Route Safety Management through Real-time Crash Prediction Model

Ananya ROY[1], Ryosuke KOBAYASHI [2], Moinul HOSSAIN [3]
and Yasunori MUROMACHI [4]

[1]Graduate Student, Graduate School of Inter Disciplinary Science and Engineering, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail:roy.a.aa@m.titech.ac.jp
[2]Graduate Student, Graduate School of Inter Disciplinary Science and Engineering, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail: kobayashi.r.af@m.titech.ac.jp
[3]Adjunct Professor, Centre for Global Engineering, University of Toronto,
(200 College Street, Toronto, ON, M5S 3E5, Canada)
E-mail: moin.hossain@utoronto.ca
[4]Associate Professor, Graduate School of Inter Disciplinary Science and Engineering, Tokyo Institute of Technology
(Nagatsuta 4259, Yokohama, Kanagawa, 226-8502 Japan)
E-mail: ymuro@enveng.titech.ac.jp

Bayesian Network is becoming popular for building real-time crash prediction models day by day. This paper indicates the shortcomings of previous methods of crash prediction models and finds an advance way to overcome those. Success of a model largely depends on availability of adequate traffic data associated with crash data. Thus, high density traffic data (flow, speed, occupancy, detector location and other crash data) were collected from the upstream and downstream detectors closest to the location of crash, for 6 consecutive months, each minute for the entire route of Shinjuku of the busiest expressway under the jurisdiction of Tokyo Metropolitan Expressway Company Ltd., Japan. Then, we employ Bayesian Network for model building to predict future crash for each about 250 meter freeway segments through the route, within 5 minutes before the crash takes place. After that, the new model was evaluated by investigating future crash with available 1 month data. The performance validation shows that, the model can handle to predict up to 70% of future crash with 61% of normal traffic condition, if the threshold of average crash probability is set to 4.5%. Lastly, the result of prediction probability for a single day (24 hour) of the entire route is represented in a very comprehensive way for better visualization of the situation.

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

1. INTRODUCTION

Road crashes have become a global phenomenon with the increasing trend of urbanization. In order to predict future crash and their associated severities on a specific segment of a road over a certain period, several models have been built in last few decades. The main purpose of the models was to identify hazardous locations based on historical crash data related to traffic, road geometry and environmental condition [1, 2, 3]. However, these models were highly criticized because of their incapability to incorporate various factors as input variables. Sudden turbulence in traffic conditions may lead to a crash, which can't be detected with huge aggregated measures of traffic variables [4,5]. To address this problem, a new crash prediction model has been introduced where real-time traffic data are taken as input to calculate the instantaneous crash risk. Rapid improvement in Intelligent Transportation System (ITS) has made it easier to get real-time traffic data and increased the availability of real-time crash prediction models. Yet this conception of real-time prediction model is pretty new in transportation field and hasn't been implemented in practical situation. The main cause for that is lack of past real-time data in terms of both quality and quantity. In the first real-time crash prediction model, two different traffic dynamics were considered – disruptive and normal, and likelihood of future traffic flow data falling into either of these two categories was assessed [6]. 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]. The major drawback of their study was, they needed the density at the time of crash to predict crash – which is extremely difficult to obtain in real situation. 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]. Although the outcome of the study was not successful due to insufficient sample size of non-crash cases and spatial difference in the traffic flow. Another reason mentioned by the author was use of inaccurate reported crash time.

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.

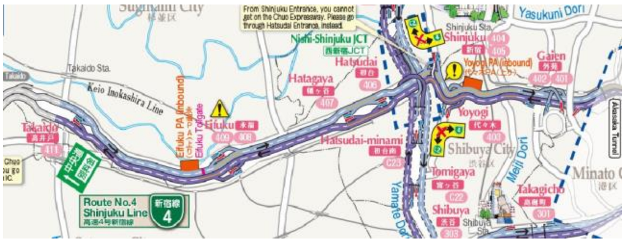
This paper proposes application of Bayesian Network (BN) to build a real-time crash prediction model considering robustness of the models from previous studies. BN is a probabilistic graphical modeling method that describes complex joint distributions of a system through a graph and local distributions, 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 calibrate itself with partially available data. It also supports sequential learning, which allows it to update itself in real-time whenever new data becomes available.

The paper is organized in four sections. This section has already addressed the problems with the present crash prediction models and indicated advantages of real-time crash prediction models over the past models. In the second section, the data collection is discussed. The third section deals with the model building and validation. Lastly, the forth 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 (Figure 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 45 detectors (about 250 meters apart) in one lane. In this paper, only inbound route is used for analysis. About 73 crash cases were reported for this direction.



Source: Official website of Tokyo Metropolitan Expressway Company Limited

Figure 1 Study area: Shinjuku 4 Tokyo Metropolitan Expressway, Japan

(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)

Table 1 Sample data for 21st August, 2014 15:26pm

parameters	values	upstream - downstream
upstream flow	49	-2
downstream flow	51	
upstream speed	66.67	0.2
downstream speed	66.49	
upstream occupancy	10.45	-0.8
downstream occupancy	11.25	

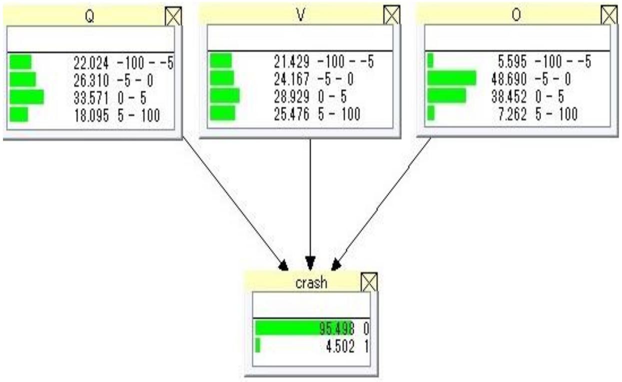


Figure 2 Bayesian Network for real- time crash prediction

and occupancy (Table 1). Crash data contained information about date, time (in minutes), location (to nearest 10 meters), crash lane, type of crash and vehicle involvement. It was assumed in this paper that the reported time is very close to the actual crash time.

Two types of data- crash and normal data are needed for the model. For crash data, 1 minute data, 5 minutes 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 02:54-2:55AM is considered as crash data. In case of selecting normal data, flow, speed and occupancy for 02:54- 2:55AM 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. Also, crash data with any missing data was ignored. After all these screening, there were a total of 67 nos of crash and 743 nos of normal data left for model building. Of the 67 nos crash data, 20 nos were used for model validation.

In order to develop a real-time crash prediction model, pairs of detectors- nearest upstream and nearest downstream of the crash location was con sidered for the entire route of the expressway. Then, difference of upstream and downstream flow, speed, occupancy were calculated. ‘1’ and ‘0’ were used to denote crash and normal condition.

3. MODEL BUILDING AND VALIDATION

(1) Model Building

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 this case, difference between upstream and downstream ‘diff_flow (Q)’, ‘diff_speed (V)’ and ‘diff_occupancy (O)’ are the information variables which we got from the detector database. On the other hand, hypothesis variable is the variable which is to be predicted, here, it is the probability of ‘crash (C)’ (Figure 2). It has dichotomous outcome – '0' for crash and '1' for non-crash. As the inter-relationship among flow, speed and occupancy is obscure, we didn’t draw any connection among these variables.

$$P(\text{Crash},\text{diff\_Speed},\text{diff\_Flow},\text{diff\_occuopancy}) = P(\text{diff\_Speed})P(\text{diff\_Flow})P(\text{diff\_occupancy})\,P(\text{Crash}|\ \text{diff\_Speed},\ \text{diff\_Flow},\ \text{diff\_occupancy}) \quad (1)$$

(2) Model Validation

The success of the model depend on its combined performance to predict crash and normal traffic conditions. Thus, for 20 nos of crash cases, both kind of data were selected in the same way mentioned in the pevious section. This model should be able to predict future crashes which will occur in the time periods of corresponding days we used while building the model. Any normal condition data corresponding to 20 nos trainging crash cases, which didn’t fall within the data set of 5 minutes prior corresponding to 47 nos crash cases, were discarded. After this, with the help of the previously built model in Hugin Expert 7.2, probability of future crash and normal condition was calculated.

Next, a threshold value is needed to calculate the success of the model based on the following equations:

$$\text{Crash} = (\text{Calculated probability over threshold/ crash sample size})*100$$
$$\text{Normal traffic} = (\text{Calculated probability below threshold/ normal sample size})*100 \quad (2)$$

Higher threshold value will reduce the model's success for crash prediction and will reduce false alarms too. The mean probability of crash was found to be 17% (for the 20 cases) and the average crash probability was found 4.5% (from Hugin Expert). If the threshold is set to 1.5% which is less than the average crash probability, the model could predict 80% crashes and 35% of normal traffic condition. If the threshold value is set to 5.5% (higher than the average crash probability), then the model can predict 65% crashes and 70% of normal conditions (Table 2). However, setting a threshold value requires expert opinion and knowledge. Experts can set the threshold value according to their priority and experience, such as cost associated with false alarm, or cost related with unable to predict a future crash. The threshold may subject to change through different time of day.

Table 2 Crash prediction result of model built with Hugin

Threshold of avg. crash probability	Crash(%)	Non-Crash(%)
1.5%	80.0	34.8
4.5%	70.0	60.9
5.5%	65.0	69.6
7.5%	60.0	73.9
10.5%	55.0	79.1

The result of the model for 24-hour dataset for the day of April 1st, 2014 is presented in Figure 3. The color gradient from light to dark was used to represent occurance of crash (darker color) and normal traffic condition (light color).

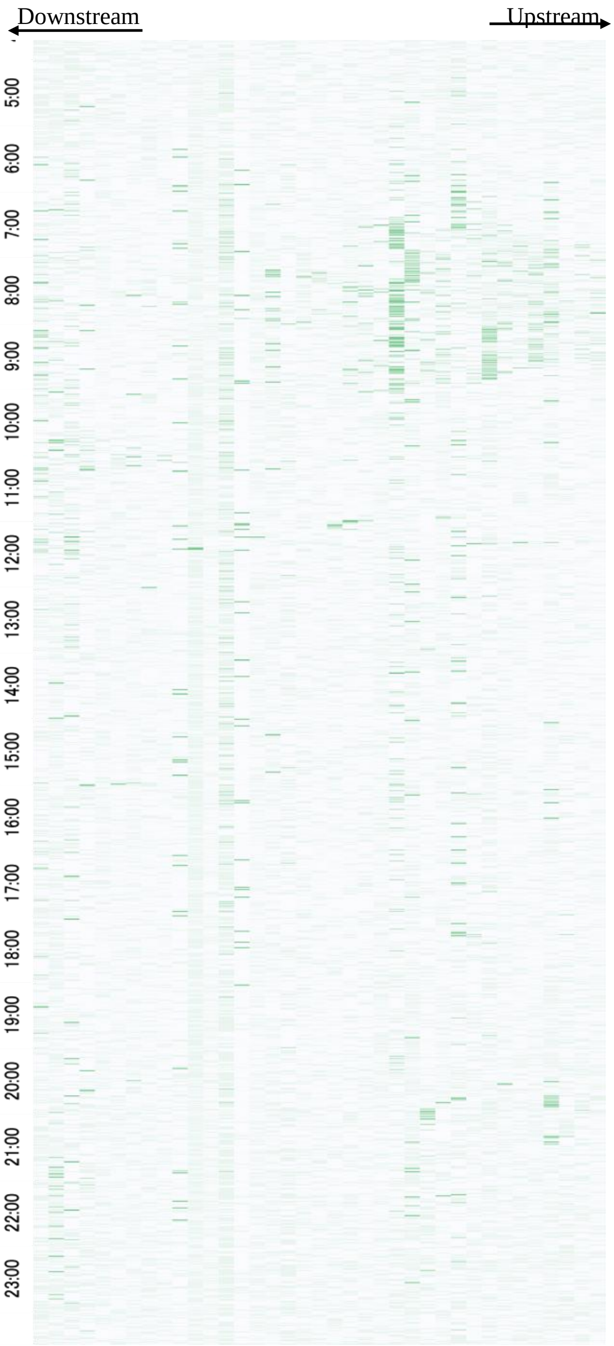


Figure 3 24- hour probability of crash and normal traffic condition for 1st April, 2014 (Shinjuku route 4)

4. CONCLUSIONS

This paper presented an approach for real-time crash prediction. At first, it addressed the lackings regarding previous classic non-real time crash prediction methods. Then, Bayesian Network a relatively new concept in transportation field was followed. Moreover, it dealt with a huge volume of quality data, which is a prerequisite for a good prediction model. After explaining the data collection process, modeling method was presented step by step. The study not only built a model, but also, demonstrated the way of evaluating the model for different threshold values. It was inferred, from the paper that the model is able to predict 70% crashes and 61% normal traffic conditions at a threshold of 4.5%. Model built with BN has advantage over the conventional statistical models of high flexibility; it can update by itself with new addition of dataset. Additionally, traffic flow variables can be added or subtracted anytime and the model could be easily modified. This model was built for Tokyo Metropolitan Expressway, with sufficient data and futher study, it could be implemented in other roads and highways as well.

Apart from this, road geometry, type of collision, weather condition were not considered in this paper. These might contribute a significant role in model building. This is a future scope of this study.

No specific intervention was proposed in this paper as it depends on the transport authority, their priorities and thorough investigation. Counter measures could be taken by indentifying the black spots, and putting signs, ramp metering, restricting speed limits etc. based on the developed model.

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.

REFERENCES

1) 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.
2) 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.
3) 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.
4) 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.
5) 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.
6) Oh, C., Oh, J., Ritchie, S., and Chang, M.: Real time estimation of freeway accident likelihood, *80th Annual Meeting of Transportation Research Board*, 2001.
7) 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.
8) Lee. C., Saccomanno, F., and Hellinga, B.: Analysis of crash precursors on instrumented freeways, *81st Annual Meeting of Transportation Research Board*, 2002.
9) 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.
10) 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.
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, *83rd Annual Meeting of Transportation Research Board*, 2004b.
12) 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.
13) Hossain, M. and Muromachi, Y.: Applicability of bayesian network in real-time crash prediction, *Proceedings of Infrastructure planning review, Japan Society of Civil Engineers (JSCE), vol. 38*, 2008.
14) 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
15) 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 24, 2015)