

文章编号: 1000-5641(2018)03-0136-10

多尺度探查偷车案件的犯罪时空热点

任浙豪^{1,2}, 张昊天^{1,2}, 刘苇航^{1,2}, 过仲阳^{1,2}

(1. 华东师范大学 地理信息科学教育部重点实验室, 上海 200241;

2. 华东师范大学 地理科学学院, 上海 200241)

摘要: 犯罪热点探查逐渐成为被动式转向主动式警务工作的关键. 现有许多研究提出多种犯罪分析方法, 得到较好成果. 本文关注时空分析中的尺度问题, 以某区偷盗“三车”案件为例, 提出分析时空热点的两种多尺度方法, 并据此分析, 为警务实务提供参考. 这两种方法与结论是: ①结合尺度法与传统的严格法整合时间数据, 在单尺度数据不足时也能探查到时间热点. 分析得到研究区偷车案件存在较显著的 30 d 周期和极显著的 20 h 周期, 警务工作可借此调整轮班时间. ②以 30 d 周期作分析, 在 DBSCAN 算法中设置评估函数(凸包案件密度中值), 自适应选择最优尺度探查空间最优热点分布, 分布的最优性由主流的 PAI 指数验证. 分析得到研究区的某些地铁站与居民区需要重点防护.

关键词: 多尺度; 犯罪分析; 小波分析; DBSCAN; 凸包

中图分类号: TP391.4 **文献标志码:** A **DOI:** 10.3969/j.issn.1000-5641.2018.03.015

Detecting spatiotemporal hotspots for vehicle thefts by multi-scale analysis

REN Zhe-hao^{1,2}, ZHANG Hao-tian^{1,2}, LIU Wei-hang^{1,2}, GUO Zhong-yang^{1,2}

(1. *Key Laboratory of Geographic Information Science (Ministry of Education),
East China Normal University, Shanghai 200241, China;*

2. *School of Geographic Sciences, East China Normal University, Shanghai 200241, China)*

Abstract: The detection of crime hot spots has become increasingly prominent in the conversion from reactive to active policing. There exist many crime analysis methods with good results. This paper focuses on scale effects in analysis. We proposed two multi-scale methods to detect temporal and spatial hotspots for vehicle thefts in a district, whose results were used for policing references. These two methods and their results are stated as follows: ① a scaling method is proposed and combined with a rigid process to aggregate temporal data. Through this combination, temporal hotspots can be detected when data are not sufficient under mono-scale. Results showed that daily hot spots (30 days) and hourly hot spots (20 hours) of vehicle thefts are significant at the study site, on which the

收稿日期: 2017-05-17

基金项目: 国家理科基地科研训练及科研能力提高项目(J1310028)

第一作者: 任浙豪, 男, 本科生, 研究领域为数据挖掘、遥感图像分析. E-mail: 13681956189@126.com.

通信作者: 过仲阳, 男, 教授, 博士生导师, 研究方向为数据挖掘、数据可视化.

E-mail: zyguo@geo.ecnu.edu.cn.

rearrangement of shift intervals can be based; ② on the basis of daily hot spots, we set a median case density of a convex hull as the evaluation function when applying DBSCAN. The optimal scale, verified by the popular Prediction Accuracy Index, was adaptively chosen. We found that several metro line stations and residence zones need key protection.

Keywords: multi-scale; crime analysis; wavelet analysis; DBSCAN; convex hull

0 引言

偷盗“三车”包括偷盗摩托车、电动车和自行车,这样的犯罪案件普遍与地理环境密切相关且具有时空集聚特性^[1-2].通过研究犯罪案件时空特征与预测方法,例如热点制图^[3-4],时间序列预测^[5-6]等能帮助警务人员更合理地安排警力部署^[7]、平衡工作重心^[8].

现有犯罪案件的分析与预测方法多样.有些研究关注时间序列分析预测,例如使用探索性数据分析寻找现有案件的时间序列特征,并据此预测犯罪率^[5];有些研究关注案件的空间分布,例如使用核密度^[8]、聚类^[9]等方法绘制犯罪热点预测图;还有些研究结合了时空两方面同时考虑,例如在时间热点基础上绘制空间热点预测图^[8]、时空分析制预测图^[3]等.但是少有研究同时关注时空尺度问题,如果找到时空最优尺度,将使绘制的热点图更准确可信,具备预测功能.本文主要关注这种时空尺度问题.

本文在前人研究的基础上,分别在时间和空间尺度对已有方法做了扩展:①探查时间热点时,提出尺度法,并结合传统的严格法来整合数据;②绘制空间热点预测图时,改进DBSCAN (Density-Based Spatial Clustering of Applications with Noise)算法,使其自适应选择最优聚类尺度.下面分别对主要技术的现有进展和本文贡献作出说明.

在确定案件发生具体时间时,有平均划分及严格划分两种常用方式^[10-11],前者充分利用数据但探测热点有偏差,而后者以舍弃部分数据为代价探测较准热点.针对这个问题,本文提出一种尺度法整合数据,先以严格法探查大尺度时间周期热点,经降维压缩到一个周期,继而在更小尺度探查热点,可在缺失数据的情况下完成周期的提取.

在绘制财产类案件的空间热点预测图时,有学者在聚类时以凸包绘制热点,使热点划分更“软”,同时具备更高的预测准确率^[4].但它难以确定最优聚类尺度,热点预测图的绘制与准确率检验仍然分离,有较多不便.本文针对财产类案件提出改进的DBSCAN聚类法(Modified DBSCAN, MDBSCAN),即在DBSCAN聚类中嵌入凸包分析.同时,以凸包案件密度中值(Median of Case Density in Convex Hull, MDCH)为评估函数,来自适应选择最优尺度,使热点预测图的绘制与准确率检验同时进行.并使用预测准确率指数(Prediction Accuracy Index, PAI)^[8-12]验证MDBSCAN方法的有效性.

1 研究区与数据

1.1 研究区

研究区位于某市中心城区的西部(由于数据涉密,研究区隐去经纬度).其西南角为一机场,故在西南角道路分布稀疏.境内道路分布密集,由东至西分别被三个高架、一个高速贯穿,并且4条地铁线路在区内形成井字形.

1.2 研究数据

如图1,本文中涉及到的数据:①研究区行政界线图 and 主要道路图;②2015年1—10月研究区偷盗“三车”案件数据,由该区公安分局提供.案件原始数据共2 805个记录,每个记录

包括案件发生地址、时间等信息.



图1 研究区数据

Fig.1 Study site data

以上原始数据通过三步数据预处理进行后期的分析:①结合百度API进行地理编码^[13],得到本文所使用的地理坐标经纬度;②筛选出研究区范围内属性不为空的事件点作为研究对象;③因地理格网体系形状规则尺度灵活,信息集成和其他单元结合方便^[14],故本文在获得各月份最优热点图后,使用格网进行最后的空间叠置分析.

2 方 法

本节将说明两种多尺度时空分析方法,研究过程如图2所示.具体涉及:以尺度法整合犯罪时间数据,利用小波分析获得犯罪时间热点(见2.1节);改进DBSCAN算法,将其尺度变量多尺度化,以绘制具有预测能力的犯罪空间热点(见2.2节).

2.1 尺度法与连续小波变换

案件记录中,通常包含“三车”拥有者最后见到自己的车以及发现车丢失的时间,这使得其中任意时刻都有可能成为真正的犯罪时间.选择合适的划分方法将使时间热点探测更准确.本文提出的尺度法如下:①以主流的严格法划分时间数据,并用小波提取案件的日尺度热点周期;②由于小波图显示案件在日尺度上发生的频率相似,故可以将其在小时尺度合并,补充小时尺度的数据量;③重复①步,以严格法划分时间数据,以小波提取小时尺度的热点周期.这解决了原本在小时尺度上案件量过少,无法探测小时尺度热点的问题.

上文中提到的连续小波变换是一种处理非线性时间序列的常用方法,它在处理不同频率的时间序列中效果更优^[15],且无论信号自身频率的高低,小波变换都能够提取该信号的特

征信息^[16]. 所以, 本文借此提取偷盗“三车”的犯罪周期.

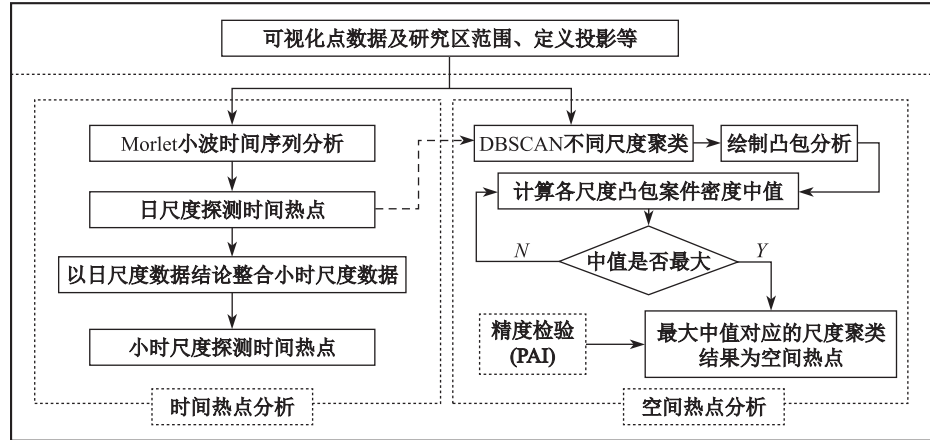


图2 研究流程图

Fig. 2 Technical flowchart

对于连续信号的连续小波变换定义如式(1)和(2)所示^[17-18], 其中, $x(t)$ 表示连续信号, $C_X(a, b)$ 为小波系数, $\psi(t)$ 为小波函数, a 与 b 分别代表了尺度与时间偏移量. 本文使用的小波函数为 Morlet 函数, 其定义如式(3)所示. 式(3)中的 ω_0 为常数, 当 $\omega_0 \geq 5$ 时, Morlet 小波近似于可接受的条件, 且其一阶导数与二阶导数都趋向于 0.

$$C_X(a, b) = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}(t) dt, \quad (1)$$

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right), \quad a, b \in \mathbf{R}, a \neq 0, \quad (2)$$

$$\psi(t) = \pi^{-\frac{1}{4}} e^{i\omega_0 t} e^{-\frac{t^2}{2}}. \quad (3)$$

在此基础上需要计算小波功率谱以获得时间热点, 其被定义为小波系数的模, 而小波系数模的求法类似于傅里叶变换, 如式(4)所示. 其中 $C_X(a, b)$ 即为小波系数, $*$ 代表其共轭复数.

$$W_X(a, b) = C_X(a, b) C_X^*(a, b) = |C_X(a, b)|^2. \quad (4)$$

2.2 多尺度 MDBSCAN 聚类

DBSCAN^[19]核心是使空间中一定距离 (Eps) 范围内的点对象数目达到最小指定数目 ($MinPt$) 的所有对象, 成为一个独立的分类; 以凸包为辅助绘制热点区域. 将凸包分析嵌入 DBSCAN 聚类, 针对财产类犯罪提出一种改进的 DBSCAN 聚类方法 (Modified DBSCAN, MDBSCAN), 在用户指定 $MinPt$ 后自适应确定最优聚类尺度 (Eps).

MDBSCAN 法的主要步骤: ① 根据指定 $MinPt$ 使用不同尺度聚类并对每个尺度下的每类单独绘制凸包; ② 分尺度遍历所有凸包案件密度, 并对各尺度的密度序列取中值 (MDCH), 若只有一个凸包, 则舍弃该尺度, 进入下一次计算; ③ 最大 MDCH 值对应尺度为最优尺度. 计算方法如式(5)所示, 其中, Eps 表示尺度, n 为该尺度下的凸包总数, C_{CH} 为凸包内案件数,

S_{CH} 为凸包面积, Med 为中值函数. 这里使用中值而非均值, 是为了衡量平均水平的同时避免极端值影响.

$$MDCH = Med\left(\frac{C_{CH_i}(Eps)}{S_{CH_i}(Eps)}\right) (i = 1, 2, \dots, n). \quad (5)$$

根据前述, 最优尺度聚类结果从热点探测角度看, 应当具备最好的热点预测准确率. 目前, PAI 被广泛采纳为热点预测精度指标^[8,20], 所以本文以 PAI 验证自适应选择最优尺度的效果. PAI 由两部分组成, 具体公式如式(6)所示, 其中 HHR 为测准案件数与总案件数之比, AP 为预测面积与研究区面积之比. 在 MDBSCAN 算法中, i 为 Eps 尺度序列, m 为尺度个数; n_i 和 a_i 分别为该尺度下以当前周期预测下一周期的测准案件数和热点区域面积; N 为下一周期的总案件数, A 为研究区面积.

$$PAI = \frac{HHR}{AP} = \frac{n_i/N}{a_i/A} (i = 1, 2, \dots, m). \quad (6)$$

3 结果与讨论

3.1 时间热点分析

偷盗“三车”类犯罪案件的时间数据从 1 月记录至 10 月, 将其整理为两个尺度: 日尺度和小时尺度, 并且提取两种不同尺度下的周期.

首先我们将原始数据整理为日尺度数据, 小波功率越大, 相应时间间隔出现的频率越高. 通过 0.05 显著度(粗黑线内部)的偷盗“三车”案件小波功率谱高值如图 3(a)所示. 8 d 以下显示的周期可能是由于随机因素导致的, 相较之下 30 d 周期与 65 d 周期有更高的波谱能量. 此处周期较明显可直接用作空间热点探查的周期. 下面以求更小的时间尺度为例, 说明如何以尺度法整理时间数据, 获得更小尺度的周期信息.

本文分析小时尺度热点时, 由于每月(30 d 为一个周期)案件发生频率具备周期性, 故而可以将每月的案件数按照相同的日期时刻进行次数统计, 得到每小时的案件数量, 从而尽可能地弥补单月造成的空值, 也使周期更具普遍意义. 如图 3(b)的小波功率谱所示, 16~24 谱段能量最高, 说明小时尺度的周期为 20 h 左右.

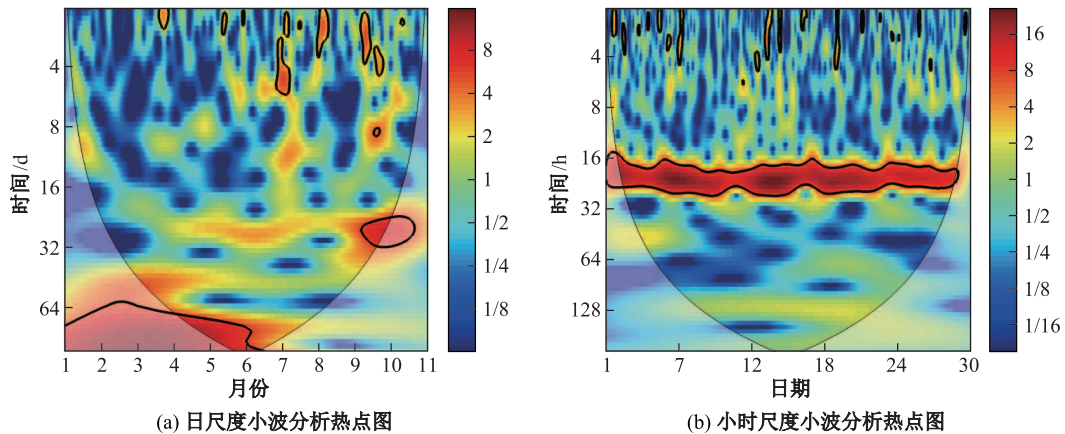


图3 小波分析所得两个尺度下的时间热点

Fig. 3 Temporal hotspots under two scales as detected by wavelet analysis

偷盗“三车”案件在日尺度的 30 d 热点与小时尺度的 20 h 热点可以很容易地理解为偷

盗“三车”事件发生的周期. 20 h 尺度, 值得警务实务工作重点关注. 后文探查空间热点时, 考虑到以 20 h 周期计算, 其耗时较多, 所以以 30 d 为周期探查空间热点.

3.2 空间热点分析

3.2.1 MDBSCAN 聚类寻找最优尺度下的空间热点

根据 2.2 节提出的 MDBSCAN, 本例中 $MinPt=10$ 是有良好聚类变化的区间, 现通过迭代法计算各个尺度下的 MDCH, 这里随机挑选 4 月为例, 以式(5)计算的结果如表 1 所示, 图4 为各尺度下的 DBSCAN 聚类与凸包分析的结果. 结合图表可知, 当 Eps 为 0.21 km 和 0.95 km 时凸包(类别)个数为 1. 前者所绘凸包过小, 则所围区域为极大热点, 虽有价值但不具备指导意义. 而后者所绘凸包过大, 可能将凸包内部的聚类点模式作为均匀分布或随机分布处理, 所以其结果不具备参考价值. 由此得出去除凸包个数为 1 的结论, 所以 Eps 为 0.32 km 时为最优尺度.

表 1 凸包内案件密度中值

Tab. 1 Median case density in convex hull								
Eps/km	0.21	0.32	0.42	0.53	0.63	0.74	0.84	0.95
案件密度中值/(起·km ²)	152.45	81.70	32.14	18.41	17.92	16.56	11.31	12.00

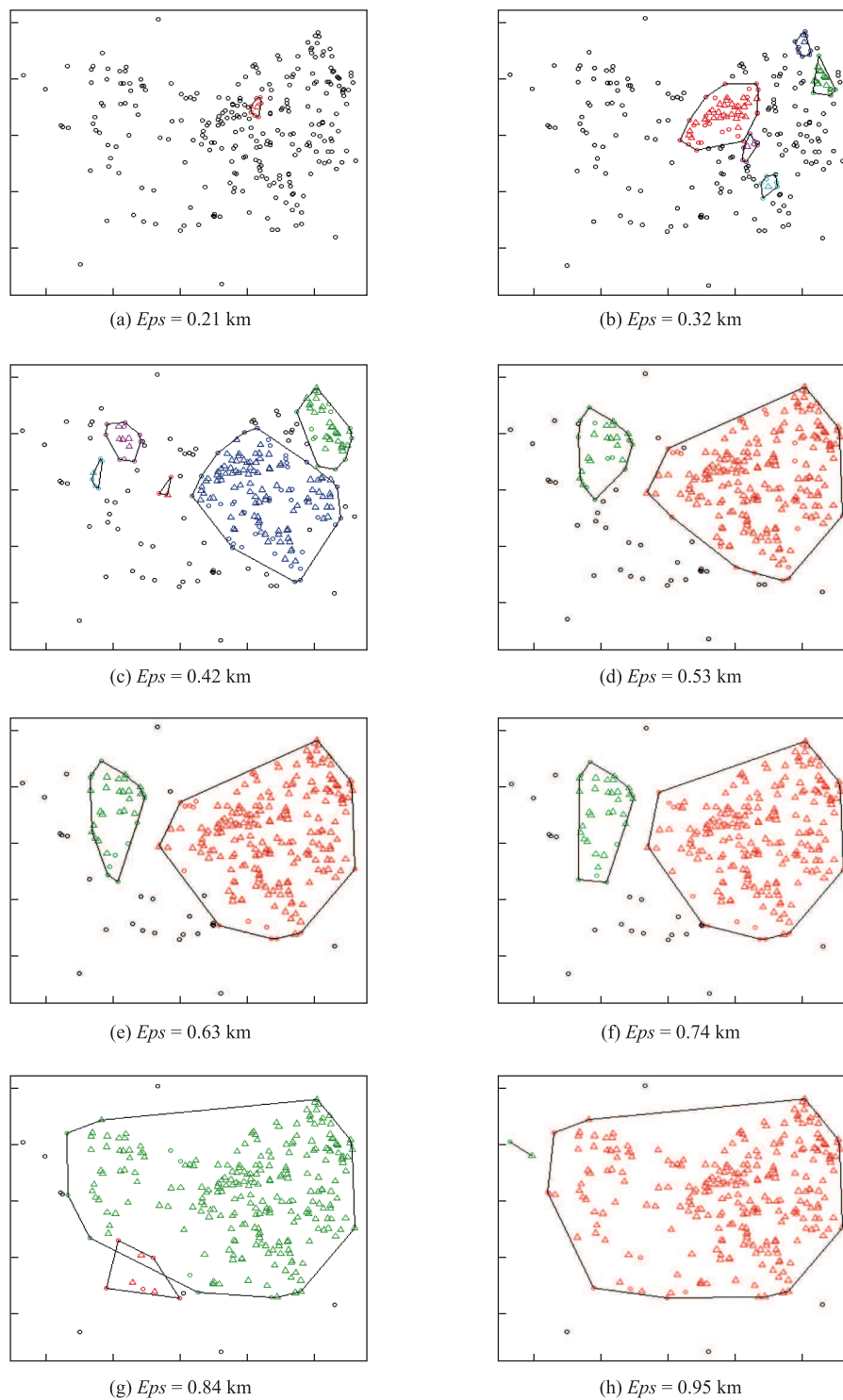
3.2.2 PAI 验证

根据式(6)易知, PAI 指数越高, 预测的准确率越高. 同以 4 月预测 5 月为例. 表 2 显示 4 月预测 5 月的 PAI 值计算列表, 可知, 当去除凸包个数为 1 对应的 Eps 后, $Eps=0.32$ km 时的 PAI 值的确最高(2.69).

图 5(a—c)展示了 4 月预测 5 月、6 月预测 7 月、9 月预测 10 月的三个例子, 说明以 MDCH 和 PAI 挑选最优 Eps 的结果一致性; 再根据表 1 与表 2 的 MDCH、PAI 验证, 可知 MDCH 与 PAI 的值变化趋势相同. 由此可见, 使用 MDBSCAN 选取的最优 Eps 的聚类结果具备优秀的热点制图预测能力. 除了 4 月外, 我们还以 MDBSCAN 算法预测了其余 8 个月的案件, 最终结果如表 3 所示.

表 2 计算 4 月预测 5 月 PAI 值

Tab. 2 PAI based on April to May							
Eps/km	凸包序号	5 月测准案件数/起	5 月实际案件数/起	4 月凸包面积/km ²	研究区总面积/km ²	PAI	
0.21	1	3	272	0.07	36.76	2.69	5.62
	1	36		1.93			
	2	8		0.26			
	3	3		0.13			
	4	4		0.14			
0.32	5	0	272	0.11	36.76	2.69	2.69
	1	8		0.05			
	2	27		1.46			
	3	99		7.64			
	4	8		0.81			
0.42	5	14	272	0.52	36.76	2.16	2.16
	1	172		13.38			
	2	31		1.81			
	1	176		14.27			
	2	35		2.25			
0.53	1	178	272	15.09	36.76	1.64	1.64
	2	37		2.58			
	1	9		14.27			
	2	234		2.25			
	1	250		26.66			



注: 因数据涉密, 图中隐去研究区经纬度

图4 4月案件不同尺度DBSCAN聚类与凸包分析结果

Fig.4 DBSCAN clustering and the convex hull analysis under different spatial scales in April

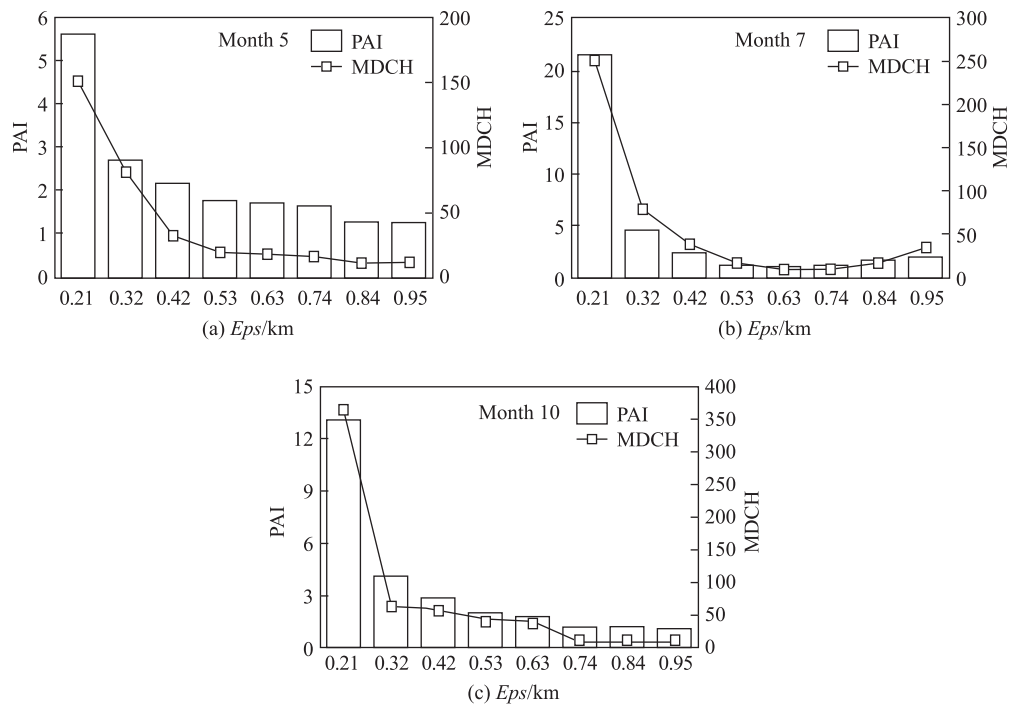


图 5 MDCH 与 PAI 的一致变化趋势

Fig. 5 Consistency in trends of MDCH and PAI

表 3 2015 年 2—10 月最优尺度下的 PAI 结果

Tab. 3 PAI from February to October with the optimal scale in 2015

月份	2	3	4	5	6	7	8	9	10
最优 Eps/km	0.63	0.53	0.42	0.32	0.53	0.32	0.42	0.53	0.32

使用 MDCH 作为评估函数, 自适应选择最优聚类尺度的有效性已被 PAI 验证. 其有效的原因在于: ① 根据式(5)与(6)可以发现, MDCH 与 PAI 都以分子衡量热点预测的精准度, 而以分母(面积)作为惩罚项, 两者相除以衡量热点的准确率. 所以 MDCH 与 PAI 实际衡量的是相似的指标; ② MDCH 与 PAI 最大的区别在于, 前者在同一周期内衡量准确率, 后者以下一周期衡量上一周期准确率. 研究已证明^[20-21], 财产类犯罪存在临近重复犯罪的现象, 所以以 3.1 时间热点的结果作为预测的周期可以弱化时间因素, 使 MDCH 有效替代 PAI. 这种替代将在聚类的时候选择最优尺度绘制犯罪热点预测图, 简化过程的同时确保了准确率.

3.2.3 热点区域制图

根据每月热点预测图可知下月易发生偷车案件的空间范围. 本文使用 2015 年 10 个月的最优 Eps 聚类结果(最优尺度 DBSCAN 聚类图), 结合案件数量格网分布来绘制犯罪热点预测图(见图 6), 供警务参考使用. 具体热点区域为研究区西北部的 A 地铁站周边, 以及东北部三块区域, B 地铁站, C 地铁站, 及 D 周边. 上述区域为案件发生最为频繁而密集的地方, 也即, 它们从时间角度看, 是 2015 年 10 个月里频数最高的地区, 从空间角度看, 是 MDSCAN 聚类密度最高的地区. 地铁站因通勤人流量大, 来往地铁站与居住地的交通方式主要为自行车, 故有大量“三车”停放. 而 D 周边有较多居住区, 除此之外, 共有四所高校在此周围, 商贩密集、大学生防护意识较差等多项因素综合形成案件热点区域.

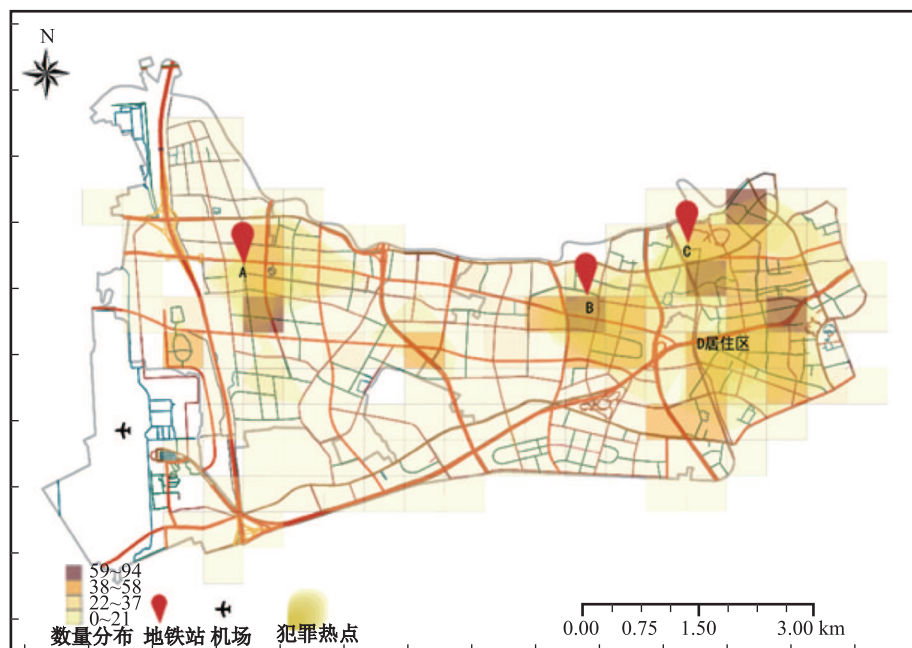


图6 研究区域偷车案件热点分布预测图

Fig.6 Predicted spatial hotspots of vehicle thefts at the study site

4 结 论

本文针对以偷盗“三车”为例的犯罪案件进行热点分析与预测,主要内容与方法包括:①提出尺度法整合时间数据,并以此法提取时间热点周期;②将DBSCAN的尺度变量多尺度化,设计针对性目标函数改进DBSCAN算法,单独或叠置使用后探查最优空间热点。

结果表明:①尺度法与严格法的结合,可以在数据量不足的情况下探测到犯罪周期,弥补了严格法的缺陷。②依据探测到的时间热点(周期),以MDBSCAN算法绘制的热点区域,能够缓解划分过硬问题,且以MDCH为评估函数时可以将空间热点探测与预测评估同时进行,提高效率,保持准确率。③尺度法整合数据后小波分析可知,研究区内的偷盗“三车”案件具有较明显的30 d周期和极明显的20 h周期。这一结论可能与警务工作人员的轮班岗位时间等有关,应当引起警务工作者注意。④MDBSCAN获取各周期最优空间热点图并叠置分析后,发现个别地铁站与居住区属于高发区,应在后期的实际警务行动中重点预防。本文将多尺度分析分别应用于时间与空间热点探查,未来还需进一步研究将其用于时空结合的分析方法中。

[参 考 文 献]

- [1] BRANTINGHAM P L, BRANTINGHAM P J. Patterns in Crime [M]. New York: Macmillan, 1984: 1-403.
- [2] GRUBESIC T H, MACK E A. Spatio-temporal interaction of urban crime [J]. Journal of Quantitative Criminology, 2008, 24(3): 285-306.
- [3] BRUNSDON C, CORCORAN J, HIGGS G. Visualising space and time in crime patterns: A comparison of methods [J]. Computers, Environment and Urban Systems, 2007, 31(1): 52-75.
- [4] GRUBESIC T H. On The Application of fuzzy clustering for crime hot spot detection [J]. Journal of Quantitative Criminology, 2006, 22(1): 77-105.
- [5] ISMAIL S, RAMLI N. Short-term crime forecasting in Kedah [J]. Procedia - Social and Behavioral Sciences, 2013, 91: 654-660.

- [6] BUNN D W, VASSILOPOULOS A I. Using group seasonal indices in multi-item short-term forecasting [J]. *International Journal of Forecasting*, 1993, 9(4): 517-526.
- [7] GORR W, HARRIES R. Introduction to crime forecasting [J]. *International Journal of Forecasting*, 2003, 19(4): 551-555.
- [8] TOMPSON L, TOWNSLEY M. (Looking) Back to the future: Using space—time patterns to better predict the location of street crime [J]. *International Journal of Police Science & Management*, 2010, 12(1): 23-40.
- [9] CRAGLIA M, HAINING R, WILES P. A comparative evaluation of approaches to urban crime pattern analysis [J]. *Urban Studies*, 2000, 37(4): 711-729.
- [10] RATCLIFFE J H. Aoristic analysis: the spatial interpretation of unspecific temporal events [J]. *International Journal of Geographical Information Science*, 2000, 14(7): 669-679.
- [11] RATCLIFFE J H, MCCULLAGH M J. Aoristic crime analysis [J]. *International Journal of Geographical Information Science*, 1998, 12(7): 751-764.
- [12] PATTEN I T V, MCKELDIN-CONER J, COX D. A microspatial analysis of robbery: Prospective hot spotting in a small city [J]. *Crime Mapping: A Journal of Research and Practice*, 2009, 1(1): 7-32.
- [13] 杜传明. 百度地图 API 在小型地理信息系统中的应用[J]. *测绘与空间地理信息*, 2011, 34(2): 152-153.
- [14] 兀伟, 邓国庆, 武晓莉, 等. 地理格网及在统计分析中应用的探讨[J]. *测绘标准化*, 2015, (3): 22-26.
- [15] JEVREJEVA S, MOORE J C, GRINSTED A. Influence of the arctic oscillation and El Niño-Southern Oscillation (ENSO) on ice conditions in the Baltic Sea: The wavelet approach [J]. *Journal of Geophysical Research: Atmospheres*, 2003, 108(D21): 1-11.
- [16] MIAO J, LIU G, CAO B, et al. Identification of strong karst groundwater runoff belt by cross wavelet transform [J]. *Water Resources Management*, 2014, 28(10): 2903-2916.
- [17] WANG W, HU S, LI Y. Wavelet transform method for synthetic generation of daily streamflow [J]. *Water Resources Management*, 2010, 25(1): 41-57.
- [18] LABAT D. Cross wavelet analyses of annual continental freshwater discharge and selected climate indices [J]. *Journal of Hydrology*, 2010, 385(1): 269-278.
- [19] ESTER M, KRIEGEL H, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise[C]. *Knowledge Discovery and Data Mining*, 1996: 226-231.
- [20] 徐冲, 柳林, 周素红. DP 半岛街头抢劫案件的临近重复发生模式 [J]. *地理研究*, 2015, 34(4): 384-394.
- [21] SHORT M B, D'ORSOGNA M R, BRANTINGHAM P J, et al. Measuring and modeling repeat and near-repeat burglary effects [J]. *Journal of Quantitative Criminology*, 2009, 25(3): 325-339.

(责任编辑: 李万会)