Transportation Research Part A 59 (2014) 384–396

Contents lists available at ScienceDirect

Transportation Research Part A

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

Highway Accident Modeling and Forecasting in Winter

Akram Khaleghei Ghosheh Balagh, Farnoosh Naderkhani, Viliam Makis *

Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, Ontario M5S 3G8, Canada

ARTICLE INFO

Article history:
Received 4 September 2012
Received in revised form 3 April 2013
Accepted 2 October 2013

Keywords:
Stochastic modeling
Logistic regression
Markov process
Road safety

ABSTRACT

Environmental attributes are critical risk factors that have proven to affect collision rates. Associated driving risks can be reduced by better maintenance of roadway infrastructure, enforcement of speed limits or other traffic laws. Given the preventive nature of these policies and regulations, accurate predictions of environmental attributes are needed. Currently, most of road safety prediction models are based on deterministic weather forecasts which are not able to capture changes in the likelihood of collision occurrence. As a result, probabilistic forecast is required to improve decision making, mainly in winter. In this paper, a stochastic approach to modeling highway collisions in winter time is considered which enables better assessment of driving conditions and a more accurate prediction. A logistic regression model with covariates is applied to crash data where environmental characteristics are modeled as a finite state space homogeneous multivariate discrete time Markov chain. After fitting the model, weather prediction as well as the conditional predictive probability distribution of collision occurrence are obtained. As the application, the ability of the proposed model to predict hourly environmental attributes and collision occurrence is examined using real highway crash data. The performance of the developed stochastic model is compared with several existing models in the literature using actual collision data. The results demonstrate that the proposed stochastic model outperforms existing models and it accurately predicts collision occurrence in the presence of stochastically changeable winter weather conditions. As a result, the proposed probabilistic forecast model can be used as a valuable tool in a decision support system.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Every year more than 1.17 million people die in road crashes around the world. Over 10 million are crippled or injured each year. The Global Burden of Disease study undertaken by the World Health Organization (WHO), Harvard University and the World Bank showed that in 1990, traffic crashes were assessed to be the world's ninth most important health problem. The study forecasts that by the year 2020 road crashes would move up to the third place in the table of leading causes of death and disability throughout the world (The Global, 1990). Deteriorating driving conditions and changes in winter due to snowfall and ice formation, significant reduction in road friction, and visibility impairment make vehicle handling more difficult and create a safety threat (Perry and Symons, 1991; Brown and Baass, 1997; Andrey et al., 2001; Eisenberg and Warner, 2005). Due to the incredible economic and emotional burden winter traffic crashes impose on society, reducing the number of vehicle crashes has long been a primary emphasis of highway agencies and motor-vehicle manufacturers.

* Corresponding author. Tel.: +1 416 9784184.
E-mail addresses: akhalegh@mie.utoronto.ca (A. Khaleghei Ghosheh Balagh), farnoosh@mie.utoronto.ca (F. Naderkhani), makis@mie.utoronto.ca (V. Makis).

0965-8564/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tra.2013.10.008

As documented in Lord and Mannering (2010), over the years, researchers have used a wide range of methodological tools to assess the impact of several risk factors for analyzing crash data. These research efforts have resulted in several widely accepted models. Typical models include the Poisson (Jovanis and Chang, 1989), zero-inflated Poisson and negative binomial (Qin et al., 2004), Poisson-Gamma (Malyshkina and Mannering, 2010b), finite-mixture/Markov switching models (Park and Lord, 2009), Bayesian neural networks (Riviere et al., 2006), and neural networks (Abdelwahab and Abdel-Aty, 2002). All above-mentioned models have been used to analyze cross-sectional data and recent literature suggests they are not capable of taking into account the effect of serial correlation often found in real data. This correlation could arise from investigating the same geographic region (spatial correlation), the same observational unit over successive time periods (temporal correlation), or a combination of the two. Ignoring the serial correlation may lead to less precise parameter estimates, resulting in imprecise statistical inferences. Different solutions have been introduced to overcome this limitation. First and simple solution to control serial correlation would be to introduce a time trend variable as an explanatory variable in the model. For instance (Noland et al., 2006) used a negative binomial model with a trend variable to study the effect of the London congestion charge on traffic casualties. However, there is no guarantee that this will explicitly account for the effect of serial correlation, specifically for the case of a long time series count data (Lord and Mannering, 2010). Incorporating the generalized estimating equation (GEE) would be another solution for handling temporal correlation. GEE model was applied for example by Lord and Persaud (2000) to describe different serial correlation structures for modeling repeatedly measured data. Their proposed model was applied to 4-legged signalized intersections in Toronto, Canada, for accident prediction. However this requires to determine or evaluate the type of temporal correlation in advance. Random and fixed-effects models are other methods which have been applied to take into account correlation among the observations. Recently, in the context of crash frequencies modeling (Guo et al., 2010) used random-effects models in the development of crash-risk maps in the state of Florida. However, these models may not be easily applied to non-temporal and spatial correlation dataset. A modified continuous dependent variable time-series methodologies such as integer-valued autoregressive Poisson models (Karlis, 2006; Quddus, 2008), integer-valued autoregressive moving average Poisson models (McKenzie, 1988), and negative binomial integer-valued generalized autoregressive conditional heteroscedastic model (Zhu, 2011; Ye et al., 2011) were also developed to account for the temporal correlation of crash data. However, extending the existing continuous dependent variable time series methods to the count data is not an easy task and it has been sparsely done. In such models, explanatory variables related to the road, environmental characteristics, human factors, vehicle type, and specific crash data were considered to estimate the crash frequencies.

Traditional statistical models used accident histories to predict current road safety conditions, and their disadvantages highlight that the complete situation of the accident cannot be completely known. Therefore, due to stochastically and dynamically changing environmental characteristics in winter season, existing traffic models fail to thoroughly capture changes in crash frequencies. Although some of these models may take weather forecasts data as an input for numerical accident prediction models, this kind of forecast is deterministic and does not properly assess uncertainty, which is important in decision making considering weather factors. Thus, driving in winter times remains a substantial hazard to drivers and demands an assessment tool to evaluate traffic safety. Having the limitation of the existing models in mind, the objective of this research is to introduce a probabilistic model capable of forecasting both the environmental characteristics and the probability of collision occurrence. The developed probabilistic model that we propose is regression-based. The main distinction between our proposed model and those in the literature is that covariates in regression models have traditionally been treated as a non-stochastic variables. In real-life situations, however, the covariate can be stochastic or non-stochastic. Some regression models with a stochastic covariate have been introduced. For instance (Oral and Gunayy, 2004) applied the binary regression model with stochastic risk factor and showed that their solutions were much more precise than those obtained by deterministic models. Later (Oral, 2006) extended their solution to a non-logistic regression model. Due to intractability of maximum likelihood estimation, a modified maximum likelihood method was utilized to obtain the parameter estimates. A stochastic linear regression model for both normal and non-normal multivariate design variables were developed by Islama and Tikub (2010). The authors showed the efficiency and the robustness of the estimators using real-life data. To the best of our knowledge, this paper is the first theoretical attempt to incorporate multiple stochastic covariates with Markov chain property into the logistic regression model.

The weather attributes form a multi-dimensional random process, and therefore could be modeled using the stochastic process theories. The Markov chain model considers transitions from one state to another over time and enables us to calculate the probability of realization of a possible situation that might arise in future not from old data but from the present data (Basawa and Prakasa Rao, 1980). Markov chain is a useful tool in modeling many practical systems such as queuing systems (Ching, 2001), manufacturing systems (Buzacott and Shanthikumar, 1993) and inventory systems (Fleischmann, 2001). The conditional probability distribution of each weather attribute in the next period given the currently available information depends not only on its own current condition, but also on the current state of other weather attributes. Therefore, the multivariate Markov chain model (MMC) is developed to explore both temporal and cross-sectional dependencies of categorical time series for weather characteristics. By exploring these relationships, better models, and hence better prediction rules can be developed. Once the parameters of Markov model are estimated, a logistic regression model is applied to crash data which depends both on the environmental characteristics as well as on deterministic attributes. Likelihood function is presented in a closed form and maximum likelihood estimates (MLEs) of the parameters are calculated. The theoretical results are then applied to real highway accident data and compared with other models in the literature. In our experiments, the proposed probabilistic method considerably outperformed other models.

Download English Version:

https://daneshyari.com/en/article/311262

Download Persian Version:

https://daneshyari.com/article/311262

Daneshyari.com