基于无监督时空状态估计的信息物理系统细粒度异常诊断

孙海丽[1]，黄炎[2]，韩兰胜[1,3]，周纯杰[2]

（1.华中科技大学网络空间安全学院，湖北 武汉 430074；2.华中科技大学人工智能与自动化学院，湖北 武汉 430074；
3.武汉金银湖实验室，湖北 武汉 430048）

摘　要：为了揭示信息物理系统工作状态中的时空依赖关系及演变机制，提出细粒度自适应多元时间序列异常检测 Transfonner（MAD-Transformer）模型以识别与诊断多元时间序列中的异常。首先，通过构建时间状态矩阵来建模并估计系统状态在时间维度上的变化规律。其次，构建空间状态矩阵捕获系统传感器间的状态关联，以准确定位异常。然后，基于原始多元时间序列和时间、空间状态矩阵，设计一个三分支结构的序列-时间-空间注意力计算模块，以同时捕获信息物理系统各传感器之间的序列、时间与空间依赖性。最后，通过构建 3 个关联对齐损失和重构损失策略以训练优化模型。实验结果表明，所提模型不仅可以准确检测和定位异常，还能够细粒度地诊断异常的持续时间。

关键词：信息物理系统；自适应异常诊断；时间状态矩阵；空间状态矩阵；状态估计

中图分类号：TP393

文献标志码：A

DOI：10.11959/j.issn.1000–436x.2025129

Unsupervised spatio-temporal state estimation for fine-grained anomaly diagnosis of cyber-physical systems

SUN Haili[1], HUANG Yan[2], HAN Lansheng[1,3], ZHOU Chunjie[2]

1. School of Cyber Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
2. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China
3. Wuhan Jinyinhu Laboratory, Wuhan 430048, China

Abstract: To reveal the spatio-temporal dependence and evolution mechanisms in cyber-physical system operational states, a fine-grained adaptive multivaviate time series anomaly diagnosis (MAD-Transformer) model was proposed for identifying and diagnosing anomalies in multivariate time series (MTS). MAD-Transformer first constructed temporal state matrixes to characterize and estimate the evolutionary patterns of system states along the time dimension. Secondly, to locate the anomalies, spatial state matrixes were constructed to capture the inter-sensor state correlation. Subsequently, a triple-branch sequence-temporal-spatial attention module was designed to simultaneously capture the sequential, temporal, and spatial dependencies among MTS. Afterwards, three associated alignment loss functions and a reconstruction loss were constructed to jointly optimize the model. The experimental results show that the MAD-Transformer can not only accurately detect and locate the anomaly, but also fine-grained diagnose the duration of the anomaly.

Keywords: cyber-physical system, adaptive anomaly diagnosis, spatio state matrix, temporal state matrix, state estimation

收稿日期：2025–04–17；**修回日期：**2025–07–07

通信作者：韩兰胜，hanlansheng@hust.edu.cn

基金项目：国家自然科学基金资助项目（No.62072200, No.62172176）；国家重点研发计划基金资助项目（No.2022YFB3103400）；国家重大科研仪器研制基金资助项目（No.62127808）

Foundation Items: The National Natural Science Foundation of China (No.62072200, No.62172176), The National Key Research and Development Program of China (No.2022YFB3103400), The National Major Scientific Research Instrument Development Project of China (No.62127808)

0　引言

随着科技的飞速发展，复杂性高的信息物理系统（CPS, cyber-physical system）在当代工业领域得到了越来越广泛的应用。CPS将计算与物理过程紧密结合，实现了信息的无缝集成与流动，从而极大地提高了生产效率、系统性能以及资源利用率。在众多行业中，如智能制造、智能交通、智能能源管理等，CPS正逐步取代传统系统，成为推动产业升级转型的重要力量。复杂CPS的广泛应用，为当代工业领域带来了前所未有的变革机遇。CPS通过实时监控生产线上的各种设备状态、产品质量和环境参数，实现了生产过程的智能化控制[1]。基于数据分析与人工智能算法，这些系统可以预测设备故障、优化生产流程、减少资源浪费，从而显著提升生产效率和产品质量。

然而，CPS在网络安全层面面临着严峻挑战。由于系统协议开放性、设备异构性及安全防护滞后性等固有缺陷，CPS极易成为网络攻击、物理攻击、数据注入攻击等多种恶意行为的目标。攻击者可通过中间人攻击篡改控制指令，利用重放攻击干扰时序逻辑，或实施虚假数据注入攻击误导决策系统，进而引发生产中断、设备损毁甚至威胁公共安全等严重后果。特别是在能源、交通、医疗等关键基础设施领域，此类攻击可能导致区域性服务瘫痪、重大人员伤亡及不可估量的经济损失。CPS持续产生大量多元时间序列数据，以水处理工厂为例，其部署的分布式传感器网络所采集的监测数据，涵盖流量参数、温度指标以及液位高度等多类型过程变量数值，因此研究者提出了基于多元时间序列分析的方法来检测CPS中的攻击行为。

由于传感器误差、数据传输波动、系统本身的波动以及外部环境的干扰，CPS生成的多元时间序列数据中往往伴随着噪声，这要求精心设计的算法能够具有很强的抗噪能力，因此高准确率、低误报率的异常检测方法对保障系统正常运行和缓解高昂的经济损失是至关重要的。另外，为了方便操作人员快速准确地排查出异常的发生原因和所在位置，通常需要精准地定位异常，即确定是哪些传感器引起了相应的异常。在现实世界的应用场景中，现代系统的容错机制意味着由时间波动或系统状态切换引起的短期异常可能不会最终导致真正的系统故障。因此，如果能够对异常的严重性进行评估，将会帮助操作人员在异常的影响和宕机带来的损失之间找到合适的平衡，毕竟工业系统与传统信息技术系统不同，工业系统的宕机也会造成严重的经济损失。据报道，一家汽车制造厂停机1 min就可能导致高达2万美元的经济损失[2]。假设异常的严重性正比于发生异常的设备（如传感器、执行器）数以及异常持续的时间，如图1所示。图1中有A_1和A_2这2个异常，A_1的异常点有5个，A_2的异常点有3个，且A_1比A_2持续的时间更长，因此认为A_1比A_2更严重。

在构建工业领域的异常检测和诊断算法时，首先遇到的一个挑战是标签样本的稀缺性，部分场景甚至完全缺失标签样本，在这种情况下，由于训练样本的数量不足，传统的有监督学习算法可能会导致过拟合或欠拟合的问题，进而使得模型的泛化能力大打折扣。近年来，一些研究学者提出基于无监督学习的异常检测方法，例如基于长短时记忆（LSTM, long short-term memory）神经网络构建的深度学习模型[3-5]，虽然能够捕获时间序列中的时

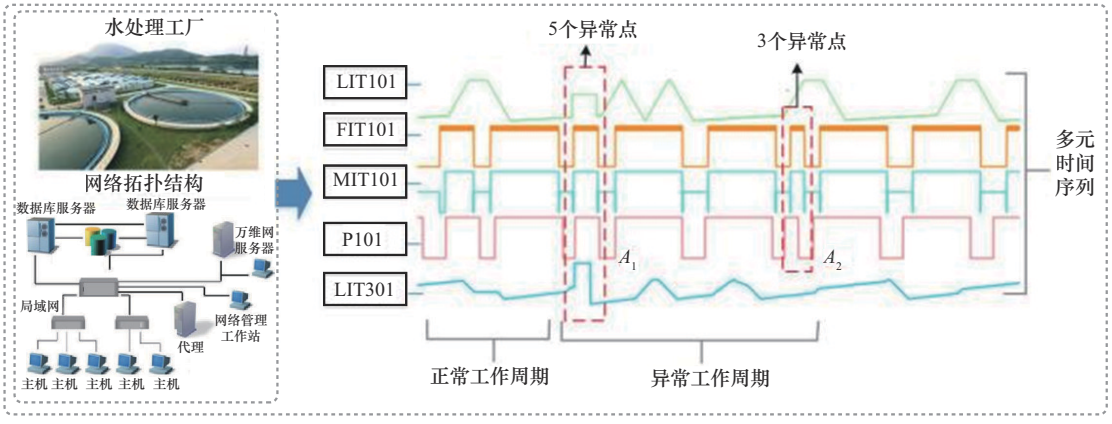


图1　异常严重性解释

间依赖关系，但大都忽视了时间序列中存在的业务空间上的依赖关系。另外，在现实世界中，为操作员提供异常发生的可能位置和异常发生严重程度的诊断说明具有较大意义。但是，现有的异常方法很少能够给出异常的严重性评估。

从多元时间序列中准确检测和诊断网络攻击等异常行为对保障工业信息物理系统稳定有效地运行至关重要。为此，要求精心设计的模型不仅能够捕获多元时间序列中的时间关联，而且能够学习不同时间序列数据之间在空间业务上的关联。此外，有必要对异常结果进行定位和自适应粒度地评估其严重程度，以提高异常的可解释性和操作人员对异常的响应速度。现有研究工作较少关注系统工作状态的逻辑依赖关系，对系统异常信号的演变机制解释不清。

针对此问题，本文提出一种细粒度自适应异常诊断方法来保障工业系统安全地工作，该方法可以同时捕获多元时间序列中的时间和空间业务上的依赖关系以及定位和细粒度地诊断异常的严重性。具体而言，设计时间状态矩阵刻画系统整体状态在不同时间步的演变，以提取多元时间序列的时间依赖关系。同时，构建空间状态矩阵来捕获系统状态中的空间依赖关系。随后，将这 2 个状态矩阵和原始时间序列输入精心构建的拥有三分支注意力的多元时间序列异常检测 Transformer （MAD-Transformer）模型中，以显式地学习多元时间序列的时间、空间和序列依赖关系。然后，设计了 3 个关联差异函数以在特征空间中对齐这 3 种依赖关系。最后，基于重构的序列、时间状态矩阵和空间状态矩阵计算出异常得分和相应的残差矩阵。其中，异常得分用以判别时间点的异常与否，残差空间状态矩阵用以判断异常发生的最可能的位置，而残差时间状态矩阵用以细粒度地诊断异常的严重性。

本文主要贡献如下。

1) 时间状态矩阵与空间状态矩阵建模：提出时间状态矩阵与空间状态矩阵，分别刻画工业 CPS 中多元时间序列在时间和空间维度的依赖关系。时间状态矩阵反映系统状态随时间的变化规律，可用于检测异常及刻画异常的持续时间；空间状态矩阵依据序列间依赖关系构建，能精准定位异常发生位置。

2) 多分支注意力与对齐损失设计：设计三分支结构的注意力模块 MAD-Attention，分别从原始输入、时间状态矩阵和空间状态矩阵中提取序列、时间及空间关联依赖模式。同时，设计 3 个关联差异对齐损失将多层特征关联组合成更具信息量的度量，使模型学习到更全面的正常数据模式，增强异常识别能力，提升检测性能。

3) 细粒度异常诊断与自适应评估：通过计算重构的时间状态矩阵和空间状态矩阵与原始矩阵之间的残差，可以识别出异常发生的最可能的位置以及细粒度地评估异常的严重性。此外，还能依据数据业务，以自适应的粒度评估异常的严重性。

在 5 个公开数据集和自制耦合油罐控制系统数据集上开展了验证实验，与 21 个基准模型相比，结果证明，本文模型在性能上的优越性、异常定位上的准确性和自适应细粒度异常诊断的能力。

1　相关工作

由于工业场景下信息物理系统中的异常通常较为稀疏，面临少样本情况[6]，不依赖标记样本的无监督多元时间序列异常检测是一个重要且富有挑战的现实问题。在过去几年，研究者们已经提出了各种各样的异常检测方法。根据异常的判定方法不同，这些方法可以被粗略地分类为：基于距离的方法、基于密度估计的方法、基于聚类的方法、基于预测的方法和基于重构的方法。

基于距离的方法通过距离函数衡量异常，例如，k-近邻算法[7]根据到 k 个最近邻居的平均距离计算每个数据样本的异常分数。对于基于密度估计的方法，经典的方法是局部离群因子（LOF）[8]和连通性离群因子（COF）[9]，分别计算局部密度和局部连通性来确定离群值。深度自编码高斯混合模型（DAGMM）[10]和混合概率主成分分析和分类分布（MPPCACD）[11]整合高斯混合模型来估计表征的密度。对于基于聚类的方法，通过计算样本到簇中心的距离作为异常得分。深度支持向量数据描述（Deep SVDD）[12]将正常数据的表示收集到紧凑的簇中。时间层次单类（THOC）[13]通过分层聚类机制融合中间层的多尺度时间特征，并通过多层距离检测异常。基于张量的集成异常检测系统（ITAD）[14]对分解张量进行聚类。

尽管上述 3 类方法已经在各种应用中验证了它们的有效性，但是它们可能不能很好地处理多元时

间序列，因为它们都没有能力刻画时间序列中的时序依赖关系。为了解决这个问题，许多基于深度学习的方法开始使用LSTM神经网络来建模时间依赖关系。这些方法可以粗略地分为基于预测的方法和基于重构的方法。

基于预测的方法旨在预测下一个时间戳的值并根据预测误差来判别异常。Hundman等[5]证明了LSTM神经网络在探测航天器异常中的可行性，并引入了一种不依赖于注释的动态设置阈值的方法。相似地，Tariq等[3]基于LSTM神经网络提出了一种数据驱动的卫星异常检测方法，该方法同时使用神经网络和概率聚类来从遥测数据中识别异常。

基于重构的方法试图通过重构损失来检测异常。Park等[15]提出了长短时记忆变分编码器（LSTM-VAE）模型，该模型使用LSTM神经网络主干进行时间建模，并使用变分自编码器进行重建。Su等[4]进一步扩展了LSTM-VAE模型，采用归一化流程，利用重构概率来检测异常。Li等[16]提出交互融合（InterFusion）将主干更新为分层变分自动编码器，同时对多个序列之间的相互依赖和内部依赖进行建模。基于对抗训练的自动编码器，Audibert等[17]提出无监督异常检测（USAD），旨在快速稳定地从多元时间序列中识别异常。

虽然这2种方法实现了比传统方法更好的泛化能力，但是它们仅仅针对时间步去捕获时间依赖关系，这限制了它们在多元时间序列中检测少样本异常的能力。

除了上述方法外，还有一些研究工作[18-20]动态地构造时间序列图，并执行图预测或重建任务。例如，进化状态图网络（EvoNet）[18]提取具有代表性的多变量时间序列片段作为节点，并在训练过程中学习它们之间的转移概率。然而，该模型旨在捕捉不同时期段之间的时间模式变化，而不是序列间关系变化。多尺度卷积循环编解码器（MSCRED）[19]采用3个不同尺度的签名矩阵对多元时间序列间的依赖关系建模，并通过这3个签名矩阵的残差去评估异常的严重性，不仅诊断结果略微粗糙，还额外增加了计算量和存储开销。Chen等[20]提出用动态图来建模多元时间序列的依赖关系，但本质上不能对异常进行诊断和定位。

与现有模型仅仅单一建模时间依赖不同，本文提出采用状态矩阵来刻画一段多元时间序列中的时间和空间依赖关系，构建了三分支注意结构的MAD-Attention来捕获多元时间序列的序列、时间和空间的关联关系，实现了同时从序列、时间和空间3个维度互补地刻画多元时间序列，极大地提升了模型抽取特征的信息量；并设计了3个对齐关联函数来在特征空间对齐这3个关联关系以获取更具表达力的特征。最后，通过计算重构的状态矩阵和输入之间的残差来识别异常、定位异常最可能发生的位置以及细粒度地诊断异常的严重性。

2 动机与定义

2.1 动机描述

为了捕获有效和可解释的异常潜在状态。一个直观的方法是把一段时间序列作为可能的模式，然后建模它们之间的序列依赖[21]。然而，工业时间序列中通常存在复杂的业务空间上的依赖关系，是刻画系统运行状态的关键要素，这对刻画系统的状态很有用[22]。因此，可以采用基于图的方法[23-24]来建模时间序列间的结构依赖关系，但是这些方法通常不能提供对异常的严重性诊断，而这对帮助操作员采取恰当的响应是必要的。因此，本文提出了新颖的时间和业务依赖状态矩阵来描述多元时间序列之间的关联关系，并设计了三分支注意力模块来捕获状态矩阵随时间演变的关联模式。最后，重构了这些状态矩阵，并且计算出输入矩阵和重构矩阵之间的重构误差用以区分、定位和诊断异常。一般来说，现实世界中异常发生的位置可能有多个且异常的持续时间大于一个时间步，而模型无法很好地重构偏离系统的正常工作模式的状态矩阵。

具体地，本文提出时间状态矩阵用以刻画系统工作状态随时间的变化关系，认为当系统正常工作时，其状态随时间的变化符合一定的工作模式，这样不仅把建模时间序列本身转换为建模系统状态随时间的变化，能够更加有效地判别异常，还能够识别异常持续的时间。此外，在现实世界中，定位异常发生的位置对于运维人员快速响应异常具有重要意义。因此，为了定位异常发生的位置，本文还设计了空间状态矩阵用以刻画传感器之间的依赖关系以区分异常发生的最可能的位置。值得注意的是，在空间状态矩阵中，每行代表一个传感器，因此异常的行数越多，则表明异常发生的位置越多。另外，现有工作[25]表明系统传感器之间的相互关系

有助于刻画系统的局部状态。因此，本文提出从时间−空间 2 个角度分别刻画系统状态随时间的演化关系，用以细粒度地诊断及定位异常。

2.2　问题定义

给定一段多元时间序列 $\boldsymbol{x}=[x_1,x_2,\cdots,x_T]\in R^{T\times n}$ 作为训练集，它是在 T 个时间步内收集到的具有 n 个特征的序列。测试集由在不同于训练集的时间范围内收集的具有 n 个特征的多元时间序列组成。目标是从多元时间序列中检测出异常行为，并对异常结果进行定位和严重程度诊断，提供异常最可能发生的一个或多个位置，以及能以任意粒度的时间尺度对异常进行严重性评估。

3　方法

3.1　总体框架

如图 2 所示，本文提出了 MAD-Transformer 来发现更有意义的关联，并通过学习系统的时间和空间状态的变化，来建模系统正常工作的模式，从而更精准地发现异常行为。技术上，本文提出一个三分支注意力 MAD-Attention 来分别学习序列−时间−空间关联，同时设计了 3 个对齐损失以在特征空间中分别对齐这 3 个关联来获得更具表达力的特征。

假定 $x_t=[x_t^1,x_t^2,\cdots,x_t^n]\in R^{1\times n}$ 为时刻 t 对应的 n 个特征；$x_{\cdot}=[x_t^1,x_t^2,\cdots,x_t^n]\in R^{1\times n}$ 表示第 i 个特征在 w 这段时间的值。首先沿着时间维度，计算出两两时间步（如 x_{t1} 和 x_{t2}）之间的关联，形成时间状态矩阵 **TM**；然后沿着特征维度，计算出两两特征（如 x^i 和 x^j）之间的空间关联，形成空间状态矩阵 **SM**。之后 **TM**、**SM** 和原始时间序列 x 同时被输入拥有三分支注意结构 MAD-Attention 中，以训练模型从时间、空间和序列 3 个维度建模多元时间序列的正常模式，最后使用训练好的模型从测试集中检测出异常行为，并提供异常最可能发生的一个或多个位置，以及以任意粒度的时间尺度对异常的严重性进行评估。

3.2　时空依赖关系建模

与现有方法只建模时间序列之间的序列依赖不同，MAD-Transformer 从原始序列和所提出的状态矩阵中捕获以下 2 种类型的依赖关系。

1) 时间影响。CPS 的当前观测值不仅受即时物理过程影响，更与历史工作状态紧密关联，呈现出时序因果链特征。以水处理工厂为例，其传感器数据的动态变化严格遵循“注水−蓄水−排水”的时序逻辑，水箱液位与压力传感器的实时数值显著依赖于系统历史操作状态。

2) 空间影响。在 CPS 运行过程中，设备间存在空间域耦合效应。传感器数据变化受物理约束与控制逻辑影响，在空间上呈现复杂的联动影响关系。例如水处理场景中，水位传感器数值上升，则流速传感器数值就会下降。

为刻画 CPS 的时间动态演化规律和设备间的空间影响关系，分别构建时间状态矩阵和空间状态矩阵。时间状态矩阵通过刻画系统工作状态随时间的变化关系，考虑了 CPS 中物理过程随时间的演变规律。以水处理工厂为例，系统不同时间的水位、流

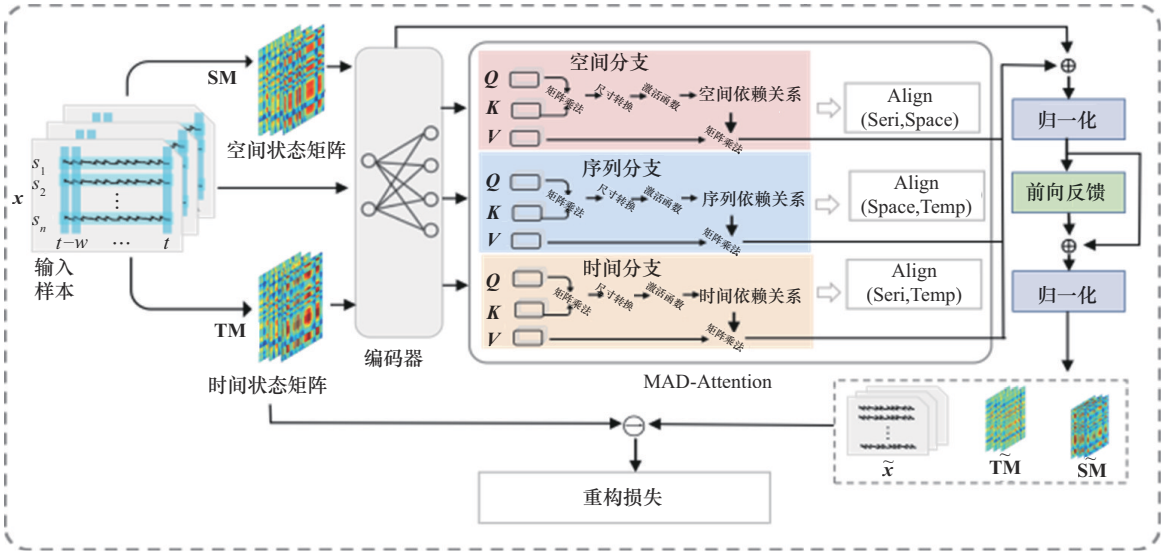


图 2　MAD-Transformer 的整体架构

速等状态变化存在内在联系，时间状态矩阵能够有效捕获这种联系，这是CPS在时间维度上物理过程动态变化在算法中的体现。空间状态矩阵用于刻画传感器之间的依赖关系。在CPS中，传感器分布于物理设备上，其数据相互关联，反映了物理系统的空间结构。如在复杂的工业生产线上，不同位置传感器监测的数据相互影响，空间状态矩阵可以精准定位异常发生位置，体现了CPS空间维度的特性。

1) 时间状态矩阵的构建。CPS作为计算进程和物理进程深度融合的系统，其运行状态在时间维度上存在着复杂且紧密的关联。为了刻画CPS在给定时间段（如长度为w的多元时间序列片段）的整体动态模式，构建了$w \times w$的时间状态矩阵**TM**，该矩阵基于此时间片段中两两时间步之间的内积进行构建。具体而言，考虑到CPS中物理过程的连续性和时间依赖性，给定长度为w的多元时间序列$\boldsymbol{x}=[x_{t-w},x_{t-w+1},\cdots,x_t]\in R^{w\times n}$，其中，$x_i=[x_i^1,x_i^2,\cdots,x_i^n]\in R^{1\times n}$表示时间步$i$（$i\in[t-w,t]$）所对应的多元时间序列，它蕴含了该时刻CPS中多个物理量（如传感器监测的温度、压力、流量等）以及相关计算处理后的综合信息。时间步i和时间步j之间的状态关联$T_{ij}\in\mathbf{TM}$的计算式为

$$T_{ij}=\frac{x_i\circ x_j}{\tau_{\mathrm{t}}} \tag{1}$$

其中，$\circ$表示点积，用于衡量2个时间步的多元时间序列在各个维度上的相似程度，反映了CPS中物理过程在不同时刻状态的关联程度；τ_{t}是时间状态矩阵的超参数，它可以根据CPS的具体特性（如系统的响应速度、物理量变化的频率等）进行调整，以更好地适配不同的CPS场景。时间状态矩阵**TM**的计算过程如图3所示。

在CPS的实际运行过程中，该矩阵每一行表示多元时间序列中相应的时间步与其他时间步之间的关联关系。由于CPS中物理过程和计算过程相互影响，当系统出现异常时，这种异常会通过物理量的异常变化或者计算逻辑的错误反馈到多元时间序列中，进而破坏时间步之间原本稳定的关联关系。在时间状态矩阵中，异常的行数越多，则意味着异常持续的时间越长。例如，在一个化工生产的CPS中，如果某个关键设备的温度传感器数据出现异常，那么在时间状态矩阵中，与该传感器数据相关的时间步所在行的关联关系会被破坏，异常持续的时间越长，涉及的异常行数就会越多。因此，通过建模多元时间序列的时间状态矩阵，不但能够发现异常，而且能够以任意时间粒度刻画异常的持续时间，这对于CPS的异常诊断至关重要。它能够紧密结合CPS中物理过程的时间特性，为后续异常严重性的细粒度诊断提供有力支持，使运维人员可以根据异常持续时间等信息，更准确地评估异常对整个系统物理运行状态的影响程度，从而采取更具针对性的措施来保障系统的稳定运行。

2) 空间状态矩阵的构建。在CPS复杂的运行环境下，空间状态矩阵的构建对于精准定位异常具有关键意义。CPS作为计算与物理过程深度融合的系统，其中大量设备相互协作，物理设备间的空间布局和相互作用关系极为复杂且紧密相关，这种特性决定了异常的发生往往与特定设备紧密相连。因此，在异常检测任务中仅仅给出异常的告警信息是不够的，还需要给出异常可能发生的位置，方便运维人员快速定位和排查原因，减少系统长时间宕机带来的损失。

为了定位异常，构建了$n\times n$的空间状态矩阵**SM**，来刻画多元时间序列的空间依赖关系。具体而言，给定长度为w的多元时间序列$\boldsymbol{x}=[x_{t-w},x_{t-w+1},\cdots,x_t]\in R^{w\times n}$，其中，$x^i=(x_{t-w}^i,x_{t-w+1}^i,\cdots,x_t^i)\in R^{w\times 1}$表示

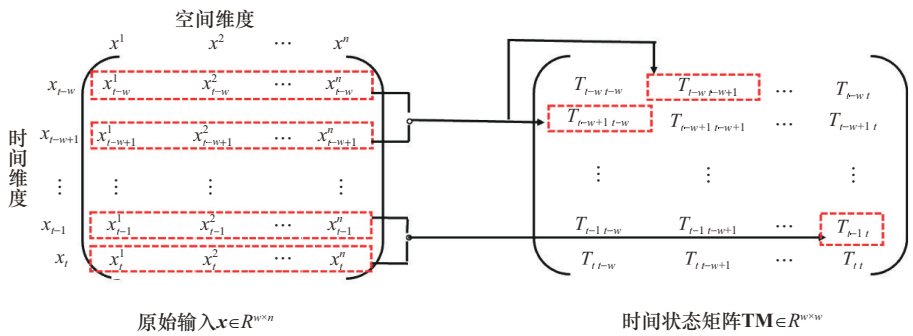


图3 时间状态矩阵的计算过程

CPS 中第 i 个序列（设备）在 w 这段时间内的状态，则序列 i 和序列 j 之间的依赖关系 $S_{ij} \in \mathbf{SM}$ 为

$$S_{ij} = \frac{x^i \circ x^j}{\tau_s} \tag{2}$$

其中，$\circ$ 表示点积，用于衡量 2 个序列在各个维度上的相似程度，这种相似程度体现了 CPS 中不同设备在运行状态上的关联紧密程度；τ_s 是空间状态矩阵的超参数，可依据 CPS 中设备的物理布局、功能耦合程度以及数据传输特性等进行调整。例如，在一个大型 CPS 中，对于空间位置相近且功能上相互配合紧密的设备，可适当调整 τ_s 的值，使空间状态矩阵更敏感地反映它们之间的依赖关系。

空间状态矩阵 **SM** 每行表示该行对应的序列（即某一设备的时间序列数据）与其他序列之间的空间关系。在 CPS 正常运行时，各设备间的关系遵循特定模式，这是由系统的物理设计和运行逻辑决定的。一旦某些设备的运行状态数据产生异常，就会打破这种正常模式下的依赖关系。例如，在一个自动化生产线上，当某一传感器出现故障时，其对应的序列数据会发生异常变化，在空间状态矩阵中，该传感器对应行与其他相关设备对应行之间的依赖关系度量值就会偏离正常范围。通过监测这些偏离情况，就可以捕获可能产生的异常点，进而定位异常发生的设备位置。图 4 给出了空间状态矩阵的计算过程，有助于理解这种空间依赖关系的构建逻辑。

3.3　多分支 Transformer

MAD-Transformer：传统的 Transformer 无法胜任多元时间序列异常检测的任务，因此本文提出 MAD-Transformer，用以检测和诊断多元时间序列中的异常，MAD-Transformer 的特点是多分支注意力块和前馈层交替叠加。这种叠加结构有利于从深层多层次特征中学习潜在的关联。

给定一段多元时间序列 $x \in R^{w \times n}$，根据式(1)和式(2)分别计算相应的时间状态矩阵 **TM** 和空间状态矩阵 **SM**。假设模型有 K 层，则第 k 层的所有式可以被形式化为

$$H^k = \mathrm{Norm}(\mathrm{MAD-Attention}([\boldsymbol{x},\mathbf{TM},\mathbf{SM}]^{k-1}) + [\boldsymbol{x},\mathbf{TM},\mathbf{SM}]^{k-1})$$

$$[\boldsymbol{x},\mathbf{TM},\mathbf{SM}]^k = \mathrm{Norm}(\mathrm{FeedForward}(H^k) + H^k) \tag{3}$$

其中，$x^k, \mathbf{TM}^k \in R^{w \times d_{\text{channel}}}$、$\mathbf{SM}^k \in R^{n \times d_{\text{channel}}}$、$k \in \{1,2,\cdots,K\}$ 表示通道的第 k 层的输出，d_{channel} 表示通道的输出维度。初始化输入 $[\boldsymbol{x},\mathbf{TM},\mathbf{SM}]^0 = \mathrm{Encoder}([x,\mathbf{TM},\mathbf{SM}])$ 表示特征的原始序列。H^k 是第 k 层的中间隐藏状态。MAD-Attention()用于计算各表征之间的关联关系。

MAD-Attention：由于单分支的自注意力机制[26]不能同时建模序列关联、时间关联和空间关联，提出具有三分支结构的时间空间注意力 MAD-Attention。其中，序列关联表示从原始输入中学习的序列级依赖关系，能够自适应地找到最有效的序列关联关系。对于时间关联，先基于两两时间步计算出时间状态矩阵，接着从该矩阵中提取出最有效的时间关联关系。同样地，对于空间关联，首先基于两两序列计算出相应的空间状态矩阵，用以表示原始的空间关联，进而从该矩阵中自适应地找到最显著的空间关联关系。请注意，这 3 种关联同时维护时间序列的序列依赖性、时间依赖性和空间依赖性，这比逐点表示能够提供更多信息。

如图 2 所示，MAD-Attention 包含 3 个分支，分别是序列分支、时间分支和空间分支。其中，序列分支用来学习时间序列的序列关联；时间分支用来从时间状态矩阵中捕获时间依赖关系；同样的，空间分支用以从空间状态矩阵显式地学习序列之间的空间相关性。这 3 个分支分别反映了时间序列中的

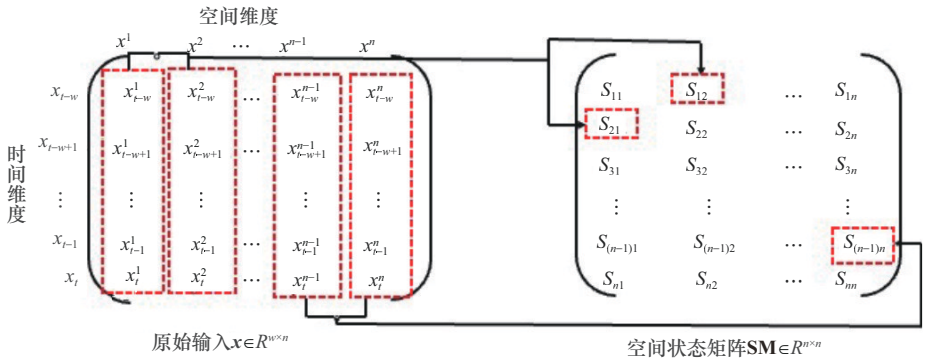


图 4　空间状态矩阵的计算过程

序列、时间和空间3个维度的关联，将三者结合可以学习到更具区分性的特征，因为违反其中的任意一种依赖关系都会引起异常。MAD-Attention的第k层输出如下。

初始化:

$$\boldsymbol{Q}_x,\boldsymbol{K}_x,\boldsymbol{V}_x = x^{k-1}\boldsymbol{w}^k_{\boldsymbol{Q}_x},x^{k-1}\boldsymbol{w}^k_{\boldsymbol{K}_x},x^{k-1}\boldsymbol{w}^k_{\boldsymbol{V}_x}$$

$$\boldsymbol{Q}_{\mathrm{TM}},\boldsymbol{K}_{\mathrm{TM}},\boldsymbol{V}_{\mathrm{TM}} = \mathbf{TM}^{k-1}\boldsymbol{w}^k_{\boldsymbol{Q}_{\mathrm{TM}}},\mathbf{TM}^{k-1}\boldsymbol{w}^k_{\boldsymbol{K}_{\mathrm{TM}}},\mathbf{TM}^{k-1}\boldsymbol{w}^k_{\boldsymbol{V}_{\mathrm{TM}}}$$

$$\boldsymbol{Q}_{\mathrm{SM}},\boldsymbol{K}_{\mathrm{SM}},\boldsymbol{V}_{\mathrm{SM}} = \mathbf{SM}^{k-1}\boldsymbol{w}^k_{\boldsymbol{Q}_{\mathrm{SM}}},\mathbf{SM}^{k-1}\boldsymbol{w}^k_{\boldsymbol{K}_{\mathrm{SM}}},\mathbf{SM}^{k-1}\boldsymbol{w}^k_{\boldsymbol{V}_{\mathrm{SM}}}$$

序列关联:$S^k = \mathrm{Softmax}\left(\dfrac{\boldsymbol{Q}_x(\boldsymbol{K}_x)^{\mathrm{T}}}{\sqrt{d_{\mathrm{model}}}}\right)$

时间关联:$\mathrm{Temp}^k = \mathrm{Softmax}\left(\dfrac{\boldsymbol{Q}_{\mathrm{TM}}(\boldsymbol{K}_{\mathrm{TM}})^{\mathrm{T}}}{\sqrt{d_{\mathrm{model}}}}\right)$

空间关联:$\mathrm{Space}^k = \mathrm{Softmax}\left(\dfrac{\boldsymbol{Q}_{\mathrm{SM}}(\boldsymbol{K}_{\mathrm{SM}})^{\mathrm{T}}}{\sqrt{d_{\mathrm{model}}}}\right)$

重构:$\tilde{H}^k_{\boldsymbol{x}} = S^k\boldsymbol{V}_x,\tilde{H}^k_{\mathbf{TM}} = \mathrm{Temp}^k\boldsymbol{V}_{\mathbf{TM}}$

$$\tilde{H}^k_{\mathbf{SM}} = \mathrm{Space}^k\boldsymbol{V}_{\mathbf{SM}} \tag{4}$$

其中，$\boldsymbol{Q}_x,\boldsymbol{K}_x,\boldsymbol{V}_x,\boldsymbol{Q}_{\mathrm{TM}},\boldsymbol{K}_{\mathrm{TM}},\boldsymbol{V}_{\mathrm{TM}} \in R^{W\times d_{\mathrm{channel}}}$、$\boldsymbol{Q}_{\mathrm{SM}},\boldsymbol{K}_{\mathrm{SM}},\boldsymbol{V}_{\mathrm{SM}} \in R^{n\times d_{\mathrm{channel}}}$分别表示序列分支、时间分支和空间分支的查询、键和值，$\boldsymbol{w}^k_{\boldsymbol{Q}_x},\boldsymbol{w}^k_{\boldsymbol{K}_x},\boldsymbol{w}^k_{\boldsymbol{V}_x},\boldsymbol{w}^k_{\boldsymbol{Q}_{\mathrm{TM}}},\boldsymbol{w}^k_{\boldsymbol{K}_{\mathrm{TM}}},\boldsymbol{w}^k_{\boldsymbol{V}_{\mathrm{TM}}},\boldsymbol{w}^k_{\boldsymbol{Q}_{\mathrm{SM}}},\boldsymbol{w}^k_{\boldsymbol{K}_{\mathrm{SM}}},\boldsymbol{w}^k_{\boldsymbol{V}_{\mathrm{SM}}}$分别表示$Q_x$、$K_x$、$V_x$、$\boldsymbol{Q}_{\mathrm{TM}}$、$\boldsymbol{K}_{\mathrm{TM}}$、$\boldsymbol{V}_{\mathrm{TM}}$、$\boldsymbol{Q}_{\mathrm{SM}}$、$\boldsymbol{K}_{\mathrm{SM}}$、$\boldsymbol{V}_{\mathrm{SM}}$在第$k$层的参数矩阵，$S^k$、$\mathrm{Temp}^k$、$\mathrm{Space}^k$分别表示序列关联、时间关联和空间关联，Softmax(·)表示沿着最后的维度规范化注意力图。因此，上述注意力图(S^k、Temp^k、Space^k)中，每行遵从了一个离散的分布。$\tilde{H}^k_{\boldsymbol{x}}$、$\tilde{H}^k_{\mathbf{TM}}$、$\tilde{H}^k_{\mathbf{SM}}$是第$k$层继MAD-Attention后的中间表示，使用MAD-Attention()来概括式(4)。

3.4 损失度量

3.4.1 关联(差异)对齐损失

采用对称的卡尔曼-莱布勒（KL）散度来对齐两两关联，这也表示这2个分布之间的信息增益[27]。对多层的关联差异进行平均，将多层特征的关联组合成一个更有信息量的度量。

$$\mathrm{Ali}(\mathrm{Seri},\mathrm{Temp}) = [\frac{1}{K}\sum_k^K(\mathrm{KL}(\mathrm{Seri}^k_{i,:}\|\mathrm{Temp}^k_{i,:}) + \mathrm{KL}(\mathrm{Temp}^k_{i,:}\|\mathrm{Seri}^k_{i,:}))]_{i=1,2,\cdots,w} \tag{5}$$

$$\mathrm{Ali}(\mathrm{Seri},\mathrm{Space}) = [\frac{1}{K}\sum_k^K(\mathrm{KL}(\mathrm{Seri}^k\|\mathrm{Space}^k) + \mathrm{KL}(\mathrm{Space}^k\|\mathrm{Seri}^k))] \tag{6}$$

$$\mathrm{Ali}(\mathrm{Temp},\mathrm{Space}) = [\frac{1}{K}\sum_k^K(\mathrm{KL}(\mathrm{Space}^k\|\mathrm{Temp}^k) + \mathrm{KL}(\mathrm{Temp}^k\|\mathrm{Space}^k))] \tag{7}$$

$$\mathrm{Ali}_{\mathrm{total}} = \|\mathrm{Ali}(\mathrm{Seri},\mathrm{Temp})\|_1 + \|\mathrm{Ali}(\mathrm{Seri},\mathrm{Space})\|_1 + \|\mathrm{Ali}(\mathrm{Temp},\mathrm{Space})\|_1 \tag{8}$$

式(5)中的KL(· ‖ ·)为对应于i和j的每一行的2个离散分布之间的KL散度，因此，Ali(Seri, Temp) $\in R^{w\times 1}$是序列关联和时间关联的逐点关联差异，结果的第i个元素对应于输入序列x的第i个时间步。式(6)和式(7)中的KL(· ‖ ·)对应的是2个分布之间的关联差异，因为空间关联的维度和序列关联以及时间关联的维度不同。式(8)中的$\|\cdot\|_1$表示L1范数。

3.4.2 重构损失

作为一个无监督任务，采用重构损失来优化模型。分别计算原始输入序列x、时间矩阵**TM**和空间矩阵**SM**的重构损失。这些重构损失将指导各个关联以找到最具信息量的信息。

$$L_R = \|\boldsymbol{x} - \tilde{\boldsymbol{x}}\|_{\mathrm{F}}^2 + \|\mathbf{TM} - \widetilde{\mathbf{TM}}\|_{\mathrm{F}}^2 + \|\mathbf{SM} - \widetilde{\mathbf{SM}}\|_{\mathrm{F}}^2 \tag{9}$$

其中，$\tilde{\boldsymbol{x}}$、$\widetilde{\mathbf{TM}}$和$\widetilde{\mathbf{SM}}$分别表示序列$\boldsymbol{x}$、时间矩阵**TM**和空间矩阵**SM**的重构，$\|\cdot\|_{\mathrm{F}}$表示Frobenius范数。

另外，请注意，时间状态矩阵和空间状态矩阵分别反映了时间序列在时间维度和空间维度的关联关系，一旦有某个时刻的序列出现异常，就必然会违反这些关联关系至少其中的一个，因此这里重构两者是为了可视化一个异常持续的时长，以及寻找哪些序列出现了异常，用以可视化异常的严重程度和确定异常发生的位置。

3.4.3 总体损失

总体损失函数包含2个部分：重构损失L_R和关联对齐损失$\mathrm{Ali}_{\mathrm{total}}$。重构损失将指导序列、时间和空间3个关联找到信息量最大的关联。同时，为了进一步放大异常和正常时间点之间的差异，本文还使用了一个额外的差异损失来对齐这3个关联，让3个关联互相影响、互为补充，从而学习到更加全面的正常数据的模式，进而使得异常重构更加困难，异常更容易被识别。根据式(8)和式(9)，输入序列x的总体损失函数可以形式化为

$$L_{\mathrm{total}} = L_R + \lambda\mathrm{Ali}_{\mathrm{total}} \tag{10}$$

其中，λ是用来平衡损失项的超参数。

4　异常诊断

异常的诊断过程如图 5 所示，**SM** 和 **TM** 分别表示空间状态矩阵和时间状态矩阵，$\tilde{x}$、$\widetilde{\mathbf{TM}}$ 和 $\widetilde{\mathbf{SM}}$ 分别表示重构的时间序列、时间状态矩阵和空间状态矩阵。et_{ij} 表示残差时间状态矩阵中的第 i 行第 j 列的误差值；同样的，es_{ij} 表示残差空间状态矩阵的第 i 行第 j 列的误差值。et_i 和 es_i 分别表示残差时间状态矩阵和残差空间状态矩阵中第 i 行的误差和。矩阵每行的误差和与相应的阈值比较，大于阈值的行被识别为异常的时间步（残差时间状态矩阵）或者异常的设备（残差空间状态矩阵）。

4.1　异常得分

由于序列关联和时间关联的差异是逐时间点计算的，为了同时利用时间序列特征和可区分的关联差异的优点，将归一化的序列关联和时间关联的差异纳入异常得分的计算。因此，给定多元时间序列 $\boldsymbol{x} \in R^{w \times n}$，其最终的异常得分为

$$\mathrm{Sc} = \|x_{i,:} - \tilde{x}_{i,:}\|_2^2 \otimes \mathrm{Norm}(-\mathrm{Ali}(\mathrm{Seri},\mathrm{Temp})) \quad (11)$$

其中，$i = [1,2,\cdots,w]$，$\otimes$ 为逐元素乘法。面对一个更好的重构，异常可能会降低关联差异，但仍然会得到更高的异常得分。因此这种设计可以使重构误差和关联差异协同工作，提高检测的准确率。

4.2　异常定位

通过计算原始空间状态矩阵和其重构之间的残差矩阵来判断异常最可能发生的位置。给定空间状态矩阵 **SM** 和重构空间状态矩阵 $\widetilde{\mathbf{SM}}$，残差空间状态矩阵为

$$\mathbf{RS} = \|\mathbf{SM} - \widetilde{\mathbf{SM}}\|_2^2 \otimes \mathrm{Norm}(-\mathrm{Ali}(\mathrm{Seri},\mathrm{Space})) \quad (12)$$

计算残差空间状态矩阵每行的误差和作为该行最终的误差。误差大于阈值的行被认为相应的设备可能是异常的。

4.3　异常严重性评估

就像上文所述，异常的严重性诊断，是通过判断异常的持续时间确定的，持续的时间越长，则异常越严重。与异常的定位类似，对异常的诊断是通过计算时间状态矩阵和其重构之间的残差时间状态矩阵来进行的。给定时间状态矩阵 **TM**，残差时间状态矩阵为

$$\mathbf{RT} = \|\mathbf{TM} - \widetilde{\mathbf{TM}}\|_2^2 \otimes \mathrm{Norm}(-\mathrm{Ali}(\mathrm{Seri},\mathrm{Temp})) \quad (13)$$

之后，对 **RT** 按行求和，作为该行的最终误差。误差大于阈值的行被认为对应的时间步是异常的，异常的时间步越多，则表明该异常越严重。

在实际应用中，可以根据系统的特性来自适应地选择判断异常严重性的粒度，例如时间状态矩阵的每行可以表示一个时间步的间隔，也可以表示 10 个时间步的间隔。MAD-Transformer 提供了异常的识别、定位和诊断 3 个功能，在实际应用中，可以根据业务需求自适应地选择其中的某一种或者两种功能。不过，将异常定位和异常诊断结合起来判断异常的严重性，结果可能会更全面、更有意义。因为直观上来说，异常持续的时间越长，发生异常的设备越多，则该异常可能越严重。

5　实验

本节进行广泛的实验以回答以下问题。

1) 异常检测。MAD-Transformer 在多元时间序列异常检测任务中是否超越了基准模型（RQ1）？MAD-Transformer 的每个组件对其检测性能的影响（RQ2）？

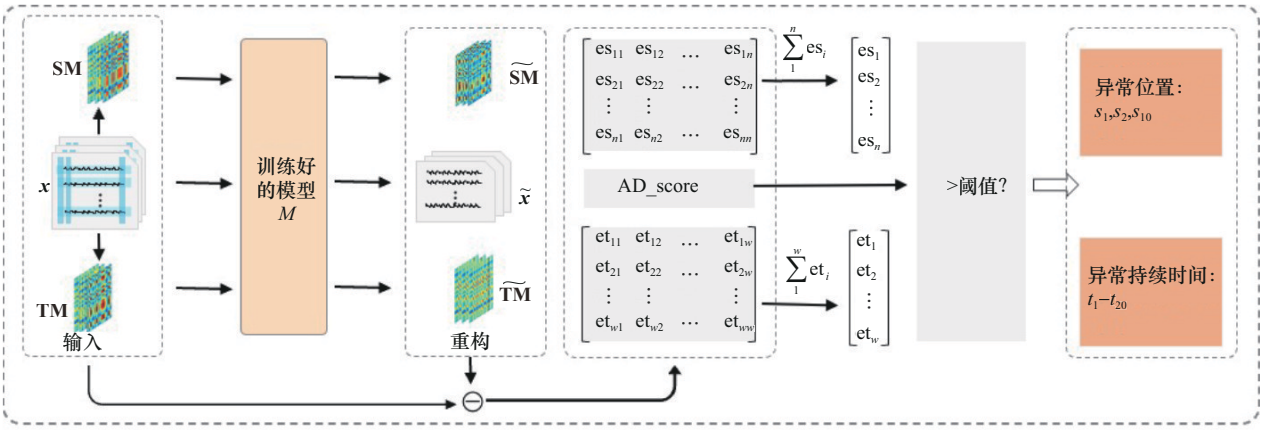


图5　异常的诊断过程

2) 异常定位和诊断。MAD-Transformer能否定位异常最可能发生的位置（RQ3），以及能否给出有效的异常严重性的判断（RQ4）?

5.1 数据集

安全水处理测试平台（SWAT）[28]是由新加坡科技设计大学的iTrust机构从包含了51个传感器和执行器的水处理测试平台系统连续运行14天获得的，其中前10天系统正常运行，后4天人为地向系统施加了模拟的物理攻击和网络攻击。该数据集的训练集包含495 000条正常记录，测试集包含449 921条正常与攻击的记录，每条记录均包含一个标签和51个执行器与传感器值的属性。

服务器机器数据集（SMD）[4]是一个从一家大型互联网公司的28个机器，收集的长达5周的数据构成的数据集，相邻两组数据的间隔为1 min。该数据集的训练集有708 405条记录，测试集有708 420条记录，每条记录包含38个维度。

土壤湿度主动被动卫星数据（SMAP）和火星科学实验室探测器数据（MSL）[5]是2个由美国航天航空局从2个不同的产生遥测数据流的航天器收集的多元时间序列数据集，收集的时间间隔是1 min，所有数据已经全被缩放到[0,1]。SMAP的训练集有135 183条记录，测试集有427 617条记录，每条记录有25个维度。MSL的训练集和测试集分别包含58 317和73 729条记录，每条记录55个维度。

池服务器指标（PSM）[29]是从易贝（eBay）的多个应用服务器节点内部收集的数据集，它的训练集和测试集分别包含132 481条和87 841条记录，每条记录包含一个标签和24个特征值，特征值表示服务器节点的各种性能指标，如中央处理器（CPU）使用率、内存使用率等。

耦合油罐控制系统平台（CTCS）收集自实验室搭建的耦合油罐控制系统仿真平台，包含7个维度。

5.2 基准方法

将MAD-Transformer与6个类别下的21个基准模型进行比较，即传统分类模型：孤立森林（IF）和一分类支持向量机模型（OCSVM）；基于重构的模型：ATransformer[30]、基于对抗Transformer的无监督异常检测模型（ATUAD）[31]、多粒动态接受域模型（MGDRF）[32]、多元异常检测生成对抗网络模型（MAD-GAN）[33]、无监督异常检测模型（USAD）[17]、交互融合模型（InterFusion）[16]、反向节律对抗生成网络模型（BeatGAN）[34]、随机循环异常检测（OMNIANOMALY）[4]、多尺度卷积循环编解码器（MSCRED）[19]，长短时记忆变分编码器（LSTM-VAE）[15]；基于密度估计模型：深度自编码高斯混合模型（DAGMM）[10]、混合概率主成分分析和分类分布（MPPCACD）[11]、局部离群因子（LOF）[8]；基于聚类的模型：基于张量的集成异常检测系统（ITAD）[14]、时间层次单类（THOC）[13]，深度支持向量数据描述（Deep-SVDD）[12]；基于自回归的模型：长短时记忆网络（LSTM）[5]、多元卷积空间异常检测（CL-MPPCA）[3]、变分异常分析（VAR）[35]；基于图的模型：基于图注意力的异常检测模型（MTAD-GAT）[10]和基于图深度网络的检测模型（GDN）[24]。

5.3 评估指标

采用精准率、召回率和F1得分3个指标来评估MAD-Transformer和基准模型的异常检测性能。为了检测异常，通过使验证数据集的r比例数据标记为异常集来得到一个阈值δ。异常分数大于该阈值的时间步被标记为异常。另外，对于主要结果，SMD数据集设置r=0.5%，SWAT数据集设置r=0.1%，其他数据集设置r=1%。

此外，遵循点调整策略，即只要在某个连续的时间段内检测到一个时间步为异常[36]，则认为该时间段内所有的异常都被正确检测到。该策略是合理的，一旦检测到异常将会引发警报，进而使整个时间段内的异常在实际应用中被注意和观察到。

5.4 实现细节

为了验证MAD-Transformer的有效性，基于单片NVIDIAA100-SXM4-80GB GPU显卡，采用PyTorch深度学习框架，在5个真实场景数据集上进行了对比实验验证。为了避免过拟合，从训练集中分离20%的样本作为验证集。根据Shen等[13]的完善方案，采用非重叠滑动窗口来获得一组子序列。对于所有数据集，滑动窗口的大小固定为100。MAD-Transformer总共包含3层，序列分支、时间分支和空间分支不共享参数。设置头为8，隐藏状态的通道数为512。对于所有的数据集，将式(10)中的λ设置为19，以权衡损失函数的2个部分。采

用 ADAM 作为优化器，初始学习率为 1×10^{-4} 批大小为 64，训练过程在 10 个批次内提前停止。

5.5　结果和讨论

5.5.1　异常检测(RQ1:与基线模型比较)

不同模型在 5 个公开数据集上的异常检测结果如表 1 所示，用加粗+下划线标注最优结果，用加粗标注次优结果。根据召回率和精确率计算 F1 得分来评估模型的性能，更高的 F1 得分表示更优的性能。

1) 公开数据集：在 5 个公开数据集上与 21 个基准模型进行了广泛的对比实验。在所有公开数据集上，MAD-Transformer 的平均 F1 得分性能优于包括 ATransformer 在内的大多数基准模型，验证了建模时间序列中空间依赖和邻接集中的关联对异常检测的有效性。

具体而言，在 SMAP、SWAT、PSM 数据集上，MAD-Transformer 取得最高 F1 得分，在 MSL 数据集上的性能与 ATUAD 相当，且显著优于其他模型。在 SMD 数据集上，MAD-Transformer 的性能略低于 ATransformer、ATUAD 和 GDN。MAD-Transformer 的检测性能在大部分数据集上优于上述模型，表明显式建模序列关联、时间关联和空间

表 1　在 5 个公开数据集上的异常检测对比结果

评估指标	SMD			MSL			SMAP			SWAT			PSM			平均
	F1 得分	精确率	召回率	F1 得分	精确率	召回率	F1 得分	精确率	召回率	F1 得分	精确率	召回率	F1 得分	精确率	召回率	F1 得分
OCSVM	56.19%	44.34%	76.72%	70.82%	59.78%	86.87%	56.34%	53.85%	59.07%	47.23%	45.39%	49.22%	70.67%	62.75%	80.89%	60.25%
IF	53.64%	42.31%	73.29%	66.45%	53.94%	86.54%	55.53%	52.39%	59.07%	47.02%	49.29%	44.95%	83.48%	76.09%	92.45%	61.22%
LOF	46.68%	56.34%	39.86%	61.18%	47.72%	85.25%	57.60%	58.93%	56.33%	68.62%	72.15%	65.43%	70.61%	57.89%	90.49%	60.94%
Deep-SVDD	79.10%	78.54%	79.67%	83.58%	91.92%	76.63%	69.04%	89.93%	56.02%	82.39%	80.42%	84.45%	90.73%	95.41%	86.49%	80.97%
DAGMM	57.30%	67.30%	49.89%	74.62%	89.60%	63.93%	68.51%	86.45%	56.73%	70.40%	89.92%	57.84%	80.08%	93.49%	70.03%	70.18%
MPPCACD	75.02%	71.20%	79.28%	69.95%	81.42%	61.31%	81.73%	88.61%	75.84%	74.73%	82.52%	68.29%	77.29%	76.26%	78.35%	75.74%
VAR	74.08%	78.35%	70.26%	77.90%	74.68%	81.42%	64.83%	81.38%	53.88%	69.34%	81.59%	60.29%	87.13%	90.71%	83.82%	74.66%
LSTM	81.78%	78.55%	85.28%	83.95%	85.45%	82.50%	83.39%	89.41%	78.13%	84.69%	86.15%	83.27%	82.80%	76.93%	89.64%	83.32%
CL-MPPCA	79.09%	82.36%	76.07%	80.44%	73.71%	88.54%	72.88%	86.13%	63.16%	79.07%	76.78%	81.50%	71.80%	56.02%	99.93%	76.66%
ITAD	79.48%	86.22%	73.71%	76.07%	69.44%	84.09%	73.85%	82.42%	66.89%	57.08%	63.13%	52.08%	68.13%	72.80%	64.02%	70.92%
LSTM-VAE	82.30%	75.76%	90.08%	82.62%	85.49%	79.94%	78.10%	92.20%	67.75%	82.20%	76.00%	89.50%	80.96%	73.62%	89.92%	81.24%
BeatGAN	78.10%	72.90%	84.09%	87.53%	89.75%	85.42%	69.61%	92.38%	55.85%	73.92%	64.01%	87.46%	92.04%	90.30%	93.84%	80.24%
OMNIANOMALY	85.22%	83.68%	86.82%	87.67%	89.02%	86.37%	86.92%	92.49%	81.99%	82.83%	81.42%	84.30%	80.83%	88.39%	74.46%	84.69%
InterFusion	86.22%	87.02%	85.43%	86.62%	81.28%	92.70%	89.14%	89.77%	88.52%	83.01%	80.59%	85.58%	83.52%	83.61%	83.45%	85.70%
THOC	84.99%	79.76%	90.95%	89.69%	88.45%	90.97%	90.68%	92.06%	89.34%	85.13%	83.94%	86.36%	89.54%	88.14%	90.99%	88.01%
MAD-GAN	85.10%	85.96%	84.25%	91.38%	85.55%	98.07%	88.14%	**94.22%**	82.79%	86.53%	90.85%	82.60%	87.90%	88.25%	87.56%	87.81%
MSCRED	78.75%	84.73%	79.11%	82.05%	87.37%	77.34%	69.33%	94.11%	54.88%	84.59%	92.69%	77.80%	83.78%	87.23%	80.60%	79.70%
MTAD-GAT	91.29%	90.24%	92.36%	90.84%	87.54%	94.40%	90.13%	89.06%	91.23%	85.50%	98.24%	75.69%	88.41%	**97.95%**	80.56%	89.23%
GDN	**92.59%**	**91.08%**	94.16%	94.62%	**94.91%**	94.34%	92.31%	91.64%	92.98%	89.57%	96.14%	83.84%	93.59%	95.67%	91.59%	92.54%
USAD	91.62%	89.14%	94.25%	92.72%	88.10%	97.86%	86.34%	76.97%	98.31%	84.60%	**98.70%**	74.02%	86.99%	96.89%	78.92%	88.45%
ATransformer	92.33%	89.40%	95.45%	93.35%	**92.00%**	94.73%	**96.25%**	93.61%	99.05%	**94.07%**	91.55%	**96.73%**	**97.89%**	96.91%	**98.90%**	**94.78%**
MGDRF	92.03%	88.64%	**95.69%**	89.99%	89.43%	90.56%	87.37%	80.45%	95.59%	82.70%	**98.32%**	71.35%	88.03%	94.56%	82.35%	88.02%
ATUAD	**95.02%**	**90.72%**	**99.74%**	**94.96%**	90.41%	**99.99%**	89.85%	81.57%	**99.99%**	81.01%	96.96%	69.57%	94.91%	96.63%	93.25%	91.15%
MAD-Transformer	92.00%	90.15%	93.93%	**94.92%**	91.98%	**98.06%**	**96.94%**	**95.02%**	**98.96%**	**95.83%**	94.65%	**97.04%**	**98.18%**	**97.45%**	**98.92%**	**95.57%**

关联对提升异常检测性能的明显优势。

深入分析发现，MAD-Transformer的优势源于对多元时间序列中先验关联与序列关联的联合建模，实现了更完备的特征空间表征。相较之下，ATUAD与MGDRF本质上仅捕获了序列特征，而未能显式建模多元变量间的空间关联关系，这可能导致其特征表示的丰富度不足，进而影响了异常检测的泛化性能。

2) CTCS数据集：在该数据集上的检测结果如图6所示，MAD-Transformer仍然可以达到最先进的性能，进一步验证了模型的有效性。

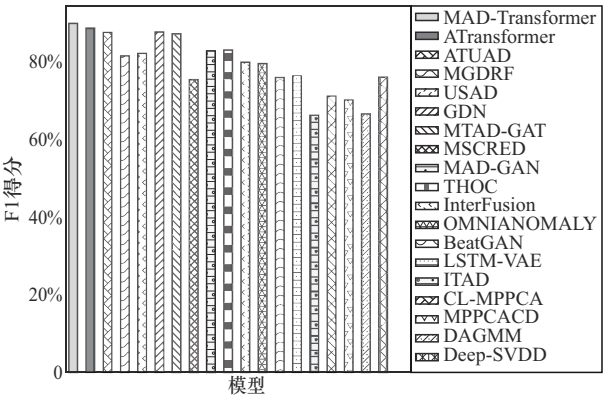


图6　在CTCS数据集上的对比实验结果

5.5.2　消融实验(RQ2:与模型的变体比较)

表2进一步研究了MAD-Transformer每个模块的影响，分别从模型中去除时间分支、去除空间分支、去除序列分支、去除对齐损失，并与纯Transformer进行性能对比。从表2中可以看出，MAD-Transformer的每个模块都可以进一步改进模型。具体而言，建模时间关联显著提升了5.01个百分点（90.56%→95.57%），建模空间关联显著提升了3.88个百分点（91.69%→ 95.57%），以及建模序列关联显著提升了5.91个百分点（89.66%→95.57%）的平均F1得分。另外，设计的对齐关联的损失也增强了模型的性能（89.66%→95.57%）。最终，MAD-Transformer比纯Transformer提升17.02个百分点（78.55%→95.57%）的绝对提升。结果验证了所设计各模块的必要性和有效性。

5.5.3　异常定位(RQ3)

异常定位依赖于良好的异常检测性能。因此，比较了MAD-Transformer和2个最佳基线（ATransformer和ATUAD）的异常检测性能。具体而言，对于ATransformer和ATUAD，使用时间序列的重构误差来表示序列的异常得分。MAD-Transformer的相同值定义为残差空间状态矩阵的特定行/列中重构不良的成对相关性的数量，因为每一行/列表示一条时间序列（设备）。对于每个异常事件，根据其异常分数对所有时间序列进行排序，并将前k个序列确定为异常发生的位置。图7为5次重复实验的平均召回率Recall@k（k=3），如图7所示，在前述的6个数据集中，MAD-Transformer平均比ATransformer和ATUAD高出约10.32%和10.92%。

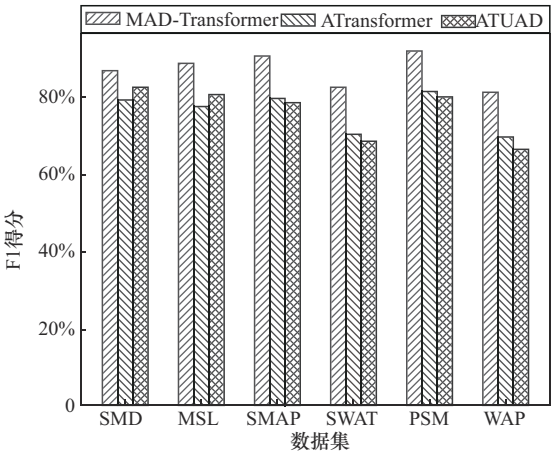


图7　6个数据集上的异常定位性能对比

表2　MAD-Transformer在5个公开数据集上的消融实验结果

数据集	MAD-Transformer	去除时间分支	去除空间分支	去除序列分支	去除对齐损失	纯Transformer
SMD	92.00%	88.25%	89.65%	86.39%	90.96%	76.89%
MSL	94.92%	89.06%	90.98%	88.56%	92.85%	80.66%
SMAP	96.94%	90.89%	92.31%	90.68%	94.64%	79.53%
SWAT	95.83%	90.25%	91.56%	90.05%	92.89%	78.72%
PSM	98.18%	94.34%	93.97%	92.61%	95.59%	76.94%
平均F1得分	95.57%	90.56%	91.69%	89.66%	93.39%	78.55%

5.5.4　细粒度的异常诊断（RQ4）

MAD-Transformer 的时间特征矩阵行数为滑动窗口的大小，可以捕获滑动窗口时间长度的系统状态。为了解释异常的严重程度，首先基于残差时间状态矩阵计算每一行的异常分数。每一行表示一个时间步，因此，MAD-Transformer 可以评估持续时间在滑动窗口（可配置）内任意长度的异常。

为了评估和展示模型对异常的检测和诊断的有效性，图 8 提供了在 CTCS 数据集上的异常诊断案例研究，图 8 上面一行是残差时间状态矩阵。

CTCS 中包含 6 个异常（A_1~A_6），每个异常持续的时间分别为 10 s、50 s、60 s、40 s、20 s 和 30 s。在这种情况下，如图 8 所示，MAD-Transformer 可以检测到所有 6 个异常。而且，注入异常事件的 6 个残差时间状态矩阵给出了对异常持续时间的诊断结果。残差时间状态矩阵中异常的行/列用粗线突出表示，每个粗线标注的行/列表示一个异常的时间步。粗线覆盖的行/列越多，表明异常持续的时间越长。如图 8 所示，MAD-Transformer 能以大约 92% 的准确率识别出异常持续的时间，而且所识别出的异常按照持续时间（即严重性）排序为 $A_3 > A_2 > A_4 > A_6 > A_5 > A_1$，与预设的结果一致。因此，在这种情况下，MAD-Transformer 可以准确地判断出各个异常的严重程度。

另外，MSCRED 虽能评估异常持续时间，但其基于 10 s、30 s、60 s 时间尺度矩阵计算残差获取异常得分判断异常。该模型仅能粗略识别持续时间大于或等于矩阵时间尺度的异常，对短于尺度的异常检测能力不足。例如，以 CTCS 数据集中 A_2（持续时间为 50 s）为例，MAD-Transformer 可识别其持续时间约 45 s，MSCRED 只能判断其持续时间大于 30 s，不能识别出其具体持续了多长时间，凸显了 MAD-Transformer 细粒度诊断的优势。

6　结束语

针对无监督多元时间序列异常检测任务，基于对时间序列的关联性分析，设计时间状态矩阵和空间状态矩阵，分别建模系统在时间和空间维度上的状态关联，提出一个三分支结构的异常检测模型 MAD-Transformer 来同时显式地捕获多元时间序列的序列-时间-空间依赖关系，利用残差时间状态矩阵和残差空间状态矩阵进行异常的定位和诊断。在 5 个公开数据集和 1 个耦合油罐控制系统数据集上广泛的实证研究表明，MAD-Transformer 不但在检测性能上优于最先进的基线方法，还能够精准地定位异常，以及给出异常持续的时间。

尽管 MAD-Transformer 在信息物理系统异常诊断方面成果显著，但仍有提升空间。未来将从以下方面优化：改进模型的结构，提升复杂场景适应性和运算速度；拓展应用场景，跨行业推广并融合边缘计算与物联网；强化数据处理能力，应对数据缺失与不完整，挖掘多源异构数据价值；增强可解释性与透明度，解释诊断结果并建立信

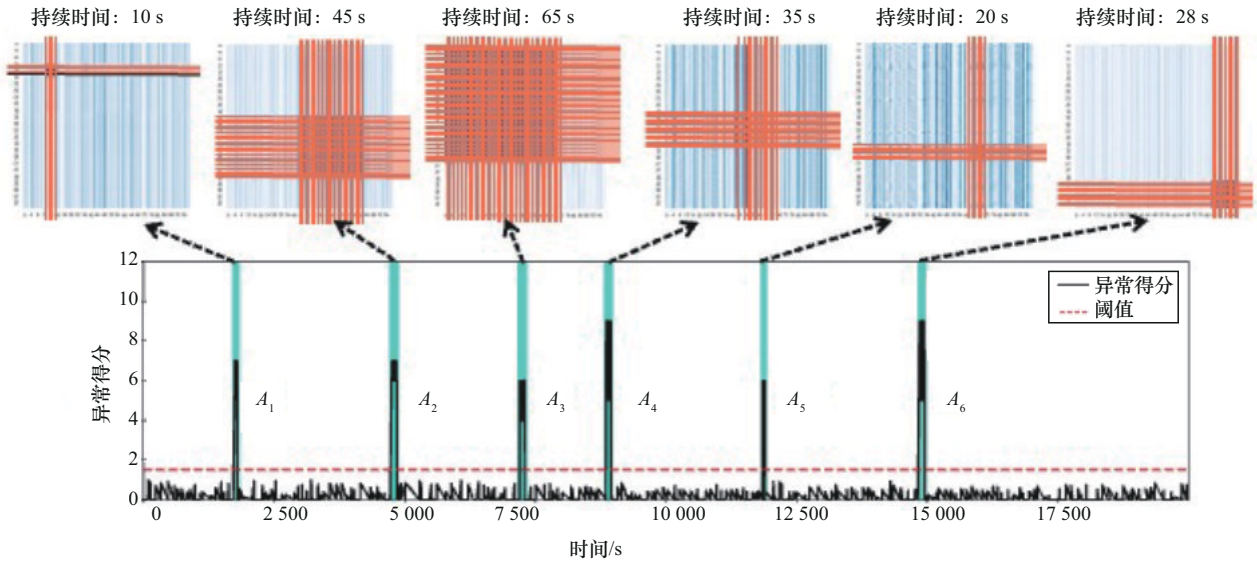


图 8　异常诊断案例分析

任评估机制，推动该技术在工业领域更广泛、更可靠地应用。

参考文献：

[1] DAUGHERTY P, BERTHON B. Winning with the industrial Internet of things: how to accelerate the journey to productivity and growth[R]. 2015.

[2] 孙海丽, 龙翔, 韩兰胜, 等. 工业物联网异常检测技术综述[J]. 通信学报, 2022, 43(3): 196-210.
SUN H L, LONG X, HAN L S, et al. Overview of anomaly detection techniques for industrial Internet of Things[J]. Journal on Communications, 2022, 43(3): 196-210.

[3] TARIQ S, LEE S, SHIN Y, et al. Detecting anomalies in space using multivariate convolutional LSTM with mixtures of probabilistic PCA[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2019: 2123-2133.

[4] SU Y, ZHAO Y J, NIU C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2019: 2828-2837.

[5] HUNDMAN K, CONSTANTINOU V, LAPORTE C, et al. Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 387-395.

[6] SUN H L, HUANG Y, ZHOU C J, et al. Space decoupled prototype learning for few-shot attack detection in cyber–physical systems[J]. IEEE Transactions on Industrial Informatics, 2024, 20(10): 12350-12362.

[7] HAUTAMAKI V, KARKKAINEN I, FRANTI P. Outlier detection using k-nearest neighbour graph[C]//Proceedings of the 17th International Conference on Pattern Recognition, 2004. Piscataway: IEEE Press, 2004: 430-433.

[8] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF[J]. ACM SIGMOD Record, 2000, 29(2): 93-104.

[9] TANG J, CHEN Z X, FU A W, et al. Enhancing effectiveness of outlier detections for low density patterns[C]//Advances in Knowledge Discovery and Data Mining. Berlin: Springer, 2002: 535-548.

[10] ZONG B, SONG Q, MIN M R, et al. Deep Autoencoding Gaussian mixture model for unsupervised anomaly detection[C]//International Conference on Learning Representations. Vancouver: ICLR, 2018: 1-19.

[11] YAIRI T, TAKEISHI N, ODA T, et al. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(3): 1384-1401.

[12] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]//Proceedings of the 35th International Conference on Machine Learning. New York: PMLR, 2018: 4393-4402.

[13] SHEN L, LI Z, KWOK J T. Timeseries anomaly detection using temporal hierarchical one-class network[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 13016-13026.

[14] SHIN Y, LEE S, TARIQ S, et al. ITAD: integrative tensor-based anomaly detection system for reducing false positives of satellite systems[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2020: 2733-2740.

[15] PARK D, HOSHI Y, KEMP C C. A multimodal anomaly detector for robot-assisted feeding using an LSTM-based variational autoencoder[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1544-1551.

[16] LI Z H, ZHAO Y J, HAN J Q, et al. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2021: 3220-3230.

[17] AUDIBERT J, MICHIARDI P, GUYARD F, et al. USAD: UnSupervised anomaly detection on multivariate time series[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2020: 3395-3404.

[18] HU W J, YANG Y, CHENG Z Q, et al. Time-series event prediction with evolutionary state graph[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2021: 580-588.

[19] ZHANG C X, SONG D J, CHEN Y C, et al. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data[J]. arXiv Preprint, arXiv: 1811.08055, 2018.

[20] CHEN K, FENG M B, WIRJANTO T S. Multivariate time series anomaly detection via dynamic graph forecasting[J]. arXiv Preprint, arXiv: 2302.02051, 2023.

[21] SUN H L, HUANG Y, HAN L S, et al. MTS-DVGAN: Anomaly detection in cyber-physical systems using a dual variational generative adversarial network[J]. Computers & Security, 2024, 139: 103570.

[22] DENG A L, HOOI B. Graph neural network-based anomaly detection in multivariate time series[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(5): 4027-4035.

[23] ZHAO H, WANG Y J, DUAN J Y, et al. Multivariate time-series anomaly detection via graph attention network[C]//Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2020: 841-850.

[24] HALLAC D, VARE S, BOYD S, et al. Toeplitz inverse covariance-based clustering of multivariate time series data[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization. Freiburg: IJCAI, 2018: 5254-5258.

[25] SONG D J, XIA N, CHENG W, et al. Deep r-th root of rank supervised joint binary embedding for multivariate time series retrieval[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2018: 2229-2238.

[26] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2017: 6000-6010.

[27] NEAL R M. Pattern recognition and machine learning[J]. Technometrics, 2007, 49(3): 366.

[28] MATHUR A P, TIPPENHAUER N O. SWaT: a water treatment testbed

for research and training on ICS security[C]//Proceedings of the 2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater). Piscataway: IEEE Press, 2016: 31-36.

[29] ABDULAAL A, LIU Z H, LANCEWICKI T. Practical approach to asynchronous multivariate time series anomaly detection and localization[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2021: 2485-2494.

[30] XU J, WU H, WANG J, et al. Anomaly transformer: time series anomaly detection with association discrepancy[C]//Proceedings of the Tenth International Conference on Learning Representations. Vancouver: ICLR, 2022: 1-12.

[31] YU X Y, ZHANG K J, LIU Y Q, et al. Adversarial transformer-based anomaly detection for multivariate time series[J]. IEEE Transactions on Industrial Informatics, 2025, 21(3): 2471-2480.

[32] CHEN L L, GAO X, LIU J, et al. A multivariate time series anomaly detection method with multi-grain dynamic receptive field[J]. Knowledge-Based Systems, 2025, 309: 112768.

[33] LI D, CHEN D C, JIN B H, et al. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks[C]// Artificial Neural Networks and Machine Learning- ICANN 2019: Text and Time Series. Berlin: Springer, 2019: 703-716.

[34] ZHOU B, LIU S H, HOOI B, et al. BeatGAN: anomalous rhythm detection using adversarially generated time series[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. Freiburg: IJCAI, 2019: 4433-4439.

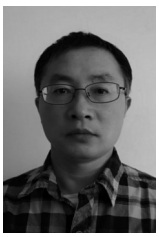
[35] ANDERSON O D. Time-Series. 2nd edn. by Maurice Kendall[J]. Journal of the Royal Statistical SocietySeries D (The Statistician), 1976, 25 (4): 308-310.

[36] EL HAJLA S, MAHFOUD E, MALEH Y, et al. Attack and anomaly detection in IoT Networks using machine learning approaches[C]//Proceedings of the 2023 10th International Conference on Wireless Networks and Mobile Communications (WINCOM). Piscataway: IEEE Press, 2023: 1-7.

[作者简介]

孙海丽（1991– ），女，湖北武汉人，华中科技大学博士生，主要研究方向为工业物联网安全、人工智能安全、入侵检测与防御、隐私保护、恶意行为识别等。

黄炎（1988– ），男，湖北武汉人，博士，华中科技大学在站博士后，主要研究方向为知识表示与推理、人工智能、工业控制系统安全、目标检测等。



韩兰胜（1972– ），男，湖北武汉人，博士，华中科技大学教授、博士生导师，主要研究方向为网络安全、大数据安全、软件安全、恶意代码检测、移动终端安全等。

周纯杰（1965– ），男，湖北黄冈人，博士，华中科技大学教授、博士生导师，主要研究方向为工业控制系统安全、网络控制系统的理论与应用、工业人工智能技术及应用等。