

面向多元时间序列的联合优化异常检测模型

吴杰辉, 柳毅*

(广东工业大学计算机学院, 广东 广州 510006)

摘要: 多元时间序列异常检测方法常被用于及时发现和识别系统中的异常模式和行为, 以提高系统的安全性和稳定性。为了解决多元时间序列内部复杂依赖关系导致的异常检测精度降低的问题, 提出一种多元时间序列异常检测模型 HGAT, 它基于图注意力网络并联合使用预测与重构方法进行优化。首先使用图注意力网络捕获多元时间序列在时间和空间维度上的依赖性; 其次采用融合变分自编码器(VAE)的 Transformer 作为重构模块, 并使用时间卷积网络(TCN)作为预测模块, 联合实现对异常序列的检测, Transformer 的自注意力机制允许重构模块在整个时间序列上建模, 从而直接考虑序列中任意 2 个位置之间的关系, 以捕捉序列的全局依赖关系, TCN 通过堆叠卷积层并且增大感受野, 可以有效地提取时间序列的局部特征; 最后通过异常分数综合考虑重构模块和预测模块, 在进行时空联合表征的基础上以全局和局部角度分析序列的整体分布。在 SMAP、MSL 和 SMD 数据集上进行实验, 结果表明, HGAT 的 F1 值分别为 96.20%、97.50% 和 92.85%, 均优于基线方法。

关键词: 多元时间序列; 异常检测; 图注意力网络; 预测; 重构

源代码链接: https://github.com/jiehuiwu/HGAT_net

中图分类号: TP391

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0069162

Jointly Optimized Anomaly Detection Model for Multivariate Time Series

WU Jiehui, LIU Yi*

(School of Computer Science and Technology, Guangdong University of Technology, Guangzhou 510006, Guangdong, China)

【Abstract】 For improving system security and stability, multivariate time series anomaly detection methods are generally used to detect and identify abnormal patterns and behaviors in systems. To mitigate reduced anomaly detection accuracy caused by complex dependencies within multivariate time series, a multivariate time series anomaly detection model, HGAT, is proposed. This model is based on a graph attention network and is optimized by combining prediction and reconstruction methods. First, graph attention networks are used to capture the temporal and spatial dependencies of multivariate time series. Second, a Transformer that integrates Variational Autoencoders (VAE) is used as the reconstruction module and a Time Convolutional Network (TCN) is used as the prediction module to jointly detect abnormal sequences. The self-attention mechanism of the Transformer enables the reconstruction module to model the entire time series by directly considering the relationship between any two positions in the sequence to capture the global dependency relationship of the sequence. TCN can effectively extract local features of the time series by stacking convolutional layers and increasing the receptive field. Finally, by comprehensively considering the reconstruction and prediction modules via abnormal scores, the overall distribution of the sequence is analyzed from both global and local perspectives based on spatiotemporal joint representation. Experiments are conducted on the SMAP, MSL, and SMD datasets, and the results show that the F1 values for the HGAT are 96.20%, 97.50%, and 92.85%, respectively. These values are superior to those for the baseline method.

【Key words】 multivariate time series; anomaly detection; graph attention network; prediction; reconstruction

0 引言

随着现代工业技术的迅速发展, 时间序列异常检测被广泛应用在工业设备状态监控、航天器状态检测、网络安全检测、金融数据欺诈分析等领域^[1-3]。异常检测的目的在于以现实世界的数据集为基础,

识别出所有与其他实例明显不同的个体^[4], 其在中企业提高效率、降低生产成本以及提高社会生活的安全性方面都发挥着重大作用。由于正常类别与异常类别之间的占比极度不平衡, 异常标签获取变得困难, 因此异常检测算法大多数为无监督机器学习算法^[5]。

收稿日期: 2024-01-03 修回日期: 2024-04-28

基金项目: 广东省重点领域研发计划(2021B0101200002)。

通信作者 E-mail: *yiliu@gdut.edu.cn

现有基于深度学习的异常检测方法已经取得了很大的进展,但仍存在缺陷,如基于预测的异常检测方法通常使用循环神经网络(RNN)及其变体网络门控循环单元(GRU)和长短时记忆(LSTM)网络,这些预测模型用于长时间序列时容易出现梯度爆炸的问题,并且它们处理下一个序列的数据需要等待前一个序列的最后输出,导致异常检测效率低下。基于重构的方法使用自编码器(AE)、变分自编码器(VAE)及生成对抗网络(GAN),时间序列通常涉及数据点之间的顺序和时间间隔等依赖特征,这些重构模型重构序列时会把时间序列中的依赖特征模糊化或丢弃,导致提取的特征不能反映原始时间序列中的重要依赖信息。此外,文献[6]表明,单独使用预测或重构的模型无法获取变量之间的复杂时空依赖关系,这种内部关系的缺失会导致模型的检测性能降低。

针对以上问题,本文提出了一种新颖的多元时间序列异常检测模型 HGAT。使用 2 种不同形式的图注意力分别提取多元时序的时间和空间依赖性,将图节点间边的注意力权重表示为时空关联程度。通过对不同查询和不同键之间的相关性进行建模,可以发现不同变量的被关注程度以及捕获时间维度的动态模式。同时,结合预测与重构方法提取时间序列的局部和全局特征。预测模块为堆叠式的时间卷积网络(TCN),其卷积操作能有效地建模序列的局部依赖关系。相较于 LSTM、GRU 等通用递归架构,TCN 具有高并行性和稳定的梯度等优点^[7]。重构模块为融合 VAE 的 Transformer,Transformer 的自注意力机制使重构模块以全局角度关注输入序列中不同位置的信息,使得重构的序列更接近原始序列的真实分布。

1 相关研究

异常检测在时间序列数据挖掘中是一个关键任务,其主要目标是从时间序列数据中识别那些与正常行为明显不同的异常模式^[8]。现有的时间序列异常检测方法大体上可分为 3 类,即统计学方法、机器学习方法和深度学习方法。

统计学方法包括指数平滑和移动平均,其中代表性的方法有自回归移动平均模型(ARIMA),它要求被检测序列为平稳随机变量^[9]。统计学方法易于理解和解释,这使得它们应用于数据分布相对稳定的场景时更具操作性。当数据的真实分布与模型的假设存在较大差异时,统计学方法可能表现不佳。

机器学习方法有基于 k -means 或基于密度的

聚类算法,同时还有决策树和支持向量机(SVM)等分类方法。经典的机器学习方法有文献[10]提出的隔离森林,它通过隔离数据点的难易程度来识别异常,路径较短的数据点可能被认为是异常点。随着现代工业技术的发展,统计学方法和传统机器学习方法已经难以应用于高维复杂的时间序列数据。

现有的深度学习时序异常检测算法大致可分为 2 类,即基于重构的方法和基于预测的方法。基于重构的异常检测方法通过对输入数据进行重建,使用重建误差来判别异常。文献[11]提出的 AE 用编码层将输入数据映射到一个低维空间,捕捉数据中的关键特征,通过比较原始数据和重构数据之间的差异,可以有效地检测出潜在的异常。文献[12]的 MAD-GAN 通过生成模型生成正常时间序列,而判别模型用于检测生成样本和实际数据之间的差异,从而识别异常。文献[13]提出 LSTM-VAE,使用 LSTM 捕获序列信息的时空特性,VAE 负责学习数据的潜在表示以重构原序列,重构误差大于阈值时则为异常。重构方法对异常点非常敏感,因此,异常点可能会引起重构误差的显著增加。由于正常点和异常点可能同时出现,这就使得重构模型在不受异常阻碍的情况下学习到正常数据的分布情况更具挑战性。

基于预测的方法通过给定时间序列来预测未来时间步的值,实际观测数据与模型预测之间的差异超过阈值时被视为异常。文献[14]使用了 SRNN,这是一种 RNN 的变体,且引入了随机性建模,使得模型在处理多元时间序列数据时更具鲁棒性,有助于捕捉时间序列中的复杂动态特征。文献[15]在航天器异常检测上使用了 LSTM 进行时间步的预测。基于 RNN 及其变体的预测模型能够学习序列中的变化,从而更好地适应和捕捉数据的突变。但其存在 2 个缺点:一是它们不能很好地捕捉长距离的时间依赖关系;二是这些方法不能对序列进行并行处理,导致异常检测效率低下。

综上可得预测与重构方法可以相互弥补两者的不足,文献[16]首次提出了联合优化基于预测和基于重构的模型,其实验结果也证明了这一点,联合优化的异常检测性能优于只选用其中一种的模型。多元时间序列的不同变量之间具有空间依赖性,序列内部也存在时间依赖性,文献[17]表明,预测与重构的模型无法对这种时间和空间依赖进行建模。为此,本文在联合预测与重构的基础上,采用 2 种不同的图注意力网络对序列的时间和空间进行特征提取,把时空关联程度表示为节点之间边的注意力权重。

2 模型设计与实现

2.1 问题定义

本文的研究专注于多元时间序列,这些时间序列包含连续相等时间间隔的观测值。设多元时间序列 $X \in R^{T \times N}$ 是一个长度为 T 、变量个数为 N 的时间序列, $X = \{x_1, x_2, \dots, x_T\}$, 其中, $x_i = \{x_i^1, x_i^2, \dots, x_i^N\}$ 。对于一个超长时间序列,选定一个长度为 W 的滑动窗口 $S = \{x_{k-w+1}, x_{k-w}, \dots, x_k\}$, 时间序列按照大小为 W 的窗口,以步幅为 1 向前滑动,通过模型输出预测向量 $y = \{y_1, y_2, \dots, y_T\}$, 其中, $y_i \in \{0, 1\}$ 表示第 i 时刻是否异常, $y_i = 1$ 表示

i 时刻发生异常。

2.2 模型细节

2.2.1 HGAT 模型

HGAT 模型的工作流程如图 1 所示。在 HGAT 模型中,时空信息提取阶段使用时间图注意力、空间图注意力捕获多元时间序列的依赖信息,接着将它们的输出与原时间序列一起拼接成一个向量,以获取总的序列特征,该向量被输入到重构与预测模块得到重构序列和预测序列,产生的结果将用于联合计算该时间戳的异常分数。最后,将异常分数与阈值进行比较,识别出该时间戳的样本是否异常。

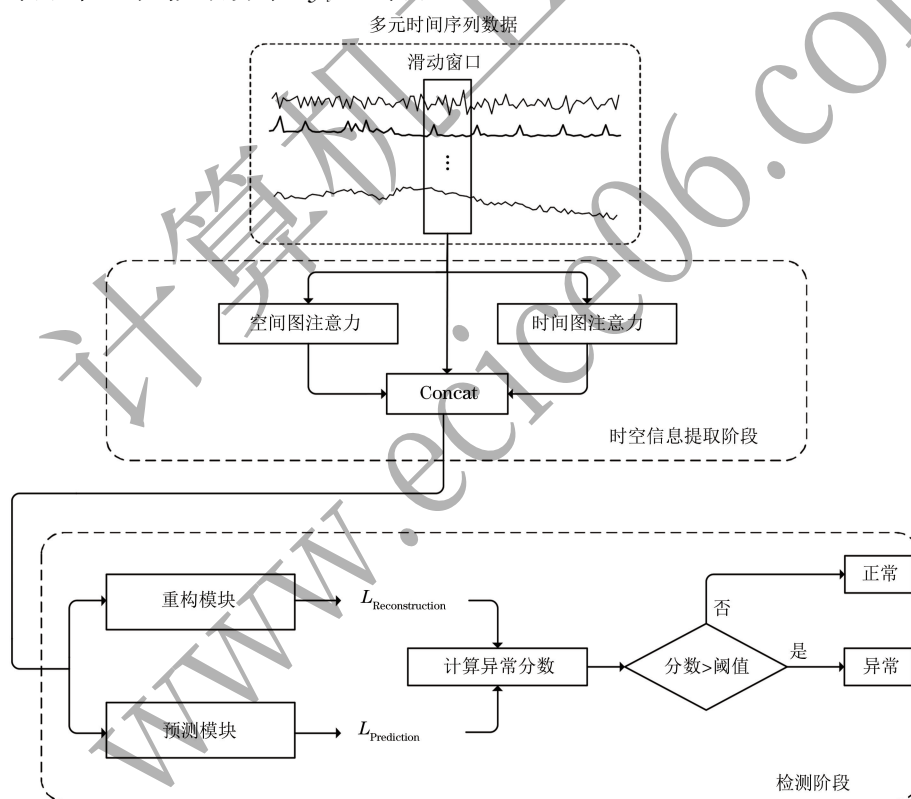


图 1 HGAT 模型工作流程

Fig.1 Workflow of HGAT model

2.2.2 图注意力网络

多元时间序列由多个单变量时间序列组成,变量内部存在时间依赖关系,这些关系可以是季节性、间隔性、周期性的。通过应用时间图注意力,模型可以捕捉到序列变量内部的模式和趋势,从而理解时间序列数据的演变和动态特征。同时变量之间也存在空间上的依赖关系,这种空间关系可以是变量之间的相似性、关联性等。多元时间序列具有复杂的时间和空间依赖性,可以使用图注意力机制来建模不同变量之间的关系^[18]。通过综合考虑序列的时间和空间依赖性,可以有效提高检测的准确性,减少误报率。

图注意力网络的工作原理如图 2 所示。设有向

图 $G = \{V, E\}$, 其中, V 是顶点集, $V = \{h_1, h_2, \dots, h_n\}$, E 是边集。图中的每个节点都会学习邻居节点分数的加权平均,节点 i 的特征输出 h'_i 可表示为:

$$h'_i = \delta \sum_{j=1}^L \alpha_{ij} h_j \quad (1)$$

式中: δ 表示 Sigmoid 激活函数; α_{ij} 是节点 i 与其邻居节点 j 的注意力分数; L 是节点 i 的所有邻居节点的个数,特征注意力分数只考虑节点 i 的所有一阶邻居。 α_{ij} 的计算如下:

$$\alpha_{ij} = \text{Softmax}(e(h_i, h_j)) = \frac{\exp(e(h_i, h_j))}{\sum_{k=1}^L \exp(e(h_i, h_k))} \quad (2)$$

$e(h_i, h_j)$ 为评分函数,其定义为:

$$e(h_i, h_j) = \text{LeakyReLU}(\mu^T [Wh_i \parallel Wh_j]) \quad (3)$$

式中:LeakyReLU 是非线性激活函数; $\mu \in R^{2m}$ 为可学习的权重向量; m 是每个节点的特征向量。

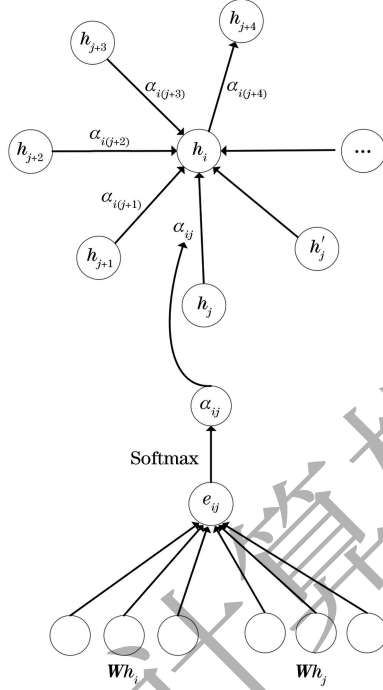


图 2 图注意力网络

Fig.2 Graph attention network

本文使用了 2 种形式的图注意力,即空间图注意力和时间图注意力。2 种不同的图注意力在多元时间序列中的节点划分如图 3 所示。

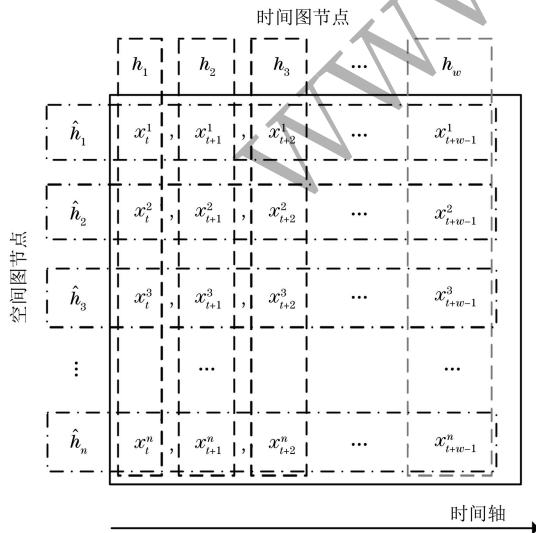


图 3 多元时间序列的时间和空间图节点

Fig.3 Temporal and spatial graph nodes of multivariate time series

1)空间图注意力。将输入的多元时间序列中特定时间戳的全部变量视为一个完整的图,图中每个变量看作一个独立的节点,节点之间的边表示 2 个

变量之间的空间依赖性。

2)时间图注意力。将一个滑动窗口内的所有时间戳视为一个完整的图,图中每个时间戳看作一个独立的节点,节点之间的边表示不同时间戳之间的时间依赖性。

2.2.3 重构模块

重构模块由 Transformer 编码器、估计网络和 Transformer 解码器组成。Transformer 编码器把特征向量压缩编码至低维潜在空间中,估计网络计算出均值与方差,得到具有序列整体特性的随机采样,Transformer 解码器的输出与随机采样融合得到重构输出。重构模块的详细工作过程如图 4 所示。

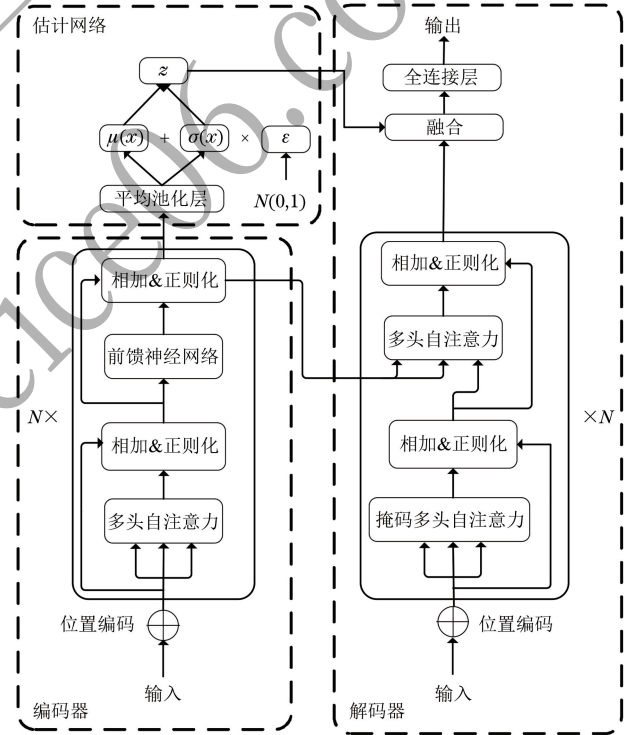


图 4 重构模块的结构

Fig.4 Refactoring module structure

首先使用正弦和余弦函数的位置编码,通过将输入向量与位置编码向量按元素相加来对数据中的顺序信息进行编码。然后定义查询向量矩阵 Q 、键向量矩阵 K 、值向量矩阵 V 这 3 个矩阵。注意力权重矩阵运算如下^[19]:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

式中: $\frac{1}{\sqrt{d_k}}$ 表示缩放因子。对于输入矩阵 Q 、 K 和 V ,先将其经过头数为 h 的前馈层,得到 Q_i 、 K_i 和 V_i ,其中, $i \in \{1, 2, \dots, h\}$,然后应用缩放点积注意力得到多头自注意力输出。多头自注意力输出表

示为:

$$\begin{aligned} \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \\ \text{Concat}(H_1, H_2, \dots, H_h) & \\ H_i &= \text{Attention}(\mathbf{Q}_i, \mathbf{K}_i, \mathbf{V}_i) \end{aligned} \quad (5)$$

通过多头自注意力,模型能够并行地学习多个不同的权重矩阵,每个矩阵关注输入序列的不同部分。多头自注意力的输出被馈送到由全连接层和 ReLU 激活函数组成的前馈子层进行计算,编码器的输出 H 如下:

$$H_{\text{encoder}} = F_2(\text{ReLU}(F_1(\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V})))) \quad (6)$$

式中: F_1 和 F_2 为 2 个不同形式的前馈函数。Transformer 编码器的输出 H_{encoder} 输入到估计网络和 Transformer 解码器的多头自注意力中, H_{encoder} 经过平均池化层得到序列的整体特征表示 h :

$$h = \frac{1}{w} \sum_{i=1}^w H_{\text{encoder}} \quad (7)$$

式中: w 为序列长度。设随机隐变量的后验分布 $q_\phi(z | x)$ 满足多元高斯分布 $q_\phi(z | x) = \mathcal{N}(\mu(x) + \sigma^2(x))$, 其数学期望 $\mu(x)$ 和方差 $\sigma^2(x)$ 可表示为:

$$\mu(x) = F_\mu(h) \quad (8)$$

$$\sigma^2(x) = \text{diag}(\exp(F_\sigma(h))) \quad (9)$$

式中: F_μ 和 F_σ 为不同形式的前馈函数; diag 为输入向量的对角矩阵。通过 Transformer 编码器输出的均值和方差计算出随机采样 z , z 的计算如下:

$$z = \sigma^2(x) \cdot \epsilon + \mu(x) \quad (10)$$

Transformer 解码器使用了掩码多头自注意力机制,其目的是屏蔽后续位置的数据信息,以确保解码器中时间序列数据点的生成仅取决于先前的数据点。设 Transformer 解码器最后一层的输出为 H_{decoder} , 将随机采样 z 与 H_{decoder} 进行融合得到隐状态 h_{out} , h_{out} 的计算如下:

$$g = \delta[H_{\text{decoder}}; z] \quad (11)$$

$$h_{\text{out}} = (1 - g) \cdot H_{\text{decoder}} + g \cdot z \quad (12)$$

式中: g 控制重构下一时刻数据时随机采样 z 所占的比重。隐状态 h_{out} 经过全连接层映射可得到下一时刻的重构值。

2.2.4 预测模块

作为序列建模的新探索,TCN 受益于 CNN,它比 RNN 具有更强的并行性和更灵活的感受野,因此本文利用 TCN 学习时间序列的表示,以预测下一个时刻的序列数据。TCN 的工作过程如图 5 所示。

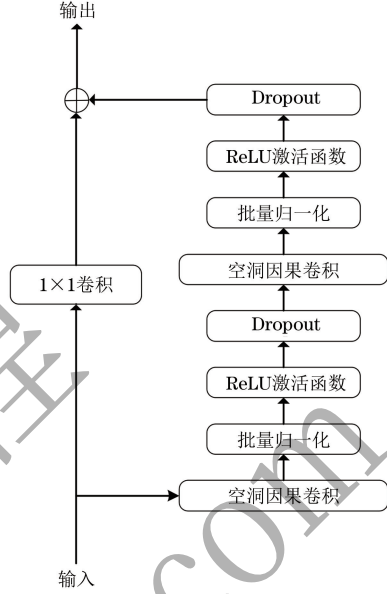


图 5 时间卷积网络的工作过程

Fig.5 Workflow of TCN

TCN 使用一维全卷积网络作为架构和零填充来保证隐藏层与输入大小相同,并且应用了因果卷积。设滤波器 $F = (f_1, f_2, \dots, f_K)$, 在 x_t 处的因果卷积为:

$$(F * X)_{(x_t)} = \sum_{k=1}^K f_k x_{t-K+k} \quad (13)$$

因果卷积仍然存在时间的建模长度受限于卷积核大小的问题,想要抓取长时间序列的依赖关系,必须线性堆叠更多的层。标准的 CNN 可以通过增加池化层来获得更大的感受野,然而,经过池化层后肯定存在信息损失的问题^[20]。为了解决这个问题,TCN 使用空洞卷积来增加感受野,空洞卷积即在标准的卷积里注入空洞,通过在卷积核移动过程中加入步长间隔来扩大因果卷积网络对历史信息感受野。空洞卷积可以减少因果卷积的深度,同时还通过在给定时间步中跳过输入值来确保输入和输出大小相同^[21]。设空洞卷积的层数为 d , 内核大小为 k , 在 x_t 处的空洞卷积为:

$$(F_d^* X)_{(x_t)} = \sum_{k=1}^K f_k x_{t-(K-k)d} \quad (14)$$

最后使用残差连接来避免梯度消失的问题,使得浅层网络更容易扩展为深层网络。残差连接的输出表示为:

$$(F^* X)_{(x_t)} = (F_d^* X)_{(x_t)} + x_t \quad (15)$$

将残差连接的结果输入到全连接层中,用于映射 TCN 编码表示,以得到下一个时刻的预测值。

2.3 联合优化

在第 i 个变量的 t 时刻,重构模块生成重构值 $\hat{x}_t^i$, 预测模块输出预测值 $\hat{y}_t^i$, 它们的损失函数均使

用均方根误差 (RMSE)。预测模块的损失函数如下:

$$L_{\text{Prediction}} = \sqrt{\sum_{i=1}^N (\hat{y}_t^i - x_t^i)^2} \quad (16)$$

重构模块的损失函数如下:

$$L_{\text{Reconstruction}} = \sqrt{\sum_{i=1}^N (\hat{x}_t^i - x_t^i)^2} \quad (17)$$

式中: x_t^i 表示 t 时刻对应的真实值; N 为变量总数。总模型优化的目标函数为预测损失与重构损失之和:

$$L = L_{\text{Reconstruction}} + L_{\text{Prediction}} \quad (18)$$

2.4 异常分数

异常分数会平衡预测与重构的权重,并且将所有变量的异常分数总和作为最终的结果。异常分数计算如下:

$$S_{\text{Anomaly scores}} = \sum_{i=1}^N (\sqrt{(\hat{x}_t^i - x_t^i)^2} + \lambda \sqrt{(\hat{y}_t^i - x_t^i)^2}) \quad (19)$$

式中: x_t^i 为第 i 个序列变量在 t 时刻的真实值; $\hat{x}_t^i$ 为重构值; $\hat{y}_t^i$ 为预测值; N 为输入数据的变量总数。设置 λ 为超参数,用于平衡预测和重构 2 个模块对模型优化的权重,通过实验得到不同 λ 值对模型性能的影响,最终得到最优的 λ 值。模型训练后得到 t 时刻的异常分数时,如果该分数大于阈值,就将该时间戳标记为异常。实验中的阈值选用峰值超过阈值 (POT)^[22] 方法来动态选择。

3 实验结果与分析

3.1 数据集

实验使用了 3 个真实世界的公开数据集,它们分别为 SMAP、MSL 和 SMD。数据集的统计情况如表 1 所示。

表 1 数据集统计信息

Table 1 Dataset statistical information

项目	SMAP	MSL	SMD
特征维度	25	55	38
训练集数目	135 183	39 312	708 405
测试集数目	427 617	73 729	708 420
异常率	0.131 3	0.102 7	0.041 6

1) 火星科学实验室数据集 MSL: NASA 对火星探测器本身的传感器和执行器的数据。

2) 土壤湿度主动被动数据集 SMAP: NASA 提供的土壤样本信息。

3) 服务器数据集 SMD: 包括服务器的 CPU 使

用率、网络使用率、内存使用率等检测量。

为了使数据集适用于模型训练,在数据预处理时把训练集和测试集中的数据标准化。以训练集为例,标准化的公式如下所示:

$$\hat{x} = \frac{x - \min(X_{\text{train}})}{\max(X_{\text{train}}) - \min(X_{\text{train}})} \quad (20)$$

式中: $\min(X_{\text{train}})$ 和 $\max(X_{\text{train}})$ 分别为训练集中的最小值和最大值。

3.2 基线方法

为验证 HGAT 的性能,选择 9 个用于时间序列异常检测的基线方法来进行比较:

1) LSTM-NDT^[15]。一种结合深度学习和非参数动态阈值技术的方法,用于航天器异常检测。使用 LSTM 对时间序列进行预测,以更好地捕捉时序模式。

2) MTAD^[16]。该模型联合重构和预测的方法应用于多元时间序列异常检测,以提高异常检测的准确性。

3) InterFusion^[23]。该模型是一种利用层次化互度量嵌入和时间嵌入的方法,以实现多元时间序列数据中异常的检测。

4) GDN^[6]。该模型基于图神经网络,将多变量时间序列数据转化为图结构,并通过学习节点表示和计算异常分数来进行异常检测。

5) AnomalyTrans^[24]。采用了 Transformer 架构,且引入关联差异的概念。关联差异涉及捕捉时间序列中不同数据点时间的关系变化,通过考虑这种关系的变化来提高异常检测的精度。

6) LSTM-VAE^[13]。该模型是为机器人辅助喂食任务而设计的基于重建模型的异常检测模型。LSTM 用于捕获序列信息的时空特性,VAE 负责学习数据的潜在表示。LSTM 和 VAE 相结合使得模型在捕捉时间序列动态特征和学习潜在表示方面具有较好的性能。

7) THOC^[25]。该模型采用时间分层的单类网络,通过比较学习到的正常模式之间的差异来进行异常检测,差异较大的时间点可能被标记为异常。

8) TimesNet^[26]。使用 TimesBlock 作为时序分析通用任务骨干,自适应地发现二维张量的多周期性,并通过 Inception 块提取二维张量的复杂时间变化。

9) GPT2(3)^[27]。利用预训练的 NLP 或 CV 模型提供一个统一的时间序列分析框架,用于处理不同的时间序列任务,其中包括时间序列异常检测。

3.3 评价指标

实验使用精确率 (Precision, P)、召回率 (Recall, R) 和 F1 值 (F_1, F_1) 来评估所有模型的检测性能:

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \quad (21)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \quad (22)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (23)$$

式中: N_{TP} (真正例) 是指模型正确地将实际为异常的样本检测为异常的情况; N_{FP} (假正例) 是指模型错误地将正常的样本检测为异常的情况; N_{FN} (假反例) 是指模型错误地将实际为异常的样本检测为正常的情况。精确率衡量了模型检测到的异常样本中有多少是真正的异常。召回率衡量了模型成功检测到真正异常样本的比例。F1 值是精确率和召回率的调和平均, 综合考虑了异常检测模型的准确性和全面性, 较高的 F1 值表示模型具有较好的异常检测能力。

3.4 实验参数

本文中的实验使用 Linux 服务器: CPU 为 AMD EPYC 7543; GPU 为 NVIDIA RTX A5000; RAM 为 128 GB。在 3 个真实世界数据集上将本文方法与 9 个时间序列异常检测基线方法进行对比,

以评估 HGAT 的异常检测性能, 对比方法包括预测方法 LSTM-NDT、重构方法 LSTM-VAE、预测与重构方法 MTAD、图神经网络方法 GDN、主流的时间序列分析方法 TimesNet、预训练方法 GPT2(3) 及其他流行方法 InterFusion、THOC 和 AnomalyTrans。本文方法使用 CUDA 12.0 的 PyTorch 2.0.1 深度学习框架来实现。模型中重要的参数设置如表 2 所示。

表 2 模型参数设置

Table 2 Model parameters configuration

参数	配置
Transformer 编码器层	5
Transformer 解码器层	5
TCN 滤波器个数	3
TCN 隐藏层	2
TCN 堆叠层	3
丢弃率	0.2
学习率	0.000 5
批处理大小	256
训练轮次	50

3.5 模型性能

表 3 中提供了 HGAT 和其他基线方法的精确率、召回率和 F1 值 (最优结果加粗标注)。从表中可以看出, HGAT 的综合检测性能优于其他基线方法。

表 3 HGAT 与其他方法的性能比较

Table 3 Performance comparison between HGAT and other methods

方法	SMAP			MSL			SMD		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
LSTM-NDT	0.864 5	0.567 3	0.685 1	0.896 1	0.639 3	0.746 2	0.673 0	0.498 9	0.573 0
LSTM-VAE	0.854 9	0.799 4	0.826 2	0.922 0	0.677 5	0.781 0	0.823 0	0.900 8	0.823 0
TimesNet	0.906 9	0.617 4	0.734 6	0.891 5	0.747 2	0.813 0	0.879 0	0.815 4	0.846 0
InterFusion	0.889 7	0.885 2	0.891 4	0.812 8	0.927 0	0.866 2	0.870 2	0.854 3	0.862 2
THOC	0.920 6	0.893 4	0.906 8	0.884 5	0.909 7	0.896 9	0.797 7	0.909 5	0.849 9
GPT2(3)	0.906 0	0.609 5	0.728 8	0.822 4	0.826 6	0.824 5	0.888 9	0.849 8	0.868 9
MTAD	0.875 2	0.899 5	0.887 2	0.868 2	0.942 5	0.903 8	0.821 0	0.921 5	0.868 3
GDN	0.748 0	0.989 1	0.851 8	0.913 5	0.861 2	0.886 6	0.877 7	0.902 2	0.889 8
AnomalyTrans	0.950 0	0.954 2	0.952 1	0.918 4	0.954 0	0.935 9	0.884 7	0.922 8	0.903 3
HGAT	0.928 8	0.997 7	0.962 0	0.969 0	0.981 0	0.975 0	0.952 1	0.906 1	0.928 5

分析表中数据可以发现: 预测方法 LSTM-NDT 采用 LSTM 来捕捉时序数据的表征, 但 LSTM 在处理长期依赖关系时可能会受到限制, 导致它在数据量越大的数据集上 F1 值表现越差; 重构方法 LSTM-VAE 对不平衡数据表现出鲁棒性并且可以快速收敛, 然而它未考虑到不同变量之间的时间相关性, 这使其在变量数最多的 MSL 数据集

上 F1 值最低; HGAT 和 GDN 都使用了图神经网络, 在 3 个数据集上的异常检测性能比大多数基线方法要好, 证明捕获多元时间序列中复杂的空间依赖性对异常检测性能有积极影响, 但 HGAT 模型在 3 个数据集上相较于 GDN 都有着更高的 F1 值, 原因是 GDN 只考虑空间相关性, 忽略了序列的时间依赖性。

在 SMAP、MSL 和 SMD 数据集上, HGAT 的 F1 值相比 MTAD 分别提高了 7.48、7.12、6.02 百分点。MTAD 虽然也联合了预测与重构方法, 但它没有以全局-局部的角度处理序列, 导致异常数据降低了其检测性能。尽管通用时间序列分析方法 TimesNet 在其他时间序列任务中表现良好, 但将其应用于异常检测时, 在 SMAP、MSL 和 SMD 数据集上的 F1 值分别仅为 73.46%、81.30% 和 84.60%, 这表明时间序列异常检测和一般时间序列表示任务存在差异, 异常检测需要模型具有较强的去噪效果。预训练方法 GPT2(3) 可以用最少的数据量训练出更优的结果, 但它在时序异常检测任务中的检测精度同样低于一般时序异常检测方法。相比于精度最高的基线方法 AnomalyTrans, HGAT 在 SMAP 上的准确率低 2.12 百分点, 在 SMD 上的召回率低 1.67 百分点。AnomalyTrans 使用关联差异进行异常检测, 其可能漏报或误报某些异常, 误报和漏报在异常报告时有时危害很大。而 HGAT 平衡了预测与重构的评分机制, 因此其在 3 个数据集上的综合检测性能 F1 值均高于 AnomalyTrans。相比于检测性能最优的基线方法 AnomalyTrans,

HGAT 模型在 SMAP、MSL 和 SMD 数据集上的 F1 值分别提高了 0.99、3.91、2.52 百分点。在这 3 个异常数据分布不均匀和变量数目差别较大的数据集中, HGAT 的平均 F1 值提高了 2.48 百分点, 体现了模型的泛化性和有效性。

3.6 可视化分析

为了展示联合预测与重构以全局-局部角度对时间序列建模的效果, 从 SMAP 和 MSL 数据集中选取了部分测试集数据进行可视化分析。图 6 展示了测试集中不同序列的波动情况, 3 个序列由上至下分别为预测序列、重构序列以及真实被检测序列。从图 6 上方可以看出, TCN 预测模块能够捕获时间序列的局部依赖性, 通过对已有序列推断后续时刻的数据, 从而有效识别突发性的数据波动, 但是其不能适应波动较大的数据, 可能会造成异常漏检。融合变分自编码器的 Transformer 重构模块更倾向于获取全局数据的依赖关系, 根据全局数据的分布情况来准确判断异常。图 6 下方清晰地展示了重构序列的稳定性, 即使在原始数据发生小波动时也不会引起很大波动, 这种特性有助于减少异常的误报率, 提高系统的稳定性和准确性。

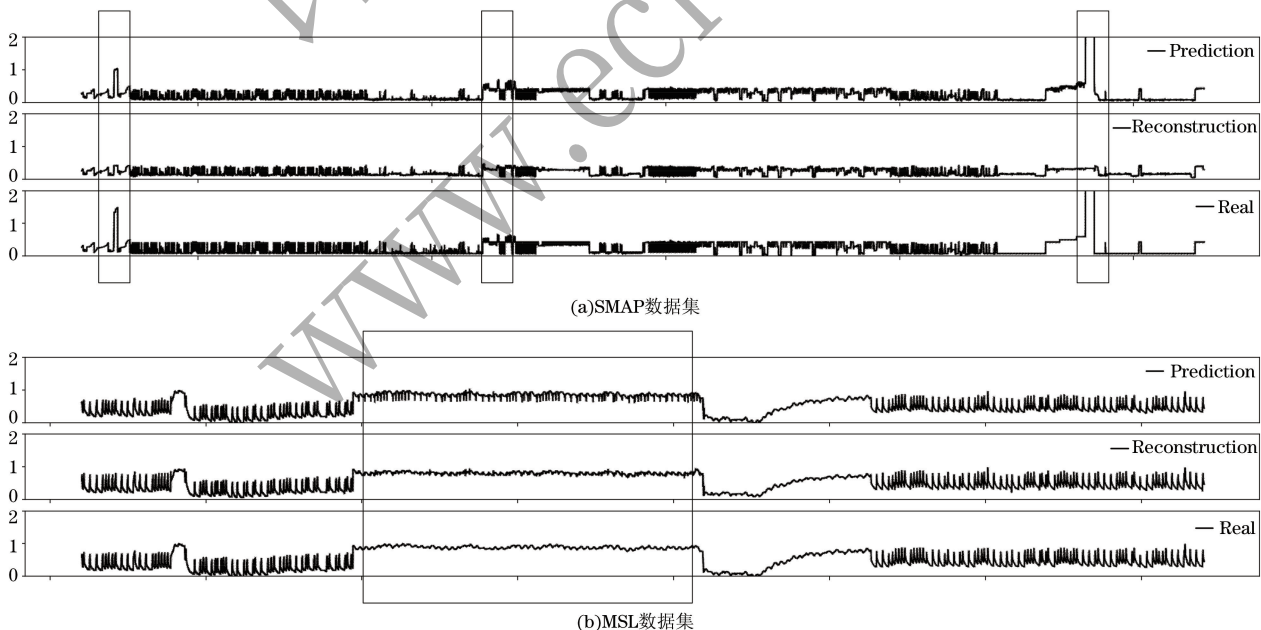


图 6 测试集中不同序列的波动情况

Fig.6 The fluctuation of different sequences in the test set

总体来说, 通过联合使用预测与重构的方法对模型进行优化, 能够以全局-局部角度处理序列的特征, 从而有效地拟合时间序列上的突变和局部小波动, 提升检测模型的鲁棒性。

3.7 参数实验

为了探索模型对超参数的敏感性, 选取 2 个最重要的超参数进行实验, 它们分别为滑动窗口长度

和联合优化权重。

选取 MSL 数据集测试滑动窗口大小对异常检测精度的影响, 对于模型中的其他参数则保持不变。设置不同的滑动窗口大小, 图 7 展示了在 MSL 数据集上滑动窗口大小对异常检测性能和时间开销的影响。从图中可以看出, 较小的滑动时间窗口会导致获取序列上下文信息的能力有限, 窗口取值为 20

和 40 时, F1 值会低于 96%。在窗口大小从 40 上升到 120 的过程中, F1 值总体呈轻微上升趋势。在窗口大小取 120 时, 模型在 MSL 上取得了最佳 F1 值 97.50%。总体来说, 重构模块中 Transformer 的多头自注意力机制可以保留序列的全局依赖, 而 TCN 可以捕获短期局部上下文信息, 使得 HGAT 对滑动时间窗口大小的敏感性较低, 在窗口大小实验中的 MSL 上 F1 值都大于 95%。虽然增加时间窗口可以使模型获取更长序列的时间依赖关系, 但所需的训练时间成本随着时间窗口的增大也会上升。

λ 是调节预测与重构联合优化权重的参数。本文在 3 个数据集上都进行大量实验来观察不同 λ 值对检测精度的影响, 图 8 展示了在 3 个数据集上不同 λ 值对应的检测性能。当 $\lambda=1.4$ 时, 模型在 3 个数据集上都取得最佳 F1 值, 在整个 λ 值敏感性实验中, SMAP 和 MSL 上的 F1 值都不低于 92%,

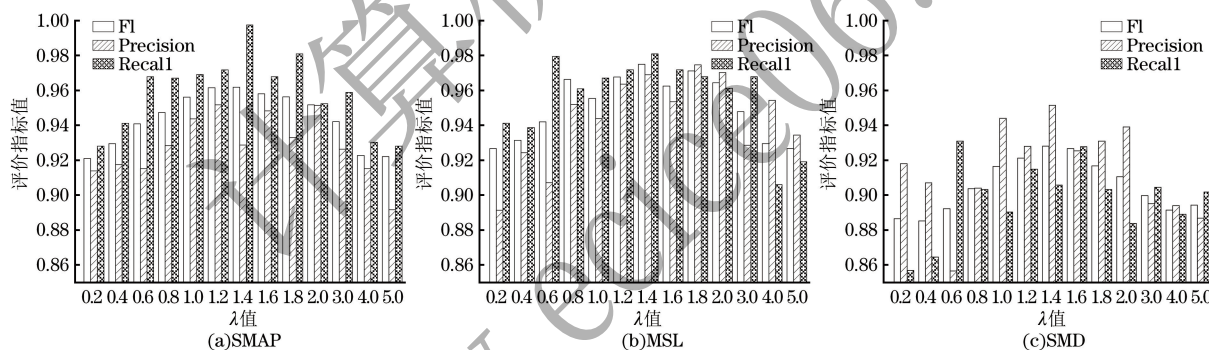


图 8 λ 值对检测性能的影响
Fig.8 The impact of λ value on detection performance

3.8 消融实验

本文的异常检测模型分为 2 个阶段: 第一阶段是序列时空信息提取; 第二阶段联合预测与重构同时检测异常。为了进一步理解模型的工作原理, 消融实验分成 2 个阶段进行, 在每个阶段移除不同的模块来观察检测性能的变化。各模块具体设置如下:

1) w/o Spatial。信息提取阶段去除空间图注意力。

2) w/o Temporal。信息提取阶段去除时间图注意力。

3) w/o Reconstruction。检测阶段去除重构模块。

4) w/o Prediction。检测阶段去除预测模块。

消融实验结果如表 4 所示, 观察表 4 可得出以下结论: 在信息提取阶段, 去除了时间图注意力模块的 HGAT 模型在 SMAP、MSL 和 SMD 数据集上 F1 值分别下降了 3.03、2.77、1.26 百分点; 去除空间图注意力模块后对应的 F1 值分别下降了 1.91、

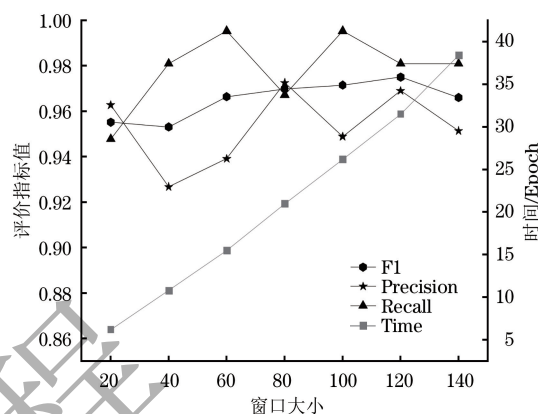


图 7 在 MSL 数据集上滑动窗口大小对检测性能及时间开销的影响

Fig.7 The impact of sliding window size on detection performance and time overhead on MSL dataset

SMD 上的 F1 值不低于 88%, 这证明模型对 λ 具有较强的鲁棒性。

4.18、3.41 百分点。

表 4 消融实验中的 F1 值

Table 4 F1 value in ablation experiment

模块	SMAP	MSL	SMD
w/o Spatial	0.942 9	0.933 2	0.894 4
w/o Temporal	0.931 7	0.947 3	0.915 9
w/o Reconstruction	0.951 2	0.953 9	0.912 8
w/o Prediction	0.929 6	0.936 4	0.890 2
HGAT	0.962 0	0.975 0	0.928 5

图 9(a) 展示了 SMAP 数据集上的空间图节点注意力分数矩阵热图。热图中较暗的区域表示 2 个空间图节点的注意力分数较高。从热图中可以看出, 所有查询节点主要关注键节点 2, 这表明其他节点在时间维度的状态更新时会与节点 2 高度相关。图 9(b) 展示了所有键节点的平均注意力分数。在异常检测的实际应用中, 空间图注意力中平均注意力分数较高的键节点应该被视为关键监测节点, 例如节点 2、8 和 16。当系统发生异常时, 这可能是由

于某些传感器变量工作状态改变引起的,这会使得各个节点间的相关性发生变化,进而改变关键监测节点,因此监测这些节点可以帮助系统及时发现可能潜在的异常情况。对于平均注意力分数较低的键节点,它们的被关注程度相对较弱,意味着这些节点在时间维度的变化趋势较为独立。在时间图注意力

中,根据时间图节点间的相关性对不相邻的时间戳进行建模,从而捕获序列的间隔性和周期性等长期依赖关系。综上所述,时间图注意力可以有效地关注到序列的时间维度信息,空间图注意力关注多变量间的空间依赖关系,时间和空间图注意力都有助于提高异常检测的性能。

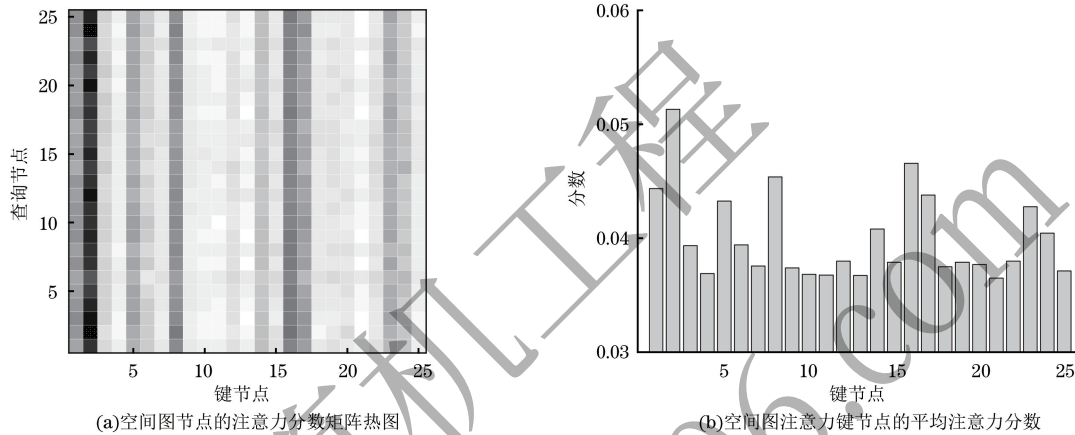


图 9 SMAP 数据集上的空间图注意力分数可视化

Fig.9 Visualization of spatial graph attention scores on SMAP dataset

在联合检测阶段,模型在去除预测或重构的情况下 F1 值都低于 HGAT 模型,原因是预测模型难以区分接近正常数据的异常数据,而重构模型不能很好地拟合突变数据,两者相结合则可以很好地解决这 2 个问题。重构模块中 Transformer 在时间序列数据中使用自注意力机制,可以帮助模型理解不同时刻数据之间的关系。预测模块中的 TCN 通过卷积操作,可以更好地捕捉时间序列的局部特征。预测与重构在时间序列中的建模能力互补,通过并联可以更好地捕捉序列中的不同模式,这也说明了联合优化在异常检测中的有效性。

4 结束语

本文提出了一种新颖的多元时序异常检测模型 HGAT,以提高时序异常检测的性能。利用 2 种形式的图注意力网络提高模型在高维度序列数据中建模复杂时空交互关系的能力。基于 TCN 的预测方法适应时间序列的突发性局部异常,使用融合变分自编码器的 Transformer 作为重构模块学习时间序列的全局分布,2 种方法的结合使得 HGAT 受突发噪声和局部波动的影响较小。在 3 个数据集上进行实验,结果表明,所提模型的 F1 值比最优的基线方法 AnomalyTrans 平均提高了 2.48 百分点,证明了模型的有效性和泛化性。在下一步工作中,将研究如何降低检测所需的计算复杂度,使模型可以更好地应用于实时的工业异常检测场景。

参考文献

- [1] 刘杰, 房俊, 雷峰津. 电能质量异常数据在线检测方法[J]. 计算机工程与应用, 2020, 56(9): 240-247.
LIU J, FANG J, LEI F J. On-line detection method for abnormal data of power quality[J]. Computer Engineering and Applications, 2020, 56(9): 240-247. (in Chinese)
- [2] 李瑞雪, 张泽旭. 数据驱动的航天器异常检测工具对未来中国空间站管理的启示[J]. 载人航天, 2021, 27(2): 244-251.
LI R X, ZHANG Z X. Enlightenment of data-driven spacecraft anomaly detection tools on future space station management in China[J]. Manned Spaceflight, 2021, 27(2): 244-251. (in Chinese)
- [3] MULINKA P, CASAS P. Stream-based machine learning for network security and anomaly detection[C]//Proceedings of the 2018 Workshop on Big Data Analytics and Machine Learning for Data Communication Networks. New York, USA: ACM Press, 2018: 1-7.
- [4] CHANDOLA V, BANERJEE A, KUMAR V. Outlier detection: a survey[J]. ACM Computing Surveys, 2007, 14: 15.
- [5] ZHANG Y X, CHEN Y Q, WANG J D, et al. Unsupervised deep anomaly detection for multi-sensor time-series signals [EB/OL]. [2023-10-05]. <https://arxiv.org/abs/2107.12626v2>.
- [6] 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.
- [7] FAN J, ZHANG K, HUANG Y P, et al. Parallel spatio-temporal attention-based TCN for multivariate time series prediction[J]. Neural Computing and Applications, 2023, 35(18): 13109-13118.
- [8] PANG G, SHEN C, CAO L, et al. Deep learning for anomaly detection: a review[J]. ACM Computing Surveys, 2021, 54(2): 1-38.
- [9] YU Q, LÜ J B, JIANG L R. An improved ARIMA-based traffic anomaly detection algorithm for wireless sensor

- networks[J]. International Journal of Distributed Sensor Networks, 2016, 12(1): 9653230.
- [10] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]//Proceedings of the 8th IEEE International Conference on Data Mining. Washington D. C., USA: IEEE Press, 2008: 413-422.
- [11] SAKURADA M, YAIRI T. Anomaly detection using autoencoders with nonlinear dimensionality reduction[C]//Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis. Washington D. C., USA: IEEE Press, 2014: 4-11.
- [12] LI D, CHEN D C, JIN B H, et al. MAD-GAN: multivariate anomaly detection for time series data with generative adversarial networks[EB/OL]. [2023-10-05]. <https://arxiv.org/abs/1901.04997>.
- [13] 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.
- [14] 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, USA: ACM Press, 2019: 2828-2837.
- [15] 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, USA: ACM Press, 2018: 387-395.
- [16] ZHAO H, WANG Y J, DUAN J Y, et al. Multivariate time-series anomaly detection via graph attention network[C]//Proceedings of the IEEE International Conference on Data Mining (ICDM). Washington D. C., USA: IEEE Press, 2020: 841-850.
- [17] 梁李芳, 关东海, 张吉, 等. 基于时空注意力机制的多元时间序列异常检测[J]. 计算机科学, 2023, 50(S2): 450-457. LIANG L F, GUAN D H, ZHANG J, et al. Anomaly detection of multivariate time series based on spatio-temporal attention mechanism[J]. Computer Science, 2023, 50(S2): 450-457. (in Chinese)
- [18] FANG R Y, WEN L J, KANG Z, et al. Structure-preserving graph representation learning[C]//Proceedings of the IEEE International Conference on Data Mining (ICDM). Washington D. C., USA: IEEE Press, 2022: 927-932.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. [2023-10-05]. <https://arxiv.org/abs/1706.03762>.
- [20] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[EB/OL]. [2023-10-05]. <https://arxiv.org/abs/1406.4729>.
- [21] LIU L, TIAN L, KANG Z, et al. Spacecraft anomaly detection with attention temporal convolution networks[J]. Neural Computing and Applications, 2023, 35(13): 9753-9761.
- [22] 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. New York, USA: ACM Press, 2017: 1067-1075.
- [23] 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, USA: ACM Press, 2021: 3220-3230.
- [24] XU J H, WU H X, WANG J M, et al. Anomaly transformer: time series anomaly detection with association discrepancy[EB/OL]. [2023-10-05]. <https://arxiv.org/abs/2110.02642v5>.
- [25] SHEN L, LI Z, KWOK J. Timeseries anomaly detection using temporal hierarchical one-class network[J]. Advances in Neural Information Processing Systems, 2020, 33: 13016-13026.
- [26] WU H X, HU T, LIU Y, et al. TimesNet: temporal 2D-variation modeling for general time series analysis[EB/OL]. [2023-10-05]. <https://arxiv.org/abs/2210.02186>.
- [27] ZHOU T, NIU P, SUN L, et al. One fits all: power general time series analysis by pretrained LM[EB/OL]. [2023-10-05]. <https://arxiv.org/abs/2302.11939>.

编辑 吴云芳