

# 基于领域本体、情感词典的商品评论倾向性分析

董丽丽 赵繁荣 张翔

(西安建筑科技大学信息与控制工程学院 陕西 西安 710055)

**摘要** 文本倾向性分析已成为当前自然语言处理领域的研究热点,其研究成果具有极高的应用价值。针对网络在线中文评论的特点,基于领域本体与情感词典对商品评论倾向性进行分析。其主要思想是首先构建面向商品论坛的领域本体;其次利用情感词典与上下文极性算法计算情感词极性;再次通过将本体与 SBV 算法相结合,实现评价对象和评价词的二元组抽取;最后完成句子的倾向性分析。实验结果表明,有效提高了句子级倾向性分析的准确率。

**关键词** 倾向性分析 领域本体 情感词典 上下文极性 评价对象抽取

中图分类号 TP391.1 文献标识码 A DOI:10.3969/j.issn.1000-386x.2014.12.024

## ANALYSING PROPENSITY OF PRODUCT REVIEWS BASED ON DOMAIN ONTOLOGY AND SENTIMENT LEXICON

Dong Lili Zhao Fanrong Zhang Xiang

(School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China)

**Abstract** Text propensity analysis has currently become the focus of research in natural language processing field, its research results have the extremely high application value. Aiming at the characteristics of online internet Chinese reviews, in this paper we analyse the tendencies of product reviews based on domain ontology and sentiment lexicon. Our main idea is to build the product forum-oriented domain ontology first. Then we calculate the polarities of sentiment words by using the sentiment lexicon and context polarity algorithm. Thirdly, by combining domain ontology with SBV algorithm we realise the extraction of 2-tuple of evaluation object and evaluation words. Finally, we complete the analysis of sentence propensity. Experimental results show that it improves the accuracy of propensity analysis in sentence level effectively.

**Keywords** Propensity analysis Domain ontology Sentiment lexicon Context polarity Evaluation object extraction

## 0 引言

随着互联网的迅速发展,面对浩如烟海的中文商品评论信息,如何自动识别商品评论的情感倾向,并从中获取有价值的信息,成为了自然语言处理领域中亟待解决的问题,因此,文本倾向性分析成为计算机领域的研究热点之一。

文献[1]将文本倾向性分析研究方法分为基于简单统计、基于相关性分析与基于机器学习三种。基于简单统计方面,主要通过求合法、向量空间法等对情感值进行简单的倾向性统计, Tsou 等<sup>[2]</sup>通过对极性元素的分布、密度及语义强度等进行分析,实现相关新闻的褒贬分类。基于相关性方面, Yi 等<sup>[3]</sup>使用语法分析器对句子进行语法分析,同时结合了情感词典对文本进行分类。朱嫣岚等<sup>[4]</sup>基于 HowNet,提出了基于语义相似度和基于语义相关场两种词汇语义倾向性计算方法。在基于机器学习的研究方面, Wang 等<sup>[5]</sup>选取形容词和副词作为特征,提出了基于启发式规则与贝叶斯分类技术相融合的评论句子语义倾向分类方法。徐琳宏等<sup>[6]</sup>通过计算词汇与基准词间的相似度,进而利用支持向量机分类器对文本的褒贬性进行分析。

通过对现有文献进行分析,可以发现目前针对特定领域的网络论坛商品评论的倾向性分析技术仍然存在以下问题:1) 针

对中文的情感词典相对较少,尤其是领域情感词典,诸如较为流行的 HowNet(知网)等在构建情感词典的过程中未考虑评价对象对情感极性的影响,也忽略了网络语言灵活性的特点;2) 基于语义规则的评价对象抽取方法还只能找出频繁的评价对象,对于非频繁评价对象抽取的准确率较低,而且可扩展性较差。

针对以上问题,本文以笔记本电脑汉语评论为主题,首先通过构建领域本体获取评论中规范的语义信息;其次给出了如何设计并利用组合情感词典与改进的情感词上下文极性算法计算情感词极性的方法,以降低统计方法在语义极性计算方面的不确定性;再次利用本体与 SBV 算法相结合来实现评价对象和评价词的二元组抽取,旨在提高评价对象抽取方法的准确率与可扩展性;最后完成句子的倾向性分析。

## 1 商品评论本体的构建

商品评论往往针对商品的某个或某些特征。例如,“神舟精盾外观大气,配置高,CPU 性能强劲”一句,其特征可能是商

收稿日期:2013-09-28。国家自然科学基金项目(51278400);陕西省自然科学基金项目(2012JM8042);陕西省教育厅专项科研项目(12JK0940);技术转移促进工程项目(CXY1348-(1))。董丽丽,教授,主研领域:数据挖掘,分布式系统。赵繁荣,硕士生。张翔,副教授。

品的某一从属属性(CPU),也可能是该商品的一个自然属性(外观、配置等),且由句法分析可知,“强劲”一词指的是 CPU 而非整机的性能。即若一个句子中同时出现多个或多种属性特征词,只有在了解特征词之间从属关系的基础上,才能得到评论准确的倾向性。因此,本文通过构建领域本体,实现了保存商品特征及其之间的从属分类关系树形结构,从而可通过遍历树获取特征词及其之间的关系。

领域本体的构建主要包括领域概念获取、概念间关系获取以及公理获取等三个部分。常用的构建领域本体的方法有:骨架法<sup>[7]</sup>、Gruninger & Fox 评价法(TOVE)<sup>[8]</sup>、Methodology 本体方法<sup>[9]</sup>及七步法<sup>[10]</sup>等。本文结合网络在线商品评论的特点,以笔记本电脑领域为例,采用七步法,利用 Protégé 实现简易笔记本电脑本体的构建。图 1 为本体的创建流程图。

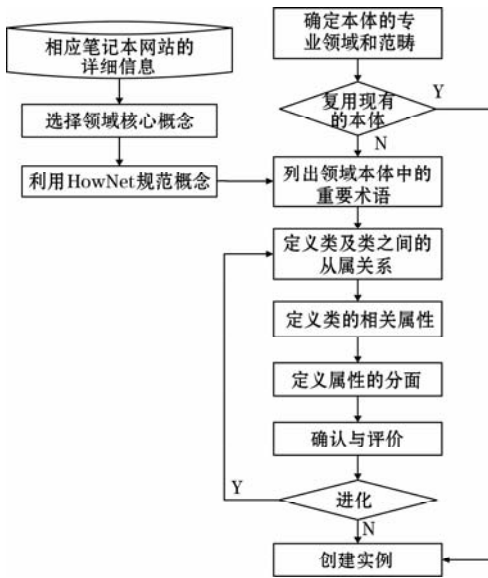


图 1 商品评论本体的创建流程图

其中,获取的领域概念包括实体与属性两部分。具体的,实体指评论中的主体对象,如 ThinkPad T430,其分类模型构成森林;属性指实体的组成与特征,如键盘与舒适性,其分类模型构成树。最终构建的笔记本电脑本体示意性片段如图 2 所示(其中, is\_a 表示对象属于该类, a\_part\_of 表示对象从属于该类的一部分, W(i) 为该对象在该类中权值)。

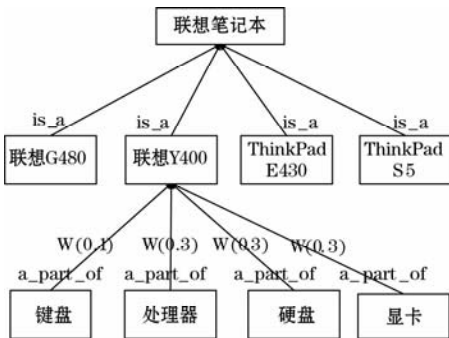


图 2 笔记本本体示意性片段

将领域本体融入到倾向性分析中,相比统计学的方法,可以获取词语间更多的语义信息。因为叶子节点的极性与父节点极性之间是相互影响的,而领域本体的构建正可以反映它们之间的从属关系。进一步的,本文将评价粒度限定在叶子节点上,并为每一个分类模型中的结点赋予一个权值,如果需要计算本体中某结点的极性值,只需要将其子结点的极性值加权,如式(1):

$$\text{TopicPolarity}(\text{父结点}) = \sum_{i=1}^n \text{TopicPolarity}(\text{子结点 } i) \times \text{Weigh}(\text{子结点 } i) \quad (1)$$

$$\text{Weight}(\text{结点 } i) = \text{hits}(\text{结点 } i) / \sum_{i=1}^n (\text{结点 } i) \quad (2)$$

其中, TopicPolarity(note) 为 note 结点的极性值; Weigh(note) 为 note 结点的权值; hits( $\omega$ ) 的值等于在搜索引擎中查询词语  $\omega$  所返回的页面数。结合图 2 可知:

$$\begin{aligned} \text{TopicPolarity}(\text{联想 Y400}) &= \text{TopicPolarity}(\text{键盘}) \times \text{Weicht}(0.1) + \\ &\quad \text{TopicPolarity}(\text{处理器}) \times \text{Weicht}(0.3) + \\ &\quad \text{TopicPolarity}(\text{硬盘}) \times \text{Weicht}(0.3) + \\ &\quad \text{TopicPolarity}(\text{显卡}) \times \text{Weicht}(0.3) \end{aligned}$$

## 2 组合情感词典

情感词的褒贬色彩和极性强度是我们关心的两个重要属性。目前,极性计算方法主要有基于语料库和基于词典两种方法<sup>[11]</sup>。基于语料库的方法在计算情感词的极性时依赖语料库的构建,对于在语料库中出现概率较低的情感词,Turney 和 Littman<sup>[12]</sup>的 PMI 等方法无法得到其正确极性。基于词典的方法多利用 HowNet 的语义相似度及语义相关场判别词语情感极性,但义原的平等性无法客观保证。

本文采用的情感词典由四部分组成:静态情感词典、动态情感词典、修饰词词典及关联强度词典。其中,前三者均以 0 为界限来区分情感词的褒和贬义。如果极性值为负数,则表示该词持贬义情感,否则为褒义,且绝对值越大表示情感越强烈。关联强度词典用以量化情感强度。为了便于描述情感词  $w$  的极性计算,本文对相关概念进行如下定义:

**定义** 语义倾向:由词语对某一事物、观点所表现的褒贬态度计算出的度量值,包括极性(词语的褒贬性)和强度(褒贬的程度)两个属性,本文用二元组来表示:

$$WS(w) = (P, I) \quad P \in \{-1, 0, 1\} \quad I \in R^+ \quad (3)$$

其中, P 中的 1、0、-1 分别表示词语褒、中、贬的情感倾向;属性 I 的大小表示褒贬感情的强弱。

### 2.1 静态情感词典

静态情感词是指具有绝对褒贬倾向的情感词,例如时尚耐用等。考虑到网络语言的灵活性,本文构建的静态情感词典又进一步分为基础情感词典、网络情感词典及未登录情感词典等三部分,如图 3 所示。

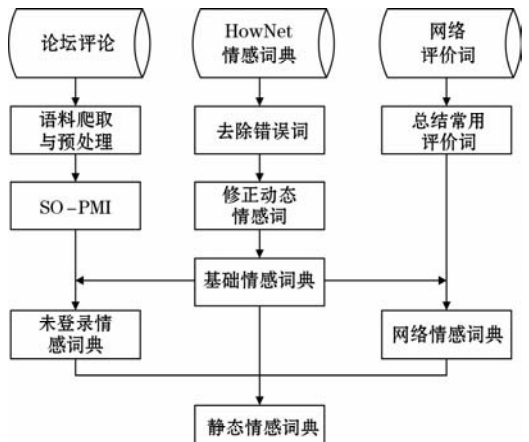


图 3 静态情感词典的构建

其中,基础情感词典针对网络用语特点,选择 HowNet 常用

情感词进行转换并人工修正得到;网络情感词典根据对 5526 条语料中使用的频率较高的情感词进行统计得到;未登录情感词典的构建通过选取基础情感词典为基准词,利用基于点互信息的情感倾向算法(SO-PMI)<sup>[13]</sup>计算候选词与基准词的相似度及识别候选词语的褒贬倾向与极性值。静态情感词典的收录情况如表 1 所示。

表 1 静态情感词典收录统计

部分词典	情感	词语(个数)
基础情感词典	正面	2363
	负面	2040
网络情感词典	正面	27
	负面	19
未登录情感词典	正面	134
	负面	102

## 2.2 动态情感词典

动态情感词是指当修饰目标不同时,其倾向会发生变化的情感词。例如:情感词“高”,在“配置高”中表现出褒义态度,在“耗电高”中则表现出贬义态度。正因为动态情感词的倾向取决于其所修饰的目标词,为充分考虑评价对象对情感极性的影响,本文构造了动态情感词典,通过把动态情感词和评价对象看作一个整体,作为动态极性值计算的标准,以期获得更高的准确率。

本文利用笔记本电脑相关论坛,通过手工并结合笔记本电脑领域常识共统计出动态情感词 32 个,并用集合  $DW$  表示。

首先,将动态情感词词典集合  $DW$  分为正趋势动态集合  $PDW$ ,例如“多”、“高”等;和负趋势动态集合  $NDW$ ,例如“少”、“低”等。他们之间的关系满足下式:

$$\begin{aligned} PDW \cap NDW &= \emptyset \\ PDW \cup NDW &= DW \end{aligned} \quad (4)$$

其次,对于主题词  $t$  的倾向值( $P_t, I_t$ )采用 SO-PMI 方法进行计算,对于动态情感词搭配的倾向计算规则如式(5)所示:

$$DynamicPolarity(\omega, t) = \begin{cases} P = P_t, I = I_t & P_t \neq 0 \text{ 且 } \omega \in PDW \\ P = -P_t, I = I_t & P_t \neq 0 \text{ 且 } \omega \in NDW \\ P = 0, I = 0 & \text{others} \end{cases} \quad (5)$$

其中,  $DynamicPolarity(\omega, t)$  表示情感词  $\omega$  关于主题词的动态极性值。由式(5)可知,在主题词  $t$  具有情感倾向的基础上,对于动态情感词  $\omega$ ,若  $\omega$  属于  $PDW$  集合,其情感倾向  $P_\omega$  及极值大小  $I_\omega$  与主题词  $t$  相同;若  $\omega$  属于  $NDW$  集合,其极性大小  $I_\omega$  与  $I_t$  相同,而  $P_\omega$  与  $P_t$  相反。在这里通过例句对上述算法加以解释。

例句:“在使用过程中,噪声越来越大”。

由于句子中  $P_{\text{噪声}} = -1$ ,  $\text{大} \in PDW$ , 所以  $P_{\text{大}} = P_{\text{噪声}} = -1$ , 从而判定在该句子中“大”表现为贬义。

## 2.3 修饰词词典

修饰词在评论文本的情感极性中也起着关键性作用。例如“不好看”和“非常好看”,如果只考虑“好看”这个情感词的极性,忽略上下文的语境就会导致极性倾向判断错误,或者导致极性强度值失真。本文在文献[14-18]等研究的基础上,结合网络评论特点,经过人工过滤,最终得到修饰词词典,其中包括否定副词 28 个(见表 2 所示)、程度副词 85 个(见表 3 所示)。

表 2 否定副词集合

否定副词
不、非、别、甬、不必、不曾、不要、没、没有、莫、不用、何必、何须、何曾、何尝、空、白、干、徒、徒然、无能、未、未曾、未尝、无需、毋庸、勿、否

表 3 程度副词集合

程度副词	极量(15)	最高为极极为极其极度极端至至为顶过过于过分分外万分
高量(48)		很更更加更其更为越越发备加愈愈加愈发愈益愈为越加格外益发太挺怪老忒非常特别相当十分好好不甚颇甚为颇为满深为蛮够多多么不胜真殊特大大为尤为何等何其尤其无比
	中量(7)	较较为比较不大不太不甚不很
低量(15)		稍稍稍许稍微稍为些微些许略略略略微略为多少有点有点儿有些

同时需要给每一个修饰词赋予一个强度值  $Strength(W)$ 。其中,程度副词的取值范围设定在  $(0.5, 2)$  之间。进一步的,本文规定“低量程度”、“中量程度”、“高量程度”、“极量程度”的强度取值分别为 0.5、1、1.5 及 2。相应的,否定词的取值定在  $(0, 1)$  之间。如果程度副词修饰情感词,短语的倾向值应该等于本身的强度值乘以程度副词的强度值。然而当有否定词修饰时,那么短语极性值与原极性相反,同时强度也会减弱,这是由于否定语句结构会在一定程度上弱化语气。例如“差”的强度值要比“不好”的强度高。

## 2.4 关联强度词典

在倾向性分析中,情感词的极性不应单纯考虑其原极性(情感词典所标注),尤其是关联关系等存在的情况下,更应该结合上下文语境考查其上下文极性。例如“虽然宏基的机子外观不好看,但是性能好”,对宏基的整体评价是偏褒义的。“好看”的原极性为 +1,但“外观”的倾向值实际为  $-1/2$ (这里之所以为  $-1/2$ ,是因为否定结构改变了情感极性,“虽然”这一关联结构进一步弱化了情感强度,如果直接对情感词取反向语义,有失合理;在情感倾向取反的同时,考虑否定词及连词的强烈程度,能更准确的表达发话人的意图,详见算法 1)。同样,对于“性能”而言,“好”的原极性为 +1,但由于转折词“但是”的存在,使得“性能”的倾向性比单纯的“好”更重。

常用关联词的关系包括并列关系、承接关系、递进关系、选择关系、转折关系、假设关系、条件关系及因果关系等 8 种。本文构建的关联强度词典共收录 89 个关联词,其词性强度由人工标注所得。各关联关系的含义、强调成分及部分关联强度词典如表 4 所示。

表 4 关联关系及部分关联强度词典

关联关系	含义及强调成分	关联词	强度(前者/后者)
并列关系	并举分句,不表强调	既...又...	1/1
承接关系	承上启下,强调前者	一...就...	1.25/1
递进关系	表进一层,强调后者	不但...而且...	1/1.5
选择关系	表非此即彼	要么...要么...	1/1
		宁可...也不...	0.75/0.5
转折关系	表转折,强调后者	虽然...但是...	0.75/1.5
假设关系	强调前者	倘若...就...	1.25/1
条件关系	强调前者	只有...才...	1.25/1
因果关系	前因后果,强调前者	既然...那么...	1.25/1

### 3 商品评论的句子级倾向性分析

本文采用哈尔滨工业大学信息检索实验室开发的汉语依存句法分析器 Deparser<sup>[19]</sup>,对评论文本进行句法分析。Deparser 可将句子由一个线性序列转化为一棵结构化的依存分析树,通过依存弧反映句子中词汇之间的依存关系,依存句法标注体系在此不再赘述。

#### 3.1 情感词极性计算

对一个句子进行句法分析后,通过分析词语之间的依存关系,根据语法知识可得,包含情感词的关系对通常为:ADV(状中结构)、DE(的字结构)及 VOB(动宾结构),包含关联结构的关系对为 CNJ(关联结构),从而找出修饰词前缀(否定前缀和程度副词前缀)及关联词,结合组合情感词典,最终计算出词语的上下文极性。本文给出的改进的情感词上下文极性计算算法<sup>[20]</sup>如算法 1。

##### 算法 1 改进的情感词上下文极性算法

输入:词语之间的依存关系对 (*RelationPair*)

输出:情感词上下文极性 (*ContextualPolarity*)

1) 结合静态情感词典,获得情感词 *word* 的原极性,并标记为 *PriorPolarity(word)*。若 *PriorPolarity(word) ≠ 0*,则转到 2)。

2) 记 *word* 的上下文极性为 *ContextualPolarity(word)*,且令  $ContextualPolarity(word) = PriorPolarity(word)$ 。

3) 查找 *RelationPair* 左侧出现情感词 *word* 的依存关系对  $RelationPair_1$ ,对于  $RelationPair_1$  而言:

① 若为 ATT(定中关系),则:

a) 若修饰词为“的”,则追溯与其关联的关系对  $RelationPair_2$ ,其中的程度副词  $DegreeAdverb_2$ ,则:

$$ContextualPolarity(word) * = Strength(DegreeAdverb_2)$$

b) 若修饰词为名词 *noun*,则结合动态情感词典有:

$$ContextualPolarity(word) = DynamicPolarity(word, noun)$$

② 若为 ADV(状中关系),且其中副词为 *adverb*,则继续追溯直到找到与 *adverb* 相关联的修饰 *adverb* 的程度副词或否定词,根据修饰词词典,则有:

a) 若为程度副词  $DegreeAdverb_1$ ,且其强度值为  $Strength(DegreeAdverb_1)$ ,则令

$$ContextualPolarity(word) * = Strength(DegreeAdverb_1)$$

b) 若为否定词 *Privative*,且强度为  $Strength(Privative)$ ,则令

$$ContextualPolarity(word) = - ContextualPolarity(word) * Strength(Privative)$$

③ 若为 CNJ(关联关系),且由关联强度词典知关联词 *Conjunctive* 的强度值为  $Strength(Conjunctive)$ ,则令

$$ContextualPolarity(word) * = Strength(Conjunctive)$$

4) 查找 *RelationPair* 右侧出现 *word* 的依存关系对  $RelationPair_3$ :

① 若为 DE(的字结构),且修饰“的”字的词语为 *adj*,则追溯所有与 *adj* 相关联的关系对  $RelationPair_4$ ,对于  $RelationPair_4$  中出现的程度副词或否定词,则按照 3) 中②的方法处理。

② 若为 VOB(动宾关系),且其中宾语为 *object*,若  $ContextualPolarity(object) ≠ 0$ ,则

$$ContextualPolarity(word) = ContextualPolarity(object)$$

例如:虽然苹果价格高,但是性能真的是无法挑剔。对于情感词“挑剔”的极性值利用算法 1 分析结果如下:

否定关系:挑剔 无法

$$\begin{aligned} ContextualPolarity(\text{挑剔}) &= -PriorPolarity(\text{挑剔}) \times Strength(\text{无法}) \\ &= -(-1 \times 0.75) = +0.75 \end{aligned}$$

程度关系:挑剔 真  $ContextualPolarity(\text{挑剔}) = ContextualPolarity(\text{挑剔}) \times Strength(\text{真}) = +0.75 \times 1.5 = +1.125$  强调关系:但是  $ContextualPolarity(\text{挑剔}) = ContextualPolarity(\text{挑剔}) \times Strength(\text{但是}) = +1.125 \times 15 = 1.6875$  因此“挑剔”的上下文极性为 +1.6875。

#### 3.2 评论的观点抽取

倾向性分析的核心问题是如何识别出评价对象和评价词。人工构建与关联规则挖掘是评价对象获取两类主要方法<sup>[21,22]</sup>。但由于其可移植性、领域相关性等存在局限,目前国内外研究多结合句法语义进行关系抽取。SBV 极性传递算法<sup>[23]</sup>虽识别句法关系效果较好,但其在评价对象抽取过程中准确率并不十分理想,而且无法获取评价对象之间的从属关系。

本文将领域本体融入到 SBV 算法当中,在结合句法规则的基础上,利用本体概念间的从属关系识别评价实体及其属性,提高了 <评价对象,评价词> 二元组抽取的准确率,同时当评论文本中存在属性概念时,评价实体的极性值是通过属性倾向值加权求和来获取,从而有效解决评价对象倾向值计算结果错误的现象。算法描述如算法 2 所示。

##### 算法 2 O-SBV 算法

输入:依存句法分析后的评论文本

输出:评论内容的极性

1) 利用领域本体判断评论中是否含有评价实体 *Entity*。

① 若无评价实体,表示此句无意义,不处理本句,结束;

② 则表示含有 *Entity*,对 *Entity* 使用 SBV 算法进行处理,得到  $ContextualPolarity(Entity)$ ,执行 2)。

2) 利用领域本体查找评论中存在的属性关键词 *Attribute*,

① 若不含属性关键词,执行 4)。

② 否则继续使用 SBV 算法处理属性关键词 *Attribute*,得到  $ContextualPolarity(Attribute)$ ,标记其中已处理的属性关键词,对于关联的情感词也打上标签,执行 3)。

3) 经过对属性关键词 *Attribute* 的 SBV 算法,

① 继续查找评论文本,将没有标记过的属性关键词记为 *UnMarkedAttribute*;

② 若 *UnMarkedAttribute* 存在,查找其 ATT(定中)关系对。对于其中出现的词 *noun*,进一步的,查找所有关联 *noun* 的关系对,直至发现含有情感词 *adj* 的关系对;

③ 如果  $ContextualPolarity(adj) ≠ 0$ ,则令:  $ContextualPolarity(UnMarkedAttribute) = ContextualPolarity(adj)$ 。

4) 记录以上所有使用过的情感词,查找没有标记过的情感词,记为 *UnMarkedPolar*,则:

① 如果 *UnMarkedPolar* 存在,在句法语义层次,向前查找最邻近的属性关键词(或评价实体),同时,将  $ContextualPolarity(UnMarkedPolar)$  作为调整参数,调整属性关键词的极性,即:

$$ContextualPolarity(Attribute) \leftarrow ContextualPolarity(UnMarkedPolar)$$

② 如果 *UnMarkedPolar* 不存在,执行 5)。

5) 利用式(1),对 *Attribute* 与 *UnMarkedAttribute* 极性值加权求和,利用所得结果对  $ContextualPolarity(Entity)$  进行调整。

结合式(1)、式(5)、算法 1 与算法 2,现对 3.1 节例句分析

如下:

$$\begin{aligned} \text{ContextualPolarity}(\text{苹果}) &= \text{ContextualPolarity}(\text{价格}) \times \text{Weight}(\text{价格}) + \\ &\quad \text{ContextualPolarity}(\text{性能}) \times \text{Weight}(\text{性能}) \\ &= \text{ContextualPolarity}(\text{高}) \times \text{Weight}(\text{价格}) + \\ &\quad \text{ContextualPolarity}(\text{挑剔}) \times \text{Weight}(\text{性能}) \\ &= -0.75 \times 0.28 + 1.6875 \times 0.36 = 0.3975 \end{aligned}$$

由计算结果可知,评论者虽然对“价格”稍有不满意,但对“性能”较为看好,整体对评价实体“苹果”的情感倾向为褒义。当一条评论文本中仅包含一个评价实体时,ContextualPolarity(Entity)即句子极性值;当含有超过多个评价实体时,各实体分别持有其ContextualPolarity(Entity),即评论者对各评价实体分别持有不同的倾向性。

## 4 实验

实验环境:CPU为Intel(R)Core™i5-2400@3.10GHz,4GB内存,Windows7操作系统,算法实现工具为VS2010。

本文实验数据使用从中关村在线(<http://www.zol.com.cn/>)选取的500条笔记本电脑相关的主观性文本作为语料,从中随机抽取了共200条评论,人工对这些评论中情感词极性和评价对象进行识别和标注,手工标注示例如表5所示。

表5 手工标注示例

序号	评论	评价对象	情感词	倾向性
1	神舟精盾外观大气,配置高,实用性强	神舟精盾外观 配置 实用性	大气 高 强	神舟精盾 + 外观 + 配置 + 实用性 +
2	S3很轻便,散热也挺好的,就是硬盘有点小	ThinkPad S3 散热 硬盘	轻便 挺好 有点小	ThinkPad S3 + 散热 + 硬盘 -
3	Yoga13当平板用键盘不是很舒服	联想 Yoga13 键盘	不是很 舒服	联想 Yoga13 - 键盘 -
4	...	...	...	...

结合手工标注,首先基于改进的上下文情感算法对200条评论中290个情感词进行分析,实验结果如表6所示。

表6 情感词极性实验结果

		算法识别	人工识别	准确率	召回率
原极性	褒	186	212	96.37%	87.74%
	贬	61	78	92.42%	78.21%
修饰词	程度副词	32	36	94.12%	88.89%
	否定副词	15	17	93.75%	88.24%
	关联词	48	56	92.31%	85.71%
上下文极性	褒	174	205	93.55%	84.88%
	贬	73	85	91.25%	85.88%

实验结果表明,引用修饰与关联词典,当句子含否定词(实验中比例为12.41%)时有效保证语义极性方向的正确性,当存在程度副词、关联词等强调成分(实验中比例为25.17%)时可以较好地还原原极性强度,从而避免了单纯统计学方法的不可靠性。组合情感词典的引入,尤其是关联强度词典与动态情感词典的引入,结合改进的情感词上下文极性算法,使得情感词的上

下文极性计算在准确率和召回率上表现良好。进一步的实验同时证明,由于融入了领域本体,与传统的SBV算法相比,OSBV算法有效降低了算法对SBV结构等依存关系的依赖,更加明确了评价对象间的依存关系,同时能更加准确地识别产品信息的专用名词和网络用语,所以在评价对象识别方面,在准确率和召回率都有了明显的提高,如表7所示。但是本文提出的算法对领域本体的依赖性较大,要求本体中的概念及其关系定义相对完善,否则会降低算法识别的准确率。

表7 评价对象实验对比结果

算法	评价对象(个数)	正确评价对象(个数)	人工识别	准确率	召回率
传统SBV算法	342	258	362	75.44%	81.21%
O-SBV算法	346	281		71.27%	77.62%

## 5 结语

本文在构建组合情感词典的过程中,通过引入网络情感词典,较大限度地减少了网络语言灵活性及多变性对倾向性分析的影响;利用动态情感词典,较充分地考虑并量化了评价对象对情感极性的影响;结合改进的情感词上下文极性算法,提高了情感词极性计算的准确率,显著降低了统计方法在语义极性计算方面的不确定性。

构建领域本体,不但保证了对非频繁评价对象的抽取,而且便于获取词语间更多地语义信息;进一步的,将领域本体引入SBV算法,在保证句法层面的极性传递正确性的基础上,更多地关注极性计算在语义结构上的完整性;提高了<评价对象,评价词>二元组抽取的准确率,有效改进了句子级的倾向性分析方法。

由于领域知识的广泛性与网络术语的灵活性,本体及组合情感词典的构建仍需要不断完善。依存关系分析过程中,仅考虑了相对规范的句法结构,因此本文中的算法还有待于进一步改进。

最后,感谢哈尔滨工业大学信息检索研究室提供了研究所需的汉语句法分析器Deparser。

## 参 考 文 献

- [1] 厉小军,戴霖,施寒潇,等. 文本倾向性分析综述[J]. 浙江大学学报:工学版,2011,45(7):1167-1174.
- [2] 邹嘉彦. 评述新闻报道或文章色彩-正负两极性自动分类的研究[C]//自然语言理解与大规模内容计算-全国第八届计算语言学联合学术会议. 清华大学出版社,2005:21-23.
- [3] Yi J, Nasukawa T, Bunescu R, et al. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques[C]//Proceedings of the 3rd IEEE International Conference on Data Mining. Piscataway, NJ, USA: IEEE, 2003: 427-434.
- [4] 朱嫣岚,闵锦,周雅倩,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.
- [5] Wang C, Lu J, Zhang G. A semantic classification approach for online product reviews[C]//Proc of the 2005 IEEE / WIC / ACM International Conference on Web Intelligence. 2005: 276-279.
- [6] 徐琳宏,林鸿飞,杨志豪,等. 基于语义理解的文本倾向性识别机制[J]. 中文信息报,2007,21(1):96-100.

续表 1

序号	计算坐标		实际坐标		残差
	X	Y	$x^{real}$	$y^{real}$	
10	13.8	7.7	13.7	7.8	0.14
11	11.2	12.6	11	12.8	0.28
12	0.4	-0.06	0.1	0	0.31
13	3.7	-0.06	3.5	0	0.21
14	-0.1	4.4	-0.1	4.2	0.2
15	2.15	7.6	2	7.8	0.25
16	-15.4	-7.4	-15	7.8	0.41
17	3.3	7.75	3.5	7.9	0.25
18	5.81	2.3	5.9	2.3	0.09
19	-0.7	1.9	-0.8	2	0.14
20	2.2	-1.7	2	-1.8	0.22

表 2 实验耗时数据表

序号	耗时/ms	序号	耗时/ms
1	1	11	1
2	0	12	1
3	0	13	0
4	0	14	0
5	1	15	0
6	1	16	0
7	0	17	1
8	0	18	1
9	0	19	0
10	0	20	0

## 4 结 语

在视觉定位系统中,为了获得较大的视场范围,有时不可避免地需要使用广角摄像机来提取周围环境中的图像。由于广角摄像机畸变的干扰,使广角摄像机在视觉定位系统中无法精确定位的问题日益凸显。本文提出的中心迭代法与区域插值法在像素坐标上消除了摄像机的畸变,并使用小孔模型得到目标物体的摄像机实际坐标。通过使用中心点迭代法,可以保证在5次迭代内找到无畸变的像素坐标范围,充分保证了系统的实时性。同时使用区域插值法使得定位更为准确。实验结果也证明了定位的准确性以及实时性。

该成果在沈阳新松机器人股份有限公司生产的工业移动机器人上有了实际的工业应用。长时间的实验测试证明通过广角摄像机实现移动机器人视觉定位在准确性和实时性都能达到工业应用的要求,可以克服现有导航方式制作成本高、现场施工复杂等缺点。

## 参 考 文 献

- [1] Zhengyou Zhang. A Flexible New Technique for Camera Calibration [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(11): 1330-1334.
- [2] Duane C Brown. Close-Range Camera Calibration [J]. Photogrammetric Engineering, 1971, 37(8): 855-866.
- [3] Tsai R Y. A versatile camera calibration technique for high accuracy 3D

machine vision metrology using off-the-shelf TV camera and lenses [J]. IEEE Journal of Robotics and Automation, 1987, 3(4): 323-344.

- [4] 傅丹,周剑,邱志强,等. 基于直线的几何不变性标定摄像机参数 [J]. 中国图像图形学报, 2009, 14(6): 1058-1063.
- [5] 王志良,张琼,迟健男,等. 基于直线射影特征的摄像机参数标定方法 [J]. 计算机科学, 2011, 38(8): 272-274.
- [6] C Brauer-Burchardt. A simple new method for precise lens distortion correction of low cost camera systems [C]//DAGM, 2004, LNCS 3175: 570-577.
- [7] Gary Bradski, Adrian Kaehler. Learning OpenCV [M]. O'Reilly Media, 2009.
- [8] Wang Hongzhi, Li Meijing, Zhang Liwei. The Distortion Correction of Large View Wide-angle Lens for Image Mosaic Based on OpenCV [C]//International Conference on Mechatronic Science, Electric Engineering and Computer. China, 2011: 1074-1077.

## (上接第 108 页)

- [7] Uschold M, King M. Towards a methodology for building ontology [C]//Workshop on basic ontological issues in knowledge sharing: International Joint Conference on Artificial Intelligence, 1995: 373-380.
- [8] Chandra S Amaravadi. The Dimensions of Process Knowledge [J]. Knowledge and Process Management, 2005, 12(1): 16-19.
- [9] Subramanian Muthu, Larry Whitman, et al. Business Process Reengineering: A Consolidated Methodology [C]//Proceeding of the 4th Annual International Conference on Industrial Engineering Theory. November 1999: 17-20.
- [10] Ontology Development 101: A Guide to Creating Your First Ontology [OL]. [http://protege.stanford.edu/publications/ontology\\_development/ontology101.pdf](http://protege.stanford.edu/publications/ontology_development/ontology101.pdf).
- [11] 赵妍妍,秦兵,刘挺. 文本情感分析 [J]. 软件学报, 2010, 21(8): 1834-1848.
- [12] Turney P, Littman M L. Measuring praise and criticism: Inference of semantic orientation from association [J]. ACM Trans. On Information Systems, 2003, 21(4): 315-346.
- [13] 陈晓东. 基于情感词典的中文微博情感倾向分析研究 [D]. 武汉: 华中科技大学, 2012.
- [14] 姚天防, 裴青阳, 李建超, 等. 一个用于汉语汽车评论的意见挖掘系统 [C]//中文信息处理前沿进展—中国中文信息学会二十五周年学术会议论文集. 北京: 清华大学出版社, 2006: 260-281.
- [15] 蔺磺, 郭慧慧. 程度副词的特点范围与分类 [J]. 山西大学学报, 2003, 26(2): 71-74.
- [16] 郝雷红. 现代汉语否定副词研究 [D]. 北京: 首都师范大学, 2003.
- [17] 王兰成, 徐震. 基于情感本体的主题网络舆情倾向性分析 [J]. 信息与控制, 2013, 42(1): 46-52.
- [18] 杨超, 冯时, 王大玲, 等. 基于情感词典扩展技术的网络舆情倾向性分析 [J]. 小型微型计算机系统, 2010, 4(4): 691-695.
- [19] [http://ir.hit.edu.cn/phpwebsite/index.php?module=pagemaster&PAGE\\_user\\_op=view\\_page&PAGE\\_id=147&MMN\\_position=52:48](http://ir.hit.edu.cn/phpwebsite/index.php?module=pagemaster&PAGE_user_op=view_page&PAGE_id=147&MMN_position=52:48)
- [20] 姜德成, 姚天防. 汉语句子语义极性分析和观点抽取方法的研究 [J]. 计算机应用, 2006, 26(11): 2622-2625.
- [21] 刘鸿宇, 赵妍妍, 秦兵, 等. 评价对象抽取及其倾向性分析 [J]. 中文信息学报, 2010, 24(1): 84-88.
- [22] 宋晓雷, 王素格, 李红霞. 面向特定领域的产品评价对象自动识别研究 [J]. 中文信息学报, 2010, 24(1): 89-93.
- [23] 姚天防, 姜德成. 汉语语句主题语义倾向性分析方法的研究 [J]. 中文信息学报, 2007, 21(5): 73-79.