



计算机科学

COMPUTER SCIENCE

基于多粒度特征融合的新型图卷积网络用于方面级情感分析

邓入茵, 张清华, 黄帅帅, 高满

引用本文

邓入茵, 张清华, 黄帅帅, 高满. 基于多粒度特征融合的新型图卷积网络用于方面级情感分析[J]. 计算机科学, 2023, 50(10): 80-87.

DENG Ruhan, ZHANG Qinghua, HUANG Shuaishuai, GAO Man. [Novel Graph Convolutional Network Based on Multi-granularity Feature Fusion for Aspect-based Sentiment Analysis](#) [J]. Computer Science, 2023, 50(10): 80-87.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于方面语义和门控过滤网络的方面级情感分析](#)

Aspect-based Sentiment Analysis Based on Aspect Semantic and Gated Filtering Network

计算机科学, 2023, 50(10): 193-202. <https://doi.org/10.11896/jsjcx.220900192>

[基于多尺度Transformer融合多域信息的伪造人脸检测](#)

Forgery Face Detection Based on Multi-scale Transformer Fusing Multi-domain Information

计算机科学, 2023, 50(10): 112-118. <https://doi.org/10.11896/jsjcx.220900048>

[融合跟踪器:融合图像特征和事件特征的单目标跟踪框架](#)

Fusion Tracker: Single-object Tracking Framework Fusing Image Features and Event Features

计算机科学, 2023, 50(10): 96-103. <https://doi.org/10.11896/jsjcx.220900075>

[介粒度空间中的最优粒度选择和属性约简](#)

Optimal Granularity Selection and Attribute Reduction in Meso-granularity Space

计算机科学, 2023, 50(10): 71-79. <https://doi.org/10.11896/jsjcx.230500218>

[覆盖多粒度下的形式概念更新方法](#)

Method of Updating Formal Concept Under Covering Multi-granularity

计算机科学, 2023, 50(10): 18-27. <https://doi.org/10.11896/jsjcx.230600049>

基于多粒度特征融合的新型图卷积网络用于方面级情感分析

邓入茵^{1,2,3} 张清华^{2,3} 黄帅帅^{1,2,3} 高 满^{1,2,3}

1 重庆邮电大学计算机科学与技术学院 重庆 400065

2 重庆邮电大学计算智能重庆市重点实验室 重庆 400065

3 重庆邮电大学大数据智能计算重点实验室 重庆 400065

(s210231030@stu.cqupt.edu.cn)

摘 要 方面级情感分析是情感分析中的细粒度任务,旨在检测给定句子中方面词的情感极性。随着图卷积网络的兴起,通过依赖树构建的图卷积网络模型被广泛用于该任务,并取得了令人满意的效果。但大多数研究只获取图卷积网络最后一层输出作为分类层的输入,忽略了其他层的节点特征,且深层图卷积网络存在节点平滑问题。近年来,有研究者将图卷积网络的多层节点特征进行集成,提高了情感分类模型的性能。文中结合自适应特征融合与高速公路网络,提出了一种基于多粒度特征融合的高速公路图卷积网络模型,用于方面级情感分析。首先,该模型通过句法依赖结构和双向的上下文信息构建图卷积网络;同时,在图卷积网络引入高速公路网络缓解深层图卷积网络过平滑的问题,加深图卷积网络的深度。然后,使用自适应融合机制从不同深度图卷积网络获得多粒度节点信息。最后,在公共数据集上进行实验,实验结果表明,与基准模型相比,所提模型能更好地捕获更多粒度的句法信息和长距离依存关系。

关键词: 多粒度;特征融合;图卷积神经网络;高速公路网络;方面级情感分析

中图法分类号 TP391

Novel Graph Convolutional Network Based on Multi-granularity Feature Fusion for Aspect-based Sentiment Analysis

DENG Ruhan^{1,2,3}, ZHANG Qinghua^{2,3}, HUANG Shuaishuai^{1,2,3} and GAO Man^{1,2,3}

1 School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2 Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

3 Key Laboratory of Big Data Intelligent Computing, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract Aspect-based sentiment analysis (ABSA) is a fine-grained task in sentiment analysis that aims to detect the emotional polarity of aspects in given sentence. Due to the rise of deep learning and graph convolutional networks (GCNs), GCN constructed over dependency tree has been widely applied to ABSA and achieved satisfactory results. However, most studies only acquire the last layer node features of graph convolutional network (GCN) as input to the classifier, while ignoring other layer node features and GCNs have over-smoothing problem. In recent years, some researchers ensembled the multilayer node features of GCN, improving the performance of sentiment classification models. A model combines adaptively spatial feature fusion and highway networks, namely highway graph convolutional network based on multi-granularity feature fusion (MGFF-HGCN) is proposed for ABSA in this paper. First, this model constructs GCN by syntactic dependency structure and bidirectional context information, and highway networks is introduced for alleviating the deep GCN over-smoothing problem, deepening the depth of GCN. Then, an adaptive fusion mechanism is effectively employed to fuse the more comprehensive and multi-granularity node feature information obtained from various highway GCN (HGCN) layers. Finally, experimental results on public datasets show that the proposed method is comparable to the benchmark models and be able to capture more granular syntactic information and long-range dependencies relationship accurately.

Keywords Multi-granularity, Feature fusion, Graph convolutional networks, Highway networks, Aspect-based sentiment analysis

到稿日期:2023-04-10 返修日期:2023-06-05

基金项目:国家自然科学基金(62276038);重庆英才计划项目(CQYC20210202215);重庆市研究生教育教学改革研究项目(YJG203079);重庆市高等教育教学改革研究重大项目(201020)

This work was supported by the National Natural Science Foundation of China (62276038), Talent Program Project of Chongqing (CQYC20210202215), Research Project on Teaching Reform of Graduate Education in Chongqing (YJG203079) and Major Project on Teaching Reform Research of Higher Education of Chongqing (201020).

通信作者:张清华(zhangqh@cqupt.edu.cn)

1 引言

互联网时代是一场信息革命,人们的言论自由表达得到了多样化的发展,同时也受到了学术界和工业界的广泛关注,如在医疗保健、政治、经济等领域^[1],通过机器判断人类情感倾向成为了自然语言处理(NLP)领域的研究热点。最早的情感分析任务仅预测给定文本的整体情感倾向,随着情感分析任务的细化,方面级情感分析(Asspect-Based Sentiment Analysis, ABSA)被提出,该任务旨在检测给定文本中不同方面的情感极性(如积极的、中性的或消极的)。例如:“Decor is nice though service can be spotty.”,方面词“Decor”情感倾向被认为是积极的,但方面词“service”情感倾向被认为是消极的。

大多数早期 ABSA 任务利用神经网络,通过方面词上下文信息来判断方面词情感倾向,将这些模型分为基于简单神经网络的模型和基于注意力的神经网络模型。基于简单神经网络的模型,如 RNN 的各种变体^[2-3]、CNN^[4]等。单一神经网络模型对整个句子的情感分析更准确,而在复杂的句子中,其句法具有复杂性,简单的神经网络模型无法对复杂句式进行分析。随着注意力机制^[5]在神经网络中的广泛应用,大量的研究者将注意力机制与简单神经网络模型或其他领域的方法相结合,用于 ABSA 任务,如 MemNet^[6]和 AOA^[7]等。然而,这些基于注意力的神经网络仍然存在忽略文本上下文及其语法,缺乏提取上下文词与方面词之间的情感依赖关系功能的问题。例如,图 1 给出了句子“Decor is nice though service can be spotty.”的依存关系,即通过单词间的语法关系表示句子结构。对于该句子,单一的神经网络可能会错误地认为“service”更靠近情感倾向词“nice”,其情感倾向是正向的。

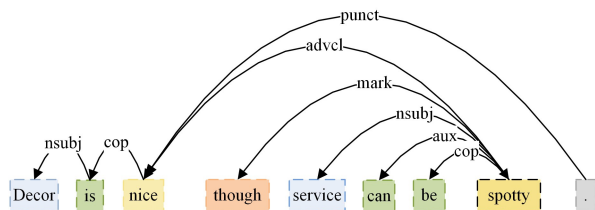


图 1 语法依赖结构示意图

Fig. 1 Illustration of grammar dependency tree

随着 GCNs 的发展,其因具有能根据不同文本的节点数目动态适应并处理各种长度文本等优点,被广泛应用于文本分类任务^[8]。Zhang 等^[9]基于句法依赖以及文本编码器对文本及其句法进行表示,通过 GCN 构建上下文词与方面词之间的关系,首次将 GCN 模型运用于 ABSA 任务。但大多数 GCN 模型仅取其最后一层特征信息作为分类器的输入,在节点聚合过程中会丢失部分词序信息或忽略短语,且深层 GCN 存在信息损失和过平滑问题^[10-11]。在 NLP 中表现为深层 GCN 使得单词节点的区分性变差,进而降低了方面级情感分析模型的性能。有学者研究在缓解神经网络过拟合的同时加深神经网络学习的层数。例如,高速公路网络(Highway Network)^[12]是一种特殊的神经网络框架,它允许特征信息在高速、自由地传递信息的同时加深神经网络的深度。Li 等^[13]提出了 3 种 DeepGCN 算法,实验结果验证了其良好的性能。Pang 等^[14]消除了噪声等不相关单词,提出了基于高速网络

的局部 GCN 模型。

粒计算是人类在自然启发中悟得的处理信息的方法,Pe-drycz 等^[15]指出信息粒度能表示和处理基本信息块。Yao^[16]提出了分层的粒度结构对应不同粒度的多层次理解。Wang 等^[17]提出可将数据挖掘过程看作是知识在不同粒度上的转换。粒计算中的多粒度思想被广泛应用于各领域,在 NLP 领域的情感分析任务中,根据粗粒度到细粒度划分该任务,可以分为文档级、句子级、方面级情感分析。

基于上述调查,从多粒度思想的角度出发,融合每一层 GCN 特征信息,本文提出了基于多粒度特征融合的高速公路图卷积网络(Highway Graph Convolutional Network Based on Multi-granularity Feature Fusion, MGFF-HGCN)模型。该模型旨在利用图卷积神经网络能够表现单词之间的语义关系、处理不同长度文本等特点,基于句法依存树和双向长短期记忆网络(BiLSTM)构建 GCN,获取与方面词相关的语法词;同时,在 GCN 中引入高速公路网络,构成高速公路图卷积网络(Highway Graph Convolutional Network, HGCN)卷积块,使得较深层 GCN 在缓解过平滑问题下扩大每个节点的感受野,即加深 GCN 深度使每个节点聚集更远距离的相邻词信息,缓解长距离语法依赖问题。由于每一层 GCN 节点信息聚合相邻词的词距离不同,不同深度 GCN 的特征信息呈现多粒度特点,本文从 Liu 等^[18]提出的融合机制中得到启发,提出多粒度特征融合机制(Multi-granularity Feature Fusion, MGFF),用于融合多粒度文本特征信息。最后,在 5 个公共数据集上评估所提模型,证明了该模型的有效性。

本文的主要贡献如下:

- 1)提出了一种基于多粒度特征融合的新型图卷积网络模型,用于方面级情感分类任务。
- 2)提出将高速公路网络结合图卷积网络,以缓解深层 GCN 过平滑问题,从而获取更多粒度特征,为后面利用多粒度特征信息做铺垫。
- 3)根据不同深度 GCN 学习到的节点特征呈现粒度特点,将自适应特征融合机制用于融合不同深度 GCN 节点特征,得到包含更多粒度的文本特征表示向量。
- 4)在公共数据集上进行实验,实验结果证明了所提出的高速公路网络结合 GCN 以及多粒度特征融合机制模型的有效性。

2 相关工作

早期的文本情感分析大多是文档级、句子级情感分析,基于情感词典^[19]、SVM 等传统机器学习^[20]来实现。Mullen 等^[20]利用支持向量机(SVM)聚集各种来源的潜在相关信息,包括词汇、短语、情感词典等,提出了一个融合策略来整合特征并进行分类。然而,诸如此类的机器学习方法需要大量的人工处理数据,具有特征工程依赖性,且对样本数据的分布、噪声敏感。

随着深度学习的兴起,深度神经网络成为解决情感分析任务的主流方法,包括简单神经网络模型、基于注意力的神经网络模型以及基于图卷积神经网络模型。一些简单神经网络模型用于情感分析,例如 Tang 等^[21]提出将递归神经网络

(RNN)和 LSTM 应用于情感分类任务,开启了 LSTM 在情感分析领域的研究之门。由于注意力机制的广泛使用及其优越的效果,因此出现了融合单一网络和注意力机制的网络,例如 Tang 等^[6]通过精确捕获带有注意力权重的上下文信息,来预测给定方面的情感极性。Huang 等^[7]引入多层注意力机制来计算每个单词对方面词的重要性。Wang 等^[22]认为句子情感极性受到句子中方面词的影响,提出了基于注意力机制的 ATAE-LSTM 模型。Tay 等^[23]吸收 ATAE-LSTM 中注意力机制的想法,提出了方面融合层学习方面词与上下文关系。但这些计算注意力权重的方法容易因噪声数据或来自其他方面词的意见词而误判,于是, Hu 等^[24]改变计算注意力机制的方法,利用预训练模型 BERT 来学习句子和方面之间的深层关联以及解决句子中远距离依赖问题,确定意见词开始和结束的位置,并对这两个位置之间的词进行情感倾向判断。然而,预训练模型需要在大规模数据集上训练,且需要大量的计算资源和显存。而医疗保健、政治、经济等领域的情感分析任务对可解释性需求较大,预训练模型难以对模型的决策过程进行解释,不利于模型在这些领域的推广和应用。

近年来,图卷积网络受到广泛关注,它能很好地处理图结构数据,在自然语言处理各个任务中表现优异,如文本分类^[8]、机器翻译^[25]等。基于 Zhang 等^[9-10]的研究, Sun 等^[26]通过识别意见词的情感极性,并利用 GCN 进一步加强给定方面词的特征。考虑全局和局部上下文, Zhu 等^[27]提出了全局局部依赖图卷积网络。然而,这些方法忽略了给定方面与情感词的依赖关系。为了解决这个问题, Huang 等^[28]利用单词之间的依赖关系,从方面词的语法上下文传递情感特征信息。Wang 等^[29]在 GCN 中融合注意力机制,以关注给定方面词与情感词之间的关系。Liang 等^[30]基于公共知识库选择与方面词相关的词并计算其重要性。Yu 等^[31]设计了两种加权方法,提出了一种多权重图卷积网络,用于缓解注意力机制更偏向检测方面词相关语义以及忽略长距离依赖的问题。Tian 等^[32]将依赖关系矩阵与依赖类型权重矩阵送入 GCN,提出利用注意力集成层(ALE)学习关系类型来感知图卷积网络的不同层。Li 等^[33]结合 SenticNet 来加强句子依赖关系表示,进而改进 GCN 构建。

然而,大多数方法没有关注到不同深度 GCN 的粒度特征不同以及深度 GCN 存在的节点平滑问题,受 Pang 等^[14]、Tian 等^[32]的启发,本文认为不同深度 GCN 的节点特征信息呈现多粒度特征,因此提出了一种基于多粒度特征融合的高速公路图卷积网络模型(MGFF-HGCN)。该模型将高速公路网络与图卷积网络结合,缓解节点平滑问题,并获取更多粒度信息。然后,使用多粒度特征融合机制,充分利用不同深度图卷积网络层的节点信息,从而提高情感分类模型的性能。

3 模型描述

本章介绍了所提模型的各个层,其整体结构如图 2 所示。它主要由输入层、上下文信息编码层、高速公路图卷积网络层、多粒度特征融合层、特定方面掩码和注意力层、情感分类和训练层构成。

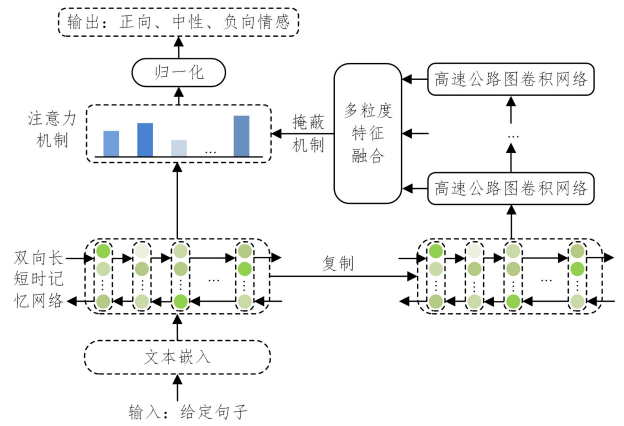


图 2 MGFF-HGCN 模型的整体框架

Fig. 2 Overall framework of MGFF-HGCN model

3.1 问题定义

给定长度为 n 的输入句子 $S = \{s_1, s_1, \dots, s_n\}$, 该句子包含预判断的 m 词方面类别 $a = \{s_{\tau+1}, s_{\tau+2}, \dots, s_{\tau+m}\}$, 以及情感倾向标签集合 $l \in T, T = [-1, 0, 1]$. 其中, 方面类别 a 是句子 S 中的一部分, 标签集合 T 中的 $-1, 0, 1$ 分别表示该方面词的消极、中性、积极情感。

3.2 输入和上下文信息编码层

在输入层,即文本嵌入层中,首先将每个输入词 s_i 映射到固定维度的低维向量 e_i 中,并构成向量矩阵 $E \in \mathbb{R}^{d_s \times |V|}$, 其中 $|V|$ 表示词汇量的大小, d_s 是词向量的维数。得到给定句子的文本嵌入后,将其送入 BiLSTM 来产生隐藏状态向量表示:

$$\mathbf{H}^b = \{h_1^b, h_2^b, \dots, h_{\tau+1}^b, \dots, h_{\tau+m}^b, \dots, h_{n-1}^b, h_n^b\} \quad (1)$$

其中, $h_i^b \in \mathbb{R}^{2d_h}$, d_h 表示单向 LSTM 输出的隐藏状态向量的维数。然后,将 \mathbf{H}^b 送入到下一层神经网络中。

3.3 高速公路图卷积网络构建层

在文本分类中,图卷积网络能够很好地捕获节点与相邻节点之间的语法信息。受 Zhang 等^[9]的工作的启发,对给定句子构造依赖树后有向图的形式保存。其中,假设每个单词为自身相邻,即在每个句子的依赖图矩阵中,对角线值都为 1。并输入 BiLSTM 得到隐藏状态向量,最后将其馈送到 GCN 中。将 GCN 中间的隐藏层简单写为以下非线性函数表现形式:

$$\mathbf{H}^{l+1} = f(\mathbf{H}^l, A) \quad (2)$$

其中,输入层 $\mathbf{H}^0 = \mathbf{H}^b$, l 是当前层层数, $f(\cdot, \cdot)$ 表示不同的聚合函数。

具体过程表示为:

$$\text{pos}_i^{l-1} = \mathcal{F}(h_i^{l-1}) \quad (3)$$

$$\tilde{h}_i^l = \sum_{j=1}^n \mathbf{A}_{ij} \mathbf{W}^l \text{pos}_j^{l-1} \quad (4)$$

$$h_i^l = \text{ReLU}(\tilde{h}_i^l / (d_i + 1) + b^l) \quad (5)$$

其中, $\mathcal{F}(\cdot)$ 表示分配位置权重的函数^[9]; h_i^{l-1} 表示第 $(l-1)$ 层 HGCN 输出节点信息,同时还可以表示第 l 层 HGCN 输入节点信息,此时的 $\tilde{h}_i^l \in \mathbb{R}^{2d_h}$ 表示为 HGCN 第 l 层的输出向量; $d_i = \sum_{j=1}^n \mathbf{A}_{ij}$ 表示依赖图中第 i 个节点的度数; 权重矩阵 \mathbf{W}^l 和偏置 b^l 是可训练参数。

高速公路网络^[12]允许部分信息不受阻碍地经过网络,其包含两个非线性转换层,分别是 Transformer 门和传输门, T 表示输入信息通过网络转换的部分,将 C 称为剩余部分输入单元。

$$y = H(x, \mathbf{W}_H) \cdot T(x, \mathbf{W}_T) + x \cdot (1 - T(x, \mathbf{W}_T)) \quad (6)$$

其中, $\mathbf{W}_H, \mathbf{W}_T$ 和 \mathbf{W}_C 是通过训练优化的参数矩阵。

本文将高速公路网络与 GCN 结合,如图 3 所示。该网络结构允许一部分输入特征经过 GCN 变换,另一部分输入特征无阻碍地流过每一层网络。对于经过变换后的输入特征,将 ReLU 作为激活函数。被卷积部分 $H(x, \mathbf{W}_H)$ 被替换为图卷积网络:

$$H(x, \mathbf{W}_H) \rightarrow f(\mathbf{H}^l, \mathbf{A}) \quad (7)$$

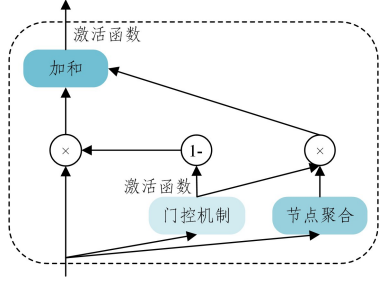


图 3 HGCN 中的卷积块

Fig. 3 Convolution blocks in HGCN

其具体传播过程表示为:

$$f(\mathbf{H}^l, \mathbf{A}) = \sum_{j=1}^n \mathbf{A}_{ij} \mathbf{W}^T \text{pos}_j^{l-1} \quad (8)$$

$$\tilde{\mathbf{h}}_i^l = \sum_{j=1}^n \mathbf{A}_{ij} \mathbf{W}^T \text{pos}_j^{l-1} \cdot T(x, \mathbf{W}_T) + \text{pos}_j^{l-1} \cdot (1 - T(x, \mathbf{W}_T)) \quad (9)$$

$$\mathbf{h}_i^l = \text{ReLU}(\tilde{\mathbf{h}}_i^l / (d_i + 1) + \mathbf{b}_i) \quad (10)$$

其中, $\mathbf{H}^l = \{\mathbf{h}_1^l, \mathbf{h}_2^l, \dots, \mathbf{h}_a^l, \dots, \mathbf{h}_{i+m}^l, \dots, \mathbf{h}_n^l\}$ 作为第 l 层 HGCN 的输出,保留每一层 HGCN 特征向量到向量 $\mathbf{H} = [\mathbf{H}^1, \mathbf{H}^2, \dots, \mathbf{H}^l]$ 中,并将其送入下一层网络。

3.4 多粒度特征融合层

不同深度 GCN 编码的上下文信息呈现多粒度特征,为了利用这一特点,本文提出了多粒度融合机制,全面地学习不同深度 GCN 输出的多粒度特征信息。

具体来说, $\mathbf{H}^{\text{Fusion}}$ 是不同层 HGCN 输出的特征向量的自适应加权融合,此时,每一层特征的重要性权重系数 α_i 被网络自适应地学习到。

$$\alpha_i = \frac{e^{w_i}}{\sum_{j=1}^l e^{w_j}}, i=1, 2, \dots, l; j=1, 2, \dots, l \quad (11)$$

其中, $\sum \alpha_i = 1, \alpha_i \in [0, 1], w_i$ 为初始化权重系数。然后,根据得到的权重系数计算融合后的特征向量 $\mathbf{H}^{\text{Fusion}}$:

$$\mathbf{H}^{\text{Fusion}} = \mathbf{H}^1 \times \alpha^1 + \mathbf{H}^2 \times \alpha^2 + \dots + \mathbf{H}^l \times \alpha^l \quad (12)$$

其中, \mathbf{H}^l 表示第 l 层 HGCN 获得的特征向量。

3.5 特定方面掩码和注意力层

为了更好地获得方面词的权重,我们屏蔽了非方面词的隐藏状态向量,得到只包含给定方面词的特征向量 $\mathbf{h}_{\text{mask}}^l = \{0, \dots, \mathbf{h}_{t+1}^l, \dots, \mathbf{h}_{t+m}^l, \dots, 0\}$,将其送入基于检索的注意力机制^[9],为每个上下文词提取方面词的注意力权重,得到的最终表示 $\mathbf{H}^{\text{Final}}$ 为:

$$\mathbf{h}_i^f = 0, 1 \leq i < \tau, \tau + m < i \leq n \quad (13)$$

$$\theta_j = \sum_{i=\tau+1}^{\tau+m} (\mathbf{h}_j^f)^\top \mathbf{h}_{\text{mask}_i}^f \quad (14)$$

$$\gamma_j = \frac{\exp(\theta_j)}{\sum_{i=1}^n \exp(\theta_i)} \quad (15)$$

$$\mathbf{H}^{\text{Final}} = \sum_{j=1}^n \gamma_j \mathbf{h}_j^f \quad (16)$$

其中, $\mathbf{h}_{\text{mask}}^f$ 表示多粒度特征融合后的只包含方面词的特征向量。然后,我们将最终表示 $\mathbf{H}^{\text{Final}}$ 输入到分类层。

3.6 情感分类和模型训练层

从特定方面掩码和注意力层得到表示向量 $\mathbf{H}^{\text{Final}}$,将其馈送到 softmax 归一化层进行情感分类:

$$p = \text{softmax}(\mathbf{W}_p \mathbf{H}^{\text{Final}} + \mathbf{b}_p) \quad (17)$$

其中, $\mathbf{p} \in \mathbb{R}^{d_p}$ 是情感极性概率分布, d_p 的维度与情感类别维度相同, $\mathbf{W}_p \in \mathbb{R}^{d_p \times 2d_h}$ 和 $\mathbf{b}_p \in \mathbb{R}^{d_p}$ 分别表示学习的权重和偏差。

该模型采用交叉熵损失和 L2 正则化的标准梯度下降算法进行训练和优化。

$$\text{Loss} = - \sum_{(d, \hat{p}) \in D} \log p_{\hat{p}} + \lambda \|\Theta\|_2 \quad (18)$$

其中, D 表示数据集的集合, $p_{\hat{p}}$ 表示预测值的第 \hat{p} 个元素, λ 表示 L2 正则化系数, Θ 表示所有可训练参数。

4 实验结果及分析

4.1 实验数据

实验基于 5 个广泛使用的 SemEval 数据集,其中包括由 Dong 等^[33] 基于 Twitter 帖子构建的数据集 LAP14 and REST14^[34], REST15^[35], REST16^[36],其基本内容是 twitter 文本。本文参照文献^[9]删除了 4 个极性冲突或没有明确指定方面词的数据。所有数据集的句子中包含句子、方面词、极性标签,其中情感极性标签分为积极、中性和消极。数据集信息如表 1 所列。

表 1 数据集统计数据

Table 1 Statistics of datasets

Datasets	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Twitter	1561	173	3127	346	1560	173
Laptop14	994	341	464	169	870	128
Restaurant14	2164	728	637	196	807	196
Restaurant15	912	326	36	34	256	182
Restaurant16	1240	469	69	30	439	117

4.2 实验设置和评估标准

在 MGFF-HGCN 模型中,我们将 300 维预训练 GloVe^[37] 向量初始化词嵌入作为该模型的文本嵌入。该模型权重初始化为均匀分布,文本嵌入和隐藏状态维度设置为 300 维,采用 Adam 优化器,学习率设置为 0.002,捕捉短距离情感词的浅层 GCN 层数设置为 2 或者 3,捕捉长距离情感词的 HGCN 层数设置为 4。为了避免过拟合,在 BiLSTM 层后设置 dropout 值为 0.3, BiLSTM 的隐藏向量维度为 300 维。模型在 NVIDIA RTX A5000, Intel(R) Core(TM) i7-10700 CPU 和 PyTorch 1.10.0 上进行训练。使用方面级情感分类任务中广泛使用的方法,即准确率 Acc 和 F1 值作为评价指

标来评判模型的最终分类效果。

4.3 基线模型

为了更好地说明模型的有效性,我们将所提算法与一些经典的模型进行比较。

1) 简单神经网络模型

SVM^[20]:将情感分类任务转化为多标签分类问题,每个分类器专注于分类一个情感标签。

LSTM^[21]:情感极性通过 LSTM 最后一个隐藏状态向量预测出来。

2) 融合注意力机制的模型

MemNet^[6]:将记忆力机制和 LSTM 网络相结合,实施不同的注意力策略,充分利用内容和位置注意学习上下文权重和文本表示。

AOA^[7]:将机器翻译中的 attention-over-attention 应用于情感分类任务。

IAN^[38]:上下文及方面词通过交互注意力网络,拼接从句子到方面的注意和从方面词到句子的注意。

TNet-LF^[39]:用注意力机制替代 CNN 网络,从双向 RNN 层产生变换后的单词表示中提取重要特征,并使用新机制保存其原始上下文信息。

3) 基于图卷积网络的模型

ASGCN^[9]:考虑情感分析中的语法和长距离词依赖问题,基于语法依赖树构建图卷积网络。

ASCNN^[9]:在 ASGCN 的基础上,用相同层数的 CNN

代替 2 层 GCN 得到的简化模型。

TD-GAT^[29]:通过图注意力网络来学习单词之间的依赖关系,通过 LSTM 学习方面词与上下文关系。

BiGCN^[10]:通过建立层次句法图和层次词汇图,聚合不同类型依赖关系和词汇词对。

MVGCN(Glove)^[31]:两种加权方法被提出,用于创建局部上下文加权邻接图。

DAGCN(Glove)^[40]:通过多头注意力和依赖感知语法信息学习依赖树中的相关信息,结构化地融合多个条件随机域,以更好地捕捉方面词及意见词。

4.4 实验结果

实验结果如表 2 所列,可以观察到 MGFF-HGCN 在这 5 个数据集中达到了最优或次优的 Acc 和 F1。根据表中的结果可知,大部分基于图卷积的模型优于简单神经网络模型,前人的实验结果也证明了大部分图卷积法依存树构建 GCN 的模型可以缓解部分长距离语法依赖问题。BiGCN 模型在 Twitter 及 Restaurant15 数据集上的性能优于大多数模型,说明词法图有助于更好地分类方面词情感。MWGCN 以及 DAGCN 模型的性能优于先前提到的模型。与给出的基准模型相比,所提模型在所给出的公共数据集上取得了良好的效果,且在部分数据集上优于基于图卷积网络的模型,展现了 MGFF-HGCN 模型融合不同深度 HGCN 节点特征信息的优越性,说明了加深 GCN 层次以及多粒度特征融合方法在提取到更多粒度特征上的有效性。

表 2 不同模型对比的实验结果

Table 2 Comparison results of different models

(单位:%)

模型分组	模型	数据集									
		Twitter		Laptop14		Restaurant14		Restaurant15		Restaurant16	
		Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
简单神经网络	SVM	63.40	63.30	70.49	—	80.16	—	—	—	—	—
	LSTM	69.56	67.70	69.28	63.09	78.13	67.47	77.37	55.17	86.80	63.88
融合注意力模型	MemNet	71.48	69.90	70.64	65.17	79.61	69.64	77.31	58.28	85.44	65.99
	AOA	72.30	70.20	72.62	67.52	79.97	70.42	78.17	57.02	87.50	66.21
	IAN	72.50	70.81	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21
	Tnet-LF	<u>72.98</u>	<u>71.43</u>	74.61	70.14	80.42	71.03	78.47	59.47	89.07	70.43
基于 GCN 模型	ASCNN	71.05	69.45	72.62	66.72	81.73	73.10	78.47	58.90	87.39	64.56
	ASGCN	72.15	70.4	75.55	71.05	80.77	72.02	79.89	61.89	88.99	67.48
	BiGCN	74.16	73.35	74.59	71.84	81.97	73.48	81.16	64.79	88.96	70.84
	MWGCN(Glove)	72.86	70.73	<u>76.73</u>	<u>72.28</u>	<u>82.56</u>	74.56	<u>80.39</u>	<u>63.91</u>	<u>89.22</u>	71.51
	DAGCN(Clove)	72.78	70.93	75.58	71.20	81.42	74.62	79.70	61.88	89.76	67.81
本文模型	MGFF-HGCN	72.68	70.74	76.85	72.96	82.61	75.20	80.25	63.46	89.12	<u>71.19</u>

4.5 消融实验及其分析

为了进一步检验所提模型中不同组件对模型性能的贡献,我们在给出的公共数据集上进行了消融实验,实验结果如表 3 所列。其中,加粗为最优结果,下划线为次优结果。

消融实验模型分别移除了高速公路网络模块(HGCN)以及多粒度特征融合模块(MGFF)。移除了多粒度特征融合的模块,将 HGCN 最后一层特征信息输入分类器,移除了高速公路图卷积网络的模块被替换为图卷积网络模块,分别比较基于 GCN 和 HGCN 在不同深度(取第 2—10 层)下对模型性能的影响。为了较好地展现模型层数的最优选择,表 3 列出了第 2—6 层的实验结果。更深层 GCN 模型在 Restaurant15

和 Restaurant16 数据集上的表现比 2 层 GCN 模型更好,说明对于包含更多复杂句式的数据集,更深层 GCN 可以展现出其优势和有效性。通过移除两个组件的实验结果表明,所提出的组件对该模型的良好性能都起到了不可否认的作用。

以往大多数研究表明,2 层 GCN 能实现更好的总体性能,但因数据集特征具有各异性(句子长度、句子中对方面词情感判断有作用的短语长度、长距离依赖等问题),所以不同数据集在不同 GCN 深度取得最优实验结果,实验结果符合我们的预期。根据表 3 中 GCN 和 HGCN 深度设置为 2—6 层的对比实验结果可知,基于普通 GCN 结合 MGFF 机制的模型,在大多数数据上深度为 2 层时能达到最好的效果。

4层 MGFF-HGCN 在所有数据集上表现的性能优于 2 层、3 层 MGFF-HGCN,且优于 2 层基于普通 GCN 结合 MGFF 机制的模型,说明更多层 GCN 模型能够获取到比 2 层、3 层 GCN 模型离方面词更远距离词的信息,缓解了情感分析任务

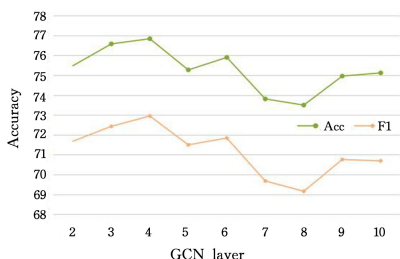
中长距离语法依赖问题。为了更直观地展示 4 层 MGFF-HGCN 模型的优越性,这里随机选择两个数据集,将 GCN 深度设置为 2—10 层,通过 Acc 和 F1 这两个评价指标来评估模型性能,结果如图 4 所示。

表 3 消融实验结果

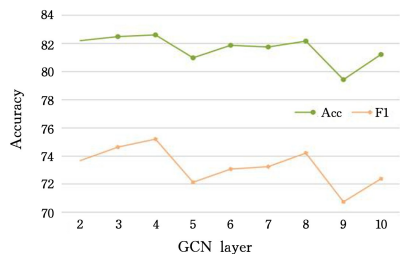
Table 3 Ablation experimental results

模型	Twitter		Laptop14		Restaurant14		Restaurant15		Restaurant16	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
原模型	72.15	70.40	75.55	71.05	80.77	72.02	79.52	61.39	88.99	67.48
HGCN	72.49	70.59	76.12	71.66	82.47	74.25	80.38	63.15	89.23	71.40
MGFF-GCN(2 layers)	72.68	70.64	75.80	71.21	81.78	73.04	79.95	62.82	88.74	68.47
MGFF-GCN(3 layers)	71.29	69.46	<u>75.07</u>	<u>70.44</u>	<u>82.14</u>	<u>73.83</u>	79.08	62.54	88.96	71.18
MGFF-GCN(4 layers)	<u>72.10</u>	<u>70.40</u>	71.83	66.30	81.63	73.09	79.82	61.66	88.79	69.01
MGFF-GCN(5 layers)	71.01	70.25	70.63	65.74	81.69	73.51	80.13	63.66	88.79	70.54
MGFF-GCN(6 layers)	71.62	69.76	70.54	65.68	80.41	70.49	79.70	<u>63.53</u>	<u>89.12</u>	<u>70.73</u>
MGFF-HGCN(2 layers)	71.91	70.27	75.49	71.69	82.20	73.67	79.58	63.98	88.85	70.08
MGFF-HGCN(3 layers)	<u>72.63</u>	70.61	<u>76.59</u>	<u>72.44</u>	<u>82.49</u>	<u>74.63</u>	<u>80.25</u>	63.25	88.96	<u>69.50</u>
MGFF-HGCN(4 layers)	72.68	<u>70.74</u>	76.85	72.96	82.61	75.20	80.25	63.46	<u>89.12</u>	71.19
MGFF-HGCN(5 layers)	72.63	70.87	75.28	71.51	80.98	72.13	79.58	64.19	88.36	67.35
MGFF-HGCN(6 layers)	71.82	70.07	75.91	71.85	81.87	73.07	79.58	62.81	89.03	69.46

(单位:%)



(a)Laptop14



(b)Restaurant14

图 4 MGFF-HGCN 模型在不同 GCN 深度下的性能表现

(Semeval14)

Fig. 4 Performance of MGFF-HGCN model at different GCN depths (Semeval14)

为了观察 HGCN 不同层对模型的贡献,图 5 给出了给每层 HGCN 分配权重值 α 以识别不同层 GCN 的贡献。其中,第 1—4 层 α 权重值分别用不同图案的条形图进行表示,分别对应 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ 权重值。通过得到的 α 结果可以看出,每一层网络对方面词的情感倾向判断都有一定的贡献,这符合我们的猜想,证明了 GCN 各层信息的贡献。通过统计每个数据集句子长度出现的频次发现,Restaurant14 数据集中较短长度句子占比频次远多于其他数据集,对于较短长度句子中方面词与情感词距离较近,2 层 HGCN 就能获取到有效信息。同时,从图 4 中可以发现,2 层 HGCN 与 4 层 HGCN 在 Restaurant14 数据集上的 Acc 差值比在 Laptop14 数据集上的相差更小,说明 2 层 HGCN 能获取到有效信息,但 3 层、

4 层 HGCN 能获取到更多粒度信息,比 2 层 HGCN 表现更好。因此,在 Restaurant14 数据集上,2 层 HGCN 对该数据集的贡献更突出。在 Restaurant15 和 Restaurant16 数据集中第 4 层的贡献优于或次优于其他层,验证了本文提到的这两个数据集包含比其他数据集更多的复杂句式,需要更深层次的 GCN。

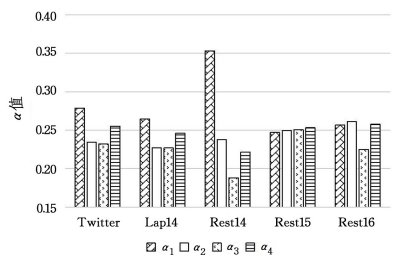


图 5 不同数据集上 α 值分布

Fig. 5 Distribution of α on different datasets

结束语 本文提出了一种基于多粒度特征融合的高速公路图卷积网络,用于方面级情感分析。该模型能够获取更远距离的句法信息,更加准确地对方面词进行情感分析。首先,将高速公路网络结合到 GCN 中,缓解深层 GCN 过平滑及长距离语法依赖问题。其次,构建多粒度特征融合机制,用于学习并融合不同深度 GCN 的粒度特征,并更加准确地预测情感倾向判断。最后,MGFF-HGCN 在几个公共数据集上进行了测试,与经典 GCN 模型进行了比较,MGFF-HGCN 取得了良好的效果,消融实验验证了模型中不同模块的有效性。在未来的研究中会尝试通过增强文本表示,如使用 BERT,来更好地动态获取给定句子句法依赖关系特征等,并将情感分析任务结果反被于构建知识图谱的部分模块中,辅助医疗决策及患者导诊。

参考文献

[1] D'ANIELLO G, GAETA M, LA ROCCA I. KnowMIS-ABSA:

- An overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis[J]. *Artificial Intelligence Review*, 2022, 55(7): 5543-5574.
- [2] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
 - [3] CHEN P, SUN Z Q, BING L D, et al. Recurrent Attention Network on Memory for Aspect-based Sentiment Analysis [C] // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2017: 452-461.
 - [4] YANN L, LÉON B, YOSHUA B, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324.
 - [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: NIPS, 2017: 6000-6010.
 - [6] TANG D Y, QIN B, LIU T. Aspect Level Sentiment Classification with Deep Memory Network [C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2016: 214-224.
 - [7] HUANG B X, OU Y, CARLEY K M. Aspect Level Sentiment Classification with Attention-over-attention Neural Networks [C] // International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation. Berlin: Springer, 2018: 197-206.
 - [8] LIANG Y, MAO C S, LUO Y. Graph Convolutional Networks for Text Classification [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 2019: 7370-7377.
 - [9] ZHANG C, LI Q C, SONG D W. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks [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). Stroudsburg, PA: ACL, 2019: 4567-4577.
 - [10] ZHANG M, QIAN T Y. Convolution over Hierarchical Syntactic and Lexical Graphs for Aspect Level Sentiment Analysis [C] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2020: 3540-3549.
 - [11] KIPF T N, WELING M. Semi-supervised Classification with Graph Convolutional Networks [J/OL]. <https://arxiv.org/abs/1609.02907>. pdf.
 - [12] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J. Highway Networks [J/OL]. <https://arxiv.org/pdf/1505.00387>. pdf.
 - [13] LI G H, MÜLLER M, QIAN G C, et al. Deepgens: making gens go as deep as cnns [J]. *arXiv*:1904.03751, 2019.
 - [14] PANG S G, YAN Z H, HUANG W H, et al. Highway-based Local Graph Convolution Network for Aspect-based Sentiment Analysis [C] // Natural Language Processing and Chinese Computing. Berlin: Springer, 2021: 544-556.
 - [15] PEDRYCZ W. Granular Computing: an Introduction [C] // Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569). Piscataway: IEEE, 2001: 1349-1354.
 - [16] YAO Y Y. A triarchic theory of granular computing [J]. *Granular Computing*, 2016(1): 145-157.
 - [17] WANG G Y, HU J, ZHANG Q H, et al. Granular Computing based Data Mining in the Views of Rough Set and Fuzzy Set [C] // 2008 IEEE International Conference on Granular Computing. Piscataway: IEEE, 2008: 67-67.
 - [18] LIU S T, HUANG D, WANG Y H. Learning Spatial Fusion for Single-shot Object Detection [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2020: 4016-4025.
 - [19] LIU S M, CHEN J H. A multi-label classification based approach for sentiment classification [J]. *Expert Systems with Applications*, 2015, 42(3): 1083-1093.
 - [20] MULLEN T, COLLIER N. Sentiment Analysis using Support Vector Machines with Diverse Information Sources [C] // Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP). Barcelona, Spain: ACL, 2004: 412-418.
 - [21] TANG D Y, QIN B, FENG X C, et al. Effective Lstms for Target-dependent Sentiment Classification [C] // Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. Berlin: COLING, 2016: 3298-3307.
 - [22] WANG Y Q, HUANG M L, ZHU X Y, et al. Attention-based Lstm for Aspect-level Sentiment Classification [C] // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2016: 606-615.
 - [23] TAY Y, TUAN L A, HUI S C. Learning to Attend Via Word-aspect Associative Fusion for Aspect-based Sentiment Analysis [C] // Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park, CA: AAAI, 2018: 5956-5963.
 - [24] HU M T, ZHAO S W, GUO H L, et al. Learning to Detect Opinion Snippet for Aspect-based Sentiment Analysis [C] // Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). Stroudsburg, PA: ACL, 2019: 970-979.
 - [25] BECK D, HAFFARI G, COHN T. Graph-to-sequence Learning Using Gated Graph Neural Networks [C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018: 273-283.
 - [26] SUN K, ZHANG R C, MENSAH S, et al. Aspect-level Sentiment Analysis Via Convolution over Dependency Tree [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). Stroudsburg, PA: ACL, 2019: 5679-5688.
 - [27] ZHU X F, ZHU L, GUO J F, et al. Gl-gcn: Global and local dependency guided graph convolutional networks for aspect-based sentiment classification [J]. *Expert Systems with Applications*, 2021, 186: 115712.
 - [28] HUANG B X, CARLEY K. Syntax-aware Aspect Level Senti-

- ment Classification with Graph Attention Networks[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). Stroudsburg, PA: ACL, 2019; 5469-5477.
- [29] WANG K, SHEN W Z, YANG Y Y, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2020; 3229-3238.
- [30] LIANG B, SU H, YIN R D, et al. Beta Distribution Guided Aspect-aware Graph for Aspect Category Sentiment Analysis with Affective Knowledge[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2021; 208-218.
- [31] YU B G, ZHANG S W. A novel weight-oriented graph convolutional network for aspect-based sentiment analysis[J]. The Journal of Supercomputing, 2023, 79(1): 947-972.
- [32] TIAN Y H, CHEN G M, SONG Y. Aspect-based Sentiment Analysis with Type-aware Graph Convolutional Networks and Layer Ensemble[C]// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2021; 2910-2922.
- [33] LI S, XU B, HAN Y K, et al. SS-GCN: Aspect-based Sentiment Analysis with Graph Convolutional Networks over Dependency Awareness[J]. Computer Science, 2023, 50(3): 3-11.
- [34] DONG L, WEI F R, TAN C Q, et al. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification [C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2014; 49-54.
- [35] PONTIKI M, GALANIS D, PAVLOPOULOS J, et al. SemEval-2014 Task 4: Aspect based Sentiment Analysis[C]// Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Stroudsburg, PA: ACL, 2014; 27-35.
- [36] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2015 Task 12: Aspect based Sentiment Analysis[C]// Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Stroudsburg, PA: ACL, 2015; 486-495.
- [37] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 Task 5: Aspect based Sentiment Analysis[C]// Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Stroudsburg, PA: ACL, 2016; 9-30.
- [38] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global Vectors for Word Representation[C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Stroudsburg, PA: ACL, 2014; 1532-1543.
- [39] LI X, BING L D, LAM W, et al. Transformation Networks for Target-oriented Sentiment Classification [C]// Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: ACL, 2018; 946-956.
- [40] WANG X, LIU P, ZHU Z, et al. Aspect-based Sentiment Analysis with Graph Convolutional Networks over Dependency Awareness[C]// 2022 26th International Conference on Pattern Recognition. Piscataway: IEEE, 2022; 2238-2245.



DENG Ruhan, born in 1999, postgraduate. Her main research interests include granular computing and natural language processing.



ZHANG Qinghua, born in 1974, Ph.D., professor, Ph.D supervisor. His main research interests include rough sets, fuzzy sets, granular computing and uncertain information processing.

(责任编辑:何杨)