基于关系图注意力网络和宽度学习的负面情绪识别方法

彭三城[1]，**陈广豪**[1]，**曹丽红**[1,*]，**曾嵘**[2]，**周咏梅**[3]，**李心广**[1]
1.语言工程与计算实验室，广东外语外贸大学，广东广州，510006
2.信息光电子科技学院，华南师范大学，广东广州，510006
3.信息科学与技术学院，广东外语外贸大学，广东广州，510006

摘要

对话文本负面情绪识别主要是从对话文本中识别出每个话语的负面情绪，近年来已成为了一个研究热点。然而，让机器在对话文本中识别负面情绪是一项具有挑战性的任务，因为人们在对话中的情感表达通常存在上下文关系。为了解决上述问题，本文提出一种基于关系图注意力网络(Rational Graph Attention Network, RGAT)和宽度学习(Broad Learning, BL)的对话文本负面情绪识别方法，即RGAT-BL。该方法采用预训练模型RoBERTa生成对话文本的初始向量；然后，采用Bi-LSTM对文本向量的局部特征和上下文语义特征进行提取，从而获取话语级别的特征；采用RGAT对说话者之间的长距离依赖关系进行提取，从而获取说话者级别的特征；采用BL对上述两种拼接后的特征进行处理，从而实现对负面情绪进行分类输出。通过在三种数据集上与基线模型进行对比实验，结果表明所提出的方法在三个数据集上的weighted-$F1$、macro-$F1$值都优于基线模型。

关键词： 对话文本；负面情绪；关系图注意力网络；宽度学习；预训练模型

Negative Emotion Recognition Method Based on Rational Graph Attention Network and Broad Learning

Sancheng Peng[1], **Guanghao Chen**[1], **Lihong Cao**[1], **Rong Zeng**[2], **Yongmei zhou**[3], **Xinguang Li**[1]
1.Laboratory of Language Engineering and Computing,
Guangdong University of Foreign Studies, China.
2.School of Information and OptoelectronicScience and Engineering,
South China Normal University, China.
3.School of Information Science and Technoogy,
Guangdong University of Foreign Studies, China.

Abstract

Negative emotion recognition in textual conversations aims to identify the negative emotion of each utterance from textual conversations, which has become a hot research topic in recent years. However, enabling machines to recognize negative emotions in textual conversations is a challenging task, because there are contexts for peoples'emotional expression in conversations. To address the problem, we propose a method for negative emotion recognition based on rational graph attention network (RGAT) and broad learning (BL), namely RGAT-BL. We use pre-training model Roberta

收稿日期：2022-08-01 **定稿日期**：2020-08-15
基金项目：本课题得到到国家自然科学基金资助项目（编号：61876205，61877013），教育部人文社科项目（19YJAZH128，20YJAZH118）。
作者简介：彭三城（1974—），博士，教授，主要研究领域为情绪计算、宽度学习。陈广豪（1998—），硕士研究生，主要研究方向为负面情绪识别；曹丽红（1985-），硕士，讲师，主要研究方向为情绪计算；曾嵘（1998—），硕士研究生，主要研究方向为情绪原因识别；周咏梅(1971—)，硕士，教授，主要研究领域为情感计算；李心广(1962—)，博士，教授，主要研究领域为语音识别、情感计算

to generate the initial vector for textual conversations. Then, we adopt Bi-LSTM to extract local features and context semantic features of textual vectors, so as to obtain utterance-level features. Thirdly, we employ RGAT to extract long-distance dependency among speakers, so as to obtain the speaker-level features. At last, we use BL to process the above two connected features, so as to conduct the classified output of negative emotions. Compared with baseline models on three datasets, the experimental results show that the weighted-$F1$ and macro-$F1$ values of the proposed method are better than the baseline model on the three datasets.

Keywords: Textual conservations , negative emotion , rational graph attention network , broad learning , pre-training model

1 绪论

对话文本是由多个说话者交替说话而产生的，其全局语义是由多个用户在对话的语境中共同构建的，话语之间以及说话者之间在情绪表达上具有很强的关联性，上述特点使得对话文本的情绪识别成为了自然语言处理的一个研究热点 (彭韬et al., 2021)。情绪是指人们对外界刺激所作出的反应，而负面情绪是指人们对负面事件所作出的主观消极情绪反应 (赖河蒗et al., 2022a)。如何从这些海量的对话文本中自动地识别出携带负面情绪的信息，对于社交网络 (X et al., 2020; Peng et al., 2019)安全具有重要的意义。

现有的方法大都是针对短文本的情绪分类 (Sancheng et al., 2021)；同时，由于对话文本中话语之间以及说话者之间存在着一定的依赖关系，与短文本的情绪识别任务相比，对话文本的情绪识别任务无疑更具有挑战性。现有的针对对话文本的情绪识别方法主要包括：基于对话序列的方法 (Hazarika et al., 2018; Jiao et al., 2019; Majumder et al., 2019; Poria et al., 2016; Jin et al., 2020)和基于图神经网络的方法 (Zhang et al., 2019; Ghosal et al., 2019; Zhong et al., 2019; Shen et al., 2021b)。基于对话序列的方法大都采用长短时记忆网络(Long Short-Term Memory, LSTM)、门控循环单元(Gated Recurrent Unit, GRU)等深度学习模型来对话语级别的特征进行提取。这些模型虽能捕获长距离特征和话语之间的关联性，但该类方法忽略了说话者在对话中的相互影响和作用。

基于图神经网络的方法大都是采用图卷积神经网络对对话文本进行建模，从而对说话者之间的影响与联系进行更好地刻画。图卷积网络能捕获文本的结构特征以及单词间非连续和长距离的依赖关系。但是，图卷积网络还存在以下不足 (郑诚et al., 2022b)：一是由于是把单词表示为图的节点，采用邻接矩阵来表示节点的邻域信息，没有考虑对话文本的顺序结构，从而导致不能捕获文本的上下文语义信息；二是对局部特征信息的提取存在不足。

由于Bi-LSTM (Schuster and Paliwal, 1997)具有能有效捕获局部特征和上下文语义特征等优点，关系图注意力网络(Rational Graph Attention Network, RGAT) (Schlichtkrull et al., 2018)具有能有效捕获对话文本的结构特征和单词间的长距离依赖关系等优点，宽度学习具有网络结构简单、训练时间短、泛化能力强等特点。

因此，为了解决上面所提到的问题，提出一种基于RGAT和宽度学习(Broad Learning, BL) (Chen and Liu, 2017) 的对话文本负面情绪识别方法，即RGAT-BL。主要的贡献如下：采用Bi-LSTM能有效捕获局部特征和上下文语义特征的优点来提取话语级别的特征；然后，采用RGAT能有效捕获对话文本的结构特征和单词间的长距离依赖关系等优点来提取说话者级别的特征；最后，将对话语级别和说话者级别的特征进行拼接后，采用BL对负面情绪分类输出。在IEMOCAP、MELD和EmoryNLP三个对话文本数据集上，与基线模型进行对比实验，结果表明所提出的模型在每个数据集上的性能都优于基线模型。

2 相关工作

2.1 对话文本情绪分类模型

现有对话文本的方法大都针对情绪进行分类，主要可分为两种：基于对话序列的模型和基于图神经网络的模型。

(1) 基于对话序列的模型

Hazarika等人 (Hazarika et al., 2018)提出ICON模型，它将GRU和Multihop Attention机制相结合，学习话语之间的相关性。Jiao等人 (Jiao et al., 2019)提出HiGRU模型，该模型采用层两级GRU，一层用于提取每个话语内的上下文关系，另外一层用于提取话语级别的特征。Majumder等人 (Majumder et al., 2019)提出DialogueRNN模型，该模型采用CNN提取对话文本的上下文特征，用RNN提取话语级别的特征。Poria等人 (Poria et al., 2016)提出基于卷积多核学习的分类器以及基于上下文的层次Bi-LSTM来对多模态情绪进行识别。Jin等人 (Jin et al., 2020)提出一种层次的多模态Transformer，采用局部感知注意力机制和说话者感知注意力机制来分别捕捉说话者的局部语境和情绪惯性。

(2) 基于图神经网络的模型

Zhang等人 (Zhang et al., 2019)构建一种基于GCN的情绪识别模型来捕捉对话文本中的话语和说话者之间的关联。Ghosal等 (Ghosal et al., 2019)提出DialogueGCN模型，该模型分别采用双向GRU和GCN以提取对话级别特征和说话者级别特征。Zhong等人 (Zhong et al., 2019)提出KET模型，即采用自注意力机制提取话语的上下文关系，采用图注意力机制提取全局话语的特征。Shen等人 (Shen et al., 2021b)提出DAG-ERC模型，该模型通过构建有向无环图来对对话文本进行建模和训练。

2.2 宽度学习

BL是陈俊龙教授等人在2018年提出来的，主要由输入层、特征节点、增强节点和输出层四部分组成。BL需要训练的参数较少，一般只包括输出层的权重，即每个分类标签的对应权重。它可以通过岭回归算法 (Hoerl and Kennard, 1970)来快速获取。因此，BL具有结构简单、参数量少、训练时间短等特点，在许多分类任务上均得到了应用，如图像分类 (Chu et al., 2021)、视觉识别 (Jin et al., 2021)、情绪分类 (Peng et al., 2021)等。因此，本文采用BL作为分类器，有助于在短时间内得出更准确的分类结果，提升模型性能。

3 对话负面情绪识别

3.1 问题定义

给定一轮对话$U=[u_1,u_2,\cdots u_N]$以及M个说话者$S=[s_1,s_2,\cdots s_M]$，其中，N表示一轮对话中的话语个数，即M个人在本轮对话中总共说了N句话，其中，$u_i\in\mathbb{R}^d$ 表示第i句话的特征向量，d表示向量的维度。对话文本负面情绪识别任务的核心就是通过给定的一轮对话U以及M个说话者，从而预测出该轮对话中每个话语对应的负面情绪类别(如悲伤、生气)。

3.2 模型框架

本文所提出的RGAT-BL的框架结构主要包括文本编码层、话语级别编码层、说话者级别编码层和情绪分类层四个部分。文本编码层采用RoBERTa将句子中的每个单词映射到连续的低维向量空间中。话语级别编码层采用Bi-LSTM来提取话语级别的特征。说话者级别编码层采用RGAT来提取说话者级别的特征。情绪分类层采用BL对上述两种特征进行拼接后，并通过softmax输出负面情绪的分类结果。RGAT-BL具体的结构如图1所示。

3.3 文本编码层

在对对话文本进行预处理后，使用RoBERTa对对话文本的特征进行抽取，生成词向量；然后，对RoBERTa进行微调，使得模型每层的参数满足以下关系：

$$q_n^i=q_{n-1}^i-\mu^m\times\nabla_{q^l}H(q) \tag{1}$$

其中，q^i表示模型的第i层参数，n表示时间步长，$\nabla_{q^l}H(q)$表示模型目标函数的梯度，μ^m表示第m层的学习率，且需满足以下关系：

$$\mu^{m-1}=\delta\times\mu^m \tag{2}$$

其中，δ表示学习率的衰减速率且$\delta\leq 1$，当$\delta\leq 1$时，学习率逐层衰减，当$\delta=1$时，学习率则不做衰减，即每层的学习率保持不变。

利用微调后的RoBERTa对每条语句进行预训练以生成词向量，取词符“[CLS]”对应的输出向量作为句向量，从而得到句向量矩阵$U=[u_1,u_2,\cdots u_N]\in\mathbb{R}^{N\times d}$，其中，$N$表示一轮对话的话语数量，$u_i\in\mathbb{R}^d$表示第$i$句话的句向量，$d$表示句向量的维度。

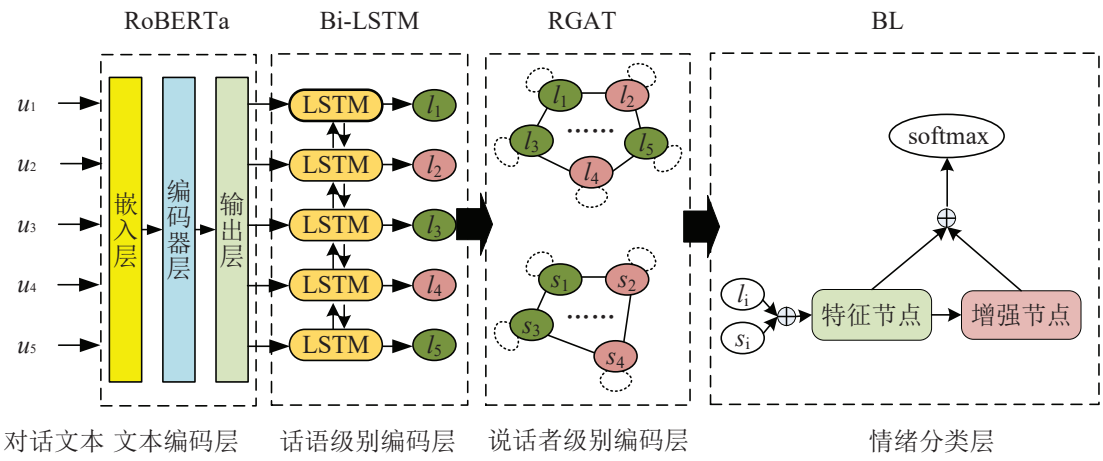

图 1. RGAT-BL的框架

3.4 话语级别编码层

由于对话文本是一个连续的话语序列，前后的话语之间往往有很强的关联性，即文本的上下文信息，这对于话语的情绪识别也至关重要。因此，采用Bi-LSTM能有效捕获上下文语义特征的优点来对话语级别的特征进行提取。Bi-LSTM的基本思想就是利用两个方向相反的LSTM来分别处理正向和反向序列，以获取特征之间在两个方向上的关联，从而将两个方向的关联信息输出作为LSTM的前、后向特征；然后，将前、后向特征进行拼接。使得Bi-LSTM能有效捕获文本的上下文信息。具体的过程表示如下：

$$\overrightarrow{l_i} = \overrightarrow{\text{LSTM}}(\overrightarrow{l_{i-1}}, u_i) \tag{3}$$

$$\overleftarrow{l_i} = \overleftarrow{\text{LSTM}}(\overleftarrow{l_{i+1}}, u_i) \tag{4}$$

$$l_i = [\overrightarrow{l_i}, \overleftarrow{l_i}], i = 1, 2, \cdots, N \tag{5}$$

其中，$\overrightarrow{l_i}$和$\overleftarrow{l_i}$分别表示第i句话的前向和后向LSTM特征。

因此，对于一轮话语，话语级别的特征可表示为：

$$L = [l_1, l_2, \cdots, l_N] \in \mathbb{R}^{N \times d_l} \tag{6}$$

其中，d_l表示话语级别特征的维度，在Bi-LSTM中，d_l表示其中一个LSTM层的神经单元数的两倍。

3.5 说话者级别编码层

由于RGAT能有效捕获对话文本的结构特征和单词间的长距离依赖关系等优点，因此，本文采用RGAT来对说话者级别进行编码。首先构造一个有向图用于保存对话者之间的对话以及情绪交互关系；然后，采用RGAT来获取对话文本的结构特征和单词间的长距离依赖关系。具体的构建过程如下。

假设有向图$G = \{V, E, R, W_G\}$用来表示一个具有N个话语的对话，其中，V表示节点集合，E表示边集合，R表示边的关系类型集合，W表示边的注意力权重集合。

V：每个话语对应G的一个节点$l_i \in V, i = 1, 2, \cdots N$，每个节点都是采用RoBERTa生成的面向话语的句向量$u_i \in \mathbb{R}^d$来表示。假设$p(\bullet)$表示话语到说话者之间的映射关系，即$p(l_i) \in S$表示节点l_i所对应的说话者$s_j, j = 1, 2, ..., M$。

E：如果节点l_i与l_j之间的边可表示为$e_{ij} \in E, i, j = 1, 2, \cdots N$，那么$e_{ij}$可用来表示话语之间的上下文关系。如果要表示$l_i$与其他话语的上下文关系时，那么$l_i$ 需要与G中的所有节点进行连接，即G成为了一个全连通图。

R：不同的边表示话语之间可能具有不同的上下文关系，共包括五种关系，可表示为$R=\{1,2,3,4,5\}$。当$p(l_i)=p(l_j),i<j$时，$R=1$；当$p(l_i)=p(l_j),i>j$时，$R=2$；当$p(l_i)\neq p(l_j),i<j$时，$R=3$；当$p(l_i)\neq p(l_j),i>j$时，$R=4$；当$i=j$时，$R=5$。每条边表示每个说话者的该话语受到其他说话者的影响或自身的影响，从而有效对对话文本的结构特征和单词间的长距离依赖关系进行提取。

W_G：节点l_i对的l_j边e_{ij}的权重可表示为$w_{ij}\in W_G$，其数值的大小表示了话语l_i对与话语l_j的影响程度。

下面给出基于RGAT的说话者级别特征的提取过程。

先对边的注意力权重w_{ij}进行求解，采用一个单层前馈神经网络来计算l_j对l_i的注意力系数c_{ij}，具体表示如下：

$$c_{ij}={l_i}^{\mathrm{T}}W_e\left[l_i,l_j\right],j=i-p,\cdots,i+f \tag{7}$$

其中W_e表示权重矩阵。

为了更好地表示不同节点对l_i的影响程度，采用softmax函数将注意力系数进行归一化，即l_j对l_i的边的注意力权重w_{ij}可表示如下：

$$w_{ij}=\mathrm{softmax}(c_{ij})=\frac{\exp\left({l_i}^{\mathrm{T}}W_e\left[l_i,l_j\right]\right)}{\sum_{k=i-p}^{i+f}\exp\left({l_i}^{\mathrm{T}}W_e\left[l_i,l_k\right]\right)},j=i-p,\cdots,i+f \tag{8}$$

采用两层RGAT进行来对说话者进行编码。对于第1层RGAT，通过聚合邻居节点信息将节点l_i转化为说话人相关的特征向量h_i，具体过程如下：

$$h_i=\sigma(\sum_{r\in R}\sum_{j\in N_i^r}\frac{w_{ij}}{c_{i,r}}W_r^1l_j+w_{ii}W_0^1l_i),i=1,2,\cdots,N \tag{9}$$

其中σ表示非线性激活函数，W_r^1和W_0^1表示权重矩阵，N_i^r表示节点l_i在关系$r\in R$下的邻居节点集合，$c_{i,r}$表示归一化常量，通常取值为$|N_i^r|$，即节点l_i的邻居节点个数。

对于第2层RGAT，是在第1层的基础上采用相同的特征转化方法，将h_i转化为特征向量o_i，具体过程如下：

$$o_i=\sigma(\sum_{j\in N_i^r}W^2h_j+W_0^2h_i),i=1,2,\cdots,N \tag{10}$$

其中W^2和W_0^2表示权重矩阵，p表示历史话语的窗口大小，f 表示将来话语的窗口大小。

通过式(9)和式(10)使得模型能有效聚合各节点的邻居节点信息，从而获取说话者之间的长距离依赖关系，即说话者级别的特征可表示为：

$$O=[o_1,o_2,\cdots,o_N]\in\mathbb{R}^{N\times d_s} \tag{11}$$

其中，d_s表示说话者级别的特征维度，即RGAT的隐藏单元数。

例如：假设一个具有6个话语的对话l_1,l_2,l_3,l_4,l_5,l_6，其中l_1,l_3,l_4为说话者s_1所说，即$s_1=p(l_1)=p(l_3)=p(l_4)$；l_2,l_5,l_6为说话者s_2所说，即$s_2=p(l_2)=p(l_5)=p(l_6)$，设$p=2$，$f=2$，则每个说话者的话语之间的关系如表1所示。

3.6 情绪分类层

先对话语级别的特征和说话者级别的特征进行拼接，由式(6)、(11)可得：

$$G^*=[L|O]\in\mathbb{R}^{N\times(d_l+d_s)} \tag{12}$$

然后，采用BL设计负面情绪分类器以对G^*进行分类，并预测每个话语的情绪。负面情绪分类器的设计过程如下。将上面所提取的特征G^*进行线性映射以生成BL的多组特征节点，即对G^*进行线性映射操作，生成k组特征节点；第i组特征节点表示如下：

$$Z_i=\varphi(G^*W_{ei}+\beta_{ei})\in\mathbb{R}^{N\times q},i=1,...,k \tag{13}$$

表 1. 每个说话者的话语之间的关系

关系类型	e_{ij}	说话者	i与j关系
5	e_{11}, e_{33}, e_{44}	s_1, s_1	$i = j$
1	e_{13}, e_{34}	s_1, s_1	$i < j$
2	e_{31}, e_{43}	s_1, s_1	$i > j$
5	e_{22}, e_{55}, e_{66}	s_2, s_2	$i = j$
2	e_{56}	s_2, s_2	$i < j$
1	e_{65}	s_2, s_2	$i > j$
3	$e_{12}, e_{35}, e_{45}, e_{46}$	s_1, s_2	$i < j$
4	e_{32}, e_{42}	s_1, s_2	$i > j$
3	e_{23}, e_{24}	s_2, s_1	$i < j$
4	$e_{21}, e_{53}, e_{54}, e_{64}$	s_2, s_1	$i > j$

其中φ表示线性激活函数，W_{ei}表示随机生成的权重矩阵且$W_{ei} \in \mathbb{R}^{(d_l+d_s)\times q}$，$\beta_{ei}$表示随机生成的偏置矩阵且$\beta_{ei} \in \mathbb{R}^{N\times q}$，$q$表示每组特征节点的数量，$k$表示特征节点的组数。

因此，k组特征节点可表示为$Z^k = [Z_1, Z_2, \cdots Z_k] \in \mathbb{R}^{N\times kq}$。将$Z^k$进行非线性映射以生成BL的多组增强节点，即对$Z^k$进行非线性映射操作，生成$m$组增强节点；第$j$组增强节点表示如下：

$$H_j = \xi(Z^k W_{hj} + \beta_{hj}) \in \mathbb{R}^{N\times r}, j = 1, ..., m \tag{14}$$

其中ξ表示非线性激活函数，W_{hj}表示随机生成的权重矩阵且$W_{hj} \in \mathbb{R}^{kq\times r}$，$\beta_{hj} \in \mathbb{R}^{N\times r}$表示随机生成的偏置矩阵且，$kq$表示所有特征节点的数量，$r$表示每组增强节点的数量。

因此，m组增强节点可表示为$H^m = [H_1, H_2, \cdots, H_m] \in \mathbb{R}^{N\times mr}$。将$k$组特征节点和$m$组增强节点进行拼接，可得$A = [Z^k | H^m] \in \mathbb{R}^{N\times(kq+mr)}$，再通过$A$来计算输出层的权重$W$。根据$Y = AW$，有$W = A^+Y$，由于$A$在大多数情况下都不是方阵，因此，可用$A^+$表示$A$的广义逆矩阵。为了更快速地计算$W$，同时增强模型的泛化能力，可采用岭回归算法[23]来求解W，具体表示如下：

$$\underset{W}{\operatorname{argmin}} \left(\left\| \left[Y - \hat{Y} \right] \right\|_2^2 + \lambda \left\| W \right\|_2^2 \right) \tag{15}$$

其中λ表示正则化系数，$\hat{Y} \in \mathbb{R}^{N\times(kq+mr)}$表示BL的近似输出。

W的全局最优解可表示为：

$$W = (\lambda I + A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} Y \tag{16}$$

4 实验及分析

将本文所提出的方法在三个对话文本的数据集上与基线模型进行对比实验。本次实验所采用的设备是一台搭载NVIDIA RTX 8000 48G显卡的Dell服务器。

4.1 评价指标

先采用weighted-$F1$值作为评价指标来比较各种方法在所有情绪标签上的性能；然后，采用macro-$F1$值作为评价指标来比较各种方法在负面情绪识别上的性能。weighted-$F1$、macro-$F1$具体的表示如下：

$$\text{weighted}-F1 = \frac{\sum_{i=1}^{C} (N_i \times F1_i)}{C} \tag{17}$$

$$\text{macro}-F1 = \frac{\sum_{i=1}^{C_{neg}} F1_i}{C_{neg}} \tag{18}$$

$$F1_i = \frac{2 \times P_i \times R_i}{P_i + R_i}, i = 1, 2, \cdots, C \tag{19}$$

其中，P_i和R_i分别表示第i类情绪的精确率和召回率，$F1_i$表示第i类情绪的$F1$值，N_i表示包含第i类情绪的样本数，C表示所有情绪的类别数，C_{neg}表示负面情绪的类别数。

4.2 数据集

为了验证所提出的模型有效性，在IEMOCAP (Busso et al., 2008)、MELD (Poria et al., 2019)和EmoryNLP (Zahiri and Choi, 2018)三个公开数据集上进行了实验。每个数据集的具体描述如下：

IEMOCAP：该数据集一个多模态数据集，包括文本、音频和视频，本文使用该数据集中的文本，该数据集包含六类情绪标签，分别是中性、开心、伤心、愤怒、沮丧和激动，其中负面情绪有伤心、愤怒和沮丧。

MELD：该数据集从美剧《老友记》中收集，包含七类情绪标签，分别是中性、开心、惊讶、恐惧、伤心、厌恶和愤怒，其中负面情绪有恐惧、伤心、厌恶和愤怒。

EmoryNLP：该数据集同样从美剧《老友记》中收集，包含七类情绪标签，分别是中性、开心、恐惧、愤怒、伤心、强烈和平静，其中负面情绪有恐惧、愤怒和伤心。

数据集的统计情况如表2所示。

表 2. 数据集的统计情况

数据集	对话数(训练/验证/测试)	话语数(训练/验证/测试)	类别数
IEMOCAP	100/20/31	4810/1000/1523	6
MELD	1038/114/280	9989/1109/2610	7
EmoryNLP	659/89/79	7551/954/984	7

4.3 基线模型

每种基线模型所采用的预训练模型都是RoBERTa，它们的具体描述如下：

SVM (Debnath et al., 2004)：SVM是一种传统机器学习模型，用于提取文本的特征。它使用的核是径向基函数，分类的策略是“一对多”；

TextCNN (Kim, 2014)：一种用于文本分类的CNN模型，用于提取文本的特征并对文本的负面情绪进行分类。卷积核的大小分别为3、4和5。核的维度为100；

Bi-LSTM (Schuster and Paliwal, 1997)：一种用于提取文本特征的双向LSTM。模型层数为2，每个层的隐藏单元数为32；

Bi-LSTM-ATTN (Zhou et al., 2016)：在Bi-LSTM模型的隐藏层的输出上增加了一个注意力层；

DialogueRNN (Majumder et al., 2019)：一种基于RNN的模型，包含三个GRU模块，其中两个GRU被用于记录说话者的状态和全局对话环境，另外一个GRU被用与对话过程中的情绪变化；

HiTrans (Li et al., 2020)：一种基于Transformer的对话情绪识别模型，利用Transformer提取序列特征的优良特性对序列上下文特征以及说话者情绪特征进行训练；

DialogXL (Shen et al., 2021a)：一种基于XLNet的对话情绪识别模型，采用XLNet对说话者自身以及说话者之间的依赖关系进行抽取，从而完成对话情绪识别任务；

DialogueGCN (Ghosal et al., 2019)：一种基于Bi-LSTM和GCN的对话情绪识别模型，采用Bi-LSTM对话语序列特征进行提取，用GCN对说话者级别特征进行提取。

4.4 参数设置

由于实验采用了三种数据集，在每种数据集上的实验参数稍微有所不同，具体情况如表3所示。

由于文本最大长度和LSTM隐层单元数对于话语级别特征的提取性能具有重要的影响，因此，在说话者级别上下文编码器中，包括说话者个数和话语窗口大小等参数。说话者个数可决定图中边的数量，其关系是$n_r = 2 \times 2^{n_s}$，n_s表示说话者个数，n_r表示边的数量。

4.5 实验性能对比

将本文所提出的RGAT-BL与基线模型进行对比，它们在三种数据集上的weighted-$F1$值对比结果如表4所示。

表 3. 参数设置

参数名称	IEMOCAP	MELD	EmoryNLP
文本最大长度	200	200	200
LSTM隐层单元数	150	150	150
说话者个数	2	12	10
话语窗口大小	10	4	6
RGAT隐层单元数	100	100	100
BL特征节点组数	10	10	10
BL特征节点个数/组	50	100	100
BL增强节点组数	10	10	10
BL特征节点个数/组	50	50	50
BL正则化参数	0.1	10	1
训练轮数	3	3	3

表 4. 不同模型在三种数据集上的weighted-$F1$值(%)

模型	IEMOCAP	MELD	EmoryNLP
SVM	36.12	48.75	24.32
TextCNN	47.65	54.14	32.46
Bi-LSTM	47.15	55.26	31.73
Bi-LSTM-ATTN	47.61	55.39	32.15
DialogueRNN	62.75	57.03	35.36
HiTrans	64.5	61.94	36.75
DialogXL	65.94	62.41	34.73
DialogueGCN	64.18	58.1	36.29
RGAT-BL	**66.13**	**64.83**	**37.94**

由表4可以看出，本文提出的RGAT-BL在数据集IEMOCAP、MELD和EmoryNLP上的性能均优于其它基线模型，weight-$F1$值分别为66.13%、64.83%和37.94%。其主要原因是RGAT-BL能有效地结合基于话语序列的模型和基于图神经网络的模型的优点，在话语之间的特征和说话者之间的特征提取上都发挥了重要作用，且基于BL的情绪分类器比传统的全连接层分类器能动态地计算情绪标签权重，从而获取更好的分类性能。

为了进一步验证模型对于对话负面情绪的识别性能，在三个数据集上对负面情绪的识别结果进行了对比实验，结果分别如表5、6、7所示。

由表5、6、7可知，本文提出的模型在三个数据集上针对负面情绪macro-$F1$值均优于其它基线模型。对于数据集IEMOCAP，“伤心”和“沮丧”的$F1$值相比其他基线模型更高，分别达到87.34%和67.20%；对于数据集MELD，“恐惧”、“伤心”和“厌恶”的$F1$值为最高，分别达到17.46%、36.71%和23.28%；对于数据集EmoryNLP，“愤怒”和“伤心”的$F1$值为最高，分别达到36.20%和26.06%。可见，该模型在负面情绪识别上同样表现良好。其主要原因是BL能将每个负面情绪类别通过特征节点和增强节点来得到合适的权重，从而确保了RGAT-BL在单个负面情绪类别上都能取得很高的$F1$值。

总之，与基线模型相比，本文所提出的模型在三个数据集上均有良好的情绪识别性能。采用Bi-LSTM与RGAT相结合的方式不仅可以提取话语的上下文特征，而且能充分考虑说话者之间的情绪交互和影响，从而能有效地对话语级别和说话者级别这两个级别的特征进行提取。在分类器方面，相比传统的全连接层分类器，基于BL的分类器具有三层网络架构以及岭回归优化的最小二乘法，也使得模型在分类性能上有所提升。

4.6 消融实验

为了进一步验证RGAT-BL的各部分的有效性，本文在三个数据集上分别进行了消融实

表 5. 不同模型在IEMOCAP数据集上负面情绪的识别性能

模型	$F1$值(%)			macro-$F1$ (%)
	伤心	愤怒	沮丧	
SVM	45.11	48.35	44.68	46.05
TextCNN	51.56	56.12	53.23	53.64
Bi-LSTM	52.98	55.45	57.61	55.35
Bi-LSTM-ATTN	53.5	57.19	58.74	56.48
DialogueRNN	78.8	65.28	58.91	67.66
HiTrans	80.23	**66.49**	60.16	68.96
DialogXL	77.10	61.59	64.67	67.79
DialogueGCN	84.54	64.19	66.99	71.91
RGAT-BL	**87.34**	62.54	**67.20**	**72.36**

表 6. 不同模型在MELD数据集上负面情绪的识别性能

模型	$F1$值(%)				macro-$F1$ (%)
	恐惧	伤心	厌恶	愤怒	
SVM	9.64	22.77	3.01	21.63	14.26
TextCNN	14.13	26.06	15.70	45.95	25.46
Bi-LSTM	13.56	24.15	16.06	43.44	27.88
Bi-LSTM-ATTN	13.81	25.35	16.19	44.01	24.84
DialogueRNN	9.62	34.08	13.16	41.87	24.68
HiTrans	16.84	35.03	20.45	**53.46**	31.45
DialogXL	10.32	33.16	9.33	49.93	25.69
DialogueGCN	9.71	34.76	19.35	42.71	26.63
RGAT-BL	**17.46**	**36.71**	**23.28**	52.13	**32.40**

表 7. 不同模型在EmoryNLP数据集上负面情绪的识别性能

模型	$F1$值(%)			macro-$F1$ (%)
	恐惧	愤怒	伤心	
SVM	22.16	16.01	7.37	15.18
TextCNN	29.41	25.13	11.53	22.02
Bi-LSTM	27.60	27.81	10.14	21.85
Bi-LSTM-ATTN	28.85	28.02	12.19	23.02
DialogueRNN	28.89	26.73	22.06	25.89
HiTrans	27.15	22.89	21.43	23.82
DialogXL	**37.38**	35.81	21.90	31.70
DialogueGCN	34.88	30.33	25.95	30.39
RGAT-BL	33.67	**36.20**	**26.06**	**31.98**

表 8. RGAT-BL在三个数据集上的消融实验

模型	macro-F1 (%)		
	IEMOCAP	MELD	EmoryNLP
RGAT-BL	**72.36**	**32.40**	**31.98**
BL	54.73	26.29	23.67
RGAT	71.62	31.76	31.35
GCN-BL	70.55	29.67	30.15

验：BL、RGAT 和GCN-BL，其中GCN-BL将RGAT-BL中的RGAT替换为GCN。具体的实验结果如表8所示。

由表8可知，RGAT-BL在三个数据集上的macro-F1均比BL、RGAT和GCN-BL高。其中，RGAT-BL比BL分别高出17.90%、6.11%和8.31%，这表明RGAT能通过图形注意力网络有效地对话语之间的关系进行建模，使得RGAT-BL能更好地识别话语在上下文语境下的情绪；RGAT-BL比RGAT分别高出0.74%、0.64%和0.63%，这表明了BL能提高RGAT-BL的性能，主要原因是BL能通过特征节点和增强节点，进一步对话语级别的特征和说话者级别的特征提取深层语义信息；RGAT-BL比GCN-BL分别高出1.81%、2.73%和1.83%，这表明RGAT通过引入话语之间的注意力权重，比GCN能更好地刻画话语之间的影响程度。

5 结论

本文主要研究了RGAT，Bi-LSTM和BL对对话文本负面情绪识别的重要性。通过与现有的方法进行比较，发现RGAT和Bi-LSTM与BL的结合有利于完成对话文本负面情绪识别这一任务。该方法通过结合深度学习和宽度学习的优点，旨在提供一种更直观的方法来提取话语中的局部上下文信息（即话语级别），以及对话中的全局上下文信息（即说话者级别）。最后，在三个对话文本数据集上进行了大量的实验，结果表明，话语层面和说话者层面的语境都有利于负面情绪识别；同时在大多数测试数据集上，该方法在加权平均$F1$值上都优于基线模型。在未来的工作中，计划将所提出的方法和其他深度学习模型进行结合，以更有效地对对话文本中的负面情绪进行识别。

参考文献

彭韬, 杨亮, 桑钟屹, 唐雨, and 林鸿飞. 2021. 基于异构二部图的对话情感分析. 中文信息学报, 35(11):135–142.

赖河蒨, 李玲俐, 胡婉玲, and 颜学明. 2022a. 一种基于层次化r-gcn的会话情绪识别方法. 计算机工程, 48(01):85–92.

郑诚, 陈杰, and 董春阳. 2022b. 结合图卷积的深层神经网络用于文本分类. 计算机工程与应用, 58(7):206–212.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. Iemocap: Interactive emotional dyadic motion capture database. *Language resources and evaluation*, 42(4):335–359.

CL Philip Chen and Zhulin Liu. 2017. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE transactions on neural networks and learning systems*, 29(1):10–24.

Yonghe Chu, Hongfei Lin, Liang Yang, Shichang Sun, Yufeng Diao, Changrong Min, Xiaochao Fan, and Chen Shen. 2021. Hyperspectral image classification with discriminative manifold broad learning system. *Neurocomputing*, 442:236–248.

Rameswar Debnath, Nogayama Takahide, and Haruhisa Takahashi. 2004. A decision based one-against-one method for multi-class support vector machine. *Pattern Analysis and Applications*, 7(2):164–175.

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 154–164, Hong Kong, China.

Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018. ICON: Interactive conversational memory network for multimodal emotion detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2594–2604, Brussels, Belgium.

Arthur E Hoerl and Robert W Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Wenxiang Jiao, Haiqin Yang, Irwin King, and Michael R. Lyu. 2019. HiGRU: Hierarchical gated recurrent units for utterance-level emotion recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 397–406, Minneapolis, Minnesota.

Xiao Jin, Jianfei Yu, Zixiang Ding, Rui Xia, Xiangsheng Zhou, and Yaofeng Tu. 2020. Hierarchical multimodal transformer with localness and speaker aware attention for emotion recognition in conversations. In *Proceedings of the 9th Natural Language Processing and Chinese Computing International Conference*, pages 41–53.

Junwei Jin, Yanting Li, Tiejun Yang, Liang Zhao, Junwei Duan, and CL Philip Chen. 2021. Discriminative group-sparsity constrained broad learning system for visual recognition. *Information Sciences*, 576:800–818.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October.

Jingye Li, Donghong Ji, Fei Li, Meishan Zhang, and Yijiang Liu. 2020. Hitrans: A transformer-based context-and speaker-sensitive model for emotion detection in conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4190–4200.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6818–6825.

Sancheng Peng, Guojun Wang, Yongmei Zhou, Cong Wan, Cong Wang, Shui Yu, and Jianwei Niu. 2019. An immunization framework for social networks through big data based influence modeling. *IEEE transactions on dependable and secure computing*, 16(6):984–995.

Sancheng Peng, Rong Zeng, Hongzhan Liu, Guanghao Chen, Ruihuan Wu, Aimin Yang, and Shui Yu. 2021. Emotion classification of text based on bert and broad learning system. In *Proceeding of the Asia Pacific Web and Web-age Information Management Joint International Conference on Web and Big Data*, pages 382–396.

Soujanya Poria, Iti Chaturvedi, Erik Cambria, and Amir Hussain. 2016. Convolutional mkl based multimodal emotion recognition and sentiment analysis. In *Proceedings of 16th International Conference on Data Mining*, pages 439–448.

Soujanya Poria, Devamanyu Hazarika, Navonil Majumder, Gautam Naik, Erik Cambria, and Rada Mihalcea. 2019. Meld: A multimodal multi-party dataset for emotion recognition in conversations. In *Proceedings of the 57th Conference of the Association for Computational Linguistics*, pages 527–536.

Peng Sancheng, Lihong Cao, Yongmei Zhou, Zhouhao Ouyang, Aimin Yang, Xinguang Li, Weijia Jia, and Shui Yu. 2021. A survey on deep learning for textual emotion analysis in social networks. *Digital Communications and Networks*.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The Semantic Web-15th International Conference*, pages 593–607.

Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing*, 45(11):2673–2681.

Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. 2021a. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition. In *Proceedings of the 35th Association for the Advancement of Artificial Intelligence*, volume 35, pages 13789–13797.

Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. 2021b. Directed acyclic graph network for conversational emotion recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 1551–1560, Online.

Han X, Li B, and Wang Z. 2020. An attention-based neural framework for uncertainty identification on social media texts. *Tsinghua Science and Technology*, 251:117–126.

Sayyed M Zahiri and Jinho D Choi. 2018. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In *Proceedings of the 32nd Association for the Advancement of Artificial Intelligence*, pages 44–52.

Dong Zhang, Liangqing Wu, Changlong Sun, Shoushan Li, Qiaoming Zhu, and Guodong Zhou. 2019. Modeling both context-and speaker-sensitive dependence for emotion detection in multi-speaker conversations. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 5415–5421.

Peixiang Zhong, Di Wang, and Chunyan Miao. 2019. Knowledge-enriched transformer for emotion detection in textual conversations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 165–176, Hong Kong, China.

Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 207–212.

第二十一届中国计算语言学大会论文集，第485页-第496页，南昌，中国，2022年10月14日至16日。