Analytic Methods in Accident Research 10 (2016) 12–25

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Modeling nonlinear relationship between crash frequency by severity and contributing factors by neural networks

Qiang Zeng [a,b], Helai Huang [b,*], Xin Pei [c], S.C. Wong [d]

[a] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong 510641, PR China
[b] Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, PR China
[c] Department of Automation, Tsinghua University, Beijing, PR China
[d] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

Article history:
Received 18 September 2015
Received in revised form
20 March 2016
Accepted 21 March 2016
Available online 6 April 2016

Keywords:
Crash frequency by severity
Neural network
Over-fitting
Structure optimization
Rule extraction

ABSTRACT

This study develops neural network models to explore the nonlinear relationship between crash frequency by severity and risk factors. To eliminate the possibility of over-fitting and to deal with black-box characteristic, a network structure optimization and a rule extraction method are proposed. A case study compares the performance of the modified neural network models with that of the traditional multivariate Poisson-lognormal model for predicting crash frequency by severity on road segments in Hong Kong. The results indicate that the trained and optimized neural networks have better fitting and predictive performance than the multivariate Poisson-lognormal model. Moreover, the smaller differences between training and testing errors in the optimized neural networks with pruned input and hidden nodes demonstrate the ability of the structure optimization algorithm to identify insignificant factors and to improve the model's generalizability. Furthermore, two rule-sets are extracted from the optimized neural networks to explicitly reveal the exact effect of each significant explanatory variable on the crash frequency by severity under different conditions. The rules imply that there is a nonlinear relationship between risk factors and crash frequencies with each injury-severity outcome. With the structure optimization algorithm and rule extraction method, the modified neural network models have great potential for modeling crash frequency by severity, and should be considered a good alternative for road safety analysis.

1. Introduction

In the past decade, there has been a quantity of research on predicting crash frequency by certain categories, such as injury severity (e.g., property damage only, possible injury, non-incapacitating injury, incapacitating injury or fatality) (Park and Lord, 2007), the number of vehicles involved (e.g., single vehicle, two vehicles, or three or more vehicles) (Venkataraman et al., 2013) or collision type (e.g., angle, head-on, rear-end, sideswipe or pedestrian-involved) (Ye et al., 2009). The first kind of classification covers most concerns, because crash injury severity is an important aspect in assessing safety

* Corresponding author.
E-mail addresses: 641459622@qq.com (Q. Zeng), huanghelai@csu.edu.cn (H. Huang), peixin@mail.tsinghua.edu.cn (X. Pei), hhecwsc@hku.hk (S.C. Wong).

http://dx.doi.org/10.1016/j.amar.2016.03.002
2213-6657/

performance, in addition to the crash frequency (AASHTO, 2010). Compared with conventional crash prediction models (referred to as "safety performance functions"), modeling crash frequency by severity identifies the effects of observed risk factors (such as the traffic, geometrical and environmental characteristics of sites) on the frequency of accidents with a particular injury-severity outcome. The expected crash frequencies at each level of severity provide deeper insights on the safety situation of a certain road entity (road segment, intersection, etc.). Therefore, while crash totals may not reveal a site deficiency, over exposure of a specific crash severity may uncover otherwise undetected deficiencies. Moreover, the models have been employed to rank road sites with promise for safety improvement, a critical step of network screening in the roadway safety management process (AASHTO, 2010), as injury severity and its associated costs are primary concerns in many programs (Miaou and Song, 2005).

Methodologically, there are mainly two groups of approaches to crash frequency by severity prediction: joint and separate modeling. In the former group, correlation between crash frequencies at various severity levels is the most important issue. To deal with it, a series of techniques have been investigated, such as multivariate regression models (Aguero-Valverde and Jovanis, 2009; Anastasopoulos et al., 2012; Barua et al., 2014, 2016; Bijleveld, 2005; El-Basyouny and Sayed, 2009; El-Basyouny et al., 2014; Ma and Kockelman, 2006; Ma et al., 2008; Park and Lord, 2007), simultaneous equations (Ye et al., 2009, 2013), a joint-probability approach (Pei et al., 2011), two-stage bivariate/multivariate models (Wang et al., 2011; Xu et al., 2014) and multinomial-generalized Poisson models (Chiou and Fu, 2013, 2015; Chiou et al., 2014). The multivariate Poisson regression proposed by Ma and Kockelman (2006) adds a common error term into the Poisson distributions of univariate regressions to account for their correlation, but it does not allow for the commonly observed over-dispersion, and it assumes the identical and positive covariances across crash frequencies (Park and Lord, 2007). In order to improve it, a multivariate Poisson-lognormal regression has been developed (Ma et al., 2008), which is able to accommodate over-dispersion and provides a fully general covariance structure. To account for the spatial correlation among neighboring sites, error terms with Gaussian conditional auto-regressive distribution have been introduced into the multivariate Poisson-lognormal model (Barua et al., 2014). Based on it, Barua et al. (2016) have proposed a multivariate random parameters count model to further capture unobserved heterogeneity across observations.

Compared with multivariate regression models, the formulation of simultaneous equations, the joint probability model and the two-stage bivariate/multivariate models are less complicated (Pei et al., 2011; Wang et al., 2011; Xu et al., 2014; Ye et al., 2009, 2013). Besides, the computation burden of simultaneous equations is lighter, because their coefficients are calibrated by a simulated likelihood estimation method (Ye et al., 2009), while the others are calibrated by Markov chain Monte Carlo simulation, a typical Bayesian inference method. On the contrary, the multinomial-generalized Poisson models (Chiou and Fu, 2013; Chiou et al., 2014), especially the extension with accommodating spatio-temporal dependence (Chiou and Fu, 2015), are even more complicated than multivariate count models.

Although the correlation across severity levels is significant in many studies, the advantage of joint modeling over separate modeling is not "theoretical" but rather "empirical", as noted by Ma et al. (2008). In the comparative analysis conducted by Lan and Persaud (2012), univariate models are found to fit the crash data better than the multivariate model. Consequently, some researchers continue to separately model crash frequencies at each severity level. For example, Venkataraman et al. (2013) advocate univariate random parameter models to individually predict crash frequency by severity, or other aggregation types, by accounting for heterogeneities across unobserved or unobservable factors. All of the above-mentioned models are based on a generalized linear function framework and certain assumed distributions of crash data. However, in some cases, these assumptions may be violated and thereby result in biased inferences (Li et al., 2008).

Relative to the statistical models, without any prior knowledge or assumption on model structure, some artificial intelligence models can be used to approximate the underlying nonlinear relationship between crash frequency by severity and safety predictors (Haykin, 2009). As a common class of artificial intelligence models, neural network models have been successfully used in many fields of transportation research (Karlaftis and Vlahogianni, 2011). For highway safety analysis, a number of studies have investigated the performance of neural network models in predicting crash frequency or injury severity (Abdelwahab and Abdel-Aty, 2001; Chang, 2005; Huang et al., 2016; Zeng and Huang, 2014b). The results show that neural network models outperform some traditional statistical models, such as the negative binomial model of crash frequency prediction and the ordered logit/probit models of crash injury severity prediction. To the best of our knowledge, neural networks have not yet been employed to predict crash frequency by severity.

Moreover, with the development of neural network techniques, the commonly criticized weaknesses of crash prediction, the over-fitting problem and the black-box characteristic, have been mostly eliminated. Advanced methods for network training and structure optimization can establish generalized neural network models that effectively approximate the relationship between crash frequency by severity and explanatory variables (Haykin, 2009). In addition, piecewise linear rules extracted from the developed neural networks are able to clearly illustrate the effects of risk factors (Setiono and Thong, 2004).

In summary, this study attempts to develop advanced neural networks for modeling the nonlinear relationship between crash frequency by severity and risk factors, and to clarify the effects of factors on the outcomes by extracting rules from the developed neural networks. To demonstrate the proposed methods, the neural network models are compared with the multivariate Poisson-lognormal model with regard to fitting and predictive performance. Accordingly, the remainder of this paper is organized as follows. The next section specifies the proposed models and methods. The collected data for model demonstration are described in Section 3. Section 4 introduces the detailed implementation of the proposed models and discusses the results. Finally, conclusions and recommendations for future research are presented in Section 5.