Dynamics of Continuous, Discrete and Impulsive Systems
Series **B:** *Applications & Algorithms* **18 (2011)** 347-364
 http://www.watam.org

A GRID CLUSTERING TECHNIQUE BASED ON MINIMAL SPANNING TREES

Xiyu Liu[1], Wenping Wang[2], Feng Qi[3] and Wenxing Sun[4]

School of Management and Economics
Shandong Normal University, Jinan, Shandong 250014, P.R. China

[1]E-mail: sdxyliu@163.com; xyliu@sdnu.edu.cn
[2]wenping216@126.com; [3]qfjordan@163.com; [4]swenxing@163.com

Abstract. Spatial clustering is an important cluster problem. Typical algorithms include $k-$medoids, density and gravity based clustering, and the minimum spanning tree based clustering. Although there have appeared extensive studies with proposed algorithms and applications, one of the basic computing architecture is that they are all at the level of data objects. The purpose of this paper is to propose a new clustering technique based on grid architecture. Our new technique integrates minimum spanning tree and grid clustering together. A time complexity analysis shows that the new technique has $O(N \log N)$ computing time which is better than many proposed algorithms. This new technique is simple and will apply for large scale clustering problems potentially.

Keywords. Cluster analysis; Minimum spanning tree; Grid clustering.

AMS (MOS) subject classification: 68T30; 68T35.

1 Introduction

Cluster analysis is as an important area in data mining which can explore the hidden structures of business databases [7]. Traditionally, clustering methods can be categorized as several classes. Partitioning method works by constructing various partitions and then evaluating them by some criterion. Hierarchy method creates a hierarchical decomposition of the set of data (or objects) using some criterion. Density-based method is based on connectivity and density functions, Grid-based method is based on a multiple-level granularity structure. Model-based method is to construct a model and to find the best fit model.

Along these lines many techniques and algorithms have been proposed in the literature. For example, Ester et al [6] present the density-based clustering algorithm that uses an Eps-neighborhood for a point containing at least a minimum number of points. Raphael Bar-Or and Christiaan van Woudenberg [1] presented a gravity-based clustering method intended to find clusters of data in $n-$space. The most classical clustering technique is due to Raymond T. Ng and Jiawei Han [16] who developed a CLARANS which aims to use randomized search to facilitate the clustering of a large number of objects. More recent work include agglomerative fuzzy K-Means clustering algorithm by introducing a penalty term to the objective function to make the clustering process not sensitive to the initial cluster centers