Transportation Research Part C 54 (2015) 176–186

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data

Jie Sun, Jian Sun *

Department of Traffic Engineering & Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

A R T I C L E I N F O

Article history:
Received 31 October 2014
Received in revised form 4 March 2015
Accepted 4 March 2015
Available online 1 April 2015

Keywords:
Urban expressway
Real-time crash prediction
Dynamic Bayesian network
Traffic states
Speed conditions data

A B S T R A C T

Traffic crashes occurring on freeways/expressways are considered to relate closely to previous traffic conditions, which are time-varying. Meanwhile, most studies use volume/occupancy/speed parameters to predict the likelihood of crashes, which are invalid for roads where the traffic conditions are estimated using speed data extracted from sampled floating cars or smart phones. Therefore, a dynamic Bayesian network (DBN) model of time sequence traffic data has been proposed to investigate the relationship between crash occurrence and dynamic speed condition data. Moreover, the traffic conditions near the crash site were identified as several state combinations according to the level of congestion and included in the DBN model. Based on 551 crashes and corresponding speed information collected on expressways in Shanghai, China, DBN models were built with time series speed condition data and different state combinations. A comparative analysis of the DBN model using flow detector data and a static Bayesian network model was also conducted. The results show that, with only speed condition data and nine traffic state combinations, the DBN model can achieve a crash prediction accuracy of 76.4% with a false alarm rate of 23.7%. In addition, the results of transferability testing imply that the DBN models are applicable to other similar expressways with 67.0% crash prediction accuracy.

1. Introduction

Traffic data collection has become more convenient and efficient with the development of advanced transportation information systems (ATIS). Thus, a number of studies assessing the real-time risk of traffic flow operation on freeways and urban expressways have been conducted using the traffic data collected from the ATIS, primarily the fixed-point data (e.g., loop detector, microwave radar and automatic vehicle identification (AVI) data). Consequently, numerous proactive crash prediction models have been developed (Oh et al., 2001; Lee et al., 2003; Abdel-Aty et al., 2004, 2005, 2012; Abdel-Aty and Pande, 2005; Pande et al., 2005, 2011, 2012; Pande and Abdel-Aty, 2006; Hossain and Muromachi, 2012, 2013b; Ahmed and Abdel-Aty, 2012; Ahmed et al., 2012; Golob et al., 2008; Zheng et al., 2010; Xu et al., 2013, 2014a, 2014b; Sun et al., 2014a). These models can distinguish crash-prone traffic conditions from normal conditions, which can be applied in safety promotion strategies such as proactive warning with in-car devices or variable message signs (VMS) and other traffic flow smoothing management strategies such as speed harmonization (Lee et al., 2004; Abdel-Aty et al., 2006; Allaby et al., 2007) to avoid crashes or decrease the likelihood of crashes.

* Corresponding author. Tel.: +86 21 69583650.
E-mail address: sunjian@tongji.edu.cn (J. Sun).

http://dx.doi.org/10.1016/j.trc.2015.03.006
0968-090X/

In general, numerous variables including traffic flow data (i.e., volume, speed and occupancy) and their combinations collected from detectors, road geometry alignment and environmental parameters were used in the development of a real-time crash prediction model. In this paper, regarding roads that lack of fixed point traffic flow detector data, the feasibility of predicting real-time crash risk utilizing only traffic speed condition data, which may be obtained from sampled floating cars or smart phones, was investigated. However, due to the lack of floating car data or smart phone data, the speed data collected from the dual-loop detectors were used as speed condition data instead. If this approach is successful, it can be deduced that the speed data collected via other methods are also effective for crash prediction. Moreover, computational complexity and over-fitting issues caused by redundant variables can be avoided by using relatively fewer speed condition variables.

Although several previous studies have adopted speed-only data collected from AVI sensors or loop detectors for crash prediction (Ahmed and Abdel-Aty, 2012; Li et al., 2013), it still makes sense to consider speed condition data for crash prediction with other concerns, as the traffic state before a crash is an essential factor in developing crash prediction models. It has been indicated that for different combinations of upstream and downstream traffic states, the crash involvement rates and crash risk ratios (ratio of crash cases and non-crash cases) are inconsistent (Yeo et al., 2013; Hossain and Muromachi, 2013a). When different traffic states were considered in previous studies, separated models were usually built under different levels of traffic conditions (Abdel-Aty et al., 2005; Xu et al., 2013, 2014b; Li et al., 2013). However, models employing the value of different traffic state combinations as input parameters might be more efficient and less complex. Thus, two types of state determining approaches were used and compared in this study to identify the better approach.

Regarding the correspondence of traffic data to crash data, there are several time interval data computed with 5-min aggregation that related to the crash occurrence (Pande et al., 2012; Sun et al., 2014a). Generally, only one time interval of traffic data was used in one model for real-time crash prediction, and the time interval (5–10 min before crash) traffic data show the most significant relationship with crashes (Abdel-Aty et al., 2004; Xu et al., 2013, 2014a). However, considering that a crash can be induced by the disturbance of traffic flow before the crash occurs, time series traffic data consisting of several time intervals should be used to illustrate the dynamic process of traffic flow before crash occurrence. Thus, it is essential to establish a single model that can address such time series data and the evolving process of traffic flow.

To address the above issues, a dynamic Bayesian network (DBN) model that can handle time series data was proposed in this study. First, with speed condition data collected in several time intervals, the congestion levels upstream and downstream of the crash location were determined and considered as the explanatory variables of the prediction model. Then, with a matched case-control dataset that includes speed data corresponding to 551 crash cases and 2755 matched non-crash cases collected from two expressways in Shanghai, the DBN models were calibrated and evaluated. Meanwhile, both the DBN models with traffic flow detector data (i.e., volume, speed and occupancy) and a static Bayesian model with speed data were built for comparison purposes. Finally, the transferability of DBN models was also tested to examine the ability to implement it directly on other expressways.

2. Data sources

2.1. Study area and crash data

The real-time crash prediction model established in this study aims to investigate the relationship of traffic flow characteristics and crash risk on urban expressways. Thus, to establish the crash prediction model, crash data and corresponding traffic data should both be collected. In this study, crash data and traffic data were collected from 3 segments on the Yan-an expressway and 3 segments on the North–South expressway in Shanghai, China, which have similar road geometry and on/off-ramp arrangement. Thus, the road geometries are not additional influencing factors on the crash prediction model. All of these sites are three-lane expressway segments, with detectors spaced at approximately 300–500 m. Considering the different quantities of crashes on the two expressways, the 411 crash cases collected on the Yan-an expressway were used for the training and testing of the crash prediction model because of the larger number of crashes which occurred on this road, whereas the 140 crash cases collected on the North–South expressway were used to examine the transferability of the crash prediction model. The sites and crash statistics of segments are presented in Table 1.

Table 1
Summary of study sites and crash data.

Expressway	Segment on expressway	Length/km	Number of detectors	Number of crashes	% of total crashes
Westbound Yan-an expressway	Yandong Interchange to Maoming Road	0.8	4	103	18.7
Eastbound Yan-an expressway	Hongxu Road to Loushanguan Road	1.3	5	202	36.7
	Jiangsu Road to Huashan Road	1.5	6	106	19.2
Northbound North–South expressway	Yanchang Road to Guangzhong Road	1.8	6	28	5.1
Southbound North–South expressway	Guangzhong Road to Luochuang Road	2.0	7	52	9.4
	Gongjiang Road to Changzhong Road	2.2	7	60	10.9
Total		n/a	n/a	551	100