

## 基于多模态生理数据的情感识别综述



刘颖<sup>1</sup>, 袁莉<sup>1\*</sup>, 祖钰迪<sup>1</sup>, 范有腾<sup>2</sup>, 谢宁<sup>2</sup>, 杨阳<sup>2</sup>

(1. 军事科学院 军队政治工作研究院, 北京 100091; 2. 电子科技大学 计算机科学与工程学院, 成都 611731)

**摘要** 情感识别是人工智能和人机交互领域的重要研究方向, 对提高用户体验和应用的智能性具有重要意义。基于多模态生理数据的情感识别由于其数据来源的客观性和多样性, 能够更准确地捕捉个体的情感状态, 成为近年来的研究热点。首先介绍了情感计算的基本概念和情感理论模型。其次总结了基于生理数据的情感识别方法。再重点介绍了基于多模态生理数据的情感识别流程, 包括生理数据预处理、传统机器学习方法以及深度学习方法。最后分析了基于多模态生理数据的情感识别面临的主要挑战和对未来的展望。

**关键词** 情感识别; 生理数据; 深度学习; 机器学习

中图分类号 TP391 文献标志码 A DOI 10.12178/1001-0548.2024176

## Emotion Recognition Based on Multimodal Physiological Data: A Survey

LIU Ying<sup>1</sup>, YUAN Li<sup>1\*</sup>, ZU Shuodi<sup>1</sup>, FAN Youteng<sup>2</sup>, XIE Ning<sup>2</sup>, and YANG Yang<sup>2</sup>

(1. Military Political Work Research Institute of the Academy of Military Science, Beijing 100091, China;

2. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

**Abstract** Emotion recognition is an important research direction in the fields of artificial intelligence and human-computer interaction. It has significant implications for enhancing user experience and the intelligence of applications. Emotion recognition based on multimodal physiological data has become a research hotspot in recent years due to the objectivity and diversity of its data sources, which enable more accurate capture of an individual's emotional state. Firstly, the basic concepts of affective computing and emotion representation models are introduced. Secondly, emotion recognition methods based on physiological data are summarized. Then, the focus shifts to the process of emotion recognition based on multimodal physiological data, including physiological data preprocessing, traditional machine learning methods, and deep learning methods. Finally, the main challenges faced by emotion recognition based on multimodal physiological data are analyzed, and future prospects are discussed.

**Key words** emotional recognition; physiological data; deep learning; machine learning

人类的情感是在对外界刺激和内部状态的反应中所产生的心理和生理体验<sup>[1]</sup>, 不仅包括主观的感觉, 还伴随着生理反应和行为表现。影响着我们的决策、行为、社交互动以及整体的心理健康。

在人机交互的过程中, 计算机能否识别出人的情感并作出相应的决策呢? 这就需要计算机实现对人类情感的计算。情感计算 (Affective Computing, AC) 研究涉及多种学科, 目的是致力于开发能够识别人类情感, 并对其进行解释和模拟的计算系

统。核心理念是使计算机能够与人类以更自然、更直观的方式互动, 理解和表达情感, 通过获取情感的变化信息来辅助人类完成任务, 从而实现更智能的人机交互。情感计算包含多个方面的研究, 如情感识别、情感合成、情感建模和情感表达。

情感计算的概念最初是在 1995 年由 Picard 提出, 他认为忽视情感的计算系统是片面的, 无法真正理解和适应人类的需求。虽然在此之前人机交互和人工智能领域的研究已经取得了一定进展, 但这

收稿日期: 2024-07-24; 修回日期: 2024-08-28

基金项目: 国家自然科学基金 (62306067)

作者简介: 刘颖, 博士, 副研究员, 主要从事社交媒体内容分析方面的研究。

\*通信作者 E-mail: 1054138110@qq.com

些研究主要集中在逻辑推理、数据处理和任务执行等方面, 情感因素往往被忽略<sup>[2]</sup>。随着研究的深入, 科学家们逐渐认识到情感在决策、学习和社交互动中的重要作用, 这促使他们开始探索如何将情感引入计算系统。

情感计算现如今已经在学术、医疗、娱乐、教育等领域得到了广泛关注<sup>[3]</sup>, 它结合了计算机科学、心理学、认知科学和神经科学等多个领域的知识, 旨在使计算机能够识别、理解、处理和模拟人类情感。情感识别在情感计算中扮演着至关重要的角色, 它是使计算机能够理解、处理和回应人类情感的核心技术<sup>[4]</sup>。

近几十年来, 研究人员探索了不同的方法来赋予机器类似人的情感状态感知。目前的自动情感识别方法可以根据所使用的信号分为两大类型。一种是使用人类的行为信号, 如面部表情<sup>[5-7]</sup>、语音<sup>[8]</sup>, 另一种方法则是使用生理信号。早期, 由于可穿戴技术和生理数据采集技术不够成熟, 研究者们通常采用非生理数据来进行情感识别工作。最早的人脸表情研究是将表情划分为6种情感<sup>[9]</sup>, 包括悲伤(Sad)、高兴(Happy)、恐惧(Fear)、厌恶(Disgust)、惊讶(Surprise)和愤怒(Angry)。并在之后的研究中通过对人脸面部表情的识别设计了基于运动单元(Action Unit, AU)的人脸表情动作解码系统(Facial Action Coding System, FACS), 通过分析面部运动单元的运动特征及其对应的面部区域来说明与之联系的相关表情<sup>[10]</sup>。文献[11]利用卷积神经网络(Convolutional Neural Networks, CNN)提出了一种多目标的时空特征表示学习方法, 通过结合空间和时间信息, 并优化多个目标, 显著提高了面部表情识别的鲁棒性和准确性, 特别是在表情强度变化的情况下, 展示了良好的适应能力和性能提升。文献[8]利用短时间对数频率功率系数来表示语音信号, 并利用离散隐马尔可夫模型进行语音情感识别。文献[12]使用深度神经网络(Deep Neural Network, DNN)架构的方法对语音进行了情感识别。文献[13]采用面部和语音混合的多模态信息判断被试的情感状态。

但是, 由于人们在特定情况下会隐藏自己的情感, 因此基于面部表情和语音的情感识别等信息来辨别一个人当前的情感会存在可信度低的情况。随着可穿戴技术的成熟, 研究者们逐渐采用用户的生

理数据进行情感识别, 如心电图(ECG)、脑电图(EEG)、肌电图(EMG)、皮肤电反应(GSR)和体温等。

与行为信号不同, 生理信号来源于中枢神经系统(Central Nervous System, CNS)和自主神经系统(Autonomic Nervous System, ANS)的活动, 这是一种非自愿的反应, 人们很难有意地对其进行控制或隐藏<sup>[14]</sup>。因此, 基于生理信号的情感识别方法在识别人类<sup>[15-16]</sup>的内在情感方面更为可靠和适用。

## 1 情感理论模型

情感分类提供了情感识别的理论框架和参考标准, 通过明确情感分类的类别和标准, 情感识别系统可以更有针对性地设计和优化识别算法。常见的情感理论模型如表1所示。

表1 常见情感理论模型

名称	类别	空间维度
基础情感理论 <sup>[9]</sup>	离散	二维
Plutchick情感轮盘 <sup>[17]</sup>	离散	二维
PAD模型 <sup>[18]</sup>	连续	三维
VA模型 <sup>[19]</sup>	连续	二维
VAD模型 <sup>[4]</sup>	连续	三维
PANA模型 <sup>[20]</sup>	连续	二维
ET模型 <sup>[21]</sup>	连续	二维
四维情感模型 <sup>[22]</sup>	连续	四维

### 1.1 基本情感理论

基本情感理论是最早的一种情感分类方法, 主要由心理学家Ekman在20世纪60年代提出<sup>[9]</sup>。他通过对跨文化研究发现, 人类面部表情有6种基本情感: 快乐、悲伤、愤怒、恐惧、惊讶和厌恶。他认为这些基本情感是普遍存在的, 不受文化影响。文献[23]指出人类和动物在表达情感时有许多相似之处, 这些表情具有生物学上的适应性功能。Ekman的基本情感理论不仅验证了Darwin的观点, 还通过实验研究提供了证据支持。

此外, 文献[17]也提出了另一种基本情感理论。他认为情感是由8种基本情感构成的, 具有强度、相似性和两极性3个维度, 这些情感可以两两结合, 形成复杂情感, 如图1所示。如愤怒和厌恶结合形成蔑视, 快乐和惊讶结合形成欣喜。文献[18]的情感轮模型通过几何图形直观地展示了情感之间的关系。



种生理信号密切相关,这种关系是双向互动的。在情感体验的过程中,身体会发生一系列生理变化,这些变化包括心率、血压、呼吸频率、皮肤电反应和荷尔蒙水平等,这些生理信号由自主神经系统调节,需要多个器官系统的协作,理解这些系统之间的相互关系是研究者们一直探索的课题<sup>[24]</sup>。

生理信号的变化不仅反映了情感的表层表现,更深层次地揭示了情感在人体内部的复杂反应机制。这种复杂的生理反应网络,为研究者提供了宝贵的数据基础,使得通过外部设备捕捉这些信号,进而理解和分析情感成为可能。随着技术的发展,尤其是可穿戴设备的进步,科学家们能够以更加精确和实时的方式记录和分析这些生理数据,进一步推动了情感识别领域的研究。

可穿戴传感器是近年来技术进步的重要领域之一,这些设备能够连续监测和收集用户生理数据,使得基于生理数据的情感识别得到了前所未有的发展机会<sup>[25]</sup>。神经科学理论研究表明,人的情感产生和变化与大脑活动息息相关,从大脑采集到的EEG包含了大量与情感变化相关的信息<sup>[26]</sup>。研究者们采用单一生理信号对情感进行识别,如文献<sup>[27]</sup>根据受试者在听音乐过程中自我报告的情感状态对脑电图动态进行分类,以探究情感状态与大脑活动之间的联系,并采用支持向量机(Support Vector Machine, SVM)对4种情绪状态(喜悦、愤怒、悲伤和愉悦)进行分类,在26名受试者中平均分类准确率为82.29%。文献<sup>[28]</sup>将多通道脑电图中提取的微分熵特征作为输入,将深度信念网络与隐马尔可夫模型进行集成,能够达到87.62%的准确率。文献<sup>[26]</sup>提出了一种基于经验模态分解的求解方法,用于检测心电图上动态演变的情绪模式,该研究表明当诱导方法对受试者有效时,心电图对情绪的反应性更高。文献<sup>[29]</sup>通过CNN和LSTM融合的方法对EEG信号进行分析,在SEED数据集上的识别准确率为93.74%。文献<sup>[30]</sup>使用深度卷积神经网络(Deep Convolutional Neural Network, DCNN)自动检测和诊断EEG信号中的癫痫发作,达到了88.76%的准确度。

生理信号在识别情感表达中起着关键作用。如心率的变化常与压力或兴奋相关<sup>[31]</sup>,皮肤电导水平的变化可以反映焦虑或放松状态<sup>[32-33]</sup>,而肌电图所记录的面部肌肉活动(如微笑或皱眉)电信号则直接表现出快乐或愤怒等情绪<sup>[34]</sup>。因此使用多种不同

的生理信号进行综合分析比只使用单一类型的生理信号能更准确地理解和识别个体的情感状态。文献<sup>[35]</sup>通过设计并实现了一个专门的实验方案诱导特定的情感状态,同时获取ECG、GSR、RESP这3种外周生理信号,采用非线性方法提取特征替代标准特征,提高了识别成功率。文献<sup>[36]</sup>通过视频对被试人员进行情感诱发,并采集他们的EEG、ECG和GSR,使用多种机器学习算法和深度学习模型对生理特征进行了情感和人格分类。文献<sup>[37]</sup>采用ECG和GSR的组合特征集对快乐、悲伤、中性3种情感状态进行分类,评估了GSR和ECG在情绪状态识别中的有效性。文献<sup>[38]</sup>提出了一种基于多层长短期记忆递归神经网络(LSTM-RNN)的具有两种注意力机制的情绪识别模型,该方法在Mahnob-HCI数据库上实现了更高的精度。文献<sup>[39]</sup>根据HR、ST、EDA信号,使用LSTM进行自然对话中的情感识别。文献<sup>[40]</sup>使用EEG和ECG对演奏音乐家进行情感识别,结果表明,EEG和ECG特征实现了最佳分类性能,相比单独使用ECG更准确。文献<sup>[41]</sup>利用ECG和EDA信号进行融合对音乐诱导的情感进行识别,比单模态方法准确率高4%~6%。

国内的情感识别研究起步较晚,相较于国外仍存在一定差距。然而,近年来随着生活水平的提升,心理和情绪问题在国内日益凸显,加之深度学习在各个领域的广泛应用,这些因素推动了国内在情感识别领域的研究。中科院情感研究小组研究了人类在人机交互过程中存在的情感识别、计算等问题,提出了面向特殊人群的表情识别模型方法,同时也研究了表情和情绪之间的内部机制和联系,并且开发了一套系统用来进行微表情识别<sup>[42]</sup>。文献<sup>[43]</sup>从时域和频域两个维度对ECG、EMG、GSR等信号进行特征融合,并利用ReliefF进行特征降维,能够达到98.4%的情感识别准确率。文献<sup>[44]</sup>采用自注意力机制和双向长短期记忆网络,对EEG和ECG等信号进行特征融合和分类,能够较为准确地识别情感。文献<sup>[45]</sup>融合了DEAP数据集中脑电信号、外周生理信号的生理信息特征以及它们的统计特征,将这三类特征进行融合分类,相比于单一的脑电信号特征分类精度提高了7%~10%。

### 3 基于多模态生理数据的情感识别

最近的研究发现,多种生理信号比单一信号<sup>[8]</sup>

能更有效地反映情感变化。不同的生理信号反映了情感状态的不同方面，并且在反映情感状态时具有不同的敏感性和特异性。通过结合多种生理信号，可以更好地适应个体差异，增强情感识别系统的泛化能力，因此目前很多研究者们采用多模态生理数据进行情感识别。基于多模态生理数据的情感识别流程如图 4 所示。

数据预处理之前要得到生理数据，一般需要对被试人员进行情感诱发，并采集被试人员的生理数据，但目前基于生理数据的情感识别已经有了较为深入的研究，国际上已经有一些公开的生理数据集，如 DEEP、SEED 等，因此很多研究者们直接采用生理数据集中的数据进行情感识别研究。表 3 概述了部分生理数据集，包括数据集名称、生理信号种类、被试人数。

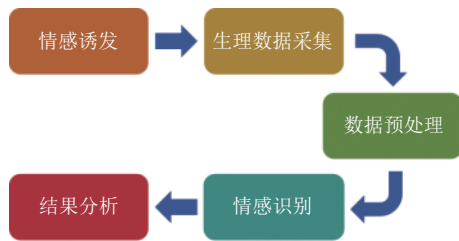


图 4 基于生理数据的情感识别流程图

表 3 生理数据集

数据集	生理信号种类	被试人数
DEAP <sup>[15]</sup>	EEG、EOG、EMG、GSR、RESP	32
MAHNOB-HCI <sup>[46]</sup>	EEG、ECG、ST、RESP	30
SEED <sup>[47]</sup>	EEG	15
MPED <sup>[48]</sup>	EEG、ECG、GSR、RSEP	23
AMIGOS <sup>[49]</sup>	EEG、GSR、ECG	40
WESAD <sup>[50]</sup>	EMG、RESP、ECG	15
ASCERTAIN <sup>[33]</sup>	EEG、ECG、GSR	58

在基于生理信号的情感识别流程中的数据采集部分，研究者们会采集多种人体的生理数据进行研究。之后对生理数据进行预处理并进行情感识别，在情感识别的过程中，一般有传统的机器学习方法和深度学习方法。

### 3.1 生理数据预处理

由于生理信号会存在来自于仪器和人体生理伪影产生的噪声，因此需要采用相应的数据去噪方法对生理数据进行去噪。目前常见的生理数据预处理方法有滤波法 (Filtering)、主成分分析法 (Principle Component Analysis, PCA)<sup>[51]</sup>、归一化法 (Normalization)、独立成分分析法 (Independent

Component Analysis, ICA)<sup>[52]</sup> 和经验模态分解 (Empirical Mode Decomposition, EMD)<sup>[53]</sup> 等。

滤波法是最简单直接的，通过低通滤波器滤除生理数据中 50 Hz 的设备本身所带来的工频干扰。文献 [54] 采用了滤波的方法对 ECG 进行了预处理去噪。归一化法通过将各个体的生理信号数据进行标准化处理，使得这些数据可以在同一尺度上进行比较和分析，从而消除个体之间的生理差异，将数据统一映射到 [0, 1] 区间，公式为：

$$\text{data}_{\text{normalized}} = \frac{\text{data}_n - \text{data}_{\text{min}}}{\text{data}_{\text{max}} - \text{data}_{\text{min}}} \quad (1)$$

式中， $\text{data}_{\text{normalized}}$  为数据归一化后的结果； $\text{data}_{\text{max}}$  为原始数据中的最大值； $\text{data}_{\text{min}}$  为原始数据中的最小值； $\text{data}_n$  为原始数据中某一个值。文献 [55] 在对 PPG 信号删除基线值后，采用归一化方法，从而消除个体之间的生理差异。文献 [56] 采用了 ICA 的方法，将复杂的多变量信号分解成多个相互独立的子成分，通过保留部分成分对生理信号进行了去噪。文献 [57] 采用了 EMD 方法将复杂的信号分解成多个本质上是振荡的本征模态函数 (Intrinsic Mode Functions, IMFs)，通过分解和去除信号中的高频噪声成分对 PPG 和 ECG 进行了去噪处理。

PCA 是一种常用的降维技术，旨在通过线性变换将原始数据从高维空间映射到低维空间，同时尽可能保留数据的主要特征和信息。通过对数据进行中心化处理，对于每个特征，减去其均值，使得数据的均值为零，以消除不同特征之间的量纲差异，使得数据分布更加集中。中心化后的数据可以用来计算协方差矩阵，协方差矩阵的每个元素表示两个特征之间的相关性。如果数据集有  $n$  个特征，那么协方差矩阵是一个  $n \times n$  的矩阵。对协方差矩阵进行特征值分解，得到特征值和对应的特征向量。特征值表示主成分的方差，特征向量表示主成分的方向。根据特征值的大小排序，选择前  $k$  个特征值所对应的特征向量，这些特征向量组成的子空间就是降维后的新特征空间，这些特征向量即为数据的主要成分。将原始数据投影到选定的主成分上，得到降维后的数据。这些新的特征 (即主成分) 是原始特征的线性组合，且这些主成分之间是正交的。

### 3.2 机器学习方法

基于多模态生理数据的情感识别流程中第一种

情感识别方法为机器学习方法。机器学习方法首先对经过预处理后的生理数据进行手工特征提取, 之后将提取到的不同生理数据特征进行简单的特征融合, 最后将融合后的生理数据特征通过特征分类器进行情感识别分类。

### 3.2.1 特征提取

生理信号中包含了大量的信息, 其中就有与情感相关的信息。生理信号常用的特征包括时域特征、频域特征和时频域特征。时域特征主要关注信号在时间上的变化, 频域特征主要关注信号频率成分, 而时频域特征能够同时描述信号在时间和频率上的变化, 这对于分析生理信号这种非平稳信号有着重要意义, 因为这种非平稳信号的频率成分也会随着时间变化。常用的提取时频域特征的方法有短时傅里叶变换 (Short-Time Fourier Transform, STFT)、小波变换 (Wavelet Transform, WT)<sup>[58]</sup> 和希尔伯特-黄变换 (Hilbert-Huang Transform, HHT)<sup>[53]</sup> 等。

短时傅里叶变换在傅里叶变换的基础上引入了窗函数, 通过对每个窗口内的信号  $x(t)\omega(t-\tau)$  进行傅里叶变换, 得到该窗口内的频谱, 这里的  $\tau$  表示窗口中心的时间位置:

$$X(\tau, f) = \int_{-\infty}^{\infty} x(t)\omega(t-\tau)e^{-j2\pi ft} dt \quad (2)$$

式中,  $X(\tau, f)$  是在时间  $\tau$  和频率  $f$  处的 STFT 系数;  $x(t)$  是原始信号;  $\omega(t)$  是窗口函数;  $f$  是频率。虽然短时傅里叶变换在每个时间窗口内, 信号被假设为平稳的, 适用于分析非平稳信号。但窗口长度越短, 时间分辨率越高, 频率分辨率越低, 反之亦然。因此, 需要选择合适的时间窗长度。

小波变换是对 STFT 方法的一种改进, 小波变换的窗函数会随着频率的变化而改变, 因此相比 STFT, 在时域和频域都能够达到较好的分辨率。小波变换通过使用具有有限长度并且会衰减的小波基, 取代了傅里叶变换中的无限长三角函数基。 $\varphi(t)$  为基本小波函数, 将  $\varphi(t)$  进行尺度和平移变换, 尺度对应于频率, 平移量对应于时间, 得到小波变换函数:

$$WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\varphi\left(\frac{t-\tau}{a}\right) dt \quad (3)$$

式中,  $a$  代表尺度;  $\tau$  代表平移量。小波变换虽然能够提供多分辨率分析, 但其时间和频率分辨率之

间的权衡 (即不确定性原理) 仍然存在限制。在高频区域, 具有较高的时间分辨率和较低的频率分辨率; 在低频区域, 具有较高的频率分辨率和较低的时间分辨率。

HHT 结合了 EMD 和希尔伯特变换 (Hilbert Transform), 能够有效地处理复杂信号, 提取其瞬时频率和能量分布。该方法首先采用 EMD 将复杂信号分解为一组振幅-频率调制的分量, 即 IMFs。IMF 需要满足极值个数与过零点个数相等或相差至多一个, 且由极值和极值定义的包络线的均值必须为零。因此输入的原始生理信号可以表示为:

$$f(t) = \sum_{i=1}^k IMF_i(t) + r(t) \quad (4)$$

式中,  $f(t)$  表示输入的生理信号;  $IMF_i(t)$  表示本征模函数;  $r(t)$  表示生理信号减去 IMF 后的余项。

希尔伯特变换为:

$$IMF_i(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{IMF_i(\tau)}{t-\tau} d\tau \quad (5)$$

HHT 在经验模态分解过程中会出现不同频率成分的混叠现象, 导致分解结果不理想。

### 3.2.2 特征分类

经过特征提取之后, 通过简单拼接或加权平均方法对不同的生理数据特征进行简单融合。之后通过分类方法对提取的特征进行情感特征分类, 常用的传统机器学习分类方法有 SVM<sup>[59]</sup>、K-近邻算法 (K-Nearest Neighbor, KNN)<sup>[60]</sup>、朴素贝叶斯 (Naive Bayes, NB)<sup>[61]</sup>、随机森林 (Random Forest, RF)<sup>[62]</sup> 等。

SVM 是一种用于分类和回归分析的非概率的两元线性分类器。SVM 突出的泛化能力, 连同其最优解和判别力, 现在已经广泛应用于二元分类或多元分类<sup>[63]</sup>。SVM 的原理是找到一个能够将不同类别的数据点分开的超平面 (在二维情况下是直线), 同时最大化分类边界两侧到超平面的最小距离, 如图 5 所示, 上间隔边界上方的样本属于正类, 下间隔边界下的样本属于负类。位于间隔边界上的正类和负类样本为支持向量 (support vector), 两个间隔边界的距离被定义为边距。

SVM 能够处理高维数据且在小样本情况下表现良好, 有明确的几何解释, 能够提供一个最优的分类边界, 但在大样本数据上存在困难。

KNN 是一种广泛用于分类和回归任务的简单、直观的机器学习算法。它的基本思想是：给定一个新样本，找到距离它最近的  $K$  个样本点，然后根据这  $K$  个样本点的类别来决定新样本的类别：

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad i = 1, 2, \dots, n \quad (6)$$

KNN 是情感识别研究中最简单有效的机器学习方法之一。文献 [64] 通过诱发被试情感，采集被试人员的生理信号数据，之后采用 K-近邻算法对害怕、愤怒、悲伤等情绪进行分类识别。K-近邻算法可以处理多分类问题，而不仅限于二分类。但是 K-近邻算法需要计算所有样本点的距离，所以在处理大规模数据集时，计算开销很大。

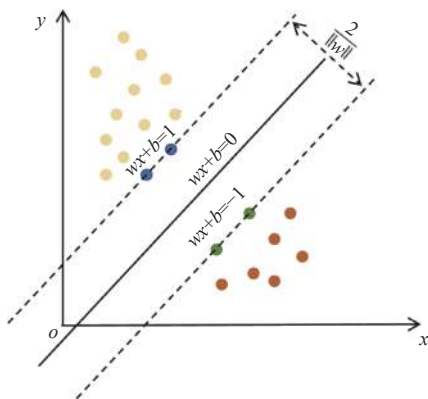


图 5 线性支持向量机模型

### 3.3 深度学习的方法

另一种情感识别方法则是深度学习的方法。尽管机器学习的方法在情感识别上表现出了较好的效果，但深度学习技术在情感识别领域更具有优势<sup>[65]</sup>。深度学习能够自动提取和学习复杂的特征，而不需要人为设计特征，这使得其在处理复杂的情感识别任务时表现更加优异。此外，深度学习模型可以在大量的数据中自动发现情感特征的潜在模式，有效捕捉到生理数据中所隐含的情感特征间的关联性，从而提高识别的准确率和鲁棒性。此外，深度学习还具备处理大规模数据和实时数据分析的能力，这使得情感识别系统能够在大数据环境中学习和预测情感状态，从而在实时应用中具有更高的实用性。

#### 3.3.1 特征分类

CNN 是一种前馈神经网络，其特点是每个神

经元只与前一层的部分神经元相连接，这个小部分的连接区域被称为感受野。局部连接使得网络能够专注于输入数据的局部特征，从而提取出有用的局部信息。这对于生理信号中的时间序列数据来说特别有用，因为生理信号的特征往往是局部的。同时，卷积操作的平移不变性使得卷积神经网络对输入信号的平移具有鲁棒性，这使得生理信号的特征可能会在时间轴上有所移动，但它们的模式和结构保持不变。

卷积层是卷积神经网络的核心，能够提取出数据隐含的特征。卷积层通过卷积操作，将输入数据与多个可学习的卷积核进行卷积计算，提取局部特征。对于输入数据  $X$  和卷积核  $W$ ，卷积操作的输出  $Y$  表示为：

$$Y[i, j] = \sum_m \sum_n X[i+m, j+n] W[m, n] \quad (7)$$

在卷积操作后，通常会使用非线性激活函数（如 ReLU）对卷积输出进行非线性转换，从而增强模型的表达能力。对于卷积输出  $Y$ ，ReLU 激活函数的输出  $Z$  表示为：

$$Z[i, j] = \text{ReLU}(Y[i, j]) = \max(0, Y[i, j]) \quad (8)$$

之后用池化层进行下采样，进一步压缩数据量，同时保留重要的信息，进而降低计算复杂度，并增强模型对位置变化的鲁棒性。文献 [66] 提出一种融合二维卷积神经网络模型，提取脑电信号特征用于情感识别。

除此之外，递归神经网络（Recurrent Neural Network, RNN）<sup>[67]</sup> 也常常被用于生理信号数据的特征提取，这是因为 RNN 是一种擅长处理序列数据的神经网络，它能够捕捉到生理信号中的时间依赖关系，但也存在一些问题，如梯度消失、爆炸，而 LSTM 成功解决了 RNN 的这些问题。LSTM 通过引入输入门、遗忘门和输出门等机制，有效地保留和利用长期依赖信息，因此在处理生理数据任务中表现出色<sup>[68]</sup>。LSTM 网络模型如图 6 所示。

遗忘门决定了上一时刻的单元状态有多少保留到单元状态  $c_t$ ， $\sigma$  是 sigmoid 函数， $W_f$  是遗忘门的权重矩阵， $[h_{t-1}, x_t]$  表示把两个向量连接成一个向量， $b_f$  是遗忘门的偏置项：

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

输入门决定了有多少输入数据  $x_t$  保存到单元状

态  $c_t$ ,  $W_i$  是输入门的权重矩阵,  $b_i$  是输入门的偏置项:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (10)$$

之后根据当前输入和先前的隐藏状态来计算当前输入的单元状态  $\tilde{C}_t$ :

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (11)$$

结合遗忘门和输入门的信息, 更新当前时刻的单元状态:

$$c_t = f_t * c_{t-1} + i_t * \tilde{C}_t \quad (12)$$

式中,  $*$  表示按元素乘。这就将当前记忆  $\tilde{C}_t$  和长期记忆  $c_{t-1}$  组合到了一起, 形成了新的单元状态  $c_t$ 。遗忘门保存先前的信息, 输入门控制当前的输入信息, 输出门控制长期记忆对当前输出的影响:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (13)$$

而输出门和单元状态决定了 LSTM 的最终输出:

$$h_t = o_t * \tanh(C_t) \quad (14)$$

文献 [69] 使用 LSTM 从 EEG 信号中提取情感特征, 并通过全连接层将这些特征分类为低/高唤醒度、效价和喜好, 该方法在 DEAP 数据集上平均准确率达到 85.65%、85.45% 和 87.99%。

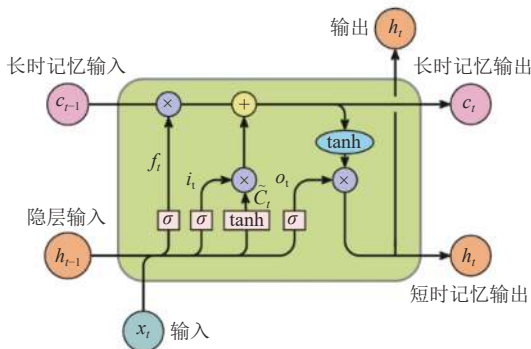


图 6 LSTM 网络模型

### 3.3.2 特征融合

多模态融合可以根据融合时间的先后分为早期的数据层融合、中期的特征层融合以及后期的决策层融合。

早期数据层融合是在模型的早期阶段将来自不同模态的数据经过预处理后进行合并, 将合并后的数据输入到统一的模型中提取多模态特征, 如图 7

所示。

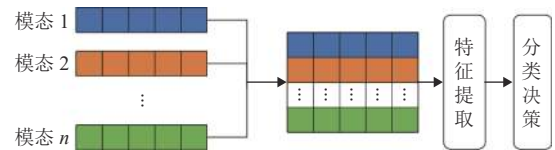


图 7 早期融合

中期特征层融合是用各自的神经网络或特定的特征提取算法对不同模态的数据进行独立的特征提取, 之后将不同模态的特征进行融合, 能够保留个模态数据的独特信息, 如图 8 所示。文献 [70] 通过 CNN 提取脑电信号情感信息, 之后利用 LSTM 提取外周生理信号情感信息, 通过中期融合进行特征融合。

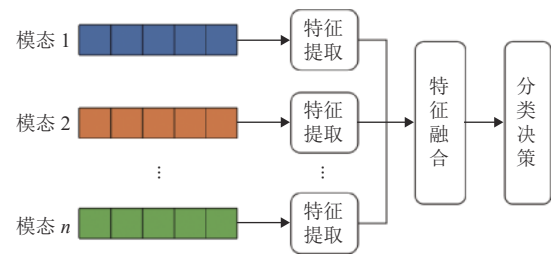


图 8 中期融合

后期决策层融合是先对每个模态的数据独立进行处理和分析, 生成各自的预测结果或决策, 然后再对这些独立的决策结果进行融合, 以得到最终的决策, 如图 9 所示。文献 [71] 提出了一种混合特征提取方法, 从生理信号中提取统计相关特征和时间相关特征, 之后采用自适应决策融合方法, 整合信号模态进行情感分类。

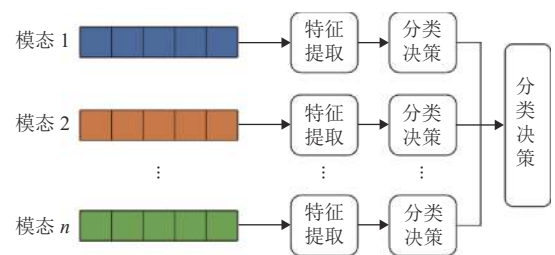


图 9 后期融合

在 3 种多模态数据融合中, 早期数据层融合是将原始生理数据直接进行融合, 这种方法对各模态生理数据全部信息进行特征提取, 能够充分利用生理数据的完整信息, 但这种方式存在较多冗余信息, 很难提取到有效的特征信息<sup>[72]</sup>。中期特征层融合是对各模态生理数据进行特征提取后进行特征融

合, 该方法不但减少了生理数据的冗余信息, 并且为不同生理数据选择合适的特征提取模型, 更好地捕捉生理数据的信息特征<sup>[73]</sup>, 通过融合实现不同模态生理数据信息之间的局部交互, 得到融合特征用于最后决策, 这也是大多数研究者选择的方式。后期决策层融合仅仅是在分类决策时考虑到了不同模态生理数据信息之间的相关性, 没有充分利用跨模态的局部和全局关联信息<sup>[74]</sup>。

## 4 结束语

基于多模态生理数据的情感识别是一项具有重要意义且充满挑战的研究领域。多模态情感识别系统结合了多种生理信号(如 ECG、EEG、GSR 等)以提高情感识别的准确性和鲁棒性。以下是该领域面临的主要挑战及未来展望。

### 4.1 生理数据采集设备

大多数用于基于生理数据的情感研究的情感数据库(如 DEAP, MAHNOB-HCI)中的数据获取通常依赖于医疗级的设备, 因为其功能支持性, 如高灵敏度和高采样率。虽然高质量的生理数据可以通过这种方式收集, 但这些设施昂贵且部署起来较为复杂, 阻碍了情感识别技术被整合到日常设备中, 并被部署到现实环境中。而便携式生理采集设备虽然便于穿戴, 能够用在日常生活中, 但这些设备所采集的生理数据精度也较低, 且数据信息也不够全面。设计和制造便携、易用的采集设备, 提高设备的集成度以及传感器的精度和抗噪能力, 是基于多模态生理数据的情感识别发展的必然趋势。

### 4.2 个体差异

基于多模态生理数据的情感识别中的个体差异影响是一个复杂而重要的问题, 显著影响着情感识别系统的准确性和泛化能力。个体之间的生理差异主要体现在生理信号的基线水平、响应模式和变化幅度上。不同个体的情感表达方式和强度也存在显著差异, 同样的情感刺激可能在不同个体中引发不同的生理反应, 这增加了情感识别模型的复杂性和不确定性。深入理解和有效处理个体差异, 建立通用的情感模型, 将是基于多模态生理数据的情感识别发展方向之一。

### 4.3 计算复杂度

多模态生理数据融合后通常会导致高维特征空

间, 增加了计算复杂度和模型训练难度。在实际应用中, 情感识别系统往往需要在低延迟、高效率的条件下运行, 尤其是在交互式应用中, 因此要求高效的算法和计算资源。开发计算高效的情感识别算法, 有效控制和优化计算复杂度, 满足实时处理要求, 将是基于多模态生理数据的情感识别另一发展方向。

## 参考文献

- [1] BARRETT L F, MESQUITA B, OCHSNER K N, et al. The experience of emotion[J]. *Annual Review of Psychology*, 2007, 58: 373-403.
- [2] LANDAUER T K. Research methods in human-computer interaction[M]//Handbook of Human-Computer Interaction. Amsterdam: Elsevier, 1988: 905-928.
- [3] MAJOR T C, CONRAD J M. A survey of brain computer interfaces and their applications[C]//Proceedings of the IEEE SOUTHEASTCON. New York: IEEE, 2014: 1-8.
- [4] COWIE R, DOUGLAS-COWIE E, TSAPATSOLIS N, et al. Emotion recognition in human-computer interaction[J]. *IEEE Signal Processing Magazine*, 2001, 18(1): 32-80.
- [5] COHEN I, GARG A, HUANG T S. Emotion recognition from facial expressions using multilevel HMM[C]//Neural Information Processing Systems. State PA: Citeseer, 2000, 2.
- [6] SARSENBAYEVA Z, MARINI G, VAN BERKEL N, et al. Does smartphone use drive our emotions or vice versa? a causal analysis[C]//Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2020: 1-15.
- [7] YANG K N, WANG C F, SARSENBAYEVA Z, et al. Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets[J]. *The Visual Computer*, 2021, 37(6): 1447-1466.
- [8] NWE T L, FOO S W, DE SILVA L C. Speech emotion recognition using hidden Markov models[J]. *Speech Communication*, 2003, 41(4): 603-623.
- [9] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion[J]. *Journal of Personality and Social Psychology*, 1971, 17(2): 124-129.
- [10] EKMAN P, FRIESEN W V. Facial action coding system (FACS): A technique for the measurement of facial actions[J]. *Rivista Di Psichiatria*, 1978, 47(2):126-138.
- [11] KIM D H, BADDAR W J, JANG J, et al. Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition[J]. *IEEE Transactions on Affective Computing*, 2019, 10(2): 223-236.
- [12] HARAR P, BURGET R, DUTTA M K. Speech emotion recognition with deep learning[C]//Proceedings of the 4th International Conference on Signal Processing and Integrated Networks. New York: IEEE, 2017: 137-140.

- [13] DE SILVA L C, MIYASATO T, NAKATSU R. Facial emotion recognition using multi-modal information [C]//Proceedings of the Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications. New York: IEEE, 1997: 397-401.
- [14] SHU L, XIE J Y, YANG M Y, et al. A review of emotion recognition using physiological signals[J]. *Sensors*, 2018, 18(7): 2074.
- [15] KOELSTRA S, MUHL C, SOLEYMANI M, et al. DEAP: A database for emotion analysis: Using physiological signals[J]. *IEEE Transactions on Affective Computing*, 2012, 3(1): 18-31.
- [16] ZHU J J, WEI Y X, FENG Y F, et al. Physiological signals-based emotion recognition via high-order correlation learning[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2019, 15(3s): 1-18.
- [17] PLUTCHIK R, KELLERMAN H. Theories of emotion[J]. Academic Press, 1980, DOI:10.7551/mitpress/10398.003.0004.
- [18] MEHRABIAN A, RUSSELL J A. An approach to environmental psychology[M]. Cambridge, Mass: MIT 1974.
- [19] RUSSELL J A. A circumplex model of affect[J]. *Journal of Personality and Social Psychology*, 1980, 39(6): 1161.
- [20] WATSON D, TELLEGEN A. Toward a consensual structure of mood[J]. *Psychological Bulletin*, 1985, 98(2): 219-235.
- [21] THAYER R E. Toward a psychological theory of multidimensional activation (arousal)[J]. *Motivation and Emotion*, 1978, 2(1): 1-34.
- [22] IZARD C E. The psychology of emotions[M]. New York: Plenum Press, 1991.
- [23] 达尔文. 人类和动物的表情[M]. 周邦立, 译. 北京: 科学出版社, 1958.  
DARWIN C. The expression of the emotions in man and animals[M]. translated by ZHOU B L. Beijing; Science Press, 1958.
- [24] LI L, CHEN J H. Emotion recognition using physiological signals from multiple subjects[C]//Proceedings of the International Conference on Intelligent Information Hiding and Multimedia. New York: IEEE, 2006: 355-358.
- [25] KIM D G, HONG K S, CHUNG K W. Implementation of portable multi-channel EEG and head motion signal acquisition system[C]//Proceedings of the 8th International Conference on Computing and Networking Technology. New York: IEEE, 2012: 370-375.
- [26] AGRAFIOTI F, HATZINAKOS D, ANDERSON A K. ECG pattern analysis for emotion detection[J]. *IEEE Transactions on Affective Computing*, 2012, 3(1): 102-115.
- [27] LIN Y P, WANG C H, JUNG T P, et al. EEG-based emotion recognition in music listening[J]. *IEEE Transactions on Bio-Medical Engineering*, 2010, 57(7): 1798-1806.
- [28] ZHENG W L, ZHU J Y, PENG Y, et al. EEG-based emotion classification using deep belief networks[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. New York: IEEE, 2014: 1-6.
- [29] RAMZAN M, DAWN S. Fused CNN-LSTM deep learning emotion recognition model using electroencephalography signals[J]. *The International Journal of Neuroscience*, 2023, 133(6): 587-597.
- [30] ACHARYA U R, OH S L, HAGIWARA Y, et al. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals[J]. *Computers in Biology and Medicine*, 2018, 100: 270-278.
- [31] QUINTANA D S, GUASTELLA A J, OUTHRED T, et al. Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition[J]. *International Journal of Psychophysiology*, 2012, 86(2): 168-172.
- [32] LANG P J, GREENWALD M K, BRADLEY M M, et al. Looking at pictures: Affective, facial, visceral, and behavioral reactions[J]. *Psychophysiology*, 1993, 30(3): 261-273.
- [33] SUBRAMANIAN R, WACHE J, ABADI M K, et al. ASCERTAIN: Emotion and personality recognition using commercial sensors[J]. *IEEE Transactions on Affective Computing*, 2018, 9(2): 147-160.
- [34] NAKASONE A, PRENDINGER H, ISHIZUKA M. Emotion recognition from electromyography and skin conductance[C]//Proc of the 5th International Workshop on Biosignal Interpretation. [S.l.]: Citeseer, 2005: 219-222.
- [35] VALENZA G, LANATA A, SCILINGO E P. The role of nonlinear dynamics in affective valence and arousal recognition[J]. *IEEE Transactions on Affective Computing*, 2012, 3(2): 237-249.
- [36] ABADI M K, SUBRAMANIAN R, KIA S M, et al. DECAF: MEG-based multimodal database for decoding affective physiological responses[J]. *IEEE Transactions on Affective Computing*, 2015, 6(3): 209-222.
- [37] DAS P, KHASNOBISH A, TIBAREWALA D N. Emotion recognition employing ECG and GSR signals as markers of ANS[C]//Proceedings of the Conference on Advances in Signal Processing. New York: IEEE, 2016: 37-42.
- [38] LIU J M, SU Y Q, LIU Y H. Multi-modal emotion recognition with temporal-band attention based on LSTM-RNN[C]//Pacific Rim Conference on Multimedia. Cham: Springer, 2018: 194-204.
- [39] ZITOUNI M S, PARK C Y, LEE U, et al. LSTM-modeling of emotion recognition using peripheral physiological signals in naturalistic conversations[J]. *IEEE*

- Journal of Biomedical and Health Informatics*, 2023, 27(2): 912-923.
- [40] TURCHET L, O'SULLIVAN B, ORTNER R, et al. Emotion recognition of playing musicians from EEG, ECG, and acoustic signals[J]. *IEEE Transactions on Human-Machine Systems*, 2024 (99): 1-11.
- [41] MU J J, QIAO Y H, LIU G Y. Research on emotion recognition strategy based on electrocardiogram and electrodermal activity signals induced by music[C]// *Proceedings of the 5th International Seminar on Artificial Intelligence, Networking and Information Technology*. New York: IEEE, 2024: 1574-1578.
- [42] WU Q, SHEN X, FU X. The machine knows what you are hiding: an automatic micro-expression recognition system[C]// *4th International Conference on Affective Computing and Intelligent Interaction*. Memphis: [s.n.], 2011: 152-162.
- [43] 潘礼正, 尹泽明, 余世刚, 等. 基于 FCA-ReliefF 的融合生理信号情绪识别研究[J]. *计算机测量与控制*, 2020, 28(2): 179-183.
- PAN L Z, YIN Z M, SHE S G, et al. Emotion recognition based on physiological signal fusion and FCA-ReliefF[J]. *Computer Measurement & Control*, 2020, 28(2): 179-183.
- [44] 陈田, 蔡从虎, 袁晓辉, 等. 基于多尺度卷积和自注意力特征融合的多模态情感识别方法[J]. *计算机应用*, 2024, 44(2): 369-376.
- CHEN T, CAI C H, YUAN X H, et al. Multimodal emotion recognition method based on multiscale convolution and self-attention feature fusion[J]. *Journal of Computer Applications*, 2024, 44(2): 369-376.
- [45] 马壮, 甘开宇, 尹钟. 基于脑电信号和周围生理信号的多模态融合情感识别[EB/OL]. [2024-05-12]. <https://doi.org/10.16180/j.cnki.issn1007-7820.2025.02.008>.
- MA Z, GAN K Y, YIN Z. Emotion recognition based on multimodal fusion of the EEG and peripheral physiological signals[EB/OL]. [2024-05-12]. <https://doi.org/10.16180/j.cnki.issn1007-7820.2025.02.008>.
- [46] 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.
- [47] DUAN R N, ZHU J Y, LU B L. Differential entropy feature for EEG-based emotion classification[C]// *Proceedings of the 6th International IEEE/EMBS Conference on Neural Engineering*. New York: IEEE, 2013: 81-84.
- [48] SONG T F, ZHENG W M, LU C, et al. MPED: A multimodal physiological emotion database for discrete emotion recognition[J]. *IEEE Access*, 2019, 7: 12177-12191.
- [49] MIRANDA-CORREA J A, ABADI M K, SEBE N, et al. AMIGOS: A dataset for affect, personality and mood research on individuals and groups[J]. *IEEE Transactions on Affective Computing*, 2021, 12(2): 479-493.
- [50] SCHMIDT P, REISS A, DUERICHE R, et al. Introducing WESAD, a multimodal dataset for wearable stress and affect detection[C]// *Proceedings of the Proceedings of the 20th ACM International Conference on Multimodal Interaction*. New York: ACM, 2018: 400-408.
- [51] KARL P. On lines and planes of closest fit to systems of points in space[J]. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1901, 2(11): 559-572.
- [52] COMON P. Independent component analysis, A new concept?[J]. *Signal Processing*, 1994, 36(3): 287-314.
- [53] HUANG N E, SHEN Z, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. *Proceedings of the Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, 1998, 454(1971): 903-995.
- [54] CHETTUPUZHAKKARAN P, SINDHU N. Emotion recognition from physiological signals using time-frequency analysis methods[C]// *Proceedings of the International Conference on Emerging Trends and Innovations in Engineering and Technological Research*. New York: IEEE, 2018: 1-5.
- [55] LEE M S, LEE Y K, PAE D S, et al. Fast emotion recognition based on single pulse PPG signal with convolutional neural network[J]. *Applied Sciences*, 2019, 9(16): 3355.
- [56] PAMUNGKAS Y, WIBAWA A D, RAIS Y. Classification of emotions (positive-negative) based on EEG statistical features using RNN, LSTM, and Bi-LSTM algorithms[C]// *Proceedings of the 2nd International Seminar on Machine Learning, Optimization, and Data Science*. New York: IEEE, 2022: 275-280.
- [57] SHAHID H, BUTT A, AZIZ S, et al. Emotion recognition system featuring a fusion of electrocardiogram and photoplethysmogram features[C]// *Proceedings of the 14th International Conference on Open Source Systems and Technologies*. New York: IEEE, 2020: 1-6.
- [58] GROSSMANN A, MORLET J. Decomposition of hardy functions into square integrable wavelets of constant shape[J]. *SIAM Journal on Mathematical Analysis*, 1984, 15(4): 723-736.
- [59] CORTES C, VAPNIK V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [60] COVER T, HART P. Nearest neighbor pattern classification[J]. *IEEE Transactions on Information Theory*, 1967, 13(1): 21-27.
- [61] WEBB G I. Naïve Bayes[M]// *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2011: 713-714.
- [62] BREIMAN L. Random forests[J]. *Machine learning*, 2001, 45: 5-32.
- [63] LIANG X P, ZHU L, HUANG D S. Multi-task ranking SVM for image cosegmentation[J]. *Neurocomputing*, 2017, 247: 126-136.

- [64] NASOZ F, ALVAREZ K, LISETTI C L, et al. Emotion recognition from physiological signals using wireless sensors for presence technologies[J]. *Cognition, Technology & Work*, 2004, 6(1): 4-14.
- [65] TANG H, LIU W, ZHENG W L, et al. Multimodal emotion recognition using deep neural networks[C]// *Proc Int Conf Neural Inf Process*. [S.l.]: Springer, 2017: 811-819.
- [66] KWON Y H, SHIN S B, KIM S D. Electroencephalography based fusion two-dimensional (2D)-convolution neural networks (CNN) model for emotion recognition system[J]. *Sensors*, 2018, 18(5): 1383.
- [67] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[EB/OL]. [2024-03-21]. <http://arxiv.org/abs/1409.2329v5>.
- [68] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [69] ALHAGRY S, ALY A, REDA A. Emotion recognition based on EEG using LSTM recurrent neural network[J]. *International Journal of Advanced Computer Science and Applications*, 2017, DOI:10.14569/IJACSA.2017.081046.
- [70] LIAO J X, ZHONG Q H, ZHU Y S, et al. Multimodal physiological signal emotion recognition based on convolutional recurrent neural network[J]. *IOP Conference Series: Materials Science and Engineering*, 2020, 782(3): 032005.
- [71] YAN M S, DENG Z, HE B W, et al. Emotion classification with multichannel physiological signals using hybrid feature and adaptive decision fusion[J]. *Biomedical Signal Processing and Control*, 2022, 71: 103235.
- [72] MARTÍNEZ H P, YANNAKAKIS G N. Deep multimodal fusion: Combining discrete events and continuous signals[C]// *Proceedings of the Proceedings of the 16th International Conference on Multimodal Interaction*. New York: ACM, 2014: 34-41.
- [73] ZHANG L, ZHANG D. Visual understanding via multi-feature shared learning with global consistency[J]. *IEEE Transactions on Multimedia*, 2016, 18(2): 247-259.
- [74] MURPHY R R. Computer vision and machine learning in science fiction[J]. *Science Robotics*, 2019, DOI: 10.1126/scirobotics.aax742.

编辑 税红