

一种基于句法分析的情感标签抽取方法*

■ 李纲 刘广兴 毛进 叶光辉

[摘要] 指出情感标签由评价对象和情感词组成,包含评论的关键要素,能清楚地表达评价者的观点意见。提出一种针对产品网络评论的情感标签抽取模型,利用依存句法分析设计情感标签抽取算法,通过情感极性计算对抽取出的情感标签进行过滤。通过放宽的抽取规则与情感极性过滤相结合,以提高情感标签的召回率,实现潜在评价对象的抽取。最后用网络抓取的产品评论语料作为测试数据集对模型进行测试,获得较高的抽取准确率和召回率,并对模型中存在的问题进行总结,作为模型改善的指导。

[关键词] 情感标签 观点挖掘 依存句法分析 产品评论

[分类号] TP391.3

DOI: 10.13266/j.issn.0252-3116.2014.14.002

1 引言

随着电子商务的快速发展,商品或服务的交易越来越多地在互联网上发生。根据2014年1月中国互联网信息中心(CNNIC)发布的第33次《中国互联网络发展状况统计报告》,截至2013年12月,我国网民网络应用中网络购物用户规模达3.02亿,使用率达到48.9%。随着网络购物的不断发展和用户量的增多,利用Web 2.0的强交互性,顾客在互联网交易平台上发表产品或服务评论的行为越来越频繁,线上产品或服务的评论信息量急剧增加。这些评论包含顾客对产品的功能、性能等方面的观点,具有巨大的价值。线上商家和顾客可以通过分析评论获取有用的信息以辅助决策^[2],但是由于评论信息存在数量大、内容繁杂、非结构化等特点,对评论信息的分析需要运用自然语言处理技术中观点挖掘的相关理论及技术,抽取出关于产品的评价观点。

观点挖掘的研究主要分为两个方面:①观点词或观点句抽取方法及技术;②观点词或观点句情感极性计算。观点词的抽取与情感极性的判定是观点句抽取与情感极性计算的基础,在此本文重点关注。观点词抽取包括观点词本身的抽取和评价对象与观

点词关系的抽取。观点词的抽取有两种方式:一种是基于词性的直接抽取,包括词组模式^[3]、情感词典构建^[4-5]等方法;另一种是基于观点词评论对象识别的间接抽取,包括基于共现规则的抽取方法^[2]、基于句法分析的抽取模型^[6]。两种抽取方式存在着一定的性能差异,主要表现为:直接抽取不容易遗漏情感词汇,准确性也比较高,但是不容易得到观点词的评论对象及相互之间的关系;间接抽取可以一步完成“评论对象-观点词”意见对的抽取,但是在评价对象隐含的情况下容易遗漏观点词,而且由于评论对象种类的不确定性,评论对象词库构建起来比情感词典复杂很多。观点词情感极性的研究则集中关注如何通过规则或方法的设计来量化情感倾向及其程度,目前主要有3种方法:基于词典的语义距离计算方法^[7-8]、基于语料库的语义关联方法^[9-10]和基于机器学习的文本情感分类方法^[11-12]。

目前在观点挖掘的研究中,对篇章和语句整体情感极性的计算方面已经取得了比较好的效果。然而,对语句中各语义成分及其对应关系的识别方面,还没有形成十分有效的解决方案。具体的问题在于:情感词和评价对象的识别精度和鲁棒性不高;情感的上下文极性研究不足;隐式评价对象识别的研究欠缺

* 本文系国家自然科学基金项目“科研团队动态演化规律研究”(项目编号:71273196)、北京市财政项目“大数据环境下情报服务规范化体系建设”(项目编号:PXM2013_178214_000010)和武汉大学自主科研项目(人文社会科学)“网络视角下的应急情报体系建设主题研究”(项目编号:274014,得到“中央高校基本科研业务费专项资金”资助)的研究成果之一。

[作者简介] 李纲,武汉大学信息资源研究中心副主任,教授,博士生导师;刘广兴,武汉大学信息管理学院硕士研究生,通讯作者,E-mail:lgx1431@126.com;毛进,武汉大学信息资源研究中心博士研究生;叶光辉,武汉大学信息管理学院博士研究生。

收稿日期:2014-04-10 修回日期:2014-05-12 本文起止页码:12-20 本文责任编辑:刘远颖

等^[13]。相较于篇章和语句整体情感极性的计算,语句语义层面的挖掘有着更深远的价值,例如文章自动摘要、新闻事件识别、本体构建自动化等方面都依赖于句子语义层面的挖掘。意见挖掘的最终目标是挖掘评论者对产品属性、功能的态度或评价,因此在句子粒度上识别各语义成分及其相互关系是产品评论挖掘的一个研究重点,但目前要自动处理文本并给出较为完整的语法和语义分析,还有一定的难度。而在现有句法分析研究的基础上,结合情感知识,可以获得有一定成效的观点挖掘的模型和方法^[14]。

借鉴上述研究基础,本文将基于依存句法分析的词对抽取算法与基于 HowNet 的情感极性计算方法相结合,提出一种从产品评论语料中抽取情感标签的方法。该方法能够识别出评论语料中显式和部分隐式的评论对象,并细粒度、结构化地抽取评论对象、程度副词、否定词和观点词,这些元素共同构成了评论的情感标签。根据依存句法分析结果中的核心词展开宽松规则的抽取,以识别出语句中的隐式评价对象,并尽可能地扩大抽取范围,提高抽取结果的召回率。同时,针对宽松抽取规则抽取到的大量错误结果,利用对情感词的情感极性计算进行过滤,以保证抽取结果的精度。

2 基于句法分析的情感标签抽取框架

2.1 情感标签抽取任务

情感分析主要是运用自然语言处理的方法对文本的意见、情感进行计算处理。产品评论的情感分析目的是获取评论文本中的意见,从而了解消费者对于产品和服务的具体认知^[12]。情感标签抽取的任务就是从评论语句中获取消费者的具体意见,具体的计算处理任务为识别产品网络评论语句中意见持有者对于一个或多个评论对象的具体评价,并将其按照对应的关

系细粒度地抽取出来。

S. M. Kim 等认为一个观点 (opinion) 由主题、意见持有者、情感描述项和褒贬倾向性 4 个部分来描述,即意见持有者针对某一特定主题表达出的具有某种褒贬倾向的情感描述^[15]。在产品评论的观点挖掘中,意见持有者基本上是消费者,并且评论语句大多数不包含意见持有者信息,所以本文不考虑意见持有者的抽取。褒贬倾向来自于对应的情感描述项,所以主题与情感描述项是产品评论中最核心的两个元素,本文的主要任务就是识别和抽取产品评论的主题以及对应的情感描述项。主题由产品评论语句中产品的评价特征 (即评论对象) 表示,情感描述项由评论语句中情感词、程度副词、语句否定词三者共同表示。

因此,情感标签的结构由观点句中的评价对象、情感词、表达情感词程度的副词及否定词组成,这些结构元素共同表达了评论者的情感倾向及程度。情感标签是最终抽取的成果,对其结构的描述有助于在本方法中做出针对性的设计与调整,情感标签的结构如表 1 所示:

表 1 情感标签结构

评价对象	否定词	副词	情感词
屏幕	-	很	漂亮
价格	不	-	高

2.2 情感标签抽取流程

图 1 给出了从评论信息中抽取情感标签的整体流程。抽取模型首先对观点句进行句法解析,将每一个观点句解析得到观点句的句法依赖关系树,然后根据句法规则和评价特征从依赖关系树中抽取情感标签,再根据情感极性值计算和错误词组特征对情感标签进行过滤,最后生成情感标签。由于网络评论普遍存在语法不规范现象,本文采取抽取规则放宽、过滤环节严格筛选的策略,目的在于尽可能全面地识别出语句中的情感标签。

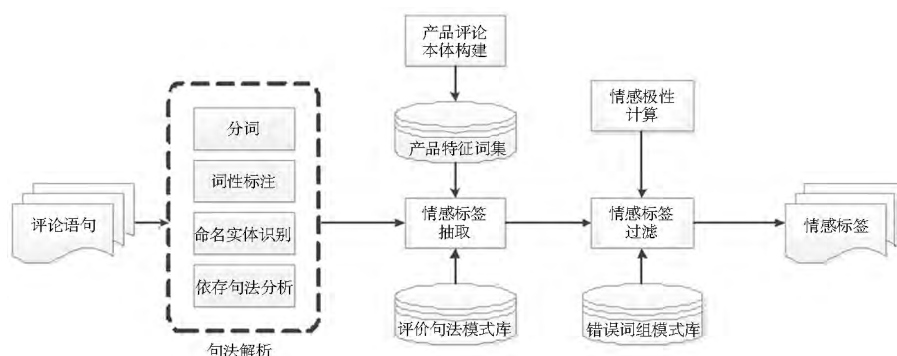


图 1 产品评论语料情感标签抽取模型

总体上讲, 整个抽取流程包括 4 个核心环节: 句法解析、产品评论本体构建、情感标签抽取和情感标签过滤。

2.2.1 句法解析 评论语句的句法解析, 主要通过对评论语句依次进行分句、词性标注、命名实体识别、依存句法分析, 生成对应的句法依赖关系树, 提供给下一环节进行情感标签的抽取。

本文利用哈尔滨工业大学的 LTP(Language Technology Platform) ^[16] 工具对评论语料依句进行句法解析, LTP 能够自动对一条语句进行分句、词性标注、命名实体识别、依存句法分析, 生成一个含有节点以及节点间关系的信息流。此信息流就是一个句法依赖关系树, 树中节点是观点句句后的词语, 词语之间根据语法形成不同的关系, 在树中显示成某一种关联关系, 这些关联关系就是情感标签抽取的基础。

2.2.2 产品评论本体构建 此环节属于线下行为, 主要针对目标评论产品类型, 构建该类产品的评论本体, 生成评论产品的特征词集, 为情感标签的抽取环节服务。

J. Yi 等认为, 对于在线商品评论, 评论的对象一般是商品的属性。商品的属性一般包括: ①商品名称; ②商品的组成部分; ③商品的特点和功能; ④商品属性的特点和功能^[17]。参考上述分类, 产品评论的特征可以分为以下几种: 产品本身、产品的组成部分、产品及其组成部分的属性。为了很好地组织这些特征及其之间的关系, 本文引入本体的组织方法。利用本体中 Part - of(部分) 关系可以表示产品及其组成部分的整体与部分关系; 利用 Kind - of(种类) 关系可以表示产品概念的种类关系; 利用 Attribute - of(属性) 关系可以表示产品及其组成部分的属性关系; 利用 Instance - of(实例) 关系可以表示产品本身、产品组成部分、产品及其组成部分的属性指标的具体实例关系。利用本体来表示产品特征以及特征之间的关系除了能够科学地组织产品特征信息, 也能表示特征之间的语义关系, 为情感标签抽取之后的功能拓展(如评论本体构建、推理功能等) 构建本体库基础。

利用领域语义词典和评论语料构建产品评论领域本体, 形成产品评论的特征词集, 用于情感标签抽取环节中“评价对象 - 情感词”词对抽取策略中的核心词匹配。

2.2.3 基于评价句法模式和评论特征的情感标签抽取 情感标签的抽取环节是整个抽取过程中最核心的环节, 主要目的为从句法依赖关系树中识别和抽取情

感标签。具体的抽取过程为: 根据句法解析中生成的句法依赖关系树, 归总评价对象、情感词、程度副词、否定词之间的语法关系, 设计出能够识别出情感标签的评价句法模式, 形成评价句法模式库, 对每一个观点句的句法依赖关系树进行匹配, 抽取成形如 < 屏幕, 不是, 很, 高 > 的情感标签。此阶段抽取出来的情感标签会包含一些错误标签, 满足评价句法模式中语义关系的词语集合并不都是“评价对象 - 情感词”词对, 需要在下一环节进行过滤筛选。

2.2.4 基于情感极性计算和错误词组模式的情感标签过滤 为了提高情感标签抽取的召回率, 情感标签抽取环节放宽了评价句法模式的识别规则, 将所有符合评价句法的句法规则均纳入抽取模式库, 这势必造成一些非情感标签的错误抽取。因此, 在此环节识别和过滤错误的情感标签: 利用基于 HowNet^[18] 的情感极性值计算, 去除没有情感倾向的情感标签; 根据抽取出的非情感标签的特点, 总结出非情感标签的词组模式, 并将其作为错误词组模式对情感标签进行过滤筛选, 进而得到最终的情感标签库。

3 融合句法结构和评价特征的情感标签抽取规则

在情感标签抽取环节中, 根据句法分析所得评价句的句法依赖树, 设计如下几个方面的句法抽取规则: 评价对象与情感词的抽取规则、副词抽取规则、否定词抽取规则和定语抽取规则。其中定语抽取主要用于识别评论对象的定语, 如语句“手机屏幕的色彩很鲜艳”中, 评论对象为“色彩”, 需要将色彩的定语“屏幕”、“屏幕”的定语“手机”都抽取出来, 合成新的评价对象“手机屏幕色彩”, 使得评价对象更精确。

3.1 评价对象与情感词抽取规则

通过句法分析, 可以更加深入地挖掘评价对象和情感词之间的修饰关系, 相较于单纯停留在词表面或句子形式表面的抽取具有更高的精确度^[19]。从生成的句法依赖关系树中, 根据词与词之间在语义层面上的修饰关系, 可以识别和抽取评论语句中的“评价对象 - 情感词”词对。

规则一: 若核心词(HED 指向词) 词性是“形容词(用 a 表示)”或“其他名词修饰符(用 b 表示)”, 则认为该核心词为情感评价词, 然后寻找左依赖于该核心词的语法元素, 若它们之间依赖关系为 SBV(主谓关系), 则认为该语法元素为评价对象词(见图 2)。

规则二: 若不满足规则一题设要求, 则寻找左依赖



图 2 抽取规则一示例

注: 本文图示中句法关系呈现工具为 2013 哈尔滨工业大学社会计算与信息检索研究中心在线演示平台(<http://www.ltp-cloud.com/demo>)。

于核心词的语法元素,若它们之间依赖关系为 SBV(主谓关系)则认为该语法元素为评价对象词。同时,寻找右依赖于核心词的语法元素,若其词性为“形容词”、“动词”或“其他名词修饰符”,则认为该语法元素为情感评价词。如图 3 所示:



图 3 抽取规则二示例

规则三: 匹配本体特征集,识别出评价语料中的特征词以及特征词所依赖的语法元素。若依赖关系为 ATT(定中关系)、ADV(状中结构)或 SBV(主谓关系),而且其依赖的语法元素词性为“形容词”、“其他

名词修饰符”、“副词”、“动词”或者“习语”,则认为该特征词为评价对象,其依赖的语法元素对应的词为情感评价词。如图 4 所示:



图 4 抽取规则三示例

规则四: 若规则三没有识别出情感评价词对,则认为该特征词所依赖的词为非情感评价词,继续寻找依赖于非情感评价词的语法元素,即寻找间接关联对象。寻找规则可描述为: 寻找非情感评价词右依赖的语法元素,若它们之间的依赖关系为 SBV(主谓关系)、ATT(定中关系)、COO(并列关系)、VOB(动宾关系)或 CMP(动补结构),且该语法元素词性为“形容词”、“其他名词修饰符”、“副词”、“动词”或者“习语”,则认为该特征词为评价对象,间接关联对象所对应的词为情感评价词。如图 5 所示:



图 5 抽取规则四示例

表 2 评价对象与情感词抽取规则

规则	评价语法模式	评价对象词	情感评价词
规则一	$T \xleftarrow{SBV} H$		$pos(C) \in \{a, b\}$
规则二	$T \xleftarrow{SBV} H \rightarrow S$		$pos(S) \in \{a, b, p\}, pos(C) \notin \{a, b\}$
规则三	$T \xleftarrow{SBV, ADV, ATT} S$	$pos(T) \in \{n, nd, ws, k\}$	$pos(S) \in \{a, b, d, p, i\}$
规则四	$T \xleftarrow{M} \xrightarrow{SBV, ATT, COO, VOB, CMP} S$	$pos(T) \in \{n, nd, ws, k\}$	$pos(S) \in \{a, b, d, p, i\}$

为了更加清楚地表述 4 种规则,笔者采用表 2 中的模式来结构化呈现抽取过程。表 2 中, T 表示评价对象, H 表示核心词, S 表示情感评价词, M 表示中间词, $pos(X)$ 表示元素 X 的词性。词性含义与依赖关系说明参见表 3 和表 4。

表 3 词性参照表

标识符	含义	示例
n	名词	鼠标
nd	方向名词	右侧
ws	外来词	CPU
k	后缀	界、率
a	形容词	高、快
b	其他名词修饰符	大型、西式
d	副词	很、非常
v	动词	跑
i	习语	百花齐放

表 4 依赖关系参照

标识符	关系类型	示例
SBV	主谓关系	做工精美(做工←精美)
ATT	定中关系	这是个好评盘(好←键盘)
COO	并列关系	CPU 和 RAM(CPU←RAM)
VOB	动宾关系	价格不是很高(是→高)
ADV	状中关系	速度非常快(非常←快)
CMP	动补关系	电脑跑起来很顺畅(跑→顺畅)
HED	核心关系	整句的核心

由表 2 可知,规则一和规则二是由核心词展开的抽取规则,主要用于潜在评价对象(未加入特征词集的评价对象)的抽取;规则三和规则四是由特征词展开的抽取规则,较为全面,能够解决一条语句中含有多个评价对象的问题。大多数情况下规则一和规则二抽取出的情感标签会与规则三、规则四的抽取结果重复,在抽

取结束后需要对抽取结果去重。因为一条语句中可能存在多个评价对象, 这些评价对象中有可能存在潜在评价对象, 而规则三、规则四无法识别, 因此规则一与规则二的设计是必要的。

3.2 副词抽取规则

查找左依赖于情感评价词的语法元素, 如果该元素词性为“副词”且依赖关系为 ADV(状中关系), 则认为该词为修饰情感评价词的副词。

3.3 否定词抽取规则

识别间接关系中的中间语法元素的左依赖词, 若为“副词”并存在于否定词集中, 则该词为“否定词”。这里所指否定词是对意见评价起到意见反转作用的词, 而作为副词的否定词不属于此类情况。例如, “这个手机不是很好看”中的“不”为句子的否定词, 而“我认为这个手机不漂亮”中的“不”为“漂亮”的副词。

3.4 评价对象定语抽取规则

寻找左依赖于评价对象词的语法元素, 如果依赖关系为 ATT(定中关系), 且该语法元素的词性为“名词”或“动词”, 则该元素为评价对象的定语, 将该元素与原评价对象词合并, 构成新评价对象。

4 情感标签过滤方法

在情感标签抽取环节中, 根据评价句法规则抽取出来的情感标签存在着重复、词语及词语间关系错误抽取的问题, 如评价对象与情感词对应错误、情感词识别错误等。针对该问题, 运用情感极性筛选的方法去除其中情感词不具有明显极性的情感标签。针对重复词组和错误词组的问题, 本节构建错误词组模式库, 对情感标签进行匹配筛选。

4.1 情感极性筛选

利用基于 HowNet 的情感极性值计算^[6]方法计算情感标签中情感词的情感极性。HowNet 将每一个词用一系列的义原来描述, 可以将词理解为多个义原的集合, 这些义原通过上下位关系组织成一个树状义原层次体系^[20]。如词语“高兴”被表示为义原集合 < aValue | 属性值, circumstances | 境况, happy | 福, desired | 良 >; 词语“北京”被表示为 < place | 地方, capital | 国都, ProperName | 专, (China | 中国) > 的义原集合。通过树状义原层次体系中义原之间的位置关系, 可以计算出义原之间的相似度, 而通过词的义原集合, 可以计算词与词之间的相似度。义原相似度计算公式如式 (1) 所示:

$$\text{sim}(p1, p2) = \frac{\alpha}{d + \alpha} \quad (1)$$

式 (1) 中, $\text{sim}(p1, p2)$ 表示义原 $p1$ 和义原 $p2$ 的相似度, d 表示树状义原层次体系中的路径距离, α 为可调节参数。在式 (1) 基础之上, 可推出词语间的相似度计算公式, 如公式 (2) 所示:

$$\text{sim}(w1, w2) = \max_{i=1 \dots n, j=1 \dots m} \text{sim}(s1i, s2j) \quad (2)$$

在公式 (2) 中, $\text{sim}(w1, w2)$ 表示词语 $w1$ 和词语 $w2$ 的相似度, 其中词语 $w1$ 有 n 个义原 ($s11, s12, s13, \dots, s1n$), 词语 $w2$ 有 m 个义原 ($s21, s22, s23, \dots, s2m$)。

通过构建褒义基准词库和贬义基准词库, 利用词语间相似度来计算情感词的情感极性, 计算公式如式 (3) 所示:

$$\text{orientation}(w) = \sum_{i=1}^n \text{sim}(pw_i, w) - \sum_{j=1}^n \text{sim}(nw_j, w) \quad (3)$$

公式 (3) 中, $\text{orientation}(w)$ 代表词语的情感极性值, 褒义基准词库中有 n 个基准词, pw_i 代表第 i 个褒义基准词; 贬义基准词库中有 n 个基准词, nw_j 代表第 j 个贬义基准词。

利用公式 (3) 计算出情感标签中情感词的情感极性值, 设定阈值区间, 将符合阈值区间条件的情感词选出, 去除情感词情感极性值未满足阈值区间条件的情感标签。

4.2 错误词组过滤

情感标签抽取模式中存在着一些规则, 这些规则辅助本文方法从评价语料中抽取正确的情感标签, 但同时由于设计规则无法穷尽所有情形或规则分类细化不够, 还有可能会抽取错误的情感标签, 错误词组的一部分可通过情感极值计算过滤, 另一部分则不能。针对这些错误抽取出来而未被情感极值筛选排除的词组, 需要构建错误词组模式库, 对情感标签进行错误词组模式匹配筛选。

模式一: 一个句子中抽取出的情感标签中, 含有两个评价对象在句中存在着 ATT(定中关系) 依赖。

如图 5 的例子中, 根据规则三可抽取 < 鼠标大小, -, -, 合适 > 的标签, 根据规则一可抽取 < 大小, -, -, 合适 > 的标签, 根据规则四可抽取 < 鼠标, -, -, 合适 > 的标签, 而“大小”是“鼠标”的属性, 第二、三个标签有误, 且情感词“合适”具有强情感极性, 在情感值筛选中无法排除。对此, 解决的策略为: 当一个句子中出现两个评价对象存在着 ATT(定中关系) 依赖的时候, 去除依赖者和被依赖者的情感标签。

模式二: 同一个情感词有多个副词。

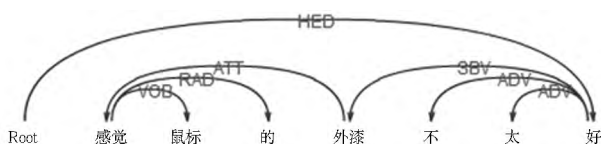


图 6 模式二语料示例

以图 6 语料为例,根据副词的抽取规则,会抽取两个标签, < 外漆, -, 不, 好 > 和 < 外漆, -, 太, 好 >, 两个标签都是不正确的, 所要的情感标签应该是 < 外漆, -, 不太, 好 >。对此, 解决的策略为: 当一句话中某一情感词含有两个以上的副词时, 首先识别各个副词的先后顺序, 然后将副词按照句中顺序合并为一个副词, 其他元素不变, 生成新情感标签, 去除原来的两个或两个以上标签。

模式三: 一个句子中抽取出的情感标签有重复。

在评价对象和情感词的抽取规则中, 规则一、二与规则三、四的抽取结果会重复, 以图 5 语料为例, 规则一和规则三都会抽取出来 < 大小, -, -, 适合 > 的情感标签, 在同一语句中重复。对此, 解决策略为: 当一句话中出现了两个以上的评价对象和情感词都一样的情感标签时, 只留下一个。

5 实验分析

5.1 实验数据

实验所用语料来自网络爬取的关于手机、数码相机、鼠标、显示器、路由器 5 类电子产品的中文评论, 总共近 2 825 条, 单句共 6 415 个, 人工识别每句情感标签, 语料共包含情感标签近 5 283 条, 各类语料数据分布情况见表 5。综合参考《中国分类主题词表》、百度百科以及淘宝评论, 从产品组成部件、功能、属性 3 个方面构建 5 类产品的本体特征词集, 共收录特征词 399 个。利用朱嫣岚等人^[6]选取的 40 对褒贬义基准词作为本实验 HowNet 情感值计算中褒贬基准词, 所选基准词具有强烈情感倾向和一定代表性。

表 5 语料数据

产品类型	评论条数	评论单句个数	情感标签个数	句均字符个数
鼠标	431	464	423	25
手机	609	1 767	1 461	85
数码相机	861	1 838	1 743	59
显示屏	587	1 762	1 255	104
路由器	337	584	401	51
总计	2 825	6 415	5 283	67.8

评论语料中的语句, 大多数包含多个分句, 甚至多个整句, 句子或者句段中包含 0 个、1 个或 1 个以上的情感标签。语料中也包含没有明确情感倾向的句子, 所以表 5 中鼠标语料数据中情感标签数少于评论语句数是合理的。句均字符数表示语料中语句的平均字数, 用来衡量语料句子的长短。5 类产品语料中, 鼠标语料句子相对较短, 其中短句较多; 显示屏语料句子相对较长, 其中长句或者句段较多。综上, 所取语料较为平衡, 满足对模型测试的要求。

5.2 评价指标

本文使用准确率和召回率以及 F-Measure 对实验结果进行评估, 同时将人工识别出的情感标签作为标准情感标签结果。准确率是指抽取结果中的正确情感标签数与抽取出的情感标签总数的比值, 召回率是抽取结果中的正确情感标签数与观点句库中实际存在的情感标签总数的比值, F-Measure 则是综合了准确率和召回率的评价指标。公式如下:

$$P(\text{准确率}) = \frac{A}{A+B} \quad (4)$$

公式(4)中 A 表示抽取结果中正确个数, B 表示抽取结果中不正确个数;

$$R(\text{召回率}) = \frac{A}{A+C} \quad (5)$$

公式(5)中 A 表示抽取结果中正确个数, C 表示遗漏个数;

$$F\text{-Measure} = \frac{(\beta^2 + 1)PR}{\beta^2 P + R} \quad (6)$$

公式(6)中 β 是参数, 用来衡量准确率和召回率的相对重要性, P 是准确率, R 是召回率。在本实验分析中, 本文假定准确率和召回率同样重要, 故 β 取 1 值, 此时 F-Measure 为 F1:

$$F1 = \frac{2PR}{P+R} \quad (7)$$

5.3 实验结果分析

表 6 中显示的是鼠标产品评论语料实验结果中语句 364 到 370 的抽取结果, 表中左边为原始语料, 右边为抽取出的情感标签。其中语句 365、369 均包含两个情感标签, 语句 365 存在抽取错误。由于语句 365 省略主语, LTP 在判断句法依赖关系时把“定位”判断为主语, 把“适合”判断为谓语, 且两者具有依赖关系, 通过规则三中被抽取出来, 而在情感极性计算中, “适合”一词的情感极性值超过阈值, 判定为含有情感极性的情感词, 系统最终把 < 定位, 适合 > 作为情感标签保留下来。此错误主要由原句缺乏主语以及谓语具有强极性造成。

表 6 情感标签抽取结果样例

观点语料		情感标签抽取结果				
语句编号	观点句	语句编号	评价对象	否定词	副词	情感词
364	鼠标线很长,有 1.8 米,无论机箱离得多远都很方便了	364	鼠标线	-	很	长
365	定位准确,尤其适合做图	365	定位	-	-	准确
		365	定位	-	尤其	适合
366	发光管有节能设计,很实用	366	发光管	-	很	实用
367	这样的价格是适中的了,我觉得	367	价格	-	-	适中
368	使用时感觉精度不错,而且左右移动时不用设置	368	精度	-	-	不错
369	鼠标的外形很新颖,酷似苹果机,而且它的外观感觉很酷	369	鼠标外形	-	很	新颖
		369	外观感觉	-	很	酷
370	在表层覆盖着一层光滑的材质,而且与手部的结合部位没有缝隙,特别便于清理	370	材质	-	-	光滑

5.3.1 抽取准确率与召回率分析 表 7 展示了情感标签抽取实验最终的统计结果,可以看到 5 类语料的情感标签抽取准确率均在 70% 以上,召回率均在 80% 以上,总体结果理想。准确率方面,鼠标与显示屏语料的抽取准确率相对较低。对比表 5,鼠标语料的句均字符数最低,只有 25 字/句,人工识别过程中发现鼠标语料中短句较多,主语省略情况较频繁,这是造成其准确率较低的主要原因;显示屏语料的句均字符数非常高,达到 104 字/句,人工识别过程中发现显示屏语料的长句或句段较多,指代关系复杂且一些语句存在语法不规则、语法错误的现象,这是造成其准确率较低的主要原因。手机与数码相机语料的抽取准确率相对较高,主要是由于这两类语料的语句长度适中,语法相对规范且样本数据量较大,避免了部分语法错误句、主语省略的影响。召回率方面,5 类语料的召回率都比较高,且相对稳定,说明情感标签抽取规则可以识别出大部分的评论语法结构,同时构建的本体特征词集也相对完整。综合考虑,语料总体的准确率和召回率均相对较高,F1 值接近 0.8,总体的抽取效果比较理想。

表 7 实验结果

产品类型	抽取总条数	正确条数	未识别条数	准确率	召回率	F1
鼠标	484	348	75	0.719	0.823	0.767
手机	1 544	1 216	245	0.788	0.832	0.809
数码相机	1 861	1 442	301	0.775	0.827	0.800
显示屏	1 514	1 061	194	0.701	0.845	0.766
路由器	474	343	58	0.724	0.855	0.784
总计	5 877	4 410	873	0.750	0.835	0.790

5.3.2 HowNet 情感值计算筛选分析 表 8 展示了情感标签过滤中情感词情感值计算筛选过程的统计结果,其中筛选前条数指的是在错误词组模式筛选之后的抽取条数,避免了错误词组模式筛选的干扰。表中可看到各类语料的情感值筛选正确率均达到 90% 以上,筛选率均达到 10% 以上,且均相对稳定,说明情感标签过滤中情感值筛选效果十分理想。在情感标签抽取规则三、四中,由于允许抽取词性为“动词”的情感词,使得抽取结果包含一定数量的错误标签。如语句“鼠标使用起来挺顺手”中会抽取出来使用“起来”的错误标签。其中“起来”一词不具有较强情感极性,可以通过情感极性值的计算及阈值设定筛选排除。实验中,10% 以上的筛选率以及 90% 以上的筛选正确率使得情感值的极性筛选对于情感标签抽取的最终准确率贡献很大。

表 8 情感值筛选情况

产品类型	筛选前条数	筛选条数	正确条数	正确率	筛选率
鼠标	567	83	80	0.964	0.146
手机	1 723	179	168	0.939	0.104
数码相机	2 115	254	235	0.925	0.120
显示屏	1 746	232	220	0.948	0.133
路由器	535	61	57	0.934	0.114
总计	6 686	809	760	0.939	0.121

5.3.3 抽取规则覆盖率与精确度对比 表 9 展示了各抽取规则的召回率和准确率。在 4 个抽取规则中,规则一与规则二都依据句中核心词进行观点抽取,且规则二建立在规则一条件为非的基础上,故将规则一与规则二合成抽取策略一进行测试;同样,规则三和规则四都依据特征词展开抽取,且规则四建立

在规则三条件为非的基础上,故将规则三与规则四合成抽取策略二进行测试。策略三为 4 个规则共同抽取。从表 9 中结果可以看到,策略一召回率较低,策略二召回率相对较高,达到总召回量的 67%,说明规则三、四共同抽取策略的覆盖率比较高。通过策略一与策略二的召回总和超出总召回量的比值,得到策略一、二的重复量约占总召回量的 15%。去除重复的召回数据,策略一的召回量还约占总召回量的 25%,说明策略一为总抽取模型提供了约 25% 的潜在特征词。在准确率方面,基于特征词抽取的策略二精确度较高,而规则放宽的策略一精确度偏低,都处在合理范围之内。策略一的召回率偏低的主要原因在于规则一二只能抽取语句中的一个情感标签,而测试语料中存在大量包含多个情感标签的长句,导致了策略一的遗漏抽取。而策略一在一条语句中只能抽取一个情感标签的局限是由其基于核心词的抽取策略决定的,故除开策略二对策略一的弥补,想要单独提升策略一的召回率,可以从分解长句入手,这是本文今后要研究的重点之一。

表 9 抽取规则召回率与精确度统计

抽取策略	抽取规则	召回率	准确率
策略一	规则一 + 规则二	0.340	0.686
策略二	规则三 + 规则四	0.622	0.795
策略三	综合	0.835	0.750

5.4 结果讨论

5.4.1 影响本文方法流程的因素

(1) 产品本体构建形成的特征词集的完整性。在抽取算法中,规则三和规则四以产品特征词为核心来抽取情感词,词集如果不够完整会导致某些评价对象被忽略,从而降低召回率。

(2) 实验语料中存在的不规范文法。产品网络评论中普遍存在的句法不规范问题会导致 LTP 进行句法解析的准确率降低,同时影响实验的准确率和召回率。

(3) HowNet 知识库的更新程度。网络评论中常出现一些新兴的评论词,如“给力”、“杯具”等,这些新词及其义原在 HowNet 知识库中未能收入,导致情感误判,在情感极性筛选中被错误筛选出来,影响实验结果。

5.4.2 本文方法存在的不足

(1) 否定词漏判。算法中对于否定词的抽取算法只适用于一般否定语句中的语法结构,对于一些特殊的否定表述(如否定词连接副词)则会被错误识别,出

现否定词漏判的情况。

(2) 评价对象错误识别。由于规则一与规则二是根据句法依赖结构树中的核心词来抽取评价对象和情感词,一个句子中有可能有多个评价对象被抽取,存在 ATT 依赖关系的评价对象可以由错误词组模式筛选出,而其余评价对象未能筛选,造成抽取过程中出现评价对象与情感词不对应的情况。

(3) 抽取算法中,规则三和规则四以本体特征词为核心来抽取情感词,特征词集目前尚需要人工构建,容易导致特征词覆盖不全面的问题。

(4) 产品网络评论语句普遍存在主语省略、指代不明、语法错误、错别字、口语化等现象,评论句法识别在这些方面的处理上尚显不足。

(5) 实验中,抽取规则的算法实现与错误词组模式筛选算法在程序上融合,导致不能统计错误词组模式筛选的情况。

6 结语与展望

本文在句法分析的基础上构建产品网络评论的情感标签抽取模型,获取产品评论中的情感标签。情感标签由评论语句中的评价对象、情感词、否定词、副词组成的词语集合表示,能够代表评论者对某一产品特征的观点意见。本文运用本体方法构建产品特征词集,利用依存句法关系识别和抽取评论语句中的评论对象、情感词,同时细粒度地抽取情感词对应的副词以及句中否定词,并运用 HowNet 计算情感极性值对情感标签进行过滤筛选。本文采用抽取规则放宽、筛选功能增强的策略,在提高召回率的同时保证了较高的准确率。在实验部分,对 5 类电子产品评论语料的情感标签抽取得到了较高的准确率和 F 测试值。针对本文存在的不足,未来的研究工作主要包括如下几个方面:①着重研究评价对象的自动识别,利用统计自然语言处理的方法识别评价对象,并将之与人工方法相结合,提升评价对象识别的精准率;②更加细致地制定评论句法规则,特别是针对评论语料、微博语料等用户生成内容,进一步提高规则覆盖率;③情感标签抽取能够挖掘网络评论文本、网民意见文本等中的带有情感倾向的观点信息,针对电商情感标签而言,可进一步服务于网商、生产企业以及电商平台等多种主体,而针对网民观点而言,可从细粒度角度分析网民口碑和舆情信息。

参考文献:

[1] 中国互联网信息中心(CNNIC). 第 33 次《中国互联网络发展状

- 况统计报告》[EB/OL]. [2014-02-01]. <http://www.cnmc.cn/hlwzzyj/hlwzxbg/hlwztjbg/201403/P020140305346585959798.pdf>
- [2] Popescu A M, Etzioni O. Extracting product features and opinions from reviews[M]//Kao A, Poteet S R. Natural Language Processing and Text Mining. London: Springer, 2007: 9-28.
- [3] Hatzivassiloglou V, McKeown K R. Predicting the semantic orientation of adjectives[C]//Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics. Stroudsburg: ACL, 1997: 174-181.
- [4] Li Zhuang, Feng Jing, Zhu Xiaoyan. Movie review mining and summarization[C]//Proceedings of the 15th ACM International Conference on Information and Knowledge Management. New York: ACM, 2006: 43-50.
- [5] Pak A, Paroubek P. Twitter as a corpus for sentiment analysis and opinion mining[C]//Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10). Valletta: ELRA, 2010: 1320-1326.
- [6] Hu Mingqing, Liu Bing. Mining and summarizing customer reviews[C]//Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2004: 168-177.
- [7] 朱嫣岚, 闵锦, 周雅倩, 等. 基于 HowNet 的词汇语义倾向计算[J]. 中文信息学报, 2006, 20(1): 14-20.
- [8] Kim H D, Zhai Chengxiang. Generating comparative summaries of contradictory opinions in text[C]//Proceedings of the 18th ACM Conference on Information and Knowledge Management. New York: ACM, 2009: 385-394.
- [9] Turney P D. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg: ACL, 2002: 417-424.
- [10] Sun Wenjun, Pan Mingyang, Ye Qiang. Comparative study on objective and subjective emotional tendencies of online reviews from different sources[C]//Internet Technology and Applications (iTAP), 2011 International Conference on. Wuhan: IEEE, 2011: 1-4.
- [11] Mullen T, Collier N. Sentiment analysis using support vector machines with diverse information sources[C]//Barcelona: EMNLP, 2004: 412-418.
- [12] Pang B, Lee L. Opinion mining and sentiment analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1/2): 1-135.
- [13] 姚天昉, 程希文, 徐飞玉, 等. 文本意见挖掘综述[J]. 中文信息学报, 2008, 22(3): 71-80.
- [14] 张紫琼, 叶强, 李一军. 互联网商品评论情感分析研究综述[J]. 管理科学学报, 2010, 13(6): 84-96.
- [15] Kim S M, Hovy E. Determining the sentiment of opinions[C]//Proceedings of the 20th International Conference on Computational Linguistics. Stroudsburg: ACL, 2004: 1367-1373.
- [16] LTP[EB/OL]. [2013-10-24]. <http://www.ltp-cloud.com/download/>.
- [17] Yi J, Nasukawa T, Bunescu R, et al. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques[C]//Proceedings of 2003 IEEE International Conference on Data Mining. Melbourne: IEEE, 2003: 427-434.
- [18] HowNet[EB/OL]. [2013-10-24]. http://www.keenage.com/html/c_index.html.
- [19] 刘鸿宇. 情感标签抽取相关技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2010.
- [20] 刘群, 李素建. 基于《知网》的词汇语义相似度计算[J]. 中文计算语言学, 2002, 7(2): 59-76.

A Sentiment Label Extraction Method Based on Dependency Parsing

Li Gang¹ Liu Guangxing² Mao Jin¹ Ye Guanghui²

¹Center for the Studies of Information Resource, Wuhan University, Wuhan 430072

²School of Information Management, Wuhan University, Wuhan 430072

[Abstract] As a collection of evaluation object and sentiment words, sentiment label contains key elements of on user reviews and can effectively reflect their core contents. This paper proposes a model of extracting sentiment label from products reviews in Web. Based on dependency parsing technology, it designs the algorithm of sentiment label extraction, and filters extracted sentiment labels by setting the threshold for emotional polarity. Combined with the relaxed rules of extraction and emotional polarity filter, it gets higher recall and extracts the potential target of reviews. Finally, it captures online reviews of product as the test data set to test the model and receive a higher precision and recall. To improve the model, the problems in the model are also summarized.

[Keywords] sentiment label opinion mining dependency parsing product review