



ISSN 1000-3428
CN 31-1289/TP
CODEN JISGEV

计算机工程

Computer Engineering

www.ecice06.com

第48卷 第1期 2022年1月15日
Vol.48 No.1 15 January 2022

2022 1

■北大中文核心期刊 ■中科院CSCD来源期刊 ■中国科技核心期刊 ■RCCSE中国核心学术期刊 ■工业和信息化部精品期刊

第48卷 第1期 2022年1月15日

www.ecice06.com



华东计算技术研究所 主办
上海市计算机学会

计 算 机 工 程 (月 刊)

Jisuanji Gongcheng

第 48 卷 第 1 期 2022 年 1 月 15 日

目 次

• 热点与综述 •

量子计算模拟及优化方法综述·····	喻志超,李扬中,刘 磊,冯圣中 (1)
深度学习方法在兴趣点推荐中的应用研究综述·····	汤佳欣,陈 阳,周孟莹,王 新 (12)
基于区块链非同质化代币的软件订阅模型·····	张展鹏,张 亮,彭凌祺,阚海斌 (24)
基于区块链的物联网数据服务信誉评估模型·····	黄志清,解鲁阳,张严心,尹泽明 (33)
基于区块链的微电网数据安全共享方案 ·····	张利华,曹 宇,张赣哲,黄 阳,陈世宏 (43)

• 人工智能与模式识别 •

一种优化初始点与自适应半径的密度聚类算法 ·····	王治和,曹旭琰,杜 辉 (51)
基于反向可达集的影响力最大化算法 ·····	邓心惠,宾 晟,孙更新 (60)
基于多通道脑电信号的因果网络情绪识别 ·····	王 斌,王忠民,张 荣 (69)
面向置换检验的冗余对比模式过滤算法 ·····	吴 军,欧阳艾嘉,张 琳 (75)
一种基于层次化 R-GCN 的会话情绪识别方法 ·····	赖河菴,李伶俐,胡婉玲,颜学明 (85)
基于级联加性噪声模型的因果结构学习算法 ·····	乔 杰,蔡瑞初,郝志峰 (93)
基于改进图节点的图神经网络多跳阅读理解研究 ·····	舒 冲,欧阳智,杜逆索,何 庆,魏 琴 (99)
融合用户信任度的概率矩阵分解群组推荐算法 ·····	宋玉龙,马文明,刘彤彤 (105)
基于自适应分层阈值判断的神经网络模型压缩·····	卢 鹏,万 莹,邹国良,陈金宇,郑宗生,王振华 (112)
基于用户意图的微博文本生成技术研究·····	高永兵,黎预璇,高军甜,马占飞 (119)

• 先进计算与数据处理 •

基于位置的路网 Skyline 查询处理研究 ·····	白 梅,裴仕涵,王习特 (127)
车联网中视频内容理解任务的计算卸载决策研究·····	冯 浩,郭彩丽 (135)
面向国产平台的 LLVM 自动向量化移植与优化 ·····	李嘉楠,韩 林,柴赞达 (142)

Computer Engineering (Monthly)

Vol.48

No.1

15 January 2022

Contents

• Research Hotspots and Reviews •

Survey of Quantum Computing Simulation and Optimization Methods	YU Zhichao, LI Yangzhong, LIU Lei, FENG Shengzhong (1)
A Survey of Studies on Deep Learning Applications in POI Recommendation	
.....	TANG Jiaxin, CHEN Yang, ZHOU Mengying, WANG Xin (12)
Software Subscription Model Based on Non-Fungible Tokens in Blockchain	
.....	ZHANG Zhanpeng, ZHANG Liang, PENG Lingqi, KAN Haibin (24)
Blockchain-based Reputation Evaluation Model for Data Services in Internet of Things	
.....	HUANG Zhiqing, XIE Luyang, ZHANG Yanxin, YIN Zeming (33)
Blockchain-based Security Data Sharing Scheme for Microgrid	
.....	ZHANG Lihua, CAO Yu, ZHANG Ganzhe, HUANG Yang, CHEN Shihong (43)

• Artificial Intelligence and Pattern Recognition •

A Density Clustering Algorithm for Optimizing Initial Points and Adaptive Radius	WANG Zhihe, CAO Xuyan, DU Hui (51)
Influence Maximization Algorithm Based on Reverse Reachable Set	DENG Xinhui, BIN Sheng, SUN Gengxin (60)
Emotion Recognition Using Causal Network Based on Multi-Channel EEG Signal.....	WANG Bin, WANG Zhongmin, ZHANG Rong (69)
Redundant Contrast Pattern Filtering Algorithm for Permutation Testing	WU Jun, OUYANG Aijia, ZHANG Lin (75)
Method for Conversational Emotion Recognition Using Hierarchical R-GCN	LAI Helang, LI Lingli, HU Wanling, YAN Xueming (85)
Causal Structure Learning Algorithm Based on Cascade Additive Noise Model	QIAO Jie, CAI Ruichu, HAO Zhifeng (93)
Research on Multi-Hop Reading Comprehension Based on Graph Neural Network with Improved Graph Nodes.....	
.....	SHU Chong, OUYANG Zhi, DU Nisuo, HE Qing, WEI Qin (99)
Group Recommendation Algorithm Incorporating User Trust with Probabilistic Matrix Factorization	
.....	SONG Yulong, MA Wenming, LIU Tongtong (105)
Neural Network Model Compression Based on Adaptive Hierarchical Threshold Judgment	
.....	LU Peng, WAN Ying, ZOU Guoliang, CHEN Jinyu, ZHENG Zongsheng, WANG Zhenhua (112)
Research on Weibo Text Generation Technology Based on User Intention	GAO Yongbing, LI Yuxuan, GAO Juntian, MA Zhanfei (119)

• Advanced Computing and Data Processing •

Research on Location-based Skyline Queries Processing in Road Network	BAI Mei, CHANG Shihan, WANG Xite (127)
Research on Computing Offloading Decision for Video Content Understanding Tasks in Internet of Vehicles ...	FENG Hao, GUO Caili (135)
Automatic Vectorization Transplant and Optimization of LLVM for Domestic Processors	LI Jia'nan, HAN Lin, CHAI Yunda (142)
Parallel Solution and Optimization of Large-Scale Sparse Linear System in GRAPES Dynamic Framework	
.....	ZHANG Kun, JIA Jinfang, YAN Wenxin, HUANG Jianqiang, WANG Xiaoying (149)
A Nonlinear Static Scheduling Strategy for Linear Loop Structure.....	LI Yapeng, PANG Jianmin, XU Jinlong, NIE Kai (155)



一种基于层次化 R-GCN 的会话情绪识别方法

赖河蓂^{1,2}, 李伶俐², 胡婉玲¹, 颜学明³

(1. 华南师范大学 计算机学院, 广州 510631; 2. 广东司法警官职业学院 信息管理系, 广州 510520;

3. 广东外语外贸大学 信息科学与技术学院, 广州 510006)

摘 要: 会话情绪识别包括说话者自身情绪影响以及说话者之间情绪影响这2个重要因素, 为了充分考虑上述影响因素以提高会话情绪识别效果, 提出一种基于层次化关系图卷积神经网络(HRGCN)的识别方法。使用一个基础神经网络对会话序列的特征数据进行优化, 按照不同的说话者划分出2个不同的会话子序列, 采用2个局部关系图卷积神经网络(R-GCN)分别对2个子序列进行局部建模, 按照会话发生的时间顺序重新整合局部建模后的2个子序列, 并利用全局 R-GCN 对其进行全局建模。通过对输入的多模态特征数据的分层次建模, 使得会话序列捕获到更多的上下文信息。在 IEMOCAP 数据集上的实验结果表明, 与当前流行的循环神经网络 LSTM、GRU 等相比, HRGCN 方法的会话情绪识别性能较高, 准确率与 F1 值分别达到 84.48% 与 84.40%。

关键词: 基础神经网络; 关系图卷积神经网络; 会话; 情绪识别; 人工智能

开放科学(资源服务)标志码(OSID):



中文引用格式: 赖河蓂, 李伶俐, 胡婉玲, 等. 一种基于层次化 R-GCN 的会话情绪识别方法[J]. 计算机工程, 2022, 48(1): 85-92.

英文引用格式: LAI H L, LI L L, HU W L, et al. Method for conversational emotion recognition using hierarchical R-GCN[J]. Computer Engineering, 2022, 48(1): 85-92.

Method for Conversational Emotion Recognition Using Hierarchical R-GCN

LAI Helang^{1,2}, LI Lingli², HU Wanling¹, YAN Xueming³

(1. School of Computer Science, South China Normal University, Guangzhou 510631, China;

2. Department of Information Management, Guangdong Justice Police Vocational College, Guangzhou 510520, China;

3. School of Information Science and Technology, Guangdong University of Foreign Studies, Guangzhou 510006, China)

[Abstract] Conversational emotion recognition should consider not only the emotions of the speakers themselves, but also the emotions passing between speakers. This paper proposes an emotion recognition method based on Hierarchical Relational Graph Convolutional Network (HRGCN), which considers both two types of emotions to improve the recognition performance. The method employs a Basic Neural Network (BNN) to optimize the feature data of the conversational sequence, and divides the sequence into two different conversational subsequences according to the speaker. Two local Relational Graph Convolutional Networks (R-GCN) are used for local modelling of these two subsequences respectively, and the two locally modeled subsequences are reconcatenated in chronological order of the conversation. Furthermore, the global R-GCN is used to model the reconcatenated sequence globally. through hierarchical modeling of the input multimodal feature data, HRGCN can capture more contextual information. The experimental results on the IEMOCAP dataset show that HRGCN displays an accuracy of 84.48% and a F1 score of 84.40%, higher than LSTM, GRU and other mainstream recurrent neural networks.

[Key words] Basic Neural Network (BNN); Relational Graph Convolutional Network (R-GCN); conversation; emotion recognition; artificial intelligence

DOI: 10. 19678/j. issn. 1000-3428. 0060346

基金项目: 国家自然科学基金青年基金项目(62006053); 广东省教育厅特色创新类项目(2018KQNCX072); 2020年度广东省普通高校青年创新人才项目(2020KQNCX186); 2020年度广东省普通高校特色创新项目(2020KTSCX273); 广东省高等教育学会“十三五”规划2019年度高校青年教师高等教育学研究课题(19GGZ070); 广东司法警官职业学院第四届院级课题(2020YB16)。

作者简介: 赖河蓂(1985—), 男, 博士研究生, 主研方向为情感分析、服务计算理论与技术; 李伶俐(通信作者), 教授; 胡婉玲, 硕士研究生; 颜学明, 博士。

收稿日期: 2020-12-21 **修回日期:** 2021-01-21 **E-mail:** mmonly@163.com

0 概述

随着人工智能技术的快速发展,推荐系统、舆情预测、财务预测、智能搜索、对话系统、人机交互等相关应用得到普及,其中,人机交互受到科研人员的广泛关注。用户在交互过程中会产生各种形式的生理和行为数据,人工智能追求的目标是人机可以无缝衔接,即当用户与机器进行情感交互时,机器可以全面准确地分析交互数据并洞察、理解、表达各种情绪。因此,需要构建能够识别和理解情绪的机器,使得它们可以像人类一样进行自然和友好的互动。

情绪是人们受到外界刺激时在感觉、思想或行为等方面产生的一些心理或生理反应^[1],如生气、开心、伤心、兴奋、沮丧、害怕、惊讶等。情绪在人们的生活中发挥着极其重要的作用,其会影响人们的日常交流、行为举止、判断、决策等。情绪识别是指通过对表达的信息(通常该信息单元是一个“话语”)进行分析,进而判断该“话语”属于哪一种情绪。虽然当前已经有较多关于情绪识别方面的研究,但是较少有学者对会话情绪识别进行研究。随着神经网络、机器学习、深度学习^[2-3]等人工智能技术的发展,会话情绪识别成为研究热点,在对话系统中,情绪识别极其重要,其要求机器人能识别并透彻地理解会话中的情绪,从而给出带有感情的响应或回复。然而,由于会话中说话者之间存在很强的相互依赖性,使得会话情绪识别成为一项具有挑战性的任务^[4]。在一个交互的会话中,有2个重要因素影响情绪动态变化:一为说话者自身的情绪影响,指个人在某一时刻的情绪传递到另一时刻;二为说话者之间的情绪影响^[5-6],指个人情绪受到了另一说话者的情绪影响。

本文考虑上述2个影响因素,提出一种基于层次化关系图卷积神经网络(HRGCN)的会话情绪识别方法,其层次化地使用关系图卷积神经网络(R-GCN)对会话中产生的序列“话语”进行建模,从而识别会话中的情绪。具体地,用多模态表示会话中的每一个“话语”,将整个序列输入到一个基础神经网络(该网络可以是LSTM或GRU等),经过基础神经网络输出后,按照不同的说话者将整个序列划分成不同的子序列(在本文中划分为2个子序列)。不同的子序列输送到局部R-GCN(隶属于不同的说话者)进行计算,并将计算结果按照会话发生的时间顺序相级联,级联后的序列“话语”送到全局R-GCN中进行计算,计算后得到的全局特征表示传送到完全连接层进行降维操作,完全连接层的输出即为情绪分类预测结果。

1 相关工作

近年来,随着对话系统应用的普及,会话情绪识别逐渐成为人机交互领域中的一个研究热点,同时也受到国内外研究人员的广泛关注。会话情绪识别是一个涉及信号处理、心理学、自然语言处理、认知科学等多个交叉学科的研究领域。

PORIA等^[7]提出基于卷积多核学习的分类器以及基于上下文的层次化双向LSTM模型^[8],以进行多模态情绪识别以及情感分析。HAZARIKA等^[4]首先提出会话存储网络模型,该模型使用2个不同的GRU来捕获不同说话者所表达“话语”的上下文信息,随后,他们改进该模型,通过增加一个GRU对说话者之间进行建模,该GRU可以看作跟踪整个会话的记忆存储^[9],此外,他们又提出迁移学习的方法,该方法先训练一个神经对话生成模型,然后利用该模型的参数来初始化目标情绪分类器^[10]。MAJUMDER等^[11]描述一个基于3个GRU的循环神经网络,这3个GRU分别用于追踪会话中说话者的个人状态、全局上下文信息以及情感状态,该神经模型能有效捕获说话者之间的关系并扩展到多个说话者的会话中。JIN等^[12]将层次化的多模态转换器作为基础网络,设计一个局部感知的注意力机制和一个说话者感知的注意力机制,以分别捕获局部上下文以及说话者的情绪惯性。高玮军等^[13]建立一种基于注意力机制的深度学习模型AT-DPCNN,其通过注意力矩阵重点关注文本序列中对情感走向影响较大的部分,解决了传统卷积神经网络在提取特征过程中存在的信息丢失问题。PORIA等^[14]对会话情绪识别中存在的挑战、数据集以及最新进展进行了详细综述。

由于图神经网络(GNN)、图卷积神经网络(GCN)和R-GCN在捕获相邻信息时具有优势以及在预测分类方面效果较好,因此在情感分析和情绪识别领域得到广泛应用。ZHANG等^[15]提出基于图卷积神经网络的情绪识别模型,用于解决会话中的上下文依赖以及说话者相互依赖这2个挑战性问题。ZHONG等^[16]设计一个知识丰富型的转换器,利用层次化结构的自我关注来诠释上下文中的“话语”,并通过一个上下文感知的情感图注意力机制来动态地使用相关常识。XU等^[17]使用图卷积神经网络进行特征提取以及面部表情识别分类。WANG等^[18]提出基于锁相值的图卷积神经网络,用于多通道脑电情绪识别。SONG等^[19]设计一个动态图卷积神经网络来进行多通道脑电情绪识别研究。LO等^[20]提出一种基于关系建模的图卷积神经网络,并将其用于微表情识别。GHOSAL等^[21]建立一个对话图卷积神经网络,用于捕获说话者自身以及说话者之间的相互依赖关系。

上述方法都取得了较好的识别效果,但它们大多没有考虑不同说话者自身表达的特点以及历史上下文和将来上下文信息存在不同权重的问题。本文将不同说话者在会话中所表达的“话语”分开进行局部建模,然后再级联完成全局建模,在建模过程中采用非对等窗口技术表示历史上下文和将来上下文信息所隐含的权重。

2 层次化R-GCN模型

为了有效捕获上下文之间的依赖关系,需要同时考虑说话者自身的情绪影响以及说话者之间的情绪影响。图1所示为IEMOCAP(Interactive Emotional dyadic Motion Capture database)数据集中抽取出的一个对话片段,其体现了上述2个因素对情绪变化的影响。

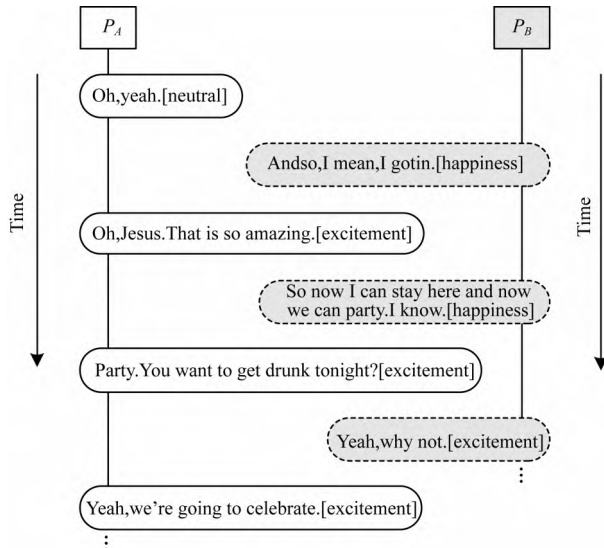


图1 情绪影响示例

Fig.1 Examples of emotional impact

本文提出一种HRGCN方法进行会话情绪识别,所研究的会话是二元会话,即在会话交互过程中有2个不同的说话者。由于不同说话者具有各自表达的特点,本文将整个会话序列划分成2个不同的子序列(分别对应2个说话者各自表达的“话语”序列),并用2个局部关系图卷积神经网络对2个子序列进行独立建模。

文献[21]在划分边关系类型时没有考虑“话语”在会话序列中的位置,未区分当前测试“话语”与其他“话语”之间的距离大小。与文献[21]不同,本文认为不同时刻的“话语”所产生的情绪依赖影响是不同的,即不同时刻的“话语”含有不同的权重,在划分边关系类型时,应该将当前测试“话语”与其他“话语”之间的距离大小作为划分结果的主要影响因素之一。在会话过程中,当前测试“话语”的情绪受历史上下文的影响较大,受将来上下文的影响较小。为了强调当前“话语”的历史上下文和将来上下文所占权重不同,本文使用非对称距离大小技术来表示权重,即将历史和将来2个不同方向的距离设置为不同的值。例如,令测试“话语” u_i 与其他“话语”之间的距离大小为 d ,在 u_i 的历史上下文方向设置一个距离划分点,假设距离大小为 D'_p ,在 u_i 的将来上下文方向设置一个距离划分点,假设距离大小为 D'_f ,然后讨论 d, D'_p, D'_f 三者之间的关系,最后得到边关系类型以及每条边所属的类型。

图2所示为HRGCN结构框架,其主要模块有基础神经网络(BNN)、局部关系图卷积神经网络(Local R-GCN)以及全局关系图卷积神经网络(Global R-GCN)。

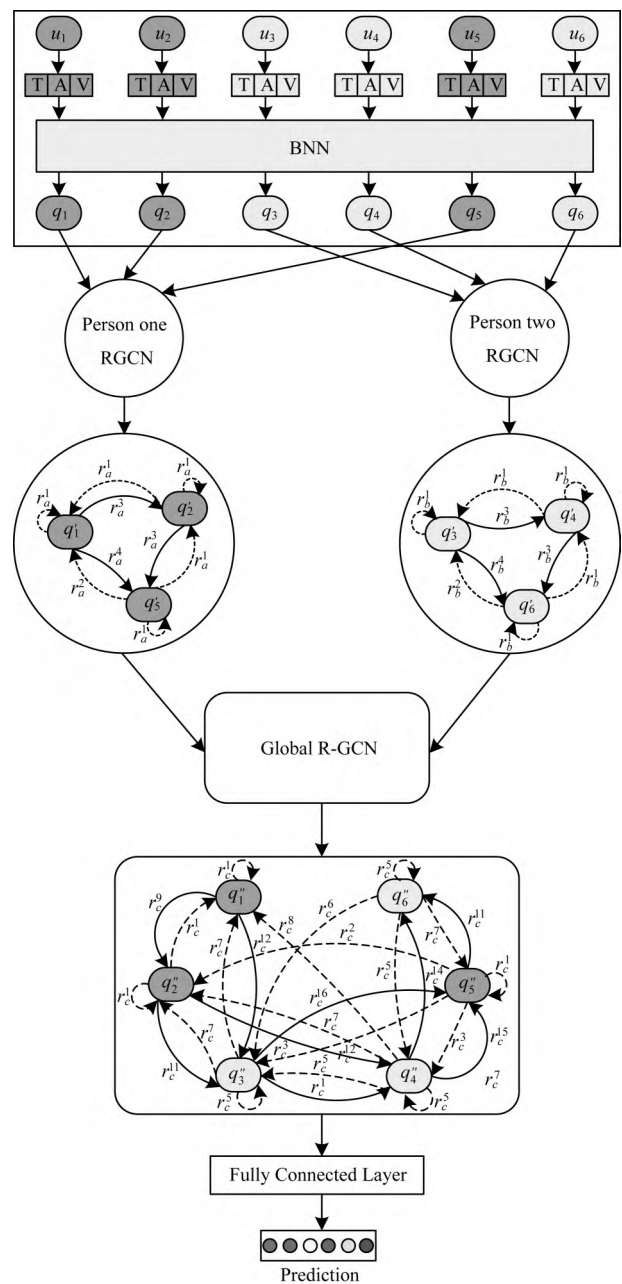


图2 HRGCN结构框架

Fig.2 HRGCN structural framework

2.1 基础神经网络

基础神经网络的目的是使多模态数据在输入到关系图卷积神经网络之前获得较好的特征表示,即为后续的关系图卷积计算打下基础。该模块可以选择不同的神经网络,如循环神经网络LSTM或GRU。在本文实验部分,分别采用这2种不同的神经网络作为基础神经网络,并对比实验效果。

假设有一个二元会话,其“话语”序列是 $u_1, u_2, u_3, u_4, u_5, u_6$,其中, u_1, u_2, u_5 是第一个说话者表

达的“话语”, u_3, u_4, u_6 是第二个说话者表达的“话语”。将整个序列输入到基础神经网络时, 具体的计算过程为: 首先将序列中每个“话语”的各个模态数据(包括文本、声音和图像)级联起来; 然后输入到双向 LSTM 或 GRU 网络进行计算, 得到输出序列 $q_1, q_2, q_3, q_4, q_5, q_6$ 。

2.2 局部关系图卷积神经网络

特征数据在经过基础神经网络之后, 将在局部关系图卷积神经网络模块进行局部关系图卷积计算: 首先, 根据不同的说话者对序列“话语”进行划分; 然后, 将这 2 个不同的子序列输送到 Person one RGCN 和 Person two RGCN 这 2 个不同的局部关系图卷积神经网络。本文所提方法基于局部邻居信息(即一定范围内的邻居“话语”)进行卷积处理, 此外, 考虑到历史上下文和将来上下文隐含的权重不同, 本文引入距离大小非对等技术。

2.2.1 图结构

一个具有 N 个“话语”的会话, 其关系图网络结构可以用一个有向图 $G=(V, E, R)$ 来表示: 图中每个顶点 $v_i \in V$ 表示一个“话语”; 顶点 v_i 和顶点 v_j ($i, j \in [1, 2, \dots, N]$) 之间的边为 $(v_i, v_j) \in E$, 边的关系类型为 $r \in R$ 。

2.2.2 顶点划分

在对自身情绪影响进行建模时, 考虑到不同说话者具有不同的特点, 本文将不同说话者在会话中所表达的“话语”进行划分, 如图 2 所示, 将序列 $q_1, q_2, q_3, q_4, q_5, q_6$ 划分为 q_1, q_2, q_5 和 q_3, q_4, q_6 。划分后将这 2 个子序列分别送到 Person one RGCN 和 Person two RGCN 进行卷积运算。

2.2.3 基于非对等上下文窗口的边集合

边集合中具体包含哪些边取决于建模过程中考虑的上下文内容范围。在一个会话中, 如果每个“话语”的上下文内容考虑的范围是其他所有的“话语”, 那么会产生一个完全连接图, 此时每个顶点与其他所有顶点(包括自身)都有一条边。由于这种情况会花费大量的计算时间, 因此本文只考虑局部邻居信息进行卷积处理, 即在收集局部邻居信息时使用上下文窗口来限制范围。例如, 将历史上下文窗口设置为 W_p , 将来上下文窗口设置为 W_f , 图中任意一个“话语”为 $v_i \in V$, 则该“话语” v_i 对应的边集合中包含 v_i 自身、 W_p 个历史“话语”(在 v_i 之前的 $v_i-1, v_i-2, \dots, v_i-W_p$) 以及 W_f 个将来“话语”(在 v_i 之后的 $v_i+1, v_i+2, \dots, v_i+W_f$)。

进一步区分 Person one RGCN 和 Person two RGCN 网络的上下文窗口, 将其分别设置为 W_{p1}, W_{p2} 与 W_{f1}, W_{f2} 。为了进行简化, 在图 2 所示的模型中, 这 4 个窗口的大小均设置为 2, 需要注意的是, 在优化实验参

数时, 可以将隶属于不同说话者的 W_p 和 W_f 窗口大小设置为非对等数值。在窗口值范围内, 说话者 Person one 的“话语”序列 q_1, q_2, q_5 所产生的边有 $\langle 1, 1 \rangle$ 、 $\langle 1, 2 \rangle$ 、 $\langle 1, 5 \rangle$ 、 $\langle 2, 2 \rangle$ 、 $\langle 2, 1 \rangle$ 、 $\langle 2, 5 \rangle$ 、 $\langle 5, 5 \rangle$ 、 $\langle 5, 1 \rangle$ 、 $\langle 5, 2 \rangle$ 。同理, 说话者 Person two 的“话语”序列 q_3, q_4, q_6 所产生的边有 $\langle 3, 3 \rangle$ 、 $\langle 3, 4 \rangle$ 、 $\langle 3, 6 \rangle$ 、 $\langle 4, 4 \rangle$ 、 $\langle 4, 3 \rangle$ 、 $\langle 4, 6 \rangle$ 、 $\langle 6, 6 \rangle$ 、 $\langle 6, 3 \rangle$ 、 $\langle 6, 4 \rangle$ 。

2.2.4 基于非对等距离大小的关系划分

在二元会话中, 2 个“话语”之间的时序关系依赖于它们在会话中的位置。在划分每条边所属关系时, 本文采用基于非对等距离大小的方法来产生不同时序“话语”之间的边关系类型。

在局部关系图卷积神经网络 Person one RGCN 中, 2 个非对等距离分别用 D_{p1} 和 D_{f1} ($0 < D_{p1} < W_{p1}, 0 < D_{f1} < W_{f1}$) 表示, 隶属于第一个说话者的会话子序列中的任意 2 个“话语”分别用 v_{i1} 和 v_{j1} 表示: 当 v_{i1} 发生的时间不早于 v_{j1} 且它们之间相隔的“话语”个数不大于 D_{p1} 时, 令它们之间的边关系类型为 r_a^1 ; 当 v_{i1} 发生的时间不早于 v_{j1} 且它们之间相隔的“话语”个数大于 D_{p1} 时, 令它们之间的边关系类型为 r_a^2 ; 当 v_{i1} 发生的时间早于 v_{j1} 且它们之间相隔的“话语”个数不大于 D_{f1} 时, 令它们之间的边关系类型为 r_a^3 ; 当 v_{i1} 发生的时间早于 v_{j1} 且它们之间相隔的“话语”个数大于 D_{f1} 时, 令它们之间的边关系类型为 r_a^4 。所有的边关系类型用集合 R_a 表示, 即 $R_a = \{r_a^1, r_a^2, r_a^3, r_a^4\}$ 。

同理, 在局部关系图卷积神经网络 Person two RGCN 中, 2 个非对等距离分别用 D_{p2} 和 D_{f2} ($0 < D_{p2} < W_{p2}, 0 < D_{f2} < W_{f2}$) 表示, 隶属于第二个说话者的会话子序列中的任意 2 个“话语”分别用 v_{i2} 和 v_{j2} 表示: 当 v_{i2} 发生的时间不早于 v_{j2} 且它们之间相隔的“话语”个数不大于 D_{p2} 时, 令它们之间的边关系类型为 r_b^1 ; 当 v_{i2} 发生的时间不早于 v_{j2} 且它们之间相隔的“话语”个数大于 D_{p2} 时, 令它们之间的边关系类型为 r_b^2 ; 当 v_{i2} 发生的时间早于 v_{j2} 且它们之间相隔的“话语”个数不大于 D_{f2} 时, 令它们之间的边关系类型为 r_b^3 ; 当 v_{i2} 发生的时间早于 v_{j2} 且它们之间相隔的“话语”个数大于 D_{f2} 时, 令它们之间的边关系类型为 r_b^4 。所有的边关系类型用集合 R_b 表示, 即 $R_b = \{r_b^1, r_b^2, r_b^3, r_b^4\}$ 。

图 2 中的 Person one RGCN 和 Person two RGCN 模块展示的是 $D_{p1}, D_{f1}, D_{p2}, D_{f2}$ 均设置为 1 时的划分情况。

2.2.5 基于局部序列的特征转换

根据 2 个不同的说话者划分出的 2 个不同子序列“话语”, 分别被送到 Person one RGCN 和 Person two RGCN 进行特征转换运算。本文使用受文献[21-23]启发的特定于关系的转换方法来进行转换运算, 通过收集基于非对等上下文窗口($W_{p1}, W_{p2}, W_{f1}, W_{f2}$)的邻居上下文信息(“话语”), 可以将某一

“话语” v_x 的特征 q_x 转换为新的特征表示 q'_x 。

在 Person one RGCN 中,运算公式如下:

$$q'_x = \sigma \left(\sum_{r \in R_a} \sum_{y \in N_x^r} \frac{1}{c_x} W_r q_y + W_{o_1} q_x \right) \quad (1)$$

在 Person two RGCN 中,运算公式如下:

$$q'_x = \sigma \left(\sum_{r \in R_b} \sum_{y \in N_x^r} \frac{1}{c_x} W_r q_y + W_{o_2} q_x \right) \quad (2)$$

其中: $x=1,2,\dots,N$; N_x^r 表示“话语” v_x 在关系 $r \in R_a$ 或 $r \in R_b$ 下的邻居“话语”的索引集合; c_x, r 是一个特定于问题的规范化常量,可以通过学习获得或预先设置(如 $c_x, r = |N_x^r|$); σ 是激活函数; W_r, W_{o_1}, W_{o_2} 是可以训练的转换参数。

2.3 全局关系图卷积神经网络

在局部关系图卷积神经网络 Person one RGCN 和 Person two RGCN 中分别计算出局部序列的新特征后,HRGCN 依据会话中发生的时序将 2 个局部序列进行级联,形成一个具有 N 个“话语”的序列(如图 2 中的序列 $q'_1, q'_2, q'_3, q'_4, q'_5, q'_6$),并输送到全局关系图卷积神经网络。

2.3.1 基于非对等上下文窗口的边集合

在全局关系图卷积神经网络中也采用非对等上下文窗口技术。例如,将历史上下文窗口设置为 W_p' ,将来上下文窗口设置为 W_f' 。在图 2 的模型中, W_p' 和 W_f' 的大小分别设置为 3 和 2,在这 2 个窗口值范围内,“话语”序列 $q'_1, q'_2, q'_3, q'_4, q'_5, q'_6$ 所产生的边有 $\langle 1,1 \rangle, \langle 2,2 \rangle, \langle 5,5 \rangle, \langle 2,1 \rangle, \langle 5,2 \rangle, \langle 5,3 \rangle, \langle 5,4 \rangle, \langle 3,3 \rangle, \langle 4,4 \rangle, \langle 6,6 \rangle, \langle 4,3 \rangle, \langle 6,4 \rangle, \langle 6,3 \rangle, \langle 3,1 \rangle, \langle 3,2 \rangle, \langle 4,2 \rangle, \langle 6,5 \rangle, \langle 4,1 \rangle, \langle 1,2 \rangle, \langle 2,3 \rangle, \langle 5,6 \rangle, \langle 1,3 \rangle, \langle 2,4 \rangle, \langle 3,4 \rangle, \langle 4,6 \rangle, \langle 4,5 \rangle, \langle 3,5 \rangle$ 。

2.3.2 基于非对等距离大小的关系划分

在全局关系图卷积神经网络中,边关系类型的划分比较复杂,需要分情况讨论。为了确定某条边的关系类型,本文考虑的因素包括“话语”的时序关系、“话语”的间隔距离关系以及“话语”的说话者关系。

类似于局部关系图卷积神经网络的划分,本文在间隔距离关系方面同样采用基于非对等距离大小的方法。2 个非对等距离分别用 D_p' 和 D_f' ($0 < D_p' < W_p', 0 < D_f' < W_f'$) 表示,任意 2 个“话语”分别用 v_i' 和 v_j' 表示 ($i, j' \in [1, 2, \dots, N]$),“话语” v_i' 和 v_j' 所对应的说话者分别用 $p_s(v_i')$ 和 $p_s(v_j')$ 表示,会话中的第一个说话者和第二个说话者分别用 p_{one} 和 p_{two} 表示。“话语” v_i' 和 v_j' 之间相隔的距离设置为 d 。在图 2 的模型中, D_p' 和 D_f' 分别设置为 2 和 1,则可以得到以下 2 种划分情况:

1) 当 v_i' 的时序不早于 v_j' 时,可得到如表 1 所示的关系划分结果。

表 1 图 2 中 Global R-GCN 关系划分结果 1

Table 1 Global R-GCN relationship division result 1 of fig.2

R	$p_s(v_i'), p_s(v_j')$	$d \leq D_p'$	(v_i', v_j')
r_c^1	p_{one}, p_{one}	Yes	$\langle 1,1 \rangle, \langle 2,2 \rangle, \langle 5,5 \rangle, \langle 2,1 \rangle$
r_c^2	p_{one}, p_{one}	No	$\langle 5,2 \rangle$
r_c^3	p_{one}, p_{two}	Yes	$\langle 5,3 \rangle, \langle 5,4 \rangle$
r_c^4	p_{one}, p_{two}	No	None
r_c^5	p_{two}, p_{two}	Yes	$\langle 3,3 \rangle, \langle 4,4 \rangle, \langle 6,6 \rangle, \langle 4,3 \rangle, \langle 6,4 \rangle$
r_c^6	p_{two}, p_{two}	No	$\langle 6,3 \rangle$
r_c^7	p_{two}, p_{one}	Yes	$\langle 3,1 \rangle, \langle 3,2 \rangle, \langle 4,2 \rangle, \langle 6,5 \rangle$
r_c^8	p_{two}, p_{one}	No	$\langle 4,1 \rangle$

2) 当 v_i' 的时序早于 v_j' 时,可得到如表 2 所示的关系划分结果。

表 2 图 2 中 Global R-GCN 关系划分结果 2

Table 2 Global R-GCN relationship division result 2 of fig.2

R	$p_s(v_i'), p_s(v_j')$	$d \leq D_f'$	(v_i', v_j')
r_c^9	p_{one}, p_{one}	Yes	$\langle 1,2 \rangle$
r_c^{10}	p_{one}, p_{one}	No	None
r_c^{11}	p_{one}, p_{two}	Yes	$\langle 2,3 \rangle, \langle 5,6 \rangle$
r_c^{12}	p_{one}, p_{two}	No	$\langle 1,3 \rangle, \langle 2,4 \rangle$
r_c^{13}	p_{two}, p_{two}	Yes	$\langle 3,4 \rangle$
r_c^{14}	p_{two}, p_{two}	No	$\langle 4,6 \rangle$
r_c^{15}	p_{two}, p_{one}	Yes	$\langle 4,5 \rangle$
r_c^{16}	p_{two}, p_{one}	No	$\langle 3,5 \rangle$

所有的边关系类型用集合 R_c 表示,即 $R_c = \{r_c^1, r_c^2, r_c^3, r_c^4, r_c^5, r_c^6, r_c^7, r_c^8, r_c^9, r_c^{10}, r_c^{11}, r_c^{12}, r_c^{13}, r_c^{14}, r_c^{15}, r_c^{16}\}$ 。

2.3.3 基于全局序列的特征转换

与局部序列的特征转换计算类似,全局序列的特征转换使用特定于关系的转换方法,基于非对等上下文窗口 (W_p' 和 W_f') 收集相邻上下文“话语”的信息,然后转换局部特征 q'_x 为全局特征 q''_x 。

Global R-GCN 的运算公式如下:

$$q''_x = \sigma \left(\sum_{r \in R_c} \sum_{y \in N_x^r} \frac{1}{c_x} W_r q'_y + W_{o'} q'_x \right) \quad (3)$$

2.4 预测输出

输入的特征数据经过上述层次化的关系图卷积神经网络计算(式(1)~式(3))之后,序列中的每个“话语”都积累了其相邻“话语”的相关信息。最后,将全局关系图卷积神经网络的计算结果输入到一个完全连接层,经过降维后进行预测分类。计算公式如下:

$$h'_x = \text{ReLU}(W_l q''_x + b_l) \quad (4)$$

$$h''_x = \text{softmax}(W_{\text{smax}} h'_x + b_{\text{smax}}) \quad (5)$$

$$\hat{z}_x = \arg\max_k (h''_x[k]) \quad (6)$$

在训练过程中,使用带 L2 正则化的交叉熵损失函数来计算损失值:

$$L_{\text{Loss}} = -\frac{1}{N} \sum_{x=1}^N \sum_{y=1}^C z_{x,y} \log_a(h''_{x,y}) + \omega \|\theta\|_2 \quad (7)$$

其中: N 表示训练集中所有“话语”的总个数; C 是情绪分类的数目; $z_{x,y}$ 是训练集中第 x 个“话语”属于第 y 类别的真值; $h''_{x,y}$ 是预测第 x 个“话语”属于第 y 类别的概率值; ω 是 L2 正则化的权重值; θ 是所有训练参数的集合。

3 实验结果与分析

3.1 数据集

本文实验基于 IEMOCAP 数据集进行^[24], 该数据集包含总长约 12 h 的多元会话视频数据, 这些视频共有 5 个会话, 每个会话又有若干个 5 min 左右的交互视频, 每个交互视频被进一步划分为若干个切片, 这些切片也称为“话语”, 所有的“话语”都用多模态(文本、声音以及图像的级联)进行表示, 并至少由三名标注人员标注一个情绪标签。本文考虑的情绪标签包括生气(anger)、开心(happiness)、伤心(sadness)、中立(neutral)、兴奋(excitement)、沮丧(frustration)共 6 类。表 3 所示为 IEMOCAP 数据集中各种情绪的分布情况。

表 3 IEMOCAP 数据集中各种情绪的分布情况

Table 3 Distribution of various emotions in IEMOCAP dataset

情绪分类	数量
anger	1 103
happiness	648
sadness	1 084
neutral	1 708
excitement	1 041
frustration	1 849

3.2 实验设置

参考文献[4, 9, 11, 21]对特征数据集进行预处理, 且同一说话者可以出现在训练集和测试集中。在实验过程中, 首先将下载的多模态特征数据集随机打乱, 然后随机划分出 80% 作为训练集和验证集, 20% 作为测试集。本文 HRGCN 方法所使用的优化器是 Adam^[25], 学习率设为 0.001。为了防止过拟合, 采用 Dropout^[26]机制。在训练阶段设置早停机制, 如果验证集的损失值连续 10 轮不再下降, 则停止训练。实验采用的评估指标是准确率(记为 ACC)和 F1 分数(记为 F1)。

本文在 IEMOCAP 数据集上进行会话情绪识别测试。实验中的对比模型如下:

1) LSTM^[27], 即长短期记忆神经网络, 是为了解决一般 RNN 存在的长期依赖问题而专门设计出的时间循环神经网络。

2) GRU^[28], 即门控循环单元, 它是为解决长期记忆和反向传播中的梯度等问题而设计出的时间循环神经网络。

3) Local R-GCN, 本文 HRGCN 方法中的 Local R-GCN 模块, 即 Person one RGCN 和 Person two RGCN 这 2 个部分的组合。

4) Global R-GCN, 本文 HRGCN 方法中的 Global R-GCN 模块。

5) HRGCN(BNN=GRU), 在本文 HRGCN 方法的 BNN 中装载 GRU 神经网络。

6) HRGCN(BNN=LSTM), 在本文 HRGCN 方法的 BNN 中装载 LSTM 神经网络。

3.3 结果分析

不同模型在 IEMOCAP 数据集上的实验结果如表 4 所示。从表 4 可以看出, HRGCN(BNN=LSTM)模型的识别效果最好, 其次是 HRGCN(BNN=GRU)模型, 效果较差的是 Local R-GCN 模型。

表 4 不同模型在 IEMOCAP 数据集上的实验结果

Table 4 Experimental results of different models on

IEMOCAP dataset		%
模型	ACC	F1
LSTM	81.28	81.37
GRU	80.21	80.30
Local R-GCN	78.41	78.59
Global R-GCN	81.88	81.79
HRGCN(BNN=GRU)	82.21	82.10
HRGCN(BNN=LSTM)	84.48	84.40

本文 HRGCN 方法的理论依据是: 在会话交互的过程中, 一个“话语”的情绪既受到说话者自身情绪变化的影响, 也受到说话者之间情绪变化的影响^[5-6]。局部 R-GCN 模块对说话者自身情绪的变化进行建模, 考虑到不同说话者会受到自身情绪变化的影响(有各自不同的表达特点), 局部 R-GCN 模块使用 2 个独立的关系图神经网络进行特征转换。全局 R-GCN 模块对说话者之间情绪的变化进行建模。实验结果表明, 特征数据经过局部 R-GCN 以及全局 R-GCN 之后, 可以更好地捕获说话者自身以及说话者之间的情绪影响, 从而提高会话情绪识别的性能。仅使用局部 R-GCN 或全局 R-GCN 都不能全面地捕获以上 2 个方面的影响, 其中的原因可能是: 单一的局部建模不能很好地反映会话是一个交互现象, 无法很好地捕获说话者之间的情绪影响; 单一的全局建模没有考虑到不同说话者各自的特点, 无法较好地捕获说话者自身的情绪变化影响。

4 超参数分析

本文对 HRGCN(BNN=LSTM)模型的相关超参数进行分析。令局部 R-GCN 计算输出结果 q'_x 的特征维度大小为 localSize, 全局 R-GCN 计算输出结果 q''_x 的特征维度大小为 globalSize, 本文使用随机参数搜索方法, 在实验中得到最优参数组合为 localSize=176, globalSize=88。当 globalSize 的大小固定为 88 时, 表 5 所示为 localSize 取不同值时 ACC 和 F1 的变化情况。当 localSize 的大小固定为 176 时, 表 6 所示为 globalSize 取不同值时 ACC 和 F1 的变化情况。

表5 localSize取不同值时ACC和F1的变化情况
Table 5 The change of ACC and F1 when localSize takes different values

localSize	ACC/%	F1/%
128	83.14	83.14
136	84.14	84.11
144	83.08	83.02
152	84.01	83.95
160	83.21	83.08
168	83.74	83.67
176	84.48	84.40
184	84.41	84.30
192	83.01	82.93
200	82.41	82.32
208	83.14	83.01

表6 globalSize取不同值时ACC和F1的变化情况
Table 6 The change of ACC and F1 when globalSize takes different values

globalSize	ACC/%	F1/%
40	83.28	83.19
48	83.34	83.29
56	84.14	84.02
64	82.81	82.66
72	83.41	83.32
80	82.94	82.95
88	84.48	84.40
96	83.14	83.03
104	83.28	83.19
112	83.21	83.11
120	82.94	82.82

此外,为了验证预测输出阶段的降维操作是否对情绪识别效果产生影响,本文分析维度取不同值时ACC和F1的变化情况。令预测输出阶段的维度大小为hiddenSizeFC,表7所示为hiddenSizeFC取不同值时ACC和F1的变化情况。

表7 hiddenSizeFC取不同值时ACC和F1的变化情况
Table 7 The change of ACC and F1 when hiddenSizeFC takes different values

hiddenSizeFC	ACC/%	F1/%
28	82.61	82.51
38	83.68	83.63
48	83.94	83.79
58	84.48	84.40
68	84.41	84.32
78	83.28	83.20
88	83.01	82.86
98	82.81	82.70
108	82.94	82.83
118	83.81	83.70
128	83.48	83.34

从表7可以看出:在开始阶段,随着hiddenSizeFC的增大,ACC和F1均略有提高;当hiddenSizeFC取值58时,ACC和F1达到最好效果;随后,在继续增大hiddenSizeFC时,ACC和F1均略有下降。由此可见,预测输出阶段的维度大小对HRGCN(BNN=LSTM)模型的分类效果具有一定影响。

5 结束语

本文提出一种基于层次化R-GCN的会话情绪识别方法。该方法利用2个不同的局部关系图卷积神经网络对输入特征数据进行局部建模,然后使用全局关系图卷积神经网络对所得局部建模结果进行全局建模,从而充分考虑不同说话者自身的特点以及相互之间的依赖关系。在情绪识别基准数据集IEMOCAP上进行实验,结果表明,相比循环神经网络,该方法的情绪识别性能较高,准确率高达84.48%。但是,本文所提方法仍存在一定的局限性,如涉及的超参数较多、层次化神经网络组合运算成本较高、运算实时性不佳等。因此,下一步将优化参数组合并通过调整神经网络结构等方法来降低模型的运算复杂度。

参考文献

- [1] 牛耘,潘明慧,魏欧,等. 基于词典的中文微博情绪识别[J]. 计算机科学,2014,41(9):253-258,289.
NIU Y, PAN M H, WEI O, et al. Emotion analysis of Chinese microblogs using lexicon-based approach [J]. Computer Science, 2014, 41(9): 253-258, 289. (in Chinese)
- [2] 周锦峰,叶施仁,王晖. 基于深度卷积神经网络模型的文本情感分类[J]. 计算机工程,2019,45(3):300-308.
ZHOU J F, YE S R, WANG H. Text sentiment classification based on deep convolutional neural network model [J]. Computer Engineering, 2019, 45(3): 300-308. (in Chinese)
- [3] 蔡林森,彭超,陈思远,等. 基于多样化特征卷积神经网络的情感分析[J]. 计算机工程,2019,45(4):169-174,180.
CAI L S, PENG C, CHEN S Y, et al. Sentiment analysis based on multiple features convolutional neural networks [J]. Computer Engineering, 2019, 45(4): 169-174, 180. (in Chinese)
- [4] HAZARIKA D, PORIA S, ZADEH A, et al. Conversational memory network for emotion recognition in dyadic dialogue videos[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics. [S. l.]: ACL, 2018: 2122-2132.
- [5] MICHAEL W M, DACHER K. How emotions work: the social functions of emotional expression in negotiations[J]. Research in Organizational Behavior, 2000, 22: 1-50.
- [6] LIU F, MAITLIS S. Emotional dynamics and strategizing processes: a study of strategic conversations in top team meetings[J]. Journal of Management Studies, 2014, 51(2): 202-234.
- [7] PORIA S, CHATURVEDI I, CAMBRIA E, et al. Convolutional MKL based multimodal emotion recognition and sentiment analysis [C]//Proceedings of 2016 IEEE

- International Conference on Data Mining . Washington D. C. , USA ;IEEE Press, 2016; 439-448.
- [8] 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. [S. l.] : ACL, 2017; 873-883.
- [9] HAZARIKA D, PORIA S, MIHALCEA R, et al. ICON: interactive conversational memory network for multimodal emotion detection[C]// Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Washington D. C. , USA ;IEEE Press, 2018; 2594-2604.
- [10] HAZARIKA D, PORIA S, ZIMMERMANN R, et al. Emotion recognition in conversations with transfer learning from generative conversation modeling[EB/OL]. [2020-11-02]. <https://arxiv.org/pdf/1910.04980v1.pdf>.
- [11] MAJUMDER N, PORIA S, HAZARIKA D, et al. DialogueRNN: an attentive RNN for emotion detection in conversations[C]// Proceedings of AAAI Conference on Artificial Intelligence. [S. l.] : AAAI Press, 2019; 6818-6825.
- [12] JIN X, YU J, DING Z, et al. Hierarchical multimodal transformer with localness and speaker aware attention for emotion recognition in conversations[C]// Proceedings of CCF International Conference on Natural Language Processing and Chinese Computing. Washington D. C. , USA ;IEEE Press, 2020; 41-53.
- [13] 高玮军, 杨杰, 张春霞, 等. 基于 AT-DPCNN 模型的情感分析研究[J]. 计算机工程, 2020, 46(11): 53-60.
GAO W J, YANG J, ZHANG C X, et al. Sentiment analysis reserach based on AT-DPCNN model [J]. Computer Engineering, 2020, 46(11): 53-60. (in Chinese)
- [14] PORIA S, MAJUMDER N, MIHALCEA R, et al. Emotion recognition in conversation; research challenges, datasets, and recent advances[J]. IEEE Access, 2019, 7: 100943-100953.
- [15] ZHANG D, WU L Q, SUN C L, et al. Modeling both context-and speaker-sensitive dependence for emotion detection in multi-speaker conversations[C]// Proceedings of the 28th International Joint Conference on Artificial Intelligence. Washington D. C. , USA ;IEEE Press, 2019; 5415-5421.
- [16] ZHONG P X, WANG D, MIAO C Y. Knowledge-enriched transformer for emotion detection in textual conversations [C]// Proceedings of Conference on Empirical Methods in Natural Language Processing. Washington D. C. , USA ;IEEE Press, 2019; 165-177.
- [17] XU X, RUAN Z, YANG L. Facial expression recognition based on graph neural network[C]// Proceedings of 2020 IEEE International Conference on Image, Vision and Computing. Washington D. C. , USA ;IEEE Press, 2020; 211-214.
- [18] WANG Z, TONG Y, HENG X. Phase-locking value based graph convolutional neural networks for emotion recognition[J]. IEEE Access, 2019, 7: 93711-93722.
- [19] SONG T, ZHENG W, SONG P, et al. EEG emotion recognition using dynamical graph convolutional neural networks[J]. IEEE Transactions on Affective Computing, 2020, 11(3): 532-541.
- [20] LO L, XIE H X, SHUAI H H, et al. MER-GCN: micro expression recognition based on relation modeling with graph convolutional network [EB/OL]. [2020-11-02]. <https://arxiv.org/pdf/2004.08915.pdf>.
- [21] GHOSAL D, MAJUMDER N, PORIA S, et al. DialogueGCN: a graph convolutional neural network for emotion recognition in conversation[EB/OL]. [2020-11-02]. <https://aclanthology.org/D19-1015.pdf>.
- [22] KIPF T N, WELLMING M. Semi-supervised classification with graph convolutional networks[EB/OL]. [2020-11-02]. <https://arxiv.org/pdf/1609.02907.pdf>.
- [23] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks [C]// Proceedings of European Semantic Web Conference. Berlin, Germany : Springer, 2018; 593-607.
- [24] CARLOS B, MURTAZA B, CHI-CHUN L, et al. IEMOCAP: interactive emotional dyadic motion capture database[J]. Language Resources and Evaluation, 2008, 42(4): 335-359.
- [25] KINGMA D P, JIMMY B. Adam: a method for stochastic optimization[EB/OL]. [2020-11-02]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [26] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [27] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [28] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[EB/OL]. [2020-11-02]. <https://arxiv.org/pdf/1412.3555v1.pdf>.

编辑 吴云芳