

基于方面级情感分析的深度语义挖掘模型

张换香^{1,3}, 彭俊杰^{1,2*}

(1. 上海大学计算机工程与科学学院, 上海 200444; 2. 上海先进通信与数据科学研究院, 上海 200444;
3. 内蒙古科技大学创新创业教育学院, 内蒙古包头 014010)

摘要: 方面级情感分析是一种细粒度的情感分类任务, 具有广泛的应用前景, 正因为如此, 得到了广泛关注与研究, 尤其是近年来, 基于依赖树的图神经网络和基于注意力的网络模型的研究取得了较大进展. 但是, 由于在线评论表达的复杂性和依赖关系不易解析使得这些方法在情感分析的性能上得不到有效提升. 为了克服这些挑战, 本文提出了一种同时考虑句法语义和上下文语义的深度语义挖掘模型 (Deep Semantic Mining Model, DSMM). 具体地, 为了深度挖掘句法背后隐含的深度语义, 模型采用并行的图卷积和多头注意力机制挖掘丰富的语义; 为了充分利用句法语义和上下文语义的内在关联关系, 采用了关联注意力机制获取句法语义和上下文语义的相关性, 并且采用自适应方面路由机制有效获取方面的情感语义, 并在此基础上, 通过引入基于依赖树的语义位置嵌入, 进一步增强方面-意见词的关联. 在三个公共数据集上的实验结果表明, 该模型在复杂句情感分析中既能从不同语义空间挖掘句子的语义特征, 也能有效利用句法特征强化句子的语义表征, 在分类准确率和泛化能力上的表现优于相关工作.

关键词: 方面级情感分析; 图卷积神经网络; 多头注意力机制; 关联注意力; 句法; 上下文语义

中图分类号: TP391.1

文献标识码: A

文章编号: 0372-2112(2024)07-2307-13

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230037

A Deep Semantic Mining Model Based on Aspect-Level Sentiment Analysis

ZHANG Huan-xiang^{1,3}, PENG Jun-jie^{1,2*}

(1. School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China;

2. Shanghai Institute for Advanced Communication and Data Science, Shanghai 200444, China;

3. School of Innovation and Entrepreneurship Education, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, China)

Abstract: Aspect level sentiment analysis is a fine-grained sentiment classification task, which has a wide range of application prospects. Therefore, it has been widely concerned and researched, especially in recent years, the graph neural network based on dependency tree and the network model based on attention have made great progress. However, these studies are limited by factors such as the difficulty in parsing dependency and the complex expression of online reviews. To overcome these challenges, this paper proposes a deep semantic mining model (DSMM) that considers both syntactic and contextual semantics. Specifically, in order to mine deep semantic hidden behind the syntax, the model uses parallel graph convolution and multi-head self-attention to mine rich semantic. In order to make full use of the intrinsic correlation between syntactic semantics and contextual semantics, we used the relevance attention mechanism to obtain the correlation between syntactic semantics and contextual semantics, and we used the adaptive aspect routing mechanism to obtain the sentiment semantics of aspects effectively. Moreover, we introduced the semantic location embedding based on dependency tree to further enhance the aspect-opinion word correlation. The experimental results on three public datasets show that our model can not only mine the semantic features of sentences from different semantic spaces, but also effectively use the syntactic features to strengthen the semantic representation of sentences in sentiment analysis of complex sentence, and the performance of classification accuracy and generalization ability is better than that of related work.

Key words: aspect level sentiment analysis; graph convolutional neural network; multi-head attention mechanism; relevance attention; syntax; contextual semantics

1 引言

情感分析是分析用户在文本中所表达的情感^[1-6],分为粗粒度和细粒度的情感分析,前者是指对句子或段落进行整体的情感分析^[7],后者是指对句子或段落中具体实体或特定方面的深刻见解的挖掘,称作方面级情感分析,这样的分析对于商业决策是非常宝贵的,可以更好地理解用户行为和意图,提升产品性能及服务质量。

方面级情感分析一般由三个子任务组成,即方面词提取、意见词提取、基于方面的情感分类。本文研究句子相对于特定方面的情感分类,即在给定句子方面词的前提下,分析句子相对于此方面的情感极性。例如句子“The ambience was nice but service was poor”(来自数据集 Restaurant 的评论),相对于方面词“ambience”的情感极性是积极的,相对于方面词“service”的情感极性是消极的。因此,方面级情感分析可精确地识别用户对某一方面的感情倾向,而不是笼统地给定整个句子的情感极性。

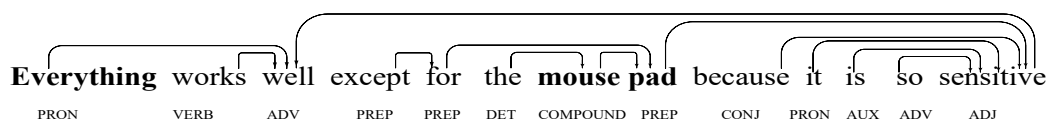
方面级情感分析的关键是对给定方面词与意见信息之间的依赖关系进行建模。以往的研究(Cai等^[8],Gu等^[9])利用各种带有注意机制的循环神经网络生成面向方面词的句子表示,并取得了不错的效果。然而,用户的评论中可能存在复杂的表达,比如,句法不规范、表述非正式、语义较深刻而非表面化等。如图1(a)所示,句子相对方面词“mouse pad”的情感极性是消极的。然而,注意机制的内在缺陷使得句子相对“mouse pad”方面的情感受到噪声影响而判断错误。意见词“well”可能比“sensitive”更受关注。但是,“well”是另一个方面“everything”的情感意见词。近年来,基于依赖树的图卷积网络和图注意网络^[10]被广泛应用在方面级情感分析中,这些方法明确地利用了句子的句法结构生

成面向方面的句子表示。然而,在复杂句中,尤其在非正式的复杂表达中,由于不易准确获取依赖关系解析结果,往往会导致情感分析错误。如图1(b)所示,句子相对方面词“service”的情感极性是消极的。但是,方面词“service”和意见词“good”之间存在依赖关系使得句子相对于方面词“service”的情感极性误判为积极的。综上分析,对于表述复杂的句子,传统方法会导致情感分析误差,只有深入分析上下文语义才能正确识别句子相对于特定方面的情感极性。为准确挖掘方面级语义及提高其情感分析的准确率,本文提出一种深度语义挖掘模型(Deep Semantic Mining Model, DSMM),通过构建基于句法的图卷积网络(Graph Convolutional Network, GCN)挖掘句子中所蕴含的丰富的句法结构信息,通过多头注意力机制(Multi-Head Attention, MHA)从不同维度挖掘上下文语义特征,利用关联注意力机制连接 GCN 和 MHA 之间相匹配的特征,将捕获的语义丰富的特征封装到特征胶囊^[11]。进一步,通过设计具有自适应方面路由功能的语义胶囊网络模块,将特征胶囊聚合为一组与方面相关的句子级语义胶囊,最终作为类胶囊的输入预测句子相对方面的情感极性。

本文的主要贡献可以概括如下:

(1)提出了一个深度语义挖掘的方面级情感分析模型,模型通过图卷积网络(Graph Convolution Network, GCN)和多头注意力(Multi-Head Attention, MHA)同时考虑了句法结构特征和不同于句法的上下文语义特征的抽取,通过关联注意力计算它们的相关性、融合句法结构及其背后的语义特征,深度捕获复杂句中蕴含的丰富且深层的上下文语义特征。

(2)设计了包含自适应方面路由的胶囊网络模块,解决方面与其意见信息的耦合,具体地,将深度语义特



Labeled aspect: mouse pad Sentiment: Negative

(a) 来自数据集 Laptop 的复杂句



Labeled aspect: service Sentiment: Negative

(b) 来自数据集 Restaurant 的复杂句

图1 复杂句示例

征封装到特征胶囊,并将其聚合为一组与方面相关的句子级语义胶囊。

(3)在 SemEval 2014 的 Restaurant 和 Laptop 及 Twitter 3 个数据集上进行了广泛的实验。实验结果证明了 DSMM 模型的有效性。

2 相关工作

方面级情感分析是自然语言处理中的一个重要任务,旨在识别评论性句子相对某一特定方面的情感极性。开展方面级情感分析研究可有效地理解用户行为,从中获得的数据可用于舆论分析、商品服务和质量的提升、商品价格预测、民众心理分析等。因此,方面级情感分析的研究具有深远的应用价值。

在早期研究中,基于机器学习的方法(Kiritchenko 等^[12],Jiang 等^[13])是常用方法,通常使用机器学习算法构建情感分类器,主要集中在手工构建特征和输入分类器来判断方面的情感极性,对特征的质量依赖程度较高,同时无法对给定方面与其上下文之间的依赖关系进行建模。

深度学习作为一种强大的机器学习技术,被广泛用于构建复杂的方法,在方面级情感分析研究中同样取得了良好结果,研究初期主要以基于卷积神经网络、循环神经网络的方法和基于时序模型的方法为主。对于第一类方法,主要根据给定的方面挖掘和捕获不同的情感特征,例如,Xue 等人^[14]引入了一种基于卷积神经网络和门控机制的新模型,根据给定的方面有选择地输出情感特征,进行情感分类。Dong 等人^[15]基于自动解析的推文,提出了一种新的递归神经网络,以自适应地将情绪信号传播到特定目标。为进一步改进模型性能,Vo 等人^[16]不是依赖于自动解析而是使用分布式词表示和神经池函数来提取丰富的自动特征集,从而提高了模型的效率。为了更准确、高效地捕捉方面的上下文,Li 等人^[17]将卷积神经网络和循环神经网络结合使用,提取方面的上下文特征,当时取得了较好的性能。由于人类语言中往往包含一些时序信息,研究人员提出了基于长短时记忆网络(Long Short-Term Memory network, LSTM)的时序模型来捕获不同时间步的信息。例如,Tang 等人^[18]设计了 TD-LSTM 和 TC-LSTM 两个改进模型,将句子分为前语境、方面词和后语境三个部分。使用两个 LSTMs 提取前上下文和后上下文的特征。然而,基于 LSTM 的模型也存在一定的局限性,词的输入存在时序性,加上遗忘门有选择地保留信息,在每个时间点获取的状态表征只包含以前一些局部浅显的特征,细节信息容易丢失。为了改善上述问题,刘全等人^[19]提出一种结合区域卷积神经网络和分层长短期记忆网络的深度分层网络模型,既可以捕获局部特征,

也可以获取全局语义特征。然而,对于复杂的句子,由于其语义较深刻,本类方法很难捕捉深层的语义信息,情感分析仍然是一个具有挑战性的问题。

近年来,人们提出了各种基于注意力的神经网络^[20]对方面及其上下文的语义关系进行建模,以捕获匹配的意见信息。例如,Huang 等人^[21]引入了分层注意力网络来识别与给定方面相关的重要情感信息。曾义夫等人^[22]提出一种基于双记忆注意力机制的方面级别情感分类模型,用于捕获语句中与给定方面词相关的词级别和短语级别信息。谢珺等人^[23]提出一种基于双指导注意力网络,同时关注到文本的上下文信息和远距离依赖信息,提高了模型对于方面级别情感特征的代表学习能力。为解决上下文信息丢失问题,Xu 等人^[24]运用多注意力网络模型,采用层次内和层次间的注意机制来捕获方面和上下文之间不同粒度的交互信息。类似地,Wang 等人^[25]提出了一种具有多重注意机制,捕捉上下文与方面词之间的交互信息,进行方面级情感分析。此外,预训练的 Transformers 双编码表示(Bidirectional Encoder Representations from Transformers, BERT)语言模型^[26]在许多 NLP 任务中都取得了显著的表现,包括方面级情感分析。Sun 等人^[27]通过构造辅助句,基于 BERT 模型将方面级情感分析任务转化为句子对分类任务。Xu 等人^[28]提出了一种 BERT 的后训练方法,以提高方面级情感分析任务微调阶段的性能。上述方法在一定程度上提高了模型的分类能力,然而在复杂句中,噪声信息会占据一定的注意分值,往往会导致注意机制捕捉到错误的意见信息而影响模型分析准确度。

此外,基于依赖树的情感分析逐渐显现出其优势。研究人员在依赖树上引入了图卷积网络来捕获与方面相关的句法结构特征。例如,Liang 等人^[29]提出了一种基于依赖树的图卷积网络 GCN 来分析句法信息。Yang 等人^[30]提出了一种基于句法依赖注意模型,用以阐述注意推理中词语间的句法依赖线索,获取上下文信息。为了增强上下文信息表示,Tang 等人^[31]采用了依赖图增强双 Transformer 网络,通过迭代交互方式联合考虑从 Transformer 学习的平面表示和从相应依赖图学习的基于图的表示。由于句子中可能有多个方面的存在,现有的方法往往会混淆联系。为了解决这一问题,Li 等人^[10]提出了一种双图卷积网络模型,该模型同时考虑了句法结构的互补性和语义相关性。An 等人^[32]提出了一种异构方面图神经网络从句间关系中学习结构和语义知识,用于情感预测。考虑到上下文句间关联的影响,Huang 等人^[33]提出了一个具有特定方面的上下文位置信息的方面级情感分析模型。该模型通过设计两个不对称的上下文位置权重函数,根据方面词在句子中

的位置调整上下文词的权重,缓减了方面词两侧字数差异对情感极性判断的干扰.Zhang等人^[34]提出了一种新的句法语义增强图卷积网络模型,不仅可以学习到与方面相关的语义相关性,还可以学习到句子的全局语义.上述研究主要基于句子的依赖关系进行建模,然而,没有考虑到在复杂句中,尤其是非正式表述句,会出现句法混乱的情况,依赖关系解析必然会出现一定的误差,影响模型的分析准确度.

综合上述讨论的方面级情感分析方法,本文针对复杂句的方面级情感分析,提出DSMM模型,通过图卷积网络捕获词间句法依赖特征的同时还通过多头注意力机制挖掘不同语义空间的上下文特征,并采用关联注意力机制将二者深度融合,获取深层的语义特征,从而提高情感分析的性能.

3 DSMM 模型

为了解决文本的表述复杂导致方面级情感分析面临挑战的问题,本文提出了一种面向方面级情感分析的深度语义挖掘模型DSMM.模型的体系结构如图2所示.该模型由六部分组成,分别是句子输入(Sentence

Input, SI)、句子嵌入表示(Embedding Representation, ER)、句子特征抽取(Feature Learning, FL)、特征融合(Feature Fusion, FF)、方面级句子语义表示(Semantic Representation, SR)和情感预测(Sentiment Prediction, SP).ER通过预训练语言模型BERT和LSTM获取句子嵌入,并通过融合每个词相对方面的位置嵌入强化面向方面的句子表示.FL用于从不同维度学习句子特征.FF用于融合所学到的特征.SR用于获取句子相对方面的语义表示.SP是将得到的句子语义信息作为预测层的输入,得到最终的情感标签.对于复杂句,特征的获取是分析其情感的关键,为此,在FL中,采用GCN和MHA并行地提取句法特征和不同语义空间的上下文特征,在FF中,使用关联注意力机制计算GCN和MHA所获取特征的相关性,从不同语义空间将特征有效融合为句子的深层语义特征并将其封装到特征胶囊.为有效获取句子相对方面的最终特征表示,在SR中,设计了自适应方面路由机制,将特征胶囊聚合为一组与方面相关的句子级语义胶囊,将其作为类胶囊的输入预测句子相对方面的感情极性.

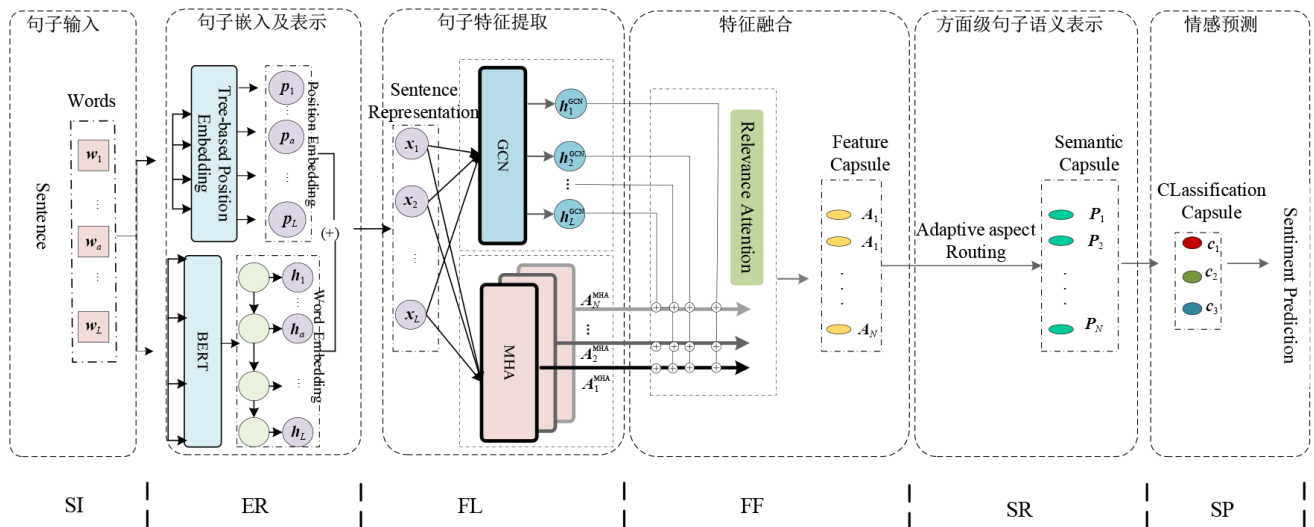


图2 深度语义挖掘模型的结构

3.1 句子嵌入及表示

对于输入的由多个词构成的句子 S ,使用预训练的BERT^[23]进行 d_s 维句子嵌入.在每个句子 S 中,单词序列的长度是不同的.为此,本文采用填充和截断的方法,将句子最终长度设为 L ,即 $S=[w_1, \dots, w_a, \dots, w_L]$,其中, w_a 表示方面词序列.句子长度 L 分两步计算,首先,计算所有句子的平均长度及标准差,然后取 λ 倍标准差和平均值的和作为 L 的值.根据 3σ 准则, λ 的值取3为最佳.为了充分挖掘句子中单词之间的语义关系,使用LSTM网络为每个时间步生成融合特征.将最终的隐藏

状态输出作为句子嵌入 $[h_1, \dots, h_a, \dots, h_L] \in \mathbf{R}^{d_s \times L}$,其中, h_a 表示句子中方面的隐藏状态输出, h_a 是通过对方面词序列嵌入求平均值所得.

此外,本文使用了另一个位置嵌入层.通过计算每个上下文词 w_i 到方面词 w_a 的语义依赖距离^[31]可以得到句子额外的位置嵌入 $[p_1, \dots, p_a, \dots, p_L] \in \mathbf{R}^{d_s \times L}$,其中, p_a 是零向量,表示方面 w_a 的位置嵌入.

通过相加融合以上两种嵌入,每个单词 w_i 的最终嵌入表示为 $x_i = h_i + p_i \in \mathbf{R}^{d_s \times L}$,因此,得到句子 S 的最终隐藏表示为 $X=[x_1, x_2, \dots, x_L] \in \mathbf{R}^{d_s \times L}$.

3.2 句子语义特征提取及融合

3.2.1 句法结构信息编码

为了对句子句法结构信息进行编码,本文使用图卷积神经网络 GCN^[35],GCN 可以对直连节点进行卷积运算,对局部信息进行编码. 通过多层 GCNs 的消息传递,图中的每个节点可以了解到更多的全局信息. 给定一个有 n 个节点的图,该图可以表示为邻接矩阵 $A \in \mathbf{R}^{n \times n}$, A 中的元素 A_{ij} 表示第 i 个节点是否连接到第 j 个节点. 具体来说,如果第 i 个节点连接到第 j 个节点,则 $A_{ij}=1$,否则 $A_{ij}=0$. 本文引入先进的依赖解析模型 LAL-Parser^[36],构建所有依赖弧的概率矩阵 $A \in \mathbf{R}^{L \times L}$. 依赖概率矩阵可以通过提供所有潜在的句法结构来捕获丰富的结构信息. 在 GCN 中,对于第 l 层的第 i 个节点,其隐藏状态表示形式为 h_i^l ,计算如式(1)所示:

$$h_i^l = \sigma \left(\sum_{j=1}^n A_{ij} W^l h_j^{l-1} + b^l \right) \quad (1)$$

其中, W^l 为权重矩阵, b^l 为偏置项, σ 为激活函数(如 ReLU). 将 3.1 节中隐藏状态向量 X 输入 GCN 模块,作为句法图中的初始节点表征,使用依赖弧的概率矩阵 $A \in \mathbf{R}^{L \times L}$ 对句子编码,通过式(1)获取句子的句法结构图表征信息 $H^{GCN} = [h_1^{GCN}, h_2^{GCN}, \dots, h_L^{GCN}]$,其中 h_i^{GCN} 表示第 i 个节点的隐藏状态表示.

3.2.2 语义信息编码

MHA 引入了缩放点积和多头计算,可在不同的特征子空间中获取相关信息. 由于这些优点,它被广泛应用于许多其他的 NLP 任务. 为了进一步提取深层语义特征,本文采用 MHA 模型对 3.1 节中隐藏状态向量 X 进行编码. MHA 注意机制的本质可以描述为查询(Q)到键(K)-值(V)对的映射. 为了增强模型的拟合能力,本文将矩阵 X 设为查询矩阵 Q 、键矩阵 K 和值矩阵 V 的初始值,即 $Q=K=V=X$. 在 MHA 中,每个头的 Q 、 K 和 V 首先由线性转换层处理,如式(2)所示:

$$\begin{cases} \bar{Q}_h = X \times W_h^Q \\ \bar{K}_h = X \times W_h^K \\ \bar{V}_h = X \times W_h^V \end{cases} \quad (2)$$

其中, W_h^Q 、 W_h^K 和 W_h^V 分别是 Q 、 K 和 V 的参数矩阵, h 表示具体的头部.

首先通过 Q 和 K 的点积来计算相似度,然后除以 \sqrt{d} (d 是矩阵 X 的维数),这样得到的点积计算结果不会太大. 然后通过 softmax 函数对结果进行归一化,再乘以矩阵 V 得到每个头部的信息表示 A_h^{MHA} . 运算过程如式(3)所示. 为从不同空间维度挖掘语义信息,本文区别于传统的 MHA 使用方式,并没有将多头表征进行连接,而是将独立的多头信息单独保留,聚集得到多个子空间上丰富的语义表征 A^{MHA} ,如式(4)所示,其中 N 表示头数.

$$A_h^{MHA} = \text{softmax} \left(\frac{\bar{Q}_h \times (\bar{K}_h)^T}{\sqrt{d}} \right) \times \bar{V}_h \quad (3)$$

$$A^{MHA} = \{A_1^{MHA}, A_2^{MHA}, \dots, A_N^{MHA}\} \quad (4)$$

3.2.3 特征融合

为了在挖掘句子语义特征的同时也获取句子的句法结构特征,本文设计了关联注意力机制(relevance attention),计算 GCN 和 MHA 所捕获特征的相关性,将句法结构信息和语义信息高度关联,并引入共享掩码^[37]共同作用获取共享的相关特征. 为了直观地描述这个方法,在图 3 中展示了这个过程. 具体地,建立了一个注意图,边的值表示 A^{MHA} 相对于 H^{GCN} 的注意权值 α ,权值计算如式(5)和式(6),其中 $\alpha_{MHA-GCN}^{i,j}$ 表示 A^{MHA} 的第 i 个特征相对于第 j 个隐藏状态的注意值. 为获取和句法相关度较高的语义特征,将注意权重值进行排序,对于每个 H^{MHA} 节点,只保留具有较大值的边(大于预先设定的阈值 T),并删除其它边. 将这个注意图映射到共享掩码 M_S 矩阵,如果 MHA 和 GCN 之间有一条边,则 M_S 为 1,否则 M_S 为 0. 为聚焦 MHA 中共享特征,将共享掩码作用于 $\alpha_{MHA-GCN}^i$,具体过程如式(7)和式(8).

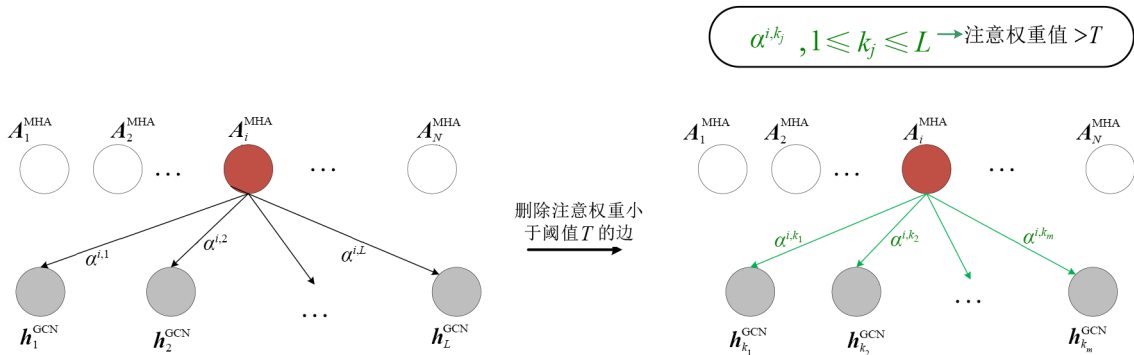


图 3 获取共享掩码

$$s_{\text{MHA-GCN}}^{i,j} = \mathbf{W}_3 (\tanh(\mathbf{W}_2 ([\mathbf{A}_i^{\text{MHA}}, \mathbf{h}_j^{\text{GCN}}] + \mathbf{b}_2))) \quad (5)$$

$$\alpha_{\text{MHA-GCN}}^{i,j} = \frac{e^{s_{\text{MHA-GCN}}^{i,j}}}{\sum_{j=i}^L e^{s_{\text{MHA-GCN}}^{i,j}}} \quad (6)$$

$$\alpha_{\text{MHA-GCN}}^i = [\alpha_{\text{MHA-GCN}}^{i,1}, \alpha_{\text{MHA-GCN}}^{i,2}, \dots, \alpha_{\text{MHA-GCN}}^{i,L}] \quad (7)$$

$$\alpha_{\text{MHA-GCN}}^i = \alpha_{\text{MHA-GCN}}^i \odot \mathbf{M}_S \quad (8)$$

其中, \mathbf{W}_2 和 \mathbf{W}_3 是可训练参数, \mathbf{b}_2 为偏置, \odot 表示对应位置元素相乘.

为了获取句法和语义的共享特征, 将 $\mathbf{A}^{\text{MHA}} = \{\mathbf{A}_1^{\text{MHA}}, \mathbf{A}_2^{\text{MHA}}, \dots, \mathbf{A}_N^{\text{MHA}}\}$ 中每一个表征与 \mathbf{H}^{GCN} 中的关联特征进行连接得到每个子空间上的特征表示 \mathbf{A}_i , 计算过程如式(9):

$$\mathbf{A}_i = [\mathbf{A}_i^{\text{MHA}}; \alpha_{\text{MHA-GCN}}^i \mathbf{H}^{\text{GCN}}] \quad (9)$$

基于胶囊网络处理特征的优越性^[11], 模型中引入特征胶囊层, 将每一个特征 \mathbf{A}_i 封装到特征胶囊, 则得到最终特征胶囊的输出序列为 $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_N]$.

3.2.4 方面级句子语义表示

对于方面级情感分析, 确定方面和上下文之间的关系是任务的关键. 为此, 本文设计了一种自适应方面路由方法来计算特征 \mathbf{A} 相对方面的权重. 通过式(10)计算得到方面路由权重 a_i :

$$a_i = \text{sigmoid}(\mathbf{A}_i + \mathbf{W}_a \mathbf{x}_a + \mathbf{b}_a) \quad (10)$$

其中, \mathbf{x}_a 是方面权重, \mathbf{W}_a 是将 \mathbf{x}_a 转换成标量的转换矩阵, \mathbf{b}_a 是偏置. 生成的路由权重 $a_i \in [0, 1]$ 将方面信息与其上下文特征相融合, 自适应地聚合上下文特征. 它控制当前上下文有多少信息可以传输到下一层. 基于 N 个特征胶囊通道, 可得整个方面路由权重 $\mathbf{a} = [a_1, a_2, \dots, a_N]$, 最后, 使用这些权重对特征胶囊进行路由, 得到与方面关联度较高的语义胶囊, 具体如式(11):

$$\mathbf{P}' = \mathbf{A} \odot \mathbf{a} \quad (11)$$

其中, \mathbf{P}' 为方面语义胶囊, \odot 表示元素对应相乘.

为了使得每个语义胶囊 \mathbf{P}'_i 的长度表示 \mathbf{P}'_i 的语义在当前输入中出现的概率, 本文使用非线性压缩函数 squash^[11] 将其长度限制到 $[0, 1]$, 具体如式(12):

$$\mathbf{P}_i = \frac{\|\mathbf{P}'_i\|^2}{1 + \|\mathbf{P}'_i\|^2} \frac{\mathbf{P}'_i}{\|\mathbf{P}'_i\|} \quad (12)$$

3.2.5 情感预测

为了获得有效的情感分类结果, 本文使用类胶囊来表示类, 类胶囊长度作为分类概率^[38]. 因本文数据共标注3种情感极性, 所以设置3个类胶囊. 每个类胶囊都拥有自己的路由权重, 以自适应地聚合来自前一层的语义胶囊. 从语义胶囊 $i (1 \leq i \leq N)$ 到类胶囊 $j (1 \leq j \leq N)$ 会产生一个预测向量 $\widehat{\mathbf{P}}_{j|i}$:

$$\widehat{\mathbf{P}}_{j|i} = \mathbf{W}_{ij} \mathbf{P}_i \quad (13)$$

其中, \mathbf{W}_{ij} 是权重矩阵, \mathbf{P}_i 为语义胶囊 i 的向量表示 (参见式(12)). 每个类胶囊 j 的向量表征可通过所有语义胶囊预测向量加权和获取, 具体如下:

$$\mathbf{c}_j = \sum_{i=1}^N \varphi_{ij} \widehat{\mathbf{P}}_{j|i} \quad (14)$$

其中, φ_{ij} 是通过从语义胶囊到类胶囊及预测概率整个过程迭代 r 次更新所得的耦合系数. 对于类胶囊 \mathbf{c}_j 再次使用非线性压缩函数 squash 得到最终表示 \mathbf{v}_{c-j} :

$$\mathbf{v}_{c-j} = \text{squash}(\mathbf{c}_j) \quad (15)$$

其中, j 的取值为 $1, 2, 3$, \mathbf{v}_{c-j} 的值限制在 $[0, 1]$, 表示情感预测概率.

4 实验结果与分析

4.1 数据集

在3个公共标准数据集 Restaurant、Laptop 和 Twitter 数据集上进行了实验. 统计数据如表1所示, 其中*表示本类数据集中包含复杂句的数量.

表1 数据集的统计

单位: 个

数据集	数据分配	积极	消极	中性
Restaurant	训练集	2 164(*206)	807(*84)	737(*6)
	测试集	727(*106)	196(*24)	196(*3)
Laptop	训练集	976(*182)	851(*149)	455(*4)
	测试集	337(*106)	128(*16)	167(*3)
Twitter	训练集	1 507(*312)	1 528(*298)	3 016(*18)
	测试集	172(*34)	169(*17)	336(*3)

简要介绍如下:

(1) Restaurant 数据集

Restaurant 数据集是在 SemEval 2014^[39] 中公开的数据集. 该数据集包含 4 827 个评论, 涉及对食品、餐具、用餐环境及服务的评论, 包含复杂句 429 个, 其中训练集共 296 个, 测试集共 133 个.

(2) Laptop 数据集

Laptop 数据集同样来源于 SemEval 2014^[39]. 该数据集包含 2 148 个评论, 涉及对笔记本电脑各部件的功能、体验感及服务的评论, 包含复杂句 460 个, 其中训练集共 335 个, 测试集共 125 个. 相较于 Restaurant 数据其存在更多复杂句, 再加上需要对计算机常识的理解, 语义及情感分析相对困难.

(3) Twitter 数据集

Twitter 数据集是推文的集合^[40], 该数据集包含 6 728 个评论, 涉及对在各种各样的话题如品牌、名人、产品和公共事件等的评论, 包含复杂句 682 个, 其中训练集共 628 个, 测试集共 54 个.

以上3个数据集集中的每个评论都手动标注了一个

方面词及其对应的情感极性:消极(-1)、中性(0)和积极(1). 数据集中训练集占70%,测试集占30%.

(4)评估指标

本文使用常用的情感分类评估指标来验证模型性能,分别为分类准确度(Accuracy)和兼顾准确率和召回率的 F_1 分数(F_1 -score). 评估指标值越高(↑)代表模型性能越好.

4.2 基线模型

将本文模型与一些先进的基线模型进行比较. 模型整体上分为基于注意力机制的模型,如MAMN(Multiple Attention Mechanism Network)、RAM(Recurrent Attention on Memory)等,基于句法依赖的模型,如CDT(Convolution over a Dependency Tree)、DualGCN(Dual Graph Convolutional Network)等,以及基于BERT的模型,如R-GAT+BERT(Relational Graph Attention network+BERT)等.

MAN(Multi-Attention Network)^[24]:MAN运用多注意力网络模型,采用层次内和层次间的注意机制来捕获方面和上下文之间不同粒度的交互信息;

MAMN(Multiple Attention Mechanism Network)^[25]:MAMN提出了一种具有多重注意机制,捕捉上下文与方面词之间的交互信息,进行方面级情感分析;

RAM(Recurrent Attention on Memory)^[41]:RAM运用多重注意网络和记忆网络学习句子表示;

TNet(Transformation Networks)^[42]:TNet将双向LSTM嵌入转换为特定目标的嵌入,并使用CNN提取最终的嵌入进行分类;

CPA-SA(Sentiment Analysis with aspect-specific Context Position)^[33]:CPA-SA设计了一种多粒度的注意机制来捕获方面和上下文之间的字级交互;

CDT(Convolution over a Dependency Tree)^[43]:CDT利用依赖树上的GCN学习具有语法信息的方面表示;

BiGCN(Bidirectional Graph Convolution Network)^[11]:BiGCN使用层次图结构整合词共现信息和依赖类型信息;

InterGCN(Interactive Graph Convolutional Network)^[29]:InterGCN利用GCN在依赖树上学习具有语法信息的方面表示;

R-GAT(Relational Graph Attention network)^[44]:R-GAT提出了一个面向方面的依赖树结构,然后用关系GAT编码新的依赖树;

DGEDT(Dependency Graph Enhanced Dual-Transformer network)^[31]:DGEDT通过联合考虑平面表示和基于图的表示,提出了一种依赖图增强的双向Transformer网络;

DualGCN(Dual Graph Convolutional Network)^[10]:

DualGCN设计了一种双图卷积网络捕获方面和上下文信息;

DualGCN+BERT(Dual Graph Convolutional Network+BERT)^[10]:DualGCN+BERT是使用预训练的BERT代替双向LSTM作为编码器的DGEDT模型;

R-GAT+BERT(Relational Graph Attention network+BERT)^[45]:R-GAT+BERT是使用预训练BERT的GAT模型.

4.3 网络参数设置

本文采用BERT预训练的768维向量作为词嵌入. 使用先进的LAL-Parser^[36]对句子进行依赖解析,在每个数据集上运行模型5次,并报告测试集上的平均性能. 在训练过程中,考虑调优以下超参数:学习率(lr=0.001),批大小(bs=128),dropout(0.3),注意力头(h=16),r=4. 我们在所有实验中使用初始学习率的Adam优化器^[43],并在20个epochs停止.

4.4 模型对比结果分析

DSMM和基线模型在准确率(Accuracy)和 F_1 -score的实验结果如表2所示,加粗数据为最优结果.

本文提出的DSMM模型在Restaurant、Laptop和Twitter数据集上的表现始终优于所有基于注意力和基于句法的方法. 这些结果表明,DSMM有效地整合了句法结构知识和语义信息,适合包含正式、非正式或复杂评论的数据集情感分析. DSMM与基于注意力的方法如RAM、TNet、MAN、MAMN和CPA-SA相比,DSMM模型利用句法知识建立单词之间的依赖关系,适当地避免了由注意机制引入的噪声干扰,取得了更好的性能. 同样可发现,DSMM性能超越了所有的基于句法的方法和使用BERT的方法,如CDT、BiGCN、R-GAT、DGEDT、R-GAT+BERT、DualGCN+BERT等,出现这种现象的原因之一是DSMM模型在初始化单词嵌入向量时受益于预训练的BERT. 其次,DSMM使用MHA和GCN联合挖掘句子的语义信息,确保模型能够学习到更多抽象的语义信息. 还发现,在基于注意力机制的方法中,MAMN性能表现最好,超越了其它所有的基于注意力机制的方法,也超越了大部分基于句法的方法. 分析其原因,一方面得益于BERT模型,另一方面,它使用多重注意机制使得特定方面的长短距离信息都被充分挖掘,但由于忽略了单词之间的语义相关性,其性能仍然未能超越DSMM,在准确率指标上也同样逊色于部分基于句法的方法如DualGCN+BERT. 在基于句法的方法中,DualGCN+BERT性能表现最佳,能超越所有基于注意力的方法,主要原因是其既考虑到词嵌入的准确性也考虑到句法的影响,同样也考虑到了上下文语义信息的重要性. 但是这一方法针对特别的复杂表达及非正式表达的评论,不是完全有效,所以仍然未能超越

表 2 3个公开数据集的实验结果

单位:%

模型	Restaurant \uparrow		Laptop \uparrow		Twitter \uparrow		Source
	Accuracy	F_1 -score	Accuracy	F_1 -score	Accuracy	F_1 -score	
RAM ^[41]	80.23	70.80	74.49	71.35	69.36	67.30	ACL 2017
TNet ^[42]	80.69	71.27	76.54	71.75	74.90	73.60	ACL 2018
CPA-SA ^[33]	82.66	74.41	75.18	71.50	—	—	KBS 2022
MAN ^[24]	84.29	71.36	78.21	72.98	76.70	72.41	Neucom 2020
MAMN ^[25]	86.52	81.57	81.35	77.83	76.59	75.27	KBS 2021
CDT ^[45]	82.30	74.02	77.19	72.99	74.66	73.66	ACL 2019
BiGCN ^[11]	81.97	73.48	74.59	71.84	74.16	73.35	ACL 2020
InterGCN ^[29]	82.23	74.01	77.86	74.32	—	—	COLING 2020
R-GAT ^[44]	83.30	76.08	77.42	73.76	75.57	73.82	ACL 2020
DGEDT ^[31]	83.90	75.10	76.80	72.30	74.80	73.40	ACL 2020
R-GAT+BERT ^[44]	86.60	81.35	78.21	74.07	76.15	74.88	ACL 2020
DualGCN+BERT ^[10]	86.91	81.16	81.40	78.10	77.40	76.02	ACL 2021
DualGCN ^[10]	84.27	78.08	78.48	74.74	75.92	74.29	ACL 2021
DSMM	87.82	83.21	83.84	79.54	80.01	78.83	

注:“—”表示在本研究中没有公开结果也没有公开实现方式。

DSMM. 此外,在基线模型中,由于RAM仅使用简单注意机制和记忆网络结构,因此较其它模型性能不佳. 由于TNet设计了一种目标特定转换策略可以更好地将目标信息集成到单词表示中,性能略优于RAM. 而CDT将依赖树与神经网络集成增强特征学习,性能不仅优于RAM并且超越了后来的BiGCN. 更值得注意的是MAN虽然只用到了注意机制,但得益于其巧妙的设计策略,使用3个转换编码器层来获取上下文和方面词的隐藏表示. 因此,MAN的性能优于多数的模型. 基于BERT的模型,如DualGCN+BERT不仅使用预先训练好的BERT进行初始词嵌入,还使用双向的图卷积,同时

设计了新颖正则化策略捕获语义信息,性能表现较突出. 最后,整体来看MAMN,R-GAT+BERT和DualGCN+BERT在数据集Restaurant上与本文方法性能相当,但是在另外两个数据集上性能相差较大,尤其是Accuracy结果充分证明本文方法适用面较广,有着良好的通用性.

尽管性能是模型最重要的指标,但时间开销也是反映算法模型真正实用性的重要指标. 为了验证所提出方法的实用性,将本文方法和先进的基线模型在训练时间上作了对比分析,为了公平比较,所有方法都是在相同的处理器Nvidia A100 GPU上训练的,各种方法的计算训练时间如表3所示,加粗数据表示最优结果.

表 3 3个公开数据集上各种方法的计算时间

单位:s

模型	Restaurant	Laptop	Twitter
	Total training time	Total training time	Total training time
RAM ^[41]	60.4	47	98.3
TNet ^[42]	175.5	93.7	253.3
MAN ^[24]	139.1	75.6	175.8
MAMN ^[25]	145.3	91.4	201.6
CPA-SA ^[33]	80.5	53.2	131.4
CDT ^[45]	122.4	72.8	156.5
BiGCN ^[11]	5 220.3	2 802.3	6 054.1
InterGCN ^[29]	5 100.4	2 721.7	6 410.3
R-GAT ^[44]	8 900	5 430	13 320.9
DGEDT ^[31]	6 354.8	4 002.6	7 103.4
DualGCN ^[10]	138.6	102.5	259.8
R-GAT+BERT ^[44]	1 1720	7 850.4	14 032.3
DualGCN+BERT ^[10]	189.4	133.6	323.5
DSMM	176.6	103.6	260.2

表3结果表明,整体而言,大部分基于句法依赖的方法训练时间高于基于注意力机制的方法.因为基于句法依赖的方法需要解析词间依赖关系,需要耗费大量的时间.本文方法的训练时间虽然超过了所有基于注意力的方法,但是,在 Accuracy 和 F_1 -score 性能上突显了其优势,同时,除了 CDT 和 DualGCN 方法外,其训练时间低于所有基于句法依赖的方法和基于 BERT 的方法. CDT 利用 Bi-LSTM 生成上下文嵌入,并使用 GCN 层进一步增强嵌入, DualGCN 使用两个独立的 GCNs 来处理句法和语义知识,较这两种方法而言,本文方法除了用到 GCN 获取句法特征外还用到了多头注意力和关联注意机制,增加了模型的计算量.但是本文方法将句法依赖解析任务提前进行了预处理,训练时间大大低于其它基于句法和基于 BERT 的方法.

最后,综合来看,本文方法在 3 个数据集上的准确率都超越了所有基线模型,其中,在 Laptop 上准确率超过表现最好的基线模型 2.44%,同时训练时间缩短了 22.46%. 在 Twitter 上准确率超过表现最好的基线模型 2.61%,训练时间缩短了 19.57%. 在 Restaurant 上准确率超过表现最好的基线模型 0.91% 上,训练时间缩短了 6.76%. 以上结果充分证明本文方法有着良好的泛化性,并且有着广泛的应用前景.

4.5 消融实验与分析

为了验证 DSMM 中各个组成部分的合理性及其作用,本文进行了广泛的消融实验.结果如表4所示,加粗数据表示最优结果,GCN+Capsule 是指通过 GCN 从句子中提取词间关联特征,作为特征胶囊的输入,再通过自适应方面路由获取特定方面的上下文语义得到句子相对特定方面的语义表征,进一步通过分类胶囊层进行情感预测. MHA+Capsule 与 GCN+Capsule 相比唯一变化的就是特征提取部分用 MHA 取代,其它保持不变. GCN+position+Capsule 是在 GCN+Capsule 的基础上引入语义位置嵌入强化句子相对特定方面的嵌入表示. MHA+position+Capsule 是在 MHA+Capsule 的基础上引入语义位置嵌入. GCN+MHA+Capsule 是将特征提取部分用并行的 GCN 和 MHA 联合方式取代独立使用 GCN 或独立使用 MHA 方式. 从表4可以看出,DSMM 模型在公共数据集 Restaurant、Laptop 和 Twitter 上都获得了最佳性能. GCN+MHA+Capsule 较 MHA+Capsule 和 GCN+Capsule 而言性能得到了大幅度的提升,究其原因,联合 GCN 和 MHA 不仅能捕获词间关联语义特征而且能够从不同语义空间深度挖掘句子的深层语义特征,强化了模型学习复杂句所表达的语义信息的能力.此外,语义位置嵌入在与 GCN 的组合中,性能提升更加显著,这可能是词位置信息对句法更加敏感,得到较好的信息强化作用.

表4 消融实验结果

单位: %

模型	Restaurant ↑		Twitter ↑		Laptop ↑	
	Accuracy	F_1 -score	Accuracy	F_1 -score	Accuracy	F_1 -score
GCN+Capsule	79.40	74.05	71.12	65.31	72.31	68.54
MHA+Capsule	82.60	75.12	74.96	72.68	75.01	70.23
GCN+position+Capsule	81.94	75.49	75.16	72.35	76.91	71.53
MHA+position+Capsule	82.68	75.12	74.97	72.77	76.99	72.10
GCN+MHA+Capsule	86.89	82.27	79.59	75.68	82.98	77.88
DSMM	87.82	83.21	80.01	78.83	83.84	79.54

4.6 阈值 T 对模型性能的影响

为了验证共享掩码(第3.2.3节)中注意权重阈值的大小对模型准确性的影响,在 Restaurant 和 Laptop 数据集上评估了不同阈值的 DSMM 模型.如图4所示,在 Restaurant、Laptop 和 Twitter 上的阈值分别为 0.1、0.15 和 0.15 时模型表现最好.一方面,当阈值较小时,上下文语义表征会受到噪声信息的严重影响;另一方面,如果阈值太大,会滤除过多的有用信息而削弱特征表示.此外发现,曲线前面部分起伏较大,后面部分趋于平缓,表明阈值的合理设置可适当抑制噪声的影响,有效获取重要的情感信息,但是超过一定值对模型性能的影响几乎为0.

4.7 实例研究分析

为了直观地理解 DSMM 模型较其它模型的优越

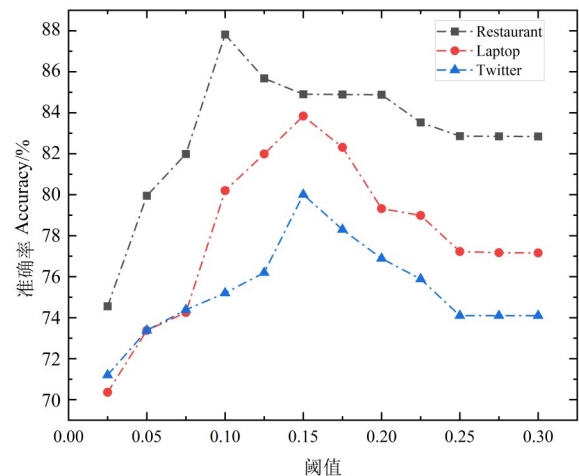


图4 共享掩码中权重阈值对模型性能的影响

性,本文通过几个实际案例在不同模型中的结果进行分析比较,具体如表5所示.表中用积极、消极和中性分别表示3种相应的情感.表中5个实际案例选自数据

集Laptop、Twitter和Restaurant.选取性能较好的基于注意力机制模型MAMN和基于句法依赖的模型DualGCN与本文方法进行比较分析.

表5 在DSMM模型与先进的基线模型上的案例研究

序号	方面	句子	真实性	基线模型及分析结果		
				DSMM	MAMN	DualGCN
1	fan	the i3 processor doesn't seem to run hot and the fan rarely turns on.	积极	积极	消极	积极
2	DDR3	It could be a perfect laptop if it would have faster system memory and its radeon 5850 would have DDR5 instead of DDR3.	消极	消极	积极	消极
3	color	Perfect, got, verry stylish dress, but expect one with better color	消极	消极	消极	积极
4	protector	I bought a protector for my key pad and it works great :)	中性	积极	积极	积极
5	service	Oh, my God, so delicious pizza, but wait two hours, expect good service	消极	消极	消极	积极

基于注意力的方法更倾向于关注句子中的部分单词,没有其它的限制方法,导致噪声词也会被重点关注,MAMN虽是性能最好注意力模型,但在复杂句中分析中仍然存在一定的困难,如案例1,模型更关注于就近的意见词“hot”,引发严重误判.此外,虽然句法依赖关系可以在一个方面与某些词之间建立直接联系,但在语序较混乱的复杂句中,方面与就近的噪声观点词之间会建立关联,如案例3,方面词与就近的描述词“better”关联,案例5中,方面词与相邻的“good”建立直接关联,都导致模型判断失败.5个案例中,本文模型在4个中突显其优势,案例1、2、3、5显著表明,本文DSMM模型充分考虑了句法知识和语义信息的互补性,尤其是其内部的关联注意力机制,使得模型从不同的语义空间深度捕获句法和语义特征,可有效处理复杂的句子.

5 结论

考虑到在线评论的复杂性表达导致方面级情感分析准确率不高的问题,本文设计了一个新的模型DSMM来改进方面级的情感分析性能.DSMM采用预先训练好的BERT模型来初始化词嵌入,使得其比Word2vec或Glove等常用方法更好.通过GCN和MHA并行提取句子特征代替胶囊网络中CNN特征提取功能,并且在本层中设计了关联注意力方法,得到丰富语义特征封装到特征胶囊.在方面语义胶囊层设计自适应方面路由机制捕获方面的上下文特征,最后输入分类胶囊层进行情感预测.实验结果表明,在复杂句情感分析中,DSMM模型可以从不同语义空间挖掘句子的语义特征,结合丰富的句法特征得到句子的强有力的语义表征,在性能上的表现优于相关工作.未来,将尝试添加一些辅助信息,使用更多融合机制融合更多外部知识和背景知识,使得模型更深度地挖掘句子背后隐藏的特殊语义,为情感分析提供更多的基准结果.

致谢 感谢2022年上海市服务业发展引导资金项目和上海大学上海市科学与工程计算专业技术服务平台的支持.

参考文献

- [1] 黄发良,李超雄,元昌安,等.基于TSCM模型的网络短文情感挖掘[J].电子学报,2016,44(8):1887-1891.
HUANG F L, LI C X, YUAN C A, et al. Mining sentiment for web short texts based on TSCM model[J]. Acta Electronica Sinica, 2016, 44(8): 1887-1891. (in Chinese)
- [2] 黄发良,冯时,王大玲,等.基于多特征融合的微博主题情感挖掘[J].计算机学报,2017,40(4):872-888.
HUANG F L, FENG S, WANG D L, et al. Mining topic sentiment in microblogging based on multi-feature fusion[J]. Chinese Journal of Computers, 2017, 40(4): 872-888. (in Chinese)
- [3] 黄发良,于戈,张继连,等.基于社交关系的微博主题情感挖掘[J].软件学报,2017,28(3):694-707.
HUANG F L, YU G, ZHANG J L, et al. Mining topic sentiment in micro-blogging based on micro-blogger social relation[J]. Journal of Software, 2017, 28(3): 694-707. (in Chinese)
- [4] 何炎祥,孙松涛,牛菲菲,等.用于微博情感分析的一种情感语义增强的深度学习模型[J].计算机学报,2017,40(4):773-790.
HE Y X, SUN S T, NIU F F, et al. A deep learning model enhanced with emotion semantics for microblog sentiment analysis[J]. Chinese Journal of Computers, 2017, 40(4): 773-790. (in Chinese)
- [5] 饶元,吴连伟,王一鸣,等.基于语义分析的情感计算技术研究进展[J].软件学报,2018,29(8):2397-2426.
RAO Y, WU L W, WANG Y M, et al. Research progress on emotional computation technology based on semantic analysis[J]. Journal of Software, 2018, 29(8): 2397-2426.

- (in Chinese)
- [6] 李卫疆, 漆芳, 余正涛, 等. 基于多通道特征和自注意力的情感分类方法[J]. 软件学报, 2021, 32(9): 2783-2800.
LI W J, QI F, YU Z T, et al. Sentiment classification method based on multi-channel features and self-attention[J]. Journal of Software, 2021, 32(9): 2783-2800. (in Chinese)
- [7] 郭贤伟, 赖华, 余正涛, 等. 融合情绪知识的案件微博评论情绪分类[J]. 计算机学报, 2021, 44(3): 564-578.
GUO X W, LAI H, YU Z T, et al. Emotion classification of case-related microblog comments integrating emotional knowledge[J]. Chinese Journal of Computers, 2021, 44(3): 564-578. (in Chinese)
- [8] CAI Y, HUANG Q B, LIN Z J, et al. Recurrent neural network with pooling operation and attention mechanism for sentiment analysis: A multi-task learning approach[J]. Knowledge-Based Systems, 2020, 203: 105856.
- [9] 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. New York: ACL, 2018: 774-784.
- [10] 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. Stroudsburg: ACL, 2021: 6319-6329.
- [11] CHEN Z, QIAN T Y. Transfer capsule network for aspect level sentiment classification[C]//Proceedings of the 57th Conference of the Association for Computational Linguistics. Stroudsburg: ACL, 2019: 547-556.
- [12] KIRITCHENKO S, ZHU X D, CHERRY C, et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews[C]//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). Stroudsburg: ACL, 2014: 437-442.
- [13] JIANG L, YU M, ZHOU M, et al. Target-dependent Twitter sentiment classification[C]//The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2011: 151-160.
- [14] XUE W, LI T. Aspect based sentiment analysis with gated convolutional networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2018: 2514-2523.
- [15] DONG L, WEI F, 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: ACL, 2014: 49-54.
- [16] VO D, ZHANG Y. Target-dependent twitter sentiment classification with rich automatic features[C]//Proceedings of the Twenty Fourth International Joint Conference on Artificial Intelligence. New York: AAAI Press, 2015: 1347-1353.
- [17] LI X, BING L D, LI P J, et al. A unified model for opinion target extraction and target sentiment prediction[C]//The Thirty-Third AAAI Conference on Artificial Intelligence, The Thirty-First Innovative Applications of Artificial Intelligence Conference. New York: AAAI Press, 2019: 6714-6721.
- [18] TANG D Y, QIN B, FENG X C, et al. Target-dependent sentiment classification with long short term memory [EB/OL]. (2015-12-01)[2023-01-10]. <http://arxiv.org/abs/1512.01100>.
- [19] 刘全, 梁斌, 徐进, 等. 一种用于基于方面情感分析的深度分层网络模型[J]. 计算机学报, 2018, 41(12): 2637-2652.
LIU Q, LIANG B, XU J, et al. A deep hierarchical neural network model for aspect-based sentiment analysis[J]. Chinese Journal of Computers, 2018, 41(12): 2637-2652. (in Chinese)
- [20] 严莉, 张凯, 徐浩, 等. 基于图注意力机制和Transformer的异常检测[J]. 电子学报, 2022, 50(4): 900-908.
YAN L, ZHANG K, XU H, et al. Anomaly detection based on graph attention mechanism and transformer[J]. Acta Electronica Sinica, 2022, 50(4): 900-908. (in Chinese)
- [21] HUANG F L, LI X L, YUAN C G, et al. Attention-emotion-enhanced convolutional LSTM for sentiment analysis[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4332-4345.
- [22] 曾义夫, 蓝天, 吴祖峰, 等. 基于双记忆注意力的方面级别情感分类模型[J]. 计算机学报, 2019, 42(8): 1845-1857.
ZENG Y F, LAN T, WU Z F, et al. Bi-memory based attention model for aspect level sentiment classification[J]. Chinese Journal of Computers, 2019, 42(8): 1845-1857. (in Chinese)
- [23] 谢珺, 王雨竹, 陈波, 等. 基于双指导注意力网络的属性情感分析模型[J]. 计算机研究与发展, 2022, 59(12): 2831-2843.

- XIE J, WANG Y Z, CHEN B, et al. Aspect-based sentiment analysis model with Bi-Guide attention network[J]. *Journal of Computer Research and Development*, 2022, 59(12): 2831-2843. (in Chinese)
- [24] XU Q N, ZHU L, DAI T, et al. Aspect-based sentiment classification with multi-attention network[J]. *Neurocomputing*, 2020, 388: 135-143.
- [25] WANG X D, TANG M W, YANG T, et al. A novel network with multiple attention mechanisms for aspect-level sentiment analysis[J]. *Knowledge-Based Systems*, 2021, 227: 1-12.
- [26] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//*Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg: ACL, 2019: 4171-4186.
- [27] SUN C, HUANG L Y, QIU X P. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence[C]//*Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg: ACL, 2019: 380-385.
- [28] XU H, LIU B, SHU L, et al. BERT post-training for review reading comprehension and aspect-based sentiment analysis[C]//*Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg: ACL, 2019: 2324-2335.
- [29] LIANG B, YIN R D, GUI L, et al. Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis[C]//*Proceedings of the 28th International Conference on Computational Linguistics*. Stroudsburg: International Committee on Computational Linguistics, 2020: 150-161.
- [30] YANG T, YIN Q, YANG L, et al. Aspect-based Sentiment analysis with new target representation and dependency attention[J]. *IEEE Transactions on Affective Computing*, 2022, 13(2): 640-650.
- [31] TANG H, JI D H, LI C L, et al. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification[C]//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg: ACL, 2020: 6578-6588.
- [32] AN W B, TIAN F, CHEN P, ZHENG Q H. Aspect-based sentiment analysis with heterogeneous graph neural network[J]. *IEEE Transactions Computer Society System* 2023, 10(1): 403-412.
- [33] HUANG B, GUO R Y, ZHU Y M, et al. Aspect-level sentiment analysis with aspect-specific context position information[J]. *Knowledge-Based Systems*. 2022, 243: 1-11.
- [34] ZHANG Z, ZHOU Z, WANG Y. SSEGNCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis[C]//*Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics*. Stroudsburg: ACL, 2022: 4916-4925.
- [35] 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*. Stroudsburg: ACL, 2019: 4568-4578.
- [36] MRINI K, DERNONCOURT F, BUI T, et al. Rethinking self-attention: An interpretable self-attentive encoder-decoder parser[EB/OL]. (2019-11-03) [2023-01-10]. <http://arxiv.org/abs/1911.03875>.
- [37] WU Y, LIN Z J, ZHAO Y, et al. Text-centered shared-private framework via cross-modal prediction for multimodal sentiment analysis[C]//*Findings of the Association for Computational Linguistics*. Stroudsburg: ACL, 2021: 4730-4738.
- [38] SABOUR S, FROSST N, HINTON G. Dynamic routing between capsules[C]//*Annual Conference on Neural Information Processing Systems*. New York: ACM, 2017: 3856-3866.
- [39] 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*. Stroudsburg: ACL, 2014: 27-35.
- [40] 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: ACL, 2014: 49-54.
- [41] 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*. Stroudsburg: ACL, 2017: 452-461.
- [42] LI X, BING L D, LAM W, et al. Transformation net-

works for target-oriented sentiment classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Stroudsburg: ACL, 2018: 946-956.

- [43] KINGMA D, BA J. Adam: A method for stochastic optimization[EB/OL]. (2015-05-07) [2023-01-10]. <http://arxiv.org/abs/1412.6980>.
- [44] WANG K, SHEN W Z, YANG 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: ACL, 2020: 3229-3238.
- [45] 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. Stroudsburg: ACL, 2019: 5678-5687.

作者简介



张换香 女,1979年9月出生于内蒙古包头市。现为上海大学计算机工程与科学学院博士生。主要研究方向为自然语言处理、情感分析。
E-mail: zhanghuanxiang@imust.edu.cn



彭俊杰 男,1977年1月出生于湖北省黄冈市。现为上海大学计算机工程与科学学院教授、博士生导师。主要研究方向为数据挖掘与分析、自然语言处理。在国内外发表学术论文130余篇。
E-mail: jjie.peng@shu.edu.cn