Accident Analysis and Prevention 94 (2016) 59–64

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Utilizing the eigenvectors of freeway loop data spatiotemporal schematic for real time crash prediction

Shou'en Fang[a], Wenjing Xie[a], Junhua Wang[a,*], David R. Ragland[b]

[a] School of Transportation Engineering, Tongji University, 4800, Cao'an Highway, Shanghai, China
[b] School of Public Health, UC Berkeley, Safe Transportation Research and Education Center (SafeTREC), University of California, Berkeley, United States

ARTICLE INFO

Article history:
Received 3 September 2015
Received in revised form 7 March 2016
Accepted 12 May 2016
Available online 31 May 2016

Keywords:
Real time accident model
Crash precursor
Spatiotemporal schematics
Road safety

ABSTRACT

The concept of crash precursor identification is gaining more practicality due to the recent advancements in Advanced Transportation Management and Information Systems. Investigating the shortcomings of the existing models, this paper proposes a new method to model the real time crash likelihood based on loop data through schematic eigenvectors. Firstly, traffic volume, occupancy and density spatiotemporal schematics in certain duration before an accident occurrence were constructed to describe the traffic flow status. Secondly, eigenvectors and eigenvalues of the spatiotemporal schematics were extracted to represent traffic volume, occupancy and density situation before the crash occurrence. Thirdly, by setting the vectors in crash time as case and those at crash free time as control, a logistic model is constructed to identify the crash precursors. Results show that both the eigenvectors and eigenvalues can significantly impact the accident likelihood compared to the previous study, the proposed model has the advantage of avoiding multicollinearity, better reflection of the overall traffic flow status before the crash, and improving missing data problem of loop detectors.

1. Introduction

Recent advances in Intelligent Transportation System (ITS) allow traffic safety studies to extend from historic data-based analyses to real-time applications. This study employs the real-time capability of Advanced Traffic Management and Information Systems (ATMIS) for safety enhancement. In the last 20 years, researchers have developed crash prediction models to relate crash risks to some real-time traffic flow parameters collected from loop detectors such as traffic occupancy and vehicle speed variances. Madanat and Liu (1995) proposed the concept of "Real-time Incident Likelihood Prediction" to analyze the crash similarity by modelling the crash data on freeway with the detected flow data and weather data. Hughes and Council (1999) were the first to use loop detector data to explore the relationship between freeway safety and operations during peak periods. They proposed that real time detected speed variance was significant in crash occurrence and macroscopic measures were not suitable for real time safety analysis.

Since then, quite a few research studies followed Oh et al. (2001) employed a Non-Parametric Bayesian Classification Model to estimate accident likelihood. This study demonstrated that speed variation detected 5 min before accident could be used to estimate an accident effectively by using loop detector data and by setting different threshold level different percentage of accidents can be identified. Lee et al. (2002) proposed the concept "Crash Precursor" for traffic flow characteristics observed prior to crash occurrence. An aggregate log-linear model was developed which proved that variation of speed and traffic density were statistically significant predictors of crash frequency. In the coming year, Lee et al. (2003) developed a probabilistic real-time crash prediction model with a rational method to select crash precursors and optimal observation time slice durations. The study found that the difference between the speed at the upstream detector and the speed at the downstream detector was significantly higher.

Abdel-Aty et al. (2004) proposed that 5 min speed data before the accident was too short to be applied to real time traffic management. They utilized data from 7 loop stations which were closest to the accident site. Meanwhile, 6 periods of 5 min data of each loop before the crash were collected and case-control logistic regression method was used, and the model could achieve more than 69% crash identification. Abdel-Aty et al. (2005) and Abdel-Aty and Pande (2005) applied Bayesian classifier based methodology and probabilistic neural network (PNN) to predict crash likelihood. This study suggested that coefficient of variation measured in 5 min time slices of 10–15 min prior to the crash time in some upstream loops

* Corresponding author.
E-mail addresses: fangsek@tongji.edu.cn (S. Fang), 100931@tongji.edu.cn (W. Xie), benwjh@163.com (J. Wang), davidr@berkeley.edu (D.R. Ragland).

http://dx.doi.org/10.1016/j.aap.2016.05.013

affected crash occurrence most significantly. Abdel-Aty et al. (2005) suggested that utilizing hazard ratio values about 15 min before the accident occurrence could effectively reduce the impending risk. His method analyzed the flow data 90 min before and after the accident occurrence for four loop detectors upstream and two loop detectors downstream. Pande and Abdel-Aty (2006) also used Kohonen clustering algorithm, classification tree, multilayer perceptron (MLP), and normalized radial basis function (NRBF) neural networks to achieve 75% prediction of rear-end crashes.

Golob et al. (2004) used cluster analyses to prove that mean volume, median speed and temporal variations in volume and speed are the key traffic flow elements affecting safety. Golob et al. (2008) also tried to use a single loop data to evaluate safety performance. He converted the 36 redundant traffic flow parameters into 8 factors to develop a logit model for accident severity, type of collision, collision lane and number of vehicles involved. Oh et al. (2005) employed a Nonparametric Bayesian estimated accident likelihood, although they used only 52 accident samples and 5 min standard deviation of speed as indicators. Xu et al. (2012) divided the traffic flow into 5 states, and evaluated the safety performance associated with each state. This helps the safety researches better understand the traffic flow state before the crash. Vlahogianni et al. (2012) also developed neural network models by analyzing different variables associated with secondary accident likelihood and suggested that traffic speed/duration of the primary accident, hourly volume, rainfall intensity and number of vehicles involved in the primary accident are the top associated factors. Hossain and Muromachi (2012) pointed out some of the major shortcomings of the previous models including location of detectors, variable space and modelling methods which result in the implemented scenario being impractical. And he addressed the aforementioned shortcomings by proposing a Bayesian belief net based framework to develop real-time crash prediction models.

AVI (Automatic Vehicle Identification) data had been utilized in trial to analyze the crash potential in recent years (Ahmed et al., 2012; Ahmed and Abdel-Aty, 2012). These studies found that AVI systems could provide a measure of the risk of a crash in real time, with the accuracy 75.93% for rear-end crashes. Moreover, when fusing the AVI data and RTMS (Remote Traffic Microwave Sensors) data, the model with the data fusion framework had a higher estimation accuracy, robustness and reliability.

Meanwhile, transferability and robustness of real-time freeway crash risk assessment were emphasized in recent works (Shew et al., 2013). Xu et al. (2014) developed a Bayesian updating approach with data from the I-880N freeway in 2002 and 2009 and from the I-5N freeway in 2009 to improve transferability, and found it effective.

Generally, these findings point to the potential use of detector data and crash data in the field of traffic safety. However, none of these models have been implemented in practical scenario so far, which may result from the some of the major shortcomings of the existing models as follows.

(1) Most of the previous research took 5 min volume, speed and occupancy data as variable in Logit model. However, the upstream loop data in the earlier period and downstream loop data in the later period are usually correlated. The correlation coefficients between loops next to each other are shown in Table 1. High correlation between nearby loop data was found. Also, as the time slice gets longer, the relativity gets severer. We believe that the correlation between parameters influent the models estimates. This can cause multicollinearity.
(2) Large bunch of loop data in the series stations fail to explain the likelihood while just some of the loop data variables are significant which is too abstract to directly reflect the traffic flow status. Thus it is hard to be applied in the real traffic management.
(3) Most of the accident precursor study utilized 6 loop detectors (4 loops upstream and 2 loops downstream) or 7 loop detectors (the nearest detector to the accident with the other 4 loops upstream and 2 loops downstream) (Abdel-Aty et al., 2005). This kind of dataset is not necessary the best choice to reflect traffic follow characteristics as accident precursors since none of the research showed that the speed, volume and occupancy of the 1st and 6th loop significantly impacted crash likelihood.
(4) Most of the previous study utilized the single slice time period data before the accident occurrence as the precursor in modelling. While the crash record time in the accident database is always not so accurate that the significant variable in certain time slice in one location may not reflect the real traffic situation before the accident happens.

This study addresses the aforementioned shortcomings by proposing a freeway loop data schematic eigenvectors based framework to develop real-time crash prediction models. Firstly, traffic volume, occupancy and density spatiotemporal schematics in certain duration before accident occurrence were constructed to describe the traffic flow status. Secondly, the eigenvectors and eigenvalues of the spatiotemporal schematics were extracted to represent traffic volume, occupancy and density situation before the crash occurrence. Thirdly, by setting the vectors in crash time as case and those at crash free time as control, a logistic model is constructed to identify the crash precursors.

2. Methodology

2.1. Step1 construction of spatiotemporal loop data schematic

Spatiotemporal loop data schematic is proposed to contain the traffic flow parameters of the turbulence stage prior to the crash time. In this study, a 6×6 matrix which focuses on both upstream and downstream section of crash location and contains 6 time slices (5 min or 2 min each) is proposed. Tables 2 and 3 show examples of spatiotemporal traffic flow schematics for 5-min data and 2-min data. Due to the eigenvector methodology proposed in this study, the spatio-temporal matrix should be square matrix. Then the time period cover by the matrix is depended on the number of loop detectors.

Abdel-Aty et al. (2004) suggested that it was not enough to take 5 min ahead of the accident if it came to real-time traffic management. In this research, 10 min was taken as the time offset prior to the accident occurrence. Therefore the time duration in the 6×6 schematic is 10–40 min before the crash. Additionally, it should be mentioned that loop detectors In California are installed approximately between 500 m and 800 m apart.

2.2. Step2 extraction of matrix eigenvectors and eigenvalues from the schematics

In linear algebra, an eigenvector of a square matrix is a vector that points in a direction which is invariant under the associated linear transformation. As eigenvectors can represent the character of the matrix, it is widely used as pattern recognition tool in numerous fields. Taking image recognition as a typical example, eigenvectors are used to represent the image pixel matrix, which is similar to the proposed method in this paper as follows. Eigenvectors of a square matrix represent the space characters in different directions. And eigenvalues is the degree of the stretch along the according direction. These figures are independent with each other from the