doi:10.1631/FITEE.1800038

题目：FAAD：一种无监督快速准确的数据流上多维序列异常检测方法

概要：最近，序列异常检测被广泛应用于许多领域。这些领域中的序列数据在数据流上通常是多维的。设计同时满足检测精度和速度的数据流上多维序列异常检测方法是一个挑战。因为：（1）序列数据和庞大状态空间的维度冗余导致序列建模能力较差；（2）异常检测无法适应数据流的高速性，尤其是概念漂移会降低检测率。一方面，大多数现有序列异常检测方法集中在单维序列。另一方面，多维序列研究主要集中在静态数据集而非数据流。为提高数据流上多维序列异常检测性能，提出一种新型无监督快速和准确异常检测（fast and accurate anomaly detection，FAAD）方法，该方法包括 3 种算法。首先，采用一种“信息计算和最小生成树聚类”（information calculation and minimum spanning tree cluster，IMC）方法减少冗余维度。其次，为加速模型构建确保数据流上序列的检测率，提出一种“基于索引概率后缀树的随机抽样和子序列划分”（random sampling and subsequence partitioning based on the index probabilistic suffix tree，RSIPST）方法。最后，“基于模型动态调整的异常缓冲”（anomaly buffer based on model dynamic adjustment，ABMDA）方法显著降低数据流中概念漂移的影响。在流平台 Storm 上实施 FAAD 检测多维日志审计数据。与现有异常检测方法相比，FAAD 在检测精度和速度方面不受概念漂移影响，具有良好性能。

关键词：数据流；多维序列；异常检测；概念漂移；特征选择