汉语意见型主观性语句类型分析*

黄高辉　姚天昉　刘全升

上海交通大学计算机科学与工程系 上海 200240

fangao2620@163.com　yao-tf@cs.sjtu.edu.cn　xubylin@sjtu.edu.cn

摘要：目前，意见挖掘已经成为文本挖掘的一个热门研究方向，其主要研究对象是意见型主观性语句。本文首先介绍了汉语意见型主观性语句的定义和特点，并依据三种分类标准，即主题和情感的形式、数量以及对应关系，对汉语意见型主观性语句的类型进行分类。然后本文结合例句，分别从词汇层、句法层和语义层对汉语意见型主观性语句的特征进行分析归纳，并提出了一些需要注意的问题。

关键词：意见型主观性语句，词汇层，句法层，语义层

The Analysis for Chinese Opinioned-subjective Sentence

Huang Gaohui　Yao Tianfang　Liu Quansheng

Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240

fangao2620@163.com　yao-tf@cs.sjtu.edu.cn　xubylin@sjtu.edu.cn

Abstract: Nowadays, Opinion Mining has become a hot research topic in Text Mining. Its main research object is opinioned-subjective sentence. In this paper, firstly the definition and main features of opinioned-subjective sentence are introduced. And then we present three classification methods for opinioned-subjective sentence, which are based on the form, number and relation of topic and sentiment respectively. Then with some examples, we analyze the features of opinioned-subjective sentence in lexical, syntactic and semantic level, and also we propose some questions needed to pay attention.

Keyword: opinioned-subjective sentence, lexical level, syntactic level, semantic level

1. 概述

随着电子商务的飞速发展，网上商品交易成为一种普遍现象，越来越多的消费者通过网络购买商品，并通过网络对其发表评论。这些反馈信息对于厂家和潜在的消费者都有着重要的参考价值。但是，目前网络上信息数量急剧膨胀，使得客户很难对产品有一个正确的认识，商家也很难有效地追踪客户的反馈信息。同时，这些评论信息往往具有明显的主观色彩，以往的基于客观性文本的挖掘技术已经开始显得不具有针对性。面对这样的现实问题，一种新的挖掘领域——意见挖掘[1]孕育而生。利用意见挖掘技术，我们可以有效地从客户反馈信息中挖掘出客户对产品的情感倾向——支持还是反对，赞美还是批评等，这在电子商务领域中已经凸显出巨大的应用价值。

意见挖掘技术作为一种新颖的语言技术，不仅可以运用于自然语言接口、自然语言生成等方面，还可以应用于现实生活中的许多方面，如电子商务、电子学习、信息监控、民意调查等[1]。国外在这方面的研究起步较早，也产生了一些应用系统。比如美国伊利诺斯大学 Liu 等人开发的系统 Opinion Observer 可以处理网上在线顾客产品评价，对相关产品各种特征的优缺点进行统计，并采用可视化方式对产品特征的综合质量进行比较[2]。国内对于汉语文本意见挖掘的研究起

* 国家自然科学基金项目（60773087）

步较晚。如香港城市大学 Yuan 等人对汉语极性词地自动获取进行了研究[3]；国立台湾大学 Lun-Wei Ku 等人利用 NTCIR 的数据构建了一个语料库，并且基于该语料库提出了汉语文本意见抽取的各种算法[4]。

虽然国内外对于意见挖掘理论和应用研究已取得了一定的成果，但都存在着一些尚未解决的问题。特别是汉语意见挖掘理论的研究起步较晚，还没有出现比较完善的应用系统。意见挖掘的主要研究对象是意见型主观性语句。因此，很有必要首先从语言现象的角度对汉语意见型主观性语句的类型进行分析，为以后更加深入地研究汉语意见挖掘打下基础。

2. 汉语意见型主观性语句

主观性文本，是指对于非事实进行描述，基于断言或评论的文本。主观性文本可以分为非意见型主观性文本和意见型主观性文本。非意见型主观性文本是指不明显带有情感倾向的主观性文本。而所谓意见型主观性文本则是指明显含有个人、群体或者组织等的意见、态度和情感倾向的主观性文本，这种文本往往含有许多意见型主观性语句。下面将重点介绍意见型主观性语句的定义、特点和主要类型。

2.1 意见型主观性语句的定义

目前，对于“意见”还没有一个统一的定义。一般情况下，采用南加州大学 Kim 和 Hovy 的定义[5]，即一个意见是由持有者、主题、陈述和情感构成的四元组（也称为四个意见元素）。其中，意见持有者发表针对某主题的陈述，并且通常具有情感倾向。

结合 Kim 和 Hovy 的定义，我们认为一般情况下，意见型主观性语句至少要具有主题和情感，否则称为不完整的意见型语句，比如有些句子只含有情感，但没有显式给出主题。

2.2 意见型主观性语句的特点

首先，意见型主观性语句表达了个人或者团体对某个主题的意见和情感倾向，具有主观性。也就是说，说话人在说出一段话的同时表明自己对这段话的立场、态度和情感，从而在话语中留下自我的印记。因此，它的表达方式（包括语言和句式）往往是非规范的。它往往使用具有感情色彩的词汇来表达对事物的情感倾向，包括显式情感词和隐式情感词。这些情感词具有褒贬性，并且其强度是可以比较量化的。同时，与情感对应的主题有可能是显式主题，也有可能是隐式主题。表 1 总结了意见型主观性的主要特点。

意见型主观性语句主要特点	表现形式
主观性	说话者表现自我的印记
情感倾向	显式情感或者隐式情感
表达主题	显式主题或者隐式主题
非规范性	非规范性语言或者非规范性句式

表 1 意见型主观性语句的主要特点

2.3 意见型主观性语句的类型

在 2.2 节中分析了意见型主观性语句的特点，这些特点决定了意见型主观性语句的类型多种多样，存在多种分类标准。首先，可以根据情感和主题（包括子主题）的数量，将意见型主观性语句划分为单主题单情感、单主题多情感、多主题单情感、多主题多情感四种类型。

(1). 单主题单情感

单主题单情感表明一个句子只有一个情感和主题，这是最基本的类型。比如句子“我非常喜欢这辆车。”中只含有一个主题“车”和一个情感“喜欢”。

(2). 单主题多情感

单主题多情感表明一个句子中只有一个主题，但是有多个情感，这些情感可以直接修饰主题，也有可能修饰隐式主题。比如句子“保时捷动感大气，非常快。”中，主题为“保时捷”，情感为“动感”、“大气”和“快”，其中“动感”和“大气”直接修饰主题“保时捷”，“快”则修饰隐式主题“速度”。有时情感的贬褒性可能是相反的，例如句子“保时捷动感大气，但是价格实在太贵。”中，“动感大气”为褒义，“太贵”为贬义。

(3). 多主题单情感

一个句子中可能有多个主题，但只有一个情感，此时这个情感可能修饰所有主题，也有可能只修饰某些主题。比如句子“宝马和奔驰都是成功人士的最爱。”中情感“爱”修饰两个主题。多主题之间的关系有时可能是主题和子主题的关系，例如句子“奔驰的外形太漂亮了。”中，主题“外形”是主题“奔驰”的子主题。

(4). 多主题多情感

多主题多情感即一个句子有多个主题和多个情感，这种类型是最常见的，它的子句可能属于前三种类型。比如句子“奥迪和大众很实惠，前者质量好，后者性价比高。”中，主题为“奥迪”，“大众”，“质量”和“性价比”，情感为“实惠”，“好”和“高”，并且第一个子句属于多主题单情感类型，后两个子句属于单主题单情感类型。

同时，我们也可以根据主题和情感之间的对应关系，将意见型主观性语句划分为主题情感一对一、一对多、多对一和多对多四种类型，分别如图 1、图 2、图 3、图 4 所示。

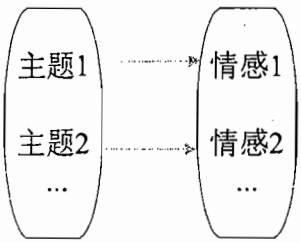

图 1 主题情感一对一

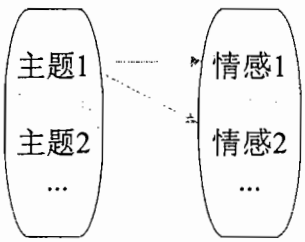

图 2 主题情感一对多

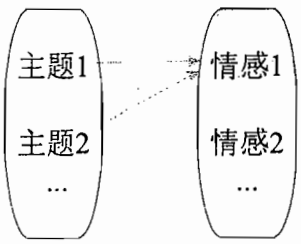

图 3 主题情感多对一

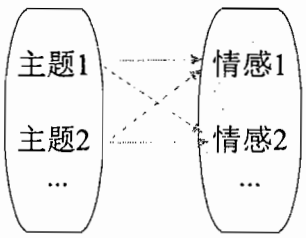

图 4 主题情感多对多

由图 1~图 4 可知，这种分类标准不再对句子中主题和情感的数量做任何假设，而是依据句子中主题元素集合与情感元素集合的映射关系进行分类。

最后，我们注意到，有些意见型主观性语句只有情感，却没有显式给出主题，此时需要结合上下文环境才可以确定主题。如句子“很好，很漂亮”中，只有情感，没有主题。一般情况下，意见型主观性语句含有情感词，但是也有些句子表达了情感，却没有使用情感词，这种情感称为隐式情感。比如句子“吉利车加满油也只能跑一天，简直就是油老虎”，表达了一种对吉利车耗油太大的不满情绪，但是句中没有情感词。因此，我们可以根据主题和情感的形式，即显式和隐式，将意见型主观性语句划分为：显式主题显式情感，显式主题隐式情感，隐式主题显式情感以及隐式主题隐式情感四种类型。其中，隐式主题和隐式情感往往需要结合上下文的信息才能确定，处理起来比较困难。

3. 汉语意见型主观性语句示例

第二章从宏观上分析了汉语意见型主观性语句的特点和类型。这一章将从三个具体的层次，即词汇层、句法层和语义层，对意见型主观性语句进行分析归纳，并给出一些示例。

3.1 词汇层

词汇是构成句子的基础，因此首先从词汇层分析意见型主观性语句的特点。通过观察和分析一些意见型主观性语句，我们提出以下几个类型特征。

1. 情感词

情感词是一类带有情感倾向的词，用于表达说话者的情感态度，是判断意见型语句的最重要特征。情感词的词性包括形容词、动词和名词。形容词是对事物属性的一种描述，能十分有效地表达说话人的情感，比如“好”与“坏”，“高”与“低”等。有些动词和形容词一样具有主观性，也能强烈地传达情感，比如“喜欢”与“讨厌”等。有些名词也带有情感色彩，比如“英雄”，“汉奸”等，由于名词较形容词和动词主观性最弱，因此通常它表达的情感强度也较低。

例 1：“影星郭富城的这辆法拉利实在是太漂亮了，我好喜欢，它就是跑车中的精灵。”在这个例句中，形容词“漂亮”、动词“喜欢”以及名词“精灵”都表达了一种肯定的情感，它们都属于情感词。

2. 非规范性语言

网络上的评论等意见型主观性语句常常含有大量的非规范性语言。当使用非规范性语言时，往往可能会伴随着一种或悲或喜，或褒或贬的情感。非规范性语言可以分为非规范词和非规范符号。非规范词通常包括中文谐音，拼音缩写，英文缩写，特殊数字等。非规范符号包括非规范标点符号和表情符号等。

例 2：“555, 我买了一辆尼桑，结果性能很差，没几天就出问题了，很郁闷啊!!!!!!真是一辆破车!!!”在这个例句中，特殊数字“555”是一种非规范词，表达了一种悲伤的情感，非规范标点符号“!!!!!!”表达了一种强烈的失望情绪，句末的非规范标点符号“!!!”则表达了一种愤怒的情感。

3. 否定词和强调词

在意见型主观性语句中，否定词和强调词经常作为情感词的修饰成分，与情感词同时出现。

否定词能够使情感词的极性反向，一般是副词或动词，如“不”，“不能”，“没有”等；强调词则用于加强或者削弱情感词的情感强度，一般是副词，如“最”，“更”，“很”等。

例3：“我不喜欢日本车，不是很耐用。”这个例句中的前半句中，否定词“不”修饰情感词“喜欢”，使得最后的情感倾向为否定；后半句中，情感词“耐用”前面则同时有否定词“不是”和强调词“很”，最终的情感倾向也为否定。

4. 命名实体

由于意见型主观性语句通常是领域相关的，因此在句子中往往包含许多领域相关的命名实体。命名实体是指在文本中具有特殊定义的实体，主要包括人名、地名、机构名、专有名词、时间、数量短语等。在意见型主观性语句中，主要的命名实体为领域专有名词，比如在汽车领域中，命名实体为一些汽车的品牌，如“凯迪拉克”，以及一些专业术语等。

5. 其它一般主观性语句的特征

当然，意见型主观性语句还具有一般主观性语句的一些特征。比如意见型主观性语句里可能会含有一些建议性动词，如“认为”，“觉得”等；在句末包含一些感叹词，或者在句末增加一些特殊的标点符号，如感叹号，问号等，表明一种特殊的情感，并且这种情感往往属于隐式情感。

3.2 句法层

3.2.1 意见元素依存关系

根据Kim和Hovy对“意见”的定义，我们在句子层上将意见型主观性语句分割为陈述、意见持有者、主题、情感等部分。关于陈述定界的研究目前主要有三种思路：粗分法、细分法和基于模板的方法[6]。在这里，我们采用粗分法，但是约定一个陈述里只有一个持有者。

例4：“专家们都说国产车质量差，没有市场前景，我却要说国产车终究会步入正轨。”在例4中有两个意见持有者，即“专家们”和“我”，因此它可以分为两个陈述，即前两个子句作为一个陈述，最后一个子句也单独作为一个陈述。

在一个陈述里，有意见持有者、主题和情感。持有者和主题可以是单个词汇，也可以是复合词。情感一般由情感词和情感修饰部分（否定词和强调词）组成。这些意见元素之间存在着内在的联系，即意见持有者针对某主题发表了具有情感的陈述。这种内在联系在句法层上体现为持有者和主题、持有者和情感以及主题和情感之间的依存关系。利用哈尔滨工业大学信息检索实验室依存句法分析器（http://ir.hit.edu.cn/demo/ltp/），我们可以很容易得到这些依存关系。

例5：“我很喜欢这辆车，性价比很高。”在例5中，意见持有者是“我”，主题“车”对应的情感为“喜欢”，主题“性价比”对应的情感为“很高”。其句法依存关系树如图5所示。

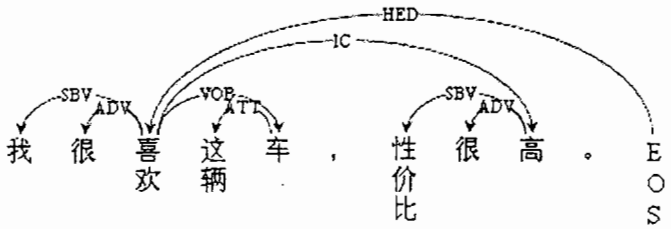

图5 句法依存关系树

由图 5 可以看出，持有者“我”和情感“喜欢”的依存关系为 SBV（主谓关系），主题“车”和情感“喜欢”的依存关系为 VOB（动宾关系），主题“性价比”和情感“很高”的依存关系为 SBV。利用这些依存关系，我们可以确定意见持有者、主题和情感之间的对应关系，为以后的语义分析提供依据。当然，有时意见元素之间也可能不存在直接的依存关系。

3.2.2 非规范性句型

在 2.2 节中已经提到过汉语意见型主观性语句具有非规范性，表达方式比较随意，常常采用非规范性句型，比如成分省略等。同时，为了突出表达某些特殊的情感，也经常会使用一些特殊的句式，比如句子成分倒置等。

例 6：“太有震撼力了，保时捷！”在例 6 中，主语“保时捷”倒置，放在句末，突出表达一种肯定的情感。此外，意见型主观性语句主要来源于网络，表达随意，常常存在语法错误。例如下面这种句式经常在网络上出现。

例 7：“通过这次车展，使我对法拉利的奢侈天价有了一个感性认识。”这句话采用了“通过…使…”的句式，存在语法错误。意见型主观性语句的这些不规范性表达方式给句子的分析带来了困难，需要在分析前做一些规范化预处理。

3.3 语义层

意见型主观性语句的语义层包括词汇语义、短语语义以及句义。词汇语义又可以细分为情感词语义、一般词汇语义以及领域术语语义等。一般词汇里，否定词表明了一种极性取反的语义，强调词则表达了一种强度加强或者减弱的语义。情感词的语义就是特定的情感倾向，是支持还是反对，肯定还是否定。需要注意的是，有一类词汇，在一般情况下并不具有情感，但当它在修饰某些主题时，出现了情感性，并且它的语义倾向是变化的，即在修饰某些主题时极性为正，而在修饰另外一些主题时极性则为负。还有一类词汇与领域密切相关，其极性也是动态的[1]，即在通常情况下没有极性，只有在修饰某些主题时才具有极性。

例 8：德国车价格高，性价比也高。

例 9：这辆车颜色黑，价格更黑。

在例 8 中，形容词“高”在通常情况下不具有情感。但在修饰主题“价格”和“性价比”时，出现了情感性，前者极性为负，后者极性为正，即极性是动态的。在例 9 中，形容词“黑”在通常情况下也不具有情感，但是在修饰主题“价格”时，它出现了情感性，并具有负极性，它的极性也是动态的。

词汇语义是构成短语语义的基础。短语包括命名实体、意见元素等。持有者和主题由一般词汇或者领域术语组成。情感元素则由情感词、否定词和强调词构成，这三种词汇的语义最终决定了整个情感元素的极性和强度，确定了情感元素的语义。

句义是指整个句子的语义，它由所有意见元素的语义以及意见元素之间的相互依存关系确定。句义分析的目的是为了确定整个意见型主观性语句的意见倾向，包括总极性与总强度。由 2.3 节可知，一个句子可能包含许多主题和情感元素，这些情感的极性可能为正，也可能为负。因此，最终整个句子的极性可能为正，也可能为负，甚至可能为中性。比如：

例 10：“奥迪 A6 的价格即不高也不低。”在例 10 中，有两个情感元素，情感“不高”的极性为正，“不低”的极性为负。但是，例 10 的总体极性为中性。

Gamon 等人曾提出[7]，很多句子是无法被归类为褒义或者贬义范畴的，它们或者没有极性，或者同时表现了褒义和贬义，或者由于缺乏上下文环境或领域知识而无法判断极性。这个问题在汉语中尤其明显，因为汉语句子中经常出现连续逗号的现象，同一个句子中出现多个主题或者多个意见倾向，如果只通过简单的正负叠加来计算整句的极性，则必然存在许多问题。因此，对于汉语意见型主观性语句的句义还需要进一步深入地研究。

4. 结束语

近年来，对于描述非事实的主观性文本处理方面的研究十分活跃，其中对意见型主观性文本进行意见挖掘是主要的研究方向之一。在这方面，汉语意见挖掘的研究起步较晚，还没有一套完整的理论。因此，本文从语言学的角度，分析和归纳了汉语意见型主观性语句的特点和类型，希望对以后的进一步深入研究有所帮助。

根据 Kim 和 Hovy 对“意见”的定义，在意见型主观性语句中，意见持有者对某个主题发表了具有情感倾向的评论。因此它具有主观性和非规范性。汉语意见型主观性语句的类型多种多样，可以根据主题和情感的形式（显式或者隐式）进行分类，也可以根据主题和情感的数量或者对应关系进行划分。最后，本文重点从词汇层、句法层和语义层对意见型主观性语句做了详细的分析。词汇层主要包括情感词、非规范性语言等；在句法层重点分析了意见元素的依存关系和非规范性句型；在语义层则分析了词汇语义、短语语义和句义。本文在对意见型主观性语句的类型特点进行分析归纳时，都给出了详细的例句。

感谢：此研究工作得到了国家自然科学基金会和上海交通大学中德语言技术联合实验室资助。此外，哈尔滨工业大学信息检索研究室所开发的句法分析器为我们的研究提供了帮助。笔者在此深表谢意。

参考文献

[1] 姚天昉，聂青阳，李建超，李林琳，娄德成，陈珂，付宇. 一个用于汉语汽车评论的意见挖掘系统. 见：曹右琦，孙茂松主编，中文信息处理前沿进展-中国中文信息学会成立二十五周年学术年会论文集. 清华大学出版社，北京，2006 年 11 月.

[2] B.Liu, M. Hu, and J.Cheng.2005.Opinion observer: analyzing and comparing opinions on the Web. In Proc.of WWW '05, the 14th international conference on World Wide Web, pages 342-351. Chiba, Japan.

[3] R.Yuan et al. 2004. Morpheme-based Derivation of Bipolar Semantic Orientation of Chinese Words. In Proc.of the 20th International Conference on Computational Linguistics (COLING-2004), pages 1008-1014. Switzerland.

[4] Lun-Wei Ku, Tung-Ho Wu, Li-Yang Lee and Hsin-His Chen. Construction of an Evaluation Corpus for Opinion Extraction. In Proc. of NTCIR-5 Workshop Meeting,December 6-9,2005,Tokyo,Japan,2005,pages 513-520.

[5] Kim S-M, Hovy E. Determining the Sentiment of Opinions. In Proc. of COLING-04: The Conference on Computational Linguistics (COLING-2004), 1367-1373, Geneva, Switzerland, Aug. 23-27, 2004.

[6] Tetsuya Nasukawa and Jeonghee Yi.2003.Sentiment analysis: Capturing favorability using natural language process. .In Proc. of the 2nd International Conference on Knowledge Capture(K-CAP 2003),pp.70-77,Sanibel Island,Florida.

[7] Gamon,M.,A.Aue,S.Corston-Oliver and E.Ringger. Pulse: Mining Customer Opinions from Free Text. Lecture Notes in Computer Science,Vol.3646:121-132,Springer Verlag,IDA 2005.