

# 一种基于多标记的局部离群点检测算法

钱景辉, 梁 栋

(南京工业大学 计算机科学与技术学院, 江苏 南京 211816)

**摘 要:** 基于密度的局部离群点检测算法 (LOF) 不适用于解决高维度、多义性的数据集检测.通过对 LOF 算法的分析,提出了一种基于多标记学习(Multi-Label Learning, MLL)的局部离群点检测算法 MLL-LOF (a local outlier factor based on multi-label learning).该算法采用 MLL 框架,首先将真实对象数据拆分成多示例包形式,然后运用退化策略及相应的权重调整,计算最终离群点因子,判别离群点.并运用实际企业的监控数据将 MLL-LOF 算法与其他经典的局部离群点检测算法进行了对比实验,结果表明提出的 MLL-LOF 算法检测的精准率、召回率、F1 值以及时间效率均优于传统的局部离群点检测算法.

**关键词:** 机器学习; 局部离群点; 多标记; 类别权重

## Local Outlier Detection Algorithm Based on

### Multi-Label Learning

QIAN Jing-hui, LIANG Dong

(College of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China)

**Abstract:** The density-based local outlier detection algorithm (LOF) is not suitable for detecting the data set which is high dimension and polysemous. In this paper, we propose a local outlier detection algorithm based on multi-label learning (MLL-LOF). The main idea of the MLL-LOF algorithm is as follows: Firstly, the real object data is divided into multi-instance by using an MLL framework, then the MLL-LOF calculates the final outlier factor and detects outliers by using degradation strategy and weight adjustment. We compare the MLL-LOF algorithm with other classical local outlier detection algorithms by using actual data set which is comes from enterprise monitoring. Experimental results show that the precision, recall, F1 and time efficiency of the MLL-LOF algorithm are superior to the traditional local outlier detection algorithm.

**Key words:** machine learning; local outlier; multi-label learning framework; class weights

**作者简介:**

钱景辉 男, (1978 -), 博士, 讲师.研究方向为计算机控制自动化.

梁 栋(通讯作者) 男, (1991 -), 硕士研究生.研究方向为机器学习、离群点检测. E-mail: js\_liangdong@163.com.