

doi: 10.11830/ISSN.1000-5013.201605025



采用时空重排扫描统计量的 犯罪集聚可视分析

张永田, 吴升

(福州大学 福建省空间信息工程研究中心, 福建 福州 350003)

摘要: 首先, 基于时空重排扫描统计量方法, 对福州市 2014 年案事件数据进行日尺度与小时尺度的时空重排扫描统计, 得到不同尺度下的犯罪时空集聚区. 然后, 结合数据可视化技术, 对分析结果进行多尺度、多视角的交互式可视化设计. 结果表明: 用户可以通过选择日期尺度, 按月或按月日期段、周日期段查看犯罪热点空间分布; 也可以通过选择小时尺度, 按时段查看周日期或节假日犯罪热点的空间分布.

关键词: 扫描统计; 犯罪热点; 交互式; 可视分析; 空间尺度; 时空聚集

中图分类号: P 208 **文献标志码:** A **文章编号:** 1000-5013(2016)05-0649-05

Visual Analysis of Crime Clusters Based on Space-Time Permutation Scan Statistic

ZHANG Yongtian, WU Sheng

(Spatial Information Research Center of Fujian, Fuzhou University, Fuzhou 350003, China)

Abstract: Firstly, based on space-time permutation scan statistic method, this paper analyzes crime of Fuzhou City in 2014 in days and hours, the space-time crime clusters in different scales are obtained. Then, combine with the data visualization technology and propose interactive visualization design for the results of the analysis in multi-scale and multi perspective. The results show that the method above enables users to view crime hot space distribution by date scale in monthly, month-date segment or week-date segment and by hour scale in week-date segment or holiday segment.

Keywords: scan statistic; crime hot spots; interactive; visualization; space scale; spatio-temporal aggregate

随着公安部“金盾工程”二期的完成, 全国公安通信网络和信息系统应用水平得到进一步的提升, 案件信息每年以 100~120 万条的速度递增^[1]. 面对不断增长的海量信息, 缺乏有效的数据挖掘及可视化分析手段使公安部门陷入了“库中有数而心中无数”的困境. 传统的犯罪空间数据挖掘方法将时间和空间信息割裂开, 无法全方位、立体地挖掘更深层的、富有价值的信息. 此外, 犯罪可视化与犯罪数据分析出现相互脱节的现象^[2], 特别是交互式可视化的设计方面尤为欠缺. 扫描统计法是空间统计学的重要方法之一, 众多学者对其进行了相关研究^[3-11]. 这些研究得到的犯罪集聚热点空间范围广, 时间跨度大, 对犯罪预防的指导意义不大. 本文选取更小的时间、空间尺度, 针对目前犯罪时空数据挖掘方法与犯罪时空数据交互式可视分析的不足^[12], 应用时空重排扫描统计方法探测犯罪的时空集聚特性.

收稿日期: 2015-09-09

通信作者: 吴升(1972-), 男, 教授, 博士, 主要从事时空数据分析与可视化、数字区域与智慧城市、应急信息系统的研究. E-mail: ws0110@163.com.

基金项目: 国家“863”重大科技项目(2012AA12A208)

1 方法原理

运用时空重排扫描统计法探测案事件的集聚区域. 利用一个圆柱体扫描窗口, 其底面中心对应于探测的地理区域中的一点, 高则对应于相应的时间区间. 底面半径逐渐增大, 使扫描的空间区域不断扩大, 直至设定的最大半径. 同时, 圆柱体的高度也逐渐增加, 每次增加一个规定的时间单位, 直至预设的上限为止. 扫描窗口的变化过程在整个研究区域内重复进行, 最终将生成数量庞大的扫描窗口. 采用广义似然比(GLR)判断每个扫描窗口内的案事件数量是否异常, 其值反映该窗口为集聚的可能性, 具有最大函数值的窗口即最有可能是集聚的^[13]. 然后, 利用蒙特卡罗假设检验方法对得到的候选集聚区域进行非随机性的置信度分析, 最终得到合理的犯罪集聚区^[8]. 具体的建模过程^[14]如下所示.

某一研究区 z 在 d 天中的案件数为 $C_{z,d}$, 则所有区域在所有时间范围内的总案件量 C 为

$$C = \sum_z \sum_d C_{z,d}. \tag{1}$$

每个单位区域每天的预期案件量 $\mu_{z,d}$ 为

$$\mu_{z,d} = \frac{1}{C} (\sum_z C_{z,d}) (\sum_d C_{z,d}). \tag{2}$$

式(2)中: $\sum_z C_{z,d}$ 为 d 天中整个研究区的案件量; $\sum_d C_{z,d}$ 为区域 z 中整个时间范围内的案件量.

则圆柱体扫描窗口 A 内的预期案件量为

$$\mu_A = \sum_{(z,d) \in A} \mu_{z,d}. \tag{3}$$

令 C_A 为圆柱体窗口 A 中的实际案件量, C_A 服从均数为 μ_A 的超几何分布, 其概率函数为

$$P(C_A) = \frac{\binom{\sum_{z \in A} C_{z,d}}{C_A}}{\binom{C}{C_A}} \frac{\binom{C - \sum_{z \in A} C_{z,d}}{C - C_A}}{\binom{C}{C - C_A}}. \tag{4}$$

当 $\sum_{z \in A} C_{z,d}$ 和 $\sum_{d \in A} C_{z,d}$ 相对于 C 而言非常小时, C_A 近似服从均值为 μ_A 的泊松分布. 基于此, 可采用泊松广义似然函数判定圆柱体窗口 A 中的案件量是否为集聚, GLR 的表达式为

$$GLR = \left\{ \frac{C_A}{\mu_A} \right\}^{C_A} \left\{ \frac{C - C_A}{C - \mu_A} \right\}^{C - C_A}. \tag{4}$$

广义似然函数值反映窗口为集聚的可能性. 因此, 在所有的候选集聚窗口里, 最大的窗口最不可能是随机发生的, 最有可能是有效集聚的. 通过计算扫描统计量能够得出大量的可能为集聚的窗口, 但为了验证探测到的集聚窗口的非随机性, 还需要对这些区域进行置信度分析. 扫描窗口统计量的无效假设为: 案件量在时间和空间上的分布完全随机. 要得到扫描窗口统计量的概率分布极为困难, 即使对一维的时间扫描统计量而言, 其概率分布也十分复杂.

因此, 通常情况下, 可运用蒙特卡罗假设检验方法进行 P 值的计算. 首先, 根据总的案件数生成 N 个随机分布的数据集, 每个数据集都能够得出各自最大的似然函数值, 再将真实数据集中窗口的 GLR 与 N 个随机生成的数据集的 N 个最大似然函数值进行由小到大的排序. 若真实数据集某窗口的 GLR 排在第 Q 位, 则 P 值为 $Q/N+1$. 因此, Q 的排序越靠前, P 值越小, 说明集聚窗口属于随机的可能性越小. 例如, 利用蒙特卡罗方法生成 999 个随机分布数据集, 某窗口的广义似然比排在第 30 位, 则 $P=30/999+1=0.03$, 则该集聚窗口为随机的概率为 3%; 若排序为 10, 则随机的概率为 1%.

2 多时间尺度时空集聚可视分析方法

研究区域为福州市行政管辖区中的鼓楼区、仓山区、台江区全境, 以及晋安区与马尾区的部分街道(北纬 $25^{\circ}57' \sim 26^{\circ}24'$, 东经 $119^{\circ}10' \sim 119^{\circ}39'$). 研究区域面积大约为 408 km^2 , 常住人口大约为 276.3 万^[13]. 数据为经过脱敏脱密处理后的 2014 年福州市案件数据.

2.1 日尺度犯罪时空

利用 SaTScan 9.3 软件对 2014 年福州各月案事件数据进行时空重排扫描统计分析. 扫描窗口的最

大空间上限为 300 m,即底面半径最大为 300 m;窗口的时间跨度为 1~15 d,时间变化步长为 1 d;蒙特卡罗法模拟次数为 999,当 $P<0.05$ 时,认为犯罪集聚.以 1 月份盗窃案件为例,时空重排扫描统计后的结果,如表 1 所示.表 1 中: r 为半径; n 为案件数.

表 1 1 月份盗窃电动车案件聚集区探测结果

Tab.1 Detection results of stealing electric bikes clusters in January

序号	中心经纬度坐标/(°)	r/m	开始日期	结束日期	n	P
1	119.458 28,25.993 40	114.3	2	2	3	0.027 2
2	119.304 13,26.092 97	168.3	15	16	6	0.009 0
3	119.305 90,26.096 02	236.0	22	22	4	0.028 2
4	119.355 92,26.065 74	226.5	25	25	3	0.031 8
5	119.285 44,26.067 16	220.1	26	29	4	0.022 0
6	119.285 27,26.094 48	84.6	31	31	3	0.018 0

由表 1 可知:通过蒙特卡罗方法计算得出的 6 个犯罪热点的 P 值为 0.009 0~0.031 8,表明这些热点集聚区域为随机的概率是 0.90%~3.18%,即集聚区属于随机的可能性非常小,因此,具有很高的非随机性,是有意义的集聚区.

为了从多角度查看月尺度下各个犯罪热点的时空分布情况,设计以下 3 种交互可视方式.

1) 按月查看犯罪热点空间分布.如图 1 所示.用户可以通过左上部分的“月份轴”或“上月”、“下月”箭头选择某一月份.然后,选取案件类别中的某一类案件(缺省为所有种类案件).此时,甘特图用空心矩形表示不同犯罪热点所在的日期段,右侧地图用空心圆形表示该月所有犯罪热点的空间分布,圆的大小代表了热点范围(扫描半径).

用户可以在甘特图上选取该月的某个日期段,右侧地图以灰色圆形显示所选日期对应的热点位置;或者在右侧地图上选择某个热点,可查看该热点的所属辖区、日期、热点半径、案事件数量等详细信息,左侧甘特图上将用灰色矩形显示该热点发生的日期段.

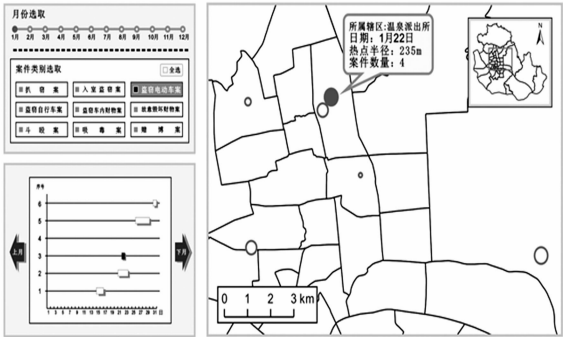


图 1 1 月份盗窃电动车案件聚集区可视分析

Fig.1 Visual analysis of stealing electric bikes clusters in January

2) 按月日期段查看犯罪热点空间分布,如图 2 所示.用户可以通过左上部分的“月日期轴”选取日期段(如 6~8 日),然后,选取案件类别中的某一类案件(缺省为所有种类案件).

此时,年历图上所有 6~8 日出现犯罪热点的都用矩形框加以强调.右侧地图用空心圆形表示当年所有 6~8 日出现犯罪热点的空间分布.用户也可在地图上选取某一点查看该热点详情,年历图将以灰底矩形框着重显示该热点对应的月日期段.

3) 按周日期段查看犯罪热点空间分布,如图 3 所示.用户可以通过左上部分的“周日期轴”选取一定日期段(如周一),也可选取案件类别中的某一类案件(缺省为所有种类案件).

此时,年历图上当年在周一出现赌博犯罪热点的日期用矩形框加以强调,右侧地图用空心圆形表示当年在周一出现犯罪热点的空间分布.用户也可在地图上选取某一点查看该热点详情,年历图将以灰底矩形框着重显示该热点对应的周日期段.

2.2 小时尺度犯罪时空

利用 SaTScan 9.3 软件对 2014 年福州案事件数据按“小时”进行时空重排扫描统计分析.半径上限为 300 m,时间跨度为 1~3 h,时间变化步长为 1 h;蒙特卡罗法模拟次数为 999,当 $P<0.05$ 时,认为犯罪集聚.以周六案事件数据为例,时空重排扫描统计后的结果,如表 2 所示.

由表 2 可知:8 个犯罪热点的 P 值为 0.013 0~0.049 9,表明了这些热点集聚区域为随机的概率是 1.30%~4.99%,属于随机的可能性很小,具有较高的非随机性.

在时段热点的可视化设计方面,可以通过时钟图直观地反映热点的案发时段,如图 4 所示.将圆分

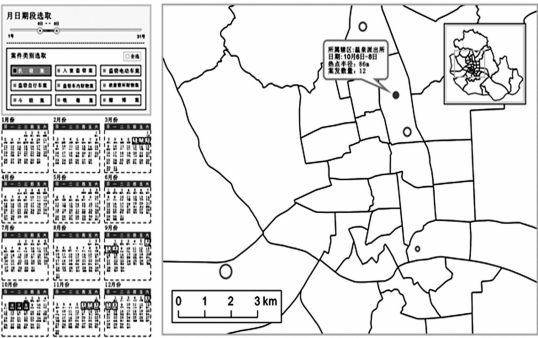


图 2 6~8 日扒窃案件聚集区可视分析
Fig. 2 Visual analysis of pickpocket clusters on 6 to 8

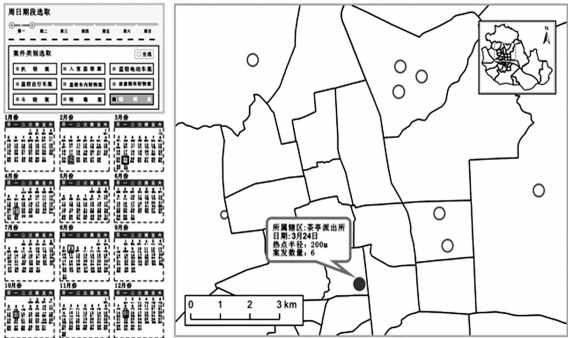
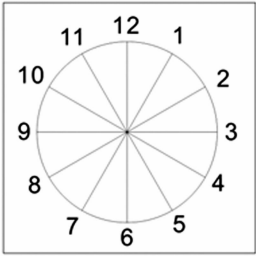


图 3 周一赌博案件聚集区可视分析
Fig. 3 Visual analysis of gambling clusters on Monday

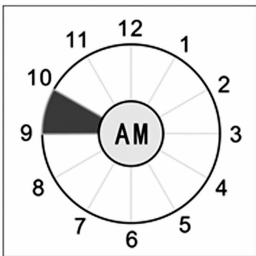
为 12 个区域,如图 4(a)所示. 图 4(b),(c)中:灰色区域表示热点时段;圆环内部的“AM”表示午前 0~12 时,“PM”表示午后 12~0 时. 由图 4(b),(c)可知:午前犯罪热点出现在 9~10 时;午后犯罪热点出现在 7~8 时.

表 2 周六入室盗窃案件聚集区探测结果
Tab. 2 Detection results of burglary clusters on Saturday

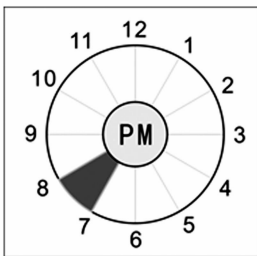
序号	中心经纬度坐标/(°)	r/m	开始时间	结束时间	n	P
1	119.288 92,26.061 87	126.8	0:00	0:59	2	0.013 0
2	119.305 51,26.060 88	220.1	2:00	3:00	2	0.024 0
3	119.277 99,26.077 98	102.9	4:00	4:59	2	0.026 5
4	119.293 38,26.038 51	65.3	11:00	11:59	2	0.049 3
5	119.309 91,26.110 10	253.2	15:00	16:00	4	0.027 9
6	119.326 91,26.034 65	142.2	20:00	20:59	3	0.030 7
7	119.302 24,26.064 45	77.9	22:00	22:59	2	0.039 7
8	119.264 08,26.109 07	55.3	23:00	23:59	2	0.049 9



(a) 时段热点划分



(b) 午前热点时段



(c) 午后热点时段

图 4 时段热点可视化设计

Fig. 4 Visual design of time periods hotspot

为了方便用户从多个角度对小时尺度的犯罪热点进行时空分析,设计按周日期查看犯罪热点空间分布,如图 5 所示.

由图 5 可知:用户可通过左上部分的“时钟图”设置某一时段范围(如 15~16 时),可选取某一周日期(缺省为所有周日期)或是将选项卡切换到节假日,选取某一节假日,再选取某一案件类别(缺省为所有种类案件),右侧地图用空心圆形表示当年在周六 15~16 时出现犯罪热点的空间分布. 用户也可选取地图上的某一热点查看详情.

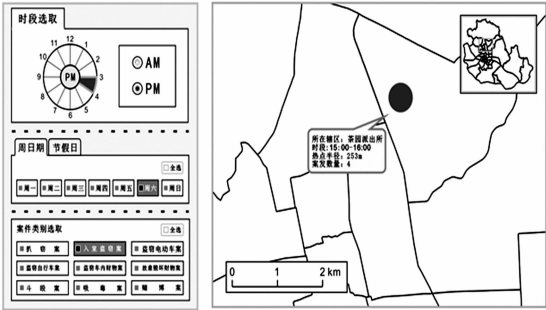


图 5 周六 15~16 时入室盗窃案件聚集区可视分析
Fig. 5 Visual analysis of burglary clusters from 15 to 16 pm on Saturday

3 结束语

基于时空数据挖掘理论、信息可视化技术,运用时空重排扫描统计方法,对福州市主城区2014年案事件进行时空聚集性分析.对案事件数据分别进行日尺度及小时尺度的时空重排扫描统计,将得到的分析结果运用信息可视技术对犯罪聚集区进行表达.通过对数据挖掘的结果进行多尺度、多视角的交互可视化设计,方便用户对犯罪热点进行多角度的查看分析,为公安部门科学决策提供技术支持.下一步将研究犯罪的预测,将研究区域格网化,分析各种场景(如商场、医院、KTV、网吧等)对犯罪的影响并结合有效的预测模型,对各个网格犯罪的案发概率进行计算,为公安部门进一步细化警力配置提供依据.

参考文献:

- [1] 程春惠.公安犯罪案件文本挖掘关键技术研究[D].杭州:浙江大学,2010:1-2.
- [2] 李代超,吴升.面向不同主题的犯罪大数据可视分析[J].地球信息科学学报,2014(5):735-745.
- [3] NAUS J I. The distribution of the size of the maximum cluster of points on a line[J]. Journal of the American Statistical Association,1965,60(310):532-538.
- [4] KULLDORFF M, HEFFERNAN R, HARTMAN J. A space time permutation scan statistic for disease outbreak detection[J]. Plos Medicine,2005,2(3):216-224.
- [5] NAKAYA T, YANO K. Visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics[J]. Transactions in GIS,2010,14(3):223-239.
- [6] LEITNER M, HELBICH M. The impact of hurricanes on crime: A spatio-temporal analysis in the city of Houston, Texas[J]. Cartography and Geographic Information Science,2011,38(2):213-221.
- [7] 王占宏.基于扫描统计方法的上海犯罪时空热点分析[D].上海:华东师范大学,2013:89-97.
- [8] ZEOLI A, PIZARRO J, GRADY S, et al. Homicide as Infectious disease: Using public health methods to investigate the diffusion of homicide[J]. Justice Quarterly,2014,31(3):609-632.
- [9] 王培安,罗卫华,白永平.基于空间自相关和时空扫描统计量的聚集比较分析[J].人文地理,2012(2):119-127.
- [10] 唐咸艳,李峤,黄秋兰,等.时空扫描统计量三维可视化的实现[J].中国卫生统计,2015(4):693-694.
- [11] 李小洲,王劲峰.空间扫描统计量方法中候选聚集区域生成的快速算法[J].地球信息科学学报,2013(4):505-511.
- [12] 福州市统计局.福州统计年鉴[M].北京:中国统计出版社,2012:22-23.
- [13] 殷菲.时-空扫描统计量在传染病早期预警中的应用研究[D].成都:四川大学,2007:20-21.
- [14] LEITNER M. Crime modeling and mapping using geospatial technologies[M]. Dordrecht: Springer Netherlands, 2013:320-436.
- [15] 周丽君,张兴裕,马越,等.前瞻性时空扫描统计量与时空重排扫描统计量在传染病聚集性探测中的适用性探讨[J].现代预防医学,2012(5):1068-1070.

(责任编辑:钱筠 英文审校:吴逢铁)