

文章编号: 1003-0077(2010)01-0084-05

评价对象抽取及其倾向性分析

刘鸿宇, 赵妍妍, 秦 兵, 刘 挺

(哈尔滨工业大学 计算机学院 信息检索研究室, 黑龙江 哈尔滨 150001)

摘 要: 情感分析近年来已经成为自然语言处理领域的热点问题, 该文对情感分析中的两项关键技术——评价对象抽取和倾向性判断进行了深入研究。在评价对象抽取阶段, 首先使用句法分析结果获取候选评价对象, 继而结合基于网络挖掘的 PMI 算法和名词剪枝算法对候选评价对象进行筛选。在倾向性判断阶段, 通过分析情感句句型, 归纳相应的分析规则, 使用无指导的方法完成评价对象在情感句中的倾向性判断。该系统参加了 COAE2008 任务三的评测, 取得了较好成绩。

关键词: 计算机应用; 中文信息处理; 情感分析; 评价对象; 倾向性判断; 句法分析

中图分类号: TP391 **文献标识码:** A

Comment Target Extraction and Sentiment Classification

LIU Hongyu, ZHAO Yanyan, QIN Bing, LIU Ting

(Information Retrieval Laboratory, School of Computer Science and Technology,
Harbin Institute of Technology, Harbin, Heilongjiang 150001, China)

Abstract The research on sentiment analysis is a hot issue in natural language processing. This paper makes an intensive study of the two stages of sentiment analysis: the comment target extraction and the corresponding sentiment classification. For the first task, we use the syntactic analysis to obtain the candidates and then combine PMI based on web mining and NN filtering algorithm to decide the targets. For the second task, we design certain heuristic rules by analyzing subjective sentences and then apply these rules to predict the orientation of opinion in the sentences. This method performs well in Task Three of the COAE2008 i.

Key words: computer application; Chinese information processing; sentiment classification; target; orientation judgment; syntactic analysis

1 引言

情感分析主要是针对主观性文本单元自动获取有价值的意见信息, 是一个新颖且非常有应用价值的研究课题。情感分析技术可被广泛应用于多种自然语言处理问题中, 如: 问答系统^[8], 信息抽取系统^[9]等。同时, 情感分析也产生了许多有挑战性的

相关子方向。例如: 文本主客观分析, 旨在识别文本单元的主客观性; 情感分类, 旨在识别主观文本单元是褒义、中性还是贬义。

本文致力于研究情感句中的评价对象抽取及其情感倾向性判断任务。该任务可分为两个主要阶段: (1)自动识别情感句中的评价对象; (2)判别情感句中评价对象的情感倾向性。例如: 对于某一评论“这款相机的电池寿命很长, 但价格太高。”, 系统首先识

收稿日期: 2009-05-26 定稿日期: 2009-11-06

基金项目: 国家自然科学基金资助项目(60675034); 863 计划探索类专题资助项目(2008AA01Z144); 国家自然科学基金青年基金资助项目(60803093)

作者简介: 刘鸿宇(1986—), 男, 硕士生, 主要研究方向为情感分析; 赵妍妍(1983—), 女, 博士生, 主要研究方向为情感分析; 秦兵(1968—), 女, 教授, 主要研究方向为信息抽取, 多文档文摘。

别出评论中的评价对象(如:“电池寿命”和“价格”),然后结合修饰评价对象的评价词语(如:长,高)给出两个评价对象的相应情感倾向性,即“电池寿命—长”→“电池寿命,褒义”,“价格—高”→“价格,贬义”。

关于评价对象的获取,已有的方法主要可以归为两类:人工构建^[3]和关联规则挖掘^[4-5],人工构建的缺点在于需要大量人力,且可移植性较差;关联规则挖掘的缺点在于没有充分考虑短语评价对象的结构特征以及评价对象的领域相关性,会产生很多噪声。

关于评价对象的倾向性判断,已有的工作包含有指导和无指导两类方法。关于有指导方法, Kim 和 Hovy 使用词、位置以及情感词三类特征来对情感句进行分类^[6];赵军等人使用 CRF 和冗余标签对句子序列进行情感倾向性标注^[7],有指导方法的缺陷在于需要人工标注大量语料,不利于领域切换。无指导方法目前主要是基于句法规则的方法^[3,8],

此类方法的优点在于能够准确的描述情感词和评价对象之间的修饰关系,缺点在于各种修饰关系需要人工统计,导致召回率不高。Minqing Hu 和 Bing Liu 基于句子的情感词以及上下文信息来判定倾向性^[8],但他们的的方法主要处理英文,系统中的部分技术无法直接向中文移植,同时有以下两类问题没有很好的解决,一是句子中褒、贬情感词数目相同,二是含有上下文相关情感词的句子。

鉴于以上方法的不足,本文使用无指导方法进行评价对象抽取和倾向性分析。评价对象抽取上,使用基于句法分析技术的评价对象抽取方法,同时加入三种过滤技术。在倾向性分析阶段,将情感句分为四类,采用分治策略对各类别情感句制定相应的判定规则,从而完成评价对象的倾向性判定。该方法在第一届中文倾向性分析评测中取得了较好的成绩。图 1 给出了系统框架图。

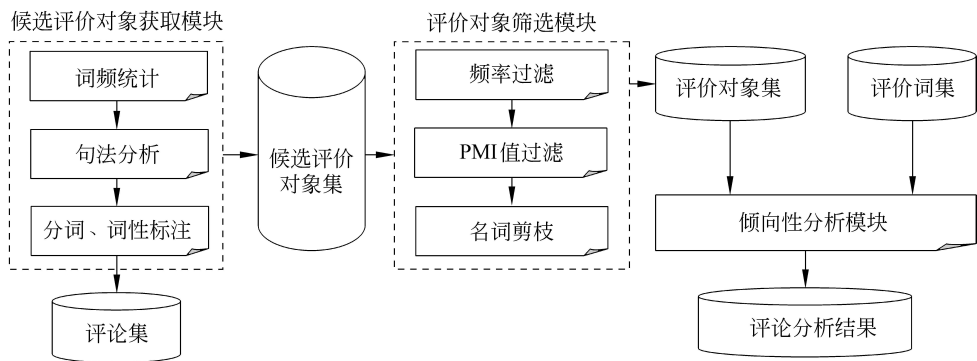


图 1 系统框架图

2 基于句法路径的评价对象抽取

本文使用基于句法分析的评价对象抽取技术。对于给定语料,首先对其进行分词、词性标注以及句法分析等处理,然后提取其中的名词(NN)和名词短语(NP)得到候选评价对象;继而对候选评价对象使用频率过滤、PMI 算法和名词剪枝等算法进行筛选得到最终的评价对象表。

2.1 候选评价对象抽取

评价对象可以是名词或名词短语。关联规则挖掘方法提供了一种抽取语料中频繁项的方法,可以有效的抽取语料中的评价对象。但其存在一定问题,就是在判定两个词能否形成短语时仅考虑了两个单词的共现次数,并没有考虑到短语的句法结构,

为此,我们引入了句法分析技术来抽取评价对象。考察情感句“这款电脑的续航能力很强,但价格太高。”其句法结构如图 2 所示。使用关联规则挖掘方法,若“能力”和“价格”的共现次数足够多,则“能力价格”会被识别成评价对象。按照我们的句法限制,此句的评价对象为“续航能力”、“续航”、“能力”、“价格”,可以看出,通过使用句法信息,我们能够避免抽取“能力价格”这种错误评价对象,但同时也会产生一定的噪声(续航、能力),为此,我们引入了三种过滤技术,下面对其进行详细介绍。

2.2 评价对象筛选

由句法分析得到的候选评价对象集存在一定的噪声,为此,我们加入了相应的过滤技术。评价对象筛选流程如下:首先,使用词频信息进行过滤;其次,使用基于网络挖掘的 PMI 算法进行过滤;最后,

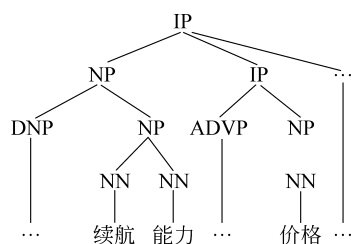


图 2 句法分析示例

使用名词剪枝技术解决单个词的冗余现象。下面分别对这三项技术进行介绍:

1) 词频过滤:

词频过滤即将语料中出现次数比较少的 NN 或 NP 过滤掉。词频信息过滤的加入主要基于两点考虑: 1. 评价对象更倾向于在评论中多次出现, 一些不相关的 NN 或 NP 应该在商品中很少出现, 如“有限公司”、“多图”等。2. 词频信息过滤可能会过滤掉一些评价对象, 但这对系统的结果影响不会很大, 因为那些出现次数较少的评价对象并不被大多数人所关心, 属于次要属性。

2) PMI 过滤 (Pointwise Mutual Information):

PMI 值能够量化词与词之间的关系, 在一定的文本集合中, 词 a 和 b 的 PMI 值定义如下:

$$PMI_{a-b} = N_{ab} / (N_a \times N_b)$$

其中, N_{ab} 表示既包含 a 又包含 b 的文本数, N_a 表示含有 a 的文本数, N_b 表示含有 b 的文本数。从公式中可以看出, PMI 值的计算使用了统计的思想, 同时基于这样一个假设: 两个单词共现的次数越多, 则它们之间的联系也就越大。PMI 值计算的难点在于大规模文本集合的获取, 理论上讲, 文本数越多, 则统计效果越明显, PMI 值的计算也应该越准确。

本文使用 PMI 值来进一步挖掘评价对象的领域相关性。为了得到足够大的语料, 本文选取百度的搜索结果作为语料库。方法如下: 针对每一领域, 选取最具代表性的词语 w_a , 计算候选评价对象 w 与相关领域 w_a 的 PMI 值, 值越大, 则说明 w 的相关性越强, 更可能成为一个评价对象。最后通过设定阈值的方法进行过滤, 实验表明这种方法取得了比较好的效果。

3) 名词剪枝:

此技术主要应用于冗余名词的过滤。为了说明什么是冗余, 本文首先定义 $s\text{-support}$: 对于名词 t , 设包含 t 的句子数为 s , 在 s 个句子中, t 单独作为评价对象出现的句子数为 k (这 k 个句子中不含有包

含 t 的短语), 则 $s\text{-support} = k/s$ 。对于 $s\text{-support}$ 值小于 0.5 的名词评价对象, 本文认为它是冗余的, 作过滤处理。

3 评价对象倾向性判断

针对该任务, 本文首先分析情感句的结构, 将其分为四类; 继而针对各类制定相应的倾向性判断规则, 最终基于无指导的方法完成评价对象的倾向性判断。下面对其进行详细介绍。

3.1 情感句型分类

通过观察语料, 本文将句子分为四类, 具体定义如下:

类别一: 句子带有明显的倾向性, 即句子中带有一种倾向性(褒义或贬义)的上下文无关情感词明显多于另一种。例如: “这款数码相机的质量不错外形也很漂亮”, 这句话的情感词为“不错”和“漂亮”, 均为褒义, 则此句明显含有褒义的倾向性。

类别二: 句子不带有明显的倾向性, 但句子中含有情感词, 且褒义和贬义情感词的个数相同。例如: “这款相机的质量不好但外形很漂亮”, 这句话中含有“不好”和“漂亮”两个情感词, 极性分别为贬义和褒义, 但不能说此句的倾向性是褒义还是贬义。

类别三: 句子不带有明显的倾向性, 且句子中没有情感词, 但其上下文的句子带有明显倾向性。例如: “这款相机的质量非常好, 照出的照片也一样”, “照出的照片也一样”这句话本身不含有情感词, 但此句的前一个句子“这款相机的质量非常好”带有明显的倾向性。

类别四: 句子不带有明显的倾向性, 句子中没有情感词, 且其上下文的句子也不带有明显倾向性。例如: “我购买相机主要看中质量。”、“锐志和皇冠有很多类似的地方, 如外型、内饰等”。

3.2 倾向性判定规则

针对前三类句子, 本文分别定义规则对其进行处理, 这三个规则也是本文倾向性判断的基础; 对于第四类句子来说, 无法找到一个通用的规则, 为此本文引入了上下文相关极性词的概念, 下面对具体规则进行详细介绍:

规则一: 对于第一类句子, 句子中评价对象的

极性与句子的极性相同, 如“这款相机的质量不错外形也漂亮”, 这个句子的倾向性为褒义, 则其中的评价对象“质量”和“外形”均为褒义。

规则二: 对于第二类句子, 找出句子中的评价对象, 针对每个评价对象, 选取评价对象 8 个字(实验表明这个窗口大小比较合适)的范围内与其最近的情感词作为直接修饰它的情感词, 评价对象的极性即与修饰它的情感词的极性相同。

规则三: 对于第三类句子, 使用上下文信息进行判断, 当前句子优先与前一个句子的极性相同, 如其前一个句子也不存在明显的倾向性, 则认为当前句子与其后一个句子的倾向性相同。句子的倾向性判定之后, 则句子中所有评价对象的极性与此句子的极性相同。

规则四: 最后一类句子中包含了剩余的所有句子, 为了进一步挖掘此类句子中的情感句, 本文引入了上下文相关情感词的概念。考虑“这款相机的电池寿命很长。”这个句子, “长”这个词本身并不含有极性, 但在修饰“寿命”的时候, 就含有了一定的极性, 我们将这种词称为上下文相关情感词。针对这种情况, 相应的评价对象倾向性判定方法如下: 首先找出句子中的评价对象, 然后查找距离其最近的上下文相关情感词, 如果二元对(情感词, 评价对象)含有极性, 则抽取结果, 否则过滤掉。上下文相关情感词表较小, 并搭配评价对象而产生倾向性, 该词表由人工构建。

这四类句子的处理优先级与上面介绍的顺序相同, 即对一个句子, 依次判断它是否属于某一类, 属于则按照相关规则处理, 给出倾向性结果, 若无倾向性, 则直接过滤掉。

4 实验结果及分析

本文通过第一届中文倾向性分析的评测结果来说明本文方法的有效性, 对系统所使用技术的评价包括三个方面: 1. 系统的总体性能, 即评价对象抽取以及倾向性分析的结果; 2. 三种候选评价对象过滤方法对系统指标的影响; 3. 情感句分类精度。下面对这三个方面进行详细介绍。

4.1 系统总体性能

本系统参加了第一届中文倾向性分析评测, 在评测中取得了较好的成绩。此次评测的语料主要涉及汽车、手机、数码相机以及笔记本四个领域, 共有 473 篇文档(3 000 个句子, 10 000 个具有倾向性评价的评价对象)。评测方法有精确评价和覆盖评价两种, 精确评价是指抽取的实体与答案完全匹配, 覆盖评价是指抽取的实体与答案有重叠就看作正确匹配, 其中重叠部分大小可以调整, 用多个参数进行评价。表 1 给出了系统在这两种情况下评价对象抽取的性能指标。

表 1 评价对象提取的评测结果对比

		属性抽取的结果 (Strict)			属性抽取的结果 (Lenient)		
RunID	DataSet	Precision	Recall	F-measure	Precision	Recall	F-measure
HIT_IR	Car	0.312 6	0.412 7	0.355 7	0.451 3	0.595 8	0.513 6
	Camera	0.321 9	0.400 6	0.356 9	0.455 6	0.567	0.505 2
	Phone	0.347 2	0.381	0.363 3	0.497 6	0.546	0.520 6
	NoteBook	0.337 9	0.456 6	0.388 3	0.469 3	0.634 2	0.539 4
	All	0.327 5	0.405 8	0.362 5	0.467	0.578 8	0.516 9
所有评测结果平均值		0.287 749	0.226 97	0.233 079	0.442 581	0.347 527	0.357 563
所有评测结果最大值		0.564 1	0.417 2	0.397 6	0.720 6	0.578 8	0.516 9

由结果可以看出, 本文的评价对象提取部分在 Strict 评价和 Lenient 评价中均远远高于平均水平, 且在 Lenient 评价中取得了最好的成绩, 这也进一步证明了本文方法的有效性。但尽管如此, 我们发现该系统还是存在的问题。通过观察可以看

出, Lenient 下的属性抽取结果明显要优于 Strict 下的属性抽取结果, 这说明了系统对评价对象的定位准确, 但边界识别存在的问题, 造成边界识别错误的原因主要有以下两点: 1. 评价对象本身为复杂短语, 如“空调系统面板的布局”、“P. A. T. S. 发动

机电子控制防盗模块”等; 2. 预处理模块(分词、句法分析)带来的噪音。因此, 如何解决复杂短语以及预

处理模块的纠错是我们今后研究的主要方向。

表 2 评价对象倾向性判断的评测结果对比

		属性极性判断结果(Strict)			属性极性判断结果(Lenient)		
RunID	DataSet	Precision	Recall	F-measure	Precision	Recall	F-measure
H IT_ IR	Car	0. 284 2	0. 375 2	0. 323 4	0. 409 4	0. 540 6	0. 466
	Camera	0. 297 4	0. 370 2	0. 329 9	0. 421 1	0. 524 2	0. 467
	Phone	0. 318	0. 348 9	0. 332 7	0. 458 3	0. 502 9	0. 479 5
	NoteBook	0. 302 9	0. 409 3	0. 348 1	0. 425	0. 574 3	0. 488 5
	All	0. 299	0. 370 5	0. 330 9	0. 427 4	0. 529 6	0. 473
所有评测结果平均值		0. 230 635	0. 188 586	0. 191 876	0. 353 313	0. 286 639	0. 292 517
所有评测结果最大值		0. 370 6	0. 370 5	0. 336 7	0. 470 5	0. 529 6	0. 473

表 2 给出了评价对象倾向性判断的评测结果。评价对象的倾向性分析是基于评价对象抽取结果的, 存在错误累加效果。通过对比表 1 和表 2, 我们发现倾向性分析的 F 值(0. 336 7、0. 473)并没有较评价对象抽取的 F 值(0. 397 6、0. 516 9)下降很多, 这说明对于绝大部分被正确抽取出的评价对象, 我们的倾向性分析结果都是正确的, 这充分说明了我们的倾向性分析规则的合理性。

4.2 评价对象过滤技术性能

选取笔记本领域的语料对评价对象过滤技术的

有效性进行验证, 笔记本领域共有语料 56 篇, 1 036 个具有倾向性评价的评价对象, 选取的评价方式为精确评价。表 3 给出了在笔记本领域三种过滤技术对系统指标的影响。表的第一行为未对候选评价对象集进行过滤所得到的结果, 接下来的三行依次为加入每种过滤方法得到的结果。

由表 3 可以看出, 随着加入各种过滤技术, 系统指标不断上升, 在加入 PMI 值过滤技术后, 系统指标明显上升, 这也与本文的设想一致, 即通过 PMI 值挖掘领域信息, 能够有效的过滤候选评价对象集中的噪声。

表 3 过滤技术对系统指标的影响

	属性极性判断结果(Strict)			属性极性判断结果(Lenient)		
	Precision	Recall	F-measure	Precision	Recall	F-measure
未对过滤候选评价对象集	0. 085 3	0. 278 0	0. 130 5	0. 077 6	0. 252 9	0. 118 8
仅使用频率过滤方法	0. 090 4	0. 358 1	0. 144 4	0. 082 8	0. 328 2	0. 132 2
频率过滤+ 名词剪枝	0. 117 0	0. 404 4	0. 181 5	0. 105 6	0. 364 9	0. 163 8
使用三种过滤方法	0. 337 9	0. 456 6	0. 388 3	0. 302 9	0. 409 3	0. 348 1

4.3 情感句型识别精度

本文将情感句分成四类, 对于每种句型, 均有相应的处理方法来判定其中评价对象的情感倾向性, 因此, 情感句型分类的精度将直接影响评价对象倾向性判定的精度。本文选取第一届中文倾向性分析的语料对情感句型识别进行评价, 分别从这四个领域各随机抽取 50 个句子, 人工对这 200 个句子进行标注, 表 4 给出了这两部分语料的情感句分类精度。

表 4 情感句分类精度

	数码相机	手机	笔记本	汽车
精确率	0. 78	0. 76	0. 84	0. 84

可以看出, 情感句类别精度在各个类别大体相当, 均为 80%左右。情感句型识别是以情感词表为基础, 而情感词表本身会存在噪声和不完全两类问题, 即情感词表中的词可能是错误的, 且部分情感词

(下转第 122 页)

- H. Isahara. Morphological analysis of the spontaneous speech corpus[C] // Proceedings of COLING, 2002.
- [10] M. Asahara. Corpus-based Japanese morphological analysis[D] . Japan: NAIST, 2003.
- [11] T. Nakagawa. Chinese and Japanese word segmentation using word-level and character-level information[C] // Proceedings of COLING, 2004.
- [12] Y. Zhang and S. Clark. Chinese segmentation with a word-based perceptron algorithm[C] // Proceedings of ACL, 2007.
- [13] H. Ng and J. Low. Chinese part-of-speech tagging: one-at-a-time or all-at-once? Word-based or character-based? [C] // Proceedings of EMNLP, 2004.
- [14] Y. Zhang and S. Clark. Joint word segmentation and POS tagging using a single perceptron [C] // Proceedings of ACL, 2008.
- [15] W. Jiang, L. Huang, Q. Liu, Y. Lu. A cascaded linear model for joint Chinese word segmentation and part-of-speech tagging[C] // Proceedings of ACL, 2008.

(上接第 88 页)

可能并未在情感词表中, 就会造成情感句分类不准的情况, 如“按键设计凹凸有质。”这个句子, 本应属于第一类情感句, 但由于情感词“凹凸有质”并不在情感词表中, 导致此句子会被错误分为第三类情感句。因此, 建立更加准确、完整的情感词表也是我们下一步的主要工作之一。

5 结论与未来工作

本文实现了一个评价对象抽取及其倾向性判断系统, 可分为评价对象抽取和倾向性判断两个部分, 系统在 COAE2008 的评测中取得了较好的成绩。由评测的总体结果可以看出, 情感分析技术目前还处于初级阶段, 因此还有很广阔的研究空间。如何更准确的确定评价对象边界, 如何挖掘出更多的情感句等等, 都将成为我们下一步的工作。

参考文献

- [1] Hong Yu, Vasileios Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences[C] // Proceedings of EMNLP-2003, 2003: 129-136.
- [2] Ellen Riloff, Janyce Wiebe, William Phillips. Exploiting subjectivity classification to improve information extraction [C] // Proceedings of AAAI-2005, 2005: 1106-1111.
- [3] 姚天昉, 等. 一个用于汉语汽车评论的意见挖掘系统 [C] // 中文信息处理前沿进展—中国中文信息学会二十五周年学术会议论文集. 北京: 清华大学出版社, 2006: 260-280.
- [4] Mingqing Hu, Bing Liu. Mining opinion features in customer reviews [C] // Proceedings of AAAI-2004, 2004: 755-760.
- [5] 倪茂树, 林鸿飞. 基于关联规则和极性分析的商品评论挖掘 [C] // 第三届全国信息检索与内容安全学术会议, 2007: 635-642.
- [6] Soo-Min Kim, Eduard Hovy. Automatic detection of opinion bearing words and sentences[C] // Proceedings of IJCNLP-2005, 2005: 61-66.
- [7] Jun Zhao, Kang Liu, GenWang. Adding redundant features for crfs-based sentence sentiment classification [C] // Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, 2008: 117-126.
- [8] Mingqing Hu, Bing Liu. Mining and summarizing customer reviews [C] // Proceedings of KDD-2004, 2004: 168-177.