大规模高维数据集中局部异常数据挖掘算法

张凯斐 1，刘继华 1，张菊芳 2
（1 吕梁学院 计算机科学与技术系，山西 吕梁 033000;
2 山西能源学院 基础部，山西 太原 030000）

摘 要：提出一种基于 FFD 的大规模高维数据集中局部异常数据挖掘算法.将 FFD 首次应用在挖掘中，通过引用无线传输技术，将所提方法的宗旨定为对作业级与任务级的实现，以提高局部异常数据抗干扰能力.所提方法利用 FFD 的强控制能力实现无线传输技术与挖掘进程的数据互通，利用 FIFO 挖掘思想依次进行数据本地化与挖掘，并对挖掘流程与目标函数进行了重点设计.实验结果证明，所提方法的可靠性强，挖掘效率高，挖掘任务完成量大.

关键词： 大规模高维数据集；局部异常数据；挖掘算法；本地化
中图分类号：TP311.52
文献标识码：A
文章编号：1000-7180(2018)03-0116-04

Local Outlier Mining Algorithm for Large Scale High Dimensional Data Set

ZHANG Kai-fei1,LIU Ji-hua1, ZHANG Ju-fang2

(1 Department of Computer Science and Technology,Lvliang University,Lvliang 033000,China;2 Basic Courses,Shanxi Institute of Energy,Taiyuan 030000,China)

Abstract：This paper proposes a mining algorithm of large scale and high dimensional data based on FFD local concentration of abnormal data. The FFD was first applied in mining, by referencing the wireless transmission technology, the proposed method is the aim for the realization of the working class and the task level, in order to improve the anti-jamming ability of local abnormal data. The data exchange method based on robust control ability of FFD wireless transmission technology and the data mining process, followed by local mining and using FIFO mining method and the mining process and the objective function of the key The experimental results show that the proposed method has strong reliability, high mining efficiency and large amount of mining tasks.

Key words：Large scale high dimensional data set; Local outlier data; Mining algorithm; Localization

作者简介：
张凯斐男，（1981-），硕士研究生，讲师.研究方向为数据挖掘等.E-mail：zhangkaifei035@163.com.
刘继华女，（1975-），硕士研究生，副教授.研究方向为软件工程、云计算.
张菊芳女，（1977-），硕士研究生，副教授.研究方向为非线性泛函分析.