

基于历史命中率的时空重排扫描最大搜索半径选取方法及应用实验

杨威, 龙华*, 王美

(昆明理工大学信息工程与自动化学院, 云南 昆明 650500)

摘要:最大搜索半径是控制时空重排扫描统计方法性能的关键参数,而搜索半径过大会造成资源浪费。为此,该文提出了一种基于历史命中率的时空重排扫描最大搜索半径选取方法。通过迭代不同最大搜索半径对历史性时空数据集进行时空重排扫描统计,根据不同最大搜索半径所得扫描结果计算预警命中率,选取命中率最高的最大搜索半径作为前瞻性扫描的最大搜索半径,避免了因最大搜索半径过大导致的资源浪费,提升了模型的性能;基于旧金山地区火灾数据集的验证结果证明了该方法的有效性。

关键词:时空重排扫描;命中率;搜索半径

中图分类号:TP311.13 **文献标识码:**A **文章编号:**1672-0504(2020)02-0022-06

0 引言

随着时空数据逐渐增多,从异常的时空数据中探索事物发展的特殊规律,对于相关领域的研究与发展至关重要;但基于时空数据的特殊性,传统的数据分析方法无法直观理解其中蕴含的信息。因此,时空异常探测方法^[1]作为新兴的时空数据挖掘手段,受到了时空数据研究人员的广泛关注。

相对于 Gi、LISA^[2]等纯空间的不规则聚集区探测方法,时空异常探测方法同时考虑了空间因素和时间因素,使得异常探测的过程更贴近实际,探测结果也更可靠。扫描统计方法^[3]是一种常用的时空异常探测方法,广泛应用在公共卫生^[4]、公共安全^[5]以及气象^[6]等领域。Kulldorff 等为提升该方法的性能,在原扫描模型的基础上提出了可变大小的圆形扫描窗口^[7],继而加入时间参量生成圆柱扫描窗口,提出了时空扫描统计方法^[8],通过划定某一区域,不断改变扫描范围和扫描中心点位置,以搜索整个研究区域内的时空异常区域;并进一步提出了时空重排扫描方法^[9],通过新的概率模型减少了时空扫描统计方法对风险人口数的依赖性,以探测和评估时空域内存在的异常区域,从而解释空间或时空随机性无法解释的问题。

在时空重排扫描方法中,圆柱体扫描窗口大小由窗口形状及最大扫描半径决定,虽然 Kulldorff 等

对模型进行了不断修改^[7-9],但扫描窗口对于不规则大型实际数据集的探测仍有局限性。在最初引入时空扫描统计方法的流行病学和公共卫生领域,研究人员依据研究区域面积 50% 所对应的圆形半径来设定扫描窗口的最大搜索半径^[9],但当研究区域面积较大时,过大的搜索半径会导致较大的资源消耗。为此,相关学者对优化扫描窗口的选取方法进行了研究。例如, Takahashi 等提出了一种灵活时空扫描统计方法,通过组合相邻子区域生成具有不规则形状的窗口进行扫描^[10],该算法扩展了扫描统计方法对不规则形状区域的扫描,并在算法中加入集群尺寸(一般为整个研究区域面积的 10%~15%)控制,可进行较小尺寸异常点的检测。为更迅速地识别出空间聚集, Walther 在此基础上提出了一种对空间聚集快速检测和优化的方法,依据扫描窗口的大小对扫描窗口进行分组,然后将不同的临界值赋予各组,通过该校准方式提升不规则区域探测速度^[11]。为提升不规则区域异常探测效果, Han 等提出了一种利用基尼系数对扫描窗口扫描结果进行修正的方法^[12],能更精确地对扫描窗口内非重叠集群进行判别,从而发现较小的非重叠集群与包含在大集群内的较小集群,提升了扫描窗口探测效果;万幼等在此基础上提出一种改进的不规则形状时空异常聚类模式挖掘方法,通过空间网格结构与蚁群优化算法的集合,进一步提升了时空扫描方法对于不规则形状

收稿日期:2019-07-30; 修回日期:2019-10-22

基金项目:地区科学基金项目(61761025)

作者简介:杨威(1992-),男,硕士研究生,研究方向为信息处理。* 通讯作者 E-mail:longhua@kmust.edu.cn

时空异常的探测精度^[13]。为解决实际研究中不规则形状对扫描的影响, Katragadda 等提出了一种利用多边形传播算法进行时空聚集探测的方法^[14], 通过对多个三角形组成的多边形区域进行扫描, 从而进行异常检测, 但该算法的时间复杂度与观测数量均为非线性, 对于较大的数据集不能很好地进行扩展。

综上, 虽然相关学者对圆柱体扫描窗口的选取方法进行了优化, 但对于扫描窗口最大搜索半径的选取, 依然是一个悬而未决的问题。因此, 本文提出了一种基于历史命中率的时空重排扫描最大搜索半径的选取方法。

1 研究方法

1.1 时空重排扫描方法

时空重排扫描方法同基于泊松分布的时空扫描统计方法一致^[9], 在设定扫描窗口最大搜索半径和扫描时间阈值的基础上, 均是利用圆柱体扫描窗口对研究区域进行扫描, 从而获取研究区域内事件潜在爆发的时空区域。该方法具体实现过程为: 1) 选择并迭代研究区域内有限数量的事件观测点, 以每个观测点作为圆柱底面的中心点进行扫描, 同时圆柱体的底面半径从零增加至最大搜索半径, 圆柱体的高度表示扫描开始时间与结束时间的间隔, 从一天增加至用户所设定的最大天数; 2) 对于研究区域内每个事件观测点, 通过迭代所有的时间长度及搜索半径, 最终形成若干个圆柱体覆盖的研究区域, 计算研究区域内外预期事件发生数与实际发生数的对数似然比, 通过比较二者的差异判断该区域是否为时空异常区域。

对于从局部开始蔓延至整个区域的事件爆发, 由于局部事件数量的增加不足以使整个地区的事件数呈现上升趋势, 因此, 单纯的时间或空间性探测方法对于此类事件的聚集探测存在滞后性, 而时空重排扫描方法对单纯的时间或空间聚集性进行了修正, 充分利用了数据中的空间及时间信息, 从而完成了更为有效的早期预警^[15]。时空重排扫描方法的特点是: 1) 采用了新的概率模型, 自动纠正了单纯空间或时间聚集性, 识别了时空交互作用; 2) 解决了研究人员在无法获取研究区域风险人口数据情况下时空扫描方法使用的问题^[15]。因此, 在时空重排扫描方法中, 只需依据事件发生数就可进行时空扫描。

假设研究区域内规定时段内发生事件总数 M 的计算公式为:

$$M = \sum_z \sum_d M_{z,d} \quad (1)$$

式中: $M_{z,d}$ 是观测点 z 在 d 天所发生的事件数; z 是空间点位置 (x_i, y_j) 在研究区域内映射的标记值, $z \in \{Z_1, Z_2, \dots, Z_m, \dots, Z_n\}, 1 \leq m \leq n, n \in \mathbf{R}; d$ 是规定时段内的时间点, $d \in \{1, 2, \dots, a, \dots, D\}, 1 \leq a \leq D, D \in \mathbf{R}$ 。

进而可以得到每个观测点每天的预期事件发生量 $\mu_{z,d}$:

$$\mu_{z,d} = (\sum_z M_{z,d} \sum_d M_{z,d}) / M \quad (2)$$

设观测点集合中任意一点是底面半径为 R_A 的圆柱体扫描窗口 A 的圆心, 则圆柱体扫描窗口 A 内预期的事件数 μ_A 是该圆柱体扫描窗口内所有观测点在研究时段内预期事件数的总和:

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

假设圆柱体扫描窗口 A 中实际事件总量为 M_A , 在没有发生时空交互时, M_A 服从均值为 μ_A 的超几何分布^[9], 其概率函数 $P(M_A)$ 的计算公式为:

$$P(M_A) = \left(C_{\sum_{z \in A} M_{z,d}}^{M_A} C_{M - \sum_{z \in A} M_{z,d}}^{\sum_{z \in A} M_{z,d} - M_A} \right) / C_M^{\sum_{z \in A} M_{z,d}} \quad (4)$$

当 $\sum_{z \in A} M_{z,d}$ 远小于研究区域内规定时段所发生的事件总数 M 时, M_A 可以近似服从泊松分布, 其均值为 μ_A ^[16]。基于该条件, 使用泊松广义对数似然比 (GLR) 作为圆柱体 A 中事件聚集程度的度量工具, 其计算公式为:

$$GLR = \ln \left(\frac{M_A}{\mu_A} \right)^{M_A} \left(\frac{M - M_A}{M - \mu_A} \right)^{M - M_A} I \left(\frac{M_A}{\mu_A} > \frac{M - M_A}{M - \mu_A} \right) \quad (5)$$

式中: 阶跃信号 I 为指示函数, 如果 $\frac{M_A}{\mu_A} > \frac{M - M_A}{M - \mu_A}$, 则 $I = 1$, 否则 $I = 0$ 。通过使用指示函数确保所搜索到的异常区域为热点聚集。

比较所有候选扫描圆柱体的 GLR , GLR 值越大, 其扫描圆柱体的聚集性越强, 则该区域的热点聚集性越强。为减少扫描结果的随机性, 采用蒙特卡罗假设检验方法对所得热点区域进行显著性检验, 并选取 p 值不为 1 的候选扫描圆柱体内的观测点作为事件预警点。

1.2 历史命中率

圆柱体 A 的最大底面半径依据最大搜索半径 $R_{A_{\max}}$ 而定, 一般为研究区域面积 50% 所对应的圆形半径。但当研究区域面积较大时, 过大的搜索半径会导致极大的资源消耗。故本文通过对历史性数据进行时空重排扫描, 根据历史命中率选取最大搜索半径, 然后利用该搜索半径对地区事件进行前瞻性时空重排扫描。相较于依据研究区域面积 50% 所对应的圆形半径来设定最大搜索半径, 该方法灵活性更好, 可减少因最大搜索半径过大引起的资源损耗。

历史命中率($Rate_A$)因其简单且易于理解,常被用于评价预警准确性^[17],其定义为在给定空间覆盖范围内,通过时空重排扫描方法得到的预警观测点数量 $N_{prospective}$ 与通过观察得到的实际事件的观测点数量 N_{true} 的百分比(式(6))。图 1 中表现为圆柱体扫描区域内实心圆点的数量与所有圆点数量的比值。

$$Rate_A = N_{prospective} / N_{true} \times 100\% \quad (6)$$

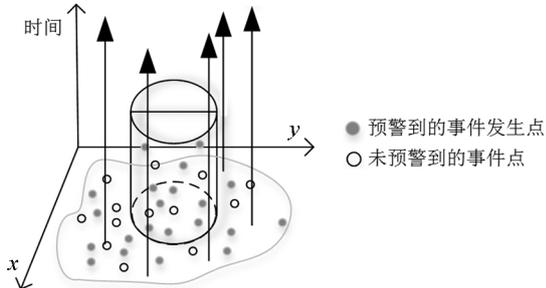


图 1 扫描圆柱体的命中率示意

Fig. 1 Schematic of the hit ratio of scanned cylinder

1.3 基于历史命中率的时空重排扫描最大搜索半径选取方法

在具有蔓延性的时空事件中,同一地区在相同的时间间隔内,事件所造成的最大蔓延距离是相对固定的^[18],即可根据历史时段预警结果来选取未来相同时间间隔内时空重排扫描的最大搜索半径。基于此原理,本文提出了基于历史命中率的时空重排扫描最大搜索半径选取方法,其具体实施步骤(图 2)为:通过迭代不同最大搜索半径,对历史时段数据集进行时空重排扫描统计,来预警后一时段事件的发生;根据式(6)计算后一时段的历史命中率,选取命中率最高的扫描结果所对应的最大搜索半径作为该数据集前瞻性扫描的最大搜索半径。

2 实验与结果分析

2.1 实验数据

相关研究^[19-22]证明,火灾事件可以使用时空扫描方法进行时空异常探测,同时,为便于实验结果的验证,本研究基于旧金山地区数据协调网站(<https://datasf.org/opendata/>)提供的“Fire Department Service”数据集,从中提取 2016—2019 年的火灾事件数据进行实验。由于公共安全事件的预警通常为短期预警,因此,本文选取 1 周作为扫描的时间阈值,即将 7 d 作为观测与预警时间阈值,从而增强事件间的相关性^[23]。为减小实验结果的偶然性和随机性,本研究共设计了 3 组实验进行验证;同时,为了更好地展现不同搜索半径下的探索效果,实验中引入 q 统计量作为分层异质性的探测方法^[24]。

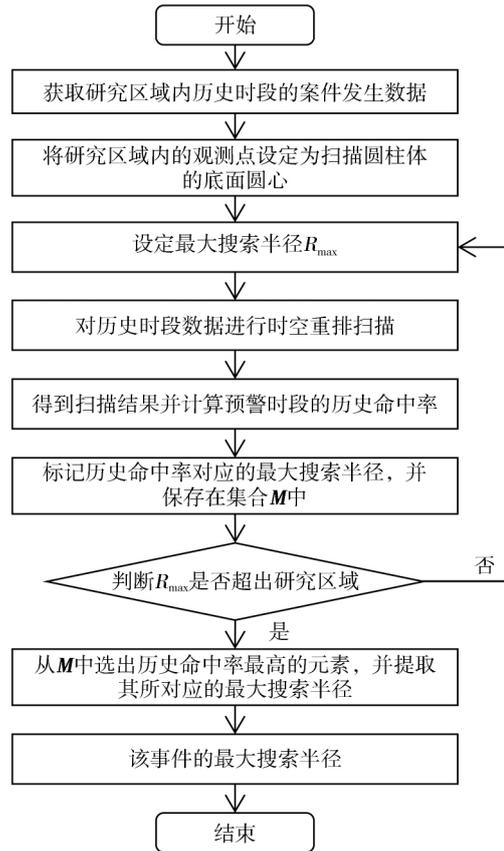


图 2 最大搜索半径选取方法流程

Fig. 2 Flow chart of the selecting method for maximum search radius

为方便时空重排扫描及历史命中率的计算,对 42 个观测点进行编号(表 1),并对每天不同观测点的火灾事件数进行统计,生成如表 2 所示的时空重排扫描数据集。

表 1 观测地点对应编号
Table 1 The corresponding number of observation sites

地点名称	地点编号
Bayview Hunters Point	1
Bernal Heights	2
Upper Market	3
Chinatown	4
Excelsior	5
Financial District	6
Glen Park	7
Golden Gate Park	8
...	...
Sea cliff	42

表 2 火灾事件数量统计结果
Table 2 Statistical results of the number of fire incidents

地点编号	日期	事件数量
1	2016/02/01	2
2	2016/02/01	2
3	2016/02/01	2
8	2016/02/01	17
13	2016/02/01	2
20	2016/02/01	2
21	2016/02/01	1
23	2016/02/01	1
...
39	2019/06/07	1

2.2 实验结果与分析

第一组实验选取 2016—2018 年每年 2 月 1 日—7 日的火灾事件数据作为历史性扫描数据进行时空重排扫描, 以预警 2016—2018 年每年 2 月 8 日—14 日的火灾事件发生数量; 为便于计算历史命中率, 同时从实验数据集中获取 2016—2018 年每年 2 月 8 日—14 日旧金山地区 42 个火灾观测点的火

灾实际发生情况(图 3, 彩图见封 2), 并对发生火灾的观测点数量进行统计(表 3), 最终得到该时段在不同搜索半径下的历史命中率(表 4)。可以看出, 时空重排扫描的最大搜索半径为 5 km 时, 时空重排扫描模型的命中率达到最大。因此, 在该事件集中, 在时间阈值为 1 周的研究范围内, 最大搜索半径为 5 km。

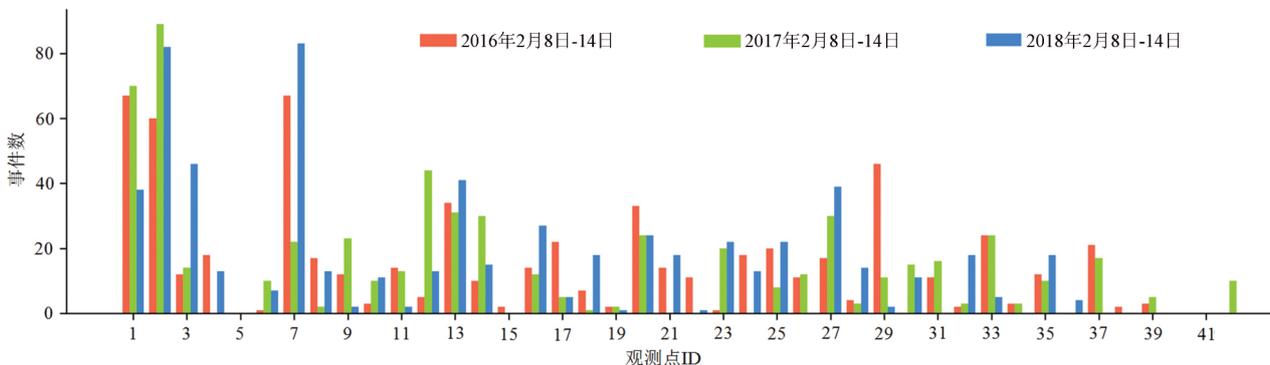


图 3 2016—2018 年 42 个观测点火灾实际发生情况(2 月 8 日—14 日)
Fig. 3 Actual fire situation at 42 observation sites from 2016 to 2018 (February 8th-14th)

表 3 2016—2018 年实际发生火灾的观测点情况(2 月 8 日—14 日)
Table 3 Observation sites of actual fires from 2016 to 2018 (February 8th-14th)

时间	实际发生火灾的观测点	观测点数量
2016/02/08—2016/02/14	[1,2,3,4,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,31,32,33,34,37,38,39]	35
2017/02/08—2017/02/14	[1,2,3,6,7,8,9,10,11,12,13,14,16,17,19,20,23,25,26,27,28,29,30,31,32,33,34,35,37,39,42]	31
2018/02/08—2018/02/14	[1,2,3,4,6,7,8,9,10,11,12,13,14,16,17,18,19,20,21,22,23,24,25,27,28,29,30,32,33,35,36]	31

表 4 2016—2018 年时空重排扫描结果命中率情况
Table 4 The hit ratio of space-time permutation scan results from 2016 to 2018

时间	搜索半径									
	1 km	2 km	3 km	4 km	5 km	6 km	7 km	8 km	9 km	10 km
2016/02/08—2016/02/14	25.7%	28.5%	28.5%	28.5%	31.4%	31.4%	31.4%	31.4%	31.4%	31.4%
2017/02/08—2017/02/14	22.6%	32.3%	32.3%	38.7%	41.9%	41.9%	41.9%	41.9%	41.9%	41.9%
2018/02/08—2018/02/14	22.6%	29.0%	41.9%	51.6%	61.3%	61.3%	61.3%	61.3%	61.3%	61.3%
2016/03/08—2016/03/14	20.6%	41.3%	41.3%	44.8%	48.2%	44.8%	44.8%	44.8%	44.8%	44.8%
2017/03/08—2017/03/14	15.1%	21.2%	24.2%	36.3%	45.5%	45.5%	45.5%	45.5%	45.5%	45.5%
2018/03/08—2018/03/14	12.5%	28.1%	37.5%	40.6%	43.7%	43.7%	43.7%	43.7%	43.7%	43.7%
2016/04/08—2016/04/14	22.2%	25.0%	36.1%	36.1%	38.8%	38.8%	38.8%	38.8%	38.8%	38.8%
2017/04/08—2017/04/14	14.7%	20.5%	29.4%	32.3%	52.9%	52.9%	52.9%	52.9%	52.9%	52.9%
2018/04/08—2018/04/14	18.7%	37.5%	50.0%	62.5%	65.6%	65.6%	65.6%	65.6%	65.6%	65.6%

第二组、第三组实验分别选取 2016—2018 年每年 3 月 1 日—7 日、每年 4 月 1 日—7 日的数据进行时空重排扫描, 重复第一组的实验过程, 得到如表 4 对应时段所示的命中率结果。可以看出, 时空重排扫描的最大搜索半径为 5 km 时, 时空重排扫描模型的命中率仍达最大。

进一步对不同最大搜索半径下的空间聚集情况进行实验(图 4), 可以看出, 3 组实验数据期间, q 统计量均先随着最大搜索半径的增大而递增, 在最大搜索半径为 5 km 时, q 统计量不再发生变化, 表明此时空间分异特性非常显著。

为验证本研究中提出的最大搜索半径选取方法的优越性, 使用原时空重排扫描最大搜索半径选取

方法进行对比实验。原方法中最大搜索半径为研究区域面积 50% 所对应的圆形半径, 本文中研究区旧金山总面积约为 600 km², 则近似选取最大搜索半径为 10 km。分别以 2019 年 2 月 1 日—7 日、3 月 1 日—7 日、4 月 1 日—7 日的数据作为扫描样本数据, 2019 年 2 月 8 日—14 日、3 月 8 日—14 日、4 月 8 日—14 日的数据作为验证数据, 使用两种最大搜索半径进行时空重排扫描, 并对扫描耗时进行统计(表 5), 发现采用 5 km 作为最大搜索半径耗时较少。

通过 3 组对比实验可以得出: 1) 当采用原始最大搜索半径选取方法时, 通常在达到最大搜索半径之前, 扫描结果的预警命中率已经达到最高, 说明该研究区域内, 在相同实验条件下, 得到最大预警命中

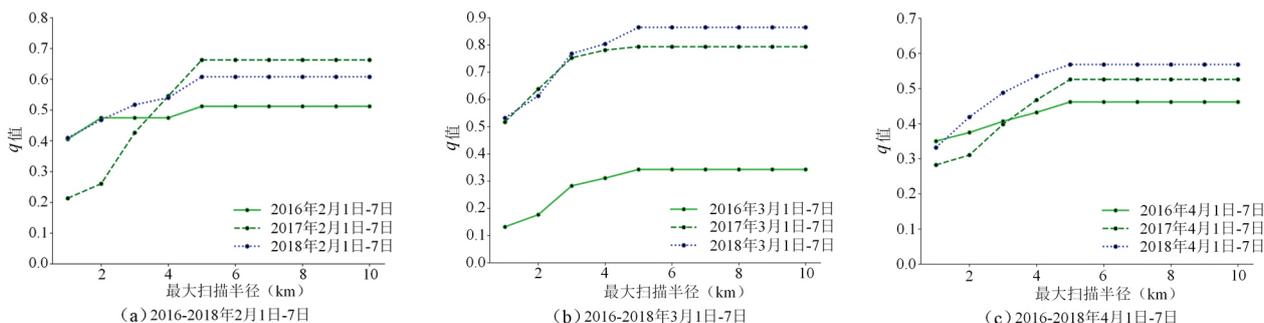


图 4 不同最大搜索半径空间异质性情况

Fig. 4 Spatial heterogeneity of different maximum search radii

表 5 2019 年 2—4 月时空重排扫描结果命中率情况
Table 5 The hit ratio of space-time permutation scan results from February to April 2019

时间范围	最大搜索半径(km)	历史命中率	耗时(s)
2019/02/08—2019/02/14	5	26.5%	320.14
	10	26.5%	633.81
2019/03/08—2019/03/14	5	32.0%	296.74
	10	32.0%	587.95
2019/04/08—2019/04/14	5	41.9%	321.82
	10	41.9%	607.61

率的最大搜索半径,要比采用原始最大搜索半径选取方法得到的最大搜索半径小。2)基于历史命中率的最大搜索半径选取方法可以更为快速、准确地获取该事件的最大搜索半径,并且在搜索半径处空间分异特性非常显著,减少了搜索半径过大所引起的资源损耗,提升了时空重排扫描方法的效能。

3 结论与展望

时空重排扫描统计方法是一种新兴的点事件时空异常探测方法,本文中被用作短期事件的热点事件预警。从蔓延性的角度看,在具有蔓延性的时空事件中,当其他因素都相同,只考虑时间与空间因素时,同一地区在相同的时间间隔内,事件所造成的蔓延距离是相对固定的。依据这一特性,本研究提出了基于历史命中率的时空重排扫描最大搜索半径的选取方法。该方法的重要性在于,在不影响扫描结果预警精度的情况下,合理选取了最大搜索半径,虽然历史命中率的计算额外增加了一定的资源消耗,但从模型长远使用角度考虑,合理的最大搜索半径可以减少今后对于该区域时空重排扫描时,因搜索半径过大造成的持续资源浪费。

由于在选取最大搜索半径时,需要先对该事件进行历史性时空重排扫描,故该方法对于数据较少的事件适用性不强,未来会对该问题进行优化。

参考文献:

[1] 邓敏,石岩,龚健雅,等. 时空异常探测方法研究综述[J]. 地理与地理信息科学,2016,32(6):43—50.

[2] 张王剑,杜志成,郭貌,等. 常用不规则聚集区探测方法[J]. 中国卫生统计,2016,33(1):175—178.

[3] GLAZ J, NAUS J, WALLENSTEIN S. Scan Statistics[M]. Berlin: Springer, 2001. 755—788.

[4] LU Y, DENG X J, CHEN J H, et al. Risk analysis of African swine fever in Poland based on spatio-temporal pattern and Latin hypercube sampling, 2014—2017 [J]. BMC Veterinary Research, 2019, 15(1): 160—171.

[5] 李智,李卫红. 点模式条件下的犯罪嫌疑人时空同现模式挖掘与分析[J]. 地球信息科学学报, 2018, 20(6): 827—836.

[6] 张婷,程昌秀,杨山力,等. 时空聚集性探测方法在极端高温事件聚集分析中的应用研究[J]. 地理与地理信息科学, 2019, 35(3): 51—57.

[7] KULLDORFF M. A spatial scan statistic[J]. Communications in Statistics-Theory and Methods, 1997, 26(6): 1481—1496.

[8] KULLDORFF M. Prospective time periodic geographical disease surveillance using a scan statistic[J]. Journal of the Royal Statistical Society; Series A (Statistics in Society), 2001, 164(1): 61—72.

[9] KULLDORFF M, HEFFERNAN R, HARTMAN J, et al. A space-time permutation scan statistic for disease outbreak detection [J]. PLoS Medicine, 2005, 2(3): e59.

[10] TAKAHASHI K, KULLDORFF M, TANGO T, et al. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring[J]. International Journal of Health Geographics, 2008, 7(1): 14—27.

[11] WALTHER G. Optimal and fast detection of spatial clusters with scan statistics[J]. The Annals of Statistics, 2010, 38(2): 1010—1033.

[12] HAN J H, ZHU L, KULLDORFF M, et al. Using Gini coefficient to determining optimal cluster reporting sizes for spatial scan statistics[J]. International Journal of Health Geographics, 2016, 15(1): 27—37.

[13] 万幼,周脚根,翁敏. 点集数据不规则形状时空异常聚类模式挖掘研究[J]. 武汉大学学报(信息科学版), 2017, 42(7): 924—930.

[14] KATRAGADDA S, CHEN J, ABBADY S. Spatial hotspot detection using polygon propagation[J]. International Journal of Digital Earth, 2019, 12(7): 825—842.

[15] 唐威艳,周红霞. 扫描统计及其在流行病学中的应用[J]. 中国卫生统计, 2011, 28(3): 332—337.

- [16] FORBES C, EVANS M, HASTINGS N, et al. Statistical Distributions (4th Edition)[J]. John Wiley & Sons, 2010.
- [17] BOWERS K J, JOHNSON S D, PEASE K. Prospective hot-spotting: The future of crime mapping?[J]. British Journal of Criminology, 2004, 44(5): 641–658.
- [18] FINNEY M A. An overview of FlamMap fire modeling capabilities[A]. Management-How to Measure Success: Conference Proceedings[C]. 2006, 41: 213–220.
- [19] LEE S, DAVIDSON R, OHNISHI N, et al. Fire following earthquake-Reviewing the state-of-the-art of modeling[J]. Earthquake Spectra, 2008, 24(4): 933–967.
- [20] GARDNER W D, THOMSON C R. Flame spread properties of forest products. Comparison and validation of prescribed Australian and North American flame spread test methods[J]. Fire and Materials, 1988, 12(2): 71–85.
- [21] TUIA D, RATLE F, LASAPONARA R, et al. Scan statistics analysis of forest fire clusters[J]. Communications in Nonlinear Science and Numerical Simulation, 2008, 13(8): 1689–1694.
- [22] VADREVU K P. Analysis of fire events and controlling factors in eastern India using spatial scan and multivariate statistics[J]. Geografiska Annaler: Series A, Physical Geography, 2008, 90(4): 315–328.
- [23] CORRAL Á, TELESKA L, LASAPONARA R. Scaling and correlations in the dynamics of forest-fire occurrence[J]. Physical Review E, 2008, 77(1): 016101.
- [24] WANG J F, ZHANG T L, FU B J. A measure of spatial stratified heterogeneity[J]. Ecological Indicators, 2016, 67: 250–256.

Method and Application Experiment for Selecting Maximum Search Radius of Space-Time Permutation Scan Based on Historical Hit Ratio

YANG Wei, LONG Hua, WANG Mei

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: The maximum search radius is a key parameter to control the performance of the space-time permutation scanning statistical method. But the oversized search radius will lead to a resource loss when the study area is too large. Therefore, this paper proposed a method to select the maximum search radius of space-time permutation scan based on the historical hit ratio. Firstly, the historical space-time dataset was scanned with the space-time permutation scanning statistical method by iterating different maximum search radii. Then, early warning hit ratios were calculated with the scan results. At last, the largest search radius with the highest hit ratio was selected as the maximum search radius for the prospective scans. This method avoided the resource waste caused by the oversized search radius and improved the performance of the model. This method was validated according to the fire dataset in San Francisco area.

Key words: space-time permutation scan; hit ratio; scan radius