

电子商务用户反馈文本数据的可视化研究与实现

何媛, 陈亮, 李丰妤, 张志远
(华南理工大学, 广州 510006)

摘要: **目的** 针对数以万计的电子商务交易量背后产生的海量评价数据, 给用户带来的信息过载、检索低效等问题, 探索合理的可视化方案, 辅助用户进行购买决策, 帮助商家提取有效反馈以帮助产品优化。**方法** 以京东平台商品的评论数据为例, 提出一种电子商务用户反馈文本可视化方法, 以口红为目标商品, 采用网络爬虫技术从电商平台抓取用户评论文本数据并进行整理分析, 针对文本数据采取自然语言处理技术进行语义分析并归类, 得出可直接用于可视化的文本数据, 进一步提取了口红色号、价格趋势、关注词频率三个评论数据种类, 价格对比、口红色号可视化等用户指标, 借助 Tableau 可视化工具对数据进行映射, 建立了趋势图和分布图, 得到直观的可视化图表, 为用户提供了更美观且具有交互功能的信息浏览方式。**结论** 所提出的可视化方案采用视觉认知的方式提取关键信息, 能有效降低用户浏览文本信息的时间成本, 此方案的评估结果也获得了较高的评分, 对推进电商平台数据可视化具有一定指导意义。

关键词: 可视化设计; 用户反馈数据; 电商平台; 语义分析

中图分类号: J511 **文献标识码:** A **文章编号:** 1001-3563(2020)10-0228-07

DOI: 10.19554/j.cnki.1001-3563.2020.10.036

Research and Implementation of Visualization of Text Data Fed Back by E-commerce Users

HE Yuan, CHEN Liang, LI Feng-yu, ZHANG Zhi-yuan
(South China University of Technology, Guangzhou 510006, China)

ABSTRACT: The work aims to explore reasonable visualization schemes to assist users in making purchasing decisions and help merchants extract effective feedback for product optimization, with respect to such problems as overloaded information and inefficient retrieval brought to users by huge amounts of evaluation data generated by tens of millions of e-commerce transactions. With the comment data of the products of Jingdong platform as an example, a method for visualizing the feedback text of e-commerce users was proposed. Taking the lipstick products as the target goods, the web crawler technology was used to capture the user's comment text data from the e-commerce platform for their analysis. The text data were semantically analyzed and classified by natural language processing technology, and the text data directly used for visualization were obtained. Three kinds of comment data (lipstick color code, price trend and frequency of attention word), price comparison, lipstick color code visualization and other user indicator were extracted. Tableau visualization tools were used to map data, create trend and distribution maps, and get intuitive visual charts to provide users with more beautiful and interactive information browsing. The proposed visualization scheme can extract key information and effectively reduce the time cost for users to browse text information by means of visual cognition. The evaluation result of this scheme also obtains a high score. The proposed scheme has certain guiding significance for advancing the data visualization of e-commerce platforms.

KEY WORDS: visual design; user feedback data; e-commerce platform; semantic analysis

收稿日期: 2020-03-21

作者简介: 何媛(1994—), 女, 湖南人, 华南理工大学硕士生, 主攻设计学工学(信息与交互设计方向)。

通信作者: 陈亮(1975—), 男, 山东人, 博士, 华南理工大学教授, 主要研究方向为信息与交互设计。

信息技术的快速发展,使得互联网成为人们日常购物非常普遍的渠道之一,为人们日常采购带来了极大的便利。中国互联网络信息中心(CNNIC)已完成第 44 次《中国互联网络发展状况统计报告》,报告中详细分析了中国网民规模情况,截至 2019 年 6 月,我国网络购物用户规模达 6.39 亿,较 2018 年底增长 2 871 万,占网民整体的 74.8%^[1]。电子商务平台拥有过亿的用户背景,而庞大交易量数据背后的海量评论数据不容忽视。大量研究显示,在线商品评论对用户购买决策具有重要的影响^[2-3],但与此同时,体量巨大、类型繁多、时效性高的评论数据给用户带来了信息重复与过载问题,人类的记忆和信息处理能力都是有限的,大量冗余的评论信息严重干扰了用户对目标信息的准确定位和快速选择。对于想要准确购买目标产品的用户或者一些想要优化产品和服务的商家,面对现有的评论数据多文本展示、大量无效评论、过度的主观描述等一系列问题,想获取准确信息较为困难。

目前国内外对商品评论领域已有广泛的研究, Ghose A, Ipeiritis P G 基于文本挖掘和信息质量评价理论的视角,认为在线商品评论内容中的观点主观性有助于预测评论的效用^[4],孙妮使用数据抓取处理工具和数据分析工具研究 B2C 购物网站消费者的在线评价信息,分析了网络购物群体的行为规律^[5]。复旦大学林钦和提出了基于情感计算的商品评论分析系统,该系统能自动采集商品评论数据进行情感分析,并将研究分析得出的对应结果展示给用户,从而帮助用户做出更合理的购物决策^[6]。以上研究仍然大都以文本或简单图表的形式将这些评论信息展示出来,目前可视化在商品评价和用户反馈领域的研究中并不常见,需进一步拓展。

1 大数据可视分析

当大数据以直观的可视化的图表形式展示在用户面前时,用户往往能快速获取数据背后隐藏的信息^[7]。可视分析(Visual Analytics)是通过交互式可视化界面

来辅助用户对大规模复杂数据集进行分析推理的科学与技术^[8]。数据可视化即将数据信息转化成视觉信息,利用相应计算机技术或工具提供可交互的方式^[9]。图形展示信息的效率比文本高得多,人们从外部获取信息的方式,视觉渠道占比高达 80%^[10],人眼对图像感知的敏感度要高于文本,图像认知的速度和效率优于文本阅读,数据可视化能让人们在攫取数据价值、特征和关系的过程中,提升准确性、提高效率,获得更好的用户体验^[11]。

目前国内外现有的数据可视化技术经过二十余年的迅猛发展,已经成为人们分析复杂数据、解决棘手问题的强有力工具^[12]。现有的可视化工具如 Tableau、Power BI 和 Echarts 等已经得到广泛使用,借助这些可视化工具可制作出多样化的数据可视化图表,各类可视化工具的优缺点分析见表 1。

口红商品的可视化,在界面美观度以及可交互性方面会有较高的要求,Tableau 可完成自定义的网页交互界面,同时其免费共享能力便于后期的可用性评估测试,故本研究选择此可视化工具进行数据的可视化映射。借助 Tableau 的设计流程如下:通过文本数据量化、指标化的方法,将电子商务用户反馈文本数据通过可视化图表的方式向用户展示,旨在帮助用户更快速地获取商品评论中的关键信息,缓解信息过载的问题。前期通过用户研究工作挖掘用户需求,采用自然语言处理技术对获取的研究数据进行处理并分析,为可视化设计提供数据,后期研究结果为前期的用户研究工作提供验证,最终进行评估。主要设计流程见图 1。

2 数据分析

通过对评论数据的分析,可分别从局部的角度和整体的角度挖掘用户掌握商品真实情况的潜在信息,给用户或者商家带来有效的综合信息,从而掌握商品整体状况。

表 1 部分可视化工具优缺点比较

Tab.1 Comparison of advantages and disadvantages of some visualization tools

名称	优点	缺点
Tableau	1.免费共享能力 2.混合数据源 3.支持多维数据集	1.财务报告应用程序不及其他工具 2.无版本控制的概念 3.免费版本的数据是公开的
QlikView	1.灵活的用户界面 2.快速渲染图型和表格 3.支持各种来源的数据导入	1.容易无意间组合一些数据 2.编程复杂 3.价格昂贵
Klipfolio	1.灵活的外观,实时数据分析简单 2.持续开发新功能 3.与第三方软件和工具包轻松集成	1.自定义 CSS 样式价格昂贵 2.无临时分析或临时报告功能 3.计算复杂
Power BI	1.支持多方式整合或导入数据 2.多平台支持(Web, 移动)	1.基本功能操作复杂 2.缺乏数据准备和清理工具

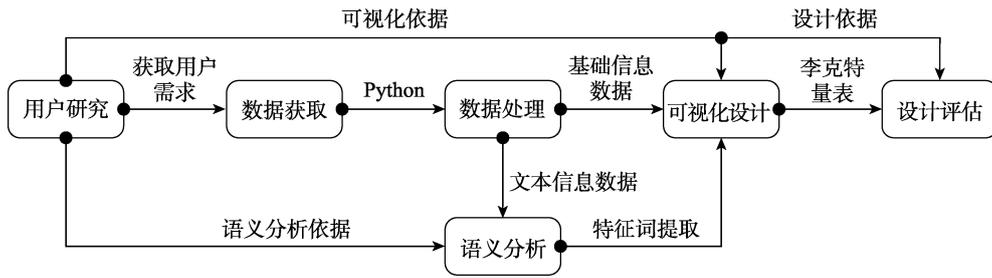


图1 设计流程
Fig.1 Design flow chart

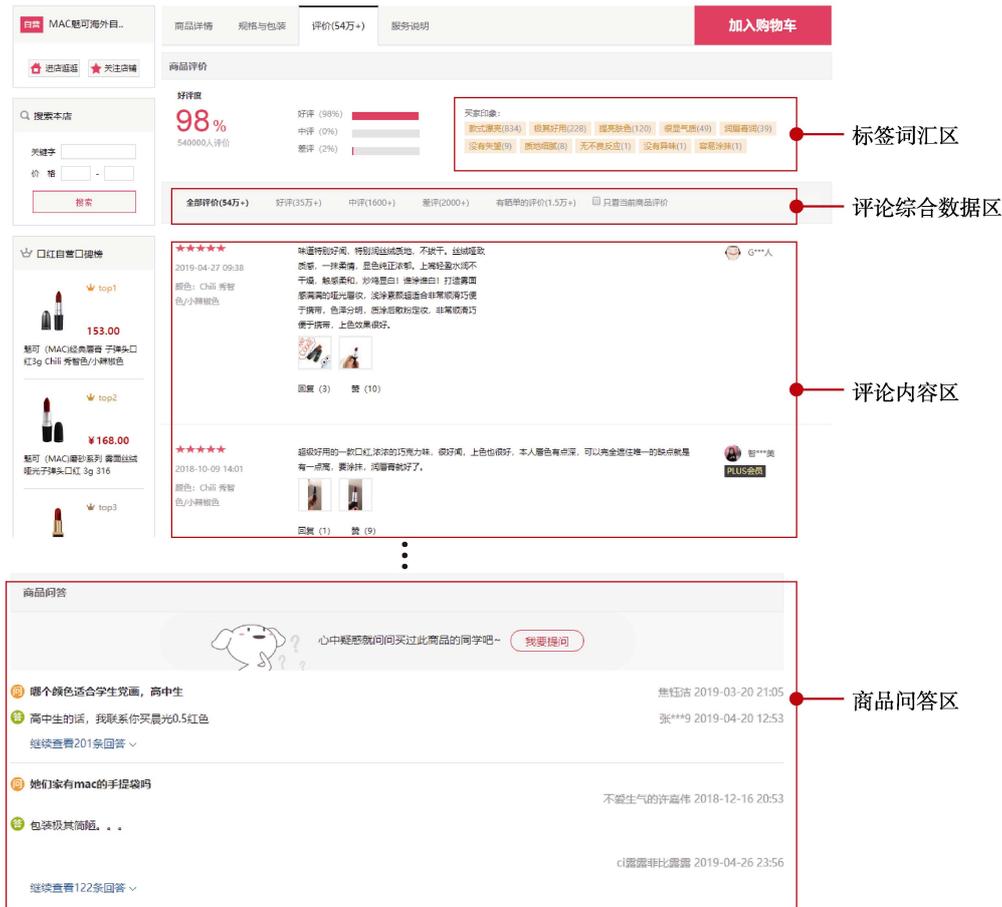


图2 MAC子弹头口红商品评论区界面分析
Fig.2 Commodity comment area interface analysis of MAC Bullet Lipstick

2.1 商品评论区分析

此研究的对象是商品评论，具体以魅可（MAC）经典唇膏子弹头口红为研究案例。由于口红商品被视为较为廉价的奢侈品，即使处于经济不景气的场景下，人们对口红商品依然有强烈的消费欲望，且此类产品的分类较多，属于高决策型商品，所以选择该类商品作为研究案例对其他商品具有更为广泛的参考意义。本次分析是以分析商品评价向用户展示信息的方式，在京东商城上浏览MAC子弹头口红的商品评论区，对其进行界面分析，见图2，该界面大致可分为四个区域，即标签词汇区、评论综合数据区、评论内容区与商品问答区，通过分区便于后期合理进行数

据提取。

标签词汇区域是根据频数统计得出的多个标签词汇组合区域，目的是向用户展示买家对商品的印象。目前这些标签的作用一方面向用户展示买家对商品的印象，如“很显气质”相关的评论有四十九条，用户可以根据个人兴趣进行浏览，但是如“质地细腻”、“质地柔软舒适”等相似的词，会因为不同用户对同一词汇的不同认知描述，而造成“买家印象”栏的信息重复等问题，在类别及准确性上与消费者所关心的指标不符。这种展示信息的方式虽然在一定程度上可以帮助用户掌握一些信息，但是由于数量局限，重要信息展示不全面，而且存在部分属性为无效信息的标签，例如“极其好用”等词汇，并非对商品的具体描

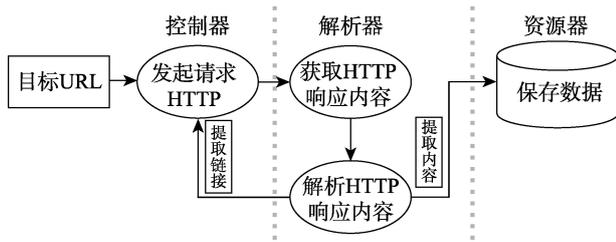


图 3 评论数据获取的流程
Fig.3 Crawler process of comment data

述，所以实际参考意义不大。评论综合数据区是将用户评价的各类数据信息整合展示给用户，主要分为好评、中评、差评和有晒单的评价。通过统计评论内容区域的页面数，发现每页最多向用户展示十条评论数据，共计可以显示一百页的评论数据，即一千条评论数据。面对如此庞大的文本数量，用户很少愿意消费大量时间成本去查看这些评论来获取自己想要的信息。最能展示用户需求的是商品问答区，用户可以在看了评价之后主动对已买过该商品的用户进行提问，但并不会得到即时反馈，因此导致用户在商品选购的过程中花费较大的时间成本。只有将全部评论中有用的信息进行统计，合理归类并分析，才能给用户提供更好的购买决策辅助功能。

2.2 评论数据获取

以京东 MAC 子弹口红网页的用户反馈信息为案例，使用 Python 进行网络爬虫，获取了六百九十九条口红评论数据。网络爬虫是一个自动提取网页的程序，它为搜索引擎从 Web 上下载网页 [13]。评论数据获取的流程见图 3。

由于通过网络爬虫所获取的原始数据都是以键/值对(也可理解为字典)的形式存储在 JSON 文件中，所以需要通 Python 编码语言把评论数据转变为 Excel 的 CSV 文件，成为后期 Tableau 可视化工具的数据连接格式。处理所得的预处理数据包含以下内容：用户名称、会员等级、评论内容、评论创立时间、好差评分、支持数量、反对数量、追加评论、商品色

号、商品型号以及回复数量等，这里给出部分“评论内容”和“商品颜色”的数据内容，见图 4。

2.3 文本语义特征词提取

语义是信息对事物的描述和逻辑表示 [14]。特征词提取 [15] 也称为关键词提取，是指从文章或者段落中提取相关具有代表性的词语，它是语义分析中的重要环节，在文献搜索、自动文摘和文本聚类或分类中应用较为广泛。本研究通过词频-逆文档频率 (TF-IDF) 算法来提取商品评论中的关注词。TF-IDF 是一种数值的统计方式，它通过单词的出现频率来反映该词在文档中的重要性，如果该词在语料库中的其他文档中出现的频率低，则文档频率 (DF) 低，逆文档频率 (IDF) 高，可以理解为该词具有较高的识别度，即可认为该词是对应文档的关键词 [16]。本设计研究对象是用户的反馈文本数据，反馈文本数据是对商品的一个或多个特征描述，将全部的反馈数据汇总作为语料库，则可以得出商品特征在语料库中的不同频数，具体操作步骤如下：首先导入 extract_tags() 函数；然后将初始评论语料置入列表中；最后按照 n、ng、nr、vn、nt、nz、x 词性来提取。

通过对预处理的评论数据中名词与形容词的提取，按提及频率从高到低筛选出包括“显白、实惠、包装、质感、持久、气色、湿润、黑皮”等词汇，反映出用户购买口红这个商品所关注的都集中在口红的这些指标上，为下一步的可视化功能定位提供数据基础。

3 可视化设计

3.1 可视化流程

以前期的用户研究为基础，获得直观层面的用户需求，通过对商品评论数据的分析，剖析现有数据带来的问题，挖掘用户更多层面的需求，包括探索用户的潜在需求，继而进行精准的设计定位，然后对归类好的数据进行可视化图表的映射，通过 Tableau 工具

评论内容	商品颜色
做工还行，不知道真假。没有礼盒有点low,反正媳妇挺喜欢的	秀智色/小辣椒色
昨天刚到的，关注了很长一段时间，还是买了这个小辣椒太喜欢了，物流很快没一个星期。	秀智色/小辣椒色
最下面那个是番茄红，涂了几遍的效果，上层是红润，很好看的颜色。	番茄红
专业的就是不一样 有质量保证 值得信赖	秀智色/小辣椒色
种草许久啦啦啦，去店里试了颜色，没有活动，然后转身来这下单，实惠好用，颜色大爱，不挑皮肤	摩卡奶茶色
种草许久啦啦啦，去店里试了颜色，没有活动，然后转身来这下单，实惠好用，颜色大爱，不挑皮肤	豆沙红
种草很久了一支口红，终于在京东买到，味道很好闻，薄涂厚涂，颜色都很漂亮，是我想要的那种。	豆沙红
种草很久，终于买了，色号真心好看。显皮肤白	秀智色/小辣椒色
质量非常好，与描述的完全一致，非常满意，真的很喜欢，完全超出期望值，发货速度非常快，包装非常仔细、严实，物流公司服务态度很好，运送速度很快，很满意的一次购质量很好，希望更多的朋友信赖。店主态度特好，我会再次光顾的。	秀智色/小辣椒色
知道选色的烦恼，拍几张高清的让要买的美女参考，忽略皮肤！没图唇膏，稍微有点干，也不行，超显色，只涂了一点用手指抹开的，没有我想象的橘，还可以，薄涂很日常	亮珊瑚橘
只能说终于买到了，找了很久，正品吧	复古哑光红
只能说慎拍吧！因为真的是太次了，别和我说什么不退！你说为什么！	秀智色/小辣椒色

图 4 部分评论数据
Fig.4 Partial comment data

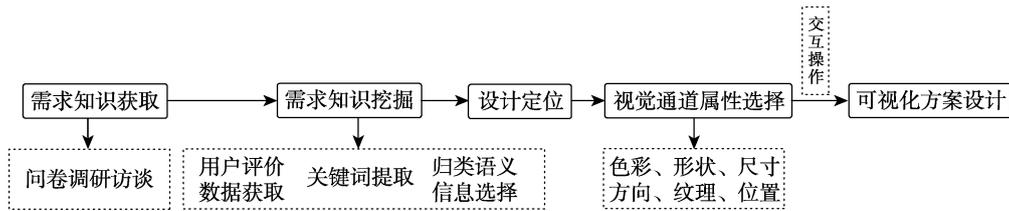


图5 可视化设计流程

Fig.5 Visual design process

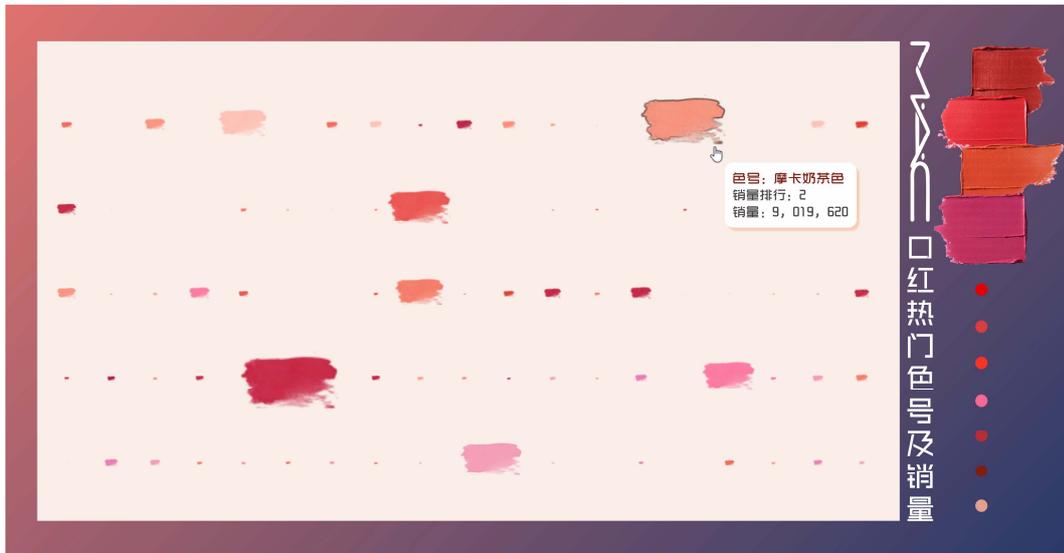


图6 口红色号及销量可视化设计

Fig.6 Visual design of lipstick color code and sales

实现新的可视化图表方案，可视化设计流程见图5。

3.2 可视化设计方案

现有商品评论区数据量太大，获取真实的评论信息难且干扰评论过多，许多用户会在“商品问答区”耗时来提问，用户更希望直接获取在商品评论区所没有得到的信息，现有评论区缺少具体综合指标分析。

通过问卷调查与访谈得出的需求以及 TF-IDF 算法所提取的关键词，挖掘出用户希望获得大众用户客观的口红色号推荐，根据多层次可视化设计、概括化简和评价数据信息表达设计原则，进行设计定位：

(1) 口红色号及销量的可视化；(2) 价格趋势的可视化展示；(3) 用户关注词的频率可视化展示。

关于第一点的设计定位，具体设计方案见图6，遵循多层次可视化方法，通过笔刷式色条的不同大小展示销量及热门程度，鼠标悬停在色条上即可显示口红色号名称、销量、以及销量排名。

基于第三点的设计定位，具体设计方案见图7。选择与口红呼应的红色渐变色系，价格从高到低遵循灰度等级进行可视化展示，同时可进行品牌筛选，对用户而言可以直观地看到适合自己价位的口红价格趋势，单击某品牌可以高亮查看单一的价格趋势，直观呈现品牌的价格浮动区间。

对应第四点的设计定位，具体设计方案见图8。遵循概括化简原则，设计高频词可视化交互展示，当

鼠标悬停在代表某一项关键词的图标上时，界面就会显示点赞最多的用户评论，这是评价数据信息的表达；界面内部较小的图标，是包含同一项关键词的其他点赞数没有那么多的评论，但每一个形状都会在鼠标悬停时出现“被提及次数”及评论，与用户进行交互。商家也可根据用户的关注重心设计出相对应的活动策划并获取更明确的产品迭代方向。

4 可用性测试

为了进一步验证本研究可视分析方法的可用性和实用性，通过线上将 Tableau 中的图表保存到 Tableau Public 然后以链接形式分享给用户进行测试，以及线下用户测试两种方式，调查了三十四位有网购经历的用户（五位线下测试，二十九位线上测试），分别给出案例口红的色号及销量图、价格趋势图与评论数据可视化图表，在用户自由探索三种视图之后，用户需要完成调查问卷上的问题。

本文使用五等级李克特量表进行评估。李克特量表由美国社会心理学家李克特 (R. A. Likert) 于 1932 年提出，量表由一组与主题相关的问题或陈述组成，用来表明受访者对某一事物的态度、评价或意向。量表通常采用五级测量层次，即对量表中每一题目/陈述均给出表示态度积极程度等级的五种备选选项（如“非常不同意”、“不同意”、“不确定”、“同意”、“非常

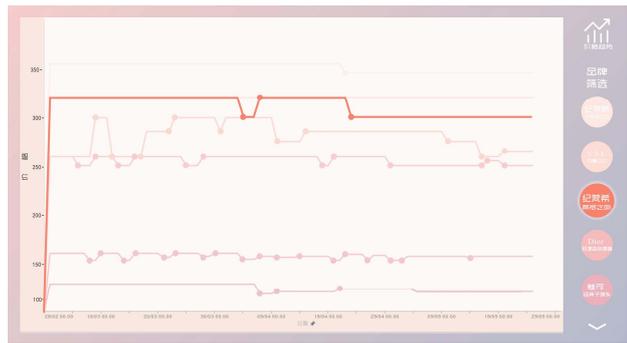
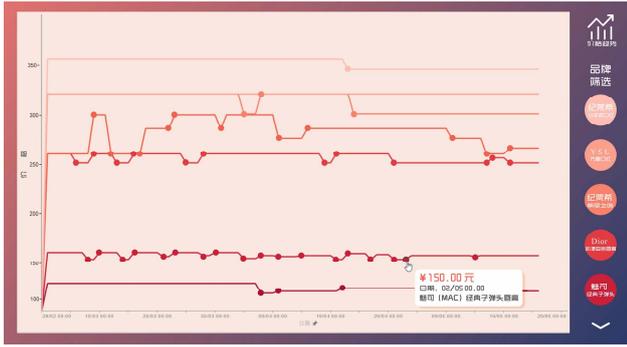


图 7 价格趋势可视化设计
Fig.7 Visual design of price trend



图 8 评论数据可视化设计
Fig.8 Visual design of comment data

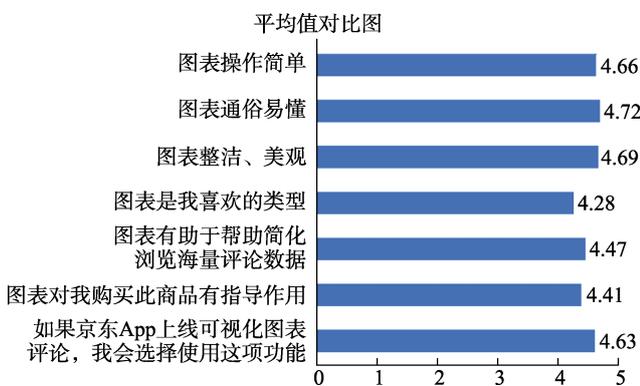


图 9 李克特量表得分
Fig.9 Likert scale score

同意”等)，并用 1~5 分别为五种答案量化计分，量表总分越高则表明受访者态度越积极^[17]。本评估问卷将从操作的简易性、构成的简易性、视觉效果、反应速

表 2 李克特量表问题设计
Tab.2 Likert scale problem design

问题 1: 图表操作简单
问题 2: 图表通俗易懂
问题 3: 图表整洁、美观
问题 4: 图表是我喜欢的类型
问题 5: 图表有助于帮助简化浏览海量评论数据
问题 6: 图表对我购买此商品有指导作用
问题 7: 如果京东 APP 上线可视化图表评论，我会选择使用这项功能

度、好感度和是否有用这六个要素展开问题陈述，量表问题设计见表 2。

三十四位受访者填写问卷后的各项平均得分见图 9。结果显示，大部分参与者对本文可视化视图结果的简易性、实用性和有用性表示满意，对视觉效果非常满意。

5 结语

本文面向电子商务平台的用户反馈数据提出了一种基于语义分析的用户评论数据可视化方法，通过界面优化，形式创新，改变了以往评论数据的呈现方式，实现了数据可视化，方便浏览与查阅数据。研究提出了电子商务领域信息可视化设计的完整流程框架，即通过用户调研、数据获取、语义分析再到可视化展示四个流程模块，不同产品可延伸出针对性的可视化方案。研究以口红商品为例，展开了对电商平台用户反馈信息可视化设计探究，做出了具体的具有合理交互操作且视觉认知更为美观的可视化图例，可视化分析结果及用户评价进一步验证了本文可视化设计的有效性和实用性，但该研究仍需要更多维度的探索，关注商品问答区的数据，思考差评数据如何避开主观臆断得到合理筛选，从而让用户获得更好的购物体验。在这个大数据的时代，从语义分析的角度提出对大数据信息的可视化设计建议，可以运用到更多领域。

参考文献:

[1] 中商产业研究院. 2019 上半年我国网络购物用户规模达 6.39 亿使用率达 74.8%[EB/OL]. (2019-08-31) [2019-11-23]. <https://www.askci.com/news/chanye/20190831/1112151152112.shtml>.
China Business Industry Research Institute. In the First Half of 2019, the Number of Online Shopping Users in China Reached 639 Million and the Utilization Rate Reached 74.8%[EB/OL]. (2019-08-31) [2019-11-23]. <https://www.askci.com/news/chanye/20190831/1112151152112.shtml>.

[2] GILBERT E, KARAHALIOS K. Understanding Deja Reviewers[C]. Savannah: Proceedings of the 2010 ACM Conference, 2010.

- [3] HSU L H. The Effects of User Reviews on Online Purchasing Behavior across Multiple Product Categories[C]. Hawaii State: IEEE Photovoltaic Specialist Conference, 2006.
- [4] GHOSE A, IPEIROTIS P G. Designing Novel Review Ranking Systems: Predicting Usefulness and Impact of Reviews[C]// Proceedings of the Ninth International Conference on Electronic Commerce. ACM, 2007.
- [5] 孙妮. B2C 购物网站商品评价的效应研究[D]. 北京: 对外经济贸易大学, 2014.
SUN Ni. Effect of Product Evaluation on B2C Shopping Website[D]. University of International Business and Economics, 2014.
- [6] 林钦和, 刘钢, 陈荣华. 基于情感计算的商品评论分析系统[J]. 计算机应用与软件, 2014, 31(12): 39-44.
LIN Qin-he, LIU Gang, CHEN Rong-hua. Analysis System of Product Reviews Based on Emotional Computing[J]. Computer Applications and Software, 2014, 31 (12): 39-44.
- [7] 任磊, 杜一, 马帅, 等. 大数据可视分析综述[J]. 软件学报, 2014, 25(9): 1909-1936.
REN Lei, DU Yi, MA Shuai, et al. A Review of Big Data Visual Analysis[J]. Journal of Software, 2014, 25(9): 1909-1936.
- [8] THOMAS J J. Illuminating the Path: the Research and Development Agenda for Visual Analytics[M]. United States: IEEE Computer Society, 2005.
- [9] 刘勘, 周晓峥, 周洞汝. 数据可视化的研究与发展[J]. 计算机工程, 2002(8): 1-2.
LIU Kan, ZHOU Xiao-zheng, ZHOU Dong-ru. Research and Development of Data Visualization[J]. Computer Engineering, 2002 (8): 1-2.
- [10] ARABNIA H. Reading in Information Visualization: Using Vision to Think (Media Review)[J]. IEEE Multimedia, 1999, 6(4): 93-93.
- [11] 田凯燕. 基于语义映射的大数据视觉信息可视化设计研究[D]. 南京: 东南大学, 2018.
TIAN Kai-yan. Visual Information Visualization Design of Big Data Based on Semantic Mapping[D]. Nanjing: Southeast University, 2018.
- [12] 黄朝晖, 李晓梅. 数据可视化技术及现状[J]. 计算机工程与科学, 1993(2): 39-43.
HUANG Zhao-hui, LI Xiao-mei. Data Visualization Technology and Current Status[J]. Computer Engineering and Science, 1993(2): 39-43
- [13] 刘金红, 陆余良. 主题网络爬虫研究综述[J]. 计算机应用研究, 2007(10): 26-29.
LIU Jin-hong, LU Yu-liang. Summarization of Research on Theme Web Crawlers[J]. Application Research of Computers, 2007(10): 26-29.
- [14] 赵哲峰. 基于语义分析方法的视频流媒体大数据技术研究[D]. 太原: 太原理工大学, 2013.
ZHAO Zhe-feng. Video Streaming Media Big Data Technology Based on Semantic Analysis Method[D]. Taiyuan: Taiyuan University of Technology, 2013.
- [15] 李丕绩, 马军, 张冬梅, 等. 用户评论中的标签抽取以及排序[J]. 中文信息学报, 2012, 26(5): 14-19.
LI Pi-ji, MA Jun, ZHANG Dong-mei, et al. Label Extraction and Sorting in User Reviews[J]. Chinese Journal of Information Science, 2012, 26(5): 14-19.
- [16] 严立军. 电子商务用户反馈信息的可视化研究[D]. 广州: 华南理工大学, 2019.
YAN Li-jun. Visualization of E-commerce User Feedback Information[D]. Guangzhou: South China University of Technology, 2019.
- [17] 亓莱滨. 李克特量表的统计学分析与模糊综合评判[J]. 山东科学, 2006(2): 18-23.
QI Lai-bin . Statistical Analysis and Fuzzy Comprehensive Evaluation of Likert Scale[J]. Shandong Science, 2006(2): 18-23.