APPLICABILITY OF BAYESIAN NETWORK IN REAL-TIME CRASH PREDICTION*

By Moinul HOSSAIN** and Yasunori MUROMACHI***

1. Introduction

The attempts to predict crashes on freeways through statistical modeling involving capacity driven measures of traffic flow (e.g., AADT) and road geometry have spanned for more than two decades. However, success in crash prediction involving these static data has so far been limited. In recent times, some researchers made efforts to accommodate the weather conditions and seasonal effects to better predict crashes through time series analysis. So far, these models have shown their incapability to accommodate the ever complex human factors, which, to a great extent, are believed to be directly or indirectly responsible for most of the crash cases. One of the latest additions to the crash prediction modeling is the approach to predict crashes based on real-time traffic data which considers the human factors in macroscopic level from the traffic engineering perspective. The concept of real-time crash prediction modeling is gaining momentum due to its proactive nature of application and the growing implementation of ITS, which ensures the future availability of real-time traffic data. However, being at its primitive stage and due to scarcity of past real-time data, the present models are yet theoretical and largely prone to unrealistic data requirements and lack of reliability. This paper commences with enumerating the recent attempts of predicting crashes in real-time, their findings and drawbacks. Later, the paper emphasizes on the competitive edges of Bayesian Network as a method to be used in predicting crashes in real-time and presents a methodology for that. Lastly, the paper explains how the outcome of a real-time crash prediction model can be used for initiating proactive safety measures.

2. Real-time Crash Prediction - Progress Till Date

Real-time crash prediction modeling is a concept younger than a decade in the road safety arena. This section briefly explains the initiation and onward progress in real-time crash prediction by summarizing five prominent research studies in this area.

The study by Oh et al.[1)] was the first of its kind to assess the possibility to classify a traffic condition as accident prone to estimate the likelihood of potential crashes. In that study, they separated traffic dynamics into two categories – disruptive and normal. In a recent update of the same study (Oh et al.[2)]) they defined the same terms as hazardous and normal traffic conditions. A hazardous traffic condition is defined as that potentially leading to an accident occurrence whereas a normal traffic condition does not instigate accidents. This subsection focuses on their two latest publications (Oh et al.[2), 3)]) as they present the latest state of their research study. Between these two studies, in the first study (Oh et al.[2)]) they employed a nonparametric Bayesian approach to identify the real-time crash likelihood. In the later study (Oh et al. 2005[3)]) they applied Probabilistic Neural Network (PNN) to accomplish their estimation purpose. Normal condition was then specifically defined as a 5 minute period occurring at 30 minutes prior to the crash and the disrupted condition was defined as the 5 minute time period just before the crash. A total of 52 crashes had adequate real-time traffic data to be matched with. Later on, they employed t-test on the mean and deviation of three variables – occupancy, flow and speed, to identify the crash indicators. However, there was no suggestion explaining if they have tested the data for normality as t-test is applicable only with the assumption that the data follow normal distribution. Oh et al.[1)—3)] identified standard deviation of speed to be the most significant variable although most of the variables came out as significant in the t-test. Later, they defined the two traffic conditions with two probability density functions (PDF) and used those to identify the posterior probability of a traffic condition to belong to either of these two traffic conditions and determine crash probability thereby.

*Keywords: Crash Prediction, Real-time, Turbulent Traffic Condition, Bayesian Network

**Student Member of JSCE, M. Engg.., Dept. of Built Environment, Tokyo Institute of Technology (Nagatsuta-machi 4259, Midori-ku, Yokohama-shi, Kanagawa,226-8502 Japan, Telefax: +81-(0)45-924-5524, E-mail address: moinul048i@yahoo.com)

***Member of JSCE, Ph.D., Dept. of Built Environment, Tokyo Institute of Technology (Nagatsuta-machi 4259, Midori-ku, Yokohama-shi, Kanagawa, 226-8502 Japan, Telefax: +81-(0)45-924-5524)

The motivation for Golob et al.[4] to develop a real-time crash prediction model exuded from the thirst to understand the complex relationship between traffic flow and traffic crashes rather than to develop a proactive highway safety system. The final outcome of the research study was a software tool named FITS (Flow Impacts on Traffic Safety) which takes data stream of 30-second observations sensed through a single inductance loop detector as input and envisage the type of crashes likely to occur based on the monitored traffic flow condition. Golob et al.'s[4] study results were limited within a specific weather condition and that was, dry roadways during daylight and dusk-dawn conditions. Four statistical measures (median of the ratio of volume to occupancy, difference of the 90^{th} percentile and the 50^{th} percentile in the ratio of volume to occupancy, mean volume over the entire 27.5 minutes period preceding the accident and standard deviation of 30-second volumes) for three types of lanes (left, interior and right) resulted in 12 variables which were used to apply Principal Component Analysis (PCA) to develop the model. Later, 6 out of 12 variables were found to explain 86.8% of the variation in traffic conditions. They performed K-means Clustering analysis in conjunction with Non-linear Correlation Analysis and detected eight traffic flow clusters explaining specific crash types in a sunny day with dry road condition. It was expected that further development of the model may become successful in forecasting crash rates in terms of vehicle miles of travel, for vehicles exposed to different traffic conditions. The greatest criticism of the model by Golob et al.[4] was its unrealistic data requirements.

Lee et al.[5]–[6] conducted two studies to predict crash risk in real time where the second study basically reduced the number of assumptions made in the first study to make it more acceptable. In their latest model they selected speed variations along a lane, traffic queue and traffic density at given road geometry, weather condition and time of the day as predictors and applied aggregated first order log-linear model to predict crash. Unlike the other studies, Lee et al. took extra care in correcting the actual time of crash occurrence by studying the speed profile in the upstream and the downstream with time. One major drawback with Lee's model to be implemented in real-time is that it requires the density at the time of crash to predict crash. Thus the model is applicable in crash detection rather than in crash prediction.

The initial study by Abdel-Aty et al.[7] concentrated on classifying speed patterns to predict crashes in real-time. Their later study (Abdel-Aty et al.[8]) incorporated road geometry into the model using Generalized Estimating Equations for correlated data. This research was the first of its kind where a series of crash and non-crash traffic classification analysis were conducted to identify the predictors. They used Probabilistic Neural Network method to separate traffic patterns leading to crash from those not leading to crash. The variables found significant in their model were mean and variance of volume, occupancy and speed. Later on, they employed Generalized Estimating Equations to predict crashes. The study further calculated the false alarm rate, too. It was suspected that Abdel-Aty et al.'s model may generate high false alarm as they considered the traffic variables independently ignoring their interactions.

The research project by Luo and Garber[9] is the latest addition in the attempt to predict crashes in real-time. They commenced their research with intensive investigation of previous research studies and tried not to repeat their mistakes. Luo and Garber analyzed each crash case independently to identify crash leading patterns and factors describing the patterns. Like Abdel-Aty et al., they also concluded mean and variance of occupancy, speed and volume to be suitable for predicting crashes in real-time. However, they limited their study within identifying traffic patterns distinguishing crash and non-crash situations. They applied three pattern recognition methods: K-means clustering method, Naive-Bayes method and Discriminant Analysis and compared the outcomes. However, all the methods failed to associate a certain traffic pattern with crash with 50% or more accuracy. The Naive-Bayes method demonstrated better prediction capability than the other two methods.

From the literature review, it can be fathomed that predicting crashes in real-time is in its infancy. Two of the most important drawbacks of the present models are the unrealistic data requirements and the appropriateness of the method. Another problem associated with the present models is their stage of the research. To elaborate this, the study by Golob et al.[4] is in the stage of predicting the pattern of crash where the study by Luo et al.[9] is investigating the predictors that can define a crash with high reliability. These studies can be considered to be in a very early stage to predict crashes in real-time. Lee et al.[6] requires real-time density information as an input to their model, which is unrealistic in case of practical implementation. The accuracy of Oh et al.'s [3] model heavily relies on its accuracy to distinguish between disrupted traffic condition and the normal traffic condition. However, this was achieved through a simple t-test without shedding enough light on the normal distribution assumption of the data. The model by Abdel-Aty et al.[8] stands out to be the most acceptable among the present models, however, it lacks in explaining and considering the internal relationships among its precursors. One more criticism of this model can be the use of Probabilistic Neural Network (PNN) at such an early stage of the research avenue as the researchers remain unaware of the interpretations of the hidden layers in a PNN. Moreover, PNN requires the 'weights' and 'thresholds' to be known to model the system, which, again, may be too much optimistic at such an early stage. Finally, modeling crashes in real-time demands the use of a substantial amount of subjective probabilities which in course of time require to be updated. However, the route of inference is fixed in the construction of a PNN. The prior information and the information that is expected to be calculated by the system are fixed prior to modeling, which makes the method substantially inflexible considering its stage of development.

This paper suggests the development of a new model considering easily obtainable predictors and using Bayesian Network (BN) as the modeling method. The study by Luo et al.[9] used Naive-Bayesian method, which is the simplest form of a Bayesian Network (BN) where interactions among variables are not considered, exhibited comparatively better prediction capability. BN is highly applicable in situations where the belief regarding events may change frequently and the system needs to capitalize on the benefits induced by the new information. Moreover, BN represents the relationships among the variables in graphical manner which is quite important for a model which is in its early stage. BN is a well understood method in the arena of Artificial Intelligence (AI) which did not receive much attention as it requires huge computing power. However, the recent developments in the processing power of the computers have revived the use of BN in various disciplines. Recently, BN is becoming a prominent choice in developing military intelligence, testing the effectiveness of medicines, identifying the reliability of systems, etc.

A comparative standing of the present real-time crash prediction models and the proposed model is presented in Figure 1 with reference to three factors – the applicability of the method, suitability of the predictors and the stage of the development.

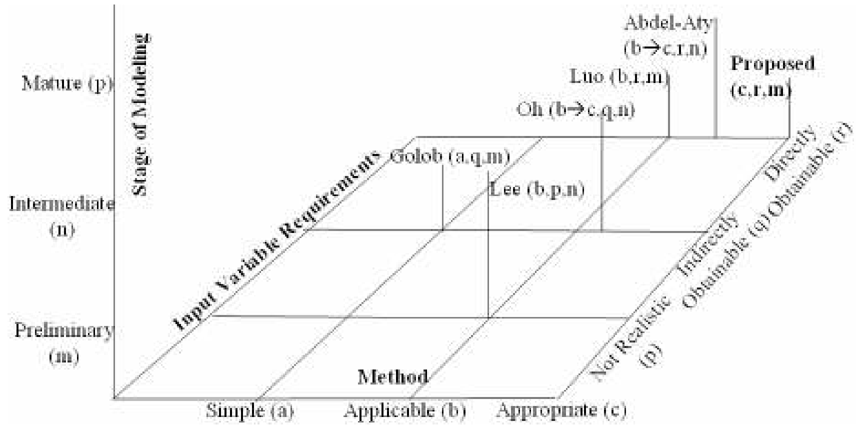

Figure 1. Comparative standing of the Models

3. The Bayesian Network (BN) Approach to Predict Crashes in Real-time

The first step in BN involves identifying a hypothesis variable where each of its states (mutually exclusive and collectively exhaustive) will be tested. Afterwards, it is important to identify the predictors of the hypothesis variable, normally referred as information variable in BN. In a simple case of crash prediction model, the hypothesis variable can be 'Crash' with two states – 'Yes' and 'No'. This is a dichotomous variable providing information regarding the instantaneous crash risk based on aggregated predictor data. Considering the previous research studies on real-time crash prediction, three information variables are selected for the model namely variation of 'Flow', 'Speed' and 'Occupancy' – which can be easily obtained from the basic outputs of most of the real-time traffic data collection equipment. Later on, the variables can be discretized into suitable number of categories after aggregating them into an interval 't'. Therefore, the final variables can be represented as Flow ($Fl\{x_1, x_1, \ldots, x_m\}$), Speed ($Sp\{x_1, x_1, \ldots, x_n\}$) and Occupancy ($Occ\{x_1, x_1, \ldots, x_p\}$). The second step in modeling is to establish direct links among the variables to form a causal network. Now, it is expected that variation of flow will have impacts on speed and occupancy and vice versa. A common way to solve this problem is by introducing an intermediate variable called turbulent traffic condition ('TTC' with values 'yes' or 'no'), known as mediating variable, in between the hypothesis variable and the information variables. In that case, even if the hypothesis variable is instantiated, i.e., information about it is known, information can flow among the information variables through the mediating variable. Figure 2 presents a simple diagram of BN for predicting crash in real-time. Afterwards, the crash probability when information regarding the variables Fl, Sp and Occ are known can be obtained from Equation 1.

$$P(Cr \mid Fl, Sp, Occ) = \frac{\sum_{TTC} P(Cr, Fl, Sp, Occ)}{P(Fl, Sp, Occ)} \quad \text{(1)}$$

Now, $\sum_{TTC} P(Cr, Fl, Sp, Occ)$ can be calculated using the Chain rule of BN theorem and marginalizing the values for TTC:

$$\sum_{TTC} P(Cr, Fl, Sp, Occ) = P(Cr)P(TTC)P(Fl \mid TTC)P(Sp \mid TTC)P(Occ \mid TTC) \quad \text{(2)}$$

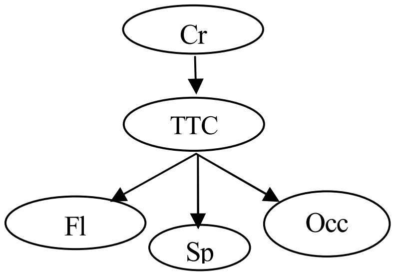

Figure2. A Model for Real-time Crash Prediction.

The values of P(Fl|TTC), P(Sp|TTC) and P(Occ|TTC) will be three matrices with dimensions 'm by 2', 'n by 2' and 'p by 2' respectively. The values of probability distributions of all these variables can be obtained from collecting real-time traffic data along with crash data and separating the for turbulent and non-turbulent traffic conditions associated with crash and non-crash situations.

4. Conclusion and Recommendation

The outcome of the real-time crash prediction model can be utilized in conjunction with the variable message sign (VMS) by providing instantaneous safety condition on the road to the users. Moreover, VMS can also be used to control the variables: Flow, Speed and Occupancy through various methods such as variable speed limit, restriction on lane changing, controlling the flow at toll plazas, etc. Moreover, when the technology attains maturity, it can be used in internalizing the external cost of road crash accommodating it into the toll price.

The paper provided a brief overview on the present attempts to predict road crashes in real-time and demonstrated why and how Bayesian Network can be a potential method to overcome the existing prediction hindrances. This is a part of an on-going research study being conducted by the authors. The next step in the study is to collect real-time traffic data and the crash data to develop a prototype model and calibrate it as necessary.

References

1) Oh, C., Oh, J., Ritchie, S., and Chang, M. (2001). Real Time Estimation of Freeway Accident Likelihood. 80th Annual Meeting of Transportation Research Board, 2001.

2) Oh, J., Oh, C., Ritchie, S., and Chang, M. (2005a). Real Time Estimation of Accident Likelihood for Safety Enhancement. ASCE Journal of Transportation Engineering, Vol. 131, No. 5, pp 358-363.

3) Oh, C., Oh, J., Ritchie, S., and Chang, M. (2005b). Real Time Hazardous Traffic Condition Warning System: Framework and Evaluation, IEEE Transactions on Intelligent Transportation Systems, Vol. 6. No.3, pp 265-272.

4) Golob, Thomas F., Golob, Will Recker, Alvarez, Veronica M. (2003). A Tool to Evaluate the Safety Effects of Changes in Freeway Traffic Flow. 82nd Annual Meeting of Transportation Research Board, 2003.

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

6) Lee. C., Saccomanno, F., and Hellinga, B. (2003). Real-Time Crash Prediction Model for the Application to Crash Prevention in Freeway Traffic. 82nd Annual Meeting of Transportation Research Board, 2003.

7) Abdel-Aty, Mohamed and Pande, Anurag. (2003a). Classification of Real-time Traffic Speed Patterns to Prediction Crashes on Freeways. 83rd Annual Meeting of Transportation Research Board, 2003.

8) Abdel-Aty, M., and Abdalla, M. F. (2003b). 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, 2003.

9) Luo, L., and Garber, N. J. (2006). 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.