

First applying the LS-SVMs in non-peak period forecasting, the case studies comprehensively compare the performance of two parametric and two nonparametric techniques. An hourly TTI time series is used in our experiments to demonstrate the effectiveness of our model, only as an example and because hourly traffic data are available to us. The LS-SVMs predictor shows its superiority because of its extraordinary ability of converging rapidly and avoiding local minimum. The good adaptability to forecast traffic status in non-peak hours evidences the potential applicability of the approach in real-time traffic forecasting.

REFERENCES

[1] B. Van Arem, H. R. Kirby, M. J. M. Van Der Vlist, and J. C. Whittaker, "Recent advances and applications in the field of short-term traffic forecasting," *Int. J. Forecast.*, vol. 13, no. 1, pp. 1-12, 1997.
[2] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term forecasting: Overview of objectives and methods," *Transport Rev.*, vol. 24, no. 5, pp. 533-557, 2004.
[3] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg, "Different methods of traffic forecast based on real data," *Eur. J. Oper. Res.* vol. 155, no. 3, pp. 558-568, 2004.
[4] W. H. K. Lam, Y. F. Tang, K. S. Chan, and M. L. Tam, "Short-term hourly traffic forecasts using Hong Kong annual traffic census," *Transp.*, vol. 33, no. 3, pp. 291-310, 2006.
[5] W. H. K. Lam, Y. F. Tang, and M.L. Tam, "Comparison of two non-parametric models for daily traffic forecasting in Hong Kong," *J. Forecast.*, vol. 25, no. 3, pp. 173-192, 2006.
[6] W. H. K. Lam, and J. Xu, "Estimation of AADT from short period counts in Hong Kong—A comparison between neural network method and regression analysis," *J. Adv. Transp.*, vol. 34, no. 2, pp. 249–268, 2000.
[7] C. H. Wu, J. M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 276-281, 2004.
[8] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. Part C.*, vol. 10, no. 2, pp. 85-98, 2002.
[9] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. Part C*, vol. 10, no. 4, pp. 303-321, 2002.
[10] B. L. Smith, and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transp. Eng.*, vol. 123, no. 4, pp. 261-266, 1997.
[11] S. Lee, Y. I. Lee, and B. Cho, "Short-term travel speed prediction models in car navigation systems," *J. Adv. Transp.*, vol. 40, no. 2, pp. 123-139, 2006.
[12] M. S. Ahmed, and A. R. Cook, "Analysis of freeway traffic time-series data by using Box-Jenkins techniques," *Transportation Research Record*, no. 722, pp. 1-9, 1979.
[13] D. Park, and L. R. Rilett, "Forecasting multiple-period freeway link travel times using modular neural networks," *Transportation Research Record*, no. 1617, pp. 63-70, 1998.
[14] A. Ding, X. Zhao, and L. Jiao, "Traffic flow time series prediction based on statistics learning theory," in *Proc. IEEE 5th Int. Conf. Intell. Transp. Syst.*, 2002, pp. 727–730.
[15] L. Vanajakshi, and L. R. Rilett, "A comparison of the performance of artificial neural network and support vector machines for the prediction of traffic speed," in *Proc. IEEE Intell. Vehicles Symp.*, 2004, pp. 194-199.
[16] J. A. K. Suykens, J. Vandewalle, and B. De Moor, "Optimal control by least squares support vector machines," *Neural Netw.*, vol. 14, no. 1, pp. 23-35, 1998.
[17] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. Singapore: World Scientific, 2002.
[18] T. Van Gestel, J. A. K. Suykens, D. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor, and J. Vandewalle, "Financial Time Series Prediction using Least Squares Support Vector Machines within the Evidence Framework," *IEEE Trans. Neural Netw.*, vol. 12, no. 4, pp. 809-821, 2001.
[19] R. E. Turochy, "Enhancing short-term traffic forecasting with traffic condition information," *J. Transp. Eng.*, vol. 132, no. 6, pp. 469-474, 2006.
[20] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik. "Support vector regression machines," *Adv. Neural. Info. Proc. Sys.*, vol. 9, NIPS, pp. 155-161, MIT Press, 1996.

[21] J. Q. Fan, and Q. W. Yao, *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer-Verlag, New York, 2003.
[22] S. Chen, C. F. N. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. Neural Networks*, vol. 2, no. 2, pp. 302-309, 1991.
[23] V. Vapnik, *Statistical Learning Theory*, New York, John Wiley, 1998.
[24] PeMS, Available: http://pems.eecs.berkeley.edu.
[25] LS-SVMlab Matlab/C toolbox,
Available: http://www.esat.kuleuven.ac.be/sista/lssvmlab.