Accident Analysis and Prevention 115 (2018) 160–169

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Vehicle manoeuvers as surrogate safety measures: Extracting data from the gps-enabled smartphones of regular drivers

Joshua Stipancic[a,*], Luis Miranda-Moreno[b], Nicolas Saunier[c]

[a] Department of Civil Engineering and Applied Mechanics, McGill University, Room 391, Macdonald Engineering Building, 817 Sherbrooke Street West, Montréal, Québec, H3A 0C3, Canada
[b] Department of Civil Engineering and Applied Mechanics, McGill University, Room 268, Macdonald Engineering Building, 817 Sherbrooke Street West, Montréal, Québec, H3A 0C3, Canada
[c] Department of Civil, Geological and Mining Engineering, Polytechnique Montréal, C.P. 6079, succ. Centre-Ville, Montréal, Québec, H3C 3A7, Canada

ARTICLE INFO

Keywords:
Surrogate safety
Smartphone
GPS
Urban
Probe vehicles
Vehicle manoeuvres
Naturalistic data

ABSTRACT

Network screening is a key element in identifying and prioritizing hazardous sites for engineering treatment. Traditional screening methods have used observed crash frequency or severity ranking criteria and statistical modelling approaches, despite the fact that crash-based methods are reactive. Alternatively, surrogate safety measures (SSMs) have become popular, making use of new data sources including video and, more rarely, GPS data. The purpose of this study is to examine vehicle manoeuvres of braking and accelerating extracted from a large quantity of GPS data collected using the smartphones of regular drivers, and to explore their potential as SSMs through correlation with historical collision frequency and severity across different facility types. GPS travel data was collected in Quebec City, Canada in 2014. The sample for this study contained over 4000 drivers and 21,000 trips. Hard braking (HBEs) and accelerating events (HAEs) were extracted and compared to historical crash data using Spearman's correlation coefficient and pairwise Kolmogorov-Smirnov tests. Both manoeuvres were shown to be positively correlated with crash frequency at the link and intersection levels, though correlations were much stronger when considering intersections. Locations with more braking and accelerating also tend to have more collisions. Concerning severity, higher numbers of vehicle manoeuvres were also related to increased collision severity, though this relationship was not always statistically significant. The inclusion of severity testing, which is an independent dimension of safety, represents a substantial contribution to the existing literature. Future work will focus on developing a network screening model that incorporates these SSMs.

1. Introduction

The safety of urban road networks is a serious concern that requires the continuous monitoring of crash risk. Considering that parties involved in improving road safety have finite budgets, the common approach is to identify the most dangerous sites in the network and to prioritize them for remediation in order to maximize the efficiency of countermeasures. In this process, known as network screening, candidate high-risk sites are identified as those locations where design or operation "create an increased risk of unforeseeable accidents" (Agerholm et al., 2012). Traditional screening methods have used observed crash frequency or severity ranking criteria, despite the fact that crash-based methods are reactive (Agerholm et al., 2012), require long collection periods (Lee et al., 2006), are subject to errors in collision databases, and are sensitive to underreporting (Kockelman and Kweon, 2002). A screening method to replace the crash-based approach requires a new data source from which crash risk can be computed. Naturalistic driving data is collected unobtrusively in crashes, near crashes, and normal conditions, provides information difficult to observe by other techniques (Bagdadi, 2013; Wu and Jovanis, 2013), and supports safety assessments based on surrogate safety measures (SSMs) rather than crash data. SSMs are any non-crash measures that are physically and predictably related to crashes (Tarko et al., 2009) and have the potential to reduce dependency on crash data in network screening (Laureshyn et al., 2009).

Naturalistic approaches typically yield large volumes of data from which surrogate indicators must be identified (Bagdadi, 2013). Popular methods for surrogate safety analysis include event-based techniques, behavioural techniques, and techniques based on measures of traffic flow. Event-based techniques consider traffic conflicts, interactions between road users, or vehicle manoeuvres. Traffic conflicts were first studied in the late 1960s based on human observation. Though data

* Corresponding author.
E-mail addresses: joshua.stipancic@mail.mcgill.ca (J. Stipancic), luis.miranda-moreno@mcgill.ca (L. Miranda-Moreno), nicolas.saunier@polymtl.ca (N. Saunier).

https://doi.org/10.1016/j.aap.2018.03.005
Received 3 February 2017; Accepted 4 March 2018

volumes were limited and analysis potentially subjective (Laureshyn et al., 2009), human observation provided a level-of-detail beyond what was possible through objective (crash-based) techniques (Laureshyn et al., 2009). Video-based sensors and computer vision techniques have improved objectivity and increased the amount of data that can be processed. Though video-based sensors provide high temporal resolution (Agerholm et al., 2012) and rich positional data beyond counts and speed (Bahler et al., 1998), the analysis of video data is potentially resource intensive and has spatial limitations (Laureshyn et al., 2009), leading to a desire to implement event-based techniques using other data sources. Behavioural techniques aim to identify individual driver behaviours not related to conflict or crash avoidance, such as infractions and yielding (Dingus et al., 2006). Traffic flow techniques, which use measures of volume, mean speed, or density to estimate risk (Yan et al., 2008), typically require roadside point sensors such as loops, radar, or other sensors (Oh et al., 2001; Golob et al., 2004; Lee et al., 2002). Though successful on freeways, it is impractical and costly to implement roadside sensors across an entire urban network (Herrera et al., 2010), and traffic flow measures have yet to be proven as reliable SSMs in urban networks with at-grade intersections.

For SSMs to be useful in screening applications, data must be captured continuously as drivers move through the network. Thanks to the advent of instrumented vehicles, extracting SSMs across an urban network is now possible. Instrumented vehicles (or probe vehicles) act "as moving sensors, continuously feeding information about traffic conditions" (El Faouzi et al., 2011). GPS devices are reliable sources of naturalistic driving data (Jun et al., 2007) and may be complemented by additional vehicle kinematics from accelerometers or gyroscopes and environmental factors collected by external sensors such as radars. These sensors provide long periods of continuous data for a relatively small sample of road users (Agerholm et al., 2012). Though limited in terms of the studied population of drivers, the spatial coverage of GPS data makes it ideal for network screening applications. Smartphones are inexpensive, simple, and user-friendly data collection devices that eliminate the need for external sensors (Eren et al., 2012; Johnson and Trivedi, 2011). Despite the advent of smartphone GPS data in the last years, few studies have investigated its use in the network screening process. Accordingly, the purpose of this study is to examine vehicle manoeuvres extracted from probe vehicle data collected by the GPS-enabled smartphones of regular drivers. More specifically, this research contributes to the field by introducing a methodology for collecting, cleaning, and analyzing GPS travel data, extracting vehicle manoeuvres from the GPS data of regular drivers, and investigating the relationship between vehicle manoeuvres and historical collision frequency and severity across different facility types.

2. Literature review

Although probe vehicles have primarily been used in spatio-temporal applications such as traffic monitoring and origin-destination studies (Herrera et al., 2010), they have also been applied less frequently, in studies of road safety. This underutilization can be partially attributed to the difficulties of collecting large volumes of data using dedicated GPS devices that are installed for a specific research purpose. Studies using probe vehicles with dedicated GPS tracking devices must overcome low penetration rates which may be insufficient for providing "an exhaustive coverage of the transportation network" (Herrera et al., 2010). This shortcoming has been apparent in the field of automated incident detection (AID), which involves the identification of "non-recurring events such as accidents" through pattern classification of traffic flow (Dia and Thomas, 2011). In AID studies, dedicated GPS devices have been supplemented by roadside sensors (El Faouzi et al., 2011) or simulation. Sethi et al. (1995) found that probe vehicles were useful in AID, though only when using a proportion of probe vehicles beyond what could be expected in practice (Sethi et al., 1995). Dia and Thomas (2011) found the best results when probe vehicles comprised 20% of the traffic flow.

Despite this, several studies have attempted to extract vehicle manoeuvres from probe vehicles as SSMs. Jun et al. (2007) analyzed spatio-temporal driving activity and crash involvement using dedicated GPS devices and self-reported safety data for 460 light-duty vehicles. The study found that drivers involved in crashes tended to travel longer distances and at higher speeds and "engaged in hard deceleration events" (greater than $2.7\,m/s^2$) more frequently (Jun et al., 2007). Though failing to show a causal link, the authors suggest that decelerations 'may be employed as roadway safety surrogate measures' (Jun et al., 2007). This study highlights an additional shortcoming of dedicated devices; that drivers behave more safely when monitored (Johnson and Trivedi, 2011). In studies using dedicated devices, installations are often biased towards a specific segment of the population (such as light-duty vehicle drivers or taxi drivers). Ellison et al. (2013) studied 106 drivers using dedicated GPS devices along with demographic surveys for each driver. By controlling for temporal and spatial factors including geometry, weather, time of day, trip purpose, and vehicle occupancy, the authors found that the road environment was a significant influencer of driver behaviour (Ellison et al., 2013). Although speed is often regarded as an important surrogate measure, changes in speed (acceleration, the first derivative of velocity, or jerk, the second derivative) may be more important (Laureshyn et al., 2009). Agerholm et al. (2012) collected data from six drivers over a 3-month period using GPS devices and accelerometers. The authors stated that 'braking was the evasive action […] in 88% of the accidents in built-up areas' (Agerholm et al., 2012), making decelerations a logical indicator to extract. Jerk was found to be correlated with accident occurrence (Agerholm et al., 2012). Bagdadi (2013) noted that the most common crashes are rear-end collisions. The study used GPS, accelerometer, and radar data from 109 participants and found that jerk could correctly identify self-reported near misses at an 86% success rate (Bagdadi, 2013).

The use of smartphones as a potentially rich source of safety data became popular in the early 2010s. Johnson and Trivedi (2011) developed a system to distinguish non-aggressive and aggressive driving behaviour. Their system fused accelerometer, gyroscope, magnetometer, GPS, and video data from smartphones to monitor drivers. However, at the time of publication, the system had only been installed in three vehicles. Eren et al. (2012) similarly studied manoeuvering using the smartphone accelerometer and gyroscope data of 15 drivers. Guido et al. (2012) attempted to evaluate time-to-collision (TTC) and deceleration rate as measured from smartphone GPS data as possible SSMs for rear-end collisions on a two-lane rural highway. The study used only three drivers and no attempt was made to correlate the results to actual collision risk. Fazeen et al. (2012) used smartphone accelerometer data to classify 'safe' accelerations and decelerations from 'unsafe' ones (approximately $3\,m/s^2$ or greater), though failed to demonstrate whether 'unsafe' behaviour led to increased collision risk and used only a single smartphone.

Several shortcomings are apparent in the existing literature, which this study attempts to address. First, there has been no attempt to derive SSMs from smartphone-collected GPS data of regular drivers alone. Existing studies have used dedicated probe vehicles (resulting in sample sizes of 100 drivers or less) or dedicated GPS devices with supplemental accelerometer data. Studies using smartphones have used extremely few drivers, despite the potential for application to the population at large. Second, there has been no comprehensive comparison of GPS-based SSMs to large quantities of crash data at the network scale. Instead, studies have compared indicators to sample safety data, which is often self-reported.