Computers, Environment and Urban Systems 59 (2016) 164–183

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

# Adaptive detection of spatial point event outliers using multilevel constrained Delaunay triangulation

Yan Shi, Min Deng *, Xuexi Yang, Qiliang Liu

*Department of Geo-informatics, Central South University, Changsha, PR China*

A R T I C L E I N F O

*Article history:*
Received 7 July 2015
Received in revised form 12 June 2016
Accepted 12 June 2016
Available online 15 July 2016

*Keywords:*
Spatial point events
Spatial outliers
Multilevel constrained Delaunay triangulation

A B S T R A C T

Spatial outlier detection is a research hot spot in the field of spatial data mining. Because of the lack of specific research on spatial point events, this study presents an adaptive approach for **s**patial **p**oint **e**vents **o**utlier **d**etection (SPEOD) using multilevel constrained Delaunay triangulation. First, the spatial proximity relationships between spatial point events are roughly captured by Delaunay triangulation. Then, three-level constraints are described and used to refine spatial proximity relationships with the consideration of statistical characteristics. Finally, those spatial point events connected by remaining edges are gathered to form a series of subgraphs. Those subgraphs containing very few point events are regarded as spatial outliers. Experiments on both synthetic and real-world spatial data sets are used to show that the proposed SPEOD algorithm can detect various types of spatial point event outliers with high efficiency. Moreover, there is no need to input any parameter in SPEOD.



## 1. Introduction

In recent years, spatial outlier detection has become an important research branch in the field of spatial data mining (Tan, Steinbach, & Kumar, 2006). The aim of spatial outlier detection is to discover those small parts of spatial entities deviating from the global or local distribution in the massive spatial data sets. Correspondingly, spatial outliers are also divided into global and local outliers. Spatial outliers may indicate potential, unknown, and important knowledge rather than only noise in many application domains, for example, geographic information systems, environmental science, meteorology, social economics, and urban traffic. Indeed, spatial outlier detection has been playing an important role in abnormal climate event detection (Telang, Deepak, & Joshi, 2014), abnormal distribution detection of crime and disease events (Cheng & Adepejue, 2013; Janeja & Atluri, 2009), traffic congestion identification (Pan, Zheng, Wilkie, & Shahabi, 2013), and so on.

The concept of outliers was proposed by Hawkins, who defined an outlier as "an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism" (Hawkins, 1980). This definition was expanded from the application of traditional transaction data sets to spatial data sets (Shekhar, Lu, & Zhang, 2001) and described as spatially referenced entities, whose nonspatial attributes are significantly different from those in its spatial neighborhood. Spatial data sets can be divided into two categories: spatial point event data sets and spatial data sets with nonspatial attributes. The former contain only spatial location attributes (e.g., spatial coordinates $X$ and $Y$), while the latter have both spatial location attributes and nonspatial thematic attributes (e.g., precipitation). Existing spatial outlier detection methods were developed mainly for detecting spatial outliers in the spatial data sets with nonspatial attributes. However, there is still a lack of detection methods specially designed for spatial point event data sets; in particular, for complicated spatial point event data sets having arbitrary geometrical shapes and/or different densities. Further, the applications on complicated spatial point events bring new demands for spatial outlier detection methods. Therefore, this study aims to develop spatial outlier detection methods for complicated spatial point events.

The remaining of this study is organized as follows: In Section 2, related works are reviewed in detail and our strategy for detecting spatial point event outliers is presented; and in Section 3, the SPEOD algorithm is presented and fully described. Experiments on both simulated and real-life data sets are described in Section 4. Finally, in Section 5, the interesting findings are summarized and future research works are highlighted.

## 2. Related works and our strategy for detecting spatial point event outliers

### 2.1. Related works

Spatial outlier detection can be seen as an extension of traditional outlier detection from transaction data sets to spatial data sets. This study combines the idea of both traditional outlier detection and current spatial outlier detection to detect outliers from spatial point events.

* Corresponding author at: Dept. of Geo-informatics, Central South University, Changsha, Hunan, PR China.
*E-mail addresses:* csu_shiy@126.com (Y. Shi), dengmin028@yahoo.com (M. Deng), studyang@sina.cn (X. Yang), qiliang.liu@csu.edu.cn (Q. Liu).

http://dx.doi.org/10.1016/j.compenvurbsys.2016.06.001

Therefore, these two types of outlier detection methods are successively reviewed.

Traditional outlier detection methods can be classified into six types as being based on statistics, distance, density, angle, depth, and cluster. Initially, some scholars used statistical methods to detect outliers (Barnett & Lewis, 1994). Because statistics-based methods require a large amount of experimental data to determine the type of data distribution, they lack in applicability. Distance-based methods identify those entities far away from most of the other entities as outliers (Knorr & Ng, 1998; Ramaswamy, Rastogi, & Shim, 2000). These methods are more suitable for the detection of global outliers. Density-based methods use the local density of entities to detect outliers (Breunig, Kriegel, Ng, & Sander, 2000). Density-based methods are better able to identify both global and local outliers than distance-based methods. A series of modified methods based on local outlier factors were then proposed to discover outliers in data sets with uneven density (Chiu & Fu, 2003; Jin, Tung, Han, & Wang, 2006). It is worth noting that all these methods require a large amount of input parameters that unavoidably affect the detection results. Angle-based methods use the angle formed by the entity and its neighbors to measure the outlier degree of each entity (Kriegel, Schubert, & Zimek, 2008). Depth-based methods assign each entity a depth value by using a partition strategy and then organize the data set into different layers (Johnson, Kwok, & Ng, 1998). Those entities in the lower layer are identified as outliers. However, neither angle-based nor depth-based methods can detect outliers with high efficiency and stability. Cluster-based methods define those isolated entities that do not belong to any clusters as outliers (Al-Zoubi, Al-Dahoud, & Yahya, 2010; Han, Kamber, & Tung, 2001). With respect to this method, the detection results depend on the clustering technique involved.

In addition, an extensional research called "rare category detection" aims to locate those rare classes in an unlabeled noisy set (Pelleg & Moore, 2004). However, two characteristics of rare category detection make it different from the outlier detection performed in this study (Vatturi & Wong, 2009). On the one hand, the aim of rare category detection is to identify at least one representative sample that has the highest probability of belonging to each rare class, rather than detecting all members of each rare class. On the other hand, the identified samples under an undiscovered class need to be labeled with the help of experts. Therefore, this study is classified as outlier detection rather than rare class detection.

Most spatial outlier detection methods are developed based on the previous definition of spatial outliers proposed by Shekhar (Shekhar et al., 2001) and can be divided into six types as based on diagram (Haslett, Brandley, Craig, Unwin, & Wills, 1991), distance (Chen, Lu, Kou, & Chen, 2008; Shekhar, Lu, & Zhang, 2003), density (Chawla & Sun, 2006), cluster (Birant & Kut, 2006; Cheng & Li, 2004; Telang et al., 2014), graph (Liu, Lu, & Chen, 2010; Lu, 2011), and model (Cai, He, & Man, 2013; Chen, Lu, & Boedihardjo, 2010). Existing spatial outlier detection methods are designed to identify local differences of nonspatial attribute in spatial neighborhoods; therefore, they are not suitable for spatial point event data sets that have only spatial location attributes. However, they can provide some important thoughts for the outlier detection of spatial point events, for example, detection of local outliers.

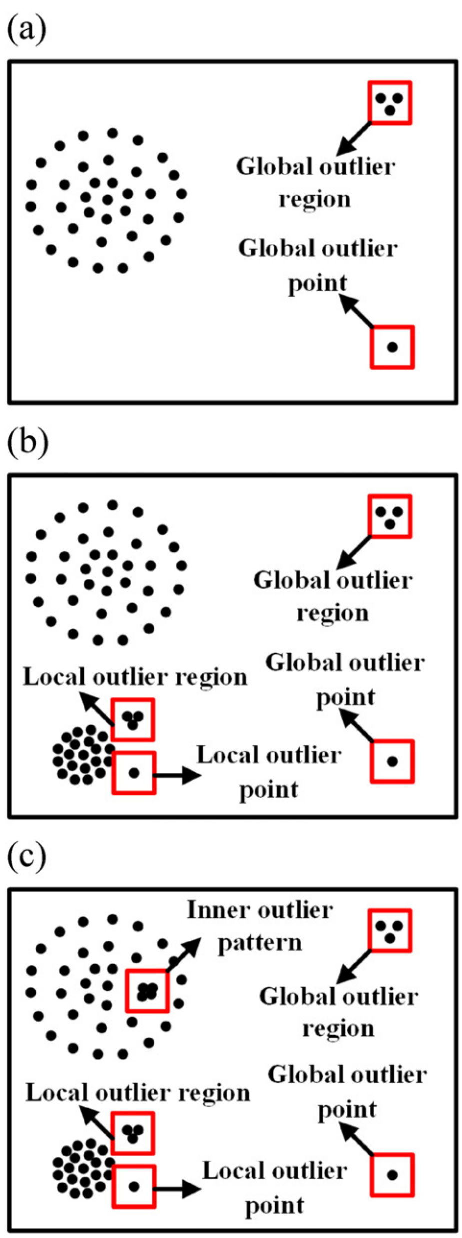


**Fig. 1.** Types of spatial outliers. (a) Global spatial outliers; (b) local spatial outliers; and (c) inner spatial outliers.

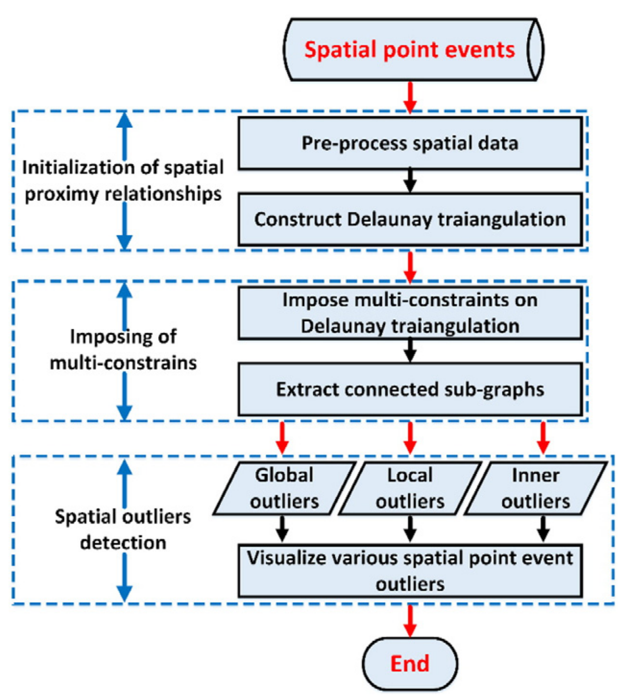


**Fig. 2.** Framework of detection spatial outliers in spatial point event data sets.