Proceedings of the 11th International IEEE
Conference on Intelligent Transportation Systems
Beijing, China, October 12-15, 2008

A Multiple SVR Approach with Time Lags for Traffic Flow Prediction

Tianshu Wu, Kunqing Xie, Guojie Song, Cheng Hu

Abstract—**A multiple support vector regression (SVR) model with time lags was proposed for short term traffic flow prediction. Time lags between current traffic flow and upstream traffic flow were estimated in order to make better use of spatial-temporal correlation between the upstream and the downstream. The time lags could help identify the upstream flow series most similar to that of the current road and to be used as the model input. A global SVR model with a time lag was constructed and we found it performed not so well during some time intervals where the traffic flow was dramatically fluctuant. Local SVR models with time lags were constructed especially for those intervals and improved the performance. Combining both of the global and the local models, the multiple model was applied to 5-minute freeway data observed by loop detectors in the project Freeway Performance Measurement System (PeMS) of California. Comparisons with several other methods showed that the multiple SVR model with time lags was a promising and effective approach for traffic flow prediction.**

I. INTRODUCTION

Recent years, many strategies are proposed to ease traffic congestion, such as Advanced Traffic Management Systems (ATMS) and Advanced Vehicle Control Systems (AVCS), which are components of Intelligent Transportation Systems (ITS). All of the systems need accurate traffic flow prediction in order to work properly.

Short term traffic flow prediction is usually defined as prediction about traffic flow in future 5 to 30 minutes [1]. There are some methods for this issue. Time series analysis methods such as AR, MA, ARMA, ARIMA etc are used for the prediction [2][3]. However, they are easy to be disturbed by noise data, and as a result, perform not so well when applied to prediction of short term traffic flow with much randomness. Being capable to deal with nonlinear problem, neural networks are widely used in this research field [4][5][6]. Zheng et al. proposed a forecasting approach that combined back propagation and the radial basis function neural networks based on the theory of conditional probability and Bayes' rule [7]. Vlahogianni et al. proposed an optimization framework for construction and working process of neural networks in the context of short-term traffic flow prediction [8]. Genetic algorithms were used to refine neural network and locally weighted regression models to predict short-term traffic with better accuracy [9]. There are also many other methods, such as Kalman filter theory [10], nonparametric methods [11], local regression models [12], and Markov chain models [13] etc. Sun et al. used an approach based on Bayesian networks for traffic flow forecasting [14]. Ding et al. used ordinary support vector regression in predicting traffic flow [15]. Chaos Theory was also used for traffic flow prediction as complicated traffic system reflected chaotic characteristics [16].

Traffic flow on the object road is affected by flow on the adjacent roads (mainly upstream roads). Some of the approaches above did not utilize the information from the adjacent roads and other methods used traffic flow of the adjacent roads without careful analysis of the temporal relation between traffic flow on the current road and the upstream roads. Another drawback of the existing methods is that the character of traffic flow is variable and it is difficult to predict a whole day traffic flow in different states by the same single model, leading to lower accuracy.

In this paper, we use a multiple Support Vector Regression (SVR) model to make prediction of short term traffic flow. Compared to neural networks, the nonlinear SVR model based on statistical learning theory could obtain the global optimal results. It has strong generalization ability and faster convergence speed. We focus on applying SVR to short term traffic flow prediction based on traffic flow of the current road, the upstream roads and historical average flow of the current road. Similarities between the temporal series of the current flow and its upstream flow with different time lags are estimated to find the suitable time lags. Upstream flow with a proper time lag would guaranty a better prediction to the object road. A global SVR model with a time lag is constructed to make a whole day prediction and an error analysis is given, based on which local SVR models with time lags are utilized to improve the local performance. Experiments show that the multiple SVR model with time lags is applicable for traffic flow forecasting.

The rest of the article is organized as follows. Section 2

Manuscript received June 13, 2008. This work was supported by the National Natural Science Foundation of China under Grant No. 60703066 and supported by the National High-Tech Research and Development Plan of China (863) under Grant No.2006AA12Z217.

Tianshu Wu is with the Key laboratory of Machine Perception (Peking University), Ministry of Education, Beijing, 100871. e-mail: wutianshu@ cis.pku.edu.cn

Kunqing Xie is with the Key laboratory of Machine Perception (Peking University), Ministry of Education, Beijing, 100871. e-mail: kunqing@ cis.pku.edu.cn

Guojie Song (the corresponding author) is with the Key laboratory of Machine Perception (Peking University), Ministry of Education, Beijing, 100871. e-mail: gjsong@ pku.edu.cn

Cheng Hu is with the Key laboratory of Machine Perception (Peking University), Ministry of Education, Beijing, 100871. e-mail: hucheng@ cis.pku.edu.cn

gives a brief introduction to SVR. The multiple model is constructed in section 3 and section 4 reports the experiments based on the realistic data of the project Freeway Performance Measurement System (PeMS) [17][18]. Section 5 concludes the paper.

II. BRIEF INTRODUCTION TO SUPPORT VECTOR REGRESSION

Traffic flow is time-variant and has characters of non-linear and randomness. The SVR method can be used to make prediction in nonlinear systems, like artificial neural networks, but it has stronger generalization ability because the support vector machine theory is based on the minimization principle to structure risk. On the other hand, SVR could work well with small samples, so it has good potential to make online prediction.

For a given set of data ($\boldsymbol{x}$, y), the SVR algorithm [19] tries to find the function

$$y = f(\boldsymbol{x}) = w^T\phi(\boldsymbol{x}) + b \quad (1)$$

It can be transferred to a optimization problem as follows.

$$\begin{aligned} \min \quad & \frac{1}{2}w^T w + \frac{C}{2}\sum_{i=1}^{k}\xi_i^2 + \frac{C}{2}\sum_{i=1}^{k}(\xi_i^*)^2 \\ s.t. \quad & -\xi_i^* - \varepsilon \le w^T\phi(x_i) + b - y_i \le \xi_i + \varepsilon \\ & \xi_i, \xi_i^* \ge 0, i = 1\cdots k \end{aligned} \quad (2)$$

The first term in formula (2) objects at maximize the classification interval, which will make the function smoother. The rest terms of formula (2) compose the loss function to minimize the errors. Constant C (C>0) is the soft margin parameter, weighting the penalization of errors lager than e. Function ϕ mapping the data in the original space to a high dimension space in order to perform the liner regression.

After some reformulation based on Lagrange optimization and duality principle, the optimization problem is then transformed into

$$\begin{aligned} \min \quad & \frac{1}{2}(\alpha-\alpha^*)^T \tilde{K}(\alpha-\alpha^*) + \sum_{i=1}^{k}(\alpha_i-\alpha_i^*)y_i + \varepsilon\sum_{i=1}^{k}(\alpha_i+\alpha_i^*) \\ s.t. \quad & \sum_{i=1}^{k}(\alpha_i-\alpha_i^*) = 0,\ 0 \le \alpha_i, \alpha_i^*,\ i = 1\cdots k \end{aligned} \quad (3)$$

where α_i, α_i^* are Lagrange multipliers, solutions of the problem. $K(x, y)$ is a kernel function with the form $K(x_i, y_j) = \phi(x_i)^T\phi(x_j)$. In our experiments, we used the Gaussian Radial Basis Function (RBF) kernel:

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\delta^2}\right) \quad (4)$$

Then we can estimate a new point by $f(\boldsymbol{x}) = w^T\phi(\boldsymbol{x}) + b$, and w and b can be obtained by

$$\begin{aligned} & w = \sum_{i=1}^{k}(\alpha_i - \alpha_i^*)\phi(x_j) \\ & b = \begin{cases} y_l + (\tilde{K}(\alpha-\alpha^*))_l - \varepsilon & \alpha_l > 0 \\ y_l + (\tilde{K}(\alpha-\alpha^*))_l + \varepsilon & \alpha_l^* > 0 \end{cases} \end{aligned} \quad (5)$$

III. MULTIPLE MODEL CONSTRUCTION

In order to make better use of spatial-temporal correlation between upstream and downstream, the time lags between current traffic flow and upstream traffic flow are estimated, which could identify the upstream flow series most similar to that of the current road. The similar flow series on the upstream would be helpful to the prediction of the current road traffic flow.

Traffic conditions vary a lot in different time intervals. It is difficult to make prediction only by one global model. In the later experiments, we find that the global model performs not so well during some time intervals when the traffic flow is varying dramatically, so we construct local models for those intervals and improve the result.

Traffic data is usually collected at a fixed interval, for example 5 minutes. The data points ordered by time in a whole day compose a temporal series and the serial number can represent the corresponding time. Follows are some definitions. For convenience of statement, here we assume there is only one upstream road. It is easy to extend the model to more than one upstream roads.

TABLE 1
SYMBOL DEFINITIONS

C_i	The temporal series composed of data points in the i-th day on the current road
U_i	The temporal series composed of data points in the i-th day on the upstream road
H	The temporal series of historical data, computed by data in the past several days, for instance from the s-th day to the e-th day. $H = \frac{1}{s-e+1}\sum_{i=s}^{e} C_i$
$C_i[m, n]$	The subsequence from the m-th point to the n-th point in C_i
$U_i[m, n]$	The subsequence from the m-th point to the n-th point in C_i
$H[m, n]$	The subsequence from the m-th data point to the n-th data point in H

A. Time Series Similarity

Traffic conditions vary considerably in the whole road network. Traffic flow in the current road has relation with flow in all other roads in different extent. It is impracticable to use information in all the other roads in all the time intervals to make prediction of flow in the current road due to the complexity of both the model and the computation. It is reasonable to assume that flow in current road is only affected by flow in the adjacent roads (mainly the upstream roads) in limited time intervals. Then the key point is to identify the time lag with which the upstream flow series would be most relevant, or similar to the current flow series.

The similarity between two temporal series of equal length $C_i[m_1,n_1]$ and $U_i[m_2,n_2]$ denoted as $Sim(C_i[m_1,n_1], U_i[m_2,n_2])$ can be estimated as follows. Standardize the series by

$$C_i[m_1,n_1]_{std} = \frac{C_i[m_1,n_1] - C_i[m_1,n_1]_{min}}{C_i[m_1,n_1]_{max} - C_i[m_1,n_1]_{min}} \tag{6}$$

where $C_i(m_1,n_1)_{min}$ is the minimum in series $C_i[m_1,n_1]$ while $C_i(m_1,n_1)_{max}$ is the maximum in it. $U_i[m_2,n_2]$ is standardized in the same way. Then

$$Sim(C_i[m_1,n_1], U_i[m_2,n_2]) = \begin{cases} 1 - \frac{1}{n_1-m_1+1}\left|C_i[m_1,n_1]_{std} - U_i[m_2,n_2]_{std}\right| & 0 \le m_1 - m_2 \le T \\ 0 & others \end{cases} \tag{7}$$

We assume that two series with a lag larger than T (in our work set as 15) is uncorrelated in formula (7) in order to simplify the problem. The time lag between flow on the current road and the upstream road during time interval [m, n] is $Lag_{C,U}[m,n]$. Given historical data from the s-th day to the e-th day(s and e are identified by user),

$$Lag_{C,U}[m,n] = \{\, l \mid \forall x \sum_{j=s}^{e} \frac{Sim(C_j[m,n], U_j[m-l,n-l])}{e-s+1} \ge \sum_{j=s}^{e} \frac{Sim(C_j[m,n], U_j[m-x,n-x])}{e-s+1} \} \tag{8}$$

B. *The Multiple SVR Model*

The process of the multiple SVR model construction is shown in figure 1. A global SVR model with a time lag is constructed. Time intervals in which local SVR models are needed are identified by a leave-one-out method and then local models are constructed. The multiple SVR model is obtained by combining the global and the local models.

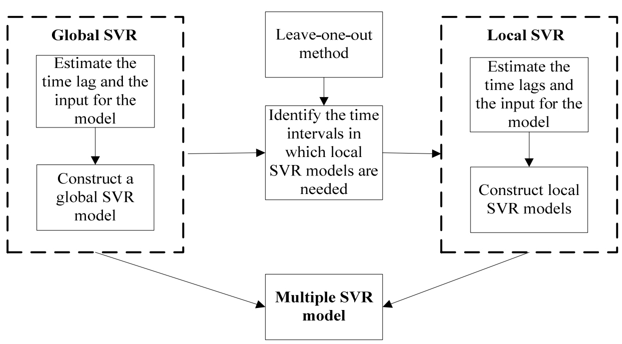

Fig. 1. Process of multiple model construction

Traffic conditions at the k-th point in the temporal series are represented by the temporal series [k-3,k], for this configuration could provide enough information for traffic flow forecasting empirically. So the future traffic flow at the next point in time of the current road $C_i(k+1)$ is forecasted based on the traffic condition of the current road $C_i[k-3,k]$, upstream road $U_i[u-3,u]$ and the historical condition of the current road $H[k-3,k]$, where $u=k-l+1$ and l is the time lag between flow of current road and upstream roads. Then the target is to construct a model f by the SVR method, such that

$$\tilde{C}_i(k+1) = f(C_i[k-3,k], U_i[u-3,u], H[k-3,k]) \tag{9}$$

Assume there is traffic flow data for fifteen days. We use the average value of the former seven days as the historical temporal series H. $H=\frac{1}{7}\sum_{i=7}^{7} C_i$. The later seven days are utilized to train the global SVR model. So the training set is $\{C_i, U_i, H \mid i = 8 \text{ to } 14\}$. The last day is to be predicted to test the model. That is, comparing the computed $\tilde{C}_i(k+1)$ to the observed data $C_i(k+1)$. The time lag $l=Lag_{C,U}([T, End])$ where T is the constant in formula (7) and End is the last point in a whole day temporal series.

There're three adjustable parameters in the SVR method. In formula (2), C is the regularization parameter; for C low minimizing of the complexity of the model is emphasized and for C high, good fitting of the training data points is stressed. δ^2 is the parameter of the kernel; in the common case of an RBF kernel, a large δ^2 indicates a stronger smoothing. ε estimates the tolerance to the error during the regression. We use "Leave-one-out" (LOO) method [20] to estimate the parameters for the global SVR.

The global SVR model is trained by data of whole days. It reflects the primary trend of the traffic flow in a whole day. The structure of the model with a relatively small parameter C is too smooth to have a accurate prediction to the flow varying dramatically. Traffic flow variation between adjacent sample points is introduced to measure the fluctuation of the flow. The fluctuation in time interval [m, n] is defined as $flu[m,n] = \sum_{i=m}^{n-1}|v_{i+1} - v_i|$ where v_i is in the time interval in which the flu is computed. The later experiment shows that the prediction accuracy tends to be low when the fluctuation is large.

Local SVR models are constructed especially for those time intervals to improve performance in fluctuant intervals. We use leave-on-out method in the training days to identify the time intervals for which local SVR models are needed. For each turn, take out one day for the global SVR prediction, leaving the rest days for training. From the prediction result of each turn, we obtain the average *RMSE* (refer to formula (11)) values for each hour and then the local SVR intervals could be identified. Once the intervals have been settled, we don't have to recompute it every time before prediction until the accuracy turns low.

The training set for the local SVR model is $\{C_i[m,n], U_i[m-l', n-l'], H[m,n] \mid i = 6 \text{ to } 9\}$ where [m, n] is the fluctuant interval. The time lag $l'=Lag_{C,U}([m,n])$. As stated in the later experiments, data points in [m, n] are more fluctuant. As a result, the parameter C of the local SVR turns out to be larger than that of the global model, meaning the local models being more fluctuant than the global model. The multiple model combines predictions of the global and the local models, as

$$\tilde{C}_i(k+1) = \begin{cases} f_{global}(C_i[k-3,k], U_i[u-3,u], H) & k \notin [m,n] \\ f_{local}(C_i[k-3,k], U_i[u'-3,u'], H) & k \in [m,n] \end{cases} \tag{10}$$

where $u=k-l+1$ and $u'=k-l'+1$.

IV. Experiments

A. Data Description

We used the real traffic volume data in California, USA. The data was from the project Freeway Performance Measurement System (PeMS), which collected historical and real-time freeway data from freeways in the State of California. In PeMS, California was divided into twelve districts in which the seventh district Los Angeles had notorious traffic. We chose a two-way six lanes freeway labeled "I5-S" crossing the seventh district on which there were two loop detectors with the number 716911 and 716902. As shown in figure 2, 716911 and 716902 were on the same direction, south-east, of the two-way freeway I5-S and 716911 was on the upstream of 716902. They both recorded the volume travel though them in the south-east three lines every 30 seconds and PeMS provided 5 minutes volume by aggregation on the raw 30 seconds data.

The experiment dataset was composed of volume data of 15 consecutive days, April 7 to 21, year of 2008. The first seven days were used to compute the historical volume, the latter seven days were for training the SVR model while the last day was for prediction to test the model's performance.

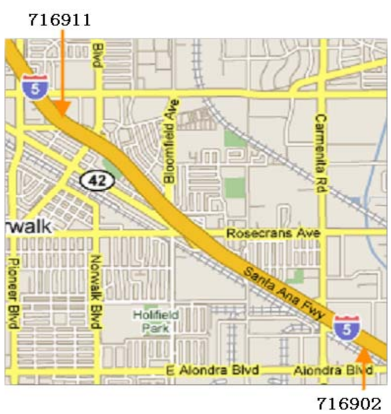

Fig. 2. Data source for experiments

Mean Absolute Percentage Error (*RMSE*) was used to estimate the accuracy of the prediction, defined as

$$RMSE(y, y') = [\frac{1}{N}\sum_{n=1}^{N}(y(n) - y'(n))^2]^{\frac{1}{2}} \tag{11}$$

where y was the actual volume and y' was the corresponding prediction while N was the total number of prediction values. The smaller *RMSE* was, the better accuracy the model had.

B. Global Model

The time lag between the upstream road and the current road was computed by the data of the first seven days. Figure 3 showed average similarities of two flow series under different time lags. The time lag of was set as 1.

The result was shown in figure 4. *RMSE* of the global prediction was 29.75, which is a little better than the ordinary SVR without a time lag as shown in the latter experiment in figure 8, where the average *RMSE* is 30.42.

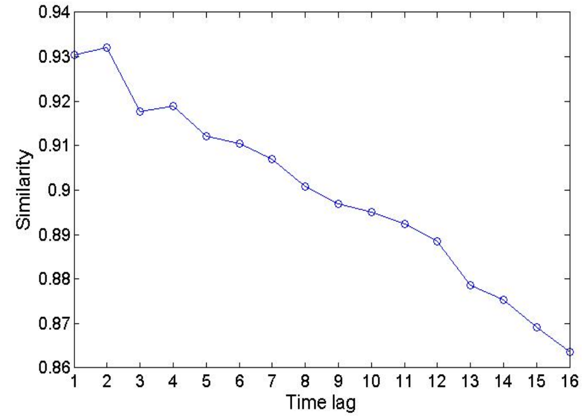

Fig. 3. Similarities under different lags

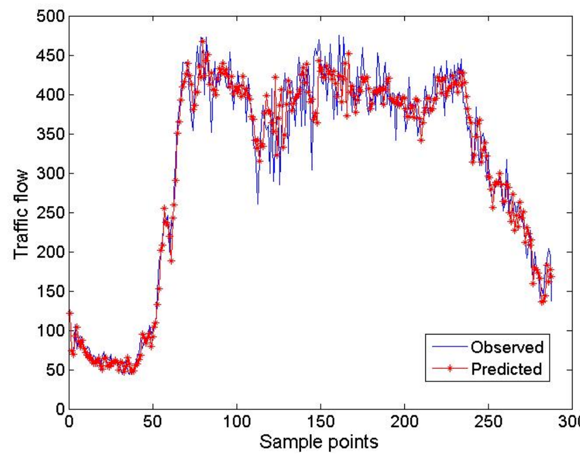

Fig. 4. The result of the global model

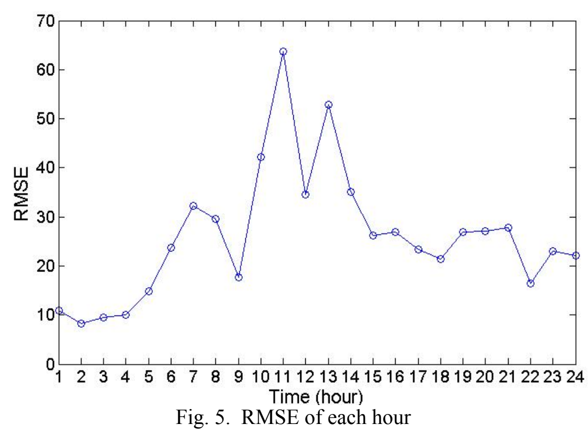

Fig. 5. RMSE of each hour

In figure 5, the *RMSE* of each hour, computed by every 12 sample points, was quite different from each other. The global SVR model performed not so well from 6 o'clock to 8 o'clock and from 9 o'clock to 14 o'clock, where the *RMSE* reached higher values than average. We computed the fluctuation of the data, shown in figure 6. Compared with *RMSE* in figure 5, it could be found that the prediction accuracy tended to be low when the fluctuation was large.

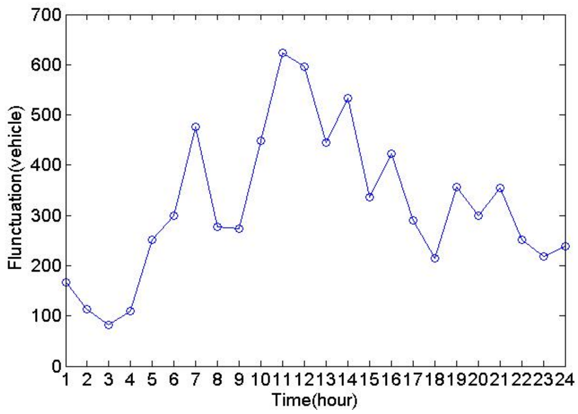

Fig. 6. Fluctuation of each hour

C. Multiple Model

In this section, we concentrated on the time intervals in which the global model not performed well and constructed local SVR models for them. Then we combined the global and the local SVR models together to make more accurate prediction.

We used data of the eighth to the fourteenth day and leave-one-out method to identify the time intervals for which local SVR models were needed. The time interval was from 9 o'clock to 14 o'clock, and that was [108, 168]. A local SVR model was constructed for this time interval, which meant all the data used to train the model was in this time interval of different days. Furthermore, we recomputed the time lag using the historical data in this time interval. Then we make a comparison of predictions for the time interval obtained from the global and the local SVR model, as shown in figure 7. The result of the local model was better than that of the global model in this time interval. For the multiple model, the RMSE of prediction for the whole day was 28.2.

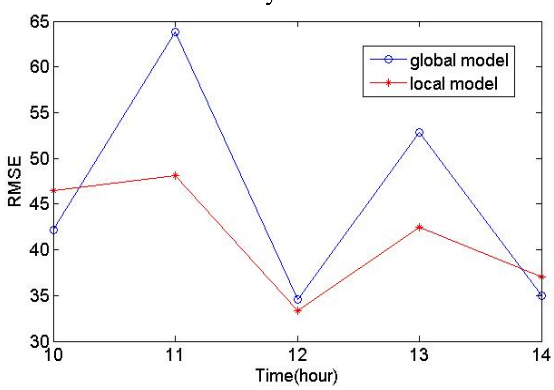

Fig. 7. Comparison between global and local model during the certain time interval

We randomly chose other 5 groups of data, each containing fifteen consecutive days. The neural network, ordinary SVR and chaotic methods were adopted as the comparison methods. The neural network was a typical method to create nonlinear mapping between input and output. In this model, $C_i[k\text{-}3,k]$, $U_i[k\text{-}3,k]$, $H[k\text{-}3,k]$ were used as input data while $C_i(k+1)$ was the expected output data. The number of hidden layer and learning rate were obtained by a grid search. The ordinary SVR method was implemented according to Ref.[15] A radial kernel was used and the parameters were obtained by a grid search. The chaotic method in our experiment was the same as that in Ref.[16] and the delayed time was 20s while the embedding dimension was 9. The *RMSE* values of each methods were shown in figure 8 from which we could see the multiple SVR model with time lags was more accurate than other methods.

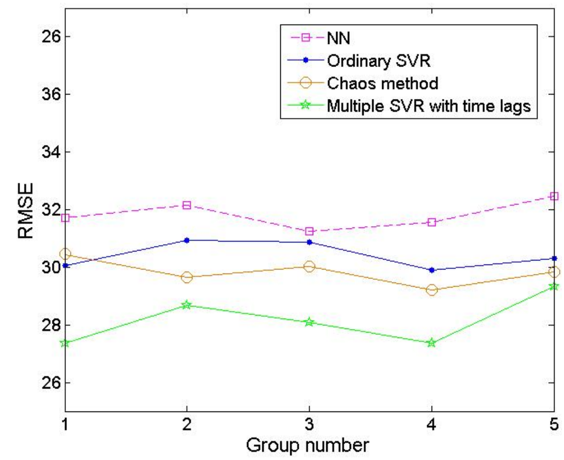

Fig. 8. Comparisons among several models

Neural networks are thought to be practical in most research. The SVR method is slower than neural networks, so we suggest that the number of models in this work is between 2 to 4 to control the computing complexity and that was enough for most cases according to our experience. Then the running time of our approach is about three times longer than that of neural networks, which is acceptable considering the rapidly improvement of computer performance. Optimizing the approach in accuracy and speed is our future work.

V. Conclusion

Traffic flow of the current road has great correlations with flow of its upstream roads. We analyzed the similarities between flow of the current road and its upstream roads with different time lags to find the time lag with which the two series were most correlated. Then traffic flow of the current road at next time point was predicted using current traffic flow, upstream traffic flow with a time lag and historical average flow of the current road. A global SVR model was constructed to make a whole day prediction and an error analysis was given, based on which local SVR models are used to improve the local performance. Experiments for the combined SVR model were carried out on the data of a traffic project PeMS. The results showed that the multiple SVR model with time lag was a promising and effective approach for traffic flow prediction.

Acknowledgment

The authors would like to thank X.P. Gao for his sincere help on realization of the SVR algorithm and some constructive suggestions. The authors are also grateful to the anonymous editor and reviewers for giving valuable comments.

REFERENCES

[1] S.l. Sun, C.S. Zhang, "The Selective Random Subspace Predictor for Traffic Flow Forecasting" *IEEE Transactions on Intelligent Transportation System*, vol.8, no.2, pp.367-373, 2006.

[2] C. K. Moorthy, B. G. Ratcliffe, "Short term traffic forecasting using time series methods," *Transp. Plan. Technol.*, vol. 12, pp. 45–56, 1988.

[3] B Ghosh, B Basu, M. Mahony, "Bayesian Time-Series Model for Short-Term Traffic Flow Forecasting," Journal of Transportation Engineering, vol. 133(3), pp.180-189, 2007.

[4] E.S. Yu, C.Y.R. Chen, "Traffic prediction using neural networks," *Global Telecommunications Conf.*,vol.2, pp.991-995, 1993.

[5] B. Park, C.J. Messer, T.Urbanik II, "Short-term freeway traffic volume forecasting using radial basis function neural network", *Transportation Research Board*, Washington D C: TRB, pp. 39-46, 1998.

[6] B. Abdulhai, H. Porwal, W. Recker, "Short-term freeway traffic flow prediction using genetically optimized time-Delay-based neural networks", *Transportation Research Board*, Washington, pp. 1-14, 1999.

[7] W.Zheng, Lee, D.-H., Q.Shi, "Short-term freeway traffic flow prediction : Bayesian combined neural network approach," *Journal of Transportation Engineering*, 132(2), 114–121, 2006.

[8] E. I. Vlahogianni, M. G. Karlaftis, J. C. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research* Part C, 211–234, 2005.

[9] M. Zhong, S. Sharma, P. Lingras, "Short-term traffic prediction on different types of roads with genetically designed regression and time delay neural network models," *Journal of Computing in Civil Engineering*, 19(1), 94–103, 2005.

[10] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filter theory," *Transp.Res. Part B: Methodol*, vol.18, pp. 1-11, 1984.

[11] Y. F. Tang, William H. K. Lam, Mei-Lam Tam. "Comparison of two non-parametric models for daily traffic forecasting in Hong Kong," *Journal of Forecasting*, vol.25, pp. 173-192, 2006.

[12] E. B. L. Smith and M. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *Journal of Transportation Engineering*, vol. 123, no. 4, pp. 261–266, 1997.

[13] G. Yu, J. Hu, C. Zhang, L. Zhuang, J. Song, "Short-term traffic flow forecasting based on Markov chain model," *IEEE Proceedings Intelligent Vehicles Symposium*, pp. 208- 212, 2003.

[14] S. Sun, C. Zhang, G. Yu, "A bayesian network approach to traffic flow forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol.7(1), 124-132, 2006

[15] A. Ding, X. Zhao, L. Jiao, "Traffic flow time series prediction based on statistics learning theory," *The IEEE 5th International Conference on Intelligent Transportation Systems*, 727- 730, 2002.

[16] L. Zang, L. Jia, L. Yang, T. Liu, "Chaotic Time Series Model of Real-time Prediction of Traffic Flow," *China Journal of Highway and Transport*, vol. 20(6), 95-99, 2007.

[17] P. Varaiya, "California's performance measurement system: Improving freeway efficiency through transportation intelligence," *TR News*, 2002.

[18] UC Berkeley, PeMS, http://pems.eecs.berkeley.edu.

[19] AJ. Smola, B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, 2001.

[20] M. Chang, C. Lin, "Leave-one-out Bounds for Support Vector Regression Model Selection," *Neural Computation*, vol. 17, pp. 1188-1222, 2005.