Unsupervised Learning Based Performance Analysis of ν-Support Vector Regression for Speed Prediction of A Large Road Network

M. T. Asif†, J. Dauwels†, C. Y. Goh†, A. Oran‡,
E. Fathi†, M. Xu†, M. M. Dhanya†, N. Mitrovic† and P. Jaillet§

***Abstract*— Many intelligent transportation systems (ITS) applications require accurate prediction of traffic parameters. Previous studies have shown that data driven machine learning methods like support vector regression (SVR) can effectively and accurately perform this task. However, these studies focus on highways, or a few road segments. We propose a robust and scalable method using ν-SVR to tackle the problem of speed prediction of a large heterogenous road network. The traditional performance measures such as mean absolute percentage error (MAPE) and root mean square error (RMSE) provide little insight into spatial and temporal characteristics of prediction methods for a large network. This inadequacy can be a serious hurdle in effective implementation of prediction models for route guidance, congestion avoidance, dynamic traffic assignment and other ITS applications. We propose unsupervised learning techniques by employing k-means clustering, principal component analysis (PCA), and self organizing maps (SOM) to overcome this insufficiency. We establish the effectiveness of the developed methods by evaluation of spatial and temporal characteristics of prediction performance of the proposed variable window ν-SVR method.**

I. INTRODUCTION

Data driven learning techniques like support vector regression (SVR) are effective time series predictors [1]–[3]. These techniques have found applications in many diverse fields such as financial sector [4], packet networks [5], and weather forecasts [6]. Methods employing SVR are particularly suitable for road traffic parameters prediction and estimation, due to prevalent non-linear relationships amongst traffic variables. Machine learning techniques like artificial neural networks (ANN) and SVR consistently provide better results than traditional regression methods for prediction of different traffic parameters like travel time, flow and speed [7]–[22]. These studies, however, concentrate on custom scenarios like highways or a few intersections. Practical road networks are much more complex. Intelligent transportation systems (ITS) applications like route guidance, or congestion avoidance would require prediction results for generic networks. In this paper, we examine whether SVR based prediction method can be applied to a more practical road network environment, comprising of thousands of road segments with different capacities, speed limits, and lanes, for different prediction horizons. This problem will be referred as large scale prediction problem. Any algorithm or architecture which deals with this problem should be modular, easily scalable and robust. We propose a temporal window based SVR method to perform large scale prediction and compare its results with prediction performance of ANN and Holt's exponential smoothing predictors.

Secondly we develop novel techniques for temporal and spatial performance evaluation of a prediction algorithm. Prediction performance is usually evaluated using mean absolute percentage error (MAPE) and root mean square error (RMSE) [7]–[24]. For large scale prediction, these prevailing point estimation methods provide little insight into actual performance of the model. To overcome this inadequacy, we propose novel methods utilizing k-means clustering, principal component analysis (PCA) and self organizing maps (SOM) for performance evaluation. To the best of our knowledge this is the first attempt to address this problem.

Singapore's land transportation authority (LTA) provided the data set for experimental purposes. It contains two months of speed data (March-April, 2011) for each road segment in Singapore. LTA collected the data using a range of on site sensors. We consider in this paper a subnetwork that consists of a continuous stretch of road network from Outram to Changi (Fig. 1). The selected area contains different types of roads. It spans over parts of Singapore's main highways (Pan Island Expressway and East Coast Park highway). The area also includes some other major roads in the downtown area, and urban arterial roads with significant traffic volumes. We did not include road segments for which little data is available. Using this criteria, we selected a total of 5024 road segments for the study.

The paper is structured as follows. In Section II we develop the architecture of variable window ν-SVR method. Performance comparison with other time series prediction methods is provided in Section III. Sections IV and V deal with development and evaluation of unsupervised learning techniques for spatial and temporal prediction performance evaluation of variable window ν-SVR. At the end of the paper (Section VI), we summarize our contributions and suggest topics for future work.

II. VARIABLE WINDOW BASED SVR FOR SPEED PREDICTION

Definition 1: (Road Network) A road network is defined as a directed graph $\mathbf{G} = (N,E)$, where $E = \{s_i | i = 1,...,m\}$

†School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798; jdauwels@ntu.edu.sg, muhammad89@e.ntu.edu.sg, cygoh@pmail.ntu.edu.sg

‡Center for Future Urban Mobility, Singapore-MIT Alliance for Research and technology, Singapore, 117543; aoran@smart.mit.edu

§Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 02139; jaillet@mit.edu

represents the set of road segments/links.

Definition 2: (Speed Profile) A Speed profile for road segment s_i is set of speed values for that link such that $\mathbf{U}(t_j,s_i)$ represents average speed of the link during interval $(t_j - t_0,t_j)$. $\tilde{\mathbf{U}}_k(t_j,s_i)$ is the predicted speed of the link during interval $(t_j - t_0,t_j)$ for k^{th} prediction horizon. t_0 is the sampling interval for data, which is 5 minutes for the data set at hand.

In this section, based on above definitions, we develop a robust and scalable SVR architecture, to deal with the problem of large scale prediction. The objective of the method is to perform prediction of future speed profiles for individual links based on current and past speed trends. To make the architecture modular, we will perform prediction for each road segment individually.

A. SVR based time series prediction - Theoretical Overview

We use SVR to extract the relationship between given and future speed values from training data to perform speed prediction. For a link s_i, consider a set of vectors of given speed values $\{\mathbf{x}_v \in \mathbb{R}^n | v = 1,...,l\}$, and corresponding future speed values $\{y_v \in \mathbb{R} | v = 1,...,l\}$. We will perform SVR training by feeding SVR with given and target speed value pairs $(\mathbf{x}_v,y_v)$. SVR will then try to find a function to replicate these trends. Each input $\mathbf{x}_v$ contains "n" input features. To exploit the relationship between these features, each input vector is mapped into higher order space using a kernel function. If $\chi = \{\mathbf{x}_1,...,\mathbf{x}_l\}$ is the input feature set, then the chosen kernel function $\phi(\bullet)$ defines this mapping ϕ: $\chi \rightarrow \mathcal{F}$. In this section, we provide a brief overview of SVR methodology. A more rigorous treatment of the topic can be found in [25]–[29]. For ε-SVR, this problem can be formally stated as in [27]:

$$min : \mathrm{z} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}(\xi_i + \xi_i^*), \tag{1}$$

where ξ_i and ξ_i^* are slack variables, introduced for constraint optimization [27]. Support vector method employs so called ε insensitive lose function, which only penalizes input points outside soft margin defined by bounds $\pm\varepsilon$. C is the cost function, associated with the training errors larger than the bound $|\varepsilon|$. $\mathbf{w}$ is the hyperplane.

Difficulty in choosing the appropriate value of error bound (ε) led to the development of a new class of SVR called ν-SVR [29]. With introduction of ν, the minimization problem in (1) takes the form [29]:

$$min : \mathrm{z} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C(\nu\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)). \tag{2}$$

It can be shown that ν encapsulates upper and lower bounds for fraction of training errors and support vectors respectively [29]. Applying the Lagrange multiplier technique to (1) and (2), will yield resultant function [27], [28]:

$$f(x) = \sum_{i=1}^{l}(\alpha - \alpha^*)\phi(x_i)^T\phi(x) + b. \tag{3}$$

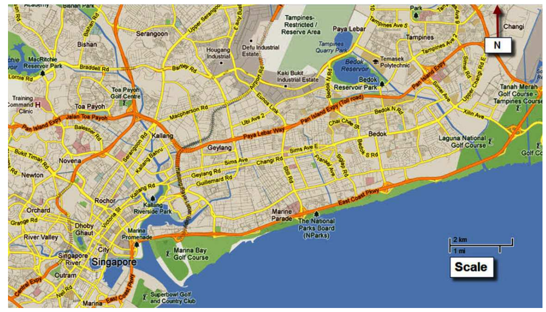


Fig. 1: The map of region for speed prediction

The goal of training SVR is to find $f(x)$ which can provide most suitable representation of data set [27].

B. Feature selection using temporal variable window

Future state of traffic parameters depends upon historical behavior of the given road segment and its neighboring links. These relationships have been utilized for traffic parameter prediction by both machine learning techniques [30], [31] as well as alternative methods [23], [24]. If both spatial and temporal relationships are taken into consideration then:

$$\mathbf{Y}(t + nt_0, s_k) = \mathcal{F}[\mathbf{U}(t,s_k),...,\mathbf{U}(t - mt_0,s_j)]. \tag{4}$$

In (4), $\mathcal{F}[\bullet]$ defines the relationship between historic speed values of given road segment, its neighbors $\mathbf{U}(t,s_k)$,...,$\mathbf{U}(t - mt_0,s_j)$, and future speed values of the given segment $\mathbf{Y}(t + nt_0,s_k)$. By neglecting the spatial features in (4), the equation reduces to window method for feature selection [22]:

$$\mathbf{Y}(t + nt_0, s_k) = \mathcal{F}[\mathbf{U}(t,s_k),...,\mathbf{U}(t - mt_0,s_k)]. \tag{5}$$

Spatial relationship are more difficult to extract and require more computations [30], [31]. This complexity strongly limits the overall scalability of the prediction system. Min et al. proposed spatial and temporal correlations, as a measure to find relevant neighboring links [23]. However, correlation methods fails to capture the non-linear relationships [30]. As a result, many prediction studies utilize only past trends of the road segment for prediction [8], [12], [18], [20]. The resulting prediction methods are scalable. For instance, assume that we wish to extend our subnetwork with additional nodes $\tilde{E} = \{g_k | k = 1,...,r\}$: $g_k \notin E$. $\tilde{E}$ may represent neighboring area of the test network. Since all predictions only need local information, we can just learn predictors for the extra nodes, without having to re-calibrate the existing predictors. However, methods which exploit spatial relations would need to be re-calibrated in this case [23], [30], [31]. Instead of choosing a fixed temporal window nt_0 for all prediction horizons [8], [12], [18], [20], [22], we couple the length of temporal window to the prediction horizon by choosing $n = m$ in (5).

C. Parameter selection for SVR

Kernel functions define the mapping of input data into feature space. Popular choices of kernel function include

radial basis function (RBF), linear kernel, and polynomial kernel function. The choice however, mostly remains an outcome of experimental results [8].
The RBF kernel is defined as:

$$\phi(x,\tilde{x}) = \exp\Big(-\frac{\| x-\tilde{x} \|}{2\sigma^2}\Big), \gamma = \frac{1}{2\sigma^2} \tag{6}$$

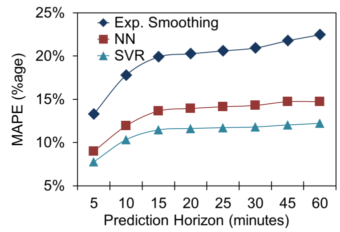
RBF is highly effective in mapping non-linear relationships. It can be shown that linear kernel behaves as special case of RBF kernel for certain parameter values [32]. Due to these attributes we used RBF kernel for SVR training. Efficiency of SVR strongly depends proper parameter selection (C,γ). Wrong selection can give rise to issues like data over fitting. To avoid these issues, we used k-fold cross validation (CV) technique using ε-SVR for parameter selection. CV however, adds additional computational cost to the system. ε-SVR offers flexibility of choosing value of ε, depending upon setup requirements. A loose error bound (ε) can reduce the CV cost substantially. For traffic prediction, loose error bound will provide inaccurate results. Hence, we used ν-SVR for speed prediction to mitigate the uncertainty associated with error bound (ε). We implemented the SVR by means of the LIBSVM software package [33].

D. Speed prediction using hybrid off-line training and on-line prediction mechanism

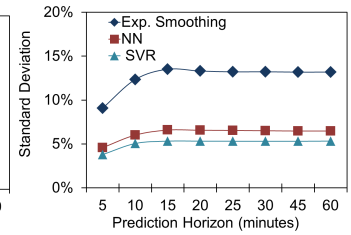
Support vector regression is computationally intense, and consequently, it does not scale well for large data sets. The online version of SVR resolves this problem by performing incremental online training. This method however, fails to provide high prediction accuracy [15]. To balance both constraints, we use the hybrid off-line training and on-line prediction based SVR model. Traffic speed behaves as a time series. This makes continuous training of SVR unnecessary. The same set of support vectors can be used for prediction for a significant portion of time (t_f). For experimental purposes we set t_f to 10 days. Each prediction horizon can be treated as an independent function estimation problem. We utilize this property by implementing parallel SVR architecture for each prediction horizon. Parallel SVR architecture can also be extended across multiple road segments. This would however, require optimization of hardware and software resources [34], [35]. We will not focus on resource optimization in this study. CV makes accuracy analysis more robust to any outliers, or bias in data. K-fold CV is less computationally intense. In many cases it provides similar performance evaluation results to more taxing leave one out cross validation (LOOCV) [36]. We performed $k = 6$ trials for each road segment. For each trial, we separated the data differently into 50 days of training data and 10 days of test data. Prediction error was calculated as the average error across the six trials.

III. PERFORMANCE COMPARISON WITH OTHER TIME SERIES PREDICTION METHODS

In this section we compare the prediction performance of proposed algorithm with performances of the artificial neural networks (ANN) and Holts exponential smoothing,



(a) Mean error



(b) Standard deviation of MAPE

Fig. 2: Performance comparison of different prediction methods

by calculating MAPE for each technique. For a link (s_i), MAPE for k^{th} prediction horizon $e_s(s_i,k)$ is defined as:

$$e_s(s_i,k) = \frac{1}{l}\sum_{j=1}^{l} \frac{|\,\mathbf{U}(t_j,s_i) - \tilde{\mathbf{U}}_k(t_j,s_i)\,|}{\mathbf{U}(t_j,s_i)}. \tag{7}$$

For the whole road network G, MAPE for k^{th} prediction horizon is calculated as $e(k) = E\{e_s(s_i,k)\}$. $E\{\bullet\}$ denotes the expectation value (approximated as mean value). The standard deviation for each prediction horizon (σ_k) is calculated as:

$$\sigma_k = (E\{(e_s(s_i,k) - e(k))^2\})^{\frac{1}{2}}. \tag{8}$$

1) Multilayer feed forward neural networks: Different architectures of ANN have been extensively used for short term traffic prediction [7], [10], [11], [13], [14], [17], [21], [22]. Multilayer feedforward (MLF) networks, possess highly desirable property of universal approximators [37]. This property makes MLF preferred ANN architecture for prediction of road network traffic parameters [15]. We performed ANN training using back-propagation algorithm.

2) Holt's exponential smoothing: Exponential Smoothing methods are more commonly known as Holt's exponential smoothing models. These models have been efficaciously applied for road network traffic parameter prediction [38]. We chose decay rates (λ_k), for each prediction horizon (k), using MAPE based CV [15]. Fig. 2 shows the results for the different prediction methods. SVR outperformed both ANN and exponential smoothing method for all prediction horizons. SVR based method has lower mean (Fig. 2a) and performance standard deviation (Fig. 2b). As expected, all methods perform better for smaller prediction horizons. The mean error and error variance increase with the prediction horizon. ANN performs only slightly worse compared to SVR.

IV. PERFORMANCE EVALUATION USING UNSUPERVISED LEARNING TECHNIQUES

Point estimation measures like MAPE fail to capture the spatial and temporal performance trends for a large network. Spatial trends can provide detailed insight into relative performance of road segments. Temporal trends can provide information related to variation in prediction performance due to changes in daily and hourly traffic patterns. We can also utilize temporal trends to analyze performance of prediction algorithm at micro level. Many ITS applications can benefit from such analysis. For a small data set such inferences are trivial. Large-scale prediction problems, however, require more sophisticated data mining and learning methods. In this section, we will develop novel unsupervised learning techniques to address above mentioned issues.

A. Analysis of performance across space

We cluster road segments based on their prediction performance, across different prediction horizons. We represent each road segment s_i by a vector $\mathbf{e}_{si} \in \mathbb{R}^n$ such that $\mathbf{e}_{si} = [e_s(s_i,t_0) \ \dots \ e_s(s_i,nt_0)]^{\mathbf{T}}$, where $e_s(s_i,k)$ is MAPE for kth prediction horizon for segment s_i. The distance (Δ) between any two road segments s_i and s_j, is defined as:

$$\Delta = \sqrt{(\mathbf{e_{si}} - \mathbf{e_{sj}})^{\mathbf{T}}(\mathbf{e_{si}} - \mathbf{e_{sj}})}. \tag{9}$$

For the given road network (G), optimal number of clusters are not explicitly known. That is a typical problem in clustering known as validation [39]. We use Silhouette index [40] as internal validation criteria to find the optimal number of clusters (θ). Silhouette value $\zeta(s_i)$ for a road segments s_i is calculated as:

$$\zeta(s_i) = \frac{\beta(s_i) - \alpha(s_i)}{max(\beta(s_i), \alpha(s_i))}, -1 \leq \zeta(s_i) \leq 1 \tag{10}$$

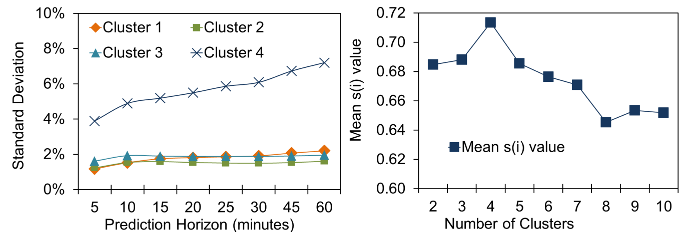
where $\alpha(s_i)$ is the mean distance of road segment(s_i) with all other road segments in the same cluster. $\beta(s_i)$ is distance of s_i with nearest neighboring cluster.

B. Analysis of performance across time

1) Analysis by PCA based clustering: The performance of predictors tends to depend on the time of the day and on the day of the week. Also large social events (e.g., New Year celebrations) may affect the traffic conditions, and result in larger prediction errors. Some links may have stable prediction errors in time. For other links, the prediction error may vary substantially over time. We wish to identify the consistent and relatively inconsistent links. To this end, we apply principal component analysis (PCA) to find such clusters of links. PCA is a well known method employed for statistical inference. Principal components provide an implicit visualization of inherent similar subsets within the data [41]. To evaluate daily variance for link s_i, we compute daily averages of the prediction error for different prediction horizons. Similarly, we computed hourly averages of the prediction error, across prediction horizons to evaluate the hourly variation. We apply PCA individually to the resulting two matrices. We represent each road segment (s_i) as a 2-tuple (δ_{di}, δ_{hi}), where δ_{di} and δ_{hi} are number of principal components explaining 80% of daily and hourly variance respectively. If the number of principal components is small (large), the prediction error varies little (substantially) across time. We apply k-means clustering to those 2-tuples in order

TABLE I: Performance for different clusters of segments

Prediction Horizon	Cluster 1	Cluster 2	Cluster 3	Cluster 4
5 minutes	2.69%	6.79%	11.06%	17.18%
10 minutes	3.24%	9.16%	14.60%	23.04%
15 minutes	3.66%	10.38%	15.90%	24.58%
20 minutes	3.86%	10.54%	16.01%	24.81%
25 minutes	4.05%	10.64%	16.05%	24.98%
30 minutes	4.20%	10.72%	16.08%	25.10%
45 minutes	4.61%	10.95%	16.18%	25.41%
60 minutes	4.89%	11.14%	16.28%	25.62%



(a) Intra cluster standard deviation (b) Mean silhouette value

Fig. 3: Spatial performance indices

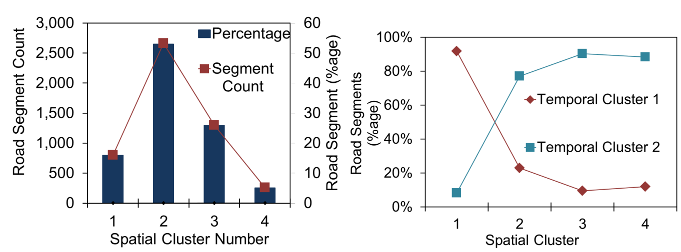
to group the segments. To this end, we define distance measure between two segments s_i and s_j as:

$$\Delta_{PCA} = \sqrt{(\delta_{di} - \delta_{dj})^2 + (\delta_{hi} - \delta_{hj})^2}. \tag{11}$$

2) Visualization of prediction error across time by SOM: Self-organizing maps (SOM) represent high-dimensional data on a low-dimensional manifold (typically two-dimensional). Different groups of segments appear as points ("centroids") on the SOM, where similar groups are located nearby on the map, and dissimilar groups are farther apart [42], [43]. For a given link (s_i), day to day performance may vary, across prediction horizons. We utilize SOM to map this higher dimensional daily performance data into two-dimensional cluster space. This would help identify days with similar performance for a given link.

V. RESULTS: PERFORMANCE EVALUATION ACROSS SPACE AND TIME

In this section, we analyze the performance of the variable window SVR algorithm by the tools proposed in Section IV. For spatial clustering, we found that θ = 4 clusters provided

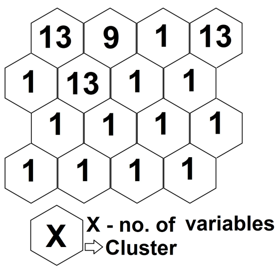


(a) Spatial clustering distribution (b) Temporal clustering distribution

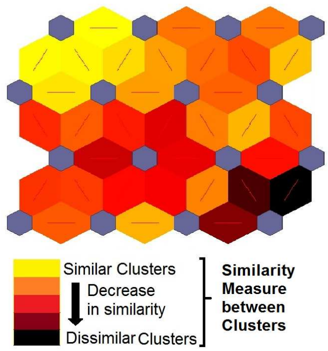
Fig. 4: Spatial and temporal clustering patterns

TABLE II: Centroid count of principal components for optimal temporal structure

Cluster	δ_d	δ_h	Number of segments
Cluster 1	5	6	1500
Cluster 2	13	12	3524



(a) Daily performance trends



(b) Inter cluster variation

Fig. 5: Daily performance trends for a single road segment

maximum mean silhouette value (Fig. 3b). The performance for the 4 different groups of segments is summarized in Table I. Spatial cluster 1 (SC1) contains road segments with smallest prediction error, mostly corresponding to highways. Most road segments (about 79%, Fig. 4a) belong either to spatial cluster 2 (SC2) or 3 (SC3), representing segments with low and intermediate prediction error. These clusters represent the mean upper and lower bounds for error performance of majority of road segments. Spatial Cluster 4 (SC4) contains the residual subset of road segments (5%, Fig. 4a), with largest prediction error. The intra cluster variance in SC1, SC2 and SC3 is relatively small, as shown in Fig 3a. The segments within each cluster have comparable prediction error, consistently for all prediction horizons. In contrast, SC4 has large variability in prediction error.The clustering method allowed us to separate the road segments in different classes, depending on prediction performance. For segments in SC4, alternative prediction procedures may be applied.

For temporal performance clustering, $\theta_{PCA} = 2$ provided maximum silhouette value. Parameters of the two temporal clusters are summarized in Table II. The first temporal cluster TC1 (TC2) contains links with little (much) temporal variation in the prediction error. Fig. 4b shows the percentage of segments belonging to those temporal clusters, separately for each spatial cluster. In each spatial cluster, links belong to either two temporal clusters. Temporal cluster 1 (TC1) contains ”highly consistent” links, where as cluster 2 (TC2) contains ”relatively inconsistent links”. Most of the links in SC1 are highly consistent. SC2 has a mix of both TC1 and TC2 (Fig. 4b). This implies that links with similar overall error performance (Fig. 3a) may possess quite different temporal characteristics. It is quite interesting that even worst performing spatial cluster has some links with “consistent”, temporal characteristics (Fig. 4b). Those links have large prediction errors at any point in time.

SOM structures for a typical single link with $\delta_{di} = 6$ are shown in Fig. 5. Fig. 5a shows variable count for each cluster, and Fig. 5b shows dissimilarity between clusters. Dark colored paths between hexagons in Fig. 5b, represent greater dissimilarity, and vice versa. We refer to j^{th} cluster in i^{th} row in Fig. 5 as ρ_{ij}. As evident from the graphs, SOM can provide detailed inference regarding temporal performance behavior. For this particular link, we observe four major groups (ρ_{14}, ρ_{11}, ρ_{12}, ρ_{22}) of daily performance variations (Fig. 5a). Cluster ρ_{14} mostly contains Saturdays and Sundays (weekends). Where as, the clusters ρ_{11}, ρ_{12} and ρ_{22} contain weekdays. Although there are three main clusters incorporating weekdays, they show similar prediction performance (Fig. 5b). However, there is a certain dissimilarity between prediction performance of ρ_{11}, ρ_{12}, ρ_{22} (weekdays) and ρ_{14} (weekends) (Fig. 5b). The rest of the clusters contain outliers (individual days with relatively dissimilar prediction performance). However, the overall daily performance variation for this particular road segment is quite small. This can be inferred from light colored paths connecting the major clusters and relatively low number of principal components ($\delta_{di} = 6$).

VI. CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a modular, robust, and scalable technique, termed as variable window ν-SVR method, to deal with the problem of speed prediction of a large and heterogenous road network. To establish prediction efficiency of the method, we compared its performance with other prediction methods including ANN and exponential smoothing. The comparison was performed for horizons of 5 min to 60 min. Performance comparison showed that variable window ν-SVR consistently provided better prediction accuracy than ANN and exponential smoothing.

Traditional methods like MAPE provide little or no information about the underlying spatial and temporal characteristics of prediction results. To assess the performance of our predictors across space and time, we applied unsupervised learning algorithms, including k-means clustering, PCA and SOM. We demonstrated the effectiveness of these methods by applying them to the prediction results of variable window ν-SVR method.

As a next step, one may couple our large-scale prediction algorithms with prospective applications like dynamic route guidance and congestion avoidance mechanisms. Moreover, unsupervised learning can also prove to be useful for assessing the performance of alternative traffic predictors, e.g., agent based traffic models.

VII. ACKNOWLEDGMENTS

The research described in this project was funded in part by the Singapore National Research Foundation (NRF) through the Singapore MIT Alliance for Research and Technology (SMART) Center for Future Mobility (FM).

References

[1] S. Mukherjee, E. Osuna, and F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines," in *Neural Networks for Signal Processing [1997] VII. Proceedings of the 1997 IEEE Workshop*. IEEE, 1997, pp. 511–520.

[2] K. Müller, A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik, "Predicting time series with support vector machines," *Artificial Neural NetworksICANN'97*, pp. 999–1004, 1997.

[3] R. Collobert and S. Bengio, "Svmtorch: Support vector machines for large-scale regression problems," *The Journal of Machine Learning Research*, vol. 1, pp. 143–160, 2001.

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

[5] P. Bermolen and D. Rossi, "Support vector regression for link load prediction," *Computer Networks*, vol. 53, no. 2, pp. 191–201, 2009.

[6] F. Yongyi, "Application of support vector machine regression method in weather forecast [j]," *Meteorological Monthly*, vol. 1, 2005.

[7] J. McFadden, W. Yang, and S. Durrans, "Application of artificial neural networks to predict speeds on two-lane rural highways," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1751, no. -1, pp. 9–17, 2001.

[8] C. Wu, J. Ho, and D. Lee, "Travel-time prediction with support vector regression," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 5, no. 4, pp. 276–281, 2004.

[9] C. de Fabritiis, R. Ragona, and G. Valenti, "Traffic estimation and prediction based on real time floating car data," in *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*. IEEE, 2008, pp. 197–203.

[10] C. van Hinsbergen, J. Van Lint, and H. Van Zuylen, "Bayesian committee of neural networks to predict travel times with confidence intervals," *Transportation Research Part C: Emerging Technologies*, vol. 17, no. 5, pp. 498–509, 2009.

[11] D. Park and L. Rilett, "Forecasting freeway link travel times with a multilayer feedforward neural network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 14, no. 5, pp. 357–367, 1999.

[12] L. Vanajakshi and L. Rilett, "Support vector machine technique for the short term prediction of travel time," in *Intelligent Vehicles Symposium, 2007 IEEE*. IEEE, 2007, pp. 600–605.

[13] E. Vlahogianni, M. Karlaftis, and J. Golias, "Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach," *Transportation Research Part C: Emerging Technologies*, vol. 13, no. 3, pp. 211–234, 2005.

[14] B. Smith and M. Demetsky, "Short-term traffic flow prediction models-a comparison of neural network and nonparametric regression approaches," in *Systems, Man, and Cybernetics, 1994.'Humans, Information and Technology'., 1994 IEEE International Conference on*, vol. 2. IEEE, 1994, pp. 1706–1709.

[15] M. Castro-Neto, Y. Jeong, M. Jeong, and L. Han, "Online-svr for short-term traffic flow prediction under typical and atypical traffic conditions," *Expert systems with applications*, vol. 36, no. 3, pp. 6164–6173, 2009.

[16] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.

[17] M. Dougherty and M. Cobbett, "Short-term inter-urban traffic forecasts using neural networks," *International journal of forecasting*, vol. 13, no. 1, pp. 21–31, 1997.

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

[19] Y. Zhao-sheng, W. Yuan, and G. Qing, "Short-term traffic flow prediction method based on svm [j]," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 6, 2006.

[20] L. Vanajakshi and L. Rilett, "A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed," in *Intelligent Vehicles Symposium, 2004 IEEE*. Ieee, 2004, pp. 194–199.

[21] J. Park, D. Li, Y. Murphey, J. Kristinsson, R. McGee, M. Kuang, and T. Phillips, "Real time vehicle speed prediction using a neural network traffic model," in *Neural Networks (IJCNN), The 2011 International Joint Conference on*. IEEE, 2011, pp. 2991–2996.

[22] C. Quek, M. Pasquier, and B. Lim, "Pop-traffic: A novel fuzzy neural approach to road traffic analysis and prediction," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 7, no. 2, pp. 133–146, 2006.

[23] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, 2010.

[24] B. Smith, B. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 10, no. 4, pp. 303–321, 2002.

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

[26] V. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Advances in Neural Information Processing Systems 9*, 1996, pp. 281–287.

[27] A. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[28] C. Chang and C. Lin, "Training v-support vector regression: theory and algorithms," *Neural Computation*, vol. 14, no. 8, pp. 1959–1977, 2002.

[29] B. Scholkopf, P. Bartlett, A. Smola, and R. Williamson, "Shrinking the tube: a new support vector regression algorithm," *Advances in neural information processing systems*, pp. 330–336, 1999.

[30] Z. Sun, Y. Wang, and J. Pan, "Short-term traffic flow forecasting based on clustering and feature selection," in *IEEE International Joint Conference on Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 577–583.

[31] X. Jin, Y. Zhang, and D. Yao, "Simultaneously prediction of network traffic flow based on pca-svr," *Advances in Neural Networks–ISNN 2007*, pp. 1022–1031, 2007.

[32] S. Keerthi and C. Lin, "Asymptotic behaviors of support vector machines with gaussian kernel," *Neural computation*, vol. 15, no. 7, pp. 1667–1689, 2003.

[33] C. Chang and C. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[34] D. Culler, J. Singh, and A. Gupta, *Parallel computer architecture: a hardware/software approach*. Morgan Kaufmann, 1999.

[35] P. Boncz, S. Manegold, and M. Kersten, "Database architecture optimized for the new bottleneck: Memory access," in *PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES*, 1999, pp. 54–65.

[36] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *International joint Conference on artificial intelligence*, vol. 14. LAWRENCE ERLBAUM ASSOCIATES LTD, 1995, pp. 1137–1145.

[37] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural networks*, vol. 2, no. 5, pp. 359–366, 1989.

[38] B. Williams, P. Durvasula, and D. Brown, "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1644, no. -1, pp. 132–141, 1998.

[39] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, vol. 17, no. 2, pp. 107–145, 2001.

[40] P. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.

[41] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning: Data mining, inference, and prediction," *BeiJing: Publishing House of Electronics Industry*, 2004.

[42] G. Pölzlbauer, "Survey and comparison of quality measures for self-organizing maps," in *Proceedings of the Fifth Workshop on Data Analysis (WDA04)*, 2004, pp. 67–82.

[43] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.