Bayesian Prediction of the Duration of Non-recurring Road Incidents

Banishree Ghosh
Nanyang Technological University
Singapore 639798
Email: banishre001@e.ntu.edu.sg

Muhammad Tayyab Asif [1]
Nanyang Technological University
Singapore 639798
Email: muhammad89@e.ntu.edu.sg

Justin Dauwels
Nanyang Technological University
Singapore 639798
Email: jdauwels@ntu.edu.sg

***Abstract*—Traffic incidents such as accidents or vehicle breakdowns are one of the major causes of traffic congestion in urban areas. Consequently, accurate prediction of duration of these incidents is considered as one of the most important challenges by traffic management authorities. Although data-driven regression methods can predict the duration of these incidents with reasonable precision. However, the prediction performance may vary considerably from one to another. Hence, it is important to provide some measure of confidence associated with the forecast duration of the incidents. Such measures can prove to be highly useful in planning real-time response. To address this issue, we propose Bayesian Support Vector Regression (BSVR), which gives error bars as the measurement of uncertainty along with the predicted duration of incidents. We also evaluate sensitivity and specificity for different error tolerance limit to assess the performance of BSVR.**

I. Introduction

Disruptive events such as accidents and vehicle breakdowns often lead to reduction in road capacity and hence, disruption in normal traffic flow. Accurate prediction of duration of these incidents is critical for advanced traffic management systems. Therefore, this topic has accumulated considerable attention in the field of transportation [1]. Data-driven regression models such as Artificial Neural Networks [2] and Decision Trees [3] have been used of late to obtain the relationship between the external factors such as time of day, affected lanes, weather condition and the incident duration. However, the predicted values obtained by these methods are subject to uncertainty because the prediction performance varies with different test conditions. We propose to solve this issue by applying a method, that can anticipate the uncertainty associated with prediction error. The Bayesian Support Vector Regression (BSVR) technique combines both Support Vector Regression [4] and Bayesian inference [5] and therefore, it can estimate variance of the prediction errors (denoted as error bars) with the predicted values [6].

The time duration associated with a traffic incident can be divided into four components: (1) reporting time (r_t): the time when the incident has been reported, (2) response time (s_t): the time between reporting of the incident and arrival of the response team, (3) clearance time (c_t): the time required by the response team to clear the road, and (4) recovery time (v_t): the time taken by the traffic condition to restore back to normal [1][7]. The incident duration T that we consider here is the sum of all of these stages:

$$T = r_t + s_t + c_t + v_t. \tag{1}$$

Next we will briefly discuss related research works in this area. Van *et al.* applied neural network in Bayesian framework to predict travel times with confidence intervals [8]. However, neural network algorithms tend to converge on local minima rather than global minima, whereas Support Vector Regression method does not suffer from this drawback. Therefore, we choose BSVR over Bayesian Neural Network for estimation of error bar associated with prediction. Wu et al. considered 1853 incidents for a five-month interval (May–Sept, 2015) from Utrecht, a central city in the Netherlands and applied support vector regression technique to predict the incidents duration [4]. However, their approach does not provide any information about the uncertainty corresponding to the predicted duration. On the other hand, Ahn *et al.* predicted traffic flow in highways of Korea using BSVR approach [9]. However, they did not predict duration of non-recurrent incidents in their work, which is our area of concern.

The remainder of this paper is organized as follows. In the next section, we describe our dataset. In Section III, we briefly discuss the basic principle of Bayesian SVR, whereas we analyze the prediction performance of BSVR in Section IV. We determine the error bars associated with the predicted values of duration and perform sensitivity-specificity analysis in Section V. Finally, Section VI provides concluding remarks.

II. Description of the Data

The dataset used in this study is comprised of (1) historical records of incidents provided by the Land Transport Authority (LTA) of Singapore, and (2) weather information from the National Environmental Agency (NEA) of Singapore.

The historical records of traffic incidents contain the following information: Type of incident (vehicle breakdown or accident), time (start-time and end-time in terms of month, date, hour and minute), types of affected lanes, name and direction of the expressway along which the incident happened, and the location of incident (road segment id, latitude & longitude). The type of affected lane is represented by a serial number according to its position from right to left as lane 1,

[1] Muhammad Tayyab Asif was with Nanyang Technological University when this work was done. Currently, he is with IBM Research, Singapore.

2, 3 etc. Singapore's entire road network has 11 expressways, which are divided into 2156 road segments for analysis. We consider 8399 breakdowns and 2052 accidents recorded on those expressways in the period of four months (Aug–Nov 2014). Furthermore, the weather data contain the rain intensity information across the island of Singapore. These images have a time resolution of 5 minutes [10] and each pixel corresponds to an area of about 100 $\times$ 100 meters.

In this study, we consider the following nine features for each traffic incident i: day of week ($w_i \in \{0, 1\}$, where 0 represents weekend and 1 represents week-day), time of day ($t_i \in \{0, 1\}$, where 0 represents off-peak and 1 represents peak-hour), total number of lanes ($n_i \in \{1, 2, 3, 4, 5\}$), shoulder affected or not ($s_i \in \{0, 1\}$, where 0 represents not affected and 1 represents affected), number of lanes affected ($l_i \in \{0, 1, 2\}$), the type of affected lane ($a_i \in \{0, 1, 2, 3, 4, 5, 6\}$, where 0 represents no lane affected and 1, 2, 3, ... represent the serial number of the affected lane according to its position from right to left), expressway ($e_i \in \{0, 1, 2, ..., 11\}$), direction ($d_i \in \{0, 1\}$, where 0 represents upstream and 1 represents downstream), and rainfall effect ($r_i \in \{0, 1\}$, where 0 represents no rainfall and 1 represents strong rainfall).

III. Bayesian Support Vector Regression as Prediction Method

As Bayesian SVR can provide the confidence intervals along with the predicted values, we apply this method to model the relationship between traffic factors and incident duration.

Let us define a vector $\mathbf{x}_i$ containing input features as $\mathbf{x}_i = [w_i, t_i, n_i, s_i, l_i, a_i, e_i, d_i, r_i]^T$. The basic idea of SVR is to find the optimal hyperplane $\mathbf{w}$ so that the input vector $\mathbf{x}_i \in \mathbb{R}^n$ is mapped from low-dimensional space to high dimensional space by a nonlinear mapping $\phi(\mathbf{x}_i)$ to get a linear decision function [4]:

$$f(\mathbf{x}_i) = \mathbf{w}^T.\phi(\mathbf{x}_i) + b \quad b \in \mathbb{R}. \tag{2}$$

Let us further extend the SVR formulation to a probabilistic framework called BSVR [6][11][12]. For BSVR, we consider following regression model:

$$y_i = f(\mathbf{x}_i) + \delta_i \quad \mathbf{x}_i \in \mathbb{R}^n, \quad \mathbf{y}_i \in \mathbb{R}. \tag{3}$$

where f is the relationship function with i.i.d noise samples δ_i [13]. Now, let us consider $\mathbf{f} = [f(\mathbf{x}_1)...f(\mathbf{x}_N)]^T$, where N is total number of incidents. In Bayesian framework, we consider $\mathbf{f}$ as a random vector with prior probability $P(\mathbf{f})$. Therefore, the probability of $\mathbf{f}$ for a given training data set D is obtained by applying Bayes' theorem:

$$P(\mathbf{f}|D) = \frac{P(D|\mathbf{f})P(\mathbf{f})}{P(D)}. \tag{4}$$

which is defined as *a posteriori* distribution. The uncertainty in prediction can either arise due to the noise in the data δ_i or the limitations in model generalization $P(\mathbf{f}|D)$. We compute error bars by incorporating both of these uncertainties [11]. Let

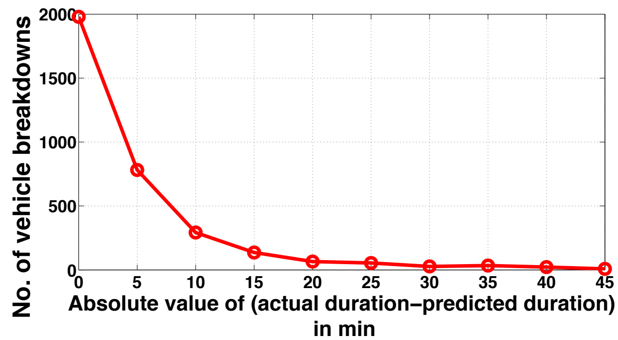

Fig. 1: Absolute error distribution of vehicle breakdowns in Singapore.

us assume the variance due to noise is σ_n^2 and the variance coming from model fitting issue will be σ_D^2. Therefore, the error bar will be estimated by $\sqrt{\sigma_n^2 + \sigma_D^2}$ at the time of prediction [13].

We follow the similar steps mentioned in [6] to implement BSVR. To evaluate the prediction performance, we calculate the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}}, \tag{5}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N} |e_i|}{N}, \tag{6}$$

where N is total number of incidents and e_i is the error between the actual and predicted duration d_i and $\hat{d}_i$ respectively:

$$e_i = d_i - \hat{d}_i. \tag{7}$$

Later, we perform sensitivity and specificity analysis to evaluate the detection performance of BSVR.

IV. Prediction Performance of BSVR

In this section, we analyze the performance of Bayesian SVR method in predicting the duration of non-recurring road-incidents in the network of Singapore. Fig. 2 shows the distributions of the absolute error for vehicle breakdowns and traffic accidents. The prediction performance in case of accidents seems to have higher uncertainty, in comparison with vehicle breakdowns. For some accidents, the prediction error is even larger than 90 min. The RMSE and MAE values obtained for vehicle breakdowns and accidents in Singapore are mentioned in Table I. From the tables we can conclude that it is generally more difficult to predict the duration of accidents than breakdowns.

TABLE I: Prediction error for incidents in Singapore.

	Vehicle breakdowns	Accidents
RMSE	14.97 min	24.77 min
MAE	10.69 min	20.28 min

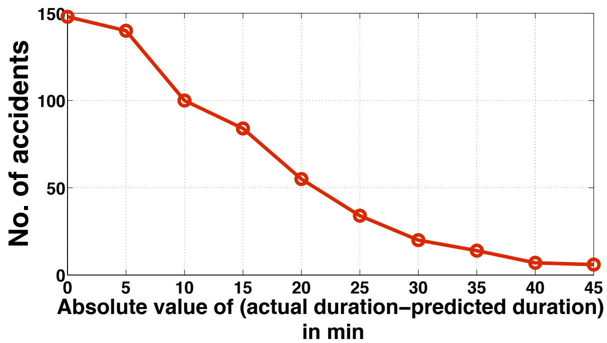

Fig. 2: Absolute error distribution of accidents in Singapore.

V. Anticipating Uncertainty Associated with Prediction Errors

In this section, we determine the error bars associated with the predicted duration of incidents. The predicted values obtained in the previous section tend to have uncertainty because the errors are subject to variations in test data-set. Therefore, we apply BSVR to anticipate the variations in prediction accuracy. Moreover, we analyze sensitivity and specificity to evaluate the detection performance of BSVR.

A. Correspondence of the Error-bars with Prediction Errors

In this subsection, we show the correspondence of the error bars with predicted incident duration for both vehicle breakdowns and accidents in Singapore. To this end, we group them separately according to the absolute value of prediction errors in three different categories: $0-10$ min, $10-20$ min and $20-30$ min. Further, we compute the average of the error-bars associated with the incidents of each range and plot it against the ranges in Fig. 3 separately for vehicle breakdowns and accidents. From Fig. 3, we can conclude that on an average the

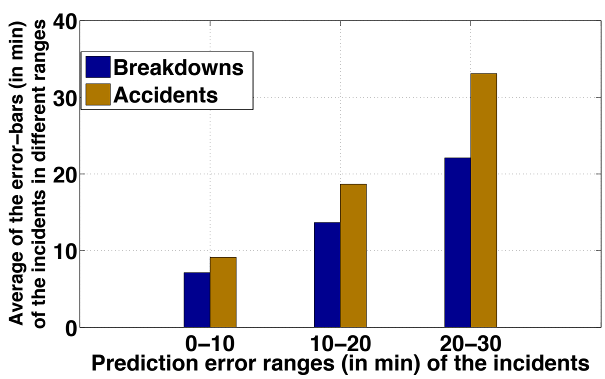

Fig. 3: Average of the error-bars (in min) associated with different ranges of prediction errors (in min) for vehicle breakdowns and accidents in Singapore.

values of error bars are higher for accidents, which indicates the accident data is more volatile.

B. Sensitivity and Specificity Analysis

In this subsection, we perform sensitivity and specificity analysis to estimate the detection performance of BSVR. For this purpose, we consider different tolerance values τ_d of absolute prediction errors, for example $\tau_d = \{4\sigma, 5\sigma\}$; where σ is the standard deviation of the prediction errors. The errors which are higher than the tolerance limit τ_d are considered to be positive events and vice versa. Our goal is to anticipate these large errors in prediction by utilizing information provided by the error bars. To this end, we consider this problem as a detection problem. If the magnitude of error bar is higher than a pre-specified threshold, then we anticipate that the prediction performed will be highly uncertain. This detector threshold is represented by γ_d, where γ_d is the mean of error-bars.

There are four possible outcomes in this case: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). We define the two parameters sensitivity and specificity as:

$$\text{Sensitivity} = \frac{\text{number of true positives (TP)}}{\text{number of positive events (TP+FN)}}. \tag{8}$$

$$\text{Specificity} = \frac{\text{number of true negatives (TN)}}{\text{number of negative events (TN+FP)}}. \tag{9}$$

Our only constraint is to keep False Positive rate (F.P.R.) low ($\leq 30\%$), where

$$\text{F.P.R.} = 1 - \text{Specificity}. \tag{10}$$

We demonstrate the specificity-sensitivity profiles in Fig. 4 and Fig. 5 for vehicle breakdowns and accidents respectively and analyze whether we can obtain high sensitivity with our constraint. The blue line is termed as the no-discrimination line, which represents the performance of a detector that randomly selects an event to be either positive or negative. Other curves should remain above this line for a detector to be useful. In Fig. 4 and Fig. 5, BSVR can detect around

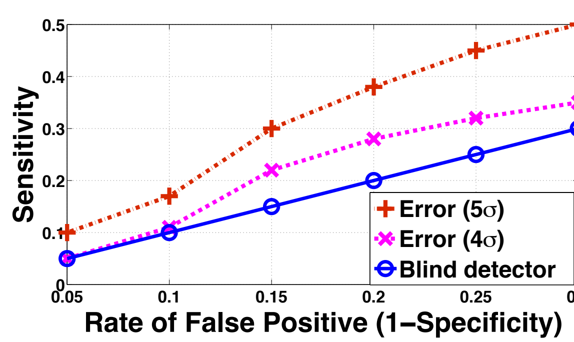

Fig. 4: Specificity-sensitivity profile for the vehicle breakdowns in Singapore.

50% instances of large error for the tolerance level 5σ. For this level of sensitivity, it only reports false alarms in around $25\%-30\%$ of incidents (i.e. specificity $70\%-75\%$). However,

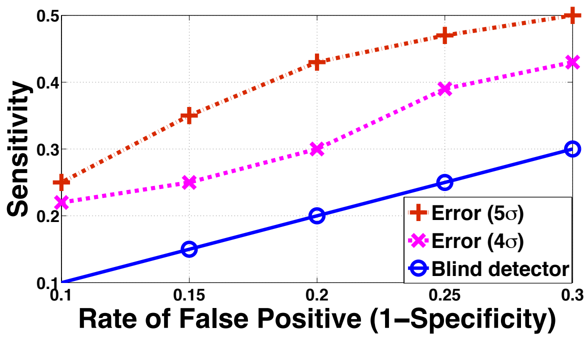

Fig. 5: Specificity-sensitivity profile for the accidents in Singapore.

for tighter error tolerance (prediction error 2σ), we observe degraded sensitivity.

Let us now show the positions of the vehicle breakdowns and accidents in Fig. 6 and Fig. 7 to locate the training data-points and the positive events (i.e. the incidents which have large prediction error) in the map of Singapore. We observe in

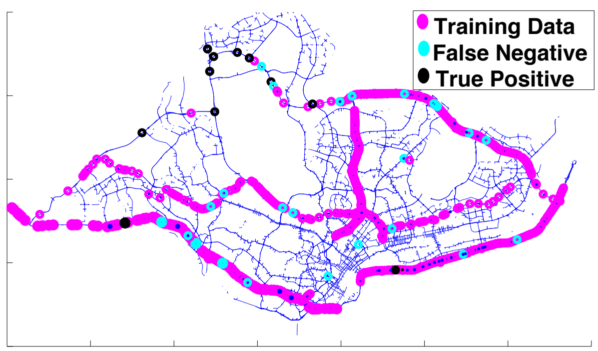

Fig. 6: Location of test data-points of vehicle breakdowns corresponding to the positive events.

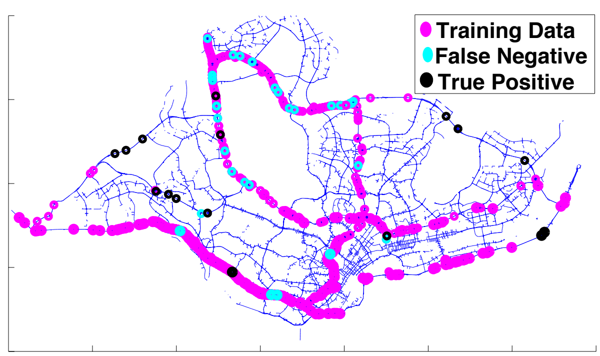

Fig. 7: Location of test data-points of accidents corresponding to the positive events.

Fig. 6 and Fig. 7 that the false negative incidents are located in the region having high density of training data-points, because the detector did not expect large prediction error for these incidents. On the contrary, true positive events are located away from training data-points, where the input features of these incidents are different from the training data-set. Hence, the detector was able to detect these incidents.

VI. Conclusion

In this paper, we proposed BSVR to provide information about uncertainty in predicting duration of incidents in real-time. To this end, we considered traffic incidents data from Singapore. We performed sensitivity and specificity analysis to evaluate the detection efficiency of BSVR and found that it can detect variations in prediction error with reasonable accuracy. In future work, we plan to analyze larger datasets and apply Gaussian Process to compare the results with Bayesian SVR for anticipating the uncertainty associated with the prediction error values.

Acknowledgment

The authors wish to thank the Land Transport Authority and the National Environmental Agency of Singapore for providing incident and rainfall data from the expressways of Singapore.

References

[1] Q. He, Y. Kamarianakis, K. Jintanakul, and L. Wynter, “Incident duration prediction with hybrid tree-based quantile regression,” in *Advances in Dynamic Network Modeling in Complex Transportation Systems*. Springer, 2013, pp. 287–305.

[2] Y. He, S. Blandin, L. Wynter, and B. Trager, “Analysis and real-time prediction of local incident impact on transportation networks,” in *Data Mining Workshop (ICDMW), 2014 IEEE International Conference on*. IEEE, 2014, pp. 158–166.

[3] L. Ruimin, Z. Xiaoqiang, Y. Xinxin, L. Junwei, C. Nan, and Z. Jie, “Incident duration model on urban freeways using three different algorithms of decision tree,” in *Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on*, vol. 2. IEEE, 2010, pp. 526–528.

[4] W. Wu, S. y. Chen, and C. j. Zheng, “Traffic incident duration prediction based on support vector regression,” *Proceedings of the ICCTP*, pp. 2412–2421, 2011.

[5] M. Seeger, “Bayesian model selection for support vector machines, gaussian processes and other kernel classifiers,” in *Proceedings of the 13th Annual Conference on Neural Information Processing Systems*, no. EPFL-CONF-161324, 2000, pp. 603–609.

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

[7] D. Chimba, B. Kutela, G. Ogletree, F. Horne, and M. Tugwell, “Impact of abandoned and disabled vehicles on freeway incident duration,” *Journal of transportation engineering*, vol. 140, no. 3, p. 04013013, 2013.

[8] C. I. 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.

[9] J. Ahn, E. Ko, and E. Y. Kim, “Highway traffic flow prediction using support vector regression and bayesian classifier,” in *2016 International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2016, pp. 239–244.

[10] L. W. Leong, K. Lee, K. Swapnil, X. Li, H. Y. T. Victor, N. Mitrovic, M. T. Asif, J. Dauwels, and P. Jaillet, “Improving traffic prediction by including rainfall data,” *ITS Asia-Pacific Forum*, vol. 14, 2015.

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

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

[13] G. Gopi, J. Dauwels, M. T. Asif, S. Ashwin, N. Mitrovic, U. Rasheed, and P. Jaillet, “Bayesian support vector regression for traffic speed prediction with error bars,” in *Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on*. IEEE, 2013, pp. 136–141.