Computers & Geosciences 56 (2013) 104–118

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

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

# Adaptive spatial clustering in the presence of obstacles and facilitators

Qiliang Liu [a], Min Deng [b,*], Yan Shi [b]

[a] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong
[b] Department of Surveying and Geo-informatics, Central South University, Changsha, China

ARTICLE INFO

*Article history:*
Received 30 July 2012
Received in revised form
18 February 2013
Accepted 7 March 2013
Available online 15 March 2013

*Keywords:*
Spatial clustering
Delaunay triangulation
Obstacle
Facilitator
Adaptive

ABSTRACT

An intersection-and-combination strategy for clustering spatial point data in the presence of obstacles (e.g. mountain) and facilitators (e.g. highway) is proposed in this paper, and an adaptive spatial clustering algorithm, called ASCDT+, is also developed. The ASCDT+ algorithm can take both obstacles and facilitators into account without additional preprocessing, and automatically detects spatial clusters adjacent to each other with arbitrary shapes and/or different densities. In addition, the ASCDT+ algorithm has the ability to find clustering patterns at both global and local levels so that users can make a more complete interpretation of the clustering results. Several simulated and real-world datasets are utilized to evaluate the effectiveness of the ASCDT+ algorithm. Comparison with two related algorithms, AUTOCLUST+ and DBRS+, demonstrates the advantages of the ASCDT+ algorithm.


## 1. Introduction

As one of the main tasks of spatial data mining, spatial clustering aims to separate a spatial dataset into a series of meaningful groups (also called clusters) without prior labeling (Liu et al., 2012). Clustering spatial points, known as a powerful technology for exploratory data analysis, has been widely applied to epidemic monitoring, geographic customer segmentation, crime hotspot analysis, land use detection, seismicity research, and so on (Miller and Han, 2009). Most spatial clustering algorithms utilize Euclidean distance to measure the proximity of spatial points, and a spatial cluster is usually defined as a set of spatial points that Euclidean distances among them are relatively small. However, there are often some obstacles and facilitators in real applications which make the commonly-used Euclidean distance measure ineffective. Obstacles (e.g. mountains, rivers and lakes) are physical objects which can hinder straight reachability among points, while facilitators (e.g. bridges, highways and high-speed railways) are physical objects which can enhance straight reachability among points.

Taking the synthetic dataset in Fig. 1(a) as an example, where the points can be assumed to be the locations of houses, the rivers are obstacles, and the highway is a facilitator. Most existing spatial clustering algorithms can obtain the clustering result shown in Fig. 1(b). When only the obstacles are considered, the clustering result in Fig. 1(c) can be obtained. If the obstacles and the facilitator are taken into account, the clustering result in Fig. 1(d) can be obtained. In Fig. 1(b), three clusters are detected, though the Euclidean distance among all clusters is uniform and entities in the same cluster have different levels of reachability. In Fig. 1(c), the obstacles are considered; however, the facilitator (highway) between cluster $C_2$ and $C_6$ is ignored. Fig. 1(d) is indeed a good interpretation of the clustering patterns with consideration of both the obstacles and the facilitator.

Spatial clustering in the presence of obstacles and facilitators belongs to the field of constraint-based spatial clustering. The consideration of obstacles and facilitators is able to increase the effectiveness of spatial clustering and capture application semantics. In the research field of facility locations, when planning the location of ATMs or supermarkets, the reachability between residential area and these facilities is seriously influenced by obstacles and facilitators (Tung et al., 2001). In the research field of crime hot spot analysis, the clustering results may be useless or distorted if obstacles and facilitators are ignored (Wang et al., 2011). In the research field of image processing, clustering with certain kinds of pixels as obstacle can improve the effectiveness of image segmentation (Estivill-Castro and Lee, 2004). In addition, the result of spatial clustering that considers obstacles and facilitators may be further used in the fields of spatial association rules mining, cartographic generalization and geographic customer segmentation (Li, 2007; Miller and Han, 2009). Thus, though a constraint-based spatial clustering algorithm is more complex than a traditional spatial clustering algorithm, it is indeed helpful for exploratory spatial analysis. Moreover, a good spatial clustering

* Corresponding author. Tel.: +86 135 0746 7258; fax: +86 731 8883 6783.
*E-mail addresses:* qiliang.liu@connect.polyu.hk (Q. Liu), dengmin028@yahoo.com (M. Deng), shiyan0401060322@126.com (Y. Shi).

0098-3004/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2013.03.002

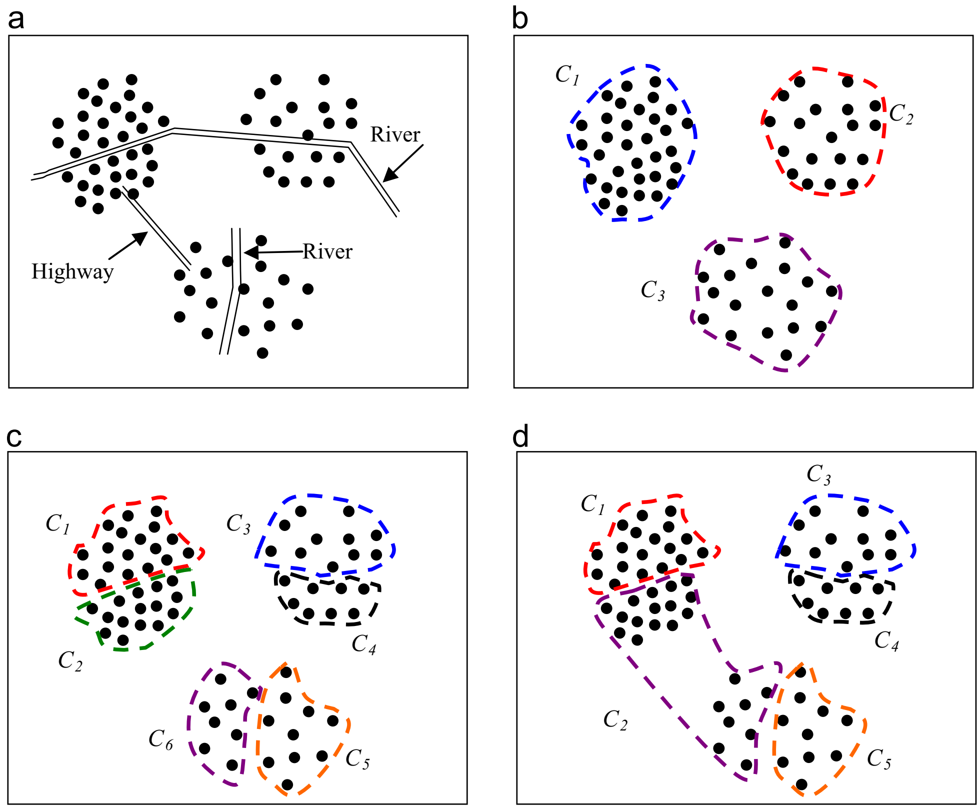


**Fig. 1.** Spatial clustering in the presence of obstacle and facilitator. (a) Spatial data set with obstacles (river) and facilitator (highway); (b) Spatial clustering result ignoring obstacles and facilitator; (c) Spatial clustering result considering obstacles; (d) Spatial clustering result considering both obstacles and facilitator.

algorithm is also expected to have the following characteristics (Ester et al., 1996; Estivill-Castro and Lee, 2002a; Deng et al., 2011):

- Adaptiveness (less input parameters, local parameter rather than global parameter);
- Multi-level (discover clusters at both global and local levels);
- Identify clusters with different densities and arbitrary shapes;
- Clustering the whole dataset rather than a sample;
- High efficiency and effectiveness.

Based on the above, a new strategy for considering obstacles and facilitators in this study, and a novel spatial clustering algorithm is developed with the help of the ASCDT algorithm (Deng et al., 2011), named ASCDT+. The rest of this paper is organized as follows. In Section 2, an overview of related work is first provided, and then the strategy for spatial clustering with obstacles and facilitators is illustrated. In Section 3, the principle of the ASCDT algorithm is briefly mentioned. The ASCDT+ algorithm is fully elaborated in Section 4. In Section 5, the ASCDT+ algorithm and two related algorithms, AUTOCLUST+ and DBRS+, are evaluated using four simulated datasets, and two real world datasets are utilized to test the practicability of the ASCDT+ algorithm. Conclusions and main findings are presented in Section 6.

## 2. Related work and the strategy for spatial clustering with obstacles and facilitators

### 2.1. Related work

Currently, only a few algorithms consider obstacles and/or facilitators in the spatial clustering process. In what follows, a brief overview of some previous algorithms will be provided.

COD-CLARANS (Tung et al., 2001) was the first spatial clustering algorithm designed to consider obstacles in a spatial database. COD-CLARANS is an extension of the classic partitioning-based spatial clustering algorithm called CLARANS (Ng and Han, 1994). The COD-CLARANS algorithm involves three main procedures. The first is to calculate 'unobstructed distance' between points based on a visibility graph. The second is to use the pre-clustering method to obtain micro-clusters. Finally, the CLARANS algorithm is utilized to construct spatial clusters. The COD-CLARANS algorithm can take obstacles into account. However, it has similar limitations to the CLARANS algorithm. Specifically speaking, the COD-CLARANS algorithm is sensitive to density variation and outliers, and it cannot discover clusters with arbitrary shapes. In addition, the algorithm does not consider facilitators.

There are some improved density-based spatial clustering algorithms, such as DBCluC (Zaïane and Lee, 2002), DBRS_O (Wang and Hamilton, 2005) and DBRS+ (Wang et al., 2011). The DBCluC algorithm extends the concepts of density-reachable and density-connected of DBSCAN (Ester, et al., 1996), and is an extended constraint-based DBSCAN algorithm. To model the obstacles, the DBCluC algorithm uses internal edges to fill the visible space of obstacles, called obstruction lines. For an entity, its neighbor should be visible to this entity itself by the obstruction lines. Correspondingly, if two points are not density-connected but there is a facilitator crossing their neighborhoods, they will be regarded as density-connected. Based on these extensions, the DBCluC algorithm is able to consider both obstacles and facilitators. However, the DBCluC algorithm has similar limitations to the original algorithm DBSCAN. Global parameters used by DBCluC make it unable to discover clusters of different densities. However, clusters of different densities usually exist in the real world. For example, the density of houses in the core region of a city is usually higher than that in the suburbs of a city. Furthermore, the