Transportation Research Part C 95 (2018) 431–441

Contents lists available at ScienceDirect

Transportation Research Part C

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

Where are the dangerous intersections for pedestrians and cyclists: A colocation-based approach

Yujie Hu[a,*], Yu Zhang[b], Kyle S. Shelton[c]

[a] School of Geosciences, University of South Florida, Tampa, FL 33620, USA
[b] Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL 33620, USA
[c] Kinder Institute for Urban Research, Rice University, Houston, TX 77005, USA

A R T I C L E I N F O

Keywords:
Traffic safety
Safety countermeasures
Spatial patterns
Colocation
Houston

A B S T R A C T

Pedestrians and cyclists are vulnerable road users. They are at greater risk for being killed in a crash than other road users. The percentage of fatal crashes that involve a pedestrian or cyclist is higher than the overall percentage of total trips taken by both modes. Because of this risk, finding ways to minimize problematic street environments is critical. Understanding traffic safety spatial patterns and identifying dangerous locations with significantly high crash risks for pedestrians and cyclists is essential in order to design possible countermeasures to improve road safety. This research develops two indicators for examining spatial correlation patterns between elements of the built environment (intersections) and crashes (pedestrian- or cyclist-involved). The global colocation quotient detects the overall connection in an area while the local colocation quotient identifies the locations of high-risk intersections. To illustrate our approach, we applied the methods to inspect the colocation patterns between pedestrian- or cyclist-vehicle crashes and intersections in Houston, Texas and we identified among many intersections the ones that significantly attract crashes. We also scrutinized those intersections, discussed possible attributes leading to high colocation of crashes, and proposed corresponding countermeasures.

1. Introduction

Active transportation modes such as walking and bicycling are critical elements of a sustainable urban transportation system. Many U.S. cities, including Houston, are grappling with how to reduce vehicular traffic, energy consumption, and greenhouse gas emissions that come with car travel. Beyond their clear environmental benefits, the large-scale embrace of active modes of travel could lead to significant gains in public health (Krizec, 2007; Miranda-Moreno et al., 2011).

Automobile-centric street design of most U.S. cities, however, leaves pedestrians and cyclists vulnerable. For example, pedestrians were reported to be 23 times more likely to be killed than vehicle occupants in terms of distance traveled (Pucher and Dijkstra, 2003; Moudon et al., 2008). According to the National Highway Traffic Safety Administration (2015), an average of 4500 pedestrians were killed and 66,000 injured in automobile-related crashes each year between 2004 and 2013. Nationally, pedestrians are involved in 14% of all fatality crashes, despite the fact that walking remains a minimal part of the overall share of trips taken (National Highway Traffic Safety Administration, 2012a). As for cyclists, the percentage of cyclist fatalities increased from 1.5% in 2003 to 2.2% in 2012 (National Highway Traffic Safety Administration, 2012b).

The majority of pedestrian/cyclist crash injuries and fatalities occurred in cities, where activity is concentrated and where cycling

* Corresponding author at: School of Geosciences, University of South Florida, 4202 E Fowler Ave, NES 107, Tampa, FL 33620, USA.
E-mail address: yhu@usf.edu (Y. Hu).

https://doi.org/10.1016/j.trc.2018.07.030
Received 20 November 2017; Received in revised form 29 June 2018; Accepted 27 July 2018
0968-090X/ Published by Elsevier Ltd.

and walking rates are at their highest (National Highway Traffic Safety Administration, 2008). Due in part to high transportation costs, re-urbanization is occurring in metro regions across the U.S. and more people are moving back to cities. As urban populations grow and the streetscape maintains the same car-centric design, these vulnerable road users will still likely face high crash risks.

Some studies have employed density analysis to look for crash clustering patterns and to identify the most dangerous locations (e.g., intersections) based on historical crash incidents. One such example is to plot the histogram of absolute value of crashes or crash rates and take the locations with statistically high values as the most dangerous locations (e.g., Pulugurtha et al., 2007; Anderson, 2009; Ferreira and Couto, 2015). Another set of studies take that approach a step further by focusing on understanding the connection between elements of the built environment (e.g., intersections) and crashes by using macro-level regression analysis (e.g., Moudon et al., 2008; Ukkusuri et al., 2012). Most of these studies pre-define the connection between crashes and intersections by assigning crashes to intersections and take one intersection as one data point in the regression analysis. Both these methods have limitations that, if used to inform policy or funding, may lead to ineffective safety countermeasures being implemented.

Using a technique called colocation pattern mining within Geographic Information Systems (GIS), this paper examines the spatial correlation pattern between crashes and intersections to offer a novel way to understand at which intersections crashes are more likely to occur. To illustrate our approach, two colocation measures, global and local, are used to analyze the spatial relationship between crash incidents and intersections in the city of Houston based on the pedestrian- and cyclist-vehicle crash records between 2010 and 2016. Our analysis provides an alternative way to understand the spatial links between crash incidents and intersections. In particular, our local colocation method is able to present stakeholders with site-specific patterns and identify the most dangerous locations.

2. Literature review

Although traffic crashes have been widely studied by the transportation research community, crashes that involve automobiles and either pedestrians or cyclists are still understudied (Ukkusuri et al., 2012). Most existing research in this area has relied on clustering analysis. For example, Pulugurtha et al. (2007) employed the kernel density estimation (KDE) technique to detect pedestrian crash zones, and Anderson (2009) discovered and classified crash hotspots for pedestrians, cyclists, and vehicles, based on KDE and K-means clustering approach. However, Dai and Jaworski (2016) criticized the application of the above planar KDE in analyzing traffic accidents that occur in a linear space like a road, and instead offered a network-based KDE method to locate pedestrian crash hotspots. Other clustering methods used to detect pedestrian or cyclist crash hotspots in past research include the global and local spatial autocorrelation (e.g., Truong and Somenahalli, 2011), the spatiotemporal Bernoulli model (Dai, 2012), and the binary Probit model (Ferreira and Couto, 2015).

Another line of crash research shifted the focus from locating crash hotspots to studying the links between the urban built environment and crashes. Commonly examined built environment elements include urban land use type or mix, road network characteristics, and accessibility to public transit systems and educational facilities. Typical techniques for analyzing the links between crashes and the built environment are regression models such as the negative binomial model, the log-linear model, the Poisson model, and the Bayesian model (e.g., Li et al., 2007; Moudon et al., 2008; Pulugurtha and Sambhara, 2011; Ukkusuri et al., 2012; Chen and Zhou, 2016). For example, Loukaitou-Sideris et al. (2007) reported that areas proximal to educational facilities and with higher fractions of commercial land use are correlated with more pedestrian crashes. Similarly, Ukkusuri et al. (2012) found that census tracts with denser commercial or industrial land use types had more crashes than those with a greater portion of residential land use. In addition, they discovered that road characteristics such as the width and the number of lanes are other related factors for pedestrian crashes. Chen (2015) revealed that traffic analysis zones (TAZ) with a greater number of road signals and street parking signs were likely to see more bicycle crashes. Further, Chen and Zhou (2016) reported that higher densities of four-way and signalized intersections, mixed land use, and more bus stops in a TAZ would result in a higher pedestrian crash frequency; in contrast, the density of sidewalk was negatively related to the crash frequency. Besides purely built environment elements, some studies also considered interactions between built environment variables and socio-demographic characteristics such as proportions of different age groups (e.g., children and senior population) or race-ethnicity groups like the black and Hispanic population (Cho et al., 2009; Ukkusuri et al., 2012; Chen, 2015).

Despite the increasing attention given to studying the impacts of the built environment on crashes, most of these efforts are still focused on the macro level–TAZs, census tracts or even counties (e.g., Loukaitou-Sideris et al., 2007; Wier et al., 2009; Huang et al., 2010; Ukkusuri et al., 2012; Narayanamoorthy et al., 2013; Chen, 2015; Chen and Zhou, 2016). Analyses at the micro level such as individual intersections can provide more detailed and localized patterns than area-wide analyses. Such close looks can help decision-makers propose site-specific strategies for safety improvements (Son et al., 2011; Ukkusuri et al., 2012). In addition, analyses at a disaggregated level can lead to more consistent and reliable results due to the well-known modifiable areal unit problem (MAUP) in geospatial studies, which would cause inconsistent findings in quantitative measures and statistical tests when analyzing problems at different scales such as TAZs and tracts (e.g., Hu and Wang, 2015; 2016). A few studies have attempted to analyze the impacts of the built environment on the crash frequency at the micro level using statistical approaches (e.g., Lee and Abdel-Aty, 2005; Lee et al., 2017; Murphy et al., 2017); however, they only focused on a sample of signalized intersections. Additionally, the statistical approaches—which model the crash frequency by taking crash site attributes as explanatory variables—are often used to infer causation from the association and to predict future trends by controlling for certain attributes. Lord and Mannering (2010) reviewed more than one hundred papers that applied statistical approaches to model crash-frequency data and discussed the strengths and limitations of different statistical models (refer to Table 2 on page 295 of Lord and Mannering, 2010). They concluded that although crash-frequency data analysis has been advanced over the years, the work was still inherently limited by the available data. To impel the