

采用无标注语料的动词和形容词主观性评级*

徐戈^{1,2,3}, 蒙新泛^{1,2}, 王厚峰^{1,2}

¹(北京大学 计算语言学研究, 北京 100871)

²(计算语言学教育部重点实验室(北京大学), 北京 100871)

³(闽江学院 计算机科学系, 福建 福州 350108)

通讯作者: 徐戈, E-mail: xuge@pku.edu.cn

摘要: 为了自动区分中文主观词和客观词, 采用主观性线索和汉字的主观性两种手段对动词和形容词进行主观性度量. 主观性的线索进一步被分成级差(gradability)线索和主体(subject)线索; 根据这些线索, 使用基于图的算法进行评级(ranking). 在汉语主观性词表构建中, 提出使用主体线索和汉字主观性. 5 个标注人员对随机选择的 500 个单词进行主观性标注, 据此构建主客观标准集, 并将其用于各种设置下的实验结果评估. 实验结果显示, 当被标注的单词出现频率较高时, 所提出的方法能够超过或者匹配人工标注. 此外, 尽管文中只使用了无标注的数据, 但还有更多的先验知识(如语义词典等)可以被引入到该方法中.

关键词: 主观性; 无标注语料; 级差线索; 主体线索; 评级; 汉字主观性

中图法分类号: TP391 文献标识码: A

中文引用格式: 徐戈, 蒙新泛, 王厚峰. 采用无标注语料的动词和形容词主观性评级. 软件学报, 2013, 24(5): 1036-1050. <http://www.jos.org.cn/1000-9825/4315.htm>

英文引用格式: Xu G, Meng XF, Wang HF. Subjectivity ranking of verbs and adjectives with an unlabeled corpus. Ruan Jian Xue Bao/Journal of Software, 2013, 24(5): 1036-1050 (in Chinese). <http://www.jos.org.cn/1000-9825/4315.htm>

Subjectivity Ranking of Verbs and Adjectives with an Unlabeled Corpus

XU Ge^{1,2,3}, MENG Xin-Fan^{1,2}, WANG Hou-Feng^{1,2}

¹(Institute of Computational Linguistics, Peking University, Beijing 100871, China)

²(Key Laboratory of Computational Linguistics (Peking University), Ministry of Education, Beijing 100871, China)

³(Department of Computer Science, Minjiang University, Fuzhou 350108, China)

Corresponding author: XU Ge, E-mail: xuge@pku.edu.cn

Abstract: In this paper, aiming to automatically distinguish subjective words from objective ones in Chinese, the study performs a subjectivity ranking test on Chinese adjectives and verbs. The paper exploits subjectivity clues and the subjectivity of Chinese characters. The subjectivity clues are further divided into gradability clues and subject clues. The study then uses graph-based algorithms to calculate the subjectivity originated from subjectivity clues. The subject clues and subjectivity of Chinese characters are novel ideas in such tasks. Five annotators are asked to label subjectivity of 500 words, from which the gold standard is built upon and evaluates rankings in various settings. It is shown that when words to be ranked occur frequently, this approach can outperform or match some human annotators. Furthermore, although the study only an unlabeled corpus, more prior knowledge can be incorporated into the graph-based approach.

Key words: subjectivity; unlabeled corpus; gradability clue; subject clue; ranking; subjectivity of Chinese characters

文献[1]指出, 主观性表达是用来表示观点、情绪、评价、推测等的单词或者短语. 主观性这个概念的一个常用术语是私有状态(private state)——“一个无法被客观观察和验证的状态”.

* 基金项目: 国家自然科学基金(91024009); 国家高技术研究发展计划(863)(2012AA011101); 国家社会科学基金(12&ZD227)

收稿时间: 2011-10-12; 修改时间: 2012-03-08; 定稿时间: 2012-07-03

一些关注主观性分析的工作直接涉及极性,隐含地假设“主观性=正极性+负极性”。例如,大部分的主观词表构建工作实际上都是极性词表的构建,将单词分成正极性和负极性两部分。然而,存在着一些无法推断极性的主观性。比如,尽管“惊讶”可以被用在正极性、负极性、甚至中性场合^[2]，“惊讶”本身不携带极性。许多认知、动机、推测的单词属于这个范畴。

一些构建主观性词典的工作采用了句法分析器^[3]、信息抽取^[4],或者语义词典(比如 WordNet^[5,6])。对于中文而言,这样的资源或工具要么不存在,要么不够充分。比如,我们就没有类似于 WordNet 的可以免费使用的中文语义词典。

在本文中,我们使用无标注的语料对中文动词和形容词进行主观性的度量,考虑了来自上下文的主观性线索(包括级差线索和主体线索)和汉字的主观性。级差线索指程度副词,它们修饰动词或者形容词;主体线索包括人称代词等,通常和形容词或动词出现在一个句子中。我们注意到,中文里,汉字是一个独立的语义单位,因此本身携带主观性。当使用汉字的主观性来计算单词的主观性时,性能得到了提高。

1 研究动机

文献[7]指出,把主观和客观进行区分有如下两个原因:

- (1) 两阶段的方法往往是有效的。也就是说,首先进行主客观的区分,然后进行极性分析。
- (2) 我们也许需要更广义的主观性,诸如动机、思维、推测等,而不仅仅是正负极性。

在其他实验中(未出现在本文中,比如产品评论分析),我们采用了两阶段的处理。即先进行句子主客观的分类,然后对主观句子集合再进行产品特征抽取、极性判定等任务。结果显示,经过主观性的过滤,能够降低后续处理的数据规模,且不损害甚至提高性能。另外,非极性的主观性工作已经有研究者涉及,比如,文献[8]中提出了一种自动发现“许愿”语料中的愿望模板的方法。该方法找出的模板如:“I wish”,“I hope”,“I want”,“if only”,“would be better if”,“would like if”等,然后根据这些模板抽取出人们所期望的主题。

对于汉语而言,已经存在一些情感词表,使用比较广泛的有:

- (1) HowNet 情感词表(<http://www.keenage.com/download/sentiment.rar>)。HowNet 本身是中文本体词典,HowNet 情感词表是在本体词典的基础上构建而成。
- (2) NTUSD(NTU sentiment dictionary)(<http://nlg18.csie.ntu.edu.tw>)。NTUSD 的构建则是以初始的人工标注种子词表(来自中文同义词词典)通过 Sinica Bilingual Ontological WordNet^[9]自动扩充而成。

然而,这些词表都是极性词表,即词表分为正、负极性两部分。虽然可以将其作为主观性词表来使用,但是对于大规模的语料而言,仅仅包含极性词表的词表对主观性单词的覆盖程度仍然不够。在本文的实验结果分析中(见第 5.3 节),我们会对由这两个极性词表合并而成的词表(作为主客性词表的替代)进行分析,并与本文方法构建的词表进行比较。

构建汉语主观性词表是一项非常有挑战性的任务:一方面,主观性的判定对于人而言都是困难的任务,这一点在人工标注一致性分析中已有体现(见第 5.1.2 节);另一方面,无论采用何种方法构建主观性词表,当切换到一个全新的应用领域时,这份词表通常仍然是不够充分的。

鉴于此,我们并不期望本文能够提供一份完整的中文主客观词表。本文的工作是,给定一个无标注语料,对其中包含的所有动词和形容词进行主观性度量。首先,该方法将主观性词表的构建限定在特定的语料上,降低了词表构建的难度,同时又不失实用性;其次,动词和形容词是表示主观性的主要词类,并且与主观性线索的关系相对简单,适合采用自动的方法对单词的主观性进行度量。

2 相关工作

本节首先介绍一些情感词表构建的工作,然后对这些工作进行分析,并指出本文工作的不同之处。

情感词表构建工作中的一部分是通过手工完成的。在汉语中,有王国璋编写的《汉语褒贬义词语用法词典》^[10];在英语中,有 General Inquirer(GI)词表^[11]、OpinionFinder 词表(<http://www.cs.pitt.edu/mpqa/>)等。这些工作

往往是后续情感词表构建的基础,虽然有着较高的质量,但是其规模通常难以满足大规模处理的需要,且对新词、不规则用语(常见于网络)并未收录.近些年,通过已有资源构建情感词表的方法不断出现.下面介绍其中的一些典型工作,了解其大致思路.

文献[12]针对连词,从大规模语料中找出并校验在所连接形容词的正负语义极性上的连词约束,根据连词约束将形容词聚为两类.真实数据上的评测以及模拟实验显示出其较好的性能.文献[13]提出了一种判定句子主观性的方法.文中使用一个浅层句法分析器抽取所有的形容词,然后手工构建了包含 73 个程度副词和名词短语的级差修饰词.文献[6]使用了基于 PageRank 的算法对 WordNet 中的同义词集(synset)作正、负极性的评分.文中把 WordNet 看成是一个图,其中的节点是同义词集,同义词集通过各自的注解建立连接.为了对句子进行主观性的标注,文献[3]通过分布相似度对单词聚类,用少量的人工标注作为种子.这些特征通过极性和级差性进一步地细化.文献[7]介绍了对罗马尼亚语进行单词主客观性分类的实验.文中从英语主观性词典入手,利用双语词典将其翻译成罗马尼亚语,得到了含有 4 983 个词条的罗马尼亚词表.文献[14]介绍了如何使用少量种子主观词、在线词典和少量的裸语料构建主观性词表的方法,同时采用一个自展(bootstrapping)过程,根据相似度对新的候选词进行主观性评分.文献[15]描述了一个集成的自动语义主观性标注的方法.文中利用诸如 WordNet 等词典资源中的层级和领域信息以及其他类型的特征来度量注释的相似度以及语义相关单词的重合程度.为了给荷兰语构建主观性词表,文献[16]使用了类似 PageRank 的算法从一个英语词表的荷兰语翻译结果开始自展,利用 WordNet 中单词关系网络对词典中的单词进行主观性评分.

关于如何构建情感词表的工作还有很多,此处不再细述.对以上工作进行分析,并结合所要处理的任务(构建主观性词表),我们发现了如下的一些问题:

- (1) 目前,情感词表构建的工作主要还是针对极性的词表构建^[3,6,10-13,15,16],专门为区分主客观而构建的词表较少^[7].
- (2) 尽管人工词表的质量较高^[10,11],但是其覆盖率却较低,通常没有收录新词和非规范的用法(如网络表达等).
- (3) 对于从种子词开始进行扩张的情感词表构建方法而言,如何选择种子词以及如何构建单词之间关系都会影响词典的质量.事实上,选择种子词是一个高度经验化的工作,不同的种子词集合得到的结果可能差异很大.

本文采用基于图的信息扩散算法对动词和形容词的主观性(而非极性)进行度量.虽然基于图的方法在上述工作中已经被广泛采用,但本文方法并没有选择主观性种子词,而是选择了主观性线索的种子.相对而言,主观性线索的集合较为封闭,而且不依赖于特定的领域.并且,本文方法完全从无标注语料中挖掘主观性信息,不依赖于任何其他资源,如语义词典等.当然,本方法能够很容易地融合这些语言学资源,但在本文中,为了衡量自动方法的性能,有意不采用无标注语料以外的语言资源.最后,利用汉字的主观性判断单词的主观性对于低频单词而言尤其有效.

3 我们的方法

对目标词(形容词或动词)的主观性计算分为两个部分:(1) 使用主观性线索;(2) 使用汉字的主观性.

3.1 主观性线索

对于一个目标词,主观性线索(subjectivity clues)是那些能够指示该目标词主观性的单词.我们把主观性线索分成级差线索(gradability clues)和主体线索(subject clues),并选择了基于图的算法来对目标词的主观性进行计算.

3.1.1 级差线索

级差线索是程度副词.人们已经注意到,级差性和主观性有着很强的关联.比如,我们可以说“非常高兴”,其中,“高兴”是主观词,而“非常”是程度副词;但是我们几乎不能说“非常塑料的”,其中,“塑料的”是客观词,倾向于不受程度副词的修饰.

文献[3,13]中使用了包含 73 副词和名词短语的程度修饰词集合,然后构建主观性分类器用于句子的主观性分类.文中指出,级差性是很好的主观性指示器.文献[17]指出,“强化(intensification)”能够获得主观性,这是因为说话者的视角被投影到了实体上.文献[18]中使用否定性副词结构(包括 5 个程度副词)来寻找中文情感种子词.

在中文里人们发现,程度副词总是被修饰词的前一个或者后一个词^[19],尤其是出现在被修饰词前的情况占大多数.因此,在构建副词和潜在主观词(本文为形容词或动词)之间关系的时候,只考虑潜在主观词的前一个副词.这些信息在使用级差线索来度量主观性的时候非常有用.

3.1.2 主体线索

第 2 个主观性线索是主体线索,该线索在自然语言处理的主观性分析中受到的关注较少.

文献[20]指出,人称代词是诱导出语言中主观性的第一步.文献[21]研究了不同语言中被用来表达情感的语言资源,提供了一个与情感有关的详细的线索列表.文中指出,人称代词、反身代词、限定词、体、时态、重复都能指示情感.甚至语音信息和篇章结构也能显示说话者的主观性,详见文献[21].主体线索是这样一些单词,它们协助说话者构造对话环境,并把自己假定为一个谈话的主体.人称代词、限定词是典型的主体线索.

文献[22]认为,主观性的表达是那些用来表达观点、情绪、评价、推测等的单词或者短语.基于这个定义可知,在主观性会话中,表达观点、情绪、评价和推测必须首先要存在一个主体.因此,有理由假定主观性更有可能出现在那些具有主体的对话环境中.比如,虽然单词“这个”本身并不直接传达任何主观性,“这个”隐含地指出了人(主体)的存在性,而该主体使用“这个”来指称一个实体.根据我们的假设,此时主观性相对于客观性而言更容易出现.为了检验我们的假设,选择了 5 个中文主观词(高兴、喜欢、愚蠢的、希望、怀疑)和 5 个中文客观词(中国的、塑料的、吃饭、死的、振动)在北京大学汉语语言研究中心(<http://ccl.pku.edu.cn>)提供的搜索工具上进行实验,结果见表 1.

Table 1 An example of how subject clues indicate subjectivity

表 1 主体线索指示主观性的例子

单词	C1	C2	$R=C1/C2$
高兴	951	27 331	0.034 8
喜欢	1 667	28 778	0.057 9
愚蠢的	66	997	0.066 2
希望	2 653	79 802	0.033 2
怀疑	581	12 070	0.048 1
中国的	880	39 741	0.022 1
塑料的	1	115	0.008 7
吃饭	265	10 544	0.025 1
死亡的	75	2 808	0.026 7
振动	22	650	0.033 8

注:C1 为单词和“这个”在固定长度窗口共现的频次;C2 为单词在整个语料中的频次.

从表 1 可以看出,平均来说,主观词的 R 值(表示“这个”与目标单词的相关程度)比客观词的 R 值要大.这个趋势表明,当“这个”(一个主体线索)出现在上下文中的时候,主观词更容易出现.此外,我们有各种主体线索,当它们同时使用的时候,这种趋势会被进一步地加强.

与级差线索相比,主体线索对主观性的指示较弱,但是主体线索更加容易获得.

3.1.3 实验中采用的主观性线索

在实验中,去掉一些对于中文而言不存在或者采用目前的 NLP 工具很难获得的主观性线索(比如,中文不存在词缀问题,比较级无形态变化),并将这些线索进行归纳,见表 2.

Table 2 Subjectivity clues used in our experiments

表 2 实验中所采用的主观性线索

主观性线索	子类
级差线索	程度副词
主体线索	人称代词
	反身代词
	限定词
	叹词
	情态词
	感叹号

3.1.4 基于图的算法

图 1 中使用了 3 种类型的单词(节点):

- 级差线索候选(gradability clue candidates,简称 GCC);
- 主体线索候选(subject clue candidates,简称 SCC);
- 形容词和动词(adjectives and verbs,简称 AV).

AV 节点是被评级的目标单词,由于动词和形容词与主观性线索有着相似的关系,将它们合在一组中.GCC 节点和 SCC 节点是级差线索和主体线索的候选,详见表 2.

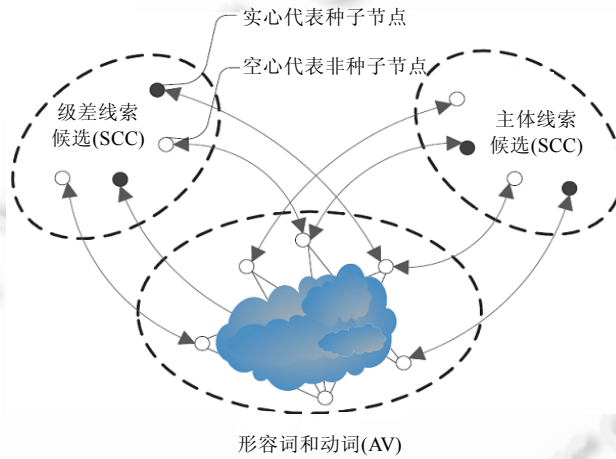


Fig.1 Graph model of subjectivity propagation

图 1 传播主观性的图模型

把图 1 表示为 $G=(V,E)$,其中,节点集合 $V=\{V_{AC},V_{GCC},V_{SCC}\}$,边集合 $E=\{E_{SCC \rightarrow AC},E_{AV \rightarrow SCC},E_{GCC \rightarrow AC},E_{AV \rightarrow GCC},E_{AV}\}$.每个节点集合的大小记为 $n_{GCC}=|V_{GCC}|,n_{SCC}=|V_{SCC}|,n_{AV}=|V_{AV}|$.对于 5 个边集合,定义 5 个矩阵 $\{M_{SCC \rightarrow AC},M_{AV \rightarrow SCC},M_{GCC \rightarrow AC},M_{AV \rightarrow GCC},M_{AV}\}$.矩阵的每个元素代表了节点间的关系.比如, $M_{SCC \rightarrow AC}$ 是 $n_{SCC} \times n_{AV}$ 的矩阵,其元素代表 SCC 节点和 GCC 节点之间的关系; M_{AV} 是 $n_{AV} \times n_{AV}$ 方阵,因为该矩阵表示 AV 节点之间的关系.

各个矩阵的构建过程见第 4.4 节.

接下来,定义 5 个向量(见表 3),向量中的每一个元素代表一个节点的分数.分数的值介于 0~1 之间,1 表示这是一个种子节点.

迭代算法如算法 1 所示.

算法 1. 计算 AV 节点的分数.

Input: $M_{SCC \rightarrow AC},M_{AV \rightarrow SCC},M_{GCC \rightarrow AC},M_{AV \rightarrow GCC},M_{AV}$,种子.

Output: av_score .

1. 使用种子对分数向量 $grad_clue, sub_clue$ 进行初始化
2. repeat
3. $av_grad = MAV \rightarrow GCC \times grad_clue$
4. $av_sub = MAV \rightarrow SCC \times sub_clue$
5. $av_grad = av_grad / \max(av_grad)$
6. $av_sub = av_sub / \max(av_sub)$
7. $grad_clue = MGCC \rightarrow AV \times av_grad$
8. $sub_clue = MSCC \rightarrow AV \times av_sub$
9. $grad_clue = grad_clue / \max(grad_clue)$
10. $sub_clue = sub_clue / \max(sub_clue)$
11. 将分数向量 $grad_clue, sub_clue$ 中种子的分值重新赋值为 1
12. until 满足终止条件
13. $av_score(i) = av_grad(i) \times av_sub(i); i \in [1, \dots, n_{AV}]$

Table 3 Score vectors

表 3 分数向量

节点类型	向量名称	向量维度	分数说明
GCC	$grad_clue$	$n_{GCC} \times 1$	作为级差性线索的质量
SCC	sub_clue	$n_{SCC} \times 1$	作为主体线索的质量
AV	av_grad	$n_{AV} \times 1$	级差性分数
	av_sub	$n_{AV} \times 1$	主体性分数
	av_score	$n_{AV} \times 1$	主观性分数

这里对算法的基本思路进行介绍.根据表 3 中的定义,算法 1 用来计算 AV 节点的主观性分数(av_score).在运行这个算法之前,级差线索和主体线索都已经确定相应的种子词.这些种子词作为人工标注和语言学知识的结果,具有很高的质量.算法 1 的基本思想是:和高质量的主观性线索频繁共现的 AV 节点是高质量的主观词;而和高质量的主观词频繁共现的主观性线索是高质量的线索.初始的高质量主观性线索就是种子,我们使用一个相互强化的方法来对 AV 节点和线索词(GCC 节点和 SCC 节点)同时进行评级.最终,对于第 i 个 AV 节点,当从级差线索和主体线索获得其两个分数(即 $av_grad(i)$ 和 $av_sub(i)$)的时候,返回这两个分数的乘积作为最终第 i 个节点的主观性分数($av_score(i)$).在算法 1 中,第 5 步、第 6 步以及第 9 步、第 10 步是对分数向量的归一化,以保证算法收敛.

在算法 1 中, $grad_clue, sub_clue$ 也是有用的.比如,可以对 $grad_clue$ 进行排序,找出那些分数较高但不是种子词的级差线索候选,然后手工地更新级差线索种子词集合.

一旦用算法 1 获得了 AV 节点的主观性分数,就可以使用 AV ranking 算法在 AV 节点之间传播主观性的分数(见算法 2).其中, Y 是分数向量的初始赋值, M 是归一化的相似度矩阵, α 是权衡初始赋值和相邻节点影响的主观性分数常数.

算法 2. AV ranking.

Input: Y, M, α .

Output: F .

1. $F = Y$
2. repeat
3. $F = (1 - \alpha)M \times F + \alpha Y$
4. until 收敛
5. return F

算法 2 的基本思想是:相似的节点应该有相似的主观性分数.一个节点的主观性分数由其初始赋值和相邻

的那些节点的主观性分数所决定.

简而言之,算法 1 参考了 HITS 算法^[23]对异质的节点(在 AV 节点和 GCC 节点之间以及在 AV 节点和 SCC 节点之间)进行评级,算法 2 参照文献[24]的工作来对同质节点(AV 节点之间)的内部进行分数传播.

3.2 基于汉字的主观性

对于语料中的一个低频目标词(一个 AV 节点),为其抽取出的主观性线索(级差线索和主体线索)往往是不充分的,因此很难计算出有统计意义的主观性分数.在这种情况下,我们转向单词的内部.

一个中文单词通常由 1 个或者多个汉字构成.汉字经常充当独立的语义成分,从而本身也携带主观性.因此,我们希望通过一个单词包含的汉字的主观性来推断这个单词的主观性.

比如,假设不知道“敏慧”是否是一个主观词,但已经知道“敏”和“慧”都是主观的,因此能够推断“敏慧”是一个主观词.

本文使用带约束的线性回归模型来估计汉字和主观性的联系,该模型如下式所示:

$$\min_x |C \cdot x - d|_2 \text{ 且 } 0 \leq x \leq 1.$$

我们将所涉及的汉字个数表示为 n_c ,所涉及的中文单词(所有的动词和形容词)个数表示为 n_w , x 是 n_c 维的向量,而 n_w 维的向量 d 是所有动词和形容词的主观性分数. C 是一个 $n_w \times n_c$ 矩阵,其每一行表示一个单词中汉字的出现情况.更准确地说, $C_{ij}=1$ 表明第 j 个汉字出现在第 i 个中文单词中;否则, $C_{ij}=0$.

一旦获得了 x ,我们就认为它的元素值表示汉字的主观性.给定一个中文单词,返回该单词所包含的汉字的平均主观性作为整个单词的主观性.

下面给出一个例子.假设语料中出现了 6 个单词{敏感、聪慧、感冒、振动、失聪、敏慧},所涉及的汉字是{敏、感、聪、慧、智、冒、失}.表 4 中各个单词的主观性分数 d 来自于前面的基于主观性线索的方法; C 表示单词中汉字的出现情况.

Table 4 An example for calculating the subjectivity of Chinese characters

表 4 计算汉字主观性的例子

单词	d	C						
		敏	感	聪	慧	智	冒	失
敏感	0.925 7	1	1	0	0	0	0	0
聪慧	0.874 6	0	0	1	1	0	0	0
智慧	0.866 3	0	0	0	1	1	0	0
感冒	0.015 2	0	1	0	0	0	1	0
失聪	0	0	0	1	0	0	0	1
敏慧	0	1	0	0	1	0	0	0

对带约束的回归模型 $\min_x |C \cdot x - d|_2$ 进行求解,得到每个汉字的主观性分值,记录在 x 中.对于本例, x 中包含 7 个元素,每个元素对应一个汉字的主观性,其值介于 0~1 之间.对于“敏慧”这个词,可能由于语料中出现的次数较少,没有捕捉到主观性线索,从而主观性分数为 0.但是通过对汉字的主观性计算发现,“敏”(0.1964)和“慧”(0.1606)具有主观性.因此,基于汉字,“敏慧”的主观性修改为 $0.1785 \left(\frac{0.1964 + 0.1606}{2} \right)$.

4 实验设置

本节描述了如何处理语料并构建矩阵以及选择主观性线索种子词的方法.这些种子词是我们的实验中主观性的来源.

4.1 数据集

选择 1997 年~2004 年人民日报的无标注语料作为实验语料.该语料约 700M 字节,包含约 450 万个句子,单词(包括词性标记)空间的规模是 80 000 左右.动词和形容词的数量是 21 126,它们是将要被评级的目标单词.

4.2 分词和词性标注

使用 ICTCLAS(从 <http://www.ictclas.org/>下载)工具包来进行分词和词性标注,采用北京大学语料处理规范^[25].

对于 GCC,SCC 和 AV 节点,按照表 5 选择单词.

Table 5 Selecting nodes based on specification for corpus processing at Peking University

表 5 基于北京大学语料处理规范的节点选择

节点类型	PKU 编码	含义
级差线索候选(GCC)	D	副词
	Ad	副形词
	Vd	副动词
	Dg	副语素
主体线索候选(SCC)	R	代词
	Y	语气词
	!	感叹号
	E	叹词
动词和形容词(AV)	A	形容词
	V	动词

4.3 选择种子词

在本文方法中,所有的主观性都来自 GCC 和 SCC 节点的种子词.因此,给每一类节点选择高质量的种子词非常关键.

4.3.1 选择 GCC 种子词

对于 GCC 节点而言,从文献[19]的工作入手,其中给出了几位不同作者收集的程度词.

此外,我们手工标注了表 5 中的所有 2 597 个副词.在标注过程中,使用 3 个标准来鉴别程度副词:

- (1) 程度副词应该是中性的.因此,“非常”是程度副词,而“恶意地”不是程度副词.
- (2) 程度副词不能连接客观词.即程度副词不能修饰诸如“吃饭”、“奔跑”等客观词,“非常吃饭”是不合理的.
- (3) 如果一个副词能够同时被用作程度副词,并且还有其他含义,则不选入.比如,“特别”在“特别高兴”中是程度副词,而在“特别报导”中具有其他的意思,所以“特别”不收入我们的程度副词集合.

最终,收集到了 110 个程度副词,作为种子节点.

4.3.2 选择 SCC 种子词

对于 SCC 节点,我们将所有的情态词和感叹词都作为种子.情态词的例子如“也好”、“呀”等;感叹词的例子如“哎呀”、“哈哈”等.

对于 SCC 节点中的代词,我们手工选择人称代词、反身代词和限定词作为主体线索的种子词.

感叹号也是种子词.

最后,收集到 263 个 SCC 种子词.

4.4 构建矩阵

对于中文,大多数情况下,程度副词是被修饰词的前一个词(见第 3.1.1 节).因此,当构建 GCC 节点(副词)和 AV 节点(动词和形容词)之间关系的时候,只考虑形容词或者动词的前一个词.

为了使用无标注语料构建 $M_{AV \rightarrow GCC}$,我们进行了如下步骤的操作:首先,在整个语料中,统计一个 GCC 节点有多少次出现在一个 AV 节点之前,这样就有了 $n_{AV} \times n_{GCC}$ 的矩阵 $M'_{AV \rightarrow GCC}$;然后,对于矩阵中的第 i 行(这行记录着第 i 个 AV 节点和 GCC 节点共现的次数),用第 i 个节点在整个语料中的频次除第 i 行的所有元素得到矩阵 $M_{AV \rightarrow GCC}$.应该注意到, $M_{AV \rightarrow GCC}$ 的每一行之和并不是 1,因为 GCC 节点只是第 i 个 AV 节点前能够出现的节点集合的子集.为了得到 $M_{GCC \rightarrow AV}$,将 $M'_{AV \rightarrow GCC}$ 进行转置后,做类似的处理.

对于一个 AV 节点,SCC 节点能够出现的位置更加灵活,因此我们把共现的窗口调整为一个句子.为了构建 $M_{AV \rightarrow SCC}$,在整个语料中,统计一个 SCC 节点和一个 AV 节点有多少次在一个句子中共现,这样得到了 $n_{AV} \times n_{SCC}$ 矩阵 $M'_{AV \rightarrow SCC}$.其他步骤与构建 $M_{AV \rightarrow GCC}$ 或 $M_{GCC \rightarrow AV}$ 类似,最终得到矩阵 $M_{AV \rightarrow SCC}$ 和 $M_{SCC \rightarrow AV}$.

M_{AV} 与上述的 4 个矩阵不同,因为它是相同节点之间的矩阵.本文使用 3 种相似度来构建 M_{AV} .以下 3 个相似度矩阵同维度,即都为 $n_{AV} \times n_{AV}$ 方阵.构建过程简述如下:

- 1) 参照文献[12]的工作,使用连词来获得单词之间的主观相似性.为了避免噪声,要求被一个连词连接的两个单词具有相同的词性.本文选择了 14 个连词,包括“和”、“但”等.初始的相似度矩阵为零矩阵.假设在语料中,出现“聪明但懒惰”3 次;若“聪明”的索引为 i 、“懒惰”的索引为 j ,则该相似度矩阵中的 (i,j) 和 (j,i) 两个元素的值分别增加 3.在语料中,遍历所有的单词对和连词后,得到最终相似度矩阵.
- 2) 文献[26]中提到了 4 种词汇重复,见表 6.本文认为,重复和它的原始单词具有相似的主观性.同样,初始的相似度矩阵为零矩阵.构建此相似度矩阵,只需要单词列表(长度为 n_{AV}),不需要语料.比如说,“红红火火”对应的原始单词为“红火”,若“红红火火”的索引为 i 、“红火”的索引为 j ,则该相似度矩阵中的 (i,j) 和 (j,i) 两个元素的值分别增加 1.遍历所有的单词对后,得到最终相似度矩阵.
- 3) 本文使用目标词的前一个词和后一个词作为目标词的特征来计算任意两个目标词(AV 节点)之间的相似度,这个相似度是对分布相似度的简化和近似.对任意一个单词 X 而言,有一个特征向量 W 与之对应(该特征向量的长度为 $2 \times n_{AV}$).向量的前半半记录语料中 X 的前一个单词的分布情况.即,若单词 W_i 在语料中出现在 X 的左侧且相邻的次数为 n ,则特征向量的第 i 个元素为 n ;采用相同的方式,向量的后半半记录语料中 X 的后一个单词的分布情况.对于任意两个单词,其相似度为两个单词对应的两个特征向量的余弦相似度.

Table 6 Four types of duplication

表 6 4 种词汇重复

重复类型	例子
AB-AABB	红火-红红火火
A-AA	暗-暗暗
AB-ABAB	通红-通红通红
AB-A 里 AB	罗嗦-罗里罗嗦

当以上矩阵构建完成以后,进行对称归一化^[24]并将它们线性组合^[27].

5 评测

5.1 人工标注

5.1.1 标注方案

我们从 21 126 个动词和形容词中随机选择了 500 个,然后让 5 个标注者(X, M, H, P 和 W)以 3 个标记进行主观性标注.3 个标记为:

- SS(strongly subjective):强主观(2);
- WS(weakly subjective):弱主观(1);
- OB(objective):客观(0).

注意,这里的 0,1,2 是标注时采用的符号,同时也可以看成是主观性的得分.

本文提供两个主观性的定义来辅助标注工作:

- 主观性是一种无法被客观观测和验证的心理状态;
- 主观表达是用来表示观点、情绪、评价、推测等的单词和短语^[22].

为了帮助标注者区分强主观和弱主观,我们从 OpinionFinder 词表(<http://www.cs.pitt.edu/mpqa/>)中选择了一些动词和形容词的强主观和弱主观词作为参考.另外,如果一个单词有多个语义,考虑其主观性最强的那个

语义.

5.1.2 标注一致性分析

表 7 是在 3 个标记上 5 个标注者的标注一致性分析.在表 8 中,把 SS 标记和 WS 标记合成一个标记(subjective),这样只有 2 个标记(subjective 和 objective).

Table 7 Agreement on three tags (SS, WS, and OB)

表 7 在 3 个标记(SS,WS 和 OB)上的一致性

		<i>X</i>	<i>M</i>	<i>H</i>	<i>P</i>	<i>W</i>
<i>X</i>	A	1.00	0.59	0.63	0.61	0.66
	κ	1.00	0.24	0.41	0.41	0.40
<i>M</i>	A	0.59	1.00	0.58	0.47	0.80
	κ	0.24	1.00	0.23	0.14	0.53
<i>H</i>	A	0.63	0.58	1.00	0.53	0.60
	κ	0.41	0.23	1.00	0.28	0.30
<i>P</i>	A	0.61	0.47	0.53	1.00	0.53
	κ	0.41	0.14	0.28	1.00	0.25
<i>W</i>	A	0.66	0.80	0.60	0.53	1.00
	κ	0.40	0.53	0.30	0.25	1.00

A:一致性; κ :Kappa 值;5 个标注者:*X,M,H,P* 和 *W*.

Table 8 Agreement on two tags (subjective and objective)

表 8 在 2 个标记(subjective 和 objective)上的一致性

		<i>X</i>	<i>M</i>	<i>H</i>	<i>P</i>	<i>W</i>	<i>Gold</i>
<i>X</i>	A	1.00	0.73	0.77	0.75	0.79	0.90
	κ	1.00	0.44	0.53	0.51	0.57	0.81
<i>M</i>	A	0.73	1.00	0.66	0.63	0.85	0.79
	κ	0.44	1.00	0.33	0.32	0.62	0.54
<i>H</i>	A	0.77	0.66	1.00	0.70	0.68	0.81
	κ	0.53	0.33	1.00	0.39	0.37	0.62
<i>P</i>	A	0.75	0.63	0.70	1.00	0.69	0.80
	κ	0.51	0.32	0.39	1.00	0.41	0.61
<i>W</i>	A	0.79	0.85	0.68	0.69	1.00	0.84
	κ	0.57	0.62	0.37	0.41	1.00	0.67
<i>Gold</i>	A	0.90	0.79	0.81	0.80	0.84	1.00
	κ	0.81	0.54	0.62	0.61	0.67	1.00

A:一致性; κ :Kappa 值;5 个标注者:*X,M,H,P* 和 *W*;

Gold 是 5 个标注结果混合而成的主客观标准集.

通过比较表 7 和表 8 我们发现,区别强主观和弱主观比起区别主观和客观的难度更大,对于人而言都是比较困难的任务.这种认识也促使在构建主客观标准集时使用 2 个标记(subjective 和 objective)而不是 3 个(SS,WS,和 OB).

5.1.3 构建主客观标准集

经过 5 个标注者人工标注后,得到了 5 个标注列表,列表的每个元素是 0(OB),1(WS)或者 2(SS).尽管事先给出一些标注指导,但是不同标注者的标注结果仍然存在很大的差异,这体现出不同标注者对主观性的不同理解.从表 9 中可以很容易地看出,标注者 *M* 倾向于把单词标为客观的,而标注者 *X* 具有较为平衡的分布.

为了得到更为一般性的主观性标准,我们把 5 个标注列表加在一起,得到一个混合的列表(其中,每个元素的值是介于 0~10 之间的整数).将值大于等于 3 的元素对应的单词定为主观词,把值小于 3 的元素对应的单词定为客观词,这样就得到了一个元素取值二元(主观/客观)的列表.该列表即为实验中的主客观标准集(Gold).它与其他 5 个标注者的一致性见表 8.从表中可以看出,Gold 是一个更加一般性的主观性标注,因为它与其他标注者的一致性更高.

Table 9 Agreement of annotator X and M 表 9 标注者 X 和 M 的一致性

	OB	WS	SS	总计(X)
OB	250	14	0	264
WS	89	42	2	133
SS	34	66	3	103
总计(M)	373	122	5	500

5.2 评测指标

实验中采用两个评测指标: τ_p (p -normalized Kendall tau distance) 和 $MaxAcc$.

5.2.1 τ_p

对于 τ_p , 一个评级 (ranking) 被定义为一组对象上的偏序. 任意两个对象, 要么有序 (ordered), 要么平手 (tied). 如果两个有序对 (pair) 涉及的对象相同但是序不同, 则称这两个对为不和谐对 (discordant pairs). τ_p 定义为

$$\tau_p = \frac{n_d + p \times n_u}{Z}$$

其中,

- n_d : 预测评级和标准评级之间不和谐对的个数;
- n_u : 在标准评级中有序, 但在预测评级中是平手的对的个数;
- Z : 标准评级中有序对的个数;
- p : 对平手的惩罚因子.

我们把 p 设置成 0.5, 与文献[5,10]的工作保持一致, 这相当于假设平手意味着 50% 的不和谐对. τ_p 的取值从 0 (完全匹配) 到 1.

给定一个标准评级, τ_p 适合用来对其他评级进行评估. 即分数越接近于 0, 则与标准评级越匹配. 然而, τ_p 只是反映两个评级 (ranking) 中的有序对的一致程度, 其结果对于读者而言并不直观.

5.2.2 $MaxAcc$

为了对评级有直观的评价, 我们希望把预测评级转化为二元标记 (主客和客观) 的列表, 然后比较其与主客观标准集的一致程度. 在将预测评级转换为主客观两部分时, 需要选择一个门限值来对预测评级进行切分. 为了避免人工选择门限值, 本文采用 $MaxAcc$ 来进行评估.

$MaxAcc$ 的含义是: 以主客观标准集为正确分类, 当调节门限值的大小时, 能够达到的最大准确度.

比如, 对于单词列表 {A,B,C} 的预测评级为 {0.7,0.3,0.9}; 标准答案为 {1,0,1}, 1 代表主观, 0 代表客观, 即, A,C 为主观单词, B 为客观单词.

按照预测评级排序后, 主观性从高到低依次为 C,A,B, 则可能切分出的结果分别是:

- (1) C,A,B 均为客观词, 此时正确率 (accuracy) 为 33.3%;
- (2) C 为主观词, A,B 为客观词, 此时正确率为 66.6%;
- (3) C,A 为主观词, B 为客观词, 此时正确率为 100%;
- (4) C,A,B 均为主观词, 此时正确率为 66.6%.

对于该预测评级, $MaxAcc$ 为正确率最高的情况, 即 100%. 该结果说明, 该预测评级对于主客观的度量与标准答案一致, 当将预测评级从某个门限值划分成主观和客观两个部分时, 与标准答案相同.

5.3 实验结果

在表 10 中:

- baseline 将所有的单词分类成客观词 (在随机选择的 500 个词中有 284 个客观词).
- hownet&ntusd: 将第 1 节中的两部常用汉语极性词表合并后去重, 得到共计含 17 488 个单词的词表. 将出现在该词表中的单词视为主观词, 而将未出现在此词表中的单词视为客观词.

- gradability:表示使用级差(gradability)线索进行主观性度量.
- subject:表示使用主体(subject)线索进行主观性度量.
- gradability&subject:表示同时使用级差线索和主体线索进行主观性度量.
- gradability&subject&char:表示在使用级差线索和主体线索进行主观性度量之后,再使用汉字的主观性对单词的主观性进行度量.
- AV ranking:在获得了单词的主观性以后,再采用算法 2 在 AV 节点之间进行主观性传播.

Table 10 Results of experiments

表 10 实验结果

方法类型	无 AV ranking		有 AV ranking	
	τ_p	MaxAcc	τ_p	MaxAcc
baseline	-	0.568	-	-
hownet&ntusd	-	0.714	-	-
gradability	0.395	0.670	0.393	0.680
subject	0.326	0.654	0.273	0.692
gradability&subject	0.371	0.694	0.367	0.696
gradability&subject&char	0.259	0.728	0.258	0.728

根据第 5.2.2 节的定义, τ_p 值越小,说明预测评级和主客观标准集(Gold)越一致.虽然 τ_p 对两个评级之间的相似性描述得更加精细,但是与 MaxAcc 相比却不够直观.在接下来的分析中,主要采用 MaxAcc 作为评测指标.

从表 10 可以看出,就 MaxAcc 而言,级差线索(0.670)和主体线索(0.654)都是有效的,与 baseline(0.568)相比有明显的提升.把两个分数乘起来作为主观性分数(0.694),比单独使用两个分数要好.

采用汉字的主观性能够提升性能(对于 MaxAcc,从 0.694 到 0.728).汉字主观性和词频的关系将在第 5.4.1 节中进行分析.

对于 AV ranking 方法,如果单独使用级差线索或者主体线索得到的分数,可以看到 1%(从 0.670 到 0.680)或 3.8%(从 0.654 到 0.692)的 MaxAcc 的提升.但是我们也注意到,gradability&subject 和 AV ranking 的组合并没有能够带来 MaxAcc 的进一步提升.一种可能的原因是,AV ranking 所使用的相似度矩阵质量还有待提高.

hownet&ntusd 方法也比 baseline 有明显的提高(从 0.568 到 0.714),但是有几点需要注意:

- 采用 hownet&ntusd 进行比较是一种权宜之计.目前来讲,并不存在公认的区分主观和客观的词表.为了与本文提出的方法比较,只能将已有的两个常用极性词表中的词都看成是主观词,而把其余单词看成是客观词.
- 虽然 hownet&ntusd 的词表规模已经不算小(含 17 488 个单词),但测试集中的 500 个单词只有 107 个出现在该词表中.因此,hownet&ntusd 词表用来进行主、客观分类的正确率严重受到语料中单词规模的影响.考虑极端的情况,假设所有的测试集单词都不出现在 hownet&ntusd 词表中,那么它们都会被判定为客观词,这种判断退化成了 baseline.而我们提出的方法能够对语料中出现的任意一个形容词或动词进行主观性的度量,并且能够保证达到明显高于 baseline 的准确度.
- hownet&ntusd 作为主要由人工构建的词典,其主观性判定受频率影响较小.而本文方法当被判定词在语料中出现的频率较大时,主观性度量的可靠性会进一步升高(见第 5.4.2 节),可以达到 80%左右.

5.4 词频的影响

下面,我们分析词频对预测评级性能的影响.在图 2 和图 3 中,采用最小词频(minimum word frequency)从 500 个单词中选择单词.比如,如果最小词频是 20,则只考虑那些在语料中词频大于等于 20 的单词.

5.4.1 汉字主观性与词频的关系

在图 2 中,当最小词频小于 20 时,使用汉字的主观性能够达到较好的效果;当最小词频大于等于 20 时,不使用汉字,主观性表现得更好.

这说明,当一个词出现的频率较小时,其统计信息不是十分稳定,需要通过汉字的主观性来辅助;当一个词

频繁出现时,其统计信息足够估计其主观性.

比如,“敏慧”只在整个语料中出现两次,来自级差线索的分数(0)和来自主体线索的分数(0.035 9)都很低.但是,许多包含“敏”或者“慧”的其他单词具有较高的主观性.这样,“敏”和“慧”这两个汉字相应地也有较强的主观性.当使用这两个汉字的平均主观性作为“敏慧”的主观性时,“敏慧”在整个目标词列表(21 126 个动词和形容词)中排名 152,表明其具有很强的主观性.

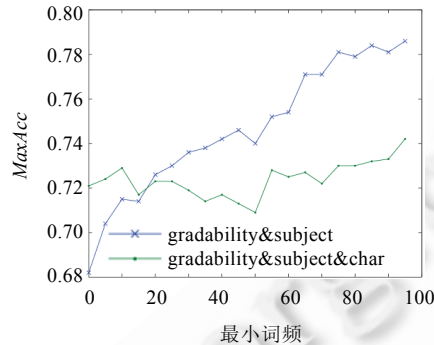


Fig.2 Subjectivity of Chinese characters

图2 汉字主观性的影响

5.4.2 本文方法与人工标注的比较

图3 考察当最小词频改变时,本文的自动方法(gradability&subject)与人工标注者的比较.

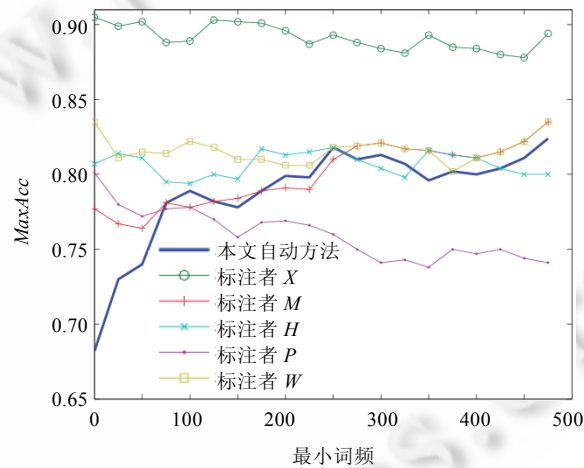


Fig.3 Comparison between automatic method and human annotation

图3 自动方法和人工标注的比较

我们观察到,人工标注对于词频不敏感(5 个标注者的曲线差不多是平的),这是因为人工标注者是通过词义判断单词的主观性,而不是单词的频率.与此对照,本文的自动方法当最小词频上升时性能变好.这说明,拥有的统计信息越多,自动方法的效果越好.

当最小词频超过 250 时,本文方法比标注者 *P* 要好,与标注者 *W*, *M* 和 *H* 相当,仅仅比标注者 *X* (本文第一作者,对于主观性有一般性的理解)差.

以上信息具有一定的实用性.这说明,对于中文动词和形容词,只要抽取足够的上下文,就能自动地对单词的主观性进行较高质量的标注,而且性能超过或者和一些人工标注者相当.

6 结论与未来的工作

本文提出了采用无标注语料对其中出现的形容词与动词进行主观性度量的方法.该方法有如下优点:

- (1) 只依赖大规模的无标注语料,不需要复杂的 NLP 工具;
 - (2) 所选择的主观性线索,诸如人称代词、程度副词等,从语料中很容易获得.汉字的主观性在统计信息偏少的时候尤其有效;
 - (3) 单词主观性以及单词相似度的先验知识能够很容易被本文方法所采用;
 - (4) 如果目标词在所采用的无标注语料中频繁出现,本文方法能够超过或者匹配人工标注者.
- 将来,我们将会根据主观性对中文名词进行评级,这是一项更有挑战性的任务.

致谢 在此,我们向对本文的工作提出了各项宝贵意见和建议的匿名评审者表示感谢.通过对他们所指出的错误和不足进行反复修改,本文得到了进一步的完善.

References:

- [1] Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 2005,39(2-3):165–210. [doi: 10.1007/s10579-005-7880-9]
- [2] Ortony A, Turner TJ. What's basic about basic emotions! *Psychological Review*, 1990,97(3):315–331.
- [3] Wiebe J. Learning subjective adjectives from corpora. In: Kautz HA, ed. *Proc. of the American Association for Artificial Intelligence*. Cambridge: MIT Press, 2000. 735–740.
- [4] Riloff E, Wiebe J. Learning extraction patterns for subjective expressions. In: *Proc. of the Conf. on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2003. 105–112. [doi: 10.3115/1119355.1119369]
- [5] Kim SM, Hovy E. Determining the sentiment of opinions. In: *Proc. of the Int'l Conf. on Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2004. Article No.1376. [doi: 10.3115/1220355.1220555]
- [6] Esuli A, Sebastiani F. Pageranking wordnet synsets: An application to opinion mining. In: *Proc. of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2007. 424–431.
- [7] Mihalcea R, Banea C, Wiebe J. Learning multilingual subjective language via cross-lingual projections. In: *Proc. of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2007. 976–983.
- [8] Goldberg AB, Fillmore N, Andrzejewski D, Xu ZT, Gibson B, Zhu XJ. May all your wishes come true: A study of wishes and how to recognize them. In: *Proc. of the HLT-NAACL*. Stroudsburg: Association for Computational Linguistics, 2009. 263–271.
- [9] Ku LW, Chen HH. Mining opinions from the Web: Beyond relevance retrieval. *Journal of the American Society for Information Science and Technology*, 2007,58(12):1838–1850. [doi: 10.1002/asi.20630]
- [10] Wang GZ. *A Dictionary of Chinese Praise and Blame Words*. Beijing: Sinolingua, 2001 (in Chinese).
- [11] Stone PJ. *The General Inquirer: A Computer Approach to Content Analysis*. Cambridge: The MIT Press, 1966.
- [12] Hatzivassiloglou V, McKeown K. Predicting the semantic orientation of adjectives. In: *Proc. of the Joint ACL/EACL Conf.* Stroudsburg: Association for Computational Linguistics, 1997. 174–181. [doi: 10.3115/976909.979640]
- [13] Hatzivassiloglou V, Wiebe J. Effects of adjective orientation and gradability on sentence subjectivity. In: *Proc. of the Int'l Conf. on Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2000. 299–305. [doi: 10.3115/990820.990864]
- [14] Banea C, Mihalcea R, Wiebe J. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In: *Proc. of the Int'l Conf. on Language Resources and Evaluation*. Paris: European Language Resources Association, 2008. 2764–2767.
- [15] Gyamfi Y, Wiebe J, Mihalcea R, Akkaya G. Integrating knowledge for subjectivity sense labeling. In: *Proc. of the HLT-NAACL*. Stroudsburg: Association for Computational Linguistics, 2009. 10–18.
- [16] Jijkoun V, Hofmann K. Generating a non-English subjectivity lexicon: Relations that matter. In: *Proc. of the 12th Conf. of the European Chapter of the Association for Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2009. 398–405.

- [17] Athanasiadou A. On the subjectivity of intensifiers. *Language Sciences*, 2007,29(4):554–565. [doi: 10.1016/j.langsci.2007.01.009]
- [18] Zagibalov T, Carroll J. Automatic seed word selection for unsupervised sentiment classification of Chinese text. In: *Proc. of the 22nd Int'l Conf. on Computational Linguistics*. Stroudsburg: Association for Computational Linguistics, 2008. 1073–1080.
- [19] Zhao J. The study on the highest degree adverbs in mandarin [MS. Thesis]. Shanghai: Shanghai Normal University, 2006 (in Chinese).
- [20] Benveniste E. Subjectivity in language. In: Gables C, ed. *Proc. of the Problems in General Linguistics*, Vol.18. Miami: University of Miami Press, 1958. 223–230.
- [21] Ochs E, Schieffelin BB. Language has a heart. *Text*, 1989,9(1):7–25.
- [22] Wiebe JM, Riloff E. Creating subjective and objective sentence classifiers from unannotated texts. In: *Proc. of the Conf. on Computational Linguistics and Intelligent Text Processing*. Mexico: Springer-Verlag, 2005. 475–486. [doi: 10.1007/978-3-540-30586-6_53]
- [23] Kleinberg J. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 1999,46(5):604–632. [doi: 10.1145/324133.324140]
- [24] Zhou D, Bousquet O, Lal TN, Weston J, Scholkopf B. Ranking on data manifolds. In: *Proc. of the 18th Annual Conf. on Neural Information Processing System*. Cambridge: MIT Press, 2003. 169–176.
- [25] Yu SW, Duan HM, Swen B, Chang BB. Specification for corpus processing at Peking university: Word segmentation, pos tagging and phonetic notation. *Journal of Chinese Language and Computing*, 2003,13(2):121–158 (in Chinese with English abstract).
- [26] Zhu DX. *Lecture Notes on Chinese Grammar*. Beijing: The Commercial Press, 1982 (in Chinese).
- [27] Tong H, He J, Li M, Zhang C, Ma W. Graph based multi-modality learning. In: *Proc. of the 13th Annual ACM Int'l Conf. on Multimedia*. New York: ACM Press, 2005. 862–871. [doi: 10.1145/1101149.1101337]

附中文参考文献:

- [10] 王国璋.汉语褒贬义词语用法词典.北京:华语教学出版社,2001.
- [19] 赵军.极性程度副词研究[硕士学位论文].上海:上海师范大学,2006.
- [25] 俞士汶,段慧明,朱学锋,孙斌,常宝宝.北大语料库加工规范:切分·词性标注·注音.汉语语言与计算学报,2003,13(2):121–158.
- [26] 朱德熙.语法讲义.北京:商务出版社,1982.



徐戈(1978—),男,浙江淳安人,博士,讲师,主要研究领域为自然语言处理,情感分析.
E-mail: xuge@pku.edu.cn



王厚峰(1965—),男,博士,教授,博士生导师,CCF高级会员,主要研究领域为自然语言处理.
E-mail: wanghf@pku.edu.cn



蒙新泛(1984—),男,博士生,主要研究领域为自然语言处理,情感分析.
E-mail: mxmf@pku.edu.cn