

在线评论细粒度情感分析

蓝天广, 万岩

(北京邮电大学经济管理学院, 北京 100876)

摘要: 在线评充满了大量决策信息使得消费者与制造商受益。然而, 随着商品属性的多样化, 只对评论整体分类的传统的情感分析已经不能满足消费者与制造商的需求, 他们需要细粒度的属性分类与分析: 消费者需要得到其所关心的目标商品属性的分析, 制造商需要从评论中了解自身产品优劣以及顾客口味。本文进行细粒度情感分析进行商品属性分析, 并且通过通用的评论抓取方法、归纳女式连衣裙商品属性方法及基于词性规则的隐式属性发现方法使得本研究更加具有通用性。

关键词: 细粒度; 情感分析; 属性分析; 隐式属性

中图分类号: TP391

Fine-grained Sentiment Analysis Of Online Reviews

Lan Tianguang , WAN Yan

(School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876)

Abstract: A huge number of online reviews which are valuable voice of customer and full of decision information benefit consumers and product designers. However, with commodity attributes becoming more diverse, the traditional sentiment analysis of overall emotional tendency already can't satisfy the needs of buyers and manufacturers, because they need fine-grained attributes analysis. Consumers hope to get analysis results of target attributes and manufacturers hope to master strengths and weaknesses of products and also consumers' taste. This study conducts a fine-grained sentiment analysis research to solve this problem and we increase generality of the application by the general methods we use to crawl reviews, induce general category of product attributes in women' dress and find implicit features based on POS rules.

Key words: Fine-grained; Sentiment analysis; Attitude analysis; Implicit attributes

0 引言

在线评论 (online review) 是网络口碑的一种重要形式, 一般指潜在或实际消费者在电子商务或第三方评论等网站上发表产品或服务的正面或负面观点^[1], 通过在线方式传递给其他消费者和生产商。从买方来讲, 评论阅读者可能会依据在线评论来判断商品的质量和商家的可信度, 从而决定是否购买; 从卖方来讲, 生产商和销售商可以了解消费者对其商品和服务的反馈信息, 从而改进产品改善服务, 赢得竞争优势^[2]。然而, 面对爆炸式增长的在线评论, 生产商和消费者遇到了信息爆炸时代的普遍问题——信息过载^[3]。

一种有效的组织在线评论的方法是文本分类。情感分析可视为文本分类的一项工作, 其按照情感的极性进行分类, 例如正向、负向^[4]。目前, 情感分析的绝大部分研究都将焦点集中于文本情感极性分类。但是, 随着评论的日渐增加和商品属性的日趋多样化, 整体情感倾向已然不能满足购买者或商家的需求。传统的情感分析未能更细粒度的在属性层面帮助消费者减轻信息过载和认知成本, 而生产商也不能更加深入地了解产品的优缺点, 不能在互联网背景下了解小众需求。更进一步的问题是, 有些评论对其评价的属性未在语句中直接点明,

作者简介: 蓝天广 (1989-), 男, 硕士研究生, 主要研究方向: 信息管理

通信联系人: 万岩 (1966-), 女, 教授, 主要研究方向: 信息管理. E-mail: wanyan@bupt.edu.cn

45 对其的理解需要根据评论的上下文进行。Mingqing Hu 和 Bing Liu 把这类问题称为显式属性和隐式属性的问题^[5]。这类研究不多，对隐式属性的提取是在线评论研究的一个重点和难点^[2]。

本研究针对传统情感分析不能细化消费者口味和明确商品优缺点的不足，以女装连衣裙为例，阐述细粒度情感分析研究。包括：1、对评论的抓取、预处理；2、在归纳了电商女装衣服领域通用的商品属性类后对商品属性归类，3、判断每条评论中每个属性的情感倾向；4、总结消费者偏好和商品优缺点。

本研究的贡献有：1、发现产品优缺点，辅助生产者或者消费者决策；2、基于词性组合的方法，简捷地解决了隐式属性发现这个问题；3、在加入隐式属性后，商品属性情感倾向精确度有了提升；4、首次指明通过隐式属性发现小众需求，这对于依靠长尾理论营销的在线商家有重大意义。

1 细粒度情感分析

按照处理文本的粒度不同，情感分析可分为词语级、短语级、句子级、篇章级几个研究层次^[6]。本研究对象商品属性是词语级和短语级文本，相对于传统情感分析的对象——一般针对句子或者篇章级文本，其文本粒度较细，所以商品属性情感分析是细粒度的情感分析。商品属性是指商品性质或者功能，例如连衣裙属性有价格、款式、穿衣效果、衣服质量等。商品属性有显式和隐式属性，显式属性是直接出现在产品评论中描述产品的性能或功能的名词或名词短语。隐式属性没有在语句中直接进行描述，需要根据上下文，对句子进行语义理解才能得到^[5]，例如，对于衣服价格这个属性来说，在文本中直接出现“价钱”、“价格”，或者其同义词，其可以清楚表明评论内容指向衣服价格这个属性；在文本中出现“便宜”、“划算”、“不贵”等词，暗示了谈论内容为价格属性。对商品进行细粒度情感分析有重大意义：

首先，一个产品具有多个属性（Feature），一段产品评论也可能涉及产品的多个属性，评论者会针对其所关注的一个或多个产品属性“一分为二”地发表看法，所以通过细粒度情感分析，我们能帮助生产商和消费者从商品属性角度发现产品优缺点。其次，对于商品的在线评论集合中，即使是某个的属性中的不同内容部分也不可能完全一样。对于衣服颜色这个属性，可能评论文本数量没有款式多，但是在衣服颜色在个属性中，消费者除了从总体来评论某个属性好坏，例如说衣服颜色好，颜色不错等，还有更为细致明确对色差评论的词，通过这些属性内以不同形式谈论的具体内容，可以发掘长尾市场中对该商品的小众需求。

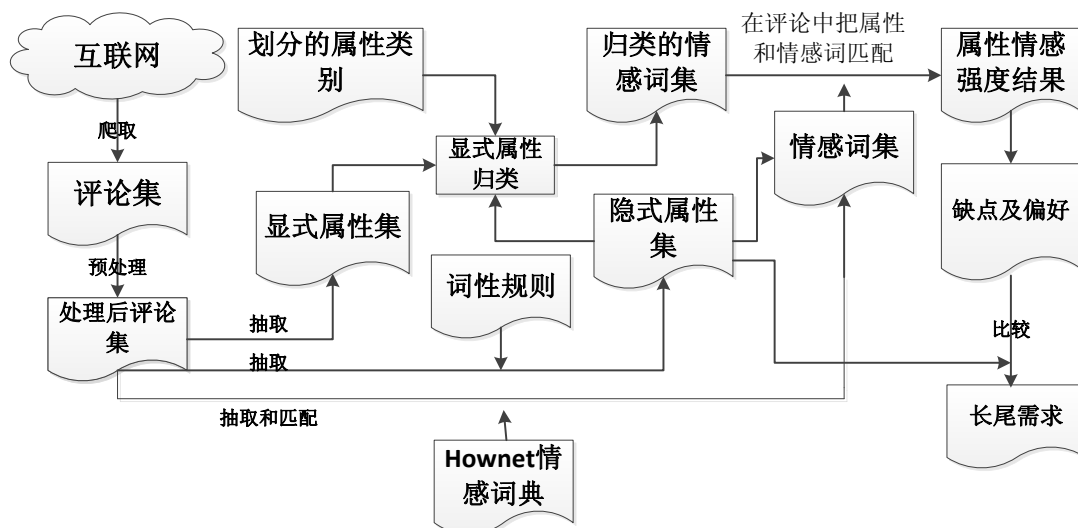


图1 研究总体流程

图1具体说明引言结尾部分提到的研究步骤。首先要从Internet上抓取在线评论，对文本的预处理是为了之后提取属性做准备。由于商品的属性可以由多个词表示，所以我们要对属性进行抽取和归类，这样便于分析文本中究竟涉及到哪些属性。一方面，我们采用了利用词性找到显式属性以及情感词，然后用语义方法对显式属性进行归类；另一方面，我们利用词性组合找到其他情感词，从中提取隐式属性。最后，我们用情感词对属性情感强度打分，每个修饰属性的正向情感词记为对属性情感强度值+1，每个修饰属性的负向情感词记为对属性情感强度值-1。然后分正向负向加总各个属性情感强度值，从中发现产品弱点和消费者偏好；同时我们对隐式属性此表，从中总结小众需求。下面是对研究步骤的具体阐述。

1.1 文本爬取以及预处理

首先我们要爬取在线评论，因为本文要分析的某商品评论接近7300条，不可能一条条去复制；并且为使得本研究更加具有通用性，我们研究了通用的在淘宝网上抓取评论的方法。具体为：1、用网络封包分析软件找到淘宝网在线评论加载网络地址(url)；2、从研究的淘宝商品的url中用正则表达式提取出商品id号，然后使在线评论加载网络地址中id为所要抓取商品id号请求服务器；3、使用JAVA网络输出流获取在线评论网络流，用正则表达式提取中文部分。

其次，预处理，主要包括：（1）去除重复评论：因为有些消费者会重复提交评论；（2）分词：中文词与词之间没有空白格，所以要把句子文本分词，本研究使用ICTCLAS分词器分词，因为其可以标注词性和导入词典辅助算法分词；（3）错误分词和错误词性的调整。由于本研究在发现隐式属性的时候要基于词性规则发现领域情感词，所以错误词性也要调整。

1.2 女装上衣通用商品属性的确定

淘宝上网，一般会总结商品评论比较集中谈论点，如图2：



100图 2 淘宝网上对评论的总结

淘宝对于评论的总结忽略了一些重要属性，如物流、服务等；并且其把同一属性多种表达编在一起如：“颜色漂亮”、“不会褪色”、“色差”都是属于颜色属性；再次，其抽取的文本一般是属性与情感词连接在一起的，但是在评论中属性不一定和情感词连接在一起。我们通过对数款销量比较大评论比较多的商品进行词频统计，然后结合专家意见总结出了通用女装上衣商品属性，我们把商品属性分为三大类如图 3：

105

商品属性划分	
产品属性	款式
	颜色
	尺码
	材质
	衣服局部
	效果
	薄厚
	质量
服务	店家客服
	物流服务
市场	价格

图 3 通用属性



图 4 词素“价”

110 1.3 属性提取及归类

商品属性包括显式和隐式属性，所以要对这两类属性分别提取及归类。

1、显式属性提取与归类。显式属性基本是名词，所以我们用正则表达式从分词并词性标注后的评论中提取这些名词，然后本研究参照文献【7】通过词素方法来进行归类，因为具有相同词素的词在概念上也相近。图 4 说明了通过词素“价”找到的同类显示属性。文献

115 【7】还用 hownet 相似度来找到同类属性, 本文在研究过程中发现很多名词相似度得分非常接近, 但其类别不一样, 这导致通过相似度进行属性归类不准确。

由于“薄厚”属性相关词频都是形容词或者副词, 所以我们暂时只给出属性划分, 相关词频将在隐式属性总结时给出。如下图:

产品属性							服务			
款式	颜色	尺码		材质		质量	薄厚	店家客服	物流服务	价格
款式 = 566 款 = 40 样式 = 39 样子 = 36 新款 = 14 版型 = 10 版型 = 10 式样 = 6 款型 = 5 薄款 = 4 型号 = 2	色差 = 162 颜色 = 557 色差 = 162 绿色 = 50 色彩 = 7 粉色 = 5 黄色 = 3 亮色 = 3 红色 = 3	大小 = 121 号 = 45 尺码 = 44 码 = 33 M号 = 31 S码 = 30 M码 = 25 L码 = 22 尺寸 = 20 号码 = 18 L号 = 13 S号 = 12 码子 = 11 码数 = 10	小号 = 8 中码 = 8 小码 = 8 大码 = 7 大码 = 7 s码 = 6 m号 = 6 长短 = 5 长度 = 5 s号 = 5 m码 = 3 XL码 = 2 大号 = 2 XXL大号 = 1 XXL码 = 1	料子 = 139 面料 = 95 布料 = 53 手感 = 40 材质 = 38 质感 = 11 质地 = 10 料 = 7 雪纺 = 7 雪纺 = 7 材料 = 5	裙子 = 436 外套 = 374 腰带 = 126 围巾 = 91 上衣 = 66 裙 = 36 裙摆 = 26 外衣 = 21 连衣裙 = 21 袖子 = 19 裤 = 18 衣领 = 8	皮带 = 8 上身效果 = 100 背心 = 8 效果 = 98 套装 = 7 整体效果 = 20 套裙 = 7 总体效果 = 2 领子 = 6 围脖 = 6 蕾丝 = 6 长袖 = 6 丝巾 = 5 扣子 = 5 丝巾 = 5 装饰 = 5 吊坠 = 5 下摆 = 5 配饰 = 5	质量 = 1417 做工 = 86 线头 = 24	服务 = 1650 态度 = 407 包装 = 391 客服 = 217 售后 = 24 包装盒 = 20 售后服务 = 4 售前 = 4	物流 = 815 发货 = 529 快递 = 530 速度 = 427 顺丰 = 189 运费 = 35 包裹 = 34 邮费 = 28	价格 = 186 价 = 67 价钱 = 60 性价比 = 44 价位 = 27 特价 = 12 差价 = 9 原价 = 7 折扣 = 6

120 图 5 显式属性归类

2、隐式属性提取与归类。

文献【7】尝试着用改进的 PMI 方法通过显式属性来找隐式属性, 由于隐式属性一般是直接指明属性的情感词, 所以评论中一般不会出现同时出现显示属性和隐式属性。文献【8】通过创新算法找到隐式属性, 但是对于低频词处理为噪声, 往往忽略重要的隐式属性。我们发现隐式属性大多与领域有关, ICTCLAS 分词器算法不能很好识别, 其把这些意思完整的词切成词素(字)。文本利用这个规律用基于词性组合的规则发现这些隐式属性。规则如下图:

组合1	(>1)/v	(>1)/n
组合2	(>1)/d	(>1)/a
组合3	(>1)/a	(>1)/a
组合4	(>1)/v	(>1)/v

图 6 词性规则

上述(>1)代表大于 1 个中文字。我们通过这些规则得到图 7 中情感词:

中文负面评价词语			中文正面评价词语		
有色差	可气	刮毛	没有色差	显皮肤	给力
有差异	坑爹	勾线	有气质	挺好	有效果
显肚子	遭报应	有褶子	无色差	很好	收身
起球	气死人	掉线	有创意	还好	有品味
显腿	哑巴吃黄连	有皱	有心	挺好	有活力
开线	太气人	冤枉钱	衬皮肤	才好	遮小肚子
有差距	含水分	刺皮肤	有心意	刚好	遮住肚子
有味道	有气味	挂纱	没问题	挺好看	显青春
走形	勾毛线	掉毛	超出期望值	蛮好	显身条
走光	起蕾	挑皮肤	有型	极好	有朝气
有区别	脱纱	磨毛	无色差	正合身	毫无色差
有问题	起纱	拉丝	没色差	不肥	适合皮肤
挂毛	刮丝	刮纱	有质感	很棒	正品
勾丝	散线	太差	垂感	最合适	有情调
没什么档次	破线	略短	坠感	正好	失个性
掉皮	开缝	非常一般	显身材	好嫩	衬肤色
有线头	吭货	不对称	上档次	适合	养眼
退色	不一样	有点糙	有档次	显档次	起效果
起毛	偏大	没诚信	有香味	修身	有个性
跳线	太大	偏黄	有心思	适合肤色	提高档次
没有质感	略大	偏绿	有亲和力	有风格	显腰
脱线	偏小	挺俗	合心意	减龄	没什么色差
有偏差	偏暗	臭美	显气质	有特色	
勾纱	偏胖	好暗	收腰	显肤色	
有误差	有点暗	显岁数			
走线	太胖	掉颜色			
没有档次	太短	太薄			
挂丝					

图 7 词性规则结果

基于词性规则我们找到图 7 中的情感词，然后我们从图 7 中选出隐式属性词，因为在线评论中隐式属性词往往为指明属性的情感词。然后我们用正则表达式提取分词后的形容词以及动词，从它们中找到其他隐式属性词。汇总之后得到文本全部隐式属性词。如图 8：

产品属性							市场	
款式	颜色	尺码	材质	效果		质量	薄厚	价格
好看 = 557 挺好看 = 108 洋气 = 47 清新 = 29 大气 = 15 靓丽 = 10 惊艳 = 10 难看 = 9 美观 = 3 性感 = 1 鲜亮 = 1 整洁 = 1 苗条 = 1 超帅 = 1 高贵 = 1 好嫩 = 1	没有色差 = 112 无色差 = 59 有色差 = 54 没色差 = 28 毫无色差 = 1 掉色 = 30 褪色 = 12 退色 = 6 掉颜色 = 2 脱色 = 1	小 = 1066 合身 = 298 合适 = 287 短 = 152 长 = 91 偏大 = 89 太短 = 33 宽松 = 15 正合身 = 3 窄 = 1 略大 = 1 不肥 = 1 肥大 = 1 最合适 = 1 宽余 = 1 贴身 = 1 量身订做 = 1	结实 = 1 有点糙 = 1 质优 = 1 平整 = 1 柔和 = 1 柔顺 = 1	显瘦 = 78 修身 = 64 洋气 = 47 有气质 = 32 显气质 = 28 上档次 = 27 显身材 = 19 有型 = 8 显老 = 8 有档次 = 6 显肚子 = 4 显档次 = 3 减龄 = 3 有质感 = 3 衬皮肤 = 2 显肤色 = 2 显腿 = 1	适合肤色 = 1 遮小肚子 = 1 没有档次 = 1 遮住肚子 = 1 没有质感 = 1 显青春 = 1 显身条 = 1 显肥 = 1 适合皮肤 = 1 显岁数 = 1 失个性 = 1 没什么档次 = 1 衬肤色 = 1 养眼 = 1 起效果 = 1 有个性 = 1 显皮肤 = 1	勾丝 = 15 掉皮 = 1 皱 = 12 缩水 = 1 起毛 = 10 挂纱 = 1 起球 = 10 掉毛 = 1 变形 = 6 刮毛 = 1 抽丝 = 5 磨毛 = 1 脱线 = 5 刮纱 = 1 钩丝 = 4 有褶子 = 1 开线 = 4 勾纱 = 1 钩丝 = 4 勾毛线 = 1 挂丝 = 3 起雷 = 1 拉丝 = 3 脱纱 = 1 勾线 = 3 走形 = 1 抽纱 = 2 起球球 = 1 掉线 = 2 起纱 = 1 挂毛 = 2 破线 = 1 跳线 = 2 开缝 = 1 走线 = 2 皱皱的 = 1 起皱 = 2 刺皮肤 = 1 有皱 = 1	薄 = 362 太薄 = 55 厚 = 31 单薄 = 15 厚实 = 6 轻薄 = 1	贵 = 251 便宜 = 54 划算 = 39 物有所值 = 39 超值 = 34 实惠 = 27 实用 = 27 物超所值 = 17 物美价廉 = 6 廉价 = 6 物美价廉 = 6 价廉物美 = 5 货真价实 = 2 合算 = 2 省钱 = 1 划不来 = 1 值 = 131 优惠 = 67 降价 = 15 便宜 = 14 优惠 = 1 划得来 = 1

图 8 所有隐式属性

1.4 领域情感词典建立

本研究的词典资源来源于三部分（1）、用 Hownet 词典里面的正负向情感词匹配分词后的评论得到词，并从中剔除掉不对领域属性修饰的情感词，如“欣喜若狂”、“自信十足”等修饰消费者自我心情的情感词；（2）2.3 中通过词性规则找出来的情感词、（3）评论中出现但是 Hownet 不包括的情感词，如“价廉物美”、“物超所值”等。这三部分词分正负向保存为两个词集。这就是我们自建的领域情感词典。

1.5 情感强度计算

目前，在线评论情感强度的分析相对薄弱^[2]。在中文情感强度方面比较有代表性的是：Guohong Fu 和 Xin Wang 通过运用模糊集思想基于字的情感计算能够让对情感强度进行量化^[9]；施寒潇在基于字的情感计算基础上把情感词分类，然后根据情感词的类别计算词的情感强度，使得情感强度计算量化更加细致^[10]。以上方法都从不同的角度取得一定效果，但是略显复杂。

在中文在线评论中一般比较倾向“价格有点贵”而不是“很高的价格”这种表达^[7]。淘宝网上自身也是利用这种规则来总结的，如图前述图 2 中的“款式漂亮”、“材质很好”等。其他的就是直接指明属性的隐式词汇“很修身”等。本研究利用此规律计算属性情感强度。具体来说算法是：

（1）先遍历每条分词后文本，匹配文本中所有出现的显式属性，在显式属性之间找到从正负两个情感词集匹配相应情感词，一个正向词记情感强度+1、一个负向词记情感强度-1，得到显式属性情感强度； 此时的情感词典要除去图 8 中的隐式属性情感词；

（2）用图 8 中的词直接匹配文本，一个正向词记情感强度+1、一个负向词记情感强度

-1, 得到隐式属性情感强度;

160 (3) 然后在文本中找到各个否定词, 使得对应属性情感倾向反转;

说明: ; 对于有并列关系的属性, 我们取其情感强度为相同; 对于有几个显式属性串联在一起的属性, 本文算法中取最后一个显式属性作为最终属性; 否定词从 Hownet 词典里面提取; 为了不重复计算情感强度, 我们在为显式属性匹配情感词的时候需要去掉隐式属性情感词。

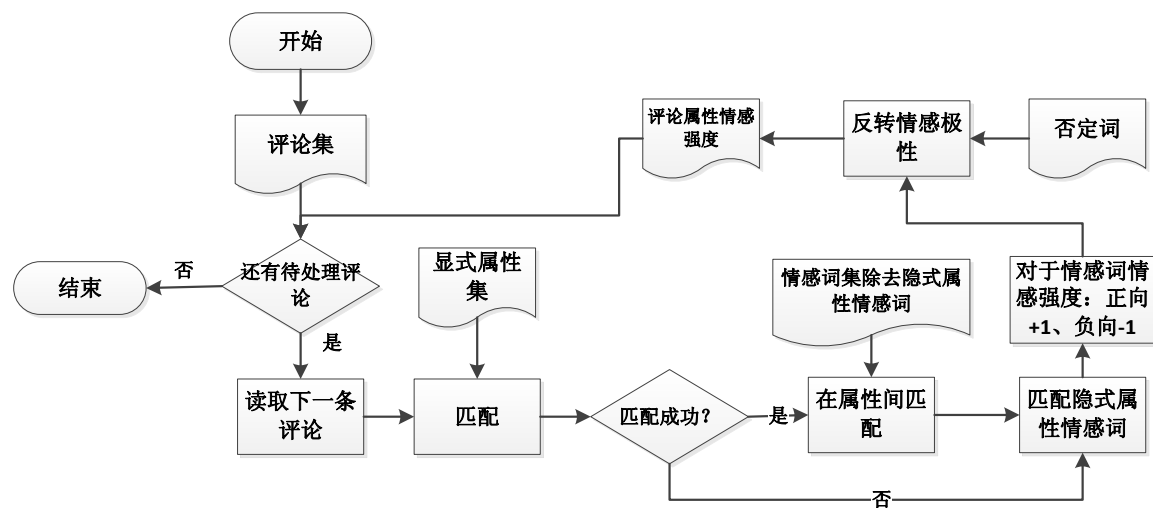


图 9 属性情感算法

2 结果以及分析

为了对比隐式属性的有效性, 我们按照 Fig10 的算法, 计算出显式属性情感强度, 以便于比较 1.5 小节的 Fig.9. Fig10 的算法相对于图 9, 在匹配显式属性情感词的时候没有去掉隐式属性情感词, 并且在其后也不计算隐式属性。这样, 隐士属性词会被当作普通情感词用来计算显式属性情感强度。

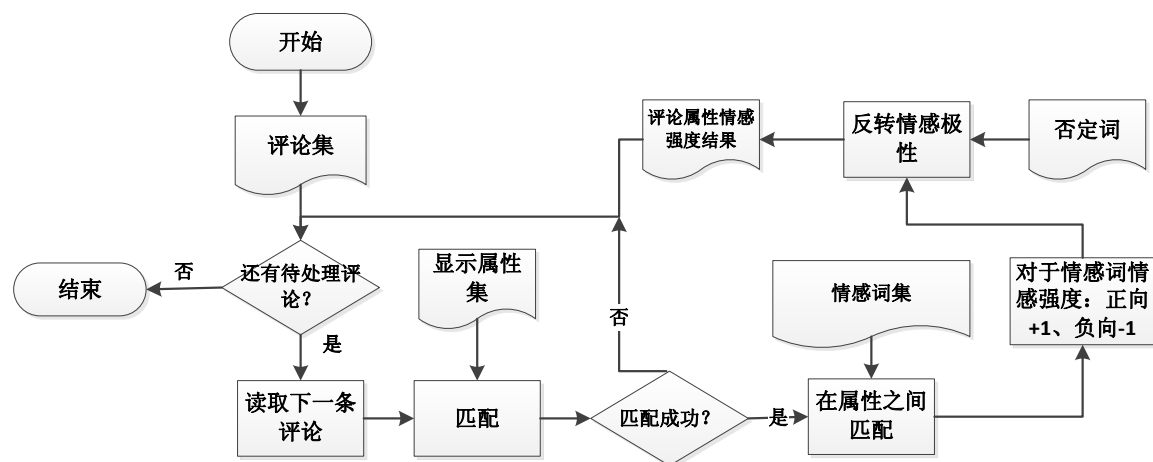


图 10 对比算法

图 11 为图 9 算法和图 10 算法结果对比, 其中折线图为图 9 算法, 其得到的是显式和隐式属性汇总的正向结果, 柱状图为图 10 算法, 其得到的是正向显式属性情感强度。横轴为我们划分的通用商品属性, 其纵轴为每个属性在情感强度值上的文本数量, 每个属性会有三种情感强度得分: 柱形图从左至右分别表示图 10 算法得到的某属性情感强度得分为 1、2、3 的文本数量; 折线图从上至下表示图 9 算法得到的某属性情感强度得分为 1、2、3 的文本数量。图 12 为两种算法在负向情感结果, 其含义与图 11 相同。

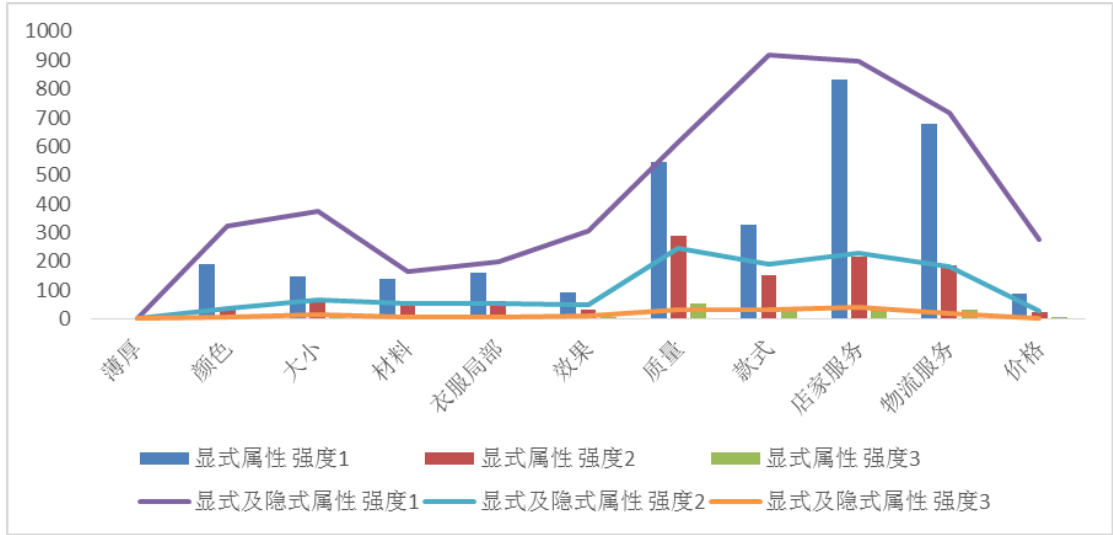


图 11 正向情感对比

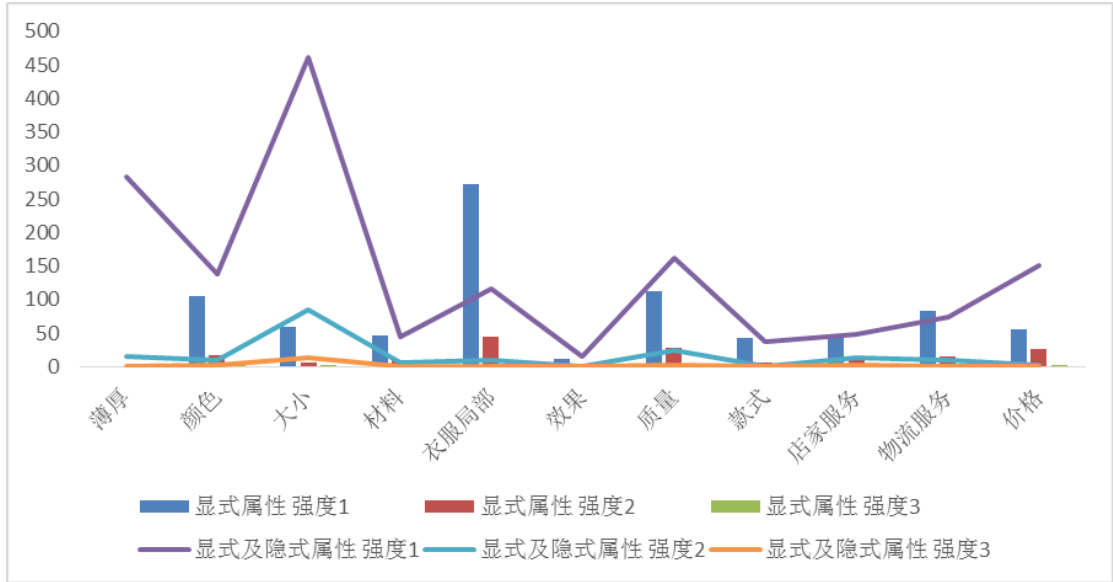


图 12 负向情感对比

在图 11 和图 12 中，每个属性的总得分，为该属性的三种情感强度值乘以其对应文本数量，然后在求和得到，如下图：

类别	属性	属性类别	情感文本数量		百分比 (%)		总的情感得分
			正向	负向	正向	负向	
产品属性	薄厚	显式属性					
		显式和隐式属性	4	300	1.31	98.69	-314
	颜色	显式属性	240	124	65.93	34.07	157
		显式和隐式属性	363	152	70.48	29.52	243
	尺码	显式属性	225	69	76.53	23.47	245
		显式和隐式属性	457	563	44.8	55.2	-122
	材质	显式属性	204	56	78.46	21.54	218
		显式和隐式属性	224	52	81.16	18.84	231
	衣服局部	显式属性	236	319	42.52	57.48	-46
		显式和隐式属性	265	127	67.6	32.4	205
	效果	显式属性	132	12	91.7	8.3	169
		显式和隐式属性	365	16	95.8	4.2	418
	质量	显式属性	892	142	86.27	13.73	1125
		显式和隐式属性	894	190	81.72	18.28	992
	款式	显式属性	506	49	91.17	8.83	660
		显式和隐式属性	1141	39	96.7	3.3	1361
服务	店家服务	显式属性	1106	57	95.1	4.9	1385
		显式和隐式属性	1180	63	94.93	5.07	1464
	物流服务	显式属性	900	101	89.91	10.09	1041
		显式和隐式属性	924	85	91.57	8.43	1061
	市场	显式属性	117	85	57.92	42.08	39
		显式和隐式属性	304	155	66.23	33.77	174

图 13 属性情感强度得分

图 13 便是我们要得出的细粒度情感分析的结果。

首先，我们看到，在加入了隐式属性情感词计算后，细粒度情感分析变得更加准确。例如对于“尺码”这个属性，从图 13 可以看到，对于图 10 算法得出的总情感强度为+245，图 9 算法得出的总情感强度为-122，这是由于图 10 算法在计算情感强度的时候，没有把隐式属性情感词划分到其本身指明的属性下去，导致了属性情感强度计算错误造成的，图 9 算法针对此问题做了改进，可以看出对隐式属性进行计算，可使细粒度情感分析变得更加准确。另外一个属性“衣服局部”在加入隐式属性后，也出现了这种情感强度值正负转向，也是前述道理。同时，由于“薄厚”属性在文本中基本都靠上下文来判断，在加入了隐式属性情感词计算后，细粒度情感分析变得更加准确。

其次，从图 13 我们可以清楚了解到了各个属性正负向文本数量以及总的情感得分（显式和隐式属性结果），并且结果也接近我们真实购物——这款连衣裙的“款式”、“衣服效果”、“店家服务”、“物流服务”这四个属性让消费者非常满意，差评率分别都低于 10%；“衣服尺码”和“衣服薄厚”这两个属性总得分均为负，说明在选择这件商品的时候要根据自身情况向店家询问清楚后做出决策以避免失败的购买——我们已经达到了目的，即减少了信息载荷和决策风险。

再次，“质量”、“衣服局部”及“价格”这三属性负面评价是仅次于“尺码”及“薄厚”属性的（图 12）。消费者在评价这三个属性时会比“尺码”及“薄厚”的评价更加理性，以为后者主观性更强，所以这三个属性负面评价对于生产商来讲改进意义更大。虽然这三个属性总体情感强度值为正，但是“价格”和“衣服局部”负面评论都占比超过该属性总评论数 30%，显然要引起生产商改进的；另外，“质量”属性负面评论数量为这三个属性中最高的，达到 161 条，说明质量合格率有待提高。我们通过这些属性情感强度分析确实可以帮助生产商明确商品改进之处。

最后，生产商不仅可以从图 13 以及 Fig.9~10 中结果了解哪些商品属性需要改进，还可以通过图 8 中对隐式属性的评价发现小众需求，这是由于图 8 中隐式属性情感词都是针对属性的具体评价。从图 8 的“效果”属性我们发现，“修身”、“显瘦”、“显气质”等占多数，反映消费者对商品需求为衣服修身、时尚；对于“质量”属性，绝大部分评价是“钩

丝”、“走线”，反映消费者对商品需求为衣服做工更加精细。通过这样分析，生产商可以明确小众需求，进而做出改进决策：如设计多款修身、时尚的商品，满足巨大的长尾市场

3 结论

本研究以淘宝网上某款连衣裙为例，阐述了细粒度情感分析巨大价值：一方面，不仅可以帮助生产商明确具体改进的地方（商品属性），还可以使其挖掘小众需求；另一方面，帮助消费了解商品各个属性优劣，从而更加明智地决策。本研究贡献在于指明了加入隐式属性后，确实对细粒度情感分析准确性有提高；同时通过隐式属性评论的统计有助于生产商发现小众需求，这对于拥有巨大长尾市场的电商来讲，意义重大；最后，由于我们所使用的在线评论抓取方式、领域通用商品属性划分、基于词性规则的方法发现隐式属性等方法的通用性得本研究情感强度分析方法更加具有普遍性。

225 [参考文献] (References)

- [1] Chen Y B, Xie J H. Online consumer review: word of mouth as a new element of marketing communication mix[J]. Management Science, 2008, 54(3): 477-491.
- [2] 王洪伟, 郑丽娟, 尹裴. 在线评论的情感极性分类研究综述[J]. 情报科学, 2012, 30 (8) : 1263-1271.
- [3] E Brynjolfsson, M D Smith. Frictionless commerce? a comparison of internet and conventional retailers[J]. Management Science, 2000, 46(4): 563-585.
- [4] A Duric, Song F. Feature selection for sentiment analysis based on content and syntax models[J]. Decision Support Systems, 2012, 53(4): 704-711.
- [5] Hu M Q, Liu B. Mining Opinion Features in Customer Reviews[A]. Hu M Q. In proceedings of the national conference on artificial intelligence[C]. Menlo Park: Aaai
- [6] 来亮, 钱屹. 文本情感分析综述[J]. 计算机光盘软件与应用, 2012, 15 (18) : 74-75.
- [7] Zhang W H, Xu H, Wan W. Weakness finder: find product weakness from chinese reviews by using aspects based sentiment analysis[J]. Expert Systems With Applications, 2012, 39(11): 10283-10291.
- [8] Liu L Z, Lv Z X, Wang H S. Extract product features in chinese web for opinion mining[J]. Journal Of Software, 2013, 8(3): 627-632.
- [9] Fu G H, Wang X. Chinese sentence-level sentiment classification based on fuzzy sets[A]. Fu G H. In proceedings of COLING[C]. Beijing : Coling
- [10] 施寒潇. 细粒度情感分析研究[D]. 苏州: 苏州大学, 2013.