

文章编号: 1003-0077 (2011) 00-0000-00

## 领域相关的汉语情感词典扩展\*

宋佳颖, 贺宇, 付国宏

(黑龙江大学 计算机科学技术学院, 黑龙江 哈尔滨 150080)

**摘要:** 动态情感知识的获取, 特别是领域相关极性词典的构建一直是意见挖掘和情感分析系统在开放应用时面临的主要挑战之一。本文面向产品评价文本提出一种汉语情感极性词典扩展方法。该方法首先采用序列标注方法从意见文本中抽取产品意见要素, 同时构建属性-评价对; 然后, 对抽取的属性-评价对进行正规化, 以减少词典扩展中的复杂性和噪声; 最后, 改进 PolarityRank 算法的构图方式以使其适用于汉语文本, 从而完成词典扩展。在汽车和手机两个领域的意见文本的实验结果表明领域相关的情感极性词语的扩展有利于情感极性分类性能的提高。

**关键词:** 情感分析; 情感词典扩展; PolarityRank; 意见要素一致化

中图分类号: TP391

文献标识码: A

## Automatic Expansion of Domain-Specific Sentiment Lexicon for Chinese

Jiaying Song, Yu He, Guohong Fu

(School of Computer Science and Technology, Heilongjiang University, Harbin, Heilongjiang 150080, China)

**Abstract:** Dynamic sentiment knowledge acquisition, especially domain-specific sentiment lexicon construction is still a challenge for developing practical opinion mining and sentiment analysis systems for open applications. In this paper we incorporate opinion element normalization with the PolarityRank algorithm and thus propose a semi-supervised approach for building a domain-specific opinion lexicon for Chinese sentiment polarity classification. We first extract a set of attribution-evaluation pairs from opinionated texts. In order to reduce the complexity and noises in sentiment lexicon expansion, we further normalize the extracted product attributions and their relevant evaluations using Jaccard coefficient and rules, respectively. Finally, we modify the PolarityRank algorithm to expand the original sentiment lexicon. Experimental results over two sets of reviews on automobile and mobile phone products show that using the expanded domain-specific dynamic polar words helps improve sentiment classification performance.

**Key words:** Sentiment analysis; opinion lexicon expansion; PolarityRank; opinion element normalization

### 1 引言

随着 Web2.0 的兴起和社会媒体的迅速发展, 情感分析(亦称意见挖掘)已成为自然语言处理研究的一个热点, 并在近年得到快速的发展, 出现不同的情感分析系统。但这些系统在大规模开放应用时依然面临情感词典覆盖度<sup>[1-2]</sup>、动态极性识别<sup>[3-5]</sup>和领域适应性<sup>[6-10]</sup>等挑战。一方面, 词语极性是句子情感分类的基础。因此, 大多数句子情感分类系统一般都带有一个情感词典以帮助确定词语极性。但事先构造的情感词典很难覆盖开放意见文本中的所有情感词语。另一方面, 许多情感词语或短语在不同的上下文或领域中呈现不同的极性。例如, “油耗高”和“屏幕分辨率高”分属汽车和手机的评价, 其中同一个情感词“高”却分别表达负

\* 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:** 宋佳颖(1990—), 女, 研究生, 主要研究方向: 自然语言处理; 贺宇(1988—), 男, 研究生, 主要研究方向: 自然语言处理, 意见摘要; 付国宏(1968—), 男, 教授, 主要研究方向: 自然语言处理、文本挖掘。

向和正向情感极性；在“噪音大”和“驾驶空间大”两个例子中，它们虽然同属汽车领域且使用同一个情感词“大”，因其修饰不同的汽车属性而表达了不同的情感倾向性。因此，情感词典的自动构造和扩展近年来引起人们的广泛关注<sup>[1-10]</sup>。

本文在前人工作基础上，融合意见要素正规化信息，提出一种面向产品意见文本的基于PolarityRank算法的领域相关的汉语情感词典扩展方法。主要步骤如下：首先将意见要素抽取看作序列标注问题，采用条件随机场（CRFs）得到各个评论句中的意见属性和评价词。而后综合考虑共现词频和词间距离两个特征，将抽取的属性和评价词进一步匹配成属性-评价对。为了减少词典扩展的复杂性和噪声，我们还分别采用Jaccard系数以及修饰词、否定词缩减规则对抽取的属性及其评价进行了正规化；最后，改进基于连接图的PolarityRank算法扩展极性词典，使其适用于中文产品评论，从而提升词典扩展效果。

本文的组织结构如下：第2节是相关研究的总结；第3节详细介绍领域相关的情感词典扩展方法；第4节给出本文的实验结果及分析；最后一节是本文的结论和展望。

## 2 相关研究

意见要素抽取往往作为情感词典构造和扩展的预处理，其主要任务是从给定的意见文本中抽取组成意见的要素，包括评价对象及其属性、评价词等。典型的意见要素抽取方法主要有规则方法<sup>[11]</sup>、基于依存知识的方法<sup>[12]</sup>和机器学习方法<sup>[3]</sup>。Hu和Liu<sup>[10]</sup>首先利用关联规则挖掘的方法从产品评论中抽取高频的名词及短语作为属性，扩展已有属性周边名词作为低频属性，这种方法由于规则限定，结果的召回率不佳。Zhang等<sup>[11]</sup>以属性词典和评价词词典为基础，利用依存关系匹配抽取属性-评价对。Jakob和Gurevych<sup>[12]</sup>在CRFs框架下探讨了领域内和跨领域产品属性抽取。此外，Wang等<sup>[13]</sup>以评价词词典为基础，采用自举迭代策略，依据上下文关联度获取属性及其评价词。

情感词典自动扩展研究目前主要围绕词典覆盖度和动态情感极性获取两个问题展开，并形成各具特色的扩展方法。Kanayama和Nasukawa<sup>[1]</sup>将句内同现信息和跨句子的同现信息结合起来学习词和短语的极性，从而构造领域相关的词典。这种方法强烈依赖于已知的种子情感词，对于种子词典不包含的未知情感词，缺少抽取能力。Qiu等(2011)<sup>[2]</sup>采用双重繁殖算法利用属性-情感词关系来扩展情感词典。为了获取上下文相关的动态极性知识，Esuli和Sebastiani<sup>[3]</sup>利用词典中词语的解释来确定词语的关联关系，并以此构建词语网络，进而通过连线权重来确定词语的极性。Wilson等(2009)<sup>[4]</sup>深入分析了影响词语动态极性的各种因素，并尝试在短语层面融合各种特征解决词语动态极性问题。此外，Wu和Wen(2010)等<sup>[5]</sup>研究了汉语形容词的极性消歧问题。虽然这些方法主要关注动态极性词所在的局部上下文，忽视了相关的领域，因而不能识别领域相关的情感极性词。领域相关的动态极性词识别对意见挖掘系统的领域移植至关重要。为此，Andreevskaia和Bergler(2008)<sup>[6]</sup>、Tan等(2007)<sup>[7]</sup>和吕韶华等人(2012)<sup>[8]</sup>分别从不同角度探讨了情感分类系统跨领域移植问题。为了领域动态极性词的极性判定问题，Cruz等人(2011)<sup>[14]</sup>以属性-评价对为单位，提出基于PageRank<sup>[15]</sup>的随机游走排列算法PolarityRank，并用连词结构作为桥梁从未加工的意见语料对种子词典进行扩展。本文以PolarityRank算法为基础，面向产品评价文本展开汉语领域动态极性词典扩展研究。与Cruz等人<sup>[10]</sup>的研究不同，本文改进PolarityRank算法连接图构图方式以适应汉语的特点，同时融合产品属性及其评价的正规化信息，以提高词典扩展质量。

## 3 情感词典扩展方法

### 3.1 情感词典扩展任务

在本文研究中，情感词典扩展的任务是在种子词典基础上从特定领域的产品评价文本中

自动抽取未知情感词语及其属性，并根据改进的 **PolarityRank** 算法确定相应的情感极性，从而完成相应词条的构造。如图 1 所示，本文的情感词典扩展系统主要包括以下四个模块：

(1) **预处理模块**的主要任务是对给定的产品评价文本进行分词和词性标注，为后续的意见要素识别和抽取作准备。为了提高词法分析的可靠性，本文采用基于语素的分词和词性标注一体化的系统<sup>[15]</sup>完成预处理任务。

(2) **意见要素识别模块**的主要任务是在词法分析基础上识别产品属性及其评价等产品意见的主要要素。本文将这一问题视作序列标注问题，并采用条件随机域方法完成这一任务。

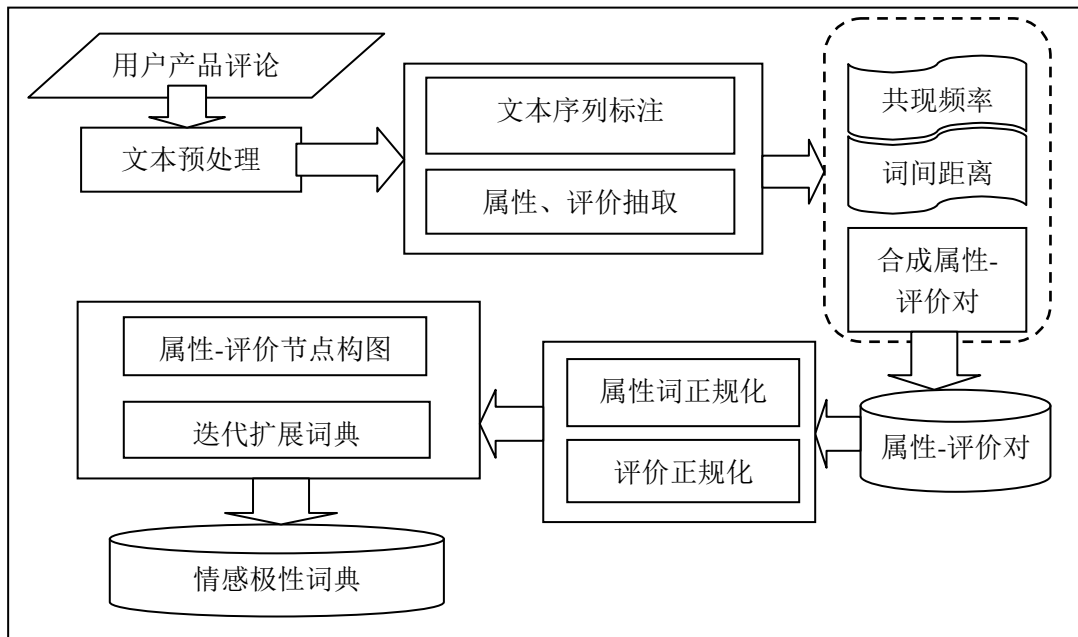


图 1 词典扩展流程图

(3) **属性-评价对匹配/合成模块**的任务是从抽取的属性和评价中根据其在意见句中的共现关系判断他们是否形成修饰关系。存在修饰关系的属性-评价对实际上是情感词典潜在的一个扩展词条。

(4) **属性-评价对正规化模块**的任务是对属性-评价对中存在的大量互为复述关系的属性词及其评价分别进行正规化处理，为同一词意的词确定一致化标准，从而获取更多的属性-评价对情感极性，降低词典扩展的难度。

(5) **基于 PolarityRank 的情感词典扩展模块**的任务是将步骤(4)正规化后的属性-评价对作为候选词条，构造相应的无向连接图，并通过迭代确定相应的情感极性。

为了描述情感词语的动态极性，本文情感词典的每个词条结构为<eval, attr, polar, domain>。其中，eval 代表评价词，attr 代表产品属性，polar 代表情感极性，domain 代表所在领域标记。情感极性  $polar \in \{-1, 1\}$ ，-1 表示负面(negative)极性，1 表示正面极性。产品属性 attr 可以是相应产品的某一部分、组件或性能指标，评价词就是针对属性的意见词或短语。例如：配置-高，其中配置是手机的属性，高是针对属性的评价词，对应正向情感极性。由于动态情感词的存在，例如：油耗-高，此处的高是负向情感。因此词条不能是简单的评价词，而需要有属性来限定上下文，从而明确表达情感极性。

3.2-3.6 节将分别介绍相应模块的基本原理。

### 3. 2 意见要素识别

在词典扩展中，意见要素识别的主要任务是识别给定意见句子中的属性及相应的评价。本文把意见要素识别看作是一个序列标注问题，鉴于条件随机场模型在解决序列标注问题时的良好表现，我们以词作为序列标注的基本单位，应用条件随机场（conditional random fields, CRFs）框架完成属性、评价词的识别。

如图 2 所示，每个意见要素在标注中获得一个形如 x-y 的标记。其中，x 代表意见要素类别标记，y 代表相应词语在所在意见要素的位置标记。考虑到词典扩展的实际，本文定义三种意见要素类别标记，即 {A, E, O}，分别用来表示产品属性、评价和非意见要素词语。至于位置标记，本文采用常见的四标记体系，即 {S, B, M, E}，分别表示单个词语构成的意见要素和意见要素首词、中间词和尾词。此外，在四标记 SBME 基础上，本文还引入一个标记 I 表示意见要素首个中间词，形成五标记体系 SBIME，以提高意见要素标注性能。

(a) 原始输入：iphone 手机的屏幕的分辨率一般。
(b) 词性标注：iphone/nx 手机/n 的/u 屏幕/n 的/u 分辨率/n 一般/a 。/w
(c) 意见要素标注(4 标记)：iphone/O 手机/O 的/S 屏幕/B-A 的/M-A 分辨率/E-A 一般/S-E 。/O
(d) 意见要素标注(5 标记)：iphone/O 手机/O 的/S 屏幕/B-A 的/I-A 分辨率/E-A 一般/S-E 。/O

图 2 意见要素标注样例

考虑到属性及其评价的长度，在意见要素序列标注中我们选取当前词语前后 5 个词语窗口内词形和词性以及相应的一元、二元和三元上下文特征来构造特征模板。

### 3. 3 属性-评价对合成

给定一个意见句，它可能蕴含多条意见信息，即一个意见句子可能有多个属性-评价对。属性-评价对合成的任务是从意见句子潜在的多个属性和评价组合中抽取合适的属性-评价对。如公式(1)所示，为了简化问题，本文在意见要素标注基础上，考虑属性和评价的共现频度及其距离来确定属性 attr 和评价 eval 匹配的可能性。

$$P(attr,eval) = Count(attr,eval)/Distance(attr,eval) \quad (1)$$

其中， $Count(attr,eval)$ 表示属性 attr 和评价词 eval 在所有产品评论中的共现频率， $Distance(attr,eval)$ 表示在对应句中的属性和评价词之间的字数，实际应用时为防止分母为 0，对 Distance 进行了加 0.05 的调整，对于中间含有标点符号的情况做了加 4 的调整。 $P(attr,eval)$ 实际上反映了当前的属性和评价词匹配成对的可能性。对于同一属性的每个候选属性-评价对，我们选取其中 P 值最大的作为正确的属性-评价匹配对。

### 3. 4 属性-评价对正规化

在分词和意见要素标注时，评价词语边界确定存在不一致的现象，加之用户生成的意见文本行文比较自由，评价复述非常丰富。因此，有必要对属性和评价进行正规化，以进一步提高词典扩展的质量。为了简化问题，本文采用以下方法进行意见要素正规化：

- (1) **属性正规化**：由于存在多个属性词表示相同意见对象的情况，比如：“屏幕分辨率”和“分辨率”，“油耗”、“耗油”和“耗油量”等，如果这些属性词搭配了相同的评价词，则可以看做是同意的属性评价对，我们可将其进行一致化处理，从而降低词典扩展的难度，减少未知极性词条的数量。因此我们用公式(2)所示的 Jaccard 系数计算属性词间的相似度，以获取产品属性的同意属性词簇，为属性词生成互为复述的属性词集合，选择集合中在评价文本中出现频率最高的属性词作为标准，对其他属性词进行一致化处理。

$$J_{xy} = \frac{\sum_{k=1}^L (m_{xk} m_{yk})}{\sum_{k=1}^L m_{xk}^2 + \sum_{k=1}^L m_{yk}^2 - \sum_{k=1}^L m_{xk} m_{yk}} \quad (2)$$

其中,  $m_{xk}$  表示字符  $k$  在短语  $x$  中出现的总次数,  $m_{yk}$  则表示字符  $k$  在短语  $y$  中出现的总次数。通过多组实验为相似度判定确定阈值。

(2) **评价正规化**: 针对抽取得到评价词切分不规范及含有冗余信息现象, 我们采用下面规则处理评价信息。

**Rule 1**: 由于抽取得到的一些极性词前面还含有修饰词, 这为后续的词典扩展带来不利的影响, 例如: 非常硬 $\rightarrow$ 硬, 有点高 $\rightarrow$ 高, 这两个词对的左右两边实际表示了相同的评价意义。在词典扩展阶段, 由于抽取得到的很多评价短语含有冗余的修饰部分, 会造成连接图中大量的同类节点无法合并, 这将直接影响情感预测结果。因此我们在扩展前先对所有极性词进行修饰词的过滤, 以达到正规化的效果。本文过滤的修饰词包含收集所得的程度副词、肯定副词和少量语气副词共 76 个。实验阶段构建连接图时, 有效减少了边数, 提高极性预测精度。

**Rule 2**: 由于抽取得到的极性词有些还被否定词所修饰, 这种情况也会对极性判定和词典扩展产生干扰, 我们又用 8 个否定前缀依次对极性词进行过滤, 对去除否定修饰的极性词再次查找已知词典, 从而获得更多的已知极性的属性-评价对, 减小词典扩展的任务难度。

### 3.5 种子词典

我们根据训练部分的极性信息, 得到初始的种子情感词典, 即其中的属性-评价对词条都含有对应的情感极性, 本节任务是完成对测试语料中的词条的极性预测, 从而实现情感词典的扩展。从训练语料中抽取得到领域内动态情感词, 汽车领域 22 个, 手机领域 20 个, 在词典扩展时要对两类极性词分别考虑。

### 3.6 PolarityRank

Cruz 等人<sup>[14]</sup>所提出的 PolarityRank 算法主要用连词结构作为桥梁来进行情感词典扩展。然而, 汉语产品评价文本中连词结构并不多见, 通常有关关系的特征就会出现在同一个短句中。为此, 我们对 PolarityRank 算法进行了改进, 通过词语共现关系及词频构建无向连接图, 以适合汉语特性。情感词典扩展步骤具体如下:

- 根据训练语料和测试语料中的属性-评价对在评论语句中的共现关系, 构建无向图, 出现在同一句评价中的两个属性-评价对可以连接成一条边, 属性-评价对为节点。构建初始图时, 静态评价词和动态评价词没有区别, 在简化图阶段, 节点合并时两者的判定会有差别。静态评价词节点只要评价词相同就算作相同节点, 动态评价词节点需要对应评价和属性都完全相同才可以看做是相同节点。共现频率作为边的权重, 节点合并后边的权重也会有相应增加。
- 未知极性节点初始极性为 0, 根据对每个点计算得到的  $PR$  值为节点更新极性, 从而完成对无向图的节点极性更新。
- 迭代计算每个点的  $PR$  值直到没有新的节点极性值产生, 完成词典扩展。这里, 节点  $v_i$  的  $PR^+(v_i)$  和  $PR^-(v_i)$  值可分别采用公式(3)和(4)计算。

$$PR^+(v_i) = (1-d)e_i^+ + d \left( \sum_{j \in E(v_i)} \frac{p_{ji}}{\sum_{k \in E(v_j)} |p_{jk}|} PR^+(v_j) + \sum_{j \in E(v_i)} \frac{-p_{ji}}{\sum_{k \in E(v_j)} |p_{jk}|} PR^-(v_j) \right) \quad (3)$$

$$PR^-(v_i) = (1-d)e_i^- + d \left( \sum_{j \in E(v_i)} \frac{p_{ji}}{\sum_{k \in E(v_j)} |p_{jk}|} PR^-(v_j) + \sum_{j \in E(v_i)} \frac{-p_{ji}}{\sum_{k \in E(v_j)} |p_{jk}|} PR^+(v_j) \right) \quad (4)$$

式中,  $d$  是一个自定参数, 本文采用 PageRank<sup>[16]</sup> 推荐的 0.85。  $e_i^+$  代表与当前节点  $v_i$  关联的极性为正的节点个数,  $e_i^-$  代表与  $v_i$  关联的极性为负的节点个数,  $E(v_i)$  表示与  $v_i$  相连的边集,  $p_{ji}$  表示节点  $v_i$  和节点  $v_j$  形成的边的权重。

$PR^+(v_i)$  和  $PR^-(v_i)$  值可分别作为  $v_i$  所代表的评价词为正向极性词、负向极性词的可能性概率值。由公式(2)和(3)可以看出, 每个节点的极性计算都会考察关联的所有节点的极性, 这也是共现策略的重要应用。节点  $v_i$  的极性  $Polarity(v_i)$  可由以下三条规则确定:

- (1) 若  $PR^+ > PR^-$ ,  $Polarity(v_i) = 1$ ;
- (2) 若  $PR^+ < PR^-$ ,  $Polarity(v_i) = -1$ ;
- (3) 若  $PR^+ = PR^-$ ,  $Polarity(v_i) = 0$ 。

```

Build a graph G( V , E );
for each unknown  $v_i$  in G
    for each  $v_j$  associated with  $v_i$ 
        if  $Polarity(v_j) > 0$  then  $e_i^{+++}$ ;
        else if  $Polarity(v_j) < 0$  then  $e_i^{--}$ ;
        end if
    Compute  $PR^+(v_i)$  and  $PR^-(v_i)$ ;
    if  $PR^+(v_i) > PR^-(v_i)$  then  $Polarity(v_i) := 1$ ;
    else if  $PR^+(v_i) < PR^-(v_i)$  then  $Polarity(v_i) := -1$ ;
    else  $Polarity(v_i) := 0$ ;
    end if
end for
end for
Update the current graph.

```

图 3 词典扩展算法

图 3 给出词典扩展算法的伪码描述。该算法的基本思想为: 每次迭代将新得到极性的属性评价对作为已知词条, 继续下一轮计算, 直到获得极性的属性-评价对不再增加为止。

## 4 实验结果与分析

为了验证上述方法的有效性, 我们构建一个领域相关的情感词典扩展系统, 并分别应用于汽车和手机两种产品评论的情感极性分析。本节将给出相应的实验结果及其分析。

### 4.1 实验数据

如表 1 所示, 本文实验所用语料来自汽车、手机两个领域的网络用户评价并进行了意见要素的标注, 标注内容包括意见对象、产品属性、评价和属性-评价对对应的情感极性。意见要素标注形式如下图:

```

<sentence1>我的三星手机屏幕分辨率一般。</sentence1>
<target>三星手机</target>
<polarity1>NEG</polarity1>
<attr1>屏幕分辨率</attr1>
<opin1>一般</opin1>

```

图 4 语料标注样例

下面表 1 给出实验语料的整体统计数据。

表 1 语料统计信息

	汽车		手机	
	训练	测试	训练	测试
句子	2014	1007	2044	1023
词语	81197	42723	19063	9341
属性词	2605	1354	2884	1466

表 2 给出了测试语料中未知极性词语的统计信息。

表 2 测试语料未知极性词语

分类	动态极性词	静态极性词
汽车	37	210
手机	33	271

#### 4. 3 属性评价词抽取结果

为了验证意见要素标注性能对词典扩展质量可能产生的影响,测试了不同标记下不同领域的意见要素标注效果,结果如表 3 所示。本实验的测试指标为精确率(Precision, P)、召回率(Recall, R)和 F-测度(F)。

从表 3 的结果可以看出:五标记系统在两个领域的抽取结果都明显优于相应的四标记系统。因此,后续的实验均采用五标记系统抽取所得的属性和评价词。对汽车和手机两个领域测试语料中所识别的属性和评价进一步采用 3.3 节的匹配原则,分别得到 687 和 1135 个属性-评价对。

表 3 意见要素标注结果

	汽车				手机			
	属性		评价		属性		评价	
	四标记	五标记	四标记	五标记	四标记	五标记	四标记	五标记
P	76.72%	80.86%	69.87%	73.44%	87.04%	88.78%	80.03%	83.03%
R	48.01%	50.99%	36.86%	42.71%	77.69%	78.33%	69.73%	72.28%
F	59.06%	<b>62.54%</b>	48.26%	<b>54.01%</b>	82.10%	<b>83.23%</b>	74.53%	<b>77.28%</b>

#### 4. 4 情感词典扩展结果

为了考察不同因素对情感词典扩展的影响,我们在实际情感词典扩展测试中设计了以下五组实验:

- 实验 1: 应用 3.6 节的词典扩展算法直接预测测试语料中未知极性词的极性;
- 实验 2: 针对人工标注的测试语料中的各个属性-评价对 (金标数据), 应用 3.6 节的词典扩展算法判定属性-评价对极性;
- 实验 3: 先用 3.4 节的属性正规化方法对所有属性进行标准化处理, 再用 3.6 节词典扩展算法预测测试语料中未知极性词的极性;
- 实验 4: 先用 3.4 节的属性正规化方法对所有属性进行标准化处理以及评价正规化规则 1 过滤修饰词, 再应用 3.6 节词典扩展算法预测测试语料中未知词极性;
- 实验 5: 先通过 3.4 节的属性正规化方法标准化属性词, 再用评价正规化规则 1 为极性词过滤修饰词, 用规则 2 过滤否定词, 最后应用 3.6 节算法扩展词典。

词典扩展的构图阶段在实验 1 时共得到汽车领域属性-评价对节点 3292 个, 边数为 2417 条。手机领域节点 4019 个, 边数为 1414 条, 而经过正规化后实验 5 的汽车领域边数减少为 2213 条, 手机领域的边数减少为 1279 条, 有效合并了相同意见节点。属性一致化时本着使尽可能多的有共指关系的属性能被识别的原则, 确定 3.4 节的相似度计算阈值为 0.5。评测部分采用常用指标准确率 (P)、召回率 (R) 和 F-测度 (F), 为了考察对两类极性词词典的扩展效果, 首先对动态极性和静态极性词进行了分别评测, 然后对合并的极性词典完成评测。实验结果如表 4 和表 5 所示。

表 4 汽车领域词典扩展结果

	动态极性词			静态极性词			全部		
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
实验 1	33.33%	21.62%	26.23%	69.01%	23.33%	34.87%	60.00%	23.08%	33.34%
实验 2	<b>77.27%</b>	<b>51.52%</b>	<b>61.82%</b>	71.17%	32.66%	44.77%	71.62%	33.65%	45.79%
实验 3	34.62%	24.32%	28.57%	69.44%	23.81%	35.46%	60.20%	23.89%	34.21%
实验 4	44.44%	36.36%	40.00%	89.08%	77.94%	83.14%	80.82%	69.82%	74.92%
实验 5	46.43%	41.94%	44.07%	<b>91.94%</b>	<b>89.76%</b>	<b>90.84%</b>	<b>83.55%</b>	<b>80.38%</b>	<b>81.93%</b>

从表 4 和表 5 的实验结果可以看出: (1) 对比实验 1、3、4 和 5 的结果, 经过正规化处理后, 两个领域的极性词典扩展性能均得到较大提高; (2) 同等条件下, 实验 2 的扩展性能优于相应的实验 1 的结果, 这说明意见要素识别以及属性-评价对匹配性能直接影响到词典扩展性能。此外, 同等条件下, 手机领域的词典扩展结果优于汽车领域, 其原因可能是汽车评价文本质量本身较差, 相应的意见要素识别性能较低(如表 3 所示), 从而导致最后的词典扩展性能不佳。

表 5 手机领域词典扩展结果

	动态极性词			静态极性词			全部		
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
实验 1	77.78%	42.42%	54.90%	86.84%	24.35%	38.03%	85.11%	26.58%	40.51%
实验 2	<b>88.24%</b>	57.69%	69.77%	87.50%	33.25%	48.19%	87.57%	34.68%	49.68%
实验 3	78.95%	50.00%	61.23%	86.84%	24.35%	38.03%	85.26%	26.91%	40.91%
实验 4	83.33%	86.21%	<b>84.75%</b>	95.04%	74.03%	83.23%	92.98%	75.71%	83.46%
实验 5	75.86%	<b>88.00%</b>	81.48%	<b>96.15%</b>	<b>90.36%</b>	<b>93.17%</b>	<b>95.68%</b>	<b>92.67%</b>	<b>94.15%</b>



#### 4. 5 扩展的情感词典对情感极性分类的影响

为了进一步说明领域词典扩展对情感分析的影响，我们从上述两个领域的语料中随机抽取 1000 句，并采用以下三组不同的词典进行句子级情感极性分类测试。

- 词典一：知网(HowNet)+ 台湾大学 (NTUSD) + 清华大学 (汉语情感词极值表)；
- 词典二：词典一 + 本文构建的静态情感词典(4.4 节实验 5)；
- 词典三：词典二 + 本文构建的动态情感词典(4.4 节实验 5)。

注意：本实验的情感极性分类采用简单的基于词袋模型的分方法。实验结果如表 6 所示。

表 6 采用不同词典时的情感极性分类结果

词典	汽车			手机		
	<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
词典一	<b>47.97%</b>	42.30%	44.96%	<b>67.13%</b>	47.86%	55.88%
词典二	47.43%	44.21%	45.76%	63.26%	52.83%	57.58%
词典三	47.79%	<b>46.18%</b>	<b>46.97%</b>	62.34%	<b>56.76%</b>	<b>59.42%</b>

表 6 的实验结果显示：加入扩展的领域静态情感词语后，虽然极性分类的精确度稍有下降，但在召回率和 F-值方面都有显著的提高，这从侧面说明增加情感词典的覆盖度有助于提高情感极性分类的召回率和整体性能。相比于加入扩展的领域静态词语，在加入扩展的动态情感词语后，情感极性分类的召回率和 F-值提升幅度更为明显，这进一步验证了领域情感词典对于领域相关的情感分析和意见挖掘的重要意义。

#### 5 结论与展望

针对汉语产品评论的特点，本文融合意见要素正规化信息，提出一种面向汉语产品评价的基于 PolarityRank 的领域相关的情感词典扩展方法。在汽车和手机两个领域的产品评价语料上的实验结果表明意见要素标注以及正规化直接影响到情感词典的质量。相应的句子情感极性分类实验表明引入扩展的领域相关的情感词语可以有效提高情感分类性能。

虽然本文实验取得预期结果，证明了领域相关的情感词典对于汉语情感分析的意义。但由于本文研究的重点限制，所采取的意见要素正规化、属性-评价对抽取和情感极性分类方法均比较简单，不够系统，直接影响到词典扩展性能。在将来的研究中，我们将系统探索上述这些问题，同时扩大语料规模和领域范围，以进一步提高领域相关的情感词典扩展性能。

#### 参考文献

- [1] Hiroshi Kanayama, Tetsuya Nasukawa. Fully automatic lexicon expansion for domain-oriented sentiment analysis[C]. Proceedings of EMNLP'06, 2006: 355-363.
- [2] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. Opinion word expansion and target extraction through double propagation. Computational Linguistics, 2011, 37(1): 9-27.
- [3] Andrea Esuli, Fabrizio Sebastiani. Determining the semantic orientation of terms through gloss classification[C]. Proceedings of CIKM'05, 2005: 617-624.
- [4] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics, 2009, 35(3): 399-434.
- [5] Yunfang Wu, Miaomiao Wen. Disambiguating dynamic sentiment ambiguous adjectives[C]. Proceedings of COLING'10, 2010: 1191-1199.

- [6] Alina Andreevskaia, and Sabine Bergler. When specialists and generalists work together: Overcoming domain dependence in sentiment tagging. Proceedings of ACL '08, 2008, pp.290-298.
- [7] Songbo Tan, Gaowei Wu, Huifeng Tang, and Xueqi Cheng. A novel scheme for domain-transfer problem in the context of sentiment analysis. Proceedings of CIKM '07, 2007, pp.979-982.
- [8] 吕韶华, 杨亮, 林鸿飞. 基于SimRank的跨领域情感倾向性分析算法研究[J]. 中文信息学报, 2012, 26(6): 38-44.
- [9] Azniah Ismail, Suresh Manandhar. Bilingual lexicon extraction from comparable corpora using in-domain terms[C]. Proceedings of COLING'10, 2010: 481-489.
- [10] Mingqing Hu, Bing Liu. Mining opinion features in customer reviews [C]. Proceedings of AAAI'04, 2004: 755-760
- [11] Li Zhang, Feng Jing, Xiaoyan Zhu. Movie review mining and summarization[C]. Proceedings of CIKM'06. 2006: 43-50.
- [12] Niklas Jakob, Iryna Gurevych. Using anaphora resolution to improve opinion target identification in movie reviews[C]. Proceedings of ACL'10, 2010: 263-268.
- [13] B.Wang, H.Wang. Bootstrapping both product features and opinion words from Chinese customer reviews with cross-inducing[C]. Proceedings of IJCNLP'08, 2008: 289-295.
- [14] F.L. Cruz, J.A. Troyano, F.J.Ortega, et al. Automatic expansion of feature-level opinion lexicons[C]. Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, 2011: 125-131.
- [15] Guohong Fu, Chunyu Kit, and Jonathan J. Webster. Chinese word segmentation as morpheme-based lexical chunking[J]. Information Sciences, 2008, 178(9): 2282-2296.
- [16] L. Page, S. Brin, R. Motwani, et al. The PageRank citation ranking: bringing order to the web[J]. 1999-66: Stanford Digital Library Technologies Project.