Transportation Research Part C 36 (2013) 72–82

Contents lists available at ScienceDirect

Transportation Research Part C

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

A novel visible network approach for freeway crash analysis

Jianjun Wu [a], Mohamed Abdel-Aty [b], Rongjie Yu [b,d,*], Ziyou Gao [c]

[a] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
[b] Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, United States
[c] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China
[d] School of Transportation Engineering, Tongji University, 4800 Cao'an Road, 201804 Shanghai, China

ARTICLE INFO

Article history:
Received 8 March 2013
Received in revised form 24 June 2013
Accepted 12 August 2013

Keywords:
Freeway safety
Crash data
Safety analysis
Visible network

ABSTRACT

Freeway crashes have attracted considerable attention in recent years leading to the development of various methodologies to unveil the crash occurrence mechanisms including two general modeling approaches: parametric and non-parametric. In this paper, a novel visible network approach has been proposed to analyze crash characteristics with real-time traffic and weather data. In the suggested model, traffic states prior to crash occurrence have been extracted from real-time data; and crashes are mapped as nodes on the network. Each node contains information for the most hazardous factors relate to crash occurrence selected by random forest algorithm. With the help of transferring technology, links are connected between the nodes according to the state values. Therefore, complete freeway crash evolution networks can be obtained by analyzing one year crash data (including real-time weather and traffic variables) on I-70 in the state of Colorado. Additionally, the suggested method is also used to analyze single- and multi-vehicle crashes separately to identify their distinct characteristics. Compared with the traditional analysis methods, the proposed visible approach has the advantages of easy to be extended, transferred, and applied; easy to identify the effects of the various contributing factors on a traffic crash and to visually inspect the model. Moreover, the crash contributing factors identified in this study is beneficial for designing advanced early-warning and risk assessment systems in the context of real-time highway management.

1. Introduction

With the increase of densities of roadways and vehicles, real-time crash risk analysis has received much attention in recent years. In addition, timely prediction of incidents has also become a critical field in freeway traffic management. To improve the highway safety conditions, researchers have developed various methods to incorporate different types of data and concluded a variety of countermeasures (Yu et al., 2013). In this direction, detailed crash data, e.g., the crash type, the crash location (Sobhani et al., 2013; Abdel-Aty et al., 2007), the weather condition, the road condition, and effective statistical techniques for analysis of crash frequency data would better enable the identification of the crash contributing factors and probabilities.

To model the crash characteristics, a wide variety of methods have been employed. The most popular one is the Poisson regression method suggested by Jovanis and Chang (1986) which has been extended after many years since being proposed. In addition, various methods have been developed to solve different issues caused by the crash data (over-dispersion, under-dispersion, etc.). These methods include Poisson-Gamma (Oh et al., 2006), generalized estimating equation (Lord and Persaud, 2000), generalized additive (Xie and Zhang, 2008), random-effects Poisson (Johansson, 1996; Wang et al., 2009),

* Corresponding author.
E-mail address: rongjie.yu@gmail.com (R. Yu).

0968-090X/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.trc.2013.08.005

bivariate Poisson (Miaou and Lord, 2003; Aguero-Valverde and Jovanis, 2009; N'Guessan, 2010), neural network, Bayesian neural network, and support vector machine (Abdelwahab and Abdel-Aty, 2002; Chang, 2005; Riviere et al., 2006; Xie et al., 2007; Li et al., 2008; Gregoriades and Mouskos, 2013), and so on. More detailed review can be found in Lord and Mannering (2010).

In order to analyze the crash characteristics accurately, various types of factors contributing to traffic crashes are considered. Among which the most important ones are weather conditions and traffic variables. Caliendo et al. (2007) used hourly rainfall data and transformed it into a binary indicator of daily pavement surface status (dry and wet). Miaou et al. (2003) also used a surrogate variable to indicate wet pavement conditions. Malyshkina et al. (2009) investigated the effects of precipitation, snowfall amounts and temperature on crashes. In addition, traffic variables always play a vital role in crash occurrence studies. It has been proven that traffic density (Kononov et al., 2011), different traffic flow scenarios (Noland and Quddus, 2004), Annual Average Daily Traffic (Chang and Chen, 2005) were the key determinants for freeway crash frequencies. The results are useful in the analysis of freeway crash frequency. For example, it was found that when some critical traffic density is reached, the crash occurrence likelihood would increase at a faster rate with an increase in traffic density. However, using only weather related variables or aggregated traffic data would lead to the loss of the most valuable information of the pre-crash traffic status. Therefore, to improve the crash frequency analysis, both real-time weather conditions and traffic variables should be considered. Yu et al. (2013) employed a Bayesian inference approach with random effects Poisson models to establish safety performance functions; it was found that the weather condition variables, especially precipitation, play a key role in the crash occurrence models.

Although those traditional methods of analyzing crash-frequency data had been applied over the years, all of them have their own limitations. First of all, the abovementioned methods need appropriate hypothetical models with adjustable parameters which are different by datasets. Therefore, these methods could not be easily extended or transferred to be used for another roadway. In addition, they are difficult to visually inspect and complex to carry out in the treatment of huge quantity of real-time crash data. It is difficult for us to identify which situation has a higher probability of crash occurrence under the combined influence factors, such as the combination of crash location, weather, crash time, and road geometrics.

In recent years, network analysis technology has become an important tool to discover the complexities in the transportation network. Among them, the flow properties of the transported entities become of primary interest, e.g., traffic flow. Recently, there has been some research on the relationship between network and time series. Zhang and Small (2006) developed an interesting analysis of complex network and pseudoperiodic time series, where different time series corresponded to different network structures. From the perspective of dynamics, scale-free structure has recently been found to spontaneously appear in coupled logistic maps when linking is based on dynamical considerations (Gong and van Leeuwen, 2003) rather than the notion of preferential attachment usually cited in the evolution of scale-free networks. But these results are difficult to be applied in the real data analysis. Li et al. (2007) proposed a network model based on the traffic flow states generated by the Nagel and Schreckengerg's traffic model and modified comfortable driving traffic model. Wu et al. (2008) built a connection between the chaos time series and complex network in the car following model. Research towards the development of such techniques has particular relevance for a number of reasons. First, in the emerging topic of network complexity, one can view the road crash as a complex system which is affected by many factors. These factors can be obtained from the real-time traffic data. Therefore, the techniques developed in this paper may be viewed as a step on the way toward a novel statistical approach in real-time road crash data analysis.

In this study, we attempt to develop a visible network model and apply it to the freeway crash data analysis via mapping to the network from real-time freeway crash data. The network model is based on a simple and fast computational method, the visible graph algorithm which converts the data into a graph. Due to the visible (here means it can be represented by a graph form) and modeling simplicity, the network approach has its own advantages in real-time applications with rapid development.

2. Data description

This study focuses on a 15-mile mountainous freeway on I-70 in Colorado. It had been demonstrated that the weather condition, e.g., snow, visibility, has a great influence on the crash frequencies (Ahmed et al., 2011). Yu et al. (2013) employed real-time weather data (visibility, precipitation and temperature), freeway geometric characteristics data and real-time traffic data (speed, volume and lane occupancy) in developing safety performance functions. Our research is based on four datasets as listed below:

(1) **Dataset** D_1. This data set represents 1 year of crash data from August 2010 to August 2011 provided by Colorado Department of Transportation (CDOT). The data contains crash time, crash date, crash direction, mile post, etc. A total of 252 crashes were documented within the study period.
(2) **Dataset** D_2. The related road segment geometric characteristics data captured from the Roadway Characteristics Inventory (RCI). The 15-mile freeway section was split into 120 homogenous segments (60 in each direction) according to the major segmentation criterion of roadway alignment homogeneity and the Roadway Characteristics Inventory (RCI) data. Both horizontal and vertical alignments were scrutinized. A minimum-length of 0.1 mile was used to avoid the low exposure problem and the large statistical uncertainty of the crash rates in short segments.