

论文

采样特异性因子及异常检测

孙静宇,余雪丽,陈俊杰,李鲜花

太原理工大学计算机科学与技术学院, 山西 太原 030024

摘要:

特异性因子是数据的重要特征之一,常通过累计数据之间的差异得到,是面向特异性挖掘的核心概念,然而遇到了计算时间复杂度过高的问题。本文在分析已有特异性因子定义特点及其计算算法时间复杂度的基础上,指出应该基于采样的方法定义特异性因子。给出了一种基于采样的特异性因子定义,即采样特异性因子(sampled peculiarity factor,SPF),并提出了一种基于SPF的异常检测算法。在真实数据集上进行对比实验,结果表明:该算法在检测异常数据时,精度降低不明显,而运行效率得以较大提高,这说明基于采样定义特异性因子的方法可行和更为合理。本文还指出采用合适的采样方法可经进一步优化SPF的计算过程,进而节约占用CPU时间和满足实时性要求高的应用。

关键词: 采样 特异性因子 异常检测 数据挖掘 时间复杂度 实时性

Sampled peculiarity factor and its application in anomaly detection

SUN Jing-yu, YU Xue-li, CHEN Jun-jie, LI Xian-hua

College of Computer Science and Technology, Taiyuan University of Technology, Taiyuan 030024, China

Abstract:

The peculiarity factor (PF), an important feature of data and obtained by accumulating differences between data, is a core concept of peculiarity-oriented mining (POM). But it meets a higher computational time complexity for any algorithm. A sampled approach firstly was suggested to define PF through analyzing current versions of PF and computational complexities of algorithms to compute it. A sampled PF (SPF) was proposed to meet real time requirement and a SPF outlier detection algorithm was given. Experiments using real datasets show that the SPF-outlier detection algorithm is efficient with losing a few of precisions through contrasting with two baseline algorithms and it is a feasible and right approach to define PF by sampling. Furthermore, some right sampling methods could be used to compute SPFs in order to meet real-time requirement.

Keywords: sampling peculiarity factor outlier detection data mining time complexity real-time

收稿日期 2010-05-20 修回日期 网络版发布日期

DOI:

基金项目:

山西省自然科学基金资助项目(200821024)

通讯作者:

作者简介: 孙静宇(1975-),男,博士生,讲师,主要研究方向为Web Search、案例推理、个性化推荐技术、数据挖掘。

E-mail: whitesunpersun@163.com

作者Email:

PDF Preview

参考文献:

本刊中的类似文章

1. 李莉,杨忠,邢建芳,沈春林.面向立体显示的点采样栅格优化策略及其性能分析[J]. 山东大学学报(工学版), 2008,38(3): 1-6
2. 李丽萍1,杨静2,刘常春1,刘澄玉1,张庆广1,李可1.重采样对脉搏间期序列谱分析的影响[J]. 山东大学学报(工学版), 2011,41(2): 102-106

扩展功能

本文信息

Supporting info

PDF(362KB)

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

采样

特异性因子

异常检测

数据挖掘

时间复杂度

实时性

本文作者相关文章

PubMed

3. 王心一¹,杜光^{2*}.降采样固定时延估算法在声回波对消系统中的应用[J].山东大学学报(工学版),2011,41(3):42-45

Copyright by 山东大学学报(工学版)