[打印本页]　[关闭]

机器学习与数据挖掘

基于属性约简和相对熵的离群点检测算法

胡云1,2,李慧1,施珺1,蔡虹1

1.淮海工学院计算机工程学院, 江苏 连云港 222000; 2.南京大学计算机科学与技术系, 江苏 南京 210000

摘要：

本研究结合信息熵与粗糙集理论中的属性约简技术,提出了一种新颖的离群点检测算法。这种方法通过在更小的属性子空间去获得相同或相近的离群数据集,使对离群数据的分析更加集中于较小的目标域。该算法对原属性空间进行划分,通过分析计算将具有最大相对熵与负相对势的对象集合判定为离群点集合。为了验证算法的有效性,还在通用数据集上进行了测试,理论分析和实验结果表明该离群点检测算法是有效可行的。

关键词： 属性简约　相对熵　离群点检测

An outlier detection algorithm based on attribute reduction and relative entropy

HU Yun1,2, LI Hui1, SHI Jun1, CAI Hong1

1. School of Computer Engineering, Huaihai Institute of Technology, Lianyungang 222000, China;
2. Department of Computer Science and Technology, Nanjing University, Nanjing 210000, China

Abstract:

A new outlier detection algorithm combining a rough set and information entropy technology was proposed. This approach could obtain similar outlier sets by means of searching in an attributes subspace, which could lead the analysis of outlier detection to focus better on narrow and specific object fields. This algorithm divided the original attribute space into several segments, which filtered out those subjects with largest relative entropy negative relative cardinality as the outliers. To prove this algorithm's effectiveness, experiments on a real world dataset were conducted. Theoretical analysis and experimental results showed that this method of outlier detection was efficient and effective.

Keywords: attribute deduction　relative entropy　outlier detection

通讯作者：

作者简介：胡云(1977-),女,江苏连云港人,讲师,博士研究生,主要研究方向为数据挖掘,智能信息处理. E-mail: huyunzhang@yahoo.com.cn
作者Email:

PDF Preview

参考文献：

本刊中的类似文章

1．罗玉盘 商琳.基于多粒度周期模式的时序离群点检测算法[J]. 山东大学学报(工学版), 2009,39(3): 11-15

2．黄添强1,2,陈智文1.基于双向运动矢量的数字视频篡改鉴定[J]. 山东大学学报(工学版), 2011,41(4): 13-19

扩展功能

本文信息

▸ Supporting info
▸ PDF(1310KB)
▸ 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

本文关键词相关文章

▸ 属性简约
▸ 相对熵
▸ 离群点检测

本文作者相关文章

PubMed