混合量子与图神经网络的多模态情感分析方法

李兴广,蔡禹健,崔 炜,李劲松,张莹瑀
(长春理工大学电子信息工程学院,吉林长春 130022)

摘 要: 多模态情感分析(Multimodal Sentiment Analysis,MSA)是人工智能情感计算领域最具应用潜力的技术之一. 视觉、语音与文本中包含了人类多数真实情感特征,融合三种模态获得更精细的情感多维度主观表达以保障情感分析结果准确依然面临诸多挑战. 三种模态各自提取的情感特征子集中元素数量和时序不一致时,各模态选取代表性情感特征的良好策略是避免特殊情感特征被忽略或过度提取,以及保证后续融合分析时情感计算结果可信的关键. 三种模态代表性情感特征直接融合分析时模态间情感信息的传递机制与互补机制未被充分利用,导致情感分析结果仅关联于某一模态代表语义特征,造成模型过拟合,分类输出结果错误. 此外,人类的情感表达具有模态异构性与不一致性,常导致情感特征分布不均及模态极性歧义问题. 算法模型不仅要捕获不同模态间的互补信息与细粒度关联,还要抑制冗余特征对情感判别的干扰,避免数据融合过程存在"语义鸿沟",使结果稳定性受限. 本文基于多尺度时序表征与量子比特多态表征思想,提出了混合量子与图神经网络的多模态情感分析方法. 首先,构建代表性序列的拓扑表征图网络捕捉各特征节点之间的图结构动态关系,并在图网络中添加多头图注意力机制自适应调整节点与边权重,保证特殊情感特征可信选取. 然后,设计情感特征量子计算网络,将多模态特征按量子编码映射至高维希尔伯特空间,基于量子叠加与纠缠机制进一步促进模态间特征的深层次耦合与相互依赖建模,通过量子测量过程将叠加态坍缩至特定的本征态,实现量子态与情感特征的对应映射,获得更具判别性的多模态融合情感表征. 最终,将单模态与多模态预测作为多个子任务形成多任务协同优化机制,生成伪标签与共享表征提高每个任务的性能,结合多任务损失函数缓解模态表征不一致性,增强了模型的泛化性. 在CMU-MOSI、CH-SIMS和CMU-MOSEI基准数据集上的系列实验结果表明,相较常用基线模型,方法情感二分类准确率提高了1.5%~8.7%、五分类准确率提高了3.3%~10.7%、七分类准确率提高了1.5%~14.5%、F1分数最高提升8.5、皮尔逊相关系数最高提升0.146和平均绝对误差最高下降0.304.

关键词: 多模态情感分析;图神经网络;量子机器学习;跨模态信息融合;多任务优化
基金项目: 吉林省科技厅项目(No.20250102225JC)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2025)11-3983-13
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250554

A Hybrid Quantum-Graph Neural Network for Multimodal Sentiment Analysis

LI Xing-guang, CAI Yu-jian, CUI Wei, LI Jin-song, ZHANG Ying-yu
(*School of Electronic Information and Engineering*, *Changchun University of Science and Technology*, *Changchun*, *Jilin* 130022, *China*)

Abstract: Multimodal sentiment analysis (MSA) is one of the most promising technologies in the field of affective computing. Visual, acoustic, and textual modalities encode most human emotional features. Integrating them yields a finer, multidimensional representation of subjective affect. However, achieving accurate and robust sentiment analysis still faces significant challenges. When the sentiment feature subsets extracted from each modality differ in element quantity or temporal alignment, an effective strategy for selecting representative emotional features is essential to prevent key features from being overlooked or over-extracted, thereby ensuring the reliability of subsequent fusion analysis. Direct fusion of representative features across modalities often fails to fully exploit information transmission and complementarity, which can cause excessive reliance on a single modality's semantic representation and lead to overfitting or misclassification. Furthermore, human emotional expression exhibits modality heterogeneity and inconsistency, often resulting in uneven feature distributions and polarity ambiguity. Algorithmic models must not only capture cross-modal complementary information and fine-

收稿日期:2025-06-25;录用日期:2025-11-04;责任编辑:覃怀银

grained correlations but also suppress redundant features that interfere with emotional discrimination. The presence of a "semantic gap" in data fusion further limits result stability. To address these issues, this paper proposes a hybrid quantum-graph neural network, inspired by multi-scale temporal representation and qubit-based polymorphic encoding. First, a topological graph network of representative sequences is constructed to capture dynamic relationships among feature nodes, and a multi-head graph attention mechanism is introduced to adaptively adjust node and edge weights, ensuring reliable selection of critical sentiment features. Then, a quantum sentiment feature computation network is designed, mapping multimodal features into a high-dimensional Hilbert space via quantum encoding. Leveraging quantum superposition and entanglement, the model enhances deep intermodal coupling and dependency modeling. Through quantum measurement, superposed states collapse into specific eigenstates, establishing a correspondence between quantum states and sentiment features, and yielding more discriminative multimodal fusion representations. Finally, single-modal and multimodal predictions are formulated as multiple subtasks under a multitask collaborative optimization framework. Pseudo-label generation and shared representations improve task-specific performance, while a joint multitask loss mitigates inconsistencies among modality representations, enhancing the model's generalization ability. Experimental results on the CMU-MOSI, CH-SIMS, and CMU-MOSEI benchmark datasets demonstrate that, compared with conventional baselines, the proposed method improves binary classification accuracy by 1.5%~8.7%, five-class accuracy by 3.3%~10.7%, and seven-class accuracy by 1.5%~14.5%. The F1 score increases by up to 8.5 points, the pearson correlation coefficient improves by up to 0.146, and the mean absolute error decreases by up to 0.304.

Key words: multimodal sentiment analysis; graph neural network; quantum machine learning; cross-modal information fusion; multitask optimization

Foundation Item(s): Jilin Provincial Department of Science and Technology Project (No.20250102225JC)

1 引言

人类智慧,由"智商"和"情商"有机结合而成,除了具有认知客观事物和逻辑计算的能力[1],还包含受情感驱动的决策和行动. 情感分析技术通过对带有情感色彩的、以多种感官为载体的主观性表达进行特征提取与融合,赋予了机器感知和理解人类情感的能力. 准确感知情感能使机器更全面地服务人类,目前已应用于智能座舱[2]、互联网教育[3]、医疗健康[4]、人机交互[5]等多个领域,具有广泛发展前景和研究价值.

早期情感分析技术仅聚焦于单一模态数据[6],忽略了人类情感表达的多维度性,并且在面部遮挡、声学噪声强、语义反讽、光照变化等条件下难以从单一模态中准确提取情感特征. 随着传感器技术和计算机处理能力提高,基于人类情感信息模态间关联性,研究人员提出了多模态情感分析[7],通过提取并融合多源异构数据的情感信息,更精准地感知情感. 多模态情感分析主要包含模态内部情感特征提取和模态间信息交互两个主要过程:模态内情感特征提取注重单一传感器获取的空间或时序信息[8],得到更高维度的单模态表征,是提高模态间信息融合效率的基础;模态间信息交互利用异构数据情感信息的传递特性和互补特性,为情感分析任务提供更加准确鲁棒的结果[9].

模态内情感特征提取常用方法为循环神经网络如LSTM和GRU等[10],加强了序列特征提取效果,在模态内特征尺度不统一条件下,仍有许多特殊的情感特征面临被忽略或过度提取的风险. Zedeh等人[11]提出的TFN模型采用LSTM提取模态内情感特征,并构建张量融合层模拟模态间交互,提高了模态间信息融合能力,若添加模态内不同层次情感信息的表征,可进一步稳定融合后的计算结果. 对于模态间信息交互的过程,许多模型基于注意力机制来融合多模态特征[12],Tsai等人[13]提出的MulT模型采用多头注意力机制来关注不同模态间的情感信息,提高了模型综合性能指标,若进一步融合模态内高维度特征可使模型更充分利用单模态特征,并权衡模态内与模态间情感表征的交互.

综上,目前多模态情感分析模型仍然面临诸多挑战:许多模态内序列信息的代表性情感特征缺少有效提取,直接融合会导致这些独特的情感信息被忽略,情感分析准确率受限;模态间情感信息的传递特性与互补特性未被充分利用,模型预测结果更趋向于某一模态所代表的语义特征;日常生活中人类的表达在各模态的情感倾向存在不一致性,数据融合过程存在"语义鸿沟",模型需要平衡模态内与模态间的情感关系,准确分析出融合模态所表达的真实情感.

针对以上问题,本文提出了一种混合量子与图神经网络的多模态情感分析方法. 基于多尺度时序特征提取思想,构建序列拓扑表征图网络提取单模态时序情感表征,融合多头注意力机制兼顾模型节点属性与边属性,使模态内情感特征得到有效提取. 基于量子比特多态表征思想,设计量子计算网络实现多模态特征融合,引入量子叠加、纠缠、坍缩测量等计算过程实现

模态特征交互与深层次耦合,降低情感倾向差异导致分析任务性能受限,充分利用模态间情感传递与互补特性. 设计多任务学习与协同优化策略,通过伪标签生成与子任务共享表征提高模型泛化性能,平衡各模态情感倾向,对各模态情感表征不一致的困难样本也能实现稳定分析.

2　相关工作

2.1　多模态情感分析

早期单模态情感分析主要集中在文本模态[14],仅依靠文本数据无法满足日常生活中分析更加复杂情感的需求,因此多模态情感分析被提出,除了文本模态外还包含丰富的语音和视觉信息[15],通过融合多模态信息使模型预测出隐含的真实情感极性成为可能,一定程度上克服了单模态表征的局限. 然而,人类的情感表征是一个多模态的时序过程,体现在模态内情感特征时序不一致性和模态间情感极性不一致性,因此模型需要更深层次关联. Yu等人[16]剪辑中文影视片段构建了CH-SIMS多模态情感分析数据集,许多样本同样呈现出异构数据的不一致性,如图1所示.

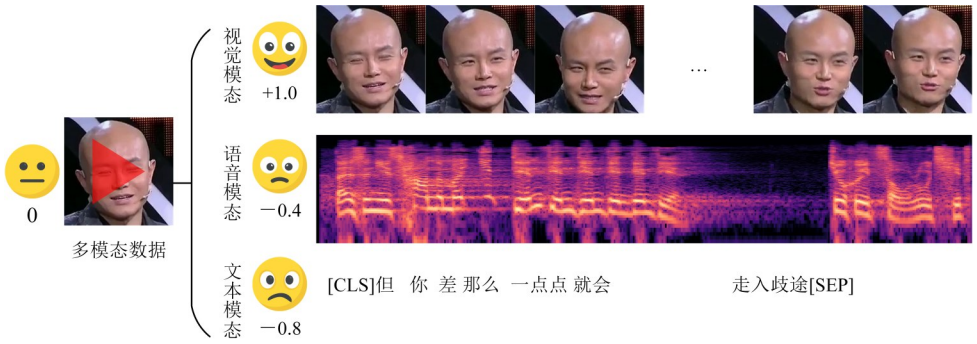

图1　CH-SIMS数据集中样本单模态标签与多模态标签示例

Hazarika等人[17]设计了两种编码器结构,将特征投影在模态情感不变子空间和模态情感特征突变子空间,有助于模型获得更加全面的单模态情感表征. 图网络可以从数据的拓扑结构与节点特征中提取情感信息,并具有非欧几里得特性,在异构数据处理中具有显著优势. Zeng等人[18]构建图卷积神经网络与自监督学习方法来整合异构知识并平衡模态间的独立信息与互补信息. 蒋昆等人[19]提出超图神经网络,利用多节点连接特性模拟情感交互并挖掘数据间深层次情感表征. 在模态交互与信息融合阶段,Sun等人[20]在特征提取过程中评估并调整模态中不变、特定、互补的信息比例优化了模态间协作,增强了模态融合性能. Zhang等人[21]构建了一种对抗网络减少跨域情感分析中表达风格的转移并获取模态间情感信息的联合表示,促进情感信息在不同域之间的传递. 然而模型仍然面临模态表征不平衡和多尺度信息融合带来的挑战,在拓扑关系提取,标签信息增强和训练损失优化等方面有改进空间.

2.2　量子机器学习

量子机器学习是计算科学中一个快速发展的领域,因其并行计算和纠缠叠加表示的特点在解决复杂计算问题上已展现出巨大潜力[22],通过引入量子比特实现同时处于0和1的叠加态表征,使系统包含多个状态. 将经典信息编码为量子信息能使其具有更复杂和通用的特性,可以体现在如图2所示的布洛赫球中.

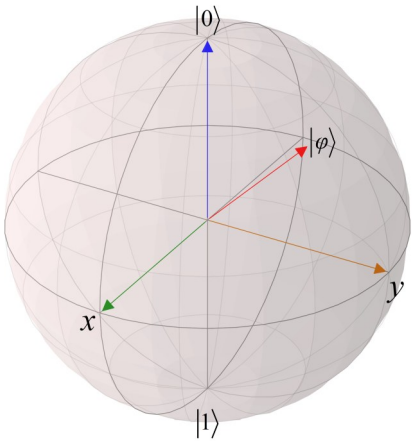

图2　单量子比特的布洛赫球表征

复合量子系统由多个独立希尔伯特空间内的子系统通过张量积构成,当叠加态的量子比特无法完全表述单比特时,出现量子纠缠[23],用于捕获更丰富和复杂的数据结构,最后通过量子测量将叠加量子系统坍缩至某个确定类别的本征态,用于算法的分类或回归. Li等人[24]将原始数据分批输入量子电路,设计变分量子电路来处理文本语音模态的高维数据以减少情感信息丢失和噪声干扰. Phukan等人[25]利用量子叠加、纠缠、干涉等特性捕获特征的跨模态相关性交互,构建混合量子框架实现了更加精准的讽刺与情绪分类. 融合经典与量子网络[26]能互补传统机器学习和量子特征的优势,降低量子噪声影响和传统算法关联提取效能

的限制.

相较于传统神经网络依赖线性或有限层次的特征映射,量子计算网络能够通过叠加态实现多模态特征在高维希尔伯特空间中的并行表示,使不同模态特征在同一量子态中形成耦合表征. 量子纠缠机制进一步强化了模态间的关联映射,使网络能够捕获跨模态的深层依赖关系与语义一致性. 同时,量子干涉特性在特征测量阶段能够抑制冗余信息、放大显著差异,以实现信息融合的非线性增强. 因此,本文构建量子机器学习模型并结合监督策略与模型优化方法,实现多源异构数据情感表征的稳定提取与关联,提高多模态情感分析准确性.

3 模型设计

本文构建的模型整体框架如图3所示,主要由序列拓扑表征图网络、量子计算网络和多任务协同优化模块组成. 首先通过序列拓扑表征图网络获取单模态特征,利用图结构有效捕捉节点之间动态关系,添加多头注意力机制自适应调整权重,增强模型的情感表征能力;然后将拼接后的单模态特征输入量子计算网络提取多模态情感特征,通过量子编码、纠缠、测量等操作实现模态间情感信息的高效共享与传递;最终利用单模态与多模态分析结果进行多任务学习,基于自监督伪标签生成和模态权重调整策略实现模型协同优化,缓解模态表征不一致性.

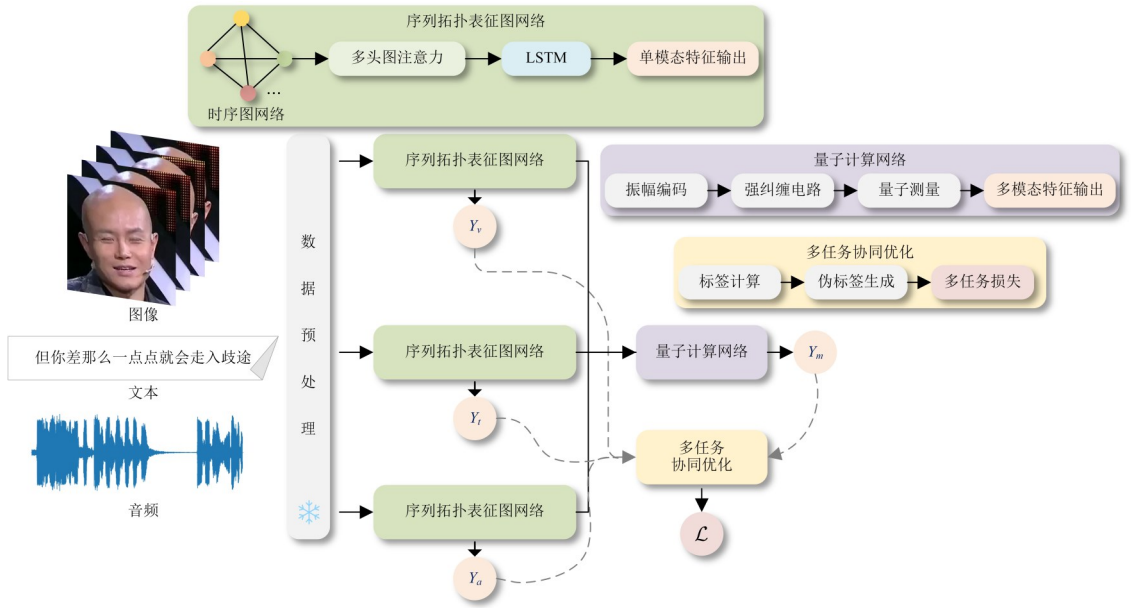

图3 混合量子与图神经网络的多模态情感分析模型整体框架

3.1 任务定义

本文的多模态情感分析任务指通过视频信息预测人的情绪极性与强度,包含视觉v、文本t和音频a三种输入模态,定义为$\boldsymbol{X}_i \in \mathbb{R}^{(T_i \times d_i)}, i \in \{v,t,a\}$,其中$T_i$为样本序列长度,$d_i$为当前时间步内的特征. 通过序列拓扑表征图网络输出的单模态情感分析结果表示为$Y_i, i \in \{v,t,a\}$,经过量子计算网络融合后多模态情感分析结果表示为Y_m,经过多任务协同优化后的损失函数为$\mathcal{L}$.

3.2 序列拓扑表征图网络

本文序列拓扑表征图网络结构如图4所示. 输入数据以图形式表示为$G=(V,E)$,其中$v_p \in V$为图网络节点,$e_{pq}=(v_p,v_q) \in E$表示节点v_p至v_q的边特征,节点v_p的邻域定义为$N(v_p)=\{v_q \in V | (v_p,v_q) \in E\}$,邻接矩阵$\boldsymbol{A}$尺寸为$T_i \times T_i$,用于判断序列节点间是否存在边连接,如式(1)所示:

$$\begin{cases} \boldsymbol{A}_{pq}=1, & e_{pq} \in E \\ \boldsymbol{A}_{pq}=0, & e_{pq} \notin E \end{cases} \tag{1}$$

标准化后的邻接矩阵如式(2)所示:

$$\tilde{\boldsymbol{A}}=\boldsymbol{D}^{-1/2}(\boldsymbol{A}+\boldsymbol{I})\boldsymbol{D}^{-1/2} \tag{2}$$

其中,$\boldsymbol{I}$为单位矩阵,用于添加节点自身信息;$\boldsymbol{D}$为度矩阵,如式(3)所示:

$$\boldsymbol{D}_{pp}=\sum_q(\boldsymbol{A}_{pq}) \tag{3}$$

将样本至图卷积网络(Graph Convolutional Networks,GCN),以不同时序特征为节点构建边连接,如式(4)所示:

$$\mathrm{GCN}(\boldsymbol{X}_i,\tilde{\boldsymbol{A}})=\sigma(\tilde{\boldsymbol{A}}\boldsymbol{X}_i\boldsymbol{W}+b) \tag{4}$$

其中,$\boldsymbol{X}_i$为输入样本节点特征矩阵;$\tilde{\boldsymbol{A}}$体现边的连接关系;$\boldsymbol{W}$为训练的权重矩阵;b为偏置项;σ为非线性激活函数.

由于单模态情感的时序表征具有随机性,为所有图节点赋予相同的权重忽略了其在情感分析中的重要性差异,在图卷积过程前增加多头注意力机制为节点分配不同权重,增加图神经网络的情感表征能力,使模型能够关注到更具代表性的特征. 图神经网络的多头注意力过程在经典注意力机制方法基础上将输入特征拆分为多个[27],兼顾节点属性与边属性,提高模型的表达和学习能力,添加多头注意力机制后的图网络节点更新如式(5)所示:

$$\boldsymbol{h}'_p = /\!/_{\text{head}=1}^{\text{head}} \sigma\left(\sum_{q \in N(p)\cup p} a_{pq}^{(\text{head})} \boldsymbol{W}^{(\text{head})} \boldsymbol{h}_q \right) \tag{5}$$

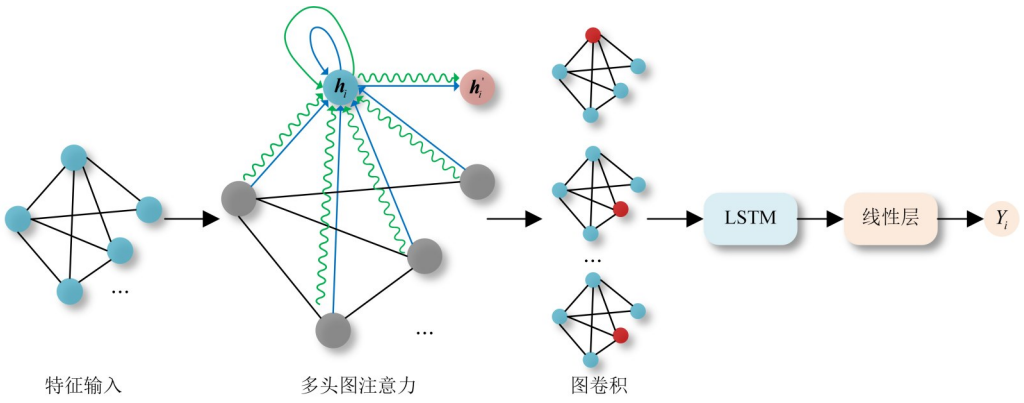

图4　序列拓扑表征图网络

其中,$\boldsymbol{h}'_p$表示添加多头注意力后节点p的新特征表示;$/\!/$表示向量拼接;head表示网络设定的注意力头数,本文设置为2;a_{pq}为注意力权重,用于衡量节点p与邻接节点q间的连接强度,如式(6)所示:

$$a_{pq} = \text{softmax}\left(g\left(\omega^{\mathrm{T}} [\boldsymbol{W}_p \boldsymbol{h}_p /\!/ \boldsymbol{W}_q \boldsymbol{h}_q] \right) \right) \tag{6}$$

其中,$g(\cdot)$为LeakyReLu激活函数;softmax$(\cdot)$表示归一化计算;ω为可学习的权重参数;$\boldsymbol{h}_p$、$\boldsymbol{h}_q$、$\boldsymbol{W}_p$、$\boldsymbol{W}_q$分别代表节点p与节点q的特征表示及权重矩阵.

经过改进图网络的单模态输出表示为$\boldsymbol{G}_i \in \mathbb{R}^{(T_i \times (\text{head} \times d_i))}, i \in \{v,t,a\}$,添加LSTM层进一步提高模型对于序列上下文依赖关系提取,实现单模态情感拓扑关系和多尺度时序表征. LSTM包含隐藏状态、输入输出门、遗忘门等结构,过程简化为

$$\boldsymbol{X}_i^{\text{LSTM}} = \text{LSTM}(\boldsymbol{G}_i) \tag{7}$$

经过序列拓扑表征图网络处理后的单模态特征$\boldsymbol{X}_i^{\text{LSTM}}$用于后续网络进行模态融合,为了进行多任务学习以优化模型模内模间情感表征提取性能,额外添加线性层生成单模态情感分析标签Y_i,如式(8)所示:

$$Y_i = \text{softmax}\left(\boldsymbol{W}\boldsymbol{X}_i^{\text{LSTM}} + b \right) \tag{8}$$

3.3　情感特征量子计算网络

由于情感表达的模态异构性与不一致性常导致情感特征分布不均及模态极性歧义等问题,在多模态情感分析中,模型不仅需要捕获不同模态间的互补信息与细粒度关联,还需抑制冗余特征对情感判别的干扰.

针对上述问题,本文构建基于量子计算的融合特征提取网络,如图5所示,通过量子编码将多模态特征映射至高维希尔伯特空间,使各模态信息在量子态叠加中实现并行表示. 量子叠加与纠缠机制进一步促进模态间特征的深层次耦合与相互依赖建模,使不同模态的情感信息能够在量子空间中实现非线性融合. 最终,通过量子测量过程将叠加态坍缩至特定的本征态,实现量子态与情感特征的对应映射,从而获得更具判别性的多模态融合情感表征. 相较于传统的线性加权或注意力机制融合方式,量子计算网络在高维量子空间中能够以指数级特征表示能力捕获模态间的复杂交互关系,从而在融合阶段显著提升情感表征的有效性.

模态融合后的特征如式(9)所示:

$$\boldsymbol{X}_m = \text{concat}(\boldsymbol{X}_v^{\text{LSTM}}, \boldsymbol{X}_t^{\text{LSTM}}, \boldsymbol{X}_a^{\text{LSTM}}) \tag{9}$$

其中,concat$(\cdot)$表示特征拼接;$\boldsymbol{X}_v^{\text{LSTM}}$、$\boldsymbol{X}_t^{\text{LSTM}}$、$\boldsymbol{X}_a^{\text{LSTM}}$分别表示视觉、文本、语音信息经过序列拓扑图网络的情感特征输出矩阵.

将融合特征编码为多个量子比特输入至量子计算网络,量子比特类似于经典比特,但具有更丰富和随机的表征特性,单个量子比特$|\psi\rangle$如式(10)所示:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{10}$$

其中,$\alpha,\beta \in \mathbb{C}$并满足$|\alpha|^2 + |\beta|^2 = 1$;$|\cdot|$为复数的模;$|\alpha|^2$和$|\beta|^2$分别表示量子比特处于$|0\rangle$和$|1\rangle$状态概率.

本文采用振幅量子编码方法,在编码过程中不需要添加额外的量子门,降低了数据高维表征的复杂度,经过振幅量子编码后的叠加态$|\psi\rangle'$如式(11)所示:

$$|\psi\rangle' = \sum_{c=1}^{N} x_c |q_c\rangle, \quad N = 2^n \tag{11}$$

其中,x为量子计算网络输入;$|q\rangle$为计算基;N为输入特

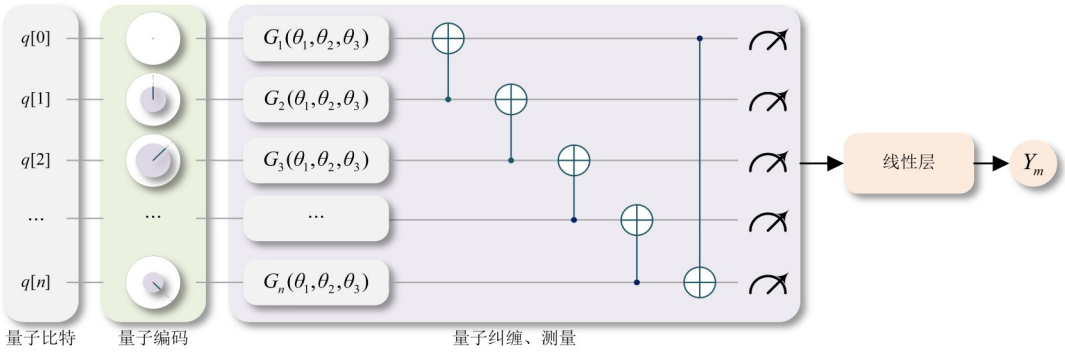

图5 情感特征量子计算网络

征数;n为量子比特数. 每批样本经过量子编码后得到振幅因子γ,如式(12)所示:

$$\gamma=A_{\text{norm}}\left((x_1^1,x_2^1,\cdots,x_N^1),(x_1^2,x_2^2,\cdots,x_N^2),\cdots,(x_1^M,x_2^M,\cdots,x_N^M)\right) \tag{12}$$

其中,M为批次样本数量;A_{norm}为归一化因子,使振幅因子γ满足式(13):

$$\|\gamma\|^2=\sum_{j=1}^{M}\sum_{c=1}^{N}\left|a_c^j\right|^2=1 \tag{13}$$

批次样本$\mathcal{X}=\left\{\left(x_1^1,x_2^1,\cdots,x_N^1\right),\left(x_1^2,x_2^2,\cdots,x_N^2\right),\cdots,\left(x_1^M,x_2^M,\cdots,x_N^M\right)\right\}$的量子集合表征如式(14)所示:

$$|\mathcal{X}\rangle=\alpha_c^j\left|q_c^j\right\rangle,\quad c\in[1,N],j\in[1,M] \tag{14}$$

将信息编码为量子态后,构建量子纠缠电路U映射希尔伯特空间内的批次量子特征$|\mathcal{X}\rangle$,用于提取每个样本经过模态间情感特征交互耦合的量子输出$|n\rangle$,如式(15)所示:

$$\left|n^j\right\rangle=U_\theta|\mathcal{X}\rangle,\quad j\in[1,M] \tag{15}$$

其中,j表示批次中第j个样本;n为量子比特数;θ表示量子电路中参数化变量. 整体量子纠缠电路可分解为多次酉变换,如式(16)所示:

$$U=U_1,U_2,\cdots,U_\ell \tag{16}$$

其中,U_ℓ表示第ℓ个单比特门或双比特门. 系统中单量子比特作用如式(17)所示:

$$U_\ell=\mathbb{I}_1\otimes\cdots\otimes G_k\otimes\cdots\otimes\mathbb{I}_n \tag{17}$$

其中,n表示系统量子比特数;$\mathbb{I}$为单位操作;G_k表示对第k个量子比特进行处理. 处理过程表示如式(18)所示:

$$G(\theta_1,\theta_2,\theta_3)=\begin{pmatrix}\mathrm{e}^{c\theta_2}\cos\theta_1 & \mathrm{e}^{c\theta_3}\sin\theta_1\\ -\mathrm{e}^{-c\theta_3}\sin\theta_1 & \mathrm{e}^{-c\theta_2}\cos\theta_1\end{pmatrix} \tag{18}$$

其中,$\theta_1,\theta_2,\theta_3$为可学习的角度参数.

经过单量子比特门后,添加双量子比特门将乘积态映射至非乘积态,实现量子纠缠,典型的受控门$C(G)$表示为

$$C_a\left(G_b\right)|x\rangle\left|y\right\rangle=|x\rangle\otimes G^x|y\rangle \tag{19}$$

当量子a的态x为0时,不进行纠缠,$G^0=\mathbb{I}$;当x为1时,在目标比特b应用多比特量子门. 本文采用**CNOT**门实现量子纠缠,如式(20)所示:

$$\textbf{CNOT}=\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&0&1&0\\0&1&0&0\end{bmatrix} \tag{20}$$

多个量子比特纠缠电路U可以最终表示为如式(21)的单比特与双比特量子门组合的形式:

$$U=\prod_{k=1}^{n}C_k(G_{k'})\prod_{k=1}^{n}G(k) \tag{21}$$

其中,k'为目标量子比特,由纠缠控制距离μ决定,如式(22)所示:

$$k'=\left((k+\mu-1)\bmod n\right)+1 \tag{22}$$

其中,mod表示模运算;$n=\lceil\log_2 N\rceil$为系统总量子比特数;N为输入量子网络的样本特征数;本文将μ设置为1. 单个样本经过量子演化后可以得到样本量子态$|n\rangle$,对量子态进行测量获得可观测的期望值$\hat{H}$,量子测量过程U_m如式(23)所示:

$$\mathrm{U}_m|n\rangle=\left\langle\hat{H}\right\rangle=\mathrm{Tr}\left(|n\rangle\langle n|\hat{H}\right) \tag{23}$$

其中,n表示第n个量子比特;$|n\rangle\langle n|$为密度矩阵计算;Tr表示迹运算,量子计算网络最终输出$\boldsymbol{H}$如式(24)所示:

$$\boldsymbol{H}=\left(\left\langle\hat{H}_1\right\rangle,\left\langle\hat{H}_2\right\rangle,\cdots,\left\langle\hat{H}_n\right\rangle\right) \tag{24}$$

最后将$\boldsymbol{H}$输入至线性层获取多模态预测标签Y_m,如式(25)所示:

$$Y_m=\mathrm{softmax}(\boldsymbol{W}\boldsymbol{H}+b) \tag{25}$$

3.4 多任务协同优化

受情感表征不一致性的影响,样本单模态标签与多模态标签往往存在差异,仅依赖融合特征进行的情感分析易表现过拟合. 受不同模态情感表征存在互补与深度关联的启发,本文构建多任务学习优化机制提高模型性能,将单模态与多模态预测作为多个子任务,通过伪标签生成与共享表征提高每个任务的性能,结合多任务损失函数,实现更好的泛化. 对于缺少单模态标签

的样本,首先进行单模态伪标签生成,定义情感分析任务训练过程中积极与消极样本标签中心C如式(26)所示:

$$C_{i,m}^{o}=\frac{\sum_{j=1}^{M}I\left(Y'_{i,m}(j)\right)\cdot F_{i,m}(j)}{\sum_{j=1}^{M}I\left(Y'_{i,m}(j)\right)} \tag{26}$$

其中,$o\in\{\text{pos},\text{neg}\}$,$i\in\{v,t,a\}$分别表示情感的两种极性与单模态类别;$m$表示融合模态;$F_{i,m}$为批次样本$j$经过序列拓扑表征图网络和量子计算网络输出的各模态特征;$I(\cdot)$为指示函数. 当计算积极标签中心$C_{i,m}^{\text{pos}}$时设定$Y'_{i,m}>0$,计算消极标签中心$C_{i,m}^{\text{neg}}$时设定$Y'_{i,m}\leqslant 0$. 计算单个样本特征与标签中心的欧式距离用于评估各个模态与标签中心差异r,表示为

$$r_{i,m}=\frac{\left\|F_{i,m}-C_{i,m}^{\text{pos}}\right\|_2-\left\|F_{i,m}-C_{i,m}^{\text{neg}}\right\|_2}{\left\|F_{i,m}-C_{i,m}^{\text{neg}}\right\|_2} \tag{27}$$

其中,$F_{i,m}$为当前样本输出的各单模态与融合模态特征;单模态与多模态距离差异(r_i-r_m)与标签值成正比,因此单模态伪标签Y_i'生成过程如式(28)所示:

$$Y_i'=Y_m'+\kappa(r_i-r_m),\quad i\in\{v,t,a\} \tag{28}$$

其中,Y_m'为已知的样本多模态标签;κ为比例因子,本文设置为1,表明使用数据集样本的单模态中心距离与融合模态中心距离的实际差异进行伪标签生成.

通过组合单模态和多模态情感分析子任务,利用MAE计算多任务损失函数,如式(29)所示:

$$\begin{aligned}L&=L_i+L_m\\&=\frac{1}{M}\sum_{j=1}^{M}\left|Y_i'(j)-Y_i(j)\right|+\frac{1}{M}\sum_{j=1}^{M}\left|Y_m'(j)Y_m(j)\right|\end{aligned} \tag{29}$$

其中,$Y_i'(j)$和$Y_m'(j)$分别表示单模态伪标签/真实标签与多模态真实标签;$Y_i(j)$和$Y_m(j)$表示网络预测结果. 所提算法整体流程如算法1所示.

算法1　混合量子与图神经网络的多模态情感分析方法

输入: 单模态特征$\boldsymbol{X}_i,i\in\{v,t,a\}$、标签/伪标签$Y'_{i,m},i\in\{v,t,a\}$与模型超参数

输出: 模型预测标签$Y_{i,m},i\in\{v,t,a\}$

模型:

1. FOR each sample in batch DO:
2. 依据式(1)~(6)计算添加多头注意力机制后的图网络输出$\boldsymbol{G}_i$
3. 依据式(7)计算序列上下文依赖关系$\boldsymbol{X}_i^{\text{LSTM}}$
4. 依据式(8)预测单模态情感分析标签Y_i
5. 依据式(9)~(14)计算批次样本的量子编码集合$|\mathcal{X}\rangle$
6. 依据式(15)~(22)计算各模态量子纠缠表征集合$|n\rangle$
7. 依据式(23)和式(24)计算量子网络输出$\boldsymbol{H}$
8. 依据式(25)预测融合模态情感分析标签Y_m
9. 依据式(26)~(28)对MOSI与MOSEI数据集样本生成单模态伪标签Y_i'
10. 依据式(29)计算多任务损失L并更新网络参数
11. END FOR

4　实验

4.1　数据集与评价指标

本文在多模态情感分析数据集CMU-MOSI[28]、CH-SIMS[16]与CMU-MOSEI[29]上进行系列实验,CMU-MOSI数据集搜集了在线分享网站的2 199个视频片段,旨在突出视频中观点信息多样性、模态间互补性和情绪强度重要性,时长在2~5 min范围内,每个视频片段使用[−3,3]的情感极性标签进行标记. CH-SIMS涵盖了来自60余个中文影视作品的2 281个视频片段,人物背景丰富,包含文本、音频、无音视频、视频的多种模态标签,用于支持研究人员进行多模态或单模态情感分析. CMU-MOSEI数据集包含了来自250个不同主题以及1 000余个不同演讲者的3 000余个视频,每个视频都包含了人工文本转录以及音素层面对齐,剪辑后的20 000余个视频片段采用多种方法标注,显著提升训练样本的多样性. 数据集样本量如表1所示,为确保实验结果的可比性,训练、验证及测试过程均采用发布方提供的官方划分集合. CH-SIMS与CMU-MOSEI数据集在训练验证过程的数据占比约80%,测试过程的数据占比约20%. CMU-MOSI数据集在训练验证过程的数据占比约70%,测试过程的数据占比约30%.

表1　实验数据集的详细统计数据

数据集	训练集	验证集	测试集	总计	范围
CMU-MOSI	1 284	229	686	2 199	[−3,3]
CH-SIMS	1 368	456	457	2 281	[−1,1]
CMU-MOSEI	16 326	1 871	4 659	22 865	[−3,3]

实验以二分类准确率Acc2、五分类准确率Acc5、七分类准确率Acc7、平均绝对误差MAE、皮尔逊相关系数Corr、F1作为评价指标,对模型进行多角度定量评估以验证算法有效性.

4.2　实验设置

本文使用Pytorch深度学习框架训练模型,实验学习率设置为0.005,训练的批处理大小为32,训练轮次为50,实验过程使用Adam(Adaptive moment estimation)优化器进行参数优化. 便于与其他基线模型对比,本文遵循多模态情感分析基线模型的信号处理流程,对图像、文本及语音模态数据进行预处理.

视觉模态方面,CMU-MOSI与CMU-MOSEI数据集采用Facet[30]提取面部表情特征,CH-SIMS数据集使用OpenFace[31]工具包进行预处理. Facet首先在视频帧中检测人脸及关键点(眉峰、鼻尖、嘴角等),随后提取20维动作单元、面部标志点及眼距等特征. OpenFace与Facet提取过程相近,同样包含人脸关键点检测、头部姿态估计、面部动作单元识别和眼动追踪等过程,特征序

列以时间顺序拼接. 文本模态方面,三个数据集使用不同语言的BERT[32]预训练模型提取中文文本与英文文本的词向量表征. 语音模态方面,CMU-MOSI与CMU-MOSEI数据集使用COVAREP[33]提取声学特征,CH-SIMS数据集使用Librosa[34]工具包进行预处理. COVAREP提取包含12维梅尔频率倒谱系数(Mel-Frequency Cepstral Coefficients,MFCCs)、音高、清浊分段特征、声门源参数(Glottal Source Parameters,GSPs)等特征. Librosa提取包含20维梅尔频率倒谱系数、对数基频(Logarithmic fundamental Frequency,Log F0)、恒定Q变换谱图(Constant-Q Transform,CQT)等特征.

4.3 实验结果与分析

4.3.1 基线模型对比实验

基线模型对比实验结果如表2所示,其中加粗字体表示最优结果. 为保证对比实验公平性,本文在相同实验环境下对基线模型进行复现,本文模型相较于常见的基线模型在各个评价指标均取得显著提升. 与基于单任务学习的LMF、MFN和Graph-MFN方法相比,在CMU-MOSI数据集的二分类准确率分别提高了约6.4%、7.4%和8.7%,七分类准确率提高了约13.1%、11.3%和12.2%. 在CH-SIMS数据集的二分类准确率分别提升了约1.5%、3.2%和3.5%,五分类准确率分别提升了约7%、7.2%和5.3%. 在CMU-MOSEI数据集的二分类准确率分别提高了约4.8%、6.1%和3.8%,七分类准确率提高了约2.3%、2.8%和1.7%. 结果验证了多任务学习优化策略能够共享子模态的情感信息进而提高模型性能的优势,通过引入更丰富特征来降低仅依靠融合模态情感分析的过拟合风险,在无单模态标签数据集如CMU-MOSI的标签稀缺情感分析任务中提升更加明显.

与MulT、MISA、Self-MM和MMIM方法相比,在CMU-MOSI数据集的二分类准确率分别提高了约6.8%、3.3%、2.6%和3.2%,七分类准确率提高了约14.5%、7.7%、3.9%和8.6%. 在CH-SIMS数据集的二分类准确率分别提高了约2.6%、1.5%、1.5%和4.6%,五分类准确率分别提高了约10.7%、7.2%、3.3%和6.2%. 在CMU-MOSEI数据集的二分类准确率分别提高了约5.5%、7.4%、3.2%和3.6%,七分类准确率分别提高了约1.7%、1.6%、1.5%和2.1%. 结果验证了序列拓扑表征图网络能够更准确地提取样本模态内时序情感表征,并结合量子计算网络实现模态间情感交互与深层次耦合. 综合表2的对比结果可知,本文所提方法和对比方法在CMU-MOSI数据集的表现效果优于CH-SIMS,可能由于CMU-MOSI数据集样本时长为2~5 min,与CH-SIMS数据集约10 s的样本时长相比,具有更多可用情感信息供模型提取与融合. LMF模型在CH-SIMS数据集上的相关性系数表现最佳,但其余指标表现相对一般,这一现象可能源于模型出现过拟合,使预测结果趋于某一固定值. 在样本标签分布不均衡条件下,此类集中预测会导致某一指标被动提高,从而掩盖模型在整体预测性能上的不足. 因此,需要多维度评价指标以对模型的综合性能进行更全面的分析. 此外,不同方法在CMU-MOSEI数据集的多分类准确率相较于其他两个数据集均有明显提升,可能是由于CMU-MOSEI数据集样本规模较大,对比CH-SIMS与CMU-MOSI增加了近十倍,因此具有更丰富的可学习特征并为模型训练提供更充分的数据支撑. 本文所提模型在小样本数据集的多分类准确率提升显著高于二分类准确率提升,验证算法模型能较好地平衡模态内与模态间的情感信息关联,在更复杂的情感分析任务中具有显著优势.

表2 本文模型与基线模型在CMU-MOSI、CH-SIMS与CMU-MOSEI数据集上的对比结果

模型	CMU-MOSI					CH-SIMS					CMU-MOSEI				
	Acc2/%	F1	MAE	Corr	Acc7/%	Acc2/%	F1	MAE	Corr	Acc5/%	Acc2/%	F1	MAE	Corr	Acc7/%
LMF[35]	80.0	80.1	0.952	0.671	35.7	77.9	78.2	0.429	**0.603**	39.2	80.2	80.8	0.565	0.730	51.9
MFN[36]	79.0	79.0	0.926	0.685	37.5	76.2	76.5	0.445	0.563	39.0	78.9	79.7	0.572	0.719	51.4
G-MFN[29]	77.7	77.9	0.994	0.665	36.6	75.9	76.4	0.432	0.592	40.9	81.2	81.7	0.563	0.728	52.5
MulT[13]	79.6	79.6	0.920	0.694	34.3	76.8	76.7	0.448	0.582	35.5	79.5	80.1	0.559	0.735	52.5
MISA[17]	83.1	83.2	0.784	0.772	41.1	77.9	77.2	0.447	0.570	39.0	77.6	78.5	0.553	0.755	52.6
Self-MM[37]	83.8	83.9	0.758	0.779	44.9	77.9	77.8	0.420	0.574	42.9	81.8	82.3	0.545	0.758	52.7
MMIM[38]	83.2	83.3	0.787	0.784	40.2	74.8	73.9	0.468	0.479	40.0	81.4	81.7	0.576	0.721	52.1
本文模型	**86.4**	**86.4**	**0.690**	**0.811**	**48.8**	**79.4**	**79.5**	**0.415**	0.591	**46.2**	**85.0**	**84.6**	**0.538**	**0.770**	**54.2**

4.3.2 消融实验

为验证所提算法中各模块的有效性,本文在CMU-MOSI、CH-SIMS与CMU-MOSEI数据集上进行系列消融实验,实验结果如表3所示,其中加粗字体表示最优结果. 相比于完整算法模型,去除序列拓扑表征图网络后模型在CMU-MOSI数据集的情感分析准确率分别下降2.4%和3.6%,相关性系数下降至0.789. 在CH-SIMS数据集的二分类准确率维持不变,五分类准确率下降

12.3%,相关性系数下降至0.557. 在CMU-MOSEI数据集的情感分析准确率分别下降0.5%和0.6%,相关性系数下降至0.759. 结果证明序列拓扑表征图网络有助于准确提取模态内时序情感特征,为后续模态融合提供基础. 去除量子计算网络后模型在CMU-MOSI数据集的情感分析准确率分别下降1.3%和4.3%,相关性系数下降至0.791. 在CH-SIMS数据集的情感分析准确率分别下降3.5%和3.7%,相关性系数保持不变. 在CMU-MOSEI数据集的情感分析准确率分别下降4.2%和0.5%,相关性系数下降至0.742. 结果表明量子计算网络能借助量子编码、纠缠与测量有效融合不同模态间情感特征,提高多模态情感分析任务准确率. 去除多任务协同优化策略后模型在CMU-MOSI数据集的情感分析准确率分别下降1.8%和0.6%,相关性系数下降至0.787. 在SIMS数据集的情感分析准确率分别下降6.1%和4.4%,相关性系数下降至0.525. 在CMU-MOSEI数据集的情感分析准确率分别下降0.7%和0.8%,相关性系数下降至0.745. 进一步表明多任务协同优化策略能缓解各模态情感特征歧义对结果的影响,使模型在样本各模态情感表征不一致的情况下保持较好的预测效果. 综合表3的消融实验结果可知,模型在CMU-MOSEI数据集上的消融结果相对稳定,这可能得益于较大的样本量在一定程度上弥补了去除单一模块后导致的特征提取能力不足. 此外,由于不需要过于复杂的时序情感特征与模态交互,去除单一网络模块后造成二分类情感分析任务准确率小幅度下降,与基线模型相比仍具备较高性能,但对于需要更深层次特征建模的多分类任务,性能下降更为明显.

表3　模型在CMU-MOSI、CH-SIMS与CMU-MOSEI数据集上的消融实验

模型变体	CMU-MOSI				CH-SIMS				CMU-MOSEI			
	Acc2/%	F1	Corr	Acc7/%	Acc2/%	F1	Corr	Acc5/%	Acc2/%	F1	Corr	Acc7/%
wo图网络	84.0	84.0	0.789	45.2	**79.4**	**79.5**	0.557	38.5	84.5	84.5	0.759	53.6
wo量子网络	85.1	85.0	0.791	44.5	75.9	76.5	0.574	42.5	80.8	81.1	0.742	53.7
wo协同优化	84.6	84.6	0.787	47.4	73.3	73.8	0.525	41.8	84.3	84.4	0.745	53.4
完整模型	**86.4**	**86.4**	**0.811**	**48.8**	**79.4**	**79.5**	**0.591**	**46.2**	**85.0**	**84.6**	**0.770**	**54.2**

为了进一步验证多任务学习中不同模态子任务对算法模型的贡献,本文在CMU-MOSI、CH-SIMS与CMU-MOSEI数据集上进行了融合模态与不同单模态情感分析任务组合消融实验,结果如表4所示,其中M、A、T、V分别表示融合模态、语音模态、文本模态和视觉模态. 随着组合的子任务数量增加,模型效果整体性能呈上升趋势,但在CH-SIMS数据集的五分类结果波动较大,可能由于样本时间较短,子任务间提取情感交互信息不足,易出现过拟合现象. 综合所有结果,视觉模态与文本模态给模型带来的增益更加明显,与语音模态相比,视觉模态包含多个面部区域运动与微表情等更丰富的特征,而文本模态由于经过人类语言的抽象与压缩,具有更鲜明的语义信息. 然而,受性格或文化差异等多重因素影响,音频模态存在更大的情感特征提取偏差. 此外,在仅使用融合模态子任务情况下,本文模型的情感分析结果仍优于多数对比模型,进一步验证了所提方法的整体有效性.

表4　利用不同子任务进行多模态情感分析实验结果

模态	CMU-MOSI				CH-SIMS				CMU-MOSEI			
	Acc2/%	F1	Corr	Acc7/%	Acc2/%	F1	Corr	Acc5/%	Acc2/%	F1	Corr	Acc7/%
M	84.8	84.8	0.807	46.5	75.9	76.3	0.552	41.1	84.1	84.2	0.757	53.7
M&A	85.5	85.5	0.806	47.7	76.4	76.8	0.572	40.3	83.8	83.9	0.758	52.9
M&T	85.2	85.2	0.803	48.4	75.5	76.1	0.547	43.1	83.5	83.5	0.757	51.6
M&V	85.5	85.5	0.803	47.1	77.7	77.5	0.568	44.2	83.1	83.3	0.755	53.0
M&A&T	84.9	85.0	0.800	**48.8**	78.3	78.5	0.545	42.7	83.1	83.3	0.761	53.1
M&A&V	85.4	85.3	0.808	46.9	76.6	76.7	0.548	42.9	84.6	84.7	0.754	52.2
M&T&V	83.8	83.9	0.794	48.4	79.0	79.0	0.558	41.6	84.9	**84.9**	0.759	54.1
全模态	**86.4**	**86.4**	**0.811**	**48.8**	**79.4**	**79.5**	**0.591**	**46.2**	**85.0**	84.6	**0.770**	**54.2**

4.3.3　模型超参数与复杂度分析

为验证模型设计过程中超参数设置的合理性,本文在CMU-MOSI、CH-SIMS与CMU-MOSEI数据集上分别针对序列拓扑表征图网络中图注意力头数head以及多任务协同优化中控制单模态伪标签生成的比例因子κ进行了超参数对比实验. 不同图卷积注意力头数的多模态情感分析结果如表5所示,其中加粗字体表示最优结果. 随着图注意力头数增加,CMU-MOSI与CH-SIMS数据集的评价指标出现不同程度下降,尽管增加注意力头能够增强模型的特征提取能力,但在样本数量有限的条件下更易导致过拟合. 在CMU-MOSEI数据集的实验结果变化趋于稳定,说明在大规模数据条件下模

型对注意力头数的敏感性相对较低. 综合不同数据集的实验结果,本文最终将注意力头数设定为2,以兼顾小样本数据集的稳定训练与模型对于单模态情感特征表征的需求,同时尽可能降低模型参数量.

使用不同比例因子大小进行伪标签生成的多模态情感分析结果如表6所示,其中加粗字体表示最优结果. 当比例因子设定为1.0时,伪标签生成依据单模态中心距离与融合模态中心距离的真实偏差进行,此时模型在CMU-MOSI与CMU-MOSEI数据集综合表现最佳. 随着比例因子的增大或减小,模型综合评价指标呈现不同程度的下降. 结果表明,比例因子在伪标签生成过程中能够有效调节单模态伪标签的情感倾向,使其更加偏向积极或消极,从而影响模型对于多任务损失的评估与参数更新方向.

表5 不同图卷积注意力头数的多模态情感分析结果

图注意力头数head	CMU-MOSI				CH-SIMS				CMU-MOSEI			
	Acc2/%	F1	Corr	Acc7/%	Acc2/%	F1	Corr	Acc5/%	Acc2/%	F1	Corr	Acc7/%
2	**86.4**	**86.4**	**0.811**	**48.8**	**79.4**	**79.5**	**0.591**	**46.2**	**85.0**	84.6	**0.770**	54.2
3	81.9	82.0	0.775	46.9	77.7	77.9	0.582	41.4	**85.0**	**85.0**	0.757	52.7
4	84.5	84.4	0.798	47.5	78.9	79.0	0.588	40.7	84.0	84.1	0.763	**54.3**
5	83.5	83.6	0.788	47.9	77.5	77.7	0.568	43.3	84.7	84.7	0.766	53.2

表6 不同比例因子大小的多模态情感分析结果

比例因子κ	CMU-MOSI				CMU-MOSEI			
	Acc2/%	F1	Corr	Acc7/%	Acc2/%	F1	Corr	Acc7/%
0.6	82.3	82.4	0.779	46.1	84.9	84.9	0.753	53.3
0.8	83.2	83.3	0.785	46.2	**85.0**	**85.1**	0.760	52.8
1.0	**86.4**	**86.4**	**0.811**	**48.8**	**85.0**	84.6	**0.770**	**54.2**
1.2	83.2	83.3	0.777	45.5	83.0	83.3	0.762	53.0
1.4	82.9	83.0	0.780	46.1	84.3	84.4	0.758	54.1

为进一步评估融合量子计算与图神经网络方法对模型复杂度的影响,本文统计了所提模型及多个基线模型的参数规模,结果如表7所示. 早期基线模型虽然参数量较少、复杂度较低,但其在多模态情感分析任务中的性能有限,预测结果稳定性亦相对不足. 随着网络结构深度的增加及硬件算力的提升,近年提出的模型在各数据集上的整体性能得到显著改善. 本文所提模型的参数量与近期代表性基线模型处于同一量级,且在三个常用多模态情感分析数据集上的各项评价指标均取得更优结果,表明所提方法在保证模型复杂度合理的同时,能够有效提升特征提取与融合能力,实现更高精度的多模态情感分析.

表7 利用不同子任务进行多模态情感分析参数规模

模型	模型参数量/M	模型	模型参数量/M
LMF	0.505	MISA	110.6
MFN	2.173	Self-MM	109.6
G-MFN	2.725	MMIM	109.7

4.3.4 可视化分析

为验证模型整体有效性,本文对数据集的随机测试样本进行可视化定性分析,图6展示了模型在CMU-MOSI、CH-SIMS与CMU-MOSEI数据集上的多模态情感分析结果,其中CMU-MOSI和CMU-MOSEI数据集仅包含融合模态标签. 在视觉模态中,模型将“撇嘴”和“皱眉”等代表负面情绪的面部动作预测为−0.4,将“点头”和“嘴角轻微上扬”等偏向积极情绪的面部动作预测为0.22,对将标注为中性的“摇头”和“抬眼”预测为−0.18;对于音频模态,将语调语速下降和常规语调分别预测为−0.4和−0.09,将语速较快、能量分布均匀的语音特征预测为0.41;对于文本模态,将“still boring”“竭尽所能”“呈现好”等关键词分别预测为−1.7和0.85,将“really good”“enjoy”等代表积极的关键词预测为1.52. 这些单模态预测结果体现了序列拓扑表征图网络能够准确提取模态内时序特征关联与代表信息,降低模态内噪声干扰.对于融合模态,模型能够准确将标签−1.8、0.6和2.00预测为−1.8、0.56和1.98,验证了量子计算网络能够有效融合各模态特征并较好地平衡互补特征与不一致特征. 此外,相比于仅以融合模态预测为目标任务的模型,多任务协同优化策略不仅提升了各模态的情感分析效果,也有效降低了预测结果受单一模态影响而产生的过拟合风险.

5 结语

本文针对模态内与模态间情感关系不均衡以及情感倾向不一致导致模型准确率受限的问题,提出了一种混合量子计算与图神经网络的多模态情感分析方法. 本文构建序列拓扑表征图网络以提取单模态时序情感特征关系,利用量子计算网络实现模态间互补特征深

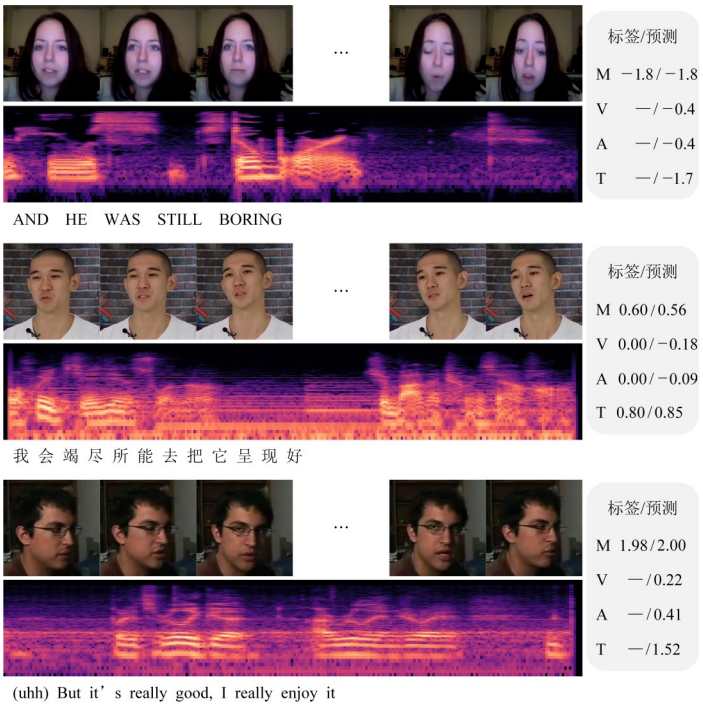

图6　本文模型多模态情感分析结果可视化

层次耦合,并减小冗余信息对分析结果的干扰. 同时,设计多任务协同优化策略,通过子任务标签生成与共享表征提升模型泛化性能. 整体模型在基准数据集CMU-MOSI、CH-SIMS和CMU-MOSEI的测试结果表明,能够有效克服多模态情感分析的"语义鸿沟",并超越基线对比模型. 此外,通过一系列消融实验验证各个模块有效性,并分析不同模态子任务对结果提升的贡献. 然而,情感分析仍有诸多方向值得探索,未来可以融合更多模态数据(如各种生理信号),以实现更加精准和多维的身心状态感知.

参考文献

[1] DAS R, SINGH T D. Multimodal sentiment analysis: A survey of methods, trends, and challenges[J]. ACM Computing Surveys, 2023, 55(13s): 1-38.

[2] LI W B, WU L, WANG C, et al. Intelligent cockpit for intelligent vehicle in metaverse: A case study of empathetic auditory regulation of human emotion[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(4): 2173-2187.

[3] ZHAI G L, YANG Y, WANG H, et al. Multi-attention fusion modeling for sentiment analysis of educational big data[J]. Big Data Mining and Analytics, 2020, 3(4): 311-319.

[4] ALAMOODI A H, ZAIDAN B B, ZAIDAN A A, et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review[J]. Expert Systems with Applications, 2021, 167: 114155.

[5] PANTIC M, ROTHKRANTZ L J M. Toward an affect-sensitive multimodal human-computer interaction[J]. Proceedings of the IEEE, 2003, 91(9): 1370-1390.

[6] 赵力, 将春辉, 邹采荣, 等. 语音信号中的情感特征分析和识别的研究[J]. 电子学报, 2004, 32(4): 606-609.
ZHAO L, JIANG C H, ZOU C R, et al. A study on emotional feature analysis and recognition in speech[J]. Acta Electronica Sinica, 2004, 32(4): 606-609. (in Chinese)

[7] GANDHI A, ADHVARYU K, PORIA S, et al. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions[J]. Information Fusion, 2023, 91: 424-444.

[8] 邵志文, 周勇, 谭鑫, 等. 基于深度学习的表情动作单元识别综述[J]. 电子学报, 2022, 50(8): 2003-2017.
SHAO Z W, ZHOU Y, TAN X, et al. Survey of expression action unit recognition based on deep learning[J]. Acta Electronica Sinica, 2022, 50(8): 2003-2017. (in Chinese)

[9] ZHU T, LI L D, YANG J F, et al. Multimodal sentiment analysis with image-text interaction network[J]. IEEE Transactions on Multimedia, 2023, 25: 3375-3385.

[10] CAI Y, HUANG Q B, LIN Z J, et al. Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: A multi-task learning approach[J]. Knowledge-Based Systems, 2020, 203: 105856.

[11] ZADEH A, CHEN M H, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis[EB/OL]. (2017-07-23)[2025-06-25]. https://arXiv.org/abs/1707.07250.

[12] WANG H B, REN C, YU Z T. Multimodal sentiment analysis based on multiple attention[J]. Engineering Applications of Artificial Intelligence, 2025, 140: 109731.

[13] TSAI Y H, BAI S J, LIANG P P, et al. Multimodal transformer for unaligned multimodal language sequences[C]// Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2019: 6558-6569.

[14] 张换香, 彭俊杰. 基于方面级情感分析的深度语义挖掘模型[J]. 电子学报, 2024, 52(7): 2307-2319.
ZHANG H X, PENG J J. A deep semantic mining model based on aspect-level sentiment analysis[J]. Acta Electronica Sinica, 2024, 52(7): 2307-2319. (in Chinese)

[15] ZHU L N, ZHU Z C, ZHANG C W, et al. Multimodal sentiment analysis based on fusion methods: A survey[J]. Information Fusion, 2023, 95: 306-325.

[16] YU W M, XU H, MENG F Y, et al. CH-SIMS: A Chinese multimodal sentiment analysis dataset with fine-grained annotation of modality[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 3718-3727.

[17] HAZARIKA D, ZIMMERMANN R, PORIA S. MISA: Modality-invariant and-specific representations for multimodal sentiment analysis[C]//Proceedings of the 28th ACM International Conference on Multimedia. New York: ACM, 2020: 1122-1131.

[18] ZENG Y F, LI Z X, TANG Z J, et al. Heterogeneous graph convolution based on in-domain self-supervision for multimodal sentiment analysis[J]. Expert Systems with Applications, 2023, 213: 119240.

[19] 蒋昆, 赵征鹏, 普园媛, 等. 基于跨模态超图优化学习的多模态情感分析[J]. 计算机科学, 2025, 52(7): 210-217.
JIANG K, ZHAO Z P, PU Y Y, et al. Cross-modal hypergraph optimisation learning for multimodal sentiment analysis[J]. Computer Science, 2025, 52(7): 210-217. (in Chinese)

[20] SUN H, NIU Z W, WANG H Y, et al. Multimodal sentiment analysis with mutual information-based disentangled representation learning[J]. IEEE Transactions on Affective Computing, 2025, 16(3): 1606-1617.

[21] ZHANG Y H, ZHANG Y, GUO W Y, et al. Learning disentangled representation for multimodal cross-domain sentiment analysis[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(10): 7956-7966.

[22] BIAMONTE J, WITTEK P, PANCOTTI N, et al. Quantum machine learning[J]. Nature, 2017, 549(7671): 195-202.

[23] HAVLÍČEK V, CÓRCOLES A D, TEMME K, et al. Supervised learning with quantum-enhanced feature spaces[J]. Nature, 2019, 567(7747): 209-212.

[24] LI Y C, QU Y, ZHOU R G, et al. QMLSC: A quantum multimodal learning model for sentiment classification[J]. Information Fusion, 2025, 120: 103049.

[25] PHUKAN A, PAL S, EKBAL A. Hybrid quantum-classical neural network for multimodal multitask sarcasm, emotion, and sentiment analysis[J]. IEEE Transactions on Computational Social Systems, 2024, 11(5): 5740-5750.

[26] 于瑞祺, 张鑫云, 任爽. 基于变分量子电路的量子机器学习算法综述[J]. 计算机研究与发展, 2025, 62(4): 821-851.
YU R Q, ZHANG X Y, REN S. A review of quantum machine learning algorithms based on variational quantum circuit[J]. Journal of Computer Research and Development, 2025, 62(4): 821-851. (in Chinese)

[27] YE Y, JI S H. Sparse graph attention networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 905-916.

[28] ZADEH A, ZELLERS R, PINCUS E, et al. MOSI: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos[EB/OL]. (2016-08-12)[2025-06-25]. https://arXiv.org/abs/1606.06259.

[29] ZADEH A B, LIANG P P, PORIA S, et al. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2018: 2236-2246.

[30] ROSENBERG E L, EKMAN P. What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) [M]. Oxford: Oxford University Press, 1997.

[31] BALTRUSAITIS T, ZADEH A, LIM Y C, et al. OpenFace 2.0: Facial behavior analysis toolkit[C]//2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). New York: ACM, 2018: 59-66.

[32] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Stroudsburg: ACL, 2019: 4171-4186.
[33] DEGOTTEX G, KANE J, DRUGMAN T, et al. COVAREP: A collaborative voice analysis repository for speech technologies[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2014: 960-964.
[34] MCFEE B, RAFFEL C, LIANG D, et al. Librosa: Audio and music signal analysis in python[J]. SciPy, 2015, 2015: 18-24.
[35] LIU Z, SHEN Y, LAKSHMINARASIMHAN V B, et al. Efficient low-rank multimodal fusion with modality-specific factors[EB/OL]. (2018-05-31) [2025-06-25]. https://arXiv.org/abs/1806.00064.
[36] ZADEH A, LIANG P P, MAZUMDER N, et al. Memory fusion network for multi-view sequential learning[J]. Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018, 32(1): 59-66.
[37] YU W M, XU H, YUAN Z Q, et al. Learning modality-specific representations with self-supervised multi-task learning for multimodal sentiment analysis[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(12): 10790-10797.
[38] HAN W, CHEN H, PORIA S. Improving multimodal fusion with hierarchical mutual information maximization for multimodal sentiment analysis[EB/OL]. (2021-09-16)[2025-06-25]. https://arXiv.org/abs/2109.00412.

作者简介

李兴广　男,1976年8月出生于吉林省农安县. 现为长春理工大学电子信息工程学院教授、副院长、博士生导师,吉林省拔尖创新人才、突贡专家. 获吉林省科技进步二等奖两项. 主要研究方向为多模态信息处理、微波毫米波技术、主动健康系统等.
E-mail: lixingguang@cust.edu.cn

蔡禹健　男,1999年4月出生于吉林省长春市. 现为长春理工大学电子信息工程学院博士研究生. 主要研究方向为计算机视觉与雷达信号处理.
E-mail: yujiantsai@mails.cust.edu.cn

崔　炜　女,1978年6月出生于吉林省长春市. 现为长春理工大学电子信息工程学院教授. 主要研究方向为信号处理技术、室内定位技术、机器人感知技术等.
E-mail: cuiwei@cust.edu.cn

李劲松　男,1998年3月出生于吉林省四平市. 现为长春理工大学电子信息工程学院博士研究生. 主要研究方向雷达信号处理.
E-mail: lijinsong@mails.cust.edu.cn

张莹瑀　女,1997年7月出生于内蒙古自治区兴安盟乌兰浩特市. 现为长春理工大学电子信息工程学院博士研究生. 主要研究方向为医学信号处理.
E-mail: zhangyingyu@mails.cust.edu.cn