

Abstract

Khaled Mahar

using grid for accelerating density-based clustering,

clustering analysis is a primary method for data mining. the ever increasing volumes of data in different applications forces clustering algorithms to cope with it. dbscan is a well-known algorithm for density-based clustering. it is both effective so it can detect arbitrary shaped clusters of dense region efficient especially in existence of spatial indexes to perform the neighborhood queries efficiently. in this paper we introduce a new algorithm gridbscan to enhance the performance of dbscan using grid partitioning merging, yielding a high performance with the advantage of high degree of parallelism. we verified the correctness of the algorithm theoretically experimentally, studied the performance theoretically using experiments on both real synthetic data. it proved to run much faster than original dbscan. we compared the algorithm with a similar algorithm, enhanced dbscan, which is also an enhancement to dbscan using partitioning. experiments showed the new algorithm's superiority in performance degree of parallelism.