

全局情境约束和局部多因素融合的对话情感识别

曹卫¹, 赵新元^{1*}, 薛煜阳²

(1. 新疆师范大学 计算机科学技术学院, 新疆 乌鲁木齐 830054; 2. 新疆财经大学 信息管理学院, 新疆 乌鲁木齐 830012)

摘要: 文本对话情感识别是自然语言处理领域中的一项重要研究任务,旨在自动识别对话文本各语句情感。然而,现有研究多侧重于对话语句的语义表征或对话人关系建模,忽略了对话交互过程中情感动态演变的影响因素。文章提出一种全局情境约束和局部多因素融合的对话文本情感识别方法,该模型不仅考虑了对话全局语义,还深入挖掘和建模了对话情感演变的多影响因素。在公开数据集上的实验结果表明该方法识别对话情感的有效性。

关键词: 对话情感识别;全局情境;局部多因素融合

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1008-9659(2024)03-0011-10

对话情感识别是指对社交媒体交互场景下的对话文本进行情感语义理解,以识别对话各子句的情感类别。该工作有助于商业领域的智能客服管理和科研领域的情感交互研究等。目前的相关研究大多使用神经网络基于对话文本深度语义表征或对话人关系建模来完成此任务,如 DialogueGCN^[1]、ConGCN^[2]和 DAG-ERC^[3]等。此外,也有学者融入常识来改进该任务,如 KET^[4]和 COSMIC^[5]等。这些工作都对对话情感识别研究有较大贡献。

对话文本交互性强,其语义和情感状态动态演变、十分复杂。探究其语义交互和情感演变过程中的各影响因素对准确识别其情感很重要^[6]。首先,对话存在多个说话者,其文本语句来自不同对话人,这不同于一般篇章。对话各子句的语义和情感与其所隶属的对话整体语义有较大关联,对话全局情境对各语句有较强的语义约束作用。其次,对话人背景文化、性格特点以及结构信息的差异均会对对话语句的情感状态产生影响。可见,对话人信息对对话情感也有较大影响。值得注意的是,在一个对话文本的局部邻近语句序列内,对话人进行“你来我往”言语互动,必然存在话题承转和信息流动^[7]。邻近位置的语句间以及蕴含直接话题承转的语句间的语义关联更强,相互的情感影响更大,即对话的局部邻近上下文及邻近上下文内部存在直接话题承转的语句对对话各语句的情感有较大影响。可见,对话文本的全局情境信息、各语句的局部对话人信息、局部情境信息和局部邻接情境信息都对对话情感有不可忽视的作用。

综上,文章同时考虑对话情感的全局语义信息和局部影响因素,高效识别对话各语句情感。该工作主要面临两大挑战:一是如何构建对话的全局情境,实现对各语句的语义约束;二是如何准确建模和融合影响对话情感的多个局部因素,促进对话语义理解和情感识别。为解决上述问题,文章提出了全局情境约束和局部多因素融合的对话情感识别模型(GCC-LMI)。针对挑战一,设计全局情境约束模块,利用对话语句间的序列语义特性建模对话全局语义,保证对话各语句的语义表征不偏离其所隶属的全局情境。针对挑战二,设计局部多因素叙述表征模块,基于对话语句的叙述连贯特性建模对话语句的各情感影响因素(局部情

[收稿日期] 2024-01-27

[修回日期] 2024-02-25

[基金项目] 新疆维吾尔自治区自然科学基金项目(2022D01A228);四川省区域创新合作项目(2023YFQ0066)。

[作者简介] 曹卫(1986-),女,讲师,主要从事人工智能、情感计算方面研究,E-mail:273060839@qq.com。

* [通讯作者] 赵新元(1974-),男,副教授,主要从事人工智能应用方面研究,E-mail:1621000@qq.com。

境信息、局部对话者信息、局部邻接情境信息),学习多因素融合下对话语句的情感状态,促进对话语义理解和情感识别。文章主要贡献如下:

第一,基于对话交互性导致的语义和情感动态演变,同时考虑了对话情感的全局情境信息和局部影响因素(局部情境信息、局部对话者信息、局部邻接情境信息),提出了全局情境约束和局部多因素融合的对话情感识别方法;

第二,基于对话文本的序列语义特点进行全局语义建模和局部叙述表征,获得对话文本的全局语义信息和影响对话情感的局部影响因素,高效理解对话情感语义;

第三,在公开基准数据集上的实验结果验证了所提方法的有效性。

1 相关工作

对话情感识别研究伴随着神经网络技术的发展而不断深入,陆续经历了基于传统循环神经网络的方法、基于知识融合的神经网络方法和基于图神经网络的方法等阶段。

1.1 传统的基于循环深度神经网络的对话情感识别

最初的对话情感识别研究大多利用循环神经网络对文本句子进行深度语义表示和上下文建模,实现对话文本的情感识别。如,Hsu等人利用卷积神经网络编码各对话语句的语义信息,再借助LSTM长短时记忆网络建模各语句的上下文语义关联,实现对话情感识别^[8]。相关模型基于循环神经网络的DialogueRNN^[9]和BiERU^[10]等。前者从对话过程中的信息流动角度考虑,利用注意力机制建模对话句子间的关联;后者使用擅长时序建模的RNNs模型从对话上下文中获取信息。Hazarika和Sukhbaatar等人考虑了对话文本的复杂语义,借助记忆网络对对话上下文进行建模来改善语义表征效率^[11-13],实现本任务。随着研究的深入,学者们注意到对话人的重要性,指出对话者语义交互会造成对话的语义关系发生变化。如,Yeh等人提出交互感知注意力网络(IANN),借助内在对话者关系建模对话语义,并同时进行对话行为识别和对话情感分类^[14]。

1.2 基于知识融合的对话情感识别

人类的对话活动离不开人类在生产、生活中的各种主客观知识。例如,对话中的回复语句一般是由回复者基于自身经验和知识对该语句的上文信息进行理解而产生的。这些知识参与对话交互过程,与对话动态语义和情感密切相关,可作为先验知识辅助对话情感语义理解。因此,相关学者引入人类先验知识来辅助对话情感识别。例如,Zhong等人引入外部知识增强预训练模型的语言表征能力,提升了对话文本的语义表征效率和情感识别效果^[4]。在此基础上,Ghosal等人提出对话情绪识别常识引导框架,引入常识知识学习话语者性格、意图及对话相关事件等,促进对话语义理解和情感状态建模^[5]。Li等人注意到心理状态在情绪活动中的重要性,指出对话者心理状态影响其行为和意图,进而导致其情绪状态演变^[15]。因此,该学者引入心理知识学习话语间的心理互动来改进该任务。类似工作还有CONSK-GCN^[16]和KI-Net^[17]等,前者联合常识知识和多模态信息辅助对话情感识别;后者引入情感极性强度预测的辅助任务,以多任务学习的思想推动对话情感识别研究。

1.3 基于图神经网络的对话情感识别

考虑到对话文本中话语者相互影响的复杂关系,一些学者将擅长关系建模的图神经网络^[18]引入该任务中。如DialogueGCN^[1]、ConGCN^[2]和RGAT^[6]等。其中,DialogueGCN考虑了对话者间相互影响和对话者个性差异导致的话题和语境的动态变化,用图神经网络建模对话者对句子语义的不同关系,并使用图神经网络的信息传播机制学习各语句间的语义关联和建模对话的动态上下文,实现对话情感识别。ConGCN考虑了对话上下文和对话者信息,将二者均看作图神经网络的节点来建模对话语义。RGAT则基于DialogueGCN并融合话语句子的位置信息,增强图神经网络的对话语句的关系建模,实现对话情感的识别。类似的工作还有DAG-ERC^[3],其使用有向无环图建模对话文本语义,并侧重于学习话语上下文和基于对话人身份的话语关系信息,以便更准确地理解对话语义及其情感。此外,由于Transformer的自注意力在一定意义上可看作是全连通图,一些学者尝试使用Transformer解决对话情感研究中复杂语义关系建模的问题^[19-20]。

2 问题定义

文章对话情感识别工作的正式定义如下:输入是对话 U 的所有文本句子,其中 U 中的每个句子都默认有一个情感类别(数据集样本特点,共涉及6个情感类别:快乐、悲伤、生气、沮丧、兴奋和中性,分别对应于 $\{0, 1, 2, 3, 4, 5\}$ 中的一个数字)。这里, U 表示如下

$$U = \{u_1, u_2, \dots, u_i, \dots, u_n\} \quad (1)$$

其中,下标 n 表示对话 U 的轮数, u_i 为 U 的第 i 个话语句子,其表示如下

$$u_i = \{w_1^i, w_2^i, \dots, w_j^i, \dots, w_m^i\} \quad (2)$$

其中,下标 m 为话语句子 u_i 的词语数目, w_j^i 为 u_i 的第 j 个词语。文章的对话情感识别任务是学习一个函数 F ,识别对话文档 U 中每个话语句子 u_i 的情感标签 y_i

$$y = F(U) \quad (3)$$

其中, $y = \{y_1, y_2, \dots, y_i, \dots, y_n\}$ 表示对话 U 中各子句的情感标签, $y_i \in \{0, 1, 2, 3, 4, 5\}$ 。

文章关于对话情感识别工作的定义可以通过图1中的实例来说明。如图1所示,就输入形式而言,输入信息是包含多个子句的一个对话文档。就文本内容而言,输入文档的各子句均对应一个情感类别。在此基础上,文章旨在识别该对话文档的各个子句的情感类别,并标记为 $\{0, 1, 2, 3, 4, 5\}$ 其中一个数字,分别对应于本数据集的6个情感类别(快乐、悲伤、生气、沮丧、兴奋和中性)中的一种。

对话文档:	说话人	情感类别
子句 1: 我想我做不到了。	A	沮丧
子句 2: 嗯,或许,你还没有足够努力。	B	中性
子句 3: 已经三年了。我已经尝试了一切。	A	沮丧
子句 4: 也许你不够聪明。	B	中性
子句 5: 你为什么不再继续尝试呢?	B	生气
子句 6: 我足够聪明。我真的很擅长我所做的。我只是不知道如何让别人看到这一点。	A	生气

图1 对话情感识别任务举例

3 全局情境约束和局部多因素融合的对话情感识别

文章构建了全局情境约束和局部多因素融合的对话情感识别模型(GCC-LMI)。图2展示了该模型的整体架构,主要包括五个部分:(1)输入向量化;(2)全局情境约束;(3)局部多因素叙述建模;(4)局部多因素融合;(5)对话情感预测。下面将对这五个模块进行具体介绍,并叙述模型学习的过程。

3.1 输入向量化

参考前人的研究^[3,5,21],文章采用对话情感分析领域广泛使用的RoBERTa预训练模型^[22]得到对话文本的初始向量化表征。具体来说,对每个输入句子 $u_i = \{w_1^i, w_2^i, \dots, w_j^i, \dots, w_m^i\}$,在其形式化标记前加上一个特殊的起始标记[CLS]。至此,输入句子 u_i 变为

$$s_i = \{[CLS], w_1^i, w_2^i, \dots, w_j^i, \dots, w_m^i\} \quad (4)$$

随后,将序列 $\{s_1, s_2, \dots, s_i, \dots, s_n\}$ 输入到RoBERTa预训练模型进行微调,并取[CLS]在最后一层的嵌入作为各对话子句的初始向量化表示。对于对话文本 U 而言,其所有子句的初始化向量表示共同构成了 U 的句子级别的初始语义向量 S_0 为

$$S_0 = \text{RoBERTa}(s_1, s_2, \dots, s_i, \dots, s_n) \quad (5)$$

其中, $S_0 \in \mathbf{R}^{m \times d}$, d 是RoBERTa的输出向量的维数, s_i 表示第 i 个语句的初始语义表征, n 为该对话中的语句数目。

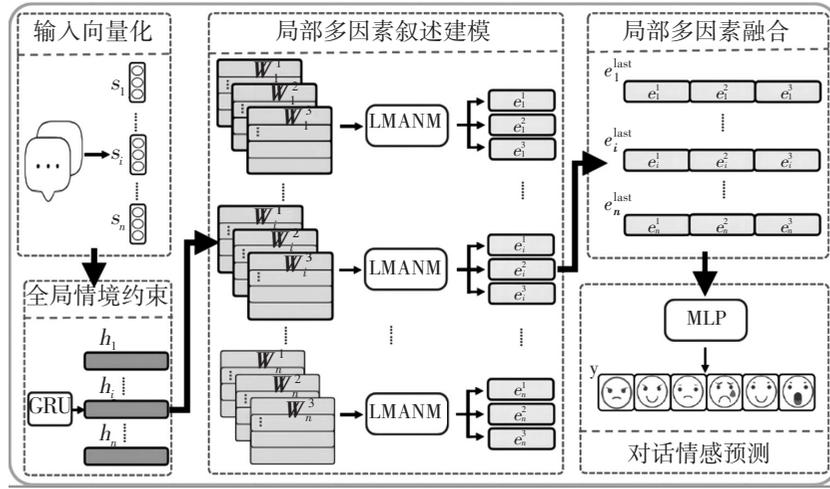


图2 全局情境约束和局部多因素融合的对话情感识别模型(GCC-LMI)的整体架构

3.2 全局情境约束

考虑到门控循环单元(GRU)^[23]在序列信息建模,尤其是在双向信息理解方面的优势,本节采用GRU进行对话语句的全局情境建模,以增强模型对各语句语义的全局情境理解,使其不偏离所隶属的对话全局上下文。

$$S_h = GRU(S_0) \quad (6)$$

其中, $S_h = \{h_1, h_2, \dots, h_i, \dots, h_n\} \in R^{m \times d_h}$ 为GRU处理的对话初始向量 S_0 中每个子句的隐藏状态, d_h 是GRU中隐藏状态的维数。至此,每个对话句子的语义表征蕴含其对话文本的全局上下文信息,句子语义的全局情境信息得以增强。

3.3 局部多因素叙述建模

本节设计实现了局部多因素叙述表征模块(LMANM),以建模影响对话情感的多个局部因素,即局部的情境信息因素、局部的对话者信息因素、局部的邻接情境信息因素,LMANM结构如图3所示。

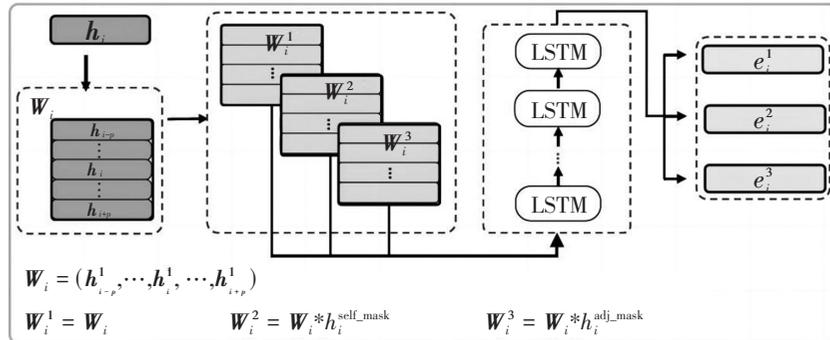


图3 局部多因素叙述建模模块(LMANM)

对话的话语轮转中,在语义呼应上各语句局部的邻近上下文比远距离上下文更具确定性。因此,LMANM先构建各对话语句的邻近上下文窗口,然后基于该窗口的语义信息进行对话情感的局部多影响因素的叙述表征,以准确学习影响对话情感的多个因素。以对话文本 U 为例,以该对话各语句为中心,引入各语句的邻近上下文窗口,则 U 的任意一个话语句子向量 h_i 所对应的邻近上下文窗口 W_i 可形式化为

$$W_i = \begin{cases} \{h_0, \dots, h_i, \dots, h_{i+p}\} & , i < p \\ \{h_{i-p}, \dots, h_i, \dots, h_{i+p}\} & , p \leq i \leq n-p \\ \{h_{i-p}, \dots, h_i, \dots, h_n\} & , i > n-p \end{cases} \quad (7)$$

其中, p 为对话语句的邻近上下文窗口的半径。如前文所述,基于对话语句的邻近上下文窗口 W_i , LMANM 模块将进行对话文本的局部多情感因素的叙述建模,相关原理如下。

考虑到局部多因素叙述建模模块(LMANM)的基础是局部多因素上下文,下面分别从基于局部情境信息因素的上下文、基于局部对话者信息因素的上下文、基于局部邻接情境信息因素的上下文三方面进行介绍。

由对话文本的语义和情感的动态交互特性可知,距离对话语句更近的局部对话内容更能影响该语句的情感状态。因此,本模块利用各语句的近邻上下文窗口来建模该句子基于局部情境信息的情感影响因素。基于这一思路,话语 h_i 的近邻上下文窗口 W_i 可直接作为该语句的局部情境上下文 W_i^1 ,即

$$W_i^1 = W_i \quad (8)$$

显然,对话语句的局部情境中对话者话语信息对各语句的情感影响更大。因此,同上,对话者信息影响因素的上下文也建立在各语句局部近邻上下文的基础上。即以对话语句的近邻上下文窗口 W_i 为基础,对对话者信息进行区分,得到基于局部对话者信息的上下文情境 W_i^2 ,即

$$W_i^2 = W_i * h_i^{self_mask} \quad (9)$$

其中, $h_i^{self_mask}$ 是对话语句 h_i 的对话者信息掩码,用于区分 h_i 的近邻上下文窗口 W_i 中的局部对话者角色,是根据数据集中关于对话者角色的标注信息转换得到的。在转换过程中,话语语句 h_i 的对话者信息掩码 $h_i^{self_mask}$ 为一个0/1向量,向量长度等同于 h_i 近邻上下文窗口 W_i 的语句数目。 $h_i^{self_mask}$ 中各元素的值根据语句 h_i 和其近邻上下文窗口 W_i 中的各语句的说话者信息来定。即如果 h_i 与 W_i 中的某个语句同属一个话语者,就将 $h_i^{self_mask}$ 对应位置的元素记为“1”,否则,就记为“0”。

同理,对话语句 h_i 的邻接情境信息上下文也以其近邻上下文窗口 W_i 为基础,并结合话语者邻接信息掩码来构建,即

$$W_i^3 = W_i * h_i^{adj_mask} \quad (10)$$

其中, W_i^3 为 h_i 的基于局部邻接情境信息的上下文, $h_i^{adj_mask}$ 为 h_i 的邻接情境掩码向量,其向量长度为 h_i 的近邻上下文窗口 W_i 的语句数目。 $h_i^{adj_mask}$ 的计算参考了 DAG-ERC^[3] 中的思路,即利用已知的对话者局部邻接信息和话语位置信息认知对话语句之间的信息传播流向,确定对话语句的直接近邻。具体来说,对 h_i 近邻上下文窗口 W_i 中的各语句而言,如果该语句和 h_i 同属一个说话者,且该语句在对话 U 中的位置处于 h_i 之前,则表明该语句与 h_i 之间存在确定的信息传播关系。对于这种与 h_i 存在确定信息传播的邻近语句,需要确定其在近邻上下文窗口 W_i 中的位置,并将 h_i 邻接信息掩码 $h_i^{adj_mask}$ 相应位置的元素记为“1”,否则,记为“0”。至此,文章获得了影响对话语句 h_i 情感的基于局部情境信息的上下文 W_i^1 、基于局部对话者信息的上下文 W_i^2 、基于局部邻接情境信息的上下文 W_i^3 。

基于话语 h_i 的各局部情感影响因素的上下文局部信息 W_i^1 、 W_i^2 和 W_i^3 ,利用局部多因素叙述建模模块 LMANM 进行 h_i 各局部因素影响下的情感状态表征,相关结构如图3所示。考虑到对话文本的语义连贯性比一般文本更强^[24],LMANM 利用 LSTM^[25] 进行时序语义建模,逐步学习 h_i 各局部情感因素上下文所对应的局部语句块的语义。以 h_i 局部情境因素上下文 W_i^1 为例,基于对话文本的连贯语篇所表现出来的语义一致性,LMANM 将 LSTM 处理语句块 W_i^1 的前 $2p - 1$ 个时间步的情感语义作为导致其最后一个时间步的情感语义的先导信息。对应地,LSTM 最后一个时间步输出的语义信息可看作语句块 W_i^1 在前 $2p - 1$ 个时间步的情感语义影响下的情感结果状态,将此情感结果状态近似为语句块 W_i^1 中心句子 h_i 的情感结果状态。同样地, h_i 局部情感影响因素上下文 W_i^2 和 W_i^3 也分别经由 LMANM 处理,并输出最后一个时间步的信息。

$$e_i^1 = LSTM(W_i^1) \quad (11)$$

$$e_i^2 = LSTM(W_i^2) \quad (12)$$

$$e_i^3 = LSTM(W_i^3) \quad (13)$$

其中, $e_i^1 \in d_w$ 、 $e_i^2 \in d_w$ 、 $e_i^3 \in d_w$ 分别代表对话语句 h_i 在其局部情境信息、局部对话者信息、局部邻接情境信息等因素影响下的情感状态表征。

3.4 局部多因素融合

本阶段将融合对话语句 h_i 的上述三种情感结果状态信息,得到最终情感状态。经多次实验探索,本节采用张量拼接的计算方法进行 e_i^1 、 e_i^2 和 e_i^3 的融合。

$$e_i^{last} = LSTM(e_i^1, e_i^2, e_i^3) \quad (14)$$

其中, $e_i^{last} \in d_w * 3$ 为局部多因素叙述表征的融合向量, 可视为 h_i 在多个局部情感影响因素影响下的最终情感表征。

3.5 对话情感预测

本模块通过多层感知机 *MLP* 和 *Sigmoid* 函数预测对话文本语句的情感, 即对话 U 的第 i 条语句的语义表征向量 e_i^{last} 被输入到 *MLP*, 并经过 *Sigmoid* 函数处理, 获得该语句的情感类别的概率向量 p_i 为

$$p_i = Sigmoid(MLP(e_i^{last})) \quad (15)$$

其中, $p_i \in R^n$ 表示对话 U 的第 i 条语句的情感概率向量。

3.6 模型学习

GCC-LMI模型包含两部分的学习过程, 即文本预训练和对话情感分类。预训练阶段, 文章采用 RoBERTa 语言模型处理对话句子(详情参见第 3.1 小节)。对话情感分类阶段, 考虑到本工作为分类问题, 作者使用交叉熵损失函数 *cross-entropy* 作为损失函数。

$$Loss = -\frac{1}{k} \frac{1}{n} \sum_{j=1}^k \sum_{i=1}^n y_i^j \log p_i^j \quad (16)$$

其中, k 为数据集中对话文本的数量, n 表示任意一个对话文本的句子数目, y_i^j 是第 j 个对话的第 i 条语句 c_i 的实际情感标签。为最小化 *Loss*, 文章采用 *Adam* 优化器更新模型参数。

4 实验与结果分析

4.1 数据集介绍

文章在对话情感识别的三个公开数据集^[26]上进行了实验, 数据细节如下。

(1) IEMOCAP 数据集^[27]。本数据集为多模态数据集, 其每一段对话均来自于两位演员根据剧本进行的表演, 文章实验仅使用了其中的文本模态。该数据集的数据样本涉及 6 种情感类别: 快乐、悲伤、生气、沮丧、兴奋和中性。考虑到 IEMOCAP 数据集无验证集, 遵循 DAG-ERC^[1] 中的做法, 对训练集中的数据进行分割, 取其中后 20 个对话进行模型验证。

(2) MELD 数据集^[28]。此数据集对话内容源自电视节目《老友记》, 蕴含多个模态的数据内容。与 IEMOCAP 不同的是, MELD 为多人对话形式, 共 7 种情感类别, 即快乐、悲伤、生气、惊讶、厌恶、恐惧和中性。

(3) EmoryNLP 数据集^[29]。同 MELD 一样, EmoryNLP 基于《老友记》, 涉及多人对话, 蕴含 7 种类别。但 EmoryNLP 在场景选择上与 MELD 有所不同, 蕴含的 7 个情感类别也与 MELD 稍有差别, 分别是快乐、强大、平和、悲伤、疯狂、恐惧和中性。

上述数据集的数据统计分析如表 1 所示。由表 1 可见, 各数据集在对话轮数、对话者数、情感类别数目等方面均存在差异, 可能会对实验结果产生一定的影响。

表 1 情感对话数据集的关键数据比较

数据集	对话总数	情感类别数	对话人数均值	对话子句数均值
IEMOCAP	151	6	2	66.8
MELD	1432	7	9	9.6
EmoryNLP	13118	7	2	7.9

4.2 实验设置

在数据集方面, 考虑到三个数据集中的对话人数差异, 文章基于双方对话这一假定进行对话人信息建模, 并借鉴 DAG-ERC^[3] 的数据划分方法划分数据, 以保证训练公平性。同时, 文章遵循前人的研究^[6,20], 采用加权平均指标 *weighted_avg* 进行性能评估。在实验配置上, 输入句子最大长度限定在 200 个字符之内, 采用预训练语言模型 RoBERTa 进行文本向量化, 通过对预训练模型进行微调, 得到维数 d_w 为 1024 的初始文本向量。在训练过程中, 初始化的权值采用均匀分布, 所有偏差初始化为零^[30], 普通隐藏层的维数 d_w 设置为 128, 激活函数选用 *Sigmoid* 函数, 使用学习率为 $1e-3$ 的 *Adam* 优化器在 NVIDIA Tesla K80 GPU 上进行训练。为更

高效地发挥模型的性能,文章设定 Dropout 为 0.5,以缓解可能出现的过拟合现象。

4.3 基准模型

为全面评估 GCC-LMI,基准模型选取了本领域较领先、较经典的代表性方法(如 DialogueRNN、DialogueGCN 等)和特别具有开创性的工作(如 KET、RGAT 和 DAG-ERC 等)。特别说明的是,考虑到 GPT 大语言模型强大的生成和理解能力,文章还选取大语言模型相关方法作为对比模型。

4.3.1 基于大语言模型的对话情感识别方法

ChatGPT Method^[31]基于具有强大语言处理能力的大语言模型 ChatGPT,并结合指令微调技术来完成对话情感识别。该方法巧妙地利用了 ChatGPT 的广泛语言知识,并通过指令微调技术进行模型训练和调参学习,使模型更好地建模对话文本中的内容信息,辅助模型高效理解对话中的情感语义信息,从而准确识别对话文本每个语句的情感类别。

4.3.2 基于深度神经网络的对话情感识别模型

DialogueRNN^[9]是一种具有代表性的基于序列建模框架的对话情感识别方法,其使用 3 个独立的 GRU 网络建模对话人以及历史上文的情境信息和情感信息来促进对话文本的情感语义理解,完成对话情感识别任务。DialogRNN+RoBERTa^[5]是 DialogueRNN 方法结合 RoBERTa 预训练语言模型的对话情感识别方法,该方法使用 RoBERTa 来进一步完善对话文本的语义表征质量,促进对话情感识别的效率提升。

4.3.3 融合知识的对话情感识别模型

KET^[4]是一种知识增强的 Transformer 框架,其动态利用外部常识知识和注意力机制建模和理解对话情感。COSMIC^[5]是基于知识的对话情感识别模型,利用外部常识性知识提高对话情感识别性能。TODKAT^[32]是基于知识的对话情感识别模型,同时融合了常识知识和对话主题信息。

4.3.4 基于图神经网络的对话情感识别模型

DialogueGCN^[1]是一种经典的基于图神经网络的对话情感识别模型,该方法基于对话者的互动信息构建图网络,并结合对话的全局上下文进行对话情感预测。DialogueGCN+RoBERTa^[19]是在 DialogueGCN 的基础上使用预训练语言模型 RoBERTa 的对话情感识别方法。DialogXL^[20]对话情感模型利于预训练语言模型 XLNet^[33]获得语义向量,并侧重于处理对话文本更长的历史上下文。DAG-ERC^[3]利用有向无环图建模对话文本中较远距离和较近距离的上下文信息流来学习对话中特有的语义结构,促进情感识别。

4.4 实验结果比较

表 2 GCC-LMI 及各个对比模型的性能比较

模型	IEMOCAP	MELD	EmoryNLP
ChatGPT Method	44.97	57.30	37.47
DialogueRNN	62.75	57.03	—
DialogueRNN+RoBERTa	64.76	63.61	37.44
KET	59.56	58.18	33.95
COSMIC	63.05	64.28	37.10
TODKAT	61.33	65.47	43.12
DialogueGCN	64.18	58.10	—
DialogueGCN+RoBERTa	64.91	63.02	38.10
RGAT	65.22	60.91	34.42
RGAT+RoBERTa	66.36	62.80	37.89
DialogXL	65.94	62.41	34.73
DAG-ERC	68.03	63.65	39.02
GCC-LMI	68.90	63.18	39.02

表2列出了文章提出的GCC-LMI模型与所有基准模型在3个公开数据集上的实验结果。需说明的是,对于经典的循环神经网络框架 DialogueRNN^[9]、图神经网络框架 DialogueGCN^[1]和 RGAT^[7],分别选择这三个模型在GCC-LMI模型所使用的RoBERTa预训练模型上的表现,以实现公平比较。

由表2可见,大模型在对话情感识别任务上表现不太突出,可能是由于对话文本特殊的语义交互和情感演变现象限制了大模型理解对话情感语义的效率。GCC-LMI在三个数据集上均表现出具有竞争力的性能,尤其在IEMOCAP数据集上表现出色。这可能是由于IEMOCAP数据集为双方对话数据集,对话人较少,文本语义理解相对容易。相反,数据集MELD和EmoryNLP均蕴含多于两人的对话人信息,语料中蕴含的语义交互和话题承转更普遍,文本语义理解难度较大。因此,GCC-LMI在这两个数据集上的表现相对来说不是很突出。

值得注意的是,模型TODKAT在数据集MELD和EmoryNLP上表现较为突出,这可能得益于该模型融入了有利于对话情感理解的常识知识和主题信息等辅助信息。由表2可见,在没有引入外部知识的情况下,文章的GCC-LMI模型仍可取得与TODKAT模型相当的实验结果。这验证了GCC-LMI在对话情感识别任务上的有效性。值得注意的是,在上述非知识融合的对话情感识别的基准模型中,文章提出的GCC-LMI模型在数据集IEMOCAP上表现最为突出,在EmoryNLP上的表现与当前最优秀的图网络对话识别框架DAG-ERC相当,在MELD上的表现略逊于DAG-ERC。这些表现充分说明文章模型在对话情感识别任务上具有较大优势。

4.5 消融实验

GCC-LMI模型的关键在于对话文本的各局部情感影响因素的叙述建模。为探讨与模型性能相关的各局部情感影响因素(局部情境信息、局部对话者信息、局部邻接情境信息)的作用,本节通过移除GCC-LMI上述各情感影响因素来进行消融实验,相关结果如图4所示。GCC-LMI(w/o F1)、GCC-LMI(w/o F2)、GCC-LMI(w/o F3)分别为GCC-LMI去除了影响对话情感的局部邻接情境信息、局部对话者信息、局部情境信息的叙述表征的演化模型。由图4可知,文章提出的GCC-LMI模型在去除上述三者之一以后,性能均有所退化。这表明,局部多因素叙述表征的联合建模能更全面地促进GCC-LMI模型的对话语义理解和情感预测。

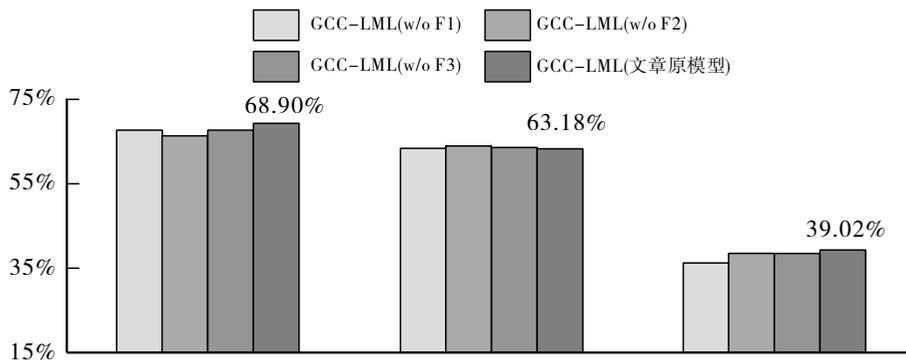


图4 GCC-LMI模型在各数据集上的消融实验结果

5 结论

文章基于对话文本的交互特性所导致的语义和情感动态演化,从影响对话情感的全局语义和局部多因素的角度来理解对话文本情感,进而提出了一种全局情境约束和局部多因素融合的对话情感识别模型——GCC-LMI。首先,通过全局情境建模,保证对话语义的动态演化下各语句的语义不偏离整体情境信息。在此基础上,GCC-LMI通过局部多因素叙述表征模块学习对话各语句情感的局部多因素理解下的最终情感状态表征,实现对话语句情感的高效预测。最后,在公开数据集上的实验结果显示了文章工作的有效性。

现有各社交媒体平台的对话数据往往蕴含多种模态的数据信息,而且,相较于单一的对话文本数据,这些关联模态信息与对应文本往往在语义上互为补充。因此,下一步的研究重点是高效理解和建模对话数据的多模态关联信息,并探索这些多模态信息与对话文本的跨模态统一表征方式。

参考文献:

- [1] 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 (EMNLP-IJCNLP), 2019: 154-164.
- [2] ZHANG D, WU L, SUN C, et al. Modeling both Context and Speaker-sensitive Dependence for Emotion Detection in Multi-speaker Conversations[C]//IJCAI, 2019: 5415-5421.
- [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 and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021: 1551-1560.
- [4] 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), 2019: 165-176.
- [5] GHOSAL D, MAJUMDER N, GELBUKH A, et al. COSMIC: ComonSense Knowledge for Emotion Identification in Conversations[C]//Findings of the Association for Computational Linguistics, ACL 2020; EMNLP 2020. Association for Computational Linguistics (ACL), 2020: 2470-2481.
- [6] TU G, LIANG B, LYU X, et al. Do Topic and Causal Consistency Affect Emotion Cognition? A Graph Interactive Network for Conversational Emotion Detection[C]//In The 26th European Conference on Artificial Intelligence (ECAI'23), 2023: 2362-2369.
- [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 (EMNLP), 2020: 7360-7370.
- [8] HSU C C, CHEN S Y, KUO C C, et al. EmotionLines: an Emotion Corpus of Multi-party Conversations[C]//Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), 2018: 1597-1601.
- [9] MAJUMDER N, PORIA S, HAZARIKA D, et al. DialogueRNN: an Attentive Rnn for Emotion Detection in Conversations[C]//Proceedings of the Thirty-third AAAI Conference on Artificial Intelligence and Thirty-first Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, 2019: 6818-6825.
- [10] LI W, SHAO W, JI S, et al. BiERU: Bidirectional Emotional Recurrent Unit for Conversational Sentiment Analysis[J]. Neurocomputing, 2022, 467: 73-82.
- [11] HAZARIKA D, PORIA S, ZADEH A, et al. Conversational Memory Network for Emotion Recognition in Dyadic Dialogue Videos[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 2018: 21-22.
- [12] HAZARIKA D, PORIA S, MIHALCEA R, et al. Icon: Interactive Conversational Memory Network for Multimodal Emotion Detection[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 2594-2604.
- [13] SUKHBAATAR S, WESTON J, FERGUS R. End-to-end Memory Networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 2440-2448.
- [14] YEH S L, LIN Y S, LEE C C. An Interaction-aware Attention Network for Speech Emotion Recognition in Spoken Dialogs[C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).IEEE, 2019: 6685-6689.
- [15] LI J, LIN Z, FU P, et al. Past, Present, and Future: Conversational Emotion Recognition Through Structural Modeling of Psychological Knowledge[C]//Findings of the Association for Computational Linguistics: EMNLP 2021. 2021: 1204-1214.
- [16] FU Y, OKADA S, WANG L, et al. CONSK-GCN: Conversational Semantic and Knowledge-oriented Graph Convolutional Network for Multimodal Emotion Recognition[C]//2021 IEEE International Conference on Multimedia and Expo (ICME), IEEE, 2021: 1-6.
- [17] XIE Y, YANG K, SUN C J, et al. Knowledge-interactive Network with Sentiment Polarity Intensity-aware Multi-task Learning for Emotion Recognition in Conversations[C]//Findings of the Association for Computational Linguistics: EMNLP 2021. 2021: 2879-2889.
- [18] KIPF T N, WELLING M. Semi-supervised Classification with Graph Convolutional Networks[J]. In proc International Conference on Learning Representations. arXiv: 1609.02907.2017.
- [19] THOST V, CHEN J. Directed Acyclic Graph Neural Networks[J]. arXiv:2101.07965, 2021.

- [20] 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, 2021, 35(15): 13789-13797.
- [21] ZHAO W, ZHAO Y, LU X. Cauain: Causal Aware Interaction Network for Emotion Recognition in Conversations[C]//Proceedings of the Thirty-first International Joint Conference on Artificial Intelligence, IJCAI, 2022: 4524-4530.
- [22] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach[J]. arXiv preprint arXiv, 2019: 1907.11692.
- [23] CHUNG J, GULCEHRE C, CHO K, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling[J]. arXiv Preprint arXiv: 3555, 2014.
- [24] 周云晓. 多轮对话文本连贯性研究[D]. 上海: 华东师范大学, 2019.
- [25] MEMORY L S T. Long Short-term Memory[J]. Neural Computation, 2010, 9(08): 1735-1780.
- [26] HU G, LU G, ZHAO Y. FSS-GCN: A Graph Convolutional Networks with Fusion of Semantic and Structure for Emotion Cause Analysis[J]. Knowledge-Based Systems, 2021, 212: 106584.
- [27] BUSSO C, BULUT M, LEE C C, et al. IEMOCAP: Interactive Emotional Dyadic Motion Capture Database[J]. Language Resources and Evaluation, 2008, 42: 335-359.
- [28] 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, 2019: 527-536.
- [29] ZAHIRI S M, CHOI J D. Emotion Detection on Tv Show Transcripts with Sequence-based Convolutional Neural Networks[C]//Workshops at the Thirty-second Aai Conference on Artificial Intelligence, 2018: 44-51.
- [30] MONTAVON G, ORR G, MLLER K R. Neural Networks: Tricks of the Trade[M]. 2nd ed. London: Springer, 2012.
- [31] ZHAO W, ZHAO Y, LU X, et al. Is ChatGPT Equipped with Emotional Dialogue Capabilities?[J]. arXiv Preprint arXiv: 2304.09582, 2023.
- [32] ZHU L, PERGOLA G, GUI L, et al. Topic-driven and Knowledge-aware Transformer for Dialogue Emotion Detection[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), 2021: 1571-1582.
- [33] YANG Z, DAI Z, YANG Y, et al. XLNet: Generalized Autoregressive Pretraining for Language Understanding[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019: 5753-5763.

Global Contextual Constraints and Local Multi-factor Integration for Conversational Emotion Recognition

CAO Wei¹, ZHAO Xin-yuan^{1*}, XUE Yu-yang²

(1.School of Computer Science and Technology, Xinjiang Normal University, Urumqi, Xinjiang, 830054, China;

2.School of Information Management, Xinjiang University of Finance & Economics, Urumqi, Xinjiang, 830012, China)

Abstract: Conversational Emotion Recognition is a vital task in the field of Natural Language Processing (NLP), aiming to automatically identify the sentiment of each utterance in a conversation. However, most existing studies have focused on the semantic representation of utterances or modeling of speakers' relations, ignored the influencing factors of emotional dynamics evolution during conversational interaction. This paper presents a new approach to Conversational Emotion Recognition that incorporates global contextual constraints and local multi-factor integration. This model not only considers the global semantic information of utterances in the conversation, but also deeply explores and models multiple influencing factors of emotional semantics during the dialogue. Experimental results on public datasets demonstrate the superiority of this new method for Conversational Emotion Recognition.

Keywords: Dialogue sentiment recognition; Global context; Local multi-factor fusion