

“噪声标签”下的运动想象多尺度时空特征学习

刘卓恒, 杨丰, 詹长安

(南方医科大学生物医学工程学院, 广州 510515)

摘要: 在运动想象脑电信号采集过程中, 因受试者注意力不集中而未严格遵从提示进行对应的运动想象, 导致所采集脑电数据与提示(标签)不一致, 即出现“噪声标签”, 降低了模型捕捉关键特征的能力, 影响模型在新受试者上的泛化。基于此, 本文提出一种“噪声标签”下多尺度时空特征学习的运动想象分类方法。首先, 采用卷积神经网络提取脑电信号多尺度局部时间特征, 降低个体间差异性影响; 其次, 在时空维度上分块划分特征图, 作为Transformer模块输入, 利用时空特征融合模块, 优化全局时空特征; 最后, 引入对称交叉熵损失, 将交叉熵计算方式扩展到所有类别, 降低“噪声标签”的影响。在PhysioNet和BCI IV 2a运动想象数据集上的实验结果表明, 本文方法的平均准确率优于其他方法, 其中在PhysioNet数据集上引入对称交叉熵损失, 二、三和四分类的平均准确率分别提升0.09%、0.65%和0.66%。此外, 在不同比例的“噪声标签”干扰下, 无需增加模型参数量和计算量, 对称交叉熵损失就能改善模型的分类性能与鲁棒性。

关键词: 运动想象; 脑电信号分类; Transformer; 对称交叉熵损失; 脑机接口

中图分类号: TP391

文献标志码: A

Multi-scale Spatio-Temporal Feature Learning for Motor Imagery in “Noisy Labels”

LIU Zhuoheng, YANG Feng, ZHAN Chang'an

(School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China)

Abstract: During the collection process of electroencephalogram (EEG) signals for motor imagery, the subjects' lack of concentration and failure to strictly follow instructions for corresponding motor imagery result in EEG data that does not match the instructions (labels), leading to the emergence of “noisy labels”. The presence of “noisy labels” reduces the model's ability to capture key features and affects the model's generalization on new subjects. Therefore, this paper proposes a method for motor imagery classification under “noisy labels” condition using multi-scale spatio-temporal feature learning. Firstly, a convolutional neural network is used to extract multi-scale local temporal features from EEG signals, reducing the impact of inter-subject variability. Secondly, feature maps are partitioned in spatio-temporal dimensions and served as input to the Transformer module, with a spatio-temporal feature fusion module used to optimize global spatio-temporal features. Finally, symmetric cross entropy loss is introduced, extending the calculation of cross entropy to all categories to reduce the impact of “noisy labels”. Experimental results on the PhysioNet and BCI IV 2a motor imagery datasets demonstrate that the average accuracy of the proposed method is superior to those of other methods. On the PhysioNet dataset, the

introduction of symmetric cross entropy loss improves the average accuracy for two-, three-, and four-class classifications by 0.09%, 0.65%, and 0.66%, respectively. Moreover, symmetric cross entropy loss can improve the model's classification performance and robustness under different proportions of "noisy labels" interference without increasing the model's parameter quantity and computational complexity.

Key words: motor imagery; classification of electroencephalogram (EEG) signals; Transformer; symmetric cross entropy loss; brain-computer interface (BCI)

引 言

脑机接口(Brain-computer interface, BCI)旨在建立人脑与外部设备之间直接连接通路,无需依赖于正常输出通路(周围神经系统和肌肉运动),便可使人通过脑电信号(Electroencephalogram, EEG)或其他脑部生物信号来控制外设^[1]。BCI应用领域非常广泛,包括但不仅限于医疗康复^[2]、辅助设备^[3]、娱乐^[4]和通信^[5]。在医疗领域中,为传统控制方式受限的人群提供了一种新的交互和控制方式,这将改善他们的生活质量,以及增加日常生活独立性。因此,准确解码大脑信号对BCI广泛应用具有重要意义。

运动想象(Motor imagery, MI)是BCI常用范式之一,受试者通过想象不同肢体的运动来激发大脑皮层中的不同区域,能引发和实际肢体运动相似的皮层活动^[6]。对于非侵入式BCI,EEG因其成本低、易用便捷和时间分辨率高,已成为BCI研究和应用中首选的神经信号记录方式^[7]。MI-EEG准确解码有助于患者康复和控制外设。然而,EEG具有非平稳、低信噪比和个体差异大的特点^[1],这将显著增加MI-EEG分类任务的难度。

在对MI-EEG分类时,尽管卷积神经网络(Convolutional neural network, CNN)模型被广泛使用,但CNN受限于感受野(Receptive field)的大小,需要逐渐增加卷积层数才能获取全局的信息。不同于CNN,Transformer能通过自注意力机制,同时考虑输入序列上下文信息^[8]。Transformer在自然语言处理、计算机视觉领域大放异彩。以视觉Transformer(Vision Transformer, ViT)^[9]为例,这项研究将Transformer应用于图像分类,其核心思想是:对输入图像进行分块,再输入Transformer,计算不同图像块之间的相关性,可大大提升模型捕获全局信息的能力。相关研究开始应用在脑电信号分类上,包括情绪识别^[10]、睡眠分期^[11]等任务。在MI-EEG分类上,Song等^[12]所提出的S3T模型采用共空间模式(Common spatial pattern, CSP)进行空间滤波后输入Transformer学习全局特征。宋耀莲等^[13]则在S3T的基础上将CSP替换为滤波器组共空间模式(Filter bank CSP, FBCSP)。上述方法都忽视了局部特征学习,且受试者依赖(Subject-dependent)的训练策略难以有效泛化到新受试者上。Ali等^[14]提出ConTraNet,Xie等^[15]利用CNN学习局部时间特征。但是,由于单一卷积核难以有效地适应个体间差异性较大的EEG信号变化,因此影响了在新受试者上应用的性能。此外,ConTraNet和刘月峰等^[16]仅沿时间维度进行分块,忽略了多通道EEG的空间信息。

上述MI-EEG分类方法在受试者保持良好心理状态、严格遵从提示等理想状态下能取得很好的实验效果,即受试者在每次实验中都能按照实验要求完成规定的任务。在认知神经科学实验中,确切地知道受试者真实地在想什么或在做什么是具有挑战性的^[17-18],这也是采集MI-EEG数据最主要的不确定因素。在采集EEG脑电信号过程中,受试者因困倦、走神及注意力不集中等影响,未能完全遵循提示做出相应的反应^[17-18],导致所采集的数据与真实的标签不对应,即出现“噪声标签”,从而影响模型分类性能。为了减轻“噪声标签”影响,Li等^[19]在情绪识别任务中使用联合优化策略,交替更新网络参数和纠正“噪声标签”,较复杂的学习过程增加了计算成本。周鹏^[20]设计了一种基于决策融合的训练样本选择算法,剔除受试者主观因素所引起的不利样本,但决策阈值选择不当会排除有效样本。杜晨晓^[21]提

出的样本质量评估方法同样依赖于阈值的选择。上述方法都难以无缝整合到现有的深度学习架构中，且增加了计算成本。目前，有关运动想象任务中的“噪声标签”研究刚刚起步。

上述研究的局限性在于忽视了局部特征学习与“噪声标签”影响，其模型捕捉关键特征能力降低，从而影响了模型泛化性能。本文提出了充分利用 CNN 和 Transformer 互补优势的混合模型 (Multi-scale spatio-temporal Transformer net, MST-TraNet)。该方法中的多尺度时间卷积模块提取局部时间信息，减轻个体间差异性影响。借鉴 ViT 分块思想，对卷积后多通道 MI-EEG 信号，分别沿时间维度和空间(电极)维度进行分块，再利用自注意力机制学习全局时间和空间特征。借鉴 Lawhern 等^[22]所提出的 EEGNet，空间卷积采用卷积核大小等于通道电极数的卷积核融合空间信息。本文设计一个时空特征融合 CNN 模块，其卷积核大小等于分块的数量，用于融合各个分块的信息，并引入对称交叉熵 (Symmetric cross entropy, SCE) 损失^[23]，提高模型对“噪声标签”的鲁棒性，能无缝集成到现有模型中且几乎不增加计算成本。

1 本文方法

本文提出的 MST-TraNet 模型的示意图如图 1 所示。它主要包括提取局部时间信息的多尺度时间卷积模块、学习全局时空特征的时空 Transformer 模块、聚合优化特征的时空特征融合模块和分类头 4 部分，其中 Transformer 具体结构如图 2 所示。

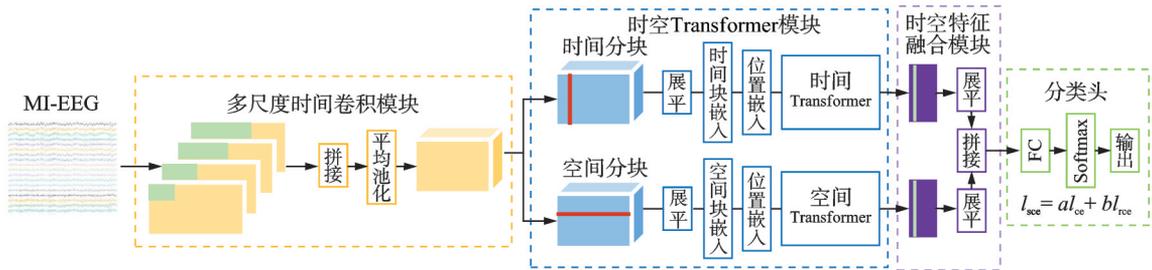


图1 MST-TraNet模型示意图

Fig.1 Diagram of MST-TraNet model

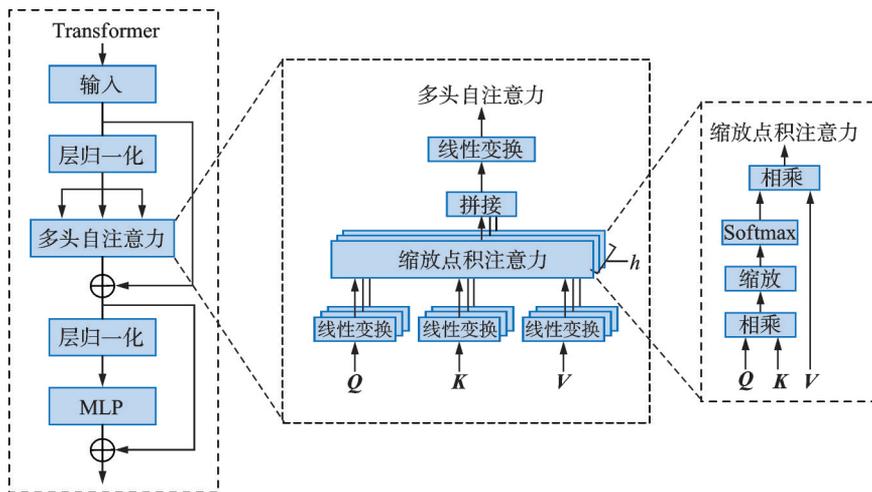


图2 Transformer模型结构示意图

Fig.2 Diagram of Transformer model structure

1.1 多尺度时间卷积模块

EEG 信号不仅有跨个体变异性,而且有跨时间变异性,这给 MI-EEG 准确解码和 BCI 应用带来了挑战。从卷积核的角度, Jia 等^[24]说明 EEG 的跨个体与跨时间变异性,即不同个体和同一个体不同会话 (Session) 的最优卷积核大小是不同的。显然,单一卷积核会限制模型跨个体和跨时间泛化能力。因此,本文采用不同大小的卷积核来提取多尺度时间特征。具体而言,4 条支路卷积核大小分别为 7、15、31 和 63,记为 k_1 至 k_4 。为了保持卷积前后特征图尺寸不变,步长设置为 1,4 条支路填充分别设置为 3、7、15 和 31。在多尺度时间卷积模块中,添加批归一化 (Batch normalization, BN) 层和 Dropout 层缓解过拟合,其中 Dropout 比例设置为 0.5。激活函数的选择与 EEGNet^[22]一致,采用 ELU。为了降低计算复杂度,减小时空 Transformer 模块前输入特征图大小,各分支使用 1×1 卷积减少特征图通道数。最后,各分支特征图沿通道维度拼接,并通过平均池化减少时间维度,同时保留对分类有用的重要特征。

1.2 时空 Transformer 模块

Transformer 的应用始于自然语言处理, ViT 将 Transformer 的应用扩展到计算机视觉领域。在本研究中,借鉴 ViT 对输入图像进行分块的思想,对多尺度时间卷积模块输出的特征图分别沿时间和空间 (通道电极) 维度进行分块,再输入 Transformer 中,利用自注意力机制,分别学习 MI-EEG 信号的全局时间及空间相关性,捕获信号远程依赖关系。

将原始 MI-EEG 信号或经卷积后得到的特征图记为: $X \in \mathbf{R}^{N_a \times N_s \times C}$ 。当输入为原始 MI-EEG 信号时, C 取值为 1; N_{ch} 表示 MI-EEG 信号的电极数,本文无通道选择, $N_{ch} = 64$; N_s 表示输入序列的采样点个数,运动想象 EEG 数据的前 3 s 且采样率为 160 Hz, $N_s = 480$ 。经卷积后输出特征图,其 C 代表特征图的通道数, N_{ch} 和 N_s 分别代表特征图空间和时间维度的大小。特征图分割成若干个子块后,经过展平和线性变换,产生块嵌入 (Patch embedding),加上位置编码 (Position encoding),作为时空 Transformer 模块的输入。位置编码采用可学习的位置编码,让模型学习到输入序列中各个位置的重要性。时空 Transformer 模块包含时间 Transformer 和空间 Transformer。对于时间 Transformer,将 X 分割成 N_s/N 个不重叠的块,每个块的大小为 $(64, N)$ 。每个时间块包含 64 个电极中时间长度为 N 的信息,由时间 Transformer 中的自注意力计算每个时间片段之间的相关性。对于空间 Transformer,将 X 分割成 N_{ch} 个不重叠的块,每个块的大小为 $(1, N_s)$ 。每个空间块只包含单个电极信息,由空间 Transformer 中的自注意力计算每个电极之间的相关性。

在 Transformer 中,最核心的结构是多头注意力层 (Multi-head attention layer)。首先,将输入矩阵分别与 3 个不同的权重矩阵相乘,即输入矩阵通过线性变换得到大小相同的 3 个矩阵:查询 (Query, Q) 矩阵、键 (Key, K) 矩阵和值 (Value, V) 矩阵;接着 Q 和 K 矩阵的转置相乘得到不同标记 (Token) 之间的相关性,设计比例因子 \sqrt{d} ,避免梯度爆炸,保证训练稳定;然后将结果输入到激励函数 (Softmax),计算加权矩阵,获得注意力分数 (Attention score);最后,注意力分数矩阵乘以 V 进行加权。上述过程可表示为

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

式中 d 表示 Q 和 K 的维度。多头注意力将注意力机制应用于多个不同的子空间,从不同的角度学习到更加全面丰富的特征,主要区别在于:多头注意力会根据注意力头的数目 h 对 Q 、 K 、 V 矩阵进行均等拆分,即 $d_{q'} = d_{k'} = d_{v'} = d/h$ 。模型通过分别学习多组 Q 、 K 、 V 参数,每组称为一个头,从而使模型能够关注不同位置的不同特征。每个头计算出一个输出,然后这些输出通过拼接进行融合。该过程可表示为

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^O \quad (2)$$

$$\text{head}_i = \text{Attention}(\mathbf{Q}W_i^Q, \mathbf{K}W_i^K, \mathbf{V}W_i^V) \quad (3)$$

式中:Concat表示对矩阵进行拼接;head_{*i*}代表第*i*个头的输出;*h*代表注意力头的个数; W^O 表示对最终输出进行线性变换; W_i^Q 、 W_i^K 和 W_i^V 分别表示 \mathbf{Q} 、 \mathbf{K} 和 \mathbf{V} 在head_{*i*}上的线性变换。

1.3 时空特征融合模块

EEGNet^[22]中的空间卷积层融合所有电极的空间信息,其卷积核大小与输入脑电信号的电极数相等。受此启发,本文设计时空特征融合模块,对时空Transformer模块输出的时空特征进行聚合优化,其卷积核大小等于分块的数量。在空间Transformer和时间Transformer中的分块数量分别为64和60,因此对应支路卷积核大小分别为64和60。

1.4 对称交叉熵损失

在MI-EEG采集过程中,受试者需要根据屏幕上的提示执行特定的运动想象,受试者在运动想象时必须保持专注。现有研究往往假设受试者都能按照实验要求完成任务,但长时间实验、外界环境干扰会使受试者疲劳、走神,未严格遵从实验要求进行相应的运动想象,或未进行运动想象,导致采集数据包含“噪声标签”,最终降低深度学习模型的泛化性能。分类常用的交叉熵(Cross entropy, CE)损失函数对“噪声标签”不鲁棒^[25],这可能导致模型学习到错误模式,影响其性能和泛化能力。故本研究引入对称交叉熵损失,将传统交叉熵损失的计算范围从正确类别扩展至所有类别。通过对“噪声标签”鲁棒的反向交叉熵(Reverse cross entropy, RCE),减轻“噪声标签”的负面影响。损失函数的相关公式为

$$l_{ce} = - \sum_{k=1}^K q(k|x) \log p(k|x) \quad (4)$$

$$l_{rce} = - \sum_{k=1}^K p(k|x) \log q(k|x) \quad (5)$$

$$l_{sce} = al_{ce} + bl_{rce} = -a \sum_{k=1}^K q(k|x) \log p(k|x) - b \sum_{k=1}^K p(k|x) \log q(k|x) \quad (6)$$

式中: K 为分类类别数, x 为样本, $q(k|x)$ 为样本 x 的真实标签分布为且 $\sum_{k=1}^K q(k|x) = 1$, $p(k|x)$ 为模型预测的分布; l_{ce} 为交叉熵损失, l_{rce} 为反向交叉熵损失, l_{sce} 为对称交叉熵损失, a 用来调整 l_{ce} 的拟合程度, b 用来探索 l_{rce} 对“噪声标签”的鲁棒性。

2 数据集及实验设置

2.1 数据集

本研究使用PhysioNet的MI-EEG数据集^[26]和BCI IV 2a数据集^[27]验证所提方法的有效性。PhysioNet数据集通过BCI2000系统记录109名受试者的运动和运动想象脑电数据,其通道数为64,采样率为160 Hz。由于部分数据未按规定的长度记录,因此排除第88、92、100和104位受试者的数据。BCI IV 2a数据集记录了9名受试者的运动想象脑电数据,其通道数为22,采样率为250 Hz。

目前研究大多为每位受试者单独训练模型,即受试者依赖的训练策略,这有助于实现更高的分类精度,然而这种训练策略在个体间的泛化能力较差。因此对于PhysioNet数据集,本研究使用运动想象的前3 s数据,按受试者独立(Subject-independent)的训练策略,进行个体间的二、三和四分类实验。二分类为区分想象左手及右手(L/R),三分类为区分想象左手、右手和睁眼基线(L/R/O),四分类为区分想象左手、右手、睁眼基线和双脚(L/R/O/F)。考虑到想象左右手和双手会包含相同的特征,所以四分类不选择想象双手运动的数据。参考文献[28]中的试次选择标准,为了类别数据平衡,本研究同样选择每轮实验

中各类别的前7试次。每位受试者各类别均进行了3轮实验,即每位受试者各类别选取的试验次数均为21。对105位受试者进行五折交叉验证,每一折使用84位受试者的数据作为训练集,其余21位受试者的数据则作为测试集。获得5组分类准确率后,计算其平均值作为最终分类准确率。对于BCI IV 2a数据集,本研究使用从第2 s到第6 s的4 s数据,按受试者独立的训练策略,进行个体间的四分类实验(左手/右手/双脚/舌头)。对9位受试者进行九折交叉验证,每一折使用1位受试者的数据作为训练集,其余8位受试者的数据则作为测试集。获得9组分类准确率后,计算其平均值作为最终分类准确率。

2.2 实验设置

为了平衡模型复杂性和分类准确率,采用两层Transformer,分块线性变换的维度设置为32。MLP层采用GeLU激活函数,维度设置为64。多头注意力层的头数设置为4。多尺度时间CNN模块的Dropout值设为0.5。所提方法在Pytorch和GeForce GTX 1080Ti GPU上实现,使用Adam优化器训练50轮。

3 实验及结果分析

3.1 分类准确率

在PhysioNet数据集按受试者独立的训练策略,将MST-TraNet与几种代表性基线模型进行比较,结果如表1所示。基线模型的简要描述如下:EEGNet是采用深度可分离卷积的轻量级CNN模型;DeepConvNet和ShallowConvNet是具有不同数量卷积层的CNN模型;ConTraNet和F-CTrans都是单一尺度CNN与Transformer相结合的混合模型。

从表1可以看出,MST-TraNet在3种分类场景(L/R、L/R/O和L/R/O/F)的平均准确率分别为83.65%、76.21%和67.34%,均优于其他方法。这表明MST-TraNet在解码运动想象任务时具有更好的性能。对于二分类,MST-TraNet准确率与ConTraNet差别不大,说明二分类无需太复杂的模型。观察不同分类场景下的准确率,随着分类类别的增加,模型的性能相应下降,这表明实现更多指令的外设控制是具有挑战性的。结合Transformer的方法在准确率上明显优于纯CNN的方法,这说明混合模型能有效结合Transformer

和CNN的互补优势,学习到更加全面丰富的特征表示。相比ConTraNet和F-CTrans,MST-TraNet通过学习多尺度时间特征,降低个体间差异性影响,提高了在新受试者上的分类准确率。

在BCI IV 2a数据集上按受试者独立的训练策略,将MST-TraNet与3种开源的CNN模型进行比较,结果如表2所示。MST-TraNet得益于其多尺度CNN能够学习到多尺度局部时间特征,以及时空Transformer能够捕捉全局时空特征,因此在BCI IV 2a数据集上的平均准确率超过了这3种CNN模型。

表1 不同模型在PhysioNet数据集上的分类结果

Table 1 Classification results of different models on the PhysioNet dataset %

模型	准确率		
	L/R	L/R/O	L/R/O/F
EEGNet ^[22]	81.79	71.94	61.63
DeepConvNet ^[29]	80.93	72.44	62.62
ShallowConvNet ^[29]	80.52	71.40	59.41
F-CTrans ^[15]	82.95	74.44	64.22
ConTraNet ^[14]	83.61	74.38	65.44
MST-TraNet	83.65	76.21	67.34

表2 不同模型在BCI IV 2a数据集上的分类结果

Table 2 Classification results of different models on the BCI IV 2a dataset %

模型	EEGNet ^[22]	DeepConvNet ^[29]	ShallowConvNet ^[29]	MST-TraNet
准确率	60.05	63.35	57.5	66.4

3.2 消融实验

为了验证模型中各个模块的有效性,在 PhysioNet 数据集上进行一系列消融实验。以时间 Transformer 模块(Model 0)为基准,逐步添加空间 Transformer 模块(Model 1)、多尺度时间卷积模块(Model 2)、时空特征融合模块(Model 3)和 SCE 损失(Model 4),实验结果如表 3 所示。表 3 中的“噪声标签”比例设置为 0,即使用原始标签,不额外添加标签噪声,但原始标签中可能存在“噪声标签”。

Model 0 是仅使用时间 Transformer 的基准模型,在 3 个不同的分类场景下,分别取得 82.27%、73.61% 和 63.90% 的准确率,已经优于对比的纯 CNN 模型,表明 Transformer 模型具有更强的特征学习能力。在 Model 0 基础上引入空间 Transformer 后,准确率分别提高了 0.9%、0.87% 和 1.09%,表明空间 Transformer 学习到的全局空间特征能有效补充时间 Transformer 学到的全局时间特征,有效提高性能。Model 2 在 Model 1 的基础上引入多尺度时间卷积,在四分类任务中取得了最显著的提升,提高了 1.43%。这表明多尺度时间卷积学习的局部时间特征对模型在更复杂的分类场景中的性能提升起到重要作用。Model 3 在 Model 2 基础上引入时空特征融合模块,在三分类任务上提升最显著,提高了 0.85%。这说明时空特征融合模块有助于优化特征,提高模型分类性能。Model 4 在 Model 3 基础上引入 SCE 损失,在 3 个不同的分类场景下,准确率分别提升 0.09%、0.65% 和 0.66%,这表明相较于 CE 损失,SCE 损失能有效减轻“噪声标签”的影响。值得注意的是,“噪声标签”主要集中在睁眼基线和想象双脚这两个类别中,两分类由于类别少出错概率更低。综上所述,通过依次引入不同模块,在不同分类场景下分类准确率逐步得到提升,这验证了各个模块的有效性,也为选择最终模型提供了指导,使其在多个场景下均能表现出色。

3.3 模型复杂度与计算成本

为了评估模型在实际应用中的可行性,本文以 PhysioNet 数据集中的四分类任务为例,采用参数量(Params)和计算量(Floating point operations, FLOPs)作为评估指标,讨论模型复杂度和计算成本。表 4 列举了 MST-TraNet 中各模块与 3 种开源 CNN 模型在参数量和计算量上的对比结果。

在时间 Transformer(Model 0)基础上加入空间 Transformer 得到 Model 1,模型的参数量和计算量分别变为原来的 1.99 倍和 2.06 倍。尽管与使用深度可分离卷积的 EEGNet 相比,参数量并没有优势,但计算量比 EEGNet 低。在 Model 1 基础上增加包含 4 条支路的多尺度 CNN 得到 Model 2,虽然仅增加 1.11×10^3 的参数量就能提高 1.43% 的准确率,但计算量却变为原来的 4.11 倍。因此,为了平衡准确率和计算量,可以考虑使用单尺度 CNN。虽然这会牺牲对多尺度时间特征的学习,但能有效控制

表 3 MST-TraNet 在 PhysioNet 数据集上的消融实验结果

Table 3 Ablation experimental results of MST-TraNet on the PhysioNet dataset %

模型	准确率		
	L/R	L/R/O	L/R/O/F
Model 0	82.27	73.61	63.90
Model 1	83.17	74.48	64.99
Model 2	83.38	74.71	66.42
Model 3	83.56	75.56	66.68
Model 4	83.65	76.21	67.34

表 4 不同模型在 PhysioNet 数据集的参数量和计算量

Table 4 Params and FLOPs of different models on the PhysioNet dataset

模型	Params/ 10^3	FLOPs/ 10^6
EEGNet ^[22]	2.07	14.78
DeepConvNet ^[29]	193.30	32.54
ShallowConvNet ^[29]	112.96	63.59
Model 0	98.40	5.54
Model 1	196.23	11.39
Model 2	197.34	46.81
Model 3	205.41	47.07
Model 4(MST-TraNet)	205.41	47.07

计算量的增加。在 Model 3 基础上引入 SCE 损失得到 Model 4, 能在不增加参数量和计算量的前提下改善分类准确率, 提高模型对“噪声标签”的鲁棒性。MST-TraNet 的参数量中有 95.53% 来自时空 Transformer 模块, 这表明未来可以通过优化该模块来减少参数量。虽然 MST-TraNet 在参数量和计算量上都不是最优的, 在参数量上比 DeepConvNet 高 12.11×10^3 , 在计算量上比 ShallowConNet 低 16.52×10^6 , 但它的平均准确率最高。

3.4 参数设置

本文以 PhysioNet 数据集中的四分类任务为例, 考察 Transformer 层数、输入矩阵线性变换(投影)维度、MLP 维度以及注意力头个数对模型性能、参数量和计算量的影响, 其中 MLP 维度设定为投影维度的 2 倍, 实验结果如表 5 所示。结果表明, 增大 Transformer 中的参数设置可以略微改善准确率, 但会显著增加参数量和计算量。综合考虑准确率和模型复杂度, 本研究选择 2 层 Transformer、线性变换维度为 32 以及注意力头数为 4 的参数配置。这一选择是在保证准确率的同时, 有效控制模型参数, 避免单纯增加注意力头个数带来冗余。F-Ctrans 使用 3 层双支路的 Transformer, 注意力头个数为 8, 而 ConTraNet 使用了 1 层时间 Transformer, 注意力头个数为 8。因此, MST-TraNet 在参数量上相对于 F-Ctrans 有优势, 但高于 ConTraNet。

表 5 Transformer 中不同参数的分类准确率、参数量和计算量

Table 5 Classification accuracy, Params and FLOPs for different parameters in Transformer

层数	线性变换维度	MLP 维度	注意力头个数	准确率/%	Params/ 10^3	FLOPs/ 10^6
1	32	64	4	66.51	131.17	42.37
1	32	64	8	66.70	196.70	46.53
2	32	64	4	67.34	205.41	47.07
2	32	64	8	66.83	336.48	55.38
3	32	64	8	66.75	476.26	64.23
3	64	128	8	67.09	992.48	95.54

参考 SCE 损失论文^[23]的参数设置, 对 SCE 损失中的参数 a 和 b 进行实验, 结果如图 3 所示, 其中绿色虚线表示使用交叉熵损失的结果。在图 3(a) 中, 将参数 b 固定为 1, 然后调整参数 a , 当 a 设置为 3 时结果最好; 在图 3(b) 中, 将 a 固定为 3, 然后调整参数 b , 当 b 设置为 1 时结果最好。结果表明 SCE 损失能提高分类准确率, 且参数 b 比 a 更敏感。

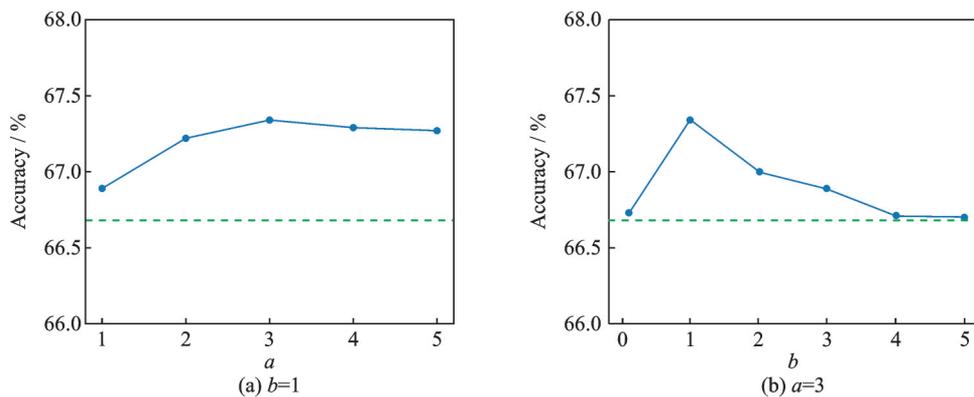


图 3 调整 SCE 损失中参数 a 和 b 的实验结果

Fig.3 Experimental results of adjusting parameters a and b in SCE loss

3.5 “噪声标签”

考虑到不同受试者专注程度不同,本文以 PhysioNet 数据集中的四分类任务为例,在训练集中使用不同比例的“噪声标签”进行实验,验证引入对称交叉熵损失的有效性,以及进一步分析模型中各模块对“噪声标签”的鲁棒性,并与3种开源 CNN 模型进行了对比,结果如表6所示。参照文献[19],“噪声标签”使用的是对称标签噪声。从表6可以看出,随着各模块的加入,在不同“噪声标签”比例 p 下,Model 0 到 Model 4 的性能逐渐提高。各模型的性能随着“噪声标签”比例的增加而下降,这表明“噪声标签”对模型性能有负面影响。具体而言,当“噪声标签”比例从0增加到30%时,EEGNet、DeepConvNet、ShallowConvNet、Model 0 到 Model 4 的准确率分别下降3%、4.47%、3.81%、5.07%、5.75%、4.9%、5.01%和3.89%。包含时间 Transformer 和空间 Transformer 的 Model 1 准确率下降最多,局部时间信息的丢失降低了模型对“噪声标签”的鲁棒性。与 EEGNet 和 ShallowConvNet 相比,引入 SCE 损失的 Model 4 在准确率下降方面表现次之。3个 CNN 模型的准确率低于 Transformer 模型,但它们的准确率下降幅度较小。这表明,虽然 CNN 模型的准确率较低,但它们对“噪声标签”的过拟合风险也相对较低。此外,与 Model 3 相比,Model 4 在所有“噪声标签”比例下的准确率都有提高,这表明在处理带有标签噪声的数据时,SCE 损失相对于 CE 损失具有更好的鲁棒性和性能。

表6 不同“噪声标签”比例下各模型的分​​类准确率

Table 6 Classification accuracy of various models under different proportions of “noisy labels” %

模型	准确率			
	$p=0\%$	$p=10\%$	$p=20\%$	$p=30\%$
EEGNet ^[22]	61.92	61.43	60.36	58.92
DeepConvNet ^[29]	62.34	61.11	59.67	57.87
ShallowConvNet ^[29]	59.33	58.71	58.03	55.52
Model 0	63.90	62.06	60.61	58.83
Model 1	64.99	63.59	61.56	59.24
Model 2	66.42	65.43	63.80	61.52
Model 3	66.68	65.59	64.24	61.67
Model 4(MST-TraNet)	67.34	66.05	64.76	63.45

4 结束语

针对“噪声标签”影响模型分类性能以及 Transformer 特征学习不充分的问题,本文提出了一种“噪声标签”下结合 CNN 和 Transformer 互补优势的运动想象脑电信号分类模型 MST-TraNet。MST-TraNet 是端到端算法,仅需简单的标准化预处理,通过各个模块自动提取和联合优化特征,提高了分类的整体效率和性能。首先,采用多尺度卷积模块捕获局部时间特征,减轻个体间差异性影响。其次,利用 Transformer 的自注意力机制关注时间序列中的长期依赖关系,学习全局时空特征。然后,受 EEGNet 启发,设计了一个时空特征融合模块来优化时空 Transformer 模块的输出。最后,引入对称交叉熵损失,提高模型对“噪声标签”的鲁棒性。因此,与 Transformer 模型相比,该方法能学习到更丰富的判别表示。实验结果表明,该方法在受试者独立的训练策略下准确率优于现有算法,并且在运动想象脑电采集过程中确实存在“噪声标签”,特别是在睁眼和想象双脚任务中。此外,在 BCI 设备实际推广中,引入 SCE 损失可以进一步提高性能。然而,本文算法仍有一些局限性。虽然模型在目前的实验中表现优异,但模型中的线性变换和全连接层有大量参数,增加了训练成本。因此,未来将优化模型结

构,减少参数量。此外,将利用数据增强技术增加训练样本,进一步改善分类结果。最后,脑电信号表现出显著的个体间和个体内差异。考虑到源域和目标域数据分布的差异,将探索域自适应和域泛化技术,提高模型在新受试者中的泛化能力。

参考文献:

- [1] ALTAHERI H, MUHAMMAD G, ALSULAIMAN M, et al. Deep learning techniques for classification of electroencephalogram (EEG) motor imagery (MI) signals: A review[J]. *Neural Computing and Applications*, 2023, 35: 14681-14722.
- [2] CANTILLO-NEGRETE J, CARINO-ESCOBAR R I, CARRILLO-MORA P, et al. Motor imagery-based brain-computer interface coupled to a robotic hand orthosis aimed for neurorehabilitation of stroke patients[J]. *Journal of Healthcare Engineering*, 2018. DOI:10.1155/2018/1624637.
- [3] AL-QURAISHI M S, ELAMVAZUTHI I, DAUD S A, et al. EEG-based control for upper and lower limb exoskeletons and prostheses: A systematic review[J]. *Sensors*, 2018, 18(10): 3342.
- [4] DJAMAL E C, ABDULLAH M Y, RENALDI F. Brain computer interface game controlling using fast Fourier transform and learning vector quantization[J]. *Journal of Telecommunication, Electronic and Computer Engineering*, 2017, 9(2/3/4/5): 71-74.
- [5] ZHANG X, YAO L, SHENG Q Z, et al. Converting your thoughts to texts: Enabling brain typing via deep feature learning of EEG signals[C]//*Proceedings of 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. Athens, Greece: IEEE, 2018: 1-10.
- [6] CHAUDHARY U, BIRBAUMER N, RAMOS-MURGUIALDAY A. Brain-computer interfaces for communication and rehabilitation[J]. *Nature Reviews Neurology*, 2016, 12(9): 513-525.
- [7] MILLÁN J R, RUPP R, MÜLLER-PUTZ G R, et al. Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges[J]. *Frontiers in Neuroscience*, 2010, 4: 161.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*. [S.l.]: [s.n.], 2017: 6000-6010.
- [9] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03)[2024-12-01]. <https://arxiv.org/abs/2010.11929v2>.
- [10] 张学军, 王天晨, 王泽田. 基于多域信息融合的卷积Transformer脑电情感识别模型[J]. *数据采集与处理*, 2024, 39(6): 1543-1552.
ZHANG Xuejun, WANG Tianchen, WANG Zetian. Convolutional Transformer EEG emotion recognition model based on multidomain information fusion[J]. *Journal of Data Acquisition and Processing*, 2024, 39(6): 1543-1552.
- [11] KIM D K, KIM D, LEE J G, et al. Deep learning application to clinical decision support system in sleep stage classification[J]. *Journal of Personalized Medicine*, 2022, 100: S293.
- [12] SONG Y, JIA X, YANG L, et al. Transformer-based spatial-temporal feature learning for EEG decoding[EB/OL]. (2021-06-11)[2024-12-01]. <https://arxiv.org/abs/2106.11170>.
- [13] 宋耀莲, 殷喜喆, 杨俊. 基于时空特征学习Transformer的运动想象脑电解码方法[J]. *南京大学学报(自然科学)*, 2023, 59(2): 313-321.
SONG Yaolian, YIN Xizhe, YANG Jun. Transformer based on temporal-spatial feature learning for motor imagery electroencephalogram signal decoding[J]. *Journal of Nanjing University (Natural Science)*, 2023, 59(2): 313-321.
- [14] ALI O, SAIF-UR-REHMAN M, GLASMACHERS T, et al. ConTraNet: A hybrid network for improving the classification of EEG and EMG signals with limited training data[J]. *Computers in Biology and Medicine*, 2024, 168: 107649.
- [15] XIE J, ZHANG J, SUN J, et al. A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification[J]. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022, 30: 2126-2136.
- [16] 刘月峰, 刘好峰, 王越, 等. 基于改进Transformer模型的运动想象脑电分类方法研究[J]. *计量学报*, 2023, 44(7): 1147-1153.

- LIU Yuefeng, LIU Haofeng, WANG Yue, et al. Research on motor imagery EEG classification method based on improved Transformer[J]. Acta Metrologica Sinica, 2023, 44(7): 1147-1153.
- [17] BANVILLE H, CHEHAB O, HYVÄRINEN A, et al. Uncovering the structure of clinical EEG signals with self-supervised learning[J]. Journal of Neural Engineering, 2021, 18(4): 046020.
- [18] CHEN X, LI C, LIU A, et al. Toward open-world electroencephalogram decoding via deep learning: A comprehensive survey[J]. IEEE Signal Processing Magazine, 2022, 39(2): 117-134.
- [19] LI C, HOU Y, SONG R, et al. Multi-channel EEG-based emotion recognition in the presence of noisy labels[J]. Science China Information Sciences, 2022, 65(4): 60-75.
- [20] 周鹏. 基于运动想象的脑机接口的研究[D]. 天津: 天津大学, 2007.
ZHOU Peng. The research of brain-computer interface based on motor imagery[D]. Tianjin: Tianjin University, 2007.
- [21] 杜晨晓. 运动想象 BCI 训练样本优化及集成分类方法研究[D]. 合肥: 安徽大学, 2022.
DU Chenxiao. Research on training samples optimization and ensemble classification methods for motor imagery BCI[D]. Hefei: Anhui University, 2022.
- [22] 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.
- [23] WANG Y, MA X, CHEN Z, et al. Symmetric cross entropy for robust learning with noisy labels[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE, 2019: 322-330.
- [24] JIA Z, LIN Y, WANG J, et al. MMCNN: A multi-branch multi-scale convolutional neural network for motor imagery classification[C]//Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD). Cham: Springer International Publishing, 2021: 736-751.
- [25] GHOSH A, KUMAR H, SASTRY P S. Robust loss functions under label noise for deep neural networks[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI, 2017: 1919-1925.
- [26] GOLDBERGER A L, AMARAL L A N, 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.
- [27] TANGERMANN M, MÜLLER K R, AERTSEN A, et al. Review of the BCI competition IV[J]. Frontiers in Neuroscience, 2012. DOI:10.3389/fnins.2012.00055.
- [28] WANG X, HERSCHE M, TÖMEKCE B, et al. An accurate eegnet-based motor-imagery brain-computer interface for low-power edge computing[C]//Proceedings of 2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA). Bari, Italy: IEEE, 2020: 1-6.
- [29] SCHIRMEISTER 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.

作者简介:



刘卓恒(1999-),男,硕士研究生,研究方向:运动想象、脑机接口及生物医学信号处理, E-mail:535320621@qq.com。



杨丰(1965-),通信作者,男,教授,研究方向:模式识别、机器学习及生物医学信号处理等, E-mail: yangf@smu.edu.cn。



詹长安(1972-),男,研究员,研究方向:智能医学仪器、神经信息工程及生物医学信号处理。

(编辑:王静)