

一种预测驾驶注意力的多尺度注意力模型

蒋超, 郜东瑞, 李芃锐, 赵长名

(成都信息工程大学 计算机学院, 四川 成都 610225)

摘要: 近年来,许多研究致力于利用EEG与EOG多模态数据预测驾驶注意力,但有效融合这两种模态数据仍是一项充满挑战的任务。为此,提出一个基于多模态的多尺度通道注意力回归模型(MMCAR-Net)预测驾驶注意力。首先,通过多尺度感知单元(Multi-Scale Inception)从EEG、EOG模态数据中分别提取多尺度特征;其次,在多个尺度上有序合并EEG与EOG特征以增强融合特征的多样性;最后,引入多尺度通道注意力机制为多尺度特征赋予差异化权重,以强化与注意力预测相关的特征,提升模型对驾驶注意力相关特征的敏感性和表达能力。在SEED-VIG数据集上的实验表明,所提模型在个体内实验组中取得的PCC与RMSE分别为0.959和0.064,在跨被试实验组中对应数值为0.892和0.112。

关键词: 驾驶注意力预测;多尺度感知单元;多尺度通道注意力机制;特征融合

DOI:10.11907/rjtk.241158

开放科学(资源服务)标识码(OSID):

中图分类号:TP391

文献标识码:A

文章编号:1672-7800(2025)004-0018-07



A Multi-Scale Attention Model for Driving Attention Prediction

JIANG Chao, GAO Dongrui, LI Pengrui, ZHAO Changming

(School of Computer Science, Chengdu University of Information Science and Technology, Chengdu 610225, China)

Abstract: In recent years, many studies have been dedicated to the prediction of driving attention by utilizing multimodal data derived from EEG and EOG. However, effectively fusing these two modal data remains a challenging task. In this paper, a methodology named the multi-modal multiscale channel attention regression network for driving attention prediction was introduced. The model employs a multi-scale inception unit to extract multi-scale features from both EEG and EOG data. Subsequently, it sequentially merges EEG and EEG features at different scales, enhancing the diversity of fused features. Concurrently, a multi-scale channel attention mechanism is introduced to assign differentiated weights to multi-scale features to strengthen the features related to attention prediction, and to improve the model's sensitivity and expressiveness to driving attention-related features. Through a series of experiments conducted on the SEED-VIG dataset, the proposed model achieved a PCC of 0.959 and RMSE of 0.064 in the intra-individual experimental group, and corresponding values of 0.892 and 0.112 in the cross-subjects experimental group, showcasing exceptional performance.

Key Words: driving attention prediction; multi-scale inception unit; multiscale channel attention mechanism; feature fusion

0 引言

一个人在专注于某一事物或活动时体现出的心理状态被称为注意力^[1]。当注意力分散时,驾驶者在驾驶车辆

时会感到疲倦、判断力下降,进而导致交通事故^[2]。国家统计局数据显示,驾驶注意力不集中是导致重大事故的三大主因之一,占比超过40%^[3,4]。因此,准确检测并预警司机的注意力状态对维护道路安全至关重要。

注意力预测会对对被试者的面部表情、肢体动作和生

收稿日期:2024-02-26

扫描二维码阅读全文:



基金项目:国家重点研发计划项目(2021YFF1200605);LOST 2030 脑科学项目(2022ZD0208500);成都信息工程大学科研基金项目(KYQN202208、KYQN202206);四川省科技计划项目(2023NSFSC0499);国家自然科学基金项目(62272067)

作者简介:蒋超(1997-),男,成都信息工程大学计算机学院硕士研究生,研究方向为人工智能、脑机接口;郜东瑞(1987-),男,博士,成都信息工程大学计算机学院副教授、硕士生导师,研究方向为人工智能、脑机接口;李芃锐(1998-),男,成都信息工程大学计算机学院硕士研究生,研究方向为人工智能、脑机接口;赵长名(1984-),男,博士,成都信息工程大学计算机学院副教授,研究方向为边缘计算、高性能分布式计算。本文通讯作者:郜东瑞。

理信号等数据进行特征提取和回归^[5]。在多种模态信号中,生理信号能更准确地反映被试者的注意力状态^[6]。其中,脑电图(Electroencephalogram, EEG)提供了客观、功能性的脑活动映射,被视为最直接、有效的用于预测驾驶员注意力的信号^[7,8]。同样,眼电图(electro-oculogram, EOG)记录了人眼位置的电位数据,携带了关于个体的注意力信息^[9]。目前,无线和可穿戴设备可收集EEG、EOG数据,以实时预测驾驶员的注意力状态^[10]。

近年来,在注意力预测的研究领域,大多数研究侧重于采用单一模态的EEG数据进行注意力预测,并在此方向上取得了不错的效果^[11,12]。Du等^[13]提出的多尺度卷积模型IENet在脑电数据处理中取得了显著效果,使用了3个不同尺度的卷积结构丰富特征多样性,同时引入一个并行的最大池化函数分支和一个并行的一维卷积函数分支来有效提升模型抗干扰能力。Gao等^[14]提出的CSF-GTNet通过创新的融合策略解决了脑电信号不稳定性 and 复杂性问题。在模拟和真实驾驶环境中,Zheng等^[15]的方法取得了出色的平均相关系数,分别为71.18%和66.20%。

然而,人体注意力缺失是一个涉及复杂生理过程的现象,使用单模态数据进行预测时存在可靠性不足的问题^[16]。研究表明,将EEG、EOG两种模态进行多模态融合后用于注意力预测具有显著优势^[17,18]。Zheng等^[19]提出一种结合EEG与EOG的多模态方法预测驾驶注意力,可显著提升性能。Wu等^[20]提出一种具有子网络神经元的早期融合方法,将EEG与EOG数据特征合并成一个单一表示,但未能充分挖掘各模态数据的显著特征。Zhang等^[21]采用长短时记忆网络(Long Short-Term Memory, LSTM)网

络处理EOG的36维特征,取得了不错的效果,但所用特征缺乏明显的时间序列,因此使用LSTM结构可能存在一些缺陷。邱达锋^[22]提出一个用于驾驶注意力检测的特征融合网络,在SEED-VIG上取得了良好的准确率,但该方法仅从单一尺度对EEG与EOG数据进行特征融合,未充分考虑多模态融合特征的多样性。

为此,本文提出一个基于多模态的多尺度通道注意力回归模型(MMCAR-Net)。首先在文献[13]启发下,本文模型多尺度感知单元包含了一个浅层和深层结构。其中,深层结构仍由3个不同尺度的卷积函数构成,浅层结构包含两个不同尺度的卷积函数。这一设计的目的在于从局部和全局层面提取特征。其次受文献[20]的启发,本文模型采用多个多尺度感知单元堆叠以深化特征提取模块,旨在增强提取特征的显著性,使模型有效学习每个模态的抽象特征。再次参考文献[22]对EEG、EOG数据进行了5个不同尺度的特征提取,以有序合并这些尺度的特征来全面融合多模态数据。最后,引入多尺度通道注意力机制学习融合后的多模态特征。综上,本文所提设计旨在丰富融合特征多样性,提升模型对驾驶注意力相关特征的敏感性和表达能力。

1 本文方法

1.1 模型总体架构

图1为本文模型整体结构,Multi-Scale Inception代表多尺度感知单元,Multi-Scale Cooperative Module表示多尺度协作模块,Multi-Scale Channel Attention Module表示多

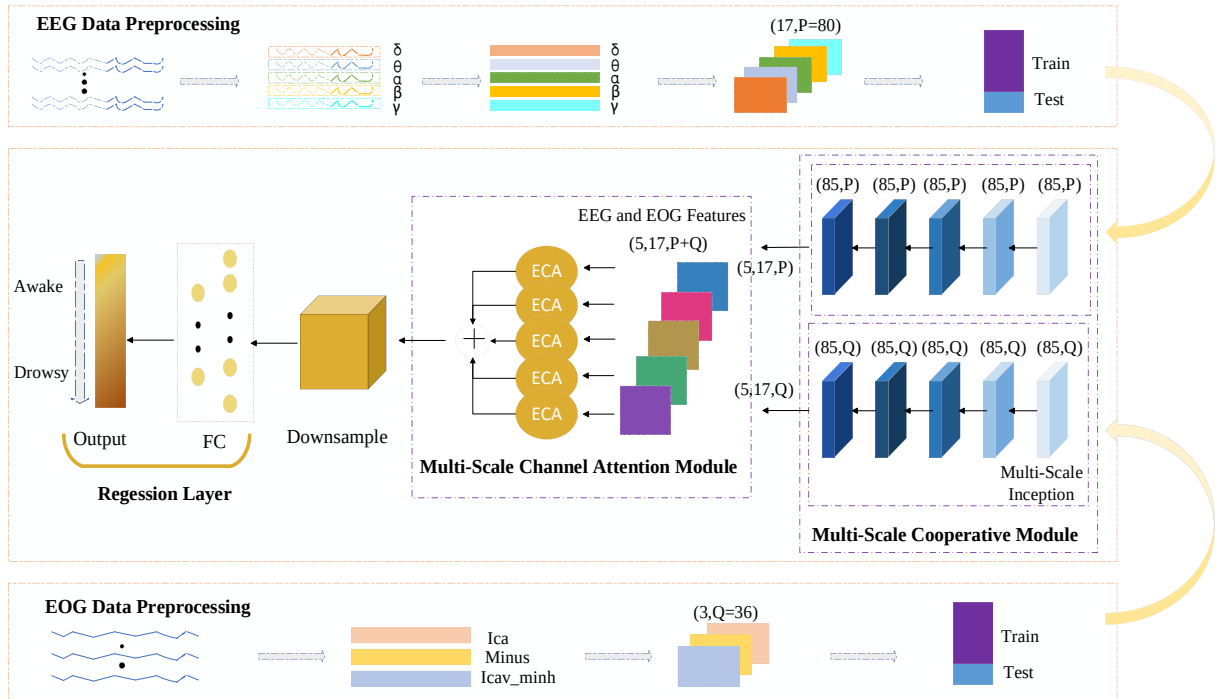


Fig. 1 Overall model architecture

图1 整体模型架构

尺度通道注意力模块, ECA 为高效通道注意力模块, Downsample 为降采样模块, FC 表示全连接层, Output 表示模型的预测结果, FC 和 Output 组成了模型的回归层。数据处理过程为: 首先预处理被试者的 EEG 数据后保存为 $N \times 17 \times 80$ (样本数 $\times$ 电极通道数 $\times$ 微分熵特征); 其次通过合并频带数和微分熵特征全面描绘脑电信号的时域、频域特征; 最后输出预处理后的 EOG 数据格式为 $N \times 3 \times 36$ (样本数 $\times$ 方法数 $\times$ 眼电特征)。

由图 1 可见, 首先多尺度协作模块从 EEG 和 EOG 数据中提取了 5 个不同尺度的特征进行有序合并, 丰富了融合特征的多样性, 构建了更全面、综合的特征表示; 其次将融合特征输入多尺度通道注意力模块, 通过通道加权强化关键的通道; 再次将处理后的各尺度特征进行均值融合以综合不同尺度的信息, 以进一步提升模型的表达能力和泛化能力; 最终将融合后的均值特征依次送入降采样模块进行处理, 再输入回归层得到最终预测结果。

1.2 脑电数据预处理

本文对 EEG 数据的预处理分为频率分段、微分熵特征提取两个步骤。EEG 数据包括多个特征频率范围, 每个具体的频率范围与大脑不同功能状态相关^[14]。本文使用带通滤波器将 EEG 数据划分为 5 个频段: Delta、Theta、Alpha、Beta 和 Gamma, 频率范围依次为 0.5~4 Hz、4~8 Hz、8~12 Hz、12~30 Hz 和 30~50 Hz。其中, EEG 数据的微分熵特征注意力预测中具有良好的性能, 计算公式如式 (1) 所示^[23]。

$$h(x) = - \int_{-\infty}^{\infty} f(x) \log(f(x)) dx \quad (1)$$

式中: x 为一个服从高斯分布的随机变量; $f(x)$ 为 x 的概率密度函数。

如果时间序列随机变量 x 服从高斯分布 $N(\mu, \sigma^2)$, 则微分熵特征可定义为:

$$h(x) = - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right) dx \\ = \frac{1}{2} \log(2\pi e \sigma^2) \quad (2)$$

在实验过程在, 每个被试者的总采样时长为 7 080 s, EEG 数据采样频率为 200 Hz, 因此原始脑电数据的格式为 $1\,416\,000 \times 17$ (采样点数 $\times$ 电极通道)。由于每隔 8 s 计算一次标签值, 被试者的 EEG 数据格式可划分为 $885 \times 1\,600 \times 17$ (样本数 $\times$ 单个采样点数 $\times$ 电极通道), 在带通滤波器划分频段后数据格式变为 $885 \times 5 \times 1\,600 \times 17$ (样本数 $\times$ 频段数 $\times$ 单个采样点数 $\times$ 电极通道)。由于实验每隔 100 个采样点计算一次微分熵特征, 因此被试者的 EEG 数据格式最终为 $885 \times 5 \times 16 \times 17$ (样本数 $\times$ 频段数 $\times$ 微分熵特征 $\times$ 电极通道)。

1.3 眼电数据预处理

在 SEED-VIG 数据集中, 被试者的 EOG 数据采用 3 种不同方法生成了 36 维眼电特征, 涵盖眨眼、扫视和注视 3 个眼球运动类别, 字段 "features_table_ica"、"features_table_minu" "features_table_icav_minh" 分别对应独立成分分析 (Independent Component Analysis, ICA)、减法方法进行垂直眼电 (Vertical Electrooculography, VEO)、水平眼电 (Horizontal Electrooculography, HEO) 分离方法。每个被试者的 EOG 数据在经过特征提取后的数据格式为 $885 \times 3 \times 36$ (样本数 $\times$ 方法数 $\times$ EOG 特征数)。

1.4 多尺度感知单元

本文受文献[14]的启发, 在多尺度感知单元设计了深层、浅层结构, 如图 2 所示。

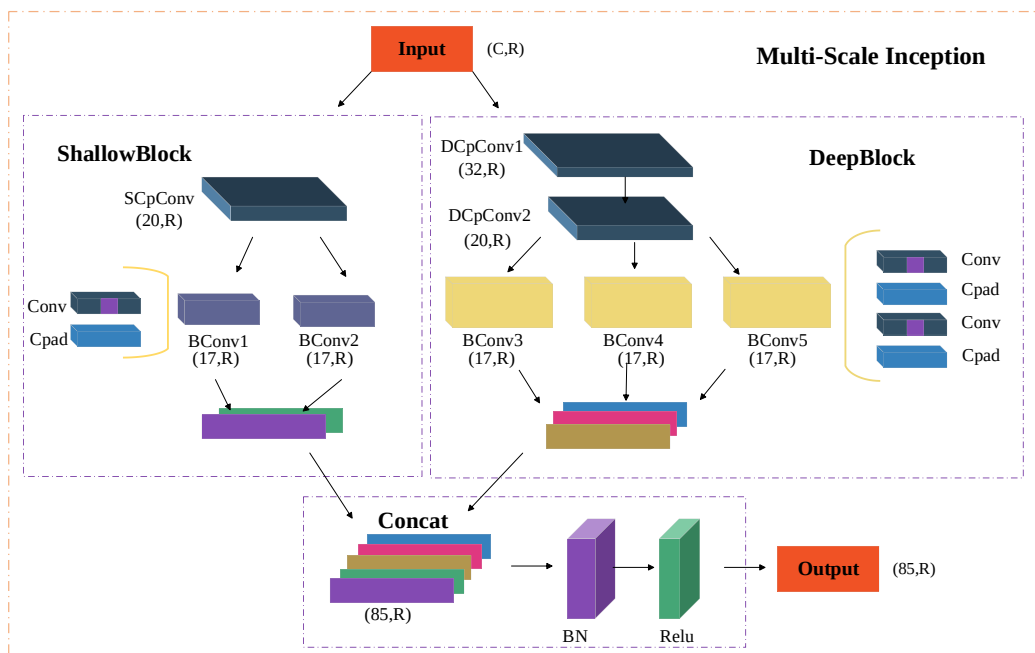


Fig. 2 Multi-scale perception unit

图 2 多尺度感知单元

图2中的深层结构(DeepBlock)由3个不同尺度的卷积函数组成,浅层结构(ShallowBlock)包含两个不同尺度的卷积函数,旨在局部和全局层面提取特征,丰富模型提取特征的多样性。EEG信号记录了大脑在不同时间、空间上的活动,多尺度感知单元能从多个尺度捕捉脑电信号中的时频和空间特征,以全面反映大脑活动的多样性。EOG信号记录了眼睛运动情况,多尺度特征提取能全面捕捉眼动过程中眨眼、扫视及注视的关键信息。值得注意的是,虽然EEG与EOG信号均使用多尺度感知单元进行特征提取,但两者在参数设置上存在一些差异。

多尺度感知单元包括ShallowBlock、DeepBlock。ShallowBlock由压缩层(SCpConv)和并行卷积层(BConv1和BConv2)组成,专注于提取局部特征。其中,压缩层通过减少通道数和模型参数来预防模型发生过拟合;并行卷积层包含两个并行卷积块(BConv1和BConv2),通过不同尺寸的卷积核扩展特征提取的视野。

DeepBlock由两个压缩层和并行卷积层(BConv3、BConv4和BConv5)组成,以提取全局特征。其中,两个压缩层用来增加网络深度,进一步强化全局特征提取效果;并行卷积块包含3个不同尺寸卷积核的卷积函数以扩展特征提取视野,并且这3个并行卷积块由两个相同卷积函数和两个相同常量填充函数组成,以进一步加深网络结构。

实验中,本文样本数据的输入格式为 $X=(x_1, x_2, \dots, x_N) \in R^{N \times C \times R}$, N 表示批量大小, C 表示数据的通道数, R 表示特征维度。首先,样本特征通过ShallowBlock的压缩卷积层得到得到 V_x^1 ,数据格式为 $N \times 20 \times R$ 。压缩卷积层的计算公式为:

$$V_x^1 = \text{Conv}^1(X) \quad (3)$$

式中: Conv^1 表示卷积核为1的卷积函数。

其次,特征 V_x^1 进入到ShallowBlock的并行卷积层得到两个特征 $[T_x^1, T_x^2]$,数据格式均为 $N \times 17 \times R$ 。对于EEG数据而言,该层的卷积函数卷积核大小分别为5和10。对于EOG数据而言,该层的卷积函数的卷积核大小分别为4和8。ShallowBlock的并行卷积层的运算定义如式(4)~式(5)所示。

$$T_x^1 = \text{lpad}_b^a(\text{Conv}(V_x^1)) \quad (4)$$

$$T_x^2 = \text{lpad}_b^a(\text{Conv}(V_x^1)) \quad (5)$$

式中: lpad_b^a 表示常量填充函数; Conv 表示卷积函数。

再次,样本特征经过DeepBlock的压缩卷积层处理后得到特征 V_x^2 ,数据格式为 $N \times 20 \times R$ 。

$$V_x^2 = \text{Conv}^1(\text{Conv}^1(X)) \quad (6)$$

接下来, V_x^2 进入并行卷积层得到3个特征 $[T_x^3, T_x^4, T_x^5]$,数据格式均为 $N \times 17 \times R$ 。对于EEG数据而言,该层卷积函数的卷积核大小分别为10、20和40。对于EOG数据而言,该层卷积函数的卷积核大小分别为5、10和20。DeepBlock的并行卷积层的计算公式为:

$$T_x^3 = \text{lpad}_b^a(\text{Conv}(\text{lpad}_b^a(\text{Conv}(V_x^2)))) \quad (7)$$

$$T_x^4 = \text{lpad}_b^a(\text{Conv}(\text{lpad}_b^a(\text{Conv}(V_x^2)))) \quad (8)$$

$$T_x^5 = \text{lpad}_b^a(\text{Conv}(\text{lpad}_b^a(\text{Conv}(V_x^2)))) \quad (9)$$

最后,在多尺度感知单元合并ShallowBlock和DeepBlock提取的特征,通过归一化函数和激活函数分别进行处理再传递到下一层,输出数据格式为 $N \times 85 \times R$ 。

$$Y_x = \text{Relu}(\text{BN}(\text{M}(T_x^1, T_x^2, T_x^3, T_x^4, T_x^5))) \quad (10)$$

式中: M 表示合并操作; BN 表示Batch Normalization 归一化函数; Relu 表示Relu激活函数。

1.5 多尺度通道注意力模块

本文通过多尺度协作模块将多尺度特征按5个尺度调整为新的数据格式: $X_{EEG}=(x_1, x_2, \dots, x_N) \in R^{N \times 5 \times 17 \times P}$, $X_{EOG}=(x_1, x_2, \dots, x_N) \in R^{N \times 5 \times 17 \times Q}$ 。其中, N 表示样本数,5表示尺度数,17表示通道数, P 、 Q 分别表示EEG与EOG的特征维度。

本文受文献[16]的启发,首先合并EEG与EOG数据得到多模态数据格式为 $X_{multi-modal} \in R^{N \times 5 \times 17 \times (P+Q)}$;其次按5个尺度划分得到特征组 $G_{multi-modal}=(g_1, g_2, \dots, g_5)$, $g_i \in R^{N \times 17 \times (P+Q)}$, $i \in \{1, 2, 3, 4, 5\}$;最后将特征组送入多尺度通道注意力模块进行通道注意力加权。具体计算公式为:

$$output_i = g_i \odot \sigma(r * \text{Conv}(\text{avg_pool}(g_i)) + b), i \in \{1, 2, 3, 4, 5\} \quad (11)$$

式中: $\text{avg_pool}()$ 为平均池化函数; σ 为Sigmoid激活函数; r 、 b 为模块可学习参数; $\odot$ 表示逐元素相乘。

综上,本文模型首先通过自适应平均池化平均每个通道的特征以获取全局通道信息。其次,利用一维卷积整合自适应平均池化后的通道信息以获取更丰富的通道关系。再次,使用Sigmoid激活函数激活整合后的通道信息计算通道注意力权重,以加权原始输入中的每个通道,使模型更聚焦于具有重要信息的通道。最后,将注意力权重应用于原始输入得到最终输出,在获得各尺度的加权特征后均值融合5个尺度得到特征 $Output \in R^{N \times 17 \times (P+Q)}$ 。

$$Output = \frac{1}{5} \sum_{i=1}^5 output_i, i \in \{1, 2, 3, 4, 5\} \quad (12)$$

1.6 降采样模块

降采样模块主要由自适应平均池化函数(AdaptiveAvgpool1d)、降维函数(Flatten)、激活函数(Relu)和Dropout函数组成,如图3所示。池化函数用于维度压缩;Dropout函数防止模型过拟合,目的在于压缩最终提取的特征尺寸,减少回归层计算的数据量。输入数据格式为 $X=(x_1, x_2, \dots, x_N) \in R^{N \times 17 \times (P+Q)}$, N 表示批量大小,数据通道数为17, $P+Q$ 表示多模态数据的特征维度。降采样模块计算公式如式(13)所示。

$$Z_x = D(\text{Relu}(\text{F}(\text{Avg}(X)))) \quad (13)$$

式中: F 表示降维函数 Flatten; Avg 表示自适应平均池化函数; D 表示 Dropout 函数。

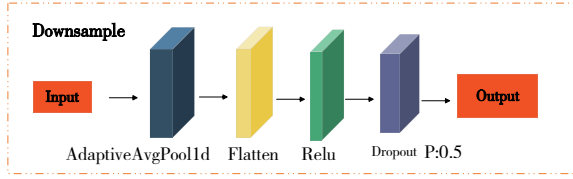


Fig. 3 Downsampling module

图3 降采样模块

1.7 回归层

数据经过降采样模块后,将得到的特征送入回归层进行注意力预测,并利用 MSE 损失函数优化模型。

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (14)$$

式中: N 为样本数量; y_i 为实际值; $\hat{y}_i$ 为模型预测值。

2 实验结果与分析

2.1 数据描述

本文采用上海交通大学仿脑计算与机器智能研究中心(BCMI)提供的 SEED-VIG 数据集,包含来自 21 名被试者的模拟驾驶系统采集的 EEG 和 EOG 数据。EEG 数据共包括 17 个电极通道(FT7, FT8, T7, TP7, TP8, T8, CP1, CP2, P1, Pz, P2, PO3, POz, PO4, O1, Oz, O2),总采样时长为 7 080 s,采样频率为 200 Hz。EOG 数据集涵盖 3 种不同方法提取的 36 维眼电特征,眼电数据中 3 个标签值“features_table_ica”“features_table_minus”“features_table_icav_minh”分别对应前额眼电 3 种不同的分离方法。

每位被试者的实验时长为 7 080 s,每 8 s 计算一次 PERCLOS (Percentage of Eye Closure in the Longest Blink) 值,每位被试者的 perclos_labels 文件包含 885 个样本标签。其中,PERCLOS 用于度量眼睛在特定时间段内持续闭合的指标,其值表示眼睛处于闭合状态的百分比,在驾驶注意力研究中常用于评估驾驶员的注意力水平,取值范围为 (0, 1),具体计算公式如式 (15) 所示。

$$\text{PERCLOS} = \frac{\text{blink} + \text{CLOS}}{\text{blink} + \text{CLOS} + \text{fixation} + \text{saccade}} \quad (15)$$

式中: blink 表示眨眼时间; CLOS 表示闭眼时间; fixation 表示注视时间; saccade 表示扫视时间。

2.2 实验参数

本文实验操作系统为 Windows 10, CPU 为 AMD Ryzen 7 3700X 8 核, GPU 为 NVIDIA GeForce RTX 2060, 采用深度学习框架 Pytorch 搭建模型, Batch Size 设置为 32, 最大训练次数设为 100, 模型训练采用 EarlyStopping 策略, patience 为 20, 即在连续训练 20 轮后若损失值未显著减小则终止训练, 初始学习率为 0.01。

本文为了动态调整学习率引入 PyTorch 框架中的 lr_scheduler 模块, 根据训练轮数以指数衰减方式自动调整学习率来加快模型收敛速度并提升泛化能力。随着训练次数增加, 模型学习率以初始学习率乘以 0.99 的训练轮数次进行迭代更新, 模型损失函数为 MSE, 优化器为 SGD。SGD 为 PyTorch 框架中的一个优化器, 用于实现随机梯度下降, 通过计算训练数据中每个样本的梯度平均值来更新模型参数, 所涉及的激活函数包括 ReLU、Sigmoid。

2.3 实验评价指标

本文引入均方根误差 (Root Mean Square Error, RMSE)、Pearson 相关系数 (PCC) 来评估模型性能。其中, RMSE 值越小表示模型预测越准确, 与实际观测值之间的差距越小; PCC 用于测量两个变量间线性关系强度与方向的统计量, 取值范围为 (-1, 1), 具体计算公式如式 (16)~式 (17) 所示。

$$\text{RMSE}(Y, \hat{Y}) = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (16)$$

$$\text{PCC}(Y, \hat{Y}) = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}} \quad (17)$$

式中: $Y, \hat{Y}$ 分别表示样本实际和预测的 PERCLOS 值输出向量; $y_i, \hat{y}_i$ 分别代表单个样本的实际和预测的 PERCLOS 值; $\bar{y}, \bar{\hat{y}}$ 则分别表示所有样本的实际和预测的 PERCLOS 均值。

同时, 本文设计了两个实验设来评估模型性能, 即被试者内部实验和跨被试实验。其中, 在被试者内部实验中, 将每个被试者数据分为 5 等份, 采用 5 折交叉验证, 使用其中 4 份进行训练、1 份进行测试; 跨被试实验采用留一被试法, 即每次将一个被试者的数据用于测试, 其他被试者的数据用于训练。

2.4 比较实验

为了评估所提模型性能, 在被试者内部试验组中选择 SVR^[19]、CCRF^[19]、CCNF^[19]、DAE^[24]、GELM^[4]、LSTM^[10]、DNNSN^[20]、MPCL^[9] 等模型进行比较, 在跨被试试验组中选择 DANN^[1]、ADDA^[1]、CNN-LSTM^[21]、CapsNet^[21] 等模型进行比较, 实验结果如表 1 所示。

由表 1 可知, 在被试者内部试验组中所提模型表现最佳, RMSE、PCC 分别为 0.064、0.959。相较于其他模型而言, MPCL 的 PCC 指标最高为 0.893, GELM 的 RMSE 指标最高为 0.071, 但仍与本文模型指标存在一定差距。在跨被试试验组方面, 本文所提模型表现出色, RMSE、PCC 值分别为 0.112、0.892。实验表明, 所提模型通过多尺度感知单元和多尺度通道注意力模块全面融合了多模态数据, 加强了与注意力预测相关的特征, 在驾驶注意力预测中效果显著。

Table 1 Experimental comparison results
表 1 实验比较结果

实验类别	模型	RMSE	PCC
个体内实验	SVR	0.100	0.830
	CCRF	0.100	0.840
	CCNF	0.095	0.845
	DAE	0.094	0.852
	GELM	0.071	0.808
	LSTM	0.080	0.830
	DNNSN	0.080	0.860
	MPCL	0.092	0.893
	本文模型	0.064	0.959
跨被试实验	DANN	0.143	0.840
	ADDA	0.141	0.844
	CNN-LSTM	0.130	0.860
	CapsNet	0.132	0.859
	本文模型	0.112	0.892

2.5 消融实验

2.5.1 不同多尺度感知单元数量

为了验证多尺度感知单元数量对模型性能的影响,本文分别设置了多尺度感知单元数量数量为 1、2、3、4、5、6 共 6 个实验组合进行消融实验,其余参数部分保持一致,实验结果如表 2 所示。由此可知,随着多尺度感知单元数量逐渐增加,本文模型在被试者内部试验组和跨被试实验组上的性能逐步提升,在多尺度感知单元数量达到 5 时模型取得最佳结果,但多尺度感知单元数量达到 6 时模型性能略微下降。实验表明,通过堆叠多个多尺度感知单元来增加特征提取模块的深度,有助于加强所提特征的显著性,使模型学到更抽象的表示,以有效捕捉每个模态的显著特征,但模型深度过深可能导致过拟合现象,进而导致模型性能下降。

Table 2 Performance comparison of models with different numbers of multi-scale perception units

表 2 不同多尺度感知单元数量的模型性能比较

实验类别	多尺度感知单元数量	RMSE	PCC
个体内实验	1	0.085	0.925
	2	0.075	0.944
	3	0.070	0.955
	4	0.069	0.956
	5	0.064	0.959
	6	0.066	0.957
跨被试实验	1	0.121	0.845
	2	0.117	0.881
	3	0.115	0.888
	4	0.114	0.890
	5	0.112	0.892
	6	0.113	0.889

2.5.2 不同注意力模块

为了验证多尺度通道注意力模块对模型性能的影响,设置了无注意力、SE 和 ECA 共 3 个实验组,除了注意力机制外各结构和参数完全一致。其中,SE、ECA 是通道注意

力机制的两种实现形式,相较于 SE 注意力机制,ECA 注意力机制采用一维卷积函数捕捉不同通道间的关系。

表 3 为个体内实验和跨被试实验结果。由此可知,未使用注意力机制的情况下模型实验表现最差,引入 SE、ECA 通道注意力模块后性能提升明显,尤其是引入 ECA 通道注意力模块后性能提升最优,证明了多尺度注意力模块通过关注不同尺度的通道,以自适应方式为不同滤波器学到的特征分配不同权重,增强了与注意力预测相关的特征,提升了模型性能。特别的,在个体内实验组中 SE 模块的 RMSE、PCC 分别为 0.075 和 0.938;在跨被试实验组中对应数值为 0.123 和 0.887,均低于 ECA 模块。ECA 模块采用可学习的卷积核计算通道注意力权重,相较于 SE 模块能更灵活地捕捉多模态融合数据中特定通道间的关系。此外,ECA 模块设计考虑了参数效率,采用了轻量级操作使模型更易于优化,降低了模型过拟合的风险。

Table 3 Comparison of model performance with different attention modules

表 3 不同注意力模块的模型性能比较

实验类别	注意力模块	RMSE	PCC
个体内实验	无注意力模块	0.092	0.922
	SE	0.075	0.938
	ECA	0.064	0.959
跨被试实验	无注意力模块	0.133	0.870
	SE	0.123	0.887
	ECA	0.112	0.892

2.5.3 不同模态性能比较

为了验证多模态融合策略在注意力预测中的有效性,本文对单一模态的 EEG、EOG 及多模态融合数据进行比较,实验结果如表 4 所示。

Table 4 Comparison of different modal performance

表 4 不同模态性能比较

实验类别	模态	RMSE	PCC
个体内实验	EEG	0.082	0.932
	EOG	0.068	0.947
	多模态	0.064	0.959
跨被试实验	EEG	0.136	0.856
	EOG	0.124	0.872
	多模态	0.112	0.892

由此可见,在个体内和跨被试实验中多模态实验组的表现相较于单一模态更优越,EEG、EOG 模态中使用注意力预测的信息密切相关且具有互补性,相较于与仅使用 EEG 或 EOG 的数据而言更具优势。特别的,在个体内实验组中单一模态 EOG 的 RMSE、PCC 分别为 0.068 和 0.947,在跨被试实验组中 RMSE、PCC 分别为 0.124 和 0.872,均高于单一模态的 EEG,原因为 EOG 信号主要来源于眼球运动,相对清晰,而 EEG 信号涉及更多生理过程,包括大脑电活动,可能受到更多因素影响,从而降低了其信噪比。

3 结语

本文提出一个基于多模态的多尺度通道注意力回归模型预测驾驶注意力。首先,通过多尺度感知单元分别从 EEG、EOG 数据中提取多尺度特征,并在多个尺度中有序合并两者特征以增加融合特征的多样性。其次,引入多尺度通道注意力机制以自适应方式关注不同尺度通道,为多尺度特征赋予差异化权重,增强与注意力预测相关的特征,以提升模型对驾驶注意力相关特征的表达能力。

实验表明,本文模型能有效融合两种模态数据来提取更多关键特征,并在驾驶注意力预测领域取得了显著成效。然而,目前研究尚存在一些局限性:①在跨被试的场景中模型性能仍需改进;②仅在离线环境进行注意力预测,实现驾驶注意力的实时预测仍然面临多方挑战,这也是未来研究的关键方向。

参考文献:

- [1] LI H, ZHENG W L, LU B L. Multimodal vigilance estimation with adversarial domain adaptation networks[C]// 2018 International Joint Conference on Neural Networks, 2018: 1–6.
- [2] WU W, SUN W, WU Q M J, et al. Multimodal vigilance estimation using deep learning[J]. IEEE Transactions on Cybernetics, 2020, 52(5): 3097–3110.
- [3] LI P, ZHANG Y, LIU S, et al. An EEG-based brain cognitive dynamic recognition network for representations of brain fatigue[J]. Applied Soft Computing, 2023, 146: 110613.
- [4] HUO X Q, ZHENG W L, LU B L. Driving fatigue detection with fusion of EEG and forehead EOG[C]// 2016 International Joint Conference on Neural Networks, 2016: 897–904.
- [5] GE W J, CHEN L. Fatigue driving detection method based on random forest and multi-source information fusion[J]. Software Guide, 2021, 20(10): 73–77.
葛文杰, 陈龙. 基于随机森林与多源信息融合的疲劳驾驶检测方法[J]. 软件导刊, 2021, 20(10): 73–77.
- [6] LI B, CHEN L. Driver fatigue state classification method based on regularized extreme learning machine[J]. Software Guide, 2020, 19(10): 121–124.
李冰, 陈龙. 基于正则极限学习机的驾驶员疲劳状态分类方法[J]. 软件导刊, 2020, 19(10): 121–124.
- [7] CHEN Y, LI J T, XU J L, et al. Design of driver fatigue warning system based on eye movement features[J]. Software Guide, 2020, 19(5): 116–119.
陈瑜, 李锦涛, 徐军莉, 等. 基于眼动特征的驾驶员疲劳预警系统设计[J]. 软件导刊, 2020, 19(5): 116–119.
- [8] PAN J, CAI X, MO D, et al. Residual attention capsule network for multimodal EEG-and EOG-based driver vigilance estimation[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1–12.
- [9] ZHANG M, LUO Z, XIE L, et al. Multimodal vigilance estimation with modality-pairwise contrastive loss[J]. IEEE Transactions on Biomedical Engineering, 2023, 71(4): 1139–1150.
- [10] ZHANG N, ZHENG W L, LIU W, et al. Continuous vigilance estimation using LSTM neural networks[C]// 23rd International Conference on Neural Information Processing, 2016: 530–537.
- [11] MA B Q, LI H, LUO Y, et al. Depersonalized cross-subject vigilance estimation with adversarial domain generalization[C]// 2019 International Joint Conference on Neural Networks, 2019: 1–8.
- [12] GAO Z, WANG X, YANG Y, et al. EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(9): 2755–2763.
- [13] DU Y, LIU J. IENet: a robust convolutional neural network for EEG based brain-computer interfaces[J]. Journal of Neural Engineering, 2022, 19(3): 36031.
- [14] GAO D, LI P, WANG M, et al. CSF-GTNet: a novel multi-dimensional feature fusion network based on Convnext-GeLU-BiLSTM for EEG-signals-enabled fatigue driving detection[J]. IEEE Journal of Biomedical and Health Informatics, 2023, 28(5): 2558–2568.
- [15] ZHENG W L, GAO K, LI G, et al. Vigilance estimation using a wearable EOG device in real driving environment[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(1): 170–184.
- [16] SONG K, ZHOU L, WANG H. Deep coupling recurrent auto-encoder with multi-modal EEG and EOG for vigilance estimation[J]. Entropy, 2021, 23(10): 1316.
- [17] WU E Q, XIONG P, TANG Z R, et al. Detecting dynamic behavior of brain fatigue through 3-D-CNN-LSTM[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2021, 52(1): 90–100.
- [18] NEMCOVA A, SVOZILOVA V, BUCSUHÁZY K, et al. Multimodal features for detection of driver stress and fatigue[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3214–3233.
- [19] ZHENG W L, LU B L. A multimodal approach to estimating vigilance using EEG and forehead EOG[J]. Journal of Neural Engineering, 2017, 14(2): 26017.
- [20] WU W, WU Q M J, SUN W, et al. A regression method with subnetwork neurons for vigilance estimation using EOG and EEG[J]. IEEE Transactions on Cognitive and Developmental Systems, 2018, 13(1): 209–222.
- [21] ZHANG G, ETEMAD A. Capsule attention for multimodal EEG-EOG representation learning with application to driver vigilance estimation[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 1138–1149.
- [22] QIU D F. A method for fatigue detection based on multimodal fusion neural network[J]. Modern Computer, 2020(35): 32–36, 42.
邱达锋. 一种基于多模态融合神经网络对疲劳度检测的方法[J]. 现代计算机, 2020(35): 32–36, 42.
- [23] CHENG X, WEI W, DU C, et al. VigilanceNet: decouple intra-and inter-modality learning for multimodal vigilance estimation in RSVP-Based BCI[C]// 30th ACM International Conference on Multimedia, 2022: 209–217.
- [24] DU L H, LIU W, ZHENG W L, et al. Detecting driving fatigue with multimodal deep learning[C]// 8th International IEEE/EMBS Conference on Neural Engineering, 2017: 74–77.

(责任编辑:刘嘉文)