中文网络评论的情感特征项选择研究*

王洪伟　郑丽娟　刘仲英　霍佳震
(同济大学经济与管理学院，上海 200092)

摘　要　本文采用统计机器学习方法，对面向情感分类的中文网络评论特征项的选择进行研究。选取词性、词性组合、N-gram作为情感文本的潜在特征项，利用文档频率法对特征项实施降维处理，采用布尔权重法构建特征向量，并采用SVM分类器进行网络评论的情感分类。最后，以手机网络评论为对象进行实验分析，并采用卡方检验测试实验结果的差异显著性。结果表明，中文网络评论的情感分类中，将形容词作为特征项可以获得较高的分类准确率和效率；选用N-gram作为特征项时，分类准确率随着阶数的增加而下降；选取训练语料和特征项的数量对分类效果也有显著影响，但并非数量越多准确率越高。

关键词　网络评论，情感分类，特征项选择，统计机器学习

中图分类号　C931.6，H042

网络评论(online review)反映了用户通过互联网对产品功能或性能发表的看法。与商家发出的促销信息相比，在线评论具有独立性、非商业性，因此深得用户信赖。与此同时，由于缺少线下体验，更多的用户倾向于先看评论，后做决策。Deloitte's Consumer Products Group调查显示，有67%的网民会浏览在线评论，其中82%认为在线评论直接影响了他们的购买意愿。因此，对在线评论进行情感分类，挖掘消费者的偏爱喜好，对商家以及潜在消费者都具有重要的意义。

情感分类研究分两个流派：基于情感词汇语义特性和基于统计自然语言处理。相对于语义方法，基于统计的方法分类准确率较高，有较强的通用性和应用性，已成为首选方法。在统计自然语言处理中，文本被抽象为特征向量，因此哪些特征能够反映文本的情感属性非常关键。特别是，对于中文评论而言，由于语法结构以及情感流露方式的复杂性，已有的成果不能直接应用于中文情感分类，因此本文将研究中文网络评论的特征项选择对情感分类的影响。

1　情感分类相关研究综述

1.1　情感分类基本流程

网络评论的情感分类是通过对非结构化的网络评论文本进行分析，自动将其判断为正面评价或负面评价，从而识别消费者的观点是"赞同"还是"反对"、态度是"肯定"还是"否定"。基于统计自然语言处理的方法，基本过程如图1所示，即经过预处理、文本表示(特征项选择、特征项降维、特征项权重计算)、分类器处理，最终得到一个有关情感类别的输出。

* 基金项目：国家自然科学基金资助项目(70971099，91024023)，中央高校基本科研业务费专项资金资助。
通信作者：王洪伟(1973—　)，男，汉族，大连人，博士，副教授，E-mail：hwwang@tongji.edu.cn，研究方向：商务智能，本体建模，情感计算。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

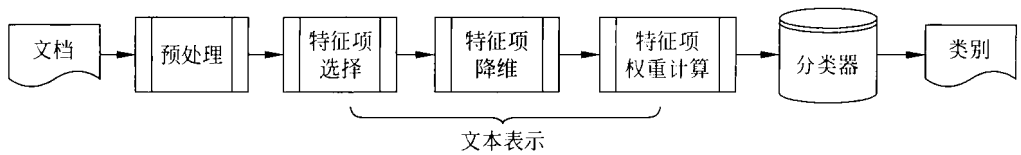

图 1 情感分类的过程示意

1.2 文本表示方法

向量空间模型(vector space model,VSM)是一种有效的文本表示方法。基本过程是：对文本进行分词处理,然后根据训练样本集生成特征项的序列 $\boldsymbol{T}=T(t_1,t_2,\cdots,t_n)$,再根据 $\boldsymbol{T}$ 对训练样本集和测试样本集中的文档进行赋值,生成向量 $\boldsymbol{D}=D(t_1,w_1,t_2,w_2,\cdots,t_n,w_n)$,简记为 $\boldsymbol{D}=D(w_1,w_2,\cdots,w_n)$。其中 w_k 为特征项 t_k 的权重。VSM 模型涉及三个问题：特征项选择、特征项降维和特征项权重计算。

1.2.1 特征项选择

特征项选择,即选取什么语义单元作为特征项,这是决定情感分类效果的重要因素。特征项既要真实地反映文档的情感信息,也要对不同文档有较强的区分能力,可以是词、词的组合、N 元组等。已有一些研究选取词、词的组合、N 元组等作为特征项,但对其分类效果存在较大争议。

(1) 选取词或词的组合。徐军等利用朴素贝叶斯和最大熵方法对新闻语料进行情感分类研究。实验结果显示,选择形容词和名词作为特征项时,具有较高的分类准确率,且分类性能明显好于只选择形容词[1]。周杰等对网络新闻评论的特点进行了归纳总结,并在此基础上选取不同的特征集、特征维度、词性进行分类测试,研究结果显示,名词和动词的分类效果好于形容词和副词[2]。

(2) 选取 N-gram。Pang 等分别以词频作为权重的 Unigrams、以布尔值作为权重的 Unigrams、Bigrams、Unigrams + Bigrams、最前面 2633 的 Unigrams 等作为情感特征项。实验表明,使用布尔值作为权重的 Unigrams 作为特征的分类效果最好,使用 Bigram 作为特征不能达到预期的分类准确率[3]。与之相反,Zhang Z. Q. 等选取 N-gram 作为情感特征项,针对餐饮评论进行研究,结果显示,Bigram 的分类效果好于 Unigram 和 Trigram[4]。Cui 等指出 Pang 的研究语料较小,无法体现 N-grams($n\geqslant 3$)的优势。然后分别令 n 取 1,2,3,4,5,6,实验对比显示,高阶 N-gram 项能够提高情感分类准确率[5]。与 Cui 的结论相反的是,Ng 等发现将 Bigram 和 Trigram 加入 Unigram 项后能够提高 SVM 的分类性能,但如果分别单独使用 Unigram、Bigram 或 Trigram 作为特征项,分类准确率随着阶数的增加反而下降[6]。

1.2.2 特征项降维

特征项降维方法有：基于文档频率(document frequency,DF)法、信息增益(information gain,IG)法、统计量(chi-square statistic,CHI)法、互信息(mutual information,MI)法等。已有一些研究对特征项降维方法进行比较。刘鳓等对 DF,IG,CHI 进行比较,实验结果显示 DF 法优于 CHI 和 IG[7]。Yao 等对 DF、MI、CHI 和 IG 进行比较,实验结果显示,DF 方法的分类效果较好,同时发现 MI 方法不适用于情感特征项的降维[8]。

1.2.3 特征项权重计算

特征项权重计算方法有布尔权重、绝对词频(TF)、倒排文档频度(IDF)、词频-逆文档频率(TF-IDF)等。Pang 等采用布尔权重法进行实验,情感分类准确率达到 82.9%,优于其他权重设置法[3]。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

这是因为语言的褒贬倾向主要取决于正面或负面词语在语言中是否出现，而不是出现的次数。大多数情况下，带有情绪信息的特征项出现几次并不重要，重要的是它是否出现，在哪个类别中出现。

1.3 分类器的选择

文本分类常用的分类器包括支持向量机（support vector machines，SVM）、最大熵（maximum entropy，ME）、朴素贝叶斯（Naïve Bayes，NB）等。

文本分类研究发现，SVM 的分类效果较好，尤其在训练样本有限的情况，效果优于其他分类器。为此，SVM 也成为情感分析首选的分类器。Xia H. S. 以虚拟社区中的旅馆评论为语料库，使用 SVM 进行情感分类，实验结果显示，随语料库内评论数量的增加，SVM 分类准确性有所提高[9]。Li J 以中文电影评论为对象，采用 144 条电影评论作为训练集和测试集，因为训练样本有限，使用 SVM 分类器进行情感分类，准确率高达 85.4%[10]。Shi W 采用 SVM 对中文书评进行情感分类，并与之前在英文评论的分类研究进行比较，实验结果表明，SVM 在中文情感分析方面表现较英文情感分类更优异[11]。Phienthrakul T 对 SVM 的核函数进行研究，认为选用正确的核函数会提高分类的准确性。实验采用产品评论作为情感分类的对象，结果显示多项式核比单核的表现好[12]。Pang 人工标记电影评论中常有的特征情感词，以特征项在文本中出现的频率作为分类特征，采用 NB、ME 和 SVM 三种分类器进行对比实验，结果证明 SVM 分类效果最好[3]。叶强、张紫琼以旅游博客上的评论作为语料库，对朴素贝叶斯和 SVM 的分类效果进行比较，实验结果显示 SVM 优于朴素贝叶斯[13]。叶强、李一军比较了 SVM 和语义方法的分类效果，结果证明 SVM 方法优于基于语义的方法[14]。基于上述分析，本文选择 SVM 作为中文网络评论的情感分类器。

1.4 研究评述

总体而言，在情感分类研究中，分类算法相对成熟，特征项选取方面仍存在不足。网络评论中的词性（如形容词、副词、名词、动词）及词性的组合、*N*-gram 等都有可能成为潜在的情感特征，如何从网络评论文本中选择特征，将直接影响情感型文本的表示，并决定了最终的情感分类效果。本文选用分类效果较好的 SVM 分类算法，对词性、词性组合、*N*-gram 作为特征项的分类效果进行了比较。

2 实验设计

2.1 实验基本流程

实验的基本流程如图 2 所示。本文采用词（形容词、副词、名词、动词）、词的组合、*N*-gram 作为情感文本的潜在特征项，对搜集到的训练语料数据进行处理，并且在各种阈值下获得多种可供实验的向量。

2.2 语料库

从京东网站（www.360buy.com）选取手机评论作为语料库，并将其分为四个语料集，分别为：作为训练语料的正面评论，作为训练语料的负面评论，作为测试的正面评论，作为测试的负面评论。如下例。

正面评论：外观漂亮，屏幕够大，使用方便。

负面评论：电池不好，触摸屏不够灵敏。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

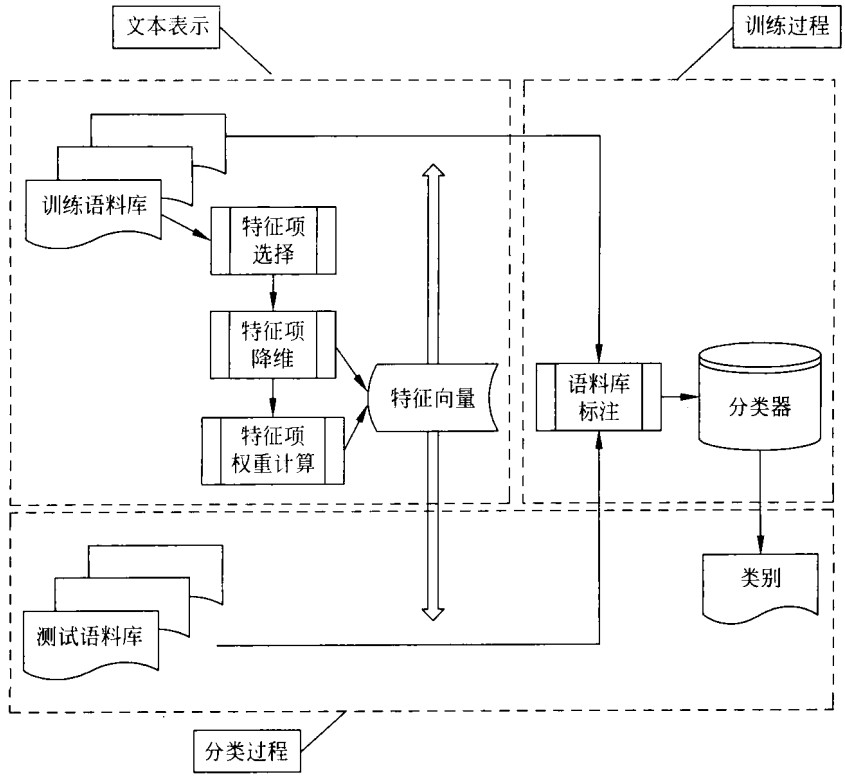

图 2 实验基本流程

训练语料的数量对于监督学习算法的效果至关重要。语料数量不够，分类器将难以有效地反映两类情感极性的特征，造成预测准确率不够。为了分析训练语料数量对分类准确率的影响，从而选择合适的训练语料数量，本文分别选取 300、600、900、1 200 条的平衡语料（正负语料数目相同）作为实验的训练集。

2.3 特征项选取

2.3.1 词和词的组合

采用中国科学院计算技术研究所开发的 ICTCLAS 系统（http://ictclas.org/），进行分词和词性标注，并通过人工方式调整。从训练集中的评论提取表 1 所示的四类词作为备选特征，供进一步算法进行抽取。

表 1 选取的四类特征项

词 性	举 例	词 性	举 例
名词	功能、性价比	动词	用、看
形容词	不错、漂亮	副词	很、也

在提取特征项的过程中，应按照以下原则处理。

（1）为了考虑否定词在评论中的重要影响，如果评论中出现类似“不”、“不是”等否定词，则将否定词与该否定词修饰的词（通常为名词、形容词、副词或者动词）组合在一起作为一个词考虑。比如“我

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

不喜欢手机的外形，看上去不舒服”，两次出现了否定词“不”，就需要分别将其所修饰的词语“喜欢”和“舒服”连在一起提取，因此提取出的特征词应该是“不喜欢”和“不舒服”，这种处理可以很大程度地将否定词考虑进去，从而能更好地反映评论的实际情感。

（2）中文表达方式比较丰富，有时，一个词在句子中的不同位置会表现出不同的词性，比如“这个显示屏效果很出色”中的“很出色”是形容词，而“这个功能很出色地满足日常需要”中的“很出色”则为副词。为不影响分析不同词性的分类效果，本文将表现为不同词性的同一词当做不同的特征项处理，即上面的“很出色”既出现在形容词中，又出现在副词中。这样不仅不会影响在不分词性情况下所有特征词的统计量，而且能最大限度地考虑每种词性下能出现的各种特征。

（3）当形容词、副词后出现“的”、“得”和“地”时，不将这些词和形容词、副词考虑在一起。比如出现“好看的”只选取“好看”作为形容词。表2给出两个特征选取的例子。

表2 特征选取示例

文　件	名词	频数	形容词	频数	动词	频数	副词	频数
按键后总感觉顿一下才有反应，电池不耐用，两天充一次	键	1	不耐用	1	按	1	总	1
	一下	1			感觉	1	才	1
	电池	1			顿	1		
	两天	1			充	1		
看电影很方便，听歌感觉不错	电影	1	方便	1	看	1	很	1
	歌	1	不错	1	听	1		
					感觉	1		

2.3.2 N-gram

N-gram算法的基本思想是将文本内容按字节流进行大小为*N*的滑动窗口操作，形成长度为*N*的字节片段序列，每个字节片段称为Gram，对全部Gram的出现频率进行统计，按照事先设定的阈值进行过滤，形成关键的Gram列表，即为该文本内容的特征向量空间，列表中的每一种Gram均为一个特征向量的维度。例如当$N=2$，即两个汉字的长度时，文本“向量空间模型”可以被切分为“向量”、“量空”、“空间”、“间模”和“模型”。

N-gram算法的优点有：①语种无关性，可同时处理中英文、繁简体文本；②不需对文本内容进行语言学处理，无须分词和词性标注；③对拼写错误的容错能力强，对输入文本的先验知识要求低；④无须词典和规则。

本文使用VBA程序实现长度为*N*的滑动窗口操作，从训练集中提取表3中1-gram、2-gram、3-gram作为备选特征，并通过人工方式调整（调整原则见下文），供进一步算法进行抽取。

表3 选取*N*-gram特征项

文　件	1-gram	2-gram	3-gram
性价比不错	性	性价	性价比
	价	价比	价比不
	比	比不	比不错
	不	不错	不错
	错	错	

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

特征项的提取遵循以下原则：在提取 1-gram、2-gram、3-gram 作为特征项时，长度为 N 的滑动窗口会将标点符号也识别为一个 gram 片段。这些 gram 片段尽管出现次数较多，但对情感分类的作用却微乎其微，因此为获得较准确的分类准确率，应将这些符号人工剔除。

选用分类准确率较高的 DF 法，对选取的特征项进行降维。布尔权重法设置权重后，通过 VBA 程序，实现了所有语料的自动标注，将语料库中非结构化的、机器无法识别的无序化数据转化成了结构化的、机器可识别的有序数据。标注完的语料库会生成一个待处理文档，用于下一阶段的分类器输入处理。

2.4 分类器选取

如前所述，SVM 情感分类的效果较好，为此选用 SVM 分类器。核函数的类型会影响 SVM 分类器的性能。常用的核函数有多项式和函数、径向基核函数(radius basis function，RBF)和 Sigmoid 核函数。其中 RBF 应用最广泛，文献[15]提到 RBF 是一个普适的核函数，通过选择合理的参数可以适用于任意分布的样本，同时文献[16]指出 RBF 在解决小样本问题时的优势。因此采用 RBF 作为核函数。

本文选取 Chih-Chung Chang 和 Chih-Jen Lin 的 LIBSVM-A Library for Support Vector Machines (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)所提供的 SVM 分类器进行实验操作，具体操作步骤如图 3 所示。

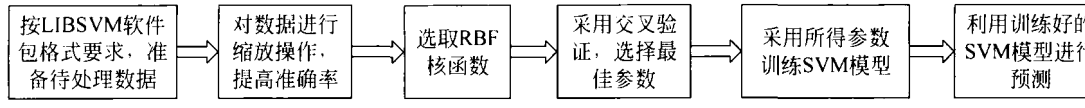

图 3 SVM 操作步骤

3 实验结果分析

3.1 训练集规模对分类性能的影响

分别选取 300、600、900 和 1 200 条语料作为训练集，选取名词、动词、形容词、副词的组合(NVAA)作为特征项，采用 DF 法抽取不同数量(150、200、250)的特征进行实验。实验结果如表 4 和图 4 所示。卡方检验显示，不同训练集规模的分类准确率存在显著差异。

表 4 训练集规模对分类性能影响

特征项数量	训练集规模				χ^2	P
	300	600	900	1 200		
150	86.00%	91.33%	96.32%	94.67%	24.28	0.00
200	81.33%	92.33%	94.65%	95.67%	46.53	0.00
250	81.67%	91.33%	95.65%	94.67%	42.60	0.00

实验结果显示，训练集规模较小时，增加训练集的规模可以提高分类准确率。但当训练语料增加到一定数量时，精确度不再有明显提高，甚至会出现下降，并且训练集规模的增加会导致训练时间的增加。因此，选择训练语料数量时，需要平衡效率和准确率二者的关系。本实验中，训练集规模为 900 时出现拐点，为此之后的实验都选取 900 条平衡语料作为训练集，300 条平衡语料作为测试集。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

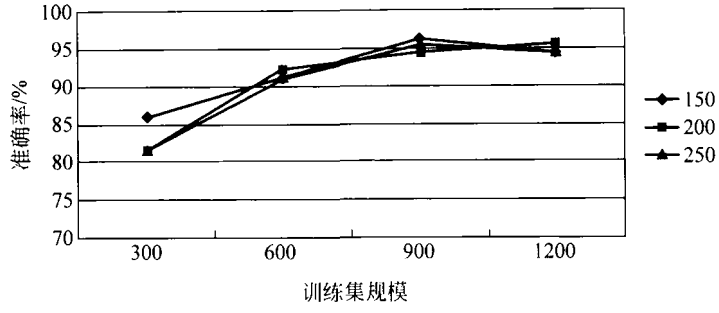

图 4 训练集规模对分类性能影响

3.2 NVAA、N-gram 分类性能比较

分别采用 NVAA 和 *N*-gram 方法选取特征项，进行分类性能比较，实验结果如表 5 和图 5 所示。卡方检验显示，选择 NVAA 或 *N*-gram 作为特征项时的分类准确率存在显著差异，选择不同数量特征时的分类准确率也存在显著差异。

表 5 NVAA 和 *N*-gram 分类性能比较

特征数	准确率				χ^2	*P*
	NVAA	1-gram	2-gram	3-gram		
50	87.29%	89.67%	81.33%	62.33%	86.99	0.00
100	93.31%	93.00%	85.67%	65.67%	113.60	0.00
150	96.32%	95.00%	85.67%	68.00%	127.10	0.00
200	94.65%	94.00%	86.33%	70.33%	95.81	0.00
250	95.65%	94.00%	86.33%	71.33%	95.03	0.00
χ^2	22.95	7.09	3.76	6.96		
P	0.00	0.13	0.44	0.14		

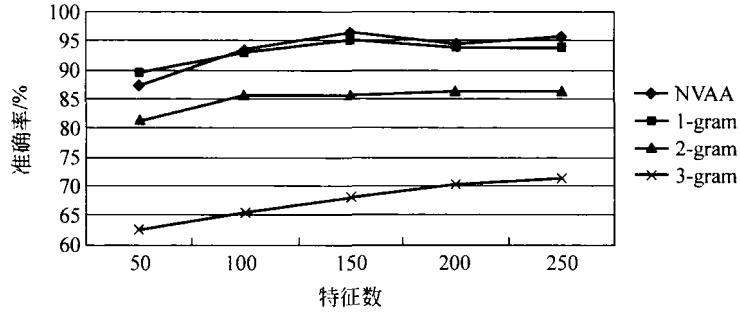

图 5 NVAA 和 *N*-gram 分类性能比较

由图 5 可以看出以下两点。

(1) 通常认为，中文词汇中 2 字短语出现频率最大，所以 2-gram 的分类效率最为理想。但结果显示，选用 *N*-gram 作为特征项，分类准确性随着阶数的增加而下降。3-gram 分类准确率最低，1-gram 的分类效果优于 2-gram，与 NVAA 的分类准确率接近。原因可能是以下几点。

① 由 DF 法提取的 NVAA 中，存在很多单个字，如大、高、快、难、紧，这些字情感倾向明显，对分

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

类的作用较大。1-gram是选择出现频率较高的字作为特征项，所以采用1-gram和NVAA方法选择出的特征项存在很多相同的内容，这可能是1-gram和NVAA分类准确率最接近的一个原因。

② 采用2-gram提取特征项，可能低估特征项的重要性。例如，表达相同情感的两个特征项“大方”和“大气”，分别出现10次和15次，使用DF法排序时，分别排在150位和100位，如果按照1-gram统计，“大”出现25次，使用DF法排序时，则会排在100位以内。这种排序在一定程度上更准确地反映了提取出的特征项的重要性。

③ 2-gram中出现频率较高的特征项，1-gram中肯定出现频率也较高。如2-gram选取出“没有”，1-gram选取出“没”和“有”，所以2-gram中对分类贡献最大的特征项在1-gram中都有所体现。

④ 3-gram和2-gram的分类效果不理想，其原因是采用这种方法提取特征项会产生大量的冗余数据，占用大量空间的同时也导致了时间和效率上的损失。

(2) 选取的特征项的数量对分类准确率有一定影响。从图5可以看出，开始随着特征项数量的增加，分类准确率上升较多，但继续增加分类准确率的上升幅度变小，大概在150或200时出现拐点。为平衡效率和准确率的关系，本文之后的实验选取150个特征。

3.3 不同词性的特征项对分类的贡献

前面实验表明，采用NVAA方法进行特征项提取，分类效果较理想。对于海量语料，如果抽取所有的名词、动词、形容词、副词，将会占用很多时间。为此，需要进一步分析各种词性对分类的贡献。为排除单一验证的随机性对实验结果的影响，实验采用交叉验证的方法。将1200条数据集分成4份，每一份300条评论(正负各150条)，轮流将其中的3份作为训练语料，1份作为测试语料，4次结果的均值作为对算法准确率的估计。

分别选取名词、形容词、副词、动词和四种词性的组合作为特征项，采用DF法选取150个特征进行分类实验，结果如表6所示。卡方检验显示，选择某种词性的词或NVAA作为特征项时的分类准确率存在显著差异。

表6 四种词性和词性组合分类性能

词性	准确率				
	124-3	123-4	234-1	341-2	平均值
形容词	89.67%	93.33%	93.67%	92.67%	92.33%
副词	82.00%	91.67%	90.00%	90.00%	88.42%
动词	74.33%	80.67%	79.67%	77.00%	77.92%
名词	75.67%	70.00%	73.33%	78.67%	74.42%
NVAA	89.33%	94.67%	96.32%	92.33%	93.16%
χ^2					69.34
P					0.00

从表6所示的平均值可以看出，形容词的分类效果比较理想，接近考虑全部词性的精确度，这和实验前所预期的形容词包含更多情感信息的假设相吻合。相比之下名词和动词所能达到的精确度差强人意，两者的抽取效果基本持平。75%左右的准确率表明名词和动词在抽取中的贡献非常少，甚至在一定程度上影响了总体准确率。剔除动词和名词后，进一步实验，分类准确率的变化如表7所示。卡方检验显示，选择形容词、NVAA、VAA和AA作为特征项时的分类准确率不存在显著差异。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表7 词性验证

词 性	准确率				
	124-3	123-4	234-1	341-2	平均值
形容词	89.67%	93.33%	93.67%	92.67%	92.33%
NVAA	89.33%	94.67%	96.32%	92.33%	93.16%
VAA(剔除名词)	93.67%	94.33%	96.33%	93.33%	94.42%
AA(剔除名词和动词)	90.67%	92.00%	94.00%	94.33%	92.75%
χ^2					1.07
P					0.78

从表7所示的平均值可以看出以下两点。

(1) 剔除名词后，将三种词性(形容词、副词与动词)共同作为特征集合，筛选后所得的特征向量分类性能较好。原因：语料中的名词多是“功能”、“电池”等中性词，这些词对分类作用不大，出现的频率却很高，因此采用DF法进行降维时，位置比较靠前，从而排除了其他一些可能对分类更有效果的词。

(2) 剔除名词和动词后的词，即采用形容词和副词的组合作为特征项，分类准确率也较高。原因：比如“非常漂亮”、“很不错”等形容词和副词的组合带有极其强烈的感情色彩，比起单个的副词更能体现评论者的情感。

3.4 实验结论

本实验得出以下结论。

(1) 选用N-gram作为特征项，分类准确率随阶数的增加而下降，即1-gram>2-gram>3-gram，其中1-gram与NVAA的分类准确率接近。

(2) 选用词性和词性组合作为特征项时，由于形容词的分类效果比较理想，在对分类准确率没有太高要求，只是希望快速达到可接受的分类准确率时，可以采用形容词作为特征项。

(3) 训练语料和特征项数量并非越多越好，因此选择训练集规模和特征数量时应平衡效率和准确率的关系。

4 结束语

以中文网络评论为研究对象，分别采用词性、词性组合、N-gram方法选择特征项表示文本，并通过实验研究特征项选择对情感分类准确率的影响。今后将在以下几方面进一步探讨。

(1) 本文研究了采用词、词的组合、N-gram作为特征项时的情感分类准确率。在选择特征项时，本文发现一些符合句法结构的特征更能体现褒贬倾向，如符合主谓结构的“性价比高”、“操作简单”等，因此在已有研究的基础上，加入句法结构相关特征进行实验，是一个值得研究的问题。

(2) 本文采用文档频率法进行特征降维。已有特征降维的研究显示，在采用词、词的组合作为特征项时，文档频率法的分类准确性较高。该结论是否也适用于N-gram作为特征项的情况，即N-gram作为特征项时，采用不同特征降维方法，对分类的准确率将产生怎样的影响，是一个值得探讨的问题。

(3) 除了单纯地识别情感极性，网络评论情感分析还需与其他文本挖掘技术结合，得到比单独的褒或贬的情感极性更有价值的信息。其中情感极性和情感对象的关系抽取是一个应用价值非常广泛的课题。本文的实验结果显示，名词对情感分类的贡献很小，但情感对象多是名词，如“键”、“电池”。因

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

此，是否可以提取名词作为情感对象，再按照不同的情感对象进行情感分类是一个值得研究的问题。

参考文献

[1] 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，2007，21：95-100.

[2] 周杰，林琛，李弼程. 基于机器学习的网络新闻评论情感分类研究[J]，计算机应用，2010，30(4)：1011-1014.

[3] Pang B，Lee L and Vaithyanathan S. Sentiment classification using machine learning techniques[C]. Proceedings of the Conference on Empirical Methods in Natural Language Processing. Philadelphia，US，2002：79-86.

[4] Zhang Z Q，Ye Q，Zhang Z L and Li Y J. Sentiment classification of Internet restaurant reviews written in Cantonese[J]. Expert Systems with Applications，2011，38：7674-7682.

[5] Cui H，Mittal V and Datar M. Comparative experiments on sentiment classification for online product reviews [C]. Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)，Boston，USA，2006：1265-1270.

[6] Ng V，Dasgupta S and Arifin S M N. Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews [C]. Proceedings of the COLING/ACL Main Conference Poster Sessions，Morristown，NJ，USA：Association for Computational Linguistics，2006：611-618.

[7] 刘鎏. 基于统计自然语言的中文评论情感极性分类研究 [D]. 上海：同济大学硕士学位论文，2011.

[8] Yao J N，Wang H W and Yin P. Sentiment feature identification from Chinese online reviews [C]. Communications in Computer and Information Science，2011，201 CCIS：315-322.

[9] Xia H S and Peng L Y. SVM-based comments classification and mining of virtual community：For case of sentiment classification of hotel reviews [C]. Proceedings of the International Symposium on Intelligent Information Systems and Applications (IISA'09)，2009.10：507-511.

[10] Li J. An approach of sentiment classification using SVM for Chinese texts [C]. Proceedings of 2006 International Conference on Artificial Intelligence—50 years' achievements，future directions and social impacts，2006：759-761.

[11] Shi W，Qi G Q and Meng F J. Sentiment classification for book reviews based on SVM model[C]. Proceedings of the 2005 International Conference on Management Science & Engineering. 2005：214-217.

[12] Phienthrakul T，Kijsirikul B，Takamura H and Okumura M. Sentiment classification with support vector machines and multiple kernel functions [C]. ICONIP，2009：583-592.

[13] Ye Q，Zhang Z Q and Law R. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches [J]. Expert Systems with Applications，2009，36(3)：6527-6535.

[14] Ye Q，Lin B and Li Y J. Sentiment classification for Chinese reviews：A comparison between SVM and semantic approaches. Machine Learning and Cybernetics [C]. Proceedings of 2005 International Conference，2005，4：2341-2346.

[15] 李晓宇，张新峰，沈兰荪. 一种确定径向基核函数参数的方法[J]. 电子学报，2005，33(B12)：2459-2463.

[16] 边肇祺. 模式识别[M]. 北京：清华大学出版社，1998.

Sentiment Feature Selection from Chinese Online Reviews

WANG Hongwei，ZHENG Lijuan，LIU Zhongying，HUO Jiazhen

(School of Economics and Management，Tongji University，Shanghai，200092)

Abstract Using statistical machine learning methods for sentiment classification of Chinese online reviews feature selection research. Select words，various combinations of words，N-gram as the potential sentimental feature. Use the

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

Document Frequency to reduce dimensionality, adopt Boolean Weighting method to structure vectors and SVM classifier to classify online reviews. At last, have an experimental analysis based on online reviews of mobile phone. The results showed that: sentiment classification of Chinese online reviews will obtain the highest accuracy when taking adjectives, adverbs and verbs together as the feature. When taking N-gram as the feature, the results showed that low order N-grams can achieve a better performance than high order N-grams. Different training corpus size and feature size have distinct impact on classification, but not the more the better.

Key words Online reviews, Sentiment analysis, Feature selection, Statistical machine learning

作者简介

王洪伟(1973—),男,汉族,大连人,博士,副教授,研究方向:商务智能、本体建模、情感计算。E-mail: hwwang@tongji.edu.cn。

郑丽娟(1983—),女,汉族,山东人,博士研究生,研究方向:商务智能、面向 UGC 的数据挖掘。E-mail: zhenglijuan83@163.com。

刘仲英(1943—),女,汉族,上海人,教授/博导,研究方向:信息管理与信息系统。E-mail: zhongyingliu@gmail.com。

霍佳震(1962—),男,汉族,上海人,博士,教授/博导,研究方向:服务运营、管理信息系统、物流与供应链。E-mail: huojiazhen@163.com。

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net