

分类号 C93/69
U D C _____

密级 公开
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 融合语法信息的交互式方面级情感
分析研究

研究生姓名: 王文瑞

指导教师姓名、职称: 李强 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: 2022年5月29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： 王文瑞 签字日期： 2022.5.29

导师签名： 李强 签字日期： 2022.5.29

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，_____（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： 王文瑞 签字日期： 2022.5.29

导师签名： 李强 签字日期： 2022.5.29

Aspect-Level Sentiment Analysis Based on Syntactic Information of Interactive Attention

Candidate : Wang WenRui

Supervisor: Li Qiang

摘 要

随着互联网在各年龄段人群的快速普及,越来越多的人加入到了社交平台和电商购物行列,人们在网络上交流的同时留下了大量的观点和评论,这些带有主观情感态度的网络文本对于消费者而言可以作为消费决策的依据,因此文本情感分析应运而生,方面级情感分析以判断评价实体所对应的情感极性为主要任务,根据这些情感极性判断产品的满意程度,从而为其他潜在消费者提供参考价值,网络上大量的数据为深度学习的应用提供了长足的动力,也为情感极性的判断开辟了新思路。

现存在的研究方法是在方面级情感分类过程中使用注意力机制,给文本特征分配权重,来提取语义层面的信息,忽略了语法信息的使用,导致不能很好地利用不同方面词的意见来进行情感分类,而且以往的研究对上下文和方面实体没有进行单独建模,当文本中有多个方面实体时,不同方面实体与所对应的情感极性无法做到很好地匹配,因此,针对以上问题,本文提出了基于语法信息的交互式情感分析模型(Aspect-Level Sentiment Analysis Based on Syntactic Information of Interactive Attention, 以下简称为 SICA 模型)。模型主要包括两大部分内容:一部分是文本上下文特征的提取;另一部分是基于特定方面词特征的提取;最后将提取的上下文文本特征与特定方面词特征进行拼接输入交互注意力层进行交互,得到情感分类结果。上下文文本特征提取部分主要包括词嵌入层、语法信息提取层、卷积层和双向长短期记忆神经网络层,本文采用 BERT 预训练语言模型作为上下文信息和方面词的词向量模型,语法信息层是以 BERT 模型的中低层作为主要信息来源,以带有依赖关系的依存句法树作为辅助信息,构成语法信息层,卷积层用来获取全局的语义信息,在情感分析任务中,尤其在中文领域容易出现一词多义的情况,双向长短期记忆神经网络含有丰富的语义信息,能够有效地捕捉到一词多义的情况,而单向的长短期记忆神经网络无法处理。特定方面词的特征提取部分,利用依存句法树作为辅助提取方面词,不同方面的方面词不能很好地利用,用卷积神经网络提取特征。交互注意力模块同时设定阈值,交互权重过小时,可进行二次交互,能够全面地利用文本信息,最后,通过情感输出层输出结果,以此判断文本的情感极性。

本文所提出的 SICA 模型的主要创新点有以下几点：

第一，利用 BERT 模型的中低层短语结构，结合标明依赖关系的依存句法树引入句法结构组成模型的语法信息层，作为整个语法信息的抽取层；

第二，以往的研究没有将方面词与上下文信息分开单独建模处理，本文将抽取基于特定的方面词序列特征，以依存句法树作为辅助信息进行提取；

第三，设置交互机制，将基于特定方面词序列与上下文序列进行交互，设置打分函数，二次计算交互权重，全面提取相关的词语信息。

将本文提出的 SICA 模型在 3 个公开数据集 Restaurant、Laptop 和 Twitter 进行实验验证，通过对不同模型之间的性能对比和该模型在上述 3 个数据集上的准确率和 F1 值证明了提出的 SICA 模型的有效性，证明语法信息在提高模型的性能方面有明显贡献。

关键字：BERT 预训练语言模型 语法信息 方面词提取 交互注意力

Abstract

With the rapid popularization of Internet in all ages, more and more people join the social platform and electric business shopping, people in the network of communication left a lot of views and comments at the same time, the network text with subjective emotional attitude to consumers can be used as the basis for the consumer decision-making, so text sentiment analysis arises at the historic moment, Emotional level analysis to determine evaluation entity of the emotional polarity as the main task, according to these emotional polarity judgment's satisfaction with the products, so as to provide reference value for other potential customers, a large amount of data on network provides a great incentive for the application of deep learning, as well as emotional polarity judgment opened up a new way of thinking.

The existing research approach uses attention mechanism in the aspect-level sentiment classification process, which assigns weights to text features to extract semantic-level information and ignores the use of syntactic information, resulting in the inability to make good use of the opinions of different aspect words for sentiment classification, and previous studies do not model context and aspect entities separately, and when there are multiple aspect entities in the text, the different aspect entities cannot be well matched with the corresponding sentiment polarity. Therefore, to address the above problems, this paper proposes the Aspect-Level Sentiment Analysis Based on Syntactic Information of Interactive Attention (hereinafter referred to as SICA model). The model mainly consists of two major parts: one part is the extraction of text contextual features; one part is the extraction based on aspect-specific word features; finally, the extracted contextual text features and aspect-specific word features are stitched together and input into the interactive attention layer

for interaction, and the sentiment classification results are obtained. Contextual text feature extraction part mainly includes word embedding layer, grammatical information extraction layer, convolutional layer and two-way long and short-term memory neural network layer, this paper adopts BERT pre-trained language model as the word vector model of contextual information and aspect words, grammatical information layer is the middle and low layers of BERT model as the main information source, and the dependent syntactic tree with dependencies as the auxiliary information, which constitutes the grammatical information layer, and the convolutional layer is used to obtain global semantic information. In sentiment analysis tasks, especially in the Chinese domain, where multiple meanings of words are likely to occur, the bidirectional long- and short-term memory neural network contains rich semantic information and can effectively capture multiple meanings of words, which cannot be handled by the unidirectional long- and short-term memory neural network. The feature extraction part of specific aspect words, using the dependent syntax tree as an aid to extract aspect words, different aspects of aspect words cannot be well utilized, and the features are extracted by convolutional neural network. The interaction attention module sets the threshold value at the same time, and the interaction weight is too small for secondary interaction to be able to fully utilize the text information, and finally, the results are output through the emotion output layer, which is used to judge the emotional polarity of the text.

The main innovations of the SICA model proposed in this paper are as follows:

Firstly, the syntax information layer of the BERT model is used as the extraction layer of the whole syntax information by using the middle and lower level phrase structure of the BERT model and the dependency

syntaxtree which indicates the dependency relationship.

Second, previous studies did not model aspect words and context information separately. In this paper, features based on specific aspect word sequences will be extracted with dependency syntax tree as auxiliary information.

Third, set the interaction mechanism, based on the specific aspect word sequence and context sequence interaction, set the scoring function, secondary calculation of interaction weight, comprehensive extraction of relevant word information.

The SICA model proposed in this paper was verified experimentally in three public data sets, Restaurant, Laptop and Twitter. The effectiveness of the proposed SICA model was proved by comparing the performance of different models and the accuracy and F1 values of the model in the above three data sets. It is proved that syntactic information contributes significantly to improving the performance of the model.

Keywords: BERT pre-trained language model; Grammatical information; Aspectual word extraction; Interactive attention

目 录

1 绪论	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	2
1.2.1 基于传统方法的方面级情感分析.....	3
1.2.2 基于深度学习的方面级情感分析.....	5
1.3 论文研究内容.....	8
1.4 论文组织结构.....	9
1.5 本章小结.....	10
2 相关技术与理论	11
2.1 文本情感分析概述.....	11
2.2 词向量模型.....	12
2.2.1 One-Hot 编码.....	12
2.2.2 CBOW 语言模型.....	12
2.2.3 Skip-gram 语言模型.....	13
2.2.4 BERT 模型	14
2.3 卷积神经网络.....	16
2.4 循环神经网络.....	18
2.4.1 循环神经网络相关理论.....	18
2.4.2 双向循环神经网络.....	19
2.4.3 长短期记忆神经网络.....	20
2.4.4 双向长短期记忆神经网络.....	22
2.5 注意力机制.....	23
2.6 本章小结.....	25
3 SICA 模型框架介绍	26
3.1 模型构建.....	26
3.2 上下文文本序列提取部分.....	27
3.2.1 词嵌入层.....	27

3.2.2 语法信息层.....	28
3.2.3 卷积层.....	31
3.2.4 BiLSTM 层	31
3.3 特定方面词提取部分.....	32
3.3.1 方面词的提取.....	32
3.3.2 卷积层.....	33
3.4 交互注意力层.....	33
3.5 情感输出层.....	34
3.6 本章小结.....	34
4 实验与分析	36
4.1 目标函数与评价指标.....	36
4.1.1 目标函数.....	36
4.1.2 评价指标.....	36
4.2 实验数据及实验环境介绍.....	37
4.2.1 实验数据.....	37
4.2.2 实验环境.....	38
4.3 实验设置.....	38
4.3.1 实验参数设置.....	38
4.3.2 对比实验设置.....	39
4.3.3 消融实验设置.....	40
4.4 实验结果分析.....	40
4.4.1 模型预测结果分析.....	41
4.4.2 实例分析.....	43
4.5 本章小结.....	45
5 总结与展望	46
5.1 总结.....	46
5.2 展望.....	47
参考文献	48

致 谢.....54

攻读硕士期间发表的论文及科研情况.....55

1 绪论

1.1 研究背景及意义

随着在线社交网络的快速发展,互联网使用成本逐渐随之降低,计算机网络的各类活动已经融入人们的日常生活之中,同时互联网也变成了人们获取信息的主要来源。近年来,使用互联网用户数目不断增加,中国网民的规模已经在 2021 年 6 月达到 10.11 亿,相比去年 12 月增加了 2175 万,与此同时互联网普及率已达到 71.6%,较 2020 年 12 月提升 1.2 个百分点。中国网民使用手机进行上网的比例已达到 99.6%,网民中有 80.3% 的用户使用网络购物,其数目高达 8.12 亿,相比于 2020 年 12 月增长了 2965 万。同时,网民中有 46.4%^[1]的人使用互联网订外卖,其数目相比 2020 年 12 月增长了 4976 万,数目高达 4.69 亿。数据显示,网络购物和外卖用户大幅度增加,通过各类网络平台发表对服务和体验等各类事物的见解,产生了大量的评论数据,这些数据大多以文本居多,而评论文本是信息的载体,其内容丰富复杂,网络评论文本信息涉及教育、科研和日常生活各个方面,很大程度上满足了不同用户群体的需求。同时,大数据时代,商品购买数据以及服务评论数据可以通过开放公开的方式获取,当用户想要购买商品或者了解某个产品时,评论数据可以为用户提供更真实、更具有参考价值的信息。而当产品评论数量高达上千上万时,太多的评价容易产生混乱,误导我们最终的判断。因此利用计算机将所有的评论进行系统性的评价判断,得到评论的情感态度十分有必要,这种对评论文本进行分析的方法称之为自然语言处理(Natural Language Processing ,NLP)^[2]。

从文本评论数据中挖掘消费者的情感及偏好并进行归纳分析的过程称为文本情感分析(Sentiment Analysis ,SA),又称为观点挖掘(Opinion Mining),是自然语言处理领域的一个分支^[3]。文本情感分析,顾名思义就是对给定的文本内容进行情感态度的判定,文本情感分析在生活中被应用于许多领域,在电商网站上对商品评价、社交媒体中对品牌、产品、政策的评价等。对于商家而言,可以根据用户对商品的评价,分析得出用户对哪些产品满意,对哪些产品不满意,从而进

行改进,提高用户的产品体验感,从而吸引更多的消费用户。对于个人而言,在进行选品或者选择某一项服务之前,用户可以根据评价来进行选择购买的商品或者服务。但是,海量的评论数据带给用户极大的不便,对评论逐条进行浏览显得不切合实际,运用情感分析技术可以有效处理这类问题。对于政府而言,政府相关部门可以依据网络媒体中群众对于某项政策的情感反映,有针对性地辅助修订政策,有利于维护社会稳定。

文本情感分析可以根据粒度进一步细分为3类,分别是判断整个文本的情感态度的篇章级情感分析,判断文本中某一句情感态度的句子级情感分析和判断文本中具体某一方面的情感态度^[4]。在日常生活中,尤其是查看商品评论时,情感态度经常是以句子形式体现的,而当我们只想要了解商品的具体某一个属性时,只需要关注给定方面的评价即可。例如顾客想要找一个环境较好的餐厅进行用餐时,就需要确定餐厅评论中环境这一方面的情感极性是否为正面,并且忽视掉食品的口味、价格等其他方面,这种细粒度的情感分析就称为方面级情感分析(Asspect Category Sentiment Analysis, ACSA)。

不同粒度的文本情感分析对应不同的文本数据结构,句子情感分析需要着重对句子的整体信息进行关注,以此来把握句子整体的情感极性。目标级情感分析以捕捉与实体目标相关的信息,以及方面级情感分析中不一定会出现需要分析的方面词,这是整个分析过程中信息有效捕捉的困难之处,并且文本评论往往是不规则的、没有固定结构形式,并且更替速度非常快,对信息采用人工标注、提取工作已经不现实。传统的机器学习也需要耗费大量的人力和物力,方面级情感分析对于数据标注有着极高的要求,需要大规模的数据集进行训练,面对当前大量未标注的数据,如何让从信息主体中挖掘出精确的方面信息,为人们提供参考价值,是方面级情感分析亟待解决的问题。随着研究人员进一步的研究和发现,利用深度学习模型来提升预测准确率已经成为自然语言处理领域的热门方向。

1.2 国内外研究现状

方面级情感分析是情感分析中粒度最细的一项任务,于2010年被提出^[5],其任务是对某给定文本的实体的若干方面进行挖掘,提取出各个方面所包含的情

感极性^{[6],[7]}。首先需要对方面进行识别,在文本中准确的方面信息有时需要人们去识别挖掘,这为文本情感分析工作带来了困难。例如,评论“这家餐厅氛围很好,很安静,并且工作人员很热情”这包含环境和服务两个方面,但是文中并未明确出现包含“环境”和“服务”等方面词,而是通过“氛围很好,很安静”和“工作人员很热情”等隐式的表达来表示,因为方面词信息没有非常准确出现在预测文本中,模型很难精确的识别。目前情感分析的方法主要有两种:传统情感分析方法和使用深度学习的情感分析方法,接下来便分别对两种方法做介绍。

1.2.1 基于传统方法的方面级情感分析

方面级情感分析的传统方法研究有情感字典构建和机器学习方法。对于用语义规则进行方面级情感分析任务而言,在最终进行情感极性判断时,需要构建相关的情感字典。目前研究者们常用的情感字典有台湾大学的中文情感极性词典(NTUSD)^[8]和知网情感词典(HowNet)^[9]。

将文本内容与词典内容进行匹配,从而判断文本的情感极性。情感字典的缺点是在人工上成本较大,大大地限制了其设计的范围,在跨领域研究中有局限性。由于人工构建词典并不能罗列出所有相关的词,可以根据算法自动对已有的人工词典进行扩展,进一步丰富情感词典。在国外研究中最先出现的词典资源是 SentiWordNet^[10],将含义相同的词语结合到一起,为每个词语提供一个情感分数,用于反应用户的真实情感态度。栗雨晴等人^[11]提出一种基于双语词典的多类情感分析模型,该模型有效解决了使用微博时人们语言单一的问题。李祥祥等人^[12]利用网络爬虫技术扩充了语料库,收集英语和汉语等多条语句,还加入了网络中出现的新鲜词语,对语料库中的英语和汉语分别进行分词处理,形成双语词典的初始版本,且后期互联网中出现的新词语可通过爬虫技术进行扩充。李枫林等人^[13]针对情感词典在某一领域存在一词多义、准确率不高和对某些新词无法准确判断情感极性的问题,对现有的情感字典进行有针对性的改进,实验结果表明,进一步的对特定领域构建相关词典可以得到更好的效果。Cai 等人^[14]针对在提取情感特征时,不同语境下情感词具有不同的情感倾向,并且传统的情感词典无法考虑到错误的句子,在利用分类器进行分类,实验证明基于堆叠的混合模型,如 SVM

与 GBDT 堆叠在一起，可以达到更好的性能。

通过构建词典来预测情感分析态度对词典丰富度有着极强的依赖，这项工作耗时又费力。研究者们便根据语义规则对文本进行情感分析。在情感分类过程中需要依赖于特征工程，而特征工程的构建首先需要得到特征向量，然后通过特征向量输入句子在特定方面的语义信息，最后进行情感极性预测得到情感分类结果。例如吴杰胜等^[15]在对中文微博进行情感分析研究时采用基于情感词典的方法和基于深度学习的方法，以已有的情感词典为基础并扩充了六部情感词典，赋予情感值给每一个词典的每一个词，并且引入了语义规则集（句型分析规则和句间分析规则），其进一步提高了中文微博情感准确率以在实验中得以证明。

基于机器学习的方面级情感分析主要是指用统计学习算法来进行抽取特征，对标签定义极性，最后再进行情感分析。Huang 等人^[16]提出了一种基于海量评论数据的无监督学习的情感分析模型，针对微博等社交软件中口语化的汉语单字情感词多义的现象，首先得到文本的情感信息集合，然后利用集合预测文本的情感极性。Turney 等人^[17]提出了一种简单的无监督学习算法 PMI，将微博评论按照推荐和不推荐进行分类，分类依据按照评论中的词性进行预测。在语义相似度计算方面，Liu 等人^[18]提出了一种相似度计算方法，根据知网提供的单词进行快速划分未知词的方法来计算未知词的语义相似度，实验取得不错的结果。孙艳等人^[19]提出的基于无监督学习的 LDA 主题混合模型，该模型在预处理过程中不需要进行样本标注，就可以对每个句子序列的词进行情感标签和主题标签的标注，从而获得主题情感词，在基于无监督学习算法模型中取得了很好的效果。基于无监督的学习算法虽然已经取得了不错的分类效果，但是研究者们常用的方法仍然是有监督学习法，然而有监督的学习算法通过给定训练集带有情感极性的样本，并依赖于人工标注的样本，在测试集上进行情感分类。为验证传统机器学习方法在情感分析中的优劣，Pang 等人^[20]利用监督算法中的朴素贝叶斯、最大熵分类和支持向量机等三种算法在电影数据集上进行实验，结果显示，支持向量机的分类效果更好。预测准确率更高。Kiritchenko 等人^[21]在 SemEval-2014 任务 4 中将监督的机器学习算法运用于检测方面词和方面类别，新构建情感词典来获得表层特征、词典特征，利用 SVM 算法进行方面级的情感极性判断，取得了不错的预测效果。D 等人^[22]对各种情感方法（混合基础意见挖掘、支持向量机分析模式等）

进行了分析,采用语音化、标注化和限制化方法将评论分为正面和负面,并且检测出了垃圾评论,证明了方面数据在分类中具有重要作用。Wang 等人^[23]针对训练数据需要依赖人工标注,提出了一种渐进式机器学习(GML)范式进行方面级情感分析,可以精确的进行机器标注,无需手工标注,在情感分析任务中,从简单的实例进行机器标注,通过迭代因子推理图由易到难学习,实验证明准确度比DNN模型更高。Kamal 等人^[24]提出了一种结合规则和机器学习的情感分析系统以识别特征-意见,用以在不同用户评价中进行情感极性的划分。孙翌博等人^[25]提出了一种SD-LS-SVM算法情感评论模型,构建词网并用上下文分析技术计算分词评分,提出特征向量,对LS-SVM进行置信区间动态优化向量修剪,最后由SD-LS-SVM算法进行情感极性分类,准确率达到70%~85%。黄晨晨等人^[26]利用SVM分类器对藏文数据集进行情感分析,对新浪微博上的藏文微博进行测试,取得了准确率为91%的结果。Liu 等人^[27]在研究中国微博立场检测时采用了基于监督和半监督的机器学习方法,在任务数据集上比较了线性SVM、朴素贝叶斯和随机森林等分类器,实验结果表明,效果最好的是基于集成学习且具有情感词特征的方法,尽管基于机器学习的方面级情感分析取得了不错的效果,但是依赖于大量人工标注的文本特征,随着互联网技术的发展,产生的海量数据仅仅依靠人工进行构建特征显得不现实,并且机器学习方法对于不同数据集所需的特征不同,泛化性差。近年来,深度学习模型在其他领域都有新的突破,研究者们也不断研究将深度学习应用到情感分析中。

1.2.2 基于深度学习的方面级情感分析

基于传统学习的方面级情感分析忽略了文本的语义信息,在不同语境下相同的单词有不同的意思,传统的机器学习无法捕捉到,基于深度学习的方法能够有效的捕捉单词间的语义信息和联系,在文本情感分析任务中效果显著。目前,由于人工神经网络的优越性能,在文本情感分析中广泛使用各种神经网络模型,研究方法中最为常见的有单纯使用神经网络算法、多种神经网络算法的融合使用,神经网络算法融合注意力机制的网络模型以及使用预训练模型^[28]。Bengio 等人^[29]为了解决测试和训练中单词序列不同的问题,引入了神经网络语言模型,并将

文本的上下文信息输入神经网络语言模型来预测下一个单词的内容,有效了解决了传统机器学习过程中忽略上下文语义的问题。Zhang 等人^[30]利用卷积神经网络(Convolutional Neural Network, CNN)的特点,在卷积层中加入词袋转换,提出了 Bow-CNN 和 Seq-CNN 模型,最终获得了较高的准确率。Kim 等人^[31]将多个卷积核组合使用提出了 TextCNN 模型,该模型能深度获取句子的 n-gram 特征,因此在短文本领域中能得到较高的准确率。针对评论语料内容短且涉及范围广泛的问题,王煜涵等人^[32]根据卷积神经网络提出一种针对 Twitter 的情感分析模型。模型首先通过预训练模型对数据进行训练得到初始化数据,在通过深度卷积网络将文本的深层语义进行分析,取得了 F1 值为 82.3%的结果。苏灵松等人^[33]提出了双通道卷积记忆神经网络来获取更好的文本特征和利用上下文语义信息,在输入层中将通道 1 中获取的原始词向量和通道 2 中对结果影响较大的 5 种特征(情感词、词性、程度副词、否定词和标点符号)作为扩展特征向量融合,再从卷积层中获取上下文信息特征,有效的提高了文本分析的准确率。虽然卷积神经网络在方面级情感分析中文本特征提取效果不错,但是卷积神经网络的输出只考虑了前一输入的影响而无法考虑到其他时刻的输入,在文本情感分析过程中,上下文对于输出结果有很大的影响。基于以上问题,研究者们将循环神经网络(Recurrent Neural Network, RNN)结构用于文本特征学习,希望进一步考虑到上下文对分析的影响。Socher 等人^[34]针对单字向量空间模型无法捕捉长短语,引入了一个新模型 RecNN,利用语法结构解析树获得单词本身内在含义以及相邻单词或短语的含义,能够捕捉到更细粒度的语义关系,在电影评论数据集上的实验结果显著提高。Zhang 等人^[35]为了得到包含语义和序列特征的特征向量,对不同长度的句子使用循环神经网络进行向量训练。但是训练过程中,面对长序列文本时,RNN 需要不断计算并储存之前学习的信息,非常容易出现梯度消失和梯度爆炸,这会极大的影响最终的预测结果。因此,在文本情感分析中引入了长短时记忆(Long Short-Term Memory, LSTM)网络。Wang 等人^[36]提出的基于注意力机制的长短时记忆网络可以让注意力机制在不同的方面输入时,使得句子的不同部分被集中注意,在 SemEval 2014 数据集上进行了实验,该模型在方面级情感分析方面达到了先进的性能。在文本情感分析任务中,训练数据中只能获得显性知识,对于外部知识的挖掘比较欠缺。针对这一问题,Hu 等人^[37]提出了一种基于扩展注意

力机制的 LSTM 方面级情感分析模型,该模型通过词语嵌入扩展知识连接,利用 LSTM 的输入和输出门来过滤多余的信息,神经网络模型学习丰富的扩展知识,帮助改进方面级情感分析模型的情感极性计算,实验结果证明,该扩展模型具有较好的性能。Yin 等人^[38]提出了层次迭代注意力模型,通过文本与方面问题之间频繁和重复的交互来构建面向方面的表示,由与方面相关的关键字和方面评级来构成伪问题答案,来进行方面情感信息的提取。Ma 等人^[39]提出了交互式注意力网络 (IAN),针对以往建模中忽略了上下文信息的问题,该模型将获得的信息表征,进行拼接作为方面的情感特征,大大优化了模型的预测性能。Hazarika 等人^[40]针对方面词内部之间存在关联会对情感分析结果产生影响的方面进行建模,使用 LSTM 来处理方面词之间的关联性,通过研究方面词的关联性,得出影响方面词之间的关联性的一个原因是方面在文本中的顺序特征的结论。在以往的建模过程中,研究者们忽略了文中个别单词的重要程度。针对这一问题,张克等人^[41]提出了一种基于内联关系的方面级情感分析模型,使用门控机制对方面信息和单词进行融合分布式表达,利用基于注意力机制的 LSTM 和查询机制来增加内联关系的权重,进行情感分类,实验证明取得了不错的分类效果。通过显式方面提取和隐式方面提取可以获得方面词,现有的隐式提取技术存在泛化能力差的问题。汪兰兰等人^[42]针对这一问题,提出了基于依存句法分析与交互注意力机制的隐式方面提取模型,利用预训练语言模型 BERT 提取初始表征,通过依存句法的交互注意力层,充分利用文中词的语义信息。在方面级情感分析任务中,显式方面提取过程中,当方面词由多个单词构成时,单词的词性可能不同,并且方面与极性对之间存在关联。针对这一问题,卜令梅等人^[43]提出了利用关系抽取技术来识别方面极性对,将方面-极性对之间的关联映射成一元关系抽取问题,通过序列解码捕捉,实验证明,目前性能超过了现有的模型。方面级情感分析过程中,对于同时提取全局的上下文信息和局部的情感信息无法进行, YE 等人^[44]提出了基于 ALBERT-CNN 的方面级情感分析模型,该模型中使用了 TransConvs 双向编码层,TransConvs 特征提取器主要提取全局句子特征信息和位置特征信息,以及本地方面的信息和情感信息来自文本向量,再通过卷积层来增强特征提取器的输出,能够很好地同时提取出全局的上下文信息和局部的情感信息。词性的不同会影响分类结果的准确度,薛芳等人^[45]针对不同词性的贡献程度的不同提出了基于双

层词性感知和多头交互注意力机制的方面级情感分析模型,充分考虑了局部特征和全局特征的交互作用,双层词性感知降低了噪音词的影响,利用多头交互注意力机制充分挖掘交互特征,在输出层使用了动态特征融合层,在三个公开数据集上进行了实验,准确率有所提升。在充分考虑全局上下文信息建模过程中,随着参数量的增加,计算量也逐渐增大,夏鸿斌等人^[46]提出了基于多头注意力机制的局部与全局特征融合网络模型,使用双向的门控机制对上下文编码,用方面项的语义相关距离代替相关度较小的上下文词,再通过多头注意力机制对局部和全局信息进行提取并融合。

1.3 论文研究内容

本文的研究内容主要是方面级情感分析模型性能的提升,其是近年来自然语言处理的任务之一,受到众多研究者的关注,其主要的任务是判断目标对象某一方面的情感极性,常见的为三分类问题,即分为积极、消极或中立。早些研究学者利用传统的机器学习来进行实验,虽然取得了一些不错的结果,但是全靠人工构建特征,使用深度学习算法能够自动构建文本特征,可以自动从句子学习获取到丰富的文本特征信息,再通过注意力机制捕捉到对结果贡献较大的信息,使用 LSTM 模型建模的深度学习模型非常多,但是数据集的扩大和模型复杂度的增加,导致模型参数量增多,模型通常会出现过拟合等问题,通常上下文中所包含的词性信息、语义信息、句子结构等信息对情感分析结果的影响很大,但是以往的研究中,往往忽略了语法信息的使用,会导致模型难以正确匹配方面与对应的情感表达,因此,针对上述问题,本文提出了基于语法信息的交互式方面级情感分析模型。首先,采用了 BERT 预训练词向量, BERT 模型中含有丰富的语法信息,拥有多个层,每个层所学习到的内容不同,又引入了带有依赖关系的依存句法树作为辅助信息,带有依赖关系的依存句法树更直接地描述词与词之间的联系关系,构成了语法信息层,将语法信息层经过双向长短期记忆神经网络和卷积神经网络后与方面词特征进行交互,上下文与方面词地正确匹配能提高分类效果,使用依存句法树将特定方面词提取特征再经过卷积神经网络输出与带有语法信息的上下文信息进行交互就构成了整个模型框架,本文使用的 3 个数据集是公开数据集

Restaurant、Laptop 和 Twitter 进行实验，实验结果显示提出的 SICA 模型在以上 3 个数据集上的评级指标均有所提升，从而证明了模型的有效性，但是模型仍然存在不足，结果有待继续优化。

1.4 论文组织结构

本文的主要内容包括五大部分，具体详细内容如下：

第一章：绪论部分。第一章绪论部分首先介绍了方面级情感分析的背景及意义，通过简单介绍了解到什么是文本情感分析以及为什么需要对文本进行细粒度的情感分析，接着介绍方面级情感分析的国内外研究现状，分为两个方面：基于传统方法的方面级情感分析和基于深度学习方法的方面级情感分析现状，介绍着重介绍了本文的研究内容，最后对本章节进行总结。

第二章：相关技术与理论部分。本章节详细介绍了基于深度学习方法的方面级情感分析的理论基础，分别从文本情感分析概述、词向量模型、神经网络模型和注意力机制四个方面展开详细介绍。

第三章：基于语法信息的交互式方面级情感分析模型框架(SICA 模型)介绍。本章主要详细阐述了该模型构建的过程和方法，主要分为两部分内容：基于语法信息的文本序列特征提取部分和基于特定方面的提取过程，最后用交互注意力机制使得上下文信息与方面词更好地匹配，以提高模型分类效果。

第四章：实验与分析部分。本章主要详细分析了模型实验结果，并且进行基线实验对比，消融实验对比，详细阐述了模型各个模块对整体模型的性能影响，以此来验证本文提出模型的有效性。

第五章：总结和展望部分。对本文的全部内容进行一个全面详细的总结，并且思考与分析当前存在的问题以及接下来的解决方案，最后对方面级情感分析的研究后续工作进行展望和思考。

本文的主要框架内容如图 1.1 所示。

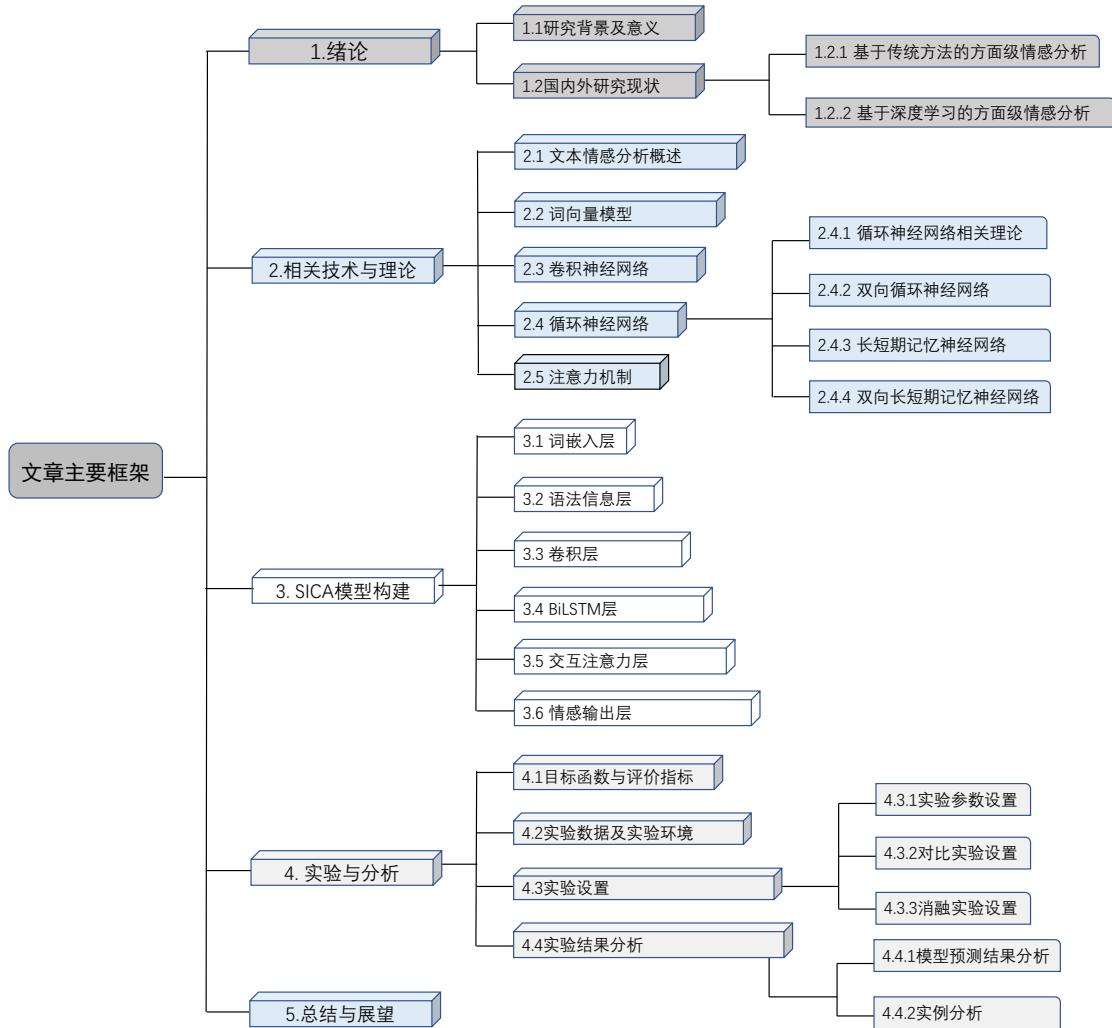


图 1.1 论文框架图

1.5 本章小结

本章分别阐述了文本情感分析的研究背景及意义、方面级情感分析的国内外研究现状、研究内容和论文组织结构几部分，详细描述了基于深度学习方法的方面级情感分析现状，为本文的下一步研究奠定了基础。

2 相关技术与理论

本章主要将方面级情感分析的理论和技术进行介绍，共分为四个部分，分别为词向量模型、神经网络模型、注意力机制和文本情感分析概述等。

2.1 文本情感分析概述

作为自然语言处理的重要组成部分，文本情感分析的主要研究内容是利用计算机技术对自然语言文本的进行准确的分析和自动分类、抽取和计算文本的情感属性，情感分析具有极高的研究价值，从社会发展的各个方面来说，人类独特的主观情感在认知和感知事物过程中发挥着极其重要的作用。在文本情感分析任务中，将情感极性划分为三类：分别为正面，负面和中性，并且根据粗细粒度不同，可以进一步具体划分为目标级情感分析、句子级情感分析和方面级情感分析这三种类型，判断给定文本中的一个或者多个目标的实体的情感极性是目标级情感分析的主要任务，对于句子级情感分析任务，即判断文本的整个给定句子的情感极性，而方面级情感分析与上述的不同在于，是判断给定文本中的一个或者多个方面实体的情感极性。

对于句子级情感分析的表述如下：对于给定文本中的句子 $S = \{x_1, \dots, x_n\}$ ， x_i 表示文本中的第 i 个单词。句子级情感分析的任务就是判断 S 的整体情感极性。对于目标级情感分析来说，首先要给定目标实体，可以是一个目标或者 K 个目标实体，表述如下：对于给定文本 $S = \{x_1, \dots, x_n\}$ ，目标实体 $T = \{t_1, \dots, t_k\}$ 。 T 中每个 t_i 都对应于文本 S 中的序列，目标级情感分析的任务就是判断文本 S 中所对应的每个 t_i 的情感极性。对于方面级情感分析来说，与目标级情感分析的任务类似，把给定目标换成给定方面就成了方面级情感分析，具体表述如下：对于给定文本 $S = \{x_1, \dots, x_n\}$ ，给定 K 个方面实体 $D = \{d_1, \dots, d_k\}$ 。方面级情感分析的任务就是判断文本中 S 对于给定 D 中的每个方面实体 d_i 的情感极性。由于方面级情感分析能够更加全面地考虑到文本的上下文信息，因此方面级情感分析在近几年逐渐被学者们关注，进行情感分析的第一步就是对文本进行预处理，将文本转化成机器能够识别的特征表示，因此，文本预处理获取地文本特征的好坏会影响最终的分

类效果。因此，我们对文本的文本表征用到的技术进行详细的介绍。

2.2 词向量模型

文本情感分析过程中，词向量是文本情感分析的基础，文本特征的质量好坏影响着后期模型的性能，所以首先要获取文本的词向量。本小节从 One-Hot 编码、CBOW 语言模型、Skip-gram 语言模型、和 BERT 四种方式展开介绍。

2.2.1 One-Hot 编码

在机器学习任务中，特征分为离散特征和连续性特征，将离散特征转化成数字表征时无法直接运用到分类器中，为了解决上述问题，在离散特征处理过程中使用 One-Hot 编码。One-Hot 编码过程中将词语表示成一个高维稀疏向量，在该向量中所有的元素都是 0 和 1 组成，当该向量中有对应词出现时对应的元素表示为 1，其余都为 0。用 One-Hot 编码文本时虽然过程简单，但是一方面会造成维度灾难，文本中词汇量的增加会导致文本中的词所对应的向量维度会增加，计算机的运算能力会下降，一方面，对于需要计算特征之间的距离，判断变量之间有没有偏序关系时，One-Hot 编码就无法有效处理，One-Hot 编码是根据词频进行排序，词语在文中的顺序被忽略，而且词语之间是互相独立的，没有相关性，进而导致上下文语义的信息丢失。所以近年来，多使用单词的分布式向量表示来解决此类问题。

2.2.2 CBOW 语言模型

分布式词向量表示可以大幅度地降低向量维度，将向量各维度信息的分布式存储，并且在高维度向量中只存在一个维度的向量去描述词的语义，因此，分布式向量表示可以增强词语之间的联系。连续词袋模型(Continuous Bag-of-Words, CBOW)^[47]的处理过程是以上下文词去预测目标词，是一种统计模型。该模型结构如图 2.1 所示。

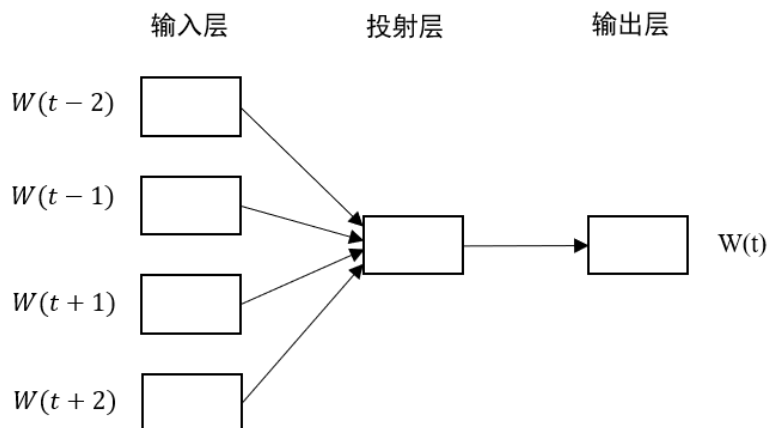


图 2.1 CBOW 模型结构图

在 CBOW 模型具体预测过程中，首先确定目标词所在的窗口，然后得到该窗口所在的上下文词向量，最后得到目标词的具体向量表示。输入一段训练文本 $S = \{w_1, \dots, w_t\}$ ，滑动窗口大小为 C ，即以目标词为中心，以 C 为半径范围内的上下文单词的词向量，中间部分的投射层就是将高维度的词向量映射为具体指定维度大小的词向量，输出层在投射层输出的词向量的条件下预测为目标词的条件概率分布值 $P(w_t | w_{t-c} : w_{t+c})$ ，最后计算模型的训练损失 L （除目标词以外的所有词的对数概率分布求均值）。具体计算公式如式(2.1)(2.2)：

$$L = \frac{1}{T} \sum_{t=1}^T \log P(w_t | w_{t-c} : w_{t+c}) \quad (2.1)$$

$$P(w_t | w_{t-c} : w_{t+c}) = \frac{\exp(\bar{v}^T v_{w_t})}{\sum_{n=1}^N \exp(\bar{v}^T v_n)} \quad (2.2)$$

式中 T 表示训练文本中词的个数， N 表示大维度向量大小， C 表示滑动窗口的大小， v 表示输入词 w 的词向量。在该模型的预测过程中，每个周围词都会轮流作为目标词进行预测从而获得整个文本的词向量，因此整个文本向量获取的次数与词库的数量几乎相同，导致模型的复杂度升高。

2.2.3 Skip-gram 语言模型

而 Skip-gram 模型是以目标词来预测周围词，这个预测过程恰好与 CBOW 模型相反，该模型的结构如图 2.2 所示。

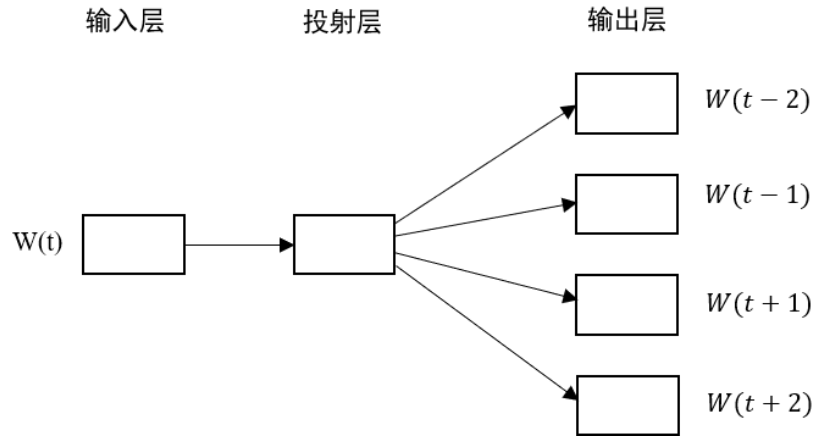


图 2.2 Skip-gram 模型结构图

Skip-gram 模型已知某个目标词 w_t 去预测周围的上下文词 ($w_{t-c}, \dots, w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2}, \dots, w_{t+c}$)。图 2.2 中滑动窗口大小为 2，输入一段训练文本 $S = \{w_1, \dots, w_t\}$ ，滑动窗口大小为 C ，目标词设置为 w_t ，得到周围词的条件概率分布值为 $P(w_{t+c}|w_t)$ ，最后计算模型的训练损失 L ，目标损失 L 利用随机梯度下降算法优化和更新模型参数，具体计算公式如式(2.3)(2.4)：

$$L = \frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log P(w_{t+c} | w_t) \quad (2.3)$$

$$P(w_{t+j} | w_t) = \frac{\exp(w_{t+j}^T w_t)}{\sum_{n=1}^N \exp(w_{t+j}^T w_n)} \quad (2.4)$$

式中的 T 表示训练文本中词的个数， N 表示大维度向量大小。在 Skip-gram 模型中，由于是一个目标词对应多个周围词，每个周围词对目标词进行训练，预测结果比 CBOW 模型准确些，但是耗时更长，模型复杂度更高。但无论是 CBOW 模型还是 Skip-gram 模型都属于静态词向量模型，其词向量是不会随着文本单词的上下文语境的变化而变化，一般是通过训练大型语料库得到的，单词的词向量表示是唯一的，无法考虑到文本的语义信息，因此我们接下来介绍动态词向量中的 BERT 模型。

2.2.4 BERT 模型

BERT^[48]作为一种动态词向量语言模型，在 2018 年提出，全称是 Bidirectional

Encoder Representation from Transformers, BERT 模型结构如图 2.3 所示。

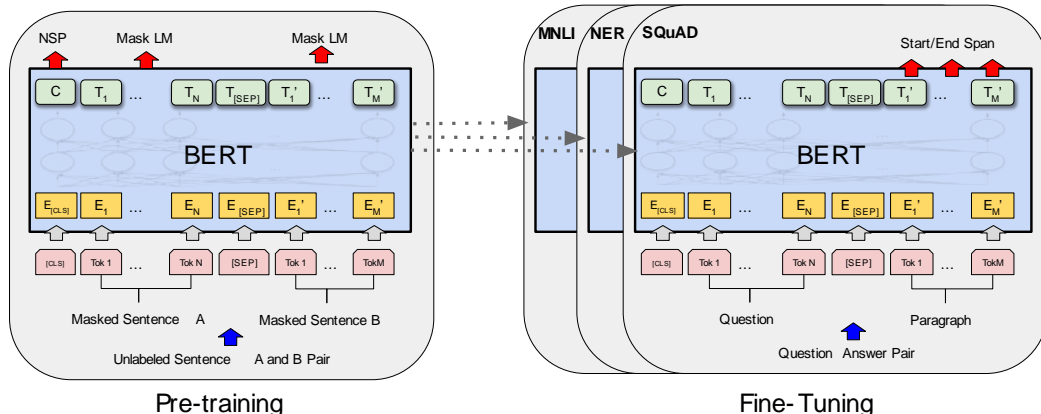


图 2.3 BERT 模型结构图

该模型采用了双向 Transformer 的编码器，在进行预训练的过程中采用了两个新的无监督预测任务：Masked LM 模型 和 Next Sentence Prediction 模型，用 Masked LM 预训练模型方法学习词语在上下文中的表示，用 Next Sentence Prediction 模型学习句子级的表示。

Masked LM 预训练模型方法学习词语在上下文中的表示，Masked LM 预训练模型在预训练过程中，首先将每个单词序列中 15% 进行替换，如果所替换的单词要被预测时，模型利用剩余单词的上下文进行预测被替换的单词，因此最终的损失函数计算只有被替换的单词被计入。由于 Masked LM 预训练过程中不知道具体哪些单词被 Masked，所以模型对于每个词的关注度是一样的，增加了文本的情境意识。但是 Masked LM 预训练过程中句子之间的关系无法考虑，而句子之间的关系对于文本情感分析非常重要，因此 Next Sentence Prediction 任务就是一个简单的二分类，将句子 A 和 B 链接起来，预测文本中的句子 A 是不是在句子 B 的前面，原始文本中的前后关系包括具体训练过程中 50% 的输入，剩下的 50% 是从语料库中随机生成的。在自然语言处理的 11 个任务中都获得了大幅度的提升，因此在方面级情感分析任务中使用该模型可以提升文本表征能力。该模型的基本思想是语言模型处理分为上游和下游任务，上游任务是通过训练文本取得词向量，下游任务是利用词向量得到单词的最佳上下文表示，获得的词向量要包含丰富的上下文信息，才能更多地利用语境信息，因此预训练模型从左到右训练和从右到左训练串联起来组成双向的语言模型来完成下游任务。

BERT 模型的 Embedding 由三部分组成, 分别为 Token Embeddings 嵌入、位置嵌入(Position Embedding)和分割嵌入(Segment Embedding), Token Embeddings 嵌入层可以将各个词转换成固定维度, 在将文本送入 Token embeddings 嵌入层之前, 要对文本进行 tokenization 处理。例如输入一段文本为: “I like to eat pasta in restaurants”, 在 tokenization 的结果的开头 ([CLS])和结尾 ([SEP])会插入两个特殊的 token。其是为了将句子对进行划分开, 以便完成分类。经过处理, 句子就被转化成 “[CLS] I like to eat [SEP] pasta in restaurants [SEP]”。位置嵌入(Position Embedding)层是将位置在文本序列中的位置信息进行编码得到位置向量, 能够将同一个单词在不同位置进行不同的向量表示, 是通过训练得到的。分割嵌入(Segment Embedding)是判断两个句子先后顺序, 用来区分句子, 前一个句子的每个 token 都用 0 表示, 后一个句子的每个 token 都用 1 表示, 例如 “[CLS] I like to eat [SEP] pasta in restaurants [SEP] ” 表示成 “0 0 0 0 0 0 1 1 1 1 ”。在预训练结束后, 需要依据不一样的下游任务, 处理 BERT 的不同位置, 即微调(Fine-tuning)功能, 例如对于常规分类任务, 只需要在 Transformer 的输出之前加入一个分类层。BERT 模型采用了双向 Transformer 编码器, Transformer 编码中运用了自注意力机制(Self-Attention), 帮助机器来识别词之间的相互关系, 例如 “I am a student, You and me are students ” 我们可以很容易地判断 “I” 和 “me” 都是指同一个人, 但是机器很难识别, 自注意力机制就是根据权重重新调整句子, 得到新的文本特征表示, 因此在方面级情感分析任务中使用 BERT 模型可以充分考虑单词之间上下文之间的语义信息。

2.3 卷积神经网络

卷积神经网络(Convolutional Neural Networks, CNN), 最先运用于图像识别领域, 作为深度学习算法中的一个重要组成部分, 于 1988 年提出, 本质前身是感知机, 是一种前馈神经网络, 主要组成部分包括输入层、卷积层、激活层、池化层和全连接层, 模型结构如图 2.4 所示。

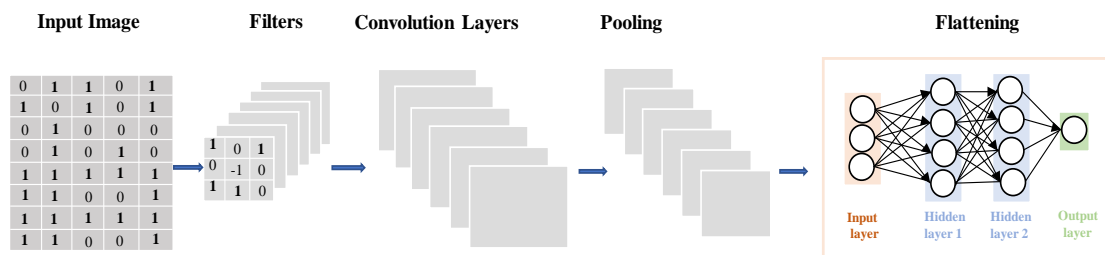


图 2.4 卷积神经网络模型结构图

作为神经网络文本分类的一个例子，首先，将一个词向量模型处理过的固定大小的矩阵输入到输入层里。每一个词语是一个一维向量，但是作为输入层，我们必须得到一个二维矩阵，矩阵中的每一行对应不同的词，不同的词用不同的向量表示，例如：输入 $100 * 256$ 的矩阵，将二维卷积窗口的第二维固定成 256，就是将一维卷积看作二维卷积的简易版。卷积层的作用就是将卷积核矩阵和对应的输入层中的特征矩阵进行点积相乘，卷积核通过权重共享的方式，根据步幅大小进行周围滑动提取特征，在文本特征提取过程中，从句子中提取特征时，卷积的窗口宽度应该覆盖单个词向量，即卷积核的宽度必须等于词向量的维度。在图像处理过程中，需要对输入层进行 padding 操作，即将输入层的四周以 0 元素填充来确保图像边缘特征可以完整提取到，在文本情感分析领域，由于在词嵌入时就已经将训练集中的所有句子都进行了截长补短操作，因此不需要进行 padding 操作，设置卷积核的数量进行提取文本的多层次特征。池化层中文本分类一般采用最大池化，最大池化是在窗口滑入特征区域时选择该区域的最大值作为输出。最后将提取到的向量进行按行拼接成最终向量，而后传入带有激活函数的全连接层进行分类。

卷积神经网络模型可防止过拟合现象，因为其权值共享的操作可以有效地减少网络参数数量，直接减少连接数量。由于输入文本之间的空间联系是局部的，不需要对所有的文本进行感受，只需要感受局部特征，最后只要将局部综合起来得到全局即可，因此可以减少连接的数目。卷积操作过程中共享同一个卷积核的方式就是权值共享，在输入矩阵的不同位置使用完全相同的函数，可以共享同一组参数，在处理文本特征提取过程中，通过权重共享可以检测出现在文本不同位置的完全相同的词，并且通过权值共享可以实现平移操作，被称为“平移不变性”，也就是说，在整个模型的训练过程中想要减少模型的网络参数，避免出现模型的

过拟合问题，卷积神经网络的“平移不变性”起重要作用，因此近几年在自然语言处理领域中得到快速发展。

2.4 循环神经网络

循环神经网络(recurrent neural network, RNN)^[49]，是一种人工神经网络，与卷积神经网络不同的是卷积神经网络的前馈传播是无记忆的，而循环神经网络最大的特点是可以存储和计算以往的信息，带有记忆性，对带有时间序列的数据非常有效，并且可以挖掘数据中的时间序列和相关语义信息，被广泛运用于时间序列预测、机器翻译等领域。本节首先介绍循环神经网络的相关理论，接下来对双向循环网络、长短期记忆网络和双向长短期记忆网络进行仔细的介绍。

2.4.1 循环神经网络相关理论

在自然语言处理中，对于不同句子中相同的单词在文本中出现时，如果使用全连接神经网络时，我们只考虑该单词的特征向量在模型中出现的次数，无法结合文本的上下文去训练数据，例如：“I like eating apple”和“The Apple is a great company!”两个句子中“apple”代表的意义不同，但是机器却无法分别，从而RNN就应运而生，循环神经网络的结构简单，其组成主要包含三部分：输入层、隐藏层和输出层，其基本结构如图 2.5 所示。

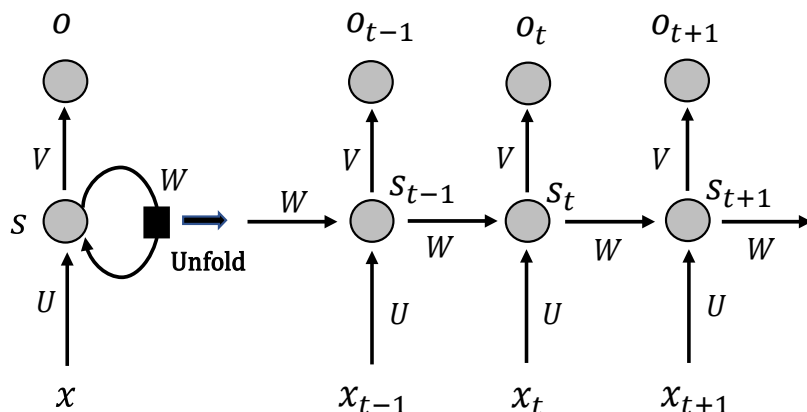


图 2.5 循环神经网络模型结构图

如图所示,左边部分 RNN 的结构简图,右边部分是时间序列网络的展开图;例如,“I like it”这句中,“I”的单词向量是由图中 X_{t-1} 表示,“like”的单词向量是由 X_t 表示,“it”的单词向量是由 X_{t+1} 表示,而每个时间点之间的权重矩阵由 W 表示,在每个间隔的隐藏层不仅由当时的输入层决定,同时也是由前一时期的隐藏层决定的。时间长度 t 是由文本输入序列的长度所决定的,RNN 向前传播时的状态更新方程如下所示:

$$O_t = \text{softmax}(V \cdot S_t) \quad (2.5)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1}) \quad (2.6)$$

公式(2.5)中, t 时刻的输出由 O_t 代表,而 t 时刻隐藏层的值则用 S_t 代表,公式(2.6)中, f 代表非线性激活函数,通常为 \tanh 函数或者 Relu 函数。 U 是用来调节输入 X_t 的权值矩阵, W 是前一个时刻隐藏状态的权值矩阵。隐藏状态 S_t 看作是循环神经网络的记忆单元,用来捕获上一状态的相关信息, O_t 输出的结果取决于时间 t 的隐藏状态 S_t ,对应的权值矩阵为 V ,神经网络的参数共享机制,有效的降低了模型的训练成本,然而,循环神经网络处理序列数据的能力是有限的,这会导致梯度消失或梯度爆炸的问题,从而导致模型训练失败,因此为了解决模型的不足,研究人员提出了更为复杂的神经网络模型,如:双向循环神经网络(Bidirectional RNN)、长短期记忆网络(LSTM)和双向长短期记忆网络(Bidirectional LSTM)。

2.4.2 双向循环神经网络

在单向循环神经网络中,下一次的输出只能由上一个时间段的时序信息来预测,但当前时间的输出不仅与上一个状态有关,还同时与未来状态相关,需要真正地利用上下文信息,因此提出了一个由向前和向后两个组合起来的循环神经网络,即双向循环神经网络(Bi-directional RNN),其基本结构如图 2.6 所示。

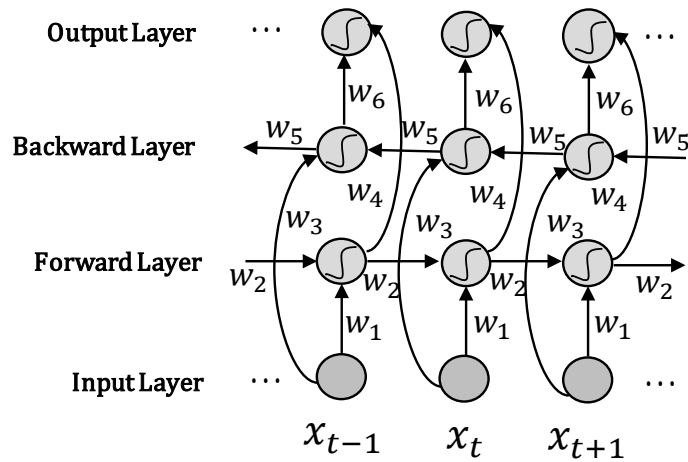


图 2.6 双向循环神经网络模型结构图

t 时刻的输出结果可以同时使用 $t - 1$ 时刻的信息和 $t + 1$ 时刻的信息，双向结构允许序列的每一个时刻都能捕获到过去和未来的上下文信息，具体的计算过程如式(2.7)(2.8)(2.9)所示：

$$\vec{S}_t = f(\vec{W}X_t + \vec{V}\vec{S}_{t-1} + \vec{b}) \quad (2.7)$$

$$\overleftarrow{S}_t = f(\overleftarrow{W}X_t + \overleftarrow{V}\overleftarrow{S}_{t-1} + \overleftarrow{b}) \quad (2.8)$$

$$S_t = g(U[\vec{S}_t; \overleftarrow{S}_t] + b) \quad (2.9)$$

式中，前向的隐藏层由 \vec{W} 表示，隐藏层的权值矩阵由 \vec{V} 表示，前向隐藏层的偏置由 \vec{b} 表示。 \overleftarrow{W} 和 \overleftarrow{V} 分别代表后向的隐藏层和隐藏层的权值矩阵， \overleftarrow{b} 代表后向隐藏层的偏执， f 和 g 代表的是激活函数， U 表示拼接后的向量。

若要充分考虑到文本的上下文信息，则在处理文本建模时，运用双向循环神经网络，双向循环神经网络在预测文本中一段缺失序列单词时，就可以根据上文和下文的单词去预测判断当前状态的序列，与单向循环神经网络相比，双向循环神经网络可以更完整地对孩子进行建模，从而提高情感分析的效果。

2.4.3 长短期记忆神经网络

基于最基本循环神经网络结构固有的缺陷，如当网络结构较复杂时，会出现梯度爆炸和梯度消失的现象。长短期记忆神经网络(Long Short-term Memory

Networks ,LSTM))就能有效的处理这一问题。LSTM 模型结构更为复杂，LSTM 在传统的 RNN 中加入了输入门、遗忘门和输出门，新加入的门结构能够保留对后续任务有效的信息，同时遗忘不需要的信息。其基本结构如图 2.7 所示。

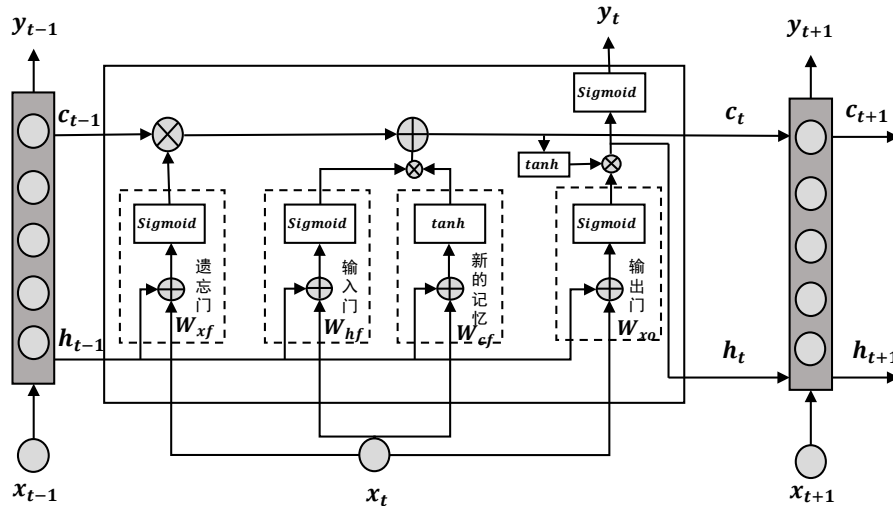


图 2.7 长短期记忆神经网络模型结构图

LSTM 会对经过的数据信息分三个阶段进行处理，首先是对数据进行有选择的更新，即保留有效信息，过滤无效信息；接着是对进一步保留的信息进行储存记忆，如果信息重要，则将信息储存下来；最后阶段是把前一阶段的需要使用的信息输出。经过三个阶段的处理，就可以长序列文本进行分析从而避免普通循环神经网络的缺点。

LSTM 的记忆单元就相当于一个门控单元，可以根据输入信息的重要程度进行控制信息流，根据神经网络中权值分配，权值占比小时信息不重要会被丢弃，权值比重高时会保留此信息，这一功能通过sigmoid函数 σ 来实现，sigmoid函数的输出范围是[0,1]的数值，其中 0 代表所有的信息都会被丢弃，1 代表所有的信息都会被保留下来并输出到下一个时刻的状态中。神经网络参数的更新过程如下。

一、遗忘门的更新过程由当前时刻的信息 x_t 和前一时刻的输出值 h_{t-1} 决定，具体的更新计算过程如式(2.10)所示：

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (2.10)$$

式中， f_t 代表遗忘门， σ 为 sigmoid 激活函数， W_{xf} 、 W_{hf} 和 W_{cf} 分别代表输入、隐藏状态和记忆单元的参数权值矩阵， b_f 代表遗忘门的偏置向量。

二、LSTM 中的输入门可以控制信息的存储与否，有用的信息将会保存在记忆细胞中。首先，信息的存储与否是由 sigmoid 激活函数决定的，进一步的，需要储存的信息将与遗忘门中有用的信息一起通过 tanh 函数储存到新的记忆单元中，具体更新计算过程如式(2.11)(2.12)所示：

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2.11)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (2.12)$$

式子(2.11)和(2.12)中 W_{xi} 代表输入的参数权重矩阵， W_{hi} 代表上一隐藏状态的参数权重矩阵， W_{ci} 代表记忆单元的参数权重矩阵， σ 代表的是 sigmoid 激活函数，(2.12)式中代表意义与(2.11)相同， b_i 和 b_c 分别代表输入门和记忆单元的偏置矩阵。

三、LSTM 将根据新的记忆单元在输出门得到新的信息，分为两个步骤，首先利用常用的激活函数 \tanh 对相应记忆细胞的信息进行更新，接着与上一步中得到的结果累乘，最终实现一个输出的范围为[-1,1]的信息。具体的更新计算过程如式(2.13)(2.14)所示：

$$O_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o) \quad (2.13)$$

$$h_t = o_t \tanh(c_t) \quad (2.14)$$

式中 W_{xo} 代表输入门与 x_t 的权重矩阵， W_{ho} 代表输出门与隐藏层的权重矩阵， b_o 代表输出门与记忆单元的偏置向量。

在大多数使用 RNN 的任务中，LSTM 都能取得不错的效果。在本文情感分析领域中，由于语料数据的复杂性，LSTM 常被应用到该领域。

2.4.4 双向长短期记忆神经网络

LSTM 的出现解决了长序列文本在后续任务中的梯度消失和爆炸的问题，对于情感分析领域来说，一个句子的情感态度是由句子前后文共同决定的，而 LSTM 只能记忆到前期的预料信息，无法获取到后面的信息，这就很难准确判断出文本所要表达的情感态度。针对这一问题，提出了双向长短期记忆神经网络(Bidirectional Long Short-Term Memory, Bi-LSTM),其基本结构如图 2.8 所示。

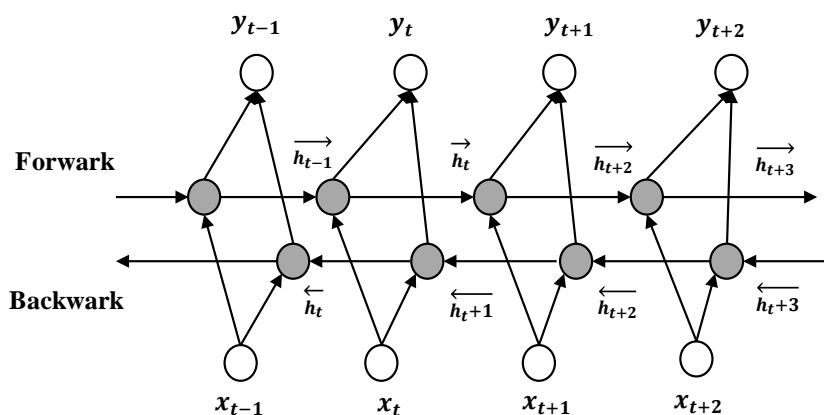


图 2.8 双向长短期记忆神经网络模型结构

Bi-LSTM 是在 LSTM 基础上进一步改进的，通过在一个 LSTM 末端在链接一个方向相反的 LSTM，信息从两端输入，正方向的 LSTM 按照正常顺序对信息处理，反方向的 LSTM 则从语料末端开始往前处理，然后将两个正序和逆序输出的特征向量进行拼接形成最终的特征表达，并且两个 LSTM 是相互独立的，这种模型对文本的特征提取率和性能明显高于单个的 LSTM 模型，但是 LSTM 和 Bi-LSTM 本质上都是循环神经网络结构，循环神经网络固有的模式就是按照顺序处理，并不能有选择性的处理信息，还是有很大的局限性。随着深度学习的发展，研究人员模拟人的思考特点提出了注意力机制，注意力机制很好的避免的这一问题，并在实际应用中取得了更好的效果。

2.5 注意力机制

现实生活中，当我们想要了解一个人的学习情况时，我们只会关注与他学习相关的事情，而忽略与学习无关的事情，例如：资产、工作等方面都会过滤掉。注意力机制就是模拟了我们处理事务的过程，注意力机制首先应用于计算机视觉领域，并在该领域取得了当时最好的成绩，Bahdanau 等人^[50]首次在自然语言处理领域引入了注意力机制，将注意力机制加入到发言语翻译中，并取得了更准确的翻译结果。注意力模型能关注重点的功能在于编码-解码器(Encoder-Decoder)框架，其基本模型结构如图 2.9 所示。

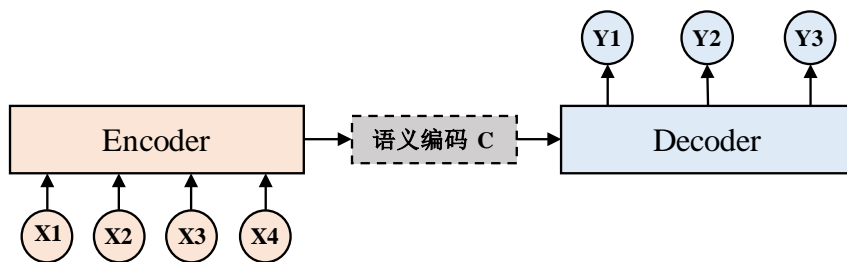


图 2.9 Encoder-Decoder 模型结构图

Encoder-Decoder 框架由两部分组成，首先将文字作为输入信息， Encoder 将句子 $S = (x_1, \dots, x_n)$ 进行非线性的语义编码，表示为相关语义 C ，语义信息是由多维的向量表示的，这些语义信息经过 Decoder 结构，Decoder 结构会将语义信息进行解码操作，最终输出目标文本向量。计算过程如式(2.15)所示：

$$C_i = \sum_{i=1}^m a_{ij} h_i \tag{2.15}$$

式中， a_{ij} 即分配权重， i 对应输出句子中的单词， j 表示输入句子中单词， h_i 为隐藏向量，是一种通用化模型。在对不同文字信息编码时，语义编码 C 都是一样的，就类似于人类看到外界的画面时，眼睛里没有焦点一样，在 Encoder-Decoder 框架中加入注意力机制可以对键值对索引。注意力机制如图 2.10 所示。

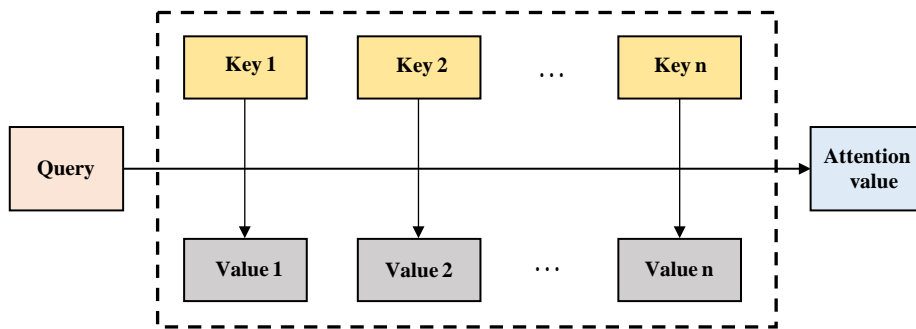


图 2.10 注意力机制

通过对序列与键值对相似度计算得到其权重，具体分为三个阶段：第一阶段：通过计算权重系数大小来判断其重要程度；第二阶段：相似度和相关性判断，最后再进行归一化；第三阶段：将权重系数与 Value 值进行加权计算。

第二过程中计算 Query 和 Key 的相似度的方法有：

- 点积

$$\text{Similarity}(\text{Query}, \text{Key}_i) = \text{Query} \cdot \text{Key}_i \tag{2.16}$$

- MLP 网络

$$\text{Similarity}(\text{Query}, \text{Key}_i) = \text{MLP}(\text{Query}, \text{Key}_i) \quad (2.17)$$

- Cosine 相似性

$$\text{Similarity}(\text{Query}, \text{Key}_i) = \frac{\text{Query} \cdot \text{Key}_i}{\|\text{Query}\| \|\text{Key}_i\|} \quad (2.18)$$

对于方面级情感分析任务来说,引入注意力机制在生成向量表示的过程中能够关注到文本的上下文信息,从而能够将方面特征信息与上下文信息充分交互,使得模型能够关注到与结果关联程度较高的部分,最终提高模型性能。

2.6 本章小结

本章对应用于文本情感分析领域的深度学习方法做出解释,首先对文本情感分析做了概述,引入最早期的词向量技术,接着分别介绍了卷积神经网络和循环神经网络,基于以上网络结构的缺点与不足,最后介绍了注意力机制。同时简单介绍了典型静态词向量模型(CBOW 词袋模型和 Skip-gram 模型)和谷歌提出来的动态词向量模型 BERT、长短期记忆神经网络等具体计算过程,为后续模型的建立奠定了坚实的理论基础。

3 SICA 模型框架介绍

方面级情感分析的子任务包括方面实体的情感分析(aspect-term sentiment analysis)和方面类别的情感分析(aspect-category sentiment analysis)。目前大多数方面级情感分析模型建模目标方面和上下文信息,再引入注意力机制,但是忽略了语法信息在文本表征中的影响,句子短语结构和词性信息等可以提高模型的性能。基于上述问题本文提出了 SICA 模型,首先介绍模型的整体框架结构,随后详细阐述模型结构的相关构建方法。

3.1 模型构建

本节主要介绍所提出的 SICA 模型。模型结构图如图 3.1 所示,主要包括两大部分,一部分是基于语法信息上下文文本序列的提取,包括词嵌入层、语法信息层、卷积层、BiLSTM 层,一部分是特定方面序列的提取。本文借鉴了 Zheng 等人^[51]提出的借助依存句法树转换成以特定方面词为根节点的依存树,首先遍历方面序列中的每一个单词,遍历过程中翻转边的方向来进行,最后经过交互注意力层将文本序列与方面序列进行交互,最后进行情感分类。

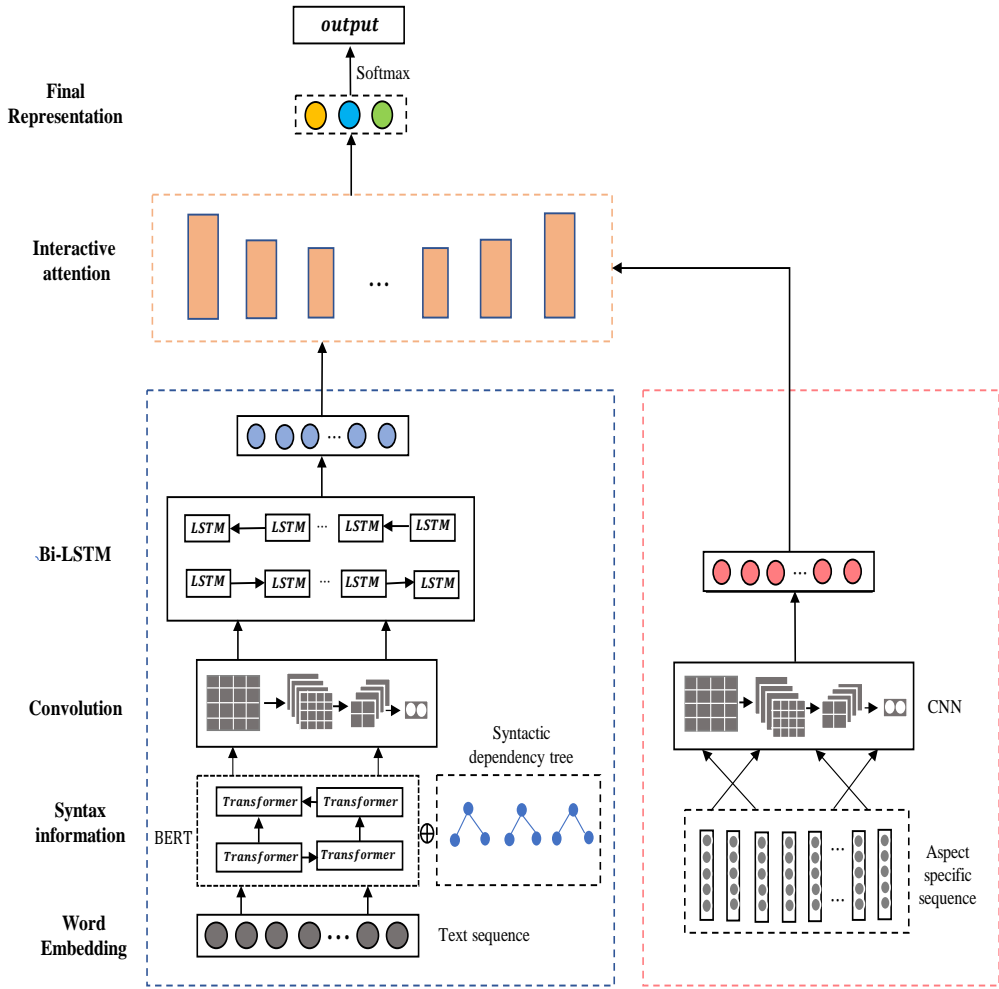


图 3.1 SICA 模型结构图

3.2 上下文文本序列提取部分

本小节主要介绍基于语法信息的上下文文本序列特征提取部分，主要包括词嵌入层、语法信息层、卷积层、BiLSTM 层，重点介绍语法信息层的构建过程，也是本文的亮点。

3.2.1 词嵌入层

本文使用 2.2.4 小节中介绍的词向量模型 BERT，该模型自提出以来，大量的研究者将 BERT 最后一层作为上下文相关词的表示输入到模型，但是这忽略了其它层的语法信息，一定程度上降低了模型预测的准确率。Jawahar 等人^[52]用实验说明了 BERT 不同层次结构的信息，低层次结构往往包含了语料的结构信息，

中层次结构主要包含着语料的语法信息，高层则主要包含了语料的语义信息。因此重点关注 BERT 模型中间层次能得到文本的语法信息，与本文方面级情感分析任务的侧重点相契合。

BERT 模型的输入向量中包含了词的位置向量、词所在句子的分段向量和当前的词语向量，具体的结构如图 3.2 所示。

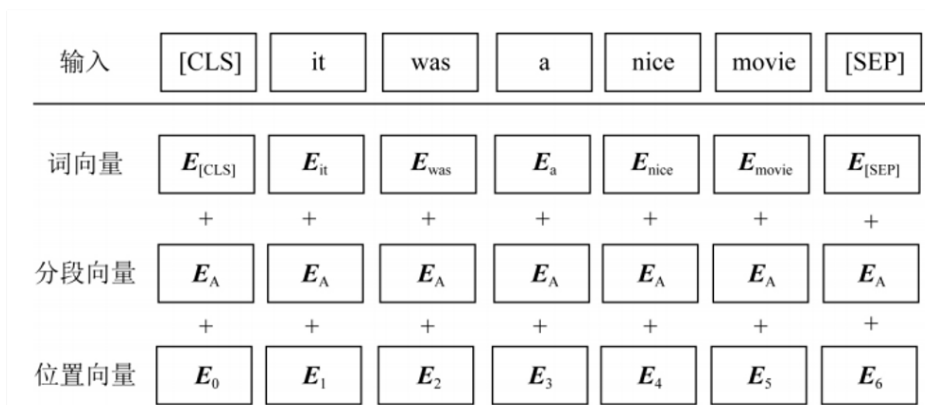


图 3.2 BERT 的词向量表示

3.2.2 语法信息层

BERT 模型的优势在于可以对输入的单词建模，建模之后的词向量的信息体现在维度上。本节提到 BERT 模型的不同层可以偏向学习的结构不同，我们将引入该模型的中低层的短语结构，借助依存句法树来引入句子的结构信息，句子结构信息可以避免长序列文本信息结构混乱的问题，同时也能帮助快速找到长序列文本中方面词与情感词对应的关系。并且在冗长的句子中词的依赖信息容易丢失，而句法信息可以使其信息完整保留，还有利于缩短方面实体和方面词之间的距离。如图 3.3 所示。

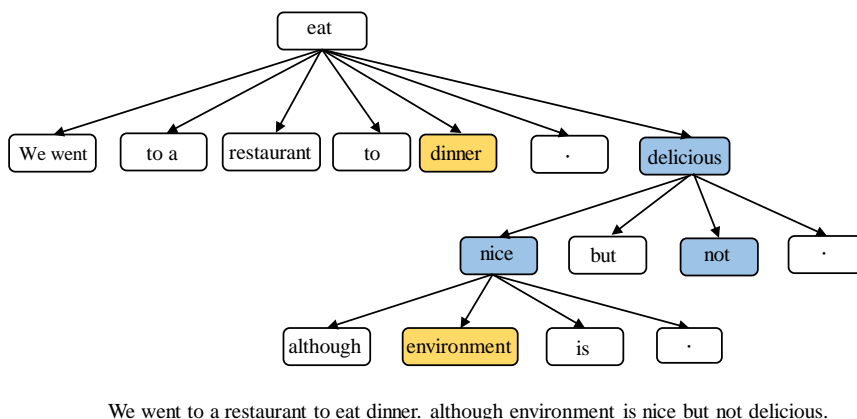


图 3.3 依存句法树示例

句子 “we went to a restaurant to eat dinner. although environment is nice but not delicious” 的依存句法树中，方面序列 “dinner” 与 “not delicious” 所表达的情感距离比 “nice” 近，而在利用注意力机制处理线性的句子时，“dinner” 与 “nice” 的距离更近，因此单纯利用注意力机制匹配方面项所对应的实体时往往会出现很大的误差，尤其是引入位置信息来进行编码时，出现的误差更大，因此借助依存句法树来处理句子的结构信息，可以有效的降低分类误差，提高模型性能。但是只是单纯的利用单词之间的依赖弧，不考虑单词之间的依赖类型，会降低单词之间的关联性，因此在本文中采用带有依赖关系的依存句法树结构，图 3.4 展示了一个句子带有依赖关系的依存树结构。

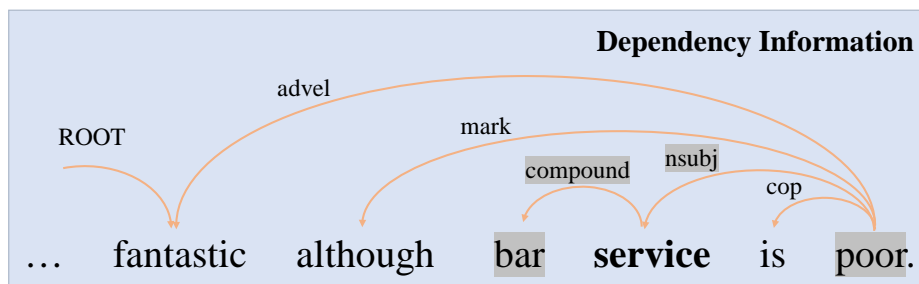


图 3.4 带有依赖关系的依存句法树结构

通过使用 Stanford parse 工具获得语句内各单词之间的依赖关系，这些依赖关系就是依存句法树的核心结构，而根节点就是改语句内的方面词和情感态度词，即语句内的重点。具体构成如下：首先从语料库中解析得到所有词与词之间的关系，形成关键集，并且将相关的依赖关系进行映射到价值集，矩阵 K 和 V 分别对应于关键集和价值集，每个向量都代表了一个键和值，在给定一个输入文本时，都可以得到其依赖性解析结果。图 3.5 表示词依赖性关系的提取过程。

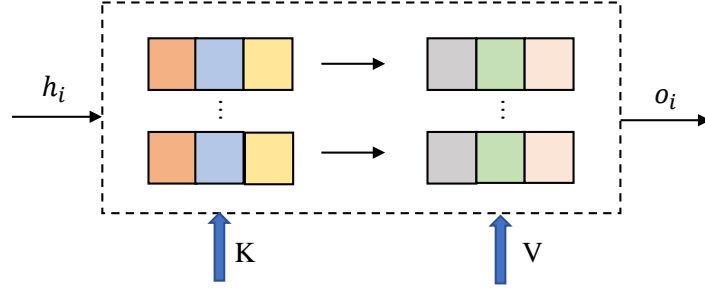


图 3.5 词依赖关系提取

将关键集矩阵 $K = (k_{i,1}, k_{i,2}, \dots, k_{i,l})$ 和价值集矩阵 $V = (v_{i,1}, v_{i,2}, \dots, v_{i,p})$ 代表的每个项目表示进行分配权重计算，分配权重是为了区分对句子的贡献程度，例如句子之间的主谓关系的修饰关系的重要程度不同，具体计算过程如式(3.1)和(3.2)所示。

$$p_{i,j} = \frac{\exp(h_i \cdot e_{i,j}^k)}{\sum_{j=1}^l \exp(h_i \cdot e_{i,j}^k)} \quad (3.1)$$

$$o_i = \sum_{j=1}^p p_{i,j} e_{i,j}^v \quad (3.2)$$

其中， h_i 表示隐藏向量， $e_{i,j}^k$ 和 $e_{i,j}^v$ 分别表示关键集和价值集中每个嵌入的项目表示， o_i 表示最后输出的带有依赖关系类型的向量表示。最后在该模型中，将 BERT 学习到的带有语法信息文本信息的邻接矩阵与带有依赖关系的依存树所处理的邻接矩阵进行拼接输入卷积层，BERT 每一层对应产生的语法信息为：

$$B^l = \{h_1^l, h_2^l, \dots, h_n^l\} \quad (3.3)$$

$$C = \{B^1, B^3, B^5, B^9\} \quad (3.4)$$

其中， $B^l \in \mathbb{R}^{n \times d_{BERT}}$ 代表着 BERT 第 l 层的输出， $h_i^l \in \mathbb{R}^{d_{BERT}}$ 代表着第 i 个单词在第 l 层的表示， C 表示各个层输出的集合。

将 C 进行归一化处理，用 D 表示带有依赖关系的句法树的度矩阵，具体的计算如式(3.5)(3.6)所示：

$$\tilde{C} = D^{-\frac{1}{2}} C D^{-\frac{1}{2}} \quad (3.5)$$

$$H = \tilde{C} * D \quad (3.6)$$

因此,语法信息层将带有 BERT 模型和带有依赖关系的依存句法树的输出表示输入卷积层进行全局信息的提取。

3.2.3 卷积层

卷积神经网络所用的权重和偏置的参数较少,具有平移不变性并能够有效地捕捉文本连续的特征表示,在文本情感分析研究领域被广泛使用,选取特定大小卷积核 $W_c \in \mathbb{R}^{d \times k}$ 进行周围滑动提取特征 c_i , 来获取含有全局语义信息的隐藏输出表示层, 获得文本序列的特征表示 $C = [c_1, c_2, \dots, c_n]$ 。

$$c_i = f(H_{ii+K} * W_c + b_c) \quad (3.7)$$

其中, f 代表一个非线性激活函数, $*$ 代表卷积操作, 使用 K 个大小相同的卷积核, 得到特征矩阵 C , $C \in \mathbb{R}^{K \times nK}$ 。

3.2.4 BiLSTM 层

双向长短期记忆网络模型能够获取含有丰富语义信息的特征向量,在文本情感分析领域,特别文本为中文时,可能会有一词多义的现象,在同一文本中相同的词在不同上下文语境中所表达的意思完全不同。为更准确得到语义信息,采用 Bi-LSTM 分别从前向和后向获取上下文信息,卷积层输出向量 $C = (c_1, c_2, \dots, c_t)$ 经过双向长短期记忆神经网络模型后输出正反方向的隐藏状态 $\overrightarrow{h^1}$ 和 $\overleftarrow{h^1}$, 数学表达形式如式(3.6)(3.7)所示:

$$\overrightarrow{h^1} = \overrightarrow{\text{LSTM}}(C) \quad \overrightarrow{h^1} \in \mathbb{R}^{n \times d_1} \quad (3.8)$$

$$\overleftarrow{h^1} = \overleftarrow{\text{LSTM}}(C) \quad \overleftarrow{h^1} \in \mathbb{R}^{n \times d_1} \quad (3.9)$$

其中, n 代表文本句子的长度, d_1 代表长短期记忆网络的隐藏状态维度。在该语言模型中,正反方向的 LSTM 通过隐藏状态预测下一个字,具体计算过程如式(3.8)(3.9):

$$p(c_{t+1} | \overrightarrow{h^1}) = \text{softmax}(\overrightarrow{W}_d \overrightarrow{h^1}) \quad (3.10)$$

$$p(c_{t+1}|\bar{h}^1) = \text{softmax}(\bar{W}_d h^1) \quad (3.11)$$

双向长短期记忆神经网络大大提升了模型对于上下文语义信息提取的性能，很好地解决了一词多义的问题，最后将正反向隐藏状态 \bar{h}^1 、 \bar{h}^1 拼接起来得到 H^1 ，作为下一层的输入向量，数学公式如式(3.10)所示：

$$H^1 = [\bar{h}^1, \bar{h}^1] = [\overline{\text{LSTM}}(C), \overline{\text{LSTM}}(C)] \quad H^1 \in R^{n \times 2d_1} \quad (3.12)$$

3.3 特定方面词提取部分

本小节主要介绍特定方面词的特征提取部分，主要包括方面词的提取过程和卷积层的简单介绍，方面词的提取最简单的方式就是根据规则来提取，例如方面词一般为名词或名词短语，而评价词一般为形容词或成语。

3.3.1 方面词的提取

方面级情感分析就是能够精确的判断特定方面的情感极性，在近年研究中受到重视，目前的大多研究模型是将句子经过词向量提取语义信息后送入神经网络模型进行情感分类从而得到结果，而后引入注意力机制对特定目标进行划分效果明显提高，只有当方面词与上下文表示真正匹配的时候，才能真正提高分类效果^[53]，因此基于方面词的特征提取得好坏影响着模型的性能。而在方面词的提取过程中，利用依存句法树作为辅助信息，句子间依存关系的表示与依存句法树的结构表示有关，根节点不同，所得到的树结构不同，而根节点又由单词所决定。本文借鉴了 zheng 等人所提出来的方式，采用单词逐一遍历的方式，并且翻转边的方向进行遍历。这使得方面序列中的第一个单词成为了根节点，将用特殊的标记来区分发生转变的边信息，以此来代表含有不同的性质。例如原本的未转换边被标记为“obj”，在转换后标记为“rev#obj”。以上操作处理得到了基于特定方面词为根节点的依存树数据。将蕴含了句子中词间相关联信息的 E_d 进行转置操作得到 E_d^T 作为最终向量表示送入卷积层。

3.3.2 卷积层

基于特定方面词的特征提取过程中所用到的卷积神经网络与小节 3.2.3 相同, 选取特定大小卷积核 $W_c \in \mathbb{R}^{d \times k}$ 进行周围滑动提取特征 c_d , 来获取含有全局语义信息的隐藏输出表示层, 获得文本序列的特征表示 $C_d = [c_1, c_2, \dots, c_{dn}]$ 。

$$c_d = f(E_{d:d+k}^T * W_c + b_c) \quad (3.13)$$

其中, f 代表一个非线性激活函数, $*$ 代表卷积操作, 使用 K 个大小相同的卷积核, 得到 $C_d = \{e_{d1}^c, e_{d2}^c, \dots, e_{dn}^c\} C \in R^{b \times m \times l}$, 表示每个单词针对于特定方面的权重值。通过对 C_d 进行最大池化及一层全连接操作得到最终权值表示 $T_{avg} = (t_1, t_2, t_3)$, 作为整个基于特定方面词特征提取的输出向量, 最后与双向长短期记忆神经网络输出的上下文特征向量 H^1 一同送入交互注意力层。

3.4 交互注意力层

为了放大上下文信息与特定方面词之间的联系, 将输入的文本序列与方面序列进行交互, 使方面词与上下文表示真正地进行匹配, 以融入语法信息的上下文信息和以特定方面词的特征进行交互学习, 交互注意力机制计算了上下文信息对方面词的权重影响, 同时又计算了方面词对上下文信息的权重影响, 设置打分机制, 用来衡量每个单词隐藏状态在句子中的权重程度, 同时设定阈值, 低于阈值的权重返回进行二次交互, 防止损失全面有关的词语信息从而影响模型性能, 具体的建模过程如下所示:

给定上下文句子信息的每个词的隐藏状态序列 $[h_c^1, h_c^2, \dots, h_c^n]$, 通过注意力机制与方面词的表示 t_{avg} 来计算上下文句子的注意力向量 α_i , 如式(3.12):

$$\alpha_i = \frac{\exp(\varphi(h_c^i, t_{avg}))}{\sum_{j=1}^n \exp(\varphi(h_c^j, t_{avg}))} \quad \alpha_i \geq 50\% \quad (3.14)$$

注意力向量 α_i 表示上下文句子中每个单词对方面词的注意力权重, 同时设定阈值范围, 当计算结果小于 50% 时, 表示上下文信息与方面词的关联匹配程度不

大, 进行二次注意力计算, 其中 φ 表示打分函数, 衡量每个单词的隐藏状态在句子中的重要程度, 打分函数的定义如式(3.13)

$$\varphi(h_c^i, t_{avg}) = \tanh(h_c^i \cdot W_a \cdot t_{avg}^T + b_a) \quad (3.15)$$

W_a 和 b_a 分别为权重矩阵和偏置向量, \tanh 为非线性函数。

同样, 给定方面词的每个词的隐藏状态序列 $[h_t^1, h_t^2, \dots, h_t^m]$, 通过注意力机制和上下文句子的表示 c_{tag} 来计算方面词的注意力向量 β_i , 如式(3.14):

$$\beta_i = \frac{\exp(\varphi(h_t^i, c_{avg}))}{\sum_{j=1}^n \exp(\varphi(h_t^j, c_{avg}))} \quad \beta_i \geq 50\% \quad (3.16)$$

和上述相同, 如果计算结果小于 50%时, 进行二次计算, 最终得到上下文和方面词的最终表示如式 (3.15) (3.16):

$$T_r = \sum_{i=1}^m \beta_i h_t^i \quad (3.17)$$

$$C_r = \sum_{i=1}^n \alpha_i h_c^i \quad (3.18)$$

3.5 情感输出层

交互注意力层的输出结果向量作为情感输出层的输入, 将上下文和方面词的最终表示向量 T_r 和 C_r 进行拼接为向量 d , 再进行非线性变换后经过 $Softmax$ 函数计算得到情感类别的概率, 具体计算过程如下所示:

$$h = \tanh(w_l \cdot d + b_l) \quad (3.19)$$

$$p_i = \text{softmax}(w_i \cdot h + b_i) \quad (3.20)$$

其中得到 p_i 值最大的即为最终的结果。

3.6 本章小结

本章主要介绍了模型的具体构建过程, 主要包括上下文特征信息的提取、基

于特定方面词特征的提取和交互注意力层、最终的情感输出层，上下文特征的提取过程包括词嵌入层、语法信息层、卷积层和 BiLSTM 层。特定方面词的提取过程包括特定方面词的提取和卷积层，最后再经过交互注意力层进行交互，输出情感分类结果。

4 实验与分析

模型性能的好坏需要在方面级情感分析任务中进行验证,本章主要从以下几个方面进行阐述:要确定模型的好坏首先要确定一个目标函数,目标函数确定以后确定评价指标,介绍实验数据、确定实验环境、整体模型实验参数设置,模型对比实验设置,最后再进行结果分析。

4.1 目标函数与评价指标

4.1.1 目标函数

交叉熵损失函数是一种可以详细地描述两个概率分布的相似程度,用于多分类问题,后期可以不断地调整参数来使得预测值与真实输出值相近,因此本文模型采用带有L2正则化的交叉熵函数作为损失函数,具体的计算方式如式(4.1)(4.2)所示:

$$Loss = -\frac{1}{N} \sum_{i=1}^N y_{ik} \log p[k] + \lambda \|\theta\| \quad (4.1)$$

$$y_{ik} = \begin{cases} 1 \\ 0 \end{cases} \quad (4.2)$$

公式(4.1)中 N 代表输入样本的数量, i 为对应的索引值, y_{ik} 代表训练数据的真实标签值, k 可以代表二分类或三分类,本文中取 $k = 3$, λ 代表正则化权重系数,使用随机梯度下降中的 Adam 算法,将学习率设置为 0.0001。

4.1.2 评价指标

评价指标是为了更好地反映出评价对象的特性和相互联系之间的关联程度,在方面级情感分析任务中用准确率 Accuracy 和 $F1$ 值来反映模型的性能好坏,其中准确率 Accuracy 代表模型预测正确数量所占总量的比例。精准率反映了预测结果中正确的样本数量,而召回率反映了样本被预测正确的数量,在某一情况下,两者会出现矛盾,而 $F1$ 值能够解决这一问题, $F1$ 值是两者的调和平均数。

$$Acc = \frac{TP + TN}{N} \quad (4.3)$$

$$F1 = \frac{2TP}{N + TP - TN} \quad (4.4)$$

其中式子中 TP 代表本来预测为 1，实际为 1，预测正确，既本来在数据集中是正样本，被预测为正样本， TN 代表预测为 0，实际为 0，预测正确，就是本来数据集中是负样本，被预测为负样本， N 代表为数据集样本的总数。

4.2 实验数据及实验环境介绍

4.2.1 实验数据

表 4.1 数据集的测试与样本划分

数据集	Postive		Neutral		Negtive	
	Train	Test	Train	Test	Train	Test
Laptop	968	341	530	178	808	169
Restaurant	2164	748	807	196	637	196
Twitter	1568	175	3127	346	1565	175

社交媒体的迅猛发展带动了电商平台的发展，电商平台的发展产生了海量的评论数据，常用于方面级情感分析任务的公开数据集有 Sem Eval2014-2016 数据集，本章实验采用的数据集为 Sem Eval2014¹和 Twitter²公开数据集，Sem Eval2014 数据集包括餐厅评论(Restaurant)和笔记本电脑(Laptop)评论数据，其数据集包括三个标签，分别为正向(Postive)、负向(Negtive)和中性(Neutral)，如表 4.1 所示 Laptop 数据集中的正向样本训练集包含 968 条，测试集中包含 421 条，负向样本训练集包含 808 条，测试集包含 169 条，中性样本训练集包含 520 条，测试集包含 178 条，Twitter 数据集的数据分布比较均衡，三种数据的比例分布分别为 25%、

¹ 数据来源: <https://github.com/Noahs-ARK/semEval-2014>

² 数据来源: <https://github.com/IsaacChanghau/DL-NLP-Readings>

25%和 50%，数据集的分布如表 4.1 所示。

4.2.2 实验环境

本章所用的实验环境如下表 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	Ubuntu20.04.3LTS
CPU	Intel® Xeon(R) Silver 4208 CPU @ 2.10GHz× 32
显卡	Tesla P100 PCIe 16GB
内存	7.0TB
编程语言	Python 3.8.3
深度学习框架	PyTorch
开发平台环境	Anaconda

4.3 实验设置

为了让模型达到最优的预测效果，要对参数进行设置，为了验证模型的有效性，需要与其他方面级情感分析模型进行对比实验，最后再进行消融实验对模型的各个模块进行分析，详细讨论各个模块对整体模型性能的影响程度，因此本节从以下三个方面进行介绍，实验参数设置、对比实验设置和消融实验设置。

4.3.1 实验参数设置

在实验参数的设置过程中，上下文语法信息特征提取和特定方面词特征提取的词向量均使用 BERT 词向量进行初始化，词向量训练过程中会出现过拟合现象，本实验过程中使用 BERT-Base 模型，是 Google 的一种开源模型，其有 12 个 Transformer 编码层，12 个多头注意力，设置维度为 768，BERT 模型的学习率设置为 2×10^{-5} ，并且通过标签值来对 BERT 预训练语言模型进行微调，设置 LSTM 的隐藏状态维度为 300，CNN 的卷积核为 2、3，依存句法树用 Stanford

Parser 开源工具生成, 每次输入数据的批量大小 Batch-size 为 32, 其余学习率设置为 0.0001, 设置随机失活率 dropout 的值为 0.5, 正则化系数设置为 0.0001, 优化算法使用梯度下降算法的 Adam 函数, 训练轮回次数设置为 50。

4.3.2 对比实验设置

为了对所提出的方面级情感分析模型的性能进行对比评估, 将与以下模型进行对比:

- **LSTM**: 该模型将文本的词向量作为输入, 经过 LSTM 模型处理后得到隐藏层句子向量表示, 然后将向量表示进行全连接处理, 最后得到情感分类。
- **BiLSTM**: 基于双向长短期记忆神经网络模型与上述过程一致, 模型使用了双向的 LSTM 模型来获取前后向的隐藏层向量。
- **TD-LSTM^[54]**: 该模型使用两个 LSTM, 一个从前到后输入左侧文本和方面词, 另一个从后到前输入右侧文本和方面词; 然后, 作者将两个 LSTM 最后一个时间步的隐态向量拼接起来, 送入 softmax 进行分类。
- **ATAE-LSTM^[55]**: 该模型利用方面嵌入和词嵌入的串联作为 LSTM 层的输入, 然后添加一个公共注意层以获得与方面有关的表示, 提高模型对方面词的权重。
- **IAN^[56]**: 该模型提出了交互注意力机制, 将方面词和上下文信息单独建模, 并且其中都加入注意力机制, 对方面词和上下文信息之间进行交互, 能够充分的提取特征, 匹配程度进一步提高, 这样最终得到的方面词和上下文向量表示融合了二者的交互信息。
- **RAM^[57]**: 该模型首先使用将预训练得到的向量输入到一个双向 LSTM 中, 得到每个隐含层的输出, 之后用多重注意对每个隐藏层赋予权重, 最后将注意力结果用 GRU 非线性结合输出, 以预测目标情绪。
- **MemNet^[58]**: 该模型用递归单元和门控单元组成一个深层记忆网络, 用以储存放下文和方面词的位置信息, 接着融入注意力机制对上下文信息分配权重,

加权求和后将其输入到情感分类层进行情感极性分析。

- ASGCN^[59]: 该模型在句法依存树上建立图卷积神经网络, 对方面词和语义信息同时建立表达形式, 之后用注意力机制进行二者的交互, 用以捕获相关句法约束和远距离单词依存关系。
- AEGCN^[60]: 该模型提出了注意力增强图卷积神经, 采用了多头注意力对 GCN 进行改进, 更好地结合语义和语法信息, 同时, 增加了一个关注机制提高其性能。

4.3.3 消融实验设置

设置消融实验的目的是验证语法信息提取模块, 特定方面词提取模块, 交互注意力模块三部分分别对整体模型性能的影响程度, 本实验过程中具体将整体模块与各部分模块进行了对比, 实验结果如表 4.4 所示。在整个模块中, 文本嵌入层模块、双向长短期记忆神经网络模块、卷积神经网络模块和情感输出模块不参与消融实验对比。特别说明, 对比消融实验中语法信息提取模块只考虑是否加入了带有依赖关系的依存句法树结构, 用 SICA-pos 表示整体模块去除带有依赖关系的依存句法树之后的模块; 对于特定方面词模块的消融对比实验设置为去除整个模块, 即上下文信息特征的提取和方面词的提取同时提取, 用 SICA-A 表示整体模块去除特定方面词特征的提取部分以外的模块; 对于交互注意力模块的消融对比实验设置为三部分: 用 ATT-t/c 交互注意力模块, 第一, SICA-ATT-t 表示交互注意力机制仅使用了方面对上下文的交互, 第二, SICA-ATT-c 表示交互注意力机制仅使用了上下文对方面的交互, 第三, SICA-ATT-t/c 表示整个去除交互注意力模块的剩余模块模型。

4.4 实验结果分析

为了验证模型的性能, 本节对模型的结果进行分析, 分为两部分: 第一是通过和其他模型的准确率和 F1 值的对比分析, 消融实验结果来验证来证实模型各部分模块的有效性。第二是进行实例分析, 通过对 Restaurant 数据集中的两个句

子的示例作为案例进行分析，证明语法信息在方面级情感分析过程中的贡献。

4.4.1 模型预测结果分析

SICA 模型与其他模型对比的实验结果如表 4.3 所示。

表 4.3 各对比模型的准确率和 F1 值

Models	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
LSTM	0.6650	0.6123	0.6735	0.6498	0.6289	0.6753
BiLSTM	0.6462	0.6190	0.6603	0.6589	0.6345	0.6709
TD-LSTM	0.6480	0.6357	0.6750	0.6602	0.6498	0.6770
ATAE-LSTM	0.6865	0.6509	0.6935	0.7054	0.6987	0.7086
IAN	0.7076	0.6588	0.7180	0.6985	0.7065	0.7109
RAM	0.7123	0.6696	0.7308	0.7091	0.7124	0.7098
MemNet	0.6832	0.6432	0.6575	0.6403	0.6576	0.6434
ASGCN	0.7454	0.6563	0.7314	0.7280	0.7408	0.6908
AEGCN	0.7473	0.7078	0.7456	0.7482	0.7698	0.7093
SICA*	0.7621	0.6593	0.7490	0.7637	0.7569	0.7109

从表 4.3 中可以看出，本文提出的模型在 3 个领域的数据集上的准确率和 F1 值分别为 0.7621、0.7490、0.7569 和 0.6593、0.7637、0.7109，结果均优于其他各对比模型，实验结果证明，在建模过程中使用语法信息能够提升模型的文本表征能力，从而提升性能，进一步证实了模型的有效性。

在各对比模型中，LSTM 模型的准确率和 F1 值最低，造成这一结果的原因是模型只是粗略地将文本的上下文信息和方面词信息输入到模型中，没有做进一步的区分，提取的信息相对较少，BiLSTM 模型的实验结果高于 LSTM 模型，双向 LSTM 从正向和反向同时提取文本向量信息，更加全面地提取特征信息，减少信息缺失，提高模型分类效果。TD-LSTM 模型使用两个 LSTM 神经网络，提取

了方面词的左上文信息和右下文信息,使得信息的提取更为集中,实验结果的准确率优于 BiLSTM,ATAE-LSTM 模型的准确率和 F1 值与前面几个模型相比较,准确率比 TD-LSTM 提升了 2%–3%左右。因为在模型中添加了一个公共注意层以获得与方面有关的表示,提高了模型对方面词的权重。在基于注意力机制的模型中,IAN 模型和 RAM 模型都引入了注意力机制,其实验结果都超过了前面几种模型,表明了注意力机制会加强关注与方面词有交互的词, RAM 模型的性能表现优于 IAN 模型,因为 RAM 模型使用多重注意力机制与循环神经网络结合,更能有效地捕捉方面词与上下文之间的关系。MemNet 模型中没有采用传统的 seq2seq 模型,但是因为 Mem Net 的模型参数较少,且没有对语义嵌入层的隐向量进行非线性建模,最后输出的结果本质上只是词嵌入的线性组合,因此它的输出结果效果不好。

ASGCN 模型的实验结果明显有所提升,其利用依存句法树建立图卷积神经网络,对方面词和语义信息同时建立表达形式,之后用注意力机制进行二者的交互,用以捕获相关句法约束和远距离单词依存关系,但是没有利用句子中的语法信息,AEGCN 模型的实验结果在 Restaurant 和 Twitter 数据集上的表现比本文提出的 SICA 模型的结果稍有提高,造成这一结果的原因是 SICA 模型受噪声影响,而 AEGCN 模型在处理噪声信息和依赖树性能局限方面具有优越性,但是缺少了上下文信息与方面词的交互过程。本文提出的 SICA 模型,准确率和 F1 值均优于 ASGCN 模型,在 3 个数据集上准确率较 ASGCN 模型分别提高 2.24%、2.41%、2.17%,进一步证实了模型的有效性。

SICA 模型的消融实验结果如表 4.4 所示。

表 4.4 消融实验的准确率和 F1 值

Models	Restaurant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
SICA-pos	0.7490	0.6438	0.7265	0.7508	0.7350	0.7045
SICA-A	0.7546	0.6358	0.7356	0.7567	0.7407	0.6993
SICA-ATT-t	0.7589	0.6431	0.7361	0.7424	0.7406	0.7094
SICA-ATT-c	0.7578	0.6453	0.7349	0.7460	0.7420	0.7024
SICA-ATT-t/c	0.7370	0.6327	0.7281	0.7319	0.7319	0.6904
SICA*	0.7621	0.6593	0.7490	0.7637	0.7569	0.7109

由表 4.4 所示可以看出，当去掉带有依赖关系的依存句法树部分后，语法信息层的信息来源就是完全由 BERT 模型来提供，虽然句子中仍然包括丰富的语义信息，但是带有依赖关系的依存句法树可以增强词与词之间联系，因此去掉该部分后，得到的实验结果明显比完整模型训练得到的结果低，这表明了带有依赖关系的依存句法树的重要性。

当模型去掉基于特定方面词的特征提取部分后，上下文信息特征的提取和方面词特征的提取同时进行，实验结果证明模型的准确率和 F1 值都有所降低，表明了基于特定方面词的特征提取模块的重要性。

当交互注意力模块中仅使用单一方面的注意力时，实验结果证明无论是使用方面对上下文的交互还是上下文对方面的交互，准确率和 F1 值的大概变化都基本相同，但是相对于原本模型结果还是有一定的影响，当整个去除掉交互注意力模块时，准确率和 F1 值明显下降，上下文信息和方面词的正确匹配能够提高模型的分类效果，其实验表明了交互注意力层对模型的重要性。

4.4.2 实例分析

为了验证本文所提出的模型在处理复杂情感表达方面的性能，明确了解语法

信息在模型中所起的作用，本文对 Restaurant 数据集中的两个句子作为实例进行案例分析，用可视化的方式进行绘制热图，直观地展示有无语法信息对模型的影响，绘制的热图如图 4.1 所示。图中的颜色越深，说明注意力的权重越大，模型对它的关注就越大。

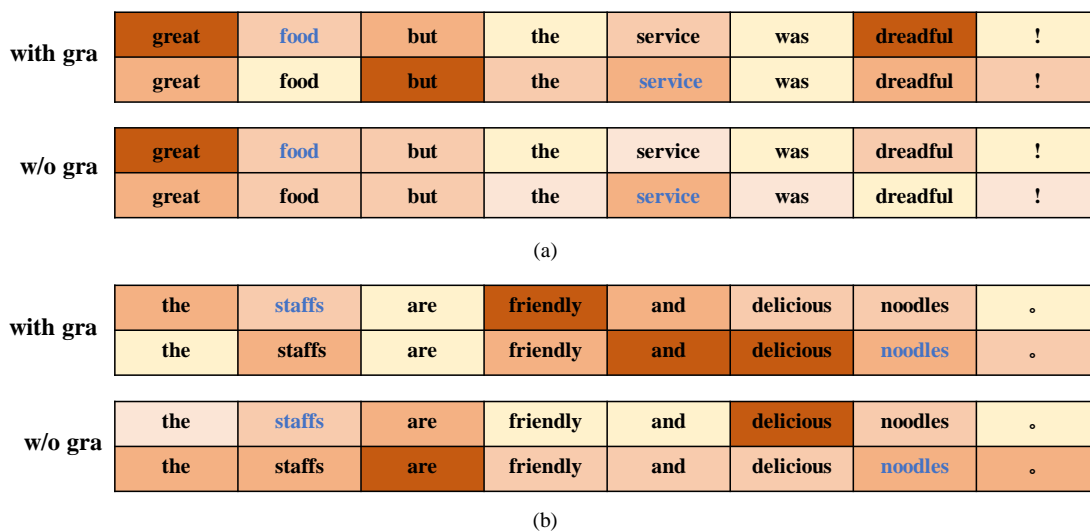


图 4.1 具有用语法信息和不具有语法信息模型的注意力权重的图示

图中的(a)和(b)表示 Restaurant 数据集中的两个示例，(a)中句子“great food but the service was dreadful”中包含两个方面实体 food, service, (b)中句子“the staffs are friendly and delicious noodles”中包含两个方面实体 staffs, noodles, 从第一示例可以看出, without gra 模型分别孤立地通过关注“great”一词来预测“food”的情感极性, 通过关注单词“dreadful”一词去预测方面“service”的情感极性, 没有具体分析句子之间的语法信息, 忽略了两个方面实体之间的联系, 而对于 with gra 模型除了关注情感词“great”以外, 句子中含有转折词“but”, 表示方面实体“food”和“service”具有相反的情感极性。第二个示例中可以看出, without gra 模型中通过关注单词“delicious”来预测方面实体的情感极性, 而 with gra 模型中通过关注“friendly”和“delicious”来预测方面实体的情感极性, 加入语法信息后, 形容词一般修饰名词, “friendly”修饰名词“staffs”, 单词“and”表示前后的意思相近, 如果“staffs”方面实体的情感极性判断为正向, 那么“delicious”修饰“noodles”方面实体的情感极性一定为正向。

从例子中可以看出, 具有语法信息的方面级情感分析模型可以捕捉到词与词

之间的联系关系类型，在处理复杂语境方面更具有优势。

4.5 本章小结

本章主要介绍了模型性能验证的完整过程，首先介绍了模型的目标函数和评价指标，分别为准确率和 F1 值、实验数据以及实验环境，为了对模型性能进行验证，重点介绍了实验的设置过程，包括实验参数的设置，对比实验的设置和消融实验的设置，最后进行实例分析。本文提出的模型在 3 个数据集上进行测试，最后结果表现优于其他模型，证明了其有效性。

5 总结与展望

5.1 总结

随着互联网的迅猛发展，互互联网已经深入到人们的日常生活之中，各大电商平台的崛起导致人们从在社交媒体上自由发表言论，因此分析评论观点是当下自媒体时代必不可少的一部分，方面级情感分析任务是自然语言处理的任务之一，是一种细粒度的文本情感分析，更能帮助社会舆情监控，帮助用户分析产品价格和服务。

本文针对方面级情感分析任务中神经网络模型忽略句子的语法信息，上下文与方面词特征提取不完全的问题，无法利用不同方面的意见，上下文信息与方面词无法正确匹配等问题，提出了基于语法信息的交互式方面级情感分析模型，文章中简称为 SICA 模型，模型的主要内容分为两大部分：第一，上下文信息的提取部分，第二，特定方面词的提取部分，最后将带有语法信息的上下文文本特征与基于特定方面词特征拼接入交互注意力层，第一部分的上下文信息提取部分主要包括词嵌入层、语法信息提取层、卷积层和双向长短期记忆神经网络，本文中上下文信息与方面词同使用 BERT 预处理语言模型，语法信息层是以 BERT 模型的中低层作为主要信息来源，以带有依赖关系的依存句法树作为辅助信息，构成语法信息层，卷积层用来获取全局的语义信息，在情感分析任务中，尤其在中文领域容易出现一词多义的情况，双向长短期记忆神经网络含有丰富的语义信息，能够有效地捕捉到一词多义的情况，而单向的长短期记忆神经网络无法处理。特定方面词的特征提取部分，利用依存句法树作为辅助提取方面词，不同方面的方面词不能很好地利用，用卷积神经网络提取特征。交互注意力模块是将上下文特征与方面词特征进行交互，同时设定阈值，交互权重过小时，可进行二次交互，能够全面地利用文本信息，最后，通过情感输出层输出结果，以此判断文本的情感极性。

本文为了验证模型的性能，在 3 个公开数据集 Restaurant、Laptop 和 Twitter 上进行实验，进行了基线实验对比，实验结果有所提升，证明语法信息在句子中起重要作用，再对模型进行消融实验，分别对模型的语法信息层、交互注意力层

和特定方面词的特征提取三部分进行消融对比实验,交互注意力层分为三小部分进行对比,整体去除和只单一使用方面对上下文的交互或上下文对方面的交互,结果证明这三部分模块对模型性能的提升起重要作用,最后再进行实例分析,通过对 Restaurant 数据集上的两个例子进行分析,发现带有语法信息的方面级情感分析模型的性能更好。

5.2 展望

本文提出的 SICA 模型性能在方面级情感分析任务中取得了不错的效果,在 3 个数据集上的预测结果比基线模型的稍有提高,但是对于含有大量噪声信息的文本数据, SICA 模型的预测准确率提升幅度有限。SICA 模型中使用了依存句法树结构,但是依存句法树在生成过程中所出现的错误问题依旧没有得到有效解决,将会使得模型在进行方面词与上下文文本信息匹配过程中考虑到不相关的信息,影响最终的分类效果,而且在进行方面词与上下文信息交互过程中,由于二次交互使得算法的运算时间和复杂度上升。未来的工作过程中,将继续优化算法过程以及继续研究提取过程中的信息匹配问题,以进一步提升模型预测。

参考文献

- [1] 中国互联网络信息中心 CNNIC.第 48 次《中国互联网络发展状况统计报告》[R].北京:中国互联网络信息中心 CNNIC,2021: 1-2.
- [2] Liu B. Sentiment analysis and opinion mining[J]. Synthesis lectures on human language technologies, 2012, 5: 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. 2003: 519-528.
- [4] 赵品龙. 基于深度学习的不同粒度下的文本情感分析研究[D].天津大学,2020.
- [5] Chen Z, Mukherjee A, Liu B. Aspect extraction with automated prior knowledge learning[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2014: 347-358.
- [6] Poria S, Cambria E, Gelbukh A. Aspect extraction for opinion mining with a deep convolutional neural network[J]. Knowledge-Based Systems, 2016, 108: 42-49.
- [7] Thet T T, Na J C, Khoo C S G. Aspect-based sentiment analysis of movie reviews on discussion boards[J]. Journal of information science, 2010, 36(6): 823-848.
- [8] Chen W T, Lin S C, Huang S L, et al. E-HowNet and automatic construction of a lexical ontology[C]//Coling 2010: Demonstrations. 2010: 45-48.
- [9] 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.
- [10] Baccianella S, Esuli A, Sebastiani F. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining[C]//Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). 2010.

- [11] 栗雨晴, 礼欣, 韩煦, 宋丹丹, 廖乐健. 基于双语词典的微博多类情感分析方法[J]. 电子学报, 2016, 44(09): 2068-2073.
- [12] 李祥祥, 石刚. 基于深度学习的双语词典构建方法研究[J]. 现代计机, 2021(21): 8-14.
- [13] 李枫林, 范雅娴. 领域情感词典构建方法研究[J]. 图书馆理论与实践, 2019(12): 60-65+112. DOI: 10.14064/j.cnki.issn1005-8214.2019.12.012.
- [14] 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.
- [15] 吴杰胜. 基于多部情感词典和深度学习的中文微博情感分析研究[D]. 安徽理工大学, 2020.
- [16] Huang Z, Zhao Z, Liu Q, et al. An unsupervised method for short-text sentiment analysis based on analysis of massive data[C]//International Conference of Young Computer Scientists, Engineers and Educators. Springer, Berlin, Heidelberg, 2015: 169-176.
- [17] Turney P D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[J]. arXiv preprint cs/0212032, 2002.
- [18] Zhang P Y. Word similarity computation based on WordNet and HowNet[C]//Applied Mechanics and Materials. Trans Tech Publications Ltd, 2013, 336: 2115-2118.
- [19] 孙艳, 周学广, 付伟. 基于主题情感混合模型的无监督文本情感分析[J]. 北京大学学报(自然科学版), 2013, 49(01): 102-108.
- [20] Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques[J]. arXiv preprint cs/0205070, 2002.
- [21] Kiritchenko S, Zhu X, Cherry C, et al. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews[C]//Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). 2014: 437-442.

- [22] Shubham D, Mithil P, Shobharani M, et al. Aspect level sentiment analysis using machine learning[C]//IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2017, 263(4): 042009.
- [23] Wang Y, Chen Q, Shen J, et al. Aspect-level sentiment analysis based on gradual machine learning[J]. Knowledge-Based Systems, 2021, 212: 106509.
- [24] Kamal A, Abulaish M. Statistical features identification for sentiment analysis using machine learning techniques[C]//2013 International Symposium on Computational and Business Intelligence. IEEE, 2013: 178-181.
- [25] 孙翌博, 濮泽堃, 徐玉华, 胡冰. 应用 SD-LS-SVM 算法的评论情感分析模型[J]. 软件导刊, 2021, 20(04): 43-48.
- [26] 黄晨晨, 索朗拉姆, 拉姆卓嘎, 群诺. 基于 SVM 的藏文微博文本情感分析研究与实现[J]. 高原科学研究, 2020, 4(01): 92-96.
- [27] Liu L, Feng S, Wang D, et al. An empirical study on Chinese microblog stance detection using supervised and semi-supervised machine learning methods[M]//Natural Language Understanding and Intelligent Applications. Springer, Cham, 2016: 753-765.
- [28] 王婷, 杨文忠. 文本情感分析方法研究综述[J]. 计算机工程与应用, 2021, 57(12): 11-24.
- [29] Bengio Y, Ducharme R, Vincent P. A neural probabilistic language model[J]. Advances in Neural Information Processing Systems, 2000, 13.
- [30] Johnson R, Zhang T. Effective use of word order for text categorization with convolutional neural networks[J]. arXiv preprint arXiv:1412.1058, 2014.
- [31] Chen Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.
- [32] 王煜涵, 张春云, 赵宝林, 袁肖明, 耿蕾蕾, 崔超然. 卷积神经网络下的 Twitter 文本情感分析[J]. 数据采集与处理, 2018, 33(05): 921-927. DOI:10.16337/j.1004-9037.2018.05.017.
- [33] 苏灵松, 应捷, 杨海马, 肖昊琪. 双通道卷积记忆神经网络文本情感分析[J]. 软件导刊, 2019, 18(07): 32-36.

- [34] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces[C]//Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. 2012: 1201-1211.
- [35] Zhang Y, Jiang Y, Tong Y. Study of sentiment classification for Chinese microblog based on recurrent neural network[J]. Chinese Journal of Electronics, 2016, 25(4): 601-607.
- [36] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 conference on empirical methods in natural language processing. 2016: 606-615.
- [37] Hu B, Liu Y, Sun M, et al. An Extended Attention-based LSTM with Knowledge Embedding for Aspect-level Sentiment Analysis[J]. Journal of Computers, 2019, 30(4): 176-184.
- [38] Yin Y, Song Y, Zhang M. Document-level multi-aspect sentiment classification as machine comprehension[C]//Proceedings of the 2017 conference on empirical methods in natural language processing. 2017: 2044-2054.
- [39] Ma D, Li S, Zhang X, et al. Interactive attention networks for aspect-level sentiment classification[J]. arXiv preprint arXiv:1709.00893, 2017.
- [40] Hazarika D, Poria S, Vij P, et al. Modeling inter-aspect dependencies for a 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 (Short Papers). 2018: 266-270.
- [41] 张克, 张文俊, 朱蕴文, 邢毅雪. 基于内联关系的方面级情感分析方法[J/OL]. 上海大学学报(自然科学版):1-12[2022-03-03]. <http://kns.cnki.net/kcms/detail/31.1718.N.20210604.1116.008.html>.
- [42] 汪兰兰, 姚春龙, 李旭, 于晓强. 结合依存句法分析与交互注意力机制的隐式方面提取[J]. 计算机应用研究, 2022, 39(01): 37-42. 1001-3695.2021.06.0249.
- [43] 卜令梅, 陈黎, 卢永美, 于中华. 利用关系抽取技术联合识别文本中的方面-极性对[J]. 四川大学学报(自然科学版), 2022, 59(01): 43-50.

- [44] Ye X, Xu Y, Luo M. ALBERTC-CNN Based Aspect Level Sentiment Analysis[J]. IEEE Access, 2021, 9: 94748-94755.
- [45] 薛芳,过弋,李智强,王家辉.基于双层词性感知和多头交互注意机制的方面级情感分析[J/OL].计算机应用研究:1-8[2022-03-01].
- [46] 夏鸿斌,李强,刘渊.局部与全局特征融合的方面情感分析网络模型[J/OL].计算机科学与探索:1-13[2022-03-01].<http://kns.cnki.net/kcms/detail/11.5602.tp.20211014.2239.016.html>.
- [47] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [48] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [49] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural networks, 2015, 61: 85-117.
- [50] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [51] Zheng Y, Zhang R, Mensah S, et al. Replicate, walk, and stop on syntax: an effective neural network model for aspect-level sentiment classification[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(05): 9685-9692.
- [52] Jawahar G, Sagot B, Seddah D. What does BERT learn about the structure of language?[C]//ACL 2019-57th Annual Meeting of the Association for Computational Linguistics. 2019.
- [53] Dong L, Wei F, Tan C, et al. Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification[C]. meeting of the association for computational linguistics, 2014: 49-54.
- [54] D Tang, Qin B, Feng X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[J]. Computer Science, 2015.

- [55] Wang Ye-quan, Huang Min-lie, Zhu Xiao-yan, et al. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
- [56] Ma De-hong, Li Su-jian, Zhang Xian-dong, et al. Interactive attention networks for aspect-level sentiment classification[C]//26th International Joint Conference on Artificial Intelligence, 2017.
- [57] Tang Du-yu, Qin Bing, Liu Ting. Aspect level sentiment classification with deep memory network[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016:214-224.
- [58] Zhang Zhong-lin, Li Lin-chuan, Zhu Xiang-qi, et al. Aspect sentiment analysis combining ON-LSTM and self-attention mechanism[J]. Journal of Chinese Computer Systems, 2020, 41(9):1839-1844.
- [59] Nguyen Huy-Thanh, Nguyen Minh-Le. Effective attention networks for aspect-level sentiment classification[C]//10th International Conference on Knowledge and Systems Engineering(KSE), 2018:25-30.
- [60] Xu G, Liu P, Zhu Z, et al. Attention-enhanced graph convolutional networks for aspect-based sentiment classification with multi-head attention [J]. Applied Sciences, 2021, 11 (8): 3640.

致 谢

相见时难别亦难，行笔至此，在 496 号度过了人生中很美好的 7 年，依稀记得 18 岁第一次踏进校园时的青涩，到 25 岁时离开校园步入社会，在这里见证了太多成长的时刻，大四在图书馆备考时的场景依稀在目，下午 5 点的阳光和 10 点图书馆的铃声永远定格在记忆中，始于 2015 年秋天，终于 2022 年的盛夏，总觉得来日方长，殊不知人生是减法，好在我在这里遇到的都是很温暖的人。

树高千尺不忘根深沃土，研究生 3 年中我最感谢的是我的导师李强教授，是一位温暖又有趣的人，感谢他三年对我生活中的照顾、学业上的帮助和人生路上的开导，教导我应该要有开阔的眼界和独立的思想，学生只能以简单的言语以表谢意，同时感谢学院各位老师的关心与照顾，感谢电子商务实验室的各位老师的帮助与关心！

寸草之心，难报三春之晖，父母在人生尚有来处，感谢父母二十多年的养育之恩，生于农村，父母给予了他们能力范围内最好的条件给我，在他们面前永远是个小朋友，一路支持我走到现在，把最好的永远留给我，只希望你们身体健康，未来为你们遮风挡雨！感谢我的哥哥嫂嫂，长兄如父，长嫂如母，像疼爱自己孩子一样，包容和支持。

山水一程，三生有幸，感谢 520 的小伙伴们，两个非常优秀的师兄翟亮、何多魁，在学习上帮我解惑，帅气的师弟黄建民和赵金雨，办事可靠踏实，还有刘伟同学的三年革命友情，美丽又开朗的师妹王雪绒和李聪聪，带给我很多快乐和温暖，虽然我年长于他们，但是在生活中更照顾我，感谢我的舍友张杉杉、李彤、宋文琴，三年相处过程愉快又温馨，感谢遇到这么合拍又有趣的你们，感谢我认识十多年的挚友陈莉，见证彼此的成长，感谢丁宁，相识于年少，多亏六年一路的包容与体谅，相识于此，实属幸运，愿此去之路，少年永存！

二十余年寒窗经历渐进尾声，如大梦苏醒，亦似劫后重生，一路波折忐忑，许多不易，功不唐捐，玉汝之成，最终与自己和解。

最后，感谢强大的祖国在新冠疫情肆虐中给予我们的坚强后盾，愿疫情早日散去，山河无恙，国泰民安！

攻读硕士期间发表的论文及科研情况

- [1] 李强, 王文瑞等, Aspect-level Emotion Analysis based on BERT Fusion Multimodule[C].ConferencePublishingServices.DOI10.1109/MLBDBI54094.2021.00055.
- [2] 项目名称: 基于大数据的个性化自适应学习模式研究, 项目来源: 校级一般项目, 项目编号: Lzufe2018C-12, 主要参加者.
- [3] 专利: 李强, 王文瑞等, 一种机械壁式沙漠植树机及系统, 专利号: ZL202120025078.4, 授权日期: 2021年1月6日.
- [4] 著作出版: 《信息系统基础(第八版)》(中国人民大学出版社), ISBN:978-7-300-29974-7, 参与翻译其中第九章.