

基于KAN的无监督多元时间序列异常检测网络

王成, 金城

引用本文

王成, 金城. 基于KAN的无监督多元时间序列异常检测网络[J]. 计算机科学, 2026, 53(1): 89-96.

WANG Cheng, JIN Cheng. [KAN-based Unsupervised Multivariate Time Series Anomaly Detection Network](#) [J]. Computer Science, 2026, 53(1): 89-96.

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)

[EvR-DETR:融合事件与RGB图像的轻量级端到端目标检测](#)

EvR-DETR:Event-RGB Fusion for Lightweight End-to-End Object Detection

计算机科学, 2026, 53(1): 153-162. <https://doi.org/10.11896/jsjcx.250300021>

[基于深度学习的OCT/OCTA视网膜图像分析方法综述](#)

Review of Retinal Image Analysis Methods for OCT/OCTA Based on Deep Learning

计算机科学, 2026, 53(1): 128-140. <https://doi.org/10.11896/jsjcx.241100047>

[图嵌入学习研究综述:从简单图到复杂图](#)

Review of Graph Embedding Learning Research:From Simple Graph to Complex Graph

计算机科学, 2026, 53(1): 58-76. <https://doi.org/10.11896/jsjcx.250300081>

[基于异构图和指令序列的智能合约字节码漏洞检测方法](#)

Smart Contract Bytecode Vulnerability Detection Method Based on Heterogeneous Graphs and Instruction Sequences

计算机科学, 2025, 52(12): 367-373. <https://doi.org/10.11896/jsjcx.241100076>

[人工智能在金融领域的的数据预测方法综述](#)

Survey of Data Prediction Methods Using Artificial Intelligence in the Financial Sector

计算机科学, 2025, 52(12): 271-284. <https://doi.org/10.11896/jsjcx.250700166>

基于 KAN 的无监督多元时间序列异常检测网络

王 成 金 城

复旦大学计算机科学技术学院 上海 200438

(whiletruedo@163.com)

摘 要 时间序列数据在金融、医疗、工业和交通等领域中广泛存在,异常检测对确保系统稳定和安全至关重要。由于异常样本的收集十分困难,当前大多数时间序列异常检测方法是无监督的。然而,这些方法普遍存在过泛化问题,即模型不仅能重建正常样本,还能很好地重建异常样本。这一问题使得异常检测效果不佳。因此,提出了一种基于 Kolmogorov-Arnold 表示理论的时间序列异常检测方法 TS-KAN,利用其参数高效性与局部可塑性,使模型更好地拟合正常样本并缓解过泛化问题。此外,提出了局部特征增强层 Local-KAN,以增强时域特征的表达能力,提高上下文异常检测能力。在 5 个主流时间序列异常检测数据集上的实验表明,TS-KAN 的异常检测能力显著优于现有方法。

关键词: 时间序列异常检测;KAN;Transformer;记忆模块;深度学习

中图分类号 TP391

KAN-based Unsupervised Multivariate Time Series Anomaly Detection Network

WANG Cheng and JIN Cheng

School of Computer Science, Fudan University, Shanghai 200438, China

Abstract Time series data is widely present in fields such as finance, healthcare, industry, and transportation. Time Series Anomaly Detection (TSAD) is crucial for ensuring system stability and safety. Most current time series anomaly detection methods are unsupervised due to the difficulty in collecting anomaly samples. However, these methods commonly face the problem of over-generalization, where the model can not only reconstruct normal samples, but also effectively reconstruct anomaly samples, leading to poor anomaly detection performance. Therefore, this paper proposes a time series anomaly detection method based on Kolmogorov-Arnold representation theory, called TS-KAN. TS-KAN leverages its parameter efficiency and local plasticity to better fit normal samples and alleviate the overgeneralization problem. Additionally, this paper introduces a local feature enhancement layer, namely Local-KAN, to enhance the representation of temporal features and improve contextual anomaly detection capability. Experiments on five mainstream time series anomaly detection datasets demonstrate that TS-KAN significantly outperforms existing methods in anomaly detection capability.

Keywords Time series anomaly detection, KAN, Transformer, Memory module, Deep learning

1 引言

时间序列异常检测是指从时间序列中识别出与正常模式显著不同的异常样本,在金融分析、医疗诊断、工业控制、交通管理等众多领域中具有重要的应用价值。然而,时间序列异常检测面临着诸多挑战:异常样本数量少且稀有,标注代价高昂,异常样本在形态上往往存在较大的差异性,导致传统有监督学习方法难以有效应用。现有研究多采用无监督的方法,通过仅依赖正常样本的训练数据来检测异常样本。

目前主流的时间序列异常检测方法主要基于重建损失来判断异常,即利用模型重建输入数据的能力,将重建误差较大的时间序列视为异常样本。然而,这类方法通常存在过度泛

化的问题,即模型的编码器能力过强,不仅能准确重建正常样本,而且能较为准确地重建部分异常样本,进而导致检测效果不佳。为缓解此问题,一些研究提出在自编码器结构中引入记忆模块,通过记忆正常样本的特征来区分正常与异常样本。然而,这些方法的记忆模块往往基于相似度进行更新,这使得模块更倾向于记住简单正常样本的共性特征,而忽略了困难正常样本的独特特性,从而容易将困难样本与异常样本混淆。

为了解决上述问题,本文提出了一种基于 Kolmogorov-Arnold 表示理论的时间序列异常检测方法 Time Series Kolmogorov-Arnold Network (TS-KAN)。该方法将 Kolmogorov-Arnold 表示引入时间序列异常检测领域,并提出了局部特征增强模块 Local-KAN,结合 Transformer 结构在多维时间序

到稿日期:2024-12-26 返修日期:2025-03-12

基金项目:国家自然科学基金(62472097)

This work was supported by the National Natural Science Foundation of China(62472097).

通信作者:金城(jc@fudan.edu.cn)

列上进行长程特征融合。此外,还设计了一种基于困难样本增强的记忆模块,通过对困难样本特征的深度挖掘,迫使模型记住困难正常样本的更多有价值的独特特征,提高模型区分正常与异常样本的能力,进而提升异常检测的性能。本文的主要贡献如下:

1)提出了基于 Kolmogorov-Arnold 表示理论的时间序列异常检测方法 TS-KAN 用于高效拟合正常样本,利用 KAN 的局部可塑性缓解过泛化问题,有效提升了异常检测性能;

2)设计了用于局部特征增强的 Local-KAN 模块,并提出了基于困难样本增强的记忆模块,通过迫使模型记住更多困难正常样本的独特特征,显著提升了模型对困难正常样本和异常样本的区分能力;

3)在多个基准数据集上进行了广泛的实验,实验结果表明,所提方法在 5 个主流时间序列异常检测数据集上的检测性能均优于现有的主流方法。

2 相关工作

2.1 无监督时间序列异常检测

由于异常数据具有难收集、差异大等特点,主流的时间序列异常检测方法大多是无监督的,即这些方法的训练集不包含异常数据,异常数据仅存在于测试集中。它们普遍通过建模正常数据的特征与分布或设立一个代理任务,来间接地将异常数据检测出来,即根据其在特征与分布上的差异或在代理任务上的表现给出其异常得分。根据建模方式的不同,无监督时间序列异常检测方法可以分为密度估计法、聚类法、自回归法和重建法 4 类。

密度估计法首先在训练集上拟合正常样本的分布。测试时,这类方法首先估计给定样本属于正常样本的概率密度,再将概率密度低于特定阈值的样本预测为异常样本。DAGMM^[1]和 MPPCAD^[2]采用高斯混合分布来建模正常样本分布。MTGFlow^[3]采用图结构学习和标准化流来更准确地建模正常样本分布。KDE-AE^[4]利用核密度估计函数来提高正常样本和异常样本之间的分布差异。

聚类法首先将正常样本特征聚类成若干类,再通过与最近聚类中心的距离来判断给定样本是否为异常样本。THOC^[5]将单分类网络中的若干个中间层特征分别进行聚类。ITAD^[6]将特征解耦后再分别进行聚类。STS-AD^[7]通过对子时间序列进行聚类来更准确地定位异常样本区间。

自回归法将时间序列预测作为代理任务,将预测不准确的点作为异常样本。LSTMs-AD^[8]采用 LSTM 进行时间序列预测,并设置动态误差阈值来确定异常样本。CL-MPPCA^[9]采用 LSTM 来进行时间序列预测,并采用混合概率主成分分析来根据预测结果计算异常得分。

重建法是目前最主流的一类方法。重建法将时间序列重建作为代理任务,将样本的重建损失作为其异常得分。LSTM-VAE^[10]采用 LSTM 进行特征提取,并采用变分自编码器进行重建。OmniAnomaly^[11]在 LSTM-VAE 的基础上增加标准化流来提升正常样本与异常样本之间的差异。InterFusion^[12]在时间维度和多元变量间分别进行特征融合,再利用变分自编码器进行样本重建。AnomalyTransformer

(A. T.)^[13]引入 Transformer 在多元变量之间进行特征交流,取得了不错的表现,证明了 Transformer 在多元变量特征交流上的有效性,以及其在时间序列领域的潜力。

重建法由于其出色的表现被广泛应用于各种异常检测任务中。目前重建法最大的挑战是过泛化问题,即编码器具有过强的泛化能力,使其不仅能够重建正常样本,也能很好地重建异常样本,进而使这部分异常样本被漏检。Sun 等^[14]提出了 UMTS-Mixer,该方法利用 MLP 的顺序敏感特性,将其用于捕获时间相关性和跨通道相关性,进而准确区分正常样本和异常样本。Ye 等^[15]提出了 WMAD,该方法利用融合小波分解,从长时间的趋势变化和短时间的瞬时变化角度捕捉复杂时间模式,进而提高了精度。MemAE^[16]提出一个外置记忆模块来记忆正常样本,以此缓解模型过泛化到异常样本上的问题。MNAD^[17]提出了一种两阶段的外置记忆模块更新方法,避免了记忆模块的趋同化记忆,提升了对正常样本分布的覆盖率。MEMTO^[18]采用 KNN 聚类结果作为外置记忆模块的初始化,并提出门控更新单元来进一步避免记忆模块的趋同化记忆。这些方法均采用 Transformer 结构作为其特征提取器,利用了 Transformer 在多元变量特征交流上的强大能力。然而,Transformer 缺乏归纳偏置,因此其具有极强的泛化能力,虽然这些方法通过正则化、记忆正常样本等手段来限制其泛化能力,但是过泛化问题仍然很严重。

2.2 KAN

KAN(Kolmogorov-Arnold Network)^[19]基于 Kolmogorov-Arnold 表示理论,设计了基于三阶均匀 B 样条的小规模网络替代 MLP。KAN 具有良好的局部可塑性、参数高效性等特点^[20],在函数拟合、图像分类等多个基础任务上具有超过 MLP 的表现,展现出了在众多任务上替代 MLP 的潜力。CheyKAN^[21],fKAN^[22]和 Wav-KAN^[23]在 KAN 的基础上分别尝试利用切比雪夫多项式、雅克比基函数、小波函数等作为三阶 B 样条的替代,也获得了不错的表现。U-KAN^[24]将 KAN 应用在医学图像分割和生成领域,取得了 SOTA 的结果。KAN-TS^[25]将 KAN 应用于单元时间序列预测任务上,利用 KAN 进行时域关系建模,取得了不错的效果。TKAN^[26]将 KAN 与 LSTM 结合,利用 KAN 进行多元变量关系建模,并利用 LSTM 进行时域关系建模,在短期时间序列预测任务上取得了具有竞争力的结果,但 LSTM 存在遗忘问题,其在长期时间序列预测上的效果不佳。

考虑到 KAN 具有局部可塑性和参数高效性,可以为 Transformer 提供异常检测所需的归纳偏置,本文创新地将 KAN 应用于多元时间序列异常检测任务,并提出了 KAN-TSAD。该网络将所提的 Local-KAN 融合到 Transformer Encoder 中,利用 Local-KAN 为 Transformer 提供额外的归纳偏置,缓解了时间序列异常检测中的过泛化问题。这一做法使得网络不仅具有 KAN 的局部可塑性和参数高效性,还具有 Transformer 在多元时间序列进行长距离特征融合的优势。

3 本文方法

3.1 问题描述

多元时间序列是将多个采样结果按照时间排列而成的序

列。不失一般地,时间序列可定义为:

$$S = [S_1, S_2, \dots, S_T], S_i \in \mathbb{R}^N, i=1, 2, \dots, T \quad (1)$$

其中, T 表示时间序列 S 的总长度, N_i 表示时间序列的变量数, S_i 表示第 i 个采样数据。

时间序列异常检测问题定义为:对于一个给定的时间序列 $S \in \mathbb{R}^{T \times N}$, 为每个时间点给出一个异常判定结果 $r^a \in \{0, 1\}^T$, 用于表示该样本是否是异常样本。而无监督时间序列异常检测问题则是在时间序列异常检测的基础上额外要求模型不采用任何异常样本作为其训练数据。

本文所使用符号的含义如表 1 所列。

表 1 符号及其含义
Table 1 Symbols and their meaning

| 符号 | 含义 |
|---------------|-----------|
| \oplus | 拼接操作 |
| \triangleq | 定义为 |
| \rightarrow | 函数映射 |
| \circ | 函数复合运算 |
| \otimes | 逐元素相乘 |
| $\psi(\cdot)$ | ReLU 激活函数 |
| $\ \cdot\ _2$ | 欧几里德范数 |

3.2 前置知识

传统的多层感知机 (Multi-Layer Perceptron, MLP) 基于通用近似理论^[27]。该理论认为具有有限个神经元的两层全连接层可以在 \mathbb{R}^n 的紧子集上近似连续函数。而 KAN 则基于 Kolmogorov-Arnold 表示理论^[28]。该理论认为任意多元连续函数可以表示为单元连续函数与加法的组合, 即对于一个光滑的多元连续函数 $f: [0, 1]^n \rightarrow \mathbb{R}$, 存在一组单元连续函数 $\phi_{p,q}: [0, 1] \rightarrow \mathbb{R}$ 和 $\Phi_q: \mathbb{R} \rightarrow \mathbb{R}$, 使得:

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^n \phi_{q,p}(x_p) \right) \quad (1)$$

Kolmogorov-Arnold 表示理论从理论上证明了仅使用加法作为多元聚合函数即可将任意多元连续光滑的函数转换为两层单元函数与加法的组合。Liu 等^[19] 认识到, 看似不同的外部函数 Φ_q 和内部函数 $\phi_{q,p}$ 可以通过他们提出的 KAN 层进行统一, 从而设法将其扩展到任意深度, 提升了其实用性。一般地, 一个深度为 L 的 KAN 由 L 个堆叠的 KAN 层组成, 其形状可以表示为 $\text{KAN}[n_0, n_1, \dots, n_L]$, 其中 n_l 表示第 l 层的特征维度。第 l 层 KAN 层将输入的向量 $\mathbf{x}_{l-1} \in \mathbb{R}^{n_{l-1}}$ 变换为向量 $\mathbf{x}_l \in \mathbb{R}^{n_l}$, 即:

$$\mathbf{x}_l = \underbrace{\begin{bmatrix} \phi_{l,1,1} & \phi_{l,1,2} & \dots & \phi_{l,1,n_{l-1}} \\ \phi_{l,2,1} & \phi_{l,2,2} & \dots & \phi_{l,2,n_{l-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{l,n_l,1} & \phi_{l,n_l,2} & \dots & \phi_{l,n_l,n_{l-1}} \end{bmatrix}}_{\Phi_l} \mathbf{x}_{l-1} \quad (2)$$

则一个 L 层的 KAN 可以表示为:

$$\text{KAN}(\mathbf{x}) = (\Phi_L \circ \Phi_{L-1} \circ \dots \circ \Phi_1)(\mathbf{x}) \quad (3)$$

具体而言, 在 KAN 中 ϕ 采用三阶均匀 B 样条作为基函数进行拟合, 并增加了残差连接的设计, 即:

$$\phi(\mathbf{x}) = \mathbf{w}_b \text{SiLU}(\mathbf{x}) + \mathbf{w}_s \text{spline}(\mathbf{x}) \quad (4)$$

其中, \mathbf{w}_b 和 \mathbf{w}_s 为两个可学习的权重参数, $\text{SiLU}(\cdot)$ 表示 Sigmoid 线性单元 (Sigmoid Linear Unit) 激活函数^[29],

$\text{spline}(\cdot)$ 则为三阶均匀 B 样条函数 $B(\cdot)$ 的线性组合:

$$\text{spline}(\mathbf{x}) = \sum_i \mathbf{c}_i B_i(\mathbf{x}) \quad (5)$$

其中, \mathbf{c}_i 为可学习的权重参数。

3.3 网络结构

TS-KAN 采用包含记忆模块的自编码器结构。该网络由局部特征增强的编码器、基于困难样本增强的记忆模块和解码器 3 个部分组成。其网络结构如图 1 所示。首先将 TS-KAN 的输入 S 送入基于 KAN 的编码器 \mathcal{E} , 得到编码后的查询特征 $\mathbf{q} \in \mathbb{R}^{T \times C}$ 。之后, 该特征将被送入基于困难样本增强的记忆模块 \mathcal{M} , 利用相关的记忆项特征进一步增强该查询特征。增强后的特征将被送入解码器 \mathcal{D} , 用于回归出原始输入 S' 。在训练时, 查询特征将用于不断更新所提记忆模块, 从而使记忆模块记住更多的正常样本特征。在测试时, 为防止记忆模块错误地记住异常样本, 所提记忆模块仍然会提供相关记忆项的特征来增强查询特征, 但记忆模块将不再被更新, 取而代之的是, 该记忆模块将计算该查询特征同其最相关的记忆项特征之间的差异大小, 作为最终异常判定的依据之一。

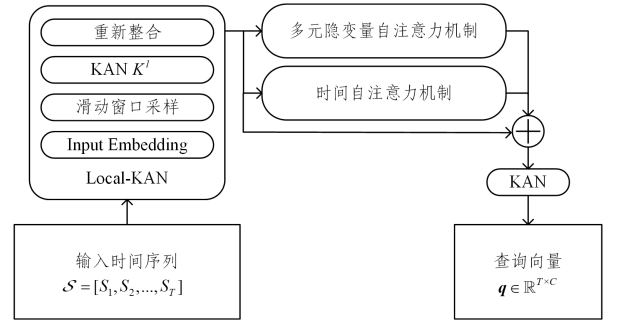


图 1 TS-KAN 的基本结构

Fig. 1 Structure of TS-KAN

3.4 局部特征增强的编码器

如图 2 所示, 局部特征增强的编码器首先将多元时间序列变换到高维隐特征空间中; 然后采用提出的 Local-KAN 模块来进行时间局部特征增强, 用于更好地区分出时间序列中常见的上下文异常; 之后利用自注意力机制在时间维度和多元隐变量特征维度分别提取长程和全局依赖关系; 最后采用 KAN 对这些不同维度的特征进行交流和聚合, 得到查询特征 $\mathbf{q} \in \mathbb{R}^{T \times C}$ 。其中, C 表示隐空间的特征维数。

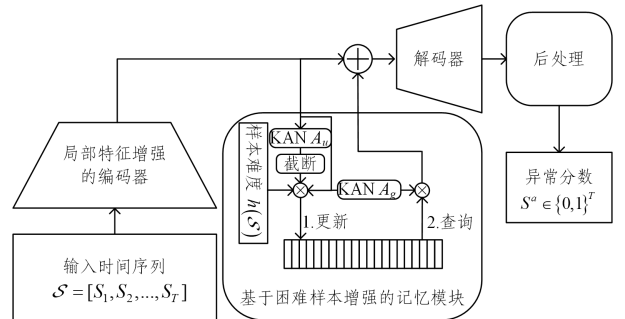


图 2 局部特征增强的编码器

Fig. 2 Structure of local-enhanced encoder

具体而言, 所提编码器首先通过输入编码将输入 $S \in \mathbb{R}^{T \times N}$ 变换到高维空间中, 得到编码后的特征 $\mathbf{E}' \in \mathbb{R}^{T \times C}$ 。为了增强编码特征对局部变化模式的感知, TS-KAN 提出了一

个用于局部特征增强的 Local-KAN 网络。具体而言, TS-KAN 采用一个窗口大小为 w 的滑动窗口对特征 \mathbf{E}^s 进行采样, 得到 $\mathbf{E}^w \in \mathbb{R}^{(T-w+1) \times w \times C}$ 。该特征随后将被送入一个 KAN 进行局部特征建模, 得到 $\mathbf{E}^l \in \mathbb{R}^{(T-w+1) \times w \times C}$, 即:

$$\mathbf{E}_{i,j}^l = K_j^l(\mathbf{E}_{k,j}^w) \quad (7)$$

其中, $K_j^l: \mathbb{R}^w \rightarrow \mathbb{R}^w \triangleq \text{KAN}[\tau, 2w, w]$ 用于进行局部特征增强。之后利用平均池化将增强后的特征重新整合为 $\mathbf{E}^e \in \mathbb{R}^{T \times C}$, 即:

$$\mathbf{E}_i^e = \frac{1}{\sigma(i)} \max_{k=i-w+1}^{\min(i, T-w+1)} \sum_{j=k}^{T-w+1} \mathbf{E}_j^l \quad (8)$$

$$\sigma(i) = \max(i, T-w+1) - \min(i-w+1, 1) + 1 \quad (9)$$

被局部特征增强之后的特征 \mathbf{E}^e 将分别通过时间自注意力机制 (Temporal Self-Attention, TSA) 和多元隐变量自注意力机制 (Multivariate Latent Self-Attention, MLSA), 用于提取长程时域依赖关系和多元隐变量之间的依赖关系, 即:

$$\mathbf{E}^a = \text{LN}(\mathbf{B} \oplus \text{TSA}(\mathbf{E}^e) \oplus \text{MLSA}(\mathbf{E}^e)^T) \quad (10)$$

其中, LN 为层归一化 (Layer Normalization) 操作, \oplus 表示通道层面的拼接操作, 而 TSA 和 LSA 均是 Transformer^[30] 中标准的自注意力机制。TSA 和 LSA 的区别在于: TSA 在时间维度上进行特征关系建模, 而 LSA 建模的是不同隐变量之间的依赖关系。 $\mathbf{E}^a \in \mathbb{R}^{T \times 3C}$ 随即被送到 $K^q: \mathbb{R}^{T \times 3C} \rightarrow \mathbb{R}^{T \times C} \triangleq \text{KAN}[3C, 2C, C]$ 中, 用于整合不同来源的特征。

$$\mathbf{q} = \text{LN}(K^q(\mathbf{E}^a)) \quad (11)$$

随后, 该查询特征 $\mathbf{q} \in \mathbb{R}^{T \times C}$ 将被送入基于困难样本增强的记忆模块, 用于查询和更新相关记忆项, 并在测试时计算其同所记忆的正常样本特征之间的距离, 以此作为异常判定的依据之一。

3.5 基于困难样本增强的记忆模块

现有方法^[16-18]普遍采用记忆模块来帮助区分正常样本和异常样本。然而, 正常样本中简单样本广泛存在且特征相似, 而困难样本数量少且特征差异大。实际上, 简单正常样本往往较容易同异常样本区分开, 相反, 困难正常样本由于与异常样本的差异较小, 更容易混淆。但现有记忆模块的相似度计算机制使其更倾向于挖掘简单样本的共性特征, 而忽略了困难样本的独特特征, 导致其难以区分困难正常样本和异常样本。为了解决这一问题, 本文提出了一种基于困难样本增强的记忆模块。该模块利用 KAN 代替点积相似度来指导相关记忆项的更新, 从而扩展不同记忆项之间的差异性, 并引入困难样本挖掘机制, 迫使模块挖掘困难样本中更丰富的独特特征。

具体而言, 该模块包含 N_m 个记忆项 $\mathbf{M}_i \in \mathbb{R}^C, i \in \{1, \dots, N_m\}$ 。训练时, 对于从编码器中得到的每个查询特征 $\mathbf{q} \in \mathbb{R}^{T \times C}$, 外置记忆模块会首先将查询特征送入一个两层的 KAN, 来指导相关记忆项的更新:

$$\mathbf{M} \leftarrow \text{Norm}(\mathbf{M} + h(S)^\gamma \delta_\lambda(A_n(\mathbf{q}))^T \mathbf{q}) \quad (12)$$

$$A_n(\mathbf{q}) = \text{KAN}[C, C, N_m](\mathbf{q}) \quad (13)$$

其中, Norm 为 L2 归一化; $A_n(\cdot)$ 为一个两层的 KAN, 用于计算相关项特征同查询特征之间相关性; δ_λ 为截断函数, 用于防止不相关的记忆项被更新。

$$\delta_\lambda(x) = \begin{cases} 0, & x < \lambda \\ x, & x \geq \lambda \end{cases} \quad (14)$$

$h(S)$ 表示输入时间序列的样本难度, 样本难度越大, 其对记忆项的贡献就越大。 γ 为一个可学习的超参数, 表示困难样本的重要系数。该系数越大, 则困难样本相较于简单样本对记忆项的贡献比重就越大。 $h(S)$ 的计算式如下:

$$h(S) = \begin{cases} 1, & \text{epoch} = 1 \\ \frac{\mathcal{L}_{\text{last}}(S)}{\max_{S' \in S^{\text{train}}} \mathcal{L}_{\text{last}}(S')}, & \text{epoch} > 1 \end{cases} \quad (15)$$

其中, $\mathcal{L}_{\text{last}}(S)$ 表示该样本在上一个 epoch 中的损失函数值, S^{train} 为全体训练集样本。

更新后的记忆项 \mathbf{M} 将用于进一步增强查询向量。

$$\hat{\mathbf{q}} = \mathbf{q} \oplus \text{LN}(A_e(\mathbf{q})\mathbf{M}) \quad (16)$$

$$A_e(\mathbf{q}) = \text{KAN}[C, C, N_m](\mathbf{q}) \quad (17)$$

其中, \oplus 表示通道层面的拼接操作; $\hat{\mathbf{q}} \in \mathbb{R}^{T \times 2C}$ 为增强后的特征; $A_e(\cdot)$ 为两层的 KAN, 用于计算查询向量与更新后各记忆项特征之间的相关性。式 (16) 与式 (12) 的不同之处在于, 其在更新记忆项时需要利用截断操作来限制记忆项的更新数量, 而在增强查询向量时, 需要最大程度地保留信息以提高增强后特征的代表能力。因此式 (16) 不再使用截断操作, 并采用拼接操作取代逐元素加法, 来最大程度地保留信息。增强后的特征 $\hat{\mathbf{q}}$ 包含从输入时间序列中提取出的特征 \mathbf{q} , 以及来源于记忆模块的知识 $A_e(\mathbf{q})\mathbf{M}$ 。该特征 $\hat{\mathbf{q}}$ 随后将被送入解码器, 用于重建输入时间序列 S 。

3.6 解码器与损失函数

在异常检测任务中, 解码器存在一个潜在问题, 即当解码器的能力过强时, 可能会导致异常样本被高质量地重构, 从而造成异常样本的漏检。为了解决这一问题, TS-KAN 参考了 MEMTO^[18] 的思路, 提出了一种弱解码器架构, 该解码器由两层全连接层构成, 即:

$$S' = \psi(\hat{\mathbf{q}}\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2 \quad (18)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{2C \times C}, \mathbf{W}_2 \in \mathbb{R}^{C \times N_s}, \mathbf{b}_1 \in \mathbb{R}^C$ 和 $\mathbf{b}_2 \in \mathbb{R}^{N_s}$ 分别为两层全连接层的权重和偏差, ψ 是 ReLU 激活函数。

在训练时, TS-KAN 采用重建损失作为损失函数。该损失的计算过程如下。

首先, 计算样本的每一帧 S_t 同其重建结果 S_t' 之间的均方误差 (Mean Square Error, MSE), 记作重建误差 d_t^{re} :

$$d_t^{\text{re}} = \|S_t' - S_t\|_2^2 \quad (19)$$

然后, 将各帧重建误差之和作为最终的损失函数:

$$\mathcal{L} = \sum_{t=1}^T d_t^{\text{re}} \quad (20)$$

3.7 模型训练与异常判定

模型训练分为两个阶段, 在第一个阶段训练结束后, 将计算训练集所有样本的查询特征 \mathbf{q} , 并使用 K-means 聚类算法对查询特征进行聚类, 用于初始化第二阶段的各记忆项特征, 具体流程如算法 1 所示。其中, N_{e_1} 表示第一阶段训练的 epoch 数, N_{e_2} 表示第二阶段训练的 epoch 数。 ϕ_E, ϕ_M, ϕ_D 分别表示编码器 \mathcal{E} 、记忆模块 \mathcal{M} 和解码器 \mathcal{D} 的参数, $\phi = \{\phi_E, \phi_M,$

ϕ_D 表示模型的所有参数。

算法1 模型训练流程

输入:训练集 $S^{\text{train}} = \{\mathbf{x} | \mathbf{x} \in \mathbb{R}^{T \times N_c}\}$

输出:训练得到的模型参数 ϕ

1. 随机初始化模型 f 的参数 ϕ
2. (阶段一)重复 N_{e1} 次:
 - 对于训练集的每一个样本 $\mathbf{x} \in S^{\text{train}}$:
 - 2.1. 计算重建结果 $\mathbf{y} = \text{TSKAN}_{\phi}(\mathbf{x})$
 - 2.2. 计算重建误差 $\mathcal{L}(\mathbf{y}, \mathbf{x})$,并利用反向传播更新模型参数 ϕ
3. 计算训练集所有样本的查询特征 $Q = \{\mathbf{q} \in \mathbb{R}^{T \times C} | \mathbf{q} = \mathcal{E}_{\phi_x}(\mathbf{x}), \mathbf{x} \in S^{\text{train}}\}$
4. 使用K-means算法将所有查询特征 Q 聚类为 N_m 类,使用第 i 个聚类中心 C_i 重新初始化记忆项 \mathbf{M}_i 的参数 $\phi_{M_i} = C_i$
5. (阶段二)重复 N_{e2} 次:
 - 对于训练集的每一个样本 $\mathbf{x} \in S^{\text{train}}$:
 - 5.1. 计算重建结果 $\mathbf{y} = \text{TSKAN}_{\phi}(\mathbf{x})$
 - 5.2. 计算重建误差 $\mathcal{L}(\mathbf{y}, \mathbf{x})$,并利用反向传播更新模型参数 ϕ
6. 返回训练得到的模型参数 ϕ

测试时,遵循主流做法^[15,17-18],首先用模型估计出一个异常得分 s^a ,然后利用训练集的所有样本动态地计算出一个阈值 δ_{th} ,最后将测试集上所有异常得分大于该阈值的样本判定为异常。

具体而言,TS-KAN的异常得分由隐空间记忆偏差 \mathbf{d}' 和重建误差 \mathbf{d}^{rec} 两部分组成。隐空间记忆偏差 $\mathbf{d}' \in \mathbb{R}^T$ 定义为每帧查询向量 $\mathbf{q}_t \in \mathbb{R}^C$ 同最近的记忆项之间的均方误差:

$$\mathbf{d}'_t = \min_i \|\mathbf{q}_t - \mathbf{M}_i\|_2^2 \quad (21)$$

则TS-KAN的异常得分 $s^a \in \mathbb{R}^T$ 可以表示为:

$$s^a = \text{softmax}(\mathbf{d}') \otimes \mathbf{d}^{\text{rec}} \quad (22)$$

其中, \otimes 为逐元素乘法。

阈值 δ_{th} 的选取对异常判定至关重要,过高的阈值将使得异常样本大量被漏检,而过低的阈值将使正常样本被错检为异常样本。然而,不同数据集、不同模型的最优阈值差异非常大,导致阈值选取非常困难。因此,TS-KAN采用动态阈值选取方法^[31]。首先计算训练集 S^{train} 上所有样本的异常得分 s ,并取训练集上第 $\alpha\%$ 大的异常得分作为阈值 δ_{th} 。然后TS-KAN将测试集上所有异常得分大于 δ_{th} 的时间点标记为异常。

4 实验

4.1 实验设置

本文实验的系统环境为Ubuntu 24.04 64位。采用的深度学习框架为PyTorch 2.1,Python的版本为3.10。服务器包括1块Intel Xeon Silver 4210R CPU,1块Nvidia Titan RTX 4090 GPU,256GB内存和8TB硬盘。

TS-KAN的隐藏层特征维度 C 设置为512, τ_w 设置为10,温度超参数 τ 为0.1,截断系数 λ 为0.1,记忆项数量 N_m 为2048。训练采用Adam作为优化器,初始学习率为 1×10^{-4} ,批次大小为512。 N_{e1} 和 N_{e2} 均为100。

4.2 数据集

本文在5个公开数据集SMD^[32],MSL^[33],SMAP^[33],SWaT^[34]和PSM^[35]上进行对比实验和消融实验。

SMD(Server Machine Dataset)是一个大规模数据集,该

数据集包含来自28台机器的38个传感器的数据。SMAP(Soil Moisture Active Passive)数据集和MSL(Mars Science Laboratory)数据集是两个广泛使用的基准数据集。其中MSL包含火星探测器的执行器和传感器数据,共有55个维度,而SMAP包含土壤样本和火星探测器遥测数据,共有38个维度。SWaT(Secure Water Treatment)数据集包含一个水处理测试平台上51个传感器的测量数据。PSM(Pooled Server Metrics)数据集包含不同应用服务器节点上26个维度的数据。

与主流方法^[15,18]一致,本文采用精确率(Precision,P)、召回率(Recall,R)和F1分数作为评估指标。具体定义如下。

精确率表示在被模型判定为异常的样本中,实际为异常的样本的比例。

$$\text{Precision} = \frac{TP}{TP + FP}$$

其中, TP 表示被模型正确预测为异常的样本, FP 表示被模型错误预测为异常的正常样本。

召回率表示实际异常样本中,模型能够正确识别出来的比例。

$$\text{Recall} = \frac{TP}{TP + FN}$$

其中, TP 表示被模型正确预测为异常的样本, FN 表示被模型错误预测为正常的异常样本。

4.3 对比实验

表2—表6列举了本文方法在5个主流数据集上同6个主流模型之间的性能对比。不难看出,TS-KAN都取得了最好的结果。本文方法相较于次优方法MEMTO,在SMD,MSL和SWaT数据集上具有较为明显的性能提升(F1分数分别提升了1.75个百分点、0.95个百分点和0.87个百分点)。而在SMAP和MSL数据集上,本文方法虽有所提升,但提升不明显(F1分数分别提升0.23个百分点和0.33个百分点)。本文猜测这可能是数据集固有问题导致的,即“数据集的质量和有效性受到一些缺陷的影响,例如平凡的异常、错误的标注以及异常密度不切实际等^[31,36]。在这些数据集质量不佳的场景下,TS-KAN与MEMTO显著优于其他方法,原因可能是两者包含的记忆模块区分出了这些数据集中的平凡异常。在PSM和SMAP数据集上,受异常阈值大小的影响,TS-KAN在PSM和SMAP数据集上的精确率略低于最优方法,但在召回率上远超次优方法。从综合性评价指标F1分数上看,TS-KAN在5个主流数据集上均为最优,充分说明了其有效性。

表2 SMD数据集上的对比结果

Table 2 Comparison results on SMD dataset

| Methods | P ↑ | R ↑ | F1 ↑ |
|--------------|--------------|--------------|--------------|
| LSTM-VAE | 75.76 | 90.08 | 82.30 |
| BeatGAN | 72.90 | 84.09 | 78.10 |
| OmniAnomaly | 83.68 | 86.82 | 85.22 |
| InterFusion | 87.02 | 85.43 | 86.22 |
| A. T. | 87.96 | 94.68 | 91.20 |
| MEMTO | 89.13 | 98.40 | 93.54 |
| UMTS-Mixer | 86.81 | 88.16 | 87.48 |
| WMAD | 87.68 | 79.19 | 83.22 |
| TS-KAN(Ours) | 91.95 | 98.89 | 95.29 |

表3 MSL数据集上的对比结果

Table 3 Comparison results on MSL dataset

| Methods | $P \uparrow$ | $R \uparrow$ | $F1 \uparrow$ |
|--------------|--------------|--------------|---------------|
| LSTM-VAE | 85.49 | 79.94 | 82.62 |
| BeatGAN | 89.75 | 85.42 | 87.53 |
| OmniAnomaly | 89.02 | 86.37 | 87.67 |
| InterFusion | 81.28 | 92.70 | 86.62 |
| A. T. | 91.13 | 90.12 | 90.63 |
| MEMTO | 92.07 | 96.76 | 94.36 |
| UMTS-Mixer | 91.15 | 91.55 | 91.35 |
| WMAD | 86.77 | 85.02 | 85.88 |
| TS-KAN(Ours) | 93.19 | 97.54 | 95.31 |

表4 SMAP数据集上的对比结果

Table 4 Comparison results on SMAP dataset

| Methods | $P \uparrow$ | $R \uparrow$ | $F1 \uparrow$ |
|--------------|--------------|--------------|---------------|
| LSTM-VAE | 92.20 | 67.75 | 78.10 |
| BeatGAN | 92.38 | 55.85 | 69.61 |
| OmniAnomaly | 92.49 | 81.99 | 86.92 |
| InterFusion | 89.77 | 88.52 | 89.14 |
| A. T. | 93.96 | 98.45 | 96.15 |
| MEMTO | 93.76 | 99.63 | 96.61 |
| UMTS-Mixer | 95.91 | 69.98 | 80.92 |
| WMAD | 89.78 | 62.49 | 73.68 |
| TS-KAN(Ours) | 94.01 | 99.85 | 96.84 |

表5 SWaT数据集上的对比结果

Table 5 Comparison results on SWaT dataset

| Methods | $P \uparrow$ | $R \uparrow$ | $F1 \uparrow$ |
|--------------|--------------|---------------|---------------|
| LSTM-VAE | 76.00 | 89.50 | 82.20 |
| BeatGAN | 64.01 | 87.46 | 73.92 |
| OmniAnomaly | 81.42 | 84.30 | 82.83 |
| InterFusion | 80.59 | 85.58 | 83.01 |
| A. T. | 85.85 | 100.00 | 92.39 |
| MEMTO | 94.18 | 97.54 | 95.83 |
| WMAD | 92.20 | 93.26 | 92.73 |
| TS-KAN(Ours) | 95.57 | 97.86 | 96.70 |

表6 PSM数据集上的对比结果

Table 6 Comparison results on PSM dataset

| Methods | $P \uparrow$ | $R \uparrow$ | $F1 \uparrow$ |
|--------------|--------------|--------------|---------------|
| LSTM-VAE | 73.62 | 89.92 | 80.96 |
| BeatGAN | 90.30 | 93.84 | 92.04 |
| OmniAnomaly | 88.39 | 74.46 | 80.83 |
| InterFusion | 83.61 | 83.45 | 83.52 |
| A. T. | 96.81 | 98.63 | 97.71 |
| MEMTO | 97.46 | 99.23 | 98.34 |
| UMTS-Mixer | 97.63 | 88.66 | 92.93 |
| WMAD | 98.65 | 92.95 | 95.72 |
| TS-KAN(Ours) | 98.00 | 99.34 | 98.67 |

4.4 消融实验

本文方法的特点在于局部特征增强层 Local-KAN、时间维度与多元隐变量维度的特征融合机制以及基于互注意力机制的记忆模块的设计。为验证上述模块设计的必要性和有效性,在5个主流数据集上对本文方法进行了消融实验。

4.4.1 特征增强层

为了验证特征增强层 Local-KAN 的有效性,提出了3个变体方法,分别是采用一维卷积层(Conv)、全连接层(MLP)

代替 Local-KAN 进行特征增强的模型,以及不进行特征增强(None)。表7列出了3种基线方法同 TS-KAN 在5个数据集上的 F1 分数。表8列出了不同时间窗口 τ 下,Local-KAN 的实验结果。

表7 Local-KAN 的消融实验

Table 7 Ablation study of Local-KAN

| methods | SMD | MSL | SMAP | SWaT | PSM |
|---------|--------------|--------------|--------------|--------------|--------------|
| +CNN | 94.90 | 94.35 | 96.85 | 96.15 | 98.15 |
| +MLP | 93.67 | 90.13 | 96.41 | 94.71 | 97.98 |
| None | 93.35 | 89.57 | 96.23 | 93.47 | 97.65 |
| TS-KAN | 95.29 | 94.95 | 97.21 | 96.70 | 98.67 |

表8 时间窗口的消融实验

Table 8 Ablation study of window size

| τ | SMD | MSL | SMAP | SWaT | PSM |
|--------|--------------|--------------|--------------|--------------|--------------|
| 5 | 94.37 | 92.57 | 96.83 | 96.67 | 97.95 |
| 10 | 95.29 | 94.95 | 97.21 | 96.70 | 98.67 |
| 15 | 94.95 | 94.67 | 96.77 | 96.75 | 98.55 |

实验表明,所提方法相较于不进行特征增强(None)或使用 MLP 进行特征增强的模型,能够大幅度提升模型的表现,原因在于局部特征增强能够帮助模型更好地建模局部上下文关系,从而更好地检测出上下文异常。Local-KAN 比 CNN 具有更好的表现,原因在于 Local-KAN 基于 KA 表示理论,因此其比 CNN 具有更好的拟合能力,能够编码更丰富的局部特征模式。在 SMD,MSL,SMAP 和 PSM 数据集上,时间窗口为 10 的 TS-KAN 具有最好的表现。而在 SWaT 数据集上,时间窗口为 15 的 TS-KAN 模型具有最好的表现,但不同时间窗口下的表现差异不大。因此,本文将模型的时间窗口统一设置为 10。

4.4.2 特征融合机制

为了验证所提特征融合机制的有效性,本文提出了3个变体融合方法,分别为仅采用多变量特征融合(MLSA)、仅采用时域特征融合(TSA)和采用普通多层感知机取代 KAN 进行特征整合(MLP)。表9列出了3种基线方法同 TS-KAN 在5个数据集上的 F1 分数。

表9 特征融合机制的消融实验

Table 9 Ablation study of feature fusion

| Methods | SMD | MSL | SMAP | SWaT | PSM |
|----------|--------------|--------------|--------------|--------------|--------------|
| 仅采用 MLSA | 93.81 | 93.40 | 95.78 | 95.19 | 97.15 |
| 仅采用 TSA | 93.89 | 93.53 | 95.69 | 95.30 | 97.10 |
| 采用 MLP | 94.51 | 94.32 | 96.41 | 96.16 | 98.21 |
| TS-KAN | 95.29 | 94.95 | 97.21 | 96.70 | 98.67 |

实验结果表明,仅采用多变量特征融合(MLSA)、仅采用时域特征融合(TSA)和采用 MLP 进行特征整合都会导致 F1 分数下降,说明所提特征融合机制能够从多变量、时域两个方面进行特征融合并利用 KAN 将其有效地整合,验证了所提方法的有效性。

4.4.3 记忆模块

为了验证所提记忆模块的有效性,本文提出了两个变体方法,分别为不使用记忆模块和使用基于注意力机制的记忆模块(Attn.)^[16]。表10列出了两种基线方法同 TS-KAN 在5个数据集上的 F1 分数。

表 10 基于困难样本增强的记忆模块的消融实验

Table 10 Ablation study of KAN-based memory module

| methods | SMD | MSL | SMAP | SWaT | PSM |
|---------|--------------|--------------|--------------|--------------|--------------|
| None | 89.50 | 88.76 | 90.75 | 89.10 | 91.34 |
| Attn. | 93.10 | 94.40 | 96.71 | 95.40 | 98.25 |
| TS-KAN | 95.29 | 94.95 | 97.21 | 96.70 | 98.67 |

实验结果表明,两种记忆模块相较于没有记忆模块的变体,都有大幅度的性能提升,而基于困难样本增强的记忆模块比基于注意力机制的记忆模块具有更好的表现,充分说明了所提记忆模块的有效性。

4.5 可视化分析

如图 3 所示,本文还在 PSM 数据集上对异常得分进行了可视化分析。可以发现,所提方法在异常样本上的异常得分显著高于正常样本,能够很好地区分正常样本和异常样本。

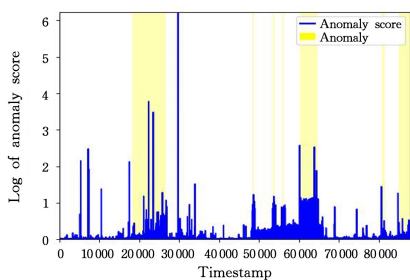


图 3 PSM 上的可视化分析

Fig. 3 Visualization on PSM

结束语 本文提出了一种基于KAN的时间序列异常检测网络 TS-KAN。首先利用所提的 Local-KAN 进行局部特征融合;然后从多变量和时域两个方面进行特征融合,从而得到更具辨别性的融合特征;最后利用基于互注意力机制的外置记忆模块记住大量正常样本特征,从而更好地区分异常样本和正常样本。对比实验结果表明,TS-KAN 在 5 个主流时间序列异常检测数据集上均取得了最优成绩。本文从特征增强层、特征融合机制、外置记忆单元这 3 个方面开展了消融实验,充分证明了所提方法的有效性。

参考文献

- ZONG B, SONG Q, MIN MARTIN R Q, et al. Deep autoencoding gaussian mixture model for unsupervised anomaly detection [C]// International Conference on Learning Representations. 2018.
- 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.
- ZHOU Q H, HE S B, LIU H Y, et al. Label-free multivariate time series anomaly detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(7):3166-3179.
- FREHNER R B, WU K S, SIM A, et al. Detecting Anomalies in Time Series Using Kernel Density Approaches[J]. IEEE Access, 2024, 12:33420-33439.
- 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.
- SHIN Y J, 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. 2020:2733-2740.
- DONG C, TAO J F, CHAO Q, et al. Subsequence time series clustering-based unsupervised approach for anomaly detection of axial piston pumps[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72:1-12.
- 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.
- TARIQ S, LEE S, SHIN Y J, 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. 2019:2123-2133.
- PARK D, HOSHI Y, KEMP CHARLES 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.
- SU Y, ZHAO Y J, NIU C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019:2828-2837.
- LI Z H, ZHAO Y J, HAN J Q, et al. Multivariate time series anomaly detection and interpretation using hierarchical intermetric and temporal embedding[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021:3220-3230.
- XU J, WU H, WANG J, et al. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy[C]// International Conference on Learning Representations. 2022.
- SUN Y Y, CHEN Z D, FENG C, et al. UMTS-Mixer: Time Series Anomaly Detection Based on Temporal Correlation and Channel Correlation[J]. Computer Systems and Applications, 2024, 33(1):127-133.
- YE L, HE Z. Multiscale time series anomaly detection incorporating wavelet decomposition[J]. Journal of Computer Applications, 2024, 44(10):3300-3306.
- GONG D, LIU L Q, LE V, et al. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:1705-1714.
- PARK H J, NOH J, HAM B. Learning memory-guided normality for anomaly detection[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:14372-14381.
- SONG J, KIM K, OH J, et al. MEMTO: Memory-guided transformer for multivariate time series anomaly detection[J]. arXiv:

- 2312.02530,2023.
- [19] LIU Z M, WANG Y X, VAIDYA S, et al. KAN: Kolmogorov-arnold networks[J]. arXiv:2404.19756,2024.
- [20] LIU Z M, MA P C, WANG Y X, et al. KAN 2.0: Kolmogorov-Arnold Networks Meet Science[J]. arXiv:2408.10205,2024.
- [21] SIDHARTH S S. Chebyshev polynomial-based kolmogorov-arnold networks: An efficient architecture for nonlinear function approximation[J]. arXiv:2405.07200,2024.
- [22] AGHAEI A A. fKAN: Fractional Kolmogorov-Arnold Networks with trainable Jacobi basis functions[J]. arXiv:2406.07456,2024.
- [23] BOZORGASL Z, CHEN H. Wav-KAN: Wavelet kolmogorov-arnold networks[J]. arXiv:2405.12832,2024.
- [24] LI C X, LIU X Y, LI W Y, et al. U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation[J]. arXiv:2406.02918,2024.
- [25] VACA-RUBIO C J, BLANCO L, PEREIRA R, et al. Kolmogorov-arnold networks (kans) for time series analysis[J]. arXiv:2405.08790,2024.
- [26] GENET R, INZIRILLO H. TKAN: Temporal Kolmogorov-Arnold Networks[J]. arXiv:2405.07344,2024.
- [27] SONODA S, MURATA N. Neural network with unbounded activation functions is universal approximator[J]. Applied and Computational Harmonic Analysis, 2017, 43(2): 233-268.
- [28] LAI M J, SHEN Z M. The kolmogorov superposition theorem can break the curse of dimensionality when approximating high dimensional functions[J]. arXiv:2112.09963,2021.
- [29] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. Neural networks, 2018, 107: 3-11.
- [30] SUBAKAN C, RAVANELLI M, CORNELL S, et al. Attention is all you need in speech separation[C]// 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2021:21-25.
- [31] SARFRAZ M S, CHEN M Y, LAYER L, et al. Position: Quo Vadis, Unsupervised Time Series Anomaly Detection? [C]// Proceedings of the 41st International Conference on Machine Learning. 2024:43461-43476.
- [32] SU Y, ZHAO Y J, NIU C H, et al. Robust anomaly detection for multivariate time series through stochastic recurrent neural network[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 2828-2837.
- [33] 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.
- [34] LI D, CHEN D C, JIN B H, et al. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks[C]// International Conference on Artificial Neural Networks. 2019:703-716.
- [35] ABDULAAL A, LIU Z H, 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] WU R, KEOGH E J. Current time series anomaly detection benchmarks are flawed and are creating the illusion of progress [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(3): 2421-2429.



WANG Cheng, born in 1985, Ph.D candidate. His main research interest is time series anomaly detection.



JIN Cheng, born in 1978, Ph.D, professor, Ph.D supervisor, is a member of CCF (No. 08763S). His main research interests include computer vision and multimedia information retrieval.

(责任编辑:何杨)