DOI: 10.19650/j.cnki.cjsi.J2209745

基于耦合自适应距离的高维异常检测算法*

周金洽,于劲松,宋 悦,梁思远
(北京航空航天大学自动化科学与电气工程学院 北京 100191)

摘 要:距离聚类方法是航天器等复杂系统实现遥测参数异常检测的常用方法之一,但在面对高维遥测数据进行异常检测任务时,往往会暴露出效率低下、精度劣化等严重问题。针对基于高维遥测数据的航天器异常检测难题,提出了一种基于耦合自适应的改进距离定义,并针对归纳监视系统(IMS)算法这一经典距离聚类算法进行了改进。该方法利用历史数据的分布特征,在进行聚类的同时,对于参数耦合性进行动态挖掘,并将挖掘到的知识高效地投入到异常检测任务。最后,采用运载火箭电源系统的真实高维遥测数据对所提方法进行了应用验证。在与多种传统基于 IMS 的异常检测方法的对比实验中,该改进算法检测效率与准确率较另两类 IMS 算法中的最优方法分别提升了 41.83% 和 69.03%,验证了运用该距离定义的检测方法在效率与精确率上的优越性。

关键词: 航天器;异常检测;高维数据;距离聚类;关联性挖掘
中图分类号: TH707 **文献标识码**: A **国家标准学科分类代码**: 460.4

High-dimensional anomaly detection algorithm based on coupling-adaptive distance

Zhou Jinhan, Yu Jinsong, Song Yue, Liang Siyuan
(*School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China*)

Abstract: The distance-based clustering is one of the common methods to realize the anomaly detection of telemetry parameters in complex systems, such as spacecraft. However, when it is applied to high-dimensional remote measurement data, it often exposes serious problems, such as low efficiency and degraded accuracy. To overcome the difficulty in anomaly detection on the high-dimensional telemetry data, this article proposes an improved distance definition based on coupling adaptation. The inductive monitoring system (IMS) algorithm which is a classical distance clustering algorithm is improved. Based on the intrinsic distribution characteristics of historical telemetry data, this method mines dynamically the couplings among telemetry parameters while clustering. Then, it takes efficiently advantage of this mined knowledge of telemetry parameters' couplings into the following task of anomaly detection. Finally, this article evaluates the application of the proposed method on a high-dimensional telemetry data of a real rocket power supply system. Compared with a variety of classic high-dimensional anomaly detection methods based on IMS algorithms, this article demonstrates its advantages for high-dimensional anomaly detection as well, which is 69.03% and 41.83% better than the best method in other two categories of IMS algorithms respectively on efficacy and accuracy of anomaly detection. It shows the superiority of the detection methods using the proposed distance definition in efficiency and accuracy.

Keywords: spacecraft; anomaly detection; high-dimensional data; distance-based clustering; correlation-mining

0 引 言

基于遥测参数的航天器异常检测,是根据遥测数据采取有效数据分析方法,实现航天器在轨运行状态地面监控的一项重要工作[1],对于航天器在轨运行的安全性与稳定性具有重要意义。随着航天器的内部结构日益复杂化,基于专家经验的传统检测方法局限性愈发明显,数

收稿日期:2022-05-07 Received Date: 2022-05-07
*基金项目:国家重点研发计划(2018YFB1403300)、国家自然科学基金(51875018)项目资助

据驱动方法因此被逐渐引入到了航天器的异常检测任务中,用以突破专家有限认知能力对传统方法检测能力的限制[2]。

而在众多的数据驱动方法中,基于距离聚类的检测方法不仅对指令响应不匹配等多参数耦合故障有一定的检测能力,也凭借其实现的便捷性与原理的可解释性,在当前航天器异常检测任务中具有较为广泛的应用。文献[3]基于放射性聚类方法建立了液体火箭发动机稳态过程的检测模型。文献[4]基于移动最小欧氏距离聚类在卫星雷达测高波形分类问题上获得了很好的效果。文献[5]基于 K 均值(K-means)算法这一经典的欧氏距离聚类方法,实现了对于绝缘栅双极型晶体管(insulated gate bipolar transistor, IGBT)的健康状态管理。美国航天局(NASA)于 2004 年提出的基于归纳式监视算法(inductive monitoring system, IMS)[6]则基于切比雪夫距离,通过增量式的自主学习方式突破了传统聚类方法在超参数配置、在线模型更新等方面的局限性,形成了一种成熟的增量式时序异常检测方法。

然而,随着航天器结构功能的快速演进,所涉及的遥测参数规模不断膨胀,参数间相关性日益复杂。以我国新一代空间站为例,核心舱监测参数规模已经超过了 30 000,建成后空间站系统实时监测参数总量会超过 100 000[7]。因此,面对遥测数据在参数层面上的维度膨胀趋势,异常检测算法也需要由低维数据向高维数据扩展。但在扩展的过程中,由于相关团划分不确定性上升、相关团内参数耦合关系模糊等问题,传统距离聚类算法及其现有改进版本在检测实时性与准确率等指标上出现严重的劣化问题,亟需基于高维异常检测的特征对距离聚类算法做出针对性的改进。

本文针对于高维数据耦合性复杂的关键特征,向传统距离聚类的距离定义中引入参数耦合性的自主感知机制,提出一种基于耦合自适应的距离定义,并以归纳式监视算法为代表,通过所提出的距离定义对这一经典的距离聚类方法进行面向高维异常检测任务的改进,从而实现该距离聚类算法在高维异常检测实时性和准确率等指标上的显著提高。

1　遥测参数耦合性的基本假设与应用

1.1　遥测参数耦合性的基本假设

为了更好刻画高维异常检测任务在遥测参数耦合性层面的特征,基于耦合自适应的改进距离定义首先需要从工程实际出发,明确遥测参数耦合性所满足的基本假设。而为了与满足交换律的传统相关性定义相区别,本文采用“耦合性”特指满足如下假设的一种有向关系。

假设 1　参数 A 对于参数 B 具有耦合性,则在参数 A 的不同取值下,参数 B 有确定的取值范围。假设来源于传统聚类基于距离的相似性度量方法[8]。若参数 A 对于参数 B 有耦合性,那么参数 A 应该能够与参数 B 本身的聚类形成映射关系,即能够对参数 B 的取值进行较好地划分。耦合性不是因果性。如参数 A 与参数 B 是所监视部件的一组同步控制指令,则虽然参数 A 与参数 B 间不存在因果性,但参数 A 对于参数 B 具有耦合性,参数 B 对于参数 A 具有耦合性。

假设 2　耦合性不满足交换律。聚类前提下的耦合性挖掘更多是针对于参数分布的相似性分析。而同一事件对于不同参数的影响力度是不一样的。如,参数 B 对应了事件 α 的响应,而参数 A 则对应了事件 α 和 β 的触发,则参数 A 因为反映了事件 α 的状态,对于参数 B 具有耦合性;而参数 B 因为不能反映事件 β 的状态,对于参数 A 不具有耦合性。这一理念在传统距离相似性度量中有所体现,诸如 KL(Kullback-Leibler)散度[9]等相似性指标也同样具有非对称性。

假设 3　参数频繁的同步复现性是参数耦合性的重要表征。耦合性是规律的,耦合参数表征出的时序变化也因此是大致同步的,具有较频繁的同步复现性。因此,遥测参数的耦合性可以通过统计遥测参数复现的频繁性而得到间接的评估。

假设 4　与某一参数同时复现的参数越少,这些参数对于该参数具有越确定的耦合性。实际的训练样本而历史在一段时间中的采样结果,其中不乏局部稳定的工况。因此,无关的参数组合也可能出现一定程度的复现性,而该复现性则不一定是由于参数间存在耦合关系导致的。因此,某一时刻同时复现的参数越多,越可能是处于复现的已归纳工况下,越可能存在包含无关的参数耦合信息,对于耦合性挖掘的参考价值也就越小;反之,同时复现的参数越少,存在无关参数耦合信息的可能性就越低,这些关系在耦合性挖掘的优先级也应该越大。

1.2　遥测参数耦合性的形式化

为了更方便地阐述基于耦合自适应的改进距离定义,本文预先将遥测参数耦合性及其涉及的概念形式化。记待处理的单帧数据(即训练中为尚未学习的数据、检测中则为待检测的数据)为 $\boldsymbol{X} \in \boldsymbol{R}^M$,$M$ 为待处理数据包含的参数规模,则有如下定义。

定义 1　簇群 $\boldsymbol{C} = \{\boldsymbol{C}_i\}_{i\in\{1,\cdots,N_c\}}$,历史数据通过聚类方法形成的 N_c 个簇的集合。各簇对应的上下限集合则分别为 $\boldsymbol{T}^u \in \boldsymbol{R}^{N_c\times M}$ 和 $\boldsymbol{T}^l \in \boldsymbol{R}^{N_c\times M}$。

定义 2　逐维距离矩阵 $\boldsymbol{D}(\boldsymbol{X},\boldsymbol{C}) \in \boldsymbol{R}^{N_c\times M}$,待处理数据各参数相对于簇群中各簇上下限的超限距离所形成的二维矩阵。

$$\boldsymbol{D}_i^{(j)}(\boldsymbol{X},\boldsymbol{C}) = \max(0,\boldsymbol{X}^{(j)} - \boldsymbol{T}_i^{u(j)},\boldsymbol{T}_i^{l(j)} - \boldsymbol{X}^{(j)}) \quad (1)$$

式中:矩阵的第 i 行代表数据 $\boldsymbol{X}$ 与簇 $\boldsymbol{C}_i$ 的关系,称为簇行,簇行第 j 个参数的分量表示为 $\boldsymbol{D}_i^{(j)}(\boldsymbol{X},\boldsymbol{C})$。

定义 3 同步复现性,待处理数据 $\boldsymbol{X}$ 中的部分参数同时出现在同一个已有簇 $\boldsymbol{C}_i$ 内,则称这些参数在 $\boldsymbol{X}$ 中出现了同步复现性。具有同步复现性的所有参数组合称为该簇行的关联组合 $\boldsymbol{I}$,关联组合规模称为该簇行的重合维度。

$$\boldsymbol{I}_i^{\text{train}} = \{j \in \{1,\cdots,M\} \mid \boldsymbol{D}_i^{(j)}(\boldsymbol{X},\boldsymbol{C}) = 0\} \tag{2}$$

$$\boldsymbol{I}_i^{\text{test}} = \{j \in \{1,\cdots,M\} \mid \boldsymbol{D}_i^{(j)}(\boldsymbol{X},\boldsymbol{C}) < T_{th}\} \tag{3}$$

式中:$\boldsymbol{I}^{\text{train}}$ 与 $\boldsymbol{I}^{\text{test}}$ 分别代表训练与检测阶段的关联组合;T_{th} 代表检测阶段正常数据超限距离的阈值。

定义 4 先验最小耦合维度 $N_R^* \in \boldsymbol{N}$,训练阶段针对参数耦合性层面的知识缺失,为训练样本中存在频繁复现性的最小重合维度设置的预估值。

定义 5 参数耦合性矩阵 $\hat{\boldsymbol{R}} \in \boldsymbol{R}^{M\times M}$,聚簇过程中利用参数同步复现性对参数耦合性进行动态估计而得到的经验耦合性矩阵。

定义 6 后验最小耦合维度 $\boldsymbol{N}_R \in \boldsymbol{R}^M$,表示对于每一参数具有耦合性的参数数目进行相应统计所得的估计值。由参数相关性矩阵得到:

$$\boldsymbol{N}_R^{(j)} = \sum_{j^*} \hat{\boldsymbol{R}}_{j,j^*} \tag{4}$$

该矢量将替代依赖于直觉的先验最小耦合维度,在检测过程中为不同参数提供更为适合的差异化最小重合维度指标。

定义 7 逐维距离矢量 $\boldsymbol{d}(\boldsymbol{X},\boldsymbol{C}) \in \boldsymbol{R}^M$,逐参数根据先验(训练阶段)或后验(检测阶段)最小耦合维度过滤逐维距离矩阵 $\boldsymbol{D}(\boldsymbol{X},\boldsymbol{C})$ 中重合度低的簇行后取对应参数列中最小值所组成的距离矢量。逐维距离矢量的切比雪夫距离即为待处理数据与当前簇群的点 - 簇群距离。

2 传统距离聚类方法的不足

距离聚类方法类型多样,在异常检测任务中的适用范围也各有差异,但本质上都依赖于点-簇群距离定义来判定待处理数据与现有簇群的具体关系。在训练过程中,距离聚类方法要求待训练数据在其指定的簇上下限定义下,只有当所有参数严格处于同一个簇内(即要求待训练数据在其指定的簇上下限定义下,与现有簇群的点-簇群距离为 0),才能载入现有簇群;而在检测过程中,检测阈值则在对数据超限状态量化分析的基础上,对于待检测数据的异常判读进行了松弛,而对数据超限状态的量化分析仍依赖于该点-簇群距离定义。因此,点-簇群距离定义是支持距离聚类方法训练与检测阶段的重要基础。

对于常见的距离聚类方法而言,点-簇群距离定义可形式化描述为:

$$d(\boldsymbol{X},\boldsymbol{C}) = \min_i \| \boldsymbol{D}_i(\boldsymbol{X},\boldsymbol{C}) \|_\infty \tag{5}$$

点—簇距离定义只注重了各数据纵向的时序分布关系,将数据的所有参数默认为一个同步变化的整体,而忽视了数据横向的各参数间存在强弱差异的耦合关系。因此,随着检测参数规模的进一步膨胀,高维参数内部的参数耦合关系则越发复杂,该点-簇群距离定义的不足也逐渐显现。

高维数据的参数一般能够按照其耦合性进一步分成更小的参数组合,这些内部耦合性强的参数组合之间则在时序上近似呈现出互相基本独立的分布。如图 1 所示,以航天器的多组推进装置为例,各组推进系统根据航天器变轨、入轨、转向等具体工况,各自采取对应的加速减速操作;而这些操作由不同指令控制,所涉及的参数间互相不存在直接的耦合关系。假设每组推进装置及其控制指令的簇群都只有分布代表加速与减速状态的两个簇。又因为这些参数在物理意义上均属于推进系统的同一类型组件,易于被归入同一个参数团。那么,随着参数团中涉及的独立推进装置组合增加,参数团表征也随着多个推进装置独立的加减速切换,呈现为多个耦合组合的独立变化,所有参数综合聚类所得到的簇群簇数因此呈指数增长特征。更一般地,如果各耦合参数组合独立聚类所得到的簇群都至少包含两个簇,则所有参数综合聚类所得到的簇群簇数 N_c 与耦合关系组合个数 N 的关系满足 $N_c = O(2^N)$。这一现象是归纳式监控算法在对高维数据进行归纳时,发生簇数爆炸问题的重要原因之一。

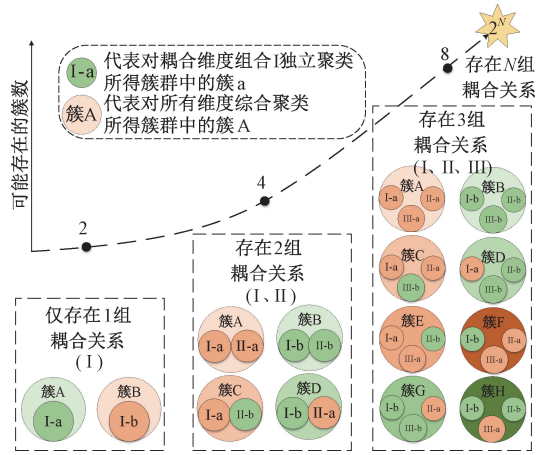


图 1 样本各参数耦合关系导致的簇数爆炸问题

Fig. 1 Cluster-number explosion caused by coupling between sampled telemetry parameters

而传统距离聚类方法也正因为点-簇群距离定义中对各参数间耦合关系的忽视,只能笼统地统计这些分布组合形成的联合分布,因此对联合分布间若干参数组合

的独立分布也设置了冗余的耦合性约束。而这些独立分布之间本身就在复杂的工况下具有时序层面的无关性,因此,该冗余约束不但对异常检测任务没有意义,更因为既要求训练样本包含无关组合在时序上联合出现的所有可能性、又要求算法记录无关组合这些可能出现的联合状态,使得传统距离聚类方法表现出聚类簇数虚高、检测精度下降等问题。

面对上述问题,传统的距离聚类算法亟需得到改进,特别是需要针对各参数间耦合关系的差异在算法设计上做出调整。最为直接的方法则是将参数按照统计学层面的关联关系进行再分组,即在构建距离聚类模型前,先对于输入的高维参数通过统计学的相关性指标进行进一步划分,使划分出的各参数组合的参数规模降低到适合传统距离聚类方法的水平。改进策略主要面临如下 4 个问题。

1) 进一步划分相关团的依据一般建立在某种统计学假设上,该假设不一定契合距离聚类方法内部的机理,不合理的参数关联划分方法可能导致部分对检测有意义的耦合关系因参数分散到不同参数组合中而丢失。

2) 海量的时序数据使得相关团划分过于低效,需要采取一定的数据预采样操作。直接截取某段子序列的方法忽视了其他序列蕴含的参数耦合性信息,因此相关团划分的实际效果较差;而基于时序分布进行的数据预采样也需要花费大量的时间,因此使得改进方法的训练效率低下。

3) 考虑到系统运行机理内部的复杂因果逻辑,参数间的影响不一定是互相的。即使将参数划分为不同的相关团,正向耦合关系有效但反向无效的参数组合仍然存在,影响距离聚类方法的准确率。

4) 实际训练数据不一定能覆盖系统的全生命周期,划分操作使得训练样本覆盖外的参数关联信息因为涉及参数已经分别处于不同参数组合内,在后续模型的检测和增强过程中不再能被距离聚类方法所自主感知并在对应的模型得到增量式的修正。

由此可见,基于参数关联再分组的改进方法未真正从传统距离聚类方法在高维参数检测任务的缺陷出发,并不能真正对传统方法进行有效的改进。因此,要真正在高维参数检测任务上实现对距离聚类方法的有效改进,必须从参数间的复杂耦合关系出发,凸显距离聚类方法在聚类过程中对于参数内部耦合性的自主感知潜能。具体而言,考虑到传统距离聚类方法对参数耦合性差异的忽视主要蕴含在点-簇群距离定义中,针对性的改进策略也应该从优化点-簇群距离定义出发,在距离定义内部引入对于参数耦合性的感知与权衡机制,进而支持各类距离聚类方法向高维参数异常检测任务的有效迁移。

3　基于耦合自适应的距离定义

基于耦合自适应的点-簇群距离的计算过程如算法 1 所示,其中 $|\cdot|$ 代表输入集合的元素个数。

算法 1　基于耦合自适应的点-簇群距离

输入:数据 $\boldsymbol{X}$,簇群 $\boldsymbol{C}$,参数耦合维度 $\boldsymbol{N}_{\boldsymbol{R}}^{*}$
参数:离簇阈值 T_{th}
输出:点-簇群距离 d

$\boldsymbol{D}=\boldsymbol{D}(\boldsymbol{X},\boldsymbol{C})$;
$\boldsymbol{d}=[T_{th},T_{th},\cdots,T_{th}]$; % 初始化距离矢量
$\boldsymbol{I}_i=\{j\,|\,\boldsymbol{D}_i^{(j)}<T_{th}\}$; % 计算各维度重合组合
FOR　$j=1:M$
　$\boldsymbol{J}=\{i\,|\,|\boldsymbol{I}_i|>\boldsymbol{N}_{\boldsymbol{R}}^{*(j)}\}$; % 滤除无效簇行
　if $|\boldsymbol{J}|\neq 0$, then
　　% 若 $\boldsymbol{X}$ 在参数 j 的分量在某簇内
　　$\boldsymbol{d}^{(j)}=\min_{i\in\boldsymbol{J}}\boldsymbol{D}_i^{(j)}$; % 更新参数 j 的分量
$d=\|\boldsymbol{d}\|_{\infty}$; % 根据距离矢量计算距离

训练阶段,该距离定义选取先验最小耦合维度作为各参数的参数耦合维度,根据待训练数据与现有簇群的点-簇群距离是否为 0 来判断现有簇群是否可以完整表征待训练数据。检测阶段,该距离定义选取后验最小耦合维度对各参数的参数耦合维度做出差异化的设置,从而通过点-簇群距离与检测阈值的比较判断待检测数据异常与否。

如图 2 所示,基于耦合自适应的点-簇群距离定义在合适的参数耦合维度下,能够有效削弱传统距离定义中无关参数耦合约束所引入的消极影响。

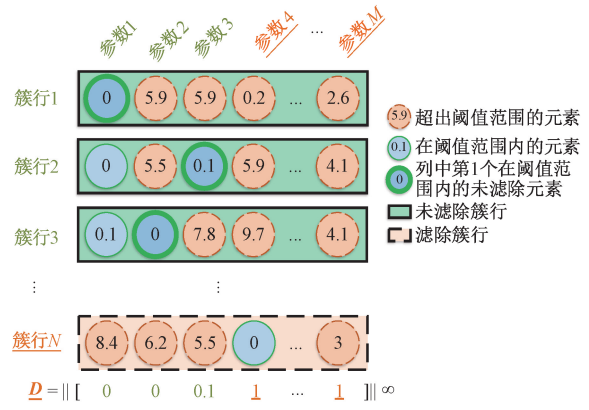


图 2　针对检测的点-簇群距离定义

Fig. 2　Point-cluster distance definition for detection

一方面,重合维度对于簇行的筛选,保证了筛选后的有效簇行能让簇中关联组合内的参数仍满足充分的耦合性约束。具体而言,如果参数对应的数据从未出现过,则所有簇行均为无效簇行,使数据在该参数维度上不能归入现有簇群中的任何簇中;如果参数对应的数据出现过,但不满足所假设的耦合关系,各簇行也会重合维度未超过参数耦合维度这一约束被视为无效簇行被滤除,使数据在该参数维度上也不能归入现有簇群中的任何簇中(图 2 中的参数 4 虽然本身在簇行 N 等簇行对应的参数上下限内,但该簇行并没有其他参数支持距离定义所需的重合维度约束,因此也被滤除,从而使得参数 4 最终并未归入任何簇中);只有参数所在的簇行中有充分多的其他参数支撑起距离定义所需的重合维度约束,数据才会在该参数上被归入该簇中,也只有当所有参数都能归入现有簇群的某一簇中,才能认为待处理数据属于现有簇群。这一机制一定程度上保证以联合控制指令为代表的参数耦合性在距离定义中能够得到有效体现,不降低距离聚类方法对于依赖多参数联合判别故障的敏感度。

另一方面,在筛除无效簇行后,所有剩余簇行的耦合性约束都通过具有充分规模的参数关联组合予以保证,不同参数因此可以自主归入不同的簇。这一机制有效削弱了传统距离定义由于要求所有参数归入同一簇而引入的无关参数耦合约束,使参数本身忽视耦合关系弱的参数,根据有效簇行自主归入最合适的关联参数组合,从而实现各参数对于本身参数耦合性的自主感知。

综上所述,相较于传统的点-簇群距离定义,在合适参数耦合维度的约束下,基于耦合自适应的点-簇群距离定义在计算过程中在保证参数间强耦合性约束被满足的情况下,消除对弱耦合性参数间的冗余耦合约束,使待处理数据与现有簇群在各参数维度上的距离计算更贴合参数间内秉的耦合性特征,进而弥补传统距离聚类算法在高维检测任务上由于忽视参数间耦合性差异而导致的簇数爆炸、精度恶化等问题。

4 基于耦合自适应距离的改进 IMS 算法

4.1 IMS 算法的优势

为了更直观地说明耦合自适应距离在实际异常检测的应用方法,本文将以 IMS 算法这一航天器异常检测领域较为经典的距离聚类方法为例,利用基于耦合自适应的点-簇群距离在高维检测任务上对该方法进行针对性的改进,进而更直观地介绍基于耦合自适应的点-簇群距离在具体距离聚类方法上的应用方法。

IMS 算法是一种面向时间序列设计的自主聚类方法。与以邻近算法(K-nearest-neighbor, KNN)[10]、谱密度聚类算法(density-based spatial clustering of applications with noise, DBSCAN)[11-12]等其他传统距离聚类方法不同,该聚类方法摆脱了对于先验簇数估计的依赖,更倾向于在学习过程中按照时间顺序增量地对数据进行分簇,更适合监视对应具体系统操作任务的时间序列,尤其对当前航天器复杂控制下异常检测的多工况问题具有良好的适用性。同时,增量式的学习机制也使该聚类方法在不遗忘已有知识的前提下,实现基于未知工况数据的自我修正,从而在不妨碍已知工况检测精度的前提下,扩充未知工况的检测机理,进而满足航天器在实际运行任务中的全生命周期异常检测需求。

在工程中应用方面,基于 IMS 算法的异常检测方法易于实现,运算复杂度低,检测效果好,逐渐成为了支撑航天领域各类异常检测任务的一项成熟技术。NASA 的詹森航天中心(Johnson space center, JSC)利用战神一号-X 火箭(Ares I-X)[13]等项目验证了 IMS 算法的有效性,并将 IMS 算法配备于国际空间站姿态控制系统与热控制系统的长期运行检测任务中[14];美国空军研究实验室则与 NASA 联合开发了嵌入 IMS 算法的飞行器综合诊断系统,并将其应用于实际的飞行检测任务中[15]。

4.2 IMS 算法的基本原理

IMS 算法[16]主要涉及的超参数有如下 4 个:1)初始簇半径 T_c 代表创建新簇时,簇上下限的初始值;2)簇增长百分比 K_c 代表算法用于判别待处理数据能否触发簇群中某簇扩张的松弛上下限(即扩张上下限)相对于对应簇本身上下限的松弛比例;3)簇膨胀系数 K_e 代表簇根据某帧数据进行扩张时,上下限根据各参数维度上超限情况所扩张的具体程度;4)检测阈值 T_{th} 代表检测时正常数据在扩张上下限判别下,点-簇群距离的上限。与其他基于自主学习机制的算法一样,该算法也分为线下训练和线上检测两个阶段。

在训练阶段,该算法通过对正常历史数据逐条进行增量式自主学习,以簇群的形式归纳并记录正常数据的大致分布。若待处理数据在现有簇群中某一簇的上下限内,则不更新该簇;若待测数据超出某一簇上下限,但所有维度超限部分的相对比例不超过 K_c(即待测数据仍处于该簇扩张上下限内),则在各参数维度上按照超出的比例以 K_e 单边扩张该簇的上下限;否则,以该数据为中心建立一个新簇,并通过 T_c 初始化该簇的上下限。

在检测阶段,该算法将待检测数据与训练得到的簇群进行比较,根据比较得到的点-簇群距离评估待测数据与正常分布的贴合程度,进而根据检测阈值 T_{th} 判别待检测数据的异常与否。

4.3 基于耦合自适应距离的改进 IMS 算法

参数间耦合关系的一个重要表征是耦合参数在时序数据上频繁的同时复现性。耦合关系不明显的参数则会

在多样的运行模式下逐渐丧失同时复现的频繁性。因此,根据该特征在 IMS 算法中引入耦合性机制是提高 IMS 算法在高维检测任务中准确率的关键。

检测过程与训练过程中运用参数耦合性的条件和目的存在差异。训练过程中,只能通过大致估计得到的先验最小耦合维度对参数耦合性进行较为模糊的描述,但训练数据的置信度高,蕴含大量可挖掘的参数耦合信息;因此,该阶段的目的主要是实现集约度高的聚类并挖掘置信度高的参数耦合性矩阵。检测过程中,算法已经挖掘出了参数耦合性矩阵,能够对各参数所涉及的耦合关系做出更为明确的刻画,其目的也因此转为利用参数耦合性矩阵,指导算法在各参数维度上完成对待检测数据与现有簇群更适合的距离评估。因此,两阶段中点-簇群距离的具体应用方法也需要针对这一区别,进行差异化设计。

基于以上特征,本文通过基于耦合自适应的点-簇群距离,分别在 IMS 算法的训练与检测机制上做出了具体的改进。

1)改进 IMS 算法的训练机制

在训练过程中,改进的 IMS 算法除利用基于耦合自适应的点-簇群距离,改写用于判断数据入簇与否的传统点-簇群距离定义外,还在增量式聚类过程中通过动态更新的耦合性矩阵,量化统计不同参数同时复现的频繁性,从而真正发挥出该算法聚类过程潜在的耦合性感知能力。

如图 3 所示,算法在根据第 1 帧训练数据构建初始簇群、并以单位阵初始化耦合性矩阵后,开始逐帧载入后续训练数据,增量式地更新由簇群与耦合性矩阵组成的模型。

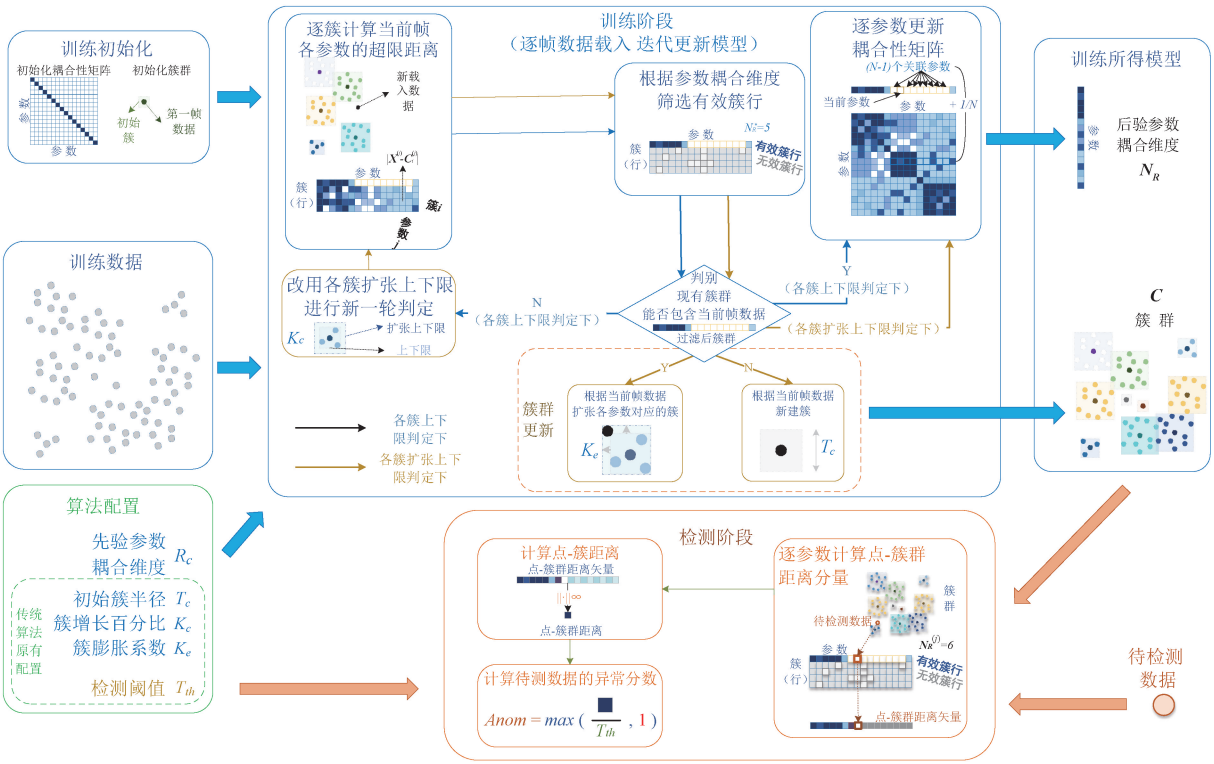


图 3　基于耦合自适应距离的 IMS 算法原理

Fig. 3　Principle of IMS algorithm based on coupling-adaptive distance

在增量更新模型的过程中,每载入一帧训练数据,该算即采用定义 7 的点-簇群距离方式代替传统距离定义,从而判别该帧数据在上下限(或扩张上下限)的判定下是否包含于现有簇群中。(1)对于不同情况,新增数据仍会与传统 IMS 算法一样对现有簇群做出扩张或新建簇等适合的簇群更新操作;(2)在改进算法独有的耦合性挖掘机制下,该算法还会根据当前帧数据更新耦合性矩阵,即:该算法逐参数寻找当前参数在上下限(或扩张上下限)内且重合维度最小的有效簇行,并将此次总量为 1 的增益权重平均分配给该簇行重合维度中包括参数本身在内的所有关联参数,从而在当前参数所在的耦合性矩阵行上更新对应参数的权重。

在增量更新结束后,该算法得到的则是较传统算法结果而言更为精简的簇群和统计性质的耦合性矩阵。为进一步得到反映参数间耦合性(取值在 0~1)的实际耦合性矩阵,需要利用参数与自身耦合性为 1 这一性质,将耦合性矩阵修正为:

$$\hat{\boldsymbol{R}}_{j,j^*} = \frac{\boldsymbol{R}_{j,j^*}}{\boldsymbol{R}_{j,j}} \tag{6}$$

根据定义 6 估算出比先验最小耦合维度 N_R^* 更符合参数实际特征的后验最小耦合维度矢量 $\boldsymbol{N}_{\boldsymbol{R}}$。簇群与后验最小耦合维度则组成了该算法训练的最终模型,从而支撑起后续的检测任务。

改进的训练机制主要具有两方面优势:(1)在先验最小相关维度的松弛耦合约束和下,数据各参数都尽可能找到适合自己的簇,从而挖掘分布在现有簇群不同簇中的若干耦合参数组合,摆脱不同组合间彼此的弱耦合性在传统算法中引入的无效耦合约束,最终联合表征出当前帧数据,提高了同等规模簇群的数据表征能力,抑制了该类数据在无效耦合限制下用现有簇群所表征而导致的簇数爆炸风险;(2)耦合性矩阵参数较好地利用耦合参数复现性更为频繁这一特征,通过平均分配权重与选择最小重合维度的方法在增量过程中逐渐凸显耦合性较强的参数耦合关系,从而使算法进一步通过数据本身得出了较人工经验的先验最小耦合维度更为适合的后验最小耦合维度,既对不同的参数形成更个性化的衡量标准,更使得后续的检测机制与归纳式监控算法所内秉的聚类机理更为契合。

2)改进 IMS 算法的检测机制

在检测阶段,根据训练阶段得到的簇群和后验最小耦合维度矢量 $\boldsymbol{N}_{\boldsymbol{R}}$,该算法利用基于耦合自适应的点-簇群距离则可以直接计算实时数据距现有簇群的超限距离,从而根据检测阈值 T_{th} 快速判别对应部件的异常与否,并进一步量化评估出反应异常演化过程的异常分数。其中,异常分数为 0,则代表数据未超限,即对应部件完全正常;异常分数为 1,则代表数据超限距离已经超出检测阈值,即对应部件发生严重故障;异常分数处于 0~1 时,则代表数据超限但超限距离为超出检测阈值,即对应部件正处于从正常向严重故障逐渐演化的过程中。

改进的检测机制通过基于耦合自适应的点-簇群距离,发挥簇群较传统模式更为强大的表征能力,一方面有效规避点-簇群距离计算过程中无关参数间无效耦合约束所导致的虚警偏高问题,从而提升算法在实际训练样本有限背景下的检测准确率;另一方面则通过较传统模式更为精简的簇群,显著提高了检测效率,从而真正满足实际异常检测任务的实时性需求。

5 实验验证

5.1 实验设置

实验采用运载火箭电源系统的实际运行数据集[17]进行实验验证,数据集包含状态控制指令、母线电压电流、支路电流与定期自检结果在内的 242 个参数维度。其中,定期自检结果属于时钟编码,用于校准系统的时间同步,与其他参数的耦合性较弱;母线电压电流、支路电流设计两条母线,每条母线各具有 4 支路的电路;状态控制指令则统一控制母线所在的电路,以实现不同工况的切换。

实验将完整数据集按器上时进行排序、采样、拼接等预处理操作,形成同一连续时段下的 15 000 帧数据,共涵盖不同指令下的 6 个运行工况。该数据集被均匀划分为包含 10 000 帧数据的训练集与 5 000 帧测试集。训练集与测试集无时间重合数据,均覆盖了原有数据集所包含的完整火箭发射过程。其中,在测试集中,实验根据由专家门限与工况描述信息组成的参数异常判据表对火箭电源系统多类型故障进行模拟,共注入分散在 15 个连续时段下的 696 帧异常,涵盖单帧参数独立超限异常、单帧多参数工况异步异常、单帧参数剧烈扰动异常等常见故障类型。

为更明显地体现改进 IMS 算法的优势,实验针对 IMS 算法在高维数据异常检测任务中的效率低下和精度劣化问题,设计了多元检测指标。检测指标包含使用效率和检测精度两个部分。使用效率主要由训练速率与检测速率两个直接指标进行衡量,训练速率反映了算法在从获取历史数据到正式上线过程中所耗费的时间成本,而检测效率则反映了算法在线上检测过程中的实时性能力;检测精度则分为精确率 P、召回率 R、准确率 A 3 个指标,分别反映了算法在检测过程中的漏检概率、虚警概率和推理错误概率。

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$A = \frac{TP}{TP + FP + FN + TN} \tag{9}$$

式中:TP 为被检测为正常且标签为正常的样本数目;TN 为被检测为异常且标签为异常的样本数目;FP 为被检测为正常且标签为异常的样本数目;FN 为被检测为异常且标签为正常的样本数目。

5.2 异常检测结果分析

1)检测指标分析

对实验分别通过传统 IMS 算法、基于参数关联再分组的改进 IMS 方法和基于耦合自适应距离的改进 IMS 算法这 3 类异常检测方法对于火箭电源系统运行数据进行了异常检测。其中,考虑到统计学相关系数的多样性,实验选取皮尔森(Pearson)相关系数、斯皮尔曼(Spearman)相关系数、肯德尔秩(Kendall Rank)相关系数、最大信息系数(maximal information coefficient, MIC)[18-21]这 4 类常

见统计相关系数,以 0.6 作为相关性阈值对参数进行划分,进而构建对应的 4 种多 IMS 算法,实现 4 种较有代表性的基于参数关联再分组的改进 IMS 方法,为对比试验提供有效的参考。实验结果如表 1 所示。

表 1　实验结果

Table 1　Results of experiment

算法	使用效率指标			检测精度指标		
	聚类簇数(簇)	每帧训练用时/ms	每帧检测测用时/ms	精确率	召回率	准确率
IMS	10 001	26.178	78.356	1.000	0.000	0.140
多 IMS(MIC)	10 397(110 个 IMS)	205.142	40.333	1.000	0.022	0.158
多 IMS(Pearson)	10 517 (100 个 IMS)	40.688	30.694	0.998	0.122	0.244
多 IMS(Spearman)	10 291 (101 个 IMS)	24.282	16.368	0.996	0.496	0.564
多 IMS(Kendall)	9 540(103 个 IMS)	22.180	18.100	1.000	0.525	0.591
改进 IMS	218	2.954	10.529	1.000	0.999	0.999

由实验结果可知,基于传统 IMS 的检测方法精确率为 100%,而召回率却几乎为 0,反映了传统 IMS 算法在高维数据异常检测下的高虚警率。具体而言,在传统点-簇群距离定义下,高维参数组合内部大量弱耦合关系的参数使得传统 IMS 算法的聚簇统计包含了与参数耦合关系不相符的冗余耦合关系,严重影响了簇群对于待处理处理的表征能力,进而影响该算法对各参数异常分数的正确评估。

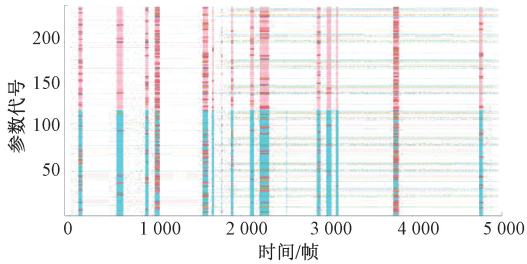
而 4 种基于相关性指标的多 IMS 检测方法都在维持传统 IMS 算法高精确率的前提下,在召回率上有所提升,一定程度上反映了无效耦合关系对于 IMS 算法等距离聚类方法的实际消极影响。同时,不同相关性指标下召回率的提升程度也存在着明显区别。基于 MIC 指标与基于 Pearson 指标的多 IMS 检测方法召回率分别仅有 2.2% 和 12.2%,虽略优于传统 IMS 方法,但虚警问题仍然严重。基于 Spearman 指标的多 IMS 检测方法召回率到达 49.6%,相较于传统 IMS 方法有了显著提升,但是在精确率上却下降到 99.6%,出现了漏检现象。基于 Kendall 指标的多 IMS 检测方法效果最为理想,不仅保持了 100% 的精确率,也将召回率提升到了 52.5%。这些方法均通过细化参数之间的关联组合划分,虽然一定意义上能够减少异常分数评估过程中冗余的无效耦合性约束以提升检测效果,但一方面,对原参数组合的进一步划分也不可避免地破坏原有参数组合内部的部分有效耦合性约束,因此在检测精确率方面也有性能恶化的风险;另一方面,这种划分方法只从基础的统计学假设出发,未真正切合距离聚类方法对耦合性挖掘的内秉需求,因此划分后仍存在无效耦合性约束,因此对召回率的提升也存在明显的瓶颈。

本文利用基于耦合自感知的点-簇群距离而改进出的 IMS 算法,精确率维持在 100% 的理想水平上,召回率也同时上升到 99.9%。该改进 IMS 算法真正从 IMS 算法原理出发,针对性地制定了参数耦合性的自主感知机制,在保留参数间有效耦合性约束的同时,有效滤除了聚类中无效耦合性约束对于检测效果的消极影响,因此在精确率与召回率等评价指标上远优于其他类型的 IMS 算法。

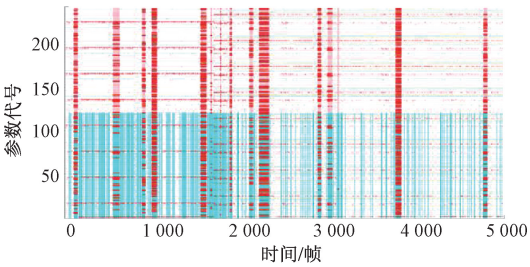
同时,就各检测方法的效率而言,传统 IMS 算法由于无效耦合约束导致簇群对于待处理数据的表征能力不理想,最终使得簇群规模快速膨胀,极大影响了训练与检测效率;基于参数关联再分组的改进 IMS 方法则由于关联性划分过细,产生的更小参数组合导致了对应的 IMS 检测模型数量过多,使得最终的多 IMS 算法的簇群总规模仍然维传统方法同样的水平;而本文所提出的改进 IMS 算法则依赖于基于耦合自适应的点-簇群距离定义,有效提高了簇群对待检测数据的表征能力,最终仅生成 218 个簇,较传统 IMS 算法的 10 001 个簇有明显的优化,与基于参数关联再分组的改进 IMS 方法所需要的 100 个左右的 IMS 检测模型相比也更显精简,进而使得改进 IMS 算法在训练速率与检测速率上明显优于其他 IMS 算法。

为进一步验证本文提出的改进 IMS 算法的异常检测效果,本文对全部 242 维遥测参数的异常检测结果进行可视化分析。改进 IMS 算法与基于 Kendall 相关性指标的多 IMS 算法对高维数据的检测结果如图 4 所示。图 4 所有参数均进行了归一化处理,检测样本标签标记的异常区域在图 4(a)由红色区域标记,检测算法检测出的异常区域在图 4(b)由蓝色区域标记,具体参数异常则由红色圆点标记。

从实验结果可以看出,基于 Kendall 的相关性指标在前 2 000 帧的范围内虚警密集,在后 3 000 帧内仍然有间歇性的虚警,而虚警中的异常参数组合均都基本呈现为几个固定的参数组合,且组合内的参数在虚警段的波形也大致相同。这一现象也同样出现在传统 IMS 算法和基于其他相关性指标的多 IMS 算法中,很好地体现了各



(a) 改进IMS算法
(a) Improved IMS algorithm



(b) 基于Kendall的多IMS算法
(b) Multi-IMS algorithm based on Kendall

图 4 改进 IMS 算法与基于 Kendall 相关性指标的多 IMS 算法的异常检测结果

Fig. 4 Result of anomaly detection respectively by the improved IMS algorithm and multi-IMS algorithm based on Kendall rank correlation coefficient

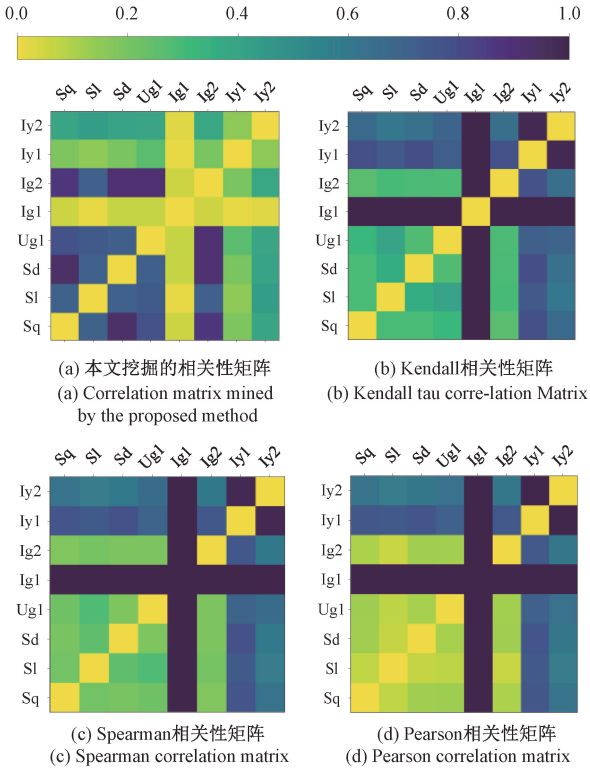
IMS 检测模型在传统点-簇群距离定义下,参数组合内部残存的无效耦合约束对于检测结果的消极影响。而利用基于耦合自适应的点-簇群距离而改进的 IMS 算法则通过距离定义内引入的参数耦合性自主感知机制,有效规避了这些虚警,进而使得该方法检测出的异常区域与样本标签标记的异常区域基本重合,促成该改进 IMS 算法在异常检测准确率上的优越性。

2)参数耦合性挖掘结果分析

利用基于耦合自适应的点-簇群距离而改进的 IMS 算法与传统统计学相关性指标各自挖掘出的部分参数相关性结果如图 5 所示。其中,为了与其他指标下的对称相关性矩阵形成直观的对比,本文所提出的改进 IMS 算法也对于挖掘出的耦合性矩阵进行对称化处理,即取各参数对双向耦合性系数的较大值作为参数对的相关性系数,从而形成相关性矩阵 $\hat{\boldsymbol{R}}^*$:

$$\hat{\boldsymbol{R}}^*_{j,j*} = \hat{\boldsymbol{R}}^*_{j^*,j} = \max(\hat{\boldsymbol{R}}_{j,j*}, \hat{\boldsymbol{R}}_{j^*,j}) \tag{10}$$

如图 5 所示,相关性矩阵不同参数对之间按照相关性高低进行标色。相关性为 1 时,矩阵对应元素呈最亮的淡黄色;相关性为 0 时,矩阵元素呈最暗的深蓝色。涉及的 8 个参数的物理意义分别为自检时间编码 Sq 与 Sl,母线 1 转电状态控制指令 Sd,母线 1 电压 Ug1,母线 1 电流 Ig1,母线 2 电流 Ig2,母线 1 下两条支路的电流 Iy1 与 Iy2。因此,与电路无关的参数 Sq 和 Sl 应该与其余参数的相关性微弱;唯一与母线 2 相关的参数 Ig2 和 Ug1 等与母线 1 相关的参数因为电路连接存在一定的相关性,但相关性弱于母线 1 内部参数间的相关性;母线电流 Ig1 与其支路电流 Iy1 和 Iy2 的相关性应该最大。



(a) 本文挖掘的相关性矩阵
(a) Correlation matrix mined by the proposed method

(b) Kendall相关性矩阵
(b) Kendall tau corre-lation Matrix

(c) Spearman相关性矩阵
(c) Spearman correlation matrix

(d) Pearson相关性矩阵
(d) Pearson correlation matrix

图 5 本文所提出改进 IMS 算法挖掘的相关性矩阵与传统指标下的部分参数相关性矩阵

Fig. 5 Correlation matrix mined respectively by the proposed improved IMS algorithm and by some classic correlation coefficients

结合物理意义对比图 5 的相关性矩阵,可以发现:本文所提出的改进 IMS 算法与传统统计学相关性指标在重要参数组合的相关性判别上具有明显的差异性,且前者挖掘出地相关性矩阵更符合上述物理层面的描述。

对于自检时间编码 Sq 与 Sl 而言,由于这些参数实际只存在 0 和 1 两个离散值,且编码中 1 的比例较少,在传统指标看来其与大部分参数均容易满足统计学层面的相关性假设,从而建立这些参数间的无效相关性,而本文提出的改进 IMS 算法则从参数同步复现性的加权统计出发,使得离散的指令量与连续物理量的时序不同步现象得到充分的感知,从而较为有效地筛去这一类无关信息。而对于参数 Iq2 而言,由于电路的连接,不同母线的电压电流存在一定的线性关系,从而在传统指标下满足相关性假设;但是在时序同步性上,由于母线间存在一定程度的独立性,不同母线的同步性具有一定的差异,本文所提出的改进 IMS 算法因此可以筛除到 Ug1 与 Ig2 的冗余耦

合性。同理,对于 Iy1 和 Iy2 而言,本文所提出的改进 IMS 算法也可以准确感知到 Ig1 和 Ig2 对于这些支路参数的耦合性强弱。

综合上述分析,改进的 IMS 方法对于参数间的耦合关系更为敏锐,能够很好地解决传统 IMS 方法在高维数据异常检测中检测精度与检测效率上的不足,实现对于高维数据异常的实时准确检测。

6 结　论

本文针对于距离聚类方法在高维检测任务中面临的新挑战,提出一种基于耦合自适应的点-簇群距离定义。该距离定义立足于高维检测任务背后参数的复杂耦合关系,实现聚类距离评估中对于参数耦合性的自主感知,有效地弥补传统距离聚类算法在高维异常检测中的效率低下、虚警频繁等问题,进而实现以 IMS 算法为代表的距离聚类方法在高维数据异常检测的应用。

本文的研究仍存在进一步的研究空间。传统距离聚类方法在高维异常检测存在局限性外,对于故障机理的时域分析也不够充分。此外,基于耦合自适应的距离定义能够使距离聚类方法同样通过数据在时间维度上的窗口化处理实现兼容时序异常的异常检测[22],但针对于窗口中参数不同帧间耦合性的特殊性质,仍需要进一步对耦合自适应机制进行相应调整。

参考文献

[1] 彭喜元, 庞景月, 彭宇, 等. 航天器遥测数据异常检测综述[J]. 仪器仪表学报, 2016, 37(9): 1929-1945.
PENG X Y, PANG J Y, PENG Y, et al. Review on anomaly detection of spacecraft telemetry data[J]. Chinese Journal of Scientific Instrument, 2016, 37(9): 1929-1945.

[2] 沈毅, 李利亮, 王振华. 航天器故障诊断与容错控制技术研究综述[J]. 宇航学报, 2020, 41(6): 647-656.
SHEN Y, LI L L, WANG ZH H. A review of fault diagnosis and fault-tolerant control techniques for spacecraft[J]. Journal of Astronautics, 2020, 41(6): 647-656.

[3] 张翔, 徐洪平, 安雪岩,等. 基于聚类分析的液体火箭发动机稳态过程故障程度评估方法[J]. 导弹与航天运载技术, 2015(4): 24-26, 35.
ZHANG X, XU H P, AN X Y, et al. Discussion on the fault level evaluation method for the steady process of liquid propellant rocket engine based on cluster analysis[J]. Missiles and Space Vehicles, 2015(4): 24-26, 35.

[4] 汪海洪, 岳迎春, 邹贤才,等. 基于聚类分析的卫星雷达测高波形分类研究[J]. 武汉大学学报(信息科学版), 2010,35(7):833-836.
WANG H H, YU Y CH, ZOU X C, et al. Classification of radar altimeter waveforms based on cluster analysis[J]. Geomatics and Information Science of Wuhan University. 2010,35(7):833-836.

[5] 王志远, 孙鹏菊, 王海波,等. 基于聚类分类算法的IGBT 健康状态分类研究[J]. 电工电能新技术, 2021, 40(11):1-8.
WANG ZH Y, SUN P J, WANG H B, et al. Research on IGBT state classification based on cluster and classification algorithm[J]. Advanced Technology if Electrical Engineering and Energy, 2021, 40(11):1-8.

[6] IVERSON D, MARTIN R, SCHWABACHER M, et al. General purpose data-driven system monitoring for space operations[J]. Journal of Aerospace Computing, Information, and Communication, 2012, 9(2): 26-44.

[7] 李瑞雪, 张泽旭. 数据驱动的航天器异常检测工具对未来中国空间站管理的启示[J]. 载人航天, 2021, 27(2): 244-251.
LI R X, ZHANG Z X. Enlightenment of Data-driven Spacecraft Anomaly Detection Tools on Future Space Station Management in China[J]. Manned Space-flight, translated. 2021, 27(2): 244-251.

[8] CHAKRABARTI S. Similarity and clustering[J]. Mining the Web, 2003, DOI: 10.1016/B978-155860754-5/50005-7.

[9] POLANI D. Kullback-Leibler Divergence[M]. Newyork: Springer, 2013.

[10] YU L, DING W. A KNNS based anomaly detection method applied for UAV flight data stream[C]. Prognostics & System Health Management Conference, IEEE, 2016.

[11] THANG T M, KIM J. The anomaly detection by using DBSCAN clustering with multiple parameters[C]. IEEE, 2011:1-5.

[12] BOONCHOO T, XIANG A, HE Q. An efficient density-based clustering algorithm for higher-dimensional data[J]. Computer Science, 2018, DOI: 10.48550/

arXiv. 1801. 06965.

[13] SCHWABACHER M, MARTIN R, WATERMAN R, et al. Ares I-X ground diagnostic prototype[C]. AIAA Infotech Aerospace, 2010.

[14] IVERSON D, MARTIN R, SCHWABACHER M, et al. General purpose data-driven system monitoring for space operations [J]. Journal of Aerospace Computing, Information, and Communication, 2012, 9(2): 26-44.

[15] FRANK J, AASENG G B. Transitioning autonomous systems technology research to a flight software environment[C]. AIAA SPACE Forum, 2016.

[16] IVERSON D L, MARTIN R, SCHWABACHER M, et al. General purpose data-driven monitoring for space operations [J]. Journal of Aerospace Computing Information & Communication, 2012, 9(2):26-44.

[17] YU J, SONG Y, TANG D, et al. Telemetry data-based spacecraft anomaly detection with spatial-temporal generative adversarial networks[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-9.

[18] QIAO Z, HE J, CAO J, et al. Multiple time series anomaly detection based on compression and correlation analysis: a medical surveillance case study[C]. Asia-Pacific International Conference on Web Technologies and Applications, 2012:294-305.

[19] HAUKE J, KOSSOWSKI T. Comparison of values of pearson's and spearman's correlation coefficients on the same sets of data[J]. Quaestiones Geographicae, 2011, 30(2):87-93.

[20] BAUMGARTNER R, SOMORJAI R, SUMMERS R, et al. Assessment of cluster homogeneity in fMRI data using Kendall' s coefficient of concordance. [J]. Magnetic Resonance Imaging, 1999, 17 (10): 1525-1532.

[21] 孙广路, 宋智超, 刘金来,等. 基于最大信息系数和近似马尔科夫毯的特征选择方法[J]. 自动化学报, 2017, 43(5): 795-805.
SUN G L, SONG ZH CH, LIU J L, et al. Feature selection method based on maximum information coefficient and approximate Markov blanket[J]. Acta Automatica Sinica, 2017, 43(5): 795-805.

[22] KEOGH E, LIN J. Clustering of time-series subsequences is meaningless: Implications for previous and future research [J]. Knowledge & Information Systems, 2005, 8(2):154-177.

作者简介

周金洽,2020 年于北京航空航天大学获得学士学位,现为北京航空航天大学研究生,主要研究方向为自动化测试和故障预测与健康管理技术。
E-mail: jinhan_zhou@buaa.edu.cn

Zhou Jinhan received his B.Sc. degree from Beihang University in 2020. He is currently a M.Sc. candidate at Beihang University. His main research interests include automatic testing and prognostic and health management.

于劲松(通信作者),2004 年于北京航空航天大学获得博士学位,现为北京航空航天大学教授,主要研究方向为自动化测试和故障预测与健康管理技术。
E-mail: yujs@buaa.edu.cn

Yu Jinsong (Corresponding author) received his Ph.D. degree from Beihang University in 2004. He is currently a professor at Beihang University. His main research interests include automatic testing and prognostic and health management.