

基于时频域注意力的时间序列异常检测模型

徐敬涛, 杨燕, 江永全

引用本文

徐敬涛, 杨燕, 江永全. 基于时频域注意力的时间序列异常检测模型[J]. 计算机科学, 2026, 53(2): 161-169.

XU Jingtao, YANG Yan, JIANG Yongquan. [Time-Frequency Attention Based Model for Time Series Anomaly Detection](#) [J]. Computer Science, 2026, 53(2): 161-169.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于背景结构感知的小样本知识图谱补全](#)

Background Structure-aware Few-shot Knowledge Graph Completion

计算机科学, 2026, 53(2): 331-341. <https://doi.org/10.11896/jsjcx.250100107>

[深度融合句法和语义特征的情感三元组片段级抽取方法](#)

Method for Span-level Sentiment Triplet Extraction by Deeply Integrating Syntactic and Semantic Features

计算机科学, 2026, 53(2): 322-330. <https://doi.org/10.11896/jsjcx.250100061>

[语义引导的红外与可见光图像混合交叉特征融合方法](#)

Semantic-guided Hybrid Cross-feature Fusion Method for Infrared and Visible Light Images

计算机科学, 2026, 53(2): 253-263. <https://doi.org/10.11896/jsjcx.250100123>

[基于注意力机制的音频驱动数字人脸视频生成方法](#)

Attention-based Audio-driven Digital Face Video Generation Method

计算机科学, 2026, 53(2): 245-252. <https://doi.org/10.11896/jsjcx.241200067>

[双支特征融合的带约束的多损失视频异常检测](#)

Constrained Multi-loss Video Anomaly Detection with Dual-branch Feature Fusion

计算机科学, 2026, 53(2): 236-244. <https://doi.org/10.11896/jsjcx.250300103>

基于时频域注意力的时间序列异常检测模型

徐敬涛 杨燕 江永全

西南交通大学计算机与人工智能学院 成都 611756

可持续城市交通智能化教育部工程研究中心 成都 611756

(kk_xjt@163.com)

摘要 时序数据中复杂的时间依赖关系和有限的异常标签数据,使得时序异常检测成为一项十分具有挑战性的工作。以往大多数工作的聚焦点局限于数据的时域建模上,而忽略并缺失了对时序数据频域信息的提取与利用,造成了一定程度的性能瓶颈。以此为突破点,提出一种基于时域注意力和频域注意力的异常检测模型——TFA-TSAD。该模型首先创新性地对输入数据进行渐进式分解,进而探索数据在不同模式下的异常情况;然后利用精心设计的时频域建模模块,以注意力机制为核心,分别对时序数据的时域信息与频域信息进行高效提取来提高异常检测的性能;此外,在传统误差损失的基础上,采用均方误差加和平均的方式作为损失函数,进一步提高了模型的性能。经过大量的实验验证,该方法在多个数据集上的性能表现卓越,与其他13个基线模型相比,异常检测效果提升显著。

关键词: 时序异常检测;时域;频域;注意力机制;模式分解

中图分类号 TP391.1

Time-Frequency Attention Based Model for Time Series Anomaly Detection

XU Jingtao, YANG Yan and JIANG Yongquan

School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

Engineering Research Center of Sustainable Urban Intelligent Transportation, Ministry of Education, Chengdu 611756, China

Abstract Time series anomaly detection is a challenging task due to complex temporal dependencies and limited labeled anomaly data. Previous methods have predominantly focused on modeling in the time domain, overlooking valuable information contained in the frequency domain, resulting in a certain degree of performance bottleneck. Taking this as a breakthrough, this paper proposes a Time-Frequency Attention Based Model for Time Series Anomaly Detection—TFA-TSAD, which firstly innovatively performs progressive decomposition of the input data to explore the anomalies of the data in different modes, and then utilizes the well-designed time-frequency domain modeling module to efficiently extract the time-domain information and frequency-domain information of the origin data respectively, by taking the attention mechanism to improve the performance of anomaly detection. Finally, based on the traditional error loss, a loss function incorporating average MSE is utilized to further improve model performance. Extensive experimental results on multiple datasets demonstrate that the proposed model outperforms 13 other benchmark models with a significant performance.

Keywords Time series anomaly detection, Time domain, Frequency domain, Attention mechanism, Decomposition

1 引言

时序异常检测旨在找出时序数据中处于正常分布外的点或区段。现实生活中诸多类型的数据都可以看作时序数据,这使得时序异常检测具备广泛的实际应用需求与研究价值。例如,对工业机械的工作数据进行异常检测,可以为设备检修和维护提供便捷途径^[1];对大脑电波的数据实施异常检测,可以为癫痫的发作提供预警^[2];对股市数据进行异常检测,可为股市投资进行风险预警和规避^[3]。

异常检测技术虽具有广泛适用性,但在复杂时序数据中实现精准的异常识别仍面临关键性挑战:1)时序数据可能包含多个复杂的数据模式,包括但不限于季节性、趋势性和周期性等,如何有效地识别这些不同类型的模式,并从中捕获异常并非易事;2)异常事件在真实场景下并非经常发生,这使得获取足量且高质量的已标注训练数据相对困难;3)时序数据中隐含的时间和特征之间的依赖关系也应该被充分考虑,异常检测模型能否正确处理这些依赖关系也十分重要。

异常数据在数量上明显少于正常数据,且往往隐藏在大

到稿日期:2024-12-16 返修日期:2025-03-17

基金项目:国家自然科学基金(61976247)

This work was supported by the National Natural Science Foundation of China(61976247).

通信作者:杨燕(yyang@swjtu.edu.cn)

量正常数据中难以被区分,这种双重特性导致异常数据难以被发现。为了解决异常数据和异常标签缺失的问题,研究人员提出了大量基于无监督学习的方法。Isolation Forest (IF)^[4]通过构建随机决策树来隔离数据,其中正常数据点需要更多的分割才能被隔离开来,而异常点因为与其他数据点相对孤立而容易被隔离开来。One-Class SVM^[5]通过找到数据在高维空间中的最大边界,试图将正常数据点与原点隔离开来,异常点则位于这个边界之外。K-Means clustering^[6]将数据点划分为 k 个簇,每个数据点分配到最近的簇中心,离簇中心较远的点可以被认为是异常点。Local Outlier Factor (LOF)^[7]通过比较数据点局部邻域的密度来识别异常点,异常点通常具有较低的局部密度,与周围点的密度有显著差异。上述基于无监督的传统机器学习方法采用不同的正常值和异常值划分策略,将异常值区分开来,并取得了一定成效。但是这些传统方法并没有考虑时序数据中的时间依赖关系。此外,它们在处理大规模数据集时具有较高的计算复杂度且对方法中的参数格外敏感,这些缺陷导致它们的检测性能遇到瓶颈。

基于无监督的方法并不局限于传统的机器学习方法中,目前大量的深度学习模型也采用了无监督的方法来实现时序数据的异常检测,这些无监督的方法有一个共同点,即寻找两种不同的角度,使得正常数据和异常数据通过这两种不同的角度得到的特征空间表示或者重建结果有较大差异,从而区分正常点和异常点。Fu等^[8]设计了一个双分支结构,分别从时间和空间维度对输入数据进行建模,采用KL散度度量时间上的关联差异和空间上的关联差异,从而实现正常点和异常点的区分。DCdetector^[9]采用基于对比学习的双分支注意力结构,将输入数据进行分段处理,通过段间和段内视图学习排列不变表示,以扩大正常点和异常之间的表示差异。

受上述方法的启发,正常样本在多视角特征学习与重构过程中均能保持表征一致性,而异常样本因内在模式的不稳定性,其跨维度的表征学习或进一步重建都难以得到一致或近似的结果。同时,根据时间序列表示的不确定性原理^[10],在一个时域或频域中具有显著表征优势的异常数据,其在对应的另一域中的表征效果难以达到较高水平。基于此,本文提出了一种基于时频域注意力的异常检测模型(Time-Frequency Attention Based Model for Time Series Anomaly Detection, TFA-TSAD),该模型同时在时域和频域两个角度对输入数据进行特征表示和重建,通过对比数据在两个不同角度下的重建结果差异来捕获异常。除此之外,为了能够充分利用输入数据的不同数据模式,该模型利用Hodrick-Prescott (HP)滤波器对输入数据进行渐进式的分解,将数据分解为趋势项和周期项,其中趋势项代表时间序列中的长期变化趋势,周期项代表时间序列中的短期波动。最后,为了更有效地区分正常数据和异常数据,该方法改变了以往仅在最后阶段计算重建损失的策略,而是对重建过程中的每个阶段的损失进行加和平均。这一改进显著提升了模型的最终性能。

本文的主要贡献如下:

1)提出了一种新颖的结合了时域和频域信息的异常检测模型,该方法分别从时域和频域两个角度对输入数据执行

重建,通过度量数据在两个角度下的重建结果来区分正常数据和异常数据;

2)采用HP滤波器进行数据的渐进式分解,以充分探索输入数据中不同模式下的异常情况;

3)采用对重建过程中每个阶段的均方误差加和平均的方式来度量不同层之间频域和时域的重建输出,该方式赋予了模型更为优质的检测性能;

4)相较于其他基线模型,所提出的模型在多个包含单变量和多变量的数据集上取得了更好的结果。

2 相关工作

2.1 基于预测的异常检测模型

基于预测的异常检测模型的核心是利用先验数据来训练模型,以预测未来的数据,并通过比较预测序列和原始时间序列来检测异常。某个数据点在两个序列中的差异值超过某个阈值,则该数据点被视为异常。THOC^[11]利用多尺度扩张卷积循环神经网络和跳跃连接来捕获时序数据中的时间动态,并通过分层聚类过程获得多个超球体,从而定义了一种名为Multiscale Vector Data Description的单类目标。Hundman等^[12]提出一种基于长短期记忆网络(LSTM)的预测模型,这种循环神经网络结合一种互补的无监督、无参数的异常阈值方法,实现了对航天数据的异常数据监测。Ren等^[13]设计了一种基于光谱残差(SRs)和卷积神经网络(CNNs)的时间序列异常检测算法,用于Microsoft服务中的异常检测。Ahmad等^[14]提出一种称为分层时间记忆(HTM)方法的无监督模型,该方法基于在线序列记忆算法,用于数据流的实时异常检测。尽管基于预测的方法在建模下一个时间戳方面效果显著,但这些模型容易受到历史数据集的干扰。同时,这些模型对异常的定义较为敏感,可能无法捕捉一些潜在的非传统异常。

2.2 基于重建的异常检测模型

基于重建的异常检测模型通过设计适当的编码器将原始时间序列映射到潜在空间中,并设计对应的解码器根据上述潜在表示重建时间序列。如果原始序列和重建序列在某个时间点的差异值超过阈值,则该数据点被视为异常。变分自动编码器(Variational Auto-Encoders, VAE)^[15]是一个经典的基于重构的模型,最初有研究者提议使用有向概率模型执行有效的近似推理和学习过程。VAE被广泛应用于去噪、压缩、异常检测等应用。DAGMM^[16]是一种无监督异常检测模型,该模型利用深度自编码器为每个输入数据点生成低维表示和重构误差。变分自动编码器在基于重建的方法中应用广泛,对于编码器的设计也多种多样。Wang等^[17]采用简单的MLP(Multilayer Perceptron)和RNN作为网络的编码器,实现对空气质量数据的异常检测。Zhang等^[18]基于时间序列数据的时频域分析,在时域和频域以及趋势项和残差项分别对数据进行重建,采用TCN(Temporal Convolutional Network)作为编码器的主要组分,在多个真实数据集上实现了良好的异常检测效果。尽管上述方法已经取得了一定的成效,但多数方法仅在时域内建模,忽略了频域信息的作用。

2.3 基于 Transformer 的异常检测模型

Transformer^[19]自提出以来在自然语言处理领域取得了显著的成果,并在计算机视觉领域的后续工作中广泛应用,如今越来越多基于 Transformer 的模型也被用于时序任务。Anomaly Transformer^[20]提出了一种名为 Anomaly-Attention 的注意力机制,同时建模先验关联和序列关联,以体现关联差异,进而对正常数据和异常数据进行区分。Nam 等^[21]采用外部窗口和内部窗口嵌套的方式对齐时域重建和频域重建的异常分数,使得模型能够更好地识别模式异常的边界。TranAD^[22]提出了一个两阶段的编码器-解码器结构,并利用对抗训练得到一个稳定的重构模型。AnomalyBert^[23]提出了 4 种类型的退化方式,通过合成异常值来模拟异常行为,解决了现实世界中异常值较少的问题。

2.4 基于频域分析的时间序列模型

频域分析提供了对时间序列数据分析的不同视角,通过揭示数据的频率成分,可以更好地理解和利用数据。FreTS^[24]在频域中应用 MLP 进行时间序列预测,实现了全局感受野和频率聚焦。TimesNet^[25]利用快速傅里叶变换(FFT)从输入数据中提取最显著的 k 个周期,从而将一维时间序列转换为二维时频表示。FEDformer^[26]在 Transformer 结构中提出了傅里叶增强块和小波增强块,通过频域映射捕捉时间序列中的重要结构。CoST^[27]采用离散傅里叶变换(DFT)将中间特征映射到频域,促进表征内部的交互。本文最相关的工作是 TFAD^[18],它采用了融合时域和频域的双分支方法,并采用 TCN 作为主干网络。尽管频域分析在时序分析中应用广泛,但其在异常检测中的运用仍然有限。

2.5 基于时域分析的时间序列模型

时序数据通常涵盖多种复杂的数据模式,而时序分解技术有助于提取其独特的组成部分,从而更深入地理解和分析数据中的时间特征。Lei 等^[28]设计基于时序分解和频谱分析的异常检测新型深度框架。Autoformer^[29]通过设计内部分解块为深度预测模型赋予内在的渐进分解能力。MICN^[30]采用不同尺度的核大小对输入数据进行分解,验证了单独建模复杂和可变输入数据的必要性和有效性。TFDNet^[31]提出专注于解决长时序输入问题的输入序列分解方法,提高了预测的效率和稳健性。D3R^[32]提出针对长周期多变量时序的动态分解方法,有效利用外部信息来克服局部滑动窗口的局限性。

3 异常检测模型

3.1 问题定义

给定一段时间序列数据 \mathbf{X} :

$$\mathbf{X} = [x_{i-T+1}, x_{i-T+2}, \dots, x_i] \quad (1)$$

其中, $x_t \in \mathbb{R}^D$ 是这段序列中时刻 t 的观测值; D 是维度,即变量的个数; T 是这段时间序列的长度。要解决的问题是:对于给定的不含异常值的训练数据 \mathbf{X} ,用其训练异常检测模型,使得该模型能够识别出另外一段未知序列 $\mathbf{X}' = [x'_{i-T+1}, x'_{i-T+2}, \dots, x'_i]$ 中的每一个观测值是否为异常值,即 $x'_t \in \{0, 1\}$,其中 1 代表异常值,0 代表正常值。

3.2 整体架构

TFA-TSAD 的整体架构如图 1 所示,主要包含数据嵌入、渐进式模式分解模块、时域和频域建模块以及异常诊断模块。整个模型分 4 步来执行异常检测。

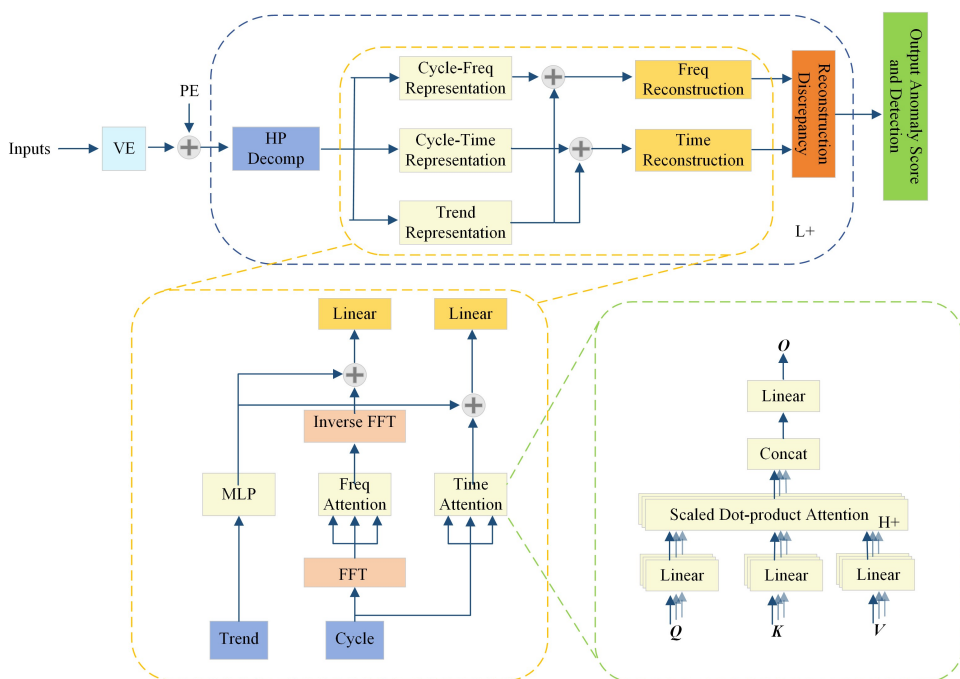


图 1 TFA-TSAD 模型的整体架构

Fig. 1 Overall architecture of TFA-TSAD model

第一步 原始输入数据首先通过嵌入层进行值嵌入(Value Embedding, VE)和位置嵌入(Positional Encoding, PE)。

第二步 将执行嵌入后的数据输入渐进式分解模块,该

模块将时序数据分解为周期项和趋势项。

第三步 将分解后的周期项送入时频域注意力模块,获取其在频域空间以及时域空间内的特征表示;趋势项被送入一个多层感知机(MLP)中得到其特征表示。随后将上述周

期项在时域和频域的特征表示分别与趋势项的特征表示进行加和,并送至时域和频域的重建层,分别得到时域和频域的重建输出,并度量时频域的重建差异。

第四步 重复第二步和第三步 L 次,其中 L 是超参数。最后一层输出异常分数和异常检测结果。

3.3 渐进式分解模块

时序数据可能包含复杂的数据模式,所以模型中采用了模式分解以便充分利用这些不同的模式,从找到不同模式中存在的异常。如图 2(a)和图 2(b)所示,蓝色线条表示原始数据,黄色线条表示趋势项,而绿色线条表示周期项,检测到的异常点用红点表示。显然,分解后异常更容易被检测到。

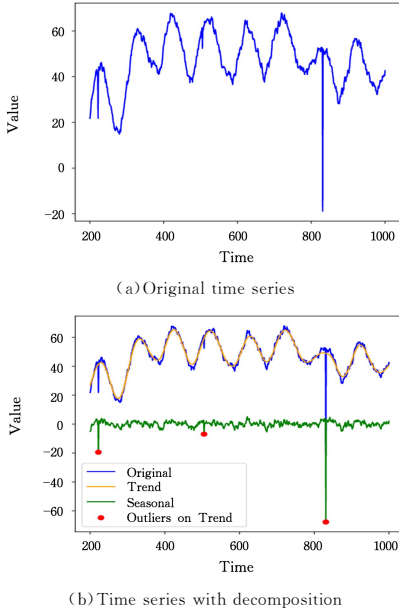


图 2 分解策略协助异常检测的实例(电子版为彩图)

Fig. 2 Example of decomposition helps time series anomaly detection

时序分解的方式有很多种,经典的分解方法包含加法分解、乘法分解和 STL 分解等。本文采用加法分解模型中的

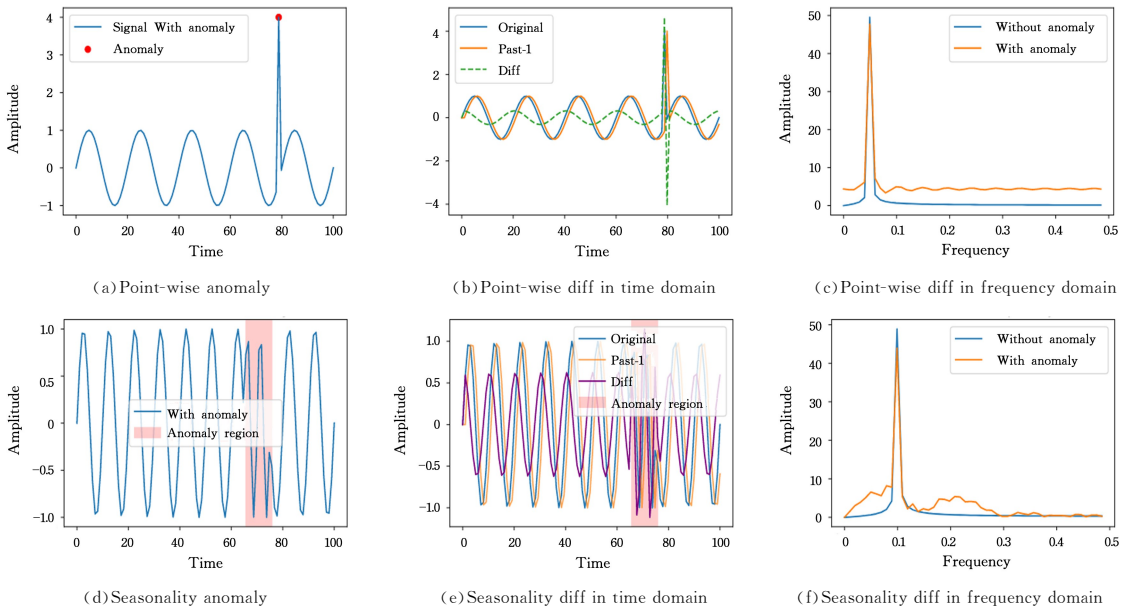


图 3 不同异常在时域和频域中的检测情况

Fig. 3 Detection of different anomalies in time and frequency domains

Hodrick-Prescott(HP)滤波器作为时序分解的方法,因为它分解方式简便,并且对于真实数据集的分解效果较好。

假定有一段时间序列 $y_t, t \in \{1, 2, \dots, T\}$ 包含了趋势项 τ_t 和周期项 c_t , 即 $y_t = \tau_t + c_t$ 。HP 滤波器是一个趋势滤波器,可以通过最小化式(2)得到其中的趋势项:

$$\min_{\tau} \left(\sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \right) \quad (2)$$

其中, λ 是平滑参数,用于控制趋势项和周期项的平滑程度。

与之前的方法不同,模型并不是只在开头使用一次分解,而是在每一层的开始都使用一次分解,即渐进式分解,从而充分利用数据在趋势项和周期项中的异常信息。这一设计也使模型能够深度挖掘数据中存在的异常点,使得正常数据和异常数据之间的区别更加明显。每一层分解后的周期项会送至时频域注意力中得到其在时域和频域中的特征表示,趋势项部分送至 MLP 中得到趋势项的特征表示,再分别与上述周期项在时域和频域中的特征表示进行加和后,送至时域和频域的重建模块得到对应域中的重建输出。

$$\mathbf{X}_i^t, \mathbf{X}_i^f = \text{HPDecomp}(\mathbf{X}_i) \quad (3)$$

其中, i 为当前编码器的层数; \mathbf{X}_i 为当前层的输入; $\mathbf{X}_i^t \in \mathbb{R}^{n \times d}$ 和 $\mathbf{X}_i^f \in \mathbb{R}^{n \times d}$ 分别为当前层的输入经过 HP 滤波器分解后得到的周期项和趋势项, d 为自注意力的嵌入维度, n 为序列的长度。

3.4 时频域建模块

时域和频域表示数据中的不同方面。时域中主要关注输入数据随时间的变化,即更关注整体趋势;频域则关注输入数据的频率分布和频率成分强度等,即更关注周期性。异常数据在时域中会表现为突然的波动或不连续性,而在频域中,这些异常会导致频率成分的变化,尤其是引入高频成分,从而使得从频域重建的结果与时域重建的结果表现出显著差异。而对于正常数据,从时域和频域重建的结果差异相对较小。

同时,针对不同种类的异常,在时域和频域中获取到异常信息的难易程度不同,如图 3 所示。

如图 3(b)和图 3(e)所示,时域中对点异常的捕捉更为敏感,因为它与正常值之间的差异较大,即图中的 diff 较大;而针对季节性或周期性的异常来看,它在时域中与正常值的差距并不明显,即图中 diff 相对较小。而在频域中对于季节性或周期性异常的捕捉会更加容易,因为有异常和无异常的数据有明显的不同数量的峰值,而在时域中差别并不明显,如图 3(c)和图 3(f)所示。因此,为了尽可能检测到多的异常,需要结合时域与频域进行异常检测。

本文设计了一个结合时域和频域进行异常检测的模块,如图 1 所示。在输入数据经过 HP 滤波器分解为周期项 \mathbf{X}_i^c 和趋势项 \mathbf{X}_i^t 后,前者通过时频域注意力模块得到周期项的特征表示,后者通过一个 MLP 得到趋势项的特征表示。上述周期项在时域和频域的特征表示分别与趋势项的特征表示进行相加,并送至时域和频域的重建层,分别得到时域和频域的重建输出。

3.4.1 时域注意力

对于时域分支,采取了经典的多头自注意力机制,多头自注意力提供了自注意力对时序数据不同角度的关注,其示意图如图 1 所示。对于第 i 层输入时域分支的周期项 $\mathbf{X}_{\text{time},i}^c \in \mathbb{R}^{n \times d}$,经过多头自注意力机制后得到对应的特征表示输出:

$$\mathbf{Z}_{\text{time},i} \in \mathbb{R}^{n \times d}, \mathbf{Q}_j = \mathbf{X}_{\text{time},i}^c \mathbf{W}_j^Q, \mathbf{K}_j = \mathbf{X}_{\text{time},i}^c \mathbf{W}_j^K, \mathbf{V}_j = \mathbf{X}_{\text{time},i}^c \mathbf{W}_j^V \quad (4)$$

$$\mathbf{Z}_j = \text{softmax}\left(\frac{\mathbf{Q}_j \mathbf{K}_j^T}{\sqrt{d_k}}\right) \mathbf{V}_j \quad (5)$$

$$\mathbf{Z}_{\text{time},i} = \text{Concat}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_h) \mathbf{W}^O \quad (6)$$

其中, h 表示多头注意力的头数; $\mathbf{W}_j^Q \in \mathbb{R}^{d \times \frac{d_k}{h}}, \mathbf{W}_j^K \in \mathbb{R}^{d \times \frac{d_k}{h}}, \mathbf{W}_j^V \in \mathbb{R}^{d \times \frac{d_v}{h}}$ 分别表示第 j 个注意力头的可学习参数矩阵; $\mathbf{W}^O \in \mathbb{R}^{d_o \times d}$ 表示多头自注意力的输出权重矩阵; $\mathbf{Q}_j \in \mathbb{R}^{n \times \frac{d_k}{h}}, \mathbf{K}_j \in \mathbb{R}^{n \times \frac{d_k}{h}}, \mathbf{V}_j \in \mathbb{R}^{n \times \frac{d_v}{h}}$ 分别代表查询、键、值矩阵。

周期项在时域分支通过多头自注意力后得到的特征表示,将会与趋势项通过 MLP 后得到的特征表示进行相加,用于计算当前层在时域中的重建输出。

3.4.2 频域注意力

对于频域分支,首先要将时域中的数据转换至频域。转换的方式有多种,如小波变换和傅里叶变换等,本文采用傅里叶变换。对于时域数据 $\mathbf{H}(v)$,经过傅里叶变换后,得到对应的频域表示 $\mathcal{H}(f)$,如式(7)所示:

$$\begin{aligned} \mathcal{H}(f) &= \int_{-\infty}^{\infty} \mathbf{H}(v) e^{-j2\pi f v} dv \\ &= \int_{-\infty}^{\infty} \mathbf{H}(v) \cos(2\pi f v) dv + j \int_{-\infty}^{\infty} \mathbf{H}(v) \sin(2\pi f v) dv \end{aligned} \quad (7)$$

其中, f 是频率变量; v 是积分变量; j 是虚数单位,定义为 -1 的平方根; $\int_{-\infty}^{\infty} \mathbf{H}(v) \cos(2\pi f v) dv$ 是 \mathcal{H} 的实部,记作 $\text{Re}(\mathcal{H})$; $\int_{-\infty}^{\infty} \mathbf{H}(v) \sin(2\pi f v) dv$ 是 \mathcal{H} 的虚部,记作 $\text{Im}(\mathcal{H})$ 。在频域中的建模结束后,需要将频域中学习到的信息转换到时域中,如式(8)所示:

$$\begin{aligned} \mathbf{H}(v) &= \int_{-\infty}^{\infty} \mathcal{H}(f) e^{j2\pi f v} df \\ &= \int_{-\infty}^{\infty} (\text{Re}(\mathcal{H}(f)) + j \text{Im}(\mathcal{H}(f))) e^{j2\pi f v} df \end{aligned} \quad (8)$$

为了方便表示,将时域到频域的转换记作 FFT,将频域到时域的转换记作 IFFT。与时域中类似,对于第 i 层输入频域分支的周期项分量 $\mathbf{X}_{\text{freq},i}^c \in \mathbb{R}^{n \times d}$,在频域中采用多头自注意力机制得到对应的特征表示输出 $\mathbf{Z}_{\text{freq},i}^c \in \mathbb{R}^{n \times d}$ 。上述过程表示如下:

$$\mathbf{Q}_j = \mathbf{X}_{\text{freq},i}^c \mathbf{W}_j^Q, \mathbf{K}_j = \mathbf{X}_{\text{freq},i}^c \mathbf{W}_j^K, \mathbf{V}_j = \mathbf{X}_{\text{freq},i}^c \mathbf{W}_j^V \quad (9)$$

$$\begin{cases} \mathbf{Q}_f = \text{FFT}(\mathbf{Q}_j) \\ \mathbf{K}_f = \text{FFT}(\mathbf{K}_j) \\ \mathbf{V}_f = \text{FFT}(\mathbf{V}_j) \end{cases} \quad (10)$$

$$\mathbf{Z}_j = \text{softmax}\left(\frac{\mathbf{Q}_f \mathbf{K}_f^T}{\sqrt{d_k}}\right) \mathbf{V}_f \quad (11)$$

$$\mathbf{Z}_{\text{freq},i}^c = \text{Concat}(\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_h) \mathbf{W}^O \quad (12)$$

$$\mathbf{Z}_{\text{freq},i}^c = \text{IFFT}(\mathbf{Z}_{\text{freq},i}^c) \quad (13)$$

其中, $\mathbf{Q}_f, \mathbf{K}_f, \mathbf{V}_f$ 分别为 $\mathbf{Q}_j, \mathbf{K}_j, \mathbf{V}_j$ 经过傅里叶变换在频域中的表示。

3.4.3 时频域重建输出

在得到周期项在时域的特征表示输出 $\mathbf{Z}_{\text{time},i} \in \mathbb{R}^{n \times d}$ 和频域中的特征表示输出 $\mathbf{Z}_{\text{freq},i}^c \in \mathbb{R}^{n \times d}$ 后,将其分别加上趋势项经过 MLP 得到特征表示输出 $\mathbf{Z}_i \in \mathbb{R}^{n \times d}$,得到周期项在当前层最终的时域特征表示输出 $\mathbf{Z}_{\text{time},i} \in \mathbb{R}^{n \times d}$ 和频域中的特征表示输出 $\mathbf{Z}_{\text{freq},i} \in \mathbb{R}^{n \times d}$:

$$\mathbf{Z}_i = \text{MLP}(\mathbf{X}_i^t) \quad (14)$$

$$\mathbf{Z}_{\text{time},i} = \mathbf{Z}_{\text{time},i}^c + \mathbf{Z}_i \quad (15)$$

$$\mathbf{Z}_{\text{freq},i} = \mathbf{Z}_{\text{freq},i}^c + \mathbf{Z}_i \quad (16)$$

随后,设置一个简单的线性层作为解码器,分别得到当前层的时域重建输出 $\mathbf{O}_{\text{time},i} \in \mathbb{R}^{n \times d}$ 和频域重建输出 $\mathbf{O}_{\text{freq},i} \in \mathbb{R}^{n \times d}$,并在当前层中计算上述时域和频域的重建差异,如式(17)和式(18)所示:

$$\mathbf{O}_{\text{time},i} = \text{Linear}(\mathbf{Z}_{\text{time},i}) \quad (17)$$

$$\mathbf{O}_{\text{freq},i} = \text{Linear}(\mathbf{Z}_{\text{freq},i}) \quad (18)$$

3.5 异常诊断模块

为了度量在时域和频域中的特征表示及重建结果的差异,以及在异常检测中扩大正常值与异常值之间的差异,本文在每一个重建层之间都采用均方误差度量当前层的时域重建结果和频域重建结果之间的差异,并将所有层的均方误差进行相加和平均,表示为:

$$\text{Loss} = \frac{1}{LT} \sum_{i=1}^L \sum_{t=1}^T (\mathbf{O}_{\text{time},i,t} - \mathbf{O}_{\text{freq},i,t})^2 \quad (19)$$

其中, T 为输入时间序列的长度, L 为重建层的数量, $\mathbf{O}_{\text{time},i,t}$ 和 $\mathbf{O}_{\text{freq},i,t}$ 分别代表第 i 层对应的时域和频域的重建结果。

在计算异常得分时,同样采样上述求解均方误差的方式,当求解的误差结果大于阈值时,判定该点为异常点。其中阈值的计算结果是根据模型训练集和测试集上的计算结果拼接而成的,是一个动态的值。

4 实验

4.1 数据集

本文采用 5 个多变量的真实数据集、1 个单变量的真实数据集和 2 个多变量的合成数据集来测试模型的性能,这些数据集的信息如表 1 所列。其中 SMD(Server Machine Data-

set) 是一个为期 5 周的数据集, 收集自一家互联网公司计算集群, 包含了 28 台机器的资源利用率访问记录; MSL (Mars Science Laboratory dataset) 由 NASA 收集, 展示了火星探测器的传感器和执行器数据的状态; SMAP (Soil Moisture Active Passive Dataset) 也是由 NASA 收集的, 展示了火星漫游车使用的土壤样本和遥测信息, 与 MSL 相比, SMAP 数据集包含更多的点异常; SWaT (Secure Water Treatment) 是一个基于 51 个维度传感器的数据集, 收集自持续运行中的关键

基础设施系统; PSM (Pooled Server Metrics Dataset) 是来自 eBay 服务器机器的公开数据集, 包含 25 个维度; UCR 是由 KDD2021 的多数据集时间序列异常检测竞赛提供的, 包含来自各种自然来源的 250 个子数据集; NIPS-TS-GECCO 是一个用于物联网的饮用水质量数据集, 发布于 2018 年的遗传与进化计算会议; NIPS-TS-SWAN 是一个开放访问的综合多变量时间序列基准数据集, 提取自 Spaceweather HMI 活动区图像系列中的太阳光球矢量磁图。

表 1 数据集描述

Table 1 Description of each dataset

数据集	来源	维度	训练集	验证集	测试集	异常率
SMD	Internet Server Machine	38	566 724	141 681	708 420	0.042
MSL	NASA Space Sensors	55	46 653	11 664	73 729	0.105
SMAP	NASA Space Sensors	25	108 146	27 037	427 617	0.128
SWaT	Infrastructure System	51	396 000	99 000	449 919	0.121
PSM	eBay Server Machine	25	105 984	26 497	87 841	0.278
NIPS-TS-GECCO	Water Quality for IoT	9	55 408	13 852	69 261	0.011
NIPS-TS-SWAN	Space(Solar) Weather	38	48 000	12 000	60 000	0.326
UCR	Various Natural Sources	1	1 790 679	447 670	6 143 541	0.006

4.2 基线和评价指标

将本文模型与 13 个基线模型进行对比, 主要包括传统方法 OCSVM^[5] 和 IForest^[4], 基于聚类的方法 ITAD^[33] 和 THOC^[11], 自回归模型 CL-MPPCA^[34] 和 LSTM^[12], 以及基于重建的方法 InterFusion^[35], OmniAnomaly^[36] 和 LSTM-VAE^[37]。除此之外, 本文还额外引入了 4 个基于 Transformer 的异常检测方法来验证模型的异常检测性能, 包括 Anomaly Transformer^[20], DCdetector^[9], TranAD^[22] 以及一个目前最新的模型 TFMAE^[38]。TFMAE 模型使用两个基于 Transformer 的自编码器, 分别结合了基于窗口的时间掩蔽策略和基于幅度的频率掩蔽策略, 在没有异常偏差的情况下学习知识, 并通过提取的正常信息重建异常。

本文采用广泛使用的 3 个指标, 即精确率 (Precision)、召回率 (Recall) 和 F1 分数来评估模型的性能。

$$P = \frac{TP}{TP + FP} \quad (20)$$

$$R = \frac{TP}{TP + FN} \quad (21)$$

$$F1 = \frac{2 \times P \times R}{R + P} \quad (22)$$

其中, TP 为真正例, 即模型将真实的异常样本正确地识别为异常的样本数量; FP 为假正例, 即模型将实际上是正常的样本错误地识别为异常的样本数量; FN 为假负例, 即模型将实际上是异常的样本错误地识别为正常的样本数量。

4.3 实验设置

与 Anomaly Transformer 等基线模型中的设置一致, 本文将所有数据集均分成训练集、测试集和验证集。所有数据集的窗口大小统一设置为 100, 除了 SWaT 数据集的 batch size 设置为 128 外, 其余数据集的 batch size 均设置为 256。模型总共包含 3 层编码器, 自注意力中的嵌入维度 d 均设置为 256, 自注意力的头数均设置为 8。模型训练过程中采用 Adam 作为优化器, 初始学习率设置为 0.0001。为了进行公平的对比, 所有数据集都只训练 3 轮。本文实验中的所有模

型训练均在一张 NVIDIA GeForce GTX 3090 (24 GB) GPU 上进行, 且模型均使用 PyTorch 框架搭建。

另外, 在最终计算准确率、召回率和 F1 得分时, 采用广泛使用的异常调整策略^[9,20]。该调整策略的核心机制是当连续异常区间内检测到至少一个有效异常点时, 即可判定该异常段的完整检测。该策略的合理性源于实际监测需求, 即在持续异常状态下, 单个异常点的有效识别足以触发预警机制。

4.4 实验结果

4.4.1 多变量的真实数据集

为了验证本文模型的有效性, 首先在 5 个真实数据集上进行验证, 并采用 13 个基线模型进行对比验证。为了避免偶然性, 所有实验均重复 3 次取平均值, 实验结果如表 2 所列, 其中加粗表示最优结果, 下划线表示次优结果。

结果表明, TFA-TSAD 在所有数据集上均取得了最优或者次优的结果。特别地, 在 MSL, SMAP, SWaT 和 PSM 数据集上, TFA-TSAD 的 F1 表现均比最新的模型 TFMAE 更加出色。值得注意的是, 模型在 SMD 上的结果普遍偏低, 这可能是由于该数据集中的数据量在 5 个数据集中最大, 同时它对应的异常率在所有数据集中最小, 所以相对于另外 4 个数据集, 在该数据集上检测结果的精确率、召回率和 F1 分数都有不同程度的下降, 即便如此, TFA-TSAD 在该数据集上仍然取得了有竞争力的结果。

DCdetector 与 Anomaly Transformer 都可以看作从时域中的局部信息和全局信息两个角度出发, 在时域中通过度量上述两个角度下的特征表示差异来进行正常点和异常点的区分。与上述两个方法不同的是, 本文采用时域和频域两个角度结合的方式, 且取得了更优的结果, 证明了所提出的时域和频域结合的方法的有效性。同时, TFMAE 也是采用时域和频域结合的方法, 但是采用 KL 散度量时域和频域中的特征表示差异, 与 TFMAE 的结果进行比较, 证明了本文方法度量时域和频域两个角度下的重建差异对于区分正常点和异常点是有效的, 且后续消融实验结果也表明了所选择的度量方式的有效性。

表2 在所有数据集上的结果对比
Table 2 Comparison of results on all datasets

数据集		SMD			MSL			SMAP			SWaT			PSM		
模型/指标	Venue	P	R	F1	P	R	F1	P	R	F1	P	R	F1	P	R	F1
OCSVM	MachLearn-04	44.34	76.72	56.19	59.78	86.87	70.82	53.85	59.07	56.34	45.39	49.22	47.23	62.75	80.89	70.67
IForest	ICDM-08	42.31	73.29	53.64	53.94	86.54	66.45	52.39	59.07	55.53	49.29	44.95	47.02	76.09	92.45	83.48
ITAD	CIKM-20	86.22	73.71	79.48	69.44	84.09	76.07	82.42	66.89	73.85	63.13	52.08	57.08	72.80	64.02	68.13
THOC	NeurIPS-20	79.76	<u>90.95</u>	84.99	88.45	90.97	89.69	92.06	89.34	90.68	83.94	86.36	85.13	88.14	90.99	89.54
CL-MPPCA	KDD-19	82.36	76.07	79.09	73.71	88.54	80.44	86.13	63.16	72.88	76.78	81.50	79.07	56.02	99.93	71.80
LSTM	KDD-18	78.55	85.28	81.78	85.45	82.50	83.95	89.41	78.13	83.39	86.15	83.27	84.69	76.93	89.64	82.80
InterFusion	KDD-21	87.02	85.43	86.22	81.28	92.70	86.62	89.77	88.52	89.14	80.59	85.58	83.01	83.61	83.45	83.52
OmniAno	KDD-19	83.68	86.82	85.22	89.02	86.37	87.67	92.49	81.99	86.92	81.42	84.30	82.83	88.39	74.46	80.83
LSTM-VAE	RAL-18	75.76	90.08	82.30	85.49	79.94	82.62	92.20	67.75	78.10	76.00	89.50	82.20	73.62	89.92	80.96
AnomalyTrans	ICLR-22	<u>88.83</u>	91.76	90.27	91.94	95.22	93.55	93.61	96.05	94.76	88.23	<u>100.00</u>	93.73	97.11	97.53	97.32
TranAD	VLDB-23	74.30	81.65	77.80	90.72	94.73	92.68	93.12	71.33	80.78	<u>94.23</u>	94.36	94.29	<u>97.44</u>	98.19	<u>97.92</u>
DCdetector	KDD-23	85.80	83.63	84.70	<u>92.27</u>	<u>97.48</u>	<u>94.80</u>	94.29	<u>98.52</u>	<u>96.36</u>	<u>92.97</u>	99.89	<u>96.30</u>	97.07	97.89	97.47
TFMAE	ICDE-24	91.54	87.46	<u>89.46</u>	92.60	92.69	92.64	<u>94.36</u>	98.03	96.16	97.36	95.16	96.24	97.40	98.16	97.78
TFA-TSAD(ours)	—	86.50	87.58	87.04	91.74	98.34	94.92	94.43	99.08	96.70	93.25	100.00	96.51	97.53	<u>98.81</u>	98.17

4.4.2 多变量的合成数据集

为了进一步验证模型处理不同异常情况的能力,本文在NIPS-TS-GECCO和NIPS-TS-SWAN两个合成数据集上对模型进行验证。这两个数据集比上述5个真实数据集具有更多的异常种类,对应的异常检测也更具有挑战性。尽管这两个数据集拥有最低的(NIPS-TS-GECCO:1.1%)和最高的异常比(NIPS-TS-SWAN:32.6%),但是本文模型的检测性能优于其他方法且有较大幅度提升,说明了本文模型在多变量数据集上表现良好,能有效应对数据的多样性。特别地,本文模型在NIPS-TS-GECCO数据集上的F1分数在次优模型的基础上提升了15.2%,实验结果如表3所列。

表3 在NIPS-TS数据集上的结果对比

Table 3 Comparison of results on NIPS-TS dataset (%)

DataSet Metric	NIPS-TS-GECCO			NIPS-TS-SWAN		
	P	R	F1	P	R	F1
AnomalyTrans	23.17	58.08	33.12	90.73	47.42	62.28
Dcdetector	38.25	59.73	46.63	95.48	59.55	73.35
TFMAE	25.36	32.86	28.62	94.23	57.65	71.53
TFA-TSAD	43.91	69.18	53.72	95.49	59.92	73.63

4.4.3 单变量的真实数据集

上述数据集均为多变量数据集,而单变量数据集中缺少数据不同维度之间的依赖关系,进一步提升了检测难度。为了验证所提模型对于单变量数据的异常检测能力,还对模型在UCR和其增强数据集UCR-AUG中的每个子数据集上的异常检测表现进行了测试,实验结果如表4所列。

表4 在UCR数据集上的结果对比

Table 4 Comparison of results on UCR dataset (%)

DataSet Metric	UCR-AUG			UCR		
	P	R	F1	P	R	F1
AnomalyTrans	88.63	77.55	82.72	33.18	60.32	42.81
TFMAE	73.98	63.74	68.48	32.86	45.05	38.00
TFA-TSAD	90.28	85.94	88.06	33.37	61.79	43.34

结果显示,本文模型在单变量数据集上相对于基线模型也能取得最好的结果,且在UCR-AUG数据集上的F1分数较次优模型提升了6.5%。

4.5 消融实验

为了验证模型中使用到的渐进式分解策略和所设计的对每一层重建误差做加和平均方式的有效性,设计了基于TFA-TSAD的多个变体模型,结果比较如表5所列。

I与II的实验结果表明,与不使用分解策略相比,在模型中使用HP滤波器进行模式分解的方式是有效的,因为模式分解能使模型分别发现周期项和趋势项下的异常。

II与III的实验结果表明,与使用分解但不使用渐进式分解相比,在模型的每一层重建网络中使用渐进式分解更加有效。

III与IV的对比结果说明,采用度量重建差异的方式进行异常检测比仅使用KL散度量特征空间的表示差异效果更好。而IV与TFA-TSAD的结果对比显示,采用传统的只在模型最后输出计算时频域重建差异的效果比对每一层的重建差异进行加和平均的效果差,体现了本文模型设计的有效性。

表5 不同模型变体的F1分数比较

Table 5 Comparison of F1 scores for different model variants (%)

No.	Model					SMD	MSL	SMAP	SWaT	PSM	Avg F1
	Decomposition	Progressive decomposition	Loss with KL	Loss with mse	Loss with avg mse						
I					✓	85.84	94.58	94.05	96.24	97.79	93.81
II	✓				✓	85.60	94.32	95.89	96.30	97.76	94.13
III		✓	✓			86.11	92.89	91.55	96.34	97.96	93.24
IV		✓		✓		85.29	94.61	96.23	96.19	97.89	94.20
TFA-TSAD		✓			✓	87.04	94.92	96.70	96.51	98.17	94.67

4.6 参数影响

本文探究了不同窗口大小、不同的编码器层数以及注意力中不同的嵌入维度对5个真实数据集F1分数的影响,实验结果如图4所示。

图4(a)展示了在5个数据集上采用50,100,150,200和250的窗口大小对模型异常检测效果的影响。实验结果显示,本文模型的检测性能表现为:在多数数据集上对窗口大小的变化是相对稳定的,并且当窗口大小为100时便能提供相对其他窗口大小而言最好的性能。过小的窗口会带来检测性能不佳,而过大的窗口则带来更大的内存开销和更长的计算时间。

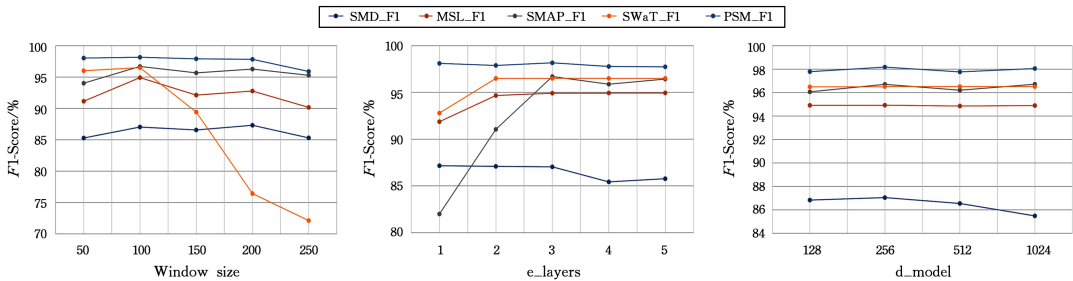


图4 模型中主要超参数的敏感性研究

Fig. 4 Parameter sensitivity studies of hyper-parameters in the proposed model

结束语 本文提出了一种基于时域和频域注意力的异常检测模型 TFA-TSAD。该模型首先创新性地利用渐进式的分解策略在每一层中都对输入数据进行分解,以便充分捕捉输入数据中不同模式的异常。其次以多头自注意力为核心,从时域和频域两个角度对输入数据进行特征表示和重建。最后使用均方误差加和平均的方式度量时域和频域中的重建差异,判定数据中是否存在异常值。TFA-TSAD在多个真实数据集中均取得了优于其他模型的性能,表明了该模型的有效性。

时序数据之间不仅存在时间上的依赖关系,还可能存在特征维度之间的依赖关系,也可能存在空间上的依赖关系。在未来的工作中,应该进一步探索特征维度和空间上的依赖关系对于异常检测的影响,提高异常检测性能。同时,可以从不同角度出发,对正常点和异常点的差异性做更进一步的探索,找到针对输入数据更有效的异常检测方式。

参考文献

- [1] LAI K H, ZHA D, XU J, et al. Revisiting time series outlier detection: Definitions and benchmarks [C] // Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track. 2021.
- [2] LI C, LAMMIE C, DONG X, et al. Seizure detection and prediction by parallel memristive convolutional neural networks [J]. IEEE Transactions on Biomedical Circuits and Systems, 2022, 16(4): 609-625.
- [3] GOLMOHAMMADI K, ZAIANE O R. Time series contextual anomaly detection for detecting market manipulation in stock market [C] // 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2015: 1-10.
- [4] LIU F T, TING K M, ZHOU Z H. Isolation forest [C] // 2008 Eighth IEEE International Conference on Data Mining. IEEE,

图4(b)显示了不同编码器层数对5个数据集上的F1分数的影响,结果表明,不同的编码器层数对SMAP数据集上的性能影响最大,对PSM的性能影响最小。总体来看,多数数据集在编码器层数为3时即可达到性能饱和。过少的层数可能无法达到最好的检测效果,而过多的层数则会带来更大的计算消耗。

图4(c)显示了嵌入维度 d_{model} 对5个数据集上F1分数的影响,可以看出,嵌入维度并没有对模型性能产生较大的影响,这不仅说明了本文模型的稳定性,也说明了选择 d_{model} 为256是对性能表现和计算复杂度的权衡考虑。

2008:413-422.

- [5] VAPNIK V, CHERVONENKIS A Y. A class of algorithms for pattern recognition learning [J]. Avtomatikai Telemekhanika, 1964, 25(6): 937-945.
- [6] MACQUEEN J. Some methods for classification and analysis of multivariate observations [C] // Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 1967: 281-297.
- [7] OEHMCKE S, ZIELINSKI O, KRAMER O. Event detection in marine time series data [C] // KI 2015: Advances in Artificial Intelligence, 38th Annual German Conference on AI. Springer, 2015: 279-286.
- [8] FU S, GAO X, ZHAI F, et al. A time series anomaly detection method based on series-parallel transformers with spatial and temporal association discrepancies [J]. Information Sciences, 2024, 657: 119978.
- [9] YANG Y, ZHANG C, ZHOU T, et al. Dcdetector: Dual attention contrastive representation learning for time series anomaly detection [C] // Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023: 3033-3045.
- [10] GRÖCHENIG K. Uncertainty principles for time-frequency representations [M] // Advances in Gabor Analysis. 2003: 11-30.
- [11] 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.
- [12] HUNDMAN K, CONSTANTINO 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. 387-395.
- [13] REN H, XU B, WANG Y, et al. Time-series anomaly detection

- service at microsoft[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 3009-3017.
- [14] AHMAD S, LAVIN A, PURDY S, et al. Unsupervised real-time anomaly detection for streaming data [J]. *Neurocomputing*, 2017, 262: 134-147.
- [15] KINGMA D P. Auto-encoding variational bayes [J]. arXiv: 1312.6114, 2013.
- [16] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]//International Conference on Learning Representations, 2018.
- [17] WANG Z, FENG J, FU Q, et al. Quality control of online monitoring data of air pollutants using artificial neural networks [J]. *Air Quality, Atmosphere & Health*, 2019, 12: 1189-1196.
- [18] ZHANG C, ZHOU T, WEN Q, et al. TFAD: A decomposition time series anomaly detection architecture with time-frequency analysis[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022: 2497-2507.
- [19] VASWANI A. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000-6010.
- [20] XU J. Anomaly transformer: Time series anomaly detection with association discrepancy [J]. arXiv:2110.02642, 2021.
- [21] NAM Y, YOON S, SHIN Y, et al. Breaking the Time-Frequency Granularity Discrepancy in Time-Series Anomaly Detection [C]//Proceedings of the ACM on Web Conference, 2024: 4204-4215.
- [22] TULI S, CASALE G, JENNINGS N R. Tranad: Deep transformer networks for anomaly detection in multivariate time series data [J]. arXiv:220107284, 2022.
- [23] JEONG Y, YANG E, RYU J H, et al. Anomalybert: Self-supervised transformer for time series anomaly detection using data degradation scheme [J]. arXiv:2305.04468, 2023.
- [24] YI K, ZHANG Q, FAN W, et al. Frequency-domain MLPs are more effective learners in time series forecasting [J]. arXiv: 2311.06184, 2023.
- [25] WU H, HU T, LIU Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis [J]. arXiv: 2210.02186, 2022.
- [26] ZHOU T, MA Z, WEN Q, et al. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting[C]//International Conference on Machine Learning, PMLR, 2022: 27268-27286.
- [27] WOO G, LIU C, SAHOO D, et al. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting [J]. arXiv:2202.01575, 2022.
- [28] LEI T, GONG C, CHEN G, et al. A novel unsupervised framework for time series data anomaly detection via spectrum decomposition [J]. *Knowledge-Based Systems*, 2023, 280: 111002.
- [29] WU H, XU J, WANG J, et al. Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting [J]. arXiv:2106.13008, 2021.
- [30] WANG H, PENG J, HUANG F, et al. Micn: Multi-scale local and global context modeling for long-term series forecasting [C]//The Eleventh International Conference on Learning Representations, 2023.
- [31] LUO Y, LYU Z, HUANG X. TFDNet: Time-Frequency Enhanced Decomposed Network for Long-term Time Series Forecasting [J]. arXiv:230813386, 2023.
- [32] WANG C, ZHUANG Z, QI Q, et al. Drift doesn't matter: dynamic decomposition with diffusion reconstruction for unstable multivariate time series anomaly detection [C]//Proceedings of the 37th International Conference on Neural Information Processing Systems, 2024: 10785-10774.
- [33] 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, 2020: 2733-27340.
- [34] 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, 2019: 2123-21233.
- [35] LI Z, ZHAO Y, HAN J, 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.
- [36] SU Y, ZHAO Y, NIU C, 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.
- [37] 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.
- [38] FANG Y, XIE J, ZHAO Y, et al. Temporal-Frequency Masked Autoencoders for Time Series Anomaly Detection [C]//2024 IEEE 40th International Conference on Data Engineering (ICDE). IEEE, 2024: 1228-1241.



XU Jingtao, born in 1999, postgraduate. His main research interest is time series anomaly detection and forecasting.



YANG Yan, born in 1964, Ph.D, professor, Ph.D supervisor, is a member of CCF(No. 06877D). Her main research interests include artificial intelligence, big data analysis and mining.