

引文格式:王明磊,王莉. 基于时序分解的时序数据通用表征学习[J]. 太原理工大学学报, 2026, 57(2): 365-377.

WANG Minglei, WANG Li. Universal time series representation learning based on temporal data decomposition [J]. Journal of Taiyuan University of Technology, 2026, 57(2): 365-377.

基于时序分解的时序数据通用表征学习

王明磊, 王莉

(太原理工大学 计算机科学与技术学院(大数据学院), 山西 晋中)

摘要:【目的】时间序列在金融、医疗、工业和气象等多个领域中具有广泛的应用价值。时间序列具有强时间依赖性。建模时间序列的多尺度特性往往忽视了数据中的细粒度结构。为了解决这些问题,提出了一种基于时序分解的通用表示对比学习框架(STURL)。【方法】通过对季节项进行层次化对比学习,能够在多尺度和细粒度层面上有效捕捉时间序列的多维特征。此外,还引入了频率混合策略,通过傅里叶变换增强季节项的表示,进一步提高了模型的鲁棒性和稳定性。【结果】实验结果显示,STURL在各下游任务上展现出了卓越的性能,并超越了基线方法。这些结果表明,该框架在提高模型准确性、减少预测误差以及增强异常检测能力方面的优越性,显示了其在实际应用中的潜力和可行性。

关键词:时间序列;表征学习;时序分解;对比学习;多任务一致性;时序数据通用表征

中图分类号:TP183 **文献标识码:**A

DOI: 10.16355/j.tyut.1007-9432.20250192

文章编号: 1007-9432(2026)02-0365-13

Universal Time Series Representation Learning Based on Temporal Data Decomposition

WANG Minglei, WANG Li

(College of Computer Science and Technology (College of Data Science), Taiyuan University of Technology, Jinzhong, Shanxi, China)

Abstract: 【Purposes】Time series data are widely used in fields such as finance, healthcare, industry, and meteorology. Existing methods often focus on instance-level learning, with limited success in modeling the multi-scale and fine-grained characteristics of time series. Moreover, many methods rely on inductive biases from computer vision or natural language processing, which are unsuitable for time series due to their strong time dependencies. This paper proposes a time-series decomposition-based contrast learning framework (Seasonal-Trend Universal Time Series Representation Learning, STURL) that addresses these challenges by leveraging hierarchical contrast learning of seasonal terms to capture multi-dimensional features. 【Methods】Additionally, a frequency mixing strategy based on Fourier transform enhances the representation of seasonal terms, improve model robustness and stability. 【Results】The experimental results demonstrate that STURL exhibits outstanding performance across various downstream tasks and surpasses baseline methods. These findings highlight the framework's superiority in improving model accuracy, reducing prediction errors, and enhancing anomaly detection capabilities, underscoring its potential and feasibility for practical ap-

收稿日期: 2025-03-25; 修回日期: 2025-05-12

基金项目: 国家重点研发计划(2021YFB3300503); 国家自然科学基金区域创新发展联合基金(U22A20167)

第一作者: 王明磊, 硕士, (E-mail) m15035399943@163.com

通信作者: 王莉, 博士, 教授, 主要从事数据挖掘、人工智能等研究, (E-mail) wangli@tyut.edu.cn

plications.

Keywords: time series; representation learning; temporal decomposition; comparative learning; multi-task consistency; generic representation of temporal data

时序数据广泛存在于多个领域,如气象^[1]、医疗^[2]、工业监控^[3]和金融^[4]等,这些领域中的数据通常是高维的、非平稳的,并且有多种不同的波动模式和季节性变化。这些复杂性使得时序数据分析和表示学习变得尤为重要。传统的时序数据分析方法(如基于ARIMA^[5]模型、滑动平均^[6]等统计方法)无法处理这些复杂的特性,特别是他们不能有效地捕捉到数据的多尺度特性和长期趋势。与此同时,深度学习方法虽然能够自动提取特征,但往往在处理具有强噪声、非平稳或局部特征的数据时会出现问题。现有的时序数据表示学习方法大多集中在实例级别的学习,即学习整个时间序列的表示。然而,这种方法通常忽视了数据中细粒度的局部信息和多尺度特征,这对处理一些特定任务(如预测、异常检测)是非常重要的。另外,传统的深度学习模型通常依赖于计算机视觉或自然语言处理领域的归纳偏置(如变换不变性、裁剪不变性等),这些策略难以适应时序数据的时间依赖性和动态特征,如现有的数据增强方法中的翻转和置换,常引入不适当的归纳偏差,破坏了时序数据的固有规律。

为了应对时序通用表征学习面临的一系列挑战,本研究提出了一种基于时序分解的通用表示对比学习框架(STURL)。该方法受到序列分解和对比学习思想的启发,将时间序列分解为季节项和趋势项两部分,其中季节项捕捉时间序列的周期性,趋势项则反映时间序列的趋势性。分别对每一部分训练对应的编码器,并将其聚合成最终的通用表示,能够生成更高质量的时间序列表示。该方法主要贡献如下:

1)提出了一种基于时序分解的对比学习框架。通过时序数据分解,STURL能够将时间序列分解为季节项和趋势项,并提取各自的特征,最后对其叠加以得到最终的通用表示。对于季节项,该框架引入了3个维度的对比学习训练编解码器,通过实例级、时间级和分层对比学习,以得到更好的表示。

2)提出了一种基于季节项S的数据增强方法,主要通过引入傅里叶变换(Fourier Transform)将原始数据转移到频率域进行处理。这一方法的核心思想是对数据进行频域变换,然后只替换频率中的

高频部分,保留低频成分。这种策略能够有效地增强模型对于时序数据中的趋势性、周期性以及其他重要特征的鲁棒性,尤其是在处理高噪声数据和非平稳时间序列时,具有显著的优势。

3)通过大量的实验验证了所提出模型(STURL)在时序数据通用表征方面的优势。利用四个公开数据集进行评估,并与当前主流方法进行了对比。实验结果显示,STURL在各下游任务上展现出了卓越的性能,并超越了基线方法。这些结果表明,该框架在提高模型准确性、减少预测误差以及增强异常检测能力方面的优越性,显示了其在实际应用中的潜力和可行性。

1 相关工作

基于一致性的通用表示学习方法通常通过引入不同的对比学习策略,以确保学习到的表示能够有效地捕捉时间依赖性、上下文一致性、迁移一致性以及多尺度特征等关键信息。根据不同的侧重点,这些方法可以分为多个类别,包括基于时间一致性、上下文一致性、迁移一致性、分层和跨尺度一致性等方向的研究工作。

1.1 基于时间一致性的通用表示学习

TNC^[7]利用时间邻域作为ADF测试的阳性样本,并通过调整样本权重来缓解采样偏差。EL-DELE et al^[8]提出了一种新的时间对比模块,通过设计硬交叉视图预测任务来学习时间依赖性,强迫模型通过更难的预测任务来学习鲁棒的表示。HA-JIMORADLOU et al^[9]引入了沿时间和实例维度的相似性蒸馏来进行预训练。YANG et al^[10]提出了TimeCLR,通过最小化样本的两个增强视图之间的相似性来学习不变表示。WANG et al^[11]结合FC图构造与移动池GNN,能够学习时间序列的空间和时间特征,确保时间一致性并考虑空间相关性,这对下游任务至关重要。MORGADO et al^[12]使用视听空间对齐作为360°视频数据的接口任务,涉及空间错位的音频和视频剪辑。HARESH et al^[13]则通过联合使用时间正则化项来鼓励不同帧映射到嵌入空间的不同点。CHEN等人^[14]提出了序列对比损失,利用重叠子序列进行正负样本采样。最近的TempCLR研究^[15]通过DTW技术在视频段落对齐

中探索时间动态,并通过基于时间粒度的新型负采样策略更有效地捕捉时间动态。

这些方法虽然在不同的任务中取得了不同的成功,但它们通常侧重于单一方面的时间特性,忽略了更为复杂的多维时间依赖性,这也是本研究中所着力解决的问题。

1.2 基于上下文一致性的通用表示学习

TimeAutoML^[16]通过AutoML框架实现自动配置和超参数优化,利用在实例范围内引入随机噪声来创建负样本。CARL^[14]通过对齐增强视频视图之间的序列相似性,采用序列对比损失,保持帧间的时间连贯性和上下文一致性。这使得特征表示在视频长度和内容变化时仍具备鲁棒性,与本研究关注的时间一致性问题高度相关。TS2Vec^[17]利用随机重叠片段捕获多尺度上下文信息,采用时间和实例对比损失,在捕捉时间依赖性方面表现优异。ZHANG et al^[18]提出的TS CoT算法通过联合训练增强不同视图表示的全局一致性,对本研究的多维时间依赖性建模具有启发意义。SHIN et al^[20]提出的基于重叠窗口的一致性正则化框架,以软标签作为共享优化目标,能在一定程度上提升对时间一致性的捕捉能力。

这些方法针对不同任务均取得了进展,但普遍缺乏对多时间尺度和跨任务一致性建模的综合考量,这正是本研究着重优化的方向。

1.3 基于分层和跨尺度一致性的通用表示学习

YUE et al^[17]提出了TS2Vec,该方法通过上下文一致性学习,在随机重叠的子序列中,采用两种对比策略来计算时间和实例对比损失。与其他方法不同,TS2Vec通过层次化方式在增强的上下文视图上开展对比学习,为每个时间戳提供鲁棒的上下文表示。该方法能通过相应时间戳上的最大池化生成全局特征表示,并能够有效捕捉多尺度上下文信息。NGUYEN et al^[25]设计了一种融合层次损失和三重损失的新型损失函数,用于增强时间序列表示的多粒度特性。CSL^[26]则通过学习多粒度对比和多尺度对齐的时间序列表示来捕捉跨时间范围的依赖信息。COMET^[27]结合医学时间序列的四级对比损失,通过对每个损失引入超参数系数,权衡多层次对比损失,从而获得更稳健和全面的特征表示。DUAN et al^[28]利用分层结构捕捉不同频率的时间序列特征,并通过强制各组间的一致性约束,生成在不同频率下都有效且鲁棒表示。

这些方法为时间序列表示的多尺度学习提供

了多种有效策略,但普遍聚焦于单一尺度或频率的建模,忽略了更复杂的多维度时间依赖性,这也是本研究的核心创新之一。

1.4 基于多任务一致性的通用表示学习

FEAT^[29]通过分层时间对比、特征对比和重建损失的联合学习,实现特征一致性与时间一致性的协同优化。CHOI和KANG^[30]引入了不确定性加权方法,通过建模多个任务的同方差不确定性,权衡多个对比损失,增强时间和转换一致性的建模能力。FOCAL^[31]强调模态一致性和变换一致性,学习跨模态共享特征;同时考虑局部性零星偏差,通过松散排名损失对样本对施加约束。ZHANG et al^[32]提出时频一致性建模,认为同一时间序列中基于时间和频率的表示应该在时频潜在空间中更接近,从而引入一致性损失以最小化嵌入之间的距离。TimesURL^[33]通过构造双优解并引入对比学习联合优化目标,捕捉分段级和实例级信息。LEE et al^[34]提出SoftCLT,通过软对比损失实现细致的学习,基于时间序列实例之间的距离进行实例对比,重点处理时间戳差异。LI et al^[35]揭示了时间序列中表示偏差与光谱距离的正相关关系,并通过优化对比损失,减少数据增强中的偏差,提高了嵌入的稳健性和普适性。

这些方法提供了多种有益的策略来增强时间一致性建模,但大多侧重于单一维度的对比学习,未能充分考虑多层次和多尺度的时间依赖性,这也是本研究的核心贡献之一。

2 相关定义

定义1 问题定义

给定一个时间序列集合 $X = \{x_1, x_2, \dots, x_n\}$,每个序列 $x_i \in \mathbb{R}^T$,长度为 T ,目标是学习一个表征函数 $f(x_i; \theta)$,将序列 x_i 映射到潜在空间中的特征向量 $h_i = f(x_i; \theta)$ 。

定义2 时序数据分解。设输入时间序列为 $x_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,T}\}$,其中 T 表示时间序列长度。在数据预处理阶段会将 x_i 分解为两部分:时间序列中的季节项 x_i^S 和趋势项 x_i^T 。分解的数学形式为:

$$x_i = x_i^S + x_i^T. \quad (1)$$

定义3 表示向量。在两部分编码阶段,会得到季节项表征 h_i^S 和趋势项表征 h_i^T ,最终的表示的数学形式为:

$$h_i = h_i^S + h_i^T. \quad (2)$$

3 方法

在本节中,提出了一种基于时序分解的对比学习框架 STURL,该模型旨在通过解耦时间序列的趋势和季节性成分,学习更加鲁棒的表征,充分挖掘时间序列的多尺度季节性特征,生成适配多任务的通用表征。

3.1 STURL 模型

STURL 该模型结构由多个关键模块组成,每个模块采用了精细化分解方法,以有效捕获时间序列中的多种特征。输入的时间序列数据首先经过时序分解模块,被分解为季节项 x_i^S 和趋势项 x_i^T ,其

中季节项通过频率混合策略处理,趋势项通过简单移动平均操作平滑,该分解方法遵循目前主流的分解框架^[43]。通过这些操作,数据实现了有效的平滑和频域转换,为后续的建模过程提供了支持。在编码阶段,模型通过时序编码器(SEncoder)处理季节性特征 h_i^S 并通过时间编码器(TEncoder)提取趋势特征 h_i^T 。随后,对比学习模块对这些特征进行进一步学习,生成高质量的特征表示,最终得到通用表示 $\bar{h}$ 。在解码阶段,时序解码器(TDecoder)会根据时序特征进行预测,同时采用基于 Kullback-Leibler 散度的优化策略来调整模型输出,使得生成的时间序列具有更好的拟合性。

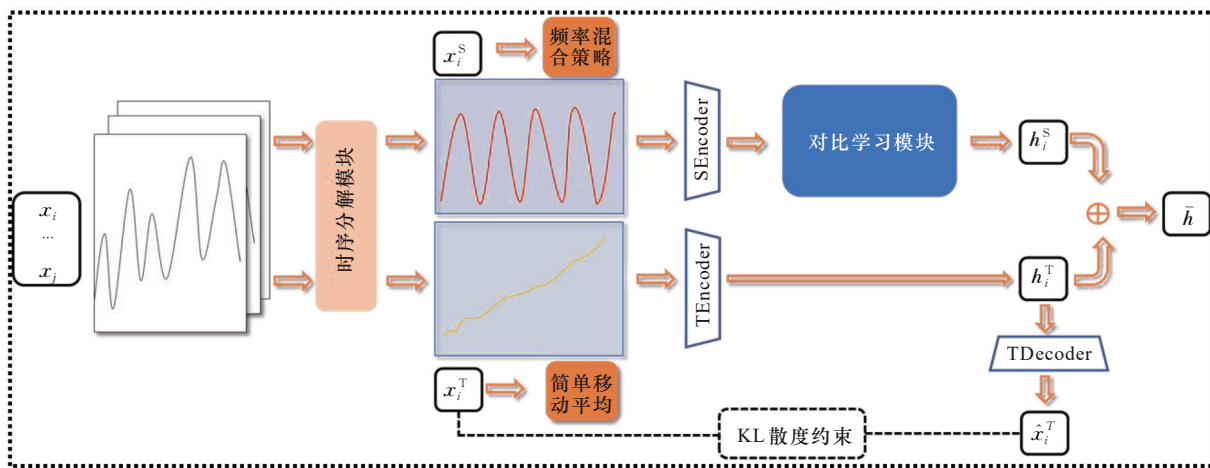


图1 STURL 模型结构图

Fig.1 STURL structure

3.2 频率混合策略

季节项数据的增强基于快速傅里叶变换(FFT)。如图2所示,对输入序列 x_i 和另一端输入序列 x_k 进行快速傅里叶变换(FFT),得到其频域表示,其中,傅里叶变换频段划分采用随机裁剪策略,与 TimesURL 模型的频段划分保持一致。从频域中选择高频部分与另一段序列 x_k 的高频部分进行替换混合,通过逆傅里叶变换(iFFT)以重构新的增强数据,同时保留信号的核心特征。

傅里叶变换是一种经典的数学工具,它能够时将域信号转换为频域信号,从而揭示出信号中的不同频率成分,且计算复杂度也仅为 $O(N\log N)$,其中 N 是数据点的数量。在时序数据中,通常可以将数据分解为不同频率的成分,其中低频成分代表了长期的趋势和周期性变化,而高频成分则通常与短期波动、噪声和细节信息有关。

对时序数据进行傅里叶变换后,通过设定预定义的频率阈值,将高频部分从原始信号中提取出

来,并与另一信号中的高频部分进行交换。替换比例是根据实验设计预先设定的,本实验中采用了30%的高频部分替换。本文明确定义了混合规则,即仅替换频率幅度超过阈值的部分,同时保持相位信息不变。该策略有助于保留时序信号的整体结构,同时引入跨信号的高频特征信息。

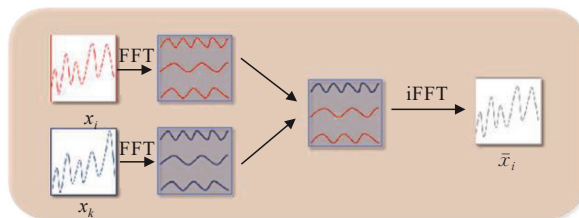


图2 频率混合策略结构图

Fig.2 Frequency-mixing strategy structure

3.3 对比学习模块

对比学习模块结构如图3所示,其主要目标是通过自监督学习优化模型的特征表示。在时间序列数据中,模型不仅需要学习数据的全局特征,还

需要捕捉局部特征的细节。通过构建正负样本对,模型能够学习到时间序列中不同部分(如不同时间步、不同时间尺度、不同实例)的相似性和差异性。对比学习模块的核心在于时间一致性对比学习、实例区分对比学习和多尺度池化与对比学习的结合。

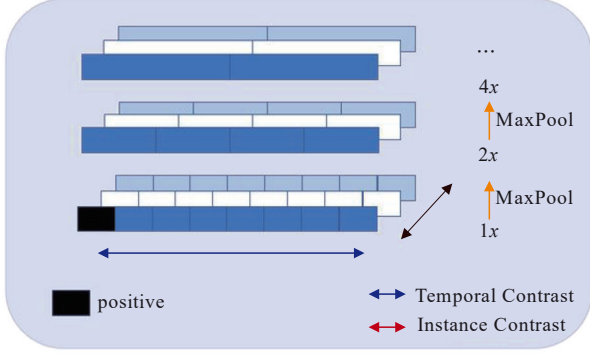


图3 对比学习模块结构图

Fig.3 Contrastive learning module structure

3.3.1 时间一致性对比学习

时间一致性对比学习旨在确保不同时间尺度的特征表示保持一致性,尤其是全局特征与局部特征间的对齐。对于同一时间序列的不同尺度特征 r_i, r_i' (例如全局特征和局部特征),目标是最大化它们的相似性,同时区分其他时间步的特征。

在多个基于对比学习的表示学习方法中,欧氏距离被广泛使用并验证了其有效性。例如,AMC-Loss^[41]和 SimOLoss^[42]中均采用了欧氏距离作为主要的相似度量,并取得了良好的效果。结合前人的研究成果,本文认为欧氏距离在本研究中是合理且有效的选择。在这里,我们使用欧氏距离作为相似度量。欧氏距离可以计算两个嵌入表示之间的距离,公式如下:

$$L(r_a, r_b) = \|r_a - r_b\|_2. \quad (3)$$

对于时间一致性对比学习,目标是最大化同一时间序列在不同时间步之间的特征表示相似性,同时最大化不同时间序列之间的特征表示差异性。损失函数 $\mathcal{L}_T$ 的表达式为:

$$\mathcal{L}_T = \sum_{t \neq t'} \|r_{i,t} - r_{i,t'}\|_2^2 + \sum_{t' \neq t} \sum_{j=1}^N 1_{t \neq t'} \|r_{i,t} - r_{j,t'}\|_2^2. \quad (4)$$

其中, $r_{i,t}$ 和 $r_{i,t'}$ 是时间序列 i 在时间步 t 和 t' 处的特征表示; $\|r_{i,t} - r_{i,t'}\|_2^2$ 是两个特征向量之间的欧氏距离的平方,用来衡量它们之间的差异。 $\sum_{j=1}^N \|r_{i,t} - r_{j,t'}\|_2^2$ 是对比不同时间步(负样本)的部分,确保同一时间序列中的不同时间步的特征表示之间的距离最小化,而与其他时间序列(负样本)的距离最大化。

3.3.2 实例区分对比学习

实例区分对比学习旨在区分不同时间序列实例的特征,确保同一时间步的特征在同一实例内保持相似性,而与其他时间序列实例的特征保持显著差异。同样使用欧氏距离来度量不同实例间特征表示的相似性,其目标是最大化同一时间步的正样本对之间的相似性,同时最小化它们与其他序列的特征的相似性。损失函数可以表示为:

$$\mathcal{L}_I = \sum_{i=1}^B \|r_{i,t} - r_{i,t'}\|_2^2 + \sum_{j \neq i} \|r_{i,t} - r_{j,t'}\|_2^2. \quad (5)$$

其中, $r_{i,t}$ 和 $r_{j,t'}$ 是来自不同序列 i 和 j 的特征表示, $i \neq j$ 。 $\|r_{i,t} - r_{i,t'}\|_2^2$ 是正样本对之间的欧氏距离,鼓励相同实例(时间序列)中相同时间步的特征表示尽量接近。 $\|r_{i,t} - r_{j,t'}\|_2^2$ 是负样本对之间的欧氏距离,确保来自不同时间序列的特征之间的距离较大。

这两种损失函数在功能上是互为补充的。例如,在处理来自多个地区的天气时序数据时,实例区分对比学习能够捕捉到不同地区间的特征差异,而时间一致性对比学习则专注于揭示随时间变化的动态模式。这种结合使得模型不仅能区分不同实例(如地区间的差异),还能够深入挖掘同一实例内随着时间推移的趋势变化。

3.3.3 多尺度池化与对比学习的结合

在模型中,时间序列数据经过多尺度池化操作后输入到对比学习模块。这意味着模型首先通过多尺度池化提取了时间序列的不同尺度特征(例如短期波动、长期趋势等),然后通过对比学习进一步优化这些特征的表示。具体的算法流程见算法1。

在模型复杂度方面,尽管这些技术在一定程度上增加了模型的计算开销,但其复杂度相较于同类任务中的其他深度学习模型并不算高。多尺度池化通过不同尺度的池化操作来捕捉不同时间粒度的特征,这一操作相较于全连接层或其他重参数化方法,其计算复杂度是可以控制的,且在实验中没有发现显著的计算瓶颈。

通过这种结合,模型能够从全局到局部、从粗到细地逐步学习和提取时间序列中的多层次、多尺度的特征,使得最终生成的特征表示更加全面且具有更高的区分度。

3.3.4 其他模块

1) SMA 模块。SMA 是一种基于历史数据的平滑方法,通过计算一个固定窗口内数据点的均值来

算法 1 多尺度池化与对比学习结合的算法流程

```

1. 初始化阶段:
   计算  $\mathcal{L}_{\text{总损失}}(r, r')$  并初始化为  $\mathcal{L}_{\text{多尺度池化}}$ 
   设置参数  $d = 1$ 
2. 逐层计算:
   while  $r > 1$  do
        $r = \text{maxpool1d}(r, \text{kernel\_size} = \{1, 2, 4, 8\})$ ;
        $r' = \text{maxpool1d}(r, \text{kernel\_size} = \{1, 2, 4, 8\})$ ;
        $\mathcal{L}_{\text{多尺度池化}} = \mathcal{L}_{\text{多尺度池化}} + \mathcal{L}_{\text{总损失}}(r, r')$ ;
3. 结束阶段:
   当序列长度缩小到 1 时, 循环结束。此时, 所有层的损失已经计算完毕;
    $\mathcal{L}_{\text{多尺度池化}} = \mathcal{L}_{\text{多尺度池化}} / d, d = d + 1$ 
4. 返回结果:
   返回最终的  $\mathcal{L}_{\text{多尺度池化}}$  作为输出。

```

减少随机波动。SMA 能够有效平滑短期的噪声, 使得序列中的长期趋势更加明显。对于时间序列表征学习, 长期趋势的学习比短期波动更为重要, 尤其是在需要捕捉趋势变化的任务中, 如股市预测或气象预测。

在时间序列数据中, 噪声通常来源于许多短期波动, 这些波动可能对模型的学习造成干扰。SMA 通过去除这些短期波动, 使得输入数据更加稳定, 从而提高模型的预测能力和鲁棒性。这有助于模型更好地理解数据的底层规律, 避免过度拟合短期波动。

在 STURL 模型数据预处理阶段, SMA (Simple Moving Average, 简单移动平均) 模块用于对分解后的趋势项数据进行平滑处理, 以减少噪声干扰, 从而更好地捕捉时间序列的趋势性特征。具体来说, SMA 模块通常用于去除时间序列中的短期波动, 使得模型能够更专注于长期趋势与周期性变化。

2) 季节编码器。在 STURL 模型中, SEncoder 专注于处理时序分解后的季节项, 其主要目标是提取季节性特征, 并将其映射为潜在空间中的有效表示。SEncoder 采用变分自编码器 (VAE)。该方法有助于将季节性成分 (作为潜在变量) 建模为连续的概率分布。

具体而言, VAE 通过引入隐变量和变分推断, 可以有效地学习时间序列的隐含结构。VAE 编码器输出的是季节项的潜在分布 (均值和方差), 而不是直接输出季节项的值。这种方法能在学习季节性特征的同时, 捕捉到不确定性和潜在模式, 尤其是当季节性变化具有复杂性时, VAE 能够帮助更好地建模这种不确定性。VAE 允许模型通过自监督学习方式训练, 从而在没有标签的情况下学习季节项的特征。同时, VAE 通过对潜在空间的正则化,

使得模型能够从数据中学习更通用的特征, 而不是简单地拟合训练数据。通过最大化变分下界 (ELBO), 模型能够自动从季节项中提取有用的特征表示。ELBO 公式如下:

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q(z|x)} [\log p(x|z)] - D_{\text{KL}}[q(z|x) \| p(z)]. \quad (6)$$

其中, $q(z|x)$ 是编码器的输出, 即给定输入 x 时潜在变量 z 的条件概率分布。 $p(x|z)$ 是解码器的输出, 即给定潜在变量 z 时, 生成数据 x 的条件概率。 $p(z)$ 是潜在变量的先验分布 (通常选择标准正态分布 $N(0, I)$)。 $D_{\text{KL}}[q(z|x) \| p(z)]$ 是 Kullback-Leibler 散度, 用于衡量编码器输出的潜在分布 $q(z|x)$ 与先验分布 $p(z)$ 之间的差异。

3) 趋势编码器。TEncoder 的输入是时序分解后的趋势项 ($x_1^T, \dots, x_i^T, \dots, x_n^T$), 这些趋势项由时序分解从原始时间序列中提取得到。每个趋势项包含不同时间尺度的变化, 捕捉了时间序列的长期趋势特征。

TEncoder 采用了多个自实现回归专家, 每个自回归专家基于一维因果卷积, 能够建模时间序列数据中的因果关系。每个专家输出的卷积结果具有不同的卷积核大小, 这使得每个专家能够专注于学习时间序列的不同尺度的趋势特征。

每个自回归专家 ($H(T, i)$) 会输出一个矩阵。随后, 通过对所有自回归专家输出的结果进行平均池化操作, 从多个不同时间尺度中提取出全局的趋势特征。这一步的作用是将多个专家的结果融合为统一的发展趋势表示, 从而得到最终的趋势项表示:

$$H(x_i^T) = \text{平均池化}(H(T, 0), H(T, 1), \dots, H(T, L)). \quad (7)$$

其中, $H(x_i^T)$ 就是 TEncoder 输出的最终趋势项表示。

4 实验结果与分析

为了评估 STURL 的可用性, 本文在 4 个下游任务上进行了广泛的实验, 并在不同下游任务选择不同的基线模型进行对比, 具体基线模型在本文 4.3 中介绍。实验环境为配备 NVIDIA RTX 3090 24 GB 显卡的服务器, 在 CUDA 11.1 下使用 Pytorch 1.9.1 库进行代码设计。常用超参数设置如下: 对比学习的温度参数为 1.0, 季节项表示和趋势项表示的维度大小设置为 64, 批量大小的取值为 $\{4, 8, 16, 32\}$, 学习率设定为 0.001。在实验中, 我

们测试了模型训练时间,结果表明 STURL 模型在计算资源的消耗上与现有的主流时序模型相当,能够在合理的时间范围内完成各个下游任务。

4.1 数据集与基线介绍

4.1.1 时间序列分类数据集与实验设置

由于分类和聚类任务都旨在识别时间序列样本所属的真实类别,现有的研究通常使用同一组基准数据集来评估模型性能。由不同应用领域的异构时间序列组成的精心设计的基准测试,如 UCR^[36] 和 UEA^[37],是目前使用最广泛的,因为它们可以全面评估所测模型的泛化能力。除了 DTW 方法之外,其他所有分类方法的表示维度都设定为 320,并且我们遵循与 TS2Vec 相同的协议,使用具有 RBF 核的 SVM 分类器对表示进行分类训练。

表 1 实验基准汇总
Table 1 Experiment benchmarks

任务	基准	评价指标
预测	ETT(4个子集)	MSE, MAE
插值	ETT(4个子集)	MSE, MAE
分类	UEA(30个子集)	Accuracy
	UCR(128个子集)	
异常检测	KPI 和 Yahoo	Precision, Recall, F1-Score

UCR 时间序列归档(UCR Time Series Classification Archive)是时间序列分类领域的经典基准数据集,包含 128 个单变量时间序列数据集,涵盖医疗、工业、运动等多个领域。每个数据集包含训练集和测试集,时间序列长度固定。类别数从 2 到 60 不等,适用于二分类和多分类任务。代表性数据集包括 ECG200(心电图分类)、GunPoint(手势识别)、Coffee(光谱分类)等。UEA 多变量时间序列分类归档(UEA Multivariate Time Series Classification Archive)是 UCR 的扩展,专注于多变量时间序列分类任务。包含 30 个多变量数据集,每个时间序列包含多个变量(通道)。应用场景包括动作识别(如 BasicMotions)、医疗监测(如 Libras)等。由于数据长度和变量数各异,该数据集具有更高的复杂性。

4.1.2 时间序列预测数据集与实验设置

由于预测任务的输出是数值序列,因此大多数研究对这项任务使用相同的基准数据集。常用的数据集来自多个应用领域和服务,本文使用电力数据集 ETT^[38], ETT 数据集的目标通常是预测电力变压器温度的变化,尤其是在多操作条件下的表现。ETT 数据集在很多时间序列预测任务中都被用作基准数据集,具有较高的挑战性,因为它涉及

到对长时间序列的预测以及复杂的时序动态。

4.1.3 时间序列异常检测数据集与实验设置

使用两个公共数据集来评估我们的模型。Yahoo 是一个用于异常检测的基准数据集,包括 367 个小时级采样的时间序列,且带有标记的异常点。涵盖了各种异常类型,如异常值和变化点。KPI^[39] 是 AIOPS Challenge 发布的竞赛数据集,包括多家互联网企业的多组精细采样的真实 KPI 曲线。

4.1.4 时间序列插值数据集与实验设置

在实际情况下,经常会出现不规则和异步采样,这可能导致数据缺失,从而给下游任务带来挑战。插补是缓解这一问题的一种简单且常用的方法。我们在电力场景中使用 ETT 数据集^[38]进行实验,该数据集普遍存在数据缺失问题,实验遵循与 TimesNet^[40]相同的设置,该模型使用 MLP 网络进行下游任务。

4.2 评价指标

针对时间序列分类任务,本文采用准确性、精确度、召回率和 F1 分数来评估分类任务性能,针对类别不平衡的情况额外使用精确召回曲线下的面积(AUPRC)作为评估指标。

关于时间序列预测与插值任务的评估指标,鉴于预测结果的数值性质,最常用的度量是均方误差(MSE)和平均绝对误差(MAE)。MAE 是一个直观的指标,它计算了预测值与实际值之间差异的绝对值的平均大小。MAE 的计算公式如下:

$$\text{MAE} = \frac{1}{T} \sum_{i=1}^T |y_i - \hat{y}_i|. \quad (8)$$

其中 $\hat{y}$ 是真实值, y 是模型输出的预测值, T 表示时间序列的长度。MAE 值越小表示预测结果与实际情况的偏差越小,预测的准确性越高。

MSE 则考虑了预测误差的平方,对较大误差赋予了更高的权重。MSE 的计算公式如下:

$$\text{MSE} = \frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2. \quad (9)$$

这里的符号定义与 MAE 相同。MSE 值越小,表示预测误差越小,模型的预测性能越好。

关于时间序列异常检测任务的评估指标,准确性、精确度、召回率和点调整 F1 分数是时间序列异常检测中使用最广泛的指标。

4.3 实验结果分析

4.3.1 时间序列插值

在时间序列插值实验中,基线方法包括 DLinear, TimesURL, InfoTS 和 TS2Vec,其中 TimesURL

是2024年时序数据的通用表征研究的SOTA结果，TS2Vec为2022年时序数据的通用表征研究的成果。本文提出的方法在多个数据集和不同缺失数据比例下展现了显著优于基线方法的性能。如表2

所示，实验结果表明，STURL模型在均方误差(MSE)和平均绝对误差(MAE)这两个关键评价指标上均具有较为明显的优势。

表 2 多变量时间序列插补结果
Table 2 Multivariate time series interpolation results

数据集	方法	STURL		DLinear		TimesURL		InfoTS		TS2Vec	
	屏蔽比例	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	0.125	0.647	0.638	1.518	2.674	0.659	0.640	0.717	0.666	0.690	0.658
	0.250	0.659	0.641	1.804	2.925	0.679	0.648	0.726	0.674	0.710	0.668
	0.375	0.698	0.652	2.150	3.180	0.702	0.656	0.726	0.676	0.728	0.676
	0.500	0.704	0.685	2.570	3.470	0.712	0.693	0.783	0.695	0.751	0.682
ETTh2	0.125	1.425	0.841	1.004	2.157	2.455	1.215	2.491	1.199	2.866	1.288
	0.250	1.198	0.898	1.269	2.470	2.560	1.239	2.644	1.244	2.792	1.271
	0.375	1.569	1.175	1.583	2.760	2.673	1.269	2.757	1.266	2.793	1.271
	0.500	1.782	1.897	1.834	2.990	2.701	1.281	2.844	1.283	2.769	1.267
ETTm1	0.125	0.563	0.684	0.580	1.620	0.644	0.650	0.702	0.651	0.726	0.663
	0.250	0.614	0.693	0.803	1.930	0.699	0.668	0.732	0.664	0.719	0.664
	0.375	0.671	0.715	1.030	2.190	0.718	0.686	0.753	0.674	0.728	0.670
	0.500	0.694	0.723	1.320	2.480	0.716	0.680	0.759	0.677	0.739	0.669
Avg.		0.935	0.854	1.455	2.571	1.327	0.860	1.386	0.864	1.418	0.871

在各个缺失数据比例(0.125,0.25,0.375)下，STURL模型的表现显著优于其他方法，尤其在较高缺失比率(0.375)情况下，STURL依然保持较低的MSE和MAE。这表明该方法能够有效应对时间序列数据中的缺失问题，展示了较强的鲁棒性和稳定性。本文方法相比次优模型TimesURL在所有数据集和屏蔽比例下的平均提升为17.06%。相比之下，其他对比方法(如TS2Vec、InfoTS等)在处理相同缺失数据比例时，虽然能够取得一定的性能，但其误差值普遍较高，表明这些方法可能未能充分利用时间序列的细粒度信息和多尺度特性。本文提出的STURL模型通过多尺度的特征提取和分层对比学习策略有效地捕捉了时间序列数据中的多尺度特征，从而在面对缺失数据时依然能够保持较高的准确性。尤其是在高缺失比率下，STURL模型能够较好地恢复缺失数据，并保持预测结果的稳定性。

4.3.2 时间序列分类

本文方法在UCR数据集上平均准确率和平均排名上均表现突出，如图4所示，表明该方法在处理复杂的时间序列数据时，能够更好地从数据中提取深层次の時序特征。相较于其他模型，STURL模型在特征提取和模型结构设计上具有更强的表征能力，从而在分类任务中取得更好的结果。与

LaST和TimesURL等方法相比，STURL模型在准确率上的微小优势(0.847对0.837/0.845)源于其在模型训练过程中对时间序列特征的更好学习和表示。虽然InfoTS和TS2Vec在准确率上接近Ours，但它们的平均排名(分别为2.047和2.625)表明其在多个数据集上的整体表现较差。

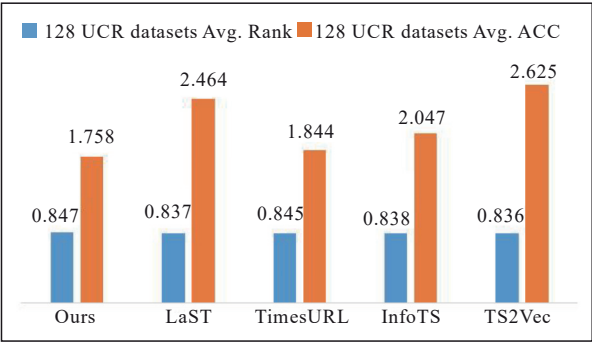


图 4 时间序列分类在UCR数据集上的表现
Fig.4 Performance of time series classification on UCR dataset

在图5中展示了UEA数据集(10个子集)上时间序列分类任务的实验结果，图表对比了多个模型在该任务中的分类准确率。图中的不同颜色代表了不同的模型，其中本论文模型的表现通过蓝色条形图显示。其他模型包括VSFormer、SVP-T、DA-Net和OS-CNN，其中VSFormer为2024年的最优模型。

本文所提方法在多项任务中取得高准确率,相较于 VSFormer 的平均提升为 1.80%,表明该方法在提取时序数据中的长期和短期依赖关系方面具有优越性。相比其他模型,本论文采用了更高效的网络架构和特征选择策略,使其能够更好地处理时间序列中的复杂性。

STURL 模型在 UEA 数据集上的时间序列分类任务中表现出色,远超其他对比模型,特别是在准确率上取得了显著优势。该模型的高效特征提取和优越的学习能力使其能够在多个任务中展现强大的性能。尽管 VSFormer 在部分数据集上具有一定优势,但本论文模型凭借其在多个数据集上的一致性和高效性,验证了其在时间序列分类任务中的广泛适用性和鲁棒性。

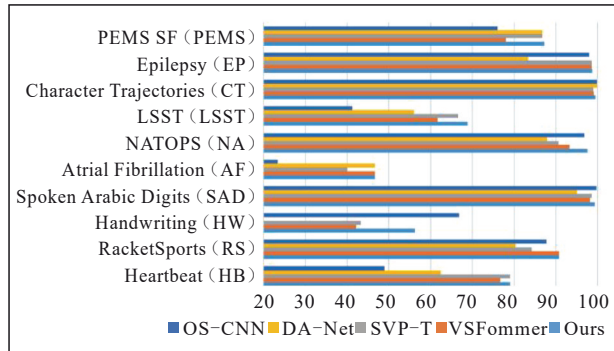


图 5 时间序列分类在 UEA 数据集上的表现
Fig.5 Performance of time series classification on UEA dataset

4.3.3 时间序列异常检测

在时间序列异常检测实验中,本文对比了常用的异常检测方法(包括 SPOT、DSPOT、DONUT、和 SR)的结果,同时对比了时序数据通用表征研究方法的結果。在异常检测实验结果中,表 3 展示了不同模型在 Yahoo 和 KPI 数据集上的表现,评估指标包括 F1 分数、精确度(Precision)和召回率(Recall)。实验结果表明,本文提出的模型(STURL)在所有关键指标(F1 分数、精确度和召回率)上均优于其他先进模型,尤其在 Yahoo 和 KPI 数据集上。相较次优模型 TimesURL,本文在 Yahoo 数据集上 F1 提升 0.27%, Precision 提升 1.20%, Recall 提升 0.27%;在 KPI 数据集上 F1 提升 0.58%, Precision 提升 0.65%, Recall 提升 1.47%。在保持良好的插值、分类和预测性能的同时,异常检测任务上也有略微的提升。这凸显了该方法在处理复杂数据集时在异常检测任务中的有效性以及表征的通用性。STURL 的优势主要源自其多尺度的特征建模、对比学习框架、上下文一致性机制、时间序列的分解

学习以及频域增强策略。这些创新的设计使得该模型能够在异常检测任务中表现出色,尤其是在复杂的数据集和多种异常类型下,能够有效提高模型的精确度和召回率。

表 3 时间序列异常检测 STURL 较其他基线的表现
Table 3 Time-series anomaly detection of STURL performance over other baselines

检测方法	Yahoo			KPI		
	F1	Prec.	Rec.	F1	Prec.	Rec.
SPOT	0.338	0.269	0.454	0.217	0.786	0.126
DSPOT	0.316	0.241	0.458	0.521	0.623	0.447
DONUT	0.026	0.013	0.825	0.347	0.371	0.326
SR	0.563	0.451	0.747	0.622	0.647	0.598
TS2Vec	0.745	0.729	0.762	0.677	0.929	0.533
TimesURL	0.749	0.748	0.750	0.688	0.925	0.546
Ours	0.751	0.757	0.752	0.692	0.931	0.554

4.3.4 时间序列预测

在时间序列预测实验中,本文对比了表征学习方法(其中 LaST 和 CoST 均为 2022 年优秀成果)的结果,同时对比了端到端方法的结果。由于预测任务的表格内容较多,我们将其拆分为表 4 和表 5。本文所提方法在多个任务中表现出较为优越的预测能力,尤其在 ETTh2 和 ETTm1 数据集上,MSE 和 MAE 表现均优于其他模型,特别是在 ETTh2 任务上,MSE 和 MAE 的数值显著低于其他方法,表明我们的模型在该数据集上的预测精度较高。TimesURL 作为对比模型,在 ETTh1 数据集的 MSE 降低了 14.99%,MAE 降低了 5.78%;ETTh2 数据集的 MSE 降低了 0.33%,MAE 降低了 9.35%;ETTm1 数据集的 MSE 降低了 48.14%,MAE 降低了 15.96%;Electricity 数据集的 MSE 降低了 19.40%,MAE 降低了 15.41%。总体来看,Ours 在所有数据集上的平均 MSE 降低了 20.72%,平均 MAE 降低为 11.63%。尽管在某些任务上 TimesURL 和 Autoformer 表现优于 STURL,但在其他任务和总体上,STURL 展示了其在长序列预测任务中的优势。

需要注意的是,预测步长从 24 增加到 720 时,无论是 STURL 还是其他模型,随着预测步长的增大,预测精度都会有所下降。因为在长时间步预测的情况下,模型面临的预测任务更加复杂,尤其是长期趋势、突变或周期扰动等因素的影响,导致预测精度自然下降。

实验结果表明,基于我们的时序数据分解与对比学习框架,提出的模型在时序数据预测任务中优

表 4 STURL与表征学习方法对比结果
Table 4 Comparison results between STURL and other representation learning methods

方法	STURL			TimesURL		LaST		CoST		TS2Vec	
	H	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	24	0.030	0.125	0.036	0.142	0.030	0.131	0.040	0.152	0.039	0.151
	48	0.050	0.158	0.054	0.146	0.051	0.169	0.060	0.186	0.062	0.189
	168	0.074	0.205	0.096	0.233	0.078	0.211	0.097	0.236	0.142	0.291
	336	0.098	0.232	0.121	0.267	0.100	0.246	0.112	0.258	0.160	0.316
	720	0.132	0.288	0.145	0.307	0.138	0.298	0.148	0.306	0.179	0.345
ETTh2	24	0.814	0.197	0.083	0.219	0.082	0.204	0.079	0.207	0.091	0.230
	48	0.117	0.237	0.116	0.219	0.126	0.275	0.118	0.259	0.124	0.274
	168	0.172	0.308	0.175	0.323	0.187	0.342	0.174	0.329	0.198	0.355
	336	0.198	0.356	0.188	0.347	0.229	0.389	0.206	0.364	0.205	0.364
	720	0.206	0.395	0.186	0.352	0.238	0.425	0.214	0.345	0.208	0.371
ETTm1	24	0.010	0.069	0.013	0.084	0.011	0.077	0.015	0.117	0.016	0.093
	48	0.019	0.099	0.024	0.177	0.021	0.108	0.025	0.167	0.028	0.126
	96	0.030	0.134	0.037	0.145	0.033	0.134	0.036	0.141	0.045	0.162
	288	0.069	0.197	0.080	0.214	0.069	0.197	0.081	0.207	0.095	0.235
	672	0.093	0.214	0.114	0.255	0.100	0.239	0.125	0.253	0.142	0.290
Electricity	24	0.148	0.262	0.250	0.284	0.151	0.277	0.243	0.264	0.260	0.288
	48	0.178	0.300	0.197	0.319	0.186	0.307	0.292	0.300	0.313	0.321
	168	0.227	0.336	0.259	0.371	0.243	0.346	0.405	0.375	0.429	0.392
	336	0.262	0.367	0.295	0.416	0.286	0.379	0.560	0.473	0.565	0.478
	720	0.305	0.414	0.337	0.452	0.322	0.422	0.889	0.645	0.863	0.651

表 5 STURL与端到端方法对比结果
Table 5 Comparison results between STURL and End-to-End methods

方法	STURL			Informer		Autoformer		N-BEATS		TCN	
	H	MSE	MAE	MAE	MSE	MSE	MAE	MSE	MAE	MSE	MAE
ETTh1	24	0.030	0.125	0.098	0.247	0.057	0.189	0.094	0.238	0.104	0.254
	48	0.050	0.158	0.158	0.319	0.07	0.207	0.210	0.367	0.206	0.366
	168	0.074	0.205	0.183	0.346	0.108	0.26	0.232	0.391	0.462	0.586
	336	0.098	0.232	0.222	0.387	0.119	0.281	0.232	0.388	0.422	0.564
	720	0.132	0.288	0.269	0.435	0.109	0.264	0.322	0.490	0.438	0.578
ETTh2	24	0.081	0.197	0.093	0.240	0.088	0.207	0.198	0.345	0.109	0.251
	48	0.117	0.237	0.155	0.314	0.121	0.330	0.234	0.386	0.147	0.302
	168	0.172	0.308	0.232	0.353	0.219	0.329	0.331	0.453	0.209	0.366
	336	0.198	0.356	0.263	0.417	0.236	0.327	0.437	0.508	0.237	0.391
	720	0.206	0.395	0.277	0.431	0.243	0.425	0.437	0.517	0.200	0.367
ETTm1	24	0.010	0.069	0.030	0.137	0.022	0.115	0.054	0.184	0.027	0.127
	48	0.019	0.099	0.069	0.203	0.032	0.138	0.103	0.361	0.034	0.154
	96	0.030	0.134	0.075	0.141	0.045	0.168	0.083	0.353	0.036	0.208
	288	0.069	0.197	0.401	0.554	0.071	0.207	0.360	0.362	0.353	0.455
	672	0.093	0.214	0.512	0.644	0.102	0.254	0.197	0.368	0.445	0.570
Electricity	24	0.148	0.262	0.251	0.275	0.290	0.411	0.243	0.367	0.243	0.367
	48	0.178	0.300	0.346	0.339	0.310	0.408	0.283	0.397	0.283	0.397
	168	0.227	0.336	0.544	0.424	0.435	0.490	0.357	0.449	0.357	0.449
	336	0.262	0.367	0.713	0.512	0.646	0.606	0.355	0.446	0.355	0.446
	720	0.305	0.414	1.182	0.806	0.609	0.587	0.387	0.477	0.387	0.477

于现有的主流方法,尤其在长时间跨度的预测中具有较好的鲁棒性和精度,展现了较强的实际应用潜力。

4.4 消融以及变体实验

在整篇论文中,我们强调了适当的频率混合、对比学习和季节趋势编码器三个策略,为了验证上述五个模块在 STURL 中的有效性,表中显示了 UEA 数据集中 30 个数据上的完整 STURL 及其六个变体之间的比较,其中 OV1 为替换 SMA 为 EMD 分解。结果表明,STURL 的上述所有组成部分都是不可或缺的。

对于四个去掉不同组件的实验,通过对表 6 中的数据进行分析,可以明显观察到完整模型的平均准确率为 0.748,这是包含所有策略和组件的基线模型表现。优于去除四个组件的结果。去除频率混合策略后,模型的准确率下降至 0.737,较完整模型减少了 0.011。这表明频率混合策略对模型的表现具有一定的提升作用,尤其是在处理数据中的频率成分时,频率混合策略有助于提升模型的准确性。

去除实例对比学习后,准确率降至 0.742,较完整模型降低了 0.006。这表明实例对比学习对于增强模型在实例层面上的表现有一定的作用,去除该策略后,模型的表现有所下降。

表 6 消融实验结果比较
Table 6 Ablation experiment results comparison

	Avg.Accuracy
STURL	0.748
w/o 频率混合策略	0.737
w/o 实例对比学习	0.742
w/o 时间对比学习	0.740
w/o 多尺度池化(SEncoder)	0.739
w/o VAE	0.742
OV1 SMA→EMD	0.735

去除时间对比学习后,准确率为 0.740,较完整模型下降了 0.008。这说明时间对比学习对于模型的时间序列建模有显著影响,去除后模型的表现略有下降。

去除多尺度池化(SEncoder)后,模型准确率降至 0.739,相比完整模型减少了 0.009。这表明多尺度池化在捕捉时间序列中不同尺度特征方面起到

了积极作用,去除该组件后,模型的准确度有所降低。

去除 VAE 后,准确率降至 0.742,较完整模型降低了 0.006。这表明 VAE 在生成潜在空间的表示方面具有显著作用,去除该策略后,模型的表现有所下降。

替换 SMA 为 EMD 分解后准确率降至 0.732,较完整模型降低了 0.013。这表明在趋势项处理上 SMA 的性能较好,替换该策略后,模型的表现有所下降。

综上所述,通过实验结果可以得出 STURL 方法在整体上表现最佳,其次是去除实例对比学习,最后是频率混合策略。这也进一步证实了在时间序列通用表示学习中,数据增强对于整个模型的提升作用最大。

5 结语

本文提出了一种新的基于时序数据分解的对比学习框架——STURL,用于时序数据的通用表征学习。通过将时间序列分解为季节项和趋势项,结合分层对比学习策略,STURL 能够有效捕捉时序数据的多尺度特性和细粒度结构,提供更加鲁棒的语义表示。实验结果表明,STURL 在多个任务中,如时间序列预测、分类、插值和异常检测等,均显著优于现有的主流方法,验证了其在复杂时序数据处理中的优越性能。尤其在面对高噪声、非平稳数据以及数据缺失的情况下,STURL 展现出了强大的鲁棒性和稳定性。通过频率混合策略、上下文一致性机制和多尺度池化等设计,本文所提模型能够有效增强模型对时序数据的适应能力,极大提升了在多任务场景下的表现。

尽管 STURL 在多个实验中取得了良好效果,仍存在一些可以改进的空间。未来的研究可以进一步探索如何更高效地处理更大规模的时序数据,探讨分解误差对模型性能的影响,探讨模型在极端情况下的表现。总之,STURL 为时序数据的表征学习提供了一种新的视角,其所提出的框架和方法不仅具有较强的理论价值,也为实际应用中时序数据分析的优化提供了可行的解决方案。

参考文献:

[1] CAILLAULT É P,BIGAND A.Comparative study on univariate forecasting methods for meteorological time series[C]//2018 26th European Signal Processing Conference (EUSIPCO).Rome,Italy,IEEE,2018:2380-2384.

- [2] BASU S, MUKHERJEE A, KLIVANSKY S. Time series models for internet traffic[C]//Proceedings of IEEE INFOCOM'96. Conference on Computer Communications. San Francisco, CA, USA, IEEE, 1996, 2: 611-620.
- [3] MAKRIDIS G, KYRIAZIS D, PLITSOS S. Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry[C]//2020 IEEE 23rd international conference on intelligent transportation systems (ITSC). Rhodes, Greece, IEEE, 2020: 1-8.
- [4] SEZER O B, GUDELEK M U, OZBAYOGLU A M. Financial time series forecasting with deep learning: A systematic literature review: 2005—2019[J]. Applied soft computing, 2020, 90: 106181.
- [5] SHUMWAY R H, STOFFER D S, SHUMWAY R H, et al. Time series regression and ARIMA models[J]. Time Series Analysis and Its Applications, 2000: 89-212.
- [6] 尹光志, 张卫中, 张东明, 等. 基于指数平滑法与回归分析相结合的滑坡预测[J]. 岩土力学, 2007(8): 1725-1728. DOI: 10.16285/j.rsm.2007.08.005.
- [7] TONEKABONI S, EYTAN D, GOLDENBERG A. Unsupervised representation learning for time series with temporal neighborhood coding[J]. arXiv preprint arXiv:2106.00750, 2021.
- [8] ELDELE E, RAGAB M, CHEN Z, et al. Time-series representation learning via temporal and contextual contrasting[J]. arXiv preprint arXiv:2106.14112, 2021.
- [9] HAJIMORADLOU A, PISHDAD L, TUNG F, et al. Self-supervised time series representation learning with temporal-instance similarity distillation[C]//First Workshop on Pre-training: Perspectives, Pitfalls, and Paths Forward at ICML, 2022.
- [10] YANG X, ZHANG Z, CUI R. Timeclr: A self-supervised contrastive learning framework for univariate time series representation[J]. Knowledge-Based Systems, 2022, 245: 108606.
- [11] WANG Y, XU Y, YANG J, et al. Fully-connected spatial-temporal graph for multivariate time-series data[C]//Proceedings of the AAAI conference on artificial intelligence. Vancouver, British Columbia, Canada, 2024, 38(14): 15715-15724.
- [12] MORGADO P, LI Y, NVASCONCELOS N. Learning representations from audio-visual spatial alignment[J]. Advances in Neural Information Processing Systems, 2020, 33: 4733-4744.
- [13] HARESH S, KUMAR S, COSKUN H, et al. Learning by aligning videos in time[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE/CVF, 2021: 5548-5558.
- [14] CHEN M, WEI F, LI C, et al. Frame-wise action representations for long videos via sequence contrastive learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE/CVF, 2022: 13801-13810.
- [15] YANG Y, MA J, HUANG S, et al. Tempclr: Temporal alignment representation with contrastive learning[J]. arXiv preprint arXiv:2212.13738, 2022.
- [16] JIAO Y, YANG K, DOU S, et al. Timeautoml: autonomous representation learning for multivariate irregularly sampled time series[J]. arXiv preprint arXiv:2010.01596, 2020.
- [17] YUE Z, WANG Y, DUAN J, et al. Ts2vec: Towards universal representation of time series[C]//Proceedings of the AAAI conference on artificial intelligence. AAAI, 2022, 36(8): 8980-8987.
- [18] ZHANG K, WEN Q, ZHANG C, et al. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects[J]. IEEE transactions on pattern analysis and machine intelligence, 2024, 46(10): 6775-6794.
- [19] XU M A, MORENO A, WEI H, et al. Retrieval-based reconstruction for time-series contrastive learning[J]. Arxiv-abs231100519, 2023.
- [20] SHIN Y, YOON S, SONG H, et al. Context consistency regularization for label sparsity in time series[C]//International Conference on Machine Learning. Honolulu, Hawaii, USA, PMLR, 2023: 31579-31595.
- [21] ZHONG S, SONG S, ZHUO W, et al. A multi-scale decomposition mlp-mixer for time series analysis[J]. arXiv preprint arXiv:2310.11959, 2023.
- [22] CHOWDHURY R R, LI J, ZHANG X, et al. Primenet: Pre-training for irregular multivariate time series[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Washington D C, USA, 2023, 37(6): 7184-7192.
- [23] CHEN P, HUANG D, HE D, et al. Rspnet: Relative speed perception for unsupervised video representation learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Atlanta, Georgia, USA, 2021, 35(2): 1045-1053.
- [24] JENNI S, JIN H. Time-equivariant contrastive video representation learning[C]//Proceedings of the IEEE/CVF international conference on computer vision. Shen Zhen, IEEE, 2021: 9970-9980.
- [25] DUY NGUYEN A, TRAN T H, PHAM H H, et al. Learning Robust and Consistent Time Series Representations: A Dilated Inception-Based Approach[J]. arXiv e-prints arXiv:2306.06579, 2023.
- [26] LIANG Z, ZHANG J, LIANG C, et al. A shapelet-based framework for unsupervised multivariate time series representation learning[J]. arXiv preprint arXiv:2305.18888, 2023.
- [27] WANG Y H, YU H, WANG H S, et al. 2023. Contrast everything: A hierarchical contrastive framework for medical time-se-

- ries. In NeurIPS[C]//Proceedings of the 37th Conference on Neural Information Processing Systems. New Orleans, LA: Curran Associates Inc, 2023, 36: 35694–55717.
- [28] WANG Y, HAN Y, WANG H, et al. Contrast everything: A hierarchical contrastive framework for medical time-series[J]. Advances in Neural Information Processing Systems, 2023, 36: 55694–55717.
- [29] KIM S, CHUNG E, KANG P. FEAT: A general framework for feature-aware multivariate time-series representation learning[J]. Knowledge-Based Systems, 2023, 277: 110790.
- [30] CHOI H, KANG P. Multi-task self-supervised time-series representation learning[J]. Information Sciences, 2024, 671: 120654.
- [31] LIU S, KIMURA T, LIU D, et al. Focal: Contrastive learning for multimodal time-series sensing signals in factorized orthogonal latent space[J]. Advances in Neural Information Processing Systems, 2023, 36: 47309–47338.
- [32] ZHANG X, ZHAO Z, TSILIGKARIDIS T, et al. Self-supervised contrastive pre-training for time series via time-frequency consistency[J]. Advances in neural information processing systems, 2022, 35: 3988–4003.
- [33] LIU J, CHEN S. Timesurl: Self-supervised contrastive learning for universal time series representation learning[C]//Proceedings of the AAAI conference on artificial intelligence. Vancouver, BC: AAAI Press, 2024, 38(12): 13918–13926.
- [34] LEESEUNGHAN, PARKTAEYOUNG, LEEKIBOK. Soft Contrastive Learning for Time Series in ICLR[C]//Proceedings of the 12th International Conference on Learning Representations. kigali: Openview, 2024.
- [35] LEE S, PARK T, LEE K. Soft contrastive learning for time series[J]. arXiv preprint arXiv: 2312.16424, 2023.
- [36] DAU H A, BAGNALL A, KAMGAR K, et al. The UCR time series archive[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(6): 1293–1305.
- [37] BAGNALL A, DAU H A, LINES J, et al. The UEA multivariate time series classification archive[J]. arXiv preprint arXiv: 1811.00075, 2018.
- [38] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the AAAI conference on artificial intelligence. Virtual Event: AAAI Press, 2021, 35(12): 11106–11115.
- [39] 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. Anchorage, AK, USA: ACM, 2019: 3009–3017.
- [40] WU H, HU T, LIU Y, et al. Timesnet: Temporal 2d-variation modeling for general time series analysis[J]. arXiv preprint arXiv: 2210.02186, 2022.
- [41] CHOI H, SOM A, TURAGA P. AMC-Loss: Angular Margin Contrastive Loss for Improved Explainability in Image Classification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 3659–3666.
- [42] BOUHSINE T, EL AAROUSSI I, FAYSAL A, et al. SimO loss: anchor-free contrastive loss for fine-grained supervised contrastive learning[J]. arXiv e-prints arXiv: 2410.05233, 2024.

(编辑:贾丽红)