

# 基于BERT的双特征融合注意力的方面情感分析模型

李 锦<sup>1</sup>, 夏鸿斌<sup>1,2+</sup>, 刘 渊<sup>1,2</sup>

1. 江南大学 人工智能与计算机学院, 江苏 无锡 214122

2. 江苏省媒体设计与软件技术重点实验室, 江苏 无锡 214122

+ 通信作者 E-mail: hbxia@jiangnan.edu.cn

**摘要:** 方面情感分析旨在预测句子或文档中一个特定方面的情感极性, 现阶段大部分的研究都是使用注意力机制对上下文进行建模。然而, 目前情感分类模型在使用BERT模型计算表征之间的依赖关系抽取特征时, 大多未根据不同的语境背景考虑上下文信息, 导致建模后的特征缺乏上下文的语境信息。同时, 方面词的重要性未能得到充分的重视而影响模型整体分类的性能。针对上述问题, 提出双特征融合注意力方面情感分析模型(DFLGA-BERT), 分别设计了局部与全局的特征抽取模块, 充分捕捉方面词和上下文的语义关联。并将一种改进的“准”注意力添加到DFLGA-BERT的全局特征抽取器中, 使模型学习在注意力的融合中使用减性注意力以削弱噪声产生的负面影响。基于条件层规泛化(CLN)设计了局部特征和全局特征的特征融合结构来更好地融合局部和全局特征。在SentiHood和SemEval 2014 Task 4数据集上进行了实验, 实验结果表明, 与基线模型相比该模型在融入了上下文语境特征后取得了较明显的性能提升。

**关键词:** 情感分析; 自然语言理解; 准注意力机制; 上下文注意

**文献标志码:** A **中图分类号:** TP391

## Dual Features Local-Global Attention Model with BERT for Aspect Sentiment Analysis

LI Jin<sup>1</sup>, XIA Hongbin<sup>1,2+</sup>, LIU Yuan<sup>1,2</sup>

1. School of Artificial Intelligence and Computer, Jiangnan University, Wuxi, Jiangsu 214122, China

2. Jiangsu Key Laboratory of Media Design and Software Technology, Wuxi, Jiangsu 214122, China

**Abstract:** Aspect-based sentiment analysis aims to predict the sentiment polarity of a specific aspect in a sentence or document. Most of recent research uses attention mechanism to model the context. But there is a problem in that the context information needs to be considered according to different contexts when the BERT model is used to calculate the dependencies between representations to extract features by sentiment classification models, which leads to the lack of contextual knowledge of the modelled features. And the importance of aspect words is not given more attention, affecting the overall classification performance of the model. To address the problems above, this paper proposes a dual features local-global attention model with BERT (DFLGA-BERT). Local and global feature extraction modules are designed respectively to fully capture the semantic association between aspect words and context. Moreover, an improved quasi-attention mechanism is used in DFLGA-BERT, which leads to the model using minus attention in the fusion of attention to weaken the effect of noise on classification in the text. The feature fusion structure of local and global features is designed to better integrate regional and global features based on conditional layer normalization (CLN). Experiments are conducted on the SentiHood and SemEval 2014 Task 4

**基金项目:** 国家自然科学基金(61972182)。

This work was supported by the National Natural Science Foundation of China (61972182).

**收稿日期:** 2022-10-08 **修回日期:** 2023-01-09

datasets. Experimental results show that the performance of the proposed model is significantly improved compared with the baselines after incorporating contextual features.

**Key words:** sentiment analysis; natural language understanding; quasi-attention mechanism; contextual attention

基于方面的情感分析(aspect-based sentiment analysis, ABSA)旨在对文本如用户评论进行细粒度的情感分析<sup>[1]</sup>,是对文本中一个或多个方面的情感极性进行分类,在学术界受到广泛关注的同时也在工业界得到了广泛的实施。它通过亚马逊等网站上的在线评论提供的实时反馈,获取客户对特定产品或服务的意见。如图1所示,在句子“This book is a hardcover version, but the price is a bit high.”中“version(封面)”的情感是积极的,而“price(价格)”的情感是消极的。由于这两个方面词表达了相反的情感,仅仅分配一个句子级别的情感极性是不合适的。于是,Jo等人<sup>[2]</sup>和 Pontiki等人<sup>[3]</sup>提出了基于方面的情感分析(ABSA),旨在识别针对特定方面的细粒度情感极性。细粒度的方面情感分析根据方面目标实体数量的不同可以分为单目标方面情感分析任务和多目标方面情感分析任务。ABSA任务通常只有一个目标实体,而为了处理多个目标实体在预先给定的方面集合中对多个不同的方面产生不同的情感倾向的问题。Saeidi等人<sup>[4]</sup>又引入了基于目标方面的情感分析(target aspect-based sentiment analysis, TABSA)任务,旨在识别句子中一个指定实体(target, aspect term)的一个指定方面类别(aspect category)的情感。本文中提到的方面词、方面类别以及方面集合等概念较易混淆,特举例说明。例如“这块牛排的味道非常好吃,但价格有点贵。果汁既便宜又好喝”在给定的关于餐厅的目标方面集合{味道,服务,价格,位置,服务,装修等}中,方面词“牛排”“果汁”分别提到了它们在“价格”和“味道”两个方面类别中的情感倾向。为了全面地对句子进行细粒度的情感分析,本文分别针对一个实体的多个方面类别(ABSA任务)以及多个实体的多个方面类别

(TABSA任务)进行研究。

最初,研究者对ABSA任务使用神经网络模型的研究主要集中在使用深度神经网络。近年来,随着强大的自注意力机制<sup>[5-7]</sup>的出现,以自注意力机制为内核的Transformer<sup>[8]</sup>和BERT(bidirectional encoder representations from transformers)<sup>[9]</sup>因其强大的特征抽取能力被Sun等人<sup>[10]</sup>相继应用于ABSA,使模型的效果得到显著增强。Zeng等人<sup>[11]</sup>提出了一种不同于上述模型的局部上下文关注模型(local context focus, LCF),LCF模型使用句子序列中单词距离作为语义相关距离(semantic-relative distance, SRD)来得到局部上下文表示。该模型注意到一个方面的情感极性与接近自身的上下文词更相关,且远离方面词的上下文词可能会对特定方面极性的预测精确度产生消极影响。吴仁彪等人<sup>[12]</sup>参考图像处理领域的胶囊网络,将胶囊网络运用于情感分析任务上,有效降低了模型在特征迭代中信息丢失的问题。赵志影等人<sup>[13]</sup>提出了GCAN(GCN-aware attention network),引入了基于句法分析树的图卷积神经网络,以挖掘句子的句法特征和结构信息,从而更好地将全局信息整合到特定方面的向量表示中。Wu等人<sup>[14]</sup>用Context-Aware Self-Attention Networks作为模型的特征抽取层,提出了面向方面情感分析的CG-BERT(context-guided BERT for targeted aspect-based sentiment analysis)模型,获得了比基础的BERT模型更好的结果,上下文感知的自注意力网络的引入,使得深度学习模型能够更好地抽取输入序列的全局语义特征,在多个数据集上都取得了前沿性的成果。

尽管现有的模型在方面词情感分析任务上有着不俗的表现,但仍存在三个方面的不足。其一,在很多研究工作中,方面词的重要性未能得到充分的重

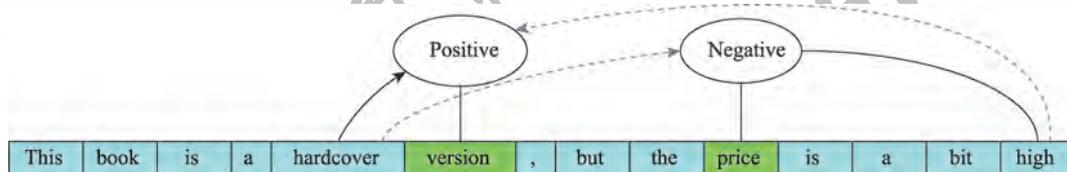


图1 上下文词对情感极性预测的不同影响

Fig.1 Different influences of context words on sentiment polarity prediction

视。Jiang 等人<sup>[15]</sup>在 Twitter 数据集上评估了一个情感分类器的效果,结果发现 40% 的分类错误都缘于没有考虑方面词的指示作用。这说明,如何充分使用方面词提供的位置特征和语义特征将是影响模型效果的重要因素。其二,现有的模型在提取文本特征的过程中未能充分发挥局部特征和全局特征的优势,方面词的情感极性更多的相关于与之在绝对位置和逻辑位置相近的词语,距离较远的词语往往会成为情感分类的干扰项。有些模型虽然发现了局部特征或全局特征的重要性,却未能将两者的特征进行很好的结合进而消除干扰项的不良影响。同时,在局部特征和全局特征进行融合的过程中使用拼接或相加的方式容易造成特征损失和特征覆盖的问题。其三,现阶段大部分模型使用 Softmax 激活函数来计算每个位置的注意力分数,由于 Softmax 函数的值域为  $[0,1]$ ,这导致每个特征都会对注意力的计算产生积极的影响,从而产生一定的噪声干扰。

针对上述不足,本文受 Zeng 等人<sup>[14]</sup>提出的局部注意力和 Yang 等人<sup>[16]</sup>提出的情境感知网络的启发,提出了基于 BERT 的双特征融合注意力模型(dual features local-global attention with BERT, DFLGA-BERT),借助 BERT 来获取高质量的语义编码,为突出方面词的重要性,本文根据上下文动态掩码(context dynamic mask, CDM)设计了局部特征抽取模块聚焦与方面词在语义上更相关的特征,在此基础上计算局部特征的注意力分数。同时,为了弥补局部特征容易忽略远距离特征的缺陷和挖掘多维度的全局特征,本文模型使用情景感知网络设计了全局特征抽取模块模型。在获取深度全局语义信息的同时尽可能地降低干扰项对情感分类的影响。同时受 Tay 等人<sup>[17]</sup>的启发,本文将“准”注意力融入到全局注意力的计算中,将注意力分数的值域由原来的  $[0,1]$  变换到  $[-1,1]$ 。通过训练线性层参数让模型学习在不同的上下文语境中分配合适的注意力,(正向、零、负向)分别对应(增加注意力、忽略、降低注意力),从而降低文本噪声对于模型的影响。在特征融合阶段,模型采用多个条件层规范化(conditional layer normalization, CLN)<sup>[18]</sup>的组合结构,在强调平衡局部特征和全局特征的同时,避免了融合阶段的特征损失,强化了模型的拟合能力。实验证明,DFLGA-BERT 在 SentiHood 和 SemEval 2014 Task 4 数据集上都取得了优异的成绩。同时,在消融实验中进行了子模块的有效性验证,也充分证明了模型主要组成部分的设计合理性。

## 1 本文模型

对于方面情感分类,为模型准备的输入序列通常由上下文序列和方面词序列组成,方面词序列通常为单个方面词或多个词组成的方面词组,这使得模型能够学习到与方面相关的信息。假设  $s = \{w_i | i \in [1, n]\}$  是一个包含方面的上下文序列,  $s' = \{w_i | i \in [1, m]\}$  是一个方面序列,其中  $s'$  是  $s$  的子序列,且  $m \geq 1$ 。如图 1,整个句子为上下文序列  $s$ ,而“version”和“price”是两个不同的方面序列  $s'$ 。本文模型的目的是预测  $s'$  在  $s$  中所表现的情感极性。模型总体架构如图 2 所示,宏观上可以分为三层:输入层、特征抽取层以及输出层。特征抽取层又可以分为局部特征抽取模块和全局特征抽取模块,分别用以获取方面词局部特征以及全局上下文语境动态加权特征。模型的介绍将围绕输入层、局部特征抽取模块、全局特征抽取模块以及输出层这四个部分展开。

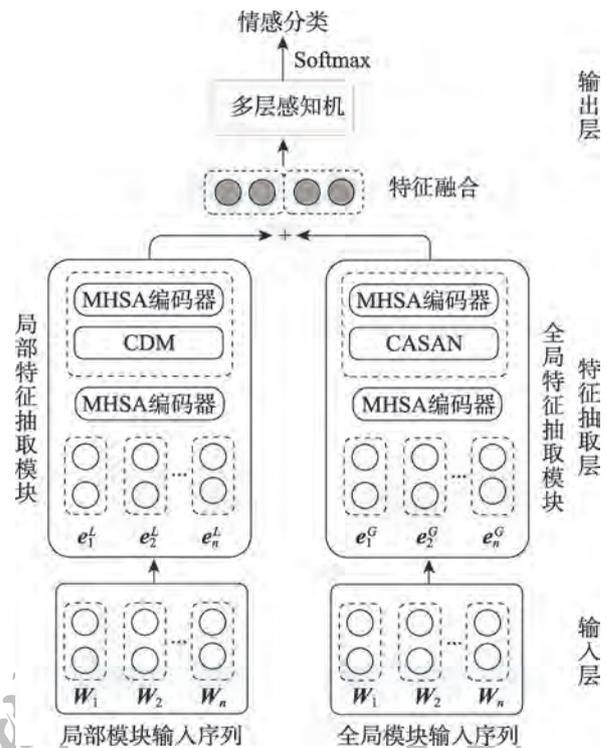


图2 DFLGA-BERT 模型整体架构

Fig.2 Overall architecture of DFLGA-BERT model

### 1.1 输入层

输入层是 DFLGA-BERT 模型的第一层,其目的是将上下文序列转化为向量化表示然后输入到模型中。在 DFLGA-BERT 模型中,为了充分发挥 BERT 在模型训练中的作用,本文将输入序列处理成“[CLS]+上下文序列+[SEP]+方面词序列+[SEP]”的形式。

其中,“[CLS]”和“[SEP]”是BERT使用的两种标记符号,“[CLS]”是特殊分类向量标记符,聚集了分类相关的信息,“[SEP]”是分隔符,在输入多个序列时用于分隔不同序列。将原输入序列处理成BERT在处理文本分类任务时所要求的输入格式,从而最大程度地发挥BERT的效果。

## 1.2 特征抽取层

### 1.2.1 局部特征抽取器

#### (1)词嵌入层

将句子输入到词嵌入层进行特征表示,词嵌入层可以将分好的词变为稠密的向量。在DFLGA-BERT中,由于BERT具有强大的表示能力,为了加快进程和获得更强性能,使用经过预训练的BERT词向量作为词嵌入层。假设 $L \in \mathbb{R}^{d_e \times |V|}$ 是BERT向量, $d_e$ 是嵌入向量的维度, $|V|$ 是词库的大小,每个上下文词 $W_i$ 对应一个向量 $V_i \in \mathbb{R}^{d_e \times |V|}$ 。

#### (2)MHSA编码器

经过词嵌入层的表示后,输入序列的隐层向量已经包含了所提及的语义特征,为了提取其中有助于判别方面情感的特征,使用多头注意力(multi-head self-attention, MHSA)来对特征进行提取。多头注意力与单头注意力的不同之处在于进行了多次计算而不仅仅算一次,这样可以允许模型在不同的表示子空间里学习到相关的信息。MHSA的主要思想是通过多个注意力函数来对每个单词计算注意力数值,将这多个注意力数值拼接在一起后进行激活函数运算,其结果即为输入序列的语义表示。令词嵌入层的输出为 $E^l$ ,单头注意力的计算方式为:

$$Q = E^l W^q \quad (1)$$

$$K = E^l W^k \quad (2)$$

$$V = E^l W^v \quad (3)$$

$$H(E^l) = \text{Softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (4)$$

式中, $W^q$ 、 $W^k$ 、 $W^v$ 均为权重矩阵, $W^q = \mathbb{R}^{d_h \times d_q}$ , $W^k = \mathbb{R}^{d_h \times d_k}$ , $W^v = \mathbb{R}^{d_h \times d_v}$ 。这三个矩阵中的权重属于模型的可训练参数。 $d_q$ 、 $d_k$ 、 $d_v$ 是矩阵维度,三者的值都是 $d_h/h$ 。其中, $d_h$ 是隐藏层维数, $h$ 是多头注意力机制中的头数。根据每个注意力头的 $H$ 计算结果,可得到整个MHSA的输出 $Out(E^l)$ :

$$H = H^0 \oplus H^1 \oplus \dots \oplus H^{h-1} \quad (5)$$

$$Out(E^l) = \text{Tanh}(HW^{\text{MHSA}}) \quad (6)$$

式中, $\oplus$ 表示向量的拼接操作; $W^{\text{MHSA}}$ 是一个向量矩

阵, $W^{\text{MHSA}} \in \mathbb{R}^{d_h \times d_h}$ ;Tanh是激活函数。通过以上步骤,多头注意力将基础的词嵌入向量编码为包含丰富信息的语义向量表示。

#### (3)语义相关距离SRD

语义相关距离(SRD)就是指同一句子中两个词互相之间的语义影响度,一般来说,两个单词之间的语义相关距离越近,则两个单词之间的语义相关程度越高。SRD一般分序列距离和依存距离两种,序列距离就是两个词在句子中位置的绝对距离。依存距离是指两个词在句法依存树中对应节点之间的最短距离。如果方面词由多个词语组成,则SRD被计算为方面词每个组件单词和输入单词之间的平均距离。SRD(序列距离)的计算过程如下:设输入的原上下文序列为 $W = \{w_1, w_2, \dots, w_i, w_{i+1}, \dots, w_{i+m-1}, w_{i+m}, \dots, w_n\}$ ,其中 $\{w_i, w_{i+1}, \dots, w_{i+m-1}\}$ 为方面词序列,对上下文序列中的任意单词 $w_j (1 \leq j \leq n)$ ,SRD的计算方法为:

$$SRD_j = \left| j - \left( i + \left\lfloor \frac{m}{2} \right\rfloor \right) \right| - \left\lfloor \frac{m}{2} \right\rfloor \quad (7)$$

经过实验测试,本文模型中使用序列距离的效果更好,因此选用序列距离作为词语的语义相关距离。

#### (4)上下文动态掩码CDM

通过计算句子中各词与方面词之间的语义相关距离,然后根据与方面词的SRD大小可以掩蔽掉一些对方面词的情感倾向无关的部分上下文词。给定全局上下文向量 $H$ 以及语义相关距离的阈值 $a$ 。具体流程为:对于输入的上下文序列,本文首先计算上下文词与方面词的语义相关距离,再根据阈值 $a$ 将上下文中与方面词的语义相关距离超过 $a$ 的词用全零矩阵进行覆盖,进而计算上下文的掩码向量 $M$ 。

$$m_i = \begin{cases} E, & SRD_i \leq a \\ O, & SRD_i > a \end{cases} \quad (8)$$

$$M = [m_1, m_2, \dots, m_n] \quad (9)$$

$$Out^{\text{local}} = Out(E^l) \circ M \quad (10)$$

式中, $O$ 和 $E$ 分别表示全零向量和全一向量; $O, E \in \mathbb{R}^{d_h}$ ,其中 $d_h$ 为上下文隐层向量的维度, $\circ$ 表示哈达玛积(即两个相同维度的矩阵相同位置两两相乘),通过式(10)将与方面词的语义相关距离超过 $a$ 的词用全零矩阵进行覆盖即得到局部特征嵌入层的输出 $Out^{\text{local}}$ 。最后通过注意力机制提取之后输出局部上下文对方面词的注意结果如式(11)。

$$O^L = \text{ATT}(Out^{\text{local}}) \quad (11)$$

### 1.2.2 全局特征抽取器

其中词嵌入部分、MHSA编码器部分和局部特

征抽取器相同,就不再赘述。

### (1) 上下文感知注意力编码

与传统的注意力编码不同,本文模型采用 Yang 等人<sup>[16]</sup>提出的上下文语境融合网络,其核心思想是在保持自注意力模型并行计算和特征抽取的基础上充分利用语料的内在信息去挖掘句子中词语之间的联系。传统注意力编码的计算<sup>[16]</sup>和深度全局语境融合网络编码<sup>[16]</sup>的区别见图3和图4所示。

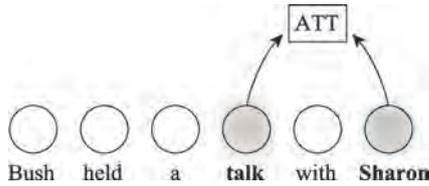


图3 传统注意力计算

Fig.3 Conventional attention calculating

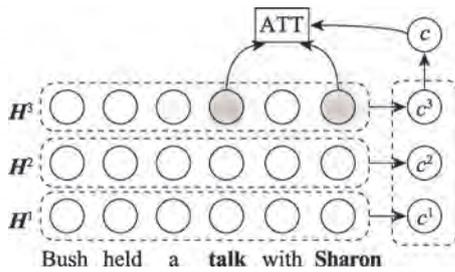


图4 深度全局上下文注意力计算

Fig.4 Deep-global context attention calculating

在传统的注意力编码中主要是建模了词与词之间的关联,得到整个输入序列对某一个词的权重序列。全局上下文注意是指用一个在层与层之间共享的向量  $c$  来建模序列的全局语义,帮助模型理解整句的整体语义。深度是指模型内部具有多层的隐藏层,在深度学习模型中通常会使用多层的隐藏层来建模多个方面的信息,例如每个词在不同方面的句法信息和语法信息, Yang 等人<sup>[16]</sup>的研究表明,底层的隐藏层通常会蕴含句法相关的信息,较高层的隐藏层用来捕获上下文相关的方面信息。在深度全局上下文注意力的计算当中结合了全局上下文和深度上下文,充分利用全局上下文和深度上下文的优势,跨层融合全局上下文,将词与词之间的词汇信息、句法信息、语义信息一起融入到特征向量中。

### (2) 深度全局上下文注意

深度全局上下文注意分数计算是基于上下文感知的注意力编码,为了缓解自注意力模型上下文信息的缺乏,模型将上下文的信息嵌入到编码层的查

询矩阵  $Q$  和键值矩阵  $K$  中,在避免了显著增加计算量的同时保持了原有的并行计算,具体的方式见图(12)所示:

$$\begin{bmatrix} \hat{Q}^h \\ \hat{K}^h \end{bmatrix} = \begin{bmatrix} 1 - \begin{bmatrix} \lambda_Q^h \\ \lambda_K^h \end{bmatrix} \\ \begin{bmatrix} \lambda_Q^h \\ \lambda_K^h \end{bmatrix} \end{bmatrix} \begin{bmatrix} Q^h \\ K^h \end{bmatrix} + \begin{bmatrix} \lambda_Q^h \\ \lambda_K^h \end{bmatrix} \begin{bmatrix} C^h U_Q \\ C^h U_K \end{bmatrix} \quad (12)$$

式中,  $C^h \in \mathbb{R}^{n \times d}$  是全局上下文特征矩阵;  $\{U_Q, U_K\} \in \mathbb{R}^{d \times d}$  是与之关联的可训练的参数矩阵,用于变换  $C^h$  线性层权重。为了更好地融入上下文的特征,本文用权重系数  $\{\lambda_Q^h, \lambda_K^h\}$  来对上下文的特征向量进行加权,通过这样的方式将上下文特征嵌入到全局特征中,以适应不同的语境对注意力分数计算的影响。Britz 等人<sup>[19]</sup>指出,当矩阵  $Q$  和矩阵  $K$  过大的时候可能会导致 Softmax 函数(式(4))的梯度消失,为了避免这种影响,本文使用一对门控标量  $\{\lambda_Q^h, \lambda_K^h\}$  来适当调节矩阵  $Q$  和矩阵  $K$  的大小,如式(13)所示:

$$\begin{bmatrix} \lambda_Q^h \\ \lambda_K^h \end{bmatrix} = \sigma \left( \begin{bmatrix} Q^h \\ K^h \end{bmatrix} \begin{bmatrix} V_Q^h \\ V_K^h \end{bmatrix} + C^h \begin{bmatrix} U_Q \\ U_K \end{bmatrix} \begin{bmatrix} V_Q^c \\ V_K^c \end{bmatrix} \right) \quad (13)$$

式中,  $\{V_Q^h, V_K^h\} \in \mathbb{R}^{d \times 1}$  和  $\{V_Q^c, V_K^c\} \in \mathbb{R}^{d \times 1}$  是可训练的线性层权重;  $\sigma(\cdot)$  表示 sigmoid 激活函数。

### (3) “准”注意力分数计算

在常规的注意力计算中,由于使用了 Softmax 函数注意力分数在  $[0,1]$  之间,也就是说,它只允许每个位置的隐层向量进行凸加权组合,对注意力分数的贡献只能是相加的,这会导致即使是跟目标方面词不相关的词也会得到正的注意力分数,从而对最终方面词的情感分类造成一定程度的干扰。因此本文模型将 Tay 等人<sup>[17]</sup>在 2019 年提出的“准”注意力<sup>[17]</sup>注入其中,同时允许加性注意和减性注意来克服这一限制。为了导出“准”注意力矩阵,本文首先定义了“准”上下文查询矩阵  $C_Q^h$  和“准”上下文键值矩阵  $C_K^h$ ,如式(14)所示:

$$\begin{bmatrix} C_Q^h \\ C_K^h \end{bmatrix} = C^h \begin{bmatrix} Z_Q \\ Z_K \end{bmatrix} \quad (14)$$

式中,  $\{Z_Q, Z_K\} \in \mathbb{R}^{d \times d_h}$  是用于转换原上下文矩阵线性层的权重变成“准”注意力的权重;得到全局注意力分数矩阵  $H^h(E^s)$ ,定义为式(15):

$$H^h(E^s) = \alpha \cdot \text{sigmoid} \left( \frac{f_\varphi(C_Q^h, C_K^h)}{\sqrt{d_h}} \right) \quad (15)$$

式中,  $\alpha$  是一个比例因子;  $f_\varphi(\cdot)$  是相似度度量函数,用来捕捉  $C_Q^h$  和  $C_K^h$  之间的相似度,本文使用点积函数

来参数化  $f_\varphi$ 。根据每个注意力头的  $H(E^g)$  计算结果, 可得到整个MHSA的输出  $Out(E^g)$ :

$$H(E^g) = H^0 \oplus H^1 \oplus \dots \oplus H^{h-1} \quad (16)$$

$$Out(E^g) = \text{Tanh}(HW^{\text{MHSA}}) \quad (17)$$

式中,  $\oplus$  表示向量的拼接操作;  $W^{\text{MHSA}}$  是一个向量矩阵,  $W^{\text{MHSA}} \in \mathbb{R}^{hd_h \times d_h}$ ;  $\text{Tanh}$  是激活函数。将全局上下文的权重系数  $\lambda_A^h$  的值表示为双向的, 得到式(18):

$$\lambda_A^h = 1 - (\beta \cdot \lambda_Q^h + \gamma \cdot \lambda_K^h) \quad (18)$$

式中,  $\{\beta, \gamma\}$  是控制权重的标量。用来控制“准”注意力权重系数  $\lambda_A^h$ , 进而调整全局特征在总体特征中权重占比。由式(13)可知  $\{\lambda_Q^h, \lambda_K^h\}$  的值域为  $[0, 1]$ 。当  $\beta = 1, \gamma = 1$  时,  $(\beta \cdot \lambda_Q^h + \gamma \cdot \lambda_K^h)$  的值域为  $[0, 2]$ , 则  $\lambda_A^h$  的值域为  $[-1, 1]$ 。此时, 局部特征和全局特征对模型总体的贡献为 1:1, 由此来平衡局部特征和全局特征。通过注意力机制进行提取之后得到全局上下文对方面词的注意力分数的结果, 全局特征抽取模块的输出可表示为式(19):

$$O^C = \lambda_A^h \text{ATT}(Out(E^g)) \quad (19)$$

### 1.3 特征融合

在局部特征和全局特征进行融合时, 普通的融合方式如线性层、拼接、和、积等, 很容易损失和覆盖特征信息, 最终导致模型的性能下降。

为了解决这个问题, 本文基于 CLN 设计了新的特征融合结构, 张龙辉等人<sup>[18]</sup>曾在对序列进行层规范化 (layer normalization, LN) 的同时将额外的条件  $c_1$  和  $c_2$  融入到 LN 的过程中, 来增强特征, 提出了 CLN。LN 和 CLN 的计算过程如下:

$$\text{LN}(x) = \frac{x - E[x]}{\sqrt{\text{Var}[x] + \varepsilon}} * \varphi + \eta \quad (20)$$

$$\text{CLN}(y, c_1, c_2) = \frac{y - E[y]}{\sqrt{\text{Var}[y] + \varepsilon}} * W_\varphi c_1 + W_\eta c_2 \quad (21)$$

式(20)为 LN 的计算公式, 其中  $E[x]$  为求  $x$  的均值,  $\text{Var}[x]$  为求  $x$  的方差,  $\varepsilon$  为保持分母不为零的极小常量,  $\varphi$  和  $\eta$  分别为缩放变量和平移变量。式(21)为 CLN 的计算公式, 其中  $W_\varphi, W_\eta$  为可训练参数,  $c_1$  和  $c_2$  为待融合条件。对比式(20), CLN 将条件  $c_1$  和  $c_2$  通过线性变换映射到两个不同的空间, 作为缩放变量  $\varphi$  和平移变量  $\eta$  加入到  $y$  的规范化过程中。

CLN 中变量的不同输入顺序, 同样会导致不同的融合结果, 能够体现变量间的方向性特征。本文利用多个 CLN 的组合结构来强化阶段间的特征传输

过程, 丰富语义特征的表达, 保留原有特征的信息以及方向性特征, 进而达到特征融合的目的。为了最大程度地保留特征抽取模块抽取的特征信息, 本文首先将全局特征作为辅助信息融入到局部特征中, 过程如下:

$$O^{L,G} = \text{CLN}(O^L, O^C, O^G) \quad (22)$$

然后将局部特征作为辅助信息融入到全局特征中, 过程如下:

$$O^{G,L} = \text{CLN}(O^G, O^L, O^L) \quad (23)$$

最后取式(22)和式(23)的均值得到双特征融合向量  $O_{\text{fusion}}$ , 如式(24)所示:

$$O_{\text{fusion}} = \frac{O^{L,G} + O^{G,L}}{2} \quad (24)$$

### 1.4 输出层

得到特征抽取层输出的双特征融合向量  $O_{\text{fusion}}$  后, 将其输入到线性分类层去做关于方面词的情感分类, 使用 Softmax 层预测情感极性。计算过程如式(25)和式(26)所示:

$$O_{\text{final}} = \text{MLP}(O_{\text{fusion}}) \quad (25)$$

$$p = \text{Softmax}(W_p O_{\text{final}} + b_p) \quad (26)$$

式中,  $\text{MLP}$  (multilayer perceptron) 表示多层感知机,  $p$  与情绪标签的维数相同, 而  $W_p \in \mathbb{R}^d$  和  $b_p \in \mathbb{R}^d$  分别是学习权重和偏置。

## 2 实验和结果

在本章中, 为了验证模型的有效性, 首先介绍了用于评估的数据集和用于比较的基线方法。然后, 针对细粒度情感分析任务 ABSA 以及它的子任务 TABSA 分别进行了实验, 并从不同的角度报告了实验结果。

### 2.1 实验环境及数据集

本文实验环境如下: CPU 为 Intel Core i7 8700K, GPU 为 GeForce GTX 1080Ti, 内存大小为 DDR4 16 GB, 开发环境为 win10 64 位操作系统, Pytorch1.5.0。

为了更加全面地对本文模型进行评估, 本文采用两个公开的英文数据集对 ABSA 任务和 TABSA 任务进行了实验评估。对于 TABSA 任务, 本文使用了由雅虎问答构建的 Sentihood 数据集 5, 它由 5 215 个句子组成, 其中 3 862 个句子包含一个目标方面, 1 353 个句子包含多个目标方面。对于 ABSA 任务, 本文使用的数据集是 SemEval 2014 Task 4, 这些数据集中包含多个方面项及其对应的情绪极性 (积极、中立或消

极)。表1显示了每个类别中的训练和测试实例的数量。此外,本文在参数选择过程中随机划分出20%的训练集作为开发集。本次实验所有复现的模型以及本文模型都使用了同样的数据集处理方式。

表1 SemEval数据集的统计信息  
Table 1 Statistics of SemEval datasets

Dataset	Positive		Neutral		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant	2 164	728	637	196	807	196
Laptop	994	341	464	169	870	128
Twitter	1 561	173	3 127	346	1 560	173

与学者 Pontiki 等人<sup>[3]</sup>的研究一样,本文为每个数据集定义了两个子任务:(1)方面词的检测;(2)基于方面的情感分析。对于方面词的检测,如果它是 TABSA 任务,则检测在目标实体为  $t$  的输入句子中是否提到了预先给定的方面集合中的方面类别  $a$  (如果为 None,则意味着没有提到),如果是 ABSA 任务则直接检测输入句子中是否提到了预先给定的方面集合中的方面类别  $a$ 。对于方面的情感分析,本文给出了模型存在的方面(即忽略掉在句子中没有提到的方面),让模型预测方面的情感极性(Sentihood 数据集是积极或消极,SemEval 2014 Task 4 数据集是积极、中立、消极或冲突)。

## 2.2 实验参数设置

在本文的实验中,对于方面词的检测(输入的句子中是否存在给定的方面,如不存在则为 None),本文采用分类任务常用的准确度  $Acc$ 、 $F1$  以及 AUC 值作为评价指标,对于基于方面的情感分析任务本文使用召回率  $F1$  和 AUC 值作为评价指标。本文使用预先训练好的 BERT 进行微调,特征向量的层数为 12,维度为 768。在本文的模型训练中,使用 Adam 作为优化器,初始学习率为  $2E-5$ ,batch\_size 为 4,dropout 概率为 0.1,epoch 为 20,在两个数据集中单次的训练时间为 19 h。

## 2.3 实验 1: TABSA 任务

基于目标方面的情感分析 TABSA 旨在识别句子中针对一个指定实体(target, aspect term)的一个指定方面(aspect category)的情感。例如:“这块牛排的味道非常好吃,但价格有点贵。果汁既便宜又好喝”在给定的关于餐厅的目标方面集合{味道,服务,价格,位置,服务,装修等}中表现出了对“牛排味道”“果汁味道”“果汁价格”这个方面的积极情感,对“牛排价

格”这个方面的消极情感,模型需要正确识别出目标实体在目标方面的情感极性。相较于关于方面的情感分析 ABSA 任务,TABSA 中可能出现的实体更多,预测的是实体-方面类别词对的情感极性。

为了综合评价本文模型,将它们与一系列基线和最先进的模型进行比较,如下所示:

(1)LR:一种具有多分类语法和位置标签的逻辑回归分类器。

(2)LSTM-Final<sup>[20]</sup>:Ruder 等人使用分层双向长期记忆网络(long short-term memory networks,LSTM)对评论中句子的相互依存关系进行建模。

(3)LSTM-Loc<sup>[21]</sup>:Tang 等人将目标位置相关联的状态融入特征的双向长短期记忆网络模型。

(4)LSTM+TA+SA<sup>[22]</sup>:Ma 等人在长短期记忆网络中引入了目标层注意和句子层注意的分层注意机制。情感相关概念的常识知识被纳入用于情感分类的深度神经网络的端到端训练中。

(5)SenticLSTM<sup>[22]</sup>:Ma 等人将 SenticNet<sup>[23]</sup>的外部信息引入到模型之中,是 LSTM+TA+SA 模型的升级版。

(6)Dmu-Entnet<sup>[24]</sup>:Liu 等人受机器阅读记忆增强模型的最新进展的启发,提出了一种新的双向 Entnet<sup>[25]</sup>体系结构,利用外部“内存链”和延迟的内存更新机制来跟踪实体。

(7)BERT-single<sup>[9]</sup>:Devlin 等人设计为通过联合限制所有层的左右上下文来预先训练来自未标记文本的深层双向表征。仅使用 BERT 对自然语言表征进行编码后进行目标方面情感分析。

(8)BERT-pair-QA-M<sup>[10]</sup>:Sun 等人利用 BERT 做神经表征编码的基础上,通过从方面词构造一个助句,将方面情感分析任务转换为问答。

(9)BERT-pair-NLI-B<sup>[10]</sup>:通过从方面词构造一个助句,将方面情感分析任务转换为自然语言推理。

(10)LCF-BERT<sup>[11]</sup>:Zeng 等人使用语义-关系距离(SRD)的局部上下文-关注设计来丢弃无关的情感词。

(11)LCFS-BERT<sup>[26]</sup>:Phan 等人结合词性嵌入、基于依赖关系的嵌入和上下文嵌入来提高方面抽取器的性能。提出了句法相对距离来淡化不相关词的负面影响,提高了方面情感分类器的准确性。

(12)Dual-MRC<sup>[27]</sup>:Mao 等人基于三元组的抽取任务,构造了两个机器阅读理解(machine reading comprehension,MRC)问题,并通过联合训练两个参数共享的 BERT-MRC 模型来解决端到端框架的所有子任务。

(13)DualGCN<sup>[28]</sup>:Li等人提出双重图卷积网络模型,分别捕获句法知识和语义相关性。

(14)CG-BERT<sup>[14]</sup>:Wu等人基于上下文动态分配自注意力权重提出上下文引导的BERT模型CG-BERT,可以针对不同的上下文背景自适应地调节上下文语境对方面词的影响。

(15)QACG-BERT<sup>[14]</sup>:Wu等人对CG-BERT模型进行改进,提出“准”注意力的CG-BERT模型,使得模型在学习的过程中支持值为负数的注意力分数。

(16)DFLGA-BERT:本文模型,基于双特征融合注意力分别设计了局部与全局的特征抽取模块,充分捕捉方面词和上下文的语义关联。并将一种改进的“准”注意力添加到DFLGA-BERT中,使模型学习在注意力的融合中使用减性注意力以削弱文本噪声产生的负面影响。

表2展示了模型在Sentihood数据集5上的实验结果。如表2所示,在普通的基线模型中,基于特征工程的逻辑回归分类器首先开始尝试方面级情感分析任务。时间最早,但在情感分类上的效果却出人意料得好,超过了许多不含BERT加持的基线模型。但在传统的机器学习方法中由于手工制作特征的复杂性,在不考虑人工成本的情况下依旧是一个不错的选择。在长短期记忆网络问世之后,Ruder等人<sup>[20]</sup>

表2 实验结果对比(TABSA任务)

Table 2 Comparison of experimental results (TABSA task) 单位:%

Model	Aspect categorization			Sentiment	
	Acc	F1	AUC	Acc	AUC
LR	—	39.3	92.4	87.5	90.5
LSTM-Final	—	68.9	89.8	82.0	85.4
LSTM-Loc	—	69.3	89.7	81.9	83.9
LSTM+TA+SA	66.4	76.7	—	86.8	—
SenticLSTM	67.4	78.2	—	89.3	—
Dmu-Entnet	73.5	78.5	94.4	91.0	94.8
BERT-single	73.7	81.0	96.4	85.5	84.2
BERT-pair-QA-M*	79.4	86.4	97.0	93.6	96.4
BERT-pair-NLI-B*	79.8	87.5	96.6	92.8	96.9
LCF-BERT*	79.0	87.4	95.8	92.8	94.6
LCFS-BERT*	79.2	87.1	97.0	92.5	95.0
Dual-MRC*	78.1	86.9	95.8	93.6	96.0
DualGCN*	79.6	87.7	97.0	93.4	96.2
CG-BERT	79.7	87.1	97.5	93.7	97.2
QACG-BERT*	79.9	88.6	97.3	93.8	97.8
DFLGA-BERT	<b>80.6</b>	<b>89.7</b>	<b>97.8</b>	<b>94.9</b>	<b>97.8</b>

注:每个数据集的最佳结果以粗体突出显示。本文用原论文公开的代码在本文实验环境下复现的结果用“\*”表示,其他结果引用自原论文。本文使用“—”来表示未报告的实验结果。

和Ma等人<sup>[22]</sup>先后利用分层双向LSTM以及将目标层注意和句子层的注意引入到LSTM中,模型整体在方面类别抽取子任务上的性能得到了提升。由于LSTM的结构比较简单,且未考虑方面词的有关信息,分类的效果依旧不是很理想。BERT-pair-QA-M和BERT-pair-NLI-B模型利用BERT模型在自然语言推理和问答任务上的优势,巧妙地通过构造助句的方法,将情感分析任务转换为问答任务和自然语言推理任务,在方面情感分析任务上的Acc指标提升了0.07~0.11。LCFS-BERT、CG-BERT、QACG-BERT也都设计了合适的网络结构和子模块来引导整个模型满足方面词情感分析任务的要求,充分发挥了BERT强大的优势,取得了令人瞩目的成绩。DFLGA-BERT则是凭借整合局部特征和深度全局上下文特征的优势,在情感分类的Acc和AUC上,都获得了比上述基线更理想的结果。证明了整体模型在解决方面词情感分析任务上的优秀能力。

## 2.4 实验2:ABSA任务

基于方面的情感分析任务ABSA目的是给定文本以及文本中可能出现的方面,预测文本是否提到该方面词,并预测该方面词在文本中的情感极性。

在ABSA任务中,将本文的DFLGA-BERT模型与在SemEval-2014数据集上表现优秀的几个模型进行比较。在方面检测任务中本文采用分类常用的Precision、Recall、F1值作为衡量指标,在方面情感分析方面,本文采用Acc作为评价指标,结果展示了模型在二分类(积极,消极)、三分类(积极,中性,消极)、四分类(积极,中性,消极,冲突)上的表现。实验结果如表3所示。

为了综合评价本文模型,将它们与一系列基线和最先进的模型进行比较,如下所示:

(1)XRCE<sup>[29]</sup>:Brun等人介绍了为SemEval-2014任务4开发的专门用于基于方面的情感分析的系统。该系统基于一个健壮的解析器,该解析器为不同的分类器提供信息,为不同的分类器提供专用于方面类别和方面类别极性分类的语言特征。

(2)NRC-Canada<sup>[30]</sup>:Kiritchenko等人基于SemEval-2014任务4设计的关于方面检测以及方面情感分析的分类器。

其他对比模型同实验1中模型,此处不予赘述。

表3展示了各个模型在SemEval-2014数据集上的表现。如表3所示,XRCE和NRC-Canada模型是Brun等人<sup>[29]</sup>和Kiritchenko等人<sup>[30]</sup>基于SemEval-2014

表3 实验结果对比(ABSA任务)

Table 3 Comparison of experimental

Model	results (ABSA task) 单位:%					
	Aspect categorization			Sentiment		
	Precision	Recall	F1	Binary	3-way	4-way
XRCE	83.23	81.37	82.29	—	—	78.1
NRC-Canada	91.04	86.24	88.58	—	—	82.9
BERT-single	92.78	89.07	90.89	93.3	86.9	83.7
BERT-pair-QA-M*	92.87	90.24	91.54	95.4	89.3	85.3
BERT-pair-NLI-B*	93.57	90.83	92.18	95.1	88.7	84.6
LCF-BERT	93.12	90.61	91.77	—	88.4	85.7
LCFS-BERT	93.03	90.22	92.17	—	88.9	86.2
DualGCN*	93.22	90.60	91.79	95.0	88.4	86.2
CG-BERT	93.02	90.00	91.49	94.3	89.9	85.6
QACG-BERT*	94.38	90.97	<b>92.64</b>	95.6	<b>90.1</b>	86.8
DFLGA-BERT	<b>95.26</b>	<b>92.04</b>	92.60	<b>96.0</b>	<b>90.1</b>	<b>87.8</b>

注:每个数据集的最佳结果以粗体突出显示。本文用原论文公开的代码在本文实验环境下复现的结果用“\*”表示,其他结果引用自原论文。本文使用“—”来表示未报告的实验结果。

任务4设计的关于方面类别检测以及方面情感分析的分类器,虽然没有用复杂的深度学习网络,但依靠手工编码特征依旧取得了不错的成绩,为之后的学者提供了可参考的实验结果。Sun等人<sup>[10]</sup>在利用BERT做神经表征编码的基础上,通过从方面词构造一个助句,将方面情感分析任务转换为问答和自然语言推理,充分利用BERT模型在分析句子对任务上天然的优势。在二分类的情感分类任务上达到了95.4%的准确率,该模型的性能已近乎达到其潜在最优效果的极限。LCFS-BERT则基于依赖关系提出了句法相对距离来淡化不相关词的负面影响,在三分类和四分类的任务上提升了1个百分点。本文模型在方面词检测的精确度和召回率上都表现出了比基线模型更优的实验结果,Macro-F1分数比Wu等人<sup>[14]</sup>的QACG-BERT模型略低4E-4,在情感分类任务上也相对占优,证明了整体模型在解决方面词情感分析任务上的优秀能力。

## 2.5 实验结果分析

综合两个实验来看,本文提出的基于双特征融合注意力的方面情感分析模型,针对局部特征抽取和全局特征抽取分别设计了局部特征和全局特征的抽取模块,能够显著提升BERT在方面词检测和方面情感分析任务上的表现,在TABSA任务和ABSA任务上都有比较好的效果。其一,本文将模型的输入处理成BERT在处理文本分类任务时所要求的输入格式,最大程度地利用BERT在语义抽取方面的优

势。其二,ABSA任务和TABSA任务中由于有来自方面词(aspect)和目标项(target)的额外信息,分类场景比普通的SA(sentiment analysis)任务更加复杂,直接微调传统的BERT预训练模型并不能实现模型性能的增长。本文根据不同的上下文场景,设计局部特征抽取模块减少注意力的感知范围,屏蔽距离较远的词对方面词产生的干扰,强化方面词的范围感知能力。同时设计全局特征抽取模块,去弥补局部特征抽取时忽略的距离较远但也和方面词比较相关的词。以内部的隐藏状态作为上下文的特征向量,在不增加外语义信息的条件下深度抽取全局语义信息。其三,本文在特征融合阶段采用多个条件层规范化的组合结构,丰富了语义特征的表达,避免了特征融合阶段的特征损失。在方面词的检测和情感分类上都取得了相比于之前的基线模型更好的结果。

## 2.6 实验参数分析

为了检验DFLGA-BERT模型学习到的全局参数的有效性,本文用训练好的DFLGA-BERT模型将测试集中的一个句子进行了自注意力和准注意力的可视化展示,如图5所示。在这个句子中本文正确地预测了“book version”的情感极性为正,high相对于方面词version也是一个积极情感的词,因此在自注意力的计算中high的自注意力分数和hardcover都很高。但是就全句上下文语境而言,high与price更相关,并且表示负面的情感极性,因此可以看到在“准”注意力分数上high的分数接近负1.0,这也会在本文最终的

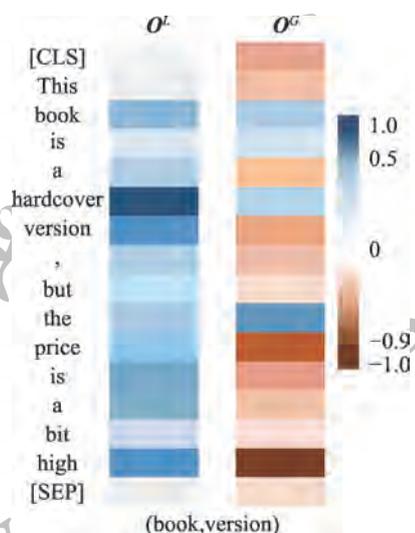


图5 测试集中对(book, version)词对的注意力权重示例

Fig.5 Example of attention weights extracted from test set of pair (book, version)

关注度中将对应的注意分数减去,最终得到 version 更多关于 hardcover 并且情感极性为正。由于准注意力的存在,很明显看到在特征融合之后部分词的注意力分数为负, Tay 等人<sup>[7]</sup>在论文“Compositional De-Attention Networks”第四章对自然语言处理任务的分析中得出结论,“准”注意力的使用能有效帮助模型学习在不同的上下文语境中分配合适的注意力(正向、零、负向),这些分别对应(增加注意力、忽略、降低注意力),因此方面词会降低对“high”的关注度。这也证明了 DFLGA-BERT 模型能有效地排除大部分的噪声干扰,能够忽略输入中一部分不相关的信息,有效提升后期情感分类任务的准确度。

### 2.7 自注意力和“准”注意力对比实验

从图 6 可知,在传统的自注意力(attention)中,方面词“book version”将注意力主要集中于上下文的“a hardcover”以及“a bit high”中,由于自注意力分数的计算受到 softmax 函数的限制,注意力分数的数值会分布在[0,1]区间内。通俗地理解为方面词会按照注意力分数的大小分配自身的自注意力,而所有上下文词的自注意力分数都大于 0,这表示它们或多或少都会“被关注”,因此才会造成一定的噪声问题。例如图中的例子,从自注意力的角度而言,“a hardcover”和“a bit high”的关注度都很高,那么“a bit high”也有可能到最后进行情感分类的时候被误认为是“book version”的观点项,从而导致模型预测结果不准确。而从“准”注意力(de-attention)的角度而言,“准”注意

力中学习了全局上下文的整体语义,它“学”到“a bit high”也饱含情感,但与“book version”在语境上不符,因此降低对它的关注度,最终会将注意力集中于“a hardcover”,将它作为情感分类的观点项,这正是本文模型所追求的。

### 2.8 SRD 参数的选择

为了探究不同的语义相关距离对 DFLGA 模型的影响,以及寻求 DFLGA 模型在各个数据集上的最佳 SRD 阈值参数,本文在两个数据集上进行了关于 SRD 阈值  $a$  和序列距离以及依存距离的实验,以评估在不同情况下的最佳语义相关距离和阈值  $a$ 。相应模型和数据集的 SRD 范围为 0 到 8,如果 SRD 阈值  $a$  为 0,则局部上下文等于方面词本身。结果如图 7 和图 8 所示,序列距离为 Seq,依存距离为 Deq。在本次的比较实验中,除 SRD 阈值外,所有的参数和超参数均一致。由于实验结果波动,本文在每个数据集上进行了多次实验,以得到的平均结果进行比较。由图 7 和图 8 可知,无论是在 Sentihood 数据集还是

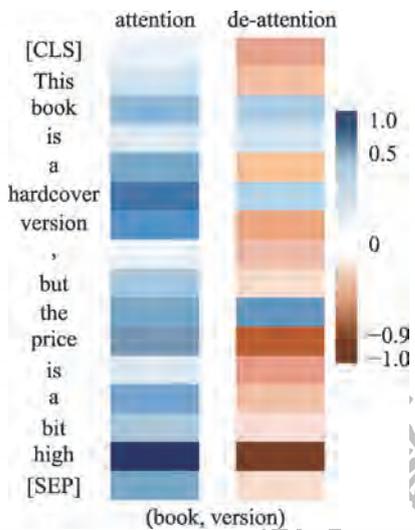


图 6 自注意力与“准”注意力可视化效果对比 (book, version)

Fig.6 Comparison of visualization between attention and de-attention (book, version)

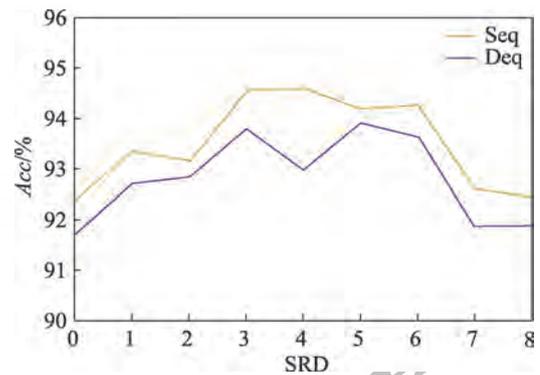


图 7 不同  $a$  下 DFLGA 的精度 (Sentihood)  
Fig.7 Accuracy of DFLGA under different  $a$  (Sentihood)

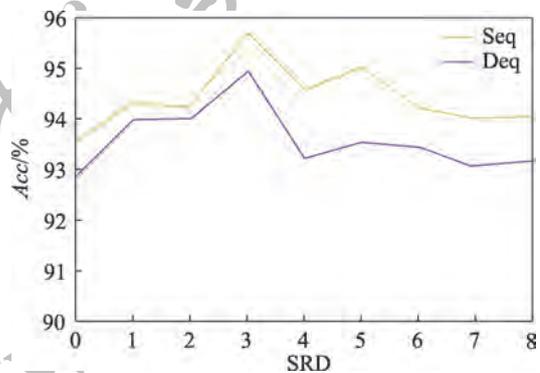


图 8 不同  $a$  下 DFLGA 的精度 (SemEval-2014)  
Fig.8 Accuracy of DFLGA under different  $a$  (SemEval-2014)

SemEval-2014 数据集中,使用序列距离的效果都优于使用依存距离。当阈值为3附近的时候取得本次实验的最佳结果。

## 2.9 消融实验

为了分析DFLGA-BERT模型每个组件的重要性,本文设计了DFLGA-BERT关于TABSA任务和ABSA任务的消融实验,实验结果列于表4和表5。

表4 DFLGA-BERT 消融实验(TABSA 任务)

Table 4 DFLGA-BERT ablation

experiment (TABSA task) 单位:%

Ablations	Aspect categorization			Sentiment	
	Acc	F1	AUC	Acc	AUC
w/o CDM	78.8	86.5	97.5	93.1	97.0
w/o L	79.0	87.0	97.5	93.0	96.8
w/o G	78.5	86.8	97.0	92.6	96.5
CDM&L&G	<b>80.6</b>	<b>89.7</b>	<b>97.8</b>	<b>94.9</b>	<b>97.8</b>

注:“w/o”表示去除,“G”代表全局特征提取器,“L”代表局部特征提取器,“CDM”代表上下文动态掩码。

表5 DFLGA-BERT 消融实验(ABSA 任务)

Table 5 DFLGA-BERT ablation

experiment (ABSA task) 单位:%

Ablations	Aspect categorization			Sentiment		
	Precision	Recall	F1	Binary	3-way	4-way
w/o CDM	93.27	90.12	92.24	95.6	89.2	86.0
w/o L	93.41	90.05	91.97	95.1	89.4	85.7
w/o G	93.83	89.78	92.06	94.4	88.7	85.4
CDM&L&G	<b>95.26</b>	<b>92.04</b>	<b>92.60</b>	<b>96.0</b>	<b>90.1</b>	<b>87.8</b>

注:“w/o”表示去除,“G”代表全局特征提取器,“L”代表局部特征提取器,“CDM”代表上下文动态掩码。

由表4和表5可以清楚地看到,缺少了任意一个模块,DFLGA的效果都会出现明显的下降。全局特征抽取模块对模型整体的效果影响最大,这表明了全局语义特征对自然语言理解任务的重要性,相对于普通句中方面词和情感指示词都相距较近,比较容易判断。最难判断的句子往往是情感指示词出现在模型意想不到的地方,甚至是局部特征抽取会直接忽略的地方,此时就只能依靠全局语义特征来进行判断。在模型整体中,局部语义特征和全局语义特征相辅相成,全局语义特征能够估计全面的语义但不够突出,局部特征能充分捕捉邻近的语义特征但容易忽略较远的词语,两个特征抽取模块都对DFLGA模型非常关键。通过上述消融实验,可以验证DFLGA各个子模块在本任务中的有效性。

## 3 结束语

本文针对当前方面词情感分析任务研究中存在的不足,设计了DFLGA模型。该模型首次将局部注意力模型LCF与全局特征抽取的深度上下文感知网络相结合,能够高效地抽取指定方面词的情感特征。全局和局部特征抽取器的设计使模型能够兼顾全面的同时聚焦局部特征,同时注入“准”注意力来削弱文本中噪声对分类的影响。为了加快模型的训练获得更好的结果,本文使用了BERT的预训练模型,在TABSA任务的Sentihood数据集5、ABSA任务的SemEval 2014中,进行了实验验证,取得了比之前的模型更好的结果。此外,本文还设计了消融实验以验证模型各组件的有效性。尽管DFLGA模型在多个数据集上都有不错的表现,但仍然存在一定的提升空间。在未来,将考虑使用额外的领域内语料资源来增强BERT在该领域的语义表示能力,从而帮助模型获得更优秀的情感判断能力。

## 参考文献:

- [1] YADAV A, VISHWAKARMA D K. Sentiment analysis using deep learning architectures: a review[J]. Artificial Intelligence Review, 2020, 53(6): 4335-4385.
- [2] JO Y, OH A H. Aspect and sentiment unification model for online review analysis[C]//Proceedings of the 4th ACM International Conference on Web Search and Data Mining, Hong Kong, China, Feb 9-12, 2011: 815-824.
- [3] PONTIKI M, GALANIS D, PAPAGEORGIOU H, et al. SemEval-2016 task 5: aspect based sentiment analysis[C]//Proceedings of the 2016 International Workshop on Semantic Evaluation. Stroudsburg: ACL, 2016: 19-30.
- [4] SAEIDI M, BOUCHARD G, LIAKATA M, et al. Sentihood: targeted aspect based sentiment analysis dataset for urban neighbourhoods[J]. arXiv:1610.03771, 2016.
- [5] TANG D, QIN B, LIU T. Aspect level sentiment classification with deep memory network[J]. arXiv:1605.08900, 2016.
- [6] WANG S, MAZUMDER S, LIU B, et al. Target-sensitive memory networks for aspect sentiment classification[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Jul 15-20, 2018. Stroudsburg: ACL, 2018: 957-967.
- [7] LIU J, ZHANG Y. Attention modeling for targeted sentiment [C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg: ACL, 2017: 572-577.
- [8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30, Long Beach, Dec 4-9, 2017: 5998-6008.
- [9] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [J]. arXiv:1810.04805, 2018.

- [10] SUN C, HUANG L, QIU X. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence [J]. arXiv:1903.09588, 2019.
- [11] ZENG B, YANG H, XU R, et al. LCF: a local context focus mechanism for aspect-based sentiment classification[J]. Applied Sciences, 2019, 9(16): 3389.
- [12] 吴仁彪, 乔晗, 贾云飞, 等. 基于胶囊网络的中长微博情感分析[J]. 信号处理, 2022, 38(8): 1632-1641.
- WU R B, QIAO H, JIA Y F, et al. Sentiment analysis of mid-length microblog based on capsule network[J]. Journal of Signal Processing, 2022, 38(8): 1632-1641.
- [13] 赵志影, 邵新慧, 林幸. 用于方面情感分析的结合图卷积神经网络的注意力模型[J]. 中文信息学报, 2022, 36(7): 154-163.
- ZHAO Z Y, SHAO X H, LIN X. GCN-aware attention networks for aspect-based sentiment analysis[J]. Journal of Chinese Information Processing, 2022, 36(7): 154-163.
- [14] WU Z, ONG D C. Context-guided BERT for targeted aspect-based sentiment analysis[C]//Proceedings of the 2021 AAAI Conference on Artificial Intelligence, Feb 2-9, 2021. Menlo Park: AAAI, 2021: 14094-14102.
- [15] JIANG L, YU M, ZHOU M, et al. Target-dependent Twitter sentiment classification[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: ACL, 2011: 151-160.
- [16] YANG B, LI J, WONG D F, et al. Context-aware self-attention networks[C]//Proceedings of the 2019 AAAI Conference on Artificial Intelligence, Hawaii, Jan 27-31, 2019. Stroudsburg: ACL, 2019: 387-394.
- [17] TAY Y, LUU A T, ZHANG A, et al. Compositional de-attention networks[C]//Advances in Neural Information Processing Systems 32, Vancouver, Dec 8-14, 2019: 6132-6142.
- [18] 张龙辉, 尹淑娟, 任飞亮, 等. BSLRel: 基于二元序列标注的级联关系三元组抽取模型[J]. 中文信息学报, 2021, 35(6): 74-84.
- ZHANG L H, YIN S J, REN F L, et al. BSLRel: a binary sequence labeling based cascading relation triple extraction model[J]. Journal of Chinese Information Processing, 2021, 35(6): 74-84.
- [19] BRITZ D, GOLDIE A, LUONG M T, et al. Massive exploration of neural machine translation architectures[J]. arXiv: 1703.03906, 2017.
- [20] RUDER S, GHAFARI P, BRESLIN J G. A hierarchical model of reviews for aspect-based sentiment analysis[J]. arXiv: 1609.02745, 2016.
- [21] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[J]. arXiv:1512.01100, 2015.
- [22] MA Y, PENG H, CAMBRIA E. Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, Feb 2-7, 2018. Stroudsburg: ACL, 2018: 5876-5883.
- [23] CAMBRIA E, PORIA S, BAJPAI R, et al. SenticNet 4: a semantic resource for sentiment analysis based on conceptual primitives[C]//Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, Osaka, Dec 9-14, 2016: 2666-2677.
- [24] LIU F, COHN T, BALDWIN T. Recurrent entity networks with delayed memory update for targeted aspect-based sentiment analysis[J]. arXiv:1804.11019, 2018.
- [25] HENAFF M, WESTON J, SZLAM A, et al. Tracking the world state with recurrent entity networks[J]. arXiv:1612.03969, 2016.
- [26] PHAN M H, OGUNBONA P O. Modelling context and syntactical features for aspect-based sentiment analysis[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACL, 2020: 3211-3220.
- [27] MAO Y, SHEN Y, YU C, et al. A joint training dual-MRC framework for aspect based sentiment analysis[J]. arXiv: 2101.00816, 2021.
- [28] LI R, CHEN H, FENG F, 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.
- [29] BRUN C, POPA D N, ROUX C. XRCE: hybrid classification for aspect-based sentiment analysis[C]//Proceedings of the 8th International Workshop on Semantic Evaluation, Dublin, Aug 23-24, 2014. Stroudsburg: ACL, 2014: 838-842.
- [30] KIRITCHENKO S, ZHU X, CHERRY C, et al. NRC-Canada-2014: detecting aspects and sentiment in customer reviews [C]//Proceedings of the 8th International Workshop on Semantic Evaluation, Dublin, Aug 23-24, 2014. Stroudsburg: ACL, 2014: 437-442.



李锦(1997—),男,湖北天门人,硕士研究生,主要研究方向为自然语言处理、机器学习。

**LI Jin**, born in 1997, M.S. candidate. His research interests include natural language processing and machine learning.



夏鸿斌(1972—),男,安徽庐江人,博士,副教授,CCF会员,主要研究方向为个性化推荐、自然语言处理、网络优化。

**XIA Hongbin**, born in 1972, Ph.D., associate professor, CCF member. His research interests include personalized recommendation, natural language processing and computer network optimization.



刘渊(1967—),男,江苏无锡人,博士,教授,博士生导师,CCF会员,主要研究方向为网络安全、社交网络。

**LIU Yuan**, born in 1967, Ph.D., professor, Ph.D. supervisor, CCF member. His research interests include network security and social network.