



计算机科学

COMPUTER SCIENCE

DCDAD:考虑上下文依赖差异化的时间序列异常检测模型

廖思睿, 黄飞虎, 战鹏祥, 彭舰, 张凌浩

引用本文

廖思睿, 黄飞虎, 战鹏祥, 彭舰, 张凌浩. [DCDAD:考虑上下文依赖差异化的时间序列异常检测模型](#)[J]. 计算机科学, 2025, 52(6): 106-117.

LIAO Sirui, HUANG Feihu, ZHAN Pengxiang, PENG Jian, ZHANG Linghao. [DCDAD:Differentiated Context Dependency for Time Series Anomaly Detection Method](#) [J]. Computer Science, 2025, 52(6): 106-117.

相似文献推荐 (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)

[一种基于直接反馈对齐的精确脉冲时间学习规则](#)

Learning Rule with Precise Spike Timing Based on Direct Feedback Alignment

计算机科学, 2025, 52(3): 260-267. <https://doi.org/10.11896/jsjcx.240100195>

[基于时间依赖和变量交互的多元时间序列预测](#)

Multivariate Time Series Forecasting Based on Temporal Dependency and Variable Interaction

计算机科学, 2025, 52(2): 67-79. <https://doi.org/10.11896/jsjcx.240100167>

[面向工业品缺陷检测的对比表示学习](#)

Contrastive Representation Learning for Industrial Defect Detection

计算机科学, 2025, 52(1): 210-220. <https://doi.org/10.11896/jsjcx.240100202>

[融合多图卷积与层级池化的文本分类模型](#)

Text Classification Method Based on Multi Graph Convolution and Hierarchical Pooling

计算机科学, 2024, 51(7): 303-309. <https://doi.org/10.11896/jsjcx.230400164>

[计及风电的发电商报价多智能体模型](#)

Multi-agent Based Bidding Strategy Model Considering Wind Power

计算机科学, 2024, 51(6A): 230600179-8. <https://doi.org/10.11896/jsjcx.230600179>

DCDAD:考虑上下文依赖差异化的时间序列异常检测模型

廖思睿¹ 黄飞虎¹ 战鹏祥¹ 彭舰¹ 张凌浩²

¹ 四川大学计算机学院 成都 610065

² 国网四川省电力公司电力科学研究院 成都 610000

(2022223045124@stu.scu.edu.cn)

摘要 时间序列异常检测旨在检测时间序列中与正常数据不符的时间点或片段。如何充分利用时间序列中的上下文信息以提升检测精度,是目前构建异常检测模型的关键。然而,现有方法未充分考虑数据中上下文依赖关系的差异性,也缺乏对异常样本的建模,导致正常和异常样本区分度不明显,检测效果欠佳。因此,提出了一种考虑上下文依赖差异化的异常检测(Differentiated Context Dependency for Time Series Anomaly Detection, DCDAD)模型用于时序异常检测。DCDAD模型通过自注意力捕捉时间维度的上下文依赖,并在聚类过程中学习用于区分正、异常样本的超球面。采用异常注入思想对数据集进行扩充,解决异常样本稀缺的问题,并针对性地设计了差异化学习的目标函数,扩大正、异常样本的差异性,进而提升异常检测性能。在5个真实时序数据集上进行了大量实验,在 F_1 分数上相比于现有最先进的算法提升了约1.2%,证实了以差异化方式学习上下文依赖关系可提升模型的异常检测效果,同时参数敏感性分析和消融实验的结果也验证了DCDAD模型的稳定性以及有效性。

关键词: 时序异常检测; 差异化表示学习; 上下文依赖关系; 超球面

中图分类号 TP183

DCDAD: Differentiated Context Dependency for Time Series Anomaly Detection Method

LIAO Sirui¹, HUANG Feihu¹, ZHAN Pengxiang¹, PENG Jian¹ and ZHANG Linghao²

¹ College of Computer Science, Sichuan University, Chengdu 610065, China

² State Grid Sichuan Electric Power Research Institute, Chengdu 610000, China

Abstract Time series anomaly detection aims to identify data points or segments in a time series that deviate from normal patterns. Enhancing detection accuracy by effectively utilizing contextual information in time series is a key role in constructing anomaly detection models. However, existing methods inadequately consider the differential context dependency in the data and lack explicit modeling of anomalous samples, resulting in poor discrimination between normal and anomalous samples and suboptimal detection performance. Therefore, this paper proposes a model that considers differential context dependency for time series anomaly detection (DCDAD), which enhances to learn the differential representations of context dependency. The DCDAD model captures temporal context dependency using self-attention mechanisms and learns hyperspheres for discriminating between normal and anomalous samples during the clustering process. By adopting the concept of anomaly injection, the dataset is augmented to address the issue of limited anomalous samples. Additionally, a targeted objective function for differentiated learning is designed to amplify the differences between normal and anomalous samples, thereby improving the anomaly detection performance. Extensive experiments conduct on five real-world time series datasets, and the results show an improvement of approximately 1.2% in terms of the F_1 score compared to state-of-the-art algorithms, validating the effectiveness of learning context dependency in a differentiated manner for improving the anomaly detection performance of the model. Furthermore, sensitivity analysis of parameters and ablation experiments validate the stability and effectiveness of the proposed model.

Keywords Time series anomaly detection, Differentiated representation learning, Context dependency, Hypersphere

到稿日期:2024-06-03 返修日期:2024-09-01

基金项目:四川省重点研发计划(2023YFG0112, 2023YFG0115);国家自然科学基金(12401682, 62072320);四川省重点实验室开放课题(SCIT-LAB-20001);四川大学宜宾市合作项目(2020CDYB-30);四川大学自贡市合作项目(2022CDZG-6)

This work was supported by the Sichuan Science and Technology Program(2023YFG0112, 2023YFG0115), National Natural Science Foundation of China(12401682, 62072320), Intelligent Terminal Key Laboratory of Sichuan Province(SCITLAB-20001), Cooperative Program of Sichuan University, Yibin(2020CDYB-30) and Cooperative Program of Sichuan University, Zigong(2022CDZG-6).

通信作者:黄飞虎(hd808080@126.com)

1 引言

时间序列异常检测(Time Series Anomaly Detection, TS-AD)旨在检测时间序列中与正常数据不符的数据点或片段^[1],在许多领域中都有应用,如工业、航空航天、互联网等。准确地识别和定位时序数据中的异常,对于保障系统安全、提高运行效率至关重要。然而,实际应用中产生的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^[2]。

时间序列数据通常在时间维度上展现出上下文依赖关系,即过去和现在的观察值与未来的观察值之间存在相关性,并且这种关系是潜在、动态的^[3],充分建模这种上下文依赖关系对实现准确的异常检测至关重要^[4]。传统方法^[5-8]通常只考虑单个时刻的数据,而忽略了上下文信息,在时序异常检测任务上表现不佳。不少深度方法基于循环神经网络(Recurrent Neural Network, RNN)迭代地输入数据^[9-14],这些方法尽管融合了过去所有时刻的上下文信息,但缺乏时刻级别的依赖关系表示。文献^[15-16]计算自注意力权重矩阵,用于表示时刻级别的上下文依赖关系,并将该关系作为异常检测依据的特征。文献^[15]认为异常时刻关注局部信息,对应的上下文依赖关系类似于单峰明显的高斯分布。但是,当异常点的上下文与高斯分布假设不符时,该方法的有效性可能会受到影响^[17]。文献^[16]采用对时间序列划分 patch 的思想,并构建了一种 patch 内与 patch 间的双视图注意力对比学习方法,该方法的检测性能对 patch 的划分方式比较敏感。

基于注意力机制的异常检测方法^[15-16]通过强调上下文依赖关系在正常和异常样本之间的差异性,取得了更好的检测性能。受其启发,本文也利用注意力机制来学习这种差异性的表示。进一步地,现有方法大多遵循无监督学习的范式,没有充分利用异常样本的信息,并且由于神经网络具有较强的泛化性,它们学习到的正、异常数据之间的差异性被掩盖,难以得到有效区分正、异常样本的边界^[18]。因此,如果要充分利用上下文依赖关系在样本间的差异,就需要考虑如何实现正、异常表示间的差异化,在保留有效信息的前提下增加正、异常样本之间的区分度,以提高异常检测的准确性。

为解决上述问题,一个自然的想法是构造辅助的异常数据集,通过引入具有明确异常标签的数据样本来帮助模型学习区分正常和异常数据的特征。然而真实的异常样本非常稀少,因此本文通过人工合成的方式,将异常样本注入数据集,进而把无监督学习转换为有监督学习。该思路类似于异常暴露^[19],尽管合成的异常与真实的异常有所区别,但对于学习差异化的表示是有用的^[20],4.5.3节中的实验也验证了该想法的有效性。

本文提出上下文依赖差异化的异常检测(DCDAD)模型,该模型学习时序数据的上下文依赖关系的聚类。具体而言,利用自注意力机制提取样本的上下文依赖关系作为用于聚类的特征表示,目的是在该特征表示空间中学习可区分正、异常的超球面;通过异常注入的方式对训练数据集进行扩充,并设计差异化表示学习的损失函数,学习包围正常样本的超球面,并且在聚集正常样本的同时,迫使异常样本远离超球中

心,使正、异常样本在特征表示空间中呈现出明显的分离趋势。

本文的主要贡献总结如下:

1)提出通过差异化表示学习的方式增强样本的上下文依赖关系挖掘,并在聚类过程中学习可区分正、异常的超球面,解决现有异常检测模型对于正、异常样本区分度不明显的问题;

2)提出通过异常注入的方式,解决缺少监督信号的问题,同时设计差异化目标函数,促进正、异常表示分离,使超球面成为更准确的划分边界;

3)在多个时序数据集上进行实验,实验结果表明,该模型在异常检测任务上取得了优异的性能,验证了该模型的有效性。

本文第2章介绍无监督异常检测领域中具有代表性的各类工作;第3章详细描述提出模型的实现过程;第4章展示和分析实验结果;最后总结全文并展望未来。

2 相关工作

时间序列异常检测是当前数据挖掘研究领域备受关注的重要现实问题,在广泛的研究中得到了深入探索。由于数据集集中缺乏有标记的异常点,目前时序异常检测算法多为无监督算法,按照异常判断标准,现有方法可以分为基于密度估计、聚类、预测以及重构的方法,最近也有基于对比学习实现异常检测的研究。

1) 基于密度估计的方法

基于密度估计的方法通过估计数据的密度分布来识别异常样本。LOF^[5]是一种非参数化方法,通过考虑数据点与其邻近点的相对密度来确定异常值。LOF 主要关注样本在局部区域的离群程度,检测具有局部异常行为的样本;参数化方法 DAGMM^[21] 和 MPPCAD^[22] 则分别通过自编码器(Autoencoder, AE)、概率主成分分析(Probabilistic Principal Component Analysis, PPCA)对数据进行降维,然后用混合高斯模型(Gaussian Mixture Model, GMM)拟合数据分布,从而计算样本的概率密度。上述方法没有对时序依赖性进行建模,忽略了时序信息,导致对时序数据中的异常识别能力欠佳。

2) 基于聚类的方法

基于聚类的方法主要通过计算数据点之间的相似性或距离来判断异常样本。iForest^[6]构建随机分割树对数据逐步分割,基于样本在树中的深度测量其孤立程度,从而识别出异常;OC-SVM^[7], Deep-SVDD^[8]在高维特征空间寻找最优的超平面或超球体来包围正常样本,其中 OC-SVM 基于支持向量机(Support Vector Machine, SVM)将正常样本数据映射到高维特征空间中,通过在特征空间中找到一个超平面,将正常样本与异常样本分开;Deep-SVDD 也是基于 SVM 的方法,但与 OC-SVM 不同的是,它采用了深度神经网络来学习数据的紧致表示。Deep-SVDD 通过将正常样本映射到一个低维的紧致球形区域,对模型参数的优化通过最小化样本到球心的平均距离来实现,最终得到将异常样本与该区域分开的超球面。iForest, OC-SVM, Deep-SVDD 存在的缺陷与 LOF 等方法相似,即未对时序依赖性进行建模。THOC^[14]受到 Deep-SVDD

的启发,利用具有跳跃连接的 RNN 提取时序数据中的多尺度时间特征,并通过层次聚类将各个尺度下的中间层特征融合在一起,从而捕获时序数据中的复杂特征,然后将正常样本的模式表示成多个超球体,经过端到端的训练后,通过评估识别出偏离超球体的离群异常点。

3) 基于预测的方法

基于预测的方法作为近年来的主流研究方向,建立模型来学习时间序列的正常模式,然后使用该模型来预测未来的数据点。VAR^[23]将每个时间序列变量的当前值与过去时间步长的值以及其他相关变量的值相关联,进行线性组合,从而构建一个多变量的自回归模型来进行预测,但是该方法假设变量间的关系是线性的,未考虑非线性关系;LSTM^[9],GRU^[24]是常用的循环神经网络(RNN)的变体,能够学习和建模时间序列变化的非线性关系,也具有捕捉序列数据中长期依赖关系的能力,在基于预测^[3,10,17]和重构^[9,12-13,25-26]的时序异常检测方法中得到了广泛应用。

在用于异常检测任务时,基于预测的模型首先学习历史序列数据的变化模式和规律,然后根据当前观测值的预测与实际观测值之间的差异来识别异常^[27]。CL-MPPCA^[17]构建了一种混合概率主成分分析(Mixtures of Probabilistic Principal Component Analyzer, MPPCA)的多变量卷积 LSTM,以克服单通道 LSTM 模型的限制,它训练了一个多元时序异常检测模型,计算模型的预测与观测值的误差,并对误差应用平滑技术来减少尖峰。然而,预测方法依赖于模型的假设空间,当数据的非平稳性较强或者存在自相关误差时,难以保障模型的预测能力^[28],导致异常检测的准确性下降。

4) 基于重构的方法

基于重构的异常检测模型利用已知的正常数据进行训练,然后用于重构新的未知数据,其识别异常的前提在于:异常数据与正常数据之间存在较大差异,导致异常难以重构而产生较大重构误差。这类模型通常在正常样本上自监督地进行训练,常用框架包括自编码器 AE、生成对抗网络(Generative Adversarial Network, GAN)^[29],以及变分自编码器(Variational Autoencoder, VAE)^[30]。

BeatGAN^[11]首先利用动态时间规整(Dynamic Time Warping, DTW)技术对数据集进行增强扩充,然后在自编码器中添加 GAN 框架,用对抗的方法执行正则化,使模型具有更好的重构能力,最终将重构结果中偏离原始序列的部分解释为异常。MAD-GAN^[31]在 GAN 框架下构建了基于 LSTM 的时序异常检测模型,利用 LSTM 捕捉时序数据中的时序相关性,利用 GAN 训练的生成器和鉴别器来检测基于重构和鉴别损失的异常。

文献^[9,12-13]利用了 VAE 或改进的变体与 RNN 相结合:LSTM-VAE^[9]利用 VAE 得到局部窗口的嵌入,然后将 LSTM 作用于该嵌入,以对长期趋势进行建模,这种分层结构允许模型检测短期和长期的异常;OmniAnomaly^[12]是一种结合了时序依赖性和随机性的方法,它利用 GRU 和 VAE 分别对时序数据中的时序依赖性和随机性进行建模。该方法还使用线性高斯状态空间模型来捕捉随机变量之间的时间相关性。接着,通过使用潜在表示对输入进行重构,最终能够将与

正常模式差异较大的观测结果识别为异常。InterFusion^[13]利用 GRU 和卷积层分别捕捉多维时序数据中的时序依赖、属性间依赖信息,形成两个随机隐变量,并在分层 VAE 中对数据的正常模式进行建模,在测试阶段通过重构概率计算异常分数。该方法还在模型中引入了预过滤的模块,以减少在训练数据集时可能存在的异常过拟合的情况,并且提出使用马尔可夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC)填补方法对待检测的数据进行“纠正”。上述方法大多基于 RNN 对时序依赖进行建模,尽管在每个时刻的隐变量中融合了过去所有时刻的上下文信息,但是难以从中观测到各个时刻之间真实的依赖关系,即缺乏时刻级别的依赖关系表示;而且重构误差由逐点计算得到,这种误差无法对上下文信息进行全面描述^[15]。

文献^[15,32-33]利用 Transformer 更好的长期依赖关系建模和并行计算的能力,构建了具有良好重构能力的模型,并取得了不错的效果。Anomaly Transformer^[15]利用自注意力权重来表征时序维度的上下文依赖关系(也被称为序列关联),并将其用于新的异常检测指标计算:在重构误差的基础上融合了序列和高斯先验之间的关联差异作为异常分数。该模型基于一个前提:当异常发生时,当前点的周围点极可能也发生了异常,其序列关联更加关注局部信息,因此采用了具有单峰特性的正态分布作为先验。但是当实际情况不满足该先验时,模型的效果会受到影响,比如仅单个时间点脱离正常模式(点异常)。

5) 基于对比学习的方法

对比学习也是一种自监督学习方法,旨在通过学习样本之间的相似性和差异性来学习有用的表示。DCdetector^[16]进一步扩展 Anomaly Transformer 的工作,采用了对比学习和对时间序列划分 patch 的思想^[34]:从 patch 内和 patch 间这两个视图构造出正样本对,然后用对比损失进行训练,测试时则将两视图间的序列关联差异作为异常分数。该方法的检测性能对 patch 的划分方式比较敏感。

总结来看,针对时序数据,现有方法通常考虑时间维度的上下文依赖关系,并融合该信息以提升模型的表示能力。过去的方法没有充分考虑数据中上下文依赖关系的差异性;而最新的两个方法 Anomaly Transformer^[15]和 DCdetector^[16]虽然考虑了差异性,但它们没有对异常样本进行建模来突出这种差异性。Transformer 中的自注意力机制提供了一种用于提取时序维度的上下文依赖关系的有效手段。因此,本文利用自注意力机制提取这一特征,通过差异化正、异常样本表示学习的方式来增强样本的上下文依赖关系挖掘,并在聚类过程中学习可区分正、异常的超球面,解决现有异常检测模型对正、异常样本区分度不明显的问题。

3 考虑上下文依赖差异化的异常检测

DCDAD 是一个基于自注意力学习的时序异常检测模型,整体框架如图 1 所示,主要包括异常注入、上下文依赖表示学习、差异化表示学习 3 个模块。1)异常注入:在数据预处理阶段利用滑动窗口构造出训练集后,利用异常注入向训练数据集中加入额外的标记异常数据。2)上下文依赖表示学

习:模型采用编-解码器的架构,编码器对输入窗口进行编码,得到窗口内的上下文依赖关系以及编码后的潜在向量,解码器则根据潜在向量生成重构结果。3)差异化表示学习:计算

依赖关系的表示到超球中心的距离,构造用于差异化学习的目标函数,增大正、异常的可区分性,学习包围正常样本的超球面并使得超球面成为更准确的划分边界。

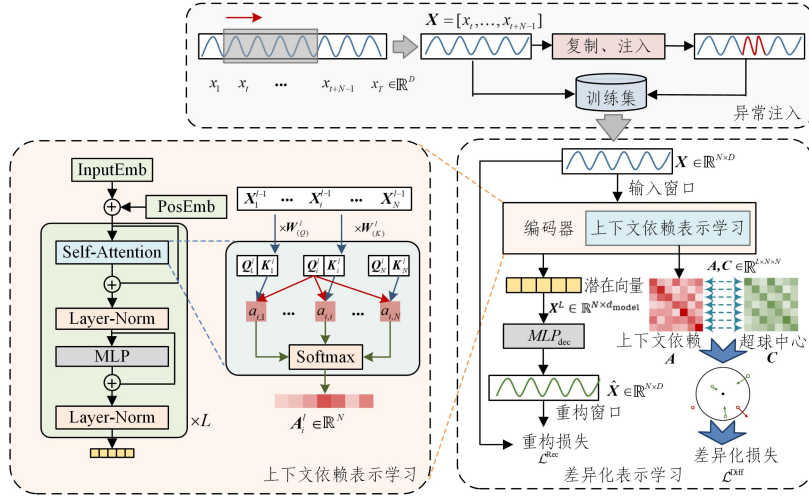


图1 DCDAD 整体框架

Fig. 1 Overall framework of DCDAD

3.1 问题描述

给定长度为 T 的多元时间序列数据 $X = [x_1, \dots, x_T] \in \mathbb{R}^{T \times D}$, D 表示数据的维度,每个时刻都具有标签,表示为 $Y = [y_1, \dots, y_T]$,其中标签为 0 表示正常,为 1 表示异常。

异常检测的目标是预测每个时刻 ($t=1, 2, \dots, T$) 的异常分数 $score = [score_1, \dots, score_T]$,与阈值 θ 比较来判定该序列的各时刻是否异常:

$$\hat{y}_t = \begin{cases} 1, & \text{if } score_t > \theta \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

为了充分利用时序数据的上下文依赖关系,以进行更好的特征表示,使用滑动窗口对 X 进行划分。滑动窗口长度为 N ,滑动步长为 S ,构成窗口数据集 $\mathcal{D} = \{(X^t, Y^t) | t=1, \dots, \frac{T-N}{S} + 1\}$,其中 $X^t \in \mathbb{R}^{N \times D}$, $Y^t \in \mathbb{R}^N$ 表示每个窗口的数据和标签(训练集和测试集分别为 \mathcal{D}_{train} , \mathcal{D}_{test})。标签是时间点(时刻)级别的,在本文中将交替使用时刻、时间点和样本这 3 个术语。

3.2 异常注入

由于异常数据的稀缺性,目前用于时序异常检测的基准数据集通常假定训练集只有正常数据,不包含异常标记,因此训练集中所有窗口的各时刻的标签均为 0。这种现状导致了现有异常检测模型大多在无监督的框架下进行构建,而没有对异常样本进行建模。为利用和挖掘异常样本中的信息,本文采用多种异常注入方式构造多样化的异常数据集,包括迁移注入、高斯注入、常数注入。

异常注入不需要将真实异常的模式完全覆盖^[18],只是提供一种启发式的手段来指导模型突出学习异常与正常样本之间的差异性。通过模拟多种类型的异常注入,允许有监督的学习,可以帮助模型学习到正常数据和异常数据的差异性,从而进一步提高模型对异常数据的识别准确性。4.5.3 节的实验结果也验证了异常注入的有效性。以一个异常窗口的构造为例,具体的实现方式如下:

给定一个来自于训练集的正常窗口 $(X^t, Y^t) \in \mathcal{D}_{train}$,索引为 t ,将 X^t 拷贝作为异常窗口的模板,即 $X^{Ano} = Copy(X^t)$,然后将窗口内的子序列用其他序列替换,以得到新的异常窗口。从该窗口中选定长度为 p 的子序列,满足 $p < N$,再从 3 种方式中选择其一,以完成异常片段的注入:1)迁移注入:用索引为 k 的另一个窗口 $X^k \in \mathcal{D}_{train}$ 内等长的片段替换 X^{Ano} 的片段;2)高斯注入:用高斯噪声序列替换 X^{Ano} 的片段,噪声的每个值均服从高斯分布;3)常数注入:从均匀分布中采样一个常数,将子序列每个时刻的值替换为该常数。

通过上述异常注入方式,构造额外的异常窗口,注入异常在单个属性上的结果示例如图 2 所示。

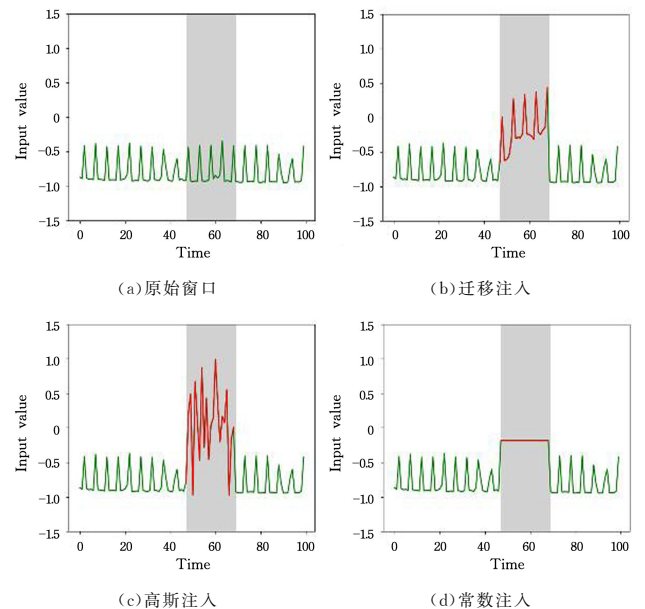


图2 异常注入效果示例

Fig. 2 Examples of anomaly injection

通过注入新的序列可以破坏原本窗口中时序数据的连续性,构成一种状态突变的异常类型,例如突发事件、异常操作、

缺失数据等。由于异常片段使窗口内形成了新的依赖关系,因此将窗口内的各时刻均标记为异常 $\mathbf{Y}^{\text{Ano}} = [1]_{j=1}^N$,而不是仅标记被替换的片段。

最后将额外异常窗口添加到训练集 $\mathcal{D}_{\text{train}}$ 中,即 $\mathcal{D}_{\text{train}} \leftarrow \mathcal{D}_{\text{train}} \cup \{(\mathbf{X}^{\text{Ano}}, \mathbf{Y}^{\text{Ano}})\}$ 。

3.3 上下文依赖表示学习

为了学习时序数据中的上下文依赖关系,利用自注意力机制,自适应地计算不同时刻之间的依赖。本文采用了编-解码器的架构,其中编码器负责对输入序列进行编码得到潜在向量,同时利用自注意力层提取上下文依赖关系,即自注意力权重矩阵;解码器则负责根据潜在向量生成重构结果。

编码器由嵌入层以及多个堆叠的编码层构成。对于一个输入窗口数据 \mathbf{X}' ,为方便表示,在下文中简称为 \mathbf{X} ,嵌入层将窗口的原始输入进行编码,提取特征信息,并加上位置编码以保留序列中的顺序信息:

$$\mathbf{X}^{\text{emb}} = \text{PosEmb}(\mathbf{X}) + \text{InputEmb}(\mathbf{X}) \quad (2)$$

其中, $\text{PosEmb}(\cdot)$ 和 $\text{InputEmb}(\cdot)$ 分别表示位置嵌入和输入特征嵌入。 $\mathbf{X} \in \mathbb{R}^{N \times D}$ 经过嵌入得到表示 $\mathbf{X}^{\text{emb}} \in \mathbb{R}^{N \times d_{\text{model}}}$,并作为第一个编码层 ($l=1$) 的输入, d_{model} 表示隐藏层大小。

编码层有 L 层,每一层结构相同,上一层的输出即为该层的输入,其计算过程如下:

$$\mathbf{Z}^l = \text{Layer-Norm}(\mathbf{X}^{l-1} + \text{Attention}(\mathbf{X}^{l-1})) \quad (3)$$

$$\mathbf{X}^l = \text{Layer-Norm}(\mathbf{Z}^l + \text{MLP}(\mathbf{Z}^l)) \quad (4)$$

其中, $\mathbf{X}^l \in \mathbb{R}^{N \times d_{\text{model}}}$, $l \in \{1, 2, \dots, L\}$ 表示第 l 编码层的输出表示, $\mathbf{Z}^l \in \mathbb{R}^{N \times d_{\text{model}}}$ 则是第 l 编码层的中间表示, MLP 和 Layer-Norm 分别表示多层感知机以及层归一化计算, $\text{Attention}(\cdot)$ 用于计算在时间维度上进行自注意力操作后输出的表示,该模块允许在窗口间任意两个时刻间建立关联^[15],具体表示为权重矩阵。本文所关注的上下文依赖表示即为该矩阵,其计算过程如下:

$$\mathbf{Q}_i^l = \mathbf{X}_i^{l-1} \mathbf{W}_Q^l, \mathbf{K}_i^l = \mathbf{X}_i^{l-1} \mathbf{W}_K^l \quad (5)$$

$$\mathbf{K}^l = [\mathbf{K}_1, \mathbf{K}_2, \dots, \mathbf{K}_N] \in \mathbb{R}^{N \times d_{\text{model}}} \quad (6)$$

$$\mathbf{A}_i^l = \text{Softmax}\left(\frac{\mathbf{Q}_i^l (\mathbf{K}^l)^T}{\sqrt{d_{\text{model}}}}\right) \quad (7)$$

其中,对于第 l 个自注意力层, $\mathbf{Q}_i^l, \mathbf{K}_i^l \in \mathbb{R}^{1 \times d_{\text{model}}}$ 分别表示第 i 时刻的 query 和 key 向量, $\mathbf{A}_i^l \in \mathbb{R}^N$ 表示第 i 时刻关于整个窗口中各时刻的依赖权重,通过计算 i 时刻的 query 向量与各个时刻的 key 向量的相似度并经由 $\text{Softmax}(\cdot)$ 对相似度进行归一化得到。

本文将各编码层通过式(7)计算的上下文依赖关系记为 $\mathbf{A}_i = [\mathbf{A}_i^1, \dots, \mathbf{A}_i^L] \in \mathbb{R}^{L \times N}$,表示窗口中第 i ($i=1, 2, \dots, N$) 时刻的依赖关系,并将其嵌入到下一节提出的目标函数中,用于改进表示学习。

解码器是一个多层感知机,将编码得到的潜在变量作为输入,经过解码器得到最终的输出。

$$\hat{\mathbf{X}} = \text{MLP}_{\text{dec}}(\mathbf{X}^L) \quad (8)$$

$\hat{\mathbf{X}} \in \mathbb{R}^{N \times D}$ 是模型对输入窗口 \mathbf{X} 的重构输出,它试图在尽可能保留原始数据特征的同时,通过模型的学习来捕捉数据中的模式和结构。

3.4 差异化表示学习

3.4.1 不含差异化的学习目标

正常点通常具有相似的特征和模式,而异常则可能与之存在显著差异。基于该假设,通过在表示空间中找到一个最优的超球面,将正常点包围在内,从而识别出不在超球面内的异常。与 Deep-SVDD 中的基本思想类似,Deep-SVDD 在无监督的情况下训练神经网络,以最小化特征表示到超球中心 \mathbf{c} 的距离 $\sum_{\mathbf{x} \in \mathcal{X}} \|\phi(\mathbf{X}; \boldsymbol{\omega}) - \mathbf{c}\|^2$,实现最小化划分超球面,超球中心可以被视为这些正常点的共享潜在模式或聚类中心。

对于一个包含 N 个连续时间点的输入窗口数据 $\mathbf{X} \in \mathbb{R}^{N \times D}$,受文献[15]启发以及通过 4.5.3 节的实验验证,异常样本通常具有与正常样本不同的依赖关系,自注意力权重可以捕捉到这种依赖关系,从而帮助模型区分正常和异常样本。

文献[15]假定异常时刻具有更关注于邻近时刻的局部效应并将其作为识别异常的依据,这种假定的先验性很强,在具有复杂异常模式的现实数据中不一定适用(例如点异常就不具备局部效应,它和邻近的点没有强相关性)。而本文提出将依赖关系用于对正常样本进行聚类,自适应地学习正常样本的共享依赖关系,再将不满足该共享依赖关系的样本视为异常,从而避免了上述先验性的问题。

用包裹所有正常样本的超球面来表征正常样本的共享依赖关系,在不考虑训练集中异常点的情况下,为引导模型将正常样本的表示映射到体积最小的超球面内,本文首先构造以下目标函数:

$$\mathcal{L}^{\text{Hyper}} = \|\text{Dis}(\mathbf{A}, \mathbf{C}; \mathbf{X})\|_1 \quad (9)$$

其中, $\mathbf{C} \in \mathbb{R}^{L \times N \times N}$ 是一个超球中心集合, $\text{Dis}(\mathbf{A}, \mathbf{C}; \mathbf{X}) \in \mathbb{R}^N$ 表示窗口中各时刻点 i 的依赖关系 \mathbf{A} 到超球中心 \mathbf{C} 的径向距离,即 $[\text{Dis}(\mathbf{A}, \mathbf{C}; \mathbf{X})]_{i=1}^N$ 。

式(9)最小化各正常样本到超球中心的距离,从而完成正常样本的聚类。同时,包围正常样本的超球面作为分割正、异常样本的划分边界,将会随着优化的迭代而逐渐变小。但是该目标函数未对异常样本进行建模,因此将在 3.4.2 节中联合式(9)以及异常注入,完成构建差异化学习的目标函数。

3.4.2 上下文依赖关系的差异化学习

本节将进一步阐述式(9)中依赖关系与超球中心距离的计算方式,以及差异化学习的实现。

1) 依赖关系的距离度量

对于超球中心集合 $\mathbf{C} \in \mathbb{R}^{L \times N \times N}$,其中 $\mathbf{C}_i \in \mathbb{R}^{L \times N}$ 表示第 i 个超球中心,与第 i 时刻的依赖关系 \mathbf{A}_i 的维度相同, i 是矩阵 \mathbf{C} 和矩阵 \mathbf{A} 的第二维度上的索引。索引 i 衡量了不同时刻在窗口内的相对位置,不同相对位置的依赖特征存在差异,为了保留这种差异性,本文设计了超球中心集合 \mathbf{C} ,窗口内的各个相对位置的依赖关系 \mathbf{A}_i 分别对应每个超球中心 \mathbf{C}_i 。

依赖关系 \mathbf{A}_i 本质上是一个时刻与所有时刻之间的注意力权重分布。为了计算依赖关系到超球中心的径向距离,本文选择用基于 Kullback-Leibler(KL)散度的距离度量来衡量两者的总体差异:

$$\text{Dis}(\mathbf{A}_i, \mathbf{C}_i; \mathbf{X}) = \sum_{i=1}^L \left[\frac{1}{2} \text{KL}(\mathbf{A}_i^l \parallel \mathbf{C}_i^l) + \frac{1}{2} \text{KL}(\mathbf{C}_i^l \parallel \mathbf{A}_i^l) \right] \quad (10)$$

其中, $Dis(\mathbf{A}_i, \mathbf{C}_i; \mathbf{X}) \in \mathbb{R}^l$ 表示各编码层的依赖特征 \mathbf{A}_i^l 到超球中心 $\mathbf{C}_i^l \in \mathbb{R}^N$ 的径向距离之和。

对于输入窗口 $(\mathbf{X}, \mathbf{Y}) \in \mathcal{D}_{\text{train}}$, 考虑其标签 \mathbf{Y} , 则窗口中各时刻的距离计算式如下所示:

$$\begin{aligned} Dis_N(\mathbf{A}, \mathbf{C}; \mathbf{X}, \mathbf{Y}) &= [Dis(\mathbf{A}_i, \mathbf{C}_i; \mathbf{X}) \cdot (1 - y_i)]_{i=1, \dots, N} \\ Dis_A(\mathbf{A}, \mathbf{C}; \mathbf{X}, \mathbf{Y}) &= [Dis(\mathbf{A}_i, \mathbf{C}_i; \mathbf{X}) \cdot y_i]_{i=1, \dots, N} \end{aligned} \quad (11)$$

其中, $Dis_N(\cdot), Dis_A(\cdot)$ 分别表示窗口中正、异常样本的径向距离, y_i 表示第 i 时刻的异常标签。

2) 差异化学习

为了实现最小化划分超球面以及对正、异常表示进行差异化, 本文基于目标函数式(11)进一步提出了以下目标:

$$\begin{aligned} \mathcal{L}^{\text{Diff}}(\mathbf{A}, \mathbf{C}, \lambda; \mathbf{X}, \mathbf{Y}) &= (1 - \lambda) \cdot \|Dis_N(\mathbf{A}, \mathbf{C}; \mathbf{X}, \mathbf{Y})\|_1 - \\ &\quad \lambda \cdot \min(\|Dis_A(\mathbf{A}, \mathbf{C}; \mathbf{X}, \mathbf{Y})\|_1, \\ &\quad \gamma \cdot N) \end{aligned} \quad (12)$$

其中, $\lambda \in [0, 1]$ 分别表示正、异常时刻的损失权重; γ 是一个超参数; N 是窗口长度, 用于将异常点径向距离之和的范围裁剪到 $[0, \gamma \cdot N]$ 内。

对模型参数进行优化时, 最小化式(12)可以对正常样本表示进行聚类, 使其向超球中心靠近, 得到正常的共享特征表示, 同时还将异常表示排除在超球面外, 实现正、异常表示的差异化。

此外, 通过 \min 函数实现裁剪操作, 可以截断超过范围的异常样本距离的梯度, 防止异常样本的距离在优化过程中无限制的增大。如果不对其进行截断, 异常点的上下文依赖关系表示将可能无限远离超球中心, 导致自注意力学习崩溃或者超球中心的偏移, 得到无效表示。

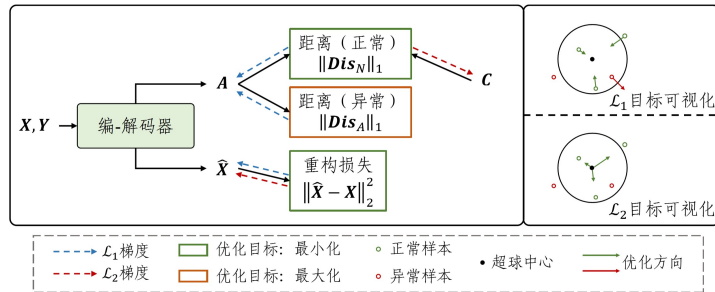


图3 优化策略可视化

Fig. 3 Visualization of optimization strategy

3.5.2 异常检测

在异常检测(推理)阶段, 本文将超球中心视为正常样本的共享依赖关系, 通过计算待检测样本与超球中心的距离, 将远离超球中心(超球面外)的离群点判定为异常。测试数据中, 对滑动窗口数据 $\mathbf{X} \in \mathbb{R}^{N \times D}$, 计算其各个时刻的异常分数, 计算方法如下:

$$\text{score}(\mathbf{X}) = \text{Softmax}([Dis(\mathbf{A}_i, \mathbf{C}_i; \mathbf{X})]_{i=1, \dots, N}) \quad (15)$$

其中, $\text{score}(\mathbf{X}) \in \mathbb{R}^N$ 是时间点级别的异常分数, 该分数仅使用了各点的上下文依赖关系关于超球中心的距离作为异常衡量指标。算法的完整训练及测试流程框架如算法1所示。

算法1 DCDAD 训练、测试过程

输入: 时间序列 \mathcal{X} (包括训练数据和含标签的测试数据)

输出: 异常分数 score

3.5 模型优化与异常检测

3.5.1 优化策略

为了使模型能够学习到输入序列窗口 $\mathbf{X} \in \mathbb{R}^{N \times D}$ 的内部结构和特征, 本文通过基于重构的自监督学习方式, 构造了额外的目标函数:

$$\mathcal{L}^{\text{Rec}}(\hat{\mathbf{X}}; \mathbf{X}) = \|\hat{\mathbf{X}} - \mathbf{X}\|_2^2 \quad (13)$$

其中, $\hat{\mathbf{X}}$ 表示模型对输入窗口 \mathbf{X} 的重构输出结果。该自监督学习的目标在于最小化重构损失, 从而辅助模型捕捉序列中的长距离依赖性, 并更好地理解上下文信息。

除了模型的参数权重, 超球中心 \mathbf{C} 也是一个需要优化的参数, 可通过梯度下降进行更新。为了联合训练模型参数和超球中心参数, 本文利用 Stop-Gradient 操作使得模型参数和超球在训练过程中相互独立地更新。具体来说, 基于式(12)和式(13)构造以下两个目标损失, 以执行梯度计算和参数更新:

$$\mathcal{L}_1 = \mathcal{L}^{\text{Rec}}(\hat{\mathbf{X}}; \mathbf{X}) + \mathcal{L}^{\text{Diff}}(\mathbf{A}, \mathbf{C}_{\text{detach}}, \lambda; \mathbf{X}, \mathbf{Y}) \quad (14)$$

$$\mathcal{L}_2 = \mathcal{L}^{\text{Rec}}(\hat{\mathbf{X}}; \mathbf{X}) + \|Dis_N(\mathbf{A}_{\text{detach}}, \mathbf{C}; \mathbf{X}, \mathbf{Y})\|_1$$

其中, 下标 detach 表示将阻断对应变量的梯度反向传播, \mathcal{L}_1 和 \mathcal{L}_2 主要分别负责依赖关系的学习以及超球中心参数的调整。图3(左)展示了所提优化策略对应的梯度反向传播路径, 图3(右)则对目标函数的作用进行可视化。总而言之, \mathcal{L}_1 通过差异化学习的目标 $\mathcal{L}^{\text{Diff}}$ (包含 Dis_N, Dis_A) 对编码器部分进行更新, 使其学习到正常样本的共享表示并将异常样本的表示推离超球中心; \mathcal{L}_2 则通过正常点的径向距离 Dis_N 迭代更新超球中心的参数, 逐步缩小它到所有正常样本的距离, 从而得到最优的中心。

1. 经过数据预处理以及划分窗口得到训练集 $\mathcal{D}_{\text{train}}$ (设标签为 0) 以及测试集 $\mathcal{D}_{\text{test}}$
2. 通过异常注入构造额外的异常窗口(标签为 1), 将其加入训练集中
3. 从训练集中取出训练窗口 $(\mathbf{X}, \mathbf{Y}) \in \mathcal{D}_{\text{train}}$
4. 将训练窗口数据 \mathbf{X} 输入模型, 得到重构结果 $\hat{\mathbf{X}}$ 以及上下文依赖关系 \mathbf{A}
5. 根据式(14)更新模型参数(包括超球中心 \mathbf{C})
6. 输入测试窗口数据(推理阶段), 根据式(15)计算异常分数

4 实验

4.1 数据集

本节在 5 个公开的真实时间序列数据集上进行实验, 验证了 DCDAD 模型的异常检测性能。

1) 服务器机器数据集 (Server Machine Dataset, SMD)^[12]: 是从一家大型互联网公司收集的, 包含来自 28 台服务器和 38 个传感器的 5 周的数据。前 5 天仅包含正常数据, 最后 5 天间歇性注入异常数据。

2) 池化服务器指标 (Pooled Server Metrics, PSM)^[35]: 是从 eBay 的多个应用程序服务器节点内部收集的。有 13 周的训练数据和 8 周的测试数据。

3) 火星科学实验室 (Mars Science Laboratory, MSL) 和土壤湿度主、被动 (Soil Moisture Active and Passive, SMAP)^[36-37] 是 NASA 收集的公共数据集, 包含来自航天器监测事件意外异常 (ISA) 报告的遥测异常数据系统。数据集分别有 55 和 25 维。

4) 安全水处理 (Secure Water Treatment, SWaT)^[38]: 是在 11 天内从具有 51 个传感器的小型水处理测试台收集的。在后 4 天内, 使用不同的攻击方法注入了多个异常, 而在前 7 天内仅生成正常数据。

数据集的统计信息如表 1 所列, 训练 #、测试 # 对应于训练集、测试集中的时间点总数, 测试集中包含有标记的异常点, 异常率是计算出的测试集中标记异常点数量关于所有时间点数量的比率。

表 1 数据集信息 (预处理前)

Table 1 Information of the datasets (before preprocessing)

数据集	应用	维度	训练 #	测试 #	异常率 / %
SMD	服务器	38	708 405	708 420	4.16
PSM	服务器	25	132 481	87 841	27.76
MSL	空间	55	58 317	73 729	10.48
SMAP	空间	25	135 183	427 617	12.79
SWaT	水	51	495 000	449 919	12.14

4.2 实验设置

DCDAD 在不同数据集上的实验参数设定如表 2 所列。对于训练数据集, 设置不同的窗口大小 N 和滑动步长 S 。异常注入的窗口数量与正常窗口数量相同, 每个窗口内, 随机选择迁移注入、高斯注入、常数注入这 3 种注入方式之一。注入时被替换的片段长度 p 满足 $(p/N) \sim U(0.1, 0.4)$, 高斯注入的均值和方差分别为 0 和 0.5, 常数注入的常数从 $[-1.5, 1.5]$ 中均匀随机采样产生。目标损失函数中对正、异常的损失权重分别表示为 λ_1 和 λ_2 , 训练轮次为 Epoch, 截断目标函数中用于裁剪的超参数 γ 设为 1。

表 2 各数据集训练实验参数

Table 2 Experimental parameters for training on each dataset

数据集	Epoch	Batchsize	N	S	λ
SMD	10	512	300	20	0.3
PSM	3	512	100	10	0.3
MSL	3	512	100	10	0.5
SMAP	3	512	200	20	0.5
SWaT	3	512	200	10	0.4

测试阶段, 窗口的滑动步长 S 设置为等于窗口大小 N , 即滑动窗口彼此不重叠, 从而对每个时刻计算出一个异常分数。异常分数的阈值 θ 是使得验证集中比例 $r_0\%$ 的点被划分为异常的值, r_0 的值在 5 个数据集上分别设置为 0.25, 0.6, 0.8, 0.5, 0.35, 此外, 本文设计的验证集仅被用于评估异常分数分布从而确定异常分数阈值, 由 20% 的原始训练集数据以及等量的注入异常数据共同构成。模型通过 PyTorch 框架

实现, 所有实验均在单个 NVIDIA GeForce RTX 3090 24 GB 显卡上运行。

DCDAD 的编码器包括 3 层, 注意力头数量为 1, 表 3 列出了模型各部分的具体参数, 其中 D 表示输入数据的维度大小。

表 3 DCDAD 模型架构及相应参数信息

Table 3 DCDAD model architecture and corresponding parameter information

		Conv1D(C, 256, (kernel_size: 3, stride: 1, padding: 1))	
3 * 编码层	InputEmb		
	Attention	$W_{(Q)}$	FC(256, 256)
		$W_{(K)}$	FC(256, 256)
		$W_{(V)}$	FC(256, 256)
		$W_{(O)}$	FC(256, 256)
Layer-Norm	LayerNorm(256, $\epsilon_{ps}=1 \times 10^{-5}$)		
MLP	FC(256, 128), gelu		
Layer-Norm	FC(128, 256)		
Layer-Norm	LayerNorm(256, $\epsilon_{ps}=1 \times 10^{-5}$)		
MLP _{dec}	FC(256, D)		

4.3 评价指标

为了评估算法的性能, 采用广泛使用的点调整策略^[14, 16, 39]: 如果检测到某个连续时间的异常段中的任意时间点, 就将该异常时间段中的所有异常视为被正确检测到。从观察到异常时间点会引起警报并进一步使整个分段在实际应用中引起注意的观察来看, 该策略是合理的。本文主要关注点调整后的 F_1 分数, 计算式如下:

根据真实标签 $\mathbf{Y} = [y_1, \dots, y_T]$ 以及模型推断的异常分数 $\mathbf{score} = [score_1, \dots, score_T]$, 首先给定分数阈值 θ , 计算预测标签 $\hat{\mathbf{Y}} = [\hat{y}_1, \dots, \hat{y}_T]$, 其中 $\hat{y}_t = 1_{score_t > \theta}$ 。进行点调整策略后, 再利用预测标签 $\hat{\mathbf{Y}}$ 分别计算真正例 TP 、假正例 FP 以及假负例 FN :

$$\begin{aligned} TP &= \{t | \hat{y}_t = 1, y_t = 1\} \\ FP &= \{t | \hat{y}_t = 1, y_t = 0\} \\ FN &= \{t | \hat{y}_t = 0, y_t = 1\} \end{aligned} \quad (16)$$

随后, 分别计算精确率 P 以及召回率 R , 再计算 F_1 分数:

$$\begin{aligned} P &= \frac{n(TP)}{n(TP) + n(FP)} \\ R &= \frac{n(TP)}{n(TP) + n(FN)} \\ F1 &= \frac{2PR}{P + R} \end{aligned} \quad (17)$$

4.4 对比基线

为对比不同模型的性能, 本文采用的基线方法有: 基于对比学习的模型 DCdetector^[16] (2023); 基于重构的模型 Anomaly Transformer^[15] (2022), InterFusion^[13] (2021), OmniAnomaly^[12] (2019), LSTM-VAE^[19] (2018); 基于密度估计的方法 DAGMM^[21] (2018), LOF^[5] (2000); 基于聚类的方法 THOC^[14] (2020), Deep-SVDD^[8] (2018), iForest^[6] (2008), OC-SVM^[7] (2004); 基于预测的模型 CL-MPPCA^[17] (2019), LSTM^[3] (2018)。

4.5 实验结果与分析

4.5.1 异常检测实验

本文在5个真实数据集上与13个基线模型进行了比较,结果如表4所列。其中加粗、下划线分别表示最优、次优的结果。从 F_1 分数来看,DCDAD模型在大部分数据集上都取得

了最佳的检测效果,在MSL数据集上也取得了接近基线的最优结果。将时间维度的上下文依赖关系作为区分异常发生的重要参考,在这个表示空间里,DCDAD通过差异化的目标函数扩大了正正常样本表示的差异性,并优化了划分超球面的学习,显著地提升了异常检测的准确性。

表4 在5个数据集上的实验结果
Table 4 Experiment results on five datasets

Dataset	SMD			PSM			MSL			SMAP			SWaT		
	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1	P	R	F_1
OC-SVM	44.34	76.72	56.19	62.75	80.89	70.67	59.78	86.87	70.82	53.85	59.07	56.34	45.39	49.22	47.23
iForest	42.31	73.29	53.64	76.09	92.45	83.48	53.94	86.54	66.45	52.39	59.07	55.53	49.29	44.95	47.02
LOF	56.34	39.86	46.68	57.89	90.49	70.61	47.72	85.25	61.18	58.93	56.33	57.60	72.15	65.43	68.62
Deep-SVDD	78.54	79.67	79.10	95.41	86.49	90.73	91.92	76.63	83.58	89.93	56.02	69.04	80.42	84.45	82.39
DAGMM	67.30	49.89	57.30	93.49	70.03	80.08	89.60	63.93	74.62	86.45	56.73	68.51	89.92	57.84	70.40
LSTM	78.55	85.28	81.78	76.93	89.64	82.80	85.45	82.50	83.95	89.41	78.13	83.39	86.15	83.27	84.69
CL-MPPCA	82.36	76.07	79.09	56.02	99.93	71.80	73.71	88.54	80.44	86.13	63.16	72.88	76.78	81.50	79.07
LSTM-VAE	75.76	90.08	82.30	73.62	89.92	80.96	85.49	79.94	82.62	92.20	67.75	78.10	76.00	89.50	82.20
OmniAnomaly	83.68	86.82	85.22	88.39	74.46	80.83	89.02	86.37	87.67	92.49	81.99	86.92	81.42	84.30	82.83
InterFusion	87.02	85.43	86.22	83.61	83.45	83.52	81.28	92.70	86.62	89.77	88.52	89.14	80.59	85.58	83.01
THOC	79.76	90.95	84.99	88.14	90.99	89.54	88.45	90.97	89.69	92.06	89.34	90.68	83.94	86.36	85.13
AnomalyTrans	<u>89.40</u>	95.45	92.33	96.91	98.90	97.89	92.09	95.15	93.59	94.13	99.40	94.07	91.55	96.73	94.07
DCdetector	83.59	91.10	87.18	<u>97.14</u>	<u>98.74</u>	<u>97.94</u>	93.69	99.69	96.60	95.63	<u>98.92</u>	<u>97.02</u>	<u>95.72</u>	98.93	97.30
DCDAD	94.91	<u>93.80</u>	94.35	98.60	98.30	98.45	<u>93.02</u>	95.12	<u>94.06</u>	96.70	97.84	97.27	97.38	<u>98.44</u>	97.90

注:加粗、下划线分别表示最优、次优的结果。

将模型利用时序上下文信息的方式作为划分依据,进一步对基线方法进行分类和对比。经典的通用异常检测方法(OC-SVM,iForest,LOF,Deep SVDD,DAGMM)未考虑时序上下文信息,因此在时序数据上的性能欠佳;使用循环网络建模时序上下文的相关性的方法包括基于聚类的模型(THOC)、基于预测的模型(LSTM,CL-MPPCA)、基于重构的模型(LSTM-VAE,OmniAnomaly,InterFusion),这些模型中上下文信息仅用于增强模型的表示学习能力,相比于通用的异常检测方法有一定优势,但在各指标上的表现仍不如最新的两项研究;最新研究(Anomaly Transformer,DCdetector)将时序上下文信息用于异常指标的计算,取得了更好的检测效果,但它们没有建模异常本来突出上下文依赖之间的差异性。

了不同的指标计算,并且探索了将时序上下文依赖的表示用于聚类学习。进一步地,通过差异化表示学习的方式来增强样本的上下文依赖关系挖掘,方便学习到更加准确的超球面作为划分边界。实验结果显示,在各个数据集上的整体性能优于这两个模型。综上,本文提出的基于上下文依赖关系的差异化学习框架在时序异常检测任务是有效的。

图4展示了在模型训练过程中所得正、异常样本的上下文依赖关系表示到超球中心的平均径向距离随着迭代次数的增加而变化的过程,可以看出,随着模型的不断优化,正常点到超球中心的距离逐渐缩小并趋于平稳,而异常点在差异化学习的作用下不断远离中心,直到其距离稳定在某个常数附近。该现象验证了本文提出的差异化的目标损失有利于提高正、异常表示之间的差异性。

相比于 Anomaly Transformer 和 DCdetector,本文设计

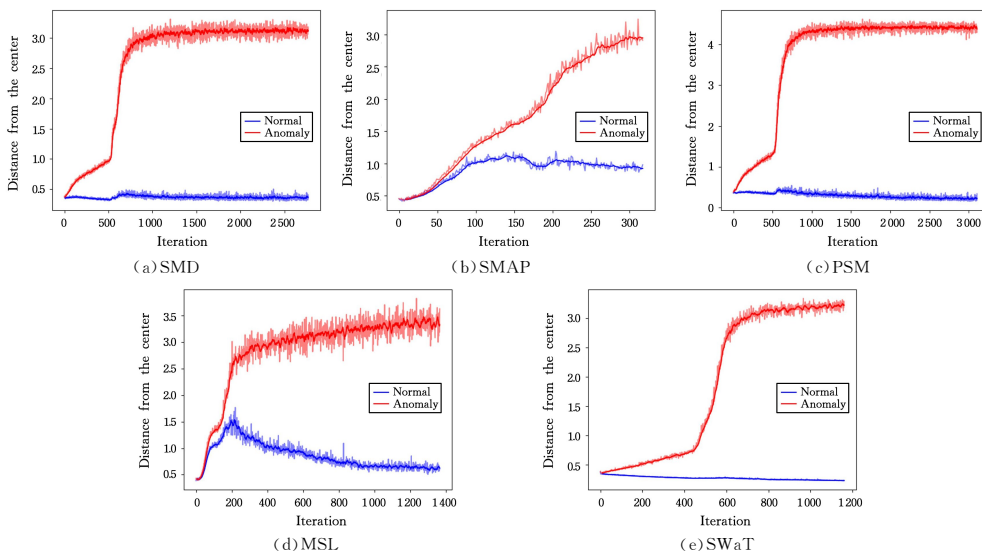


图4 训练过程中正、异常点关于超球中心的径向距离

Fig. 4 Radial distance of normal and anomalous points from the hypersphere center during training

4.5.2 参数敏感性分析

窗口长度 N 是影响上下文依赖关系跨度的重要参数,更大的 N 意味着更长期的依赖关系以及更大的计算复杂度;权重 λ 是影响上下文依赖关系差异化学习的参数,其大小衡量了优化异常样本远离超球中心的程度,相对应的 $(1-\lambda)$ 则表示对正常样本收紧的程度。在本节中,通过固定其中一个参数,然后对另一参数的不同取值进行实验,以验证该参数对模型性能的影响。对于 λ :在 $[0, 0.5]$, N 固定(参照表 2 中 N 的取值);对于 N :在 $[50, 300]$, λ 固定(参照表 2 中 λ 的取值)。在 5 个数据集上分别进行了 5 次重复实验,取结果的均值。图 5 为固定 N 时,模型在不同 λ 取值下的实验结果。图 6 为固定 λ 时,模型在不同 N 取值下的表现。

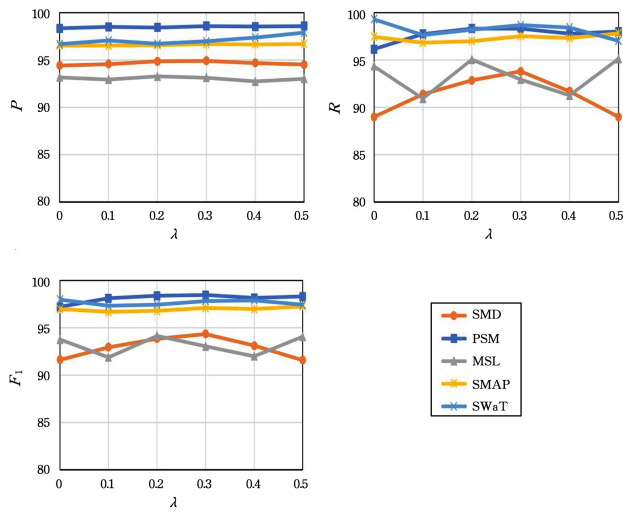


图 5 参数 λ 的敏感性实验

Fig. 5 Sensitivity experiment of λ

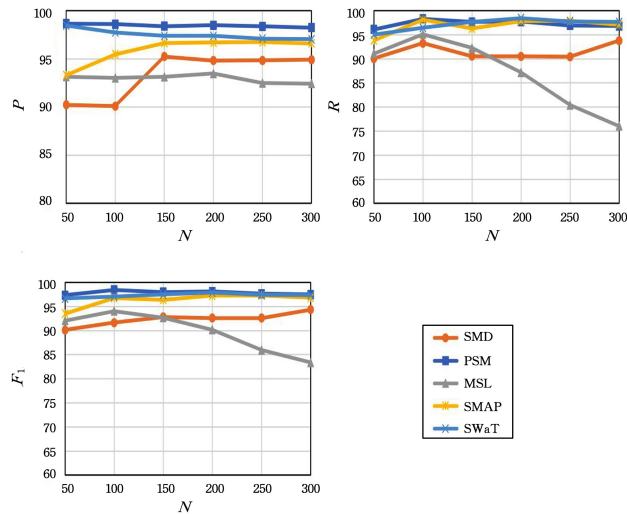


图 6 参数 N 敏感性实验

Fig. 6 Sensitivity analysis of N

从图 5 中可以看到,精确率 P 总是维持着相当稳定的水平, F_1 分数主要与召回率 R 相关,并且两者的波动不大。5 个数据集上的 F_1 分数的浮动范围分别保持在 2.76%, 1.24%, 2.26%, 0.60%, 0.65% 以内,可以看出模型在各种取值下的性能都比较稳定。模型的差异化目标包含了在超球空间中对

正常表示进行压缩,以及将异常表示推离球心,权重 λ 则分别控制了两者在优化过程中的重要性,受注入合成异常与测试集中真实异常的分布差异以及数据量等因素的影响,该权重在不同的数据集中具有不同的理论最优值。整体而言,随着 λ 的改变,模型在各个数据集上均保持稳定且良好的性能。

从图 6 中可以看到,在 5 个数据集上的 F_1 分数的波动范围分别保持在 4.22%, 1.10%, 10.69%, 3.79%, 1.23%, 因此大部分数据集的性能相对于窗口大小 N 的变化都较为稳定,而 MSL 受到数据量相对较少的制约,其表现出召回率 R 和 F_1 分数随着 N 的增大而减小。要达到好的性能,应当同时考虑数据量以及异常模式的关系,以确定滑动窗口大小。

4.5.3 消融实验

1) 模块消融实验

为了验证模型中差异化学习的有效性,本节使用不同的设计在各个数据集上进行模块消融实验。实验参数同 4.2 节,模型的不同设计如下:

(1)DCDAD(without diff):保留 DCDAD 模型的主干网络结构,但去掉差异化学习模块,具体实现为不进行异常注入并且修改损失函数(即式(12)不包含第二项)。

(2)DCDAD(without clip):保留 DCDAD 的主干网络结构,以及差异化学习模块,但是不使用裁剪操作,具体实现为在式(9)中的差异化目标函数时。

(3)DCDAD 模型,即本文最终模型。

模块消融实验的结果如表 5 所列,不含有差异化(without diff)和截断目标函数(without clip)的 F_1 分数在 5 个数据集上均略微逊色于最终模型 DCDAD,说明差异化的学习目标以及截断操作能够分别对异常检测任务带来一定的改进,也进一步表明了提出模型各个模块的有效性。

2) 指标消融实验

为了验证模型中所使用的异常指标的有效性,即上下文依赖关系与超球中心的径向距离 dis ,本节使用不同的异常指标在各个数据集上进行指标消融实验。实验参数同 4.2 节,模型的不同设计如下:

(1)DCDAD(only rec):使用重构损失作为异常指标,将式(15)的右式替换为各时刻的重构损失(rec)作为最终的异常分数。

(2)DCDAD(with rec):在异常指标中添加重构损失,将式(15)计算结果与各时刻的重构损失进行点乘的结果($rec \odot dis$)作为最终的异常分数。

(3)DCDAD 模型,使用式(15)计算结果作为异常指标,即本文最终模型采用的指标。

该指标消融实验的结果如表 6 所列。仅使用 rec 在本文训练的模型上表现不佳,而将 rec 与 dis 相结合的效果不如仅使用 dis ,表明了所提出的指标具备较好的异常识别能力,而重构误差的加入会干扰异常检测的效果。对于多元时间序列,维度的数量和不同维度上量纲的差异会使得重构误差变得不稳定,可能导致异常分数的计算结果受到噪声或无关特征的影响,而上下文依赖关系与超球中心的径向距离在一定程度上避免了这种影响。

表 5 模块消融实验结果

Table 5 Module ablation experiment results

算法\数据集	diff	clip	SMD	PSM	MSL	SMAP	SWaT	平均 F_1
DCDAD(without diff)	×	×	90.55	97.74	92.16	96.96	96.53	94.79
DCDAD(without clip)	✓	×	93.65	98.13	93.31	97.06	97.59	95.96
DCDAD	✓	✓	94.35	98.45	94.06	97.27	97.90	96.41

注:加粗、下划线分别表示最优、次优的结果。

表 6 指标消融实验结果

Table 6 Indicator ablation experiment results

算法\数据集	指标	SMD	PSM	MSL	SMAP	SWaT	平均 F_1
DCDAD(only rec)	rec	77.24	83.76	87.74	70.58	79.28	79.72
DCDAD(with rec)	rec⊙dis	94.22	96.53	92.08	96.43	91.18	94.09
DCDAD	dis	94.35	98.45	94.06	97.27	97.90	96.41

注:加粗、下划线分别表示最优、次优的结果。

4.5.4 可视化分析

1) 样本特征降维可视化

为了探索差异化表示学习的效用,本节首先对上下文依赖这一特征进行可视化比较。具体而言,本实验将 DCDAD (without diff)、DCDAD 学习到的依赖特征作为 t-SNE 方法的输入,从 SWaT 数据集中随机选择 5000 个样本点,计算其在二维空间中的表示并可视化,得到如图 7 和图 8 所示的结果。其中紫色对应正常数据,蓝色对应异常数据。

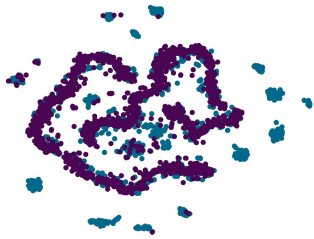


图 7 DCDAD(without diff)特征可视化(电子版为彩图)

Fig. 7 Feature visualization on DCDAD(without diff)

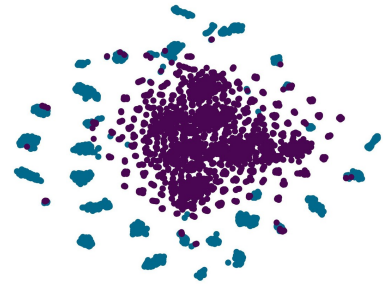


图 8 DCDAD 特征可视化

Fig. 8 Feature visualization on DCDAD

DCDAD(without diff)的结果显示,在未经过差异化表示学习得到的正常数据分布集中,相当多的异常数据与正常数据分布重合;而 DCDAD 的结果中,异常数据和正常数据的重合度更低,异常数据在正常数据的分布外显著地聚集。从对比结果可以看出,DCDAD 由于考虑了注入异常样本的特征,以及差异化的表示学习,使得异常样本和正常样本的表示更具区分性,提升了异常检测的效果。

2) 异常分数可视化

本节对真实数据集中的输入数据以及本文方法输出的异常分数进行可视化。可视化结果如图 9 所示。第 1 行是截取的序列数据,用灰色背景和红色曲线标注了异常片段,红点标注了点异常,第 2 行是模型输出的异常分数。

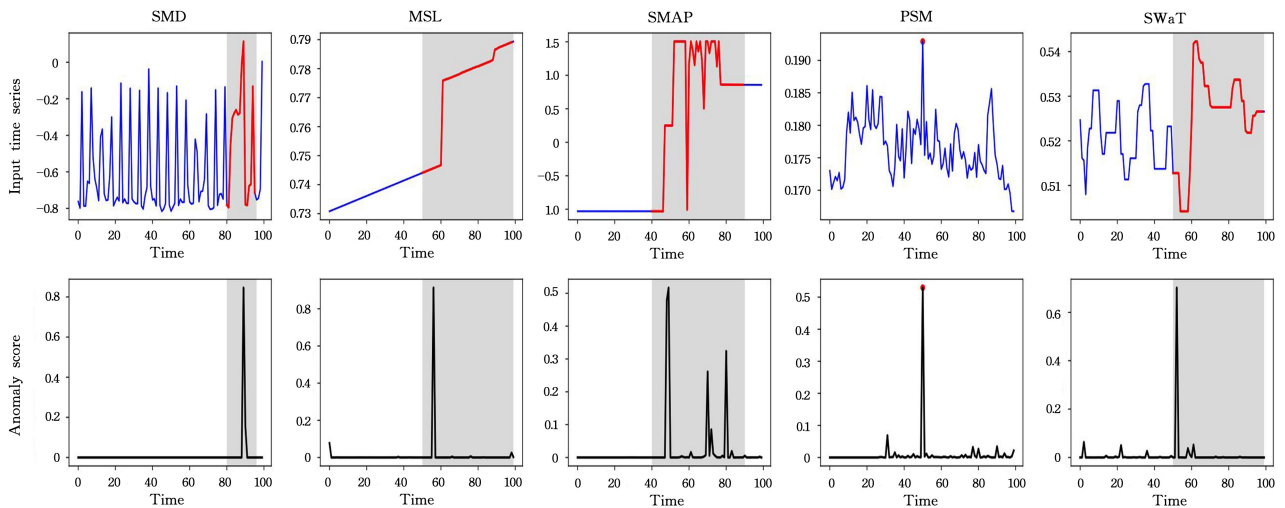


图 9 测试集窗口输入和模型输出的异常分数

Fig. 9 Anomaly scores of window input and model output on test set

实验中,分别选取了各数据集中具有代表性的异常片段。(1)SMD 数据集中的季节性异常:在时间序列中出现的与周期性季节模式不一致的异常;(2)MSL 和 SMAP 数据集中的集合异常:某个数据子集的整体行为与整体数据模式不一致的异常;(3)PSM 数据集中的点异常:单个数据点与其周围数

据点显著不同的异常;(4)SWaT 数据集中的趋势性异常:与长期趋势不一致的异常。

可视化结果显示,在各种类型的连续异常片段上,本文方法在接近异常发生的时刻附近输出了较高的异常分数,具备较好的早期检测能力,同时也能够精确定位到突出的点异常。

4.5.5 模型大小与推理速度对比

本小节对模型的计算量、参数量、推理时间进行分析。以MSL数据集为例,从中取出一个窗口样本作为输入示例,其维度为[1,100,55],3个取值分别对应batchsize、窗口长度、属性数量。本小节将输入示例输入模型进行推理并计算出对应的模型计算量、参数量、推理时间,与最新的模型Anomaly Transformer,DCdetector进行对比,其结果如表7所列。

表7 模型大小与推理速度对比

Table 7 Comparison of model size and inference speed

算法\指标	FLOPs	Params	FPS/ms
AnomalyTransformer	485.79×10^6	4.86×10^6	<u>7.23</u>
DCdetector	1265.50×10^6	0.68×10^6	11.74
DCDAD	104.65×10^6	<u>1.05×10^6</u>	3.74

注:加粗、下划线分别表示为最优、次优的结果。

表7中的结果表明,DCDAD模型在计算量和推理速度上都领先于Anomaly Transformer和DCdetector,提出的模型选用了较小的注意力的头数和隐藏层维度,在保持较好测试性能的情况下降低计算量并提升推理速度。DCDAD在参数量上多于DCdetector,主要是因为DCdetector运用了patch的思想,patch上的注意力模块参数量显著低于整个窗口上的。

结束语 时序异常检测在各个领域中都具有重要的应用。时间序列数据具有时间维度的上下文依赖关系,然而,当前的大多数研究只是将融合了时序上下文的特征用于增强模型的重构或者预测能力,而没有考虑数据所蕴含的上下文依赖关系的差异性;并且它们都是无监督的方法,缺乏对异常样本的建模。这些问题导致了模型在确定异常和正常的界限时表现模糊,检测结果欠佳。

针对以上问题,本文提出了DCDAD模型,重点关注时序数据的上下文依赖关系,通过差异化正、异常样本表示学习的方式来增强样本的上下文依赖关系挖掘,并在聚类过程中学习可区分正、异常的超球面。通过异常注入的方式扩充训练集,以及设计差异化的目标函数,促使正常和异常样本在超球空间中的分离性更加显著,解决现有异常检测模型对于正、异常样本区分度不明显的问题。最终利用超球面作为划分边界,完成异常检测任务。同时,通过在5个真实时序数据集上的大量实验,证实了差异化地学习上下文依赖关系可显著提升模型对异常的检测效果,验证了该方法的有效性。

未来将在本文方法的基础上,研究如何将模型扩展到各个领域的大规模数据集上,改进模型以提取和适应更加复杂的上下文依赖关系。其次,我们将探索如何设计高效的模型和算法,以实现实时的异常检测,应对大规模数据和高速数据流的挑战。此外,现实世界中的异常往往涉及多个传感器或数据源,需要考虑将各个维度的信息融合起来,进一步提升时间序列数据上的检测精度以及泛化性能。

参考文献

- [1] BLÁZQUEZ-GARCÍA A, CONDE A, MORI U, et al. A review on outlier anomaly detection in time series data[J]. ACM Computing Surveys, 2021, 54(3):1-33.
- [2] CHOI K, YI J H, PARK C, et al. Deep learning for anomaly detection in time-series data: review, analysis, and guidelines[J]. IEEE Access, 2021, 9:120043-120065.
- [3] GREFF K, SRIVASTAVA R K, KOUTNÍK J, et al. LSTM: A search space odyssey[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(10):2222-2232.
- [4] ZHANG G H, YAN X F, GUAN D H. Anomaly Detection of Time-series Based on Multi-modal Feature Fusion[J/OL]. https://www.jsjx.com/CN/article/openArticlePDF.jsp?id=21650.
- [5] BREUNIG M M, KRIEGEL H P, NG R T, et al. LOF: identifying density-based local outliers[C]// Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2000:93-104.
- [6] LIU F T, TING K M, ZHOU Z H. Isolation-based anomaly detection[J]. ACM Transactions on Knowledge Discovery from Data, 2012, 6(1):1-39.
- [7] SCHÖLKOPF B, PLATT J C, SHAWE-TAYLOR J, et al. Estimating the support of a high-dimensional distribution. [J] Neural Computation, 2001, 13(7):1443-1471.
- [8] RUFF L, VANDERMEULEN R, GOERNITZ N, et al. Deep one-class classification[C]// Proceedings of The 37th International Conference on Machine Learning. PMLR, 2018:4393-4402.
- [9] 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.
- [10] TAO T, ZHOU X, MA B, et al. Abnormal time series data detection of gas station by Seq2Seq model based on bidirectional long short-term memory [J]. Computer Applications, 2019, 39(3):924-929.
- [11] ZHOU B, LIU S H, HOUI B, et al. Beatgan: Anomalous rhythm detection using adversarially generated time series[C]// Proceedings of the 28th International Joint Conferences on Artificial Intelligence. AAAI, 2019:4433-4439.
- [12] 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. ACM, 2019:2828-2837.
- [13] WANG L, ZHANG X Y, XV B W, et al. InterFusion: Interaction-based 4D radar and LiDAR fusion for 3D object detection [C]// 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2022:12247-12253.
- [14] SHEN L F, LI Z C, KWOK J. Timeseries anomaly detection using temporal hierarchical one-class network[J]. Advances in Neural Information Processing Systems, 2020, 33:13016-13026.
- [15] XU J H, WU H X, WANG J M, et al. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy [C]// Proceedings of the 10th International Conference on Learning Representations(Virtual). 2022:1-20.
- [16] YANG Y Y, ZHANG C L, ZHOU T, et al. DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection[C]// Proceedings of the 29th ACM SIGKDD International Conference on Knowledge Discovery & Data Mi-

- ning. ACM,2023:3033-3045.
- [17] 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. ACM,2019:2123-2133.
- [18] CARMONA C U, AUBET F X, FLUNKERT V, et al. Neural contextual anomaly detection for time series[C]//Proceedings of the 31st International Joint Conferences on Artificial Intelligence. 2022:4433-4439.
- [19] HENDRYCKS D, MAZEIKA M, DIETTERICH T. Deep anomaly detection with outlier exposure[J]. arXiv:1812.04606, 2018.
- [20] RUFF L, VANDERMEULEN R A, FRANKS B J, et al. Rethinking assumptions in deep anomaly detection[J]. arXiv:2006.00339,2020.
- [21] ZONG B, SONG Q, MIN M R, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection[C]//Proceedings of the 6th International Conference on Learning Representations. 2018:1-19.
- [22] YAIRI T, TAKEISHI N, ODA T, et al. A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction[J]. IEEE Transactions on Aerospace and Electronic Systems, 2017, 53(3):1384-1401.
- [23] ANDERSON O D. Time-Series[J] Journal of the Royal Statistical Society Series D: The Statistician, 1976, 25(4):308-310.
- [24] CHO K, VAN MERRIENBOER B, GULCEHRE C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. 2014:1724.
- [25] YIN C Y, ZHOU L W. Unsupervised time series anomaly detection model based on re-encoding[J]. Journal of Computer Applications. 2023, 43(3):804-811.
- [26] ZHAO H, WANG Y J, DUAN J Y, et al. Multivariate time-series anomaly detection via graph attention network[C]//Proceedings of the 2020 IEEE International Conference on Data Mining. 2020:841-850.
- [27] ZENG W R, WU J, YAN F. Time Series Anomaly Detection Model Based on Hierarchical Temporal Memory[J]. Acta Electronica Sinica, 2018, 46(2):325-332.
- [28] ZHANG S Q, ZHANG Z Y, JIANG Y, et al. Time Series Theory and Algorithm of Predictable Learnability Based on Error Truncation Assumption[J]. Chinese Journal of Computers, 2022, 45(11):2279-2289.
- [29] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2014, 63:139-144.
- [30] KINGMA D P, WELING M. Auto-encoding variational Bayes[J]. arXiv:1312.6114, 2013.
- [31] LI D, CHEN D C, JIN B H, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks[C]//Proceedings of the 28th International Conference on Artificial Neural Networks. 2019:703-716.
- [32] CHEN Z K, CHEN D S, ZHANG X, et al. Learning graph structures with transformer for multivariate time-series anomaly detection in IoT[J]. IEEE Internet of Things Journal, 2021, 9(12):9179-9189.
- [33] TULI S, CASALE G, JENNINGS N R. TranAD: deep transformer networks for anomaly detection in multivariate time series data[J]. Proceedings of the VLDB Endowment, 2022, 15(6):1201-1214.
- [34] NIE Y, NGUYEN N H, SINTHONG P, et al. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers[C]//Proceedings of the 11th International Conference on Learning Representations. 2023:1-24.
- [35] ABDULAAL A, LIU Z, LANCEWICKI T. Practical approach to asynchronous multivariate time series anomaly detection and localization[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021:2485-2494.
- [36] ENTEKHABI D, NJOKU E G, O'NEILL P E, et al. The soil moisture active passive(SMAP) mission[J]. Proceedings of the IEEE, 2010, 98(5):704-716.
- [37] 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. 2018:387-395.
- [38] GOH J, ADEPU S, JUNEJO K N, et al. A dataset to support research in the design of secure water treatment systems[C]//Proceedings of the 11th International Conference on Critical Information Infrastructures Security. 2016:88-99.
- [39] XU H W, CHEN W X, ZHAO N W, et al. Unsupervised anomaly detection via variational auto-encoder for seasonal kpis in web applications[C]//Proceedings of the 2018 World Wide Web Conference. 2018:187-196.



LIAO Sirui, born in 2000, master candidate. His main research interests include spatio-temporal data mining and anomaly detection.



HUANG Feihu, born in 1990, Ph.D, lecturer. His main research interests include spatio-temporal data mining, prognostics and health management and so on.