

特色专题

2024 年人机情感交互领域热点回眸

谷雨, 任福继*

摘要 “人机情感交互”研究旨在创建一种能感知、识别和理解人的情感,并能针对人的情感作出智慧、灵敏、自然回应的智能系统,是实现机器人自然化、拟人化、人格化的基础技术和重要前提。2024年,大语言模型的进步给该领域带来了深刻的变革。从交互理论、情感认知、情感调节及交互效果评估层面回眸该领域年度研究进展,并探讨在情感认知可计算模型、情感调控可进化交互环路、情感交互可解释评估体系等方面存在的理论与技术问题。指出了人机情感交互领域未来研究主要方向为情感认知可计算建模方法、情感调控可进化交互环路、情感交互可解释评估体系等。

关键词 人机情感交互;多模态大模型;情感认知;情感调节;交互效果评估

情感被誉为人类社会生活的文法(grammar of social living),也是人类作为高级智能体的主要表征之一。如何赋予机器以类人智能?1970年,人工智能创始人之一、ACM(美国计算机协会)图灵奖获得者、麻省理工学院教授明斯基(Minsky)在《脑智社会》(《The Society of Mind》)专著中就智能与情感的辩证关系提出了论断:“问题不在于智能机器能否有情感,而在于没有情感的机器能否实现智能(The question is not whether intelligent machines can have any emotions, but whether machines can be intelligent without emotions)”。该问题开启了计算机学界赋予智能机器人情感自然交互能力的漫漫征程。

2024年,以ChatGPT为代表的大语言模型(large language model, LLM)在交互能力上有了长足的进步,出现了Emohaa^[1]、MeChat^[2]等具有一定情感支撑能力的对话系统。这背后的原因在于诸如情绪生成理论^[3]、情绪支撑方法^[4]等理论和方法层面的进步。

从社会工程学出发,详述交互理论、情感认知、情感调节及情感交互能力评测这一情感交互闭环在2024年的研究回眸。

1 人机情感交互领域研究回眸

1.1 交互理论重要进展

人机情感交互理论研究自然人与机器(包括机器人)之间通过语言(文字、语音)与非语言(面部表情、手势、体姿、心跳、呼吸、皮肤电等)通道来相互作用与传递情感信息的理论框架与方法体系。情感交互理论是认知-调节-评估人机情感交互环路的理论基础,具有指导这3个环节研究方向与技术导向的重要意义。Ren等^[5]在2024年提出了“可进化心状态转移模型”(evolutionary mental state transition model, EMSTM),旨在通过可观察行为和历史情感来模拟追踪自然人的心理状态转变过程。如图1所示,它拓展了“心状态转移网络”^[6],融合了模式识别方法,创新地定义了情感表达强度(external emotional energy,

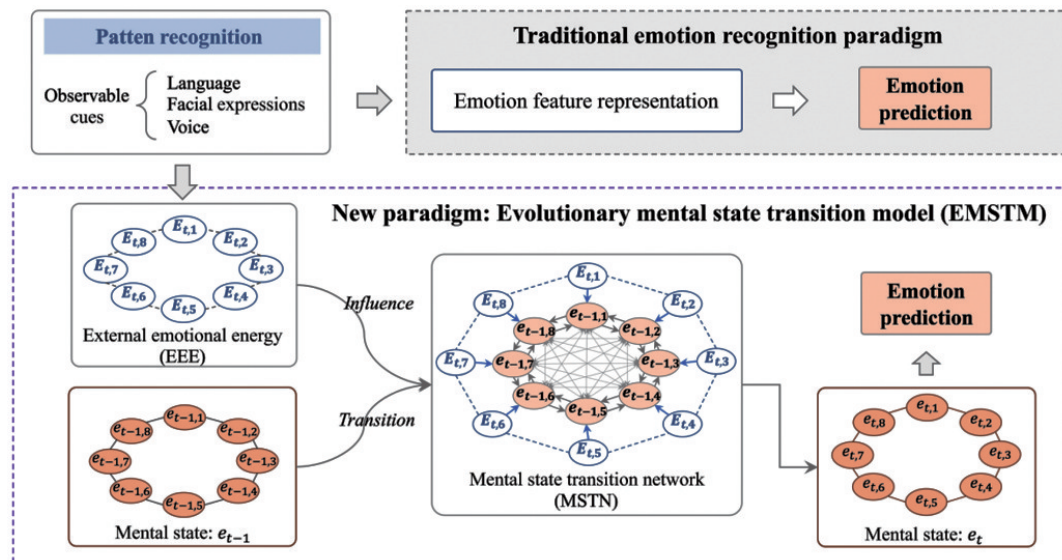
电子科技大学计算机科学与工程学院,成都 611731

收稿日期:2024-12-03;修回日期:2025-01-05

基金项目:国家自然科学基金联合重点项目(U24A20250)

作者简介:谷雨,研究员,研究方向为情感计算、人机交互、普适计算等,电子信箱:guyu@uestc.edu.cn;任福继(通信作者),教授,日本工程院院士、欧盟科学院院士、俄罗斯工程院外籍院士,研究方向为先进智能、情感计算、智能机器人等,电子信箱:renfuji@uestc.edu.cn

引用格式:谷雨,任福继. 2024年人机情感交互领域热点回眸[J]. 科技导报, 2025, 43(1): 132-142; doi: 10.3981/j.issn.1000-7857.2025.01.00035



与传统情感识别范式依赖情感表达的特征工程不同,该理论通过定义情感表达强度这一概念,以统一刻画不同模态情感外在表达

图1 可进化心状态转移模型

EEE)这一概念,用以统一刻画不同模态情感外在表达,奠定了人机情感交互的可计算理论基础。

在此框架下,美国俄勒冈州立大学的 Sanchez 等^[7]着重关注了机器人手势与自然人情感在公共交互场景下的关联。结果显示,交互中非言语线索对自然人的情绪有着显著影响,这揭示了机器人手部动作与人类情感体验之间的复杂相互作用。此外,鉴于体态在人类社交情绪表达和识别中的关键作用,法国巴黎综合理工学院的 Bal 等^[8]探索了如何从机器人自带的传感器中获取人类手臂的姿势和运动信息,以便在人机物理交互场景中区分人类的情感状态。具体来说,他们关注的是情感状态的“力量”维度,它提供了关于在特定交互情境下控制感的信息。该研究中“力量”维度的定义与 EMSTM 模型中 EEE 概念高度契合。

此外,作为影响交互内容、方式与结果重要因素的交互场景也得到了重视。2024年10月,美国罗切斯特理工学院的 Xu 等^[9]针对工业场景,探讨了将增强现实技术融入人机协作以提高情境感受的可能性。该方向构成了任福继 EMSTM 模型的自然扩展。

交互理论的进展为人机情感交互带来了新的理论支撑及分析手段。在此基础上,人机情感交互的关键技术在2024年也有了稳健的进步。

1.2 情感认知

《Science》2002年11月发表题为《Emotion, cognition, and behavior》的论文中指出,情感认知能力在人类社会中泛指感知、识别和理解他人情感的能力^[10]。但受限于对人机交互机理的认知不够清晰,在计算机视域下该能力收敛于“情感计算”。该概念由麻省理工学院媒体实验室 Picard^[11]于1997年提出,旨在以计算为手段,研究和开发能够识别、解释和处理人类情感系统的一个跨学科领域,涉及计算机科学、心理学和认知科学。

情感计算的基础是情感量化模型^[12],主要用于将无形的、抽象的人类情感转化为机器人可以理解、表达和处理的数据形式,可细分为离散模型与连续模型。前者将情感状态描述为离散形式,即基本情绪,如快乐、悲伤、愤怒、惊讶等。基本情绪被认为是先天的,具有独立的神经生理学机制、内在经验和外在表现,以及不同文化、场景等的情境适应性。目前使用最广泛的是心理学家 Ekman^[13]于1992年提出的6分类模型(怒、恐、恶、喜、悲、惊)。离散模型在计算层面有2个问题:同类别情绪可比较性差和跨类别的关联性差。为此,后者(连续模型)从1个或多个维度来定义情绪。1897年,现代心理学奠基人 Wundt 提出,情绪可以通过3个维度来描述:“愉快的与不愉快

的“激起或抑制的”,以及“紧张或放松的”^[14]。它引导出当前计算机学界主流的2个模型:2维的 Valence-Arousal Model^[15]和3维的 Pleasure-Arousal-Dominance Model^[16]。

2024年,量化模型方面一个重要的工作是 Collobetti 等^[17]关于效价和唤醒度之间关系的讨论,在 Ekman 等前人工作中,这2个维度一直被认为是正交的。基于此,Smith 等^[18]回顾了唤醒度在生理学中的起源,指出了它是一种非特异性激活状态的概念,因此可能并不是一个很好的量化维度,甚至被质疑为“比无用更加糟糕”^[19]。

此外,当前情感量化模型主要针对自然人之间交互,当交互主体的一方由自然人变为机器人时,交互情境性质的改变(如外形、语言因素多变性、非语言因素匮乏性等)会改变自然人的情感认知。因此,如何优化、扩展量化模型(如增加 Trust 维度^[20])来更好捕捉情感认知在新情境中量化性质和规律^[21],是赋予智能机器人情感自然交互能力的一个基本挑战。

情感计算的主要计算手段是数据挖掘方法,基本原理是挖掘情感表达(语言因素与非语言因素)的感知数据来量化识别情感。从机器人感知角度,可以根据数据类型分为:生理信号(心电模态^[22]、脑电模态^[23]、皮电模态^[24])和物理信号(文本模态^[25]、语音模态^[26]、视觉模态^[27]);从学习方法角度,可以分为机器学习方法和深度学习方法;从计算角度,可以分为单模态与多模态计算^[28]。

单模态计算较为关注学习算法(包括大规模数据库构建),2024年度的研究服从如下趋势^[29]:从人工定义(专家知识)特征结合传统机器学习方法,到以CNN(图像)与LSTM(文本)等为代表的特征自学习深度神经网络,再到Transformer架构(注意力机制为核心)的学习框架^[30]。近来更是将对比学习、迁移学习等无监督或半监督方法引入,以减轻对监督信号(情感量化标签)的依赖,提升模型鲁棒性^[31]。多模态计算关注的焦点在模态组合和融合机制,2024年度的研究趋势总结如下:多媒体社交网络的盛行让物理信号内部的模态组合逐渐固定(如:视觉-声音,文本-声音,图像-声音-文本)、情绪的生理反应让皮肤电成为生理信号模态组合中的重要选择^[32];融合机制从特征层/决策层融合过渡到模型层融合^[33],以便

利用诸如隐式马尔科夫模型等方法增加融合的可解释性与效果^[34]。

近期,ChatGPT等基于Transformer结构的大语言模型语义理解能力的阶跃式提升对单模态情感认知产生了冲击^[35],而类CLIP(contrastive language-image pre-training)结构也被设计出来提升多模态计算(文本-图像)的效能^[36]。如图2所示,利用GPT-4o测试了路怒症前摄调节方法中路况视觉诱因理解的准确性,发现GPT-4o可以较为准确地提供逐帧的场景理解,并且在时序上可以做到部分的路怒诱因推理工作。另外一个最新的例子是《Nature》2024年4月26日发表的论文《A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM》^[37]。该文提出了一种创新的多模态跨语言情感分析方法。该方法结合了Transformer模型和大型语言模型(LLM),通过神经机器翻译和模型集成,有效地提高了跨语言情感分析的准确性和鲁棒性。

Prompt

From the perspective of an experienced driver, analyse the scene content and relative changes of each frame and determine whether there are factors that may trigger the driver's road rage.



GPT-4o

From the provided sequence of images, I will analyze the situation to evaluate the driving conditions, potential triggers for road rage, and frame-to-frame changes that might exacerbate driver frustration or aggression.

Frame 1: The vehicle ahead appears to be maintaining a consistent speed, while a larger truck is present in the adjacent lane. No immediate conflict or erratic behavior is visible yet, but the presence of multiple vehicles close together could create potential tension.

Frame 2: A new vehicle (black SUV) is moving into the lane from the right side. This SUV appears to be cutting into the lane in front of the driver without adequate space or signaling. Such abrupt lane changes can provoke annoyance or alarm in the driver behind, as they may perceive it as reckless.

...

Frame 5 to 7: The black SUV continues forward without further aggression but does not show signs of creating additional safe space. The truck remains nearby in the adjacent lane. If the driver behind the SUV felt "boxed in," this scenario might escalate frustration.

Factors That May Trigger Road Rage:

1. Abrupt Lane Change: The black SUV's entry into the lane without proper signaling or spacing could be seen as inconsiderate or unsafe.

2. Crowding/Perceived Threat: The proximity of the SUV, the truck, and the forward vehicle may create a feeling of entrapment or danger.

3. Lack of Safe Distance: Tailgating or lack of sufficient stopping distance between vehicles increases tension.

4. Perceived Lack of Courtesy: The SUV cutting in could be interpreted as selfish driving behavior, especially if the driver recording the footage had to adjust speed abruptly.

图2 多模态大模型GPT-4o在路怒诱因理解和推测案例

在具体应用领域,Xing^[38]基于明斯基(Minsky)的心智情感理论,提出了一种具有异构LLM代理的设计框架,该框架使用语言学和金融学的先验指导知识实例化多个异构专业代理,并通过聚合代理进行总结,为金融情感分析领域探索了新的方向。Venerito等^[39]研究了纤维肌痛患者在疼痛表达的细微差异,提出了一种用于诊断纤维肌痛病症的LLM驱动情感分析方法,通过检测能够反映纤维肌痛相关的大脑感知与负面影响的语言和情感线索,模型能够有效地辅助临床医学诊断。Hellwig等^[40]探索了LLM在细粒度方面级情感分析的数据标签生成能力,在低资源场景下,合成数据增强在方面级类别情感分析上表现出了显著增益。

总结来说,情感认知能力是智能机器人情感交互能力提升的前置条件。当前研究在量化模型、学习方法、融合机制等技术层面的提升并不能掩盖一个主要问题:“情感计算”将“认知”简化为“计算”,忽视了情感是一个时序变量,而不是单一独立的量化状态。它的混合性、主观性、动态性等特点决定了其对诸如时间、地点、环境、生理状态、偏好、习惯,甚至是场合、文化等交互情境知识(context,又称上下文)

的依赖关系。这种依赖关系没有被很好地建模、量化,没有体现在计算过程中,因此,当前研究仍然受到“知其然”,不能“知其所以然”的困扰。

1.3 情感调节

大语言模型如ChatGPT在语义理解与对话生成方面的进步极大地增强了机器对自然人情感调节的能力(图3),能够有策略、有目的地影响情绪形成。历史上心理学研究已经发现,向人类施加与期望、自信和社交影响相关的情绪刺激,往往能够产生正面的影响^[41]。例如,在教育领域^[42]和健康促进^[43]中,通过鼓励性和正面的话语来提升学生的成功感。同时情绪调节并不局限于通过言语交流实现,非言语因素如音乐、环境调整 and 身体接触也起到关键作用。例如,在治疗和教育环境中,舒缓的背景音乐和温馨的环境布置可以显著提高个体的情绪状态,促进学习和康复^[44]。同时,肢体语言、面部表情,以及视觉和触觉的刺激,也是调节情绪的重要手段,这些非言语的交流方式能够绕过语言的局限,直接影响情绪的感知和表达^[45-46]。总的来说,情绪刺激可以通过语言因素通道(如文本、语音等)和非语言因素通道(如肢体动作、面部表情等)两大通道实现交互过程中的情感调节。

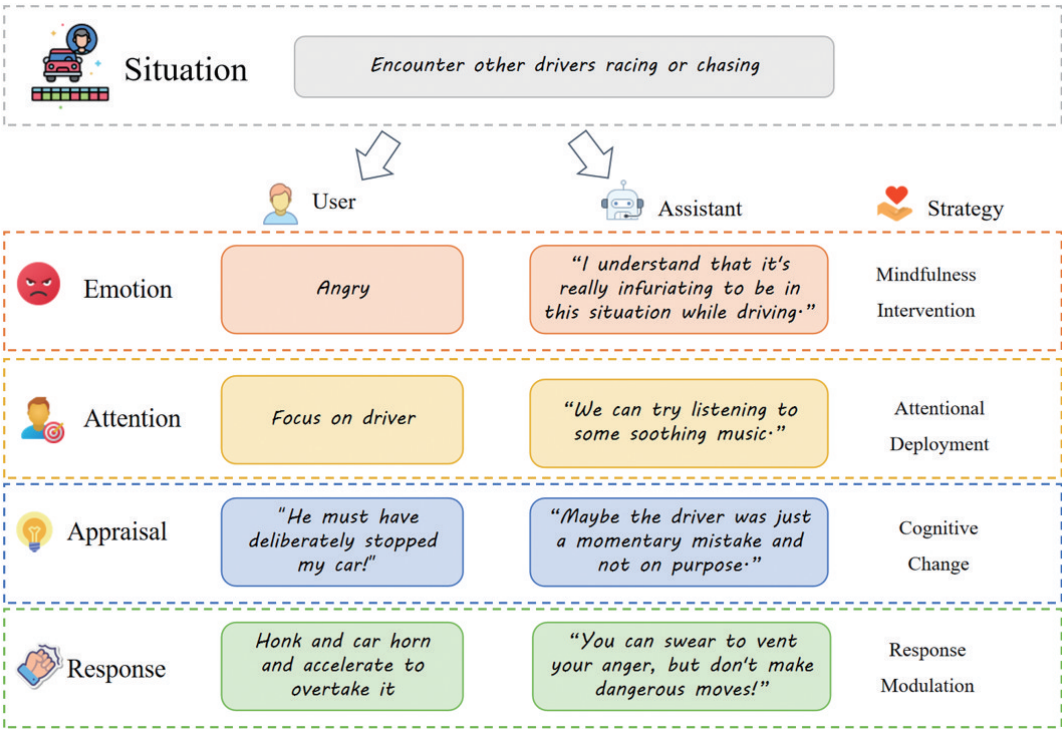


图3 针对路怒症的对话调节案例

近期,大语言模型在各种领域的任务中均展示了惊人的表现力,甚至被冠以“智能涌现”的现象级描述。智能机器人与大语言模型的结合,带来了其交互能力的阶跃式增长,特别是考虑到大语言模型在实现人工通用智能(AGI)方面具有巨大的潜力^[47],基于大语言模型来赋予智能机器人以自然流畅的情感交互能力正在逐渐成为可能^[48-49]。

在语言因素的调节通道中,以 ChatGPT 为代表的大语言模型,在现有的基准数据集上,对比以往的 SOTA(state of the art)模型表现极其优越,模型能够理解用户的经历和感受,并生成适当的共情性回复(图 2)^[50]。当前较为主流做法是通过扩展真实的心理互助 QA(question answering),支持多轮对话,从而提高通用大语言模型在情感调节方面的表现能力。例如,结合专业的心理咨询理论,研究人员开发了新一代共情陪伴大模型 Emohaa^[1],模型能够根据用户需求提供针对性的回复方案,善用疑问和引导的方式,和用户共同构建安全温暖的聊天环境,提供深层次情绪支持。中文心理健康支持对话大模型 MeChat^[2]则致力于使用户获得及时、有效的倾听和支持。Na^[51]提出了专门为认知行为治疗技术设计的大语言模型 CBT-LLM,它在基于认知行为疗法(CBT)的中文心理健康问答数据集上进行微调,擅长在心理健康支持任务中生成结构化、专业化和高度相关的回复。情感引导多模态对话模型(ELMD)^[52]将整体框架分解成情绪检索模块(ERM)、情绪反应预测模块(REP)和情绪增强的反应生成模块(EEPG),通过两阶段训练策略来理解多模态信息背后的微妙情绪,生成更加细粒度的富有情感的对话。在文本到音频(text-to-audio, TTA)领域,基于与大语言模型相似的 Transformer 架构, Bark 模型组合可以端到端地生成接近人类水平的语音,流畅、清晰且富有情感。

此外,基于非语言因素(如手势^[53])的智能机器人情感调节能力也获得了相当的重视。《Science》子刊《Robotics》在 2024 年 3 月发表了题为《Human-robot facial coexpression》的工作,创建了一个人形面部机器人 Emo,通过机器人面部的 26 个电机和硅胶皮肤,以及 2 个预测网络组合实现了逼真的面部表情表达^[54]。Llanes-Jurado 等^[55]创建了一个基于大语言模型的能够进行半引导对话的实时对话虚拟人,该虚

拟人融合了实时语音合成、实时声唇同步、实时面部表情生成、大语言模型驱动的对话系统,实验证明能够有效激发人类情绪。

然而,综合多个情绪调节通道建立一个多模态智能体来对用户进行多通道的情感调节依旧进展缓慢,主要难点在于大模型的情感调节能力是一个领域性问题,需要引入心理学知识,建立相应有效的调节策略(如质疑、复述、自我揭示和再保证等),并克服以困惑度为唯一标准带来的公式化回应。此外,语言因素和非语言因素的协同^[56],也是自然流畅的人机情感交互体验的一个重要难点。因此,建立一个基于一致情感驱动的以大语言模型为核心的多模态智能体^[57],增强交互过程中机器人对情感的认知和调节,是当前需要研究的重点。

1.4 情感交互能力评测

智能机器人的交互能力评测是近年来人工智能领域关注的焦点之一。在智能机器人领域,情感交互能力不仅涉及情绪的识别和响应,还包括在复杂社交环境中的适应性和应对策略。目前,尽管智能机器人在多种任务中表现出越来越高的技能,如自然语言处理和图像识别,但其在人机交互中的情感能力评估中仍面临诸多挑战。在情感交互能力的测量上,心理学界主要采用特质测量和能力测量 2 种方法^[58]。特质测量通常通过自我报告问卷完成,评估个体如何在特定情境中应对情绪。这种方法适用于个人的自我感知分析,但对于智能机器人来说并不适用。能力测量则更加重视个体情感理解和表现,通常采用更科学的评估方式,如 Mayer-Salovey-Caruso 情绪智力测试(MSCEIT)^[59],以及情绪理解和管理情境测试(STEU 和 STEM)^[60],这些测试通过设计多项选择题来评估能力的不同维度。这些测试不仅衡量了个体对情绪的感知和理解能力,还考查了其在复杂情境中管理和运用情绪的能力。

智能机器人交互能力评测方法可以分为 2 类:基于任务的评测和基于交互的评测。基于任务的评测通常通过设计特定的情感识别或情感响应任务,评估机器人完成这些任务的能力。这些任务可能包括情绪分类、情绪强度判断或情感驱动的决策等。基于交互的评测则更加注重评估机器人在实时人机交互中的表现,特别是其适应复杂社交环境的能力。

这类评测通常涉及长时间的交互,评估机器人如何理解和适应用户的情绪变化,以及如何在交互中持续调整其行为以维持或改善用户的情绪状态。

为有效评估大模型在复杂交互方面的能力,研究人员从不同角度提出多种评测基准。EmoBench是首个专门用于评估多模态大模型在5种主流情感任务中情感交互能力的综合基准,涵盖了通用情感任务(如多模态情绪识别和意图理解)和情感应用任务(如社交媒体应用中的特定挑战),共计287000余条多模态指令,如图4^[61]所示。传统情绪识别及原因认知方法中,忽视了对对象感知价值与个人情绪及其强度的相关性(这里的强度契合任福继EMSTM模型中EEE概念)。而EmoBench可以准确测试出对象模型在这个层面的能力。

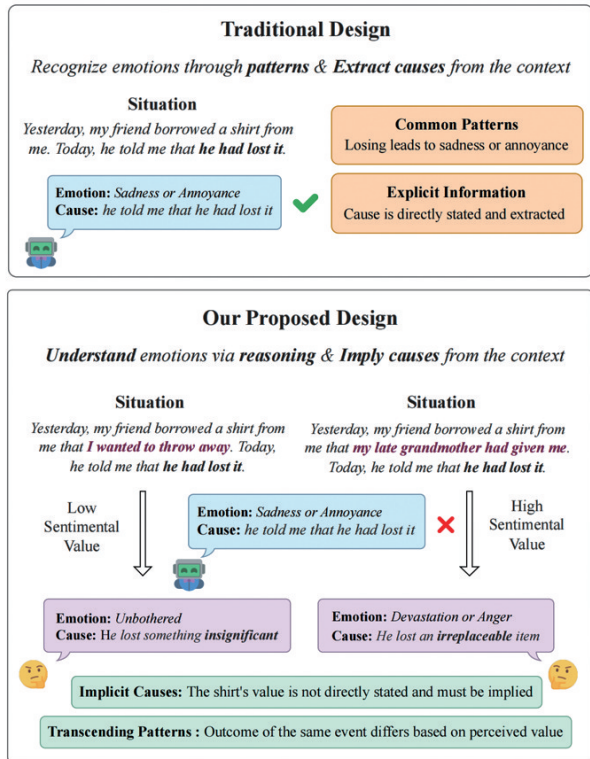


图4 首个专门用于评估大模型在主流情感任务中情感交互能力的EmoBench基准

FOFO^[62]是首个将LLM的指令遵循(instruction-following)分为内容遵循(content-following)和格式遵循(format-following),并对格式遵循能力进行评估的基准,它通过人工和AI协作的三步走策略,构建了涵盖50个领域、包含248种数据格式和494条格式化指令的复杂分层结构布局。Mizrahi等^[63]深入研究了单

提示评估的脆弱性,发现不同指令模板导致模型性能迥异,因而提出了一组基于多指令改写的多样化指标,有助于促进LLM评估工作的一致性和可比性。

总结来说,当前针对情感交互能力的评测方法大多基于人工设计量表,往往依赖交互主体的主观反应。然而,人机交互主体的变化(从自然人到机器人)使得这种评测模式不具可行性。而针对智能机器人情感交互能力的全面评估应涵盖情感理解和情感表达2个维度。与情感认知面临的困境类似,当前基于学习式的黑盒评分方法只能知其然,不能知其所以然。因此,需要建立一个综合多个维度的可解释评估策略,不仅要能量化个体对情绪的感知和反应,还应包括对情绪的深层次理解和在实际情境中的应用能力。

2 人机情感交互领域研究挑战

表1总结了人机情感交互环路中情感认知、情感调节与交互能力评测3方面在多模态大模型支撑下的研究回眸。

公共安全、国防军事、医疗照护等国计民生领域对 人机情感交互能力有着重大共性需求,但当前“感—析—互”机械式人机交互体系中存在的情感模态少、感知范围小、情绪知悉慢、自然交互难等缺陷,难以满足应用需求。

如表1所示,相较于传统深度学习模型,大语言模型在情感认知、情感调节等方面的极大进步给实现人机情感交互奠定了基础。然而,为营造“融感—协析—情互”自然人机交互环境,基于大模型的情感交互架构仍需应对以下挑战。

1) 感知挑战。人类情感表达具有多通道并发特征,交互是通过视觉、听觉、肢体动作等多种通道来进行情感交流的。然而,当感知设备将多通道交互信息采集转化成多模态感知数据时,面临数据完整性、一致性挑战。

2) 语义挑战。人类交互过程中天然存在表达二义性与语境依赖问题,这些问题在感知过程中(多通道达到多模态数据)由于采集缺陷和信道噪声而加剧,因此,如何扩展大语言模型的语义计算能力至多模态数据,以应对语义计算过程中存在含噪多模

表 1 多模态大模型在情感交互领域 2024 年度研究回眸

研究方向	研究内容
情感认知	Xing ^[38] 提出了一种用于金融情感分析领域的具有异构 LLM 代理的设计框架,使用语言学和金融学的先验指导知识实例化多个异构专业代理,并通过聚合代理进行总结
	Venerito 等 ^[39] 提出了一种用于诊断纤维肌痛病症的 LLM 驱动情感分析方法,通过检测能够反映纤维肌痛相关的中枢感知与负面影响的语言和情感线索,有效地辅助临床医学诊断
	Hellwig 等 ^[40] 探索了 LLM 在细粒度方面级情感分析的数据标签生成能力,在低资源场景下,合成数据增强在方面级类别情感分析上表现出了显著增益
情感调节	Na ^[51] 提出了专门为认知行为治疗技术设计的大型语言模型 CBT-LLM 模型,擅长在心理健康支持任务中生成结构化、专业化和高度相关的回复
	情感引导多模态对话模型 ^[52] 将整体框架分解成情绪检索模块、情绪反应预测模块和情绪增强的反应生成模块,通过两阶段训练策略来理解多模态信息背后的微妙情绪,生成更加细粒度的情感对话
	Llanes-Jurado 等 ^[53] 创建了一个基于大语言模型的半引导对话的实时对话虚拟人,融合了实时语音合成、实时声唇同步、实时面部表情生成、大语言模型驱动的对话系统,能够有效激发人类情绪
情感交互能力测评	EmoBench 是首个专门用于评估多模态大模型在 5 种主流情感任务中的情感交互能力的综合基准,涵盖了通用情感任务和情感应用任务 ^[61]
	FOFO 是首个将 LLM 的指令遵循分为内容遵循和格式遵循,并对格式遵循能力进行评估的基准,通过人工和 AI 协作的三步走策略,构建了涵盖多个领域和数据格式的复杂分层结构布局 ^[62]
	Mizrahi 等 ^[63] 深入研究了单提示评估的脆弱性,提出了一组基于多指令改写的多样化指标,有助于促进 LLM 评估工作的一致性和可比性

态数据到统一语义表示等面临挑战。

3) 交互挑战。人机交互过程中机器人、数字人等载体存在交互模态少、共情能力差等问题,致使人类作为交互主体容易产生交互“不自然、不流畅、不温情”的主观感受。在基于大模型的情感交互架构中,需要引入情感外在表达的知识,并通过强化学习等手段,在真实交互场景中嵌入人类反馈来优化模型。

3 人机情感交互领域未来研究方向

基于上述挑战,为赋予智能机器人人机情感自然交互能力,未来研究可以聚焦在以下具体方面。

3.1 情感认知可计算建模方法

人类情感状态具有混合性、主观性、动态性等特点,从人机交互的视角出发,它们会给机器端(智能机器人)带来情感量化、情境依赖、状态时变等计算问题。因此,如何建模人类情感状态在交互过程中产生的固着、反向、逃避和共情等认知规律(动力学),以准确量化计算情感状态,不但做到“知其然”,更进一步做到“知其所以然”,是实现人机情感自然交互所面临的基础理论挑战。

3.2 情感调控可进化交互环路

人机自然交互包含多样性、适应性、演化性等关键特性,需要高效快速地处理并响应多模态交互数据、情境变化数据等复杂数据。因此,如何形成一种有效的情感调节环路,并基于高效能的情境感知计算、情感调节策略指导的大语言模型推理、多样化的交互呈现,以及强化学习迭代进化策略,实现更自然、更流畅人机交互体验,是目前面临的关键技术挑战。

3.3 情感交互可解释评估体系

建立一个能够全面评估和解释智能机器人在人机交互中表现的评测体系,是提高交互质量和用户满意度的关键。有效的评测体系需要能够详细定义、精确量化,并清晰解释机器人在认知、生成、表达环路中的交互能力。这包括但不限于情景认知能力、情感表达能力,以及多场景适应能力,还需考虑交互过程中的伦理规范,确保机器人的行为符合伦理道德标准。因此,如何建立一个全面、高效且用户友好的可解释人机交互评测体系,是实现人机情感自然交互所面临的重要评估挑战。

解决上述“认知”→“调节”→“评估”问题,有望形成增强人机情感自然交互能力的研究闭环,其最

显著的理论研究意义在于探索了多学科融合研究的新范式:以数据为驱动,以计算为核心,融合心理学和认知科学知识,提升研究的可解释性与效能。

4 总结与展望

“人机情感交互”源于“情感计算”技术,20世纪90年代起源于美国,30年来一直是国际研究的前沿热点。任福继课题组长期致力该领域研究,提出“人机情感交互”方向,并入选中国科学技术协会信息科技领域6大“硬骨头”重大科学问题和工程技术难题。经诸多研究人员不懈努力,中国成为该领域最重要的崛起力量之一。《Nature》子刊《人文与社会科学通信》(《Humanities and Social Sciences Communications》)2021年第8卷刊文《情感计算学术研究与中国的崛起:从近25年文献计量学数据来看》(《Affective computing scholarship and the rise of China: A view from 25 years of bibliometric data》)指出了该领域中国的崛起。

国家《“十四五”机器人产业发展规划》中,也多次将“人机情感交互”列为主要攻关任务,各地市紧跟国家导向,陆续出台相应政策:《上海市人工智能产业发展“十四五”规划》《广州市科技创新“十四五”规划》(前沿技术与重点产业领域部分)等文件明确“情感计算”为当地“十四五”科技创新核心技术。

该方向的研究具有很强的研究、社会与现实意义,其实施有望填补该项基础研究理论空白,使中国进入自然交互和机器人平台技术的国际前列,取得具有重大国际影响力的自然交互研究成果,加速推进我国信息技术与产业发展,使其在健康医疗、公共服务、科技文教等领域产生显著的社会和经济效益。

参考文献(References)

- [1] Sabour S, Zhang W, Xiao X Y, et al. A chatbot for mental health support: Exploring the impact of Emohaa on reducing mental distress in China[J]. *Frontiers in Digital Health*, 2023, 5: 1133987.
- [2] Qiu H C, He H L, Zhang S, et al. SMILE: Single-turn to multi-turn inclusive language expansion via ChatGPT for mental health support[C]//*Proceedings of Findings of the Association for Computational Linguistics: EMNLP 2024*. Stroudsburg, PA: USAACL, 2024: 615-636.
- [3] Ryan O, Dablander F, Haslbeck J M B. Toward a generative model for emotion dynamics[EB/OL]. (2024-09-11)[2025-01-03]. <https://osf.io/preprints/psyarxiv/x52ns>.
- [4] Deng Y, Liao L Z, Zheng Z H, et al. Towards human-centered proactive conversational agents[C]//*Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, 2024: 807-818.
- [5] Ren F J, Zhou Y Y, Deng J W, et al. Tracking emotions using an evolutionary model of mental state transitions: Introducing a new paradigm[J]. *Intelligent Computing*, 2024, 3: 75.
- [6] Ren F J. Affective information processing and recognizing human emotion[J]. *Electronic Notes in Theoretical Computer Science*, 2009, 225: 39-50.
- [7] Sanchez L, Crocker D, Oo T M, et al. Robotic gestures, human moods: Investigating affective responses in public interaction[C]//*Proceedings of Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. New York: ACM, 2024: 935-939.
- [8] Bal B S, Pitti A, Cohen L, et al. Assessing the sense of control during affective physical human robot interaction[C]//*In 19th Annual ACM/IEEE International Conference on Human Robot Interaction (HRI)*. Boulder: ACM, 2024.
- [9] Xu W, Gao Z F. Applying HCAI in developing effective human-AI teaming: A perspective from human-AI joint cognitive systems[J]. *Interactions*, 2024, 31(1): 32-37.
- [10] Dolan R J. Emotion, cognition, and behavior[J]. *Science*, 2002, 298(5596): 1191-1194.
- [11] Picard R W. *Affective computing*[M]. Cambridge: MIT Press, 2000.
- [12] Belkaid M, Pessoa L. Modeling emotion to enable intelligent behavior in robots[J]. *Intellectica-La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo)*, 2023, 79: 109-128.
- [13] Ekman P. An argument for basic emotions[J]. *Cognition & Emotion*, 1992, 6(3/4): 169-200.
- [14] Wundt W M, Judd C H. *Outlines of psychology*[M]. Leipzig: W.Engelmann, 1902.
- [15] Russell J A. A circumplex model of affect[J]. *Journal of Personality and Social Psychology*, 1980, 39(6): 1161-1178.

- [16] Bakker I, van der Voordt T, Vink P, et al. Pleasure, arousal, dominance: Mehrabian and Russell revisited[J]. *Current Psychology*, 2014, 33(3): 405–421.
- [17] Colombetti G, Kuppens P. How should we understand valence, arousal, and their relation?[M]//*Emotion Theory: The Routledge Comprehensive Guide*. New York: Routledge, 2024: 599–620.
- [18] Smith K E, Woodard K, Pollak S D. Arousal may not be anything to get excited about[J]. *Emotion Review*, 2024: 17540739241303499.
- [19] Sander D. Is "arousal", as a scientific concept, worse than useless?[J]. *Emotion Review*, 2024: 17540739241303501.
- [20] Uddin M T, Yin L J, Canavan S. Spatio-temporal graph analytics on secondary affect data for improving trustworthy emotional AI[J]. *IEEE Transactions on Affective Computing*, 2024, 15(1): 30–49.
- [21] Stappen L, Baird A, Schumann L, et al. The multimodal sentiment analysis in car reviews (MuSe-CaR) dataset: Collection, insights and improvements[J]. *IEEE Transactions on Affective Computing*, 2023, 14(2): 1334–1350.
- [22] Casado C Á, Cañellas M L, López M B. Depression recognition using remote photoplethysmography from facial videos[J]. *IEEE Transactions on Affective Computing*, 2023, 14(4): 3305–3316.
- [23] Huang W C, Wang W L, Li Y Q, et al. FBSTCNet: A spatio-temporal convolutional network integrating power and connectivity features for EEG-based emotion decoding[J]. *IEEE Transactions on Affective Computing*, 2024, 15(4): 1906–1918.
- [24] Wickramasuriya D S, Faghih R T. A Bayesian filtering approach for tracking arousal from binary and continuous skin conductance features[J]. *IEEE Transactions on Bio-Medical Engineering*, 2020, 67(6): 1749–1760.
- [25] Kang X, Shi X F, Wu Y N, et al. Active learning with complementary sampling for instructing class-biased multi-label text emotion classification[J]. *IEEE Transactions on Affective Computing*, 2023, 14(1): 523–536.
- [26] Ren F J, Liu Z, Kang X. An efficient framework for constructing speech emotion corpus based on integrated active learning strategies[J]. *IEEE Transactions on Affective Computing*, 2022, 13(4): 1929–1940.
- [27] Xue F L, Wang Q C, Tan Z C, et al. Vision transformer with attentive pooling for robust facial expression recognition[J]. *IEEE Transactions on Affective Computing*, 2022, 14(4): 3244–3256.
- [28] Wang Y, Song W, Tao W, et al. A systematic review on affective computing: Emotion models, databases, and recent advances[J]. *Information Fusion*, 2022, 83: 19–52.
- [29] Li S, Deng W H. Deep facial expression recognition: A survey[J]. *IEEE Transactions on Affective Computing*, 2022, 13(3): 1195–1215.
- [30] 李路宝, 陈田, 任福继, 等. 基于图神经网络和注意力的双模态情感识别方法[J]. *计算机应用*, 2023, 43(3): 700–705.
- [31] Zhang W J, Song P, Zheng W M. Joint local-global discriminative subspace transfer learning for facial expression recognition[J]. *IEEE Transactions on Affective Computing*, 2023, 14(3): 2484–2495.
- [32] 任福继, 于曼丽, 胡敏, 等. 融合表情和BVP生理信号的双模态视频情感识别[J]. *中国图象图形学报*, 2018, 23(5): 688–697.
- [33] Gu Y, Zhang X, Yan H, et al. WiFE: WiFi and vision based unobtrusive emotion recognition via gesture and facial expression[J]. *IEEE Transactions on Affective Computing*, 2023, 14(4): 2567–2581.
- [34] Hossain M S, Muhammad G. Emotion recognition using deep learning approach from audio-visual emotional big data[J]. *Information Fusion*, 2019, 49: 69–78.
- [35] 邓佳文, 任福继. 2023年生成式AI大模型发展热点回眸[J]. *科技导报*, 2024, 42(1): 266–285.
- [36] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//*International Conference on Machine Learning*. Online: PMLR, 2021: 8748–876.
- [37] Miah M S U, Kabir M M, Sarwar T B, et al. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM[J]. *Scientific Reports*, 2024, 14(1): 9603.
- [38] Xing F. Designing heterogeneous LLM agents for financial sentiment analysis[EB/OL]. [2024–12–15]. <https://arxiv.org/abs/2401.05799v1>.
- [39] Venerito V, Iannone F. Large language model-driven sentiment analysis for facilitating fibromyalgia diagnosis[J]. *RMD Open*, 2024, 10(2): e004367.
- [40] Hellwig N C, Fehle J, Wolff C. Exploring large language models for the generation of synthetic training samples for aspect-based sentiment analysis in low resource settings[J]. *Expert Systems with Applications*, 2025, 261: 125514.

- [41] Li C, Wang J D, Zhu K J, et al. Large language models understand and can be enhanced by emotional stimuli[EB/OL]. (2021-12-26) [2025-01-03]. <https://arxiv.org/abs/210.00020>.
- [42] Mitiadoulou M, Savenye W C. Applying social cognitive constructs of motivation to enhance student success in online distance education[J]. *AACE Review (formerly AACE Journal)*, 2003, 11(1): 78-95.
- [43] Bandura A. Health promotion from the perspective of social cognitive theory[J]. *Psychology & Health*, 1998, 13(4): 623-649.
- [44] 鲍艳伟, 任福继. 人脑信息处理和类脑智能研究进展[J]. *科技导报*, 2023, 41(9): 6-16.
- [45] 黄忠, 任福继, 胡敏, 等. 基于Transformer架构和B样条平滑约束的机器人面部情感迁移网络[J]. *机器人*, 2023, 45(4): 395-408.
- [46] 黄忠, 任福继, 胡敏, 等. 基于双LSTM融合的人机机器人实时表情再现方法[J]. *机器人*, 2019, 41(2): 137-146.
- [47] Bubeck S, Chandrasekaran V, Eldan R, et al. Sparks of artificial general intelligence: Early experiments with GPT-4 [EB/OL]. (2023-03-22) [2025-01-03]. <https://arxiv.org/abs/230.1271>.
- [48] 王国庆, 裴云强, 杨阳, 等. 多模可信交互: 从多模态信息融合到人-机器人-数字人三位一体式交互模型[J]. *中国科学: 信息科学*, 2024, 54(4): 872-892.
- [49] 邵雷, 石峰. 生成式人工智能对社交机器人的影响与治理对策[J]. *情报杂志*, 2024, 43(7): 154-163.
- [50] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[EB/OL]. [2025-01-03]. <https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800a0fe0b668a1cc19f2ec95b5003d0a50-35>.
- [51] Na H. CBT-LLM: A chinese large language model for cognitive behavioral therapy-based mental health question answering[C]//In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation. Torino, Italia: ELRA and ICCL, 2024: 2930-2940.
- [52] Liu C X, Xie Z Y, Zhao S R, et al. Speak from heart: An emotion-guided LLM-based multimodal method for emotional dialogue generation[C]//Proceedings of the 2024 International Conference on Multimedia Retrieval. New York: ACM, 2024: 533-542.
- [53] Hei X X, Yu C, Zhang H, et al. A bilingual social robot with sign language and natural language[C]//Proceedings of Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. New York: ACM, 2024: 526-529.
- [54] Hu Y H, Chen B Y, Lin J, et al. Human-robot facial coexpression[J]. *Science Robotics*, 2024, 9(88): eadi4724.
- [55] Llanes-Jurado J, Gómez-Zaragoza L, Minissi M E, et al. Developing conversational Virtual Humans for social emotion elicitation based on large language models[J]. *Expert Systems with Applications*, 2024, 246: 123261.
- [56] Chang C J, Sohn S S, Zhang S, et al. The importance of multimodal emotion conditioning and affect consistency for embodied conversational agents[C]//Proceedings of the 28th International Conference on Intelligent User Interfaces. New York: ACM, 2023: 790-801.
- [57] Kim C Y, Lee C P, Mutlu B. Understanding large-language model (LLM)-powered human-robot interaction[C]//Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction. New York: ACM, 2024: 371-380.
- [58] Ashkanasy N M, Daus C S. Rumors of the death of emotional intelligence in organizational behavior are vastly exaggerated[J]. *Journal of Organizational Behavior*, 2005, 26(4): 441-452.
- [59] Salovey P, Mayer J D. Emotional intelligence[J]. *Imagination, Cognition and Personality*, 1990, 9(3): 185-211.
- [60] MacCann C, Roberts R D. New paradigms for assessing emotional intelligence: Theory and data[J]. *Emotion*, 2008, 8(4): 540-551.
- [61] Yang Q, Ye M, Du B. EmoLLM: Multimodal emotional understanding meets large language models[EB/OL]. [2025-01-03]. <https://arxiv.org/abs/2406.16442v2>.
- [62] Xia C Y, Xing C, Du J S, et al. FOFO: A benchmark to evaluate LLMs' format-following capability[C]//Annual Meeting of the Association for Computational Linguistics, 2024.
- [63] Mizrahi M, Kaplan G, Malkin D, et al. State of what art? A call for multi-prompt LLM evaluation[J]. *Transactions of the Association for Computational Linguistics*, 2024, 12: 933-949.

Review of research hotspots in affective human computer interaction in 2024

GU Yu, REN Fuji*

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

Abstract Research on affective human-machine interaction aims to create an intelligent system capable of perceiving, recognizing, and understanding human emotions, and responding with wisdom, sensitivity, and naturalness. It serves as a fundamental technology and a crucial prerequisite for achieving naturalization, anthropomorphism, and personalization in robotics. In 2024 advancements in large language models brought profound changes to this field. This article reviews the progress in this field in 2024 in terms of interaction theory, emotional cognition, emotional regulation, and evaluation of interaction effects. Additionally, it discusses the theoretical and technical challenges that exist in areas such as computational models of emotional cognition, evolutionary interaction loops for emotional regulation, and explainable evaluation systems for emotional interaction.

Keywords human-machine emotional interaction; multimodal large models; emotional cognition; emotional regulation; interaction effect evaluation ●



(责任编辑 王微)