计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(2):1−11 [doi: 10.15888/j.cnki.csa.007777]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于人脸图像和脑电的连续情绪识别方法①

李瑞新, 蔡兆信, 王冰冰, 潘家辉

(华南师范大学 软件学院, 佛山 528225)
通讯作者: 潘家辉, E-mail: panjh82@qq.com

摘　要: 基于多模态生理数据的连续情绪识别技术在多个领域有重要用途, 但碍于被试数据的缺乏和情绪的主观性, 情绪识别模型的训练仍需更多的生理模态数据, 且依赖于同源被试数据. 本文基于人脸图像和脑电提出了多种连续情绪识别方法. 在人脸图像模态, 为解决人脸图像数据集少而造成的过拟合问题, 本文提出了利用迁移学习技术训练的多任务卷积神经网络模型. 在脑电信号模态, 本文提出了两种情绪识别模型: 第一个是基于支持向量机的被试依赖型模型, 当测试数据与训练数据同源时有较高准确率; 第二个是为降低脑电信号的个体差异性和非平稳特性对情绪识别的影响而提出的跨被试型模型, 该模型基于长短时记忆网络, 在测试数据和训练数据不同源的情况下也具有稳定的情绪识别性能. 为提高对同源数据的情绪识别准确率, 本文提出两种融合多模态决策层情绪信息的方法: 枚举权重方法和自适应增强方法. 实验表明: 当测试数据与训练数据同源时, 在最佳情况下, 双模态情绪识别模型在情绪唤醒度维度和效价维度的平均准确率分别达 74.23% 和 80.30%; 而当测试数据与训练数据不同源时, 长短时记忆网络跨被试型模型在情绪唤醒度维度和效价维度的准确率分别为 58.65% 和 51.70%.

关键词: 连续情绪识别; 迁移学习; 多任务卷积神经网络; 跨被试型模型; 长短时记忆网络; 决策层信息融合

引用格式: 李瑞新,蔡兆信,王冰冰,潘家辉.基于人脸图像和脑电的连续情绪识别方法.计算机系统应用,2021,30(2):1–11. http://www.c-s-a.org.cn/1003-3254/7777.html

Continuous Emotion Recognition Based on Facial Expressions and EEG

LI Rui-Xin, CAI Zhao-Xin, WANG Bing-Bing, PAN Jia-Hui

(School of Software, South China Normal University, Foshan 528225, China)

Abstract: Continuous emotion recognition based on multimodal physiological data plays an important role in many fields. However, it needs more physiological data to train emotion recognition models due to the lack of subjects' data and subjectivity of emotion, and it is largely affected by homologous subjects' data. In this study, we propose multiple emotion recognition methods based on facial expressions and EEG. Regarding the modality of facial images, we propose a multi-task convolutional neural network trained by transfer learning to avoid over-fitting induced by small datasets of facial images. With respect to the modality of EEG, we propose two emotion recognition models. The first is a subject-dependent model based on support vector machine, possessing high accuracy when the validation and training data are homogeneous. The second is a cross-subject model for reducing the impact caused by the individual variation and non-stationarity of EEG. It is based on a long short-term memory network, performing stably under the circumstance that validation and training data are heterogeneous. To improve the accuracy of emotion recognition for homogeneous data, we propose two methods for decision-level fusion of multimodal emotion prediction: Weight enumeration and adaptive

① 基金项目: 广州市科技计划重点领域研发计划 (202007030005); 广东省自然科学基金面上项目 (2019A1515011375); 广东大学生科技创新培育专项资金 ("攀登计划"专项资金) (pdjh2020a0145)
Foundation item: Research and Development Plan of Key Areas, Science and Technology Plan of Guangzhou Municipality (202007030005); General Program of Natural Science Foundation of Guangdong Province (2019A1515011375); Special Fund for Scientific and Technological Innovation and Cultivation for Students of Higher Education of Guangdong Province (Special Fund for Climbing Plan) (pdjh2020a0145)
收稿时间: 2020-06-13; 修改时间: 2020-07-10; 采用时间: 2020-07-23; csa 在线出版时间: 2021-01-27

boost. According to the experiments, when the validation and training data are homogeneous, under the best circumstance, the average accuracy that multimodal emotion recognition models reached in both arousal and valence dimensions were 74.23% and 80.30%; as the validation and training data are heterogeneous, the accuracy that the cross-subject model reached in both arousal and valence dimensions are 58.65% and 51.70%.

Key words: continuous emotion recognition; transfer learning; multi-task convolutional neural network; Long Short-Term Memory (LSTM) network; cross-subject model; decision-level fusion

1 引言

1.1 研究背景

情绪 (emotion) 是人对客观事物的态度体验和相应的行为反映[1], 是一种由感觉、思想与行为综合而成的复杂的心理和生理状态, 它与大脑许多内部和外部活动相关联, 在生活中的各个方面都起重要作用. 情绪识别在心理学研究、安全驾驶、犯罪测谎、远程教育、人机交互、数字医疗等领域有着重要的影响和需求. 情绪识别技术涵盖人工智能、自然语言处理、认知与社会科学等领域的方法和技术[2]. 但是, 目前情绪识别的量化精度不高, 又囿于被试生理数据的缺乏以及情绪的主观性, 目前情绪识别在技术层面仍然需要克服数据集小、跨被试性能差等问题. 基于此背景, 本文采用 Posner 提出的情绪的二维模型[3] 量化情绪, 将情绪分为效价 (valence) 和唤醒度 (arousal) 两个维度,每个维度的分数范围为 1–9. 同时, 本文基于人脸图像和脑电技术, 提出了多个情绪识别模型.

1.2 研究现状

(1) 人脸表情识别的相关研究

人脸的表情是一种重要的情绪交流方式, 1971 年, Ekman 等[4] 首次将表情划分为 6 种基本形式: 悲伤 (sad)、高兴 (happy)、恐惧 (fear)、厌恶 (disgust)、惊讶 (surprise) 和愤怒 (angry). 而人脸表情识别 (Facial Expression Recognition, FER) 技术则将生理学、心理学、图像处理、机器视觉 与模式识别等研究领域进行交叉与融合, 是近年来模式识别与人工智能领域研究的一个热点问题[5]. 传统的人脸表情识别方法重视特征提取和表情分类. 2016 年, Meng 等[6] 将 Roweis 研究团队提出的 LLE 方法[7] 与神经网络进行结合, 提出了 LLENET 特征提取算法, 显著提高了算法的性能. 2009 年, 朱明旱等[8] 结合二维 Fisher 线性判别分析 (Two Dimensional Fisher Linear discriminant Analysis, 2DFLA) 与局部保持投影算法识别表情, 显著提升了识别效率. 基于深度学习的人脸表情识别方法能够同时提取特征并分类表情. Mollahosseini 等[9] 将 AlexNet 与 GoogleNet 模型结合, 构建了一个 7 层的卷积神经网络 (Convolutional Neural Networks, CNN) 用于人脸表情识别, 得到了较好的识别效果. 目前人脸表情识别的研究有如下难点: ① 表情量化方式不精确; ② 表情适用范围小; ③ 被试数据量不足, 难以训练更复杂的深度学习模型.

(2) 脑电情绪识别的相关研究

相较于人脸表情, 脑电信号具有的更高的客观性, 难以伪装. 因此, 脑电信号在情绪识别领域备受关注. 现有的脑电情绪识别技术大多针对时域特征、频域特征、时频域特征和空间域特征 4 个方向进行特征值挖掘[10], 以达到更好的分类效果. 1924 年, St. Louis 等[11] 首次提出在实践中应用脑电技术, 后来该技术被应用于情绪识别领域. 2009 年, Yazdani 等[12] 利用贝叶斯线性判别分析, 基于脑电模态对喜悦、愤怒、厌恶、悲伤、惊讶、恐惧等 6 种情绪进行分类, 实验表明准确率超过 80%. 2015 年, Georgieva 等[13] 采用 6 种无监督算法构建被试内和被试间的情绪模型, 实验表明模糊 C 均值据类算法的效果最佳. 2020 年, 郑伟龙等[14] 用异质迁移学习构建跨被试脑电情感模型, 利用眼动信号作为量化被试间域差异的标准, 初步实现了跨被试情绪识别, 准确率达到 69.72%. 目前基于脑电的情绪识别研究有如下难点: ① 脑电信号具有非平稳性, 难以挖掘合适的特征值; ② 脑电信号个体差异性显著, 大多模型为被试依赖型模型, 难以在保证准确率的情况下实现跨被试型情绪识别模型.

(3) 多模态情绪信息融合的相关研究

基于不同的生理模态, 情绪识别的研究方法有很多. 但是, 单一模态的情绪识别往往准确率比多模态情绪识别低. 正如前文所述, 人脸表情容易伪装, 脑电情绪的跨被试性能差, 各个生理模态的信息有不同的优缺点. 因此, 近年来, 针对不同层次的模态信息融合算法也在快速发展中. 通过融合多种互补的模态生理信

息, 能够切实提高情绪识别的准确率和适用范围.

针对以上研究现状, 本文的工作是: ① 对于人脸表情识别, 利用迁移学习技术训练多任务卷积神经网络模型, 以避免因数据量少而导致的过拟合现象. ② 对于脑电情绪识别, 本文提出了两种互相独立的方法, 第一种是准确率较高的被试依赖型模型—支持向量机 (Support Vector Machine, SVM); 第二种是适用范围广的跨被试型模型—长短时记忆网络 (Long Short-Term Memory, LSTM) 网络. ③ 融合人脸图像模态和脑电信号模态的决策层信息 (情绪得分), 以提高情绪识别的准确率. 其中, 针对被试依赖型模型, 我们将 SVM 和 CNN 子分类器进行模态融合, 而对于基于脑电信号的跨被试模型 LSTM, 则作为单模态情绪识别模型, 独立进行实验.

2 基于人脸图像的情绪识别

2.1 基于人脸图像识别表情的基本流程

在基于人脸图像识别表情的模块中, 通过系统调用摄像头以 4 Hz 的频率对视频进行图像数据采样, 使用文献 [15] 的基于 Haar 特征值的自适应增强 (Adaptive Boost, AdaBoost) 算法[16] 检测人脸, 并将提取的人脸图像信息转换为宽和高皆为 48 像素的矩阵, 将该矩阵输入至多任务卷积神经网络 (Multi-Task Convolutional Neural Networks, MTCNN)[16] 中, 以预测人脸表情的效价和唤醒度得分. 其系统运行流程如图 1 所示.

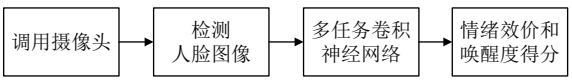

图 1 人脸表情识别系统运行流程

2.2 人脸检测

本文采用基于 Haar 特征值的 AdaBoost 模型进行人脸检测. 对于 AdaBoost 算法而言, 用式 (1) 假定一个训练数据集 T, 用式 (2) 假定权值系数D_i.

$$T=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\} \tag{1}$$

$$D_i=(\omega_{11},\cdots,\omega_{1i},\cdots,\omega_{1N}),\omega_{1i}=\frac{1}{N},i=1,2,\cdots,N \tag{2}$$

其中, $x_i\in X\subseteq R^n$为实例, X 是实例空间, Y 是标记集合. 最终分类器$G(x)$由多个弱分类器线性组合而成. 弱分类器$y_i\in Y=\{-1,+1\}$的分类误差率e_m由式 (3) 表示, 弱分类器$G_m(x)$的系数α_m由式 (4) 表示.

$$e_m=\sum_{i=1}^{N}P(G_m(x)\neq y_i)=\sum_{i=1}^{N}\omega_m I(G_m(x)\neq y_i) \tag{3}$$

$$\alpha_{m+1}=\frac{1}{2}\ln\frac{1-e_m}{e_m} \tag{4}$$

每次计算出更新的训练数据集的权值分布D_{m+1}如式 (5) 所示, 权值向量中的每个权值由式 (6) 表示. 式 (6) 中的Z_m是规范化因子.

$$D_{m+1}=(\omega_{m+1,1},\cdots,\omega_{m+1,i}\cdots,\omega_{m+1,N}) \tag{5}$$

$$\omega_{m+1,i}=\frac{\omega_{mi}}{Z_m}\exp(-\alpha_m y_i G_m(x_i)),i=1,2,\cdots,N \tag{6}$$

$$Z_m=\sum_{i=1}^{N}\omega_{mi}\exp(-\alpha_m y_i G_m(x_i)) \tag{7}$$

通过不断地训练, 可以得到如式 (8) 所示的最终分类器. AdaBoost 算法执行流程如图 2 所示.

$$G(x)=sign\left(\sum_{m=1}^{M}\alpha_m G_m(x)\right) \tag{8}$$

本文采用 AdaBoost 算法在人脸检测及模态信息融合模块中进行分类预测. 在人脸检测中, 使用 OpenCV 开源框架中已训练的分类模型, 该模型通过 (*.xml) 文件存储信息, 是用于检测人脸及前额的 AdaBoost 检测方法.

2.3 利用迁移学习技术训练 CNN 人脸表情识别模型

(1) 多任务卷积神经网络

我们利用迁移学习技术, 训练一个多任务卷积神经网络来进行人脸图像的特征提取和特征分类. 具体来说, 训练网络的过程分为 2 步. 第 1 步, 先将网络在一个具有图像级别标注的大数据集 (Fer2013) 进行训练[17]. 第 2 步, 将模型所有卷积层参数固定, 以相对比较小的学习率 (0.001) 在小数据集 (我们的目标数据集的划分) 上再进行二次训练 (微调), 这样才能完成模型的训练.

得到充分训练的 CNN 模型之后. 对于一个视频, 在找出视频中的人脸之后, 我们将多个人脸分别输入模型得到多个子结果, 通过这些子结果的投票, 我们得到这个视频的基于脸部的情绪结果 (valence 和 arousal 的分类).

CNN 得到子结果的过程是通过神经网络的一个从输入端到输出端的前向传播. 具体过程如下, 对于一个 48×48 的灰度图, 首先被模型的 3 个卷积层提取图像特征, 第 1 个卷积层为 32 个 3×3×1 的卷积核. 第 2 个卷积层是具有 32 个大小为 3×3×32 的核. 第 3 个卷积层有 64 个大小为 3×3×32 的核. 提取出来的特征经过铺平后送到第 4 层与 64 个神经元完全连接. 所有卷积层和全连接层, 都应用 ReLU 激活函数[18]. 网络随后分为两个分支预测任务. 本文所提出的卷积神经网络结构如图 3 所示.

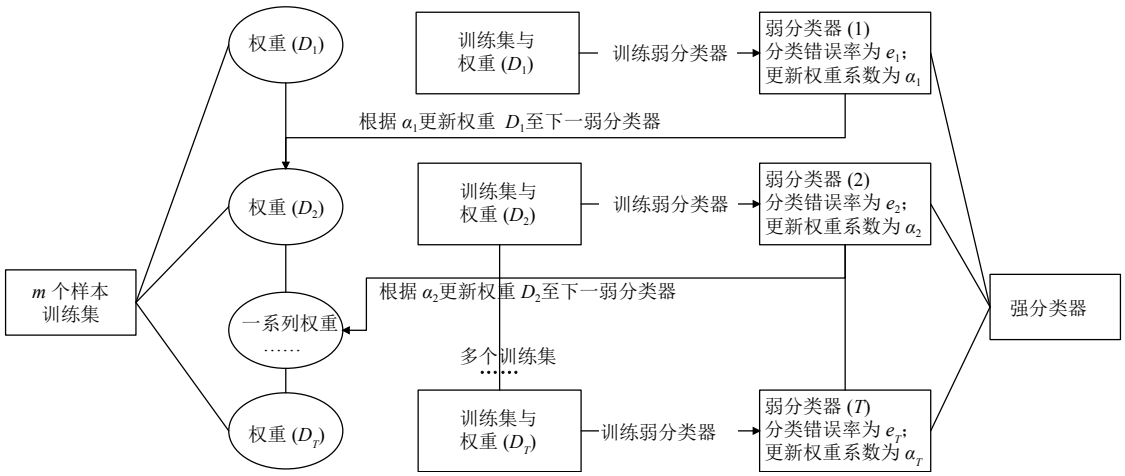

图 2 AdaBoost 算法执行流程

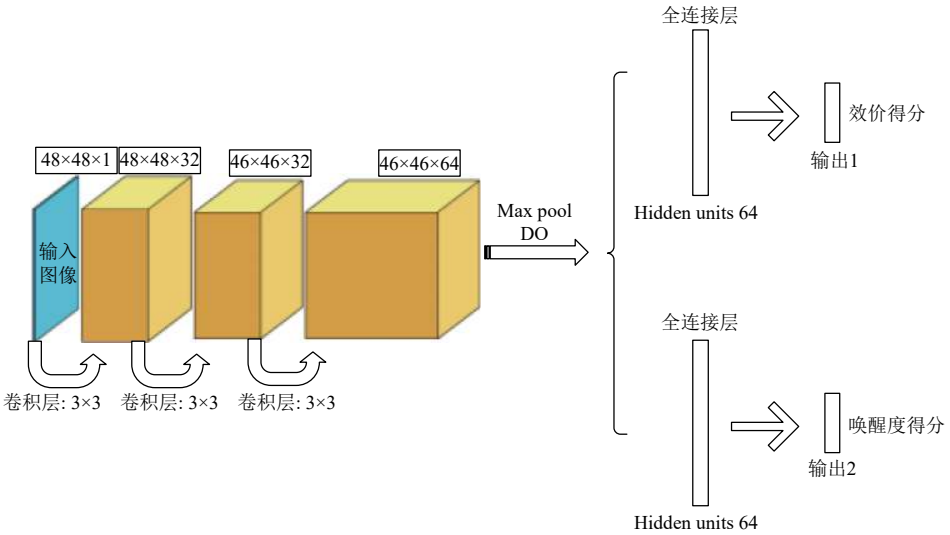

图 3 多任务卷积神经网络模型, 其中“DO”为 CNN 的 dropout 层

(2) 基于卷积神经网络的情绪回归计算

第一个分支学习计算效价得分, 它包含两个全连接的大小为 64 和 1 的层. 然后将输出输入到 Sigmoid 函数中, 并最大限度地减少交叉熵损失 L_1:

$$L_1 = -\sum_{i=1}^{m}(1-y_{1i}\log y_{1i}+y_{1i})\log(1-y_{1i}) \tag{9}$$

其中, y_{1i}表示第 i 个样本效价的真实标签 (ground-truth labels), $\hat{y}_{1i}$表示第 i 个样本对应于情绪效价的模型输出, m 表示训练样本的大小. 第二个分支是针对唤醒度进行预测的, 它包含两个全连接 i 的大小为 64 和 1 的层. 输出被馈送到 Sigmoid 函数, 我们再次最小化交叉熵损失 L_2:

$$L_2 = -\sum_{i=1}^{m}(1-y_{2i}\log y_{2i}+y_{2i})\log(1-y_{2i}) \tag{10}$$

其中, y_{2i}表示第i个样本中唤醒度的真实标签, $\hat{y}_{2i}$表示第 i 个样本对应于唤醒度的模型输出, m 表示训练样本的大小. 最终, 我们最小化 L_1 和 L_2的联合损失.

$$L = \sum_{p=1}^{2}\alpha_p L_i \tag{11}$$

其中, α_p是的线性权重, 也是模型需要确定的超参数. 如果我们将第二个权重设置为 0, 模型将退化为传统的单任务学习方法. 在模型充分训练完之后, 我们可通过式 (12) 从网络的输出值S_{face}中得到情绪效价和唤醒度分类的结果如下:

$$r_{\text{face}} = \begin{cases} \text{high}, & S_{\text{face}} \geq 0.5 \\ \text{low}, & S_{\text{face}} < 0.5 \end{cases} \tag{12}$$

例如, 如果上分支效价得分的输出为$S_{\text{face}} = 0.8$, 那么认为它对应的效价结果属于 high 一类. 对于表情数据的回归计算, 本文的损失函数不再是交叉熵, 而是均

方差误差. 然后分别预测效价和唤醒度的数值连续大小.

3 基于脑电信号的情绪识别

3.1 基于脑电信号识别情绪的基本流程

在基于脑电信号识别情绪模块中, 使用 Emotiv Eopc+的脑机接口采集生理数据, 并利用小波变换提取特征值, 选取好特征值后再利用 SVM 或者 LSTM 识别情绪. 其中, 基于 SVM 的情绪识别方法为被试依赖型模型, 基于 LSTM 的情绪识别方法为跨被试型模型. 基于脑电信号的情绪识别系统运行流程如图 4 所示.

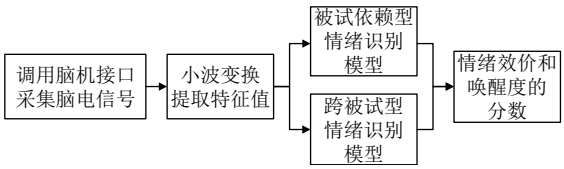

图 4 脑电情绪识别系统运行流程

3.2 利用小波变换提取脑电信号特征值

在特征值提取与选取阶段, 利用小波变换从原始 EEG 数据中获得功率谱密度 (Power Spectral Density, PSD) 特征. 小波变换适用于多尺度分析, 这意味着可以使用不同的频率和时间尺度检查信号. 本文采用 Daubechies 的小波变换系数[19] 进行特征提取, 小波变换公式如下所示:

$$\omega_f(s,\tau)=\int_{-\infty}^{\infty} f(t)\varphi_{s\tau}(t)\mathrm{d}t \tag{13}$$

其中, $\omega_f(s,\tau)$表示一维连续小波变换, φ表示小波母函数, s 表示尺度参数, t 为平移参数. 而连续小波逆变换的公式如下所示:

$$f(t)=\frac{1}{C_\varphi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W_f(s,\tau)\varphi_{s\tau}(t)\frac{\mathrm{d}t\mathrm{d}s}{s^2} \tag{14}$$

其中, $C_\varphi=\int_{-\infty}^{\infty}(|\varphi(u)|^2/|u|\mathrm{d}u)$, $\varphi(u)$为$\varphi(t)$的傅里叶变换.

在提取特征值后, 情绪识别模型分为两种情况: 第一种为一个模型仅适用于一个被试, 即模型依赖于被试 (subject dependence). 此时训练数据集和测试数据集为同源数据, 来自于同样的被试, 没有域差异. 第二种则为一个模型适用于所有被试, 即模型不依赖于被试 (subject independence). 此时训练数据集和测试数据集来自于完全不同的被试, 有一定的域差异. 针对情况一, 为构建被试依赖型模型 (subject dependent models), 使用递归特征消除算法 (Recursive Feature Elimination, RFE) 进一步选择了提取的特征, 并将所选特征再通过 SVM 进行分类以获得基于脑电信号的情绪状态. 针对情况二, 为构建跨被试型模型 (cross-subject models), 可以通过构建长度为 10 s 的时序特征, 将所有特征利用长短时记忆网络模型模型进行预测, 从而跨被试预测脑电情绪状态.

3.3 利用 SVM 构建被试依赖型脑电情绪识别模型

本文的算法使用 14 个通道 (AF3, F3, F7, FC5, T7, P7, O1, AF4, F4, F8, FC6, T8, P8, O2) 进行特征提取. 使用的 5 个频率波段分别为 theta (4 Hz<f<8 Hz)、slow alpha (8 Hz<f<10 Hz)、alpha (10 Hz<f<12 Hz)、beta (12 Hz<f<30 Hz) 以及 gamma (30 Hz<f<45 Hz), 共有 14×5=70 个特征.

在第二步分类中, 在最终的特征被选择之后, 本文用一个应用于高斯核的 SVM 进行分类, 且该 SVM 的惩罚系数$C=1.0$. 当惩罚系数$C=1.0$时该模型能够达到较好的效果, 弱数值过大, 则易导致过拟合, 若数值过小则容易欠拟合. 为训练模型, 我们去除了权重最低的 10% 的特征数据, 并使用 10 倍交叉验证分割训练数据集[20]. 训练完模型之后, 对于不同的任务 (预测 valence 和 arousal), 我们分别用不同的 SVM 进行预测. 每个对应的 SVM 预测出得分S_{EEG}. 我们再根据这个得分, 通过式 (13) 获得基于脑电波的结果r_{EEG}.

$$r_{\mathrm{EEG}}=\begin{cases}\text{high}, & S_{\mathrm{EEG}}\geq 0.5\\ \text{low}, & S_{\mathrm{EEG}}<0.5\end{cases} \tag{15}$$

3.4 利用 LSTM 构建跨被试型脑电情绪识别模型

本文提出利用 LSTM 构建跨被试脑电情绪识别模型. 该过程分为两步, 第 1 步先进行构造时序特征, 第 2 步再使用 LSTM 进行回归预测.

在构建跨被试模型时所有特征值仍然如前文所述, 但选取方式有了变化. 在构造时序特征时, 以 10 s 作为一个样本, 以 50% 的重叠率采样. 并且, 以每一秒作为一个时间单元, 比如说对于离线实验, 一秒有 85 个特征, 那么本文的一个样本是一个二维矩阵第 1 维是 10, 而第 2 维是 85. 而不是一个大小为 850 的一维向量. 样本的构造跟 LSTM 的结构有关.

在使用 LSTM 进行预测时, 网络首先是两层 LSTM 层, 跟着一个全连接层, 然后接着是输出层. 第一个 LSTM 层由 10 个 LSTM 单元 (LSTM cell) 组成, 每个单元包含 128 个神经元. 第二层 LSTM 层由 10 个 LSTM 单元 (LSTM cell) 组成, 每个单元包含 64 个神经元. 全连接层包含 54 个神经元. 输出层由 2 个神经元构成代

表情绪的效价得分和唤醒度得分. 每个层都应用了 0.5 的 dropout. 每层都应用了 ReLU 激活函数以及在每层之间, 本文都进行了数据归一化. 采用均方差作为网络损失函数.

4 融合双模态决策层信息的情绪识别

4.1 融合双模态决策层信息的算法流程

本系统通过调用摄像头和脑机接口设备采集两个模态的生理数据, 并在各模态情绪识别模型的决策层进行信息融合, 以提高情绪识别准确率. 图 5 为本系统进行双模态情绪识别的运行流程图. 首先对采集的人脸图像信息和脑电信息进行预处理, 并提取特征值, 然后分模块各自进行情绪量化计算, 并最终融合两个模态的情绪得分.

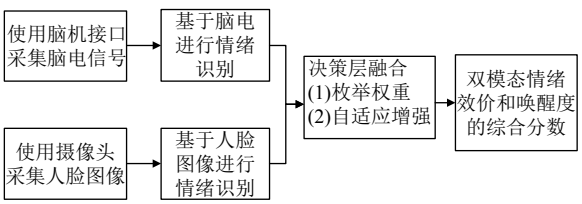

图 5 双模态情绪识别系统运行流程

4.2 利用枚举权重算法融合信息

在获取了基于脑电波和人脸表情 2 个分类器给出的情绪得分之后, 通过枚举 2 个单模态分类器输出的线性组合权重, 来找到一个参数 k, 使得两个模态情绪输出的线性组合在训练集上取得最好的表现: 对于分类, 找出最大准确率; 对于回归, 找出真实值和预测值的最小绝对值. 具体来说, 先通过式 (16) 来进行融合输出情绪得分, 并通过式 (17) 输出结果. 问题的关键在于找出合适的 k, 以 0.01 的步长枚举 k, 并且每一次枚举, 都计算融合后的准确率, 选取一个 k, 使得融合后在训练集上准确率最大.

$$S_{\text{emum}}=kS_{\text{face}}+(1-k)S_{\text{EEG}} \tag{16}$$

$$r_{\text{enum}}=\begin{cases}\text{high}, & S_{\text{enum}}\geq 0.5\\ \text{low}, & S_{\text{enum}}<0.5\end{cases} \tag{17}$$

其中, r_{enum}代表量化的情绪分数融合后预测的分类结果 (high 或 low) 而S_{enum}则代表融合后预测的连续值结果, S_{face}和 S_{EEG}分别代表人脸表情和脑电波的输出, 而 k 代表人脸表情的重要程度, 相应地, $(1-k)$ 代表脑电波的重要程度. 我们应用这个方法在两个不同的任务 (效价和唤醒度) 上, 也就是说, 两个任务的 k 是不同的.

4.3 利用自适应增强算法融合信息

对于第二种方法, 我们使用 AdaBoost 技术, 将两个分类器作为 AdaBoost 的子分类器进行融合. 该方法的目标是为每一个子分类器寻找$w_j(j=1,2,\cdots,n)$和获得最终的输出, 如式 (18), 式 (19).

$$S_{\text{boost}}=1\Big/\left(1+\exp\left(-\sum_{j=1}^{n}w_j s_j\right)\right) \tag{18}$$

$$r_{\text{boost}}=\begin{cases}\text{high}, & S_{\text{boost}}\geq 0.5\\ \text{low}, & S_{\text{boost}}<0.5\end{cases} \tag{19}$$

其中, r_{boost}代表自适应增强融合方法的预测的结果 (high 或 low), $s_j\in\{-1,1\}(j=1,2,\cdots,n)$代表对应的子分类器的输出. 比如说, S_1是基于脑电的情绪分类器的输出而S_2代表基于人脸图像的情绪分类器的输出. 而要获取$w_j(j=1,2,\cdots,n)$的方法如下所述: 对于一个含 m 个样本的训练集, 我们先用$s(x_i)_j\in\{-1,1\}$表示第 j 个分类器对于第 i 个样本的输出, 用y_i表示第 i 个样本的真实标签. 我们首先用式 (20) 初始化每个样本的训练权重:

$$\alpha_i=1/m \tag{20}$$

其中, α_i代表第 i 个样本的权重系数. 训练权重体现在训练数据的时候, 如果用到当前数据点, 那么数据点的数据要先乘以这个权重系数. 然后进行子分类器的训练如之前所述, 训练完之后用式 (21) 计算错误率ε_j.

$$\varepsilon_i=\sum_{i=1}^{M}t_i\alpha_i \tag{21}$$

其中, t_i通过式 (22) 确定.

$$t_i=\begin{cases}0, & s(x_i)_j=y_i\\ 1, & s(x_i)_j\neq y_i\end{cases} \tag{22}$$

最终, 用式 (23) 得到需要计算的子分类器权重:

$$w_i=\ln((1-\varepsilon_j)/\varepsilon_j)/2 \tag{23}$$

随后, 还需根据式 (24) 更新每个数据点的权重系数, 用于下一个分类器更加针对性地训练.

$$\alpha_{j+1,i}=\begin{cases}\dfrac{\alpha_{j,i}\exp(-w_j)}{\sum_{i=1}^{m}\alpha_{j,i}\exp(-w_j)}, & s(x_i)_j=y_i\\ \dfrac{\alpha_{j,i}\exp(w_j)}{\sum_{i=1}^{m}\alpha_{j,i}\exp(w_j)}, & s(x_i)_j\neq y_i\end{cases} \tag{24}$$

与枚举权重融合方法相同, 我们应用这个方法在两个不同的任务 (效价和唤醒度) 中, 为两个任务训练出不同的参数.

5 实验设置与结果分析

5.1 离线实验

(1) 被试依赖型模型的离线实验结果

本实验选用的数据集为 DEAP 数据集来验证被试依赖型模型的有效性. 图 6 是 4 种被试依赖型模型在 DEAP 数据集上的表现. 根据实验结果可知, 人脸表情识别的准确率较高, 但在部分被试上仍表现出较低的准确率, 其分别为: 被试 1、被试 3、被试 5、被试 11、被试 12. 各模型的平均最高准确率如表 1 所示.

(2) 被试依赖型模型实验的显著性分析

对被试依赖型模型进行数据的显著性检验: 首先对 4 种方法的结果 (EEG, 脸部图像, 枚举权重融合方法和自适应增强融合方法) 进行正态分布检验 (normality test), 正态分布检测的结果小于 0.05, 因而认为其符合正态分布. 对符合正态分布的数据接着进行 t 方检验, t 方检验的 P 值小于 0.05, 因而可以认为有显著的差异; 而对于不符合正态分布的数据, 则进行 Nemenyi 检验, 其 P 值小于 0.05, 因而也可以认为其有显著差异. 进一步地说, 显著的差异意味着准确率的显著提升. 在 DEAP 数据集的 valence 空间和 arousal 空间中, 各个融合方法之间未体现出显著性差异.

(3) 跨被试情绪识别模型的实验与分析

对于跨被试情绪识别模型—基于 LSTM 识别脑电情绪模型, 利用 MAHNOB-HCI 数据集训练并验证. 该数据集采集自 30 名被试, 此处我们仅使用脑电数据集, 并对该模型进行了如下两组实验: 验证数据与训练数据部分同源实验、验证数据与训练数据完全非同源实验. 记验证数据与训练数据部分同源组为 A 组, 验证数据与训练数据完全非同源实验为 B 组.

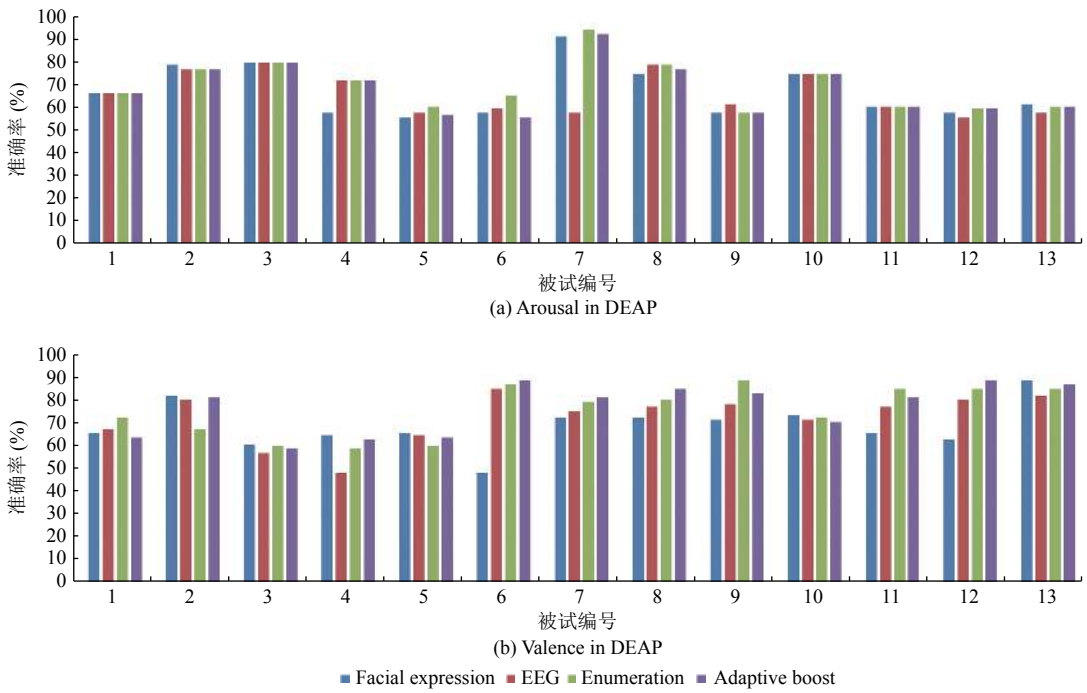

图 6 被试依赖型模型在 DEAP 数据集上的准确率

表 1 被试依赖型模型在 DEAP 数据集上的准确率 (%)

维度	人脸表情识别	脑电情绪识别	枚举权重融合方法	自适应增强融合方法
效价	72.31±12.02	75.38±12.16	80.30±11.37	80.00±12.40
唤醒度	71.15±11.62	68.85±10.02	74.23±10.34	71.54±11.16

当验证数据与训练数据部分同源时, 我们对数据集进行划分: 选取 1 至 23 号被试的数据, 1 号至 20 号被试的数据作为训练集, 21 号至 23 号被试作为验证集. 当已训练的模型预测 1 至 23 号被试的数据时, 模型在 valence 维度的平均准确率为 78.56%, 回召率为 68.18%; 而模型在 arousal 维度的平均准确率为 77.22%, 回召率为 69.28%.

当验证数据与训练数据完全非同源时, 即我们使用被试 1 至 20 号的数据训练模型, 而模型却预测 21 至 30 号的数据. 最终模型在效价维度的平均准确率

为 51.70%, 回召率为 47.13%; 而模型在唤醒度维度的平均准确率为 58.65%, 回召率为 33.62%.

关于跨被试模型情绪识别的损失函数最终值如表 2 所示 (表中 loss 值对应实验组的训练损失函数, val_loss 值代表实验组的验证损失函数); 而关于跨被试模型情绪识别的准确率、回召率和均方根误差 (Root Mean Square Error, RMSE) 如表 3 所示.

表 2 跨被试模型在 MAHBON-HCI 数据集上的损失函数最终值

组别	效价维度		唤醒度维度	
	loss	val_loss	loss	val_loss
A	3.16	3.35	2.71	3.30
B	3.05	6.20	2.66	6.88

表 3 跨被试模型在 MAHBON-HCI 数据集上的情绪识别准确率和回召率

组别	效价维度			唤醒度维度		
	准确率(%)	回召率(%)	RMSE	准确率(%)	回召率(%)	RMSE
A	78.56	68.18	1.83	77.22	69.28	1.82
B	51.70	47.13	2.50	58.65	33.62	2.63

由此可见, 虽然当模型预测非同源数据时准确率和回召率均有下降, 损失函数最终值较高, 但在面对预测连续情绪这种较为复杂多样的情绪的情况下仍能保持超过 50% 的准确率, 情绪识别性能具有一定的稳定性.

(4) 模型比较与分析

情绪识别相关研究有很多. 本文使用两种信息融合算法, 将双模态情绪识别信息融合, 在唤醒度和效价维度平均准确率分别可以达到 74.23% 和 80.30%. 而本文提出的基于脑电的跨被试情绪识别模型, 在使用 MAHNOB-HCI 数据集验证的情况下, 在唤醒度和效价维度最高准确率分别可以达到 77.22% 和 78.56%.

2019 年, Chao 等[21] 基于脑电信号提出了多频段特征矩阵 (Multiband Feature Map, MFM) 和胶囊网络 (Capsule Networks, CapsNet) 模型, 在使用 DEAP 数据集验证的情况下, 该模型在唤醒度和效价维度最高分别能够达到 68.28% 和 66.73% 的准确率. 同年, Huang 等[22] 基于脑电和其他生理信号提出利用集成卷积神经网络 (Ensemble Convolutional Neural Network, ECNN) 识别情绪, 该算法利用 DEAP 数据集进行验证, 对情绪的四分类准确率最高能够达到 82.92%. 2017 年, Yin 等[23] 提出迁移特征递归消除跨被试模型, 在使用 DEAP 数据集验证的情况下, 在唤醒度和效价维度准确率分别达到 78.67% 和 78.75%.

由此可见, 本文提出的被试依赖型模型与其他模型相比同样具有较高的准确率. 本文的跨被试模型与目前已有的跨被试情绪识别模型相比具有相近的准确率.

5.2 在线实验

(1) 实验步骤

图 7 概述了本文实验的工作流程. 一开始使用视频来诱发被试的情绪, 同时记录面部图像和 EEG 信号. 在视频结束时, 要求被试报告他们的效价维度 (valence) 和唤醒度 (arousal) 维度的分数, 也即时情绪状态——模型要预测的目标. 积极程度和唤醒程度的值为 1 到 9 之间的离散值.

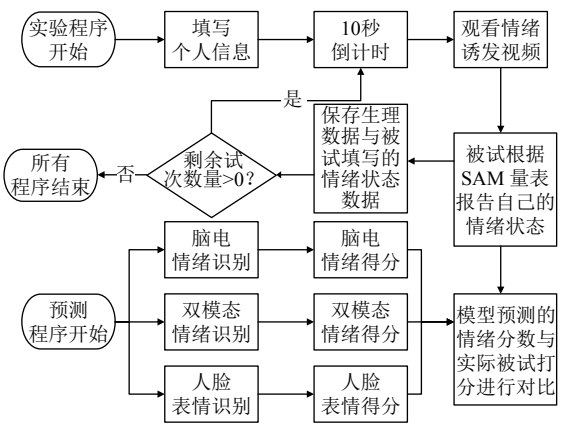

图 7 在线实验流程图

在线实验包含 20 名被试 (50% 男性, 50% 女性), 年龄范围从 7 到 75(平均值=34.15, 标准差=22.14). 实验过程如下, 首先向被试介绍了 valence 和 arousal 的含义, 接着被试观看视频并在每个视频结束时报告他们的情绪指标 (valence 和 arousal). 在实验期间, 被试坐在舒适的椅子上并被指示尽量避免眨眼或移动他们的身体. 期间还进行了设备测试并校正了相机位置, 以确保拍摄对象的面部出现在屏幕中央.

在进行实验之前需要选择用于诱发情绪的材料: 从大量商业电影中手动选择 40 个视频进行剪辑, 再将他们分为 2 部分用于采集训练时展示和采集测试数据中展示. 每个部分包含 20 个视频. 影片剪辑的持续时间为 69.00 到 292.00 s (平均值=204.06, 标准差=50.06).

在进行测试之前, 首先需要数据来训练模型. 因此, 实验首先进行训练数据的收集. 对于每个被试收集 20 组实验的数据. 在每组实验开始时, 屏幕中央都会有 10 秒倒计时, 以吸引被试的注意力, 并作为视频开始的提示. 倒计时结束后, 屏幕上开始播放电影视频用于诱发情绪. 在此期间使用摄像机每秒收集 4 个人脸图像,

并使用 Emotiv Epoc+移动设备每秒收集 10 组 EEG 信号. 每个影片持续 2~3 分钟. 在每组试验结束时, 情绪自评量表 (Self-Assessment Manikins, SAM)[24] 出现在屏幕中央, 以收集被试的 valence 和 arousal 标签. 指示被试填写整个表格并单击“提交”按钮以进行下一个试验. 在两次连续的情绪恢复试验中, 屏幕中央还有 10 秒的倒计时. 收集的数据 (EEG 信号, 面部图像和相应的化合价和唤醒标签) 用于训练上述模型.

在测试阶段, 每个被试进行 20 组实验. 每次实验的过程与训练阶段数据收集的过程类似. 这里使用不同于训练采集数据时的视频对被试进行刺激, 因为相同的视频会引发相同的生理状态从而导致无法判别生理状态是由情绪产生还是由视频产生. 在每次试验结束时, 使用 4 种不同的检测器 (面部表情检测器, EEG 检测器, 枚举权重融合方法和自适应增强融合检测器) 来得到结果. 通过比较预测结果和真实标签来统计准确率.

(2) 实验结果与显著性分析

图 8 展示了 20 个实验对象的测试过程中准确率.

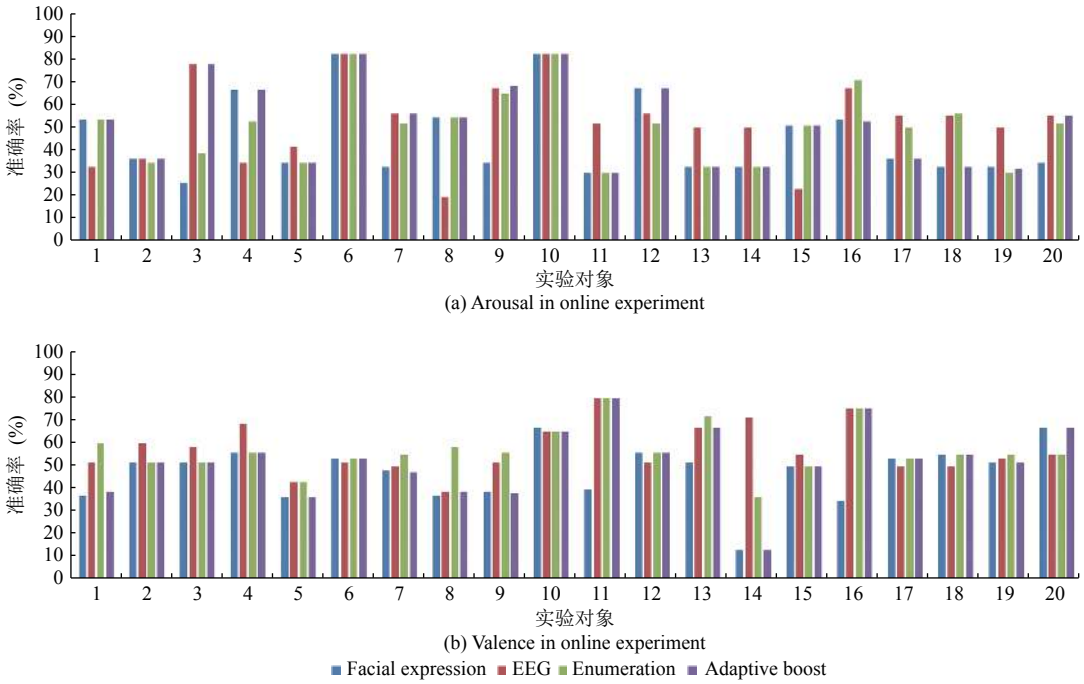

图 8 在线实验不同对象各种方法准确率

表 4 展示了测试过程中各种方法的平均准确率. 可以看到, 除了在线实验中唤醒度维度中枚举权重融合方法相对于脑电情绪识别的准确率, 所有融合方法的准确率都比单一模态高. 并且, 由于在线实验无法进行超参的调整, 使得我们的模型普适性更高. 在线实验中, 我们只针对被试依赖型情绪识别模型进行实验.

表 4 在线实验情绪识别准确率 (%)

维度	人脸表情识别	脑电情绪识别	枚举权重融合方法	自适应增强融合方法
效价	55.75±19.25	69.25±11.96	69.75±12.79	68.00±17.39
唤醒度	54.00±20.28	64.00±22.39	61.75±19.76	70.00±20.51

对于在线实验, 在效价维度中, 枚举权重融合方法相对于脸部图像的结果有显著差异$P = 0.026$. 而且自适应增强融合方法与人脸表情识别方法, 在效价维度和唤醒维度均有显著性差异, 效价维度中的 P 为 0.026 而唤醒度维度中的 P 为 0.007.

5.3 改进情绪识别方法的有效性分析

(1) AdaBoost 融合双模态信息的有效性分析

为了融合双模态决策层信息, 本文提出利用 AdaBoost 算法融合人脸表情识别分类器和脑电情绪识别分类器, 以达到提高双模态情绪识别准确率的效果. 实验表明, AdaBoost 算法的表现优于枚举权重.

AdaBoost 相对于枚举权重算法的优点主要体现在两点: ① 对多组训练数据集赋予不同权值; ② 子分类器权重精度更高.

本文所提的两种融合方法都是根据错误率的降低的思路来找到最优解的. 枚举权重算法仅设置了一定精度的步长 (本文为 0.01), 通过步长的增加, 子分类器

的权重遍历范围为 [0, 1] 之间的数值, 从而找到最低错误率对应的子分类器权重. 而 AdaBoost 算法首先赋予多组训练数据集默认权值, 然后计算分类误差率, 接着通过分类误差率计算子分类器的权重, 最后更新训练数据集的权值分布, 开始下一个分类误差率和子分类器权重的计算. 在这个过程中, AdaBoost 算法要求计算规范化因子, 并结合上一组的训练数据集权值、规范化银子、子分类器系数、ground-truth 标签和子分类器权重, 计算下一组权值分布, 使得该权值成为一个概率分布——对于重要的训练数据集, 权值更高. 这种方法区别于默认数据集为均匀权值分布的枚举权重算法, 更符合实际中训练数据集是非均匀分布的这一情况. 同时, 在计算过程中, 因为没有固定的步长. 由此可得, 子分类器的权重精度高于枚举权重算法.

(2) 整体情绪识别算法复杂度分析

对于模型的算力需求和时间复杂度, 整个算法的计算主要集中在卷积神经网络的部分, 相比于卷积神经网络的参数, SVM 的参数极少. 而实际上, 我们的卷积神经网络共有 831 074 个参数, 使用 GeForce GTX 950 显卡中, 进行一次单样本前向传播的时间是 0.0647 s. 对于基于 LSTM 的脑电情绪识别模型, 我们首先提取了脑电特征值, 然后进行训练, 每个 epoch 训练时间均不超过 3 s.

在模型融合方面, 第一种枚举权重融合方法被许多多模态融合的研究广泛使用, 该种方法比较简单但是其计算损失却随着模态的增多指数上升. 因为第一种融合方法的复杂度为 O(100mn), 其中 m 是样本个数而 n 是模态个数. 而第二种方法 AdaBoost 的时间复杂度却是 O(nm). 也就是说第二种方法随着模态的增多计算损失的增加是线性的, 因此在更多模态的条件下第二种融合方法更加适合.

(3) 改进的情绪识别机制

在情绪识别的基准值方面, 区别于传统的离散情绪识别方法, 本文引入了连续情绪的概念, 利用效价 (valence) 和唤醒度 (arousal) 两个维度的得分量化情绪, 分数为整数, 范围为 [1, 9].

在算法方面, 本文重点介绍了两种方法, 分别用于解决情绪识别的两个难题: 准确率不高、跨被试性能差.

为了提高准确率, 本文结合了人脸表情识别技术和脑电情绪识别技术. 在人脸图像模态, 我们采用端到端的多任务卷积神经网络, 以计算效价和唤醒度得分. 由于在通过被试获取数据集时, 人脸图像往往数据集过少, 因此利用迁移学习技术, 首先用 Fer2013 数据集预训练模型, 然后再用采集的被试的数据微调模型. 在脑电模态, 我们利用了分类效果最好的支持向量机算法, 根据效价和唤醒度得分是否大于 5, 来进行二分类. (若大于 5 则为高分, 否则属于低分). 而为了融合两个模态的数据以进一步提高准确率, 我们探究了枚举权重融合方法和 AdaBoost 方法. 通过在决策层的信息融合提高情绪识别准确率. 实验表明, 在融合更多模态数据的时候, AdaBoost 表现出优于枚举权重算法的性能.

然而, 上述算法依然无法做到一个模型识别多个被试的情绪, 而是针对每个被试训练一组特定的模型, 由此本文称之为被试依赖型模型, 其适用范围不广.

为了提高情绪识别的跨被试性能, 本文在脑电模态提出了基于 LSTM 的跨被试情绪识别方法. 通过构建长短时记忆网络达到一个模型识别多个被试情绪的目的. 实验表明, 该方法具有一定的跨被试性能.

6 总结

本文基于人脸图像和脑电信号提出了多种情绪识别方法. 本文使用情绪的二维模型量化情绪, 根据连续情绪的效价和唤醒度两个维度的得分量化情绪. 在人脸图像模态, 本文利用迁移学习技术训练多任务卷积神经网络以识别人脸表情. 在脑电信号模态, 对于与训练数据同源的数据, 本文采用支持向量机进行情绪识别; 对于非同源数据, 则采用长短时记忆网络. 为了提高情绪识别的准确率, 本文提出使用枚举权重模型和自适应增强模型融合人脸表情模型和脑电情绪模型的决策层信息以提高准确率.

本文进行的实验可验证各情绪识别方法的有效性. 其中跨被试脑电情绪模型在预测非同源数据时准确率仍然高于传统算法, 一定程度上保证了模型的稳定性和有效性. 对于多模态情绪识别来说, 本文的最终实验涵盖了情绪的效价和唤醒度, 即愉悦度和强度. 该量化情绪的指标有效、可行, 且在识别较多种类情绪的情况下依然体现出了较高的准确率. 下一步的工作即针对跨被试型脑电情绪识别模型进行优化, 通过结合其他生理模态信息的方法, 为不同被试源的情绪信息衡量域差异, 并根据域差异来进一步利用迁移学习提升跨被试脑电情绪识别模型的性能.

参考文献

1 Barrett LF. Solving the emotion paradox: Categorization and

the experience of emotion. Personality and Social Psychology Review, 2006, 10(1): 20–46. [doi: 10.1207/s15327957pspr1001_2]
2 Poria S, Cambria E, Bajpai R, *et al*. A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 2017, 37: 98–125. [doi: 10.1016/j.inffus.2017.02.003]
3 Posner J, Russell JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Development and Psychopathology, 2005, 17(3): 715–734. [doi: 10.1017/S0954579405050340]
4 Ekman P, Friesen WV. Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 1971, 17(2): 124–129. [doi: 10.1037/h0030377]
5 叶继华, 祝锦泰, 江爱文, 等. 人脸表情识别综述. 数据采集与处理, 2020, 35(1): 21–34. [doi: 10.16337/j.1004-9037.2020.01.002]
6 Meng D, Cao GT, He ZH, *et al*. Facial expression recognition based on LLENet. Proceedings of 2016 IEEE International Conference on Bioinformatics and Biomedicine. Shenzhen, China. 2016. 1915–1917. [doi: 10.1109/BIBM.2016.7822814]
7 Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000, 290(5500): 2323–2326. [doi: 10.1126/science.290.5500.2323]
8 朱明旱, 罗大庸. 2DFLD 与 LPP 相结合的人脸和表情识别方法. 模式识别与人工智能, 2009, 22(1): 60–63. [doi: 10.16451/j.cnki.issn1003-6059.2009.01.011]
9 Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision. Lake Placid, NY, USA. 2016. 1–10. [doi: 10.1109/WACV.2016.7477450]
10 张冠华, 余旻婧, 陈果, 等. 面向情绪识别的脑电特征研究综述. 中国科学: 信息科学, 2019, 49(9): 1097–1118. [doi: 10.1360/N112018-00337]
11 St. Louis EK, Frey LC, Britton JW, *et al*. ElectroEncephaloGraphy (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants. Chicago: American Epilepsy Society, 2016. [doi: 10.5698/978-0-9979756-0-4]
12 Yazdani A, Lee JS, Ebrahimi T. Implicit emotional tagging of multimedia using EEG signals and brain computer interface. Proceedings of the First SIGMM Workshop on Social Media. Beijing, China. 2009. 81–88. [doi: 10.1145/1631144.1631160]
13 Georgieva O, Milanov S, Georgieva P, *et al*. Learning to decode human emotions from event-related potentials. Neural Computing and Applications, 2015, 26(3): 573–580. [doi: 10.1007/s00521-014-1653-6]
14 郑伟龙, 石振锋, 吕宝粮. 用异质迁移学习构建跨被试脑电情感模型. 计算机学报, 2020, 43(2): 177–189. [doi: 10.11897/SP.J.1016.2020.00177]
15 Viola P, Jones MJ. Robust real-time face detection. International Journal of Computer Vision, 2004, 57(2): 137–154. [doi: 10.1023/B:VISI.0000013087.49260.fb]
16 江伟坚, 郭躬德, 赖智铭. 基于新 Haar-like 特征的 Adaboost 人脸检测算法. 山东大学学报 (工学版), 2014, 44(2): 43–48. [doi: 10.6040/j.issn.1672-3961.1.2013.003]
17 Goodfellow IJ, Erhan D, Carrier PL, *et al*. Challenges in representation learning: A report on three machine learning contests. Neural Networks, 2015, 64: 59–63. [doi: 10.1016/j.neunet.2014.09.005]
18 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84–90. [doi: 10.1145/3065386]
19 Bhatnagar G, Wu QMJ, Raman B. A new fractional random wavelet transform for fingerprint security. IEEE Transactions on Systems, Man, and Cybernetics —Part A: Systems and Humans, 2012, 42(1): 262–275. [doi: 10.1109/TSMCA.2011.2147307]
20 Duan KB, Rajapakse JC, Wang HY, *et al*. Multiple SVM-RFE for gene selection in cancer classification with expression data. IEEE Transactions on Nanobioscience, 2005, 4(3): 228–234. [doi: 10.1109/tnb.2005.853657]
21 Chao H, Dong L, Liu YL, *et al*. Emotion recognition from multiband EEG signals using CapsNet. Sensors, 2019, 19(9): 2212. [doi: 10.3390/s19092212]
22 Huang HP, Hu ZC, Wang WM, *et al*. Multimodal emotion recognition based on ensemble convolutional neural network. IEEE Access, 2020, 8: 3265–3271. [doi: 10.1109/ACCESS.2019.2962085]
23 Yin Z, Wang YX, Liu L, *et al*. Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. Frontiers in Neurorobotics, 2017, 11: 19. [doi: 10.3389/fnbot.2017.00019]
24 Bradley MM, Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry, 1994, 25(1): 49–59. [doi: 10.1016/0005-7916(94)90063-9]