



基于域间Mixup微调策略的跨被试运动想象脑电信号分类算法

蒋云良, 周阳, 张雄涛, 苗敏敏, 张永

引用本文:

蒋云良, 周阳, 张雄涛, 苗敏敏, 张永. 基于域间Mixup微调策略的跨被试运动想象脑电信号分类算法[J]. 智能系统学报, 2024, 19(4): 909–919.

JIANG Yunliang, ZHOU Yang, ZHANG Xiongtao, et al. Cross-subject motor imagery EEG classification based on inter-domain Mixup fine-tuning strategy[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(4): 909–919.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202208017>

您可能感兴趣的其他文章

样本仿真结合迁移学习的声呐图像水雷检测

Detection of underwater mine target in sidescan sonar image based on sample simulation and transfer learning
智能系统学报. 2021, 16(2): 385–392 <https://dx.doi.org/10.11992/tis.202101030>

基于迁移学习的无监督跨域人脸表情识别

Unsupervised cross-domain expression recognition based on transfer learning
智能系统学报. 2021, 16(3): 397–406 <https://dx.doi.org/10.11992/tis.202008034>

基于分类差异与信息熵对抗的无监督域适应算法

Unsupervised domain adaptation algorithm based on classification discrepancy and information entropy
智能系统学报. 2021, 16(6): 999–1006 <https://dx.doi.org/10.11992/tis.202010020>

基于改进型BP神经网络的手部动作识别

Hand-motion recognition based on improved BP neural network
智能系统学报. 2018, 13(5): 848–854 <https://dx.doi.org/10.11992/tis.201703018>

混合脑电信号及视觉信息的智能轮椅人机交互系统

On a hybrid electroencephalograph and visual information intelligent wheelchair human-machine interactive system
智能系统学报. 2016, 11(5): 648–654 <https://dx.doi.org/10.11992/tis.201511004>

基于最小最大概率机的迁移学习分类算法

Transfer learning classification algorithms based on minimax probability machine
智能系统学报. 2016, 11(1): 84–92 <https://dx.doi.org/10.11992/tis.201505024>

DOI: 10.11992/tis.202208017

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240314.1602.005>

基于域间 Mixup 微调策略的跨被试运动想象 脑电信号分类算法

蒋云良^{1,2,3}, 周阳^{1,2}, 张雄涛^{1,2}, 苗敏敏^{1,2}, 张永^{1,2}

(1. 湖州师范学院 信息工程学院, 浙江 湖州 313000; 2. 湖州师范学院 浙江省现代农业资源智慧管理与应用研究重点实验室, 浙江 湖州 313000; 3. 浙江师范大学 计算机科学与技术学院, 浙江 金华 321000)

摘要: 为了缓解传统微调算法的灾难性遗忘问题, 本文提出了一种基于域间 Mixup 微调策略的跨被试运动想象脑电信号分类算法 Mix-Tuning。Mix-Tuning 通过预训练、微调的二阶段训练方式, 实现跨领域知识迁移。预训练阶段, Mix-Tuning 使用源域数据初始化模型参数, 挖掘源域数据潜在信息。微调阶段, Mix-Tuning 通过域间 Mixup, 生成域间插值数据微调模型参数。域间 Mixup 数据增强策略引入源域数据潜在信息, 缓解传统微调算法在样本稀疏场景下的灾难性遗忘问题, 提高模型的泛化性能。Mix-Tuning 被进一步应用于运动想象脑电信号分类任务, 实现了跨被试正向知识迁移。Mix-Tuning 在 BMI 数据集的运动想象任务达到了 85.50% 的平均分类准确率, 相较于被试-依赖和被试-独立训练方式的预测准确率 58.72% 和 84.01%, 分别提高 26.78% 和 1.49%。本文分析结果可为跨被试运动想象脑电信号分类算法提供参考。

关键词: 域间 Mixup; 预训练; 微调; 脑电信号; 运动想象; 跨被试知识迁移; 卷积神经网络; 正则化

中图分类号: TP18 **文献标志码:** A **文章编号:** 1673-4785(2024)04-0909-11

中文引用格式: 蒋云良, 周阳, 张雄涛, 等. 基于域间 Mixup 微调策略的跨被试运动想象脑电信号分类算法 [J]. 智能系统学报, 2024, 19(4): 909-919.

英文引用格式: JIANG Yunliang, ZHOU Yang, ZHANG Xiongtao, et al. Cross-subject motor imagery EEG classification based on inter-domain Mixup fine-tuning strategy[J]. CAAI transactions on intelligent systems, 2024, 19(4): 909-919.

Cross-subject motor imagery EEG classification based on inter-domain Mixup fine-tuning strategy

JIANG Yunliang^{1,2,3}, ZHOU Yang^{1,2}, ZHANG Xiongtao^{1,2}, MIAO Minmin^{1,2}, ZHANG Yong^{1,2}

(1. School of Information Engineering, Huzhou University, Huzhou 313000, China; 2. Zhejiang Province Key Laboratory of Smart Management & Application of Modern Agricultural Resources, Huzhou University, Huzhou 313000, China; 3. School of Computer Science and Technology, Zhejiang Normal University, Jinhua 321000, China)

Abstract: In order to alleviate the catastrophic forgetting problem of vanilla fine-tuning algorithms, we propose a cross-subject motor imagery EEG classification method based on inter-domain Mixup fine-tuning strategy, i.e., Mix-Tuning. Mix-Tuning realizes cross-domain knowledge transfer through a two-stage training manner consisting of pre-training and fine-tuning. In the pre-training stage, Mix-Tuning uses the source domain data to initialize the model parameters and mine potential information of the source domain data. In the fine-tuning stage, Mix-Tuning generates inter-domain interpolation data to fine-tune the model parameters through inter-domain Mixup. Inter-domain Mixup data enhancement strategy introduces latent information of the source domain data, which alleviates the catastrophic forgetting problem of Vanilla Fine-tuning in sparse sample scenarios and improves the generalization performance of the model. Mix-Tuning is further applied to the motor imagery EEG classification task and achieves cross-subject positive knowledge transfer. Mix-Tuning achieved an average classification accuracy of 85.50% on motor imagery task BMI dataset. Compared with 58.72% and 84.01% for Subject-specific and Subject-independent training manner, Mix-Tuning increased by 26.78% and 1.49%, respectively. The analysis results in this paper can provide a reference for cross-subject motor imagery EEG classification algorithm.

Keywords: inter-domain Mixup; pre-training; fine-tuning; electroencephalogram; motor imagery; cross-subject knowledge transfer; convolutional neural network; regularization

收稿日期: 2022-08-17. 网络出版日期: 2024-03-14.

基金项目: 国家自然科学基金项目 (61771193, 62101189, 62376094, U22A20102); 浙江省教育厅科研项目 (Y202146028).

通信作者: 张雄涛. E-mail: 02032@zjhu.edu.cn.

基于脑电 (electroencephalogram, EEG) 信号的运动想象任务^[1-2]通过解码大脑运动想象过程中的放电信号, 实现对外部设备的直接控制, 是脑

机接口领域^[3-5]的一个重要研究方向。然而,受肌电、情绪等诸多因素影响,被试者之间数据分布不一致^[6-8]。传统微调算法(vanilla fine-tuning)^[9]通过预训练、微调的二阶段策略实现跨领域知识迁移,在计算机视觉、自然语言处理等领域得到了广泛的应用,显著提高了下游任务的预测精度。由于神经网络不具备记忆功能,拟合下游任务数据分布的过程中常伴随着对上游任务的遗忘^[10]。在样本稀疏的场景下,灾难性遗忘问题常常导致模型过拟合下游任务。运动想象 EEG 分类任务数据采集和标注成本昂贵^[11],单一被试仅有少量标注样本。传统微调算法直接应用于 EEG 运动想象任务,往往由于灾难性遗忘问题^[12-14]导致模型在目标任务上过拟合^[14]。

传统微调算法^[9]预训练所得到的模型参数保留了源域数据潜在信息。因此,主流的方法通过抑制微调阶段参数更新^[15-19]缓解传统微调算法在微调过程中的灾难性遗忘问题。Top-K^[15]仅仅微调最后 K 层网络参数,减少了可调参数数量,固定参数保留了源域数据潜在信息。权重衰减(weight decay)^[16]算法添加模型参数与预训练参数之差的 L2 损失函数作为正则化项,抑制微调阶段参数更新。RecAdam^[17]基于权重衰减算法^[16],针对 L2 损失构建了随循环次数不断衰减的权重项,提高了模型对下游任务的拟合能力。Child-Tuning^[19]借鉴了随机失活^[20]的思想,在微调过程中仅更新深度神经网络部分参数。抑制参数更新的方式,虽然能够缓解传统微调算法在目标域样本稀疏情况下的灾难性遗忘问题,却牺牲了模型对下游任务的拟合能力。

邻域风险最小化策略^[21]基于标注数据邻域值训练模型参数,可以在不牺牲模型拟合能力的基础上,提高模型的泛化性能。Mixup^[22]凸组合域内数据,构造标注数据邻域分布,缓解了经验风险最小化策略所引发的过拟合问题。为了缓解传统微调算法在目标域样本稀疏场景下的灾难性遗忘问题,同时提高模型对下游任务的拟合能力,本文提出基于域间 Mixup 微调策略的跨被试运动想象脑电信号分类算法 Mix-Tuning,在微调阶段引入源域数据潜在信息。

Mix-Tuning 沿用了传统微调算法预训练、微调的二阶段策略。预训练阶段, Mix-Tuning 使用源域数据初始化模型参数,挖掘源域数据潜在信息;微调阶段, Mix-Tuning 通过域间 Mixup 得到合成样本训练模型参数。合成样本引入了源域数据潜在信息,缓解了微调过程中的灾难性遗忘问

题,提高了模型的泛化性能。Mix-Tuning 被进一步应用于运动想象脑电信号分类任务,实现了跨被试正向知识迁移。

本文的贡献点主要有:

1)提出了域间 Mixup 微调策略,缓解了传统微调算法在目标域样本稀疏场景下的灾难性遗忘问题。

2)域间 Mixup 微调策略通过数据合成对模型施加正则化约束,为缓解微调阶段灾难性遗忘问题贡献了一种新的思路。

3)将 Mix-Tuning 方法应用于运动想象脑电信号分类任务,在 BMI 数据集达到了 85.50% 的平均分类准确率。

1 运动想象 EEG 训练方式

运动想象脑机接口^[3-4]近年来在运动康复、游戏等领域取得了诸多实质性的进展,受到了学术界和工业界越来越多的关注。然而,脑电信号采集和标注过程繁琐、代价昂贵^[11],单一被试往往只有少量样本。此外,脑电信号作为生物信号的一种,不同被试数据分布不一致。运动想象脑电信号数据稀疏、跨被试数据分布不一致的特点,促使脑电信号的深度学习研究主要采用被试-依赖^[23-26]、被试-独立^[26-28]和被试-自适应^[29]3种训练方式。

被试-依赖^[23-26]使用目标被试标注数据训练模型参数,其经验风险公式为

$$R_{\text{spe}} = E_{(x_i, y_i) \in D_i} \ell(\mathcal{F}(x_i; \theta), y_i)$$

式中: R_{spe} 为被试-依赖训练方式经验风险; $\mathcal{F}(\cdot; \theta)$ 为预测模型, θ 为模型的参数; ℓ 为风险函数; D_i 为目标被试标注数据; (x_i, y_i) 为 D_i 中第 i 个样本的数据和标签; E 为数学期望。

由于单一目标被试数据非常稀少,为了降低过拟合风险,采用被试-依赖训练方式的模型往往容量较小。文献[26]模仿滤波器组共空间模式(filter bank common spatial pattern, FBCSP)^[2]的特征提取过程,设计了模型 Shallow ConvNet。Shallow ConvNet 使用时域卷积和空域卷积提取特征,并使用 1 层全连接层完成分类任务,其在多个数据集上取得了超越传统 EEG 信号处理方法的精度。文献[23]引入深度可分离卷积(depthwise separable convolution)^[30],提出了紧凑型的卷积神经网络 EEGNet。EEGNet 在保证模型表达能力的基础上,进一步减少了模型参数。然而,受限于单个被试数据规模,被试-依赖训练方式容易导致模型过拟合^[23-26]。

被试-独立^[26-28]忽略不同被试者数据分布的差异,组合多个被试者数据得到训练数据,其经验风险表示为

$$R_{\text{ind}} = E_{(x_i, y_i) \in (D_s \cup D_t)} \ell(\mathcal{F}(x_i; \theta), y_i)$$

式中: R_{ind} 为被试-独立训练方式经验风险, D_s 为非目标被试标注数据, (x_i, y_i) 为 $(D_s \cup D_t)$ 中第 i 个样本的输入和标签。

为了匹配训练数据规模,采用被试-独立训练方式的模型通常有着更大的容量。文献[26]提出一个层次化特征提取模型,称作 Deep ConvNet。Deep ConvNet 使用 4 个卷积-池化块实现特征提取和降维,并使用 1 层全连接层完成分类任务。文献[27]提出了模型 TIDNet,引入卷积网络中的残差模块(residual block)^[31]和密集模块(dense block)^[32],进一步加深了模型的深度。然而,被试-独立训练方式假设不同被试 EEG 数据服从独立同分布假设,不符合 EEG 真实数据分布^[27-28]。

被试-自适应^[29]通过预训练、微调的二阶段策略,挖掘源域数据潜在信息,实现跨被试知识迁移,有望解决被试-依赖和被试-独立训练方式所面临的困境。被试-自适应训练方式预训练和微调阶段的经验风险公式为

$$\begin{aligned} \text{预训练: } R_{\text{pre}} &= E_{(x_i, y_i) \in D_s} \ell((x_i; \theta), y_i) \\ \text{微调: } R_{\text{tuning}} &= E_{(x_i, y_i) \in D_t} \ell((x_i; \theta), y_i) \end{aligned} \quad (1)$$

式中: R_{pre} 为预训练阶段经验风险, R_{tuning} 为微调阶段经验风险。

文献[29]通过调低学习率和早停策略成功将传统微调算法应用于 EEG 领域,实现了跨被试正向知识迁移。然而,EEG 标注数据获取成本昂贵,早停策略需要划分验证集^[33]。这将进一步减少标注样本的数量,增加模型的过拟合风险。为了缓解传统微调算法在目标域样本稀疏时出现的过拟

合问题,本文提出了基于域间 Mixup 微调策略的跨被试运动想象脑电信号分类算法 Mix-Tuning。微调阶段, Mix-Tuning 通过域间 Mixup,生成域间插值数据微调模型参数。域间 Mixup 引入了源域数据潜在信息,缓解了微调阶段的过拟合问题。

2 域间 Mixup 微调算法

在 EEG 运动想象跨被试知识迁移的场景中,数据分别来自源域 $D_s = \{(x_i^s, y_i^s)\}_i^{n_s}$ 和目标域 $D_t = \{(x_i^t, y_i^t)\}_i^{n_t}$, n_s 和 n_t 分别表示源域和目标域样本数量。目标域数据又可分为标注数据 $D_l = \{(x_i^t, y_i^t)\}_i^{n_l}$ 和未标注数据 $D_u = \{(x_i^t)\}_i^{n_u}$, n_l 和 n_u 分别表示目标域标注样本和未标注样本数量。源域数据和目标域数据共享输入空间 $\mathcal{X} \in \mathbf{R}^{C \times T}$ 和标签空间 $\mathcal{Y} \in \mathbf{R}^z$ 。源域数据和目标域数据分别服从数据分布 P 和 Q , 并且 $P \neq Q$ 。源域含有大量标注数据,而目标域仅有少量样本, $n_s \gg n_t = n_l + n_u$ 。

2.1 模型介绍

本文选择 Deep ConvNet^[26]作为主干网络。Deep ConvNet 由 4 个卷积-池化块和 1 层全连接层构成。第 1 个卷积-池化块使用时域卷积和空域卷积分别提取 EEG 原始数据时序特征和空间特征,并通过最大值池化对特征进行降采样。后 3 个卷积-池化块由时域卷积层和最大值池化层构成,进一步提取任务相关特征。EEG 数据经过 4 个卷积-池化块完成特征提取后,送入最后 1 层全连接层。全连接层神经元数量与类别数量相等,其输出经过 Softmax 函数,转化为对每个类别的预测概率。Deep ConvNet 网络结构如图 1 所示,其中, λ 为域间插值系数, $\lambda \sim \text{Beta}(\alpha, \alpha)$, α 为贝塔分布参数, L_{CE} 表示交叉熵损失函数。

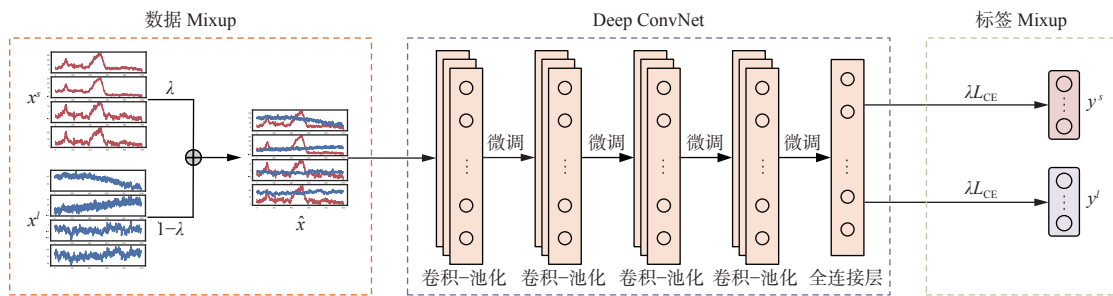


图 1 Mix-Tuning 微调过程

Fig. 1 Mix-Tuning fine-tuning process

2.2 训练过程

EEG 信号跨被试数据分布不一致。Mix-Tuning 通过预训练、微调的二阶段策略,实现跨被试

知识迁移。预训练阶段, Mix-Tuning 使用源域数据初始化模型参数,挖掘源域数据潜在信息;微调阶段,源域数据和目标域数据通过域间 Mixup

生成域间插值数据微调模型参数。其中,数据 Mixup 通过按维度加权相加得到,标签 Mixup 通过分类损失函数的加权和替代。Mix-Tuning 微调阶段训练过程如图 1 所示。Mix-Tuning 预训练-微调训练方式的训练步骤如下所示。

1) 预训练。源域数据与目标域数据共享输入空间 $\mathcal{X} \in \mathbf{R}^{C \times T}$ 和标签空间 $\mathcal{Y} \in \mathbf{R}^z$, 且数据之间有着接近的数据分布 $P(x_i^s, y_i^s) \approx P(x_i^l, y_i^l) = P(x_i^u, y_i^u)$ 。卷积神经网络浅层部分提取通用的特征, 网络浅层部分参数可迁移至相关任务。为了充分利用源域数据潜在信息, 提高模型的泛化性能, Mix-Tuning 使用源域数据预训练模型参数。预训练阶段损失函数表示为

$$L_{\text{pre}} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \sum_{c=1}^z \mathbb{I}_{(y_i^s=c)} \log(\mathcal{F}(x_i^s; \theta))$$

式中: \mathbb{I} 为指示函数, 当 $y_i^s = c$ 时, \mathbb{I} 取值为 1, 否则取值为 0; z 为类别数量; n_s 为源域样本数量, (x_i^s, y_i^s) 为源域的第 i 个样本。

2) 微调。EEG 标注数据获取成本昂贵, 单一被试无法提供充足的样本微调模型参数。经验风险最小化倾向于记忆目标域标注数据, 所训练的模型泛化性能较差。Mix-Tuning 通过 Mixup 凸组合域间数据, 构造了源域数据和目标域标注数据的邻域分布 $P_v(\hat{x}, \hat{y})$ 。

$$P_v(\hat{x}, \hat{y}) = \frac{1}{n_l n_s} \sum_{i=1}^{n_l} \sum_{j=1}^{n_s} \mathbb{H}_i[\delta(\hat{x}, \hat{y})]$$

$$\hat{x} = \lambda x_i^l + (1 - \lambda) x_j^s$$

$$\hat{y} = \lambda y_i^l + (1 - \lambda) y_j^s$$

式中: δ 为狄拉克函数, $\mathbb{H}_i[\cdot]$ 为取全体 λ 取值的集合。

Mix-Tuning 采样邻域分布构建了一个新的数据集 $\mathbf{D}_v := \{(\hat{x}_k, \hat{y}_k)\}_k^m$, m 表示样本数量, (\hat{x}_k, \hat{y}_k) 表示第 k 个合成样本, $\hat{y}_k = y_k^s + y_k^l$ 。微调阶段, 每次迭代独立采样数据集 \mathbf{D}_v 得到训练数据。确定交叉熵需要标签为独热码方可计算, Mix-Tuning 微调阶段分别计算预测标签 $\mathcal{F}(\hat{x}_k; \theta)$ 与源域标签 y_k^s 和目标域标签 y_k^l 的交叉熵函数, 二者加权和构成微调阶段的损失函数为

$$L_{\text{tuning}} = -\frac{1}{m} \sum_{k=1}^m \lambda \sum_{c=1}^z \mathbb{I}_{(y_k^s=c)} \log(\mathcal{F}(\hat{x}_k; \theta)) -$$

$$\frac{1}{m} \sum_{k=1}^m (1 - \lambda) \sum_{c=1}^z \mathbb{I}_{(y_k^l=c)} \log(\mathcal{F}(\hat{x}_k; \theta))$$

式中: Mix-Tuning 通过域间 Mixup 微调策略, 实现跨被试正向知识迁移, 伪代码如算法 1 所示。

算法 1 Mix-Tuning

输入 源域数据 $\mathbf{D}_s = \{(x_i^s, y_i^s)\}_i^{n_s}$, 目标域标注数

据 $\mathbf{D}_l = \{(x_i^l, y_i^l)\}_i^{n_l}$, Deep ConvNet 参数 θ , 超参数 α , 预训练迭代次数 T_1 , 微调迭代次数 T_2 ;

输出 Deep ConvNet 最优参数 $\hat{\theta}$;

- 1) 随机初始化 θ
- 2) WHILE $t < T_1$ DO
- 3) $t \leftarrow t + 1$
- 4) $(x_i^s, y_i^s) \leftarrow \text{RANDOM}(\mathbf{D}_s)$
- 5) $\theta_{t+1} \leftarrow \theta_t - \nabla_{\theta} \mathcal{L}(\mathcal{F}(x_i^s, \theta), y_i^s)$
- 6) ENDWHILE
- 7) WHILE $t < T_2$ DO
- 8) $t \leftarrow t + 1$
- 9) $(x_i^s, y_i^s) \leftarrow \text{RANDOM}(\mathbf{D}_s)$
- 10) $(x_i^l, y_i^l) \leftarrow \text{RANDOM}(\mathbf{D}_l)$
- 11) $\lambda \sim \text{Beta}(\alpha, \alpha)$
- 12) $\hat{x} \leftarrow \lambda x_i^l + (1 - \lambda) x_j^s$
- 13) $\theta_{t+1} \leftarrow \theta_t - \lambda \nabla_{\theta} \mathcal{L}(\mathcal{F}(\hat{x}, \theta), y_i^l) -$
 $(1 - \lambda) \nabla_{\theta} \mathcal{L}(\mathcal{F}(\hat{x}, \theta), y_i^s)$
- 14) ENDWHILE

2.3 邻域风险最小化

目标域真实数据分布 $P(x, y)$ 通常是未知的。传统微调算法在微调阶段基于目标域标注数据, 通过经验风险最小化^[12] 搜索最优假设。微调阶段经验风险 $R_{\text{tuning}}(\theta)$ 如式 (1) 所示。经验风险最小化策略^[12] 在样本稀疏场景下倾向于简单记住训练数据, 导致模型泛化性能不足^[21,33]。域间 Mixup 数据增强策略通过凸组合域间数据得到合成样本微调模型参数, 将经验风险最小化转变为邻域风险最小化, 从而降低了模型微调过程中的过拟合风险。邻域风险 $R_v(\theta)$ 为

$$R_v(\theta) = \frac{1}{m} \sum_{k=1}^m \ell(\mathcal{F}(\hat{x}_k; \theta), \hat{y}_k)$$

邻域风险 $R_v(\theta)$ 可以进一步划分为目标域标注数据邻域风险 $R_l(\theta)$ 和源域数据邻域风险 $R_s(\theta)$:

$$R_v(\theta) = R_l(\theta) + R_s(\theta)$$

$$R_l(\theta) = \frac{1}{m} \sum_{k=1}^m \lambda \ell(\mathcal{F}(\hat{x}_k; \theta), y_k^l)$$

$$R_s(\theta) = \frac{1}{m} \sum_{k=1}^m (1 - \lambda) \ell(\mathcal{F}(\hat{x}_k; \theta), y_k^s)$$

$R_s(\theta)$ 、 $R_l(\theta)$ 分别促使模型拟合源域和目标域邻域数据分布。 $R_s(\theta)$ 可以看作 Mix-Tuning 微调阶段的正则化项, 促使模型偏好于有效拟合源域数据分布的参数。Mix-Tuning 通过域间 Mixup 生成域间插值数据微调模型参数, 在微调阶段引入源域数据邻域风险最小化策略, 缓解了传统微调算法在微调阶段的灾难性遗忘问题, 提高了模型的泛化性能。

3 实验结果与分析

3.1 数据集和数据预处理

本文采用韩国高丽大学大脑与认知工程系发布的 BMI 数据集^[11]的运动想象部分作为实验数据。BMI 数据集通过 62 个电极,以 1000 Hz 频率采样被试脑电信号。数据集采集自 54 个被试(S1~S54),执行左、右手 2 类运动想象任务。每个被试数据分为 2 个会话,每个会话包含 200 个类别平衡样本。目标被试第 1 个会话数据作为目标域训练集 $D_l = \{(x_i^l, y_i^l)\}_{i=1}^n$, 目标被试第 2 个会话数据作为目标域测试集 $D_u = \{(x_i^u)\}_{i=1}^n$, 非目标被试的其余 53 个被试所有数据组合构成源域数据 $D_s = \{(x_i^s, y_i^s)\}_{i=1}^n$ 。

本文所选择的对比模型包括深度学习模型^[23, 26-27]和传统脑电数据处理模型^[1-2]。对于深度学习模型,保留数据集所有电极采样的脑电信号,截取被试执行运动想象任务的 4 s 采集的脑电信号作为输入数据,并将其频率降采样为 250 Hz。因此,深度学习模型输入数据 $X \in \mathbf{R}^{C \times T}$, $C = 62$, $T = 1000$ 。传统的脑电模型对噪声比较敏感^[34],本文沿用数据集发布方^[11]所选择与运动想象任务相关的 20 个电极作为输入信号,并将数据采样频率降采样至 250 Hz。为了进一步提高数据的信噪比,本文选择 8~30 Hz 的带通滤波器对数据滤波。因此,传统的机器学习方法输入数据 $X \in \mathbf{R}^{C \times T}$, $C = 20$, $T = 1000$ 。

3.2 实验设置

本文算法 Mix-Tuning 基于 Pytorch 机器学习库实现。所有实验均在同一设备运行,处理器为 Intel(R) Xeon(R) Silver 4214,图形处理器为 Nvidia Tesla V100,操作系统为 CentOS Linux release 7.9.2009。本文中深度神经网络均采用小批次梯度下降算法训练,除样本数量实验外,批次大小(batch size)统一设置为 50。优化器选择 Adam^[35],相应的参数分别设置为 0.9 和 0.99。本文中所有采用被试-独立训练方式的模型,学习率和迭代次数统一设置为 0.01 和 200。本文中采用被试-自适应训练方式的算法学习率和迭代次数保持一致。预训练阶段,学习率和迭代次数分别被设置为 0.01 和 200;微调阶段,学习率和迭代次数分别被设置为 0.003 和 300。本文采用准确率和 F_1 值作为评价指标, F_1 值是精确率和召回率的调和平均。

3.3 对比实验

为了验证 Mix-Tuning 的有效性,本文选择采用被试-依赖、被试-独立和被试-自适应 3 种训

练方式的多个算法与本文算法进行比较。

1)被试-依赖:共空间模式(common spatial pattern, CSP)^[1]通过最大化 2 类信号的方差值,得到多通道脑电信号在空间中的最优投影。FBCSP^[2]首先划分脑电数据为多个不同的频带,然后针对每个频带的数据使用共空间模式提取特征。共空间模式和滤波器组共空间模式完成 EEG 数据的特征提取,线性判别分析对所提取特征执行分类任务。

2)被试-独立:EEGNet^[23]、Shallow ConvNet^[26]、Deep ConvNet^[26]和 TIDNet^[27]采用被试-独立训练方式,作为 Mix-Tuning 的对比算法。EEGNet、Shallow ConvNet、Deep ConvNet 和 TIDNet 除失活概率^[20]外,均采用默认参数设置。EEGNet、Shallow ConvNet 和 Deep ConvNet 所有失活层^[20]的失活概率均设置为 0.5。TIDNet 包括 5 层失活层,其失活概率分别设置为 {0, 0, 0.4, 0.4, 0}。

3)被试-自适应:运动想象 EEG 领域采用被试-自适应训练方式的算法还比较稀少。为了验证 Mix-Tuning 方法的有效性,本文选用了部分已经在计算机视觉和自然语言处理领域取得成功的预训练-微调算法,迁移至运动想象 EEG 领域,作为 Mix-Tuning 的对比算法。

权重衰减算法^[16]:添加模型当前参数与预训练参数之差的 L2 范数作为正则化项,抑制微调阶段参数更新,从而缓解模型在微调过程中出现的灾难性遗忘问题,提高模型泛化性能。对比实验中,文中通过网格搜索超参数的最优值。

RecAdam^[17]:在权重衰减算法的基础上,借鉴模拟退火公式,构建随迭代次数逐渐衰减的 L2 正则化项,提高了模型对下游任务的拟合能力。对比实验中,文中通过网格搜索 RecAdam 超参数的最优值。

Child-Tuning^[19]:借鉴了随机失活^[20]方法,在微调过程中仅更新部分参数,从而达到抑制参数更新的目的。Child-Tuning 设定一个概率值,每一次迭代依据概率值生成伯努利矩阵 M_i , M_i 维度与参数维度相匹配。对比实验中,文中在 {0.1, 0.3, 0.5, 0.7, 0.9} 中网格搜索概率值的最优值。

本次实验中所有的被试-自适应方法均采用 Deep ConvNet^[26]作为模型的主干网络。为了能够公平地比较不同算法的实验结果,除算法特定参数外,学习率、迭代次数等参数的设置均保持一致。被试-依赖、被试-独立、被试-自适应 3 种训练方式的模型所对应 BMI 数据集^[11]上 54 个被试的平均准确率和 F_1 值如表 1,括号中的内容表示标准差。

表1 不同的算法在BMI数据集(54被试)上的准确率和 F_1 值的平均值(标准差)Table 1 Accuracy and F_1 mean (standard deviation) of different algorithms on BMI dataset (54 subjects)

%

训练方式	算法	准确率	F_1 值
被试-依赖	CSP	64.67(15.84)	64.67(15.84)
	FBCSP	64.70(13.85)	64.90(13.98)
	Deep ConvNet	58.72(9.60)	54.89(16.51)
被试-独立	EEGNet	79.77(10.81)	79.35(11.20)
	Shallow ConvNet	81.55(10.90)	82.29(10.11)
	Deep ConvNet	84.01(10.92)	85.50(9.35)
	TIDNet	75.15(11.32)	73.56(12.74)
被试-自适应 (Deep ConvNet)	预训练	80.71(11.62)	81.11(12.03)
	传统微调算法	77.23(13.44)	77.29(14.79)
	权重衰减算法	78.66(12.67)	78.70(14.28)
	RecAdam	78.25(12.38)	78.10(14.16)
	Child-Tuning	79.06(12.56)	78.53(15.06)
	Mix-Tuning	85.50(10.49)	85.72(10.92)

注: 加黑代表本文模型的正确率, 下同。

被试-依赖实验中, 由于目标域数据训练集和测试集来自不同的会话, 数据分布之间存在一定的差异, 传统的脑电数据处理方法 CSP 和 FBCSP 表现较差。由于标注数据样本极其稀疏, Deep ConvNet 采用被试-依赖训练方式出现了严重的过拟合问题, 仅取得了 58.72% 的平均准确率。被试-独立实验中, Deep ConvNet 取得了最高的准确率 84.01%。EEGNet 和 Shallow ConvNet 网络规模较小, 拟合能力较弱, 二者预测准确率低于 Deep ConvNet。TIDNet 由于参数量过大, 出现了过拟合问题, 导致准确率偏低。被试-自适应实验采用 Deep ConvNet 网络结构, 预训练模型在 54 个被试测试数据上的平均准确率为 80.71%。本次实验中, 采用了 BMI 数据集作为数据集。其中, 源域数据整合了 53 个非目标被试 2 个会话的数据, 样本数量达到 21 200 个。目标域仅选取单一被试第 1 个会话的 200 个标注样本。受到数据规模的影响, 预训练模型展示了较强的泛化性能, 同时, 模型在微调阶段将面临较高的过拟合风险。传统微调算法未采用任何正则化策略, 在目标域微调过程中出现了严重的过拟合问题, 平均准确率降为 77.23%。相较于预训练模型, 传统微调算法微调之后平均准确率下降了 3.48%。权重衰减算法、RecAdam 和 Child-Tuning 通过抑制微调过程中的参数更新, 缓解了模型微调过程中的灾难性遗忘问题。相较于传统微调算法, 其准确率均有小幅度提高。然而, 权重衰减算法、RecAdam 和 Child-Tuning 在微调后平均准确率仍低于预训练模型准确率。Mix-Tuning 通过域间 Mixup 生成域间插值数据微调模型参数, 其平均准确率达到

了 85.50%。相较于预训练模型, Mix-Tuning 平均准确率提高了 4.79%, 实现了目标域样本稀疏场景下跨被试正向知识迁移。

不同的被试-自适应算法在微调过程中的训练集交叉熵损失函数(权重被置为 1)和测试集准确率随迭代次数变化情况如图 2 所示, 算法{传统微调算法, 权重衰减算法, RecAdam, Child-Tuning, Mix-Tuning}分别简记为{V-F, W-D, Rec, C-T, M-T}。图 2 中, 曲线均采用指数移动平均平滑化处理, 平滑因子设置为 0.8。交叉熵损失函数均值随迭代次数变化情况如图 2(a)所示, 传统微调算法方法未采用正则化方法, 其损失函数曲线随迭代次数快速下降, 直至逼近于 0。权重衰减算法、RecAdam 和 Child-Tuning 添加了正则化项, 限制了微调阶段的参数更新, 其损失函数曲线下降至一定程度后便不再继续下降。Mix-Tuning 采样源域和目标域邻域分布, 在微调阶段引入源域数据邻域风险最小化, 其损失函数曲线下降幅度最小, 缓解了传统微调算法在样本稀疏场景下的过拟合问题。目标域测试数据平均准确率随迭代次数变化曲线如图 2(b)所示。预训练模型在 BMI 数据集上的平均准确率为 80.71%, 在图 2 中以虚线表示, 作为被试-自适应方法的基准模型。随着迭代次数的增加, 传统微调算法、权重衰减算法、RecAdam 和 Child-Tuning 的准确率均经历了短暂的上升, 随后由于过拟合开始下降。权重衰减算法、RecAdam 和 Child-Tuning 通过抑制参数更新的策略, 缓解了微调阶段的过拟合问题, 准确率下降幅度低于传统微调算法。然而, 由于目标被试仅有 200 个标注样本, 在标注样本如此稀

疏的场景下微调模型参数有很高的过拟合风险。权重衰减算法、RecAdam 和 Child-Tuning 的准确率仍低于预训练模型,微调过程反而导致模型预测精度下降。Mix-Tuning 的准确率随着迭代次数的增加逐渐升高,直至趋于稳定。Mix-Tuning 很好地缓解了模型在目标域样本稀疏场景下的过拟合问题,提高了模型的泛化性能。

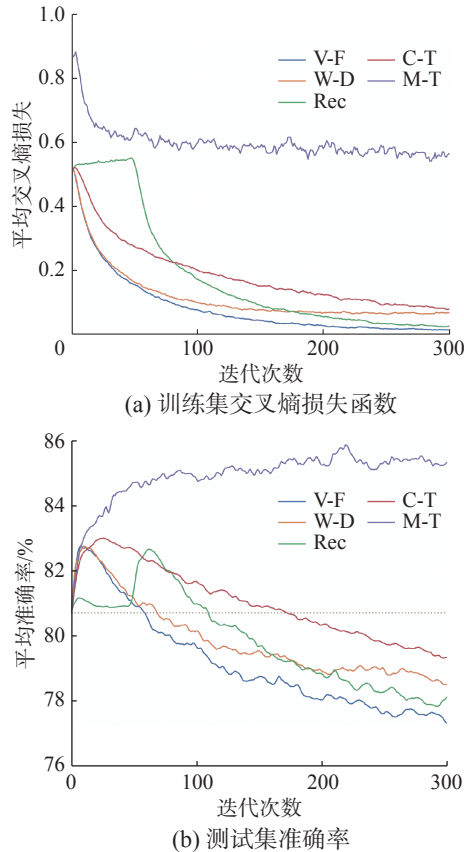


图2 被试-自适应算法损失函数和准确率变化曲线

Fig. 2 Subject-adaptive algorithm loss function and accuracy curves

3.4 样本数量实验

EEG 标注数据的成本昂贵,某些现实场景中单一被试无法提供 200 个标注数据。为了观察本文算法 Mix-Tuning 在目标被试标注样本极端稀疏时的表现,本文组织了样本数量实验,从目标域标注数据 D_t 随机选取 {200, 128, 64, 32, 16, 8} 个

类别平衡样本进行实验。为了保证模型在目标域微调过程中参数更新次数一致,此次实验中批次大小被设置为样本数的四分之一。即当样本分别为 {200, 128, 64, 32, 16, 8} 时,批次大小分别设置为 {50, 32, 16, 8, 4, 2}。不同的被试-自适应算法在不同样本数量下的平均准确率绘制折线如图 3 所示。不同样本数量下被试-自适应算法的平均准确率如表 2 所示。

如图 3 所示,随着样本数量的减少,传统微调算法在目标域测试集上的准确率极速衰减。权重衰减算法、RecAdam 和 Child-Tuning 抑制微调过程中参数更新,缓解了微调过程中的过拟合问题。随着样本数量的减少,权重衰减算法和 Child-Tuning 下降幅度明显低于传统微调算法。然而,当样本数取值为 {200, 128, 64, 32, 16, 8} 时,权重衰减算法、RecAdam 和 Child-Tuning 微调后的准确率普遍低于预训练模型准确率,微调过程反而导致预测准确率下降。随着样本数量的减少,模型过拟合风险逐渐增加, Mix-Tuning 平均准确率也随之下降。当样本数量仅为 16 时, Mix-Tuning 仍取得了 80.91% 的平均准确率,略高于预训练模型 80.71% 的准确率,实现了跨被试正向知识迁移。样本数量实验表明,本文算法 Mix-Tuning 能够应对目标域标注样本极其稀疏的场景,实现跨被试正向知识迁移。

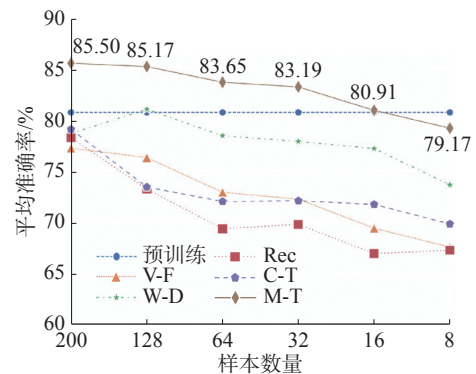


图3 不同样本数量下被试-自适应算法准确率对比
Fig. 3 Comparison of the accuracy of subject-adaptive algorithm under different sample sizes

表2 被试-自适应算法在不同样本数量下准确率和标准差对比

Table 2 Comparison of accuracy and standard deviation of subject-adaptive algorithms with different sample sizes %

样本数量	传统微调算法	权重衰减算法	RecAdam	Child-Tuning	Mix-Tuning
200	77.23 (13.44)	78.66 (12.67)	78.25 (12.38)	79.06 (12.56)	85.50 (10.49)
128	76.31 (13.13)	81.01 (11.62)	73.24 (13.29)	73.44 (12.68)	85.17 (9.89)
64	72.93 (14.08)	78.44 (11.78)	69.38 (12.95)	72.03 (13.31)	83.65 (13.32)
32	72.26 (13.68)	77.88 (11.78)	69.81 (13.33)	72.11 (12.70)	83.19 (10.88)
16	69.43 (12.53)	77.19 (13.65)	66.97 (13.09)	71.75 (13.78)	80.91 (11.36)
8	67.60 (14.54)	73.66 (13.51)	67.29 (12.51)	69.85 (13.71)	79.17 (12.29)

3.5 自适应网络层实验

为了观察本文算法 Mix-Tuning 应用于具有不同拟合能力的模型时的表现, 本文设计了网络自适应层的实验。网络自适应实验通过固定 Deep ConvNet 部分参数, 调节 Deep ConvNet 对下游任务的拟合能力。Deep ConvNet 由 4 个卷积-池化块和 1 层全连接层组成, 其网络参数 θ 由卷积-池化块参数 $\{\theta_1, \theta_2, \theta_3, \theta_4\}$ 和全连接层参数 $\{\theta_c\}$ 构成, $\theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_c\}$ 。微调阶段, 本文固定 Deep ConvNet 预训练模型的前 h 个卷积-池化块参数, 得到具有不同拟合能力的模型 Frozen- h , $h \in \{0, 1, 2, 3, 4\}$ 。 h 的取值决定了微调阶段模型可调参数数量, 影响模型对下游任务的拟合能力。本文统计了当输入空间 $\mathcal{X} \in \mathbf{R}^{62 \times 100}$ 和标签空间 $\mathcal{Y} \in \mathbf{R}^2$ 时, Frozen- h 可调参数数量如表 3 所示。

当 $h = 4$ 时, Deep ConvNet 前 4 个卷积-池化块参数被固定, 只剩下最后 1 层全连接层参数可调, 可调参

数量为 2802。随着 h 取值变小, 模型可调参数数量逐渐增加, 模型对下游任务拟合能力逐渐增加; 与此同时, 微调阶段过拟合风险也随之增加。本文设计了网络自适应层实验, 观察不同被试-自适应算法在 Frozen- h 下的表现, $h \in \{0, 1, 2, 3, 4\}$ 。不同被试-自适应算法在模型 Frozen- h 的准确率绘制箱线图如图 4 所示, 54 个被试平均准确率和标准差如表 4 所示。

表 3 Frozen- h 可调参数

Table 3 Frozen- h tunable parameters

模型	参数	参数数量
Frozen-4	θ_c	2802
Frozen-3	θ_4, θ_c	203202
Frozen-2	$\theta_3, \theta_4, \theta_c$	253402
Frozen-1	$\theta_2, \theta_3, \theta_4, \theta_c$	266002
Frozen-0	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_c$	305077

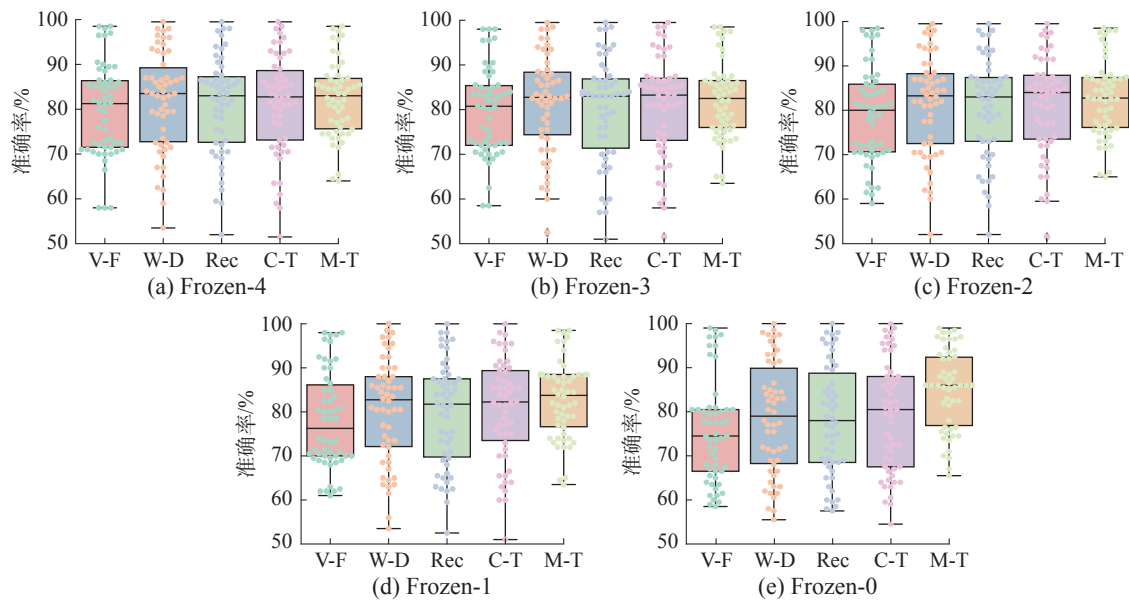


图 4 被试-自适应算法在 Frozen- h 下的表现, $h \in \{0, 1, 2, 3, 4\}$

Fig. 4 Performance of different subject-adaptive algorithms with different Frozen- h , $h \in \{0, 1, 2, 3, 4\}$

表 4 被试-自适应算法在 Frozen- h 下准确率和标准差对比

Table 4 Comparison of accuracy and standard deviation of subject-adaptive algorithms with different Frozen- h %

模型	传统微调算法	权重衰减算法	RecAdam	Child-Tuning	Mix-Tuning
Frozen-4	79.79(10.29)	81.88(11.02)	81.01(10.97)	81.21(11.29)	82.32(8.36)
Frozen-3	79.67(9.77)	81.49(10.97)	80.07(11.76)	80.81(11.39)	82.19(8.54)
Frozen-2	78.95(10.60)	81.36(11.08)	80.55(10.30)	81.11(11.01)	82.63(8.47)
Frozen-1	78.07(10.89)	80.47(11.68)	79.71(10.79)	80.58(11.48)	82.80(8.63)
Frozen-0	77.23(13.44)	78.66(12.67)	78.25(12.38)	79.06(12.56)	85.50(13.49)

如图 4 所示, 本文算法 Mix-tuning 在 Frozen- h 中正确率分布更为集中, 有着更高的准确率下

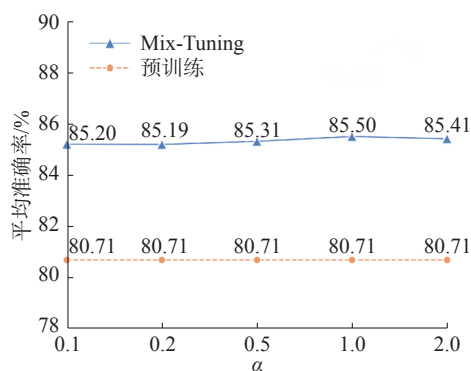
界。如表 4 所示, 随着 h 取值变小, Frozen- h 对下游任务拟合能力增强, 过拟合风险逐渐增加。传

统微调算法、权重衰减算法、RecAdam、Child-Tuning 由于缺少有效的正则化策略,随着模型可调参数增加,正确率却随之下降。从 Frozen-4~Frozen-0,传统微调算法、权重衰减算法和 Child-Tuning 平均正确率分别下降了 2.56%、3.22% 和 2.15%。与此同时, Mix-Tuning 平均正确率却随着模型拟合能力的增强而逐渐增加。Mix-Tuning 在 Frozen-4 的正确率仅为 82.32%, 在 Frozen-0 时却取得了最高正确率 85.50%。网络自适应层实验表明,本文算法 Mix-Tuning 在模型参数数量较大、过拟合风险较高的场景下,所训练模型仍保持了良好的泛化性能。

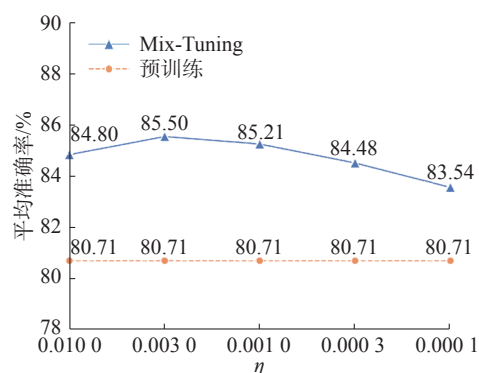
3.6 参数敏感性实验

Mix-Tuning 包括贝塔分布中 α 的取值和模型微调阶段学习率 η 共 2 个超参数。参数敏感性实验通过手动设置 α 和 η 不同取值,寻找 Mix-Tuning 在 BMI 数据集上的最优参数,同时观察 Mix-Tuning 方法对参数的敏感程度。参数 α 决定贝塔分布密度函数,进而影响 λ 分布情况。固定学习率为 0.003, 当 α 的取值分别为 {0.1, 0.2, 0.5, 1.0, 2.0} 时, Mix-Tuning 的平均准确率如图 5(a) 所示。为了提高模型泛化性能, Mix-Tuning 预训练阶段学习率被设置为 0.01。微调阶段,为了充分拟合目标域数据分布,学习率 η 需要设置为一个小于 0.01 的值。当贝塔分布的参数 α 被固定为 1, 微调阶段学习率 η 分别设置为 {0.01, 0.003, 0.001, 0.000 3, 0.000 1} 时, Mix-Tuning 的平均准确率如图 5(b) 所示。

图 5(a) 表明, 平均准确率随 α 取值变动幅度很小, Mix-Tuning 方法对贝塔分布的参数 α 不敏感。图 5(b) 显示, 当学习率 η 设置为 0.003 时, 模型得到最高的准确率 85.50%; 当学习率高于 0.003 时, Mix-Tuning 不能够充分拟合目标域数据; 学习率 η 低于 0.003 时, 模型参数更新步长过小, 模型参数不能得到充分训练。本文其他实验, 学习率 η 和 α 分别固定为 0.003 和 1。



(a) 参数 α 对准确率的影响



(b) 学习率 η 对准确率的影响

图 5 Mix-Tuning 参数敏感性实验

Fig. 5 Mix-Tuning parameter sensitivity experiment

4 结束语

本文提出了一种基于域间 Mixup 微调策略的跨被试运动想象脑电信号分类算法 Mix-Tuning。Mix-Tuning 基于预训练、微调的二阶段策略训练模型参数。域间 Mixup 引入源域数据潜在信息,缓解了传统微调算法在样本稀疏场景下的灾难性遗忘问题,提高了模型的泛化性能,实现了 EEG 运动想象跨被试正向知识迁移。在 BMI 数据集运动想象部分实验结果验证了本文所提算法的有效性。

Mix-Tuning 在预训练和微调阶段均需要源域数据的参与,与其他微调算法一样,其预测精度受源域数据规模影响。此外,域间 Mixup 通过凸组合得到域间插值数据,要求源域和目标域共享输入空间和标签空间。然而,采样自不同设备的 EEG 数据通常数据维度不一致。因此, Mix-Tuning 方法无法直接实现跨设备、跨任务知识迁移。这将是本文未来的一个研究方向。

参考文献:

- [1] RAMOSER H, MULLER-GERKING J, PFURTSCHELLER G. Optimal spatial filtering of single trial EEG during imagined hand movement[J]. *IEEE transactions on rehabilitation engineering*, 2000, 8(4): 441–446.
- [2] ANG Kaikeng, CHIN Z Y, WANG Chuanchu, et al. Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b[J]. *Frontiers in neuroscience*, 2012, 6: 39.
- [3] GAO Xiaorong, WANG Yijun, CHEN Xiaogang, et al. Interface, interaction, and intelligence in generalized brain-computer interfaces[J]. *Trends in cognitive sciences*, 2021, 25(8): 671–684.
- [4] REZEIKA A, BENDA M, STAWICKI P, et al. Brain-

- computer interface spellers: a review[J]. *Brain sciences*, 2018, 8(4): 57.
- [5] 李景聪, 潘伟健, 林镇远, 等. 采用多路图注意力网络的情绪脑电信号识别方法[J]. *智能系统学报*, 2022, 17(3): 531–539.
- LI Jingcong, PAN Weijian, LIN Zhenyuan, et al. Emotional EEG signal recognition method using multi-path graph attention network[J]. *CAAI transactions on intelligent systems*, 2022, 17(3): 531–539.
- [6] CRAIK A, HE Yongtian, CONTRERAS-VIDAL J L. Deep learning for electroencephalogram (EEG) classification tasks: a review[J]. *Journal of neural engineering*, 2019, 16(3): 031001.
- [7] ZHAO Liming, YAN Xu, LU Baoliang. Plug-and-play domain adaptation for cross-subject EEG-based emotion recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2021, 35(1): 863–870.
- [8] NI Ziyi, XU Jiaming, WU Yuwei, et al. Improving cross-state and cross-subject visual ERP-based BCI with temporal modeling and adversarial training[J]. *IEEE transactions on neural systems and rehabilitation engineering*, 2022, 30: 369–379.
- [9] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks?[C]//Proceedings of the Advances in Neural Information Processing Systems. Montréal: MIT Press, 2014, 27: 3320–3328.
- [10] WANG Yaqing, YAO Quanming, KWOK J T, et al. Generalizing from a few examples: a survey on few-shot learning[J]. *ACM computing surveys*, 53(3): 1–34.
- [11] LEE M H, KWON O Y, KIM Y J, et al. EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy[J]. *GigaScience*, 2019, 8(5): giz002.
- [12] VAPNIK V N. An overview of statistical learning theory[J]. *IEEE transactions on neural networks*, 1999, 10(5): 988–999.
- [13] PERKONIGG M, HOFMANNINGER J, HEROLD C J, et al. Dynamic memory to alleviate catastrophic forgetting in continual learning with medical imaging[J]. *Nature communications*, 2021, 12: 5678.
- [14] ZHANG Baosheng, GUO Yuchen, LI Yipeng, et al. Memory recall: a simple neural network training framework against catastrophic forgetting[J]. *IEEE transactions on neural networks and learning systems*, 2022, 33(5): 2010–2022.
- [15] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-efficient transfer learning for NLP[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach: IMLS, 2019, 97: 2790–2799.
- [16] LI Xuhong, GRANDVALET Y, DAVOINE F. Explicit inductive bias for transfer learning with convolutional networks[C]//Proceedings of the 35th International Conference on Machine Learning. Stockholm: IMLS, 2018, 80: 2825–2834.
- [17] CHEN Sanyuan, HOU Yutai, CUI Yiming, et al. Recall and learn: fine-tuning deep pretrained language models with less forgetting[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2020: 7870–7881.
- [18] LEE C, CHO K, KANG Wanmo. Mixout: effective regularization to finetune large-scale pretrained language models[C]//Proceeding of the 8th International Conference on Learning Representations. Addis Ababa: ICLR, 2020: 1–17.
- [19] XU Runxin, LUO Fuli, ZHANG Zhiyuan, et al. Raise a child in large language model: towards effective and generalizable fine-tuning[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 9514–9528.
- [20] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. *Journal of machine learning research*, 2014, 15: 1929–1958.
- [21] CHAPELLE O, WESTON J, BOTTOU L, et al. Vicinal risk minimization[C]//Proceedings of the Advances in Neural Information Processing Systems. Denver: MIT Press, 2000, 13: 1–7.
- [22] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mixup: beyond empirical risk minimization[C]//Proceedings of the International Conference on Learning Representations. Vancouver: ICLR, 2018: 1–13.
- [23] 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.
- [24] LI Yang, GUO Lianghui, LIU Yu, et al. A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding[J]. *IEEE transactions on neural systems and rehabilitation engineering*, 2021, 29: 1534–1545.
- [25] CAI Siqi, SU Enze, XIE Longhan, et al. EEG-based auditory attention detection via frequency and channel neural attention[J]. *IEEE transactions on human-machine systems*, 2021, 52(2): 256–266.

- [26] 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.
- [27] KOSTAS D, RUDZICZ F. Thinker invariance: enabling deep neural networks for BCI across more people[J]. *Journal of neural engineering*, 2020, 17(5): 056008.
- [28] KWON O Y, LEE M H, GUAN Cuntai, et al. Subject-independent brain-computer interfaces based on deep convolutional neural networks[J]. *IEEE transactions on neural networks and learning systems*, 2019, 31(10): 3839–3852.
- [29] ZHANG Kaishuo, ROBINSON N, LEE S W, et al. Adaptive transfer learning for EEG motor imagery classification with deep convolutional neural network[J]. *Neural networks*, 2021, 136: 1–10.
- [30] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818–2826.
- [31] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [32] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.
- [33] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. Cambridge: MIT press, 2016.
- [34] 章杭奎, 刘栋军, 孔万增. 面向跨被试 RSVP 的多特征

低维子空间嵌入的 ERP 检测 [J]. *智能系统学报*, 2022, 17(5): 1054–1061.

ZHANG Hangkui, LIU Dongjun, KONG Wanzeng. ERP detection of multi-feature embedding in the low-dimensional subspace for cross-subject RSVP[J]. *CAAI transactions on intelligent systems*, 2022, 17(5): 1054–1061.

- [35] KINGMA D P, BA J L. Adam: a method for stochastic optimization[C]//3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings. San Diego: ICLR, 2015: 1–15.

作者简介:



蒋云良,教授,博士生导师,博士,主要研究方向为智能信息处理、地理信息系统。享受国务院政府特殊津贴。获发明专利授权 26 项。发表学术论文 63 篇,出版学术著作 2 部, E-mail: jy12022@zjnu.cn。



周阳,硕士研究生。主要研究方向为深度学习、迁移学习、脑电信号处理。 E-mail: 3189269614@qq.com。



张雄涛,副教授,博士,主要研究方向为人工智能与模式识别、机器学习。承担国家和省部级科研课题 10 余项,获发明专利授权 7 项,发表学术论文 20 余篇。 E-mail: 02032@zjhu.edu.cn。