

基于误差序列的离线数据无监督快速异常检测方法研究

谷阳阳, 张东

(中国人民解放军 92941 部队 45 分队, 辽宁 葫芦岛 125000)

摘要: 针对利用多维度、多模态的遥测参量实现快速定位检测故障需求的问题, 在分析和归纳飞行器遥测参量特点及故障检测方法的基础上, 研究了基于误差序列的离线数据无监督快速异常检测方法; 该方法改变了传统的主要依靠人工分析进行故障判定模式, 结合历史数据的使用, 通过无监督自适应阈值的设定, 实现自动适配并检测遥测参量异常记录状态的功能, 降低了需要大量人员分析的时间成本, 减少了遂行处理工作中快速定位故障的压力, 解决了故障定位通用性不强的问题; 经过实际应用, 效果良好, 满足数据处理要求, 具有较好的自适应性, 易于实现。

关键词: 遥测数据处理; 自适应阈值; 误差序列; 自动检测; 数据驱动方法

Research on Unsupervised Fast Anomaly Detection for Off-line Data Based on Error Sequences

GU Yangyang, ZHANG Dong

(Sub unit 45, Unit 92941, PLA, Huludao 125000, China)

Abstract: To address the requirements for rapid fault location and detection through multi-dimensional and multi-modal telemetry parameters, by analyzing and summarizing the characteristics of aircraft telemetry parameters and fault detection methods, research on an unsupervised fast anomaly detection method for offline data based on error sequences is conducted. This method transforms the traditional approach that primarily relies on manual fault analysis and judgment. By incorporating historical data and setting an unsupervised adaptive threshold, it achieves automatic adaptation and detects the abnormal recording state of the telemetry parameter. This reduces the time required for a large number of personnel to analyze, achieves the rapid fault positioning pressure during parallel processing, and solves weak generality in fault locations. Practical application meets the data processing requirements with good adaptability and easy implementation.

Keywords: telemetry data processing; adaptive threshold; error sequences; automatic detection; data-driven method

0 引言

飞行器在飞行过程中会产生大量的遥测数据, 地面遥测系统和飞行器端通过遥测数据的形式进行沟通, 包括接收飞行器端发出的下行遥测数据和发送处理指令的上行遥测数据, 通过对这些数据的分析可以了解和掌握飞行器的运行和设备状况。随着计算机算力、存力和运力大幅提升, 飞行器遥测测量数据的体量不断增大, 针对这些海量测量数据的相关信息处理及挖掘, 如果还是依靠传统的人工分析、处理方法, 已远远无法满足快速量化分析、支撑关联决策的需求。同时, 立足装备信息

化、智能化的趋势, 需要我们针对飞行器故障检测的处理及分析模式也随之向高效、灵活、敏捷发展。异常检测(Anomaly Detection)技术在数据挖掘领域无论在理论层面还是应用方面都已经发展得极为成熟, 其定义为在数据中识别不符合预期模式或与数据集中其他项目显著不同的观测值的过程^[1-2]。在实际系统中, 异常数据的产生是对故障的描述, 故障的发生伴随着异常数据的产生, 因此可使用数据的异常检测的方法和理论研究故障检测技术^[3-4]。

随着人工智能领域的发展, 如何在故障检测领域有效的利用数据挖掘技术协助处理人员进行数据分析与故

收稿日期:2024-12-12; 修回日期:2025-02-13。

作者简介:谷阳阳(1982-), 女, 大学本科, 工程师。

引用格式:谷阳阳, 张东. 基于误差序列的离线数据无监督快速异常检测方法研究[J]. 计算机测量与控制, 2025, 33(11):111

障检测, 实现有人工到自动分析决策是现在研究的热点之一。国内外在卫星故障检测技术领域利用数据驱动进行决策已展开了研究, 在个别相关方面取得了一系列的成果。E Carlos 等设计了遥测数据时间序列相似度度量是通过不同概率密度函数之间的偏差作为规则^[5], 从度量时间序列数据之间的相似性完成规则分类, 更适用于度量简单样本, 对于随着时间序列变化复杂, 或者变化机动性较大的数据样本, 遥测数据的故障检测效果很难保证。Zirije Hasani 对海量数据流的异常检测进行了研究, 并对 Hold-Winters、分层时间记忆、移动平均等现有检测方法进行了改进^[6], 但是对于突变的有效捕捉还是有一定的滞后性, 同时窗口的选择也会直接影响到检测效果; 由于飞行器遥测数据异常检测面临时序数据维度大、异常数据不平衡、异常样本缺乏、异常标签少等问题, 刘切等提出一种基于迁移学习的深度异常检测模型^[7], 但是需要设计鲁棒性更强的模型挑战巨大, 模型的泛化和检测能力直接体现了检测效果的好坏。面对多维航空时序数据维度大、噪声多、缺失率高、采样间隔不规则等特点, 北京理工大学提出一种非规则采样多维时序数据异常检测 (IMAD) 算法^[8], 但是航空数据的特点是异常数据远远少于正常数据, 数据的不平衡性会影响模型的准确性, 需要设计泛化性强的模型, 同时计算的复杂度和精确性较高。

为了满足各种遂行处理任务需求, 我们迫切地需要针对自身分析处理需求, 结合上述故障诊断的优缺点和适应性, 在分析遥测数据特点的基础上, 研究更具通用性、实用性、高效性的故障检测与分析方法, 在大数据量的遥测待分析数据中能够迅速定位异常数据, 将相关分析预测提供给处理人员, 实现自动或辅助检测与分析功能, 提高处理与分析的能力与效率是我们亟待解决的问题。

1 飞行器遥测数据特点及规律

遥测数据是飞行器监测和记录与其相关的各系统工作状态的参量^[9-10], 通过传感器和变换器将所需测量的物理量转变为电量等数据集, 测量系统通过参量的变化对飞行器实现飞行状态及工作过程的追踪、监测和控制, 完成预定的工作。遥测数据是工作人员评估飞行器运行状态、鉴定其设计功能是否有效及安全性、可靠性的依据。由于飞行器自身就是一个复杂系统, 所以遥测数据反映出是一种多元时间序列, 所以在遥测数据分析和异常检测中需要结合其自身数据的特点及变化规律进行分析处理, 充分考虑到数据与飞行器运行机理的对应性, 从而提高异常检测效率和鲁棒性。

1.1 遥测历史数据特点

1) 历史数据集是真实试验环境下采集到的数据集,

包含试验过程中完整的环境信息、产品系统内部和外部信息, 机器学习需要将大量的历史数据集分为训练集和测试集, 繁复标签标注划分正常数据与异常数据后形成训练集, 通过大量迭代训练才能得到标准的训练模型; 而直接利用该阶段试验中成功的一次历史数据集, 作为异常检测的标准模型是容易获得的, 避免了繁杂的验前准备工作, 同时可以快速处理的需求。

2) 遥测数据处理方法随型号以及遥测系统的变化而变化, 同型号不同批次的试验也会有所不同, 因此需要相对应的遥测数据分析的检测技术也要适应这一特点而具有很大的灵活性和适应性, 而历史数据完整的将每次试验特性及变化以数据的形式保留了下来, 对历史数据加以利用, 可以有效的适应遥测系统的灵活变化的特点。

3) 遥测系统复杂度高、关联性强, 无法通过具体某一个或几个参数就能够将该系统描述出来, 通过计算机准确的将系统的物理模型转换成数学模型进行描述难度较大。因此, 基于模型的故障检测方法在遥测系统级故障检测运用上效果和难度都很大。基于数据驱动的故障检测方法偏向于收集能够间接描述系统特征的数据, 如正常数据与异常数据, 基于统计的故障检测方法主要挖掘对象的统计学特征, 所以基于数据驱动的故障检测方法和基于统计的故障检测方法相结合, 将较为适合遥测复杂的软、硬件系统异常检测^[11]。以数据驱动故障检测方法的基础是大量的数据信息, 其中包括大量的历史数据信息, 所以直接利用历史数据进行数据统计挖掘, 可以有效的保留遥测系统特征, 降低数学模型的模拟难度。

1.2 遥测参数变化规律

遥测数据种类繁多, 按照参数所属分系统可分为: 控制系统参数、动力系统参数、通信系统参数、测量系统参数等; 按照记录形式可分为: 模拟量参数、数字量参数等; 按照实效性可分为实时数据和历史数据; 按照是否译解处理可分为遥测原码和事后处理结果数据。对飞行器遥测数据进行分析时, 通常可以得到如下几种典型的数据变化规律。

1) 数据在规定的范围内上下波动, 呈现平稳变化趋势:

当外界环境正常且飞行器处于平稳工作状态下时, 这类数据呈现平稳变化趋势, 即使在记录过程中存在小范围的波动, 但都在预设的允许误差范围内, 可认定为正常变化趋势^[12-14]。这类参数主要包括飞行器工作中的温度、电压等传感器的测量量。这类参数的变化特点是, 同一产品系统内环境正常工作时, 其测量量均在设计产品的取值范围上下限内, 不会随飞行状态的改变剧烈波动。

2) 周期变化参数:

从参数变化趋势上看, 这类参数呈固定周期变化, 每个周期内变化趋势基本相同, 局部可能会有微小的变动, 但是在设定的误差范围内。这类参数的变化特点是, 同一产品系统内环境正常工作时, 其测量量均在设计产品的取值周期内, 不会随飞行状态的改变剧烈波动; 若同一产品不同阶段变化周期和幅值会有所不同, 但变化量呈固定状态。

3) 飞行器工作中遥测数据有突变非平稳变化:

遥测参数记录的完整飞行过程中, 存在记录变化剧烈的段落或时刻。需要专家根据产品的设计需求判断该变化段落或时刻是否正常, 可能是故障发生时刻或飞行异常段落, 也可能是根据设计需求变化的正常飞行段落或变化时刻, 需要专家根据新的记录数据依据判读文件具体判别。

4) 飞行器工作中遥测数据缓慢平稳变化:

遥测参数记录的完整飞行过程中, 存在记录遥测数据缓慢平稳变化, 可能是产品本身设计需求对应的参数记录正常状态, 也可能这种变化趋势或者变化量并非产品设计需求, 而是由于故障关联导致的缓慢变化。此时需要专家根据产品设计需求进行关联性分析, 判定该参数的记录数据是否是异常数据。

通过上述对飞行器遥测数据特点及规律的分析, 可以看出飞行器遥测数据异常类型主要有三种形式: 一种是伴随着周期变化的异常; 一种是突变非平稳变化表现出的点异常或突变; 一种是与故障关联的偏离正常状态异常。针对典型的异常状态, 多数是通过人工和专家知识进行判别的, 但是面对大数据、多维度、实时性的处理需求, 需要结合自身特点, 研究泛化性和鲁棒性更好的快速异常检测方法。

2 基于误差序列的无监督快速异常检测

在实际的飞行器故障检测分析中, 最常用、最直接的是人工检测方法。传统人工故障检测可靠性强、操作简单, 伴随着测量数据的体量增长, 缺点也随之出现: 不仅需要了解该型号的技术人员参与检测, 各分系统涉及人员较多, 人工成本较高, 而且遥测参量众多, 短时间确定故障的压力较大, 依托数据决策的时间成本过高。基于专家知识的飞行器故障检测方法主要是将已有知识和经验编码形成规则和推理机制进行故障检测^[15]。该方法更多的应用于飞行器故障检测, 但是系统对于专家的经验和知识要求十分苛刻, 会直接影响到检测效果, 并且不同的飞行器之间具有结构和功能上的细微差异。基于解析模型的故障检测方法是通过数学模型描述飞行器, 利用构造模型与测量输出的残差, 反映模型系统与测量系统的一致性来判断系统是否出现了故障。该

方法需要精确模拟运行模型, 为了检测效果良好需要对飞行器各分系统建立独立的模型, 存在代价大, 扩展性差的现实问题^[16]。基于数据驱动的故障检测方法利用人工智能或统计分析方法通过大量数据的处理, 提取与故障相关的特征信息, 建立数学模型, 实现对数据的检测和预警^[17-21]。但是现有状况是不同飞行器的设计需求和运行模式差异较大, 通过大量历史数据的处理、训练、建模等环节最终形成相对准确的模型, 只能映射该种型号飞行器的检测和诊断模型, 不能用于其他型号的自动检测与诊断; 即使同一型号的飞行器, 在不同的研制阶段中会涉及到多种不同形式的检验与考核内容, 飞行状态、控制过程、反馈形式等过程的有很多不同之处, 导致同一型号一种模型无法在短时间内训练形成兼容性较好的的模型, 若要进行数据异常检测, 需要大量的时间和精力训练同型号适应不同考核阶段的模型。再者, 若飞行器在工作过程中存在异常, 需要在试验结束后迅速地对任务数据进行分析, 定位故障点及产生原因, 为指挥决策提供依据, 需要训练出的模型准确度和精确度要求较高, 否则不能在第一时间找到异常数据, 辅助决策作用发挥不充分。为了满足多维度、大数据量、多模态的故障检测需求, 在基于数据驱动的故障检测方法之上, 结合遥测参数的特征, 研究设计了一种适应变化多样遥测参量的无监督异常检测方法。

离线数据序列异常检测的主要思想是: 将相似批次飞行器记录的遥测历史数据序列作为验前信息与测量待处理数据列进行一致性分析, 当历史数据序列中某时间段的值和此次实测值有较大的偏差时, 那么大概率该时间段飞行器发生了异常情况。

飞行器在完成某次飞行后形成此次测量遥测原始数据, 根据相应处理要求对测量遥测原始数据进行处理形成处理结果, 完成由电量到物理量的译解。处理人员就是根据处理完成后的参量信息作为依据进行分析、鉴定工作, 那么该信息作为离线待测数据序列进行异常检测分析。检测处理过程需要选取相似批次飞行器记录的历史数据序列与离线待测数据值进行比较, 形成序列误差, 那么两个数据序列的差异性就可以通过序列误差表征。通过对序列误差进行滤波处理, 进而突出序列误差特征, 同时通过设定该参量的异常阈值对平滑误差序列进行异常检测, 通过平滑误差序列异常点表征离线待测数据的异常点, 完成该参量的异常检测。

之所以将相似批次飞行器记录的历史数据序列作为验前信息使用, 是由于相似批次考核的大部分环境参量是相似的, 间接关联和监测参量变化趋势是相似的, 这些参量的历史数据与离线待测数据的误差序列变化量较小或趋于恒定, 所以适合通过误差序列进行检测, 只有涉及到不同考核课目的相关引导和控制参量与历史数据

值会有很大的不同,但是通过异常检测算法可以快速的从成百上千的处理结果文件中快速剥离出可能变化异常的参量,定位到可能异常的数据变化段落及时刻,降低数据分析人员对数据的分析量,提升了数据分析的效率及准确度,为第一时间查找故障原因提供了辅助决策功能。

2.1 误差序列的异常检测

定义在 t 时刻,历史数据值为 $y'(t)$, 离线待测数据值为 $y(t)$, 则该时刻的序列误差 $e(t)$ 计算方法为:

$$e(t) = |y(t) - y'(t)|$$

每一时刻产生的误差 $e(t)$ 最终形成一个一维的误差序列: $e(t) = [e(0), e(1), \dots, e(t)]$, 当离线待测序列中存在异常的序列段时,由于历史数据序列作为正常状态的取值,则异常区域序列的误差将会比无异常区域序列的误差值更大。

由于误差序列 $e(t)$ 结果会存在大量的尖峰毛刺,会导致信号的主要特征表现的不明显。为了更好的体现误差序列特征,有效过滤噪声,同时保留信号的主要特征,选用高斯加权移动 (GWMA, gaussian weighted moving average filtering) 对 $e(t)$ 进行数据平滑处理。对原始误差序列通过高斯加权移动平滑滤波后形成的 $e_g(t)$ 进行异常序列检测。当平滑误差序列 $e_g(t)$ 中误差值超出了设定的异常阈值时,则将该段序列标记为异常序列,那么异常阈值的设定直接影响到异常序列检测的效果。

2.2 无监督的异常阈值选取

在相近试验条件下的多次考核中历史数据和离线待测数据本身均还有相近的系统误差,通过分析平滑误差序列 $e_g(t)$ 的期望 $\lambda(e_g)$,可以有效地消除测量误差对异常检测的影响。同时随机误差的数学期望为零,而系统误差的数学期望就是它本身,即使改变试验条件,系统误差的数学期望会呈现一个常数。则序列 $e_g(t)$ 的异常阈值 k 定义为:

$$k = \lambda(e_g) \pm \gamma \cdot \sigma(e_g)$$

式中, $\sigma(e_g)$ 为误差序列的标准差, $\gamma \cdot \sigma(e_g)$ 表示序列偏离数学期望的程度。该式中通过调节 γ 的取值来影响异常阈值 k 的取值,使得调节 k 的取值适应不同参数的检测需求,不会造成误判漏判的情况发生,因此 γ 需满足条件:

$$\left\{ \begin{array}{l} \text{argmax}_{\gamma} \frac{\Delta \sigma^2(e'_g) / \sigma^2(e_g)}{|e'_g|} \\ \Delta \sigma^2(e'_g) = \sigma^2(e_g - e'_g) - \sigma^2(\{e_g \in e_g \mid k_- < e_g < k_+\}) \\ e'_g = \{e'_g \in e_g \mid k_+ > e_g, e_g < k_-\} \end{array} \right.$$

当平滑误差序列 $e_g(t)$ 中超过阈值的部分被剔除后,使得 $\sigma^2(e_g - e'_g)$ 的方差和 $\sigma^2(e_g)$ 方差形成最大的衰减比,以此判断异常判别的效果是否最佳。上式中

$\Delta \sigma^2(e'_g)$ 为异常值剔除后计算的误差序列的方差变化量,便于分析人员判断变化程度。 $|e'_g|$ 记录检测出异常值的数量,便于统计分析。通过上述方法,在不用获取异常序列标签的情况下,自适应确定该数据序列的异常阈值 k ,进而自动检测出待测参量的异常序列。

综上所述,离线数据序列异常检测方法是将原有故障分析对象由离线参量物理值转变为对其处理后的序列误差值,通过异常阈值的自适应选取对序列误差进行再加工,在无人为操作的情况下,自动识别出异常误差序列,将识别出异常误差序列对应到待检测的离线待测数据系列,从而完成该遥测参量的故障检测。该方法通过巧妙引用历史数据作为验前信息参与到异常值检测中,大大减少了数据驱动型故障检测需要大量数据的计算,提取与故障相关的特征信息,建立数学模型等需求,同时无监督的异常阈值选取又契合了遥测数据多维度的特点,有效提高了异常检测方法的泛化性和鲁棒性。

3 实验结果与分析

同一型号的飞行器在研制考核的不同阶段,遥测参数的记录状态有的是不变的,如反应内环境的温度、电压、状态监测的参量;有的是变化的,如综控、导航、导引头、干扰等参量随着设计考核方案的不同而变化。基于误差序列的离线数据快速异常检测方法是针对飞行器遥测数据的变化特点,充分利用该型号的历史数据信息的基础上,从成百上千的遥测解析参数中,迅速定位异常参量,实现快速的异常检测定位,自动筛选出异常参数文件,作为专家分析故障的数据支撑,提高人工检测的实效性,同时保证了最终检测结果的可靠性。

3.1 实验环境

离线数据无监督快速异常检测环境需要以下关键组件如下。

1) 硬件如表 1 所示。

表 1 系统硬件环境说明

项目	详细要求
计算资源	高性能 CPU 和 GPU,如 Intel i7,32 GB RAM, NVIDIA RTX 3080,1 TB SSD
内存	至少 16 GB,建议 32 GB DDR4
存储	高速 SSD,可根据容量数据规模而定

2) 软件环境如表 2 所示。

表 2 系统软件环境说明

项目	详细要求
操作系统	Linux 或 Windows7/10/11
编程语言	Python 3.7 或更高的版本
开发工具	Jupyter Notebook、PyCharm 等
数据处理	Pandas、NumPy 等库
可视化工具	Matplotlib、Plotly、Seaborn 等

3) 服务器如表 3 所示。

表 3 系统服务器环境说明

项目	详细要求
操作系统	服务器操作系统:Ubuntu Server20.04 LTS
服务器软件	Web 服务器:Apache 2.4 数据库服务器:MySQL 8.0 应用服务器:Tomcat 9.0

3.2 测试数据

表 4 测试数据样本

参量名称	历史数据	待测数据	异常阈值 K	是否异常
XX 壁温/℃	23.156	22.699 4	0.312 5	否
	23.156	23.094 6		
	23.156	22.618 1		
	30.306 5	29.525 7		
	30.306 5	29.782 8		
	30.306 5	29.750 5		
	30.306 5	30.173 0		
供电电压/V	0.493 5	是
	2.553	1.808 90		
	2.563	0.859 76		
	3.565	1.807 16		
	3.564	1.636 83		
	3.574	2.216 91		
	3.583	2.155 98		
	3.586	2.706 59		
XX 补偿量/g	0.103 5	否
	0.936 5	0.924 5		
	0.935 5	0.925 5		
	0.936 5	0.925 5		
	0.936 6	0.946 6		
	0.938 5	0.940 5		
	0.934 4	0.925 4		
	0.935 8	0.943 8		

测试样本数据选取 3 种典型遥测数据变化类型, XX 壁温在传感器加电工作时, 由 23.156 ℃ 阶跃变化为 30.306 5 ℃ 平稳变化在规定的范围内上下波动, 呈现平稳变化趋势, 包含了节约变化和平稳变化形式; 供电电压参量属于有突变非平稳变化的形式, 在实际数据集中这种形式占比较大, 所以测试结果分析中以此进行了具体详细论述; XX 补偿量呈固定周期变化, 每个周期内变化趋势基本相同, 局部可能会有微小的变动不会随飞行状态的改变剧烈波动, 呈平稳变化形式。异常阈值从统计表中可以看到是根据不同参量自适应变化的, 在有效地保证了异常检测的泛化性的同时兼顾了检测效率, 保证了大数据集的检测处理效果。

3.3 实验步骤及方法

输入: 相同批次同型号的历史数据集和译解后待检

测数据集。

输出: 存在异常的参量集。

- 1) 根据导入的历史数据集和待检测数据集, 选择参量名称相同的历史数据值 $y'(t)$ 、待检测数据值 $y(t)$;
- 2) 计算出同时刻的序列误差 $e(t)$, 形成误差序列;
- 3) 通过高斯加权移动对 $e(t)$ 进行数据平滑处理, 形成平滑误差序列 $e_g(t)$;
- 4) 通过计算平滑误差序列 $e_g(t)$ 的期望 $\lambda(e_g)$, 确定该参量的异常阈值 k ;
- 5) 当平滑误差序列 $e_g(t)$ 中超出了设定的异常阈值 k 时, 标记出该时刻序列, 将该参量标记为异常参量;
- 6) 当数据集中所有参量均完成检测后, 该数据集的异常检测结束。

某一参量具体检测方案流程如图 1 所示。

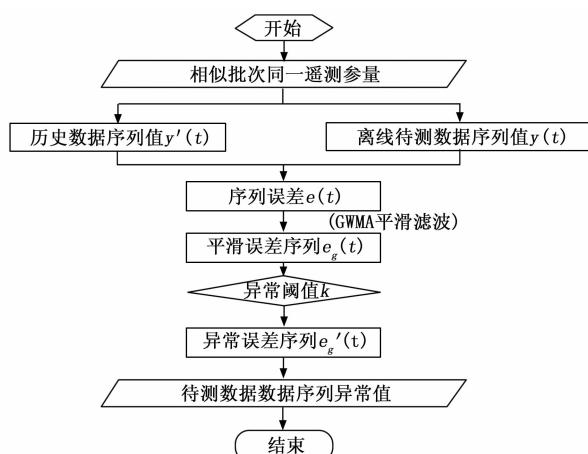


图 1 离线数据序列异常检测方案流程

3.4 测试结果分析

某一型飞行器模拟测试数据结果为例进行仿真测试, 以其中的供电电压参量为例, 通过曲线图的形式加以分析, 历史数据序列和待测离线数据序列如图 2 所示, 历史数据作为验前信息 $y'(t)$, 待测数据序列 $y(t)$ 是需要检测是否存在异常序列的数据集, 两个数据集分别模拟同一型号飞行器在相似环境下进行的不同批次试验译解后的参量信息。从图示中可以直观看出历史数据记录的情况要好于待测数据, 待测数据记录的电压量不是很稳定, 译解出的参量数据中有大量的毛刺, 不知道状态是否正常。

为了自动检测该待测数据集记录信息是否存在异常序列, 利用误差序列的异常检测方法, 将两次的数据集做差值, 得到误差序列 $e(t)$ 如图 3 中所示, 可以看到误差序列 $e(t)$ 的尖峰和毛刺较多, 信号的变化趋势和特征容易被淹没, 同时也为了消除测量误差带来的影响, 利用高斯加权移动方法对误差序列 $e(t)$ 进行平滑滤波, 形成平滑误差序列 $e_g(t)$, 从图 3 中可以看出 $e_g(t)$ 去除了原序列中的高斯噪声, 同时有效的保留住信号的主要

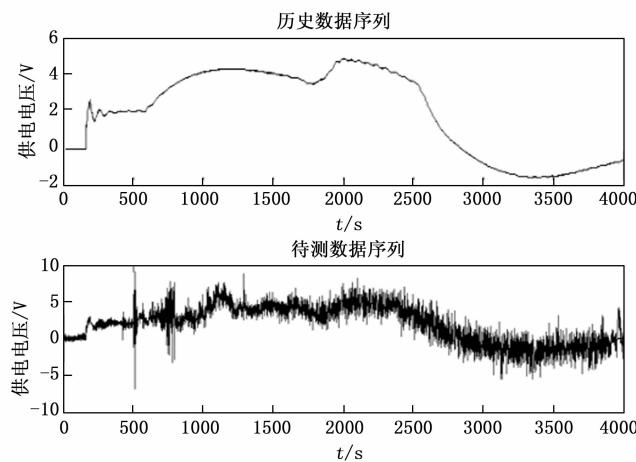


图 2 历史数据序列和待测数据序列

变化特征。对平滑误差序列 $e_g(t)$ 求期望 $\lambda(e_g)$ ，可以看出平滑误差序列 $e_g(t)$ 中偏离期望的程度是不同的，通过判断平滑误差序列 $e_g(t)$ 偏离期望的程度可以有效的识别异常点或序列发生的时刻或段落。

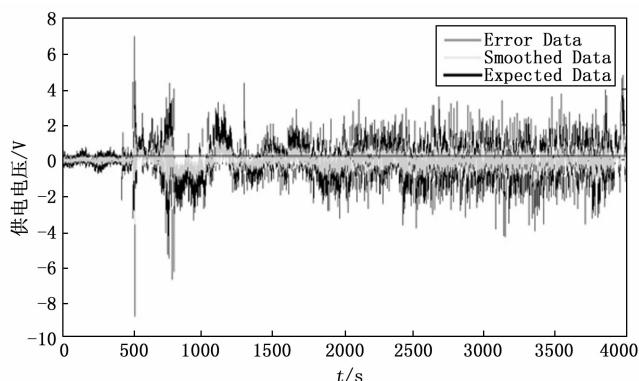


图 3 误差序列、平滑序列和期望

将图 3 中的平滑序列和期望放大如图 4 所示，该平滑误差序列的误差分布集中在 0.2 附近，通过无监督的异常阈值选取，迭代计算确定序列的异常阈值 k 如图 4 所示，从图 4 可以看出超出 K -Limit 的部分都是自动判断为数据异常的段落，该数据集在 0~450 的时间段落内是工作正常的，450~500 的时间段落只有个别数据点超出了阈值，不能严格地确定是异常点，也可能是由于记录野值引起的扰动，但 500 开始突然有大量的数据点偏离阈值，同时后续异常段落明显增加，所以可以判断出从 500 开始该传感器测量的对应参量开始发生异常，同时后续测得的异常段落也印证了该参量在异常状态下的工作情况，因为飞行器工作时各传感器间是相互关联、实时反馈的，所以如果某个器件工作异常，通常都是同一系统内耦合性强的参量先异常，再影响与其对应反馈器件或系统异常，所以有些参量异常点非常明显，而有些参量反应出的是异常段落。该参量通过误差

序列无监督快速异常检测方法得到起始的异常段落为 450~500 的时间段落，500 对应时刻已经出现了明显的异常数据。

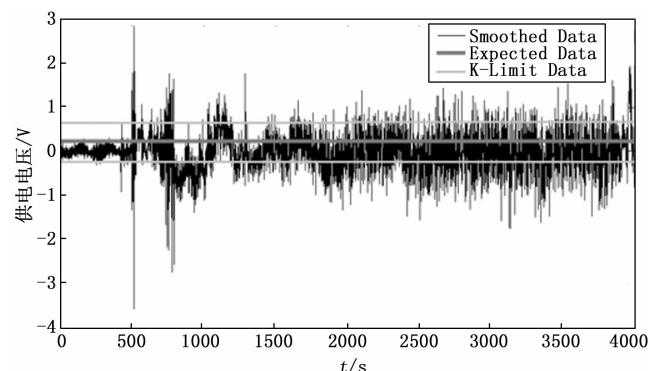


图 4 误差序列、平滑序列和期望

4 结束语

本文针对飞行器故障检测需求的现状，结合遥测数据的特点及变化规律，从实际处理需求出发，将历史数据加以利用，通过设定自适应阈值对与其对应的误差序列进行处理，设计了适用于飞行器的基于误差序列的离线数据无监督快速异常检测方法，实现无监督快速异常检测遥测参量。从实际仿真测试效果来看，结果准确，通过判断异常段落或异常点的发生，可以迅速定位出与该参量相关联的飞行器异常工作系统或故障发生的器件；同时，通过无监督阈值设定，实现了在成百上千个遥测参量中适配各自参量记录的数据信息，从而自动筛选出记录异常的参量，快速地从异常参量中提取出异常段落或异常点。将自动检测出的异常段落和异常点作为数据处理结果的一部分提供给分析人员，通过分析参量异常段落的分布或时序，可以帮助处理人员全面了解掌握故障发生的过程；通过异常段落持续状态的再分析，利用系统间的耦合关系可以相互印证故障原因有效性，有效减少人工分析的工作量及压力，大大提高了故障分析处理的速度与效率，满足了快速定位故障数据、分析故障原因需求。在飞行器遥测数据多样化，数据维度不断增加，数据量呈现指数式爆发的形式下，提高数据处理系统的通用性和可扩展性，满足日益增长的遥测数据处理需求，有效地提升数据处理的综合能力，为遥测数据处理通用性平台的建设奠定基础；同时，该方法是基于对历史数据利用，提高了对历史数据利用率，有效提升了“用数据说话的能力”，为数据驱动性的大数据建设奠定基础，具有很好的现实意义。

参考文献：

- [1] SHANGGUAN D, CHEN L, DING J. A digital twinbased approach for the fault diagnosis and health monitoring of a

- complex satellite system [J]. Symmetry, 2020, 12 (8): 1307.
- [2] 李楠, 强懿耕, 樊瑞, 等. 基于异常因子的航空器飞行轨迹异常检测研究 [J]. 安全与环境学报, 2021, 21 (2): 643–648.
- [3] GE H, LI B, JIA S, et al. LEO enhanced global navigation satellite system (LeGNSS): progress, opportunities, and challenges [J]. Geo-spatial Information Science, 2022, 25 (1): 1–13.
- [4] ZHANG L. Development and prospect of chinese lunar relay communication satellite [J]. Space: Science & Technology, 2021, 1: 241–254.
- [5] CARLOS E, MELLO, ST ANDRÉ, et al. Time series classification via divergence measures between probability density functions [J]. Pattern Recognition Letters, 2019, 125: 42–48.
- [6] HASANI Z, KRRABAJ S. Survey and proposal of an adaptive anomaly detection algorithm for periodic data streams [J]. Journal of Computer and Communications, 2019, 7: 33–55.
- [7] 刘切, 上官子卓, 李嘉玺. 基于迁移学习的航天器遥测数据异常检测技术 [J]. 空间控制技术与应用, 2023, 49 (4): 76–85.
- [8] 闫廷锦, 夏元清, 张宏伟, 等. 一种非规则采样航空时序数据异常检测方法 [J]. 航空学报, 2021, 42 (4): 1–11.
- [9] 陈俊夫, 皮德常, 张强. 一种基于迁移学习的遥测数据异常检测方法 [J]. 宇航学报, 2021, 42 (4): 522–530.
- [10] 张晓, 孙超, 王曼燕, 等. 基于大数据技术的实时卫星数据监视系统的设计与实现 [J]. 自动化技术与应用, 2022, 41 (9): 63–68.
- [11] 赵玉炜, 苏举. 基于参数自适应优化聚类的卫星状态异常检测方法 [J]. 空间科学学报, 2023, 43 (5): 927–933.
- (上接第 110 页)
- [12] SAN J, WANSHOU J, WEI H, et al. UAV-Based oblique photogrammetry for outdoor data acquisition and off-site visual inspection of transmission line [J]. Remote Sensing, 2017, 9 (3): 278.
- [13] 王舒梦, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 航拍轻量化小目标检测算法: PECS-YOLO [J/OL]. 计算机工程, 1–16 [2024-07-20].
- [14] 王磊, 郝涌汀, 潘明然, 等. 电力巡检中改进 YOLOv5s 的缺陷检测算法研究 [J]. 计算机工程与应用, 2024, 60 (10): 256–265.
- [15] 贵向泉, 秦庆松, 孔令旺. 基于改进 YOLOv5s 的小目标检测算法 [J]. 计算机工程与设计, 2024, 45 (4): 1134–1140.
- 937.
- [12] 熊泰然, 陈雯雯, 陈宏宇, 等. 基于自适应 KF 动态虚拟陀螺数据融合算法的研究 [J]. 上海航天 (中英文), 2021, 38 (6): 124–130.
- [13] 于峰, 郭珈. 一种靶场试验遥测数据处理与分析系统的设计与实现 [J]. 遥测遥控, 2022, 43 (1): 98–104.
- [14] QU S, ZHENG Y, YAN H, et al. Operational effectiveness evaluation of antimissile early warning based on combination weighted TOPSIS [J]. Journal of Physics: Conference Series, 2021, 1927 (1).
- [15] 史晓云, 陈军, 郭小红, 等. 基于回归分析的航天器异常检测方法 [J]. 遥测遥控, 2022, 43 (2): 57–62.
- [16] YUN K, CHONG Y, ENZHE S, et al. Fault diagnosis method of diesel engine injector based on hierarchical weighted permutation entropy [C] // 2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2021: 1–6.
- [17] 李桢煜, 宋宇晨, 彭喜元, 等. 基于对比序列重构的卫星遥测数据异常检测方法 [J]. 仪器仪表学报, 2024, 45 (4): 17–26.
- [18] 彭会湘, 刘光辉, 陈韬亦, 等. 一种基于离散时间段的测运控数据异常检测方法 [J]. 无线电工程, 2021, 51 (6): 458–469.
- [19] HU Y J, ZHANG L L, ZHAO K, et al. Android privacy leak detection method based on static taint analysis [J]. Journal of Cyber Security, 2020, 5 (5): 144–151.
- [20] HRISTOZOV S, WETTERMANN M, HUBER M. A toctou attack on dice attestation [J]. Advances in Neural Information Processing Systems, 2022, 24: 59–75.
- [21] 王岩, 李少波, 张仪宗, 等. 数据驱动的无人机异常检测算法综述 [J]. 无线电工程, 2024, 54 (6): 1407–1420.
- [22] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, Real-Time object detection [C] // Computer Vision & Pattern Recognition, IEEE, 2016.
- [23] 吴涛. 基于深度学习的航拍绝缘子缺陷检测研究 [D]. 大连: 大连理工大学, 2019.
- [24] 郝紫霄, 王琦. 基于 YOLO-v7 的无人机航拍图像小目标检测改进算法 [J]. 软件导刊, 2024, 23 (1): 167–172.
- [25] 郭敬东, 陈彬, 王仁书, 等. 基于 YOLO 的无人机电力线路杆塔巡检图像实时检测 [J]. 中国电力, 2019, 52 (7): 17–23.
- [26] 杨利波, 杨嘉妮, 贺学敏. 基于 YOLO 目标检测算法的输电线路检测 [J]. 电力信息与通信技术, 2022, 20 (8): 99–105.