

基于依存分析的中文微博情感分析研究

孙艳[†] 周学广 李中远

海军工程大学信息安全系, 武汉 430033

[†] 通信作者, E-mail: zyg196610@hotmail.com

摘要 有效地管理和使用微博评论信息是当前的迫切需求。在依存分析的基础上对中文微博进行了情感分析, 参加了 NLP&CC2012 中文微博情感分析评测中的所有 3 个任务。总结了新词出现的原因, 通过对成词的规则制定和串频统计等, 提出一种基于规则和统计的新词识别算法。通过对现有情感词库的整合, 构建了自己的情感词库。对情感词和评价对象的依存模式进行了分析, 总结出 6 种评价单元模式, 结合依存树的剪枝和归并策略, 提出了基于依存分析的观点句识别方法和基于依存分析的情感要素抽取方法。在任务 3 的情感要素抽取评测中成绩排名第 3, 效果理想。

关键词 情感分析; 依存分析; 情感要素抽取; 新词识别

中图分类号 TP391

Research on Chinese Weibo Sentiment Analysis Based on Dependency Parsing

SUN Yan[†], ZHOU Xueguang, Li Zhong yuan

Department of Information Security, Naval University of Engineering, Wuhan 430033

[†]Corresponding author, E-mail: yzिसun@126.com

Abstract Effectively manage and to use microblog comments become a hotspot. In this paper, the methods based on dependency parsing in all three tasks of NLP&CC2012 the Chinese microblog sentiment analysis evaluation are detailed introduced. The appeared reasons of new words were summarized. Based on the rulemaking into words and string frequency statistical, a new words recognition algorithm is proposed. Through the integration of existing sentiment words lexicons, we build an own sentiment lexicon. Through six kinds of summarized evaluation unit mode, combining the dependency tree pruning and merging strategy, the opinion sentence identification method and the sentiment feature extraction method based on dependency parsing were proposed. The evaluation results, especially in the task 3, proved the effectiveness of the methods we proposed.

Key words sentiment analysis; dependency parsing; sentiment information extraction; new words identification

网民的增加和微博的发展使得大量评论信息迅速膨胀^[1], 仅靠人工的方法难以应对微博海量信息的收集和处理, 如何有效地管理和使用这些微博评论信息成为当前的迫切需求。对这些微博评论信息进行情感分析, 是正确引导其它用户决策的关键, 为政府进行舆情监控提供帮助, 具有较高的应用和商业价值。

在情感分析的各任务中, 情感要素抽取是最重要且最难的。情感要素包括情感词和评价对象, 作为一个评价单元, 利用句法规则可以获取到对方^[2]。在Hatzivassiloglou等人的情感词扩展中^[3], 形容词之间的并列关系或转折关系表示的是句法关系, 引申到整个句子, 可以利用句法分析进行情感词和评价对象获取。Kobayashi等人通过句法分析查找(特征, 评价词)对来获取评价对象^[4]。Qiu等人利用情感词与评价对象的依存关系获取评价对象^[5-6]。Zhang等人提出一种基于双向图的迭代算法, 称为MRE(Mutual Reinforcement based on Expected Values)算法, 来获取隐性的评价对象和评价词^[7]。赵等人通过自动获取句法路径来描述评价对象及其评价词语之间的修饰关系, 再计算句法路径编辑距离抽取评价单元^[8]。

中文微博情感分析评测是中国计算机学会中文信息技术专业委员会(CCF TCCI)主办的公开评测, 它是2012年CCF自然语言处理与中文计算会议(NLP&CC 2012)的一部分^[9]。中文微博情感分析评测对象是面向中文微博的情感分析核心技术, 包括观点句识别、情感倾向性判断和情感要素抽取, 分别为任务

1、任务 2 和任务 3，其中任务 1 是必选任务，任务 2 和任务 3 都是基于任务 1 的。本课题组参加了所有 3 个任务。本文将详细阐述每个任务的技术和方法，并对评测结果进行分析。

1 新词识别

本文提出的基于依存分析的中文微博情感分析需要对文档进行依存句法分析，而句法分析需要在分词和词性标注的基础上进行。随着时代的发展，新事物的出现，大量新词不断出现，影响分词效果，进而影响句法分析效果。本文在已有新词识别基础上^[10]，借助于 ICTCLAS 中文分词工具，通过对成词的规则制定和串频统计等，提出一种基于规则和统计的新词识别方法。

步骤如下：

1) 结合词性过滤和停用词表过滤，设计一个语素表，只有被语素表包含的词才与周围词结合成候选新词加入候选新词列表，语素是指可能成为新词的分词词库中的词。

2) 候选新词剪枝

规则的制定虽然可以过滤掉大部分的垃圾串，但是对于子串为垃圾串的情况却没有考虑。因此需要对语素数目大于 2 的序列的所有候选新词子序列进行剪枝。

3) 采用式 (1) 计算成词可能性权重：

$$w = p_{trans} * df \quad (1)$$

其中 p_{trans} 为转移频率， df 为文档频数， p_{trans} 的值等于语料库中为一个分词切片序列的出现频数除以该序列第一个切片的出现频数。

根据成词权重大小选取大于阈值的候选新词为新词识别结果。对于 5 字及 5 字以上的候选新词噪音非常大，仅录用 2 字、3 字和 4 字的新词。2 字、3 字和 4 字的成词权重阈值是不同的。相较于 3 字、4 字结果，对于一个较低的概率，当 3 字、4 字结果中的噪音比较大时，2 字结果的噪音数量依然较小。因此，我们根据字数来设定 3 个结果表，分别保存 2 字、3 字和 4 字结果。选出每个结果表中权重最大的若干结果作为新词识别结果。

2 依存树的修剪

根据对具有依存关系的两个成分进行分析，发现有些依存词对是无意义的依存词对，对于倾向性分析没有意义，影响文本倾向性分析精度，因此采用以下策略对依存树进行修剪，圆圈加大写字母代表词语。

1) 剪枝

对依存树的剪枝操作是指将当前节点进行删除，其作用域仅仅局限于当前节点中，其子节点并不发生作用。图 1 表示的是对依存树的剪枝操作，即直接将其子节点 B 依靠在父节点 A 上，当前节点不加入依存词对的生成分析，剪枝操作通常应用于助词及介词标签如 DEC、DEG、P 等，可将具有间接依存关系的两个词变成具有直接依存关系。

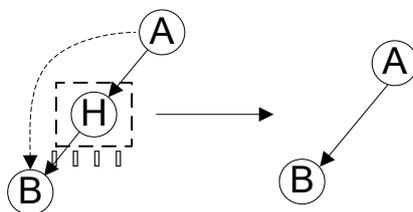


图 1 依存树的剪枝

Fig.1 Dependency tree pruning

2) 归并

依存树的归并是将子节点和父节点归并为一个节点，新的节点对应的词是子节点和父节点的结合。图 2 表示的是对依存树的归并操作，即将节点 B 和 H 归并为 B' 节点。归并操作用于否定词和情感词中，将否定词和情感词归并为一个新的情感词，新的情感词的极性是原情感词的极性取反。通过归并可将具有间接依存关系的两个词变成直接依存关系。

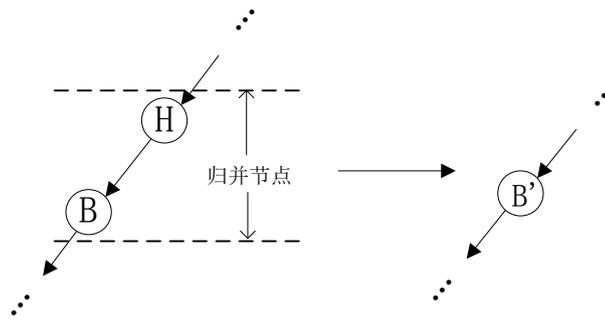


图 2 依存树的归并

Fig.2 Dependency tree merge

3 情感词库

① 情感词库

以《知网》公布的情感词表、《常用褒贬义词语详解词典》、《学生褒贬义词典》、《褒义词词典》和《贬义词词典》为基础，删除其中使用频率很低的情感词，增加网络用语和口语情感词，构建了一个情感词库，其中含褒义词 5554 个，贬义词 6321 个。

② 极性词库

有部分情感词极性非常强烈，特别是一些骂人的贬义词，在观点句的识别时，只要出现这些词，就将观点句极性判定为该词的极性（否定句式取反）。为区别于大的情感词库，将这类词库称为极性词库，其中包含褒义词 16 个，贬义词 262 个。情感词库与极性词库是包含关系，即情感词库包含极性词库中的所有词。

③ 否定词库

否定词库对于情感词的极性判断很重要。否定词库含“不、未、没有、欠失”等否定词 20 个。

④ 能愿动词库

能愿动词库用于观点句的识别中。能愿动词库含“祝、希望、应该、必须”等能愿动词共 16 个。

⑤ 条件连词库

条件连词库用于观点句的识别中。条件连词库含“如果、要是、假如、也许”等条件连词共 11 个。

4 观点句的识别

按照中文微博情感分析评测中观点句的定义，结合新词识别方法和词库，提出基于依存分析的观点句识别方法。

- ① 利用第 1 节提出的新词识别方法，对评测数据集进行新词识别，得到 13300 个新词，加入分词词典文件 dict 中；
- ② 采用 ctbparser 对句子进行中文分词、词性标注和句法分析；
- ③ 若含有极性词库中的词，则认为该句子是观点句，否则转④；
- ④ 主语为第一人称代词，谓语为主张动词，状语为表达“喜、怒、哀、乐”的情感词的句子认为不是观点句；
- ⑤ 含有能愿动词库中的词作谓语的句子认为不是观点句；
- ⑥ 由条件连词库中的词作连词的句子认为不是观点句；
- ⑦ 其它含有情感词且符合 5 节中 6 种模式的句子认为是观点句。

5 情感要素的抽取

本文提出基于依存分析的情感要素抽取方法。基本思想是先得到候选情感词和候选评价对象，再对候选的评价单元进行筛选。将待分析句子中所有在情感词库中出现的词列为候选情感词，所有名词或代名词列为候选评价对象。

按照表 1 中的评价单元模式对候选的评价单元进行筛选。

表 1 评价单元模式表

Table 1 Evaluation unit mode table

评价单元模式	情感词词性	评价对象词性
$\{\mathbf{NP}\} \xleftarrow{SUB} \{\mathbf{AV}\}$	VA, VV	NN, NR, NT, PN
$\{\mathbf{NP}\} \xleftarrow{SUB} \{\mathbf{VC}\} \xrightarrow{PRD} \{\mathbf{NP}\} \xrightarrow{NMOD} \{\mathbf{JV}\}$	VA, JJ	NN, NR, NT, PN
$\{\mathbf{AV}\} \xrightarrow{OBJ} \{\mathbf{NP}\}$	VA, VV	NN, NR, NT, PN
$\{\mathbf{NP}\} \xleftarrow{SUB} \{\mathbf{VV}\} \xrightarrow{VMOD} \{\mathbf{AV}\}$	VA, VV	NN, NR, NT, PN
$\{\mathbf{JV}\} \xleftarrow{NMOD} \{\mathbf{NP}\} \xleftarrow{SUB} \{\mathbf{VV}\}$	VA, JJ	NN, NR, NT, PN
$\{\mathbf{VV}\} \xrightarrow{OBJ} \{\mathbf{NP}\} \xrightarrow{NMOD} \{\mathbf{JV}\}$	VA, JJ	NN, NR, NT, PN

6 实验结果与分析

数据集采用评测公开数据集，于 2012 年 7 月发布评测数据集。该评测数据全部来自腾讯微博。评测数据包括 20 个话题，每个话题约 2000 条微博，数据集大小为 5.6M，共包含 31673 条微博。数据采用 unicode (utf-16) 编码的 xml 格式，已经预先切分好句子。

实验所用的依存分析工具采用复旦大学的 ctbparser 工具包，词性标注采用中文宾州树库标准，评测实验在 Visual studio 2010 上采用 C++ 语言实现。

6.1 任务 1 评测结果与分析

任务 1 为观点句的识别。观点句的识别需要针对每条微博中的各个句子，判断出该句是观点句还是非观点句。它是情感倾向性判断和情感要素抽取的基础。

任务 1 共有 34 个单位提交 53 组结果，表 2 给出了本组结果与任务 1 所提交的 53 组结果的平均值和 53 组结果中的每个评价指标的最好值比较。

表 2 观点句识别评测结果比较表

Table 2 The evaluation results of opinion sentences identification comparison table

结果编号	微平均			宏平均		
	准确率	召回率	F值	准确率	召回率	F值
17	0.828	0.537	0.651	0.826	0.520	0.629
平均值	0.727	0.615	0.647	0.727	0.607	0.634
最好值	0.835	0.959	0.784	0.836	0.96	0.783

从和表 2 中可以看出，基于依存分析的观点句识别方法在评测结果中准确率较高，但召回率偏低。在微平均评测结果中准确率排名第 2，与排名第 1 的仅差 0.007，在宏平均评测结果中准确率排名第 3，与排名第 1 的相差 0.01。但是召回率结果在所有结果中处于中下，在微平均评测结果中比平均值低 0.078，在宏平均评测结果中比平均值低 0.087。虽然准确率较高，但综合考虑准确率和召回率，基于依存分析的观

点句识别方法的 F 值仅与平均值约持平，在微平均评测结果中比平均值略高，约 0.004，在宏平均评测结果中略低于平均值，约差 0.005。

通过观点句识别评测结果，分析基于依存分析的观点句识别方法进行观点句识别的步骤，发现造成召回率不高的原因有两个：一是情感词库不够全，二是对于不含情感词库中的词的观点句不能识别出来。

6.2 任务 2 评测结果与分析

情感倾向性判断评测要求判断微博中每条观点句的情感倾向。该任务在观点句识别的基础进行，针对每条微博中的各个句子，先进行观点句识别再进行观点句的倾向性判断。观点句的情感倾向可以分为正面 (POS)，负面 (NEG) 和其它 (OTHER)。OTHER 表示中性以及其它无法明确归为正面或者负面的倾向。

由于观点句和观点句中的情感词极性已经确定，故采用最简单的线性相加法进行句子的情感倾向性判断，步骤如下：

统计观点句中情感倾向性值，句子从左往右，含褒义情感词+1，含贬义情感词-1，出现“但、但是”连词，将情感倾向性值清零，从转折后继续计算，最后得到句子的倾向，大于 0 判断为 POS，小于 0 判断为 NEG，等于 0 为 OTHER。

任务 2 共有 32 个单位提交了 48 组结果，表 3 给出了本组结果与任务 2 所有组的结果平均值和最好值比较。

表 3 情感倾向性判断评测结果比较表

Table3 The evaluation results of opinion polarity judgment comparision

结果编号	微平均			宏平均		
	准确率	召回率	F值	准确率	召回率	F值
17	0.772	0.415	0.54	0.776	0.404	0.523
平均值	0.704	0.46	0.546	0.702	0.454	0.536
最好值	0.881	0.698	0.741	0.878	0.702	0.733

线性相加法操作、简单实现快速。从表 3 我们可以看出，线性相加法在任务 2 中的情感倾向性判断评测中准确率比平均值略高，召回率低于平均值，导致 F 值比平均值略低。对比表 2 和表 3，情感倾向性判断评测准确率比观点句识别准确率在微平均评测中下降了 0.056，在宏平均评测中下降了 0.05；召回率在微平均评测中下降了 0.122，在宏平均评测中下降了 0.116。这个结果在可接受范围内，但是鉴于情感倾向性判断的最好结果是观点句识别结果持平，从这一点看，线性相加法仍存在可改进空间。

6.3 任务 3 评测结果与分析

情感要素抽取评测要求找出微博中每条观点句作者的评价对象，即情感对象，同时判断针对情感对象的观点极性。情感要素抽取也是在观点句识别的基础上进行的，需要先进行观点句识别，再对观点句进行情感要素抽取。

任务 3 共有 15 个单位提交了 22 组结果，表 4 给出了严格评价指标下本组结果与任务 3 所有组的结果平均值和最好值比较。

表 4 情感要素抽取（严格评价指标）评测结果比较表

Table 4 The strict evaluation results of sentiment feature extraction comparision

结果编号	微平均			宏平均		
	准确率	召回率	F值	准确率	召回率	F值
17	0.311	0.177	0.225	0.324	0.174	0.22
平均值	0.172	0.098	0.112	0.174	0.098	0.112
最好值	0.485	0.275	0.288	0.474	0.265	0.278

从表 4 中可以看出，基于依存分析的情感要素抽取方法在 22 组提交结果中名列前茅，微平均和宏平均的准确率、召回率和 F 值都排名第 3，F 值正好是平均值的 2 倍。这个结果表明了基于依存分析的情感

要素抽取方法通过定义 6 种评价单元模式并对情感对象进行归并能够抽取出完整的情感要素。

表 5 给出了宽松评价指标下本组结果与任务 3 所有组的结果平均值和最好值比较。

表 5 情感要素抽取（宽松评价指标）评测结果比较表
Table 5 The lenient evaluation results of sentiment feature extraction comparison

结果编号	微平均			宏平均		
	准确率	召回率	F值	准确率	召回率	F值
17	0.404	0.223	0.287	0.42	0.219	0.279
平均值	0.289	0.149	0.18	0.29	0.149	0.177
最好值	0.636	0.356	0.371	0.643	0.344	0.359

根据评测大纲，在抽取情感对象时，要求抽取出完整和明确的对象。例如在例句“ipad 的屏幕很棒！”中，按照严格评价指标，要求抽取出“ipad 的屏幕”，而不仅是“屏幕”。但是在宽松评价中，抽取出“屏幕”就算正确。

对比表 4 和表 5 我们发现，本文的基于依存分析的情感要素抽取方法在宽松评价指标评测中，微平均的 F 值是严格评价中 F 值的 1.276 倍，宏平均是 1.268 倍。计算 22 组结果的微平均 F 值平均值在宽松评价中是严格评价中的 F 值的 1.607 倍，宏平均是 1.580 倍。这两组结果说明了本文的基于依存分析的情感要素抽取方法在宽松评价指标评测中的提高率比平均值的提高率低，进一步表明了基于依存分析的情感要素抽取方法采用依存树的归并来得到完整的情感对象是有效的。

7 结语

本文主要介绍了参加 NLP&CC2012 中文微博情感分析评测的基本情况。提出了基于依存分析的观点句识别方法、线性相加法和基于依存分析的情感要素抽取方法。在任务 3 中的情感要素抽取评测中取得了理想的结果，但在任务 1 观点句识别和任务 2 情感倾向性判断中召回率方面还需深入研究。

参考文献

- [1] Thelwall M, Buckley K, Paltoglou G. Sentiment in Twitter events[J]. Journal of the American society for Information Science and Technology, 2011, 62(2): 406-418.
- [2] Zhai Zhong wu, Liu Bing, Zhang Lei, et al. Identifying Evaluative Sentences in Online Discussions[C]. In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, Cambridge: AAAI Press, 2011: 933-938.
- [3] Hatzivassiloglou V, Mckeown K. Predicting the Semantic Orientation of Adjectives[C]. In Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics. Stroudsburg: ACL, 1997: 174-181.
- [4] Kobayashi N, Inui K, Matsumoto Y. Extracting Aspect-evaluation and Aspect-of Relations in Opinion Mining[C]. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Stroudsburg: ACL, 2007: 1065-1074.
- [5] Qiu Guang, Liu Bing, Bu Jia jun, et al. Opinion Word Expansion and Target Extraction through Double Propagation [J]. Computational Linguistics, 2009, 37(1): 9-27.
- [6] Qiu Guang, Liu Bing, Bu Jia jun, et al. Expanding Domain Sentiment Lexicon through Double Propagation[J/OL]. <http://ijcai.org/papers09/Papers/IJCAI09-202.pdf>, 2012-9-11.
- [7] Zhang Lei, Liu Bing. Extracting Resource Terms for Sentiment Analysis[C]. In Proceedings of the 5th International Joint Conference on Natural Language Processing, Chiang Mai: Thailand, 2011: 1171-1179.
- [8] 赵研研, 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, 2011,22(5): 887-898.
- [9] 中文微博情感分析评测[DB/OL]. <http://tcci.ccf.org.cn/conference/2012/pages/page04eva.html>, 2012-9-12.
- [10] 王洋. 网络舆情分析技术的研究与实现[D], 武汉: 海军工程大学, 2011.