

基于AU的多任务学生情绪识别方法研究

张笑云, 赵晖

引用本文

张笑云, 赵晖. 基于AU的多任务学生情绪识别方法研究[J]. 计算机科学, 2024, 51(10): 105-111.

ZHANG Xiaoyun, ZHAO Hui. [Study on Multi-task Student Emotion Recognition Methods Based on Facial Action Units](#) [J]. Computer Science, 2024, 51(10): 105-111.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[融合Inception与SE-Attention的加密流量移动业务识别](#)

Identification of Mobile Service Type of Encrypted Traffic Based on Fusion of Inception and SE-Attention

计算机科学, 2024, 51(10): 399-407. <https://doi.org/10.11896/jsjcx.230900103>

[基于跳表的secGear性能优化方法](#)

Optimum Proposal to secGear Based on Skiplist

计算机科学, 2024, 51(6A): 230700030-5. <https://doi.org/10.11896/jsjcx.230700030>

[基于SEIR模型的网络热搜话题传播仿真研究](#)

Study on Communication Simulation of Online Hot Search Topics Based on SEIR Model

计算机科学, 2024, 51(6A): 230500107-6. <https://doi.org/10.11896/jsjcx.230500107>

[基于多任务联合训练的长文本多实体情感分析](#)

Long Text Multi-entity Sentiment Analysis Based on Multi-task Joint Training

计算机科学, 2024, 51(6): 309-316. <https://doi.org/10.11896/jsjcx.230400001>

[融合语义解释和DeBERTa的极短文本文层次分类](#)

Very Short Texts Hierarchical Classification Combining Semantic Interpretation and DeBERTa

计算机科学, 2024, 51(5): 250-257. <https://doi.org/10.11896/jsjcx.231100134>

基于 AU 的多任务学生情绪识别方法研究

张笑云¹ 赵 晖²

1 新疆大学软件学院 乌鲁木齐 830046

2 新疆大学信息科学与工程学院 乌鲁木齐 830046

(1004905193@qq.com)

摘 要 智能教育快速发展,运用人工智能提升教育质量和效率已成为趋势。学生作为教育的核心,其情绪状态对教育成效具有至关重要的影响。为了深入研究学生情绪,收集了课堂场景中的学生学习视频,包括听课和小组讨论两种情境,并据此建立了一个多任务学生情绪数据库。面部作为内在情绪状态的直接外在体现,显示出 AU 与情绪之间的紧密关联。在此基础上,提出了一个基于多任务学习的学生情绪识别模型 Multi-SER。该模型通过结合 AU 识别和学生情绪识别两项任务,挖掘各个 AU 与学生情绪之间的关联关系,进而提升模型在学生情绪识别方面的性能。在多任务实验中,Multi-SER 模型的情绪识别准确率达到了 80.87%,相比单情绪识别任务模型 SE-C3DNet+,效果提升了 3.11%。实验结果表明,通过多任务学习挖掘 AU 和情绪之间的关联关系,模型在分类各种情绪方面的性能得到了提升。

关键词: 学生情绪识别;多任务学习;C3D;SE;面部单元

中图分类号 TP391.1

Study on Multi-task Student Emotion Recognition Methods Based on Facial Action Units

ZHANG Xiaoyun¹ and ZHAO Hui²

1 College of Software, Xinjiang University, Urumqi 830046, China

2 School of Information Science and Engineering, Xinjiang University, Urumqi 830046, China

Abstract With the rapid development of intelligent education, it has become a trend to use artificial intelligence to improve the quality and efficiency of education. The emotional state of students, who are at the center of education, has a crucial impact on educational effectiveness. In order to study students' emotions in depth, this paper collects students' learning videos in classroom scenarios, including two contexts of listening to lectures and group discussions, and builds a multi-task students' emotion database accordingly. The face serves as a direct outward manifestation of the internal emotional state, which shows a strong correlation between AU and emotions. Based on this, this paper proposes a multi-task learning-based student emotion recognition model Multi-SER. The model explores the association relationship between individual AUs and students' emotions by combining the two tasks of AU recognition and students' emotion recognition, thereby improving the performance of the model in students' emotion recognition. In the multi-task experiment, the Multi-SER model achieves an accuracy of 80.87% in emotion recognition, which improves the effect by 3.11% compared to the single emotion recognition task model SE-C3DNet+. The experimental results show that the performance of the model in categorizing various emotions is improved by mining the correlations between AUs and emotions through multi-task learning.

Keywords Student emotion recognition, Multi-task learning, C3D, SE, Facial action units

随着信息技术和人工智能(AI)的迅速发展,在互联网的引领下,传统行业正在向数字化、智能化方向迈进。国务院颁布的《中国教育现代化 2035》^[1]和教育部推出的《教育信息化 2.0 行动计划》^[2]文件都强调了推进信息技术、人工智能与教育教学融合的重要性,这意味着教育部门和相关机构将致力于利用人工智能技术来创新教学模式,探索智能化教学工具的应用,以提高教学质量和效果。

学生的情绪对教学成果有着决定性影响。掌握学生在学习过程中的情绪,对于深入理解学生的学习体验、提高教学效果至关重要。首先,教师可以通过观察学生的情绪,洞察其学习动机、兴趣、注意力等,进而调整教学方法,使之更贴合学生的实际需求,以此改进教学效果。其次,对学生情绪的分析也为教学评价和教学反思提供了重要依据,教师可以根据学生情绪的反馈来评估教学内容的吸引力、难度和有效性,并进行

到稿日期:2024-03-11 返修日期:2024-07-04

基金项目:国家自然科学基金(62166041)

This work was supported by the National Natural Science Foundation of China(62166041).

通信作者:赵晖(zherry@126.com)

必要的调整和改进^[3]。

因此,本研究旨在结合人工智能和教室硬件技术,通过建立一个课堂环境下的学生情绪数据集,并利用 AU 作为辅助信息提升情绪识别的准确性。这样的方法旨在辅助教师更精确地理解学生的情绪状态,从而提升教学质量。

1 学生情绪识别研究现状

课堂技术和智能课堂的快速发展使得研究人员开始探索如何准确地捕捉学生在学习过程中的情绪,以更好地理解他们的学习体验并优化教学效果。一些研究人员参考了保罗·艾克曼(Paul Ekman)提出的愤怒(Anger)、厌恶(Disgust)、害怕(Fear)、高兴(Happiness)、悲伤(Sadness)和惊讶(Surprise)的 6 种基本情绪作为学生情绪类别^[4-5]。如 Sumalakshmi 等^[6]借鉴了这 6 种基本情绪,并在 FER-2013 数据集上进行了实验。Dai 等^[7]提出了一种基于多头注意力的方法,用于对学生情绪进行分类。他们将学生的情绪分为快乐、悲伤、愤怒、惊讶、恐惧、厌恶和中性,并在静态面部数据 JAFFE 和 OLSFED 数据集上进行了测试。Jiao 等^[8]构建了一个全局多尺度和局部注意力网络(MA-Net)的表情识别模型,该模型将学生情绪划分为愤怒、厌恶、恐惧、高兴、中性、悲伤、惊讶 7 类。为了验证模型的准确性和可靠性,他们在公开的面部表情数据集 KDEF、FED-RO 和 Pose-AffectNet 上进行了实验。

随着研究的深入,研究者们发现 Ekman 的情绪理论不完全适合描述学生的情绪,同时注意到非课堂环境下采集的数据集与学生在环境中的情绪表现之间存在显著差异。为了更好地理解学生的情绪,研究者们采用问卷调查和访谈等方法确定学生情绪,并开始构建针对学生情绪的专用数据集。Han 等^[9]通过问卷调查和访谈,对学生在听课过程中的情绪状态进行了分类,将其划分为倾听、疑惑、理解、抗拒和不屑 5 种。他们通过结合现有监控设备,并应用多姿态人脸检测以及面部表情识别技术,实时获取学生在学习过程中的情绪。同时,通过分析学生眉毛两端与眼睛中点夹角、头部角度和唇部 3 个特征点的变化来判断学生的情绪状态。Gao^[10]通过总结相关研究成果,与一线教师交流,并结合课堂实地考察,主要对愉悦、困惑、惊奇、疲倦和中性 5 种常见的学习情绪进行研究。此外,还构建了一个自发数据集,但在录制时对志愿者设置了不戴眼镜、刘海不遮挡眉毛的要求。Chen^[11]将学生情绪分为专注、愉悦、困惑、沮丧和厌倦。他设计了特定的数学测试和智力挑战游戏,目的在于诱发目标情绪反应,并邀请学生参与实验,以采集和分析这些情绪。Xu^[12]在两所中学和一所高校展开了结构化观察,记录学生在课堂学习和在线学习过程中出现的情绪和次数,并对 5 名教师进行访谈,了解他们对学生情绪的看法。最终根据结构化观察和访谈结果,将学生的情绪分为常态、高兴、愤怒、悲伤、惊恐、专注、走神。此外,他创建了一个学生表情数据库,并在数据收集前对 70 名研究生进行了培训,目的是让他们在自然状态下尽可能准确地表达和展现不同的情绪。

尽管这些研究为我们理解学生情绪提供了新视角,但通过结构化观察、访谈、问卷调查和实地考察等方法来识别学生情绪可能会受到个人认知偏见和情绪波动的影响,这可能会

损害数据的客观性和精确性。此外,特定课堂环境下收集的数据可能因非自发性或某些限制条件而与学生 in 自然课堂环境中的情绪表达不一致,进而影响数据的代表性和广泛适用性。本文的主要贡献如下:

1)在综合现有研究成果的基础上,结合 Plutchik 的情绪理论,同时考虑到课堂情境中学生情绪的温和特质,选择了 Plutchik 情绪轮^[13]中的温和情绪类别来对学生情绪进行分类。这些情绪包括兴趣(Interest)、接受(Acceptance)、宁静(Serenity)、忧虑(Apprehension)、涣散(Distraction)、沉思(Pensiveness)、无聊(Boredom)和烦恼(Annoyance)。

2)为了深入探析学生的情绪,收集了学生在教学场景下的学习视频数据,覆盖了学生在听课和小组讨论这两种不同的互动情景。基于这些数据,构建了一个多任务学生情绪数据库,包含学生情绪和 AU 的标注。

3)提出了一种基于多任务学习的学生情绪分类模型,综合考虑了 AU 识别和情绪识别两项任务。通过引入 AU 的辅助信息并利用信息共享机制,进一步提升了模型在学生情绪识别方面的性能。

2 相关研究

面部活动单元(Facial Action Units,简称 AU)起源于 Ekman 等^[14]开发的脸部行为编码系统(Facial Action Coding System,FACS),它是基于对面部肌肉动作的详细测量和分析,将面部划分为不同的 AU,每个 AU 对应特定的面部肌肉动作,以描绘和衡量面部肌肉的变化。

AU 与情绪的紧密联系体现在面部表情通常是情绪内在状态的外在反映^[15]。首先,从生理学角度来看,AU 和面部表情都是以面部肌肉运动为研究对象^[16]。AU 是通过对面部肌肉运动的系统观察和分析而定义的一系列动作单元,它能够精确地描述面部表情的细微变化,从而为我们提供了一种客观、量化的方式来研究面部肌肉与情绪之间的关系。每个 AU 都对应着特定的面部肌肉或肌肉群的运动,这些运动在不同情绪状态下会有不同表现。其次,从心理学角度来看,不同的情绪状态会导致不同的心理过程和认知变化,这些变化会进一步影响面部表情的形成。例如,当一个人感到快乐时,面部通常会展现出 AU12(嘴角上扬)和 AU6(脸颊提升)。相反,在感受到烦恼或忧虑时,则可能表现出 AU4(眉头皱起)。

面部表情是人类用来传达情绪最有力、最自然、最普遍的信号之一^[17]。AU 作为面部表情的量化指标,与情绪之间存在着紧密的生理学 and 心理学联系。这种对 AU 的观察为我们深入理解和预测情绪提供了宝贵的线索^[18]。

单任务学习(Single-Task Learning,STL)用于实现单一任务或目标,模型通过学习特定的目标函数来完成该任务。多任务学习(Multi-Task Learning,MTL)是一种机器学习方法,旨在通过同时学习多个相关任务来提升模型的性能^[19]。在多任务学习中,一个模型被训练来完成多个任务,而不是仅针对单一任务进行训练。这些任务可以是相关的、互补的,并通过共享底层表示来提升性能。Zhang 等^[20]提出了一种情绪识别模型,该模型基于多任务学习框架,特别融入了跨模态

注意力机制。这种机制的引入使得模型能够高效地学习和区分不同模态(如视觉、听觉)中的特有特征以及跨模态间的共享特征。通过这种方式,模型优化了对各个模态数据的处理和理解,同时促进了不同信息源之间特征的整合,从而在情绪识别任务上实现了更高的准确度和鲁棒性。Zhang 等^[21]将情感分析和情绪识别视为互相关联的两个任务,并提出了一种创新的多任务学习模型,该模型采用跨任务注意力机制。这种机制使得模型能够在执行情感分析时注意到情绪识别中的相关信息,从而优化了两个任务之间的信息流动和特征共享。通过这种方法,模型能够更有效地理解和处理复杂的情感和情绪数据,为提高两种任务的模型性能提供了新的策略和视角。Li 等^[22]指出人格特征与情绪存在显著的相关性,并基于这一观察,设计了一个融合迁移学习理念的多任务学习框架。该框架旨在并行预测个人的人格特征和情绪,并深入探讨了这两个任务之间的信息共享策略。

3 数据集构建

3.1 数据采集

课堂环境对于准确识别学生情绪至关重要。因此,本文采集课堂场景中学生的自然表现(没有施加任何外部约束或指导),构建了一个名为 EmoAU 的数据库,包括学生上课的学习视频和学生小组讨论的视频,涵盖了课堂和小组讨论两种情景。虽然研究者们对学生情绪分类的看法各不相同,但情绪词义之间确实存在相似之处。本文在深入分析现有研究

成果并细致观察课堂上学生情绪表现的基础上,考虑到学生情绪在教育环境下通常展现出温和的特质,特别选取了 Plutchik 情绪轮中的 8 种温和情绪,即感兴趣、接受、宁静、忧虑、涣散、沉思、无聊和烦恼,作为学生情绪分类的依据。本数据集涵盖了 74 名学生,年龄在 23~30 岁之间,包括本科生、研究生和博士生。课堂视频涉及不同的学科内容,而小组讨论则围绕计算机研究方向和机器学习两个主题展开。

3.2 数据处理

本研究收集了 77 个学生在上课和讨论过程中的视频,总数据量约为 156GB,所有视频均已转换为统一的 MP4 格式。随后,招募了志愿者来提取并标注视频中的学生情绪。为了提高标注的准确性,志愿者接受了全面的培训,包括情绪识别的基本知识、Plutchik 情绪轮理论、需识别的情绪类型,以及如何使用剪映软件提取视频片段的操作指南。之后,志愿者借助剪映软件,手动提取视频中的情绪数据,以确保数据的准确性。在预处理视频时,考虑到个体差异的重要性,本文没有对视频片段的时长和尺寸设定任何限制。

3.3 情绪数据标注

在进行视频剪辑标注的过程中,志愿者会密切观察学生的情绪表现。一旦发现目标情绪表达,立即暂停视频,通过调整时间轴前后位置来仔细分析情绪变化,并进行初步情绪状态的判断。在初步评估完成后,标注人员会依据表 1 中的情绪判定标准进行进一步评估。该表综合了现有文献、图书、摄影领域中关于情绪描述的关键特征。

表 1 情绪特征
Table 1 Emotional features

Emotion	Face gesture	Body gesture
Interest ^[23]	Cheeks up	Lean forward
	Raise eyebrows	Head tilt
	Crinkle at the corner of the eye	Adjustment of body posture
	Open your mouth wide or show your teeth	
Acceptance ^[11]	Lift the upper eyelid	Nod of the head
	The corners of one's mouth lift	
	Corner of mouth stretching	
	Lips pressing against each other	
Serenity ^[24-25]	There are no distinct facial features	No distinctive physical features
	Glower	Lift one's chin
Boredom ^[24]	Droopy eyelid	Face in hand
	Close one's eyes	Hands palms under the chin, elbows on table
	The corners of the mouth go down	Body changes direction and moves to the right/left yawn
		To turn one's head
Distraction ^[23]	Eyes right/left/up/down	Play with one's cell phone
	Mouth opens/closes naturally	Looking out the window, classmates
Apprehension ^[24]	Knit one's brow	Bite one's finger
	Bite one's lip	Tap the table with your fingertips
	Ptosis of the eyelids	Head tilt Frequent movements
	Eyes right/left/up/down	
Pensiveness ^[26]	Knit one's brow	Scratch one's chin
	Raise one's head	Head tilt
	Scrunch up one's nose	Bow one's head Climb on the table Hands behind one's head Bury one's head in the sand
Annoyance ^[27]	Knit one's brow	
	Drooping eyebrows	
	Lift the upper lip and lower lip up	

为了提高标注结果的准确性和可靠性,标注人员将结合初步判断和标准表评估来确定最终的情绪类别。每个视频片段将由 5 名志愿者独立评估,以保证标注结果的一致性并降低个

人主观判断的影响。当出现分类上的不确定性或分歧时,如新出现的特征、多重特征表达或标注人员意见不一致,本文将采取以下措施:若某个类别获得超过 60% 标注者的支持,则

采纳多数意见为最终分类;若无类别达到此标准,则进行集体讨论以决定类别;讨论无果的样本将被排除。

最终,本文从原始数据集中成功提取了 7704 个学生情绪视频片段。其中,宁静个数为 1 744,涣散个数为 1 946,无聊个数为 1 546,感兴趣个数为 1 080,沉思个数为 988,接受个数为 298,烦恼个数为 98 个,忧虑个数为 4 个。鉴于忧虑情绪样本数量过少,本研究不对其进行深入探讨。针对烦恼和接受情绪样本不足的问题,本文采用了数据扩充技术以增强模型的泛化能力。鉴于细微的情绪变化可能会影响情绪分类,本文选择了水平翻转作为数据扩充的手段。这种方法可以产生与原始视频片段相似但又略有差异的新样本,这对于维持情绪表达的一致性来说是有效且适宜的。

3.4 AU 标注

在标注学生 AU 的过程中,人工标注是一项挑战,这项任务不仅复杂,而且耗时。要成为一名合格的标注人员,需要经过大约 100h 的专业培训,以确保其能够精确地识别和标注各种 AU。针对 AU 标注的挑战,本文选择了一种更为实用的方法。这种方法着重于标注那些容易被直观识别的 AU,标注人员不必花费大量时间进行细致的分析,从而减少错误的可能性。同时,因直观的面部特征在不同标注者之间的主观判断较为一致,故此方法也有助于提高数据标注的一致性。此外,研究特别关注与课堂行为紧密相关的 AU。例如,感兴趣通常表现为脸颊抬高和嘴角上扬,而沉思和烦恼的表达则可能表现为眉头皱紧和嘴角下拉。因此,本文选择了 AU4, AU6, AU12, AU15, AU17, AU27, AU28, AU43 这几种容易识别的 AU 进行标注,如表 2 所列。

表 2 EmoAU 数据集中的 8 种 AU
Table 2 8 AUs in EmoAU dataset

AU	English Name
AU4	Brow lowerer
AU6	Cheek raiser
AU12	Lip corner puller
AU15	Lip corner depressor
AU17	Chin raiser
AU27	Mouth stretch
AU28	Lip suck
AU43	Eyes Closed

为确保 AU 标注的准确性,标注人员接受了专门的培训,深入学习了每个 AU 的具体位置和特征动作。同时,我们为标注工作创造了一个安静且光线充足的环境,以确保标注者能够清晰观察和记录 AU 信息。经过对面部表情进行详细观察和分析,标注人员能够准确识别出这 8 种 AU。为了消除不同标注人员在评估 AU 强度时可能产生的主观差异,本文采用了二进制的标注方法,仅记录 AU 是否出现。如果样本中特定 AU 出现,则标为 1;如果未出现,则标为 0。

4 基于 AU 的多任务情绪识别方法研究

为了更精准地识别学生情绪,本研究构建了一个基于多任务学习的网络模型,即 Multi-SER,该模型能够同时执行 AU 和情绪的识别两项任务。模型的输入是一个包含 16 帧图像的序列,每帧都是 3 通道的彩色图像,且每帧的尺寸大小为 112×112 。Multi-SER 整体框架分为共享层和输出层,共享层主要负责通道特征融合和时空特征提取,输出层则用于各个任务的具体输出。详细架构细节已在图 1 中展示。

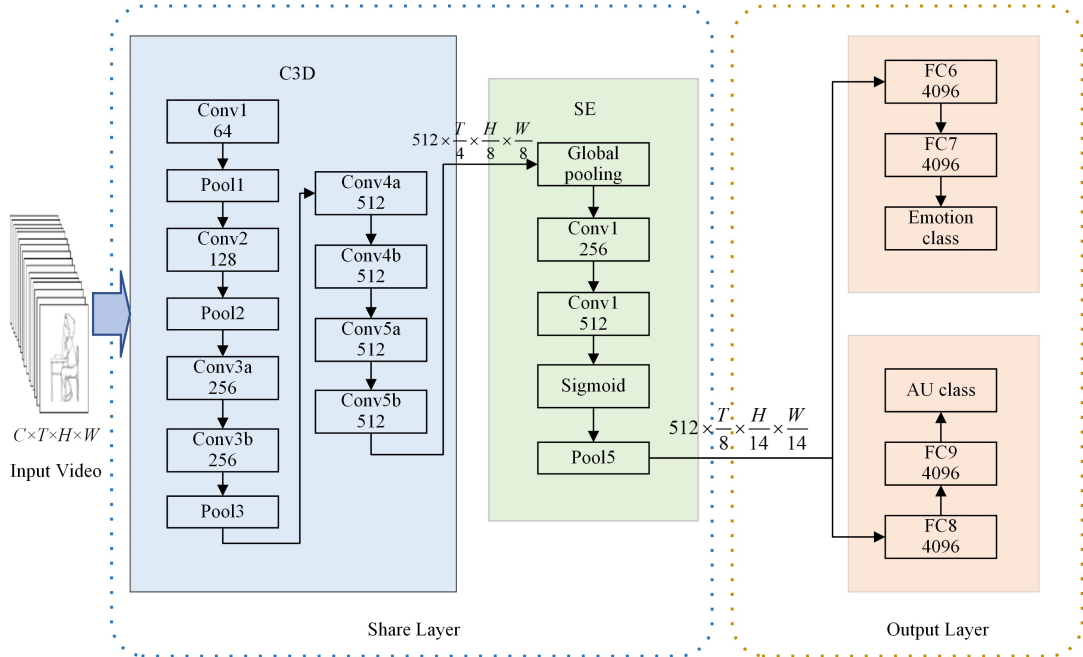


图 1 Multi-SER 网络结构图
Fig. 1 Network architecture of Multi-SER

4.1 共享层

在 Multi-SER 网络架构中,本文设计了一个共享层,该共享层结合了 SE(Squeeze-Excitation)和 C3D 模型,以实现对特征的处理和提取。其中,C3D 模型负责从视频数据中捕获

时空特征,为模型提供丰富的时空信息;而 SE 模块自适应地调整网络中不同通道的权重,从而突出重要特征,抑制不重要的特征。具体的计算式如式(1)和式(2)所示:

$$x_1 = C3D(h)$$

(1)

$$\tilde{x}=SE(x_1)$$

(2)

其中, h 表示输入的特征, x_1 表示经过 C3D 模块后得到的输出; $\tilde{x}$ 表示经过 SE 模块后得到的输出。

在共享层部分,本研究采用参数的硬共享机制,这一机制使得多个目标模型能够联合训练,并共享主体部分参数。这一策略不仅大幅缩减了模型参数的总规模,而且有效地避免了过拟合现象的发生,同时降低了模型的复杂度,提高了模型的泛化能力和学习效率。

4.2 输出层

Multi-SER 网络的输出层划分为学生情绪识别通道和 AU 识别通道,每个任务都拥有其专门的输出层,确保了任务的独立性。这种多通道设计的输出层使得模型能够从共享的特征表示中精确提取出不同任务所需的关键信息,进而实现精准预测。这样的设计使得 Multi-SER 网络具有更强的灵活性和适应性,能够同时解决多个任务,并且在每个任务上都能取得良好的性能。具体的实现公式如式(3)–式(6)所示:

$$x_{em_i}=W_{em_i}\tilde{x}+b_{em_i}$$

(3)

$$x_{au_i}=W_{au_i}\tilde{x}+b_{au_i}$$

(4)

$$p_{em}=\text{softmax}(W_{em}x_{em_i}+b)$$

(5)

$$p_{au}=\text{softmax}(W_{au}x_{au_i}+b)$$

(6)

其中, W_{em_i} 和 W_{au_i} 表示第一层激活线性层权重, W_{em} 和 W_{au} 表示第二层激活层的权重; b 为偏置项; p_{em} 表示模型在学生情绪上的概率; y_{au} 表示模型在 AU 识别上的概率。

4.3 损失函数设计

本文中情绪识别和 AU 识别两个任务均采用了交叉熵损失函数来训练网络,具体损失函数如式(7)和式(8)所示:

$$l_{em}=-\frac{1}{N}\sum_i\sum_{c=1}^ky_{em}^i\log(p_{em}^i)$$

(7)

$$l_{au}=-\frac{1}{N}\sum_i\sum_{c=1}^my_{au}^i\log(p_{au}^i)$$

(8)

其中, k 和 m 分别代表学生情绪识别和面部行动单元类别的数目; y_{au}^i 和 y_{em}^i 分别代表学生情绪识别和面部行动单元识别第*i*个样本类别*c*的标签; p_{em}^i 和 p_{au}^i 分别代表学生情绪识别和面部行动单元识别第*i*个样本类别*c*的概率。

为平衡情绪识别和 AU 识别这两个任务,本文采用了加权损失策略,通过为各任务分配不同权重,调节它们在总损失中的权重占比,从而提升学习效率和识别精度。这种方法不仅解决了任务重要性差异的问题,还实现了多任务学习框架下的任务平衡,如式(9)所示:

$$l=\alpha l_{em}+\beta l_{au}$$

(9)

其中, l 代表总损失, α 和 β 分别代表情绪识别和 AU 识别任务的权重。

5 实验结果与分析

5.1 实验参数设置

本文将 7704 个样本按照 8:2 的比例划分为训练集和测试集,然后将训练集中的数据再次以 8:2 的比例划分为训练集和验证集。为保证数据划分的一致性,本文使用了固定的随机种子来确保每次分配的样本都相同。随后,在 EmoAU 数据集上,本文运用 Multi-SER 模型进行了一系列实验,并采用

SGD 作为优化器,模型的具体网络参数配置如表 3 所列。

表 3 模型参数

Table 3 Model parameters

Parameters	Value
Learning rate	1×10^{-3}
Batch_size	64
momentum	0.9
Weight_decay	5×10^{-6}
Number of epochs for training	100
α	0.7
β	0.3
Dropout	0.1

5.2 学生情绪识别结果

本实验采用 Multi-SER 模型在 EmoAU 数据集上进行情绪识别,相关实验结果如表 4 所列。“涣散”情绪的准确率达到了 92.18%,F1 得分也达到了 88.72%。能获得这一良好的表现可能是因为学生在“涣散”状态下的动作变化较为明显,如频繁转头或低头玩手机。相较之下,“接受”情绪的精确度不高,但是其召回率达到了 90%,这表明该模型能有效识别出大部分真正表现为“接受”情绪的样本。然而,较低的精确率在分类时可能将许多非“接受”情绪的样本误分类为该类别,进而造成了较高的假阳性率。根据图 2 的混淆矩阵,“宁静”和“接受”情绪之间存在 20.45%的混淆概率,这很可能是因为这两种情绪在外部表现上具有高度的相似性。具体而言,“宁静”通常没有显著的外在表征,而“接受”情绪通常仅伴随细微的面部动作变化,这使得区分这两种情绪变得更具挑战性。这一发现提示我们,在未来的模型训练中需要对这两种情绪之间的区分度给予更多关注。

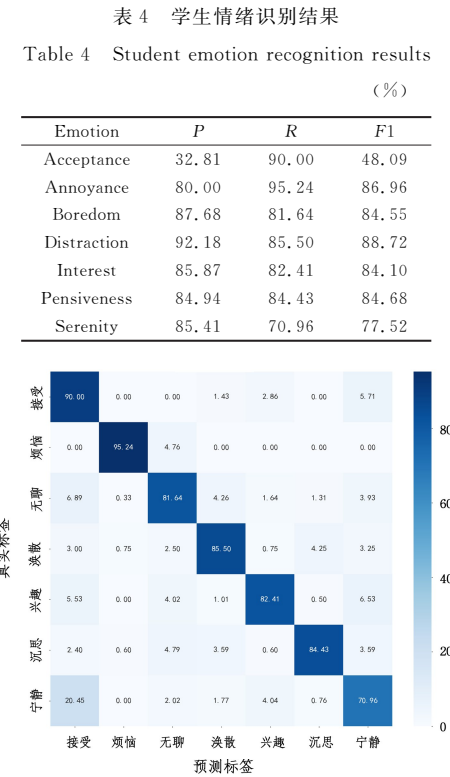


图 2 学生情绪识别混淆矩阵图

Fig. 2 Diagram of students' emotion recognition confusion matrix

5.3 模型对比实验

为了全面评估 Multi-SER 模型在情绪分类任务中的性能,本文选取了一系列模型进行对比实验,包括 3D 卷积神经网络 C3D、混合网络 ResNet18+LSTM、3D ResNet18、R(2+1)D18、P3D 和 STT 模型。此外,为了验证多任务情绪识别模型的有效性,本文还设计了一个用于单情绪识别任务的模型 SE-C3DNet+,并将其纳入对比实验中。该模型使用 C3D 进行特征提取,随后引入 SE 模块,自适应地调节不同通道之间的重要性。实验结果如表 5 所列。在基线模型对比中,C3D 在准确率上的表现优于其他基线方法;同时,本文提出的 Multi-SER 模型在准确率和 F1 得分上均达到了最优,分别为 80.87% 和 74.76%,优于单情绪识别任务 SE-C3DNet+ 模型。这一结果证明了虽然情绪识别和 AU 识别是两项不同的任务,但通过多任务学习,模型能够利用 AU 信息为情绪识别提供有价值的辅助信息。

表 5 模型对比实验
Table 5 Model comparison experiment (%)

Method	Acc	F1
C3D ^[28]	65.27	51.36
3D ResNet18 ^[29]	63.77	59.30
R(2+1)D18 ^[30]	50.19	47.34
P3D ^[31]	53.20	44.28
ResNet18+LSTM ^[32]	64.72	56.27
STT ^[33]	59.10	53.79
SE-C3DNet+	77.76	63.23
Multi-SER	80.87	74.76

5.4 AU 辅助对基线模型情绪识别的影响

为了探索引入 AU 辅助是否能提高学生情绪识别的准确率,本文在 R(2+1)D 和 P3D 模型中分别加入了 AU 识别任务,旨在验证该策略在不同架构下的有效性。实验结果如图 3 所示,在未加入 AU 辅助的情况下,R(2+1)D 模型和 P3D 模型的情绪识别准确率分别为 50.19% 和 53.20%;而引入 AU 辅助后,R(2+1)D 模型的准确率提高至 56.21%,P3D 模型的准确率提升至 55.33%,分别提升了 6.02% 和 2.13%。这些数据表明,多任务学习策略能够促进 AU 和情绪信息源之间特征的整合,从而提高情绪识别的准确性。

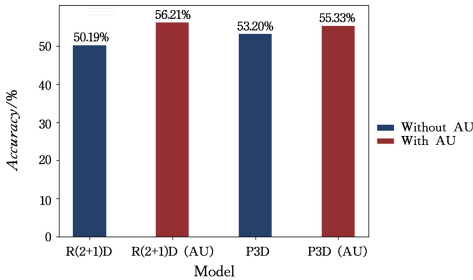


图 3 有无 AU 辅助基线模型的情绪识别效果

Fig. 3 Effect of baseline model on emotion recognition with or without AU

结束语 为了研究课堂环境中学生的情绪,本文构建了 EmoAU 数据集,该数据集包含了 AU 和情绪标签。在此基础上,本文提出了一种基于 AU 的多任务学生情绪识别模型,该模型将情绪识别和 AU 识别整合为一个联合学习任务,

实现了二者的并行处理。实验结果表明,采用多任务学习的方式,模型能够充分利用 AU 和情绪之间的相互关联,进而提升模型在情绪分类方面的性能。然而,当前的研究仍局限于课堂场景,未来我们将进一步拓展研究范围,探索不同教育场景下的学生情绪识别问题,如讲座、在线学习等场景。同时,还将考虑对男生女生不同情绪表达进行分类,以期更全面地理解和分析学生情绪的变化。

参 考 文 献

[1] 中共中央国务院印发《中国教育现代化 2035》[N]. 人民日报, 2019-02-24(001).

[2] 教育部关于印发《教育信息化 2.0 行动计划》的通知[J]. 中华人民共和国教育部公报,2018,(4):118-125.

[3] WEI Y T, LEI F, HU M J et al. A review of student expression recognition[J]. China Educational Informatization, 2020 (21): 48-55.

[4] YANG H, PENG C, SUN R, et al. Student Expression Recognition in Smart Education Environment based on Convolutional Neural Network[C]//2022 9th International Conference on Dependable Systems and Their Applications(DSA). 2022.

[5] PANDIMURUGAN V, SINGH A, TIWARI A. Facial Emotion Recognition for Students Using Machine Learning[C]//2023 International Conference on Computer Communication and Informatics(ICCCI). IEEE, 2023; 1-4.

[6] SUMALAKSHMI C H, VASUKI P. Fused deep learning based Facial Expression Recognition of students in online learning mode[J]. Concurrency and Computation: Practice and Experience, 2022, 34(21): e7137.

[7] DAI X, WEI P, ZENG Y, et al. Students' Facial Expression Recognition Based on Multi-head Attention Mechanism[J]. Journal of Physics: Conference Series. IOP Publishing, 2022 (2493): 012004.

[8] JIAO S, YAN Y X. Research on Students' Emotional State Based on Expression Recognition Technology—Taking the Junior Middle School Classroom as an Example[J]. Modern Information Technology, 2022(18): 6.

[9] HAN L, LI Y, ZHOU Z J et al. Teaching Effect Analysis Based on the Facial Expression Recognition in Classroom[J]. Modern Distance Education Research, 2017(4): 8.

[10] GAO Q. Spontaneous learning facial expression recognition based on RealSense research [D]. Wuhan: Central China normal university, 2020.

[11] CHEN Z J. Research on learning emotion recognition method and application based on facial expression in online learning environment [D]. Wuhan: Central China normal university, 2020.

[12] XU Z G. Emotion Recognition of learning picture in Intelligent Learning Environment and its application [D]. Jinan: Shandong Normal University, 2019.

[13] PLUTCHIK R. A general psychoevolutionary theory of emotion [M]//Theories of Emotion. Academic Press, 1980; 3-33.

[14] EKMAN P, FRIESEN W V. Facial action coding system [J/OL]. <https://www.paulekman.com/facial-action-coding-system/>.

[15] COSTA W, TALAVERA E, OLIVEIRA R, et al. A survey on datasets for emotion recognition from vision: Limitations and in-the-wild applicability[J]. Applied Sciences, 2023, 13(9): 5697.

[16] ZHAO H, WANG Z L, LIU Y F. A Survey of Automatic Facial Action Units Recognition[J]. Journal of Computer-Aided Design & Computer Graphics, 2010, 22(5): 894-906.

[17] LIU C, ZHANG X, LIU X, et al. Facial Expression Recognition Based on Multi-Modal Features for Videos in the Wild[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5871-5878.

[18] HOSSAIN S, UMER S, ROUT R K, et al. Fine-grained image analysis for facial expression recognition using deep convolutional neural networks with bilinear pooling[J]. Applied Soft Computing, 2023, 134: 109997.

[19] ZHANG Y, YANG Q. A survey on multi-task learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 34(12): 5586-5609.

[20] ZHANG S, YIN C, YIN Z. Multimodal sentiment recognition with multi-task learning[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 7(1): 200-209.

[21] ZHANG Y, RONG L, LI X, et al. Multi-modal sentiment and emotion joint analysis with a deep attentive multi-task learning model[C] // European Conference on Information Retrieval. Cham: Springer International Publishing, 2022: 518-532.

[22] LI Y, KAZEMEINI A, MEHTA Y, et al. Multitask learning for emotion and personality traits detection[J]. Neurocomputing, 2022, 493: 340-350.

[23] BIAN C, ZHANG Y, YANG F, et al. Spontaneous facial expression database for academic emotion inference in online learning [J]. IET Computer Vision, 2019, 13(3): 329-337.

[24] GUNES H, PICCARDI M. A bimodal face and body gesture database for automatic analysis of human nonverbal affective behavior[C] // 18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 1: 1148-1153.

[25] WALLBOTT H G. Bodily expression of emotion[J]. European Journal of Social Psychology, 1998, 28(6): 879-896.

[26] EL KALIOUBY R A. Mind-reading machines: automated inference of complex mental states[R]. University of Cambridge, Computer Laboratory, 2005.

[27] KOSTI R, ALVAREZ J M, RECASENS A, et al. Emotion re-

cognition in context[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1667-1675.

[28] RASTGOO R, KIANI K, ESCALERA S, et al. Multi-modal zero-shot dynamic hand gesture recognition[J]. Expert Systems with Applications, 2024, 247: 123349.

[29] HARA K, KATAOKA H, SATOH Y. Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet? [C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018.

[30] WANG H, LI B, WU S, et al. Rethinking the Learning Paradigm for Dynamic Facial Expression Recognition[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 17958-17968.

[31] ROMEO L, CAVALLO A, PEPA L, et al. Multiple instance learning for emotion recognition using physiological signals[J]. IEEE Transactions on Affective Computing, 2019, 13(1): 389-407.

[32] JIANG X, ZONG Y, ZHENG W, et al. DFEW: A Large-Scale Database for Recognizing Dynamic Facial Expressions in the Wild[C] // Proceedings of the 28th ACM International Conference on Multimedia. 2020: 2881-2889.

[33] MA F, SUN B, LI S. Spatio-temporal transformer for dynamic facial expression recognition in the wild[J]. arXiv: 2205. 04749, 2022.



ZHANG Xiaoyun, born in 1996, post-graduate, is a member of CCF (No. T6014G). Her main research interests include natural language processing and computer vision.



ZHAO Hui, born in 1972, Ph.D, professor. Her main research interests include artificial intelligence, affective computing, speech and digital image processing.

(责任编辑:何杨)