

同濟大學

TONGJI UNIVERSITY

毕业设计（论文）

课题名称	基于深度学习技术的细粒度情感分析研究
副标题	
学 院	电子信息与工程学院
专 业	数据科学与大数据技术
学生姓名	郑海东
学 号	1953174
指导教师	苗夺谦
日 期	2023 年 6 月 7 日

基于深度学习技术的细粒度情感分析研究

摘要

细粒度情感分析是自然语言处理中的热点任务，与句子级情感分析相比，细粒度情感分析旨在分析句子中不同对象的情感极性。细粒度情感分析任务可以分为情感对象的抽取和情感极性识别两个子任务。细粒度情感分析到现在已经有了大量的相关研究工作，从早期的机器学习单模型到后期的各类深度学习模型，从早期的人工建立词典映射到后期具有高泛用性的预训练模型。其中不同的模型有其优越性也有其不足之处，本文的工作就是探讨研究几种不同的思路方法在细粒度情感分析任务中的表现，并提出一些改进思路和方法。主要包括：

1. 对细粒度情感分析的早期方法进行了有效性探讨和实践，将细粒度情感分析任务划分为方面词抽取和方面词分类两个子任务，并结合机器学习模型，包括随机向量场，支持向量机等训练预测，并对比其他经典模型效果进行讨论。
2. 针对机器学习单模型中比较困难的方面词分类问题，使用了卷积神经网络等的深度学习方法进行处理，并使用门控单元完成方面词的筛选识别，实现将两个子任务集成到一个网络中进行，并取得了良好的分类效果。
3. 针对卷积神经网络对于上下文提取过于依赖句子结构的问题，提出门控单元进行信号筛选，配合图卷积神经网络进行特征抽象和分类的思路，并介绍了语义依赖图的提取和表示方法，最后将得到的情感分类结果与卷积神经网络等网络的结果相比较，验证了语义依赖图的结构在文本分类中的良好抽象能力，并讨论了门控单元的选择问题和解释性问题。

关键词：自然语言处理，机器学习，深度学习，细粒度情感分析，图卷积网络

ABSTRACT

Aspect-Based Sentiment Analysis is a classic problem in natural language processing. From the perspective of sentiment objects, its granularity is primarily reflected in the sentiment polarity classification of each aspect within a sentence. The essence of this task lies in identifying the sentiment objects in a sentence and determining their corresponding semantically relevant context. There has been a significant body of research conducted on Aspect-Based Sentiment Analysis, encompassing diverse approaches ranging from early machine learning single-model techniques to advanced deep learning models. Additionally, the field has witnessed the evolution from manual dictionary mapping to the utilization of highly versatile pre-trained models. Each model has its own strengths and limitations. This paper aims to delve into Aspect-Based Sentiment Analysis by exploring and proposing several distinct approaches. Furthermore, improvement ideas and methods will be presented. The main contents of this paper include the following:

1. We focus on efficient exploration and practical implementation of early methods for Aspect Based Sentiment Analysis. The task is divided into two subtasks: aspect term extraction and aspect term classification. Machine learning models, specifically random vector fields and support vector machines, are employed for training and prediction. The performance of these models is compared and discussed in comparison to other classical models.
2. in single model machine learning, deep learning methods such as convolutional neural networks are employed for processing. Gated units are utilized to filter and identify aspect terms, integrating the two subtasks into a unified network and achieving superior classification performance.
3. Considering the problem of convolutional neural networks relying heavily on sentence structure for context extraction, graph convolutional neural networks with gated units used for signal selection are proposed. They are combined for feature abstraction and classification. The extraction and representation methods of semantic dependency graphs are introduced. Finally, the sentiment classification results are compared with convolutional neural networks, demonstrating the strong abstraction ability of the structural semantic dependency graph in text classification. Besides, we discuss how to choose the gate units and explain the choices.

Key words: Natural Language Processing, Machine Learning, Deep Learning, Aspect Based Sentiment Analysis, Graph Convolutional Networks

目 录

1 绪论	1
1.1 研究背景及意义	1
1.2 研究现状和问题	2
1.3 本文的主要内容	3
1.3.1 基于机器学习方法的 ABSA 任务	3
1.3.2 基于门控卷积神经网络的细粒度情感分析	3
1.3.3 基于门控图卷积神经网络的细粒度情感分析	4
1.4 本文的组织结构	4
2 基于传统机器学习方法的细粒度情感分析	5
2.1 引言	5
2.2 基于 CRF 模型的方面项提取任务	5
2.2.1 任务概述	5
2.2.2 特征工程	5
2.2.3 随机向量场模型原理概述	7
2.2.4 关于不同词向量方法的讨论	7
2.3 基于机器学习模型的方面项分类任务	9
2.3.1 任务概述	9
2.3.2 上下文依赖关系的表示	9
2.3.3 基于机器学习分类器的文本分类任务	10
2.4 实验	10
2.4.1 实验数据	10
2.4.2 实验方案	12
2.4.3 评价指标	13
2.4.4 实验结果及分析	14
2.5 本章小结	15
3 基于门控卷积神经网络的细粒度情感分析	16
3.1 引言	16
3.2 卷积神经网络概述	16
3.3 基于门控机制的关键词捕捉	17
3.4 引入门控机制的卷积神经网络	17
3.5 实验	19
3.5.1 实验数据	19
3.5.2 实验设计	19

3.5.3 评价指标	19
3.5.4 实验结果及分析	20
3.6 本章小结	20
4 基于门控图卷积神经网络的细粒度情感分析	21
4.1 图卷积神经网络概述	21
4.2 依赖图的获取	22
4.3 引入门控机制的图卷积神经网络	23
4.4 实验	24
4.4.1 实验数据	24
4.4.2 实验设计	25
4.4.3 评价指标	25
4.4.4 实验结果及分析	25
4.5 本章小结	26
5 总结与未来工作展望	28
5.1 本文工作总结	28
5.2 不足与展望	28
参考文献	29
谢辞	31

1 绪论

1.1 研究背景及意义

作为一种探索用户意见的研究，网络平台上的用户生成评论可以作为商业智能的数据来源。事实上，许多服务销售行业非常在意平台与游客之间的互动交流。诸如国内的美团，京东等平台，国外的 Amazon，都有产品的留言区和讨论社区等。通过分析用户的互动信息，企业确定其产品或服务的改进领域。通过了解客户对特定方面的看法，公司可以优先考虑增强功能，改进功能，并优化他们的产品以满足客户的期望。这有助于产品/服务创新、客户满意度和竞争优势。因此，合理利用互联网信息交互的特点，挖掘出文本语言之中的信息反馈可以帮助不同的人群，组织进行更高效的交流。

情感分析是文本分析的一个重要分支，主要目标是提取文本的情感特征，并基于这些特征进行分类。情感分析根据研究对象的粒度不同可分为文本级、句子级、方面级等，分别对相应单位的文本进行情感倾向分析。细粒度的情感分析一般指方面级情感分析（Aspect-based Sentiment Analysis, ABSA），下面举一个例子来说明情感分析的不同粒度。比如一个顾客在餐馆用餐后评论，“The pizza is good but its price is high.” 这段文本是比较符合客观评价的一个样例，我们不难发现，首先它并没有非常明显的情感倾向，而且对于已评论的两个方面，pizza 味道好（正面），价格贵（负面），如图 1.1 所示 也就是针对不同的方面的评价可能不一样。这样的评论文本很有研究意义，因为它更客观更

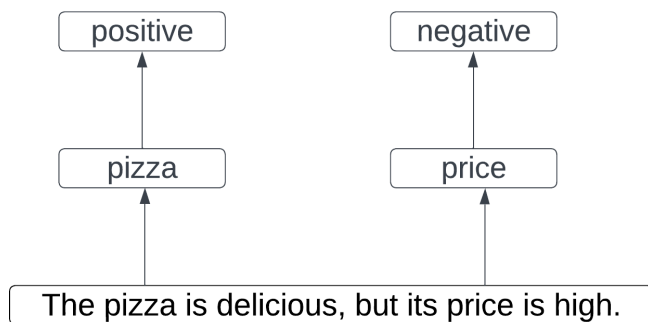


图 1.1 细粒度情感分析举例

精准。用此时的视角看待一个句子时，不再讨论或者分析这个人到底是处于一个正面或者负面的情绪，而是只针对其中的方面项方面项（下文统称方面项）来讨论其中的情感色彩。这种分析的视角的意义在于它绕开了评论者和旁观者的不同立场带来的主观色彩。比如在上面的例子中，对于一个有钱人来说，可能他会认为饭菜的美味比价格更重要，所以他会认为这条评论是在推荐这家餐厅；反之对于一个没有钱的人可能会觉得价格贵是难以接受的，因此认为这家餐厅是不被推荐的。所以当我们把目光放在更加细粒度的视角上，就不再去考虑整个文本的情感倾向。上述的场景就是一个典型的基于方面项的情感分析任务，也就是细粒度情感分析的主要研究问题。

细粒度情感分析（Aspect-Based Sentiment Analysis），是自然语言处理（Natural language processing）中是自然语言处理领域中一个热门的子任务。“粗粒度”的文档级、句子级情感分析通常为整个评论或句子分配一个唯一的情感极性标签，在实际生活中，这种方式是不合理的。因为事实上，人们在发表评论的时候，通常不是只从其中一个方面去评价，另外不同的人往往也用不同的角度去看待评论。要抓住不同的角度和不同的评价，需要同时把握事物的属性和评价，才能较大程度上还原评价的内容。

对于企业来说，其可以从这些评论数据中提取有价值的见解，以了解客户的意见，改进产品和服务，并做出明智的决策。ABSA 通过将情感与特定方面联系起来，可以对用户意见进行更详细的分析。例如，在产品评论中，方面可以包括“电池寿命”、“屏幕质量”或“客服态度”等方面。研究 ABSA 问题的目的就是 from 文本中自动识别和分类这些方面从而更加精准地了解顾客的诉求和企业的问题。

正因为 ABSA 问题场景的普遍存在，使其具有了巨大的商业价值和学术价值，一经提出，就一直是自然语言处理领域的热点，与其相关的工作更是引人入胜。从提出到今天经过大概十年的研究探索，已经累积了不少的学术经验和研究贡献。

1.2 研究现状和问题

方面级情感分析任务最早由 Liu 等人提出^[1]。在早期的研究方法主要依靠词典或者规则定义的方式对情感进行极性判断，此类方法的鲁棒性较差，泛化能力也较低。在 2014 年，伴随着 SemEval2014 竞赛的开展，ABSA 问题成为了自然语言处理领域的热门话题之一。在该竞赛上，主办方将 ABSA 问题拆分成方面项抽取（task2）和方面项分类（task4）两个子任务，这个划分思路在此后的各类机器学习方法和深度学习方法都被广泛使用。在此基础上，不同的机器学习方法被应用到 ABSA 任务中，比如 Mubarak 等人^[2]采用了朴素贝叶斯方法，Joachims 等人^[3]以及 Mullen 等人^[4]采用了基于支持向量机的方法。

从单分类器延伸到深度学习网络的过程是迅速的，基于神经网络的 ABSA，如 Dong 等人^[5]利用 RNN 来提取文本的特征，展现出了其在捕捉上下文关系的良好性能。Wang 等人^[6]提出了通过 LSTM 的经典架构来分别处理方面项的筛选问题和语义抽象的功能，并采用了 Attention 机制对上下文关系特征进行增强，在 ABSA 任务中取得了良好的效果。Duyu^[7]等人提出了基于位置编码提取方面项位置信息的方法进一步完善了注意力机制在感知方面项位置信息的功能。但是 Mulyo 等人^[8]认为基于循环神经网络的文本特征抽象结构在时间消耗是比较大的，在此基础上，Xue 等人^[9]尝试使用 CNN 来替代 LSTM 提取上下文关系，也取得了良好的分类结果。Chen 等人^[10]则提出了利用位置编码结合 GCN 来提取上下文关系，相较 CNN，GCN 能捕捉到非连续的上下文依赖关系。Li 等人^[11]在语义依赖树的提取方面做出改进，在表达文本同时考虑语法结构和语义依赖，让模型对上下文的语义关系表达上更加准确而丰富。

另一方面，模型预训练也一直是 ABSA 任务的重要步骤，其主要目标是将文本语言转化为向量

形式，而其中的编码方式也要尽可能多地提取出文本中的单词本身特征，语义上下文特征。预训练方法从早期的词典编码到 Onehot 法 Word2Vec，GloVe^[12] 等，到 Peters 等人^[13] 提出的 ELMo 嵌入捕获不同语言上下文中的单词含义，从而提高了包括 ABSA 在内的各种 NLP 任务的性能。而 Devlin 等人^[14] 于 2018 年推出的 BERT，通过引入一种基于 Transformer 的架构，在大量未标记文本上预训练深度双向语言模型，彻底改变了预训练领域，近两年大量基于 BERT 的研究，如 Li 等人的端到端 BERT 模型^[15] 和 Xiao 等人的 BERT4GCN 模型^[16]，都表现了预训练模型在情感分析任务上的强大性能和重要地位。

此后开展了更多基于深度学习的 ABSA 任务研究，从而扩展出了 ABSA 的很多新的研究方向，如弱监督的 ABSA 任务^[17]，多语种 ABSA 任务^[18]，多模态的 ABSA 模型^{[19][20]}，语义关系提取^[21] 的研究，基准数据集的建立和评估等^[22]。总的来说，ABSA 任务是自然语言处理中的热点问题，围绕其可以衍生出很多新的课题，这也是本文研究的目标和意义。

1.3 本文的主要内容

本文的主要工作分为以下三点：

1.3.1 基于机器学习方法的 ABSA 任务

首先受到早期 ABSA 任务思路的启发，将该任务划分为方面项抽取和方面项分类两个子任务，探讨并完成了基于机器学习方法的细粒度情感分析问题，主要步骤包括数据处理，特征工程，分类器训练，结果分析等。将该任务划分为方面项抽取和方面项分类两个子任务后，前者的主要目标就是给定一个句子文本，找到其中的方面项的坐标，或偏移值；后者的主要目标就是基于抽取基础的短句子情感分析。结果分析发现该思路在方面项抽取的任务上取得了比较好的效果，但是在方面项分类任务上效果却并不令人满意，而机器学习方法的实验步骤繁琐，容易受到操作者主观经验影响等缺点也确实让实验者感到困难重重。该工作为后续工作思路的转变和模型构建提供了经验和启示。

1.3.2 基于门控卷积神经网络的细粒度情感分析

本文尝试了一种带门控单元的 CNN 模型来解决细粒度情感分析问题，主要步骤包括数据处理，模型搭建，神经网络训练，结果分析等。CNN 处理文本分类工作时具有可并行，效率高等特点，所以在该工作中选取其为主要模型，采用门控单元来进行文本和方面项相关性的计算和判断，以完成对输出信号的筛选和过滤。实验表明，该模型在细粒度情感分析任务常用的几个数据集上都表现出了较高的准确率，具有良好的分类处理效果，验证了 CNN 结构在处理文本特征时的强大计算能力和抽象能力。

1.3.3 基于门控图卷积神经网络的细粒度情感分析

本文提出了门控单元结合图卷积层的模型对于细粒度情感分析问题的方法，主要步骤包括数据处理，模型搭建，神经网络训练，结果分析等。GCN 处理文本分类工作时采用语义依赖图来表达句子中不同成分之间的关系，相比于基于 CNN 网络的结构，GCN 能表示的信息量更大，能更精确地刻画自然语言的语义关系，所以在本章工作中选取其为主要模型。而采用门控单元来进行文本和方面项相关性的计算和判断，以完成对输出信号的筛选和过滤，这在上一个工作中被证明是有效的，固仍然采用门控单元为重要组成成分的卷积层结构，只改变卷积方式和预处理模块。实验表明，该模型在细粒度情感分析任务常用的几个数据集上都表现出了较高的准确率，具有良好的分类处理效果，并且在多项指标上相较门控 CNN 模型有了一定的提升，验证了 GCN 结构在抽象文本特征时更精确完整的优越特性。此外，本章探讨了门控单元的选择问题对细粒度情感分析问题的影响，并结合实验进行了猜想的验证，最后在实验后对门控单元的表现讨论了其可解释性。

1.4 本文的组织结构

本文的组织结构如下：

1. 绪论。其中简单介绍细粒度情感分析工作的主要任务和场景，应用和意义。并罗列了相关的研究情况，并对本文的思路和结构进行了概述。
2. 基于机器学习方法的 ABSA 任务。其中对 ABSA 的早期方法进行了有效性探讨和实践，将 ABSA 任务划分为方面项抽取和方面项分类两个子任务，并结合机器学习模型进行训练预测，并对比经典模型效果进行讨论。
3. 基于门控卷积神经网络的细粒度情感分析。其中针对机器学习中比较困难的方面项分类问题，使用了卷积神经网络进行处理，并使用门控单元完成方面项的筛选识别，将两个子任务集成到一个网络中进行，并取得了良好的分类效果。
4. 基于门控图卷积神经网络的细粒度情感分析。其中提出门控单元进行信号筛选，配合图卷积神经网络进行特征抽象和分类，并介绍了语义依赖图的提取和表示方法，最后将分类结果与卷积神经网络等比较，验证了语义依赖图的结构在文本分类中的良好抽象能力，并讨论了门控单元的选择问题和解释性问题。

2 基于传统机器学习方法的细粒度情感分析

2.1 引言

本部分任务主要的工作可以划分为方面项的抽取任务和方面项的分类任务。其中方面项抽取指从一个文本中抽取出其中的方面项列表；而方面项分类任务则指对应方面项列表中的每一个方面项，找出每一个方面项对应的情感标签。之所以这样划分任务是因为机器学习的模型在分类问题上具有较高的性能，这样两步完成的思路让模型的选择上变得简单，而且我们可以直观对比各种模型之间的性能（比如我们的模型如果整体效果比较差，我们可以直观地找出该模型在具体哪一个分类任务上效果比较差）。事实上，在 SemEval14 的竞赛上，主办方在设计题目时候，将方面项提取作为四个小子任务的第二项，而方面项分类作为第四项，也启发众多参赛者将这个问题作为两个子任务来一一完成。

2.2 基于 CRF 模型的方面项提取任务

2.2.1 任务概述

由上文，我们现在讨论 ABSA 的第一个子任务，即方面项抽取任务，这个任务中的输入为一个文本数据，输出为句子对应的方面项的列表。在本章中我们基于 CRF 模型，通过特征的构造和选择，以方面项在句子中的偏移量为标签，来实现方面项的预测。

2.2.2 特征工程

特征工程是指从原始数据中选择、转换和创建新特征或变量，以提高机器学习模型性能的过程。在机器学习程序中，特征工程的作用至关重要，因为它可以显著影响模型的准确性和可靠性。以下是特征工程在机器学习中的主要作用：

A. 选择相关特征

特征工程涉及从可用数据集中选择最相关的特征，这些特征可以对模型的结果产生显著影响。这有助于消除噪声和不相关的数据，从而实现更准确和有效的模型。

B. 提高模型性能

通过创建新特征来捕捉数据中的基本模式和关系，特征工程可以显著提高机器学习模型的性能。这可以帮助提高准确性、减少错误和提高整体效率。

C. 处理缺失数据

特征工程技术可以帮助处理缺失数据，例如插补值、删除不相关的特征或使用高级技术（例如插值或机器学习算法）来预测缺失值。

D. 减少过拟合

特征工程还可以帮助减少过拟合，当模型过于复杂并学习数据中的噪声而不是基本模式时就会发生过拟合。通过选择最相关的特征，特征工程可以帮助预防过拟合，提高模型的泛化能力。

其中本文主要提取了比较重要的几类特征：

A. 偏移量

也就是单词在句子中的位置。单词出现在句子的前中后不同的位置可能会影响到它在句子中的作用，就像我们通常把宾语放在主语后面的例子一样。

B. 单词开头是否大写

考虑当前词是否为开头大写，因为大写开头的单词通常以专有名词以及特指词居多，这些词作为情感表达对象的几率较高。

C. 语义关系

语义依赖关系是一个被广泛使用的文本语义分析方法，它将一个句子构造成一棵依存句法树，依存句法树中每个节点对应句子中的一个单词，其中，根节点为句子的中心词；分支节点的词与其叶子节点的词之间有直接的依赖关系，分别被称为支配词与被支配词。举例来说，现有一个英文句

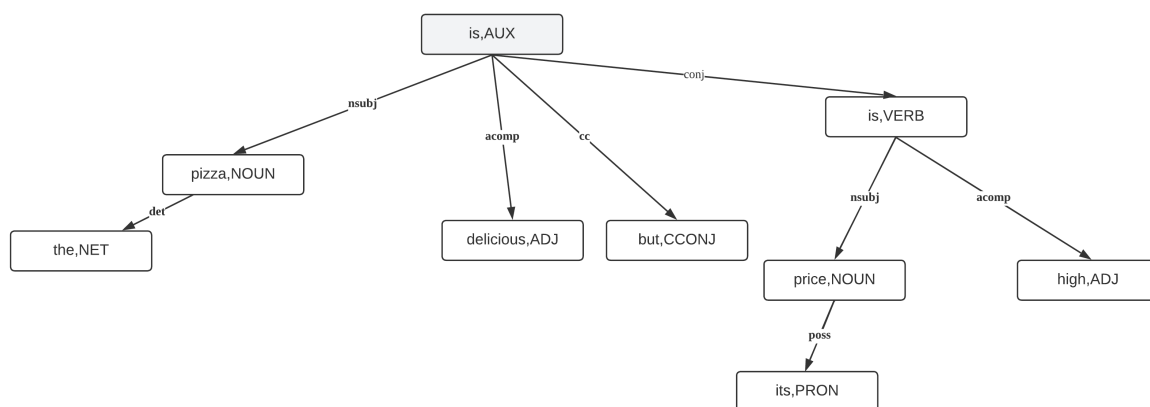


图 2.1 语义依赖图

子 “The pizza is delicious but its price is high.”，如图 2.1 所示的依存句法依赖树。其中句子的中心词（根节点）为 “is”，我们可以发现，每个单词之间的关系是复杂的，而且单词并不是只有或者一定和其相邻的单词存在语义上的联系，要准确统一地分析这种关系非常复杂，这里考虑几种常见的单词间联系，如表 2.1 所示。

表 2.1 关系标签种类

关系	关系描述
nsubj	动词与主语之间
dobj	动词与直接宾语之间
amod	形容词与被修饰词之间
advmod	副词与被修饰词之间
conj	两个并列关系之间
ccomp	动词与其补语之间
acomp	形容词与其补语之间

表例本文提取了与位置相关的几组比较简单而直观的依存句法特征，分别为：amod（修饰关

系), nsubj (主语与谓语), dobj (谓语与宾语)。

D. 词性

由于情感表达对象多为名词等特定词性的单词, 而形容词可能是表达情感态度的重要单词, 因此单词词性是很重要的情感特征之一。

E. 词向量的聚类特征

使用词向量扩展训练数据量的方法是一种常见的解决训练集数据规模过小的策略。词向量的训练方法包括 word2vec 方法, 对应 skip-gram 和 CBOW 两种策略; 以及 Glove 方法。这两种方法的原理和对比讨论将放在下文进行。词向量的训练将文本语言转化为更容易被计算机处理的数字信号。但是考虑到词向量的维度通常会比较大 (通常是几百维的向量来表示一个单词), 在机器学习方法中模型训练的开销是巨大的, 因此本文将两种方法得到的词向量分别聚类, 得到两个聚类模型和单词在聚类模型中的分类特征, 作为两个位置特征加入 CRF。

2.2.3 随机向量场模型原理概述

CRF 模型 (Conditional Random Fields, 条件随机场) 是一种用于序列标注的概率图模型。具有非常良好的预测性能和可解释性。它可以用于解决许多自然语言处理任务, 例如命名实体识别、词性标注、语义角色标注等。CRF 模型的原理是基于最大熵模型和隐马尔可夫模型的思想, 它假设序列中每个位置的标签 (例如命名实体类型、词性等) 都是相互依赖的, 即当前位置的标签不仅仅取决于当前位置的特征, 还取决于整个序列的上下文信息。CRF 模型使用了特征函数来表示每个位置的特征, 它计算当前位置标签和序列上下文的条件概率。CRF 模型的目标是最大化训练数据的对数似然函数, 通过学习特征函数的权重来最大化训练数据的似然概率。CRF 模型的训练使用了随机梯度下降等优化算法, 它在训练过程中不断地更新特征函数的权重, 使模型能够更好地捕捉数据中的模式和关系。在测试时, CRF 模型使用动态规划算法来计算每个位置的最优标签序列, 从而对输入序列进行标注。

2.2.4 关于不同词向量方法的讨论

A. Word2Vec

Word2vec 是一些用来产生词向量的训练模型。在这之前最简单的单词转向量模型是 one-hot 法, 但是 one-hot 法得到的向量太过于稀疏了, 而且词与词之间没有任何逻辑关系。而更高级的词向量要尽可能保留原始语义中的语义相关性和语义信息。例如, 有两个词 A 和 B, 他们的前后文的非常接近, 那么他们训练得到的向量应该也是非常接近的, 比如 "he" 和 "she" 的前后文就可能非常相似, 表示的向量也应该非常接近。所以说这样得到的向量会考虑进单词与单词之间的关系。Word2vec 模型的一个优势在于, 在相应的训练完成之后, 其模型可用来映射每个词到一个向量, 可用来表示词对词之间的关系。

CBOW (Continuous Bag-of-Words Model) 和 Skip-gram (Continuous Skip-gram Model), 是 Word2vec

的两种训练模式。skip-gram 方法是 Word2Vec 模型的一个基本组成部分，旨在通过预测给定目标词的上下文词来学习词嵌入。其基本思想是，出现在相似语境中的单词可能具有相似的含义。通过训练神经网络来预测上下文词，skip-gram 模型可以为词生成密集的、低维的向量表示。在 skip-gram 方法中，训练数据由单词对组成：一个目标单词和一个上下文单词。目标词是从文本中选择的，上下文词是出现在其附近的词。上下文窗口可以定义为目标词前后固定数量的单词。

skip-gram 模型采用了一个神经网络架构，包括输入层、隐藏层和输出层。输入层表示目标词，输出层表示预测的上下文词。隐藏层作为中间层，将输入的单词转换为有意义的表示。在训练过程中，skip-gram 模型通过优化神经网络的参数进行学习，使正确预测上下文词的概率最大化。这是通过使用 softmax 函数将输出层的激活转换为概率来实现的。该模型通过反向传播和梯度下降来调整其参数，旨在最大限度地减少预测概率与实际上下文词之间的损失。把单词表示成向量之后，skip-gram 就构建了这么一个语言模型，一个单词出现在一个句子中，那么这个单词就是和它的上下文有联系的，因此，我们可以通过一个单词以及语料库，推导它最可能的上下文，比如我们有单词“是”，同时我们的语料库有“今天”、“开心”、“星期一”，如果我们要推断“是”的前后文，很容易就能知道答案是：“今天是星期一”。

而 CBOW 方法是 Word2Vec 模型的另一个重要组成部分，用于从大型文本语料库中学习词嵌入。与 skip-gram 方法类似，CBOW 的目的是通过根据周围的上下文词预测目标词来捕获词之间的语义关系。与 skip-gram 相比，CBOW 方法采用的方法略有不同。CBOW 不是根据给定的目标词预测上下文词，而是根据上下文预测目标词。CBOW 背后的直觉是，一个单词的意思可以从它周围的单词中推断出来。CBOW 的训练数据由上下文窗口及其对应的目标词组成。上下文窗口是一个固定大小的围绕目标单词的单词窗口。例如，如果上下文窗口大小为 3，则 CBOW 将考虑目标单词前后的三个单词作为上下文。CBOW 采用了一个包含输入层、隐藏层和输出层的神经网络架构。输入层表示上下文词，这些词被映射到它们各自的词嵌入。嵌入通常在训练期间与 CBOW 模型一起进行预训练或学习。隐藏层作为中间层，处理上下文词嵌入并将其转换为有意义的表示。隐藏层中神经元的数量决定了隐藏表示的大小。输出层表示预测的目标词。它使用 softmax 函数将隐藏层的激活转换为概率。每个输出神经元对应于词汇表中的一个特定单词，概率反映了每个单词成为目标单词的可能性。在训练过程中，CBOW 模型调整其参数，以最大限度地提高在给定上下文词的情况下预测正确目标词的概率。这是通过最小化预测概率和实际目标单词之间的交叉熵损失来实现的。

总的来说，Word2Vec 中的 CBOW 方法与 skip-gram 方法一起，对词嵌入的发展做出了重大贡献，彻底改变了自然语言处理领域，使机器能够更好地理解和处理人类语言。

B. GloVe

GloVe (Global Vectors for Word Representation) 模型是一种思路创新的词嵌入技术，用于学习词的密集向量表示。GloVe 与 Word2Vec 等其他词嵌入模型的不同之处在于，它将全局共现统计与基于本地上下文窗口的方法相结合，以捕获词义。GloVe 模型背后的关键思想是利用大量文本语料库中单词的共现统计。GloVe 认为词义可以从单词在不同语境中同时出现的频率模式中推断出来。它

旨在学习捕获这些共现模式的词向量，并可以编码词与词之间的语义关系。GloVe 的训练过程是基于语料库中的词之间的共现统计构造一个词-词共现矩阵。矩阵的每个元素表示两个单词在指定上下文窗口内同时出现的次数。上下文窗口可以定义为目标单词前后固定数量的单词。然后 GloVe 对这个共现矩阵应用分解方法，试图将其分解为两个低秩的因子矩阵。一个因子矩阵的行向量表示目标词的词向量，另一个因子矩阵的列向量表示上下文词的词向量。GloVe 的训练目标是最小化加权最小二乘损失函数，该函数捕获词向量及其共现统计量之间的关系。优化过程迭代调整词向量，寻求提高共现统计与词向量之间的拟合。从 GloVe 中获得的结果词向量是密集的、低维的表示，可以捕获全局和局部的词义。GloVe 的一个显著优势是它能够整合全局统计数据，捕捉单词之间的长期依赖关系。这与 Word2Vec 等方法不同，后者仅依赖于本地上下文窗口。此外，GloVe 可以有效地处理大规模文本语料库，并在各种自然语言处理任务中表现出良好的性能。

2.3 基于机器学习模型的方面项分类任务

2.3.1 任务概述

定位了情感对象的位置后，下一步任务就是对情感对象进行“情感分析”了。事实上，这个任务的本质和文本级文本情感分析没有太大差异。不同点在于，在文本级任务中，你是对整篇文本做特征向量表示，而对情感对象做特征表示时，你应该考虑的是那些“与当前情感对象有直接关联的文本”。根据文本某一位置的情感对象，我提取了两种文本：窗口上下文和语义上下文，基于此上下文作文本分类工作。

2.3.2 上下文依赖关系的表示

A. 窗口上下文

比较经典的上下文提取方式，比如句子 $[W_1, W_2, \dots, W_n]$ 中， W_4 为情感对象，当窗口大小为 3 的时候，如图 2.2 所示，其窗口上下文为 $[W_1, W_2, W_3, W_4, W_5, W_6, W_7]$ 。

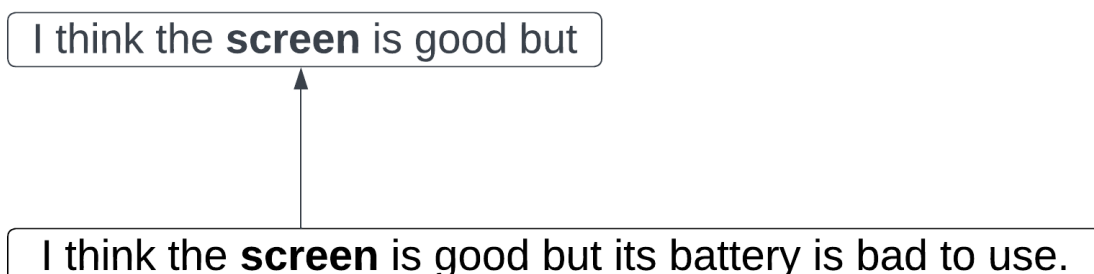


图 2.2 上下文关系表示

B. 语义依赖上下文

窗口上下文有个很大的问题，那就是对情感对象表达了观点的词极可能在位置上离情感对象很远，那么这里我利用了前面提到的语义依赖分析，对 W_4 的窗口上下文，提取了其在依赖树上有关系的单词。通过句法分析，screen 这个单词和 the 与 good 具备依存关系，那么我们把它们提取出来做相关文本。但对 screen 做依存词抽取肯定是不够的，这里我对 I, think, the, screen, is, good, but 这 7 个单词都抽取了依赖词作为相关文本。

2.3.3 基于机器学习分类器的文本分类任务

基于上下文的依赖关系作为标签，情感分类作为特征，我们在本文中探讨了几种常见分类器的分类效果。

A. LR

LR (Logistic 回归) 模型是机器学习和统计学中广泛使用的监督学习算法。逻辑回归模型是一个线性模型，尽管它最早用于二分类任务，其目标是预测实例属于两个类之一的概率，但它也可以扩展到处理多类分类问题。LR 模型在 NLP 任务上的优势是其具有良好的可解释性，体现在其可以检查 LR 模型中学习到的权重，以了解不同特征或单词对分类决策的重要性的贡献。这种可解释性使使用者能够深入了解模型的决策过程，并有助于解释模型的预测，可以扩展到处理多类分类问题。

B. SVM

SVM(支持向量机) 模型是一种强大的监督学习算法，用于机器学习中的分类和回归任务。其主要思想是找到一个最优的超平面，该超平面能最好地分离特征空间中不同类别的数据点。支持向量机在解决复杂 NLP 分类问题时特别有效，不同类之间有明显的分离。它们可以通过将输入特征映射到高维特征空间来处理线性可分和非线性可分数据。支持向量机因其强大的理论基础和处理高维数据的能力而受到广泛欢迎。

2.4 实验

2.4.1 实验数据

A. Rest14

Rest14 数据集，也被称为 restaurant -2014 数据集，是一个在餐厅领域广泛使用的 ABSA 基准数据集。它是为 SemEval-2014 的基于方面的情感分析任务创建的，SemEval-2014 是一个关于语义评估的研讨会。它包含了顾客对餐馆各个方面的评论和意见，比如食物、服务、氛围和价格。数据集来自多个来源的在线评论集合，包括 Yelp、TripAdvisor 和 OpenTable 等网站。这些评论是由人类注释者手动注释方面和情感极性的。Rest14 数据集提供方面级别的注释，这意味着评审中的方面或实体被标识和标记。每个方面对应于一个特定的方面类别，如“食物”、“服务”、“环境”、“价格”等。注释还包括与每个方面相关联的情感极性标签（积极、消极、中性）。Rest14 数据集由大量数据组成，具体的数据分布统计数据见所示 其中每条数据会标明其中方面项 在句子中出现的位置，和其对应的情感极性标签。后面介绍的数据集的数据结构都与此结构类似或完全相同，后续不再赘述。我们

表 2.2 数据集的统计分布情况

数据集	Pos	Neg	Neu
Rest14	2892	1001	829
Lap14	1328	994	629
Rest15	1238	408	70
Rest16	1695	551	95
Twitter	1560	1712	3420

的工作目标就是针对一个新的评论文本，我们能以此为标准，找到其中的方面项和其情感分类。

B. Lap14

Lap14 数据集，也称为笔记本电脑-2014 数据集，是笔记本电脑领域另一个广泛使用的基于方面的情感分析 (ABSA) 基准数据集。它也是为 SemEval-2014 的基于方面的情感分析任务创建的。Lap14 数据集特别关注笔记本电脑领域。它包含了客户对笔记本电脑各个方面的评论和意见，比如性能、设计、电池寿命和价格。与 Rest14 数据集类似，Lap14 数据集来自各种来源的在线评论集合，包括电子商务网站、技术评论平台和其他来源。评论由人类注释者手动注释，以识别方面和情感极性。Lap14 数据集在方面级别提供注释，类似于 Rest14。评审中的方面或实体被识别和标记，每个方面对应于一个特定的方面类别，如“性能”、“设计”、“电池寿命”、“价格”等。数据集还包括与每个方面相关的情绪极性标签 (积极，消极，中性)。与 Rest14 一样，Lap14 数据集包含多种语言的评论，主要是英语和西班牙语。这使得研究人员可以在不同的语言中探索基于方面的情感分析，并研究跨语言迁移学习技术。

C. Twitter

基于方面的情感分析 (ABSA) 上下文中的 Twitter 数据集是指专门从 Twitter 上收集的数据集。这些数据集捕获包含用户对各个方面或实体的意见的推文，使其成为研究短文本和非正式文本情感分析的宝贵资源。Twitter 数据集涵盖了广泛的领域和主题，因为 Twitter 是一个用户可以就不同主题表达意见的平台。研究人员根据他们的研究重点，收集与特定领域 (如电影、产品、事件或一般主题) 相关的 Twitter 数据。Twitter 数据集通常是通过使用 Twitter API 收集的，它允许研究人员根据特定的搜索查询、标签、用户配置文件或其他标准访问和检索 tweet。数据收集过程包括选择相关关键字、定义搜索查询和检索符合所需标准的 tweet。由于 tweet 的非正式性质，通常需要预处理步骤来处理噪声，例如删除 url，用户名，标签，表情符号和标点符号。标记化、小写和删除停止词也是常用的预处理技术来清理 tweet 文本。Twitter 数据集通常涉及手动注释，人工注释者在 tweet 中标记情感和方面/实体。情绪标签可能包括积极、消极、中性或更细粒度的选项。方面/实体标签表示 tweet 中正在讨论的特定方面或实体。

D. Rest15

类似 Rest14 数据集，最早收集在 SemEval-2015 竞赛上，数据格式为 XML，标注方式和背景内容同 Rest14。

E. Rest16

表 2.3 CRF 模型参数

参数	取值
L1 coefficient	0
L2 coefficient	1.0
Learning rate	0.1
epoch	75

同上 Rest14 数据集，最早收集在 SemEval-2016 竞赛上，数据格式为 XML，标注方式和背景内容同 Rest14。

2.4.2 实验方案

由本章可知，机器学习方法来实现 ABSA 任务时，将其拆分为两个子任务来进行试验。试验分为方面项抽取和方面项 Term 情感分类两个子试验。

A. 方面项 抽取试验

在抽取试验中，首先根据本章第一节中的特征构造与选择的讨论，选择了如下表的 12 个特征，而其中的符号含义也一一列出表例然后根据上述特征表对数据进行处理，将特征列表输入到 CRF 中进行训练，本实验中的 CRF 参数如表 2.3 所示：

- L1 正则化 (Lasso) 参数: L1 正则化的正则化参数 (lambda) 被设置为零。这意味着没有向 CRF 模型的目标函数中添加稀疏性诱导惩罚。
- L2 正则化 (Ridge): L2 正则化的正则化参数 (lambda) 的默认值通常设置为 1.0。该值控制 L2 正则化的强度，较大的值会增加对大权重的惩罚。
- 学习率: 学习率设置为 0.1。在使用基于梯度的优化算法的训练过程中，学习率决定了更新模型参数的步长。
- 最大迭代: 最大迭代数设置为 75。该值决定了训练过程中允许的最大迭代次数这里的设置是为了让学习过程早点结束。

模型的输出为相应的方面项的位置，即索引，这样根据索引可以反向定位到具体的词语，也方便我们找到方面项对应的上下文的信息。

B. 方面项情感分类

根据第一节中的概述，方面项的分类工作在抽取工作之后完成，此时我们默认已经成功抽取到了句子中所有的方面项的列表，基于此去寻找每个方面项的情感分类。而在 SemEval2014 的比赛中，情感分类是作为 task4 出现，题目中给出了文本和句子中的方面项 列表，在此基础上完成对每一个方面项的情感分类。具体见上一节的思路概述，我们找出某一个方面项在语义上的上下文，然后将这一个单词的集合按语义顺序排列，组成一个“句子”，这里加引号的含义是说明这里的句子实际上只是一个语义上互相依赖的单词的集合，并不是一个有通顺语义的，自然语义上一定容易理解的句子。然后该子任务就转化为了对这个构造的抽象“句子”进行文本三分类工作，那么我们将

其单词转化为词向量作为特征加入到常见的机器学习模型中进行学习，训练一个输出为-1, 0, 1 的分类器。最后根据此分类器对应每一个方面项 做分类预测。

2.4.3 评价指标

当处理机器学习三分类问题时，其中输出涉及三个不同的类或标签，通常使用的评估度量是用于二元分类问题的评估度量的扩展。以下是一些常见的评价指标。为了比较准确统一地定义几个评价标准，这里给出分类问题的混淆矩阵作为基准，将分类的情况分为以下四类：

- TP (True Positive): 真阳性，指模型预测结果和真实标签都是指定的正向情绪。
- TN (True Negative): 真阴性，指模型预测结果和真实标签都不是指定的正向情绪。
- FN (False Negative): 假阴性，指模型预测结果不是指定正向情绪，但是真实标签是指定正向情绪。
- FP (False Positive): 假阳性，指模型预测结果是指定正向情绪，结果也不是指定正向情绪。

那么可以定义几种常见的评价标准的数学形式如下：

A. Accuracy

Accuracy 的计算式如下：

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

准确性度量模型在所有三个类中预测的总体正确性。它计算正确预测的实例与实例总数的比率。然而，仅靠准确性可能无法提供完整的图像，特别是当类不平衡时。

B. Precision

Precision 的计算式如下：

$$Precision = \frac{TP}{TP + FP} \quad (2.2)$$

C. Recall

Recall 的计算式如下：

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

D. macro-F1-score

首先定义单分类问题的 F_1 -score 计算式为

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (2.4)$$

然后考虑本问题中的情感极性为 3，那么在考虑 F1-score 时，要对每个情感分类分别计算 Precision 和 Recall，然后分别求平均值，对平均 Precision 和平均 Recall 上求 F1-score，这样的结果表示为 macro-F1-score。

由于 macro-F1-score 已经考虑到了精度和 Recall 的综合结果，所以不再单独展示 Precision 和 Recall 的结果，在后续的结果分析中仅考虑 Accuracy 和 macro-F1-score。

2.4.4 实验结果及分析

在 SemEval2014 的比赛中，一共有四个任务，任务 2 的任务就是方面项提取，任务 4 的任务则是方面项分类。本文虽然也将 ABSA 任务划分成两个子任务。但是和在当时的比赛时候不同的是，选手在做方面项 的分类的时候，是默认前一个子任务，也就是方面项 抽取的准确率是 100%的，然后其他的准确率等评价指标也是以此为基础的，而事实上抽取工作是没有办法做到 100%的准确率保证的，因此后面在讨论不同模型之间的效率比较时候，也应该明确在计算的前提一致的情况下进行。

A. 方面项 提取

这里展示了在方面项提取的子任务上本文采用的方法（CRF+Word2vec）与其他几种在 SemEval2014 的 Task2 任务中表现出色的模型的实验结果对比分析。如表 2.4 所示，可以看出 Rest14 和 Rest16 上都有良好的抽取效果达到了 85.41%和 83.20%，而且抽取任务的准确率普遍比较高，尽管在 Laptop 数据集上的效果略低，但是也在比较良好的范围之内，而这种准确率的浮动，其实很难从模型上进行解释，因为对于不同主题的数据集而言，抽取任务的难度本就不同。本实验中没有对参数配置进行深层次研究，但是我们已经看到在一个粗略的模型上抽取任务已经具有良好的效果。

表 2.4 方面项抽取结果

Models	Rest14		Rest15		Rest16		Lap14	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
CRF	85.41	82.80	83.20	71.87	86.18	74.91	75.62	66.20

B. 方面项 分类

这里展示了在方面项情感分类的子任务上本文采用的方法（SVM、LR 等）和其他几种在 SemEval2014 的 Task4 任务中表现出色的模型在几个指标上的效果。如本节上文所述，本子任务是针对所有的句子中所有的方面项进行的评价，其中方面项分类成功是建立在方面项提取成功的基础上，也就是说，如果某一文本中方面项提取错误，则直接默认这一条数据的此分类结果也错误。而评价指标是所有的结果都是在此条件下进行的，而在后续章节中，也难免需要考虑到提取的效率问题，在后文中会明确说明计算的条件和假设。具体的结果见表 2.5，举例来说其中 LR 和 SVM 在 Rest14 数据集上分别得到了 63.40%和 61.85%的准确率。而综合所有数据集上的表现来看，我们的结论是在本文的思路下使用 LR 或者 SVM 并没有明显差别，而这个分类工作的效果显然是不太令人满意的，因此在后文中我们将讨论其他的解决思路。

表 2.5 方面项分类结果

Models	Rest14		Rest15		Rest16		Lap14	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
LR	61.85	57.36	61.60	58.59	60.08	58.17	57.60	59.63
SVM	63.40	62.24	63.30	58.45	61.96	59.45	60.79	57.71
LSTM	68.13	67.47	67.37	55.17	76.80	63.88	69.28	63.09

* 粗体表示在该指标上的最优结果

2.5 本章小结

本章简介而完整地完成了常规的机器学习方法对于细粒度情感分析问题的尝试，主要步骤包括数据处理，特征工程，分类器训练，结果分析等。其中受到 SemEval2014 主办方的启示，将该任务划分为方面项抽取和方面项分类两个子任务，在方面项抽取的任务上取得了比较好的效果，但是在方面项分类任务上效果却并不令人满意，而机器学习方法的实验步骤多，容易受到操作者主观经验影响等缺点也确实让实验者感到困难重重。总的来说，本章节在讨论机器学习方法在细粒度情感分析问题上表现的同时，也为接下来的章节起到了启示的作用。

3 基于门控卷积神经网络的细粒度情感分析

3.1 引言

3.2 卷积神经网络概述

卷积神经网络（Convolutional Neural Networks, CNNs）最初是为计算机视觉任务开发的，但近年来已经被用于自然语言处理问题。使用 CNNs 进行自然语言处理的基本思想是将文本视为一个二维图像，其中一个维度是句子长度，另一个维度是词向量的大小。一般来讲，一个完整的处理 NLP 问题的神经网络包括内容如下：

A. 嵌入层

网络的第一层接收输入文本作为一系列词嵌入。因为神经网络无法处理文本数据，所以要将文本语言映射到向量空间，具体的映射方法讨论见上一章的第三节。此时对于一个句子，其长度为 L ，那么假设每个单词经过嵌入层之后映射到一个 D 维词向量空间，那么句子经过嵌入层的输出如下：

$$X = [W_1, W_2, \dots, W_L], W_i \in R^D \quad (3.1)$$

B. 卷积层

卷积层对输入序列应用一组具有一定大小，如 $D \times k$ 的过滤器，其中把 k 称为卷积核的大小。每个过滤器滑动过输入序列，计算过滤器权重和相应的词嵌入序列之间的点积。如下所示。

$$c_i = f(X_{i:i+k} * W_c + b_c), W_c \in R^{D \times k} \quad (3.2)$$

所有的 c_i 组合得到的结果是一组特征映射 C 。每个特征映射表示输入序列中某种模式的存在。 C 的长度与 L 和 k 都有关，可以简单记作 $f(L, k)$ 。

C. 池化层

池化层通过对每个特征映射应用池化函数（例如最大池化）来减少特征映射的维数。此操作捕获每个特征映射中最重要的特征，以滤除噪声，从而帮助模型更加鲁棒地处理输入的变化。

D. 全连接层

全连接层获取池化层的输出，并将其通过一组与每个输入神经元相连接的神经元。该层对池化数据映射到概率空间进行最终分类以实现所需的类别（例如正面或负面情感）。

E. 输出层

输出层产生模型的最终预测。

尽管 CNN 最早并不应用于 NLP 任务，但它可以学习输入文本词向量中的词义特征和语义依赖关系，从而成功地应用在文本分类工作中。而由于 CNN 卷积层可以并行的特点，在 NLP 的问题中，时间开销上往往优于 LSTM 等串行处理的网络。

3.3 基于门控机制的关键词捕捉

CNN 中的门控机制有选择地过滤网络不同层之间传递的信息。CNN 通常由多层卷积滤波器和池化层组成，这有助于从输入数据中提取特征。然而，并非所有特征对于最终的分类任务都是同等重要的，有些特征甚至可能是无关或嘈杂的。当门控单元在 CNN 中使用，可以有选择地过滤掉无关或嘈杂的特征，同时允许重要特征通过。这是通过引入门单元到网络中实现的，这些单元类似于在 RNN 中使用的门控单元。门控单元通常实现为 sigmoid 或 ReLU 激活函数，它们以卷积层的输出作为输入，并为每个特征图输出一个介于 0 和 1 之间的门值。

然后，门控值与特征图逐元素相乘，以有选择地过滤掉信息。例如，如果门值接近 0，则相应的特征图被有效地“关闭”，其信息被过滤掉。另一方面，如果门值接近 1，则相应的特征图被完全“打开”，其信息被允许通过到下一层。门控机制是 CNN 的一种重要手段，因为它允许网络有选择地学习和使用相关特征，同时过滤掉无关或嘈杂的特征。这可以提高网络在图像分类、目标检测和语义分割等任务中的效率和准确性。

3.4 引入门控机制的卷积神经网络

类似章节 2 中机器学习方法中的分步策略，其主要结构如图 3.1 所示：

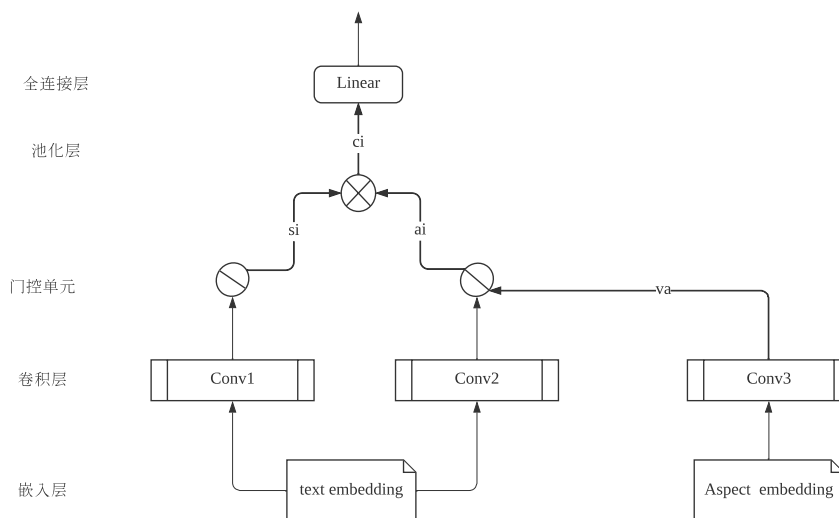


图 3.1 门控 CNN 网络的结构

A. 嵌入层

如上一节所述，嵌入层将句子文本转化为 $L \times D$ 的词向量，其中 L 为句子的长度， D 为词向量的维度。

B. 卷积层

其中对于卷积层网络结构的讨论可以分为两个主要的部分：方面项提取层，方面项分类层，如

图 3.1 所示，这里分别说明其作用：

方面项提取网络结构为一个卷积层和一个池化层，它的主要作用是将词向量的输入进行卷积处理，从而提取一个 t 时刻的方面项抽象特征，之所以称之为抽象特征，是因为这个特征是否有意义，需要结合门控单元进行有效性验证。比如一个句子中的方面项是 “sushi rolls”，那么为了统一这种情况下的方面项提取输出，最简单的手段就是进行卷积和池化来提取已知维度下的特征，如图 3.1 中经过池化后的输出部分理论上保留了两个单词的共同特征，记为 v_a 。而在形式上，它和一个单词的词向量并无区别。

方面项分类层核心思想是使用了一个 GTRU 的门控结构控制了方面项的筛选，从而找出最符合某一个方面项的分类。所谓 GTRU 门，指的是一个 Tanh 门和一个 ReLU 门，而其处理的输入和输出的方式也和通常情况下的卷积神经网络不同。从而找出最符合某一个方面项的分类，其中图 3.1 中的 Tanh 门在示意图中与 Conv1 相连用于捕捉文本中的情感特征，其输出为

$$s_i = \text{Tanh}(X_{i:i+k} * W_s + b_s) \quad (3.3)$$

ReLU 在示意图中与 Conv2 相连，用于捕捉方面项信息，其输出为

$$a_i = \text{ReLU}(X_{i:i+k} * W_s + V_a v_a + b_s) \quad (3.4)$$

Xue 等人第一次尝试通过 GTRU 的门控单元组合来提取方面项的结构^[9]，但在论文中并没有提到 GTRU 为何使用并且高效地完成了分类工作中的特征捕捉任务，在本问题中，后续试验部分也将探讨其他几种不同门控机制在本任务中的表现。

最后将两个门控单元的结果做哈达玛积，即

$$c_i = s_i * a_i \quad (3.5)$$

得到某一个方面项筛选下的分类结果，在该过程中，对应一个方面项，其中的某一个与方面项密切关联的词语可能会获得比较重要的权重，而另一个关联度较低的单词则会获得较低的权重甚至 0 权重。

C. 池化层

我们已经知道，在卷积层中，若存在 n_k 个大小为 k 的卷积核对一个长度为 L 的句子做卷积运算，实际上可以得到一个 $n_k \times f(L, k)$ 大小的卷积层输出 W_i 。在池化层中，我们取这个 W_i 每行的最大值，这个过程称为 max-pooling，作用是对卷积输出做一个过滤和抽象，大幅度保留其特征的同时对矩阵进行降维。经过 max-pooling，我们得到一个 nk 的池化层输出。

D. 全连接层

最后将运算结果输入到全连接层进行分类结果输出。假设情感标签数量为 T ，那么此时的全连

接层为一个含大小为 $T \times n_k$ 的可学习参数矩阵 W 的映射运算：

$$\text{logit} = WH, W \in R^{T \times n_k} \quad (3.6)$$

再通过 softmax 函数将上述的运算结果 logit 再映射到 0 和 1 之间，且总和为 1：

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^T e^{z_j}} \quad (3.7)$$

其中 z_i 为 logit 中的向量，总长度为 T （情感标签数量），此时输出 P 为本次模型分类工作对于所有情感标签的概率空间。再根据当前模型计算出交叉熵：

$$\text{Loss} = - \sum_{i=1}^T y_i \log P_i \quad (3.8)$$

然后根据梯度方向传播以减小交叉熵为目标更新参数，并进行循环迭代一定次数，最后对于该背景下的问题获得良好的分类效果。

3.5 实验

3.5.1 实验数据

同章节 2 中的数据集，包括 Twitter, Rest14, Lap14, Rest15, Rest16。

3.5.2 实验设计

本实验中，对单词采用的是章节 2 中的 GloVe 方法进行预训练，其中的训练参数如表 3.1 所示，其中 $D_{embedding}$ 代表词向量的维度，Kernel number 代表卷积核的数量，指同时有多少个卷积核同时对句子进行卷积运算，Kernel size 代表卷积核的大小，代表卷积核一次卷积能同时处理句子中的多少单词，batch size 代表的是送入模型的一轮训练所使用的数据量，在文本处理中等于句子的长度。learning rate 是指在最小化交叉熵的梯度选择器中学习率的大小，代表拟合的速度。在本实验中，采用的优化器是 Adam，其内部的超参数主要包括两个：第一个是 β_1 ，代表指数衰减率，控制更新优化状态时候梯度的权重，第二个是 β_2 ，控制更新优化状态时候梯度的平方项的权重。这里采用的都是优化器的默认初始化值。epoch 指迭代次数，指训练循环的次数，seed 是实验中的随机数种子，用于随机初始化实验中的模型，dropout 是隐藏层的参数，代表在某一时刻隐藏状态的参数被随机置 0 的比例，这是为了减少模型过拟合的一种训练机制。

3.5.3 评价指标

同章节 2 的中的评价指标，包括 Accuracy 和 macro-F1-score。

表 3.1 GCNN 模型参数

参数	取值
$D_{embedding}$	300
Kernel number	100
Kernel size	3
batch size	32
learning rate	$1e^{-3}$
β_1	0.9
β_2	0.999
epoch	30
seed	776
dropout	0.2

3.5.4 实验结果及分析

在本节中，通过门控 CNN 模型进行细粒度情感分析工作，其结果如表 3.2 所示，其中 SVM 中的 N/A 表示空缺，SVM 方法代表的是章节 2 中实现的基于 SVM 模型对方面项进行分类工作，其中没有在 Twitter 数据集上进行试验，而在其他相关论文中也没有找到该相关的数据，或可复现源码，固空缺。LSTM 是长短期记忆网络，采用循环计算的思路计算最终状态的情感极性，MemNet 是细粒度情感分析的经典模型，它提取方面项的上下文作为记忆模块，通过多跳机制来优化模型的分类效果，IAN 也是比较常见的基于方面项的文本分类模型，通过上下文交互的方式提取上下文的语义依赖关系，在情感分类工作中效果显著。这三种方法都是细粒度情感分析中被大量引用和探讨的方法，而 GCNN（Gated CNN）则是本章采用的模型，在 Rest14，Rest15，Rest16 上都表现良好。

表 3.2 模型效果比较

Models	Rest14		Rest15		Rest16		Lap14		Twitter	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
SVM	63.40	62.24	63.30	58.45	61.96	59.45	60.79	57.71	N/A	N/A
LSTM	68.13	67.47	67.37	55.17	76.80	53.88	69.28	63.09	69.56	67.70
MemNet	69.61	64.64	67.31	58.28	75.44	52.99	70.61	65.17	71.48	69.90
IAN	69.26	60.09	68.54	52.65	74.74	55.21	72.05	67.38	71.50	70.81
GCNN	70.52	62.19	68.34	58.94	77.45	54.52	71.69	66.72	71.09	68.95

* 粗体表示在该指标上的最优结果

3.6 本章小结

本章尝试了门控 CNN 模型对于细粒度情感分析问题的尝试，主要步骤包括数据处理，模型搭建，神经网络训练，结果分析等。CNN 处理文本分类工作时具有可并行，效率高等特点，所以在本章工作中选取其为主要模型，而采用门控单元来进行文本和方面项相关性的计算和判断，可以完成对输出信号的筛选和过滤。实验表明，该模型在细粒度情感分析任务常用的几个数据集上都表现出了较高的准确率，具有良好的分类处理效果，验证了 CNN 结构在处理文本特征时不逊色于 RNN 的性能以及门控单元的有效性。

4 基于门控图卷积神经网络的细粒度情感分析

4.1 图卷积神经网络概述

图卷积是自然语言处理中用于建模单词或其他语言单位之间结构依赖关系的技术。在 NLP 中，图通常用于表示单词之间的关系，例如语义相似性或上下文关系。在图卷积中，输入数据被表示为图结构数据，其中节点表示单词或其他语言单位，而边表示它们之间的关系。图卷积网络（GCN）直接在图形上应用卷积操作，以提取捕获节点之间结构依赖关系的特征。一种常见的方法是使用图卷积层，该层将图卷积操作应用于图中每个节点及其邻居。卷积操作的输出然后在图中所有节点上汇总，以生成一组新的节点特征。图卷积已应用于各种 NLP 任务，例如情感分析、命名实体识别和关系提取。通过建模单词之间的结构依赖关系，图卷积可以捕获更复杂的语言模式，并提高 NLP 模型的性能。从数学语言上来说，图卷积的基本方法是基于图的拉普拉斯矩阵进行特征值分解的操作。首先考虑上述的拉普拉斯矩阵，假设有一个 N 个节点的无向图和它对应的邻接矩阵 A ，可以定义它的拉普拉斯矩阵基本形式为 $L=D-A$ ，其中 D 是以每个节点的度，即每个节点连接的边的数量，为对角线元素的对角矩阵。

正则化拉普拉斯矩阵之后， L 可以如下表示：

$$\begin{aligned} L &= D^{-\frac{1}{2}} L D^{\frac{1}{2}} \\ &= I_N - D^{-\frac{1}{2}} A D^{\frac{1}{2}} \\ &= U \Lambda U^T \end{aligned} \quad (4.1)$$

此时 L 可以表达为

$$L = \begin{cases} 1 & i=j \text{ and } \deg(v_i) \neq 0 \\ -\frac{1}{\sqrt{\deg(v_i)\deg(v_j)}} & i \neq j \text{ and } v_i \text{ and } v_j \text{ is adjacent} \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

定义对角滤波器 $g_\theta = \text{diag}(\theta)$ 并使用该滤波器对图的拉普拉斯矩阵做卷积，即

$$g_\theta * x = U g_\theta(\Lambda) U^T X \quad (4.3)$$

其中 g 可以展开为多项式 T_k 来估计到 k 阶：

$$g_\theta * x \approx \sum_{k=0}^K \theta_k T_k(\tilde{L}) x \quad (4.4)$$

其中

$$\frac{2}{\lambda_{max}}L - I_N \quad (4.5)$$

其中 λ_{max} 是 L 的最大特征值。接下来我们尝试用切比雪夫多项式来化简该方程。由切比雪夫多项式的递推公式可知：

$$T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), T_0(x) = 1, T_1(x) = x \quad (4.6)$$

在一个简单的 GCN 中，可以取 2，我们将切比雪夫多项式展开到一阶，那么有

$$\begin{aligned} g_\theta * x &\approx \theta'_0 x + \theta'_1 (L - I_N)x \\ &= \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{\frac{1}{2}} \end{aligned} \quad (4.7)$$

再令 $\theta = \theta'_0 = -\theta'_1$ ，此时有

$$g_\theta * x \approx \theta(I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}})x \quad (4.8)$$

给图加上自环，此时新的邻接矩阵为

$$\tilde{A} = A + I_N \quad (4.9)$$

此时的度矩阵为

$$\tilde{D}_{ii} = \sum_{j=0}^N \tilde{A}_{ij}$$

加入激活函数后，可以写成

$$H^{l+1} = \text{ReLU}(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l) \quad (4.10)$$

这个卷积方法对应 GCN 网络中的一层卷积，是比较简单基本的形式，在别的论文中有作者对图卷积的运算进行了改进的各种尝试，其中的转移矩阵运算和算法过程可能略有不同，但是整体的思路和原理是相似的，在本文中不会对此进行推导上的更深层次讨论。

4.2 依赖图的获取

在章节 2 中已经提到，想要提取出文本中的上下文关系，我们需要通过提取其语义上的依赖关系，比如下面的一个句子：

The pizza is delicious but the price is high.

在这个句子中，经过依赖分析可以得到其每个词语在句子中扮演的成分和其他各词语之间的联系，这个关系是可以用一个有向图来表示的。其中每个节点都是一个词汇，词汇之间的关系就代表节点之间的连线。

假如用一个箭头来表示一个指向关系的话，那么箭头的起始端则是这个关系的父节点，而箭头

指向端则是这个关系的子节点。而没有父节点的节点则成为 ROOT 节点。这里面我们会发现，每一个单词都会对应另外的单词指向它，并且一个单词只被一个单词指向，这样每一个单词都会找到一条唯一的线路回溯到 ROOT 节点。那么通常的依赖图的寻找方法，往往需要训练一个分类器，将词语两两加入分类器，然后每一次输出一个判断变量，如果判断的结果为两个节点之间有关系，那么则将两个节点之间连上一条边，证明两者之间有句法依赖关系，按照这个逻辑，这个判断算法的时间消耗应该是在 N^2 的量级上的。为了将图的依赖关系表示加入卷积神经网络中进行运算，还需要对其进行一定的转化，这里选择将其转化为邻接矩阵，邻接的解释含义就是词语在这个句子中的语义依赖图中存在依赖关系。那么根据图的性质我们可以讨论分为有向图和无向图，这两种图在解释性上都有各自的道理，比如，对于一个主谓宾的句子，有向图可能认为，主语是一个更支配更主动的地位，而宾语尽管和主语都是名词，却在句子中的作用靠后一些；而无向图可能认为，句子的主语和宾语没有语义上的“先后”和“主被动”之分，那么显然这两者在解释性上都有一定的道理，那么在后文中也会根据两种情况的训练效果进行评估。本文利用 Python 的 Spacy 工具包来生成依赖图的模型，然后通过遍历图计算出图的邻接矩阵 A ，其中如果 $A_{ij}=0$ ，说明节点 i 和 j 之间没有依赖关系；反之，如果 $A_{ij}=1$ ，则说明 i 和 j 之间存在依赖关系。比如图 2.1 中的依赖结构，可以通过上述规则得到其依赖图的邻接矩阵如所示

$$A_{9 \times 9} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \quad (4.11)$$

至于依赖的关系的种类，在本文中只宽泛地认为这些节点都是在语义上“相邻”的。而依赖关系的种类从自然语言的角度出发，是有很多种不同的情况，而其所表达的语义和信息也会各有侧重，但是这个讨论显然超过了本文的范畴，本文仅简单带过，不做更深探讨。

4.3 引入门控机制的图卷积神经网络

图例本节中介绍引入门控机制的图卷积神经网络。其结构和章节 3 中的门控卷积网络类似，具体如图 4.1 所示

A. 嵌入层

与上一章(章节 3)的嵌入层一致。

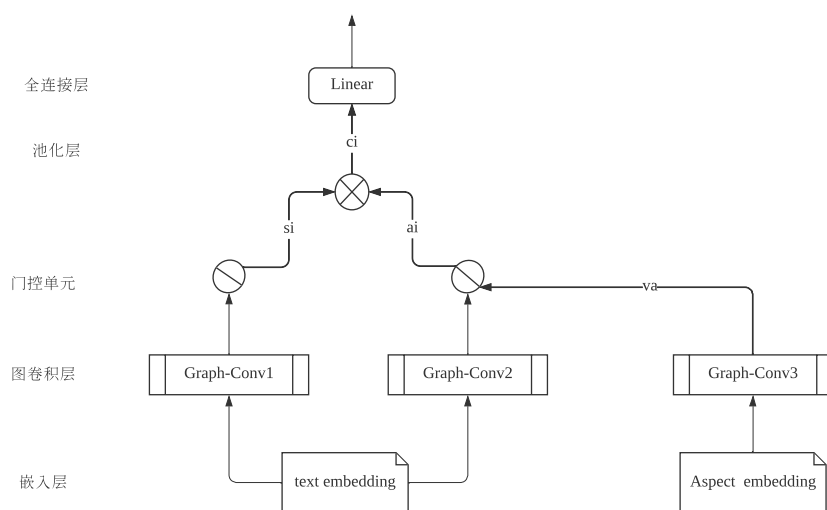


图 4.1 门控 GCN 网络的结构

B. 图卷积层

由上一节可知，我们通过图卷积层来捕获图节点之间的依赖关系，其中图的依赖结构通过图的邻接矩阵 A 实现。在经过图卷积运算（参考上一节），记为 $f(,)$ ，该映射的输入是上一个状态下的各个节点的隐藏状态 H_k 和依赖图的邻接矩阵 A ，输出则为更新后的节点状态 H_{k+1} 。具体如下：

$$H^{k+1} = f(H^k, A) \quad (4.12)$$

C. 门控机制

与上一章的门控机制结构一致，通过 GTRU 门来选择与方面项关联度高的情况进行信号输出，其中 Tanh 门提取文本的依赖图的结构信息，Relu 门提取方面项依赖图的信息。在本章中，还进一步探索不同的门控单元组成的表现力问题，其中，将探究 Tanh-ReLU 门（GTRU 门）和另外三种门控结构的表现力，包括 Tanh-Tanh 门，ReLU-ReLU 门，ReLU-Tanh 门，分别记为 GTTU 门，GRRU 门，GRTU 门。

D. 损失函数

与上一章（章节 3）中的全连接层一致。

4.4 实验

4.4.1 实验数据

同章节 2 中的数据集，包括 Twitter，Rest14，Lap14，Rest15，Rest16。

表 4.1 GGCN 模型参数

参数	取值
$D_{embedding}$	300
Layer dimension	2
batch size	32
learning rate	$1e^{-3}$
β_1	0.9
β_2	0.999
epoch	30
seed	776
dropout	0.2

4.4.2 实验设计

本实验中，对单词采用的是章节 2 中的 GloVe 方法进行预训练，其中的训练参数如表 4.1 所示，部分参数可参考章节 3。其中 $D_{embedding}$ 代表词向量的维度，Layer dimension 代表图神经网络中的层数，batch size 代表的是送入模型的一轮训练所使用的数据量。learning rate 是指学习率的大小。在本实验中，采用的优化器是 Adam，其内部的超参数主要包括两个指数衰减率： β_1 和 β_2 。这里采用的都是优化器的默认初始化值。epoch 指迭代次数，seed 是实验中的随机数种子，dropout 是隐藏层的参数，代表在某一时刻隐藏状态的参数被随机置 0 的比例。

4.4.3 评价指标

同章节 2 的中的评价指标，包括 Accuracy 和 macro-F1-score。

4.4.4 实验结果及分析

在本节中，通过门控 CNN 模型进行细粒度情感分析工作，其结果如表 4.2 所示，同章节 3，其中 SVM 中的 N/A 表示空缺，LSTM，MemNet，IAN 等方法都是细粒度情感分析中被大量使用的方法，而 GCNN（Gated CNN）则是章节 3 采用的模型，GGCN（Gated GCN）则是本章采用的模型结构，其在本文所列举的所有五个数据集上都表现出了良好的效果。与 GCNN 相比，在 Rest14 上提升了 0.08%的准确率，Rest15 上提升了 0.74%的准确率，1.21%的 F1 score，在 Rest16 上提升了 0.27%的准确率，3.41%的 F1 score，在 Lap14 上提升了 2.86%的准确率，2.42%的 F1 score，在 Twitter 上提升了 1.04%的准确率，1.40%的 F1 score。在 10 个指标中有 9 个指标都获得了提升，仅有 1 个指标小幅度下降，充分体现了在同种网络架构下图卷积层对于文本语义的抽象能力是更强的。

值得注意的是，在 Restaurant 系列数据集上，门控 GCN 相较于门控 CNN 的提升幅度普遍偏小，甚至个别指标小幅下降，可能是由于 Restaurant 的评论，普遍比较简短而且情感态度浓烈，语义结构简单，因此图卷积在考虑句子语义依赖时发挥不出其在抽象远距离单词之间关系的能力，反而可能增加了冗余信息干扰了分类器判断。以上分析仅从自然语义上进行解释，本文没有在理论上更进一步证明，只从实验数据上提出一些猜想供以后工作提供经验。

此外，在大部分讨论和使用门控单元的论文中并没有解释门控单元使用上的选择问题。这里给出一个自然语言上的解释：由于 \tanh 的值域在 $[-1,1]$ ，所以其在三分类的任务中有比较好的抽象能力，因为三分类任务也是一个 $-1,0,1$ 的映射任务；而 ReLU 函数在方面项提取任务中表现出色，其原因是该函数在某一时刻 t 只有取值为正时候输出为正，取值为负的时候都没有输出，这样的整流作用让经过该门的输出只在与方面项有一定关联度的时候才会进行下一步的运算，否则输出为 0，从而完成了方面项的筛选工作，这里没有非常严格的数学证明，只是从运算过程上给出一个符合逻辑的推断。因此，本文在实验中还测试了不同门控单元的表现效果，其中包括 GTRU 门，GTTU 门，GRRU 门，GRTU 门。他们的表现如表 4.3 所示，其中可以发现，GTRU 和 GRRU 门的表现相近，而他们两个会略高于另外两种门控单元。因此，在本任务中，ReLU 门用于提取方面项特征时候是具有优势的，而 Tanh 提取文本特征时候的表现和 ReLU 门相近，在众多数据集上没有明显优势。从实验上也说明了门控单元的选择对于分类结果存在影响，而且在不同的任务上这个影响的大小也不一样，此时，从可解释性上来分析门控单元的选择可能是有效的。

表 4.2 模型效果比较

Models	Rest14		Rest15		Rest16		Lap14		Twitter	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
SVM	63.40	62.24	63.30	58.45	61.96	59.45	60.79	57.71	N/A	N/A
LSTM	68.13	67.47	67.37	55.17	76.80	53.88	69.28	63.09	69.56	67.70
MemNet	69.61	64.64	67.31	58.28	75.44	52.99	70.61	65.17	71.48	69.90
IAN	69.26	60.09	68.54	52.65	74.74	55.21	72.05	67.38	71.50	70.81
GCNN	70.52	62.19	68.34	58.94	77.45	54.52	71.69	66.72	71.09	68.95
GGCN	70.58	62.02	69.08	60.15	77.78	57.93	74.55	69.14	72.13	70.35

* 粗体表示在该指标上的最优结果

表 4.3 门控单元效果比较

Gate	Rest14		Rest15		Rest16		Lap14		Twitter	
	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1	Accuracy	F1
GTRU	70.58	62.02	69.08	60.15	77.78	57.93	74.55	69.14	72.13	70.35
GTTU	68.49	60.25	68.38	59.45	75.96	57.51	73.79	67.71	71.59	67.62
GRRU	70.61	63.74	68.87	60.17	77.89	57.88	74.28	69.09	72.56	67.85
GRTU	69.71	62.64	67.27	58.91	76.44	56.39	72.61	68.17	71.48	69.90

* 粗体表示在该指标上的最优结果

4.5 本章小结

本章尝试了门控 GCN 模型对于细粒度情感分析问题的尝试，主要步骤包括数据处理，模型搭建，神经网络训练，结果分析等。GCN 处理文本分类工作时采用语义依赖图来表达句子中不同成分之间的关系，相比于基于 CNN 网络的结构，GCN 能表示的信息量更大，能更精确地刻画自然语言的语义关系，所以在本章工作中选取其为主要模型。而采用门控单元来进行文本和方面项相关性的计算和判断，以完成对输出信号的筛选和过滤，这在章节 3 中被证明是有效的，固仍然采用门控单

元为重要组成成分的卷积层结构，只改变卷积方式和预处理模块。实验表明，该模型在细粒度情感分析任务常用的几个数据集上都表现出了较高的准确率，具有良好的分类处理效果，并且在多项指标上相较门控 CNN 模型有了一定的提升，验证了 GCN 结构在抽象文本特征时更精确完整的优越特性。此外，本章探讨了门控单元的选择问题对细粒度情感分析问题的影响，并结合实验进行了猜想的验证，最后在实验后对门控单元的表现讨论了其可解释性。

装

订

线

5 总结与未来工作展望

5.1 本文工作总结

在本文中，主要围绕细粒度情感分析问题的概念，常见解决思路和优化方向进行了研究和呈现。其中细粒度情感分析问题和一般文本情感分类问题的不同之处是，其视角为句子中的情感对象方面项，基于每个方面项分析其情感倾向，所以其视角更小维度，更细粒度。而本文的主要工作主要以尝试进行细粒度情感分类工作为目标推进，首先尝试了模型较为简单的机器学习模型，通过构造特征来转化文本语言为数字信号并抽象了句子中的重要上下文语义关系，然后将细粒度情感分析任务划分为方面项抽取和分类两个分离的步骤，分别使用 CRF 和 SVM 等常见机器学习模型进行训练。

通过分析结果，以优化分类准确率为导向，使用了卷积神经网络等深度学习技术来替代手工提取特征，并使用门控单元实现了方面项识别和分类的功能集成。通过此门控 CNN 网络的模型训练和预测，大幅度提升了情感分类的效果。

最后考虑继续优化深度学习模型在复杂上下文中的抽象能力，引入图卷积网络层，通过语义依赖图结构来替代原始文本词向量信号，通过门控机制来实现方面项的匹配和筛选，最后在多个数据集的多项指标上得到了性能的提升，验证了图卷积神经网络在表达文本语义特征上的优越性。

此外，本文还讨论了门控单元的选择问题对情感分析问题结果的影响，进行了实验验证，并结合实验结果对门控单元的解释性问题做了进一步探讨和猜想。

5.2 不足与展望

本文探讨了多种不同方法对于解决细粒度情感分析问题的思路和流程，并对比了其在常见数据集上的效果和表现，并对其其中的一些不足之处提出了改进的思考和猜想，但是受时间、资源、能力所限，本文仍然存在许多不足之处值得在后续工作中去深入研究和讨论。

首先是本文的方法，不管是语义依赖关系的提取和表示，还是通过门控机制来实现筛选，都相对来说比较粗糙，从很多方面都可以更细化地进行分类，比如在语义依赖关系上更加具体地去探究不同的语义关系对其情感分类上的影响大小等，当然上述工作也需要大量的优质数据和模型训练作为支持，效果不一定是立竿见影的。

其次是在细粒度情感分析领域，优质的数据来源实在太少，泛用的数据集如 SemEval2014 等，都存在不同程度的标签分布不均衡，这会一定程度上影响某些分类标签的判断和分类。通用的数据集如 SemEval2014 等都是为了语义分析大赛而人工标注的数据集，其涉及的领域大多比较单一，比如针对食物，电脑等，比如我们在本文中已经探讨过同样的方法在针对不同的数据对象和内容时的分析难度是不一样的，因此细粒度情感分析的众多方法是否在更广泛的领域中依然能保证稳定和高效率，我们尚不可知。

参考文献

- [1] LIU B, ZHANG L. A survey of opinion mining and sentiment analysis[M]//Mining Text Data. Springer, 2012: 415-463.
- [2] MUBAROK M S, ADIWIJAYA, ALDHI M D. Aspect-based sentiment analysis to review products using naïve bayes[C]//Proceedings of the AIP Conference. AIP Publishing LLC, 2017: 020060.
- [3] JOACHIMS T. Learning to classify text using support vector machines[M]. Springer Science & Business Media, 2002.
- [4] MULLEN T, COLLIER N. Sentiment analysis using support vector machines with diverse information sources[C]//Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2004: 412-418.
- [5] DONG L, WEI F, TAN C, et al. Adaptive recursive neural network for target-dependent twitter sentiment classification[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short papers). 2014: 49-54.
- [6] 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.
- [7] TANG D, QIN B, LIU T. Aspect level sentiment classification with deep memory network[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2016: 214-224.
- [8] MULYO B M, WIDYANTORO D H. Aspect-based sentiment analysis approach with cnn[C]//2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). 2018: 142-147.
- [9] 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). Association for Computational Linguistics, 2018: 2514-2523.
- [10] 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. Association for Computational Linguistics, 2019: 4568-4578.
- [11] 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). Association for Computational Linguistics, 2021: 6319-6329.
- [12] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]//

Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014: 1532-1543.

- [13] PETERS M E, AMMAR W, BHAGAVATULA C, et al. Semi-supervised sequence tagging with bidirectional language models[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2017: 1756-1765.
- [14] 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). Association for Computational Linguistics, 2019: 4171-4186.
- [15] LI X, BING L, ZHANG W, et al. Exploiting BERT for end-to-end aspect-based sentiment analysis[C]//Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019). Association for Computational Linguistics, 2019: 34-41.
- [16] XIAO Z, WU J, CHEN Q, et al. BERT4GCN: Using BERT intermediate layers to augment GCN for aspect-based sentiment classification[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021: 9193-9200.
- [17] DAI H, SONG Y. Neural aspect and opinion term extraction with mined rules as weak supervision[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2019: 5268-5277.
- [18] ISLAM K I, KAR S, ISLAM M S, et al. SentNoB: A dataset for analysing sentiment on noisy Bangla texts [C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021: 3265-3271.
- [19] LING Y, YU J, XIA R. Vision-language pre-training for multimodal aspect-based sentiment analysis[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, 2022: 2149-2159.
- [20] HU J, LIU Y, ZHAO J, et al. MMGCN: Multimodal fusion via deep graph convolution network for emotion recognition in conversation[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). Association for Computational Linguistics, 2021: 5666-5675.
- [21] ZHOU Y, LIAO L, GAO Y, et al. To be closer: Learning to link up aspects with opinions[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021: 3899-3909.
- [22] ORBACH M, TOLEDO-RONEN O, SPECTOR A, et al. YASO: A targeted sentiment analysis evaluation dataset for open-domain reviews[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2021: 9154-9173.

谢辞

在一段学生生涯画上句号之时，我感到非常幸运，有在本科阶段里帮助过我的人如此之多，同时也遗憾在这里没法记下每一个想要去感谢的人。

首先，感谢我的导师苗夺谦老师，以及实验室里的其他几位老师，对于我而言，自然语言处理是一个全新的领域，刚开始是极其茫然的，苗老师在为我提供思路和规划毕设方面给了我很大的帮助，同时他的严谨的治学态度和自律上进的作风也给了我激励，让我在研究工作中也能高效而务实。

其次，我要感谢我的家人，我的父母为我提供了良好的生活环境，当我在遇到工作学习上的困难时候，他们总会是给予我陪伴的人。我的室友，同学，社团朋友们，他们在生活中给我鼓励 and 安慰，让我的本科生涯温暖而多彩。感谢实验室的学长学姐们，在每个星期四大家都会聚在一起讨论各自的领域中的想法，对我而言，带来的更多的是启发和激励，让我以他们为榜样，努力地学习工作。

最后，感谢同济大学里的四年时光，这是一段难以忘怀的美好回忆。