

融合句法距离与方面注意力的方面级情感分析

张隆基, 赵晖

引用本文

张隆基, 赵晖. 融合句法距离与方面注意力的方面级情感分析[J]. 计算机科学, 2023, 50(12): 262-269.

ZHANG Longji, ZHAO Hui. [Aspect-level Sentiment Analysis Integrating Syntactic Distance and Aspect-attention](#) [J]. Computer Science, 2023, 50(12): 262-269.

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)

[融合多头注意力机制和孪生网络的语义匹配方法](#)

Semantic Matching Method Integrating Multi-head Attention Mechanism and Siamese Network

计算机科学, 2023, 50(12): 294-301. <https://doi.org/10.11896/jsjcx.221000083>

[SemFA:基于语义特征与关联注意力的大规模多标签文本分类模型](#)

SemFA:Extreme Multi-label Text Classification Model Based on Semantic Features and Association Attention

计算机科学, 2023, 50(12): 270-278. <https://doi.org/10.11896/jsjcx.230300239>

[基于可信细粒度对齐的多模态方面级情感分析](#)

Aspect-based Multimodal Sentiment Analysis Based on Trusted Fine-grained Alignment

计算机科学, 2023, 50(12): 246-254. <https://doi.org/10.11896/jsjcx.221100038>

[一种融合CNN和Swin Transformer的医学显微图像分割模型](#)

Medical Microscopic Image Segmentation Model Based on CNN Structure and Swin Transformer

计算机科学, 2023, 50(11A): 230200119-8. <https://doi.org/10.11896/jsjcx.230200119>

[基于注意力机制和ConvLSTM的船舶交通流量预测算法](#)

Ship Traffic Flow Prediction Algorithm Based on Attention Mechanism and ConvLSTM

计算机科学, 2023, 50(11A): 230800067-7. <https://doi.org/10.11896/jsjcx.230800067>

融合句法距离与方面注意力的方面级情感分析

张隆基¹ 赵 晖²

1 新疆大学软件学院 乌鲁木齐 830000

2 新疆大学信息科学与工程学院 乌鲁木齐 830000

(13009626824@163.com)

摘要 目前,基于句法依存树的图卷积网络面临着卷积层数过深而产生过平滑的问题,无法提取句法依存树的全局节点信息。虽然搭配序列模型可以提取到语句的上下文的信息,但是序列模型依赖时序的特点导致图卷积网络无法有效地区分上下文特征对方面项的贡献度。针对上述问题,提出了一种基于句法距离和方面关注注意力机制的新型图卷积网络模型。首先,该模型利用双向长短期记忆网络分别学习语句和方面项的上下文信息,同时结合图卷积网络学习语句的句法依存信息。其次,依据句法依存树计算所有节点之间的句法依存距离,设定阈值削弱长距离特征的权重占比,提高图卷积模型区分上下文特征的能力。最后,设计具有残差连接的注意力机制,指导方面项自动聚焦于语句中的重要信息。实验结果表明,相较于基线方法,所提模型在多个公开数据集上展现出了较好的分析性能,在 Twitter 数据集和 Laptop 数据集上的情感分类准确率分别高达 75.94% 和 78.59%,表明了所提方法的有效性。

关键词: 图卷积网络;句法依存树;句法依存距离;注意力机制;方面级情感分析

中图法分类号 TP391

Aspect-level Sentiment Analysis Integrating Syntactic Distance and Aspect-attention

ZHANG Longji¹ and ZHAO Hui²

1 College of Software, Xinjiang University, Urumqi 830000, China

2 School of Information Science and Engineering, Xinjiang University, Urumqi 830000, China

Abstract Currently, the over-smoothing problem arises from deep convolution in syntactic dependency tree-based graph convolutional networks. This problem prevents the convolutional graph network from extracting the global node information of the syntactic dependency tree. Although the sequential model can extract information about the context of the sentence, the timing-dependent nature of the sequential model leads to the inability of the graph convolutional network to effectively distinguish the contribution of context features to aspect terms. This paper proposes a novel graph convolutional network model based on syntactic distance and aspect focus attention mechanisms to address the above problems. First, the model learns the contextual information of sentences and aspect terms separately using a bidirectional long short-term memory network and uses a convolutional graph network to learn the syntactic dependency information of sentences. Secondly, this model calculates the syntactic dependency distance among all nodes based on the syntactic dependency tree, sets a threshold to weaken the weight share of long-distance features, and improves the ability of the graph convolution model to distinguish context features. Finally, this paper also designs attention mechanisms with residual connectivity to automatically guide the aspect terms to focus on the critical information in the sentence. Experimental results demonstrate that the model exhibits better analytical performance on several publicly available datasets compared to the baseline approach, with sentiment classification accuracy as high as 75.94% and 78.59% on the Twitter and Laptop datasets, demonstrating the effectiveness of the proposed approach.

Keywords Graph convolutional networks, Syntactic dependency tree, Syntactic dependency distances, Attention mechanisms, Aspect-level sentiment analysis

到稿日期:2022-10-12 返修日期:2023-03-15

基金项目:国家自然科学基金(62166041)

This work was supported by the National Natural Science Foundation of China(62166041).

通信作者:赵晖(zhaohui@xju.edu.cn)

1 引言

方面级情感分类 (Aspect-level Sentiment Classification, ALS) 是一种细粒度情感分类任务,在电商口碑分析、舆情监控、内容营销^[1-3]等场景中创造出了极高的商业价值。ALS 的研究对象对应于语句中的某个单词或单词组,也被称为方面项。相较于篇章级情感分析和语句级情感分析,它提供了更为精确的情感定位。比如例句 1 “The food was well prepared, but the service is very poor” 中,方面项 “food” 对应的情感极性是 “积极”,因为语句表达了对 “food” 的积极看法;方面项 “service” 对应的情感极性是 “消极”,因为语句表达了对 “service” 的消极看法。同一语句中两个方面项却蕴含了完全相反的情感极性,因此,方面情感分类的难点在于如何区分上下文特征对特定方面项的贡献度,以及排除无关情感信息对特定方面项的干扰。

早期,序列模型常被用来建模语句的上下文信息。循环神经网络^[4] (Recurrent Neural Networks, RNN) 基于序列位置链式提取上下文信息,却无法支持长序列特征记忆。长短期记忆网络^[5] (Long Short-Term Memory, LSTM) 和门控循环单元网络^[6] (Gate Recurrent Unit, GRU) 设计独特的门控单元结合短期记忆与长期记忆,却依旧无法摆脱时序性限制,缺乏区分不同上下文单词重要性的能力。正如例句 1 中方面项 “service” 更应该关注 “poor” 而不是 “prepared”, 序列模型受限于串联结构,会不可避免地将 “prepared” 的信息传递给方面项 “service”, 导致模型给出错误的分类结果。虽然众多研究者引入注意力机制以区分句中单词对方面项的重要性,但是这类方法都忽略了对句法依存信息的利用,导致注意力机制错误地引导模型给予不相关单词过高的关注度^[7]。因此,搭配句法依存树的图卷积网络^[8-9] (Graph Convolutional Networks, GCN) 被用来提取句中单词之间的句法依存信息。首先,基于传统句法树的集成模型较为经典,该类模型联合学习句子上下文信息和单词之间的句法信息,缓解了序列模型面临的困境。其次,传统的 GCN 模型即使借助序列模型读取到句子的上下文信息,却依然不合理地将所有邻居节点放在同等的地位。为此,一些研究采用信息增强的手段提高 GCN 模型对全局节点的敏感度,另外一些研究利用图注意力 (Graph Attention Networks, GAT) 模型替换 GCN 模型,通过多头节点拼接的方式计算节点之间的相关性系数^[10-11]。

然而,基于信息增强的手段都是针对外部全局单词进行改变,缺乏对句法树本身结构特性的充分利用。基于图注意力网络的模型虽然起到了区分邻接节点贡献度的作用,但是其采用多头注意力拼接的方式,只用到了二阶邻居节点,当卷积高阶邻居节点时,往往会产生过平滑的问题。文献^[12]虽然同时考虑了句法依存树结构和单词的句法距离,但其对卷积后的上下文表示执行权重矩阵相乘的操作,更类似于注意力机制对上下文信息进行加权优化,反而没有考虑到卷积过程中邻接矩阵对 GCN 卷积范围的限制。

针对上述问题,本文提出了一种结合句法距离和方面关注注意力机制的新型图卷积神经网络模型,用于打破 GCN

模型无法有效提取依存树全局节点信息的困境。一方面,考虑到卷积次数限制了传统 GCN 模型提取全局特征的能力,该模型借助句法距离划分全局节点的分配比例,并通过阈值削弱长距离无关项对最终输出的干扰。另一方面,该模型在设计方面关注注意力机制、学习方面项和上下文单词的交互作用,进一步提高模型识别上下文单词重要性的能力。

本文的主要贡献如下:

1) 构建基于阈值限定的单词句法距离矩阵,在有限的卷积层次内,帮助 GCN 模型有区别地提取依存树全局节点信息,同时降低无关特征的干扰。

2) 分别建模语句和方面项,并设计了基于方面关注的注意力机制,使得方面项着重关注语句中与之相关的情感特征,进一步增加模型处理复杂情况的表达能力。

3) 在 3 个公开数据集上评估模型的性能,实验结果表明,该模型对不同数据集的准确率与 F1 值均有所提升,进一步证明了所提方法的有效性。

2 相关工作

2.1 句法依存树

传统序列模型依赖句中单词的时间顺序,最终的输出特征包含了整个时间序列的特征信息。为了控制不同时间片对最后输出的信息流入,Tang 等^[13]和 Wang 等^[14]将方面项添加到 LSTM 模型的训练过程中,用以区分不同时间片的上下文单词对特定方面的影响。Ma 等^[15]则认为方面项与上下文单词享有同等级别,并设计了注意力机制交互学习方面项和上下文特征之间的关系。Chen 等^[16]考虑到情感特征距离方面项较远的情况,设计多层注意力机制捕获长距离情感特征。Fan 等^[17]设计了细粒度的注意力机制,避免在多单词组成方面项的情况下出现信息流失。Xing 等^[18]则在 LSTM 的门控单元内部增加方面项过滤机制,使得模型保留方面项相关的信息,进一步去除噪音的影响。在考虑方面项的前提下,Gu 等^[19]进一步表明序列位置信息有助于区分方面项上下文的重要性,距离方面项较近的单词应享有较高的权重,距离方面项较远的单词则分配较低的权重。例如图 1(a) 中句子 “It’s a perfect place to have an amazing indian food”, 方面项是 “Indian food”, 对应的情感极性是 “积极”。观察句子时间顺序,距离方面项较近的单词 “amazing” 对分析方面项的情感极性非常重要,距离方面项较远的单词 “to” 和 “have” 对于分析方面项的情感极性贡献微弱。然而,序列位置信息并不适用于所有语句,尤其是在情感词距离方面项较远的情况下,模型将错误地为情感词分配较低的权重。例如图 1(b) 中句子 “amplifier you gave me played music so loudly” 的方面项是 “amplifier”, 对应的情感极性是 “消极”。情感词 “loudly” 距离方面项较远,但对分析方面项的情感极性贡献较大,“you” 和 “gave” 等距离方面项较近的单词反而贡献较小。随着句法依存信息的引入,句法依存树较好地缓解了序列位置信息的尴尬局面。如图 1 所示,节点之间的曲线代表方面词与情感词之间的依存距离,也就是 $n:hot$, n 代表获取信息所需跨越的节点数。

图 1(a) 中方面项“Indian food”距离情感词“amazing”的序列距离为 1, 句法依存距离也为 1。图 1(b) 中方面项“amplifier”与情感词“loudly”的序列距离为 7, 句法依存距离仅为 2。由此可见, 句法依存树不仅解决了情感词距离方面项较远且权重较小的问题, 同时也保留了近距离单词与方面项的紧密性。

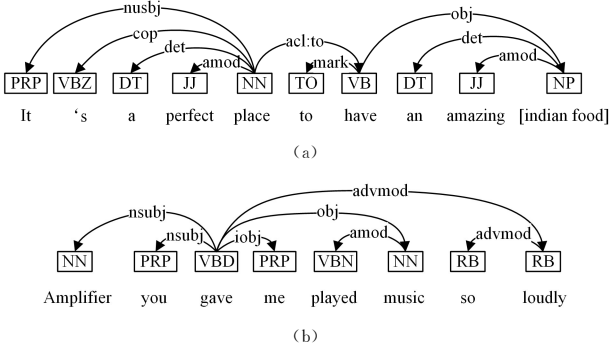


图 1 序列信息与句法依存树

Fig. 1 Sequence information and syntactic dependency trees

2.2 图卷积网络

区别于处理欧氏空间数据的传统神经网络, GCN 是一种专注于处理图结构的神经网络, 也被称为基于度感知的消息传递网络^[20] (Message Passing Neural Networks, MPNN)。GCN 模型的卷积算法具有并行卷积节点信息的优势, 并借助多个卷积层达到信息连续传递的目的。其信息传递算法设计如下:

$$\mathbf{h}^i = \sigma(c_i \sum_{j=1}^n \mathbf{A}_{ij} (\mathbf{W}^{l-1} \mathbf{h}^{j-1} + \mathbf{b}^{l-1})) \quad (1)$$

其中, c_i 代表节点 i 度数的倒数; $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是由依存树生成的邻接矩阵, $\mathbf{A}_{ij} = 0$ 表示节点 i 和节点 j 之间无连接边, $\mathbf{A}_{ij} = 1$ 表示二者之间有连接边; \mathbf{W}^{l-1} 和 \mathbf{b}^{l-1} 分别代表权重矩阵和偏置, σ 代表 relu 非线性激活函数。

在方面情感分析任务中, 基于 GCN 的一系列模型在提取句法依存信息的任务中应用广泛。比如, Sun 等^[21] 基于句法依存树提出 GCN 和 LSTM 的集成模型, 联合学习语句上下文信息句法依存信息; Zhang 等^[22] 利用句法依存信息约束注意力机制关注方面相关的情感信息; Lu 等^[23] 利用方面项指导句法信息编码, 设计了基于门控机制的图卷积神经网络; Chen 等^[24] 基于大脑认知原理学习语句信息和句法信息。此外, 为了增强单词的特征表示, Chen 等^[25] 利用不同的注意力机制生成潜在图结构, 弥补单一句法依存树的信息缺失问题。Zhang 等^[26] 则利用单词共现率构建单词共现矩阵, 增加节点特征之间的全局注意力。Pang 等^[27] 优化融合句法信息和语义信息, 从而弥补句子缺乏显著句法结构的缺陷。Li 等^[28] 通过精化句法解析的方式来提高句法信息提取的准确度。

3 模型构建

本文提出的 SA-GCN 模型如图 2 所示, SA-GCN 主要由输入层、Bi-LSTM 层、GCN 层、注意力层、方面掩码层和输出层 6 部分组成, 模型的每部分将在本章分别介绍。

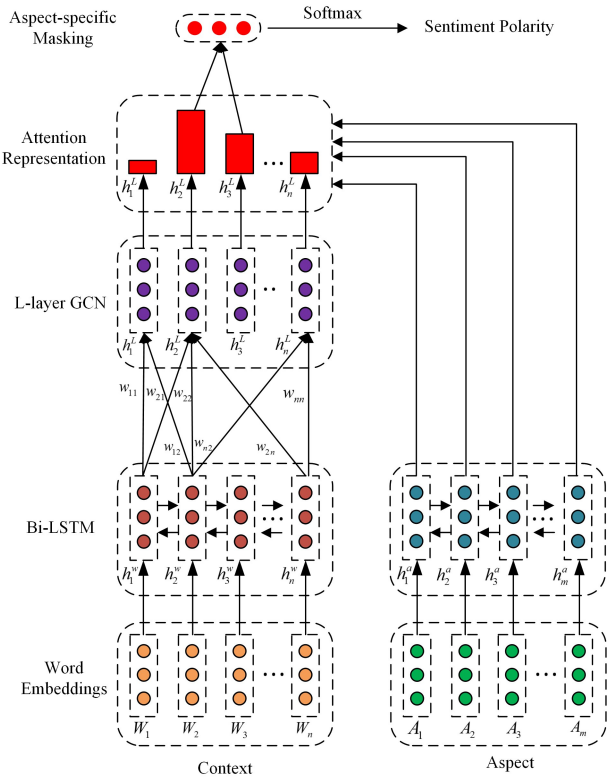


图 2 SA-GCN 模型结构

Fig. 2 Structure of SA-GCN model

3.1 输入层与 Bi-LSTM 层

给定长度为 n 的语句, 其嵌入向量表示为 $\mathbf{Sentence} = \{\omega_1, \omega_2, \dots, \omega_n\}$, 包含所有单词的嵌入表示, 其中 $\mathbf{Aspect} = \{a_1, a_2, \dots, a_m\}$ 是句向量中长度为 m 的方面项子序列。本文采用 Glove 预处理词向量提供的嵌入矩阵 $\mathbf{G} \in \mathbb{R}^{|\mathcal{V}| \times d_e}$ 将文本特征转化为可供神经网络训练的低维稠密特征向量, 其中 $|\mathcal{V}|$ 表示语料库词表的大小, d_e 表示词嵌入的维度。随后, 本文利用 Bi-LSTM 分别提取语句和方面项的时间序列信息, 将语句特征和方面项嵌入特征分别作为 Bi-LSTM 的输入, 得到隐藏状态向量 \mathbf{h}_s 和 \mathbf{h}_A , 其中 $\mathbf{h}_s \in \mathbb{R}^{n \times d_w}$, $\mathbf{h}_A \in \mathbb{R}^{m \times d_w}$, d_w 为双向 LSTM 的输出维度。

3.2 GCN 层

本文将语句特征经过 Bi-LSTM 模型得到的输出向量 \mathbf{h}_s 作为 GCN 模型的输入, 进一步提取语句中单词之间的句法关系。传统的 GCN 模型依赖邻接矩阵仅能获取有限范围内的节点特征, 使得模型无法充分提取对方面项极为重要的情感信息。基于此, 本文以单词之间的句法距离矩阵为媒介, 在卷积过程中帮助 GCN 模型提取依存树全节点信息, 并设定阈值来削弱无效特征的干扰。

本文采用深度优先搜索算法计算当前单词到树中其他单词之间的句法距离, 如算法 1 所示。算法的输入为无向图, 输出为当前节点与图中其他节点的句法依存距离。给定长度为 n 的语句, 通过深度优先搜索算法生成维度为 $n \times n$ 的句法依存距离矩阵。第 2-9 行表示循环访问无向图中相连节点的过程, 第 4 行代表距离计算算法, 假设当前节点为 v_i , 访问无向图中的邻接节点 v_j , 此时 $depth$ 记为 1。若 v_j 在无向图中存在 v_i 之外的邻接节点 v_k , 则第 5 行递归调用算法更新 v_i 与

v_k 之间的距离,即 $depth(v_i, v_k) = 2$;若节点 v_k 仍有子节点连接,则一直访问到该分支的叶子节点为止,否则返回双亲节点继续搜索。值得注意的是,本文在计算距离时设定了阈值边界 t ,依存树上并非所有节点都对当前节点起到辅助作用,距离当前节点较远距离的节点可能对当前节点毫无贡献,甚至会产生噪音干扰。

算法 1 Dfs

Input: Graph = $\{v_i | (v_1, v_2, \dots, v_j)\}$

Output: Dependency distance vector

```

1. Void dfs(nodes, parent=None):
2. For child in Graph(nodes) do
3.   If child is not parent then
4.     depth ← distance(child, nodes)
5.     dfs(child, nodes)
6.   Else
7.     continue
8.   End if
9. End for
10. Return depth

```

因此,一旦句法距离超过阈值,便将句法距离设为定值,限定长距离节点的权重占比。句法距离计算公式如下:

$$D_i = \begin{cases} D_{ij}, & \text{if } D_{ij} < t \\ t, & \text{if } D_{ij} \geq t \end{cases} \quad (2)$$

其中, $D_i \in R^n$ 代表节点 i 到依存树中其他节点的句法依存距离。然后,根据距离矩阵计算不同节点之间的相关性系数,计算公式如下:

$$S = 1 - \frac{D}{t} \quad (3)$$

其中, $D \in R^{n \times n}$ 表示节点之间的句法依存距离矩阵。 $t \in R^+$ 代表阈值,为本文设置的超参数。 $t < \max(D_{ij})$, 保证阈值在节点最大距离范围之内。本文将依存距离转换为范围在 $[0, 1]$ 内的权值,当其他节点与当前节点的句法距离超过阈值范围时,权值记为 0,表示该节点对当前节点无贡献,最大程度降低无效特征的干扰。由此,计算得到的分数矩阵 $S \in R^{n \times n}$,其中 S_{ij} 表示单词 j 对于单词 i 的相关性分数。即在句法依存关系上,距离方面项越近的单词享有越高的分数,距离方面项越远的单词享有越低的权重。

最后,本文利用分数矩阵区分不同单词对目标单词的贡献度,并迭代参与 GCN 的卷积运算,最终得到上下文输出向量 $h_G \in R^{n \times d_h}$ 。以上下文单词 i 为例,其第 $l+1$ 层特征向量的计算式如下:

$$h_i^{(l+1)} = \sigma(c_i \sum_{j=1}^n W^{(l)} A_{ij} (S_{ij} \cdot h_j^{(l)}) + b^{(l)}) \quad (4)$$

其中, $h_i^{(l+1)} \in R^{d_h}$ 代表第 $l+1$ 层上下文单词 i 的隐藏状态输出, $h_j^{(l)} \in R^{d_h}$ 代表第 l 层上下文单词 j 的隐藏状态向量。 A_{ij} 代表邻接矩阵元素, S_{ij} 代表分数矩阵元素, $W^{(l)}$ 和 $b^{(l)}$ 分别表示权重矩阵和偏置。

3.3 注意力层

尽管句法距离能够规避长距离无效特征对最终输出的干扰,方面项却无法合理地区分相同句法距离的有效特征和无效特征。因此,本文设计了方面关注注意力机制,首先利用

双向 LSTM 学习方面项的特征表示,与基于句法信息的句子上下文相交互。其次,分别从方面项和上下文的角度出发,计算彼此的相关性系数。最终,该模块采用残差机制增强网络的表达能力。注意力机制的细节详情如图 3 所示,将经过 GCN 网络的语句上下文隐藏层输出与经过 Bi-LSTM 的方面项隐藏层输出相乘,得到二者的交互矩阵。计算式如下:

$$I = h_G \cdot h_A^T \quad (5)$$

其中, h_G 代表语句上下文的隐藏层输出, h_A 代表方面项的隐藏层输出。

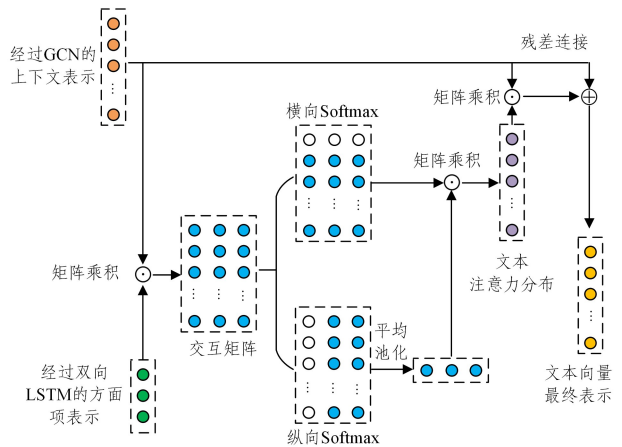


图 3 注意力模型结构

Fig. 3 Structure of attention model

然后,利用 softmax 函数计算不同角度下交互矩阵中上下文单词和方面项之间的注意力系数,分别得到上下文对方面项的注意力矩阵 $M \in R^{n \times m}$ 和方面项对上下文的注意力矩阵 $N \in R^{n \times m}$,计算式如下:

$$M_{ij} = \frac{\exp(I_{ij})}{\sum_i \exp(I_{ij})} \quad (6)$$

$$N_{ij} = \frac{\exp(I_{ij})}{\sum_j \exp(I_{ij})} \quad (7)$$

其中, M_{ij} 表示上下文单词与方面项中每个单词的相关性系数, N_{ij} 表示方面项与上下文中每个单词的相关性系数。

然后,利用平均池对注意力矩阵 N 进行处理,计算方面项中每个方面词对整个上下文文本的注意力值 $\bar{N} \in R^m$,计算式如下:

$$\bar{N} = \frac{1}{n} \left(\sum_{i=1}^n N_i \right) \quad (8)$$

将得到的结果与上下文对方面项的注意力矩阵 M 相乘,计算整个文本的上下文注意力向量 $\bar{M} \in R^n$,计算式如下:

$$\bar{M} = M \cdot \bar{N}^T \quad (9)$$

最后,将最终的注意力向量与上下文特征相乘,并借助残差机制计算上下文输出向量,计算式如下:

$$\bar{h}_G = \bar{M}^T h_G + h_G \quad (10)$$

其中, $h_G \in R^{n \times d_h}$ 代表 GCN 模型的上下文输出向量, $\bar{h}_G \in R^{n \times d_h}$ 代表基于注意力机制的上下文向量表示。

3.4 掩码层

本文采用方面掩码机制削弱无效特征对最终输出的

干扰,保留方面词本身的最终特征表示,公式如下:

$$mask_i = \begin{cases} 1, & \text{if } w_i \in \mathbf{Aspect} \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$h_F = \frac{1}{m} (\sum mask \cdot \bar{h}_G) \quad (12)$$

其中, $mask \in R^n$ 代表掩码序列,如果单词 i 位于方面项之内,那么对应的 $mask_i$ 等于 1,否则为 0. $h_F \in R^{d_h}$ 代表方面项特征向量之和,将其作为最终输出。

3.5 输出层

本文将获得的最终输出 h_F 输入到全连接层,计算式如下:

$$f = W_F \cdot h_F + b_F \quad (13)$$

其中, W_F 和 b_F 分别代表权重矩阵和偏置, f 为全连接层输出。最后利用 softmax 函数对线性层输出情感分类,公式如下:

$$P(y=c) = \frac{\exp(f_c)}{\sum_{i \in C} \exp(f_i)} \quad (14)$$

其中, $c \in C$ 代表方面项的情感标签, C 代表情感极性决策空间。

最后,模型采用交叉熵损失函数进行训练,公式如下:

$$loss = - \sum_{i \in C} I(y_i = c) \log(P(y_i = c)) \quad (15)$$

其中, $I(\cdot)$ 是指示函数,表示真实概率分布。

4 实验

4.1 数据集与实验平台

为了验证模型的有效性,本文在 3 个公开数据集上进行了实验,分别是 Twitter 数据集^[29]、semval2014^[30] 的 Laptop 数据集和 Restaurant 数据集。这些数据集包含 3 类标签,分别是积极(Positive)、中性(Neutral)和消极(Negative),数据集内部情感极性分类细节如表 1 所列。本文采用的实验平台如表 2 所列,实验主要采用 Pytorch1.1 深度学习框架,使用 python3.8 实现。

表 1 数据集分类

Table 1 Dataset classification

数据集	积极		中性		消极	
	train	test	train	test	train	test
Twitter	1507	172	3016	336	1528	169
Laptop	976	337	455	167	851	128
Restaurant	2164	727	637	196	807	196

表 2 实验平台信息

Table 2 Experimental platform information

实验环境	具体信息	
操作系统	Windows 10 家庭中文版	
CPU/GHz	Inter(R) Core(TM) i7-4710MQ CPU @ 2.5 2.50	
内存/GB	12.0	
显卡	GTX 950	
显存/GB	4.0	

4.2 参数设置

本文采用 300 维的 Glove 词向量初始化词嵌入,模型所有的权重参数采用均匀分布初始化。其中,词嵌入特征维度为 300,序列位置嵌入特征维度为 30,词性嵌入特征维度为

30. Bi-LSTM 模型隐藏层维度为 100,GCN 模型隐藏层维度为 50。输入层特征 dropout 为 0.7,每层隐藏层随机丢弃 10% 的神经元。模型采用 Adam 优化器,学习率为 0.01,epoch 为 100,训练批次为 32。经过实验测试,GCN 卷次层数为 2,阈值为 4 时,模型效果最佳。

实验结果取随机化运行 10 次的平均值,并采用分类准确率 Acc 和 F1 值作为评价指标,比较模型在不同数据集上的效果优劣。

4.3 对比模型

为了更好地评估模型性能,本文将其与一系列基准模型进行比较,各个模型的简要介绍如下。

TD-LSTM:以方面项为界,利用 LSTM 分别建模方面项前后上下文特征。

ATAE-LSTM:在 LSTM 序列建模的基础上,设计注意力机制计算方面项和不同上下文的权重。

IAN:对方面项和上下文特征分别建模,设计交互注意力机制交互学习二者的特征表示。

RAM:利用多重注意力机制捕获长距离情感特征。

AA-LSTM:将方面项加入到 LSTM 的门控单元中,动态过滤与方面项无关的信息,保留与方面项相关的信息。

CDT:基于 LSTM 和 GCN 的集成模型,联合学习语句的上下文信息和句法依存信息。

AS-GCN:基于 LSTM 和 GCN 的集成模型,利用句法依存信息区分 LSTM 输出的上下文权重。

AG-GCN:基于方面门控的 GCN 网络,指导面向方面的信息编码。

Kuma-GCN:设计潜在图结构补充句法特征,弥补句法依存树结构不完整的缺陷。

Bi-GCN:将单词共现率考虑到句法建模中,弥补文本处理工具错误分析的缺陷。

IGAT:利用图注意力网络设计交互注意力机制,交互学习方面项和上下文信息。

DGCN:利用句法距离加权优化上下文最终表示。

4.4 实验结果与分析

表 3 列出了各模型在 3 个数据集上的性能对比。

表 3 各模型性能对比

Table 3 Performance comparison of each model

(单位:%)

模型	Twitter		Laptop		Restaurant	
	Acc	F1	Acc	F1	Acc	F1
TD-LSTM	70.80	69.00	68.10	—	75.60	—
ATAE-LSTM	—	—	68.70	—	77.20	—
RAM	69.36	67.30	74.49	71.35	80.23	70.30
IAN	72.50	70.81	72.05	67.38	79.26	70.09
CDT	74.66	73.66	77.19	72.99	82.30	74.02
ASGCN	72.15	70.40	75.55	71.05	80.77	72.02
Bi-GCN	74.16	73.35	74.59	71.84	81.79	73.01
Kuma-GCN	72.45	70.77	76.12	72.42	81.43	73.64
AG-GCN	73.64	72.20	73.53	68.99	84.37	73.82
IGAT	75.29	73.40	76.02	72.05	82.32	73.99
DGCN	72.87	71.21	74.63	71.03	81.24	72.84
SA-GCN	75.94	74.39	78.59	74.46	82.83	74.61

由表 3 可以看出,一系列以 LSTM 为基础的模型(TD-LSTM,ATAE-LSTM)在 3 个数据集上性能相近,TD-LSTM 和 ATAE-LSTM 都利用方面项区分上下文的重要性,但是添加方面项的方式效果一般。RAM 虽然考虑到了多重注意力机制,但并未认识到方面项的重要性。IAN 认为方面项应被独立建模,并设计交互注意力机制耦合方面项与上下文特征,在 Twitters 数据集和 Restaurants 数据集上优于基于 LSTM 的模型。实验结果表明,独立建模方面项并搭配注意力机制,对于提取方面项与上下文单词的关系有着良好的效果。

利用句法依存信息后,CDT 和 ASGCN 两类集成模型在 3 个数据集上的性能得到了不错的提升,证明了句法依存信息对方面情感分类的重要性。此外关于 GCN 模型的变种,AG-GCN 依靠方面门控机制,Bi-GCN 模型构建单词贡献矩阵,Kuma-GCN 采用多种注意力机制设计潜在的图结构,这些模型并未对原始的句法依存关系做出调整。IGAT 则对句法依存关系做出调整,采用 GAT 区分节点之间的相关性,但是 GAT 采用节点拼接的方式,在脱离图结构的同时也丢失了深层卷积的优势。DGCN 虽然考虑到句法距离,但也仅仅起到类似注意力权重优化的作用,并未从卷积内部出发,增强 GCN 模型有区别地提取全局节点信息的能力。

因此,本文在上述一系列模型的基础上提出了新的模型 SA-GCN。实验结果表明,SA-GCN 在 3 个数据集上的整体性能优于其他模型。相较于集成模型 CDT,在 Twitter 数据集上,SA-GCN 的 *Acc* 提升了 1.28%,*F1* 值提升了 0.73%;在 Laptop 数据集上,*Acc* 提升了 1.4%,*F1* 值提升了 1.47%;在 Restaurant 数据集上,*Acc* 提升了 0.53%,*F1* 值提升了 0.39%。实验结果表明,独立建模方面项特征和调整句法依存规则都有利于发挥集成模型的优势。相较于 AG-GCN 模型,SA-GCN 模型在 Twitter 数据集和 Laptop 数据集上实现了性能的全面领先,尤其是在 Laptop 数据集上,*Acc* 提升了 5.06%,*F1* 值提升了 5.47%,表明句法依存关系的调整有助于进一步识别方面项的情感极性。在 Restaurant 数据集上,SA-GCN 模型的 *Acc* 降低了 1.54%,*F1* 值提高了 0.79%,说明 AG-GCN 在 LSTM 层加入方面项嵌入信息实现了一定的信息过滤作用,但是也增加了 LSTM 模型本身的参数复杂度。相较于 IGAT 模型,SA-GCN 模型在 3 个数据集上性能提升明显,尤其在 Laptop 数据集上,*Acc* 提升了 2.57%,*F1* 值提升了 2.41%,表明 GAT 利用多头注意力机制计算节点相关性的方式并不能发挥图结构中节点迭代卷积的优势。而相较于 DGCN,SA-GCN 模型同样取得了较好的结果,证明本文提出的句法距离矩阵参与内部卷积操作的方式,很好地缓解了 GCN 节点信息获取受限的问题。

4.5 消融实验

本文对整个模型框架执行消融测试,分别去除权重距离矩阵模块和方面关注注意力机制模块,以及同时去除两个模块,测试去除每个模块后模型的性能,实验结果如表 4 所列。

如表 4 所列,SA-GCN w/o A&D 模型表示 SA-GCN 模型同时去除方面关注注意力机制模块和权重距离模块;SA-

GCN w/o A 模型表示 SA-GCN 模型去除方面关注注意力机制,保留权重距离矩阵;SA-GCN w/o D 模型表示 SA-GCN 模型去除权重句法距离,保留方面关注注意力机制。首先,比较 SA-GCN w/o A 模型和 SA-GCN w/o A&D 模型,SA-GCN w/o A 模型在 3 个数据集上性能均有优势,表明权重距离矩阵能够帮助 GCN 模型有效地识别句法依存树的其他节点对当前节点的贡献度。二者在 Twitter 数据集上性能基本持平,原因在于 Twitter 数据集语句较为简短单一,权重距离矩阵更适合处理长序列语句。其次,比较 SA-GCN w/o D 模型与 SA-GCN w/o A&D 模型,SA-GCN w/o D 模型依旧在 3 个数据集上取得了较好的结果,其中,在 Restaurant 数据集上 *Acc* 提高 1.70%,*F1* 提高 2.88%。实验结果表明,方面关注注意力机制能够较好地帮助方面项着眼于语句上下文中的重要信息,而残差连接也进一步保证了上下文信息的饱和度。最后,与 SA-GCN w/o D 和 SA-GCN w/o A 模型相比,SA-GCN 模型达到了最优的效果,表明两种模块起到了互相辅助的效果,进一步提高了模型处理复杂语句的能力,增强了模型的泛化性能。

表 4 模型框架消融对比

Table 5 Comparison of model framework ablation

(单位:%)

模型	Twitter		Laptop		Restaurant	
	<i>Acc</i>	<i>F1</i>	<i>Acc</i>	<i>F1</i>	<i>Acc</i>	<i>F1</i>
SA-GCN	75.94	74.39	78.59	74.46	82.83	74.61
SA-GCN w/o D	75.08	73.84	77.45	74.36	82.73	74.55
SA-GCN w/o A	74.66	73.36	77.60	73.84	81.75	72.90
SA-GCN w/o A&D	74.66	73.06	76.35	72.66	81.03	71.76

4.6 案例分析

给定例句 1:“the food was definitely good but I just couldn't justify it for the price which including 2 drinks, \$ 100/person”。图 4(a)是例句 1 的句法依存树结构图,其中“food”和“price”和“drinks”是方面项,对应的情感极性分别是“积极”“消极”和“中性”。图 4(b)具体展示了方面项“drinks”到其他单词之间的句法距离,由于本文通过设定阈值来削弱长句法距离无关特征的干扰,因此图中仅画出 3:hot 之内的关系词。图 5 则利用注意力可视化展示了 SA-GCN 模型和 SA-GCN w/o D 模型之间的区别,通过注意力关注度测试句法距离矩阵的作用,颜色从浅到深代表单词受关注的程度从低到高。其中 SA-GCN w/o D 模型错误地将方面项“drink”的情感极性分类为“积极”,SA-GCN 模型则正确地判定方面项“drink”的情感极性为“中性”。在 LSTM 模型建模上下文信息时,由于时序上的依赖性,单词“good”得到了较高的关注度,而 GCN 模型将 LSTM 模型的输出作为输入时并不能区分其他节点对方面项的贡献度,使得单词“good”的情感色彩残留在最终的输出向量中,导致模型给出错误的分类结果。而添加句法距离后的 SA-GCN 模型通过设定阈值将单词“good”的信息排除在方面项的关系网之外,正如图 4(b)中方面项“drink”排除了“good”的干扰,集中关注句法距离较近的单词域,使得模型给出正确的分类结果。

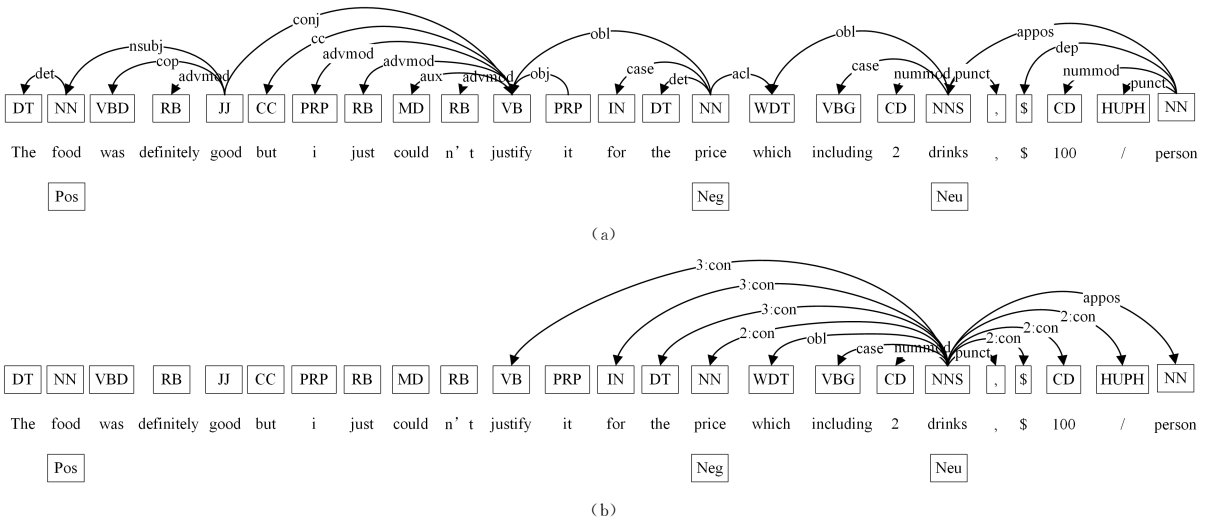


图 4 例 1 句法依存结构
Fig. 4 Syntactic dependency structure of example 1



图 5 例 1 单词注意力分布
Fig. 5 Word attention distribution of example 1

给定例句 2:“great food but the service was dreadful”, 方面项分别是“food”和“消极”, 对应的情感极性分别为“积极”和“消极”。如图 6 所示, 方面项“food”到“great”和“dreadful”的距离相同, 二者都是“food”的邻接节点。图 7 的注意力分布则展示了 SA-GCN 模型和 SA-GCN w/o A 模型之间的区别, 通过注意力关注度测试方面关注注意力机制的作用。其中, SA-GCN w/o A 针对依存距离相同的节点无法做更具体的划分, 致使方面项“food”较多地关注到“dreadful”的情感信息, 导致模型分类错误。而 SA-GCN 模型通过方面关注注意力机制, 指导方面项“food”着重关注上下文中的情感词“great”, 进一步区分方面项与上下文单词之间的相关性。

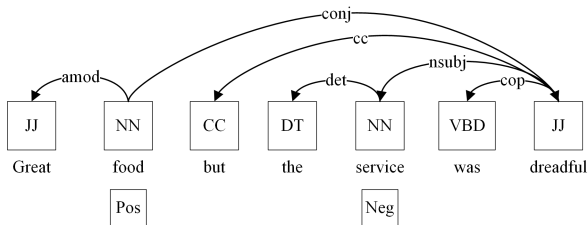


图 6 例 2 句法依存结构
Fig. 6 Syntactic dependency structure of example 2

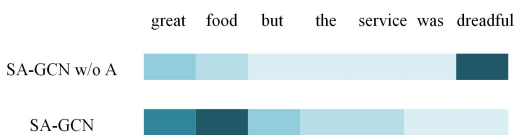


图 7 例 2 单词注意力分布
Fig. 7 Word attention distribution of example 2

结束语 针对目前关于图卷积网络无法区分依存树上全局节点贡献度的问题, 本文提出了强化句法依存规则的

方案, 并设计了基于句法依存距离和方面关注注意力机制的图卷积网络模型 SA-GCN。该模型在 GCN 卷积过程中引入句法距离矩阵, 使得 GCN 模型摆脱了邻接矩阵的范围限制, 有区别地提取到句法依存树的深层节点信息。此外, 考虑到同句法距离信息干扰的状况, 本文设计了带有残差网络的方面关注注意力机制, 使得方面项进一步关注语句中的相关情感信息, 提高了模型的泛化能力。实验结果表明, 该模型在多个公开数据集上有着较大的竞争力, 说明了句法距离和方面关注注意力机制的融合方案有助于提高方面情感分析的准确性。

参考文献

- [1] DONTU N, KUMAR S, PANDEY N, et al. Mapping the electronic word-of-mouth (eWOM) research: A systematic review and bibliometric analysis [J]. Journal of Business Research, 2021, 135: 758-773.
- [2] DANDREA E, DUCANGE P, BECHINI A, et al. Monitoring the public opinion about the vaccination topic from tweets analysis [J]. Expert Systems with Applications, 2019, 116: 209-226.
- [3] KULKARNI K, KALRO A D, SHARMA D, et al. A typology of viral ad sharers using sentiment analysis [J]. Journal of Retailing and Consumer Services, 2020, 53: 101739.
- [4] MEDSKER L R, JAIN L C. Recurrent neural networks [J]. Design and Applications, 2001, 5: 64-67.
- [5] GRAVES A, MOHAMED A, HINTON G. Speech recognition with deep recurrent neural networks [C] // 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013: 6645-6649.
- [6] DEY R, SALEM F M. Gate-variants of gated recurrent unit (GRU) neural networks [C] // 2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS).

- IEEE,2017:1597-1600.
- [7] HE R D,LEE W S,NG H T, et al. Effective Attention Modeling for Aspect-Level Sentiment Classification[C]// International Conference on Computational Linguistics. Association for Computational Linguistics,2018:1121-1131.
- [8] KIPF T N,WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907,2016.
- [9] ZHANG Y H,QI P,MANNING C D. Graph convolution over pruned dependency trees improves relation extraction[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:2205-2215.
- [10] VELIČKOVIĆ P,CUCURULL G,CASANOVA A, et al. Graph attention networks[J]. arXiv:1710.10903,2017.
- [11] HAN H,WU Y H,QIN X Y. An interaction graph attention networks model for aspect-level sentiment analysis[J]. Journal of Electronics and Information,2021,43(11):3282-3290.
- [12] WU H S,MIAO Y Q,ZHANG W Z, et al. Aspect level sentiment analysis based on distance and graph convolution network [J]. Journal of Application Research of Computers, 2021, 38(11):3274-3278.
- [13] TANG D Y,QIN B,FENG X, et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv:1512.01100,2015.
- [14] 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. 2016:606-615.
- [15] MA D H,LI S J,ZHANG X D, et al. Interactive attention networks for aspect-level sentiment classification[J]. arXiv:1709.00893,2017.
- [16] CHEN P,SUN Z Q,BING L D, et al. Recurrent attention network on memory for aspect sentiment analysis[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017:452-461.
- [17] FAN F,FENG Y S,ZHAO D Y. Multi-grained attention network for aspect-level sentiment classification[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018:3433-3442.
- [18] XING B W,LIAO L J,SONG D, et al. Earlier attention? aspect-aware LSTM for aspect-based sentiment analysis[J]. arXiv:1905.07719,2019.
- [19] GU S Q,ZHANG L P,HOU Y X, et al. A position-aware bidirectional attention network for aspect-level sentiment analysis [C]// Proceedings of the 27th International Conference on Computational Linguistics. 2018:774-784.
- [20] GILMER J,SCHOENHOLZ S,RILEY P F, et al. Message passing neural networks[M]// Machine Learning Meets Quantum Physics. Cham: Springer,2020:199-214.
- [21] 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). 2019:5679-5688.
- [22] 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). 2019:4568-4578.
- [23] LU Q,ZHU Z F,ZHANG G Y, et al. Aspect-gated graph convolutional networks for aspect-based sentiment analysis[J]. Applied Intelligence,2021,51(7):4408-4419.
- [24] CHEN J P,HUANG Z H,XUE Y. Bilateral-brain-like Semantic and Syntactic Cognitive Network for Aspect-level Sentiment Analysis[C]// 2021 International Joint Conference on Neural Networks(IJCNN). IEEE,2021:1-8.
- [25] CHEN C H,TENG Z Y,ZHANG Y. Inducing target-specific latent structures for aspect sentiment classification[C]// Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). 2020:5596-5607.
- [26] 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(EMNLP). 2020:3540-3549.
- [27] PANG S G,XUE Y,YAN Z H, et al. Dynamic and multi-channel graph convolutional networks for aspect-based sentiment analysis[C]// Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. 2021:2627-2636.
- [28] LI R F,CHEN H,FENG F X, et al. Dual graph convolutional networks for aspect-based sentiment analysis[C]// Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 6319-6329.
- [29] LI D,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(volume 2:Short papers). 2014: 49-54.
- [30] PONTIKI M,GALANIS D,PAPAGEORGIOU H, et al. SemEval-2016 task 5: Aspect based sentiment analysis[C]// International Workshop on Semantic Evaluation. 2016:19-30.



ZHANG Longji, born in 1996, postgraduate. His main research interests include natural language processing, aspect based sentiment analysis, etc.



ZHAO Hui, born in 1972, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include artificial intelligence, affective computing, speech and digital image processing.