

文章编号: 2095-2163(2023)07-0027-06

中图分类号: TP391.4

文献标志码: A

基于掩码时间注意力和置信度损失函数的序列数据早期分类方法

陈慧玲, 张 晔, 田奥升, 赵晗馨

(国防科技大学 电子科学学院, 长沙 410073)

摘要: 序列数据的早期分类对于高时效性应用具有重要意义。该任务的目标是在满足预期分类精度的前提下, 尽快地对持续输入的时间序列进行分类。目前, 深度学习已经在序列数据早期分类任务中得到了广泛应用。现有的深度方法通常利用递归神经网络来适应流数据的长度变化, 并通过设置分类概率阈值退出分类过程。然而这些方法忽视了流数据的关键识别区域随信息量的增加持续变化。为了解决该问题, 本文提出了一种基于掩码时间注意力机制的时间卷积网络来动态关注关键识别区域。此外, 考虑到正确类别的分类概率分数应随模型观察到更多数据单调不递减, 本文设计了一个置信度损失函数惩罚不符合该条件的模型, 进一步促使模型提取更有区分性的特征。在8个公开数据集的实验结果表明了所提方法优越的早期分类性能。

关键词: 序列数据早期分类; 掩码时间注意力; 置信度损失函数; 时间卷积网络

Early time series classification based on masked time attention and confidence loss function

CHEN Huiling, ZHANG Ye, TIAN Aosheng, ZHAO Hanxin

(College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China)

[Abstract] Early time series classification (ETSC) is significant for time-sensitive applications. This task aims to classify a continuous input time series as quickly as possible and meet the expected classification quality. Deep learning has been extensively explored in ETSC. Existing deep learning methods usually leverage recurrent neural networks to adapt to the length variation of flow data, and set the threshold for early quitting. However, these methods underestimate the continuous change of critical identification regions of flow data. To solve this problem, this paper proposes a temporal convolutional network based on a masked time attention mechanism. In this way, the model can dynamically focus on key recognition regions. Besides, considering that the classification probability of the correct category should be monotonically non-decreasing as the model observes more data, this paper designs a confidence loss function to penalize models that do not meet this condition, further prompting the model to extract more discriminative features. Experimental results on 8 public datasets demonstrate the superior early classification performance of the proposed method.

[Key words] early time series classification; masked time attention; confidence loss function; temporal convolutional network

0 引言

随着传感器技术的飞速发展, 序列数据分类在健康监测、智能家居控制、设备监控等领域得到了广泛的应用。一些对时效性要求较高的现实应用, 例如灾难预测、气体泄漏检测、故障检测^[1-2]等, 都需要提早地对序列数据进行分类。因此, 序列数据的早期分类具有重要研究价值^[3]。然而, 真实世界中的早期分类的流数据输入形式为分类增加了难度,

并且难以设置合适的停止条件退出分类。

近年来, 传统方法取得了较好的早期分类效果^[4-6], 然而手工设计的特征需要大量的专家经验。此外, 这些方法还需要为不同长度的数据训练多个不同的分类器。深度方法由于其自动分类方案以及有效的特征提取能力在序列数据分类领域取得了卓越性能^[7]。部分研究人员逐渐利用深度学习的方法解决序列数据早期分类任务^[8]。因此, 本文主要关注基于深度学习的方法。

作者简介: 陈慧玲(1998-), 女, 硕士研究生, 主要研究方向: 序列数据分类、信号处理技术; 张 晔(1990-), 男, 博士研究生, 主要研究方向: 序列数据分类、信号处理技术; 田奥升(1998-), 男, 硕士研究生, 主要研究方向: 序列数据分类、信号处理技术; 赵晗馨(1999-), 女, 硕士研究生, 主要研究方向: 视觉目标跟踪、信号处理技术。

收稿日期: 2022-09-07

哈尔滨工业大学主办 ◆ 学术研究与应用

在时间早期分类领域,基于深度的方法主要分为一阶段的方法和二阶段的方法。其中,一阶段的方法是指同时对分类过程及退出过程进行优化。这类方法通常设置分类子网和提前退出子网并对其联合优化。然而,分类子网和退出子网的优化具有一定的冲突^[8]。这是由于在时间推移过程中,分类准确率随信息量的增长递增的同时早期性不断降低。因此,一阶段的方法难以同时对 2 个子网进行优化。二阶段的方法通过将分类过程与退出过程分离来缓解这种冲突,首先单独对分类器进行训练,然后通过制定设置阈值或一定的退出规则来退出分类。递归神经网络(recursive neural networks, RNN)由于其贯序的输入方式被广泛应用于早期分类^[9],然而其局部特征提取能力不强。Huang 等学者^[10-12]利用卷积神经网络(Convolutional Neural Networks, CNN)良好的局部特征提取能力,结合 CNN 和 RNN 构建混合模型来同时提取局部特征及时序信息,然后利用分类概率计算置信度,并制定了一定的退出规则。Hsu 等学者^[13]同样使用了混合分类模型,并引入了注意力机制以增强模型的可解释性。此外,现有方法也未考虑到识别正确的概率随时间变化的规律。

为了解决这些问题,本文提出了掩码时间注意力机制以及置信度损失函数。首先,本文利用基于掩码时间注意力机制的时间卷积网络对于不同长度的数据产生自适应的注意力权重,从而动态地抑制无关信息,并更加关注关键区域的有效信息,增强特征图的信息表达能力。然后,本文受到正确类别的概率随时间推移而递增^[14-15]的启发,设计了置信度

损失函数。通过对不满足该条件的概率进行惩罚,使得正确类别的概率随数据长度增加而平滑地增加,利于退出阈值的设置。

1 相关知识

1.1 注意力机制

深度模型的表达能力随着参数的增加而不断提升。然而参数的增加带来了更大的计算量,同时也增加了大量的冗余信息。因此,注意力机制被引入深度模型对网络参数进行调制。该机制的核心思想是抑制无关信息,使模型关注更有效的关键特征^[16]。注意力机制的核心思想是通过一定的变换来学习不同特征重要性的差异,显著提高了信息处理与应用的效率,此外,还具有通用性、直观性和可解释性等优点。因此被广泛应用于机器翻译、文本分类、语音识别、图像处理等多个领域。根据注意力机制插入的位置,注意力机制可分为时间注意力、空间注意力、通道注意力等^[17]

实现注意力机制,通常首先对特征图进行非线性变化得到注意力分数,然后对其进行 *SoftMax* 归一化当成注意力权重,最后将相应的权重作用于原特征图进行加权或者逐点相乘获得新的特征表示。

1.2 时间卷积网络

由于 RNN 不能大规模并行以及具有长时间遗忘的缺陷,Bai 等学者^[18]提出了具有时序处理能力的时间卷积网络(Temporal Convolutional Network, TCN)。膨胀因果卷积为 TCN 的主要组成部分,其结构如图 1 所示。

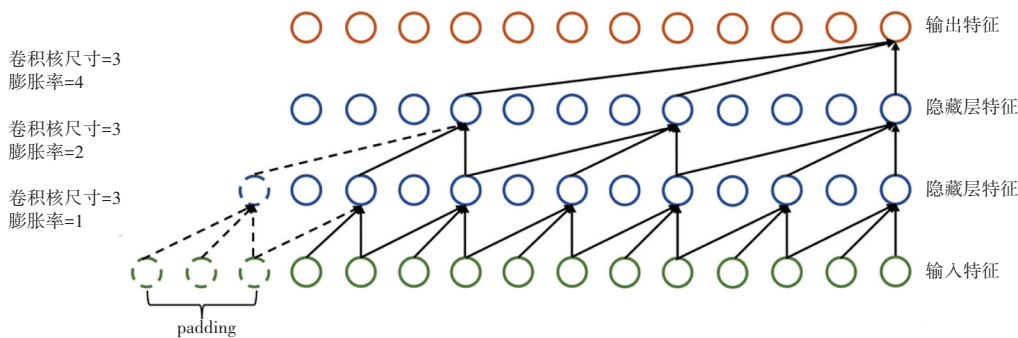


图 1 膨胀因果卷积示意图

Fig. 1 Dilated causal convolution

由图 1 可知,膨胀因果卷积具有严格的时间约束,这是由于因果卷积的应用使得某一时刻的特征只能观察到该时刻之前的数据^[19]。相比于常规卷积,因果卷积防止了未来信息的泄漏。单纯的因果卷积受限于卷积核大小,难以有效提取全局特征。

常规 CNN 通过引入 pooling 层来增加感受野,然而 pooling 层会造成一定的信息损失。因此 Chen 等学者^[20]提出膨胀卷积,通过对卷积时的输入间隔采样来增加感受野。采样率,即膨胀率,指的是 kernel 的间隔数量(标准的 CNN 中膨胀率为 1)。通常,膨

胀率随着深度模型的层数加深而增大,因此膨胀卷积使得感受野大小随着层数呈指数型增长。除膨胀因果卷积外,TCN 使用 *WeightNorm* 和 *Dropout* 来正则化网络,并且对不同卷积层进行残差连接以更好地对网络进行训练。

2 方法

为了解决模型难以动态关注流数据的关键识别区域的问题,本文提出了掩码时间注意力机制。此外,考虑到识别正确的概率随时间推移而增加,本文提出了置信度损失函数。

2.1 掩码时间注意力机制

为了使模型能够自适应地关注不同长度数据的关键特征区域,利用有限的参数提取更有效的特征,本文为序列数据的早期分类设计了基于掩码的时间注意力机制。常规的注意力机制对于所有时刻的特征计算其注意力分布,然而对于早期分类持续输入的序列数据,在某一时刻只能观察到该时刻之前的特征。因此,本文将掩码引入常规的时间注意力机制以防止未来信息的泄露。具体的掩码时间注意力过程如图 2 所示。

首先,本文将输入特征经过线性层以及 Tanh 激

活函数变换得到不同时刻的注意力分布,其大小为 $1 \times T$ (T 为完整序列数据的长度)。然后,本文对注意力分数进行扩充,对其重复 T 次,并拼接在一起,得到一个大小为 $T \times T$ 的注意力矩阵。同时,输入特征也被采取同样的扩充,得到大小为 $C \times T \times T$ 的特征矩阵 (C 为特征图的通道数)。接下来,将注意力矩阵的上三角填充为负无穷 (该步骤简称为掩码,即图 2 中的 Mask 操作),使得注意力矩阵经过 *SoftMax* 操作后上三角的注意力值为 0。这表示在 t 时刻,模型只会关注 t 时刻之前的时刻特征。此后,对特征矩阵和经过掩码的注意力矩阵进行逐元素乘法,得到不同时刻的动态特征 (大小为 $C \times T \times T$)。最后,本文对该动态特征使用平均池化得到 $C \times T$ 的融合特征,每个时刻的特征都是由该时刻前的特征通过加权相应的注意力权重得到。

将提出的掩码时间注意力机制嵌入 TCN 中,得到基于掩码时间注意力的 TCN 网络,该网络结构如图 3 所示。首先,完整的序列数据被输入到多个时域卷积块提取出局部时序特征。然后,网络利用掩码时间注意力机制对不同时刻的局部特征进行动态加权,输出各时刻的动态融合特征。最后,这些动态特征通过线性层以及 *SoftMax* 函数得到分类概率。

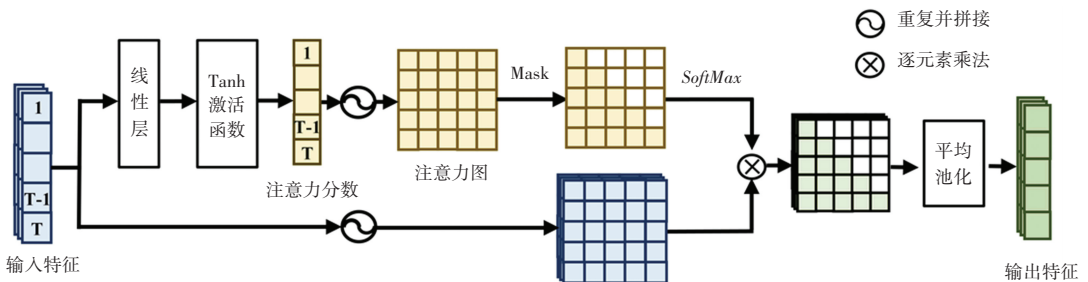


图 2 掩码时间注意力机制结构图
Fig. 2 Masked time attention mechanism

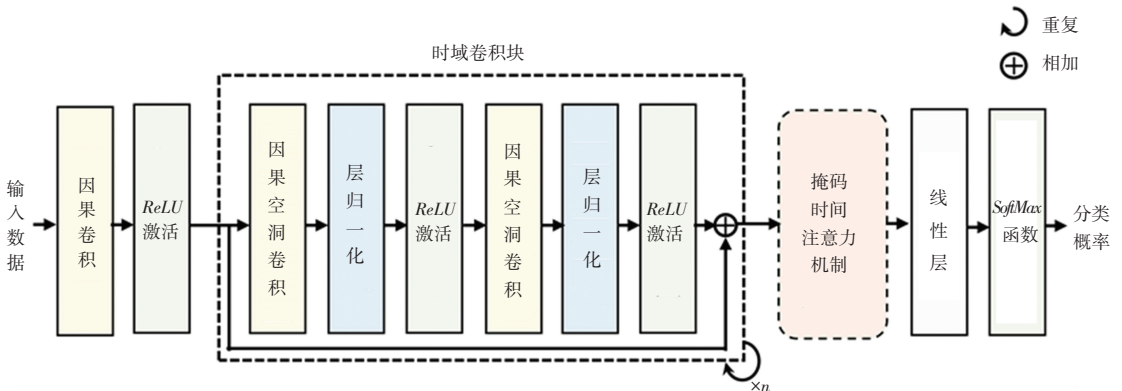


图 3 基于掩码时间注意力机制的时间卷积网络结构图
Fig. 3 The architecture of temporal convolutional network based on masked time attention mechanism

2.2 置信度损失函数

考虑到当分类器观察到更多的信息时应该对正确的活动类别有更大的影响,本文引入了对正确类别分类概率随时间的约束,即随着数据长度的增加,正确的类别输出更高的概率分数。

具体地,本文设计了一个置信度损失,该损失定义为:

$$L(\theta) = \mu * L_c(\theta) + L_p(\theta) \quad (1)$$

其中, θ 为模型的所有参数, $L_p(\theta)$ 、 $L_c(\theta)$ 分别表示常规的交叉熵损失函数和本文设计的违背时间约束的惩罚损失,对此求得的数学定义见如下公式:

$$L_p(\theta) = \frac{1}{N * T} \sum_{t=1}^T \sum_{i=1}^N (-\log P_i^t) \quad (2)$$

$$L_c(\theta) = \frac{1}{N * T} \sum_{t=1}^T \sum_{i=1}^N l_i^t \quad (3)$$

$$l_i^t = \min(0, (P_i^t - P_{\max_i}^t)) \quad (4)$$

$$P_{\max_i}^t = \max(P_i^1, P_i^2, \dots, P_i^t) \quad (5)$$

其中, P_i^t 表示分类器输入第 i 个样本的前 t 个数据得到的输出分类概率, N 为训练集的样本总数。为了便于理解,本文在图 4 中对该损失函数做进一步说明。

图 4 绘制了一个样本的正确类别分类概率随时间变化的示意图。图 4 中,在 t_a 时刻之前,概率 P 一直单调递增,该现象符合正确类别概率随时间递增的约束。因此,本文提出的置信度损失不对其进行惩罚,即此时 l_i 为 0。在 t_a 时刻之后,正确类别的分类概率开始下降。例如在 t_b 时刻,正确类别的概率 $P_i^{t_b}$ 低于其在 t_b 时刻前的最大正确类别概率 ($P_{\max_i}^{t_b}$),这不满足本文提出的置信度约束。因此,该样本在 t_b 时刻的损失通过 t_b 时刻之前的最大正确类别概率减去 t_b 时刻的正确类别概率计算得到,具体参见式 (4)。 $L_c(\theta)$ 的设计将正确类别的检测分数限制为随着活动的进展而单调地不减少。

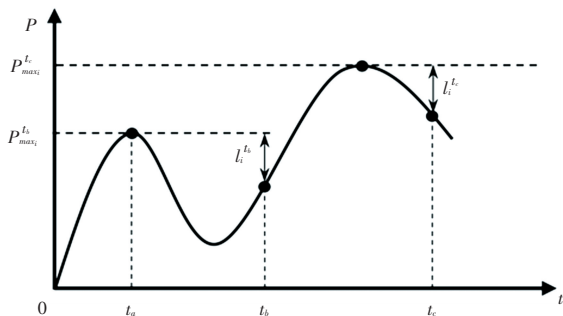


图 4 正确类别分类概率随时间变化曲线

Fig. 4 The classification probability curve of correct class regard to time

2.3 训练及测试流程

本节对具体的训练及测试流程进行介绍,设计研发过程如图 5 所示。

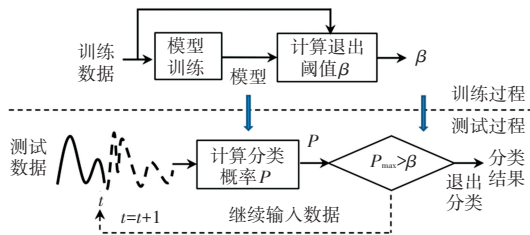


图 5 训练及测试过程

Fig. 5 Training and testing process

在训练阶段,利用训练集数据对提出的基于掩码时间注意力机制的时间卷积网络进行训练。随后,将训练集的所有序列数据输入到之前训练的模型中,得到所有样本不同时刻的分类概率。利用这些分类概率,采用 Sharma 等学者^[14]提出的退出规则计算出该数据集的退出阈值 β 。

在测试阶段,将测试数据随时间逐渐输入到训练好的模型中。在 t 时刻,将长度为 t 的数据输入到模型得到该时刻的分类概率,当该分类概率的最大值大于阈值 β 时,则停止继续输入更多的数据,将 t 时刻的分类结果作为该样本的分类结果,并将 t 时刻作为提前退出的时刻。通过该分类结果和该退出时刻来计算准确率及早期性。如果 t 时刻的分类概率的最大值小于阈值 β ,则继续输入数据,重复测试过程,直至不能再观测到任何数据。保留最后时刻的分类结果作为该样本的分类结果,且该样本的早期性为 1。

3 实验

3.1 实验数据

为了验证提出方法的有效性,本文采用了公开的 UCR^[21] 存储库提供的单变量数据集,从中选取了不重复的 8 个数据集。UCR 存储库中的序列数据从诸多现实应用采集而来,包括电气设备监控数据、心电图数据、动作识别数据以及其他传感器数据等。UCR 存储库依据一定的规则将这些数据集划分了训练集和测试集,并对数据进行了归一化。

3.2 实验条件

实验采用的深度学习框架 Pytorch1.9.0,所使用的硬件环境为 NVIDIA RTX 3080 GPU。实验中,使用 Adam 优化器对模型参数进行训练,学习率衰减为原来的一半。采用训练集损失最小的模型作为测试模型。每个模型的训练迭代次数设置为 200。所

有模型均使用了3个时域卷积模块,卷积核尺寸为3,隐藏层的通道数为64。根据经验,本文将2.2节中式(1)中的参数 μ 设置为6。为了衡量早期分类性能,研究使用准确率和早期性的调和平均值(harmonic mean, HM)作为评价指标, $HM^{[21]}$ 具体定义为:

$$HM = \frac{2 * accuracy * (1 - earliness)}{accuracy + 1 - earliness} \quad (6)$$

据式(6)可知, HM 的值随早期分类性能的提升而增加。

3.3 对比实验

为验证本文提出方法的有效性,将本文提出的模型与ECLN^[22]、ETMD^[14]、EARLIE^[23]进行对比,4种方法在8个数据集上的测试结果见表1。

表1 在8个数据集上的对比实验结果

Tab. 1 The comparative experimental results on 8 datasets

| 数据集 | 方法 | | | |
|------------------|-------|-------|--------|--------------|
| | ECLN | ETMD | EARLIE | 本文 |
| ChlorineCon | 69.44 | 69.35 | 69.85 | 86.24 |
| CricketX | 16.33 | 26.02 | 23.84 | 62.97 |
| FaceAll | 16.84 | 7.65 | 45.22 | 49.98 |
| MedicalImages | 52.08 | 4.86 | 67.81 | 82.15 |
| SyntheticControl | 22.83 | 29.59 | 68.04 | 78.18 |
| TwoPatterns | 17.42 | 42.84 | 38.21 | 53.85 |
| UWaveZ | 41.03 | 27.87 | 44.85 | 60.23 |
| Wafer | 86.01 | 94.00 | 93.92 | 98.81 |

从表1可以看出,对比其他3种方法,本文提出的模型在8个数据集上均取得了最优的早期分类结果,证明了本文提出方法的先进性。

3.4 消融实验

为了分别验证本文提出的掩码时间注意力机制以及置信度损失函数的有效性,本节分别对这2个部分进行消融。本文设置的基线模型为去除了掩码时间注意力机制的时间卷积网络,并使用经典的交叉熵损失函数对模型进行训练。首先,为了证明提出的掩码时间注意力机制的效果,本文将该模块添加到基线模型进行第一个消融实验。该实验在8个数据集上实验结果见表2第2、3列。其次,本文将交叉熵损失函数替换为提出的置信度损失函数,以证明该损失函数的有效性。该实验结果见表2第3、4列。

观察表2第2、3列,相比于基线方法,添加了掩码时间注意力的基线方法分别将8个数据集上的 HM 分数提高了1%,5.48%,10.56%,3.57%,0.51%,

2.33%,1.81%,0.02%。因此,本文提出的掩码时间注意力机制显著提升了模型的早期分类性能。

表2 在8个数据集上的消融实验结果

Tab. 2 The ablation experimental results on 8 datasets

| 数据集 | 基线 | 基线+掩码时间注意力 | 本文 |
|------------------|-------|------------|--------------|
| ChlorineCon | 83.57 | 84.57 | 86.24 |
| CricketX | 55.73 | 61.21 | 62.97 |
| FaceAll | 36.12 | 46.68 | 49.98 |
| MedicalImages | 75.48 | 79.05 | 82.15 |
| SyntheticControl | 75.54 | 76.05 | 78.18 |
| TwoPatterns | 46.85 | 49.13 | 53.85 |
| UWaveZ | 53.59 | 55.40 | 60.23 |
| Wafer | 98.40 | 98.42 | 98.81 |

为了进一步说明添加了掩码时间注意力的模型能实现更有效的分类,本文在图6中绘制了Synthetic Control数据集的分类准确性随数据长度变化的结果。图6中,本文提出的掩码时间注意力机制提高了几乎所有长度的数据的分类性能。

观察表2的第3、4列,用本文提出的置信度损失替换经典的交叉损失函数后,在8个数据集上的 HM 分数分别提高了1.67%,1.76%,3.3%,3.1%,2.13%,4.72%,4.83%,0.39%。这表明,使用了本文提出的置信度损失函数训练模型使得模型的早期分类性能得到了显著的提升。

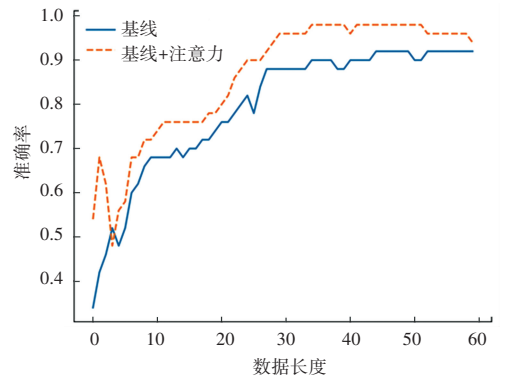


图6 注意力机制对Synthetic Control数据集的不同长度数据准确率的影响

Fig. 6 The effect of the attention mechanism on the accuracy of varied-length data on the Synthetic Control dataset

4 结束语

本文提出了基于掩码时间注意力机制的时间卷积网络,提高了模型对不同长度数据的自适应能力。此外,本文通过设计的置信度损失函数促使正确类别的概率随信息量的增加递增,有利于设置更合理的退出阈值。在8个公开数据集上的实验结果证明

了本文提出的方法的有效性。然而,固定阈值难以适应难度程度不同的数据,该问题将在未来进行更深入的探讨研究。

参考文献

- [1] GANDHIMATHINATHAN A, LAVANYA R. Early fault detection in safety critical systems using complex Morlet wavelet and deep learning[M]// RANGANATHAN G, FERNANDO X, SHI F. Inventive Communication and Computational Technologies. Singapore: Springer, 2022, 311: 515-531.
- [2] LIANG Haopeng, CAO Jie, ZHAO Xiaoqiang. Multi-scale dynamic adaptive residual network for fault diagnosis[J]. Measurement, 2022, 188: 110397.
- [3] ACHENCHABE Y, BONDU A, CORNUÉJOLS A, et al. Early Classification of Time Series is Meaningful[J]. arXiv preprint arXiv:2104.13257, 2021.
- [4] ZHANG Wenjing, WAN Yuan. Early classification of time series based on trend segmentation and optimization cost function[J]. Applied Intelligence, 2022, 52(6): 6782-6793.
- [5] MORI U, MENDIBURU A, DASGUPTA S, et al. Early classification of time series by simultaneously optimizing the accuracy and earliness[J]. IEEE Transactions on Neural Networks and Learning Systems, 2017, 29(10): 4569-4578.
- [6] SHARMA A, SINGH S K. Early classification of multivariate data by learning optimal decision rules[J]. Multimedia Tools and Applications, 2021, 80(28-29): 35081-35104.
- [7] FAOUZI J. Time series classification: A review of Algorithms and Implementations[J]. Machine Learning (Emerging Trends and Applications), 2022.
- [8] GUPTA A, GUPTA H P, BISWAS B, et al. Approaches and applications of early classification of time series: A review[J]. IEEE Transactions on Artificial Intelligence, 2020, 1(1): 47-61.
- [9] HUANG Zhengjie, YE Zi, LI Shuangyin, et al. Length adaptive recurrent model for text classification[C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. Singapore: ACM, 2017: 1019-1027.
- [10] HUANG Huaishuo, LIU C L, TSENG V S. Multivariate time series early classification using multi-domain deep neural network[C]// 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA). Turin, Italy: IEEE, 2018: 90-98.
- [11] SHEKHAR S, ESWARAN D, HOOI B, et al. Benefit-aware early prediction of health outcomes on multivariate EEG time series[J]. Biomedical Informatics, 2023, 139: 104296.
- [12] MIN Rui, WANG Xing, ZOU Jie, et al. Early gesture recognition with reliable accuracy based on high-resolution IoT radar sensors[J]. IEEE Internet of Things Journal, 2021, 8(20): 15396-15406.
- [13] HSU E Y, LIU C L, TSENG V S. Multivariate time series early classification with interpretability using deep learning and attention mechanism[C]// Pacific - Asia Conference on Knowledge Discovery and Data Mining. Macau, China: Springer, 2019: 541-553.
- [14] SHARMA A, SINGH S K, UDMALE S S, et al. Early transportation mode detection using smartphone sensing data[J]. IEEE Sensors Journal, 2021, 21(14): 15651-15659.
- [15] MA Shugao, SIGAL L, SCLAROFF S. Learning activity progression in lstms for activity detection and early detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1942-1950.
- [16] NIU Zhaoyang, ZHONG Guoqiang, YU Hui. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.
- [17] CHAUDHARI S, MITHAL V, POLATKAN G, et al. An attentive survey of attention models[J]. ACM Transactions on Intelligent Systems and Technology, 2021, 12(5): 1-32.
- [18] BAI Shaojie, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271, 2018.
- [19] HARELL A, MAKONIN S, BAJIĆ I V. Wavenilm: A causal neural network for power disaggregation from the complex power signal[C]// ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, UK: IEEE, 2019: 8335-8339.
- [20] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. arXiv preprint arXiv:1412.7062, 2014.
- [21] SCHÄFER P, LESER U. TEASER: Early and accurate time series classification[J]. Data Mining and Knowledge Discovery, 2020, 34(5): 1336-1362.
- [22] RUBWURM M, TAVENARD R, LEFÈVRE S, et al. Early classification for agricultural monitoring from satellite time series[J]. arXiv preprint arXiv:1908.10283, 2019.
- [23] HARTVIGSEN T, SEN C, KONG Xiangnan, et al. Adaptive-halting policy network for early classification[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Anchorage, Arras, USA: ACM, 2019: 101-110.