

基于语义任务辅助的方面情感分析

吴肇真 新疆大学 软件学院/ 乌鲁木齐 1452078416@qq.com	赵晖 新疆大学 信息科学与工程学院/ 乌鲁木齐 zhmerry@126.com	谷体泉 新疆大学 信息科学与工程学院/ 乌鲁木齐 gtq@stu.xju.edu.cn	曹国义 新疆大学 软件学院/ 乌鲁木齐 cgy1599@163.com
--	--	--	--

摘要

方面情感分析(Asspect-Based Sentiment Analysis, ABSA)任务旨在判断一句话中不同方面的细粒度情感极性。如何有效的捕获句子的语义信息是该任务的关键。现有的大多数分类方法通过引入外部知识并设计复杂的模块来理解句子的语义信息,而忽略了外部解析器的噪音及模型的复杂化。在本文中,我们提出了一种基于语义理解的多任务学习网络,它旨在通过多任务学习从原始语料中捕获句子的语义信息。本文考虑从多任务角度出发,在具有共享参数的原始数据集中,分别提出了两个语义辅助任务:方面上下文顺序预测任务和方面上下文句法依存预测任务。然后,将辅助任务与原始的方面情感分类任务进行多任务的训练得到增强了语义理解的编码器,最后将该编码器用于方面情感分类任务。实验结果表明,模型在三个主要的公开数据集Rest14、Lap14和Twitter上的准确率和Macro-F1值都有较好的表现。

关键词: 方面情感分析; 多任务; 方面上下文顺序信息; 方面上下文句法依存信息

Semantic Task-assisted Aspect-based Sentiment Analysis

Wu Zhaozhen Xinjiang University School of Software / Urumqi 1452078416@qq.com	Zhao Hui Xinjiang University School of Information Science and Engineering/ Urumqi zhmerry@126.com	Gu Tiquan Xinjiang University School of Information Science and Engineering/ Urumqi gtq@stu.xju.edu.cn	Cao Guoyi Xinjiang University School of Software Urumqi/ cgy1599@163.com
--	--	--	---

Abstract

The Aspect-Based Sentiment Analysis (ABSA) task aims to determine the fine-grained sentiment polarity of different aspects in a sentence. How to effectively capture the semantic information of a sentence is the key to this task. Most existing classification methods understand the semantic information of sentences by introducing external knowledge and designing complex modules, ignoring the noise of external parsers and the complexity of models. In this paper, we propose a multi-task learning network based on semantic understanding, which aims to capture the semantic information of sentences from the original corpus through multi-task learning. This paper considers auxiliary semantic tasks from a multi-task perspective. Two auxiliary semantic tasks are proposed in the original dataset with shared parameters: an aspect and context order prediction task and an aspect and context syntactic dependency prediction task, respectively. Then, the auxiliary tasks are trained with the original aspect sentiment classification task in a multi-task to obtain an encoder with enhanced semantic understanding. Finally, the encoder is used for the aspect sentiment classification task.

The experimental results show that the model performs well in terms of accuracy and Macro-F1 values on three major publicly available datasets Rest14, Lap14, and Twitter.

Keywords: Aspect-based sentiment analysis , Multi-task , Order information of Aspect-Context , Syntactic dependency information of Aspect-Context

1 引言

情感分析又被称为意见挖掘(王婷, 杨文忠, 2021)是自然语言处理(Natural Language Processing, NLP)领域的一项基础任务。方面情感分析是对实体的特定方面(属性)的情感倾向判断, 本文将这样的实体的特定方面(属性)定义为方面项, 它是句子中连续的方面词序列, 用于情感倾向判断的词统称为情感词。一般说来, 将句子中不同方面的情感类别分为积极(Positive)、中性(Neutral)和消极(Negative)三类。

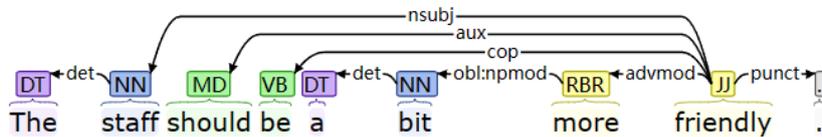


Figure 1: 依存解析样例

ABSA任务上的最近进展都与深度学习的发展息息相关。然而深度学习的基础来自于神经网络, 使用神经网络来处理ABSA任务会产生这样的倾向: 模型在遇到情感词时, 会直接根据情感词输出情感极性, 而不再更深层地理解句子的语义。例如, 在句子“The staff should be a bit more friendly .”中, 传统的神经网络会直接根据情感词“friendly”从而将方面项“staff”判定为积极的情感极性, 事实上这里的虚拟语气词“should”表示否定的含义, 因此方面项“staff”的真实情感极性应该是消极的。另一方面, 在最近的新方法中(Zhang et al., 2019; Wang et al., 2020a), 为了能够更好地提取语义特征, 不得不需要一些外部的知识, 例如句法依存信息, 然而为了使用这种外部知识, 不得不在网络中设计相应的模块, 同时这些外部知识中常常包含着对于目标任务ABSA无关的信息, 因而对目标任务来说是一种噪音。

考虑到上述问题, 本文从多任务的角度分别设计了两个辅助任务: 方面上下文顺序预测任务(Aspect-Context Order Prediction, ACOP)和方面上下文句法依存预测任务(Syntactic Dependency of Aspect-Context Prediction, SDACP)。然后在预训练模型RoBERTa(Liu et al., 2021)的基础上将辅助任务与目标任务ABSA进行多任务训练得到增强了语义理解和提取了句法结构知识的编码器, 最后将该编码器用于目标任务ABSA。总的来说, 本文的主要贡献如下:

- (1) 本文探索了多任务框架下的方面情感分析, 从而定义了两种语义增强的辅助任务: ACOP和SDACP。
- (2) 基于RoBERTa模型, 融合ACOP任务提出了RoBERTa-ACOP模型, 融合SDACP任务提出了RoBERTa-SDACP模型。
- (3) 基于辅助任务提出的模型在公开数据集上实验, 实验结果证明了本文模型的有效性。

2 相关工作

现有的方面情感分析方法分为两种: 基于机器学习的方法和基于深度学习的方法。基于机器学习的方法通过考虑与任务相关的特征来完成任务, 需要人工参与许多重要的特征设计。Wagner(Wagner et al., 2014; Kiritchenko et al.,)使用基于支持向量机的模型来完成方面情感分类任务。Gupta等人(Kumar Gupta et al., 2015)使用了基于粒子群优化算法来更好的提取与ABSA任务相关的特征, 通过对粒子的不断迭代演化去除与任务特征无关的噪声, 从而得到更高质量的特征表示, 然后将特征输入条件随机场训练, 得到更好表现的模型。Akhtar等

人(Akhtar et al., 2017)使用最大熵模型、条件随机场和支持向量机构建分类器,然后整合特征。Jo等人(Jo and Oh, 2011)引入无监督训练的方法提出了一种基于句子潜在Dirichlet分布的概率生成模型。

基于深度学习的方法大致分为三个阶段。第一阶段是使用结合注意力机制的循环神经网络(RNN)来生成方面项特定的句子表示(Wang et al., 2016; Tang et al., 2016; Ma et al., 2017; Chen et al., 2017; Fan et al., 2018; Huang et al., 2018; Gu et al., 2018),并取得了令人满意的结果。例如,Wang等人(Wang et al., 2016)提出了基于注意机制的长短期记忆网络(Long Short-term Memory, LSTM)用于方面级情感分类。Tang等人(Tang et al., 2016)和Chen等人(Chen et al., 2017)都引入层级注意力网络来识别与给定方面相关的重要情感信息。Fan等人(Fan et al., 2018)使用多粒度注意机制来捕获方面项及其上下文之间的词级交互信息。

第二阶段是是基于图神经网络(Graph Neural Network,GNN)特别是图卷积网络(Graph Convolution Network,GCN)和图注意力网络(Graph Attention Network,GAT)的方法(Zhang et al., 2019; Wang et al., 2020a; Sun et al., 2019; Huang and Carley, 2019; Zhang and Qian, 2020; Chen et al., 2020; Liang et al., 2020; Wang et al., 2020b; Tang et al., 2020; Li et al., 2021)。Zhang等人(Zhang et al., 2019)和Sun等人(Sun et al., 2019)认为语法结构信息对于模型理解文本非常重要,都各自提出了图卷积网络。Huang等人(Huang and Carley, 2019)则在句子语法结构的基础上进一步引入注意力机制。Zhang等人(Zhang and Qian, 2020)利用全局词共现频率信息和词性信息来增强模型使得模型更好地理解文本。Chen等人(Chen et al., 2020)利用潜在推导图,在解析出的句法结构基础上获取更多的潜在关系,从而获取更多节点间的关系信息。Liang等人(Liang et al., 2020)通过注意力机制和图卷积网络在改进的依存树解析结果上取得了不错的效果。Wang等人(Wang et al., 2020a)对GAT方法进行了扩展,使用方面项和上下文之间的句法依存关系来计算注意力,由此提出了关系图注意力网络(Relational Graph Attention Network,R-GAT)。Li等人(Li et al., 2021)则同时考虑了句子的语义信息和句法信息,并融合了两种信息,提取到更好的特征。

第三阶段是大规模预训练模型的出现,在ABSA任务上带来突破性的成果(Devlin et al., 2018; Xu et al., 2019; Song et al., 2019),例如Google的Devlin等人(Devlin et al., 2018)提出了BERT(Bidirectional Encoder Representations from Transformers)模型,该模型包括MLM(Masked Language Model)任务和NSP(Next Sentence Prediction)任务,实现了阅读理解多项任务上的突破。Xu等人(Xu et al., 2019)设计了BERT-PT,它探索了一种基于BERT模型的新型的后训练方法。Song(Song et al., 2019)进一步提出将句子的方面项附加到上下文句子中,以形成对应于文本对分类模型BERT-SPC的输入。另一方面,大规模预训练模型的出现给我们的启发不仅是多项自然语言处理任务上的突出表现,同时我们看到多任务的训练为模型的优异表现提供了坚实的基础,这为本文使用多任务方法提供了思路 and 方向。

3 多任务情感分析

3.1 任务定义

3.1.1 ACOP辅助任务

ACOP任务是指给出一个句子,判断该句子是否是正确的顺序,如果是乱序并识别出乱序的模式。本文提出ACOP任务的主要目的是使得模型能够理解句子的语义及方面项与上下文之间的结构信息。本文考虑了两种方式来实现ACOP任务:全局实现和局部实现。在全局实现中,本文将方面项作为边界将整个句子划分为左上上下文,方面项和右上上下文,然后我们基于以上三部分的全排列得到6种类型的顺序,如表1所示。

另外一种局部实现中,将方面项随机地插入到上下文中生成负样本,将句子的原始顺序作为正样本,如表2所示。由于当上下文足够长时,使用这种方式会产生非常多的负样本,给实验带来极大的计算代价,因此本文设置窗口大小(window_size)限制方面项随机插入的范围,和负样本数(neg_num)限制生成的乱序句子数量来解决这个问题。

3.1.2 SDACP 辅助任务

SDACP任务是指判断句子中每一个单词节点的关系类型,该任务是一个序列分类任务。本文提出SDACP任务的目的是希望模型能够以多任务方式去利用句法依存信息从而更好地学习句子的语义信息。本文的具体实现是在图2这样的句法依存树分析的基础上,由于生成的句

Aspect	Sentence	Output class
the food	Nevertheless the food itself is pretty good.	0
	Nevertheless itself is pretty good the food.	1
	The food nevertheless itself is pretty good.	2
	The food itself is pretty good nevertheless.	3
	Itself is pretty good nevertheless the food.	4
	Itself is pretty good the food nevertheless.	5

Table 1: ACOP任务的全局实现

Aspect	Sentence	Output class	Window size
the food	Nevertheless the food itself is pretty good .	0	0
	The food nevertheless itself is pretty good .	1	1
	Nevertheless itself the food is pretty good .	2	
	The food nevertheless itself is pretty good .	3	2
	Nevertheless itself the food is pretty good .	4	
	Nevertheless is itself the food pretty good .	5	

Table 2: ACOP任务的局部实现

法依存树不是以方面项为根(root)的, 在该例中根是“was”, 因此需要将依存树重新塑造成以方面项“staff”为根。然后在重塑的依存树的基础上为每一个上下文词生成句法依存标签, 如表3所示, 其中con_n表示上下文词与方面项不具有直接的句法依存关系, 而是通过上下文词与方面项的句法依存距离为n这一关系相联系。

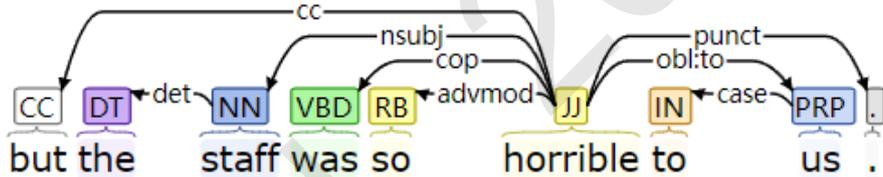


Figure 2: 句法依存树样例

在句法依存标签的生成过程中, 本文使用变量cap_sd(captured syntactic distance)来控制需要捕获的句法依存距离信息范围, 在表3的示例中, 将该值设置为2, 这意味着只捕获句法依存距离为2以内的信息(保留原来的句法依存信息), 而将距离为2以上的句法依存距离信息的上下文词统一分配给一个新的类别ncon, 标点符号分配给另一个新的类别pun。

3.2 模型结构

使用辅助任务来帮助完成原ABSA任务的具体模型结构包含输入层、编码层、多任务层。模型整体架构如图3所示。

模型的具体作用过程如下: 首先通过RoBERTa模型输出预训练的方面项词向量和整个句子词向量拼接, 并将拼接后的向量输入多任务层训练, 最后再将多任务层训练后的向量输入到分类层, 输出情感极性。

3.2.1 输入层

该模型的输入是句子 $S = \{w_{s1}, w_{s2}, \dots, w_{sls}\}$ 和相应的方面项 $A = \{w_{a1}, w_{a2}, \dots, w_{ala}\}$, 其中 ls 和 la 分别表示句子长度和方面项的长度。然后我们使用预训练模型RoBERTa将每个词转为一个实值向量。具体来说, 对于句子 S , 我们构造RoBERTa模型的输入 “<s>” + S + “</s>”, 然后将句子 S 表示为RoBERTa隐藏状态 $s = \{s_i \mid i = 1, 2, \dots, ls + 2\}$; 对于方面项, 我们构造相应的输入 “<s>” + A + “</s>”, 并将其转为对应的隐藏状态表示 $a = \{a_j \mid j$

Sample	but	the	staff	was	so	horrible	to	us	.
Syntactic dependency	con_2	det	root	nsubj	con_3	con_2	con_3	con_4	-
Tag	con_2	det	root	nsubj	ncon	con_2	ncon	ncon	pun

Table 3: 句法依存标签生成

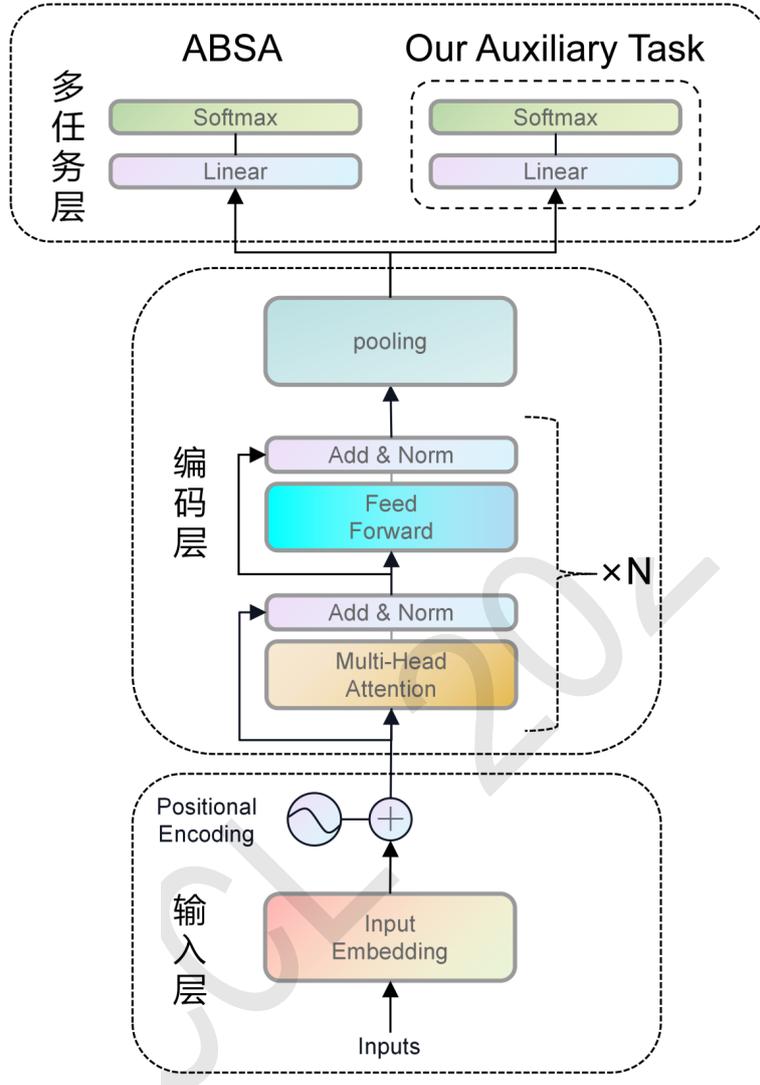


Figure 3: 模型整体架构

$= 1, 2, \dots, la + 2\}$ 。为了使模型能够感知到句子的方面项，我们将上述的句子向量和方面项向量进行拼接得到感知方面项的句子表示 S^A ，计算过程如下所示。

$$S^A = \{s_i^a | i = 1, 2, \dots, l_s, \dots, l\} = s \oplus a \tag{1}$$

其中 $l = l_s + l_a + 4$ 。

3.2.2 编码层

在编码层部分需要同时对目标任务ABSA和辅助任务完成编码和池化工作。我们使用RoBERTa编码器编码,RoBERTa编码器部分像BERT模型一样仍然是基于Transformer的编

数据集	分类	积极	中性	消极	总和
Twitter	Train	1561	3127	1560	6248
	Test	173	346	173	692
Laptop	Train	994	464	870	2328
	Test	341	169	128	638
Restaurant	Train	2164	637	807	3608
	Test	728	196	196	1120

Table 4: 数据集统计

码器。RoBERTa的编码器部分，用下列公式进行概括。

$$\mathbf{m} = \text{RoBERTa_Encode}(\mathbf{S}^A) \quad (2)$$

池化操作就是从提取到的特征中进行选择从而更好地完成后面的任务，我们直接截取第一个词的特征进行池化。在经过编码器部分后，得到特征表示 $\mathbf{m} = \{m_i | i = 1, 2, \dots, l\}$ ，然后进行池化得到更加重要的特征，如下公式所示。

$$\mathbf{r} = \text{Pooling}(m_i | i = 1, 2, \dots, l) \quad (3)$$

3.2.3 多任务层

在原ABSA任务中，给出集成了方面项的句子信息 \mathbf{S}^A 输入到多任务层，我们得到特征表示 \mathbf{r}_{absa} ，对于我们的辅助任务，给出句子信息 \mathbf{s} 输入到多任务层，得到特征表示 \mathbf{r}_{aux} ，然后分别对两个任务进行分类得到相应的情感倾向概率分布，计算过程由下列公式给出。

$$\mathbf{p}_{absa} = \text{softmax}(\mathbf{r}_{absa} \mathbf{W}^{absa} + \mathbf{b}^{absa}) \quad (4)$$

$$\mathbf{p}_{aux} = \text{softmax}(\mathbf{r}_{aux} \mathbf{W}^{aux} + \mathbf{b}^{aux}) \quad (5)$$

其中 \mathbf{W}^{absa} ， \mathbf{W}^{aux} 是可训练权重矩阵， \mathbf{b}^{absa} ， \mathbf{b}^{aux} 是偏置项。我们采用标准的交叉熵函数和 L_2 正则优化训练，计算过程如下所示。

$$L(\mathbf{p}) = -\sum_{s \in \mathcal{S}} \sum_{c \in \mathcal{C}} \hat{p} \log p \quad (6)$$

$$\text{Loss} = \lambda_1 L(\mathbf{p}_{absa}) + \lambda_2 L(\mathbf{p}_{aux}) + \lambda_3 \|\Theta\|_2 \quad (7)$$

其中 \mathcal{S} 包含所有的训练句子， \mathcal{C} 表示情感类别的集合， \mathbf{p} 表示预测结果， $\hat{\mathbf{p}}$ 表示标签， λ_1 ， λ_2 ， λ_3 表示正则系数， Θ 表示所有的可训练参数。

4 实验与分析

4.1 数据集与参数设置

数据集：本文采用国际语义评估竞赛SemEval于2014年发布的任务4中的Restaurant、Laptop数据集(Maria Pontiki et al., 2014)以及Twitter数据集(Dong et al., 2014)。Restaurant、Laptop数据集分别是关于餐厅和笔记本电脑相关的评论，以下将分别简称为Rest14和Lap14数据集，Twitter数据集则是用户在社交平台推特上的评论数据，每一个样本数据都标注好了方面项和情感类别，情感类别包括积极(Positive)、消极(Negative)、中性(Neutral)三种类别，本文在实验中采用的评价指标是准确率Acc和Macro-F1值。数据集相关统计情况如表4所示。

参数设置：本文使用的是RoBERTa-base英文版本，该模型包含12个隐藏层、12个注意力头，每层的隐藏状态维度是768维。使用Adam(Kingma and Ba, 2014)作为模型的优化器，学习率初始化为0.000002，正则化系数 λ_1 设置为1， λ_2 根据数据集有所不同，在主体实验中，

本文在Rest14数据集上设置为0.1, 在Lap14数据集上设置为0.001, 在Twitter数据集上设置为1, 衰减率 λ_3 设置为0.01。除了衰减率的影响外, 学习率在预热阶段(warm up)首先增加, 然后保持在设置的学习率, 最后随着当前步数的增加而降低。批量改组(Batch Shuffling)应用于训练集。RoBERTa的词汇量为50,265。所有模型的批量大小(batch size)都设置为32。针对ACOP任务局部实现中, window_size的范围是1到10, 当window_size足够大时neg_num设置为5, 否则我们将所有的句子作为负样本。针对SDACP任务的超参数cap_sd在主要实验中设置为5, 在研究该参数影响的扩展实验中, 将其设置为1到8, 该任务中使用的句法依存分析的工具是spacy(Srinivasa-Desikan, 2018)。本文提出的模型通常训练不到30个epoch, 就提前停止, 尽管设置了100个epoch。

4.2 对比实验与分析

该小节将本文模型的表现与一些基线模型进行对比来证明本文模型的有效性。本文将对比的基线方法分为基于机器学习的方法、基于注意力的方法、图神经网络方法及预训练方法, 在下面具体列出。

基于注意力模型:

ATAE-LSTM(Wang et al., 2016): 使用LSTM并结合注意力机制的情感分类模型。

IAN(Ma et al., 2017): 使用互动注意力模型识别方面项和对应上下文的联系。

MemNet(Tang et al., 2016): 使用多层记忆网络计算上下文中每个单词的情感贡献程度。

AOA(Huang et al., 2018): 使用多层注意力机制识别方面项和对应上下文的联系。

MGNet(Fan et al., 2018): 从词级识别上下文中每个单词与方面项之间的联系, 是一个细粒度的情感分类模型。

TNet(Li et al., 2018): 使用CNN模型提取来自于RNN中的突出特征完成情感分类任务。

基于图神经网络模型:

ASGCN-DG(Zhang et al., 2019): 基于图卷积神经网络利用句法依存信息提升情感分类任务的效果。

BiGCN(Zhang and Qian, 2020): 利用全局词共现频率信息和词性信息来增强模型。

TD-GAT(Huang and Carley, 2019): 基于图注意力模型聚合句子中的情感信息。

R-GAT(Wang et al., 2020a): 基于图注意力模型利用句法依存信息提升情感分类任务的效果。

Dual-GCN(Li et al., 2021): 基于图卷积神经网络融合句子的语义信息和句法依存信息。

基于预训练模型:

Fine-tune BERT(Devlin et al., 2018): 在情感分类任务上微调BERT模型。

BERT-SPC(Xu et al., 2019): 方面项和上下文对作为BERT模型的句子对任务的微调BERT模型。

Fine-tune RoBERTa(Liu et al., 2021): 在情感分类任务上微调RoBERTa模型。

RGAT-BERT(Wang et al., 2020a): 基于图注意力模型利用句法依存信息并使用BERT模型向量化句子表示提升情感分类任务的效果。

实验结果如表5所示, 从表中有以下观察结果。(1)基于预训练的方法在两个指标上的表现都优于大多数基于注意力的方法(例如AOA和TNet)和基于GNN的方法(例如ASGCN-DG和TD-GAT), 这表明预训练语言模型强大的语义表示能力。这一现象也表明, 预训练模型已经成为NLP的趋势, 其中也包括在ABSA任务上的应用。(2)具有特定任务的基于预训练的方法(例如RGAT-BERT)优于纯预训练(BERT、RoBERTa)的方法, 说明将特定任务集成到预训练模型可以改善模型表现, 我们对此的解释是纯预训练模型是基于整体语义从而判断出情感, 而基于任务的预训练模型通常可以捕捉到句子的方面项和对应上下文的联系。这也表明细粒度的属性级ABSA相对于句子级和文档级情感分析中是更加有益的, 因为方面项在情感分析中是一个关键信息。基于以上分析, 本文提出两个新的辅助任务(ACOP和SDACP)并集成到预训练模型从而改善在方面情感分析上的表现。(3)尽管之前的模型表现出色, 但本文的模型在准确性和F1分数两个指标上相较于最先进的基线仍有不同程度地改善, 例如, 我们的模型相较于最好的基线模型(BERT-SPC、RGAT-BERT)在Rest14和Lap14数据集上的准确度和F1值分别提升了1.97%、1.76%和4.58%、5.16%, 在Twitter数据集上提升了1.74%、2.38%, 因此证明了ACOP任务和SDACP任务在提取方面项与上下文的结构信息和句法结构信息方面的有效性, 从而增强模型的语义理解, 得到更好的性能。

分类	模型	Restaurant		Laptop		Twitter	
		Acc	F1	Acc	F1	Acc	F1
Attention.	ATAE-LSTM	76.58	67.39	68.57	64.52	67.27	66.43
	IAN	76.88	68.36	70.84	65.73	68.74	67.61
	MemNet	78.12	68.99	72.32	67.03	70.19	68.22
	AOA	79.42	70.43	74.56	68.77	71.68	69.25
	MGNet	81.28	72.07	75.37	71.26	72.54	70.78
	TNet	80.69	71.27	76.54	71.75	74.93	73.60
GNN.	ASGCN-DG	80.77	72.02	75.55	71.05	72.15	70.40
	BiGCN	81.97	73.48	74.59	71.84	74.16	73.35
	TD-GAT	81.20	-	73.40	-	-	-
	R-GAT	83.30	76.08	77.42	73.76	75.57	73.82
	Dual-GCN	84.27	78.08	78.48	74.74	75.92	74.29
Pre-trained.	BERT	82.40	73.17	77.29	73.36	73.42	72.17
	BERT-SPC	84.46	76.98	78.99	75.03	74.13	72.73
	RoBERTa	84.38	77.26	78.83	74.53	74.57	74.02
	RGAT-BERT	<u>86.60</u>	<u>81.35</u>	78.21	74.07	<u>76.15</u>	<u>74.88</u>
Ours.	RoBERTa-SDACP	85.18	77.66	79.78	75.87	77.89	77.26
	RoBERTa-ACOP	88.57	83.11	83.57	80.19	77.60	76.94

Table 5: 多任务语义模型RoBERTa-ACOP和RoBERTa-SDACP在三个公认数据集上性能比较。表中的“Acc”表示准确率，“F1”表示Marco-F1值，表中各项数值的单位均为%，以前方法中的最优结果使用下划线标记。

模型	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
RoBERTa	84.38	77.26	78.83	74.53	74.57	74.02
RoBERTa-SDACP	85.18	77.66	79.78	75.87	77.89	77.26
RoBERTa-LACOP	86.61	80.22	82.62	78.99	77.60	76.94
RoBERTa-GACOP	88.57	83.11	83.57	80.19	77.02	75.55

Table 6: 消融结果。表中数据的单位均为%，模型中的“G”和“L”分别表示实现ACOP任务的两种方式全局实现(Globally)和局部实现(Locally)。

4.3 任务级消融实验

为了验证本文提出的辅助任务ACOP和SDACP的有效性，本文在任务级别上进行了消融实验进行对比，实验结果如表6所示。从表中可以得到，将ACOP任务和SDACP任务集成到RoBERTa模型的RoBERTa-GACOP、RoBERTa-LACOP和RoBERTa-SDACP都优于纯RoBERTa模型。例如，在Lap14数据集上，相比于纯RoBERTa模型，RoBERTa-SDACP模型在准确度和F1上分别提升了0.95%和1.34%，RoBERTa-LACOP模型在准确度和F1上分别提升了3.79%和4.46%，RoBERTa-GACOP模型在准确度和F1上分别提升了4.74%和5.66%，在另外两个数据集上也有不同程度的提升。另外，全局实现ACOP任务的RoBERTa-GACOP总体优于局部实现的RoBERTa-LACOP模型。以上分析证明了本章提出的ACOP任务和SDACP任务在提取语义方面的有效性，同时说明使用多任务方法相对于设计相应的模块引入外部知识处理目标任务ABSA确实是一个不错的选择。

4.4 参数研究

4.4.1 局部实现ACOP窗口大小的影响

在RoBERTa-LACOP模型的局部实现中，窗口大小(window_size)是一个重要的超参数，由于该参数会影响局部替换方面项和上下文片段时的范围，因此本文对窗口大小设置为1到10时

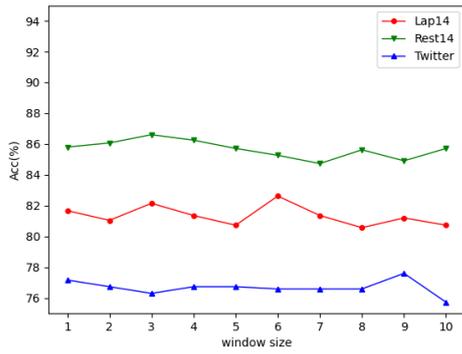


Figure 4: 窗口大小对准确率的影响

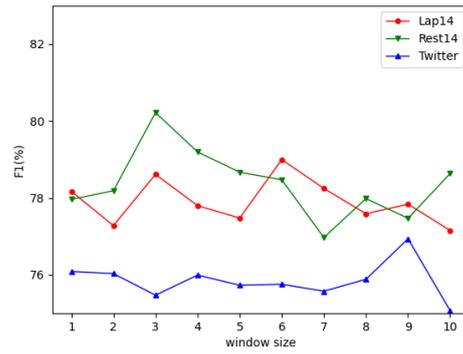


Figure 5: 窗口大小对F1值的影响

在Rest14、Lap14和Twitter数据集上进行了实验探索窗口大小参数的影响，结果如图4和图5所示。从图中，可以发现模型的性能不会随窗口大小单调变化。然而，我们观察到在Rest14数据集上窗口大小为3时达到相对最优的性能，在Lap14数据集上窗口大小为6时达到相对最优的性能，在Twitter数据集上窗口大小为9时达到相对最优的性能，这说明参数窗口大小确实会影响我们模型的性能。

4.4.2 捕捉句法依存距离的影响

在RoBERTa-SDACP模型的实现中，捕捉句法依存距离(*cap_sd*)是一个重要的超参数，由于该参数会影响模型捕捉句法信息的范围，因此本文做了扩展实验观察该参数设置为1到8时整个模型的表现，从而研究该参数对RoBERTa-SDACP模型的影响，结果如图6、图7和图8所示。

从图中，可以发现捕捉句法依存距离的变化确实会影响模型的性能。在三个数据集上都发现捕捉句法依存距离的参数设置为5时，模型性能达到最好的效果。本文对该实验结果可能的解释是：当该值设置得过小时，由于模型捕捉到的句法结构距离较短，不能充分地理解到观点词所表现的情感，当该值设置过大时，由于捕捉到的句法结构较大，虽然能够捕捉到观点词的情感信息，但是引入的过多的噪音会削弱模型对方面项对应的观点词的正确理解。

4.5 案例分析

为了说明本章提出的方法的有效性，本节对句子评论样本使用不同方法后的果进行比较，结果如表7所示，其中P表示积极情感，N表示消极情感，O表示中性情感。

	句子样本	方面项	情感标签	ATAE-LSTM	ASGCN-DG	RoBERTa-SDACP	RoBERTa-ACOP
1	Great food but the service was dreadful!	food, service	P,N	P,N	P,N	P,N	P,N
2	The staff should be a bit more friendly.	staff	N	P	N	N	N
3	Food was okay, just so so.	food	O	P	P	O	O

Table 7: ATAE-LSTM、ASGCN-DG、RoBERTa-SDACP和RoBERTa-ACOP的预测样例

观察表7可以得到：对于样本1，三个模型：ATAE-LSTM、ASGCN-DG、RoBERTa-SDACP和RoBERTa-ACOP都给出了正确的预测结果，这表明当样本语法结构简单且语义清楚的情况下，四种方法都能准确给出方面项对应的情感极性。对于样本2，结合了注意力机制的方法ATAE-LSTM做出了错误的判断，这是因为传统的基于注意力的方法不能识别到“should”是一个虚拟语气词，这个样本实际上表达了一个否定的含义，然而融合了语义信息的模型ASGCN-DG、RoBERTa-SDACP和RoBERTa-ACOP都能准确地给出方面项对应的情感

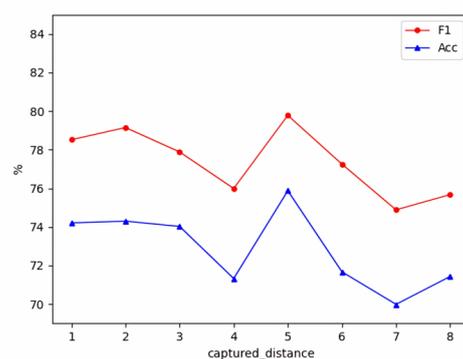
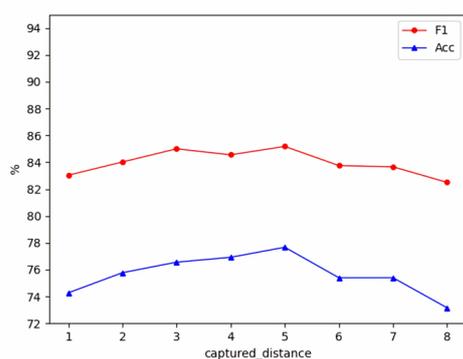


Figure 6: Rest14数据集上捕捉句法依存距离的影响 Figure 7: Lap14数据集上捕捉句法依存距离的影响

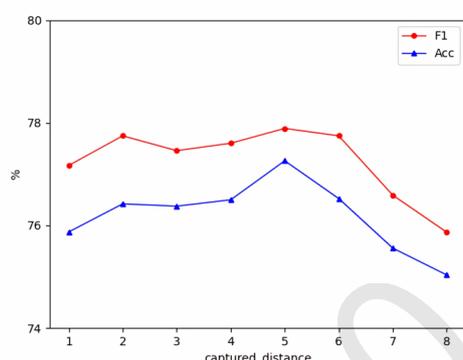


Figure 8: Twitter数据集上捕捉句法依存距离的影响

极性。对于样本3，习惯用语“just so so”削弱了“okay”情感词的积极情感，整个句子实际上表达的是一种中性的情感倾向，前两个模型ATAE-LSTM和ASGCN-DG都不能正确判断方面项对应的情感极性，只有模型RoBERTa-SDACP和RoBERTa-ACOP正确给出了情感极性，这说明了本章的模型对一些习惯用语也能提取到其语义特征。

5 总结

本文放弃了设计相应的模块引入外部知识而是从多任务的角度来解决原来的ABSA任务。本文分别提出了两个辅助任务：方面上下文顺序预测(ACOP)任务和方面上下文句法依存预测(SDACP)任务，并分别将辅助任务与原ABSA任务使用多任务方法进行训练，使得模型更好地提取句子的语义和语法特征。实验结果表明，模型在三个主要的公开数据集Rest14、Lap14和Twitter上的准确率和Macro-F1值都有较好的表现。本研究还可以从以下几个方面进一步改进:1)在ACOP任务的局部实现中，由于本文采用随机交换的方式探索方面项和上下文之间的顺序关系，引入了很多噪音，理论上会对ABSA任务带来一定的影响，因此采用一种启发式的方法避免这些噪音带来的影响。2)探索多个任务联合辅助完成ABSA任务的可能性。

参考文献

- Md Shad Akhtar, Deepak Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2017. Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems*, 125:116–135.
- Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory

- for aspect sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 452–461.
- Chenhua Chen, Zhiyang Teng, and Yue Zhang. 2020. Inducing target-specific latent structures for aspect sentiment classification. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 5596–5607.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent twitter sentiment classification. In *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, pages 49–54.
- Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3433–3442.
- Shuqin Gu, Lipeng Zhang, Yuexian Hou, and Yin Song. 2018. A position-aware bidirectional attention network for aspect-level sentiment analysis. In *Proceedings of the 27th international conference on computational linguistics*, pages 774–784.
- Binxuan Huang and Kathleen M Carley. 2019. Syntax-aware aspect level sentiment classification with graph attention networks. *arXiv preprint arXiv:1909.02606*.
- Binxuan Huang, Yanglan Ou, and Kathleen M Carley. 2018. Aspect level sentiment classification with attention-over-attention neural networks. In *Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11*, pages 197–206. Springer.
- Yohan Jo and Alice H Oh. 2011. Aspect and sentiment unification model for online review analysis. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 815–824.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- S Kiritchenko, X Zhu, C Cherry, and SM Mohammad. Detecting aspects and sentiment in customer reviews. In *8th International Workshop on Semantic Evaluation (SemEval)*, pages 437–442.
- Deepak Kumar Gupta, Kandula Srikanth Reddy, and Asif Ekbal. 2015. Pso-asent: Feature selection using particle swarm optimization for aspect based sentiment analysis. In *Natural Language Processing and Information Systems: 20th International Conference on Applications of Natural Language to Information Systems, NLDB 2015, Passau, Germany, June 17-19, 2015, Proceedings 20*, pages 220–233. Springer.
- Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. *arXiv preprint arXiv:1805.01086*.
- Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6319–6329.
- Bin Liang, Rongdi Yin, Lin Gui, Jiachen Du, and Ruifeng Xu. 2020. Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis. In *Proceedings of the 28th international conference on computational linguistics*, pages 150–161.
- Zhuang Liu, Wayne Lin, Ya Shi, and Jun Zhao. 2021. A robustly optimized bert pre-training approach with post-training. In *Chinese Computational Linguistics: 20th China National Conference, CCL 2021, Hohhot, China, August 13–15, 2021, Proceedings*, pages 471–484. Springer.
- Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.
- DG Maria Pontiki, HP John Pavlopoulos, and SM Ion Androutsopoulos. 2014. Semeval-2014 task 4: Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014), Dublin, Ireland*, pages 23–24.

- Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.
- Bhargav Srinivasa-Desikan. 2018. *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python, Gensim, spaCy, and Keras*. Packt Publishing Ltd.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 5679–5688.
- Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. *arXiv preprint arXiv:1605.08900*.
- Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 6578–6588.
- Joachim Wagner, Piyush Arora, Santiago Cortes, Utsab Barman, Dasha Bogdanova, Jennifer Foster, and Lamia Tounsi. 2014. Dcu: Aspect-based polarity classification for semeval task 4. In *SemEval@COLING*, pages 223–229.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020a. Relational graph attention network for aspect-based sentiment analysis. *arXiv preprint arXiv:2004.12362*.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020b. Relational graph attention network for aspect-based sentiment analysis. *arXiv preprint arXiv:2004.12362*.
- Hu Xu, Bing Liu, Lei Shu, and Philip S Yu. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis. *arXiv preprint arXiv:1904.02232*.
- Mi Zhang and Tiejun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 3540–3549.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. Aspect-based sentiment classification with aspect-specific graph convolutional networks. *arXiv preprint arXiv:1909.03477*.
- 王婷, 杨文忠. 2021. 文本情感分析方法研究综述. *计算机工程与应用*, 57(12):11–24.