

# 基于模糊集合的汉语主观句识别\*

宋洪伟, 贺宇, 付国宏

(黑龙江大学计算机科学技术学院, 黑龙江省哈尔滨市 邮编 150080)

**摘要:** 主观句识别的工作在诸如情感分类和意见摘要等意见挖掘系统中占有很重要的地位。在本文中, 我们提出一种基于情感密度的模糊集合分类器以识别汉语主观句。首先, 我们利用优势率方法从训练语料中抽取主观性线索词。然后, 为了能更好的表达一个句子的主观性, 我们利用抽取出的主观性线索词计算出每个句子的情感密度。最后, 我们结合情感密度的特点实现了一个三角形隶属度函数的模糊集合分类器以识别主观句。我们在 NTCIR-6 中文数据中做了两组实验。实验结果表明我们的方法具有一定的可行性。

**关键词:** 主观句识别; 情感密度; 模糊集合; 优势率

**中图分类号:** TP391

**文献标识码:** A

## Chinese Subjective Sentence Recognition Based on Fuzzy Sets

Hongwei Song, Yu He, Guohong Fu

(School of Computer Science and Technology, Heilongjiang University, Harbin city, Heilongjiang province zip code 150080, China)

**Abstract:** Subjectivity recognition plays an important role in many opinion mining systems such as sentiment classifiers and opinion summarization systems. In this paper, we present a sentiment density-based fuzzy sets classifier for Chinese subjectivity classification. In this study, we first employ the odds ratio technique to extract subjective cues from training data. Then, we calculate sentiment density using the extracted subjective cues to represent sentence subjectivity. Finally, we implement a triangular fuzzy sets classifier with sentiment density as features for subjectivity classification. We conduct two experiments on the NTCIR-6 Chinese opinion data, showing the feasibility of the proposed method.

**Key words:** Subjectivity recognition; sentiment density; fuzzy sets; odds ratio

### 1 引言

近几年来, 随着网络上用户生成数据爆发式的增长, 意见挖掘在自然语言处理领域中已经越来越受关注<sup>[1][2][3]</sup>。作为意见挖掘中的一个子任务, 主观句识别的主要目的是判定一个给定的句子是主观句还是客观句。对于许多意见挖掘系统, 比如情感分类、意见摘要和意见问答等系统, 预先将主观句从客观句中识别出来可以降低相关问题的复杂度, 而且还能够提高系统的性能。

虽然近几年来主观句识别技术已经有了很大进步, 但是对于面向网络文本的意见挖掘应用来说, 主观句识别问题仍然没有得到很好的解决。很大一部分原因是由于主观句的语言特性过于灵活多变<sup>[2]</sup>。一方面, 人们总是用各式各样的方法表达主观信息; 另一方面, 主观信息通常是上下文相关或领域相关的<sup>[2]</sup>。这使得抽取大量主观性线索及更好的描述主观句变得十分困难。此外, 由于意见挖掘相关研究工作仍处于早期阶段, 所以没有足够的标注语料用于主观句识别模型的训练。因此, 如果能发现主观性文本的本质特征并据此提出一种简洁的模型, 对于主观句识别工作甚至是意见挖掘领域的其他工作都将具有重大的意义。

针对以上问题, 本文在前人工作基础上提出一种基于情感密度的模糊集合分类器来识别

---

\* **收稿日期:**

**定稿日期:**

**基金项目:** 国家自然科学基金 (60973081, 61170148); 黑龙江省人力资源和社会保障厅留学人员科技活动项目 (1154hz26); 黑龙江大学研究生创新科研项目重点项目 (YJSCX2014-017HLJU)

**作者简介:** 宋洪伟 (1989—), 男, 研究生, 主要研究领域为自然语言处理。Email: songhongwei@live.cn; 贺宇 (1988—), 男, 研究生, 主要研究领域为自然语言处理。Email: heyucs@yahoo.com; 付国宏 (1968—), 男, 博士/教授, 主要研究领域为自然语言处理、文本挖掘。Email: ghfu@hotmail.com。

汉语主观句。首先,我们利用优势率方法在已分好词的训练数据中抽取主观性线索词。为了更好的描述主观句,我们用已抽取出的线索词计算训练语句的情感密度。最后,我们把训练语料的情感密度作为识别主观句的特征,并以此实现了一个三角形隶属度函数的模糊集合分类器。我们认为,相对于传统的分类方法,基于模糊集合的分类方法提供了一个更为直接的方式来区分出主观句与客观句在概念外延上的细微差别。这一点在 NTCIR-6 中文数据<sup>[3]</sup>上的初步实验结果中得到验证。

本文接下来的安排如下:第 2 节简要介绍了相关工作及背景。第 3 节描述我们方法的具体细节。第 4 节给出了 NTCIR-6 数据上的实验结果。最后,第 5 节给出了本文工作的结论以及未来研究的展望。

## 2 相关工作

之前的研究工作大部分都是把主观性识别问题看作将一个给定句子分成主观的或客观的这样一种二分类任务。为了从描述性文本中分离出意见句,Yu 和 Hatzivassiloglou(2003)提出了三个不同的方法,分别叫做基于句子相似度的方法、融合多特征的朴素贝叶斯分类器和多重朴素贝叶斯分类器<sup>[4]</sup>。他们的实验结果显示多特征和多重分类器的融合对主观性识别有很大的帮助。与 Yu 和 Hatzivassiloglou 不同,Pang 和 Lee(2004)将文档级别和句子级别的主观性识别问题形式化地看作一个统一的问题,并提出了一个基于最小切割的解决方法<sup>[5]</sup>。像他们所描述的,不论是文本内的上下文信息还是传统的观点词特征都能被结合起来完成主观性识别。而 Lin 等人(2011)提出了一种称为 subjLDA 的基于潜在狄利克雷分布的分层贝叶斯模型,以此自动识别主观句。他们的方法只需要较少的领域相关的主观性线索词<sup>[6]</sup>。

如何发现恰当的主观性线索是主观性识别任务的关键问题。早期的研究工作集中探索观点词,尤其是主观性形容词,以完成主观性识别任务。Hatzivassiloglou 和 Wiebe(2000)表明形容词是主观性语句的一类很好的指示词<sup>[7]</sup>。除了主观性形容词,Riloff 等人(2003)还探索了主观性名词对主观性识别任务的影响<sup>[8]</sup>。在他们的工作中,使用 bootstrapping 算法从未标注的语料中抽取主观性名词。他们还指出,主观性名词虽然非常重要但是很少被使用。随后,Wiebe 和 Mihalcea(2009)指出词义与主观性有很紧密的关联<sup>[9]</sup>。Akkaya 和 Wiebe(2009)探索利用主观性词义消解来提高主观分析系统的性能<sup>[10]</sup>。他们表明利用主观性词义消解能显著提高情感分析任务的性能。除了词级别的线索,近期的一些研究工作还考虑了其他线索,比如在主观性识别中考虑情感模式规则<sup>[11]</sup>。Jindal 和 Liu(2006)探索使用序列模式挖掘算法来自动从语料中抽取基于类别的序列规则,然后进一步用这些序列规则识别产品评论中的主观性比较句<sup>[12]</sup>。Karamibekr 和 Ghorbani(2013)则以动词为主要线索,制定了一系列的启发式规则,然后面向社会热点评论本文抽取能够代表主观句的意见三元组<sup>[13]</sup>。除此之外,Remus(2011)假设自然语言文本的主观性与其可读性有一定的联系。他把文本的可读性度量值结合传统的主观性线索词作为特征,使用支持向量机模型识别主观句<sup>[14]</sup>。特别的,Wang 和 Fu(2010)提出了一种基于情感密度子区间的朴素贝叶斯分类器,他们将词语级别的特征融合为情感密度,进一步将情感密度划分为一系列的子区间,并以此作为朴素贝叶斯分类器的特征<sup>[15]</sup>。他们表明利用情感密度作为特征能较好的区分出主客观句之间的不同。

在本文中我们处理汉语句子级别的主观性分类问题。我们使用 Wang 和 Fu(2010)提出的情感密度的概念来描述句子的主观程度。与其不同的是,我们首先利用优势率方法抽取主观性线索词及其权重,然后使用情感密度作为模糊集合分类器的特征以判断一个给定的句子是主观句还是客观句。与现存的主观性识别系统相比较,我们的方法提供了更完善的统一框架来处理不同类型的主观性线索词,而且能够应对大量开放性文本中各种各样的主观句。

## 3 方法

在本节中,我们详细介绍我们提出的汉语主观句识别方法,包括主观性特征的选取、情感密度的定义和基于情感密度的模糊集合分类器。

### 3.1 主观性特征

我们从组成主观句的基本要素上寻找主观性特征。一般的，一个意见句包括一个观点的持有者、一个意见指示词、一个意见客体以及一个或多个表达情感极性的极性词<sup>[1]</sup>。在具体的主观句中，观点持有者通常作为命名实体或代词出现，极性词指明了情感的极性。意见持有者通常使用一些特殊的动词表达对某一对象的观点，比如“指出”和“认为”这样的意见指示动词。考虑下面的观点句“他同时指责北约和美国军方不负责任地撒谎。”在这句主观句中，“他”是表达该观点的观点持有者，“指责”是一个意见指示动词，“美国军方”是观点的客体。“不负责任地”和“撒谎”是表达两个负向极性的极性词。

通过上面的观察分析，我们主要考虑五种主观性线索词。他们是命名实体或代词、意见指示词、属性词、极性词和程度副词，如表1所示：

表1 主观性线索词示例

主观性线索词	示例
命名实体或代词	中国、他
意见指示词	指责、申明
属性词	价格、外观
极性词	漂亮、丑陋
程度副词	很、更

为了方便起见，我们将表1中的线索词统称为主观性关键词。显然，我们可以简单的从现有的情感词典中抽取相应词性的情感词作为一个主观性关键词词典。然而，这样做词典的覆盖度必然会有限制。为了弥补词典的低覆盖度，我们利用优势率的方法自动从训练数据中抽取更多与主观句有高关联性的主观性关键词，同时将利用优势率公式计算出的值作为主观性关键词的权重。

公式(1)是词 *word* 在训练语料中的优势率计算公式。

$$Score(word) = \log \frac{P(word / sub)(1 - P(word / obj))}{P(word / obj)(1 - P(word / sub))} \quad (1)$$

其中， $P(word/sub)$ 代表的是词 *word* 在主观句中出现的条件概率， $P(word/obj)$ 代表词 *word* 在客观句中出现的条件概率。为了方便起见，我们采用公式(2)和(3)所示的最大似然估计方法来计算  $P(word/sub)$ 和  $P(word/obj)$ 的值：

$$P(word / sub) = \frac{count(word, sub)}{count(sub)} \quad (2)$$

$$P(word / obj) = \frac{count(word, obj)}{count(obj)} \quad (3)$$

其中， $count(word,sub)$ 表示在训练语料中包含词 *word* 的所有主观句的数目。 $count(sub)$ 表示在训练语料中所有主观句的数目。 $count(word,obj)$ 和  $count(obj)$ 的含义与上述类似。

优势率公式主要应用于二分类问题。当词 *word* 的优势率为正数时，权值越高表明该词与主观句有越紧密的关联；对称的，当词 *word* 的优势率为负数时，权值越小表明该词与客观句有越紧密的关联，此时我们称之为客观性关键词。

### 3.2 情感密度

为了更好地描述句子的主观程度，我们在本文中采用情感密度的概念<sup>[15]</sup>。在问答系统和摘要生成领域，常采用关键词的密度来给句子打分，以选择与主题相关的有代表性的句子。像我们在3.1节中已经讨论过的，一个主观句通常包括主观性关键词。基于这点，我们采用

问答系统中关键词密度的概念来表示主观句中主观性关键词的密度。为了方便起见，我们称之为情感密度。它的定义如公式(4)：

$$SD(S) = \frac{1}{N \times (N + 1)} \sum_{i=1}^{N-1} \frac{Score(w_i) \times Score(w_{i+1})}{Distance(w_i, w_{i+1})} \quad (4)$$

这里， $N$ 是句子 $S$ 中的关键词的总数， $Distance(w_i, w_{i+1})$ 指的是句子 $S$ 中相邻的两个关键词 $w_i$ 和 $w_{i+1}$ 之间的非关键词数量。 $Score(w_i)$ 是关键词 $w_i$ 的权重，该值通过3.1节的公式(1)计算得出。

我们希望情感密度 $SD(S)$ 不仅能表示一个句子的主观程度，还能够尽可能的显示出主观句与客观句之间的不同。所以我们将公式(4)中 $w_i$ 的意义进行扩展，引入权重为负的客观性关键词。我们认为由权重为负的客观性关键词可计算出数值为负数的情感密度，此时的情感密度可代表一个句子的客观程度，这样的情感密度在同一框架下应该更能够展现主观句与客观句之间的细微差别。为了实现上述的方法，在利用公式(4)计算每个句子的情感密度时，若 $w_i$ 与 $w_{i+1}$ 权重为负值，我们就翻转他们乘积的符号。图1给出情感密度算法的伪代码描述。

```

输入：待计算情感密度的汉语语句 S
输出：汉语语句的情感密度值 sd
scores=0,sd=0,distance=0,N=0;
for each word in S:
    if word is not a keyword:
        distance+=1
    else:
        N+=1
        scores*=word.score
        If word.score <0:scores*=-1
        sd+=scores/distance
        scores=word.score
        distance=0
return sd/(N*N(+1))

```

图1 情感密度算法

### 3.3 模糊集合分类器

通过上文的介绍，我们得到句子的情感密度，其在一定程度上刻画了所属句子的主观性程度。我们通过对情感密度分布的理论分析和实验观察，发现在大规模文本中，其情感密度的分布是不均匀的。在训练语料中，大部分句子的情感密度都集中在某一较小的区间内。同时，在情感密度小于某个较小阈值的句子集中，大部分句子都是客观句；对称地，在情感密度大于某个较大阈值的句子集中，大部分句子都是主观句。但由于语言表达的灵活性使得情感密度的分布具有一定的模糊性，简单的规定某个阈值无法适应这样的特点。基于以上情感密度分布的特点，本文采用三角形隶属度函数的模糊集合分类器区分句子的主观性强度。

首先，我们将句子的情感密度划分到三个模糊集合中，分别为“低主观性强度”、“中主观性强度”和“高主观性强度”。然后，我们选择三角形隶属度函数作为以上三个主观性强度集合的隶属度函数。我们先给出三角形隶属度函数的定义，如公式(5)所示：

$$\begin{aligned}
T_{\text{low}}(x) &= \begin{cases} 1 & x \leq m_1 \\ \frac{m_2 - x}{m_2 - m_1} & m_1 < x < m_2 \\ 0 & x \geq m_2 \end{cases} \\
T_{\text{med}}(x) &= \begin{cases} 0 & x \leq m_1 \\ \frac{x - m_2}{m_2 - m_1} & m_1 < x < m_2 \\ \frac{m_3 - x}{m_3 - m_2} & m_2 < x < m_3 \\ 0 & x \geq m_3 \end{cases} \\
T_{\text{high}}(x) &= \begin{cases} 1 & x \geq m_3 \\ \frac{x - m_2}{m_3 - m_2} & m_2 < x < m_3 \\ 0 & x \leq m_2 \end{cases} \quad (5)
\end{aligned}$$

其中,  $T_{\text{low}}(x)$ 、 $T_{\text{med}}(x)$ 、 $T_{\text{high}}(x)$ 分别是句子的情感密度从属于相应三个模糊集合的隶属度;  $m_1$ 、 $m_2$ 、 $m_3$ 是相应三个隶属度函数的聚类中心。我们使用简单可靠的 K-MEANS 方法确定聚类中心的值。

K-MEANS 算法是一种得到广泛使用的基于划分的聚类算法,把  $n$  个对象分为  $k$  个类,以使类内具有较高的相似度。相似度的计算根据一个类中对象的平均值来进行。算法首先随机地从训练语料中选择三个句子的情感密度值,每个情感密度值初始地代表了一个类的中心:  $m_1$ 、 $m_2$ 、 $m_3$ 。对训练语料中剩余的每个情感密度值根据其跟各个类中心的距离,将它赋给最近的类,然后重新计算每个类的平均值。这个过程不断重复,直到所有三个聚类中心同时收敛。

至此,我们可以利用三角形隶属度函数来判断给定句子的主观性强度,并根据分数大小确定句子所属的主观性强度集合。为了识别主观句,我们采用如下规则:如果一个句子属于“中主观性强度”集合或者“高主观性强度”集合时,该句属于主观句。

基于上述原理,我们面向汉语句子主观性分类实现了一个基于情感密度的模糊集合分类器。该分类器主要包括三个步骤:首先,预处理模块对输入的句子进行分词、词性标注和命名实体识别等分析处理,以便获取后续主观性分类所需的线索词。为了完成预处理任务,该模块分别嵌入一个基于语素的汉语词法分析器<sup>[16]</sup>和一个基于 LHMM 的汉语命名实体识别器<sup>[17]</sup>。接着,计算给定句子的情感密度。最后,利用三角形隶属度函数计算该句子属于不同主观性强度集合时的分数,并根据分数大小确定句子的主观性强度,进而根据上述规则判定句子是否为主观句。

## 4 实验和结果

为了验证我们方法的有效性,我们在 NTCIR-6 中文语料上进行了两组实验,本节介绍这些实验的结果。

### 4.1 实验设置

本文实验数据来自 NTCIR-6 的中文语料。表 2 给出了实验数据的基本统计信息。为了评价系统的性能,本文采用 NTCIR-6 的 LWK 评价方法,并采用 NTCIR-6 的 Lenient 评价标准下的准确率、召回率和 F-值三个指标来评价系统的性能。

表 2 实验数据的统计信息

项目	统计信息	
	训练数据	测试数据
主题	4	28
文档数	143	700
句子数	2660	9247
宽松标准下的意见句子	62%	62%
严格标准下的意见句子	25%	25%

为了确定情感关键词及主观性特征，实验中的基础词典来自 NTU 和 CUHK 情感词典。在此基础上，我们用优势率方法从训练语料中抽取主观性线索词以扩充情感词典。最终实验所用的情感词典包含 852 个代词和名词、1832 个意见指示动词、8750 个情感词和 219 个程度副词。

#### 4. 2 实验结果

我们的第一组实验的目的是通过实验对比验证模糊集合分类器结合情感密度对主观性识别的效果，表 3 是实验的结果。

表 3 不同分类方法的主观句识别结果

方法	准确率	召回率	F-值
Wang 和 Fu(2010) <sup>[15]</sup>	62.8%	96.5%	76.1%
本文系统	67.1%	91.1%	77.3%
K-MEANS 方法	66.5%	85.8%	74.9%

在这组实验中，在情感密度的计算方法上我们采用与 Wang 和 Fu(2010)<sup>[15]</sup> 一致的策略，即仅使用权值为正数的主观性关键词计算每个句子的情感密度。与 Wang 和 Fu(2010)<sup>[15]</sup> 使用基于情感密度子区间的朴素贝叶斯分类器不同，我们使用基于情感密度的三角形模糊集合分类器。表 3 的实验结果显示我们的系统较 Wang 和 Fu(2010)<sup>[15]</sup> 的系统在总体的 F 值上提高了 1.2%。这在一定程度上说明比起朴素贝叶斯分类器，模糊集合分类器能更好的利用情感密度以区分主观性与客观性之间的差别。在 K-MEANS 方法中，我们简单地把利用 K-MEANS 聚类方法得到的“中主观性强度”集合的聚类中心值作为判别主客观句的阈值，以此识别主观句。此实验表明，简单的使用特征值划分主观句与客观句之间的界限无法得到更好的效果。

我们分析认为，实验中的情感密度以主观性线索词作为特征，将表达方式灵活多变的句子映射到情感密度的度量值上进行区分。直观上，一个句子的情感密度越大意味着该句的主观性强度越大，就越可能是一个主观句。但是，由于主观句的表达方式灵活多变以及其与客观句的差别细微，很难判断主观句与客观句之间的精确界限。因此模糊集合能够很好的描述情感密度的本质。实验也证明了我们方法的可行性。

我们的第二组实验的目的是验证主观性关键词与客观性关键词对情感密度产生的影响。在这组实验中，方案 1 仅使用权值为正数的主观性关键词作为特征计算情感密度，而方案 2 的情感密度计算公式融合了权值为正数的主观性关键词与权值为负数的客观性关键词。

表 4 不同关键词对主观句识别的影响

方案	准确率	召回率	F-值
方案 1: 主观性关键词	67.1%	91.1%	77.3%
方案 2: 混合关键词	67.9%	90.3%	77.5%

表 4 的实验结果表明，混合关键词的方案 2 准确率最高，达到了 67.9%，验证了混合关键词特征可以增强情感密度对主客观句之间细微区别的区分能力。但是召回率下降了 0.8%，最终 F 值提高了 0.2%。可能是因为抽取的特征里参杂了过多的噪音，使得某些句子的情感

密度表现异常。

我们分析认为,在包含了多个词性特征的情感密度这一统一框架下,方案1仅使用主观性关键词作为特征,在一定程度上限制了情感密度区分主观句与客观句的能力。所以我们在方案2中引入了权值为负数的客观性关键词作为特征,希望能以此增强情感密度的区分能力。实验也证明了我们方法的有效性。

表5 本文系统与 NTCIR-6 最好系统的比较

方案	准确率	召回率	F-值
UMCP-1 <sup>[18]</sup>	64.5%	97.4%	77.6%
本文系统	67.9%	90.3%	77.5%

表5比较了本文系统与 NTCIR-6 中最好系统的结果。UMCP-1<sup>[18]</sup>系统首先利用自动收集与人工校对相结合的方法构造情感词典,然后根据一个句子中情感词的个数来判别该句是否为客观句。

实验结果表明,本文系统仅比 UMCP-1<sup>[18]</sup>系统的 F 值低 0.1%。我们分析可能是由于训练语料过小,抽取出的情感词质量不够高,使得我们无法更精确地调整三角形隶属度函数的参数。相比 UMCP-1<sup>[18]</sup>系统,本文系统可自动识别主观句,无需手工校对情感词典等方式进行人工维护。这使得我们的方法具有更大的适用性,可以更好地应对大量开放性文本中各种各样的主观句。

## 5 结论

本文提出了一种融合情感密度和模糊集合的汉语主/客观句分类系统,并采用 NTCIR-6 数据对系统进行了测试。实验表明我们的方法有一定的可行性,这在一定程度上说明:在模糊集合框架下,融合主观性关键词与客观性关键词的情感密度能够很好地区分主客观句子在概念外延上的细微区别。虽然在所进行的实验中我们系统的准确率达到最高,但 F 值提升的幅度相对并不明显。我们分析可能是由于训练语料太小,抽取出的特征质量不够高。这使得情感密度分布有一定的局限性,三角形隶属度函数的参数得不到精确的划分。因此,在将来的工作中我们将研究如何提高特征词的质量,并进一步扩大训练语料库。同时,我们还将研究如何构造和选取其他特征与情感密度进行融合,以弥补情感密度受低质量特征词的影响。

## 参考文献

- [1] Liu B. Sentiment analysis and subjectivity[J]. Handbook of natural language processing, 2010, 2: 627-666.
- [2] Pang B, Lee L. Opinion mining and sentiment analysis[J]. Foundations and trends in information retrieval, 2008, 2(1-2): 1-135.
- [3] Seki Y, Evans D K, Ku L W, et al. Overview of opinion analysis pilot task at NTCIR-6[C]//Proceedings of NTCIR-6 Workshop Meeting, 2007: 265-278.
- [4] Yu H, Hatzivassiloglou V. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences[C]//Proceedings of the 2003 conference on Empirical methods in natural language processing. Association for Computational Linguistics, 2003: 129-136.
- [5] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts[C]//Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004: 271.
- [6] Lin C, He Y, Everson R. Sentence Subjectivity Detection with Weakly-Supervised Learning[C]//IJCNLP. 2011: 1153-1161.
- [7] Hatzivassiloglou V, Wiebe J M. Effects of adjective orientation and gradability on sentence subjectivity[C]//Proceedings of the 18th conference on Computational linguistics-Volume 1. Association for Computational Linguistics, 2000: 299-305.
- [8] Riloff E, Wiebe J, Wilson T. Learning subjective nouns using extraction pattern bootstrapping[C]//Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4. Association for Computational Linguistics, 2003: 25-32.
- [9] Wiebe J, Mihalcea R. Word sense and subjectivity[C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2006: 1065-1072.
- [10] Akkaya C, Wiebe J, Mihalcea R. Subjectivity word sense disambiguation[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1. Association for

- Computational Linguistics, 2009: 190-199.
- [11] Riloff E, Wiebe J, Phillips W. Exploiting subjectivity classification to improve information extraction[C]//Proceedings of the National Conference On Artificial Intelligence. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2005, 20(3): 1106.
  - [12] Jindal N, Liu B. Identifying comparative sentences in text documents[C]//Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2006: 244-251.
  - [13] Karamibekr M, Ghorbani A A. Sentence subjectivity analysis in social domains[C]//Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on. IEEE, 2013, 1: 268-275.
  - [14] Remus R. Improving sentence-level subjectivity classification through readability measurement[C]//NODALIDA-2011 Conference Proceedings. 2011: 168-174.
  - [15] Wang X, Fu G H. Chinese subjectivity detection using a sentiment density-based naive Bayesian classifier[C]//Machine Learning and Cybernetics (ICMLC), 2010 International Conference on. IEEE, 2010, 6: 3299-3304.
  - [16] Fu G, Kit C, Webster J J. Chinese word segmentation as morpheme-based lexical chunking[J]. Information Sciences, 2008, 178(9): 2282-2296.
  - [17] Fu G, Luke K K. Chinese named entity recognition using lexicalized HMMs[J]. ACM SIGKDD Explorations Newsletter, 2005, 7(1): 19-25.
  - [18] Wu Y, Oard D W. NTCIR-6 at Maryland: Chinese opinion analysis pilot task[C]//Proceedings of the 6th NTCIR Workshop on Evaluation of Information Access Technologies. 2007: 344-349.