DOI:10.3979/j.issn.1673-825X.202407070171

连贯性与语篇结构的对话情绪识别

杨上玮[1,2],李卫疆[1,2]
(昆明理工大学 信息工程与自动化学院,昆明 650500;2.昆明理工大学 云南省人工智能重点实验室,昆明 650500)

摘　要:现有的对话情绪识别模型在上下文建模时通常忽视了话语间的连贯性特征和语篇结构信息。为此,提出了基于连贯性特征与语篇结构的对话情绪识别模型。通过进行话语连贯性检测,排除弱连贯或不连贯的话语信息,采用构建连贯矩阵获取局部与全局的连贯信息;利用对话解析器构建语篇结构关系,并采用有向无环图进行语篇结构建模,同时传递语篇结构信息和说话者信息;通过交互注意力对连贯信息与语篇信息进行交互融合,生成情感标签。在 2 个公开数据集上进行了实验验证,结果表明,提出模型与现有相关模型相比,在性能指标上有一定的提升。

关键词:对话情绪识别;话语连贯性;语篇结构;图神经网络
中图分类号:TP391;TN99　　**文献标志码**:A　　**文章编号**:1673-825X(2025)05-0758-11

Dialogue emotion recognition based on coherence and discourse structure

YANG Shangwei[1,2], LI Weijiang[1,2]
(1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, P. R. China;
2. Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, P. R. China)

Abstract: The current dialogue emotion recognition models often overlook the coherence features and discourse structure information in context modeling. Therefore, this paper proposes a dialogue emotion recognition model based on coherence features and discourse structure. Firstly, discourse coherence detection is conducted to eliminate weak or incoherent discourse information, and both local and global coherent information are obtained by constructing a coherence matrix. Secondly, a dialogue parser is utilized to establish discourse structure relations, and a directed acyclic graph is employed to model the discourse structure while conveying both discourse structure information and speaker information. Finally, through interactive attention, coherent information and discourse information are interactively integrated to generate emotional labels. This paper validates the proposed model using two public datasets, with results indicating that compared to existing models, the proposed model demonstrates certain improvements in performance indices.

Keywords: dialogue emotion recognition; discourse coherence; discourse structure; graph neural network

收稿日期:2024-07-07　　**修订日期**:2025-01-03　**通讯作者**:李卫疆　hrbrichard@126.com
基金项目:国家自然科学基金项目(62066022)
Foundation Item: National Natural Science Foundation of China (62066022)

0　引　言

随着社交对话数据的公开以及自然语言处理领域的发展,对话中的情绪识别(emotion recognition in dialogue,ERC)也逐渐成为情感分析领域新兴且重要的领域,特别是随着 ChatGPT 等对话模型的公开,分析对话中说话者的情感也越来越受到关注。ERC 在人工客服问答、移情对话系统、医疗等领域有着重要作用。

目前对 ERC 进行建模的方法主要分为基于序列与图两类。Hu 等[1]设计上下文推理网络,使用多轮推理模块能够充分理解对话上下文并整合情感线索。Shen 等[2]将预训练语言模型应用于 ERC,使模型可存储更长的历史上下文,且设计对话感知注意力处理说话者之间的依赖关系。以上采用基于序列的方法虽然考虑了对话的时间顺序和邻近上下文对 ERC 的影响,但这些模型倾向于考虑与目标话语相邻的有限信息来更新话语表示,而无法对全局上下文进行有效建模。

随着图神经网络的发展,越来越多研究倾向于使用图神经网络(graph neural network,GNN)进行对话情绪识别。Shen 等[3]通过有向无环图结构传播对话历史上下文信息。Yuan 等[4]为减少信息冗余,使用单层 GNN 捕获远程上下文信息。这些方法通过对会话构建图结构,从而对会话整体有效建模。

虽然以上方法能有效提升 ERC 的性能,但均忽略了 2 个问题。

1)忽略了对话自身的交互性结构引起的话语间连贯性对 ERC 的影响。与传统文本不同,传统文本常以一段或者一篇文本的形式出现,文本整体的连贯性与逻辑性较好,而对话是交互式结构,其连贯性涉及到多方因素,例如主题转化、说话者的思维特征等。对话虽然是连续的交流过程,但连续话语之间不一定是强连贯性的话语,因为单个话语所表达的内容与情感不一定都建立在邻近话语的基础上,对话连贯性及话语依赖关系示例图如图 1 所示,当说话者 A 说第 7 句话语时,其情感并未建立在之前的话语 5 或话语 6 之上,而是建立在话语 2 的基础之上。以往方法在聚合上下文信息时由于未考虑到话语的连贯性,导致将弱相关甚至不相关的话语进行关联,从而使得模型性能下降。因此,捕获话语连贯性可以为话语选择与其强连贯的话语,从而在一定程度上解决对话交互性引起的话语连贯性较差的问题,能更精准地捕获上下文信息。

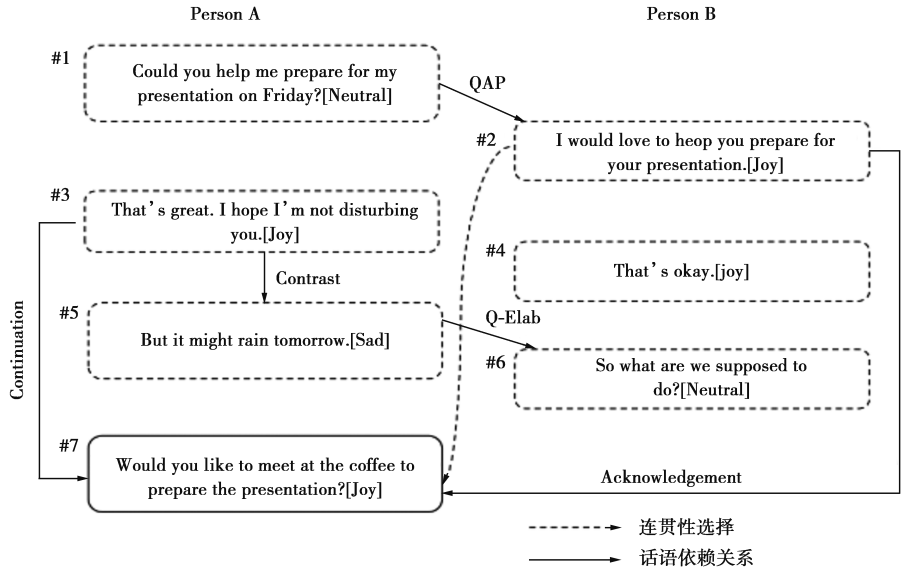

图 1　对话连贯性及话语依赖关系示例图

Fig.1　Dialogue coherence and discourse dependence diagram

2)忽略了语篇结构对 ERC 的影响。由于话语通常比较简短,一句话语中包含的有效信息较少,为了丰富话语语义信息,本文通过语篇解析器解析话语之间的关系类型,见图 1,语句 3 和语句 7 之间存在延续类型的话语结构关系,在语句 3 中,说话者 A 传达了对说话者 B 愿意帮助演讲的喜悦情感。即使语句 3 与语句 7 距离较远,但语句 3 和语句 7 存在延续性,所以当话语之间的距离较远时,依然可以通过这种结构关系识别出喜悦的情感。因此,通过对话语间的结构关系进行建模,能够丰富话语之间

的依赖关系,从而一定程度上避免缺乏依赖关系信息而导致的情绪识别错误。

综上所述,考虑到话语连贯性和语篇结构关系对 ERC 任务性能的影响,本文提出了 DialogCD 模型用于解决上述问题,本文主要贡献如下。

1)提出一种对话连贯性特征感知的方法,设计连贯性检测与选择,从而有效捕获局部连贯性信息与全局连贯性信息。

2)对会话的语篇结构进行建模,构建了话语间的依赖关系,为模型理解简短的话语信息提供了指导作用。

3)采用有向无环图对连贯信息与语篇信息建模,并采用交互注意力对 2 种信息进行有效交互融合。

1 相关工作

1.1 对话情绪识别

目前 ERC 的研究主要采用基于深度学习的方法,分为基于序列的模型、基于图的模型和引入外部知识的模型。

基于序列的模型基于时序,顺序地对 ERC 文本进行编码,例如 Zhang 等[5]在每个 LSTM 隐藏单元前面添加了一个置信门来确定前一个说话者的可信度,模拟前一个说话者的情绪影响。Majumder 等[6]采用 3 个 GRU 分别对说话人、对话上下文语境和之前话语的情感 3 个模块建模。这类方法通常依赖于来自附近话语的有限信息来更新当前话语的表示,并且序列神经网络存在梯度爆炸、梯度消失等问题致使信息丢失,从而导致模型对上下文信息的理解不完整。

由于图结构更贴切对话的整体结构,越来越多的模型使用 GNN 对会话进行建模。Ishiwatari 等[7]基于关系位置编码,将位置信息与 GNN 相结合,弥补了图神经网络不考虑序列信息的缺陷。Shou 等[8]对说话者关系和依赖句法结构建模,提高了模型获取语义信息和理解话语语法的能力。但上述研究基于固定窗口建立话语依赖关系,仍然容易考虑弱相关甚至不相关的上下文信息。

此外,有研究引入外部知识,即在基于序列和图的方法中加入额外信息辅助理解话语中的隐含信息。Zhong 等[9]融合图注意力机制动态利用常识知识。Zhang 等[10]使模型可在预训练阶段感知标签类别含义,有效区别相似情感标签。Chen 等[11]引入外部知识有效地解决了情绪转移问题。

虽然上述研究能在一定程度上提高 ERC 的效果,但均未考虑对话整体连贯性和话语局部连贯性导致的模型捕获弱相关甚至不相关上下文信息的问题,以及缺乏语篇结构依赖关系导致的话语语义信息单一的问题。

1.2 对话语篇解析

对话中的语篇解析旨在解析对话中话语的依存结构。Afantenos 等[12]首次提出对话中的语篇解析,其利用位置特征、词汇特征和解析特征这三类特征进行对话语篇解析。Li 等[13]基于矩阵嵌入解析图使模型能够自下而上和自上而下地进行语篇解析。Shi 等[14]提出的深度序列模型对对话中的基本话语单元进行顺序扫描,通过预测依存关系和联合交替构建话语结构来构建话语依存树。Bhatia 等[15]将语篇解析出的话语结构引入情感分析任务,提出了一个基于修辞结构理论结构的递归神经网络。因此,实验证明直观地利用语篇结构能够更好地编码非结构化人类对话,使模型专注于更加显著的话语,从而实现更准确的预测。

2 方法与模型

2.1 问题定义

对话 U 中包含 N 个话语 $U_i=\{u_1,u_2,\cdots,u_N\}$,给定说话人集合 $P_i=\{p_1,p_2,\cdots,p_m\}$($m\geqslant 2$),每一句话 u_i 由 p_i 中的一个说话者说出,给定情感标签集合 $y_i=\{y_1,y_2,\cdots,y_n\}$,ERC 的目标是预测每句话语 u_i 对应的情感标签 y_i。本文模型如图 2 所示。

2.2 文本特征提取

本文的模型采用 RoBERTa 模型来提取上下文无关语句级向量。对话语 $u_i=\{w_1,w_2,\cdots,w_N\}$,在话语的开头附加一个特殊标记[$CLS$],得到输入序列 $u_i=\{[CLS],w_1,w_2,\cdots,w_N\}$ 后将其输入到 RoBERTa模型中,得到最后一层隐藏状态表示 $\boldsymbol{h}_i=RoBERTa\{w_1,w_2,\cdots,w_N\}$,$\boldsymbol{h}_i\in\mathbf{R}^{d_u}$,$d_u$ 表示特征的维度,所有话语的特征表示为 $H_i\in\mathbf{R}^{N\times d_u}$。

2.3 连贯性特征感知

为了捕获话语连贯性信息,本文设计了话语连贯性特征感知模块。该模块分为 2 个步骤。

步骤 1 进行连贯性检测,通过计算话语间双向关注获取一段对话中每句话语之间的连贯性,即在获取 u_i 对 u_j 之间的关注时,也要获取 u_j 与 u_i 之

间的关注,对话语特征 $\boldsymbol{h}_i$ 采用协同注意力机制,该注意力机制能够同时关注到 2 个序列中的相关元素,并学习它们之间的交互关系,通过计算连贯特征矩阵 $\boldsymbol{s}_{ij}$ 获取话语之间的双向连贯性,连贯特征矩阵 $\boldsymbol{s}_{ij}$ 表示 u_i 与 u_j 之间的连贯性,表示为

$$\boldsymbol{s}_{ij} = W_1\boldsymbol{h}_i + W_2\boldsymbol{h}_j + W_3(\boldsymbol{h}_i \odot \boldsymbol{h}_j) \tag{1}$$

式(1)中:W_1,W_2,W_3 为可训练参数;⊙表示逐元素乘法。

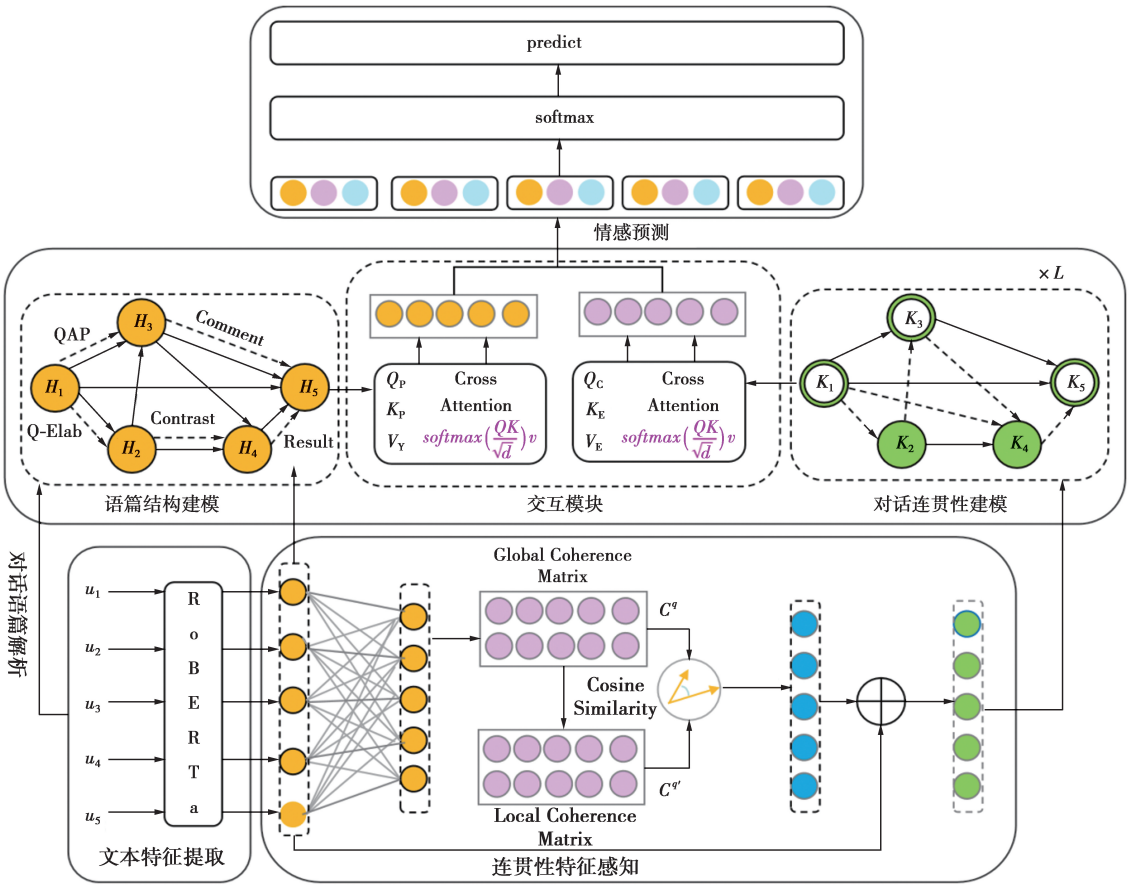

图 2　DialogCD 模型图

Fig.2　DialogCD model diagram

为了获取 u_i 的局部连贯特征向量,对 $\boldsymbol{s}_{ij}$ 检索 $i\pm1$ 范围内的话语,获取 $\boldsymbol{s}_{ij}$ 的子矩阵 $\boldsymbol{s}'_{ij}$,然后对 $\boldsymbol{s}_{ij}$ 与 $\boldsymbol{s}'_{ij}$ 应用逐行 *softmax*,得到 u_i 对其他话语的关注权重,即获取到 u_i 与全局话语之间的连贯性信息,以及 u_i 对其局部话语的关注权重,表示为

$$\alpha_{ij}^{q} = \frac{\mathrm{e}^{s_{ij}}}{\sum_{i=1}^{N}\mathrm{e}^{s_{ij}}} \tag{2}$$

$$\alpha_{ij}^{q'} = \frac{\mathrm{e}^{s'_{ij}}}{\sum_{i=1}^{N}\mathrm{e}^{s'_{ij}}} \tag{3}$$

由 α_{ij}^{q} 组合可以得到全局连贯性特征矩阵 $\boldsymbol{A}^q$,由 $\alpha_{ij}^{q'}$ 组合可以得到局部连贯性特征矩阵 $\boldsymbol{A}^{q'}$。通过将 $\boldsymbol{A}^q$ 和 $\boldsymbol{A}^{q'}$ 分别与话语上下文 $\boldsymbol{H}^q$ 嵌入相乘,可以得到话语的全局连贯特征向量矩阵与局部连贯特征向量,表示为

$$\boldsymbol{C}^q = \boldsymbol{A}^q \cdot H^q \tag{4}$$

$$\boldsymbol{C}^{q'} = \boldsymbol{A}^{q'} \cdot \boldsymbol{H}^{q'} \tag{5}$$

因此,得到话语全局关注特征向量 $\boldsymbol{c}_i \in \boldsymbol{C}^q$ 与局部关注特征向量 $\boldsymbol{c}'_i \in \boldsymbol{C}^{q'}$。

步骤 2　进行连贯性选择,排除弱连贯或不连贯的话语。为了有选择性地选择全局连贯信息,本模块将比较每个话语的全局连贯特征向量之间的相关度,即若话语 u_i 与 u_j 的连贯特征向量大于 u_j 与 u_s 之间的连贯特征向量,则认为话语 u_i 与 u_j 之间的连贯性与逻辑性更强,反之则更弱。具体来说,使用余弦相似度公式计算 2 个话语连贯特征向量之间的相似性。即对话语 u_i 来说,依次计算 u_i 的全局关注特征向量与其他话语全局关注特征向量之间的相似性,这样可以充分比较 2 个话语之间在全局上的相关性,表示为

$$y_i = \frac{\boldsymbol{c}_i \cdot \boldsymbol{c}_j}{\| \boldsymbol{c}_i \| \ \| \boldsymbol{c}_j \|} \tag{6}$$

式(6)中:$\boldsymbol{c}_i \cdot \boldsymbol{c}_j$ 表示话语上下文 u_i 和 u_j 的点积;$\| \boldsymbol{c}_i \|$ 和 $\| \boldsymbol{c}_j \|$ 表示向量 $\boldsymbol{c}_i$ 和向量 $\boldsymbol{c}_j$ 的范数;$y_i \in [-1,1]$表示 $\boldsymbol{c}_i$ 与 $\boldsymbol{c}_j$ 之间的局部连贯特征向量相似性,越接近 1 表示相似性越高,越接近-1 表示相反性越高,接近 0 表示相似性较低或基本无关系。

因此,在获取到话语非局部特征相似向量的余弦相似输出后,选取余弦相似度最接近于 1 的 $y_{\max}$,则其对应的话语非局部关注特征向量为 $\boldsymbol{c}_{\max}$,至此,本文在该模块获取了各个话语的全局关注特征向量集合 $M = \{\boldsymbol{c}_{m1}, \boldsymbol{c}_{m2}, \cdots, \boldsymbol{c}_{mn}\}$。

此外,以往模型在捕获局部上下文信息时只是简单聚合上下文,为了有选择性地聚合局部信息,同样采取余弦相似度的方法计算局部上下文之间的相关度,并选择最接近于 1 的 y'_i所对应的局部话语特征向量$\boldsymbol{c}'_{mi}$,可以表示为集合 $E = \{\boldsymbol{c}'_{m1}, \boldsymbol{c}'_{m2}, \cdots, \boldsymbol{c}'_{mn}\}$,于是 u_i 的连贯性信息为

$$k_i = \boldsymbol{c}_{mi} \oplus \boldsymbol{c}'_{mi} \tag{7}$$

式(7)中,⊕表示拼接操作,则一段对话的连贯性信息可以表示为集合 $k_i = \{k_1, k_2, \cdots, k_n\}$。

2.4 对话连贯性建模

连贯性特征感知模块能有效检测与当前话语较为连贯的局部与全局特征信息,并对该信息进行有效选择。而为了实现对会话整体的建模,本模块采用有向无环图(directed acyclic graph,DAG)结构,因为该结构满足会话的结构特点,能模拟对话信息的传播建模。但单纯的 DAG 并不考虑有选择性地聚合信息,只是重点关注邻近上下文节点,因此,本文在 DAG 的基础上考虑了 2.3 节中模块获取到的连贯特征信息,以便对会话进行全局连贯建模。

2.4.1 全局 DAG 建模

将获取到的连贯特征向量构建为有向无环图,具体地,将 DAG 定义为 $G = (V^c, E^c, R^c)$,其中,V^c 为话语节点集,即 $V^c \in k_i$;E^c 表示边信息传播,边关系类型 $R^c \in \{0,1\}$,当为同一人说话关系时,$r_{ij} = 1$,用实线连接。当为不同人说话关系时,$r_{ij} = 0$,用虚线连接。

此外,考虑到对话的交互性,为了使节点学习到更加丰富的信息,本文考虑了 2 个约束来选择 2 个话语如何进行连接。

约束 1(方向性) 为模拟对话从前向后的传播方式,信息只能由先前的话语传向未来的话语,即 $j>i, r_{ji} \notin E^c$,此约束确保对话为有向无环图。

约束 2(信息传递) 每个话语 u_i 与其先前所有话语连接,根据边关系类型 R 进行连接。

首先采用图注意力网络(graph attention network,GAT)让模型能更好地学习节点之间的信息,GAT 可以在图结构数据上运行,使用注意机制来关注邻居节点的表示。对于话语顶点 i,依次计算其与相邻顶点之间的相似系数,表示为

$$e_{ij} = a([W^{(l)} k_i^{(l)} \| W^{(l)} k_j^{(l)}]) \tag{8}$$

式(8)中:$a(\cdot)$是计算 2 个节点特征向量相关度的函数;$W^{(l)}$ 是该层节点特征变化(维度变换)的权重系数。

然后采用 *softmax* 进行归一化处理,表示为

$$\alpha_{ij}^c = softmax_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \tag{9}$$

根据注意力权重 α_{ij}^c,进行节点 i 特征加权求和

$$\boldsymbol{k}'_i = \sigma\left(\sum_{j \in N_i} a_{ij}^c W_4 k_j\right) \tag{10}$$

最后得到经过 GAT 增强后的节点特征向量 $\boldsymbol{k}'_i (i = 1,2,\cdots,n)$。

2.4.2 节点信息传播

学习节点之间的信息后,需要在 DAG 传播层进行节点信息传播,首先依次计算第一个话语到最后一个话语的隐藏状态。对目标语句 $\boldsymbol{k}'_i$,利用 $\boldsymbol{k}'_i$在当前层的隐藏状态和 $\boldsymbol{k}'_i$的前驱节点 $\boldsymbol{k}'_j$的隐藏状态,来计算 $\boldsymbol{k}'_i$与前驱节点之间在当前层的注意力权重,表示为

$$\alpha_{ij} = softmax_{j \in N_i}(W_5[\boldsymbol{k}'_j \| \boldsymbol{k}'_i]) \tag{11}$$

式(11)中:W_5 为可训练参数;$\|$ 表示拼接操作;N_i 为 c_i 的前驱集合。使用 $c_i^{(0)}$ 来初始化第 0 层语句节点表示。

为了更好地在每一层传递局部特征信息与远程信息,本文设计了关系感知特征,充分利用边的信息,在构造的 DAG 图的基础上进行图卷积。采用基于不同类型边的消息传递策略,表示为

$$\boldsymbol{k}_i'^{(l)} = RELU\left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{a_{i,j}^l}{c_{i,r}} W_6 \boldsymbol{k}_j'^{(l)} + a_{i,i}^l W_7 \boldsymbol{k}_i'^{(l-1)}\right) \tag{12}$$

式(12)中:N_i^r 表示节点 i 在边类型 $r \in R$ 下的节点集合;$a_{i,j}$和 $a_{i,i}$为权值;归一化常数为 $c_{i,r}$;N_i^r 为节点 i 在边类型 r 下的邻接节点数;W_6 和 W_7 为可训练参数。则在第 l 层可以获得 $\boldsymbol{k}'_i$的聚合表示 $\boldsymbol{k}_i'^{(l)}$。

2.4.3　图 Transformer 编码

基于 GNN 捕获节点信息时,其将节点自身的特征和每个邻居节点特征的聚合相累加,然后整体进行非线性变换,这会引起过平滑和过挤压的现象,因此,引入 L 层的图 Transformer 可以缓解该问题,且图 Transformer 处理图形数据时能够捕获节点之间的复杂关系,GNN 与图 Transofrmer 的对比如图 3 所示。

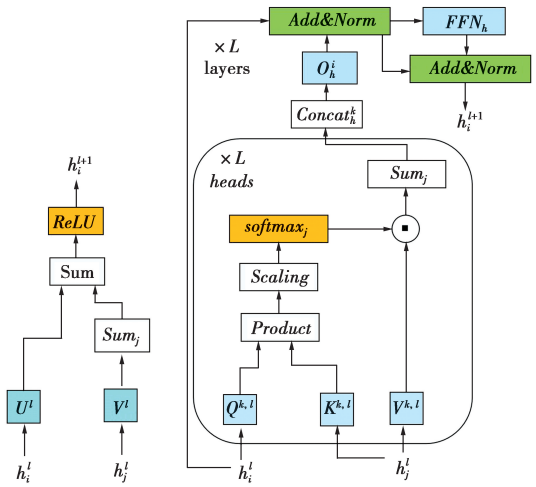

图 3　GNN 和 Graph Transformer 网络结构的对比图示

Fig.3　Comparison illustration of the GNN and Graph Transformer network architecture

图 Tansformer 由多头注意层和前馈网络层组成,在节点嵌入计算注意力分数。注意力层将文本序列映射到相同的高维空间,表示为

$$MultiHead(Q,K,V)=Concat(head_1,\cdots,head_n)\boldsymbol{W} \tag{13}$$

式(13)中:Q、K、V 分别表示查询、键、值;变量 $MultiHead(Q,K,V)$ 表示多头关注器训练的特征向量;$\boldsymbol{W}$ 是将 $MultiHead(Q,K,V)$ 线性投影到特定维度的投影矩阵。

因此,采用图 Transformer 传播每一层的信息,表示为

$$\boldsymbol{k}_i^{(l+1)}=(1-\beta_i)\Big(\sum_{j\in N_i}a_{i,j}m_j\Big)+\beta_i W_s\boldsymbol{k}_i^{\prime(l)} \tag{14}$$

式(14)中:第 i 个话语更新后的第 $l+1$ 层信息为 $\boldsymbol{k}_i^{(l+1)}$;N_i 为与目标节点 i 相连的源节点集合;m_j 为节点传递的消息;$a_{i,j}$ 为注意力得分;$\beta_i\in\mathbf{R}^l$ 为残差连接的门,可以缓解梯度消失问题,$W_s\in\mathbf{R}^{d_u\times d_u}$ 为映射权值。

2.5　语篇结构建模

在对话实际场景中,话语之间存在结构依赖关系。为了对对话文本进行结构化编码指向,同时丰富话语间的依赖关系,本模块对话语间的结构依赖关系进行建模,并使用 DAG 网络结构进行结构关系类型的传播。

2.5.1　节点信息传播

有效地利用话语结构依赖关系,使模型能够更好地对非结构化的人类对话进行编码。本文基于 Shi 等[14]的话语解析器预测对话中话语之间的依赖关系,通过预测依赖关系,联合或交替构建话语依赖树,使用该预训练的话语解析器来预测 ERC 数据集中 u_i 与 u_j 存在的话语依赖关系。从而得到关系得分四元组,即

$$\{(i,j,r_{ij},q_{ij}),\cdots\}=Parser(u_1,\cdots,u_n) \tag{15}$$

式(15)中:q_{ij} 表示每个话语依赖关系的置信度得分;r_{ij} 表示 u_i 和 u_j 之间存在 $r_{ij}\in R_2^{\mathrm{D}}$ 的结构依赖关系。

2.5.2　节点信息传播

获取到每句话语的依赖关系后,通过构建语篇结构图来传播和聚合图节点之间的话语结构信息,本文采用有向无环图结构进行语篇依赖结构建模,首先定义图 $G=(V^{\mathrm{D}},E^{\mathrm{D}},R_1^{\mathrm{D}},R_2^{\mathrm{D}})$,其中,$V^{\mathrm{D}}=(h_1,h_2,\cdots,h_n)$,$e_{ij}\in E^{\mathrm{Ds}}$ 表示 u_i 和 u_j 存在边,$r_{1ij}\in R_1^{\mathrm{D}}=\{0,1\}$ 表示话语之间说话者依赖类型的关系,$r_{2ij}\in R_2^{\mathrm{D}}=\{r_1,r_2,\cdots,r_{16}\}$ 表示话语之间的结构关系,R_2^{D} 具有 16 种关系(评论、澄清问题、阐述、确认、延续、解释、条件、问答对、交替、问答关系、结果、背景、叙述、纠正、平行、对比)。

以往研究采用普通有向图传播语篇结构信息,并不能有效对会话结构建模,而本模块在 DAG 的基础上增加边缘信息,不仅传递前驱节点的信息,还通过边传递语篇结构信息。

首先捕获 u_i 前驱节点的信息,采用 GCN 聚合当前层 u_i 与其前驱节点 N_i 的信息,表示为

$$\alpha_{ij}=softmax_{j\in N_i}(W_8[\boldsymbol{h}_j\,\|\,\boldsymbol{h}_i]) \tag{16}$$

$$\boldsymbol{h}_{gi}^{(1)}=\sigma\Big(\sum_{j\in N_i^{r_1}}\frac{a_{ij}}{c_{i,j}}W_9+a_{ii}W_{10}\boldsymbol{h}_i\Big) \tag{17}$$

式(16)—(17)中:α_{ij} 为权重;W_8、W_9、W_{10} 为可训练参数;$N_i^{r_1}$ 为在关系 R_1 下 $\boldsymbol{h}_i$ 的前驱集合,计算得到的 $\boldsymbol{h}_{gi}^{(1)}$ 则为节点 i 聚合了前驱节点后的信息。

接下来,计算话语 u_i 在说话者关系中的依赖特征向量为

$$\boldsymbol{h}_{gi}^{(l)}=\sigma\Big(\sum_{r\in R_1^{\mathrm{D}}}W_{11}\boldsymbol{h}_{gj}^{(1)}+W_{12}\boldsymbol{h}_{gi}^{(1)}\Big) \tag{18}$$

式(18)中:$\sigma(\cdot)$为激活函数;$\boldsymbol{h}_{gi}^{(1)}$为当前层中计算不同说话者依赖类型的特征向量。

获取到说话者依赖特征信息后,为了获取每个节点的语篇信息,需要先获取每个节点语篇信息的隐藏状态表示。

$$\alpha_{ij} = q_{ij} \tag{19}$$

$$\boldsymbol{h}_i^{\mathrm{D}} = \sum_{j \in N_i^{\mathrm{D}}} \alpha_{ij} W_{13} \boldsymbol{h}_j \tag{20}$$

式(19)—(20)中:α_{ij}表示节点 v_i^{D} 到其邻居节点 v_j^{D} 的边权重;N_i^{D} 表示节点 v_i^{D} 在图中邻居节点集合;$\boldsymbol{h}_i^{\mathrm{D}} \in \mathbf{R}^{d_h}$表示节点 v_i^{D} 在图网络更新后的隐藏状态;d_h 表示隐藏状态维度。每个节点在第 l 层的隐藏状态表示为 $H_i^{(l)} \in \mathbf{R}^{N \times d_h}$。

最后,为了捕获不同类型的节点对当前话语节点的重要性,对于包含不同类型的节点,通过将不同类型节点的特征空间进行连接,从而构造一个新的大型特征空间。即每个节点被表示为一个特征向量,而其他类型的无关维度的值为 0。

$$H_g^{(l+1)} = \sigma\left(\sum_{r \in R_2^{\mathrm{D}}} \widetilde{\boldsymbol{A}}_r \cdot \boldsymbol{h}_i^{(l)} \cdot \boldsymbol{W}_r^{(l)}\right) \tag{21}$$

式(21)中:$\widetilde{\boldsymbol{A}}_r$ 为矩阵,其中行表示所有节点,列表示其相邻类型为 r 的节点;节点 $H^{(l+1)}$ 的表示是利用不同的变换矩阵 $\boldsymbol{W}_r^{(l)}$,将不同类型 r 的相邻节点 H_r^l 的特征信息聚合得到的,变换矩阵 $\boldsymbol{W}_r^{(l)}$ 考虑不同特征空间的差异,并将其投影到隐式公共空间 $\mathbf{R}_q^{(l+1)}$ 中。从语篇依赖建模中可以得到 H_g。

则节点之间的传播信息为

$$\boldsymbol{h}_s = \boldsymbol{h}_{gi}^{(l)} \oplus \boldsymbol{h}_g \tag{22}$$

采用 GCN 更新节点信息

$$\boldsymbol{e}_i^l = RELU\left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{a_{i,j}^l}{c_{i,r}} W_r^l \boldsymbol{h}_{sj}^{l-1} + a_{i,i}^l W_0^l \boldsymbol{h}_{si}^{l-1}\right) \tag{23}$$

采用 l 层图 Transformer 传播信息

$$\boldsymbol{e}_i^{(l+1)} = (1 - \beta_i)\left(\sum_{j \in N_i} a_{i,j} \boldsymbol{h}_s\right) + \beta_i W e_i^l \tag{24}$$

式(24)中:$\boldsymbol{e}_i^{(l+1)}$ 表示第 i 个话语更新后的第 l+1 层信息;N_i 为与目标节点 i 相连的源节点集合,$a_{i,j}$为注意力权重;$W \in \mathbf{R}^{d_u \times d_u}$为映射权值。

2.6 交互模块

本模块采用交互注意力对 2 个模块内的输出进行信息交互,交互注意力层如图 4 所示。

首先,将连贯性建模模块与语篇依赖模块的输出作为输入序列 $P=(C,E)=\{\boldsymbol{c}_{m1},\cdots\boldsymbol{c}_{mk};\boldsymbol{e}_1,\cdots\boldsymbol{e}_n\}$,计算键值 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 分别表示为

$$\boldsymbol{K}_{\mathrm{P}} = PW^{\mathrm{K}} = (CW^{\mathrm{K}}, EW^{\mathrm{K}}) = (\boldsymbol{K}_{\mathrm{G}}, \boldsymbol{K}_{\mathrm{G}}) \tag{25}$$

$$\boldsymbol{Q}_{\mathrm{P}} = PW^{\mathrm{Q}} = (CW^{\mathrm{Q}}, EW^{\mathrm{Q}}) = (\boldsymbol{K}_{\mathrm{Q}}, \boldsymbol{K}_{\mathrm{Q}}) \tag{26}$$

$$\boldsymbol{V}_{\mathrm{Y}} = PW^{\mathrm{Y}} = (CW^{\mathrm{Y}}, EW^{\mathrm{Y}}) = (\boldsymbol{K}_{\mathrm{Y}}, \boldsymbol{K}_{\mathrm{Y}}) \tag{27}$$

式(25)—(27)中:$\boldsymbol{K}_{\mathrm{P}}$,$\boldsymbol{K}_{\mathrm{G}}$,$\boldsymbol{K}_{\mathrm{Q}}$,$\boldsymbol{K}_{\mathrm{Y}}$ 键向量;$\boldsymbol{Q}_{\mathrm{P}}$ 为查询向量;$\boldsymbol{V}_{\mathrm{Y}}$ 为值向量;C,E,P 为可学习参数;W^{K},W^{Q},W^{Y} 为权重参数。

其次,按照公式定义进行缩放点积注意力

$$Attention(Q_{\mathrm{P}}, K_{\mathrm{P}}, V_{\mathrm{P}}) = softmax\left(\frac{Q_{\mathrm{P}} K_{\mathrm{Y}}}{\sqrt{d}}\right) V_{\mathrm{Y}} \tag{28}$$

经过注意力层后,2 个模块的输出分别为

$$C' = Q_{\mathrm{C}} K_{\mathrm{C}} V_{\mathrm{C}} + Q_{\mathrm{C}} K_{\mathrm{E}} V_{\mathrm{E}} \tag{29}$$

$$E' = Q_{\mathrm{E}} K_{\mathrm{E}} V_{\mathrm{E}} + Q_{\mathrm{E}} K_{\mathrm{C}} V_{\mathrm{C}} \tag{30}$$

再次,将 C',E'发送到位置前馈子层。

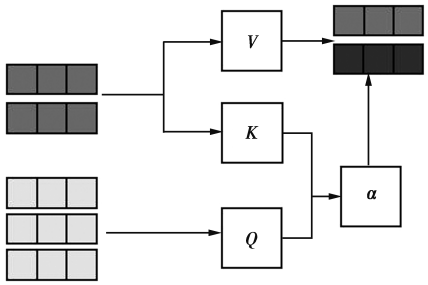

图 4 交互注意力层

Fig.4 Cross attention layer

最后,可以得到交叉关注模块中 Transformer 单元的输出,记为 $Y_{\mathrm{c}}=(H_{\mathrm{c}},E_{\mathrm{c}})$。将 2 个模块的输出进行拼接

$$\boldsymbol{Y}_{ci} = cat(\boldsymbol{c}'_{mi}, \boldsymbol{e}'_i) \tag{31}$$

2.7 情感预测模块

本模块对获取到的话语最终表示进行情感分类,通过 *softmax* 输出 u_i 情感分类的概率,表示为

$$Q_{\mathrm{out}_i} = softmax(W_{14} \cdot \boldsymbol{Y}_{ci} + b) \tag{32}$$

式(32)中:W_{14}为可训练的参数;b 为偏置。本文将情感分类定义为 3 种:消极、积极、中性,采用正则化交叉熵损失函数,表示为

$$J = -\frac{1}{N} \sum_i \sum_j y_i^j \log \hat{y}_i^j + \lambda_{\mathrm{r}} \| \theta \|^2 \tag{33}$$

式(33)中:y_i 为真实情感;$\hat{y}_i^j$ 为预测的情绪分布;i 为话语的索引;j 为类别的索引;λ_{r} 为 L2 正则化的系数。采用正则化能防止模型过拟合,并使用反向传播方法计算梯度并更新所有参数。

3　实　验

3.1　数据集

本文在 ERC 基准数据集 MELD 和 DailyDialog 上进行实验。MELD 是由 Poria 等[16]从电视节目《老友记》中收集的数据集;DailyDialog 是 Li 等[17]从英语交流学习者中收集的数据集,包含了大量日常对话。本文实验仅使用 2 个数据集的文本形式,数据描述如表 1 所示。

表 1　数据描述

Tab.1　Data description

数据集	对话数量			话语数量		
	train	val	test	train	val	test
MELD	1 038	114	280	9 989	1 109	2 610
DailyDialog	11 118	1 000	1 000	87 170	8 069	7 740

3.2　实验参数与评估指标

本文实验训练参数与评估指标如表 2 所示。

表 2　训练参数与评估指标

Tab.2　Training parameter and evaluation index

参数名称	参数值	
	MELD	DailyDialog
词向量维度	300	300
优化器	Adam	Adam
batch-size	32	64
学习率	1E-4	5E-5
dropout	0.3	0.4
传播层数	3	4
评估指标	Weighted-F1	Micro-F1

3.3　对比试验

为证明所提模型的有效性,本文从对话情绪识别现有模型中选择近两年效果较好且经典的模型进行对比,主要有以下几类模型。

3.3.1　基于序列的模型

DialogueCRN[1]:使用双向 LSTM 网络分别捕获说话者级和情景级语境信息,设计多轮推理模块,能更好地整合情感线索。

ERMC[18]:利用话语结构与说话者特定特征传播上下文信息线索,并利用门控卷积从相关话语筛选上下文信息。

3.3.2　基于图的模型

DAG-ERC[3]:利用有向无环图对对话中远距离和邻近上下文的信息流进行建模。

RBA-GCN[4]:在单层图神经网络下捕获远程上下文信息,减少信息的冗余,并进行多模态之间的交互。

DualGAT[19]:利用图注意力网络,同时考虑话语结构和说话者感知上下文信息,并使用交叉注意力机制对 2 个模块信息进行交换。

DialogueGCN[20]:利用基于图卷积网络建模对话中说话者内和说话者间的依赖关系。

DenoiseGNN[21]:设计了上下文过滤器,过滤相关性差和无信息量的话语,并通过门控机制调整特征表示中的信息内容。

ESIHGNN[22]:采用异构有向无环图神经网络动态更新和增强每个回合话语和说话者情绪状态表示。

CauAIN[23]:利用常识知识中的因果关系来建模说话者间的依赖关系。

3.3.3　基于预训练的模型

MPLP[10]:模仿人类思维过程将上下文信息与说话者背景信息输送到预训练语言模型中进行提示,并通过标签释义区别相似情绪。

CISPER[24]:利用与话语中情感表达相关的上下文信息和常识知识来构建可训练的连续提示。

本文在 2 个 ERC 数据集上评估本文模型的性能,实验结果如表 3 所示。实验结果表明,本文模型相较于以往模型的最佳效果,在 MELD 和 DailyDialog 数据集上分别提升了 0.64% 和 1.41%。说明模型具有竞争力的性能表现,且能够为 ERC 任务带来提升。在 2 个数据集上,基于图的方法总体优于基于序列的方法,表明基于图的方法更能有效对会话建模。而当引入外部辅助信息时,模型整体性能有进一步提高,例如,CauAIN 中提出的情感原因线索,使得模型能够捕捉到对话语境中更加深层且丰富的情感动态变化线索。ESIHGNN 采用异构有向无环图神经网络建模说话者情绪状态,并利用外部知识丰富图的边。这些模型虽然有效利用了外部知识、对话主题以及情感原因信息等,但均忽略了话语连贯性以及语篇结构对话语情感的影响,因此,本文模型相较于引入相关辅助信息的最佳模型相比,在 MELD 数据集提升了 1.88%,在 DailyDialog 数据集提升了 1.55%。

相较于 DAG-ERC,本文模型在 MELD 和 DailyDialog 数据集上分别提升了 4.13%和 2.80%,虽然 DAG-ERC 模型也采用 DAG 结构对会话整体进行建模,但其并未考虑到节点之间的话语连贯性以及话语依赖关系,只是简单聚合邻近上下文节点信息,相比之下,本文考虑话语连贯特征,有选择性地聚合相关节点,有效解决了话语节点考虑弱相关甚至不相关节点的问题,且利用话语依赖信息丰富了语义,从而使得本文模型效果高于 DAG-ERC。与 DualGAT 模型相比,虽然其也考虑了语篇结构对模型性能的影响,并使用有向图网络传播上下文信息,但传统有向图结构单一,并不如本文的 DAG 结构满足会话结构特点,相比之下,本文模型在 MELD 和 DailyDialog 数据集分别提升了 0.96%和 2.11%。

表 3 同领域模型对比

Tab.3 Comparison with the domain model

模型	MELD Weighted-F1/%	DailyDialog Macro-F1/%
DialogueCRN	58.39	—
ERMC	64.22	58.75
DualGAT(2023)	66.38	59.22
DialogueGCN	58.10	57.52
DAG-ERC	63.21	58.53
DenoiseGNN(2024)	66.70	59.60
ESIHGNN(2024)	63.92	59.78
RBA-GCN(2023)	65.67	58.33
CauAIN(2022)	65.46	58.21
MPLP(2023)	66.51	59.92
CISPER(2022)	66.10	59.17
DialogCD	**67.34**	**61.33**

注:模型中最佳结果用粗体表示,往年模型最佳结果用下划线表示。

3.4 消融实验

为了验证 DialogCD 每个组件的有效性,本文设计了消融实验,-feature 表示模型不再进行局部特征融合与选择性的聚合相关节点,-coherence 表示不对会话整体连贯性上下文进行建模,-structure 表示不对话语之间的结构关系类型进行建模。-interaction 表示不对 2 个模块的信息进行交互。

实验结果如表 4 所示,实验结果表明,去除模型的相关模块后,模型的性能均有所下降,说明了模型各个模块都对 ERC 任务有一定的影响。去掉话语局部特征融合后,模型性能随之下降,因为本文模型在采用 DAG 进行全局建模前,考虑了局部连贯性对话语的影响,从而能在全局建模时有效聚合更加有用的话语信息。去除语篇结构建模后,模型在 2 个数据集上分别下降了 1.52%和 1.41%,表明考虑话语间的结构关系类型能在一定程度上提升 ERC 的性能。而去掉交互注意力机制后,模型在 2 个数据集上分别下降了 0.87%和 0.75%,这是因为 2 个模块的信息无法通过交互注意力机制进行充分交互,进而导致模型性能下降。

表 4 消融实验

Tab.4 Ablation experiment

模型	MELD Weighted-F1/%	DailyDialog Macro-F1/%
DialogCD	67.34	61.33
-feature	66.13	60.39
-coherence	65.37	59.77
-structure	65.82	59.92
-interaction	66.47	60.58

3.5 Graph Transformer 层数分析

为了研究图 Transformer 层数对模型性能的影响,本文在对话连贯性建模模块和语篇依赖建模模块上分别逐渐递增图 Transformer 的层数。在 2 个数据集上得到不同层数的 Graph Transformer 所对应的 F1 值,实验结果如图 5 所示。

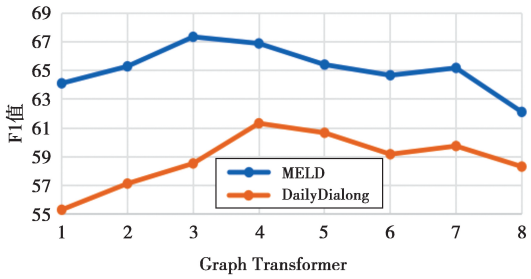

图 5 Graph Transformer 层数对应的 F1 值

Fig.5 F1 values corresponding to the number of Graph Transformer layers

从实验结果可知,2 个模块在 MELD 数据集上堆叠 3 层 Graph Transformer 时,性能达到最优,在 DailyDialog 数据集上堆叠 4 层 Graph Transformer 时,性能达到最优。这是因为当层数较少时,上下文信息与话语结构信息可能无法得到很好的提炼和交互;当层数较多时,可能会生成冗余或错误的节点表示,从而引入噪声导致模型性能下降。此外,由于 DailyDialog 数据集的对话数量比 MELD 多,所需要接收的上下文信息相对较多,如果堆叠层数较少,会

导致信息传播不完全,因此,相较 MELD 数据集需堆叠更多的层数。

4　总结和未来工作

本文综合考虑了对话的连贯性和语篇结构对会话情绪识别的影响,提出了一种基于连贯性与语篇结构建模的对话情感识别模型,通过检测局部和全局的话语连贯性,排除掉弱连贯或不连贯的话语,从而为当前话语选择强连贯特征信息,改善了以往图结构将弱相关甚至不相关的节点进行聚合的问题。通过语篇结构模块将非结构化的对话文本进行结构关系依赖指向,丰富了话语之间的依赖关系,并采用有向无环图结构模拟话语依赖关系在会话中的传播,为未具有明显情感基调的话语提供指导作用。本文设计了对比实验和消融实验验证了所提组件的有效性。

虽然本文模型对 ERC 任务的性能进行了提升,但对于一些对话中的隐性情绪,如讽刺、同情等还未进行深入研究。因此,在下一步工作中,将对话语隐性情绪检测做更深入的研究。

参考文献:

[1] HU D, WEI L, HUAI X. DialogueCRN: Contextual reasoning networks for emotion recognition in conversations [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Bangkok, Thailand: ACL, 2021: 7042-7052.

[2] SHEN W, CHEN J, QUAN X, et al. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI, 2021: 13789-13797.

[3] SHEN W, WU S, YANG Y, et al. Directed acyclic graph network for conversational emotion recognition[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics.Bangkok,Thailand:ACL,2021: 1551-1560.

[4] YUAN L, HUANG G, LI F, et al. Rba-gcn: Relational bilevel aggregation graph convolutional network for emotion recognition[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023(31): 2325-2337.

[5] ZHANG Y, TIWARI P, SONG D, et al. Learning interaction dynamics with an interactive LSTM for conversational sentiment analysis[J]. Neural Networks, 2021(133): 40-56.

[6] MAJUMDER N, PORIA S, HAZARIKA D, et al. Dialoguernn: An attentive rnn for emotion detection in conversations[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, America: AAAI, 2019: 6818-6825.

[7] ISHIWATARI T, YASUDA Y, MIYAZAKI T, et al. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations [C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Punta Cana, Dominica Republic: ACL, 2020: 7360-7370.

[8] SHOU Y, MENG T, AI W, et al. Conversational emotion recognition studies based on graph convolutional neural networks and a dependent syntactic analysis[J]. Neurocomputing, 2022(501): 629-639.

[9] ZHONG P, WANG D, MIAO C. Knowledge-enriched transformer for emotion detection in textual conversations [C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: ACL, 2019: 165-176

[10] ZHANG T, CHEN Z, ZHONG M, et al. Mimicking the thinking process for emotion recognition in conversation with prompts and paraphrasing[C]//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence.Macao,China:IJCAI,2023: 6299-6307

[11] 陈晏伊,李卫疆.融合时间序列和知识增强的双图融合对话情感识别[J].重庆邮电大学学报(自然科学版),2024,36(5):974-982.
CHEN Y Y,LI W J. Dialogue emotion recognition based on dual-graph fusion integrating time series and knowledge enhancement [J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2024, 36(5): 974-982.

[12] AFANTENOS S, KOW E, ASHER N, et al. Discourse parsing for multi-party chat dialogues[C]//Conference on Empirical Methods on Natural Language Processing (EMNLP 2015). Lisbon, Portugal: EMNLP, 2015: 928-937.

[13] LI W, ZHU L, SHAO W, et al. Task-Aware Self-Supervised Framework for Dialogue Discourse Parsing[C]// Findings of the Association for Computational Linguistics: EMNLP 2023. Singapore: EMNLP, 2023: 14162-14173.

[14] SHI Z, HUANG M. A deep sequential model for discourse parsing on multi-party dialogues[C]//Proceedings of the

AAAI Conference on Artificial Intelligence. Hawaii, USA:AAAI,2019, 33(01): 7007-7014.

[15] BHATIA P, JI Y, EISENSTEIN J. Better document-level sentiment analysis from rst discourse parsing[C]//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: ACL, 2015: 2212-2218

[16] PORIA S, HAZARIKA D, MAJUMDER N, et al. Meld: A multimodal multi-party dataset for emotion recognition in conversations[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: ACL, 2019: 527-536.

[17] LI Y, SU H, SHEN X, et al. DailyDialog: A manually labelled multi-turn dialogue dataset[C]//Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Vancouver, Canada: ACL, 2017: 986-995.

[18] SUN Y, YU N, FU G. A discourse-aware graph neural network for emotion recognition in multi-party conversation[C]//Findings of the Association for Computational Linguistics: EMNLP 2021. Punta Cana, Dominican Republic:EMNLP,2021: 2949-2958.

[19] ZHANG D, CHEN F, CHEN X. Dualgats: Dual graph attention networks for emotion recognition in conversations[C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Toronto, Canada:ACL,2023: 7395-7408.

[20] GHOSAL D, MAJUMDER N, PORIA S, et al. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China: ACL, 2019: 154-164..

[21] GAN C, ZHENG J, ZHU Q, et al. A graph neural network with context filtering and feature correction for conversational emotion recognition[J]. Information Sciences, 2024(658): 120017.

[22] ZHA X, ZHAO H, ZHANG Z. Esihgnn: Event-state interactions infused heterogeneous graph neural network for conversational emotion recognition[C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New York, USA:IEEE, 2024: 11136-11140.

[23] ZHAO W, ZHAO Y, LU X. CauAIN: Causal aware interaction network for emotion recognition in conversations[C]//Thirty-First International Joint Conference on Artificial Intelligence. Vienna, Austria: IJCAI, 2022: 4524-4530.

[24] YI J, YANG D, YUAN S, et al. Contextual information and commonsense based prompt for emotion recognition in conversation[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Vilnius, Lithuania: Cham: Springer International Publishing, 2022: 707-723.

作者简介:

杨上玮,硕士研究生,研究方向为自然语言处理、情感分析。E-mail:18250888621@163.com。

李卫疆,教授,博士生导师,博士,主要研究方向为自然语言处理、情感分析等。E-mail:hrbrichard@126.com。

(**编辑:王敏琦**)