Accident Analysis and Prevention 107 (2017) 92–101

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Modeling the impact of latent driving patterns on traffic safety using mobile sensor data

Rajesh Paleti*, Olcay Sahin, Mecit Cetin

Transportation Research Institute, Old Dominion University, 135 Kaufman Hall, Norfolk, VA 23529, USA

ARTICLE INFO

Keywords:
Mobile sensors
Crash frequency
Latent driving patterns
Generalized ordered response
Spatial dependency
Measurement error
Unobserved heterogeneity

ABSTRACT

Smartphones are now equipped with sensors capable of recording vehicle performance data at a very fine temporal resolution in a cost-effective way. In this paper, mobile sensor data from smartphones was used to identify and quantify unsafe driving patterns and their relationship with traffic crash incidences. Statistical models that account for measurement error associated with microscopic traffic measures computed using mobile sensor data were developed. The models with microscopic traffic measures were shown to be statistically better than traditional models that only control for roadway geometry and traffic exposure variables. Also, generalized count models that account for measurement error, spatial dependency effects, and random parameter heterogeneity were found to perform better than standard count models.

1. Introduction

In the United States, there were 35,092 fatalities as a result of motor vehicle crashes in 2015 and current trends show that an increase of about 8.1 percent is expected in 2016 (Nhtsa, 2015, 2017). In the year 2015, in Virginia alone, 753 people were killed and 65,029 people were injured in a total of 125,800 motor vehicle accidents (Dmv, 2014). This combined with the fact that recent vehicle miles travelled (VMT) estimate of a compound annual growth rate of about 1% through the year 2033 makes traffic safety a matter of great concern (Fhwa, 2015). These crashes not only cause injury and loss of life, but they also cost a considerable amount to the people involved. For instance, in 2010, the economic costs of motor vehicle crashes in the nation totalled $242 billion. These costs come not only from the damage to vehicles and the medical bills of the injured but also include items such as $28 billion due to congestion (Blincoe et al., 2015).

Safety engineers have relied on crash frequency modelling to inform safety policy making concerning prioritization and implementation of countermeasures to improve safety. Crash frequency modelling is an attempt to quantify the expected number of crashes in a certain period (e.g., one year) at a specific location (e.g., roadway segment or intersection) as function of variables describing the location and the traffic conditions at the location. These models are referred to as the Safety Performance Functions (SPFs) in the Highway Safety Manual (HSM) (Aashto, 2010; Farid et al., 2016). In the past, most of these SPFs for roadways only used geometry (e.g., presence of shoulder, median width etc.) and aggregate traffic measures (e.g., traffic volume) as explanatory variables. However, there is limited literature on analysing the correlation between microscopic traffic measures (e.g., high-resolution speed and acceleration) and crash risk. Lack of microscopic traffic data has been the primary impediment for limited research in this direction. In the absence of these microscopic measures, the parameter estimates in the SPFs can be biased and lead to wrong policy implications. For instance, it is possible that in the absence of microscopic traffic measures, the SPFs overestimate the impact of roadway improvements on safety because they confound the effect of driving patterns and the roadway characteristics. Also, the SPFs that lack microscopic traffic measures are not sensitive to countermeasures that are focused on changing the driving patterns (e.g., speed harmonization) rather than geometric features.

The actual driving patterns along any given road are unobserved to the analyst. However, these 'latent' driving patterns may be inferred using microscopic traffic data. For instance, smartphones are now equipped with sensors that are capable of recording highway performance data at a fine temporal resolution in a cost-effective way (Zhen and Qiang, 2014). In fact, several auto insurance firms (e.g., Progressive's Snapshot) have been experimenting with monitoring driving activity (e.g., hard-brakes per mile) through on-board diagnostic (OBD) devices to assess and valuate the crash risk of individual drivers. Recently, (Peng et al., 2017) explored the effect of reduced visibility during foggy conditions on time to collision (TTC) using real-time data collected using a new visibility and vehicle detection system. However, there is no significant research on investigating the potential use of high-resolution data from mobile sensors of smartphones in

* Corresponding author.
E-mail addresses: rpaleti@odu.edu, rajeshpaleti2014@gmail.com (R. Paleti).

http://dx.doi.org/10.1016/j.aap.2017.08.012
Received 3 March 2017; Received in revised form 29 June 2017; Accepted 7 August 2017

understanding crash risks and safety measures for highway sections. The current study aims to make use of smartphone sensors to extract microscopic traffic measures that can serve as indicators of latent driving patterns and test the relationship between these microscopic traffic measures and crash frequency along freeway segments. This study does not aim to predict crash occurrences in real-time but rather examines the relationship between annual crash occurrences and microscopic traffic data. Also, it is important to note that the current research focusses on average driving patterns across all drivers but not individual road users. So, the smartphone sensor data was used to compute microscopic traffic measures that serve as surrogates of these average driving patterns that are subsequently correlated with crash occurrences. From a methodological standpoint, the statistical models developed in this study account for two key aspects central to this modeling effort. First, the microscopic traffic measures are only indicators of real driving patterns. So, latent variable modeling techniques were used to capture the relationship between driving patterns and crash frequency. Second, driving patterns along a given road will most likely depend on the driving conditions in the close vicinity causing spatial dependency. This translates into allowing for spatial dependency of latent variables in crash frequency models.

To start-off, mobile sensor data was collected by driving along major roadways in the Hampton Roads region. Next, this data was overlaid on the transportation network to map probe data and the roadway segments. Then, several speed and acceleration metrics were calculated for each roadway using the mobile sensor data. Subsequently, these metrics were appended to the Virginia Department of Transportation (VDOT) crash data for the past one year. Supplementary data sources were used to assemble information regarding roadway inventory data and traffic exposure information. Next, statistical model estimation was undertaken to quantify the relationship between microscopic traffic measures and crash incidences along major interstates in Hampton Roads.

This paper is organized as follows. The next section describes the past literature followed by a description of the methodology and models employed in this paper. This is followed by the empirical application and model results. Conclusions and potential future work are presented at the end.

2. Literature review

Crashes are rare and random events. So, the number of observed crashes at any given location can fluctuate year-to-year even if all the observable crash causation conditions remain the same between the two years. If the observed crash frequency is very high in one year, then it is more likely to be followed by relatively lower crash frequency in the next year, and vice-versa. This effect is referred to as the 'Regression-To-Mean Bias'. This inherent variation in observed crash frequency poses a challenge to evaluating the effectiveness of different safety countermeasures. For instance, it is unclear if the reduction (or increase) in crash occurrences is due to random fluctuation or the safety countermeasure. To address this problem, safety analysts rely on estimates of the long term average crash frequency, also referred to as 'Expected Crash Frequency', as a proxy for crash risk. The observed crash frequency across several locations is used to statistically estimate the expected crash frequency. Expected crash frequency modelling is a reliable method for determining the safety of a segment of roadway.

Previous studies have looked at explanatory variables primarily in two categories, physical characteristics of the location (e.g., roadway or interchange) and aggregate traffic characteristics at that location (e.g., AADT, % of left turning traffic, % of heavy vehicle traffic etc.) (Shankar et al., 1997; Qin et al., 2005; Lord and Mannering, 2010). A majority of these early studies focused on physical characteristics of the roadway due to a lack of consistent and accurate data collection means (Ogle, 2005). Unfortunately, even though such aggregate data may have some correlation with driving patterns it is unable to capture all actual driving patterns (i.e., flow and movement of individual vehicles and their accelerations). It is difficult to develop an accurate representation of expected crash frequencies when the characteristics of the actual vehicles travelling the corridor are not considered. For instance, the overall congested crash rate in the state of Indiana is 24.1 times greater than the uncongested crash rate (Mekker et al., 2016). In addition to higher traffic volumes, there are most likely unique driving patterns that contributed to high crash rates during congested period. Simple aggregate measures (average daily traffic and truck volumes) cannot capture these differences between congested and uncongested conditions. In order to capture these differences, safety studies that used more disaggregate time-periods controlling for factors such as average hourly traffic were conducted (Zhou and Sisiopiku, 1997; Chang et al., 2000; Lord et al., 2005a). Also, real-time studies that used traffic data from loop sensors or microwave vehicle detection systems were developed to predict the probability of crash occurrence along freeway mainlines (Lee et al., 2002; Abdel-Aty et al., 2004; Pande et al., 2005), ramps (Lee and Abdel-Aty, 2006; Wang et al., 2015b), and weaving segments (Wang et al., 2015a) in real-time.

A potential source for speed data could be crash reports that were completed at the scene of an accident by the police. This would appear to be a simple way to obtain a piece of driving patterns. But, obtaining speed from a police crash report is not recommended because the police may be under a lot of stress during incident investigations and may not be able to accurately determine the speed at which the driver was going. Also, the driver may underreport the estimated speed which they were travelling in an attempt to lessen the likelihood of receiving additional infractions for an incident. Alternatively, several researchers have used speed limit as a proxy for traffic speed. Probe vehicle data, on the other hand, can be used to capture the speed and acceleration profiles that serve as reliable indicators of congested traffic conditions. Some of the previous studies have relied on simulation models to capture naturalistic driving data regarding the movements of the actual vehicle itself through space and time (Gettman and Head, 2003). This method of data collection allows the researcher to control for every aspect of the simulation while being able to alter the simulation to fit different scenarios. Multiple simulation inputs may be evaluated in a short period of time to get the most accurate results. A limitation of these methods, however, is that it is based on simulation and not driving behaviour in reality.

Recent studies have focused on obtaining and using data collected directly in the field to develop more accurate crash frequency models. GPS sensors and OBD devices are now regularly used in transportation research to obtain the aforementioned naturalistic driving behaviour data (Ogle, 2005; Jun, 2006). Another option when considering probe vehicle data is using data that is crowd-sourced, collected, and combined into a dataset by a third party source (Mekker et al., 2016). This data source has the benefit of allowing the researchers to have a more robust dataset that encompasses a greater length of time. The data can be collected and stored for multiple years rather than only being available for the duration of research period. This allows the researcher to have access to probe vehicle data that was collected around the time that actual accidents occurred. (Wåhlberg, 2004) looked at the acceleration profiles of busses as a potential indicator of crash frequency and study concluded that driver acceleration behaviour could be used as a predictor of accidents. But, due to some discrepancies between samples it was difficult to determine the validity of this finding. Also, in this study, the acceleration data was recorded on-board using a g-analyst which measured the acceleration at 10 Hz to 100th of 1 g (9.81 m/s2) accuracy. This tool did not measure the acceleration from the vehicle directly but, simply measured the g-force felt by the bus's start and stop motions. This may have resulted in errors due to the vehicle not producing the data itself. In the literature, there are studies that focus on registering driving patterns with a data logging equipment to study environmental effects of traffic such as emissions estimation (Höglund and Niittymäki, 1999; Rodríguez et al., 2016).

To summarize, past literature highlighted the importance of