Journal of Transport Geography 54 (2016) 248–256

Contents lists available at ScienceDirect

Journal of Transport Geography

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

Macro and micro models for zonal crash prediction with application in hot zones identification

Helai Huang [a], Bo Song [a], Pengpeng Xu [b], Qiang Zeng [c,*], Jaeyoung Lee [d], Mohamed Abdel-Aty [d]

[a] Urban Transport Research Center, School of Traffic and Transportation Engineering, Central South University, Changsha, Hunan 410075, PR China
[b] Department of Civil Engineering, The University of Hong Kong, Pokfulam Road, Hong Kong, China
[c] School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong 510641, PR China
[d] Department of Civil, Environment & Construction Engineering, University of Central Florida, Orlando, FL, USA

A R T I C L E I N F O

Article history:
Received 29 March 2015
Received in revised form 9 June 2016
Accepted 13 June 2016
Available online xxxx

Keywords:
Crash prediction model
Zonal safety analysis
Bayesian inference
Spatial correlation
Conditional autoregressive model
Integrated screening

A B S T R A C T

Zonal crash prediction has been one of the most prevalent topics in recent traffic safety research. Typically, zonal safety level is evaluated by relating aggregated crash statistics at a certain spatial scale to various macroscopic factors. Another potential solution is from the micro level perspective, in which zonal crash frequency is estimated by summing up the expected crashes of all the road entities located within the zones of interest. This study intended to compare these two types of zonal crash prediction models. The macro-level Bayesian spatial model with conditional autoregressive prior and the micro-level Bayesian spatial joint model were developed and empirically evaluated, respectively. An integrated hot zone identification approach was then proposed to exploit the merits of separate macro and micro screening results. The research was based on a three-year dataset of an urban road network in Hillsborough County, Florida, U.S.
Results revealed that the micro-level model has better overall fit and predictive performance, provides better insights about the micro factors that closely contribute to crash occurrence, and leads to more direct countermeasures. Whereas the macro-level crash analysis has the advantage of requirement of less detailed data, providing additional instructions for non-traffic engineering issues, as well as serving as an indispensable tool in incorporating safety considerations into long term transportation planning. Based on the proposed integrated screening approach, specific treatment strategies could be proposed to different screening categories. The present study is expected to provide an explicit template towards the application of either technique appropriately.

1. Introduction

Crash prediction model (CPM) is an essential tool in traffic safety analysis. Numerous studies have been conducted to evaluate the safety level of various types of road entities, to identify hotspots or sites with promise, and to find appropriate countermeasures. Recently, an increasing research effort is being focused on a higher aggregated level of crash analysis, which could be referred to as zonal CPM. Traffic crashes are typically aggregated at a certain spatial scale and researchers usually seek to relate safety to zone-level covariates. These macro-level CPMs may aid transportation agencies in more proactively incorporating safety consideration into long term transportation planning process (Washington et al., 2006).

Last decade has witnessed fast growing scope of scientific research to investigate crash propensity on macroscopic levels. Different area-wide characteristics were considered, including road characteristics such as intersections density (Huang et al., 2010; Xu and Huang, 2015), road length with different speed limit (Abdel-Aty et al., 2011; Siddiqui et al., 2012), road length with different functional classification (Quddus, 2008; Hadayeghi et al., 2010), junctions and roundabouts (Quddus, 2008); traffic patterns such as traffic flow and vehicle speed (Quddus, 2008; Hadayeghi et al., 2010); trip generation and distribution (Abdel-Aty et al., 2011; Dong et al., 2014, 2015); environment conditions such as total precipitation/snowfall, and number of rainy/snowy days per year (Aguero-Valverde and Jovanis, 2006); land use (Pulugurha et al., 2013); and socioeconomic factors such as population density (Huang et al., 2010; Siddiqui et al., 2012), age cohorts (Aguero-Valverde and Jovanis, 2006; Dong et al., 2015; Hadayeghi et al., 2010), household incomes (Xu and Huang, 2015) and employment (Quddus, 2008; Hadayeghi et al., 2010).

A wide array of spatial units have been employed, such as regions (Washington et al., 1999), counties (Miaou et al., 2003;

* Corresponding author.
E-mail addresses: huanghelai@csu.edu.cn (H. Huang), songbo@csu.edu.cn (B. Song), pengpengxu@yeah.net (P. Xu), zengqiang@scut.edu.cn (Q. Zeng), jaeyoung@knights.ucf.edu (J. Lee), M.Aty@ucf.edu (M. Abdel-Aty).

http://dx.doi.org/10.1016/j.jtrangeo.2016.06.012
0966-6923/© 2016 Published by Elsevier B.V.

Aguero-Valverde and Jovanis, 2006; Huang et al., 2010; Li et al., 2013), districts (Haynes et al., 2007), wards (Quddus, 2008), zip codes (Girasek and Taylor, 2010; Lee et al., 2014a), census tracts (Ukkusuri et al., 2011, 2012; Wang and Kockelman, 2013), block groups (Levine et al., 1995), and traffic analysis zones (i.e. TAZs[1]; Hadayeghi et al., 2010; Abdel-Aty et al., 2011; Siddiqui et al., 2012; Pulugurha et al., 2013; Wang et al., 2013; Dong et al., 2014, 2015; Xu and Huang, 2015). Among them, TAZs are now the only traffic-related zone system and are superior in being easily integrated with the transportation planning process, thus having been widely adopted.

However, the efficiency of macro-level traffic safety analysis may be subject to the well-known modifiable areal unit problem (Abdel-Aty et al., 2013; Lee et al., 2014b; Xu et al., 2014) and boundary issue (Siddiqui and Abdel-Aty, 2012; Cui et al., 2015). From another perspective, the safety problem is anyhow a microscopic problem and the direct contributing factors could be related to micro-level factors for a specific road segment or intersection, or the driver-vehicle units involved. In additional to macro level CPMs, an alternatively potential solution estimating zonal safety situation is to sum up crash predictions of all entities (i.e. road segments and intersections) located within the zones of interest, which could be regarded as micro level CPMs.

As road entities located in close proximity may share confounding factors, spatial correlation (or spatial dependency) tends to be a major concern. Research demonstrated that the consideration of spatial effect of adjacent road segments in crash prediction contributes to an unbiased parameter estimation, and significantly improves model predictive performance (Wang et al., 2009; Aguero-Valverde and Jovanis, 2010; Xie et al., 2013). Nevertheless, previous studies are mostly limited to individual types of road entities, i.e. either intersection or segments. Undoubtedly, spatial correlation exists not only between adjacent road segments or between adjacent intersections, but also, even more importantly, between road segments and their connected intersections. To this end, Zeng and Huang (2014) proposed a Bayesian spatial joint approach to simultaneously model crash frequency of intersections and the feeding road segments. Results revealed that the spatial correlation between segments and the connected intersections are more significant than those solely between segments or between intersections.

To our knowledge, there is no research comparing macro-level and micro-level models in predicting zonal safety levels. A comparative analysis could be interesting and beneficial to reveal the associations and differences between those two methods, as well as to provide an explicit template towards the application of either technique appropriately.

This study intends to empirically compare two types of zonal CPMs by evaluating model fitting and predictive performance, as well as identifying crash hot zones. Two state-of-art methods, i.e. the macro level Bayesian spatial model with conditional autoregressive (CAR) prior and micro level Bayesian spatial joint model are developed and empirically evaluated, respectively. The analysis is based on an urban road network with 346 segments and 198 intersections of 155 TAZs in Hillsborough County, Florida, U.S.

2. Methodology

2.1. Bayesian spatial model with CAR prior

Traditional CPMs such as Poisson lognormal model and negative binomial model have largely ignored the issue of possible spatial correlation of traffic crashes among adjacent zones, which would be misleading as this cannot reflect the true underlying data generating process (Huang and Abdel-Aty, 2010). For this reason, by incorporating an error term followed by the CAR prior into the link function, the Bayesian spatial model with CAR prior has been widely applied in current macro-level crash prediction (Miaou et al., 2003; Aguero-Valverde and Jovanis, 2006; Quddus, 2008; Huang et al., 2010; Siddiqui et al., 2012; Wang et al., 2013; Dong et al., 2014; Xu et al., 2014; Xu and Huang, 2015).

Let Y_{it} denote the number of crashes in TAZ i during period t (in years), e_{it} the exposure function, $\mathbf{X_{it}}$ the vector of explanatory variables. The CAR model could be expressed as:

$$\begin{aligned} & Y_{it} \sim \text{Poisson}(\lambda_{it}) \\ & \log\lambda_{it} = \log e_{it} + \mathbf{X}'_{\mathbf{it}}\boldsymbol{\beta} + \theta_i + \phi_i \end{aligned} \tag{1}$$

where λ_{it} is the parameter of Poisson model (i.e. the expected crash frequency in TAZ i at time t), $\boldsymbol{\beta}$ is the vector of coefficient estimates. The variables of daily vehicle miles traveled (DVMT) and total population are simultaneously selected as the measures of exposure, as the model with combination of exposure variables outperforms the counterpart with a single exposure measure (Pirdavani et al., 2012; Lee et al., 2015). Hence:

$$e_{it} = \text{DVMT}_{it}^{\alpha_1} \times \text{POP}_{it}^{\alpha_2} \tag{2}$$

in which $\text{DVMT}_{it}^{\alpha_1}$ and $\text{POP}_{it}^{\alpha_2}$ denote DVMT and total population with coefficients α_1 and α_2, respectively.

θ_i is a random effect to account for the unstructured over-dispersion, which is specified via an ordinary, exchangeable normal prior:

$$\theta_i \sim \text{normal}\left(0, \frac{1}{\tau_h}\right) \tag{3}$$

where τ_h is the precision parameter (reciprocal of the variance), which follows a prior gamma (0.5, 0.0005).

For the spatial correlation term ϕ_i, the intrinsic CAR prior proposed by Besag et al. (1991) is adopted:

$$\phi_i \sim N\left(\frac{\sum_{i \neq j}\omega_{ij}\phi_j}{\sum_{i \neq j}\omega_{ij}}, \frac{1}{\tau_c \sum_{i \neq j}\omega_{ij}}\right) \tag{4}$$

in which ϕ_i is a conditional random variable with the CAR prior, which is used to account for the spatial correlation among adjacent TAZs. ω_{ij} is the binary entries of proximity matrix, and if i and j have a shared border, $\omega_{ij} = 1$, otherwise, $\omega_{ij} = 0$. τ_c is the precision parameter also assumed to be a prior gamma (0.5, 0.0005) as suggested by Wakefield et al. (2000).

The proportion of variability in the random effects due to spatial correlation is of interest:

$$\alpha = \frac{\text{sd}(\phi)}{\text{sd}(\phi) + \text{sd}(\theta)} \tag{5}$$

where sd is the empirical marginal standard deviation function.

2.2. Bayesian spatial joint model

In a road network, compared with spatial correlation solely between segments or between intersections, the effect between adjacent segments and intersections may be more significant as they directly connect with each other. Since intersections and road segments necessarily have different sets of risk factors, the joint model employs an indicator variable r_i to suggest whether the road

[1] TAZ is the geography unit defined delineated by the state Department of Transportation and/or local Metropolitan Planning Organizations (MPOs) for collecting and reporting traffic-related census data (e.g., journey-to-work statistic) in United States. The spatial extent of zones typically varies in models, ranging from very large areas in the exurb to as small as city blocks or buildings in central business districts. Most MPOs who conduct metropolitan travel demand models use TAZs that dovetail with census geography.