

基于集成学习与聚类联合标注的多模态个体情绪识别

柯善军, 聂成洋, 王钰苗, 何邦胜

(重庆理工大学车辆工程学院, 重庆 400050)

摘要: 针对通用情绪识别模型面对不同个体时的低识别精度问题, 提出一种基于集成学习与聚类联合标注的多模态个体情绪识别方法。该方法首先基于公共数据集训练通用情绪识别模型, 然后分析公共数据集数据与个体无标签数据的分布差异, 建立跨域模型来预测和标注个体数据的伪标签。同时, 对个体数据进行加权聚类并标注聚类标签, 利用聚类标签与伪标签进行联合标注, 筛选高置信度样本进一步训练通用模型, 得到个性化情绪识别模型。实验采集3名被试的3种情绪数据并使用该方法标注, 最后优化得到的个性化模型对3种情绪的平均识别精度达到80%以上, 相比原通用模型, 至少提升了35%。

关键词: 个体情绪识别; 领域自适应; 集成学习; 聚类; 联合标注

中图分类号: TP911.7, TP183

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202401

Multimodal individual emotion recognition with joint labeling based on integrated learning and clustering

KE Shanjun, NIE Chengyang, WANG Yumiao, HE Bangsheng

Chongqing University of Technology, Chongqing 400050, China

Abstract: To address the low recognition accuracy of generic emotion recognition models when faced with different individuals, a multimodal individual emotion recognition technique based on joint labelling with integrated learning and clustering was proposed. The method first trained a generic emotion recognition model based on a public dataset, then analysed the distributional differences between the data in the public dataset and the unlabelled data of individuals, and established a cross-domain model for predicting and labelling pseudo-labels of individual data. At the same time, the individual data were weighted clustered and labelled with cluster labels, and the cluster labels were used to jointly label with pseudo-labels, and high confidence samples were screened to further train the generic model to obtain a personalized emotion recognition model. Using this method to annotate these data with the experimentally collected data of 3 emotions from 3 subjects, the final optimized personalized model achieved an average recognition accuracy of more than 80% for the 3 emotions, which was at least a 35% improvement compared to the original generic model.

Key words: individual emotion recognition, domain adaptation, integrated learning, clustering, joint annotation

0 引言

情绪是人对客观事物的态度体验以及相应的行

为反应。情绪识别通过获取个体的生理或非生理信号对人的情绪状态进行自动辨别, 是当前研究热点之一。在人机交互领域, 情绪识别是实现机器智能

收稿日期: 2023-08-31; 修回日期: 2024-01-30

通信作者: 柯善军, shanjunkc@cqut.edu.cn

基金项目: 重庆市教委科学技术研究项目 (No.2020CJZ053)

Foundation Item: The Scientific and Technological Research Program of Chongqing Municipal Education Commission (No. 2020CJZ053)

化的关键一步，能够提高用户体验、优化交互方式、改善健康状况和提高安全性，为人机交互的发展和进步带来新的机遇和挑战^[1-3]。个性化情绪识别技术能够更好地适应跨个体情绪分类任务中个体的情绪反应，有效降低个体差异的影响，增强模型处理不同个体情绪反应变化的能力^[4]。因此，建立一种适用于不同个体的个性化情绪识别模型，对于提高情绪识别的准确性和鲁棒性至关重要。然而，个性化情绪识别仍然依赖于具有准确标签的个体数据，以进行监督学习^[5]。

对于不同个体而言，通过外部观察来准确识别和标注情绪较为困难，而且为了训练个性化模型而收集大量用户特定的标记数据是一项耗时且昂贵的活动，这在实际中是不可行的。因此，一些研究者通过伪标签法来充分利用未标记数据进行半监督学习^[6-7]。Li等^[8]提出了一种自适应调整置信度阈值的方法，用于生成高质量的伪标签，该方法可以根据不同类别的难度自适应地调整置信度边界，从而更好地适应面部表情识别任务。Zhu等^[9]使用噪声学生训练（noisy student training）方法进行语音情感识别，通过对伪标签数据的过滤，提高了情感识别的性能，但是由于情感主观性，在处理包含模糊情感的未标记数据时可能存在困难。文献[10-13]在一定程度上提高了其在指定数据集上伪标签的多样性

和准确性，但却忽略了个体差异性对情绪识别的影响。

为进一步提升标签置信度，本文提出一种基于集成学习与聚类联合标注的多模态个体情绪识别方法，其基本流程如图1所示。首先基于公共情感数据集训练多模态通用情绪识别模型，然后通过分析公共情绪数据集数据与采集到的个体无标签情绪数据之间的分布差异，建立了跨域情绪识别模型，用来预测、标注个体情绪数据的伪标签。同时，对个体无标签情绪数据进行加权聚类，标注聚类标签。利用聚类标签与伪标签对无标签数据进行联合标注，提升样本的伪标签置信度。最后，利用高置信度标签的个体情绪样本训练个性化情绪识别模型，并继续循环迭代优化。实验结果表明，利用该方法筛选出的个体情绪样本具有可靠的伪标签，能够显著提高模型对不同个体的情绪识别准确率。

1 构建通用情绪识别模型

受个体情绪表达模态差异的影响，单一模态情绪识别的精度存在较大的个体差异性。本文主要使用 MAHNOB-HCI 公共情感数据集^[14]中的脑电、心电生理数据和面部表情数据训练多模态通用情绪识别模型。目前的情绪识别多模态融合有两种主流方法：1) 特征层融合^[15-18]，即将经特征提取后全部

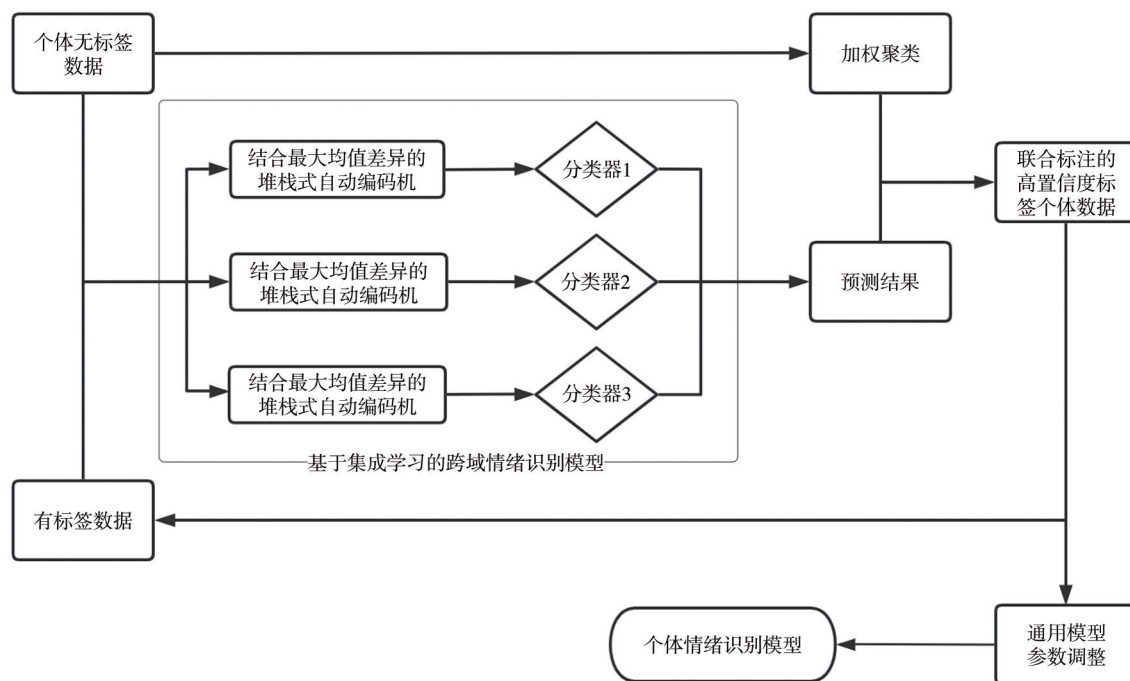


图1 基于集成学习与聚类联合标注的多模态个体情绪识别方法

模态特征数据级联为特征向量后再送入分类器；2) 决策层融合^[19-20]，依靠各个模态的可信度进行协调、联合决策。个体之间的情绪表达差异主要体现在模态层面，本文利用传统机器学习模型和深度学习模型进行各个模态的特征提取，然后在决策层进行模态融合，构建多模态情绪识别模型，其通用情绪识别模型结构如图2所示。

1.1 训练机器学习基分类器

支持向量机 (support vector machine, SVM) 是情感识别中最常用且有效的分类器^[21]，因此传统机器学习模型采用 SVM。首先，从公共数据集中提取脑电和心电模态的手工特征，其中脑电信号有8个通道，每个通道提取10个特征，共计80个特征；心电信号只有一个通道，从中提取10个特征。每组样本脑电和心电数据总共有90个手工特征。然后，将脑电、心电生理数据进行特征层融合，使用 SVM-RFE 算法^[22]训练，根据重要程度对融合特征进行排序，选取在模型性能和时间消耗方面表现最佳的前80个特征。最后采用支持向量机算法训练传统机器学习基分类器。

1.2 训练深度学习基分类器

运用神经网络来训练数据集，构造3种模态下的深度学习模型。本文采集的脑电信号有8个通道，因此使用 EEGNet^[23]来对脑电信号进行情绪识别。EEGNet采用普通2D卷积、深度卷积、深度可分离卷积神经网络组合的方式提取脑电的深度特征，最后通过 softmax 分类器对脑电情绪分类。心电信号是一维单通道信号，其特征满足局部连接和

权值共享的条件，所以采用一维卷积神经网络提取其深度特征，构建 ECGNet^[24]对其进行模型训练。在该卷积神经网络上，使用4层卷积以及两层全连接层，分类器同样为 softmax 分类器。对于面部表情模态，使用经典模型 VGG-16^[25]进行迁移学习。由于 MAHNOB-HCI 数据集的面部表情数据较少，存在过拟合风险。为了保证深度学习模型的良好性能，先用数据量较大的 fer2013 面部表情数据集^[26]构建 VGG 表情识别模型来提取其表情特征，再冻结模型卷积层保证其提取深度特征的性能，然后将全连接层中神经元个数由4096个降低为1024个，并在每层全连接层后面添加 Dropout 层，降低了过拟合风险和模型复杂度，提升了模型训练的效率。最后通过迁移学习，利用 MAHNOB-HCI 数据集将其训练为面部表情情绪识别模型。最终的面部表情情绪识别模型如图3所示。

1.3 决策层融合

本文通过在决策层对传统机器学习模型与深度学习模型构建的情绪分类器进行加权融合，即采用加权融合的方式对4个弱分类器并行联合，构建出基于公共数据集的通用情绪识别模型。最终模型预测结果是4个分类器的输出概率值乘以其分类器对应的权重后叠加的结果。权重大小根据4个弱分类器各自的情绪识别准确率进行权重分配，如4组弱分类器各自的情绪识别准确率分别为 p_1 、 p_2 、 p_3 、 p_4 ，分别计算 p_1 到 p_4 在这4种准确率中的重要程度占比 w_1 到 w_4 ，构建如下的权重矩阵 W ：

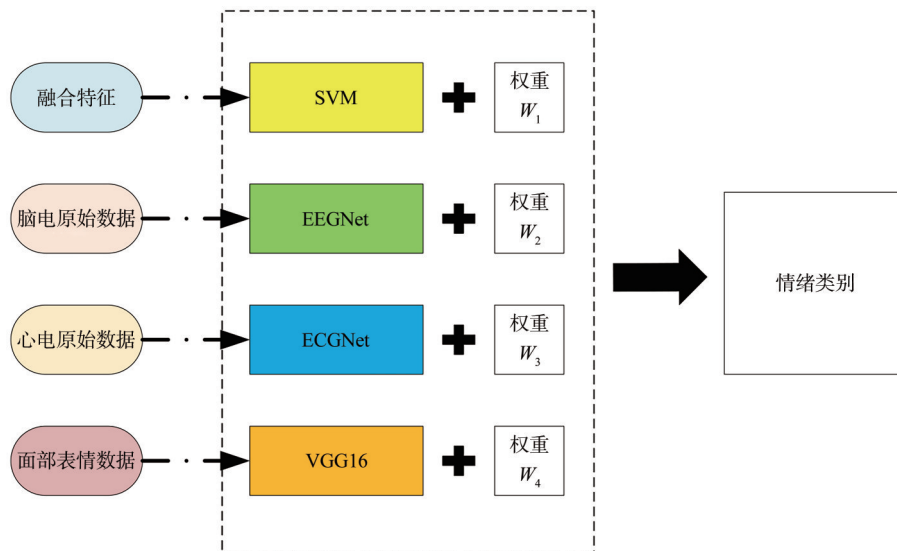


图2 通用情绪识别模型结构

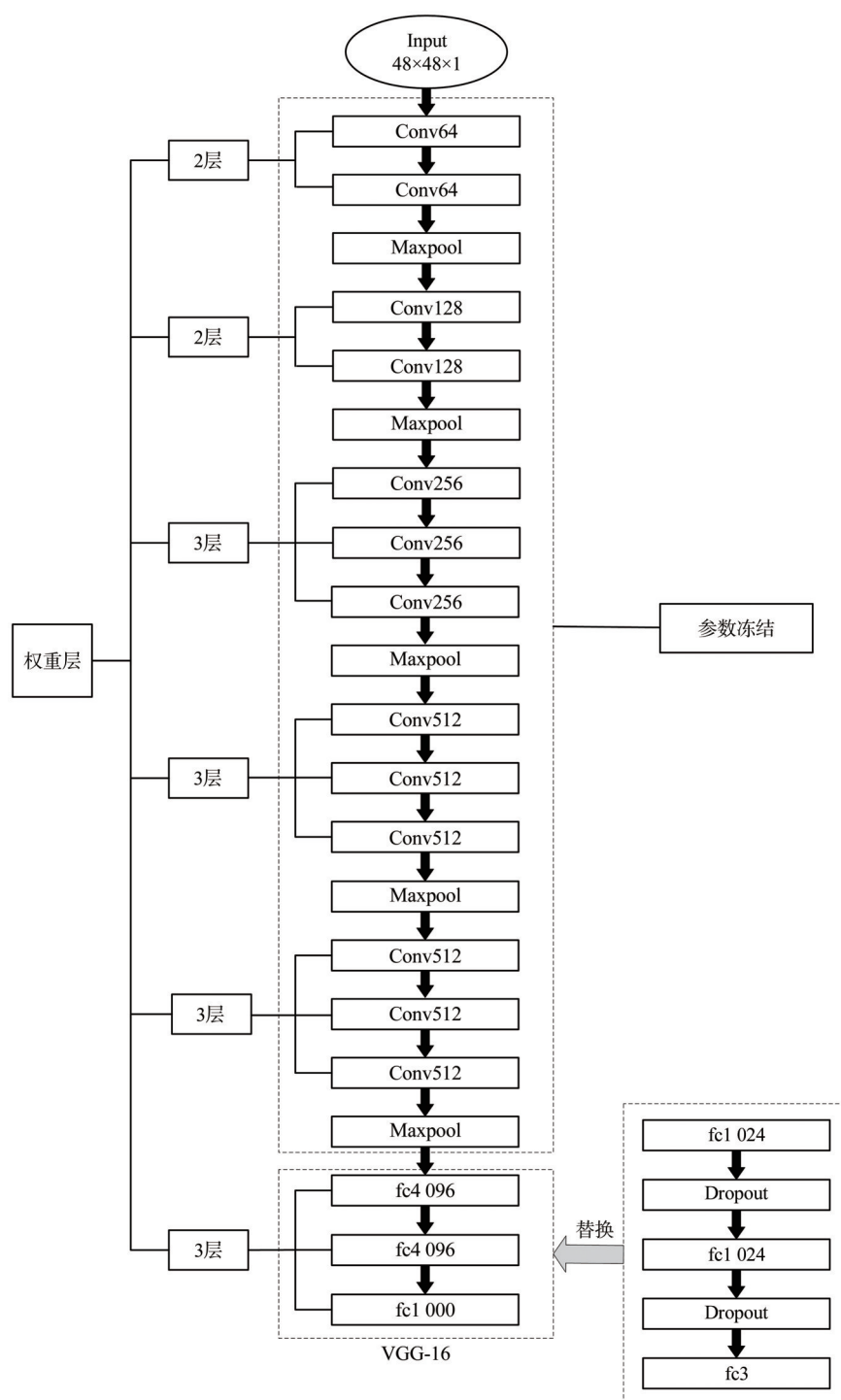


图3 面部表情情绪识别模型

$$W = \begin{pmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_4 \end{pmatrix} \quad (1)$$

根据构建的权重矩阵，对每个分类器的输出结果进行加权，得到加权后预测概率最大值对应的情绪类别。

2 集成学习与聚类联合标注方法

为了避免大量无标签数据的人工标注，并且有效提高伪标签的置信度，本文提出集成学习和聚类联合标注方法对个体情绪无标签数据进行自动标注。

2.1 领域自适应

当训练出基于公共数据集的情绪识别模型后,若将其直接用于预测数据集外的个体情绪数据标签,源域和目标域特征分布的不同可能导致模型性能下降。为解决此问题,使得模型在目标域上的性能提升到与源域相当或接近的水平,而避免在目标域进行大量标注,需使用领域自适应方法^[27-28]。特征自适应是领域自适应中常用的方法之一,通过对源域和目标域数据的特征进行调整或变换,将二者映射到能够降低它们之间分布差异的更高层特征空间,使得它们在特征空间中更加接近或对齐,以便模型能够更好地泛化到目标域^[29]。

由于脑电和心电信号特征融合后其特征维度相对较大,其中特征的非线性关系更为复杂,因此本文使用神经网络来有效拟合其非线性特征,学习源域与目标域特征之间分布不变的特征来减小两者数据集之间的分布差异。其中,常用的方法是使用自动编码器^[17]进行学习。自动编码器是一种无监督学习神经网络,它试图通过将输入数据进行编码和解码来学习数据的表示。Glorot^[30]在自动编码器的基础上提出了堆栈式自动编码器(stacked autoencoder, SAE), SAE领域自适应流程如图4所示。

SAE是由多个自动编码器按照层次结构堆叠在一起的神经网络模型,其通过增加网络层的深度来提高非线性变换,使用多层结构对数据特征进行分组来提高不变特征的权重。每个自动编码器的隐藏层输出作为下一个自动编码器的输入,形成了一个深层的编码和解码结构,使得SAE可以学习到在源域和目标域之间共享且更加抽象和深层的特征

表示。

然而,由于SAE在模型训练时不受标注信息限制,若源域和目标域的数据在特征分布上有较大差异,SAE在进行深度特征的抽取时不一定会根据原来类型进行抽取,反而有可能会通过学习数据自身的情况抽取不同的类别特征,导致源域和目标域之间的特征分布差异进一步加大。为了解决此问题,需要用相似度算法来度量抽取出来特征之间的分布差异。本文使用领域自适应中最常用的最大均值差异(maximum mean discrepancy, MMD)算法^[31]来限制源域与目标域样本抽取出来的特征差异。MMD通过最大化源域和目标域样本的特征表示之间的均值差异,使得它们的特征表示更加相似。结合最大均值差异的堆栈式自动编码器如图5所示。

通过核函数 ϕ 将不同概率分布的样本特征映射到再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS),用MMD来度量两种样本域之间的距离:

$$\text{MMD}(Z_s, Z_t) = \left\| \frac{1}{ns} \sum_{i=1}^{ns} \phi(Z_{si}) - \frac{1}{nt} \sum_{j=1}^{nt} \phi(Z_{tj}) \right\|_H \quad (2)$$

其中 Z_{si} 表示 ns 个源域样本中第 i 个样本在堆栈式自动编码器最后一层编码层的输出结果, Z_{tj} 表示 nt 个目标域样本中第 j 个样本在堆栈式自动编码器最后一层编码层的输出结果。

加入MMD后的整个堆栈式自动编码器在训练的过程中需要同时满足两个条件。第一个条件是最小化解码层的重构输出与真实输入之间的差距,鼓

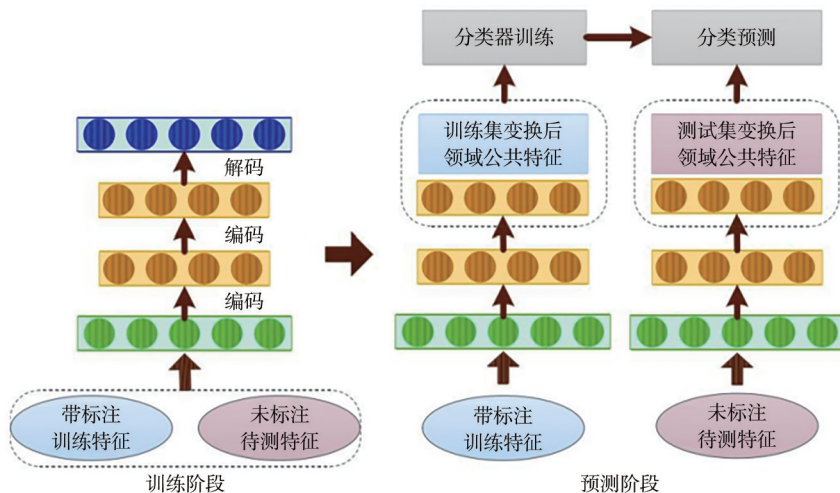


图4 SAE领域自适应流程

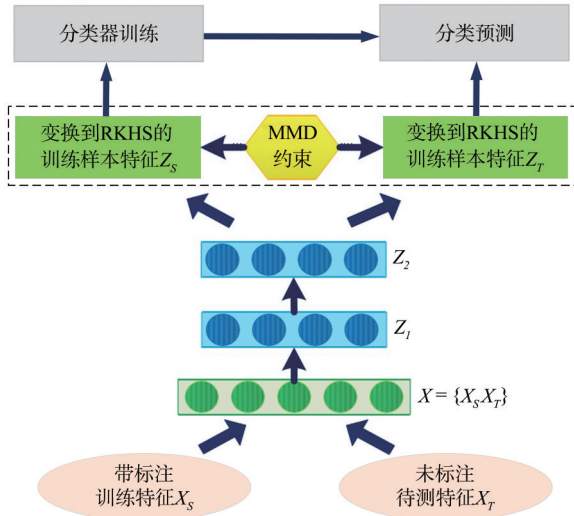


图5 结合最大均值差异的堆栈式自动编码器

励编码器学习有意义的特征表示。第二个条件是最大化源域和目标域样本在最后一层编码层的输出值之间的均值差异，从而减小源域和目标域之间的特征表示差异。所以整个网络在训练的过程中有两个损失函数，具体如下。

解码器的重构误差损失函数如式（3）所示：

$$L_{SAE} = \frac{1}{2(ns+nt)} \sum_{i=1}^{ns+nt} \|N_i - g(f(Z_i))\|^2 \quad (3)$$

其中 N_i 表示第 i 个训练样本的输入值， $g(f(Z_i))$ 表示第 i 个训练样本在网络中解码层的输出值。

最大均值差异损失函数如式（4）所示：

$$L_{MMD} = \text{MMD}(Z_S^E, Z_T^E)^2 = \left\| \frac{1}{ns} \sum_{i=1}^{ns} \varphi(Z_{si}^E) - \frac{1}{nt} \sum_{j=1}^{nt} \varphi(Z_{tj}^E) \right\|_H^2 \quad (4)$$

其中 Z_{si}^E 表示在 ns 个源域样本中第 i 个样本在最后一层编码层的输出值， Z_{tj}^E 表示在 nt 个目标域样本中第 j 个样本在最后一层编码层的输出值。由于 φ 函数是未知的，所以本文使用高斯核函数隐式地进行样本比较，将 L_{MMD} 损失函数转换改写成以下形式：

$$L_{MMD} = \frac{1}{ns} \sum_{i,j=1}^{ns} K(Z_{Si}, Z_{Sj}) - \frac{2}{nsnt} \sum_{i,j=1}^{ns,nt} K(Z_{Si}, Z_{Tj}) + \frac{1}{nt} \sum_{i,j=1}^{nt} K(Z_{Ti}, Z_{Tj}) \quad (5)$$

最终编码器在训练的过程中 L_{SAE} 损失函数用于全局的优化，包括编码层与解码层的参数优化，而 L_{MMD} 损失函数则在此基础上对编码层进行优化。

2.2 基于跨域集成学习的伪标签标注

在堆栈式自动编码器的训练过程中，我们将源域和目标域的样本数据通过编码器映射到更高层的希尔伯特空间下的特征表示，可以在一定程度上将两者之间的分布差异减小。但是由于自动编码器的训练是一个非常复杂的优化过程，初始参数可能不是最优的，不能保证每次训练的过程中都能够学习到足够好的特征。为了解决自动编码器参数初始化问题，本文提出在跨领域的基础上通过 Bagging 集成学习方法^[32]来提高预测标签置信度的方法。

在传统集成学习中，Bagging 方法通过对原始数据集进行有放回抽样，产生多个不同的子数据集。然后，针对每个子数据集训练一个弱分类器，再将所有的弱分类器进行决策层融合，形成一个强分类器，以提高整体预测性能。在当前的任务中，Bagging 方法需要对提取到的更高级的特征进行有放回抽样，但 Bagging 方法对于特征的抽样存在问题，因为我们并不确定从特征提取阶段得到的更高级特征用于情绪识别任务时是否足够好。如果提取的特征质量较差，那么使用 Bagging 方法可能导致训练出多个性能差的分类器，这将无法有效解决整体预测精度的问题^[33]。

因此，在使用集成学习应对该情况时，需要从特征提取阶段入手。具体而言，本文使用了3组加入 MMD 的堆栈式自动编码器来进行特征提取，基于集成学习的跨域情绪识别模型如图6所示。源域与目标域的数据同时输入这3组堆栈式自动编码器中，每组编码器在学习两组数据后，会输出一组使得源域与目标域样本数据分布更为相近的希尔伯特空间特征。3组编码器中隐藏层参数初始化不同，导致3组编码机会分别输出不同的特征向量。这3组特征都用来表示源域与目标域样本的原始数据。相比于只使用一组特征，使用3组特征有助于大大降低低质量特征导致的模型性能不足的概率，从而可以提高模型在目标域上的性能。

将源域与目标域数据同时输入3组编码器后进行训练，得到源域和目标域之间3组不同的高层特征表示。然后使用支持向量机算法分别训练源域数据经过3组编码器编码层输出的3组特征，得到3组分类器。将目标域经过编码器编码层输出的特征输入使用源域训练的分类器，得到目标域样本所属类别。最终3组分类器通过投票选举法决策出目标域样本最终的伪标签。

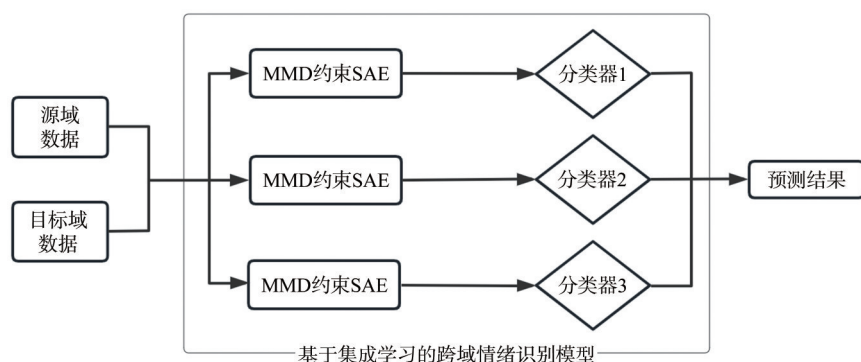


图6 基于集成学习的跨域情绪识别模型

2.3 基于信息熵的加权聚类

在使用数据集中提取出来的特征对数据进行聚类时，考虑到不同个体之间情绪表达存在的差异，从不同个体提取出来的特征的重要程度也会存在差异。在传统的聚类算法中，每个特征都被视为具有相同的重要性，所有的特征都赋予相同的权重，此时会导致聚类的效果不稳定。而在加权聚类中，我们可以为每个特征分配不同的权重，以反映它们在聚类中的相对重要性。

信息熵可以衡量特征之间的重要程度。熵值越高表示特征的取值更加分散或不确定，对聚类的影响较大，应该赋予更大的权重。反之，熵值越低则说明特征的取值更集中在某些值上，对聚类的影响较小，应该赋予其更小的权重。所以本文利用信息熵来量化不同个体情绪数据集中特征的差异程度，并据此为各特征分配不同权重，以提升聚类效果。

针对实验采集到的个体情绪数据集 X 包含80个特征，聚类数目为3，基于信息熵加权聚类^[34]的算法过程如下。

(1) 对于给定的数据集 X ，首先计算每个特征的信息熵。计算特征的信息熵是为了确定每个特征对于聚类的重要性，从而为后续的权重计算做准备。信息熵计算式如下：

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \quad (6)$$

其中 $p(x)$ 表示离散随机变量发生的概率。

(2) 根据不同特征的信息熵计算每个特征的权重，并构建权重矩阵。对聚类结果影响较大的特征被赋予更高的权重，而对聚类结果影响较小的特征被赋予较低的权重，以提高聚类算法对数据特征的敏感性。

(3) 将数据集 X 中的每个样本与权重矩阵进行加权处理，并对加权后的数据进行归一化，消除不同特征的量纲差异。

(4) 最后通过聚类算法对加权 and 归一化后的数据进行聚类，将样本归属到不同的簇中，使得簇内样本之间的相似度较高，而簇间的相似度较低，形成了最终的聚类结果。

2.4 数据联合标注

通过跨域集成学习在一定程度上提高了对无标签数据预测的准确率，但是其准确率是相对的。使用该方法预测出来的伪标签数据作为半监督学习的训练样本，是否有助于模型性能的提高依赖于伪标签的置信程度，错误的伪标签可能会误导模型，在后续迭代训练中还可能误差积累。为了提高伪标签数据的可靠性，本文通过结合跨域集成学习与个体数据加权聚类的联合标注方法来筛选高置信度的样本数据。

首先对无标签数据进行基于信息熵的加权聚类，输出聚类标签数据组，然后将聚类标签数据组通过跨域集成学习进行伪标签标注。加权聚类数据的聚类标签与基于跨域集成学习标注的伪标签共同构成伪标签混淆矩阵。最终的数据以聚类标签为主，并以每个簇为核心，将每个簇中占比低的其他两种数据样本全部舍去，筛选出该聚类标签数据组中伪标签情绪占比最大的样本作为高置信度样本数据，伪标签混淆矩阵如表1所示。

表1 伪标签混淆矩阵

伪标签	聚类标签		
	Cluster_1	Cluster_2	Cluster_3
消极情绪	TP_11	TP_12	TP_13
中性情绪	TP_21	TP_22	TP_23
积极情绪	TP_31	TP_32	TP_33

筛选出来带有高置信度标签的个体情绪数据样本后,用该数据样本对通用情绪识别模型进行参数调整,得到不同个体的个性化情绪识别模型。

3 个体情绪识别实验验证

3.1 实验设计

为了降低外部人员的影响,将实验场景设置在较为偏僻的实验室,温度控制在26℃左右,使用日光灯进行采光,并且去除所有实验所需设备以外的所有电子设备,保证数据采集稳定可靠。为体现个性化的有效性,并且保证数据量足够,实验选取了3位被试,其年龄分布在24岁到26岁之间,驾龄分布在3年到5年之间,均身体健康,且无任何生理或者心理疾病。情绪引导材料来源于北京师范大学中国情绪影像材料库^[35] (Chinese affective picture system, CAPS),包括可诱发愤怒、恐惧、悲伤、快乐、厌恶和中性等6种情绪类型的影像片段共30个,平均片段时间117 s,每种情绪的影像材料使用55寸显示器播放15分钟,分别对积极、消极和中性3种情绪进行引导。面部表情数据采用最高支持1 080 P、60帧的高清摄像头进行采集,将摄像头固定于显示器正上方朝向被试。脑电数据使用EMOTIV Epoc+脑电仪进行采集,心电数据使用Openbci设备进行采集。实验过程如图7所示。



图7 实验过程

在实验前,首先对每位被试进行相关讲解,包括实验内容、数据采集设备的适应、平复心情和减小动作幅度等。为被试穿戴好采集设备并开始实验后,让被试静坐5到10分钟,采集被试情绪数据基线,随后开始播放情绪引导材料,继续采集被试情

绪数据,播放结束后结束该类数据采集,再静坐5到10分钟,更换情绪引导材料类别继续采集数据,如此往复分别采集完3类情绪数据。整个实验过程持续90小时左右,每位被试分别30小时左右。采集的原始情绪数据经过信号分辨、去噪、剔除损坏信号等操作,剩余的每名被试每种情绪状态下的有效实验数据都在50分钟以上。对所有数据以5秒的窗口3秒的步长进行切分,得到每名被试的最终数据样本,其中1号被试数据样本有9 932组,2号被试数据样本有9 056组,1号被试数据样本有9 167组。

3.2 结果呈现

将最终的情绪数据样本进行分组,1号被试情绪样本分为4组,2号和3号被试情绪样本分为3组,每位被试的每组数据对各自的情绪识别模型进行一次优化。第一次进行联合标注时,将从MAHNOB-HCI公共情感数据集与fer2013面部表情数据集中筛选的数据作为源域,实验采集到的1号被试的第一组情绪样本作为目标域,使用本文提出的方法对目标域数据进行加权聚类 and 伪标签预测,得到目标域的聚类标签和伪标签,两者构成伪标签混淆矩阵,并从中筛选出目标域中的高置信度样本数据,结果第一次联合标注时的伪标签混淆矩阵如表2所示,1号被试第一组情绪样本的预测伪标签在聚类簇中的分布如图8所示。

表2 第一次联合标注时的伪标签混淆矩阵

伪标签	聚类标签		
	Cluster_1	Cluster_2	Cluster_3
消极情绪	597	43	175
中性情绪	67	787	505
积极情绪	78	109	639

利用上述筛选出来的个体高置信度样本对通用情绪识别模型进行训练,得到面向个体的情绪识别模型。在后续的联合标注中,再将已用于训练个性化情绪识别模型的高置信度样本与公共数据集中筛选的数据一起作为有标签的源域数据,输入基于集成学习的跨域识别模型中,对剩余组的个体无标签数据进行预测并筛选,如此迭代循环,得到被试最终的个性化情绪识别模型。

3.3 分析验证

1号被试的个性化情绪识别模型优化结果如图9所示。在优化前,该模型对于1号被试的情绪识别整体准确率只有45%左右,其中消极情绪为68%,

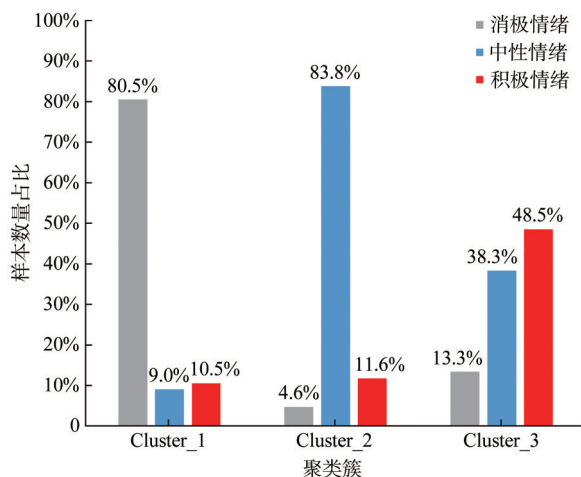


图8 1号被试第一组情绪样本的预测伪标签在聚类簇中的分布

中性情绪为50%，积极情绪识别准确率仅为18%。第一次优化后模型的性能得到大幅度提升，整体准确率达到约69%，其中积极情绪识别准确率提高了36%，中性情绪识别准确率提高了25%，消极情绪识别准确率提高了9%。第二次优化后，模型的性能也有大幅度的提高，整体准确率达到约79%。第三次优化后积极情绪识别准确率有小幅度降低，但模型的整体性能还在提升，只是提升的效果相比于前两次明显下降。第四次优化后模型的性能提升幅度也不大。

2号被试个性化情绪识别模型的优化结果如图10所示。在优化前，通用情绪识别模型对于2号

被试的情绪识别整体性能较低，仅为45%。通过第一次优化，模型性能大幅提高，且在之后的优化过程中性能持续提升，但提升幅度逐渐减小。

3号被试个性化情绪识别模型的优化结果如图11所示，表现出如前两名被试同样的趋势。

从以上3名被试的个体情绪识别模型优化结果可以观察到，由于个体的差异性，模型的初始性能和不同阶段优化后的性能存在轻微差异。但在优化前，3名被试的模型准确率情况相似，比如三者的积极情绪识别准确率最低。在经过优化后，模型则呈现出不同的变化，有的消极情绪的表达更为明显，而有的在中性或积极情绪上表达更为明显，这表明个体在情绪表达上存在一定差异。值得注意的是，使用本文提出的方法后，不同个体的情绪识别成功率均大幅提升，这证明了本文方法的可行性。

4 结论

传统的通用情绪识别模型通常基于公共数据集进行训练，但在面对不同个体时，其识别精度较低。为解决此问题，本文提出基于集成学习与聚类联合标注方法的个性化情绪识别技术。通过对通用情绪识别模型进行多次优化，建立了多模态个性化情绪识别模型，使其能够更准确地适应不同个体的情绪状态，在面向数据集外的不同个体时具有更高的识别精度。同时，采用基于集成学习与聚类的联合标注方法，对个体情绪数据进行自动伪标签标

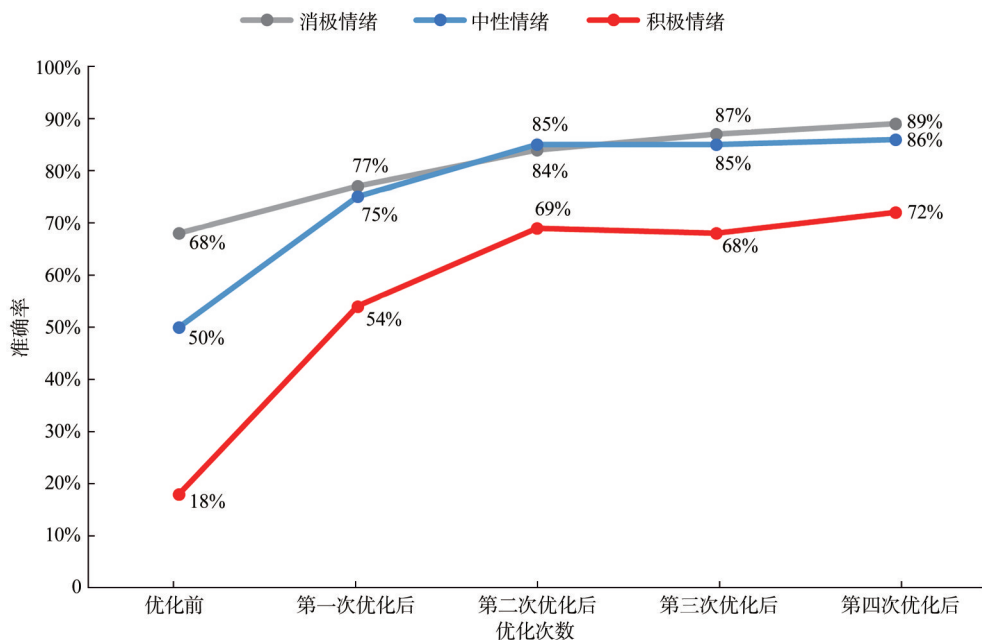


图9 1号被试情绪识别模型优化结果

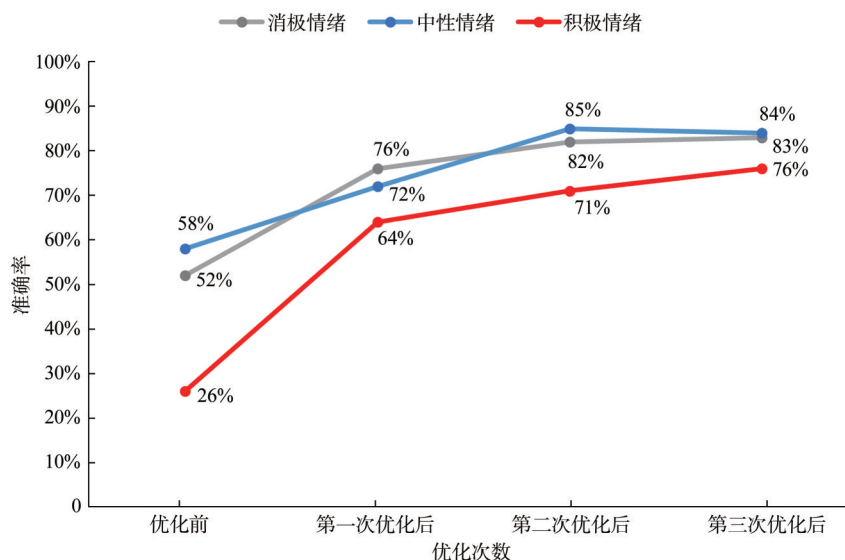


图10 2号被试情绪识别模型优化结果

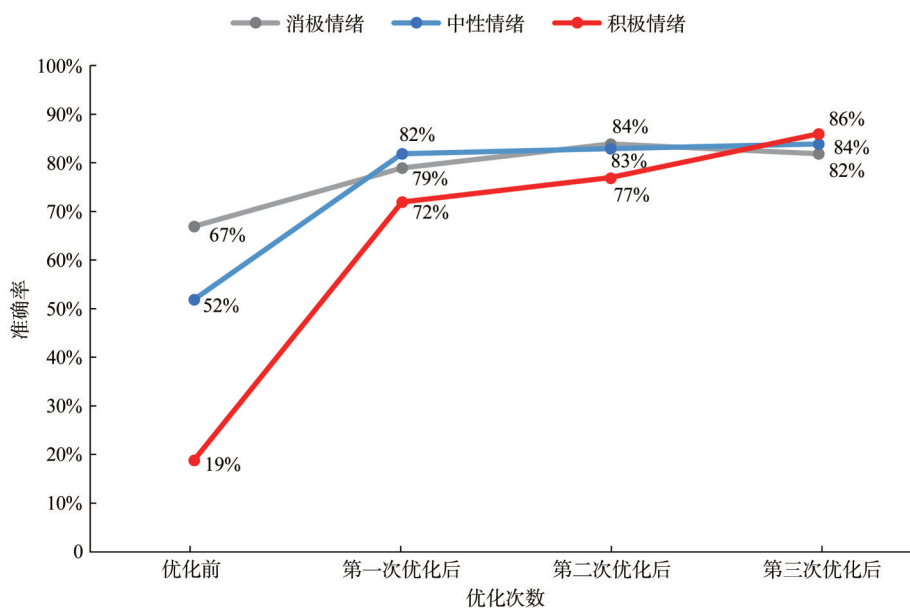


图11 3号被试情绪识别模型优化结果

注，极大降低了工作量并提高了伪标签的准确性和可靠性。实验结果表明，使用本文方法对通用情绪识别模型进行额外训练后得到的个性化情绪识别模型，在面向不同个体时，其情绪识别准确性均有大幅度的提升，证明本文方法在解决个体情绪识别时具有良好的应用潜力。在未来的工作中，将联合云端算力，构建个性化情绪识别及其模型优化系统，实现情绪状态实时预测和识别模型的持续优化。

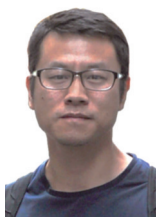
参考文献：

- [1] 殷梦馨, 倪娜, 尉怀怀, 等. 基于多模态情绪识别的研究进展[J]. 生物医学工程研究, 2023, 42(3): 285-291.

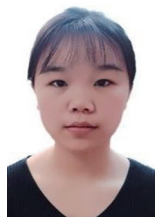
- YIN M X, NI N, YU H H, et al. Research progress based on multi-modal emotion recognition[J]. Journal of Biomedical Engineering Research, 2023, 42(3): 285-291.
- [2] CHEN M, ZHOU P, FORTINO G. Emotion communication system[J]. IEEE Access, 2016, 5: 326-337.
- [3] ZEPF S, HERNANDEZ J, SCHMITT A, et al. Driver emotion recognition for intelligent vehicles: a survey[J]. ACM Computing Surveys, 2020, 53(3): 64.
- [4] XIA Y X, LIU Y H. EEG-based emotion recognition with consideration of individual difference[J]. Sensors, 2023, 23(18): 7749.
- [5] ZHAO J M, LI R C, JIN Q, et al. Memobert: pre-training model with prompt-based learning for multimodal emotion recognition[C]//Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 4703-4707.

- [6] LEE D H. Pseudo-Label: the simple and efficient semi-supervised learning method for deep neural networks[C]//Proceedings of International Conference on Machine Learning.[S.l.:s.n.], 2013.
- [7] SOHN K, BERTHELOT D, LI C L, et al. FixMatch: simplifying semi-supervised learning with consistency and confidence[C]//Proceedings of the Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 596-608.
- [8] LI H Y, WANG N N, YANG X, et al. Towards semi-supervised deep facial expression recognition with an adaptive confidence margin[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 4156-4165.
- [9] ZHU Z, SATO Y. Speech emotion recognition using semi-supervised learning with efficient labeling strategies[C]//Proceedings of the 2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU). Piscataway: IEEE Press, 2021: 358-365.
- [10] ZHANG S, CHEN M, CHEN J C, et al. Combining cross-modal knowledge transfer and semi-supervised learning for speech emotion recognition[J]. Knowledge-Based Systems, 2021, 229: 107340.
- [11] CHEN M, HAO Y X. Label-less learning for emotion cognition[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7): 2430-2440.
- [12] ZHANG Z X, HAN J, DENG J, et al. Leveraging unlabeled data for emotion recognition with enhanced collaborative semi-supervised learning[J]. IEEE Access, 2018, 6: 22196-22209.
- [13] 李学聪, 战荫伟, 杨卓, 等. 基于标签分布学习的眼部情绪识别[J]. 计算机科学与应用, 2022(4): 1213-1225.
LI X C, ZHAN Y W, YANG Z, et al. Eye emotion recognition based on label distribution learning[J]. Computer Science and Application, 2022(4): 1213-1225.
- [14] SOLEYMANI M, LICHTENAUER J, PUN T, et al. A multimodal database for affect recognition and implicit tagging[J]. IEEE Transactions on Affective Computing, 2012, 3(1): 42-55.
- [15] CHEN G H, ZENG X P. Multi-modal emotion recognition by fusing correlation features of speech-visual[J]. IEEE Signal Processing Letters, 2021, 28: 533-537.
- [16] ZHANG Y, CHENG C, ZHANG Y D. Multimodal emotion recognition using a hierarchical fusion convolutional neural network[J]. IEEE Access, 2021, 9: 7943-7951.
- [17] SUBRAMANIAN G, CHOLENDIRAN N, PRATHYUSHA K, et al. Multimodal emotion recognition using different fusion techniques[C]//Proceedings of the 2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII). Piscataway: IEEE Press, 2021: 1-6.
- [18] 凌文芬, 陈思含, 彭勇, 等. 基于3D分层卷积融合的多模态生理信号情绪识别[J]. 智能科学与技术学报, 2021, 3(1): 76-84.
LING W F, CHEN S H, PENG Y, et al. Multi-modal physiological signal emotion recognition based on 3D hierarchical convolution fusion[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(1): 76-84.
- [19] CHEN Y S, JIA Z Y, HIROTA K, et al. A multimodal emotion perception model based on context-aware decision-level fusion[C]//Proceedings of the 2022 41st Chinese Control Conference (CCC). Piscataway: IEEE Press, 2022: 7332-7337.
- [20] ZHAO Y X, CAO X Y, LIN J L, et al. Multimodal affective states recognition based on multiscale CNNs and biologically inspired decision fusion model[J]. IEEE Transactions on Affective Computing, 2023, 14(2): 1391-1403.
- [21] PAN B, HIROTA K, JIA Z Y, et al. A review of multimodal emotion recognition from datasets, preprocessing, features, and fusion methods[J]. Neurocomputing, 2023, 561: 126866.
- [22] GUYON I, WESTON J, BARNHILL S, et al. Gene selection for cancer classification using support vector machines[J]. Machine Learning, 2002, 46(1): 389-422.
- [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] MURUGESAN B, RAVICHANDRAN V, RAM K, et al. ECGNet: deep network for arrhythmia classification[C]//Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA). Piscataway: IEEE Press, 2018: 1-6.
- [25] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB]. arXiv preprint, 2014, arXiv: 1409.1556.
- [26] GOODFELLOW I J, ERHAN D, CARRIER P L, et al. Challenges in representation learning: a report on three machine learning contests[J]. Neural Networks, 2015, 64: 59-63.
- [27] 高寒冰, 张弛, 靳明艳, 等. 基于多源域自适应的跨被试情感脑电识别[J]. 智能科学与技术学报, 2021, 3(1): 59-64.
GAO H B, ZHANG C, JIN M Y, et al. Cross-subject emotional EEG recognition based on multi-source domain adaptation[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(1): 59-64.
- [28] 吕宝粮, 张亚倩, 郑伟龙. 情感脑机接口研究综述[J]. 智能科学与技术学报, 2021, 3(1): 36-48.
- [29] LU B L, ZHANG Y Q, ZHENG W L. A survey of affective brain-computer interface[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(1): 36-48.
- [30] FARAHANI A, VOGHOEI S, RASHEED K, et al. A brief review of domain adaptation[C]//Proceedings of Advances in Data Science and Information Engineering. Cham: Springer, 2021: 877-894.
- [31] GLOROT X, BORDES A, BENGIO Y. Domain adaptation for large-scale sentiment classification: a deep learning approach[C]//Proceedings of the Proceedings of the 28th International Conference on International Conference on Machine Learning. New York: ACM, 2011: 513-520.
- [32] 柴鑫. 基于领域适配的跨个体脑电情绪识别方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2018.
CHAI X. Research on cross-individual EEG emotion recognition method based on domain adaptation[D]. Harbin: Harbin Institute of Technology, 2018.
- [33] SALAMA E S, EL-KHORIBI R A, SHOMAN M E, et al. A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition[J]. Egyptian Informatics Journal, 2021, 22(2): 167-176.
- [34] 胡毅, 瞿博阳, 梁静, 等. 进化集成学习算法综述[J]. 智能科学与技术学报, 2021, 3(1): 18-33.
HU Y, QU B Y, LIANG J, et al. A survey on evolutionary ensemble learning algorithm[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(1): 18-33.
- [35] CARDONE B, DI MARTINO F, SENATORE S. Improving the emotion-based classification by exploiting the fuzzy entropy in FCM clustering[J]. International Journal of Intelligent Systems, 2021, 36(11): 6944-6967.
- [36] 徐鹏飞, 黄宇霞, 罗跃嘉. 中国情绪影像材料库的初步编制和评定[J]. 中国心理卫生杂志, 2010, 24(7): 551-554, 561.
XU P F, HUANG Y X, LUO Y J. Establishment and assessment of native Chinese affective video system[J]. Chinese Mental Health Journal, 2010, 24(7): 551-554, 561.

[作者简介]



柯善军（1975- ），男，重庆理工大学车辆工程学院副教授、硕士生导师，主要研究方向为机电产品工业设计、用户体验与人机交互、汽车主动安全。



王钰苗（2000- ），女，重庆理工大学车辆工程学院硕士生，主要研究方向为用户体验与人机交互。



聂成洋（2000- ），男，重庆理工大学车辆工程学院硕士生，主要研究方向为驾驶员状态监测。



何邦胜（1998- ），男，重庆理工大学车辆工程学院硕士生，主要研究方向为智能座舱测试技术。