基于注意力机制的多模态情感计算关键技术

王修婷
河南工业职业技术学院　河南 南阳　473000

摘　要　本文聚焦基于注意力机制的多模态情感计算关键技术，分析了传统单模态情感计算的局限性。详细介绍了注意力机制在多模态情感计算中的应用原理，包括该机制的核心思想与计算过程。深入探讨了多模态特征提取与融合技术，以及基于注意力机制的情感模型构建方法。最后，对多模态情感计算的应用场景进行了展望，并指出了未来研究方向。

关键词　注意力机制；多模态情感计算；特征提取；特征融合；情感模型

Key Technologies of Multimodal Affective Computing Based on Attention Mechanism

Wang Xiu-ting
Henan Polytechnic Institute, Nanyang 473000, Henan Province, China

Abstract　This paper focuses on key technologies of multimodal affective computing based on attention mechanisms and analyzes the limitations of traditional unimodal affective computing. It provides a detailed introduction to the application principles of attention mechanisms in multimodal affective computing, including the core concepts and computational processes of this mechanism. The paper delves into multimodal feature extraction and fusion techniques, as well as methods for constructing affective models based on attention mechanisms. Finally, it discusses the application scenarios of multimodal affective computing and outlines future research directions.

Key words　Attention Mechanism; Multimodal Affective Computing; Feature Extraction; Feature Fusion; Affective Model

在人工智能快速发展的当下，情感计算作为人机交互的核心环节备受瞩目，其目标在于使计算机具备识别、理解与表达情感的能力，达成更自然、智能的人机交互。传统单模态情感计算仅依赖语音、文本或面部表情等单一信息源来识别情感，但鉴于情感表达的多样性与复杂性，单一模态难以全面且精准地捕捉情感信息。而多模态情感计算通过融合文本、语音、生理信号、面部表情等多种模态信息，能更全面、准确地理解与分析人类情感状态，比如社交媒体中用户发布的视频涵盖文本描述、语音表达和面部表情，多模态情感计算可综合这些信息精准判断用户情感倾向[1]。此外，注意力机制作为模拟人类注意力分配方式机制，能自动识别并聚焦重要信息，在多模态情感计算中至关重要，引入该机制可更有效地融合不同模态信息，提升情感分析的鲁棒性与准确性。

1　注意力机制在多模态情感计算中的应用原理

1.1 注意力机制的核心思想

注意力机制的核心思想是在处理信息时，有选择性地聚焦关键部分，忽略次要信息，提高模型对重要特征的捕捉能力。在多模态情感计算中，不同模态的信息对情感分析的重要性各不相同。例如，在一段包含语音和文本的视频中，语音中的语调、语速与文本中的关键词都可能对情感表达起到关键作用，而一些无关的背景噪音或冗余文字则可能干扰情感分析。注意力机制能够自动为不同模态信息分配不同权重，使模型更加关注情感相关的信息。

1.2 注意力机制的计算过程

注意力机制的计算过程可抽象为3个阶段，涉及查询（Query）、键（Key）、值（Value）三个关键要素。给定目标中的某个元素Query，通过计算Query和各个Key的相关性或相似性，得到每个Key对应Value的权重系数，然后对Value进行加权求和，即得到最终的Attention数值。

以多模态情感计算为例，假设有一个视频数据，包含文本模态和语音模态。在计算过程中，可以将文本模态中的某个特征向量作为Query，语音模态中的各个特征向量作为Value和Key。通过计算Query与每个Key的相似度，得到相应权重系数，再对Value进行加权求和，即可得到融合了文本和语音信息

[基金项目] 南阳市软科学研究计划项目，项目名称：数据要素驱动下南阳市数智中医药创新发展的路径与策略研究，项目编号：2025RKX0390。

的注意力输出。相似度计算方法有多种，如点积、缩放点积、度、加性注意力等[2]。

1.3 注意力机制在多模态情感计算中的优势

注意力机制在多模态情感计算中具有诸多优势。它能够自动识别和关注重要信息，有效提高模型对不同模态信息的融合。在处理复杂多模态数据时，传统方法可能难以准确判断不同模态信息的重要性，而注意力机制可以通过学习自动分配权重，使模型更加聚焦于关键信息。通过可视化注意力权重，还可以直观地看到模型在处理多模态信息时重点关注了哪些部分，从而更好地理解模型的决策过程，尤其适用于处理大规模的多模态数据。

2 多模态特征提取与融合技术

2.1 多模态特征提取方法

在特征提取方面，文本特征提取现阶段主流方法为基于深度学习的方法，如词到向量（Word2vec）能将词语映射为低维向量，捕捉语义关系；基于Transformer的双向编码器表征（BERT）作为基于变换器（Transformer）的预训练语言模型，通过双向编码器结构更准确表示词语含义，广泛应用于文本情感特征提取；从语言模型中嵌入（ELMO）可根据不同上下文动态调整词语表示，以捕捉多义性。语音特征提取常用梅尔频谱和梅尔频率倒谱系数（MFCC）等，梅尔频谱模拟人类听觉系统感知特性，捕捉频率信息，MFCC在其基础上经离散余弦变换压缩，得到更具代表性特征系数，在语音情感识别中应用广泛。图像特征提取常见方法有梯度特征（描述像素点灰度变化，用于边缘检测和目标识别）、颜色特征（反映整体颜色分布，对表达情感相关视觉信息有作用）。

2.2 多模态特征融合策略

1.特征级融合。特征级融合也称为早期融合。即对各模态的特征完成提取后，将各模态的特征向量进行简单拼接、相乘、相加以及复合等操作，形成组合特征，最后将早期融合后得到的组合特征输入到分类器中进行情感分类。例如，在处理包含语音和文本的多模态数据时，可以将语音特征向量和文本特征向量进行拼接，得到更长的组合特征向量，然后输入到支持向量机（SVM）或神经网络等分类器进行分类。特征级融合的优点是能够保留各模态的原始特征信息，且计算相对简单。但也存在一些问题，如不同模态的特征维度可能差异较大，直接拼接可能导致信息不平衡，影响融合效果。

2.决策级融合。决策级融合也称为后期融合，是在决策层进行特征融合的策略。即每个模态的特征先单独经过分类器的分析，将每个模态的情感分类结果进行融合，从而得到最终分类结果。例如，可分别使用文本分类器和语音分类器对文本和语音数据进行情感分类，得到两个分类结果，然后通过投票、加权平均等方法将这两个结果进行融合，得到最终的情感分类。决策级融合的优点是能够充分利用各模态分类器的优势，对单个模态的分类错误具有一定的容错能力。但决策级融合忽略了各模态特征之间的相关性，可能无法充分利用多模态信息的互补性。

3.模型依赖融合。模型依赖融合可以将不同模态特征分别输入到不同模型结构再进行特征提取，例如，可使用卷积神经网络（CNN）处理图像特征，使用长短期记忆网络（LSTM）处理语音特征，然后将二者提取的特征进行进一步融合，如通过全连接层将两个特征向量进行合并，再进行分类。模型依赖融合能够更好地适应不同模态的特征特点，提高融合效果，但模型的复杂度和训练难度也相对较高[3]。

3 基于注意力机制的情感模型构建

3.1 情感模型描述

情感模型描述是进行情绪识别的基础和前提，通过建立情感状态的数学模型，可更直观地理解和描述情感的内涵。根据其描述情感的方式，情感模型分为离散情感模型、维度情感模型与其他情感模型。在实际的情感建模应用中，离散情感模型和维度情感模型的应用最为广泛。

离散情感模型将情感划分为若干个离散的类别，如悲伤、高兴、愤怒、惊讶等。这种模型简单直观，易于理解。例如，在社交媒体情感分析中，将用户的评论分为积极、消极和中性三类，通过多模态情感计算模型判断每条评论所属的类别。维度情感模型则将情感看作是在多个维度上的连续变化，常见的维度包括效价（积极—消极）和唤醒度（兴奋—平静）。维度情感模型能够更细腻地描述情感的变化，适用于一些需要精确衡量情感强度的场景。

3.2 基于注意力机制的情感模型架构

基于注意力机制的情感模型架构通常包括注意力机制层、多模态特征输入层、特征融合层和情感分类层。在多模态特征输入层，可分别提取文本、语音、图像等模态的特征向量。注意力机制层则根据不同模态特征重要性，计算每个模态特征的注意力权重。特征融合层将加权后的各模态特征进行融合，得到一个综合的特征表示。最后，情感分类层根据融合后的特征，使用分类器（如Softmax分类器）进行情感分类。

例如，在一个基于注意力机制的视频情感分析模型中，可

首先使用BERT提取文本特征，使用MFCC提取语音特征，使用CNN提取图像特征。然后，在注意力机制层，可分别计算文本、语音和图像特征的注意力权重。通过计算查询向量（可以是某个模态的平均特征向量）与各模态键向量的相似度，得到相应的注意力权重。接着，在特征融合层，将加权后的文本、语音和图像特征进行拼接，得到一个综合的特征向量。最后，在情感分类层，将综合特征向量输入到Softmax分类器中，判断视频所属的情感类别[4]。

3.3 模型训练与优化

模型训练是使基于注意力机制的情感模型能够准确进行情感分类的关键步骤。在训练过程中，需要准备大量的多模态情感标注数据，将数据分为训练集、验证集、测试集。使用训练集对模型训练，通过反向传播算法调整模型参数，使模型在训练集上的损失函数最小化。验证集用于在训练过程中监控模型性能，防止模型过拟合。当模型在验证集上的性能不再提升时，停止训练。最后，使用测试集模型进行评估，检验模型在实际应用中性能。

4 多模态情感计算的应用场景展望

4.1 智能客服领域

在智能客服系统中，多模态情感计算可以发挥重要作用。通过集成多模态情感计算能力，智能客服能够更深入理解客户的需求，实时感知并适应客户情绪变化，提供更加个性化和贴心的服务。例如，当客户通过语音和文字与智能客服交流时，系统可以同时分析客户的语音语调、文字内容与可能的面部表情（如果客户使用视频通话），判断客户的情绪状态。如果系统识别到客户困扰或不安，会自动调整服务方式，以更为耐心细致的态度回应客户，必要时无缝衔接人工客服，确保每位客户都能获得恰如其分的关注与帮助，从而提升整个服务过程中的舒适度和满意度[5]。

4.2 教育领域

在教育领域，多模态情感计算可以为教学提供有力支持。通过分析学生在课堂上的语音、面部表情和肢体语言等多模态信息，教师可以实时了解学生的情绪变化和学习状态。例如，如果系统检测到学生在课堂上表现出困惑或厌烦的情绪，教师可以及时调整教学节奏和方式。多模态情感计算还可以用于学生心理健康监测，通过长期收集和分析学生的多模态情感数据，及时发现学生可能存在的问题，并采取相应措施[6]。

5 结束语

本文深入研究了基于注意力机制的多模态情感计算关键技术。首先，介绍研究背景与意义，分析传统单模态情感计算的局限及多模态情感计算的发展趋势；其次，阐述了注意力机制在其中的应用原理，探讨了多模态特征提取与融合技术；再次，构建了基于注意力机制的情感模型，并介绍模型训练与优化方法；最后，对应用场景进行展望。未来，该领域在模型优化上可深入研究注意力机制变体及其与其他深度学习技术结合以构建更强模型；应用场景拓展上，可探索其在医疗、交通等领域的应用，如医疗中辅助诊断监测、交通中预防危险驾驶行为；跨学科融合研究也是重要方向，加强多学科合作交流能为多模态情感计算发展提供新思路和技术支持。

参考文献

[1]李虹羽,苏锦,何俊.联合自注意力和交叉注意力的多模态方面级情感分类[J].信息技术与信息化,2025(6):133-138.

[2]赵雪峰,柏长泽,狄恒西,等.基于跨模态语义信息交互学习的多模态情感分析[J].南京师范大学学报(工程技术版),2025,25(2):43-53.

[3]冯程.基于混合注意力机制的多模态情感分析算法研究与实现[D].济南:山东交通学院,2025.

[4]欧阳宁,张恩泽,林乐平.基于协同注意力机制的多模态情感分析模型[J/OL].桂林电子科技大学学报,2025:1-7.[2025-09-29].

[5]赵杨,张雪,王玮航,等.基于多模态情感分析的图书馆智能服务用户情感体验度量[J].情报科学,2023,42(9):89-96.

[6]张换香,李梦云.基于多模态信息融合的中文隐式情感分析[J].计算机工程与应用,2025,61(2):102-110.

（本文编辑　曹　阳）

