

DOI: 10.19650/j.cnki.cjsi.J2210846

基于可解释深度学习的单通道脑电 跨被试疲劳驾驶检测*

冯笑^{1,2}, 代少升¹, 黄炼¹

(1. 重庆邮电大学通信与信息工程学院 重庆 400065; 2. 河南省高速铁路运营维护工程研究中心 郑州 450002)

摘要:脑电信号被认为是检测驾驶员疲劳状态的最佳生理信号之一。然而,由于不同被试者和不同记录时段的脑电信号差异很大,设计一个无校准的脑电疲劳检测系统仍然具有挑战性。近年来,虽然开发了许多深度学习方法来解决这个问题并取得了重大进展,但是深度学习模型的黑盒效应使得模型决策不可信赖。为此,本文提出了一种可解释深度学习模型,用于从单通道脑电信号中检测跨被试疲劳状态。该模型具有紧凑的网络结构,首先设计浅层 CNN 提取 EEG 特征,然后引入自适应特征重新校准机制增强提取特征的质量,最后通过 LSTM 网络将时间特征序列与分类相关联。模型分类决策的可解释信息则是由 LSTM 输出隐藏状态的可视化技术实现的。在持续驾驶任务的公开脑电数据集上进行大量跨被试实验,该模型分类平均准确率最高达到 76.26%。相比于先进的紧凑型深度学习模型,该模型有效降低了参数量和计算量。可视化结果表明该模型已发现神经生理学上可靠的解释。

关键词:单通道脑电;疲劳检测;紧凑网络;可视化技术;自适应特征重新校准

中图分类号: TP391 R741.044 TH79 **文献标识码:** A **国家标准学科分类代码:** 510.40

Cross-subject driver fatigue detection from single-channel EEG with an interpretable deep learning model

Feng Xiao^{1,2}, Dai Shaosheng¹, Huang Lian¹

(1. College of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. Henan High-Speed Railway Operation and Maintenance Engineering Research Center, Zhengzhou 450002, China)

Abstract: Electroencephalography (EEG) has been considered as one of the best physiological signals to detect drivers' mental fatigue. However, EEG signals vary significantly among different subjects and recording sessions, and it is still challenging to design a calibration-free system for EEG fatigue detection. In recent years, many deep learning-based methods have been developed to address this issue and achieve significant progress. However, the "black-box" nature of deep learning models makes their decisions mistrust. Therefore, an interpretable deep learning model is proposed to recognize cross-subject fatigue states from single-channel EEG signals in this article. The model has a compact network structure. Firstly, a shallow CNN is designed to extract the EEG features. Then, the adaptive feature recalibration mechanism is introduced to enhance the features extraction ability. Finally, the time series of extracted features are linked to classification with LSTM. The interpretable information of the classified decision is achieved through a visualization technique that is taking advantage of hidden states output by the LSTM layer. Extensive cross-subject experiments are implemented on an open EEG dataset with a sustained-attention driving task, and the proposed model achieve the highest average accuracy of 76.26%. In addition, compared with the advanced compact deep learning models, the parameters and computation are effectively reduced. Visualization results indicate that the proposed model has discovered neuro-physiologically reliable interpretation.

Keywords: single-channel EEG; fatigue detection; compact network; visualization technologies; adaptive feature recalibration

0 引言

疲劳是一种以注意力下降、警觉性缺失和睡意为主要特征的生理、心理状态。疲劳驾驶主要表现为驾驶员精神疲劳,明显不同于因体力劳动引起的身体疲劳。据美国国家公路交通安全管理局透露,2009~2013年发生超过7.2万起撞车事故。2018年由疲劳驾驶导致的死亡事故为775起(占死亡总数的2.1%)^[1]。开发一种能连续观察驾驶员的警觉状态,并在驾驶员产生睡意前发出预警的消费级脑电疲劳监测系统,对驾驶安全和预防交通事故具有重要意义。

脑电信号(electroencephalography, EEG)作为人脑活动的直接体现,可以快速地反映出人的生理及心理变化过程^[2],被誉为疲劳检测的“金标准”。随着EEG采集设备的不断发展,因其时间分辨率高、便携而成本低等技术优势,越来越多的研究人员将脑电信号作为生物神经信号特征引入疲劳检测^[3-7]。然而,脑电信号在不同被试间和不同记录时段存在较大差异性。大多数驾驶环境中的脑电疲劳检测研究使用被试相关模型来测试他们的方法,每一被试都需要花大量的时间来进行必要的校准步骤^[5-7],这对于新用户的使用是很不友好的^[8]。此外,由于脑电信号对噪声/伪影敏感,在被试间/被试内存在非平稳性和低信噪比特性也加剧了任务识别的难度。

1 理论分析

利用脑电信号建立一个完全无校准的疲劳检测系统仍然是一项具有挑战而困难的任务。传统的疲劳识别方法依赖于从脑电信号中提取特征,这需要专家知识或先验知识以便对感兴趣的脑电特征建模,这很可能排除了对疲劳识别非常重要的相关信息。如Ogino等^[9]提出使用单通道脑电的功率谱密度(power spectral density, PSD)特征^[10]及逐步线性判别分析(stepwise linear discriminant analysis, SWLDA)的特征选择方法。Hu等^[11]则利用样本熵、模糊熵、近似熵和谱熵共4种熵特征进行驾驶疲劳识别。

深度学习允许端到端学习而不需要手工特征提取,能够直接从原始的高维数据中学习基本特征,越来越受到脑机接口(brain computer interface, BCI)研究人员的关注^[12-14]。受滤波器组共同空间模式(filter bank common space pattern, FBCSP)算法的启发,Lawhern等^[12]专门为脑电识别任务而设计了一种通用且紧凑的卷积神经网络(convolutional neural network, CNN),该模型EEGNet使用二维卷积层对原始信号进行滤波,然后

使用空间深度卷积层和时间可分离卷积层提取特征,在多个脑机接口数据集上取得了比传统方法更高的准确率。Schirrneister等^[13]提出了一种由时间卷积、空间卷积和池化层构成的浅层CNN模型ShallowConvnet,被证明是处理脑电运动想象分类的有效方法。对于单通道脑电信号分类,Ding等^[14]利用级联CNN和注意力机制构建了一个深度学习模型,用于从移动设备捕获的单通道脑电信号中检测司机困倦。为了提取单通道脑电数据之间的时间依赖性特征,Fraiwan等^[15]提出了一种长短期记忆网络(long short-term memory networks, LSTM),来自动区分局灶性和非局灶性癫痫脑电信号。Supratak等^[16]提出了一种名为DeepSleepNet的深度学习模型来进行睡眠阶段的分类。该模型使用CNN提取特征,并使用双向LSTM学习睡眠阶段的过渡规则。然而,现有脑电信号处理领域的研究工作大多将深度学习模型看作黑盒分类器^[17],这些深度学习方法具有不透明、缺乏解释性,使得我们更加需要一个可信赖的模型。由于深度学习模型可以通过提取有意义的可解释信息来提高信任度,而且通过将所学内容的可视化过程作为验证有效性的关键,因此,建立一种可解释深度学习模型对于无校准疲劳驾驶检测至关重要。为此,Cui等^[17]提出一种具有全局池化层的紧凑且可解释CNN,来识别单通道脑电跨被试疲劳驾驶。该模型使用类激活映射(class activation mapping, CAM)方法对学习到的脑电特征可视化,但是,采用全局池化层忽略了数据序列中的时间依赖信息。

综上所述,本文提出了一种新的可解释深度学习模型架构,用于从单通道脑电信号中识别跨被试疲劳驾驶任务。该模型采用紧凑的网络结构,首先设计浅层CNN提取相应频段的脑电特征。然后,引入自适应特征重新校准机制学习特征的相互依赖性,以增强提取特征的质量。最后,通过LSTM将时间上的特征序列与分类关联起来。为了解释模型分类决策的学习模式,设计基于LSTM的可视化技术来揭示输入信号中包含重要特征的局部区域。通过在公开脑电疲劳驾驶数据集上的大量实验证明了模型不仅轻量高效,而且很好地解释了模型分类决策的共享脑电模式。

2 方法

2.1 网络总体结构

可解释深度学习模型的网络结构如图1所示,采用紧凑型结构设计,主要包括浅层CNN、自适应特征重新校准模块和LSTM网络。浅层CNN与自适应特征重新校准模块共同作为模型的特征提取端,以提高网络提取特征的代表学习能力。LSTM网络则将时间上获得的特征输出并分类预测。

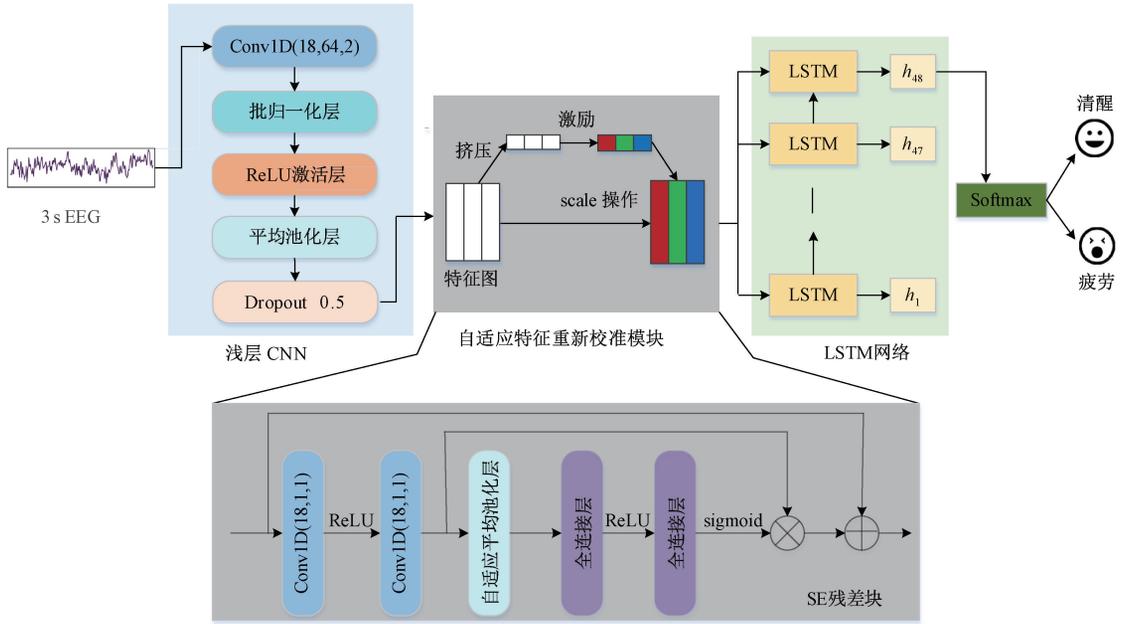


图1 整体网络结构

Fig. 1 The overall network architecture

2.2 浅层 CNN 特征提取

考虑 CNN 在处理 EEG 时间序列数据方面有广泛应用效果,本文设计浅层 CNN 网络从单通道脑电信号中提取脑电疲劳的浅层特征。如图 1 所示,浅层 CNN 网络包含有 1 维卷积层、批归一化层、激活层、平均池化层和 Dropout 丢弃层。

第 1 层为卷积层 Conv1D(18,64,2),由 N 个大小为 64、步幅为 2 的一维卷积滤波器组成($N=18$)。每个卷积核长度为 64,是信号采样率(128 Hz)的 1/2,因此可以捕获 2 Hz 以上脑电图的基本频段。接着是批处理归一化层和激活层。通过批归一化操作^[18]对每个特征维度小批量归一化以消除内部协变量移位。激活层则是使用 ReLU 函数对数据进行非线性变换。平均池化层 AvgPooling(8,4)使用内核大小为 8、步幅为 4 的平均池化操作来减少模型的参数量。然后采用 Dropout 层以进一步减小模型的过拟合。

2.3 自适应特征重新校准

为了增强浅层 CNN 提取特征的质量,采用自适应特征重新校准(adaptive feature recalibration, AFR)机制对特征之间相互依赖关系进行建模。这里通过挤压和激励(squeeze-and-excitation, SE)残差块自适应地选择最具鉴别的特征^[19],其中残差结构可以防止因网络加深而出现的梯度消失和梯度爆炸。

SE 残差网络包含两个卷积 Conv1D(18,1,1),内核大小和步幅都为 1,以及 ReLU 激活函数。假设浅层 CNN 生成的特征图为 $I \in \mathbf{R}^{N \times T}$,SE 网络的两个卷积操作为

$F = \text{Conv2}(\text{Conv1}(I))$,其中 $F = \{F_1, \dots, F_N\} \in \mathbf{R}^{N \times T}$, N 是特征总数, T 是 F_i 的长度($T=48$)。

然后采用自适应平均池化层对全局空间信息进行挤压,将 $F \in \mathbf{R}^{N \times T}$ 压缩为 $s = \{s_1, \dots, s_N\}$,其中 s_i 是 F_i 中 T 个序列特征点的平均值, $1 \leq i \leq N$ 。为了利用挤压操作所聚合的信息,引入激励操作来捕获特征之间的依赖关系。具体通过两个全连接层(fully connected layer, FC)来实现。第 1 个全连接层后使用 ReLU 激活函数进行降维,第 2 个全连接层后使用 sigmoid 激活函数进行升维,其结果如式(1)所示。

$$e = \sigma(\omega_2(\delta(\omega_1(s)))) \in \mathbf{R}^{N \times T} \quad (1)$$

其中, σ 和 δ 分别表示 sigmoid 和 ReLU 激活函数, ω_1 和 ω_2 表示这两个 FC 层。然后,特征图 F 与 e 通过 scale 操作得:

$$O = F \otimes e \in \mathbf{R}^{N \times T} \quad (2)$$

其中, \otimes 表示逐点乘法。将原始输入 I 与 SE 残差块增强的特征相结合,最终 AFR 模块输出为:

$$X = I + O \in \mathbf{R}^{N \times T} \quad (3)$$

2.4 LSTM 网络

作为一种特殊的循环神经网络(recurrent neural networks, RNNs),LSTM 网络可以捕获时间序列数据的长期依赖信息。这里引入 LSTM 网络来捕获脑电序列的长、短期依赖关系。LSTM 网络采用单向而不是双向,可以消除 EEG 后向处理所需的迭代时间,使得计算资源减少了大约 1/2。LSTM 层按照时间序列方式将一组输入元素 $X = (x_1, x_2, \dots, x_T)$ 映射到一组隐藏序列特征表示

$(h_1, h_2, \dots, h_T), h_i \in \mathbf{R}^D$, 其中 D 为每一序列特征的维度。LSTM 层可以表示为:

$$h_i \leftarrow LSTM(x_i, h_{i-1}) \quad (4)$$

通过输入 x_i 和前一个隐藏状态 h_{i-1} 之后, LSTM 层生成新的隐藏状态 $h_i, t \in (1, \dots, T)$ 。最后一个隐藏向量 h_T 被认为包含对分类有用的所有重要信息, 通常在其后引入全连接层来总结特征。为了使 LSTM 输出的隐藏状态具有可解释性, 将其隐藏向量维度 D 设定为 2, 最后将 h_T 连接 Softmax 激活函数进行分类。这种设计不仅大大减少了模型的参数量, 而且使 LSTM 隐藏状态展现出网络预测的变化过程。

2.5 可视化技术

从脑电信号中学习到的特征进行可视化分析, 能够揭示模型用于分类决策的神经生理可解释模式, 而不是受噪声或人工伪影的干扰。为此, 设计一种可视化技术来识别输入信号对于分类决策贡献最大的局部区域, 并且通过与已知晓的脑电神经生理模式相比较, 帮助理解模型学习到的脑电特征与不同精神状态之间的关系。

假设某一标签 c 样本 $\mathbf{S} = (s_1, s_2, \dots, s_n)$, 生成的热图为 $\mathbf{M}_c = (m_1, m_2, \dots, m_n)$, m_i 指示样本采样点 s_i 贡献分类标签 c 的重要性。与计算机视觉领域中利用 CAM 及其扩展技术从深度 CNN 模型中提取热图不同, 这里模型能够直接从 LSTM 隐藏状态序列表示 (h_1, h_2, \dots, h_T) 中生成可视化热图 \mathbf{M}_c 。

模型最后使用 Softmax 激活函数生成二分类预测的概率值。这里对每一隐藏状态 h_i 应用 Softmax 函数得:

$$h_i^* = \text{Softmax}(h_i) \quad (5)$$

假设 $h_{t,c}^*$ 为类 c 在时刻 t 的预测值, 序列 $\mathbf{H}_c^* = (h_{1,c}^*, h_{2,c}^*, \dots, h_{T,c}^*)$ 反映了网络如何获得最终概率值的演化过程。因此, 输入 x_i 对于网络最终分类的贡献可以由式(6)量化得:

$$\Delta h_{t,c}^* = h_{t,c}^* - h_{t-1,c}^* \quad (6)$$

其中, $t \in (1, 2, \dots, T)$, 且 $h_{0,c}^* = 0$ 。然后, 对 $\Delta \mathbf{H}_c^* = (\Delta h_{1,c}^*, \Delta h_{2,c}^*, \dots, \Delta h_{T,c}^*)$ 进行 z -score 标准化后可得 $\mathbf{Z}_c = (z_{1,c}, z_{2,c}, \dots, z_{T,c})$, 该过程类似于 CAM 方法。

根据模型的网络结构可知 AFR 模块的时间维度不变, 浅层 CNN 网络中只有平均池化层将数据的时间维度缩小到原来的 $1/8$ 。因此, 可以对每一个 $z_{t,c}$ 元素进行复制并填充 8 次, 热图 \mathbf{M}_c 即可恢复到平均池化前的数据长度, 其表示式为:

$$\mathbf{M}_c = (\underbrace{z_{1,c}, \dots, z_{1,c}}_8, \dots, \underbrace{z_{T,c}, \dots, z_{T,c}}_8) \quad (7)$$

这里 \mathbf{M}_c 被称为“相对热图”, 它反映了输入信号局部区域对最终分类决策的相对重要性。同样, 可以从序列 \mathbf{H}_c^* 得到另一热图 \mathbf{M}_c^* , 即为“累积热图”, 它反映了输入信号局部区域对类 c 预测似然值的累积贡献值, 其表

示式为:

$$\mathbf{M}_c^* = (\underbrace{h_{1,c}^*, \dots, h_{1,c}^*}_8, \dots, \underbrace{h_{T,c}^*, \dots, h_{T,c}^*}_8) \quad (8)$$

3 实验与结果分析

3.1 实验数据

为了研究在驾驶环境下的跨被试疲劳检测, 使用了来自持续驾驶任务的数据集^[20]。在实验过程中, 通过在沉浸式驾驶模拟器中进行 90 min 持续注意的夜间驾驶任务诱导疲劳和困倦。车道偏离事件被随机引入, 使汽车向左或向右偏离中央车道。参与者被要求立即对事件作出反应, 操纵方向盘将车开回中央车道。疲劳程度可以通过参与者对事件的反应时间来衡量。

实验邀请 27 名年龄在 22~28 岁之间的受试者参加实验, 使用 Scan SynAmps2 Express 系统记录脑电信号。EEG 使用 32 个 Ag/AgCl 电极的有线脑电帽采集, 其中 30 个脑电电极和 2 个参考电极(对侧乳突)。脑电电极按照 10~20 国际标准导联系统放置, 采样频率为 500 Hz。

脑电数据预处理采用 1~50 Hz 带通滤波器和伪影抑制处理^[20]。然后通过人工去除明显的眨眼干扰信号, 而且使用 EEGLAB 工具进一步去除眼部和肌肉伪影。处理后的数据被下采样至 128 Hz, 并提取每次试验在偏离事件发生前 3 s 的 EEG 样本。由于 Oz 通道被发现包含了区分疲劳和清醒脑电信号的最显著特征^[17], 研究中使用了来自 Oz 通道的脑电数据, 每个样本维度为 1 个(通道)×384 个(采样点)。

样本标签按照文献[8]中描述的方法来选择和标记 EEG 样本。计算每个样本的局部反应时间(reaction time, RT), 即被试对汽车漂移事件做出反应的时间, 以及全局反应时间, 即汽车漂移事件发生前 90 s 窗口内所有试验样本的局部反应时间(RTs)的平均值。每个 EEG session 内的基线“清醒 RT”被定义为其局部 RTs 的前 5% 时间。局部 RT 和全局 RT 均小于 1.5 倍清醒 RT 的样本被标记为清醒状态; 局部 RT 和全局 RT 均大于 2.5 倍清醒 RT 的样本被标记为疲劳状态。为了确保有足够的训练和测试样本, 每位参与者的每种状态至少选择 50 个样本以上。最后对每个参与者的清醒和疲劳标签样本进行平衡, 共提取了 11 个不同参与者的 1 011 个清醒标签和 1 011 个疲劳标签样本^[17]。

3.2 实验设置

本实验选用 Pytorch 框架, 使用 Python3.6 编程, 硬件设备使用的 CPU 为 Intel(R) Core(TM) i5-12600KF, 内存大小 16 GB, GPU 型号为 NVIDIA GeForce GTX 1080 Ti, 显存大小 11 GB。其中 EEGNet 模型代码开源^[11], 为

TensorFlow 环境运行,其他深度模型代码均在 Pytorch 环境下运行。模型训练过程中使用 Adam 优化器优化,对模型中每个参数使用相同的学习率,其中学习率固定为 0.001,迭代次数设为 17,批大小设为 20。

3.3 实验结果分析

实验在公开的脑电疲劳驾驶数据集上,通过留一法 (leave-one-subject-out) 交叉验证来比较跨被试分类结果,即其中一个被试者的 EEG 数据用于测试,而其他被试者数据用于训练分类器。这个过程反复进行,直到每名被试都被测试一次。

1) 不同方法的跨被试实验结果

该模型与传统方法及当前先进的 EEG 深度学习方法进行跨被试分类准确率的比较。对传统方法进行了测试,基线分类器为 SVM 和 LDA,特征提取方法包括 RelativePower 和 FourEntropies。其中 RelativePower 是用 Welch 方法从 EEG 中 Oz 信道计算 Delta(1~4 Hz)、Theta(4~8 Hz)、Alpha(8~12 Hz) 和 Beta(12~30 Hz) 的相对功率^[17],FourEntropies 是利用样本熵、模糊熵、近似熵和谱熵共 4 种熵特征进行驾驶疲劳识别^[11]。与当前先进

的深度学习方法比较,如 EEGNet^[12]、ShallowConvNet^[13]、DeepConvNet^[13] 以及 Compact-CNN^[17]。其中 EEGNet、ShallowConvNet 和 DeepConvNet 这 3 种模型都是将用于多通道脑电的空间卷积核改为 1×1 卷积核,以此来适应单通道脑电数据的特征提取。由于 EEGNet-8,2 网络比 EEGNet-4,2 网络具有更高的准确率,这里的 EEGNet 模型采用 EEGNet-8,2 网络。如表 1 所示跨被试测试比较结果,除了 DeepConvNet 模型外所有深度学习方法普遍要高于传统方法的测试准确率,说明深度学习方法相比于传统特征提取方法能够直接从大量高维数据中学习 EEG 疲劳特征,而 DeepConvNet 准确率低的原因很可能是随着网络加深而出现了过拟合造成的。相比于这些传统方法和先进的深度学习方法,我们的模型取得了跨被试平均准确率最高达到 76.26%,而且在每一被试单独测试中多次取得了最优或次优的结果(最优结果已黑色加粗),证明了该模型泛化性更好。该模型的准确率在每一被试单独测试中要明显高于当前先进的可解释卷积神经网络模型 CompactCNN,在可解释深度学习方法比较中性能也占据优势。

表 1 不同方法在脑电疲劳驾驶数据集上的跨被试准确率比较

Table 1 Comparison of different methods for cross-subject test accuracy on EEG driving fatigue dataset %

被试 编号	传统方法				深度学习方法				
	SVM	RelativePower+ SVM	RelativePower+ LDA	FourEntropies+ SVM	EEGNet	ShallowConvNet	DeepConvNet	CompactCNN	Ours
1	53.19	77.13	75.00	79.79	82.45	76.06	72.87	77.45	81.38
2	50.76	46.21	43.94	55.30	51.52	46.97	42.42	52.80	51.52
3	53.33	49.33	49.33	63.33	52.67	55.33	50.67	63.47	66.00
4	52.70	61.49	50.68	45.95	75.68	75.00	62.16	76.22	74.32
5	50.00	68.75	64.29	49.11	70.09	64.73	53.13	76.52	85.71
6	53.61	85.54	87.95	69.88	74.70	81.93	83.13	77.11	83.73
7	47.06	63.73	65.69	52.94	59.80	68.63	70.59	67.35	68.63
8	47.35	73.48	77.65	66.67	60.23	55.30	67.05	71.93	78.03
9	56.69	81.21	76.11	71.34	87.58	87.26	82.80	88.25	90.45
10	49.07	86.11	87.96	74.07	84.26	86.11	84.26	81.67	86.11
11	53.54	69.03	65.04	68.58	65.49	66.37	66.81	72.65	73.01
均值	51.57	69.27	67.60	63.36	69.50	69.43	66.90	73.22	76.26

2) 紧凑型深度学习模型参数量和计算量

如图 2 所示为该模型与其他紧凑型深度模型在脑电疲劳驾驶数据集上的参数量和和计算量对比,可以看出该模型的参数量达到次优值约为 2.2 K。虽然高于 EEGNet 的 1.8 K,但是所取得的平均准确率提高了 6.76%。与其他 3 种紧凑深度学习模型相比计算量明显下降,约为

5.6 MFLOPs。由此可知该模型具有轻量高效的优势,能够有效减少参数量和计算量,从而降低了算法复杂度,也避免造成过拟合现象。

3) 消融实验

为了验证该模型各部分的作用,在该脑电疲劳驾驶数据集上进行消融实验测试。该模型主要包括浅层

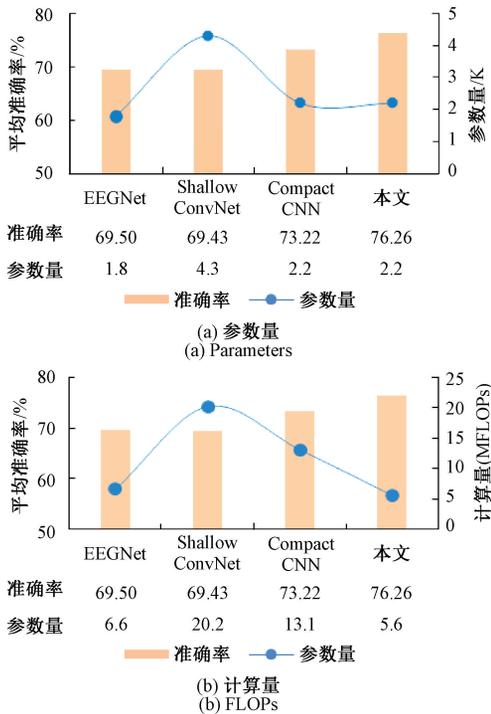


图 2 不同紧凑型深度学习模型的跨被试性能评估

Fig. 2 The evaluation of cross-subjects performance on different compact deep learning models

CNN、AFR 和 LSTM 网络 3 个模块。在消融实验对比中,CNN 表示以浅层 CNN 作为基准结构,然后添加全连接层分类;CNN+LSTM 表示在浅层 CNN 的基础上,添加 LSTM 将时间序列特征与分类相关联;CNN+AFR+LSTM 表示在 CNN+LSTM 的基础上,添加 AFR 机制。如图 3 所示,通过实验比较跨被试分类结果的均值,包括准确率、精确率、召回率以及 F1 分数,可以看出该模型各部分的有效性,特别是 AFR 机制明显提高了模型的性能。

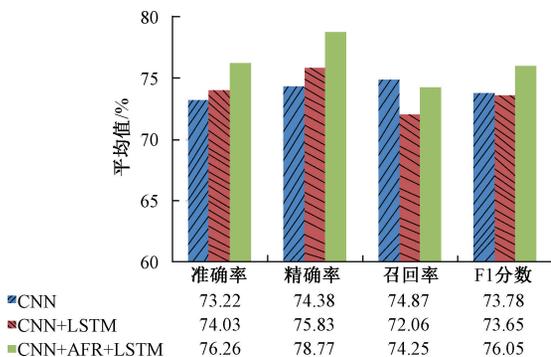


图 3 模型消融实验结果

Fig. 3 The ablation experiment results of the model

4) 实测实验

为了进一步评估该模型的疲劳检测性能,在真实驾驶场景下选用经典的脑电识别任务模型 EEGNet^[12] 作为基线进行实测实验对比。实验由一名被试者进行实车驾驶,持续时间约 2.5 h。如图 4(a) 所示为百度地图中驾驶路线。实验通过 KSS 和 PERCLOSE 算法分别对疲劳状态进行主观和客观评估。KSS 按从 1~10 依次表示非常清醒到非常困倦共 10 个等级,然后除以 10 得到疲劳评估的 KSS 值^[21]。KSS 值是通过对被试者 KSS 问卷而得到的。PERCLOS 算法是机动车辆的一种有效、便携且非接触的疲劳评估方法,它是通过实时采集人脸的眼睑图像来分析并计算闭眼时间与总监测时间的比率^[22]。如图 4(b) 所示,通过摄像头实时采集人脸图像帧,然后使用 Python 库中 Dlib 库工具获取眼睑部位的特征点。这里 PERCLOS 值是通过计算总监测时间内眨眼频率和闭眼时间所占比率,来生成相应的疲劳预测值。最后,总的疲劳评估值=0.5×KSS 值+0.5×PERCLOS 值。实验选择困倦和清醒状态的临界值 0.5 作为阈值,即当疲劳评估值≥0.5 时,驾驶员被认为处于驾驶的疲劳状态。



图 4 实测驾驶场景图

Fig. 4 Experimental driving scene diagram

实测期间以每 5 min 进行一次主、客观评估来获取本次疲劳评估状态,可以得到 30 个 0~1 区间的疲劳测量值。如图 4(b) 所示,这里使用 Bitbrain 32 通道通用无线脑电设备来采集脑电图数据,并选择 Oz 通道进行单通道无校准脑电疲劳识别的实验验证。然后经过预处理后使用这两种训练模型来提取脑电特征并生成相应的疲劳预测值。

实测结果如图 5 所示,实线表示 30 个实际测量值,EEGNet 的预测值表现出更大的波动范围。相比于 EEGNet,该模型预测值不仅可以较好的匹配实际测量值,而且预测疲劳状态的错误个数更少,即测试准确率更高,达到 80%。因此,该模型能够更加准确且稳定地识别驾驶员的疲劳状态。

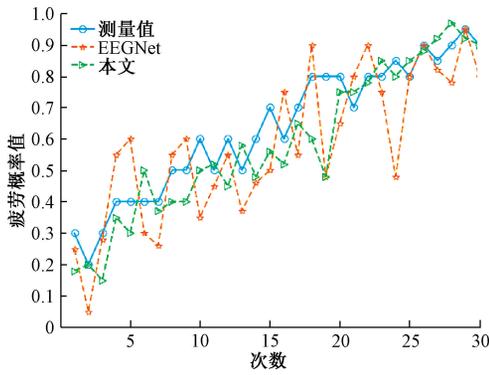


图5 实测性能比较

Fig. 5 Comparison of performances in road test

3.4 学习模式可视化

学习模式可视化分析就是探索模型学习到的不同被试间共享的脑电特征,并以此作为清醒和疲劳状态的分类指标。为了解释模型学习到的普遍神经生理模式,图6和7中展示了来自不同被试的一些代表性样本。根据2.5节式(7)和(8)分别可以计算累积热图和相对热图。如图6、7、8所示累积热图和相对热图分别位于每个图左侧顶部和中间,脑电Oz通道的相对功率图位于左侧底部,图右侧则显示了所有通道的EEG信号样本(含Oz)。

通过对高概率预测疲劳标签样本的可视化分析,可以观察到有关频率成分的两种普遍生理模式。第1类样本通常包含在Alpha波段产生高功率的纺锤状结构,如图6(a)所示;第2类样本在Delta和Theta频段占有较高比重,如图6(b)所示。从图6(a)所示样本的累积热图可以看出分类主要取决于信号的后半部分检测到的特征;从其相对热图及功率图可以看出信号在Alpha波段具有更高的振幅和相对功率,其特征表现为Alpha频带内一个狭窄的频率峰值。该多通道脑电图则显示来自其他通道的脑电信号也出现了这种纺锤状结构。实际上,捕获到脑电信号中的这种微结构被称为“ α 纺锤波”,已被证明为识别驾驶疲劳的强烈指标^[23]。如图6(b)所示,从相对热图中可以看到模型捕捉到了Theta-Delta波段爆发,并作为疲劳状态识别的证据。事实上,这些位于Theta-Delta频段的脉冲,已被文献[24]发现与嗜睡状态有很大关联。

通过对高概率预测清醒标签样本的可视化分析,可以发现这类样本通常含有较高Beta波比重,如图7(a)所示;或有较高Delta波比重,如图7(b)所示。从图7(a)所示样本的相对热图发现模型已经将脑电信号中包含较高比重的高频波作为清醒状态的判断依据。这些高频波大多是由头皮肌电(electromyography, EMG)活动引起的,通常在清醒的脑电信号中占主导地位^[24-25]。如图7(a)所示的累积热图,当信号逐渐输入到LSTM网络

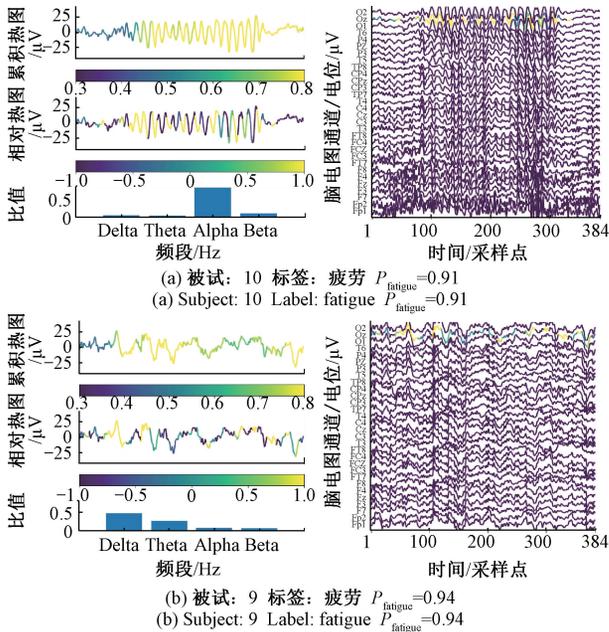


图6 模型对疲劳标签样本以高概率值正确分类的学习模式可视化

Fig. 6 Visualization of learned patterns on selected samples that are correctly classified by the network with high likelihoods of the fatigue label

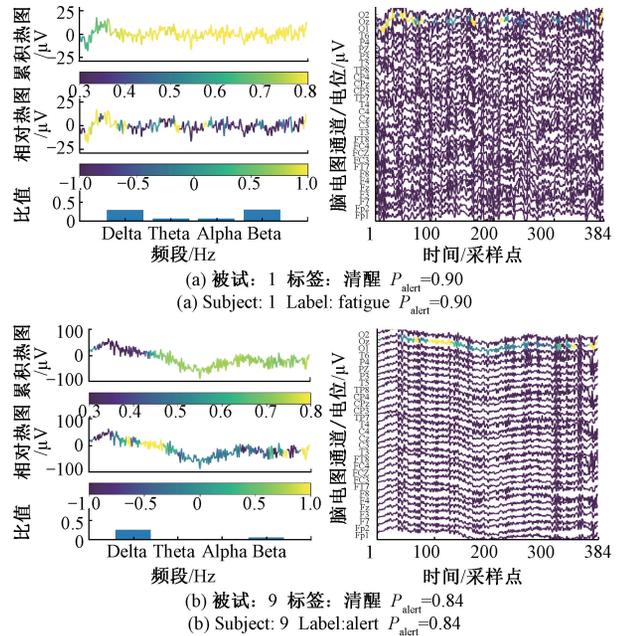


图7 模型对清醒标签样本以高概率值正确分类的学习模式可视化

Fig. 7 Visualization of learned patterns on selected samples that are correctly classified by the network with high likelihoods of the alert label

时,分类概率值是稳步增加的,这说明了相对热图中高亮显示的局部区域对分类决策的贡献值也是逐渐递增的。从图 7(b)所示的样本发现模型将较大幅度漂移的区域作为识别清醒状态的显著特征。由于 Delta 波被认为在深度睡眠阶段占主导地位,因此这些漂移特征和 Delta 波无关。从多通道脑电图观察到所有通道都出现大幅度信号漂移现象,这很可能是由传感器信号漂移或被试者运动引起的。

最后,利用可视化技术分析一些分类错误的样本特征。如图 8(a)所示的样本标签为疲劳状态,但被模型错误分类为清醒状态。该样本与图 7(a)所示的情况相似,都含有较高比重的 Beta 波,而该典型特征被模型作为识别清醒状态的证据。正如前面所述,由肌电活动引起的高 Beta 波比重通常存在于被试者清醒时的脑电信号中,也没有出现明显的疲劳特征,如 Alpha 纺锤波或 Theta-Delta 波段爆发。如图 8(b)所示样本标签为清醒状态,但被错误分类为疲劳状态。错误分类直接原因很可能是因为含有较高比重的 Alpha 波,被模型作为强烈的疲劳识别证据。根据上述分析,被试者 2 分类准确率低的原因可能为缺乏明显的疲劳相关特征,这与模型学习到的群组统计数据存在偏差。而被试者 7 测试准确率低的原因有可能是在采集样本期间,被试者正好从昏昏欲睡的状态及时警醒造成的。可视化的结果证明了模型对这两类样本的决策是合理的。

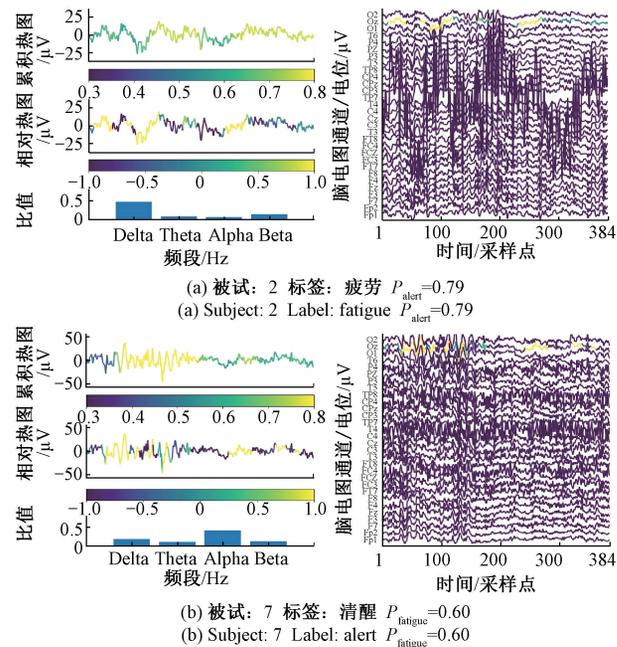


图 8 模型对样本错误分类的学习模式可视化

Fig. 8 Visualization of learned patterns on selected wrongly classified samples

4 结 论

本文提出了一种可解释性端到端的深度学习模型,用于从单通道脑电信号中检测跨被试疲劳驾驶。该模型采用紧凑网络结构设计,将深度神经网络与可解释技术相结合,实现对每一样本进行分类决策的神经生理特征可视化。通过在数据集上的实验验证和实测验证,证明了模型的轻量高效性。可视化结果表明我们的模型可以依据可解释证据进行有效的分类决策,例如疲劳状态的证据有 Alpha 纺锤波和 Theta-Delta 波段爆发,清醒状态的证据有肌肉伪影、传感器信号漂移等。针对基于 EEG 的无校准脑机接口方向,探索深度学习模型发现与不同精神状态相关且有价值的脑电模式具有重要意义。

参考文献

[1] ADMINISTRATION N H T S. 2018 fatal motor vehicle crashes: Overview [R]. Traffic Safety Facts Research Note, Report, Washington, 2019.

[2] 陈玲玲, 毕晓君. 基于残差收缩网络的睡眠脑电分期 [J]. 仪器仪表学报, 2022, 43(2): 148-155.
 CHEN L L, BI X J. Sleep EEG staging based on the residual shrinkage network [J]. Chinese Journal of Scientific Instrument, 2022, 43(2): 148-155.

[3] CUI J, JIN X, HU H, et al. Dynamic distribution alignment with dual-subspace mapping for cross-subject driver mental state detection [C]. IEEE Transactions on Cognitive and Developmental Systems, 2021.

[4] SHAHBAKHTI M, BEIRAMVAND M, REJER I, et al. Simultaneous eye blink characterization and elimination from low-channel prefrontal EEG signals enhances driver drowsiness detection [J]. IEEE Journal of Biomedical and Health Informatics, 2022, 26(3): 1001-1012.

[5] WU E Q, DENG P Y, QU X Y, et al. Detecting fatigue status of pilots based on deep learning network using EEG signals [J]. IEEE Transactions on Cognitive and Developmental Systems, 2021, 13(3): 575-585.

[6] ZHANG Y K, GUO R Q, PENG Y, et al. An auto-weighting incremental random vector functional link network for EEG-based driving fatigue detection [J]. IEEE Transactions on Instrumentation and Measurement,

- 2022, 71: 1-14.
- [7] KO W, KWANSEOK O, JEON E, et al. VIGNet: A deep convolutional neural network for EEG-based driver vigilance estimation [C]. IEEE 8th Int. Winter Conf. Brain-Comput. Interface, Gangwon, Korea, 2020: 1-3.
- [8] PAULO J R, PIRES G, NUNES U J. Cross-subject zero calibration Driver' s drowsiness detection: Exploring spatiotemporal image encoding of EEG signals for convolutional neural network classification [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 905-915.
- [9] OGINO M, MITSUKURA Y. Portable drowsiness detection through use of a prefrontal single-channel electroencephalogram [J]. Sensors, 2018, 18 (12): 4477.
- [10] 蔡靖,李玉涛,宋雪丰,等. 基于脑电两节律和 BP 神经网络的运动想象分类研究 [J]. 电子测量与仪器学报, 2020, 34(6): 176-182.
- CAI J, LI Y T, SONG X F, et al. Classification of motor imagery based on two rhythms of EEG and BP neural network [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(6): 176-182.
- [11] HU J F, MIN J J. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model [J]. Cognitive Neurodynamics, 2018, 12(4): 431-440.
- [12] LAWHERN V J, SOLON A J, WAYTOWICH N R, et al. EEGNet: A compact convolutional neural network for EEG-based brain-computer interfaces [J]. Journal of Neural Engineering, 2018, 15(5): 056013.
- [13] SCHIRRMESTER R T, SPRINGENBERG J T, FIEDERER L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization [J]. Human Brain Mapping, 2017, 38(11): 5391-5420.
- [14] DING S, YUAN Z Y, AN P F, et al. Cascaded convolutional neural network with attention mechanism for mobile EEG-based driver drowsiness detection system [C]. 2019 IEEE International Conference on Bioinformatics and Biomedicine, 2019: 1457-1564.
- [15] FRAIWAN L, ALKHODARI M. Classification of focal and non-focal epileptic patients using single channel EEG and long short-term memory learning system [J]. IEEE Access, 2020, 8: 77255-77262.
- [16] SUPRATAK A, HAO D, CHAO W, et al. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG [J]. IEEE Trans. Neural Syst. Rehabil. Eng, 2017, 25 (11): 1998-2008.
- [17] CUI J, LAN Z R, LIU Y S, et al. A compact and interpretable convolutional neural network for cross-subject driver drowsiness detection from single-channel EEG [J]. Methods, 2021, 202: 173-184.
- [18] JIANG C K, LV W P, LI J Q. Protein-protein interaction sites prediction using batch normalization based CNNs and oversampling method borderline-SMOTE [J]. ACM Transactions on Computational Biology and Bioinformatics, 2023: 1-11.
- [19] ZHOU W, ZHU H Y, SHEN N, et al. A lightweight segmented attention network for sleep staging by fusing local characteristics and adjacent information [J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 238-247.
- [20] CAO Z H, CHUANG CH H, KING J K, et al. Multi-channel EEG recordings during a sustained-attention driving task [J]. Sci. Data, 2019, 6(19): 1-8.
- [21] DU G L, LI T, LI CH Q, et al. Vision-based fatigue driving recognition method integrating heart rate and facial features [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(5): 3089-3100.
- [22] YI Y, ZHANG H, ZHANG W, et al. Fatigue working detection based on facial multifeature fusion [J]. IEEE Sensors Journal, 2023, 23(6): 5956-5961.
- [23] SIMON M, SCHMIDT E A, KINCSES W E, et al. EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions [J]. Clin. Neurophysiol. 2011, 122(6): 1168-1178.
- [24] BRITTON J W, FREY L C, HOPP J L, et al. Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children,

and infants [M]. Chicago: American Epilepsy Society, 2016.

- [25] HUSSEIN M K, SALMAN T M, MIRY A H, et al. Driver drowsiness detection techniques: A survey [C]. 2021 1st Babylon International Conference on Information Technology and Science, 2021: 45-51.

作者简介



冯笑, 2011年于郑州大学获得学士学位, 2014年于大连工业大学获得硕士学位, 现为重庆邮电大学博士研究生、河南省高速铁路运营维护工程研究中心讲师, 主要研究方向为智能感知处理、脑机接口。

E-mail: d200101006@stu.cqpt.edu.cn

Feng Xiao received his B. Sc. degree from Zhengzhou University in 2011, and received his M. Sc. degree from Dalian Polytechnic University in 2014. He is currently a Ph. D. candidate at Chongqing University of Posts and Telecom-

munications. He is also a lecturer at Henan High-Speed Railway Operation and Maintenance Engineering Research Center. His main research interests include intelligent sensitive processing and brain-computer interface.



代少升(通信作者), 1998年于昆明理工大学获得学士学位, 2001年于重庆大学获得硕士学位, 2004年于重庆大学获得博士学位, 现为重庆邮电大学教授, 主要研究方向为人工智能、图像智能处理。

E-mail: daiss@cqpt.edu.cn

Dai Shaosheng (Corresponding author) received his B. Sc. degree from Kunming University of Science and Technology in 1998, received his M. Sc. degree from Chongqing University in 2001, and received his Ph. D. degree from Chongqing University in 2004. He is currently a professor at Chongqing University of Posts and Telecommunications. His main research interests include artificial intelligence and image intelligent processing.