

数据挖掘与软件技术

一种改进的密度偏差抽样算法

张建锦¹; 吴渝¹; 刘小霞²

重庆邮电大学计算机科学与技术研究所¹

北京邮电大学²

收稿日期 2007-1-4 修回日期 2007-3-5 网络版发布日期 2007-7-6 接受日期

摘要 随机抽样技术已经广泛应用于数据挖掘的各类算法中，它在处理分布均匀的数据集时非常有效，但在处理分布比较倾斜的数据集时容易丢失小的聚类。为此提出基于网格的密度偏差抽样算法，仅需要扫描一遍数据集就可以得到近似的密度偏差抽样。经实验测试分析表明，该算法不仅提高了聚类的正确性，而且抗噪声能力强、效率高，是解决海量数据挖掘的一种有效途径。

关键词 [数据挖掘](#) [偏差抽样](#) [聚类](#) [数据约简](#) [海量数据](#)

分类号

DOI:

对应的英文版文章: [A7010042](#)

通讯作者:

张建锦 zhangjj616@gmail.com; zhangjj616@yahoo.com

作者个人主页: 张建锦 吴渝 刘小霞

扩展功能

本文信息

▶ [Supporting info](#)

▶ [PDF](#) (815KB)

▶ [\[HTML全文\]](#) (0KB)

▶ [参考文献\[PDF\]](#)

▶ [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](#)

▶ [引用本文](#)

▶ [Email Alert](#)

▶ [文章反馈](#)

▶ [浏览反馈信息](#)

相关信息

▶ [本刊中 包含“数据挖掘”的 相关文章](#)

▶ 本文作者相关文章

· [张建锦](#)

· [吴渝](#)

· [刘小霞](#)