Analytic Methods in Accident Research 5-6 (2015) 59–71

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Fuzzy modeling of freeway accident duration with rainfall and traffic flow interactions

Loukas Dimitriou [a,*], Eleni I. Vlahogianni [b,1]

[a] Department of Civil and Environmental Engineering, University of Cyprus, 75 Kallipoleos Street, P.O. Box 20537, 1678 Nicosia, Cyprus
[b] National Technical University of Athens, 5 Iroon Polytechniou Street, Zografou Campus, 157 73 Athens, Greece

ARTICLE INFO

Article history:
Received 14 March 2014
Received in revised form
23 April 2015
Accepted 24 April 2015
Available online 22 May 2015

Keywords:
Accident duration
Traffic and environmental variables
Multivariate analysis
Fuzzy modeling

ABSTRACT

Accident duration modeling has been considered as a difficult problem due to the variety of information (accident characteristics, traffic and weather information, geometry of the accident location and so on) that should be taken into account to improve predictions and explain the phenomenon. We introduce Fuzzy Rule-Based Systems to model freeway accident duration and cope with the uncertainties and complexities hindering in accident monitoring systems. The models are also compared to classical hazard-based regression models, as well as Multi-Layer Perceptrons. Results show that a Fuzzy Rule-Based System may predict accident duration with fair accuracy using limited information on traffic and weather conditions. Introducing the entire amount of information on accidents to the Fuzzy Rule-Based System leads to reduced modeling accuracy, probably due to the difficulties in converging to a solution. Nevertheless, the Fuzzy Rule-Based System with limited information may predict more accurately than classical hazard based duration models and with comparable accuracy to a Multi-Layer Perceptron, which is presented with information on accident characteristics, traffic and weather conditions, as well as the geometry of the accident location.

1. Introduction

Modeling incident duration has for long been a serious consideration in freeway traffic management; incident durations are directly related to network's efficiency and the secondary accident likelihood (Khattak et al., 2009; Vlahogianni and Karlaftis, 2012). Four main approaches have been implemented to model incident durations: simple regression; survival analysis; neural networks and Bayesian networks. The efficiency of classical linear regression models in accident duration modeling have been documented in Khattak et al. (1995, 2012) and Garib et al. (1997). Neural networks have been also applied as nonlinear regression models. Wang et al. (2005) compared the performance of MLP classifiers for incident duration prediction to fuzzy logic and underlined that both approaches provide reasonable estimates of incident duration, but the accident parameters used in modeling were limited. Wei and Lee (2007) a similar neural network, but with richer input space and detailed upstream and downstream traffic information for the point of interest (accident occurrence location). To improve the neural network regression process through optimizing the input space of the models, Lee and Wei

* Corresponding author. Tel.: +357 22892286.
E-mail addresses: lucdimit@ucy.ac.cy (L. Dimitriou), elenivl@central.ntua.gr (E.I. Vlahogianni).
[1] Tel.: +30 210 772 1369; fax: +30 210 772 1454.

http://dx.doi.org/10.1016/j.amar.2015.04.001
2213-6657/

(2010) proposed a genetically optimized feature election process prior to the neural network modeling. Ozbay and Nayan (2006) proposed a framework based on Bayesian Networks for modeling accident duration. The main advantage of such an approach is that it provides a probabilistic framework and incorporates uncertainty in modeling. Bayesian Networks have been also implemented to traffic analysis and forecasting (Vlahogianni et al., 2007; Castillo et al., 2008), secondary accidents modeling (Vlahogianni et al., 2010a, 2010b) and driver's behavior (Vlahogianni and Golias, 2012).

Survival analysis is the most popular manner to analyze duration data. It assumes that the probability of an incident to end is dependent on its duration and deals with the probability of an incident to end in the forthcoming time interval given it has lasted until now. This modeling approach has been adopted in Golob et al. (1987), Giuliano (1989), Nam and Mannering (2000) and Chung (2010). Golob et al. (1987) found that a lognormal distribution fits duration data best, whereas Jones et al. (1991) resulted in a log logistic distribution as the optimum fit. Nam and Mannering (2000) tested several distributions from exponential and Weibull to log-logistic and lognormal for modeling incident duration and found that response times are described by a 2 or 3 parameter Weibull distribution, whereas clearance times are best described by a log-logistic distribution. Chung (2010) implemented a log-logistic accelerated failure time regression model. Nam and Mannering (2000) underline that hazard-based duration models have an advantage over the rest of the modeling approaches as they allow the explicit study of duration effects of accidents. Recently, Vlahogianni and Karlaftis (2013) developed and tested a fuzzy entropy feature selection neural network model to imitate the survival analysis models with single and competing uncertainties and predict incident durations that are prone to severe congestion, the occurrence of secondary incidents, as well as their joint effect.

Apart from the methodologies proposed and the methods employed, interest has focused on revealing the factors that may influence the accidents' duration especially on freeways. Accident factors such collision type (rear-end, sideswipes etc.), severity (injuries, fatalities etc.), location, lanes blocked, number of vehicles involved in the accident, heavy vehicle involvement have been extensively studied (Golob et al., 1987; Giuliano, 1989; Nam and Mannering, 2000; Lee and Wei, 2010). Moreover, the geometry specifications, such as toll plaza, interchange etc. that is located near the accident occurrence location (Jones et al., 1991; Wei and Lee, 2007; Khattak et al., 2012) and incident detection type (call, operator etc.) (Wang et al., 2005; Chung, 2010; Khattak et al., 2012) have been also addressed.

Literature has systematically underlined traffic and weather conditions as two significant aspects of modeling accident duration. Various approaches have been adopted in literature in order to incorporate such information in modeling. Garib et al. (1997) introduced the average traffic flow upstream of an incident before its occurrence, as well as the reduction in capacity due to the incident, and stated that traffic conditions are negatively related to incident durations but this outcome may vary from site to site. Nam and Mannering (2000) related the day, month, type of period (peak or not, daytime etc.), to the accident duration. Wei and Lee (2007) and Lee and Wei (2010) introduce traffic volume and speed information upstream and downstream from the location of the accident. Although Wei and Lee (2007) have not clearly shown the influence of traffic related data to the accuracy of incident duration modeling, they stated that traffic is a feasible input of the incident duration model in terms of model effects. Khattak et al. (2012) related the Average Annual Daily Traffic (per 1000 vehicles) to the incident duration and found that AADT is positively related to incident durations. Tavassoli Hojati et al. (2013, 2014) proposed a set of parametric accelerated failure time survival models for incident duration modeling with emphasis on traffic characteristics including daily and hourly traffic volume, v/c at the time of the incident and speed.

Adverse weather conditions and especially rainfall may have a significant impact on traffic operations (Hall and Ibrahim, 1994; Vlahogianni et al., 2010a, 2010b; Tsirigotis et al., 2012; Vlahogianni and Karlaftis, 2012). Khattak et al. (1995) introduced a variable in linear regression dedicated to weather conditions; results showed a positive relationship between clear weather conditions and incident durations. Whether in terms of dry or rainy conditions during the occurrence of the accident were also introduced in Garib et al. (1997). Nam and Mannering (2000) introduced weather – rainfall, snowfall etc. – as dummy variables in the survival analysis modeling. Khattak et al. (2012) introduced the weather to the incident prediction system in the form of binary variables (bad/good weather). Vlahogianni and Karlaftis (2013) found that rainfall intensity is a highly contributing factor, while lane volume, number of blocked lanes, as well as number of vehicles involved in the incident are among the top ranking factors for determining the extent of duration. Recently, Tavassoli Hojati et al. (2013) provided findings showing no significant effects of infrastructure and weather on incident duration.

The above analysis indicates that literature is numerous regarding both the methods implemented and the critical factors of incident duration revealed. However, the observed complexities related to the relationships between traffic, weather, other incident related and infrastructure related factors and the incident duration have been rarely treated through a stochastic modeling framework, which may take into consideration the uncertainties involved in observed incident datasets. The present paper extends past research by introducing a Multiple-Input–Single-Output Hybrid Fuzzy system for incident duration modeling that is consistent with the stochasticity/uncertainty met in traffic and rainfall data. The intuition behind using a fuzzy approach centers to three issues: first, to test whether the expert knowledge may improve the modeling, second, to overcome any restrictions on the variables' distribution assumptions/properties of classical duration models, and, third, to provide both the range, as well as a crisp value reflecting the most possible outcome of an incident duration under alternative prevailing conditions. The proposed approach targets prediction, regardless if this may be at the expense of the understanding of the underlying mechanism in accident-related phenomena. For the proposed model estimation, Genetic Algorithms are used. The models will be evaluated with respect to hazard-based regression and neural network regression.