一种基于Transformer架构的多层级自动睡眠分期模型

金 峥[1,2,3],贾克斌[1,2,3*]

(1. 北京工业大学信息科学技术学院,北京 100124;2. 先进信息网络北京实验室,北京 100124;
3. 计算智能与智能系统北京市重点实验室,北京 100124)

摘 要: 睡眠是人体保持健康的重要生理过程,基于多导睡眠图(PolySomnoGraphy,PSG)的睡眠分期是诊疗睡眠疾病和评估睡眠质量的重要依据. 人工睡眠分期法在处理大规模PSG数据时存在耗时久、效率低的问题,采用深度学习模型有效表征PSG的自动睡眠分期法显现出广阔的研究前景. 针对现有模型未充分考虑PSG片段内波形信息、通道间相关性信息、片段间睡眠转换信息的问题,本文提出一种基于Transformer架构的多层级睡眠分期网络模型(Hierarchical transFormer sleep staging model,HierFormer),采用Transformer编码器有效提取片段内波形特征、通道相关性特征、片段间转换特征,并结合注意力机制综合提升模型对于PSG片段内、通道间、片段间三种视角信号特性的可解释性. 基于睡眠集-欧洲数据格式(sleep-European Data Format,sleep-EDF)扩展睡眠数据集开展的实验结果表明:本文模型利用更少的参数量取得优于多种现有基线模型的分期性能,分类准确率、宏平均精确率、宏平均召回率、宏平均F1分数、科恩卡帕系数分别可达到0.807、0.784、0.735、0.750和0.721. 通过在三种视角下不同特征编码方式的性能对比和注意力分数的可视化,本文进一步证明了所提模型良好的编码能力和可解释性. 本研究旨在为睡眠分期领域的深度学习应用提供新途径和新技术,从而辅助医生提升睡眠疾病诊疗效率.

关键词: 多导睡眠图(PSG);自动睡眠分期;深度神经网络;Transformer架构;注意力机制;模型可解释性

基金项目: 北京市自然科学基金(No.4212001)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)02-0545-13

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20240596

A Hierarchical Automatic Sleep Staging Model Based on Transformer Architecture

JIN Zheng[1,2,3], JIA Ke-bin[1,2,3*]

(1. *School of Information Science and Technology*, *Beijing University of Technology*, *Beijing* 100124, *China*;
2. *Beijing Laboratory of Advanced Information Networks*, *Beijing* 100124, *China*;
3. *Beijing Key Laboratory of Computational Intelligence and Intelligent System*, *Beijing University of Technology*, *Beijing* 100124, *China*)

Abstract: Sleep is the significant physiological process to keep healthy. Sleep stage classification based on polysomnography (PSG) is the fundamental evidence to diagnose sleep disorders and assess sleep quality. Manual sleep staging method has some typical problems when handling the large-scale PSG data, such as time-consuming and low-efficiency. The automatic sleep staging method that utilizing deep learning models to effectively learn PSG representations shows extensive researching prospects. Most existing models cannot fully consider the epoch-level waveform information, channel-wise correlations, sequence-level sleep transitions. This paper proposes a transformer-based hierarchical sleep staging model (HierFormer), which employs transformer encoder to extract valid epoch-level waveform features, channel-wise correlation features, sequence-level transition features. Meanwhile, it adopts attention mechanism to improve the model interpretability of signal properties from three views, including epoch-level, channel-wise, and sequence-level views. Experimental results on the sleep-european data format (sleep-EDF) database expanded dataset show that the proposed model achieves better sleep staging performance with less parameters compared with various baseline models. The overall accuracy, macro-aver-

收稿日期:2024-06-25;修回日期:2024-10-28;责任编辑:赵克
*通讯作者:贾克斌

aging precision, macro-averaging recall, macro-averaging F1-score, and Cohen's-kappa coefficient achieve 0.807, 0.784, 0.735, 0.750, and 0.721, respectively. According to the performance comparisons of different feature encoding methods from three views and the visualization of attention weights, this paper further demonstrates the satisfied encoding ability and interpretability of proposed model. This study aims to provide innovative deep learning approaches and technologies for the research of sleep staging applications, thus assisting sleep experts to improve the efficiency of sleep disorder diagnosis and treatment.

Key words: polysomnography (PSG); automatic sleep stage classification; deep neural network; transformer architecture; attention mechanism; model interpretability

Foundation Item(s): Natural Science Foundation of Beijing Municipality (No.4212001)

1 引言

睡眠是人体用于缓解压力和恢复机能的重要生理过程,充足的睡眠可避免注意力不集中、头痛等问题[1]. 临床上用于监测睡眠过程的"金标准"是多导睡眠图(PolySomnoGraphy,PSG)[2],其主要包含脑电信号(Electroencephalogram,EEG)、眼电信号(ElectroOculoGram,EOG)、肌电信号(ElectroMyoGram,EMG)、心电信号(ElectroCardioGram,ECG)等[3]. 基于PSG的睡眠分期是诊疗睡眠疾病和评估睡眠质量的前提. 根据美国睡眠医学学会(American Academy of Sleep Medicine,AASM)发布的睡眠分期准则,医生以30 s为一个时间段,将整夜PSG切分为连续多个片段,在分析片段具体特征后将其标记为清醒期(Wake,W)、非快速眼动1~3期(Non-Rapid Eye Movement 1~3,NREM1~3,记作N1~N3)、快速眼动期(Rapid Eye Movement,REM,记作R)共5类中的1类[4]. 睡眠医生通过肉眼分析一整夜PSG(约7~8 h)至少需要花费2 h[5],当遇到大规模PSG数据时,这种方法会出现耗时耗力、效率低下的问题. 因此,以算法模型为基础的自动睡眠分期近年来受到越来越多研究者的关注.

自动睡眠分期的核心思想是利用算法提取PSG片段的关键特征,再结合分类器对每个片段实现睡眠标定. 按照特征提取方法可将现有自动睡眠分期方法分为两大类:传统机器学习和深度学习. 传统机器学习类方法首先利用带通滤波、傅里叶变换(Fourier Transform,FT)等方式对原始PSG进行预处理[6,7];其次针对预处理后的信号提取时域特征(Hjorth参数、偏度等)、频域特征(相对谱功率、功率谱密度等)、非线性特征(香农熵、Petrosian分形维数等)等多种参数组成信号特征集[1,8~12],部分学者选用ReliefF等特征选择算法,进一步筛选出特征集内的有效特征以提升算法性能[9];最后将筛选后特征集结合支持向量机(Support Vector Machine,SVM)、随机森林(Random Forest,RF)、逻辑回归(Logistic Regression,LR)等分类方法实现睡眠分期[7,8,13]. 深度学习类方法则是利用深度神经网络直接学习PSG的高维特征,再结合softmax层实现睡眠分期. 卷积神经网络(Convolutional Neural Network,CNN)可从一维原始信号角度或二维时频图角度学习PSG片段内局部高维特征[14,15],常见的CNN架构包括UNet等[16]. 循环神经网络(Recurrent Neural Network,RNN)可在片段内或片段间角度编码PSG的时序高维特征[17],常用的RNN架构包括长短期记忆(Long Short-Term Memory,LSTM)网络和门控循环单元(Gated Recurrent Unit,GRU)网络[14,18]. 图神经网络(Graph Neural Network,GNN)可对PSG不同通道之间的关联性进行编码,从而学习通道相关性高维特征,流行的GNN架构主要为图卷积神经网络(Graph Convolutional Network,GCN)[3]. 传统机器学习方法的核心在于人为构建特征集,这一过程需要大量的理论知识基础,当特征参数量较大时仍会出现耗时久、效率低的问题. 此外,这类方法在处理病患PSG数据时,还易出现泛化性不足的问题. 而深度学习方法能利用神经网络优化信号预处理,人为设计特征参数、特征选择的整体过程,能在提升算法效率的同时取得更优异的算法性能. 因此,近年来深度学习方法逐渐成为主流的自动睡眠分期方法.

随着自动睡眠分期研究的不断深入,深度学习方法从早期的单一架构模型逐渐优化为多层级架构模型,旨在进一步贴合睡眠医生的临床分析过程. 具体而言,大多数学者先采用CNN或RNN架构提取PSG片段内高维特征(类似医生分析各类有效波形),在此基础上选用时序卷积网络(Temporal Convolutional Network,TCN)或RNN架构抓取连续多个片段间的时序转换高维特征(类似医生分析睡眠阶段转换信息)[14,18~20]. 部分学者还考虑了PSG多通道特征融合问题,对各通道高维特征采用拼接、注意力融合、GCN编码等方法实现特征聚合(类似医生整合各通道波形信息),再学习片段间时序转换信息[21~23]. 此外,很多学者在原有架构的基础上引入注意力机制,让模型从不同角度聚焦于有效特征(通道注意力、时序注意力等)[24,25]. 多层级架构能更全面地考虑PSG的本质特性(片段内波形特征、通道相关性特征、片段间转换特征)并贴合医生分析过程,但现有多数模型并未同时考虑以上三种重要特征[16,26],即使有些模型考虑到了该问题,但其无法同时

在三种视角下提供较高的模型可解释性(例如通道视角简单的特征拼接、片段间视角单一的RNN架构)[17,27].

Transformer架构是近年来深度学习领域的研究热点,不同于传统的CNN、RNN等网络架构,其完全依靠注意力机制对特征序列进行建模,在计算机视觉、自然语言处理等研究领域内的多种任务上表现优异[28]. 部分学者将Transformer架构引入自动睡眠分期领域,以替代传统的CNN或RNN架构. 文献[29]提出了SleepTransformer模型,利用Transformer架构编码PSG片段内和片段间时序高维特征,提升分期性能与模型可解释性;文献[30]设计了CNN-Transformer协作网络(CNN-Transformer Cooperation Network,CTCNet)模型[30],将多尺度CNN与膨胀卷积得到的PSG高维特征输入Transformer架构编码时序高维特征. 上述两文献只针对单通道信号输入,其忽略了重要的通道相关性特征. 文献[31]将PSG时频图切分为块序列,经过线性映射后输入Transformer架构进行编码,得到二维时频视角的片段内高维特征,但其未考虑关键的片段间转换特征. 文献[32]设计的SleepViTransformer模型同样将时频图切块后输入Transformer架构进行编码,基于得到的片段内高维特征采用双向GRU架构学习片段间转换特征,虽然该模型考虑了上述三种视角的重要特征,但其通道融合方式为简单的特征拼接操作,缺乏通道视角的模型可解释性. 总体而言,目前还未有基于Transformer架构的睡眠分期模型能够充分考虑PSG片段内波形特征、通道相关性特征和片段间转换特征,并保证较为全面的模型可解释性.

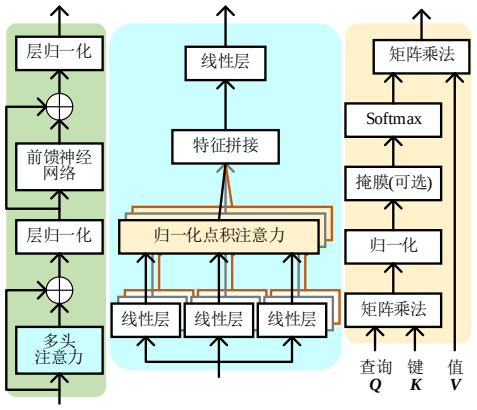
综上所述,本文提出一种基于Transformer架构的多层级睡眠分期模型(Hierarchical transFormer sleep staging model,HierFormer),旨在解决现有方法未同时考虑PSG片段内波形特征、通道相关性特征和片段间转换特征的问题,并采用注意力机制在三种视角下提升模型可解释性. 通过在公开睡眠数据集上与多种现有模型方法进行对比实验,本文验证了所提睡眠分期模型的有效性与可行性. 本文工作的研究贡献可概述为:

(1)引入Transformer架构替代传统的CNN或RNN架构,按递进关系从片段内、通道间、片段间三种视角下编码多通道PSG片段序列的有效特征.

(2)通过Transformer架构和注意力机制相结合的方式,综合提升模型在上述三种视角下针对PSG信号特性的模型可解释性.

2　Transformer架构

在计算机视觉或自然语言处理领域,Transformer架构在各类任务上的实验结果能够优于传统网络架构[29,31]. 标准的Transformer架构由一个编码器和一个解码器组成,二者具有相同的模型结构. 考虑到解码器主要用于序列生成任务[29],而睡眠分期属于分类任务,因此HierFormer模型只采用编码器架构提取PSG各个视角下的高维特征. Transformer编码器结构如图1(a)所示,其主要包含多头注意力、前馈神经网络和层归一化.



(a) 编码器　(b) 多头注意力　(c) 归一化点积注意力

图1　Transformer架构

多头注意力结构如图1(b)所示,该模块首先计算输入特征序列不同位置的元素相互之间的关联性权重值,然后通过加权求和的方式得到输出特征序列,每个值元素相应的权重由查询元素和键元素计算归一化点积注意力函数得到. 多头注意力模块通常由H个归一化点积注意力子模块组成,如图1(c)所示. 具体而言,输入特征首先经过H个线性层映射计算得到多个查询元素、键元素、值元素,再基于三种元素同时计算每个头的归一化点积注意力加权特征,随后H个头的加权特征拼接后再经过线性层映射,输出最终的多头注意力特征. 上述多头注意力计算过程可表述为

$$\boldsymbol{Q}_i=\boldsymbol{Z}\boldsymbol{W}_i^{\boldsymbol{Q}},\ \boldsymbol{K}_i=\boldsymbol{Z}\boldsymbol{W}_i^{\boldsymbol{K}},\ \boldsymbol{V}_i=\boldsymbol{Z}\boldsymbol{W}_i^{\boldsymbol{V}},\ 1\leqslant i\leqslant H \tag{1}$$

$$\boldsymbol{H}_i=\mathrm{Attention}\left(\boldsymbol{Q}_i,\boldsymbol{K}_i,\boldsymbol{V}_i\right)=\mathrm{softmax}\left(\frac{\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}_i \tag{2}$$

$$\boldsymbol{Y}=\mathrm{Concat}\left(\boldsymbol{H}_1,\boldsymbol{H}_2,\cdots,\boldsymbol{H}_H\right)\boldsymbol{W}^{\boldsymbol{Y}} \tag{3}$$

其中,$\boldsymbol{Z}\in\mathbb{R}^{l\times d}$表示长度为$l$的输入特征序列,每个特征向量维度为$d$;$\boldsymbol{Q}_i,\boldsymbol{K}_i,\boldsymbol{V}_i\in\mathbb{R}^{d/H}$表示第$i$个注意力头的查询元素向量、键元素向量、值元素向量;$\boldsymbol{H}_i\in\mathbb{R}^{l\times(d/H)}$表示第$i$个注意力头加权特征;Attention$(\cdot)$表示归一化点积注意力函数;Concat$(\cdot)$表示特征拼接操作;$\boldsymbol{W}_i^{\boldsymbol{Q}},\boldsymbol{W}_i^{\boldsymbol{K}},\boldsymbol{W}_i^{\boldsymbol{V}}\in\mathbb{R}^{d\times(d/H)}$和$\boldsymbol{W}^{\boldsymbol{Y}}\in\mathbb{R}^{d\times d}$表示可学习的权重矩阵;$\boldsymbol{Y}\in\mathbb{R}^{l\times d}$表示多头注意力输出特征序列.

前馈神经网络模块主要由两个全连接层(Fully Connected layer, FC)和 ReLU 激活函数组成. 此外, Transformer编码器还包含两个残差连接操作和两个归一化层,如图1(a)所示,以确保模型稳定训练和快速收敛. 综上, Transformer 编码器整体计算过程可表示为

$$\boldsymbol{Y}=\mathrm{MultiHeadAttention}(\boldsymbol{Z}) \tag{4}$$

$$\boldsymbol{Y}_{\mathrm{LN}}=\mathrm{LayerNorm}(\boldsymbol{Z}+\boldsymbol{Y}) \tag{5}$$

$$\boldsymbol{Y}_{\mathrm{FF}}=\mathrm{ReLU}\left(\boldsymbol{Y}_{\mathrm{LN}}\boldsymbol{W}_1+\boldsymbol{b}_1\right)\boldsymbol{W}_2+\boldsymbol{b}_2 \tag{6}$$

$$\boldsymbol{O}=\mathrm{LayerNorm}\left(\boldsymbol{Y}_{\mathrm{LN}}+\boldsymbol{Y}_{\mathrm{FF}}\right) \tag{7}$$

其中,MultiHeadAttention(·)表示多头注意力计算模块; LayerNorm(·)表示层归一化操作; $\boldsymbol{Y}_{\mathrm{LN}}\in\mathbb{R}^{l\times d}$ 表示归一化特征序列; $\boldsymbol{Y}_{\mathrm{FF}}\in\mathbb{R}^{l\times d}$ 表示前馈神经网络输出特征序列; $\boldsymbol{W}_1\in\mathbb{R}^{d\times d_{\mathrm{FF}}}$、$\boldsymbol{W}_2\in\mathbb{R}^{d_{\mathrm{FF}}\times d}$、$\boldsymbol{b}_1\in\mathbb{R}^{l\times d_{\mathrm{FF}}}$、$\boldsymbol{b}_2\in\mathbb{R}^{l\times d}$ 表示 FC 层可学习权重参数; $\boldsymbol{O}\in\mathbb{R}^{l\times d}$ 表示 Transformer 编码器输出特征序列.

3 多层级睡眠分期模型 HierFormer

HierFormer 模型的整体架构如图2所示. 具体而言,给定一个长度为 L 的 30 s PSG 片段序列 $\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_L\}$, $\boldsymbol{X}_i\in\mathbb{R}^{C\times N}$, $1\leqslant i\leqslant L$,本文模型的目标是预测出中间时刻片段 $\boldsymbol{X}_t\in\mathbb{R}^{C\times N}$, $t=(L+1)/2$ 对应的独热(one-hot)编码标签 $\hat{\boldsymbol{y}}_t\in\mathbb{R}^5$(五分类睡眠分期). 其中, t 表示序列中间时刻的索引, C 表示信号通道数量, $N=f_s\times 30$ 表示各通道数据点数, f_s 表示信号采样率. 多层级端到端模型 HierFormer 主要由 4 个子模块级联而成. (1)片段内波形特征编码模块:对 PSG 片段各通道计算短时傅里叶变换(Short-Term Fourier Transform, STFT),再从时间视角采用 Transformer 编码器学习 STFT 幅值谱(时频序列)的时序特征,结合注意力机制融合出片段内波形高维表征;(2)通道间相关性特征编码模块:从通道视角对得到的多通道片段内波形表征矩阵采用 Transformer 编码器进一步学习空域特征,再结合注意力机制融合出通道间相关性高维表征;(3)片段间转换特征编码模块:从时间视角对通道间相关性表征序列采用 Transformer 编码器进一步学习时序特征,最后结合注意力融合出片段间转换高维表征;(4)睡眠阶段分类模块:基于片段间转换表征计算得出模型预测睡眠标签,结合真实标签实现模型训练与优化.

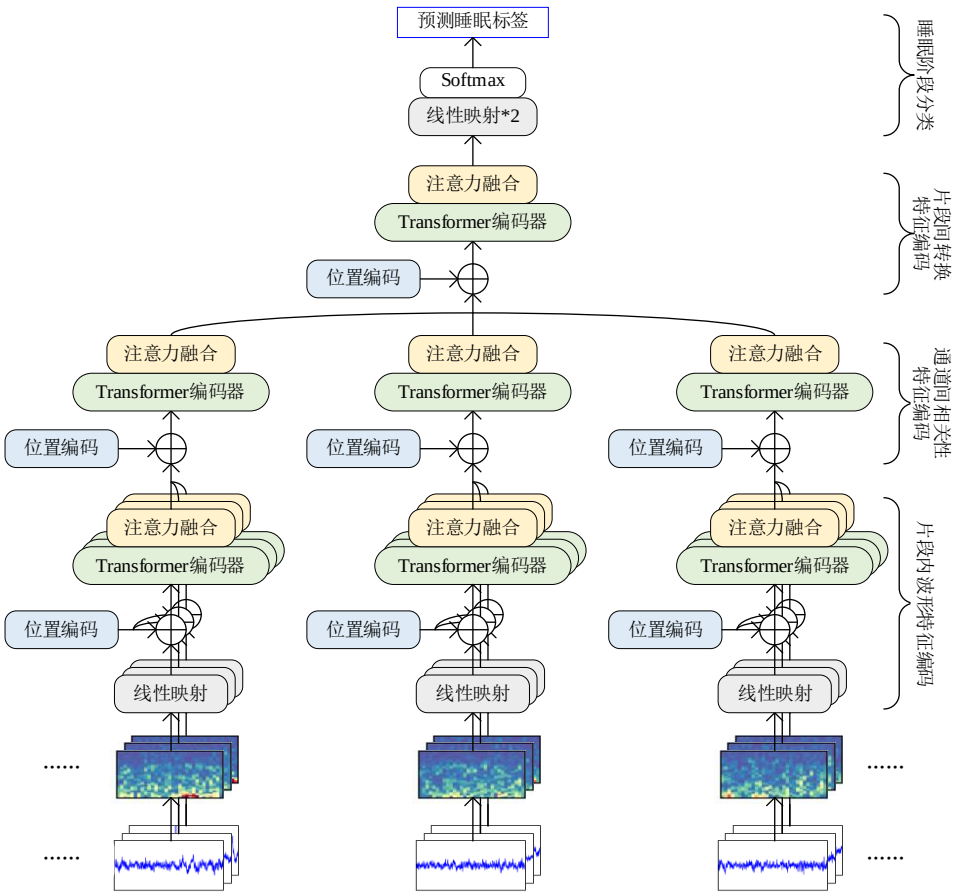


图2 多层级自动睡眠分期模型 HierFormer 整体架构

3.1 片段内波形特征编码

本文模型首先对输入序列每个PSG片段$\boldsymbol{X}_i \in \mathbb{R}^{C\times N}, 1\leqslant i\leqslant L$的各通道一维信号$\boldsymbol{x}_c \in \mathbb{R}^N, 1\leqslant c\leqslant C$计算STFT[3],窗函数选用汉宁窗,窗长2 s,重叠率50%,快速傅里叶(Fast Fourier Transform,FFT)计算点数为N_{FFT}. 对STFT计算结果取绝对幅值谱,得到二维时频图$\boldsymbol{s}_c \in \mathbb{R}^{T\times F}, 1\leqslant c\leqslant C$,$T$表示时频序列长度(时间索引总数),$F$代表频带数,$C$为信号通道数. 随后本文模型对每个时频图进行线性映射(线性层权重矩阵$\boldsymbol{W}_s \in \mathbb{R}^{m\times F}$),在频率维度将时频图的频带数$F$降为$m$(类似可学习的带通滤波器),得到降维后的时频矩阵$\bar{\boldsymbol{s}}_c \in \mathbb{R}^{T\times m}, 1\leqslant c\leqslant C$.

时频矩阵的每一列可被看作相应时间轴上不同频带的能量分布[24],因此,本文模型将时频矩阵$\bar{\boldsymbol{s}}_c$拆分为长度为T的频率信息序列,并采用Transformer编码器学习其时序特征. 因为Transformer编码器中的自注意力机制本身不包含位置的概念,所以需要引入位置编码来处理序列中重要的顺序信息. 对于时频矩阵$\bar{\boldsymbol{s}}_c$的计算过程可表示为

$$\bar{\boldsymbol{s}}_c^{\boldsymbol{P}} = \bar{\boldsymbol{s}}_c + \boldsymbol{P}_s \tag{8}$$

其中,$\boldsymbol{P}_s \in \mathbb{R}^{T\times m}$为位置编码矩阵. 各个矩阵元素由正弦和余弦函数计算得到[28],可表示为

$$\left(\boldsymbol{p}_s\right)_{i,2j} = \sin\left(\frac{i}{10\,000^{2j/m}}\right) \tag{9}$$

$$\left(\boldsymbol{p}_s\right)_{i,2j+1} = \cos\left(\frac{i}{10\,000^{2j/m}}\right) \tag{10}$$

其中,$\left(\boldsymbol{p}_s\right)_{i,2j}$表示位置编码矩阵$\boldsymbol{P}_s$第$i$行第$2j$列的元素. 对添加位置信息的时频矩阵$\bar{\boldsymbol{s}}_c^{\boldsymbol{P}} \in \mathbb{R}^{T\times m}$采用$N$层Transformer编码器学习其片段内波形时序特征,具体过程为

$$\boldsymbol{e}_c^{(N)} = \text{Transformer}_N\left(\bar{\boldsymbol{s}}_c^{\boldsymbol{P}}\right) \tag{11}$$

其中,$\boldsymbol{e}_c^{(N)} \in \mathbb{R}^{T\times m}(1\leqslant c\leqslant C)$表示最后一层(第$N$层)Transformer编码器的输出特征序列;Transformer$(\cdot)$表示式(4)~(7)的简要表示形式;Transformer$_N(\cdot)$表示N层Transformer编码器的堆叠级联形式. 本文模型利用注意力机制将该特征序列融合为单一表征向量并提升模型在片段内视角下的可解释性,计算过程为

$$\boldsymbol{f}_c = \sum_{j=1}^{T} \alpha_j \boldsymbol{e}_{cj}^{(N)} \tag{12}$$

$$\alpha_j = \frac{\exp\left[\sigma\left(\boldsymbol{W}_\alpha^{\mathrm{T}} \boldsymbol{e}_{cj}^{(N)} + b_\alpha\right)\right]}{\sum_{k=1}^{T} \exp\left[\sigma\left(\boldsymbol{W}_\alpha^{\mathrm{T}} \boldsymbol{e}_{ck}^{(N)} + b_\alpha\right)\right]} \tag{13}$$

其中,$\boldsymbol{f}_c \in \mathbb{R}^m(1\leqslant c\leqslant C)$表示第$c$个通道的片段内波形高维表征;$\alpha_j \in \mathbb{R}$表示序列内第$j$个特征向量$\boldsymbol{e}_{cj}^{(N)} \in \mathbb{R}^m(1\leqslant c\leqslant C)$对应的注意力权重;$\sigma(\cdot)$代表Sigmoid激活函数;$\boldsymbol{W}_\alpha \in \mathbb{R}^m$和$b_\alpha \in \mathbb{R}$为线性层可学习参数.

综上所述,各通道一维信号$\boldsymbol{x}_c \in \mathbb{R}^N$对应的二维时频图$\boldsymbol{s}_c \in \mathbb{R}^{T\times F}$经过上述编码过程后变为片段内波形表征$\boldsymbol{f}_c \in \mathbb{R}^m$. 而每个PSG片段$\boldsymbol{X}_i \in \mathbb{R}^{C\times N}(1\leqslant i\leqslant L)$也由此变换为多通道片段内波形表征矩阵$\boldsymbol{F}_i \in \mathbb{R}^{C\times m}(1\leqslant i\leqslant L)$.

3.2 通道间相关性特征编码

现有通道相关性特征提取方式主要有特征拼接、注意力融合等[21,22]. 本文模型则采用N层Transformer编码器和注意力机制相结合的方式来实现通道相关性空域特征编码,并提升模型在通道间视角下的可解释性. 具体而言,基于多通道片段内波形表征矩阵$\boldsymbol{F}_i \in \mathbb{R}^{C\times m}(1\leqslant i\leqslant L)$,首先对其添加位置编码信息:

$$\boldsymbol{F}_i^{\boldsymbol{P}} = \boldsymbol{F}_i + \boldsymbol{P}_c \tag{14}$$

其中,$\boldsymbol{P}_c \in \mathbb{R}^{C\times m}$表示位置编码矩阵,计算方式同式(9)~(10). 随后继续采用N层级联Transformer编码器和注意力机制对加入位置信息的表征矩阵$\boldsymbol{F}_i^{\boldsymbol{P}} \in \mathbb{R}^{C\times m}$实现特征提取与融合,计算过程表示为

$$\boldsymbol{d}_i^{(N)} = \text{Transformer}_N\left(\boldsymbol{F}_i^{\boldsymbol{P}}\right) \tag{15}$$

$$\boldsymbol{r}_i = \text{Attention}_{\text{channel}}\left(\boldsymbol{d}_i^{(N)}\right) \tag{16}$$

其中,$\boldsymbol{d}_i^{(N)} \in \mathbb{R}^{C\times m}$表示第$N$层Transformer编码器的输出特征序列;$\boldsymbol{r}_i \in \mathbb{R}^m(1\leqslant i\leqslant L)$表示序列内第$i$个PSG片段对应的通道间相关性高维表征;Attention$_{\text{channel}}(\cdot)$表示式(12)~(13)描述的注意力融合计算过程.

综上所述,多通道片段内波形表征矩阵序列$\{\boldsymbol{F}_1, \boldsymbol{F}_2, \cdots, \boldsymbol{F}_L\}$被编码为通道间相关性空域表征序列$\boldsymbol{R} \equiv \{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_L\}$.

3.3 片段间转换特征编码

与前面两个编码模块相同,本文模型继续采用N层Transformer编码器和注意力机制相结合的形式,来提取通道间相关性表征序列$\boldsymbol{R} \equiv \{\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_L\}$的片段间睡眠阶段时序转换特征,并提升模型在片段间视角下的可解释性. 基于表征序列$\boldsymbol{R} \in \mathbb{R}^{L\times m}$的整体计算过程可表示为

$$\boldsymbol{R}^{\boldsymbol{P}} = \boldsymbol{R} + \boldsymbol{P}_S \tag{17}$$

$$\boldsymbol{Z} = \text{Transformer}_N\left(\boldsymbol{R}^{\boldsymbol{P}}\right) \tag{18}$$

$$\boldsymbol{O} = \text{Attention}_{\text{seq}}\left(\boldsymbol{Z}\right) \tag{19}$$

其中,$\boldsymbol{R}^{\boldsymbol{P}} \in \mathbb{R}^{L\times m}$表示添加位置信息的特征序列;$\boldsymbol{P}_S \in \mathbb{R}^{L\times m}$表示位置编码矩阵,计算方式同式(9)~(10);$\boldsymbol{Z} \in \mathbb{R}^{L\times m}$表示第$N$层Transformer编码器的输出特征序列;$\boldsymbol{O} \in \mathbb{R}^m$表示包含睡眠阶段转换信息的片段间时序转换高维表征;Attention$_{\text{seq}}(\cdot)$表示式(12)~(13)描述的注意力融合计算过程.

3.4 睡眠阶段分类

基于最后融合出的片段间转换表征 $\boldsymbol{O}\in\mathbb{R}^m$，本文模型选用两个线性层对其进行线性映射，再采用softmax层计算出模型预测独热标签 $\hat{\boldsymbol{y}}_t\in\mathbb{R}^5, t=(L+1)/2$，该过程为

$$\hat{\boldsymbol{y}}_t=\mathrm{softmax}\left[\boldsymbol{W}_y\left(\mathrm{ReLU}\left(\boldsymbol{W}_O\boldsymbol{O}\right)\right)+\boldsymbol{b}_y\right] \tag{20}$$

其中，$\boldsymbol{W}_y\in\mathbb{R}^{5\times m}$，$\boldsymbol{W}_O\in\mathbb{R}^{m\times m}$，$\boldsymbol{b}_y\in\mathbb{R}^5$ 表示线性层可学习参数；ReLU(·)表示激活函数. 最后，本文通过计算预测标签 $\hat{\boldsymbol{y}}_t\in\mathbb{R}^5$ 与真实标签 $\boldsymbol{y}_t\in\mathbb{R}^5$ 之间的交叉熵来实现模型训练与参数优化. HierFormer模型的损失函数可定义为

$$\begin{aligned}&\mathcal{J}\left(\boldsymbol{X}_1^1,\cdots,\boldsymbol{X}_L^1;\ \cdots;\ \boldsymbol{X}_1^M,\cdots,\boldsymbol{X}_L^M\right)\\&\quad=-\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{5}\left(\left(\boldsymbol{y}_t\right)_i\right)_j\ln\left[\left(\left(\hat{\boldsymbol{y}}_t\right)_i\right)_j\right]\end{aligned} \tag{21}$$

其中，M 表示用于模型训练的输入序列数量.

4 实验与结果分析

4.1 数据集

本文从数据网站PhysioNet(www. physionet. org)下载了开源的睡眠集-欧洲数据格式扩展数据集(Sleep-European Data Format database expanded，SleepEDFx)[33,34]，该权威数据集广泛用于自动睡眠分期模型的训练与验证，其包含197个PSG记录，采集自78个健康受试者(37男41女，25~101岁)和22个具有轻微入睡障碍的受试者(7男15女，18~79岁). 所有PSG记录对应的睡眠阶段均由专业睡眠医师根据Rechtschaffen and Kales(R&K)睡眠分期准则进行标定[35]. 针对每个PSG记录，本文仅保留受试者在夜间睡眠时的数据记录(即夜间熄灯到清晨开灯的时间段)，并去除首尾长时间的清醒阶段信号采集记录[36]. 本文选用EEG Fpz-Cz、EEG Pz-Oz和EOG信号通道进行模型测试，三者的信号采样率均为100 Hz. 为了实现AASM睡眠分期准则定义的五阶段睡眠分类，本文将数据集中标记为S3和S4阶段的样本统一记为N3阶段，并且删除了大体动等无效样本[4]. 经过处理后该数据集的样本类别分布如表1所示.

表1 SleepEDFx数据集各睡眠阶段的30 s PSG片段数量

数据集	W期	N1期	N2期	N3期	R期	总计
SleepEDFx	31 789	23 826	87 317	19 222	34 113	196 267

4.2 实验参数设定

本文针对SleepEDFx数据集(197个PSG记录)采用10折交叉验证方法来对比测试HierFormer模型与基线模型. 具体而言，选用 $\left(N_d-N_d/10\right)$ 个PSG记录用于训练模型，剩余 $N_d/10$ 个PSG记录用于测试模型，N_d 表示数据集PSG记录总数量(即197). 针对每个分期模型，本文重复该过程10次以测试全部PSG记录，并将10次实验结果取平均用于模型性能对比. 所有实验基于英特尔Xeon Gold 6142 2.60 GHz(CPU)、英伟达RTX 3090(GPU)、Python 3.7和Pytorch 1.10实现[37]. 对于STFT计算参数，本文设定FFT计算点数 $N_{\mathrm{FFT}}=256$，由此得到二维时频图的时频序列长度 T=31、频带数 F=129. 对于模型结构参数，本文针对输入序列长度 L 测试多个数值(9、11、13、15)，最优结果长度值为11. 此外，本文根据频带数将时频图映射频带数设为 m=128(模型输入序列特征维度)，以实现可学习的带通滤波效果. 针对三个编码模块内Transformer编码器的层数 N、注意力头数 H、前馈神经网络隐层维度 d_{FF}，本文分别测试多个数值(N：2、3、4，H：2、4、8，d_{FF}：64、128、256、512)，三者最优值为 N=2、H=4、$d_{\mathrm{FF}}=128$. 编码模块内与后续线性层的dropout值设为0.1. 以上Transformer架构相关参数的测试值均小于标准Transformer架构的定义值(dropout值除外)，其目的在于降低模型复杂度(参数量)并提升模型训练效率. 对于模型训练参数，本文在每一折训练过程中针对整体数据集迭代训练50次，对于批训练样本数(batch size)尝试16、32、64共3个数值，训练效果最优值为64，所有模型参数选用Adam优化器进行迭代优化(学习率为 10^{-4}、$\beta_1=0.9$、$\beta_2=0.999$、$\varepsilon=10^{-8}$)[38].

4.3 模型评价指标

本文根据交叉验证预测结果展示了HierFormer模型的分类混淆矩阵，并且计算了各个类别的精确率(precision，PR = TP/(TP + FP))、召回率(recall，RE = TP/(TP + FN))和F1分数(F1-score，F1 = (2 × PR × RE)/(PR + RE))，其中TP、FP、FN分别代表真阳样本数、假阳样本数、假阴样本数. 对于HierFormer模型和其他基线模型的分期性能对比，本文选用的指标包括分类准确率(Accuracy，$\mathrm{ACC}=\sum_{c=1}^{C}\mathrm{TP}_c/N$)、宏平均精确率(Macro-averaging Precision，$\mathrm{MP}=\sum_{c=1}^{C}\mathrm{PR}_c/C$)、宏平均召回率(Macro-averaging Recall，$\mathrm{MR}=\sum_{c=1}^{C}\mathrm{RE}_c/C$)、宏平均F1分数(Macro-averaging F1-score，$\mathrm{MF1}=\sum_{c=1}^{C}\mathrm{F1}_c/C$)、科恩卡帕系数(Cohen's Kappa coefficient，$\kappa=\left(p_j-p_e\right)/\left(1-p_e\right)$)和模型参数量(仅深度学习网络模型). 其中，$\mathrm{TP}_c$、$\mathrm{PR}_c$、$\mathrm{RE}_c$、$\mathrm{F1}_c$ 分别表示第 c 个类别的真阳样本数、精确率、召回率、F1分数，C=5表示睡眠阶段类别数，N 表示测试集PSG片段样本总数，p_j 表示模型预测与人工标定分类结果一致的频率，p_e 表示上述二

者分类判定相同时的概率. 所有分期性能对比指标均由平均值与标准差相结合的方式($\mu\pm\sigma$)进行表示.

4.4　模型睡眠分期性能

HierFormer模型基于SleepEDFx数据集的睡眠分期混淆矩阵如表2所示,其展示了10折交叉验证下模型预测(横轴)与真实标签(纵轴)的结果对比,以及各睡眠阶段对应的准确率PR、召回率RE和F1分数. 由表2结果可得,本文模型对于W、N2、N3、R期的分类结果(F1分数:0.764~0.856)优于N1期(F1分数:0.466). 大量的N1期样本被错误分类为W、N2期,原因在于N1期属于W期和N2期之间的过渡阶段,其可能含有与W期或N2期相似的EEG波形(α波、K复合波等)[39]. 部分N3期样本被归类为N2期,这是因为二者具有相似的EEG睡眠纺锤波. 整体而言,HierFormer模型对于SleepEDFx数据集取得了较为满意的分期结果.

图3展示了HierFormer模型对于SleepEDFx数据集其中一个整夜PSG记录的预测结果与真实标签对比情况. 由图3可知,模型针对该条PSG记录的预测结果曲线与真实标签曲线拟合情况较好(分类准确率ACC:0.931 2,宏平均F1分数:0.892 9). 具体而言,模型在睡眠阶段稳定时期很少出现错误分类,而大多数预测错误情况(黑叉标记)发生在睡眠阶段转换之处. 结合混淆矩阵结果可得,W、N1、N2期之间的相互转换容易出现错误分类,其原因仍在于N1期的转换特性使得三者易存在相似的EEG波形. 整体结果也表明,睡眠过渡时期样本的准确分类是进一步提升模型性能的关键.

表2　HierFormer模型基于SleepEDFx数据集的睡眠分期混淆矩阵

真实\预测	W期	N1期	N2期	N3期	R期	准确率PR	召回率RE	F1分数
W期	25 360	3 023	1 043	43	496	0.847	0.846	0.847
N1期	3 726	9 088	8 781	90	2 071	0.595	0.383	0.466
N2期	302	2 351	79 924	2 799	1 879	0.803	0.916	0.856
N3期	47	14	5 441	13 716	1	0.821	0.714	0.764
R期	514	794	4 375	60	28 359	0.864	0.832	0.848

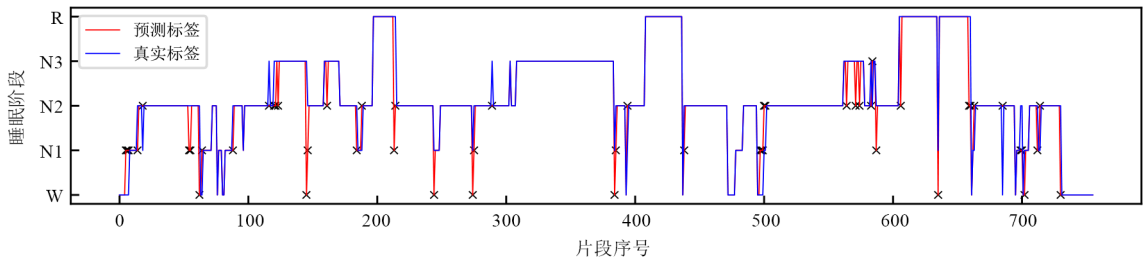


图3　睡眠分期整夜对比(SC4061EC-Hypnogram.edf,ACC: 0.931 2,MF1: 0.892 9)

4.5　基线模型性能对比

本文基于SleepEDFx数据集,上述实验参数设定复现了11种已发表的基线模型方法,以对比验证HierFormer模型的有效性. 表3展示了各个方法的算法架构、输入信号、分期性能和模型参数量.

从表3结果可看出,对于三种单通道信号输入的方法(即文献[8]、DeepSleepNet[14]、SleepTransformer[29]),两种深度神经网络模型相比于信号特征集+RF方法取得更优异的分期性能;Transformer架构相对于CNN架构参数量更少,更为准确地提取了EEG有效波形特征,进而提升分类效果. 但是相较于大部分多通道PSG信号输入的方法,上述三种方法的分期性能还存在一定差距. 对于多通道信号输入方法,文献[7]采用的信号特征集+LR方法取得了优于多种神经网络模型的分期结果,这表明有效构建特征集也可以使传统机器学习方法取得良好性能,但该方法特征集包含上千个信号参数,算法设计效率较低. GraphSleepNet[23]虽然基于多通道信号设计模型,但其仅提取原始PSG的微分熵特征(Differential Entropy,DE)输入网络模型,部分有效波形信息的缺失导致该方法结果较差. 其他多通道信号输入的神经网络基线模型均取得较好的分期性能,其中以GRU架构为主体的模型(SAGSleepNet[3]、RobustSleepNet[20]等)性能略优于CNN架构模型(文献[21]、HybridAtt[22]),这说明RNN架构的时序特征编码属性相较于CNN架构的局部特征编码属性更适用于PSG这类时序信号. 此外,注意力机制的引入也使上述神经网络能够聚焦关键的有效波形信息或睡眠转换信息,进一步改善模型分类结果. 值得注意的是,SAGSleepNet[3]与RobustSleepNet[20]的分期性能接近HierFormer模型,甚至宏平均精确率MP更高,其原因在于两种基线采用的GRU架构和本文模型采用的Transformer编码器架构均有效提取了PSG内关键的片段内

波形特征和片段间转换特征. 但结合准确率 ACC、宏平均 F1 分数和模型参数量综合来看,HierFormer 模型在采用 Transformer 编码器架构替代传统的 GRU 架构后,利用更少的参数量取得了整体更优异的分期性能(分类准确率 ACC:0.807±0.038,宏平均 F1 分数:0.750±0.046),这说明本文模型基于 Transformer 编码器和注意力机制充分聚焦了 PSG 在片段内、通道间、片段间三种视角下的有效波形信息、通道相关性信息、睡眠阶段转换信息,验证了该模型的有效性. 另外,本文模型的科恩卡帕系数 κ(0.721±0.054)处于 0.61~0.80,达到了高度一致性的分类性能级别.

表 3 HierFormer 模型与基线模型基于 SleepEDFx 数据集的睡眠分期性能对比

方法	输入信号	准确率 ACC	宏平均精确率 MP	宏平均召回率 MR	宏平均 F1 分数	科恩卡帕系数 κ	模型参数量
文献[8] (RF)	Pz-Oz	0.694±0.044	0.635±0.047	0.597±0.052	0.599±0.052	0.560±0.060	—
文献[7] (LR)	PSG	0.796±0.044	0.749±0.044	0.764±0.038	0.750±0.039	0.715±0.059	—
DeepSleepNet[14] (CNN+LSTM)	Fpz-Cz	0.711±0.080	0.712±0.068	0.630±0.097	0.630±0.115	0.593±0.108	2.671×10^7
SleepTransformer[29] (双层级 Transformer)	Fpz-Cz	0.780±0.039	0.742±0.054	0.703±0.040	0.712±0.046	0.681±0.053	1.472×10^6
文献[21] (CNN)	PSG	0.781±0.043	0.759±0.038	0.702±0.051	0.718±0.048	0.686±0.060	1.917×10^4
HybridAtt[22] (CNNAtt+GRUAtt)	PSG	0.786±0.045	0.760±0.045	0.701±0.070	0.711±0.068	0.691±0.065	5.966×10^5
GraphSleepNet[24] (STGCNAtt)	PSG	0.747±0.049	0.696±0.070	0.658±0.055	0.665±0.061	0.637±0.066	2.338×10^4
SAGSleepNet[3] (GRUAtt+GCN+GRU)	PSG	0.804±0.036	0.801±0.031	0.721±0.050	0.741±0.049	0.717±0.051	3.279×10^5
SeqSleepNet[17] (GRUAtt+GRU)	PSG	0.779±0.030	0.786±0.044	0.675±0.052	0.685±0.061	0.678±0.040	1.581×10^5
SimpleSleepNet[18] (GRUAtt(PE)+GRU)	PSG	0.786±0.038	0.754±0.054	0.711±0.049	0.723±0.052	0.692±0.053	6.401×10^4
RobustSleepNet[20] (GRUAtt+GRU)	PSG	0.806±0.033	0.809±0.026	0.707±0.054	0.729±0.058	0.718±0.046	2.548×10^5
HierFormer (多层级 TransformerAtt)	PSG	0.807±0.038	0.784±0.041	0.735±0.046	0.750±0.046	0.721±0.054	2.328×10^5

4.6 特征编码方式性能对比

为了验证 HierFormer 模型中 Transformer 编码器架构在片段内、通道间、片段间视角下特征编码的有效性,本文在固定其中两种视角编码模块的基础上,将剩余视角模块的 Transformer 编码器分别替换为两种不同的特征编码方式(均参考基线模型),再基于 SleepEDFx 数据集、实验参数设定、模型评价指标进行相同的训练与测试过程,最后展示各方式的分期性能与模型参数量.

4.6.1 片段内波形特征编码

对于模型片段内波形特征编码方式,本文将 Transformer 编码器替换为 CNN 或 RNN 架构,模型其余结构不变. 其中,CNN 架构采用 Wav2Vec 卷积特征编码模块[40],基于 7 层卷积对原始 PSG 提取高维特征;RNN 架构采用 GRU 编码模块[3],将 STFT 时频图经过线性映射后的特征序列输入 GRU 得到高维特征. 随后利用注意力机制融合相应的高维特征得到片段内波形表征. 三种编码方式的分期性能对比如表 4 所示.

由表 4 结果可知,GRUAtt 编码方式的分期性能优于 Wav2VecAtt 方式,这说明相较于从原始 PSG 角度编码局部卷积特征,基于时频图角度编码 PSG 时序特征能更有效地表征片段内有效波形时序信息,即 RNN 比 CNN 架构更适用于 PSG 编码. 而相比于 RNN 架构,本文采用的 Transformer 编码器利用更少的模型参数量进一步提升了分期性能,这表明 Transformer 结构同样具有很强的时序编码能力,其内部的自注意力机制有效抓取了时频图特征序列内各个特征向量之间的上下文时序相关性信息.

表4　基于SleepEDFx数据集不同片段内波形特征编码方式的睡眠分期性能对比

编码方式	准确率ACC	宏平均精确率MP	宏平均召回率MR	宏平均F1分数	科恩卡帕系数κ	模型参数量
GRUAtt[3]	0.803±0.038	0.780±0.043	0.728±0.049	0.742±0.050	0.715±0.054	3.818×10^5
Wav2VecAtt[40]	0.773±0.038	0.757±0.051	0.684±0.039	0.705±0.045	0.671±0.051	4.991×10^5
TransformerAtt(本文)	0.807±0.038	0.784±0.041	0.735±0.046	0.750±0.046	0.721±0.054	2.328×10^5

4.6.2　通道间相关性特征编码

对于模型通道间相关性特征编码方式,本文将Transformer编码器和注意力机制替换为特征拼接或自注意力特征融合,模型其余结构不变. 其中,特征拼接Concat将多通道片段内波形表征直接拼接,再经过线性映射得到通道相关性表征[21];自注意力特征融合SelfAtt对多通道片段内表征计算上下文相关性特征后,结合注意力机制融合出通道相关性表征[22]. 三种编码方式的分期性能对比如表5所示.

由表5可知,相较于仅采用自注意力机制的特征融合方式SelfAtt,本文Transformer编码器与注意力机制结合的方式较为明显地提升分期性能,体现Transformer编码器对于各通道片段内表征之间的相关性信息编码能力. 相比于特征拼接方式Concat,本文方式取得了更好的宏平均精确率MP、召回率MR和F1分数,这意味着HierFormer模型在应对类别不平衡的数据集时性能更为稳定. 而二者准确率ACC和科恩卡帕系数κ结果持平,其原因在于SleepEDFx数据集的PSG通道数量仅为3,导致通道相关性信息量较低,在模型内其他特征编码模块不变的情况下,特征向量直接拼接与Transformer架构计算特征向量自注意力相关性的效果相近. 但考虑到模型参数量的问题,当通道数量增加时,特征拼接方式可能会迅速拉开与本文方式的差距并出现维度灾难,从而影响模型性能.

表5　基于SleepEDFx数据集不同通道间相关性特征编码方式的睡眠分期性能对比

编码方式	准确率ACC	宏平均精确率MP	宏平均召回率MR	宏平均F1分数	科恩卡帕系数κ	模型参数量
SelfAtt[22]	0.794±0.038	0.771±0.039	0.714±0.053	0.730±0.049	0.701±0.054	2.332×10^5
Concat[21]	0.807±0.040	0.782±0.046	0.729±0.051	0.743±0.051	0.721±0.056	2.820×10^5
TransformerAtt(本文)	0.807±0.038	0.784±0.041	0.735±0.046	0.750±0.046	0.721±0.054	2.328×10^5

4.6.3　片段间转换特征编码

对于模型片段间转换特征编码方式,本文将Transformer编码器替换为CNN或RNN架构,模型其余结构不变. 其中,CNN架构采用TCN编码模块[19],将通道相关性表征序列输入TCN得到高维特征;RNN架构采用LSTM编码模块[14],将通道相关性表征序列输入LSTM得到高维特征. 随后利用注意力机制融合高维特征序列得到片段间转换表征. 三种编码方式的分期性能对比如表6所示.

由表6可知,本文TransformerAtt编码方式和TCNAtt方式的分期结果均好于LSTMAtt方式,这表明相比于关注时序相关性的LSTM来说,TCN聚焦全局依赖性与Transformer编码上下文时序相关性的结构特点,更适用于处理PSG片段序列的睡眠时序转换信息. 虽然TCNAtt方式的分期性能接近于本文方式,但其模型参数量却5倍于后者. 整体而言,相比于传统的CNN和RNN架构,Transformer编码器凭借更少的参数量有效抓取了序列内各个PSG片段之间的睡眠转换信息,从而取得最优分期性能.

表6　基于SleepEDFx数据集不同片段间转换特征编码方式的睡眠分期性能对比

编码方式	准确率ACC	宏平均精确率MP	宏平均召回率MR	宏平均F1分数	科恩卡帕系数κ	模型参数量
LSTMAtt[14]	0.796±0.038	0.762±0.050	0.727±0.043	0.736±0.046	0.707±0.052	4.315×10^5
TCNAtt[19]	0.805±0.039	0.778±0.040	0.734±0.049	0.747±0.045	0.719±0.056	1.236×10^6
TransformerAtt(本文)	0.807±0.038	0.784±0.041	0.735±0.046	0.750±0.046	0.721±0.054	2.328×10^5

4.7　注意力分数可视化

为了体现HierFormer模型在片段内、通道间、片段间视角下特征编码的可解释性,本文选取SleepEDFx数据集中某一PSG记录(SC4001E0-PSG.edf)内一个片段序列(长度为11)作为模型输入,将计算过程中三个视角特征编码的注意力分数进行可视化,同时展示原始输入信号与STFT时频图,以验证模型如何正确实现睡眠分期. 针对该序列的可视化结果如图4所示,序列长度表示共有11个时间戳(图片底部),包含中间时刻t的PSG片段与前后相邻5个片段,每个片段标记相应的睡眠阶段(图片顶部);从下至上第1~3行展示了三通道原始PSG,第4~6行展示了相对应的STFT时频图,第7~9行罗列了时频图对应的片段内波形特征编码注意力分数曲线,第10行绘制了三个通道间的相关性特征编码注意力分数柱状图,第11行展示了各片段的片段间转换特征编码注意力分数柱状图.

从原始PSG绘制结果(从下至上第1~3行)可看出,序列前半段EOG信号波动较为明显,EEG信号(Fpz-Cz和Pz-Oz)波形较为平稳,因为W与N1期属于由清醒到睡眠的入睡过程,此时眼球活动频繁;序列后半段EOG信号波动减弱而EEG信号出现明显的波动,其原因在于N2期属于睡眠过程,此时眼球活动减弱,大脑皮层神经元放电活动增强,出现*K*复合波或睡眠梭形波[4]. 三通道STFT时频图(从下至上第4~6行,各子图横轴为时间(30 s),纵轴为频带(0~12.5 Hz),颜色块表示幅值,亮度越高代表幅值越高)表明,EEG与EOG信号活动均处在低频波段,即红黄颜色块大多处于图片底部;在W→N1→N2期转换过程中,EOG时频图低频波段高能量谱逐渐减弱,对应原始信号波幅变化,EEG时频图低频波段颜色由暗到亮,能量值升高,对应原始信号出现的EEG波形. 与时频图共享时间轴的片段内波形注意力分数曲线(从下至上第7~9行)体现出,HierFormer模型有效聚焦各个时频图内出现的高能量谱时刻,即EEG与EOG信号波动位置,分配了较高的注意力权重;综合所有注意力分数曲线图可看出,序列前期EOG通道注意力权重分配相比EEG通道更为复杂,对应原始EOG的复杂波形与EEG的平稳波形,序列后期EOG通道注意力分数曲线趋于平缓而EEG通道趋于复杂,对应大量的EEG有效波形以及EOG信号波幅的减弱. 从片段内通道注意力权重柱状图(从下至上第10行)可看出,W与N1期片段的EEG Pz-Oz和EOG通道权重较高,说明模型聚焦于这些时段的眼球活动与α脑电波,而N2期片段的EEG Fpz-Cz通道权重较高,表明其存在*K*复合波或睡眠梭形波;EOG通道分数的逐渐降低与EEG通道的逐渐提升同样体现了W→N1→N2期的转换过程. 各片段注意力分数柱状图(从下至上第11行)体现出,对于中间*t*时刻的N1期片段,模型主要关注前期W与N1期片段的睡眠阶段信息,后续N2期片段则权重较低,原因在于该片段衔接前面的N1期时段,与序列前期的时序关联性比序列后期更高. 综上所述,本文模型有效聚焦了PSG片段内、通道间、片段间视角的关键信息,体现良好的可解释性.

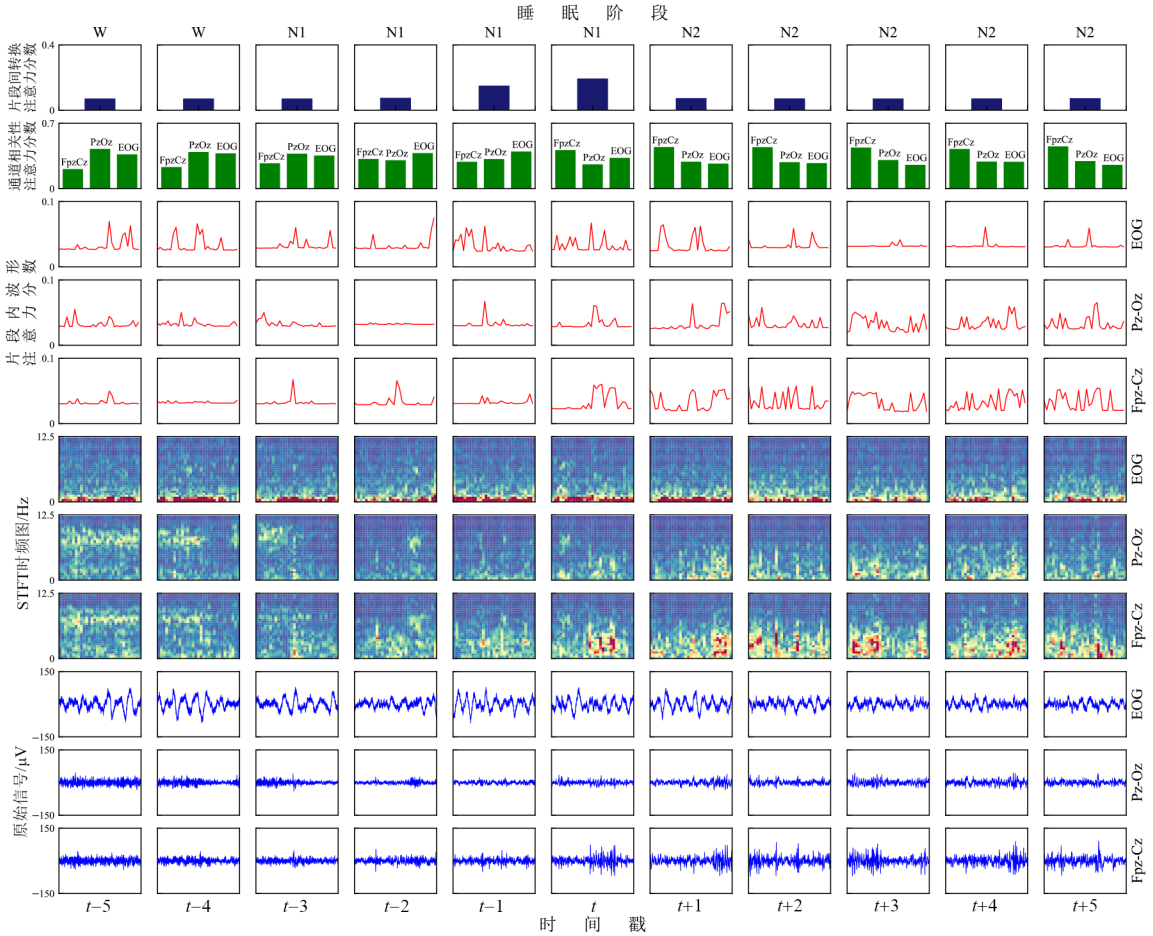


图4 HierFormer模型基于PSG记录(SC4001E0-PSG.edf)内某一片段序列(长度为11)的注意力分数可视化结果

5　讨论

现有自动睡眠分期方法大多采用传统的CNN、RNN架构来提取PSG片段内、通道间或片段间视角下的有效特征. 结合表3可看出,SimpleSleepNet[18]、RobustSleepNet[20]等RNN类模型的分期性能优于HybridAtt[22]等CNN类模型,说明RNN时序建模比CNN卷积运算更适合编码时序特性显著的PSG. 而同样以时序建模能力著称的Transformer架构采用自注意力机制对序列数据进行并行处理,相较于RNN能够利用更少的模型参数量来高效捕获序列时序关联性,因此,本文引入Transformer编码器作为主体结构来抓取PSG片段序列的片段内波形特征、通道间相关性特征、片段间睡眠转换特征. 表3结果说明,HierFormer模型相对于多种基线方法取得了最优睡眠分期性能,如分类准确率ACC为0.807、宏平均F1分数为0.750等,证明了该模型对于未来推动人工智能与睡眠分期相结合的有效性.

医生在睡眠分期决策时会综合考虑PSG片段序列的片段内波形信息、通道间相关性信息、片段间睡眠转换信息. 然而现有大多睡眠分期模型却忽略了其中一种信息,例如GraphSleepNet[23]提取的DE特征未考虑片段内波形信息,DeepSleepNet[14]或SleepTransformer[29]采用的单通道信号输入无法抓取通道间相关性信息,文献[21]采用的CNN架构未考虑片段间时序转换信息. 结合三种视角特征编码方式对比实验可看出,HierFormer模型在充分考虑上述三种关键信息并利用模型参数量更少的Transformer编码器与注意力机制提取有效特征后,取得了优于基线模型和基线编码方式的睡眠分期结果.

睡眠分期模型较低的可解释性一直是阻碍其用于临床诊断的关键,其原因在于精确实现睡眠疾病诊疗或健康监测需细致观察PSG有效信息. 本文利用注意力机制体现模型如何聚焦PSG各视角有效信息,结合权重可视化实验可看出,HierFormer模型关注了片段内有效波形信息、通道间相关性信息和片段间睡眠阶段转换信息,较大程度地贴合了AASM分期准则. 基于已有睡眠分期研究成果,现有模型还无法替代睡眠医生实现睡眠分期过程,难以消除医生或患者的怀疑态度. 根据本文思路,未来深度学习分期模型可主要起到辅助作用,即模型分期结果由医生最终裁决,从而提升诊疗效率.

虽然本文模型的分期性能与可视化结果较为满意,但仍存在一些缺陷需要改善. HierFormer模型主要基于Transformer编码器构建,Transformer架构相较于CNN、RNN等架构更为复杂,若模型参数设置不当,会导致模型训练过程占用大量计算内存. 此外,训练数据量的匮乏会导致Transformer架构无法发挥出更优异的性能,这也是深度神经网络一直存有的缺陷. 综上所述,优化模型结构以降低内存占用和利用小规模数据提升模型性能将会是未来研究的重点.

6　结论

本文针对PSG提出一种基于Transformer架构的端到端多层级自动睡眠分期模型,即HierFormer模型. 该模型引入Transformer编码器替代传统的CNN或RNN架构,按递进关系编码多通道PSG片段序列的片段内波形信息、通道间相关性信息、片段间睡眠转换信息,同时结合注意力机制提升模型在三种视角下针对PSG信号特性的模型可解释性. 实验结果表明:本文模型利用更少的模型参数量取得了优于多种基线模型的睡眠分期性能,Transformer编码器在三种视角下的性能优于多种基线特征编码方式. 此外,注意力可视化实验证明了本文模型在睡眠分期时的有效性与可解释性. 本文研究为深度学习在睡眠分期领域的发展提供创新技术,对未来改善睡眠疾病诊疗做出积极贡献.

参考文献

[1] TAI C H, LIAO T Y, CHEN S P, et al. Sleep stage classification using Light Gradient Boost Machine: Exploring feature impact in depressive and healthy participants[J]. Biomedical Signal Processing and Control, 2024, 88: 105647.

[2] 金峥, 贾克斌, 袁野. 基于混合注意力时序网络的睡眠分期算法研究[J]. 生物医学工程学杂志, 2021, 38(2): 241-248.
JIN Z, JIA K B, YUAN Y. A hybrid attention temporal sequential network for sleep stage classification[J]. Journal of Biomedical Engineering, 2021, 38(2): 241-248. (in Chinese)

[3] JIN Z, JIA K B. SAGSleepNet: A deep learning model for sleep staging based on self-attention graph of polysomnography[J]. Biomedical Signal Processing and Control, 2023, 86: 105062.

[4] IBER C, ANCOLI-ISRAEL S, CHESSON A J, et al. The AASM manual for the scoring of sleep and associated events[M]. Westchester: American Academy of Sleep Medicine, 2007.

[5] ZHANG L D, FABBRI D, UPENDER R, et al. Automated sleep stage scoring of the Sleep Heart Health Study using deep neural networks[J]. Sleep, 2019, 42(11): zsz159.

[6] SHAHBAKHTI M, BEIRAMVAND M, EIGIRDAS T, et al. Discrimination of wakefulness from sleep stage I using nonlinear features of a single frontal EEG channel[J]. IEEE Sensors Journal, 2022, 22(7): 6975-6984.

[7] VAN DER DONCKT J, VAN DER DONCKT J, DEPROST E, et al. Do not sleep on traditional machine learning Simple and interpretable techniques are competitive to deep learning for sleep scoring[J]. Biomedical Signal Processing and Control, 2023, 81: 104429.

[8] MEMAR P, FARADJI F. A novel multi-class EEG-based sleep stage classification system[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, 26(1): 84-95.

[9] SATAPATHY S K, BHOI A K, LOGANATHAN D, et al. Machine learning with ensemble stacking model for automated sleep staging using dual-channel EEG signal[J]. Biomedical Signal Processing and Control, 2021, 69: 102898.

[10] HUANG J, REN L F, FENG L F, et al. AI empowered virtual reality integrated systems for sleep stage classification and quality enhancement[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30: 1494-1503.

[11] KARIMZADEH F, BOOSTANI R, SERAJ E, et al. A distributed classification procedure for automatic sleep stage scoring based on instantaneous electroencephalogram phase and envelope features[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, 26(2): 362-370.

[12] SEKKAL R N, BEREKSI-REGUIG F, RUIZ-FERNANDEZ D, et al. Automatic sleep stage classification: From classical machine learning methods to deep learning[J]. Biomedical Signal Processing and Control, 2022, 77: 103751.

[13] ZAIDI T F, FAROOQ O. EEG sub-bands based sleep stages classification using Fourier Synchrosqueezed transform features[J]. Expert Systems with Applications, 2023, 212: 118752.

[14] SUPRATAK A, DONG H, WU C, et al. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017, 25(11): 1998-2008.

[15] ZHOU D D, XU Q, WANG J, et al. LightSleepNet: A lightweight deep model for rapid sleep stage classification with spectrograms[C]//The 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Piscataway: IEEE, 2021: 43-46.

[16] PERSLEV M, JENSEN M H, DARKNER S, et al. U-time: A fully convolutional network for time series segmentation applied to sleep staging[C]//The 33rd Conference on Neural Information Processing Systems (NeurIPS). California: NIPS, 2019: 4415-4426.

[17] PHAN H, ANDREOTTI F, COORAY N, et al. SeqSleepNet: End-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2019, 27(3): 400-410.

[18] GUILLOT A, SAUVET F, DURING E H, et al. Dreem open datasets: Multi-scored sleep datasets to compare human and automated sleep staging[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2020, 28(9): 1955-1965.

[19] KHALILI E, MOHAMMADZADEH ASL B. Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG[J]. Computer Methods and Programs in Biomedicine, 2021, 204: 106063.

[20] GUILLOT A, THOREY V. RobustSleepNet: Transfer learning for automated sleep staging at scale[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 1441-1451.

[21] CHAMBON S, GALTIER M N, ARNAL P J, et al. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2018, 26(4): 758-769.

[22] YUAN Y, JIA K B, MA F L, et al. A hybrid self-attention deep learning framework for multivariate sleep stage classification[J]. BMC Bioinformatics, 2019, 20(Suppl 16): 586.

[23] JIA Z Y, LIN Y F, WANG J, et al. GraphSleepNet: Adaptive spatial-temporal graph convolutional networks for sleep stage classification[C]//The 29th International Joint Conference on Artificial Intelligence. California: IJCAI, 2020: 1324-1330.

[24] JIN Z, JIA K B. A temporal multi-scale hybrid attention network for sleep stage classification[J]. Medical & Biological Engineering & Computing, 2023, 61(9): 2291-2303.

[25] ELDELE E, CHEN Z H, LIU C Y, et al. An attention-based deep learning approach for sleep stage classification with single-channel EEG[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 809-818.

[26] WANG H F, LU C G, ZHANG Q, et al. A novel sleep staging network based on multi-scale dual attention[J]. Biomedical Signal Processing and Control, 2022, 74:

103486.

[27] PATHAK S, LU C Q, NAGARAJ S B, et al. STQS: Interpretable multi-modal Spatial-Temporal-seQuential model for automatic Sleep scoring[J]. Artificial Intelligence in Medicine, 2021, 114: 102038.

[28] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//The 31st Annual Conference on Neural Information Processing Systems (NeurIPS). California: NIPS, 2017: 5999-6009.

[29] PHAN H, MIKKELSEN K, CHÉN O Y, et al. SleepTransformer: Automatic sleep staging with interpretability and uncertainty quantification[J]. IEEE Transactions on Biomedical Engineering, 2022, 69(8): 2456-2467.

[30] ZHANG W J, LI C, PENG H, et al. CTCNet: A CNN Transformer capsule network for sleep stage classification[J]. Measurement, 2024, 226: 114157.

[31] CHEN Z, YANG Z W, ZHU L W, et al. Automated sleep staging *via* parallel frequency-cut attention[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 1974-1985.

[32] PENG L, REN Y Z, LUAN Z H, et al. SleepViTransformer: Patch-based sleep spectrogram transformer for automatic sleep staging[J]. Biomedical Signal Processing and Control, 2023, 86: 105203.

[33] GOLDBERGER A L, AMARAL L A, GLASS L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals[J]. Circulation, 2000, 101(23): E215-E220.

[34] KEMP B, ZWINDERMAN A H, TUK B, et al. Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG[J]. IEEE Transactions on Biomedical Engineering, 2000, 47(9): 1185-1194.

[35] WOLPERT E A. A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects[J]. Archives of General Psychiatry, 1969, 20(2): 246-247.

[36] IMTIAZ S A, RODRIGUEZ-VILLEGAS E. An open-source toolbox for standardized use of PhysioNet Sleep EDF Expanded Database[C]//The 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Piscataway: IEEE, 2015: 6014-6017.

[37] PASZKE A, GROSS S, CHINTALA S, et al. Automatic differentiation in pytorch[C]//The 31st Annual Conference on Neural Information Processing Systems (NeurIPS). California: NIPS, 2017: 1-4.

[38] KINGMA D P, BA J, HAMMAD M M. Adam: A method for stochastic optimization[EB/OL]. (2017-01-30) [2024-06-25]. https://arxiv.org/abs/1412.6980v9.

[39] DAVIES H J, NAKAMURA T, MANDIC D P. A transition probability based classification model for enhanced N1 sleep stage identification during automatic sleep stage scoring[C]//The 41st Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Piscataway: IEEE, 2019: 3641-3644.

[40] BAEVSKI A, ZHOU Y H, MOHAMED A, et al. Wav2vec 2.0: A framework for self-supervised learning of speech representations[C]//The 34th Conference on Neural Information Processing Systems (NeurIPS). California: NIPS, 2020: 12449-12460.

作者简介

金　峥　男，1997年3月出生于北京市. 2019年获得北京工业大学信息学部电子信息工程专业学士学位. 目前在北京工业大学信息科学技术学院电子科学与技术专业攻读博士学位. 主要研究方向为时序信号处理(主要为生物医学信号)、机器学习、数据挖掘.

E-mail: zhengj@emails.bjut.edu.cn

贾克斌　男,1962年8月出生于新疆维吾尔自治区乌鲁木齐市. 分别于1990年和1998年获得中国科技大学信息与通信工程专业工学硕士学位和博士学位. 现为北京工业大学信息科学技术学院教授、博士生导师,并担任数字多媒体信息处理与成像技术研究团队领导人. 主要研究方向为数据挖掘、模式识别、信号处理、机器学习、生物信息处理等.

E-mail: kebinj@bjut.edu.cn