

分类号 TP391.1  
U D C                     

密级                       
编号 10741

**兰州财经大学**

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

# 硕士学位论文

论文题目 基于依存句法树的交互式方面级情感分析

研究生姓名: 王雪绒

指导教师姓名、职称: 杨春林 教授

学科、专业名称: 管理科学与工程

研究方向: 信息管理与信息系统

提交日期: 2023年6月6日

# 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 王增斌 签字日期： 2023.5.20

导师签名： 杨春林 签字日期： 2023.5.20

# 关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定，同意（选择“同意” / “不同意”）以下事项：

1. 学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2. 学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入 CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 王增斌 签字日期： 2023.5.20

导师签名： 杨春林 签字日期： 2023.5.20

# **An Interactive Aspect-Level Sentiment Analysis Model Based on a Dependency Syntax Tree**

**Candidate : Wang XueRong**

**Supervisor: Yang ChunLin**

## 摘要

随着互联网应用的快速发展,各大社交平台的兴起,人们越来越喜欢在互联网上发表自己的观点、意见或表达情绪等。通过收集、整理并分析这些带有用户情感倾向的评论文本,商家可以更好地了解客户的行为和偏好,政府也可以及时了解群众对某项政策的看法,并适当引导公众舆论,具有一定的社会和商业价值。近些年流行的方面级情感分析是先提取方面实体,再去判断该方面实体所对应的情感态度,以一种更加细致的方式来判断某一方面实体所蕴含的情感色彩,并且当句中有多个方面时,能够将不同方面与其情感极性一一匹配,因其能够满足更多的实际需要而受到广泛关注。

目前,大多数基于方面级的情感分析模型是依靠依存句法树获取句法信息,并结合神经网络来进行情感极性的预测,但是仅根据依存树获取信息一方面无法区分不同上下文词的重要程度,导致语义信息提取不充分;另一方面会因为单独计算注意力而忽略上下文信息与目标方面之间的联系,使一些重要特征信息丢失。因此,为解决上述问题,本文提出了基于依存句法树的交互式方面级情感分析模型(An Interactive Aspect-Level Sentiment Analysis Model Based on a Dependency Syntax Tree, 以下简称为 SIGAT 模型)。本模型主要包括两个部分:一部分是基于句法注意机制的目标方面提取,该模块主要是将语法信息与注意力机制相结合。首先利用语法解析器生成依存句法树,随后使用句法注意机制对依存树上的语法信息进行编码,建模上下文词的语义关系,并有选择性的关注语法路径上接近目标方面的语境词,对句法相对距离近的意见词赋予更多的关注,同时引入高斯函数降低计算的复杂度,避免因权重急剧下降导致情感信息的丢失。另一部分是基于图注意力网络的上下文特征提取,以依存句法树为辅助,不仅考虑句法信息,也考虑单词之间的依赖关系,丰富单词的表示;最后,为加强二者之间的联系,通过交互注意力机制使基于图注意力网络的上下文特征与基于句法注意机制的目标方面特征交互学习,用对方的信息来补充自身的特征信息,提高特征表达能力,经特征融合后,输入至全连接层分类输出结果,以判断目标方面的情感态度。

最后,为证明本文所提出的 SIGAT 模型的分类效果,在三个领域的数据集:SemEval-2014 Task 4 中的餐厅评论(Restaurant)、笔记本电脑评论(Laptop)

以及 ACL-14 Twitter 上进行实验。通过与其他基线模型进行对比, SIGAT 模型在数据集上的准确率 (Acc) 和 F1 值优于其他模型, 证实了该模型的可靠性。同时, 通过消融实验和对不同影响因素的分析发现, SIGAT 模型中的依存句法信息和交互注意力机制在提高模型性能方面有重要贡献。

**关键词:** 方面级情感分析 句法信息 注意力机制 特征提取

## Abstract

With the rapid development of Internet applications and the rise of major social networking platforms, people are increasingly interested in expressing their views, opinions, etc. on the Internet. By collecting, collating, and analyzing these comment texts with users' sentiment tendencies, businesses can better understand their customers' behavior and preferences, and governments can keep abreast of the public's views on a certain policy and appropriately guide public opinion, which has certain social and commercial value. Aspect-Based Sentiment Analysis, which has become popular in recent years, extracts the aspect entities first and then determines the sentiment attitudes corresponding to them, in a more detailed way, and when a sentence contains multiple aspects, it is possible to match the sentiment polarities of different aspect entities according to their The ability to match different aspectual entities to their corresponding emotional polarity when multiple aspects are included in a sentence is of great interest because of its ability to meet a wider range of practical needs.

Currently, most aspect-level based sentiment analysis models rely on dependency syntactic trees to obtain syntactic information and combine with neural networks for sentiment polarity prediction, but obtaining information based on dependency trees alone cannot distinguish the importance of different contextual words on the one hand, resulting in inadequate extraction of semantic information; on the other hand, the connection between contextual information and the target aspect will be ignored because attention is calculated separately, making some models inefficient. On the other hand, some important feature information is lost because the contextual information and the target aspect are neglected by calculating attention alone. Therefore, to solve the above problems, this paper proposes an Interactive Aspect-Level Sentiment Analysis Model

Based on a Dependency Syntax Tree (hereafter referred to as the SIGAT model). The model consists of two main parts: one is the extraction of target aspects based on syntactic attention mechanisms. The module mainly combines syntactic information with attention mechanisms, first using a syntactic parser to generate a dependency syntax tree, then using syntactic attention mechanisms to encode the syntactic information on the dependency tree, modeling the semantic relations of contextual words, and selectively paying attention to contextual words close to the target aspect on the syntactic path, giving more attention to opinion words close to the syntactic relative distance, and introducing a Gaussian function to reduce the complexity of the computation and avoid the loss of sentiment information due to a sharp drop in weighting. Another part is the graph attention network based contextual feature extraction, aided by a dependency syntactic tree, which not only considers syntactic information but also the dependency relationship between words, enriching the representation of words; and finally, through the interactive attention mechanism, the contextual features based on the graph attention network and the target aspect features based on the syntactic attention mechanism are learned interactively, using each other's information to supplement. Finally, through the interactive attention mechanism, the contextual features based on the graph attention network and the target aspect features based on the syntactic attention mechanism are learned interactively, using each other's information to supplement their own feature information and improve the expression of the features.

Finally, to demonstrate the classification effectiveness of the SIGAT model proposed in this paper, experiments were conducted on three domain datasets: Restaurant reviews (Restaurant), Laptop reviews (Laptop) and ACL-14 Twitter in SemEval-2014 Task 4. By comparing with other

baseline models, the SIGAT model outperformed the accuracy (Acc) and F1 values on the dataset, confirming the reliability of the model. Also, the ablation experiments and the analysis of different influencing factors revealed that the syntactic information and the interaction attention mechanism on the dependency tree in the SIGAT model made an important contribution to improving the model's performance.

**Keywords:** Aspect-level sentiment analysis; Syntactic information; Attentional mechanisms; Feature extraction

# 目 录

<b>1 绪论</b> .....	<b>1</b>
1.1 研究背景及意义 .....	1
1.2 研究现状 .....	2
1.2.1 基于传统机器学习方法的方面级情感分析 .....	3
1.2.2 基于深度学习方法方面的方面级情感分析 .....	5
1.3 本文主要工作 .....	10
1.4 本文创新点 .....	11
1.5 本文结构 .....	12
<b>2 方面级情感分析的相关理论与技术</b> .....	<b>13</b>
2.1 文本情感分析概述 .....	13
2.2 词嵌入与语言模型 .....	13
2.2.1 Word2Vec 词嵌入.....	14
2.2.2 GloVe 词嵌入 .....	16
2.2.3 BERT 语言模型 .....	17
2.3 依存句法分析 .....	19
2.4 相关神经网络 .....	20
2.4.1 图形神经网络 .....	20
2.4.2 循环神经网络 .....	24
2.4.3 长短期记忆网络 .....	25
2.5 注意力机制 .....	27
2.6 本章小结 .....	29
<b>3 基于依存句法树的交互式方面级情感分析模型框架构建</b> .....	<b>30</b>
3.1 任务定义 .....	30
3.2 SIGAT 模型框架介绍 .....	31
3.2.1 输入层 .....	32

3.2.2 依存句法注意层 .....	33
3.2.3 图注意力网络层 .....	35
3.2.4 交互学习层 .....	37
3.2.5 情感输出层 .....	39
3.3 本章小结 .....	39
<b>4 实验设计与结果分析 .....</b>	<b>41</b>
4.1 评价指标 .....	41
4.2 实验数据及实验环境介绍 .....	41
4.2.1 实验数据 .....	41
4.2.1 实验环境介绍 .....	42
4.3 模型训练相关设置 .....	43
4.3.1 实验参数设置 .....	43
4.3.2 对比实验设置 .....	43
4.4 实验结果分析 .....	44
4.4.1 主要结果 .....	44
4.4.2 消融实验 .....	46
4.4.3 不同因素对模型性能的影响 .....	47
4.4.4 实例分析 .....	50
4.5 本章小结 .....	52
<b>5 总结与展望.....</b>	<b>53</b>
5.1 总结 .....	53
5.2 展望 .....	54
<b>参考文献 .....</b>	<b>55</b>
<b>致谢.....</b>	<b>62</b>

# 1 绪论

## 1.1 研究背景及意义

随着移动互联网、人工智能、物联网等技术的推动下,计算机技术已经深度渗透到人们的日常生活中,无论是交通出行,还是休闲娱乐,都极大地提高了我们的生活效率。据第 50 次《中国互联网络发展状况统计报告》显示<sup>[1]</sup>,从网络使用人口看,截止于 2022 年 6 月,我国已有 10.15 亿的网民用户,相较于 2021 年 12 月,新增用户 1919 万;从互联网普及率来看,2022 年 6 月普及率高达 74.4%,比 2021 年提升了 1.4%。随着互联网应用的发展,我国短视频用户规模达 9.62 亿,占整体网民的 91.5%;网络新闻用户规模达 7.88 亿,占整体网民的 75%;在线医疗用户规模达 3 亿,占整体网民的 28.5%。从上述数据可知,互联网已经成为社会各种信息的媒介,网民通过各类社交平台发表评论,而这些评论数据往往包含着人们对产品、服务、时事以及政策的看法和意见。但是随着互联网的快速发展,网络评论文本更新速度快、内容复杂多变,且数量庞大,仅靠人工处理已经不能满足精细化的需求,因此我们利用计算机技术将这些评论文本转化为电脑可识别处理的计算机语言,并进行系统性的分析,即自然语言处理<sup>[2]</sup>(Natural Language Processing,NLP)。

作为自然语言处理任务的重要分支,文本情感分析(Sentiment Analysis,SA)的目的是通过分析评论数据来判断人们对目标方面的情感态度,亦被称作观点挖掘<sup>[3]</sup>(Opinion Mining),根据研究内容的细粒度不同又分为文档级、句子级和方面级,包含三类情感态度:积极、消极或中性。在早期的研究中,通常以文档级情感分析和句子级情感分析任务为主,其中基于文档级的情感分析是以整个篇章为研究任务,判断该文档的整体情感;基于句子级的情感分析任务是以整条句子为单位,判断其情感态度。它们假设该段文本只有一种情感极性,并且只是在处理文本的长度上有差异,二者均属于粗粒度的情感分析。当文本或句子结构复杂、包含多个方面时,它们就无法准确地判断情感极性,在推荐系统或问答系统的应用中,会计算出错误的结果。为了更加系统全面地分析文本情感极性,我们提出使用方面级情感分析方法,以特定方面为目标,寻找其对应的情感极性。例如在

句子 “This Macbook’s price is expensive but it’s function are powerful.” 中，包含 “price”和“function” 两个方面，但二者所包含的情感极性是相反的，其中“price”的意见词 “expensive” 是消极的；而 “function” 的意见词 “powerful” 是积极的。因此，面对包含不同方面不同情感极性的句子时，仅用文档级或句子级的情感标签是不合适的，只有方面级的情感分析可以识别不同的情感极性，提供更细粒度的意见挖掘。

与此同时，文本情感分析被广泛应用在多个领域中。例如，在电子商务领域中，消费者可以参考其他的评论来决定是否购买，而商家也可以根据评论去分析消费者的偏好和态度，从而对商品或服务进行调整和改进。在监测公众舆论方面，通过分析群众对某事件或某政策的观点和态度，了解群众对具体政策和事物的看法，当出现大量的负面情绪时，能够及时、精准的进行控制和干预，从而正确地引导舆论。同样在金融领域方面，通过对新闻以及用户评论的分析，预测市场的波动，主要包括股票预测、期货预测等。随着互联网应用范围的扩大，情感分析在社会科学、医疗、金融等领域同样发挥着重要作用，并满足更多的实际需要，逐渐成为自然语言处理关注的热点和难点。

方面级情感分析的主要任务是获取与特定目标相关的信息，同时避免不相关信息的干扰，这是方面级情感分析的特点也是难点。早期的方法是通过构建情感词典和语法规则对文本情感进行分类，但这需要耗费大量的人力来构建和维护；传统的机器学习方式通过训练的模型来分类情感极性，虽在一定程度上减轻了人力，但由于网络文本数据多是无规则、无固定结构形式，并且含有大量噪声的，因此对未标注的数据仍需要人工提取特征，无法达到高效率的分类。目前，多数研究是基于深度学习方法，利用神经网络提取目方面标和上下文词的特征，虽然提高了情感分类的准确率，但仍存在意见词与目标词不能精确匹配等问题，有待进一步展开研究。

综上所述，方面级的情感分析具有重要的实践意义和商业价值，也因其具有很强的技术挑战性而具有更大的研究意义。

## 1.2 研究现状

作为文本情感分析中一个重要的分支，方面级情感分析更加着重于细粒度情

感信息的提取，其实现分为两类<sup>[3]</sup>：基于传统机器学习的方法和基于深度学习的方法。

### 1.2.1 基于传统机器学习方法的方面级情感分析

基于传统机器学习的方法就是通过使用情感词典或预先制定的规则来提取特征以训练情感极性分类器（如 SVM<sup>[4]</sup>），大致可分为三种：基于情感词典的方面级情感分析、基于监督学习的方面级情感分析和基于半监督学习与神经网络相结合的方面级情感分析。

基于情感词典的方面级情感分析，首先对文本的情感词进行匹配，而匹配对象来源于现有的情感词典，随后根据汇总得到情感词进行评分，得到该文本的情感极性。目前，英文词典主要包括 WordNet<sup>[5]</sup>和通用查询程序，中文中常用的词典是知网情感词典 HowNet<sup>[6]</sup>、台湾大学开发的情感词典 NTUSD<sup>[7]</sup>和大连理工大学开发情感词汇本体库<sup>[8]</sup>，其中使用最广泛的是情感词汇本体库。早期 Hu 等人<sup>[9]</sup>通过寻找客户评价中的高频特征，然后使用 WordNet 去判别这些高频特征的情感方向，是正向还是负向，随后再寻找非高频的特征，最后输出高频和非高频特征的情感方向。赵妍妍等人<sup>[10]</sup>基于海量的微博数据构建了一个大规模的中文情感词典，首先利用情感词语种子获取表情符种子，然后利用表情符种子计算所有候选情感词语的情感分值，最后根据情感词语的分值来构建大规模情感词典。通过在情感分类任务上进行实验，结果表明，该词典的性能远超过其他的中文情感词典。Thien 等人<sup>[11]</sup>针对越南语文本情绪分析，提出了一种越南语情感词典（VNSD），该词典采用了逻辑回归、双重传播和模糊规则等技术，并包含大约 5000 个形容词、2000 个动词、300 个名词等，为越南语文本短语情感分析做出了一定的贡献。元海霞<sup>[12]</sup>提出一种基于 Word2Vec 和 HowNet 构建情感词典的方法，通过 Word2Vec 和 HowNet 对目标词以及与目标相近的 10 个词进行情感倾向判断。大量实验证明，该方法具有较高的准确性和实用性。在此基础上，杨立月等人<sup>[13]</sup>认为微博情感分析的关键是情感词典的构建，因此他们在原有情感词典的基础上新构建了开源情感词典、语气词词典和否定副词词典等，并且在进行情感分析时，引入了句型关系、句间关系原则等，利用改进后的词典，其分类准确率有显著提升。与此类似，杜振中等人<sup>[14]</sup>也在原有情感词典的基础上，将语义

规则与扩展后的情感词汇本体库、关联副词、程度副词和特定事件情感词等多种情感词典相结合，建立了有关微博用语的情感分析模型，与词频模型的 Precision 相比，该模型准确率平均提升 0.17，F1 值平均提升 0.25。颜明阳等人<sup>[15]</sup>提出一种领域特定情感词典（DSED）的扩展方法，将词性标注、情绪词典和 GPEL 作为提取特征的外部资源，在三个公开数据集（SemEval-2007 新闻数据集、微博数据集、博客数据集）上均取得较好的结果，相比于 PMI、n 元语法等模型，该模型分类准确率更优。

基于监督学习的方面级情感分析，通过对带有特征和标签的训练数据实验，让计算机找到特征和标签之间的关联，在处理只有特征没有标签的数据时，可以根据关联判断出该数据的标签。早期是 Pang 等人<sup>[16]</sup>应用了三种代表性分类器对文本进行情感分类，包括支持向量机（SVM）、朴素贝叶斯（NB）和最大熵（ME）。其中，SVM 算法得出结果比朴素贝叶斯和最大熵算法结果更好。在此基础上，Zhang 等人<sup>[17]</sup>提出了一种基于 SVM 和语义分析的新方法，该方法是将理论和情感分析相结合，并划分了四种情感模式，分别是幸福、希望、厌恶和焦虑；公司通过使用该方法对在线评论进行分析，一方面优化了 Web 2.0 下的营销策略，另一方面为改善客户关系提供一定参考。Cai 等人<sup>[18]</sup>提出一种基于堆叠的混合模型，即把 SVM 和 GBDT 组合在一起，通过在不同数据集上的实验表明，该模型的性能优于基线单模型。一部分学者开发了基于规则的方法来处理 ABSA，该方法是以句子的依赖性分析为基础，通过预定义的规则来确定每个方面的情感倾向。如 Dwivedi 等人<sup>[19]</sup>提出一种基于规则的情感词典方法（RBM），在康奈尔电影评论数据上，该方法的精确率、召回率和 F1 值均比其他已知和已建立的情感词典方法结果更好、更准确。吕妹园等人<sup>[20]</sup>提出在加权词向量的基础上融入情感信息，该方法在原有 TF-IDF 算法中引入了情感信息，通过词嵌入技术 Word2vec 得到加权词向量表示。为证明该模型分类效果，爬取河北省部分景点游客的评论文本进行试验，其中，积极文本和消极文本的准确率均有所提高，分别是 6.1%和 6.0%，证明了该模型可以有效提高准确率。同时一些学者对分类算法做出了改进，使分类效果更好。孙翌博等人<sup>[21]</sup>为解决电商平台和商家能够精准营销的问题，提出了基于 SD-LS-SVM 算法的情感分析模型，该模型首先利用最小二乘支持向量机（LS-SVM）对评论文本进行动态化地修剪，然后使用 SD-LS-SVM 算法进行

情感分类,该模型在公开数据集上的准确率高达 85%,相较于其他模型,有较好的分类效果。

基于半监督学习与神经网络相结合的方面级情感分析方法,通过神经网络对未标记的文本进行特征提取,在面对带有标记的数据集稀缺问题时,可以更好的处理该类数据,同时改善文本情感分类的结果。在神经网络模型中,序列转换模型是学习文本特征的主要方法之一。Vaswani 等人<sup>[22]</sup>提出的 Transformer 架构,用注意力机制将编码器(Encoder)和解码器(Decoder)串联,省去了递归层和卷积层,该模型同时具有更高的并行化能力和较少的训练时间。陈珂等人<sup>[23]</sup>将情感词典与 Transformer 相结合构成新的文本情感分析模型(SLTM),该模型利用情感词典的特征信息,帮助情感词编码,同时利用 Transformer 的注意力机制专注于不同位置的情感词。最后通过 NLPCC2014 数据集实验证明,该模型比普通卷积网络和单纯基于注意力机制的卷积神经网络分类效果更好。与此同时,Hochreiter 等人<sup>[24]</sup>通过大量实验证明,长短时记忆(LSTM)可以学习到最佳的文本特征。其中,Kumar 等人<sup>[25]</sup>提出一种新的基于双向长短期记忆(Bi-LSTM)的双层注意力网络,该模型将双层注意力网络和基于支持向量回归(SVR)的监督模型相结合,使用多层感知器进行情感分析。该模型在 SemEval-2017 Task 5 的任务中,其 F1 值和 ACC 值均有所提高。在医学方面,罗熹等人<sup>[26]</sup>为解决中文电子病历的临床命名实体识别的问题,提出了基于多头自注意力神经网络的模型(MHA-BiLSTM-CRF),该模型在原有 BiLSTM-CRF 模型的基础上,引入了多头自注意力机制和医学词典,从而更准确地识别特征,在中文临床命名实体方面有了很大突破。

从上述文献可以看出,基于传统机器学习的方面级情感分析方法大部分是以特征提取为主,还有部分是以选择分类器的组合为主,这类方法在判断目标方面的情感态度时会忽略上下文的语义信息,考虑不全面,因此分类效果存在一定的争议。

### 1.2.2 基于深度学习方法的方面级情感分析

基于深度学习的方面级情感分析方法主要包括以下几步:首先对语句进行预处理,利用词嵌入技术对词向量编码,再利用神经网络(如 RNN、CNN、LSTM

等)对生成的词向量进行特征提取,最后通过全连接层输出情感分类。当面对大量特征工程且无需人工定义的训练集时,该方法能够满足自然语言处理要求,同时掌握更高层次、更加抽象的语言特征。现有的针对方面级情感分析的工作可以分为两条主线:不依赖外部语法信息的方法和使用语法信息的方法。

### (1) 不依赖外部语法信息的方法

大多数不依赖语法信息的工作是将给定的句子分成两部分输入,包括目标提及和上下文,分别从每个部分提取特征,并结合特征进行预测。例如, Tang 等人<sup>[27]</sup>提出了两种与 LSTM 相关的模块:一个是与目标相关的 TD-LSTM 模块,该模块将句子分为目标和上下文左右两部分;另一个是与目标连接的 TC-LSTM 模块,该模块将目标和上下文词嵌入作为输入,使目标和上下文信息进行语义交互,最后通过两个不同的 LSTM 传输并确定情感极性。虽然 LSTM 有效,但由于它们是时间序列,无法并行训练且像其他 RNN 一样耗时。因此 Xue 等人<sup>[28]</sup>提出了基于 CNN 和门控机制的更简单、更准确、更快的模型,该模型只专注研究两项任务:目标范畴分析和目标情感分析。通过在 SemEval-2014 数据集上的实验证明,该模型比具有注意力机制的传统 LSTM 模型更简单、更快。在现阶段的研究中,还有部分工作是通过预训练模型的微调和改进来提升分类效果,如 BERT<sup>[29]</sup>和 ALBERT<sup>[30]</sup>等已经显示出它们在广泛的自然语言处理任务中的有用性。Song 等人<sup>[31]</sup>提出了注意编码网络(AEN),该模型使用 BERT 对目标以及周围上下文进行不重复的编码,并在输出前利用两个注意力机制使目标特征和上下文特征交互学习,丰富特征的表达能力,在三个领域的公开数据集上均取得较好的结果。胡任远等人<sup>[32]</sup>提出一种多层协同卷积神经网络模型(MCNN),该模型利用预训练的 BERT 生成词向量,最后使用 Bi-LSTM 进行特征融合并输出情感极性。大量实验表明,该模型的性能优于传统神经网络模型和单纯基于 BERT 的模型。Gao 等人<sup>[33]</sup>提出了全新的 TD-BERT 模型,该模型在原有 BERT 的基础上,介绍了三种用于目标情绪分类的 BERT 变体,并在 3 个公开数据集:SemEval-2014 中 Laptop、Restaurant 和 ACL-14 Twitter 上进行实验,结果表明,与早期传统的 BERT 模型相比,TD-BERT 模型实现了最先进的性能。与此同时, Sun 等人<sup>[34]</sup>提出从目标方面出发构造辅助句,将情感分类任务转化成句子对分类任务,应用于问答系统(QA)和自然语言推理(NLI)中,并对预训练的 BERT 模型进行了微调,

该方法在 Sentihood 和 SemEval-2014 Task4 数据中准确率和 F1 值都取得了很好的结果。

此外,注意力机制也被证明是有用的,通过在执行任务时可以选择性地聚焦句子的主要部分来正确学习。Basiri 等人<sup>[35]</sup>提出一种基于注意力的深度网络模型 (ABCDM),该模型将注意力机制融入到两个独立的 Bi-LSTM 和 GRU 上,同时考虑过去方向的时间流以及未来上下文的时间流。通过在三个 Twitter 数据集上的实验表明,ABCDM 模型的 F1 值和 ACC 值均优于其他基于 DNN 模型的值,证明了该模型的良好性能。Ma 等人<sup>[36]</sup>提出一种基于交互注意力的网络模型 (IAN),分别对目标方面和上下文单独建模,并利用注意力机制将二者拼接操作,最终得到的目标和上下文表示融合了二者的交互信息。对比模型 ATAE-LSTM,准确率和 F1 值均有所提高。Li 等人<sup>[37]</sup>在此基础上提出了一种基于层次注意力的位置感知网络 (HAPN),利用位置信息一方面生成句子的位置感知表示,另一方面生成上与语境词有关的特定方面表示,进一步改进了基于注意力的模型。在 SemEval-2014 Task4 数据集上其 ACC、F1 值和 Recall 均有所提升。Tang 等人<sup>[38]</sup>针对神经网络在学习过程中存在的强模式过学习和弱模式欠学习的问题,提出了一种基于神经 ASC 模型的算法,该算法通过自动挖掘语料库中的监督信息来完善注意力机制;同时在传统的训练目标中加入正则化项,使 ASC 模型能够公平的提取上下文词,降低含有误导性上下文词的权重,并在三个公开数据 (SemEval-2014 中的餐厅评论、笔记本电脑评论和 Twitter 上取得较好的结果。程艳等人<sup>[39]</sup>提出一种基于注意力机制的模型 (MC-AttCNN-AttBiGRU),该模型包含两部分特征,一部分是用 CNN 提取目标的局部特征,另一部分是用 BiGRU 提取上下文的特征,并通过引入注意力机制来提高模型的特征提取能力。最后通过大量实验证明,相比于其他情感分类模型,该模型的预测性能更好。大多数研究只应用了一个注意力,这可能无法捕获与依赖距离较长的不同目标相关的上下文词。因此邓钰等人<sup>[40]</sup>提出一种多头注意力记忆网络 (MAMN),该网络利用有序神经元长短时记忆网络 (ON-LSTM) 获取丰富的上下文特征信息,能够更好的处理短文本情感分类问题。通过在三个不同数据集:电影评论集 (MR)、斯坦福情感树 SST-1 和 SST-2 上的对比实验,结果显示, MAMN 模型比基于 RNN 和 CNN 的基线模型分类效果更好。

还有一些研究表明,不同位置的单词对识别特定目标的情绪有不同的贡献,越靠近目标方面,则对判断其情感态度产生更大的影响。如 Fan 等人<sup>[41]</sup>依据单词之间的直线距离提出了多粒度注意力网络(MGAN),该网络可以获取目标方面与语境词之间的交互,根据位置距离(PD)对上下文词分配不同的注意权重。在 SemEval-2014 数据集和 Twitter 数据集上均取得较好的结果,同时也证明了方面级交互可以带来额外的信息并进一步提高性能。Chen 等人<sup>[42]</sup>提出基于多层注意力机制的 RAM 模型,该模型利用双向的 LSTM 生成记忆切片,并使用多注意力机制来捕获位置距离较远的情感特征,对相关的信息赋予更多的关注,最后将模型的结果输入 softmax 函数得到分类结果。通过在 SemEval-2014 中餐厅和笔记本电脑的评论和中文新闻评论数据集中的测试,结果表明,该模型可以更好的处理多方面情感特征的表达。同样, Park 等人<sup>[43]</sup>针对给定的目标方面建模,提出了两个方面术语情感分类器:单向单层网络(HRT\_One)和双向双层网络(HRT\_Bi)。其中, HRT\_Bi 模型使用两个双向双层 LSTM 网络生成内存,在整合记忆时,将位置注意力与内容注意力相结合从而解决目标方面与情感词不匹配的问题。通过在三个数据集上证明,该模型取得了不错的结果,证明了该模型的有效性,并且面对包含多个方面的句子时,双向 LSTM 的性能优于单向 LSTM。

## (2) 依赖语法结构的方法

由于句法信息能够使依赖信息保存在长句中,并且有助于缩短目标词和观点词之间的距离,因此,一部分学者选择将语法信息与神经网络相结合来进行情感极性的预测。在使用语法结构的工作中,最早是由 Qiu 等人<sup>[44]</sup>提出,他们依赖于手动定义的语法规则,并使用非神经模型;并建立意见词和目标方面的联系,这种方法的优点是只需要一个初始的意见词字典即可。该方法正式将语法信息引入了 ABSA 任务中。在此基础上,部分学者对 Aspect-Aspect 关系进行建模,利用各 Aspect 之间的依赖关系进行情感分析。Zhao 等人<sup>[45]</sup>为更好地获取目标方面的特征,提出了一种基于图卷积网络模型(SDGCN),该模型利用图卷积网络的特点,在捕获不同方面之间依赖关系的同时,利用双向注意力机制得到特定方面的表示,且包含位置编码,是第一个考虑方面之间情感依赖的方面级情感分类任务。该模型在 SemEval-2014 数据集上取得了较好的结果,同时也证明了不同方面之间的依赖关系会影响方面级情感分类的结果。Hazarika 等人<sup>[46]</sup>为更好地利用方面

间的依赖信息,提出利用 LSTM 实现各方面间信息共享的神经网络,根据句子中各个方面的表示来建模它们之间的顺序关系,利用注意力机制使模型能够集中于句子的关键部分,最后反馈给 softmax 分类器来预测对应方面的情感极性,通过在 SemEval-2014 数据集进行大量实验,结果证实了该方法的有效性。Liang 等人<sup>[47]</sup>提出一种基于图卷积网络(GCN)的交互式情感分析模型(InterGCN),通过考虑上下文词和特定方面单词之间的语法依赖关系来优化关系图,推导出以方面为中心的图形并采用跨方面 GCN,在交互式地提取特定方面的情感特征时,可以兼顾重要的上下文词和目标方面,通过在 Restaurant14、15、16 以及 Laptop14 等四个公开数据集上实验,结果显示, InterGCN 模型的 ACC 值和 F1 值均高于其他模型,证明了该模型的性能能够提高分类准确率。Majumder 等人<sup>[48]</sup>提出了一种与记忆网络相关的模型(IARM),该模型不同于以前的利用上下文信息,而是将相邻方面的信息整合到目标方面。在 Restaurant 和 Laptop 数据集上进行性能检测,结果显示,该模型的 ACC 值比现有模型平均高出 1.6%。还有一些学者是根据单词和目标方面的句法相对距离(SRD)对上下文词分配权重。He 等人<sup>[49]</sup>提出了一种将句法信息与注意力机制相结合的模型(SynAtt),该模型通过将句法信息纳入注意力机制,利用句法相对距离(SRD)对上下文词分配不同的注意权重;通过使用 SemEval-2014、2015 和 2016 的数据集进行实验,结果表明,该模型相较传统的注意力模型分类效果更好。在最新的研究中,Miao 等人<sup>[50]</sup>为解决语义信息提取不足、注意力机制计算复杂度高的问题,提出了一种上下文图注意力网络(CGAT),该模型利用句法相对距离对不同的上下文词给与不同的关注,并巧妙的引入了高斯函数作为句法权重函数,降低了计算的复杂度。通过大量实验证明,该模型能够更好的利用语义信息和句法结构信息来提高情感分类的准确性。

为获得更好的结果,一些研究使用图形神经网络(GCN)<sup>[51]</sup>或图注意力网络(GAT)<sup>[52]</sup>来整合语法知识。特别是,Sun 等人<sup>[53]</sup>探索了一种用于编码句法特征的方面级情感分析模型(CDT),该模型利用双向长短期记忆(BiLSTM)学习句子特征的表示,并将 GCN 建立在句法依存树上,以帮助目标提及和上下文之间进行信息交换。与此类似,Zhang 等人<sup>[54]</sup>将依存句法树与图卷积网络相结合,提出一种基于句子依赖关系树方面级情感分析模型(ASGCN),该模型引入一种特定

于方面的 GCN, 将句法信息和与目标相关的词合并到分类模型中。在四个基准数据集(SemEval-2014 task4、SemEval-2015 task12、SemEval-2016 task5 和 ACL14-Twitter) 上的实验表明, 图卷积网络(GCN) 可以正确捕获语法信息和远距离单词依存关系。肖泽管<sup>[55]</sup>提出一种基于 BERT 和依存句法树的模型(BERT4GCN), 该模型将多种类型的句法信息融入 BERT 的隐藏层, 以依存树为辅助, 并结合多层图卷积网络, 既考虑了文本的序列特征也考虑了树结构的语法特征。在三个领域的公开数据集上进行了实验, 结果显示, 该模型比前沿的情感分析模型性能更加优越。而 Li 等人<sup>[56]</sup>提出了一种双图卷积网络(DualGCN) 模型, 该模型利用两个图卷积同时考虑了句法结构和语义相关性; 其中, SynGCN 模块包含丰富的语法知识, 并降低了因错误的依存关系而带来的影响; SemGCN 模块则是为捕获语义相关性, 引入了注意力机制; 在此基础上, 加入正交和差分正则化器, 鼓励 SemGCN 学习 SynGCN 无法捕获的语义特征, 在三个公开数据集上均取得了较好的结果。在图注意力网络方面, Huang 等人<sup>[57]</sup>提出了一种新的面向属性层次情感分类任务的目标依赖图注意力网络模型(TD-GAT), 该模型将 LSTM 与 GAT 集成用于编码; 通过使用上述依赖图, 可以使上下文中对目标方面的情感特征通过语法结构传播。在大量的实验中, 该模型优于 GloVe 嵌入的多个基线, 并在使用 BERT 后, 性能进一步提升、分类准确率进一步提高。虽然这项工作利用依赖语法进行编码, 但没有考虑依赖标签信息。因此 Bai 等人<sup>[58]</sup>提出一种新型的关系图注意力网络(RGAT), 利用单词之间的关系来帮助衡量上下文单词对目标单词的重要性, 同时用依存标签丰富单词的表示。通过在 SemEval-2014 Task4 中餐厅评论和笔记本电脑评论、ACL14-Twitter 以及 MAMAS 等数据上的实验表明, 该模型优于大部分先进的基于语法的方法, 证明了单词间的依赖关系有助于提高分类准确性。Wang 等人<sup>[59]</sup>在此基础上提出一种关系图注意力网络(R-GAT) 来进行情感预测, 并通过重塑和修剪原始依赖树来构建面向目标方面的依赖树, 这种统一的树结构不仅能够关注方面和潜在意见词之间的联系, 而且便于统一处理和并行操作。在 SemEval-2014 和 Twitter 数据集上进行了广泛的实验, 证明了关系图形注意力网络(RGAT) 可以有效提升情感预测的准确率。

### 1.3 本文主要工作

本文的研究内容是基于深度学习方法的方面级情感分析模型,针对基于依存句法信息和注意力机制的神经网络模型。Bai 等人<sup>[58]</sup>提出一种基于依存句法树的关系图注意力网络模型,在原有句法结构的基础上,考虑了单词之间的依赖关系,丰富了单词的表示,提升了模型的性能。但上述方法也存在不足之处,首先,仅使用 BERT 模型提取目标方面的特征,无法区分不同单词对于目标方面的重要程度;其次,单独计算注意力而忽略上下文信息与目标方面之间的联系,会丢失一些重要的情感信息。为解决上述问题,本文提出基于依存句法树的交互式方面级情感分析模型 (An Interactive Aspect-Level Sentiment Analysis Model Based on a Dependency Syntax Tree, 文中简称 SIGAT 模型), 主要工作如下:

(1) 首先介绍方面级情感分析的研究背景及研究意义,随后对目前已有的方面级情感分析模型进行分类整理,并详细阐述构建该模型时使用的深度学习技术。

(2) 为有效将意见词集中于目标方面,并考虑目标方面特征与上下文特征之间的联系,本文提出 SIGAT 模型。SIGAT 模型大致包含两个部分:一部分是基于句法注意力机制的目标方面提取,对句法相对距离近的上下文词给与更多的关注;另一部分是基于图注意力网络 (GAT) 的上下文特征提取,以依存句法树为辅助,在考虑句法信息的同时,有效利用单词之间的依赖关系,丰富单词的表示;最后,通过交互注意力机制,使上下文特征和目标方面特征交互学习,经特征融合后输入至全连接层,分类输出情感极性。

(3) 为证明 SIGAT 模型的有效性,在 3 个领域的公开数据集 (SemEval-2014 Task4 中 Restaurant、Laptop 和 ACL14-Twitter) 上进行了实验, SIGAT 模型的准确率 (Acc) 和 F1 值优于其他对比模型。通过消融实验、对其他不同影响因素的分析和实例分析证实了 SIGAT 模型的可靠性,依存句法信息和注意力机制能够有效提升模型性能。

## 1.4 本文创新点

在 SIGAT 模型中包含两个创新点:

第一个是基于句法注意机制的目标方面特征提取。为区分不同上下文词的重要性,本文在得到依存句法树中的句法信息后,通过句法注意机制有选择性的关

注语法路径上接近目标方面的上下文词，对句法相对距离（SRD）近的修饰词赋予更多的权重，同时为降低计算复杂度，本文引入高斯函数作为句法权重函数，避免因权重急剧下降导致情感信息的丢失。

第二个是引入交互注意力机制。为加强目标方面与上下文之间的联系，本文引入交互注意力机制将目标方面的特征和上下文特征进行交互学习，通过对方的信息补充增强自身的特征信息，从而获得包含丰富句法信息的目标特征和表达能力强的上下文特征，提高特征的信息容纳能力。

## 1.5 本文结构

本文结构划分为以下五个章节：

第一章：绪论。本章首先介绍了选题的背景及意义，接着详细阐述了基于方面级情感分析现有的模型及结果，并简述本文的主要工作内容和创新点，最后介绍了本文的整体架。

第二章：方面级情感分析的相关理论与技术。本章主要介绍用于方面级情感分析的相关理论，主要包括：词嵌入与语言模型、依存句法分析、相关神经网络和注意力机制，为下文介绍模型做好准备，最后对本章进行总结。

第三章：基于依存句法树的交互式方面级情感分析模型框架构建。本章是对提出的 SIGAT 模型以及构建部分进行详细阐述，主要包括两个部分：一部分是基于句法注意机制目标方面特征提取，另一部分是基于图注意力网络（GAT）的上下文特征提取；最后，通过交互注意力机制，使二者交互学习增强信息表达能力，经特征融合后通过全连接层分类输出情感极性。

第四章：实验设计与结果分析。本章是对实验过程及结果进行分析，验证 SIGAT 模型的性能。首先介绍评价指标、实验环境、数据以及训练的参数；然后重点分析对比实验和消融实验的结果，证实了 SIGAT 模型优于对比模型；最后对具有代表性的实例进行分析，证实 SIGAT 模型的可靠性。

第五章：总结与展望。本章对全文内容进行总结，指出研究内容的不足之处，并对方面级情感分析的未来工作进行展望与思考。

## 2 方面级情感分析的相关理论与技术

本章将主要介绍用于方面级情感分析的相关理论与技术，其中包括文本情感分析概述、词嵌入与语言模型、依存句法分析、相关神经网络和注意力机制等五个部分。

### 2.1 文本情感分析概述

自 2000 年以来，情感分析越来越受到人们的关注，正逐渐成自然语言处理的热点话题。通过对带有丰富情感的主观性文本进行解析、处理和总结，得出人们对产品、服务、组织、实体以及属性所表达的情感倾向和观点态度<sup>[3]</sup>。按照不同的分析细腻程度可分为三个级别：文档级、句子级和目标方面级，其中文档级和句子级属于粗粒度分析任务，而方面级为细粒度分析任务。在早期阶段，情感分析任务主要集中在篇章级和句子级，但随着社交媒体和电商平台的大力发展，主观性评论越来越丰富多样，例如“开胃菜还可以，但服务很慢。”其中包括“开胃菜”和“服务”两个方面，但两个方面的情感极性不同。类似上述句子，文档级和句子级的分析任务是无法准确、深入了解的，在一些推荐系统、问答系统等应用中会出现错误的结果。因此，一些学者提出对文本中的每个目标方面进行情感极性分析，从而得到更加具体的情感倾向，这就是方面级情感分析（Aspect-Based Sentiment Analysis, ABSA）。接下来本文将介绍进行方面级情感分析所需的理论和技术进行介绍。

### 2.2 词嵌入与语言模型

词嵌入与预训练的语言模型在自然语言处理（NLP）中是为下游任务（如序列标记和情感分析）提供特征向量，将自然语言中的词语转化为计算机可以处理的数值。在近十年里，已经提出多种词嵌入的方法，如无需考虑上下文信息的 Word2Vec<sup>[60]</sup>词嵌入和 GloVe<sup>[61]</sup>词嵌入，以及基于 Transformers 的 BERT<sup>[29]</sup>语言模型。

### 2.2.1 Word2Vec 词嵌入

Word2vec 是由 Google 团队在 2013 年提出的一种开源训练词嵌入向量的工具,通过训练浅层前馈神经网络来学习目标和上下文单词之间的关系,由输入层、隐藏层和输出层组成。主要包括连续词袋模型(CBOW)和跳字模型(Skip-gram),并且二者的预测方式和预测结果均不相同。

#### (1) CBOW 模型

CBOW 模型的中心思想是通过上文词和下文词来预测目标词,同时为提升训练速度,去除了非线性的激活过程,图 2.1 为 CBOW 的结构示意图。

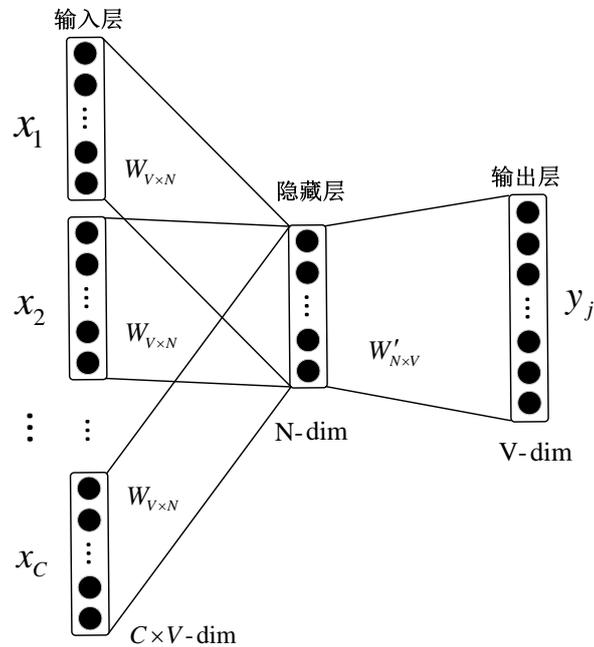


图 2.1 CBOW 模型结构图

CBOW 的输入: 设目标词  $w_i$  的上下文  $C(w_i) = \{w_j | j \in [i-R, i) \cup [i+1, i+R]\}$ , 与目标词相关的上文部分用前  $R$  个词代表, 下文部分用后  $R$  个词代表,  $e(w_i)$  表示全文每个词的词向量,  $X$  表示  $2R-1$  个词的词向量平均值, 具体表达见公式 (2.1)。

$$X = \frac{1}{2R-1} \sum_{w' \in C(w_i)} e(w') \tag{2.1}$$

其中,  $w'$  表示为文本中各个词。

CBOW 的输出：设  $P(w'|C(w_i))$  为给定上下文  $C(w_i)$  后，文本中每个词  $w'$  成为中心词的概率，并通过 softmax 函数获得，最后计算该模型的损失函数  $L$ 。具体计算见公式 (2.2) (2.3)。

$$P(w'|C(w_i)) = \frac{\exp(e_{w'}^T e_{C(w_i)})}{\sum_{v=1}^V \exp(e_v^T e_{C(w_i)})} \quad (2.2)$$

$$L = \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq i \leq c, i \neq 0} \log P(w'|C(w_i)) \quad (2.3)$$

式中  $V$  表示词典大小， $D$  表示所选词向量的长度， $e$  表示输入词  $w_i$  的词向量， $T$  表示为训练上下文词的个数 ( $T = 2R - 1$ )， $c$  表示为滑动窗口的大小。

### (2) Skip-gram 模型

与 CBOW 模型不同，Skip-gram 的中心思想是通过目标词预测周围的上下文词。即输入目标词  $w_t$ ，最后预测目标词窗口范围内的邻接词  $\{w_{t-c}, \dots, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, \dots, w_{t+c}\}$ ，具体模型如图 2.2 所示。

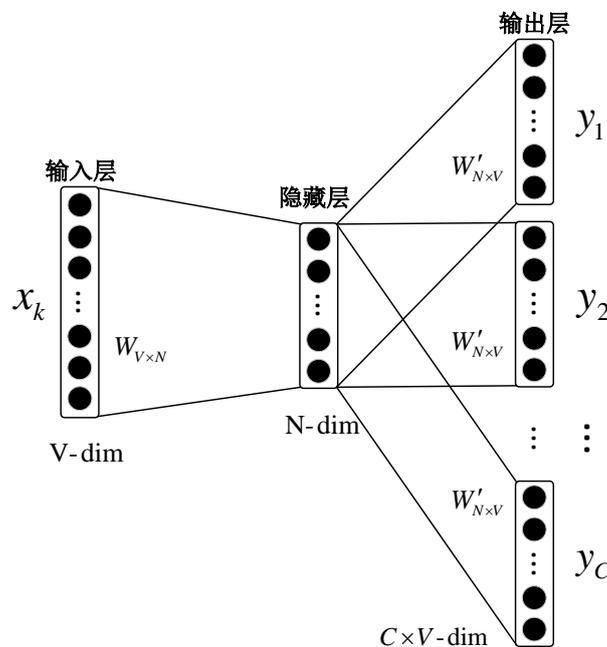


图 2.2 Skip-gram 模型结构图

假设该词在词典中的索引为  $i$ ，当为目标词时向量表示为  $v_i \in R^d$ ，上下文词

向量表示为  $u_i \in R^d$ 。设目标词  $w_c$  在词典中的索引为  $c$ ，上下文词  $w_o$  在词典中的索引为  $o$ ，通过对向量内积做 softmax 运算得到给定目标词生成上下文词的条件概率，具体计算见公式 (2.4) 和公式 (2.5)。

$$P(w_o | w_c) = \frac{\exp(u_o^T v_c)}{\sum_{i \in V} \exp(u_i^T v_c)} \quad (2.4)$$

$$E = -\sum_{t=1}^T \sum_{-m \leq j \leq m, j \neq 0} \log P(w_{t+j} | w_t) \quad (2.5)$$

式中， $E$  表示损失函数，一般是取条件概率的对数。通过梯度下降法使损失函数最小化，并假设在更新中心词的权重时，上下文词的权重是固定的，在更新上下文词的权重时，中心词的权重是固定的。在训练结束后，均可以得到两组词向量  $v_i$  和  $u_i$ ，其中  $v_i$  表示词  $i$  作为目标词的向量表示， $u_i$  表示词  $i$  作为上下文词的向量表示，具体过程见公式 (2.6)。

$$\frac{\partial \log P(w_o | w_c)}{\partial v_c} = u_o - \frac{\sum_{j \in V} \exp(u_j^T v_c) u_j}{\sum_{i \in V} \exp(u_i^T v_c)} \quad (2.6)$$

通过对比两种模型，我们发现不论是预测次数还是训练时间，Skip-gram 都要比 CBOW 多。Skip-gram 模型在预测的时候，把每个词作为目标词时都要都周围的语境词预测一次，因此训练时间会长一些。但无论是 Skip-gram 模型还是 CBOW 模型，得到的词嵌入与单词都是一一对应的，无法更好的处理容易产生歧义的词。

### 2.2.2 GloVe 词嵌入

为生成词嵌入，GloVe (Global Vectors for Word Representation) 也是一种可供选择的技术。该模型的单词表示是基于预测方法和基于计数方法(如 Word2Vec)的结合。为计算上下文  $C$  和单词  $W$  之间的关系，采用最简单的计数方法——共现计数，并且在构建损失函数时，用同样的方法计算全局向量。通过构建一个大型的共现信息矩阵，计算每一个“单词”(行)，并在大型语料库中查找该“单词”在“上下文”(列)中的频率。一般情况下，语料库的查找是在特定的范围内搜索上下文术语，而该范围是由术语前的窗口大小和术语后的窗口大小设定的，由于

“上下文”基本上是组合的，所以会产生较大的计算量。因此，为了生成便于分解的低维矩阵，对一行中每个单词的都生成一个嵌入向量，并通过“重建损失”最小化来实现，这种“重建损失”的目的就是为发现高维信息中不断变化的低维单词嵌入。GloVe 模型的具体实现步骤：

(1) 根据大型语料库构建一个共现信息矩阵，用  $X_{ij}$  (单词  $j$  出现在单词  $i$  的上下文中的次数) 表示矩阵中的每一个元素，其次再构建词向量与共现信息矩阵之间的关系，如公式 (2.7) 所示。

$$\log(X_{ij}) = w_i^T \tilde{w}_j + b_i + \tilde{b}_j \quad (2.7)$$

式中，用  $w \in R^d$  表示词向量，用  $\tilde{w} \in R^d$  表示目标词周围上下文词语的词向量， $b_i$  和  $\tilde{b}_j$  表示两个可训练的参数向量。在公式 (2.7) 的基础上，构造损失函数，具体见公式 (2.8)。

$$J = \sum_{i,j=1}^V f(X_{ij}) (w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log(X_{ij}))^2 \quad (2.8)$$

其中  $V$  表示语料库中词语数量的大小。

(2) 为保证共现频率高的词语权重大，共现频率低的词语权重小，对损失函数添加权重项，如公式 (2.9) 所示。

$$f(x) = \begin{cases} \left(\frac{x}{X_{\max}}\right)^\alpha & \text{if } x < X_{\max} \\ 1 & \text{otherwise} \end{cases} \quad (2.9)$$

通过对两种词嵌入模型的对比，Word2Vec 是采用局部上下文窗口的方法，每次都是使用一个窗口的信息来更新词向量，可能会忽略目标词与上下文的关系。相比之下，GloVe 模型是将全局的先验统计信息与局部上下文窗口结合，在控制词语权重的同时提高了模型的训练速度，并且能通过多个窗口更新词向量。

### 2.2.3 BERT 语言模型

BERT (Bidirectional Encoder Representations from Transformers) 是 Google 团队在 2019 年提出的一种基于 Transformer 的预训练语言模型，是一种语言转换的方式。它通过考虑左右语境来学习文本的深度表示，因此被称为深度双向编码，

可以用于大型语料库上训练通用语言模型，而这些模型可以很好的处理自然语言任务。BERT 模型如图 2.3 所示。

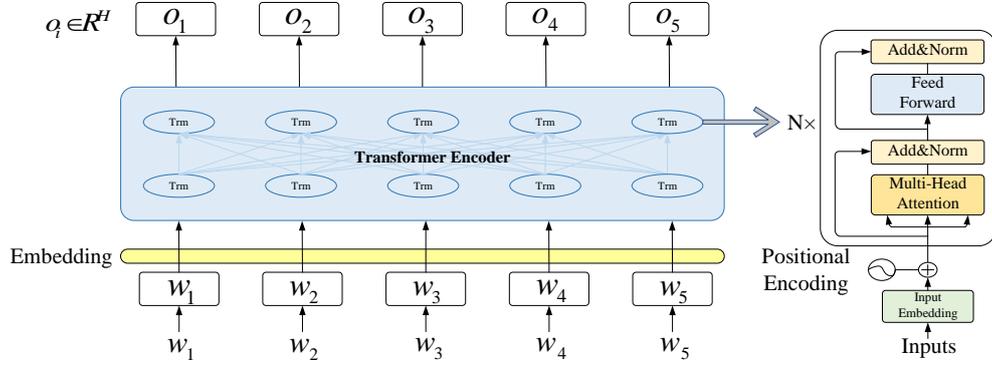


图 2.3 BERT 结构图

在使用 BERT 语言模型时，主要包括两个步骤：预训练和微调。通常情况下，我们在预训练阶段使用未标记的数据，其中包括两个任务。第一个任务是 Masked-LM (MLM)，随机屏蔽一些需要预测的单词，并用特殊符号[mask]代替，让模型根据所给的标签来学习[mask]词，此过程就是 MLM。第二个任务是预测下一句 (NSP)，即在语料库中选择句子 A 和句子 B，当句子 B 是 A 的下一句时，则将其标记为 IsNext；若与句子 A 无关系，只是语料库中的随机句，则将其标记为 NotNext，且二者的可能性各占一半；随后用[CLS]和[SEP]对全文 C 的信息进行二分类预测。大多数 NSP 任务是以两个句子的关系为基础，并非通过语言模型直接捕获。

根据 BERT 的两个预训练任务可知，其损失函数也是由两部分组成，第一部分来自 Masked-LM 任务，第二部分来自 NSP 任务。在 Masked-LM 和 NSP 的损失函数中，令集合  $M$  为被[mask]所代替的词，集合  $N$  为需要预测的句子集损失函数见公式 (2.10) (2.11) 和公式 (2.12)。

$$L(\theta, \theta_1, \theta_2) = L_1(\theta, \theta_1) + L_2(\theta, \theta_2) \quad (2.10)$$

$$L_1(\theta, \theta_1) = -\sum_{i=1}^M \log p(m = m_i | \theta, \theta_1), m_i \in [1, 2, \dots, |V|] \quad (2.11)$$

$$L_2(\theta, \theta_2) = -\sum_{j=1}^N \log p(n = n_j | \theta, \theta_2), n_j \in [IsNext, NotNext] \quad (2.12)$$

在三个公式中， $\theta$ 表示模型中 Encoder 的参数， $\theta_1$ 表示 Masked-LM 任务中 Encoder 的参数， $\theta_2$ 表示 NSP 任务中 Encoder 的参数， $|V|$ 表示该词典的大小；IsNext 表示预测为下一句的集合，NotNext 表示为语料库中随机句的集合。

BERT 的思想很大一部分是源于 CBOW 模型，但由于 BERT 是基于 Transformer 的，因此它拥有强大的语言表达能力和特征提取能力，并且可以使用更加复杂的模型以及大量的语料库数据，在处理一些自然语言任务时，准确率会比 Word2Vec 模型高。

## 2.3 依存句法分析

依存句法分析<sup>[62]</sup>也被称为依赖性分析，是自然语言处理中重要的一环，其目的是分析句子的语法结构或是分析单词之间的依存关系，并用简单易懂的图形结构（通常是树形结构）表示出来，其中一个节点表示一个单词，图的边缘标签表示节点之间的句法关系。通常情况下，我们会在依存结构中加一个根节点（ROOT），保证句中每个单词都有一个依存。如图 2.4 所示，是句子“The iced Americano at this airport tastes good”的依存句法结构树，其中，“iced Americano”表示目标方面，“good”代表情感词，若两个单词之间有箭线连接，则表示单词存在一定的依存关系，箭线上方的标签表示依存关系类型。在图 2.4 中，方面词“iced Americano”与情感词“good”直线距离较远，但在依存句法树中距离较近。

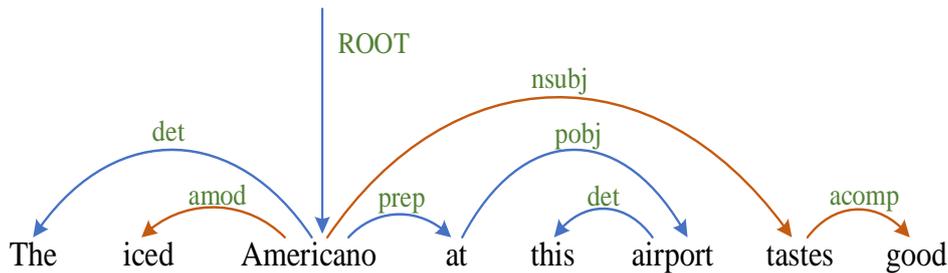


图 2.4 依存句法树示例图

句法树中部分依存关系标签及含义如表 2.1 所示。

表 2.1 关系标签含义

标签	含义
det	决定词
prep	介词修饰语
pobj	介词宾语
amod	形容词
nsubj	名词性主语
acomp	形容词补语
root	根节点

在进行句法分析时，会使用一些分析工具，如斯坦福大学开发的 StanfordCoreNLP、哈工大开发的 LTP、SpaCy 分析器以及解析树等，它们通过捕获句子语法结构的方式来提取句子的依赖性分析，并将依赖关系映射到一个有向图上。在细粒度的情感极性分析中，若单纯地关注上下文，而忽略单词之间的依存关系，可能无法正确判断情感极性，因此，依赖性分析在语义分析、方面级情感分析等自然语言处理任务中都是有益的。

## 2.4 相关神经网络

在处理大规模文本语料时，深度学习已成为自然语言处理中处理各种任务的主要方法，如循环神经网络（RNN）、卷积神经网络（CNN）和注意力机制等都已广泛用于文本序列的建模。近年来，图形神经网络（GNN）也被证明可以处理各种自然语言任务，它们利用文本序列中的句子结构信息，同时结合特定的知识来扩展原始序列数据，在各大数据集上均取得了不错的结果。接下来，本文将对一些深度学习技术进行详细介绍。

### 2.4.1 图形神经网络

图形神经网络是将 NLP 任务的文本序列数据转化为图结构的数据，而实体之间的依赖关系就是靠这些图结构数据进行编码，从而学习到更多的特征信息。

例如使用文本序列中的句子结构信息（如句法分析树），或序列数据中的语义信息（如语义解析图、信息提取图）等来增强原始序列数据，从而提高模型性能。接下来本文主要介绍图卷积网络（GCN）<sup>[51]</sup>和图注意力网络（GAT）<sup>[52]</sup>。

### （1）图卷积网络

图卷积网络（GCN）是一种用于编码非结构化数据的深度学习方法，也是传统卷积神经网络的一种自适应。在给定的图中，利用图卷积网络计算以其相邻节点为条件的节点表示，并通过叠加多层 GCN 网络聚合来自相邻节点的信息，从而更新每个节点的表示。目前，大多数自然语言处理任务都是通过图神经网络利用句法信息和单词之间的长期依赖关系进行方面级情感分类，并在各大公开数据集上均取得了较好的结果。

给定图  $G$  包含节点集  $\nu$  和边集  $\varepsilon(G(\nu, \varepsilon))$ ，通过对图  $G$  计算得到一个邻接矩阵  $A \in R^{n \times n}$ ，其中  $n$  表示节点集  $\nu$  的大小。为了方便起见，我们用  $h_j^{(l)}$  表示节点  $j$  在 GCN 第  $l$  层生成的隐藏表示，用  $h_j^{(0)}$  表示节点  $j$  的初始状态， $h_j^{(L)}$  表示节点  $j$  的最终状态。具体的节点更新操作见公式（2.13）。

$$h_i^{(l+1)} = \sigma \left( \sum_{j=1}^n A_{ij} \left( W^{(l)} h_j^{(l)} + b^{(l)} \right) \right) \quad (2.13)$$

式中， $W^{(l)}$  表示第  $l$  层中的线性变换权重， $b^{(l)}$  表示第  $l$  层中的偏置项， $\sigma$  表示非线性激活函数。 $A_{ij}$  表示节点  $i$  和  $j$  之间的邻接矩阵，若节点  $i$  与节点  $j$  相连，则  $A_{ij} = 1$ ；若节点  $i$  与节点  $j$  不相连，则  $A_{ij} = 0$ 。

由于图卷积过程只能编码直接相邻节点的信息，因此图中的节点只能受图卷积网络中  $L$  步内的相邻节点的影响。

### （2）图注意力网络

图注意力网络（GAT）是图神经网络（GNN）的一种变体，它利用掩蔽的自注意层对图结构进行编码，且不依赖具体的图网络结构。与其他类型的图神经网络（如 GCN）相比，为能够更好的聚合邻节点，GAT 引入了注意力机制，同时根据不同的邻接点特征分配不同的权值，最后通过对隐藏状态加权求合，得到目标节点的表示。在方面级情感分析中，GAT 网络是将方面的上下文语法传递到方

面目标节点，使用 L 层的 GAT 网络，并将第 L 层的特征传递到方面目标节点。如图 2.5 所示为 GAT 的注意系数计算示意图。

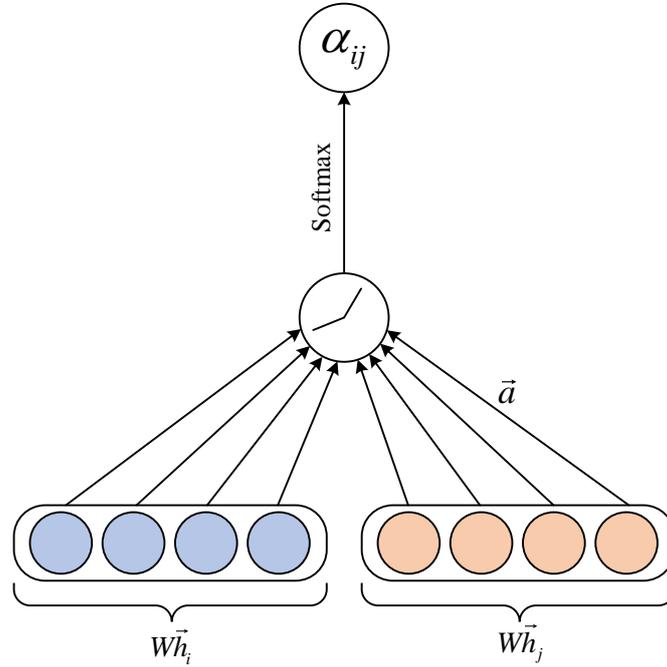


图 2.5 注意系数计算示意图

首先，逐个计算顶点  $i$  与其相邻顶点 ( $j \in N(i)$ ) 之间的相似系数，随后对相似系数进行归一化处理，得到两个顶点 ( $i$  与  $j$ ) 之间的注意力系数  $\alpha_{ij}$ ，计算过程见公式 (2.14) 和公式 (2.15)。

$$\alpha_{ij}^z = \frac{\exp(f(h_i^{l-1}, h_j^{l-1}))}{\sum_{j' \in N(i)} \exp(f(h_i^{l-1}, h_{j'}^{l-1}))} \quad (2.14)$$

$$f(h_i^{l-1}, h_j^{l-1}) = \frac{(W_Q^z h_i^{l-1})(W_K^z h_j^{l-1})}{\sqrt{d/z}} \quad (2.15)$$

在公式 (2.14) 中， $\alpha_{ij}^z$  表示节点  $i$  在  $l$  层的第  $z$  个注意头中，对邻节点  $j$  的注意系数， $h_i^{l-1}$  和  $h_j^{l-1}$  分别表示节点  $i$  和节点  $j$  在  $l$  层的隐藏状态， $f$  表示注意函数。

在公式 (2.15) 中，用  $W_Q^z, W_K^z \in \mathbb{R}^{\frac{d}{z} \times d}$  表示  $l$  层中第  $z$  个注意头的参数矩阵， $d$  表

示词向量的维数， $Z$  表示注意头的个数。

通过上述运算得到了不同顶点之间的注意力系数，随后用该注意力系数来预测每个节点的输出特征。即在  $l$  层处，给定单词  $w_i$ 、其隐藏状态  $h_i^{l-1}$  和相邻节点  $w_j \in N(i)$ ，GAT 通过多头注意力来计算  $l$  层处节点的隐藏状态，从而更新目标节点的表示。如图 2.6 所示为 GAT 的特征加权示意图。

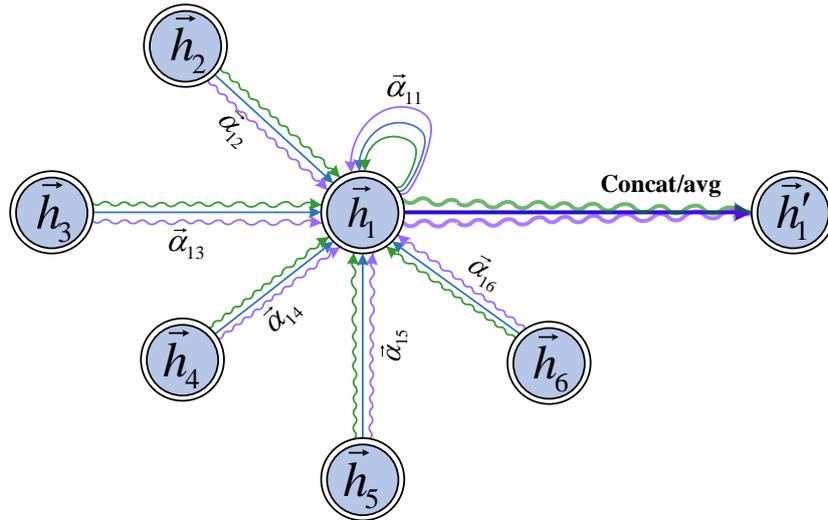


图 2.6 特征加权示意图

基于多头注意力的 GAT 聚合过程见公式 (2.16)。

$$h_i^l = \parallel \sigma \left( \sum_{j \in N(i)} \alpha_{ij}^z W_V^z h_j^{l-1} \right) \quad (2.16)$$

式中，若  $j \notin N(i)$ ，则  $\alpha_{ij}^z = 0$ ； $\parallel$  表示向量之间的串联， $W_V^z \in R^{\frac{d}{z} \times d}$  表示层  $l$  处第  $z$  个注意头的线性变换矩阵， $Z$  表示注意头的个数， $\sigma$  表示 sigmoid 激活函数。

相比图神经网络，GAT 网络将注意力机制引入了图模型，通过堆叠网络层来获取每个节点的深层特征，并为领域中不同节点分配不同的权重，优化了图神经网络中边权重分配问题，且摆脱了拉普拉斯矩阵限制，便于处理有向图结构。在归纳推理等自然语言处理任务中，GAT 网络比图神经网络表现更加出色。

### 2.4.2 循环神经网络

循环神经网络 (RNN) [9] 是一种能够处理长度可变的序列输入的前馈神经网络, 适合于包含任意时空维度的顺序预测任务, 是目前最流行的序列预测建模方法之一。在近十年里, 深度 RNN 在一些 NLP 任务中取得了最先进的性能, 如音乐生成、语音识别、视频行为识别和实体名字识别等。

RNN 使用神经网络对输入特征与未来目标变量之间的函数关系进行建模。如图 2.7 所示, 这些神经网络都是由激活的隐藏状态  $y_t$  组成, 其状态不仅受内部状态  $S_t$  的影响, 而且与当前输入状态  $x_t$  和前一内部状态  $S_{t-1}$  有关。我们给定一个时间序列  $X = (x_1, x_2, \dots, x_t)$ , RNN 通过关注  $t-1$  时刻到  $t$  时刻的隐藏状态从历史数据的训练集中不断进行递归学习, 其目标就是预测未来  $t+1$  时刻的目标变量  $y_{t+1}$ 。

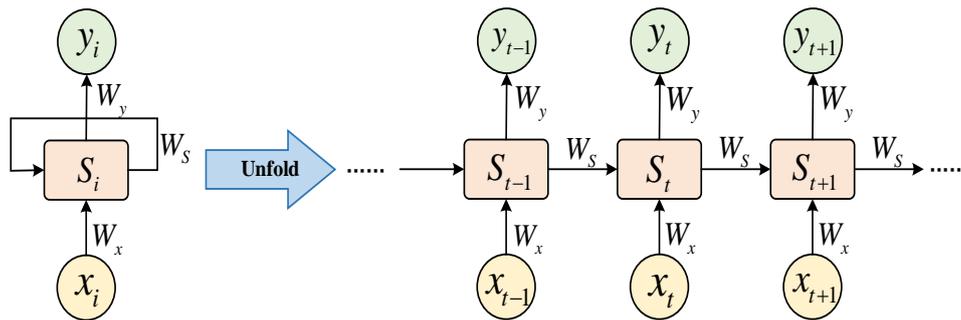


图 2.7 RNN 结构图

在 RNN 中, 基于激活单元, 需要初始化三种不同类型的权重, 若激活单元是  $\tanh$  或  $\text{sigmoid}$ , 则使用  $\text{xavier}$  或  $\text{glorat}$  初始化; 若激活单元是  $\text{Relu}$ , 则使用“ $\text{He}$ ”来初始化。因此, 最终模型是由三个参数矩阵所决定:  $W_x$ 、 $W_y$  和  $W_s$ , 与此同时还有两个有助于定义模型的偏置向量  $b_s$  和  $b_y$ 。具体计算过程见公式 (2.17) (2.18)。

$$S_t = \tanh(W_{xs} \cdot (x_t \oplus S_{t-1}) + b_s) \tag{2.17}$$

$$y_t = \sigma(W_y \cdot S_t + b_y) \tag{2.18}$$

在公式 (2.17) 和公式 (2.18) 中, 用  $x_t$  ( $x_t \in R^m$ ) 表示  $t$  时刻的输入,  $S_t$  表示

$t$ 时刻的内部隐藏状态， $y_t$ 表示 $t$ 时刻的输出；用 $W_{xs} \in R^{n \times (m+n)}$ 和 $W_y \in R^{n \times n}$ 表示矩阵参数， $n$ 代表RNN层中神经元的数量；用 $b_s \in R^n$ 和 $b_y \in R^n$ 分别表示内部状态和输出的偏置向量；用 $\sigma$ 表示sigmoid激活函数；而 $x_t \oplus S_{t-1}$ 表示 $x_t$ 与 $S_{t-1}$ 向量联级。

### 2.4.3 长短期记忆网络

长短期记忆网络(LSTM)<sup>[24]</sup>是RNN的一种形式，解决了RNN中梯度消失的问题。与其他的神经网络相比，LSTM使用分层连接的存储块来代替互相连接的神经元，以便储存来自先前时间段的信息。LSTM通过RNN单元中的门来调节信息流，这些门可以决定时间序列中保留或者丢弃哪些信息。LSTM网络为处理长序列问题，对学习长期依赖关系进行更加精细的控制，设计了遗忘门、输入门和输出门，具体模型如图2.8所示。

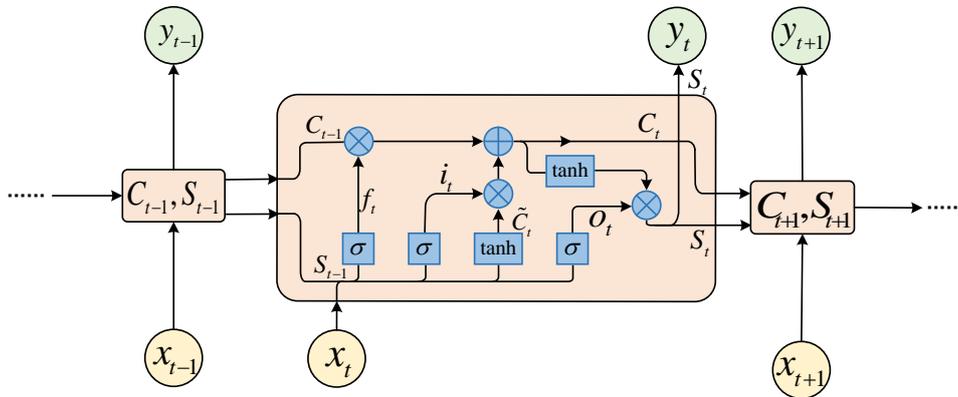


图 2.8 LSTM 模型图

(1) 遗忘门：该门可以删除或者保留单元状态  $C_{t-1}$  中的部分信息。读取先前的隐藏状态  $S_{t-1}$  和当前输入  $x_t$ ，并通过 sigmoid 函数传递结果特征，具体过程见公式 (2.19)。

$$f_t = \sigma(W_f \cdot (x_t \oplus S_{t-1}) + b_f) \tag{2.19}$$

式中，用  $x_t \in R^m$  表示  $t$  时刻的特征向量， $f_t$  表示遗忘门， $S_{t-1}$  表示  $t-1$  时刻的

隐藏状态， $\sigma$  表示 sigmoid 激活函数；用  $W_f \in R^{n \times (m+n)}$  表示遗忘门中的参数矩阵， $n$  表示 LSTM 层中的神经元数量；用  $b_f \in R^n$  表示偏置向量， $\oplus$  表示向量之间的连接。

(2) 输入门：该门可以决定当前输入  $x_t$  中哪些信息应该被保留在单元状态  $C_t$  中，控制数据进入 LSTM 的存储块。该门可以避免无关紧要的信息进入 LSTM。具体计算过程见公式 (2.20) 和公式 (2.21)。

$$i_t = \sigma(W_i \cdot (x_t \oplus S_{t-1}) + b_i) \quad (2.20)$$

$$\tilde{C}_t = \tanh(W_c \cdot (x_t \oplus S_{t-1}) + b_c) \quad (2.21)$$

式中，用  $i_t$  表示输入门， $\tilde{C}_t$  表示备选单元状态， $\sigma$  表示 sigmoid 激活函数；用  $W_i, W_c \in R^{n \times (m+n)}$  表示输入门的参数矩阵， $n$  表示 LSTM 的神经元数量；用  $b_i, b_c \in R^n$  表示输入门的偏置向量，用  $\oplus$  表示向量之间的连接。

(3) 输出门：该门决定了当前单元状态  $C_t$  中哪些信息应该被作为隐藏状态  $S_t$  的输出，根据前一存储块  $S_{t-1}$  和当前输入  $x_t$  来决定对输出执行什么操作。具体计算过程见公式 (2.22) (2.23) 和公式 (2.24)。

$$o_t = \sigma(W_o \cdot (x_t \oplus S_{t-1}) + b_o) \quad (2.22)$$

$$S_t = \tanh(C_t) \cdot o_t \quad (2.23)$$

$$y_t = \sigma(W_y \cdot S_t + b_y) \quad (2.24)$$

式中，用  $o_t$  表示输出门， $S_t$  表示  $t$  时刻的隐藏状态， $y_t$  表示  $t$  时刻的输出， $\sigma$  表示 sigmoid 激活函数；用  $W_o \in R^{n \times (m+n)}$  和  $W_y \in R^{n \times n}$  表示输入门的参数矩阵， $n$  表示 LSTM 的神经元数量；用  $b_o, b_y \in R^n$  表示偏置向量，用  $\oplus$  表示向量之间的连接。

(4) 单元状态  $C_t$ ：它可以长期存储数据，并充当额外的支持存储器，在整个过程流中跟踪相关信息。 $C_t$  由两部分组成，一部分是前一个单元状态  $C_{t-1}$  与遗忘门向量  $f_t$  的乘积，另一部分是备选单元状态  $\tilde{C}_t$  与输入门向量  $i_t$  的乘积。具体计

算过程见公式 (2.25)。

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{2.25}$$

与 RNN 相比, LSTM 网络最主要的进步就是能够保持随时间传播的信息, 使误差更加稳定并克服消失梯度问题。

## 2.5 注意力机制

注意力机制<sup>[22]</sup> (Attention Mechanism) 用于对文本情感有不同贡献的词分配不同的权重, 是目标与上下文之间的语义关联度量。一般情况下, 每个意见词对句子语义的贡献并不相等, 出现在目标附近的词或者与目标有修饰关系的词会更重要。此外, 若一个句子中有多个目标方面, 则上下文词的重要性也应该不同。例如在句子 “This Macbook’s price is suitable and it’s function are powerful.” 中, 对于方面词 “price”, 上下文中的 “suitable” 比 “powerful” 更重要, 而对于方面 “function” 来说, “powerful” 就比 “suitable” 更重要。图 2.9 为注意力机制模型图。

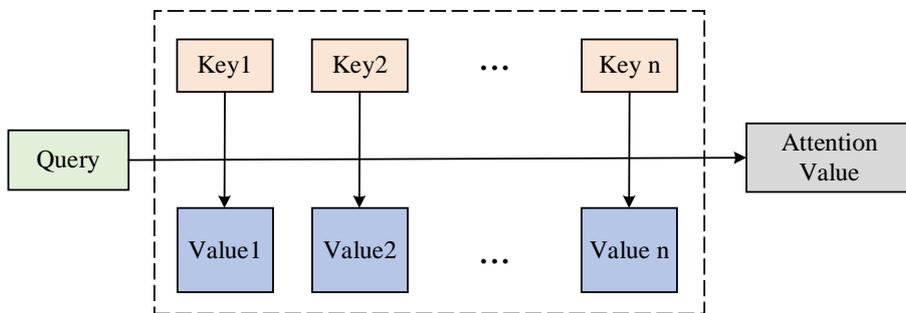


图 2.9 注意力机制结构图

Attention 机制的实质就是 “检索” 的过程。给定输入序列  $X=(x_1, x_2, \dots, x_m)$ , 通过线性函数投影得到三个不同的表示: Query、Key 和 Value。其计算过程主要包括两步: 第一步是计算 Query 和 Key 的相似性, 并通过归一化处理得到其权重系数; 第二步是根据得到的权重系数对 Value 加权求和, 最终得到注意力系数, 公式 (2.26) 展示了具体的计算过程。

$$Attention(Q, K, V) = \sum_{i=1}^M Similarity(Q, K_i) \cdot V_i \tag{2.26}$$

式中,  $Q, K, V \in R^{m \times d_m}$ ,  $d_m$  表示线性投影输出的维度,  $M$  表示输入序列长度。其中, 在计算 Query 和 Key 的相似系数时, 有三种常见的方法, 如公式 (2.27)、公式 (2.28) 和公式 (2.29) 所示:

(1) 点积法:

$$\text{Similarity}(Q, K_i) = Q \cdot K_i \tag{2.27}$$

(2) Cosine 相似性:

$$\text{Similarity}(Q, K_i) = \frac{Q \cdot K_i}{\|Q\| \cdot \|K_i\|} \tag{2.28}$$

(3) MLP 网络法:

$$\text{Similarity}(Q, K_i) = \text{MLP}(Q, K_i) \tag{2.29}$$

在 NLP 任务中, 早期的研究是将注意力机制运用到机器翻译 (NMT) 中。图 2.10 为基于 Attention 机制的 Encoder-Decoder 模型图, 在给定输入序列  $X = (x_1, x_2, \dots, x_m)$  后, 首先用 Encoder 将输入序列编码成固定维度的语义向量  $C_i$ , 其中  $C_i$  是通过隐藏层加权求和得来, 最后再用 Decoder 进行解码得到目标语言序列  $Y = (y_1, y_2, \dots, y_n)$ 。

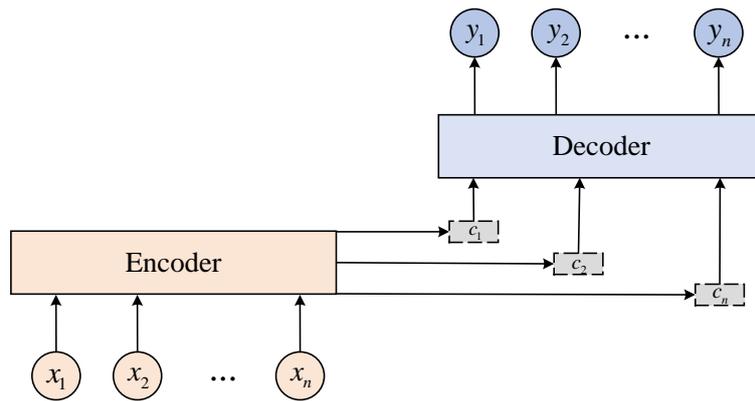


图 2.10 基于 Attention 机制的 Encoder-Decoder 模型图

具体计算过程见公式 (2.30)。

$$C_i = \sum_{j=1}^m a_{ij} h_j \tag{2.30}$$

式中,  $a_{ij}$  表示分配权重,  $i$  表示输入序列中的单词,  $j$  表示目标序列中单词,  $h_i$  表示隐藏向量。

注意力机制速度快、参数少、效果好, 在处理单元预测结果时, 循环神经网络需要等待上一个单元的预测结果, 而注意力机制可以并行处理, 这大大地提升了运行速度。在机器阅读 (NMT)、语义角色标注 (SRL) 等自然语言处理任务中, 均取得了较为出色的成果。

## 2.6 本章小结

本章对用于方面级情感分析所需的技术及原理过程进行了解释, 为后文介绍模型框架奠定了理论基础。首先介绍的是词嵌入技术与预训练语言模型: Word2Vec、GloVe 以及 BERT 模型; 接下来是对依存关系分析和相关深度学习技术的介绍, 其中包括图卷积网络 (GCN)、图注意力网络 (GAT) 和循环神经网络 (RNN); 最后介绍了热门的 Attention 机制以及基于注意力机制机制的 Encoder-Decoder 模型。

### 3 基于依存句法树的交互式方面级情感分析模型框架构建

基于方面级的情感分析旨在识别句子中特定方面的情感极性。近年来,利用图神经网络获取依赖语法信息已成为最流行的趋势,尽管它取得了成功,但仅依靠依存句法树获取信息一方面无法区分不同上下文词的重要程度,导致语义信息提取不充分;另一方面会因为单独计算注意力而忽略上下文信息与目标方面之间的联系,使一些重要特征信息丢失。针对于此,本文提出了 SIGAT 模型,首先介绍任务定义以及相关符号说明,随后详细阐述模型的整体框架及各个构建部分,最后,对本小章进行总结。

#### 3.1 任务定义

我们将方面级情感分析模型视为由三个部分组成:句子、目标方面、依存句法树。其中用  $S = \{w_1, w_2, \dots, w_{i+m}, \dots, w_n\}$  表示句子,  $n$  代表句子中单词数;用  $T = \{w_1, w_2, w_{i+1}, \dots, w_{m+i-1}\}$  表示目标方面,  $m$  代表方面个数。用  $G = (V, A, R)$  表示句子的依存句法树,其中  $V$  代表所有的节点,即单词  $\{w_1, w_2, w_3, \dots, w_n\}$ ;  $A \in \mathbb{R}^{n \times n}$  代表相邻矩阵,若单词  $w_i$  和  $w_j$  存在依赖关系,则  $A_{ij} = 1$ , 否则  $A_{ij} = 0$ ;  $R$  表示标签矩阵,当  $A_{ij} = 1$  时,  $R_{ij}$  代表  $A_{ij}$  对应的标签,若  $A_{ij} = 0$ , 则  $R_{ij} = None$ 。

在方面级情感分析 (ABSA) 中, 一个重要的分支是面向方面的情感分类, 其目标是预测每个目标方面  $w_i \in T$  的情感极性  $y_i$ ,  $y_i \in \{\text{积极}, \text{消极}, \text{中立}\}$ 。如表 3.1 所示为数据集中部分评论, 其中加粗字体表示特定方面, 蓝色下标表示该目标方面所对应的情感极性。例如在 “This Macbook’s price is expensive but it’s function are powerful.” 中, 包含 “price” 和 “function” 两个方面, 但二者情感极性是相反的, 其中 “price” 的意见词 “expensive”, 是消极的; 而 “function” 的意见词 “powerful”, 是积极的。面对包含两个方面的句子, 给一个句子级别的情感标签是不合理的, 本文的任务是依据不同方面判断其相应的情感极性, 进行细粒度的情感分析。

表 3.1 方面级情感分析示例

The <b>food</b> <sub>pos</sub> is great but the <b>service</b> <sub>neg</sub> and the <b>environment</b> <sub>neg</sub> are dreadful.
The <b>iced Americano</b> <sub>pos</sub> at this airport tastes good.
The <b>appetizers</b> <sub>pos</sub> are ok, but the <b>service</b> <sub>neg</sub> is slow.
The <b>Chinese dumplings</b> <sub>pos</sub> in this restaurant taste very good.
Can be a bit busy around peak times because of the <b>size</b> <sub>pos</sub> .

### 3.2 SIGAT 模型框架介绍

本节将介绍所提出的基于依存句法树和交互注意力机制的 SIGAT 模型。模型大致可分为两部分，一部分是基于句法注意机制的目标方面特征提取，利用该机制对依存树上的语法信息进行编码，有选择性的关注语法路径上接近目标方面的上下文词，并赋予较高的注意权重，与此同时，引入高斯函数作为其句法权重函数，降低计算复杂度；另一部分是基于图注意力网络的上下文特征提取，不仅考虑了句法信息，也考虑了单词之间的依赖关系，丰富了文本特征表示；最后通过交互注意力机制使二者互相学习，增强信息容纳能力，经特征融合后输入至全连接层分类输出情感极性。图 3.1 展示了 SIGAT 框架结构，主要包括五个模块：

(1) 输入层：利用 BERT 进行词向量的转化，将文本数据转换为计算机能够识别处理的状态。

(2) 依存句法注意层：为更好的提取目标方面的特征，本文首先利用语法解析器（Deep Biaffine Parse）生成依存句法树，在此基础上，通过句法注意机制有选择性地关注语法路径上接近目标方面的上下文词，并赋予较高的权重。为避免因权重急剧下降导致丢失情感信息，在此基础上引入了高斯函数，同时降低计算复杂度，有效的将意见词集中到了目标方面上。

(3) 图注意力网络层：以依存句法树为辅助，将句法树上的依赖关系引入注意力机制，引导信息从句子的上下文传播到目标提及本身，更好的捕获单词之间的关系，丰富单词的表示。

(4) 交互学习层：使得基于句法注意机制的目标方面特征与基于图注意力

网络的上下文特征交互学习，通过对方的信息来补充增强自身的特征表达能力。

(5) 输出层：经过交互注意学习后得到包含丰富句法信息的目标特征和表达能力强的上下文特征，经特征融合后输入至全连接层，计算出正确的方面情感极性并输出。

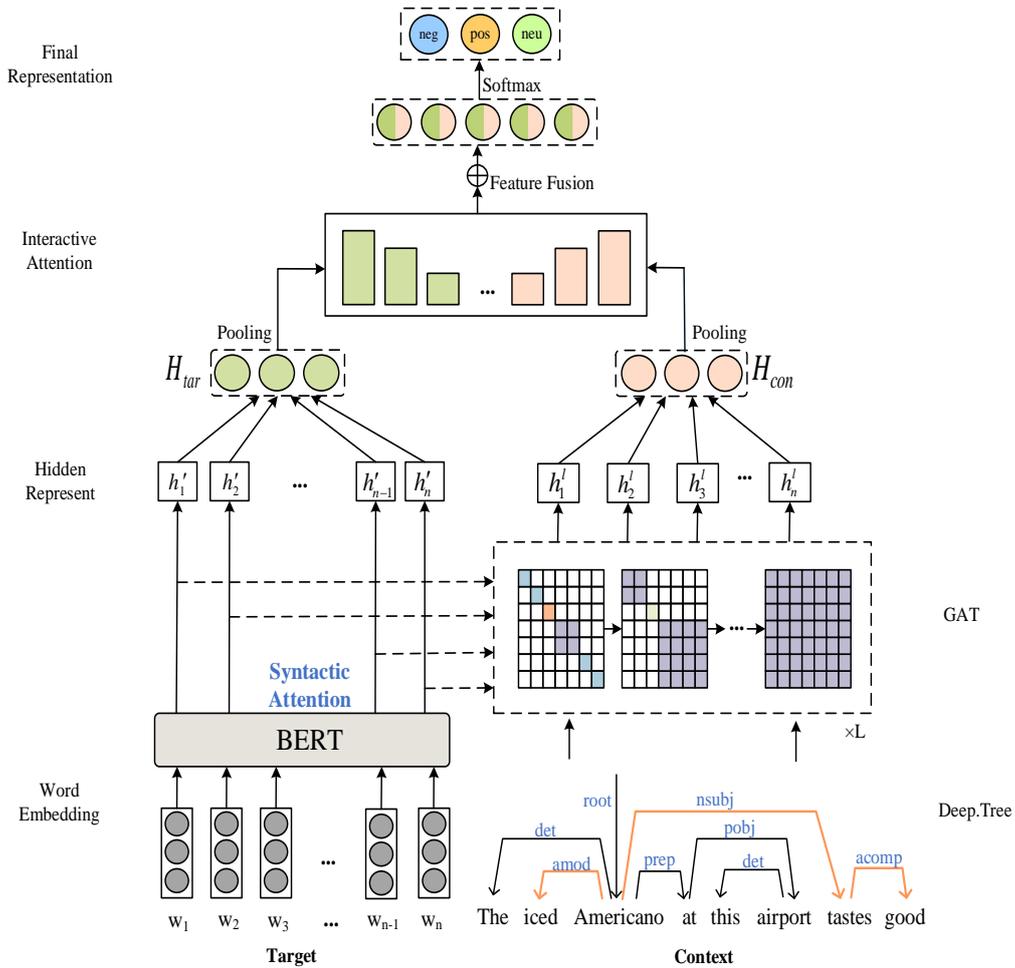


图 3.1 SIGAT 框架结构

### 3.2.1 输入层

在基于句法注意机制的目标方面提取部分，给定句子-目标方面对  $(S, T)$ ，其中用  $S = \{w_1, w_2, \dots, w_{i+m}, \dots, w_n\}$  表示一个句子，用  $T = \{w_i, w_{i+1}, \dots, w_{i+m-1}\}$  表示目标提及的序列， $n$  和  $m$  分别表示  $S$  和  $T$  的长度。为提高分类准确度，本文采用 BERT 模型生成词嵌入向量。首先将句子的序列重构为 “[CLS]+sentence+[SEP]+目标提及

+ [SEP]”，作为 BERT 的输入，在形式上将其表示为  $X = \{x_0, x_1, \dots, x_n, x_{n+1}, \dots, x_{n+2+m}\}$ ， $x_0$  和  $x_{n+1}$  分别是 “[CLS]” 和 “[SEP]” 的向量表示。在经过 BERT 模型后，输出与  $X$  长度相同的新序列  $H' = \{h'_0, h'_1, \dots, h'_n, h'_{n+1}, \dots, h'_{n+2+m}\}$  作为隐藏向量的表示，其中， $h'_0$  表示 BERT 的池化向量， $\{h'_1, h'_2, \dots, h'_n\}$  表示输入序列中的上下文词向量，而  $\{h'_{n+2}, \dots, h'_{n+2+m}\}$  表示输入序列中对应的目标提及向量，同时，我们将  $\{h'_1, h'_2, \dots, h'_n\}$  用于池化和特征融合的向量。

在基于图注意力网络（GAT）的上下文特征提取部分，首先利用 GAT 网络对词嵌入  $\{x_1, x_2, \dots, x_n\}$  进行特征转化，得到初始隐藏状态  $\{h_1^0, h_2^0, \dots, h_n^0\}$ ；随后通过不断迭代更新，得到第  $L-1$  层的隐藏状态  $\{h_1^{L-1}, h_2^{L-1}, \dots, h_n^{L-1}\}$ ，将它与相邻矩阵  $A$  作为第  $L$  层 GAT 的输入，得到第  $L$  层的特征表示  $H_{con} = \{h_1^L, h_2^L, \dots, h_n^L\}$  作为输出。

### 3.2.2 依存句法注意层

在之前的工作中，我们对语境词都是给予同等的重视，注意权重也仅仅是目标方面和语境词的关联度量。而近几年的研究中，我们发现并不是全部的上下文词都会对目标方面的情感极性产生影响。一般情况下，我们认为离目标方面较近或与目标方面有修饰关系的上下文词更重要，在判断情感极性时会产生更大的影响；而距离较远的修饰词会被抛弃，但这可能会导致我们无法准确的判断目标的情感极性。

为此，我们提出基于句法距离（SRD）的注意力机制。首先从语法解析器（Deep Biaffine Parse）生成的依存句法树中获取句法信息，并对句子的语法信息进行编码，有选择性地关注语法路径上靠近目标的上下文词，对这些词赋予较高的权重；同时为降低计算复杂度，引入高斯函数，避免因句法权重急剧下降导致信息的丢失。图 3.2 展示了一个例句的依存句法树，单词下的数字表示该词与目标方面的距离，其中 PD 表示基于位置距离的上下文词离目标方面的距离，SRD 表示基于句法结构的相对距离。从图中可以看出，目标方面“Americano”与修饰词“good”之间的位置距离（PD）是 5，属于远程；在经过句法解析后，“Americano”与“good”

之间的句法相对距离 (SRD) 是 2, 小于位置距离。这表明上下文词与目标方面之间的句法距离可以有效地强调在相对位置远而在句法上接近目标方面的词。当数量达到一定规模时, 上下文词的重要性与 SRD 呈负相关, 即 SRD 越小, 则该词的贡献就越大, 所获权重就越大。

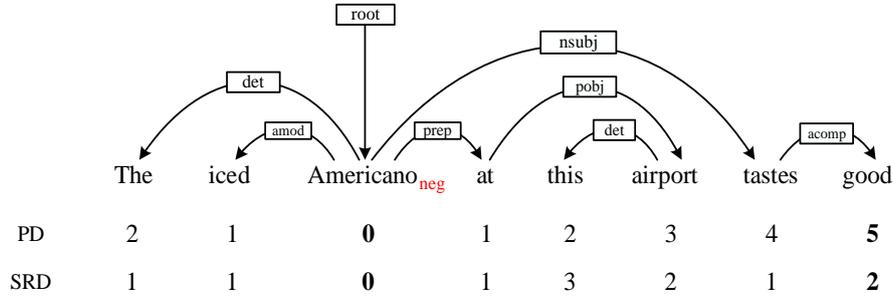


图 3.2 依存句法树结构示意图

为实现基于句法距离的权重计算, 我们将  $L$  定义为上下文词的位置, 代表该词到目标方面的句法距离, 为了使模型能有选择性的关注上下文词, 我们设置一个注意窗口并用  $ws$  表示 (在实验中, 设置  $ws = 4$ ), 并对窗口内的上下文词赋予不同的权重, 使得与目标方面句法距离近或有修饰关系的词获得更多的关注, 具体计算过程见公式 (3.1) 和公式 (3.2)。

$$p_i = g(d_i) = a * e^{-\frac{(d_i - \mu)^2}{2\sigma^2}} \quad (3.1)$$

$$d_i = \begin{cases} 0 & n = 0 \\ n & 1 \leq n < w \\ w & n \geq w \end{cases} \quad (3.2)$$

在公式 (3.1) 中, 为降低计算的复杂度, 本文引入高斯函数作为句法权重函数。其中,  $n$  表示句法相对距离,  $d_i$  表示加入高斯函数的句法注意机制的距离,  $a$  表示峰值 (实验中设  $a = 1$ ),  $\mu$  表示均值 (实验中设  $\mu = 0$ ),  $\sigma$  表示标准差 (实验中设  $\sigma = 5$ )。当  $1 \leq n < w$  时,  $d_i = n$  且句法权重随句法相对距离的增加而缓慢降低; 当  $n \geq w$  时, 令  $d_i = w$ , 且不考虑超过注意窗口  $w$  的上下文词。

句法权重计算只涉及句法距离, 并且随着句法相对距离的增加而减小。得到

每个上下文词的权重后, 连接得到一个句法权重向量  $P = \{p_1, p_2, \dots, p_n\}$ , 将  $P$  和  $H'$  相乘, 得到基于句法结构的权值表示  $H_{tar} = \{h_1, h_2, \dots, h_n\}$ , 过程见公式 (3.3)。

$$H_{tar} = P \odot H' \quad (3.3)$$

其中,  $\odot$  表示逐元素相乘。

### 3.2.3 图注意力网络层

为提取到包含句法信息的上下文的特征, 本文使用关系图注意力网络 (RGAT) 提取特征。与标准的 GAT 相比, RGAT 网络是根据句法依赖路径进行单词之间的信息交换, 可以更好的将句法结构信息聚合到目标方面, 从而更深层次的表示特征。在本节中, 我们首先介绍基线 GAT 模型, 然后介绍 RGAT 模型, 通过拓展 GAT 来编码带有标记边的图。

#### (1) 图注意力网络

通过构建依赖树将目标方面和上下文词连接, 利用图注意力网络 (GAT) 将情感信息聚合到目标方面。依赖树可以表示为具有  $n$  个节点的图结构  $G$ , 句子中每一个单词都用节点表示, 图  $G$  的边表示单词之间的依赖关系,  $N_i$  表示节点  $i$  的邻域节点。应用 GAT 和多头注意力机制聚合邻域节点, 并通过不断的迭代更新来表示每个节点。基于多头注意力的 GAT 聚合过程见公式 (3.4)。

$$h_{att_i}^{l+1} = \parallel_{z=1}^Z \sigma \left( \sum_{j \in N_i} \alpha_{ij}^{lz} W_z^l h_j^l \right) \quad (3.4)$$

其中,  $h_{att_i}^{l+1} \in R^d$  表示节点  $i$  在第  $l+1$  层的注意头表示,  $\parallel_{z=1}^Z$  表示向量拼接操作,  $\alpha_{ij}^{lz} \in R^1$  表示节点  $i$  与节点  $j$  在第  $l$  层的第  $z$  个注意头计算得到的归一化注意系数,  $W_z^l \in R^{d \times d/K}$  表示第  $l$  层第  $z$  个注意头的权值矩阵,  $d$  表示词特征的向量维数,  $Z$  表示注意头的个数,  $\sigma$  表示 sigmoid 激活函数。具体过程见公式 (3.5) 和公式 (3.6)。

$$\alpha_{ij}^{lz} = \frac{\exp \left( f \left( h_{att_i}^{l-1}, h_{att_j}^{l-1} \right) \right)}{\sum_{j' \in N(i)} \exp \left( f \left( h_{att_i}^{l-1}, h_{att_{j'}}^{l-1} \right) \right)} \quad (3.5)$$

$$f(h_{att_i}^{l-1}, h_{att_j}^{l-1}) = \left( \frac{(W_Q^{l_z} h_{att_i}^{l-1})^T (W_K^{l_z} h_{att_j}^{l-1})}{\sqrt{d/Z}} \right) \quad (3.6)$$

其中，若  $j \notin N(i)$ ，则  $\alpha_{ij}^{l_z} = 0$ 。  $f$  表示缩放点积注意函数<sup>[22]</sup>，  $W_Q^{l_z}, W_K^{l_z} \in R^{d \times d/Z}$  表示第  $l$  层第  $z$  个注意头的参数矩阵。

为了能够提高每个节点的信息容量，捕获更多的特征信息，本文在原有注意层的基础上加入了 PCT。当给定输出的隐藏状态  $h^l = \{h_1^l, h_2^l, \dots, h_n^l\}$  时，则 PCT 层的定义为见公式 (3.7)。

$$PCT(h^l) = \delta(h^l * W_{P_1} + b_{P_1}) * W_{P_2} + b_{P_2} \quad (3.7)$$

式中，  $\delta$  表示非线性激活函数，  $*$  表示卷积运算，  $W_{P_1} \in R^{d \times d}$  和  $W_{P_2} \in R^{d \times d}$  表示可训练的权重矩阵，  $b_{P_1} \in R^d$  和  $b_{P_2} \in R^d$  表示可训练的偏置向量。

### (2) 关系图注意力网络

GAT 在对邻域节点进行聚合时没有考虑依赖关系，这可能会丢失一些重要的依赖信息，因此我们使用关系图注意力网络，以依存句法树为辅助，将依赖关系与目标方面、上下文词连接，把基于句法结构的依赖关系融入到注意力的计算和聚合中，模型的结构如图 3.3 所示。

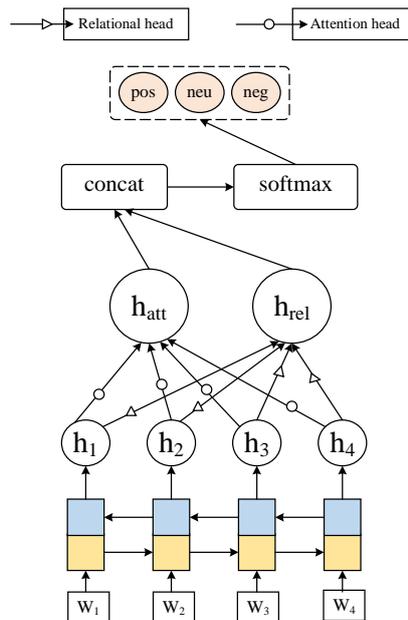


图 3.3 RGAT 结构图

首先，我们将单词  $w_i$  和  $w_j$  之间的依赖关系记为  $R_{ij}$ ，并将  $R_{ij}$  映射为向量表示  $r_{ij} \in R^{d_r}$ ，其中  $d_r$  表示关系嵌入的维数，利用拓展的 GAT 和多头注意力机制对每个依赖关系进行编码，聚合过程见公式 (3.8) 和公式 (3.9)。

$$h_{rel_i}^{l+1} = \parallel \sum_{m=1}^M \beta_{ij}^{lm} W_m^l h_{rel_j}^l \quad (3.8)$$

$$\beta_{ij}^{lm} = \text{soft max}(r_{ij} W^m + b^m) \quad (3.9)$$

其中， $h_{rel_i}^{l+1} \in R^d$  表示节点  $i$  在第  $l+1$  层的关系头表示， $\beta_{ij}^{lm} \in R^1$  表示节点  $i$  和节点  $j$  在第  $l$  层的第  $m$  个关系头中计算得到的归一化关系系数， $W_m^l \in R^{d \times d/M}$  表示第  $l$  层第  $m$  个关系头的权值矩阵， $W^m \in R^{d \times L}$ ， $b^m \in R^d$ 。最终每个节点的表示如公式 (3.10) (3.11) 所示。

$$x_i^{l+1} = h_{att_i}^{l+1} \parallel h_{rel_i}^{l+1} \quad (3.10)$$

$$h_i^{l+1} = \text{relu}(W_{l+1} x_i^{l+1} + b_{l+1}) \quad (3.11)$$

在公式 (3.10) 中， $x_i^{l+1}$  表示节点  $i$  在第  $l+1$  层的向量表示， $h_i^{l+1}$  表示节点  $i$  的最终表示， $\parallel$  表示注意力与关系的拼接操作；在公式 (3.11) 中， $W_{l+1}$ 、 $b_{l+1}$  分别表示节点  $i$  在第  $l+1$  层的可训练参数矩阵和参数向量。

为了学习深度特征，本文采用了堆叠多层 RGAT 的方法。

### 3.2.4 交互学习层

单独计算注意力而忽略上下文信息与目标方面之间的联系，会丢失一些重要的特征信息，因此，为解决上述问题，本文加入交互注意力机制使两个不同的特征向量建立联系，同时考虑上下文特征和目标方面特征，二者相互指导学习，以获得更丰富的特征表示。图 3.4 为交互学习的示意图，该机制可以同时计算目标方面的注意力和上下文的注意力，并且目标方面的特征和上下文特征通过相互指导、相互学习来更新特征向量，丰富特征向量的表示。

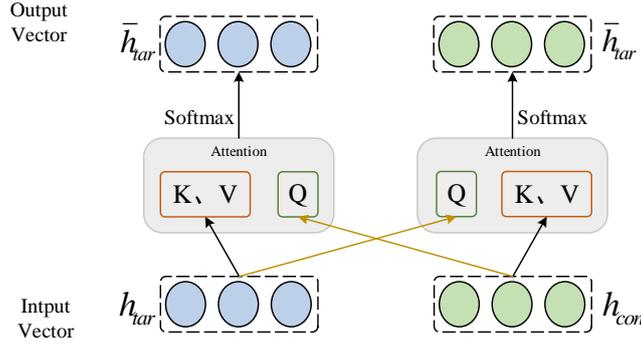


图 3.4 交互注意力结构

在进行交互学习之前,我们首先对基于句法注意机制的目标方面特征向量和基于图注意力网络(GAT)的上下文特征向量应用平均池化函数,以获得两个全局向量表示 $h_{tar}$ 和 $h_{con}$ ,其表达式见公式(3.12)。

$$\begin{aligned} h_{tar} &= pool(h_1, h_2, \dots, h_{n-1}, h_n) \\ h_{con} &= pool(h'_1, h'_2, \dots, h'_{n-1}, h'_n) \end{aligned} \quad (3.12)$$

在得到 $h_{con}$ 和 $h_{tar}$ 后,将其作为交互注意力层的输入,并映射到查询( $Q_{tar}, Q_{con}$ )、键( $K_{tar}, K_{con}$ )和值( $V_{tar}, V_{con}$ )的矩阵中。在计算基于句法注意机制的目标方面注意力时,以 $Q_{con}$ 为查询矩阵, $K_{tar}$ 和 $V_{tar}$ 为键、值矩阵,具体计算过程如公式(3.13)和公式(3.14)所示。

$$C_{tar} = soft \max \left( \frac{Q_{con} K_{tar}^T}{\sqrt{d_k}} \right) V_{tar} \quad (3.13)$$

$$\bar{h}_{tar} = LN(h_{tar} + C_{tar}) \quad (3.14)$$

其中, $LN$ 表示层标准化函数。

在计算基于图注意力网络的上下文注意力时,以 $Q_{tar}$ 为查询矩阵, $K_{con}$ 和 $V_{con}$ 为键、值矩阵,以此来计算基于该部分的注意力,具体过程见公式(3.15)和公式(3.16)。

$$C_{con} = soft \max \left( \frac{Q_{tar} K_{con}^T}{\sqrt{d_k}} \right) V_{con} \quad (3.15)$$

$$\bar{h}_{con} = LN(h_{con} + C_{con}) \quad (3.16)$$

最后, 得到  $\bar{h}_{tar}$  和  $\bar{h}_{con}$  作为交互注意层的输出, 通过利用对方的信息来强化自身的特征表达, 以便深度学习更多的特征信息。

### 3.2.5 情感输出层

在经过交互注意力层的特征学习后, 我们获得包含丰富句法信息的目标特征  $\bar{h}_{tar}$  和表达能力强的上下文特征  $\bar{h}_{con}$ , 为方便后续分类输出, 我们先将二者的部分特征进行融合得到特征向量  $h_f$ , 特征融合过程见公式 (3.17)。

$$h_f = (\bar{h}_{tar} \oplus \bar{h}_{con})W_f + b_f \quad (3.17)$$

在公式 (3.17) 中,  $W_f$  表示可训练的权重矩阵,  $b_f$  表示偏置向量,  $\oplus$  代表拼接操作。

再得到融合特征向量  $h_f$  后, 将其输入至全连接层, 并分类输出该目标方面的情感分类  $c$ , 具体过程见公式 (3.18)。

$$P(y = c) = \frac{\exp(W_h h_f + b)_c}{\sum_{c' \in C} \exp(W_h h_f + b)_{c'}} \quad (3.18)$$

式中,  $w$  和  $b$  都是可调整的参数,  $C$  为情感类的集合。

最后, 使用标准交叉熵损失作为目标函数, 见公式 (3.19)。

$$\ell = -\sum_{i=1}^N \sum_{c \in C} I \ll (y = c) \log(P(y = c)) + \lambda \|\Theta\|^2 \quad (3.19)$$

其中,  $I$  表示指标函数,  $N$  表示训练集的数量,  $\lambda$  表示正则化参数,  $\Theta$  表示模型中的所有参数集。

## 3.3 本章小结

本章是对提出的 SIGAT 模型以及构建部分进行详细阐述, 主要包括两个部分: 一是基于句法注意机制的目标方面提取, 该模块是将依存句法信息与注意力机制相结合, 有选择性的关注语法路径上接近目标方面的上下文词, 并加入高斯函数降低计算复杂度, 有效避免意见词因位置距离远而被遗漏; 二是基于图注意

力网络（GAT）的上下文特征提取，该方法有效地利用单词之间的依赖关系丰富文本的特征表达能力；最后，通过交互学习层加强二者的联系，互用对方的信息来加强自身的表达能力，经融合后分类输出情感极性。

## 4 实验设计与结果分析

在本节中,我们首先介绍评价指标、用于评估的数据集以及实验的相关环境,然后我们介绍模型训练的相关设置,与不同的模型进行对比实验和消融实验,最后,对结果进行分析并选取 Restaurant 和 Laptop 中 2 个具有代表性的实例进行分析讨论。

### 4.1 评价指标

为评估 SIGAT 模型在细粒度情感分析任务中的有效性,本文使用准确率(Acc)和宏平均 F1 值作为模型的评价指标。其中, Acc 值是描述所有预测正确的数量占总数量的比例, F1 值是反映全局性能的指标,是精确率(Precision)和召回率(Recall)调和平均数。如公式(4.1)(4.2)所示,是 Acc 值和 F1 值具体计算过程。

$$Acc = \frac{TP + TN}{N} \quad (4.1)$$

$$\begin{cases} F1 = \frac{2PR}{P + R} \\ P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \end{cases} \quad (4.2)$$

在公式(4.1)中,  $TP$  表示真正例样本数,  $TN$  表示真反例样本数,  $N$  代表样本总数; 在公式(4.2)中,  $FN$  表示假反例样本数,  $FP$  表示假正例样本数。

## 4.2 实验数据及实验环境介绍

### 4.2.1 实验数据

为了评估所提出的 SIGAT 模型, 本文选择在 3 个公开数据集上进行实验,

主要包括 SemEval-2014 Task4<sup>[1]</sup>数据中 Restaurant、Laptop, 以及 ACL-14 Twitter<sup>[2]</sup>数据集。其中 Restaurant 数据是有关餐厅的食物、味道以及服务的评价, Laptop 数据是有关笔记本电脑功能、颜色等各方面的评价, 而 Twitter 是一个人工标注的数据集。这些数据集被划分为两种状态: 训练集 (Train) 和测试集 (Test), 其中包括 3 类情感标签: 积极 (Positive)、消极 (Negative) 和中性 (Neutral), 具体统计情况如表 4.1 所示。

表 4.1 三个数据集的基本统计情况

数据集	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	727	807	196	637	196
Laptop	976	337	851	128	455	167
Twitter	1507	172	1528	169	3061	336

### 4.2.1 实验环境介绍

本文实验环境相关配置如表 4.2 所示。

表 4.2 实验环境配置表

实验环境配置表	配置信息
操作系统 CPU	Ubuntu20.04
显卡	NVIDIA GeForce 3080ti
编程语言	Python3.8
深度学习框架	pytorch
开发环境	pycharm

<sup>[1]</sup> 数据来源: GitHub - Noahs-ARK/semEval-2014

<sup>[2]</sup> 数据来源: RGAT-ABSA/dataset/Biaffine/glove/Tweets at master · goodbai-nlp/RGAT-ABSA · GitHub

## 4.3 模型训练相关设置

### 4.3.1 实验参数设置

在实验中，对于基于 BERT 的编码器，我们采用预训练的 BERT 进行微调。为方便交互层特征向量的学习，我们将词嵌入维度和关系嵌入维度均设置为 768，隐藏层的维数设置为 300，模型的丢弃率设为 0.1，正则化系数设置为 $1 \times 10^{-5}$ ，学习率设置为 $2 \times 10^{-5}$ 。对于基于图注意力网络（GAT）的编码器中，我们将注意力头  $K$  和关系头  $M$  均设为 5，丢弃率设置为 0.7， $L_2$  正则化系数也设置为 $1 \times 10^{-5}$ ，每次输入数据的 Batch-size 为 16，并使用 Deep Biaffine Parse<sup>[1]</sup>来解析依赖关系，生成语法依赖树。我们的模型使用默认配置的 Adam 优化器进行训练。

### 4.3.2 对比实验设置

为体现 SIGAT 模型性能，我们将与一些基于方面级情感分析的基线模型进行比较，主要包括以下模型：

(1) LSTM+SynAtt<sup>[49]</sup>：该模型将句法信息融入注意力机制，根据上下文单词和目标词之间的语义关联程度，对其赋予不同的注意权重，提高了注意力的有效性。

(2) IAN<sup>[36]</sup>：该模型认为在句子中，目标方面的信息和上下文词的信息均会对最终文本表示产生影响，因此提出了通过交互注意力机制获得目标方面与上下文的向量表示，匹配程度进一步提高，最后通过拼接操作输出情感结果。

(3) RAM<sup>[42]</sup>：该模型同时考虑了方面短语的语义依赖和不同方面之间的情感交互关系，通过多层注意力机制获取目标方面相关的情感信息，并根据位置距离对上下文词进行权重的分配。

(4) CGAT<sup>[50]</sup>：该模型提出了基于句法距离的注意力机制，根据句法相对距离对上下文词赋予权重，并巧妙地引入了高斯函数作为句法的权重函数，降低了

---

<sup>[1]</sup> 解析器：GitHub - yzhangcs/parser: State-of-the-art syntactic/semantic parsers with pretrained models for many languages.

计算的复杂度。

(5) ASGCN<sup>[54]</sup>: 该模型是将图卷积网络 (GCN) 设立在句子的以来书上, 不仅考虑了目标方面的表达, 也考虑了句子语义信息的表达, 并加入注意力机制对特征进行优化和重新分配权重, 最后再进行分类。

(6) CDT<sup>[53]</sup>: 该模型首先利用 Bi-LSTM 来学习句子特征, 然后在依存句法树上建立图卷积网络 (GCN), 将上下文信息和依赖关系从意见词传播到方面词, 提高了分类的准确率。

(7) R-GAT+BERT<sup>[59]</sup>: 该模型是在原始的依存句法树上进行剪枝, 得到一个有关目标方面的依存树, 扎根于目标方面, 并利用两个图注意力网络 (GAT) 对新的依赖树编码, 最后分类输出结果。

(8) DualGCN<sup>[56]</sup>: 该模型是建立了两个图卷积网络 (GCN), 一个含有句法信息的 SynGCN, 另一个是含有语义关系的 SemGCN, 同时考虑单词之间的句法结构和语义关系, 解决了依存关系解析结果的不准确性, 提高了分类效果。

## 4.4 实验结果分析

为了证实模型的有效性, 本节将对实验结果进行分析: 1.通过与基线模型的 Acc 和 F1 值对比, 证明 SIGAT 的有效性; 2.为证明 SIGAT 模型中各模块 (句法注意机制、交互注意层、关系图注意力网络) 重要性, 设计三组消融实验; 3.分析影响模型分类准确率的因素, 其中包括: 注意窗口大小  $w_s$ 、不同峰值、不同关系类型和交互注意层深度; 4.选取数据集中两例具有代表性的句子进行实例分析, 并对句法权重进行可视化。

### 4.4.1 主要结果

各模型的 Acc 值和 F1 值对比结果如表 4.3 所示, SIGAT 模型在 Restaurant、Laptop 和 Twitter 数据集上都取得了较高的结果, 表格中的数字均为百分比值。实验结果表明, 在建模过程中使用句法信息和注意力机制能够增强文本特征的表达能力, 从而提高分类准确率, 进一步证实了模型的有效性。

表 4.3 各对比模型的 Acc 值和 F1 值

Models	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
LSTM+SynAtt	69.84	65.61	68.35	64.37	65.10	61.90
IAN	70.21	66.53	70.65	66.12	67.88	63.53
RAM	71.96	68.01	72.13	68.35	69.21	65.35
ASGCN	73.61	70.88	74.31	70.05	70.09	66.40
CDT	75.43	71.26	75.62	72.89	71.05	67.66
R-GAT+BERT	77.32	73.90	77.49	74.01	73.66	69.02
DualGCN	77.03	73.17	79.27	75.32	74.46	70.67
CGAT	78.81	<b>75.30</b>	80.35	76.50	<b>76.06</b>	71.50
<b>SIGAT*</b>	<b>78.35</b>	74.21	<b>81.69</b>	<b>77.31</b>	75.60	<b>71.78</b>

从表 4.3 中可知，在 3 个领域的数据集中，SIGAT 模型的准确率（ACC）和 F1 值分别达到 78.35%、81.69%、75.60%和 74.21%、77.31%、71.78%，结果优于其他基线模型。其中，与 LSTM+SynAtt、IAN、RAM、ASGCN 和 CDT 模型相比，SIGAT 模型的 ACC 和 F1 值均高于这些对比模型，主要是因为 SIGAT 是基于 BERT 编码的，与 LSTM 相比，它可以根据上下文环境自适应建模，更加有效的提取语义信息。

在基于注意力机制构建的模型中，LSTM+SynAtt、IAN 和 RAM 模型通过注意力机制更好的建模了目标方面的特征，但对文本特征挖掘不深。而 CGAT 模型的 ACC 和 F1 值均高于 LSTM+SynAtt、IAN、RAM 模型，一部分是因为该模型采用了基于 SRD 的句法注意机制，依据句法相对距离对上下文词赋予不同的权重，另一部分是利用两个图注意力网络（GAT）提取句法结构信息，增强了文本特征的表达能力。虽然 CGAT 模型有丰富的句法信息和语义信息，但它忽略了上下文特征和目标方面特征之间的重要联系，当二者不进行交互学习时，就变成了两个单独计算的注意力机制，无法增强特征的信息容纳能力，会造成一些信息特征的丢失。本文提出的 SIGAT 模型在增加了交互学习层后，Acc 值和 F1 值优于 CGAT 模型，也进一步证实了加入交互注意力机制可以有效增强特征表达能力，

进而提高分类准确率。

在基于句法信息构建的模型中，大部分都是将句法信息与神经网络相结合。其中，R-GAT+BERT 模型和 DualGCN 模型在 3 个数据集上的 Acc 值均高于 ASGCN 和 CDT 模型，并且 DualGCN 模型的准确率优于 R-GAT+BERT。从实验结果看出，DualGCN 模型的 Acc 值仅比我们提出的 SIGAT 模型低 1.32%，差距很小。产生这种结果的原因是 DualGCN 模型不仅考虑了语法结构的互补性，还考虑了语义的相关性，为减少单词之间的重叠、提高分类效果，该模型使用正交和差分正则化器来获取语义相关性。需要强调的是，该模型在计算注意力矩阵时，是依据句子的表征而非句法距离，这可能导致信息提取不充分，甚至会出现丢失位置距离远而句法距离近的上下文词。SIGAT 模型可以有效地克服信息提取不充分和丢失信息的问题，并且模型的 Acc 值和 F1 值高于上述模型，证实了 SIGAT 的有效性。

#### 4.4.2 消融实验

为了验证不同组成分在 SIGAT 模型中的有效性，在 Restaurant、Laptop、和 Twitter 数据集上设计三组消融实验，每组的具体描述如下：

SIGAT-S：消除了句法注意机制。

SIGAT-Att：消除了交互注意力机制。

SIGAT-R：消除了关系图注意力网络。

如表 4.4 所示，消除模型中的任意部分都会导致 Acc 值和 F1 值的降低。

表 4.4 消融实验

Models	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
SIGAT-S	76.27	71.41	79.37	75.25	73.50	69.75
SIGAT-R	77.31	72.77	80.20	76.10	74.11	70.88
SIGAT-Att	77.98	73.70	80.87	76.90	74.47	70.02
SIGAT*	<b>78.35</b>	<b>74.21</b>	<b>81.69</b>	<b>77.31</b>	<b>75.60</b>	<b>71.78</b>

由实验结果可知，特别是在消除了句法注意机制（SIGAT-S）后，SIGAT 模型在 3 个数据集上的 Acc 值明显下降，分别降低了 2.08%、2.32%和 2.10%，这足以表明基于 SRD 的句法注意机制可以更好的建模上下文词和目标方面的关系，通过分配不同的注意权重，突出靠近目标方面的词，从而提高目标方面特征的表达能力。

当消除交互注意力层（SIGAT-Att）时，模型的准确率和 F1 值有所降低。若只是单独计算上下文特征的注意力和目标方面特征的注意力，会忽略二者之间的重要联系，导致部分信息的丢失，可能会影响分类的效果，因此交互注意力机制的是不可或缺的组成部分。

当消除关系图注意力网络（SIGAT-R）时，与 SIGAT 模型相比，虽然准确率和 F1 值略有降低，但不会对模型产生巨大的影响。从实验结果可以看出，增加关系图注意力网络可以丰富单词之间的依赖关系，在考虑句法相对距离的同时也能考虑额外的依赖关系，有助于模型性能的提升。

#### 4.4.3 不同因素对模型性能的影响

为研究 SIGAT 模型的影响因素，本文从四个方面对模型进行影响因素分析。在目标特征的提取方面，我们选取注意窗口大小  $ws$  和峰值  $a$  为分析因素；在上下文特征的提取方面，我们选取不同依赖关系和交互注意层深度为分析因素，分别设计了四个系列的对比实验。

##### （1）注意窗口大小 $ws$ 的影响分析

为观察不同注意窗口大小对 SIGAT 模型的影响，本文设计了 3 组对比实验，分别选取 Restaurant、Laptop 和 Twitter 数据集，并将注意窗口大小分别设置为 3、4、5，结果如图 4.1 所示。

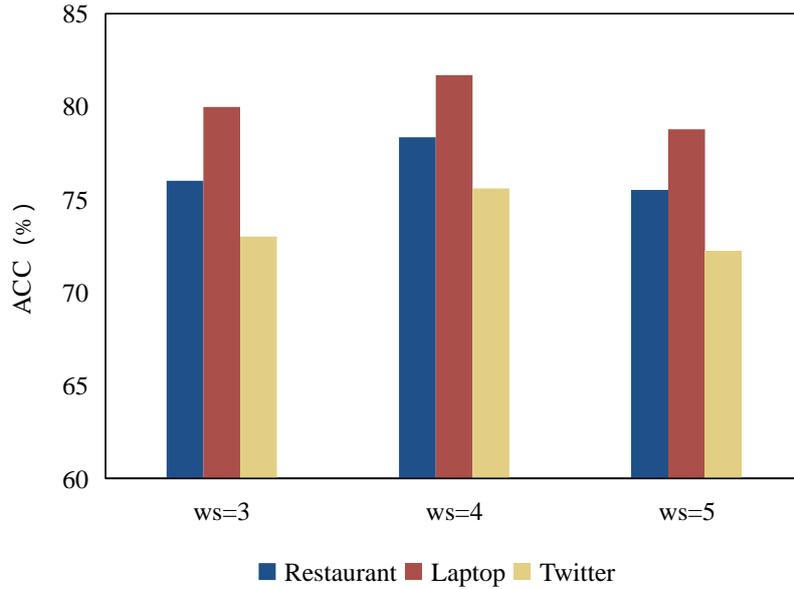


图 4.1 不同注意窗口大小  $ws$  的结果

从上图可知，在三个公开领域数据集中，当  $ws = 4$  时，模型预测的准确性最高，因此，我们将注意窗口大小  $ws$  设为 4。

(2) 不同峰值的影响分析

为观察高斯函数中不同峰值  $a$  对模型性能的影响，我们设计了 10 组对比实验，在公式 (3.1) 中，峰值  $a$  从 0.6 增加到 1.5，实验结果如图 4.2 所示。

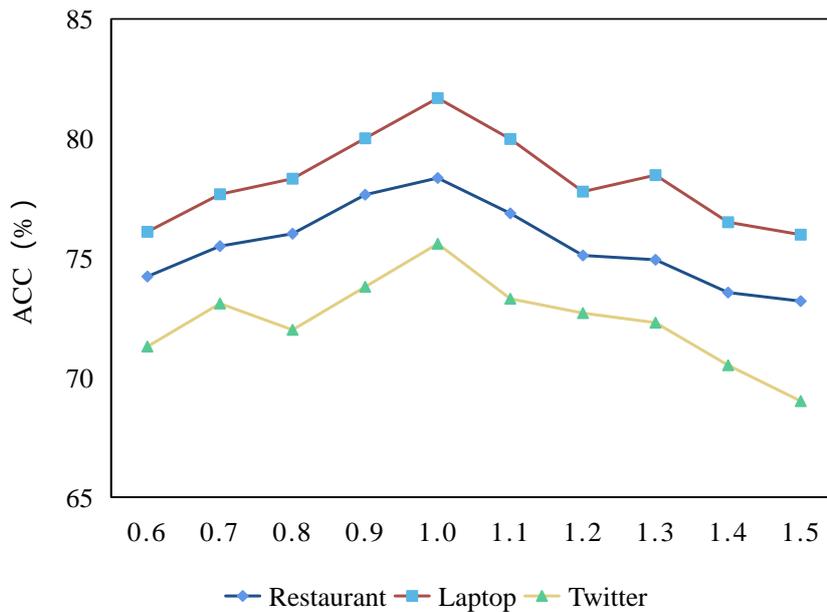


图 4.2 不同峰值的结果

由图可知，在 Restaurant、Laptop 和 Twitter 数据上，模型的 ACC 值整体趋势是先升高后降低，当峰值  $a = 1$  时，模型的分类准确率最高，因此设置峰值  $a$  为 1。

### (3) 交互注意层深度的影响分析

除了注意窗口的大小和不同关系类型会影响模型性能，交互注意层的叠加层数对 SIGAT 模型也会产生影响。为分析叠加层数对模型的影响，本文对比了叠加 1 层到叠加 6 层的实验结果，如图 4.3 所示。

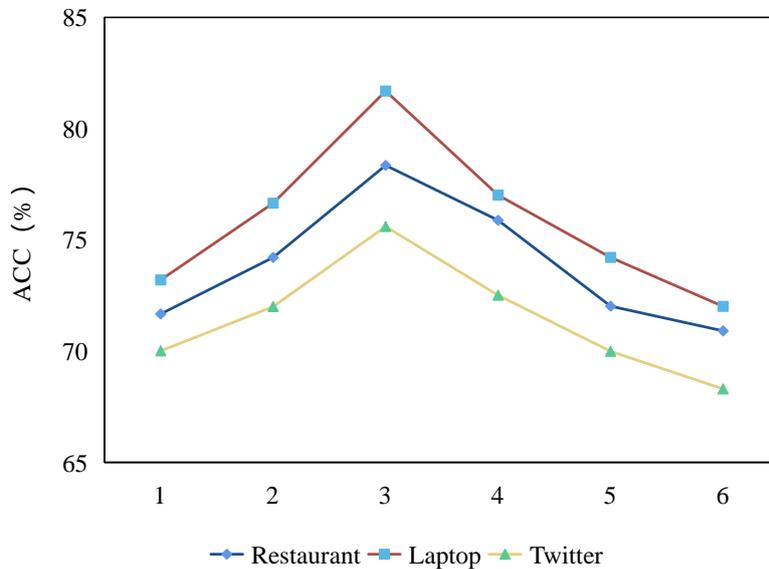


图 4.3 不同叠加层数的结果

我们从图中可看出，模型的 ACC 值是先上升后下降，当叠加层数为 3 时，在三个数据集上都取得了较高的准确率，分别达到 78.35%、81.69%和 75.60%，且二者可以利用对方的信息来增强自己的特征表示，从而提高分类效果，因此在实验中，本文将交互注意力层数设为 3 层。当叠加层数小于 3 层时，目标方面特征和上下文特征无法充分的相互学习，导致分类准确率较低；当叠加层数超过 3 层甚至更多时，模型准确率也会下降，这是因为梯度消失或梯度爆炸。

### (3) 不同关系类型的影响分析

为研究不同关系类型对模型提取上下文特征的影响，本文选取了一些出现频率较高的关系类型，主要包括：核心依存关系 {nsubj, ccomp, dobj, xcomp, acomp}、

名义依存关系 {nmod, amod, advmod, det, case} 以及其他依存关系 {conj, cc, punc}, 并在 Restaurant 和 Twitter 数据集上进行实验。如图 4.4 所示, 为 SIGAT 模型仅去除单个关系类型的模型准确率。

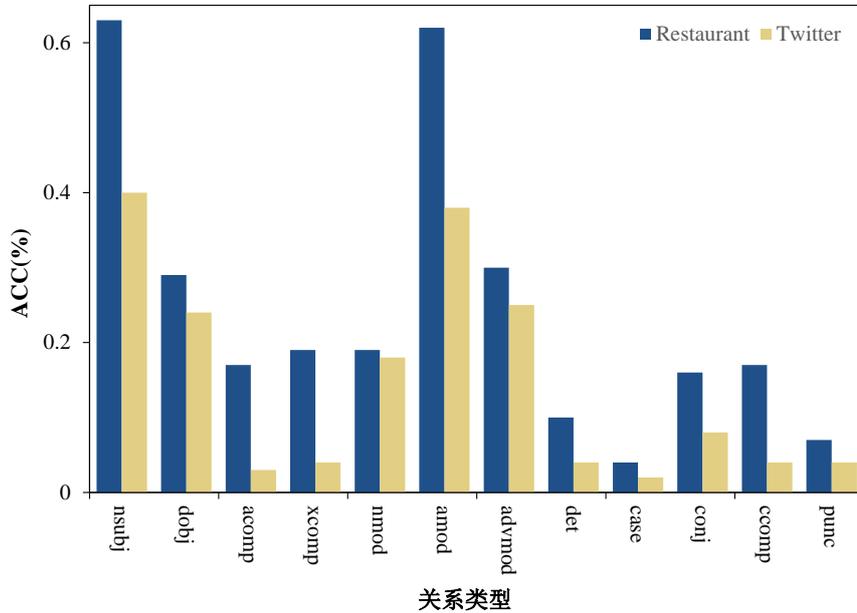


图 4.4 去除不同关系类型导致模型 ACC 值下降

由图可知, 去除 nsubj 和 amod 关系时, SIGAT 模型的准确率下降最快。当去除 nsubj 关系时, 其 ACC 值在两个数据集上分别下降了 0.67% 和 0.39%, 作为句法结构的主语, 一般与目标方面相连, 因此可以通过该依赖关系找到需要判定的目标。当去除 amod 关系时, 其 ACC 值分别下降了 0.65% 和 0.40%, amod 作为形容词的修饰语, 一般与目标方面的情感极性有关, 所以去除 amod 关系会对模型产生一定的影响。从数据可以看出, 这两种关系类型都携带了重要的分类信息, 是不可缺少的关系类型。对于关系 det、case 和 punc 等, 虽说使用频率较高, 但是对模型分类的准确率影响不大, 其 ACC 值分别下降了 0.12%、0.05% 和 0.09%, 从侧面证明了关系类型出现的频率与模型的精确率没有很强的相关性。

#### 4.4.4 实例分析

为证实本文所提出的模型在判断多方面情感级极性任务中的能力, 明确句法注意机制在模型中的作用。我们选取 Restaurant 和 Laptop 数据集中两个具有代

表性的实例进行研究，并绘制句法权重可视图，其中图 4.5 表示 Restaurant 中的实例，图 4.6 表示 Laptop 中的实例，二者均包含两个不同的目标方面，图中颜色越深则表示权重越高，反之权重越小。

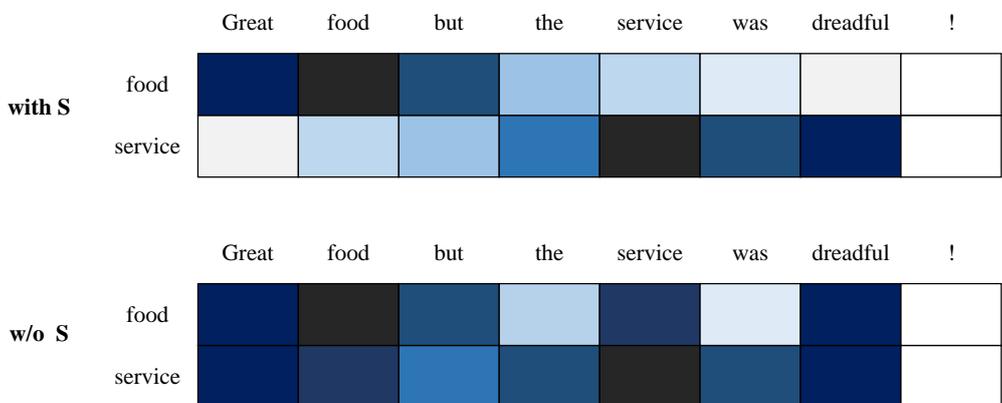


图 4.5 不同情感极性句法权重可视图

如图 4.5 所示，在例句“Great food but the service was dreadful!”中，包含“food”和“service”两个方面，并且二者的情感极性不同。从可视图看出，对于使用了句法注意机制（with S）的句子来说，可以更加准确的判断出目标方面的情感极性，“great”是方面词“food”的意见词，对其赋予较高的注意权重，颜色较深；而意见词“dreadful”、介词和标点符号等，在句法距离上离“food”较远，因此赋予较小的注意权重，颜色偏浅。通过二者的可视图对比，面对包含不同方面不同情感极性的句子，句法注意机制依旧可以捕获不同的情感极性，能够成功的区分不同意见词，提高分类效果。

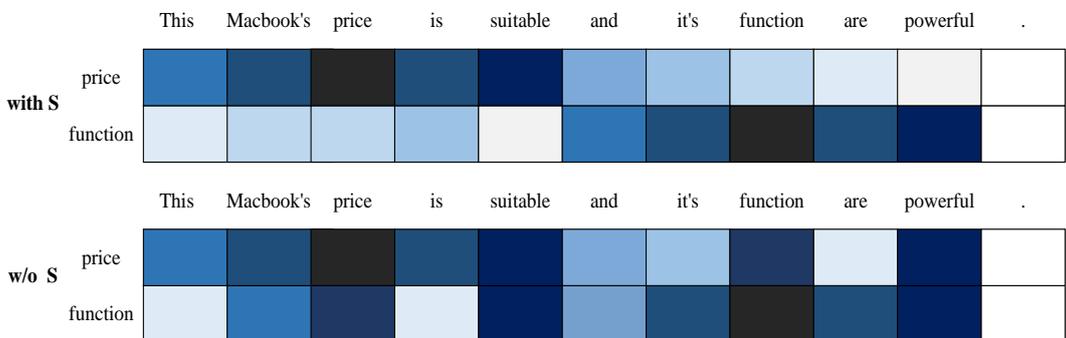


图 4.6 相同情感极性句法权重可视图

如图 4.6 所示，句子“This Macbook’s price is suitable and it’s function are

powerful.”中，包含“price”和“function”两个方面，与图 4.5 不同的是，两个方面的情感极性相同，都是积极。从可视图看出，对于不使用句法注意机制（w/o S）的句子，因情感极性相同，所以对情感词都赋予了一定的权重，无法准确将目标方面和意见词匹配；而对于使用了句法注意机制（with S）的句子，在情感极性相同情况下，依旧能够准确匹配目标方面与意见词。

从上述两个实例看出，句法注意机制能够成功的区分不同方面的意见词，在处理较为复杂的句子时更具有优势。

## 4.5 本章小结

本章的主要内容是在 Restaurant、Laptop 和 Twitter 数据集上验证 SIGAT 模型的性能。首先我们介绍了评价模型性能的指标，分别是准确率(Acc)和 F1 值；随后，我们对实验选取的数据集和相关环境配置进行了说明，为开展实验做好准备。在构建完 SIGAT 模型后，本文指出了模型训练时的参数以及各个对比的模型结构，从实验结果可知，本文所提出的 SIGAT 模型要优于其他对比模型。为进一步证明模型中不同模块的有效性，我们设计了三组消融实验，同时也从不同角度分析了影响模型性能的因素；最后，对具有代表性的实例进行分析，证明 SIGAT 模型的可靠性。

## 5 总结与展望

### 5.1 总结

随着互联网技术的不断进步，互联网已经渗透到人们的日常生活中，成为人们表达意见的重要渠道。如果能从数以千计的意见中提取并分析出具有实际价值的信息，不仅可以帮助企业根据评论了解用户对某一产品的意见从而进行改进，也可以让政府根据公众对某一公共事件或政策的意见来调整政策。面对如此精细化的需求，文本级别的情感分析是无法满足的。因此，本文以依存树上的句法信息和注意力机制为基础，侧重研究方面级别的情感分析。

为更好的区分不同上下文词对目标方面的贡献，以及加强目标方面特征与上下文特征的联系，从而减少情感信息的丢失，本文提出了基于依存句法树的交互式方面级情感分析模型，文中简称 SIGAT 模型。SIGAT 模型主要包括两大部分，第一部分是基于句法注意机制的目标方面特征提取，第二部分是基于图注意力网络的上下文特征提取；接着通过交互学习的方式，增强特征的表达能力；最后，将两个特征向量经融合后输入全连接层输出情感分类。在第一部分中，我们的重点是根据依存句法信息构建基于句法的注意模型，并依据句法信息定义上下文词的位置，用注意力机制选择性的关注一个窗口 ( $ws$ )，对于窗口内的上下文词应用不同的权重，以便接近目标的词获得更多的关注；同时为避免因权重急剧下降导致丢失情感信息，本文在原有基础上引入高斯函数作为句法权重函数，降低了计算复杂度。当一个句子中有多个目标和多个意见词时，基于句法注意机制的模型可以让我们更好地区分不同方面的情感极性。第二部分是通过图注意力网络 (GAT) 得到的上下文特征，不仅考虑了句法信息，也考虑了单词之间的依赖关系，丰富了文本特征的表达。最后，为了学习更加丰富的特征信息、加强二者之间的联系，我们设计了交互学习层，将基于句法注意机制的目标方面特征与基于图注意力网络 (GAT) 的上下文特征进行交互式学习，利用对方的信息增强自身的特征表达能力，经融合后通过全连接层分类输出情感极性。

本文为验证所提出的 SIGAT 模型的性能，在 3 个领域的公开数据集上进行实验，其准确率 (Acc) 和 F1 值优于对比模型，证明了 SIGAT 模型的有效性。

通过消融实验可知,模型中依存句法信息和注意力机制对提升分类准确率有帮助,并通过对其他影响因素的分析,找到模型适宜的参数。最后,通过对具有代表性的实例进行分析,发现带有句法注意机制的方面级情感分析模型性能更好。

## 5.2 展望

虽然,本文提出的 SIGAT 模型在 SemEval-2014 中的餐厅评论(Restaurant)、笔记本电脑评论(Laptop)和 Twitter 评论数据集上的准确率和 F1 值较其他模型有所提高,证明了 SIGAT 模型的性能。但是在面对句子结构复杂、包含多个方面且含有大量噪声的文本数据时, SIGAT 模型预测的准确度提升有限。在未来,我们将继续探索如何更好的确定目标相关的意见表达,并且在交互学习的过程中降低运算复杂度,增强句子语法知识的整合,进一步提升模型的预测性能。

## 参考文献

- [1] 中国互联网信息中心 CNNIC.第 50 次《中国互联网络发展状况统计报告》[R].  
北京: 中国互联网信息中心 CNNIC, 2022 年 9 月.
- [2] Liu B. Sentiment Analysis and Opinion Mining[J]. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers, 2012, 5(1): 1–167.
- [3] Dave K, Lawrence S, Pennock D M. Mining the peanut gallery: opinion extraction and semantic classification of product reviews[C]//Proceedings of the 12th international conference on World Wide Web, New York, NY, USA: Association for Computing Machinery, 2003: 519–528.
- [4] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273–297.
- [5] Baccianella S, Esuli A, Sebastiani F. SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.[C]//International Conference on Language Resources and Evaluation, 2010, 10(3):2200-2204.
- [6] Ku L-W, Chen H-H. Mining opinions from the Web: Beyond relevance retrieval[J]. Journal of the American Society for Information Science and Technology, 2007, 58(12): 1838–1850.
- [7] Chen W-T, Lin S-C, Huang S-L, et al. E-HowNet and automatic construction of a lexical ontology[C]//Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations, USA: Association for Computational Linguistics, 2010: 45–48.
- [8] 徐琳宏, 林鸿飞, 潘宇.情感词汇本体的构造[J].情报学报, 2008, 27(2): 180–185.
- [9] Hu M, Liu B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA: Association for Computing Machinery, 2004: 168–177.
- [10] 赵妍妍, 秦兵, 石秋慧.大规模情感词典的构建及其在情感分类中的应用[J].中文信息学报, 2017, 31(2): 187–193.

- [11] Tran T K, Phan T T. A hybrid approach for building a Vietnamese sentiment dictionary[J]. *Journal of Intelligent & Fuzzy Systems*, IOS Press, 2018, 35(1): 967–978.
- [12] 元海霞.基于 Word2Vec 和 HowNet 的情感词典构建方法[J].*现代计算机(专业版)*, 2018(4): 9–13.
- [13] 杨立月, 王移芝.微博情感分析的情感词典构造及分析方法研究[J].*计算机技术与发展*, 2019, 29(2): 13–18.
- [14] 杜振中, 万岩.融合情感词典和语义规则的微博评论细粒度情感分析[J].*情报探索*, 2020, 1(11): 34-41.
- [15] 颜明阳, 闫国梁, 李明兰.领域特定情感词典扩展方法在情感分类中的应用[J].*计算机应用与软件*, 2022, 39(6): 176–182.
- [16] Pang B, Lee L, Vaithyanathan S. Thumbs up sentiment classification using machine learning techniques[C]//*Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, USA: Association for Computational Linguistics, 2002: 79–86.
- [17] Wei Z, Sui-xi K, Yan-chun Z, et al. Sentiment classification and computing for online reviews by a hybrid SVM and LSA based approach[J]. *Cluster Computing*, 2019, 22(5): 12619–12632.
- [18] Cai Y, Yang K, Huang D, et al. A hybrid model for opinion mining based on domain sentiment dictionary[J]. *International Journal of Machine Learning and Cybernetics*, 2019, 10(8): 2131–2142.
- [19] Dwivedi R, Aggarwal M, Keshari S, et al. Sentiment Analysis and Feature Extraction Using Rule-Based Model (RBM): *Proceedings of ICICC 2018, Volume 2*[A]. *Lecture Notes in Networks and Systems*, 2019: 57–63.
- [20] 吕妹园, 张永健, 张永强, 孙胜娟.融入情感信息词向量的评论文本情感分析方法[J].*河北科技大学学报*, 2021, 42(4): 380–388.
- [21] 孙翌博, 濮泽堃, 徐玉华.应用 SD-LS-SVM 算法的评论情感分析模型[J].*软件导刊*, 2021, 20(4): 43–48.
- [22] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of*

- the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA: Curran Associates Inc., 2017: 6000–6010.
- [23] 陈珂, 谢博, 朱兴统. 基于情感词典和 Transformer 模型的情感分析算法研究[J]. 南京邮电大学学报: 自然科学版, 2020, 40(1): 55–62.
- [24] Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8): 1735–1780.
- [25] Kumar A, Kawahara D, Kurohashi S. Knowledge-Enriched Two-Layered Attention Network for Sentiment Analysis[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers) 2018: 253–258.
- [26] 罗熹, 夏先运, 安莹. 结合多头自注意力机制与 BiLSTM-CRF 的中文临床实体识别[J]. 湖南大学学报(自然科学版), 2021, 48(4): 45–55.
- [27] Tang D, Qin B, Feng X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, Japan: The COLING 2016 Organizing Committee, 2016: 3298–3307.
- [28] Xue W, Li T. Aspect Based Sentiment Analysis with Gated Convolutional Networks[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia: Association for Computational Linguistics, 2018: 2514–2523.
- [29] Devlin J, Chang M-W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171–4186.
- [30] Lan Z, Chen M, Goodman S, et al. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations[J]. 8th International Conference on Learning Representations, arXiv, 2020.
- [31] Song Y, Wang J, Jiang T, et al. Attentional Encoder Network for Targeted Sentiment

- Classification[A]//Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series: 28th International Conference on Artificial Neural Networks, Munich, Germany, Proceedings, Part IV 2019: 93–103.
- [32] 胡任远, 刘建华, 卜冠南.融合 BERT 的多层次语义协同模型情感分析研究[J]. 计算机工程与应用, 2021, 57(13): 176–184.
- [33] Gao Z, Feng A, Song X, et al. Target-Dependent Sentiment Classification With BERT[J]. IEEE Access, 2019, 7: 154290–154299.
- [34] Sun C, Huang L, Qiu X. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 380–385.
- [35] Basiri M E, Nemati S, Abdar M, et al. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis[J]. Future Generation Computer Systems, 2021, 115: 279–294.
- [36] Ma D, Li S, Zhang X, et al. Interactive Attention Networks for Aspect-Level Sentiment Classification[J].2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture (AIAM) ,2017: 4068–4074.
- [37] Li L, Liu Y, Zhou A. Hierarchical Attention Based Position-Aware Network for Aspect-Level Sentiment Analysis[C]//Proceedings of the 22nd Conference on Computational Natural Language Learning, Brussels, Belgium: Association for Computational Linguistics, 2018: 181–189.
- [38] Tang J, Lu Z, Su J, et al. Progressive Self-Supervised Attention Learning for Aspect-Level Sentiment Analysis[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy: Association for Computational Linguistics, 2019: 557–566.
- [39] 程艳, 尧磊波, 张光河.基于注意力机制的多通道 CNN 和 BiGRU 的文本情感倾向性分析[J].计算机研究与发展, 2020, 57(12): 2583–2595.
- [40] 邓钰, 李晓瑜, 崔建.用于短文本情感分类的多头注意力记忆网络[J].计算机应

- 用, 2021, 41(11): 3132–3138.
- [41] Fan F, Feng Y, Zhao D. Multi-grained Attention Network for Aspect-Level Sentiment Classification[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing ,2018: 3433–3442.
- [42] Chen P, Sun Z, Bing L, 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: Association for Computational Linguistics, 2017: 452–461.
- [43] Park H, Song M, Shin K-S. Deep learning models and datasets for aspect term sentiment classification: Implementing holistic recurrent attention on target-dependent memories[J]. Knowledge-Based Systems, 2020, 187: 104825.
- [44] Qiu G, Liu B, Bu J, et al. Opinion Word Expansion and Target Extraction through Double Propagation[J]. Computational Linguistics, 2011, 37(1): 9–27.
- [45] Zhao P, Hou L, Wu O. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification[J]. Knowledge-Based Systems, 2020, 193: 105443.
- [46] Hazarika D, Poria S, Vij P, et al. Modeling Inter-Aspect Dependencies for Aspect-Based Sentiment Analysis[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 , 2018: 266–270.
- [47] Liang B, Yin R, Gui L, et al. Jointly Learning Aspect-Focused and Inter-Aspect Relations with Graph Convolutional Networks for Aspect Sentiment Analysis[C]// Proceedings of the 28th International Conference on Computational Linguistics , 2020: 150–161.
- [48] Majumder N, Poria S, Gelbukh A, et al. IARM: Inter-Aspect Relation Modeling with Memory Networks in Aspect-Based Sentiment Analysis[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018: 3402–3411.
- [49] He R, Lee W S, Ng H T, et al. Effective Attention Modeling for Aspect-Level

- Sentiment Classification[C]//Proceedings of the 27th International Conference on Computational Linguistics , 2018: 1121–1131.
- [50] Miao Y, Luo R, Zhu L, et al. Contextual Graph Attention Network for Aspect-Level Sentiment Classification[J]. Mathematics, Multidisciplinary Digital Publishing Institute, 2022, 10(14): 2473.
- [51] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[J]. arXiv, 2016.
- [52] Veličković P, Cucurull G, Casanova A, et al. Graph Attention Networks[J]. arXiv, 2018.
- [53] Sun K, Zhang R, Mensah S, et al. Aspect-Level Sentiment Analysis Via Convolution over Dependency Tree[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China: Association for Computational Linguistics, 2019: 5679–5688.
- [54] Zhang C, Li Q, Song D. Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) , 2019: 4568–4578.
- [55] 肖泽管.一种融合多种类型语法信息的属性级情感分类模型[D].暨南大学, 2021.
- [56] 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 (Volume 1: Long Papers), Online: Association for Computational Linguistics, 2021: 6319–6329.
- [57] Huang B, Carley K M. Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) , 2019: 5469–5477.
- [58] Bai X, Liu P, Zhang Y. Investigating Typed Syntactic Dependencies for Targeted

- Sentiment Classification Using Graph Attention Neural Network[J]. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2021, 29: 503–514.
- [59] Wang K, Shen W, Yang Y, et al. Relational Graph Attention Network for Aspect-based Sentiment Analysis[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online: Association for Computational Linguistics, 2020: 3229–3238.
- [60] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[C]. //International Conference on Learning Representations, 2013.
- [61] 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), 2014: 1532–1543.
- [62] Nguyen D Q, Verspoor K. An Improved Neural Network Model for Joint POS Tagging and Dependency Parsing[C]//Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, 2018: 81–91.

## 致谢

三年转瞬一逝，如白驹过隙，当写下“致谢”二字时，意味着我的研究生生活将落下帷幕，而我也即将和“王同学”这个称呼告别。很荣幸能与兰财共度三个春夏秋冬，还记得2020年初入校园时的憧憬，殊不知时光如流水，在2023的盛夏与兰财告别。回顾三年生活，有不舍，有遗憾，但更多的是收获良师益友，我想“告别”不是再见，而是新的开始。

一朝沐杏雨，一朝念师恩。在这里首先要感谢我的导师杨春林教授，从学术上认真细致的答疑到生活中无处不在的肯定与鼓励，您温柔友善的人格魅力对我影响深远，是我孜孜以求的人生目标。同时我还要感谢李强教授在这三年对我的指导、引路以及帮助，师者若水，润物无声，我常用“温柔又有力量”来形容您，您时常鼓励我们要用开阔的眼光看世界，要树立独立思考和终身学习的意识，您对我来说，亦师亦友亦亲人。最后我要感谢信工学院的各位老师，你们所授的课程将成为我未来实践的理论基石。

养育之恩，春辉朝霭。在这里我要感谢我的母亲，二十余年里对我无微不至的照顾，在自己力所能及的范围内给我创造最好的条件，让我站在你的肩膀上，见识到外面的繁华世界。感谢您一路支持我走到现在，养育之恩，无以回报，只有不断努力，在未来为你遮风挡雨，只愿您身体健康、欢欣长久。

愿岁并谢，与友长兮。感谢山水一程，在兰财遇到志同道合的小伙伴，美丽开朗的师妹李聪聪还有踏实可靠的师弟赵金雨，能在这里和你们相遇是莫大的缘分。感谢处于五湖四海各自奔前程的老友们，你们给予我人生中重要的“归属感”，这些归属感让我回想的时候有故事，让我觉得这些细碎的时光充盈且美好，欢乐且富足。

行文至此，百感交集，感谢大家在这二十余年的陪伴，也谢谢自己有无穷无尽的热情和勇气，在成长的道路上一直前进，从未驻足。论文停留在致谢，但人生还得继续书写，我们后会有期。

最后，我要感谢所有评阅本论文以及出席论文答辩的各位专家、教授，感谢他们在繁忙的工作中给予指导。