

● 张莉<sup>1,2</sup>, 苏新宁<sup>1</sup>, 王东波<sup>1</sup>

(1. 南京大学 信息管理学院, 江苏 南京 210093; 2. 南京大学 计算机科学与技术系国家重点实验室, 江苏 南京 210093)

## 通用领域中文评论的意见挖掘研究\*

**摘要:** 相对于传统的产品领域意见挖掘研究, 文章对中文通用领域的意见挖掘各部分内容进行了尝试性研究。利用基于多种语言特征和候选评价对象的条件随机场模型进行观点表达抽取, 对有窗口限制的最近邻方法进行改进, 提出一种评价对象—观点表达对的匹配算法, 其对评价对象抽取效果也进行了进一步的修正。

**关键词:** 通用领域; 意见挖掘; 评价对象; 观点表达

**Abstract:** Relative to the traditional research on opinion mining in the product area, this paper makes an experimental study of various part of content of opinion mining in universal Chinese area. The Conditional Random Fields (CRFs) model based on multilingual characteristics and candidate evaluation object is used to extract the opinion expressions. The paper improves the window-constrained nearest neighboring algorithm, and proposes a matching algorithm for the pairs of evaluation object and opinion expressions. The paper further revises the extraction effect of evaluation object.

**Keywords:** universal area; opinion mining; evaluation object; opinion expressions

随着互联网的发展, 网络上关于产品、电影和新闻等各种评论铺天盖地, 人们希望从中获得一些感兴趣的信息。Liu Bing 等在 2004 年首次提出了利用关联规则等方法基于产品特征总结倾向性<sup>[1]</sup>, 但是对于某一特征, 用户除了想知道其他用户对它的态度外, 还常常希望知道用户是如何评论它的。用户的具体评价即观点表达 (Opinion Expressions), 如果观点表达是一个词或词组, 也常常称其为评价词。这种以抽取评价对象、观点表达以及评价对象与观点表达的搭配关系等内容为任务的技术被称为意见挖掘 (Opinion Mining), 它是情感分析技术的发展。姚天昉等认为意见挖掘的主要子任务有主题抽取、意见持有者识别、陈述的选择和情感分析<sup>[2]</sup>。而从目前的研究来看, 意见挖掘的主要任务有从主观文本中抽取评价对象 (主题)、观点表达 (陈述中的情感部分)、观点表达和极性 (陈述中的主题部分与情感部分) 之间的关系和判断观点的倾向性 (情感语义确定), 除新闻领域外, 意见的持有者常常是评论者本人, 所以一般不进行持有者的抽取。意见挖掘结果可应用于互联网的多个领域中, 对网络用户起着重要的作用, 如有害信息过滤、社会舆情分析和商品性

能改进等领域, 具有很大的理论意义和应用价值, 对经济和政治均起着重要作用。由于网络文本形式和内容的多样性, 致使意见挖掘研究较为困难, 在研究中其与多种语言技术有关, 如文本挖掘、信息检索、信息抽取、语言学和统计学等; 涉及多个语言层面, 如词汇层、句法层和语义层等。

由于产品领域的的数据形式较为简单, 且有较明显的模式, 所以目前国外意见挖掘的研究绝大多数在产品领域, 如手机、DC、MP3 和汽车等, 主要着眼于对产品属性和观点表达以及它们之间搭配关系的抽取上, 而中文意见挖掘系统同样大多限制在产品领域, 且观点表达抽取研究工作甚少, 通常只判断极性<sup>[3]</sup>。Wiebe 等在 2005 年提出了抽取主观表达的重要性, 并将主观表达分为两类: 直接主观表达 (DSEs) 和非直接主观表达 (ESEs)<sup>[4]</sup>; 姚天昉等利用本体、极性词典和匹配规则构建了一个汉语汽车评论意见挖掘系统<sup>[2]</sup>; Kobayashi 等在 2007 年结合上下文线索和统计方法抽取特征和特征以及特征和评价词等之间的关系<sup>[5]</sup>; E. Breck 等把抽取主观表达任务作为标记问题<sup>[6]</sup>, 利用条件随机场模型<sup>[7]</sup> (Conditional Random Fields, 简称为 CRFs 或 CRF) 进行 DSEs 和 ESEs 的抽取; 张姝等在 2008 年第一届中文倾向性分析评测会 (COAE2008) 上提出使用 CRF 一体化抽取特征—评价词对的方法<sup>[3]</sup>; Wei Jin 等在 2009 年提出了基于多种语言特征的 HMMs 自学习系统 OpinionMiner 来抽取特征—观点表

\* 本文为教育部人文社会科学重点研究基地基金项目“基于智能信息处理的知识挖掘技术及应用研究” (项目编号: 08JJD870225) 和国家社会科学基金项目“基于语言特征的中文意见挖掘研究” (项目编号: 11CYY031) 的成果。

达对<sup>[8]</sup>；Wu 等在 2009 年提出了利用短语句法依存树结构，基于词性、候选评价对象和候选观点表达词典及依存关系等特征依靠 SVM 寻找特征、观点表达和特征—观点表达<sup>[9]</sup>。

目前国内外对于其他如新闻、旅游和医学等领域的观点表达抽取相关研究鲜见于报道，导致这些领域中隐藏在数据背后的观点无法被获知，这将在很大程度上限制普通用户和政府机构等理解这些领域中的对象并做出正确的决策，例如博客（包含微博）和新闻评论中有大量对于各类事件（包括事物）的评论，这些评论涉及各行各业，对于它们的理解十分有利于舆情的监控和分析，因此着手进行与领域无关的意见挖掘，特别是观点表达和评价对象—观点表达对的抽取十分必要和迫切。笔者将结合探讨现有的基于产品领域所提出的意见挖掘方法是否适用于通用领域（即各类领域而非特指产品领域）数据，对难度更大的基于通用领域的中文评论句的意见挖掘各部分进行研究。评价对象抽取目前大多使用机器学习方法，前期笔者提出了基于核心句和句法模式利用 CRFs 进行评价对象抽取的方法，实验结果证明该方法效果不错，领域适应能力强，具体可参考文献 [10]。本文着重对观点表达和评价对象—观点表达对的抽取进行研究，在获得观点表达抽取结果后对评价对象进行修正。

## 1 观点表达抽取

产品领域的产品属性通常为一个名词或名词词组，而观点表达通常为单个的词组，多数由一个褒义或贬义词构成，如：

例句 1：尼康 D7000 性能很强大，高感很适用，手感较差，套头配置令人失望。

例句 1 中的评价对象为“尼康 D7000”，其产品属性分别为“性能”、“高感”、“手感”和“套头配置”，而观点表达则要与评价对象或产品属性匹配。对于“尼康 D7000”，则其观点表达为“性能很强大”和“高感很适用”等，而对于“性能”则其观点表达为“很强大”，“高感”其观点表达为“很适用”，在以往的研究中观点表达常常指与产品属性匹配的词组。

而如新闻和旅游等领域的网络评论中通常不包含评价对象的属性，这些评论中的观点表达常常为一个或多个句子，如：

例句 2：Linux 主要由学术界和爱好者们发展起来，并一贯遵循开放、共享的原则，因此非常适合于教育。

例句 2 中的评价对象为“Linux”，观点表达为“一贯遵循开放、共享的原则”和“非常适合于教育”。

以上分析比较说明单纯依靠情感词词典和规则匹配等

方法并不适用于非产品领域评论句的观点表达抽取，因此需要寻找一种适合通用领域（包括产品领域）的抽取方法。

### 1.1 条件随机场模型

条件随机场模型是由 Lafferty 在最大熵马尔可夫模型（MEMM）基础上提出来的<sup>[8]</sup>。假定  $X$  和  $Y$  为随机变量，对于  $(X, Y)$ ，定义  $X = (x_1, x_2, \dots, x_n)$  为观察序列，定义  $Y = (y_1, y_2, \dots, y_n)$  为状态序列，即观察序列  $X$  的输出结果序列，随机变量  $X$  和  $Y$  是联合分布的，根据观察序列和状态序列对可构建一个条件模型  $P(Y|X)$ 。由于 MEMM 对状态序列的计算是局部的，所以会带来标记偏置（Label Bias）问题，而 CRFs 是一个在给定观察序列的条件下计算整个观察序列的联合条件概率分布的无向图模型，它对整个标记序列求解联合概率且在序列范围内归一化，因此可克服标记偏置问题。

CRFs 定义的条件概率分布为：

$$p(y|x) = \frac{1}{Z(x)} \exp \left\{ \sum_{i=1}^n \sum_k \lambda_k f_k(y_{i-1}, y_i, x, i) \right\}$$

其中  $Z(x)$  是只依赖于观察序列的归一化因子。

CRFs 最简单的形式为线性链式，对于一个训练样本的观察序列  $X$ ，CRFs 标注的目标就是找出训练样本对应的最可能的标注结果序列  $y$ ，即  $y = \arg \max P_\lambda(Y|X)$ ，其可以基于整个观察序列求解一个最优的标记序列，所以非常适合用来进行序列标注。CRFs 除良好的标注能力外，其也不像规则匹配方法那样对数据领域有特定的依赖，所以利用 CRFs 模型进行通用领域中评价句的观点表达抽取。

### 1.2 语言特征

可基于多种语言特征利用 CRFs 进行训练获得学习模型。词本身和词性是两种最基本的语言特征，由于一些观点表达本身就是词和词组，并且它们常常包含形容词和副词，所以词性特征对于这种类型的评价句是有利的。一些观点表达（如例句 2）也常常是以某些情感词为中心的句子，为此可考虑利用情感词、程度词和否定词等词典资源。另外对于这类句式观点表达其句法特征也会具有一定的规律，进一步可将句子的句法特征作为 CRFs 的语言特征。

### 1.3 评价对象作为特征

除考虑语言特征作为 CRFs 的特征外，以往有些研究还使用诸如领域词典和搭配词典等资源进行评价词的抽取。这种方式能较好地适用单个领域，但对于通用领域来说却不是一种好的方法。网络评论中常常出现评价对象与观点表达之间存在一定的关联的句子，例如：

例句 3：他的性格近乎完美，因此当我读到这则消息

的时候,我简直不敢相信(评价对象:他的性格 观点表达:近乎完美)。

鉴于这类句子的特征,我们将评价对象作为特征交由 CRFs 进行训练,这样可以进一步提高 CRFs 模型的效果,从而获得较好的标注结果。

因此 CRFs 模型所使用的特征选择如下:①词;②词性;③评价对象;④情感词典;⑤程度词词典;⑥否定词词典;⑦句子的原始句法,每一项均使用 Uni-gram 和 Bi-gram 特征。

#### 1.4 修正评价对象

利用若干特殊符号和观点表达结果可以对文献[10]中使用的评价对象抽取方法做进一步修正,修正方法为:①如果标注的评价对象出现在“《》”或“< >”中且只标出其内的部分内容则扩充评价对象为整个其内部分。②如果某一句中未标注了观点表达但未标出评价对象,则将距观点表达最近的名词或名词词组(若不含名词则不标注)标注为评价对象,标注后若出现方法①中所述情况,则按方法①进行扩充标注。

### 2 评价对象—观点表达对抽取

以往对于产品属性和观点表达对的抽取,不管是简单的词之间的线性窗口或是依存树之间的窗口,常常使用有窗口限制的最近邻原则,即将某一长度范围内(文献[8]的窗口大小为5)最相邻的产品属性和观点表达进行配对,如例句1中的产品属性“性能”、“高感”和“手感”与其匹配的观点表达都是其近邻“很强大”、“很适用”和“较差”。这种方式对于产品领域的产品属性—观点表达对抽取非常适合,而对于产品领域的评价对象—观点表达对以及其他领域的评价对象—观点表达对抽取并不适合。笔者从 COAE 2009 提供的数据包中随机抽取了20条产品领域的句子和20条不限领域的句子,前者产品属性与观点表达的词之间的窗口以及后者评价对象与观点表达的词之间的窗口比较如图1所示。

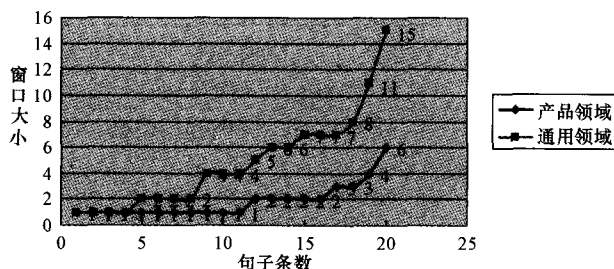


图1 领域搭配窗口大小比较

从图1和大量语料的观察可以发现,产品领域的产品属性和观点表达的窗口大小多数为1和2,少量会超过5,

所以以往搭配关系抽取中常限制窗口大小为5;与之相比,通用领域的评价对象和观点表达搭配的窗口大小明显变大。根据图1中随机抽取的40条句子来看,产品领域相关搭配的平均窗口大小为1.85,而通用领域则为4.75,如果按照以往只匹配窗口大小5之内的关系,则将有近50%的观点表达丢失。

通过观察和以上分析可以得知,各领域评价对象和观点表达紧邻的情况较为常见,另外评价对象距观点表达较远或者一个评价对象有多个观点表达的情况较多。对此提出一种改进的算法,即将未匹配的观点表达与其最近的评价对象匹配而不设置窗口限制,这样既适合与产品领域语言形式类似的领域,也适合一般的通用领域。另外也对标出了观点表达但未标注评价对象的句子进行处理,将与观点表达最靠近的名词或名词词组作为匹配的评价对象,并考虑了名词或名词词组出现在“《》”或“< >”中的情况。算法在寻找评价对象—观点表达对的过程中还对含观点表达的句子中未匹配的评价对象进行删除,进一步提高了评价对象抽取的效果。算法的具体步骤如下:

输入:

多条标注了评价对象和观点表达的句子

输出:

评价对象—观点表达对

Fun (S)

for 每一个句子 S 中未匹配的观点表达 O do

if 该句中不包含评价对象 P then

将 O 与其最近的 P 进行匹配,如果 O 与两个 P 之间的距离一样,则选择位于其前的 P 进行匹配  
else

寻找 O 最紧邻的名词或名词词组作为其匹配的 P,若名词或名词词组出现在“《》”或“< >”中则 P 为整个其内的内容

end if

end for

for 每个句子 do

if 句子 S 中含观点表达且有 P 未匹配 then

丢弃所有未匹配的 P

end if

end for

return (P, S)

以下分别以例句4、例句5、例句6为例来说明匹配算法的实现情况。

例句4:郭树清在内蒙古插过队,酒量不小,性格耿直,非常直率,很有看法,也很固执,是个思辨型学者,学术功底很深,也是个务实的人。

例句 5：其实，说起本片确有点怪怪的感觉。

例句 6：他也认为这些缺陷是极为严重的弱点。

例句 4 抽取出的评价对象为正确的“郭树清”，观点表达为正确的“酒量不小”、“性格耿直”和“非常直率”等，按照改进算法可将所有搭配抽取出来，而如果利用窗口限制最近邻方法则会丢掉若干搭配。例句 5 未抽取评价对象，抽取的观点表达为正确的“确有点怪怪的感觉”，按照改进算法将观点表达前的名词“本片”作为例句 5 的评价对象，搭配关系为 {本片，确有点怪怪的感觉}，此为正确的抽取和搭配结果。例句 6 抽取出的评价对象为“他”和“缺陷”，观点表达为“是极为严重的弱点”，根据改进算法确定的搭配关系为正确的 {缺陷，是极为严重的弱点}，将错误的评价对象“他”丢弃。从几个例句的分析可以看到，改进算法除了能进一步提高搭配数目外，还能提高评价对象的召回率。

### 3 实验与分析

实验由数据预处理、利用核心句抽取候选评价对象、标注观点表达、观点表达抽取及评价对象—观点表达抽取共 5 个阶段组成。

#### 3.1 数据预处理

采用的数据集为 COAE 2009 任务 4 已标注的数据，涉及体育、电影和手机等多个领域，每个句子只包含一个评价对象，如文献 [10] 所述对其进行了处理后获得句子 2 723 条，其中最长的句子含 400 多个字符，而最短的句子仅含几个字符。在前期的实验过程中发现，在 2 723 条句子中某些句子并不止一个评价对象，为此重新进行了一次扩展标注。

#### 3.2 利用核心句抽取候选评价对象

文献 [10] 中提出了基于 CRFs 模型，利用核心句和句法模式抽取评价对象的方法，由于数据集发生改变，必须重新抽取核心句和评价对象。仍然利用哈尔滨工业大学的 LTP 2.0 工具，用 5 折交叉验证方法获得评价对象抽取结果，抽取结果用精确率（Precision）、召回率（Recall）和  $F$  值（ $F$ -measure）来表示，其定义为：

$$\text{Precision} = \frac{\text{标出结果中包含的正确数目}}{\text{系统提交的数目}}$$

$$\text{Recall} = \frac{\text{标出结果中包含的正确数目}}{\text{答案数目}}$$

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

实验结果的精确率、召回率和  $F$  值分别为 0.7344、0.4636 和 0.5683。将 5 次的评价对象标注作为候选评价对象，在后续实验中其将作为 CRFs 的一项特征。

#### 3.3 标注观点表达

对补充标注过的 COAE 2009 任务 4 的数据集进行观点表达的标注，标注采用 COAE 类似的形式，标注者共有 3 名，取其共同部分，并由其中一人进行检查得到最终答案。共同部分的标注结果共有 1 965 条，标注方法仍然使用 IOB 2 形式，在标注时借助于最初标注的评价对象，并且在标注时采用相同的原则。如：

例句 7：哈林顿获得 33 分，是该队得分最高的球员。

例句 8：在 NBA 众多前锋中，马丁是防守最出色的一人，是受人尊敬的对手。

例句 9：从目前的效果来看，条例起到的作用是非常关键的，谣言不攻自破。

例句 10：黄浩然的表演似乎有些生硬，但陈小春已俨然吴镇宇般酷劲十足，其实我觉得应该由外表不羁的陈小春来饰演男主角可乐，这样才会令可乐是否报杀父之仇更有悬念。

例句 7 和例句 8 中标注出的评价对象分别是“哈林顿”和“马丁”，在标注观点表达时，如果例句 7 是“是该队得分最高的球员”，则例句 8 应该标“是防守最出色的一人”和“是受人尊敬的对手”，而不应标为“防守最出色”和“受人尊敬的对手”。

例句 9 和例句 10 在评价对象和观点表达标注时也需要有相似的标准，如例句 9 如果标出的评价对象为“条例”则例句 10 第一子句应标出的评价对象为“黄浩然”而不是“黄浩然的表演”，观点表达也需要根据评价对象的标注标准作相应规范，如例句 9 中标注“起到的作用是非常关键的”还是“是非常关键的”。

在标注过程中应该尽量使用相似的标准，因为规范的标注能提升 CRFs 模型的学习能力。

#### 3.4 观点表达抽取实验结果及分析

3.4.1 实验结果 根据 1.3 节所述，共选用了 7 种语言特征，在实验过程中为进行对比试验，笔者将所有特征组合成 5 种，具体形式如表 1 所示。

表 1 语言特征组合

语言特征名	语言特征
WPO	词 词性 观点表达的标注 tag
WPCDEO	词 词性 主张词 程度词 情感词 观点表达的标注 tag
WPKO	词 词性 候选评价对象 观点表达的标注 tag
WPRKO	词 词性 真正标注的评价对象 观点表达的标注 tag
WPCDEKSO	词 词性 主张词 程度词 情感词 候选评价对象 原始句法 观点表达的标注 tag

编写程序将句子按照表 1 所示的 5 种形式分别进行标

注, 然后利用 CRFs 模型进行学习, 实验结果如表 2 所示。实验结果按照 COAE 的方式采用精确评价 (Strict) 和覆盖评价 (Lenient) 两种形式, 对于精确评价, 要求抽取结果与答案完全匹配正确才算正确, 而覆盖评价则要求抽取结果与答案有重合就表示正确。本实验对覆盖评价做了更严格的要求, 即除精确标注外标出部分比原标注对象序列长的才算正确。另外根据方法比较的确定性有部分模式的覆盖评测未进行实验, 在实验中涉及的词典主要是 HowNet 词典<sup>[11]</sup>, 包括主张词 38 个, 程度词 219 个, 情感评价词褒贬合并后共 8 936 个。

表 2 观点表达抽取结果

特征组合名	观点表达抽取的结果 (Strict)			观点表达抽取的结果 (Lenient)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
WPO	0.4360	0.2701	0.3333	0.5709	0.3896	0.4630
WPCDEO	0.4365	0.2820	0.3422	/	/	/
WPKO	0.4595	0.3105	0.3702	0.6334	0.4262	0.5092
WPRKO	0.5850	0.4389	0.5014	/	/	/
WPCDEKSO	0.4420	0.3220	0.3723	0.6736	0.4897	0.5667

3.4.2 基准 (Baseline) 由于目前意见挖掘研究的方法所针对的数据领域基本为产品领域, 根据观点表达抽取的分析, 该领域的观点表达常常为一个含褒贬含义的词或词组, 所以以往提出的方法也都是适合这类数据的, 但对于研究通用领域的观点表达抽取不具有参考性。为了进行对比, 使用文献 [8] 中提出的根据词性确定评价对象和观点表达抽取的 Baseline, 并对其中的匹配模式进行了扩充, 将其中如“副词 形容词 名词/名词词组”中的形容词改成一个或多个连续的形容词, 其他出现形容词的模式也作相应调整。实验结果 (包含评价对象) 如表 3 所示。

表 3 评价对象和观点表达抽取结果

抽取对象	抽取结果 (Strict)		
	Precision	Recall	F-measure
评价对象	0.0517	0.0335	0.0406
观点表达	0.0242	0.0105	0.0146

3.4.3 实验结果分析 根据表 2 和表 3 的实验结果和前文有关观点表达的分析可以发现:

1) 观点表达抽取的覆盖评测结果与精确评测相比有较大的性能提高, 其中 WPO 模式覆盖评测比精确评测的  $F$  值提高了 12.97%, WPCDEKSO 模式的  $F$  值提高了 19.44%。这主要是因为通用领域的观点表达常常为一个或多个句子, 在 CRFs 学习后某些情况下会漏掉一两个非关键的字, 如“的”, 另外也常把两个连续的句子作为一

个观点表达 (标注时除少量特殊情况外通常将由逗号或其他标点符号分隔的几个句子作为几个观点表达), 这样的情况在覆盖评测中都进行了修正。

2) 除词和词性外其他几个语言特征对于观点表达的抽取均有作用, WPCDEO 与 WPO 相比精确评测的  $F$  值提高了近 1%, WPKO 与其相比  $F$  值提高了近 4%, WPCDEKSO 与 WPKO 相比其  $F$  值也略有提高; 而覆盖评测 WPKO 与 WPO 相比  $F$  值有了 4% 的提高, WPCDEKSO 与其相比  $F$  值有了 10% 的提高。图 2 所示为实验结果与文献 [9] 中提出的产品领域的实验结果的比较。

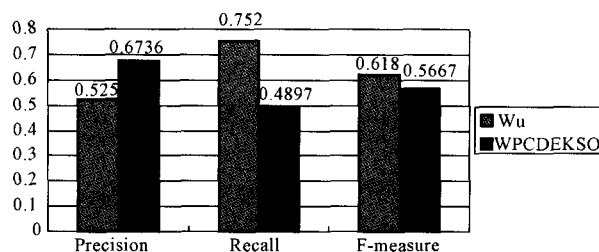


图 2 实验结果比较

在图 2 中, Wu 等是利用收集到的产品领域的 8 221 条观点表达与句子进行匹配从而获得句子中的观点表达, 这种方法需要较多的人工干预且领域适应度较差, 而笔者提出的方法领域适应度好, 另外尽管是通用领域, 但  $F$  值仍能达到 0.5667, 这也可以说明本文所选用的 CRFs 的语言特征较好以及基于 CRFs 这类机器学习的方法抽取通用数据的观点表达方法的有效性。但是也可以看到通用数据领域中观点表达形式多样, 情况复杂, 所以抽取效果与产品领域相比还有一定的差距。

3) 实验结果显示, 评价对象对于观点表达的抽取结果有重要影响。WPKO 是利用候选评价对象抽取观点表达的模式, 其召回率和  $F$  值与 WPO 相比仅提高 1% 左右, 而 WPRKO 是利用真正人工标注的评价对象进行抽取, 召回率和  $F$  值与 WPO 相比提高了近 16%, 所以提高通用领域的评价对象抽取效果与产品领域一样, 同样可以对观点表达抽取带来很大方便。

4) 从表 3 的 Baseline 实验结果可以看到, 基于产品领域的词性模式匹配在通用领域的效果非常不理想, 观点表达抽取结果的  $F$  值仅为 1.46%, 而该方式在产品领域的  $F$  值通常能达到 60% 左右, 由此可见通用数据领域的的数据复杂度较高, 抽取观点表达很难利用模式匹配或构建固定搭配数据库的方法来完成。相比而言, 利用机器学习模型方法进行抽取具有显著优势。

### 3.5 评价对象—观点表达对抽取实验结果及分析

1) 实验结果和 Baseline。利用 WPCDEKSO 模式的学

习结果作为抽取评价对象—观点表达对的来源数据，有窗口限制的最近邻方法和我们的方法实验结果比较见表4。表4中结果由匹配个数和匹配率即 $P$ 值表示。

表4 评价对象—观点表达对抽取结果

方法	匹配数
窗口限制最近邻	31.38% (904/2881)
改进的最近邻	40.65% (1171/2881)

两种方法对于原观点表达的抽取结果影响不一，笔者提出的方法观点表达的抽取结果基本没改变，而窗口限制最近邻法在处理完后因为删除了较多的观点表达，其抽取结果的精确率提高，但召回率降低，其覆盖评测的精确率为0.6858(WPCDEKSO模式覆盖评测的精确率为0.6736)，召回率为0.3924(WPCDEKSO模式覆盖评测的召回率为0.4897)。

另外，如1.4节所述，可以根据观点表达与评价对象的搭配关系以及特殊字符的扩展进一步修正评价对象，表5所示为评价对象修正后与修正前抽取结果的比较。

表5 评价对象抽取结果比较

形式	抽取结果比较 (Lenient)		
	Precision	Recall	F-measure
修正前	0.7344	0.4636	0.5683
修正后	0.6496	0.6086	0.6283

2) 实验结果分析。表4的实验结果显示窗口限制的最近邻方法应用在通用数据的观点表达抽取上效果不佳，而笔者提出的改进最近邻方法在匹配数上有了10%的提高，证明了该方法的可行性。但同时也可以看到，目前的匹配效果并不是很好(产品领域关系匹配的 $P$ 值一般在50%~60%之间)，究其原因，主要是因为通用领域的评价对象和观点表达抽取本身有较大难度，抽取效果不是非常理想，导致匹配数目不是很高。从表5的实验结果可以看到评价对象修正后其召回率提高了近14%，由此可以证明修正方法是有效的。

#### 4 结束语

本文主要研究了中文通用数据领域的意见挖掘方法，对于其中的观点表达抽取方法和评价对象—观点表达对搭配算法进行了重点研究，实验结果证明了本文所述方法的可行性。但同时也看到通用数据领域的意见挖掘非常困难，相比产品领域的相应任务更有难度，尚有很多工作要做。在后续的工作中，将在以下几方面进行进一步的研究。

1) 因为评价对象能较大程度地影响观点表达的抽取

结果，所以提高评价对象的抽取精确率十分重要，后续将寻找更多语言特征和算法来进一步提升通用领域中评价对象的抽取结果。

2) 对评价对象—观点表达对的搭配算法做后续的修正以提高搭配数。

3) 对于目前提出的各种方法进行改进，以提升其领域适应度。

4) 研究发现，因为词性标注和分词等基础工作的失误，会对意见挖掘各部分研究带来一定问题，因此在进行意见挖掘研究中也关注自然语言基础的改善工作。

感谢哈尔滨工业大学信息检索研究室为本研究提供了LTP 2.0工具和COAE 2009提供的语料。□

#### 参考文献

- [1] HU Mingqing, LIU Bing. Mining and summarizing customer reviews [C]. KDD' 04, 2004: 168-177.
- [2] 姚天昉, 聂青阳, 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25周年学术会议论文集, 2006: 260-281.
- [3] 张姝, 贾文杰, 夏迎炬, 等. 基于CRF的评价对象抽取技术研究 [C] // 第一届中文倾向性分析评测 (COAE 2008) 论文集, 2008: 70-76.
- [4] WIEBE J, WILSON T, CARDIE C. Annotating expressions of opinions and emotions in language [C] // Language Resources and Evaluation. 2005, 39 (2/3): 165-210.
- [5] KOBAYASHI N, INUI K, MATSURNOTO Y. Extracting aspect-evaluation and aspect-of relations in opinion mining [C]. EMNLP-CoNLL 2007, 2007: 1065-1074.
- [6] BRECK E, CHOI Y. Claire cardie: identifying expressions of opinion in context [C]. IJCAI-2007, 2007: 1-6.
- [7] LAFFERTY J, McCALLUM A, PEREIRA F. Conditional random fields: probabilistic models for segmenting and labeling or sequence data [C]. ICML, 2001: 282-289.
- [8] WEI Jin, HO H H, SRIHARI R K. Opinion miner: a novel machine learning system for Web opinion mining and extraction [C]. KDD' 09, 2009: 1195-1203.
- [9] WU Yuanbin, ZHANG Qi, HUANG Xuanjing, et al. Phrase dependency parsing for opinion mining [C]. ACL/EMNLP' 09, 2009: 1533-1541.
- [10] 张莉, 钱玲飞, 许鑫. 基于核心句及句法结构的评价对象抽取 [J]. 中文信息学报, 2011 (5): 23-29.
- [11] <http://www.keenage.com>.

作者简介: 张莉, 女, 1976年生, 博士生, 讲师。研究方向: 文本挖掘, 自然语言处理。

苏新宁, 男, 1955年生, 教授, 博士生导师。研究方向: 信息处理与检索, 信息分析与评价。

王东波, 男, 1981年生, 博士生。研究方向: 自然语言处理。

收稿日期: 2011-10-10