Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 25 (2017) 664–676

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016

Freeway crash frequency modeling under time-of-day distribution

Yu-Chiun Chiou*, Yu-Chun Sheng, Chiang Fu

Department of Transportation and Logistics Management, National Chiao Tung University, 4F, 118, Sec. 1, Chung-Hsiao W. Rd. Taipei, 100, Taiwan

Abstract

This study aims to identify key factors affecting crash frequencies under various times of the day, so as to propose effective and time-specific countermeasures. Two approaches are proposed and compared. The clustering approach combines a crash count model to predict total number of crashes and a clustering model to divide segments into clusters according to their time-of-day distribution patterns of crash frequency. The multivariate approach treats the crash frequencies of various times of the day as target variables and accommodates potential correlation among them. Crash datasets of Taiwan Freeway No.1 are used to estimate and validate the models. Four times of the day, late-night/dawn (24-06), morning/noon (07-13), afternoon/evening (14-19), and night (20-23) are formed according to crash count distribution. In terms of Adj-MAPE and RMSE, the clustering approach performs better than the multivariate approach. According to the clustering results, segments in metropolitan areas have higher crash frequency in the afternoon/evening, while those in rural areas have higher crash frequency in late-night/dawn, suggesting the time-of-day distributions of crash frequency markedly differ among segments. Time-specific countermeasures are then proposed accordingly.

Keywords: *Time-of-day crash frequency distribution; negative binomial regression; clustering; multivariate modeling approach.*

1. Introduction

Many studies have developed crash frequency models to identify factors contributing to crash counts at roadway segments or at intersections during a certain time period (usually one year) (e.g., Jones et al., 1991; Miaou, 1994;

* Corresponding author. Tel: +886-2-23494940
E-mail address: ycchiou@mail.nctu.edu.tw (Y.C. Chiou)

2352-1465 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.
10.1016/j.trpro.2017.05.450

Fridstrom et al., 1995; Shankar et al., 1995; Poch and Mannering, 1996; Shankar et al., 1997; Milton and Mannering, 1998; Ivan et al., 1999; Ivan et al., 2000; Abdel-Aty and Radwan, 2000; Khattak et al., 2002; Wang and Nihan, 2004; Lord, 2006; Wong et al., 2007; Malyshkina and Mannering, 2010). However, only a few of studies further examined risk factors contributing to crash counts at various times of the day (e.g., morning, afternoon, and night) which can definitely provide more useful information for proposing effective and time-specific countermeasures. For example, Doherty et al. (1998) studied the distribution of crash frequency at times of the day. They showed a very high crash frequency for drivers aged 16 to 19 to drive at night. Clarke et al. (2006) investigated how age, driving experience, and time of day affect the crash frequency of young drivers and the results suggested that the problems of accidents in darkness are not a matter of visibility, but a consequence of the way young drivers use the roads at night. Qin et al. (2006) examined the relationship between crash occurrence and hourly traffic. The results revealed how the relationship between crashes and hourly traffic varies by time of day, thus improving the accuracy of crash occurrence predictions. The results show that even accounting for time of day, the hourly traffic is indeed non-linear of crash occurrence, implying that at any time of day, the crash occurrence is not proportional to the hourly traffic. Marquis (2014) analyzed the truck-related crash occurrences in Manhattan, New York over four time blocks: the morning peak (6:00-10:00), the mid-day (10:00-15:00), the afternoon peak (15:00-19:00), and the night time (19:00-6:00) by using zero-inflated negative binomial models. The study found that both the built environment and the traffic flows contribute to the temporal variation of truck-related crash occurrence.

Most of the above studies examined the effect of time of day on crash counts by introducing a time variable into the model or by modelling crash counts at various time periods separately. The former has difficulty in investigating the different effects of risk factors on crash counts at various time periods and the latter ignores the potential correlation among crash counts at various time periods. Thus, this study aims to simultaneously model the crash counts at various time periods. The model framework is similar to those modelling crash frequencies by severity levels (e.g., property damage only, injury, and fatality) and collision types (e.g., rear-end, head-on, sideswipe, and right angle) or by collision types (Milton et al., 2008; Ye et al., 2009; Naderan and Shahi, 2010; Aguero-Valverde, 2013; Chiou and Fu, 2013; Ye et al., 2013; Chiou et al., 2014; Venkataraman et al., 2014; Chiou and Fu, 2013). The remainder of this paper is organized as follows. Section 2 presents the proposed models. Section 3 addresses data collection and descriptive statistics of the study crash dataset. Section 4 presents the model estimation results and comparisons. Concluding remarks and suggestions are then given in Section 5.

2. Model

To identify the key factors contributing to crash counts at various times of day, two approaches are proposed, clustering and multivariate modeling approaches, as narrated below, respectively.

2.1. Clustering approach

The proposed clustering approach contains two stages. The first stage is to predict total crashes on each segment by using commonly adopted Poisson (PO) and Negative binomial (NB) models. The second stage is to divide freeway segments into finite clusters according to their time-of-day crash frequency distribution patterns. The average time-of-day crash frequency distribution of each cluster is then used to represent the segments belonging to it.

The PO model is the most fundamental count model. The probability function of PO model (Miaou, 1994; Jones et al., 1991) is expressed as Eq. (1).

$$PO(y_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!} \tag{1}$$

where PO (y_i) is the probability of y_i accidents occurring on roadway segment i under a specific time period. λ_i is the expected number of accidents on roadway segment i at the time period, which is defined in non-negative numbers. For estimation purposes, λ_i is usually specified as Eq. (2):

$$\lambda_i = \exp(\beta' X_i) \tag{2}$$