Bayesian Support Vector Regression for Traffic Speed Prediction with Error Bars

Gaurav Gopi, Justin Dauwels, Muhammad Tayyab Asif, Sridhar Ashwin,
Nikola Mitrovic, Umer Rasheed, Patrick Jaillet

Abstract— **Traffic prediction algorithms can help improve the performance of Intelligent Transportation Systems (ITS). To this end, ITS require algorithms with high prediction accuracy. For more robust performance, the traffic systems also require a measure of uncertainty associated with prediction data. Data driven algorithms such as Support Vector Regression (SVR) perform traffic prediction with overall high accuracy. However, they do not provide any information about the associated uncertainty. The prediction error can only be calculated once field data becomes available. Consequently, the applications which use prediction data, remain vulnerable to variations in prediction error. To overcome this issue, we propose Bayesian Support Vector Regression (BSVR). BSVR provides error bars along with the predicted traffic states. We perform sensitivity and specificity analysis to evaluate the efficiency of BSVR in anticipating variations in prediction error. We perform multi-horizon prediction and analyze the performance of BSVR for expressways as well as general road segments.**

I. INTRODUCTION

Traffic management systems require algorithms that can deal with varying traffic conditions [1]–[3]. Data driven prediction algorithms such as Support Vector Regression (SVR) tend to provide on average high-prediction accuracy for road networks [4]–[6]. This makes data driven methods useful for many Intelligent Transportation Systems (ITS) applications. These methods help ITS to deal with variations in traffic conditions. However, the predicted values are also prone to uncertainty. Prediction accuracy may vary for different test conditions. Actual prediction error can only be calculated once field data becomes available. For instance, consider 15 minutes prediction horizon. In this case, we can only calculate prediction error after 15 minutes. Large variations in prediction error can severely degrade the performance of applications which utilize prediction data. Algorithms that can anticipate such variations in prediction accuracy will be highly desirable. These algorithms will allow ITS applications such as route guidance to take into account the uncertainty associated with predicted network state.

We propose Bayesian Support Vector Regression (BSVR) to provide error bars [7], [8] for the predicted traffic states.

This technique combines advantages of SVR and Bayesian inference [8]. SVR is highly suitable for traffic prediction as it can map non-linear relationships between past and future traffic conditions for SVR [4], [9]. Bayesian methods for SVR are useful for selecting suitable hyperparameters [10]. They can also estimate variance (termed as error bar, [7], [8]) associated with each prediction. The estimated variance can be considered as a confidence measure for the corresponding predicted traffic state. This additional information is helpful for many ITS applications. One proposed approach to estimate variance in prediction error is to combine prediction from different neural networks [11], [12]. This approach has two drawbacks. Neural network training algorithms tend to suffer from the problem of local minima [13]. Secondly, we are required to train multiple models [11]. This increases the computational complexity of such algorithms. BSVR overcomes both these issues [8].

In this study, we analyze the efficiency of BSVR in forecasting the behavior of prediction error for multiple prediction horizons. For this purpose, we perform traffic speed prediction on two different sub-networks. Singapore Land Transportation Authority (LTA) provided speed data for the study. The confidence levels provided by BSVR will help us to identify predicted states with high uncertainty. We expect to incur high prediction error for these states. We consider this as a detection problem. To be useful, BSVR should anticipate whether or not the predicted state will have large error. Therefore, we perform sensitivity and specificity analysis on prediction data of BSVR. We analyze prediction data for different scenarios and prediction horizons. The analysis shows that BSVR can anticipate large prediction errors with reliable sensitivity and low false alarm rate.

The paper is structured as follows. In section II, we briefly explain Bayesian Support Vector Regression. In section III, we explain the data set. In section IV, we explain the setup for sensitivity and specificity analysis. In section V, we analyze the detection performance of BSVR. In section VI, we summarize our contributions and suggest topics for future work.

II. BAYESIAN SUPPORT VECTOR REGRESSION FOR TRAFFIC PREDICTION

In this section, we briefly explain the principle of Bayesian Support Vector Regression for traffic prediction. We represent the test network by a directed graph $G = (N, E)$. The links/road segments $\{s_i\}_{i=1}^p$ represent the edges of the graph such that $\{s_i \in E\}_{i=1}^p$. We represent the weight of each

link by the average speed $z(s_i, t_j)$ of the link during the time interval $(t_j - t_0, t_j)$. The sampling interval is $t_0 = 5$ minutes.

Future traffic states of a road segment may depend upon certain past states of the road. These states are called input features. Let $\mathbf{x}_j$ be a vector containing input features at time t_j such that $\mathbf{x}_j = [z(s_i, t_j) \ldots z(s_i, t_j - mt_0)]^T$. Furthermore, $y_{jk} = z(s_i, t_j + kt_0)$ is the k step ahead speed value. Let us consider the training data set $D = \{(\mathbf{x}_j, y_{jk})\}_{j=1}^d$ for link s_i and k^{th} prediction horizon. We start by explaining traditional SVR. In SVR training, our goal is to find the optimal hyperplane $\mathbf{w}$. In case of SVR, the relationship function takes the following form:

$$f_k(\mathbf{x}_j) = \mathbf{w}^T \phi(\mathbf{x}_j) + b, \tag{1}$$

where $\phi(\mathbf{x}_j)$ represents the non-linear mapping of input features $\mathbf{x}_j$ into some high dimensional space [7], [14], [15]. To train SVR, we solve the following optimization problem [15]:

$$\min \; \frac{1}{2}\mathbf{w}^T\mathbf{w} + C \sum_{j=1}^r (\xi_j + \xi_j^*),$$
$$\text{subject to } \begin{cases} y_{jk} - f_k(\mathbf{x}_j) \leq \varepsilon + \xi_j \\ f_k(\mathbf{x}_j) - y_{jk} \leq \varepsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0, \end{cases} \tag{2}$$

where ξ_j, ξ_j^* are slack variables and ε is the parameter for Insensitive Loss Function (ILF) with cost C [15]. Now, let us extend the SVR formulation to bayesian framework called BSVR. For BSVR, we consider following regression model:

$$y_{jk} = f_k(\mathbf{x}_j) + \delta_j \quad \mathbf{x}_j \in \mathbb{R}^n, \; y_{jk} \in \mathbb{R}, \tag{3}$$

where f_k is the relationship function for k^{th} prediction horizon with i.i.d noise samples $\{\delta_j\}_{j=1}^d$ [7], [8]. The extension of (2) to Bayesian framework will help us to find optimal hyperparameters for SVR. More importantly, we will obtain error bars corresponding to the predicted traffic states. For this purpose, let us consider $\boldsymbol{f}_k = [f_k(\mathbf{x}_1) \ldots f_k(\mathbf{x}_d)]^T$. It contains the values of relationship functions $\{f_k(\mathbf{x}_j)\}_{j=1}^d$ for the training data set D. In BSVR, we consider $\boldsymbol{f}_k$ as a random vector with prior probability $P(\boldsymbol{f}_k)$. The probability of obtaining $\boldsymbol{f}_k$, for a given speed data set D is:

$$P(\boldsymbol{f}_k | D) = \frac{P(D | \boldsymbol{f}_k) P(\boldsymbol{f}_k)}{P(D)}, \tag{4}$$

where $P(D | \boldsymbol{f}_k)$ is the likelihood of speed data D, given function $\boldsymbol{f}_k$ and is calculated as $\prod_{j=1}^d P(y_{jk} - f_k(\mathbf{x}_j))$ or $\prod_{j=1}^d P(\delta_j)$. It is usually assumed that $P(\delta_j)$ is of exponential form [7], [8], so:

$$P(D | \boldsymbol{f}_k) \propto \exp - \left(C \sum_{j=1}^d L(y_{jk} - f_k(\mathbf{x}_j)) \right), \tag{5}$$

where $L(\cdot)$ is called the loss function and C is the associated cost. In BSVR, we further assume that prior probability $P(\boldsymbol{f}_k)$ follows a multivariate Gaussian distribution such that $\boldsymbol{f}_k \sim N(0, \Sigma)$ [7]. Hence, The Maximum a Posteriori (MAP) estimate for the relationship function can be obtained as [7], [8]:

$$\min_{\boldsymbol{f}_k} \quad S(\boldsymbol{f}_k) = C \sum_{j=1}^d L(y_{jk} - f_k(\mathbf{x}_j)) + \frac{1}{2}\boldsymbol{f}_k^T \Sigma^{-1} \boldsymbol{f}_k. \tag{6}$$

For BSVR, we choose Soft Insensitive Loss Function (SILF) [8]. Hence, by incorporating Bayesian framework to SVR, we obtain the following formulation [8]:

$$\min \; C \sum_{j=1}^d \left(\Psi(\xi_j) + \Psi(\xi_j^*) \right) + \frac{1}{2}\boldsymbol{f}_k^T \Sigma^{-1} \boldsymbol{f}_k,$$
$$\text{subject to } \begin{cases} y_{jk} - f_k(\mathbf{x}_j) \leq (1-\beta)\varepsilon + \xi_j \\ f_k(\mathbf{x}_j) - y_{jk} \leq (1-\beta)\varepsilon + \xi_j^* \\ \xi_j, \xi_j^* \geq 0, \end{cases} \tag{7}$$

where β and ε are the parameters for SILF. The function $\Psi(\cdot)$ is defined as [8]:

$$\Psi(a) = \begin{cases} \dfrac{a^2}{4\beta\varepsilon} & a \in [0, 2\beta\varepsilon) \\ a - \beta\varepsilon & a \in [2\beta\varepsilon, \infty). \end{cases} \tag{8}$$

The extension of SVR in (7) is termed as BSVR. For implementation, we follow the procedure provided in [8]. For error bar (confidence level) estimation, consider an input feature vector $\mathbf{x}_t$ for k^{th} prediction horizon. Suppose the speed predicted by BSVR is $\hat{y}_{tk}$. The uncertainty in predicted speed can arise from two factors (see (3)). It can be due to either $P(f_k(\mathbf{x}_t) | D)$ or noise δ_t. The variances due to $P(f_k(\mathbf{x}_t) | D)$ and δ_t are σ_t^2 and σ_n^2 respectively. Consequently, the variance estimated by BSVR is $\sigma_t^2 + \sigma_n^2$. We refer to $\sqrt{\sigma_t^2 + \sigma_n^2}$ as error bar. The variance σ_n^2 only depends upon the choice of loss function [7], [8]. The value of σ_t^2 depends upon the input feature vector $\mathbf{x}_t$ and off bound support vectors [7], [8], [16].

We perform speed prediction for multiple horizons by applying BSVR. In this study, we consider horizons from 5 minutes up till 30 minutes. We train separate predictors for each link s_i and prediction horizon k. For analysis, we choose two different road networks. We explain the data set in the next section.

III. DATA SET

In this section, we explain the data set for performance analysis. For this purpose, we choose two different road networks (see Fig. 1). The network $G_1 = (N_1, E_1)$ consists of road segments from Pan Island Expressway (PIE) in Singapore. We perform multi-horizon prediction on 20 links in the network. The network $G_2 = (N_2, E_2)$ consists of 20 segments from arterial roads in the vicinity of Lavender Mass Rapid Transit (MRT) and Boon Keng MRT stations.

For this study, the speed data was provided by Singapore Land Transportation Authority (LTA). The data set has averaging intervals of 5 minutes. We consider data from the months of March and April, 2011. We perform BSVR training on speed data of March and evaluate performance on the data from April.

TABLE I: Prediction Performance for Pan Island Expressway (G_1)

Error Measure	Prediction Horizon					
	5 min	10 min	15 min	20 min	25 min	30 min
MAPE	4.26%	5.28%	6.07%	6.68%	7.19%	7.50%
MAE (km/hr)	2.83	3.36	3.76	4.05	4.28	4.46

TABLE II: Prediction Performance for Lavender Area (G_2)

Error Measure	Prediction Horizon					
	5 min	10 min	15 min	20 min	25 min	30 min
MAPE	6.94%	9.51%	11.08%	11.32%	11.45%	11%
MAE (km/hr)	2.05	2.80	3.25	3.31	3.34	3.21

Fig. 1: Regions for performance Analysis.

Fig. 2: Absolute Error (normalized) with error bars for a given road segment. The prediction horizon is 5 minutes. Increase in error bar value, in most cases, corresponds to large prediction error.
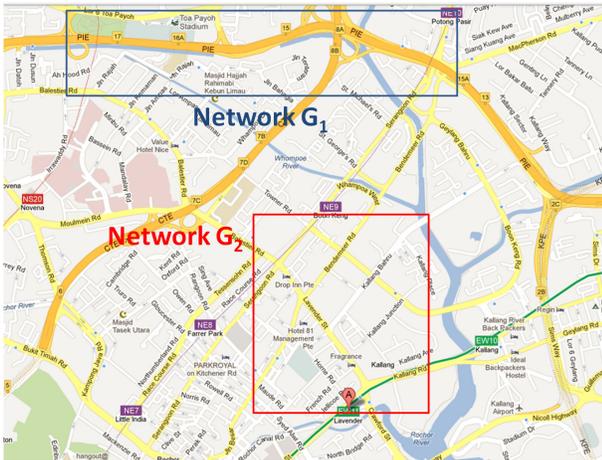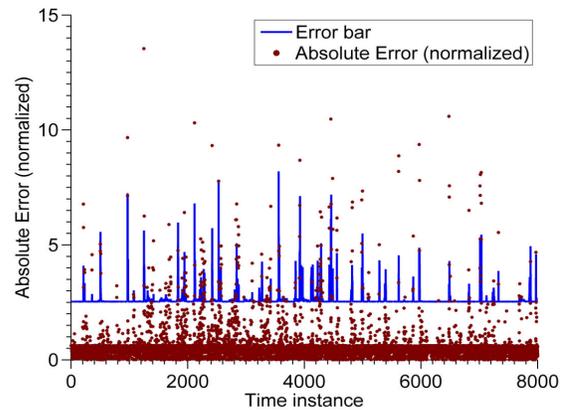
We calculate Mean Absolute Percentage Error (MAPE) to evaluate prediction performance of BSVR. We define MAPE $e(s_i, k)$ for link s_i and k^{th} prediction horizon as:

$$e(s_i, k) = \frac{1}{l} \sum_{j=1}^{l} \frac{|\hat{z}(s_i, t_j) - z(s_i, t_j)|}{z(s_i, t_j)}, \tag{9}$$

where l is the size of the test data set. The speed values $\hat{z}(s_i, t_j)$ and $z(s_i, t_j)$ represent the predicted and actual speed values for the interval $(t_j - t_0, t_j)$ respectively. We calculate MAPE for the whole network $\{e(G_i, k)\}_{i \in \{1,2\}}$ as:

$$e(G_i, k) = \frac{1}{p} \sum_{j=1}^{p} e(s_j, k), \tag{10}$$

where the network G_i contains p road segments. We also calculate Mean Absolute Error (MAE) for the two road networks [8]. MAE for link s_i and prediction horizon k is calculated as:

$$d(s_i, k) = \frac{1}{l} \sum_{j=1}^{l} |\hat{z}(s_i, t_j) - z(s_i, t_j)|. \tag{11}$$

We calculate MAE for the whole network G_i as:

$$d(G_i, k) = \frac{1}{p} \sum_{j=1}^{p} d(s_j, k). \tag{12}$$

The main advantage of BSVR is that it can provide confidence measures (error bars) for the predicted traffic states. With error bars, we can anticipate variations in prediction error (see Fig. 2). Traffic management systems require information about discrete states of prediction error (such as high/low). The state itself depends upon the threshold of prediction error, which a system can tolerate. For this purpose, we treat BSVR as a detection algorithm. To evaluate the detection performance of BSVR, we apply standard sensitivity and specificity analysis [17], [18]. We explain the setup for sensitivity and specificity analysis in the next section.

IV. SENSITIVITY AND SPECIFICITY ANALYSIS FOR BSVR

In this section, we explain the setup to evaluate detection performance of BSVR. For this purpose, we choose different tolerance values τ_d of MAE. We categorize those prediction errors which are higher than the tolerance limit τ_d as positive events. Positive events are associated with large prediction errors (MAE). During negative events, prediction error will remain below the tolerance limit. The role of BSVR is to anticipate such events. We represent the detector threshold by γ_d. If the value of error bar is larger than γ_d, then we expect to observe large prediction error and vice versa. For instance, suppose we perform prediction for $t_j + kt_0$ at t_j. With error bars, we should be able to tell at time t_j whether
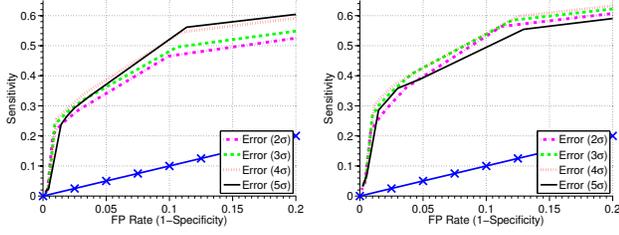
(a) Prediction Horizon: 5 minutes. (b) Prediction Horizon: 10 minutes.
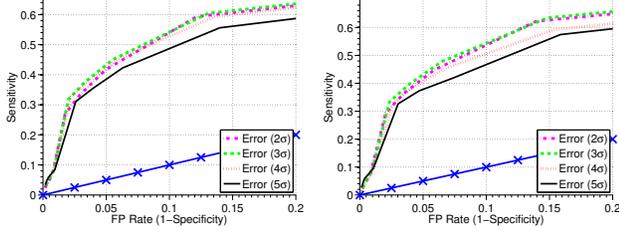
(c) Prediction Horizon: 15 minutes. (d) Prediction Horizon: 20 minutes.
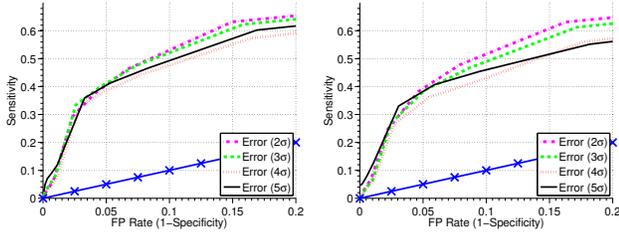
(e) Prediction Horizon: 25 minutes. (f) Prediction Horizon: 30 minutes.

Fig. 3: ROC plots for Pan Island Expressway (PIE) with different tolerance limits $\tau_d \in \{2\sigma,...,5\sigma\}$, where σ is the standard deviation of prediction error. In this study, we keep the False Positive (FP) rate below 20%. The blue line (shown as -x-) represents the no-discrimination line.

(a) Prediction Horizon: 5 minutes. (b) Prediction Horizon: 10 minutes.

(c) Prediction Horizon: 15 minutes. (d) Prediction Horizon: 20 minutes.

(e) Prediction Horizon: 25 minutes. (f) Prediction Horizon: 30 minutes.

Fig. 4: ROC plots for Lavender Area with different tolerance limits $\tau_d \in \{2\sigma,...,5\sigma\}$, where σ is the standard deviation of prediction error. In this study, we keep the False Positive (FP) rate below 20%. The blue line (shown as -x-) represents the no-discrimination line.

the prediction error at $t_j + kt_0$ will remain within the tolerance limit or not. We expect four possible outcomes, which are True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). BSVR should be able to identify time instances when we observe large prediction errors (True Positives). We define the sensitivity of the detector as:

$$\text{Sensitivity} = \frac{\text{number of true positives (TP)}}{\text{number of positive events (TP + FN)}}. \quad (13)$$
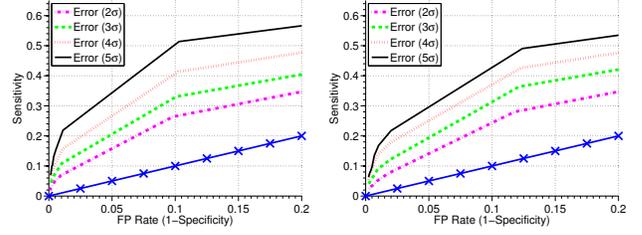
It should also accurately identify time instances, when prediction error is small (True Negatives). The specificity of the detector is defined as:

$$\text{Specificity} = \frac{\text{number of true negatives (TN)}}{\text{number of negative events (TN + FP)}}. \quad (14)$$
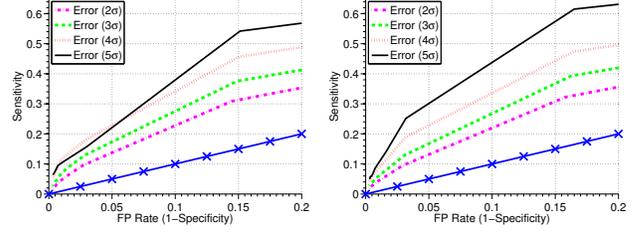
Consequently, the proportion of false positives (FP) is calculated as:

$$\frac{\text{FP}}{\text{TN + FP}} = 1 - \text{Specificity}. \quad (15)$$

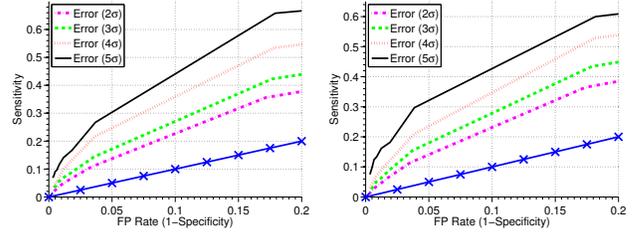Our primary concern is to keep false positive rate low ($< 20\%$). We analyze whether we can achieve high sensitivity with this constraint.

We apply these criteria to assess the detection efficiency of BSVR. In the next section, we perform multi-horizon prediction on the test networks and analyze the results.

V. RESULTS AND DISCUSSION

In this section, we analyze the performance of BSVR. We refer to network G_1 as Pan Island Expressway (PIE) and network G_2 as Lavender area.

Table I and II summarize the prediction performance of the two networks. As expected, expressways (PIE) have smaller MAPE as compared to arterial roads (Lavender area). However, MAE values seem to contradict this trend (see Table I and II). This is due to different speed profiles of the two networks. On expressways, we observe high speed traffic and consequently get large MAE. For MAPE, we normalize prediction errors with observed speed and arguably obtain a better comparison measure.

Let us now analyze the detection performance of BSVR for the two networks. Detection performance is typically analyzed by keeping the false positive rate below a certain limit [18]. In this study, we keep the false positive rate below 20%. We only consider those sensitivity values for which

TABLE III: Sensitivity and Specificity of BSVR for a fixed threshold of the detector for PIE. In this case, if the error bar value is above its average then the corresponding instance will be detected as positive event by BSVR.

Prediction Error	5 min		10 min		15 min		20 min		25 min		30 min	
	Sen.	Spec.	Sen.	Spec.	Sens.	Spec.	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.
Error tolerance (1σ)	44%	91%	53%	90%	56%	89%	58%	88%	59%	87%	60%	85%
Error tolerance (2σ)	47%	90%	56%	89%	59%	88%	62%	86%	63%	85%	63%	84%
Error tolerance (3σ)	50%	89%	58%	88%	60%	87%	63%	85%	62%	84%	61%	83%
Error tolerance (4σ)	55%	89%	60%	87%	59%	86%	59%	85%	57%	84%	56%	82%
Error tolerance (5σ)	56%	89%	55%	87%	56%	86%	57%	84%	60%	83%	55%	82%

TABLE IV: Sensitivity and Specificity of BSVR for a fixed threshold of the detector for Lavender Area. In this case, if the error bar value is above its average then the corresponding instance will be detected as positive event by BSVR.

Prediction Error	5 min		10 min		15 min		20 min		25 min		30 min	
	Sen.	Spec.	Sen.	Spec.	Sens.	Spec.	Sen.	Spec.	Sen.	Spec.	Sen.	Spec.
Error tolerance (1σ)	21%	91%	21%	89%	24%	86%	25%	85%	28%	84%	29%	84%
Error tolerance (2σ)	26%	90%	28%	88%	31%	86%	32%	84%	36%	83%	36%	83%
Error tolerance (3σ)	33%	90%	36%	88%	38%	85%	39%	84%	42%	82%	43%	82%
Error tolerance (4σ)	41%	90%	43%	88%	46%	85%	47%	84%	53%	82%	53%	82%
Error tolerance (5σ)	51%	90%	49%	88%	54%	85%	62%	83%	66%	82%	60%	82%

we obtain specificity of 80% or more (false positive rate < 20%). We plot each Receiver Operating Characteristic (ROC) curve by varying detector threshold γ_d. This provides us with different sensitivity and specificity values for each error tolerance level.

Fig. 3 shows the ROC for different prediction horizons for expressway links (PIE). In Fig. 4, we plot ROC curves for the road segments in the Lavender area. The ROC curves in these figures represent the average detection performance of BSVR across a particular network. We plot each curve by setting a certain tolerance limit for prediction error. We plot ROC curves for different tolerance levels $\tau_d \in \{2\sigma, ... , 5\sigma\}$, where σ is the standard deviation of prediction error. For analysis, we keep the False Positive (FP) rate below 20%. The blue lines in Fig. 3 and 4 are called the no-discrimination lines. These lines represent the performance of a detector that randomly selects an event as either positive or negative. The ROC curve should remain above the no-discrimination line for a detector to be useful. ROC curves clearly remain above the no-discrimination line for both networks (see Fig. 3 and 4). ROC curves for expressway links remain far above the no-discrimination line for high as well as low error tolerance levels (see Fig. 3). In case of general roads, ROC curves tend to move closer to the no-discrimination line for low error tolerance levels.

For a given false positive rate, we achieve higher sensitivity for expressways as compared to general roads (lavender area). In case of expressways (PIE), BSVR can detect around 60% instances of large error for all the prediction horizons. For this level of sensitivity, it only reports false alarms in around 15% of time instances (specificity $\sim$ 85%). More importantly, it provides similar detection performance for small as well as large tolerance levels (different error sigma levels, see Fig. 3).

For general roads, we observe slightly degraded detection performance. In the Lavender area, average sensitivity varies from 30% to around 60% for different error tolerance levels and prediction horizons. We achieve these sensitivity

levels with specificity of around 85% (see Fig. 4). Still, we can detect large errors with relatively high sensitivity and specificity. For tighter error tolerance (prediction error $\leq 2\sigma$), we observe degraded sensitivity (see Fig. 4).

Let us now perform specificity and sensitivity analysis for a fixed threshold γ_d of BSVR error bars. In this analysis, we expect large prediction error, if the corresponding error bar value is above its average. We select positive events by choosing different tolerance levels for prediction error. We treat these events as ground truth. Table III shows the sensitivity and specificity values across multiple prediction horizons for expressway links. We find that the detector sensitivity remains around 60% for expressways across different prediction horizons. The detector specificity degrades from 90% for small prediction horizons to around 82% for large prediction horizons. The detector performance does not vary much for different error tolerance levels for expressway links (PIE, see Table III). False positives rate remains below 20% for all tolerance levels and prediction horizons. We summarize the results from Lavender area in Table IV. Detector specificity remains above 82% for all tolerance levels and prediction horizons. We observe low detection rate (sensitivity) for small error tolerance levels ($\leq 2\sigma$). Sensitivity of BSVR improves considerably for large error tolerance levels (from 25% to 60%, see Table IV). We also find that detector sensitivity improves slightly for large prediction horizons. This improvement however comes with higher false alarm rate (see Table IV). False alarm rate increases from around 10% for 5 min prediction horizons to 18% for 30 min prediction horizon. We observe this trend for small as well as large prediction errors ($1\sigma, ... , 5\sigma$, see Table IV).

BSVR provides confidence measure for each prediction in terms of error bars. For certain traffic conditions (input features), we may not have enough historical data. In such cases, BSVR will anticipate large prediction error for that time instances. However, for similar input patterns, we might still observe different future traffic trends. Such instances
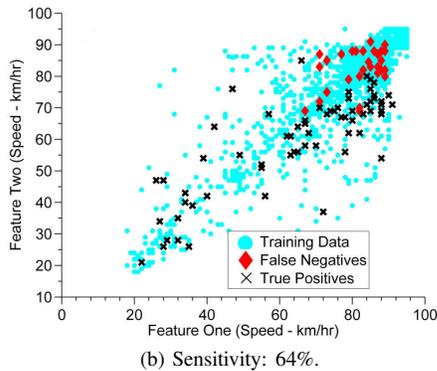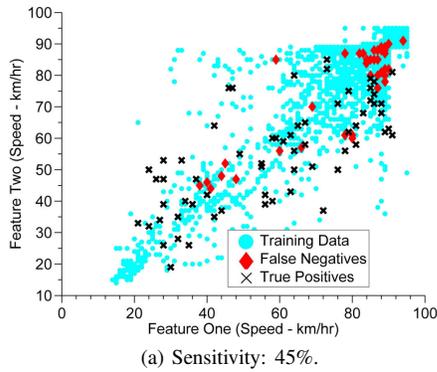
(a) Sensitivity: 45%.

(b) Sensitivity: 64%.

Fig. 5: Location of input features for test data corresponding to positive events for two road segments (from PIE) with different sensitivity values. In both cases, false negative events are embedded in the region which has high density of training data points. Consequently, the detector did not expect large prediction error for such instances.

might occur due to changes in on ground conditions. BSVR will not be able to anticipate behavior of prediction error during these time intervals. As an illustration, Fig. 5 shows feature plots for two different road segments. In these figures, we plot input features (current speed and speed observed 5 min earlier) of training data points and positive events (high prediction error) for 5 minute prediction horizon. Fig. 5a shows the plot for a road segment with sensitivity of 45%. In Fig. 5b, we plot the input features of a link with sensitivity of 64%. In both cases, input features of true positive events are located away from training data points. On the other hand, false negative events have similar input features as many other training data points.

We can conclude that BSVR can detect variations in prediction error with good sensitivity ($\sim 60\%$, see Fig. 3, 4) for expressways as well as general roads. We achieve these values at high level of specificity (85% to 90%, see Fig. 3, 4) for both networks. The detection performance remains consistent across small and large prediction horizons.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed Bayesian Support Vector Regression (BSVR) to provide error bars associated with the prediction errors. Uncertainty measures for prediction data can be useful for many ITS applications. We apply BSVR to provide error bars alongside predicted traffic states. These error bars can be considered as confidence measures for the corresponding predicted traffic states. With these confidence measures, we can anticipate the behavior of prediction error before field measurements become available. We analyzed the detection efficiency of BSVR by performing sensitivity and specificity analysis. To this end, we performed speed prediction on expressways and arterial roads in downtown area in Singapore for multiple prediction horizons. We found that BSVR can detect variations in prediction error with low false alarm rate and reasonable detection accuracy for expressways as well as general road segments.

In the future, traffic management systems can incorporate such measures to achieve more robust performance.

REFERENCES

[1] C. Coffey, A. Pozdnoukhov, and F. Calabrese, "Time of arrival predictability horizons for public bus routes," in *Proceedings of the 4th ACM SIGSPATIAL International Workshop on Computational Transportation Science*. ACM, 2011, pp. 1–5.

[2] J. Zhang, F. Wang, K. Wang, W. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1624–1639, 2011.

[3] J. Wang, K. Wong, and Y. Chen, "Short-term travel time estimation and prediction for long freeway corridor using nn and regression," in *2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2012, pp. 582–587.

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

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

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

[7] J. B. Gao, S. R. Gunn, C. J. Harris, and M. R. Brown, "A probabilistic framework for svm regression and error bar estimation," *Machine Learning*, vol. 46, no. 1-3, pp. 71–89, 2002.

[8] W. Chu, S. S. Keerthi, and C. J. Ong, "Bayesian support vector regression using a unified loss function," *IEEE Transactions on Neural Networks*, vol. 15, no. 1, pp. 29–44, 2004.

[9] M. Seeger, "Bayesian model selection for support vector machines, gaussian processes and other kernel classifiers," *Advances in neural information processing systems*, vol. 12, pp. 603–609, 2000.

[10] P. Sollich, "Bayesian methods for support vector machines: Evidence and predictive class probabilities," *Machine learning*, vol. 46, no. 1-3, pp. 21–52, 2002.

[11] 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.

[12] L. Fu and L. Rilett, "Estimation of time-dependent, stochastic route travel times using artificial neural networks," *Transportation planning and technology*, vol. 24, no. 1, pp. 25–48, 2000.

[13] M. Gori and A. Tesi, "On the problem of local minima in backpropagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 1, pp. 76–86, 1992.

[14] C.-J. Lin, R. C. Weng *et al.*, "Simple probabilistic predictions for support vector regression," *National Taiwan University, Taipei*, 2004.

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

[16] D. J. MacKay, "A practical bayesian framework for backpropagation networks," *Neural computation*, vol. 4, no. 3, pp. 448–472, 1992.

[17] D. Heeger, "Signal detection theory," *Dept. Psych., Stanford Univ., Stanford, CA, Teaching Handout*, 1997.

[18] S. M. Kay, *Fundamentals of Statistical signal processing, Volume 2: Detection theory*. Prentice Hall PTR, 1998.