Spatial outliers detection considering distances among their neighbors

LI Guang-qiang[1], DENG Min[1,2], ZHU Jian-jun[1], CHENG Tao[3], LIU Qi-liang[1]

1. *School of Info-physics and Geomatics Engineering, Central South University, Hunan Changsha* 410083, *China*;
2. *Geomatics and Applications Laboratory, Liaoning Technical University, Liaoning Fuxin* 123000, *China*;
3. *Department of Geomatic Engineering, University College London U, K*

Abstract: Detection of spatial outliers has been one of the hot issues in the field of spatial data mining and knowledge discovery. So far, the detection of spatial outliers is determined by spatial outlier factor in most of the existing methods, while geometrical distances among their corresponding spatial neighbor are ignored. In this case, these existing methods are inappropriative to the spatial inhomogeneous distribution. To overcome this limitation, this paper presents a new method for spatial outlier detection, named as spatial outlier measure method (SOM for short). At first, some concepts related to the SOM are defined, such as the attribute gradient, the inverse distance weight and the degree of spatial outlier. The algorithm of the SOM is further presented. One can easily find that the new method considers the distances among the neighborhood and their effects on the attribute values of the target entities, and the degree of spatial outlier is used to check spatial outliers. Finally, a practical example is employed to demonstrate the validity of the method proposed in the paper, where the Cr concentration data of soil in a southern city of China are utilized.

Key words: spatial outlier, spatial nearest neighbor, spatial outlier measure, inverse distance interpolation

CLC number: TP751.1 **Document code**: A

1 INTRODUCTION

Spatial outlier pattern is more valuable than the universal one, which implicates much unexpected information and stands for the specificity of the geographical phenomena or the geographical processes. Spatial outlier plays significant role in many applications fields, such as geological disaster monitoring, mineralization forecasting, geological movement research, mines hydrological monitoring, earth geophysical and geochemical data analysis, public safety and health, remote sensing image data processing and others. By definition, the spatial outlier detection is to find a small part of data which deviate from universal pattern in the spatial database. Hawkins defines spatial outlier (also called isolation) as "which deviate so much from other observations so as to arouse suspicions that they were generated by a different mechanism" (Hawkins, 1980). After that, many scholars have proposed various definitions of outlier and methods for the outlier pattern detection. These research results may be divided into six categories, the distribution-based, the depth-based, the distance-based, the density-based, the cluster-based and the super-map-based (Hautamaki *et al.*, 2004; Cheng & Li, 2006; Wei *et al.*,2002). Since the traditional outlier detection methods don't distinguish spatial dimension and non-spatial dimension, and ignore the spatial characteristics of observations, they are not suitable for the spatial outliers detection. Shekhar firstly definite the spatial outlier as "a spatially referenced object whose non-spatial attribute values are significantly different from those of other spatially references objects in its spatial neighborhood." Informally, a spatial outlier is a local instability, whose non-spatial attributes are extreme relative to its neighbors, while they may not be significantly different from the entire trend (Shekhar *et al.*, 2001, 2003). At present, spatial outlier detection methods includes, in general, qualitative and quantitative types. The qualitative method is a graphic-based (such as Variable clouds and plot) visualization approach, which discovers outlier data in the graph. Indeed, this method has been rarely used due to many disadvantages (Huang *et al.*, 2006). The quantitative methods can be classed into the distance-based method, the density-based method, the convex-shell-depth-based method, etc (Shekhar *et al.*, 2001). These detection methods only focus on the local difference of non-spatial

Received date: 2007-09-26; **Accepted date**: 2007-11-20

Foundation: Supported by the Major State Basic Research Development Program of China (973 Program), No. 2006CB701305; the Scientific Research Foundation of Jiangsu Key Laboratory of Resources and Environmental Information Engineering (China University of Mining and Technology) (Grant No. 20080101) and Open Research Fund Program of the Geomatics and Applications Laboratory, Liaoning Technical University, Grant No. 2007001.

First author biography: LI Guang-qiang (1972—), male, lecturer, received the BSc. and MSc. degree in China University of Geosciences (Wuhan). He is currently the candidate of PhD thesis in Central South University. His research interests include spatial data mining and spatio-temporal outliers detection.

attribute values based on the spatial neighbors, while the influence of the distances among the neighbors is ignored. Therefore, this paper will propose a spatial outliers measure method (SOM for short), which takes the distances among the spatial neighbors into account.

2 RELATED WORK AND OUR STRATEGY IN THIS PAPER

2.1 Related Work

Shekhar *et al.*, (2001, 2003) employ the statistics aggregation function to detect the outlier points in the spatial neighbors, whose main idea is to utilize function $S(X) = f(X) - f^{N}_{aggr}(X)$ to define the spatial outliers, where $f(X)$ is the non-spatial attribute value of the X, $N(X)$ is the spatial neighbors of X, $f^{N}_{aggr}(X) = E_{y \in N(X)}(f(Y))$, and usually, E is the average function. That is, spatial outlier is a spatial entity whose non-spatial attribute values are significantly different from the non-spatial values average of the spatial neighbors. After computing the non-spatial attribute values average $g(X)$ of the spatial neighbors, the formula $h_i = h(X_i) = f(X_i) - g(X_i)$ is used to generate a set $\{h_1, h_2, \cdots, h_n\}$. In the set $\{Y_1, Y_2, \cdots, Y_n\}$ generated by $Y_i = |(h_i - \mu)/\sigma|$, the top m entities form the spatial outliers set (Lu *et al.*, 2003). A spatial local outlier measurement (SLOM) is presented by Chawla & Sun (2006), basing on the distances of the spatial entities attribute to their spatial neighbors. In Huang & Qin's research (2004), the local density is computed according to the attribute distances in the spatial neighbors, and the ratios of the entity's local density to its near neighbors' densities are computed, too. Then, the ratios' average value of spatial neighbors of each entity is computed. The spatial outliers are detected according to the average values. Similarly, Ma & He (2006), Zhou *et al.*, (2003) and Huang *et al.* (2006) also utilize the non-spatial attribute distances to detect spatial outliers. In general, the existing spatial outlier detections procession can be divided into 3 steps: (1) to employ the entities' coordinates to define the spatial neighbors, (2) to compute the non-spatial attribute deviations factors of the entity from their spatial neighbors according to the non-spatial attribute distances, (3) to identify the spatial outliers depending on the non-spatial attribute deviation factors. Spatial neighbors are usually constructed depending on the spatial relationships (such as distance and joint), where the most general method is the k-Nearest Neighborhood (Hautamaki *et al.*, 2004).

To identify the spatial outliers, all existing methods commonty consider that the entity is equally influenced by its neighbors, while the entity's distances to its spatiat neighbors are negtected. Thus, when the spatial entities distribution is unevenly, the distances of the entity to its spatial neighbors are prominently varied, and the neighbors' effects to it are very different. For instance, in Fig. 1, the entity A's distances to its spatial neighbors are far less than B's. Clearly, the impacts among the entities around A are far more than those around B. Since the spatial autocorrelation and continuity of the non-spatial attributes are the inherent properties of the spatial entity, "the first law of geography" (Tobler, 1970) shows that the near entities attribute difference is less than distant entities. In Fig. 1, on the assumption that a pollution source (denoted by ⊕) locates in the south of B, the impact in the region, which is a function of the inverse distance, only relates to the distance from the pollution source. I. e. the pollution changes equably. Obviously, the monitoring result around B is significantly higher than that around A. If employing the existing detection methods to identify the spatial outliers, there are a little difference among the monitoring results around A since A is far away from the pollution source, and the A's distances to its neighbors are nearly equal. Thus, it's difficult to identify the spatial outlier around A. However, around B, because the B's distances to its neighbors are quite different, the monitoring results are much deviated from its neighbors. Then, B is likely to be regarded as a spatial outlier. Obviously, the neglect of the distances within the neighbors may lead to some drawbacks. That is, when we compute the spatial outlier factors based on the non-spatial attributes of the entities in the neighbors, the distances among the neighbors can not be ignored.

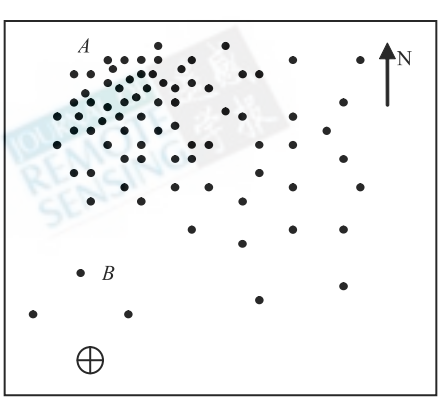

Fig. 1 Illustration of inhomogeneous distribution of spatial entities

2.2 Our strategy in this paper

For a spatial entity O, let $N(O) = \{X_1, X_2, \cdots, X_n\}$, where $N(O)$ is O's spatial neighbors, and $f(X_i)$ is the non-spatial attribute value. In order to identify whether O is a spatial outlier, the inverse distance weighted (IDW for short) interpolation formula is employed to compute the attribute value of O, noted as $f^{C}(O)$, which can be represented as follows:

$$f^C(O) = \frac{\sum_{i=1}^{n} \frac{f^R(X_i)}{D_s(O,X_i)}}{\sum_{i=1}^{n} \frac{1}{D_s(O,X_i)}} \quad (1)$$

where n is the number of entities in O's neighbors, $f^R(X_i)$ is the observed value of X_i, $D_s(O,X_i)$ is the distance between O and X_i (Wang, 2006). Because IDW interpolation depends on the inverse distances among the spatial entities, the formula (1) takes the interaction weight among the spatial entities into account so that the impact from the near entity is stronger than that from the distant entity. Then, $|f^R(O) - f^C(O)| > \varepsilon$ (ε is a given threshold) shows that O is significantly different from it's neighbors, and O is a spatial outlier. In other words, it illuminates that the O's attribute value is impacted not only by its neighbors but also by other factors. So in this paper, $|f^R(O) - f^C(O)|$ can be taken for a spatial outlier measure, which is a non-spatial attribute deviation factor considering the entity's distances to its neighbors. The existing detection methods' flaws for spatial inhomogeneous distribution are solved by this method.

3 A SPATIAL OUTLIER MEASURE CONSIDERING THE DISTANCES AMONG THE NEIGHBORS

To compute the degree of spatial outlier, some concepts, such as the non-spatial attribute gradient, the inverse distance weight factor and the degree of spatial outlier, are defined strictly.

Definition 1: Non-spatial attribute gradient is the ratio of the absolute margin value, which is between the spatial entity O and a certain entity X_i in the O's neighbor, to the distance between them. It is described as:

$$G(O,X_i) = \frac{|f^R(O) - f^R(X_i)|}{D_s(O,X_i)} \qquad \forall X_i \in N(O) \quad (2)$$

Definition 2: The inverse distance weight factor is the inverse sum of the inverse spatial entity O's distances to all other entities in its neighbor. It is described as:

$$\delta(O) = \frac{1}{\sum_{i=1}^{n} \frac{1}{D_s(O,X_i)}} \qquad \forall X_i \in N(O) \quad (3)$$

Definition 3: Spatial Outlier Measure (SOM for short) is the product of the attribute gradient and the inverse distance weight factor, which is described as:

$$\mathrm{SOM}(O) = \sum_{i=1}^{n} G(O,X_i) \times \delta(O) \qquad \forall X_i \in N(O) \quad (4)$$

Definition 4: Spatial outlier is a entity whose SOM exceeds the threshold ε. Then, $S_{\mathrm{Outliers}} = \{\forall X: \mathrm{SOM}(X) > \varepsilon\}$, where S_{Outliers} is the spatial outliers set.

In practice, since the value of ε is difficult to be established, in general, the methods to generate the outliers set include (1) given the number (α) of the outliers, the top α entities in the SOM set sorted by descending order compose the outliers set (Huang *et al.*, 2006; Lu *et al.*, 2003; Chawla & Sun, 2006; Huang & Qin, 2004; Ma & He, 2006), (2) given the threshold value (ε) of the SOM, all entities whose SOMs are greater than ε form the outliers set (Zhou *et al.*, 2003). The former is utilized in this paper. Let the spatial entities set SD = $\{X_1, X_2, X_3, \cdots, X_n\}$. Each entity's SOM value is computed. Let $S_{\mathrm{SOM}} = \{\mathrm{SOM}(X_1), \mathrm{SOM}(X_2), \cdots, \mathrm{SOM}(X_n)\}$. If $i_1, i_2, \cdots, i_\alpha$ are code of the top α entities in the descending order SOM set, the spatial outliers set $O_{\mathrm{outliers}} = \{X_{i1}, X_{i2}, X_{i3}, \cdots, X_{i\alpha}\}$.

Since the non-spatial attribute value gradients take the spatial entity's interaction from its neighbors into account, SOM based on the non-spatial attribute gradient considers the spatial entity's distances from its neighbors. This method reflects the non-spatial attribute values' spatial autocorrelation and continuity.

4 DESIGN AND ANALYSIS OF THE SOM ALGORITHM

4.1 Algorithm of SOM dsign

The algorithm of SOM can be divided into five steps:

(1) to generate the spatial neighbors of each spatial entity;

(2) to compute the non-spatial attribute gradient between the entity and its neighbors;

(3) to compute the inverse distance weight factor of each spatial entity;

(4) to compute the SOM value of each entity, and

(5) to sort the SOMs set by descending order and picking out the top α entities from the set to compose spatial outliers set.

While generating the spatial neighbors, the algorithm-complexities of the fixed-number method or fixed-distance method are very high, and the extrapolation impacts greatly on the interpolation results, etc (Du & Xiao, 2005). Since the Delaunay triangulated irregular network (D-TIN for short) is employed to construct the spatial neighbors easily (Du & Xiao, 2005; Adam *et al.*, 2004), this paper utilizes the nodes inter-connectivity in the D-TIN to form the spatial neighbors. Namely, the spatial neighbor of the node contains all the nodes linking directly to it. Besides, the two methods, which include the given number of outlier and the given threshold of the SOM value, can be utilized to determine the number α of the spatial outliers. This paper chooses the former method to find out the spatial outliers. The SOM algorithm is described in detail as follows:

Input: Spatial entities set SD , the number α of spatial outliers

Output: Spatial outliers set S_{Outliers}

```
Sub Main( )
{
  CreateTIN( SD) ;
//Employs Mirko algorithm ( Mirko & Borut, 2005 ) to build D-TIN for spatial entities
  CreateNeighbor( ) ;//Creates spatial neighbors of each node
S_SOM = φ;
for each O ∈ SD
{
G( O) = 0;
//Initializes the cumulative variable of the gradient
δ( O) = 0;
//Initializes the cumulative variable of the inverse distance weight factor
for each X ∈ N( O)
{
G( O,X) = |f^R( O) - f^R( X) |/D_s( O,X) ;
//Computes the attribute value gradient
G( O) = G( O) + G( O,X) ;
//Cumulates the attribute value gradient
δ( O) = δ( O) + 1/D_s( O,X) ;
//Cumulates the inverse distance
}
δ( O) = 1/δ( O) ;
//Computes the IDW factor
SOM( O) = G( O) × δ( O) ;
//Computes the SOM value
S_SOM = S_SOM ∪ { SOM( O) } ;
//Add the SOM of O to S_SOM
}
S'[ ] = SortSOM( S_SOM) ;
//Sorts the S_SOM by descenting order and saves to the array S'
S_Outliers = { ∀ X: SOM( X) = S'[ k] ∧ 1 ≤ k ≤ α}
//Fetches the top α SOM from the array S' to compose S_Outliers
Return S_Outliers ;
}
//The procedure of the spatial neighbor Creation
Sub CreateNeighbor( )
{
for each edge in Edges
//Edges is a edges set of TIN
{
//Adds an endpoint of an edge to another endpoint's neighbor
AddNeighbor( N( edge. point1) , edge. point2) ;
AddNeighbor( N( edge. point2) , edge. point1) ;
D_s( edge. point1, edge. point2) = edge. Length;
//Saves the distance of a edge to the distance list
}
}
```

4.2 Analysis of the complexity of the SOM algorithm

In general, there are four factors needing to be considered for the complexity of the algorithm, including:

1) The complexity of generating the D-TIN. This paper employs Mirko's algorithm to generate the D-TIN, which runs in $O(n)$ time (Mirko & Borut, 2005)

2) The complexity of forming the spatial neighbors. If n denotes the number of nodes, the maximum of the edges of the D-TIN is $(3n\text{-}6)$ (He *et al.*, 2006). This step runs in $O(3n)$ time.

3) The complexity of computing the SOM values. In the D-TIN, since the maximum of the edges is $(3n-6)$ and each node has 6 directly connective nodes on average, this step runs in $O(6n)$ time.

4) The complexity of sorting the SOM values is $O(n)$ (Zhou, 2006).

Summarizing the complexity of the four above steps, the total complexity of SOM algorithm is $O(11n)$.

5 EXAMPLE

The 103 observed data of the soil Cr concentration in a Chinese Southern city is employed to prove the feasibility and validity of the algorithm proposed in this paper. Fig. 2 shows the D-TIN, which has 295 edges, built by the 103 soil monitoring points. On average, each node has 5.7 directly connective nodes. Table 1 enumerates some observation points and their spatial neighbors. The points' distances and the concentration margin absolute values to their spatial neighbors are listed, too. In Table 2, the SOM values of each point are listed in descending order.

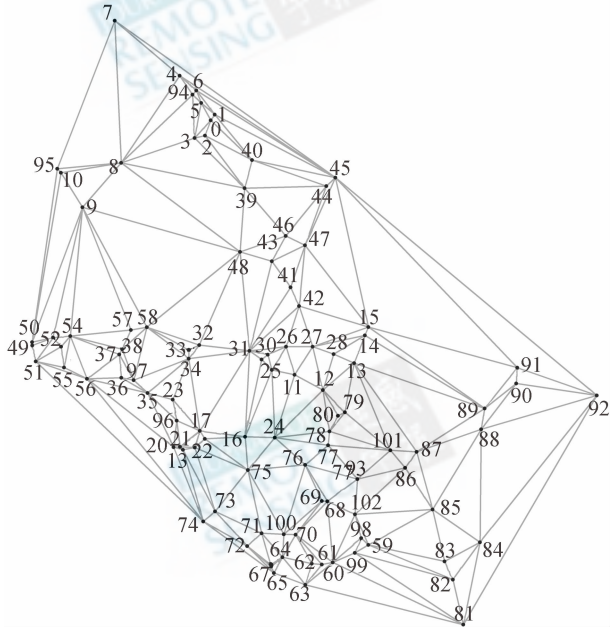

Fig. 2 D-TIN of observation points of soil

Table 1 Spatial neighbours of observation points and the distances among them

Observed point ID	Near point ID	Distance/m	Cr concentration margin absolute values	Observed point ID	Near point ID	Distance/m	Cr concentration margin absolute values
0	1	1590.32	33.9	2	40	12071.23	7.53
0	40	13416.78	43.68	2	1	3722.83	17.31
0	94	4081.17	52.69	2	3	2454.77	25.84
0	5	6943.48	55.74	2	39	15003.59	47.33
0	45	31061.54	40.43	3	39	16173.23	21.49
1	40	13084.92	9.78	3	6	9926.11	33.03
1	94	4524.85	18.79	3	2	2454.77	25.84
1	0	1590.32	33.9	3	1	5453.38	8.53
1	3	5453.38	8.53	3	94	8144.36	27.32
1	2	3722.83	17.31	…	…	…	…

Table 2 Descending order of SOM

Observed point ID	SOM	Observed point ID	SOM
96	60.67	34	32.88
70	59.43	100	30.89
21	52.66	47	29.79
69	47.16	30	29.73
39	45.12	20	29.7
33	43.21	19	29.6
0	41.63	36	29.04
68	39.67	…	…

If we select five spatial outliers from Table 2, the spatial outliers set $S_{outliers} = \{96, 70, 21, 69, 39\}$. The locations of the outliers are showed in Fig. 3.

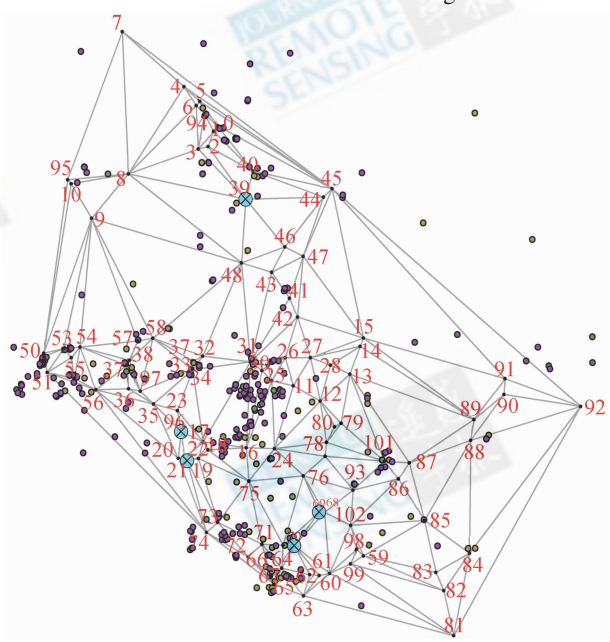

Fig. 3 Location of Outliers (⊗—outliers, ◆—pollution source)

In order to analyze the cause of the outliers, outliers and their neighbors' land used types and elevators are enumerated in Table 3. Clearly, the prominent differences of land used types between the 96, 70, 21, 39 and their spatial neighbors are the causes of outliers. In somewhere, Mountains obstruct some nodes. Though the land used types of the 69 is consistent with its nearest neighbors, its elevator is quite different from its neighbors, and mountains obstruct them. Fig. 3 demonstrates little relation between the pollution sources and the outliers' distribution. Through the above analysis, the main causes of the spatial Cr concentration outliers include (1) the differences of the land used types; (2) topographical differences, such as elevator or mountain barriers.

Table 3 Outliers and their nearest neighbors' soil types and elevators

Outlier points			Neighbors		
ID	Land used type	Elevator/m	ID	Land used type	Elevator/m
96	Vegetable field	47.94	23	Paddy field	48.36
			17	Vegetable field	264.12
			20	Paddy field	133.00
			19	Corn field	96.54
			35	Litchi Field	15.21
70	Vegetable field	36.47	68	Vegetable field	89.74
			69	Paddy field	132.33
			100	Vegetable field	25.11
			60	Litchi Field	28.33
			62	Litchi Field	36.11
			61	Vegetable field	62.00
			64	Paddy field	41.85
21	Vegetable field	83.52	19	Corn field	96.54
			74	Vegetable field	100.25
			20	Paddy field	133.00
			22	Paddy field	124.41
			73	Paddy field	100.51
69	Vegetable field	132.33	100	Vegetable field	25.11
			68	Vegetable field	89.74
			70	Vegetable field	36.47
			76	Vegetable field	20.36
39	Vegetable field	103.28	40	Paddy field	45.01
			2	Vegetable field	153.78
			3	Paddy field	137.65
			48	Paddy field	13.65
			46	Vegetable field	17.22
			44	Paddy field	36.32
			8	Paddy field	75.64

6 CONCLUSIONS AND FURTHER WORK

The spatial outlier detection is useful to discover the underlying laws about the geographical phenomenon

development, which is one of hot issues in the field of spatial data mining and knowledge discovery. Most existing methods find out the spatial outliers, which are the top n non-spatial attribute deviation factor, from their spatial neighbors without considering the entities' distances to their spatial neighbors. Especially, in the case of spatial inhomogeneous distribution, where the non-spatial attribute values change continuously and are of spatial autocorrelation, the distances need to be considered while we compute the non-spatial attribute values deviation factors. To solve this problem, the SOM method proposed in this paper considers the entity's distances to its spatial neighbors. This method employs Delaunay TIN to generate the spatial neighbors. The SOM algorithm is described in detail, which runs in $O(11n)$ time, and is less than the SLOM (Chawla & Sun, 2006). However, the SOM method has two limitations: (1) when the entity's distance to its spatial neighbors is close to zero, the SOM value may be exceptional, and (2) the SOM error of the points locating on the TIN boundary may be high. Thus, the further research work is to solve these two problems.

REFERENCES

Adam N R, Janeja V P and Atluri V. 2004. Neighborhood based detection of anomalies in high dimensional spatio-temporal sensor datasets. *Proceedings of the* 2004 *ACM symposium on Applied computing table of content*: 576—583

Chawla S and Sun P. 2006. SLOM: A new measure for local spatial outliers. *Knowledge and Information Systems*, **9**(4): 412—429

Cheng T and Li Z L. 2006. A multiscale approach to spatio-temporal outlier detection. *Transactions in GIS*, **10**(2): 253—263

Du Y J and Xiao D Y. 2005. A moving neighborhood kriging algorithm using delaunay-fixed distance neighborhood selection. *Journal of Engineering Graphics*, **2**: 64—68

Hautamaki V, Karkkainen I and Franti P. 2004. Outlier detection using K-Nearest neighbor graph. *Proceedings of the* 17*th International Conference on Pattern Recognition*, ICPR (3): 430 — 433

Hawkins D. 1980. Identification of outliers. London: Chapman and Hall

He J, Dai H, Xie Y Q, *et al.* 2006. Fast improved delaunay triangulation algorithm. *Journal of System Simulation*, **18**(11): 3055—3057

Huang T Q and Qin X L. 2004. Detecting outliers in spatial database. *Proceedings of the Third International Conference on Image and Graphics* (*ICIG'*04), IEEE 2004

Huang T Q, Qin X L and Wang Q M. 2006. Spatial outlier model and detection algorithm with leapingly sampling. *Journal of Image and Graphics*, **11**(9): 1230—1236

Huang T Q, Qin X L and Wang Q M. 2006. New approach of spatial outliers measurement and detection in spatial databases. *Journal of Image and Graphics*, **11**(7): 982—989

Lu C T, Chen D C and Kou Y F. 2003. Algorithms for spatial outlier detection. *Proceedings of the* 3*rd IEEE International Conference on Data Mining* (ICDM'03), Melbourne, USA, 597—600

Ma R H and He Z Y. 2006. Fast mining of spatial outliers from GIS database. *Geomatics and Information Science of Wuhan University*, **31**(8): 679—682

Mirko Z and Borut Z. 2005. An almost distribution-independent incremental delaunay triangulation algorithm. *The Visual Computer*, **21**(6): 384—396

Shekhar S, Lu C T and Zhang P S. 2003. A unified approach to detecting spatial outliers. *GeoInformatica*, **7**(2): 139—166

Shekhar S, Lu C T and Zhang P S. 2001. Detecting graph-based spatial outliers: algorithms and applications. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California

Tobler W. 1970. A computer movie simulating urban growth in the detroit region. *Economic Geography*, **46**(2): 234—240

Wang J F. 2006. Spatial Analysis. Beijing: Sciences Press

Wei L, Gong X Q, Qian W N, *et al.* 2002. Finding outliers in high-dimensional space. *Journal of Software*, **13**(2): 280—290

Zhou L K, Wang L Z and He J. 2003. Research on the algorithm of detecting graph-based spatial outliers. *Journal of Yunnan University*, **25**(5):398—400

Zhou J Q. 2006. Super Quick Sort Algorithm. *Computer engineering and applications*, **42**(29): 41—42

一种顾及邻近域内实体间距离的空间异常检测新方法

李光强[1],邓 敏[1,2],朱建军[1],程 涛[3], 刘启亮[1]

1. 中南大学 信息物理工程学院,湖南 长沙 410083;
2. 辽宁工程技术大学 地理空间信息技术与应用实验室,辽宁 阜新 123000;
3. 英国伦敦大学 地理信息工程系 伦敦

摘 要: 空间异常检测已成为空间数据挖掘和知识发现的一个重要研究内容.空间异常蕴含着许多意想不到的知识,现有的空间异常检测方法大多依据空间邻近域的非空间属性差异来计算偏离因子,忽略了邻近域内空间实体间距离的影响。本文首先讨论了空间邻近域内实体间距离对空间异常检测的影响,在此基础上,提出了一种顾及邻近域内实体间距离的空间异常度量方法——SOM 法,并分析了它的复杂度。由于该方法是利用实体非空间属性的加权内插值与实测值的差值作为度量空间异常程度的参数,从而顾及了邻近域内所有实体相互间距离对非空间属性偏离的影响,并且克服了现有检测方法在不均匀分布空间实体集内寻找空间异常的缺陷。最后,通过一个实际算例验证了所提方法的可行性和正确性。

关键词: 空间异常,空间邻近域,空间异常度,距离倒数加权插值