The Development of Robust Real-time Crash Prediction Models with Bayesian Network

Ananya ROY[1], Yasunori MUROMACHI [2]

[1]Graduate Student, School of Environment and Society, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail:roy.a.aa@m.titech.ac.jp
[2]Associate Professor, School of Environment and Society, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail: ymuro@enveng.titech.ac.jp

The RTCP model is expected to make inference with partially available data, because traffic flow variables are highly correlated with some missing values. RTCP models with Bayesian Network (BN) which is a probabilistic graphical modeling method with a certain degree of robustness. It has two inbuilt learning process using Expectation Maximization (EM) Algorithm and Adaptation Algorithm which enables BN based models to deal with missing values.But to build an advanced RTCP model using BN, it is wise to identify the most influential traffic parameters and their combinations. Considering these facts, this study proposes BN based RTCP models with twenty four combinations of twelve traffic parameters. After modelling, their performances were validated and compared to identify the more preferable combinations of input variables. Then different combinations of the variables were applied to find the best possible input parameters to build the model namely-difference between upstream and downstream Congestion Index, flow, speed and upstream Congestion Index which proved to be the most effective combination of input variables.

Key Words : *Real-Time Crash Prediction, Urban Expressway, Bayesian Network*

1. INTRODUCTION

One of the greatest contributions of the advancement of the Intelligent Transportation System (ITS) and Advanced Transportation Information Systems (ATIS) is the in the rising field of Real-time Crash Prediction (RTCP) Models. Although, ITS has made instantaneous data collection very easy and efficient, there are often some discrepancies in the data due to instrumental or sensor malfunction causing missing data. Hence, the RTCP model is expected to make inference with partially available data. Additionally, traffic flow variables are highly correlated in nature, thus a flexible modeling method should be chosen which could be able to overcome these problems and at the same time could update itself with time without re-building or calibrating from initial level. Hence, a practical real-time crash prediction model should be able to handle missing data. Also, in future, with the introduction of modern sensors, the model should have the ability to incorporate new variables without requiring it to be built from the scratch. It should also be able to make predictions when information on all the variables is not available. Finally, the existing models are built based on the available detector spacing and traffic condition specific to the concerning expressway. This emphasizes the importance of having a model that have the ability to update itself in real-time and customize itself to match with traffic condition of the expressway where it will be employed, i.e., they must be transferrable. Considering these requirements, this paper proposes Bayesian Belief Net (BBN) as a platform to build real-time crash prediction models [1,2)]. BBN is a probabilistic graphical modeling method that describes complex joint distributions of a system through a graph and local distributions, i.e., conditional probabilities. It has the inherent capability to handle missing data while model building, it can accommodate new variables and update itself accordingly even after it is fully built and develop itself with partially available data. It also supports sequential learning, which allows it to update itself in real-time whenever new data becomes available.

Missing data is a critical issue that often plagues any real-time system. Most models developed in the literature are based on an assumption of no missing data. Missing just one item in the critical data stream may prevent the system from functioning properly.

The proposed Bayesian Network based prediction method in this study, which has two inbuilt learning process using Expectation Maximization (EM) Algorithm and Adaptation Algorithm which enables BN based models to calculating the expected sufficient statistics and then maximizing its likelihood of any information variable that is missing, due to detector malfunctions and/or communication problems. This method's embedded missing data estimation module and the incident detection module have significantly increased the RTCP model's operational reliability.

But to build a practical and advanced RTCP model using BN, it is wise to identify the most influential factors i.e. traffic parameters. It is necessary to know which traffic parameters should be used while building a model, or if any parameter is missing from the input data set, which are the next best combination of input parameters. The descriptive statistics (e.g., average, standard deviation, variance, coefficient of variation, etc.) of the basic traffic flow variables (e.g., vehicle count, speed, occupancy, etc.) yielded by the traffic sensors form a very large variable space. In previous studies it was found that the standard deviation of speed is the most suitable variable to distinguish between normal and disrupted traffic condition. They applied Bayesian statistics and, in a later study [7)], they used Probabilistic Neural Network (PNN) method, and found standard deviation of both speed and occupancy to be suitable predictors. Later on, another study applied first order log-linear models to predict crash at given road geometry, weather condition and time of the day using speed variations along a lane, traffic queue and traffic density as predictors [8,9)]. Afterwards, PNN was chosen as method and mean and variance of volume, occupancy and speed as predictors to build real-time models [10)]. Also, Generalized Estimating Equation method was applied in some models where road geometry was included as variable as well [11)]. The latest study applied three different methods – K means clustering, Naïve Bayes method and Discriminant analysis including the joint effect of two or more traffic variables to identify traffic patterns leading to crash [12)].

Later on, Bayesian Network (BN) was introduced and successfully used to build real-time crash model where flow, speed and occupancy data were used as inputs [13)]. Afterwards, the authors conducted study on urban expressway and applied BN with flow and speed as traffic variables. This model could predict up to 54% crashes with a prediction success of 85% depending on the threshold value [14)]. In another study, they worked with a huge data set and did a thorough analysis on the basic freeway segment (BFS) of an urban expressway, where they used Random Multinomial Logit (RMNL) to select appropriate predictors and then applied BN for real-time prediction model building [15)]. Their study introduced a term 'congestion index' as a variable along with other traffic variables like difference in speed and occupancy between upstream and downstream. The model was robust enough to predict 66% crash cases with only 20% of false alarm.

The paper is organized in five sections. This section has already addressed the problems with the present crash prediction models and indicated advantages of BN based real-time crash prediction models over the past models. In the second section, the data collection is discussed. The third section explains the methodology of this study and the forth section deals with the model building and validation. Lastly, the fifth section draws conclusion and future scope of this study.

2. DATA

(1) Study Area

Shinjuku 4 route of the Tokyo Metropolitan Expressway was selected for this study. It is about 13.5km long and one of the busiest expressways in Japan. Shinjuku 4 route is connected with the Chuo expressway starting at the point on the boundary of the Tokyo Metropolitan area (**Fig. 1**). Two types of data– detector data and crash data– were collected from Tokyo Metropolitan Expressway Company Limited for six months (March 2014 to August 2014). The expressway has two lanes in each directions with 74 detectors (about 250 meters apart) in one lane. In this paper, only inbound route is used for analysis. About 101 crash cases were reported for this direction.

(2) Data Collection and Processing

Six months detector data and crash data were extracted. Detector data consists of detector location (kilo post), speed (1min average speed), flow (1min) and occupancy. Crash data contained information about date, time (in minutes), location (to nearest 10

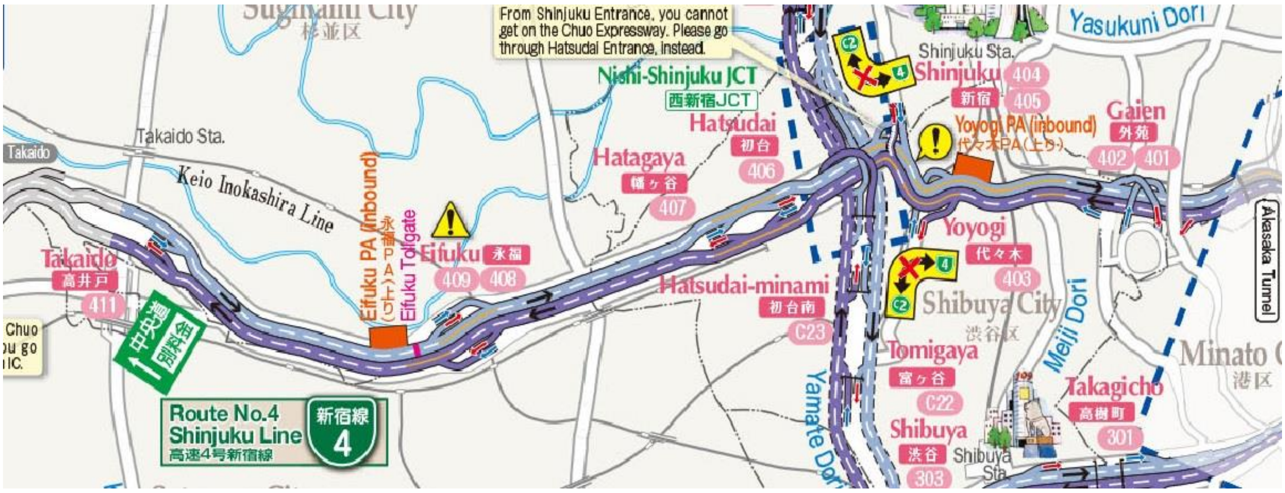

Figure 1 Study area: Shinjuku 4 Tokyo Metropolitan Expressway, Japan
Source: Official website of Tokyo Metropolitan Expressway Company Limited

meters), crash lane, type of crash and vehicle involvement.

Two types of data- crash and normal data are needed for the model. For crash data, 1 minute data, before the crash was extracted and for normal data, the same was done for the same day of all other weeks. For example, if a crash was reported on 26th March (Wednesday) 03:00AM, then 02:58- 2:59AM is considered as crash data. In case of selecting normal data, flow, speed and occupancy for 02:58-2:59AM time period for every other Wednesday were collected. However, crashes might occur in other Wednesdays during the same time period selected. To avoid misleading data, we removed all normal condition data where a crash took place on the same date before or after 1 hour of the selected time period. After all these screening, there were a total of 101 nos of crash and 1732 nos of normal data left for model building. Two separate datasets were used for model building and validation.

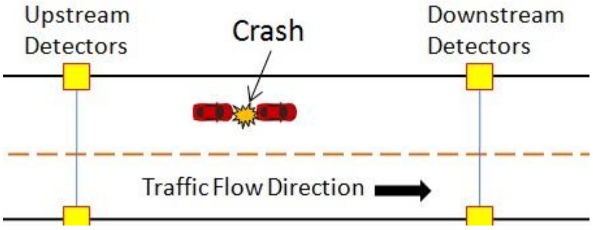

Figure 2 Data Collection Procedure

In order to develop a real-time crash prediction model, pairs of detectors (**Fig. 2**)- nearest upstream and nearest downstream of the crash location was considered for the entire route of the expressway[3]. Then, difference of upstream and downstream flow (q), speed (v), occupancy(o) and congestion index (CI) were calculated.

3. METHODOLOGY

In this section, the basis of Bayesian Network and its concept of inference is briefly explained. Bayesian Network, is a probabilistic graphical modeling method where we represent a system with a graph and a joint probability distribution compacted with the notion of conditional independence. Later, we can use this model of system to understand the dynamics within the system and also to predict the state of variables in lights of the evidence on any one or more variables.

Fig. 3 presents a simple BBN involving five variables. Here, each variable is represented with a node and the influence of one variable on others is demonstrated with directed edges (may or may not represent causality). We would like to mention here that these graphs are acyclic in nature and are called acyclic directed graph (DAG).Hence, the BBN in **Fig. 3** can be written as:

$$P(A,B,C,D,E)$$
$$= P(A)P(B|A,C)P(C)P(D|B,C)P(E|C) \quad (1)$$

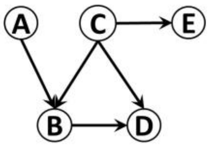

Figure 3 Bayesian Network (Batch Learning (EM-Algorithm) and Sequential Learning (Adaptation Algorithm))

The task of EM-Algorithm [18)] in BBN is to determine the conditional probability tables (CPTs) for nodes based on prior probabilities and availability of new N number of records. The algorithm has two steps – calculating the expected sufficient statistics and then maximizing its likelihood. To elaborate more, if no probability is assigned to a variable for which we are estimating the parameters, a uniform distribution is assumed. Then, with presence of a batch of data, the new parameter is estimated in such way that first, the expected sufficient statistics under that parameter is calculated and then the log-likelihood of that parameter under the expected sufficient statistics is maximized. This is an iterative process and it stops when one of these two criteria are satisfied – i) the maximum number of iteration specified by the user has exceeded, or, ii) the relative log-likelihood between two successive iterations is smaller than the preset minimum difference value. Here, it is important to mention that the EM-algorithm does not need data on each of the variables to update the model. The adaptation algorithm [19] is similar to the EM-algorithm with the exception that here the evidence from each record is propagated throughout the network and the parameters for each of the variables are updated accordingly.

There are two types of variables in Bayesian Network- information variable and hypothesis variable. Information variables are those, the values of which are to be expected to be obtained to calculate the probability of the hypothesis variable. In thiscase, the information variables used are shown in Table 1. Whereas the hypothesis variable is the likelihood of crash or not.

Table 1 Information Variables Used in Model Building

Informatio Variables	*Respective Desccription*
u_flow, d_flow, q	Upstream flow, downstream flow, flow difference between upstream and downstream
u_speed, d_speed, v	Upstream speed, downstream speed, speed difference between upstream and downstream
u_occ, d_occ, o	Upstream occupancy, downstream occupancy, occupancy difference between upstream and downstream
u_CI, d_CI, CI	Upstream Congestion Index (CI), downstream CI, CI difference between upstream and downstream

Twenty four models (**Table 2**) were built with different combinations of above mentioned information and hypothesis variables. Out of 101 crash cases and 1732 normal cases, 71 crash cases and 1190 normal cases were used for 24 model building and the rest were used in model validation process.

All these crash prediction models were later on validated with missing variables and randomly selected partially missing traffic data. This way, for each model, atleast four cases were developed to check the adaptability of the model and influence of the information variables. For example, in case of model-1, in first case of evaluation, it was assumed that the detectors providing downstream flow (d_flow) are broken or generating erroneous flow data and only speed (v), upstream flow (u_flow) data is available. Hence, this case calculates P(*Crash*| *v, u_flow, CI*). Similarly, in second case, it was assumed that only the speed data was available, thus this case calculates P(*Crash*| *v,CI*). All cases are listed in **Table 3**. Congestion Index (CI) is calculated applying the following equation:

Table 2 List of Twenty Four Crash Prediction Models

Model No.	*Information Variables*	*Model No.*	*Information Variables*
Model-1	q, v, CI, u_flow	Model-13	o, v, CI, u_occ
Model-2	q, v, CI, d_flow	Model-14	o, v, CI, d_occ
Model-3	q, v, CI, u_speed	Model-15	o, v, CI, u_CI
Model-4	q, v, CI, d_speed	Model-16	o, v, CI, d_CI
Model-5	q, v, CI, u_occ	Model-17	CI, q, o, u_flow
Model-6	q, v, CI, d_occ	Model-18	CI, q, o, d_flow
Model-7	q, v, CI, u_CI	Model-19	CI, q, o, u_speed
Model-8	q, v, CI, d_CI	Model-20	CI, q, o, d_speed
Model-9	o, v, CI, u_flow	Model-21	CI, q, o, u_occ
Model-10	o, v, CI, d_flow	Model-22	CI, q, o, d_occ
Model-11	o, v, CI, U_speed	Model-23	CI, q, o, u_CI
Model-12	o, v, CI, d_speed	Model-24	CI, q, o, d_CI

Note: See Table 1 for description of the information variables.

Congestion Index (CI)
= (Free Flow Speed - Speed)/Free Flow Speed; when CI>0
= 0; when CI<=0 (2)

Table 3 Twenty Four Crash Prediction Models and Coresponding Cases with Missing Information Variables

Model No.	*Description of Missing Variables*	*Model No.*	*Description of Missing Variables*
Model-1 P(*Crash*| *q, v, CI, u_flow*)	*Case 1:* d_flow missing *Case 2:* u_flow missing *Case 3:* u_speed or d_speed missing *Case 4:* all variables are present	Model-13 P(*Crash*| *o, v, CI, u_occ*)	*Case 1:* u_occ missing *Case 2:* d_occ missing *Case 3:* u_speed or d_speed missing *Case 4:* all variables are present
Model-2 P(*Crash*| *q, v, CI, d_flow*)	*Case 1:* u_flow missing *Case 2:* d_flow missing *Case 3:* u_speed or d_speed missing *Case 4:* all variables are present	Model-14 P(*Crash*| *o, v, CI, d_occ*)	*Case 1:* d_occ missing *Case 2:* u_occ missing *Case 3:* u_speed or d_speed missing *Case 4:* all variables are present
Model-3 P(*Crash*| *q, v, CI, u_speed*)	*Case 1:* d or, u_flow missing *Case 2:* d_speed missing *Case 3:* u_speed missing *Case 4:* all variables are present	Model-15 P(*Crash*| *o, v, CI, u_CI*)	*Case 1:* o missing *Case 2:* u_speed missing *Case 3:* d_speed missing *Case 4:* all variables are present
Model-4 P(*Crash*| *q, v, CI, d_speed*)	*Case 1:* d or, u_flow missing *Case 2:* u_speed missing *Case 3:* d_speed missing *Case 4:* all variables are present	Model-16 P(*Crash*| *o, v, CI, d_CI*)	*Case 1:* o missing *Case 2:* d_speed missing *Case 3:* u_speed missing *Case 4:* all variables are present
Model-5 P(*Crash*| *q, v, CI, u_occ*)	*Case 1:* u_occ missing *Case 2:* u_speed or d_speed missing *Case 3:* q missing *Case 4:* all variables are present	Model-17 P(*Crash*| *CI, q, o, u_flow*)	*Case 1:* o missing *Case 2:* u_speed or d_speed missing *Case 3:* u_flow missing *Case 4:* d_flow missing *Case 5:* all variables are present
Model-6 P(*Crash*| *q, v, CI, d_occ*)	*Case 1:* d_occ missing *Case 2:* u_speed or d_speed missing *Case 3:* q missing *Case 4:* all variables are present	Model-18 P(*Crash*| *CI, q, o, d_flow*)	*Case 1:* o missing *Case 2:* u_speed or d_speed missing *Case 3:* d_flow missing *Case 4:* u_flow missing *Case 5:* all variables are present
Model-7 P(*Crash*| *q, v, CI, u_CI*)	*Case 1:* u_spped missing *Case 2:* d_speed missing *Case 3:* q missing *Case 4:* all variables are present	Model-19 P(*Crash*| *CI, q, o, u_speed*)	*Case 1:* o missing *Case 2:* u_speed missing *Case 3:* d_speed missing *Case 4:* q missing *Case 5:* all variables are present
Model-8 P(*Crash*| *q, v, CI, d_CI*)	*Case 1:* d_spped missing *Case 2:* u_speed missing *Case 3:* q missing *Case 4:* all variables are present	Model-20 P(*Crash*| *CI, q, o, d_speed*)	*Case 1:* o missing *Case 2:* d_speed missing *Case 3:* u_speed missing *Case 4:* q missing *Case 5:* all variables are present
Model-9 P(*Crash*| *o, v, CI, u_flow*)	*Case 1:* o missing *Case 2:* u_speed or d_speed missing *Case 3:* u_flow missing *Case 4:* all variables are present	Model-21 P(*Crash*| *CI, q, o, u_occ*)	*Case 1:* u_speed or d_speed missing *Case 2:* q missing *Case 3:* u_occ missing *Case 4:* d_occ missing *Case 5:* all variables are present
Model-10 P(*Crash*| *o, v, CI, d_flow*)	*Case 1:* o missing *Case 2:* u_speed or d_speed missing *Case 3:* d_flow missing *Case 4:* all variables are present	Model-22 P(*Crash*| *CI, q, o, d_occ*)	*Case 1:* u_speed or d_speed missing *Case 2:* q missing *Case 3:* d_occ missing *Case 4:* u_occ missing *Case 5:* all variables are present
Model-11 P(*Crash*| *o, v, CI, U_speed*)	*Case 1:* o missing *Case 2:* u_speed missing *Case 3:* d_speed missing *Case 4:* all variables are present	Model-23 P(*Crash*| *CI, q, o, u_CI*)	*Case 1:* u_speed missing *Case 2:* d_speed missing *Case 3:* u_occ missing *Case 4:* o missing *Case 5:* all variables are present
Model-12 P(*Crash*| *o, v, CI, d_speed*)	*Case 1:* o missing *Case 2:* d_speed missing *Case 3:* u_speed missing *Case 4:* all variables are present	Model-24 P(*Crash*| *CI, q, o, d_CI*)	*Case 1:* d_speed missing *Case 2:* u_speed missing *Case 3:* u_occ missing *Case 4:* o missing *Case 5:* all variables are present

4. MODEL BUILDING AND VALIDATION

Twenty four real-time crash prediction models mentioned in **Table 2**, were built using Bayesian Network with the help of Hugin Researcher software.

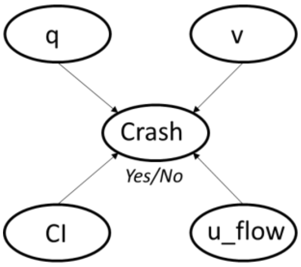

Figure 4 Structure of Model-1 built with Bayesian Network

Note: See Table 1 for description of the information variables

In the aforementioned model, four information variables (parents) i.e. relative flow between upstream and downstream *(q),* relative speed between upstream and downstream *(v),* relative congestion index between upstream and downstream *(CI), and* upstream flow *(u_flow)* are directed towards the single hypothesis variable (child), crash. This hypothesis variable denotes the crash likelihood, which has only two outputs- 'yes' and 'no'. The Directed Acyclic Graph (DAG) model is depicted in **Fig. 4**, where

- *crash* is parent to none, and
- *q, v, CI, u_flow* are directed towards the node *crash* with edges or arcs

Each model was built with 1-minute data. And, for each model structure, there were three models. For example, Model-1 has three corresponding models namely-

- Model-1 (t-1): information variables data were collected 1 minute before crash
- Model-1 (t-2) : information variables data were collected 2 minute before crash
- Model-1 (t-3) : information variables data were collected 3 minute before crash

Therefore, there were actually (24*3=) 72 models which were evaluated with four (and five) cases of missing values and variables.

The model's mathematical expression is shown in equation (3). After running the EM-algorithm, the marginal probability of crash is found to be 8.65, 9.41 and 8.92% respectively for Model-1 (t-1), Model-1 (t-2) and Model-1 (t-3).

$$P(crash,q,v,CI,u\_flow)=P(q).P(v|q).P(CI|v,q).P(u\_flow|v,q,CI)\ .\ P(crash|v,q,CI,u\_flow) \qquad (3)$$

Next, we used these models to evaluate the prediction performance when information on different variables is missing. For that, we alternatively entered the crash and normal traffic condition data into Model-1. Finally, 8.65, 9.41 and 8.92% were set as the cut point or threshold to distinguish crash from normal traffic condition for Case 1 to Case 4. The similar procedure was followed for rest of the models and all cases. The overall performances of all models are shown in **Fig. A-1 through Fig. A-9** (Appendix).

The success of the model depend on its combined performance to predict crash and normal traffic conditions.

$$\text{Crash} = (\text{Calculated probability over threshold/ crash sample size})*100$$
$$\text{Normal traffic} = (\text{Calculated probability below threshold/ normal sample size})*100 \qquad (4)$$

Let's take evaluation performance of Model-14 (t-1) as an example. From **Table 5**, it can be observed that at a threshold value of 8.13% the model can identify 43.3% of crashes and 81.2% normal situation with a false alarm of 18.8%, when all the information variables are present (Case-4), which gives an overall prediction performace of 79.2%. Even, when upstream and/or downstream speed data were unavailable due to sensor problem (Case- 3) it can predict upto 46.7% crash and 78.4% normal situations which refers to 76.75% of overall accuracy. Again, Case-1, which means downstream occupancy value was missing, the model can identify around 43.3% crash and 67.9% normal situations with an overall performance of 66.61%. Thus, when upstream occupancy data, and upstream and/ or downstream

Table 4 Overall Performance Evaluation of Model-14

	Missing Data	*Model-14 (t-1), %*	*Model-14 (t-2), %*
Case-1	d_occ	66.61	75.87
Case-2	u_occ	78.15	83.39
Case-3	u_speed and/ or d_speed	76.75	75.35
Case-4	Null	79.20	84.27

Table 5 Performance Evaluation of Model-14 (t-1) and Model-14 (t-2)

	Model-14 (t-1), %		*Model-14 (t-2), %*	
	crash	normal	crash	normal
Case-1	0.433	0.679	0.400	0.779
Case-2	0.433	0.801	0.200	0.869
Case-3	0.467	0.784	0.367	0.775
Case-4	0.433	0.812	0.433	0.865

speed data isn't available, the model is able to perform almost as well as with all the available data. This could be summarized as downstream occupancy is the most influential information variable for Model-14 (t-1). Similar kind of analogy could be given for other models separately.

5. CONCLUSIONS

Being a relatively new method, Real-Time Crash Prediction Models (RTCP) face hindrances which need to be overcomed. Among all the obstacles, the major are- i) small sample size: its difficult to get high resolution dense traffic data and corresponding crash informations,
ii) unavailability, or missing values: As difficult it is to get high resolution data, it's even quite common for sensors to go out of order and not being able to provide information on all the variables,
iii) current Real-Time Crash Prediction (RTCP) Models are built based on data obtained from fixed sensor/ detector layout: transferring a model for a different sensor/ detector layout is necessary.

To master these hardles, in this paper, RTCP Models built using Bayesian Network is proposed and its robustness is examined in terms of missing variable and data. Various models were created to evaluate the capability of BN to update itself both by adding new variables as well as the parameters of the models when data on new variable as well as existing variable become available in future. In addition, performance of the models were evaluated when some of the sensors go out of order and cannot yield data on all the variables. From (24*3=) 72 models, it was perceived that the models perform faily good when there are absence of information variables and missing traffic data. The overall performance of models in case of missing values is very close to the situation when all data are available (Appendix). Moreover, for separate models, influencing information variables could be identified based on the models prediction performance with and without their existence. This study will help to improve RTCP Models in future and will assist to apply these model in reality.

ACKNOWLEDGMENT: We would like to thank Tokyo Metropolitan Expressway Company Limited for providing necessary data to conduct this study. This study is also supported by JSPS KAKENHI Grant Number 25420535.

APPENDIX

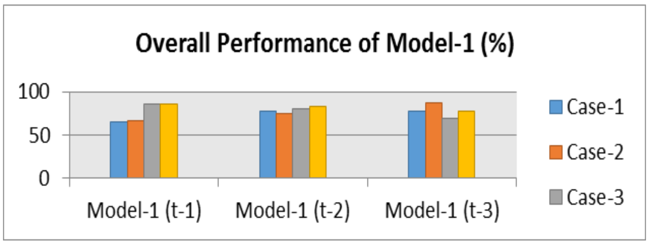

Figure A1: Overall Prediction Performance of Model -1

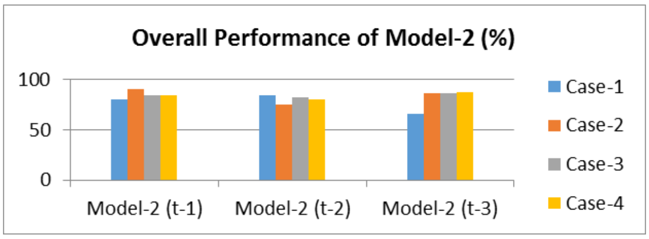

Figure A2: Overall Prediction Performance of Model -2

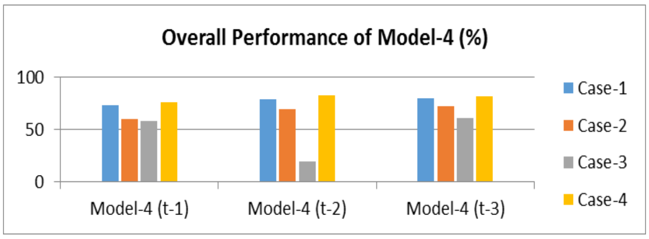

Figure A3: Overall Prediction Performance of Model -4

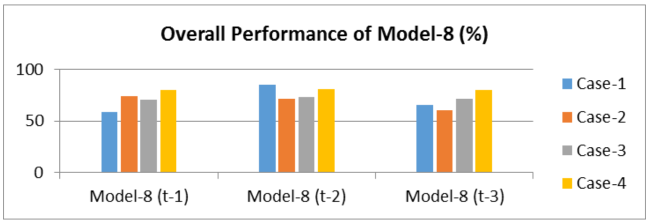

Figure A4: Overall Prediction Performance of Model -8

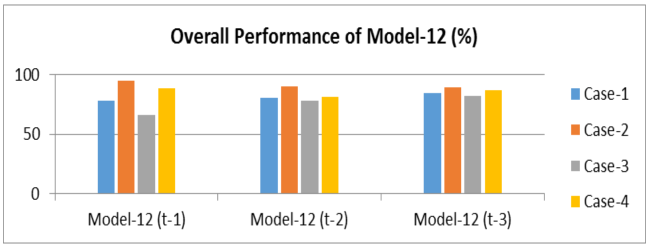

Figure A5: Overall Prediction Performance of Model -12

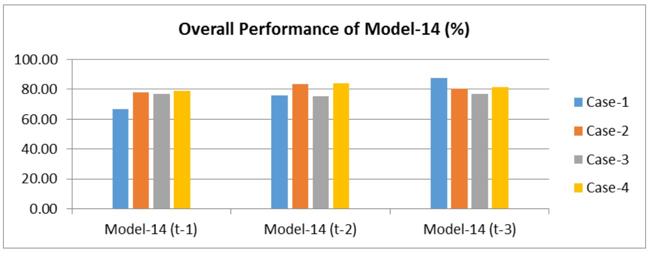

Figure A6: Overall Prediction Performance of Model -14

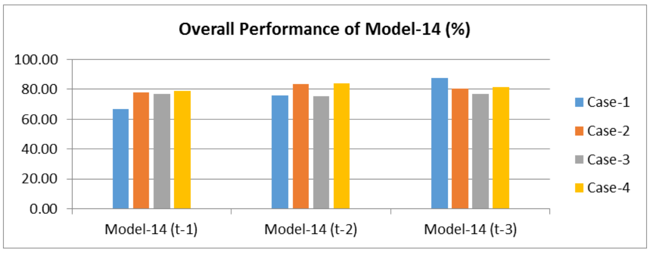

Figure A7: Overall Prediction Performance of Model -14

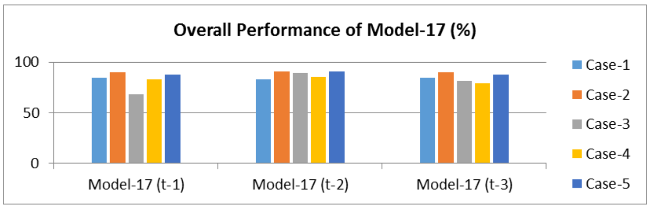

Figure A8: Overall Prediction Performance of Model -17

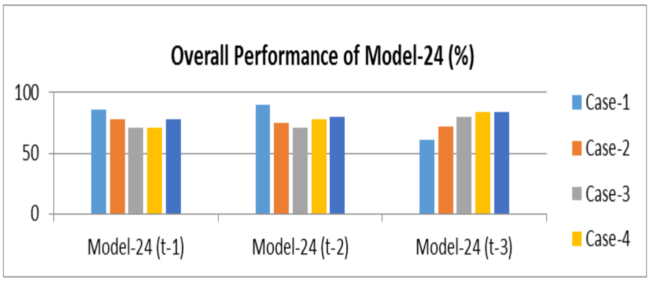

Figure A9: Overall Prediction Performance of Model -24

REFERENCES

1) Jensen, F. V. and T. D. Nielsen. *Bayesian network and decision graphs*. 2nd Edition, Springer, New York, USA, 2007.

2) Kjaerulff, U. B., and A. L. Bayesian Networks and Influence Diagrams: A guide to construction and analysis. Springer, New York, USA, 2008.

3) Hossain, M., Muromachi, Y.: A framework for Real-Time Crash Prediction: Statistical approach vesus artificial intelligence, *土木計画学研究・論文集　Vol.26 no.5　2009年9月*

4) Chen, H., Grant-Muller, S., Mussone, L. and Montgomery, F., (2001) "A Study of Hybrid Neural Network Approaches and the Effects of Missing Data on Traffic Forecasting," Neural Computing and Applications, Vol. 10, pp. 277– 286.

5) Rice, J., and van Zwet, E., (2004) "A Simple and Effective Method for Predicting Travel Times on Freeways," IEEE Transactions on Intelligent Transportation Systems, Vol. 5(3), pp. 200-207.

6) Thomas A. Dingusa,1, Feng Guoa,b, Suzie Leea, Jonathan F. Antina, Miguel Pereza, Mindy Buchanan-Kinga, and Jonathan Hankeya (2016) "Driver crash risk factors and prevalence evaluation using naturalistic driving data" Procedeedings of the NAtional Academy of Sciences of the United States of America.

7) Heydecker, B.G., Wu, J.: Identification of sites for road accident remedial work by Bayesian Statistical Methods: an example of uncertain inference, *Journal of Advances in Engineering Software*, Vol 32, pp. 859-869, 2001.

8) Kumara, S.S.P., H.C. Chin, and W.M.S.B. Weerakoon: Identification of accident causal factors and prediction of hazardousness of intersection approaches, *Transportation Research Record 1840, TRB, National Research Council, Washington DC*, pp. 116-122, 2003.

9) Tarko, A.P., and M. Kanodia: Effective and fair identification of hazardous locations, *Transportation Research Record: Journal of the Transportation Research Board, No. 1897, Transportation Research Board of the National Academics, Washington, D.C.*, pp. 64-70, 2004.

10) Abdel-Aty, M. and Pande, A.: Classification of real-time traffic speed patterns to prediction crashes on freeways, *Preprint No. TRB 04-2635, 83rd Annual Meeting of Transportation Research Board, Washington, D.C., USA*, 2004.

11) Abdel-Aty, M., and Abdalla, M. F.: Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes using generalized estimating equations for correlated data, *Transportation Research Record: Journal of the Transportation Research Board, No. 1897, TRB, National Research Council, Washington, D.C.*, pp 106–115, 2004.

12) Oh, C., Oh, J., Ritchie, S., and Chang, M.: Real time estimation of freeway accident likelihood, *80th Annual Meeting of Transportation Research Board*, 2001.

13) Oh, J., Oh, C., Ritchie, S., and Chang, M.: Real time estimation of accident likelihood for safety enhancement, *ASCE Journal of Transportation Engineering*, Vol. 131, No. 5, pp 358-363, 2005.

14) Lee. C., Saccomanno, F., and Hellinga, B.: Analysis of crash precursors on instrumented freeways, *81st Annual Meeting of Transportation Research Board*, 2002.

15) Lee. C., Saccomanno, F., and Hellinga, B.: Real-time crash prediction model for the application to crash prevention in freeway traffic, *82nd Annual Meeting of Transportation Research Board*, 2003.

16) Abdel-Aty, Mohamed and Pande, Anurag: Classification of real-time traffic speed patterns to prediction crashes on freeways, *83rd Annual Meeting of Transportation Research Board*, 2004a.

17) Abdel-Aty, M., and Abdalla, M. F.: Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes using generalized estimating equations for correlated data, *83rd Annual Meeting of Transportation Research Board*, 2004b.

18) Lauritzen, S. L. The EM algorithm for graphical association models with missing data. Journal of Computational Statistics & Data Analysis, Vol. 19, 1995, pp. 191-201

19) Spiegelhalter, D. J., and L. Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. Journal of Networks, Vol. 20, No. 5, 1990, pp. 579-605.

20)

21) Luo, L., and Garber, N. J.: Freeway crash prediction based on real-time pattern changes in traffic flow characteristics, *A Research Project Report for the Intelligent Transportation Systems Implementation Center, UVA Center for Transportation Studies*. Research Report No. UVACTS-15-0-101, January, 2006.

22) Hossain, M. and Muromachi, Y.: Applicability of bayesian network in real-time crash prediction, *Proceedings of In-*

frastructure planning review, Japan Society of Civil Engineers (JSCE), vol. 38, 2008.

23) Hossain, M. and Muromachi, Y.: Development of a real-time crash prediction model for urban expressway, *Journal of the Eastern Asia Society for Transportation Studies, Vol.8*, 2009

24) Hossain, M. and Muromachi, Y.: A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways, *Accident Analysis and Prevention 45 (2012) 373–381*, 2011.

(Received April 22, 2016)