



广东工业大学

本科毕业设计（论文）

基于 BERT 模型的文本情感分析

学 院	计算机学院
专 业	网络工程
	（高速网络与云计算方向）
年级班别	2016 级（2）班
学 号	3116004982
学生姓名	赵舒宇
指导教师	王卓薇

2020 年 3 月

摘 要

TODO

关键词: BERT, 情感分析

Abstract

TODO

KEY WORDS: BERT, Sentiment Analysis

目 录

1	引言	1
1.1	研究背景及意义	1
1.1.1	研究背景	1
1.1.2	研究意义	2
1.2	国内外研究现状	2
1.2.1	国外研究现状	3
1.2.2	国内研究现状	5
1.3	本文研究内容	7
1.4	论文组织结构	8
2	相关技术介绍	9
2.1	文本预处理	9
2.1.1	分词技术	9
3	基于情感词典的情感分析	10
3.1	情感词典构建	10
3.1.1	基础情感词典构建	10
3.1.2	基础情感词典的扩展	12
3.2	基于情感词典的情感分析过程	13
3.3	实验结果与分析	14
3.4	本章小结	15
4	基于机器学习的情感分析	16
	结论	17
	参考文献	18

致谢	24
----------	----

1 引言

1.1 研究背景及意义

1.1.1 研究背景

得益于摩尔定律^[1]，生产电子元件的成本越来越低，电子产品的性能越来越好；同时，网速的高速发展也促进了移动互联网的进步。这两者都使得我国网民数量持续增长，用户在互联网上扮演的角色已经从单纯的信息接收者，变成现在信息的生产者。在社交媒体上，电商网站上，数亿的用户产出海量的数据，且这些数据是以指数级增长的，人工分析利用这些数据需要耗费大量的时间和人力，属于不可能的任务。但这些数据有着重要的社会价值和商业价值，如在诸如微博的这类社交媒体上，分析用户针对不同社会话题发布的微博内容，可以有效的监控网络舆情^[2]；而在淘宝这类电商平台上，分析用户给予不同商品的评价，商业公司可以快速了解这一商品的受欢迎程度。所以，如何通过自动化工具正确，快速分析利用这些数据，成为当今计算机科学研究领域重要的话题。

文本情感分析（Text Sentiment Analysis）作为自然语言处理（Natural Language Processing）领域的基本研究分支之一，主要任务是对文本的主客观性，观点，情绪，喜好的检测，分析，挖掘。文本情感分析作为多学科交叉研究领域，涉及语言学，统计学，机器学习，数据挖掘，人工智能等多个领域。近年来，随着机器学习的发展，学术界在情感分析技术上取得长足进步，工业界也开发许多情感分析技术落地的项目。自然语言处理于二十世纪中叶于美国兴起，主要应用场景都是针对英语。中文自然语言处理起步晚于英语，同时，中文自然语言处理也与英语有许多不同，主要表现在以下几个方面：①中文需要进行分词；②中文没有明显的屈折变化（时态，单复数等）；③中英文句法结构上存在不同。但随着技术的发展和数据的爆炸式增长，作为中文自然语言处

理的一个子类，提取分析非结构化文本的情感分析技术也愈发成熟。

1.1.2 研究意义

在像微博这样的社交媒体上，大量用户针对不同话题广泛发表自己的见解。这些文本通常数量庞大，依靠人力根本无法进行分析。但这些数据又有着重要的意义^[2-3]。利用这些数据，可以迅速准确的把握微博平台上重要事件的情感倾向，有效的进行舆情监控，对政府，商业公司维持舆论稳定有着极大帮助，这时就彰显出情感分析技术的重要性。

而对于淘宝，京东这样的电商网站，它们拥有着大量的商家和消费者用户。消费者们乐于对自己喜欢的商品给予好评，对于不满意的产品也会留下差评。而这些评论对商家的发展十分重要^[4]。根据评论，商家可以了解到顾客对于产品的喜好程度和产品的不足之处。针对这些客户意见，商家可以有针对性的对商品进行改善，提高客户满意度的同时，吸引更多的潜在客户，提高商业利润，而情感分析技术正能有效的帮助商家实现这些目的。

鉴于上述情况，本文选取微博文本数据，针对该数据集进行情感分析方法研究。尝试不同的方法来研究情感分析技术，并设计实验比较各种方法的分类结果，发掘出性能较好的分类模型，进而改善情感分类效果。

1.2 国内外研究现状

文本情感分析的核心问题是情感分类，主要研究任务分类两类，一是区分主客观文本，降低因客观信息对情感分析性能的干扰；二是对主观性文本进行情感分类^[5]。根据对文本划分的粒度不同，又可分为对词，句，篇章的情感分析。根据情感划分的粒度不同，可分为：①二元分类，包括消极态度和积极态度。②多元分类，根据人类的多种情感进行进一步细分，包括“快乐，悲伤，褒扬，贬斥，信息，意外”等十大类^[6]。

而对于主观文本情感分类，现在流行的方法包括以下三种：

- 1) 基于情感词典的方法。这一方法最符合人类的直觉，首先构建包含大量基本词汇的词典，词典中的词已经标记好词的类别。如“喜欢”会被标记为积极，而“讨厌”会被标记为消极。之后将输入的文本与词典进行匹配来判断其情感极性。这一方法有着局限性，通常与其他方法一同使用。
- 2) 基于机器学习的方法。这一方法关键点在于选择有效特征组合利用分类器进行情感分类^[7]。这一方法可以取得不错的成果，但对数据集要求很高，往往需要大量人工标注，成本不菲。
- 3) 基于深度学习的方法。这一方法通过建立深度神经网络，模拟人脑结构，在处理情感分类这种问题上有着优异的表现，是目前情感分析领域的主流方法。

1.2.1 国外研究现状

对于文本主客观分类，Wiebe 等^[8]人于 1999 年就将代词，形容词，基数词等作为特征值，设计了对文本主客观进行分类的朴素贝叶斯分类器 (Naive Bayes Classifier)。在 2003 年，Riloff 与 Wiebe^[9]提出 bootstrapping 文本主客观分类算法。该算法提高分类召回率的同时，只损失部分的精度。之后，Wiebe 等^[10]人还针对篇章粒度下的文本主客观分类进行了研究。

在情感词典技术方面，Turney^[11]设计了点互信息的方法 (Point Mutual Information, PMI) 来计算两个词之间的语义相关性。通过计算目标词与情感词之间的 PMI，用语义倾向 (Semantic Orientation, SO) 来表示该词的情感极性。公式如下：

$$PMI(word_1, word_2) = \log_2 \left[\frac{p(word_1 \& word_2)}{p(word_1) p(word_2)} \right] \quad (1.1)$$

$$SO(phrase) = PMI(phrase, \text{“excellent”}) - PMI(phrase, \text{“poor”}) \quad (1.2)$$

Hamouda 等^[12] 人又将在社交网络上被广泛应用的表情符号加入情感词库用于情感分析。Hu 和 Liu^[13] 在已有的情感词库的基础上, 通过 WordNet 的同义, 近义关系来判断目标情感词的情感极性。

在机器学习技术方面, Pang 等^[14] 人第一次将机器学习方法引入情感分析领域, 实验对比了三种机器学习方法: 朴素贝叶斯 (Naive Bayes), 最大熵模型 (Maximum Entropy) 和支持向量机 (Support Vector Machine)。尽管实验结果距离这些方法应用于主题分类时的表现有些差距, 但结合 SVM 和 Unigrams 特征, 准确率依然有 82.9%。Kim 和 Hovy^[15] 设计算法, 在给出一个主题的情况下, 算法可以自动找出该主题下人们的不同意见以及意见的情感倾向。Godbole 等^[16] 人构建了完整的情感分析系统。这一系统可以识别新闻和博客中的实体, 并且计算实体对应的情感倾向, 在不同数据集下的测试准确率在 82.7% 至 99.1% 之间。Lin 和 He^[17] 提出一种与以往其他研究者不同的方法, 他们基于隐含狄利克雷分配模型 (Latent Dirichlet Allocation, LDA) 设计了一个完全非监督的模型 (joint sentiment/topic model, JST)。该模型将情感信息与 LDA 模型结合, 实现了主题和主题相关情感信息的联合发现。通过挖掘文本中的情感和隐含的主题分布, 提炼出文本的情感倾向, 在不同语料下都取得很好的成果。

在深度学习技术方面, Hinton 等^[18] 人于 1986 年就提出了 Distributed Representation 的方法, 应用于自然语言处理场景, 该方法通常被成为词向量 (Word Embedding), 该方法能让相似, 相关的词在距离上更为接近, 效果优于之前常用的 N-grams 技术和 N-Pos 技术。2013 年, Mikolov 等^[19] 人终于将该方法实现, 称为 Word2Vec。这一方法在自然语言处理领域有着划时代的意义, 后续大量研究人员的工作基本都与 Word2Vec 离不开关系。其中 Kim^[20] 将卷积神经网络 (CNN) 与 Word2Vec 结合, 进行文本情感分类。使用 Word2Vec 训练词向量, 并在模型训练过程中保持微调, 取得了良好的结果 (准确率最高可达 89.6%)。之后又有新的词向量方法出现, 包括 Glove^[21] 和 fastText^[22]。应用于情感分类的其他深度学习模型还有循环神经网络 (RNN) 和 RNN 的变种: 长短时记忆网络 (LSTM)。Socher 等^[23] 人实现了一种矩阵向量的 RNN 网络 (Matrix-vector Recursive

Neural Network, MV-RNN), 该网络将每个词与矩阵相联系, 并表示为树形结构, 在长文本上结果比普通的词向量效果要好, F1 值达到 82.4%, 远超其他方法。Xu 等^[24] 人实现了一种缓冲长短时记忆网络 (Cached Long Short-Term Memory neural network, CLSTM) 对篇章级的文本进行情感分析。通过将模型中的记忆网络分组, 每组有不同的遗忘率, 低遗忘率的组保持全局语义特征, 高遗忘率的组用以记忆短距离语义特征。Bahdanau 等^[25] 人于 2014 年第一次将 Attention 机制引入自然语言处理领域, 用以处理机器翻译任务。这给其他学者以启发, 也开始将 Attention 机制用于自然语言处理领域的其他问题。Zhou 等^[26] 人将 Attention 机制与 LSTM 结合处理跨语言长文本的情感分析。在训练资源丰富的英语语料集上进行训练, 使用训练资源较少的中文语料集来测试分类效果, 实验显示结果良好。2018 年, Peters 等^[27] 人发表论文, 提出了 ELMo(Embeddings from Language Models), 一种基于大规模语料集上的预训练语言模型。这一论文让学者们发现了预训练语言模型的强大威力, 随后又有像 GPT(Generative Pre-training model)^[28] 这类模型的诞生。GPT 利用 Google 提出的 Transformer^[29] 框架中的 Decoder, 在各项自然语言处理任务中都取得良好的成果。直到 2018 年末, Google 发布了 BERT(Bidirectional Encoder Representations from Transformers)^[30], 刷新了自然语言处理领域内 11 项任务的记录, 是自然语言处理领域又一里程碑式的成果。

1.2.2 国内研究现状

国内主客观文本分类研究相较国外, 起步较晚。姚天昉和彭思崑^[31] 利用情感形容词, 人称代词, 不规范的标点符号和带情感的标点符号等作为特征, 比较了四类分类算法应用于主客观分类时的性能。叶强等^[32] 人提出基于 2-POS 模型的中文文本主观性判断方法, 分类性能接近英文类似研究结果。张博^[33] 将句法关系模块, 依存关系模块与 SVM 分类器结合, 在实验中取得良好结果, F-measure 值达到 88.9%。

在情感词典方面, 由于上文提到的 WordNet 中包含的中文词条数目远少于英文词条, 因此国内多基于董振东先生、董强先生开发的知网 (HowNet) 进行研究。朱嫣岚

等^[34] 基于 HowNet 提供的语义相似度和语义相关场功能进行了实验。实验发现语义相似度比语义相关场能更好的进行情感分类，实验的判别准确率最高达到 85.5%。同时朱嫣岚等^[34] 还发现提高基准词数量和进行词频加权都可以提高准确率。柳位平等^[35] 人基于 HowNet 提供的情感词语集，通过计算词语相似度的方法建立可以跨领域使用的基础情感词语库。在这一基础情感词语库的基础上，使用改进后的 TF-IDF 公式 (1.3) 进行情感分类实验，准确率达到 82.1%。

$$w(t, \bar{d}) = \frac{tf(t, \bar{d}) \times \log(N/n_t + 0.01) \times \bar{w}_t}{\sqrt{\sum_{i=d} [tf(t, \bar{d}) \times \log(N/n_t + 0.01) \times w_t]^2}} \quad (1.3)$$

李钝等^[36] 人从语言学角度出发，分析短语中词组合方式的特点，提出中心词概念来计算各词的倾向性，再由此计算短语的情感倾向性。实验结果表明，该方法对短语的倾向分类效果较好，为更大粒度的文本情感倾向识别打下基础。

在机器学习技术方面，唐慧丰等^[37] 人使用 n-gram，名词，形容词，动词，副词等作为文本表示特征，使用信息增益，CHI 统计量，文档频率等作为特征选择方法，在中心向量法，KNN，Winnow，Naive Bayes，SVM 这些机器学习方法上测试了情感分类效果。结果表示，结合 BiGrams，信息增益和 SVM 的分类效果明显高于其他方法。同时，提高训练集大小和选取适当数量的特征均能提升表现。徐军等^[38] 人使用朴素贝叶斯模型和最大熵模型对新闻及评论语料集进行了情感分类研究，其中最大熵模型结果更好。实验发现选择具有语义倾向的词汇和使用二值作为特征项权重能有效提高分类准确率。韦航和王永恒^[39] 使用 SVM 中的卷积树核函数获取语法树的结构化特征，对语法树进行基于主题的剪枝，可以有效提高情感分类准确度，在两个数据集上准确度均可达到 86% 以上。

在深度学习技术方面，国内学者也大多采用国外学者类似的方法，针对中文语言进行特色性优化。华中科技大学的曾江峰^[40] 提出一种可以感知句法语义的文本情感算法，在篇章级的文本上取得良好效果。该算法首先对每个句子的依存句法树生成 Tree-LSTM，再利用基于 Attention 机制的 LSTM 网络来获取篇章级的情感倾向。李

然^[41]实现了基于深度神经网络 (Deep Neural Network, DNN) 的短文本情感分类算法, 在大规模电商评论数据集上的实验结果优于主题模型。胡朝举和赵晓伟^[42]基于词向量技术, 将 CNN 与 LSTM 融合形成混合神经网络, 其中 LSTM 被更换为基于 Attention 机制的双向 LSTM 网络 (BiLSTM), 在 IMDB 数据集上取得了 89.6% 的准确度。张仰森等^[43]人提出一种基于双重注意力机制的微博情感分析方法。该方法结合微博特有的情感符号和注意力机制, 有效提升了微博文本情感信息的获取。在自然语言处理与中文计算会议 (NLPCC) 于 2014 年发布的微博情感测评公共数据集上, 相较当时已知的最好模型, 将宏平均和微平均的 F1 值分别提升了 2.02% 和 2.21%。赵亚欧等^[44]人将 ELMo 模型应用于中文情感分析场景, 提出一种融合 ELMo 模型的多尺度卷积神经网络 (Multi-Scale Convolutional Neural Network, MSCNN)。使用不同尺度的卷积核对文本进行卷积处理, 利用最大池化技术筛选获取到的特征, 将这些特征融合以获得最终的文本特征。该模型在酒店评论数据集和 NLPCC 于 2014 年提供的任务 2 数据集上均表现良好, 与之前学者在相同数据集上的训练效果相比, 有着长足进步。蔡国永等^[45]人使用 BERT 模型抽取句子语义表示为向量, 通过 CNN 来获取句子局部特征, 最后使用域对抗神经网络让源领域和目标领域抽取的特征更为相似, 以实现跨领域情感分析的目的。在针对亚马逊产品评论数据集的实验下表现良好, 准确率可达 89.9%。

1.3 本文研究内容

本文主要工作包括以下几个方面的内容: 基于情感词典的情感分析方法, 基于机器学习的情感分析方法, 基于深度学习的情感分析方法。本文使用的数据集为带标注的开源微博数据集^[46], 数据量约为 120,000 条。实验验证, TODO。三种方法的具体内容如下:

- 1) 基于情感词典的方法。基于该方法的情感分析依赖于情感词典的大小, 本文在基础情感词典上, 针对本文使用的微博数据集, 对情感词典进行针对性扩展, 有效改善了分类效果。

2) 基于机器学习的方法。

3) 基于深度学习的方法。

1.4 论文组织结构

本文共分为六个章节，论文的具体结构和每个章节涵盖的内容如下：

第一章：引言。本章节简单介绍了文本情感分析的背景，研究意义和对国内外学者针对文本情感分析做出的研究。接着简单介绍本文的主要研究内容，最后介绍论文的组织结构。

第二章：相关技术介绍。该章介绍了本文使用到的一些文本情感分析技术，并对原理进行简单阐述。

第三章：基于情感词典的情感分析。本章介绍了使用情感词典进行情感分析的方法。在该章中我们建立和扩展了基础情感词典，通过实验对比，发现扩展情感词典规模可以有效提高情感分类效果，在对照实验中，规模最大的情感词典准确率达到75.35%，比起最小的情感词典，准确率提高约了6个百分点。

第四章：基于机器学习的情感分析。本章研究了如何使用支持向量机（SVM）进行情感分析，TODO：实验结果

第五章：基于 BERT 模型的情感分析。本章使用自然语言处理最前沿的 BERT 预处理语言模型进行情感分析，TODO：实验结果

第六章：实验设计和成果分析。基于数据集设计实验，分析比较上述方法在数据集上的表现效果，对结果进行分析。

2 相关技术介绍

2.1 文本预处理

2.1.1 分词技术

中文分词技术

3 基于情感词典的情感分析

在情感分析领域，情感词典在早期的研究中有着重要地位，对提升分类效果有着很大帮助。基于情感词典的情感分类技术主要是利用句中每个情感词的情感值，进行加权累加得出整句情感得分，最后根据情感得分来划分情感极性。情感词典技术的一大弊端在于，如果情感词典的规模不够大，对于情感词的覆盖面不够广，对情感分析的帮助有限。本章基于多种情感词典，结合出一个规模相对较大的情感词典，在实验结果上表现良好，证明扩大情感词典规模对提升情感分类效果有着很大帮助。

3.1 情感词典构建

我们利用互联网上的各种开源中文情感字典，结合一些辅助情感字典，再进行针对本文特定数据集，构建出用于本文情感分析的情感字典。

3.1.1 基础情感词典构建

在基础情感字典上，我们选取中文自然语言处理常用的三个情感字典，包括：知网情感字典 (HowNet)^[47]，台湾大学开发的简体中文情感极性词典 (NTUSD)^[48] 和清华大学李军^[49] 开发的中文褒贬义词典。在对三个词典进行交叉去重后，我们发现，有的词典中的积极词汇在别的词典中则被标记为消极。因此我们选用一种根据情感极性和词典权威度的投票筛选来确定这些词的情感极性^[50]，公式如下：

$$\text{Polarity} = \begin{cases} -1, & \sum P(w) \leq -1 \\ 0, & \text{others} \\ +1, & \sum P(w) \geq 1 \end{cases} \quad (3.1)$$

如果是属于 *others* 的情况，则根据词典权威度（HowNet > NTUSD > 中文褒贬义词典）来决定该词的情感极性。最终我们构建出基础情感词典的雏形，共含 11,189 个积极词汇，15,288 个消极词汇。

在上面构建好的词典雏形的基础上，我们继续丰富基础情感词典，使用的词典如

下：

(1) 程度词词典

中文中常在情感词前来增加或削弱该情感词的情感强度，这类词对分析文本情感倾向有着重要意义，因此，我们建立了一种包含六类程度词的词典，部分程度词及其对情感词的加成系数如下表 3.1：

表 3.1 程度词词典及其加成系数表

程度	加成系数	示例	个数
极其 /most	2	绝对、完全、异常	89
很 /very	1.75	不少、出奇、很	88
较 /more	1.5	更、还、比较	46
稍 /ish	1.25	略、挺、有点	42
微 /insufficiently	0.8	不怎么、丝毫、微	12
超 /over	0.6	超、过、过于	30

(2) 否定词词典

否定词其实是一种特殊的程度词，它可以将整个词的情感极性反转，对于否定词的加成系数，取决于情感词前的否定词个数，若是偶数，则为 1，即双重否定为肯定；若是奇数，则为 -1。我们选用了常见的 24 个否定词，包括不是、没、无、非、不曾、绝不等。

(3) 连接词词典

在中文中常有一些可以影响句子整体上的情感的连接词。比如：他虽然很帅，但是人品很差。如果忽略连接词的作用，“很帅”和“很差”的情感倾向相互抵消，整句话就会被判定为中性，但是实际情感更倾向于消极。因此，不能忽视连接词对整句话情感倾向影响，所以我们引入了连接词词典，部分连接词和其加成系数如

下表 3.2:

表 3.2 连接词和其加成系数表

属性	加成系数	示例	个数
让步	1.2	不过、虽然、不料	9
递进	1.5	也、其次、不仅	5
因果	2	但是、况且、然而	10
转折	3	甚至、尤其、居然	3

(4) 停用词词典

在进行分词处理后，每个句子都变成一串词组成的列表，除去文本预处理时已经完成的工作，添加停用词词典能更好的进行文本情感分析，并且提高处理速度。我们设置的停用词词典为哈尔滨工业大学发表的停用词词表。

3.1.2 基础情感词典的扩展

在完成上述工作后，我们发现情感词典的包含的情感词数量已经足够，但仍需要对本文所用的数据集进行针对性添加。因此我们添加了包含网络用语的针对微博语境的情感词典和微博情感词典。

在基础情感词典中已经包含了日常生活中常出现的情感词，但是微博作为社交网络，拥有着大量网络用语，新兴词汇和非标准简称，如果不能有效的判断出这些词汇的情感词属性和情感极性，会影响情感分类的准确性。因此我们添加了开源情感词典：BosonNLP^[51]，该词典包含 114,767 条词汇，极大的丰富了情感词典的内容。

用户在微博上发表的文本往往并非纯文字，而是包含各种表情符号的文本，而这些表情符号也包含着丰富的情感信息^[3]，添加微博表情符号词典能有效帮助我们进行情感分析。我们选用微博上常用的 20 个表情添加到表情符号词典中，其中积极和消极各 10 个，具体表情符号如下表 3.3:

表 3.3 表情符号词典

情感	示例	个数
积极	[可爱]、[太开心]、[嘻嘻]	10
消极	[怒]、[衰]、[骷髅]	10

3.2 基于情感词典的情感分析过程

上一节中，我们提出了如何构建基础情感词典，并针对数据集，有目的性的对基础情感词典进行扩充。下面详细解释如何将构建好的情感词典用于情感分析。主要流程分为三步：1、对文本进行预处理，2、构建情感词典，并设置不同规模的情感词典作为对照组，3、编写代码对数据集进行分析，最后比较训练模型的结果与预先标记好的句子情感极性。流程图见图3.1。文本预处理和情感词典构建的详细过程分别见章节2.1和章节3.1

在计算整句情感极性时，我们设计了公式3.2, 其中 S_{w_i} 为该句中第 i 个词的情感值， p_j 为该词前面连接词 / 程度词的加成系数。若无连接词 / 程度词，则 $\prod_{j=1}^m p_j$ 默认为 1。 d 为该词前面的否定系数，若否定词数量为偶数，则 $\prod_{j=1}^m d_j$ 为 1，否则为 -1。 m 通常取 3，若该词前面词数量不足 3，则视情况选取 (若该词属于句首，则 m 取 0)。最后将这句话中每个情感词乘以加成系数和否定系数后的情感值相加，若是大于 0，则为积极，否则为消极。

$$\text{Polarity}_{total} = \begin{cases} 1, & \sum_{i=1}^n (S_{w_i} \prod_{j=1}^m p_j d_j) > 0 \\ 0, & \sum_{i=1}^n (S_{w_i} \prod_{j=1}^m p_j d_j) \leq 0 \end{cases} \quad (3.2)$$

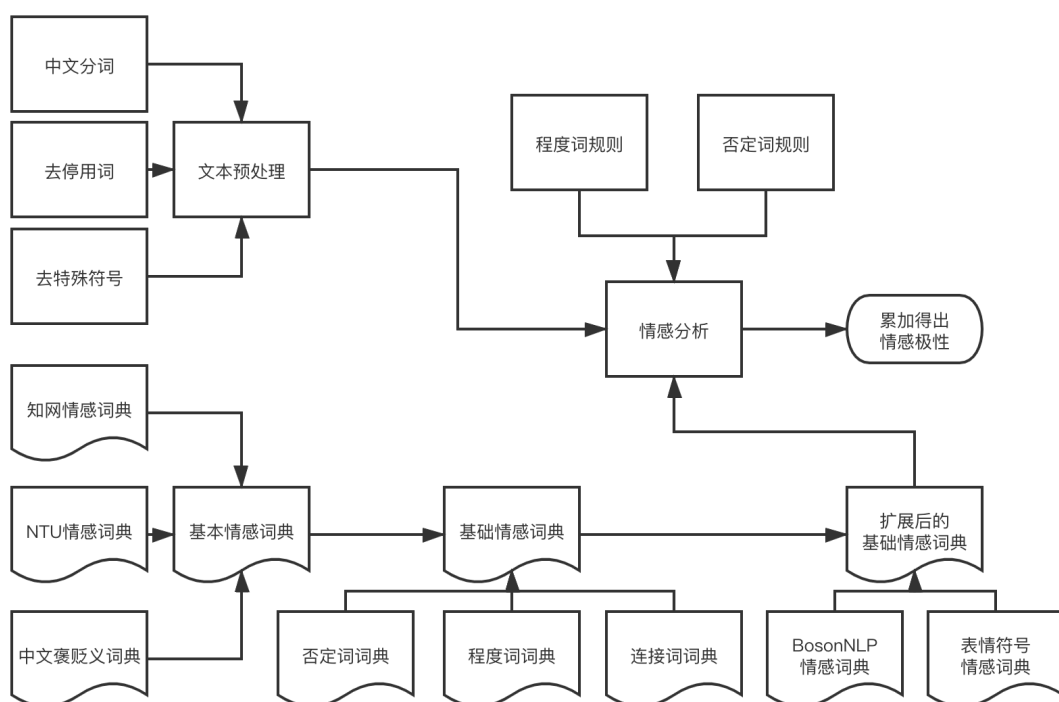


图 3.1 基于情感词典的情感分析过程流程图

3.3 实验结果与分析

本文中使用的数据集为互联网开源的微博数据集，共包含 119,988 条带标注的微博文本数据。其中积极文本和消极文本各占一半。情感词典包括知网情感词典、台湾大学简体中文情感极性词典、中文褒贬义词典、BosonNLP 开源情感词典、哈工大停用词表、自建表情符号情感词典、连接词词典、否定词词典、程度词词典。实验平台运行在 Google Colab(Jupyter Notebook) 上，编程语言为 Python，使用的开源库包括 Jieba, Numpy, Pandas 等。该实验的情感分析测评指标采用国际通用的四个标准，即准确率 (accuracy)、精确率 (precision)、召回率 (recall) 和 F1 值。

本实验的三个对照组分别为不同规模的情感词典，为方便描述，我们将他们标记为 BD, BD-A, BD-AE。分别为基准情感词典 (含知网情感词典，台湾大学简体中文情感极性词典，中文褒贬义词典)，在基准情感词典上加入连接词词典，否定词词典，程

度词词典等，在 BD-A 的基础上加入扩展情感词典（包括 BosonNLP 情感词典，表情符号情感词典）。实验结果如表3.4。

表 3.4 不同情感词典下的实验结果

情感词典	Accuracy	Precision	Recall	F1 值
BD	69.17%	71.58%	66.77%	69.09%
BD-A	69.28%	71.88%	66.67%	69.18%
BD-AE	75.35%	92.46%	58.24%	71.46%

从上表3.4结果看出，BD-AE 有效的改进了情感分类效果，基本在各项指标上均优于其他数据集，验证了扩大情感词典规模在提升实验表现上的有效性。但是 BD-AE 在召回率上表现不佳，即在处理消极文本上仍有可提升空间，也显示出了使用情感词典方法处理情感分析问题上的局限性。该方法只是简单的将句中的每个词的情感值加权相加，根据最终整句的情感值来进行分类，没有考虑到句法，语法上的深层语义。即使继续扩大情感词典规模，可能收益也不会很大。相反，更大规模的情感词典意味着在检索上需要消耗更大的计算资源和时间，得不偿失。

3.4 本章小结

本章介绍了构建中文情感词典及基于基础情感词典进行拓展的过程。首先描述了情感词典在情感分析领域的价值，之后进行了基础情感词典的构建和拓展，然后基于不同规模的情感词典设置对照组进行实验。实验结果表明，扩展后的情感词典一定程度上可以有效提升情感分类的效果。

4 基于机器学习的情感分析

结 论

总结

参 考 文 献

- [1] MOORE G E, et al. Progress in digital integrated electronics[C]//Electron devices meeting: volume 21. [S.l.: s.n.], 1975: 11-13.
- [2] 王安君, 黄凯凯, 陆黎明. 基于 Bert-Condition-CNN 的中文微博立场检测[J]. 计算机系统应用, 2019, 28(11):45-53.
- [3] 周胜臣, 瞿文婷, 石英子, 等. 中文微博情感分析研究综述[J]. 计算机应用与软件, 2013, 30(3):161-164.
- [4] 崔志刚. 基于电商网站商品评论数据的用户情感分析[D]. 北京: 北京交通大学, 2014.
- [5] 杨立公, 朱俭, 汤世平. 文本情感分析综述[J]. 计算机应用, 2013, 33(06):1574-1607.
- [6] 杨小平, 张中夏, 王良, 等. 基于 Word2Vec 的情感词典自动构建与优化[J]. 计算机科学, 2017, 44(1):42-47.
- [7] 朱晓霞, 宋嘉欣, 张晓缇. 基于主题挖掘技术的文本情感分析综述[J]. 情报理论与实践, 2019, 42(11):156-163.
- [8] WIEBE J, BRUCE R, O' HARA T P. Development and use of a gold-standard data set for subjectivity classifications[C]//Proceedings of the 37th annual meeting of the Association for Computational Linguistics. [S.l.: s.n.], 1999: 246-253.
- [9] RILOFF E, WIEBE J. Learning extraction patterns for subjective expressions[C]//Proceedings of the 2003 conference on Empirical methods in natural language processing. [S.l.: s.n.], 2003: 105-112.

- [10] WILSON T, WIEBE J, HOFFMANN P. Recognizing contextual polarity in phrase-level sentiment analysis[C]//Proceedings of human language technology conference and conference on empirical methods in natural language processing. [S.l.: s.n.], 2005: 347-354.
- [11] TURNEY P D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews[C/OL]//ACL ' 02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. USA: Association for Computational Linguistics, 2002: 417-424. <https://doi.org/10.3115/1073083.1073153>.
- [12] HAMOUDA S B, AKAICHI J. Social networks' text mining for sentiment classification: The case of facebook' statuses updates in the 'arabic spring' era[J]. International Journal Application or Innovation in Engineering and Management, 2013, 2(5):470-478.
- [13] HU M, LIU B. Mining and summarizing customer reviews[C]//Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. [S.l.: s.n.], 2004: 168-177.
- [14] 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. [S.l.]: Association for Computational Linguistics, 2002: 79-86.
- [15] KIM S M, HOVY E. Determining the sentiment of opinions[C]//Proceedings of the 20th international conference on Computational Linguistics. [S.l.]: Association for Computational Linguistics, 2004: 1367.
- [16] GODBOLE N, SRINIVASAIAH M, SKIENA S. Large-scale sentiment analysis for news and blogs.[J]. Icwsm, 2007, 7(21):219-222.

- [17] LIN C, HE Y. Joint sentiment/topic model for sentiment analysis[C/OL]//CIKM ' 09: Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York, NY, USA: Association for Computing Machinery, 2009: 375 – 384. <https://doi.org/10.1145/1645953.1646003>.
- [18] HINTON G E, et al. Learning distributed representations of concepts[C]//Proceedings of the eighth annual conference of the cognitive science society: volume 1. [S.l.]: Amherst, MA, 1986: 12.
- [19] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [20] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [21] PENNINGTON J, SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]//Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). [S.l.: s.n.], 2014: 1532-1543.
- [22] JOULIN A, GRAVE E, BOJANOWSKI P, et al. Fasttext. zip: Compressing text classification models[J]. arXiv preprint arXiv:1612.03651, 2016.
- [23] SOCHER R, HUVAL B, MANNING C D, et al. Semantic compositionality through recursive matrix-vector spaces[C]//Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. [S.l.]: Association for Computational Linguistics, 2012: 1201-1211.
- [24] XU J, CHEN D, QIU X, et al. Cached long short-term memory neural networks for document-level sentiment classification[J]. arXiv preprint arXiv:1610.04989, 2016.

- [25] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [26] ZHOU X, WAN X, XIAO J. Attention-based lstm network for cross-lingual sentiment classification[C]//Proceedings of the 2016 conference on empirical methods in natural language processing. [S.l.: s.n.], 2016: 247-256.
- [27] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[J]. arXiv preprint arXiv:1802.05365, 2018.
- [28] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[J]. URL [https://s3-us-west-2. amazonaws. com/openai-assets/researchcovers/languageunsupervised/language understanding paper. pdf](https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language%20understanding%20paper.pdf), 2018.
- [29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in neural information processing systems. [S.l.: s.n.], 2017: 5998-6008.
- [30] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [31] 姚天昉, 彭思崴. 汉语主客观文本分类方法的研究[C]//第三届全国信息检索与内容安全学术会议论文集. 苏州: 中国中文信息学会, 2007: 117-123.
- [32] 叶强, 张紫琼, 罗振雄. 面向互联网评论情感分析的中文主观性自动判别方法研究[J]. 信息系统学报, 2007, 1(1):79-91.
- [33] 张博. 基于 SVM 的中文观点句抽取[D]. 北京: 北京邮电大学, 2011.
- [34] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1):16-22.

- [35] 柳位平, 朱艳辉, 栗春亮, 等. 中文基础情感词词典构建方法研究[J]. 计算机应用, 2009, 29(10):2875-2877.
- [36] 李钝, 曹付元, 曹元大, 等. 基于短语模式的文本情感分类研究[J]. 计算机科学, 2008 (04):132-134.
- [37] 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):88-94,108.
- [38] 徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类[J]. 中文信息学报, 2007, 21(6):95-100.
- [39] 韦航, 王永恒. 基于主题的中文微博情感分析[J]. 计算机工程, 2015, 41(9):238-244.
- [40] 曾江峰. 基于深度学习的文本情感计算研究[D]. [出版地不详]: 华中科技大学, 2019.
- [41] 李然. 基于深度学习的短文本情感倾向性研究[D]. [出版地不详]: 北京理工大学, 2015.
- [42] 胡朝举, 赵晓伟. 基于词向量技术和混合神经网络的情感分析[J]. 计算机应用研究, 2018, 35(12):3556-3559,3574.
- [43] 张仰森, 郑佳, 黄改娟, 等. 基于双重注意力模型的微博情感分析方法[J]. 清华大学学报（自然科学版）, 2018, 58(2):122-130.
- [44] 赵亚欧, 张家重, 李贻斌, 等. 融合 ELMo 和多尺度卷积神经网络的情感分析[J]. 计算机应用:1-9.
- [45] 蔡国永, 林强, 任凯琪. 基于域对抗网络和 BERT 的跨领域文本情感分析[J]. 山东大学学报 (工学版), 2020(01):1-7+20.
- [46] 带情感标注的中文微博数据集[EB/OL]. https://github.com/SophonPlus/ChineseNlpCorpus/blob/master/datasets/weibo_senti_100k/intro.ipynb.

- [47] 知网情感词典[EB/OL]. <http://www.keenage.com/>.
- [48] 台湾大学简体中文情感极性词典[EB/OL]. <http://nlg.csie.ntu.edu.tw/download.php>.
- [49] LI J, SUN M. Experimental study on sentiment classification of chinese review using machine learning techniques[C]//2007 International Conference on Natural Language Processing and Knowledge Engineering. [S.l.]: IEEE, 2007: 393-400.
- [50] 熊德兰, 程菊明, 田胜利. 基于 HowNet 的句子褒贬倾向性研究[J]. 计算机工程与应用, 2008(22):143-145.
- [51] BosonNLP 情感词典[EB/OL]. <https://bosonnlp.com/dev/resource>.

致 谢

致谢