

基于轻量化GE-GRU-VAE模型的多维时间序列异常检测

赵银祥, 冯慧芳

西北师范大学数学与统计学院, 甘肃 兰州

收稿日期: 2025年9月28日; 录用日期: 2025年10月21日; 发布日期: 2025年10月28日

摘要

在工业应用中, 时间序列的无监督异常检测至关重要, 因为它能显著减少人工干预的需求。时间序列数据通常具有非平稳性、高维、异常稀缺等特点, 使得对其进行异常检测具有挑战性。本文提出了一种基于注意力机制、图嵌入技术和VAE相结合的无监督多维时间序列异常检测模型GE-GRU-VAE。首先, 在GE-GRU-VAE编码器中采用MLP和多头注意力结构进行局部特征提取, 获得输入数据内在特征的分布参数。其次, 采用重参数化得到其低维图嵌入特征。然后, 在GE-GRU-VAE解码器中采用基于图嵌入与GRU的GE-GRU模块进行时间序列重构, 通过无监督学习得到最优模型。最后, 根据双阈值异常判定方法判断序列是否异常。在两个公共数据集SWaT和WADI上验证了所提模型的有效性。实验结果表明, GE-GRU-VAE不仅具有较低的时间与空间复杂度, 而且具有良好的异常检测精度。

关键词

多维时间序列, 异常检测, 图嵌入, GRU, 变分自编码, 轻量化

Anomaly Detection in Multidimensional Time Series Based on a Lightweight GE-GRU-VAE Model

Yinxiang Zhao, Huifang Feng

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: September 28, 2025; accepted: October 21, 2025; published: October 28, 2025

Abstract

In industrial applications, unsupervised anomaly detection for time series is of crucial importance,

文章引用: 赵银祥, 冯慧芳. 基于轻量化 GE-GRU-VAE 模型的多维时间序列异常检测[J]. 数据挖掘, 2025, 15(4): 330-338.
DOI: 10.12677/hjdm.2025.154029

as it can significantly reduce the need for manual intervention. Time series data typically exhibits characteristics such as non-stationarity, high dimensionality, and scarcity of anomalies, which pose challenges to anomaly detection. This paper proposes an unsupervised multi-dimensional time series anomaly detection model named GE-GRU-VAE, which combines the attention mechanism, graph embedding technology, and Variational Autoencoder (VAE). Firstly, in the encoder of GE-GRU-VAE, a Multi-Layer Perceptron (MLP) and a multi-head attention structure are employed for local feature extraction to obtain the distribution parameters of the intrinsic features of the input data. Secondly, reparameterization is used to derive its low-dimensional graph embedding features. Then, in the decoder of GE-GRU-VAE, a GE-GRU module based on graph embedding and Gated Recurrent Unit (GRU) is adopted for time series reconstruction, and the optimal model is obtained through unsupervised learning. Finally, a dual-threshold anomaly determination method is used to judge whether the series is abnormal. This paper verifies the effectiveness of the proposed model on two public datasets, SWaT and WADI. Experimental results show that GE-GRU-VAE not only has low time and space complexity but also achieves excellent anomaly detection accuracy.

Keywords

Multivariate Time Series, Anomaly Detection, Graph Embedding, Gated Recurrent Unit, Variational Autoencoder, Lightweight

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

自 20 世纪 60 年代以来,异常点检测(也称离群点或新奇点检测)一直是众多应用领域的活跃研究领域[1]。随着计算过程的发展,大数据的收集及其在人工智能中的应用得到了更好的实现,从而促进了包括异常点检测在内的时间序列分析。时间序列异常检测(TSAD)在城市管理、入侵检测、医疗风险和自然灾害等多个领域的应用日益广泛[2][3]。近年来,深度学习在学习高维数据、时序数据、空间数据和图数据等复杂数据的表达式表示方面展现出巨大的能力,突破了不同学习任务的界限。目前,已推出了大量深度异常检测方法,在解决各种实际应用中具有挑战性的检测问题时,表现出明显优于传统异常检测的性能。

基于深度学习的异常检测主要包括两种方法,即基于预测的方法和基于重建的方法。基于预测的方法使用深度学习模型,将其预测值与实际值进行比较,根据其偏差确定是否为异常值。Ding 等[4]提出了一种基于长短期记忆(LSTM)和高斯混合模型(GMM)的实时异常检测算法 LGMAD。Shen 等[5]提出了一种名为时序分层单类(Temporal Hierarchical One-Class, THOC)网络的时序单类分类模型用于时序异常检测。该模型通过使用带跳跃连接的扩张递归神经网络,能够在多个尺度上捕捉时序动态。Chen 等[6]提出了一种名为 GTA (Graph Learning with Transformer for Anomaly Detection)的新型多变量时间序列异常检测框架,该框架通过自动学习图结构、图卷积以及基于 Transformer 架构建模时间依赖性来实现。Zheng 等[7]提出了一种基于关联感知时空图学习的时间序列异常检测模型。该模型通过基于关联学习模块捕获时间序列的成对关联,利用单跳和多跳邻居信息的图卷积网络捕获变量之间的空间依赖关系,并借助膨胀卷积进一步捕获序列在时间上的长程依赖关系。Liu 等[8]提出了一种基于增强型图注意力网络模型用于多变量时间序列异常检测,该模型引入多尺度时空卷积模块和增强图注意力模型,实验结果表明该模型能够很好识别时间序列的异常。

基于重构的方法通常是通过学习时间序列特征, 构建一个编码-解码器网络对样本数据进行重构, 如果是异常样本由于其分布差异性, 会导致重构效果差, 通过重构误差作为判断序列是否为异常的标准。Li 等[9]提出了一种基于扩张卷积变换器的生成对抗网络(DCT-GAN)用于时间序列异常检测。Wang 和 Kim 等[10][11]提出了一种基于 Transformer 的无监督多维时间序列异常检测模型。Wang 等[12]通过分解与重建方法解决了时间序列漂移问题。他们采用数据-时间混合注意力机制克服局部滑动窗口方法的局限性。Wen 等[13]提出了一种名为 LGAT 的新型无监督模型用于时间序列异常检测。该模型能够自动学习图结构, 并利用增强的异常变换器架构捕获时间依赖关系。

迄今为止, 已提出多种多维时间序列异常检测方法, 基于神经网络的异常检测方法比传统统计方法表现更好[14]。然而, 大多数现有建模方法只考虑时间序列时空依赖关系, 没有充分考虑多维时间序列空间依赖关系。针对多维时间序列的时空相关性特征, 本文提出了一种基于图嵌入 GRU 变分自编码器的多维时间序列异常检测模型(Graph Embedding-Gated Recurrent Unit Based on Variational Auto Encoders, GE-GRU-VAE)。该模型是将注意力机制、图嵌入技术、GRU 和 VAE 相结合的一个混合深度学习模型, 模型通过多个单元协同工作捕获序列的关键特征, 能够有效识别多维时间序列的异常。

2. 多维时间序列异常检测模型 GE-GRU-VAE

2.1. 模型框架

本文提出的异常检测模型框架如图 1 所示。首先, 多维时间序列通过滑动窗口的方式输入模型, 经过 GE-GRU-VAE 编码器中 MLP 和多头注意力结构进行局部特征提取, 获得输入数据内在特征的分布参数。其次, 重参数化 Gumbel-Softmax 得到其低维图嵌入。然后, 将得到的图嵌入和原始序列输入到解码器中改进的 GRU, 并通过一个全连接层进行重构, 得到与原始数据相同的形式重构结果。通过损失函数优化模型参数, 最终得到最优模型。最后, 基于最优模型, 将测试数据集输入到该模型, 得到原始序列嵌入图结构, 计算其异常得分, 并根据双阈值异常判定方法判断序列是否异常。下面是对模型各个模块的详细阐述。

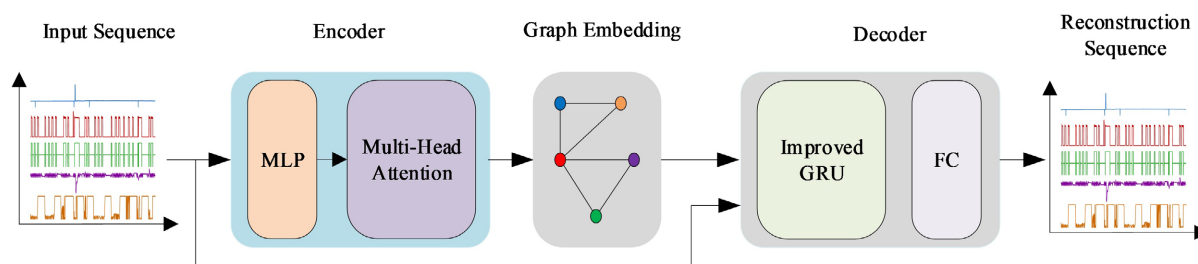


Figure 1. The framework of GE-GRU-VAE model

图 1. GE-GRU-VAE 模型框架图

2.2. 编码器

GE-GRU-VAE 模型将 MLP 和注意力机制作为 VAE 的推断网络(编码器)。编码器的输入是样本序列 S^t , 编码器的目标是学习随机变量(即潜在变量) Z^t 的近似后验分布 $q_\phi(Z^t | S^t)$, 从而得到数据的潜在分布。假设分布 $q_\phi(Z^t | S^t)$ 的参数化表示为 $\Theta^t = \{\theta_{i,j}^t, i, j = 1, \dots, N\}$ 。

在编码器部分, MLP 和多头注意力结构进行局部特征提取, 获得输入数据内在特征的分布参数。其中, 两层的 MLP 可描述为:

$$L^t = W_2 g(W_1 S^t + b_1) + b_2 \quad (1)$$

其中, S^t 为输入序列, $g(\cdot)$ 为激活函数, W_1, W_2 为权重矩阵, b_1, b_2 为偏置。

编码器中自注意力机制学习多维时间序列的依赖关系并提高潜在变量的表示能力。注意力机制是把分块操作后的数据通过不同的线性变换分别得到查询矩阵 Q^t 和键矩阵 K^t :

$$Q^t = L^t W_q, K^t = L^t W_k \quad (2)$$

其中, W_q 和 W_k 是可学习参数。通过点积计算注意力值:

$$\text{head} = Q^t (K^t)^T \quad (3)$$

通过多头注意力机制获得潜在变量的分布参数[15]:

$$\Theta^t \triangleq [\theta_{h,i,j}^t]_{H \times N \times N} = \text{Softmax}(\text{Concat}(\text{head}_1, \dots, \text{head}_H)) \quad (4)$$

其中, $\text{head}_h = Q_h^t (K_h^t)^T$, $Q_h^t = L^t W_{q,h}$, $K_h^t = L^t W_{k,h}$, $W_{q,h}$ 和 $W_{k,h}$ 是第 h 个头的可学习参数, 使用 Softmax 函数保证所有分布参数 $\theta_{h,i,j}^t$ 的概率之和等于 1。

2.3. 图嵌入

上述编码器构建了分布 q_ϕ , 生成潜在变量 $z_{i,j}^t \in R^H$, 该变量应是一个表示其所属依赖类型的 one-hot 向量。然而 $z_{i,j}^t$ 的离散化阻碍了反向传播, 并在训练过程中带来了额外的麻烦, 为此使用 Gumbel-Softmax 分类重参数化技巧:

$$z_{h,i,j}^t = \text{softmax}\left(\frac{\theta_{h,i,j}^t + g}{\tau}\right) \quad (5)$$

使得反向传播成为可能, 其中 g 是从 Gumbel(0, 1) 分布中抽取的 i.i.d. 个样本的 H 维向量, 当参数 $\tau \rightarrow 0$ 时, 连续分布收敛于预期的 one-hot 样本。我们已经得到了一组连续化的 $z_{i,j}^t$, 对其做聚合操作防止信息流失。同时, 可以构造如下矩阵:

$$a_{i,j}^t = \frac{1}{H} \sum_{h=1}^H z_{h,i,j}^t \quad (6)$$

则 $a_{i,j}^t$ 为邻接矩阵元素, $A^t \in R^{N \times N}$ 为邻接矩阵。

2.4. 解码器

解码器将潜在空间变量映射回输入空间, 得到数据输入的重构数据。本文的解码器由一个改进的 GRU (Improved GRU) 和线性层(FC)构成, 本文提出一种混合卷积 $\text{HybridConv}(\cdot, \cdot)$ 对原始 GRU 进行改进, 混合其计算公式为:

$$\text{HybridConv}(X, A) = W \left[X \parallel (AX) \parallel (A^2 X) \parallel \dots \parallel (A^{k-1} X) \right] + b \quad (7)$$

其中, W 为权重矩阵, X 和 A 分别代表特征矩阵和邻接矩阵。 $\text{HybridConv}(\cdot, \cdot)$ 混合卷积运算能够聚合不同阶邻居的信息, 以学习到节点的高层表示。改进 GRU 模块的结构如图 2 所示, 具体计算公式见(8)~(10)。

$$z_t, r_t = \text{split}\left(\sigma\left(\text{HybridConv}\left(\left[S^t, h_{t-1}\right], A\right)\right), 2\right) \quad (8)$$

$$\tilde{h}_t = \tanh\left(\text{HybridConv}\left(\left[S^t, r_t \odot h_{t-1}\right], A\right)\right) \quad (9)$$

$$h_t = \text{LayerNorm}(z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t + W_{\text{skip}} h_{t-1}) \quad (10)$$

其中, h_{t-1} 是前一时刻隐状态, S^t 为输入时间序列, A 是 t 时刻学习到的邻接矩阵, z_t, r_t 是将聚合信息经过 σ 变换后的在最后一个特征维度上平均切分形成的更新门和重置门, W_{skip} 是残差连接的参数矩阵。

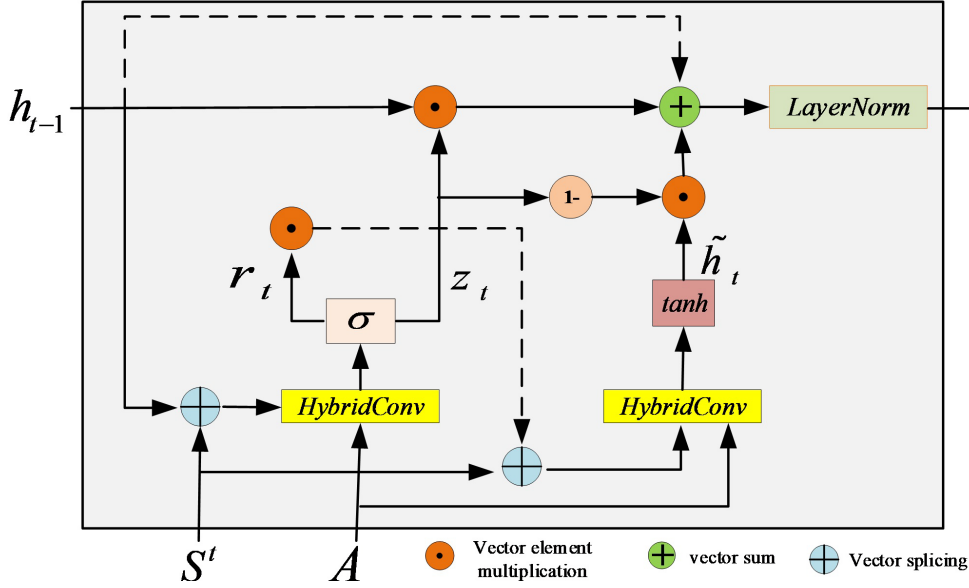


Figure 2. Improved GRU module

图 2. Improved GRU 模块

2.5. 联合优化

我们的模型是基于 VAE 的框架, 模型损失函数由两部分构成: KL 散度和重构损失, 其中 KL 散度用于度量 VAE 潜在变量分布与目标分布之间的偏离程度。重构损失度量输入与重构输出之间的差异。模型重建损失的计算公式为:

$$L_{\text{rec}} = \sum_{t=1}^T \|S^t - \hat{S}^t\|^2 \quad (11)$$

其中, S^t 和 $\hat{S}^t$ 分别为原始序列和重构序列。

KL 散度损失的计算公式为:

$$L_{\text{KL}} = \text{KL} \left[q_{\phi}(Z^t | S^t) \parallel p_{\psi}(Z^t) \right] = \sum_{k=1}^h \tilde{p}_k^t \log \left(\frac{\tilde{p}_k^t}{\tilde{q}_k} \right) \quad (12)$$

其中, $\tilde{p}_k^t = \sum_{i=1}^N \sum_{j=1}^N z_{i,j}^t(k)$ 是时间序列 i 和时间序列 j 之间的学习依赖关系属于类型 k 的概率之和。在训练过程中, 模型的损失函数包含两个优化目标, 模型整体的损失函数可以表示为:

$$L_{\text{total}} = L_{\text{rec}} + L_{\text{KL}} \quad (13)$$

3. 异常检测

多维时间序列之间存在一定耦合关系, 这种关系可用于图结构刻画[15][16]。当时间序列没有异常时, 其间的耦合关系保持稳定状态; 当序列异常时, 系统中节点间的耦合关系发生变化, 图结构特征也发生明显变化。故本文通过多维时间序列嵌入图的连接强度的变化判断时间序列是否发生异常。

通过训练集训练得到 GE-GRU-VAE 模型, 再将验证集输入该模型得到输入序列的嵌入图, 本文采用图的度的变化来衡量嵌入图连接强度的变化。因此, 根据图的度计算异常得分, 若异常得分大于阈值, 则认定该序列为异常。异常检测的步骤为:

1) 将测试数据集输入到已经训练好的模型 GE-GRU-VAE 得到嵌入图邻接矩阵 A^t 。

2) 计算图中节点 n 在 t 时刻的平均入度连接强度:

$$node_flow(t, n) = \frac{1}{N} \sum_{i=1}^N A^t(i, n) \quad (14)$$

3) 计算节点 n 在 t 时刻的异常得分:

$$score(t, n) = node_flow(t, n) - smoothed(t, n) \quad (15)$$

其中, $smoothed(t, n)$ 为移动平均平滑, 其计算公式为:

$$smoothed(t, n) = \frac{1}{w} \sum_{k=0}^{w-1} node_flow(t-k, n) \quad (16)$$

4) 计算 t 时刻所有节点的异常得分:

$$anomaly_score(t) = \frac{1}{N} \sum_{n=1}^N score(t, n) \quad (17)$$

5) 采用双阈值异常判定方法判断 t 时刻多维序列 x_t 是否异常, 如果 $label = 1$ 则序列为异常序列, $label$ 的计算公式为:

$$label = \begin{cases} 1 & \text{if } anomaly_scores(t) > \tau_1 \text{ or } anomaly_scores(t) < \tau_2 \\ 0 & \text{otherwise} \end{cases} \quad (18)$$

其中, τ_1 和 τ_2 是根据设定的百分位数得到阈值。设百分位数分别为 p 和 q , 则

$$[\tau_1, \tau_2] = prec(anomaly_scores, [p, q]) \quad (19)$$

其中, $prec(\cdot, \cdot)$ 是计算百分位数的函数。

4. 实验结果及分析

4.1. 数据集

本文采用两个真实数据集验证模型有效性, 数据集 SWaT (Secure Water Treatment) [17] 和 WADI (Water Distribution Testbed) [18] 由新加坡科技设计大学的 iTrust 机构采集并开源。数据集的基本特征如表 1 所示。两个数据集的训练集只包含正常数据, 测试数据集则包含正常和异常两种状态的数据。本文将训练集分为训练数据(80%)和验证数据(20%), 其中训练数据用于模型训练, 验证数据用于模型选择, 测试数据集用于检验模型有效性。

Table 1. Basic characteristics of the dataset

表 1. 数据集的基本特征

数据集	SWaT	WADI
训练集	496,800	1,048,571
测试集	449,919	172,801
传感器数量	51	122

续表

攻击次数	41 (36)	15
异常持续时间	100~34,208	87~1740
异常率(%)	11.97	5.99

4.2. 基线模型及评估指标

本文的基线模型为 GDN [19]、LSTM-VAE [20]、MAD-GAN [21]、Omni Anomaly [22]、Topo GDN [8] 和 GRELEN [15]。采用的性能评估指标包括精确率(Precision)、召回率(Recall)和 F1 值(F1)。

4.3. 实验设置

在实验中,我们设定了 $c_1 = 64$, $H = 4$, GRU 的所有隐藏层维度均为 64, 且其层数为 3 层。批量大小设定为 32。该方法使用 PyTorch 框架实现, 采用 Adam 优化器, 学习率为 1×10^{-3} 。所有样本均通过滑动窗口 $w = 30$ 生成。训练过程共进行了 300 个周期。 $p = 99\%$, $q = 1\%$ 。本文使用 Python 3.8 作为编程基础, PyTorch 1.11.0、cuda 11.3 实现和测试。使用 16 vCPU Intel(R) Xeon(R) Gold 6430、RTX 4090、24 GB 内存的计算机运行程序。

4.4. 实验结果

本实验在 GRELEN 基础上对模型进行了改进,表 1 为改进前后两个模型复杂度的对比。评价复杂度性能参数包括参数量 Params、浮点运算次数/秒 FLOPs、模型运行时间 time 和每个 epoch 运行时间 Avgtime/epoch。由表 2 可知,在 SWaT 和 WADI 两个数据集上,本文提出的模型 GE-GRU-VAE 的复杂度的各个指标值均明显低于 GRELEN 模型的值,尤其是刻画时间复杂度的指标 time 和 Avgtime/epoch 下降最明显。该结果表明模型 GE-GRU-VAE 在时间和空间复杂度上均优于模型 GRELEN。

Table 2. Comparison of model complexity

表 2. 模型复杂度比较

数据集	SWaT				WADI			
参数	Params	FLOPs	Time	Avgtime/Epoch	Params	FLOPs	Time	Avgtime/Epoch
GRELEN	0.47 M	21.70 G	2.65 h	31.76 s	0.47 M	51.91 G	7.78 h	93.40 s
GE-GRU-VAE	0.44 M	18.78 G	0.85 h	10.21 s	0.44 M	44.93 G	1.72 h	20.60 s

Table 3. Model performance comparison on different datasets

表 3. 在不同数据集上模型的性能比较

数据集	SWaT			WADI		
性能评价指标	Pre	Rec	F1	Pre	Rec	F1
GDN	85.46	59.31	70.02	60.8	31.64	41.62
LSTM-VAE	80.78	55.32	65.67	69.19	23.55	35.14
MAD-GAN	75.62	52.87	62.23	25.33	39.52	30.88
Omnianomaly	60.57	74.50	66.82	31.58	61.02	41.62
TopoGDN	82.92	76.74	79.71	54.08	45.79	49.59

续表

GRELEN	71	99	82	81	76	78
GE-GRU-VAE	85.28	81.35	83.27	70.12	64.25	67.06

表 3 为 GE-GRU-VAE 与基准模型性能比较。由表 3 可知, 首先, 在 SWaT 和 WADI 两个数据集上, 基线模型包括 LSTM-VAE、MAD-GAN、Omni Anomaly 和 Topo GDN 的三个性能指标的值都比较小, 说明这些模型性能较差。其次, 在 SWaT 数据集上, 与其他基线模型相比较, GE-GRU-VAE 的 F1 值最高, 其精确率 Precision 和召回率 Recall 排在次优位置, 精确率和召回率最高的模型分别是 GDN 和 GRELEN。在 WADI 数据集上, GRELEN 的三个性能最大, 说明该模型在该数据集上性能最优, GE-GRU-VAE 在该数据集上的三个性能指标都是次优。虽然在 WADI 数据集上, 本文提出的 GE-GRU-VAE 是次优模型, 但是与最优模型 GRELEN 的较高时间和空间复杂度相比较, GE-GRU-VAE 仍属于性能优良的异常检测模型。最后, 结合模型在两个数据集检测精度的性能指标值和复杂度指标值, 可知本文提出的模型 GE-GRU-VAE 对多维时间序列异常检测具有一定优势。

5. 结论

在工业应用中, 时间序列的无监督异常检测至关重要, 因为它能显著减少人工干预的需求。但多变量时间序列因其特征维度和时间维度而构成复杂挑战。本文提出了名为 GE-GRU-VAE 的多维时间序列异常检测模型, 该模型结合注意力机制、图嵌入技术和 VAE 的无监督深度学习模型。在编码器中采用 MLP 和多头注意力结构进行局部特征提取, 在解码器中采用图嵌入 GRU 进行时间序列重构, 通过训练集优化训练得到最优模型, 并根据双阈值异常判定方法判断序列是否异常。在两个真实数据集上的实验表明, GE-GRU-VAE 具有较低的时间与空间复杂度, 且具有良好的异常检测精度。未来我们不仅需要在更多数据集上检验模型的有效性, 而且要改进模型, 在降低复杂度的基础上, 提高多维时间序列异常检测的精度。

基金项目

国家自然科学基金(72461030)。

参考文献

- [1] Grubbs, F.E. (1969) Procedures for Detecting Outlying Observations in Samples. *Technometrics*, **11**, 1-21. <https://doi.org/10.1080/00401706.1969.10490657>
- [2] Esling, P. and Agon, C. (2012) Time-Series Data Mining. *ACM Computing Surveys*, **45**, 1-34. <https://doi.org/10.1145/2379776.2379788>
- [3] Carreño, A., Inza, I. and Lozano, J.A. (2019) Analyzing Rare Event, Anomaly, Novelty and Outlier Detection Terms under the Supervised Classification Framework. *Artificial Intelligence Review*, **53**, 3575-3594. <https://doi.org/10.1007/s10462-019-09771-y>
- [4] Ding, N., Ma, H., Gao, H., Ma, Y. and Tan, G. (2019) Real-Time Anomaly Detection Based on Long Short-Term Memory and Gaussian Mixture Model. *Computers & Electrical Engineering*, **79**, Article ID: 106458. <https://doi.org/10.1016/j.compeleceng.2019.106458>
- [5] Shen, L., Li, Z. and Kwok, J. (2020) Timeseries Anomaly Detection Using Temporal Hierarchical One-Class Network. *Advances in Neural Information Processing Systems*, **33**, 13016-13026.
- [6] Li, G. and Jung, J.J. (2023) Deep Learning for Anomaly Detection in Multivariate Time Series: Approaches, Applications, and Challenges. *Information Fusion*, **91**, 93-102. <https://doi.org/10.1016/j.inffus.2022.10.008>
- [7] Zheng, Y., Koh, H.Y., Jin, M., Chi, L., Phan, K.T., Pan, S., et al. (2024) Correlation-Aware Spatial-Temporal Graph Learning for Multivariate Time-Series Anomaly Detection. *IEEE Transactions on Neural Networks and Learning Systems*, **35**, 11802-11816. <https://doi.org/10.1109/tnnls.2023.3325667>

- [8] Liu, Z., Huang, X., Zhang, J., Hao, Z., Sun, L. and Peng, H. (2024) Multivariate Time-Series Anomaly Detection Based on Enhancing Graph Attention Networks with Topological Analysis. *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, Boise, 21-25 October 2024, 1555-1564. <https://doi.org/10.1145/3627673.3679614>
- [9] Li, Y., Peng, X., Zhang, J., Li, Z. and Wen, M. (2023) DCT-GAN: Dilated Convolutional Transformer-Based GAN for Time Series Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, **35**, 3632-3644. <https://doi.org/10.1109/tkde.2021.3130234>
- [10] Wang, X., Pi, D., Zhang, X., Liu, H. and Guo, C. (2022) Variational Transformer-Based Anomaly Detection Approach for Multivariate Time Series. *Measurement*, **191**, Article ID: 110791. <https://doi.org/10.1016/j.measurement.2022.110791>
- [11] Kim, J., Kang, H. and Kang, P. (2023) Time-Series Anomaly Detection with Stacked Transformer Representations and 1D Convolutional Network. *Engineering Applications of Artificial Intelligence*, **120**, Article ID: 105964. <https://doi.org/10.1016/j.engappai.2023.105964>
- [12] Wang, C., Zhuang, Z., Qi, Q., et al. (2023) Drift Doesn't Matter: Dynamic Decomposition with Diffusion Reconstruction for Unstable Multivariate Time Series Anomaly Detection. *Advances in Neural Information Processing Systems*, **36**, 10758-10774.
- [13] Wen, M., Chen, Z., Xiong, Y. and Zhang, Y. (2025) LGAT: A Novel Model for Multivariate Time Series Anomaly Detection with Improved Anomaly Transformer and Learning Graph Structures. *Neurocomputing*, **617**, Article ID: 129024. <https://doi.org/10.1016/j.neucom.2024.129024>
- [14] Pang, G., Shen, C., Cao, L. and Hengel, A.V.D. (2021) Deep Learning for Anomaly Detection. *ACM Computing Surveys*, **54**, 1-38. <https://doi.org/10.1145/3439950>
- [15] Zhang, W., Zhang, C. and Tsung, F. (2022) GRELEN: Multivariate Time Series Anomaly Detection from the Perspective of Graph Relational Learning. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, Vienna, 23-29 July 2022, 2390-2397. <https://doi.org/10.24963/ijcai.2022/332>
- [16] 姜羽, 陈华, 张小刚, 王炼红, 王鼎湘. 基于启发式时空图神经网络的多变量时序异常检测[J]. *中国科学: 信息科学*, 2023, 53(9): 1784-1801.
- [17] Mathur, A.P. and Tippenhauer, N.O. (2016) SWaT: A Water Treatment Testbed for Research and Training on ICS Security. *2016 International Workshop on Cyber-Physical Systems for Smart Water Networks (CySWater)*, Vienna, 11 April 2016, 31-36. <https://doi.org/10.1109/cyswater.2016.7469060>
- [18] Ahmed, C.M., Palleti, V.R. and Mathur, A.P. (2017) WADI: A Water Distribution Testbed for Research in the Design of Secure Cyber Physical Systems. *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, Pittsburgh, 21 April 2017, 25-28. <https://doi.org/10.1145/3055366.3055375>
- [19] Deng, A. and Hooi, B. (2021) Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 4027-4035. <https://doi.org/10.1609/aaai.v35i5.16523>
- [20] Hundman, K., Constantinou, V., Laporte, C., Colwell, I. and Soderstrom, T. (2018) Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 387-395. <https://doi.org/10.1145/3219819.3219845>
- [21] Li, D., Chen, D., Jin, B., Shi, L., Goh, J. and Ng, S. (2019) MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks. In: Tetko, I., Kůrková, V., Karpov, P. and Theis, F., Eds., *Artificial Neural Networks and Machine Learning—ICANN 2019: Text and Time Series*, Springer, 703-716. https://doi.org/10.1007/978-3-030-30490-4_56
- [22] Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W. and Pei, D. (2019) Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, 4-8 August 2019, 2828-2837. <https://doi.org/10.1145/3292500.3330672>