

# 基于三类训练两类判别框架的主客观性句子识别<sup>1</sup>

王克 张春良 高晓兴\* 朱靖波

(东北大学信息学院计算机软件所自然语言处理实验室, 沈阳 110004)

(\* 东北大学软件学院, 沈阳 110004)

[wangke@ics.neu.edu.cn](mailto:wangke@ics.neu.edu.cn)

## 摘要

主观性句子的判别是文本观点倾向性分析的基础。本文以实词作为特征, 采用 Naïve Bayes 和最大熵分类模型进行主客观句子的判别。在公开的评测语料 MPQA 上实验结果显示: Naïve Bayes 相比于最大熵模型取得了较好的性能, 两者对主客观句子的判别具有互补性: 为了改善间接主观性句子的识别性能, 本文提出了三类训练两类判别的分类框架, 有效地改善了主客观句子的分析性能。

关键词: 文本观点倾向性分析; 直接主观性句子; 间接主观性句子; 分类模型;

## Classification of Subjective and Objective Sentences under the Framework of Three-class Training and Two-class Identification

Wang Ke, Zhang Chunliang, Gao Xiaoxing, Zhu Jingbo

(Natural Language Processing Lab of Northeastern University, Shenyang 110004;)

(Software College of Northeastern University, Shenyang 110004 )

[wangke@ics.neu.edu.cn](mailto:wangke@ics.neu.edu.cn)

**Abstract:** Identification of subjective sentences is the foundation of text opinion analysis. Adopting the substantive words as the feature, this paper identifies subjective sentences and objective sentences out of MPQA Opinion Corpus, an open test corpus, with two classifiers (Naïve Bayes and Maximum Entropy). The results reveal that Naïve Bayes outperforms Maximum Entropy in this study and these two classifiers are complementary in identifying subjective and objective sentences. To boost the performance of identifying the indirect subjective sentences, this paper offers a framework of three-class training and two-class identification, which turns out to improve the performance of classification.

**Key Words:** text opinion analysis, direct subjective sentence, indirect subjective sentence, classifier

## 前言

通过互联网 Web 方式如 blog 和 BBS, 人们可以了解其他人对某一话题的讨论, 也可以发表自己的观点评价, 例如针对“周老虎照片事件”和“Nokia 手机产品的评价”等, 由此自然形成了海量主观性文本数据库。目前传统的自然语言处理技术的应用(如信息检索和信息抽取)主要针对文本中包含的事实性知识(即客观性文本)的分析和处理。最近很多研究人员和公司开始关

---

本论文工作得到了国家教育部新世纪优秀人才计划(NCET-05-0287)和国家 863 高科技计划项目(2006AA01Z154)的资助。

注文本内容包含的非事实性知识(主观性文本)的识别和分析,比较典型的应用研究是文本观点倾向性分析(Text Opinion Analysis, TOA)。例如通过 TOA 技术,股票市场分析家可以分析了解老百姓对某一个产品或某一公司的普遍评价和信心如何;娱乐公司可以分析了解老百姓对新推出的电影评价如何。

为了实现文本主观性分析,首先需要识别表达用户观点倾向性的文本片段,即主观性句子集。本文的研究任务在于将主观性文本分成主观性句子(subjective sentences)和客观性句子(objective sentences),例如一个主观性句子“这件事情做得不对”;一个客观性句子“离北京 2008 年奥运会开幕式还有 157 天”。Riloff 等(2003)和 Pang 等(2002)采用一元特征(Unigram Feature)和词袋模型(Bag-of-words, BOW),结合机器学习技术进行实现主观性检测分析,用于电影评论分类任务,实验结果显示好于采用一些复杂不同的特征(Wiebe et al., 1999),如包括采用二元特征描述、是否是段落首句子等等。Hatzivassiloglou 和 Wiebe(2000)的研究工作显示形容词(adjectives)也是非常有用的特征,能够有效改善主观性检测分析性能。Riloff 等(2003)在研究工作中采用了名词(Nouns)作为特征,也取得较好效果。Riloff 等(2003)还采用了 Bootstrapping 技术从无标注文本中自动学习主观性名词(如 trust, benefit)和模式用于主观性句子的自动判别,同时实验结果显示结合其它语段特征(Wiebe et al., 1999)可以取得更好的主观性分类性能。除了利用基于词层特征以外,Riloff 和 Wiebe 研究了采用模板层主观性线索特征进行主观性分类任务。他们采用 Bootstrapping 技术自动学习一些表达主观性的模板,如“to satisfy <obj>”。然后构建了一个高准确率的分器对无标注文本进行自动标注主观性句子和客观性句子。实验结果显示获得较高的正确率,但是召回率较低。

在主观性句子中,人们经常采用一些特定类型的词汇和表达方式来表达他们的观点、态度和评价,如褒贬、支持还是反对。通常假设语言主观性线索的最小单元是单词,可以作为主观性分析模型的特征描述。首先本文比较分析了 Naïve Bayes 和最大熵模型对主观性句子的判别性能,其中使用名词、动词、形容词和副词等实词作为特征。实验结果显示 Naïve Bayes 取得了较好性能。非常有意思的是,Naïve Bayes 模型对主观性句子判别具有较好的召回率,而最大熵模型对客观性句子判别具有较好的召回率,实验结果显示两者具有一定的互补性。

前期相关研究工作对于主客观句子的判别问题看作两类分类任务,即主观性句子和客观性句子两类。事实上,人们有很多种方式表达观点(即主观性表示),如采用直接表达方式(direct)“不好/我不喜欢”;也可以采用间接表达方式(indirect)“他对我刚才的发言表现出皱眉”,实际上隐含着“他可能反对或者有点不喜欢”。在两类分类框架下,基于机器学习(如最大熵模型)的判别技术取得了一定效果,但实验结果显示间接表达方式的主观性句子大部分被错分为客观性句子,造成主观性句子的召回率较低。由于客观性句子不表达用户的观点,实际上文本主观倾向性分析主要依赖于对所包含的主观性句子的分析,由此错分主观性句子将严重影响后续文本观点倾向性的分析,有点类似于将“用户重要邮件”识别为“垃圾邮件”进行过滤的严重后果。为此,本文将主客观性句子判别看作三类分类问题:客观性句子、直接主观性句子和间接主观性句子,并提出一种基于三类训练、两类判别框架的主客观性句子分析技术。传统的基于两类的主客观性句子分析机制可以看作作为两类训练、两类判别框架。实验结果显示基于三类训练、两类判别框架有效地改善了主客观性句子的判别性能。

## 数据说明

本文采用的数据是在观点挖掘及情感分析领域中广泛使用的公开英文新闻语料库 Multi-Perspective Question Answering (MPQA) version 1.2<sup>2</sup>该语料共包含来自不同新闻机构及不同领域的 535 篇新闻文本, 以及文本的人工标注结果。其标注结果中包括片段的 ID、起始和结束的字节位置、标注标签及其属性。共有如下六种标注标签:

- GATE\_agent: 表明观点的持有者或者是观点中针对的客观对象, 多为人名、地名、机构名等名词短语及代词。
- GATE\_expressive-subjectivity: 非直接表达主观倾向的词、短语以及句子片段, 例如: fraud, daylight robbery 等。
- GATE\_direct-subjective: 直接表达主观倾向的句子成分或者是能够表达观点、倾向以及感情等的事件。
- GATE\_inside: 表明主观倾向和报道事件包含的范围, 一般为一个整句。
- GATE\_objective-speech-event: 标注表达观点、倾向以及感情所针对的客观事件。
- GATE\_split: 表明一个句子的结束。

实验中, 首先将语料中的人工标注还原到原文本得出带标注句子, 例如: The woman(GATE\_agent nested-source="w, patient") was admitted to the hospital on Saturday after complaining(GATE\_direct-subjective nested-source="w, patient" expression-intensity="medium" polarity="negative" intensity="medium") of severe joint pains.(GATE\_inside nested-source="w")(GATE\_split), 之后按照语料说明中的标准, 根据句子中所含有的标注成分共提取原文中 11104 个句子作为实验语料。并对语料进行类别标注, 构造实验所用数据集。

根据语料说明中的如下标准:

1. 只含有一个或者多个直接表达主观成分 (direct-subjective), 此成分的强度属性 (intensity) 不为“low”和“neutral”, 且不包含 insubstantial 属性。
2. 只含有一个或多个间接表达主观成分 (expressive-subjectivity) 且此成分的强度属性 (intensity) 不为“low”;

### ● 构造所用数据集 MPQA\_Two\_Classes

定义符合标准 1 或标准 2 的句子为主观性句子, 其它为客观性句子, 整理如下:

表 1 实验用数据 MPQA\_Two\_Classes 数据集

句子总数	客观性句子	主观性句子
11104	4974	6130

### ● 构造所用数据集 MPQA\_Three\_Classes

定义只符合标准 2 的句子为间接主观性句子, 符合标准 1 或者同时符合标准 1、2 的句子为直接主观性句子, 其它句子为客观性句子, 构造实验 2 所用数据集 MPQA\_Three\_Classes, 整理如下:

<sup>2</sup> <http://www.cs.pitt.edu/mpqa/databaserelease/mailto1.2.php>

表 2 实验用数据 MPQA\_Three\_Classes 数据集

句子总数	客观性句子	主观性句子	
		直接主观性句子	间接主观性句子
11104	4974	6006	124

提取出句子之后,利用词性标注工具 Rule\_based\_Tagger Version 1.14<sup>3</sup>对所有句子进行词性标注。根据标注选取名词、动词、形容词和副词作为特征进行分类实验。

## 分类模型

本文中,对主客观句子的判别被看作单标签文本分类问题。简单来说,文本分类任务可以定义为:根据文本内容赋予一个或多个预定义类别或主题类别。核心任务为:根据给定的训练数据,构造高性能的分类器,实现对新文本的自动分类过程<sup>[5]</sup>。

首先将实验数据中禁用词去掉,然后对数据进行词性标注,之后根据标注选用语料中的名词(NN)、动词(VB)、形容词(JJ)和副词(RB)作为特征而不再进行特征选取,采用在文本分类中常用和性能较好的有监督的机器学习方法来实现句子的主客观性分析。

### 1. Naïve Bayes 分类:

朴素贝叶斯(Naïve Bayes)分类器是贝叶斯学习方法中实用性很高的一种学习器,它基于贝叶斯假设<sup>[6]</sup>,在已知先验概率和条件概率的基础上得到后验概率的分类方法,使用这种方法可以给出一个句子属于一个特定类别的概率。在本实验的应用中,贝叶斯定理被表示成如下公式:

$$C^* \equiv \arg \max p(C_j) \prod_{t=1}^{|V|} p(w_t | C_j)^{n(w_t)}$$

式(1)中 $C^*$ 为类别标签, $p(C_j)$ 为类别的先验概率, $|V|$ 为特征词典大小, $w_t$ 是特征词典中的第 $t$ 个词, $n(w_t)$ 为词 $w_t$ 在给定句子中的频数,而 $p(w_t | C_j)$ 表示了 $w_t$ 属于类别 $C_j$ 的后验概率。

### 2. 最大熵(Maximum Entropy)分类

最大熵模型是在文本分类中广泛应用的性能较好的分类模型,其原理是:在已知部分知识的前提下,关于未知分布最合理的推断就是符合已知知识的熵值最大的概率分布,而对未知的情况不做任何主观假设。在这种情况下,概率分布最均匀,预测的风险最小。最大熵分类的原理表示为如下公式:

$$p(c | s) := \frac{1}{Z(s)} \exp\left(\sum_i \lambda_{i,c} F_{i,c}(s, c)\right)$$

其中 $Z(s)$ 是归一化因子, $\lambda_{i,c}$ 为模型参数, $F_{i,c}$ 为对于特征 $w_i$ 的特征-类别函数,定义如下:

<sup>3</sup> <http://www.cs.jhu.edu/~brill/>

$$F_{i,c}(s, c') := \begin{cases} 1, & n_i(s) > 0 \text{ and } c' = c \\ 0, & \text{otherwise} \end{cases}$$

其中  $n_i(s)$  表示特征  $w_i$  在  $s$  中出现的次数。最大熵模型通过迭代来训练模型参数  $\lambda_{i,c}$ 。

## 实验结果与分析

本文使用传统的召回率、准确率、Macro F1 并结合目前广泛使用的正确度 (accuracy) 值对实验结果进行评价, 计算公式如下:

$$\begin{aligned} \text{MacroP} &= \frac{1}{n} \sum_{j=1}^n P_j & \text{MacroR} &= \frac{1}{n} \sum_{j=1}^n R_j \\ \text{MacroF1} &= \frac{\text{MacroP} * \text{MacroR} * 2}{\text{MacroP} + \text{MacroR}} \\ \text{accuracy} &= \frac{\sum_{j=1}^n \text{correct}_j}{\sum_{j=1}^n \text{all}_j} \end{aligned}$$

其中,  $n$  是类别总数,  $P_j$  为第  $j$  类的准确率,  $R_j$  为第  $j$  类的召回率,  $\text{correct}_j$  第  $j$  类分类正确的样本数,  $\text{all}_j$  为第  $j$  类包含的所有样本数。在本文的实验中,  $n = 2$ , 一个样本为一个句子。

根据实验需要选取实验数据集, 在分类器上进行十次交叉检验, 每次随机选择 80% 作为训练数据, 余下的 20% 作为测试数据, 计算十次平均的 accuracy 和 MacroF1。

### 1. 两类训练、两类判别框架

本实验是在传统的两类训练、两类判别的框架下, 选用实验数据集 MPQA\_Two\_Classes 分别在 Naïve Bayes 和最大熵模型下进行实验, 实验结果如下:

表 3 实词作特征两类分类 NB 分类器的实验结果

	Objective Sentence	Subjective Sentence	Recall (%)	Average recall
Objective Sentence	449	531	45.860	70.18
Subjective Sentence	67	1149	94.509	
Precision (%)	87.065	68.415	F1 (%) : 73.77	
Average precision	77.74			
Accuracy (%)	72.797			

表 4 实词作特征两类分类 ME 分类器的实验结果

	Objective Sentence	Subjective Sentence	Recall (%)	Average recall
Objective Sentence	639	341	65.205	69.84
Subjective Sentence	310	906	74.486	
Precision (%)	67.327	72.660	F1 (%) :	
Average precision	69.99		69.91	
Accuracy (%)	70.344			

从表 2 和表 3 中可以看出, NB 的分类结果较 ME 的分类结果 accuracy 值提高了近 2.5%, 而 F1 值提高了近 5%, 充分说明在此任务中 NB 的分类性能要比 ME 有显著提高。同时发现, 客观性句子的召回率要提高近 20%, 准确率降低 20%; 而 ME 分类模型中主观性句子的召回率与 NB 分类模型相比要降低 20.1%, 准确率提高 4%, 说明在此任务中, NB 分类模型对于主观性句子的识别要好于 ME 模型, 而 ME 模型对于客观性句子的识别与 NB 相比具有较好的性能, 可以说两者在主客观句子的判别问题上具有一定的互补性。

## 2. 三类训练, 两类判别框架

由于在传统的两类训练、两类判别的框架下, 一部分间接主观性句子被误判为客观性句子, 于是本文提出用三类来训练分类器, 但是任务中只需要将句子识别为主观和客观两类即可, 本文的判别将分类结果中的间接主观性句子和直接主观性句子集合成为主观性句子集, 采用两类结果来判别, 因此提出了三类训练、两类判别的框架。

实验中, 使用数据集 MPQA\_Three\_Classes 在两类训练、两类判别框架下性能较好的 Naïve Bayes 分类模型进行分类, 得出实验结果如下:

表 5 三类分类 NB 分类器的结果分析

	Objective Sentence	Subjective Sentence	Recall (%)	Average recall
Objective Sentence	463	517	47.24	70.79
Subjective Sentence	69	1147	94.33	
Precision (%)	87.03	68.93	F1 (%) :	
Average precision	77.98		74.21	
Accuracy (%)	73.32			

对比上表和表 3 可以看出, 在三类训练、两类判别的框架下, 相比两类分类、两类判别的框架, 分类结果的 accuracy 值提高了近 0.5%, F1 值同样提高了近 0.5%, 说明这种框架下系统的性能相比传统分类框架是所有提高的。而性能的提高是因为在此框架下, 不存在间接主观性句子被错分为客观性句子的现象, 且降低客观性句子被错分为主观性句子的概率近 2%, 使得分类的结果能够更好地区分主客观句子, 提高分类结果的 F1 值。由于实验数据是新闻语料, 使得性能提升不是很显著, 但是在实际应用中如客户对于某个事件 (尤其是一些敏感事件) 的态度, 这种间接主观性句子是大量存在的, 能够正确判别这类句子, 会显著提高主客观句子识别的性能。

## 相关讨论

从实验中发现，简单的实词对于主客观性句子识别并不具有很好的效果，因为实词不具有明显的主观倾向性特征。本文尝试利用观点识别工具 *opinionfinder*<sup>4</sup> 中提供的主观性词表中的词语作为特征进行训练，但是实验结果表明对于此任务，该特征的性能很差，客观句子的召回率只有 8% 左右。通过分析数据发现由于主观词表中的词在客观句子中的分布非常稀疏，近 1/3 的客观性句子完全为空白，近 1/10 的客观性句子只包括一个特征词，如此稀疏的数据分布对于分类性能影响是非常大的。

另外，在实验中还发现一点，间接主观性句子包含的实词表达主观性的能力较弱，造成经常被错分为客观性句子。本文认为，句子的语义信息和句法结构信息对于正确判别间接主观性句子有重要的作用，因此在下一步的工作中，将引入其他特征，综合实词、主观相关词表、句法结构以及语义信息等多种特征在间接主观性句子的识别上取得较好的性能。此外，还将针对 Naïve Bayes 和最大熵模型分类结果具有互补性这一特点，对两者的结果进行整合，以提高判别的准确性，更好地解决句子主客观性分析问题。

## 结论

主观性句子的判别是文本倾向性分析的基础，能否正确识别主观性句子对于文本倾向性分析有很重要的作用。针对主观性的判别问题，本文以实词作为特征，采用 Naïve Bayes 和最大熵分类模型在公开的评测语料 MPQA 上进行主客观句子的判别。实验结果显示：当采用句子中的实词（名词、动词、形容词和副词）作为特征对句子进行主客观分类的时候，Naïve Bayes 的分类性能相比于最大熵模型分类效果有显著的提高，且在此任务中，两个模型具有一定的互补性；在传统的两类训练、两类判别框架的基础上，引入间接主观性句子的概念，提出三类训练、两类判别的框架。实验结果证明，三类训练、两类判别的框架下系统能够更好地对主客观句子进行识别。

## 参考文献

- [1] Hatzivassiloglou, V. and McKeown, K. Prediction the semantic orientation of adjectives. In Proc. Of ACL-07, 2007.
- [2] Riloff, E. and Wiebe, J. Learning extraction patterns for subjective expression. In Proc. Of EMNLP03, 2003.
- [3] Riloff, E., Wiebe, J., and Wilson, T. Learning subjective nouns using extraction pattern Bootstrapping. In Proc. of CoNLL-03, 2003.
- [4] Pang, B. L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proc. of EMNLP02, 2002
- [5] 朱靖波, 王会珍, 张希娟. 面向文本分类的混淆类判别技术. 软件学报
- [6] 张希娟, 王会珍, 朱靖波. 基于朴素贝叶斯的文本分类 小型微型计算机系统. Vol. 27, pp. 369-370. 2006.
- [7] 唐慧丰, 谭松波, 程学旗. 基于监督学习的中文情感分类技术比较研究. 中文信息学报 Vol 21, No 6 p88-94. 2007
- [8] 徐军, 丁宇新, 王晓龙. 使用机器学习方法进行新闻的情感自动分类. 中文信息学报 Vol 21, No 6 p95-100. 2007

<sup>4</sup> <http://www.cs.pitt.edu/mpqa/opinionfinderrelease/>