中文微博情感倾向性分析特征工程*

李泽魁[1]，赵妍妍[2]，秦兵[1]，刘挺[1]

[1]哈尔滨工业大学计算机科学与技术学院信息检索研究中心，哈尔滨，150001

[2]哈尔滨工业大学机电学院媒体系，哈尔滨，150001

E-mail: {zkli, yyzhao, qinb, tliu}@ir.hit.edu.cn

摘　要：情感倾向性分析是情感分析的重要组成部分，是一种按照情感倾向对文本进行分类的任务。微博，与传统的评论文本相比更加口语化与符号化，因此对微博进行情感倾向性分析是一个非常有挑战性的任务。基于机器学习的方法是情感倾向性分析最经典的算法，核心是要进行特征的分析和选择，例如词袋特征等。然而，由于中文语言的独特性，前人很多有效的特征都是语言相关的，将其直接用于中文微博效果不佳。在中文微博语料上，还没有学者进行细致的特征工程建设。基于此，本文综合国内外诸多特征，并考虑到中文的独特性，对中文微博的褒贬中倾向性判别特征工程的词、词组、数值和句法特征分别进行了研究，并提出了基于词典规则的情感评分的新特征。最后经过大量实验与分析，得出了可靠的特征组合。实验结果表明，本文的方法能够明显提高情感倾向性分析的结果。

关键词：情感倾向性分析；中文微博；特征工程

Feature Engineering for Chinese Micro-blog Sentiment Classification

Zekui Li[1], Yanyan Zhao[2], Bing Qin[1], Ting Liu[1]

[1]Research Center for Social Computing and Information Retrieval of Computer Science and Technology School, Harbin Institute of Technology, Harbin 150001

[2] Department of Media Technology and Art, Harbin Institute of Technology, Harbin 150001

E-mail: {zkli, yyzhao, qinb, tliu}@ir.hit.edu.cn

Abstract：Sentiment classification, a basic sentiment analysis task, aims to classify a sentiment sentence into positive, negative and neutral. Sentiment analysis on microblog is challenging, which is different from it on common product reviews, due to the characteristics of microblog. Many previous works used machine learning based approaches to solve this task, the core of which is to try and select useful features, for instance, “bag of words”. However, these proposed features may not be suitable for Chinese due to linguistic differences. What is more, there is no feature engineering for Chinese microblog in details. In this paper, we do some feature engineering for Chinese microblog sentiment classification, from words, phrases, numbers, syntactic features, and new feature named dictionary-rule based sentiment score, in order to make a better performance beyond the baseline. At last, we obtain reliable feature set through a large number of experiments and analysis. Our approach significantly improves the results of sentiment classification.

Key words：Sentiment classification; Chinese micro-blog; Feature engineering

0　前言及相关研究

微博，是人们分享新鲜事、表达自己的看法和维护自己的朋友圈的一套思想交流工具。微博存在着大量的用户，他们在微博上有时对国家时事发表自己的见解，有时对某些新闻提出自己的喜恶，或者对许多产品写下自己的感受等。字里行间流露出的作者的真情实感，背后隐藏的包括舆论监控、产品口碑等等应用以及在此之上的政策改革、产品市场定位以及产品用户体验等方面的价值可以说是不可估量。

* **基金项目**：本课题受到国家自然科学基金重点项目(61133012)、国家自然科学青年基金项目(61300113)及国家自然科学基金面上项目(61273321) 资助。

作者简介：李泽魁(1992—)，男，硕士，主要研究方向为情感分析。

情感分析又称意见挖掘，按照处理文本的粒度不同可以分为词语级、短语级、句子级和篇章级等[1]，通过结合文本挖掘、信息抽取、机器学习、自然语言处理等文本处理技术对主观性文本进行分析、处理和归纳。情感分析近几年持续成为自然语言处理领域研究的热点问题，可以广泛应用到很多的自然语言处理问题中，如信息抽取[1]、自动问答[2]、产品口碑等。

倾向性分析作为文本情感分析的重要组成部分，该领域受到了国内外很多学者的关注。TREC 评测[1]、NTCIR 评测[2]以及 COAE 评测[3]推动和加速了倾向性分析研究的发展。而随着文本倾向性分析研究的深入，国内外的研究发现领域知识和上下文语境(Context)对倾向性判别至关重要，吸引了国内外研究学者的广泛关注，并纷纷开展诸如领域观点词表构建、跨领域倾向文本分类等研究[3]。在这一研究领域，主要存在两种方法：

（1）基于情感词典及规则的无监督学习方法

情感词典，是按照不同的情感倾向对情感词分类后的词典。基于情感词典及规则方法，顾名思义，即按照人工构建的情感词典和指定的相关规则来进行情感倾向性判定的方法[4]。

对于结构简单的句子，此方法会取得较好的分类效果。但是在句式复杂的某些中文微博面前，基于情感词典及规则的方法显得很乏力；与此同时，人工构建情感词典在实际操作中，受到成本和规模的限制，不适于广泛推广[5,6]。如果情感词典的规模小，则会遗漏很多情感词，无法识别文本的情感倾向，特别是对于微博这类短文本，更不易命中情感词；如果情感词典的质量不高，也会造成情感分析结果的错误。

（2）基于有监督的机器学习方法

由于情感倾向性判别的识别结果是褒义、贬义等类别，因此该任务可以采用机器学习的方法，作为分类任务完成[7]。在这种方法中，首先，一般以词汇（例如一元特征）或数字（例如情感词出现次数）等作为特征，对标注好的训练语料进行特征提取；然后结合学习模型，典型的例如朴素贝叶斯（Naive Bayes）、最大熵（Maximum Entropy）、支持向量机（Maximum Entropy），对语料进行模型训练；最后通过训练出的分类器对测试语料进行分类[8]。

但是如果缺乏充足的训练语料，有监督的机器学习方法的分类效果可能也不尽人意。对于微博这种数量庞大的互联网文本，采用人工只能标注很少的微博文本[9]，其适用领域与规模受到限制。此外，目前在英文语料上面有针对评论、微博情感倾向性分析任务的特征工程建设[10,11]，但是还没有学者在中文微博语料上进行详细的特征工程建设。

面向微博的情感倾向性分析，是情感分析的热门研究课题。基于特征提取的机器学习方法是当前倾向性判别的主流方法[7,8]，Abbasi 等人的实验[10]中对论坛帖子语料进行了情感分类任务的特征工程建设。然而，微博作为一种短文本，由于其自身特点（例如口语化、符号化等）[12]，其处理难度要高于一般句子级情感分析。Mohammad 等人针对英文微博（twitter）的特点，对微博的情感倾向性分析任务进行了特征工程建设[11]。然而，当前还没有学者在中文微博语料上进行细致的特征工程建设。

本文的主要任务是对中文微博进行褒贬中倾向性的分析。针对当前研究存在的问题，我们为了寻找最优的情感倾向性分析的特征提取方法，提出了“词特征”、“词组特征”、“数值特征”和“句法特征”四大类特征提取模板，同时提出了融合基于多词典规则的情感值的计算结果的方法，将其作为一维特征。通过对比实验，最后得到了一套可靠的特征组合，并且证明了“词典规则评分”特征的有效性，使褒贬中三类的分类效果有了进一步的提高。

---

[1] http://trec.nist.gov/

[2] http://research.nii.ac.jp/ntcir/index-en.html

[3] http://www.liip.cn/CCIR2014/pc.html

1 特征工程设计

基于有监督学习的机器学习方法，方法是利用已有资源对测试语料进行特征提取，然后供后续的机器学习方法训练模型。对于中文微博的情感倾向性分析，如果使用有监督的机器学习方法，需要对语料进行特征提取。国外学者在特征提取方面做了很多研究，但是由于语言的差异性，将其研究成果直接用于中文微博效果不佳。对于同样一组训练与测试语料，究竟哪些特征可以提取，并且他们可以对情感分析效果提升做贡献，国内还没有一个详细的总结性实验。

基于此，本节综合国内外研究中选取的诸多特征，并考虑到中文的独特性，对中文微博的褒贬中倾向性判别特征工程的词、词组、数值和句法特征分别进行了研究（详见 1.1 节）。同时，由于基于词典规则的判定方法在一些情况下也可以对微博进行正确的情感倾向性判定，本文提出了“基于词典规则的情感评分”的新特征（详见 1.2 节）。

1.1 特征分类

理论上可行的特征有很多，通过分析后我们发现这些特征可以大致分四类：分别是词特征、词组特征、数值特征和句法特征。下表更加详细地介绍了每一大类下面细分哪些特征。具体特征抽取样例见附表 1。

表 1 倾向性分析的特征

Tab.1 Sentiment classification features extraction

特征分类	子类别	特征描述	ID
词特征	单词	#一元特征 #二元特征 #HashTag #书名标题 #情感词	S_1
	词性	#POS Unigram #POS Bigram	S_2
词组特征	情感词搭配	#主干连词+情感词 #否定词+情感词，长距离否定词+情感词 #动态词+上下文 #强度副词+情感词，情感词+感叹词 #疑问词+情感词	S_3
数值特征	词频统计	#情感词词频差统计 #主干连词、否定词、动态词、强度副词、感叹词和疑问词的出现频率 #HashTag、书名标题数量	S_4
	词性统计	#几项具有代表性的 POS 的出现频率	S_5
	句法统计	#几项具有代表性的句法关系的出现频率	S_6
	基于词典规则的情感评分	#基于主干句分析与多词典联合的方法对文本进行情感评分	S_7
句法特征	频率统计	#情感词相关的句法关系的 Unigram	S_8

- **词特征**

  词特征分为单词特征和词性特征两类。

1） 单词特征

可以看作词袋模型的推广，其内容包括一元词特征、情感词等。国外已经有学者证明一元特征模型具有良好的分类效果[8, 13]。同时情感词、书名标题等单词特征同样作为一种词袋模型，据此分析它理论上也可以对判别结果有所提高。

2） 词性特征

即对文本中出现的词的词性进行统计。在 A. Agarwal 等人的面向 Twitter 的情感倾向性分析实验[12]中，曾经对情感词和非情感词的词性出现频率进行过统计，在实验中，我们尝试对每个词的 POS 作为词袋模型，然后通过实验比较其是否有效。

- **词组特征**

在对一个文本进行情感分类的时候，情感词起着很大的作用。既然一元特征可以对情感分类问题有很好的效果，那么本文按照它的思想，将情感词与它的修饰词做成词袋模型加入至训练文件中，也可能对分类效果有所提升。

- **数值特征**

数值特征分为词频统计、词性统计、句法统计和基于词典规则的情感评分。

1） 词频统计

在国内外诸多研究中[8,12,13]，词频统计一直保持着很高的应用率。在特征抽取过程中，除了常见的“否定词”、“强度副词”等的出现频率[11]，还按照实验涉及的词典特点加入了“情感词频统计差”、“动态词”等的出现频率。

2） 词性统计

Agarwal 等人和 Kouloumpis E 等人对 Twitter 进行情感分类时，对词性的频率进行了统计[12,14]，本文为了测试在中文中是否有效，也将其加入到特征工程中。

3） 句法统计

在 Abbasi 等人针对论坛帖子的情感分析实验[10]中，涉及到文本的文法特征。我们从那里获取灵感：是否可以从文本的句法统计中获得有助于分类结果的特征。于是我们将句法统计作为一维特征加入到特征工程中。

4） 基于词典规则的情感评分

基于词典的无监督学习方法作为情感分析的一种简单而基础的方法，在处理某些简单的短文本的情感倾向性的时候有着不俗的表现[15]，种种迹象表明在特征提取时应该将情感规则的倾向性评分作为一维特征加入到特征提取模板中。另一方面，基于词典规则的情感评分算法的优劣也直接决定着判别结果的好坏，本文在 1.2 节详细地描述了其算法流程。

- **句法特征**

利用句法网络关系来对文本情感进行分析，国外早有先例[16]。据此我们受到启发，将文本中与情感词有关的句法特征整理后写入特征。

1.2 基于词典规则的倾向性评分

为了尽可能地使词典规则评分这一维特征在情感倾向性判别时发挥更大的作用，本文提出了面向中文微博的词典规则情感评分的优化计算方法，主要集中在“主干句分析”与“多词典联合”两个方面。具体计算过程描述详见以下小节。

1.2.1 主干句分析

通过对 COAE2014 任务四的测试语料进行抽样统计分析，我们发现语料中微博有 23.5% 的评论包含明显的主干成分，即句子中包含明显的总结性连词或转折性连词，而且主干句的极性往往与整体极性比例一致，测试语料中高达 96.2%。

上面的数据表明，主干句分析句子极性很有必要。于是本文设计了一套主干句分析流程，即对于一条待测试语句，首先对其进行预处理；随后按照标点符号对其进行分句，得到子句集合 Clauses[1~N]，N 为子句数目；然后利用哈工大语言技术平台[17]微博分词版对其进行分词处理；如果分句中存在主干成分，整句情感值等于主干句情感值；反之，整句情感值等

于各个子句的情感值加权相加。

1.2.2 多词典辅助判别法

传统的词典评分方法为简单的褒贬词频的统计，然后求差。这种模型算法简单，但是伴随而来的是较低的准确率。为了使评分算法更具普适性，我们对于常见的但是原先算法无法正确判断的句式进行了归纳总结，并且提出了解决办法，如表 2 所示。

表 2 词典评分句式总结

Tab.2 Summary of dictionary based sentiment analysis

ID	句子示例	特点	判定依据	权值相加规则
1	这道菜好吃。	情感词	情感词词典	情感词权值
2	这道菜真好吃。	强程度副词+情感词	强程度副词词表	情感词权值*3
3	这道菜还算好吃。	弱程度副词+情感词	弱程度副词词表	情感词权值*2
4	这道菜不好吃。	否定词+情感词	否定词词表	情感词权值变为相反数
5	我不认为这道菜好吃。	长距离否定+情感词	句法关系	情感词权值变为相反数
6	这道菜哪里好吃？	疑问词+情感词	疑问词词表	情感词权值为 0
7	《好吃的菜》是一本书。	书名标题含有情感词	书名号	情感词权值为 0
8	今日话题 #好吃的菜#	话题中含有情感词	“#”符号对	情感词权值为 0

从表 2 的分析得知，使用多词典辅助判别法理论上可以解决很多单一情感词词典无法判别的情况。对于一条待测试语句，先后判定情感词以及情感词前后一定窗口大小的强度副词、否定词、长距离否定、疑问词、书名标题标记等，然后根据不同的权值规则进行情感值的计算。根据相应规则，我们归纳出该模块的计算公式，即：

$$DictBasedScore = \sum_{i=1}^{word_{count}} (B_i N_i E_i Polarity(Word_i)) \quad (1)$$

公式（1）中，$Polarity(Word_i)$代表情感词对应的基础值，可以是±1，代表情感词为褒义和贬义；E_i 代表强度副词权值，取值范围是 1,2 或 3，对应含义为上下文无程度副词、存在弱程度副词和强程度副词；N_i 代表否定修饰词权值，取值范围是±1，代表上下文中无否定词和存在否定转移；而 B_i 代表禁止词权值，如果上下文中存在表 2 中的类似书名、话题等对情感倾向判定无关的词，那么权值取 0，否则取 1。

2 实验设置

2.1 数据来源

为了使训练语料更具代表性、多样性，本文采取多个来源的语料融合的方法，褒贬中三类倾向的语料共计 5278 句。语料组成如下：

表 3 语料数据来源

Tab.3 Corpus data sources

ID	语料来源	语料标注者	倾向分类	规模
1	COAE2014 任务 4 训练语料	COAE2014 评测举办方	褒贬	2278
2	百度百科	哈工大社会计算与信息检索研究中心	中	1000
3	新浪微博	哈工大社会计算与信息检索研究中心	褒贬中	2000

2.2 词典资源构建

词典资源是情感分析任务的基础，在各个情感分析任务中起着至关重要的作用。本文使用了哈工大社会计算与信息检索研究中心的情感词典、修饰词典和连词词典资源。

（1）情感词典，分为一般情感词典与动态情感词典。其中褒义和贬义情感词是针对不同评价对象情感极性一致的情感词语，例如“高兴”、“悲伤”等；动态情感词是针对不同评价对象情感极性不一致的情感词语，例如“高”、“上涨”等。

（2）修饰词典，即与情感词搭配出现的词语，其中包含否定词、程度副词、疑问词和感叹助词。顾名思义，分别表示可以对情感词语的极性进行翻转、加强和减弱的修饰性词语。

（3）连词词典，其中包含句间的总结性连词和转折关系连词，主要用途是主干句判断。总结性连词是可以对句子的语义总结和概括的连词，例如“总之”、“所以”等；转折关系连词是可以使相邻文本的语义发生翻转的连词，例如“但是”，“然而”等。

表 4 词典资源

Tab.4 Dictionary resources

词典资源类型	词典名称	数据规模	样例
情感词典	褒义情感词	9909	高兴、美丽
	贬义情感词	11781	悲伤、差
	歧义情感词	183	高、上涨
修饰词典	否定词	117	不、绝非
	强程度副词	180	很、非常
	弱程度副词	45	有些、稍微
	疑问词	13	为什么、哪里
	感叹词	15	啊、呀
连词词典	总结性连词	19	总之、因此
	转折关系连词	20	但、但是

2.3 支持向量机简介

本文在情感倾向性判别中使用的分类器是支持向量机(SVM,Support Vector Machine)，是由 Vapnik 提出的一种非常有潜力的学习技术[18]，同时也是在统计学习理论基础上构造的一种通用学习机器，广泛的应用在人脸检测、验证和识别、文字识别、图像处理等领域。

支持向量机是有指导的机器学习，即模型在训练过程中，需要人工指定训练集的类别。其分类过程包括语料的预处理、特征提取以及模型的训练与预测，如下图所示。

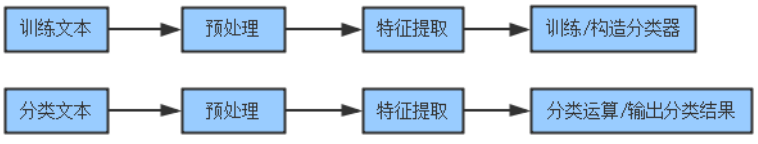

图 1 支持向量机处理流程图

Fig.1 Support vector machine flow chart

获取文本后，首先对文本进行分词处理。本实验利用哈工大语言技术平台[17]微博分词版对其进行分词处理。第二步，根据实验要求提取特征候选。第三步中，本文使用的是台湾大学林智仁(Chih-Jen Lin)教授等开发的支持向量机算法库 libsvm[4]进行模型的训练与分类。

[4] http://www.csie.ntu.edu.tw/~cjlin/

2.4 评价方法

本实验采用了交叉验证，我们关心的是在同样的测试语料下系统的分类正确性。为了直观地查看系统的分类性能，本文采用了精确率(Accuracy)来评价分类结果的整体效果。

精确率(Accuracy)为检出的相关文档数与文档数的比值，即：

$$Accuracy = \frac{|\text{检出的文档数}|}{|\text{文档数}|} \quad (2)$$

3 实验结果

在这一小节中，实验目的是对测试语料进行褒贬中的情感倾向性判定。实验中使用了在表 1 中列举的条目进行特征提取，并通过实验以及对结果的分析得出最优的特征组合。本文采用了两种方法来直观地查看特征组合是否有效（详见 3.1 节）。随后对结果的稳定性在不同领域的语料上进行了测试（详见 3.2 节）。

3.1 特征分析

实验中，我们对表 1 中所列举的所有特征组合进行分析，主要设计了两组实验，来尝试是否会得到更好的效果。第一组是在所有特征（All Features）基础上逐个删减特征组合，如果删除特征后精确率下降，说明该特征组合有效，见表 5；第二组是在基准特征（Baseline）基础上逐个增加特征组合，如果特征加入后精确率上升，说明该特征组合有效，见表 6。

表 5 所有特征组合逐个删减之间的比较

Tab.5 Comparison between different feature sets by deleting individually

ID	Features	Accuracy(%)	Change(%)
0	All Features	85.3543	——
1	All - S_1(单词特征)	76.0136	-9.3407
2	All - S_2(词性特征)	**85.5438**	**+0.1895**
3	All - S_3(词组特征)	85.2406	-0.1137
4	All - S_4(词频特征)	84.0849	-1.2694
5	All - S_5(词性频率特征)	85.2975	-0.0568
6	All - S_6(句法频率特征)	84.5775	-0.7768
7	All - S_7(词典规则评分)	84.1417	-1.2126
8	All - S_8(句法特征)	85.2217	-0.1326

通过观察表 5 的实验结果，我们不难发现，有几个特征组删除后效果下降，表明他们在提升结果精确率上有一定的效果。它们分别是单词特征（S_1 删除后 Acc.下降了 9.3407%）、词频特征（S_4 删除后 Acc.下降 1.2694%）和基于词典规则评分（S_7 删除后 Acc.下降 1.2126%）。单词特征是一元特征的扩展，包括了二元特征和情感词特征等，实验结果进一步说明了该方法的有效性；词频特征是统计各种单词在微博中的出现频率的一个特征组合，包括情感词、否定词、强度副词等特征，实验结果证明了它们对结果的正确判断有一定的效果；而基于词典规则的评分，是本次实验基于词典规则与基于机器学习两种分类方法的融合点，在正确的计算方法的基础上也对实验结果有所帮助。

而在表 5 所示的结果中，有几个特征组合删除后没有达到预期的效果，仅仅小幅度影响了 All Features 组合的性能，例如词组特征、词性频率统计、句法频率特征和句法特征（Acc.降低在 1%以下）。它们的共同点是特征不是由单词直接组成，而是词组、词性或句法等与

单词有隐含关系的特征，所以它们对微博情感分类结果有提升但不明显就理所应当了。

同时我们也发现某些特征删除后起了些性能提高，说明该组特征可能对结果有负作用，具有代表性的是词性特征（S_2 删除后 Acc.提高了 0.1895%）。

表 6 基准特征逐个增加特征之间的比较

Tab.6 Comparison between different feature sets by adding features from baseline

ID	Features	Accuracy (%)	Change(%)
0	Unigram Baseline	79.6703	——
1	+ S_1(单词特征)	81.9250	+2.2547
2	+ S_2(词性特征)	**79.0072**	**-0.6631**
3	+ S_3(词组特征)	80.2766	+0.6063
4	+ S_4(词频特征)	84.3691	+4.6988
5	+ S_5(词性频率特征)	80.5229	+0.8526
6	+ S_6(句法频率特征)	79.8598	+0.1895
7	+ S_7(词典规则评分)	84.5207	+4.8504
8	+ S_8(句法特征)	80.3524	+ 0.6821

在选取实验 Baseline 方面，本文通过调研，发现国内外诸多学者在研究微博的情感倾向性分析时广泛采用一元特征提取来训练模型[12,19]，B. Pang 等人对电影评论的情感倾向性分析实验[8]中也证明了一元特征具有出众的效果。在本实验中使用了一元特征提取的方法作为实验的 Baseline。

表 6 的实验结果中，由于是在 Baseline 上增加特征，所以精确率提升越大，表明对应的特征组合越有效。与表 5 的结果几乎相同，在表 6 中 S_1(单词特征)、S_4(词频特征)和 S_7(词典规则评分)对性能的提高有明显的效果，而 S_2(词性特征)的加入会造成对分类结果的干扰。其余特征组合对 Baseline 性能的提高效果不显著。

3.2 结果稳定性验证

由于语料是由 2.1 中所描述的三部分组成，每个部分的语料在规模、领域和平均长度等方面都或多或少地存在差异。为了验证结果的稳定性，将语料按照领域的不同分为两部分，加上所有语料共三组，分别对表 6 的实验进行了交叉验证，得出的结果如下图。

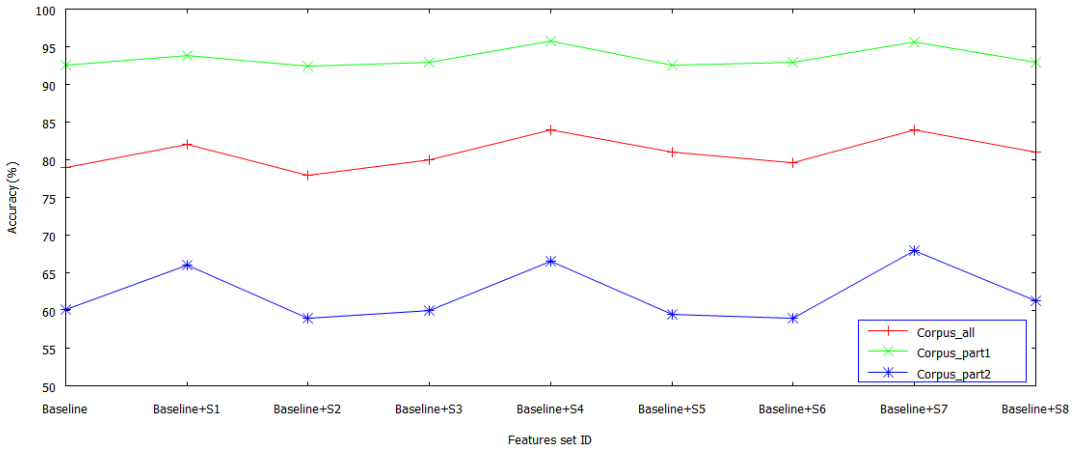

图 2 不同语料库与特征组合下的精确率

Fig.2 Accuracy under different corpus and feature sets

其中语料组成如下表：

表 7 不同测试语料的组合

Tab.7 Different sets of testing corpus

测试测试语料 ID	语料来源	倾向分类	规模
Corpus_all	Corpus_part1+ Corpus_part2	褒贬中	5278
Corpus_part1	COAE2014 任务 4 训练语料+百度百科	褒贬中	3278
Corpus_part2	人工标注的新浪微博	褒贬中	2000

图 2 中从三组语料的实验结果可以看出，三条折线同时在 S_1、S_4和 S_7组合加入后有明显上升趋势，意味着他们对结果具有一定程度的影响，而且其折线的大致走向也基本吻合。可以说明在不同语料下本文的实验结果是稳定的。

4 结论及展望

本文通过对中文微博的词特征、词组特征、数值特征、句法特征以及我们提出的词典规则情感评分特征进行调研与实验分析，初步完成了其情感倾向性判别特征工程的探究。通过分析实验结果，得出了一些可能有利于实验效果提升的特征组合，例如单词特征、词频特征和基于词典规则评分的特征等；同时也得出了某些可能对实验效果削减的特征组合，例如词性特征等。最终我们得出了一套可靠的特征组合，与一元特征的机器学习模型相比，可以更有效地对中文微博的情感倾向进行正确判断。在下一步的研究中，我们将探索新方法，以寻求更有效的特征组合，进一步提高情感倾向性判别的效果。

参 考 文 献

[1] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848.

[2] Hu M, Liu B. Opinion extraction and summarization on the web[C]//AAAI. 2006, 7: 1621-1624.

[3] Liu B. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167.

[4] Ku L W, Wu T H, Lee L Y, et al. Construction of an evaluation corpus for opinion extraction[J]. Proc. of the Fifth NTCIR Wksp. on Evaluation of Information Access Technologies: Information Retrieval, Question Answering, and Cross-Lingual Information Access, 2005: 513-520.

[5] Strapparava C, Mihalcea R. Learning to identify emotions in text[C]//Proceedings of the 2008 ACM symposium on Applied computing. ACM, 2008: 1556-1560.

[6] Neviarouskaya A, Prendinger H, Ishizuka M. SentiFul: A lexicon for sentiment analysis[J]. Affective Computing, IEEE Transactions on, 2011, 2(1): 22-36.

[7] Pang B, Lee L. Opinion mining and sentiment analysis[J]. Foundations and trends in information retrieval, 2008, 2(1-2): 1-135.

[8] 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.

[9] Jiang L, Yu M, Zhou M, et al. Target-dependent twitter sentiment classification[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 151-160.

[10] Abbasi A, Chen H, Salem A. Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums[J]. ACM Transactions on Information Systems (TOIS), 2008, 26(3): 12.

[11] Mohammad S M, Kiritchenko S, Zhu X. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets[J]. arXiv preprint arXiv:1308.6242, 2013.

[12] Agarwal A, Xie B, Vovsha I, et al. Sentiment analysis of twitter data[C]//Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011: 30-38.

[13] Bravo-Marquez F, Mendoza M, Poblete B. Combining strengths, emotions and polarities for boosting Twitter sentiment analysis[C]//Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining. ACM, 2013: 2.

[14] Kouloumpis E, Wilson T, Moore J. Twitter sentiment analysis: The good the bad and the omg![J]. ICWSM, 2011, 11: 538-541.

[15] Cambria E, Schuller B, Xia Y, et al. New avenues in opinion mining and sentiment analysis[J]. 2013.

[16] Olsher D J. Full spectrum opinion mining: Integrating domain, syntactic and lexical knowledge[C]//Data Mining Workshops (ICDMW), 2012 IEEE 12th International Conference on. IEEE, 2012: 693-700.

[17] Che W, Li Z, Liu T. Ltp: A chinese language technology platform[C]//Proceedings of the 23rd International Conference on Computational Linguistics: Demonstrations. Association for Computational Linguistics, 2010: 13-16.

[18] Vapnik V. The nature of statistical learning theory[M]. springer, 2000.

[19] Pak A, Paroubek P. Twitter as a Corpus for Sentiment Analysis and Opinion Mining[C]//LREC. 2010.

附表 1

部分特征抽取示例：

ID	分类	组	特征	句子例子	特征例子
1	词特征	S_1	一元特征	#这个牌子不错。	这个 牌子 不错 。
2			二元特征	#这个牌子不错。	这个牌子 牌子不错 不错。
3			情感词	#这个牌子不错。 #今天的天儿，=3=。	不错 =3=
4			书名标题	#《百年孤独》看了没？ #【吉利的鸿鹄之志】	百年孤独 吉利的鸿鹄之志
5			Hashtag	#今日话题#开心的一天#	开心的一天
6		S_2	一元 POS 特征	#这个牌子不错。	r n a wp
7			二元 POS 特征	#这个牌子不错。	rn na awp
8	词组特征	S_3	主干连词+情感词	#这面不好吃，但是便宜。	但是_便宜
9			否定词+情感词	#这碗面不好吃。	不_好吃
10			长距离否定词 +情感词	#我不认为这是好主意。 #绿茶没有红茶好喝。	不_好 没有_好
11			动态词+上下文	#这点钱还买车，搞笑。 #人民喜迎油价上涨。	搞笑_车 油价_涨
12			强度副词+情感词	#这款车很给力！ #这款车动力稍微不足。	太_给力 稍微_不足
13			疑问词+情感词	#算什么好车。 #相机牌子哪个好。	什么_好 哪个_好
14			情感词+感叹词	#这款车不错啊。 #这款车，好评!!	不错_啊 好评_!!
15	数值特征	S_4	情感词词频统计	#狂野中不乏细腻，美！	1 (解释：1=2-1)
16			否定词数目	#这碗面不好吃。	1 (解释：不)
17			动态词数目	#这点钱还买车，搞笑。 #人民喜迎油价上涨。	1 (解释：搞笑) 1 (解释：上涨)
18			强度副词数目	#这款车很给力！ #这款车动力稍微不足。	1 (解释：很) 1 (解释：稍微)
19			疑问词数目	#算什么好车。 #相机牌子哪个好。	1 (解释：什么) 1 (解释：哪个)
20			感叹词数目	#这款车不错啊。 #这款车，好评!!	1 (解释：啊) 1 (解释：！)
21			书名标题数目	#《百年孤独》看了没？	1 (解释：书名出现一次)
22		S_5	词性数目（例如 a）	#这个牌子不错。	1 (解释：形容词出现一次)
23		S_6	句法关系数目 （例如 SBV）	#这个牌子不错。	1 (解释：主谓结构出现一次)
24		S_7	词典规则计算值	#这家店超赞的！	3 (解释：3=3*1)
25	句法关系	S_8	情感词与修饰词的 句法依存关系	#这个牌子不错。	SBV