

一种基于密度的分布式聚类改进算法

刘恒 1, 苏静 1, 唐咸艳 2, 卢佳佳 3, 梁志胜 4, 洪月华 5

(1 广西医科大学 信息与管理学院, 广西 南宁 530021; 2 广西医科大学 公共卫生学院, 广西 南宁 530021; 3 广西医科大学 研究生院, 广西 南宁 530021; 4 广西医科大学 信息中心, 广西 南宁 530021; 5 广西经济管理干部学院 计算机系, 广西 南宁 530007)

摘要: 基于密度分布式聚类算法 (DBDC) 在分布式聚类运算中运用广泛, 但具有较高的算法时间复杂度. 本文提出了一种基于代表点交互的全局聚类方法, 利用数据网格运算方法将数据对象映射到空间网格, 改进基于密度的分布式聚类算法, 优化原算法的空间搜索过程, 从而改进生成本地聚类的效率. 同时在全局聚类层面可利用中心点作为代表点来降低聚类误差. 通过实验, 表明该方法能够改进基于密度的分布式聚类算法, 较传统分布式聚类算法提高了准确性并降低了算法时间复杂度.

关键词: 聚类; 分布式; 数据挖掘; 代表点

A Density-based Improvement for Distributed Clustering Algorithm

LIU Heng1, SU Jing 1, TANG Xian-yan2, LU Jia-jia3, LIANG Zhi-sheng4, HONG Yue-hua5

(1 School of Information and Management, Guangxi Medical University, Nanning 530021, China; 2 School of Public Health, Guangxi Medical University, Nanning 530021, China; 3 Graduate School, Guangxi Medical University, Nanning 530021, China; 4 Information Center, Guangxi Medical University, Nanning 530021, China; 5 Department of Computer Science, Guangxi Cadres University of Economics and Management, Nanning 530007, China)

Abstract: The density based distributed clustering algorithm (DBDC) has been employed widely in process of distributed clustering. A major drawback of DBDC is the high time complexity. This work presented a method based on representative point for global clustering. A data grid mapping method which mapped data object to space grid, which improved DBDC and enhance the space search process and efficiency of implementation in local clustering level. Clustering error can be reduced by a global clustering method based on representative point. Using verification experiment with real data set from open source software, results showed this algorithm can improve original DBDC in two aspects: accuracy and reducing time complexity.

Key words: clustering; distributed clustering; data mining; representative point

作者简介:

刘恒女, (1979-), 硕士, 副教授. 研究方向为数据挖掘、数据可视化和信息安全.

苏静女, (1983-), 硕士, 讲师. 研究方向为数据挖掘和人工智能.

唐咸艳男, (1982-), 博士, 副教授. 研究方向为数据挖掘、时空流行病学与时空统计学.

卢佳佳男, (1979-), 男, 硕士, 工程师. 研究方向为数据挖掘和信息安全.

梁志胜男, (1979-), 男, 硕士, 高级工程师. 研究方向为数据挖掘和信息安全.

洪月华(通讯作者)女, (1973-), 硕士, 教授. 研究方向为大数据挖掘和人工智能.

E-mail: huibian2005@163.com.