学校代号 10532　　学　　号 S141000892

分 类 号 TP393　　密　　级 普通

湖南大學
HUNAN UNIVERSITY

硕士学位论文

基于深度学习的情感分析方法研究

学位申请人姓名 吕超

培 养 单 位 信息科学与工程学院

导师姓名及职称 杨超 副教授

学 科 专 业 计算机科学与技术

研 究 方 向 自然语言处理

论文提交日 期 2017 年 5 月 22 日

学校代号：10532

学　　号：S141000892

密　　级：普通

湖南大学硕士学位论文

基于深度学习的情感分析方法研究

学位申请人姓名：　吕超

导师姓名及职称：　杨超 副教授

培　养　单　位：　信息科学与工程学院

专　业　名　称：　计算机科学与技术

论 文 提 交日 期：　2017 年 5 月 22 日

论 文 答 辩日 期：　2017 年 5 月 27 日

答辩委员会主席：　傅喜泉 教授

Sentiment Analysis Method Research Based On Deep Learning

by

LV Chao

B.E. (Hunan University) 2014

A thesis submitted in partial satisfaction of the

Requirements for the degree of

Master of Engineering

In

Computer Science and Technology

in the

Graduate School

Of

Hunan University

Supervisor

Associate Professor Yang Chao

May, 2017

湖 南 大 学

学位论文原创性声明

本人郑重声明：所呈交的论文是本人在导师的指导下独立进行研究所取得的研究成果。除了文中特别加以标注引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写的成果作品。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律后果由本人承担。

作者签名：　　日期：　　年　月　日

学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权湖南大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

本学位论文属于

1、保密□，在_____年解密后适用本授权书。

2、不保密□。

（请在以上相应方框内打“√”）

作者签名：　　日期：　　年　月　日

导师签名：　　日期：　　年　月　日

摘 要

传统的用于解决文本情感分析问题的方法包括基于情感词典和人工判定规则的无监督方法、基于机器学习的有监督方法。在数据量不大或者语义不够丰富的时候，这些方法能够取得一定的效果。但是随着数据量越来越大、表达方式越来越丰富，传统的方法已经无法有效地解决这一类问题，新的方法亟待提出。

本文根据文本情感分析的特点，结合当下应用领域广泛的深度学习算法，重点研究了基于深度学习的情感分析方法。卷积神经网络和递归神经网络是深度学习中两个主流的模型，前者能够从数据中提取出局部特征，而后者能够有效地分析时序数据、具有很强的上下文概括能力。本文的工作包括以下两个方面：

1、基于卷积神经网络和词语邻近特征的情感分析模型。目前，基于卷积神经网络的方法在情感分析任务中已经取得了不错的效果，此类方法使用词向量作为网络的输入，但是在卷积过程中每个词向量只能表征单个单词，并不蕴含上下文信息，这不利于信息传递的连续性，并且卷积操作在局部范围内可能会打乱词向量的序列性。针对这个问题，本文提出一种基于词语邻近特征的卷积神经网络模型，在卷积过程中让每个词向量携带邻近词语的特征，这样既保证信息传递的连续性也保证了词向量在局部范围内的序列性。实验结果表明，在 COAE2014、COAE2015 的情感分析任务上的准确率分别达到了 89.43%和 85.61%，说明本文提出的方法确实可行、有效。

2、基于递归神经网络和人工判定规则的情感分析模型。传统的基于情感词典和人工判定规则的分析方法是从语言学的角度出发，但是这种方法需要制定大量的情感词典和判定规则。而基于递归神经网络的分析方法可以通过不断的编码和重构训练，从大量未标注的语料中学习到先验知识。本文在参考了这两种方法之后，在第 4 章中提出了一种新颖的基于递归神经网络和人工判定规则的情感分析模型。首先利用递归神经网络并行计算出组成文本的多个子句的情感极性，然后根据极性融合规则将多个子句的情感极性进行融合，最后利用人工判定规则计算出原始文本的情感极性。该模型的优点在于：一方面递归神经网络中融入了人工判定规则，使得以往积累的人工经验得到有效利用；另一方面递归神经网络取代了情感词典，避免了情感词典的局限性。该模型在数据集 SST-C2 和 SST-C3 上的分类准确率分别达到了 87.8%、81.6%，并且整体性能均优于主流的分类模型，说明该模型不仅新颖，而且确实可行、有效。

关键词：情感分析；深度学习；卷积神经网络；递归神经网络；邻近特征；人工判定规则

Abstract

The traditional methods used to solve the problem of text sentiment analysis include unsupervised methods based on emotional dictionary and artificial judgment rules, and supervised methods based on machine learning. In the case of the amount of data is not large or semantics is not rich enough, These methods can achieve some effects. But with the increasing amount of the data and the expression is more and more rich, the traditional methods has been unable to effectively solve this kind of problem. New method needs to be put forward.

Based on the characteristics of text sentiment analysis and the deep learning algorithm for a wide range of applications in the current, this paper focuses on the method for sentiment analysis based on depth learning. Convolution neural network and recuesive neural network are the two mainstream models in deep learning. The former can extract local features from the data, and the latter can analyze the time-series data effectively, and have strong context generalization ability. The work of this paper includes the following two aspects:

1. Sentiment analysis model based on convolutional neural networks and word adjacent features. At present, the methods based on convolutional neural network has achieved good performance in the task of sentiment analysis. This method mainly uses word vectors as the input of the network, but in the convolutional process, the word vector can only characterize a single word, but does not contain contextual information, this is not conducive to the continuity of information transmission, and convolution may disrupt the sequence of word vectors in the local context. Aiming at this problem, this paper proposes a convolutional neural network model based on word adjacent features in the third chapter, which allows each word vector to carry the characteristics of its adjacent words in the convolution process, so that we can both ensure the continuity of information transmission and the sequence of word vectors in the local scope. The experimental result shows that the prediction accuracy rate on the sentiment analysis task of COAE2014 and COAE2015 reached 89.43% and 85.61% respectively, which indicates that the

proposal model is actually feasible and effective.

2. Sentiment analysis model based on recursive neural networks and artificial judgment rules. The conventional rule-based approach is from a linguistic perspective, but these method involves formulating sentiment dictionary and amount of judgment rules.However, the method based on recursive neural network can learn the prior knowledge from a large number of unlabeled corpus. After referring to these two methods, we present a novel sentiment analysis model based on recursive neural network and artificial judgment rules in the fourth chapter. Firstly, the sentiment polarity of multiple clauses that make up the original sentence are calculated by the parallel recursive networks, then the outputed polarity of the clauses are merged by polarity fusing rules to calculate the sentiment polarity of the original sentence. The advantage of this model on the one hand is that the recursive neural network model incorporates artificial judgment rules, so that artificial experience is effectively used. On the other hand, the recursive neural network replaces the sentiment dictionary, and this can avoid the limitations of the emotional lexicon. The classification accuracy on the SST-C2 and SST-C3 datasets are 87.8% and 81.6% respectively, and the overall performance is better than the mainstream classification model, which indicates that the proposed model is not only novel but also actually feasible and effective.

Key Words: Sentiment analysis; Deep learning; Convolutional neural network; Recursive neural networks; Adjacent features; Artificial judgment rules

目 录

插图索引

附表索引

第1章 绪论

1.1 研究背景及意义

随着 Web2.0 的快速发展，网络信息化时代以一种前所未有的速度迅速影响着人们的生活。与此同时，社会媒体也呈现出多样化的形态，论坛、博客以及微博等网络媒介如雨后春笋般的快速发展，网络用户与网络媒介之间的互动性也不断提高，这些都促使网络的使用方式发生了巨大变化。与过去不同，用户不再是只能被动的获取网络信息，而是更加积极主动的成为了网络信息的生产者。这种变化使得如今各种网络平台上呈现出大量的用来表达用户情绪和观点的主观性信息，而文本则无疑是最主要的呈现方式。针对这些主观性文本信息，我们该如何更加有效的利用这些海量数据，从而提取出人们感兴趣的且携带个人情感的文本信息，是迄今为止自然语言处理领域中最为重要的研究课题之一[1]。例如，在购物网站上，许多潜在用户可以通过浏览某些商品的客户评论来判断自己是否应该继续购买该商品；在微博中，政府部门可以通过对社会热点事的相关民意信息挖掘、处理和分析，从而有效地对社会舆论的整体发展趋势进行舆情监控；商业上，也有很多的生产厂商通过挖掘和分析消费者对产品给予的相关评论对该产品的未来生产计划作出相应的决定等等。

最初基于词典和规则的方法被广泛应用于自然语言处理任务中，研究人员基于海量的词典手动编写复杂的规则以赋予计算机理解人类语言的能力。由于人类语言的复杂性，这种方法仅能在小规模的数据上取得一定的成果。但是随着文本数据数量日益增多，传统的基于词典和规则的方法不足以处理大量文本。90 年代以来，各种基于统计的机器学习方法在各个自然语言处理任务中取得了非常大的进步，例如朴素贝叶斯（Naive Baye, NB）、最大熵（Maximum Entropy, ME）,支持向量机（Support Vector Machine, SVM）等等。这些机器学习算法在 NLP 的各项任务中不断发展，但是由于大量采用了人工人工特征设计，近年来却遇到了瓶颈，其发展陷入了停滞的阶段。另外，这些方法实质上都属于浅层学习，函数模型和计算方法都不复杂，实现相对简单，导致这些算法在有限的样本和计算单元下无法表达复杂的函数，泛化能力较弱。深度学习致力于探索如何从原始数据中选择合适的特征以及表达方式来解决各种复杂的任务，相对于传统机器学习的算法来说，深度学习模型具有更强的表达能力。对于自然语言处理任务，基于深度学习的方法是一种截然不同的研究模式，这类方法在建模、解释和表达能力等方面的优势非常明显。使用深度学习模型从大规模语料库中挖掘文本所蕴含的更深层次的语言信息已经成为了自然语言处理的热门研究方向。

近年来，各种深度学习方法不断成熟并且在很多领域都取得了惊人成绩。其中，基于卷积神经网络（Convolutional Neural Network, CNN）和基于递归神经网络（Recursive Neural Network, RNN）的深度学习方法是当前自然语言处理领域的研究热点。神经网络是一种模拟人脑神经元分析学习机制的复杂网络模型，它们是一种含有多个隐含层的深度网络结构。基于神经网络的情感分析方法相比于传统的基于词典和规则和基于机器学习的分析方法有其独特的优势。这主要归因于深度网络结构的表达能力强，并且深度网络模型能够自动抽取文本中深层次的抽象特征。除此之外，通过深度学习得到的句子特征表示具有应用上的通用性，它不仅能够应用于情感分析，还能应用于主题发现[2]、关键词抽取、语义分析[3]等一系列自然语言处理问题。

情感分析，也被称为观点挖掘、意见抽取或情感分类，是指利用情感分析技术从原始语料中识别和抽取出主观性信息，然后对其进行分析、处理和归纳，以了解用户对某一事物的观点和评价，或者了解一篇文章的整体情感倾向性。文本情感分析任务最常见的是二级情感分类，即将目标文本划分成具有积极情感的正类和具有消极情感的负类。按照处理粒度的不同，情感分析可分为词语级、短语级、句子级、篇章级等几个研究层次。

本文主要研究基于深度学习的情感分析方法，旨在提高基于文本情感分析的舆情监控、用户评价分析等应用系统的性能。因此，本文的研究内容具有较高的科学研究意义和实际应用价值。

1.2 国内外研究现状

情感分析问题最早的起源是 Rosalind 等人[4]于 1997 年提出“情感计算”的概念。但是，目前公认的对情感分析问题进行了比较系统且全面研究的，是 Pang 等人[5]于 2002 年提出的利用有监督机器学习算法对电影评论进行情感分析。同年，Turney[6]提出了使用基于情感词典和人工判定规则的无监督学习算法对网络评论文本进行情感分析。直到目前为止，这两种方法仍是业界普遍采用的情感分析方法。但是，在 2006 年，机器学习领域的泰斗、神经网络研究方向的重要推动者、加拿大多伦多大学教授 Hinton 等人[7]在《科学》上发表了一篇里程碑式的文章，文章中采用无监督的深层神经网络来完成分类任务，并一举大获成功。从此，学术界和工业界纷纷开启了对深度学习的研究浪潮。

1.2.1 基于情感词典的无监督情感分析

无监督分析方法不需要使用任何已标记好的样本数据来训练分类器，该类方法通常需要情感词典配合人工判定规则才能发挥作用。情感词典中包含大量已整理并分类的常用情感词，一般分为积极情感词和消极情感词两类。基于已有的情

感词典，再结合人工制定的判定规则，即可对文本所表达的情感信息进行分析和分类。最简单的判定规则是统计一段文本中积极情感词与消极情感词的个数，假如积极情感词个数大于、等于、小于消极情感词个数，则原始文本的情感可分别判定为积极、中立、消极。

Turney[6]首先将“excellent”和“poor”分别作为积极、消极情感种子词，然后利用点互信息（Pointwise Mutual Information, PMI）的方法分别计算出候选词与积极、消极情感种子词之间的语义相似度，再求出这两个相似度值的差，即可得到候选词的语义倾向值（Semantic Orientation, SO），随后分别计算出评论文本中情感短语的平均 SO 值，若平均 SO 值大于 0，则将该条评论文本判为“推荐”，否则判为“不推荐”。但是 Alistair 等人[8]认为，不能仅考虑文本中情感词的数量，还必须考虑每个情感词在当前语境中的极性转变因子（Contextual Valence Shifter, CVS）对其的影响，包括否定词、程度副词等，因为否定词会将起到反转情感词极性的作用，而程度副词则会加强或削弱情感词语的语气强度。Maite 等人[9]采用了类似 Alistair 等人提出的方法，但是处理过程更为精细，每个情感词根据语气强度的不同，分别拥有数值大小不同的情感极性分值，否定词依然起到反转情感词极性的作用，但是程度副词需要根据其增强或削弱语气的程度而对应一个百分数值，然后根据该百分数值即可对情感词的极性分值进行增减。Jinan 等人[10]分别研究了两种不同的情感词典和三种不同的打分方式对推特文本进行情感分析实验，其中打分方式包括积极和消极情感词的数量之差、词频-反向文档频率（Term frequency-Inverse document frequency, TF-IDF）和潜在狄利克雷分配（Latent Dirichlet Allocation, LDA）策略。冯时等人[11]基于句法依存关系树和情感词典抽取了博文句子中以情感词为中心词的依存关系对，通过一些人工制定的判定规则来判别这些依存关系对所蕴含的情感倾向的极性值及其强度，并将其聚合为每个句子的情感倾向值，再对这些情感句子的情感倾向值进行加权求和计算，即可得到整个篇章的情感倾向分值。

基于情感词典和人工判定规则的无监督分析方法无需对训练数据进行人工标注，易于实现和理解，并且分类精度较高，易于领域扩展。但是，情感词典的覆盖范围很有限，它无法解决大量未登录词的问题，而且人为制定的判定规则也无法处理所有情况。另外，网络文本中含有越来越多的口语化词语、网络流行语等新词，然而这些新词也常用于表达情感，但却不包含在已有的情感词典中，并且网络文本的书写相对随意，省略句成分、中英文混合表达、语法错误、错别字等现象非常普遍。因此，仅仅采用基于情感词典和人工判定规则的无监督情感分析方法对网络文本进行情感分析并不是一个好的选择。

1.2.2 基于机器学习的有监督情感分析

目前，大多数的文本情感分析技术都是基于机器学习的有监督方法，这种方法往往是将情感分析当成一种特殊的分类问题，需要大量人工标注的语料作为训练集来训练分类器。

Pang 等人[5]分别采用三种分类模型对影评文本进行情感分析——支持向量机（Support Vector Machine, SVM）、朴素贝叶斯（Naïve Bayesian, NB）、最大熵（Maximum Entropy, ME），并研究了不同的分类特征对分类效果的影响。实验结果表明，将一元词特征（unigram）作为分类特征、SVM 作为分类器能够取得的最佳分类准确率为 82.9%。Dmitry 等人[12]将微博中独有的标签符号和表情符号作为情感标签以便抽取训练样本，训练一种类似 K 近邻（K-Nearest Neighbor，KNN）模型的分类器对微博语料进行细粒度的情感分析。Hiroshi 等人[13]在进行情感分析时考虑了三种不同的文本写作风格——一种正式风格和两种非正式风格，并针对这三种写作风格分别训练了不同的分类器。Fotis 等人[14]比较了两种不同的分类特征——内容特征和上下文特征，关于内容特征作者提出一种基于字的 n 元词特征(n-gram)的图模型以及标点符号模型，上下文特征则指需要同时考虑一条微博与其他微博之间的关系，并且提出了情感极性率（Polarity Ratio）的概念，情感极性率用于衡量一个微博语料集的整体情感倾向，以便能够利用用户的关注着的整体情感倾向、话题的整体情感倾向等上下文特征来辅助具体的情感分析任问题。Efstratios 等人[15]利用了领域本体进行细粒度情感分析，该方法能对某一产品的各方面属性进行具体的分析，从而获得用户对每个属性的评价信息。在中文微博情感分析领域，谢丽星等人[16]提出一种基于 SVM 的层次结构多策略分析方法，分别提出了一步三分法和二步分类法两种策略，并将微博文本中的超链接、表情符号、情感词语、情感短语、上下文特征（形容词、动词、问号、感叹号等）等作为分类特征，取得的最佳分类准确率为 65.7%。Wang 等人[17]利用了句法分析树找出文中的观点，并将每个观点表示成四元组的形式，通过计算每个观点的情感极性值得到整个句子的情感极性值，此外还计算了每个句子情感极性值的置信度，选取置信度高的情感句作为模型的训练集，最后采用聚类算法实现了情感分类。曹海涛[18]引入了心理学中常用的 PAD 情感分析模型，结合知网语义相似度的计算方法，构建出一种基于 PAD 模型的领域情感词典，并且提出了一种基于统计信息与语义信息的权重计算方案，在一定程度上消除了特征歧义对模型性能的不良影响，最后使用一种基于 PAD 情感语义特征的 SVM 模型进行了情感分析实验。有监督情感分析方法最大的缺点是需要大量人工标注的训练集，针对这个问题，Tan 等人[19]将基于机器学习方法和基于情感词典的方法结合起来，利用情感词典的方法确定未标注样本的情

感极性，并抽取出情感极性置信度较高的样本作为基于机器学习的模型的训练集。Zhang 等人[20]在使用情感词典方法抽取训练样本时，采用了更为细致、精确的情感极性分值的计算方法，并考虑了比较句、转折句、并列句等特殊情况，再使用 SVM 分类器进行情感分类，并且同时解决了基于情感词典方法的召回率低、基于机器学习方法需要大量人工标注这两个问题。

基于机器学习的有监督情感分析方法虽然取得了不错的成绩，但由于有监督学习过分依赖大量人工标注语料和人工特征的设计，使得基于有监督学习的系统需要付出很高的人工代价，因此使用基于机器学习的有监督情感分析方法对文本进行情感分析也不是最佳方案。

1.2.3 基于深度学习的半监督情感分析

近年来，深度学习（Deep Learning）模型在计算机视觉[21]和语音识别[22]邻域取得了巨大成果。深度学习模型通常是具有多层结构的神经网络，强调从大规模的数据集学习各种现实事物的特征表示，并且这些特征能够被计算机直接应用于各种计算模型中。当前在自然语言处理领域，大多数深度学习方法通过神经网络模型学习单词的词向量表达[23]，然后在此基础上再做进一步处理。

主流的神经网络模型是递归神经网络（Recursive neural network, RNN）和卷积神经网络（Convolutional neural network, CNN）。递归神经网络将文本看作是有序的词语序列，并且考虑了词语间的关联性，可以充分利用上下文信息从而更好地进行语言建模。RNTN 模型[24]利用情感树库在二元化的句法树结构上进行语义合成，Dong 等人[25]在此模型基础上，通过对依存句法树进行变换，完成句中指定评价对象的情感极性分析。Dong 等人[26]通过添加语义合成函数的 pooling 层，使得 RNTN 可以自适应地选择最适合当前节点的语义合成函数，从而大幅提高情感分类效果。Kalchbrenner 等人[27]提出一种 DCNN 模型，通过学习句子结构来提高情感语义合成性能，Santos 等人[28]提出 CharSCNN 模型，为英语单词增加了字符表示层的特征，可以提高对于英文 Twitter 中不规则单词的表示能力。Kim[29]首次尝试将 CNN 应用到句子分类（Sentence Classification）任务中，通过词向量训练工具 word2vec 生成的词向量训练一个简单的 3 层卷积神经网络，在二分类和多分类数据集上均取得了很好的分类性能，说明预先训练的词向量是一类通用的特征向量，可以在不同的分类任务中直接使用。另外 Kim[29]采用的 MCNN 也是 CNN 的一种，能够对多个窗口大小内的词向量执行语义合成操作，在多个不同的情感数据集上也取得了很好的分类性能。何炎祥等人[30]基于 MCNN，提出一种称为 EMCNN 的模型，将基于表情符号的情感空间映射与深度学习模型 MCNN 结合，有效增强了 MCNN 捕捉情感语义的能力。Peng 等人[31]提出一种 LSTM-CNN 模型，首次尝试从子句的角度来分析文本的

情感极性，利用 LSTM（长短时记忆网络）归纳某个子句的上下文信息，然后将上下文信息拼接在子句的词向量表达形式的两侧，进而再使用 CNN 对子句进行特征提取、分类，该模型在影评数据集上相比于传统的基于情感词典的分类方法和基于机器学习的分类方法取得了更好的分类效果。

1.3 本文的研究内容

针对文本情感分析的特点，本文对深度学习模型进行改进和调整以适应特定领域的情感分析问题。本文的主要创新点和工作：

（1）研究基于卷积神经网络和词语邻近特征的情感分析模型。由于当前几乎所有的基于卷积神经网络的情感分类模型，在卷积过程中每个词向量只能孤立的表征单个词语，并不蕴含上下文信息，这不利于信息传递的连续性，并且卷积操作在局部范围内可能会打乱词向量的序列性。针对这个问题，本文提出一种基于词语邻近特征（Adjacent Feature）的卷积神经网络模型，在卷积过程中让每个词向量携带邻近词语的特征，这样既保证信息传递的连续性也保证了词向量在局部范围内的序列性。实验结果表明，在 COAE2014、COAE2015 的情感分类数据集上的准确率分别达到了 89.43%和 85.61%，并且优于未考虑邻近特征的卷积神经网络模型，说明本文提出的该模型确实可行、有效。

（2）研究基于递归神经网络和人工判定规则的情感分析模型。传统的基于情感词典和人工判定规则的分析方法是从语言学的角度出发，但是这种方法需要制定大量的情感词典和判定规则。而基于递归神经网络的分析方法可以通过不断的编码和重构训练，从大量未标注的语料中学习到先验知识。本文在参考了这两种方法之后，提出了一种新颖的基于递归神经网络和人工判定规则的情感分析模型。首先利用递归神经网络并行计算出组成文本的多个子句的情感极性，然后根据极性融合规则将多个子句的情感极性进行融合，最后利用人工判定规则计算出原始文本的情感极性。该模型的优点在于：一方面递归神经网络中融入了人工判定规则，使得以往积累的人工经验得到有效利用；另一方面递归神经网络取代了情感词典，避免了情感词典的局限性。该模型在数据集 SST-C2 和 SST-C3 上的分类准确率分别达到了 87.8%、81.6%，并且各方面性能均接近或超越主流的分类模型，说明该模型不仅新颖，而且确实可行、有效。

（3）为了继续提高情感分析的效果，本文在第 5 章中提出了几点值得进一步探究的问题：1、在计算能力允许的情况下，是否可以通过扩大邻近词的数量从而获取更精确的上下文信息，值得探究；2、标点符号和表情符号必然会影响上下文的情感极性转移，因此考虑由标点符号和表情符号而引起的情感极性转移问题，值得进一步探究；3、分类数据集不平衡会导致模型无法学习到有效的先验知识，这会在测试阶段产生了大量误判。解决分类数据不平衡问题的方法可借

鉴“重采样技术”或“成本敏感学习（Cost Sensitive Learning, CSL）”；4、在第4章中，针对2分类和3分类数据集分别设计了简单、粗糙且类似的极性融合规则，实验结果差强人意。为了进一步提高模型的性能，是否可以考虑引入模糊评价法的模糊规则，使得子句的极性融合过程更加合理。

1.4 本文的组织结构

本文的组织结构共分为5个章节：

第一章，绪论。首先介绍了文本情感分析的研究背景及意义，详细阐述了文本情感分析领域的国内外研究成果，简要说明了主流的研究方法，并提出了本文的研究内容和方向，最后介绍了本文的组织结构。

第二章，研究基础。首先介绍了数据预处理的相关技术，接着介绍了什么是词向量及词向量生成模型，另外介绍了情感分析领域针对模型性能的通用评价指标，最后介绍了深度学习的起源及当前两种主流的深度学习算法。

第三章，基于卷积神经网络和词语邻近特征的情感分析模型。首先介绍了什么是词语的邻近特征以及如何利用邻近特征，然后详细介绍了所提出的卷积神经网络模型的结构，最后通过几组实验证明了所提模型的有效性。

第四章，基于递归神经网络和人工判定规则的情感分析模型。首先阐述了基于情感词典和人工判定规则的情感分析方法的不足，然后提出了一种结合递归神经网络与人工判定规则的方法，接着详细阐述了该模型的结构，最后通过几组实验证明了该模型的有效性。

第五章，结论和展望。对本文所做的工作进行了总结，对存在的问题和今后的研究工作做出展望。

第2章 相关技术与理论基础

2.1 引言

由于文本情感分析领域涉及到很多相关技术，本章将一一介绍。首先是数据预处理部分。众所周知，无论是微博文本还是各种评论性文本，其中都包含了大量没有太多意义的停用词和特殊符号，所以需要过滤。另外由于中文文本的特殊性，中文分词几乎是所有自然语言处理任务的必经步骤。紧接着介绍了什么是词向量、词向量的意义，以及它的生成模型。然后详细介绍了针对各种情感分类模型通用的性能评价指标。最后阐述了当前主流的两个深度学习算法，从人工神经网络的起源开始说起，逐步介绍了什么是卷积神经网络、什么是递归神经网络。

2.2 数据预处理

2.2.1 过滤停用词与特殊符号

停用词（Stop Words）是指一些没有实际含义的功能词，包括语气助词、虚词、副词、介词、连接词、代词等，中文停用词包括“的”、“了”、“是”、“在”等等，英文停用词包括比如“the”、“is”、“at”、“on”、等等。停用词在文本中的出现频率高，但没有太多的实际意义，研究者认为停用词增加了特征空间的规模，却对提高模型分类准确率无益[32]，所以本文采取了去除所有停用词的手段。

另外，对于语料中包含的特殊符号如“&”、“@”、“#”、URL 等等，有些研究者认为这些特殊符号没有多大的实际意义，也应该进行过滤；但也有些研究者认为这些特殊符号也能作为情感分类特征，例如文献[33]就在进行情感分析任务中考虑了训练文本中的“&”、“@”、“#”、URL 的数量对于模型最终分类效果的影响。但是本文作者经过多轮对比实验后，发现这些特殊符号对准确率的提高并无帮助，所以本文过滤了所有实验数据集中的特殊符号。

2.2.2 中文分词

一段文本是由许多词语构成的，词语是构成文本的基本单元，整段文本的语义也是由所有词语的组合表达的。因此，如果人们需要对一段文本的语义信息进行分析，无论是采用基于统计学习的方法，还是采用句式、句法分析的方法，把一段文本转化为词语的组合序列是一项必要的前提工作，这个转化的过程就称为分词。对于英文文本的处理，由于英文单词之间都是以空格作为分隔符，所以英

文文本本身就被分成了一系列词语的组合。但是，中文文本在这一点上却和英文不同，在中文文本中，中文词语之间是没有分隔符的，因此，对中文文本进行分词是一项必要的工作。

中文文本分词的算法有很多，总体上可以分为三类：基于词典的分词方法、基于理解的分词方法和基于统计的分词方法[34]。基于词典的分词方法是根据一定的规则把中文文本中的字符串和分词词典中的词语进行匹配，如果在分词词典中匹配到了某个词语，那么这个字符串就被分割开来。基于理解的分词方法依赖大量的语言知识和信息，具体做法是在分词的过程中对中文文本的句法和语义进行分析，通过分析后的结果确定词语表达的具体含义。基于统计的分词方法是对中文文本中汉字出现的组合信息进行统计分析，主要分析出现频率等信息判断该字符串是否为一个完整的词语，根据这些统计信息进行分词。三种中文分词的算法各有利弊，但是基于词典的分词算法技术相对成熟，并且在分词速度上远远快于另外两种算法，因此常常使用基于词典的分词算法或者将基于词典的分词算法和其他两种分词算法相结合，以获取更好的分词效果。由此可以看到，分词词典的词语数量和分词的效果是密切相关的。

目前，分词技术已经相当成熟，一些中文分词系统也已经投入使用。例如中科院计算技术研究所开发的中文分词系统 NLPIR（又名 ICTCLAS），该系统基于多层隐马尔科夫模型（HMM）构建，分词的准确率达到 97.58%，它的功能除了中文分词之外，还提供了命名实体识别、词性标注、添加用户词典、关键词识别、新词发现等功能，因此它非常适用于处理网络词汇较多的微博文本或评论文本。此外，由哈工大社会计算与信息检索研究中心和科大讯飞联合研发的 LTP 语言云也是常用的分词系统，它同时提供了依存句法分析、词性标注、命名实体识别、语义角色标注等功能。其他的开源中文分词工具还包括结巴（Jieba）中文分词、IKAnalyzer、庖丁解牛、Stanford Word Segmenter 等。

2.3 词向量

2.3.1 什么是词向量

自然语言处理问题要转化为机器学习问题，最重要的一步就是将文本转化为数学符号表示。One-hot 是传统的最简洁直观的词表示方法，文本中的每个词都表示为一个高维稀疏的向量，向量的维度就是词表的大小，所有维度中仅有一个维度的值为 1，该维代表了当前词。如词“汽车”和“卡车”可以表示为：

汽车：[0, 0, 0, 1, 0, 0, 0, 0, ...]

卡车：[0, 0, 0, 0, 0, 0, 1, 0, ...]

从以上两个向量中可以看出 One-hot 表示方法存在许多问题，主要有以下两个方面：

1．“词汇鸿沟”问题。在 One-hot 表示方法中，每个词的表示都是孤立的，无法判断出两个词之间的联系，如上例中的“汽车”和“卡车”是语义相关的，但从 One-hot 表示法中却无法判断出来。

2．“维度灾难”问题。词向量的维度等于词表的大小，如果词表的大小超过了预期，就会导致词的表示维度很高，因此无论应用在传统的机器学习方法中还是深度学习方法中，都会对模型的构建带来很大负荷。

相比较 One-hot 表示方法，将词转化为分布式表示的低维实数向量，即词向量（word vector），能够很好地解决上两个问题。词向量维度是预先定义的，每一维上的数值都为实数，例如：

汽车：[0.57, 0.31, -0.69, 0.88, 0.37, ...]

卡车：[0.62, 0.41, -0.57, 0.91, 0.34, ...]

分布式的表示方法将词的表示扩展到固定维度的空间内，一方面大大的缩减了词向量的维度，不再依赖于词表大小；另一方面，可将该词本身的语义信息嵌入到词向量中，使词向量空间变为语义空间。另外，语义上相近的词语在词向量的表达上也是相关的，例如上例，通过计算两个词向量的“余弦距离”即可判断出它们的相似度。

2.3.2 词向量训练模型

常用的词向量训练模型有两种，第一种是通过主题模型训练得到词向量，该类词向量含有一定的主题信息，另第一种是通过语言模型来训练词向量，该方法得到的词向量含有一定的语义信息。在情感分类任务中，第二种训练方法是最常用的也是目前为止公认的词向量标准训练模型，因此本节着重介绍基于语言模型的词向量训练模型。

1．基于马尔科夫假设的语言模型

语言模型（Language Model，LM）[35]通常用于建立一个能够描述给定词序列出现的概率分布，其形式化描述是：给定一个长度为 T 的字符串序列 S，其自然语言的概率为 $P(w_1,w_2,w_3,\cdots,w_T)$,其中 w_j 表示 S 中的第 j 个词，则：

$$P(w_1,w_2,w_3,\cdots,w_T)=\prod_{t=1}^{T}P(w_t\mid w_1,w_2,\cdots,w_{t-1}) \tag{2.1}$$

训练语言模型的目的就在于得到上述公式中的条件概率。但是当句子中词数较多时，模型的参数就会呈指数级增长。研究者们针对这一问题提出了基于马尔科夫假设的 n 元语言模型，即词序列中第 k 个词仅与第 k-n 到 k-1 的词相关。基于以上假设，句子 S 的概率就可表示为 $P(w_1,w_2,\cdots,w_T)=\prod_{t=1}^{T}P(w_t\mid w_{t-n}^{t-1})$。n 的取值越大，则其越接近原始的统计语言模型，但模型中的参数值也随机增加。假如词表大小为 30000，二元语言模型下，参数个数为 $30000*29999\approx9*10^8$，三元语

言模型参数个数达到 $2.7*10^{13}$。当 n 增长时，“维度灾难”问题依然没有解决。同样的，由于词表数量较大，大部分词序列出现的概率为 0，“数据稀疏”的问题同样存在。

2．基于神经网络的语言模型

使用神经网络来训练语言模型，能够很好地避免上述两个问题，并且词向量可以作为副产物被输出。比较常见的基于神经网络的语言模型框架有两个，分别是 Collobert[36]提出的 C&W 模型和谷歌研究人员 Mikolov[37]提出的 word2vec 模型。C&W 借鉴了 Bengio[38]在 2003 年提出的三层神经网络语言模型，其主要目的在于使用生成的词向量完成各种自然语言处理的任务，比如词性标注、命名实体识别、短语识别、关键词抽取、主题发现、语义角色标注等。相比于 C&W 模型，word2vec 模型框架更为简单，其内部不含非线性隐藏层，模型训练复杂度大大减小。具体来说，word2vec 框架中包含有两种不同的模型实现：CBOW 模型和 Skip-gram 模型。下面分别介绍这两种模型。

CBOW：该模型的核心思想在于，由当前词的上下文（C_{ij}）预测当前词（w_{ij}）。给定文档 D，CBOW 模型的目的在于使文档的后验概率最大：

$$arg\,max_{\theta}\prod_{j}^{D}\left[\prod_{ij=1}^{T_j}p\left(w_{ij}\mid C_{ij};\theta\right)\right] \tag{2.2}$$

其中 T_j 表示文档 D 中的第 j 个句子，w_{ij} 表示第 j 个句子中的第 i 个词，C_{ij} 为 w_{ij} 的上下文，θ 为后验概率参数。条件概率 $p\left(w_{ij}\mid C_{ij};\theta\right)$ 的计算可以分为以下两个部分：

（1）将词汇表中的词汇映射成词向量矩阵 $M\in\mathbb{R}^{N\times d}$，M 中的每个行向量对应不同的词语，N 为词表大小，d 为词向量的维度，我们需要训练的就是词向量矩阵 M。

（2）给定当前词 w_{ij} 的上下文 C_{ij}，通过神经网络模型最大化当前词的条件概率 $p\left(w_{ij}\mid C_{ij},\theta\right)$。

Mikolov[37]提出两种不同方式来构建模型以计算参数 θ，分别是层次 softmax（Hierarchical softmax[39]）和负抽样（Negtive Sampling[40]）。层次 softmax 的模型结构如图 2.1 所示。该模型总共含有 3 层：

（1）输入层：该层从词向量矩阵 M 中查找出当前词的上下文所对应的词向量 $\left[w_{i-h},w_{i-h+1},\cdots,w_{i-1},w_{i+1},w_{i+2},\cdots,w_{i+h}\right]\in\mathbb{R}^{2h*d}$，将其作为输入。如图 2.1 中的最上层所示。

（2）隐含层：w_{neu} 所在的层即为隐含层。在 CBOW 模型中，该层为输入层若干向量的累加和，表现形式是维度为 d 的向量。

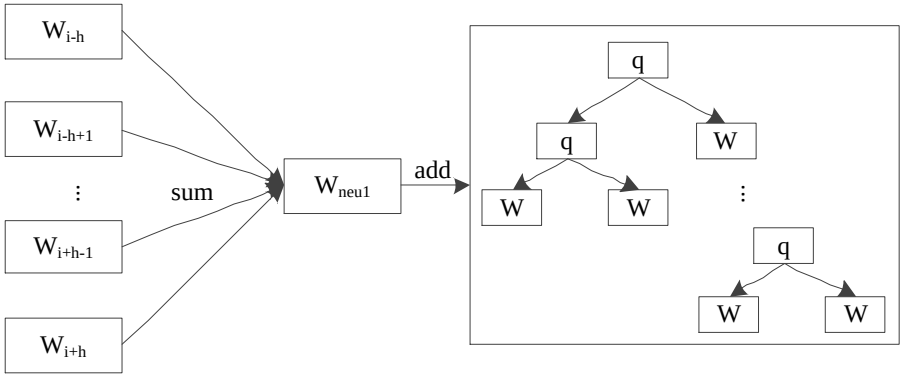

图 2.1 BOW 模型中层次 softmax 的结构

（3）输出层：图 2.1 中方框内所示即为模型输出层。与传统神经网络的输出层不同，该输出层由一个 Huffman 树构成。Huffman 树的每个叶子节点代表一个词向量，隐层的节点 w_{neu1} 与输出层二叉树的非叶结点 q 相连接，构成一颗新的 Huffman 树。其中，每一个非叶结点 q 也是一个向量，但是不代表某一个词语，代表的是某一类别。那么，当前词的条件概率可被表示为公式 2.3，其中 $q_{k_{ij}}$ 表示从 root 节点到目标叶子节点路径下的第 k 个非叶子节点向量；$d_{k_{ij}}$ 是二值表示的，为 0 时表示目标叶子节点在该非叶结点的左子树中，为 1 时表示目标叶子节点在该非叶结点的右子树中；σ 为 softmax 函数。

$$p_H\left(w_{ij}\,|\,C_{ij};\theta\right)=\prod_{k_{ij}=1}^{K_{ij}}\left(\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)^{1-d_{k_{ij}}}\cdot\left(1-\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)\right)^{d_{k_{ij}}}\right) \tag{2.3}$$

$$\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)=\frac{1}{1+e^{-q_{k_{ij}}\cdot C_{ij}}} \tag{2.4}$$

因此，模型的目标可W表示为最小化负对数似然函数：

$$L(\theta)=\prod_{j}^{S}\left(\prod_{i=1}^{T_j}p\left(w_{ij}\,|\,C_{ij}\right)\right) \tag{2.5}$$

$$-l(\theta)=-log\,L(\theta)=-\sum_{j=1}^{S}\left(\sum_{i=1}^{T_j}\left(\sum_{k_{ij}=1}^{K_{ij}}log\left(\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)^{1-d_{k_{ij}}}\cdot\left(1-\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)\right)^{d_{k_{ij}}}\right)\right)\right) \tag{2.6}$$

对于参数的更新，CBOW 模型使用随机梯度下降法（Stochastic gradient descent）。令：

$$f_H=-log\left(\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)^{1-d_{k_{ij}}}\cdot\left(1-\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)\right)^{d_{k_{ij}}}\right) \tag{2.7}$$

$$F_{Hq}\left(q_{k_{ij}}\right)=\frac{\partial f_H}{\partial q_{k_{ij}}}=-\left(1-d_{k_{ij}}-\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)\right)\cdot C_{ij} \tag{2.8}$$

$$F_{HC}\left(q_{k_{ij}}\right)=\frac{\partial f_H}{\partial C_{ij}}=-\left(1-d_{k_{ij}}-\sigma\left(q_{k_{ij}}\cdot C_{ij}\right)\right)\cdot q_{k_{ij}} \tag{2.9}$$

对应的非叶结点向量和输入词向量的迭代方法分别为公式 2.10、公式 2.11：

$$q_{k_{ij}}^{n+1}=q_{k_{ij}}^{n}-\eta F_{Hq}\left(q_{k_{ij}}^{n}\right) \tag{2.10}$$

$$w_1^{n+1}=w_1^{n}-\eta\sum_{k_{ij}=1}^{K_{ij}}F_{HC}\left(q_{k_{ij}}\right) \tag{2.11}$$

其中η为学习率，w_1表示模型输入词，在该迭代过程中仅更新输入词向量，即上下文词向量，而w_{ij}本身是不更新的。

负抽样的目的主要是为了节省计算量，该方法的本质在是将第 j 个句子的第 i 个词w_{ij}与其上下文作为正样本，将w_{ij}替换为其他的词w_k后对应的序列作为负样本。那么，其概率公式可表示为公式 2.12，于是负抽样的目标函数就可简化为最大化概率公式。

$$p_N\left(w_{ij}\mid C_{ij};\theta\right)=\sigma\left(w_{ij}\cdot C_{ij}\right)\prod_{k=1}^{K}\left(1-\sigma\left(w_k\cdot C_{ij}\right)\right) \tag{2.12}$$

假设：

$$f_N=-label\cdot log\,\sigma\left(w\cdot C_{ij}\right)-\left(1-label\right)\cdot\left(1-\sigma\left(w\cdot C_{ij}\right)\right) \tag{2.13}$$

其中正样本的 label 为 1，负样本的 label 为 0。对应的梯度计算方式如下：□□

$$F_{NW}\left(w\right)=\frac{\partial f_N}{\partial w}=-\left(label-\sigma\left(w\cdot C_{ij}\right)\right)\cdot C_{ij} \tag{2.14}$$

$$F_{NC}\left(w\right)=\frac{\partial f_N}{\partial C_{ij}}=-\left(label-\sigma\left(w\cdot C_{ij}\right)\right)\cdot w \tag{2.15}$$

在负抽样中，将所有词分为两部分，上下文词及中心词。和层次 softmax 类似，仅更新上下文词，而对于中心词及替代词的梯度，则保存在预定义的连接矩阵中。连接矩阵的迭代方式为公式 2.16，其中R_w表示连接矩阵中 W 对应的词向量，η表示学习率。上下文词的迭代方式为公式 2.17，其中w_1表示模型的输入词，

即当前词的上下文。

$$R_w^{n+1} = R_w^n - \eta F_w\left(R_w^n\right) \tag{2.16}$$

$$w_1^{n+1} = w_1^n - \eta\left(F_C\left(R_{w_1}^n\right) + \sum_{k=1}^{K} F_C\left(R_{w_k}^n\right)\right) \tag{2.17}$$

Skip-gram：不同于 CBOW 模型，Skip-gram 模型的核心思想是根据当前词预测上下文。该模型的目标在于最大化文档的后验概率：

$$arg\,max_\theta \prod_{w_{ij}}^{D}\left[\prod_{c\in C_{ij}} p\left(c \mid w_{ij};\theta\right)\right] \tag{2.18}$$

与 CBOW 类似，在 Skip-gram 模型中也实现了两种参数训练方法（层次 softmax 和负抽样）。

图 2.2 为 Skip-gram 模型的层次 softmax 方法。该模型与 CBOW 模型的层次 softmax 方法不同的是，它是根据当前词预测周围词，当能成功预测到周围词时，再直接将当前的输入词添加到输出层的 Huffman 树上。而负抽样方法与 CBOW 模型中的负抽样类似。

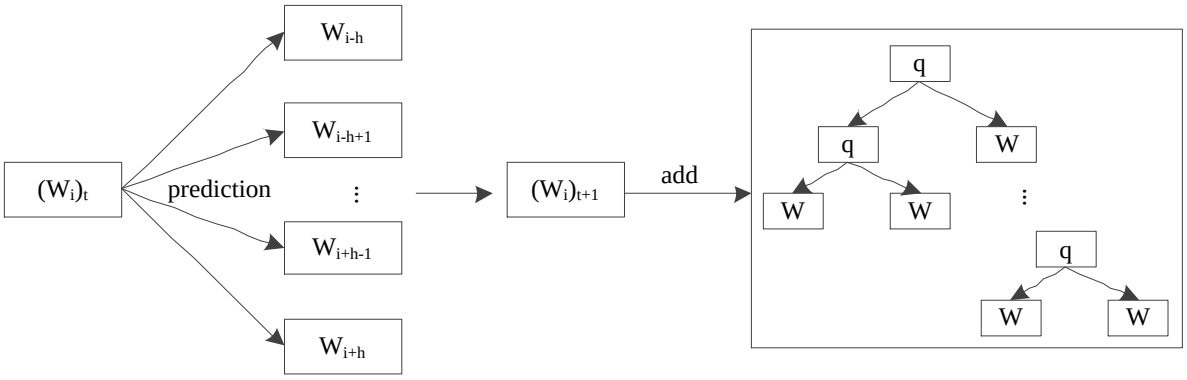

图 2.2 Skip-gram 模型中层次 softmax 的结构

公式 2.19 和公式 2.20 分别为 Skip-gram 模型下的层次 softmax 及负抽样算法的概率计算公式。其中 w_{ij} 表示当前词，而 c 表示上下文窗口内的词。

$$p_H\left(c \mid w_{ij};\theta\right) = \prod_{k_{ij}=1}^{K_{ij}}\left(\sigma\left(q_{k_{ij}}\cdot c\right)^{1-d_{k_{ij}}}\cdot\left(1-\sigma\left(q_{k_{ij}}\cdot c\right)\right)^{d_{k_{ij}}}\right) \tag{2.19}$$

$$p_N\left(c \mid w_{ij};\theta\right)=\sigma\left(w_{ij}\cdot c\right)\prod_{k=1}^{K}\left(1-\sigma\left(w_k\cdot c\right)\right) \tag{2.20}$$

假设：

$$f_H=-log\left(\sigma\left(q_{k_{ij}}\cdot c\right)^{1-d_{k_{ij}}}\cdot\left(1-\sigma\left(q_{k_{ij}}\cdot c\right)\right)^{d_{k_{ij}}}\right) \tag{2.21}$$

$$f_N=-label\cdot log\,\sigma\left(w\cdot c\right)-\left(1-label\right)\cdot\left(1-\sigma\left(w\cdot c\right)\right) \tag{2.22}$$

公式 2.22 中，正样本的 label 为 1，负样本的 label 为 0。层次 softmax 方法中非叶结点以及输入词向量的更新方式分别为公式 2.25 及公式 2.26，负抽样方法中的连接矩阵及输入词向量的迭代方式分别为公式 2.29 和 2.30。其中 R_w 表示连接矩阵中 w 对应的词向量，η 表示学习率，c 表示当前词的上下文窗口内的词。

$$F_{Hq}\left(q_{k_{ij}}\right)=\frac{\partial f_H}{\partial q_{k_{ij}}}=-\left(1-d_{k_{ij}}-\sigma\left(q_{k_{ij}}\cdot c\right)\right)\cdot c \tag{2.23}$$

$$F_{Hc}\left(q_{k_{ij}}\right)=\frac{\partial f_H}{\partial c}=-\left(1-d_{k_{ij}}-\sigma\left(q_{k_{ij}}\cdot c\right)\right)\cdot q_{k_{ij}} \tag{2.24}$$

$$q_{k_{ij}}^{n+1}=q_{k_{ij}}^{n}-\eta F_{Hq}\left(q_{k_{ij}}^{n}\right) \tag{2.25}$$

$$c^{n+1}=c^{n}-\eta F_{Hc}\left(q_{k_{ij}}^{n}\right) \tag{2.26}$$

$$F_{Nw}\left(w\right)=\frac{\partial f_N}{\partial w}=-\left(label-\sigma\left(w\cdot c\right)\right)\cdot c \tag{2.27}$$

$$F_{Nc}\left(w\right)=\frac{\partial f_N}{\partial c}=-\left(label-\sigma\left(w\cdot c\right)\right)\cdot w \tag{2.28}$$

$$R_w^{n+1}=R_w^{n}-\eta F_{Nw}\left(R_w^{n}\right) \tag{2.29}$$

$$c^{n+1}=c^{n}-\eta\left(F_{Nc}\left(R_{w_1}^{n}\right)+\sum_{k=1}^{K}F_{Nc}\left(R_{w_k}^{n}\right)\right) \tag{2.30}$$

以上我们介绍了两种基于语言模型的词向量，该类模型训练得到的词向量蕴含一定的语义信息。相似上下文的词用相似的词向量表示，例如“汽车”、“卡车”对应词向量的余弦相似度很高，非常适用于各种自然语言处理的任务中。

2.4 评价指标

在情感分类任务中，通常使用四个指标来评价分类的效果，分别是查准率（Precision）、召回率（Recall）、F1 值（F1-score）、准确率（Accuracy）。对于二分类任务，需要四个参数来计算这四个评价指标值，分别是 tp（true positive）表示判断为正面情感且实际是正面情感的测试样本数量，tn（true negative）表示判断为负面情感且实际是负面情感的测试样本数量，fp（false positive）表示判断为正面情感但实际是负面情感的测试样本数量，fn（false negative）表示判断为负面情感但实际是正面情感的测试样本数量。

2.4.1 查准率

查准率（Precision）分为积极查准率和消极查准率。积极查准率反映测试样本中被模型判定为积极情感且实际也是积极情感的样本数量占所有被模型判定为积极情感的样本总数的比例；消极查准率反映测试样本中被模型判定为消极情感且实际也是消极情感的样本数量占所有被模型判定为消极情感的样本总数的比例，积极、消极查准率分别由式 2.31、2.32 计算：

$$Precision_{pos} = \frac{tp}{tp + fp} \tag{2.31}$$

$$Precision_{neg} = \frac{tn}{tn + fn} \tag{2.32}$$

2.4.2 召回率

召回率（Recall）也分为积极召回率和消极召回率。积极召回率反映测试样本中被正确判定的积极情感的样本占所有积极测试样本的比例；消极召回率反映测试样本中被正确判定的消极情感的样本占所有消极测试样本的比例，积极、消极召回率分别由式 2.33、2.34 计算：

$$Recall_{pos} = \frac{tp}{tp + fn} \tag{2.33}$$

$$Recall_{neg} = \frac{tn}{tn + fp} \tag{2.34}$$

2.4.3 F1 值

F1-值（F-score）也分为正向 F1-值和负向 F1-值，它同时考虑了查准率和召回率两个指标，可视为是查准率和召回率的加权平均值，可由式 2.35、2.36 计算：

$$F1-score_{pos}=\frac{2\times precision_{pos}\times recall_{pos}}{precision_{pos}\times recall_{pos}} \tag{2.35}$$

$$F1-score_{neg}=\frac{2\times precision_{neg}\times recall_{neg}}{precision_{neg}\times recall_{neg}} \tag{2.36}$$

2.4.4 准确率

准确率（Accuracy）反映了分类器对全部样本的判定能力，评价的是整体的预测值与实际值之间的接近程度，计算的方式是判断正确的样本数目在所有样本中所占的比例，由式 2.37 计算：

$$Accuracy=\frac{tp+tn}{tp+fp+tn+fn} \tag{2.37}$$

上面介绍了二分类的查准率、召回率、F1-值、准确率的计算方法，而对于多分类相关指标的计算，仅与二分类在查准率、召回率上的计算稍有不同。当多分类时，公式 2.31、2.32 中的 *fp*、*fn* 应分别由式 2.38、2.39 计算：

$$fp=\sum_{c\neq pos}^{C} fp(c) \tag{2.38}$$

$$fn=\sum_{c\neq neg}^{C} fp(c) \tag{2.39}$$

同理，式 2.33、2.34 中的 *fp*、*fn* 也由上式计算。除此之外，多分类中各个分类的 F1-值以及总体的准确率的计算方式依然与二分类相同。

2.5 深度学习算法

2.5.1 起源——人工神经网络

人工神经网络（Artificial Neural Netwoks, ANN）起源于美国心理学家 Mcculloch 和数学家皮兹 Pitts 于 1943 年共同提出的 M-P 模型，该模型中以神经元作为功能逻辑器件来实现算法[41]。此后，人们对人工神经网络开展了大量研究。

作为一种模拟人脑模型特性的计算结构模型，人工神经网络试图以人脑的方

式对复杂的信息进行存储并处理，具有并行分布处理数据、高容错性、智能化和自学习能力等特征。人工神经网络本质上是由大量简单的神经元相互连接而构成的复杂网络。神经元是人工神经网络中最基本的计算单元，虽然单个神经元的结构以及计算方式并不复杂，但是大量神经元共同构成的神经网络能够拥有非常丰富的表达能力。每个输入神经元基本结构如图 2.3 所示：

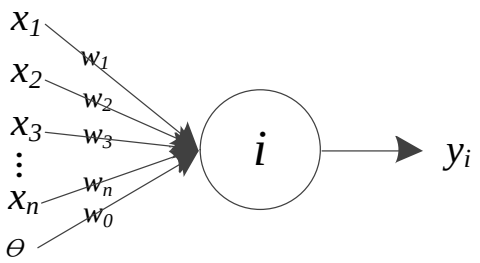

图 2.3 人工神经元结构

其中，$x_1,x_2,\ldots,x_n$ 是神经元 i 的多个输入（从其他多个神经元传输过来的值），θ 被称为阈值（threshold）或偏置（bias），用来模拟细胞膜在去极化成都超过某个阈值电位时才会被激活的状况,$w_1,w_2,\ldots,w_n$ 代表了每个输入与神经元 i 之间对应的权值，神经元的输出 y_i 可以通过以下公式求得：

$$y_i = f\left(\sum_{j=1}^{n} w_i x_j - \theta\right) \tag{2.40}$$

如果将阈值 θ 当做神经元输入 x_0，并且该阈值单元对于的权值为 w_0，则公式 2.40 可以简化为：

$$y_i = f\left(\sum_{j=0}^{n} w_i x_j\right) \tag{2.41}$$

在公式 2.41 中，f 代表非线性的激活函数（Activaition Function），线性网络中引入非线性的因素，可提高整个网络的表达能力。常用的激活函数有 Sigmoid 函数、Tanh 函数、Relu 函数等。Sigmoid 函数的表达式如下：

$$f(x) = \frac{1}{1+e^{-x}} \tag{2.42}$$

它的定义域为(-∞,+∞)，值域为(0,+1)。

Tanh 函数的表达式如下：

$$f(x) = tanh\,x = \frac{sinh\,x}{cosh\,x} = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2.43}$$

它的定义域为(-∞,+∞)，值域为(-1,+1)。

Relu 函数的表达式如下：

$$f(x)=max(0,x) \tag{2.44}$$

它的定义域为(-∞,+∞)，值域为(0,+∞)。

Sigmoid 函数、Tanh 函数、Relu 函数曲线图如图 2.4 所示：

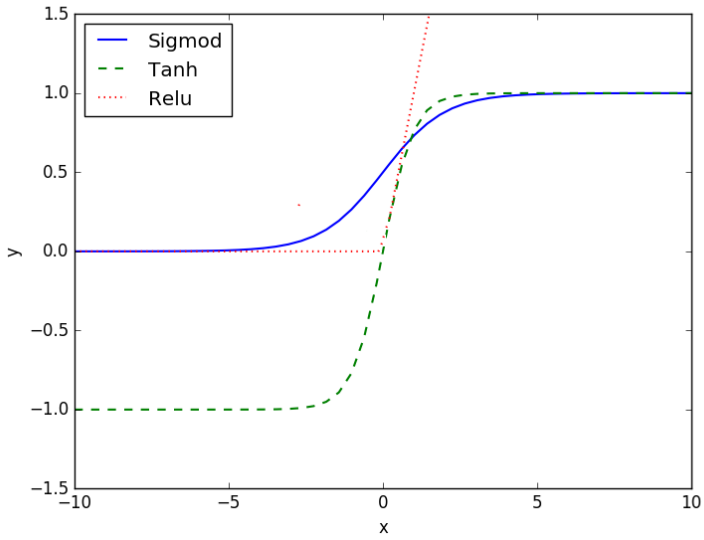

图 2.4 Sigmoid 函数、Tanh 函数、Relu 函数的曲线图

由人工神经元、输入层、隐含层、输出层、激活函数等共同构成了人工神网络，使得人工神经网络具有了逼近任意一个非线性函数或者表达某种逻辑策略的能力。一个典型的人工神经网络的结构如图 2.5 所示：

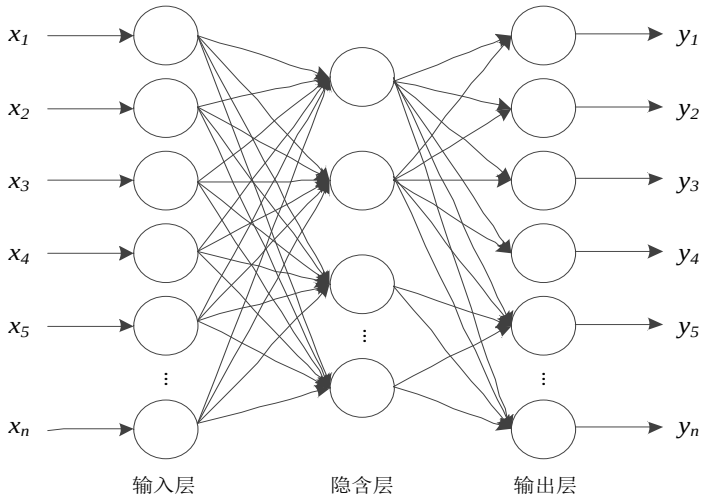

图 2.5 人工神经网络结构

经过多年发展，神经网络的功能以及表达能力得到了不断地加强，常用的神经网络有后向传播（Back Propagation, BP）神经网络[42]，径向基（Radial basis function, RBF）神经网络[43]，多层感知器[44]（Multilyaer Perceptron, MLP），自组织（Self-organization maps, SOM）神经网络[45]等等。但是这些

传统的神经网络在层数变多后遇到了以下问题：

（1）缺少大规模的数据。多层神经网络包含成千上万个参数需要训练，容易在小数据集上出现过拟合现象。

（2）模型训练时间过长，难以获得实际应用。

（3）受搜索算法控制，容易陷入局部极小点。

（4）梯度扩散（Gradient Diffusion）。梯度会随着网络层数的增加而逐步消失，无疑增加了模型的训练难度。

为了解决传统神经网络中出现的问题，近年来研究者们提出了多种深度学习的算法，并且伴随着新的网络架构以及计算能力增强的计算机硬件配置，深度学习在很多领域获得了突破性的进展。

2.5.2 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是一前馈网络，受 Hubel 和 Wiesel 对猫视觉皮层研究结果的启发[47]，该网络内神经元的连接模式与传统的神经网络的全连接方式不同，每个神经元的输入与上一层的局部感受野（Receptive filed）相连接，同时还借鉴了多层感知机中预处理少量数据的机制。卷积神经网络在图像识别[48,49]、推荐系统[50]以及自然语言处理[51,53]中获得了很多应用。目前来看，卷积神经网络在自然语言处理领域内的应用最为广泛，本节简要介绍 CNN 模型的相关知识。

传统的神经网络可分为 3 个部分，即输入层、隐藏层、输出层。其中，隐藏层可扩展为一种多层结构，相邻两层之间的神经元相互连接，而层内部神经元之间没有连接。最为经典的多层感知机（Multilayer perceptron）是上下文神经元全部相连的神经网络，而 CNN 通过变换隐藏层结构，将相邻神经元之间的交互信息加入到隐藏层中，并且通过降采样的方式，抽取出任务相关的信息。典型的卷积神经网络模型结构如图 2.6 所示，模型组成共分为五个部分：输入层、卷积层、池化层（下采样层）、全连接层和输出层。

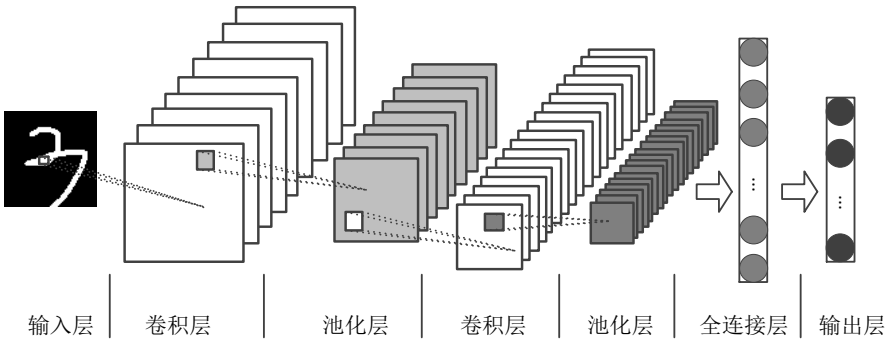

图 2.6 典型的卷积神经网络结构

通常，卷积神经网络的输入是原始图像的经过向量化后的矩阵 X，这里卷积神经网络第 i 层的特征图用 H_i 表示（H_0=X）。假设 H_i 是卷积层，那么 H_i 的计算过程可描述为：

$$H_i = f\left(H_{i-1} \otimes W_i + b_i\right) \tag{2.50}$$

其中，W_i 表示第 i 层卷积核的权重矩阵；运算符号“⊗”表示卷积核与第 i-1 层的特征图发生卷积操作，卷积结果与第 i 层的偏置向量 b_i 相加，最终通过非线性的激励函数 f(x)即可得到第 i 层的特征图 H_i。

池化层紧接在卷积层之后，网络按照一定的池化规则对特征图进行池化。池化层的作用主要有两点：筛选出特征图中最重要的特征；在一定程度上保持特征的尺寸不变性。假设 H_{i+1} 是池化层：

$$H_{i+1} = pooling\left(H_i\right) \tag{2.51}$$

输入数据在经过多个卷积层和池化层的交替操作后，卷积神经网络通过全连接层（Softmax）对提取出的特征进行分类，即可得到一个分类的概率分布图 Y（l_i 表示第 i 个标签类别）。卷积神经网络的本质是使输入矩阵（H_0）经过多个层次的数据变换或降维，映射到一个新的特征表达的数学模型，如式 2.52 所示：

$$Y_i = P\left(L = l_i \mid H_0; (W,b)\right) \tag{2.52}$$

卷积神经网络的训练目标是最小化模型的损失函数 $L(W,b)$达到一个收敛值。网络的输入 H_0 在经过前向传播后，通过计算损失函数，得到输入数据对应的实际值与期望值之间的差，称为“残差”。常用的损失函数有均方误差函数（Mean Squared Error，MSE）、负对数似然函数（Negative Log Likelihood，NLL）[54]等，如式 2.53、2.54 所示：

$$f_{MSE}(W,b) = \frac{1}{|Y|}\sum_{i=1}^{|Y|}\left(Y(i) - \hat{Y}(i)\right)^2 \tag{2.53}$$

$$f_{NLL}(W,b) = -\sum_{i=1}^{|Y|} log\, Y(i) \tag{2.54}$$

为了减轻神经网络训练极易出现过拟合的问题，损失函数在定义时通常会增加 L2 正则化约束以控制网络权重向量的过拟合，并且通过调整参数 λ（权值衰减）以控制过拟合作用的强度：

$$E(W,b) = L(W,b) + \frac{\lambda}{2} W^T W \tag{2.55}$$

卷积神经网络训练过程中，常用的优化方法是随机梯度下降（Stochastic Gradient Descent, SGD）法。网络通过随机梯度下降法将残差反向传播，逐层

更新网络的各层权重矩阵和偏置向量（W 和 b）。此外，学习率参数（η）用于控制残差反向传播的强度，如式 2.56、2.57 所示：

$$b_i = b_i - \eta \frac{\partial E(W,b)}{\partial b_i} \tag{2.56}$$

$$W_i = W_i - \eta \frac{\partial E(W,b)}{\partial W_i} \tag{2.57}$$

综上所述，我们可以看出卷积神经网络的工作原理主要分为模型的定义、训练以及预测三个部分：

1．模型的定义。网络结构的定义需根据具体任务的数据量及数据本身的特点进行设计，比如网络的深度、每一层的功能及模型的超参数，如λ、η等。对于卷积神经网络模型的设计已有不少相关研究，比如网络深度、卷积的步长、卷积核函数方面等等。此外，对于网络中的超参数选择，也存在一些有效的经验总结。但是，目前针对网络模型的理论分析和量化研究相对还比较匮乏。

2．模型的训练。卷积神经网络通过将残差反向传播的方式以对网络的参数进行训练。但是，网络训练过程中的过拟合、梯度消逝与爆炸等问题对网络的收敛性能造成了极大影响。因此，一些研究者提出了有效的改善方法：基于高斯分布的随机初始化网络参数[52]；利用经过预训练的网络参数对特定模型进行初始化。根据目前的研究趋势，卷积神经网络模型的规模正在迅速增大，而更加复杂的网络结构对相应的训练策略也提出了更高要求。

3．模型的预测。卷积神经网络的预测过程即是通过对输入数据进行前向传导，各层分别输出对应的特征图，最后利用全连接层（Softmax）输出基于输入数据的条件概率分布。目前的研究表明，卷积神经网络经过前向传导得到的高层特征具有很强的判别能力和泛化性能[53]；另外，通过迁移学习，这些特征可以直接被应用到更加广泛的领域。

2.5.3 递归神经网络

近年来，递归神经网络（Recursive Neural Networks, RNN）在比如语音识别、图片处理、自然语言处理等多个领域都取得了一定的成功。该模型最早由 Goller 等人[46]在 1996 年提出，该模型使用确定的树形结构并且词向量采用使用 one-hot 的表示方法。现在使用的 RNN 网络结构在底层通常使用通过 word2vec 训练而成的词向量表示，递归时所采用的树形结构既可以是由句子的句法结构转换而成，也可以由智能算法自动生成。RNN 可以有效利用结构化数据，例如有向无环图、二叉树等。RNN 可根据图结构自底向上适用相同的权重系数来构建神经网络。通过自底向上的遍历拓扑结构中的所有节点，即可得到根节点（如图

2.7 的节点）的向量化表示。通常有向无环图并没有具体的结构限制（例如多叉树），但是为了在进行节点结合时权重的一致性一般将句子的拓扑结构限定为二叉树，如图 2.7 所示，这样可以保证在进行遍历时结构的统一性。RNN 的结构如图 2.7 所示：

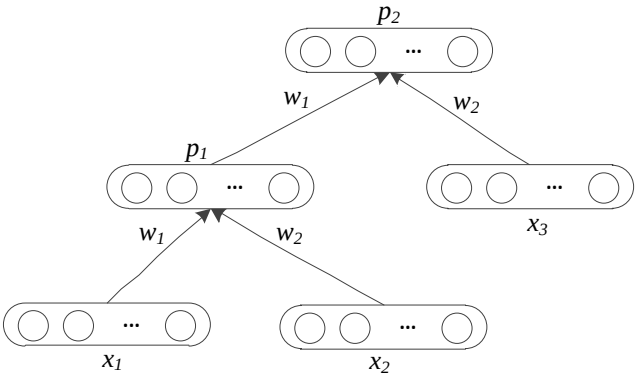

图 2.7 RNN 的网络结构

假定有句子文本 $S=\left(x_{i-1}x_ix_{i+1}\right)$，可以利用句法分析器得到其句法结构（通常句法结构为多叉树结构），将得到的句法结构二叉树化，则可得到新的句法结构 $S^{'}=\left(p_{i+1}\left(p_i\left(x_{i-1}x_i\right)x_{i-1}\right)\right)$。其中 $x_{i-1},x_i,x_{i+1}\in\mathbb{R}^d$ 分别表示句子中的词语，由多个词语组成的短语 $x_{i-1}x_i$ 可由节点 $p_i\in\mathbb{R}^d$ 表示，整个文本 S 则可以由 $p_{i+1}\in\mathbb{R}^d$ 节点表示，利用 RNN 网络结构学习文本的特征过程可用公式 2.45、2.46 表示：

$$p_i=g\left(W\begin{bmatrix}x_{i-1}\\x_i\end{bmatrix}+b\right) \tag{2.45}$$

$$p_{i+1}=g\left(W\begin{bmatrix}p_i\\x_{i+1}\end{bmatrix}+b\right) \tag{2.46}$$

其中 $W_1,W_2\in\mathbb{R}^{d\times d}$ 为权重矩阵，激活函数 g 可选取 $tanh$（双曲正切函数）。递归神经网络在学习句子文本的特征表示时，通过不断的合并词语或短语特征值（特征表示），从而可以学习到包含更多词语的短句的特征值。网络在学习特征表示的过程中，依照句子的二叉树拓扑结构自底向上的合并节点，这样可以根据句子的语法、语义特征及句子的层次结构关系更好的学习到句子的特征。

2.5.4 长短时记忆神经网络

对于传统的递归神经网络来说，它很难学习到长距离依赖信息。理论上，递归神经网络可以通过多次迭代训练挑选参数解决长距离依赖信息丢失的问题，但是实践过程中效果往往很难达到理想的效果。长短时记忆（LSTM）模型是 Hochreiter 等人[41]于 1997 年提出，并且在近期由 Graves[42]进行了改进和应用。

近些年来，在很多领域上 LSTM 都取得了巨大成功[43,44]。LSTM 模型能够帮助递归神经网络克服它的缺点，并且在很多任务上都取得了不错的效果。

LSTM 模型提出了一个新的结构叫做记忆存储单元（memory cell），该单元结构如图 2.8 所示。记忆存储单元由四个主要部分组成：输入门、自循环连通结点、遗忘门、输出门。自循环连通结点的权重为 1.0，并且保证除去外界干扰外，此记忆存储单元在时间步长从一个变化到另一个时能够保持不变。门的作用是调节记忆存储单元和其所处环境的相互作用。一方面，输入门能够允许输入信号改变记忆存储单元的状态或者阻止输入信号。另一方面，输出门能够允许记忆存储单元作用于另一个神经元或者阻止它对其他神经元的影响。最后，遗忘门能够调节记忆存储单元的自循环连接，它能够根据需要允许记忆存储单元记住或遗忘其先前的状态。

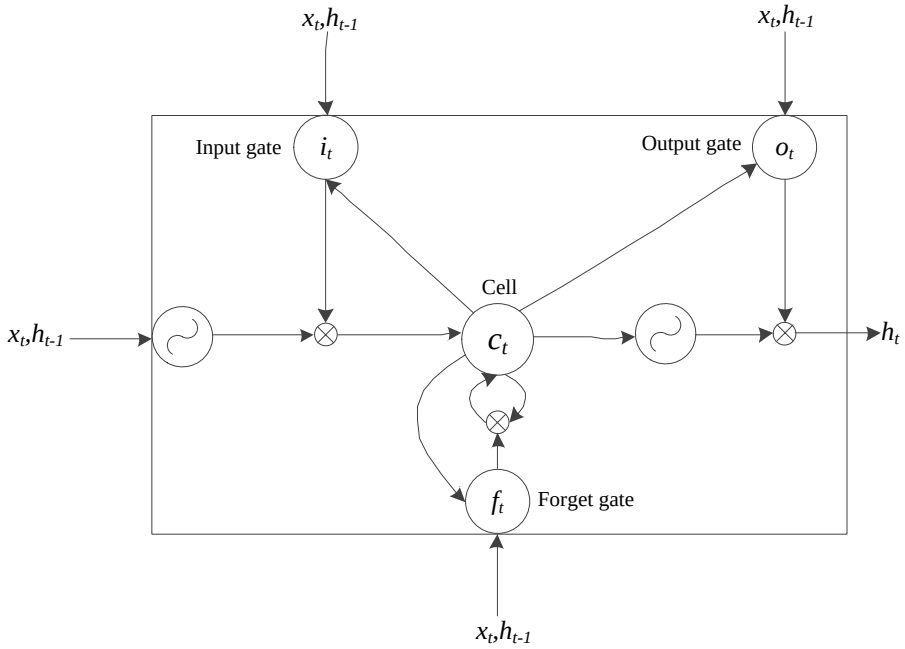

图 2.8 LSTM 单元结构

下面将通过公式形式化地描述记忆存储单元中的网络层是如何在每个时间步 t 上进行更新的。首先，计算在时间步 t 上输入门的启动值 i_t 和记忆存储单元状态的候选值 c_t：

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{2.47}$$

$$c_t = tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{2.48}$$

然后，计算在时间步 t 上记忆存储单元中遗忘门的启动值 f_t：

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2.49}$$

在计算完输入门的启动值 i_t、遗忘门的启动值 f_t 以及记忆存储单元状态的候选值 c_t 后，下一步可以计算记忆存储单元在时间步 t 上新的状态值 c_t：

$$c_t = i_t \times c_t + f_t \times c_{t-1} \tag{2.50}$$

最后，计算输出门的值和记忆存储单元的输出值：

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + V_o c_t + b_o\right) \tag{2.51}$$

$$h_t = o_t \times tanh\left(c_t\right) \tag{2.52}$$

在上面的公式中：x_t 表示时间步 t 时，记忆存储单元的输入；h_t 表示时间步 t 时，记忆存储单元中隐藏层的值；σ 表示 Sigmoid 函数，输出 0 到 1 之间的值；矩阵 U 为输入层到隐藏层的参数；矩阵 V 为隐藏层到输出层的参数；矩阵 W 为隐藏层到隐藏层的参数；b 为 Sigmoid 函数的偏置。

2.6 小结

本章介绍了关于情感分析的相关技术基础。首先介绍了数据预处理，详细阐述了中、英文数据集的特点及对应的处理方式。介绍了什么是词向量，阐述了 one-hot 词向量的缺点，然后介绍了两种生成词向量的语言模型，并详细介绍了 word2vec 框架。接着介绍了模型的通用评价指标。最后从神经网络的起源讲起，描述了传统神经网络的缺点，逐渐引出了两个主流的深度学习算法，即卷积神经网络和递归神经网络，并详细阐述了它们的优缺点、模型结构以及数学原理。

第3章 基于卷积神经网络和词语邻近特征的情感分析模型

3.1 引言

目前，基于卷积神经网络的情感分类方法得到了广泛关注。卷积神经网络通过多个并行的卷积操作获取文本的局部特征[55]，经过池化层再将筛选出的局部特征拼接然后传递给全连接层进行分类。然而大多数的卷积神经网络模型在卷积操作和特征拼接环节中，很大程度上忽略了文本中词语的有序性即词语之间的序列性，因此它不能保证信息传递的连续性。例如，“我/喜欢/自然/语言/处理/但是/它的/学习/门槛/很高”，假设卷积核的尺寸是 5，则将被卷积的分片依次是“我/喜欢/自然/语言/处理”、“喜欢/自然/语言/处理/但是”、“自然/语言/处理/但是/它的”、“语言/处理/但是/它的/学习”、“处理/但是/它的/学习/门槛”、“但是/它的/学习/门槛/很高”，那么得到的局部特征信息可能是“喜欢”、“自然”、“语言”、“但是”、“门槛”、“很高”，当然也可能是“自然”、“语言”、“它的”、“处理”、“学习”、“很高”，前一种明显表达的是负面情感，而后一种则很可能被误判为正面情感，因为“很高”带有积极情感。现实生活中，人们也常使用一些脱离语境的极性词来评估情感，比如“大”、“小”、“高”、“低”等，然而这些词在不同的语境中有着不同的含义，因此我们无法仅凭某些极性词就能分辨出整个文本的情感倾向，而必须要考虑语境。为了解决上述问题，本文提出一种基于词语邻近特征（Adjacent Feature）的卷积神经网络模型，在卷积过程中让每个词向量携带邻近词语的特征，这样既保证信息传递的连续性也保证了词向量在局部范围内的序列性。最后，本文分别在 COAE2014（2 分类）、COAE2015（3 分类）任务 4 的数据集上进行了实验，准确率分别达到了 89.43%、85.61%，并且优于未考虑邻近特征的基准神经网络模型。

3.2 词语的邻近特征

词语的邻近特征是指在当前上下文语境中，与当前词语相邻的词语的特征。这里假设原始句子为：

$$X = (w_1, w_2, w_3, \ldots, w_{i-1}, w_i, \ldots, w_n)^T \tag{3.1}$$

对应的词向量矩阵为：

$$M = (v_1, v_2, v_3, ..., v_{i-1}, v_i, ..., v_n)^T \tag{3.2}$$

其中，w_i表示组成句子的各个词语，v_i表示w_i对应的词向量特征。那么在卷积过程中，对于某个词语w_i我们采取 3 种不同的邻近特征附加策略：附加左邻近特征（Left Adjacent Feature, LAF）、附加右邻近特征（Right Adjacent Feature, RAF）、附加左右邻近特征（Left Right Adjacent Feature, LRAF），它们的作用都是为了增强词语在特定上下文中所表达的语义。如果使用 LAF，则卷积过程中使用词向量M_{LAF}，即从第一个词向量开始，每一个词向量都会与其左侧的词向量特征进行合并操作，这样做的好处是前一个词向量的信息可以传递给下一个词向量，因此下一个词向量在参加卷积操作时可以提供相对充足的情感信息供卷积层提取到有效的特征。M_{RAF}中每一个词向量与其右侧的词向量特征合并，M_{LRAI}中每一个词向量与其左、右侧的词向量特征合并。这三种策略的直观描述如图 3.1 所示，其中v_0表示零向量。

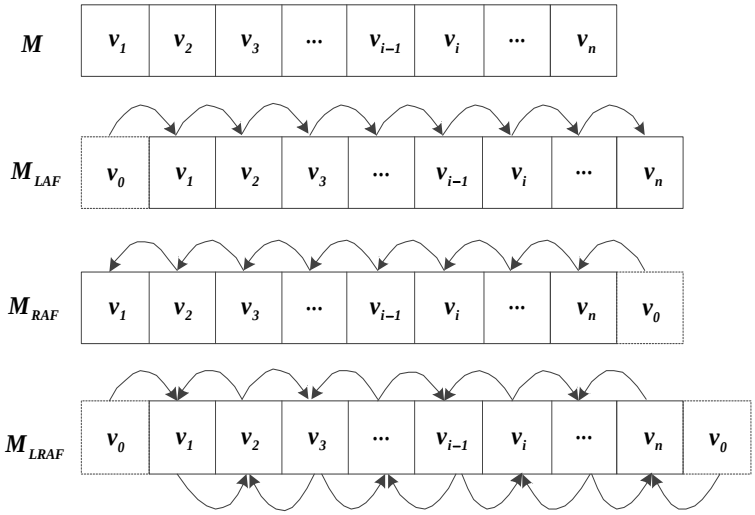

图 3.1 三种词语邻近特征附加策略

3.3 基于词语邻近特征的卷积神经网络模型

3.3.1 模型的结构

基于词语邻近特征的卷积神经网络模型如图 3.2 所示。图 3.2（图中使用的是 LRAF 附加策略）中使用的卷积核有三个尺寸：2、3、4，每个尺寸的卷积核有 2 种，每个卷积核对句子矩阵进行卷积操作并生成长度可变的特征图，然后对每个特征图进行最大池化（max-pooling）操作，即取出每个特征图中的最大值，由这些单值生成一个固定大小的特征向量，并且这个特征向量以全连接的形式传给 Softmax 层，Softmax 层使用该特征向量作为输入并且计算出各个类别上的概率分布（这里我们假设是二值分类，因此图 3.2 中只有 2 个分类状态）。

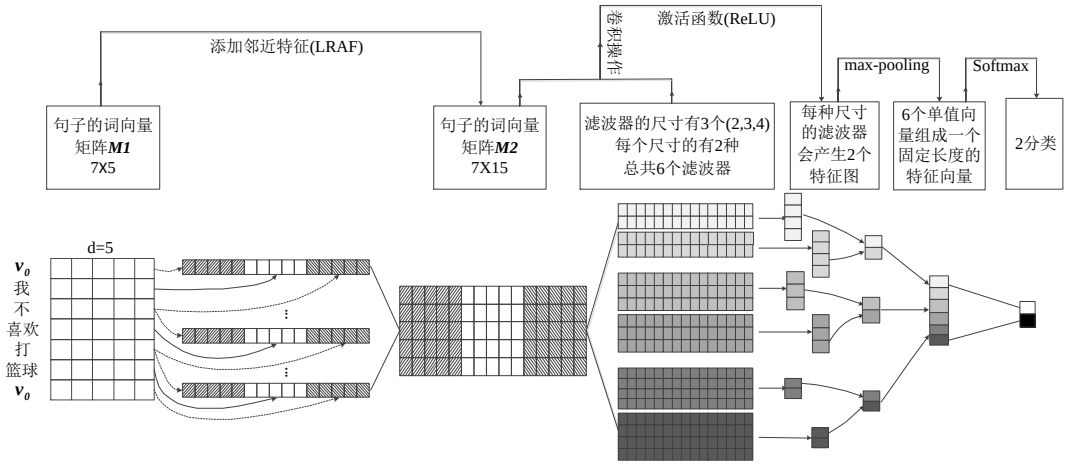

图 3.2 基于词语邻近特征的卷积神经网络模型的结构

3.3.2 模型的原理

假设输入层句子 S 的长度为 n，$v_i \in \mathbb{R}^d$ 表示 d 维的词向量，对应句子 S 中的第 i 个词语，那么句子 S 就可以表示成：

$$v_{1:n} = v_1 \oplus v_2 \oplus v_3 \oplus ... \oplus v_i \oplus ... \oplus v_n \tag{3.3}$$

其中 $\oplus$ 表示串联操作，换句话说，$v_{i:i+j}$ 表示词向量 $v_i, v_{i+1}, ..., v_{i+j}$ 的串联。如果使用邻近特征附加策略，则句子 S 会有另外 3 种表示形式（图 2 中使用的是 $v_{1:n}^{LRAF}$）：

$$v_{1:n}^{LAF} = (v_0 \uplus v_1) \oplus (v_1 \uplus v_2) \oplus (v_2 \uplus v_3) \oplus ... \oplus (v_{i-1} \uplus v_i) \oplus ... \oplus (v_{n-1} \uplus v_n) \tag{3.4}$$

$$v_{1:n}^{RAF} = (v_1 \uplus v_2) \oplus (v_2 \uplus v_3) \oplus (v_3 \uplus v_4) \oplus ... \oplus (v_i \uplus v_{i+1}) \oplus ... \oplus (v_n \uplus v_0) \tag{3.5}$$

$$v_{1:n}^{LRAF} = (v_0 \uplus v_1 \uplus v_2) \oplus (v_1 \uplus v_2 \uplus v_3) \oplus (v_2 \uplus v_3 \uplus v_4) \oplus ... \oplus (v_{i-1} \uplus v_i \uplus v_{i+1}) \oplus ... \oplus (v_{n-1} \uplus v_n \uplus v_0) \tag{3.6}$$

其中，运算符 $\uplus$ 为两个向量的合并符号，例如向量 $a=[1,2,3]$，向量 $b=[4,5,6]$，那么有 $a \uplus b = [1,2,3,4,5,6]$。卷积层利用大小为 $h \times k$ 的滤波器对输入特征矩阵进行卷积操作，即：

$$c_i = f(w \bullet v_{i:i+h-1}^{LRAF} + b) \tag{3.7}$$

其中，c_i 代表特征图中第 i 个特征值，$f(\bullet)$ 为卷积核函数，$w \in \mathbb{R}^{h \times k}$ 为滤波器，h 为滑动窗口大小，b 为偏置值。$v_{i:i+h-1}^{LRAF}$ 表示由第 i 行到第 $i+h-1$ 行组成的局部特征矩阵。因此，特征图 C 为：

$$C = [c_1, c_2, c_3, ..., c_{n+h-1}] \tag{3.8}$$

在卷积层之后，加入了 ReLU 层，将 ReLU 作为激活函数，可以加速随机梯度下降的收敛速度[30]。下采样层使用文献[56]提出的 max-pooling 方法进行特征采样，得到的特征值 $\hat{c}$ 为：

$$\hat{c} = \max\{C\} \tag{3.9}$$

卷积层和下采样层共同构成 LRAF-CNN 的特征提取层。LRAF-CNN 由多个不同尺寸的特征提取层（h有多个值）并列组成，每种尺寸的特征提取器各有 m 个，因此全连接层的特征向量 z 为：

$$z=[\hat{c}_{1,h_1},...,\hat{c}_{i,h_1},...,\hat{c}_{m,h_1},...,\hat{c}_{1,h_j},...,\hat{c}_{i,h_j},...,\hat{c}_{m,h_j},...] \tag{3.10}$$

其中，$\hat{c}_{i,h_j}$ 表示第 j 种类型的滤波器产生的第 i 个特征值。相比普通的 CNN 模型，本文提出的这种网络结构可以有效的提取出与正负面情感标签相关的词向量特征，并用于最终的情感分类。

下采样层输出的多个单值特征作为全连接层的输入，全连接层将这些单值特征进行拼接，形成特征向量，然后利用 Softmax 多分类函数输出各个分类的概率大小，从而确定分类结果，并根据训练数据的实际分类标签，采用反向传播算法对模型参数进行梯度更新，即：

$$P(y=l\,|\,V;\theta)=\frac{e^{z^T\theta_l}}{\sum_{k=1}^{K}e^{z^T\theta_k}} \tag{3.11}$$

其中，y表示句子的实际情感标签，θ表示模型的参数集合（即神经元之间的权重矩阵，训练过程中θ会不断变化，直到收敛）。通过该式，可以计算出给定句子的情感标签和模型参数集合θ的条件概率分布图，进而得出分类结果。

3.4 实验评估

3.4.1 数据集

本文使用 COAE2014、COAE2015 的任务 4 的微博数据集。中文观点倾向性分析评测（Chinese Opinion Analysis Evaluation, COAE）始办于 2008 年，是国内第一个情感分析方面的权威评测会议。COAE2014 任务 4（2 分类）的微博数据集共包含 40000 条数据，其中官方公布了 7000 条微博的极性（负面有 3224 条、正面有 3776 条），另 33000 条是混淆语料，语料涵盖手机、翡翠、保险三个领域。COAE2015 任务 4（3 分类）的微博数据集共包含 133201 条数据，其中官方公布了 20154 条微博的极性（负面有 6209 条、中性有 1639 条、正面有 12306 条），另 113047 条是混淆语料，语料涵盖汽车、电子、手机、美食、娱乐、宾馆等多个领域。两个数据集的概要信息如表 3.1 所示。

表 3.1 COAE2014、COAE2015 任务 4 的数据集概要

Data	c	l	N	V	V_{pre}	CV
COAE2014	2	100	7000	13993	13451	10
COAE2015	3	102	20154	19389	18384	10

其中，c 表示目标数据集是几分类，l 表示该数据集中最长的句子包含多少个词语，N 表示数据集中有多少条标注数据，V 表示标注数据的词汇量大小，V_{pre} 表示标注数据含有多少词汇已存在预训练词向量中，CV 等于 10 表示采用十折交叉验证法划分实验数据。

3.4.2 数据预处理和预训练词向量

在数据预处理方面，本章直接去除了文本中所有标点符号和表情符号，只留下含义较多的中、英文，然后利用 ICTCLAS 分词工具对实验数据集进行分词。在预训练词向量方面，本文使用新浪微博提供的数据抽取接口抽取了 1000 万条微博数据，经过去燥和预处理之后，再使用 Google 开源的词向量训练工具 word2vec 训练词向量。训练词向量时，word2vec 的主要参数设置如表 3.2 所示。

表 3.2 训练词向量时 word2vec 的主要参数设置

cbow	size	window	negative	hs	sample	iter
0	300	8	0	1	1e-4	10

其中，cbow 等于 0 表示使用 Skip-Gram 模型；size 指词向量的维度，本文使用 300 维的词向量；window 等于 8 表示训练考虑一个词的前 8 个和后 8 个词语；negative 等于 0、hs 等于 1 表示不使用 NEG 方法、使用 HS 方法；sample 指采样的阈值，如果一个词语在训练样本中出现的频率越大，那么就越会被采样；iter 指迭代次数。经过训练后，得到的词向量文件包含 80552 个词语，在 COAE2014、COAE2015 任务 4 的数据集上的词汇覆盖率分别是 96.13%、95.33%。

3.4.3 模型对比实验

CNN：这是本章的基准模型，使用预训练词向量对模型的输入数据进行初始化，不在预训练词向量中的词语随机初始化（随机初始化范围是[-0.25, 0.25]），并且在训练过程中所有词向量保持不变。

LAF-CNN：词向量的初始化方案与 CNN 相同；训练过程中为每个词语添加左邻近特征，如果左邻近词语不在预训练词向量中，则随机初始化左邻近特征，且在训练过程中所有词向量保持不变。

RAF-CNN：词向量的初始化方案与 CNN 相同；训练过程中，为每个词语添加右邻近特征，如果右邻近词语不在预训练词向量中，则随机初始化右邻近特征，且在训练过程中所有词向量保持不变。

LRAF-CNN：词向量的初始化方案与 CNN 相同；训练过程中，为每个词语添加左、右邻近特征，如果左、右邻近词语不在预训练词向量中，则随机初始化左、右邻近特征，且在训练过程中所有词向量保持不变。

CNN-multichannel：词向量的初始化方案与 CNN 相同；这是 Kim[29]提出的一种多通道 CNN 模型。模型在训练过程中，会同时使用了两组词向量，每组词向量被称为一个“通道”，两个通道同时参加卷积操作，但梯度只会通过其中一个通道反向传播而另一个通道保持不变。

上述四个模型除了邻近特征的附加策略不同之外，超参数设置均相同，超参数如表 3.3 所示。模型在训练过程中，采用 Zeiler[57]提出的 Adadelta Update Rule 规则进行随机梯度下降更新参数。需要说明的是，对比模型 CNN-multichannel 的超参数均与原文献中保持一致，这里不再一一列出。

表 3.3 AF-CNNs 模型的超参数

可调参数	值
激活函数	ReLU
词向量维度	300
批处理大小	64
滤波器尺寸	3,4,5
每个尺寸的滤波器个数	128
学习率	0.0001
参数更新比例	0.5
权重惩罚系数(L2)	1.0
迭代次数	200

3.4.4 实验结果与讨论

本文的基准模型是未考虑任何邻近特征的卷积神经网络。实验效果评价指标为：查准率、召回率、F 值、准确率。模型对比实验在 COAE2014 数据集上的实验结果如表 3.4 所示。由表 3.4 可知，相比基准模型 CNN，LAF-CNN、RAF-CNN、LRAF-CNN 在查准率、召回率、F1 值、准确率方面均有提升，其中准确率分别提高了 1.14%、0.57%、3%。另外，LRAF-CNN 模型相比于 Kim 的 CNN-multichannel 模型在各个指标均有一定幅度的提升，这说明考虑了邻近特征的卷积神经网络模型在各个方面的性能的确得到了提升。

表 3.4 模型对比实验在 COAE2014 上的实验结果(%)

模型	负面			正面			准确率
	查准率	召回率	F1 值	查准率	召回率	F1 值	
CNN	87.30	82.72	84.94	85.75	89.63	87.65	86.43
LAF-CNN	87.38	85.49	86.43	87.73	89.36	88.54	87.57
RAF-CNN	88.20	83.02	85.53	86.08	90.43	88.20	87.00
LRAF-CNN	90.97	85.98	88.40	88.21	92.47	90.29	**89.43**
CNN-multichannel	89.18	84.98	87.03	85.70	89.66	87.63	87.31

模型对比实验在 COAE2015 数据集上的实验结果如表 3.5 所示。由表 3.5 可知，相比基准模型 CNN，LAF-CNN、RAF-CNN、LRAF-CNN 在查准率、召回率、F1 值方面浮动比较大，这是因为各个分类的训练集数量不均衡所导致的，

但在准确率方面依然有一定提升，分别提高了 1.63%、1.09%、1.83%。与 2 分类实验结果类似，LRAF-CNN 模型相比于 Kim 的 CNN-multichannel 模型在各个指标上也都有一定幅度的提升。

表 3.5 模型对比实验在 COAE2015 上的实验结果(%)

模型	负面			中性			正面			准确率
	查准率	召回率	F1 值	查准率	召回率	F1 值	查准率	召回率	F1 值	
CNN	83.02	80.13	81.55	56.00	23.46	33.07	85.64	94.32	89.77	83.78
LAF-CNN	80.19	88.67	84.22	55.91	30.95	39.85	90.83	91.29	91.06	85.41
RAF-CNN	76.29	88.70	82.03	64.00	9.82	17.02	89.71	92.75	91.20	84.87
LRAF-CNN	84.52	83.72	84.12	50.00	21.23	29.81	87.80	94.23	90.90	**85.61**
CNN-multichannel	79.69	86.71	83.05	51.52	10.56	17.53	87.09	92.21	89.58	83.92

纵观以上两组实验的结果，我们可以看出以下几点：

1．附加邻近特征 vs. 未附加邻近特征。从两个数据集上的实验结果来看，当加入邻近特征后，模型在综合性能确实得到了一定提升，这说明在卷积过程中给每个词语添加邻近特征后，一方面可以强化词语在特定上下文中所表达的语义，另一方面相邻词语之间相互关联的关系提高了信息传递的连续性。这两点在一定程度上可确保在卷积过程中即使漏掉了个别情感词语，也不会对最终的分类结果产生多大影响。

2．左邻近特征 vs. 右邻近特征。从两个数据集上的实验结果来看，LAF-CNN 的分类准确率均略高于 RAF-CNN，这可能跟词语的序列性有关。众所周知，中文文本语义的表达是从左至右、先有上文后有下文，而 LAF-CNN 恰巧具有这种承上启下的作用。具体是否因为这个原因，还有待进一步验证。

3．左邻近特征、右邻近特征 vs. 左右邻近特征。从两个数据集上的实验结果来看，LAF-CNN、RAF-CNN 的分类准确率均略低于 LRAF-CNN，这跟词语在特定上下文中所蕴含语义有关，由于 LAF-CNN 只携带了左邻近特征、RAF-CNN 只携带了右邻近特征，而 LRAF-CNN 兼顾左右邻近特征，可以提取到更精确的语义信息，因此效果更好。

3.5 小结

本文为了在卷积过程中增强词语在当前上下文中的语义和保证词语在局部范围内的序列性，提出了一种基于词语邻近特征的卷积神经网络模型。首先介绍了什么是词语的邻近特征，并且提出了三种词语邻近特征的附加策略，分别是左邻近特征（LAF）、右邻近特征（RAF）、左右邻近特征（LRAF）。接着介绍了模型的结构，并且详细描述了模型背后的数学原理。最后，为了验证模型的各方面性能，分别在 COAE2014（2 分类）、COAE2015（3 分类）的任务 4 数据集上进行了实验。实验结果表明，本文的方法确实可行、有效，具有一定的研究价值。

第4章 基于递归神经网络和人工判定规则的情感分析模型

4.1 引言

传统的基于情感词典和判定规则的分析方法是从语言学的角度出发，但是这种方法需要制定大量的情感词典和判定规则。递归神经网络具有很强的概括、归纳能力，它可以通过不断的编码和重构训练，利用句子的层次结构特征来学习句子的特征表示，能够从大量未标注的语料中学习到先验知识。本文在基于情感词典和判定规则的方法与基于递归神经网络的方法的基础上，结合两者的优点，提出一种新颖的基于递归神经网络和人工判定规则的情感分析模型。首先利用多个并行的递归神经网络计算单元分别计算出组成文本的多个子句的情感极性，然后根据极性融合规则将多个子句的情感极性进行融合，最后利用人工判定规则计算出原始文本的情感极性。该模型的优点在于：一方面，递归神经网络中融入了专家判定规则，使得以往的专家经验得到有效利用；另一方面，递归神经网络代替了情感词典，避免了情感词典的局限性，二者相辅相成。该模型在斯坦福大学公开的数据集 SST-C2 和 SST-C3 上的分类准确率分别达到了 87.8%、81.6%，并且整体性能均优于主流的分类模型，说明该模型不仅新颖而且确实可行、有效。

4.2 情感词典和人工判定规则

情感词典是经过人工构建的带有情感色彩的词语的集合，将这样的词语及其相关属性值以一一对应的形式做成词典，以便于情感分析任务的利用。利用情感词典来判别文本情感极性是文本情感分析领域一种传统的且常见的方法。这是由于文本中带有情感极性的词语能比较直接的反映出作者所表达的情感倾向性，所以利用情感词作为重要切入点进行情感极性分析易于理解和实现。针对文本情感分析任务来说，运用情感词典判别法首先要构建出合适的情感词典并选取恰当的相关属性。目前，常用的情感词典有中国知网的 HowNet 中英文情感词典、大连理工大学情感词汇本体库、台湾大学 NTUSD 中文情感词典等等，包含的基本属性有词语名称、情感极性、情感强度、词性等。

基于情感词典的判别法是一种无监督方法，通过比对文本中所包含的褒义正向情感词、贬义负向情感词，然后按照一定的规则进行判别，仅需要用到情感词典的词语名称、词语情感极性两个基本属性。例如，对文本中包含的带有极性色

彩的情感词，可按照如下规则进行原始文本的情感极性判别：

$$\text{情感极性}=\begin{cases}\text{消极情感，褒义积极情感词个数}<\text{贬义消极情感词个数}\\ \text{中立情感，褒义积极情感词个数}>\text{贬义消极情感词个数}\\ \text{积极情感，褒义积极情感词个数}=\text{贬义消极情感词个数}\end{cases}$$

但是由于情感词典的覆盖范围有限，无法解决未登录词问题，模型的效果完全依赖算法的学习能力，往往效果不佳。另外，网络文本中含有较多的口语化词语、网络流行语等新词，这些新词往往也用于表达情感，却不包含在已有的情感词典中，因此仅采用基于情感词典和人工判定规则的无监督方法并非最佳方案。

由于递归神经网络特别善于概括上下文、挖掘文本的深度语义信息，因此针对网络文本的不规范性，RNN 成为当前主流文本情感分类的模型之一。由于普通的 RNN 模型在训练过程中是完全封闭的，而现存的专家经验又得不到有效利用，针对这个问题本文提出一种新颖的递归神经网络模型，即使用多个并行的递归神经网络代替情感词典计算出每个子句的情感极性值，然后根据极性融合规则融合所有子句的极性值，最后结合人工判定规则计算出原始文本的情感极性值。

4.3 基于人工判定规则的 LSTM 递归神经网络模型

4.3.1 模型的结构

1．基于 LSTM 和 RNN 的记忆单元结构

结合 LSTM 记忆单元和 RNN 模型即可形成一种新的记忆单元结构，称为 LSTM-RNN，如图 4.1 所示。该结构使得递归神经网络学习到的特征既包含文本的结构化信息又包含文本的历史信息。与 LSTM 记忆单元类似，LSTM-RNN 记忆单元也同样由输入门、遗忘门、输出门和记忆单元组成。

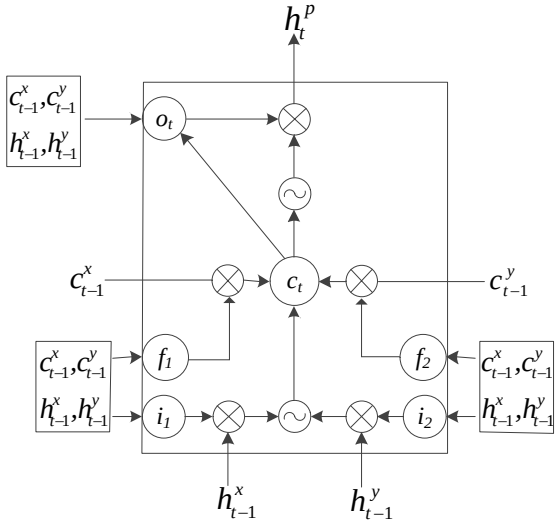

图 4.1 基于 LSTM 和 RNN 的记忆单元结构

根据 2.5.4 小结的介绍，我们知道 RNN 网络模型其实就是是一种图状拓扑结构，所以 LSTM-RNN 网络结构也是一种拓扑结构，那么在 t 时刻 LSTM-RSNN 的输入不止一个（例如对于二叉树而言，t 时刻的输入有两个：左孩子节点和右孩子节点）。因此 LSTM-RNN 的一个 LSTM 记忆单元需要两个输入门和两个遗忘门来控制对应的输入节点的信息和输出节点的信息。如图 4.1 所示，一个基于二叉树拓扑结构的 LSTM-RNN 记忆单元包含两个输入门（i_1、i_2），一个输出门（o_t），一个记忆单元（c_t）和两个遗忘门（f_1、f_2）组成。

图 4.1 中，p 表示父节点，x,y 表示两个孩子节点，h_{t-1}^x, h_{t-1}^y 分别表示孩子节点 x,y 在隐藏层的向量，而 c_{t-1}^x, c_{t-1}^y 分别表示孩子节点 x,y 的记忆向量（cell vector），那么 LSTM-RNN 的一次运算过程可以描述为：

$$i_1 = \delta\left(W_{i_1}^x h_{t-1}^x + W_{i_1}^y h_{t-1}^y + U_{i_1}^x c_{t-1}^x + U_{i_1}^y c_{t-1}^y + b_{i_1}\right) \tag{4.1}$$

$$i_2 = \delta\left(W_{i_2}^x h_{t-1}^x + W_{i_2}^y h_{t-1}^y + U_{i_2}^x c_{t-1}^x + U_{i_2}^y c_{t-1}^y + b_{i_2}\right) \tag{4.2}$$

$$f_1 = \delta\left(W_{f_1}^x h_{t-1}^x + W_{f_1}^y h_{t-1}^y + U_{f_1}^x c_{t-1}^x + U_{f_1}^y c_{t-1}^y + b_{f_1}\right) \tag{4.3}$$

$$f_2 = \delta\left(W_{f_2}^x h_{t-1}^x + W_{f_2}^y h_{t-1}^y + U_{f_2}^x c_{t-1}^x + U_{f_2}^y c_{t-1}^y + b_{f_2}\right) \tag{4.4}$$

$$c_t = f_1 \odot c_{t-1}^x + f_2 \odot c_{t-1}^y + tanh\left(W_{c_t}^x h_{t-1}^x \odot i_1 + W_{c_t}^y h_{t-1}^y \odot i_2 + b_{c_t}\right) \tag{4.5}$$

$$o_t = \delta\left(W_{o_t}^x h_{t-1}^x + W_{o_t}^y h_{t-1}^y + U_{o_t}^x c_{t-1}^x + U_{o_t}^y c_{t-1}^y + b_{o_t}\right) \tag{4.6}$$

$$h_t^p = o_t \odot tanh\left(c_t\right) \tag{4.7}$$

其中 W_*^*, U_*^* 为系数矩阵，b_* 为偏置项，i_1, i_2 分别表示两个孩子节点 x,y 对应的输入门的计算方法，f_1, f_2 分别表示两个孩子节点 x,y 对应的遗忘门的计算方法，o_t 表示父节点对应的输出门的计算方法，c_t 则表示父节点对应的记忆单元的计算方法，h_t^p 表示该节点的输出。

2．基于 LSTM-RNN 的句子级情感分析模型

本节介绍如何将提出的 LSTM-RNN 模型应用于情感分析，首先根据句子的句法结构构建网络，然后进行前馈传播学习句子的向量表示，根据学习到的表示对其进行分类判断其情感极性。具体网络结构如图 4.2 所示。其中，lx_1、lx_2 是节点 x_1、x_2 通过 LSTM-RNN 模型的输出层计算得到的节点对应的情感极性概率分布向量，同时在计算上层节点的向量表示时 lx_1、lx_2 将作为 LSTM 的输入决定偏置向量的选取，那么顶层父节点 p_2 的情感极性概率分布向量中标签为 l_k 的

概率计算方法如下：

$$P\left(l_k \mid x,\theta\right)=\frac{e^{W_k^{lp_2}x+b_k^{lp_2}}}{\sum_{i=1}^{|l|}e^{W_i^{lp_2}x+b_i^{lp_2}}} \tag{4.8}$$

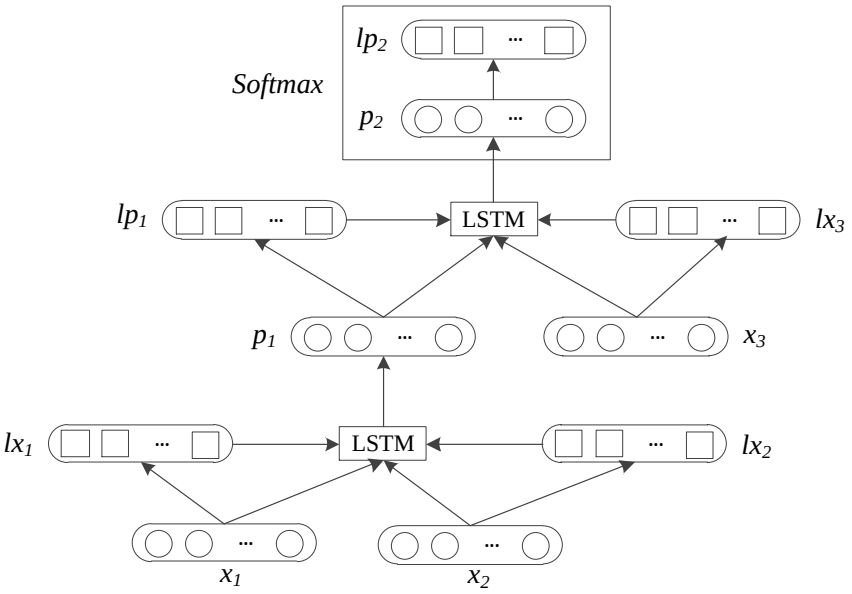

图 4.2 LSTM-RNN 模型的网络结构

3．基于 LSTM-RNN 和人工判规则的子句级情感分析模型

本文所提出的基于 LSTM-RNN 和人工判规则的情感分析模型的网络结构如图 4.3 所示（本文称之为 Rule-LSTM-RNN）。我们假设原始句子 S 的实际长度为 N，词向量维度为 d，那么原始句子就可用 N×d 的词向量矩阵表示。我们事先规定，当原始句子进入模型后，按以下标点符号对其进行子句分割：逗号、分号、句号、问号、感叹号。

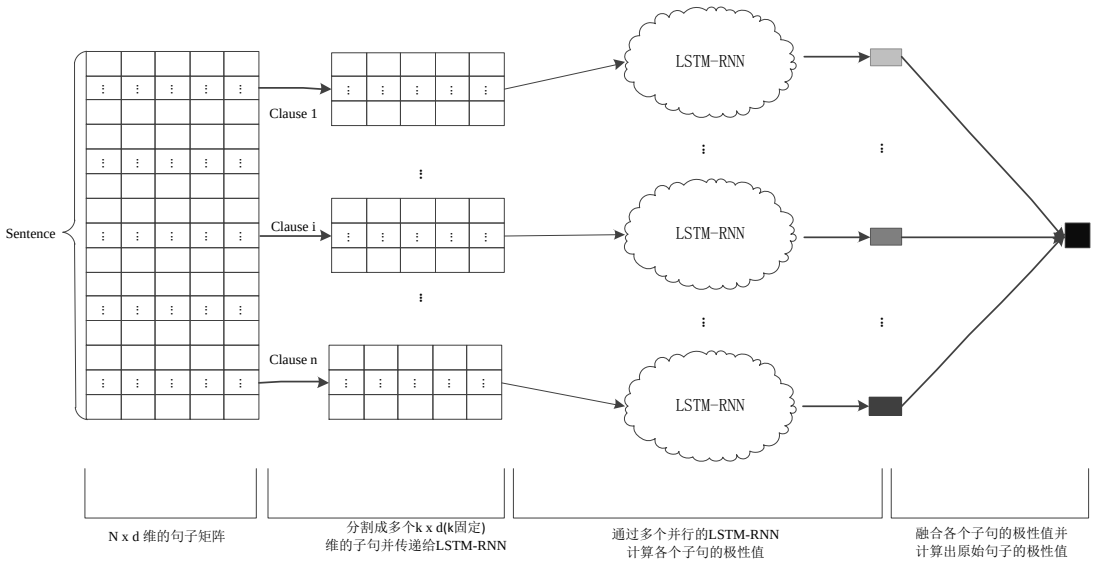

图 4.3 Rule-LSTM-RNN 的网络模型结构

图 4.3 中多条子句的情感计算任务在模型中是并行的。每个子句经过 LSTM-RNN 单元处理后均输出一个情感极性标签 l_k，最后通过预设的人工判定规则将这些子句的情感极性值进行融合，并且根据极性划分规则，从而计算出原始句子的情感分类结果。

4.3.2 模型的原理

原始句子 S 经过分割后，可表示成 $S = Clause_1 \oplus \cdots \oplus Clause_j \oplus Clause_k$，其中⊕表示串联操作，$Clause_j$表示句子 S 的第 j 条子句。我们假设子句$Clause_j$的长度为 n（必要时会进行补零向量操作，以保证每条子句长度相同），$v_i \in \mathbb{R}^d$表示 d 维的词向量，对应子句$Clause_j$中第 i 个词语，那么子句$Clause_j$就可以表示成：

$$Clause_j = v_{1:n} = v_1 \oplus v_2 \oplus v_3 \oplus \cdots \oplus v_i \oplus \cdots \oplus v_n \tag{4.9}$$

其中，$v_{1:n}$就表示词向量$v_1, v_2, v_3, \cdots, v_n$的串联，换言之，$v_{i:i+x}$表示词向量$v_i, v_{i+1}, \cdots, v_{i+x}$的串联。每条子句经过一个 LSTM-RNN 单元计算后，都会输出一个极性值P_{Clause_j}，表示分类标签概率最大的值。

1．针对 2 分类的融合规则

原始句子 S 按标点符号经过子句分割后，在模型中并行计算出多个子句的极性值$P_{Clause_j} \in \{-1, +1\}$，其中-1 代表消极，+1 代表积极。那么原始句子 S 的极性值P_S本文规定按下式计算：

$$P_S = \frac{\sum_{j=1}^{k} P_{Clause_j}}{k} \tag{4.10}$$

其中，k 表示子句的数量。经过式（4.10）计算后，P_S取值范围是[-1,+1]，P_S的极性划分按式（4.11）计算：

$$S_{polarity} = \begin{cases} neg, & -1 \le P_S < 0 \\ neg \text{ or } pos, & P_S = 0 \\ pos, & 0 < P_S \le +1 \end{cases} \tag{4.11}$$

其中，当P_S=0时，本文规定从{消极，积极}中随机选择一个极性。虽然这种做法比较粗鲁，势必会对模型的性能产生一定的负面影响，但从最后的实验效果来看，这种做法尚可接受。

2．针对 3 分类的融合规则

原始句子 S 按标点符号经过子句分割后，在模型中并行计算出多个子句的极性值$P_{Clause_j} \in \{-1, 0, +1\}$，其中-1 代表消极，0 代表中性，+1 代表积极。原始句子 S 的极性值P_S依然按照公式（4.10）计算。但是与 2 分类不同的是，3 分类P_S

的极性划分按式（4.12）计算：

$$S_{polarity}=\begin{cases} neg, & -1\le P_S<0 \\ neu, & P_S=0 \\ pos, & 0<P_S\le+1 \end{cases} \tag{4.12}$$

需要说明的是，在包含“中性”标签的多分类任务中，不存在 P_S 等于某个范围内的值而无法判定其分类的情况，因此对模型的负面影响也更小、降低了误判的概率，提高了模型的整体性能。

4.4 实验评估

4.4.1 数据集

为了验证模型的有效性，这里选用斯坦福大学在影评数据上利用众包标注开发的SST（Stanford Sentiment Treebank）数据集，该数据集包含原始句子中短语的极性，并且包含每个句子的句法分析树。原始SST数据集的标签共有5类（very negative, negative, neutral, positive, very positive），包含句子11855条，共有215154个短语，据统计平均每个句子包含19个词语。在进行实验时，针对原始SST数据集本文采用两种分类方案：SST-C2和SST-C3。SST-C2是将SST划分成2分类数据集，将数据集中情感极性为neutral的句子剔除，其中情感极性为neutral的句子约占整个数据集的20%左右，剔除掉之后的训练集包含句子6920条，开发集包含句子872条，测试集包含句子1821条。SST-C3是将SST划分成3分类数据集，将数据集中情感极性为very negative的数据归并为negative、将数据集中情感极性为very positive的数据归并为positive。经过整理后，训练集包含句子8544条，开发集包含句子1101条，测试集包含句子2210条。需要说明的是，训练集、开发集、测试集的分割方案与原始SST数据集的分割方案保持一致。

表 4.1 SST-C2、SST-C3 数据集概要

Dataset	C	Max-s-l	Max-c-l	N	V	V_{pre}	Test
SST-C2	2	53	49	9613	16185	14838	1821
SST-C3	3	53	49	11855	17836	16262	2210

其中，C表示数据集的分类标签数；Max-s-l表示数据集中句子的最大长度（单词数目）；Max-c-l表示数据集中子句的最大长度；N表示数据集的大小；V表示数据集的词汇量大小；V_{pre}表示数据集的词汇量已有多少存在预训的练词向量中；Test表示测试集大小。

另外，如前所述，SST数据集还提供了关于数据集中所有句子的短语数据集，并且短语的标签也分为5类。因此，本章对短语数据集也做了重新归类工作，并且后文网络模型的训练是同时基于短语数据集和句子数据集的，所以实际的训练语料要比表4.1中列出两个数据集的要大。

4.4.2 数据预处理和预训练词向量

在数据预处理方面，由于本章所用的方法需要先计算出子句的极性，因此训练过程中，需要根据标点符号把句子切分成多个子句，所以不能去掉原始句子中的标点符号。但是为了减小模型处理数据的复杂度，这里直接去掉了数据集中所有的字符类表情符号。此外需要说明的是，由于英文单词之间本身就是以空格隔开，因此英文数据集没有分词步骤。

在预训练词向量方面，本章直接使用 Google 开源的词向量“GoogleNews-vectors-negative300.bin”来初始化训练集。训练该词向量时，官方公布的 word2vec 的主要参数设置如表 4.2 所示。

表 4.2 开源词向量训练时 word2vec 的主要参数设置

cbow	size	window	negative	hs	sample	iter
1	300	8	25	0	1e-4	15

该词向量是基于 Google 新闻的 1000 亿个单词、连续词袋模型（CBOW）训练而成的 300 维词向量，几乎涵盖了所有领域的词汇。该词向量在 SST 数据集上的词汇覆盖率达到了 92.14%，满足了实验要求。需要说明的是，不在预训练词向量中的词汇采取随机初始化，随机初始化范围是[-0.25,+0.25]。

4.4.3 模型对比实验

LSTM-RNN：这是本章所使用的基准模型，这个模型就是普通的 LSTM 递归神经网络。LSTM-RNN 的处理单位是整个句子，不需要进行子句划分。数据初始化使用 4.4.2 节中的开源词向量，不在预训练词向量中的词汇采取随机初始化（范围是[-0.25,+0.25]）。

Rule-LSTM-RNN：这是本章提出的基于子句极性和人工判定规则的情感分类模型，本文称之为规则-LSTM 递归神经网络（Rule LSTM Recursive Neural Networks, Rule-LSTM-RNN）。如前述，Rule-LSTM-RNN 首先要将一个完整句子切分成多个子句，然后并行处理各个子句，最后通过极性融合规则计算原始句子的极性。数据初始化使用 4.4.2 节中的开源词向量，不在预训练词向量中的词汇采取随机初始化（范围是[-0.25,+0.25]）。

RNTN：这是文献[58]提出的一种名叫递归神经张量网络（Recursive Neural Tensor Networks, RNTN），该模型可以将任意长度的短语作为输入。它通过使

用词向量和解析树表示一个短语，然后使用基于张量的组合函数计算出解析树中高层节点的向量表示。该模型能够衡量出一个词的上下文语义，强大到足以凭单个词语翻转整个句子的情感极性。数据初始化使用 4.4.2 节中的开源词向量，不在预训练词向量中的词汇采取随机初始化（范围是[-0.25,+0.25]）。

MV-RNN：这是文献[59]提出的一种名叫矩阵-向量递归神经网络（Matrix-Vector Recursive Neural Networks, MV-RNN），该模型可以学习到任意句法类型和长度的短语和句子的组合向量表示。模型在训练过程中为分析树中的每个节点分配一个向量和一个矩阵：向量用于捕获组成部分的固有含义，而矩阵捕获它如何改变相邻单词或短语的含义。数据初始化使用 4.4.2 节中的开源词向量，不在预训练词向量中的词汇采取随机初始化（范围是[-0.25,+0.25]）。

DCNN：这是文献[60]提出的一种名叫动态卷积神经网络（Dynamic Convolutional Neural Networks），DCNN 能够处理可变长度的输入，网络中包含两种类型的层，分别是一维的卷积层和动态 k-max 的池化层（Dynamic k-max pooling）。该模型保留了句子中词序信息和词语之间的相对位置，同时不需要任何的先验知识，例如句法依存树等，并且模型考虑了句子中相隔较远的词语之间的语义信息。数据初始化使用 4.4.2 节中的开源词向量，不在预训练词向量中的词汇采取随机初始化（范围是[-0.25,+0.25]）。

上述 LSTM-RNN 和 Rule-LSTM-RNN 除了模型结构不同，超参数设置均相同，如表 4.3 所示。模型在训练过程中，采用 Zeiler[57]提出的 Adadelta Update Rule 规则进行随机梯度下降更新参数。需要说明的是，对比模型 MV-RNN、RNTN、DCNN 的超参数均与原文献中保持一致，这里不再一一列出。

表 4.3 LSTM-RNN、Rule-LSTM-RNN 的超参数

可调参数	值
激活函数	Sigmoid
词向量维度	300
批处理大小	64
学习率	0.0001
梯度最大值	5
权重惩罚系数(L2)	1.0
参数更新比例	0.5
迭代次数	500

4.4.4 实验结果与讨论

模型的评价指标为：查准率、召回率、F1 值、准确率。模型对比实验 SST-C2 数据集（2 分类）上的实验结果如表 4.4 所示。由表 4.4 可知，虽然 Rule-LSTM-RNN 在查准率、召回率、F1 值、准确率方面与基准模型 LSTM-RNN 相比存在一定差距，但其性能却远超其他对比模型，这说明基于子句极性和人工判

定规则的递归神经网络模型的确有效，具有一定的研究价值。

表 4.4 模型对比实验在 SST-C2 上的实验结果(%)

模型	消极			积极			准确率
	查准率	召回率	F1 值	查准率	召回率	F1 值	
LSTM-RNN	90.1	86.6	88.3	87.2	90.4	88.8	**88.5**
Rule-LSTM-RNN	88.7	86.7	87.7	87.1	88.9	88.0	87.8
MV-RNN	84.3	80.9	82.6	81.7	84.8	83.2	82.9
RNTN	86.6	83.8	85.2	84.3	87.0	85.7	85.4
DCNN	87.7	85.1	86.4	85.6	88.1	86.8	86.8

模型对比实验在 SST-C3 数据集（3 分类）上的实验结果如表 4.5 所示。由表 4.5 可知，相比基准模型 LSTM-RNN，Rule-LSTM-RNN 在查准率、召回率、F1 值方面几乎都比 LSTM-RNN 的表现要好，同时优于其他对比模型的性能。但是，各个模型在中性测试集上的评价指标都远远低于其他两个分类，这是主要因为数据集的不均衡所导致的。

表 4.5 模型对比实验在 SST-C3 上的实验结果(%)

模型	消极			中性			积极			准确率
	查准率	召回率	F1 值	查准率	召回率	F1 值	查准率	召回率	F1 值	
LSTM-RNN	84.6	85.1	84.9	62.6	57.6	60.0	85.7	88.1	86.9	81.5
Rule-LSTM-RNN	84.7	84.8	84.7	63.0	58.6	60.7	85.8	88.2	87.0	**81.6**
MV-RNN	82.9	82.6	82.8	57.1	50.9	53.8	82.9	87.1	85.0	78.9
RNTN	83.4	83.4	83.4	58.6	50.9	54.5	83.2	87.6	85.3	79.5
DCNN	84.3	86.3	85.3	61.9	52.2	56.6	84.0	87.6	85.7	80.8

纵观以上两组实验的结果，我们可以总结出以下几点：

1．整句分析 vs. 并行化子句分析。从表 4.4、表 4.5 中可以看出，Rule-LSTM-RNN 在两个数据集上都取得了不错的成绩并且在 3 分类测试中取得最佳准确率，这说明相比于直接将整个句子作为处理单元，也许以子句作为处理单元的效果更好。究其原因，可能是因为 Rule-LSTM-RNN 可以并行计算出每个子句的极性，然后进行极性融合，这种处理方式更加细致的考虑了组成句子的每个子句甚至短语的极性，可以捕获到文本更深层次的情感信息，因此效果更好。

2．多分类时子句的极性融合规则。从表 4.4 中可以看出，虽然 Rule-LSTM-RNN 的性能胜过其他对比模型，但是与 LSTM-RNN 相比还是有一定的差距。这主要是由于二分类的极性融合规则不合理引起的。我们在 4.3.2 节中针对二分类的极性融合，当$P_S = 0$时，由于没有极性为“中性”的分类，因此实验中随机从（“消极”，“积极”）中为其指定了一个分类。这种做法是粗鲁且不合理的，因此在测试模型时导致相当一部分数据有 0.5 的概率被误判。但是从表 4.5 中可以看出，Rule-LSTM-RNN 在各方面的性能都是最好，这说明在包含“中性”分类的测试任务中，不存在当$P_S = 0$时被随机判别的情况，判定规则是合理、有效的。

3．数据不平衡对整体性能的影响。从表 4.4 中可以看出，由于两个分类的

数据量差不多，各个模型在查准率、召回率、F1 值方面的表现均相差无几。但是在从表 4.5 中来看，各个模型在“中性”分类上的查准率、召回率、F1 值的表现均不如人意，这是因为“中性”分类数据的不均衡所导致的，并且由于“中性”分类相比其他两个分类的数据量太小，导致模型无法学习到有效的先验知识，这势必会对模型整体的性能产生一定的负面影响。

4.5 小结

本章首先从传统的基于情感词典和人工判定规则的无监督方分类法出发，阐述了由于情感词典的缺陷导致无监督方法的泛化能力相当受限。在此基础上，本文提出一种新颖的基于 LSTM-RNN 和人工判定规则的情感分类方法，这种方法旨在取代情感词典，将递归神经网络的深层特征学习能力与人工判定规则结合起来，共同完成情感分析的任务。最后，为了验证模型的各方面性能，分别在 SST-C2 和 SST-C3 数据集上进行了模型对比实验。实验结果表明，基于 LSTM-RNN 和人工判定规则的方法确实可行、有效，具有一定的研究价值。

结 论

随着论坛、博客以及电商等网络媒介的迅速发展，社会媒体呈现出多样化的形态，网络用户的参与性不断提高且对于网络的使用方式也发生了巨大改变。用户不再是只能被动的接受网络信息，而是更加积极主动的成为了网络信息的生产者。这样的变化使得各种网络媒介中出现大量用来表达用户情感、态度和观点的主观性信息，而文本则是最为主要的表现形式。针对这些主观性信息，如何更加有效的加以利用，提取出人们感兴趣且携带情感倾向性的文本，并且对其做出准确的分析和判断以作为商业决策时的参考，是迄今为止自然语言处理领域中的非常重要的研究课题之一，具有较高的科学研究意义和实际应用价值。

本文首先介绍了情感分析的背景和意义，接着阐述了国内外研究现状，并对现存情感分析技术的不足和难点进行了详细的说明，最后总结情感分析技术的相关技术基础。本文主要研究基于卷积神经网络的情感分析方法，旨在提高基于情感分析的舆情监控、用户评价分析等应用系统的性能。本文工作对情感分析领域的贡献主要有以下两点：

1、基于卷积神经网络和词语邻近特征的情感分析模型。当前几乎所有的基于卷积神经网络的情感分析模型在卷积过程中每个词向量只能孤立的表征单个词语，并不蕴含上下文信息，这不利于信息传递的连续性，并且卷积操作在局部范围内可能会打乱词向量的序列性。针对这个问题，本文提出一种基于词语邻近特征的卷积神经网络模型，即在卷积过程中让每个词向量携带邻近词语的特征，这样既保证信息传递的连续性也保证了词向量在局部范围内的序列性。实验结果表明，在 COAE2014（2 分类）、COAE2015（3 分类）的情感分类任务上的准确率分别达到了 89.43%和 85.61%，并且优于未考虑邻近特征的卷积神经网络模型，说明该模型确实可行、有效，具有一定的研究价值。

2、基于递归神经网络和人工判定规则的情感分析模型。传统的基于情感词典和人工判定规则的分析方法是从语言学的角度出发，但是这种方法需要制定大量的情感词典和判定规则。而基于递归神经网络的分析方法可以通过不断的编码和重构训练，从大量未标注的语料中学习到先验知识。本文在参考了这两种方法之后，在第 4 章中提出了一种新颖的基于递归神经网络和人工判定规则的情感分析模型。首先利用递归神经网络并行计算出组成文本的多个子句的情感极性，然后根据极性融合规则将多个子句的情感极性进行融合，最后利用人工判定规则计算出原始文本的情感极性。该模型的优点在于：一方面递归神经网络中融入了人工判定规则，使得以往积累的人工经验得到有效利用；另一方面递归神经网络取

代了情感词典，避免了情感词典的局限性。该模型在数据集 SST-C2 和 SST-C3 上的分类准确率分别达到了 87.8%、81.6%，并且整体性能均优于主流的分类模型，说明该模型不仅新颖，而且确实可行、有效，具有一定的研究价值。

本文基于卷积神经网络分别提出了两个具有不同研究价值情感分析模型，针对性的解决了情感信息传递不连续问题、局部范围内词序被打乱的问题以及基于情感词典和人工判定规则的无监督学习算法泛化能力弱的问题，但是在实验过程中，仍发现许多不足。针对这些不足之处，未来的完善工作可从下几点着手：

1、扩大邻近特征的挖掘范围。在第 3 章中，模型在训练过程中附加邻近特征时，只考虑了当前词语的前、后邻近词的特征，如果在计算能力允许的情况下，是否可以通过扩大邻近词的数量从而获取更精确的上下文信息，这是否有利于进一步提高模型的整体性能，值得探究。

2、由标点符号和表情符号所引起的极性转移问题。在第 3 章和第 4 章的数据预处理阶段，本文直接去掉了所有标点符号和字符型表情符号。然而，例如问号（“？”）、感叹号（“！”）、开心（“^_^”）、沮丧（“>_<”）等标点符号和表情符号必然会影响上下文的情感极性转移，这也无疑对模型最后的分类效果产生一定的负面影响，因此考虑由标点符号和表情符号而引起的情感极性转移问题是本文今后的另一个研究方向。

3、分类数据集不平衡。从第 3 章和第 4 章中的实验结果可以看出，两个模型在 3 分类数据集的“中性”分类上的性能表现均不如其他两个分类，这主要是由于“中性”分类的训练集太小，导致模型无法学习到有效的先验知识，因此在测试阶段必然会产生大量的误判。解决分类数据不平衡问题的方法可借鉴“重采样技术”或“成本敏感学习（Cost Sensitive Learning, CSL）”，这也将是本文今后的一个研究方向。

4、完善极性融合规则。在第 4 章中，我们针对 2 分类和 3 分类数据集分别设计了简单、粗糙且类似的极性融合规则，实验结果差强人意，并没有很大提高。为了进一步提高模型的性能，我们是否可以考虑引入模糊评价法的模糊规则，使得子句的极性融合过程更加合理，值得探究。

参考文献

[1] Wan X J. Co-Training for Cross-Lingual Sentiment Classification[C]. Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP. Suntec, Singapore, 2009: 235-243.

[2] WT Yih，X He，C Meek. Semantic parsing for single-relation question answering[C]. Proceedings of the ACL. Baltimore, USA: Citeseer Press, 2014: 643-648.

[3] Y Shen，X He，J Gao, et al. Learning Semantic Representations Using Convolutional Neural Networks for Web Search[C]. Proceedings of the WWW. New York, USA: ACM Press, 2014: 373-374.

[4] Picard R W. Affective Computing. Cambridge: MIT Press, 1997, 6-9.

[5] Pang B, Lee L, Shivakumar V.Thumbs up? Sentiment Classification Using Machine Learning Techniques. In: Proceedings of Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA: ACM Press, 2002, 79-86.

[6] Turney. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proceedings of the 40th Annual of the Association for Computational Linguistics. Philadelphia, USA: ACM Press, 2002, 417-424.

[7] Hinton G.E., Salakhutdinov R. R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, Vol. 313. No. 5786, 2006, 28(7): 504-507.

[8] Alistair K, Diana I. Sentiment Classification of Movie Reviews Using Contextual Valence Shifters. Computational Intelligence, 2006, 22(2): 110-125.

[9] Maite T, Julian B, Milan T, et al. Lexicon-Based Methods for Sentiment Analysis. Computational Linguistics, 2011, 37(2): 267-307.

[10] Jinan F, Osama M, Sabah M, et al. Opinion Mining over Twitterspace: Classifying Tweets Programmatically using the R Approach. In: Proceedings of the 7th International Conference on Digital Information Management. Macau, China: IEEE Press, 2012, 313-319.

[11] 冯时, 付永陈, 阳锋等. 基于依存句法的博文情感倾向分析研究. 计算机研究

与发展, 2012, 49(11): 2395-2406.

[12] Dmitry D, Oren T, Ari R. Enhanced Sentiment Learning Using Twitter Hashtags and Smileys. In: Proceedings of the 23rd International Conference on Computational Linguistics. Beijing, China: DBLP Press, 2010, 241-249.

[13] Hiroshi M, Kazutaka S, Tsutomu E. Twitter Sentiment Analysis Based on Writing Style. Advances in Natural Language Processing, 2012, 7614: 278-288.

[14] Fotis A, George P, Konstantinos T, et al. Content vs. Context for Sentiment Analysis: a Comparative Analysis over Microblogs. In: Proceedings of the 23rd ACM Conference on Hypertext and Social Media. Milwaukee, USA: ACM Press, 2012, 187-196.

[15] Efstratios K, Christos B, Theologos D. Ontology-based sentiment analysis of twitter posts. Expert System with Applications, 2013, 40(10): 4065-4074.

[16] 谢丽星，周明，孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取. 中文信息学报, 2012, 26(1): 73-83.

[17] Wang J, Song D. Liao L. The Chinese Bag-of-Opinions Method for Hot-Topic-Oriented Sentiment Analysis on Weibo. In: Proceedings of the 6th Chinese Semantic Web Symposium and 1st Chinese Web Science Conference. Shenzhen, China: DBLP Press, 2012, 357-367.

[18] 曹海涛. 基于 PAD 模型的中文微博情感分析研究: [大连理工大学硕士学位论文]. 大连: 大连理工大学计算机应用系, 2013, 1-15.

[19] Tan S, Wang Y, Cheng X. Combining Learn-based and Lexicon-based Techniques for Sentiment Detection without Using Labeled Examples. In: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development Information Retrieval. Singapore: ACM Press, 2008, 743-744.

[20] Zhang L, Ghosh R, Dekhil M, et al. Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. HP Laboratories, Tech Rep: HPL-2011-89, 2011.

[21] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks[C]. Proceedings of the NIPS. Lake Tahoe, USA: NIPS Press, 2012: 1097-1105.

[22] Graves, Mohamed A, Hinton G. Speech recognition with deep recurrent

neural networks[C]. Proceedings of the ICASSP. Vancouver, Canada: IEEE Press, 2013: 6645-6649.

[23] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]. Proceedings of the NIPS. Lake Tahoe, USA: NIPS Press, 2013: 3111-3119.

[24] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]. Proceedings of the EMNLP. Seattle, USA: Citeseer Press, 2013: 1631-1642.

[25] Dong L, Wei F, Tan C, et al. Adaptive recursive neural network for target-dependent Twitter sentiment classification[C]. Proceedings of the ACL. Baltimore, USA: ACL Press, 2014: 49-54.

[26] Dong L, Wei F, Zhou M, et al. Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis[C]. Proceedings of the AAAI. Quebec, Canada: AAAI Press, 2014: 1537-1543.

[27] N Kalchbrenner，E Grefenstette，P Blunsom. A Convolutional Neural Network for Modelling Sentences[C]. Proceedings of the ACL. Baltimore, USA: arXiv Press, 2014: 1404-2188.

[28] Dos Santos C N, Gatti M. Deep convolutional neural networks for sentiment analysis of short texts[C]. Proceedings of the COLING. Dublin, Ireland: COLING Press, 2014: 69-78.

[29] Kim Y. Convolutional neural networks for sentence classification[C]. Proceedings of the EMNLP. Doha, Qatar: arXiv Press, 2014: 1408-5882.

[30] 何炎祥, 孙松涛, 牛菲菲等. 用于微博情感分析的一种情感语义增强的深度学习模型[J]. 计算机学报, 2016, 39: 1-19.

[31] Peng Chen, Bing Xu, Muyun Yang, et al. Clause sentiment identification based on convolutional neural network with context embedding[C]. Proceedings of the ICNC-FSKD. Changsha, China: IEEE Press, 2016: 1532-1538.

[32] S.Fouzia S, Adnan R.H, Mohd A.H. Supervised Opinion Mining of Social Network Data using a Bag-of-Words Approach on the Cloud. In: Proceedings of 7th International Conference on Bio-Inspired Computing: Theories and Applications. Gwalior, India: IEEE Press, 2012, 299-309

[33] Fotis A, George P, Konstantinos T, et al. Content vs. Context for Sentiment Analysis: a Comparative Analysis over Microblogs. In:

Proceedings of the 23rd ACM Conference on Hypertext and Social Media. Milwaukee, USA: ACM Press, 2012, 187-196

[34] 奉国和,郑伟.国内中文自动分词技术研究综述[J]. 图书情报工作, 2011, 54(02): 41-45.

[35] Younger D H. Recognition and parsing of context-free languages in time n 3[J]. Information and control, 1967, 10(2): 189-208.

[36] Yang Y, Pedersen J O. A comparative study on feature selection in text categorization[C]. ICML. 1997, 97: 412-420.

[37] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space[J]. ar Xiv preprint ar Xiv:1301.3781, 2013.

[38] Huang E H, Socher R, Manning C D, et al. Improving word representations via global context and multiple word prototypes[C]. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Association for Computational Linguistics, 2012: 873-882.

[39] Aho A V, Ullman J D. The theory of parsing, translation, and compiling[M]. Prentice-Hall, Inc., 1972.

[40] Socher R, Lin C C, Manning C, et al. Parsing natural scenes and natural language with recursive neural networks[C]. Proceedings of the 28th international conference on machine learning (ICML-11). 2011: 129-136.

[41] MCCULLOCH W S, PITTS W. A logical calculus of the ideas immanent in nervous activity[J]. The bulletin of mathematical biophysics, 1943, 5(4): 115-133.

[42] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Cognitive modeling, 1988, 5(3): 1.

[43] CHEN S, COWAN C F, GRANT P M. Orthogonal least squares learning algorithm for radial basis function networks[J]. Neural Networks, IEEE Transactions on, 1991, 2(2): 302-309.

[44] RUCK D W, ROGERS S K, KABRISKY M. The multilayer perceptron as an approximation to a Bayes optimal discriminant function[J]. Neural Networks, IEEE Transactions on, 1990, 1(4): 296-298.

[45] RITTER H, MARTINETZ T, SCHULTEN K. Neural computation and self-organizing maps: an introduction[M]. Addison-Wesley Reading, MA,

1992.

[46] Goller C, Kuchler A. Learning task-dependent distributed representations by backpropagation through structure[C]. Neural Networks, 1996., IEEE International Conference on. IEEE, 1996, 1: 347-352.

[56] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks[M]. Heidelberg: Springer, 2012: 385.

[47] HUBEL D H, WIESEL T N. Receptive fields and functional architecture of monkey striate cortex[J]. The Journal of physiology, 1968, 195(1): 215-243.

[48] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]Advances in neural information processing systems. 2012: 1097-1105.

[49] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time series[J]. The handbook of brain theory and neural networks, 1995, 3361(10): 1995.

[50] VAN DEN OORD A, DIELEMAN S, SCHRAUWEN B. Deep content-based music recommendation[C]Advances in Neural Information Processing Systems. 2013: 2643-2651.

[51] DOS SANTOS C N, GATTI M. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts.[C]COLING. 2014: 69-78.

[52] ZENG D, LIU K, LAI S. Relation Classification via Convolutional Deep Neural Network.[C]. COLING. 2014: 2335-2344.

[53] HU B, LU Z, LI H. Convolutional neural network architectures for matching natural language sentences[C]Advances in Neural Information Processing Systems. 2014: 2042-2050.

[50] MURPHY K P. Machine Learning: A Probabilistic Perspective [M]. Cambridge, MA: MIT Press, 2012: 82-92.

[51] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]. Proceedings of European Conference on Computer Vision, LNCS 8689. Berlin: Springer, 2014: 818-833.

[52] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C] . Proceedings of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1106-1114.

[53] DONAHUE J, JIA Y, VINYALS O, et al. DeCAF: a deep convolutional

activation feature for generic visual recognition [J]. Computer Science, 2013, 50(1) : 815-830.

[54] MURPHY K P. Machine Learning: A Probabilistic Perspective [M]. Cambridge, MA: MIT Press, 2012: 82-92.

[55] LeCun, L. Bottou, Y. Bengio, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[56] R Collobert, J Weston, L Bottou, et al. Natural language processing (almost) from scratch[J]. Journal of machine learning research, 2011, 12(1): 2493-2537.

[57] MD Zeiler. ADADELTA: AN ADAPTIVE LEARNING RATE METHOD[J]. Computer Science, 2012, 3(4): 325-330.

[58] Socher, B. Huval, C. Manning, et al. Semantic Compositionality throughRecursive MatrixVector Spaces[C]. Proceedings of the EMNLP. PA, USA. ACM Press, 2012: 1201-1211.

[59] Socher, A. Perelygin, J. Wu, et al. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank[C]. Proceedings of the EMNLP. Seattle, USA. Citeseer Press, 2013: 282-293.

[60] N Kalchbrenner, E Grefenstette, P Blunsom. A Convolutional Neural Network for Modelling Sentences[C]. Proceedings of the ACL. Baltimore, USA: arXiv Press, 2014: 1404-2188.

致　谢

时间如白驹过隙，转眼间又到了毕业季，这是我在千年学府湖南大学的第三个年头了，回想昨日种种，依然历历在目。三年的研究生生活，我收获的不仅仅是一本学历证书，更多的是各位老师的敦敦教诲和同学之间的无私帮助，使我对生活和人生产生了新的态度和认识。借此毕业之际，我想向在我人生道路上帮助过我、关心过我的各位老师、同学、朋友、亲人表示由衷的感谢。

首先，我要感谢我的导师杨超老师。杨老师宽广的胸襟、渊博的知识和敢为人先、实事求是的态度，像榜样一样一直激励着我不懈努力。杨老师是一个实干家，每次学术讨论班杨老师总是激励我们，做学术不能一纸空文，一定要动手编程和落实成果。本论文的完成得益于杨老师的悉心教导，从当初的选题到后面的实验以及最后的论文撰写阶段，杨老师都一直不遗余力的帮我纠正错误，没有杨老师的辛勤付出，本论文就无法按时完成。此外，感谢杨老师为我提供了一个学术氛围浓厚的科研环境以及配置优良的实验器材。研究生阶段，能够成为杨老师的学生我深感荣幸，杨老师和蔼可亲、关心学生的品质一直深深影响着我，再次感谢杨老师。

感谢实验室朱立民、卢兴沄、杨竞、李万里、查理佳、秦凯强、马孔强、李鲜小伙伴们，三年来和你们一起学习、玩耍，才使得我的科研生活那么充实、不枯燥。特别是朱立民，总是带着我们到处吃喝玩乐，带我们尝尽了人生百味、玩遍了大江南北。还有卢兴沄的坚持不懈，终于把我带入王者荣耀的世界。总之感谢你们的陪伴，为了我的研究生生活添加了更多色彩。

特别感谢我的父母，是你们的关心、鼓励和支持让我取得了今天的成绩。我会带着你们的期许，继续奋斗，用更好成绩回报你们。

最后，谨向审阅本论文及答辩组的老师们致以诚挚的谢意。由于本人水平有限，文中难免有不足之处，敬请各位老师批评、指正。

吕超

2017 年 5 月 10 日

附录 A 攻读硕士学位期间发表的论文

[1] 吕超, 杨超, 李仁发. 基于卷积神经网络和词语邻近特征的情感分类模型[J]. 计算机工程，2017 (CSCD，已录用)

[2] Chao Lv, Chao Yang, Renfa Li. Convolutional Neural Networks Based On Word Context Information For Sentiment Classification[C]. The 3th International Conference of Pioneering Computer Scientists, Engineers and Educators, 2017 (EI，已录用)

[3] 吕超, 杨超, 李仁发. 基于 LSTM 递归神经网络和人工判定规则的情感分析模型[J]. 中文信息学报，2017 (CSCD，审稿中)

[4] Chao Lv, Chao Yang, Renfa Li. Convolutional Neural Networks Based On Clause Polarity And Artificial Judgement Rules For Sentiment Classification[C]. The 6th Conference on Natural Language Processing and Chinese Computing, 2017 (EI，审稿中)

附录 B 攻读硕士学位期间参与的项目

[1] 国家自然科学基金[61173036]：以汽车为例的CPS若干问题研究

[2] 湖南省科技计划[2015GK3015]：基于RFID的景区智能导航物联网系统设计与关键技术研究

[3] 国家高技术研究发展计划（863）资助项目子项[2012AA01A301-01]：行业应用市场分析和电磁计算软件研发