国防科技大学学报
Journal of National University of Defense Technology
ISSN 1001-2486,CN 43-1067/T

《国防科技大学学报》网络首发论文

题目：时空归一化流的鲁棒多元时间序列异常检测方法
作者：戴超凡，唐琪登，袁文博，马武彬，周浩浩，吴亚辉
收稿日期：2025-04-12
网络首发日期：2025-11-25
引用格式：戴超凡，唐琪登，袁文博，马武彬，周浩浩，吴亚辉．时空归一化流的鲁棒多元时间序列异常检测方法[J/OL]．国防科技大学学报.
https://link.cnki.net/urlid/43.1067.t.20251125.1139.002

网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

网络首发时间：2025-11-25 15:02:29　网络首发地址：https://link.cnki.net/urlid/43.1067.t.20251125.1139.002

国防科技大学学报
JOURNAL OF NATIONAL UNIVERSITY OF DEFENSE TECHNOLOGY

DOI:10.11887/j.issn.1001-2486.25040016

时空归一化流的鲁棒多元时间序列异常检测方法

戴超凡，唐琪登*，袁文博，马武彬，周浩浩，吴亚辉

（国防科技大学 信息系统工程全国重点实验室，湖南 长沙 410073）

摘　要：针对现有的多元时间序列异常检测模型容易受到训练集污染的影响，且难以有效建模多元时间序列中复杂的时空依赖关系的问题，提出了一种基于时空归一化流的异常检测模型。该模型利用条件归一化流，对多元时间序列中的模式进行密度估计，从而实现对训练集污染鲁棒的异常检测。同时，提出了分块长短期记忆网络来学习多元时间序列中的长时时序依赖，并提出了一种基于注意力机制的动态图学习模块，用于建模多元时间序列不同维度间的动态关系。在三个真实物理信息系统数据集上的实验结果表明，提出的模型在检测性能和鲁棒性方面均优于当前最先进的基准方法。

关键词：多元时间序列；异常检测；物理信息系统；归一化流

中图分类号：TP391.4　　**文献标志码：**A

Spatial-temporal normalizing flow for robust multivariate time series anomaly detection

Dai Chaofan, Tang Qideng, Yuan Wenbo, Ma Wubin, Zhou Haohao, Wu Yahui*

(National University of Defense Technology, National Key Laboratory of Information Systems Engineering, Changsha 410073, China)

Abstract: To address the susceptibility of existing MTS (multivariate time series) anomaly detection models to training set contamination and their limited ability to capture complex spatial-temporal correlations in MTS, a novel anomaly detection model based on spatial-temporal normalizing flow was proposed. This model employed the conditional normalizing flow to estimate the density of patterns in MTS, enabling robust anomaly detection even in the presence of contaminated training data. Additionally, a patched long short-term memory module was introduced to effectively learn long-term temporal dependencies within MTS, and a dynamic graph learning module based on attention mechanisms was devised to model the evolving correlations among different dimensions of MTS. Experimental results on three real-world cyber-physical system datasets demonstrated that the proposed model significantly outperforms state-of-the-art baselines in both detection accuracy and robustness.

Keywords: multivariate time series; anomaly detection; cyber-physical systems; normalizing flow

收稿日期：2025-04-12

基金项目：国家自然科学基金资助项目（61671233，61801208）；国家部委基金资助项目（513040106）

第一作者：戴超凡（1973—），男，湖南汨罗人，博士，研究员，E-mail：cfdai@nudt.edu.cn

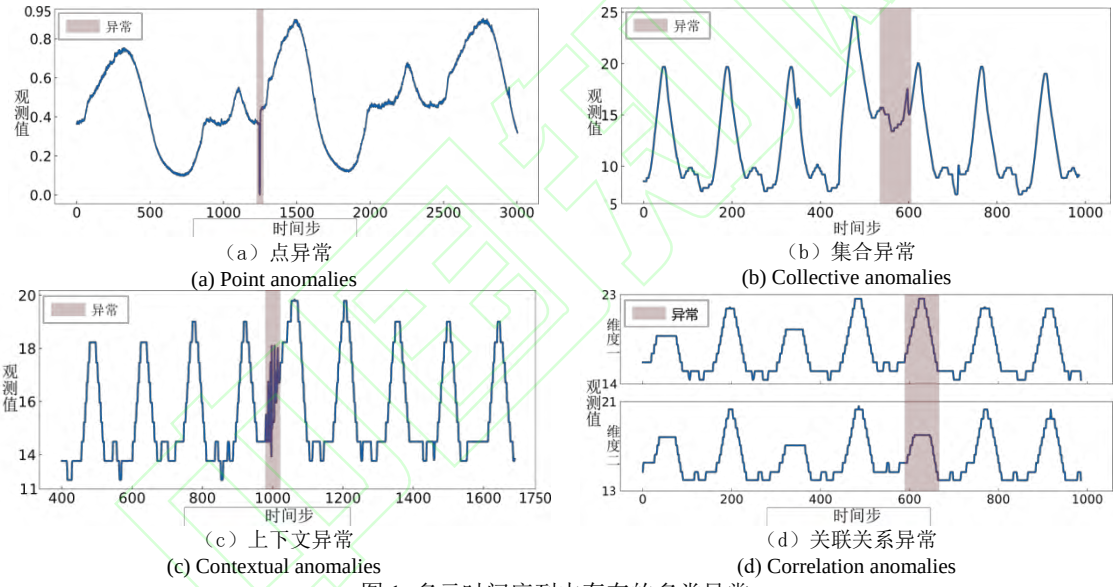
* **通信作者：**唐琪登（1999—），男，湖南衡阳人，博士研究生，E-mail：tqd18907@nudt.edu.cn；

引用格式：戴超凡，唐琪登，袁文博，等. 时空归一化流的鲁棒多元时间序列异常检测方法[J]. 国防科技大学学报，

Citation: DAI C F, TANG Q D, YUAN W B, et al. Spatial-temporal normalizing flow for robust multivariate time series anomaly detection[J]. Journal of National University of Defense Technology

随着物联网技术的快速发展，物理信息系统的复杂性和规模持续扩大，导致系统故障频率显著上升。为确保系统运行的可靠性和稳定性，运维人员部署了大量传感器设备来监控系统状态，并采集了海量多元时间序列数据。通过对这些时间序列数据进行自动化快速异常检测，能够有效预防系统故障并降低系统维护成本。多年来，研究者们提出了很多用于多元时间序列异常检测的方法。早期研究主要采用传统的机器学习方法，如孤立森林（isolation forest）[1]、单类支持向量机（OCSVM）[2]和 K 近邻（KNN）[3]。然而，这些方法由于表达能力有限，难以有效建模多元时间序列中复杂的时空依赖关系，导致检测性能欠佳。近年来，随着深度学习在文本、图像等高维数据处理中的成功应用，研究者开始将其引入多元时间序列异常检测领域。现有的深度学习方法主要分为基于预测的方法[4-6]和基于重构的方法[7-9]两类。由于实际场景中异常种类多样且发生概率低，获取充分的异常标注数据极为困难，因此当前方法主要采用无监督学习范式。这些方法通常假设训练数据仅包含正常样本，模型通过学习正常模式来识别测试集中偏离正常模式的异常样本。尽管深度学习方法在该领域取得了显著进展，但仍存在两个关键问题制约其实际应用效果：

问题 1：构建对训练集污染鲁棒的模型。现有的无监督学习异常检测方法通常假设训练集是无污染的，即训练集中仅包含正常样本。然而，实际场景中训练集不可避免地包含未标注的异常样本。当训练数据集被异常样本污染时，模型不仅会学习正常数据的模式，还会学习异常数据的模式。已有研究表明，当训练集中存在异常样本污染时，基于重构的模型和基于预测的模型都容易过拟合训练集中的异常样本[10-12]，导致模型性能显著下降。基于密度估计的方法为解决该问题提供了可行方案。这类方法将数据建模为概率



（a）点异常 (a) Point anomalies

（b）集合异常 (b) Collective anomalies

（c）上下文异常 (c) Contextual anomalies

（d）关联关系异常 (d) Correlation anomalies

图 1 多元时间序列中存在的多类异常

Fig.1 Illustration of different types of anomalies in MTS

分布，并将位于低密度区域的样本识别为异常。由于异常样本的数量远少于正常样本，即使训练集中存在异常污染，异常仍然分布在低密度区域，从而保证了检测的有效性。将这类方法应用于多元时间序列的主要挑战在于如何将多元时间序列中的时空依赖融入密度估计，以实现对多元时间序列准确的密度估计。

问题 2：建模多元时间中复杂的时空关系。如图 1 所示，多元时间序列中可能存在多种类型的异常[13]。图 1（a）和（c）中的点异常和上下文异常可通过局部时序依赖建模实现有效检测。然而，图 1（b）中集合异常的检测更为困难，因其需要准确建模长期时序依赖。基于循环神经网络（recurrent neural network，RNN）的模型难以有效学习长期时序依赖，限制了其检测持续时间较长异常的能力[14]。Transformer 架构能提取长序列中的全局依赖关系，但由于自注意力机制的排列不变特性，会不可避免地损失关键时序信息[15]。此外，多元时间序列的维度之间存在线性或非线性关系，称为空间依赖性，这对于检测图 1（d）中的关联关系异常至关重要。现有技术未能充分捕捉这些复杂且随时间变化的关系。总之，目前仍然缺乏一种能够学习多元时间序列中复杂时空依赖关系并能检测上述所有类型异常的模型。

为了解决上述两个问题，本文提出了一种基于时空归一化流（spatial-temporal normalizing flow, STNF）的无监督多元时间序列异常检测模

型。归一化流（normalizing flow, NF）是一类广泛应用于计算机视觉领域的生成模型，能够准确估计样本的概率密度[16]。本文将归一化流拓展到多元时间序列，基于时空依赖关系估算多元时间序列中样本的条件概率密度，实现鲁棒的多元时间序列异常检测。在时空依赖建模阶段，提出了一种基于分块长短期记忆网络（long short-term memory，LSTM）的模块，以学习多元时间序列中的长期依赖关系。其次，提出了一种基于注意力机制的图结构学习模块，建模多元时间序列中维度间的动态关系。在异常决策阶段，采用条件归一化流，结合获得的时空依赖，实现对时间序列中模式的细粒度密度估计。最后，根据密度估计的结果，将位于低密度区域的样本判定为异常。

1 问题描述

给定一组长度为 N 的多元时间序列：$\boldsymbol{X}=[x_1,x_2,...,x_N]$，其中 $x_t\in\mathbb{R}^M\left(1\le t\le N\right)$ 是维度为 $M\left(M>1\right)$ 的向量，代表 t 时刻的观测值。考虑到时间序列的上下文属性，采用滑动窗口方法将原始时间序列 $\mathcal{X}$ 切分为一系列长度为 T 的窗口：$W_t=\left\{x_{t-T+1},x_{t-T+2},...,x_t\right\}\left(T\le t\le N\right)$。所有子序列的集合表示为 $\mathbf{W}=\left\{W_t\mid t=T,T+1,...,N\right\}$。对 $\mathbf{W}$ 中的每一个子序列 W_t，异常检测模型需要给出对应的异常标签 $y_t\in\left\{0,1\right\}$，其中 $y_t=1$ 代表该窗口为异常，$y_t=0$ 代表该窗口为正常。

2 归一化流

归一化流是一种无监督密度估计方法，通过一系列可逆仿射变换将原始分布映射到任意目标分布。当原始数据分布 Y 的概率密度难以求解时，一种可行的方法是将其映射为简单的目标分布（如高斯分布）进行求解。假定 $y\in\mathbb{R}^D$ 为原始分布，归一化流映射方程为 $f_\theta\left(y\right):Y\mapsto Z$，将 y 转化为目标分布 $z\in\mathbb{R}^D$ 。假设目标分布的概率密度函数为 $q_Z\left(z\right)$，根据换元公式，原始分布的概率密度可以表示为 $p_Y\left(y\right)=q_Z\left(z\right)\left|\det\left(\frac{\partial f\left(y\right)}{\partial y}\right)\right|$。在归一化流中进一步引入额外的条件 C 可以实现更准确的密度估计[16]，由此引申得出的条件归一化流为 $p_Y\left(y\mid C\right)=q_Z\left(z\right)\left|\det\left(\frac{\partial f\left(y\mid C\right)}{\partial y}\right)\right|$。

3 时空归一化流异常检测方法

3.1 总体结构

STNF 的核心思想在于准确建模多元时间序列中的时空依赖关系，进而实现对多元时间序列模式的细粒度密度估计。这种准确的估计能够有效识别低密度区域，即使在训练数据集中存在异常污染的情况下，仍能保持稳定的异常检测性能。STNF 的整体框架如图 2 所示。第一部分是时间依赖学习层，将原始的多元时间序列被划分为不重叠的时间块，并采用分块 LSTM 来学习这些时间块之间的时序依赖。第二部分是空间依赖学习层，采用自注意力模块建模维度间的动态关联，并采用图卷积聚合时间块的特征向量。该步骤的输出被视为时空条件，并输入到条件归一化流中实现细粒度的密度估计。在训练阶段，整个模型通过最大似然估计（maximum likelihood estimation, MLE）进行优化。在测试阶段，将低概率密度的样本判定为异常。

3.2 基于分块 LSTM 的时间依赖建模

分块：不同于已有的时间序列异常检测模型将单个数据点作为基本的输入单元，STNF 将时间块作为基本的输入单元。时间块输入有以下三个方面的优势：1）单个数据点缺乏足够的语义信息，只有一段数据点（时间块）才包含足够的语义信息，如上升或下降趋势；2）时间序列数据通常包含由测量仪器或环境干扰引入的噪声。采用时间块可以表征整体信息，增强对局部噪声的鲁棒性；3）分块处理显著降低了模型输入序列长度，在保持较大输入窗口的同时减少了计算负担，有助于建模长期时序依赖。STNF 将多元时间序列中的每个窗口 W_t 切分为 P 个不重叠且长度为 L 的时间块：$W_t=\left\{s_i,i=1,2,...,P\right\}$，其中第 i 个时间块表示为 $s_i\in\mathbb{R}^{L\times M}$。

时序依赖学习：考虑窗口 W_t 中的一个维度 j，$W_t^j=\left\{s_1^j,s_2^j,...,s_P^j\mid s_i^j\in\mathbb{R}^L\right\}$ 为单元输入。通过连续条件化，W_t^j 的条件概率密度表示为：

$$p\left(W_t^j\right)=p\left(s_1^j\right)p\left(s_2^j\mid s_1^j\right)\cdots p\left(s_P^j\mid s_{P-1}^j,...,s_2^j,s_1^j\right) \quad (1)$$

STNF 采用 LSTM 模型建模连续概率密度 $p\left(s_i^j\mid s_{i-1}^j,s_{i-2}^j,...s_1^j\right)$，并将 $p\left(s_i^j\mid s_{i-1}^j,s_{i-2}^j,...s_1^j\right)$ 表示为 $p\left(s_i^j\mid h_{i-1}^j\right)$，其中 h_{i-1}^j 为 LSTM 模型在 $i-1$ 时刻的隐变量，包含了 i 时刻之前的所有信息。STNF 使用特征向量 h_t^j 表征每一个时间块 s_i^j，并将所有特征向量沿时间维度和空间维度连接成特征向量矩阵 $H_t\in\mathbb{R}^{P\times M\times d}$，其中 P 代表时间块的个数，M 代表维度数，d 代表 LSTM 的隐变量维度。H_t 进一步输入到空间依赖学习层，建模多元时间序列不同维度间的依赖关系。

3.3 基于自注意力机制的空间依赖建模

密度估计中引入空间依赖：为了在密度估计过程中引入空间依赖条件，本节用有向图 **A** 表示维度间的关联关系。在有向图 **A** 中，节点表示多元时间序列的维度，边表示维度间的关系，则多维时间块 $p\left(s_i^1,s_i^2,...,s_i^M\right)$ 的联合概率密度表示为：

$$p\left(s_i\right)=p\left(s_i^1,s_i^2,...,s_i^M\right)=\prod_{j=1}^{M}p\left(s_i^j\mid pa\left(s_i^j\right)\right) \quad (2)$$

其中 $pa\left(s_i^j\right)=\left\{s_i^k:\mathbf{A}_{kj}\neq 0\right\}$ 代表存在有向边指向维度 j 的维度集合。

空间依赖学习：STNF 采用自注意力机制自动学习多元时间序列维度间的关联关系图（有向图）**A**。该方法可以准确建模多元时间序列各维度间的动态依赖关系。STNF 首先将特征向量矩阵 $H_t\in\mathbb{R}^{P\times M\times d}$ 沿时间维度展平得到 $H_t'\in\mathbb{R}^{M\times Pd}$，再将 H_t' 通过线性变换得到查询矩阵 $Q_t\in\mathbb{R}^{M\times d}$ 和键值矩阵 $K_t\in\mathbb{R}^{M\times d}$。接着，STNF 对 Q_t 和 K_t 进行矩阵乘，得到自注意力矩阵 $A_t\in\mathbb{R}^{M\times M}$：

$$A_t=\mathrm{Softmax}\left(\frac{Q_t(K_t)^T}{\sqrt{d}}\right) \quad (3)$$

上式中 Softmax 函数对自注意力矩阵的每一行进行归一化处理，使得自注意力矩阵的每一行为一个和为 1 的离散分布。自注意力矩阵 A_t 表征了不同维度间的依赖强度关系，可以直接作为不同维度间的关联关系图 **A**。考虑到维度间的关系通常是稀疏的，进一步将 **A** 中的较小值重置为 0，以减少无关维度噪声的影响。

3.4 基于归一化流的密度估计

将时空依赖引入归一化流的概率密度计算过程，可以实现对多元时间序列中的模式的细粒度密度估计。这种准确的密度估计可以有效分辨出位于低概率密度处的异常模式，进而实现鲁棒的异常检测。根据上文得到的时间依赖和空间依赖，窗口 W_t 的条件概率密度可以表示为：

$$p(W_t)=p(s_1,s_2,...,s_i)=\prod_{j=1}^{M}\prod_{i=1}^{P}p\left(s_i^j\mid \mathrm{pa}\left(s_t^j\right),s_{1:i-1}^j\right) \quad (4)$$

为了同时处理时间依赖和空间依赖，STNF 将时空条件转化为一个条件向量 $c_i^j\in\mathbb{R}^d$。

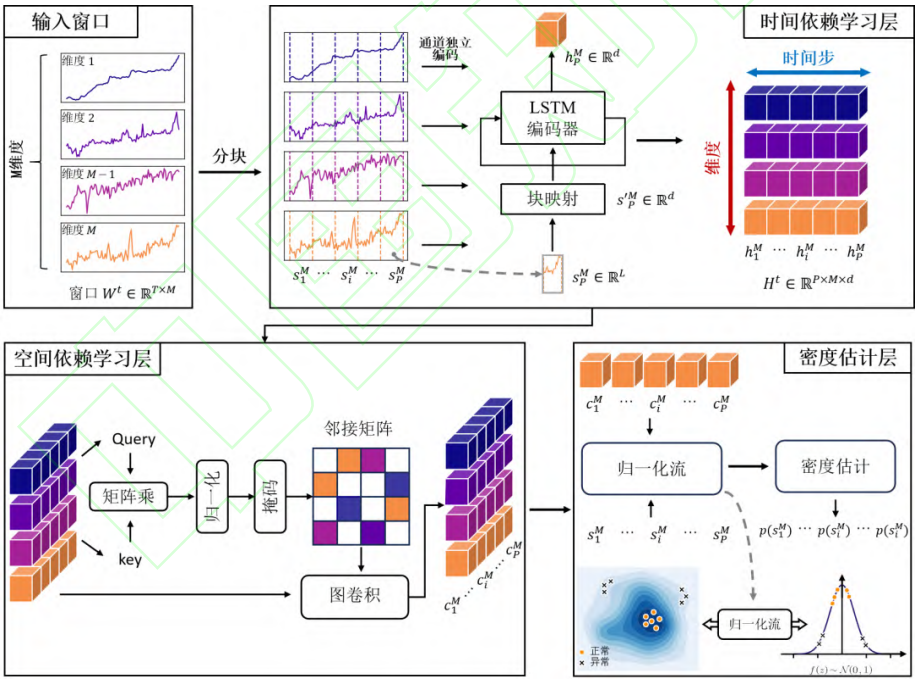


图 2 时空归一化流模型整体框架

Fig.2 The proposed architecture of the STNF

接着，窗口的条件概率密度转化为 $p\left(s_i^j\mid c_i^j\right)$，与 $p\left(s_i^j\mid \mathrm{pa}\left(s_i^j\right),s_{1:i-1}^j\right)$ 等价。为了获取条件向量 c_i^j，STNF 采用一个图卷积层汇聚学习到的空间依赖关系（spatial condition）和时间依赖关系（temporal condition），该步骤如下式所示：

$$C_t=\left\{c_i\mid c_i=\mathrm{ReLU}\left(\underbrace{A_ih_iW_1}_{\text{spatial condition}}+\underbrace{h_{i-1}W_2}_{\text{temporal condition}}\right)\cdot W_3,1\le i\le P\right\} \quad (5)$$

其中 $W_1,W_2\in\mathbb{R}^{d\times d}$ 分别为对空间依赖和时间依赖的线性变换，W_3 是用来提高条件向量的表达能力。根据参考文献[16]，STNF 将时间块 $s_i^j\in\mathbb{R}^L$ 与条件向量 $c_i^j\in\mathbb{R}^d$ 拼接作为条件归一化流的输入。由此，时间块 s_i^j 的条件概率密度可以表示为：

$$p\left(s_t^j \mid \mathrm{pa}\left(s_t^j\right), s_{t,i-1}^j\right)=q\left(f\left(\left[s_t^j ; c_t^j\right]\right)\right)\left|\det\left(\frac{\partial f\left(\left[s_t^j ; c_t^j\right]\right)}{\partial s_t^j}\right)\right| \quad (6)$$

其中 $q(z)$□ $N(z \mid 0, I)$ 为正态分布，方程 f 为掩码自回归流（masked autoregressive flow，MAF）[16]，$[\cdot;\cdot]$ 代表拼接操作。结合方程（4）和方程（6），窗口 W_t 的条件概率密度为：

$$p(W_t)=\sum_{j=1}^{M}\sum_{i=1}^{P}\left[q\left(f\left(s_t^j ; c_t^j\right)\right)\left|\det\left(\frac{\partial f\left(s_t^j ; c_t^j\right)}{\partial s_t^j}\right)\right|\right] \quad (7)$$

3.5 训练与测试

STNF 通过集成 LSTM 和自注意力机制来学习多元时间序列中的时空依赖特征。这些学习到的时空依赖特征随后被用作归一化流的输入，以实现对时间窗口的准确密度估计。为确保模型全局最优，采用端到端的联合优化策略，其目标函数遵循归一化流的经典训练范式，即最大化已观测数据的条件概率密度：

$$L_{\mathrm{MLE}}=-\frac{1}{N}\sum_{t=1}^{N}\log p\left(W_t\right) \quad (8)$$

在测试阶段，STNF 基于公式（7）计算样本的条件概率密度，并将样本的负对数密度 $-\log p\left(W_t\right)$ 作为异常分数。异常分数越高表明对应的窗口处于较低密度区域，异常的可能性越大，当样本的异常分数超过预设阈值时，模型即判定该样本为异常。需要说明的是，阈值设定可采用多种方法，如极值理论[17]和非参数扫描统计[18]。这意味着同一模型在不同阈值下会有不同的检测结果。为确保评估一致性，本研究采用文献[10-12]的方法：通过网格搜索遍历所有可能阈值，最终采用最高 F1 值所对应的阈值。

4 实验

4.1 数据集

本节在 3 个现实数据集上评估 STNF 模型的性能。这些数据集包括 WADI，SMD 和 PSM，这些数据集的统计信息如表 1 所示，对这些数据集的详细介绍如下所示：

表 1 数据集统计信息

Tab.1 Datasets statistics

数据集	维度	训练集点数	测试集点数	异常点比例/%
WADI	123	784517	172801	5.77
SMD	38	304168	304174	5.84
PSM	25	132481	87841	27.76

1）WADI 供水系统运维数据集[19]：该数据集采集自一个专为网络安全研究设计的水处理实验平台。该数据集包含 14 天系统在正常运行场景下采集的数据（训练集）和 2 天在攻击场景下采集的数据（测试集）。测试集中的异常主要包括 PLC 装置故障、网络攻击、管道泄漏和恶意化学注入。

2）SMD 云际运维数据集[20]：该数据集采集自某大型互联网公司。该数据集包含 28 个服务器连续 5 周的数据，每个服务器采集 38 个指标以描述服务器的状态（包括 CPU 负载、内存使用量、网络状态等）。测试集中的异常由系统操作员根据事件报告和领域知识进行标记。

3）PSM 服务器集群数据集[21]：该数据集采集自 eBay 的多个应用程序服务器节点。与 SMD 类似，该数据集的维度包含 CPU 利用率、内存使用率等服务器性能指标。

4.2 对比方法

本文将所提出的 STNF 方法与 7 个当前最先进的多元时间序列异常检测方法进行对比，以验证 STNF 方法的有效性：

1）DeepSVDD[22]：该方法将多元时序数据映射到低维空间，然后在该空间内找到一个最佳的超球体，使所有正常数据尽可能在该超球体内，任何远离这个球心的数据点则被视为异常。

2）ALOCC[23]：该方法采用生成对抗网络（generative adversarial networks，GAN）提升异常检测的性能。GAN 中的生成器采用自编码器模型，学习重构正常样本，判别器被当作单类分类器，区分正常数据和异常数据。

3）DROCC[24]：该方法假设正常数据点分布于一个局部线性的低维流形，显著偏离该流形结构的数据被视为异常。

4）USAD[25]：该方法假设正常时序数据可以被很好重构，而异常时序数据很难重构。该方法基于共享编码器的双自编码器架构，通过对抗训练学习正常数据的重构过程，在测试时将数据的重构误差作为异常分数。

5）MTGFlow[26]：该方法基于贝叶斯网络来学习多元时间序列维度间的相关性信息，并对多元时间序列进行密度估计。

6）DeepSAD[27]：该方法在 DeepSVDD 上添加了半监督损失函数来引入外部异常标注信息，提升异常检测效果。为了保证公平比较，本文通过在原始数据中添加高斯噪声来模拟异常样本。

7）MemAugUTransAD[28]：该方法将基于 Transformer 的 U-net 与记忆力模块结合，提高了模型的表征能力与对异常样本的鉴别能力。

4.3 评估指标

本文使用时间序列异常检测领域最常用的三个指标来评估模型的性能：1）AUC-ROC（receiver operating characteristic curve）曲线下

的面积；2）AUC-PRC（precision recall curve）曲线下的面积；3）F1 值。

4.4 实验设置

本实验基于PyTorch 1.13.1框架（CUDA 11.6版本）实现了所提出的 STNF 模型和所有对比模型。所有实验均在 NVIDIA GeForce RTX 3090 显卡上完成。STNF 模型的输入窗口大小在 WADI 数据集为 256，SMD 数据集为 512，PSM 数据集为 64。分块的长度设置为 16（PSM 数据集为 8）。LSTM、自注意力层和归一化流的隐藏层维度均设为 32。在模型训练阶段，设置样本批量训练大小为 256，学习轮次为 20，采用初始学习率为 0.001 的 ADAM 优化器对模型进行优化。为了进一步确保实验的可信性以及稳定性，所有的实验均在随机数种子 16 至 20 的范围内重复 5 次，最终报告的结果是 5 次重复实验结果的平均值和方差。

4.5 总体性能

对比实验的结果见表 2，最优结果以加粗字体标出，次优结果以下划线标出。表 2 的实验结果表明，本文所提出的 STNF 方法在 3 个多元时间序列集上都优于 7 种对比方法。多数对比方法在 SMD 数据集上表现最佳，这是因为 SMD 数据集包含模式明显的异常（例如多个指标同时出现大幅突增或者降低）。然而，部分服务器，如 1-6 号、2-7 号服务器的时间序列数据具有较强误导性，因为其中包含了大量噪声（如局部小幅波动）。这些噪声属于动态系统的固有特征，不应被判定为异常。STNF 采用的分块处理设计使模型能够聚焦全局模式而非局部波动，从而提升了对噪声的鲁棒性，降低了误报率。对比的基线方法在 PSM 数据集上表现相对较差，这是因为 PSM 训练数据集中存在较多的异常污染。然而，STNF 与 MTGFlow 在该数据集上的性能很稳定，表明基于密度估计的方法对训练数据污染具有更强的鲁棒性。大多数基线方法在 WADI 数据集上的效果最差，主要原因是该数据集有 123 个维度，且维度间存在动态变化的关联关系，而基线方法不能有效学习这些动态关系。相比之下，STNF 通过自注意力机制可以建模这些复杂的关联关系，因此在性能方面较最优对比方法提高了 3.2%的 ROC-AUC、3.0%的 PRC-AUC 和 6.2%的 F1 值。

表 2 真实数据集上的平均性能（±标准差）对比结果

Tab.2 Average performance (± standard deviation) on three datasets

%

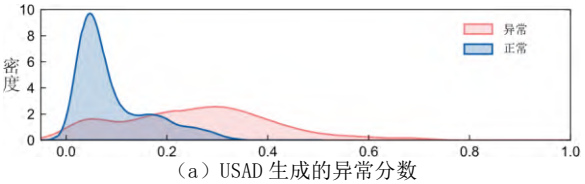
方法	WADI			SMD			PSM		
	ROC	PRC	F1	ROC	PRC	F1	ROC	PRC	F1
DeepSVDD	79.1(6.9)	30.5(8.3)	39.0(8.8)	86.0(5.4)	72.2(4.4)	66.9(3.4)	74.3(3.6)	64.8(8.1)	64.4(3.8)
ALOCC	85.3(3.2)	30.1(4.3)	30.1(5.4)	52.6(6.8)	52.8(11.2)	36.6(3.3)	71.2(8.9)	57.0(8.2)	59.6(1.7)
DROCC	85.9(9.2)	38.9(10.8)	43.9(13.1)	74.3(2.0)	51.1(1.2)	52.3(2.5)	76.3(3.6)	67.2(6.6)	67.2(4.0)
USAD	86.2(0.6)	49.2(2.3)	59.1(4.3)	88.9(7.3)	67.3(3.2)	68.3(2.7)	78.3(0.2)	67.7(1.3)	65.3(2.1)
MTGFlow	88.1(0.4)	37.3(2.3)	47.6(5.5)	88.2(2.6)	65.4(4.5)	69.4(4.4)	81.4(3.0)	71.2(4.5)	69.7(2.5)
DeepSAD	69.0(8.8)	17.8(7.3)	23.1(9.1)	88.7(1.7)	**76.1(1.4)**	68.0(2.9)	73.1.(2.7)	67.9(4.1)	61.9(2.2)
MemAug UUSTransAD	89.8(0.7)	32.1(0.8)	49.6(1.6)	78.6(1.1)	52.3(1.5)	55.4(1.9)	71.4(1.2)	55.3(1.6)	64.2(1.3)
STNF（本文方法）	**91.3(0.8)**	**52.2(1.0)**	**65.3(0.5)**	**92.2(0.9)**	72.3(2.0)	**74.1(2.2)**	**83.1(1.6)**	**73.7(3.4)**	**71.6(2.0)**

进一步，为了探究模型分辨异常的能力，本实验在 PSM 数据集上绘制了 USAD、MTGFlow 和 STNF 最大最小标准化后的异常分数的核密度估计（kernel density estimation，KDE）图。如图 3 所示，STNF 为异常窗口赋给了较大的异常值，使异常窗口和正常窗口的异常分数的重叠区域较小。这种较大的异常分数差异降低了误报率，证实了 STNF 具有更强的异常检测性能。

4.6 鲁棒性验证

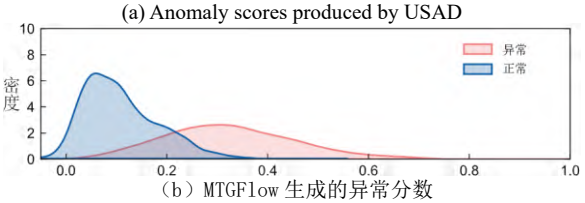
本节进一步向训练集注入异常数据来评估模型的鲁棒性。参照已有的鲁棒性验证方法[12,25]实验中向训练集随机注入高斯分布 $(\mu=0.1,\sigma=0.3)$ 值来模拟异常，注入异常的比例从 2%逐渐增加到 10%。选取表 3 中表现最好的四个模型：STNF、MTGFlow、USAD 和 DeepSVDD 进行对比分析。实验结果如图 4 所示，总体而言，随着训练集污染程度的增加，模型的性能下降程度越严重。然而，与对比模型相比，STNF 在不同污染水平下均表现出最强的鲁棒性。这是因为 STNF 采用模式密度估计的方法来检测异常，即使训练集中存在异常污染，异常的数量仍然远少于正常样本，始终分布在低密度区域，因此 STNF 能用样本的密度估计值来分辨正常样本与异常样本。即使训练集中的异常污染比例

增加到 10%，STNF 的性能下降幅度始终在 3%以内。在实际生产环境中，训练集中的异常污染比例超过 10%的情况很罕见，因此本文提出的 STNF 模型能适用于绝大多数的现实环境。



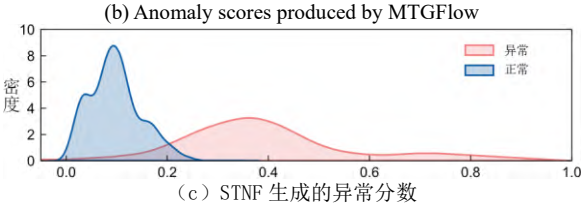
（a）USAD 生成的异常分数

(a) Anomaly scores produced by USAD



（b）MTGFlow 生成的异常分数

(b) Anomaly scores produced by MTGFlow



（c）STNF 生成的异常分数

(c) Anomaly scores produced by STNF

图 3 PSM 数据集上的异常分数分布对比图

Fig.3 Comparison of anomaly scores distribution on PSM

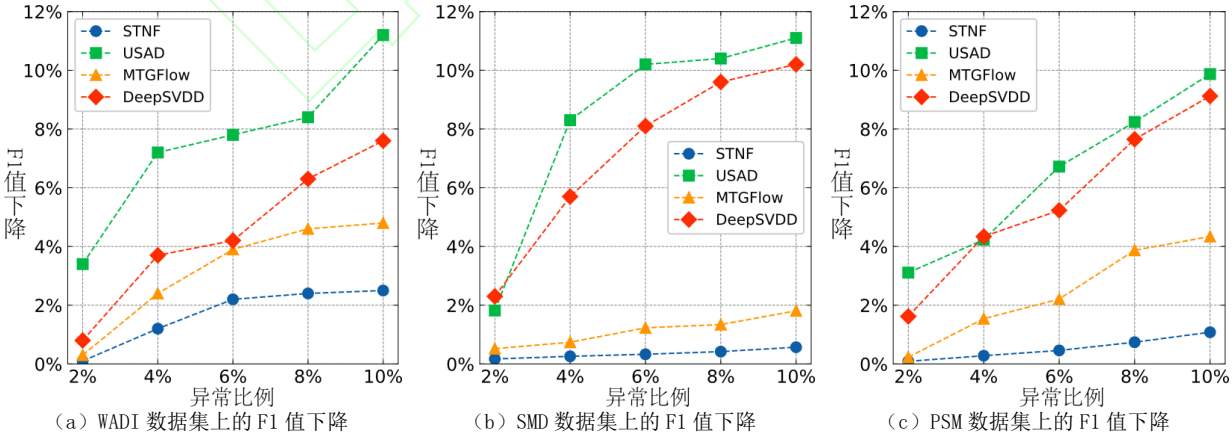
4.7 消融实验

为验证本文设计的各个模块的有效性，本节在全部数据集上进行了系统的消融实验。三种消融模型结构分别为：STNF/P、STNF/G 和 STNF/(P,G）。其中，STNF/P 将分块密度估计替换为单点密度估计；STNF/G 假设各维度相互独立，移除了用于图结构学习的自注意力模块；STNF/(P,G）同时移除了分块处理机制和自注意力模块。消融实验的实验结果如表 3 所示。STNF/(P,G）的性能最差，因其对噪声敏感且无法建模维度间的关系。STNF/P 在 PSM 数据集上的性能下降较 WADI 数据集更加明显，这是因为 PSM 数据集包含更多噪声，而分块策略有助于学习更鲁棒的全局表征。WADI 数据集单一维度的异常，会影响其关联的维度，而图结构学习模块能捕捉到正常与异常状态下的关联关系变化，因此有助于区分异常与正常模式

表 3 STNF 各消融结构的平均性能（±标准差）

Tab.3 Average performance (± standard deviation) of variants of STNF

模型种类	分块	图结构	WADI			SMD			PSM		
			ROC(%)	PRC(%)	F1(%)	ROC(%)	PRC(%)	F1(%)	ROC(%)	PRC(%)	F1(%)
STNF	✓	✓	**91.3 (0.8)**	**52.2 (1.0)**	**65.3 (0.5)**	**92.2 (0.9)**	**72.3 (2.0)**	**74.1 (2.2)**	**83.1 (1.6)**	**73.7 (3.4)**	**71.6 (2.0)**
STNF/P	✗	✓	90.3 (0.9)	49.9 (6.6)	64.4 (3.6)	89.1 (1.3)	66.1 (2.5)	70.6 (3.2)	80.7 (2.8)	72.5 (3.3)	66.7 (2.1)
STNF/G	✓	✗	90.9 (3.2)	45.1 (14.0)	58.8 (13.1)	88.8 (1.3)	66.4 (2.3)	70.8 (3.0)	81.2 (3.1)	73.0 (4.2)	70.2 (4.0)
STNF/(P, G)	✗	✗	89.8 (0.4)	37.3 (2.3)	47.6 (5.5)	88.2 (2.6)	64.3 (4.4)	67.4 (4.4)	79.2 (4.1)	68.2 (4.3)	63.8 (3.5)



（a）WADI 数据集上的 F1 值下降

(a) F1-score drop on the WADI dataset

（b）SMD 数据集上的 F1 值下降

(b) F1-score drop on the SMD dataset

（c）PSM 数据集上的 F1 值下降

(c) F1-score drop on the PSM dataset

图 4 模型在三个数据集不同污染比例下的 F1 值下降情况

Fig. 4 F1 score drop w.r.t different contamination rates on three datasets

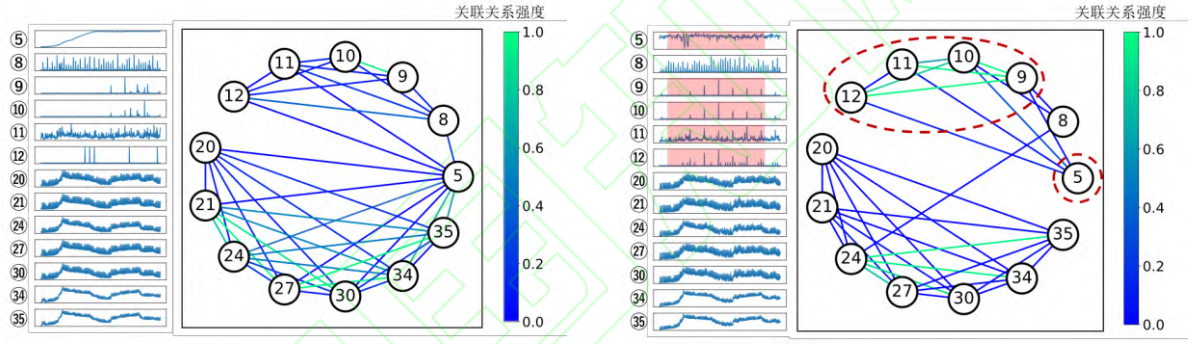
4.8 图结构学习的可视化

从隐变量学习到的图结构增强了本文提出的模型的可解释性。图 5 展示了通过注意力机制

学习到的维度间关联关系图。图 5 中每个子图的左侧表示各维度的时间序列窗口，右侧为构建的维度间关系图，其中边颜色代表维度间关系的强度，异常区域以红色高亮标注。从图 5 中可以看出，呈现相似模式的维度倾向于形成更强的连接关系。具体而言，在正常状态下，如图 5(a）所示，维度 20、21、24、27、30、34 和 35 形成一个聚类，而维度 8、9、10、11 和 12 形成另一个聚类，维度 5 与两个聚类中的维度均存在连接。在异常状态下，如图 5(b）所示，维度 5 与维度 20、21、24、27、30、34 和 35 之间的连接中断，而维度 9、10、11 和 12 之间建立了更强的连接。这种关联模式在数据集中较为罕见，位于低密度区域，因此会有更高的异常分数。

4.9 参数敏感实验

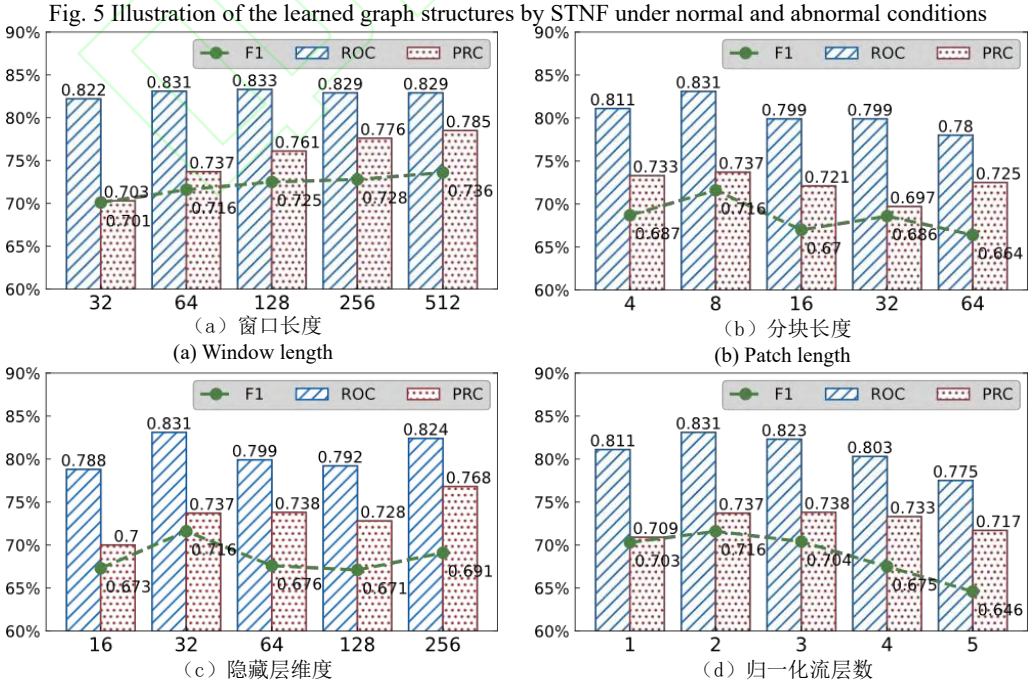
本节分析了 STNF 关键超参数对模型性能的影响。STNF 的关键参数包括窗口长度、分块长度、隐藏层维度和归一化流的层数。如图 6(a）所示，与传统基于 LSTM 的模型[7,10]随窗口长度增加，性能表现出先上升后下降的趋势不同，STNF 的性能随窗口长度增加而持续上升。这证明了分块 LSTM 模块在建模长时时序依赖的优越性。更大的窗口长度包含了更丰富的上下文信息，因此有助于检测持续时间较长的异常，如集合异常。图 6(b）展示了分块长度对模型性能的影响，当分块长度为 8 时，模型的性能最佳。这是因为过短的分块不能包含足够的语义信息，而过长的分块使时间块的模式过于复杂。图 6(c）展示了隐藏层维度对模型性能的影响，当维度设置为 32 时，模型的性能最佳，当隐藏层的维度较小时，模型难以学习多元时序数据中复杂的时空依赖关系，而维度设置过大时，由于参数量的增加，学习条件向量的过程变得更为困难。最后，本节研究了归一化流的层数对模型性能的影响，如图 6(d）所示，归一化流的层数设置为 2 时，模型的性能最好。当归一化流从层数超过 2 时，模型的性能逐渐下降，这是因为过多的归一化流层数会增加对整个分布的过拟合风险，导致异常样本也被映射到分布的高密度区域。



（a）在正常情况下学习到的图结构
(a) The graph structure learned under normal conditions

（b）在异常情况下学习到的图结构
(b) The graph structure learned under abnormal conditions

图 5 STNF 在正常情况和异常情况下学习的图结构
Fig. 5 Illustration of the learned graph structures by STNF under normal and abnormal conditions



（a）窗口长度
(a) Window length

（b）分块长度
(b) Patch length

（c）隐藏层维度

（d）归一化流层数

(c) Hidden dimension (d) Number of normalizing flow blocks

图 6 在 PSM 数据集上的参数敏感实验

Fig. 6 Hyperparameter sensitivity analysis on the PSM dataset

5 结论

本文提出了一种新颖的无监督多变量时间序列异常检测方法 STNF。该方法通过条件归一化流估计时间序列中模式的密度，并将低密度的样本识别为异常。本文提出的分块 LSTM 模块与动态图构建模块，有效建模了时间序列中复杂的时空依赖关系，实现了对时间序列中的模式的准确密度估计，进而显著提升了检测性能。在真实物理信息系统数据集上的大量实验表明，即使在训练数据存在高度异常污染的情况下，STNF 能保持优异的检测性能。在未来的研究中，计划进一步探索半监督异常检测算法，在实际生产环境中，获取少量标注的异常数据通常是可行的，希望利用少量的外部知识，进一步提升异常检测算法的性能。

参考文献

[1] LI C G, GUO L, GAO H L, et al. Similarity-measured isolation forest: anomaly detection method for machine monitoring data[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 3512512.

[2] HE J H, CHENG Z J, GUO B. Anomaly detection in telemetry data using a jointly optimal one-class support vector machine with dictionary learning[J]. Reliability Engineering & System Safety, 2024, 242: 109717.

[3] LI J B, IZAKIAN H, PEDRYCZ W, et al. Clustering-based anomaly detection in multivariate time series data[J]. Applied Soft Computing, 2021, 100: 106919.

[4] THOCKCHOM N, SINGH M M, NANDI U. A novel ensemble learning-based model for network intrusion detection[J]. Complex & Intelligent Systems, 2023, 9: 5693-5714.

[5] DING C Y, SUN S L, ZHAO J. MST-GAT: a multimodal spatial–temporal graph attention network for time series anomaly detection[J]. Information Fusion, 2023, 89: 527-536.

[6] XIAO Y H, YAO Y P, CHEN K, et al. A time-sensitive learning-to-rank approach for cloud simulation resource prediction[J]. Complex & Intelligent Systems, 2023, 9: 5731-5744.

[7] NING Z F, MIAO H, JIANG Z L, et al. Using multi-scale convolution fusion and memory-augmented adversarial autoencoder to detect diverse anomalies in multivariate time series[J]. Tsinghua Science and Technology, 2025, 30(1): 234-246.

[8] YAO Y Y, MA J H, YE Y M. Regularizing autoencoders with wavelet transform for sequence anomaly detection[J]. Pattern Recognition, 2023, 134: 109084.

[9] CHEN Z X, JIA D Q, SUN Y S, et al. Univariate time series anomaly detection based on hierarchical attention network[J]. Tsinghua Science and Technology, 2024, 29(4): 1181-1193.

[10] ZHANG Z J, LI W Z, DING W X, et al. STAD-GAN: unsupervised anomaly detection on multivariate time series with self-training generative adversarial networks[J]. ACM Transactions on Knowledge Discovery from Data, 2023, 17(5): 1-18.

[11] PAN J W, JI W D, ZHONG B, et al. DUMA: dual mask for multivariate time series anomaly detection[J]. IEEE Sensors Journal, 2023, 23(3): 2433-2442.

[12] LI L Y, YAN J C, WEN Q S, et al. Learning robust deep state space for unsupervised anomaly detection in contaminated time-series[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(6): 6058-6072.

[13] GAO C, YANG P, CHEN Y P, et al. An edge-cloud collaboration architecture for pattern anomaly detection of time series in wireless sensor networks[J]. Complex & Intelligent Systems, 2021, 7: 2453-2468.

[14] KIM J, KANG H, KANG P. Time-series anomaly detection with stacked Transformer representations and 1D convolutional network[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105964.

[15] ZENG A L, CHEN M X, ZHANG L, et al. Are transformers effective for time series forecasting[C]// Proceedings of the AAAI conference on artificial intelligence. Vol. 37. No. 9. 2023.

[16] KOBYZEV I, PRINCE S J D, BRUBAKER M A. Normalizing flows: an introduction and review of current methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(11): 3964-3979.

[17] SIFFER A, FOUQUE P A, TERMIER A, et al. Anomaly detection in streams with extreme value theory[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017: 1067-1075.

[18] LYU J M, WANG Y Q, CHEN S J. Adaptive multivariate time-series anomaly detection[J]. Information Processing & Management, 2023, 60(4): 103383.

[19] AHMED C M, PALLETI V R, MATHUR A P. WADI: a water distribution testbed for research in the design of secure cyber physical systems [C]// Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks, 2017.

[20] 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, 2019: 2828-2837.

[21] 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, 2021: 3220-3230.

[22] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]//Proceedings of the 35th International Conference on Machine Learning, 2018.

[23] SABOKROU M, FATHY M, ZHAO G Y, et al. Deep end-to-end one-class classifier[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(2): 675-684.

[24] GOYAL S, RAGHUNATHAN A, JAIN M, et al DROCC: Deep robust one-class classification [C]//Proceedings of International Conference on Machine Learning, 2020.

[25] 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, 2020: 3395-3404.
[26] ZHOU Q H, CHEN J M, LIU H Y, et al. Detecting multivariate time series anomalies with zero known label[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(4): 4963-4971.
[27] RUFF L, VANDERMEULEN R A, GÖRNITZ N, et al. Deep semi-supervised anomaly detection[C]//Proceedings of the 8th International Conference on Learning Representations, 2020.
[28] QIN S X, LUO Y C, TAO G F. Memory-augmented U-transformer for multivariate time series anomaly detection[C]//Proceedings of the ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023.