

# Spatial Outlier Detection Based on TEN Model and Weighted Attribute

Yuezhong Zhu

College of Architecture & Surveying and Mapping, Jiangxi University of Science and Technology,  
Ganzhou Jiangxi  
Email: zxc2511836601@126.com

Received: Nov. 20<sup>th</sup>, 2018; accepted: Jan. 1<sup>st</sup>, 2019; published: Jan. 8<sup>th</sup>, 2019

---

## Abstract

A Delaunay triangulation based spatial outlier detection algorithm is proposed to solve the problems of the traditional spatial outlier detection algorithm, such as the high complexity of the algorithm and the large human influence on the construction of spatial neighbors. In this algorithm, Delaunay triangulation is carried out on the spatial data points to establish the spatial neighborhood. Based on the generated spatial neighborhood, outlier factors in the spatial neighborhood are defined according to SLOF algorithm, combined with the specific drilling data of a mine in a certain area. Experimental results show that this algorithm can effectively detect spatial outliers with low complexity and low human influence.

## Keywords

Spatial Outliers, Delaunay Triangulation, SLOF Outlier Factor

---

# 基于Delaunay三角剖分的空间离群点检测算法研究

朱跃忠

江西理工大学, 建筑与测绘学院, 江西 赣州  
Email: zxc2511836601@126.com

收稿日期: 2018年11月20日; 录用日期: 2019年1月1日; 发布日期: 2019年1月8日

---

## 摘要

针对传统空间离群检测算法在空间邻居构建上存在算法复杂度过高、人为影响较大等问题, 提出一种基

于Delaunay三角剖分的空间离群检测算法。该算法通过对空间数据点进行Delaunay三角剖分,建立空间邻域,根据生成的空间邻域,依据SLOF算法定义空间邻域内离群因子。结合某地区矿山具体钻孔数据对该算法进行实验,实验结果表明该算法能有效检测出空间离群点,算法复杂度较低,人为影响较低。

## 关键词

空间离群点, Delaunay三角剖分, SLOF离群因子

Copyright © 2019 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

随着地理空间信息技术和卫星遥感等数据采集器的越发先进,空间数据的数量和质量呈海量增长。如何有效从空间数据中发掘、检测出有效的信息和知识已经成为了当前研究的重要方向。将数据挖掘技术应用到空间数据中,即空间数据挖掘,是解决当前问题的重要手段。空间数据挖掘主要功能包括关联分析、聚类分析、组合分析、离群检测等方面。

其中空间数据的离群检测作为空间数据挖掘研究的重要前提和方法,是空间数据挖掘不可或缺的部分,目前已经广泛应用在地质灾害监测、成矿预测、环境监测、矿山水文检测、疾病控制和监测[1]等众多领域。

空间数据是用来描述现实世界中的事物的大小、位置、轮廓、属性等信息的数据,空间数据相比于普通数据具有的特性是空间自相关性和异质性。因此普通的离群点检测方法不适用于空间数据的离群点检测,需要充分考虑空间数据的特性,才能有效的检测出空间异常点。

最早的空间数据离群检测是由地学统计学发展而来,主要分为空间数据可视化和属性定标两种方法,其中空间数据可视化方法主要有变差云图[2], qqplot 图[3]等方法,而属性定量的方法主要包括 Moran 散点图[4]、Z-Value 法等方法[5]。

Shekhar 等人从空间数据的本质出发,首次提出了二分法(SLZ) [6],将空间数据分为空间属性和非空间属性,根据空间属性确定空间对象的空间邻域,通过计算非空间属性确定空间对象邻域内的差异,计算离群因子。但是 SLZ 算法使用的是全局的阈值,求出的是全局的空间离群点。

对于多维非空间属性的处理,文俊浩[7]等人提出使用马氏距离计算各个维度非空间属性的差异。但是马氏距离并不能有效的表示数据的自相关性。Sun 等人在 SLZ 的算法中加入波动参数提出了 SIOM 算法[8],国内薛安荣等人提出了普适的 SLOF 算法[9],并针对混合类型的非空间属性提出了相关算法。张忠平等人提出了基于 KNN 图[10]的空间离群检测算法,利用 KNN 图来构建空间邻域,KNN 图的时间复杂度较高,不适用于海量数据。综合上訴算法可以发现存在两个问题:在空间邻域的构建方面,算法大多时间复杂度过高,不适合处理空间高维数据,人为影响较大;在非空间属性的度量方面,对空间数据的研究不够深入,离群因子的定义存在误差。

本文针对上述算法在空间邻域的构建方面存在的问题,引入 Delaunay 三角剖分的思想,通过 Delaunay 三角剖分来构建空间数据的空间邻域,利用空间。并结合某矿山钻孔数据进行实验,最终依据实验结果,从算法的检测率、时间复杂度、算法鲁棒性等方面对该算法的性能进行分析。

## 2. 基于 Delaunay 三角剖分的空间离群检测算法

本文提出的基于 Delaunay 三角剖分的空间离群检测算法的具体思路如下：首先根据  $3\sigma$  准则，选取各个维度的全局离群点，再依据 Delaunay 三角剖分生成空间邻域，然后将数据预处理中选取的离群点的值用改点邻域内对象的均值代替，建立相应的空间权重矩阵，在确定的空间权重矩阵下，依据 SLOF 离群因子的计算方法，度量空间邻域内的差异，选取 SLOF 离群因子前  $n$  个大的值作为离群点输出。

具体的算法流程如图 1 所示：

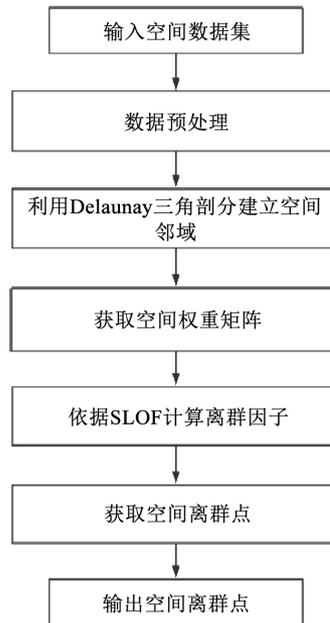


Figure 1. Spatial outlier detection algorithm based on Delaunay triangulation  
图 1. 基于 Delaunay 三角剖分的空间离群检测算法

### 2.1. 相关定义

存在空间对象集  $S = \{s_1, s_2, \dots, s_n\}$ ，由  $n$  个对象组成，其中  $\theta_p$  表示对象间在  $p$  条件下相连接，对于对象  $s \in S$ ， $k(s)$  表示空间属性， $f(s)$  表示非空间属性。

**定义 1:** (空间邻居) 当对象  $s_i$  和  $s_n$  符合条件  $\theta_p$ ，那么  $s_i$  就是  $s_n$  的空间邻居。

**定义 2:** (空间邻域) 空间对象  $s_i$  所有的空间邻居的集合称为  $s_i$  的空间邻域。

**定义 3:** ( $3\sigma$  准则) 计算数据集的分布误差，将数值分布在  $(\mu - 3\sigma, \mu + 3\sigma)$  的点设定为全局离群点。

**定义 4:** (Delaunay 三角剖分) 对空间数据进行三角网划分，其中当划分成的三角网满足以下两个特性，**a:** 空球准则，单个三角形或不规则的四面体的外接圆或外接球内部不包含除顶点外的其它点。**b:** 最大-最小角准则每个三角的角度都满足是所有可能三角形中最小的。

**定义 5:** (空间权重) 对象  $s_i$  和对象  $s_n$  是空间邻居，那么选取反距离公式  $W_{s_i, s_n} = 1/\text{dist}(s_i, s_n)$  为空间权重， $\text{dist}$  表示对象间的距离，当对象  $s_i$  和  $s_n$  不满足空间邻居关系时，那么  $W$  的值为 0。

**定义 6:** (邻居加权距离)  $\text{dist}(s_i, s_j, w)$  表示对象与对象间的邻域距离如公式 2 所示， $w_k$  表示是第  $k$  维度的权值， $f(s_{ik})$  表示  $s_i$  对象的在  $k$  维的权重。

$$\text{dist}(s_i, s_j, w) = \sqrt{\sum_{k=1}^d w_k (f(s_{ik}) - f(s_{jk}))^2} \quad (\text{公式 2.1})$$

**定义 7:** (邻域距离)  $Ndist(s, w)$  表示表示对象所有空间邻域内的邻域距离, 其中  $NB(s)$  表示邻域内所有对象的数量。

$$Ndist(s, w) = \frac{\sum_{p \in NB(s)} dist(p, s, w)}{|NB(s)|} \quad (公式 2.2)$$

**定义 8:** (离群因子)存在数学公式能够计量对象在空间邻域内的非空间属性的差异, 该数学公式可以认为是离群因子。

**定义 9:** (SLOF 离群因子) SLOF 的离群因子定义公式如下, 选择 SLOF 值最大的前  $n$  个点作为空间离群点。

$$SLOF(s) = \frac{Ndist(s, w)}{\sum_{p \in NB(o)} \frac{Ndist(s, w)}{|NB(o)|}} \quad (公式 2.3)$$

## 2.2. 数据预处理

将获取的数据集, 进行数据格式转换, 整合操作, 空间异常点会影响到局部离群因子的度量, 根据定义 3, 鉴别出空间数据集中的全局离群点, 并将其保存。

## 2.3. 通过 Delaunay 三角剖分构建空间邻域

传统的空间邻域构建方法存在参数选择困难, 构建时间复杂度过高, 人为因素较大等问题。而基于 Delaunay 的三角剖分算法, 常见的实现方法有逐点法, 三角网增长法等, 这些算法的时间复杂度最优能达到  $O(n \log n)$ , Delaunay 三角剖分本身具有严格的数据验证和坚实的理论基础。具有能有效维护空间数据的拓扑信息, 便于三维分析和显示等良好的几何特性。

## 2.4. 建立空间权重矩阵

根据生成的空间邻域, 建立相应的空间邻域矩阵, 表 1 表示生成的相应的空间对象邻域表。

**Table 1.** Generating spatial neighborhood table based on Delaunay triangulation

**表 1.** 基于 Delaunay 三角剖分生成空间邻域表

对象 ID	X 坐标	Y 坐标	Z 坐标	空间邻居 ID	数量
1	411,708.5	2,774,154	323.212	2, 8, 21, 33, 45, 55, 63, 78, 99	9
2	411,708.5	2,774,154	324.212	1, 23, 35, 66, 79, 88, 94, 123, 135, 156	10
3	411,708.5	2,774,154	325.212	5, 15, 23, 45, 56, 88, 99, 122, 167, 182	11
4	411,708.5	2,774,154	326.212	6, 12, 23, 31, 55, 67, 88, 102	8
...	...	...	...	...	...
158	411,723.396	2,774,135.456	331.655	159, 164, 168, 183, 201, 222, 238, 240	8
159	411,723.396	2,774,135.456	332.655	158, 162, 183, 203, 205, 221, 238	7
...	...	...	...	...	...

结合表 1 和定义 5 建立空间权重矩阵。示例给定简单空间数据集, 生成 Delaunay 剖分四面体如图 2(a) 所示,  $p_1 \sim p_5$  表示空间对象, 图 2(b) 为生成的相应空间权重矩阵。

## 2.5. 空间邻域离群因子的度量

依据定义 8 中的 SLOF 算法的离群因子的计算复方式, 结合上文生成的空间邻域矩阵, 生成空间离

群因子，其中空间权重选取反距离权重公式进行计算，为了防止 SLOF 的值为 0，将公式 2.3 改为公式 2.4 所示：

$$\text{SLOF}(s) = \frac{N\text{dist}(s, w) + \delta}{\frac{\sum_{p \in \text{NB}(s)} N\text{dist}(s, w)}{|\text{NB}(s)|} + \delta} \quad (\text{公式 2.4})$$

$\delta$  一般取值为  $1 * 10^{-6}$  左右，当 SLOF 的值大于 1 时，离群程度开始变高，当小于 1 时，表示离群程度正常。最终选取 SLOF 值前  $n$  个大的点作为空间离群点。

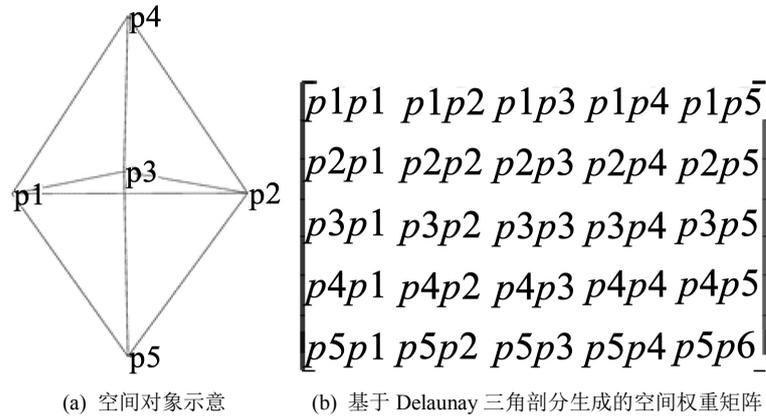


Figure 2. Schematic diagram of spatial weight matrix  
图 2. 空间权重矩阵示意图

## 2.6. 算法的描述与复杂度分析

基于 Delaunay 三角剖分的空间离群检测算法(数据包含  $N$  个对象)

**输入：** 空间数据集

**输出：** 空间离群点

**Step 1:** 对空间数据集进行预处理，甄选出全局离群点，需要求取数据集合的均值方差等，时间复杂度为  $O(N)$ 。

**Step 2:** 建立空间邻域，通过 Delaunay 三角剖分对空间数据集进行处理，建立空间邻域，时间复杂度为  $O(N \log N)$

**Step 3:** 计算空间权重矩阵，利用 Delaunay 三角剖分生成的空间邻域，将全局离群点灯值用邻域均值代替，最后建立空间权重矩阵，时间复杂度为  $O(N \log N)$ 。

**Step 4:** 计算空间邻域内的空间离群因子，结合 SLOF 算法，根据生成的空间邻域矩阵生成空间离群因子，时间复杂度为  $O(N \log N)$ 。

综上所述本文时间复杂度为  $O(N \log N)$ 。

## 3. 实验和分析

本文算法结合某地区稀土矿区具体钻孔数据，使用  $k$  邻域图法和本文所提算法做具体实验对比。实验环境为 win10 操作系统，采用 MATLAB2017b 和 C 语言编程。

### 3.1. 算法实验

结合某地矿区的钻孔的数据集对算法进行分析，其中 X, Y, Z 表示空间坐标，非空间属性为各个采

样点的品味，共有 240 个值。

通过  $3\sigma$  准则没有发现全局离群点，使用  $k$  邻域图法和 Delaunay 三角剖分建立空间邻域，根据生成的空间邻域结合 SLOF 离群因子定义，计算相应的空间离群点。

其中图 3(a)表示矿山某钻孔具体采样散点图如图 3(b)根据 Delaunay 生成的空间邻域算法的时间复杂度对比如图 4 所示。

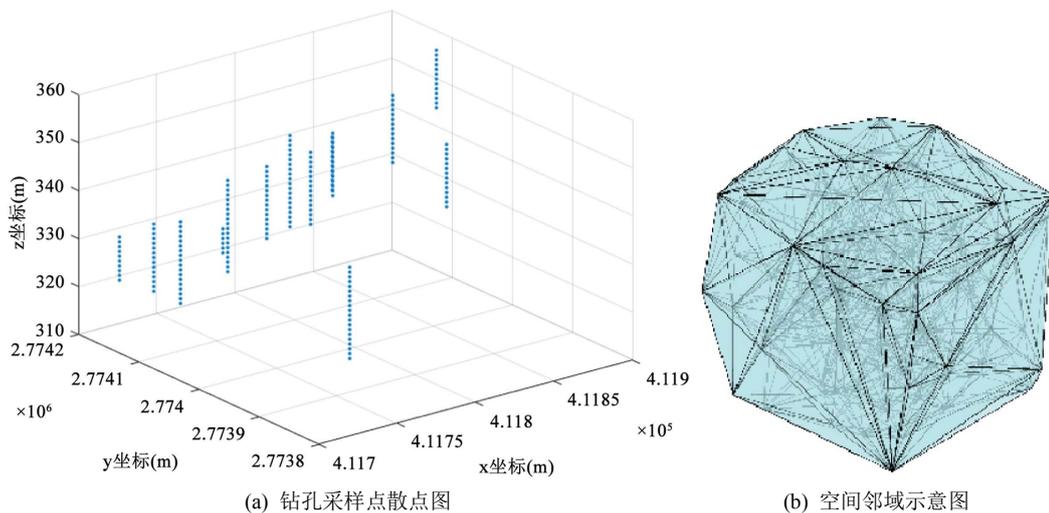


Figure 3. Scatter diagram and spatial neighborhood diagram of borehole sampling points

图 3. 钻孔采样点散点图和空间邻域示意图

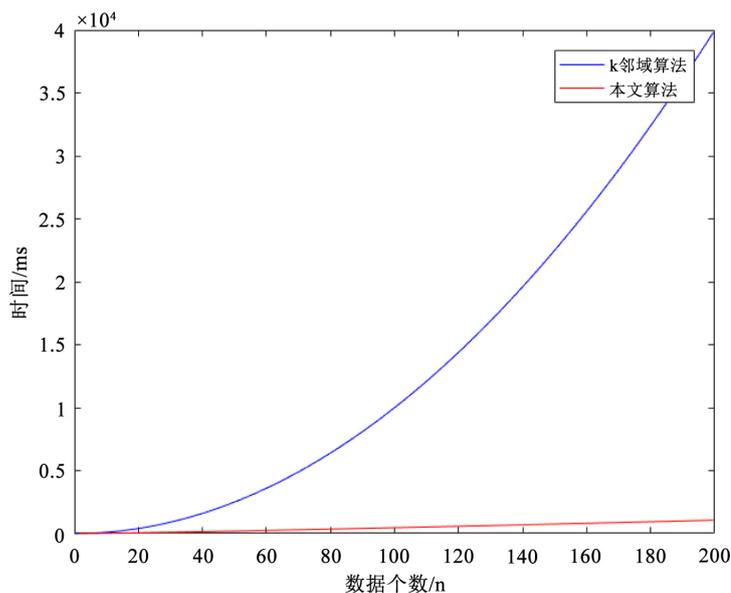


Figure 4. Comparison of time complexity between the two algorithms

图 4. 两种算法时间复杂度对比

### 3.2. 实验结果分析

表 2 为 Delaunay 三角剖分检测出的离群点，表 3 为  $K$  邻域法检测出的离群点。

1) 从检测精度对两种方法进行分析：结合现场采集数据具体情况，可以发现基于 Delaunay 三角剖分

方法检测出的离群点，其检测精度达到 80%，而  $k$  邻域的方法，其检测精度仅在  $k$  值为 5 时，最高达到 80%，因此可以得出本文所提算法的检测精度高于  $k$  邻域方法。

**Table 2.** Outliers detected from Delaunay triangulation  
**表 2.** 基于 Delaunay 三角剖分检测出的离群点

序号	离群因子	离群点 ID
1	-3.2	30
2	-2.5	61
3	-2.2	93
4	-1.7	231
5	-1.5	156

**Table 3.** SLOF algorithm detects different  $k$  values  
**表 3.** SLOF 算法不同  $k$  值检测结果

序号	$K$ 值	离群点 ID
1	3	306,193,231,156
2	5	306,093,231,155
3	7	307,593,231,156
4	9	293,290,159,210
5	15	304,969,231,156

2) 从人为因素进行分析： $k$  邻域的算法，在  $k$  值选取方面存在人为因素，最优  $k$  值难以选择。而本文算法，完全基于 Delaunay 三角剖分建立空间邻域，无需对  $k$  值进行设定和检测。

3) 从算法时间复杂度进行分析：如图 4 所示，当数据量增长较快时，本文算法时间复杂度增长较为缓慢，而  $k$  邻域算法时间复杂度增长呈指数增长，因此可得本文算法在时间复杂度方面优于  $k$  邻域算法。

## 4. 结束语

传统的基于 KNN 邻域图的空间离群检测算法，在空间邻域构建方面，存在算法时间复杂度过高，最优  $k$  值受人为因素影响大等不足，而 Delaunay 三角剖分对空间数据在空间邻域构建方面具有优势，算法本身也较为成熟，背后也具有深厚的数学理论基础。因此本文引入 Delaunay 三角剖分的方法，构建空间数据的空间邻域，并结合 SLOF 算法中对空间邻域离群因子的计算方法，求出空间数据的离群点。但是空间数据往往不是单个点存在，可能是多个点、线、面组成的，空间数据的非空间属性大多也是多维的。为此下一步将研究多维属性的复杂空间数据的空间离群点检测算法。

## 参考文献

- [1] Chen, D., Lu, C.T., Kou, Y., *et al.* (2008) On Detecting Spatial Outliers. *Geoinformatica*, **12**, 455-475. <https://doi.org/10.1007/s10707-007-0038-8>
- [2] Wackernagel, H. (2003) Variogram Cloud. *Multivariate Geostatistics*. [https://doi.org/10.1007/978-3-662-05294-5\\_6](https://doi.org/10.1007/978-3-662-05294-5_6)
- [3] Das, B.R.S.I. and Plots, Q.Q. (2008) Random Sets and Data from a Heavy Tailed Distribution. *Communications in Statistics Stochastic Models*, **24**, 103-132. <https://doi.org/10.1080/15326340701828308>
- [4] Anselin, L. (2003) Spatial Externalities, Spatial Multipliers, and Spatial Econometrics. *International Regional Science Review*, **26**, 153-166. <https://doi.org/10.1177/0160017602250972>
- [5] Shekhar, S., Lu, C. and Zhang, P. (2003) A Unified Approach to Detecting Spatial Outliers. *Geoinformatica*, **7**, 139-166. <https://doi.org/10.1023/A:1023455925009>

- 
- [6] Shekhar, S., Lu, C. and Zhang, P. (2003) A Unified Approach to Detecting Spatial Outliers. *Geoinformatica*, 7, 139-166. <https://doi.org/10.1023/A:1023455925009>
- [7] 文俊浩, 吴中福, 吴红艳. 空间孤立点检测[J]. 计算机科学, 2006(5): 186-187 + 210.
- [8] Chawla, S. and Sun, P. (2006) SLOM: A New Measure for Local Spatial Outliers. Springer-Verlag, New York.
- [9] 薛安荣, 鞠时光, 何伟光, 等. 局部离群点挖掘算法研究[J]. 计算机学报, 2007, 30(8): 1455-1463.
- [10] 张忠平, 徐晓云, 王培. 基于 KNN 图的空间离群点挖掘算法[J]. 计算机工程, 2011, 37(4): 37-39.

**知网检索的两种方式:**

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>  
期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)