

基于多示例学习的时序离群点检测算法研究

钱景辉, 卢屹韦

(南京工业大学 计算机科学与技术学院, 江苏 南京 211816)

摘要: 在 FindCBLOf 算法的基础上, 提出了一种基于多示例学习的 FindCBLOF 离群点检测算法(MIL-FindCBLOf)用于时序离群点检测.算法首先对数据进行分段聚集, 再利用多示例框架, 封装每个对象, 以此保留每个对象的属性, 然后采用全局策略计算对象的因子数值, 最后通过计算平均因子来确定离群序列.在实际企业的实时采集监控系统中, 将 MIL-FindCBLOf 算法与其它经典离群点检测算法进行实验对比, 结果表明本文提出的 MIL-FindCBLOf 算法相对其它算法提高了检测的全面性和准确性.

关键词: 机器学习; 时序离群点; 多示例学习; 聚类; 平均因子

Multi-instance-learning Based Time Series

Outlier Detection Algorithm

QIAN Jing-hui, LU Yi-wei

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

Abstract: On the basis of the FindCBLOf algorithm, this paper proposes a FindCBLOF outlier detection algorithm (MIL-FindCBLOf) for Multi instance learning based the detection of time series outliers. Firstly, the data were gathered using piecewise multi instance framework to package each object in order to preserve the properties for each object, to calculate the object factor with global strategy, to determine the outlier sequence by calculating the average factor. In the real-time monitoring system of practical enterprise MIL-FindCBLOf algorithm will be compared with other classical outlier detection algorithm by experiments.The results show that the proposed MIL-FindCBLOf algorithm compared with other algorithms in this paper can improve the accuracy and comprehensiveness of detection.

Key words: machine learning; sequential outlier; multi instance learning; clustering; average factor

作者简介:

钱景辉 男, (1978-), 硕士, 讲师.研究方向为计算机控制自动化.

卢屹韦(通讯作者) 男, (1992-), 硕士研究生.研究方向为数据挖掘、人工智能.E-mail: 362781537@qq.com.