

文章编号: 2095-2163(2023)04-0052-08

中图分类号: TP391

文献标志码: A

# 基于 Albert 和句法树的方面级情感分析

王跃跃

(上海理工大学 管理学院, 上海 200093)

**摘要:** 方面级情感分析是情感分析的子任务,具体目标是识别不同方面词的情感极性。先前的的工作大部分采用静态词向量和循环神经网络进行这个任务的建模。然而由于自然语言表达的多样性,静态词向量不能够准确地找到修饰方面词的上下文信息,并且以往的工作在对上下文的位置信息编码时存在不足。同时发现中性标签的数据表达的不确定性,本文认为会存在一定的标签不可靠情况。所以本文提出了基于预训练模型 Albert 和引入句法树的模型 Albert-DP,并且在损失函数中加入了标签平滑。通过该设计,本模型能够很好地表示方面词对象及其上下文,有助于情感分类。本文在公开的笔记本电脑数据集、餐馆数据集以及推特数据集上的实验表明,本文的方法优于传统的模型。

**关键词:** 情感分析; 静态词向量; 句法树

## Aspect-level sentiment analysis based on Albert and syntactic tree

WANG Yueyue

(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

**【Abstract】** Aspect-level sentiment analysis is a subtask of sentiment analysis, and its goal is to identify the emotional polarity of different aspect-level words. Most of the previous work used static word vectors and recurrent neural networks to model this task. However, due to the diversity of natural language expression, static word vectors could not accurately find the context information of the modified aspect-word, and the previous work had shortcomings in encoding the location information of the context. At the same time, the paper researches the uncertainty of the neutral label data, and it is believed that there will be some label unreliability. Therefore, the paper proposes a model Albert-DP based on pre-training model Albert and syntactic tree, and adds label smoothing to the loss function. Through this design, the model can well represent the object of aspect word and its context, which is helpful for emotion classification. The experiments on publicly available laptop datasets, restaurant datasets, and Twitter datasets show that the proposed approach outperforms traditional models.

**【Key words】** sentiment analysis; static word vector; syntactic tree

## 0 引言

情感分析,也称为观点挖掘<sup>[1-2]</sup>,是自然语言处理中的重要任务。与文档级情感分类任务<sup>[3-4]</sup>不同,方面级情感分类是一种细粒度的情感分类任务。具体目的是识别上下文句子中某一特定方面的情绪极性,如积极、消极、中性。例如,给出一个句子“食物很好,但服务很慢”,“食物”和“服务”这2个方面词的情绪极性分别是积极的和消极的。方面级情感分类任务克服了一个句子中出现多个方面词时文档级情感分类的局限性。在前面的例子中有两个方面词,整个句子的情感极性并不是单一的,而是掺杂着积极和消极2种情感,如果忽略了方面词信息,就很难确定一段文本中表达的情感极性。这种错误在一般的情感分类任务中普遍存在。Jiang 等学者<sup>[5]</sup>通过实验显示 40% 的情绪分类错误是由于没有考虑

方面词。

针对方面级情感分析任务,目前已经提出了很多的方法。典型的方法是通过监督训练来构建机器学习分类器。在基于机器学习的多个方法中,主要有2种不同的类型。一种是基于人工创建的特征构建分类器<sup>[3,5]</sup>。由于神经网络能够在不进行繁杂的特征工程的情况下从数据中学习数据中的知识,现已广泛应用在诸如机器翻译、问答任务等 NLP 任务中,其应用效果明显优于传统的机器学习方法。近年来,研究学界针对方面级情感分析相继做了很多工作,如 TD-LSTM<sup>[6]</sup>、TC-LSTM<sup>[6]</sup>。但是在注意力机制<sup>[7]</sup>提出之前,对于上下文中的哪些词对于方面词的情感分类虽是重要的,却一直未能获得有效解决。

通过分析以往的方面级情感分析模型和相关的数据集,并且基于神经网络的优点,本文提出了一种

**作者简介:** 王跃跃(1994-),男,硕士研究生,主要研究方向:自然语言处理、机器学习。

**通讯作者:** 王跃跃 Email: yy\_wang1929@163.com

收稿日期: 2022-11-29

基于 Albert 和句法树的方面级情感分类模型,简称 Albert-DP。前期,基于 LSTM<sup>[8]</sup>的方法主要集中于文本建模,本文的方法使用 Bi-GRU<sup>[9]</sup>同时建模方面词和上下文信息。即如前文的例子“食物很好,但服务很慢”中,这句话中有 2 个方面词:食物、服务。根据日常的语言经验,可以知道消极的词语“慢”更可能描述的是“服务”,而不是“食物”。类似地,对于方面词“服务”,“很好”和“慢”两个修饰语与“服务”的距离都很近,如果采取 PBAN<sup>[10]</sup>的位置信息的方法,这 2 个修饰语会得到很相近的权重,这会使神经网络得到错误的分类。所以应该尽可能降低“很好”的权重,引入句法树进行上下文位置信息的编码,可使这个问题得到有效缓解。同时,本文使用 Albert<sup>[11]</sup>预训练模型,而不是 Glove<sup>[12]</sup>词向量可以解决语言表达中的一词多义问题,就能更加充分地利用句子的潜在语义信息。

## 1 相关工作

### 1.1 情感分析

情感分析的目的是检测文本的情感极性。为此已陆续提出多种方法<sup>[13]</sup>。现有的研究大多使用机器学习算法、如朴素贝叶斯和支持向量机算法等,以监督的方式对文本进行分类<sup>[14-15]</sup>。这些方法中的大多数或者依赖于 n-gram 特性,或者依赖于手工设计的特性。为此,建立了多个情感词汇<sup>[4,15-16]</sup>。近年来,神经网络在情感分类方面取得了长足的进步。基于神经网络的方法自动学习特征表示,不需要大量的特征工程,为此研究人员提出了多种神经网络结构,经典的方法包括卷积神经网络<sup>[17]</sup>、递归神经网络<sup>[18-19]</sup>。这些方法在情感分析中取得了良好的效果。

### 1.2 细粒度情感分析

方面级情感分类是情感分类的分支之一,其目标是识别句子中某一特定方面的情感极性。一些早期的研究设计了几种基于规则的方面级情感分类模型<sup>[20-21]</sup>。Nasukawa 等学者<sup>[21]</sup>首先对句子进行依赖解析,然后使用预定义的规则来确定关于方面词的情感。随后,又引入了多种基于神经网络的方法来解决这一层面的情感分类问题,典型方法是基于 LSTM 的神经网络。TD-LSTM<sup>[22]</sup>通过 2 个 LSTM 网络在建模上下文过程中考虑方面词的信息,并利用 2 个 LSTM 的最后一个隐藏状态进行情绪预测。为了更好地捕捉句子中的重要部分,Wang 等学者<sup>[23]</sup>利用方面词嵌入生成注意向量,将注意力集中在句

子的不同部分上。以此为基础,Ma 等学者<sup>[24]</sup>使用 2 个 LSTM 网络分别建模句子和方面。可进一步使用句子生成的隐藏状态,通过池化操作计算对方面目标的注意,反之亦然。因此,研究中的 IAN 模型可以兼顾句子中的重要部分和目标。Zhang 等学者<sup>[25]</sup>提出了两门控神经网络。一个门用来捕获句子的语法和语义信息,另一个门建模左右上下文的交互关系。Chen 等学者<sup>[26]</sup>采用基于 Bi-LSTM 的多层架构,每一层都包含基于注意力的词特征和句子特征,该模型可以从不同的上下文中关注到对方面词分类的不同贡献程度。Fan 等学者<sup>[27]</sup>提出多粒度的注意力网络,使用细粒度的注意力网络捕获方面词和上下文之间的词级别的交互。Gu 等学者<sup>[10]</sup>认为距离方面词较近的词语对方面词的情感极性的贡献较大,反之贡献则较小。

预训练模型基于大量的文本语料训练,在微调时可以将语训练阶段学习到的语言知识运用到下游任务中,而且也陆续展开了基于预训练模型的方面级情感分析的多项工作。Xu 等学者<sup>[28]</sup>提出了一种新颖的用于问答任务的后训练方法,同时也将其成功应用在方面级情感分析任务中。BERT-SPC 模型<sup>[29]</sup>是利用 BERT 对句子对分类的优势,将上下文和方面词一起作为模型的输入进行分类。

## 2 模型

### 2.1 任务定义

在方面级情感分类问题研究中,给定一个句子序列  $w^c = \{w_1^c, w_2^c, \dots, w_n^c\}$  和一个方面序列  $w^t = \{w_1^t, w_2^t, \dots, w_m^t\}$ ,  $w^t$  是  $w^c$  的一个子序列,方面序列可以是一个单词或一个短语。该任务的目的是对句子序列中的方面词的情感极性进行分类。本文提出模型 Albert-DP 的总体框架如图 1 所示。由图 1 可知,模型中包括词嵌入层、Bi-GRU 层、PCT 层、注意力层和输出层。

### 2.2 词嵌入层

Albert 模型的全称是 A Lite Bert,类似于一个轻量级的 Bert 模型<sup>[30]</sup>。因为预训练模型动辄有几千万或者上亿的参数,使用预训练模型在下游任务中微调时计算成本很大,并且在生产环境中部署预训练模型也是一个重要问题。为了解决上述的问题,Lan 等学者<sup>[11]</sup>提出了 Albert 模型,该模型使用矩阵分解和编码层参数共享的方法降低了模型参数。以 Albert-xlarge 为例,该模型的参数量为 6 000 万, Bert-base 模型参数量为 1.08 亿,而本文中的 Albert

模型均指 Albert-xlarge, Bert 模型都指 Bert-base。Albert 嵌入层使用预训练模型 Albert 生成序列的词向量,为了便于利用 Albert 模型的预训练和微调过程,本文中则将给定的上下文和方面词信息数据转换为“[CLS] + 上下文 + [SEP]” and “[CLS] + 方面词+[SEP]”的格式。

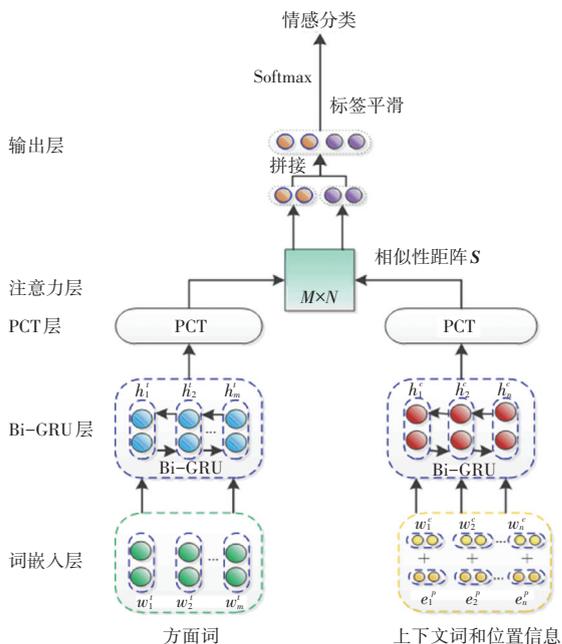


图1 Albert-DP 模型架构

Fig. 1 The architecture of the proposed Albert-DP

为了更好地与基准模型进行对比,本文使用预训练后的 Glove 词向量<sup>[11]</sup>,词嵌入矩阵  $L \in \mathbb{R}^{d_{emb} \times |V|}$ ,  $d_{emb}$  表示词向量的维度大小,  $|V|$  表示词表大小,通过从矩阵  $L$  中将每个单词  $w_i$  映射为对应的向量  $e_i$ 。

### 2.3 引入句法树的位置信息

句法树描述了句子组织的层次结构,在自然语言处理中描述句子句法结构的树状表示。通过构建句子的句法树,可以分析出单词之间的关系,更加准确地找出与方面词相关的情感词。PBAN 模型<sup>[31]</sup>出于和方面词信息更靠近的上下文更有可能在修饰方面信息的考虑,提出了上下文距离方面词位置的绝对距离,例如,在句子“The appetizers are ok, but the service is slow”中,当方面词是“appetizers”时,把方面词的权重设为 0,给定上下文的位置索引则分别表示为  $p = [1, 0, 1, 2, 3, 4, 5, 6, 7]$ ,位置矩阵  $P$  是随机初始化,参数在模型训练中不断更新。这样上下文中不同的词因为距离方面词的不同就会有不同的权重,可以更加准确地找到修饰方面词的上下

文,提高分类的准确率。由于语言表达的多样性和复杂性,并且受到 LCFs-BERT<sup>[32]</sup>的启发,引入句法知识可以更加有效判断与方面对应的情感意见表达,排除无关情感表达的干扰。语义依存树可以获得句子中不同词之间的依赖关系。通过句法树获得不同词之间的相对距离更符合语言表达的习惯,对于不同方面的情感分类会更加有效。通过句法树形成的相对语义距离可以更准确地找到修饰方面词的表达。句子“The appetizers are ok, but the service is slow”的句法解析如图 2 所示。

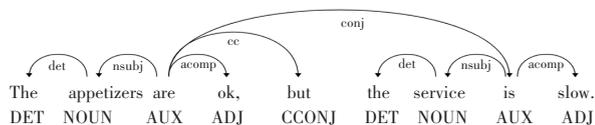


图2 句子“The appetizers are ok, but the service is slow.”句法解析图

Fig. 2 Syntactic parsing diagram of sentence “The appetizers are ok, but the service is slow.”

SpaCy 工具能够快速准确地解析句子的依存关系,SpaCy 使用 head 和 child 来描述依存关系中的连接,识别每个 token 的依存关系,通过 SpaCy 工具研究构建了每个句子的语法树,见图 2,该工具给出了句子中的每个词的词性,其中 DET、NOUN、AUX、ADJ 和 SCONJ 分别表示限定词、名词、助动词、形容词和从属连词。而且还指示出各个词语之间的语法关系。nsubj 表示名词性主语。acomp 用于动词的形容词补语,mark 表示标记语。通过构建语法树可以准确找到方面词 price 的情感表达词“reasonable”,方面词“service”的情感表达词“poor”。在建模上下文词时通过 (is, price, nsubj), (is, reasonable, acomp), (is, service, nsubj), (is, poor, acomp) 的限定语法,本次研究给予形容词词性“reasonable”、“poor”更好的权重,而如果采用计算上下文位置信息的方法<sup>[10,28]</sup>,则会赋予距离方面词“price”更近的词“The”、“is”更高的权重,而词“reasonable”将会获得更低权重,当方面词  $A = \{w'_1, w'_2, \dots, w'_m\}$  时,句子  $S = \{w_1^c, w_2^c, \dots, w_n^c\}$  中不同词与方面  $A$  之间的相对语义距离  $D = \{d_1, d_2, \dots, d_n\}$ ,所以通过构建语法树的方法可以明显提高方面级情感分析的准确率。

### 2.4 Bi-GRU 层

在得到词向量之后,将这 2 个向量集合分别送入 Bi-GRU 层,本文使用这 2 个 Bi-GRU 网络来学习句子和目标中单词的隐含语义,每个 Bi-GRU 由

2个GRU网络叠加而成。使用GRU的优点是可以避免梯度消失或爆炸的问题、能够很好地学习句子长期依赖关系,并且相对于LSTM有更少的参数和更低的计算复杂度。把方面词向量输入到左边的Bi-GRU得到前向隐层状态 $\vec{h}_i^+ \in \mathbb{R}^{d_h}$ 和后向隐层状态 $\vec{h}_i^- \in \mathbb{R}^{d_h}$ ,最终的方面词的隐层状态表示为 $h_i^t = [\vec{h}_i^+, \vec{h}_i^-] \in \mathbb{R}^{2d_h}$ ,这里 $d_h$ 表示隐层单元的数量。对于右边的Bi-GRU结构,本文把上下文词向量和位置向量的拼接,作为右边的Bi-GRU的输入。

## 2.5 PCT层

PCT层<sup>[29]</sup>是Point-Wise Convolution Transformation点式卷积转换的简称,主要的作用可以转换从Bi-GRU得到的序列信息,并且过滤到噪声信息,该卷积核的大小为1,PCT的定义如下:

$$PCT(h) = \sigma(h * W_{pc}^1 + b_{pc}^1) * W_{pc}^2 + b_{pc}^2 \quad (1)$$

其中, $\sigma$ 表示ELU激活函数,“\*”表示卷积操作; $W_{pc}^1$ 和 $W_{pc}^2$ 分别表示卷积核的可学习权重参数; $b_{pc}^1, b_{pc}^2$ 分别表示卷积核的偏置参数。

## 2.6 注意力层

传统神经网络在处理数据时往往会忽略原有数据之间的联系,特别是处理文本信息时,当输入句子较长时,中间向量难以存储足够的信息,从而限制了模型的性能。在自然语言处理过程中,注意力机制<sup>[7]</sup>允许模型根据输入文本以及其到目前为止已经生成的隐藏状态来学习要注意的内容,不同于标准神经网络将全部原文本编码成固定长度的语义向量,注意力机制将研究聚焦到特定区域,很好地解决了中间向量存储信息不足的问题。该机制灵活地为神经网络隐层单元分配不同的概率权重,使得关键节点信息得到有效关注,降低了冗余信息对情感走向的影响。利用注意力机制可以使模型更好地理解语义信息,全方位地关注更重要的内容以及整体的情感走向。在通过Bi-GRU层得到上下文和方面词隐层语义表示和PCT层转换序列信息后,由于在包含多个方面词的上下文中,不同的方面词对于相同的上下文的贡献不同,而不同的上下文的对于修饰的相同的方面词贡献大小也不一样,例如在句子“我喜欢苹果电脑的系统,但这台笔记本电脑的扬声器质量不如我那台惠普笔记本电脑”中,方面词“扬声器质量”是由多个词语组成,“质量”对于该上下文的影响应该比“扬声器”大,上下文中的词应该将更多的注意力放在“质量”这个方面词上,而对于这种注意力的量化方法,本次研究中采用注意力机

制。本文通过PCT层的上下文输出 $H$ 和方面词输出 $Q$ 构建了一个相似性矩阵 $S$ , $N$ 表示一个批次中的最大句子长度, $M$ 表示一个批次中的方面词的最大长度。 $S_{ij}$ 表示第 $i$ 个上下文的词和第 $j$ 个方面词的相关性,相似性矩阵 $S$ 的计算公式如下:

$$S_{ij} = W_s([H_i; Q_j; H_i * Q_j]) \quad (2)$$

其中, $W_s \in \mathbb{R}^{1*6d}$ 是权重矩阵;[;]表示向量的拼接;“\*”表示逐点向量元素乘积。

通过相似性矩阵 $S$ 可以计算出每一个上下文词对方面词的权重 $\alpha_i$ 以及 $m_1$ 。 $\alpha_i$ 和 $m_1$ 的计算公式具体如下:

$$\alpha_i = \text{Softmax}(\max(S_{i,:})) \quad (3)$$

$$m_1 = \sum_{i=1}^N \alpha_i H_i \quad (4)$$

同样地,通过相似性矩阵 $S$ 可以计算出方面词中的哪一个词语对于上下文的词是更加相关的,用 $m_2$ 表示,则有:

$$\beta_{ij} = \frac{\exp(S_{ij})}{\sum_{k=1}^M \exp(S_{ik})} \quad (5)$$

$$m_2 = \sum_{i=1}^M \beta_{ij} Q_j \quad (6)$$

## 2.7 输出层

在全连接层之前,如果神经元数目过大,学习能力过强,可能出现过拟合问题。本文引入dropout<sup>[33]</sup>正则化手段随机删除神经网络中的部分神经元,用来解决过拟合问题。全连接层拼接所有特征,得到一个高质量的特征图并输入到输出层。输出层通过Softmax分类器输出预测分类节点,每个节点表示一种分类,节点之和为1。接下来,又选择Adam优化器<sup>[37]</sup>优化整体网络参数,模型在使用梯度下降时,可以合理地动态选择学习速率,分配梯度下降幅度。当训练中损失函数接近最小值时减小每步的梯度下降幅度;如果损失函数曲线相对陡峭,则适当增大下降幅度。进一步地,给出数学公式见如下:

$$m = [m_1; m_2] \quad (7)$$

$$X = W_m^T m + b_m \quad (8)$$

$$y = \frac{\exp(X)}{\sum_{k=1}^c \exp(X)} \quad (9)$$

其中, $y \in R^c$ 是预测的情感极性分布, $W_m^T, b_m$ 是可学习的参数。

在标签平滑上<sup>[32]</sup>,通过观察中性标签的数据发

现中性情感是不太明确的情感表达,故本文在损失函数中加入了标签平滑,标签平滑可以通过防止神经网络在训练过程中为每个训练示例分配全概率来减少过拟合,用平滑的值(如 0.1 或 0.9)替换分类目标的 0 和 1 值。研究推得的数学公式可写为:

$$q(k|x) = (1 - \varepsilon) q(k|x) + \varepsilon u(k) \quad (10)$$

其中,  $u(k)$  表示标签的先验分布;  $\varepsilon$  是平滑参数,本文中使  $\varepsilon = 1/C$ , 这里  $C$  表示分类的个数。标签平滑等同于 KL 散度<sup>[34]</sup>, 需用到的公式如下:

$$L_{lsr} = -D_{KL}(u(k) \parallel p_{\theta}) \quad (11)$$

其中,  $u(k)$  表示标签的分布,  $p_{\theta}$  表示神经网络的预测的结果, 则本实验要优化的目标函数是:

$$L(\theta) = - \sum_{i=1}^C \hat{y}^c \log(y^c) + L_{lsr} + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (12)$$

表 1 3 个数据集的统计情况

Tab. 1 The statistics of the three data sets

数据集	积极		消极		中性	
	训练集	测试集	训练集	测试集	训练集	测试集
餐厅	2 164	728	807	196	637	196
电脑	994	341	870	128	464	169
推特	1 561	173	1 560	173	3 127	346

### 3.2 超参数设置

在实验中,首先随机选取 20% 的训练数据作为验证集来调优超参数。所有权重矩阵从均匀分布  $U(-0.1, 0.1)$  随机初始化,所有偏差项设置为零。 $L_2$  正则化系数  $\lambda = 10^{-5}$ , *Dropout* 保持率为 0.2。词嵌入使用 300 维 Glove 向量进行初始化,并在训练过程中进行固定。对于 OOV 的单词,从均匀分布  $U(-0.01, 0.01)$  中进行随机初始化。Adam 优化器<sup>[37]</sup>的初始学习率为 0.01。如果训练损失在每 5 个 *epoch* 后没有下降,将学习速率降低一半。批量大小设置为 64。Bi-GRU 隐藏状态的维数设置为 150,隐藏单元个数为 200,模型的层数为 2,采用 *ReLU* 作为激活函数。迭代次数为 10 轮,文本最大长度为一个批次中的最大句子长度。

### 3.3 模型对比

为了评估本文所提模型的有效性,将本文所提的模型与几种基线方法进行了比较,并且做了消融实验来进一步验证 Albert-DP 模型的有效性。通过在 Restaurant、Laptop 以及 Twitter 数据集上训练和评估模型,并且使用准确率和 *Macro-F<sub>1</sub>* 值来衡量模型的性能。文中用到的数学模型详述如下。

(1) TD-LSTM 模型<sup>[22]</sup>。使用 2 个 LSTM 网络对

其中,  $\hat{y}^c$  是以 one-hot 向量表示的真实标签;  $y^c$  是输出层输出的预测情感分布向量;  $\lambda$  是  $L_2$  正则化系数; “ $\odot$ ” 表示参数集合。

## 3 实验结果及分析

### 3.1 数据集

本文在 SemEval2014 任务 4 的笔记本电脑和餐厅两个领域特定的数据集<sup>[35]</sup> 以及 ACL14 推特数据集<sup>[36]</sup> 上进行了实验。这些数据集的标签有 3 个,分别是:中性、积极和消极。这 3 个数据集的训练集和测试集的一些统计情况见表 1。

围绕方面术语的前后上下文进行建模。将 2 个 LSTM 网络的最后隐藏状态连接起来,预测情绪极性。

(2) ATAE-LSTM<sup>[38]</sup>。对隐藏状态采用注意力机制,并结合注意嵌入生成最终表示。

(3) IAN<sup>[24]</sup>。使用 2 个 LSTM 网络分别对句子和方面词进行建模。利用句子中的隐藏状态生成目标的注意向量,反之亦然。基于这 2 个注意向量,则输出句子表示和分类的目标表示。

(4) RAM<sup>[26]</sup>。通过双向 LSTM 网络对上下文词学习多跳注意,并且使用 GRU 网络得到聚合向量。

(5) MemNet<sup>[39]</sup> 模型。对词嵌入进行了多重关注,所以能够从外部记忆中找出更多用于最终分类的依据,最后一层的输出被输入到 *softmax* 层进行分类。

(6) MGAN<sup>[27]</sup> 模型。在方面词是多个单词组成时或者方面词有着更长的上下文时会造成信息损失,提出多粒度的注意力机制。

(7) PBAN<sup>[28]</sup> 模型。编码方面信息时加入了上下文的位置信息,但仅仅采用距离方面信息的直线距离大小来衡量对方面信息情感极性的贡献。

(8) BERT-SPC<sup>[29]</sup>。是使用预训练模型 BERT 进行句子对分类的任务, BERT-SPC 构造的模型的

输入形式是“[CLS]”+global context + “[SEP]”+ [asp]+ “[SEP]”。

(9) LCF-BERT<sup>[40]</sup>。提出局部上下文的概念,认为方面词的情感极性应该在方面词的周围,通过人为设定一个局部上下文的阈值,不在局部上下文的上下文词会被掩码掉,不对方面词的情感极性判断做出贡献。

### 3.4 结果分析

表2给出了本模型和其他模型的对比结果。由表2分析可知,Albert-DP模型获得了最大的准确率和 $F_1$ 值的提升,这也表明本文所提模型的有效性。TD-LSTM的整体表现欠佳,因其对方面词的处理较为粗糙,在建模上下文词时没有充分利用方面词信息。ATAE-LSTM、IAN和RAM是基于注意力机制的模型,并优于TD-LSTM的表现,RAM是比其他的基于RNN的模型有更好的效果,但是在推特数据集上表现的效果却不好,可能是因为推文表达的

不规范性。IAN模型注意到方面词和其上下文交互的重要性,在建模时使用交互的注意力机制,所以IAN取得了比ATAE-LSTM更好的表现。MemNet因为从构建的外部记忆单元中获得了有用的信息,所以有关的各项衡量指标都超过了IAN模型。PBAN模型利用方面词的字符索引,并采用相对位置表示位置序列。当不同的方面词中有相同的词语时,MemNet并不能捕捉到这种差异。BERT-SPC模型和LCF-BERT模型在建模时主要利用了BERT预训练模型的强大特征表示能力,各项指标都有很大的提升。PBAN模型由于在编码上下文位置时采用的是距离方面词的绝对距离,会存在一些不足。本文所提的模型,能够有效地弥补这种不足。由于语言表达的复杂性,距离方面词很近的上下文不一定是修饰方面词,本文构建的通过句法树可以准确找到方面词的上下文词语,这更符合语言表达本身的特性。所以能够明显提高分类的准确率。

表2 3个数据集上的不同模型的表现

Tab. 2 Performance of different models on three datasets %

模型	餐厅数据集		电脑数据集		推特数据集	
	准确率	宏- $F_1$	准确率	宏- $F_1$	准确率	宏- $F_1$
TD-LSTM	75.63	-	68.13	-	70.80	69.00
ATAE_LSTM	77.20	-	68.70	-	-	-
IAN	78.60	-	72.10	-	-	-
RAM	80.23	70.80	74.49	71.35	69.36	67.30
MemNet	78.16	65.83	70.33	64.09	68.50	66.91
MGAN	81.25	71.94	75.39	72.47	72.54	70.81
PBAN	81.16	-	74.12	-	-	-
Glove-DP	81.95	72.23	75.01	73.67	73.59	71.62
BERT-SPC	85.98	78.79	80.25	77.41	75.29	73.63
LCF-BERT	86.52	80.40	82.29	79.28	76.45	75.52
ALBERT-DP	87.07	80.88	83.42	79.65	77.55	76.98

在表2中,基准模型的结果来自于已发表论文中的结果,“-”表示论文中并未使用该衡量指标。

在实验中,本文发现性能随着随机初始化的不同而有很大的波动,这是一个在训练神经网络<sup>[41]</sup>时较为常见的问题,因此,运行了本文的训练算法10次,并将得到的平均准确率和 $F_1$ 值汇总列于表2中。

### 3.5 消融实验

为了验证本模型各个模块的必要性,本文做了一系列的消融实验。借鉴AEN-BERT论文中提出

的PCT方法,在经过Bi-GRU获得的上下文和方面词语义信息后,使用PCT层过滤掉干扰分类的噪声。本文使用WP-PCT表示去掉PCT层,其他模块保持不变。为了验证引入句法树对位置信息编码的有效性,本文采用了PBAN模型中提出的位置信息方法来代替基于句法树构建的上下文位置信息的方法,简记为WO-DP。GRU相比于LSTM模型有更少的参数,训练速度更快,但这2个模型在不同的NLP任务中都有着各自的优势,本实验通过改用LSTM层来验证GRU和LSTM哪一个更适用于本文

提出的任务,简记为 W-LSTM。Albert 模型相比于 Bert 模型有很多的改进点,例如通过词嵌入层的矩阵分解、编码层的参数共享来降低参数以及将 BERT 模型中的下一个句子预测任务(NSP)改为句子顺序预测任务(SOP)等。在 Albert 论文中作者在

各大经典 NLP 任务,如阅读理解、文本分类中都进行了实验,效果高于 Bert 模型,但无法确定 Albert 模型在方面级情感分析任务中是否也能达到预期的效果,所以本文将 Albert 模型改为 Bert 模型进行验证。实验结果见表 3。

表 3 4 种消融实验结果

模型	餐厅数据集		电脑数据集		推特数据集	
	准确率	宏- $F_1$	准确率	宏- $F_1$	准确率	宏- $F_1$
WO-PCT	86.55	79.60	82.21	78.53	76.41	76.33
WO-DP	85.13	79.33	81.89	79.10	75.58	75.62
W-LSTM	86.53	80.23	82.79	79.02	77.33	74.49
W-BERT	86.62	80.59	83.06	79.31	77.12	75.58

## 4 结束语

本文提出了基于 Albert 和引入句法解析树的方面级情感分类模型 Albert-DP。以前的大部分工作都是基于静态词向量的词嵌入,不能对文本表达的多样性进行精确表示,同时对以往工作在建模上下文对方面词的位置信息存在的不足,提出了构建句子的句法树,来更准确地找到方面词的修饰语。随后,研究又发现中性标签的数据集表达的不确定性,本文认为会存在一定的标签不可靠情况,所以在损失函数中加入了标签平滑的策略。本文提出的模型在餐厅数据集(Restaurant)、电脑数据集(Laptop)和推特数据集(Twitter)上的实验表明,与那些基准方法相比,本文的模型 Albert-DP 具有更好的性能,从而进一步验证了模型的有效性。

## 参考文献

[1] PANG B, LEE L. Opinion mining and sentiment analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.

[2] LIU B. Opinion mining and sentiment analysis[M]. San Rafael, CA: Morgan & Claypool Publishers, 2012.

[3] WAGNER J, ARORA P, CORTES S, et al. DCU: Aspect-based polarity classification for SemEval task 4 [C]//SemEval@COLING. Dublin, Ireland;dblp,2014: 223-229.

[4] QIU Guang, LIU Bing, BU Jiajun, et al. Expanding domain sentiment lexicon through double propagation[C]// Proceedings of the 21<sup>st</sup> International Joint Conference on Artificial Intelligence. San Francisco, CA, USA;ACM,2009, 9:1199-1204.

[5] JIANG Long, YU Mo, ZHOU Ming, et al. Target-dependent twitter sentiment classification [C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. Portland, Oregon; ACL, 2011:

151-160.

[6] TANG Duyu, QIN Bing, FENG Xiaocheng, et al. Effective LSTMs for target-dependent sentiment classification [J]. arXiv preprint arXiv:1512.01100, 2015.

[7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems. Long Beach;NIPS Foundation, 2017, 30:5998-6008.

[8] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[9] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv preprint arXiv:1412.3555, 2014.

[10] GU Shuqin, ZHANG Lipeng, HOU Yuexian, et al. A position-aware bidirectional attention network for aspect-level sentiment analysis[C]//Proceedings of the 27<sup>th</sup> International Conference on Computational Linguistics. New Mexico, USA, : dblp, 2018: 774-784.

[11] LAN Zhenzhong, CHEN Mingda, GOODMAN S, et al. Albert: A lite bert for self-supervised learning of language representations [J]. arXiv preprint arXiv:1909.11942, 2019.

[12] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation [C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha Qatar;ACL,2014: 1532-1543.

[13] MEDHAT W, HASSAN A, KORASHY H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal,2014, 5(4): 1093-1113.

[14] WANG Sida, MANNING C D. Baselines and bigrams: Simple, good sentiment and topic classification [C]//Proceedings of the 50<sup>th</sup> Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Jeju Island; ACL, 2012: 90-94.

[15] LIU Bingwei, BLASCH E, CHEN Yu, et al. Scalable sentiment classification for big data analysis using naive bayes classifier [C]//2013 IEEE International Conference on Big Data. California, USA;IEEE, 2013:99-104.

[16] TABOADA M, BROOKE J, TOFILOSKI M, et al. Lexicon-based methods for sentiment analysis [J]. Computational Linguistics,2011, 37(2): 267-307.

- [17] KIM Y. Convolutional neural networks for sentence classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha Qatar: Association for Computational Linguistics, 2014: 1746–1751.
- [18] SOCHER R, PERELYGIN A, WU J, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C]// Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: ACL, 2013: 1631–1642.
- [19] ZHU Xiaodan, SOBIHANI P, GUO Hongyu. Long short-term memory over recursive structures [C]// International Conference on Machine Learning. Lille, France: ACM, 2015: 1604–1612.
- [20] DING Xiaowen, LIU Bing. The utility of linguistic rules in opinion mining [C]// Proceedings of the 30<sup>th</sup> annual international ACM SIGIR Conference on Research and Development in Information Retrieval. The Netherlands: ACM, 2007: 811–812.
- [21] NASUKAWA T, YI J. Sentiment analysis: Capturing favorability using natural language processing [C]// Proceedings of the 2<sup>nd</sup> International Conference on Knowledge Capture. Sanibel Island, FL, USA: ACM, 2003: 70–77.
- [22] TANG Duyu, QIN Bing, FENG Xiaocheng, et al. Effective LSTMs for target-dependent sentiment classification [J]. arXiv preprint arXiv:1512.01100, 2015.
- [23] WANG Yequan, HUANG Minlie, ZHU Xiaoyan, et al. Attention-based LSTM for aspect-level sentiment classification [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas: ACL, 2016: 606–615.
- [24] MA Dehong, LI Sujian, ZHANG Xiaodong, et al. Interactive attention networks for aspect-level sentiment classification [C]// Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17). Melbourne Australia: ACM, 2017: 4068–4074.
- [25] ZHANG Meishan, ZHANG Yue, VO D T. Gated neural networks for targeted sentiment analysis [C]// Thirtieth AAAI Conference on Artificial Intelligence. Phoenix Arizona: AAAI, 2016: 3087–3093.
- [26] CHEN Peng, SUN Zhongqian, BING Lidong, et al. Recurrent attention network on memory for aspect sentiment analysis [C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark: ACL, 2017: 452–461.
- [27] FAN Feifan, FENG Yansong, ZHAO Dongyan. Multi-grained attention network for aspect-level sentiment classification [C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Tsujii: ACL, 2018: 3433–3442.
- [28] XU Hu, LIU Bing, SHU Lei, et al. BERT post-training for review reading comprehension and aspect-based sentiment analysis [J]. arXiv preprint arXiv:1904.02232, 2019.
- [29] SONG Youwei, WANG Jiahai, JIANG Tao, et al. Attentional encoder network for targeted sentiment classification [J]. arXiv preprint arXiv:1902.09314, 2019.
- [30] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [J]. arXiv preprint arXiv:1810.04805, 2018.
- [31] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 2818–2826.
- [32] PHAN M H, OGUNBONA P O. Modelling context and syntactical features for aspect-based sentiment analysis [C]// Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics. ACL, 2020: 3211–3220.
- [33] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors [J]. arXiv preprint arXiv:1207.0580, 2012.
- [34] KULLBACK S, LEIBLER R A. On information and sufficiency [J]. The Annals of Mathematical Statistics, 1951, 22(1): 79–86.
- [35] PONTIKI M, GALANIS D, PAVLPOOULOS J, et al. Semeval-2014 task 4: Aspect based sentiment analysis [C]// Proceedings of the 8<sup>th</sup> International Workshop on Semantic Evaluation (SemEval 2014). Dublin, Ireland: dblp, 2014: 27–35.
- [36] DONG Li, WEI Furu, TAN Chuanqi, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification [C]// Proceedings of the 52<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics (volume 2: Short papers). Baltimore, Maryland: ACL, 2014: 49–54.
- [37] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. arXiv preprint arXiv:1412.6980, 2014.
- [38] WANG Yequan, HUANG Minlie, ZHU Xiaoyan, et al. Attention-based LSTM for aspect-level sentiment classification [C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, TX, USA: ACL, 2016: 606–615.
- [39] TANG Duyu, QIN Bing, LIU Ting. Aspect level sentiment classification with deep memory network [J]. arXiv preprint arXiv:1605.08900, 2016.
- [40] ZENG Biqing, YANG Heng, XU Ruyang, et al. LCF: A local context focus mechanism for aspect-based sentiment classification [J]. Applied Sciences, 2019, 9(16): 3389.
- [41] SUTSKEVER I, MARTENS J, DAHL G, et al. On the importance of initialization and momentum in deep learning [C]// International Conference on Machine Learning. Atlanta, GA, USA: PMLR, 2013: 1139–1147.

## (上接第 51 页)

- [17] HE Xiangnan, DENG Kuan, WANG Xiang, et al. LightGCN: Simplifying and powering graph convolution network for recommendation [C]// Proceedings of the 43<sup>rd</sup> International ACM SIGIR Conference on Research and Development in Information Retrieval. Xi'an, China: ACM, 2020: 639–648.
- [18] LI Changmao, FISHER E, THOMAS R, et al. Competence-level prediction and resume & job description matching using context-aware transformer models [J]. arXiv preprint arXiv:2011.02998, 2020.