

一种基于参考点的快速密度聚类算法

闫 安, 刘琪林

(合肥工业大学 管理学院, 安徽 合肥 230000)

摘 要:提出了一种基于 k 个参考点 DCUR (Density-based Clustering Using k References). 新算法以 k 个参考点反应了数据的分布状态, 然后基于参考点对数据进行聚类分析. 该算法保持了 DBSCAN 的优点, 并且可以减少区域查询次数, 降低 I/O 开销. 理论和实验证明新算法能够有效地对大规模数据库进行聚类, 且其执行效率明显高于传统的基于 R^* 树的 DBSCAN 算法.
关键词: 聚类; 密度; 簇; 参考点; 数据挖掘

A Fast Density Clustering Algorithm Based on Reference

YAN An, LIU Qi-lin

(Hefei University of Technology Anhui Province, School of Management, Hefei 230000, China)

Abstract: In this paper, a new kind of clustering algorithm that is called DCUR (density-based clustering using k references) is presented. Innovative point is that the new algorithm to k references response data Distribution, and then analyzes the data based on the k references. DCUR keeps the ability of density based clustering method's good features, and it can reach high efficiency and the execution frequency of region query can be decreased, and consequently the I/O cost is reduced, so the new algorithm can reach high efficiency and reduce the complexity of the algorithm. Both theory analysis and experimental results confirm that DCUR is effective and efficient in clustering large-scale database, and its executing efficiency is much higher than traditional DBSCAN algorithm based on R^* -tree.

Key words: clustering; density; cluster; reference; data mining

作者简介:

闫 安 男, (1977-), 博士, 副教授. 研究方向为市场营销.

刘琪林 (通讯作者) 男, (1992-), 硕士研究生. 研究方向为数据挖掘. E-mail: 1635813128@qq.com.