浙江理工大学学报，2021，45(1)：21-30
Journal of Zhejiang Sci-Tech University
DOI:10.3969/j.issn.1673－3851(n).2021.01.003

基于服装网购评论的消费热点情报分析

胡觉亮[a]，徐瑶瑶[b]，董建明[a]
（浙江理工大学，a.理学院；b.服装学院，杭州 310018）

摘　要：为有效指导服装企业生产经营决策，选取服装网购评论为数据样本和研究对象，提出了基于服装网购评论的消费热点情报分析方法，以探究消费者对所采购的服装的关注热点。采用网络爬虫技术采集服装网购评论数据并进行预处理后，利用 SnowNLP 技术进行情感倾向性分类。在关键词抽取时针对分词过程存在错分专业词汇问题，引入服装专业领域分词词典和停用词典，并结合 GooSeeker 方法和人工调整方法进一步提高关键词抽取精准度。关键词抽取后建立共词矩阵，并聚类进行社会网络分析得到消费热点情报信息。以真丝服装网购评论进行实证分析以验证方法的有效性。结果发现：真丝服装网购消费者依次易就面料、颜色、尺码、质量等热点关键词给出负面反馈；此外还得到这些热点关键词关联的负面反馈信息及与其他热点关键词之间的相互关系，如面料的负面反馈主要与薄透、褶皱、缩水和引申的丝料价格贵有关，对面料差评的消费者往往会更关注尺码、物流、价格以及退换货等信息。

关键词：服装网购；评论；消费热点；情感分类；关键词抽取；社会网络分析
中图分类号：TS941；TP391　　**文献标志码**：A　　**文章编号**：1673-3851 (2021) 01-0021-10

Consumer focus intelligence analysis based on clothing online shopping comments

HU Jueliang[a], *XU Yaoyao*[b], *DONG Jianming*[a]
(a. School of Science; b. School of Fashion Design & Engineering,
Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In order to effectively guide the production and operation decision-making of clothing enterprises, clothing online shopping comments were selected as data samples and research objects. Besides, consumer focus intelligence analysis based on clothing online shopping comments was proposed to explore consumer focus on clothing. After web crawler technology was applied to collect data on clothing online shopping; comments and preprocess them, SnowNLP technology was employed to classify emotion tendencies. During keyword extraction, there was a problem of misclassification of professional vocabulary in the process of word segmentation. Hence, the word segmentation dictionary and out-of-use dictionary in the clothing field were introduced. Meanwhile, GooSeeker method and the manual adjustment method were combined to further improve the accuracy of keyword extraction. After the keywords were extracted, a common word matrix was established and clustered for social network analysis to obtain consumer focus intelligence information. An empirical analysis was carried out for online shopping comments on silk clothing to verify the effectiveness of this method. The results show that, the online shoppers of silk clothing gave negative feedback on such hot keywords as fabric, color, size, and quality. In addition, the

收稿日期：2019－10－28　　**网络出版日期**：2020－11－05
基金项目：国家自然科学基金项目(11971435)；浙江省自然科学基金项目(LY18A010029)
作者简介：胡觉亮(1958－　)，男，浙江杭州人，教授，主要从事运筹学、服装供应链应用方面的研究。
通信作者：董建明，E-mail：djm226@163.com
(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

negative feedback information associated with these hot keywords and the interaction with other hot keywords were obtained. For example, the negative feedback of fabrics was mainly related to the thinness, wrinkles, shrinkage and the extended price of silk materials. At the same time, consumers who gave bad comments on fabrics tended to pay more attention to size, logistics, price, change and return.

Key words: Clothing online shopping; comments; consumer focus; classification of emotions; keyword extraction; social network analysis

0 引 言

近年来随着服装行业与电子商务融合得愈加紧密,服装线上零售已成为人们日常生活中不可或缺的一部分。与传统的服装线下销售方式相比,服装线上零售具有可选择种类丰富、方便快捷、价格优惠等诸多优势,与此同时各大电商平台上也产生了成千上万条服装网购评论文本。这些评论包含着消费者对服装产品的真实反馈,蕴含着大量商业竞争情报信息,如对不同品牌和不同类别服装的消费者关注热点、消费者的购买意愿、销量与评价之间的关系等信息。为了获取这些信息,传统的商业竞争情报信息获取方式主要是采访、调研问卷等,例如郭惠玲[1]通过问卷星网站和门店现场问卷调研方式,收集消费者对快时尚品牌的款式设计、质量等产品属性的打分,以此来获取消费者满意度方面的情报信息。由于问卷调查方法耗时长、成本高且信息较难获取,因此免费、公开、易获取的网购评论数据逐渐成为获取商业竞争情报信息的重要来源。如何从服装评论文本中提取情报信息以帮助服装企业了解行业热点、快速掌握消费者需求,以便企业运营者及时有效地应对消费市场变化,成为了服装企业急需解决的问题。

传统的商业情报分析遍及各大行业。Park等[2]提出了一种基于专利分析的情报系统,并通过实例证明了该系统在商业技术竞争环境下的优势。Yu等[3]将博弈论概念引入竞争情报筛选模型。方友亮等[4]将SCP范式理论引入到竞争情报分析中,并在新能源汽车领域进行了实例研究。Köseoglu等[5]针对酒店行业提出了一种整体竞争情报模型,为酒店管理人员提供形式化指导。情报分析的角度较多,例如,可以从环境角度分析企业面临的政策法规和市场竞争等外部环境,或技术、人才和企业战略等内部环境,也可以从消费者信息角度研究消费者画像(包括年龄、性别、喜好等)、消费者品牌偏好及忠诚度、消费者心理及行为等[6],由此产生了多种情报分析技术和方法,主要有SWOT分析法、文献检索法、专利分析法、反求工程法以及问卷调查、访谈法等方法。SWOT分析法[7]通过对企业内外部环境综合分析得到企业竞争态势,但在一定程度上忽略了企业的主动性且依靠情报分析人员的专业素质,客观性较弱。文献检索法[8]则是对政府公开数据、出版物及其他公开数据源进行收集分析获取情报,其缺点是工作琐碎且公开数据具有一定时滞性,不利于对未来竞争态势的分析。专利分析法[9]虽然在一定程度上能够识别竞争对手及了解技术发展方向,但是一些公司会选择技术保密,公开的专利并不能代表其实际情况。反求工程法[10]是指获取竞争产品并进行技术分析,对比自身产品优劣势进行整改,此方法较为传统且需要花费一定的金钱和精力。而问卷调查及访谈法[11]则是在对问卷内容整理分析后依靠专家经验知识分析判断,其数据真实可靠性较高且具有针对性,但耗时费力,因此局限性较高,不适合广泛开展。

随着互联网的普及特别是线上购物在国内外的盛行,自Rocker等[12]开始利用互联网获取商业情报信息的研究开始,国内外研究者开始对网购评论数据等信息加以利用和分析,以获得关于消费者、商品和市场表现等商业情报信息。Chung等[13]设计了一个基于粗糙集理论的商业智能系统,用于分析网购评论文本来获取客户所关注的产品特征和具体需求信息。靳健等[14]首先对产品评论进行特征抽取和情感分类,然后对已标记数据进行CRF模型训练和评估,最后对手机产品评论进行实证分析获取情报。聂卉等[15]利用词向量和句法依存技术提取在线评论特征词及用户观点,然后基于情感词典量化对情感值进行量化,从中得到自身及竞争对手优劣势情报。相比于传统的商业情报分析方法,基于网络评论数据的分析方法是一种更加直接有效的方法,并且正在不同行业得到广泛应用。

服装网购评论数据作为商业竞争情报分析的重要依据,越来越受到行业研究者和服装企业的重视。目前针对服装网购评论的情报分析研究主要集中在对网购评论文本从某个角度进行分析,例如从服装

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

网购评论的数量、评论长度、评论方向、信息质量、丰富度等方面研究顾客购买意愿、购买决策、感知风险和与销量之间的联系等问题。如韩立娜[16]主要研究了在线评论的数量、质量对服装消费者购买意愿的影响，以及消费者在选择服装时最关注的在线评论内容要素，旨在完善服装评论内容系统、提高评论系统的有用性，为在线经营者提供营销建议。而针对服装网络评论数据本身进行文本处理情报挖掘的研究较少，目前相关研究主要从情感分类的角度进行，例如 Zhang 等[17]以亚马逊网站上的中文服装评论文本为例，提出基于 word2vec 和 SVM^{perf}的情感分类方法，将评论文本分为积极评论和消极评论，帮助消费者和商家从互联网上大量的产品评论中得到对产品的综合和全面的评价。

本文从服装网络评论文本分析出发进行商业情报分析，设计情感分类和社会网络分析等文本挖掘方法，研究特定网络平台销售的品牌服装及评论数据，进行可视化挖掘分析，得到关于服装消费者对于所购服装的关注热点属性，以及各热点属性之间的关系，从而得到相关情报分析结果，为企业决策提供支持。

1　服装网购评论数据及相关处理技术

本文以服装网购评论文本作为研究对象，针对线上服装消费者关注热点进行商业情报分析。由于服装具有款式和面料等多种属性，且具有季节性等特点，所以服装网购评论数据与传统商品相比，不仅具有一般网购评论的特点，而且评论数据蕴含的信息更为丰富。

1.1　从情感视角区分评论数据

网购评论数据不可避免地会涉及消费者对产品本身及其他服务等方面的评价，包括正、负面的情感，在一定程度上会影响潜在消费者的购买决策。由于服装的属性众多，且通常女性的服装网购参与度较高，同时考虑到消费者在积极评论和消极评论中的关注热点属性可能不尽相同，因而难以区分商品某一方面是易得到负面评价还是正面评价。因此，本研究采用情感分析技术，先将评论文本进行积极和消极的正负面情感分类，再分别进行情报分析。

情感分析技术主要对计算文本中表达的观点和态度分析其情感表现，在网络舆情分析、消费者满意度调查等多方面具有广泛应用。情感分析技术根据研究方向一般可分为情感倾向性分类、情感强度分类和主客观分类，其中情感倾向性分类又称正负情感极性分类，即对文本中的主观性表达进行分析提取并将其归纳为正面或负面评价。目前商品评论情感倾向性分类方法主要有两种：基于情感词典的分类方法[18]和基于机器学习的分类方法[19-20]。前者需要大量可靠的情感词典对文本段落进行拆分计算，然后制定规则，计算出整个文本的情感倾向，对人工及专家的经验依赖度较高。而后者通过标注极性明显的正负文本进行训练，最终得到可使用的情感分类器进行情感倾向分类，是目前主流的情感倾向性分类方法。常见的基于机器学习的情感倾向性分类方法主要有支持向量机、决策树、朴素贝叶斯方法等。其中，支持向量机的情感倾向性分类方法难以适用大规模训练样本，且对缺失的数据较为敏感。决策树不但难以处理缺失数据，而且容易过度拟合。而朴素贝叶斯方法则原理简单，分类效果稳定且易于实现。例如，李晓东等[21]和王菲等[22]采用朴素贝叶斯方法对电脑、书籍、电器等网购商品评论进行情感分类研究，分类准确度均达到 90% 左右，显示了其在网购评论情感分类方面的优越性。因此，本研究也通过调用基于朴素贝叶斯分类方法的 Python SnowNLP 类库进行评论情感倾向性分类。SnowNLP 是一个受 TextBlob 启发用 Python 语言编写、专门用来处理中文文本内容挖掘的类库，其使用了朴素贝叶斯机器学习分类方法，在网购商品评论情感倾向性分类方面有着较高的精准率。

1.2　评论文本分析及关键词抽取

网购平台为了便于消费者自由地发表对产品服务及穿着体验的观点和看法，所以对文字评论文本一般没有格式化的规定。消费者可以各抒己见、自由评论，导致评论文本偏口语化，甚至有些消费者还会使用“颜文字”等符号。相比其他商品，由于服装受款式、面料和季节性等多种因素影响，所以服装网购评论文本受消费者个体差异的影响更大，消费者更易给出主观且感性的表达，评论文本更加错综复杂。因此，采用相关技术对服装网购评论文本进行有效的关键词提取尤其重要，例如采用 TF-IDF（词频-逆文档频率）算法[23]。TF-IDF（词频-逆文档频率）算法是一种经典的关键词抽取技术，其通过衡量词语的重要性来进行无监督关键词提取，在文本挖掘方面有着广泛应用。其中词语的重要性与其频率有关，当词语在一条评论中越是高频出现，在整个评论语料库中越是低频出现时，则表示该词语的 TF-IDF 权重越高，重要性越大。

针对服装网购评论数据特点，在数据预处理方面需要进行单独设计。比如，为避免服装领域专有

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

词汇被错误分割,可以建立自定义的服装专业领域分词词典。为剔除过滤与研究无关的信息,还扩充停用词表,减轻后续分析的工作量。经过分词过滤后的具体效果举例如图 1 所示。图中,分词效果 1 是未加载自定义分词词典和停用词表的分词表现,分词效果 2 是加载了自定义分词词典和停用词表的分词表现。两者对比不难发现,分词效果 1 在服装领域的表现结果不尽人意,比如"衬 肤色 显白显 气质"这里没有将"衬肤色""显白""显气质"这三个形容词汇正确地分割开来;"很百搭",也没有将程度副词"很"和形容词汇"百搭"进行正确分割。还有效果 1 中的"方面"作为无研究价值词汇,也没有被过滤停用。而这些方面的不足在加载了自定义分词词典和停用词表的分词效果 2 里得到了明显改善。

例 句: "款式比较简单大方,做工方面也蛮精致的,没有多余线头,颜色好看,衬肤色显白显气质,很百搭哦。"
分词效果1: 款式 比较简单 大方 做工 方面 蛮 精致 没有 多余 线头 颜色 好看 衬 肤色 显白显 气质 很百搭
分词效果2: 款式 比较 简单 大方 做工 蛮 精致 没有 多余 线头 颜色 好看 衬肤色 显白 显气质 很 百搭

图 1 分词过滤效果对比

1.3 消费者网购关注热点的关联性与社会网络分析技术

本文的目标是通过分析消费者关注的服装产品的热点属性,如服装价格、面料质感等属性,了解热点属性之间的复杂关系,以揭示消费者评论数据背后所代表的市场需求变化,帮助商家企业及时调整经营战略。所以如何通过对评论中的热点关键词及关键词之间的关系进行定量刻画,揭示出消费者网购评论热点之间的复杂关系和服装市场趋势变化是关键。

社会网络分析[24]作为一种量化关系、利用关键词构建共词网络、揭示关系结构和规律变化的分析技术,近年来被广泛应用于电子产品、家居等行业进行商业竞争情报分析。例如,徐萌[25]以华为智能手机为例,利用社会网络分析以可视化的方式判别竞争对手的优劣势。张振华等[26]以京东家居产品评论为例,运用 NLPIR 平台及关键词提取后进行共词分析,输入到 ROST CM 以及 NetDraw 软件工具中进行社会网络图绘制,挖掘物流服务问题并提供了相应的改善方法。社会网络分析技术同样也在服装行业得到了应用,例如任佩萱[27]利用社会网络分析方法分析了近十年来我国服装学科已发表文献的研究热点,为后续研究提供参考。廉同辉等[28]选取服装产业集群作为研究对象,利用社会网络分析技术对其网络密度和联结状态进行研究与分析。

本文将社会网络分析技术引入到服装网络评论数据处理中,在将非结构化的服装网购评论数据进行可视化挖掘分析后,以网络结构的方式呈现出其蕴含的情报信息。

2 研究方法设计

本文提出了一种基于情感倾向性分类、关键词抽取以及情报可视化分析等一系列文本处理技术构成的消费者关注热点研究模型,旨在对非结构化的服装网购评论文本数据进行情报挖掘分析,以获取消费者关注热点及热点关键词背后的复杂联系,帮助服装企业有针对性地调整经营战略以占据更多的消费市场。本文的具体研究方法如图 2 所示。

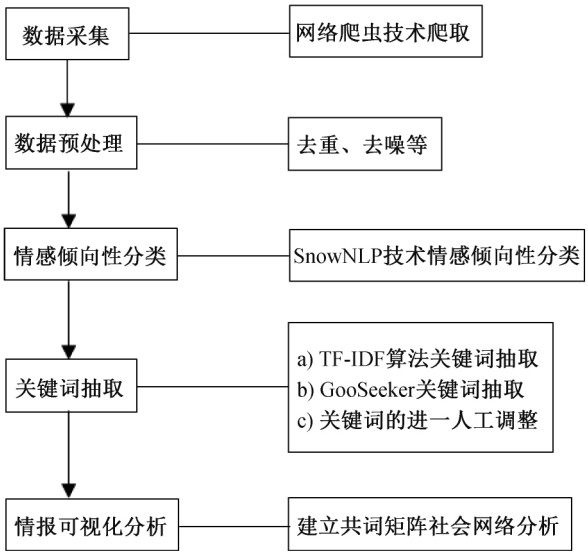

图 2 基于服装网购评论数据的消费热点情报分析研究方法

首先,本文运用网络爬虫技术爬取所需的服装评论数据,获取初始数据。为便于后续数据处理分析,需要先将杂乱的初始数据转化为高质量的一致性数据,具体工作有:数据去噪、去重、删除字数较短的评论、格式转换等。

其次,为更好地区分消极和积极评论,防止因情感主观性而造成后续情报分析在内容上的错乱,本研究引入基于朴素贝叶斯的 SnowNLP 技术进行情感倾向性分类。

再次,为了获取消费者在积极和消极评论中的关注热点,需要分别对经过情感分类的数据进行关键词抽取。本文一方面采用 TF-IDF 算法进行关键词抽取,另一方面又针对 TF-IDF 处理词频、权重、

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

生僻字与常用词等方面存在的不足，采用 GooSeeker 软件进行新词识别和关键词抽取，综合两方面提取的结果，以达到提高特征关键词精确抽取的目的。同时为提高 TF-IDF 关键词抽取的准确率，本研究还在 jieba 分词库的基础上建立了服装网购评论领域的专有分词词典和停用词表。

最后，本文引入社会网络分析技术进行情报可视化分析，对抽取到的关键词进行共词分析，并利用 k-cores、中心度分析、凝聚子群分析等常规社会网络分析技术手段，刻画消费者关注的热点之间的复杂关系和结构变化[28]。社会网络分析图由节点和连接节点的有向线段构成，其中节点代表抽取得到的关键词，有向线段则表示连接关键词的语义关系，即将单独的关键词汇转化为网络构架，通过节点和有向线段，以可视化的方式分析特征关键词的分布和相互之间的联系。

3　方法实现及应用分析

3.1　服装网购评论数据采集

国内服装线上销售平台众多，如淘宝网、京东商城和小红书等。本文选择服装品类相对更广泛、客流量与交易额较大的淘宝商城作为数据采集平台，以“真丝服装”作为关键词进行服装商品搜索，参考销量排名和综合排名，选择 100 件商品，运用网络爬虫技术实现对评论文本数据的获取，最终得到 15245 条评论文本作为初始数据，样本选取日期为 2019 年 1 月 1 日至 2019 年 3 月 1 日。

3.2　数据预处理

数据预处理的第一步是去噪，具体包括识别并过滤评论文本中的无意义符号、表情图标；识别并删除文档中的 URL 链接；辨别并删除虚假营销评论等。

由于消费者对服装评论的重复文本很多，也包括用户超过一段时间没有做出评论，系统自动评价产生的大量重复性文本，比如“好评”“此用户没有填写评价”等，还包括一些用户可能会复制粘贴别人的评论。因此，要对文本去重。本研究主要采用两种去重方法：一是直接删除去重法，二是机械压缩去重法。例如，“衣服质量和版型都很好看，很满意，好评好评好评好评好评好评好评好评好评”经机械压缩后缀去重后则变成“衣服质量和版型都很好看，很满意，好评”。

由于从语言学的角度来说，字数越少能够传达的信息越少，要想表达特定的含义一定需要相应的字数，所以字数过少的评论意义不大甚至没有意义。因此本研究设定 4 个字符的下限，即评论语料若小于 4 个字符，则将该评论语料删除。最后采用 Python 加载 re、codecs 等学习库进行操作，最终获得 13857 条能被有效使用的服装网购评论文本。

3.3　情感倾向性分类

本文采用 Python 封装的 SnowNLP 类库进行服装网购评论情感倾向性分类。本文在对实验数据进行情感分类之前，先进行 SnowNLP 情感分类模型的重新训练提高分类精准率然后再应用，步骤如下：

a）准备服装网购评论样本：积极评论 5000 条，消极评论 5000 条分别存入 pos.txt 和 neg.txt 训练文本中，以及验证集评论 2500 条（训练集 80%，验证集 20%）。

b）加载 SnowNLP 训练新模型，保存至“sentiment.marshal”文件中。

c）利用训练好的模型对验证集进行情感倾向性分析。由 precision（准确率）、recall（召回率）及 F1 值评价指标进行此分类模型的结果分析评估，结果见表 1。

表 1　SnowNLP 情感倾向性分类效果评估

评价指标	precision/%	recall/%	F1 值/%
表现效果	98.45	98.34	96.90

由表 1 可知重新训练过的情感分类模型具有较好的表现结果，可以用于本文研究。加载训练好的 SnowNLP 情感分类模型将经过预处理的实验数据情感二分类后，得到 1406 条消极评论，12451 条积极评论，经人工检验调整最终得到消极评论 1368 条，积极评论 12193 条。

3.4　关键词抽取

3.4.1　通过 TF-IDF 算法抽取关键词

本文应用 Python 语言，调用 jieba 中文分词包，读取数据源，加载服装网购评论领域自定义分词词典以及停用词表（见表 2 和表 3），循环对每一行服装网购评论进行分词过滤，然后运用 TF-IDF 算法根据权重大小排名提取出前 200 关键词，输出关键词及权重到本地文件，具体见表 4 和表 5。

表 2　自定义分词词典示例

序号	词语	词频/条
1	摸起来	2238
2	高端	2245
3	上档次	2246
4	大牌	2249
5	轻飘飘	2294

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 3 自定义停用词表示例

序号	停用词
1	光顾
2	谢谢
3	裙子
4	感觉
5	夏天

表 4 消极评论关键词抽取结果示例

序号	关键词	权重	序号	关键词	权重
1	有点	0.213	11	小贵	0.056
2	真丝	0.162	12	桑蚕丝	0.053
3	颜色	0.155	13	可以	0.053
4	面料	0.129	14	不是	0.048
5	图片	0.116	15	穿上	0.047
6	色差	0.107	16	做工	0.044
7	质量	0.083	17	价格	0.042
8	不值	0.069	18	一点	0.042
9	掉色	0.068	19	退货	0.041
10	没有	0.059	20	料子	0.040

表 5 积极评论关键词抽取结果示例

序号	关键词	权重	序号	关键词	权重
1	舒服	0.269	11	款式	0.119
2	面料	0.250	12	上身效果	0.118
3	喜欢	0.226	13	做工	0.104
4	质量	0.205	14	妈妈	0.086
5	不错	0.196	15	显瘦	0.084
6	真丝	0.176	16	颜色	0.082
7	穿上	0.163	17	柔软	0.077
8	好看	0.154	18	很漂亮	0.075
9	满意	0.144	19	合适	0.073
10	穿着	0.131	20	好评	0.073

3.4.2 GooSeeker 关键词抽取

运用上述 TF-IDF 算法进行关键词抽取可靠性较高,但也存在着一些不足,如简单将词频和权重联系起来,认为某一词汇在其中一条评论中出现的频率越高、在整个评论集中出现的频率越低时,具有较高的区分能力,因此往往一些研究意义不大的词汇会被赋予较高的权重。同样的,一些在评论集中出现频率较高的、有助于研究的词汇,不能因其区分能力低而降低它的权重。而 GooSeeker 是一款多功能的文本信息挖掘软件,在新词发现、提取关键词等方面具有较好的运用。因此,本研究除采取 TF-IDF 方法外,还利用 GooSeeker 进行特征提取来弥补 TF-IDF 算法上的不足。

实验中将情感分类后的实验数据分别导入 GooSeeker,首先利用筛词选词功能分别筛选出消极、积极评论前 200 个高频关键词,然后根据自定义添加词语功能手工补充专有词汇,将关键词分离出来,最终利用其文本分类中的特征选择和特征抽取功能,得到消极评论和积极评论关键词,分别见表 6 和表 7。

表 6 消极评论关键词提取结果示例

序号	关键词	频次	序号	关键词	频次
1	衣服	368	11	收到	93
2	有点	344	12	价格	88
3	颜色	216	13	一般	84
4	没有	171	14	色差	83
5	图片	165	15	一点	67
6	可以	137	16	裙子	66
7	质量	131	17	不值	65
8	真丝	129	18	知道	56
9	面料	124	19	效果	53
10	这个	108	20	那么	50

表 7 积极评论关键词提取结果示例

序号	关键词	频次	序号	关键词	频次
1	衣服	4538	11	裙子	1597
2	喜欢	3912	12	漂亮	1455
3	舒服	3644	13	颜色	1323
4	质量	3504	14	妈妈	1307
5	不错	3066	15	款式	1302
6	面料	2941	16	效果	1282
7	非常	2469	17	上身	1269
8	收到	2289	18	特别	1245
9	满意	2154	19	做工	1228
10	好看	1700	20	合适	1176

3.4.3 关键词的进一步人工调整

综合 TF-IDF 权重大小、GooSeeker 新词及出现频次等各方面结果发现,提取出的关键词能够很好地代表整个评论集,但还需要对其进一步有效区分。因考虑到本文研究消费者关注热点,而消费者对服装的关注热点主要是产品本身属性、物流和服务等辅助属性词,以及由此产生的相关评价。因此本文在提取关键词后,采取人工调整、删除,得到更符合本文研究的关键词,并对近义词进行转化等,最终得到消极评论关键词 141 个,积极评论关键词 138 个,部分提取信息示例如表 8 和表 9 所示。

3.5 情报可视化分析

本文对整理后的关键词进行共词分析并构建共现矩阵,关键词聚类后通过 Ucinet、NetDraw 第三方软件进行社会网络分析。共词分析[29]是一种重要的内容分析方法,是对同一条评论中某对关键词共同出现的频率进行统计,进行分层聚类以揭示这

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 8　消极评论关键词提取结果示例

序号	关键词	词频	权重	序号	关键词	词频	权重
1	真丝	129	0.162	11	做工	47	0.044
2	颜色	216	0.155	12	价格	88	0.042
3	面料	124	0.129	13	退货	35	0.041
4	图片	165	0.116	14	料子	35	0.040
5	色差	83	0.107	15	尺码	30	0.039
6	质量	131	0.083	16	差评	28	0.037
7	不值	65	0.069	17	布料	32	0.035
8	掉色	44	0.068	18	实物	42	0.035
9	小贵	42	0.056	19	客服	35	0.033
10	桑蚕丝	43	0.053	20	效果	53	0.033

表 9　积极评论关键词提取结果示例

序号	关键词	词频	权重	序号	关键词	词频	权重
1	舒服	3644	0.269	11	合身	234	0.072
2	面料	2941	0.250	12	尺码	600	0.071
3	喜欢	3912	0.226	13	料子	648	0.067
4	质量	3504	0.205	14	穿起来	548	0.067
5	不错	3066	0.196	15	手感	629	0.060
6	真丝	874	0.176	16	凉快	588	0.060
7	好看	1700	0.154	17	舒适	722	0.056
8	满意	2154	0.144	18	物流	826	0.055
9	款式	1302	0.119	19	显气质	425	0.052
10	上身效果	980	0.118	20	客服	677	0.048

些关键词之间的亲疏远近关系。一般来说，共现频次越高则代表关键词之间的关系越紧密。本文在陈农[30]和王倩倩[31]的研究基础上实现共词分析、关键词聚类等工作，构建的服装网购消极评论和积极评论关键词共现的部分矩阵如表 10 和表 11 所示。

表 10　消极评论关键词共现矩阵(部分)

关键词	肥大	颜色	图片	质量	面料	短小	小贵	色差
肥大	0	59	28	18	78	8	42	52
颜色	59	0	109	9	82	24	23	156
图片	28	109	0	12	36	16	16	92
质量	18	9	12	0	12	13	31	8
面料	78	82	36	12	0	59	73	53
短小	8	24	16	13	59	0	66	30
小贵	42	23	16	31	73	66	0	25
色差	52	156	92	8	53	30	25	0

表 11　积极评论关键词共现矩阵(部分)

关键词	喜欢	舒服	质量	面料	好看	颜色	款式	效果
喜欢	0	1025	1016	834	496	489	492	331
舒服	1025	0	916	1444	550	389	452	360
质量	1016	916	0	530	472	370	430	343
面料	834	1444	530	0	410	357	466	418
好看	496	550	472	410	0	281	323	133
颜色	489	389	370	357	281	0	220	130
款式	492	452	430	466	323	220	0	108
效果	331	360	343	418	133	130	108	0

通过 Ucinet 进行数据转化以及 NetDraw 可视化分析，辅以 analysis-centrality measures 中心度分析等考察关键词节点的相对重要程度，得到如图 3、图 4 所示的消极和积极评论社会网络分析图。

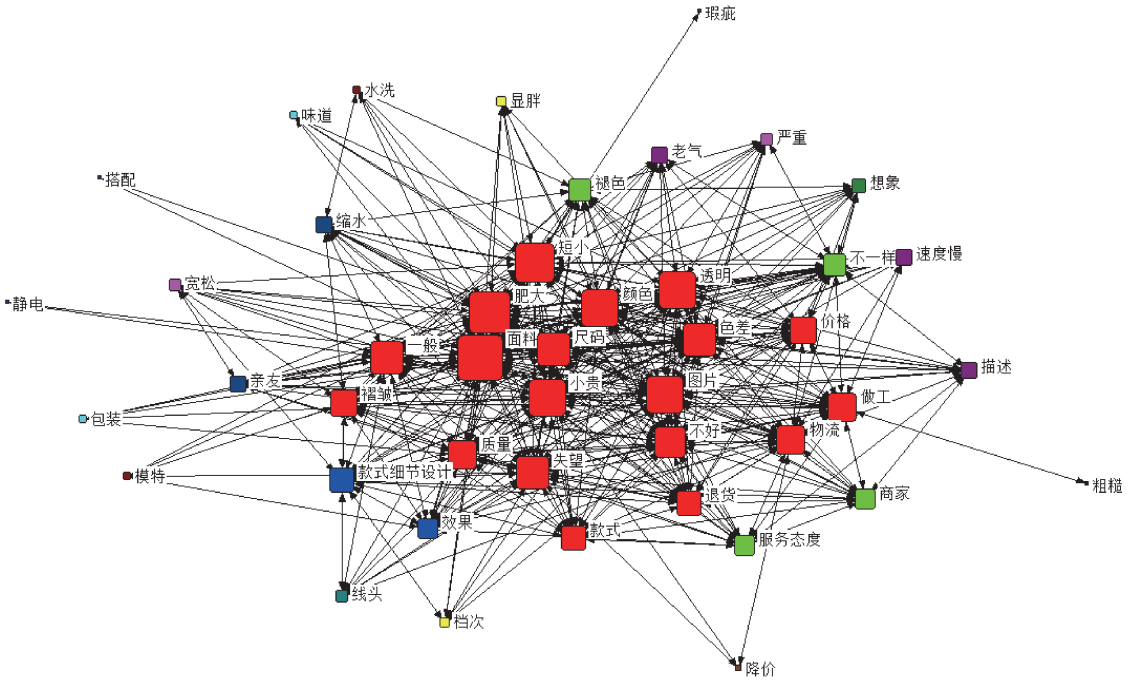

图 3　消极评论社会网络分析图

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

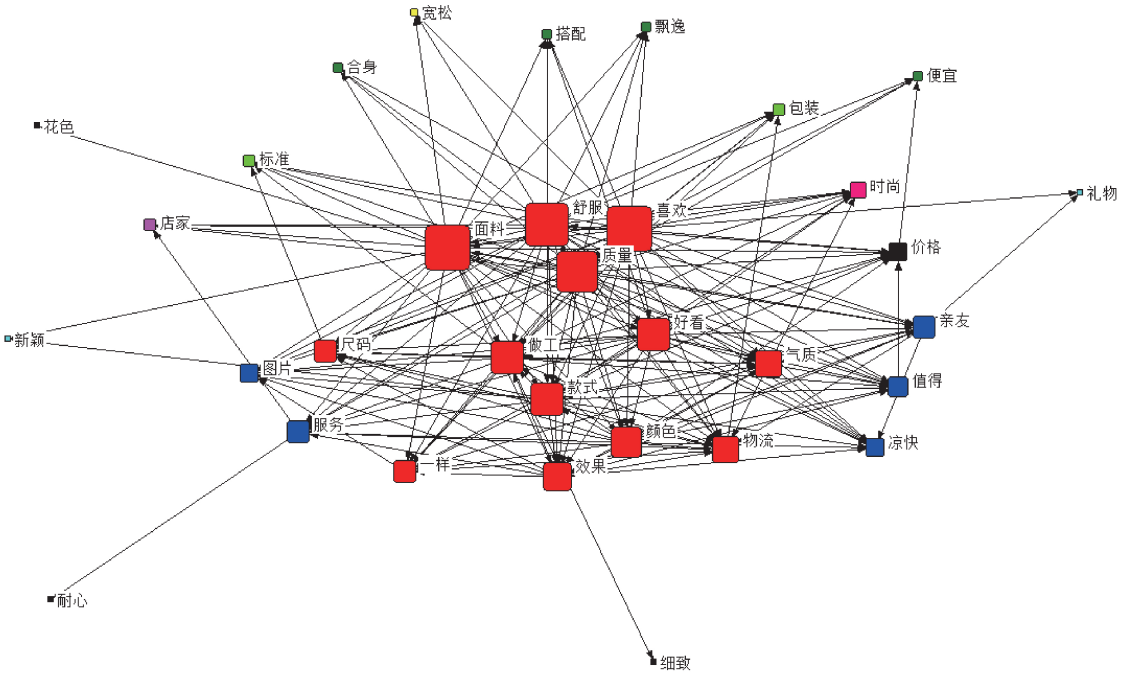

图 4 积极评论社会网络分析图

3.6 案例结果分析

因消极评论对服装网购消费者购买行为意愿影响更为突出,其潜在的商业情报信息更为有价值,所以本研究主要针对图 3 消极评论进行消费者关注热点分析。从整体来看,它是一个中间聚集、四周较为分散的网络关系图。网络结点和有向线段组成一个完整的语义网络,其中箭头的方向表示关键词之间的从属关系,以及关键词与关键词之间的关系。对该网络图进行 k-cores 解析可以发现,消费者在消极评论中提及的重要关键词结点。该网络图中结点位置越靠近网络中心、面积越大则该结点核数越大,即对整个网络的影响越大。从图 3 中可以看到属性关键词“面料”“色差”“颜色”“质量”“图片”“尺码”“物流”“做工”“款式”“价格”“褶皱”核数较大处于且处在中心位置,实现全网络之间的路径链接,促进了不同属性词与评价词之间的信息连通,在整个消极评论网络中占有重要地位。而“包装”“静电”“瑕疵”“搭配”“线头”等属性关键词则受消费者负面关注较少,与其他属性关键词之间关系疏远,处于较为边缘的位置,在整个消极评论网络中影响力较小。因此,需进一步做中心度分析和凝聚子群分析。

3.6.1 中心度分析

通过点度中心度分析可知,该网络点度中心度值为 52.63%,最大点度中心度为 42.00%(面料);其次为 39.00%(肥大)、36.00%(短小)、35.00%(小贵);最小为 1.00%(粗糙)、次小为 2.00%(瑕疵/搭配/静电)。整体网络中心度较高,关键词面料标准化中心度值为 93.33%,与其他关键词有直接联系的有 42 个,处于网络核心地位,与其他关键词的链接控制作用较强。通过中间中心度分析可知,该网络中间中心度指数为 10.12%,最大值为 111.44%,最小值为 0。其中面料、肥大、短小、做工、小贵、一般、颜色、透明中间中心度较高,依次为 111.44%、64.10%、52.49%、48.51%、38.02%、30.78%、28.95%、24.23%,说明其对消费者关注的其他关键词交流之间具有很好的链接共现,很大程度上影响整个网络的衔接与沟通。通过接近中心度分析可知该网络整体接近中心度值为 59.68%,独立性较高,最大值为 109.00%,对应关键词为粗糙,包装、味道、降价、静电、搭配、瑕疵次之,这表示关键词“粗糙”的独立性大,与其他关键词距离较远,处于网络边缘地带,在整个消极评价网络上受关注程度较小。

3.6.2 凝聚子群分析

当网络中一些关键词被消费者同时提及的频率较大时,它们会形成一个关系紧密的次级群体,这种次级群体就是凝聚子群。该网络通过凝聚子群分析得到 7 个次级子群。对 7 个次级子群做密度计算,R-square 值达到 0.643,可知模型拟合效果较好。其中子群 1 密度值最高,为 0.972。说明子群 1 中的关键词成员之间被消费者同时提及的频率较高,已知子群 1 等于(肥大、颜色、面料、短小、尺码、色

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

差、小贵、透明），分析其网络结构可知消费者在消极评论中关注价格小贵的同时往往会同时关注面料、颜色色差和尺码方面。而子群 5、6、7 密度值较小被计算归零，说明其网络内部联系较为分散，被消费者同时关注的程度低。

3.6.3　分析结果

综合以上分析可以得知，真丝服装网购消费者按相关度排序容易依次就面料、颜色、尺码、质量、物流等方面给出负面评价，这些属性在消极评论中经常被消费者提及。

具体地，由于面料的核数和中心度最大，在整个网络中处于核心位置，并对其他结点影响力较大，所以首先来分析热点关键词面料。消费者在对面料进行负面评价时，“面料-透明”“面料-小贵”“面料-褶皱”“面料-缩水”和“面料-失望”相对共现频次较高，反映了消费者在购买真丝服装时，首先会考虑真丝面料上身效果是否薄透。而真丝面料因取材原因也不可避免地导致了价格贵、易出褶皱和缩水的结果，往往让真丝服装网购消费者感到失望。对于面料给出差评的消费者往往会对真丝服装的尺码、颜色、色差、价格、物流同样给出负面评价。所以，服装企业在致力于提高消费者对真丝面料的满意度时，应同时考虑对其尺码、颜色、色差、价格和物流等方面进行相关调整。

消费者在对真丝服装颜色给予差评时，同样也常常会对尺码、物流、价格以及由此产生的退换货不满意。除了对上述属性关键词相互关联外，在真丝服装网购消极评论中颜色也常常与色差、褪色、老气、与图片描述不一样等评论联系在一起。这也反映了服装网购的弊端，即由于光线、拍照设备、显示屏等原因往往会导致服装颜色照片与实物不一致的问题。色彩被列为服装构成的三要素之一，可见其对服装整体呈现的重要性，因此颜色偏差也是服装网购消费者给出差评的重要原因。尤其对于真丝质感的服装来说，颜色太深有时易显成熟老气，颜色太浅又会薄透显露内衣。

在消极评论中，尺码又常与上身效果、服务态度和款式相关的细节设计一同被提及，消费者由此给予肥大、短小、宽松、显胖、窄等负面评价。而在质量方面给与差评的消费者同时也会在价格、物流、面料、做工等方面给出负面评价。此外，物流除了与面料、颜色、质量等属性联系紧密外，同时也会与款式、发货速度均被消费者共同关注，被消费者给出速度慢、失望等负面评价。

4　结　语

本文以服装网购评论文本为研究对象，建立了基于服装网购评论的消费热点情报分析方法。在实现情感倾向性分类和关键词抽取后，将共词分析技术和社会网络分析技术相结合，揭示了关键词背后的复杂关系。本研究分析结果以可视化的方式呈现，以帮助服装企业了解服装网购评论的主要分布和结构，从而获取消费者的核心关注信息及消费热点。为了得到更好的实际效果，本研究中不但引入了 SnowNLP 情感分类技术，同时在 jieba 分词的基础上添加了自定义的服装领域网购评论分词词典和停用词表。此外为弥补 TF-IDF 算法在关键词抽取上的不足，本研究还综合了 GooSeeker 关键词抽取和人工调整方法，进一步提高了关键词抽取的准确度，获得了可靠度较高、覆盖度较广的关键词。

为了使企业能够及时了解消费动向需求，以便调整生产经营战略，提高消费者满意度和获取更多的市场份额，本研究以真丝服装网购评论为实例进行案例分析。本研究不足之处是数据预处理步骤中虚假评论识别部分仅仅采用了简单的人工识别，随着商家反防意识提高使得虚假评论内容越加真实，人工识别越加困难且耗时耗力、成本较高。因此，未来研究需要引入虚假评论识别技术，提高虚假评论识别准确率以进一步完善情报分析结果。

参考文献：

[1] 郭惠玲.快时尚品牌顾客满意度影响因素实证研究：以快时尚服装为例[J].中国流通经济，2015，29(2)：98-106.

[2] Park H，Kim K，Choi S，et al. A patent intelligence system for strategic technology planning[J]. Expert Systems With Applications，2013，40(7)：2373-2390.

[3] Yu H D，Liu F，Luo Y F. Screening model in enterprise competitive intelligence activity[J]. Advanced Materials Research，2010，121/122：360-363.

[4] 方友亮，孙斌，张晓阳，等.基于 SCP 范式的产业竞争情报分析框架构建[J].图书情报工作，2015，59(3)：95-102.

[5] Köseoglu M A，Chan E S W，Okumus F，et al. How do hotels operationalize their competitive intelligence efforts into their management processes？Proposing a holistic model[J]. International Journal of Hospitality Management，2019，83：283-292.

[6] 袁慧慧，田园，何昕蓉.基于服装产品研发的信息情报收集[J].艺术研究(哈尔滨师范大学艺术学院学报)，2013(1)：136-137.

[7] 王知津，葛琳琳.竞争情报 SWOT 模型与 BCG 矩阵比较

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

研究[J].图书与情报,2013(3):87-93.
[8] 辛洪芹.现代企业搜集竞争情报方法的思考[J].江西农业大学学报(社会科学版),2008,7(4):199-201.
[9] 唐炜,刘细文.专利分析法及其在企业竞争对手分析中的应用[J].现代情报,2005,25(9):179-183.
[10] 林卡,李钦海,邹明霞.企业竞争情报搜集与分析方法[J].商,2016(16):21.
[11] 李嘉文. 基于中小品牌企业需求的服装设计素材库构建[D].杭州:浙江理工大学,2016:24-42.
[12] Rocker J, Roncaglia G. Using the Web for competitive intelligence (CI) gathering[J]. Nature, 2002, 437(7058): 638-645.
[13] Chung W, Tseng T L B. Discovering business intelligence from online product reviews: A rule-induction framework[J]. Expert Systems With Applications, 2012, 39(15): 11870-11879.
[14] 靳健,张黎雪,刘馨儿,等.面向用户需求分析的产品评论用例提取研究[J].情报理论与实践,2020,43(1):104-111.
[15] 聂卉,李通,何欢,等.基于在线评论的商业竞争情报自动获取[J].情报杂志,2018,37(10):167-173.
[16] 韩立娜.正面在线评论对服装消费者购买意愿影响的实证研究[D].沈阳:东北大学,2013:14-16.
[17] Zhang D W, Xu H, Su Z C, et al. Chinese comments sentiment classification based on word2vec and SVM perf[J]. Expert Systems With Applications, 2015, 42(4): 1857-1863.
[18] 陈龙,管子玉,何金红,等.情感分类研究进展[J].计算机研究与发展,2017,54(6):1150-1170.
[19] Zhao Y Y, Qin B, Liu T. Sentiment analysis[J]. Journal of Software, 2010, 21(8): 1834-1848.
[20] 曾宇,刘培玉,刘文锋,等.特征加权融合的朴素贝叶斯情感分类算法[J].西北师范大学学报(自然科学版),2017,53(4):56-60.
[21] 李晓东,肖基毅,邹银凤.基于改进的TF-IDF与隐朴素贝叶斯的情感分类研究[J].南华大学学报(自然科学版),2019,33(2):79-84.
[22] 王菲,刘云飞.基于电商平台商品评价的情感分类研究[J].信息系统工程,2017(9):115-116.
[23] 张保富,施化吉,马素琴.基于TFIDF文本特征加权方法的改进研究[J].计算机应用与软件,2011,28(2):17-20.
[24] 朱庆华,李亮.社会网络分析法及其在情报学中的应用[J].情报理论与实践,2008(2):179-183.
[25] 徐萌.基于用户评价的企业竞争情报社会网络分析[D].青岛:山东科技大学,2017:31-38.
[26] 张振华,许柏鸣.基于在线评论文本挖掘的商业竞争情报分析模型构建及应用[J].情报科学,2019,37(2):149-153.
[27] 任佩萱.基于共词分析的我国服装学科研究热点可视化分析[J].金融经济,2018(22):144-145.
[28] 廉同辉,侯瑞瑞,孙雨欢,等.产业集群的社会网络特征研究:以安徽孙村服装产业集群为例[J].统计与信息论坛,2014,29(10):103-107.
[29] 贾旭楠.基于关键词共现和社会网络分析法的我国企业竞争情报热点主题研究[J].情报探索,2019(8):114-121.
[30] Callon M, Courtial J P, Turner W A, et al. From translations to problematic networks: An introduction to co-word analysis[J]. Social Science Information, 1983, 22(2): 191-235.
[31] 陈农.在线评论研究中的主题结构:社会网络分析的视角[J].现代情报,2015,35(1):61-67.
[32] 王倩倩.基于共词分析的国内在线商品评论研究热点探讨[J].现代情报,2017,37(10):158-164.

(责任编辑:陈丽琼)

(C)1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net