

文章编号: 2095-2163(2023)11-0058-06

中图分类号: TP391.1

文献标志码: A

# 基于对话者语句交互图神经网络的对话情感分析

杨璐娴, 何庆

(贵州大学 大数据与信息工程学院, 贵阳 550025)

**摘要:** 对话情感分析任务旨在通过理解人类在对话中表达情绪的方式, 结合对话内容及对话者信息, 识别对话中的每一个语句的情感分类。不同于其余文本情感分析任务, 对话情感分析需要建模对话中的顺序上下文语境。然而, 如何更直观有效地建模对话语境, 并且充分考虑对话参与者的情绪变化, 以提高对话情感分析任务的准确率等问题仍有待探索。因此, 本文提出一种基于对话者语句交互图神经网络的对话情感分析模型。首先, 通过微调 RoBERTa 预训练语言模型提取对话文本的语句特征和对话者信息特征; 其次, 使用 Bi-GRU 建模对话文本的序列上下文语境, 获得上下文语句特征; 最后, 融合上下文语句特征和对话者信息特征构建对话者语句交互图神经网络模型。在公开数据集 MELD 上的实验结果表明, 与其他基线模型相比, 本文所提模型取得了更好的实验性能。

**关键词:** 对话情感分析; 图神经网络; 双向门控循环单元

## Conversational sentiment analysis based on speaker utterance interaction graph neural network

YANG Luxian, HE Qing

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

**Abstract:** Conversational sentiment analysis aims to identify the sentiment classification of each utterance in a conversation by understanding the way human express emotions and combining the conversation and speaker information. Different from the other text sentiment analysis tasks, conversational sentiment analysis needs to model the sequential context. However, issues such as how to model the conversation context more intuitively and effectively, and fully consider the sentiment changes of participants to improve the accuracy of the sentiment analysis remain to be explored. Therefore, this paper proposes a conversational sentiment analysis model based on speaker utterance interaction graph neural networks (SUDG). Firstly, the utterance features and speaker information features are extracted by fine-tuning the RoBERTa pre-trained language model. Then, Bi-GRU is used to encode the sequential contextual to obtain the contextual utterance features. Finally, the fusion of the contextual utterance features and the speaker information features constructs the speaker utterance interaction graph neural network model. Experimental results on the public dataset MELD show that the proposed model achieves better experimental performance compared with other baseline models.

**Key words:** conversational sentiment analysis; graph neural network; Bidirectional Gated Recurrent Unit

## 0 引言

随着大数据时代的到来, 以互联网为基础的智能化设备逐渐进入人类日常生活中的各个领域, 用户对于人机交互的需求日益增多, 解决人机交互困境成为研究热点。同时, 互联网的发展极大地提高了人类通信交流的效率, 降低了沟通成本, 丰富了互联网用户通信方式和沟通渠道。人类社交方式逐渐

多元化, 社交平台的迅速崛起为人类提供便利, 也为人工智能技术的发展提供了海量的研究数据, 对话场景下的情感分析研究应运而生<sup>[1]</sup>。

对话情感分析作为自然语言处理的新研究方向<sup>[2]</sup>, 由于其广泛的应用场景而受到研究人员的广泛关注, 该任务旨在通过理解人类在对话中表达情绪的方式, 结合对话内容及对话者信息, 识别对话中的每一个语句的情感分类<sup>[3]</sup>, 对提升人机交互系统的用户体验具有重要的研究意义, 也是人工智能领

**基金项目:** 国家自然科学基金(62166006)。

**作者简介:** 杨璐娴(1999-), 女, 硕士研究生, 主要研究方向: 自然语言处理。

**通讯作者:** 何庆(1982-), 男, 博士, 教授, 博士生导师, 主要研究方向: 深度学习、智能优化算法。Email: qhe@gzu.edu.cn

**收稿日期:** 2023-03-25

域的一个重要研究方向。对话情感分析不仅在各种社交媒体上具有极高的应用价值,在舆情分析<sup>[4]</sup>、虚假信息检测<sup>[5]</sup>、对话系统<sup>[6]</sup>和智能客服<sup>[7]</sup>等领域也得到了广泛的应用。对话是人类最基本的交流方式,具有明显的不连贯性和弱逻辑性,区别于篇章级的文本情感分析,对话文本的信息往往语言不连续且存在较大的跳跃性。对话情感分析与其余的文本情感分析任务本质上的区别在于信息的交互性。对话信息的获取依赖于对话语句发生的时序性、对话中的历史信息,对于多方多轮次的对话文本,对话信息还依赖于对话者信息及其对应语句发生的轮次等。因此,在对话的交互过程中,需要建模对话中的语句上下文语境,改变了情感分析模型的基本建模框架,给情感分析任务带来了极大的挑战<sup>[8]</sup>。

对话场景下的语句情感判定具有复杂性。对话示例如图 1 所示,该对话是在 2019 年公开的数据集 MELD<sup>[9]</sup>中选取的一个 3 方 6 轮对话文本。对话者 Chandler 的第二句话“真孩子气”在一般情况下会被判定为“愤怒”,但是在该场景中这句话实际上表达了对话者“悲伤”的情绪。由此可见,在不同对话场景下语句所表达的情感是复杂且多样的,与对话内容的历史信息、上下文语境、对话者情感状态及情绪变化等信息有极高的关联度。因此,如何准确地捕捉对话过程中的上下文信息、语句情感信息以及与之相关联的对话者信息是本文的研究重点。

在目前的对话情感分析任务模型中,主流研究模型几乎都需要对对话的上下文依赖信息进行建模,之前的工作已经提出了许多模型用于对上下文依赖信息建模,包括记忆网络<sup>[10]</sup>、递归神经网络<sup>[11]</sup>等模型,这些实验都取得了令人满意的效果,但是由于受到网络存储容量限制,导致远距离上下文信息丢失,难以有效传入当前对话中。此外,人类情绪具有惰性和干扰性两种特性。其中,情绪惰性是指在没有外界干扰的情况下,对话者自身的情绪会长期保持在同一个状态下;情绪的干扰性是指对话者保持在一个情绪状态时,由于其他对话人的话语刺激,激发了对话者的情绪变化。因此,还需要充分考虑对话者信息以便建模对话者之间的情绪交互。

针对上述问题,本文提出了一种基于对话者语句交互图神经网络的对话情感分析模型。

(1) 通过微调 RoBERTa 预训练语言模型提取语句特征,引入 Bi-GRU 建模对话文本序列上下文特征。

(2) 通过图神经网络建模对话上下文,并引入

不同类型的有向边和对话者节点建模对话者及其语句交互信息。

(3) 在 MELD 公开数据集上进行大量实验,验证了本文方法的有效性。

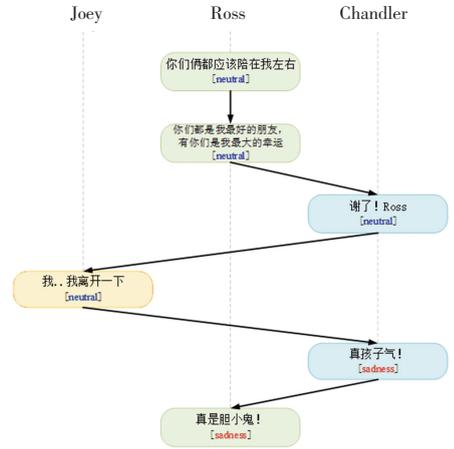


图 1 对话示例

Fig. 1 Examples of conversation

## 1 相关工作

对话场景下的文本情感分析作为一个具体问题场景,其研究历史较之其他文本情感分析问题的研究较短。目前,对话情感分析的方法主要采用基于深度学习和基于图神经网络的方法。对话情感分析任务在前期侧重于先建模对话中语句的上下文依赖关系,Poria 等学者<sup>[12]</sup>提出了一种基于 LSTM 的上下文信息提取模型,用于视频模态的情感识别任务。Zahiri 等学者<sup>[13]</sup>提出一种用于对话文本的序列的卷积神经网络模型。Hazariika 等学者<sup>[10]</sup>引入对话记忆网络,将每一位对话者的历史语句信息融入记忆网络,建模对话者之间的交互关系。Majumder 等学者<sup>[14]</sup>利用 GRU 建模对话序列信息、对话者和倾听者的情绪交互状态,再引入注意力机制捕捉长距离历史信息。由于近几年图神经网络方法在建模结构化数据中取得的巨大进展,将图神经网络方法应用于对话场景下的情感分析任务也逐渐成为研究热点。Zhang 等学者<sup>[15]</sup>通过连接语句节点之间无向边和对话者语句间的无向边建模上下文依赖和对话者情感依赖信息,提出多方对话中的情感分析方法。Hu 等学者<sup>[16]</sup>利用图卷积神经网络融合对话者内部信息和多模态对话数据,并取得较好的效果。

综上,虽然基于深度学习方法在对于对话语言顺序建模上有较好的适应能力。但是,对于参与者较多、对话篇幅较长的多方多轮对话中,深度学习方法的网络结构变得复杂,难以保留远距离历史信息。

利用神经网络对多轮对话进行建模具有更强的可解释性,可以更好地解决这一问题。

## 2 基于对话者语句交互图神经网络的对话情感分析模型

一个完整的对话由多个语句组成,在对话情感分析任务中,给定的对话文本  $U = u_1, u_2, \dots, u_n$ , 其中  $u_i$  表示对话  $U$  中的第  $i$  句话,  $n$  表示语句个数。对于多方对话文本,每个语句所对应的说话人表示为  $p(u_i) \in P, P = p_1, p_2, \dots, p_m, m$  表示说话人个数。对话文本的情感分析任务目的在于预测对话中每个独立语句  $u_i$  的情绪标签。参见图 1 的简单对话示例,对话中的第 6 句话:“真是胆小鬼!”,仅关注句子本身很难识别说话人真实的情绪,对话情感分析依赖于对话发生的顺序、对话上下文语境以及对话者人信息等。本文模型框架如图 2 所示。

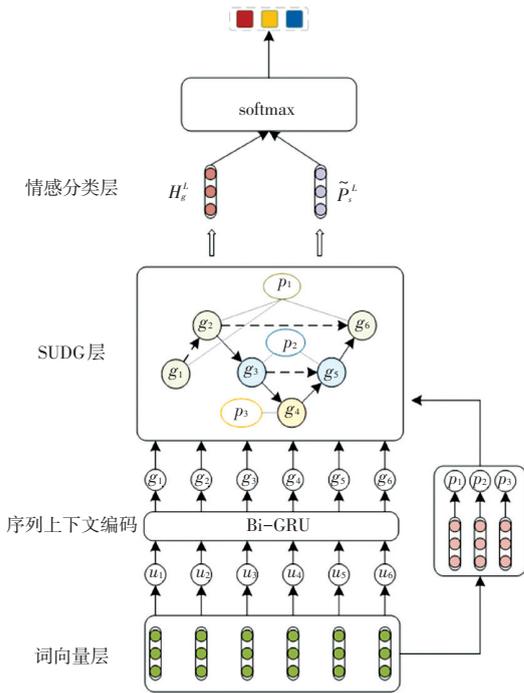


图2 模型框架

Fig. 2 Framework of the model

### 2.1 对话特征提取

对于对话语句的特征表示,本文通过微调 RoBERTa 预训练语言模型将对话文本序列转换为词向量形式。为便于训练和微调,本文将对话语句  $u_i$  定义为“[CLS],  $w_{i_1}, w_{i_2}, \dots, w_{i_{n_i}}$  [SEP]”作为模型中对话语句的输入。数学表达式如下所示:

$$u_i = \text{RoBERTa}(w_i) \quad (1)$$

对于对话者特征表示,本文将训练集中一个对

话者所说的所有语句特征的平均值设为该对话者的特征表示。在进行实验时,若某对话者只出现在测试集中,则将测试集中该对话者所说的所有语句特征平均值设置为该对话者特征表示。

### 2.2 序列上下文层级编码

对话的进行是连续且有序的,上下文信息会根据对话顺序流动,所以对话的语句顺序上下文信息是对话中的重要特征。GRU 是 RNN 的一种变体,相较于 RNN,保留了长期序列信息,并且减少了梯度消失的问题,但由于 GRU 无法编码从后往前的依赖信息,造成对话信息缺失,因此引入 Bi-GRU 来生成文本的序列上下文特征,其数学模型如下:

$$\vec{g}_i = \overrightarrow{\text{GRU}}(\vec{g}_{i-1}, u_i) \quad (2)$$

$$\overleftarrow{g}_i = \overleftarrow{\text{GRU}}(\overleftarrow{g}_{i+1}, u_i) \quad (3)$$

$$g_i = [\vec{g}_i, \overleftarrow{g}_i], i = 1, 2, \dots, n \quad (4)$$

其中,  $\vec{g}_i$  和  $\overleftarrow{g}_i$  分别表示前向和后向隐藏层的输出;  $g_i$  是序列上下文特征;  $u_i$  是上下文无关的特征。

### 2.3 对话者语句交互有向图模型 SUDG

#### 2.3.1 对话图构建

本文,将对话图定义为  $G = \{V_g, V_s, \mathcal{E}, R\}$ , 其中  $V_g = \{g_1, g_2, \dots, g_n\}$  表示对话中的语句节点,  $V_s = \{p_1, p_2, \dots, p_m\}$  表示对话者节点,  $\mathcal{E}$  表示连接这些节点和边的信息集合,  $R = \{r_1, r_2, r_3\}$  表示边的关系类型。考虑到多方对话中多名不同对话者之间的情绪交互影响,本文将边关系类型定为 3 种:

- (1) 同一对话者说出的不同语句节点间的边为  $r_1$ ;
- (2) 不同对话者说出的语句节点之间的边为  $r_2$ ;
- (3) 每个语句节点与其对应的对话者节点之间的边为  $r_3$ 。

由于对话的进行具有时序性,对话中已发生的语句可以对对话者的情绪产生影响,继而影响即将发生的语句情感,反之,未发生的语句却不能影响已发生的语句情感。因此,  $r_1$  和  $r_2$  是有向边,根据对话发生顺序模拟对话中的时间关系。  $r_3$  为无向边,用于传递整个对话中语句与其对应的对话者信息。

图 3 是根据图 1 构建的对话者语句交互图示例。对话中有 6 个语句  $\{g_1, g_2, \dots, g_6\}$ , 3 个对话者  $\{p_1, p_2, p_3\}$ , 其中  $g_1, g_2, g_6$  来自对话者  $p_1, g_3, g_5$  来自对话者  $p_2, g_4$  来自对话者  $p_3$ 。图 3 中的有向虚线,即  $r_1$ , 有向实线,即  $r_2$ , 无向实线,即  $r_3$ 。

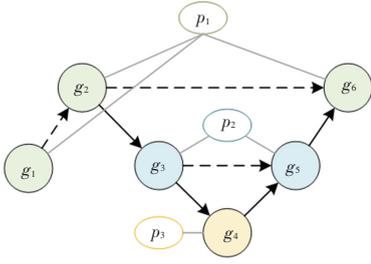


图 3 对话者语句交互图示例

Fig. 3 Sample of Speaker utterance interaction graph

### 2.3.2 SUDG 层

SUDG 模型是一个异构有向图神经网络,有 2 种不同的节点,因此需要 2 种不同的信息传播方法。对于语句节点  $V_g$  需要考虑对话序列上下文信息和对话者信息,而对于对话者节点  $V_s$  只需要考虑其对应的语句。因此邻接矩阵  $\mathbf{A}$  也需要分为不同模块:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_g \mathbf{A}_s^T \\ \mathbf{A}_s \mathbf{O} \end{bmatrix} \quad (5)$$

其中,  $\mathbf{A} \in R^{(n+m) \times (n+m)}$ ;  $\mathbf{A}_g$  表示语句之间的边;  $\mathbf{A}_s$  表示对话者节点与其语句之间的边;  $\mathbf{O}$  是 0 矩阵。同时,定义语句节点的特征矩阵为  $\mathbf{H}_g^l$ , 对话者节点的特征矩阵为  $\mathbf{P}_s^l$ 。

对话信息根据语句顺序传递,语句节点要聚合相邻节点信息和边关系类型信息,而 SUDG 构建的对话图是多关系图,语句节点间的边关系类型有  $r_1$  和  $r_2$  两种,因此本文引入基于不同关系的图注意力机制计算注意力权重。对于节点  $V_{g_i}$  通过节点在  $l$  层的隐层特征和边关系特征计算得到注意力权重,数学模型如下:

$$\alpha_{i,j}^l = \text{softmax}(\mathbf{W}_\alpha^l [\mathbf{H}_{g_i}^l \parallel \mathbf{H}_{g_j}^l \parallel \mathbf{H}_r^l]) \quad (6)$$

$j \in N_{(i)}, r \in \{r_1, r_2\}$

其中,  $\mathbf{W}_\alpha^l$  是可训练权重参数矩阵; “ $\parallel$ ” 表示拼接操作;  $\mathbf{H}_r^l$  表示边关系特征;  $N_{(i)}$  表示语句  $i$  节点的邻域。第  $l$  层中语句  $i$  基于注意力权重的邻域特征表示为:

$$\mathbf{M}_{g_i}^l = \sum_{j \in N_{(i)}} \alpha_{i,j}^l \mathbf{W}_{i,j}^l \mathbf{H}_j^l \quad (7)$$

语句节点在  $l+1$  层的最终表示如式(8)所示:

$$\tilde{\mathbf{H}}_g^{l+1} = \text{ReLU}(\mathbf{W}_g^l [\mathbf{H}_g^l \parallel \mathbf{H}_s^l \parallel \mathbf{M}_g^l] + b_g^l) \quad (8)$$

其中,  $\mathbf{H}_s^l$  是由  $\mathbf{P}_s^l$  生成的语句对应对话者特征矩阵:

$$\mathbf{H}_s^l = \mathbf{A}_s^T \mathbf{P}_s^l \quad (9)$$

给定对话者的所有语句的聚合特征可以通过邻接矩阵  $\mathbf{A}_s$  计算:

$$\mathbf{M}_s^l = \mathbf{A}_s \mathbf{H}_g^l \quad (10)$$

得到对话者及其语句聚合特征  $\mathbf{M}_s^l$  后,通过使用 GRU 更新对话者节点在  $l+1$  层的最终表示:

$$\mathbf{P}_s^{l+1} = \text{GRU}(\mathbf{P}_s^l, \mathbf{M}_s^l) \quad (11)$$

融合对话者信息的对话语句的特征表示如下:

$$\mathbf{H}_{sum}^{l+1} = [\tilde{\mathbf{H}}_g^{l+1}, \tilde{\mathbf{P}}_s^{l+1}] \quad (12)$$

### 2.3.3 情感分类

本文将所有图网络层的对话语句隐藏层状态拼接作为对话语句的最终表示  $\tilde{\mathbf{H}}_i$ , 通过一个全连接层和 *softmax* 层即可预测对话语句的情绪标签,数学模型如下:

$$\tilde{\mathbf{H}}_i = \parallel_{l=0}^L \mathbf{H}_{sum}^{l+1} \quad (13)$$

$$\mathbf{z}_i = \text{ReLU}(\mathbf{W}_H \tilde{\mathbf{H}}_i + b_H) \quad (14)$$

$$\mathbf{P}_i^e = \text{softmax}(\mathbf{W}_z \mathbf{z}_i + b_z) \quad (15)$$

$$\hat{y}_i = \underset{k \in S}{\text{Argmax}}(\mathbf{P}_{i,k}^e) \quad (16)$$

其中,  $S$  表示情绪标签合集。

本文使用标准交叉熵作为损失函数:

$$L(\theta) = - \sum_{i=1}^n \log P_i^e[y_i] \quad (17)$$

## 3 实验与结果分析

### 3.1 实验数据集

本文使用的是 2019 年公开的英文数据集 MELD<sup>[9]</sup>。MELD 来自于美剧《老友记》部分内容片段,包含 1 433 段对话,超过 13 000 句话,每句话都被标注了情绪标签,标签分为中性、愤怒、厌恶、高兴、恐惧、悲伤和惊讶 7 种情绪分类。MELD 是一个包含多名对话者的多模态数据集,本文只关注于文本模态数据,数据集具体内容见表 1。

表 1 数据集参数

Tab. 1 Parameters of datasets

数据集	MELD 数据集
对话数量	1 433
句子数量	13 708
训练集	9 989
验证集	1 039
测试集	2 610

在验证实验中,为了衡量模型对所有标签的分类精确度和有效性,本文使用宏平均  $F_1$  分数作为评价指标。

### 3.2 实验设置

本文实验在型号为 NVIDIA GTX 3090 的 GPU 服务器上进行,使用 Intel(R) Core(TM) i5-7500 处理器,内存大小为 8 GB,选用版本号为 Ubuntu 18.

04.3 LTS 的 Linux 操作系统,模型是基于 PyTorch 深度学习框架搭建。模型在训练的过程中采用 Adam 优化器来优化模型参数,通过验证集调整超参数,学习率为  $1 \times 10^{-5}$ ,丢失率为 0.3,批量大小为 32。

### 3.3 与基准模型比较

本文选用以下模型进行对比分析:

(1)KET<sup>[17]</sup>:使用一个外部知识网络增强语句信息,再结合上下文感知情感图注意力机制动态的增强上下文信息,以增强情感分类的性能。

(2)DialogueRNN<sup>[14]</sup>:使用 GRU 建模对话者状态和全局上下文信息,聚合对话者的情绪状态,建模对话者间的关系,可应用于多方对话场景。

(3)DialogueGCN<sup>[18]</sup>:使用 GCN 增强上下文信息,以句子为节点,对话图中基于固定上下文窗口连接不同关系节点,提出句子和对话者之间的有 8 种关系类型,同时考虑历史信息和未来信息。

(4)DAG-ERC<sup>[19]</sup>:通过有向无环图建模对话上下文,对话图以边关系类型建模对话者与句子关系,只关注历史信息。

(5)RGAT<sup>[20]</sup>:对不同对话者关系位置编码建模对话结构,使用图注意力网络更新节点,同时考虑历史信息和未来信息。

(6)ICON<sup>[21]</sup>:使用记忆网络储存对话者情感信息和全局状态信息,是一个多模态情感检测框架。

在 MELD 数据集上的情感分析实验结果见表 2。

表 2 不同模型结果对比

Tab. 2 Comparison with others model

编号	模型	$F_1 / \%$
1	ICON	54.65
2	DialogueRNN	57.03
3	DialogueGCN	58.10
4	KET	58.18
5	RGAT	60.91
6	DAG-ERC	63.65
7	SUDG	<b>64.76</b>

根据表 2 可知,在 MELD 数据集上,SUDG 与 ICON 模型相比  $F_1$  值提高了 10.11%,提升效果最佳,说明 SUDG 更适用于文本特征任务。与使用 RNN 和 GCN 的模型相比  $F_1$  值分别提高了 7.73%和 6.66%,说明在多方多轮对话语境中,过于复杂的模型反而会因为计算量庞大,导致上下文信息部分丢失,影响分类效果。与 KET 模型和 RGAT 模型相比,  $F_1$  值提高了 6.58%和 3.85%,说明基于图网络

的方法能更好地建模对话上下文信息,高效的传递对话历史信息。与 DAG-ERC 模型相比,  $F_1$  值提高了 1.11%,说明加入语句对话者信息能丰富对话中的上下文语境信息,同时证明了本文在图网络中加入语句对话者节点构图策略的有效性。

### 3.4 消融实验

为了验证各个组件对模型的影响,本文进行了消融实验,具体内容如下:

(1)SUDG/R:移除序列上下文编码部分和对话者交互有向图模型。

(2)SUDG/C:仅移除序列上下文编码部分,其他层保持不变。

(3)SUDG/H:仅移除对话者交互有向图模型,其他层保持不变。

(4)SUDG/S:移除对话者交互有向图模型中所有的对话者节点,引入一个随机初始化的全局节点代替原始对话者节点的位置,同时全局节点连接所有语句节点,用以验证融合对话者与对应语句关系信息的有效性,其他层保持不变。

消融实验结果对比见表 3。根据表 3 可知,相比于只使用了 RoBERTa 语言模型的 SUDG /R 而言,SUDG /H 和 SUDG /C 的  $F_1$  值分别提高了 5.89%和 11.55%,说明在多轮对话中语句情感对上下文信息依赖性强,将对话上下文信息进行顺序建模可以大幅度提高模型情感分类效果,并且基于图神经网络的方法能更有效地传递对话历史信息。相比于将对话者节点全部替换为全局节点的 SUDG/S 模型,SUDG 的  $F_1$  值提高了 1.60%,说明对话者节点与其对应语句连接可以有效提升模型对语句信息的提取,降低引入冗余信息的概率,提高模型情感分类准确率。SUDG 整体性能表现最佳,表现出全方位性能的显著提升,展示不同模型间的协同互补。

表 3 消融实验结果对比

Tab. 3 Results of ablation experiments

编号	模型	$F_1 / \%$
1	SUDG/R	49.73
2	SUDG/H	55.62
3	SUDG/C	61.28
4	SUDG/S	63.16
5	SUDG	<b>64.76</b>

## 4 结束语

本文提出一种基于对话者交互图神经网络的对话情感分析模型(SUDG)。利用 RoBERTa 预训练

语言模型提取文本特征;针对多轮多方对话文本中语句情感对语句序列的高依赖性,结合 Bi-GRU 将对话语句进行序列上下文编码;通过图结构建模对话上下文依赖关系,引入对话者信息节点,连接对话者节点及其对应语句,以提高对话情感分析中的上下文信息理解能力,传递远距离历史信息,减少冗余信息的引入,提高模型情感分类准确率。实验结果表明了 SUDG 模型的有效性,在 MELD 数据集上与其他模型相比,该模型在对话情感分析中表现出良好的分类效果。本文关注了对话者信息对语句情感的增强作用,未来将探索如何引入外部知识库增强语句情感信息。

## 参考文献

- [1] 杨宇骋. 基于层次特征的文本对话情感分析研究[D]. 重庆:重庆邮电大学,2022.
- [2] 林小颖. 多人对话场景中的多模态情绪识别研究[D]. 广州:华南理工大学,2020.
- [3] LIAN Zheng, LIU Bin, Tao Jianhua. Smin; Semi-supervised multi-modal interaction network for conversational emotion recognition[J]. IEEE Transactions on Affective Computing, 2022.
- [4] SOLEYMANI M, GARCIA D, JOU B, et al. A survey of multimodal sentiment analysis[J]. Image and Vision Computing, 2017, 65: 3-14.
- [5] SHETGAONKAR P, RODRIGUES J T, ASWALE S, et al. Fake review detection using sentiment analysis and deep learning[C]// Proceedings of 2021 International Conference on Technological Advancements and Innovations (ICTAI). Tashkent, Uzbekistan; IEEE, 2021: 140-145.
- [6] FIRDAUS M, CHAUHAN H, EKBAL A, et al. EmoSen; Generating sentiment and emotion controlled responses in a multimodal dialogue system[J]. IEEE Transactions on Affective Computing, 2020, 13(3): 1555-1566.
- [7] LI B, DIMITRIADIS D, STOLCKE A. Acoustic and lexical sentiment analysis for customer service calls[C]//Proceedings of ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, UK; IEEE, 2019: 5876-5880.
- [8] 涂耿. 对话语境中的句子级情感分析方法研究[D]. 汕头:汕头大学,2022
- [9] PORIA S, HAZARIKA D, MAJUMDER N, et al. Meld: A multimodal multi-party dataset for emotion recognition in conversations[J]. arXiv preprint arXiv:1810.02508, 2018.
- [10] HAZARIKA D, PORIA S, ZADEH A, et al. Conversational memory network for emotion recognition in dyadic dialogue videos [C]//Proceedings of the Conference of Association for Computational Linguistics. North American Chapter; NIH Public Access, 2018, 2018: 2122.
- [11] XING Songlong, MAI Sijie, HU Haifeng. Adapted dynamic memory network for emotion recognition in conversation[J]. IEEE Transactions on Affective Computing, 2020, 13(3): 1426-1439.
- [12] PORIA S, CAMBRIA E, HAZARIKA D, et al. Context-dependent sentiment analysis in user-generated videos [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (volume 1; Long papers). Vancouver, Canada; Association for Computational Linguistics, 2017: 873-883.
- [13] ZAHIRI S M, CHOI J D. Emotion detection on tv show transcripts with sequence-based convolutional neural networks [J]. arXiv preprint arXiv:1708.04299, 2017.
- [14] CHATTERJEE A, NARAHARI K N, JOSHI M, et al. SemEval-2019 task 3: EmoContext contextual emotion detection in text [C]//Proceedings of the 13th International Workshop on Semantic Evaluation. Minneapolis, Minnesota, USA; Association for Computational Linguistics, 2019: 39-48.
- [14] MAJUMDER N, PORIA S, HAZARIKA D, et al. DialogueRNN: An attentive RNN for emotion detection in conversations [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 6818-6825.
- [15] ZHANG Dong, WU Liangqing, SUN Changlong, et al. Modeling both Context- and Speaker-Sensitive Dependence for Emotion Detection in Multi-speaker Conversations [C]// Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. IJCAI. Macao, China;dblp, 2019: 5415-5421.
- [16] HU Jingwen, LIU Yuchen, ZHAO Jinming, et al. Mmgn: Multimodal fusion via deep graph convolution network for emotion recognition in conversation[J]. arXiv preprint arXiv:2107.06779, 2021.
- [17] ZHONG Peixiang, WANG Di, MIAO Chunyan. Knowledge-enriched transformer for emotion detection in textual conversations [J]. arXiv preprint arXiv:1909.10681, 2019.
- [18] GHOSAL D, MAJUMDER N, PORIA S, et al. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation[J]. arXiv preprint arXiv:1908.11540, 2019.
- [19] SHEN Weizhou, WU Siyue, YANG Yunyi, et al. Directed acyclic graph network for conversational emotion recognition [J]. arXiv preprint arXiv:2105.12907, 2021.
- [20] 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). Association for Computational Linguistics, 2020: 7360-7370.
- [21] 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. Brussels, Belgium; Association for Computational Linguistics, 2018: 2594-2604.