

基于数据集对象平均离群因子的离群点选择算法

朱付保,徐显景,白庆春,朱颢东

(郑州轻工业学院 计算机与通信工程学院,河南 郑州 450002)

摘要:针对传统的局部离群点检测算法中存在离群点判定的主观性过强的问题,通过研究局部离群点检测相关算法,提出了一种基于数据集对象平均离群因子的离群点选择算法。该算法首先求得各数据对象的邻域对象,进而根据邻域数据集合求出各自的离群因子,在进行离群点选择时计算出数据集的平均离群因子,将每个对象的离群因子与平均离群因子进行比较判断对象是否为离群点。理论分析和实验对比结果均表明,该算法在进行离群点选择时可以有效地避免离群点选择的主观性,更好地提高检测的准确率。

关键词:离群点;局部偏差因子;加权属性;数据挖掘

中图分类号:TP315

文献标识码:A

文章编号:1000-7180(2016)01-0131-04

Local Outlier Selection Algorithm Based on Average Outlier Factor

ZHU Fu-bao, XU Xian-jing, BAI Qing-chun, ZHU Hao-dong

(School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China)

Abstract: Aiming at the shortage of subjectivity existed in the traditional local outlier detection algorithm, we studied the relevant algorithm of local outlier detection, and a local outlier selection algorithm based on the average outlier factor was proposed. This algorithm calculates the neighborhood of each data object, and figures out the outlier factor according to the dataset of its neighborhood. Comparing the outlier factor of each object with the average outlier factor to determine whether the object is an outlier. Theoretic analysis and experimental results show that this algorithm can avoid the subjectivity of selection and improve the accuracy of selection.

Key words: outlier; local deviation factor; weighted attribute; data mining

作者简介:

朱付保 男,(1974-),博士,副教授,硕士生导师.研究方向为智能信息处理、空间数据库、地理信息系统。

徐显景 男,(1988-),硕士研究生.研究方向为智能信息处理、空间数据挖掘、地理信息系统。