

基于 CRFs 的评价对象抽取特征研究*

王荣洋, 鞠久朋, 李寿山, 周国栋

苏州大学 计算机科学与技术学院 自然语言处理实验室, 江苏 苏州 215006

E-mail: wangrongyang.nlp@gmail.com

摘要: 评价对象是情感分析中情感信息的一个重要组成部分。本文基于条件随机场模型, 研究多种特征在评价对象抽取任务中的表现, 并将特征归纳为词法、依存关系、相对位置、语义等四大类别。其中, 重点引入语义角色标注新特征。在实验中, 我们在三个不同的数据集上考查了各个特征及其组合对系统性能的影响, 作了详细的比较研究。另外, 实验结果表明新提出的语义角色标注特征对评价对象抽取有很好的指示作用。

关键词: 情感分析; 评价对象抽取; 条件随机场; 特征组合; 语义角色标注

Feature Engineering for CRFs Based Opinion Target Extraction

Wang Rong-yang, Ju Jiu-peng, Li Shou-shan, Zhou Guo-dong

NLP Lab, School of Computer Science and Technology, Soochow University, Suzhou 215006

E-mail: wangrongyang.nlp@gmail.com

Abstract: Opinion target is an important component of sentiment information in sentiment analysis. This paper explores Conditional Random Fields (CRFs) based opinion target extraction. After employing frequently used features in sentiment extraction, we summarize all the features into four categories, i.e. lexical, dependency, relative-position and semantic. More importantly, we propose using semantic role as a specific feature. Great efforts and detailed comparative studies have been made to evaluate the performance by exploring various features and their combination. Experimental results show that semantic role is a good indicator for opinion target.

Keywords: sentiment analysis; opinion target extraction; conditional random fields; combination of features; semantic role labeling

1 引言

随着 Web 2.0 技术的迅猛发展, 互联网上带有丰富意见的资源(如评论、博客等)正以指数级的速度增长。基于用户发布内容的情感分析已经成为自然语言处理领域中的一个研究热点。情感分析又称意见挖掘, 是指通过计算机手段, 帮助用户快速获取、整理和分析相关评价信息。近年来, 情感分析研究已经取得一系列进展, 特别是在情感信息的褒、贬极性分类方面^{[1][2]}。目前, 越来越多的研究者开始把目光转向细粒度的^[3](Fine-grained)情感分析, 如评价对象抽取。

评价对象(Opinion Target)是指某段评论中所讨论的主题, 具体表现为评论文本中评价词语所修饰的对象。评价对象抽取是情感信息抽取任务重要的研究课题之一。而且这项研究的开展有助于为上层情感分析任务提供服务。评价对象抽取任务可以被建模成信息抽取。由于条件随机场^[4]序列标注模型能较好地捕捉上下文信息, 它已经被成功应用于多个任务中, 在评价对象抽取中也得到了很好的应用^[5]。

本文基于条件随机场模型, 研究多种特征在评价对象抽取任务中的表现, 具体内容组织如下: 第二部分介绍相关研究; 第三部分描述基于 CRFs 的系统构建、详细介绍特征; 第四部分给出相关实验结果及分析; 最后总结全文。

* 本研究受国家自然科学基金(61003155), 国家自然科学基金(60873150)和模式识别国家重点实验室开发课题基金资助

2 相关研究

初期, 评价对象抽取的研究一般是基于规则/模板的方法。规则的制定通常要借助一系列自然语言理解预处理模块, 如词性标注、命名实体识别、句法分析、语义分析等。相应地, 制定的规则也包括词序列规则、词性规则、句法规则以及语义规则等形式。Kim^[3]等通过将语义角色映射到制定的模板框架, 从在线的新闻文本中抽取观点、观点持有者、以及主题。另外, 有些学者^{[6][7]}使用关联规则挖掘的方法或是基于句法分析的结果找出频繁出现的候选评价对象, 继而使用两种剪枝方法去除错误样例。

随着话题模型的逐渐兴起, 很多学者将其应用到情感分析领域。由于评价对象是蕴涵于情感文本中的某些话题, 因此可以使用话题模型来评价对象的识别。有学者^{[8][9]}采用多粒度的话题模型挖掘产品领域情感文本中的评价对象, 并将相似的评价对象进行聚类。这种方法理论上能够提高评价对象抽取的召回率。由于这是一种无指导的学习方法, 不需要人工标注的语料, 也不存在领域适应问题, 因此有着广阔的研究前景。但遗憾的是, 还没有实验将这种方法与上述传统的基于名词短语的方法进行对比。

近年来, 随着 CRFs 在中文分词、词性标注、命名实体识别等自然语言处理任务取得的进展, 特别是在情感分析领域主客观分类、观点持有者识别等任务上的成功应用, 有研究者^[3]将它应用在评价对象抽取的研究上。

3 系统构建

3.1 特征概述

本文中, 我们采用基于 CRFs 的评价对象抽取系统, 同时采用 Jakob^[5]和 Lu^[10]所用的特征, 并引入新的 SRL(语义角色标注)特征, 并在实验中组合这些特征。我们将所有特征归纳为四类, 表 1 给出这些特征的简单说明。

表 1 特征概述

词法特征	词 tk	该特征表示当前单词的字符串特征
	词性 pos	该特征表示当前单词的词性标记特征
依存关系特征	最短路径 dLn	布尔型特征, 当前词与情感词是否有直接的依存关系, 有直接的依存关系记为 1, 否则为 0
	依存关系 pRel	当前词与其父亲词之间的依存关系特征
	父亲词 pWd	当前词在依存关系中的父亲词本身特征, 用父亲词本身来表示
	父亲词词 pPos	当前词在依存关系中的父亲词的词性标记
相对位置特征	词距离 wDs	布尔型特征, 该特征指示句子中与情感词距离最近的名词/名词短语/代词
语义特征	情感词 isOE	布尔型特征, 如果当前词是情感词特征为 1, 否则为 0
	语义角色 srl	语义角色标注是浅层语义分析的一种简洁化形式, 其任务是识别出事件的施事者、受事者等角色。在 4.3 节具体讨论。

在实验过程中, 我们将词、词性特征表述为基本词法特征, 将其余特征(除语义角色特征)统称为扩展特征。

3.2 SRL 特征提取算法

从情感表达的角度, 情感词主要分为两类, 一类是形容词, 另一类是动词, 且主要为形容词。

对于形容词情感词, Arg0 (施事者) 极有可能是评价对象, 如 “While none of the features are earth-shattering”, 这里 “feature/earth-shattering” 构成了一个评价对象/情感词的搭配。对于动词情感词, 如例句 “I like the new movie I saw last night!” 中的 “like”, 其角色 Arg1 (受事者) 的中心词就是评价对象。但其中包含一个定语从句, 从句中谓语动词 “saw” 的角色 Arg1 也能提供评价对象信息。

因此, 区别对待谓词类别以及情感词词性对正确判断起到重要作用。我们还发现, 对于存在从句成分的句子, 该从句就有可能是某个谓词的角色, 而从句中通常也含有谓词, 但评价对象通常是从句中谓词的某个角色, 因此消除主句与从句中语义角色的嵌套关系也显得很重要。通过语料可以发现评价对象通常不会太长, 就算有修饰成分一般也不会超过某长度, 因此过长的角色 (>5 个词) 我们不予标注。另外, 考虑到评价对象与情感词通常相邻, 所以我们对长句作了切分, 即将长句用切分成一个个短句来分别处理。鉴于以上分析, 我们设计了如下的算法, 算法描述如图 1 所示:

Algorithm 语义角色特征提取算法

Require: 句子中词、词性、是否是情感词的信息表 **T**

Require: 语义角色标注结果表 **SRL**

Output: 语义角色特征列表 **L**

Procedure

将 **SRL** 表归并成一列: 消除谓词角色间嵌套结构, 保留最小语义角色单元的 Arg0、Arg1、Arg2 三种信息, 删除长 srl 标注结果 (>5 个词), 将 srl 用 **l*B** 的形式表示。用短句切分子程序将句子切分成若干子句。

Foreeach 子句

if 子句中 含有情感词

特征为 语义角色标记+情感词词性

else

特征为 语义角色标记

为子句中的每个词增加对应的特征 (非语义角色置特征*)

Repeat

End

图 1 SRL 特征提取算法

4 实验结果与分析

4.1 语料概述

本文用到的语料有两个来源: DSRC¹ 及 IMDb²。DSRC 语料包含 services 和 universities 两个领域, 文献^[11]给出了 DSRC 详细的标注规范及说明; IMDb 是 movies 领域评论的评价对象/评价词标注^[12]。各领域的原始语料规模统计如表 2 所示。

¹ <http://www.ukp.tu-darmstadt.de/data/sentiment-analysis/>

² <http://www.imdb.com/>

表2 原始语料规模

	services	universities	movies
句子数	7575	2911	13906
观点句的个数	1372	1012	5326
评价对象总数	1827	1143	6574
评价对象平均长度	1.37	1.48	1.56

4.2 实验设置

实验过程中, 我们进行了多组实验。首先对各种特征及其组合进行实验, 其次为了体现 SRL 的作用, 我们将引入 SRL 特征之前与引入 SRL 特征之后的性能分别作了对比。在本实验中, 条件随机场模型的实现采用 CRF++¹, 采用默认参数, 词与词性的窗口大小为 3, 其余特征用当前窗口, 使用标记的 Bigram 特征。

本文采用的 SRL 工具为我实验室开发的基于短语结构句法分析的、采用特征向量的 SRL 系统², 在 CoNLL 2005 SRL shared task 的 WSJ 测试集上获得了 78.75 的 F1。

在数据集上使用 10 倍交叉验证的方法, 最终评价标准采用 10 次的平均。我们采用的评价标准是平均准确率、平均召回率、平均 F1, 分别记作 P、R、F1。

4.3 实验结果与分析

1) 基本词法特征实验结果

表 3 是采用最基本的词法特征模板的实验结果, 记作 CRF_basic。

表3 CRF_basic 的识别结果

编号	特征	services			universities			movies		
		P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
1	tk	74.20	41.15	52.83	73.07	44.90	55.49	73.51	62.76	67.70
2	tk+pos	71.75	46.09	56.03	69.11	47.46	56.19	72.96	65.19	68.85

词形和词性特征组合能取得相对好的性能。其中 services 领域提升幅度最大, 达到 3.2 个点。

2) 单个扩展特征实验结果

在词法特征的基础上, 我们首先单独引入各种特征, 实验结果如表 4 所示, 记作 CRF_ext_uni。

可以看出, 就单个特征而言, dLn 的作用最为显著, 最大提升幅度达 4.94 个点。其次作用明显的是 wDs。但这两个特征都是间接依赖于 isOE, 因为只有先找到情感词, 才能判别某个词是否与情感词直接依存、以及是否与情感词距离最近的名词/名词短语。

表4 CRF_ext_uni 识别结果

编号	特征	services			universities			movies		
		P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
2	tk+pos	71.75	46.09	56.03	69.11	47.46	56.19	72.96	65.19	68.85
3	tk+pos+dLn	72.59	50.07	59.20	73.81	52.30	61.13	74.89	65.23	69.70
4	tk+pos+isOE	71.87	47.12	56.84	69.72	48.02	56.79	73.16	65.27	68.97
5	tk+pos+pPos	71.64	46.56	56.37	69.30	46.86	55.84	73.06	64.95	68.75
6	tk+pos+pRel	71.27	46.59	56.25	69.30	46.96	55.87	73.76	65.27	69.25
7	tk+pos+pWd	70.88	46.44	56.04	69.45	47.39	56.25	73.10	64.93	68.76
8	tk+pos+wDs	71.78	48.09	57.51	72.07	49.71	58.75	74.64	65.23	69.60

¹ <http://crfpp.sourceforge.net/>

² <http://nlp.suda.edu.cn/~jhli/>

3) 多个扩展特征组合实验结果

根据单个特征的效果,我们以四个性能较好的特征为基础,作两两、三三、以及更多的与其他特征的组合,做了对比试验,结果如表5所示,记作CRF_ext_comb。

表5 CRF_ext_comb 识别结果

编号	特征	services			universities			movies		
		P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
2	tk+pos	71.75	46.09	56.03	69.11	47.46	56.19	72.96	65.19	68.85
9	tk+pos+dLn+wDs+pWd	72.97	52.16	60.79	73.21	52.90	61.37	75.40	65.42	70.04
10	tk+pos+dLn+wDs+isOE	72.43	52.58	60.85	74.77	54.43	62.87	75.43	66.01	70.38
11	tk+pos+dLn+pWd+isOE	73.41	53.04	61.50	72.65	54.35	62.07	75.12	65.45	69.94
12	tk+pos+wDs+pWd+isOE	70.86	49.52	58.24	74.51	52.56	61.55	75.03	65.54	69.96
13	tk+pos+dLn+wDs+pWd+isOE	73.39	53.39	61.75	74.58	54.75	63.02	75.41	65.62	70.15
14	tk+pos+dLn+wDs+pWd+isOE+pRel(BEST)	73.23	54.08	62.17	74.27	54.62	62.83	75.81	65.94	70.51
15	所有特征(ALL)	72.58	53.57	61.60	73.60	54.64	62.57	75.68	65.91	70.44

随着特征的累积增加,性能会有显著的提高,其中最好的一组组合特征是编号14,较基本特征相比,提升约6个点。

4) 使用语义角色特征实验结果与分析

以基本特征、最佳系统和所有特征系统为基础,我们引入SRL特征。实验结果如表6所示:

表6 加入SRL特征后的实验结果对比

编号	特征	services			universities			movies		
		P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
2	tk+pos	71.75	46.09	56.03	69.11	47.46	56.19	72.96	65.19	68.85
16	tk+pos+ SRL	71.33	48.12	57.40	71.34	50.40	58.94	73.68	65.76	69.48
17	BEST	73.23	54.08	62.17	74.27	54.62	62.83	75.81	65.94	70.51
18	BEST+SRL	73.79	54.96	62.95	74.52	56.73	64.31	75.92	65.97	70.58
19	ALL	72.58	53.57	61.60	73.60	54.64	62.57	75.68	65.91	70.44
20	ALL+ SRL	73.34	54.57	62.53	74.40	57.21	64.55	75.60	66.05	70.49

比较引入SRL特征前、后三个系统的性能变化,可以得到如下结论:

(1) 在基础系统上,三个领域提升的幅度分别为1.37、2.75和0.63。

(2) 在最佳系统BEST上,三个领域的准确率、召回率都有提高。

(3) 在使用所有特征的系统上,提升幅度为0.93、1.98、0.05。movies领域的提升幅度最不明显,可能是由于其基线系统的性能本身已经比较高。

为直观理解SRL特征的作用,我们分析了测试结果。例句“I like the movie.”中“movie”相对于动词性谓词情感词“like”的角色是受事者Arg1,而“movie”正是评价对象。例句“The movie is wonderful.”中“movie”是谓词“be”的角色Arg0的中心词,形容词性情感词“wonderful”充当Arg2的角色。因此,在引入SRL特征时,区分动词性情感词和形容词性情感词显得尤为重要。实际上,算法的设计正是出于这一点的考虑。总体而言,SRL特征的提出改善了评价对象抽取的性能,SRL信息对评价对象识别有很好的指示作用。

5 结语

本文研究基于CRFs的评价对象抽取系统中各种特征的选择和比较,将用于情感信息抽取的特

征归纳为词法、语法、相对位置、语义等四大类别。通过大量的实验，系统地比较、研究了各类特征及其组合对系统性能的影响。此外，本文提出将 SRL 特征加入系统之中，并提出了引入的算法，实验结果表明，SRL 信息能对评价对象抽取起到很好的指示作用。

统计表明，在 movies 语料中约 9.5% 的评价对象由代词指代，在 DSRC 中这一比例高达 12.5%，因此在下一步的工作中，指代消解在评价对象抽取上的应用显得尤为重要。另外，本文通过将 SRL 作为特征加入系统中，虽然起对到评价对象起到很好的指示作用，但是其作用并未能完全发挥，因此如何将评价对象抽取融入到 SRL 框架也是一个值得研究的问题。

参 考 文 献

- [1] Pang B., Lee L., Vaithyanathan S. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. In: Proceedings of EMNLP-2002. 2002. 79-86.
- [2] Li S., Huang C., and Zong C. Multi-domain Sentiment Classification with Classifier Combination. Journal of Computer Science and Technology (JCST), 26(1): 25-33 Jan. 2011.
- [3] Kim S., Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In: Proceedings of the ACL Workshop on Sentiment and Subjectivity in Text. 2006. 1-8.
- [4] Lafferty J., McCallum A., and Pereira F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. Proceedings of ICML-2001, 2001: 282-289.
- [5] Jakob N., Gurevych I. Extracting Opinion Targets in a Single and Cross-Domain Setting with Conditional Random Fields. In Proceedings of EMNLP-2010. 2010. 1035-1045.
- [6] Hu M., Liu B. Mining Opinion Features in Customer Reviews. In: Proceedings of AAAI-2004. 2004. 755-760.
- [7] 倪茂树,林鸿飞.基于关联规则和极性分析的商品评论挖掘.第三届全国信息检索与内容安全学术会议. 2007: 628-634.
- [8] Titov L., McDonald R. Modeling Online Reviews with Multi-grain Topic Models. In: Proceedings of WWW-2008. 2008. 111-120.
- [9] Lu Y., Zhai C., and Sundaresan N. Rated aspect summarization of short comments. In Proceedings of WWW-2009. 2009. 131-140.
- [10] Lu B. 2010. Identifying Opinion Holders and Targets with Dependency Parser in Chinese News Texts. In Proc. of the NAACL HLT 2010 Student Research Workshop, pages 46-51, Los Angeles, California.
- [11] Toprak C., Jakob N., Gurevych I. Sentence and Expression Level Annotation of Opinions in User-Generated Discourse. In Proceedings of ACL-2010. 2010. 575-584.
- [12] Zhuang L., Jing F., and Zhu X. Movie review mining and summarization. In Proceedings of CIKM-2006. 43-50.