

# 基于时空相关性的多传感器数据异常检测

柳月强 张建锋\* 祝麒翔 杨会君

(西北农林科技大学信息工程学院 陕西 杨凌 712100)

**摘要** 针对环境检测系统中多传感器数据同时具有时间和空间的相关性但在进行异常检测中往往只从单一角度解决的问题,将最近邻检测与聚类检测的思想相结合,提出一种基于时空相关性的异常检测算法(ODSTC)。采用动态阈值的 K 近邻距离差算法(KNND)实现数据集在时间序列上的异常检测;采用参数优化后的 DBSCAN 聚类算法(PO\_DBSCAN)实现多密度的空间聚类异常检测。采用校试验示范站信息服务体系项目数据集进行实验,并与 DBSCAN 和 PO\_DBSCAN 等算法进行对比实验。结果表明,该算法的检测率可达 98.5%,且时间性能也有所提升。

**关键词** 数据分析 异常检测 K 近邻距离差 DBSCAN 聚类 时空相关性

中图分类号 TP391 文献标志码 A DOI:10.3969/j.issn.1000-386x.2020.10.014

## OUTLIERS DETECTION OF MULTI-SENSOR DATA BASED ON SPATIAL-TEMPORAL CORRELATION

Liu Yueqiang Zhang Jianfeng\* Zhu Qixiang Yang Huijun

(College of Information Engineering, Northwest A&F University, Yangling 712100, Shaanxi, China)

**Abstract** In the environmental detection system, multi-sensor data has temporal and spatial correlation at the same time, but in the process of outliers detection, it is often solved from a single point of view. To solve the problem, we propose an outlier detection method based on spatial-temporal correlation (ODSTC), which combines nearest neighbor detection with cluster detection. The k-nearest neighbor distance algorithm (KNND) based on dynamic threshold was used to detect the outliers in time series. Then the parameter optimized DBSCAN clustering algorithm (PO\_DBSCAN) was used to realize the multi-density spatial clustering outlier detection. The experiment used the data set of the information service system project of the University demonstration station, and compared it with DBSCAN and PO\_DBSCAN. The results show that the detection rate of our method can reach 98.5%, and the time performance of the algorithm is also improved.

**Keywords** Data analysis Outliers detection K-nearest neighbor distance DBSCAN clustering Spatial-temporal correlation

## 0 引言

传感器持续采集的数据构成数据流,每一个传感器所产生的数据流具有连续、实时、高速等特征<sup>[1]</sup>。受传感器数据概念漂移、运行环境以及传感器部署方案等因素的影响,传感器数据具有不确定性<sup>[2]</sup>。应用异常检测技术对部署在同一物理环境下的多个传感器数

据流进行分析,将在一定程度上提高传感器数据的准确性和有效性<sup>[3]</sup>。传感器在对物理环境进行感知的过程中可获得在时间上连续的离散点,数据观测值与时间之间的关联关系称之为数据的时间相关性,传感器数据流也称为时间序列。在传感器网络部署中往往会利用不同类型的多个传感器对物理空间进行全面感知,这些不同类型的传感器数据间具有的关联关系称为数据的空间相关性。根据无线传感器网络(WSN)

的特点,将异常检测的方法分为基于统计、基于分类、基于聚类、基于近邻等方法。单传感器数据流通常利用数据的时间相关性进行异常检测,多应用基于统计分析、基于最近邻距离等方法进行异常检测<sup>[4-5]</sup>。多传感器数据流同时具有时间和空间相关性,通常应用基于聚类的方法进行检测,在应用聚类的过程中往往会忽略传感器的时间相关性。

单传感器数据流异常检测是多传感器数据流异常检测的基础。段青玲等<sup>[6]</sup>应用滑动窗口和支持向量回归相结合的方法,预测传感器的测量值,确定置信区间,根据传感器的实测值是否在区间内来判断数据是否异常。苑进等<sup>[7]</sup>采用高斯过程对温室中真实传感器数据进行单步预测,通过计算数据区间差异度来判断异常,以上方法均无法对传感器数据的多维度、属性关联等因素进行预测。Sarvani等<sup>[8]</sup>应用K-means算法对具有5个属性的鸢尾花数据进行聚类分析,与传统方法相比,基于聚类的孤立点剔除(ORC)技术取得了更好的效果。而DBSCAN空间算法以其对小样本数据聚类速度快、有效处理噪声点、发现任意形状等优势备受关注,在异常检测技术领域也有研究,但目前应用较少。

本文提出一种基于时空相关性的多传感器数据异常检测方法,首先利用最近邻距离差算法实现数据流在时间序列上的孤立点和孤立群检测;再应用DBSCAN聚类算法对最近邻距离差异常处理后的数据集进行空间相关性聚类,实现多传感器数据流的异常检测。

## 1 相关定义与技术

### 1.1 异常点定义

异常点是指当前数据观测值与其他历史数据观测值之间有较大偏差的离群点,是一种由不同于正常的机制所产生的点<sup>[9]</sup>。不同的异常现象将会导致不同的数据异常,采取的措施也会有所不同。Shikha等<sup>[10]</sup>通过对物联网环境的研究,将产生异常的原因分为两种,即事件和错误。

(1) 事件:指与预定义的“通常行为”相比,物理环境状态突然发生剧烈变化。一类是人类所不能控制的自然灾害如地震、火山等;另一类是由人类自身活动导致物联网系统受到攻击而无法工作。事件通常持续时间较长,同时会改变传感器数据的历史模式。

(2) 错误:指由于传感器数据噪声、数据概念漂移,导致无线传感网络传输丢帧、网络延缓等而造成观测值不匹配且反复出现的一种变化。

本文只研究因错误导致的异常现象,将异常点定义为点异常、集体异常和关联异常,其定义如下。

点异常:指单数据的实例与数据的正常模式不同,也可称为数据跳跃。该异常点的观测值与近邻观测值差异明显,最易识别;集体异常:指一组连续的数据实例,其模式与整个正常模式不同;关联异常:指数据实例在特定场景下发生的异常,也可称为条件异常。

### 1.2 相关技术

最近邻距离差的基本思想是度量对象之间的近邻性,异常对象是指近邻度较小的对象<sup>[12]</sup>。算法的优点在于时间复杂度低,可以有效地应用于时间序列异常检测。近邻距离差算法适合用于各种维度的数据集,假设数据集 $X$ 的数据对象总数为 $N$ ,数据维度为 $M$ ,数据集 $X$ 可表达为:

$$X = \{x_i | x_i \in \mathbf{R}^M, 1 \leq i \leq N\} \quad (1)$$

其中数据对象 $x_i$ 为:

$$x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\} \quad (2)$$

相邻对象之间的距离为:

$$D_{ij} = \{|x_i - x_j|, 1 \leq i \leq N, 1 \leq j \leq N\} \quad (3)$$

该方法在判断异常的过程中需要设置数据的邻差阈值,通过邻差阈值与测量值的比较来判断是否异常,由于数据在不同的情况下波动不均衡,所以固定的阈值将会对数据的异常检测造成不好的影响。

DBSCAN是一种将足够密度区域作为聚类中心,不断生长该区域的基于密度的空间聚类算法,其聚类原理如图1所示,参数 $Eps$ 为圆半径,参数 $MinPts$ 为邻域内样本点数,在本例中 $MinPts \geq 4$ 。图中以 $A$ 点为核心点, $Eps$ 为半径,邻域 $A$ 中包含 $A, a_1, a_2, a_3$ 4个点,选择 $Eps$ 邻域的边界点继续拓展邻域得到新的边界点如点 $B$ 、点 $C$ 所示,随着样本点不断扩散, $N$ 在所有邻域之外,则被标记为离群点。图中 $A$ 在 $a_1$ 邻域内, $A$ 与 $a_1$ 直接密度可达, $A$ 同时也在 $a_3$ 邻域内,所以 $a_1$ 与 $a_3$ 密度可达。点 $C$ 与 $a_3$ 密度可达, $a_3$ 与 $a_1$ 密度可达,则可知点 $C$ 与 $a_1$ 密度相连,进而寻找到密度相连对象的最大集合即为聚类结果。由于DBSCAN算法采用的是全局参数,在处理不平衡数据时聚类效果不稳定且在大样本聚类中收敛时间长,因此降低全局参数影响是DBSCAN算法首要解决的问题<sup>[11]</sup>。

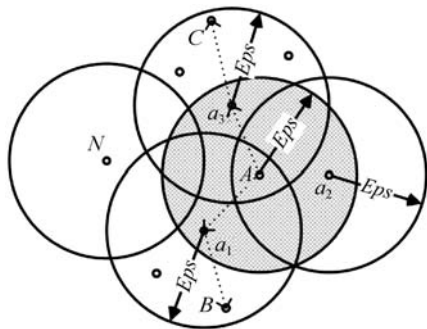


图1 DBSCAN算法原理图

## 2 基于时空相关性的异常检测算法

本文针对环境传感器数据的特点,应用参数优化来提升算法性能,将改进后的 K 近邻距离差(KNND)与聚类算法(PO\_DBSCAN)结合,从时间和空间两个角度分析多传感器数据异常检测的问题。在检测数据的时间相关性中,应用最近邻距离差算法的动态阈值克服传感器数据在时间序列上波动不均匀的问题,通过 K 个近邻点的距离差排序计算数据异常得分,然后采用得分最高点的 2K 个近邻点的距离均值来优化邻差阈值,完成异常检测。在 DBSCAN 算法中,采用滑动窗口模型为多传感器数据流设置固定的窗口大小和窗口重叠率,将数据集划分为等量小样本。DBSCAN 算法聚类的参数可根据窗口内部样本特征设置,克服算法采用全局参数、大样本收敛慢的问题,完成异常检测。本文算法包括最近邻距离差算法、DBSCAN 算法参数优化与异常检测三部分。

### 2.1 KNND 参数优化

传统的最近邻距离差算法对局部数据不均匀波动的处理具有局限性,受文献[12]的启发,本文提出一种新的基于 K 近邻距离差的异常检测方法。通过参数优化,将会有效地解决各个传感器在时间序列上的异常问题,其步骤如下。

**Step 1** 获取所有传感器数据流,按照式(3)计算数据之间的距离差  $D_{ij}$ 。

**Step 2** 设置近邻参数 K,以 K 为窗口大小,窗口滑动步长为 1,初始化异常得分  $S=0$ 。

**Step 3** 对窗口内点之间的距离差进行排序,选择距离差较大的  $q$  个点为距离优胜点,并对其异常得分  $S$  加 1,滑动窗口循环求出各个点的异常得分  $S$ 。

**Step 4** 统计标记异常得分为 K 的点,并求出该点左右相邻的  $2K-1$  个点距离差的平均值  $mean$ 。

**Step 5** 对异常得分为 K 的点进行判断,根据环境数据的不突变性和因错误导致异常的可控性,定义距离差超过均值的 1.5 倍则标记该点为异常点。

**Step 6** 由 Step 5 可以确定数据中的点异常,判断点异常处邻近点的距离差正负变化情况,正负交替则将点标记为异常点,直到有连续单一变化则异常停止,该过程捕获的连续异常点数大于 K 则将连续数据标记为集体异常,反之为点异常。

**Step 7** 多维异常数据整合处理,对点异常通过近似值代替,消除异常;集体异常不做修改。

**Step 8** 整理数据,将 K 近邻距离算法检测并处理的数据集作为聚类算法的输入。

### 2.2 DBSCAN 参数优化

基础的 DBSCAN 算法采用全局唯一参数  $Eps$  和  $MinPts$  实现聚类,受文献[13]采用局部参数进行聚类的影响,本文提出一种基于滑动窗口数据划分技术实现数据集在小样本数据上利用局部参数进行密度聚类的方法。PO\_DBSCAN 聚类算法流程如图 2 所示。

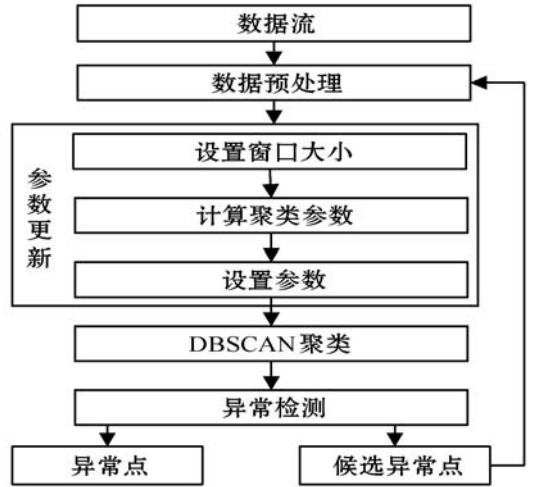


图 2 PO\_DBSCAN 算法流程

PO\_DBSCAN 聚类算法的聚类过程由三部分组成,分别是参数更新、聚类、异常检测。

参数更新过程中设置聚类窗口的大小并计算窗口内各属性间的平均距离差,以  $K$  为  $MinPts$  邻域内点的个数,以属性间的欧氏距离为半径  $Eps$ ,保证数据在单独情况下聚类无误。其公式为:

$$Eps = \sqrt{\sum_{i=1}^M y_i^2} \quad (4)$$

式中: $y_i$  为属性的平均距离差; $M$  为滑动窗口的大小。

聚类过程中采用 DBSCAN 算法进行聚类,针对属性间相关性不一致的问题,通过给每一个属性分配权重来降低对聚类效果的影响,权重  $W_{xy}$  通过相关系数计算,其公式如下:

$$r(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) Var(Y)}} \quad (5)$$

$$W_{xy} = \begin{cases} 0 & 0 \leq r(X, Y) \leq 0.1 \\ \frac{1}{r(X, Y)} & 0.1 < r(X, Y) \leq 1 \end{cases} \quad (6)$$

式中: $Cov(X, Y)$  为属性 X 与属性 Y 的协方差; $Var(X)$  为属性 X 的方差; $Var(Y)$  为属性 Y 的方差。

异常检测过程将对聚类结果进行分析,在聚类过程中将首次被标记为异常点的对象记为候选异常点,并设置异常得分加 1(初值为 0),候选异常点进入下一次循环,继续进行聚类标记,更新异常得分,异常得分

$S$  与聚类次数  $C$  (聚类次数为滑动窗口重叠率的倒数) 相等则标记为异常点, 反之为正常点。

### 2.3 ODSTC 异常检测算法

综上所述, ODSTC 算法首先利用 KNND 算法进行异常检测与处理, 然后利用 PO\_DBSCAN 算法进行聚类检测并标记, 最后对所有异常点以及异常点类型进行统一标记, 算法框架如算法 1 所示。

**算法 1** 基于时空相关性的多传感器数据异常检测

输入: 校试验示范站信息服务体系项目多传感器数据集。

输出: 通过算法检测重新标记的异常点以及异常点所属的类型。

1. 应用滑动窗口技术划分传感器数据流
2. 基于 KNND 算法对传感器数据流进行时间特性上异常检测
3. 对 KNND 算法检测出来的异常点用不同的处理方式进行处理
4. 将处理后的数据集, 再次利用滑动窗口模型划分多传感器数据流, 作为聚类的输入
5. PO\_DBSCAN 异常检测算法对多传感数据流进行异常检测
6. 对聚类异常检测出来的异常进行标记
7. 异常点统一标识, 并标记异常点类型

### 2.4 ODSTC 算法性能分析

算法的性能分析往往从时间效率和空间效率两个角度进行, 随着计算机速度与容量的不断提升, 空间效率已经不再是关注重点, 故本文着重从时间效率进行分析。

KNND 算法将数据进行划分, 然后再对滑动窗口内的点进行排序, 随着窗口的滑动来检测异常点。算法的时间复杂度可表示为:

$$f(n) = (n - w) \cdot w \cdot \log w \quad (7)$$

式中:  $n$  为算法输入规模; 常数  $w$  为滑动窗口大小, 滑动步长为 1。所以 KNND 算法的时间复杂度  $T_1(n)$  为:

$$T_1(n) = O(f(n)) = O(n - w) = O(n) \quad (8)$$

DBSCAN 算法的时间复杂度为找出半径  $Eps$  邻域中的点所需要的时间, 在最坏情况下其时间复杂度是  $O(n^2)$ 。PO\_DBSCAN 算法是采用滑动窗口对数据进行划分, 其时间复杂度可表示为:

$$g(n) = \left(\frac{xn}{w} + 1\right)w^2 \quad (9)$$

式中:  $n$  为算法输入规模,  $w$  为滑动窗口大小, 即  $w^2$  为常数,  $x$  为常数, 即滑动步长为  $w/x$ , 所以 DBSCAN 算法的时间复杂度  $T_2(n)$  为:

$$T_2(n) = O(g(n)) = O\left(\frac{xn}{w} + 1\right) = O(n) \quad (10)$$

ODSTC 算法的时间复杂度为 KNND 与 PO\_DBSCAN 算法时间复杂度之和, 由加法法则可将 ODSTC 算法的时间复杂度表示为:

$$T(n) = O(\max(f(n), g(n))) = O(n) \quad (11)$$

由此可知, ODSTC 的时间复杂度呈线性增长, 随着处理数据样本量的增大, 时间效率比基础的 DBSCAN 算法效率更高。在 KNND 算法中参数利用了时间序列的数据的不突变特性, PO\_DBSCAN 算法利用了多维数据的空间相关特性。充分考虑了数据相关性, 有效地挖掘出数据的潜在关系, 对提高算法检测的准确率具有积极意义。

## 3 实验及结果分析

为了评估 ODSTC 算法的性能, 本文采用西北农林科技大学试验示范站的环境数据进行实验。实验在 Inter(R) Core(TM) i5-6400 CPU, 主频 2.70 GHz, 内存 8 GB, 操作系统 Windows 10 环境下进行, 编程环境采用 MATLAB 2016a。分别实现了 DBSCAN、PO\_DBSCAN、KNND 以及 ODSTC 算法, 并根据不同规模数据集和不同算法进行对比实验。

### 3.1 数据集

环境数据集于 2017 年 5 月到 2017 年 10 月在西北农林科技大学千阳苹果试验示范站采集。环境传感器采集周期为 10 分钟, 采集的物理量包括空气温度、空气湿度、总辐射、风向、风速、降雨量、土壤温度、土壤湿度, 构成具有 8 个属性的数据集。本文所采集的数据是集成在一个开发板上的多个传感器分别采集的多维数据。根据数据之间的相关性强度, 本文选取物理量空气温度、空气湿度、总辐射、风速四个强相关的属性进行相关性实验。在进行对比实验中, 本文采用四组不同规模的数据集进行验证实验, 实验样本记录为 1 440 ~ 8 640 条, 涉及数据点数为 5 760 ~ 69 120 个。千阳苹果试验示范站的环境相关数据均结合实际情况经过专家标记, 异常来源均为错误, 异常点分三类依次标。数据集详细信息如表 1 所示。

表 1 实验数据集

数据集	样本量	样本属性数	数据点	点异常数	集体异常数	关联异常数
QY_1	1 440	4	5 760	72	171	72
QY_2	4 320	4	17 280	208	502	164
QY_3	8 640	4	34 560	421	997	324
QY_4	8 640	8	69 120	621	1 476	408

### 3.2 性能评价

为了衡量算法特性,本文引入性能指标检测率 TPR(True Positive Rate)、误报率 FPR(False Positive Rate)和运行时间  $T$ 。TPR 是用来衡量异常值被正确标记的概率,FPR 是用来衡量正常值被错误标记的概率,计算如下:

$$TPR = \frac{TP}{TP + FN} \times 100\% \quad (12)$$

$$FPR = \frac{FP}{FP + TN} \times 100\% \quad (13)$$

式中: $TP$  表示被正确检测出的异常值; $TN$  表示未被检测出来的异常值; $FP$  表示正常值被检测为异常值; $FN$  表示正常值被检测为正常值。

### 3.3 参数设置

基于时空相关性数据异常检测算法分为三个步骤:首先应用 KNN 检测;然后应用 PO\_DBSCAN 算法进行聚类异常检测;最后对异常检测结果进行统一标记。整个算法在运行过程中需要设置的参数有:近邻点个数  $K$ 、KNN 检测滑动窗口大小  $w_1$ 、KNN 检测窗口滑动重叠率  $d_1$ 、聚类检测滑动窗口大小  $w_2$ 、聚类检测滑动窗口重叠率  $d_2$ 。在聚类过程中因为数据的冗余影响聚类效果,滑动窗口根据数据的变化规律以及 DBSCAN 算法不适合大样本聚类的特点设定,近邻阈值参数  $Threshold$  等于  $2K$  个近邻元素距离差均值  $mean$ ,聚类参数  $MinPts$  等于  $K$ , $Eps$  计算见式(4)。算法所需设置的参数详细信息如表 2 所示。

表 2 ODSTC 算法参数

参数	KNN			PO_DBSCAN	
	$K$	$w_1$	$d_1$	$w_2$	$d_2$
值	5	5	0.8	144	0.75

### 3.4 实验结果分析

实验采用 4 个数据集对三种不同算法进行对比,检测率和误报率为数据集中三类异常点的平均值,详细如表 3 所示。可以看出,本文所提出的 ODSTC 算法在 TPR 性能和 FPR 性能上较 DBSCAN 和 PO\_DBSCAN 均有提升,随着数据集的增大,算法的 FPR 平均降低 0.3 个百分点。对比文献[14]与本文 PO\_DBSCAN 算法,算法识别的准确率提升了 2 个百分点。对比数据集 QY\_3 与 QY\_4 可以发现,当算法中不相关属性增加,算法的 TPR 约下降 6 个百分点,算法性能有所降低。由此可知本文算法在处理相关数据集时具有良好的检测效率,但是针对数据中的不相关数据集其性能

仍有待提升。综上,本文算法与其他异常点检测算法对比可得,前者的准确率提升较大,且误检率也有明显的降低。算法的检测效率满足应用需求。

表 3 异常检测算法性能对比

数据集	%					
	文献[14] DBSCAN		PO_DBSCAN		ODSTC	
	TPR	FPR	TPR	FPR	TPR	FPR
QY_1	93.4	3.6	96.6	3.4	98.4	1.8
QY_2	94.6	3.4	96.9	2.7	98.5	1.5
QY_3	94.7	3.0	96.6	2.3	98.4	1.1
QY_4	87.3	11.2	88.2	12.4	93.6	8.7

本文截取数据集 QY\_1 中空气温度的数据的检测结果进行统计,其中点异常 17 个,聚集异常 58 个,关联异常 17 个,检测率为 98.5%,误报率 1.5%。为了更清晰地描述异常点被标记情况,本文截取空气温度数据的前 180 个样本点进行绘图,检测结果如图 3 所示。空气温度数据的异常均在短时间内发生(每个样本采集间隔 10 min),导致异常的原因均为错误。

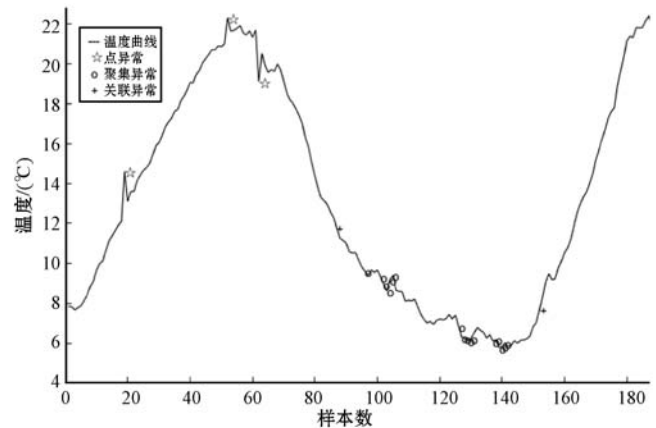


图 3 温度数据异常检测结果

实验时间效率  $T$  通过对具有相等属性,不同样本数的数据集 QY\_1、QY\_2、QY\_3 进行对比验证,对比结果如图 4 所示。DBSCAN 算法的运行时间增长速度最快,PO\_DBSCAN 与 ODSTC 算法的运行时间增长较慢,在数据集达到 8 440 小时,即数据点数达到 337 760 个时,PO\_DBSCAN 与 ODSTC 算法的运行时间少于 DBSCAN 算法的运行时间。结合前文对算法时间复杂度的分析可知,DBSCAN 算法时间复杂度为  $O(n^2)$ ,PO\_DBSCAN 与 ODSTC 的时间复杂度为  $O(n)$ 。当样本量增大到一定程度,PO\_DBSCAN 与 ODSTC 算法的时间效率较 DBSCAN 算法必然会有所提高。图中 PO\_DBSCAN 与 ODSTC 算法的运行时间差别较小,由此可知,KNN 算法时间复杂度较低。综上,ODSTC 可用于多传感器数据流的异常检测。

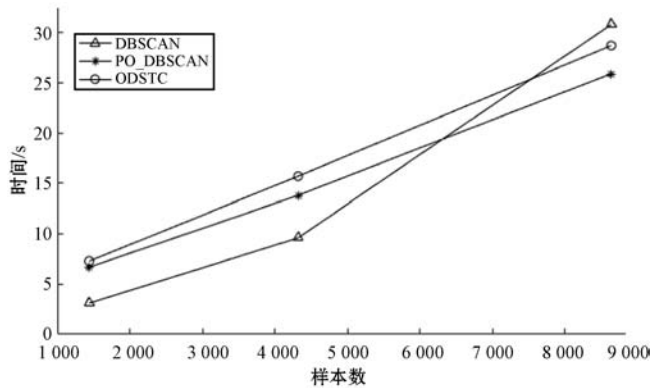


图4 异常检测算法时间效率对比

## 4 结 语

本文提出的时空相关性数据异常检测算法(ODSTC)结合了最近邻算法与聚类算法的思想,针对环境数据的特点对算法的参数进行优化,克服了最近邻距离阈值固定与聚类参数全局性及收敛速度慢等问题,提高了异常检测的效率。实验结果表明,ODSTC算法对多传感器数据流的检测准确率更高,算法的时间效率也会随着样本量的增加而提高。针对具有相关性的多传感器数据流,其效果满足现行环境的使用。从实验结果同样也可看出,当相关性较低的数据属性加入后,算法的效率有所降低,今后研究将进一步考虑不相关数据流的特征,提取频域特征,结合当地气候变化特征,从不同维度进行对照分析。

## 参 考 文 献

[1] 李道亮,杨昊. 农业物联网技术研究进展与发展趋势分析[J]. 农业机械学报,2018,49(1):8-27.

[2] Gil P, Martins H, Januário F. Outliers detection methods in wireless sensor networks[J]. Artificial Intelligence Review, 2019, 52: 2411-2436.

[3] Aggarwal CC. Proximity-based outlier detection[M]//Outlier Analysis. Springer International Publishing,2017: 111-147.

[4] Pan D W, Liu D T, Zhou J, et al. Anomaly detection for satellite power subsystem with associated rules based on kernel principal component analysis[J]. Microelectronics and Reliability, 2015, 55(9/10): 2082-2086.

[5] 余立苹,李云飞,朱世行. 基于高维数据流的异常检测算法[J]. 计算机工程,2018,44(1):51-55.

[6] 段青玲,肖晓琰,刘怡然,等. 基于SW-SVR的畜禽养殖物联网异常数据实时检测方法[J]. 农业机械学报,2017,48(8):159-165.

[7] 苑进,胡敏,Wang K S,等. 基于高斯过程建模的物联网数据不确定性度量与预测[J]. 农业机械学报,2015,46(5):265-272.

[8] Sarvani A, Venugopal B, Devarakonda N. Anomaly detection using K-means approach and outliers detection technique [C]//Soft Computing: Theories and Applications, 2018, 375-385.

[9] Yan M Y, Tong Q Y, Wang R M, et al. Outliers detection of cultivated land quality grade results based on spatial autocorrelation [C]//2016 fifth International Conference on Agro-Geoinformatics, 2016, 1-5.

[10] Shukla D S, Pandey A C, Kulhari A. Outlier detection: A survey on techniques of WSNs involving event and error based outliers [C]//2014 International Conference on Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity. IEEE, 2014:113-116.

[11] 冯振华,钱雪忠,赵娜娜. Greedy DBSCAN:一种针对多密度聚类的DBSCAN改进算法[J]. 计算机应用研究,2016, 33(9):2693-2696,2700.

[12] 侯晓晶,王会青,陈俊杰,等. 基于最近邻距离差的改进孤立点检测算法[J]. 计算机工程与设计,2013,34(4):1265-1269.

[13] Diao K J, Liang Y Q, Fan J C. An improved DBSCAN algorithm using local parameters [C]//2018 First CCF International Conference on Artificial Intelligence. Springer, 2018, 3-12.

[14] Tian H X, Liu X J, Han M. An outliers detection method of time series data for soft sensor modeling [C]//2016 Chinese Control and Decision Conference (CCDC), 2016: 3918-3922.

## (上接第46页)

[7] Leitner S, Yazdani M, Ziaei S, et al. Internal model-based active resonance damping current control of a grid-connected voltage-sourced converter with an LCL filter[J]. IEEE Transactions on Power Systems, 2018, 33(6): 6025-6036.

[8] 王华秋. 自构建小波神经网络的内模控制在深度脱硅中的应用[J]. 计算机应用与软件,2015,32(2):69-71.

[9] Singh J, Chatterjee K, Vishwakarma C B. Two degree of freedom internal model control-PID design for LFC of power systems via logarithmic approximations [J]. ISA Transactions, 2017, 72:185-196.

[10] 曹淦昆,李云峰. 结合模糊神经网络和粒子群优化的复杂装备关键备件用量预测[J]. 计算机应用与软件, 2014,31(10):167-171.

[11] 唐辉军,杨志民. 一种模糊最小二乘孪生支持向量回归机的改进算法[J]. 计算机应用与软件,2018,35(4):281-286.

[12] Dogdu G, Yalcuk A, Postalcioglu S. Application of the removal of pollutants from textile industry wastewater in constructed wetlands using fuzzy logic [J]. Environmental Technology, 2017, 38(4):443-455.