文章编号： 1673-3193（2024）03-0286-10

融合主题模型的图神经网络对话情感识别

张甜甜，李　众，谷一宽，杨晓霞

（中北大学 软件学院，山西 太原 030051）

摘　要： 对话情感识别（ERC）旨在预测对话中语句的情感类别。目前，基于图神经网络的ERC方法主要采用固定的超参数来确定图中边的连接，缺乏针对不同数据进行自适应构边的策略，且忽略了语句间的主题关系。此外，在图神经网络的训练过程中，这些方法通常采用求和叠加的方式来聚合节点信息，限制了模型的非线性能力。为此，本文将主题模型与图神经网络相融合，提出了一种新的构边方法。首先利用主题模型获取对话中语句的主题分布，然后将具有相同主题的语句相互连接。同时，引入了SwiGLU门控单元，用于调控图神经网络中层与层之间的信息流动。在边的类型方面，考虑了人物信息的差异，以更好地捕捉情感变化的内因和外因。通过在4个公开数据集（IEMOCAP、MELD、EmoryNLP、DailyDialogue）上进行的广泛实验，与当前先进的ERC方法相比，本文的方法在前3个数据集上的$F1$分数分别提升了1.69%，0.27%和0.38%。此外，本文的自适应方法在长对话上的效果提升了2.11%，优于短对话的0.8%，同时，通过引入SwiGLU有效减缓了图神经网络中的过度平滑现象。综合结果表明，本文提出的融合主题模型进行自适应构边以及引入SwiGLU门控单元的图神经网络方法，能够有效提高对话情感识别的效果，增强模型的鲁棒性。

关键词： 对话情感识别；图神经网络；主题模型；门控单元；图结构

中图分类号： TP391.1　　**文献标识码：** A　　**doi:** 10.3969/j.issn.1673-3193.2024.03.005

引用格式： 张甜甜，李众，谷一宽，等. 融合主题模型的图神经网络对话情感识别[J]. 中北大学学报（自然科学版），2024，45（3）：286-295.

ZHANG Tiantian，LI Zhong，GU Yikuan，et al. Fusion of topic models for conversational emotion recognition in graph neural networks[J]. Journal of North University of China（Natural Science Edition），2024，45（3）：286-295.

Fusion of Topic Models for Conversational Emotion Recognition in Graph Neural Networks

ZHANG Tiantian，LI Zhong，GU Yikuan，YANG Xiaoxia

（School of Software，North University of China，Taiyuan 030051，China）

Abstract： Emotion recognition in conversations（ERC）aims to predict emotional categories of utterances within a conversation. Presently，graph neural network-based ERC methods predominantly employ fixed hyperparameters to determine the connections among graph edges，lacking adaptive strategies for edge construction tailored to diverse data and ignoring thematic relationships between statements. Furthermore，during the training process of graph neural networks，these methods often utilize a summation superposi-

收稿日期： 2023-11-15

基金项目： 山西省自然科学基金资助项目（20210302123019）

作者简介： 张甜甜（1998－），女，硕士生，主要从事对话情感识别、图神经网络的研究。

通信作者： 李　众（1974－），男，副教授，博士，主要从事图形图像处理、算法分析的研究。E-mail：lizhong@nuc.edu.cn。

tion approach to aggregate node information, limiting the model's non-linear capabilities. To address these limitations, this paper integrated topic modeling with graph neural networks and proposed a novel edge construction method. Firstly, a topic model was employed to extract the thematic distribution of statements within a conversation, followed by the connection of statements sharing same themes. Meanwhile, the SwiGLU gated unit was introduced to regulate the flow of information between layers in the graph neural network. Considering differences in character information, the edge types were carefully tailored to better capture intrinsic and extrinsic factors influencing emotional changes. Through extensive experiments conducted on four publicly available datasets (IEMOCAP, MELD, EmoryNLP, DailyDialogue), our approach demonstrates significant improvements over advanced ERC methods, achieving F1 score enhancements of 1.69%, 0.27%, and 0.38% on the first three datasets, respectively. Moreover, our adaptive method exhibits a 2.11% improvement on long conversations, surpassing the 0.8% gain on short conversations. The introduction of the SwiGLU unit effectively mitigates over-smoothing phenomena in the graph neural network. Consequently, the proposed approach, which combines adaptive graph construction with topic modeling and integrates SwiGLU gated units into graph neural networks, proves to be effective in enhancing dialogue emotion recognition, thereby reinforcing the model's robustness.

Key words: emotion recognition in conversation; graph neural networks; topic model; gated unit; graphical structure

0 引 言

随着各类对话应用和数据的蓬勃发展，对话过程中每个语句的情感分类成为自然语言处理研究的一个热门议题。对话情感识别（Emotion Recognition in Conversations，ERC）广泛应用于不同系统领域，涵盖社交媒体中的意见挖掘[1]、车载对话系统[2]以及移情对话系统[3]等多个领域。在对话数据中，多个语句组成有序文本，由不同的说话者表达，并且这些语句之间存在着复杂的上下文关系。因此，与传统的情感识别方法只关注单一目标语句不同，对话领域中的情感识别需要对目标语句的上下文语境进行高质量的建模。基本原理是首先通过编码对话语句获取初始文本特征向量，然后利用图神经网络进行语境建模和特征更新，最终识别出每个语句的情感标签。

尽管图神经网络的结构化建模能力提高了对话情感识别的效果，但目前大多数图神经网络方法在构建边缘时往往缺乏根据不同对话数据进行自适应的能力。这些方法将对话中的语句视为图中的节点，并通过在语句节点之间构建边缘的方式来表示两个语句的相互关联。然而，现有的对话情感识别领域中的图神经网络方法在构建边缘时通常预先设定一个固定的超参数作为本地窗口，并假设本地窗口内的语句之间存在关联性。这种构建边缘的方式忽视了对话数据的个体差异，例如对话主题分布和对话长度。

一般认为同一主题下的语句情感分布会更加接近。当对话中的主题转变较少时，大多数语句都在讨论相似的主题，因此目标语句可能与更多的上下文存在关联。反之，当对话主题频繁变化时，目标语句的相关上下文也会更少。数据集IEMOCAP中的一段对话及情感如下："我能从你的眼睛看出来你想（开心）""想去派对（开心）""我们太高兴了（开心）""你打算贷款还是（沮丧）""好吧这可不是开心的部分，我们不想谈这个（沮丧）"。考虑对句5构边，采用固定窗口的构边方法，它将会与句2、句3、句4相连接，尽管四者的情感标签并不一致。若采用主题的自适应方法，它将会与同属于贷款主题下的句4相连接，两者均属于沮丧的情感。因此，我们认为图神经网络在为不同的对话数据构建边缘时应采用一种自适应的构建方法，通过充分利用对话语句的主题信息来自适应地构建边缘，从而提高对语境建模的效果。

图神经网络的特征更新过程实际上涉及对目标节点及其邻居节点的特征向量的聚合。一般可以将这个聚合过程划分为两步：首先对所有邻居节点的特征进行聚合，然后将目标节点的特征与聚合后的邻居节点特征相融合。通常，这两步都采用直接相加的形式。然而，简单的求和操作是一种线性操作，无法充分捕捉节点之间复杂的关

系，从而限制了模型对数据高阶特征的提取能力。为了更有效地融合信息，我们在特征更新的过程中引入了SwiGLU高效门控单元，以精确控制图神经网络中每一层节点信息的流动。这一创新旨在保持信息高效融合的同时，提升模型对数据复杂关系的抽取能力。

综合上述两点，本文提出了一种创新性的基于图神经网络的对话情感识别方法，命名为TGCN-ERC。该方法借助LDA主题模型辅助图神经网络中边缘的构建，使得模型能够根据不同特征的对话数据实现自适应的边缘构建。同时，采用了高效的门控单元SwiGLU来优化节点信息的聚合过程，使得模型能够有选择性地聚合重要信息。

本文的主要贡献如下：1）针对固定超参数设置的限制，提出了一种基于语句主题的自适应边缘构建方法；2）为应对节点信息聚合方式的局限性，将SwiGLU引入到GNN中，以增强模型的非线性表达能力；3）通过在4个公开的ERC数据集上进行的实验证明，本文提出的方法能有效提升对话情感识别的分类性能。

1 相关工作

1.1 门控线性单元

门控线性单元(Gate Linear Unit，GLU)[4]是一种基于门控思想构建的神经网络单元。该结构最早由Google Brain的研究人员于2016年提出，应用于自然语言处理任务，尤其是机器翻译和文本生成任务。GLU单元的核心特征在于引入了门控机制，该机制有助于网络学习输入数据中的关键信息，进而提升模型的性能。在GLU中，输入数据首先经过线性变换(通常为卷积层)，随后通过门控信号进行调控。门控信号一般由sigmoid函数生成，其取值范围在0～1之间。接着，通过将输入数据与门控信号相乘，得到GLU单元的输出。相较于传统的全连接层或普通卷积层，GLU能更有效地捕捉输入数据的结构信息，因而提高了模型的泛化能力和性能。

随后的研究中涌现出许多GLU单元的变种，如GTU[5]、ReGLU[6]、SwiGLU[6]等。其中，SwiGLU相较于其他变体更加平滑，能够实现更快的收敛速度，同时其非单调性也能更好地捕捉输入和输出之间的复杂非线性关系。SWiGLU是Swish函数和GLU的结合体。目前，许多大型语言模型，如Llama、Chatglm等，都采用该门控线性单元来控制信息的传播。

目前，基于图神经网络(GNN)的方法在节点特征聚合方面通常采用直接求和的方法。然而，这种线性求和操作存在一个局限，即无法捕捉到节点之间的复杂非线性关系，从而限制了模型对数据高阶特征的提取能力。因此，为了提升模型的泛化能力，我们将节点信息聚合时的求和操作替换为门控线性单元，以更有效地捕捉节点之间的复杂关系，从而增强模型对数据特征的提取和表示能力。

1.2 对话情感识别

1.2.1 基于RNN的方法

在早期的ERC研究方法中，ICON[7]、CMN[8]和DialogueRNN[9]等模型纳入了对话中语句的连续性考量，将对话语句视为一个连贯的文本序列，并采用RNN作为主干模型来处理对话数据。RNN，即循环神经网络，借助带有自反馈的神经元，能够有效处理各种长度的时序数据。然而，RNN的循环结构存在一个局限性，即仅能粗略地利用语句前后位置的偏序关系，难以准确建模复杂的语句联系。

1.2.2 基于GNN的方法

随着图神经网络(GNN)在半监督学习[10]、实体分类、链接预测、大规模知识库建模等领域的崭露头角，DialogGCN[11]是最早将图神经网络引入对话情感识别领域的先驱性工作。该方法提出将对话数据视为一种图结构数据，其中对话中的语句被看作图中的节点，而语句之间的相关性则通过节点之间的边连接进行建模。具体而言，首先利用BERT[12]、RoBERTa[13]等编码结构的模型对语句进行编码，得到对话语句的初始特征向量，这些向量即为图节点的初始特征。其次，通过构造图的边缘来建立语句节点之间的连接关系。当两个语句节点存在相关性时，系统会在它们之间构建一条边缘，这种连接关系将被映射成图的邻接矩阵。图神经网络的训练过程通过边进行信息传播，从而不断更新节点特征，这进一步强调了图神经网络中边缘构建方法的重要性。

尽管图神经网络的引入显著提升了模型的情感识别效果，但目前的相关图神经网络方法在构建边缘时通常采用固定的超参数作为本地连接窗口，用以确定节点之间的连接关系，这忽略了对话数据的个体差异。DialogGCN在构建边缘时采用了窗口内

全连接的方式，这意味着对话中的任何语句都与其窗口内的所有其他语句相连。然而，这默认了在一定窗口范围内的语句之间都存在情感联系。由于窗口大小作为超参数被人为设定，模型在处理不同对话时将使用相同大小的窗口，这种固定的连接模式显然不能适应多样且复杂的对话关系。RGAT[14]在DialogGCN的基础上引入了基于关系位置的编码，以捕捉说话者之间的依赖关系和语句之间的顺序信息，依然延续了窗口内全连接的边缘构建方法。ConGCN[15]定义了一个异质图，同时包含语句节点和说话人节点。特别地，说话人节点的初始特征是根据说话人信息的 one-hot 编码学得的。尽管ConGCN 额外引入了说话人节点，但并没有对语句节点在窗口内全连接的构边方法进行改进。为了突破这一限制，DAG[16]在构图的过程中利用说话人信息，并根据说话人的异同区分本地窗口和远程窗口，以不同类型的边连接不同类型的窗口之间的语句。这样，DAG 中的本地窗口大小实际上取决于说话人的说话顺序，但很容易退化成窗口大小为 1 的情况。

由于对话数据通常呈现主题变化、复杂语境、长度不一等特点，上述基于超参数窗口的统一构边模式难以满足针对多样化对话数据的需求，主要有两个明显的缺陷。首先，窗口大小固定，无法根据不同对话进行灵活调整。其次，该方法默认窗口内的语句都存在情感上的强关联，而实际情况可能并非如此。

为了克服先前方法中存在的构边缺陷，本研究提出了一种全新的自适应构边方法。首先引入主题模型 LDA 来识别对话语句的主题分布，然后基于语句所属的主题进行边的构建，从而克服了固定窗口对语境建模的限制。考虑到在相同主题下的语句往往具有更加相似的情感分布，这些语句之间也会呈现更强的情感关联。从主题分布的角度出发进行边的构建有助于图神经网络更好地挖掘对话中的情感变化关系，从而提高最终的情感识别的准确性。

2　融合主题模型的图神经网络

本文提出了一种创新性的图神经网络模型，旨在通过融合主题模型实现对图神经网络中边的连接进行自适应确定，以优化对话上下文语境的建模。为了增强模型的非线性能力，引入了高效的门控单元 SwiGLU，将提出的模型命名为 TGCN-ERC，它由 3 个关键部分组成：

1）图节点初始化：利用 RoBERTa 模型对对话语句进行特征提取，将提取到的特征作为图神经网络中节点的初始特征向量。

2）图网络融合 LDA[17]构边：采用 LDA 主题模型，通过分析对话语句的主题分布，将属于相同主题的语句节点相连接，以构建更为准确的图结构。

3）图网络训练及情感识别：在节点信息传播的过程中，引入了 SwiGLU 以精确控制信息的融合，从而更新节点特征。最后，读取并拼接图神经网络中各层的节点特征向量，将其送入全连接层进行情感识别。

TGCN-ERC 的整体框架如图 1 所示，图中展示了上述 3 个关键部分的流程。通过本文方法，期望能够提高模型对复杂对话上下文的建模能力，从而增强情感识别的准确性与鲁棒性。

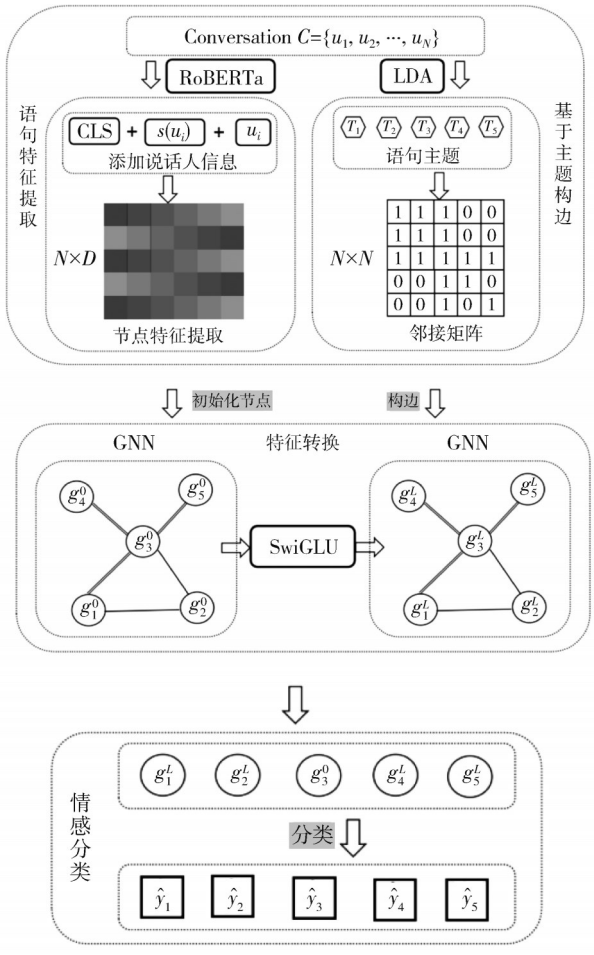

图 1　模型架构

Fig. 1　Model architecture

2.1　问题定义

对话情感识别的目标是预测出对话 $C=$

$\{u_1, u_1, \cdots, u_N\}$中包含的$N$个句子所分别对应的情感类别$e_i \in$｛生气、高兴、…、悲伤｝。其中每个句子$u_i = \{w_1, w_2, \cdots, w_{mi}\}$包含$m_i$个单词。对话中的任意一句话都存在相应的说话人$s(u_i) \in S = \{s_1, s_2, \cdots, s_p\}$。在对话场景中，任何一段对话都需要两个以上的说话人参与，即说话人集合至少包含2个元素。

2.2 图节点初始化

在对话情感识别领域，图神经网络将对话中的语句视为图中的节点。其中，对话语句的嵌入被视为图节点的初始特征向量，因此，对话语句嵌入的质量对整个模型的性能起着至关重要的作用。鉴于Transformer结构在文本表示方面的卓越能力，本文选择了一种被广泛采用的模型RoBERTa作为特征提取器。RoBERTa是一种基于Transformer编码器叠加而成的模型。对每个对话语句$u_i = w_1, w_2, \cdots, w_{mi}$进行处理，其中包括该语句的说话人$s(ui)$和句首标识[CLS]，将它们拼接在一起，得到输入向量$input_i$=[CLS],$s(u_i)$, $w_1, w_2, \cdots, w_{mi}$，随后将其输入RoBERTa模型。然后从模型的最后一层中提取出对应的[CLS]标记的嵌入，形成对话语句的嵌入e_i。因此，图节点的初始特征向量g_i^0即为对话语句的嵌入，即$g_i^0 = e_i$。对于一个包含N个对话语句的对话C，在经过RoBERTa的嵌入处理之后，可以得到对应的图的初始特征矩阵$\boldsymbol{X}$。该矩阵的维度为$N \times D$，其中D为语句嵌入的维度。这一步骤为后续的图神经网络模型提供了具体而有效的初始输入。

2.3 图网络融合LDA构边

获得图的初始特征矩阵后，下一步是构建图中节点的边缘以形成图的邻接矩阵$\boldsymbol{A}$。该邻接矩阵$\boldsymbol{A}$的维度为$N \times N$，在其中元素$a_{ij} \neq 0$时，表示节点i和节点j之间存在一条可连接的边缘。本文基于对话语句的主题分布提出了一种自适应的构边方法：首先通过LDA主题模型获得对话语句所属主题，然后基于主题进行边的连接。此方法克服了基于固定窗口内全连接的构边方法所存在的缺陷，并提高了模型的语境建模能力。

2.3.1 LDA提取语句主题

为了获取对话中语句的主题分布，需要利用概率主题模型对语句所属的主题进行分析。概率主题模型在有效识别大规模文档中潜在主题信息方面表现出色，通常被广泛应用于文本聚类。其中，隐性狄利克雷分布(Latent Dirichlet Allocation，LDA)是目前应用范围最广的概率主题模型之一，由David Blei于2003年提出，LDA构建了一个三层贝叶斯模型，涵盖了“文档-主题-词汇”的关系。该模型的基本思想是每篇文档$d \in D$都由若干个隐含的主题组成，用概率分布θ表示；每个主题k又由若干个单词组成，用概率分布φ_k表示；每个词汇的主题分配序列则用z_{dj}表示。参数φ、θ以及主题分配序列z可以通过吉布斯采样获得。具体地，LDA模型可以用图2所示的生成过程来描述。

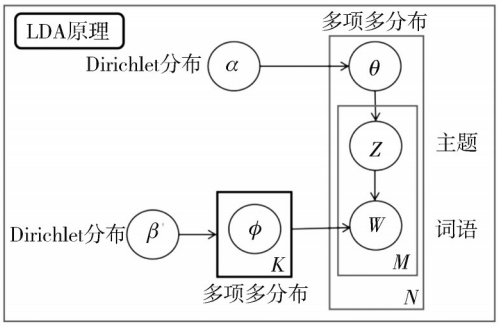

图2 LDA模型

Fig. 2 LDA model

本文采用LDA主题模型来提取对话中语句所属的主题。将包含N个语句的一段对话视为一个独立的文档，作为主题模型的输入文档d。通过LDA主题模型，挖掘出“文档-主题”的分布θ以及“主题-词汇”的分布φ。经过LDA处理后，可得到K个不同的主题，用T_i表示其中第i个主题。$S_{T_i} = \{word_1, word_2, \cdots, word_{10}\}$表示主题$T_i$下贡献度最大的前10个词汇的集合，$C_{T_i} = \{c_1, c_2, \cdots, c_{10}\}$表示这些词汇相对应的贡献度。句子$u_i$与主题$T_k$之间的相关性得分通过式(1)计算得出。

$$score(u_i, T_k) = \sum_{j=1}^{10} c_j,\ \text{if}\left(word_j \in \left(u_i \cap S_{T_i}\right)\right), \tag{1}$$

式中：$u_i \cap S_{T_i}$表示句子u_i与集合S_{T_i}中相同的所有词汇。最后，在K个不同主题中选取相关性得分最高的主题作为u_i所属的主题t_i。

$$t_i = \arg\max(u_i, T_k)。 \tag{2}$$

2.3.2 基于语句主题构边

在应用LDA主题模型获取对话中所有语句的所属主题后，通过将具有相同主题的句子相连

接来构建图的邻接矩阵$\boldsymbol{A}$。这里，句子u_i的所属主题用$Topic(u_i)$表示。对于所有相同主题下的句子$\{u_k|Topic(u_k)=Topic(u_i)\}$，将邻接矩阵$\boldsymbol{A}$中的元素$a_{ik}$设置为1，表示在$u_i$和$u_k$之间创建一条连接的边。这意味着具有相同主题的句子将在图中成为互为邻居的节点。另一方面，为了确保图的连通性，需要确保总存在一条边将u_i和u_{i+1}连接起来。最终，得到一个邻接矩阵$\boldsymbol{A}$，用于表示图中节点之间的连接关系。

2.3.3　边缘权重学习

邻接矩阵仅表示节点之间是否存在边，无法使模型在训练时有针对性地关注关键信息。因此，引入了注意力机制来计算节点之间的边缘权重，以使图神经网络能够根据邻居节点的不同重要程度来聚合节点信息。这里延续了Graph Attention Network（GAT）[18]的注意力计算公式。对于目标节点u_i和其任一邻居节点u_j，首先将它们的节点特征g_i与g_j进行拼接，形成一个$2\times F$的向量，然后通过一个全连接层得到一个标量e_{ij}，代表注意力因子。e_{ij}的计算公式为

$$e_{ij}=LeakyRELU(W[g_i\|g_j]), \tag{3}$$

式中：$\|$表示向量拼接操作，$LeakyRELU$表示泄漏整流线性函数。最终，对u_i的所有邻居节点的注意力因子进行softmax操作，以获得与邻居节点之间的边缘权重，其中α的计算公式为

$$a_{ij}=\frac{\exp(e_{ij})}{\sum_{k\in U_i}\exp(e_{ik})}, \tag{4}$$

式中：U_i表示u_i的邻居节点集合。

2.4　图网络训练及情感识别

基于上述的边缘构建方法确定图的结构后，采用图卷积来学习节点的上下文信息，并更新每个语句节点的特征向量。从空域的角度看，图神经网络的训练过程可以划分为3个关键步骤。首先，聚合目标节点u_i的邻居节点信息，得到h_i；接着，将h_i与u_i两者融合，进行节点特征的更新；最终，进行情感分类和反向传播。

2.4.1　邻居特征聚合

为了更全面地考虑说话人信息对模型效果的影响，在聚合邻居节点特征时，根据说话人信息的异同分类出不同的边类型。研究表明，在对话过程中，语句情感受到说话人自身和其他说话人的双重影响。对于语句节点u_i和u_j，若它们属于同一说话人，则边e_{ij}被归类为类型r_1，反之，属于类型r_2。前者反映了说话人自身的心理状态对其语句情感的影响，属于内在因素；而后者表示其他说话人对该说话人的影响，属于外在因素。在聚合邻居节点特征时，不同的边类型将学习不同的权重矩阵，从而区分内在和外在这两种情感变化模式。邻居节点特征的聚合公式为

$$\boldsymbol{h}_i^l=\sum_{u_k\in N_i}\alpha_{ik}\boldsymbol{W}_{r_{ik}}^l\boldsymbol{g}_k^{l-1}, \tag{5}$$

式中：$\boldsymbol{h}_i^l$是由u_i的所有邻居节点Ni聚合而成的特征向量；α_{ik}是u_i和u_k之间的边缘权重；$r_{ik}\in\{r_1,r_2\}$表示节点ui和节点u_k之间的边缘类型；$\boldsymbol{W}_{r_{ik}}^l\in\{W_{r_1}^l,W_{r_2}^l\}$是根据不同边缘类型所分配的训练参数矩阵；$\boldsymbol{g}_k^{l-1}$是节点$u_k$在模型第$l-1$层的特征向量。

2.4.2　节点特征更新

目标节点的特征更新是将邻居节点的聚合信息与自身信息进行融合。相关研究表明，门控单元能够有效地掌控信息融合过程，从而使模型中不同层之间实现更为有效的信息交互。由于图神经网络的层数过多可能导致训练过程中出现过度平滑的问题[19]，通常图神经网络的层数较少，但每一层都包含丰富的信息。本文选择使用SwiGLU作为门控单元，该单元包含额外的激活函数，更加平滑且更容易收敛。在模型中引入门控单元有助于增强模型的非线性变化能力，进而提升模型的泛化效果。目标节点的特征更新过程公式为

$$\boldsymbol{g}_i^l=SwiGLU(\boldsymbol{h}_i^l,\boldsymbol{g}_i^{l-1}), \tag{6}$$

$$SwiGLU(x,y)=x\cdot\tanh(x)\otimes y, \tag{7}$$

式中：$\otimes$为矩阵之间的元素向积；tanh为双曲正切函数；$\boldsymbol{g}_i^l$为第$l-1$层节点u_i更新后的特征向量。

2.4.3　分类情感标签

在经过多个图神经网络层的节点特征更新后，获得了语句节点最后一层的最终特征向量。为了进一步丰富语句的特征表示，在图神经网络完成特征转换后，将网络中每一层的特征向量拼接起来，从而得到语句节点u_i最终的特征向量$\tilde{\boldsymbol{g}}_i$。

$$\tilde{\boldsymbol{g}}_i=\|_{l=0}^{L}\boldsymbol{g}_i^l, \tag{8}$$

式中：L为图神经网络的层数。

将$\tilde{\boldsymbol{g}}_i$输入到两层的全连接网络当中，通过softmax函数得到句子对应的情感标签$\hat{y}_i$。

$$q_i=ReLU(W_0(\tilde{\boldsymbol{g}}_i)+b_0), \tag{9}$$

$$p_i=Softmax(W_1(q_i)+b_1), \tag{10}$$

$$\hat{y}_i=\arg\max_k(p_i[k]), \tag{11}$$

式中：$ReLU$为激活函数；W_0，W_1为全连接网络的参数；b_0，b_1为偏置项；$\hat{y}_i$为情感标签。在模型训练过程中，使用交叉熵损失CE Loss作为模型的损失函数。

3 实验结果与分析

3.1 评估数据集和基线

为验证本文提出的TGCN-ERC算法的有效性，进行了广泛的实验，使用了4个公开的对话情感识别领域的数据集：IEMOCAP[20]、MELD[21]、DailyDialog[22]、EmoryNLP[23]。

IEMOCAP是一个多模态的数据集，其中每个对话都源自两名演员基于脚本的表演。该数据集对样本进行了中性、快乐、悲伤、愤怒、沮丧、兴奋等6种情绪的评价。由于IEMOCAP中的对话数据较为有限，这里使用训练集中的最后20个对话进行验证。

MELD是一个多模态的情感识别与分类数据集，从电视剧《老友记》中收集而来。该数据集包含7种情感标签，包括中性、快乐、惊讶、悲伤、愤怒、厌恶和恐惧。

DailyDialog是从英语学习者的书面对话中收集的人工书写对话数据集，共有7种情感标签：中性、快乐、惊讶、悲伤、愤怒、厌恶和恐惧。由于该数据集缺乏说话人信息，这里将话语交替视为默认的说话人交替。

EmoryNLP的数据集包括来自《老友记》的电视剧本，与MELD相比，该数据集在场景和情感标签的选择上有所不同。情感标签包括中性、悲伤、疯狂、恐惧、强大、和平和快乐。这里仅使用该数据集的文本模态进行实验。

这些数据集中，MELD、DailyDialog和EmoryNLP的对话平均轮数为8～10轮，而IEMOCAP中对话的平均轮数为50轮，明显高于其他数据集。此外，每个数据集的情感标签也存在差异，DailyDialog中的中性情感标签占到82%，而MELD和EmoryNLP中的情感标签相对平衡。对这些数据集进行TGCN-ERC的训练，并评估了最终效果。本文将基线模型分为基于RNN的方法和基于GNN的方法两大类，并对它们的模型效果进行了比较。最终的实验结果是5次重复实验后的平均值。在数据集的划分上，按照表1所示的方式进行了训练集、验证集和测试集的划分。

表1 对话数据集划分

Tab. 1 Data set session partitioning 个

数据集	训练数量	验证数量	测试数量
IEMOCAP	120	20	31
MELD	1 000	120	312
EmoryNLP	11 000	1 000	1 118
DailyDialog	700	100	97

将本文模型与以下最先进的方法进行比较：

DialgueRNN[9]使用多个GRU作为主要结构。这些GRU用于跟踪说话人和听者的状态，捕获上下文信息和全局信息等。

COSMIC[24]使用常识变压器模型提取常识特征，并使用多个双向GRU执行情感识别任务。

DialgueGCN[11]是一种基于GNN的模型，它使用具有固定窗口的完全连接结构。在边缘类型上区分了自我依赖和说话人间依赖。

RGAT[14]提出了一种关系位置编码，可以反映关系类型的顺序信息。RGAT在固定窗口内使用完全连接的结构。

ERMC-DisGCN[25]提出了一种针对ERMC的话语感知图神经网络。设计了一个关系卷积来利用对话者的自我说话者依赖来传播上下文信息

SKAIG[26]提出了一个心理-知识感知交互图。在局部连接图中，目标话语会被过去语境推断的动作信息和未来语境暗示的意图信息所增强。

3.2 训练参数设置

对于基线模型，采用了其原始论文所描述的结果。在本文模型训练的过程中，RoBERTa编码的特征维数设置为1 024；在图神经网络中，隐藏层节点的特征维数为300；句子的截断长度限定为200；在IEMOCAP数据集上，批处理数量为16，在其余数据集上为64；Dropout设置为0.3；在优化器的选择上，采用带有动量的Adam优化器，其中动量设置为0.9。由于4个数据集的数据分布特点存在较大差异，因此，训练周期、学习率以及图神经网络堆叠的层数均通过验证集进行动态调整。在RTX 3090 GPU上进行模型的训练，Python版本为3.8，PyTorch版本为1.8.0。

3.3 实验结果及分析

在评估指标方面，对话情感识别领域的公开数据集在一定程度上存在情感类别不平衡的问题。因此，延续之前的工作，仍然选择$F1$分数作为评价指标。$F1$分数是对准确率和召回率两者的

调和分数，能够更好地反映模型的整体性能。$F1$分数的计算公式为

$$F1=\frac{2PR}{P+R},\tag{12}$$

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}},\tag{13}$$

$$R=\frac{N_{TP}}{N_{TP}+N_{FN}},\tag{14}$$

式中：N_{TP}表示模型正确预测为正类别的样本数量；FN_{FP}表示模型错误地将负类别的样本预测为正类别的数量；N_{FN}表示模型错误地将正类别的样本预测为负类别的数量。

3.3.1　对比实验

本文方法在4个数据集上的模型对比实验结果如表2所示。

表2　对比实验的$F1$分数结果

Tab. 2　Comparative experimental results of $F1$　%

方法	IEMOCAP	MELD	EmoryNLP	DailyDialog
DialogueRNN	62.75	57.03	—	—
COSMIC	63.05	64.28	37.10	56.16
DialogueGCN	64.18	58.10	34.29	49.95
RGAT	65.22	60.91	34.42	54.31
ERMC-DisGCN	—	64.22	36.38	—
SKAIG	66.96	65.18	38.88	59.75
TGCN-ERC	68.65	65.18	39.26	58.07

由表1的实验结果可以看出：1）本文模型TGCN-ERC在4个数据集上的结果均优于基于RNN结构的模型。这是因为图结构在对话语境建模方面具有天然的优势，有助于模拟对话关系和处理说话人的信息。2）相对于其他图神经网络结构的模型，本文模型在3个数据集上都取得了效果的提升。特别是在IEMOCAP上，除多模态融合的MM-DFN之外，本文模型实现了1.69%～4.47%的显著提升。这一观察结果表明，基于主题的自适应构边方法能够有效提升情感识别性能。3）在DailyDialog上，本文模型的效果略低于SKAIG。对这个数据集的对话特点进行观察，发现其中包含大量的日常对话语句，而融合了日常常识的SKAIG模型在识别日常对话方面具有更多优势。

3.3.2　消融实验及分析

为了深入探究所提出方法的有效性，本文进行了充分的消融实验，对TGCN-ERC的各个模型进行了分析，结果如表3所示。“w/o topic”表示不采用本文提出的基于主题模型的构边方法，而直接采用窗口内全连接的构边方式，窗口的大小根据RGAT中的经验被设置成了5跳。“w/o type”表示不利用说话人信息来区分边的类型。因此，在进行邻居节点特征的聚合时只学习一个共同的参数矩阵。“w/o SwiGLU”表示图网络中前一层的所有信息都直接相加传递到下一层，在更新节点特征时不经过门控单元的非线性变化。

表3　消融实验的$F1$分数结果

Tab. 3　Ablation results of $F1$　%

方法	IEMOCAP	MELD	EmoryNLP	DailyDialog
本文模型	68.65	64.35	39.26	58.07
w/o topic	66.95	63.2	38.55	57.45
w/o type	67.20	63.45	38.40	57.70
w/o SwiGLU	67.26	63.18	38.23	58.15

由表3的实验结果可以看出：1）采用全连接的图结构导致模型性能显著下降，这表明基于主题的构边方法能够自适应地建模对话中的上下文关系。模型性能大幅下降的观察结果进一步验证了对话语境建模对于情感识别的重要性。2）在不区分边类型的情况下，模型效果下降，从而证实了说话人信息对于模型性能的重要性。3）移除门控单元后，模型性能也有所降低，表明门控单元的引入可以提高模型的泛化能力，有助于融合不同层之间的特征信息。

3.3.3　对话长度的影响

多项研究表明，对于长对话，更适当的上下文多项研究均强调，在处理长对话时，上下文建模的重要性更为突出，而良好的构边方法在这类对话中的效果更为显著[27]。为了验证这一点，在DailyDialog数据集上对本研究基于主题构边的方法和传统的窗口内全连接构边方法的性能进行了详细比较。根据对话的长度将数据分为长对话（对话轮数＞15）、中对话（对话轮数＞5）和短对话（对话轮数＜5）。实验结果如图3所示。

由图3可以看出，在长对话方面，本文的构边方法性能显著提升，达到了2.11%，相比之下，在中对话上的性能提升为1.17%，而在短对话上的提升则仅有0.8%，这说明本文的构边方法在处理长对话时有更为显著的性能优势。这一结果也符合预期，因为在长对话中，语境的建模更加复杂，语句间的关联跨度更大，而本文的构边方法从语句主题的角度出发，很好地规避了长对话建模的困难问题。总体而言，本文研究结果表明，在长对话场景中，采用基于主题的构边方法能够更有效地提升模型性能，这为其在实际对话系统中的应用提供了有力的支持。

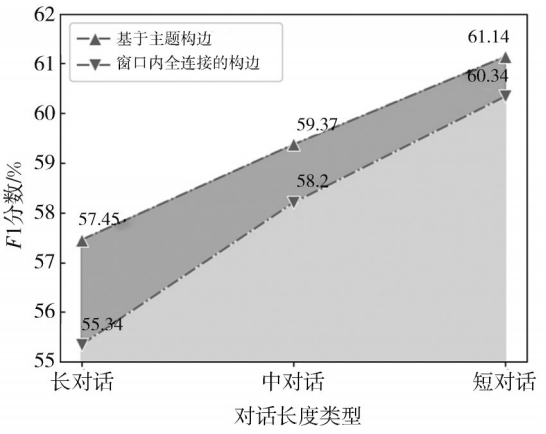

图3 主题构边方法对不同长度数据的影响

Fig. 3 The influence of topic edging method on data of different length

3.3.4 参数分析

图神经网络通过聚合邻居信息来更新节点特征，因此，当叠加的网络层数过高时，可能会出现过度平滑的现象，从而降低模型性能。为了探讨门控单元与模型层数对性能的影响，本文进行了实验，将模型层数设置在1～8之间，并与DialogGCN进行了比较。实验结果如图4所示。

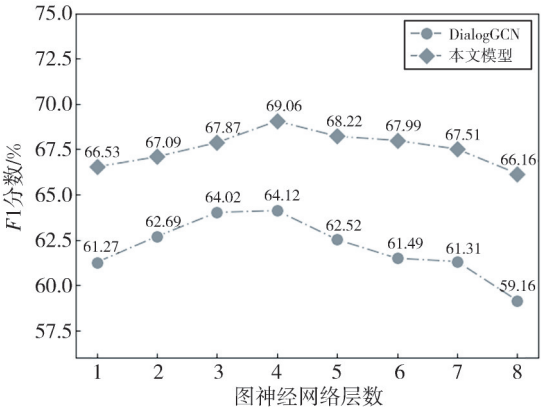

图4 图神经网络层数对模型的影响

Fig. 4 Figure the effect of the number of layers of neural network on the model

由图4可以出看出，在IEMOAP数据集上，本文方法在层数为4时取得最佳性能，而随着层数的增加，模型性能开始缓慢下降。相比之下，DialogGCN在层数增加时受到的影响更大，呈现更为明显的性能下降趋势。这表明门控单元能够有效控制层与层之间的信息流动，从而抵消过度平滑现象的影响，增强模型的鲁棒性。

4 结 语

本文针对现有图神经网络存在的固定构边问题，提出了融合主题模型的构边方法，充分利用主题关系建立了合理的语句连接。此外，还引入门控单元来增加模型的非线性能力。实验结果表明，本文模型方法在4个数据集上都达到了具有竞争力的水平。

参考文献：

[1] 洪巍，李敏. 文本情感分析方法研究综述[J]. 计算机工程与科学，2019，41(4)：750-757.
HONG Wei，LI Min. A review：Text sentiment analysis methods[J]. Computer Engineering and Science，2019，41(4)：750-757.(in Chinese)

[2] 李思贤. 基于恐惧情感强度计算的自动驾驶车辆决策机制研究[D]. 青岛：山东科技大学，2020.

[3] MAJUMDER N，HONG P，PENG S，et al. MIME：MIMicking emotions for empathetic response generation[C]//Conference on Empirical Methods in Natural Language Processing (EMNLP)，2020：8968-8979.

[4] DAUPHIN Y N，FAN A，AULI M，et al. Language modeling with gated convolutional networks[C]//International Conference on Machine Learning，PMLR，2017：933-941.

[5] LAN S，MA Y，HUANG W，et al. Dstagnn：Dynamic spatial-temporal aware graph neural network for traffic flow forecasting[C]//International Conference on Machine Learning，PMLR，2022：11906-11917.

[6] SHAZEER N. GLU variants improve transformer[DB/OL].(2020-02-12)[2023-11-15]. http://arxiv.org/abs/2002.05202v1.

[7] HAZARIKA D，PORIA S，MIHALCEA R，et al. Icon：Inter- active conversational memory network for multimodal emotion detection[C]//Conference on Empirical Methods in Natural Language Processing，2018：2594-2604.

[8] HAZARIKA D，PORIA S，ZADEH A，et al. Conversational memory network for emotion recognition in dyadic dialogue videos[C]//Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. New Orleans，2018：2122-2132.

[9] MAJUMDER N，PORIA S，HAZARIKA D，et al. Dialogue-RNN：An attentive rnn for emotion detection in conversations[C]//AAAI Conference on Artificial Intelligence，2019：6818-6825.

[10] ZHUANG C，MA Q. Dual graph convolutional networks for graph-based semi-supervised classification[C]//World Wide Web Conference，2018：499-508.

[11] GHOSAL D, MAJUMDER N, PORIA S, et al. Dialogue-GCN: 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),2020: 154-164.

[12] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019: 4171-4186.

[13] LIU Y H, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized BERT pretraining approach[DB/OL]. (2019-07-26)[2023-11-15]. http://arxiv. org/abs/1907. 11692.

[14] ISHIWATARI T, YASUDA Y, MIYAZAKI T, et al. Relati-on aware graph attention networks with relational position encodings for emotion recognition in conversations[C]//Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020: 7360-7370.

[15] ZHANG D, WU L Q, SUN C L, et al. Modeling both context-and speaker-sensitive dependence for emotion detection in multi-speaker conversation[C]//28th International Joint Conference on Artificial Intelligence, 2019: 5415-5421.

[16] SHEN W, WU S, YANG Y, et al. Directed acyclic graph network for conversational emotion recogniton [C]//The 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 1551-1560.

[17] JELODAR H, WANG Y, YUAN C, et al. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey[J]. Multimedia Tools and Applications, 2019, 78(11): 15169-15211.

[18] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[DB/OL]. (2017-10-30) [2023-11-15]. http://arxiv. org/abs/1710. 10903.

[19] YANG C Q, WANG R J, YAO S C, et al. Revisiting over-smoothing in deep GCNs[DB/OL]. (2020-03-30) [2023-11-15]. https://arxiv. org/abs/2003. 13663.

[20] 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.

[21] PORIA S, HAZARIKA D, MAJUMDER N, et al. MELD: A Multimodal multi-party dataset for emotion recognition in conversations[DB/OL]. (2018-10-05) [2023-11-15]. https://arxiv. org/abs/1810. 02508.

[22] LI Y R, SU H, SHEN X Y, et al. DailyDialog: A manu-ally labelled multiturn dialogue dataset [DB/OL]. (2017-10-11)[2023-11-15]. https://arxiv. org/abs/1710. 03957.

[23] ZAHIRI S M, CHOI J D. Emotion detection on TV show transcripts with sequence-based convolutional neural networks [DB/OL]. (2017-08-14) [2023-11-15]. https://arxiv. org/abs/1708. 04299.

[24] GHOSAL D, MAJUMDER N, GELBUKH A, et al. COSMIC: Common sense knowledge for emotion identification in conversations [C]//Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020. Association for Computational Linguistics (ACL), 2020: 2470-2481.

[25] 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: 2949-2958.

[26] 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: 1204-1214.

[27] PORIA S, CAMBRIA E, HAZARIKA D, et al. Context-dependent sentiment analysis in user-generated videos [C]//The 55th Annual Meeting of the Association for Computational Linguistics, 2017: 873-883.