

硕士学位论文

面向餐馆评论的情感分析

关键技术研究

**RESEARCH ON KEY TECHNIQUES OF
RESTAURANT COMMENT SENTIMENT
ANALYSIS**

张婷婷

哈尔滨工业大学

2017 年 6 月

国内图书分类号：TP391.3

学校代码：10213

国际图书分类号：681.37

密级：公开

工程硕士学位论文

面向餐馆评论的情感分析关键技术研究

硕 士 研 究 生： 张婷婷

导 师： 刘远超副教授

申 请 学 位： 工程硕士

学 科： 计算机技术

所 在 单 位： 计算机科学与技术学院

答 辩 日 期： 2017 年 6 月

授予学位单位： 哈尔滨工业大学

Classified Index: TP391.3

U.D.C: 681.37

Dissertation for the Master's Degree in Engineering

**RESEARCH ON KEY TECHNIQUES OF
RESTAURANT COMMENT SENTIMENT
ANALYSIS**

Candidate:	Zhang Tingting
Supervisor:	Associate Prof. Liu Yuanchao
Academic Degree Applied for:	Master of Engineering
Speciality:	Computer Technology
Affiliation:	School of Computer Science and Technology
Date of Defence:	June, 2017
Degree-Conferring-Institution:	Harbin Institute of Technology

摘 要

随着互联网与电子商务的发展,网上购物、网上订餐等方便快捷的应用日益深入人们的生活,相应地人们在这些平台上发表的评论信息也正在呈指数级的方式增长。这些信息数量庞大,拥有极其重要的研究价值。对这些评论信息进行分析,从中获得消费者对每个评价对象的评价极性,不仅能指导消费者的消费行为,而且有利于商家掌握消费者需求,从而对产品进行改进。

本文对餐馆评论领域评价对象的抽取和评价极性判别两个情感分析子任务进行研究,选择效果最好的方法应用于餐馆评论情感分析系统。具体地,本文研究内容如下:

首先,研究评价对象的抽取方法。提出基于输出依赖的双向 LSTM 模型,该模型在 LSTM 模型的基础之上通过利用两个独立的隐含层来对文本进行双向处理,从而充分利用文本上文和下文中所蕴含的有效特征,同时在输出层之间加入自连接,有效利用输出序列之间存在的依赖关系,并通过加入词性特征、句法特征、情感倾向特征和命名实体识别特征来提升模型的效果。其次,实现了条件随机场方法,主要在特征选择与组合上对模型的效果进行提升。此外,实现了基于 BLSTM-CRF 的评价对象抽取方法,将 BLSTM 的输出向量直接送入 CRF 模型中进行计算,得到最佳输出标签序列。

其次,研究评价对象极性判别方法。提出基于双向 LSTM 的评价对象极性判别模型,该模型利用两个 BLSTM 网络即 $BLSTM_L$ 和 $BLSTM_R$ 来分别收集评价对象的上文和下文语义信息,在每个时间步骤上将当前单词词向量和评价对象向量进行连接后一同送入模型,从而使模型能捕获到每个单词与评价对象之间的语义关系。该模型取得了同类模型的最好效果。此外,本文提出了基于提升的模型融合方法,该方法将支持向量机模型和随机森林模型融合,在训练完一个分类模型后,增大该模型错误分类的样本所占的权重并减小该模型正确分类的样本的权重,最后按照各模型的效果对结果加权得到最终的结果。该方法做到了将线性分类模型和非线性分类模型的优点结合。

最后,设计实现基于餐馆评论的情感分析系统。将基于输出依赖双向 LSTM 的评价对象抽取方法和基于双向 LSTM 的评价极性判别方法应用到系统中,提高了系统进行评价对象抽取与极性判别的准确性。该系统能够直观地以饼图的方式将评价对象及评价极性占比形象地表示出来。

关键字: 循环神经网络; LSTM; 评价对象抽取; 评价极性判别

Abstract

With the development of Internet and electronic commerce, applications such as online shopping and online ordering which are fast and convenient have penetrated deeply into people's daily lives. Accordingly, the comments which people delivered on these platforms are also increasing exponentially. These comments are in huge number and really have great research value. We analyze these comments to get the opinion polarity of every aspect term which not only can give consumers a lot of help on their consuming behavior but also can help businesses to understand consumer's demand and thus to improve the quality of their product.

In this thesis, we have a research on the method of aspect term extraction and aspect based sentiment analysis in the field of restaurant review. Then we apply the method which get the best result to the development of restaurant reviews sentiment analysis system. The study mainly includes the following aspects:

Firstly, this thesis has a research on the task of aspect term extraction. We implement a model named output dependent bidirectional LSTM. The model use two independent hidden layer to analyse the comment so that it can make full use of the context of the comment. Besides, the model add self-connection to output layer to take full advantage of the relationship between output labels. Moreover, this thesis implement the method of CRF and improve the result by feature selection and combination. Then, we implement the BLSTM-CRF aspect term extraction method, which put the result of BLSTM to CRF to compute, and then get the best output labels.

Secondly, this thesis focus on the method of aspect based sentiment analysis. We propose the model based on bidirectional LSTM which uses two BLSTM network that is $BLSTM_L$ and $BLSTM_R$ to get the context semantic information of the aspect term. Meanwhile, we put the current word vector and the aspect term vector together to get a new vector and the put the new vector into the model, so that the model can capture the semantic relationship between every word and aspect term. This model get the best result among models in the same class. In addition, this thesis put forward the the method of model fusion based on boosting which fuse SVM with Random Forest. The method increase the weight

of sample being incorrectly classified and decrease the sample being correctly classified after training the classifier. Finally, the final results are weighted according to the effect of each classification model. This method can combine the advantages of the linear classification model with the nonlinear classification model.

Finally, this thesis implements a sentiment analysis system based on restaurant reviews. We apply the output dependent bidirectional LSTM method and the aspect based sentiment analysis method based on bidirectional LSTM into the system which improve the accuracy of aspect term extraction and aspect based sentiment analysis in the system. The system can visually show people the proportion of the aspect term and sentiment polarity which consumer express to the restaurant by pie chart.

Keywords: recurrent neural networks, LSTM, aspect term extraction, aspect based sentiment analysis

目录

摘 要	I
Abstract	II
第 1 章 绪 论	1
1.1 课题背景及研究的意义	1
1.2 国内外研究现状	2
1.2.1 情感分析相关知识	2
1.2.2 评价对象抽取研究现状	3
1.2.3 评价极性判别研究现状	5
1.3 本文主要研究内容	6
1.4 本文内容安排	8
第 2 章 评价对象抽取方法研究	10
2.1 引言	10
2.2 问题定义与数据集	11
2.2.1 问题定义	11
2.2.2 实验数据集	11
2.3 基于条件随机场的评价对象抽取	13
2.3.1 条件随机场	13
2.3.2 特征提取	15
2.4 基于深度学习的评价对象抽取	18
2.4.1 循环神经网络	18
2.4.2 基于 BLSTM-CRF 的评价对象抽取	21
2.4.3 基于输出依赖的双向 LSTM 模型的评价对象抽取	22
2.5 实验设置与结果分析	28
2.5.1 评价指标	28
2.5.2 基线方法	29
2.5.3 实验结果及分析	29
2.6 本章小结	35
第 3 章 评价对象极性判别方法研究	36
3.1 引言	36
3.2 问题定义与数据集	37
3.2.1 问题定义	37

3.2.2 实验数据集	38
3.3 基于模型融合的评价对象极性判别	39
3.3.1 支持向量机模型	40
3.3.2 随机森林模型	41
3.3.3 模型融合方法	42
3.3.4 特征提取	44
3.4 基于深度学习的评价对象极性判别	46
3.4.1 基于 LSTM 的评价对象极性判别	46
3.4.2 基于双向 LSTM 的评价对象极性判别	47
3.4.3 句子预处理及未登录词词向量表示	48
3.4.4 模型训练	49
3.5 实验设置与结果分析	49
3.5.1 评价指标	49
3.5.2 基线方法	50
3.5.3 实验结果及分析	50
3.6 本章小结	55
第 4 章 餐馆评论情感分析系统的设计与实现	56
4.1 引言	56
4.2 餐馆评论情感分析系统的总体结构与设计	56
4.3 餐馆评论情感分析系统各模块详细设计与实现	57
4.3.1 用户评价交互模块	58
4.3.2 数据获取模块	58
4.3.3 评论文本分析模块	58
4.3.4 结果数据存储模块	59
4.3.5 统计信息展示模块	60
4.4 系统功能展示	60
4.4.1 系统登录	60
4.4.2 评价对象占比图	61
4.4.3 评价极性占比图	62
4.4.4 用户评价交互	63
4.4.5 评价信息分类浏览	64
4.5 本章小结	64
结 论	65

参考文献	67
哈尔滨工业大学学位论文原创性声明和使用权限	72
致 谢	73

第 1 章 绪 论

1.1 课题背景及研究的意义

随着互联网及电子商务的飞速发展，人们与互联网的关系越来越密切。人们不仅可以在互联网上获取他人观点也可以分享自己的观点，相关的评论信息能够分为三大类：产品及服务评论、事件评论和人物评论。这些评论信息数量庞大，难以梳理，但却有着极其珍贵的钻研价值。情感分析的任务就是对这些评论语句进行研究，然后利用自然语言处理的相关方法提取出评论语句中重要的信息。

在以上三类评论信息中，与人们日常生活更加紧密相关的是产品及服务评论，餐馆评论就属于其中一种。线上订餐、线上购物等电子商务应用具有方便和高效的特点，能轻而易举地被人们接受，已然成为一种势不可挡的潮流，人们足不出户即可利用互联网满足日常生活所需。随着电子商务的普及，众多餐饮网站上的评论呈指数级增加，这些评论信息内容丰富且有着非常重要的研究价值。高效且正确地抽取出评论信息中的评价对象及对应的情感极性具有极其重要的作用。

首先，高效准确地从评论文本中抽取出评价对象及其对应的情感极性可以使潜在的消费者更好地了解餐馆。现有的大多数网上餐饮平台只是简单地列出消费者的历史评论信息，并未做详细的归纳，潜在的消费者需要逐一阅读评论信息才能找到有价值的评论。但由于评论信息数量庞大，消费者很难在短时间内获取自己想要的信息。评价对象抽取和评价对象极性判别方法的引入可以帮助消费者快速了解餐馆的各个方面的情况，提高用户体验。

其次，评价对象抽取及评价对象极性判别技术可以帮助商家了解消费者需求从而进行产品改进。日益增长的评论信息使得商家很难人工地从中获取消费者的有价值的意见或建议，这给商家及时了解消费者反馈并做产品升级带来了困难。高效准确的评价对象抽取和评价对象极性判别方法能够使商家快速地获得消费者对餐馆的各个方面的意见，以便商家能够根据大众的需求来改善自己的菜品及服务质量，满足更多用户的需求，从而获得更多盈利。

因此，从海量数据中高效准确地挖掘出用户的评价对象及情感极性，然后采用数学方法统计分析，并将结果直观地展示出来，是一项极其有价值的研究任务。研究如何高效准确地从评论文本中进行评价对象抽取和评价对象

极性判别是一项非常有意义的课题。

1.2 国内外研究现状

情感分析在最近十几年中都是自然语言的热门研究课题。本小节首先阐述了情感分析的相关知识，然后详尽分析了评价对象抽取的研究现状，最后介绍了评价极性判别的研究现状。

1.2.1 情感分析相关知识

情感分析又称为倾向性分析、情感分类、主观性分析等，是对带有情感倾向的文本运用自然语言处理和机器学习的方法进行分析，从而辨别句子中的评价对象和情感倾向。情感分析的应用十分广泛，企业管理、电子商务、网络舆情管理等方面都用到了情感分析。例如，在电子产品的情感分析中，首先需要从评论文本中抽取出某电子产品的相关属性即评价对象，例如手机的价格、质量、售后、内存、外观等，接着利用相关的方法对每个属性的情感倾向性进行判别。

目前，国内外学者对于情感分析的研究可分为两类，即情感信息抽取和情感极性分析。情感信息主要由三个情感元素组成，情感词、情感对象和情感持有者，对于情感信息抽取的研究则主要有情感词的分析、情感作用对象的提取和情感持有者的挖掘等。也有不少学者针对上述三种情感元素的某些组合搭配做了较为深入的探索和研究。对于情感极性的分析大多都是句子级情感分析^[1]和篇章级情感分析^[2]。情感分析分为两种，一种是主客观情感信息的区分^[3]；一种是积极情感和消极情感的区分及其情感强度的分析^[4]。

按照情感分析对应的文本的粒度划分可以分为四种级别，分别是篇章级、要素级、句子级、词汇级。根据不同的粒度划分的不同的任务有不同的任务目标和应用：篇章级的任务目标是观点提取及判别文本的主观客观性；要素级的任务目标是评价对象的挖掘及观点表达者的挖掘；句子级的任务是观点句的提取；而粒度最小的词汇级任务是抽取情感词和判断情感极性。其中，电影评论等长篇幅的评论更适合篇章级情感分析，而商品评论这种短语句评论更加适合要素级的情感分析。

根据对应的文本的种类划分可以分为电影评论分析、产品评论分析、书评分析、微博分析、论坛分析等等。不同的种类的文本的特点不同，而文本的特点又决定了进行情感分析的方法。针对不同类别的文本进行情感分析所

采用的方法也会有所差异。比如，电影评论跟微博评论的差别就比较大，微博评论是针对不同事件、人物或话题的，而且微博多为口语化，情感要素的抽取会更加困难一些，而电影评论仅仅是针对某个电影，且都是用书面语表达，因此电影评论的意见挖掘的效果会更好一些。

1.2.2 评价对象抽取研究现状

评价对象抽取的目标是通过一系列有效的方法找出评论文本中的核心评论对象，即评论观点所针对的对象。如书籍评论的作者、体裁、写作手法，产品评论的价格、质量、售后，微博评论的主人公、事件等等都是评价对象。评价对象抽取作为情感分析的一个首要工作服务于情感分析的其他后续工作，是自然语言处理的研究热点，其研究方法可以归结为三类，下面进行详细介绍。

(1) 基于规则和无监督的方法

基于规则和无监督的方法的产生是由于二十一世纪初电子商务还未像现在一样普及，标注好的评论数据极其稀少。Hu 和 Liu^[7]最早提出名词挖掘技术，他们首先对语料进行词性标记获得语料中所有的名词，然后再通过经典的 Apriori 算法来获得对应评论文本中的频集，接着通过词集剪枝、稀疏剪枝和冗余剪枝来获得评价对象。Popescu 和 Etzion^[8]则利用名词短语和评价对象之间的点互信息 (Point Mutual Information) 对名词短语进行了更进一步的筛选，提升了算法的效果。Blair-Goldensohn 等人^[9]则在对评论语句赋予权重的基础上记录名词短语的频率并按权重和频率的计算结果对其排序，最后取排序中靠前的名词短语作为评价对象。Zhu 等人^[10]利用 Cvalue 度量构建了一个包括所有待筛选的评价对象的集合，然后运用 Bootstrapping 算法筛选出集合中的评价对象。

一般地，评论语句中的每一个评价对象和其对应的情感修饰词会蕴含着密切的相互依赖关系。Kessler 和 Nicolov^[11]、Somasundaran 和 Wiebe^[12]通过解析评价对象和情感词的依存关系来抽取评价对象。Qiu 等人^[13]在依存关系的基础上提出双向传播方法，将评价对象和情感词一起抽取出来。Wu 等人^[14]提出先根据评价语句中短语之间的相互依赖关系筛选出候选评价对象，之后再利用一些语言模型进行评价对象抽取。Hu 和 Liu 根据经验将评论语句中的评价对象分为显式的和隐式的。前者是指句子中的被评价词修饰的名词或名词短语，而后者是利用评价词反向推导出来的，一般是利用形容词和动词进行推导，隐式评价对象是隐含在评论文本的上文与下文的语义中的，并

没有显式地出现。例如“这家餐厅虽然位置有点偏僻，但做的很好吃。”这句话中情感词“偏僻”的评价对象是“位置”，而由动词“做”和形容词“好吃”可以推导出其评价对象是“菜品”。Su 等人^[15]提出利用聚类方法映射出隐式评价对象，通过显示评价对象与评价词的共现关系来得到他们之间的映射关系。Hai 等人^[16]首先根据显示评价对象和评价词之间的共现频率生成初步的情感词和显示评价对象的关联规则，然后通过聚类产生鲁棒性更强的评价词与显示评价对象之间的关联规则。

（2）基于机器学习的方法

近年来，互联网的飞速发展使得网络上出现了越来越多的评论信息，庞大的评论数据促进了基于机器学习的评价对象抽取方法的研究。Wong 和 Lam^[17]利用隐马尔科夫模型^[18]（Hidden Markov model, HMM）来识别出网页中的所涉及的商品属性。在 HMM 模型中，他们利用计算出的模型内部的发射概率和转移概率以及隐藏层状态来进行训练。Jin 和 Ho^[19]提出了词汇化的隐马尔科夫模型（Lexicalized HMMs）。该模型在 HMM 的前提下整合单词及其词性，使用 Viterbi 算法抽取出评价对象。由于 HMM 无法利用上文和下文的有效特征，很难获取评论文本中长距离的依赖关系，导致其对于某些情况无法给出较好的结果，而条件随机场模型^[20]（Conditional Random Field, CRF）不存在这些缺点而且能充分利用上文及下文信息同时能避免标注偏置问题。Toh 和 Wang^[21]利用标注好的语料库训练 CRF 模型，他们使用工具提取出命名实体识别特征和词聚类特征等并结合基本的句子表层特征作为 CRF 模型的特征，最终取得不错的结果。Zhang^[22]利用规则与统计相结合方法来抽取评价对象。Jakob 和 Gurevych 基于各个评论领域使用了条件随机场模型，并提出词性、句距、依存语法和意见句四个特征来提高模型效果。Li 等人^[23]将 Tree-CRF 和 Skip-CRF 两类 CRF 模型以特殊的方式融合在一起来进行评价对象抽取。这种传统的序列标注问题的方法是将句子本身作为着手点，通过句子表层特征、句法特征、语法特征出发，构造一系列特征，最后通过模型预测待标注序列的标签。Toh Z 和 Su J^[23]等人利用 CRFsuite 工具来抽取评价对象，他们提取了单词特征、词聚类特征、命名列表特征及句法分析树中每个单词的首个字母特征等，取得了较好的效果。

（3）基于深度学习的方法

深度学习技术的发展使得将自然语言处理问题与深度学习方法的结合越发成为学术界和工业界的热门话题。循环神经网络（Recurrent Neural Networks, RNN）^[25]是深度学习领域的重要模型，基于深度学习的序列标注

问题就是通过 RNN 的“记忆”功能，保证了上下文信息的可用性，进而完成对序列的标注或者预测。在评价对象抽取任务中，Irsoy 和 Cardie^[26]首先提出利用循环神经网络模型来抽取评价对象。他们分别训练了单隐含层的循环神经网络、多隐含层的深度循环神经网络和双向深度循环神经网络等。实验利用 Yang 和 Cardie^[27]的 CRF 模型作为基线系统，输入则利用 Mikolov^[28]提出的词的分布式表达，结果证明 RNN 能显著提高评价对象抽取的准确率，其中双向深度循环神经网络实验效果最佳。针对 RNN 存在的梯度爆炸问题，Liu 等人^[29]提出了含有特殊构造即门结构的 RNN 模型，包括长短期记忆网络模型^[31]（Long Short Term Memory, LSTM）和门循环网络模型^[32]（Gated Recurrent Units, GRU）和 Jordan 网络^[33]，并通过实验对比了这些模型与传统的 Elman 网络在数据集上的效果。

1.2.3 评价极性判别研究现状

评价极性的判别方法按其研究粒度可以分为粗粒度和细粒度的。粗粒度的研究方法是判断评论文本的整体的评价极性，而细粒度的研究方法是针对评论文本中的每个评价对象分别判断其评价极性。

基于统计自然语言处理的粗粒度研究方法首先对评论语句进行预处理，然后进行文本表示即特征项选择、降维和权重计算，最后通过分类器得到对应的评价极性。在特征项选择方面，Turney^[34]最早提出 5 种由副词或形容词组成的词性组合来识别情感极性。Mullen 等^[35]在 Turney 的基础上利用 SVM 分类器进行实验，特征项为通过 WordNet 计算出的形容词的 EVA(evaluative)、POT(potency)和 ACT(activity)值和价值词组的 SO(semantic orientation)值的组合。Pang^[36]等分别用以词频为权重的 Unigrams 和以布尔值作为权重的 Unigrams、Bigrams、Unigrams+Bigrams 作为特征项进行实验，最终得出分类效果最好的是使用布尔值作为权重的 Unigram 特征项。Ng^[37]提出选取句法结构作为特征项，具体做法为根据句子中的依存关系使用从句子中提取出的主谓结构和动宾结构作为特征进行情感极性分类。在特征项降维方面，唐慧丰等人^[38]分别利用互信息法、信息增益法、统计量法、文档频率法进行实验，实验表明信息增益法在评价极性判别任务中的效果较好，原因在于信息增益法同时考虑了类别信息和低频词对结果的影响。在特征项权重计算方面，Pang 等经过实验证明布尔权重法在评价极性判别中的准确率高于其他权重计算法。

基于语义特性的方法通俗来讲即利用词语的语义对整个评论文本的评价

极性进行判别。Turney 提出首先利用逐点互信息量 (PMI, Pointwise Mutual Information) 来计算两个词存在的语义相关性, 然后利用 SO 来计算词组的语义倾向性, 最终再计算进一步文本中的情感词组的 SO 值来获取评价对象的极性。

细粒度的评价极性判别方法包括传统的分类器方法和深度学习方法。前者将评价极性判别看作简单的分类问题, 通过传统的机器学习方法进行训练。Joachim Wagner 等^[39]使用 SVM 作为基础分类模型, 并使用了 Bag-of-N-gram 特征做为模型的输入, 该特征可以有效表示当前评价对象单词、词性、语义倾向等信息, 取得了不错的效果。Svetlana Kiritchenko 等^[40]创建了适合于其任务领域的情感词典并提取情感词典特征、评价对象的上下文特征、句法分析树特征等利用 SVM 作为分类器进行实验, 最终取得较好效果。

基于深度学习的方法通过神经网络能够学习评论文本中的有效特征。针对细粒度的评价极性判别任务, Duyu Tang 等^[41]提出了一种深度记忆网络模型, 该模型包含一个 Memory 和多个 Computational layer, Memory 是可读写的, 用来存储上下文的语义信息从而解决长期大量的记忆问题, Computational layer 包括一个 Attention layer 和一个 Linear layer, 他们的输入是评价对象词向量, 输出是 Memory 中比较重要的部分, 最后将他们的输出内容作求和运算作为接下来的 Computational layer 的输入, 最后一层的输出送入 Softmax 中进行分类。该模型还使用了 Attention 机制, 最终获得了不错的效果。

综上所述, 在评价对象抽取和评价对象极性判别任务中, 传统的方法通常依赖于提取的特征的好坏, 比较费时费力。基于深度学习的方法不需要提取复杂的特征, 是一种“端到端”的网络模型, 直接将词向量作为特征输入, 经过多层网络结构获取句子的语义表示, 取得了很不错的效果。因此, 本文针对两个任务分别提出传统的基于机器学习的方法和基于深度学习的方法进行实验, 并对现有模型进行优化改进从而提高模型效果, 最后对实验结果进行深入分析。

1.3 本文主要研究内容

本文针对评价对象抽取和评价对象极性判别任务分别用传统的机器学习方法和基于深度学习的方法进行了研究, 不断对模型进行优化, 给出实验结果并作深入分析, 最后利用效果最好的方法构建了一个基于 Web 的餐馆评论情感分析系统。主要研究思路如图 1-1 所示。

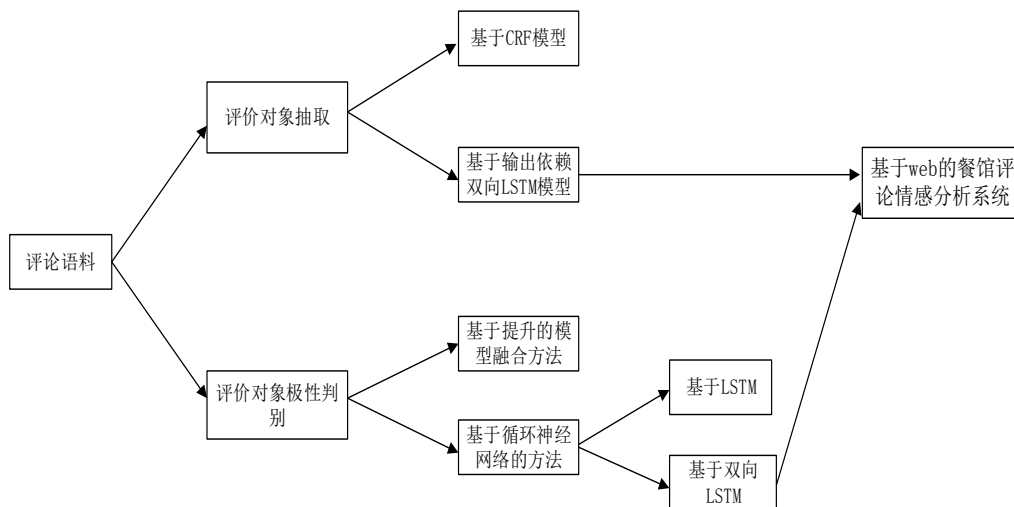


图 1-1 本文主要研究内容及思路

在评价对象抽取任务中，我们分别实现了基于特征的 CRF 模型、BLSTM-CRF 模型和基于输出依赖双向 LSTM 模型。CRF 模型是基于特征与标签的相互关系的概率模型，这就意味着特征的选择对于模型的效果极为重要，本文从特征的角度提升 CRF 模型在评价对象抽取任务中的效果，在词汇特征和词性特征两个基础特征的前提下，提出了各种有效的特征。CRF 模型需要进行繁琐的特征提取过程，费时费力，为此，本文在传统长短记忆网络模型的基础上提出“端到端”（End to End）的输出依赖的双向 LSTM 模型，该模型使用两个长短期记忆网络结构分别作用于序列的正向和反向，他们的输出将一齐传至输出层，这样在输出层就可以根据上文信息和下文信息做出综合性判断，同时在输出层之间加入自连接，充分利用输出层之间的依赖关系，除此之外，本文在输出层还加入简单的词性特征、句法分析特征、情感倾向性特征和命名实体识别特征来提升模型效果。该模型不仅解决了传统循环神经网络模型的“梯度爆炸”难题，而且能有效利用上下和下文信息来提高模型的表现力。BLSTM-CRF 模型将 BLSTM 的输出向量送入 CRF 模型中进行分类，输出最佳的标签序列，该方法也取得了不错的效果。

在评价极性判别任务中，我们提出了基于提升的模型融合方法和基于 LSTM 和双向 LSTM 的方法。其中，基于提升方法的模型融合算法使用一种新型的提升方法（Boosting）将 SVM 模型和随机森林模型进行融合。该模型依次训练各个分类模型，在每次训练得到的结果基础上改变训练数据集合的概率分布，即增加被前一分类器不正确分类的数据的权重同时降低被前一分类器正确分类的数据的权重。当所有模型训练完毕后，再按照各模型效果对

结果加权得到最终的模型结果，即加大分类误差较小的模型在最终结果中的表决权重，减小分类误差较大的模型对终极结果的表决权重。基于 LSTM 的方法利用两个 LSTM 网络，分别为 $LSTM_L$ 和 $LSTM_R$ ， $LSTM_L$ 收集评价对象的上文语义信息， $LSTM_R$ 收集评价对象的下文语义信息，然后将 $LSTM_L$ 和 $LSTM_R$ 的末尾隐含层的输出向量进行连接操作得到代表当前评价语句的向量，接着将该向量输入到 Softmax 层计算该评价对象属于每个评价极性的概率，从而得到最高的概率值所对应的极性类别。基于双向 LSTM 的方法使用两个双向 LSTM 结构，分别为 BLSTML 和 BLSTMR 来收集评价对象的上文和下文语义信息，然后分别将两个 BLSTM 的各个隐含层的输出送入均化池进行整合得到代表上文信息的向量和代表下文信息的向量，最后将两个代表上下文信息的结果向量连接后输入 Softmax 中进行分类。

通过前两章对评价对象抽取及评价对象极性判别的研究，结合 Java 开发、JavaScript 开发及 MySQL 数据库相关技术，研发了一个基于 Web 的餐馆评论情感分析系统。该系统的原始数据是利用爬虫工具从餐饮网站上爬取下来，然后经过一系列处理，利用相应的情感分析接口将评价对象及所对应的极性抽取出来并存入 MySQL 数据库。系统对消费者对餐馆的评论进行统计，以饼形图的方式直观地向用户展示所有用户评论中各个评价对象所占的比例及数量，并将每个评价对象的四种评价极性所占的比例及数量以饼形图的方式展示出来，这样用户便对其他用户对该餐馆的评论观点一目了然。此外，用户可以在系统中发表对于餐馆各个方面的意见，同时能够根据某个评价对象查询其他用户对餐馆该评对象的所有评论信息。

1.4 本文内容安排

本文共分四章，其主要研究内容与论文结构如下：

第 1 章首先阐述本文所研究课题的背景及研究意义；然后介绍课题的国内外研究现状；最后全面总结本文的主要研究思路及研究内容。

第 2 章介绍对评价对象抽取方法的研究。本章首先对评价对象抽取的定义及数据集特点进行详细介绍；然后介绍基于 CRF 模型进行评价对象抽取的一些工作，包括 CRF 模型的定义和特征的选择与提取；接着简要介绍循环神经网络模型和长短期记忆网络模型；之后介绍基于 BLSTM-CRF 模型的方法，主要介绍模型的整体结构；然后引出本文提出的基于输出依赖双向 LSTM 模型，并详细介绍模型的整体结构和隐含层结构等；最后，针对基于条件随机场的方法和基于输出依赖双向 LSTM 的方法分别设置优化对比实验，探索使

模型效果达到最佳的改进策略,同时给出基线方法、基于条件随机场的方法、基于 BLSTM-CRF 的方法及基于输出依赖双向 LSTM 的方法的实验结果,并对实验结果做对比分析。

第 3 章首先对问题定义及数据集特点进行介绍;然后介绍我们提出的基于提升的模型融合方法,包括支持向量机模型、随机森林、模型融合的方法;接着介绍本文提出的基于 LSTM 的和基于双向 LSTM 的评价对象极性判别模型,对两个模型的整体结构进行了详细介绍;最后对基于 LSTM 的模型和基于双向 LSTM 的模型中的三个关键部分的优化改进策略进行对比实验,同时给出基线方法、基于提升的模型融合方法和基于循环神经网络的方法的实验结果并作对比分析。

第 4 章利用前两章的餐馆评价对象抽取和评价极性判别的相关研究结果,结合 Java 开发、JavaScript 开发及 MySQL 数据库的相关知识,设计实现基于 Web 的餐馆评论情感分析系统,大大提高了系统进行情感分析的效果,并将所有用户评论的评价对象及评价极性以饼形图的方式展示给用户,不仅省去了用户逐条查看各个评论所花费的时间,而且更加方面直观。系统不仅对于指导用户的消费行为有重要作用,而且对于商家根据大众意见对餐馆各方面进行改善也有很重要的意义。

第 2 章 评价对象抽取方法研究

2.1 引言

用户对餐馆的评价通常会包含对餐馆的各个方面的评价,比如餐馆的环境、菜品味道、服务质量、价格等等。我们要把握用户对餐馆的看法,首先需要从评价语句中抽取出用户表达观点的相应对象,也就是评价对象,在此基础上才能针对评价对象进行情感极性的判别。评价对象就是指评价语句中用户表达观点所针对的对象,一个评价句中可能包括多个评价对象,抽取出细粒度的评价对象可以方便用户对餐馆的不同方面的性能有整体了解。

本章将详细介绍评价对象抽取方法,我们分别用机器学习和深度学习两类方法进行评价对象抽取的研究,这里机器学习的方法我们提出基于特征的 CRF 模型,深度学习的方法我们在 BLSTM-CRF 模型的基础上提出基于输出依赖的双向 LSTM 方法,经过对模型各个关键部分的探索对比实验,得到效果最好的模型用于情感分析系统。

CRF 模型在自然语言处理的各个领域都有普遍的运用且取得了较为显著的效果。CRF 模型是在 MaxEnt 和 HMM 的基础上提出来的,HMM 假设观察序列的元素之间相互独立,每个时间步骤上的观察效果仅跟当前时间上的状态有关,导致其不能考虑上下文特征,MaxEnt 只能获取局部最优值而无法得到全局最优,并且存在标记偏置问题,而 CRF 模型能够利用其较强的逻辑推理能力从文本的上下文信息中提取有价值的特征,并且解决了标记偏置问题。因此,我们选择 CRF 模型抽取评价对象,在基本特征的基础上提出新的特征,并对特征进行归纳总结,通过特征组合实验获取结果最好的组合。

近年来,深度学习方法在各个领域拥有越来越重要的地位。而深度学习中的循环神经网络以其端对端的学习模式省去了人工对原始数据的处理和特征提取等步骤,不仅节约了人力成本,也可以减少人为的误差,提升系统的准确性。因此,我们将会利用循环神经网络及对其改进后的新网络结构进行实验,接着对两种模型进行对比总结,选择效果更好的模型用于后续系统。

本章的主要内容安排如下:第 2.2 节对评价对象抽取任务及数据集进行了简要介绍;第 2.3 节介绍基于特征的条件随机场方法,主要包括 CRF 简介和特征的选择两大部分。第 2.4 节介绍了基于深度学习的评价对象抽取方法,包括对两种常见循环神经网络模型的介绍、基于 BLSTM-CRF 模型的评价对

象抽取方法的介绍及对本文提出的输出依赖双向 LSTM 的整体结构、隐含层结构、词向量训练、未登录词处理、特征选择、模型训练及实验设置的介绍。第 2.5 节首先介绍了基线方法，然分别针对基于条件随机场的方法和基于输出依赖双向 LSTM 的方法设置了不同的优化探索实验，探索能达到模型最好效果的方法，最后给出基线方法与本文提出的基于条件随机场的方法、BLSTM-CRF 模型和基于输出依赖双向 LSTM 的方法的实验结果并对实验结果进行对比分析。

2.2 问题定义与数据集

本节主要包括两个方面，一是介绍了评价对象抽取的问题定义及基本方法，二是详细介绍了本文用到的实验数据集。

2.2.1 问题定义

评价对象是指评价语句中用户表达观点所针对的对象，如书籍评论的作者、体裁、写作手法，产品评论的价格、质量、售后，微博评论的主人公、事件等等都是评价对象。评价对象抽取就是给定大量评论文本的情况下，使用自然语言处理或机器学习等技术学习出能够准确抽取给定评论文本中的所有评价对象的模型或算法，然后利用该模型或算法来抽取给定评论信息中的评价对象。评价对象抽取任务为情感分析的后续任务奠定了基础，如何正确得抽取评价对象有着非常重要的研究意义。

2.2.2 实验数据集

本节介绍了进行评价对象抽取任务的实验数据集的相关情况，包括：实验数据集的来源、数据集的表示方式、规模和数据集特点。

基于餐馆评论的语料很少，人工标注费时且成本高，因此我们选取第八届国际语义评测大会提供的评测任务中的评测数据¹，该评测数据集中全部为餐馆领域的评论数据，采用评测数据不仅节省人工标注的成本而且由于评测数据是根据统一的标准标注，其可信度及准确度较高。

该数据集详细情况如表 2-1 所示，训练集大约包括 3043 条标注好的评论语句，其中评价对象的数量大约为 3688 个，测试集大约含有 800 条评论语句，其中评价对象的数目大约 1157 个。训练集中极性为 positive 的评论文本大约

¹ <http://alt.qcri.org/semeval2014/task4/>

有 1200 多条，极性为 negative 的评论文本大约有 1000 多条，极性为 neutral 的评论文本大约有 200 多条，极性为 conflict 的评论文本大约有 600 多条。这些评论语句都是以可扩展标记语言（XML）格式存储的，其中包含的标签如表 2-2 所示。

表 2-1 数据集信息

数据集	评论数量	单词数量	评论对象数量
训练集	3043	47157	3688
测试集	800	12734	1157

表 2-2 训练数据中各个标签的名称及含义

名称	含义
Text	评价文本
AspectTerms	评价对象集
Term	评价对象
Polarity	针对评价对象的评价极性

具体某个评价语句的格式如图 2-1 所示，句子的 id 为 2983，该评论语句为“Nice atmosphere,the service was very good and the desert was nice.”，该评论文本包含的评价对象有三个，分别是 atmosphere、service 和 desert，其中，atmosphere、service 和 desert 的评价极性都是 positive。

```
<sentence id="2983">
  <text>Nice atmosphere, the service was very good and the desert was nice.</text>
  <aspectTerms>
    <aspectTerm term="atmosphere" polarity="positive" from="5" to="15"/>
    <aspectTerm term="service" polarity="positive" from="21" to="28"/>
    <aspectTerm term="desert" polarity="positive" from="52" to="58"/>
  </aspectTerms>
</sentence>
```

图 2-1 训练数据例子

餐馆语料与其他评论文本相比有其自己的特点。

（1）餐馆评论的表达方式更为口语化。如图 2-2 所示。口语化的表达会使评价极性判别任务更加艰巨。

```
<sentence id="2622">
  <text>Great food, great prices, great service.</text>
  <aspectTerms>
    <aspectTerm term="food" polarity="positive" from="6" to="10"/>
    <aspectTerm term="prices" polarity="positive" from="18" to="24"/>
    <aspectTerm term="service" polarity="positive" from="32" to="39"/>
  </aspectTerms>
</sentence>
```

图 2-2 口语化训练数据例子

(2) 评论语句的词数较少。如表 2-3 所示，对几种不同体裁的评论文本的信息进行了对比，可以发现，亚马逊书评和电影评论的词数都在 200 词以上，远多于新闻评论和餐馆评论。餐馆评论和新闻评论的长度差不多，原因在于餐馆评论和新闻评论都是评论者用一两句话发表自己的观点并没有展开来说。句子的词数少会导致在进行评价对象抽取任务时的特征稀疏问题。

表 2-3 不同体裁的评论的词数统计

类型	平均词数
亚马逊书评	271
电影评论	220
新闻评论	30
餐馆评论	16

2.3 基于条件随机场的评价对象抽取

本节介绍了利用基于特征的条件随机场方法进行评价对象抽取的一系列相关工作。首先详细介绍 CRF 模型，包括 CRF 模型的特点、应用场景及三个基本问题。然后基于基本特征的前提下，提出了一些扩展特征。

2.3.1 条件随机场

条件随机场，简称 CRF，是 Lafferty 于 2001 年首次提出的。条件随机场除了有判别式模型的优势外还拥有产生式模型能够利用上文和下文之间的转移概率，通过序列化方式来对全局参数进一步优化和解码的能力，这样，条件随机场就克服了其他判别式模型较难处理的标签中的偏差问题。

近年来，条件随机场理论在自然语言处理的诸多领域，如中文分词、歧义消解、手写体识别^[42]、命名实体识别^[43]等都有十分广泛的应用，并且取得了非常显著的效果。其整体结构如图 2-3 所示。

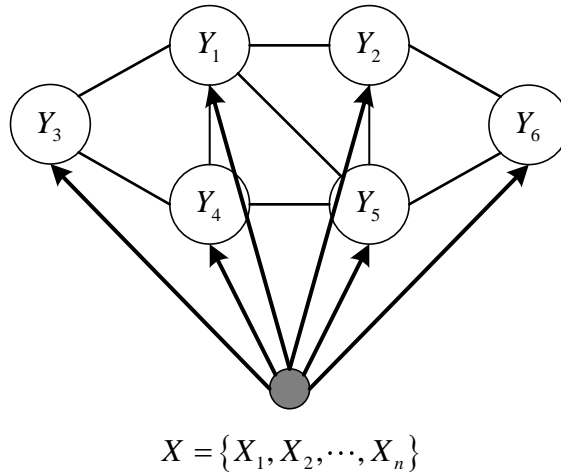


图 2-3 条件随机场 CRF 的结构图

在条件概率模型 $P(Y|X)$ 中， X 是输入变量，代表需要标注的观测序列， Y 是输出变量，代表标注序列。条件随机场利用极大似然估计方法获得相应的条件概率模型 $P(Y|X)$ ，然后针对输入序列 x ，利用学习到的条件概率模型 $P(Y|X)$ ，得到使 $P(y|x)$ 的值达到最大的序列 \hat{y} 即为输出序列。

若 $X=(X_1, X_2, \dots, X_n)$ 代表观测序列， $Y=(Y_1, Y_2, \dots, Y_n)$ 代表标注序列，线性链条件随机场的计算公式如下所示：

$$P(Y_i | X, Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n) = P(Y_i | X, Y_{i-1}, Y_{i+1}) \quad (2-1)$$

$i=1, 2, \dots, n$ （在 $i=1$ 和 n 时只考虑单边）

其参数化形式即假设 $P(Y|X)$ 为线性链条件随机场且随机变量 x 的值是 X ，随机变量 Y 的值是 Y ，则其条件概率如下式：

$$P(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} u_l s_l(y_i, x, i)\right) \quad (2-2)$$

其中，

$$Z(x) = \sum_y \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} u_l s_l(y_i, x, i)\right) \quad (2-3)$$

上式中， λ_k 和 u_l 是特征函数的权值， t_k 和 s_k 是特征函数， $Z(x)$ 是规范化因子。

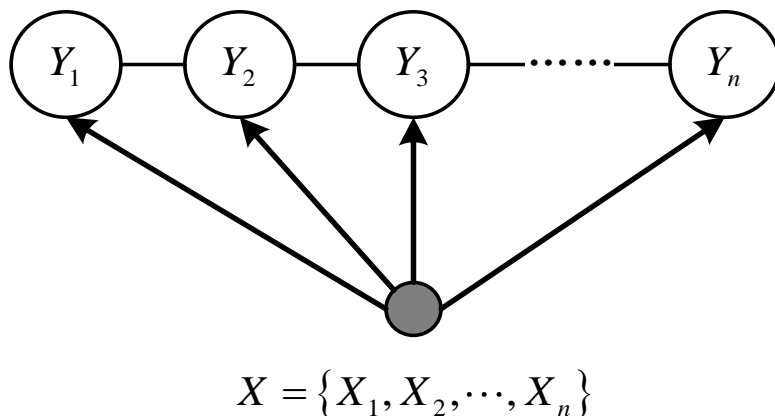


图 2-4 线性链条件随机场

条件随机场包含三个基本问题，首先是较为常见的概率计算问题，即在给出模型 $P(Y|X)$ 、输入序列 x 和输出序列 y 的前提下，采用如前向-后向算法之类的算法来计算 $P(Y_i = y_i | x)$ ， $P(Y_{i-1} = y_{i-1}, Y_i = y_i | x)$ 以及相应的数学期望的问题。其次是学习问题，即给出训练数据集的前提下估计 CRF 模型的参数，主要学习方法一般包含最大似然估计和正则化的极大似然估计。最后是预测问题，即在给出条件随机场 $P(Y|X)$ 和输入序列 x 的前提下，通过一系列算法计算出使得条件概率值达到最大的 y^* ，所使用的预测算法是维特比算法，即：

$$y^* = \arg \max_y P_w(y | x) \quad (2-4)$$

2.3.2 特征提取

CRF 模型是基于特征与标签的相互关系的概率模型，这就意味着特征的选择对于模型的效果极为重要。本节将介绍 CRF 模型在评价对象抽取任务中需要提取的特征，这些特征宏观上可以分为词汇特征、句法特征、情感倾向性特征和词向量特征。

2.3.2.1 词汇特征

词汇特征是虽然自然语言处理问题中最为常见的特征，但是它对于模型的效果却有着极其重要的作用。本小节将详细介绍实验中使用的词汇的特征。

(1) 单词特征

在评价对象抽取任务中，该特征对实验效果有着非常重要的影响。在某个具体的领域中，单词特征会直接决定其是否是评价对象。例如，单词“food”在餐馆评论文本中是评价对象，而单词“and”在所有领域的评价文本中都不会成为评价对象。因此，单词特征本身在评价对象抽取任务中是非常基础且重要的特征。

(2) 小写单词特征

实验数据是用户对餐馆的评论信息，没有统一的标准且对大小写没有限制，这就导致了评论文本中会存在许多大写形式的单词，这样，某个单词的大写形式与其小写形式会被区分为两种特征，这对于实验效果是不利的，因此我们将每个单词的小写形式作为特征。

(3) 词根特征

我们的实验数据是采用的英文语料，而英语单词的构成有个重要的特点即多数单词是由词根加词缀构成，而且词根的含义与其加上不同词缀后的含义近似。因此仅仅采用单词作为特征是不够的，因为假如仅使用单词作为特征的话，某个单词及其不同的变体会被识别为多个特征，所以我们引入了词根化特征。

(4) 词性特征

评价对象的词性一般都是名词或者是名词短语，而形容词或动词作为评价对象的可能性就很小，因此单词的词性特征对于评价对象抽取有着很好的指示作用。

2.3.2.2 句法特征

句法特征对模型的准确性有着极其关键的作用。本小节将详细介绍句法特征。

(1) 单词所在名词短语数目

该特征是在句法分析树中从起始根节点到单词节点的路径中所有的名词短语数量。如图 2-5 所示，“service”作为一个评论对象其在句法分析树中所在的名词短语数是 1，“was”不是评论对象，其在句法分析树中所在的名词短语数是 0。

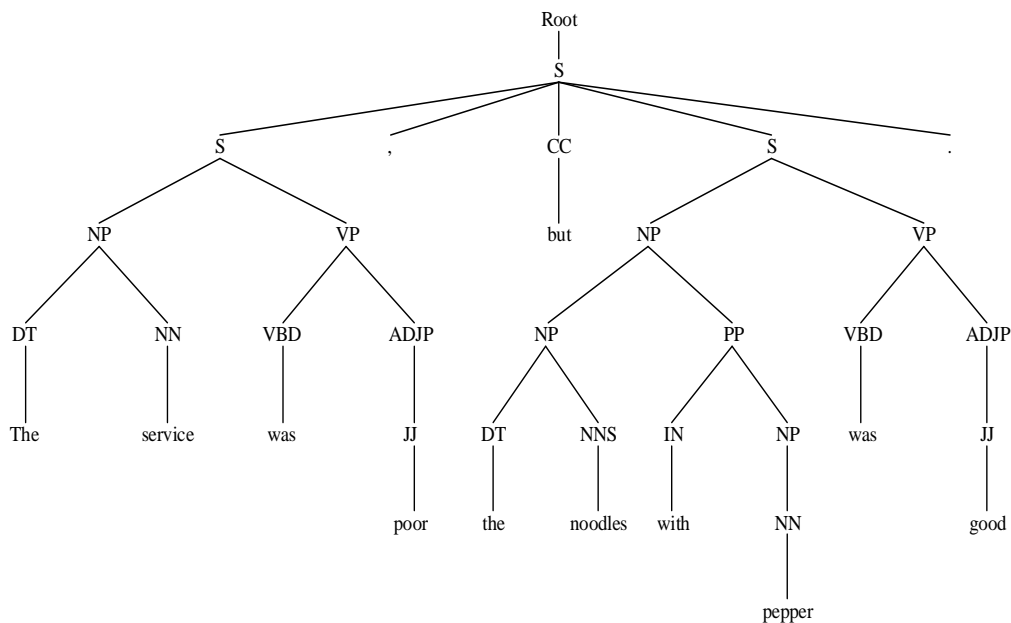


图 2-5 句法分析树实例

(2) 依存关系特征

利用依存句法分析提取出评论文本中词语间的各种关系，如主谓、修饰、介宾关系等等，这些关系能够帮助 CRF 模型“理解”当前词在句子中的作用，从而有效提升系统的准确性。

(3) 词距离特征

评价对象顾名思义即发表评论和抒发情感倾向的对象，而抒发情感倾向过程中必定会用到相应的情感修饰词，这些词的词性通常是形容词。因此，评论文本中与形容词距离较近的名词或者名词短语被判别为评价对象的可能性非常大。我们将使用形容词到单词的距离特征来对模型进行训练。

2.3.2.3 情感倾向性特征

在评价对象抽取任务中，具有情感倾向性的单词通常与评价对象单词是作为二元组的形式出现的。因此每个词的情感倾向性可以为评价对象抽取任务提供一定的信息量。因此，我们引入情感倾向性特征。

2.3.2.4 词向量特征

词向量是词语在语义空间内的特征表示，它的每个维度刻化着词语的某个隐含语义特征。词向量的用法主要包括两种，一种是作为神经网络模型的输入，一种是作为机器学习模型或算法的辅助特征。词向量作为机器学习算

法的特征在很多任务中都获得了非常好的效果。本文中，我们使用 Glove 词向量作 CRF 的一个特征来帮助提高模型的效果。

2.4 基于深度学习的评价对象抽取

在自然语言处理领域中，深度学习（Deep Learning）是目前较为主流方法，而循环神经网络（Recurrent Neural Network, RNN）^[44]及其变体长短记忆网络（Long Short Term Memory, LSTM）^[46]等在深度学习中占有重要地位，其应用都十分广泛。

本节首先简要介绍深度学习的两个常见模型即 RNN 和 LSTM，然后详细介绍基于 BLSTM-CRF 的评价对象抽取方法，接着给出本文提出的输出依赖双向 LSTM 模型的整体结构、隐含层结构，最后介绍词向量训练方法、未登录词的处理方法、特征选择、模型训练方法及实验的相关参数设置等。

2.4.1 循环神经网络

本节详细介绍了基于循环神经网络的两种常见模型，分别是循环神经网络模型和长短期记忆网络模型。

2.4.1.1 循环神经网络模型

RNN 模型与前馈神经网络的不同之处在于其增加了隐含层之间的自连接，从而将重要的历史信息沿着网络传播。其结构如图 2-6 所示。

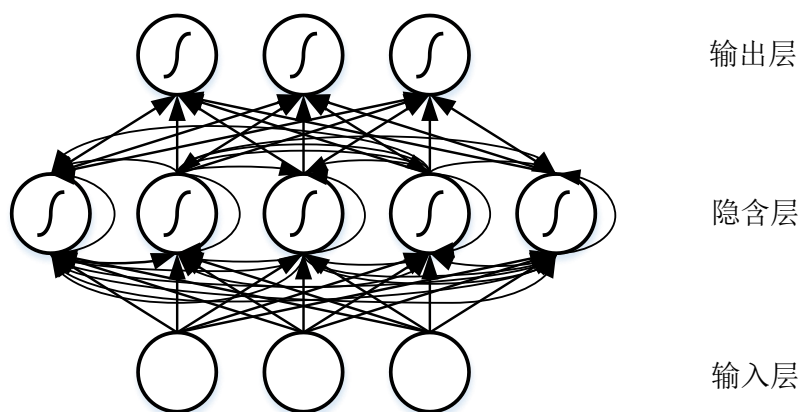


图 2-6 循环神经网络结构图

循环神经网络模型可以按输入的时间顺序来展开，我们将利用模型的展

开形式对模型进行分析，其展开形式如图 2-7 所示，循环神经网络隐含层的输入不仅来自当前时间输入层的输入还来自上一时间隐含层的输出，我们用以下公式表示：

$$\mathbf{h}^t = f_h(\mathbf{x}^t, \mathbf{h}^{t-1}) \quad (2-5)$$

式中 \mathbf{h}^t ——模型隐含层在第 t 时间的输出向量；
 \mathbf{x}^t ——模型输入层在第 t 时间的输入向量；
 f_h ——模型隐含层的激活函数，一般包括 logistic 函数，tanh 函数等。

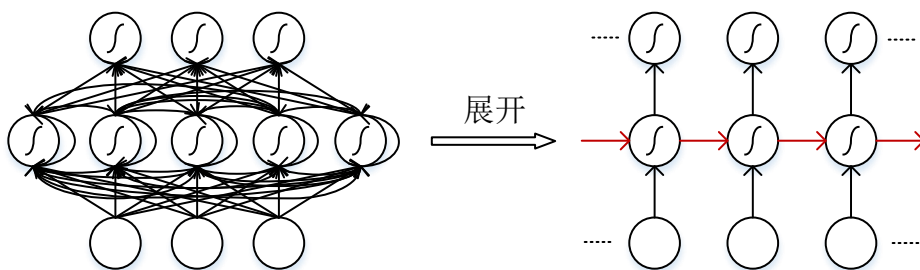


图 2-7 循环神经网络展开形式

在评价对象抽取任务中，模型的输入为 $\mathbf{x} = (\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^T)$ ，输出为 $\mathbf{y} = (\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^T)$ ，其中， \mathbf{x}^t 代表第 t 时间的输入词向量， \mathbf{y}^t 代表第 t 时间的输出标签向量。模型在第 t 时间步骤上的预测标签向量 \mathbf{y}^t 由以下式子表示：

$$\hat{\mathbf{y}}^t = \arg \max_l P_\theta(l | \mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^{t-1}) \quad (2-6)$$

式中 P_θ ——代表模型的函数；
 l ——代表模型所有可能的输出标签序列。

2.4.1.2 长短期记忆网络模型

循环神经网络模型只能考虑前几个短期时间间隔的依赖关系，无法保持较长时间间隔的依赖关系，也即存在梯度爆炸（Gradient Explosion）或梯度消失问题。为了规避循环神经网络存在的这一缺点，提出了长短期记忆网络 LSTM，它是在循环神经网络的基础上通过改变隐含层单元的计算方法实现的。其结构如图 2-8 所示。

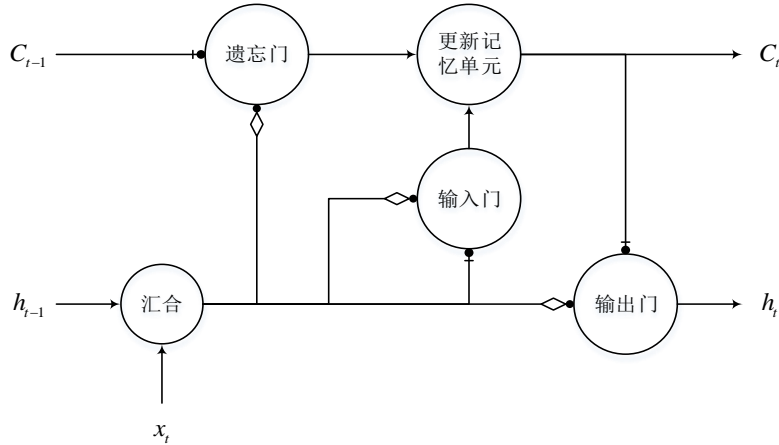


图 2-8 长短期记忆单元的结构图

由长短记忆单元的结构图可以看出，长短记忆单元存在一个贯穿始终的记忆状态单元（Memory Units） C ，该记忆单元是由输入门 i_t ，遗忘门 f_t 和输出门 o_t 一起来进行协同控制，最后根据当前记忆状态 C_t 计算当前隐含层单元的输出向量。具体的计算过程如下：

（1）当前输入记忆 \tilde{C}_t 是由前一隐含层的输出 h_{t-1} 和当前输入 x_t 进行计算，其计算公式如下：

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (2-7)$$

式中 W_c, U_c ——权值矩阵；

b_c ——偏置向量；

\tanh —— \tanh 激活函数。

（2）遗忘门 f_t 决定从历史记忆 C_{t-1} 中遗忘哪些信息，输入门 i_t 决定当前输入的哪些新信息被保留下来，输出门 o_t 决定当前隐含层输出哪些信息。其计算公式分别如下：

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2-8)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2-9)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (2-10)$$

式中 $W_f, U_f, W_i, U_i, W_o, U_o$ ——权值矩阵；

b_f, b_i, b_o ——偏置向量；

σ —— \logistic 激活函数。

（3）根据输入门和遗忘门更新当前记忆状态 C_t ，计算公式如下：

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \quad (2-11)$$

式中 $*$ ——两个向量按位相乘运算。

(4) 根据输出门和当前记忆状态 C_t 计算当前隐含层状态, 计算公式如下:

$$h_t = o_t * \tanh(C_t) \quad (2-12)$$

长短期记忆单元通过输入门、遗忘门、输出门和记忆状态单元之间的计算, 将不重要的信息遗忘掉并把重要的信息保留下来, 消除了梯度爆炸问题。

2.4.2 基于 BLSTM-CRF 的评价对象抽取

在基于 LSTM 的评价对象抽取方法中, 对输入数据经过 LSTM 进行各种运算后输出的结果向量通常都送入 SoftMax 中分类, 但是对于输出标签之间拥有依赖关系的情况, 这种利用 SoftMax 的方式效果就稍差一些。因此, 本文提出利用 LSTM 和 CRF 模型结合的方法, 该方法利用双向 LSTM 进行有效特征的提取, 利用 CRF 模型代替 SoftMax 从而捕获输出标签之间的依赖关系。

图 2-9 展示了模型的整体结构。该模型首先将评论语句中的单词送入 Embedding 层映射为其对应的词向量, 然后将单词分别送入正向 LSTM 和反向 LSTM 中进行处理, 正向 LSTM 和反向 LSTM 分别获取评价对象的上文特征信息和下文特征信息, 最后将双向 LSTM 输出的结果向量送入 CRF 中进行分类, 最终获得评论语句对应的最佳输出序列。

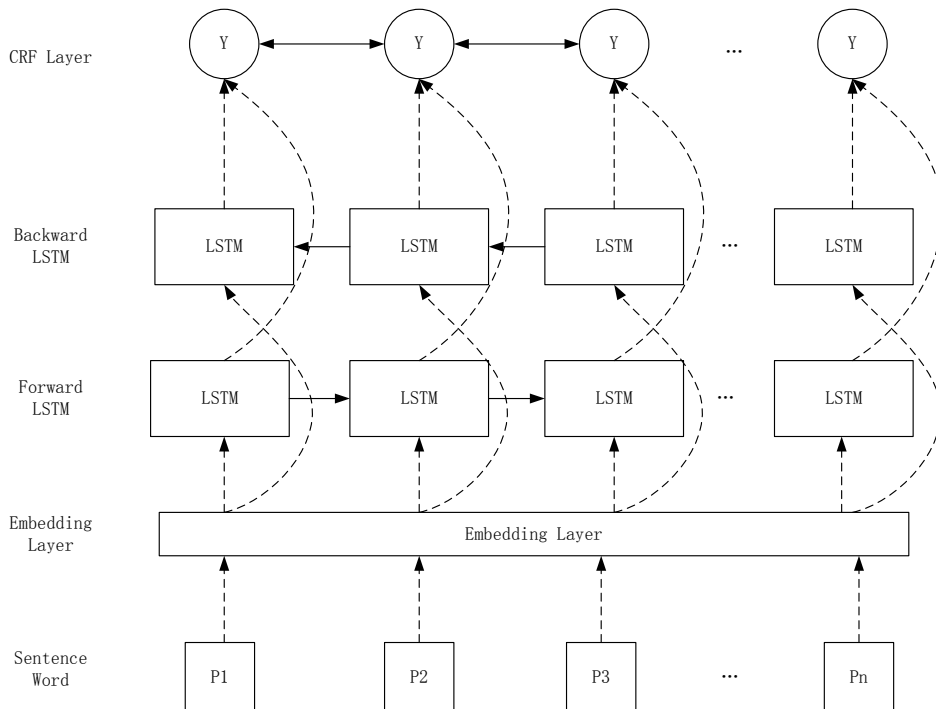


图 2-9 基于 BLSTM-CRF 的评价对象抽取模型结构示意图

BLSTM-CRF 模型的输出结果是通过动态规划算法计算出的最佳标签序列。对于输入数据 $x = (x_1, x_2, \dots, x_n)$ ，我们定义 LSTM 的输出概率矩阵为 $P_{n \times k}$ ，其中 k 是输出标签的个数。对于待预测的标签序 $y = (y_1, y_2, \dots, y_n)$ ，我们定义：

$$s(X, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=0}^n P_{i, y_i} \quad (2-13)$$

式中 A ——状态转移矩阵；

$A_{i,j}$ ——从第 i 个标签转移到第 j 个标签的概率；

$P_{i,j}$ ——评论语句中的第 i 个单词被判断为第 j 个标签的概率。

通过求得最大的 $s(X, y)$ 就可以获得评论文本对应的最优输出序列。

2.4.3 基于输出依赖的双向 LSTM 模型的评价对象抽取

在 BLSTM-CRF 模型的启发下，我们提出了基于输出依赖双向 LSTM 的评价对象抽取方法。本节对基于输出依赖的双向 LSTM 模型的评价对象抽取方法进行详细的介绍，包括输出依赖双向 LSTM 的整体结构、隐含层结构、词向量训练、未登录词处理、特征选择、模型训练方法及实验参数设置。

2.4.3.1 输出依赖双向 LSTM 整体结构

LSTM 模型利用输入门、遗忘门及输出门三种特殊的门结构对输入信息进行选择性的记忆，从而有效地解决了传统神经网络模型的梯度爆炸问题。但是问题在于 LSTM 也有自己的缺点，首先，其隐含层之间是从前向后连接的，也就是说它只能利用当前节点之前的信息而无法利用当前节点之后的信息。其次，对于评价对象抽取问题来说，输出标签之前是存在时序依赖关系的，而 LSTM 模型没有充分利用输出序列之间的这种依赖关系。针对以上两点，本文提出一种新型的模型，称为输出依赖双向 LSTM 模型。该模型首先通过在模型的隐含层加入一个正向 LSTM 结构和一个反向 LSTM 结构来分别捕捉上文特征信息和下文特征信息，解决了 LSTM 模型无法利用下文信息的缺点，其次该模型在输出层加入自连接，使得下一时刻输出层节点能够利用上一时刻输出层节点的输出信息参与计算，充分利用了输出序列之间的时序依赖关系，此外，我们提取了每个单词的词性特征、句法分析特征、情感倾向性特征和命名实体识别特征送入 SoftMax 中进行计算，提高模型效果。

如图 2-10 所示，该模型首先将句子中的单词通过 Embedding Layer 映射为词向量，然后将词向量送入双向 LSTM 结构单元中进行处理得到结果向量，接着将结果向量送到 SoftMax 里进行分类，得到分类结果。SoftMax 的分类

结果需要再输入下一时间步骤上的 SoftMax 层进行计算，这样 SoftMax 层的输入分别是隐含层的输出和上个时刻输出层的输出结果，从而充分利用输出序列间存在的相互关系，最后根据每个 SoftMax 层的结果输出其对应的词的标注结果。值得一提的是，我们提取的词性特征、句法特征、情感倾向性特征和命名实体识别特征大大提升了模型的效果。

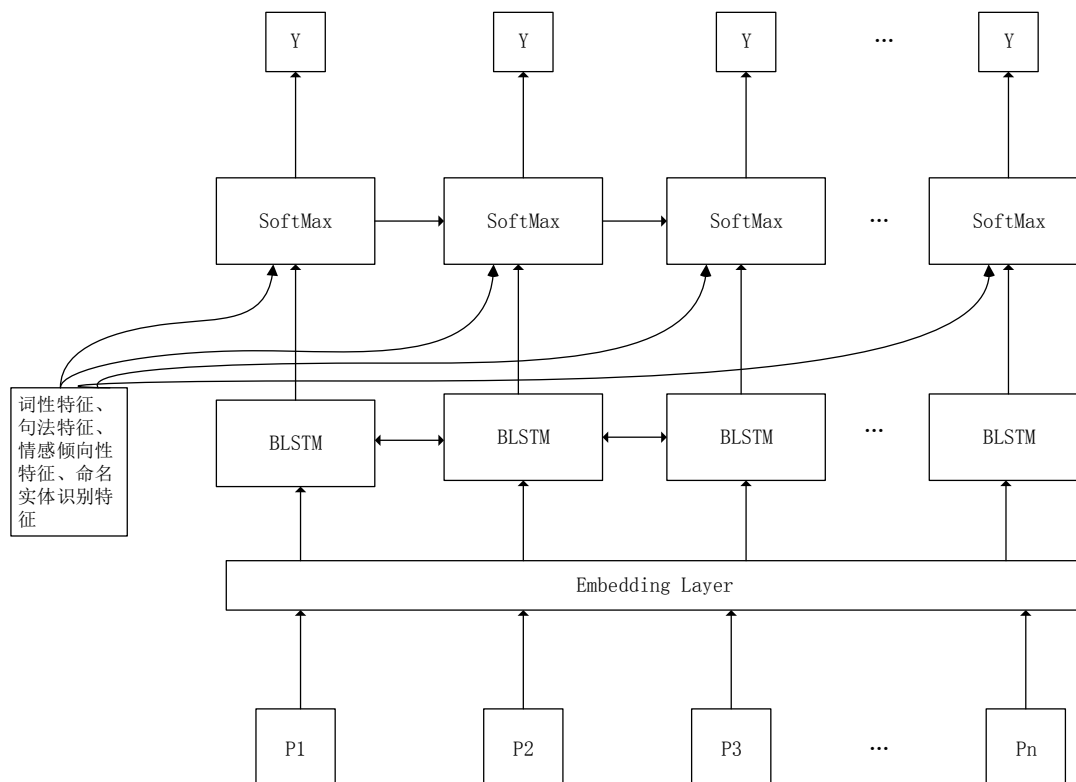


图 2-10 基于输出依赖双向 LSTM 的评价对象抽取模型结构示意图

LSTM 在 RNN 的基础上加入门结构后可以解决句子中的长距离依赖关系，但 LSTM 的问题在于其只能利用上文信息，针对这个问题双向 LSTM 模型通过在隐含层加入正向和反向 LSTM 来分别获取评论文本的上下文的信息，从而引入更多的特征，有助于提升模型的表现力。评价对象抽取任务的输出序列本身存在一定的时序依赖关系，因此为了充分利用这种依赖关系，本文将上一时间步骤上 SoftMax 层的输出结果输入到下一时间步骤上的 SoftMax 进行计算，解决了传统 RNN 模型不能有效利用输出序列中存在的相互依赖关系的问题。

2.4.3.2 输出依赖双向 LSTM 隐含层结构

模型的隐含层结构是由两个 LSTM 结构分别作用于输入序列的正向和反

向，两个 LSTM 结构之间没有权值连接，他们的输出将一齐传至输出层，这样在输出层就可以根据上文信息和下文信息做出综合性判断，提高的模型判别能力。该模型的隐含层如图 2-11 所示，词向量 X_0 分别输入到正向 LSTM 结构单元和反向 LSTM 结构单元中，然后正向 LSTM 利用其上文信息对输入信息进行处理，并将结果送入输出层与下一时间步骤的 LSTM 结构单元，反向 LSTM 利用反向的上文信息对输入信息进行处理，并将结果送入输出层与反向的下一时间步骤的 LSTM 结构单元，这样，输出层根据正向和反向 LSTM 的结果信息进行计算，输出最终结果 Y_0 。

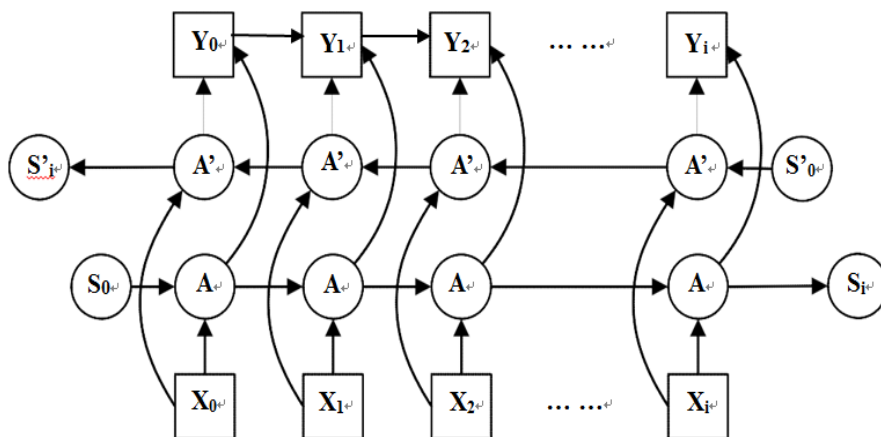


图 2-11 输出依赖双向 LSTM 的隐含层结构示意图

2.4.3.3 词向量训练

我们使用 Google 开源工具 word2vec 来进行词向量训练。由于已标注的评论信息数量有限，而训练词向量对语料的规模有很大的要求，因此我们额外加入了维基百科语料，为了使词向量更加适用于餐馆领域，我们在训练词向量的过程中加入额外的领域相关的数据进行词向量的训练，这里我们加入 Yelp 评论数据一同训练。通过 word2vec 工具训练出维度为 50、100、300 的词向量。

2.4.3.4 未登录词处理

测试数据集合中存在一些词在词向量中找不到其对应的表示，我们称这部分单词为未登录词汇，我们针对以下两种情况对于未登录词汇进行相应的处理。

首先，有些未登录词是由两个或多个单词复合在一起构成，例如

“dining-room”是未登录词，我们首先得到“dining”和“room”的词向量，然后对这两个词向量进行加和取均值得到“dining-room”的近似词向量。

其次，对于由于词汇中一个或几个字母拼写错误导致的未登录词，我们可以将词向量中与该词汇具有最小编辑距离的词汇看做该词的期望词汇。算法 2-1 列出了求两个单词最小编辑距离的算法。

算法 2-1：最小编辑距离算法

输入：

英文单词字符串 A。

英文单词字符串 B。

输出：英文单词 A 和英文单词 B 的最小编辑距离。

算法：

1. 首先建立动态规划矩阵，矩阵的长为单词 A 的长度，宽为单词 B 的长度，然后将第一行和第一列分别用自然数序列进行初始化。
 2. 从左至右，从上到下依次计算动态规划矩阵中的值：分别计算以当前字母为结尾的两个英文单词为了转成另一个单词需要的增加、删除和修改字母操作的最小次数，并取其中最小值为当前值。
 3. 最后返回动态规划矩阵中右下角的数值，该值即为所求的最小编辑距离。
-

2.4.3.5 特征选择

为了提升模型的准确率，除了利用词向量作为特征之外，我们加入了一些简单的特征，包括词性特征、句法分析特征、情感倾向特征和命名实体识别特征。

(1) 词性特征：评价对象一般是由名词或名词短语构成，而形容词或副词一般不可能是评价对象。因此，我们引入词性特征。首先，我们使用斯坦福大学的词性标注工具对句子中的单词进行词性标注，然后将名词、形容词、动词和副词作为特征，进行向量化表示，即将特征表示为一个四维向量，为 1 的维度即当前词对应的词性。

(2) 句法分析特征：句法特征即当前单词在评论文本中的句法成分，句法结构信息对于模型有效性的提升起着重要作用。我们首先利用 Stanford Parser 进行句法分析，然后将名词短语、动词短语、介词短语、形容词短语、副词短语五种句法成分作为特征，进行向量化表示，即将特征表示为一个五

维向量，为 1 的维度即当前词对应的句法成分。

(3) 情感倾向性特征：修饰评价对象的单词一般位于评价对象附近且具有情感倾向性，但评价对象单词是不具有情感倾向的，因此，我们引入单词的情感倾向性特征。

(4) 命名实体识别特征：命名实体即人名、地名、机构名及其他以名称作为标识的实体。命名实体极有可能成为评论语句中的评价对象。因此，我们提取评论语句中的命名实体，把当前单词是否位于命名实体表中作为模型的特征。

2.4.3.6 模型训练

在传统神经网络模型中，一般利用反向传播算法（Backpropagation, BP）从后向前依次计算各个参数的导数。而循环神经网络模型考虑到时间序列问题，对反向传播算法进行了改进，提出时间的反向传播算法（Backpropagation Through Time, BPTT）。算法 2-2 给出了 BPTT 算法的详细过程。

算法 2-2: BPTT (Backpropagation Through Time) 算法

输入：模型前向传播过程中每层节点的输出值。

输出：模型对各权值的梯度。

算法过程：

1. 从最后一个时间步骤从后往前进行计算：
 - 1) 计算当前输出层误差。
 - 2) 计算当前隐含层关于当前输出层的误差值。
 - 3) 计算并累加前一时间隐含层关于当前隐含层的误差。
 2. 对于每个权值：
 - 1) 计算每个时间节点关于该权值的梯度。
 - 2) 将每个时间的梯度累加，作为该权值的梯度。
-

BPTT 算法本质上就是在传统神经网络反向传播算法的基础上加入时间顺序，从最后时间步骤反向求解各个节点的误差，最后将相同节点的误差依次相加。由于时序的存在，损失函数对于隐含层的影响不仅仅来自其输出层，还来自于下个时刻的隐含层。

2.4.3.7 实验设置

模型的参数设置对于实验结果有着重要作用。首先，我们对模型各个参

数进行相应初始化，对于词向量的维度，我们分别使用 50 维、100 维和 300 维的向量进行实验，探索词向量的不同维度对模型准确率的影响。对于模型的系数我们设置为 $[-0.1, 0.1]$ 上的均匀分布，我们把模型的学习率设在 $(10^{-5}, 10^{-3})$ 之间，并在训练过程中做适当衰减，对于隐含层节点数，我们将其分别设置为 50, 100, 150 和 200 进行实验，根据实验结果选择最合适的节点数。其次，对于神经网络模型存在的过拟合问题，我们采取下面两种方法：首先利用交叉验证集，交叉验证集的设置能够较为有效地防止这类问题，因此我们从训练集中随机地抽取 10% 的样本作为验证集，剩下的 90% 的样本则当做训练集。由于我们的交叉验证集与训练集、测试集是相互独立且同分布的，这就保证了模型在交叉验证集和测试集有相似的趋势和表现，我们通过每次测试模型在交叉验证集上的结果来决定是否提前停止训练。其次，我们在模型中使用正则化方法，根据奥卡姆剃刀原理可以得出：简单的模型更有效，复杂的模型往往不会拥有很好的泛化能力，因此我们加入正则项来控制模型的参数，降低模型复杂度。我们使用的是 L2 正则化项，下式给出其计算方法：

$$J_{L2} = J + \frac{\lambda}{2n} \sum_{\theta} \theta^2 \quad (2-14)$$

式中 J_{L2} ——表示加入 L2 正则化项后的损失函数
 J ——表示原始损失函数
 λ ——表示正则项系数
 θ ——表示模型参数

理论上神经网络模型可以近似任意复杂的函数，因此，神经网络模型可以在多维空间内表达复杂的“曲面”，这就导致模型空间内有许多个局部的极小值点，而梯度下降法本身对于模型中存在的局部极值点和全局极值点无法进行辨别，因此，我们在权值迭代更新时加入冲量项，在每次进行权值更新时利用当前权值的梯度与上次权值的更新量的加权和作为本次权值的更新量，这可以帮助模型逃脱局部极值点，下式给出其计算方法：

$$\Delta \omega^n = m \Delta \omega^{n-1} - \alpha \frac{\partial J}{\partial \omega^2} \quad (2-15)$$

式中 ω^n ——表示第 n 次迭代时的 ω
 m ——表示冲量项系数
 $\Delta \omega^n$ ——表示第 n 次迭代时 ω 的更新量
 α ——表示学习率

2.5 实验设置与结果分析

本节首先给出了实验的评价指标及基线方法，然后针对基于条件随机场的方法和基于输出依赖双向 LSTM 的方法分别设置优化对比实验，探索使模型效果达到最佳的改进策略，最后给出基线方法和我们提出的基于条件随机场的方法、基于 BLSTM-CRF 的方法及基于输出依赖双向 LSTM 的方法的实验结果，同时对实验结果做对比分析。

2.5.1 评价指标

在评价对象任务中，我们利用标准的 BIO 标注方案对评价对象进行标注。表 2-4 给出了对餐馆评论的某条评论文本的标注实例。

表 2-4 餐馆评论文本标注实例

The	food	is	delicious	and	the	waiter	is	polite	.
B	I	O	O	O	B	I	O	O	O

基于 BIO 标注方案，我们可以通过比较评价对象抽取模型自动标注的标签与标准标签之间的差距来判断模型的性能。因此这里我们利用常见的精确率（Precision）、召回率（Recall）和 F1 值(F1-Measure)来作为评价指标。

精确率：正确抽取出的评价对象在所有抽取出的评价对象中占的比例。计算方式如下：

$$P = \frac{|S \cap G|}{|S|} \quad (2-16)$$

召回率：正确抽取出的评价对象在所有标准评价对象中占的比例。计算方法如下：

$$P = \frac{|S \cap G|}{|G|} \quad (2-17)$$

这里，S 代表模型标注的评价对象集合，G 代表标准的评价对象集合。精确率与召回率在某种程度上是相互制约的关系，两者通常不能两全其美，所以需要引入另一个综合度量来寻求准确率与召回率之间的平衡点，即 F1 值。

F1 值：综合了准确率和召回率的结果。计算方法如下：

$$F_1 = \frac{2 \cdot P \cdot R}{P + R} \quad (2-18)$$

2.5.2 基线方法

我们使用在 SemEval-2014 中获得最高准确率的队伍 DLIREC^[46]的方法当作基线方法。队伍 DLIREC 的系统利用无监督的评价对象抽取方法抽取出候选评价对象组成一个候选评价对象集合，然后把单词是否存在于这个集合中当成特征。

2.5.3 实验结果及分析

本节首先对基于条件随机场的相关探索实验进行介绍和分析，然后介绍基于深度学习的相关探索实验，探索使模型效果达到最佳的改进策略，最后给出总体对比实验结果并进行分析。

2.5.3.1 基于条件随机场的相关探索实验

基于条件随机场的方法中，我们选择的不同的特征对于模型实验结果的影响是不尽相同的，有的特征会较大幅度提升实验效果而有的特征对实验结果的影响则不太明显。因此，为了研究每个特征的影响，我们实验时依次加入一个特征对模型进行训练并计算相应的准确率、召回率和 F1 值。实验结果如表 2-5 所示。

表 2-5 特征分析的实验结果

特征	准确率	召回率	F1 值
Word	53.81%	45.56%	49.34%
+Small	54.65%	50.68%	52.59%
+Stem	53.54%	52.46%	52.99%
+POS	74.15%	71.17%	72.63%
+NPN	74.58%	77.36%	75.94%
+MAD	78.49%	77.25%	77.87%
+DEP	84.07%	77.89%	80.86%
+SO	84.31%	78.02%	81.04%
+EMB	85.11%	83.52%	84.31%

其中，Word 表示单词特征，Small 表示小写单词特征，Stem 表示词根特征，POS 表示词性特征，NPN 表示单词所在名词短语数目特征，MAD 表示

最近形容词距离，DEP 表示依存关系特征，SO 表示情感倾向性特征，EMB 表示词向量特征。由以上实验结果能够得出，词性特征的效果最为显著，引入词性特征后，模型的准确率、召回率和 F1 值分别提高了 20.61%，18.71%，19.64%。词向量特征对模型效果的提升也很明显，加入词向量特征后，模型的 F1 值提升了 3.27%，达到了模型的最佳结果 84.31%。

不同的特征组合方式也会对模型的效果产生一定的影响，我们通过对不同的特征进行组合来进行实验，分析特征组合的影响。实验结果如表 2-6 所示。

表 2-6 特征组合实验结果

特征	准确率	召回率	F1 值
词特征+句法特征	84.07%	77.89%	80.86%
词特征+情感倾向性特征	82.18%	72.58%	77.07%
句法特征+情感倾向性特征	58.65%	68.43%	63.16%
词特征+句法+情感倾向	84.31%	78.02%	81.04%
词+句法+情感倾向+词向量	85.11%	83.52%	84.31%

由以上实验结果能够得出，词特征和句法特征进行组合，模型的结果最佳。而词法特征和情感倾向性特征组合，模型的效果大幅度下降。这在一定程度上说明了词层面特征的重要程度，其原因在于某一特定评论领域下，常见的评价对象会经常被提及，仅单词层面的特征就能很好地抽取在训练数据中的频繁出现的评价对象。此外，对比四五个实验可以发现，加入词向量特征后模型的效果有明显提升，由此可以得出词向量特征对模型效果的提升有重要作用。

2.5.3.2 基于深度学习的相关探索实验

我们首先利用基于 2.4.1 节所介绍的两种基本的循环神经网络模型 RNN，LSTM 及 BLSTM 模型在数据集上进行了对比实验，探索三种神经网络模型在评价对象抽取任务的表现。实验结果如表 2-7 所示。

表 2-7 三种神经网络模型的对比实验结果

模型	准确率	召回率	F1 值
RNN	81.25%	75.33%	78.18%
LSTM	83.96%	76.74%	80.19%
BLSTM	84.45%	79.53%	81.92%

由表 2-7 可以看出，BLSTM 在评价对象抽取任务中的准确率、召回率和 F1 值比 RNN 和 LSTM 更高，原因可能是 BLSTM 能通过正向和反向 LSTM 结构充分利用上下文的依赖关系。因此，我们在提出的模型中使用 BLSTM 作为隐含层结构。

考虑到模型的输出标签之间是有着相互依赖关系的，而传统的网络模型无法充分利用这种依赖关系。因此，我们尝试在传统模型的基础上加入输出层到输出层的自连接。为了研究输出层之间的自连接的作用，我们在 RNN，LSTM 和 BLSTM 三个基础模型的输出层上之间加入自连接，分别记作 RNN_SC，LSTM_SC 和 BLSTM_SC。表 2-8 列出了三个基础模型输出层加入与不加入自连接的对比实验结果。

表 2-8 针对输出层的对比实验结果

模型	准确率	召回率	F1 值
RNN	81.25%	75.33%	78.18%
LSTM	83.96%	76.74%	80.19%
BLSTM	84.45%	79.53%	81.92%
RNN_SC	83.31%	77.54%	80.32%
LSTM_SC	83.28%	83.13%	83.20%
BLSTM_SC	85.85%	83.28%	84.55%

对比加入了输出层的自连接的模型与未加入自连接的模型的实验结果，我们可以看出加入输出层的自连接后模型的效果均得到了提升。因此，我们在提出的模型中的输出层之间加入自连接来提升模型准确率。

此外，我们利用基于 RNN 和 LSTM 的一系列循环神经网络模型及我们提出的输出依赖双向 LSTM 模型进行了评价对象抽取实验，探索词向量维度，隐含层节点数，额外特征、不同循环神经网络模型对实验效果的影响，同时列出 CRF 模型进行实验的结果，并作详细的对比分析。不同模型的实验结果如表 2-9 所示。

表 2-9 基于循环神经网络模型的评价对象抽取对比实验结果

模型	维度	节点数	准确率	召回率	F1 值
CRF	-	-	85.11%	83.52%	84.31%
RNN	100	50	81.25%	75.33%	78.18%
Bi-RNN	50	100	82.62%	75.91%	79.12%
Bi-RNN+Feat.	100	100	84.13%	76.83%	80.31%
LSTM	100	200	83.96%	76.74%	80.19%
Bi-LSTM	100	100	84.45%	79.53%	81.92%
Bi-LSTM+Feat.	100	100	84.96%	82.05%	83.48%
BLSTM-CRF+Feat	100	50	85.52%	84.21%	84.86%
OD-Bi-LSTM	100	50	85.85%	83.28%	84.55%
OD-Bi-LSTM+Feat.	100	150	85.64%	84.49%	85.06%

为了使对比效果更加明显，我们仅选取了每个模型的效果最好的词向量维度和隐含层节点数的结果进行展示。表 2-9 中，“Bi-”表示对模型进行双向化处理，“OD-Bi-LSTM”即表示我们提出的输出依赖双向 LSTM 模型，我们从下面五个层次对实验结果进行对比分析。

(1) 不同模型对实验结果的影响。

由实验结果可以看出，我们提出的输出依赖双向 LSTM 模型取得了最佳效果，F1 值达到了 85.06%，其次是 BLSTM-CRF 模型、Bi-LSTM 模型、LSTM 模型、Bi-RNN 模型和 RNN 模型。输出依赖双向 LSTM 模型比 Bi-LSTM 模型效果好的原因可能是模型将上一时间步骤上输出层输出的概率值输入下一时间步骤的输出层，利用了输出标签之间的依赖关系。LSTM 模型比 RNN 模型效果好的原因是 LSTM 模型隐含层的门结构能够使网络选择性地遗忘不重要的历史信息并保留新信息中的有用信息，使重要的历史信息沿网络传播，有效地学习长距离依赖关系。

(2) 对不同模型进行双向化处理对实验结果的影响

由实验结果可以看出，Bi-RNN 模型的 F1 值比 RNN 模型的 F1 值高 0.94%，Bi-LSTM 模型的 F1 值比 LSTM 的 F1 值高 1.73%，对模型进行双向化处理后的实验结果比原模型效果有所提升，其原因可能是对模型进行双向化后使两个 RNN/LSTM 结构分别作用于序列的正向和反向，能够有效利用上下文的信息。因此，我们在提出的模型中使用双向 LSTM 作为隐含层。

(3) 特征对实验结果的影响

对比实验结果能够得出,加入特征的 Bi-RNN 比不加特征的 Bi-RNN 模型的 F1 值高 1.19%,加入特征的 Bi-LSTM 比不加特征的 Bi-LSTM 模型的 F1 值高 1.56%,加入特征的输出依赖双向 LSTM 比不加特征的输出依赖双向 LSTM 模型的 F1 值高 0.51%,可见,我们提取的特征明显提高了模型的效果。原因在于词性特征可以帮助模型判断当前单词的词性是否是名词,如果是名词则当前单词是评价对象的概率就更大一些,因为评价对象一般是由名词。而句法分析特征可以帮助模型判断当前单词在评论语句中的句法成分,如果当前单词是名词短语则其判定为评价对象的概率就更大。情感倾向性特征则起到很好的指示作用,因为评价对象一般不具有情感倾向且与具有情感倾向性的单词距离较近。因此,我们在提出的模型中加入相应的特征来提升效果。

(4) 探索不同维度的词向量对模型效果的影响

实验中对于每个模型都分别利用 50 维、100 维和 300 维的词向量进行了研究,探索最佳的词向量维度。不同维度的词向量的对比实验结果没有明显规律,不是维度越大实验结果就越好,100 维的词向量几乎在各个模型中都取得了最佳效果。因此,我们在提出的模型中使用 100 维的词向量。

(5) CRF 与输出依赖双向 LSTM 的实验结果对比

由实验结果可以看出,输出依赖双向 LSTM 模型在不加特征的情况下的实验结果就比需要人工提取众多特征的 CRF 模型的效果更好。其原因可能是我们提出的输出依赖双向 LSTM 能够自主学习词向量之间的语义特征,而且能够充分利用上下文信息和输出标签之间的依赖关系。

2.5.3.3 总体对比实验结果

表 2-10 给出了基线方法、基于特征的 CRF 方法、基于 BLSTM-CRF 方法和基于输出依赖的双向 LSTM 方法在评价对象抽取任务中的实验结果并对实验结果进行了分析。

表 2-10 总体对比实验结果

方法	精确率	召回率	F1 值
基线方法	85.35%	82.72%	84.01%
条件随机场	85.11%	83.52%	84.31%
BLSTM-CRF	85.52%	84.21%	84.86%
输出依赖双向 LSTM	85.64%	84.49%	85.06%

由表 2-10 得出,我们提出的输出依赖双向 LSTM 方法在准确率、召回率

和 F1 值上都明显超过基线方法，而且比基于特征的条件随机场模型和 BLSTM-CRF 的效果更好。其原因可能是基于输出依赖双向 LSTM 的方法能够在整个句子范围能学习上下文信息，而基于条件随机场的方法只能在固定的窗口内进行学习。其次，基于输出依赖双向 LSTM 的方法能够有效地利用输出标签之间存在的依赖关系，并且可以捕捉到各个词向量间存在的语义关系，使得模型能够对具有多个词评价对象和未出现在训练集中的评价对象有更好的识别能力。此外，词性及句法分析等特征使得输出依赖双向 LSTM 能获取当前单词的词性和句法成分，结合词向量彼此存在的语义关系，提升模型进行评价对象抽取的效果。

我们还对各个方法在单词和多词评价对象上的抽取效果进行了研究，实验结果如表 2-11 所示。

表 2-11 单词与多词评价对象抽取实验结果

方法	单词评价对象			多词评价对象		
	准确率	召回率	F1 值	准确率	召回率	F1 值
基线方法	91.37%	86.77%	89.02%	74.16%	64.48%	68.98%
条件随机场	89.49%	84.26%	86.80%	80.06%	73.97%	76.89%
BLSTM-CRF	91.21%	86.72%	88.91%	81.79%	75.43%	78.48%
输出依赖双向 LSTM	91.61%	87.42%	89.47%	82.73%	75.79%	79.11%

由表 2-11 可以看出，相比基线方法而言，本文提出基于特征的条件随机场方法在由多词组成的评价对象抽取问题上准确率、召回率和 F1 值方面都取得比基线高的效果，分别比基线高 5.9%，9.49%，7.91%，但在单词评价对象的抽取问题上效果比基线方法略差些。相比而言，BLSTM-CRF 模型在单词和多词评价对象上的效果都比基于条件随机场的方法好，而本文提出的输出依赖双向 LSTM 在单词评价对象和多词评价对象的抽取任务中的实验结果都明显比基线方法和条件随机场方法及 BLSTM-CRF 模型要高，更加证明了我们提出的基于输出依赖双向 LSTM 方法的有效性。

为了验证我们提出的方法在餐馆评论领域的普适性，我们利用基线方法和本文提出的方法在 Yelp 数据集上进行了实验，表 2-12 列出了实验结果。

表 2-12 各方法在 Yelp 数据集上的实验结果

方法	精确率	召回率	F1 值
基线方法	84.25%	82.32%	83.27%
条件随机场	84.71%	82.72%	83.70%
BLSTM-CRF	85.76%	84.93%	85.34%
输出依赖双向 LSTM	86.83%	85.65%	86.24%

由表 2-12 可以看出,我们提出的方法在 Yelp 评论数据集上的效果仍然比基线方法好,尤其是输出依赖双向 LSTM 方法,准确率、召回率和 F1 值都取得了最好的结果。

2.6 本章小结

本章详细介绍了进行评价对象抽取的方法即基于条件随机场的方法、基于 BLSTM-CRF 的方法和基于输出依赖双向 LSTM 的方法。我们首先详细解释了评价对象抽取任务的概念,然后交代了数据集规模及特点。接着介绍了利用 CRF 来进行评价对象抽取的一系列工作,我们通过提取不同的特征并进行不同特征之间的组合来提高模型的效果,提取的特征大体可分为四个层面即词层面特征、句法层面特征、情感倾向性特征和词向量特征。然后介绍了基于深度学习的评价对象抽取,在简单介绍两种常见循环神经网络模型后介绍了基于 BLSTM-CRF 模型的评价对象抽取方法,然后提出了改进的输出依赖双向 LSTM,并对输出依赖双向 LSTM 的模型结构及隐含层结构进行了详细介绍,然后对模型词向量训练、未登录词处理、特征的选择、训练方法及参数设置等进行了介绍。最后,在介绍基线方法的基础上,给出针对 CRF 模型设置的对比实验和针对输出依赖双向 LSTM 模型设置的优化探索实验,探索使模型效果达到最好的方法,然后给出基线方法和我们提出的方法的实验结果并进行对比分析,得出我们提出的输出依赖双向 LSTM 模型在评价对象抽取任务中的成绩比基线方法、基于条件随机场的方法和基于 BLSTM-CRF 的方法更好。

第 3 章 评价对象极性判别方法研究

3.1 引言

互联网技术的飞速发展使得消费者接受并习惯于网上购物、网上订餐等方便快捷的消费方式。对于消费者而言，电子商务平台上其他消费者对商品各个方面的评论信息能够帮助其在购买相应商品前更加充分地了解该商品。对于商家而言，评论信息则可以使其了解消费者对商品的反馈，从而及时对商品的各个方面进行改进。但是一方面商品的相关评论信息数量庞大，而另一方面这些评论的质量又良莠不齐，这就导致从商品评论信息中获取想要了解的信息既消耗时间又消耗精力。因此，研究如何准确快捷地判别评论信息中评价对象对应的评价极性是一项极其有价值的工作。

本章提出了基于机器学习和基于深度学习的两类评价对象极性判别方法，前者即我们提出的基于提升的模型融合方法，后者即基于 LSTM 的方法和基于双向 LSTM 的方法。

基于提升的模型融合方法的核心思想就是充分利用线性分类模型有限合理性选择的特点及非线性分类模型能够解决线性不可分样本的特点，在评价对象极性判别问题上充分发挥并结合二者的优势，使得模型的准确率得到提高。图 3-1 描绘了该方法的主要研究思路所示。

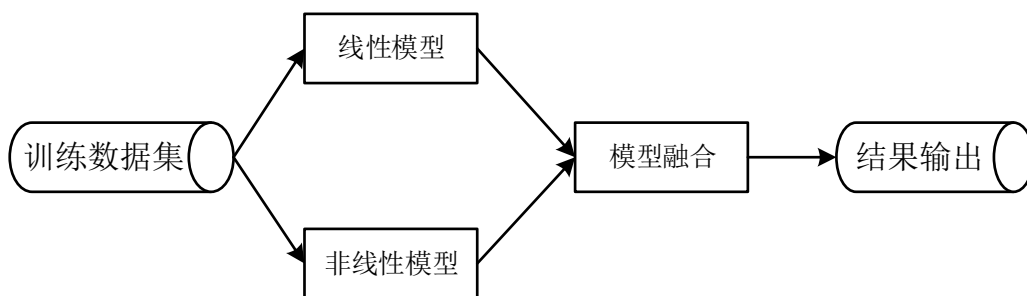


图 3-1 基于提升的模型融合方法

如图 3-1 所示，该方法将分别使用线性模型和非线性模型进行评价对象极性判别，然后在两种模型的基础上进行模型融合，进而得到最终结果。

基于双向 LSTM 的方法省去了复杂的特征提取过程，将评论文本中的词映射为词向量再将当前单词词向量和评价对象的词向量一同输入到双向

LSTM 中以便模型捕获上下文中的每个单词与评价对象的语义关系，然后将双向 LSTM 的输出结果向量输入 SoftMax 层进行分类，避免了特征抽取过程中的错误传播问题。本文提出两种模型结构，分别是基于 LSTM 的极性判别模型结构和基于双向 LSTM 的极性判别模型结构，并通过实验对比两种模型结构的优劣。

本章的其它内容安排如下。3.2 节主要介绍评价对象极性判别任务的问题定义以及相应的训练数据集情况。3.3 节首先介绍本文所使用的线性分类模型、非线性分类模型以及模型融合的具体方法，进而介绍在模型训练中所使用的特征及其相应的提取方法。3.4 节分别介绍了基于 LSTM 的模型结构与基于双向 LSTM 的模型结构及模型训练相关问题。3.5 节首先介绍评价对象极性判别的评价指标和基线方法，然后对基于 LSTM 的模型和基于双向 LSTM 的模型中的三个关键部分的改进策略进行对比实验，最后给出我们所提出的所有方法与基线方法在给定数据集上的实验结果并对实验结果进行分析。3.6 节是本章小结，简要地总结了全章内容。

3.2 问题定义与数据集

3.2.1 问题定义

在给定句子中所涉及的评价对象的条件下，评价对象极性的判别即判断评论者（情感持有者）对每个评价对象所持有的态度或情感。通常观点持有者的情感可以分为四类：积极的（positive）、消极的（negative）、中立的（neutral）、冲突的（conflict）。

积极的情感极性是用户对于餐馆的欣赏与赞美，体现了用户的积极的、正面的情绪；消极的情感极性指用户对餐馆某些方面的批评指责，反映了用户的消极、负面情绪；中立的极性是指用户仅仅在评论句提及某一评价对象，但并未对其表达任何态度或情感，因此将此类情感定义为中立的；冲突的极性是指用户表达了餐馆某个方面的不满，但同时也表达了对该方面的支持与肯定。如表 3-1 所示，第一句“The pizza was pretty good and huge.”这句话的评价对象是 pizza，“good and huge”表示评价的极性是积极的，第二句“The service is awful.”这句话评价对象是 service，awful 表示评价的极性是消极的。第三句，“Not a large place, but it's cute and cozy.”这句话表达的意思时餐馆地方不大但是却温馨舒适。首先表达了餐馆面积较小，然后转而说餐馆虽小却舒适。因此，这句话的对 place 的评价极性是冲突的。在第四局“I drank the

coffee with my friends.”中，仅仅提及“coffee”这一评价对象，但并未对其表达任何态度或情感，因词其评价极性是中立的。

表 3-1 餐馆评论语句示例

评价内容	评价对象	评价极性
The pizza was pretty good and huge.	pizza	positive
The service is awful.	service	negative
Not a large place, but it's cute and cozy.	place	conflict
I drank the coffee with my friends.	coffee	neutral

3.2.2 实验数据集

本章在评价对象极性判别任务上的数据集使用第八届国际语义评测大会（8th International Workshop on Semantic Evaluation, SemEval-2014）提供的评论信息集合¹，也即本文在评价对象抽取任务中所使用的数据集来源，在评价对象抽取任务的数据集合基础上，加入了对每个评价对象情感极性的标注。数据集详细情况如表 3-2 所示，训练集大约包括 3043 条标注好的评论语句，其中评价对象的数量大约为 3688 个，测试集大约含有 800 条评论语句，其中评价对象的数目大约 1157 个。训练集中极性为 positive 的评论文本大约有 1200 多条，极性为 negative 的评论文本大约有 1000 多条，极性为 neutral 的评论文本大约有 200 多条，极性为 conflict 的评论文本大约有 600 多条。这些评论语句都是以可扩展标记语言（XML）格式存储的

表 3-2 实验数据集

数据集	评论数量	单词数量	评论对象数量
训练集	3043	47157	3688
测试集	800	12734	1157

图 3-2 列举了本文使用的餐馆评论数据集中对评价对象及其情感极性的标注实例。其中，text 节点中列出了评论语句，opinionTargets 节点中列出了评论语句中的所有评价对象，opinionTarget 节点中的 term 属性对应评价对象，

¹ <http://alt.qcri.org/semeval2014/task4/>

polarity 属性对应评价对象的情感极性。

```
<sentence id="36244784#0#3">
  <text>Entrees include classics, where lasagna is delicious but the fettuccine Alfredo is awful.</text>
  <opinionTargets>
    <opinionTarget term="Entrees" polarity="neutral" from="0" to="7"/>
    <opinionTarget term="lasagna" polarity="positive" from="30" to="37"/>
    <opinionTarget term="fettuccine Alfredo" polarity="negative" from="70" to="88"/>
  </opinionTargets>
</sentence>
```

图 3-2 餐馆数据集中对评价对象及其情感极性的标注实例

数据集中各个极性的统计情况如图 3-3 所示,评论数据集中消费者对大多数的评价对象呈满意态度,呈消极和中立态度的评价对象占总数的 38.94%,只有极小部分评价对象的极性是冲突的。

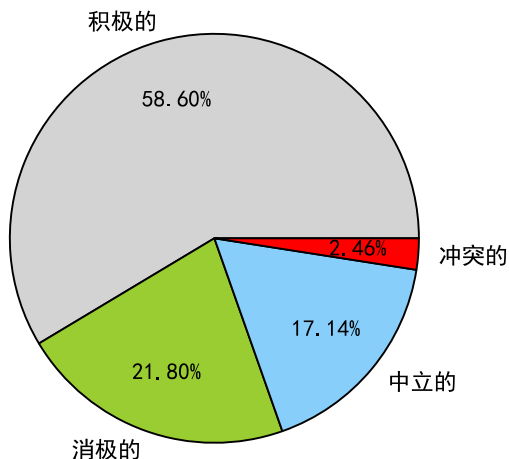


图 3-3 数据集中各个极性的占比图

3.3 基于模型融合的评价对象极性判别

本节将主要介绍基于提升的模型融合方法。首先介绍我们提出的融合方法中所使用的线性分类模型即支持向量机模型和非线性分类模型即随机森林模型。接着对模型融合的具体方法进行了详细介绍。最后详细介绍实验中所使用的特征。

3.3.1 支持向量机模型

支持向量机(Support Vector Machine, SVM)是由 Cortes 和 Vapnik 在 1995 年提出的一种分类方法,在 VC 维理论和结构风险最小定理基础上,求解模型复杂程度与其学习能力的最恰当的折中来得到最强的泛化效果。理论上来说,SVM 是属于二分类模型的,它的基础模型是特征空间中具有最大间隔的一个线性分类器。

问题描述:对于输入的样本点 $x=(x_1, x_2, \dots, x_n)^T$, 存在一个函数 $f(x)$ 可以将样本点分成两类,如果 $f(x) < 0$, 则样本 x 为负类,否则分为正类。

线性可分的支持向量机 $f(x) = \langle w \bullet x \rangle + b = \sum_{i=1}^n w_i x_i + b$ 一定有着无数个分类界面把数据正确隔开,SVM 则使用间隔最大化策略从无数个分类界面中求解出那个最优的分类超平面,即在 $y_i(w \bullet x_i + b) - 1 \geq 0, i=1, 2, \dots, N$ 约束条件下,求 $\min_{w,b} J_{svm(w,b)} = \min_{w,b} \frac{1}{2} \|w\|^2$, 使样本集的几何间隔最大。我们可以利用线性可分支持向量机的对偶算法来求解。首先将线性可分支持向量机的最优化问题作为原始最优化问题,然后利用拉格朗日对偶性,通过对对偶问题进行求解来得到原始最优化问题的最优解。

如图 3-4 所示,图中“+”表示属于正类的样例,“-”表示属于负类的样例,根据间隔最大化策略找出能正确划分两类样本同时间隔最大的直线,即最优分类超平面。

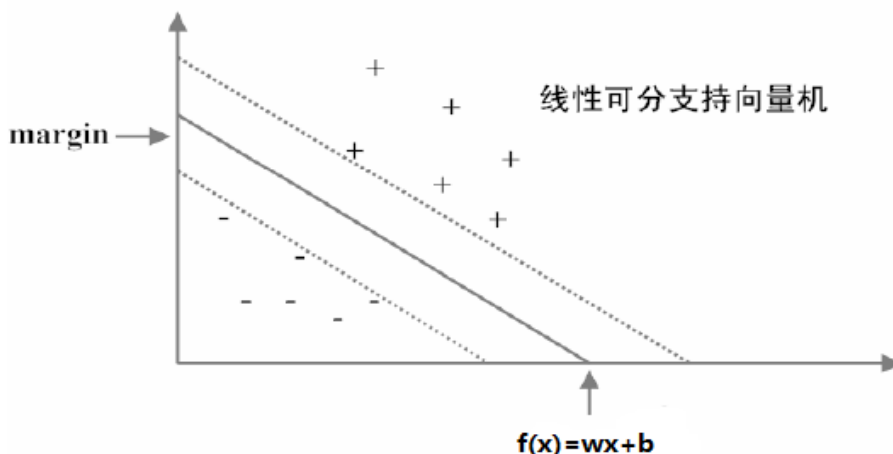


图 3-4 线性可分支持向量机的分类面

对于线性不可分支持向量机来说,所有样本在低维空间中是属于线性不可分的,此类支持向量机利用函数 ϕ 将向量映射到高维空间中,这样使线性不可分样本在高维空间中变得线性可分,即:

$$f(x) = \sum_{i=1}^n w_i \phi_i(x) + b \quad (3-1)$$

线性不可分支持向量机的分类面如图 3-5 所示：

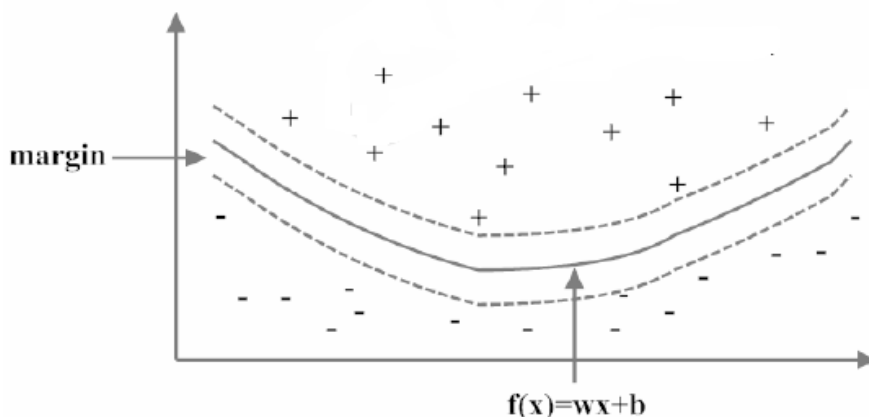


图 3-5 线性不可分支持向量机的分类面

选择合适的核函数 $K(x_i, y_i)$ ，通过利用核函数来求解，这样支持向量机就表示为

$$f(x) = \text{sign}(a_i^* y_i \phi(x_i) \bullet \phi(x) + b^*) = \text{sgn}(\sum_{i=1}^n a_i^* y_i K(x_i, x) + b^*) \quad (3-2)$$

这样，经过映射函数 ϕ 把原来的低维空间变换到能使样本线性可分的高维空间中，将低维空间中的内积 $x_i \bullet y_i$ 变换为高维中的内积 $\phi(x_i) \bullet \phi(x_j)$ ，在新的空间中学习线性支持向量机。在核函数已确定的情况下，求解非线性分类问题的支持向量机可以利用求解线性分类问题的方法。

3.3.2 随机森林模型

随机森林模型是由 Leo Breiman 和 Adele Cutler 首次提出的，对其通俗的理解就是通过多棵决策树训练样本然后进行预测的一种非线性分类模型。对于随机森林模型的每棵决策树而言，它对应的训练样本均是由总体的训练样本经过有放回的采样而最终获得的，也就是说，总体训练集合里面的每个样本都有可能数次被选到在同一决策树中，也有可能从未被选到在一棵决策树中。特征方面，随机森林模型中的每棵决策树建立过程中都是根据特定的比例从特征集合中采取无放回采样的方式进行抽取，也即对于树中任一节点，建立决策树时使用的特征集合为总体特征的一个子集。算法 3-1 描述了随机森林模型具体的训练过程。

算法 3-1 随机森林模型的训练

输入：

特征集合 F

包含决策树个数

决策树建立时使用的特征数

训练样本集

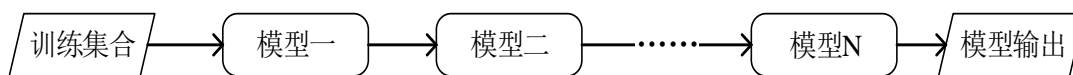
输出：随机森林

算法：

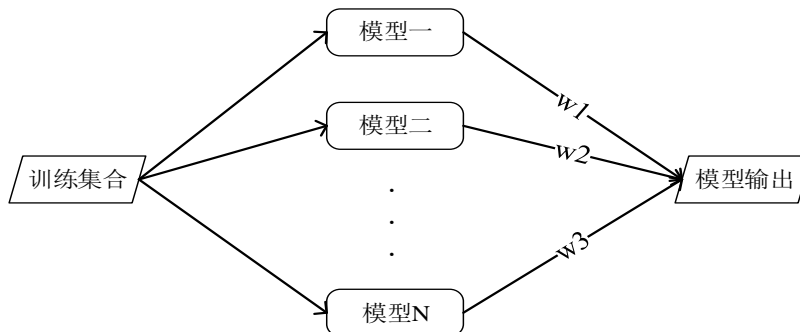
1. 从训练样本集 中以有放回的方式抽取与 相同大小的样本集合 。
2. 根据样本集合 建立决策树，建树过程中每一个节点都需要无放回地从特征集合 中抽取 个特征作为当前节点的可选特征集合。
3. 重复第 1、2 步，直到建立 棵决策树。

3.3.3 模型融合方法

在机器学习领域中常见的模型融合方法包括级联融合方法和加权融合方法。模型的级联融合就是将每个分类模型串联在一起，每个模型的输入是上个模型的输出，最末尾的分类模型所输出的结果即为模型融合的最终结果。模型的加权融合就是分别单独地训练多个分类模型，然后按照每个模型在验证集合上的表现效果对所有模型加权得到最终结果。图 3-6 描述了模型级联融合方法和模型加权融合方法的示意图。



(1) 模型级联融合方法示意图



(2) 模型加权融合方法的示意图

图 3-6 模型级联融合方法和模型加权融合方法的示意图

由图 3-6 可以看出,模型级联融合方法更适用于所使用分类模型数量较少,模型间区分程度更大的情况。相反地,模型加权融合方法则更适用于所使用分类模型数量较多且模型间区分程度较小的情况。

本文在这两种模型融合方法的基础上,使用一种提升方法(Boosting)将本文使用的线性模型(SVM 模型)和非线性模型(随机森林模型)进行融合。该方法的主要思想就是依次训练各个分类模型,在每次训练得到的结果基础上改变训练数据集合的概率分布,即增大所有被上个分类模型错误分类的数据所占的权重同时减小被上个分类模型正确分类的数据所占的权重。当所有模型训练完毕后,再按照各模型效果对结果加权得到最终的模型结果,即加大分类误差较小的模型在最终结果中所拥有的表权重,减小分类误差率较大的模型在最终结果中所拥有的的表权重。算法 3-2 描述了基于提升方法的模型融合的具体算法。

算法 3-2 基于提升方法的模型融合方法

输入:

训练样本集 S

分类器集合 M

输出: 最终分类模型 $G(x)$

算法:

1. 初始化训练集合的权值分布:

$$D_1 = (w_{11}, w_{12}, \dots, w_{1N}), w_{1i} = \frac{1}{N}, i = 1, 2, \dots, N$$

2. 对于分类器集合 M 中的每一个分类模型:

- 1) 使用具有权值分布 D_m 的训练样本训练模型,从而得到基本分类器

$$G_m(X)。$$

- 2) 计算基本分类器 $G_m(X)$ 在训练数据集合的分类误差率:

$$e_m = \sum_{i=1}^N w_{mi} I(G_m(x_i) \neq y_i)$$

- 3) 计算当前分类器的权重也即当前分类器在最终分类器中的系数:

$$\alpha_m = \frac{1}{2} \log \frac{1-e_m}{e_m}$$

- 4) 更新训练集样本中的权值分布:
-

$$w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp(-\alpha_m y_i G_m(x_i))$$

3. 将各个基本分类器组融合成最终分类器:

$$G(x) = \sum_{m=1}^M \alpha_m G_m(x)$$

3.3.4 特征提取

本小节主要介绍我们提取的特征及其相应的提取方法。

3.3.4.1 情感极性词典特征

情感极性词典中的每个单词均被赋予一定的情感倾向性,也就是说,使用情感倾向性词典可以很容易地判断出某一单词的具体情感倾向。一条评论语句中的评价对象的情感倾向性通常由其修饰词决定,而评价对象的修饰词往往与其修饰的对象在句子中的距离较近。由此,本文使用评价对象附近五个单词的情感倾向性评分作为特征。

本文利用国外的 Yelp 网站上提供的数据集创建餐馆领域的情感词典。数据集被分为 500 多个种类,我们只利用其中餐饮相关的 58 个类别,从中提取出消费者评论信息。每个评论信息都有对应的评分,评分范围在 1 到 5 分之间,本文将 1 到 2 分视为消极评论,4 到 5 分视为积极评论,3 分为中立评论。

具体地,本文计算评论集合中的词 w 的情感倾向性评分的方式如下:

$$score(w) = PMI(w, pos) - PMI(w, neg) \quad (3-3)$$

式中 pos ——代表积极评论;

neg ——代表消极评论。

PMI 的计算公式如下:

$$PMI(w, pos) = \log_2 \frac{freq(w, pos) * N}{freq(w) * freq(pos)} \quad (3-4)$$

式中 $freq(w, pos)$ ——代表单词 w 在积极评论中的出现次数;

$freq(w)$ ——代表单词 w 在整个评论集合中的出现次数;

$freq(pos)$ ——代表所有积极评论中的词数;

N ——代表整个评论集合中的词数。

$PMI(w, neg)$ 也是以同样的方式计算。除此之外,本文同时使用以下三种情感倾向词典来计算每个单词的情感倾向评分:

MPQA 语义倾向性词典。Wilson 等人为词典中的每个单词标注出其词性、情感强度（“strong”或“weak”）和语义倾向性。在计算评分时，使用-1表示强性的消极情感；-0.5表示弱性的消极情感；1表示强性的积极情感；0.5表示弱性的积极情感。

SentiWordNet 语义倾向性词典。Baccianella 等人根据 Wordnet 语义知识库中同义词集合的相互关系总结出部分单词的情感分数。在计算情感倾向性评分时，如果某单词积极情感分数大于其消极情感分数则取其积极情感分数作为其情感倾向性评分，反之则取其消极情感分数作为其情感倾向性评分。

Bing Liu 的情感倾向性词典。Bing 和 Liu 根据大规模语料总结出一系列具有情感倾向的单词。在计算评分时，如果该单词为积极的，则取其情感倾向性评分为 1，否则为-1。

本文将四个词典进行合并，得到新的情感倾向性词典，实验中将根据该情感倾向性词典得到评论信息中评价词的情感倾向性评分。

3.3.4.2 否定词特征

评论语句中的否定词一般会使其修饰的形容词的情感倾向性发生转变，比如“not very good”即表示消极的情感而不是积极情感。如果单纯的判断单词的情感倾向性而忽略了否定词，则会造成很大的误差，因此，本文引入否定词特征来避免此类问题的发生。

本文使用常见的带有否定性的单词作为否定词列表，其中包含 9 个否定词：no, nor, not, neither, none, nobody, nothing, hardly, seldom。具体地，如果某个单词出现在上述否定词列表中某个否定词的后面，则在计算其情感倾向性评分时需要修改情感倾向（取相反数）。

3.3.4.3 句法分析特征

通过句法分析技术可以准确地判断某一评价对象在评论语句中所承担的句法成分以及语句中其他词语与当前评价对象的依存关系。本文使用斯坦福大学的句法分析器¹对评论语句进行分析，并根据分析结果提取出以下几种特征：

- （1）与当前评价对象在句法分析树中距离最近的形容词情感倾向性评分。
- （2）句法分析树中所有具有积极情感的词汇与当前评价对象在句法分析树中的情感倾向性平均（句法分析树中距离加权）评分。
- （3）句法分析树中所有具有消极情感的词汇与当前评价对象在句法分析

¹ <http://nlp.stanford.edu/software/lex-parser.html>

树中的情感倾向性平均（句法分析树中距离加权）评分。

（4）句法分析树中所有词汇与当前评价对象在句法分析树中的情感倾向性平均（句法分析树中距离加权）评分。

（5）与当前评价对象具有形容词修饰关系的形容词的情感倾向性评分。

3.3.4.4 其他统计信息特征

为了给分类模型提供完整的统计信息，本文还使用了几种简单的统计特征，以描述评论语句整体的信息。本文使用的特征如下：

（1）评论语句中所有评价对象的数量。

（2）评论语句中所有具有积极情感倾向的词数占总体词数的比例。

（3）评论语句中所有具有消极情感倾向的词数占总体词数的比例。

（4）评论语句中所有具有积极情感倾向的词与当前评价对象的距离加权平均分。

（5）评论语句中所有具有消极情感倾向的词与当前评价对象的距离加权平均分。

（6）评论语句中的所有词汇与当前评价对象的距离加权平均分。

3.4 基于深度学习的评价对象极性判别

本节首先介绍基于 LSTM 的评价对象极性判别模型和基于双向 LSTM 的评价对象极性判别模型，对模型的输入输出及整体架构进行说明，然后给出句子预处理方法及未登录词的向量表示方法，最后介绍了模型训练问题。

3.4.1 基于 LSTM 的评价对象极性判别

本文实现了一种基于 LSTM 的评价对象极性判别模型，其基本结构如图 3-7 所示。LSTM 在 2.4.1.2 小节已经进行了详细介绍，这里不再赘述。在该模型中，模型的输入是评论文本和评价对象，将评论语句和评价对象送入 Embedding 层得到评论语句对应的词向量 $\{w_1, w_2, w_3, \dots, w_n\}$ 和评价对象对应的词向量 $\{w_{l+1}, w_{l+2}, w_{l+3}, \dots, w_{r-1}\}$ ，我们将评价对象对应的词向量加和取平均得到代表评价对象的向量 v_{target} 。为了利用评价对象的上文和下文信息，我们采用两个 LSTM 网络，分别为 LSTM_L 和 LSTM_R ， LSTM_L 收集评价对象的上文语义信息， LSTM_R 收集评价对象的下文语义信息，我们从左到右运行 LSTM_L ，从右到左运行 LSTM_R ，将上一个 LSTM 隐含层的输出输入到下一个 LSTM 单元中，然后将 LSTM_L 和 LSTM_R 的最末尾的隐含层的输出向量进行连接操

作得到代表当前评价语句的向量，然后将该向量输入到 Softmax 层计算该评价对象属于每个评价极性的概率，从而得到概率值最高的对应的极性类别。每个时间步骤上的输入除了该时刻上单词的词向量之外还包括评价对象的词向量 v_{target} ，这样将评价对象词向量和当前词向量进行连接后送入 LSTM 中能充分利用评价对象与上下文中每个单词之间的语义关系。

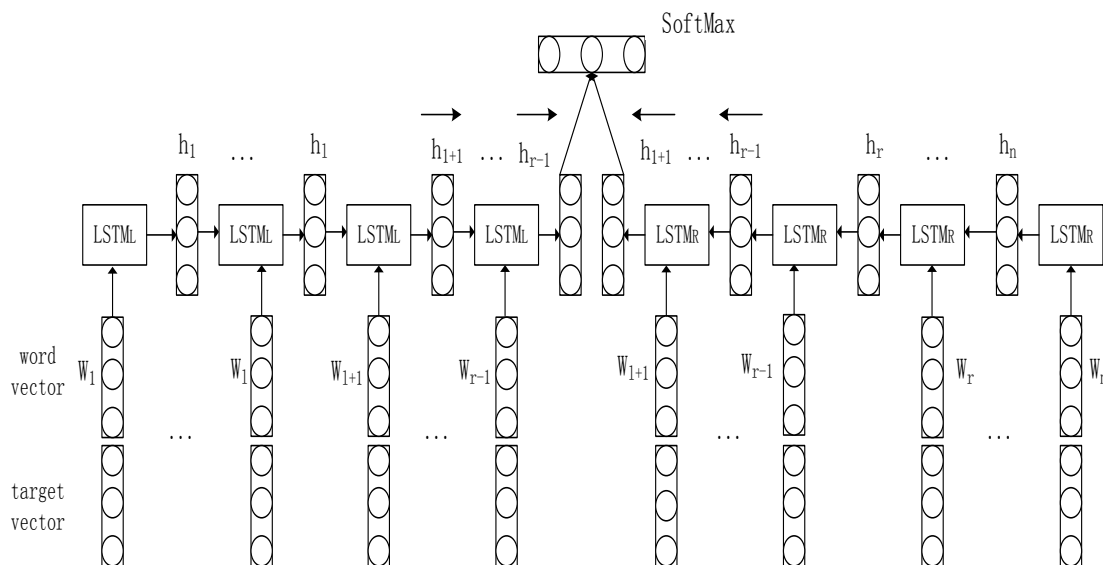


图 3-7 基于 LSTM 的评价对象极性判别网络结构

3.4.2 基于双向 LSTM 的评价对象极性判别

在基于 LSTM 的评价对象极性判别模型的基础上，本文提出了一种基于双向 LSTM(简称 BLSTM)的评价对象极性判别模型。双向 LSTM 通过在模型的隐含层加入一个正向 LSTM 结构和一个反向 LSTM 结构来分别获取评论文本的上文和下文的重要的特征信息，从而解决了 LSTM 模型无法利用下文信息的缺点。

这里我们给出基于双向 LSTM 的评价对象极性判别模型的基本结构，如图 3-8 所示，在该模型中，我们使用两个双向 LSTM 结构，分别为 $BLSTM_L$ 和 $BLSTM_R$ 来收集评价对象的上文和下文语义信息，然后分别将两个 BLSTM 的各个隐含层的输出送入均化池进行整合得到代表上文信息的向量和代表下文信息的向量，最后将两个代表上下文信息的结果向量连接后输入 Softmax 中。具体地，首先将评论语句中的词与评价对象输入到 Embedding 层得到词向量 $\{w_1, w_2, w_3, \dots, w_n\}$ 和评价对象对应的词向量 $\{w_{l+1}, w_{l+2}, w_{l+3}, \dots, w_{r-1}\}$ ，若评价对象包含多个词，则将其对应的多个词向量加和取平均得到代表当前对象的词向量 v_{target} ，然后在每个时间步骤上，将当前词的词向量与代表评价对象的

词向量进行连接并输入到双向 LSTM 中得到每个时间步骤上的隐含层输出，接着将 $BLSTM_L$ 和 $BLSTM_R$ 得到的每个时间步骤上的隐含层输出向量 $\{h_1, h_2, h_l, h_{l+1}, \dots, h_{r-1}\}$ 和 $\{h_{l+1}, \dots, h_{r-1}, h_r, h_n\}$ 分别送入 Mean Pooling 分别得到代表上文信息的向量 v_l 和代表下文信息的向量 v_r ，最后将两个向量连接后送入 Softmax 层，计算该评价对象属于各个评价极性的概率，获得概率值最高的所对应的评价极性。

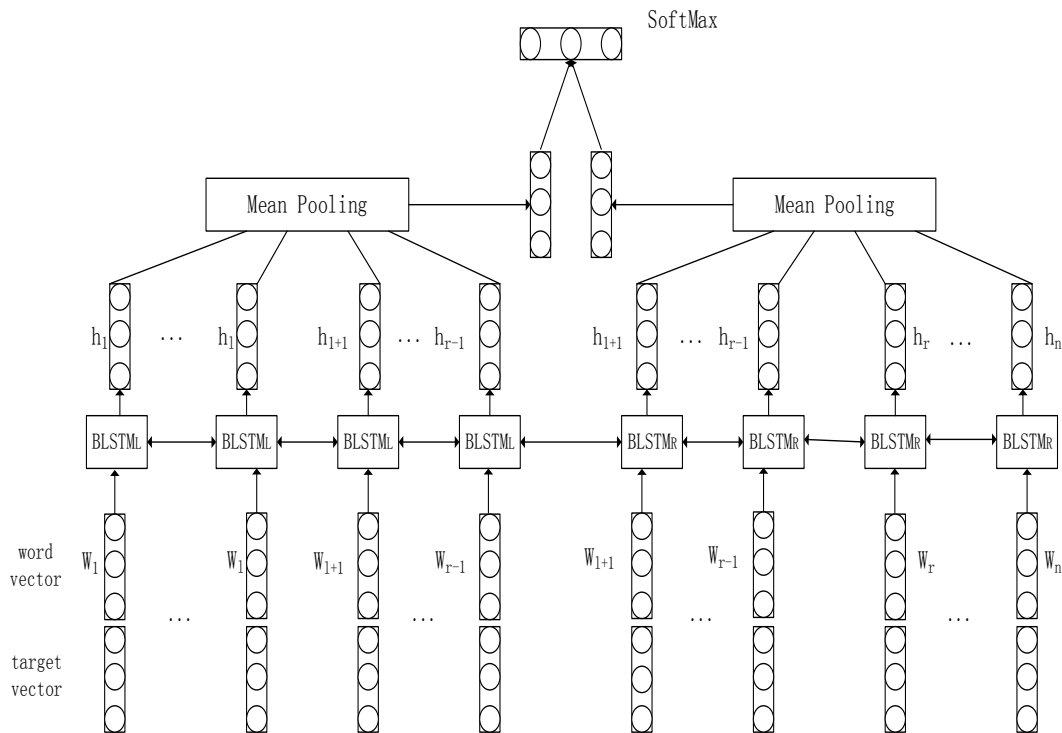


图 3-8 基于双向 LSTM 的评价对象极性判别网络结构

3.4.3 句子预处理及未登录词词向量表示

深度学习方法省去了人工提取特征的繁琐过程，直接将原始数据映射为词向量作为特征输入到模型中。在评价对象极性判别任务中，我们首先根据训练数据建立词表，然后将输入的评论语句中的词替换为其在词表中对应的 WordID。网络的 Embedding 层将这些 WordID 映射为其对应的词向量（这里我们使用在评价对象抽取任务中训练好的词向量），最后再将每个词对应的词向量送入到模型中。

在对句子中的词做映射之前，我们先对句子做简单的预处理，从而减少未登录词的数量。我们将句子中出现的特殊符号如“.”，“，”，“\”，“！”，“？”，“\$”，“（”，“）”，“+”，“#”，“&”等进行删除处理。

本文使用 2.4.3.4 节介绍的方法针对不同类型的未登录词采取不同的处理方法，最终获得所有未登录词各自对应的词向量，输入到网络中进行训练。

3.4.4 模型训练

我们使用 3.2.2 小节介绍的实验数据集进行实验，从训练集中随机地抽取 10% 的样本作为验证集，剩下的 90% 的样本则当做训练集，在实验中通过测试模型在交叉验证集上的结果来决定是否提前停止训练。

我们使用 2.4.3.6 小节介绍的 BPTT 算法进行模型训练，两个模型使用的目标函数都为交叉熵损失函数（cross-entropy loss），其公式表示如下：

$$J = - \sum_{s \in S} \sum_{c=1}^C P_c^s(s, t) \log(P_c(s, t)) \quad (3-5)$$

式中 S ——代表训练集数据；

C ——代表情感极性类别；

s ——代表一条评价语句；

t ——代表评价语句中的某个评价对象；

$P_c(s, t)$ ——评价语句 s 中的评价对象 t 的情感极性预测为类别 c 的概率；

$P_c^s(s, t)$ ——代表类别 c 是否为正确的情感极性类别，它的值为 0 或 1。

3.5 实验设置与结果分析

在本节中，我们首先介绍评价指标和基线方法。然后对基于 LSTM 的模型和基于双向 LSTM 的模型中的三个关键部分的优化改进策略进行对比实验，给出实验结果并进行分析。接着给出基线方法及本文提出的模型融合方法、基于 LSTM 的方法和基于双向 LSTM 的方法在训练集和测试集上的实验结果，并对实验结果进行分析。最后再纵向比较基线方法和我们提出的三种方法在四种情感极性上的效果，并分析实验结果。

3.5.1 评价指标

本文中，我们把评价对象极性判别看做分类问题来处理，因此我们利用常见的评测标准准确率（A，Accuracy）来评价模型的效果。准确率反应了模型正确预测的评价对象的情感极性占有所有评价对象的比例。具体地，其计算公式如下式。

$$A = \frac{Correct_Num}{Total_Num} \quad (3-6)$$

式中 $Correct_Num$ ——代表被模型正确预测出情感极性的评价对象数量；
 $Total_Num$ ——代表数据集中的所有评价对象的数量。

3.5.2 基线方法

本文选用在 SemEval-2014 评测中取得最好成绩 DCU^[46]队伍的方法作为基线方法。DCU 队伍分别使用了两种不同的方法进行评价对象极性判别。

第一种是基于规则的方法，他们使用语义倾向性词典、句子块结构距离、句法依存关系等相关规则，为每一个评价对象打出具体的分数，最后根据此分数划分评价对象的情感极性。

第二种是机器学习的相关方法，他们使用 SVM 作为基础分类模型，并使用了一种先进的 Bag-of-N-gram 特征做为模型的输入，该特征可以有效表示当前评价对象单词、词性、语义倾向等信息，表 3-3 以句子 “I charge it at night and skip taking the cord with me because of the good battery life.” 为例给出了 2640 个特征中的 7 个具体实例。

表 3-3 Bag-of-N-gram 特征具体实例

标识	N	前后词个数	N-gram	频数
L	2	2	cord with	1
AL	2	2	<aspect> with	56
ALS	1	4	<negu080>	595
ALSR	2	4	<negu080>	502
AL	2	4	and skip	1
ALSR	2	4	and <negu080>	25
ALSRP	1	4	<negu080>/vb	308

3.5.3 实验结果及分析

3.5.3.1 基于深度学习的不同维度词向量的探索实验

词向量是神经网络模型的非常重要的特征，模型最终生成的代表句子的

向量中所包含的语义信息来源于一个个输入到模型中的词向量。不同的词向量维度对实验结果的影响应该是不同的，所以我们选取维度为 50、100、300 的词向量在基于 LSTM 的评价对象极性判别模型和基于双向 LSTM 的评价对象极性判别模型上进行对比实验，实验结果如表 3-4 所示。

表 3-4 不同词向量维度对比实验结果

词向量 维度	LSTM 训练集	LSTM 测试集	BLSTM 训练集	BLSTM 测试集
50	80.9%	80.1%	81.8%	81.2%
100	81.7%	81.5%	83.4%	82.7%
300	79.9%	78.8%	80.5%	80.1%

由实验结果能够看出，词向量维度对评价对象抽取的确有一定的影响，但并不是维度越高实验效果越好，100 维的词向量在基于 LSTM 和基于 BLSTM 的评价对象极性判别实验中均取得了最好的效果。因此，我们选用 100 维的词向量进行模型训练。

3.5.3.2 LSTM 和双向 LSTM 方法中不同输入向量的探索实验

评价对象与上下文间的语义关系对评价对象极性判别有着至关重要的影响，为了获取评价对象与上下文间的语义关系，我们尝试将单词的词向量与当前评价对象的词向量进行连接后送入模型中。为此，我们按如下方式设置了对比实验：

- (1) 每个时间步骤上仅将当前单词的词向量输入到 LSTM 模型中。
- (2) 每个时间步骤上将当前单词的词向量与评价对象的词向量连接后输入到 LSTM 模型中。
- (3) 每个时间步骤上仅将当前单词的词向量输入到双向 LSTM 模型中。
- (4) 每个时间步骤上将当前单词的词向量与评价对象的词向量连接后输入到双向 LSTM 模型中。

表 3-5 不同输入向量的对比实验

输入向量	LSTM 训练集	LSTM 测试集	BLSTM 训练集	BLSTM 测试集
单词词向量	80.2%	79.7%	81.4%	80.3%
单词+评价对象词向量	81.7%	81.5%	83.4%	82.7%

对表 3-5 的对比实验结果进行分析得出，在 LSTM 模型和双向 LSTM 模型中，将当前单词的词向量与评价对象的词向量连接后一同送入模型中的效果都要好于仅输入该单词的词向量。其原因可能是将评价对象的词向量与每个单词连接后一同输入模型中能更有效地获取评价对象与上下文中每个单词之间的语义关系，使得模型能充分利用评价对象与上下文的关系来生成上文和下文的表示向量。因此，我们在模型中将单词词向量于评价对象词向量连接后再送入模型中进行计算。

3.5.3.3 LSTM 和双向 LSTM 中隐含层结果向量不同连接方式的探索实验

如何有效地利用模型生成的代表上文和下文的向量，充分地解析上下文向量中的语义信息，抽取上下文单词跟评价对象之间的语义关系对评价对象极性判别模型来说十分重要。我们在实验中尝试了四种向量整合的方式分别是向量相连、向量相加、向量相减、向量相乘。向量相连是向量整合中最常用的一种策略，它直接将两个向量连接起来，将得到的连接向量送入 Softmax 中进行进一步解析。向量相加是深度学习中向量交互的常用方法，在很多工作中都有不错的效果，它将两个向量的各个位上的值相加得到同样维度的和向量。向量相减就是将两个向量的每一位相减，另一种类似的向量操作是将向量按位相减后取绝对值。向量相乘即将两个向量中的数值按位相乘。我们首先对基于 LSTM 的评价对象极性判别模型中的 LSTM_L 的最后一个隐含层的输出向量和 LSTM_R 的最后一个隐含层的输出向量进行以上四种向量整合后送入 Softmax 中进行分类，然后对基于双向 LSTM 的评价对象极性判别模型中 BLSTM_L 和 BLSTM_R 的隐含层输出经过均化池处理后的两个向量进行以上四种向量整合，实验结果如表 3-6 所示。

表 3-6 LSTM_L 和 LSTM_R (BLSTM_L 和 BLSTM_R) 的隐含层输出向量整合方式对比实验

向量整合方式	LSTM 训练集	LSTM 测试集	BLSTM 训练集	BLSTM 测试集
	准确率	准确率	准确率	准确率
向量相连	81.7%	81.5%	83.4%	82.7%
向量相加	80.5%	79.3%	81.9%	81.3%
向量相减	81.2%	80.7%	82.8%	82.1%
向量相乘	75.9%	74.1%	77.2%	76.4%

由实验结果可以看出，向量相连在训练集合测试集的准确率最高，其次

是向量相减，向量相加的结果次之，向量相乘的结果最差。其原因可能是，向量相连能最大程度地保留上文和下文的原有的语义信息。因此，我们在模型中将向量相连然后送入 SoftMax 中解析。

3.5.3.4 总体对比实验结果

表 3-7 给出了基线方法以及本文提出的基于提升的模型融合方法、基于 LSTM 的评价对象极性判别方法和基于双向 LSTM 的评价对象极性判别方法的实验结果。

表 3-7 总体实验结果

方法	训练集（准确率）	测试集（准确率）
基线方法（规则）	69.5%	77.8%
基线方法（机器学习）	72.7%	81.0%
模型融合方法	81.7%	81.2%
LSTM	81.7%	81.5%
双向 LSTM	83.4%	82.7%

由表 3-7 可以看出，本文提出基于提升的模型融合方法、基于 LSTM 的方法和基于双向 LSTM 的方法在效果（准确率）上均超过了基于规则的基线方法和基于机器学习的基线方法，证明了我们所提出的方法在评价极性判别任务中的有效性。其中，基于提升的模型融合方法准确率超过基线方法的原因可能是模型融合能充分利用线性模型 SVM 和非线性模型随机森林的优势。基于深度学习的两种模型在效果上均高于模型融合的方法，证明了词向量表示的语义信息是极为有效的，且将评价对象词向量和上下文中的每个单词的词向量一同输入网络中最终获得的向量能够充分体现评价对象与上下文的语义关系。基于 LSTM 的方法在训练集与测试集上的准确率均比基于双向 LSTM 的方法略低，其原因可能是双向 LSTM 能充分利用上下文信息。

为了研究模型在每种极性上的表现情况，我们利用本文提出的模型在四种不同的极性的评论数据上进行了实验。表 3-8 列出了各种方法在不同极性上的实验结果。

表 3-8 各种方法在不同极性上的实验结果

方法	积极	消极	中立	冲突
基线方法（规则）	79.2%	78.4%	77.4%	41.2%
基线方法（机器学习）	82.5%	80.2%	81.3%	50.3%

表 3-8 (续表)

方法	积极	消极	中立	冲突
模型融合方法	82.7%	81.4%	81.6%	54.2%
LSTM	82.9%	81.8%	81.3%	60.4%
双向 LSTM	83.2%	82.8%	81.9%	68.4%

由表 3-8 可以看出,本文提出的基于提升的模型融合方法在各个极性上的效果均超过基线方法,并且在极性为冲突的评价对象极性判别上标签尤为突出,其准确率分别高出两种基线方法 13%和 3.9%,然而,极性为冲突的评价对象在训练数据集中占比极低(2.46%),这就说明模型融合方法相比两种基线方法来说需要更少的训练样本,并且在训练样本充足时,能够更有效地判断评价对象极性。而基于深度学习的两种评价对象极性判别方法在各个极性上的效果均超过了基于提升的模型融合方法,尤其是我们提出的基于双向 LSTM 的极性判别方法,在各个极性上取得了最高的准确率,基于 LSTM 的方法的实验结果比基于双向 LSTM 的方法略差一些,但仍然比基线方法和模型融合的方法高。

为了验证我们提出的方法在餐馆评论领域具有普适性,我们在美国最大的点评网站 Yelp 的评论数据集上进行了评价极性判别实验,实验结果如表 3-9 所示。

表 3-9 各方法在 Yelp 数据集上的实验结果

方法	训练集 (准确率)	测试集 (准确率)
基线方法 (规则)	68.3%	76.5%
基线方法 (机器学习)	70.7%	80.4%
模型融合方法	80.7%	80.5%
LSTM	81.9%	81.8%
双向 LSTM	84.2%	83.7%

由表 3-9 可以看出,我们提出的方法在 Yelp 评论数据集上的效果仍然比基线方法好,尤其是双向 LSTM 方法,在训练集和测试集上均达到了最好的实验效果。

3.6 本章小结

本章详细介绍了进行评价对象极性判别的方法，然后通过实验得出我们提出的基于深度学习的方法和基于提升的模型融合方法都比基线方法准确率高，而基于深度学习的方法比基于提升的模型融合方法效果更好，基于深度学习的方法中我们提出的基于双向 **LSTM** 的模型取得了最好的效果。另外，本章对深度学习模型中使用的词向量维度、输入向量中是否加入评价对象向量、隐含层输出结果向量的整合方式都进行了实验探索，结果表明使用 100 维的词向量模型的准确率更高，此外将评价对象向量与每个单词的词向量相连接输入模型中进行训练，模型能够有效地获得评价对象与上下文中每个单词的语义关系，使得极性判别结果更准确，而对于隐含层输出结果向量的整合方式，实验结果表明直接将两个向量相连能更好地代表上文和下文的语义信息，实验效果更好。

第4章 餐馆评论情感分析系统的设计与实现

4.1 引言

互联网及电子商务的发展使得网上的评论信息飞速增长，面对如此庞大的评论信息，用户很难人工地在短时间内从中获得对自身有意义的信息。因此实现一个能够自动进行情感分析的系统是非常有意义的。本章针对某餐饮网站的评论信息应用本文提出的评价对象抽取和评价极性判别方法实现了一个餐馆评论情感分析系统。

本章的内容安排如下：第4.2节介绍系统的总体架构；第4.3节介绍各个模块的实现；第4.4节对系统功能进行展示，通过示例证明本文提出评价对象抽取和评价对象极性判别方法的有效性；第4.5节对本章内容进行总结。

4.2 餐馆评论情感分析系统的总体结构与设计

系统使用目前较为流行的浏览器/服务器架构实现，分为前端用户交互和后台数据处理两部分，用户交互部分主要包括统计信息展示模块和用户交互模块，数据处理模块包括数据获取模块、评论文本分析模块和结果数据存储模块。本节分别从系统模块设计及系统工作流程两个方面来对系统进行介绍。

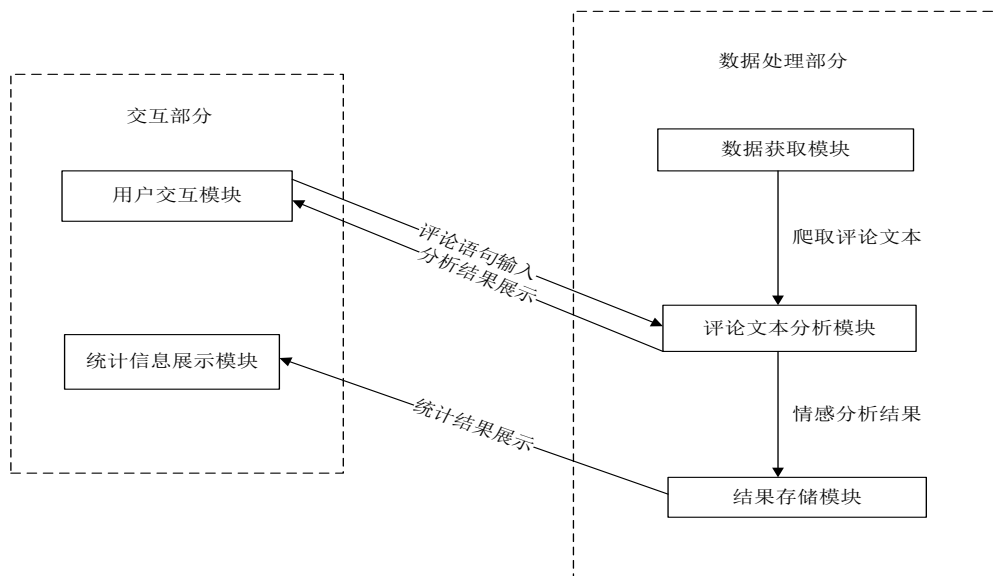


图 4-1 餐馆评论情感分析系统模块设计与交互

图 4-1 给出了系统包含的模块及模块的相互依赖关系。其中，用户评价交互模块是提供给用户的对餐馆进行评价的接口。该模块接收用户输入的评价

文本，然后通过评论文本分析模块分析出评价对象和评价对象极性，然后通过结果数据存储模块将结果存入数据库，同时将分析结果展示在前端页面。

数据获取模块对餐饮网站的网页进行爬虫，从而抽取其中有价值的评论文本信息。其方法主要是动态保存一个 URL 排队器，依次将 URL 所指向网页的内容下载到本地，然后抽取其中的评论文本及待跳转的 URL，将 URL 加入 URL 排队器。

评论文本分析模块对数据获取模块爬取的评论文本进行情感分析，包括对数据的预处理，评价对象抽取和评价对象极性判别。其中，评价对象抽取任务采用本文提出的基于输出依赖双向 LSTM 的方法，评价对象极性判别任务采用本文提出的基于双向 LSTM 的极性判别方法。

评论文本分析模块的分析结果由结果数据存储模块保存到 MySQL 数据库。统计信息展示模块对数据库中的结果数据进行统计，并以评价对象占比图和评价极性占比图的形式展示。历史评论信息的详细内容也即评论语句及抽取出的评价对象和极性可以按照评价对象和评价对象极性两个条件分类查询。

整个系统的大致工作流程是：

(1) 首先，用户交互模块中用户可以在系统中输入对餐馆的评价信息，系统接收用户评价信息并利用后台相应的接口对其进行情感分析，将分析结果展示在前端页面。

(2) 其次，数据处理模块中系统定时从餐饮网站爬取含有评论文本的网页，然后对网页内容进行处理抽取评论文本信息并以 xml 的格式存储。

(3) 系统逐条读取格式化的评论文本信息并利用本文提出的基于输出依赖双向 LSTM 的方法对其进行评价对象抽取，然后利用本文提出的基于双向 LSTM 的极性判别方法进行评价对象极性判别。

(4) 接着把获得的评价对象及评价对象极性保存到 MySQL 数据库中。

(5) 对数据库中存储的所有结果数据进行统计，生成评价对象占比图和评价极性占比图，将用户的所有评价信息进行可视化地展示。

(6) 用户可以根据评价对象和评价极性联合查询出符合条件的所有评价语句，点击某条语句可以显示该评价语句的涉及的评价对象及相应的极性。

4.3 餐馆评论情感分析系统各模块详细设计与实现

本节将具体介绍系统的五个重要模块即用户评价交互模块、数据获取模块、评论文本分析模块、结果数据存储模块和统计信息展示模块。

4.3.1 用户评价交互模块

该模块是用户与情感分析系统进行交互的模块，用来证明基于输出依赖双向 LSTM 方法在提取餐馆评价对象任务中的准确性及基于双向 LSTM 模型在评价对象极性判别任务中的有效性。用户可以在系统中输入自己对餐馆的评论然后提交到后台，后台程序除了存储用户的评论到数据库外还通过评价对象抽取接口和评价极性判别接口将对用户的评论进行分析，并将分析结果存入数据库，同时展示在页面。后续的统计分析也会对用户输入的对餐馆的评论进行统计。

4.3.2 数据获取模块

数据获取模块的作用是定时地从指定的餐饮网站中爬取评论文本信息，作为评论文本分析模块的输入数据。餐饮网站评论页面上的评论语句是按一定的结构排列在网页中的，所以要抽取每条评论语句是比较容易的。抽取出的每条评论语句都以 XML 格式进行存储，其格式如表 4-1 所示。

表 4-1 餐馆评论数据获取模块的处理结果

处理结果
<sentence id="1">It is very overpriced and not very tasty.</sentence>
<sentence id="2">The design and atmosphere is just as good.</sentence>
<sentence id="3">The price is reasonable although the service is poor.</sentence>

4.3.3 评论文本分析模块

评论文本分析模块是整个系统的主要功能实现，该模块包括三个功能即文本预处理、评价对象抽取和评价对象极性判别。首先读取从网站上爬取并以 XML 格式存储的评论文本，然后利用本文提出的基于输出依赖双向 LSTM 的方法进行评价对象抽取并利用基于双向 LSTM 的方法进行评价对象极性判别。其流程图如图 4-2 所示。

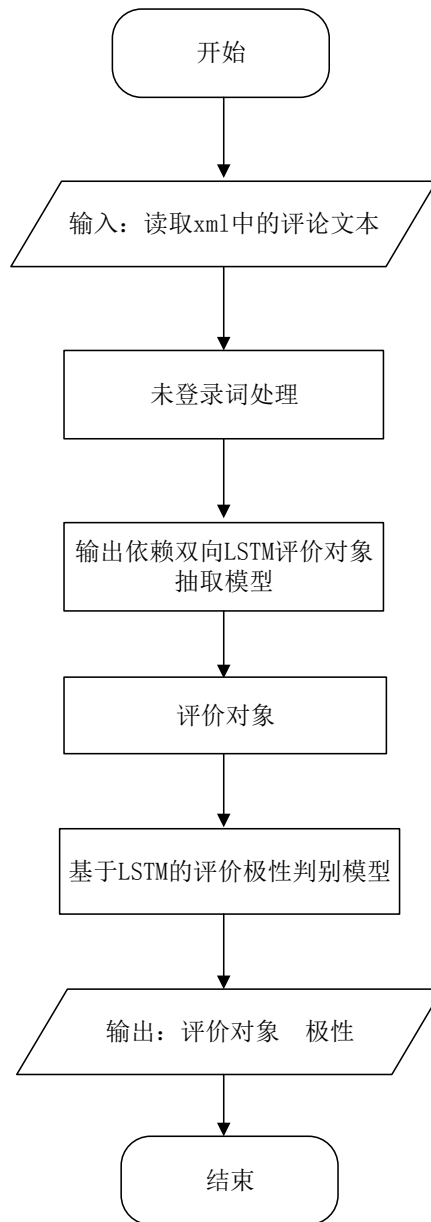


图 4-2 评论文本分析模块流程图

4.3.4 结果数据存储模块

结果数据存储模块将评价文本分析模块的分析结果及评论语句本身存入 MySQL 数据库中，方便后续的查询和统计操作。MySQL 数据库的评价语句分析涉及的数据表主要包括评价语句表和评价对象及极性表。评价语句表如表 4-2 所示，评价对象及极性表如表 4-3 所示。

表 4-2 评价语句表 t_comment 表的定义

列名	数据类型	含义
id	varchar(40)	评价 id
comment	varchar(255)	评价语句

表 4-3 评价对象及极性表 t_comment_polarity 表的定义

列名	数据类型	含义
id	varchar(40)	评价对象及极性 id
comment_id	varchar(40)	评价 id
target	varchar(255)	评价对象
polarity	varchar(255)	评价极性

4.3.5 统计信息展示模块

统计信息展示模块主要包括三部分，一是评价对象占比图，二是评价极性占比图模块，三是评价信息分类浏览模块。评价对象占比图对所有的评价文本中涉及的评价对象及其占比进行统计，以饼形图的形式将各个评价对象所占的比例进行展示。评价极性占比图统计某评价对象的四种评价极性所占的比例并以饼形图的方式进行展示。评价信息分类浏览模块的主要功能是根据评价对象和评价极性进行联合查询，查询出符合条件的所有评价语句，点击某条评论语句可以显示该语句中的所有评价对象和其对应的评价极性。页面采用 HTML+CSS 开发，并利用 JavaScript 增强展示效果，后台采购 Java 语言开发，利用较为方便快捷的 MySQL 数据库和 Apache 服务器。

4.4 系统功能展示

本文实现了一个餐馆评论情感分析系统，帮助用户更快速地了解消费者对餐馆的情感倾向。本节将从系统登录、评价对象占比图、评价极性占比图、用户评价交互、评价信息分类浏览五个方面展示系统功能，从而证明本文提出的评价对象抽取和评价极性判别方法在实际应用中的有效性。

4.4.1 系统登录

系统登录模块是登录系统的入口。系统登录界面如图 4-3 所示。



图 4-3 系统登录界面

登录成功之后，系统首页如 4-4 所示。

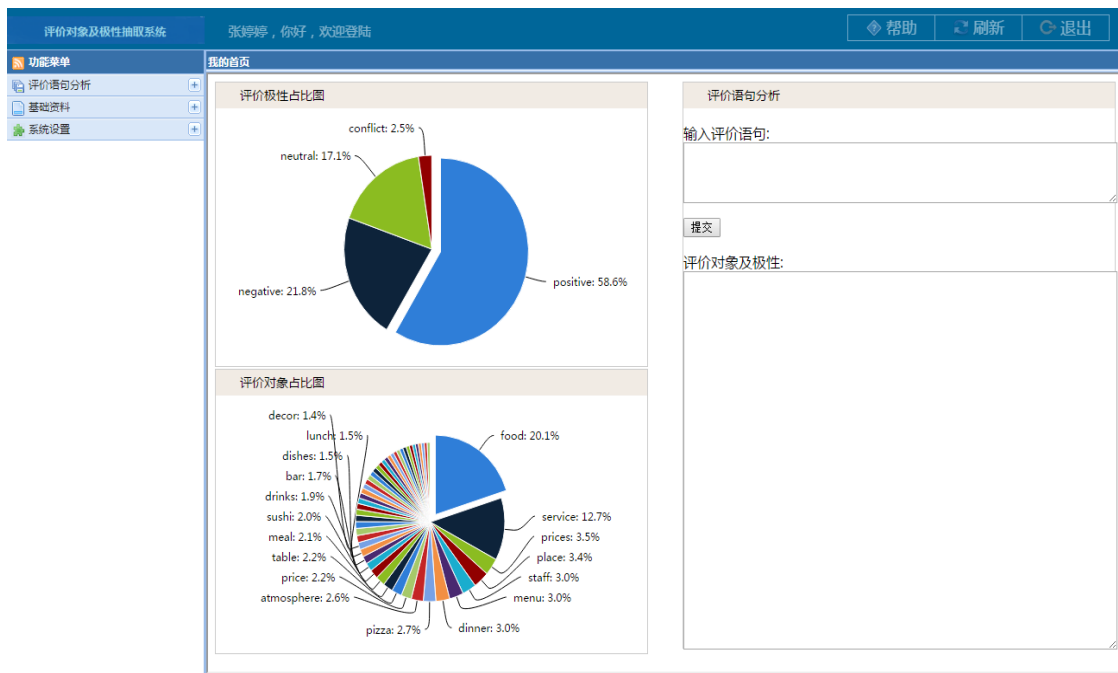


图 4-4 系统首页

4.4.2 评价对象占比图

评价对象占比图以直观的方式展示了所有评论文本中各个评价对象的数量及其所占的比例。各个评价对象及其所占的比例是系统通过对其定时爬取的评论文本利用基于输出依赖双向 LSTM 的评价对象抽取方法进行评价对象

抽取的基础上统计得到的。如图 4-5 所示，评价对象 food 所占的比例为 20.1%, service 所占的比例为 12.7%，prices 所占的比例为 3.5%等。

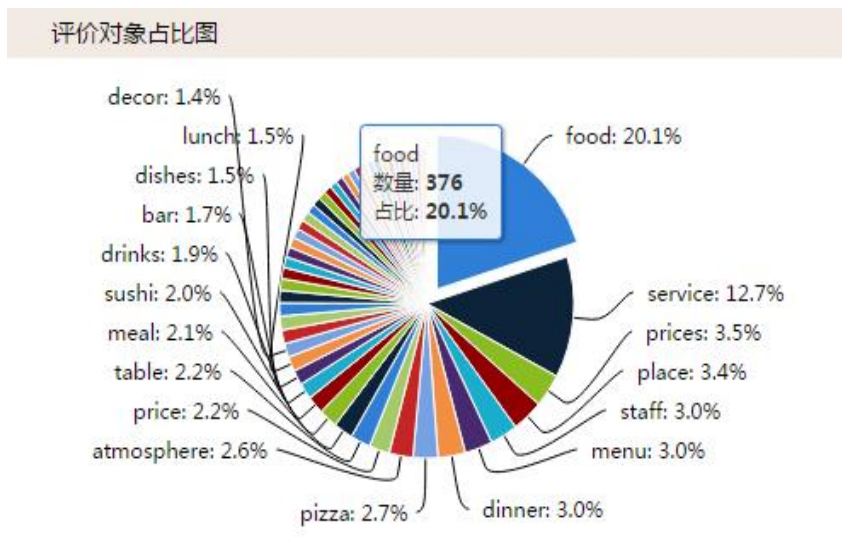


图 4-5 评价对象占比图

该占比图不仅可以显示某评价对象所占的比例还可以查看该评价对象的被评价的总次数，只需点击该评价对象在饼形图中所占的那块区域即可显示，如图 4-5 所示，food 被评价的次数为 376 次。

4.4.3 评价极性占比图

评价极性占比图显示了某评价对象在评论语句中的四种评价极性的占比图，该图与评价对象占比图是联系在一起的，点击评价对象占比图中的某个评价对象，评价极性占比图就显示该评价对象在评论语句中的四种评价极性占比。评价极性所占的比例是在利用本文提出的基于双向 LSTM 的方法进行评价极性判别的基础上统计得出的。如图 4-6 所示，点击评价对象占比图中的 food 后，评价极性占比图就会显示 food 的四种评价极性的占比，评价极性为 negative 的占 22.1%, neutral 占 17.1%, conflict 占 2.4%, positive 占 58.4%。

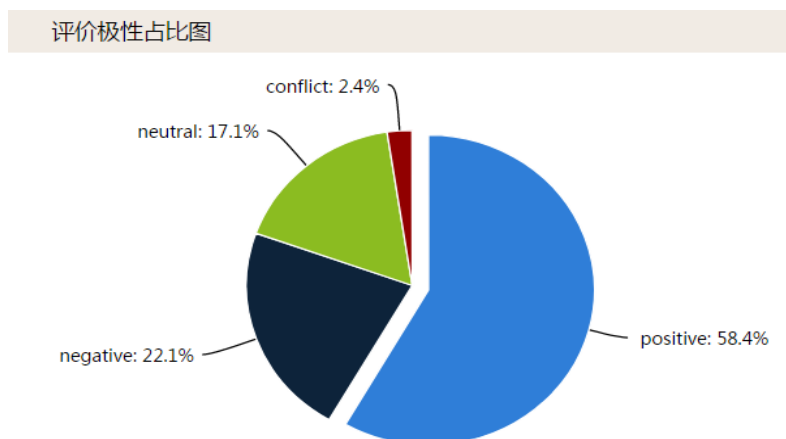


图 4-6 评价极性占比及数量

4.4.4 用户评价交互

用户评价交互功能是提供给用户的对餐馆各个方面进行评价的接口，用户在前端页面输入评价语句后系统对评价文本进行预处理并将处理后文本传到评价对象抽取接口抽取出评价对象，然后将评论文本和评价对象传到评价对象极性判别接口进行评价极性判别，最后将分析结果存入数据库的同时传回前端页面进行展示。如图 4-7 所示，输入评价语句“The food is delicious but the waiter is in bad taste”，点提交按钮，系统将保存评价语句及分析结果到 MySQL 数据库，并将分析结果展示在页面。该评价语句中有两个评价对象 food 和 waiter，分别对应的评价极性为 positive 和 negative。该结果表明了我们提出的评价对象抽取和评价极性判别方法的高效性及准确性。

评价语句分析

输入评价语句:

The food is delicious but the waiter is in bad taste.

提交

评价对象及极性:

评价对象: food 极性:positive
评价对象: waiter 极性:negative

图 4-7 用户评价交互界面

4.4.5 评价信息分类浏览

评价信息分类浏览的主要功能是按照条件查询符合条件的评价信息及其对应的评价对象和评价极性。浏览评价信息时可以按照评价对象和评价极性两个条件查询符合条件的评价语句。例如想要了解关于评价对象 food 的相关评价信息即可直接在评价对象文本框中输入“food”，然后点击查询即可查出所有关于 food 的评价语句。点击某个评价语句页面下方会显示该评价语句中涉及的评价对象和评价极性。如下图 4-8 所示，点击评价语句“The food was boring and expensive.”页面下方显示评价对象是 food，评价极性是 negative。

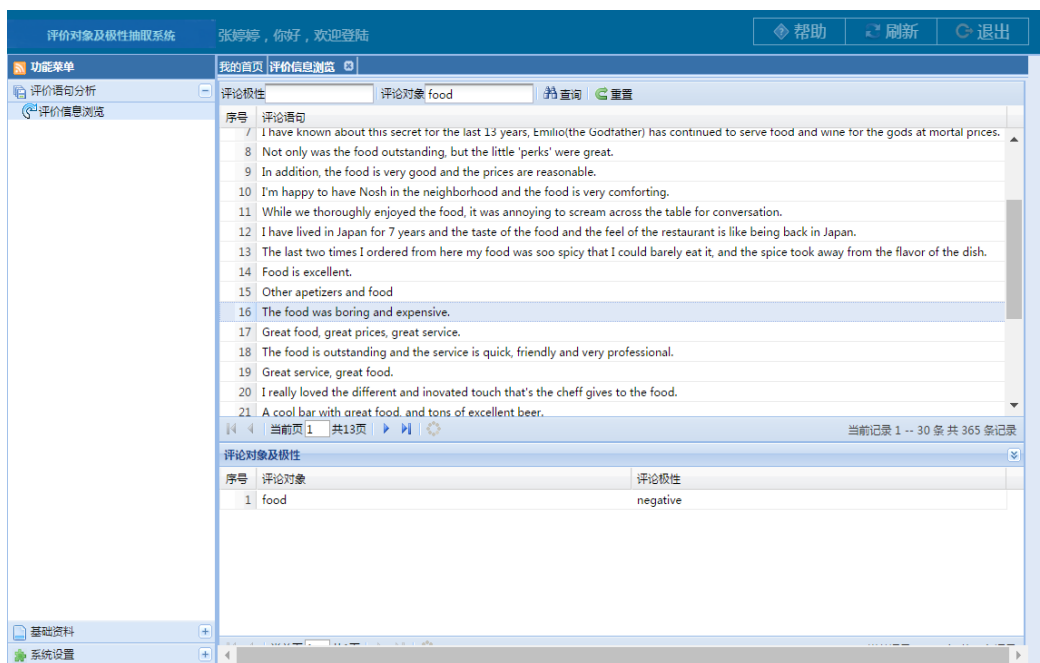


图 4-8 评价信息分类浏览

4.5 本章小结

本章主要介绍了餐馆评论情感分析系统的相关模块的设计与实现，包括各模块的实现及工作流程，以及相关分析结果数据的存储，最后以直观的图形界面展示了评论的相关分布与统计，并且可以按照评价对象和评价极性两个条件查询出相关评价语句及其分析结果，使用户可以直观地感受各个评价对象的情感趋向。通过对系统的演示，证明了文本提出的基于输出依赖双向 LSTM 的方法和基于双向 LSTM 的评价极性判别方法在评价对象抽取和评价极性判别任务中的有效性。

结 论

互联网技术的飞速发展使得电子商务在人们的衣食住行各个方面都扮演了越来越重要的角色,随之而来的是网上用户评论数据的急速增长。用户的评价信息对指导消费者行为和商家的改进计划具有非常重要的作用。因此,对这些繁杂庞大的评论信息进行分析并抽取出其中的评价对象及相应的评价极性有着重要意义。本文针对评价对象抽取和评价对象极性判别任务分别用传统的机器学习方法和基于深度学习的方法进行了研究,然后在餐馆评论情感分析系统中应用效果最好的模型对评论信息进行评价对象抽取和情感极性判别。本文的主要贡献有:

(1) 实现了基于输出依赖双向 LSTM 的评价对象抽取模型,该模型利用双向 LSTM 作为隐含层,从而有效地利用上文和下文的信息,同时在模型的输出层之间加入自连接,将上一时间步骤上输出层的输出结果送入当前时间步骤的输出层,从而有效地利用输出标签之间存在的依赖关系,此外,我们在输出层加入词性特征、句法分析特征、情感倾向性特征和命名实体识别特征来提升模型效果,实验证明,该模型的效果较基线方法及其他方法都有很大提升。其次,本文实现了基于特征的 CRF 模型,我们主要在特征选择与组合上对模型的性能进行提升,我们提取了小写单词特征、词根特征、单词所在名词短语数目、最近形容词距离、依存关系特征、词向量特征及情感倾向性特征等,通过对不同的特征组合分别进行实验,得出结果最好的特征组合。

(2) 提出了基于双向 LSTM 的评价对象极性判别模型,该模型利用两个双向 LSTM 网络,分别为 $BLSTM_L$ 和 $BLSTM_R$, $BLSTM_L$ 收集评价对象的上文语义信息, $BLSTM_R$ 收集评价对象的下文语义信息。为了模型能捕获到每个单词与评价对象之间的语义关系,我们将当前单词词向量和评价对象向量进行连接后一同送入模型中进行处理,然后将 $BLSTM_L$ 和 $BLSTM_R$ 的每个时间步骤上隐含层的输出向量送入各自的均化池中,最后将两个均化池的输出向量进行连接操作得到蕴含当前评价对象与评论语句上下文之间语义关系的向量,最后再将该向量输入到 Softmax 中进行计算,该模型的实验结果超过同类模型的最好水平。此外,本文还实现了基于 LSTM 的评价对象极性判别模型,并针对基于 LSTM 的模型和基于双向 LSTM 的模型中的三个关键部分的优化改进策略进行对比实验,探索使模型效果达到最佳的方法。其次,本文实现了一种基于机器学习的方法即基于提升的模型融合方法,利用

Boosting 提升方法将 **SVM** 模型和随机森林模型进行融合。该方法依次训练各个分类模型,在每次训练得到的结果基础上改变训练数据集合的概率分布,即增大所有被上个模型错误分类的数据所占权重,减弱所有被上个模型正确分类的数据所占的权重。当所有模型训练完毕后,再按照各模型效果对结果加权得到最终的模型结果。

(3) 构建了一个易用的餐馆评论情感分析系统。该系统通过爬虫技术获取餐饮网站上的评论信息,然后利用本文提出的基于输出依赖双向 **LSTM** 的评价对象抽取方法和基于双向 **LSTM** 的评价对象极性判别模型对这些信息进行分析处理,获得每条评论语句的评价对象及其评价极性,最终用图形化的表现方式将对评论进行情感分析的结果进行统计并展示。此外,系统还提供了用户评价交互接口,后台对评价信息进行情感分析后返回评价对象和情感极性。本文通过对系统进行实例展示证明了本文提出的评价对象抽取及评价对象极性判别方法的有效性。

尽管本文对餐馆评论领域评价对象抽取及评价极性判别的方法进行了很多尝试并取得了不错的效果,但是,研究过程中还有若干问题需要进一步探索,下一步的工作如下:

(1) 对于评价对象抽取的研究虽然达到了较好的结果,但还有提升的空间,在下一步的研究工作中可以考虑引入卷积神经网络,将卷积神经网络与输出依赖双向 **LSTM** 模型结合,此外,可以从特征的角度入手,提取更多有效的特征来提高模型效果。

(2) 针对评价极性判别的方法,可以尝试在 **LSTM** 和双向 **LSTM** 结构中引入 **Attention** 机制,利用 **Attention** 机制有效捕捉评论文本中对评价极性判别起重要作用的词进行关注,从而提高模型进行评价极性判别的准确率。

参考文献

- [1] Kim SM, Hovy E. Automatic Detection of Opinion Bearing Words and Sentences. In: Proceedings of IJCNLP-2005. 2005. 61–66.
- [2] Yu H, Hatzivassiloglou V. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In: Proceedings of EMNLP-2013. 2013. 129–136.
- [3] Zhao M, Zhang T Z, Chai J P. Judging Subjective and Objective Sentence Based on 2-POS Subjective Models[C]// IEEE, Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE, Intl Conf on Autonomic and Trusted Computing and 2015 IEEE, Intl Conf on Scalable Computing and Communications and ITS Associated Workshops. IEEE, 2015:622-625.
- [4] Whitelaw C, Garg N, Argamon S. Using appraisal groups for sentiment analysis[J]. Proceedings of Acm Sigir Conference on Information & Knowledge Management Bremen, 2015, 9(1):625-631.
- [5] Bloom K, Garg N, Argamon S, et al. Extracting appraisal expressions[C]// Proceedings of Human Language Technologies/North American Association of Computational Linguists2013:308--315.
- [6] Pontiki M, Galanis D, Papageorgiou H, et al. SemEval-2016 Task 5: Aspect Based Sentiment Analysis[C]// International Workshop on Semantic Evaluation. 2016:19-30.
- [7] Hu M, Liu B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004: 168-177.
- [8] Popescu A M, Etzioni O. Extracting product features and opinions from reviews[M]//Natural language processing and text mining. Springer London, 2007: 9-28.
- [9] Blair-Goldensohn S, Hannan K, McDonald R, et al. Building a Sentiment Summarizer for Local Service Reviews[C]// Www2008 Workshop: Nlp in the Information Explosion Era. 2008:200-207.
- [10] Zhu J, Wang H, Tsou B K, et al. Multi-aspect opinion polling from textual reviews[C]//Proceedings of the 18th ACM conference on Information and knowledge management. ACM, 2009: 1799-1802.

- [11] Kessler J S, Nicolov N. Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations[C]//ICWSM. 2009.
- [12] Somasundaran S, Wiebe J. Recognizing stances in online debates[C]//Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1. Association for Computational Linguistics, 2009: 226-234.
- [13] Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation[J]. Computational linguistics, 2011, 37(1): 9-27.
- [14] Wu Y, Zhang Q, Huang X, et al. Phrase dependency parsing for opinion mining[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3. Association for Computational Linguistics, 2009: 1533-1541.
- [15] Su F, Markert K. From words to senses: a case study of subjectivity recognition[C]//Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics, 2008: 825-832.
- [16] Hai Z, Chang K, Kim J. Implicit feature identification via co-occurrence association rule mining[M]//Computational Linguistics and Intelligent Text Processing. Springer Berlin Heidelberg, 2011: 393-404.
- [17] Wong T L, Lam W. Hot Item Mining and Summarization from Multiple Auction Web Sites[C]// IEEE International Conference on Data Mining. IEEE Computer Society, 2005:797-800.
- [18] Soni S, Sharaff A. Sentiment Analysis of Customer Reviews based on Hidden Markov Model[C]// International Conference on Advanced Research in Computer Science Engineering & Technology. ACM, 2015:12.
- [19] Jin W, Ho H H, Srihari R K. OpinionMiner: a novel machine learning system for web opinion mining and extraction[C]//Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009: 1195-1204.
- [20] Rubtsova Y, Koshelnikov S. Aspect Extraction from Reviews Using Conditional Random Fields[M]// Knowledge Engineering and Semantic Web. Springer International Publishing, 2015:158-167.
- [21] Toh Z, Wang W. Dlirec: Aspect term extraction and term polarity classification system[C]//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014: 235-240.

- [22] Zhang F, Zhang Z, Lan M. ECNU: A combination method and multiple features for aspect extraction and sentiment polarity classification[J]. 2014.
- [23] Li F, Huang M, Zhu X. Sentiment analysis with global topics and local dependency[C]// Twenty-Fourth AAAI Conference on Artificial Intelligence. AAAI Press, 2010:1371-1376.
- [24] Toh Z, Su J. NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction[C]// International Workshop on Semantic Evaluation. 2015:496-501.
- [25] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.
- [26] Irsoy O, Cardie C. Opinion Mining with Deep Recurrent Neural Networks[C]//EMNLP. 2014: 720-728.
- [27] Yang B, Cardie C. Extracting opinion expressions with semi-markov conditional random fields[C]//Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Association for Computational Linguistics, 2012: 1335-1345.
- [28] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.
- [29] Mesnil G, He X, Deng L, et al. Investigation of Recurrent-Neural-Network Architectures and Learning Methods for Spoken Language Understanding[C]. INTERSPEECH. 2013: 3771-3775.
- [30] Liu P, Joty S, Meng H. Fine-grained opinion mining with recurrent neural networks and word embeddings[C]//Conference on Empirical Methods in Natural Language Processing (EMNLP 2015). 2015.
- [31] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [32] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.
- [33] Jordan M I. Serial order: A parallel distributed processing approach[J]. Advances in psychology, 1997, 121: 471-495.
- [34] Turney P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002: 417-424.

- [35] Mullen T, Collier N. Sentiment Analysis using Support Vector Machines with Diverse Information Sources[C]//EMNLP. 2004, 4: 412-418.
- [36] Pang B, Lee L, Vaithyanathan S. Thumbs up?: sentiment classification using machine learning techniques[C]//Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002: 79-86.
- [37] Ng V, Dasgupta S, Arifin S M. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews[C]//Proceedings of the COLING/ACL on Main conference poster sessions. Association for Computational Linguistics, 2006: 611-618.
- [38] 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究[J]. 中文信息学报, 2007, 21(6): 88-94.
- [39] Wagner J, Arora P, Cortes S, et al. Dcu: Aspect-based polarity classification for semeval task 4[J]. 2014.
- [40] Kiritchenko S, Zhu X, Cherry C, et al. NRC-Canada-2014: Detecting aspects and sentiment in customer reviews[C]//Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). 2014: 437-442.
- [41] Tang D, Qin B, Liu T. Aspect Level Sentiment Classification with Deep Memory Network[J]. arXiv preprint arXiv:1605.08900, 2016.
- [42] Marti U V, Bunke H. Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system[J]. International journal of Pattern Recognition and Artificial intelligence, 2001, 15(01): 65-90.
- [43] McCallum A, Li W. Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons[C]//Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. Association for Computational Linguistics, 2003: 188-191.
- [44] Yangsen J Y Z, Tong Y. Study of Sentiment Classification for Chinese Microblog Based on Recurrent Neural Network[J]. Chinese Journal of Electronics, 2016, 25(4):601-607.
- [45] 程佳军, 张鑫, 张胜, et al. Sentiment Parsing of Chinese Microblogs Using Recurrent Neural Network[J]. Journal of Donghua University, 2016, 33(3):501-504.
- [46] Li X L. Recurrent neural network training with preconditioned stochastic gradient descent[J]. 2016.

- [47] Tang D, Qin B, Feng X, et al. Effective LSTMs for Target-Dependent Sentiment Classification[J]. Computer Science, 2016.
- [48] Tan M, Santos C D, Xiang B, et al. LSTM-based Deep Learning Models for Non-factoid Answer Selection[J]. Computer Science, 2016.
- [49] Wang Y, Huang M, Zhu X, et al. Attention-based LSTM for Aspect-level Sentiment Classification[C]// Conference on Empirical Methods in Natural Language Processing. 2016:606-615.
- [50] Wagner J, Arora P, Cortes S, et al. DCU: Aspect-based Polarity Classification for SemEval Task 4[C]// International Workshop on Semantic Evaluation. 2014.

哈尔滨工业大学学位论文原创性声明和使用权限

学位论文原创性声明

本人郑重声明：此处所提交的学位论文《面向餐馆评论的情感分析关键技术研究》，是本人在导师指导下，在哈尔滨工业大学攻读学位期间独立进行研究工作所取得的成果，且学位论文中除已标注引用文献的部分外不包含他人完成或已发表的研究成果。对本学位论文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。

作者签名：

张婷婷

日期：2017 年 6 月 28 日

学位论文使用权限

学位论文是研究生在哈尔滨工业大学攻读学位期间完成的成果，知识产权归属哈尔滨工业大学。学位论文的使用权限如下：

(1) 学校可以采用影印、缩印或其他复制手段保存研究生上交的学位论文，并向国家图书馆报送学位论文；(2) 学校可以将学位论文部分或全部内容编入有关数据库进行检索和提供相应阅览服务；(3) 研究生毕业后发表与此学位论文研究成果相关的学术论文和其他成果时，应征得导师同意，且第一署名为哈尔滨工业大学。

保密论文在保密期内遵守有关保密规定，解密后适用于此使用权限规定。

本人知悉学位论文的使用权限，并将遵守有关规定。

作者签名：

张婷婷

日期：2017 年 6 月 28 日

导师签名：

刘远超

日期：2017 年 6 月 28 日

致 谢

时光荏苒，两年的硕士生涯即将结束，驻足回首，工大浓厚的学术氛围、朴实严谨的学风、追求卓越的创新精神以及优秀的老师同学都给我留下了深刻的印象，身为工大人，我感到无比自豪的同时也从内心不断暗示自己要一直努力做一个优秀的人。在此，请允许我向所有指导我的老师、帮助我的同学、关心我的家人和朋友表示真诚的感谢。

首先，我要感谢我的导师刘远超老师。刘远超老师严谨的教学理念、持之以恒的研究精神和渊博的学识都让我由衷折服。刘老师不仅在学习上给我很多启发和指导而且在生活上也对我们关怀备至，让我们感受到亲人般的温暖。在本文的命题和撰写方面刘老师都给了我很多宝贵的意见，在刘老师的无私指导下，我顺利完成了课题的研究和论文的撰写。在此对刘老师致以最真诚的感谢。

其次，我还要感谢智能技术与自然语言处理研究中心的王晓龙老师、刘秉权老师、林磊老师、孙承杰老师、刘铭老师和单丽莉老师为实验室的付出。在王老师和刘老师的带领及各位老师的努力下，为我们营造了一个学术氛围浓厚的优秀的实验室。很高兴能在这样的一个温暖的集体中学习，并且能得到这么多优秀的老师的指导，在两年的硕士生涯中，我受益匪浅。

在这里我还要感谢各位师兄师姐及实验室小伙伴，他们都非常地优秀，一直是我学习的榜样。感谢他们在学习跟生活中给我提供的巨大帮助，感谢他们的鼓励与陪伴。

最后，我要感谢我的父母和亲人，是他们的爱与支持让我有了前进的动力，使我面对困难时毫不惧怕，面对压力时充满希望，是他们的无私付出给予我信心和希望，在此我要深深地感谢他们！

感谢所有曾经帮助过我的人！