



(12) 发明专利申请

(10) 申请公布号 CN 113705238 A

(43) 申请公布日 2021. 11. 26

(21) 申请号 202110670846.6

(22) 申请日 2021.06.17

(71) 申请人 梧州学院

地址 543000 广西壮族自治区梧州市富民三路82号

(72) 发明人 庞光垚 陆科达 玉振明 彭子真 朱肖颖 黄宏本 莫智懿 农健 冀肖榆

(74) 专利代理机构 广州三环专利商标代理有限公司 44202

代理人 卢泽明

(51) Int. Cl.

G06F 40/30 (2020.01)

G06F 40/289 (2020.01)

G06N 3/08 (2006.01)

权利要求书4页 说明书15页 附图8页

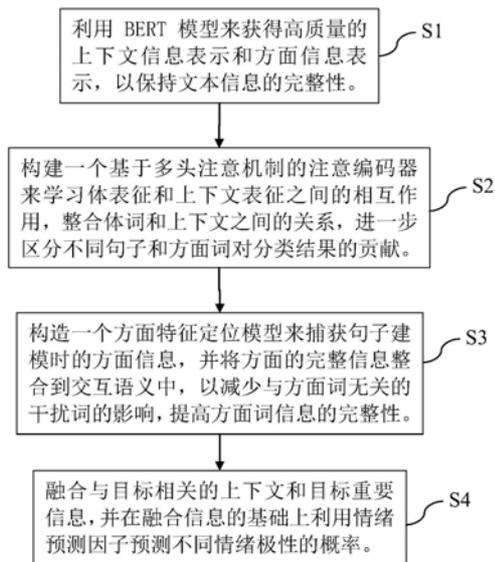
(54) 发明名称

基于BERT和方面特征定位模型的方面级情感分析方法及模型

(57) 摘要

本发明涉及一种基于BERT和方面特征定位模型的方面级情感分析方法及模型,该方法包括:首先利用BERT模型来获得高质量的上下文信息表示和方面信息表示,以保持文本信息的完整性;接着构建一个基于多头注意机制的注意编码器来学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,进一步区分不同句子和方面词对分类结果的贡献;然后构造一个方面特征定位模型来捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;最后融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率。能够更好地模拟上下文之间的隐式关系,更好地利用了方面词的信息和减少与方面词无关信息的干扰,从而获得了更高的精确度和宏F1。

CN 113705238 A



1. 一种基于BERT和方面特征定位模型的方面级情感分析方法,其特征在于,包括:

S1. 利用BERT模型来获得高质量的上下文信息表示和方面信息表示,以保持文本信息的完整性;

S2. 构建一个基于多头注意机制的注意编码器来学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,进一步区分不同句子和方面词对分类结果的贡献;

S3. 构造一个方面特征定位模型来捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;

S4. 融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率。

2. 根据权利要求1所述的方法,其特征在于,所述“利用BERT模型来获得高质量的上下文信息表示和方面信息表示”是指将预训练的BERT模型作为文本向量化机制,生成高质量的文本特征向量表示,其中所述BERT是一个预训练的语言表示模型,所述文本向量化机制是指将每个单词映射到高维向量空间,具体为:所述BERT模型通过使用深层的多层双向转换器编码器来生成文本的表示,同时通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示。

3. 根据权利要求2所述的方法,其特征在于,所述“构建一个基于多头注意机制的注意编码器来学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系”,是指基于多头注意机制实现在方面级情感分析的重要特征提取,提取上下文和目标的重要信息,具体为:首先引入转换编码器,所述转换编码器是一种基于多头注意机制和位置前馈网络的新型特征抽取器,能够在不同的特征表示子空间中学习到不同的重要信息和直接捕获序列中的长期相关性;然后通过转换编码器从BERT模型生成的方面信息表示和上下文信息表示中提取交互语义,确定对方面词的情感定性最为重要的上下文,同时以上下文的长期依赖信息和上下文感知信息作为位置前馈网络的输入数据,分别生成隐藏状态,并在均值池化操作后得到上下文交互的最终交互隐藏状态和语境与方面词的最终互动隐藏状态。

4. 根据权利要求3所述的方法,其特征在于,所述“通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示”,具体为:

所述BERT模型通过在输入序列的开头和结尾分别添加特殊的分词标记[CLS]和[SEP]对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,使输入序列的嵌入表示中包含了这三种嵌入的全部信息,最后在BERT模型中将注释文本和方面词分别转换为“[CLS]+注释文本+[SEP]”和“[CLS]+目标+[SEP]”得到上下文表示 $E_c$ 和方面表示 $E_a$ :

$$E_c = \{we_{[CLS]}, we_1, we_2, \dots, we_{[SEP]}\};$$

$$E_a = \{ae_{[CLS]}, ae_1, ae_2, \dots, ae_{[SEP]}\};$$

其中 $we_{[CLS]}$ ,  $ae_{[CLS]}$ 表示分类标记[CLS]的向量, $we_{[SEP]}$ 和 $ae_{[SEP]}$ 表示分隔符[SEP]的向量。

5. 根据权利要求4所述的方法,其特征在于,所述“通过转换编码器从BERT模型生成的

方面信息表示和上下文信息表示中提取交互语义,确定对方面词的情感定性最为重要的上下文,同时以上下文的长期依赖信息和上下文感知信息作为位置前馈网络的输入数据,分别生成隐藏状态,并在均值池化操作后得到上下文交互的最终交互隐藏状态和语境与方面词的最终互动隐藏状态”具体包括:

S201.通过组成转换编码器中多头注意机制的多个自注意机制从BERT模型生成的方面信息表示和上下文信息表示中映射出一个查询序列和一系列在并行子空间中捕捉不同的重要信息的键(K)值(V);

S202.通过注意力得分函数公式 $f_s(Q,K,V) = \sigma(f_e(Q,K))V$ 计算得到每个捕捉到的重要信息的注意力得分,其中 $\sigma(f_e(Q,K))$ 表示归一化指数函数, $f_e(Q,K)$ 是学习K和Q之间相关特征的能量函数,并通过以下公式计算:

$$f_e(Q,K) = QK^T / \sqrt{d_k}$$

其中 $\sqrt{d_k}$ 表示比例因子, $d_k$ 是查询Q和键向量K的维数;

S203.将上下文表示和方面表示输入到注意力得分函数公式 $f_{mh}(Q,K,V) = [a^1; a^2, \dots; a^i; \dots; a^{n-head}]W_d$ 中,分别获得上下文的长期依赖信息 $c_{cc}$ 和上下文感知信息 $t_{ca}$ ,以捕获上下文的长期依赖性,和确定哪些上下文对方面词的情感定性最为重要;其中, $a^i$ 表示第i个捕捉到的重要信息的注意力分数, $[a^1; a^2; \dots; a^i; \dots; a^{n-head}]$ 表示串联向量, $W_d$ 是注意力权重矩阵, $c_{cc} = f_{mh}(E_c, E_c)$ ,  $t_{ca} = f_{mh}(E_c, E_a)$ ;

S204.转换编码器分别以 $c_{cc}$ 和 $t_{ca}$ 作为位置前馈网络的输入数据,生成隐藏状态 $h_c$ 和 $h_a$ ,所述位置前馈网络PFN(h)是多层感知器的一种变体,所述位置前馈网络PFN(h)、 $h_c$ 和 $h_a$ 定义如下:

$$h_c = \text{PFN}(c_{cc})$$

$$h_a = \text{PFN}(t_{ca});$$

$$\text{PFN}(h) = \zeta(hW_1 + b_1)W_2 + b_2;$$

其中, $\zeta(hW_1 + b_1)$ 是校正后的线性单元, $b_1$ 和 $b_2$ 是偏置值, $W_1$ 和 $W_2$ 表示可学习的权重参数;

S205.在对隐藏状态 $h_c$ 和 $h_a$ 进行均值池化操作之后,得到上下文交互的最终交互隐藏状态 $h_{cm}$ 和语境与方面词的最终互动隐藏状态 $h_{am}$ 。

6.根据权利要求5所述的方法,其特征在于,所述方面特征定位模型的工作过程如下算法1:

算法1方面特征定位算法

**输入：** 上下文表示  $E_c$ ，方面词在句子中的位置  $i$ ，方面词的长度  $a_l$ ，批次  $b_s$ 。

**输出：**

```

1   for episode = 1, M do
2       选择行  $i + 1$  至  $i + 1 + a_l$  行，获得方面特征  $af$ ；
3       根据公式  $AF = Maxpooling(af, dim = 0)$  计算得到重要
        特征  $AF$ ；
4       对所有重要特征进行 dropout 运算，得到  $h_{af}$ ；
5   end for
8   util 在  $b_s$  批次的句子中完成提取重要特征.
```

具体为，该特征定位算法根据方面词的位置和长度从上下文表示  $E_c$  中提取方面词  $af$  的最重要的相关信息；同时利用最大池化从  $af$  中获取最重要特性  $AF$ ，然后对最重要特性  $AF$  执行一个 dropout 操作，并在上下文表示  $E_c$  中获得方面词的重要特征  $h_{af}$ 。

7. 根据权利要求6所述的方法，其特征在于，所述“融合与目标相关的上下文和目标重要信息，并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率和类别数量”具体包括：

S401. 使用向量拼接的方式将  $h_{cm}$ 、 $h_{am}$  和  $h_{af}$  连接起来得到整体特征  $r$ ：

$$r = [h_{cm}; h_{am}; h_{af}];$$

S402. 采用线性函数对  $r$  进行数据预处理，即：

$$x = W_u r + b_u, \text{ 其中, } W_u \text{ 是权重矩阵, } b_u \text{ 是偏置值;}$$

S403. 利用 softmax 函数计算出句子中方面词  $a$  的情感极性为  $p$  的概率  $Pr(a=p)$ ：

$$Pr(a=p) = \frac{\exp(x_p)}{\sum_{i=1}^C \exp(x_i)}, \text{ 其中, } p \text{ 表示候选情感极性, } C \text{ 是情绪极性的类别数量。}$$

8. 根据权利要求1-7中任一所述的方法，其特征在于，还包括：采用交叉熵和L2正则化作为损失函数进行训练，定义为：

$$loss = - \sum_{j=1}^D \sum_{i=1}^C y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2 ;$$

其中， $D$  表示所有训练数据， $j$  和  $i$  分别是训练数据样本和情绪类别的索引， $\lambda$  表示L2正则化的因子， $\theta$  表示模型的参数集， $y$  表示预测的情绪极性， $\hat{y}$  表示正确的情绪极性。

9. 一种方面级情感分析模型，其特征在於，包括：

文本向量化机制，利用BERT模型来获得高质量的上下文信息表示和方面信息表示，以保持文本信息的完整性；

方面级情感分析的特征提取模型，用于学习体表征和上下文表征之间的相互作用，整合体词和上下文之间的关系，以区分不同句子和方面词对分类结果的贡献，以及用于捕获

句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;

情感预测器,用于融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率。

10. 根据权利要求9所述的方面级情感分析模型,其特征在于,

所述BERT模型是一个预训练的语言表示模型,通过使用深层的多层双向转换器编码器来生成文本的表示,同时通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示;

所述方面级情感分析的特征提取模型,包括重要特征提取模型和方面特征定位模型;其中,所述重要特征提取模型是一个基于多头注意机制的注意编码器,用于学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,以区分不同句子和方面词对分类结果的贡献;所述方面特征定位模型用于捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;

所述情感预测器使用向量拼接的方式将最终交互隐藏状态、语境与方面词的最终互动隐藏状态、方面词的重要特征连接起来得到整体特征,然后采用线性函数对整体特征进行数据预处理,最后利用softmax函数计算出句子中方面词的情感极性为候选情感极性的概率。

## 基于BERT和方面特征定位模型的方面级情感分析方法及模型

### 技术领域

[0001] 本发明属于方面级情感分析技术领域,特别涉及一种基于BERT和方面特征定位模型的方面级情感分析方法及模型(ALM-BERT)。

### 背景技术

[0002] 电子商务是一个蓬勃发展的行业,对全球经济的重要性与日俱增。尤其是,随着社交媒体的迅速发展和网络社交平台的不断普及,越来越多的用户开始在各种网络平台上表达自己带情感的评论。这些评论反映了用户和消费者的情绪,为销售商和政府等提供了许多关于商品或服务质量的有价值的反馈信息。例如:在购买商品之前,用户可以在电子商务平台上浏览大量关于该商品的评论,以决定该商品是否值得购买。同样,政府和企业可以直接从互联网上收集大量的公众评论,分析用户的意见和满意度,进而满足他们的需求。因此,情感分析作为自然语言处理的一项基础性和关键性工作,引起了理论界和实践界的广泛关注。

[0003] 然而,常见的情感分析任务(如句子级情感分析)只能从整个句子中确定用户对产品或事件的情感极性(如积极、消极和中性),无法确定句子某一特定方面的情感极性。相比之下,方面级情感分析是一个更细粒度的分类任务,它可以识别句子中多个方面的情感极性。例如图9所示,提供了一些句子级情感分析和基于方面的情感分析的示例(有三个方面术语的消费者评论示例),我们可以从评论文本中看到,“它并没有附带任何安装在windows media之外的软件,但对于价格,我对它的条件和整体产品非常满意”,方面术语“软件”的情绪极性是负面的、“windows media”是中性的、“价格”和“非常满意”是正面的。

[0004] 在已有研究中,研究人员提出了各种方法来完成方面级情感分析任务。其中,大部分是基于有监督机器学习算法的方法,并且取得了一定效果。然而,这些统计方法需要在大规模数据集上精心设计手工特征,导致需要大量的人力和时间等成本。鉴于神经网络模型能够从评论文本中自动学习方面和上下文的低维表示,而不用依赖人工特征工程,使得近年来神经网络在方面级情感分析任务中受到了越来越多的关注。

[0005] 不幸的是,现有方法大多直接利用递归神经网络(RNN)或者卷积神经网络(CNN)来独立地建模,以及表达方面词(aspect word)及其上下文的语义信息,但却忽略了它们对关键部件的位置缺乏敏感性这一事实。在实践中,研究者已经证明,体词的情感极性与体词信息和语序信息高度相关,这意味着方面词的情感极性更容易受到距离方面词较近的上下文词的影响。此外,神经网络很难捕捉到方面词和上下文之间的长期依赖关系,从而造成有价值信息的丢失。

### 发明内容

[0006] 为解决现有技术中存在的上述问题,本发明提供了一种能够更好地利用方面词的信息和减少与方面词无关信息的干扰,从而获得了更高的精确度和宏F1的基于BERT和方面特征定位模型的方面级情感分析方法,以及基于该基于BERT和方面特征定位模型的方面级

情感分析方法的系统。

[0007] 为解决上述技术问题,本发明采用如下技术方案:

[0008] 本发明提供了一种基于BERT和方面特征定位模型的方面级情感分析方法,包括:

[0009] S1. 利用BERT模型来获得高质量的上下文信息表示和方面信息表示,以保持文本信息的完整性;

[0010] S2. 构建一个基于多头注意机制的注意编码器来学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,进一步区分不同句子和方面词对分类结果的贡献;

[0011] S3. 构造一个方面特征定位模型来捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;

[0012] S4. 融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率。

[0013] 进一步地,所述“利用BERT模型来获得高质量的上下文信息表示和方面信息表示”是指将预训练的BERT模型作为文本向量化机制,生成高质量的文本特征向量表示,所述BERT是一个预训练的语言表示模型,所述文本向量化机制是指将每个单词映射到高维向量空间,具体为:所述BERT模型通过使用深层的多层双向转换器编码器来生成文本的表示,同时通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示。

[0014] 其中,所述“通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示”,具体为:

[0015] 所述BERT模型通过在输入序列的开头和结尾分别添加特殊的分词标记[CLS]和[SEP]对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,使输入序列的嵌入表示中包含了这三种嵌入的全部信息,最后在BERT模型中将注释文本和方面词分别转换为“[CLS]+注释文本+[SEP]”和“[CLS]+目标+[SEP]”得到上下文表示 $E_c$ 和方面表示 $E_a$ :

[0016]  $E_c = \{we_{[CLS]}, we_1, we_2, \dots, we_{[SEP]}\}$ ;

[0017]  $E_a = \{ae_{[CLS]}, ae_1, ae_2, \dots, ae_{[SEP]}\}$ ;

[0018] 其中 $we_{[CLS]}$ ,  $ae_{[CLS]}$ 表示分类标记[CLS]的向量, $we_{[SEP]}$ 和 $ae_{[SEP]}$ 表示分隔符[SEP]的向量。

[0019] 进一步地,所述“构建一个基于多头注意机制的注意编码器来学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系”,是指基于多头注意机制实现在方面级情感分析的重要特征提取,提取上下文和目标的重要信息,具体为:首先引入转换编码器,所述转换编码器是一种基于多头注意机制和位置前馈网络的新型特征抽取器,能够在不同的特征表示子空间中学习到不同的重要信息和直接捕获序列中的长期相关性;然后通过转换编码器从BERT模型生成的方面信息表示和上下文信息表示中提取交互语义,确定对方面词的情感定性最为重要的上下文,同时以上下文的长期依赖信息和上下文感知信息作为位置前馈网络的输入数据,分别生成隐藏状态,并在均值池化操作后得到上下文交互

的最终交互隐藏状态和语境与方面词的最终互动隐藏状态。

[0020] 其中,所述“通过转换编码器从BERT模型生成的方面信息表示和上下文信息表示中提取交互语义,确定对方面词的情感定性最为重要的上下文,同时以上下文的长期依赖信息和上下文感知信息作为位置前馈网络的输入数据,分别生成隐藏状态,并在均值池化操作后得到上下文交互的最终交互隐藏状态和语境与方面词的最终互动隐藏状态”具体包括:

[0021] S201.通过组成转换编码器中多头注意机制的多个自注意机制从BERT模型生成的方面信息表示和上下文信息表示中映射出一个查询序列和一系列在并行子空间中捕捉不同的重要信息的键(K)值(V);

[0022] S202.通过注意力得分函数公式 $f_s(Q,K,V) = \sigma(f_e(Q,K))V$ 计算得到每个捕捉到的重要信息的注意力得分,其中 $\sigma(f_e(Q,K))$ 表示归一化指数函数, $f_e(Q,K)$ 是学习K和Q之间相关特征的能量函数,并通过以下公式计算:

$$[0023] \quad f_e(Q,K) = QK^T / \sqrt{d_k}$$

[0024] 其中 $\sqrt{d_k}$ 表示比例因子, $d_k$ 是查询Q和键向量K的维数;

[0025] S203.将上下文表示和方面表示输入到注意力得分函数公式 $f_{mh}(Q,K,V) = [a^1; a^2; \dots; a^i; \dots; a^{n-head}]W_d$ 中,分别获得上下文的长期依赖信息 $c_{cc}$ 和上下文感知信息 $t_{ca}$ ,以捕获上下文的长期依赖性,和确定哪些上下文对方面词的情感定性最为重要;其中, $a^i$ 表示第i个捕捉到的重要信息的注意力分数, $[a^1; a^2; \dots; a^i; \dots; a^{n-head}]$ 表示串联向量, $W_d$ 是注意力权重矩阵, $c_{cc} = f_{mh}(E_c, E_c)$ , $t_{ca} = f_{mh}(E_c, E_a)$ ;

[0026] S204.转换编码器分别以 $c_{cc}$ 和 $t_{ca}$ 作为位置前馈网络的输入数据,生成隐藏状态 $h_c$ 和 $h_a$ ,所述位置前馈网络PFN(h)是多层感知器的一种变体,所述位置前馈网络PFN(h)、 $h_c$ 和 $h_a$ 定义如下:

$$[0027] \quad h_c = \text{PFN}(c_{cc})$$

$$[0028] \quad h_a = \text{PFN}(t_{ca});$$

$$[0029] \quad \text{PFN}(h) = \zeta(hW_1 + b_1)W_2 + b_2;$$

[0030] 其中, $\zeta(hW_1 + b_1)$ 是校正后的线性单元, $b_1$ 和 $b_2$ 是偏置值, $W_1$ 和 $W_2$ 表示可学习的权重参数;

[0031] S205.在对隐藏状态 $h_c$ 和 $h_a$ 进行均值池化操作之后,得到上下文交互的最终交互隐藏状态 $h_{cm}$ 和语境与方面词的最终互动隐藏状态 $h_{am}$ 。

[0032] 进一步地,所述方面特征定位模型的工作过程如下算法1:

---

### [0033] 算法1 方面特征定位算法

---

**输入:** 上下文表示 $E_c$ , 方面词在句子中的位置  $i$ , 方面词的长

度  $a_l$ , 批次  $b_s$ .

**输出:**

```

1   for episode = 1, M do
2       选择行  $i + 1$  至  $i + 1 + a_l$  行, 获得方面特征  $af$ ;
[0034] 3       根据公式  $AF = Maxpooling(af, dim = 0)$  计算得到重要
        特征  $AF$ ;
4       对所有重要特征进行 dropout 运算, 得到  $h_{af}$ ;
5       end for
8       util 在  $b_s$  批次的句子中完成提取重要特征.
```

[0035] 具体为, 该特征定位算法根据方面词的位置和长度从上下文表示  $E_c$  中提取方面词  $af$  的最重要的相关信息; 同时利用最大池化从  $af$  中获取最重要特性  $AF$ , 然后对最重要特性  $AF$  执行一个 dropout 操作, 并在上下文表示  $E_c$  中获得方面词的重要特征  $h_{af}$ 。

[0036] 进一步地, 所述“融合与目标相关的上下文和目标重要信息, 并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率和类别数量”具体包括:

[0037] S401. 使用向量拼接的方式将  $h_{cm}$ 、 $h_{am}$  和  $h_{af}$  连接起来得到整体特征  $r$ :

[0038]  $r = [h_{cm}; h_{am}; h_{af}]$ ;

[0039] S402. 采用线性函数对  $r$  进行数据预处理, 即:

[0040]  $x = W_u r + b_u$ , 其中,  $W_u$  是权重矩阵,  $b_u$  是偏置值;

[0041] S403. 利用 softmax 函数计算出句子中方面词  $a$  的情感极性为  $p$  的概率  $Pr(a=p)$ :

[0042] 
$$Pr(a=p) = \frac{\exp(x_p)}{\sum_{i=1}^C \exp(x_i)}$$
 其中,  $p$  表示候选情感极性,  $C$  是情绪极性的类别数量。

[0043] 进一步地, 该基于 BERT 和方面特征定位模型的方面级情感分析方法还包括: 采用交叉熵和 L2 正则化作为损失函数进行训练, 定义为:

[0044] 
$$loss = - \sum_{j=1}^D \sum_{i=1}^C y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2$$
 ;

[0045] 其中,  $D$  表示所有训练数据,  $j$  和  $i$  分别是训练数据样本和情绪类别的索引,  $\lambda$  表示 L2 正则化的因子,  $\theta$  表示模型的参数集,  $y$  表示预测的情绪极性,  $\hat{y}$  表示正确的情绪极性。

[0046] 本发明又提供了一种方面级情感分析模型, 包括:

[0047] 文本向量化机制, 利用 BERT 模型来获得高质量的上下文信息表示和方面信息表示, 以保持文本信息的完整性;

[0048] 方面级情感分析的特征提取模型, 用于学习体表征和上下文表征之间的相互作用, 整合体词和上下文之间的关系, 以区分不同句子和方面词对分类结果的贡献, 以及用于捕获句子建模时的方面信息, 并将方面的完整信息整合到交互语义中, 以减少与方面词无关的干扰词的影响, 提高方面词信息的完整性;

[0049] 情感预测器,用于融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率。

[0050] 进一步地,所述BERT模型是一个预训练的语言表示模型,通过使用深层的多层双向转换器编码器来生成文本的表示,同时通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示;

[0051] 所述方面级情感分析的特征提取模型,包括重要特征提取模型和方面特征定位模型;其中,所述重要特征提取模型是一个基于多头注意机制的注意编码器,用于学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,以区分不同句子和方面词对分类结果的贡献;所述方面特征定位模型用于捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;

[0052] 所述情感预测器使用向量拼接的方式将最终交互隐藏状态、语境与方面词的最终互动隐藏状态、方面词的重要特征连接起来得到整体特征,然后采用线性函数对整体特征进行数据预处理,最后利用softmax函数计算出句子中方面词的情感极性为候选情感极性的概率。

[0053] 本发明有益效果:

[0054] 本发明通过上述技术方案,即可通过转换编码器能够更好地模拟上下文之间的隐式关系,并通过方面特征定位模型能够更好地利用了方面词的信息和减少与方面词无关信息的干扰,从而获得了更高的精确度和宏F1 (在不同长度的句子上的准确率和宏F1的均值分别高3.1%和6.56%),同时验证了BERT模型和方面信息在方面级情感分析任务中的可行性和有效性。

## 附图说明

[0055] 图1是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法实施实例的流程图;

[0056] 图2是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析系统实施实例的结构示意图;

[0057] 图3是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法在评测实验中丢弃率参数优化的实验结果曲线图;

[0058] 图4是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法在评测实验中学习率参数优化的实验结果曲线图;

[0059] 图5是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法在评测实验中L2正则化参数优化的实验结果曲线图;

[0060] 图6是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法与TD-LSTM验证实验不同长度源文本的ROUGE分数 (ROUGE-1) 的曲线图;

[0061] 图7是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法与TD-LSTM验证实验不同长度源文本的ROUGE分数 (ROUGE-2) 的曲线图;

[0062] 图8是本发明所述一种基于BERT和方面特征定位模型的方面级情感分析方法与

TD-LSTM验证实验不同长度源文本的ROUGE分数 (ROUGE-L) 的曲线图;

[0063] 图9是现有方面级情感分析的示例。

### 具体实施方式

[0064] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅用以解释本发明,并不用于限定本发明。

[0065] 如图1中所示,本发明实施例所述的一种基于BERT和方面特征定位模型的方面级情感分析方法,包括以下步骤:

[0066] 步骤S1. 利用BERT模型来获得高质量的上下文信息表示和方面信息表示,以保持文本信息的完整性;具体是指将预训练的BERT模型作为文本向量化机制,生成高质量的文本特征向量表示,所述BERT是一个预训练的语言表示模型,所述文本向量化机制是指将每个单词映射到高维向量空间,具体为:所述BERT模型通过使用深层的多层双向转换器编码器来生成文本的表示,同时通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示。

[0067] 步骤S2. 构建一个基于多头注意机制的注意编码器来学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,进一步区分不同句子和方面词对分类结果的贡献;具体是指基于多头注意机制实现在方面级情感分析的重要特征提取,提取上下文和目标的重要信息,具体为:首先引入转换编码器,所述转换编码器是一种基于多头注意机制和位置前馈网络的新型特征抽取器,能够在不同的特征表示子空间中学习到不同的重要信息和直接捕获序列中的长期相关性;然后通过转换编码器从BERT 模型生成的方面信息表示和上下文信息表示中提取交互语义,确定对方面词的情感定性最为重要的上下文,同时以上下文的长期依赖信息和上下文感知信息作为位置前馈网络的输入数据,分别生成隐藏状态,并在均值池化操作后得到上下文交互的最终交互隐藏状态和语境与方面词的最终互动隐藏状态。

[0068] 步骤S3. 构造一个方面特征定位模型来捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中,以减少与方面词无关的干扰词的影响,提高方面词信息的完整性;其中,所述方面特征定位模块是基于最大池化函数构建,即将所提取的方面词及其上下文隐藏特征分为多个区域,在每一个区域中选择最大值代表该区域,通过这种方式构建出方面特征定位模块(定位核心特征);所述方面特征定位模型的工作过程通过特征定位算法根据方面词的位置和长度从上下文表示 $E_c$ 中提取方面词af的最重要的相关信息;同时利用最大池化从af中获取最重要特性AF,然后对最重要特性AF执行一个dropout操作,并在上下文表示 $E_c$ 中获得方面词的重要特征 $h_{af}$ 。

[0069] 步骤S4. 融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率;具体为:通过使用向量拼接的方式将最终交互隐藏状态、语境与方面词的最终互动隐藏状态、方面词的重要特征连接起来得到整体特征,然后采用线性函数对整体特征进行数据预处理,最后利用softmax函数计算出句子中方面词的情感极性为候选情感极性的概率。

[0070] 如图2所示,本发明又提供了一种方面级情感分析模型,包括文本向量化机制100、方面级情感分析的特征提取模型200和情感预测器300。

[0071] 其中,所述文本向量化机制100是多角度文本向量化机制,利用BERT 模型来获得高质量的上下文信息表示和方面信息表示,以保持文本信息的完整性;所述BERT模型是一个预训练的语言表示模型,通过使用深层的多层双向转换器编码器来生成文本的表示,同时通过在输入序列的开头和结尾分别添加特殊的分词标记对给定的单词序列划分为不同的段,并为不同的片段生成标记嵌入、分段嵌入和位置嵌入,最后将注释文本和方面词分别转换、得到上下文信息表示和方面信息表示。

[0072] 所述方面级情感分析的特征提取模型200,用于学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,以区分不同句子和方面词对分类结果的贡献,以及用于捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中;具体为:该方面级情感分析的特征提取模型200包括重要特征提取模型和方面特征定位模型;其中,所述重要特征提取模型是一个基于多头注意机制的注意编码器,用于学习体表征和上下文表征之间的相互作用,整合体词和上下文之间的关系,以区分不同句子和方面词对分类结果的贡献;所述方面特征定位模型用于捕获句子建模时的方面信息,并将方面的完整信息整合到交互语义中;这样即可减少与方面词无关的干扰词的影响,从而提高方面词信息的完整性;

[0073] 所述情感预测器300用于融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率;具体通过使用向量拼接的方式将最终交互隐藏状态、语境与方面词的最终互动隐藏状态、方面词的重要特征连接起来得到整体特征,然后采用线性函数对整体特征进行数据预处理,最后利用softmax函数计算出句子中方面词的情感极性为候选情感极性的概率。

[0074] 一般来说,方面级情感分析是指以一个句子和一些预先定义的方面词作为输入数据,最后输出句子中每个方面词的情感极性的过程。在这里,我们用一些实际的评论例子来说明方面级情感分析任务。

[0075] 很明显,如表1所示,每个例句包含两个方面词,每个方面词有四种不同的情感极性,即积极、中性、消极和冲突。则方面级情感分析有如下定义:

[0076] 表1方面级情感分析的一些例子

评论	方面词	情感极性			
		积极	消极	中性	冲突
[0077] 所有的钱都花在了室内装饰上,没有一分钱给厨师。	室内装饰	√			
	厨师		√		
很棒的印度菜,服务也很棒。	印度菜	√			
	服务	√			
龙虾三明治的价格是24美元,虽然它很好吃,但远远不值这个价格。	龙虾三明治				√
	价格		√		

[0078] 定义一:在形式上,给出一个评论句子 $S = \{w_1, w_2, \dots, w_n\}$ ,其中 $n$ 为 $S$ 中的总字数。

一个方面词表 $A = \{a_1, \dots, a_i, \dots, a_m\}$ , 长度为 $m$ , 其中 $a_i$ 表示方面词表 $A$ 中的第 $i$ 个方面词, $A$ 是句子 $S$ 的子序列。 $P = \{p_1, \dots, p_j, \dots, p_c\}$ 表示候选情感极性, 其中 $C$ 表示情感极性的类别数量, $p_j$ 表示第 $j$ 个情感极性。

[0079] 问题: 方面级情感分析模型的目标是预测特定方面词最可能的情感极性, 它可以表述为:

[0080] 输入: 
$$\begin{cases} S = \{w_1, w_2, \dots, w_n\} \\ A = \{a_1, \dots, a_i, \dots, a_m\} \end{cases}$$

输出: 
$$p_k = \phi_{max}(a_i, p_j | S) \tag{1}$$

约束:  $A \in S, m \in [1, N]$

[0081] 其中 $\phi$ 表示一个函数, 该函数用于量化方面词 $a_i$ 与句子 $s$ 中的情感极性 $p_j$ 之间的匹配程度。最后, 模型输出匹配程度最高的情感极性作为分类结果。表2总结了该模型中的符号及其描述。

[0082] 表2: 使用的符号及其描述

[0083]

符号	描述
S	评论句子
A	方面词表
C	情感极性的类别数
$a_i$	方面词的长度
$p_j$	第 $j$ 个替代情绪极性
$E_c$	上下文表示
$E_a$	方面表示
$C_{cc}$	语境中的长期依赖信息
$t_{ca}$	面向方面词的上下文感知信息
$h_{af}$	方面词的重要特征
$h_{cm}$	上下文交互的最终交互隐藏状态
$h_{am}$	语境与方面词的最终互动隐藏状态
$f_s(.)$	注意力得分函数
$F_e(.)$	能量函数

[0084]

[0085] 本发明所述的基于BERT和方面特征定位模型的方面级情感分析方法为: 首先, 利用预训练BERT模型生成高质量的序列词向量, 为后续的步骤提供有效的支持; 然后, 在方面级情感分析的特征提取方法中, 基于多头注意力机制实现重要特征提取模块, 提取上下文和目标的重要信息; 接着, 提出方面特征定位模型综合考虑目标词的重要特征, 得到目标相

关特征;最后,融合与目标相关的上下文和目标重要信息,并在融合信息的基础上利用情绪预测因子预测不同情绪极性的概率。具体方法和原理如下:

#### [0086] 1、多角度文本向量化机制

[0087] 文本向量化机制主要将每个单词映射到高维向量空间。一般来说,文本向量化广泛应用Word2vec和Glove这两种基于上下文的词嵌入模型,并在方面级情感分析任务中取得了很大的成绩。然而,已有研究表明,这两个词嵌入模型不能在文本中获取足够的信息,从而导致分类准确率不足,性能降低。因此,一个高质量的词嵌入模型对提高分类结果的准确性有着重要的影响。

[0088] 实现方面级情感分析的关键在于能够有效理解自然语言处理,通常情况下高度依赖于大规模的高质量标注文本,所幸,BERT模型是一个能够有效利用无标注文本的语言预训练模型,该BERT模型使用随机屏蔽部分词汇的方式,利用深层的多层双向转换器编码器从海量未标注的文本中提取通用的自然语言识别模型,并进一步使用少量标注数据进行微调,可生成高质量的文本特征向量表示。受此启发,在本发明所提ALM-BERT方法中,对于给定的单词序列,会在输入序列的开头和结尾分别添加特殊的分词标记 [CLS]和[SEP],以便将序列划分为不同的段。即,通过该方式所输入的词嵌入向量中包括了为不同的片段生成标记嵌入、分段嵌入和位置嵌入等向量。具体地,将注释文本和方面词分别转换为“[CLS]+注释文本+[SEP]”和“[CLS]+目标+[SEP]”,最后得到上下文表示 $E_c$ 和方面表示 $E_a$ :

$$[0089] \quad E_c = \{we_{[CLS]}, we_1, we_2, \dots, we_{[SEP]}\} \quad (2)$$

$$[0090] \quad E_a = \{ae_{[CLS]}, ae_1, ae_2, \dots, ae_{[SEP]}\} \quad (3)$$

[0091] 其中 $we_{[CLS]}$ , $ae_{[CLS]}$ 表示分类标记[CLS]的向量, $we_{[SEP]}$ 和 $ae_{[SEP]}$ 表示分隔符[SEP]的向量。

#### [0092] 2、方面级情感分析的特征提取方法

[0093] 为了提取方面词及其上下文的隐藏特征,并着重考虑方面词蕴含的辅助信息,本文引入了一个转换器编码器,提出了方面词特征定位模块。其基本思想是对语境和目标词进行交互建模,以充分整合方面词和语境的信息。此外,通过获取上下文中方面词的特征信息可以提高情感分类的准确性。

##### [0094] 2.1重要特征提取模型

[0095] 转换编码器(transformer encoder)是一种基于多头注意机制和位置前馈网络的新型特征抽取器。它能够在不同的特征表示子空间中学习到不同的重要信息。不仅如此,转换编码器还可以直接捕获序列中的长期相关性,而且比递归神经网络和卷积神经网络更容易并行化,大大减少了训练时间。本发明通过转换编码器从BERT模型生成的方面信息表示和上下文信息表示中提取交互语义,确定对方面词的情感定性最为重要的上下文,同时以上下文的长期依赖信息和上下文感知信息作为位置前馈网络的输入数据,分别生成隐藏状态,并在均值池化操作后得到上下文交互的最终交互隐藏状态和语境与方面词的最终互动隐藏状态。

[0096] 直观地说,多头注意机制是由多个自注意机制(self-attention mechanism)组成的,它们可以映射到一个查询序列(Q)和一系列在在并行子空间中捕捉不同的重要信息的键(K)值(V)。注意力得分函数 $f_s(\cdot)$ 在自注意机制中计算过程如下所示:

$$[0097] \quad f_s(Q, K, V) = \sigma(f_e(Q, K))V \quad (4)$$

[0098] 其中 $\sigma(\cdot)$ 表示归一化指数函数, $f_e(\cdot)$ 是学习K和Q之间相关特征的能量函数,可以使用以下公式计算:

$$[0099] \quad f_e(Q, K) = QK^T / \sqrt{d_k} \quad (5)$$

[0100] 其中 $\sqrt{d_k}$ 表示比例因子, $d_k$ 是查询Q和键向量K的维数。

[0101] 多头注意机制的注意力得分函数 $f_{mh}(\cdot)$ 通过连接自注意机制的注意力分数获得:

$$[0102] \quad f_{mh}(Q, K, V) = [a^1; a^2; \dots; a^i; \dots; a^{n-head}] W_d \quad (6)$$

$$[0103] \quad a^i = f_s^i(Q, K, V) \quad (7)$$

[0104] 其中 $a^i$ 表示第i个捕捉到的重要信息的注意力分数,  $[a^1; a^2; \dots; a^i; \dots; a^{n-head}]$ 表示串联向量, $W_d$ 是注意力权重矩阵。

[0105] 如下公式(8) - (9)所示,将上下文表示和方面表示输入到多头注意机制中,以捕获上下文的长期依赖性,并确定哪些上下文对方面词的情感定性最为重要。

$$[0106] \quad c_{cc} = f_{mh}(E_c, E_c) \quad (8)$$

$$[0107] \quad t_{ca} = f_{mh}(E_c, E_a) \quad (9)$$

[0108] 其中, $c_{cc}$ 和 $t_{ca}$ 分别是上下文的长期依赖信息和上下文感知信息。

[0109] 然后,转换编码器分别以 $c_{cc}$ 和 $t_{ca}$ 作为位置前馈网络的输入数据,生成隐藏状态 $h_c$ 和 $h_a$ 。具体而言,所述位置前馈网络PFN(h)是多层感知器的一种变体。形式上,位置前馈网络PFN、 $h_c$ 和 $h_a$ 定义如下:

$$[0110] \quad h_c = \text{PFN}(c_{cc}) \quad (10)$$

$$[0111] \quad h_a = \text{PFN}(t_{ca}) \quad (11)$$

$$[0112] \quad \text{PFN}(h) = \zeta(hW_1 + b_1)W_2 + b_2 \quad (12)$$

[0113] 其中, $\zeta(hW_1 + b_1)$ 是校正后的线性单元, $b_1$ 和 $b_2$ 是偏置值, $W_1$ 和 $W_2$ 表示可学习的权重参数。

[0114] 在对 $h_c$ 和 $h_a$ 进行均值池化操作之后,得到上下文交互的最终交互隐藏状态 $h_{cm}$ 和语境与方面词的最终互动隐藏状态 $h_{am}$ 。

[0115] 2.2方面特征定位模型

[0116] 上述转换编码器捕获上下文的长期依赖性,并生成方面词和上下文之间的交互语义信息。为了突出不同方面词的重要性,本发明建立了方面词特征定位模型,其主要思想是从上下文特征表示中选择与方面词相关的信息,通过捕获包含方面信息的特征表示向量可以更好地整合方面信息,从而提高方面级情感分类的准确性。所述方面特征定位模型的工作过程如算法1所示:

---

### 算法1 方面特征定位算法

---

[0117] **输入:** 上下文表示 $E_c$ , 方面词在句子中的位置  $i$ , 方面词的长度  $a_l$ , 批次  $b_s$ .

**输出:**

1 for episode = 1, M do

2 选择行  $i + 1$  至  $i + 1 + a_1$  行, 获得方面特征  $af$ ;  
 3 根据公式  $AF = \text{Maxpooling}(af, dim = 0)$  计算得到重要  
 特征  $AF$ ;  
 [0118] 4 对所有重要特征进行 dropout 运算, 得到  $h_{af}$ ;  
 5 end for  
 8 util 在  $b_s$  批次的句子中完成提取重要特征.

[0119] 具体而言, 该特征定位算法根据方面词的位置和长度从上下文表示  $E_c$  中提取方面词  $af$  的最重要的相关信息; 同时利用最大池化从  $af$  中获取最重要特性  $AF$ , 如下所示:

[0120]  $AF = \text{Maxpooling}(af, dim = 0)$  (13)

[0121] 之后, 对最重要特性  $AF$  执行一个 dropout 操作, 并在上下文表示  $E_c$  中获得方面词的重要特征  $h_{af}$ 。

[0122] 3、情感预测器

[0123] 首先, 使用向量拼接的方式将  $h_{cm}$ 、 $h_{am}$  和  $h_{af}$  连接起来得到整体特征  $r$ :

[0124]  $r = [h_{cm}; h_{am}; h_{af}]$  (14)

[0125] 随后, 采用线性函数对  $r$  进行数据预处理, 即:

[0126]  $x = W_u r + b_u$  (15)

[0127] 其中,  $W_u$  是权重矩阵,  $b_u$  是偏置值。

[0128] 最后, 利用 softmax 函数计算出句子中方面词  $a$  的情感极性为  $p$  的概率  $Pr(a = p)$ :

[0129] 
$$Pr(a = p) = \frac{\exp(x_p)}{\sum_{i=1}^C \exp(x_i)}$$
 (16)

[0130] 其中,  $p$  表示候选情感极性,  $C$  是情绪极性的类别数量。

[0131] 整体而言, 本发明所述基于 BERT 和方面特征定位模型的方面级情感分析方法上述预测过程是一种端对端的运算过程。此外, 为了优化该方法的参数, 使预测的情绪极性  $y$  和正确的情绪极性  $\hat{y}$  之间的损失最小化, 还包括: 采用交叉熵和 L2 正则化作为损失函数进行训练, 定义为:

[0132] 
$$loss = - \sum_{j=1}^D \sum_{i=1}^C y_i^j \log \hat{y}_i^j + \lambda \|\theta\|^2$$
 (17)

[0133] 其中,  $D$  表示所有训练数据,  $j$  和  $i$  分别是训练数据样本和情绪类别的索引,  $\lambda$  表示 L2 正则化的因子,  $\theta$  表示模型的参数集,  $y$  表示预测的情绪极性,  $\hat{y}$  表示正确的情绪极性。

[0134] 4. 评测实验

[0135] 为了评价本发明所述基于 BERT 和方面特征定位模型的方面级情感分析方法及模型合理性和有效性, 通过以下评测实验进行了分析。

[0136] 4.1 数据集和评测指标

[0137] 我们在三个公开的英文评论数据集中构建了我们的相关评测实验。这三个数据集的详情如表 3 所示: 餐厅 (Restaurant) 和笔记本电脑 (Laptop) 数据集由 SemEval (参考文

献: Pontiki M, D Galanis, Pavlopoulos J, et al. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. Proceedings of International Workshop on Semantic Evaluation at, 2014.) 提供, 这些数据集的每一句都包含一些方面词和相应的情感极性, 标记为正、负、中性和冲突; twitter数据集由Li等人(参考文献: Li D, Wei F, Tan C, et al. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification[C]//Meeting of the Association for Computational Linguistics. 2014.) 收集的twitter 上的用户评论组成, 情绪极性被标记为正、负和中性。这三个数据集是目前流行的评论数据集, 已广泛应用于方面级情感分析任务中。

[0138] 表3数据集的统计信息

数据集	整体	积极	中性	消极	冲突
Laptop-Train	2373	994	870	464	45
Laptop-Test	654	341	128	169	16
Restaurant-Train	3699	2164	807	637	91
Restaurant-Test	1134	728	196	196	14
Twitter-Train	6248	1561	3127	1560	-
Twitter-Test	692	173	346	173	

[0141] 此外, 为了客观地评价本发明所述的基于BERT和方面特征定位模型的方面级情感分析方法及模型的性能, 我们采用了方面级情感分析任务中常用的评价指标, 即宏观F1 (macro-F1) 和准确度 (Acc)。定义为:

[0142]  $Acc = SC/N$  (18)

[0143] 其中, SC表示正确分类的样本数, N表示样本总数。一般来说, 精度越高, 模型的性能越好。

[0144] 此外, 宏观F1 (macro-F1) 用于真实反映模型的性能, 即精度和召回率的加权平均值。macro-F1根据以下公式计算:

$$[0145] \quad pre_{C_i} = \frac{T_{C_i}}{(T_{C_i} + FP_{C_i})} \quad (19)$$

$$[0146] \quad R_{C_i} = \frac{T_{C_i}}{(T_{C_i} + FN_{C_i})} \quad (20)$$

$$[0147] \quad macro-F1 = 1/C \left( \sum_{i=1}^C \left\{ \frac{2 * Pre_{C_i} * R_{C_i}}{(Pre_{C_i} + R_{C_i})} \right\} \right) \quad (21)$$

[0148] 其中, T是正确分类为情绪极性i的样本数, FP表示错误分类为情绪极性i的样本数, FN表示其情绪极性i被错误分类为其他情绪极性的样本数, C表示情绪极性的类别数,  $Pre_{C_i}$  是情绪极性i的精确度 (precision),  $R_{C_i}$  表示情绪极性i的召回率 (recall)。在我们的实验中, 为了更全面地评估我们的模型的性能, 我们将情感极性分为  $3C = \{positive,$

neutral,negative} 和4C={positive,neutral,negative,conflict}两类。

#### [0149] 4.2参数优化

[0150] 在模型的训练过程中,我们利用BERT模型来生成上下文和方面词的向量表示。具体来说,我们采用BERT模型的标准参数BERT<sub>BASE</sub>来完成模型训练。其中,在BERT<sub>BASE</sub>中的转换模块数量、隐藏神经元数量和自注意力的头数分别为12、768和12。此外,为了分析最优的超参数设置,我们提供了几个重要的超参数设置示例。

[0151] 首先,丢弃率(Dropout)是指在神经网络的训练过程中丢弃一些神经元的概率,用于解决过拟合和增强模型的泛化能力。其中,我们将dropout 的值初始化为0.3,然后以0.1的间隔搜索最佳值。实验结果如图3所示,当dropout为0.5时,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型在三个数据集上的精确度和F1值都是最好的。

[0152] 其次,学习率(learning rate)决定了目标函数是否以及何时收敛到局部极小值。在我们的实验中,我们使用Adam优化算法来更新模型的参数,并探索在 $[10^{-5}, 0.1]$ 范围内的最佳学习率参数。如图4所示,当学习速率为 $2*10^{-5}$ 时,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型的性能最好。

[0153] 最后,L2正则化参数是一个可以防止模型过度拟合的超参数。如图5 所示,当L2正则化参数的值设置为0.01时,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型的性能最佳;同时,我们通过 Glorot参数初始化方法对模型的权重进行初始化,并将批次大小设置为 16,总共训练10个迭代次数。

#### [0154] 4.3对比算法

[0155] 为了验证本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型的有效性,将该基于BERT和方面特征定位模型的方面级情感分析方法及模型与许多流行的方面级情感分析模型进行了比较,如下所示:

[0156] • TD-LSTM是一种经典的分类模型,它将方面词及其上下文的相关信息集成到基于LSTM的分类模型中,提高了分类的准确性。

[0157] • ATAE-LSTM是一种分类模型,将方面词的嵌入表示作为句子的嵌入表示输入到模型中,再应用注意力机制计算权重以实现高精度情感分类。

[0158] • MemNet是一个数据驱动的分类模型,它利用多个基于注意力机制的模型来捕获每个上下文单词的重要性,以完成情感分类。

[0159] • IAN是一种交互式的注意力网络,分别对方面词及其上下文进行建模,并生成目标和上下文的关联表示。

[0160] • RAM构建了一个基于多注意力机制的框架来捕获文本中的远距离特征,增强了模型的表示能力。

[0161] • TNet利用双向LSTM生成上下文和方面词的隐藏表示。再用CNN层代替注意力机制,从隐藏表示中提取出重要特征。

[0162] • Cabasc利用两种注意力增强机制,分别关注方面词和语境,并综合考虑了语境和方面词之间的相关性。

[0163] • AOA构建了一个双重注意力模块,将情感词与方面词联系起来。此外,双重注意力模块自动生成从方面到文本和从文本到方面的相互注意力权重。

[0164] • MGAN是一种多粒度的注意力模型,从粗到细捕捉方面词与上下文之间的交互信

息。

[0165] • AEN-BERT是一个基于注意力机制和BERT的模型,在方面级情感分析任务中表现出良好的性能。

[0166] • BERT-base是一种基于预训练BERT的方面级情感分析模型,具有完整的连接层和用于分类任务的softmax层。

[0167] 为了更准确地度量模型的性能,我们对AOA、IAN和MemNet模型进行了扩展,用BERT模型替换这些模型的嵌入层,得到了AOA-BERT、IAN-BERT 和MemNet-BERT模型。rest模型的结构与本文描述的一致。

[0168] 4.4评测实验分析

[0169] 如下表4所示,显示了情绪极性C=3时的情绪分类结果。我们可以很容易地从表中观察到,BERT-based(基于BERT预训练的方面级情感分析方法)的准确性和宏观F1显著高于基于glove和word2vec方法的模型。特别是对于餐厅数据集,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型的精度和宏F1分别比经典的IAN模型高12.77%和30.97%。这表明BERT能够更好地表达文本的语义和语法特征,同时本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型在这三个数据集上取得了最好的分类性能。具体来说,在餐厅数据集中,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法与AEN方法相比,精确度和宏F1分别提高了4.2%和8.81%。此外,不难发现,在笔记本电脑数据集上,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法的分类准确率和宏F1分别比BERT-base模型高3.29%和3.15%,这说明本发明中的方面特征定位模块在方面级情感分析中起到了积极的作用。

[0170] 表4各个对比方法的实验评测结果

[0171]

词嵌入	对比方法	笔记本数据集		餐厅数据集		推特数据集	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
Embedding	LSTM	0.6144	0.4401	0.7304	0.533	0.6474	0.6058
	ATAE-LSTM	0.6019	0.4909	0.7375	0.5725	0.6864	0.6501
	Cabasc	0.6301	0.5297	0.7241	0.5245	0.6171	0.5657
	IAN	0.6191	0.4671	0.7268	0.4897	0.6618	0.6251
	MGAN	0.5878	0.4264	0.7179	0.4974	0.6373	0.5856
	MEMNET	0.7915	0.7576	0.8241	0.7313	0.6936	0.6772
	RAM	0.5956	0.4308	0.7152	0.4656	0.6358	0.5998
	T-Net	0.7022	0.6404	0.7688	0.6269	0.6994	0.6827
BERT	IAN-BERT	0.7696	0.7179	0.808	0.722	0.7269	0.7048
	AOA-BERT	0.7774	0.7407	0.7341	0.7173	0.7341	0.7173
	Memnet-BERT	0.7915	0.7576	0.8241	0.7313	0.6936	0.6772
	AEN-BERT	0.7947	0.7544	0.8125	0.7069	0.7168	0.7052
	BERT-base	0.768	0.7288	0.8375	0.7613	0.7442	0.7271

[0172]	LCF-bert	0.7837	0.7441	0.8509	0.7894	0.7254	0.7113
	ALM-BERT	0.8009	0.7603	0.8545	0.795	0.7413	0.73

[0173] 我们从捕获评论文本中的长期依赖关系的角度,在不同长度的文本上构建了系列验证实验。

[0174] 如图6-8所示,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型总体上比TD-LSTM获得了更高的精确度和宏F1,这意味着我们构建的转换编码器比基于LSTM的编码器能够更好地模拟上下文之间的隐式关系。此外,如下面的曲线图7所示,我们还注意到ALM-BERT 模型与AEN相比,在不同长度的句子上的准确率和宏F1的均值分别高3.1%和6.56%,原因是本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型比AEN更好地利用了方面词的信息,减少与方面词无关信息的干扰。

[0175] 综上所述,这些实验表明,本发明所述基于BERT和方面特征定位模型的方面级情感分析方法及模型能够获得更高的准确率和宏F1,进一步验证了BERT模型和方面信息在方面级情感分析任务中的可行性和有效性。

[0176] 以上所述是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

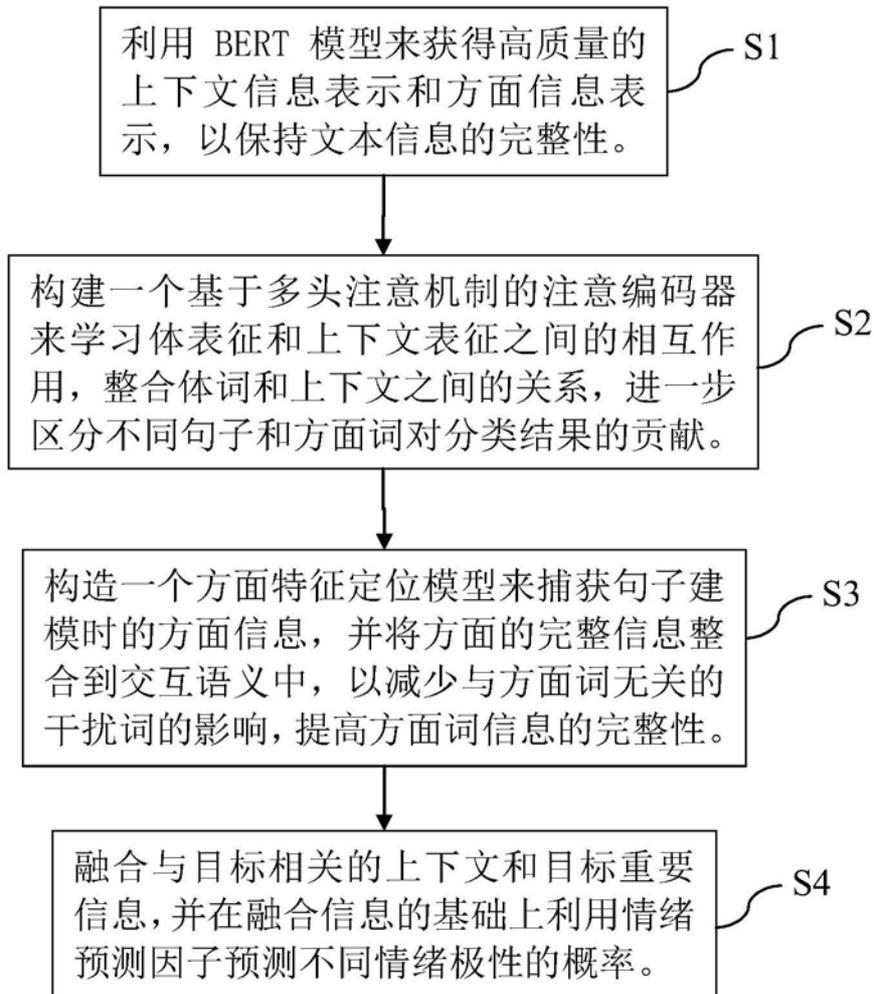


图1

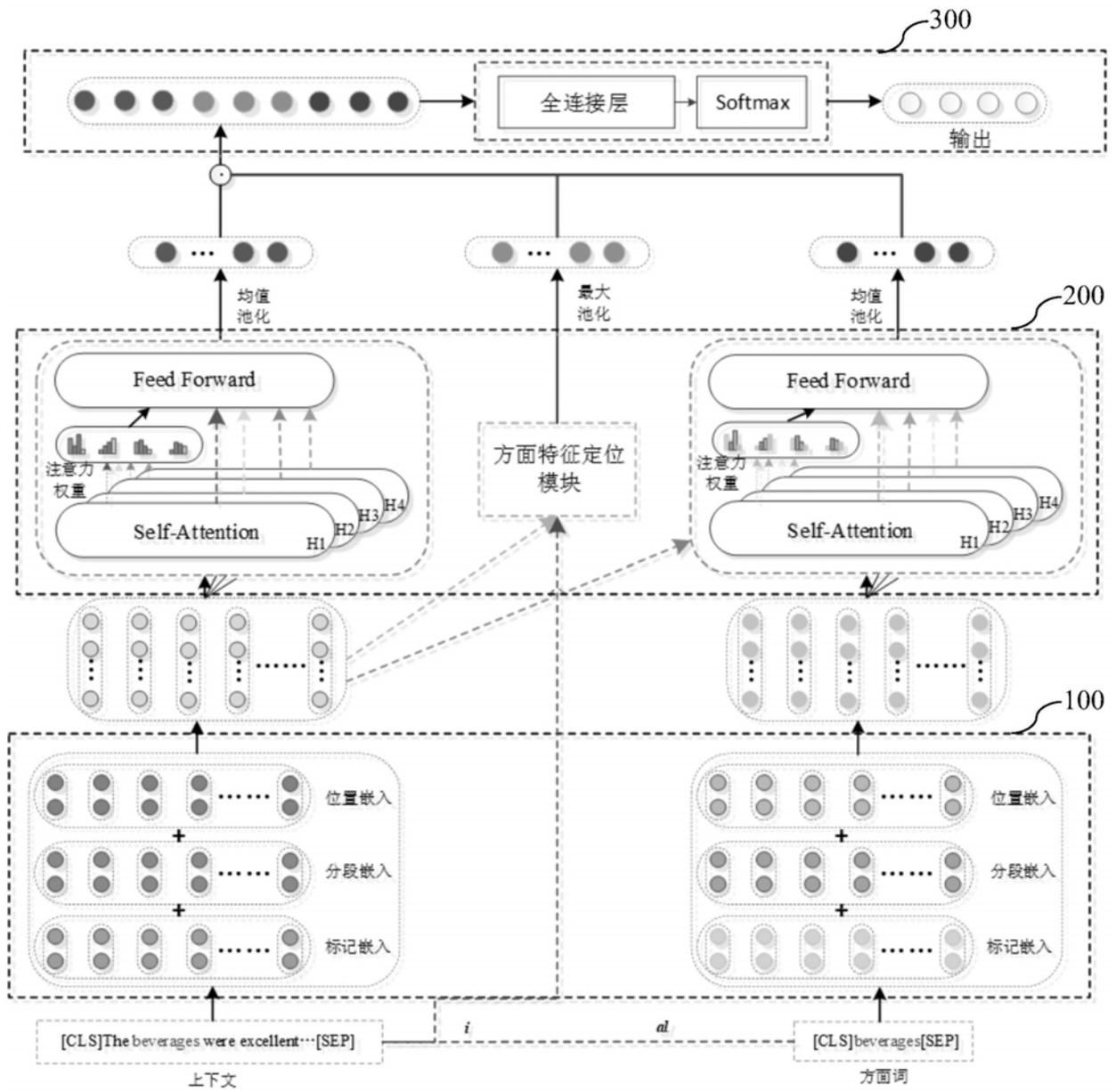


图2

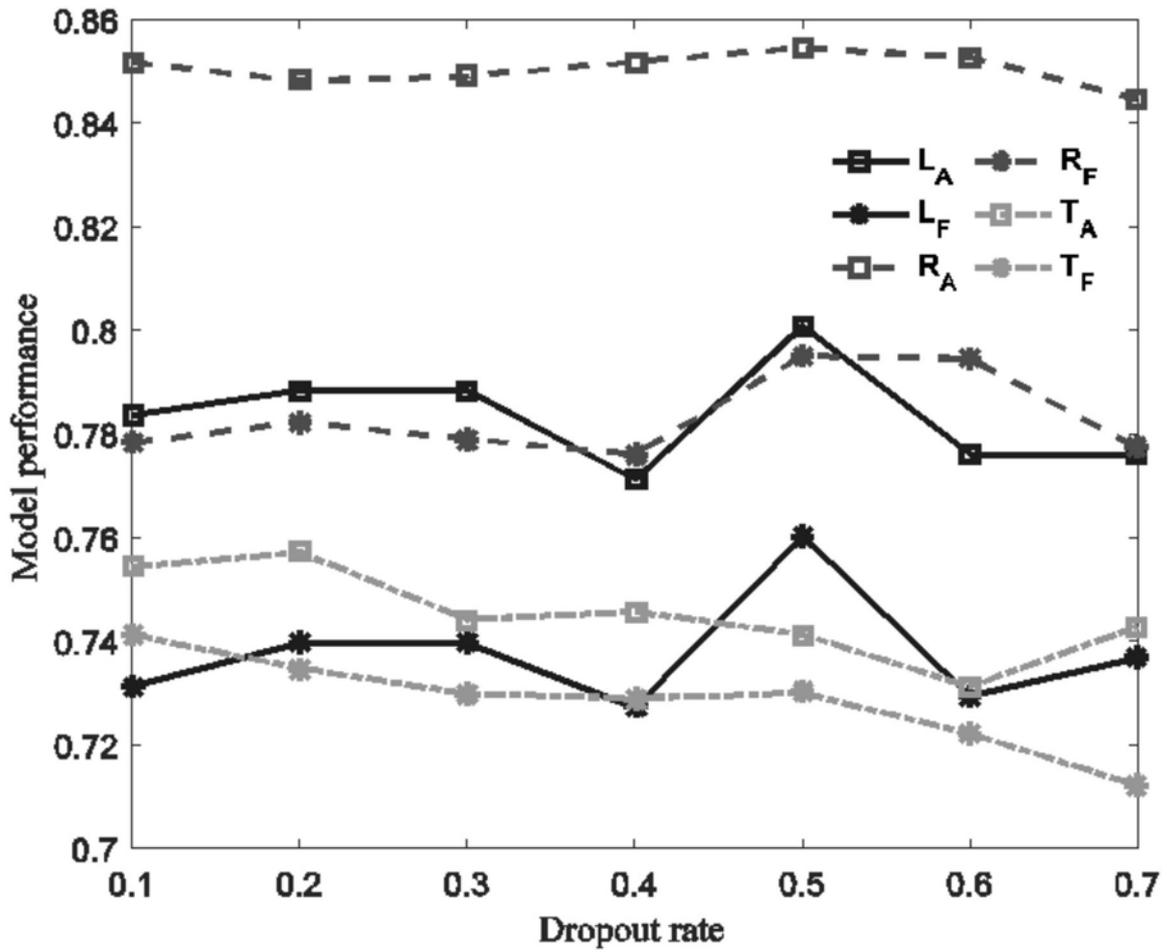


图3

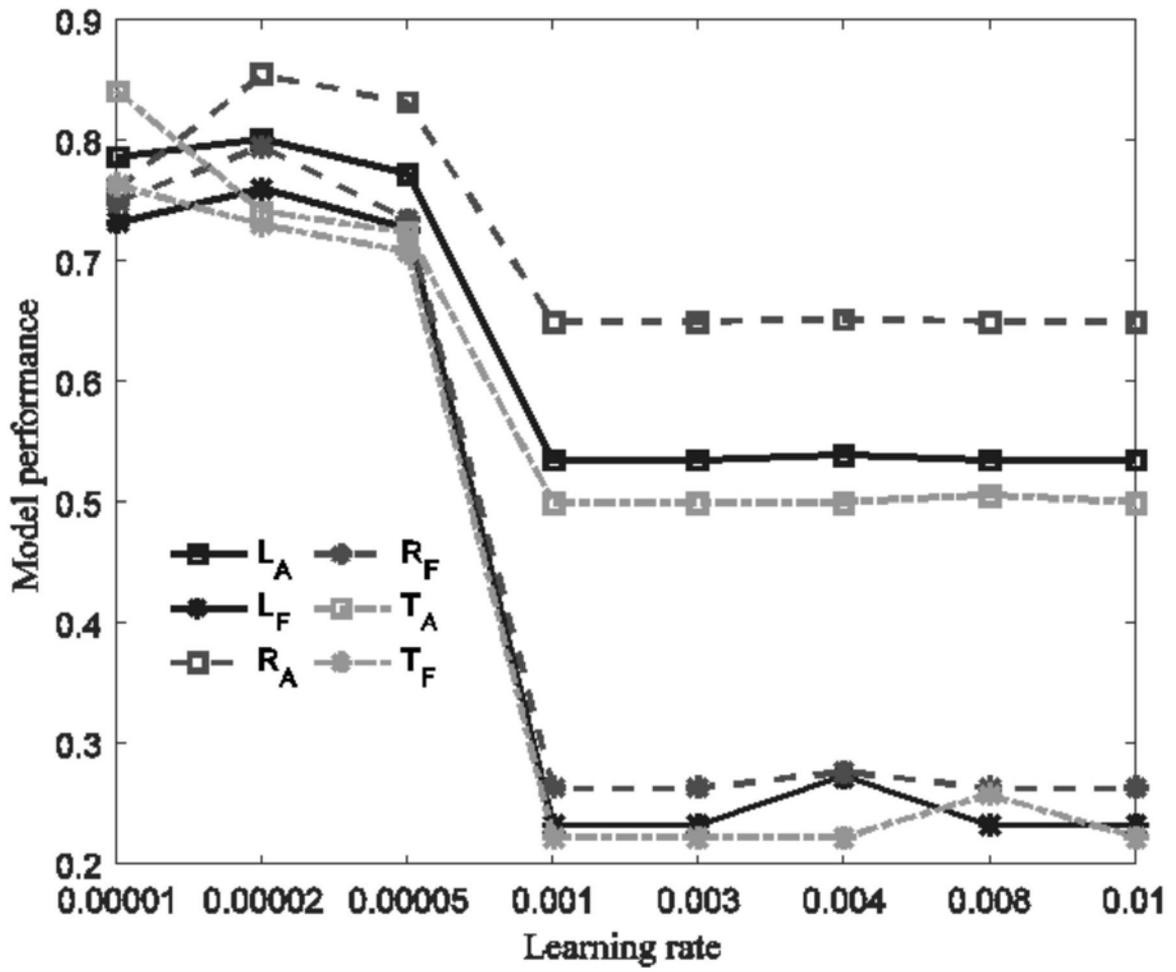


图4

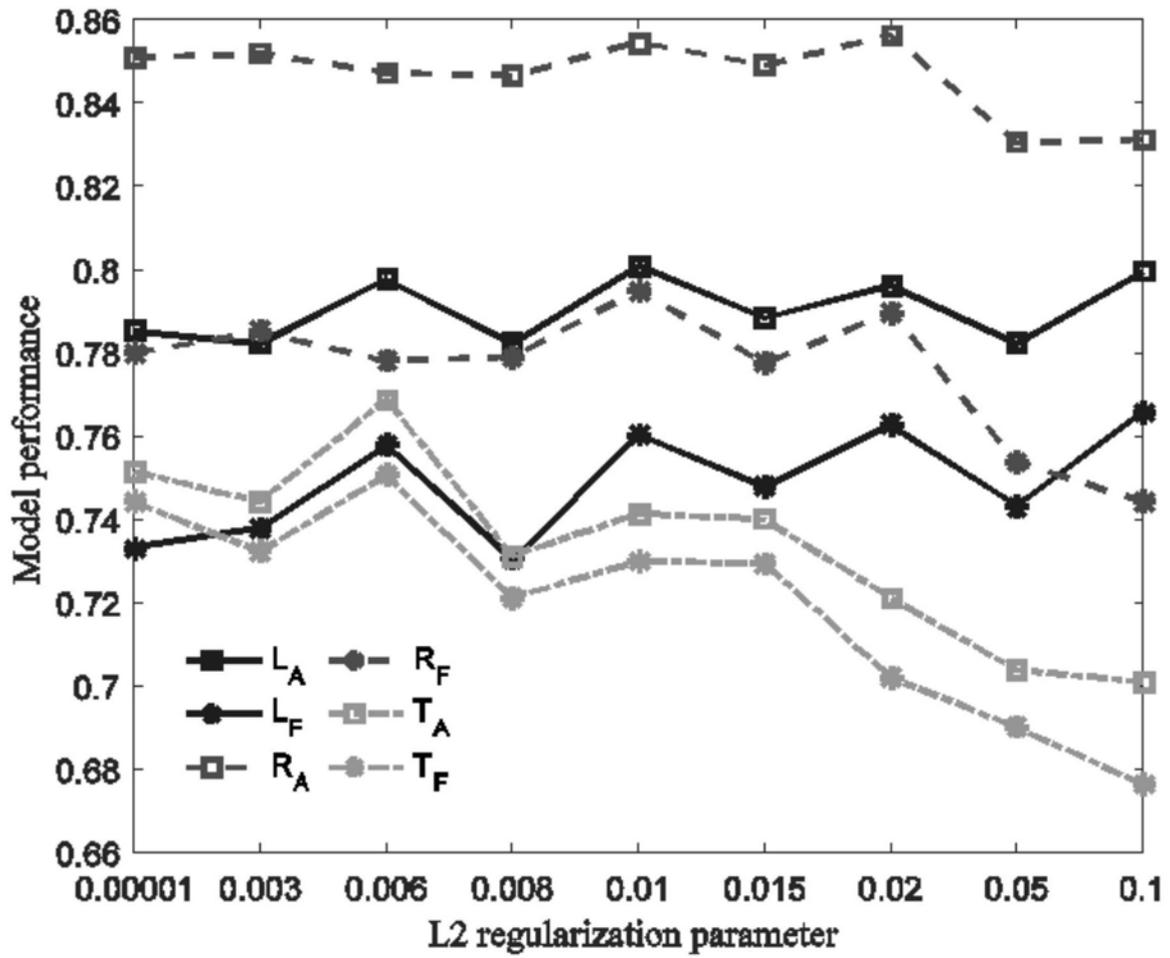


图5

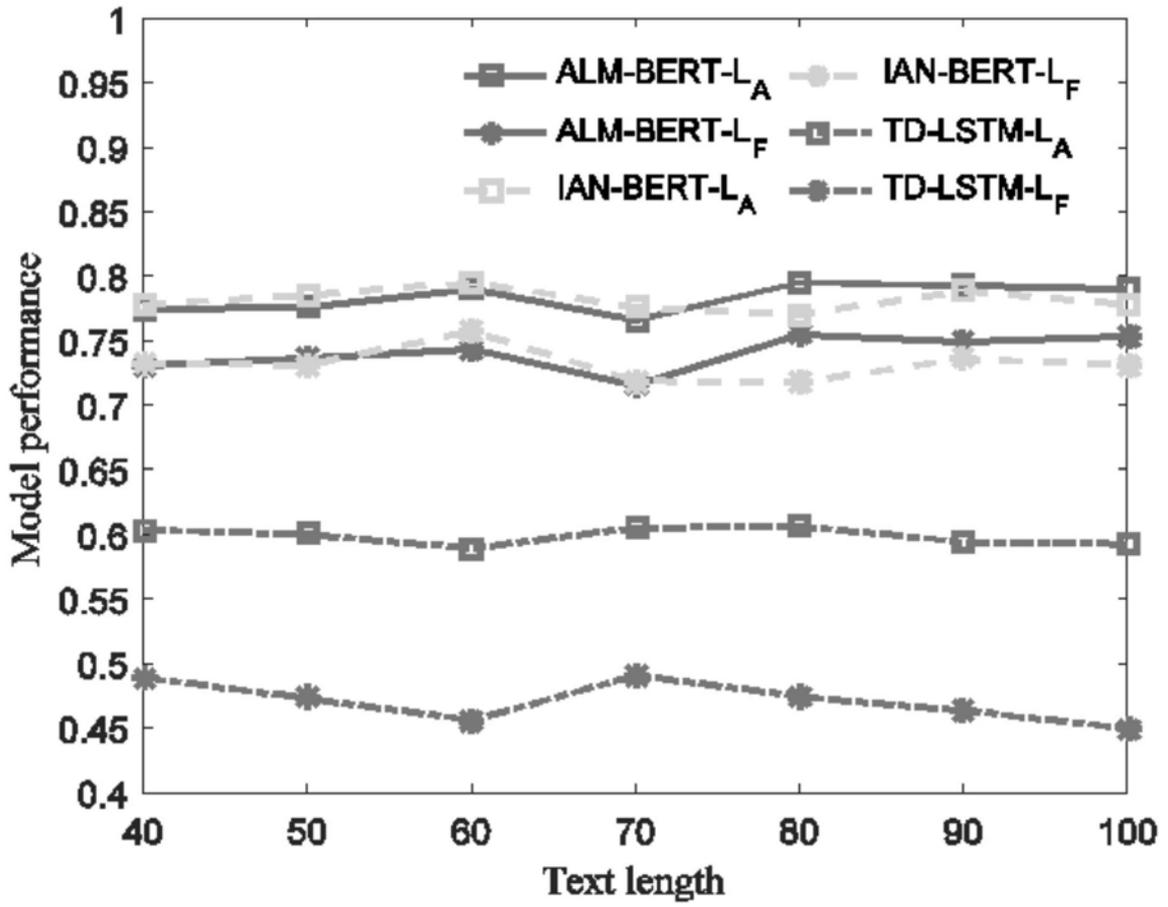


图6

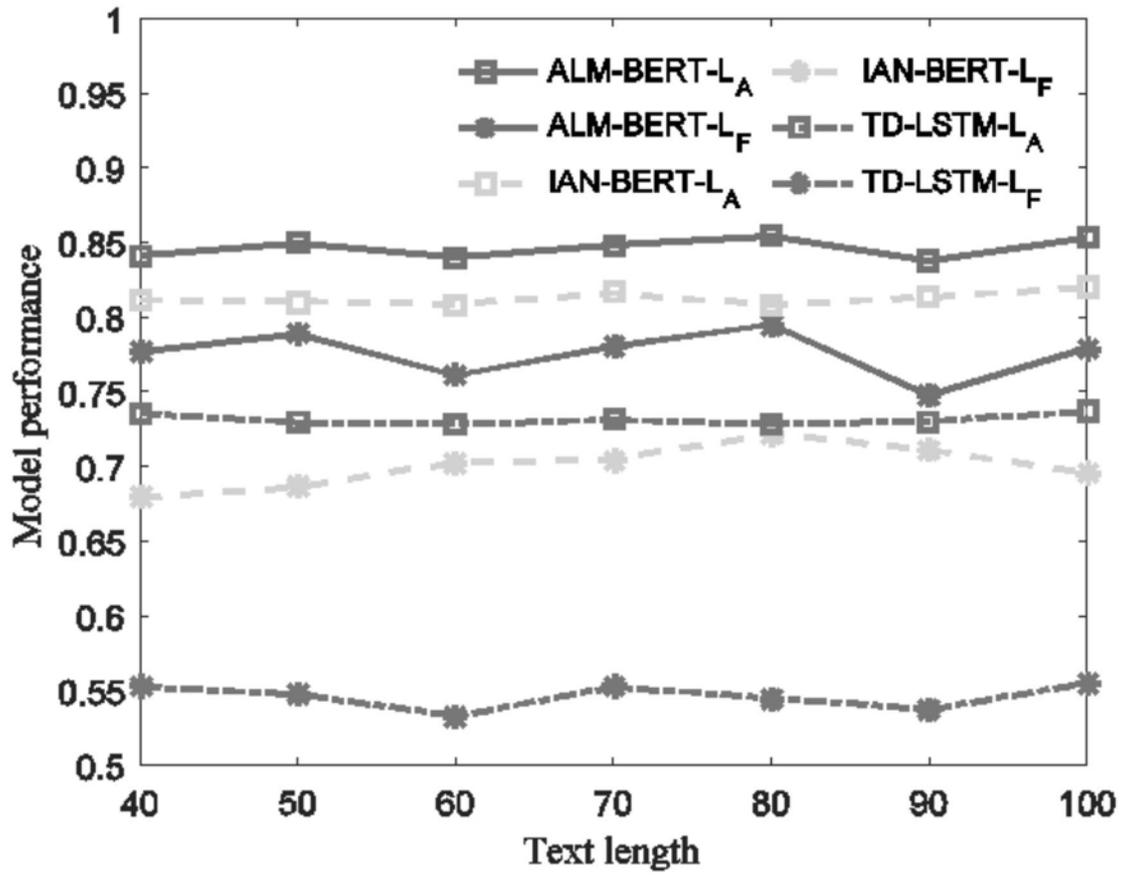


图7

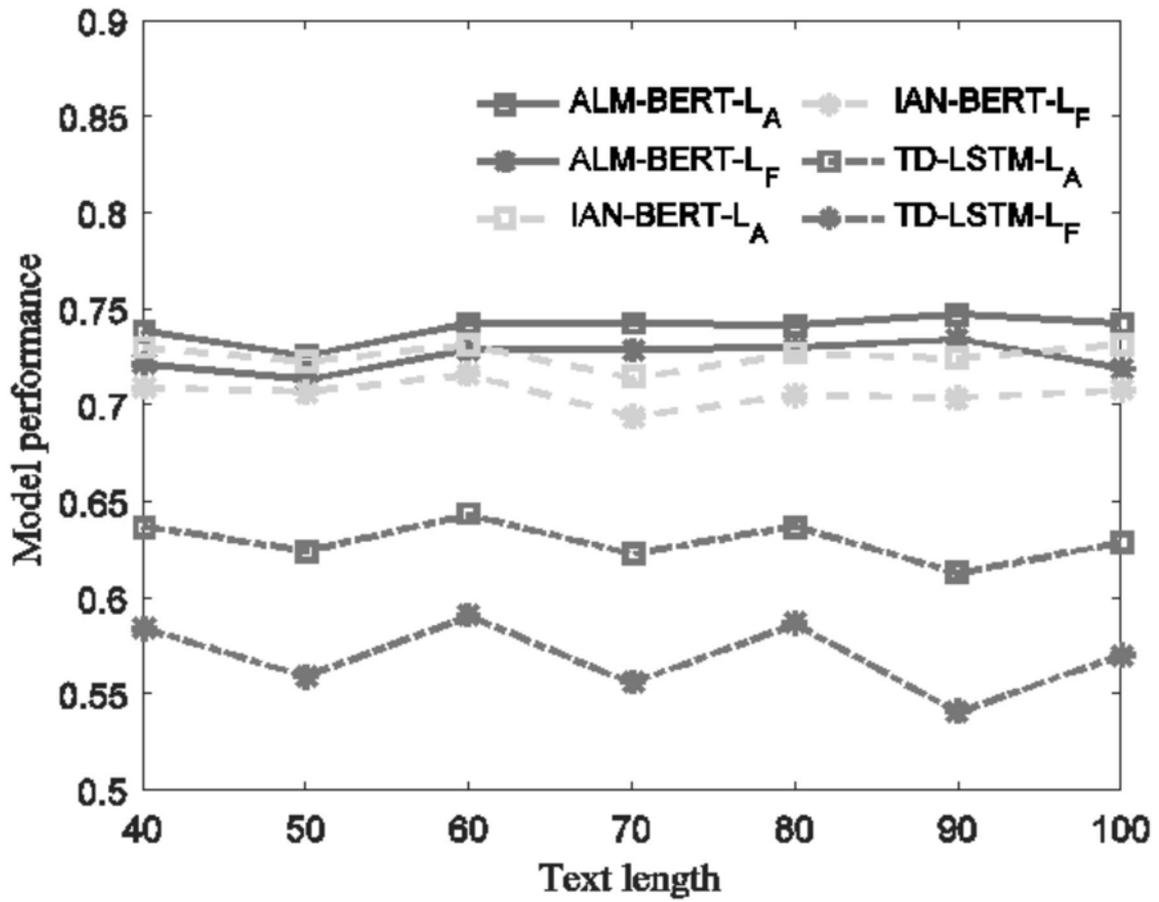


图8



图9