

数据库、信号与信息处理

## Robust的分布式k中心聚类算法的研究与实现

陶冶, 曾志勇

云南财经大学 信息学院, 昆明 650221

收稿日期 2009-4-24 修回日期 2009-6-16 网络版发布日期 2009-11-26 接受日期

**摘要** 并行处理的研究在数据挖掘中是十分必要的。在理论分析的基础上, 提出在对经典串行PAM算法进行并行时应如何从局部聚类信息生成完备的全局聚类信息, 据此提出了算法DPAM, 在提高计算性能的同时, 使聚类质量等价于相应串行PAM算法。为提高并行算法的执行效率, 还介绍了如何减小计算结点间通信的代价。最后对提出的算法进行性能分析和实验, 说明该算法是高效可行的。

**关键词** [聚类](#) [围绕中心点的划分 \(PAM\) 算法](#) [并行](#) [消息传递接口 \(MPI\)](#)

**分类号** [TP311](#)

## Robust distributed k-medoids clustering algorithm

TAO Ye, ZENG Zhi-yong

School of Information, Yunnan University of Finance and Economics, Kunming 650221, China

### Abstract

Parallel is very important in data mining. This paper proposes a distributed k-medoids clustering algorithm by analyzing how to get satisfactory clustering information from local information. Its quality is equivalent to serial PAM algorithm but its calculation performance is higher. The paper gives still the way that improves the efficiency of parallel PAM algorithm by reducing the cost of communication, analyzes the performance of the algorithm and gives the result of experiment. This explains that the algorithm is effective and reliable.

**Key words** [clustering](#) [Partitioning Around Medoids \(PAM\) algorithm](#) [parallel](#) [Message Passing Interface \(MPI\)](#)

DOI: 10.3778/j.issn.1002-8331.2009.32.039

通讯作者 陶冶 [taoye\\_yn@126.com](mailto:taoye_yn@126.com)

### 扩展功能

#### 本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(780KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献](#)

#### 服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

#### 相关信息

- ▶ [本刊中 包含“聚类”的 相关文章](#)
- ▶ 本文作者相关文章
  - [陶冶](#)
  - [曾志勇](#)