ELSEVIER

Accident Analysis and Prevention 38 (2006) 434–444

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks

Dursun Delen *, Ramesh Sharda, Max Bessonov

Department of Management Science and Information Systems, Oklahoma State University, Stillwater, OK 74106, United States

Received 25 April 2005; received in revised form 26 May 2005; accepted 15 June 2005

Abstract

Understanding the circumstances under which drivers and passengers are more likely to be killed or more severely injured in an automobile accident can help improve the overall driving safety situation. Factors that affect the risk of increased injury of occupants in the event of an automotive accident include demographic or behavioral characteristics of the person, environmental factors and roadway conditions at the time of the accident occurrence, technical characteristics of the vehicle itself, among others. This study uses a series of artificial neural networks to model the potentially non-linear relationships between the injury severity levels and crash-related factors. It then conducts sensitivity analysis on the trained neural network models to identify the prioritized importance of crash-related factors as they apply to different injury severity levels. In the process, the problem of five-class prediction is decomposed into a set of binary prediction models (using a nationally representative sample of 30 358 police-recorded crash reports) in order to obtain the granularity of information needed to identify the "true" cause and effect relationships between the crash-related factors and different levels of injury severity. The results, mostly validated by the findings of previous studies, provide insight into the changing importance of crash factors with the changing injury severity levels.

Keywords: Injury severity; Classification; Artificial neural networks; Sensitivity analysis; Problem decomposition

1. Introduction

Over 6 million traffic accidents claim more than 40 000 lives each year in the United States according, to the National Highway Traffic Safety Administration (GES, 2005). Causes of accidents and related injury severity are of special concern to researchers in traffic safety, since such research would be aimed not only at prevention of accidents but also at reduction of their severity. One way to accomplish the latter is to identify the most probable factors that affect injury severity. Understanding the circumstances under which drivers and passengers are more likely to be killed or more severely injured in an automobile accident can help improve the overall driving safety situation. Factors that affect the risk of increased injury of occupants in the event of an automotive accident include demographic or behavioral characteristics of the person (age, gender, seatbelt usage, or use of drugs or alcohol while driving), environmental factors and roadway conditions at the time of the accident occurrence (surface, weather or light conditions, the direction of impact, vehicle role, or occurrence of a rollover), as well as technical characteristics of the vehicle itself (vehicle age and body type).

The primary interest of this study is to identify which of these factors become important in influencing the probability of increased injury severity during a crash. Accidents examined herein include a geographically representative sample of multi-vehicle collision accidents, single vehicle fixed-object collisions, and single vehicle non-collision (rollover) crashes. Many of the previous studies in this domain have used regression type generalized linear models where the functional relationships between the injury severity and the crash-related factors are assumed to be linear. As noted by Mussone et al. (1999), these linear models suffer from problems related to the use of variables with non-homogenous distribution and are known to be prone to statistical problems when the correlation among the independent variables is greater than acceptable levels, resulting in unreliable models with greater error than desirable. Since artificial neural networks are capable of capturing highly non-linear relationships between the predictor variables (crash factors) and the target variable (severity level of the injuries), in this study, a series of artificial neural networks models were

* Corresponding author. Tel.: +1 918 594 8283; fax: +1 918 594 8281.
E-mail address: Delen@okstate.edu (D. Delen).

0001-4575/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2005.06.024

used to estimate the effects of significant crash factors on the level of injury severity sustained by the driver. Though artificial neural networks have been investigated by other researchers in this domain (Abdelwahab and Abdel-Aty, 2001; Mussone et al., 1999), we propose and employ some methodological innovations to develop the sensitivity analysis from the neural network models.

This paper is organized as follows. The next section summarizes previous research in accident analysis. Section 3 describes our data compilation and processing approach, as well as definition of all the variables. Section 4 presents the details of the experiments to determine the effect of the various independent variables. Section 5 includes the results of these experiments along with the discussion of the findings, and the last section summarizes and concludes the paper.

2. Prior research

There is an extensive literature devoted to analysis of traffic accidents and crash modeling. The majority of this literature references the studies that deal primarily with analysis of crash involvement and prediction of crash totals. Since this research deals with analysis of injury severity and its deriving factors, this review attempts only to present the most relevant and rigorous work in this area.

A number of studies have attempted to develop injury severity models, though most concentrate on the traffic accident records limited to a small geographic area, a particular accident type, or a certain road condition situation. The reason is to make the domain as narrow as possible so that a somewhat homogenous dataset can be used to derive accurate prediction models. What follows is a review of a number of recent studies that develop models to identify the factors most important in increasing or reducing the levels of injury severity experienced by drivers and/or passengers during traffic accidents.

Logistic regression (a type of regression where the dependent is a categorical as opposed to a numerical variable) has been the most popular technique in developing injury severity prediction models. Lui and McGee (1988) used logistic regression to analyze the probability of fatal outcomes of accidents given that the crash has occurred. The data were obtained from the Fatal Accident Reporting System, which contained accidents that included at least one fatality. Probability of a fatal outcome was modeled as a target variable, which is dependant on driver's age and gender, impact points, car deformation, use of restraint system, and vehicle weight. Their findings reveal that a heavier car weight can greatly reduce the driver's risk of dying in a two-car crash, credited to the intuition that larger cars' frames can better absorb energy from an impact, or the fact that the small and lighter cars tend to roll over more easily. These findings were also confirmed by Wood and Simms' (2002) study where they measure the impact of car size on injury risk. Lui and McGee (1988) also analytically confirmed that the weakest side of a car from a driver's standpoint is the left side, and a belted driver subject to a left front side impact is estimated to have a higher risk of dying then an unbelted driver.

In another recent study a multivariate logistic regression (Bedard and Lu, 2002) revealed that the odds ratio of a fatal outcome of a crash increases with age, reaching 4.98 for drivers aged 80+ compared to the drivers aged 40–49 years. Female gender and blood alcohol content greater than 0.30 were also to be found associated with higher fatality odds. Also, the driver side impacts doubled the odds of fatality compared to frontal impacts. This study was limited to single vehicle crashes with fixed objects. In yet another study a logistic regression approach was used to examine a contribution of individual variables to the injury severity (Al-Ghamdi, 2002). The study was limited to 560 accidents obtained from the police records in Riyadh, Saudi Arabia. The dependent variable was modeled as a dichotomous variable that could only take values of fatal or non-fatal crash outcomes. According to the logistic regression results, out of nine independent variables used in this study, only two were found to be statistically significant with respect to the injury severity: location and cause of accident. For example, the odds of being in a fatal accident at a non-intersection location are 2.64 higher than those at an intersection, and the odds of severe injury increases on accidents caused by over-speeding and entering the wrong way traffic.

Farmer et al. (1997) used a binomial regression model to investigate the impact of vehicle and crash characteristics on injury severity in two-vehicle side-impact crashes. They found that rollover or ejection from the vehicle increases the likelihood of a serious injury or death (i.e., injuries with an Abbreviated Injury Scale of at least 3) and that light-duty trucks (which they defined as pickups and utility vehicles – not vans – under 10 000 pounds of gross-vehicle weight [GVW]) were fourteen times more likely to roll than cars, when struck on the side. While gender was not a statistically significant factor in their results, the oldest drivers (aged 65 and over) were estimated to be more at risk for serious injury.

Recently, ordered probit models have become popular in analysis of multi-class injury severity data sets. For instance, an ordered probit model was employed at the University of Texas (Kockelman and Kweon, 2002) in order to examine the risk of different injury levels sustained under a variety of crash types. The results suggest that pickups and SUVs are less safe than passenger cars under single-vehicle accident conditions. However, in two-vehicle accidents, they were found to be safer for their drivers and more dangerous for the occupants of their collision partners. In a similar study, using 1994 and 1995 crash data from the state of Florida, Abdel-aty et al. (1998) used ordered probit technique to examine relationships between driver age and crash characteristics. The three injury severities in their study were no injury, injury and fatality, and their results suggest that injury severity is positively associated with age. They also concluded that middle-age drivers are more likely to be involved in some crashes, but older drivers are more likely to be involved in fatal crashes.

O'Donnell and Connor (1996) compared ordered logit and ordered probit models in assessing the probabilities of four levels of injury severity as a function of driver attributes. In logit and probit regression models, while predicting the categorical response variable, a membership probability value (in the form