

# 基于卷积树核的中文微博情感要素识别

陈 锋 巢文涵 周 庆 李舟军

(北京航空航天大学 北京 100191)

**摘 要** 情感要素识别是情感分析的关键子任务之一,其目的是识别出文本情感所作用的情感对象。文本情感要素识别属于最细粒度的情感分析,吸引了大量研究者的关注。中文微博由于其语言简短灵活、文本不规范、噪声较大等特点,给中文微博情感分析研究工作带来了新的挑战。目前大部分情感要素识别方法都是基于规则的方法或者基于扁平化特征的统计学习方法,区分噪声的能力不强,性能提升有限。针对中文微博的特点,提出一种基于卷积树核的情感要素识别算法,即首先对句子进行词性标注与依存关系分析,将句子中的名词作为候选情感要素;然后基于两种不同的修剪策略对依存树进行修剪,以获取每个候选情感要素的结构化信息;最后采用卷积树核计算依存树的相似度,并在此基础上识别句子中的情感要素。NLP&CC2012和NLP&CC2013中文微博情感分析评测任务中的实验验证了该方法的性能,其准确率相比于传统方法有显著提升。

**关键词** 情感要素识别,中文微博,卷积树核,依存树修剪

**中图分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.12.028

## Convolution Tree Kernel Based Sentiment Element Recognition Approach for Chinese Microblog

CHEN Feng CHAO Wen-han ZHOU Qing LI Zhou-jun

(Beihang University, Beijing 100191, China)

**Abstract** Sentiment element recognition is one of the key sub-tasks of sentiment analysis, and its goal is to identify the sentiment targets in the text. Sentiment target recognition is identified as the most fine-grained sentiment analysis task, and many researchers have conducted lots of research work on it. Since Chinese Microblog text is short and very flexible, which is often not standardized and contains a lot of noisy information in the text, it brings new challenges to the Chinese Microblog sentiment analysis research. At present, most of sentiment target recognition methods are based on rules or statistical learning methods using flat features, which can not distinguish between noisy information and sentiment targets very well, resulting in the low recognition performance. According to the characteristics of Chinese Microblog, a novel sentiment element recognition approach based on convolution tree kernel was proposed. Firstly, the approach analyses the part of speech (POS) and dependency relationship of the Microblog sentences, and takes the nouns in the sentences as candidate sentiment elements. Secondly, it adopts two different pruning strategies to obtain every candidate's structured information. Finally, convolution tree kernel method is used to calculate the similarity of dependency tree, which is the foundation of sentiment elements recognition. The experiments of NLP & CC2012 and NLP & CC2013 Chinese Microblog sentiment target analysis tasks show the performance of this approach is improved significantly comparing to the baseline.

**Keywords** Sentiment target recognition, Chinese microblog, Convolution tree kernel, Pruning strategy

## 1 引言

随着 Web2.0 的兴起,越来越多的用户习惯于在互联网上发布自己的观点、分享个人生活体验、与朋友交流互动,用户拥有越来越多的平台来表达他们对某件事物的意见和情感倾向。对于某件产品的建议可以帮助生产商完善自己的产品,而对于事件的情感便于研究者和媒体了解民众的娱乐导

向,所以用户对于事物的情感分类就显得尤为重要。

微博作为其中一种主要由用户分享产生内容的信息平台,越来越受到大家关注。许多微博文本有着明显的褒贬含义,体现了用户对某个事物的真实看法。但相比于传统的情感分析,微博的内容过于简短,所含噪声也远大于传统情感分析中所用的文本,对其进行情感分析要困难得多。近年来国际上有关 Twitter 的情感分析研究较为热门。但基于中文微

到稿日期:2014-02-20 返修日期:2014-03-25 本文受国家自然科学基金(61003111,61170189,61370126),高等学校博士学科点专项科研基金(20101102120016)资助。

陈 锋(1989-),男,工程师,主要研究方向为自然语言处理、文本挖掘,E-mail:chenfengbinzimu@sina.com;巢文涵(1979-),男,博士,讲师,主要研究方向为自然语言处理、机器翻译、文本挖掘;周 庆(1991-),男,硕士生,主要研究方向为自然语言处理、文本挖掘;李舟军(1963-),男,教授,博士生导师,主要研究方向为网络与信息安全、数据挖掘与文本挖掘。

博的相关工作相对较少,中文微博中表达情感的方式多种多样,甚至不合乎语法规则,这也使得中文微博的情感分析更为困难。

根据不同粒度的情感表达的文本需求,情感分析也可以分为多种粒度,如篇章级别、句子级别、词语级别等。粒度要求越精确,分析和识别的难度也越大。在用户表达情感的句子中,有时会涉及到多个事物。比如“A,B不同,前者更容易使用”。如果单纯的评价句子的情感为正面,则忽略了用户对B产品的负面评价,所以情感倾向分类需要识别其情感作用的对象,即情感要素。在句子中情感要素是用户情感作用的主要对象,有时也被称为情感对象、评价对象。情感要素有“显式”情感要素和“隐式”情感要素之分。在传统评论语料情感分析中,用户的评价对象往往集中于某个产品,或者某个产品的某个属性,这些对象是往往是显式地体现在文本中的。而微博文本的意图性和主观性很明确,但是评论范围很宽泛,并不限于某一具体的事物。并且微博中存在大量的省略和指代,如果不充分利用句子中的信息,很难自动识别出情感要素。

针对中文微博的特点,本文提出了一种基于卷积树核的情感要素识别算法,其充分利用句子的依存树结构信息,识别情感句中的一个或多个情感对象。该方法由于利用的是句子的结构信息,具有很好的领域适应性,能够区分微博中大量的噪声和情感对象。

本文第2节介绍情感要素识别的相关研究工作;第3节重点介绍基于卷积树核的情感要素识别算法,包括卷积树核计算方法和两种不同的句法树修剪策略算法;第4节给出了算法在NLP&CC2012及NLP&CC2013中文微博情感分析-情感要素识别评测子任务中的实验结果和分析;最后给出总结。

## 2 相关工作

在情感要素识别领域,一些学者进行了有益的尝试。一部分学者使用基于规则的方法。Yi<sup>[1]</sup>等使用层层递进的规则从候选情感要素中抽取真正的情感要素。Kim<sup>[2]</sup>则将句子进行语义角色标注,将语义角色映射到制定的模板中。一些学者也采用了关联规则分析方法,但是这种方法只能作用于频繁出现的评价。

Liu<sup>[3]</sup>使用了标签序列规则(label sequential rule,LSR)来匹配语料中的产品特征,此类识别方法的优点是针对性强并且可以高效地识别结构类似文章中的评价对象,缺点是需人工编写相应的模版,通用型不强。

基于规则的方法虽然可以准确地识别符合规则的模式,但是由于语言使用的多变性,基于规则的方法的召回率很低,要提高召回率需要大量的人力工作来扩展规则。

Popescu<sup>[4]</sup>等人借助搜索引擎对候选的名词短语和情感词进行搜索,获得它们的点击量,从而可以计算二者的互信息来判定二者的关联程度。通过关联程度即可确定情感词是否指向该名词短语。相对于基于规则的方法,这种方法获得了较大的提升。

Jakob<sup>[5]</sup>等则使用条件随机场方法,将情感要素识别转换为序列标注问题。除了情感分类的常用特征外,还加入了词法、短距离依存路径、词距离,以及是否为观点句等特征。然而对于不同领域,使用的情感表达并不相同,所以词法特征在不同领域是不同的。由此导致CRF方法在适应性方面交叉,

难以适应不同的领域。

鉴于以上方法面临的问题,同时针对中文微博的特点,提出的基于卷积树核方法充分考虑了句子的结构信息,同时尽量对句法树进行修剪,以尽量规避微博句子中的噪声对情感要素识别的影响,从而显著提高了识别的性能。

## 3 基于卷积树核的情感要素识别算法

### 3.1 卷积树核

卷积核(convolution kernel)是一种通过递归方式计算大结构相似性的核函数:其首先通过类似卷积的操作,将较大的结构分解成子结构;然后分别计算子结构之间的匹配度;最后对各子结构匹配的结果求和,从而计算出大结构之间的相似性。Haussler<sup>[6]</sup>和Watkins<sup>[7]</sup>都已经证明,这一计算过程满足核函数成立的对称以及半正定条件,因此,以这种方式构造的相似函数是一个核函数,称为卷积核函数。

卷积树核函数(convolution tree kernel)是卷积核函数的一个特例,其通过递归计算两个树结构的子树的相似度来确定两棵完整的树的相似度。

卷积树核在构建语法结构中起到了很重要的作用。一些研究者<sup>[9,18]</sup>应用卷积树核进行语法分析、语义分析,获得了很高的性能。卷积树核的优势在于可以生成大量的语法结构特征供学习算法选择。

给定两棵句法树 $T_1, T_2$ ,基于卷积树核计算二者相似度的典型方法为<sup>[23]</sup>:

$$K_{CTK}(T_1, T_2) = \sum_{n_1 \in N_1, n_2 \in N_2} \Delta(n_1, n_2)$$

其中, $K_{CTK}(T_1, T_2)$ 表示两者的卷积树核相似度, $N_1, N_2$ 分别代表 $T_1, T_2$ 的节点集合, $\Delta(n_1, n_2)$ 是计算以 $n_1, n_2$ 为根节点的两棵子树的相似度。常见算法如下:

1. 如果 $n_1$ 和 $n_2$ 产生式(采用上下文无关文法)不同,则 $\Delta(n_1, n_2) = 0$ ;否则转向步骤2;
2. 如果 $n_1$ 和 $n_2$ 都是叶子节点之前的一个节点,则 $\Delta(n_1, n_2) = 1 \times \lambda$ ,否则转向步骤3;
3. 递归计算:

$$\Delta(n_1, n_2) = \lambda \prod_{k=1}^{\#ch(n_1)} (1 + \Delta(ch(n_1, k), ch(n_2, k)))$$

其中, $\#ch(n)$ 是节点 $n$ 的子节点数目, $ch(n, k)$ 是节点 $n$ 的第 $k$ 个子节点,而 $\lambda(0 < \lambda < 1)$ 则是衰退因子,用来防止子树的相似度过度依赖于子树的大小。

典型的卷积树核方法对句法树中的节点不做区分,然而,中文微博文本中的噪声较大,这将对基于卷积树核方法识别情感要素造成很大影响。因此,本文将对其句法树进行修剪,以适应中文微博的特点。

### 3.2 句法树的修剪

为了识别情感要素,主要关注的是将句子中的候选情感要素用包含丰富信息的依存树表示出来,并显著区别于句子中的其他候选情感要素。如果在依存树上保留过多的信息,则不仅会引入大量噪声,而且会增加同一句子中不同候选情感要素之间的依存树表示的相似度,降低识别准确性。

因此,为了降低噪声和提高候选情感要素之间的区分度,算法将对依存树进行修剪。

#### 3.2.1 基于句子谓语的修剪策略

本文考虑了两种不同的依存树修剪方式。一种是基于句子谓语的修剪策略,它仅仅保留候选情感要素与句子谓语之间的依存路径。

根据依存句法理论,“谓语”中的动词是一个句子的中心,其他成分与该词直接或间接地产生关联。因此句子中的不同的候选情感要素与谓语之间的不同依存路径,体现了候选词在句子中的不同作用。例如给定句子“媒体与公益的结合将使社会充满爱与幸福”,其依存树如图 1 所示。

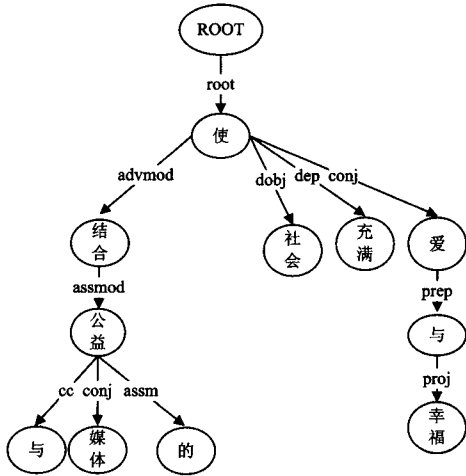


图 1 “媒体与公益的结合将使社会充满爱与幸福”依存树

句子中的谓语是“使”,名词“公益”、“媒体”、“社会”、“爱”和“幸福”都可以作为候选情感要素。它们与“使”的依存关系分别为“公益-assmod-结合-advmod 使”,“媒体-conj-公益-assmod-结合-advmod 使”,而候选词“公益”、“媒体”和“幸福”与谓语没有直接的关联。其中,“assmod”、“advmod”、“conj”等分别表示两个节点之间的依存关系类型。

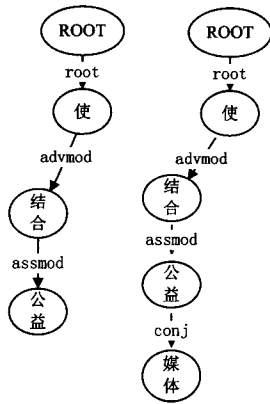


图 2 “公益”和“媒体”与“使”的依存树

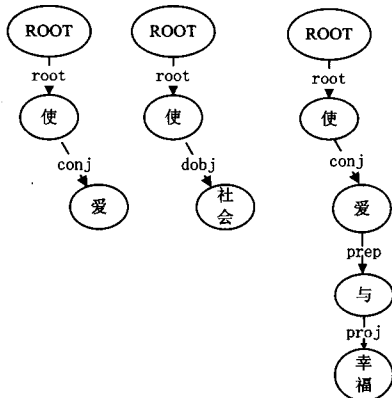


图 3 “爱”、“社会”和“幸福”与“使”的依存树

基于谓语的修剪方法过程如下:

1. 给定句子依存树  $T$  和候选情感要素词  $W$ ;

2. 寻找依存树中的谓语节点  $P$ ;

3. 在句子依存树中寻找候选词节点与谓语节点  $P$  之间最近公共祖先节点(LCA)得到最短依存路径 SPP(Shortest-Path, 记为 SPP);

4. 仅保留 SPP 上的所有节点,它们形成了候选词与谓语的依存关系。即,仅仅保留候选词到谓语之间最短依存路径上的所有节点,去掉整个句子依存树  $T$  上的其他无关节点,从而实现对依存树的修剪。

对于上述例句,经过基于谓语的修剪之后,得到的依存树如图 2 和图 3 所示。

### 3.2.2 基于句子情感词的修剪策略

根据以往的情感分析研究,情感词对于主观性文本具有重要的指示作用。因此,尝试保留了候选词与其最近的情感词之间的依存路径。

与基于谓语的修剪方法类似,基于情感词的修剪过程如下:

1. 给定句子依存树  $T$  和候选情感要素词  $W$ ;

2. 寻找候选词周围的情感词节点  $\{S\}$ ;

3. 在句子依存树中寻找候选词节点与每个情感词节点  $S$  之间的最近公共祖先节点(LCA)得到最短依存路径 SPS (ShortestPath, 记为 SPS);

4. 仅保留 SPS 上的所有节点,它们形成了候选词  $W$  与情感词  $S$  之间的依存关系。即,仅仅保留候选词到情感词之间最短依存路径上的所有节点,去掉整个句子依存树  $T$  上的其他无关节点,从而实现对依存树的修剪。

例如,句子“主裁判成为比赛主角,令人遗憾”(依存树如图 4 所示)中仅有一个情感词“遗憾”,而候选词有“主裁判度”、“主角”、“比赛”、“人”。

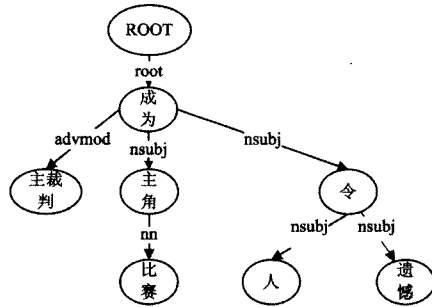
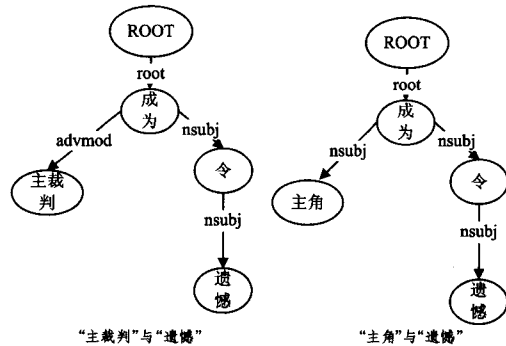


图 4 “主裁判成为比赛主角,令人遗憾”的依存树

基于情感词“遗憾”,保留各个候选词到该情感词的路径,可以分别针对候选情感词对依存树进行修剪,结果如图 5、图 6 所示。



“主裁判”与“遗憾”

“主角”与“遗憾”

图 5 “主裁判”和“主角”与“遗憾”的依存树

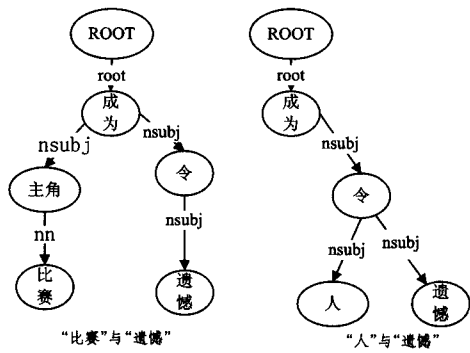


图6 “比赛”和“人”与“遗憾”的依存树

### 3.2.3 不同的依存树路径

根据修剪后的依存树是否包含依存分析中的词和依存关系,依存树分别表现为3种形态:

1. 仅包含依存关系(记为 dep,如图7(a)所示);
2. 仅包含词(记为 word,如图7(b)所示);
3. 同时包含依存关系和词(记为 dep+word,如图7(c)所示)。

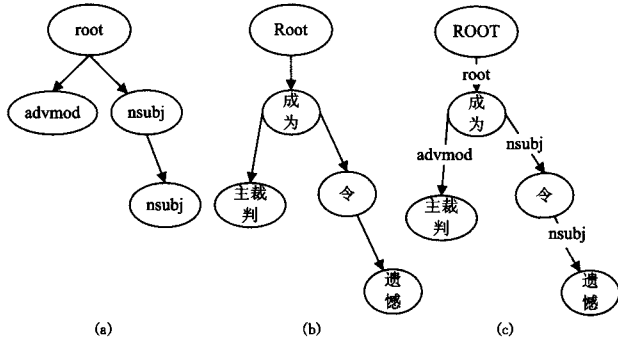


图7 依存树的3种形态图

显然,dep+word的方式包含了最多的信息,word方式次之,dep方式最少。word方式中的每个节点都是一个词,可能存在泛化能力较差的问题;而dep方式中的信息量过少,则可能会导致无法很好地容纳足够信息用以识别情感要素。后续实验中的测试结果也证明了这一点。

### 3.2.4 基于同义词的依存树扩展

如果依存树中包含词,则泛化能力会降低。因此本文使用了哈工大同义词词林(扩展版)<sup>1)</sup>进行依存树扩展。对句子中的每个词随机从同义词词典中挑选若干个词进行替换。

假设实例A生成的树为 $A_1$ ,其扩展出来的树为 $A_2, A_3, \dots, A_N$ 。实例B生成的树为 $B_1$ ,其扩展出来的树为 $B_2, B_3, \dots, B_M$ 。则核函数计算方式为:

$$K_{CTK}(A, B) = \sum_{i=1}^N \sum_{j=1}^M K_{CTK}(A_i, B_j)$$

最终,基于卷积核的情感要素识别算法分为4部分,如图8所示。

- 1) 对句子进行词性标注,将名词作为候选的情感要素,同时通过情感词词典找到句子中的情感词;
- 2) 对句子进行依存句法分析获得依存句法树;
- 3) 对依存句法树进行修剪,并将修剪后的树作为一个离

散结构,用以计算卷积核。

### 4) 基于卷积核对候选情感要素进行识别。

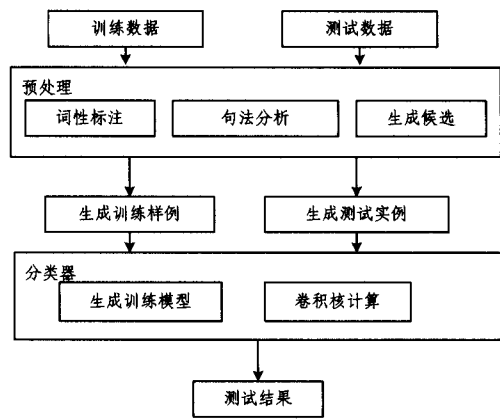


图8 流程图

## 4 实验

### 4.1 实验数据及评价方法

本文使用NLP&CC 2012以及NLP&CC 2013中文微博情感要素识别评测数据作为训练数据和测试数据。该数据集包含32个话题,3839个观点句,共4604个情感要素,平均每个句子有1.19个情感要素。实验中本文选择每个话题约70%的句子做训练,其余用来测试。情感词是情感极性判定中较为重要的考量依据。本文选用NTUSD<sup>2)</sup>、HowNet<sup>3)</sup>以及清华大学自然语言处理组情感词典的并集来构建情感词典。

参照NLP&CC评测标准,本文采用精确评价(Strict)和宽松评价(Lenient)两种评价方式,以及准确率(Precision)、召回率(Recall)、F值(F-measure)3个评价标准对实验结果进行评价。精确评价中,要求提交的情感要素和答案完全相同。其计算公式如下:

$$Precision = \frac{\# system\_correct}{\# system\_proposed}$$

$$Recall = \frac{\# system\_correct}{\# gold}$$

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

# gold是人工标注结果中情感对象的数目,# system\_correct是提交结果中与人工标注匹配的数目,# system\_proposed是提交的情感对象的数目。

宽松评价中,情感对象由起始位置和终止位置构成,首先定义两个结果之间的覆盖率c:

$$c(r, r') = \frac{|s \cap s'|}{|s|}$$

其中,s和s'为两个情感对象r和r'的区间,|\*|表示区间长度。

两个结果集合R和R'之间的覆盖率C定义为:

$$C(R, R') = \sum_{r_i \in R} \sum_{r'_j \in R'} c(r_i, r'_j)$$

<sup>1)</sup> <http://www.ltp-cloud.com/download>

<sup>2)</sup> <http://nlg18.csie.ntu.edu.tw:8080/lwku/index.html>

<sup>3)</sup> <http://nlp.csai.tsinghua.edu.cn/site2>

假设提交的结果集合为  $R'$ , 标注结果集合为  $R$ , 则准确率、召回率和  $F$  值定义为:

$$Precision = \frac{C(R, R')}{\# \text{ system\_proposed}}$$

$$Recall = \frac{C(R, R')}{\# \text{ gold}}$$

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

由于本文算法需要使用词性标注和依存句法分析, 因此本文首先使用 NLPiR 汉语分词系统工具进行词性标注, 找到候选情感对象, 再使用 stanford parser 生成依存句法树, 用修剪后的依存树表示每个候选对象。训练数据中提供了每个情感要素的标注结果, 当候选词  $s'$  与某个标注结果  $s$  有  $s \cap s' \neq \emptyset$  时, 将此候选结果置为正例, 否则置为负例。

#### 4.2 Baseline 实验

为比较本文提出方法的效果, 实验中采用已有的一种基于支持向量机平面特征的基线方法。参考 Michael<sup>[12]</sup>, 本文中手动抽取的特征有: (1) 候选词与情感词存在的依存关系; (2) 候选词左右窗口是否存在情感词; (3) 候选词左右词性标注; (4) 候选词左右窗口内是否存在程度副词和否定副词; (5) 评价对象与最近的情感词的距离; (6) 候选词与最近的谓词之间的词、标点以及词性标注等。

#### 4.3 实验结果与分析

使用 SVM-Light-TK<sup>[9,10]</sup> 进行训练, 得到表 1、表 2 列出的结果。

表 1 基于句子谓词修剪的实验结果

	宽松评价			严格评价		
	准确率	召回率	F 值	准确率	召回率	F 值
BaseLine	40.59	37.83	39.16	13.76	12.83	13.28
dep	48.47	41.14	44.51	15.76	13.38	14.47
word	<b>64.30</b>	35.29	45.57	20.69	11.35	14.66
dep+word	63.90	<b>63.34</b>	<b>63.62</b>	20.57	<b>20.39</b>	20.48
dep+word+同义词	59.88	62.69	62.25	20.37	21.33	<b>20.84</b>

表 2 基于情感词修剪的实验结果

	宽松评价			严格评价		
	准确率	召回率	F 值	准确率	召回率	F 值
Baseline	40.59	37.83	39.16	13.76	12.83	13.28
dep	41.12	28.63	33.76	14.85	10.34	12.19
word	54.55	30.80	39.37	<b>20.87</b>	11.79	15.06
dep+word	53.75	44.54	48.71	19.20	15.91	17.40
dep+word+同义词	53.28	45.19	48.90	19.27	16.34	17.68

其中表 1 是基于句子谓词修剪方法的实验结果。表 2 是基于情感词修剪的实验结果。表 1 和表 2 中的“dep+word+同义词”是做同义词替换的比较实验结果。

从以上两个表可以看出, 相对于基于常规特征的基准系统, 基于卷积树核的情感要素识别方法的性能在整体上有显著提高(宽松评价时,  $F$  值最大提高了 24.46; 严格评价时,  $F$  值最大提高了 7.56), 主要原因在于卷积树核方法考虑了句子中的结构信息, 同时通过修剪句法树, 规避了噪声的影响。

通过对比表 1 和表 2 可以发现: 对于基于句子谓词和基于情感词的修剪策略而言, 前者的效果(宽松评价  $F$  值最高为 63.62, 严格评价  $F$  值最高为 20.84)要远好于后者(宽松评价  $F$  值最高为 48.90, 严格评价  $F$  值最高为 17.68)。这是因为: 相对而言, 仅保留候选情感对象与其最近的情感词之间的

路径的策略比较粗糙, 同时由于在微博文本中用语不规范, 情感表达往往隐含在句子语义中而不是直接通过情感词体现的特点, 从而进一步降低了后者的识别性能。

不同的依存树形态同样对情感要素识别也有较大影响: 仅保留依存信息(dep)效果最差, 仅保留词信息(word)效果次之, 同时保留依存信息和词信息(dep+word)效果最好。这是因为 dep+word 方法拉长了依存路径, 提高了泛化能力, 降低了仅仅考虑词而过于具体所造成的影响。

为了进一步提高泛化能力, 在 dep+word 基础上, 对 word 进行了同义词扩展(dep+word+同义词), 实验结果表明依存树的泛化能力得到了加强, 得到了在严格测试标准下最高的  $F$  值(20.84)。

**结束语** 本文提出了一种基于卷积树核的情感要素识别方法, 该算法在对依存树进行修剪后, 使用依存树中的路径识别句子中的情感要素。实验分析了不同修剪策略和依存树形态对情感要素识别的影响。在 NLP&CC 2012 和 NLP&CC 2013 的情感要素评测数据上的实验表明, 基于句子谓词修剪的卷积树核方法能显著提高情感要素识别的性能。在今后的工作中, 一方面将尝试不同的依存树调整策略; 另一方面, 当前修剪后的依存树泛化能力仍然不强, 尤其在依存树中只保留词信息时, 召回率明显偏低。这在后期的工作中仍将会进一步改进。

#### 参考文献

- [1] Jeonghee Y, Nasukawa T, Bunescu R, et al. Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques [C] // Third IEEE International Conference on Data Mining, ICDM 2003, 2003
- [2] Kim S, Hovy E. Automatic detection of opinion bearing words and sentences [C] // Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP), 2005
- [3] Hu M, Liu B. Mining and summarizing customer reviews [C] // Seattle, WA, USA: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004
- [4] Popescu A, Etzioni O. Extracting product features and opinions from reviews [C] // Vancouver, British Columbia, Canada: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005
- [5] Jakob N, Gurevych I. Extracting opinion targets in a single- and cross-domain setting with conditional random fields [C] // Cambridge, Massachusetts: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2010
- [6] Haussler D. Convolution kernels on discrete structures [R]. Department of Computer Science, University of California at Santa Cruz, 1999
- [7] Watkins C. Dynamic alignment kernels [OL]. <http://www.kernel-machines.org/publications/wathinsonoo>
- [8] Collins M, Duffy N. Convolution kernels for natural language [C] // Advances in Neural Information Processing Systems, 2001

(下转第 142 页)

统,以为用户辨别评价信息的可信性进而选择可信商品提供辅助决策。

## 参 考 文 献

- [1] 阳振坤,张清,王勇,等. 大数据的魔力[J]. 中国计算机学会通讯,2012,8(6):17-21
- [2] Epinions[OL]. <http://www.epinions.com>
- [3] Yuan W, Guan D, Lee Y K, et al. The small-world trust network [J]. Applied Intelligence, 2011, 35(3): 399-410
- [4] Guha R, Kumar R, Raghavan P, et al. Propagation of trust and distrust[C]// Proceedings of the 13th International Conference on World Wide Web. New York, USA, 2004: 403-412
- [5] Liu X. Towards Context-Aware Social Recommendation via Trust Networks[M]// Web Information Systems Engineering-WISE 2013. Springer Berlin Heidelberg, 2013: 121-134
- [6] Ma Y, Lu H W, Gan Z B. Trust Inference Path Search with Minimum Uncertainty for E-commerce [C]// Proceedings of the 10th Web Information Systems and Applications Conference. Yangzhou, China, 2013: 133-137
- [7] Tang J, Gao H, Hu X, et al. Exploiting homophily effect for trust prediction [C]// Proceedings of the 6th ACM International Conference on Web Search and Data Mining. Rome, Italy, 2013: 53-62
- [8] Tang J, Gao H, Liu H, et al. eTrust: Understanding trust evolution in an online world [C]// Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China, 2012: 253-261
- [9] Matsuo Y, Yamamoto H. Community gravity: measuring bidirectional effects by trust and rating on online social networks [C]// Proceedings of the 18th International Conference on World Wide Web. Madrid, Spain, 2009: 751-760
- [10] Au C, Iwata T. Strength of social influence in trust networks in product review sites [C]// Proceedings of the 4th ACM International Conference on Web Search and Data Mining. HongKong, 2011: 495-504
- [11] Ziegler C N, Golbeck J. Investigating interactions of trust and interest similarity [J]. Decision Support Systems, 2007, 43(2): 460-475
- [12] O'doherty D, Jouili S, Van P. Towards trust inference from bipartite social networks [C]// Proceedings of the 2nd ACM SIGMOD Workshop on Databases and Social Networks. Scottsdale, AZ, USA, 2012: 13-18
- [13] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks[C]// Proceedings of the 19th International Conference on World Wide Web. Raleigh, North Carolina, USA, 2010: 641-650
- [14] Ma H, Zhou T C, Lyu M R, et al. Improving recommender systems by incorporating social contextual information[J]. ACM Transactions on Information Systems (TOIS), 2011, 29(2): 9
- [15] O'doherty J, Smyth B. Trust in recommender systems [C]// Proceedings of the 10th International Conference on Intelligent User Interfaces. San Diego, CA, USA, 2005: 167-174
- [16] Yuan W, Lee S, Han Y, et al. Efficient routing on finding recommenders for trust-aware recommender systems [C]// Proceedings of the 6th International Conference on Ubiquitous Information Management and Communication. Kuala Lumpur, Malaysia, 2012: 29
- [17] Wasserman S, Faust K. Social network analysis: Methods and applications [M]. Cambridge university press, 1994
- [18] Liben D, Kleinberg J. The link-prediction problem for social networks [J]. Journal of the American Society for Information Science and Technology, 2007, 58(7): 1019-1031

(上接第 137 页)

- [9] Moschitti A. A study on convolution kernels for shallow semantic parsing [C]// Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004
- [10] Moschitti A. Making Tree Kernels Practical for Natural Language Learning [C]// EACL. 2006
- [11] Joachims T. Making large scale SVM learning practical [OL]. <http://www.academic.research.microsoft.com/paper/1361548>
- [12] Wiegand M, Klakow D. Convolution kernels for opinion holder extraction [C]// The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. Los Angeles, California; Human Language Technologies. Association for Computational Linguistics, 2010
- [13] Qiu G, Liu B, Bu J, et al. Opinion word expansion and target extraction through double propagation [J]. Comput. Linguist., 2011, 37(1): 9-27
- [14] Kim Y, Kim S, Myaeng S. Extracting Topic-related Opinions and their Targets in NTCIR-7 [C]// Proc. of NTCIR-7 Workshop. 2008
- [15] Agarwal A, Xie B, Vovsha I, et al. Sentiment analysis of Twitter data [C]// Portland, Oregon; Proceedings of the Workshop on Languages in Social Media. Association for Computational Linguistics, 2011
- [16] Wilson T, Wiebe J, Hoffmann P. Recognizing contextual polarity in phrase-level sentiment analysis [C]// Vancouver, British Columbia, Canada; Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2005
- [17] Liu B, Hu M, Cheng J. Opinion observer: analyzing and comparing opinions on the web [C]// Proceedings of the 14th international conference on World Wide Web. ACM, 2005
- [18] Collins M, Duffy N. New ranking algorithms for parsing and tagging; Kernels over discrete structures, and the voted perceptron [C]// Proc. of the ACL 2002. 2002: 28-136
- [19] 赵妍妍, 秦兵, 车万翔, 等. 基于句法路径的情感评价单元识别 [J]. 软件学报, 2011(5)
- [20] 王倩, 何婷婷, 闻彬, 等. 基于依存关系的中文情感要素抽取技术研究 [C]// 第十届全国计算语言学学术会议. 烟台, 2009
- [21] 黄亿华, 濮小佳, 袁春风, 等. 基于句法树结构的情感评价单元抽取算法 [J]. 计算机应用研究, 2011, 28(9): 3229-3234
- [22] 姜德成, 姚天坊. 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): 2622-2655
- [23] 黄晨, 钱龙华, 周国栋, 等. 基于卷积核的无指导中文实体关系抽取研究 [J]. 中文信息学报, 2010, 24(4): 11-17