Accident Analysis and Prevention 66 (2014) 80–88

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Comparative analysis of the spatial analysis methods for hotspot identification

Hao Yu[a,b], Pan Liu[a,b,*], Jun Chen[a,b], Hao Wang[a,b]

[a] Jiangsu Key Laboratory of Urban ITS Southeast University, Si Pai Lou #2, Nanjing 210096, China
[b] Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Si Pai Lou #2, Nanjing 210096, China

ARTICLE INFO

Article history:
Received 1 August 2013
Received in revised form 14 January 2014
Accepted 20 January 2014

Keywords:
Hotspot
Spatial analysis
Kernel density
Empirical Bayesian

ABSTRACT

Spatial analysis technique has been introduced as an innovative approach for hazardous road segments identification (HRSI). In this study, the performance of two spatial analysis methods and four conventional methods for HRSI was compared against three quantitative evaluation criteria. The spatial analysis methods considered in this study include the local spatial autocorrelation method and the kernel density estimation (KDE) method. It was found that the empirical Bayesian (EB) method and the KDE method outperformed other HRSI approaches. By transferring the kernel density function into a form that was analogous to the form of the EB function, we further proved that the KDE method can eventually be considered a simplified version of the EB method in which crashes reported at neighboring spatial units are used as the reference population for estimating the EB-adjusted crashes. Theoretically, the KDE method may outperform the EB method in HRSI when the neighboring spatial units provide more useful information on the expected crash frequency than a safety performance function does.

1. Introduction

The identification of hazardous road segments, or hotspots, is an essential component of the highway safety improvement process. Hazardous road segments can be defined as the locations that have a higher risk of crashes than other similar locations (Elvik, 2007). Numerous methods have been proposed over the years for the identification of hazardous road segments. The methods include, but are not limited to the crash frequency (CF) method (Deacon et al., 1975), the crash rate (CR) method (Norden et al., 1956; Hauer and Persaud, 1984; Stokes and Mutabazi, 1996), the equivalent property damage only method, the empirical Bayes (EB) method (Hauer, 1997; Hauer et al., 2002; Elvik, 1997, 2008; Lord and Persaud, 2004; Lord and Park, 2008; Miaou and Song, 2005), the accident reduction potential (ARP) method, etc.

Most of the hazardous road segments identification (HRSI) methods require road segmentation. A road is divided into a set of segments with a constant length, or homogenous segments with varying lengths. One of the limitations is that road segmentation relies on researchers' subjective judgments to determine the segment length and to identify whether a segment can be considered homogeneous. Previous studies suggested that the length of road segments influences the statistical description of CF (Thomas, 1996). Until recently, however, the effects of road segmentation on hotspot identification are still not so clear.

Hazardous road segments can be considered the locations where crashes are spatially concentrated. The length of each hazardous road segment can be determined by the distribution of the local risk factors that affect the risks of crash occurrences. Spatial analysis technique has recently been introduced as an innovative approach for HRSI. The spatial analysis methods treat the road as a continuous entity with infinite numbers of spatial units. It is assumed that crashes occurring in neighboring spatial units are spatially dependent; and the local risk factors vary gradually and continuously among neighboring spatial units. The hazardous road segments are usually defined as a set of contiguous spatial units characterized by an index that reflects the spatial concentration of crashes (Flahaut et al., 2003; Anderson, 2009).

Various spatial analysis methods have been applied to HRSI (Flahaut et al., 2003; Xie and Yan, 2008; Anderson, 2009; Koohong et al., 2009; Loo et al., 2011). The most commonly used methods include: the local spatial autocorrelation (LSA) method and the kernel density estimation (KDE) method. With the LSA approach, each spatial unit is assigned with an LSA index that evaluates the level of spatial inter-dependence between the observed crashes at neighboring spatial units. A high LSA index indicates the spatial concentration of crashes. The hazardous road segments can

* Corresponding author at: Corresponding author. Tel.: +86 02583791816.
E-mail addresses: seudarwin@gmail.com (H. Yu), pan_liu@hotmail.com (P. Liu), chenjun@seu.edu.cn (J. Chen), haowang@seu.edu.cn (H. Wang).

0001-4575/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2014.01.017

be defined as a set of neighboring spatial units with an LSA index exceeding a predetermined threshold.

The KDE is a nonparametric method that is used to estimate the probability density of a random variable. The KDE method evaluates the spatial pattern by which a global degree of dangerousness is distributed over space. A local density estimate is assigned to each spatial unit as a measure of local crash risk. A spatial unit can be considered hazardous if its local density estimate is greater than a predetermined threshold. The hazardous road segments can then be defined by neighboring spatial units sharing a value of the local density estimate that is higher than a given threshold.

In this study, the performance of two spatial analysis and four conventional HRSI approaches was compared using three quantitative evaluation criteria. The authors also studied the relationship between the KDE and the EB method. The rest of the article is organized as follows: Sections 2 and 3 describe the HRSI methods and the quantitative evaluation criteria. Sections 4 and 5 describe the data and the crash predictive models used for evaluation. The results of the evaluation tests are discussed in Section 6, followed by a discussion of the transferability test in Section 7. The relationship between the KDE method and the EB method is discussed in Section 8. The article ends up with a summary and discussions of major findings in Section 9.

2. HRSI methods for comparison

2.1. Crash frequency method

The CF method is probably the simplest and most commonly used method for HRSI. The target road is divided into various segments. The safety performance of the road segments is ranked by the number of crashes reported at each road segment during a specified time period. One of the limitations is that the CF method does not consider the effects of crash exposure. As a result, the results may bias toward the locations with higher traffic volumes. In addition, using reported crash counts for safety ranking does not take into account the random fluctuation in crash counts. The results may be biased because the hazardous sites are not identified according to the long-term expected CF.

2.2. Crash rate method

Road segments are ranked by the CR to take into account the traffic exposure. Even though the method is currently being extensively used in practical engineering applications, recent studies have suggested that using CR for safety assessment mistakenly assumes that the relationship between CF and flow rate is linear. Similar to the CF method, the random fluctuation in crash counts is not considered in the CR method.

2.3. Empirical Bayes method

The EB method has been used increasingly in recent years (Hauer, 1997; Hauer et al., 2002; Elvik, 1997, 2008; Lord and Persaud, 2004; Miaou and Song, 2005; Lord and Park, 2008). The EB-adjusted crash counts were used as the performance measure for safety ranking. It is assumed that the crashes are Poisson distributed given the expected crash counts; and the expected crash counts at a group of similar locations, that is, the reference population, are gamma distributed. By combining the historical crash counts and the expected number of crashes of similar locations, the method takes into account the long-term fluctuation in crash counts. The EB estimate of the safety of a road segment i is given as (cf, Hauer, 1997):

$$\lambda_i = w_i E[\lambda_i] + (1 - w_i) x_i \tag{1}$$

where w_i can be calculated using the following equation:

$$w_i = \frac{E[\lambda_i]}{E[\lambda_i] + \mathrm{VAR}[\lambda_i]} \tag{2}$$

where λ_i represents the EB-adjusted crash counts; $E[\lambda_i]$ and $\mathrm{VAR}[\lambda_i]$ represent the expected CF and its variance; and x_i is the historical crash count at segment i.

2.4. Accident reduction potential with empirical Bayes method

Several studies suggested that only the extra crashes over the expected CF can be prevented by applying appropriate treatments (Maher and Mountain, 1988; Persaud, 1999). Accordingly, the ARP method defines the road segments with higher excess crash counts as hazardous. If the EB method is used to estimate the expected CF, the ARP for road segment i can be estimated as:

$$\mathrm{ARP}_i = w_i E[\lambda_i] + (1 - w_i) x_i - E[\lambda_i] \tag{3}$$

2.5. Local spatial autocorrelation method

The LSA approach evaluates the extent to which the crash counts in a specific spatial unit vary with the crash counts in its neighboring spatial units. With the LSA approach, the hazardous road segments can be identified by aggregating the continuous spatial units that share similar traits. The LSA approach uses an LSA index as the performance measure for safety ranking. The LSA index measures the extent to which a target spatial unit is similar to its neighboring units. Various spatial autocorrelation indexes have been proposed. One of the most commonly used spatial autocorrelation indexes is the Moran's I, which was proposed by Moran in 1948. The Moran's I index for spatial unit i can be calculated as (cf, Moons et al., 2009):

$$I_i = \frac{n}{(n-1)S^2} (x_i - \bar{x}) \sum_j w_{ij} \left(x_j - \bar{x}\right) \tag{4}$$

where x_i represents the crash counts at spatial unit i, which is defined as a 10-m road segment in this study; $\bar{x}$ represents the average CF of all the spatial units; w_{ij}, which is defined as the reciprocal value of the distance between spatial unit i and j, measures the proximity between two spatial units; n represents the total number of spatial units; and S^2 represents the variance of the observed crash counts within all spatial units.

2.6. Kernel density estimation method

The KDE method assesses the risk of crashes at a spatial unit given the crash counts at neighboring spatial units. A symmetric surface is placed on the center point of a spatial unit and the distances between the center point and the locations of crashes within the surface are evaluated. The size of the surface is determined by the bandwidth of the kernel. For each spatial unit, the kernel density is estimated and serves as the measure of safety performance for HRSI. A general density estimation function is given as follows (cf, Fotheringham et al., 2000):

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{d_i}{h}\right) \tag{5}$$

where $f_n(x)$ is the density estimate at spatial unit x; h is the predefined bandwidth; n is the number of crashes near location x within a radius of h; K is a predefined kernel density function to measure the