

基于空间自相关的城市道路事故多发点鉴别

蒋宏,方守恩,陈雨人,马柱

(同济大学 道路与交通工程教育部重点实验室,上海 201804)

摘要: 提出了基于空间自相关鉴别城市道路事故多发点的方法,并利用5年内交通事故数据进行验证。首先,以非参数核密度估计模型描述交通事故点的空间二维分布,得到事故分布密度的最优窗宽为115.2 m,借此划分路段的空间单元。然后,基于地理信息系统“面-弧”城市道路网络拓扑模型构建空间统计单元,并集计事故数和事故严重程度指数属性。最后,通过全局Moran指数检验,事故数据呈现聚集的空间分布模式,局部G统计量鉴别出属性值高低的聚类,并生成事故多发点分布图。研究表明,与事故频率建模的负二项模型相比,事故位置与属性数据一体化下的空间数据分析不但能够鉴别路段(交叉口)的事故多发点,而且能够对其属性值进行空间关联排序,交通管理部门可利用空间分析可视化结果定位事故多发点和进一步开展交通安全决策研究。

关键词: 交通安全;空间自相关;事故多发点鉴别;核密度估计;负二项模型

中图分类号: U491.3

文献标志码: A

attribute values of crash counts and severity indices were aggregated. Finally, the Globe Moran index was employed to examine spatial distribution clustered pattern of crash data. The local G-statistic was used to identify the clustering of low and high index values and to generate a crash hot spots map. Results of the research indicate that compared with the negative binomial model in modeling the crash frequency, the spatial analysis under integration of accident attribute data with position can not only identify hot spots on road segments (intersections) but also rank hot spots with spatial correlation according to the attribute values. Traffic management department can use the spatial analysis visualized results to locate hot spots and develop further traffic safety decision-making research.

Key words: traffic safety; spatial autocorrelation; hot spots identification; kernel density estimation; negative binomial model

A Spatial Autocorrelation-based Approach for Identifying Hot Spots on Urban Road

JIANG Hong, FANG Shouen, CHEN Yuren, MA Zhu

(Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China)

Abstract: An approach based on spatial autocorrelation analysis was presented to identify hot spots on urban road. The application of this methodology was illustrated by five-year's traffic crash data. Firstly, a non-parametric kernel density estimate sketched a model of two dimensional planar accident point spatial distribution. An estimation of the optimal bandwidth for accident distribution density was 115.2 m and subsequently was selected to create a spatial statistical unit on road segment. Then, the spatial statistical units were derived from a GIS-based region-arc topology model as a representation of the urban road network, and the

在事故多发点鉴别及成因分析研究方法中多数是基于统计分析经验模型开展的,主要分析对象是通过研究总体中随机选择样本观测单位而得到的数据,往往只关注它们的非空间特征和属性,未能充分挖掘事故发生所关联的地理空间属性;此外,经典的统计分析将事故的发生看成是随机过程和相互独立的,通常不考虑观测单位的地理空间信息。常见方法有事故数法、事故率法、当量总事故数法、矩阵法、临界率和质量控制法等^[1]。由于空间数据(如事故多发点的观测数据)是从特征相近的区域收集的,同一区域会具有类似的城市土地利用、道路条件及交通环境等属性。这些因素的不合理组合会孕育或导致事故多发点的产生。这也正符合地理学第一定律:所有的事物都是相互联系的,但离得越近的事物,彼此之间的联系就越强^[2]。因此,需要将地理信息系统

收稿日期: 2012-04-25

基金项目: “十一五”国家科技支撑计划(2007BAK35B03);国家自然科学基金(51078270);乌鲁木齐市科学技术计划(H101326001)

第一作者: 蒋宏(1981—),男,工程师,博士生,主要研究方向为道路交通安全、交通规划设计。E-mail: yamatuo@hotmail.com

通讯作者: 方守恩(1961—),男,教授,博士生导师,工学博士,主要研究方向为道路交通安全、道路规划与计算机辅助设计等。E-mail: fangsek@tongji.edu.cn

(geographic information system, GIS)与空间数据挖掘方法相结合来研究城市空间上的交通安全问题。

近几年,国内学者也积极开展了基于GIS的道路交通安全研究,如在事故数据的空间查询及分析结果的可视化方面,其核心“事故多发点鉴别模块”嵌入的多为当量事故次数法、模糊评价法和改进的质量控制法等,对事故空间数据信息的挖掘深度不够^[3-4]。而基于空间数据模型的交通安全研究已经逐渐成为国外研究的一个热点方向。如Kam^[5]通过GIS关联交通出行数据与事故数据基于非集计模型对事故率与出行者年龄的关系进行了研究。Yamada等^[6]根据事故数据依附于路网及距离尺度的特性,通过实例验证得出网络K函数比平面K函数对于事故点聚类模式的识别更具适用性。Pulugurtha等^[7]应用GIS系统对拉斯维加斯市行人碰撞事故的事故多发点进行了鉴别,并对事故多发点的风险进行了排序。Quddus^[8]通过非空间模型(负指数模型)和空间模型(小范围尺度空间贝叶斯模型)比较了伦敦大都市圈的事故严重度差异。Anderson^[9]运用核密度估计法判定伦敦道路伤亡事故的空间分布模式,并用K均值聚类分析方法鉴别事故多发点。

本文从事事故分布的地理学视角出发,探讨交通事故空间位置及属性的相互关系,并将核密度估计及空间自相关等空间聚类方法应用于城市道路事故多发点鉴别,为城市道路交通安全评价及管理提供了一种思路和手段。

1 基于点模式的空间聚类分析

1.1 核密度估计法的概念与方法

核密度估计法(kernel density estimation, KDE)的核心思想是:地理现象和事件可以发生在平面空间上的任何位置,但发生的概率依位置而不同。点密集的区域事件发生的概率高,点稀疏的区域则相反^[10]。核密度方程的几何意义为:密度分布在每个 s 点中心处最高,向外不断降低,当距离中心达到一定阈值范围(窗口的边缘)处密度为零,参见图1。一个常用二维平面空间的核密度函数 $\hat{f}(s)$ 定义如下:

$$\hat{f}(s) = \frac{1}{h^2} \sum_{e=1}^n K\tau(s)h^{-1} \quad (1)$$

式中: h 为窗宽; n 为研究区域 R 内的事件总数; $K(\cdot)$ 为核函数; $\tau(s)$ 为估值点 s 到第 e 个事件 s_e 处的距离。

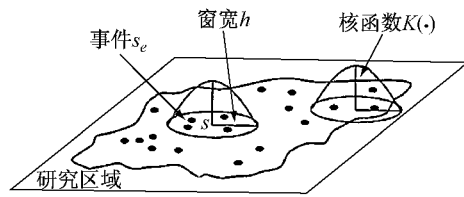


图1 核密度方法原理示意^[11]

Fig.1 Diagram of the basis for kernel density method^[11]

1.2 核函数及窗宽的选择

影响KDE的2个重要因素是 $K(\cdot)$ 的数学形式和 h 的大小。已有统计研究表明:核函数的数学形式对于密度值估计的影响较小^[12],常用的核函数有直角、三角和高斯核函数等。 h 的大小影响到密度函数估计的光滑程度以及样本点对于 s 处的密度估计,选择合适的 h 至关重要;例如,当总体服从 $N(0, \sigma^2)$ 分布,核函数 $K(\cdot)$ 为Gaussian核函数时,最优窗宽为 $\hat{h} = 1.06\sigma n^{1/5}$,其中 σ^2 为正态分布方差。

2 基于面模式的空间聚类分析

2.1 空间统计单元的定义

由于城市道路没有里程碑实体可供参考,目前交警对于交通事故现场地点的描述通常引用附近的特征点作为参考,并辅以交通事故位置相对参考点的方向和偏移量。该方法虽然灵活,但对于事故位置的描述仍不准确,且语言模糊、标准不统一。因此,本文采用地图标定的方法,用经纬度坐标精确定位事故位置,从根本上克服了上述方法的缺点。由此,交通事故观测数据由两部分构成:其一为事故空间坐标信息,每个坐标都对应了一个事故的位置;其二是与每一起事故相关联的属性数据,如事故发生时间、事故形态和事故原因等信息。

国内外研究对于事故点空间位置的搜索大都是基于平面或者线性网络的“点-弧”拓扑模型,该方法的缺点是易将点(交叉口)与弧(路段)同质化。鉴于交叉口和路段在道路几何构造,特别是在交通特性上的显著不同,美国国家公路与运输协会最新出版的《道路安全手册》中将道路事故划分为交叉口和路段2种类型,如图2所示^[13]。图中,发生于A区域内的事故定义为交叉口事故,发生于B区域内的事故定义为路段事故。因此,引入“面-弧”拓扑模型对城市道路网络进行建模。对于发生在交叉口内的事故点,可以通过“面”与“点”的包含拓扑运算规则建立交叉口与事故的空间关系;而对于路段,则需要在GIS中将弧段设置一定宽度的缓冲区,再通过空间

数据叠合运算建立空间关系。

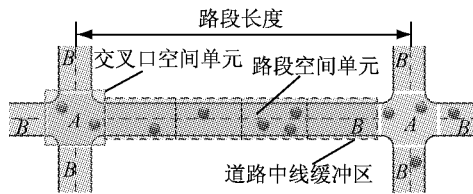


图2 空间统计单元的定义

Fig.2 Definition of spatial statistical units

2.2 空间权重矩阵

空间权重矩阵 W 表达了空间统计单元之间的空间关系,如拓扑和邻接关系等,是构成空间统计检验或模型的基础。 W 的具体形式为^[14]

$$W = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mm} \end{pmatrix} \quad (2)$$

式中, m 为空间单元总数。如果空间单元 i 与空间单元 j 具有公共边界,则表示 2 个单元具有邻接关系,即 $w_{ij} = 1$, 否则 $w_{ij} = 0$ 。道路空间单元之间恰好符合直接 4 邻域的“Rooks”相邻关系。

2.3 基于面模式的全局聚类检验

空间自相关方法结合了空间点事件的位置和相应权重值,体现了属性值在地理位置上的相关性。Moran 指数能够检验整个研究区中邻近地区间相似、相异(空间正相关、负相关)或相互独立。集计了空间属性事故单元的 Moran 指数 I 的计算公式为^[15]

$$I = \frac{m \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{(\sum \sum w_{ij})(\sum (x_i - \bar{x})^2)} \quad (3)$$

式中: w_{ij} 为反映空间单元 i 与 j 之间邻接关系的二元矩阵,对于所有的单元有 $w_{ij} = 0$; x_i 和 x_j 为空间单元 i 和 j 处的事故数或事故严重程度; \bar{x} 为事故数或事故严重程度的全局平均值。

Moran 指数的取值范围为 $[-1, 1]$, 其显著性可通过计算随机模式下期望和方差构造标准化的 Z 统计量来检验,即

$$Z = (I - E(I)) / \sqrt{V_{ar}(I)} \quad (4)$$

式中: $E(I) = -1/(m-1)$; $V_{ar}(I)$ 为理论方差。

2.4 基于面的局部聚类检验

Moran 指数为识别全局空间自相关的综合指标,无法区分高值单元聚集还是低值单元聚集。局部 G 统计量 $G_i^*(d)$ 被用来鉴别局部空间的热点区域,其公式为^[16]

$$G_i^*(d) = \left(\sum_j w_{ij}(d) x_j \right) / \sum_j x_j \quad j \neq i \quad (5)$$

式中: $w_{ij}(d)$ 为基于距离 d 的权重,且当单元 i 和 j 的距离小于 d 时, $w_{ij}(d) = 1$, 否则 $w_{ij}(d) = 0$ 。局部 G 统计量的显著性可通过构造标准化的 Z 得分值来检验,其定义为

$$Z(G_i^*) = (G_i^*(d) - E(G_i^*)) / \sqrt{V_{ar}(G_i^*(d))} \quad (6)$$

如果 $Z(G_i^*)$ 值为正且非常显著,则表明位置 i 周围的值相对较大(高于均值),表现为高值空间聚集;相反,如果 $Z(G_i^*)$ 值为负且非常显著,则表明位置 i 周围的值相对较小(低于均值),表现为低值空间聚集;而如果 $Z(G_i^*)$ 值接近于零,则表明不存在显著的空间关联模式。

3 数据来源及空间统计分析

3.1 数据来源

分析对象为某城市 2006 年至 2010 年 1 585 起事故数据,由公安部科研所开发的交通事故信息管理系统获得。该系统中的数据符合公安部制订的“道路交通事故信息采集表”的规定。事故点包含两方面信息:一是空间坐标(经纬度);二是事故属性(死伤人数和事故类型)。此外,将道路网络按照“面-弧”模型划分为 2 468 个空间统计单元。考虑到事故严重程度对于确定优先改善事故多发点的影响,根据事故严重程度对事故统计赋予权重,引入事故严重程度评价指数,定义如下^[17]:

$$F_s = \sum_{r=1}^R f_r w_r \quad (7)$$

式中: F_s 为地点 S 处的事故严重程度指数; f_{rs} 为地点 S 处事故类型 r 的频数; R 为事故类型的总数; w_r 为死亡、重伤及轻伤的权重系数,此处取值分别为 9.5, 9.5 和 3.5^[18]。

至此,空间统计单元集计的属性包括事故数及事故严重程度指数,其在空间单元上的分布如图 3 所示。

3.2 基于空间单元的事故分布统计模型

对于具有离散、随机和独立性的交通事故,可能发生在任何一个空间单元内,单元内的事故数可以用计数数据模型(count data model)来描述。其中,泊松(Poisson)分布适合于描述单位时间或空间内随机事件发生的次数,其概率函数为

$$P(y_i | u_i) = \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots \quad (8)$$

$$\mu_i = E(y_i | u_i) = \exp(u_i' \beta) \quad (9)$$

式中: y_i 为在统计时间内发生在空间单元 i 上的事故数; u_i 为与空间单元属性相关的解释变量向量; μ_i 为泊松分布参数且满足指数关系; β 为解释变量对应的系数向量。

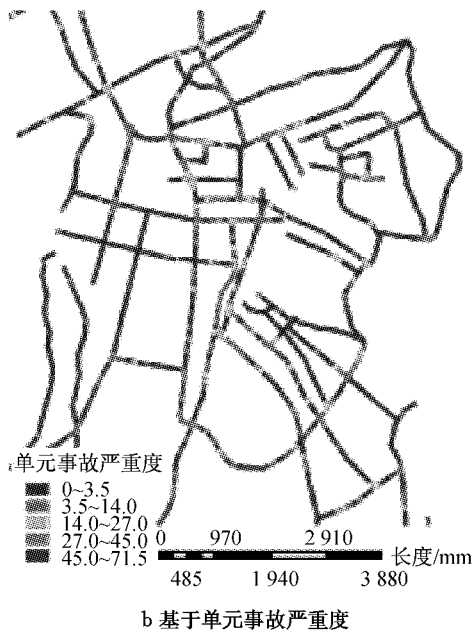
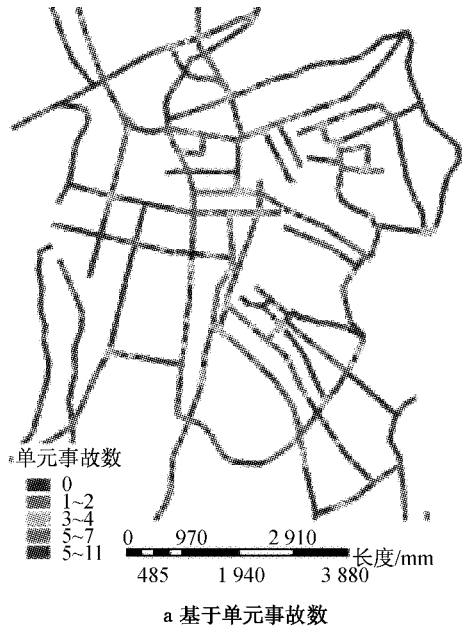


图3 空间统计单元内的事故分布

Fig.3 Spatial distribution of accidents in a spatial statistical unit

此外,为了解决数据样本方差大于均值的过离

表1 计数数据模型的参数分析

Tab.1 Analysis for parameters of calculation data model

模型	α	u_i	ω	$-2\lg L$	AIC值	BIC值
Poisson		0.660	0.660	6 225.999	6 227.999	6 233.810
NB	1.977	0.660		5 430.699	5 434.699	5 446.322
ZIP		1.476	0.552	5 598.307	5 602.307	5 613.930

散(overdispersion)现象及含零过多的离散数据情况,可分别用负二项(negative binomial, NB)和零膨胀泊松(zero-inflated Poisson, ZIP)模型来拟合数据^[19]。负二项分布的概率函数为

$$P(y_i|u_i) = \frac{\Gamma(y_i + (1/\alpha))}{y_i! \Gamma(1/\alpha)} \left(\frac{1/\alpha}{(1/\alpha) + \mu_i} \right)^{1/\alpha} \cdot \left(\frac{\mu_i}{(1/\alpha) + \mu_i} \right)^{y_i} \quad (10)$$

式中: $\Gamma(\cdot)$ 为伽马函数;非负指标 α 为离散参数。

零膨胀泊松概率函数的公式为

$$P(y_i|u_i) = \begin{cases} \omega + (1 - \omega) \exp(-\mu_i), & y_i = 0 \\ (1 - \omega) \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!}, & y_i > 0 \end{cases} \quad (11)$$

式中, ω 为零膨胀参数,表示由分布产生的零计数概率。

由图4可见,2 468个空间单元中没有事故发生的即事故次数为零的占65.48%,事故次数在6次及以下的空间单元已经达到99.39%,数据呈正偏态分布。因此考虑利用不含协变量的负二项及零膨胀分布模型对空间单元事故数的观测值进行拟合,模型整体显著性 P 值小于0.000 1,拟合结果见表1。从拟合效果来看,负二项模型明显优于普通的泊松模型及零膨胀模型,根据对数似然数值(log likelihood, $\lg L$)、赤池信息准则(Akaike information criterion, AIC)、贝叶斯信息准则(Bayesian information criterion, BIC)可知,3个模型中负二项模型的拟合效果最优,且得到空间单元事故数期望为0.660,方差为1.521。

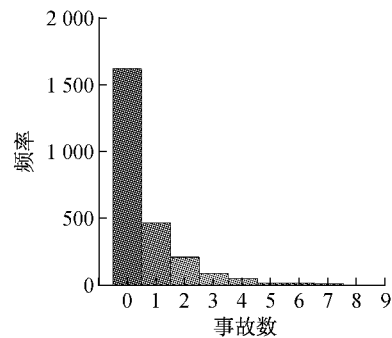


图4 事故在空间单元上的频率分布

Fig.4 Frequency distribution of accidents in a spatial unit

根据负二项分布的参数得到空间单元事故数的累计概率分布曲线,如图 5 所示. 可将累计频率大于 95% 的空间单元即实际观测事故数大于临界值 3 的空间单元确定为事故多发点.

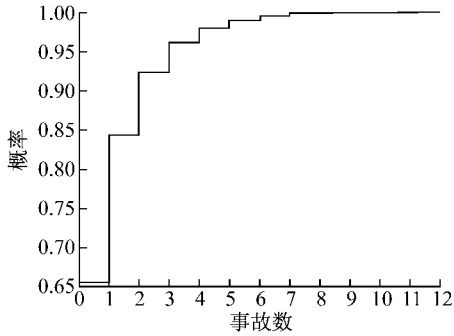


图 5 负二项分布累计概率分布函数
Fig.5 Negative binomial cumulative probability distribution function

4 空间聚类分析结果

4.1 基于点模式的核密度方法

基于点模式的事故多发点将其定义为平滑估计的截尾处,即相邻空间单元的密度值大于给定的阈值(如分位数或均值). 最终的鉴别结果取决于核密度估计的平滑程度及截尾过程. 本研究区域内共有事故样本点 $n=1\ 585$ 个,选用 Gaussian 核函数估计平面空间事故点的密度,计算得到最优窗宽为 $\hat{h}_{opt}=115.2\text{ m}$,因此,事故点 i 处窗宽范围内的事故数估计为 $c=Nh^2 \hat{f}(s_e | n)$,在最优窗宽影响下的事故数平滑估计如图 6 所示.

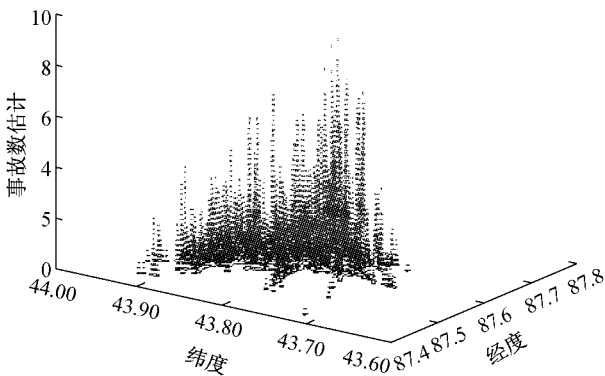


图 6 最优窗宽下的事故数估计

Fig.6 Crash counts estimator for optimal bandwidth

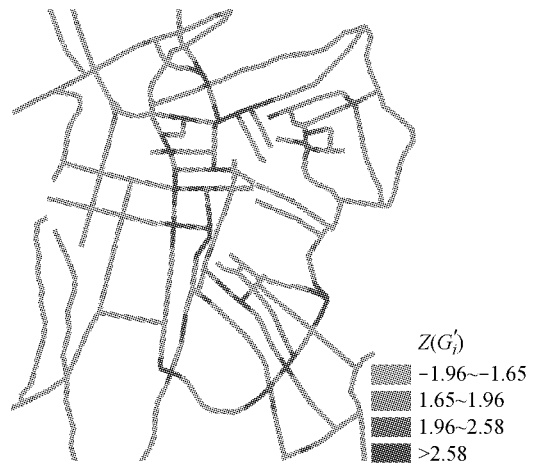
4.2 空间自相关结果分析

由于点模式的核密度方法没有考虑事故发生于道路空间的先决条件,因此再利用核密度方法得到

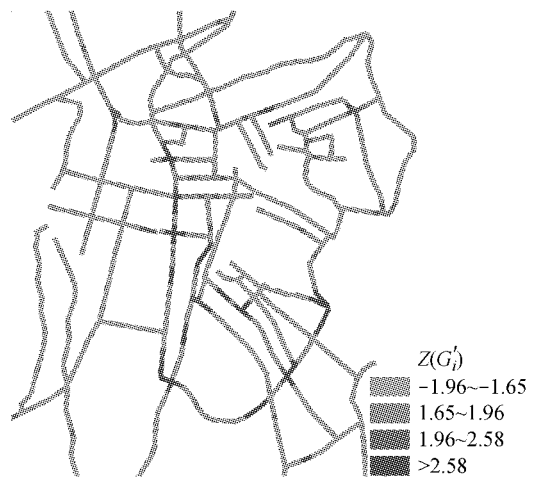
的最优窗宽将道路划分为基于面的空间统计单元,应用空间自相关方法进一步分析. 根据 Rooks 标准定义相邻的道路空间单元,并用标准化的二元空间权重矩阵计算 Moran 指数. 在随机假设 $\alpha=0.01$ 的显著性水平下,计算结果如表 2 所示,表中 Z 为基于理论 Moran 指数构造的正态分布检验统计量. Z 值大于 2.58,表明空间单元上的事故数及事故严重程度指数分布模式均与随机模式之间存在显著性差异,且呈现正的空间自相关. 空间单元的局部 $G_i^*(d)$ 统计量的标准化 Z 值分布结果如图 7 所示.

表 2 空间单元属性的全局聚类指标
Tab.2 Global clustering index for spatial unit attributes

项目	期望值/ 10^{-4}	方差/ 10^{-4}	Moran 指数	Z 值
事故数	-4.053 5	4.000 0	0.24	12.01
事故严重程度	-4.053 5	3.916 5	0.15	7.60



a 基于单元事故数



b 基于单元事故严重程度

图 7 基于标准化 Z 得分值的事故多发点鉴别结果

Fig.7 Results of hot spots identified based on the standardized Z-scores

关于事故多发点的定义各个国家和地区采用的标准不尽相同,但其核心理念是找出在一定的统计时间段内事故发生数或特征与其他正常地点相比明显突出的位置(点、路段或区域)^[20]。本文将事故多发点定义为: $Z(G_i^*) > 2.58$ 对应于 $\alpha = 0.01$ 的显著性水平,表明该单元是一个极高值的空间聚类,如图7中的相应颜色区域可判定为一级事故多发点; $1.96 < Z(G_i^*) < 2.58$ 对应于 $\alpha = 0.05$ 的显著性水平,表明该单元是一个较高值的空间聚类,相应颜色区域可判定为二级事故多发点;其他颜色的空间单元表示不存在明显的空间聚类。

与事故频率法的结果相比,基于空间自相关方法鉴别出的空间单元相对较少,其原因在于空间自相关方法同时考虑了单元之间空间邻近关系及其属性值之间高低聚集显著性,进而判别出热点空间单元。

5 结语

基于空间数据模型的支撑,从点模式和面模式2种不同的地理学视角对交通事故的空间分布规律进行了探索。点模式分析为一种非集计方法,采用二维核密度函数对最优视窗范围内的期望事故数进行了估计。由于该方法没有考虑事故类型、严重度等属性信息以及事故位置坐落于路网空间等限制条件,分析结果具有一定的误差和局限性;面模式方法将路网划分为空间统计单元,通过空间单元集计事故数及事故严重度指数,应用空间自相关方法鉴别交通事故的空间聚集性,并以此作为判定事故多发点及其严重程度排序的标准。

空间自相关的事故多发点判别方法充分利用事故地理位置关系,并反映事故属性的空间关联性,尤其适合城市道路交通安全评价研究,但在实际中也存在诸如城市道路空间统计单元划分标准及尺度确定、基于路网线性空间的非集计事故点聚类模式分析等问题需要在将来的研究中重点解决。

参考文献:

- [1] 裴玉龙. 道路交通事故多发点质量控制鉴别法的改进[J]. 哈尔滨工业大学学报, 2006, 38(1): 97.
PEI Yulong. Improvement in the quality control method to distinguish the black spot of the road[J]. Journal of Harbin Institute of Technology, 2006, 38(1): 97.
- [2] Tobler W. A computer movie simulating urban growth in the Detroit region[J]. Economic Geography, 1970, 46: 234.
- [3] 潘敏荣, 过秀成, 姜科, 等. 基于GIS的道路交通事故黑点分析处理系统研究[J]. 交通与计算机, 2007, 25(2): 67.
PAN Minrong, GUO Xiucheng, JIANG Ke, et al. GIS-based road traffic accident black-spot analysis system[J]. Computer and Communication, 2007, 25(2): 67.
- [4] 袁泉, 李一兵, 鲁光泉. 基于地理视角的交通事故多发点鉴别分析模型[J]. 交通运输工程学报, 2010, 10(1): 101.
YUAN Quan, LI Yibing, LU Guangquan. Identification analysis model of traffic accident-prone locations based on geographical view angle [J]. Journal of Traffic and Transportation Engineering, 2010, 10(1): 101.
- [5] Kam B H. A disaggregate approach to crash rate analysis[J]. Accident Analysis and Prevention, 2003, 35: 693.
- [6] Yamada I, Thill J. Comparison of planar and network K-functions in traffic accident analysis[J]. Journal of Transport Geography, 2004, 12: 149.
- [7] Pulgurtha S S, Krishnakumar V K, Nambisan S S. New methods to identify and rank high pedestrian crash zones; an illustration[J]. Accident Analysis and Prevention, 2007, 39(4): 800.
- [8] Quddus M A. Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data [J]. Accident Analysis and Prevention, 2008, 40: 1486.
- [9] Anderson T K. Kernel density estimation and K-means clustering to profile road accident hotspots [J]. Accident Analysis and Prevention, 2009, 41: 359.
- [10] Silverman B. Density estimation for statistics and data analysis [M]. London: Chapman and Hall, 1986.
- [11] Bailey T C, Gatrell A C. Interactive spatial data analysis[M]. New York: John Wiley and Sons, 1995.
- [12] O'Sullivan D, Unwin D J. Geographic information analysis [M]. Hoboken: John Wiley, 2002.
- [13] AASHTO. Highway safety manual [M]. Washington D C: American Association of State Highway and Transportation Officials, 2010.
- [14] WANG Fahui. Quantitative methods and applications in GIS [M]. Boca Raton: CRC/Taylor & Francis, 2006.
- [15] Cliff A D, J K Ord. Spatial processes [M]. London: Pion Press, 1981.
- [16] Getis A, Ord K. The analysis of spatial association by use of distance statistics[J]. Geographical Analysis, 1992, 24(3): 189.
- [17] Gholston S E, Anderson M D. A GIS accident system to accompany CARE [R]. Huntsville: University of Alabama, 2005.
- [18] PIARC. Road safety manual [R]. [S. l.]: PIARC Technical Committee on Road Safety, 2003.
- [19] Washington S P, Karlaftis M G, Mannering F L. Statistical and econometric methods for transportation data analysis[M]. Boca Raton: CRC Press LLC, 2003.
- [20] 方守恩, 郭忠印, 杨轸. 公路交通事故多发位置鉴别新方法[J]. 交通运输工程学报, 2001, 1(1): 90.
FANG Shouen, GUO Zhongyin, YANG Zhen. A new identification for accident prone location on highway [J]. Journal of Traffic and Transportation Engineering, 2001, 1(1): 90.