Available online at www.sciencedirect.com

ScienceDirect

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

Original Research Paper

Development and application of traffic accident density estimation models using kernel density estimation

Seiji Hashimoto [a], *Syuji Yoshiki* [b,*], *Ryoko Saeki* [c], *Yasuhiro Mimura* [d], *Ryosuke Ando* [e], *Shutaro Nanba* [a]

[a] *Graduate School of Environmental and Life Science, Okayama University, Okayama 700-0807, Japan*
[b] *Department of Civil Engineering, Fukuoka University, Fukuoka 814-0180, Japan*
[c] *Fukuyama Consultants Co. Ltd., Fukuoka 812-0013, Japan*
[d] *International Development Consultants Co. Ltd., Aichi 460-0008, Japan*
[e] *Toyota Transportation Research Institute, Aichi 471-0024, Japan*

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form
30 December 2015
Accepted 6 January 2016
Available online 10 May 2016

Keywords:
Traffic safety
Kernel density estimation (KDE)
Hotspots
Zone 30

ABSTRACT

Traffic accident frequency has been decreasing in Japan in recent years. Nevertheless, many accidents still occur on residential roads. Area-wide traffic calming measures including Zone 30, which discourages traffic by setting a speed limit of 30 km/h in residential areas, have been implemented. However, no objective implementation method has been established. Development of a model for traffic accident density estimation explained by GIS data can enable the determination of dangerous areas objectively and easily, indicating where area-wide traffic calming can be implemented preferentially. This study examined the relations between traffic accidents and city characteristics, such as population, road factors, and spatial factors. A model was developed to estimate traffic accident density. Kernel density estimation (KDE) techniques were used to assess the relations efficiently. Besides, 16 models were developed by combining accident locations, accident types, and data types. By using them, the applicability of traffic accident density estimation models was examined. Results obtained using Spearman rank correlation show high coefficients between the predicted number and the actual number. The model can indicate the relative accident risk in cities. Results of this study can be used for objective determination of areas where area-wide traffic calming can be implemented preferentially, even if sufficient traffic accident data are not available.

* *Corresponding author*. Tel.: +81 92 8716631x6494; fax: +81 92 8656031.
E-mail address: syoshiki@fukuoka-u.ac.jp (S. Yoshiki).

Peer review under responsibility of Periodical Offices of Chang'an University.
http://dx.doi.org/10.1016/j.jtte.2016.01.005

1. Introduction

Traffic accident frequency has been decreasing in Japan in recent years. Nevertheless, many accidents still occur on residential roads. Therefore, it is necessary to implement measures for traffic calming on roads. Area-wide traffic calming measures, including Zone 30 (Institute for Road Safety Research, 2004), are especially effective and have often been implemented throughout urban areas. However, for implementation of these measures in Japan, it is impossible to adopt a universal application methodology that merely implements these measures in an entire urban area because borders between urban and rural areas are ill-defined in Japan. In contrast to foreign countries, urban areas in Japan spread out throughout a city, because Japan is an island nation. It has not been exposed to the threat of different regional ethnic groups and it is also not necessary to surround urban areas with walls. This urban feature makes it easy to construct sprawling cities with ever-increasing populations. Consequently, urban areas have been constructed in Japan with ill-defined borders separating urban and rural areas. Moreover, it is difficult to implement measures quickly because of budget constraints. Therefore, it is necessary to ascertain areas where implementation is the most preferred.

Locations of traffic accidents are crucially important information for the implementation of traffic safety measures. Location information of arterial roads is readily available because traffic is typically heavy and traffic accidents often occur on such roads. However, insufficient information is available about such locations on residential roads because traffic accidents occur rarely and incidentally on such roads. Moreover, traffic accident data are difficult to obtain from police departments of many Japanese cities. It is difficult to ascertain the distribution of traffic accidents in all cities including residential areas. Consequently, when area-wide traffic calming measures are implemented in a city, determining which areas these measures should be implemented preferentially must depend on an experience-based subjective view.

The authors develop an estimative model of traffic accident density from GIS data, which are commonly available data by the public and private sector, including population, road factors, and spatial factors. The model enables objective and easy determination of areas to implement area-wide traffic calming preferentially, even if traffic accident data are not available.

This study examined relations between traffic accidents and city components of population, road factors, and spatial factors. Then a model was developed to estimate traffic accident density.

A sufficient amount of traffic accident data must be accumulated to develop the model because traffic accidents in residential areas occur rarely and incidentally. The possibility exists that analysis based on only a few years of data impairs the predictive accuracy. Therefore, this study specifically uses kernel density estimation (KDE) described by Silverman (1986), which can deal with comprehensive estimation of the distribution based on a finite data sample.

KDE has been used for traffic accident analysis and widely as a visualization tool. For example, parameters of traffic accident prediction models have been estimated mainly based not on KDE but on raw count data in Japan. Yu et al. (2014) recently reported that KDE outperformed other hazardous road segment identification methods. Therefore, the accuracy of traffic accident prediction model might be improved by using KDE.

This study aims to develop traffic accident density models based on KDE as an explained variable. Additionally, the contribution of these models is evaluated from practical and academic perspectives. Regarding practicality, the applicability of these models to other cities is examined. Academically speaking, this study examines the improvement in applicability of using KDE as an explained variable instead of using raw count data as an explained variable.

For this study, a model using KDE as an explained variable is a KDE model. One using raw count data as an explained variable is a raw count data model.

2. Literature review

This section presents a review of the literature about KDE application to traffic accident analysis. As explained above, traffic accidents occur rarely and incidentally in residential areas. When traffic accident hotspots are analyzed based on raw data, the possibility exists that potential hotspots are not detected. Therefore, KDE has been used to detect traffic accident hotspots.

The first report using KDE for traffic accident data was made by Banos and Huguenin-Richard (2000), who mapped the distribution of child pedestrian accidents using KDE. Similarly, several studies have identified spatial clusters of accidents through KDE (Anderson, 2009; Pulugurtha et al., 2007; Schneider et al., 2004). Furthermore, several studies have used KDE to analyze traffic accidents spatially and temporally (Blazquez and Celis, 2013; Plug et al., 2011). Krisp and Durot (2007) mapped a distribution of wildlife–vehicle accidents using KDE. In fact, KDE has been evaluated for detection of traffic accident hotspots. It has also been compared with other methods (Erdogan et al., 2008; Yu et al., 2014). Network kernel density estimation (Network KDE), a method for adapting KDE as a function of networks, has been increasing recently. Loo et al. (2011) and Xie and Yan (2008, 2013) used Network KDE to analyze traffic accidents. The literature reveals that numerous studies have investigated traffic accident analysis by using KDE. However, Xie and Yan (2013) have reported that previous studies have remained at the level of using KDE mainly as a tool for visualization.

Many previous studies have examined relations between traffic accidents and city components (Kim et al., 2006; Noland and Quddus, 2004; Pulugurtha et al., 2013; Quddus, 2008; Wier et al., 2009), and have developed models to estimate traffic accident risk (Hadayeghi et al., 2010; Marshall and Garrick, 2011; Moeinaddini et al., 2014; Rifaat et al., 2011), but the models were developed based on the prior few years of accident data, even though traffic accidents rarely occurred on residential roads. This fact can reduce the