

文章编号: 1003-0077(2007)06-0095-06

使用机器学习方法进行新闻的情感自动分类

徐军, 丁宇新, 王晓龙

(哈尔滨工业大学 深圳研究生院 智能计算研究中心, 广东 深圳 518055)

摘要: 本文主要研究机器学习方法在新闻文本的情感分类中的应用, 判断其是正面还是负面。我们利用朴素贝叶斯和最大熵方法进行新闻及评论语料的情感分类研究。实验表明, 机器学习方法在基于情感的文本分类中也能取得不错的分类性能, 最高准确率能达到90%。同时我们也发现, 对于基于情感的文本分类, 选择具有语义倾向的词汇作为特征项、对否定词正确处理和采用二值作为特征项权重能提高分类的准确率。总之, 基于情感的文本分类是一个更具挑战性的工作。

关键词: 计算机应用; 中文信息处理; 文本分类; 情感分析; 贝叶斯; 最大熵

中图分类号: TP391

文献标识码: A

Sentiment Classification for Chinese News Using Machine Learning Methods

XU Jun, DING Yu-xin, WANG Xiao-long

(Intelligence Computing Research Center, Shenzhen Graduate School,
Harbin Institute Technology, Shenzhen, Guangdong 518055)

Abstract: In this paper, we study how to apply machine learning techniques to solve sentiment classification problems. The main task of sentiment classification is to determine whether news or reviews is negative or positive. Naïve Bayes and Maximum Entropy classification are used for the sentiment classification of Chinese news and reviews. The experimental results show that the methods we employed perform well. The accuracy of classification can achieve about 90%. Moreover, we find that selecting the words with polarity as features, negation tagging and representing test documents as feature presence vectors can improve the performance of sentiment classification. Conclusively, sentiment classification is a more challenging problem.

Key words: computer application; Chinese information processing; text categorization; sentiment analysis; Naïve Bayes; maximum entropy

1 引言

随着 WWW 的高速发展, 在线的资源数量不断增多, 这些资源多是以文本方式出现, 从而文本分类成为处理和组织大量文档数据的关键技术, 所以利用计算机进行文本的自动分类成为自然语言处理和人工智能领域一项非常有意义的研究课题。目前的文本自动分类的研究主要是采用统计和机器学习的方法, 根据文本的主题和内容对文本进行分类, 例如

文献[1]等。

近年来, BBS 和 Blog 的迅速发展, 主观性的言论也越来越多, 这些言论大部分都是一些主观的论断和对事物的一些看法, 如对电子产品、汽车的使用评价、电影评论等, 那么如何判断这些评论是正面还是反面? 态度是赞成还是反对? 认为其值得推荐还是不值一文? 这就引出了文本分类领域一个新的研究方向——基于情感的文本分类。

基于情感的文本分类是近年来才兴起的一个研究方向, 主要研究如何对文本所表达的情感等主观

收稿日期: 2007-04-10 定稿日期: 2007-07-13

基金项目: 863 专题目标导向类项目(2006AA01Z197); 国家自然科学基金重点项目(60435020)

作者简介: 徐军(1980—), 男, 博士生, 主要研究方向为中文信息处理; 丁宇新(1972—), 博士, 副教授, 主要研究方向为机器学习; 王晓龙(1955—), 教授, 博导, 主要研究方向为机器学习和自然语言处理。

内容进行分类,判断其是正面还是负面。它具有很多重要的实际应用:

推荐系统:对用户对产品/服务的在线反馈进行自动分类和整理,分析和挑选出值得推荐的产品/服务,推荐给其他用户。如根据一部电影的所有相关评论,向未观看者做出是否值得观看的推介。同时,也可以帮助商业机构自动归类其产品和服务的反馈文档,避免繁琐的手工劳动。

过滤系统:对政府或商业机构的不利言论进行过滤,自动识别发文者的政治倾向或对组织机构的看法。如根据文本中的情感进行分类,对攻击个人或政府的邮件进行过滤。

问答系统:对问题答案进行自动的基于情感的文本分类,避免回答的感情色彩出现错误,适得其反。

本文主要研究如何利用机器学习的方法来对新闻及评论来进行正面和负面的分类。基于情感的文本分类的方法和基于主题的文本分类相似,不同的地方在于特征词的提取,在基于主题的文本分类中主题词很重要,而在基于情感的文本分类中具有主观性的词语更为重要,如危机、制裁、退货、优秀等。

2 相关研究

文本作为一种重要的信息交互媒介,主要的功能是传情达意,基于内容的研究已经很成熟,近些年来越来越多的研究开始集中到“传情”上——情感分析,主要的研究内容包括词语的语义倾向识别、基于情感的文本分类、观点提取、主观性分析等。

对一篇文档而言,能对其语义倾向起到决定性作用的主要是构成这篇文档所用的词语。所以,对文本进行基于情感的文本分类的基础是判定词的语义倾向。但是目前,不论是英语还是汉语,都没有一个完整的涵盖词语语义倾向的词典,也不可能有这样的一个完备的词典,因为很多的词语在不同语境中它的语义倾向也不尽相同。Hatzivassiloglou 等根据词语之间关系来判定其语义倾向^[2],他们注意到形容词的语义倾向受连接它们的连接词“and”、“but”等约束,知道其中一个词的语义倾向,就可以推测出另一个。如“excellent and X”,可以推测 X 也是褒义的。根据语言学上连接词对语义倾向约束性的分析,他们提出了一种四步法的有监督学习算法来判断一个形容词的语义倾向。Turney 在其论文^[3]中介绍了两种利用词语与具有明显语义倾向的

种子词语之间统计关系来自动识别词语语义倾向的方法:PMI-IR 和 LSA。Esuli 通过对一个词语的注释(从词典中获得)进行训练和分类,从而判断其他词语的语义倾向^[4]。

基于情感的文本分类研究大多是通过统计正面或负面语义倾向的词语特征数目来对文本进行分类。如 Turney 就使用一无指导学习方法,利用前面提到的 PMI-IR 方法计算出文本中出现的符合规则的短语的语义倾向,通过对文档中所有短语的语义倾向的平均值的正负来判断文档描述的对象是否值得推荐^[5]。Pang Bo 最早利用机器学习方法来解决基于情感的文本分类问题,应用朴素贝叶斯、最大熵、SVM 对电影评论进行分类。SVM 在几种分类方法中效果最好,分类准确率最高达到约 80%^[6]。

事实上,对一篇文章而言,它表达的情感的正面或负面是通过主观语句体现出来的,如“产品质量好!”。但是像“它的售价刚好是¥50 元!”这样的客观语句,虽然有“好”这一特征词,但并不应该能起到任何作用。但是如果区分一篇文章中的主观语句和客观语句,只对主观语句进行特征选择,会对分类的准确率有很大提高。Bruce、Wiebe 等利用贝叶斯分类器对句子的主客观性进行分类^[7,8]。Pang Bo 把主客观语句分类转换成求图的最小截问题,实现一个 Cut-based 分类器,对主客观语句进行分类^[9]。

近年,基于情感的文本分类逐渐被应用到更多的领域中。例如,微软公司开发的商业智能系统 Pulse^[10],它能够从大量的评论文本数据中,利用文本聚类技术提取出用户对产品细节的看法;产品信息反馈系统 Opinion Observer^[11],利用网络上丰富的顾客评论资源,对评论的主观内容进行分析处理,提取产品各个特征及消费者对其的评价,并给出一个可视化结果;Sanjiv 从网络上的股评信息中获取某只股票的市场评价,对股票的价格进行预测^[12]。

中文的基于情感的文本分类研究还不是很多。基于 HowNet,朱嫣岚在其论文^[13]中提出了两种词语语义倾向性计算的方法:基于语义相似度的方法和基于语义相关场的方法。Yuen 通过计算词语和具有强烈感情色彩的语素或单个汉字之间在 LI-VAC 文集上的统计关系来判定词语语义倾向^[14]。T'sou 利用词语的语义倾向来计算新闻文本的语义倾向,衡量公众对名人的评价^[15]。

3 新闻的情感分类

新闻及评论可以被分为两种类型:主观和客观。客观类型常见于新闻,主要是客观描述新闻事件。主观类型常见于新闻评论,不仅仅是对事件的客观描述,还加入作者个人的判断和观点,具有感情色彩,对新闻报导的对象加以正面或负面的评述。纯粹的客观新闻报导只是种理想情形,多数的新闻报导都会具有一定的主观性。

3.1 客观类型

即便是基于客观事实的新闻报道对于被报道的对象而言也有“好事传千里”和“家丑不可外扬”之分,这也体现出作者和媒体的观点与态度,并将会给被报道对象带来好的或坏的影响。我们把带来好的影响称为正面新闻,而带来不好影响的称为负面新闻。例如这样的几则新闻标题:

(1) 宝洁 SK-II 化妆品被查出含有违禁成分

(2) 宝洁公司向希望工程捐款四百万元并建百所希望小学

(1) 是则负面新闻,(2) 则是正面新闻。

3.2 主观类型

新闻评论中带有作者对报道对象的观点和态度,这些观点和态度表现在作者的遣词造句中。我们可以从作者使用的词语的感情色彩和语气上区分出来。所用词语对报道对象持肯定评价、有褒扬色彩的则是正面,而对报道对象予以否定评价、有贬斥感情色彩的则属负面。例如这样的几则新闻标题:

(3) 外籍 CEO 上任,索尼迎来斯丁格时代

(4) 索尼换帅成日本模式最隆重的葬礼

(5) 索尼公司任命新 CEO

(6) 索尼员工能服英国爵士 CEO 吗?

(7) 索尼效益不景气高层换马 找个老外当掌门换手气

这几则新闻都是讲述斯丁格担任索尼总裁这一客观事实,但是从作者的用词上就可以看出作者所持观点。(3)用“迎来”这一褒义词表达作者对此的肯定,属于正面;(4)用“葬礼”这一贬义词形容斯丁格的到任,属于负面;(5)则是中性;用疑问语气来表达怀疑态度,(6)是负面新闻;(7)则使用“换手气”这一俗语来形容索尼的高层变动,是一种轻蔑语气,属负面。

本文研究的对象是这两种类型中的正面和负面的新闻及评论。

4 分类方法

对新闻及评论进行基于情感的文本分类是一个两类分类问题,其目标类别只有正面和负面,我们采用在基于主题的文本分类中常用和性能较好的机器学习方法来实现基于情感的文本分类。在文本的向量空间模型表示中, $\{f_1, f_2, f_3, \dots, f_m\}$ 表示可能会在文本中出现的一个特征序列,共有 m 个, n_i 表示特征 f_i 的权重,文档 d 就被表示成向量 $d = \{n_1, n_2, n_3, \dots, n_m\}$ 。 N_j 为训练集中属于类别 c_j 的文档数目, N 为训练集中的总样本数, $n_{i,j}$ 表示训练集中特征 i 在类别 c_j 中出现的次数。

4.1 Naïve Bayes

Naïve Bayes 分类器是一种概率分类器,它利用类别的先验概率和特征分布相对于类别的条件概率来计算未知文档属于某一类别的概率。假设文档中特征分布相互独立,Naïve Bayes 分类器用数学形式来表示,即:

$$P(c_j | d) = \frac{P(c_j)P(d | c_j)}{P(d)}$$

$$= \frac{P(c_j) \prod_{i=1}^m P(f_i | c_j)^{n_i}}{P(d)}$$

对于不同的类别,上式的分母不变,故只要选择使得分子最大的类别,即认为是待分类文档的目标类别。通过对训练样本的学习,得到 $P(c_j)$ 和 $P(f_i | c_j)$ 的估计:

$$\hat{P}(c_j) = N_j / N$$

$$\hat{P}(f_i | c_j) = \frac{1 + n_{i,j}}{m + \sum_{k=1}^m n_{k,j}}$$

Naïve Bayes 分类器最根本的特点是假设文档中词语的出现是互相独立的,这样的概率方法很简单但确是一种有效的文本分类方法。基于情感的文本分类同样采用文本向量空间模型来表示一个文档,表示为属性值形式,因此可以应用 Naïve Bayes 分类器,一个文档正负面分类就是该文档中的词语在此处被观察到的概率最大的分类。

4.2 最大熵

最大熵方法的基本思想是在满足系统当前提供

的所有条件下寻求分布最均匀的模型,即熵最大的模型。将已知事实作为制约条件,求得可使熵最大化的概率分布作为正确的概率分布。它把训练集中与分类有关的数据描述为一系列的特征,这些特征多数情况下是二值函数。对于基于情感的文本分类问题,我们这里选择“特征词—类别(f_i-c)”作为一个特征,并根据应用需要确定特征值是二值还是词频,以适应文本分类这种文档层面的应用。本文中的特征函数的形式如下:

$$F_{i,c}(d,x) = \begin{cases} n_i, & \text{如果 } n_i > 0, \text{ 且 } x = c \\ 0, & \text{其他} \end{cases}$$

对于特征函数 $F_{i,c}$, 相对于经验概率分布 $\hat{P}(f_i, c)$ 的期望值与相对于模型 $P(c|f_i)$ 的期望值一致, 从而我们可以得到所求概率分布的 m 组约束:

$$E_p F_{i,c} = E_{\hat{p}} F_{i,c}$$

利用拉格朗日乘子算法,求得

$$P^*(c|d) = \frac{1}{\pi(d)} \exp\left(\sum_{i=1}^m \lambda_{i,c} F_{i,c}(d,c)\right)$$

$\pi(d)$ 是归一化因子,且

$$\pi(d) = \sum_{c \in C} \exp\left(\sum_{i=1}^m \lambda_{i,c} F_{i,c}(d,c)\right)$$

$\lambda_{i,c}$ 看成是特征函数的参数,通过在训练集上训练获得,从而得到概率分布,完成最大熵模型的构造。 $\lambda_{i,c}$ 的值我们通过 IIS 算法求得。对于基于情感的文档分类来说,文档中的特征是非常稀疏的,因而其中大部分特征函数值为 0,需要采用平滑技术来处理,本文采用的平滑方法是高斯先验法。

5 实验结果与分析

目前尚未有用于基于情感的文本分类的中文语料,我们从网络上收集最近饱受负面新闻困扰的宝洁、高露洁和索尼的近年的相关的新闻和评论 854 篇作为实验数据,这些文档中包含正面的新闻和负面新闻,但不包含中性新闻。

5.1 实验设置

本文主要采用词频(Term Frequency)和二值(Binary)作为特征项权重,在基于主题的文本分类,TF 的性能明显优于 Binary^[16]。为了研究特征项权重计算方法的不同对基于情感的文本分类的影响,设计了实验一分别对上述两种权重计算方法进行比较实验。

与基于主题的文本分类不同,“不”、“没”等否定

词不再作为停用词被删除,否定词本身是一个中性词,但是当它否定的对象是具有语义倾向的词语时,语义倾向发生反向变化。Sanjiv 在其论文^[12]中采用对否定词到第一个标点符号间的所有词语加上“_N”后缀的方法来处理否定词对一句话的语义倾向的影响,这种方法引入了太多的无关特征项,只很小程度的提高了分类的性能。通过统计上述新闻语料发现,否定词否定的对象主要是其后第一个动词(“是”、“可能”等除外)、名词、形容词、成语和习惯语。本文采用的处理方法是將否定词与否定对象结合为一个新特征。为了研究否定词的处理与否对文本的情感分类的影响,设计了实验二分别对否定词的处理方法进行实验。

表 1 常见具有语义倾向词语的词性及示例

序号	词性	标注	示例
1	形容词	A	美丽;丑陋
2	名词	N	英雄;熊市
3	动词	V	发扬;贬低
4	副词	D	昂然;暗地
5	成语	I	宾至如归;叶公好龙
6	习惯语	L	双喜临门;王八蛋

在基于主题的文本分类中主题词很重要,而在基于情感的文本分类中具有语义倾向的词语和短语更为重要。在汉语中,具有语义倾向的词语并不限于名词、形容词,具有语义倾向的词语的词性有 12 种^[17]。表 1 列出了常见的 6 种。其他的诸如某些叹词、状态词也可能具有感情色彩,但由于在新闻报导中出现的次数非常少,就不再把它们作为特征提取出来。同时,“不”、“没有”等表示否定,以及“非常”、“稍”等表示程度的副词也对文档的情感倾向起到重要作用,所以也都作为特征提取出来。

很多的词本身并不带有情感倾向,但是当词语成对共现或以短语形式出现的时候,就会具有情感倾向,或者对原有单个词语的情感倾向发生改变。例如:“走出”,是一个中性词,常与表示负面的词语成对出现,如“走出困境”、“走出危机”,虽然“危机”、“困境”都是负面的词语,但整个短语却表达的是一个正面的意思。而否定词会对词的语义倾向发生反向的变化,如“不漂亮”;有的还会令没有情感倾向的词带上正面或负面的情感倾向,如“没有感觉”。有的会令原词的情感倾向增强或减弱,如“非常美丽”。“败退”本身是个负面的词语,但是当另外的负面词语

共现时,如:“敌人败退”就表示正面的情感倾向。在基于情感的文本分类中,文档的情感主要是通过具有情感倾向的词语(或共现/关联词语)来体现出来的,这样的词语主要是有形容词、名词、动词、成语和习惯用语。实验三就是为了验证特征词提取上的差异。

实验中,中文文本的分词及词性标注方法采用实验室的 INSUN-ELUS 系统完成,使用词作为文本特征。语料被随机按 1:1 的比例随机划分为训练集和测试集,训练集和测试集不重复。在测试集上的进行 100 次分类试验,利用平均准确率进行对比实验。

5.2 实验结果

实验结果分别参见表 2、表 3 和表 4。

表 2 不同权重计算方法的分类准确率

序号	权重	准确率	
		NB	ME
1	TF	86.89	89.35
2	Binary	87.09	90.06

表 3 处理否定词后的分类准确率

序号	权重	准确率	
		NB	ME
3	TF	87.41	89.47
4	Binary	87.19	90.18

表 4 根据词性选择特征的分类准确率

序号	处理否定词?	词性	权重	准确率	
				NB	ME
5	是	A,D,V, N,I,L	Binary	87.11	90.25
6	是	A,N	Binary	85.70	87.40
7	是	A	Binary	81.27	80.22
8	否	A	Binary	80.37	79.39
9	否	A,I,L	Binary	81.77	80.07

对于基于情感的文本分类,从表 2 中可以看出,采用 Binary 作为特征项权重的分类性能稍优于采用 TF 作为特征项权重,这与基于主题的文本分类截然不同,这主要是因为两个方面:

(1) 具有正面或负面的语义倾向的词语只要在一句话或文章中出现,就决定了这句话或文章的语

义倾向。而与该词出现的次数无关。

(2) 统计发现,对于具有明显的语义倾向的文章中,这种正面或负面的词语,很少重复出现,而是多以同义词方式出现。

从表 3 中我们可以看出,对否定词进行处理能够提高文本情感分类的准确率。汉语的词汇量很丰富,特别是带语义倾向的词语,都有若干个反义词,当作者在表达情感时,常用反义词来表达,使得每篇文档中的否定词有限(平均 3 至 4 个),所以性能提高不是十分明显。比较结果 7 和 8 可以看出,在对文档特征词根据词性进行过滤时,随着特征数目的减少,处理否定词,能一定程度上有效提高分类的准确率。

由于形容词、名词、动词、成语和习惯语是最常见的具有语义倾向的词语,文档中使用的这类词语的语义倾向决定了文档的情感分类,这一点在表 3 种得到了印证。对比结果 4 和结果 5,可以看到文档中的形容词、名词、动词、成语和习惯语是决定一篇文章情感分类的主要因素。从结果 6 我们也可以看出,当特征词仅是形容词和名词时,依然有很高的准确率。比较结果 6 和 7,我们可以看出形容词和名词作为特征词的性能要明显好于只选择形容词,这是因为:

(1) 只有形容词作为特征时,特征数量太少,导致误差。

(2) 很多的形容词是名词共同出现时,才具有语义倾向,或者与不同的名词共同出现具有不同的语义倾向。如:

a. “在媒体强大的压力之下,SK-II 悄然在实际操作中放宽退货条件,真是赔了名声折了业绩”。

b. “从调查中可以看出,索尼仍然表现出强大的品牌优势”。

c. “凭借无可挑剔的技术和产品与出色的公司品牌形象、强大的市场推广能力,索尼迅速成长壮大起来”。

“强大”本身是一个正面词汇,但是在语料中,当其修饰具有不同语义倾向的名词时,却表现出不同的语义倾向。a 中其修饰负面词汇,属负面;b 中其修饰正面词汇,属正面;c 中其修饰中性词汇,属正面。

6 结论

本文中,使用朴素贝叶斯和最大熵方法进行了基于情感的文本分类研究,不足之处未采用支持向

量机方法,其中最大熵方法的准确率在多数情况下要高于贝叶斯方法。从实验的结果看,基于情感的文本分类的准确率明显低于基于主题的文本分类的准确率,而且采用 Binary 作为特征项权重的分类性能不逊于采用词频作为特征项权重,这也是与基于主题的文本分类的最大的不同。

我们在试验中也发现,只有具有正负面情感语义倾向的词语对文本的分类性能起到决定性作用,而一些中性词语则不然,所以选择有情感语义倾向的词语作为特征项将会提高分类性能。另外,如前面的分析,很多的词语在共现时才具有情感语义倾向或具有相反的语义倾向,可以设计一种这样的 Bigram 项生成方法来提高分类性能。同时,新闻和评论中还有很多的客观语句,没有任何的情感语义倾向,有效的识别这些语句并去除也将能极大提高分类的性能,这也将是情感分析研究中的一个重要方向。

准确的新闻及评论的基于情感的文本分类将会帮助个人、公司等有效地采取措施,减少负面新闻在网络等媒体上的传播和对其声誉造成的负面影响。本文使用机器学习的方法对新闻及评论进行正负面的情感分类,只是一个尝试,期望对相关的需求和应用有所帮助。

参考文献:

- [1] Kjersti Aas, Line Eikvil. Text Categorisation: A Survey [EB/OL]. Technical Report, Norwegian Computing Center, 1999.
- [2] Vasileios Hatzivassiloglou, Kathleen R. McKeown. Predicting the Semantic Orientation of Adjectives [A]. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the ACL [C]. 1997. 174-181.
- [3] Turney Peter, Littman Michael. Measuring Praise and Criticism: Inference of Semantic Orientation from Association [J]. ACM Transactions on Information Systems, 2003, 21(4): 315-346.
- [4] Esuli, Andrea, Sebastiani, Fabrizio. Determining the Semantic Orientation of Terms Through Gloss Classification [A]. In: Proceedings of CIKM-05, the ACM SIGIR Conference on Information and Knowledge Management [C]. 2005. 617-624.
- [5] Turney Peter. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews [A]. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics [C]. 2002. 417-424.
- [6] Bo Pang, Lillian Lee, Shivakumar Vaithyanathan. Thumbs up? Sentiment Classification Using Machine Learning Techniques [A]. In: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing [C]. 2002. 79-86.
- [7] Rebecca Bruce, Janyce Wiebe. Recognizing Subjectivity: A Case Study in Manual Tagging [J]. Natural Language Engineering, 1999, 5(2):1-16.
- [8] Janyce Wiebe, Ellen Riloff. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts [A]. In: Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing [C]. 2005.
- [9] Pang, B., Lee, L. A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts [A]. In: Proceedings of ACL 2004 [C]. 2004. 217-278.
- [10] M. Gamon, A. Aue, et al. Pulse: Mining customer opinions from free text [A]. In: Proceedings of the 6th International Symposium on Intelligent Data Analysis [C]. 2005. 121-132.
- [11] Bing Liu, Mingqing Hu, Junsheng Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web [A]. In: Proceedings of WWW2005 [C]. 2005. 324-351.
- [12] Sanjiv D, M. Chen. Yahoo! for Amazon: Extracting Market Sentiment from Stock Message Boards [A]. In: Proceedings of the Asia Pacific Finance Association Annual Conference (APFA) [C]. 2001.
- [13] 朱嫣岚, 闵锦等. 基于 HowNet 的词汇语义倾向计算 [J]. 中文信息学报, 2005, 20(1):14-20.
- [14] R W M Yuen, T Y W Chan et al. Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words [A]. In: Proceedings of the 20th International Conference on Computational Linguistics (COLING-2004) [C]. 2004. 1008-1014.
- [15] B K Tsou, R W M Yuen et al. Polarity Classification of Celebrity Coverage in the Chinese Press [A]. In: International Conference on Intelligence Analysis [C]. Virginia, USA; 2005.
- [16] Lan M, S Y Sung et al. A Comparative Study on Term Weighting Schemes for Text Categorization [A]. International Joint Conference on Neural Networks [C]. 2005.
- [17] 王治敏, 朱学锋, 俞士汶. 基于现代汉语语法信息词典的词语情感评价研究 [J]. Computational Linguistics and Chinese Language Processing. 2005, 10(4): 581-592.