Accident Analysis and Prevention 75 (2015) 192–201

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Competing risks mixture model for traffic incident duration prediction

Ruimin Li [a,d,*], Francisco C. Pereira [b,1], Moshe E. Ben-Akiva [c,2]

[a] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), 77 Massachusetts Avenue, Cambridge, MA 0219, USA
[b] Singapore-MIT Alliance for Research and Technology (SMART), 1 CREATE Way, #09-02 CREATE Tower, Singapore 138602, Singapore
[c] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Room 1-181, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[d] Department of Civil Engineering, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Article history:
Received 4 July 2014
Received in revised form 26 November 2014
Accepted 28 November 2014
Available online 6 December 2014

Keywords:
Traffic incident management
Traffic incident duration prediction
Competing risks
Mixture model

ABSTRACT

Traffic incident duration is known to result from a combination of multiple factors, including covariates such as spatial and temporal characteristics, traffic conditions, and existence of secondary accidents but also the clearance method itself. In this paper, a competing risks mixture model is used to investigate the influence of clearance methods and various covariates on the duration of traffic incidents and predict traffic incident duration. The proposed mixture model considers the uncertainty in any of five clearance methods that occurred. The probability of the clearance method is specified in the mixture by using a multinomial logistic model. Three candidate distributions, namely, generalized gamma, Weibull, and log-logistic are tested to determine the most appropriate probability density function of the parametric survival analysis model. The unobserved heterogeneity is also incorporated into the mixture model in a way that allows parameters to vary across observations based on the three candidate distributions. The methods are illustrated with incident data from Singaporean expressways from January 2010 to December 2011. Regression analysis reveals that the probability of different clearance methods and the duration of traffic incidents are both significantly affected by various factors, such as traffic conditions and incident characteristics. Results show that the proposed mixture model is better than the traditional accelerated failure time model, and it predicts traffic incident duration with reasonable accuracy, as shown by the mean average percent error.

1. Introduction

Traffic incidents are major causes of non-recurrent congestion on expressways (Haas, 2006) and urban arterial roads (Shao and He, 2008). An effective approach to reduce the influence of traffic incidents is the application of the Traffic Incident Management System (TIMS), which requires a timely and precise estimation of traffic incident duration. By performing reliable prediction of incident duration, traffic operators could deploy appropriate measures around the incident location and provide travelers with real-time traffic information to reduce incident-related traffic congestion. In the past two decades, significant research effort has been exerted on the analysis and prediction of traffic incident duration.

Total incident duration can be divided into the following sequential and distinct time intervals (Nam and Mannering, 2000; TRB, 2000; Valenti et al., 2010):

- Detection/reporting time: time between the time of incident occurrence and the time of response by the traffic control center operators receiving the call; usually, this period is difficult to capture.
- Preparation/dispatching time: time between operators receiving the call and dispatching the incident response team members.
- Travel time: time between incident response team members receiving the dispatch order and their arrival at the incident location.
- Clearance time: time between the incident response team members' arrival and incident clearance.

A general assumption in this work is that a model of incident duration can be discretized according to groups of "clearance methods", namely related to participation of police, tow truck or

* Corresponding author. Permanent address: Institute of Transportation Engineering, Department of Civil Engineering, Tsinghua University, Beijing 100084, China. Tel.: +86 10 62770985; fax: +86 10 62770985.
E-mail addresses: lrmin@tsinghua.edu.cn (R. Li), camara@smart.mit.edu (F.C. Pereira), mba@mit.edu (M.E. Ben-Akiva).
[1] Tel.: +65 9323 3653.
[2] Tel.: +1 617 253 5324; fax: +1 617 253 0082.

http://dx.doi.org/10.1016/j.aap.2014.11.023
0001-4575/

drivers self-driving off of the scene. From the perspective of hazard-based modeling, these methods correspond to “failure events”. Since the clearance method is not always observable, we propose a discrete mixture model.

This paper uses a hazard-based competing risks mixture model and is focused on analyzing the influence of various factors on the incident duration, as determined from the moment the operators receive the call to the total incident clearance, when the roadway capacity returns to its normal conditions; this duration is the sum of total preparation time, travel time, and clearance time but excludes the unknown period from the time of incidence occurrence to the time of reporting. The proposed model is also tested to predict the traffic incident duration.

The rest of this paper is organized as follows. The first section presents a literature review of online incident duration predictions and hazard-based models. The next section discusses the proposed parametric mixture approach used to analyze the duration of traffic incidents; parametric mixture duration models and coefficient estimation methods are also discussed in this section. The third section describes the data used in this study. The fourth section presents the calibration of the parametric mixture model, and the fifth section evaluates the prediction accuracy. Lastly, we end this paper by presenting our conclusions and suggesting future plans.

2. Literature review

In the past two decades, traffic incident duration has been investigated through various approaches, such as finding the factors that significantly affect traffic incident duration (Khattak et al., 2010; Zhang and Khattak, 2010) or predicting traffic incident duration. The majority of the literature review is focused on the latter.

Various regression models have been applied in predicting traffic incident duration. Khattak et al. (1995) applied a truncated regression model based on a simple time sequential procedure to predict the traffic incident duration by establishing a relationship between traffic incident duration and independent variables. However, they did not examine the prediction accuracy of the sequential model because not enough data were available to support the test. Peeta et al. (2000) developed a linear regression model that predicts traffic incident clearance time with time-independent variables. He et al. (2011) established an incident duration prediction model based on hybrid tree-based quantile regression to predict traffic incident duration on urban freeways. The results of the study showed that the proposed model had better prediction performance in comparison with three other kinds of prediction models. A non-parametric regression model (Smith and Smith, 2001) has also been used to predict traffic incident duration; however, the performance of the model was unsatisfactory, with an average error of more than 20 min.

Decision trees and classification trees, such as the classification and regression tree (Kim et al., 2008; Knibbe et al., 2006) and the M5P tree (Zhan et al., 2011), have also been applied to predict traffic incident duration. One study was able to predict incident duration on the basis of the time interval it occurred and obtained an overall confidence of over 80% (Kim et al., 2008). Another study showed that the M5P tree algorithm can perform a prediction with a mean average percentage error (MAPE) of 42.7% (Zhan et al., 2011).

Several kinds of Bayesian classifiers (Boyles et al., 2007; Demiroluk and Ozbay, 2011; JiYang et al., 2008; Li and Cheng, 2011; Shen and Huang, 2011) have been used to accommodate incomplete information or information received at different time points. These studies show that the Bayesian classifier has better prediction performance than other traditional models (Demiroluk and Ozbay, 2011), such as linear regression and classification and regression trees (CART); for example, the presented model outperformed the CART model with 53% accuracy rate.

A number of studies recently applied artificial neural networks (ANN) in developing prediction models of traffic incident duration. Wei et al. (Lee and Wei 2010; Wei and Lee, 2007) developed two ANN-based models that sequentially predict traffic accident duration, and the results showed that these models achieve a reasonable prediction; that is, the MAPEs of the models were mostly under 40% (Wei and Lee, 2007). Pereira et al. (2013) used radial basis function algorithm, which continuously makes predictions as new information arrives. New information arrives in the form of text messages (internal to the traffic operator and emergency response system) and is analyzed with text mining techniques (latent Dirichlet allocation) to extract a list of “topics” associated with the current situation of the incident. The overall median error decreased by 28% in the approach with topics in comparison with that without topics. Other techniques, such as genetic algorithms (GA) (Lee and Wei 2010) and fuzzy logic (Vlahogianni and Karlaftis, 2013) have been combined with ANN to obtain better prediction performance. Vlahogianni and Karlaftis (2013) applied fuzzy entropy feature selection to select the factors to be used for incident duration prediction. Hazard-based duration models, which focus on time to event data, have been used in estimating and predicting traffic incident duration. Jones et al. (1991) applied an accelerated failure time (AFT) model with log-normal distribution to examine the factors affecting the incident duration on Seattle freeways. AFT hazard-based duration models were then applied to different traffic incident duration intervals (Nam and Mannering, 2000), and the results revealed that different distributions of the hazard function are suitable for different incident duration intervals and that a wide variety of factors significantly affect incident time intervals.

Different distributions-based AFT hazard-based models, which are based on different data resources of traffic incidents, have been used to estimate and predict traffic incident duration. These models include log-logistic distribution (Chung, 2010; Chung et al., 2010; Kang and Fang, 2011; Wang et al., 2013), Weibull distribution (Alkaabi et al., 2011; Hojati et al., 2013), log-normal distribution (Chung and Yoon, 2012) and gamma distribution (Li, 2014). To find a more appropriate distribution for the hazard function, Ghosh et al. (2012) applied the flexible generalized F distribution to fit the traffic incident duration.

On basis of whether the model distinguishes multiple (and possibly latent) clearance methods, two types of hazard-based models can be considered: single and competing risks hazard-based models. The above mentioned models generally fall under the single risk category. However, in view of the high heterogeneity in incident types, driver behavior (i.e., the drivers that participate in the incident), and response strategies, significant information may be lost when all of the factors are aggregated into one type.

Competing risks hazard-based models have recently been widely used in medical research (Fürstová and Valenta, 2011; Haller et al., 2013; Lau et al., 2008, 2009, 2011; Ravani et al., 2005) and in transportation, among other fields. Gilbert (1992),Hensher (1998), and Yamamoto et al. (2004) applied competing risks hazard-based models to investigate the time spent and influencing factors in automobile transactions. Ettema et al. (1995) and Bhat (1996) investigated travel activity duration by using accelerated lifetime and proportional hazard models, respectively. Li and Guo (2014) investigated the factors effected on the duration of two incident group with proportional hazard competing risk model. Shyr and Ben-Akiva (1996) used mixture competing risks hazard models to examine rail fatigue behavior. The use of competing risks