基于浅层句法特征的评价对象抽取研究

徐 冰 [1] 赵铁军 [1] 王山雨 [1] 郑德权 [1]

摘 要 随着网络评论文本数量的快速增长, 文本情感分析越来越受到研究者的广泛关注. 句子级文本情感分析就是对主观性文本进行细粒度的挖掘, 有重要的研究价值. 评论句中的评价对象抽取是句子级情感分析要研究的关键问题之一. 为了提高评价对象抽取的性能, 本文提出在系统模型的训练过程中引入浅层句法信息和启发式位置信息, 同时在不增加领域词典的情况下, 有效提高系统的精确率. 实验结果表明, 将本文提出的特征引入到条件随机域模型和对比模型后, 系统的各项指标均有所提高, 并且条件随机域模型的结果优于对比模型. 同时, 将条件随机域模型的结果与 2008 年国内中文评测的最大值比较, 其 F 值超过最大值 5 %.

关键词 文本情感分析, 意见挖掘, 评价对象抽取, 浅层句法特征, 位置特征

DOI 10.3724/SP.J.1004.2011.01241

Extraction of Opinion Targets Based on Shallow Parsing Features

XU Bing[1] ZHAO Tie-Jun[1] WANG Shan-Yu[1] ZHENG De-Quan[1]

Abstract With the rapid development of the world wide web, more and more common users express their opinions on the web and many researchers pay more attentions to sentiment analysis. Fine-grained sentiment analysis on sentence level is very important. The extraction of opinion targets from opinion sentence is the key issue to sentence level of sentiment analysis. To improve the performance of opinion targets extraction, this paper proposes to integrate shallow parsing features and heuristic position information for modeling of the training process without introducing domain lexicon. The experiment results show that after adding the proposed features, nearly all specifications of both conditional random fields and contrast model are improved, and the results of conditional random fields are more efficient than that of the contrast model. Meanwhile, compared with the best results of the 2008 Chinese opinion analysis evaluation, the F measures of conditional random fields are 5 % higher than the maximum value.

Key words Sentiment analysis, opinion mining, opinion targets extraction, shallow parsing feature, position feature

文本情感分析是近几年在自然语言处理研究领域中的热点问题. 文本情感分析也称为观点挖掘, 就是在大量网络评论性文本的基础上对说话人的态度、观点、情感进行分析, 即对文本中的主观性信息进行分析[1−3]. 情感分析已在电影评论、产品评价、用户反馈等众多领域中得到了尝试并取得了初步成果[4−6]. 通常, 文本情感分析可根据研究对象的粒度分为词语级、句子级和篇章级等. 其中词语级和句子级的文本情感倾向性分析属于细颗粒度的文本分析, 句子级分析需要对主观性文本进行深层次挖掘, 研究难度较大[7].

句子级情感分析的处理对象就是在特定上下文中出现的语句. 其任务就是对句子中的各种主观性信息进行分析和提取, 包括判断句子的情感倾向, 以及从中提取出与情感倾向性论述相关联的各个要素. 这些要素包括情感倾向性论述的持有者、评价对象、倾向极性和强度, 甚至是论述本身的重要性等. 句子级情感分析以词语级的情感分析为基础, 为篇章级情感分析和海量文本的整体倾向性预测提供有效支持, 具有重要的研究价值.

在句子级情感分析的研究中, 已有研究主要都是针对产品评论进行细粒度的挖掘. 在产品评论中, 用户最为关心的是产品各部分属性的评价, 所以针对产品评论语料, 首要任务是找出句子中作者评价的对象, 即产品本身和产品属性, 然后才能对所评价对象的倾向性做出判别. 评价对象抽取与传统的信息抽取任务不同, 抽取的信息必须是作者评论过的, 而非客观描述的对象, 所以在抽取时要考虑主观性因素 (如评价词语) 对评价对象抽取的影响, 本文提出引入启发式的评价词位置信息来提高系统对评价对象的识别能力. 评价对象抽取另一个难点是如何保证完整地抽取出评价对象, 以产品属性抽取为例,

收稿日期 2010-11-22 录用日期 2011-03-24
Manuscript received November 22, 2010; accepted March 24, 2011
国家高技术研究发展计划 (863 计划) (2006AA010108), 国家自然科学基金 (61073130), 语言语音教育部−微软重点实验室开放基金 (HIT.KLOF.2009019) 资助
Supported by National High Technology Research and Development Program of China (863 Program) (2006AA010108), National Natural Science Foundation of China (61073130), and MOE-MS Key Laboratory Opening Funding of Natural Language Processing and Speech (HIT.KLOF.2009019)
1. 哈尔滨工业大学计算机科学与技术学院机器智能与翻译研究室 哈尔滨 150001
1. Machine Intelligence and Translation Laboratory, School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001

在产品评论中对产品属性的描述往往是比较复杂和多样的, 产品属性可以是词 (如 “屏幕”), 也可以是短语 (如 “手机外壳的接合处”), 在现有的研究中对产品属性的定位效果尚可, 但评价对象的边界识别效果较差, 其原因在于以往的特征集合中仅引入词语级别的特征, 不能有效解决短语级别的评价对象边界识别问题. 传统方法是通过人工构建产品属性词典的方法来解决这个问题, 但是人工构建词典费时费力, 而且可移植性差, 所以本文提出在特征集中引入浅层句法信息, 在不增加领域词典知识的条件下, 有效地抽取出短语级别的评价对象, 提高产品属性抽取的完整性.

综上所述, 本文提出在系统模型的训练过程中引入浅层句法信息和启发式位置信息, 在不依赖领域词典知识的情况下, 能够提高系统的抽取性能. 实验结果表明, 将本文提出的特征加入到条件随机域模型和对比模型 (最大熵模型) 后, 系统的各项指标提高明显, 说明本文提出的方法是有效的.

本文以下部分组织如下: 第 1 节介绍了产品属性抽取的国内外相关研究; 第 2 节介绍了基于条件随机域模型的产品属性抽取过程; 第 3 节描述了引入模型的特征集合; 第 4 节详细介绍了实验过程及对实验结果的比较分析, 最后是本文的结论及下一步工作.

1 产品属性抽取相关研究

在网络上的各种评论中, 用户产品评论占有很大的比例. 对于想购买某种产品的用户而言, 总希望从大量的评论文本中获得相关产品属性的有效评论信息, 所以近几年研究句子级情感分析中评价对象抽取问题时都以产品属性抽取为最主要的应用.

产品属性抽取方法主要有基于规则/模板的方法和基于统计的方法. 2005 年 Liu 等[8] 实现了用于比较用户评论中产品属性的原型系统 Opinion Observer. 该系统采用关联规则挖掘技术抽取产品属性, 并采用和抽取出的产品属性最近的形容词作为评价词的方法确定产品属性的倾向性. Popescu 等定义了评价对象和评价词模板, 采用基于规则的方法抽取产品属性[9]. 2007 年姚天昉等研制了一个汉语汽车评论的意见挖掘系统, 该系统采用领域知识和语法关系树挖掘用户对各种汽车的性能及表现的态度[10]. 基于模板与规则的方法有一定的应用价值, 但是仍然存在一定的局限性, 一般只能处理句型比较简单、评价词与评价对象距离比较近的句子. 所以, 很多学者开始采用基于统计学习的方法来抽取产品属性. Zhang 等采用有监督学习中的 CRFs 模型, 在模型训练的过程中选择词语级特征, 并引入领域词典的知识抽取产品属性, 在 2008 年中文文本倾向性分析评测中该方法获得了第一名[11]. 刘鸿宇等采用无监督方法抽取产品属性, 首先使用句法分析结果获得候选评价对象, 然后结合 PMI 算法和名词剪枝算法对候选评价对象进行筛选获得最终抽取结果, 该方法的评测结果略低于有监督方法的最好结果[12]. 与无监督方法相比较, 有监督方法需要人工标注训练语料, 但有监督方法的精确率比无监督方法的精确率高, 更具实用价值. 本文采用有监督方法实现评价对象抽取任务, 与现有研究的区别在于, 在模型的训练过程中, 不依赖领域词典知识, 在词语级语言学特征基础上, 引入浅层句法特征和启发式位置特征, 有效地提高了评价对象抽取的精确率.

2 基于条件随机域模型的产品属性抽取

2.1 条件随机域模型

条件随机域模型 (Conditional random fields, CRFs) 于 2001 年由 Lafferty 等提出[13]. 该模型已经广泛应用到人工智能的众多领域, 例如词性标注、句法分析、名实体识别等, 并且取得了非常好的效果[14−15]. CRFs 模型是最大熵模型 (Maximum entropy model, MEM) 和隐马尔科夫模型 (Hidden Markov model, HMM) 的扩展, 它结合了两种模型的优点. 该模型的主要优点是比产生式模型 HMM 更便于使用各种复杂的重叠的、非局部的特征, 来捕捉状态之间的依赖关系. CRFs 模型克服了其他判别式马尔科夫模型常见的标记偏置问题 (Label bias problem). 标记偏置问题的产生是由于有的判别式马尔科夫模型, 如最大熵马尔科夫模型 (Maximum entropy Markov model, MEMM), 是针对每个节点进行归一化, 因此有可能使条件概率值陷入局部最小, 导致标记错误, 即产生了标记偏置问题. 而 CRFs 模型针对所有的状态序列进行归一化, 克服了标记偏置问题, 因此精度又高于这些判别式马尔科夫模型.

CRFs 模型的基本思想是通过依赖于少数变量的局部函数的乘积来表示一个依赖于大量随机变量的概率分布. 图模型的优势在于可以对相互依赖的多变量进行建模, 而一般分类器只可以预测一个简单的类别变量. 由于评价对象抽取任务可以看作是一个序列标记问题, 本文采用 CRFs 模型中的线性链状结构来解决, 如图 1 所示.

在序列标注任务中, 随机变量 $X = X_1, \cdots, X_{n-1}, X_n$ 是观察序列; 随机标量 $Y = Y_1, \cdots, Y_{n-1}, Y_n$ 表示观测序列对应的标记序列, 随机变量 Y 的链式条件概率分布如式 (1) 所示.

$$P(Y|X) = \frac{1}{Z(X)} \times \exp\left(\sum_{i,k} \lambda_k f_k(y_{i-1}, y_i, x) + \sum_{i,k} \mu_k g_k(y_i, x)\right) \tag{1}$$

f_k 和 g_k 都是 0–1 函数, 其中, f_k 是从位置 i 到位置 $i-1$ 的转移特征函数, g_k 是位置 i 上的状态特征函数; 而 λ_k 和 μ_k 分别是在训练过程得到的权值. Z_x 是只依赖于观测序列 X 的归一化函数 (Normalizing factor). 在序列标注的过程中, CRFs 模型使用了 Viterbi 方法来寻找可以使得 $P(Y|X)$ 最大的 Y^*, 即 $Y^* = \max_Y P(Y|X)$, 而其训练过程采用了基于极大似然的迭代算法.

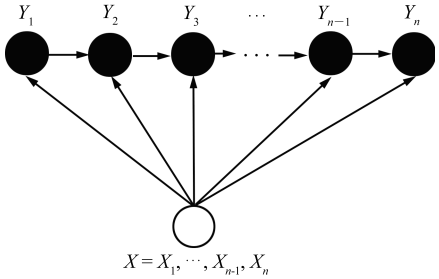


图 1　链式的无向图模型

Fig. 1　The chain of undirected graph model

2.2　产品属性抽取过程

产品属性抽取就是抽取产品本身和产品各种属性. 以手机的产品评论为例说明, 其中包含表 1 所示的几个方面.

采用 CRFs 模型来抽取产品属性的优点在于可以方便地在模型中引入丰富的特征集, 不需要考虑这些特征之间是否独立. 新的特征引入到模型中时, 不需要对模型本身做修改. 在对产品属性抽取的过程中, 本文引入了词特征、词性特征、上下文特征、浅层句法特征及位置特征和通用评价词词典信息. 抽取过程如图 2 所示.

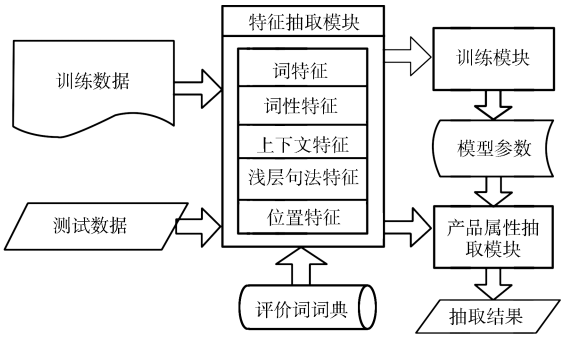


图 2　产品属性抽取过程

Fig. 2　Procedure of product attributes extraction

本文采用 CRF++–0.53 工具包来实现条件随机域模型的训练和测试. 在 CRF++–0.53 工具包中提供了两个文件: 1) crf-learn 可以根据特征模板从观察标注序列中生成 CRFs 特征函数, 并对其参数进行训练, 建立 CRFs 模型; 2) crf-test 使用建立的 CRFs 模型对输入的测试数据进行标注, 为每一个观察组选择最佳抽取结果.

3　CRFs 模型的特征集合

3.1　特征选择

3.1.1　浅层句法特征

浅层句法分析, 也叫语块分析. 与传统的完全句法分析不同, 浅层句法分析只要求识别出句子中的某些结构相对简单的成分, 如名词短语、动词短语等. 这些识别出来的结构通常被称为语块 (Chunk), 经过浅层句法分析后, 每个词都对应一个语块标记. 浅层句法分析技术已经应用在机器翻译、信息获取和自动文摘等领域, 并获得了令人满意的结果.

在产品评论中, 大部分的产品属性都是由名词和名词短语构成的. 而在传统的方法中, 在特征集合里引入词和词性特征, 仅能识别出名词性的产品属性, 对于短语级别的产品属性识别的效果较差, 结果导致了评价对象边界识别效果不好的问题. 所以, 我们提出通过浅层句法分析工具识别出产品评论中的名词短语,并通过对名词短语做标记的方式在特征模

表 1　产品属性抽取的例子

Table 1　Examples of product attributes extraction

定义	例子	抽取的评价对象
产品属性	Nokia 6600 的体积太大了	体积
产品的附属物	Nokia 3250 的屏幕很绚丽	屏幕
产品附属物的属性	Moto v3 的屏幕分辨率不是很高	屏幕分辨率
产品的相关概念	用 Nokia 3250 拍出来的照片比较清晰	照片
产品相关概念的属性	Moto v3 的摄像头的分辨率一般, 所以拍出来的照片的清晰度不尽如人意	摄像头的分辨率、照片的清晰度

板中加入浅层句法特征, 从而提高短语级别的评价对象识别能力. 从实验结果可以看到, 引入浅层句法特征有效提升了产品属性抽取的精确率.

例如, 在手机语料中, 下面的例句经过浅层句法分析处理后, 得到的形式为:

NP[外壳/n 的/ude1 接合/v 处/q] 还/d 比较/d 严实/a ./wj

该句中, 被评价的产品属性是 "外壳的接合处", 通过词性标记仅能抽取出名词 "外壳", 由于 "接合处" 没有切分成词, 所以引入词和词性特征不能抽取出正确结果. 通过浅层句法分析后, 能够标记出名词短语 "外壳的接合处", 所以引入浅层句法特征能够有效地抽取出短语级别的产品属性.

3.1.2 位置特征

在产品评论中要抽取的产品属性和其评价词通常距离很近. 根据统计, 在实验语料中大约 20 % 的产品属性可以通过和评价词就近匹配的方式抽取出来. 所以通过这种启发式位置信息, 可以进一步抽取出与评价词距离较近的产品属性. 本文将启发式位置信息称为位置特征. 首先通过通用评价词词典将语料中出现的评价词做标记, 然后对距离评价词前后最近的名词短语做标记, 这些名词短语将作为候选产品属性.

例如: 蓝牙耳机使用简单, 而且效果很好.

"简单" 和 "好" 都在评价词词典中出现, "蓝牙耳机" 和 "效果" 是距离评价词最近的名词短语, 记录其位置信息并作为特征加入特征模板.

3.1.3 词特征

由于在中文里词是表达语义的最小单位, 所以在处理中文语料时分词显得尤为重要. 在本文中我们以分词的结果作为模型输入进行处理, 更有利于系统抽取评价对象.

3.1.4 词性特征

由于产品属性主要以名词和名词短语为主, 所以名词的词性标注对产品属性抽取起到重要作用. 通过识别词性特征的标记, 能够引入更多的词性知识帮助抽取产品属性.

3.1.5 上下文特征

上下文特征用于获取词、词性和短语的上下文信息. 考虑到评论句的平均长度, 实验证明采用 $\{-4,+4\}$ 为上下文窗口获得目标词附近的语言学信息较为合适.

3.2 CRFs 模型的特征模板

根据以上语言学特征, 表 2 所示为若干组 n 元 (n-gram) 特征模板. 图中 w 表示词特征, p 表示词性特征, c 表示浅层句法分析特征, h 表示启发式位置特征. 为了验证浅层句法特征和启发式位置特征的作用, 我们考虑了四种特征组合, 分别是词语级特征模板 (T1)、在词语级特征基础上仅引入启发式位置特征的特征模板 (T2)、仅引入浅层句法分析特征的特征模板 (T3) 和同时引入两种特征的特征模板 (T4).

表 2 特征模板

Table 2 Feature templates

模板名称	特征选择	特征模板
T1	w,p	$w_n,p_n,n\in\{-4,\cdots,4\}$ $w_{n-1}w_n,p_{n-1}p_n,n\in\{0,\cdots,3\}$ $w_nw_{n+1},p_np_{n+1},n\in\{0,\cdots,3\}$
T2	w,p,h	$w_n,p_n,h_n,n\in\{-4,\cdots,4\}$ $w_{n-1}w_n,p_{n-1}p_n,h_{n-1}h_n,n\in\{0,\cdots,3\}$ $w_nw_{n+1},p_np_{n+1},h_nh_{n+1},n\in\{0,\cdots,3\}$
T3	w,p,c	$w_n,p_n,c_n,n\in\{-4,\cdots,4\}$ $w_{n-1}w_n,p_{n-1}p_n,c_{n-1}c_n,n\in\{0,\cdots,3\}$ $w_nw_{n+1},p_np_{n+1},c_nc_{n+1},n\in\{0,\cdots,3\}$
T4	$w,p,$ c,h	$w_n,p_n,c_n,h_n,n\in\{-4,\cdots,4\}$ $w_{n-1}w_n,p_{n-1}p_n,c_{n-1}c_n,h_{n-1}h_n,n\in\{0,\cdots,3\}$ $w_nw_{n+1},p_np_{n+1},c_nc_{n+1},h_nh_{n+1},n\in\{0,\cdots,3\}$

4 实验设置及结果对比分析

本文采用的实验数据来自于 2008 年中文倾向性评测 (COAE 2008) 发布的任务 3 语料[15], 主要包括笔记本电脑、手机和数码相机三个电子产品领域和汽车领域的语料, 大约 15 000 句. 表 3 中是语料的具体分布情况. 本文采用的浅层句法分析工具为哈工大机器智能与翻译实验室开发, 分词工具是中科院分词工具 ICTCLAS 3.0, 评价词词典采用知网的 "情感分析用词语集".

表 3 COAE 2008 语料分布情况

Table 3 Data of COAE 2008

名称	笔记本	手机	数码相机	汽车	总数
文本数	66	123	137	161	487
产品属性个数	1 031	2 396	1 964	2 616	8 007

评价指标采用精确率、召回率和 F 值. 计算公式分别如下:

$$\text{精确率}=\frac{\text{正确识别的产品属性个数}}{\text{系统识别出的产品属性个数}}\times 100\,\% \quad (2)$$

$$\text{召回率}=\frac{\text{正确识别的产品属性个数}}{\text{实际的产品属性个数}}\times 100\,\% \quad (3)$$

$$\text{F 值}=\frac{2\times\text{精确率}\times\text{召回率}}{\text{精确率}+\text{召回率}} \quad (4)$$

评价方法采用精确评价以及覆盖评价两种评价方式. 对于精确评价, 系统抽取结果与答案完全匹配才算正确; 对于覆盖评价, 系统抽取结果与答案有重合就算正确. 为了获得合理可信的结果, 我们结合语料的规模在电子产品领域采用了 6 折交叉验证的方式进行实验, 采用前面介绍的四种特征模板分别进行产品属性抽取的实验. 6 折交叉验证即为把实验数据随机均分为 6 个子集, 其中 5 个子集做训练集, 1 个子集做测试集, 循环 6 次. 实验结果如表 4 所示.

表 4　CRFs 模型 6 折交叉验证结果

Table 4　6-fold cross validation results of CRFs model

		精确评价			覆盖评价		
数据集	特征模板	精确率	召回率	F 值	精确率	召回率	F 值
数据集 1	T1	54.21	44.03	48.59	70.36	57.16	63.08
	T2	56.96	45.96	50.87	72.69	58.66	64.92
	T3	58.02	46.53	51.65	74.49	59.74	66.30
	T4	60.25	51.21	55.36	75.19	63.90	69.09
数据集 2	T1	53.47	43.74	48.11	70.38	57.57	63.33
	T2	54.56	44.76	49.18	70.27	57.65	63.34
	T3	59.79	49.30	54.04	76.24	62.86	68.91
	T4	61.28	51.90	56.20	76.49	64.79	70.16
数据集 3	T1	50.52	39.18	44.13	68.09	52.81	59.48
	T2	53.97	47.21	50.36	68.92	60.30	64.32
	T3	58.97	50.58	54.45	76.22	65.38	70.39
	T4	56.33	49.60	52.75	72.22	63.59	67.63
数据集 4	T1	52.46	42.00	46.65	70.04	56.09	62.29
	T2	53.77	45.00	48.99	70.65	59.13	64.38
	T3	57.18	45.18	50.48	73.85	58.34	65.19
	T4	54.94	45.82	49.97	73.39	61.22	66.75
数据集 5	T1	53.15	42.19	47.04	70.28	55.79	62.20
	T2	54.53	42.58	47.82	72.09	56.29	63.22
	T3	57.70	50.23	53.70	74.00	64.41	68.87
	T4	59.09	48.20	53.09	76.31	62.25	68.56
数据集 6	T1	54.39	41.89	47.33	75.03	57.80	65.30
	T2	57.30	46.79	51.52	71.49	58.38	64.27
	T3	58.82	49.35	53.67	76.32	64.04	69.64
	T4	56.83	46.26	51.00	76.37	62.16	68.54
平均值	T1	53.03	42.17	46.98	70.70	56.20	62.61
	T2	55.18	45.38	49.79	71.02	58.40	64.08
	T3	58.41	48.53	52.99	75.19	62.46	68.22
	T4	58.12	48.83	53.06	75.00	62.99	68.46

从表 4 的实验结果可以看到, 对于每个数据集, 在词语级特征 (T1 模板) 基础上分别引入位置特征 (T2 模板) 和浅层句法特征 (T3 模板) 后, 在精确评价和覆盖评价方式上, 精确率、召回率和 F 值都明显提高, 而且 T3 模板均比 T2 模板增加的幅度要大; 在同时加入位置特征和浅层句法特征 (T4 模板) 后, 每个数据集上的变化不一致, 对于数据集 1 和数据集 2, 加入 T4 特征模板后系统精确率、召回率和 F 值的结果达到最大值; 对于数据集 3 和数据集 6, 同时加入这两种特征后, 其各项指标有所降低, 而对应 T3 模板的实验结果达到最大值; 对于数据集 4, 加入 T4 特征模板后系统精确率下降, 召回率提高, 而对于数据集 5, 与数据集 4 的情况正好相反.

从以上的实验结果可以分析出: 1) 在词语级特征基础上引入位置特征或浅层句法特征对产品属性抽取都是有效的, 两个特征相比较, 位置特征使产品属性抽取的 F 值平均提升了 2%, 而浅层句法特征使 F 值平均提升了 6%, 所以浅层句法特征对系统抽取结果影响更大; 2) 对于每个数据集, 在词语级特征基础上同时加入位置特征和浅层句法特征后对产品属性的抽取结果影响不确定, 但从所有数据集的平均值上来看, 两种特征同时加入的实验结果略好于加入浅层句法特征的结果.

对于上述结论 2), 我们认为每个数据集中的就近匹配比例可能对位置特征的影响比较大, 因此我们对每个数据集上的就近匹配比例进行了统计, 并和两种评价方式下的 T3 模板和 T4 模板的 F 值进行了比较. 就近匹配比例即为就近匹配的正确产品属性个数与实际产品属性个数之比. 实验结果如图 3 所示.

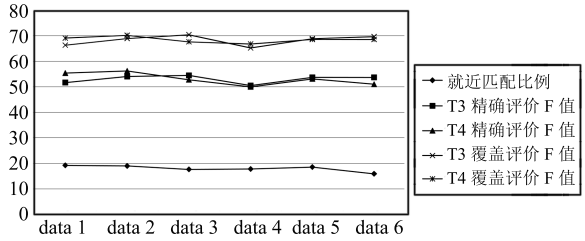


图 3　就近匹配比例对 F 值的影响

Fig. 3　Influence of nearest matching ratio on F-measure

从图 3 中可以观察到, 在随机选择的 6 个数据集上, 数据集 1、数据集 2 和数据集 5 上的就近匹配比例比较高, 数据集 3、数据集 4 和数据集 6 上的就近匹配比例相对较低; 就近匹配比例的变化趋势与精确匹配方式下 T4 模板的 F 值变化一致.

从以上观察的结果可以分析出: 1) 就近匹配的比例变化对精确匹配评价方式下的 F 值影响较大, 对覆盖匹配评价方式下的 F 值影响较小; 2) 在精确

匹配评价方式下, 就近匹配比例大的时候, T4 模板比 T3 模板的 F 值高, 说明在 T4 模板中位置特征对产品属性抽取有效, 反之, 就近匹配比例变小的时候, T4 模板比 T3 模板的 F 值略低, 说明在 T4 模板中位置特征对产品属性抽取起到了负面作用. 这个结果说明了在产品评论中, 短评论在产品评论所占比例越高, 评价词和评价对象的就近匹配比例就会越高, 在抽取产品属性的时候, 同时加入位置特征和浅层句法特征, 系统抽取的效果就越好. 实际产品评论中短评论的比例一般都高于本文测试语料中的比例, 所以在系统模型中同时引入这两个特征会更有效.

为了验证引入的特征带来的系统性能提升不依赖于模型本身, 我们采用最大熵模型作为对比模型, 在相同的数据集上用 6 折交叉验证的方式进行了实验, 实验结果见表 5.

表 5 最大熵模型平均值

Table 5 Average score of maximum entropy model

	精确评价			覆盖评价		
特征模板	精确率	召回率	F 值	精确率	召回率	F 值
T1	0.3667	0.3723	0.3694	0.5960	0.6052	0.6004
T2	0.3798	0.3755	0.3776	0.6165	0.6132	0.6148
T3	0.3849	0.3813	0.3830	0.6315	0.6254	0.6282
T4	0.3993	0.3813	0.3900	0.6430	0.6337	0.6383

从表 5 的实验结果可以看到, 引入 T2 特征模板和 T3 特征模板后, 系统的各项指标均有所提升, T4 模板与其他特征模板相比其实验效果仍是最好, 进一步验证了本文提出的浅层句法特征和位置特征的有效性. 同时, 对比最大熵模型和 CRFs 模型平均值 (表 4 的平均值) 中的各个指标可以发现, 同样的特征模板条件下 CRFs 模型的性能明显优于最大熵模型的性能, 精确评价方式 F 值提升了 14 %, 覆盖评价方式 F 值提升了约 5 %, 说明 CRFs 模型更适合评价对象抽取的研究.

最后, 我们将系统中加入 T4 特征模板的抽取结果与 2008 年 COAE 的任务 3 评测结果的最大值进行比较. 对于 3 个电子产品领域, 将任意 2 个产品评论语料做训练集, 另外 1 个做测试集, 分别得到 3 个产品领域的抽取结果. 对于汽车评论语料, 采用电子产品领域语料做训练集, 汽车领域语料做测试集, 获得汽车领域的抽取结果. 将 4 个产品属性的抽取结果取平均值与 2008 年 COAE 各个参评系统 4 个测试集的平均值中的最大值比较, 实验结果如表 6 所示.

从表 6 的实验结果可以看到, 本文的抽取结果在两种评价方式下 F 值均超过 2008 年的 F 测度最大值, 进一步证明本文提出方法的有效性. 同时从结果中我们也发现, 两种评价方式中的精确率较高, 但召回率偏低. 我们分析, 一个原因是由于汽车语料比较复杂, 导致测试结果的召回率偏低而影响了平均召回率; 另一个原因是由于系统在训练过程中没有引入人工构建的领域词典. 由于人工构建词典费时费力, 不利于领域移植, 所以本文的研究目的是探索更有效地通用特征来降低系统对领域知识的依赖, 而目前引入的相关特征还不能有效地改进召回率, 今后我们将寻找更有效的特征来解决这个问题. 另外, 实验结果中的覆盖评价方式比精确评价方式的各项指标都高, 证明系统对产品属性的定位比较准确, 但边界识别错误较多, 其原因主要是由于分词和浅层句法分析错误带来的问题. 如何避免分词和浅层句法分析的错误对系统结果的影响也是我们今后改进的主要方向.

表 6 COAE 2008 数据集实验比较结果

Table 6 Comparison of results on COAE 2008 data

	精确评价			覆盖评价		
测试集	精确率	召回率	F 值	精确率	召回率	F 值
数码相机	50.97	35.79	42.06	72.95	51.22	60.19
手机	56.79	36.31	44.30	77.61	49.62	60.54
笔记本	58.43	42.00	48.87	76.92	55.29	64.33
汽车	43.02	20.60	27.86	62.65	30.01	40.58
平均值	52.30	33.68	40.77	72.53	46.54	56.41
COAE_F 测度最大值	37.98	41.72	39.76	46.70	57.88	51.69

5 结论

本文面向电子产品领域和汽车领域, 采用条件随机域模型, 引入多种有效地语言学特征和启发式位置特征, 提高了产品属性的抽取效果. 实验结果表明, 分别引入浅层句法特征和位置特征后, 系统抽取效果均明显提升, 同时加入浅层句法特征和位置特征后, 系统的提升效果会受到语料中的评价对象和最近的评价词匹配比例的影响, 对就近匹配比例高的产品评论作用更显著. 同时本文提出的方法不依赖领域词典知识, 可以有效地应用到其他领域的产品属性抽取上.

References

1 Zhang L, Liu B, Lim S H, O′Brien-Strain E. Extracting and ranking product features in opinion documents. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters. Stroudsburg, USA: Association for Computational Linguistics, 2010. 1462−1470

2 Wan X J. Co-training for cross-lingual sentiment classifica-

tion. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language. Stroudsburg, USA: Association for Computational Linguistics, 2009. 235−243

3 Zhao Yan-Yan, Qin Bing, Liu Ting. Integrating intra- and inter-document evidences for improving sentence sentiment classification. *Acta Automatica Sinica*, 2010, **36**(10): 1417−1425

4 Pang B, Lillian L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2004. 271−278

5 Swati T, Barbara D E. Generating fine-grained reviews of songs from album reviews. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2010. 1376−1385

6 Niklas J, Iryna G. Using anaphora resolution to improve opinion target identification in movie reviews. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, USA: Association for Computational Linguistics, 2010. 263−268

7 Zhao Yan-Yan, Qin Bing, Liu Ting. Sentiment analysis. *Journal of Software*, 2010, **21**(8): 1834−1848
(赵妍妍, 秦兵, 刘挺. 文本情感分析. 软件学报, 2010, **21**(8): 1834−1848)

8 Liu B, Hu M Q, Cheng J S. Opinion observer: analyzing and comparing opinions on the Web. In: Proceedings of the 14th International Conference on World Wide Web. Chiba, Japan: ACM, 2005. 342−351

9 Popescu A M, Etzioni O. Extracting product features and opinions from reviews. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Stroudsburg, USA: Association for Computational Linguistics, 2005. 339−346

10 Yao Tian-Fang, Nie Qing-Yang, Li Jian-Chao, Li Lin-Lin, Lou De-Cheng, Chen Ke, Fu Yu. An opinion mining system for Chinese automobile reviews. In: Proceedings of the 25th Anniversary Chinese Information Processing Society of China. Beijing, China: Tsinghua University Press, 2006. 260−281
(姚天昉, 聂青阳, 李建超, 李林琳, 娄德成, 陈珂, 付宇. 一个用于汉语汽车评论的意见挖掘系统. 中国中文信息学会二十五周年学术会议. 北京, 中国: 清华大学出版社, 2006. 260−281)

11 Zhang S, Jia W J, Xia Y J, Meng Y, Yu H. Opinion analysis of product reviews. In: Proceedings of the 6th International Conference on Fuzzy Systems and Knowledge Discovery. Tianjin, China: IEEE, 2009. 591−595

12 Liu Hong-Yu, Zhao Yan-Yan, Qin Bing, Liu Ting. Comment target extraction and sentiment classification. *Journal of Chinese Information Processing*, 2010, **24**(1): 84−88
(刘鸿宇, 赵妍妍, 秦兵, 刘挺. 评价对象抽取及其倾向性分析. 中文信息学报, 2010, **24**(1): 84−88)

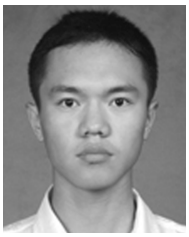
13 Lafferty J D, McCallum A, Pereira F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference Machine Learning. Williamstown, Australia: Morgan Kaufmann, 2001. 282−289

14 Zhu C H, Zhao T J, Yu M O. Chinese word segmenter based on discriminative classifiers integration. *Journal of Information and Computational Science*, 2009, **6**(3): 1155−1162

15 Zhao Jun, Xu Hong-Bo, Huang Xuan-Jing, Tan Song-Bo, Liu Kang, Zhang Qi. Overview of Chinese opinion analysis evaluation 2008. In: Proceedings of the 1st Chinese Opinion Analysis Evaluation. Beijing, China: The Professional Committee of Information Retrieval, 2008. 1−20
(赵军, 许洪波, 黄萱菁, 谭松波, 刘康, 张奇. 中文倾向性分析评测技术报告. 第一届中文倾向性分析评测研讨会. 北京, 中国: 信息检索专业委员会, 2008. 1−20)

徐 冰 博士研究生. 主要研究方向为自然语言处理和情感分析. 本文通信作者. E-mail: xb@mtlab.hit.edu.cn
(**XU Bing** Ph. D. candidate at Harbin Institute of Technology. Her research interest covers natural language processing and sentiment analysis. Corresponding author of this paper.)

赵铁军 教授. 主要研究方向为自然语言处理和人工智能应用.
E-mail: tjzhao@mtlab.hit.edu.cn
(**ZHAO Tie-Jun** Professor at Harbin Institute of Technology. His research interest covers natural language processing and artificial intelligence application.)



王山雨 硕士研究生. 主要研究方向为自然语言处理和意见挖掘.
E-mail: sywang@mtlab.hit.edu.cn
(**WANG Shan-Yu** Master student at Harbin Institute of Technology. His research interest covers natural language processing and opinion mining.)

郑德权 博士后, 副教授. 主要研究方向为自然语言处理、信息抽取和机器学习.
E-mail: dqzheng@mtlab.hit.edu.cn
(**ZHENG De-Quan** Postdoctor and associate professor at Harbin Institute of Technology. His research interest covers natural language processing, information extraction, and machine learning.)