多策略中文微博观点句识别及情感倾向性判断

郭云龙†，潘玉斌，李国祥，陆阳，余小明，李莉†

西南大学计算机与信息科学学院，重庆 400715

† 通讯作者，E-mail: 286207758@qq.com，lily@swu.edu.cn

摘要 随着社会媒体和社会网络的飞速发展，与微博相关的研究引起了广泛地关注。本文针对中文微博数据(语句)，首先采取工具模板进行标注并利用SVM特征分类方法提取微博观点句，然后采用Stanford Parser工具，并构建情感词典等方法对得到的微博观点句进行情感倾向性分析。本文参加了中国计算机学会(CCF)自然语言处理与中文计算会议（NLP&CC 2012）情感分析&词汇语义关系抽取评测。利用会议评测提供的数据，我们得到观点句准确率和情感分类准确率分别为：78.3%和 82.4%[1]。

关键词 微博；观点句；情感分析；SVM

中图分类号

Multiple Strategies Based Opinion Sentence Extraction and Sentiment Analysis in Chinese Micro Blog

GUO Yunlong†, PAN Yubing, LI Guoxiang, LU Yang， YU Xiaoming, Li Li†

School of Computer and Information Science, Southwest University, Chongqing 400715

†Corresponding Author, E-mail: 286207758@qq.com，lily@swu.edu.cn

Abstract With the development of social media and social network, micro-blog has drawn increasing attention from both academia and industry. In this paper, we elaborate on our multiple strategies approach and evaluate it based on the available dataset from NLP&CC 2012. Firstly, syntactic analysis is performed to obtain most of the opinion sentences. This is followed by SVM based approach as a complement to improve the accuracies of opinion sentences extracted. Sentiment analysis is then carried out on these sentences with Stanford Parser tool and the constructed dictionary. The dictionary, initially constructed by search the corpus we obtained from Google, is continually expanded and modified to accommodate the evolution of micro blogging. The experimental results show that the precision of extracting the opinion sentences is 78.3% and the precision of the sentiment classification is 82.4%, respectively.

Key words micro-blog; opinion sentence; sentiment analysis; SVM

1 引言

1.1 研究背景

的随着web2.0 的发展，微博[2]，即微博客（Micro Blog）的简称，一个基于用户关系的信息分享、传播以及获取平台迅速兴起。据统计，新浪微博截至到 2012 年 2 月，注册用户已突破 3 亿大关，用户每日发博量超过 1 亿条。网络社区的信息量与日俱增，这些资源共享信息为用户的学习、生活、工作带来了巨大便利，可以说，以微博为典型代表的社会媒体在国家经济、安全以及现代信息服务等领域说具有举足轻重

[1] http://tcci.ccf.org.cn/conference/2012/pages/page04_evares.html
[2] http://baike.baidu.com/view/1567099.htm

的作用。但同时，巨大的冗余信息使的人们很难在短时间内准确、迅速地获取有用的信息，如Hu(2004)[1]的文章所述。所以，针对微博数据的分析已成为国内外研究热点，针对中文微博而言，中文微博观点句抽取以及观点句的情感倾向性判也日益引起人们的关注。近年来，ACL、SIGIR、KDD等国际会议，都有相关议程探讨该领域的发展，NTCIR[1]、COAE[2]等评测也涉及该研究热点。

观点句的抽取以及文本感倾向性判断问题[2-3]，可以理解为基于数据文本的一种二分类的句子级文本分类技术。当前国内对于句子级文本分类的主要方法大致可以分为三类：（1）基于词典的方法；（2）基于有监督的机器学习方法；（3）无监督机器学习方法。

基于词典的方法：利用预先构建的词典(可以是人工标注或是机器统计的)，处理文本中出现的词语及其情感信息，进而判断其主客观性(即观点性或非观点性)。针对所得到的观点句和非观点句，利用词典中的正向情感词、负向情感词，统计待测文本中两类词语相差值，从而确定该文本情感类型。一般词典方法需结合标点符号和规则，例如，张(2011)[6]将观点句分为显性观点句与隐性观点句，集合词典方法和机器学习方法，提取微博观点句。姚等(2007)[4]以词语和标点符号作为分类特征，对特定领域的微博进行了分析。

基于有监督的机器学习方法：利用训练集，采用特定的机器学习方法，对测试集进行分类。常用的机器学习方法包括：朴素贝叶斯(Naive Bayes)、最大熵(Max Entropy)、支持向量机(Support Vector Machine）等。而针对特征值的选取，常用的方法有信息增益方法（IG)、卡方分布（CHI）值统计、文档频率(DF)、词频反文档频率（TF-IDF)。具体可参见刘等(2012)[5]的文章。

无监督机器学习方法：包括基于情感基准词的方法、基于图论分割句子并结合一定的规则方法。这对情感基准词依赖太高。

2 背景

2.1 观点句与句子的情感倾向性

NLP&CC2012 评测对观点句的要求为：(句子)只限于对特定事物或对象的评价,不包括对自我情感、意愿或心情的表达。例如“我感到很高兴。”，这样的句子是明显的情感句，但不属于本评测定义的观点句。而“我真心喜欢iphone5 的屏幕效果。”，该句属于本评测定义的观点句。本文参照张（2011）[6]一文中对观点句的分类来提取观点句。

显性观点：以指示性动词作为句子的核心谓语，明确地表达说话人观点的句子。根据文献[6]，我们修改了该文总结的指示性动词，并加入微博用户大量使用的网络指示性动词，构建了新的指示性动词表。当指示性动词出现时，人们通常是在明显地、高调地发表某些评论。此类词语的出现的句子，可以判定为观点句。这类句型具有较明显的观点句句法特征，我们将其统称为显性观点句。我们将使用工具模版标注的方法对显性观点句进行提取。

表 1　指示性动词表

Table 1　a list of indicative verbs

指示性动词表
表态 表示 承认 答称 道 感到 告诉 还说 即 表示 觉得 盼望 批评 称 说 希望 相信 形容 要求 以为 指 解释 想

隐性观点句：不含指示性动词，但整体意图是为了发表某种观点、看法或评论的句子。例如：“日本人都该死!”，该例句中不存在任何指示性动词，但通读全句后，我们可以看出这句话是在对“日本人”这个对象进行分析并做出了评价。这类观点句没有明显的观点表达标志，并且表达形式多变，对此类隐性观点句，我们将主要采用 SVM 来处理此类句子。

情感倾向性：所有的观点句都具有情感倾向，而同时有情感倾向的句子也一定是观点句。不失一般性，

[1] http://baike.baidu.com/view/1035539.htm
[2] http://www.ir-china.org.cn/coae2011.html

我们将分为三种：积极，消极和中性。通过建立的情感词典，统计其正负向情感词语在待评测观点句中的差值，从而判断该观点句的情感倾向性。

2.2 支持向量机 SVM

SVM根植于Vapnik[13]提出的统计学理论，在机器学习文本分类领域有很强的实用性，在使用时，首先选取特征，然后将文本向量化，并用足够多的示例训练分类器，然后生成分类模型并完成对测试数据的分类。具体来讲，SVM是一种监督式学习的方法，属于一般化线性分类器。该分类器的特点是能够同时最小化经验误差与最大化几何边缘区。因此SVM也被称为最大边缘区分类器。被广泛应用于文本分类领域。

SVM常用的特征选取方法[9]有： 文档频率DF、信息增益IG、卡方统计CHI、词频反文档频率（TF-IDF）等，SVM的关键性问题在于特征向量的选取、转换。本文使用Libsvm工具[1]。

3 算法设计

本文采取工具模版标注及 SVM 特征分类方法提取微博观点句。本节我们将讨论算法设计以及所使用的工具，在此之前，我们简单介绍一下文本预处理工作。

3.1 文本预处理

分词是文本预处理的重要环节，中文有别于英文，没有明显的词语分隔标志，中国科学院计算技术研究所研制出了汉语词法分析系统(ICTCLAS)能有效的快速的分隔出带有明确语义的词语。我们对其所分出的词语去噪，并采用其词性标注功能。由于ICTCLAS对繁体字支持较弱，我们先对待评测数据进行简繁体转化。这样，ICTCLAS的分词结果可直接用于斯坦福句法分析工具（Stanford Parser）[2]的输入以及SVM所选取特征向量值的计算。

3.2 观点句提取

根据前面讨论的显性观点句和隐形观点句的区别，本文使用斯坦福句法分析工具并结合构建的指示性动词词典，以及显性观点句构成的模版规则，对显性观点句进行提取；使用 SVM 并结合§3.2.1 的 8 个特征作为筛选，实现对隐形观点句的提取。

3.2.1显性观点句的提取

斯坦福句法分析工具[8]是一款以Java实现的开源句法解析工具，主要基于优化的基于概率规则集和词汇化依存句法分析方法，是一个词汇化的概率上下文无关语法分析器，同时也使用了依存分析。根据不同的语法可以输出不同的分析结果。所以，可以认为该工具是一个使用混合分析方法的剖析器。针对句子“我真心喜欢iphone5的屏幕效果。”，我们得到如下的分析结果。

```
(ROOT
  (IP
    (NP (PN 我))
    (VP
      (ADVP (AD 真心))
      (VP (VV 喜欢)
        (IP
          (VP
            (NP
              (DNP
                (QP (CD iphone5))
                (DEG 的))
              (NP (NN 屏幕) (NN 效果)))))))
    (PU 。)))
```

图 1　使用斯坦福语法分析工具的例子

Fig. 1 Results from Using Stanford Parser.

[1] http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html
[2] http://nlp.stanford.edu/software/lex-parser.shtml

具体为：先把待测数据分离出若干个分句（IP）语法树，结合构建的指示性动词表与显性观点句构成模版，对语法书中词语进行匹配，确定符合模版规则分句，从而判断该句子是否为显性观点句。进而，我们还结合句法规则，参照文[6]中的句法结构模版，对微博句子是否是显示观点句进行了判断。三种模版均以指示性动词作为匹配的起始点，具体定义如下：

(1) S_1 =<Subject>NP+<predicate> Indicative Verb+…ADJP…

解释：句子的主语是名词短语，句子的谓语动词或几个谓语动词之一为指示性动词，且此谓语动词的父节点的其他子节点中存在形容词短语。

(2) S_2=<Subject>NP+<predicate> Indicative Verb+…ADVP…

解释：句子的主语是名词短语，句子的谓语动词或几个谓语动词之一为指示性动词，且此谓语动词的父节点的其他子节点中存在副词短语。

(3) S_3=<Subject>NP+<predicate>Indicative Verb+…VC是…

解释：句子的主语是名词短语，句子的谓语动词或几个谓语动词之一为指示性动词，且此谓语动词的父节点的其他子节点中存在动词“是”。

3.2.2 隐形观点句特征值的提取

利用 SVM 模型来判定隐形观点句，基本流程包括：

① 特征值的选取：特征值的选取分析在下文中具体阐述，而向量化中用到的方法，可以是词频统计、TF-IDF 值、相似度等，下面提到的卡方统计 CHI 值的计算公式为：

$$CHI(p,c_j)=\frac{N\times(AD-CB)^2}{(A+C)\times(B+D)\times(A+B)\times(C+D)} \quad (1)$$

其中，p表示模式，c_j代表与文本类，在本文中c_j即为观点或非观点，N表示训练集中的句子总数，A表示c_j类中模式p出现的次数，B表示非c_j类中模式p出现的次数，C表示c_j类中没有出现模式p的句子数，D表示非（c_j）类中没有出现模式p的句子数。

② 向量化：通过特征值计算出的 CHI 值，作为训练集或测试集的每一列向量，向量的维数等于特征值向量维数之和。从而实现数据的向量化，输入到 SVM 中进行分类处理。

③ 训练模型。

④ 将训练好的模型用于新测试数据。

1） 情感词

即带有某种情感倾向的形容词。通常，当某个人发表了一些带有情感倾向的微博时，我们会认为它带有观点性。实验所用的情感词词表包含 9472 个情感词，它由两部分组成，一部分来自于HowNet[14]的情感词词典，另一部分是从实验语料中提取出的形容词。在后面的实验中此特征简记为F1。

对于 F1 的处理，我们首先将其看成一个 9472 维的特征向量，每一个维度的值用 CHI 值来表示。但最后发现，用此方法处理，并不好，因为一句话中能包涵的情感词是极其有限的，这将导致一句话形成的向量矩阵特别稀疏，并极度影响后面向量 SVM 的处理，因此我们将其做了如下修正：

a. 情感词典的修正：原情感词典中，有一些不符合本文的判断依据，因此对其进行修正，最后形成 8223 个词语的情感词典。

b. 特征值的修正：我们在此项中放弃使用 CHI 值而给每一情感词赋予权值，初始值都为 1，在处理训练集时，当情感词 t 在观点句中出现，则将 t 的权值加 1，若 t 在非观点句中出现，则将 t 的权值减 1，最后形成情感词的权值表。在对测试集处理时，对比该权值表，这样，每个句子只需对应一个一维的向量，该向量的值等于句中所有情感词的权值之和。直接的好处就是降低了计算的复杂度，极大地提供了特征值处理的效率。

2） 动词

动词通常是发表主观言论的标志，如“表示”、“相信”、“认为”、“预测”等，但我们发现并非所有的动词都具有明显的标志，结合前面的指示性动词表，我们使用其 23 个动词作为特征值之一。记为 F2，为防止矩阵的稀疏化，在处理方式上，对比了动词特征的总共 23 个维度 CHI 值和只有 1 个维度的频率统计

效果之后，我们选择了后者，处理方式与情感词的处理方式相同。

3） 副词

上文提取的情感几乎都是形容词性的，关于是否把副词也作为 SVM 分类特征值的问题，通过我们的观察，我们发现：微博用户在发表博文中使用副词的频率较低，或有意无意地用形容词的“的”代替了副词的“地”，由与特征值 F1 的出现，导致副词的参考价值很低，所以本文我们不考虑该词性。

4） 表情

表情，作为微博的特色之一，直观、形象地表达了用户的情感和态度，可以作为微博观点句的重要参考因素。但经过观察我们发现，用户对表情使用的随意性很大，这使得表情在体现用户观点态度上的作用大大降低。用户使用表情时往往并不是真实心态的写照，而是随意、时尚、从众心理所使。因此，本文不考虑选取表情作为特征词。

5） 网络词

微博充斥着各种网络词，新兴的网络词直观反应用户的观点。但是由于中科院分词系统目前还无法分析出网络词，不能其作为特征值之一作为特征向量机的分类特征。故提炼网络词表在现阶段基本没有意义，网络词将在下一阶段的研究的涉及。

6） 词性的统计

我们认为，观点句中的词性应该会展现出明显的特点，因此采用对观点句的词性进行统计，作为特征向量，称其为 F3。在中科院分词工具中，能将词语具体定义为 96 种词性，具体而言，我们计算 CHI 值，生成了一个 96 维向量作为特征向量。

观察发现，除单一词性外，两个连续词性也应该作为用户观点句分类的特征值，这样，我们得到特征 F4。对其进行拓展，我们选用连续词性的组合作为特征。由于连续词性的组合将会有 96*96=9216 种，我们估计会出现情感词最初始处理时的情况，即：特征矩阵稀疏的情形。实验证明了我们的猜测，所以，我们在处理训练集时，只选取了 CHI 值的前 100 个(top100)特征，形成了一个 100 维的特征向量。 实验结果表明，我们的选取效果不错。

对于连续 3 个或 3 个以上的连续词，由于中文微博字数有限，这样的情况出现频率过低，无法清晰地作为分类特征，所以，本文暂时不考虑这种情况。

7） 词语的统计

某些特定词语往往充分体现了用户的态度，这些词语具有普遍性，经过中科院分词系统词语统计能到达一定的效果，我们称其为 F5。鉴于词性的统计结果，我们直接对单个词语的特征和连续两个词语的特征都进行了。在训练 SVM 的过程中，由于可以找到的训练集有限，导致连续两个词语的效果并不好，所以，最后我们选用了单个词语的特征。以避免矩阵稀疏化，同样我们在处理训练集时，只选择了统计 CHI 值的前 200 个特征。

表 2　SVM 所选特征值

Table 2　Features to be considered for SVM

序号	类型	特征内容	描述
F1	情感词	情感词典词语个数	整理 HOWNET 情感词典后 8223 个情感词
F2	指示性动词表	指示性动词表中动词个数	所构建的 23 个指示性动词表中的动词
F3	词性	单一词性	中科院分词系统共 96 种词性
F4	双词性	两个连续词性组合	中科院分词系统共 96 种词性，9216 种双词性组合，取 CHI 值前 100 的组合
F5	词语	单个词语	中科院分词系统分词后，统计训练集中 CHI 值前 200 的词语

3.3 情感倾向性分析

情感分析目前主要使用 HowNet 和机器学习两种方法，我们首先对两种方法都进行了尝试，考虑了特征的不同组合情况，然后，我们给出本文使用的方法，下面是具体说明。

3.3.1 HowNet 方法

我们尝试了不同的实验，最后得到的结论是：HowNet 对于中文的相似度计算并不是很理想，特别是

对于情感此方面（而我们最主要又是需要这方面的），例如，该工具对于完全反义的两个情感词的相似度定义较高，这相当于认同它们在词性等方面是相似的，极不合常理，因此本文了暂时不考虑使用该方法。

3.3.2 机器学习方法

我们选用 SVM 进行机器学习。在前面完成了对观点句的提取后，现在可以进行训练/测试了。我们对不同的特征组和进行了测试：包括观点句测试、副词特征，标点符号特征等形成的向量的测试，SVM 方法除了在观点句测试阶段试用效果好之外，总体的测试结果不太理想。消极情感召回率值为 1，即全部分到了消极一方，这显然是不可能的也是不可行的。

我们分析认为：训练集太少是导致目前效果不好的主要原因所在。由于短时间内(评测时间表)无法生成大量训练集来训练 SVM 分类器，所以，经过认真考虑，本文我们将暂时不用此方法，而留到下一阶段考虑。

表 3　SVM 情感分析结果

Table 3　Sentiment Analysis based on SVM

qingan1=1	0	积极情感准确率	积极情感准确率	消极情感准确率	消极情感召回率	F 值
qingan1=1 and polarity='POS'	0	#DIV/0!	#DIV/0!			#DIV/0!
qingan1=-1 and polarity='NEG'	59			0.728395062	1	

3. 3. 3　本文采用的方法

通过分析我们所面临的问题，我们采用了与上面两种方法完全不同的方法，具体说明如下。

(1) 传统方法

在传统的方法中，有一种最为简单、直接的、无需训练集的方法，即直接统计积极词和消极词的个数的方法。当积极的词大于消极的词时就认为该句是积极的，反之则认为该词是消极的。但是该方法有很强的局限性，例如：否定词会将整个句子的情感倾向改变，但是否定词的位置和所修饰的对象也会有巨大的影响。例如：例 1:“我喜欢不高的人”。例 2:”我不喜欢高的人”。这两句中，“不”字这个否定词的修饰对象直接影响了整句的情感倾向，第一句修饰的是“高”这个形容词，但整个句子的情感倾向还是由“喜欢”这个词所决定的积极情感。第二句修饰的是“喜欢”这个词，“喜欢”是代表的积极情感，加上否定词，使整个句子成为了消极情感。同时，一个句子会有很多子句，而如果把子句作为一个句子处理，那么整体情感会不准确。

(2) 改进的算法

我们利用斯坦福的句法分析工具，将整个句子分成解成一棵树，在此基础上分析整个句子的情感极性，达到了比较好的效果。

该工具中将每一个子句都标记为一个以IP为节点的子树，我们因此可以对每一个接近叶子节点的IP（即最下层的IP）进行提取，获得最小的子句，达到分解子句的目的。完成子句分解后，进行统计工作，包括：统计积极词汇、消极词汇、以及否定词。在统计积极词汇和消极词汇时，使用之前所筛选的8223个情感词；否定词使用从训练集中归纳得到的一个列表，包含的词汇有：”不”, ”不然”, ”不行”, ”不要”, ”没”, ”没有”, ”无”, ”否”, ”非”,”不够”,”不可”,”未”,”绝非”, ”并非”。

判定算法如下：

Step1:一个子句若积极词比消极词多，否定词个数为偶数，则子句为积极，否定词个数为奇数，则子句为消极。

Step2: 一个子句若积极词比消极词少，否定词个数为偶数，则子句为消极，否定词为个数奇数，则子句为积极。

Step3: 子句若积极词与消极词相等，则子句为消极（这是因为微博上的评论，以消极的为主）。

Step4: 若积极子句个数比消极子句个数多，则整个句子为积极，否则，整个句子的情感属性都为消极。

利用上述方法，我们得到了比较好的测试结果，而且效率很高。准确率和召回率都维持在一个不错的水

平，F值也达到了0.588。由此，我们认为：在没有大量实例、而且算法对时间复杂度的要求很高的情况下，上面的方法不失为一个较好的选择　。

表 4　本文情感分析结果

Table 4　Outcomes from the Proposed Approach

Qingan2=1	29	积极情感准确率	积极情感准确率	消极情感准确率	消极情感召回率	F 值
Qingan2=1 and polarity='POS'	15	0.517241379	0.681818182			0.588235
Qingan2=-1 and polarity='NEG'	45			0.865384615	0.762711864	

4　实验结果

测试数据由NLP&CC 2012 主办方提供。具体而言，提供的评测数据来自腾讯微博[1]，全集包括 20 个话题，每个话题采集大约 1000 条微博，共约 20000 条微博。数据采用 XML格式，已经预先切分好句子，共 31675 句。利用本文的方法，对数据集进行观点句提取，并对所提取的观点句进行情感性倾向性分析，得到的结果如图 2 所示，结果同时可以从CCF网站上获取[2] (本参赛组编号为 42)。

观点句识别评测结果						
结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
42	0.783	0.338	0.472	0.792	0.337	0.452

情感倾向性判断评测结果						
结果编号	微平均			宏平均		
	正确率	召回率	F 值	正确率	召回率	F 值
42	0.824	0.279	0.417	0.802	0.28	0.404

图 **2**　中文计算会议（NLP&CC 2012）评测结果

Fig. 2 Results from NLP&CC 2012 (No. 42)

5　讨论

在评测方公布评测标注数据后[2]，由于有了更多的训练数据，因此我们选取了SVM方法，并对比了SVM和SVM+工具模板两种方法。具体为：从公布的20个话题的前10个话题中，随机抽取750个观点句和750个非观点句作为训练集，以后10个话题的数据作为测试数据，SVM方法得到的结果如表5所示, 表中列举了单一特征值的准确率，召回率，F值以及两个F值的方差。通过观测比较，组合效果较好的特征值，以达到更好的分类效果。

表 5　SVM 分类结果

Table 5　Results obtained from SVM Classifier

特征值	观点句提取数	观点句准确率	观点句召回率	非观点句准确率	非观点句召回率	观点句 F 值	非观点句 F 值	F 值方差

[1] http://baike.baidu.com/view/3264698.htm

[2] http://tcci.ccf.org.cn/conference/2012/pages/page04_evares.html (以初始版本数据为准，修订版数据出错问题已与主办方协商解决。)

F1	1157	0.701815039	0.746323529	0.533783784	0.478064	0.723385	0.504389	0.011989794
F2	1735	0.623054755	0.993566176	0.5	0.01059	0.765852	0.020741	0.138797671
F3	972	0.742798354	0.663602941	0.528957529	0.621785	0.700971	0.571627	0.004182443
F4	589	0.764006791	0.413602941	0.45	0.789713	0.536673	0.573311	0.000335599
F5	1348	0.675074184	0.836397059	0.556109726	0.337368	0.747126	0.419962	0.026759087
F2F5	1250	0.672	0.772058824	0.503006012	0.379728	0.718563	0.432759	0.020421018
F2F1	1144	0.703671329	0.739889706	0.532231405	0.487141	0.721326	0.508689	0.011303664
F2F3	962	0.738045738	0.652573529	0.519695044	0.618759	0.692683	0.564917	0.004081025
F2F4	661	0.750378215	0.455882353	0.455882353	0.750378	0.567181	0.567181	0
F1F2F4	783	0.767560664	0.552389706	0.495859213	0.72466	0.642437	0.588814	0.000718868
F1F2F3	982	0.743380855	0.670955882	0.533246415	0.618759	0.705314	0.572829	0.004388061
F1F2F5	786	0.745547074	0.538602941	0.478712357	0.697428	0.6254	0.567734	0.000831348
F1F2F3F5	933	**0.753483387**	**0.646139706**	0.528186275	0.652042	**0.695695**	0.583615	0.003140468
F1F2F3F4	857	0.758459743	0.597426471	0.50896861	0.686838	0.66838	0.584675	0.001751659
F1F2F4F5	757	0.77675033	0.540441176	0.495967742	0.744327	0.637398	0.595281	0.000443462
F1F2F3F4F5	840	0.760714286	0.587316176	0.506050605	0.695915	0.662863	0.585987	0.001477473

加入 Stanford Parser 工具模版标注后，我们预测模板标注与机器学习方法的结合能够提升分类效果，正如我们所料，结果为表 6 所示。引入 Stanford Parser 后，所找到的观点句数，观点句的准确率，召回率等值都有不同程度的提升，这说明两者结合是行之有效的。

表 6　SVM +Stanford Parser 分类结果

Table 6　Results Obtained from Combining SVM + Stanford Parser

特征值	观点句提取数	观点句准确率	观点句召回率	非观点句准确率	非观点句召回率	观点句 F 值	非观点句 F 值	F 值方差
F1	1165	0.703862661	0.753676471	0.54109589	0.478064	0.727918	0.507631	0.01213168
F2	1737	0.623488774	0.995404412	0.583333333	0.01059	0.766726	0.020802	0.139100387
F3	985	0.746192893	0.675551471	0.537958115	0.621785	0.709117	0.576842	0.004374177
F4	613	0.77324633	0.435661765	0.459507042	0.789713	0.557319	0.580968	0.000139819
F5	1350	0.675555556	0.838235294	0.558897243	0.337368	0.748154	0.420755	0.026797609
F2F5	1252	0.672523962	0.773897059	0.505030181	0.379728	0.719658	0.433506	0.020470752
F2F1	1155	0.706493506	0.75	0.542087542	0.487141	0.727597	0.513147	0.011497153
F2F3	972	0.740740741	0.661764706	0.526383526	0.618759	0.699029	0.568846	0.004236936
F2F4	666	0.752252252	0.460477941	0.457987073	0.750378	0.571266	0.568807	0.0000015108
F1F2F4	795	0.771069182	0.563419118	0.502096436	0.72466	0.651089	0.593189	0.000838098
F1F2F3	992	0.745967742	0.680147059	0.540290621	0.618759	0.711538	0.576869	0.004533977
F1F2F5	792	0.747474747	0.544117647	0.481713689	0.697428	0.629787	0.569839	0.000898438
F1F2F3F5	944	**0.756355932**	**0.65625**	0.535403727	**0.652042**	**0.702756**	0.587995	0.003292543
F1F2F3F4	871	0.762342135	0.610294118	0.517084282	0.686838	0.677897	0.589994	0.001931751
F1F2F4F5	763	0.778505898	0.545955882	0.498985801	0.744327	0.641815	0.59745	0.000492071
F1F2F3F4F5	853	0.760714286	0.587316176	0.506050605	0.695915	0.662863	0.585987	0.001477473

通过上述结果 (表 5)可以看出，当特征值采用 F1+F2+F3+F5 组合时，观点句 F 值与非观点句 F 值最高，其 F 值方差也非常理想。观点句准确率达到了 75.3%, 观点句召回率为 64. 6%。效果明显，从而说明了 SVM 方法的有效性。

加入 Stanford Parser 工具模版标注后(表 6)，情况有所改进，观点句准确率达到了 75. 6%, 观点句召回率为 65.6%, 此实验说明了结合两种方法的有效性。

存在的问题：

(1) 由于我们对评测方的观点句和积极情感的定义不是很清楚，导致人工选出的训练集有一定的误差，虽然召回率有明显的提高，但准确性较之评测方的结果，有所下降。

(2) 由于训练集样本太小，导致了SVM分类器的准确性不够高，同样由于数据大多由人工标注，衡量标准具有柔性，目前只能从评测方公布标准答案后，推知相关的标准。当然，这一切都是为了构建更大、更全面的训练集，提高准确率/召回率/F值/AUC等。

(3) 整个微博网络社区包含的话题过于宽泛，如果按照社区话题分类，在各个领域，比如：政治、经济、体育等，有针对性地分析、处理中文微博，结果会有很大的改观。

6 总结

本文探讨了中文微博观点句的提取以及对中文微博的情感倾向性分析。本文提出和实现的方法，在评测指标中，具有很高的准确率。

目前对于中文微博的研究还处于初步探索阶段，处理方法有待改进、提供。以我们在本文中的分析为例，选取特征值时，需要进一步仔细考量在第一阶段我们暂时忽略的特征。下一阶段的主要工作将会涉及：

(1) 关注微博特有的网络词汇和表情特征，充实第一阶段的特征集合。

(2) 第一阶段主要针对(微博)单句的处理。但是，我们也知道，以句子为单位进行微博信息处理和分析过于单一，如果能以每一条微博为单位，结合微博中句子的上下文关系，找出观点微博，应该更具应用价值。

(3) 引入新的模型(如条件随机场CRFs、隐马尔科夫模型HMM等)[10]或者数学统计方法（如KNN）[11-12]，或多模型学校方法，对分类方法进行改进，提高分类的精准度。

致谢：感谢 NLP&CC 2012 主办方为我们提供了这样的机会，感谢主办方老师们的辛勤劳动，感谢国家自然科学基金(61170192)资助本项目。

参考文献

[1] M.Q.Hu, B. Liu .Mining and Summarizing Customer Reviews. ACM SIGKDD 2004:168-1776

[2] Wilson, T , J. Wiebe,P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. 2005:347-354

[3] Long Jiang, Mo Yu, Ming Zhou. Target-dependent Twitter Sentiment Classification.2011.ACL

[4] 姚天防, 彭思威. 汉语主客观文本分类方法的研究.第三届全国信息检索与内容安全学术会议论文集.2007:117一123.

[5] 刘志明, 刘鲁. 基于机器学习的中文微博情感分类实证研究. 计算机工程与应用.2012,48(1):1-4.

[6] 张博. 基于SVM的中文观点句抽取[D].北京，北京邮电大学. 2011,3.

[7] 谢丽星,周明,孙茂松. 基于层次结构的多策略中文微博情感分析和特征抽取.中文信息学报.2012,1.

[8] 刘建华,张智雄. 基于Stanford Parser的实体间关系识别. DLIB&OSS 2009论文选登.2009,5.

[9] 单松巍,冯是聪，李晓明.几种典型特征选取方法在中文网页分类上的效果比较.计算机工程与应用,2003,39(22):146-148.

[10] 张华平. 基于多层隐马尔科夫模型的中文词法分析. 第41届ACL会议暨第二届SIGHAN研讨会,日本. 2003:63-70.

[11] 樊娜,安毅生，李慧贤. 基于K-近邻算法的文本情感分析方法研究.计算机工程与设计,2012,3.

[12] 张素智,孙培锋. 基于KSVM的网络评论情感分类研究.郑州轻工业学院学报，2011，6.

[13] VaPnikVN. The nature of statistieal learning theory. Springer一Verlag，New York，1995.

[14] 周德友. 基于HowNet的中文语义倾向性分析技术研究[D]. 沈阳，东北大学.2008.