

基于混合神经网络的脑电时空特征情感分类*

陈景霞, 郝为, 张鹏伟, 闵重丹, 李玥辰



(陕西科技大学 电子信息与人工智能学院, 陕西 西安 710021)

通讯作者: 陈景霞, E-mail: chenjx_sust@foxmail.com

摘要: 提出一种新的脑电图 (electroencephalograph, EEG) 数据表示方法, 将一维链式 EEG 向量序列转换成二维网状矩阵序列, 使矩阵结构与 EEG 电极位置的脑区分布相对应, 以此来更好地表示物理上多个相邻电极 EEG 信号之间的空间相关性. 再应用滑动窗将二维矩阵序列分成一个个等长的时间片段, 作为新的融合了 EEG 时空相关性的数据表示. 还提出级联卷积-循环神经网络 (CASC_CNN_LSTM) 与级联卷积-卷积神经网络 (CASC_CNN_CNN) 两种混合深度学习模型, 二者都通过 CNN 卷积神经网络从转换的二维网状 EEG 数据表示中捕获物理上相邻脑电信号之间的空间相关性, 而前者通过 LSTM 循环神经网络学习 EEG 数据流在时序上的依赖关系, 后者则通过 CNN 卷积神经网络挖掘局部时间与空间更深层的相关判别性特征, 从而精确识别脑电信号中包含的情感类别. 在大规模脑电数据集 DEAP 上进行被试内效价维度上两类情感分类实验, 结果显示本文提出的 CASC_CNN_LSTM 和 CASC_CNN_CNN 网络在二维网状 EEG 时空特征上的平均分类准确率分别达到 93.15% 和 92.37%, 均高于基准模型和现有最新方法的性能, 表明该模型有效提高了 EEG 情感识别的准确率和鲁棒性, 可以有应用到的基于 EEG 的情感分类与识别相关应用中.

关键词: EEG; 情感识别; 二维网状; 时空特征; 卷积循环神经网络; 混合模型

中图分类号: TP311

中文引用格式: 陈景霞, 郝为, 张鹏伟, 闵重丹, 李玥辰. 基于二维网状序列与混合神经网络的脑电时空特征学习与情感分类. 软件学报. <http://www.jos.org.cn/1000-9825/6123.htm>

英文引用格式: Chen J. X., Hao W., Zhang P. W., Min C. D., Li Y. C., Spatio-temporal feature learning and emotion classification from 2D mesh-like EEG sequence using hybrid deep neural networks. Ruan Jian Xue Bao/Journal of Software, (in Chinese). <http://www.jos.org.cn/1000-9825/6123.htm>

Emotion classification of spatiotemporal EEG features using hybrid neural networks

CHEN Jing-Xia, HAO Wei, ZHAMG Peng-Wei, MIN Chong-Dan, LI Yue-Chen

(College of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an 710021, China)

Abstract: In this paper, we propose a new data representation of electroencephalogram (EEG), which transforms 1D chain-like EEG vector sequences into 2D mesh-like matrix sequences. The mesh structure of the matrix at each time point corresponds to the distribution of EEG electrodes, which could better represent the spatial correlation of EEG signals among multiple physically adjacent electrodes. Then, the sliding window is used to divide the 2D meshes sequence into segments containing equal time points, and each segment is seen as an EEG sample integrating the temporal and spatial correlation of raw EEG recordings. We also propose two hybrid deep learning models: cascaded convolutional recurrent neural network (CASC_CNN_LSTM) and cascaded double convolutional neural network (CASC_CNN_CNN). Both of them use the CNN model to capture the spatial correlation between physically adjacent EEG signals from

* 基金项目: 国家自然科学基金青年基金(61806118); 陕西科技大学科研启动基金(2020BJ-30)

Foundation item: National Natural Science Foundation of China (61806118). Research Startup Foundation of Shaanxi University of Science and Technology (2020BJ-30)

收稿时间: 2020-04-02; 修改时间: 2020-05-21; 采用时间: 2020-07-20; jos 在线出版时间: 2020-12-02

the converted 2D EEG meshes. The former uses the LSTM model to learn the time dependency of the EEG sequence, and the latter uses another CNN model to extract the deeper discriminative features of local time and space. Extensive binary emotion classification experiments in valence are carried out on a large scale open DEAP dataset (32 subjects, 9,830,400 EEG recordings). The results show that the average classification accuracy of our proposed CASC_CNN_LSTM and CASC_CNN_CNN networks on spatio-temporal 2D mesh-like EEG sequence reaches 93.15% and 92.37%, respectively, which significantly outperform the baseline models and the state-of-the-art methods. It demonstrates that our proposed method effectively improves the accuracy and robustness of EEG emotion classification due to its ability of jointly learning deeper spatio-temporal correlated features using hybrid deep neural network.

Key words: EEG; emotion recognition; 2D mesh-like; spatiotemporal feature; convolutional recurrent neural networks; hybrid model

1 引言

近年来,随着深度学习、人工智能技术的发展,情感识别已经成为人机交互和情感计算领域一个备受关注的研究热点.情感识别是理解一个人所表达的情感类型的过程,开发高效、鲁棒的人类情感识别算法将对人机交互方式产生重大影响^[1].在人机交互应用程序中引入自动情感识别技术,可以显著提高用户体验的质量,带来更多感知情感的计算机交互界面,并且可以实现许多基于情绪感知、情绪调节的控制功能.脑电图(Electroencephalogram, EEG)是一种非侵入性脑成像技术,依赖于头皮的电极来测量大脑的电生理活动,其中包含了大量的生理、心理与病理信息.神经生理学、心理学大量研究发现人类情感的产生及活动与大脑皮层的活动相关度很大,人类不同的认知和情感活动能够诱发产生不同的 EEG 信号,对其进行有效的特征提取与分类可以达到情感脑-机接口控制的目的^[2-3].与人类的面部表情、语音、心电、核磁等生理信号相比,EEG 信号由于其客观性强、不易伪造、采集设备便携、易于操作、成本低等优点,正越来越多地被应用到人类情感识别中^[4].

脑电图信号的处理与识别具有极大的挑战性.首先,EEG 信号具有很低的信噪比,容易受到多种噪声的干扰,例如,敏感的记录设备很容易受到周围环境的干扰,而且肌肉活动、眼球运动或眨眼都可能带来其他不必要的噪音.其次,人们往往只对特定大脑活动相关的 EEG 信号感兴趣,但是却很难从背景中将这个信号分离出来.虽然 EEG 信号在头皮上空间分辨率较低,但它较高的时间分辨率(毫秒级)可以记录缓慢或快速变化的大脑活动动态^[5].因此,为了确定和提取 EEG 信号中与特定大脑活动或情感相关的部分,需要复杂的 EEG 信号分析与处理技术,既要考虑 EEG 信号空间的相关性,也要考虑时间上的相关性.

近年来,深度学习技术由于具有很强的端对端自学习复杂特征表示的能力,避免了大量手工特征提取,在计算机视觉、自动语音识别、自然语言处理和生物信息学等领域的识别任务中取得了巨大的成功^[6-8].卷积神经网络(Convolutional Neural Networks, CNNs)是当前识别图像和视频数据的最佳架构的核心,这主要是因为它们能够学习并提取对输入数据的部分平移和变形具有鲁棒性的特征表示^[9].而循环神经网络(Recurrent Neural Networks, RNNs)和长短时记忆网络(Long Short Term Memory, LSTM)在许多涉及时间序列动力学的应用中表现出了最先进的性能,如文本和语音识别^[7]、信息抽取与语义分析^[10]等.此外,文献^[11]中 Ng 等人将这两种类型的网络相结合用于视频分类,并且取得了较好的效果.

随着便携式脑电采集技术和设备的出现,大规模脑电数据变得更易获取, CNN、RNN 和 LSTM 等深度学习技术也越来越多地应用于脑电图时间序列的分类与识别中. bashivan 等人^[12]针对认知和精神负荷分类实验中采集的 EEG 数据,提出了一种深度递归卷积神经网络 R-CNN,从 EEG 序列中学习那些对于时、空、频三个维度的变化都不敏感的鲁棒性特征,其分类错误率比他们之前采用的支持向量机和随机森林的错误率降低了 50%以上. Ryan 等人^[13]采用基于 LSTM 的深度循环神经网络来解释认知相关脑电信号在时间上依赖性,从而显著提高跨天 EEG 特征的稳定性,其被试内的平均分类准确率达到 93%. Dalin 等人^[14]提出了一种深度卷积循环神经网络模型,通过有效地学习原始 EEG 数据流时空相关特征表示,来准确地识别人类运动意图,在 MI-EEG 数据集上取得了 98.3% 的被试依赖的分类准确率,其性能均高于其他简单分类器和别的基于深度学习的 EEG 分类模型. lawhern 等人^[15]提出了一种没有完全连接层的多层纯卷积神经网络的应用,该方法在基于 P300 的

oddball 识别任务、手指运动任务中运动相关皮层电位识别和运动想象中感官运动节律识别等范式中都取得了目前为止最好的性能.上述方法都为基于 EEG 的情感识别提供了有益的参考.

Alhagry 等人^[16]提出了一种基于 LSTM 长短时记忆结构的深度循环神经网络,在 DEAP 公开数据集上唤醒/效价/喜欢三个情感维度上依赖于被试的平均分类准确率分别达到 85.65%、85.45%和 87.99%,与传统方法相比,该方法具有较高的平均精度.Soleymani 等人^[17]提出利用 LSTM-RNN 和连续条件随机场 (CCRF) 算法,从被试的脑电信号和面部表情即时检测被试观看视频时情感状态的方法,取得了较好的效果.Salama^[18]等人采用一种三维卷积神经网络(3D-CNN),针对多通道脑电图数据进行情绪识别,在效价和唤醒上的识别准确率分别为 87.44%和 88.49%.本文作者也曾采用深度 CNN 模型^[19],基于 DEAP 数据集 EEG 信号时域、频域及其组合特征,端对端自学习这些特征的高层抽象表示,在效价和唤醒度上的分别取得了 88.5%和 86.7%的识别准确率.在文献^[20]中,我们还采用了一种基于注意力机制的多层双向 GRU (Gated Recurrent Unit) 模型对 DEAP 数据集中连续的 EEG 序列信号进行情感识别,实验证明该模型能够有效降低 EEG 序列长时非稳定性对情感识别的影响.上述这些方法都在一定程度上提高了 EEG 情感分类的性能,但是还没有一种模型能够有效地在空间、时间和频域上同步学习脑电信号情感相关的判别性特征,进一步提高基于 EEG 的情感识别的准确率和鲁棒性仍然面临较大的挑战.

本文提出一种新的脑电图 (electroencephalograph, EEG) 数据表示方法,将一维链式 EEG 序列转换成二维网状矩阵序列,使矩阵结构与 EEG 电极位置的脑区分布相对应,以此来表示 EEG 信号物理上多个相邻电极之间的空间特征.再应用滑动窗采样的方法将二维矩阵序列分成一个个等长的时间片段,作为新的融合了 EEG 时空相关性的数据表示.本文还提出级联卷积-循环神经网络(CASC_CNN_LSTM)与级联卷积-卷积神经网络(CASC_CNN_CNN) 两种深度学习模型,二者都通过 CNN 卷积神经网络从转换的二维网状 EEG 数据表示中捕获物理上相邻脑电信号之间的空间相关性,而前者通过 LSTM 循环神经网络学习 EEG 数据流在时序上的依赖关系,后者则通过 CNN 卷积神经网络挖掘局部时间与空间更深层的相关判别性特征,从而精确识别脑电信号中包含的情感类别.

本文提出的方法

二维 EEG网状特征转换

基于 EEG 的情感脑-机接口系统通常使用便携式可穿戴的多通道电极帽采集 EEG 信号,当被试者观看刺激视频时,电极帽上的传感器捕获被试大脑头皮电流的波动情况.本文提出的 EEG 信号采集与特征转换的总体流程如图 1 所示.其中,EEG 电极地图显示了一种 BCI 常用电极帽上的电极位置分布.不同 BCI 系统的脑电记录通道数量不同,其电极的分布也有所不同.EEG 采集系统获取的传感器读数表示一定采样频率下的脑电信号时间序列.通常,时间点 t 采集的原始 EEG 信号用一个一维数据向量 $X_t = [x_t^1, x_t^2, \dots, x_t^n]^T$ 表示,其中 n 表示采集系统的通道总数, x_t^n 表示第 n 个电极通道在第 t 个时间点的读数.对于观测时间段 $[t, t+N-1]$,共有 N 个一维这样的数据向量,每一个向量都包含 n 个元素,对应电极帽上第 n 个电极的读数.

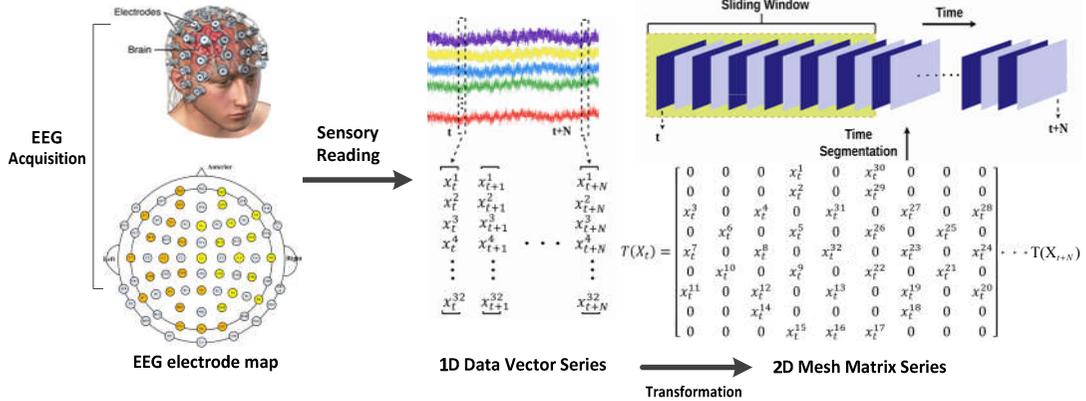


Fig.1 EEG data acquisition and pre-processing process

图 1 脑电数据采集和预处理流程

从 EEG 头皮电极分布图可以看出,每个电极在物理上都与多个电极相邻,用于测量大脑某一区域的 EEG 信号,而且不同的大脑区域对应于不同的大脑活动.上述一维链式 EEG 数据向量仅能表示两个相邻电极位置的相关性.因此,本文根据 EEG 采集系统的电极空间位置关系,将获取的一维原始脑电序列转换为二维网状矩阵序列.本文将 32 通道一维 EEG 数据向量 X_t 转换成如下所示的二维网状矩阵 Y_t :

$$Y_t = T(X_t) = \begin{bmatrix} 0 & 0 & 0 & x_t^1 & 0 & x_t^{30} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_t^2 & 0 & x_t^{29} & 0 & 0 & 0 \\ x_t^3 & 0 & x_t^4 & 0 & x_t^{31} & 0 & x_t^{27} & 0 & x_t^{28} \\ 0 & x_t^6 & 0 & x_t^5 & 0 & x_t^{26} & 0 & x_t^{25} & 0 \\ x_t^7 & 0 & x_t^8 & 0 & x_t^{32} & 0 & x_t^{23} & 0 & x_t^{24} \\ 0 & x_t^{10} & 0 & x_t^9 & 0 & x_t^{22} & 0 & x_t^{21} & 0 \\ x_t^{11} & 0 & x_t^{12} & 0 & x_t^{13} & 0 & x_t^{19} & 0 & x_t^{20} \\ 0 & 0 & x_t^{14} & 0 & 0 & 0 & x_t^{18} & 0 & 0 \\ 0 & 0 & 0 & x_t^{15} & x_t^{16} & x_t^{17} & 0 & 0 & 0 \end{bmatrix}$$

其中, t 表示某一具体的时间点,没有使用的电极位置被置为 0,在神经网络中不起作用.通过转换,观测时间段 $[t, t+N-1]$ 内的一维 EEG 向量序列 $[X_t, X_{t+1}, \dots, X_{t+N-1}]$ 被转换为二维矩阵序列 $[Y_t, Y_{t+1}, \dots, Y_{t+N-1}]$, 其中二维网状矩阵的数量仍然是 N .再通过 Z-score 算法对该二维矩阵中的非零数据进行归一化处理.这样,产生的二维网状矩阵序列既包含了时间信息,又包含了该时间点相关大脑活动的空间信息.然后,如图 1 的最后一步所示,应用滑动窗将二维网状矩阵序列划分为一个个单独的片段 P_j , 作为融合了时空特征的一个 EEG 样本,每个片段具有固定长度(窗口大小),而且相邻两个片段之间不重叠,具体表示为: $P_j = [Y_t, Y_{t+1}, \dots, Y_{t+s-1}]$. 其中, s 表示窗口的大小,即采样点的个数; $j=1, 2, \dots, q$, q 是观测时间段被划分成 EEG 样本片段的个数.本文将进一步提出一种有效的混合深度学习模型,同时学习每个样本的矩阵序列 P_j 在时间和空间上的相关性特征,并预测每个样本片段被试对应的情感类别.

Casc-CNN-LSTM 级联卷积循环神经网络模型

本文首先提出用来学习捕获 EEG 序列的空间和时间特征的 Casc-CNN-LSTM 的模型结构如图 2 所示.模型的输入是经过上述预处理的二维网状矩阵序列(如样本 P_j), 它是一个包含了空间和时间信息的三维数据结构.首先采用 CNN 网络从每一个二维网状矩阵中提取 EEG 数据的空域特征,然后将提取的空间特征序列输入到 LSTM 网络,进一步提取 EEG 数据的时域特征.最后通过一个全连接层接收 LSTM 网络最后一个时间点的输出,得到的特征向量再输入到一个 softmax 层进行最后的情感类别预测.

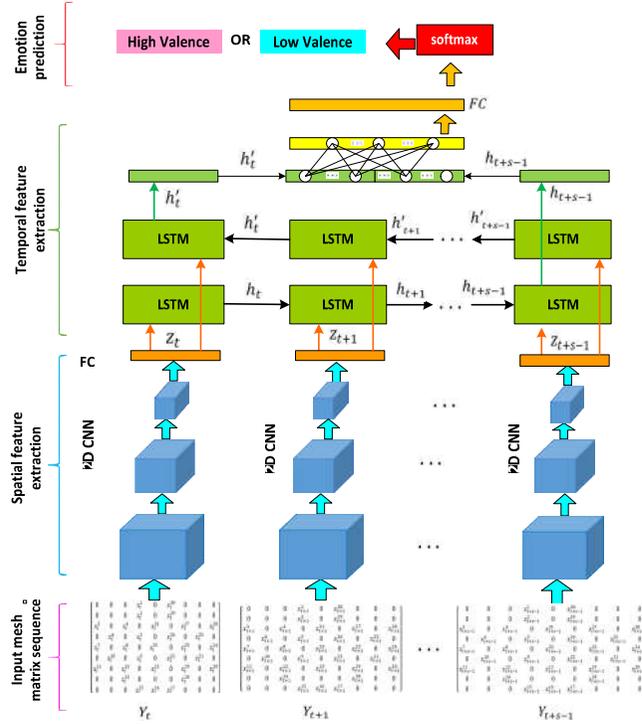


Fig.2 Combination model Casc_CNN_LSTM

图 2 Casc_CNN_LSTM 组合模型

为了提取每一个二维网状矩阵的空间特征,我们采用如图 2 所示的一个深度二维 CNN 网络进行空间特征学习.如上所述,输入模型的第 j 个 EEG 片段用二维网状矩阵序列 $P_j = [Y_t, Y_{t+1}, \dots, Y_{t+s-1}] \in R^{s \times h \times w}$ 表示,其中包含 s 个 $Y_k (k = t, t+1, \dots, t+s-1)$ 表示的二维网状矩阵元素,其大小是 $h \times w$, h 和 w 分别表示电极位置网状矩阵的高和宽.将每一个网状矩阵分别输入到一个 2D-CNN 网络,经学习得到对应的空间特征表示 $Z_k (k = t, t+1, \dots, t+s-1)$: $Z_k = CNN_{2D} Y_k, Z_k \in R^l$. Z_k 是一个包含 l 个元素的一维特征向量.由此,输入的 EEG 矩阵序列被转换成空间特征向量的序列: Cas-CNN: $P_j \rightarrow Q_j$, where $Q_j = [Z_t, Z_{t+1}, \dots, Z_{t+s-1}] \in R^{s \times l}$.该 2D-CNN 模型包含 4 个卷积层,每层分别采用 16, 32, 64, 128 个大小都是 3×3 的卷积核进行非填充的卷积操作,每一层都采用 ReLu 激活函数和 Adam 优化器,学习率取 0.0005.经过学习,第一个卷积层得到 16 个 7×7 的特征图,第二个卷积层得到 32 个 5×5 的特征图,第三个卷积层得到 64 个 3×3 的特征图,第四个卷积层得到 128 个 1×1 的特征图之后应用一个包含 128 个神经元的全连接层将 128 个特征映射转换成最终的空间特征表示 $Z_k \in R^{128}$.在将 2D-CNN 的输出结果输入 RNN 网络之前,这个全连接层是可选的,但是实验发现添加这个全连接层对于帮助模型收敛、提高整个框架的性能起着重要的作用.

通过对模型的精准性和时效性的双重筛选,本文最终使用如图 2 所示的 LSTM 单元构建一个单层双向 RNN 模型,将空间特征表示序列 Q_j 输入到该 RNN 模型进一步计算 EEG 片段时域的相关性特征.LSTM 是为了解决梯度消失和梯度爆炸问题而提出对 RNN 隐含层改进后的神经元结构.在双向 LSTM 中,顺逆序传播可以分别提取某一节点 EEG 与其前后片段的的关系,从而使该模型提取的特征更加客观和准确.

该模型的每个传播方向上都包含 s 个 LSTM 单元,LSTM 单元在当前时间点 t 处的隐状态用 h_t 表示,则 h_{t-1} 表示前一个时间点 $t-1$ 处的隐状态,同一层前一个时间点的信息被传到当前时间点,以此类推影响到最后的输出.本文使用 LSTM 单元的隐状态作为其输出,正向 LSTM 单元输出的隐状态序列 $[h_t, h_{t+1}, \dots, h_{t+s-1}]$,逆向

LSTM 单元输出的隐状态序列 $[h'_{t+s-1}, \dots, h'_{t+1}, h'_t]$. 由于我们感兴趣的是大脑在整个样本时间段内的情感类别, 所以分别从两个方向上取 LSTM 学习整个窗口内所有时间点后提取的最后一个时间点的输出 h_{t+s-1} 和 h'_t . 将 h_{t+s-1} 和 h'_t 在描述特征向量大小的维度上进行拼接, 记为 H_j . 而 H_j 作为整个 RNN 网络学习到的时序特征送入下一个全连接层, 如图 2 的最后阶段所示. EEG 网状矩阵序列 P_j 的时域特征 H_j 表示为: Cas-RNN: $h_{t+s-1}(h'_t) = RNN_{lstm}(Z_j), H_j \in R^{2i}$, 其中, i 表示一个 LSTM 单元的隐状态的大小. 最后, 在全连接层之后采用一个 Softmax 层产生最终每一类情感的概率预测值: Softmax: $C_j = \text{SoftMax}(H_j), C_j \in R^k$, 其中 k 表示模型最终要识别的情感类别数.

总的来说, 我们的框架首先将观测时间段内的 EEG 数据流转换、切分成二维的网状矩阵序列片段, 然后对每一个片段进行 k 类情感分类. 每一个片段 P_j 包含 s 个已经被转换的二维网状矩阵 $[Y_t, Y_{t+1}, \dots, Y_{t+s-1}]$. 在每一个片段内部, 应用一个 3 层 2D-CNN 网络对其中每一个网状矩阵提取其空间特征, 得到空间特征序列 $[z_t, \dots, z_{t+s-1}]$, 随后应用 RNN 网络对整个空间特征序列进行学习进而提取时域特征 H_j , 最后使用 Softmax 分类器计算每一个片段 k 种情感的概率. 原始 EEG 记录经由 Cas-CNN-RNN 模型处理的过程可以描述为: Input($s \times n$)-Trans($s \times h \times w$)-Conv($s \times h \times w \times 16$)-Conv($s \times h \times w \times 32$)-Conv($s \times h \times w \times 64$)-Conv($s \times h \times w \times 128$)-FC(l)-LSTM($s \times 2i$)-FC(l)-Softmax(k). 其中, Input($s \times n$) 表示分段大小为 s 且包含 n 个通道数据的一维原始 EEG 记录序列; Trans($s \times h \times w$) 表示将一维 EEG 记录序列转换为长度为 s 大小为 $h \times w$ 的网状矩阵序列; Conv($s \times h \times w \times m$) 表示一个卷积层从一个网状矩阵学习得到 m 个特征映射; FC(l) 表示具有 l 个神经元的全连接层; LSTM($s \times 2i$) 表示顺逆两个传播方向分别具有 s 个 LSTM 单元的隐层学习得到大小为 i 的隐状态; Softmax(k) 表示用于预测 k 个情感类别的 Softmax 层.

在级联结构的 2D-CNN 模型中, 卷积层后面都没有使用池化操作. 虽然在 CNN 体系结构中, 一个卷积层通常与一个池化层成对使用, 但池化操作并不一定是必须的, 池化操作是以牺牲一部分信息为代价来进行数据降维. 然而, 在 EEG 数据分析问题中, 数据的维度比计算机视觉研究中使用的数据维度要小得多, 为了不丢失有用的 EEG 信息, 本文没有使用池化操作, 而是直接将 4 个卷积层进行了连接.

Casc-CNN-CNN 级联双卷积神经网络模型

本文提出用来挖掘更深层局部时空特征的 Casc-CNN-CNN 模型结构如图 3 所示. 模型的输入同 Casc-CNN-RNN 的输入一样, 都是包含了空间和时间信息的三维数据结构. 模型采用 CNN 网络从每一个二维网状矩阵中提取 EEG 数据的空域特征, 然后将提取的空间特征按照时间顺序重新排列, 再次输入一个 CNN 网络继续提取 EEG 数据深层局部时空特征, 最后通过一个全连接层接收 CNN 网络的输出, 得到的特征向量再输入到一个 softmax 层进行最后的情感类别预测.

Casc_CNN_CNN 与级联模型 Casc_CNN_LSTM 中的 CNN 部分相同, 将二维网状矩阵序列 P_j 输入进一个由四个卷积层与一个全连接层组成的 CNN 网络中, 并转换成具有空间特征向量的序列 Q_j , where $Q_j = [Z_t, Z_{t+1}, \dots, Z_{t+s-1}] \in R^{s \times l}$. 其中 $Z_k (k = t, t+1, \dots, t+s-1): Z_k = CNN_{2D}(Y_k), Z_k \in R^l$ 是一个包含 l 个元素的经过 CNN 学习后的一维空间特征向量. 将 Q_j 整合成形如 $s \times l$ 的矩阵后再次使用 CNN (为区别于前文 CNN, 下将该网络称作 CNN II) 网络提取原始 EEG 的进阶空间特征. 式中 s 表示向量序列的长度, l 表示每个序列所包含的元素个数, CNN II 由两个卷积层、两个池化层、一个全连接层组成, 两个池化层分别连接在两个卷积层后, 将经过第二个池化层池化过的数据进行向量化处理后接入一个神经元数为 512 的全连接层, 最终在全连接后接入 Softmax 层产生最终每一类情感的概率预测值. 两个卷积层分别采用 32 和 64 个 3×3 大小的卷积核进行有填充的卷积操作. 两个池化层都采用 2×2 大小且步长为 2 的 Maxpooling 过滤器对卷积结果进行下采样处理. 卷积层和全连接层采用 ReLu 激活函数以及 Adam 优化器, 学习率取 0.0001.

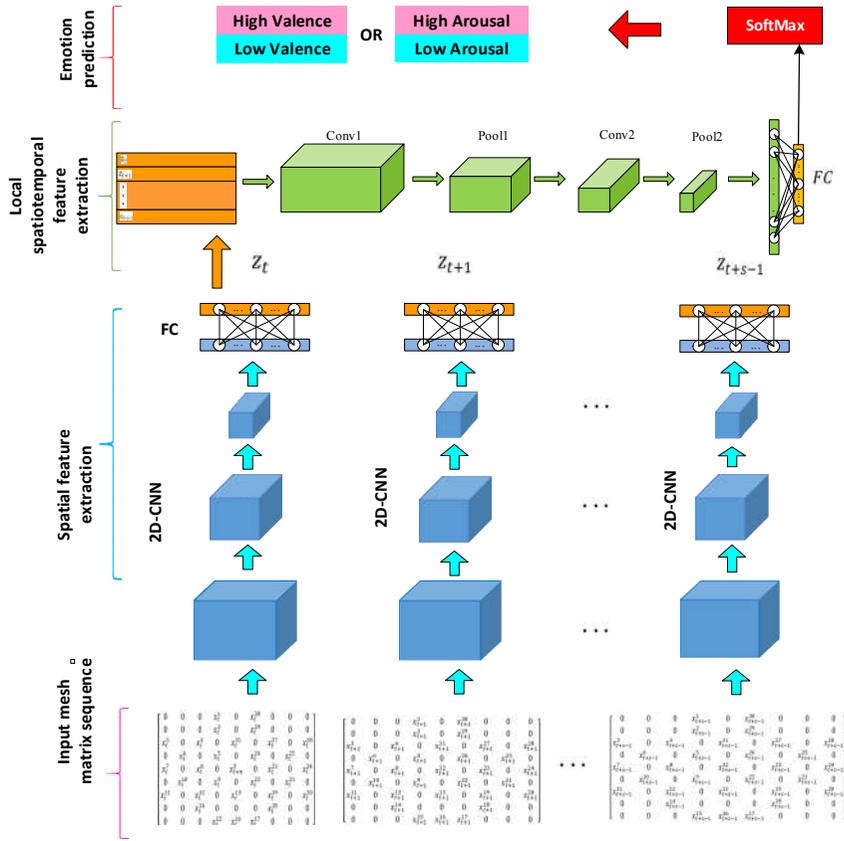


Fig.3 Combination model Casc_CNN_CNN

图3 Casc_CNN_CNN 组合模型

原始 EEG 记录经由 Cas-CNN-RNN 模型处理的过程可以描述为: $\text{Input}(s \times n) \rightarrow \text{Trans}(s \times h \times w) \rightarrow \text{Conv}(s \times h \times w \times 16) \rightarrow \text{Conv}(s \times h \times w \times 32) \rightarrow \text{Conv}(s \times h \times w \times 64) \rightarrow \text{Conv}(s \times h \times w \times 128) \rightarrow \text{FC}(l) \rightarrow \text{Cat}(b) \rightarrow \text{Trans}(s \times h \times w) \rightarrow \text{Conv}(s \times h \times w \times 32) \rightarrow \text{Pooling}(\max, 2) \rightarrow \text{Conv}(s \times h \times w \times 64) \rightarrow \text{Pooling}(\max, 2) \rightarrow \text{FC}(l) \rightarrow \text{softmax}(k)$. 其中, $\text{Input}(s \times n)$ 表示分段大小为 s 且包含 n 个通道数据的一维原始 EEG 记录序列; $\text{Trans}(s \times h \times w)$ 表示将一维 EEG 记录序列转换为长度为 s 大小为 $h \times w$ 的网状矩阵序列; $\text{Conv}(s \times h \times w \times m)$ 表示一个卷积层从一个网状矩阵学习得到 m 个特征映射; $\text{Cat}(b)$ 表示将 b 个向量按时间顺序连接; $\text{Pool}(\max, 2)$ 表示使用步长为 2, 核为 2×2 的最大池化层; $\text{FC}(l)$ 表示具有 l 个神经元的全连接层; $\text{Softmax}(k)$ 表示用于预测 k 个情感类别的 softmax 层.

实验及结果分析

首先介绍实验采用的 DEAP 数据集及 EEG 数据预处理与特征提取的方法, 然后在效价维度上进行被试内两类情感的分类实验, 以 BT、SVM 浅层分类器以及深度 CNN 模型在 1 秒 EEG 片段上的分类性能为基准, 同 Cas-CNN-RNN 和 Cas-CNN-CNN 模型的性能进行比较与分析. 同时, 通过对比实验研究了时空信息对 EEG 情感分类性能的影响.

数据集

本文基于公开的大规模 EEG 情感数据集 DEAP^[21]验证所提出级联和并行深度卷积循环神经网络模型的有效性.该数据集是由伦敦玛丽皇后大学的相关研究团队开发的包含 EEG 等多种生理信号的大规模开源数据集,它记录了 32 个被试者观看 40 个时长约为 1 分钟的、带有不同情感倾向的音乐视频所诱发的脑电、心电、肌电等生理信号,之后被试对所观看的视频在唤醒度,效价,喜好,优势度和熟悉度方面使用 1-9 的连续数值进行评价,评价值由小到大分别表示各项指标由负到正或由弱到强.40 个刺激视频中包含 20 个高效价/唤醒度刺激和 20 个低效价/唤醒度刺激.本文取其中 32 通道的 EEG 信号,采样频率降至 128Hz,为了消除直流噪声、电源噪声和其他伪迹,使用 4-45Hz 的带通滤波器进行数据过滤,再采用盲源分离技术去除了眼电干扰,得到总时长为 63 秒的 EEG 信号,包括观看视频的 60 秒和观看前静息态的 3 秒.

数据预处理与特征提取

本文提取了每个视频诱发的 60 秒的 EEG 序列做进一步分析.为了校正与刺激无关的信号随时间的变化,将观看视频之前 3 秒的 EEG 信号作为基线,从 60 秒试验信号中去除基线,得到刺激相关的序列变化.以 1 秒为窗长对每个序列进行不重叠分段,每次试验得到 60 个片段,则每个被试 40 次试验的 EEG 片段(也称样本)总数为 $40 \times 60 = 2400$ 个,每个片段包含 128 个采样点(即窗口大小为 128),每个采样点包含 32 个 EEG 通道的数据,简称 RAW 特征,其维度大小表示为 $2400 \times 128 \times 32$.实验观察发现,使用大于或小于 128 的窗口分段都会降低组合模型的性能,这也许与人类大脑情感活动的周期长度有关.再对分段以后的 EEG 数据按通道进行归一化,得到每个被试时域上的 NORM 特征.

神经科学和心理学的研究表明,EEG 信号在 delta (1Hz ~ 4Hz),theta (4Hz ~ 8Hz),alpha (8Hz ~ 13Hz),beta(13Hz ~ 30Hz) 和 gamma(30Hz 以上) 5 个频段上包含大量与情绪等心理活动密切相关的节律信息^[22].作者之前的研究^[9]也发现,EEG 信号在 4-45Hz 频带上的功率谱密度(power spectral density,PSD)特征在 2D-CNN 网络上的分类性能明显优于时域上的 RAW 特征和 NORM 特征,这也许是因为 PSD 频域特征的提取需要在较大连续采样周期内进行,而情感脑电任务也是一种相对长时的大脑活动,所以频域特征能够捕获更多情感相关的大脑动态.因此,本文在 NORM 特征基础上,在 4-45Hz 频带上利用快速傅立叶算法在 1s 的 EEG 片段的每个通道上使用 0.5s 的 Hamming 窗无重叠地滑动提取 64 个 PSD 特征,每个被试 40 次试验共提取 PSD 特征的维度是 $2400 \times 64 \times 32$.

接下来处理每个 EEG 样本的标签,基于被试对每个视频在 1-9 范围内的评价值,以中位数 5 作为阈值将效价和唤醒度上的评价值划分为两类,在某个维度上解决 2 分类问题时,大于 5 代表高类或阳性指标,用 1 表示;小于或等于 5 代表低类或阴性指标,用 0 表示.再对数据和标签做均衡化处理,使两类中每一类 EEG 数据和标签数量相同.

接着,根据 DEAP 数据集的脑电电极分布,按照公式(1)将每个 EEG 样本的一维链式向量序列转换成大小为 9×9 的二维网状矩阵序列,转换以后 EEG 样本包含了丰富的空间、时间与频域上的信息,数据和标签的维度如表 1 所示.然后,使用每个被试的所有 EEG 样本构建被试内交叉验证集,即每次随机从 2 类各取 80% 的 EEG 数据和标签来训练模型,其余 20% 的样本来验证训练模型,以此方法为每个被试构建 40 重交叉验证数据集.

Table 1 Formats of segmented 1D and 2D EEG data and labels of each subject

表 1 每个被试分段以后 1D 与 2D EEG 数据及标签的格式

特征描述	特征类型	窗长	EEG 数据形状	标签形状
1D chain-like	RAW/NORM	1-s	$2400(\text{epochs}) \times 128(\text{points}) \times 32(\text{channels}) \times 1$	2400×1
1D chain-like	PSD	1-s	$2400(\text{epochs}) \times 64(\text{points}) \times 32(\text{channels}) \times 1$	2400×1
2D mesh-like	RAW/NORM	1-s	$2400(\text{epochs}) \times 128(\text{points}) \times 9(\text{width}) \times 9(\text{height})$	2400×1
2D mesh-like	PSD	1-s	$2400(\text{epochs}) \times 64(\text{points}) \times 9(\text{width}) \times 9(\text{height})$	2400×1

基准模型

接下来简要介绍 EEG 情感识别方面目前比较先进的方法和基准模型,所有这些模型都基于与本文相同的数据集 DEAP.

Alhagry 等人^[16]采用一种基于 LSTM 长短时记忆结构的深度循环神经网络,在 DEAP 数据集的唤醒/效价/喜欢三个维度上进行了被试内两类情感分类,三个维度上二分类的平均分类准确率分别达到 85.65%、85.45% 和 87.99%.

本文作者在文献^[19]中提出使用浅层机器学习模型 BT (Bagging Tree) 对 DEAP 数据集 EEG 信号的 PSD 特征在效价上和唤醒度上进行被试内两类情感分类的准确率分别达到 86.31%和 86.18%.

Salama^[18]等人采用一种三维卷积神经网络(3D-CNN),针对多通道脑电图数据进行情感识别,在效价和唤醒度上的被试内两类情感识别准确率分别为 87.44%和 88.49%.

本文作者也曾采用深度 2D-CNN 模型,基于 DEAP 数据集 EEG 的时域与频域组合特征,端对端自学习这些特征的高层抽象表示,在效价和唤醒度上进行被试内两类情感识别的准确率分别达到 88.53%和 86.67%^[19].

除了上述最新的方法外,为了增加实验结果的可信度,本文作者对文献 17 中提出的 2D-CNN 模型加以改进,输入时将频组合特征变为分段的链式 PSD 特征,深度学习网络的结构以及超参数配置则同 Casc_CNN_CNN 中的第二个 CNN 一致,都是两组卷积池化层后接一个具有 512 个神经元的全连接层和一个 Softmax 输出层.其中两个卷积层分别采用 3*3 大小的 32 和 64 个卷积核进行有填充的卷积操作.池化层依然采用步长为 2 的 2*2 大小的过滤器对卷积结果进行降维处理.

为了深入探究混合模型与单一模型在分类性能上的优劣,本文还设计了用于对比实验的 BiLSTM(64)模型.该模型也采用分段的链式 PSD 特征作为输入,网络结构同 Casc_CNN_LSTM 中的 LSTM 网络一致,采用双向 LSTM 单元进行时域特征提取.其中正向(顺序)、逆向(倒序) LSTM 单元的隐单元个数(隐状态长度)均为 64,将双向网络各自传播尽头的两个单元的隐状态在特征个数维度上进行连接后接入一个具有 128 个神经元的全连接神经网络,最终接入 Softmax 层产生预测值.

本文研究的神经网络均采用 TensorFlow 框架,并基于 Nvidia Titan X Pascal GPU 以完全监督的方式从头开始训练实现.采用基于 Adam 更新规则的随机梯度下降法最小化模型的交叉熵损失函数.

实验结果

这一部分将展示本文所提出的级联组合模型的总体性能并系统分析 EEG 空间和时间上下文信息对实验结果的影响.并通过 NORM 特征和 PSD 特征在效价和唤醒度两个维度上二分类计算的精度对本文提出的模型以及特征表示方法进行深入的探究.

效价维度上的实验结果与分析

本文提出的级联组合模型以及各种基准模型的整体实验性能如表 2 所示.观察发现以二维网状 PSD 特征为输入的 Casc_CNN_LSTM 和 Casc_CNN_CNN 分别取得了 93.15 和 92.37%的准确率,两种组合模型均优于几种基准模型和最新方法的分类性能.这表明上下文空间和时间信息对于提升 EEG 情感分类识别性能非常关键.实验中,我们还尝试使用了较大或较小的滑动窗进行 EEG 数据分段,结果都会导致组合模型性能明显下降.在不同范围内通过改变滑动窗口大小,可以使模型适应不同类型的 EEG 信号分类,体现该方法具有较好的灵活性.而且,与以往的研究相比,该模型对原始数据的预处理较少,更适合于 BCI 等实时应用.

Table 2 Comparison of the accuracy of two types of emotion classification in the valence dimension between the benchmark model and the cascade hybrid model

表 2 基准模型与级联混合模型在效价上进行两类情感分类准确率的比较

Models	Input Features	Valence Acc
Alhagry ^[16] LSTM	1D Raw chain sequence	0.8545
Chen ^[19] BT	1D PSD chain sequence	0.8631

Salama ^[18] 3D-CNN	Segmented Raw chain sequence	0.8744
Chen ^[19] 2D-CNN	SegmentedNorm+PSD chain sequence	0.8853
2D-CNN-V2	Segmented PSD chain sequence	0.9016
BILSTM(64)	Segmented PSD chain sequence	0.8981
Casc-CNN-CNN	Segmented 2D PSD mesh sequence	0.9237
Casc-CNN-LSTM	Segmented 2D PSD mesh sequence	0.9315

除此以外,我们发现基于文献 19 中 2D-CNN 模型改进的 2D-CNN-V2 模型,分类准确率相较之前高出了 1.63%,原因一方面是输入特征的变化,正如章节 3.2 中所述,由于频域特征能够捕获更多情感相关的大脑动态,PSD 特征相较于 NORM 特征在情感分类任务中的准确率更高,而 PSD 与 NORM 的混合特征虽然扩充了数据量并在一定程度上提升了模型的泛化能力,但对于情感分类任务而言,混合特征的准确率低于单一的 PSD 特征.另一方面是由于改进后的卷积神经网络更加复杂,卷积层中卷积核的个数由 4 个上升为 32 个,提取了更多的空间信息,从而提升了模型的性能.

以 PSD 特征为输入的 BILSTM(64)模型的性能达到了 89.81%,低出 2D-CNN-V2 模型仅有 0.35% 的差距.这说明无论是在时域还是空域上,针对 DEAP 数据集,单一模型在不同维度上的情感分类任务精度相差不大.而 2D-CNN-V2 模型相较于 Casc-CNN-CNN 模型准确率降低了 2.21%,BiLSTM(64)模型相较于 Casc-CNN-LSTM 模型准确率降低了 3.34%,这说明单一模型在不同维度上的表现都不如混合模型.

与文献[16]、[18]、[19]中最新的方法相比,本文提出的 Casc-CNN-CNN 模型比 Alhagr^[16]等人提出的 LSTM 模型的性能高出 6.92%,比 Chen^[19] 等人提出的 BT 浅层分类器的性能高出 6.06%,比 Salama^[18]等人提出的 3D-CNN 模型和 Chen^[19] 等人提出 2D-CNN 模型的性能分别高出 4.93%和 3.84%.Casc-CNN-LSTM 模型也比 Alhagr^[16]等人提出的 LSTM 模型的性能高出 7.7%,比 Chen^[19] 等人提出的 BT 浅层分类器的性能高出了 6.84%,比 Salama^[18]等人提出的 3D-CNN 模型和 Chen^[19] 等人提出 2D-CNN 模型的性能分别高出 5.71%和 4.62%,性能提升都比较显著.这表明本文提出的二维网状 PSD 特征相较于一维链式特征包含了更多的时空上下文信息,而且本文提出的两种混合深度神经网络相较浅层机器学习算法和单一神经网络模型在深层时空特征联合学习上具有明显的优越性.

本文提出的以二维网状 PSD 特征为输入 Casc_CNN_CNN 模型的测试集平均分类准确率达到 92.37%,32 名被试者详细的预测精度如图 4 中所示.

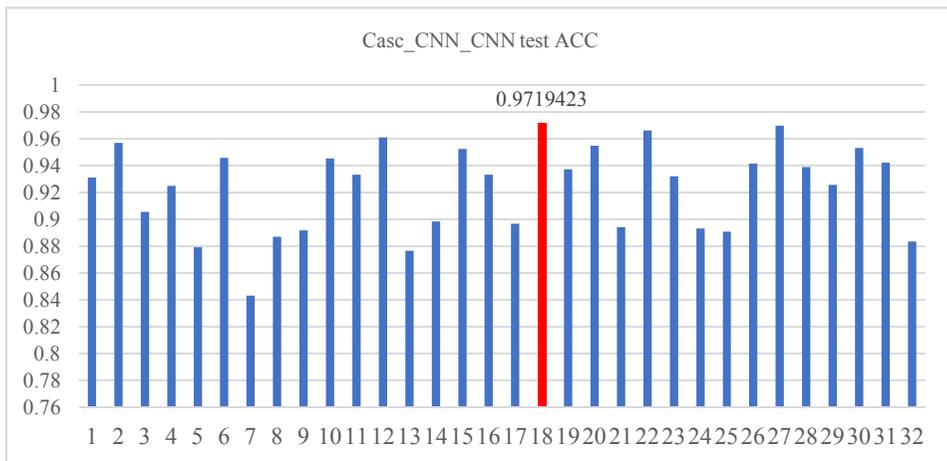


Fig.4 Statistics chart of Casc_CNN_CNN model test set classification accuracy results

图 4 Casc_CNN_CNN 模型测试集分类精度结果统计图

第 18 名被试者的分类准确率达到 97.19%,为所有被试者中最高,图 5 为其训练过程曲线,蓝线

表示训练平均误差 loss,红线表示训练准确率 acc.

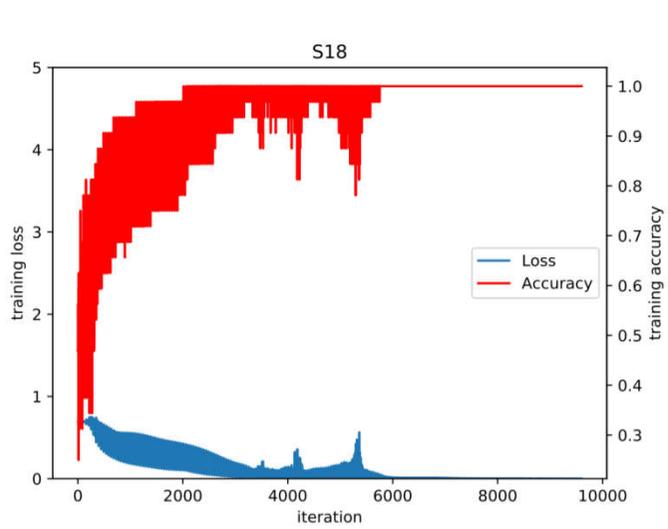


Fig.5 Casc_CNN_CNN model sub18 training process diagram

图 5 Casc_CNN_CNN 模型 sub18 训练过程示意图

通过观察上图可以得知训练过程中,随着迭代轮数 iteration 的增加,平均误差 loss 虽然出现过三次骤然上升后又急速下降的情况,但整体依然呈现下降并不断向零趋近的态势; 训练准确率 acc 整体上向 1 趋近,最终在 0.99 附近达到收敛. Iteration 从 0 增至 3500 期间, loss 初始值从 0.7 附近开始随着 iteration 的变大而以螺旋梯度下降的方式不断向 0 逼近. 而 acc 以将近每 iteration 增加 1.86×10^{-4} 的速度, 以螺旋梯度上升的方式从 0.3 升至 0.95. 在这一过程中, acc 的增速在 iteration 等于 1800~2500 左右出现过三次幅度较大的减小. 随着 acc 曲线斜率的降低, loss 值虽然不断减小, 但其下降趋势也逐步放缓, 笔者分析认为该现象的出现有两个原因, 一是由于超参数 batch_size 设置偏小, 在模型较为复杂的情况下, 每轮迭代需要更新的参数较多, 如果输入数据不足则会导致 loss 函数震荡而不收敛的情况. 二是由于超参数 learning rate 设置的偏大, 虽然在训练开始时模型的性能得到了快速的提升, 但随着 iteration 的增加, 模型难以快速找到最优解从而导致了 loss 震荡而不收敛情况的出现. 而当迭代至 3600、4200、5500 轮左右时, 模型都出现了 loss 突然增大和 acc 突然减小的情况, 但作者认为这是由于学习过程中的梯度并不是目标函数最优方向所导致的正常现象. 而随着 iteration 的继续增加, 两条曲线都趋于平稳, 斜率也逐渐收敛至 0, 最终作者以 acc 在 4000 轮内无增长作为模型终止训练的条件. 训练终止后将测试集数据放进该模型中进行分类预测, 记录预测结果并绘制成如图 4 所示柱状图.

本文提出的以二维网状 PSD 特征为输入的 Casc_CNN_LSTM 模型的测试集平均分类准确率达到 93.15%, 32 名被试者详细的预测精度如图 6 所示.

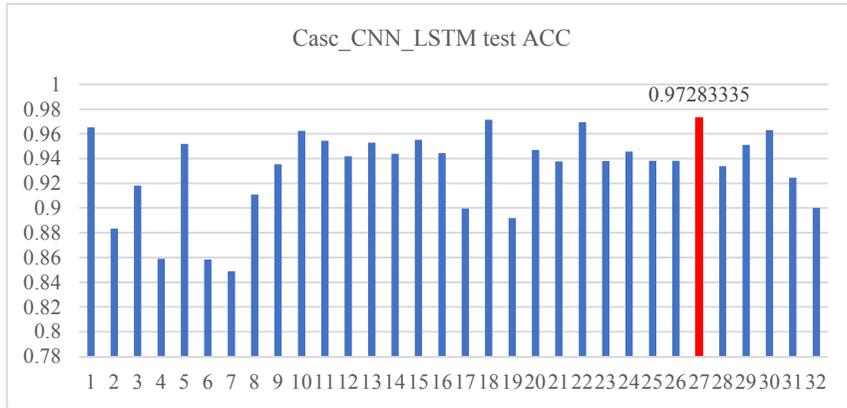


Fig.6 Statistics chart of Casc_CNN_LSTM model test set classification accuracy results

图 6 Casc_CNN_LSTM 模型测试集分类精度结果统计图

观察图 4 和图 6 不难发现,在 Casc_CNN_CNN 和 Casc_CNN_LSTM 两种模型上,效价情感分类精度的前三名是相同的 18、22、27 三名被试.可见两种混合模型性能总体上具有一致性和稳定性,也表明该三个被试的二维网状 PSD 特征中被挖掘到的情感相关的空间、时间判别性信息较少.第 27 名被试者的分类准确率达到 97.28%,为所有被试者中最高,图 7 为其训练过程曲线,蓝线表示训练平均误差 loss,红线表示训练准确率 acc.

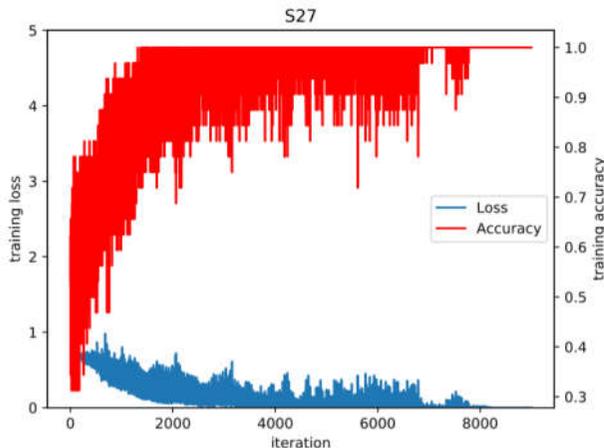


Fig.7 Casc_CNN_LSTM model sub27 training process diagram

图 7 Casc_CNN_LSTM 模型 sub27 训练过程示意图

通过观察上图可以得知训练过程中平均误差 loss 在逐步降低并收敛的同时,训练准确率 acc 也逐步收敛于 1.迭代次数 Iteration 从 0 至 3500 期间,loss 呈现螺旋梯度下降的态势,该时期内的 acc 以螺旋梯度上升的态势从 0.32 提升至 0.94 附近.而当 iteration 从 3500 增至 7000 期间,loss 和 acc 均在一定幅度内发生震荡.但在迭代 7000 轮之后,acc 和 loss 虽依然有所波动,但相较之前震幅明显变小最终 loss 曲线逐渐收敛于 0,acc 曲线逐渐收敛于 1.iteration 从 0 增至 3500 期间,acc 增幅为 0.62,增速约为每 iteration 增长 1.774×10^{-4} ,而 iteration 从 3500 增至 7000 的震荡期,acc 增幅为仅为 0.05,增速约为每 iteration 增长 1.429×10^{-5} ,衰减达到了 91.945%.这也是由于超参数 learning rate 设置偏大,模型的梯度难以按照预想的方式下降,而最终表现为 loss 曲线的震荡.训练终止后将测试集数据放进该模型中进行分类预测,预测结果如图 6 所示.

为了进一步探究二维网状 EEG 时空特征表示方法的普适性以及 CASC_CNN_CNN 和 CASC_CNN_LSTM 模型的通用性,作者以变换特征类别、变换 EEG 维度为实验方法,以模型进行二分类计算的准确率为衡量标准,对本文所提出的特征表示方法以及模型的性能进行研究.为了保证对比实验的客观性和有效性,除 PSD 特征之外,作者还选择了应用范围广泛算法原理清晰的 NORM 特征(提取方法如 3.2 中所述,数据格式如表 1 所示)在效价和唤醒度两个维度上进行二分类计算.每个特征在每个维度上分别使用 2D-CNN-V2、BILSTM(64)、CASC-CNN-CNN、CASC-CNN-LSTM 四种模型进行情感分类计算.效价维度上的计算结果如表 3 所示.

Table 3 Comparison of the accuracy of NORM features between two types of emotion classification in the valence dimension

表 3 效价维度上 NORM 和 PSD 特征两类情感分类准确率的比较

Models	Input Features	Valence Acc
2D-CNN-V2	Segmented NORM chain sequence	0.552
BILSTM(64)	Segmented NORM chain sequence	0.507
Casc-CNN-CNN	Segmented 2D NORM mesh sequence	0.6239
Casc-CNN-LSTM	Segmented 2D NORM mesh sequence	0.5521

在效价维度上,以 2D-NORM 特征作为输入的 CASC-CNN-CNN、CASC-CNN-LSTM 模型分类精度为 62.39%、55.21%,比以链式 NORM 特征作为输入的 2D-CNN-V2、BILSTM(64)模型的 55.2%、50.7%分别高出 7.19%和 4.51%.效价维度上的 NORM 特征经过维度转换后,分类精度平均提高了 5.85%.而当两个级联模型以 2D-PSD 特征作为输入时,分类准确率分别达到了 92.37%和 93.15%,比以链式 PSD 特征作为输入的 2D-CNN-V2、BILSTM(64)模型的 90.16%、89.81%分别高出 2.21%和 3.34%.效价维度上的 PSD 特征经过维度转换后,分类精度平均提高了 2.775%.效价维度上的 NORM 特征在两类情感计算任务中四种模型的平均分类精度仅有 55.875%,而 PSD 特征四种模型的平均分类精度则达到了 91.37%.

就平均分类精度而言,PSD 特征相较 NORM 特征高出了 35.495%.但就特征结构的改变对特征可分性的影响而言,NORM 特征则比 PSD 特征高出 3.075%.该组实验可以证明级联模型比单一模型分类能力更强,二维网状特征比一维链式特征的可分性更高.并且就特征的维度转换对于特征数据可分性的提升而言,NORM 特征相较于 PSD 特征的提升幅度更大.

唤醒度维度上的实验结果与分析

为了更全面的探究二维网状 EEG 时空特征表示方法的普适性,同时为了增强实验间的可对比性,结合 DEAP 数据集的特点,作者选择在唤醒度维度(Arousal)上同样对 NORM 和 PSD 特征进行四种模型的脑电分类计算.特征提取、特征结构转换方法以及分类所用模型和模型结构同上文所述完全相同,具体实验结果如表 4 所示

Table 4 Comparison of the accuracy of NORM & PSD features between two types of emotion classification in the arousal dimension

表 4 唤醒度维度上 NORM 和 PSD 特征两类情感分类准确率的比较

Models	Input Features	Arousal Acc
2D-CNN-V2	Segmented NORM chain sequence	0.5395
BILSTM(64)	Segmented NORM chain sequence	0.4957
Casc-CNN-CNN	Segmented 2D NORM mesh sequence	0.5764
Casc-CNN-LSTM	Segmented 2D NORM mesh sequence	0.5663
2D-CNN-V2	Segmented PSD chain sequence	0.8851
BILSTM(64)	Segmented PSD chain sequence	0.8889
Casc-CNN-CNN	Segmented 2D PSD mesh sequence	0.9102
Casc-CNN-LSTM	Segmented 2D PSD mesh sequence	0.9284

在唤醒度维度上,以 2D-NORM 特征作为输入的 CASC-CNN-CNN、CASC-CNN-LSTM 模型分类精度为 57.64%、56.63%,比以链式 NORM 特征作为输入的 2D-CNN-V2、BILSTM(64)模型的 53.95%、49.57%分别高出 3.69%和 7.06%.唤醒度维度上的 NORM 特征经过维度转换后,分类精度平均提高了 5.375%.而当两个级联模型

以 2D-PSD 特征作为输入时,分类准确率分别达到了 91.02%和 92.84%,比以链式 PSD 特征作为输入的 2D-CNN-V2、BILSTM(64)模型的 88.51%,88.89%分别高出 2.51%和 3.95%。唤醒度维度上的 PSD 特征经过维度转换后,分类精度平均提高了 3.23%。就特征结构的改变对特征可分性的影响而言,唤醒度维度下,NORM 特征依然比 PSD 特征高出 2.145%。

结合表 3 表 4 可以发现,效价维度上 NORM 特征经过结构转换后,其分类精度提升了 5.85%,而在唤醒度维度上 NORM 特征经过转换后提升的精度值为 5.375%,二者仅相差 0.475%。而效价和唤醒度维度上的 PSD 特征经过转换后提升的分类精度分别为 2.775%和 3.23%,两者的差值也仅为 0.455%。

以上实验说明在效价和唤醒度维度上,NORM 特征和 PSD 特征数据由一维链式结构转变为二维网状结构后,特征的可分性都会得到提升。并且特征结构的转换对于只包含时域信息的 NORM 特征的影响大于包含时频信息的 PSD 特征的影响。同时对于 NORM 特征和 PSD 特征而言,当其分别处于效价和唤醒度维度上时,一维到二维的特征结构变化为其分类精度带来的提高大致相同。

时间与空间信息对模型性能的影响

我们还比较了两种级联组合神经网络与其它几种基准神经网络的性能,来研究分析空间与时间信息对 EEG 情感识别的影响。通过比较 Casc-CNN-CNN 与 2D-CNN 发现,Casc-CNN-CNN 的总体性能优于 2D-CNN,这表明提取二维网状物理相邻传感器上的脑电信号比简单的一维链式脑电信号能更有效地捕获 EEG 的空间相关性,这更有助于识别人类大脑的情感模式。从表 2 可见,Casc-CNN-CNN 模型的性能又明显优于 3D-CNN 与 LSTM 模型的性能,比 3D-CNN 模型在效价上的分类性能提高了 4.63%,比 LSTM 模型的性能提高了 6.92%,表现出了显著的持续提升趋势。

而我们的 Casc-CNN-LSTM 级联组合模型在对每一个采样点的二维网状数据进行空间特征学习后,利用 LSTM 提取进一步学习一个 EEG 样本内连续采样点之间的全局时间动态,因此在效价上的分类性能进一步提高。然而,当使用单纯的 LSTM 模型只考虑时间相关性时,其识别率下降到 85.45%,与组合模型的 93.15%相差 7.7%,这说明时空特征的结合对脑电信号的准确分析是至关重要的。我们提出的级联卷积递归组合模型的识别率达到 93%以上,表明这种特征融合方法具有很强的时空表示能力,时空融合特征明显增强了脑电信号情感识别的准确性。

总结

本文首先提出一种新的基于网状序列的 EEG 时空特征表示方法,将每一个采样点上原始的一维链式通道信息转换成二维网状空间信息,该网状结构与 EEG 电极位置的脑区分布相对应,以此更好地表示 EEG 信号物理上多个相邻电极之间的空间相关性。再应用滑动窗将二维网状序列切成一个个等长的时间片段,作为新的融合了 EEG 时空相关性的数据表示。本文还提出两种用于 EEG 情感识别的端对端、可训练的混合深度神经网络模型,通过 CNN 网络从转换的 EEG 时空数据表示中捕获物理上相邻电极之间数据的空间相关性,通过 LSTM 网络学习 EEG 数据流在时序上的依赖关系。采用大规模数据集 DEAP 中 32 名被试在效价和唤醒度上两类 EEG 数据,来评估我们提出的 EEG 时空特征表示方法及混合深度学习模型的性能。实验结果表明,两种级联混合深度学习模型在效价上两类情感的识别的平均准确率分别达到了 93.15%和 92.37%,均明显优于目前最先进的方法,表明本文提出的方法能够利用混合神经网络有效地联合学习脑电信号在空间与时间上的相关性,进一步提高 EEG 情感识别的准确率和鲁棒性,可以有效的应用到基于 EEG 的情感分类与识别相关应用中。

References:

- [1] 卢官明,袁亮等.基于长短期记忆和卷积神经网络的语音情感识别.南京邮电大学学报(自然科学版),2018,38(05):63-69
- [2] 李娟,刘国忠,高洁.基于脑电信号的情绪分类.北京信息科技大学学报(自然科学版),2017,32(02):34-39.
- [3] 孙伟,黄进,李念龙,范向民,田丰,戴国忠,王宏安.脑-机接口辅助的动态目标选择技术.软件学报,2018,29(S2):108-119
- [4] 张家瑞,王刚.基于脑电信号的情感识别研究.计算机应用研究,2019,36(11):3306-3309.
- [5] 蒋静芳,曾颖,林志敏,童莉,闫镜.基于脑电信号的情绪评估研究综述.信息工程大学学报,2016,17(06):686-693.
- [6] Krizhevsky, A.Sutskever, I.Hinton, et al. (2012) ImageNet Classification with Deep Convolutional Neural Networks. International Conference on Neural Information Processing Systems, Lake Tahoe, 3-6 December 2012, 1097-1105.
- [7] Graves, Alex. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850 (2013).
- [8] Karpathy A, Toderici G, Shetty S, et al. Large-scale video classification with convolutional neural networks. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. IEEE: Columbus, Ohio. 2014: 1725-1732.
- [9] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [10] 钱忠, 李培峰, 周国栋, 朱巧明.基于双向 LSTM 网络的不确定和否定作用范围识别.软件学报,2018,29(8):2427-2447
- [11] Yue-Hei Ng J, Hausknecht M, Vijayanarasimhan S, et al. Beyond short snippets: Deep networks for video classification.Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE: Boston. 2015: 4694-4702.
- [12] Bashivan P, Rish I, Yeasin M, and Codella N. Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. International conference on learning representations (2015), <http://arxiv.org/abs/1511.06448>.
- [13] Ryan G.Hefron, Brett J.Borghetti,James C.Christensen,Christine M. SchubertKabban. Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation. Pattern Recognition Letters, 2017, 94(C):96-104.
- [14] Zhang D, Yao L, Zhang X, Wang S, Chen W. Eeg-based intention recognition from spatio-temporal representations via cascade and parallel convolutional recurrent neural networks. arXiv preprint arXiv:1708.06578, 2017.
- [15] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces. Journal of neural engineering, 2018, 15(5): 056013.
- [16] Alhagry S, Aly A, Reda A. Emotion Recognition based on EEG using LSTM Recurrent Neural Network. International Journal of Advanced Computer Science & Applications, 2017, 8(10).
- [17] M Soleymani, S Asghari-Esfeden, Y Fu, and M Pantic. Analysis of EEG signals and facial expressions for continuous emotion detection. IEEE Trans. Affect. Comput, vol. 7, no. 1, pp. 17–28, Jan./Mar. 2016.
- [18] Salama E S, El-Khoribi R A, Shoman M E, et al. EEG-based emotion recognition using 3D convolutional neural networks. Int. J. Adv. Comput. Sci. Appl, 2018, 9(8): 329-337.
- [19] Chen J X, Zhang P W, Mao Z J, et al. Accurate EEG-based emotion recognition on combined features using deep convolutional neural networks. IEEE Access, 2019, 7: 44317-44328.
- [20] Chen J X, Jiang D M, Zhang Y N. A hierarchical bidirectional GRU model with attention for EEG-based emotion classification. IEEE Access, 2019, 7: 118530-118540. Doi:10.1109/ACCESS.2019.2936817.
- [21] Koelstra S, Muhl C, Soleymani M, et al. Deap: A database for emotion analysis; using physiological signals. IEEE transactions on affective computing, 2011, 3(1): 18-31.2012.
- [22] Balconi M, Mazza G. Brain oscillations and BIS/BAS (behavioral inhibition/activation system) effects on processing masked emotional cues. ERS/ERD and coherence measures of alpha band. International Journal of Psychophysiology, 2009, 74(2): 158-165.

附中文参考文献:

- [1] 卢官明,袁亮等.基于长短期记忆和卷积神经网络的语音情感识别.南京邮电大学学报(自然科学版),2018,38(05):63-69
- [2] 李娟,刘国忠,高洁.基于脑电信号的情绪分类.北京信息科技大学学报(自然科学版),2017,32(02):34-39.
- [3] 孙伟,黄进,李念龙,范向民,田丰,戴国忠,王宏安.脑-机接口辅助的动态目标选择技术.软件学报,2018,29(S2):108-119.
- [4] 张家瑞,王刚.基于脑电信号的情感识别研究.计算机应用研究,2019,36(11):3306-3309.
- [5] 蒋静芳,曾颖,林志敏,童莉,闫镜.基于脑电信号的情绪评估研究综述.信息工程大学学报,2016,17(06):686-693.
- [10] 钱忠, 李培峰, 周国栋, 朱巧明.基于双向 LSTM 网络的不确定和否定作用范围识别.软件学报,2018,29(8):2427-2447