

doi: 10.3969/j.issn.1007-7375.230227

情感分析与数据驱动下面向产品迭代设计的 用户画像及建模研究

周艳杰¹, 李耀辉¹, 王宇¹, 王永胜²

(1. 郑州大学 管理学院, 河南 郑州 450001; 2. 中国烟草总公司郑州烟草研究院, 河南 郑州 450001)

摘要: 针对传统产品设计耗时耗力、效率较低等挑战, 提出了面向产品迭代设计的用户画像及建模研究, 该方法能够有效地刻画用户对产品特征的态度, 为企业进行产品迭代提供新的参考模式。首先, 通过爬虫技术获取用户信息属性、在线用户评论信息属性, 结合产品信息属性构建三维用户画像概念模型。基于产品特征本体, 采用 Word2vec 技术和情感分析方法从用户关注度和质量满意度两个维度构建画像模型。然后, 通过整体用户画像与产品细粒度特征画像, 聚焦产品核心优势特征与待优化特征, 采用 K 值法制定产品优化策略。最后, 以某相机为例进行数据驱动下的产品迭代设计案例分析。结果表明, 该方法能够有效挖掘用户核心需求, 为商家快速迭代产品提供较优产品设计方案。

关键词: 在线评论; 用户画像; 产品迭代; 数据驱动; 情感分析

中图分类号: F406.13; C93

文献标志码: A

文章编号: 1007-7375(2024)02-0074-13

User Profiles and Modeling for Iterative Product Designing with Sentiment Analysis and Data-driven Approaches

ZHOU Yanjie¹, LI Yaohui¹, WANG Yu¹, WANG Yongsheng²

(1. School of Management, Zhengzhou University, Zhengzhou 450001, China;

2. Zhengzhou Tobacco Research Institute of China National Tobacco Corporation, Zhengzhou 450001, China)

Abstract: To address the issues of time-consuming and laborious, as well as low efficiency of traditional product designing, user profiles and a modeling method for iterative product designing is proposed. The proposed method can effectively capture use attitudes towards product features, providing a new baseline for enterprises to conduct product iteration. First, the user information attributes and online user comments are obtained by web crawler technology. A conceptual model of three-dimensional user profiles is established by integrating product information attributes. Based on product features, a profile model is developed using Word2vec technology and sentiment analysis from two dimensions: user attention and quality satisfaction. Then, by analyzing the overall user profile with the fine-grained feature profile of products, a product optimization strategy is determined using the K-value method, emphasizing both core advantageous features and features for optimization. Finally, a case study is conducted using a certain camera as an example to analyze data-driven iterative product designing. Results show that the proposed method can effectively explore the core demand of users and provide enhanced product design solutions for enterprises to expedite product iterations.

Key words: online comment; user profiles; product iteration; data-driven; sentiment analysis

近年来, 我国互联网用户数量增长快速, 网络购物发展迅猛。据 CNNIC 第 52 次《中国互联网络

发展状况统计报告》显示, 截至 2023 年 6 月, 我国网民规模达 10.79 亿, 互联网普及率达 76.4%,

收稿日期: 2023-12-05

基金项目: 国家自然科学基金资助项目 (72201252, 72101240); 郑州烟草研究院青年人才托举工程项目 (602020CR0360)

作者简介: 周艳杰 (1988—), 男, 河南省人, 副研究员, 博士, 主要研究方向为启发式算法、组合优化、智能制造、智慧物流。

通讯作者: 王宇 (1990—), 男, 河南省人, 讲师, 博士, 主要研究方向为信息产品与服务管理、数字化平台运营、大数据治理。Email: ywang@zzu.edu.cn

网络购物用户规模达 8.84 亿, 占网民整体的 82.0%^[1]。可见, 网络购物受到了众多消费者的青睐, 在线商业蓬勃发展。与此同时, 用户在网购平台反馈自己对产品的看法, 产生了大量的在线用户评论。

在线用户评论是消费者基于自身使用或体验产品而形成的评价, 蕴含消费者对产品的态度、偏好与需求信息。随着企业数据监管机制的不断完善, 在线评论的真实性和可靠性也不断提升, 已然成为商家获取用户需求信息的重要渠道, 并以此进行产品改进。传统的获取用户需求信息的方法主要包括实地调研、问卷调查等, 但这些方法受到人力、物力、财力等限制, 不能快速、准确、全面地获取用户需求, 而通过在线评论获取用户需求则弥补了该方面的不足。对在线评论进行挖掘时包括粗粒度挖掘和细粒度挖掘。粗粒度挖掘基于在线评论分析用户关注的产品特征或用户对产品的整体态度, 但无法判断用户对各产品特征的态度。细粒度挖掘深入分析用户对各产品特征的情感倾向, 有利于企业更精确地挖掘用户需求偏好。经调研发现(见表 1), 目前已有较多学者通过在线评论获取用户需求并应用于产品设计, 但是现有研究对产品特征的分析仍不够细致, 且对产品特征信息进行粗粒度或细粒度挖掘后, 仅对产品改进进行了简单讨论。一个清晰可行的产品迭代方案对于企业快速迭代产品来说至关重要, 而现有研究并未深入分析并给出具体可行的产品迭代方案。

表 1 产品设计相关研究

Table 1 Previous studies on product designing

文献	用户需求获取方式		粗粒度挖掘	细粒度挖掘	产品具体迭代方案
	传统方式	在线评论			
王军等 ^[2]	√		√		
贾丹萍等 ^[3]		√	√		
Du 等 ^[4]		√	√		
黄晟青等 ^[5]		√	√		
Yang 等 ^[6]		√		√	
沈超等 ^[7]		√		√	
Joung 等 ^[8]		√	√	√	
王克勤等 ^[9]		√	√	√	
本文		√	√	√	√

用户画像可以通过大数据挖掘技术捕获用户的行为数据, 利用分类等技术建立画像模型, 概括出

用户的真实特征和偏好。在画像建模的过程中, Word2vec 是一种关注上下文逻辑的词语语义计算工具, 利用 Word2vec 技术对产品在线评论进行词向量训练可以得到产品特征词之间的语义关系。情感分析技术能够挖掘用户对不同产品特征属性的情感倾向。鉴于此, 本文使用用户画像方法, 在数据驱动下通过大数据的采集、管理和计算, 提取用户高关注度产品特征并构建产品特征本体。基于产品特征本体, 本文采用 Word2vec 技术和情感分析方法从用户关注度和质量满意度两个维度构建画像模型, 最终呈现画像结果并进行粗、细粒度相关分析, 采用 K 值法为企业快速迭代产品制定较优策略, 弥补了现有基于在线评论进行产品设计研究的不足。

本研究具有一定的理论意义和应用价值: 1) 创新性地用户画像方法应用于产品设计, 丰富了用户画像的应用领域; 2) 解决了传统方式不能快速、准确、全面地获取用户需求的问题; 3) 基于在线评论粗、细粒度挖掘产品特征信息, 剖析用户核心诉求, 为企业快速迭代产品提供了具体可行的产品迭代设计方案, 有利于提高产品的竞争力。

1 相关研究

1.1 用户需求获取研究现状

用户需求是指用户为了满足自我需要, 实现个人目的, 以自身的理解和习惯表达出来对外部环境事物的一种心理活动。充分认识和把握用户对产品的需求是企业的立足之本, 针对用户需求对产品进行优化设计, 可以帮助企业在行业竞争中获得更大的优势。早期的用户需求获取主要是通过实地调研、问卷调查、网络日志和访谈等方式。然而, 这些传统的方法受到人力、物力、财力等限制, 不能快速、准确、全面地获取用户需求。

近年来, 越来越多的学者对在线评论进行研究, 由此挖掘用户需求。吴运明等^[10]基于在线评论提出移动商务用户需求聚合与挖掘方法, 获得了比基于语义的传统聚类方法更为精准和细化的用户需求。Wang 等^[11]基于灰色粗糙模型对用户需求进行获取, 并将其有效地转化为相关特征。Son 等^[12]将用户评论进行结构化和分类, 利用基于神经网络的 SOM 分析在线评论中的用户体验, 该研究有助于将用户的潜在需求整合到产品设计中。Goldberg 等^[13]

提出了一种确定在线评论优先等级的方法,快速找到有利于公司创新的相关评论,为企业更好地利用在线评论信息提供指导。可见,在线评论已经成为很多学者获取用户需求的重要有效途径。

1.2 用户画像研究现状

由交互设计之父 Alan Cooper 提出的用户画像概念是指真实用户数据的虚拟代表。在大数据时代,用户画像以用户数据为基础,采用数据挖掘技术刻画用户特征,形成对用户等的多维度描述。

目前,国内外关于用户画像的研究主要涉及电子商务、图书馆、医疗健康、旅游业等领域。Xu 等^[4]构建了多维度的用户画像指标体系,建立了考虑情感因素的网购用户画像模型。陈添源等^[5]以图书馆具体的业务流程和用户行为路径为主线,探究了用户画像的新特征,提出实时更新且适应更广的用户画像系统框架。Jiang 等^[6]将社区心理标签和用户画像模型关联起来,建立了基于卷积神经网络的优化多模态检索模型。滕春娥等^[7]提出了一种在线医疗社区用户画像构建框架,该方法有助于信息精准推荐、改善患者体验、提升决策质量。Hu 等^[8]通过游客在线评论数据进行了新的旅游需求建模和预测,实证结果表明基于游客在线评论数据的旅游需求预测可以显著提高旅游需求模型的预测性能。

此外,用户画像在教育、科研、金融等领域也有所应用。

综上,在用户需求获取方面,在线评论和情感分析已成为获取用户需求的重要有效途径^[9],但是现有研究对产品各特征信息的挖掘不够细致,针对产品特征级进行细粒度情感分析,更有助于挖掘在线评论的价值,深入剖析用户需求;在用户画像方面,国外研究早于国内,但用户画像在应用领域、构建维度等方面还有较大发展空间。除此之外,一个具体可行的产品迭代方案对于企业快速迭代产品来说至关重要,但是由表 1 可知,现有基于在线评论进行产品设计的相关研究并未给出具体可行的产品迭代方案。针对现有研究的不足,本文利用产品在线评论构建产品特征本体,基于产品特征本体对评论语料实施情感分析;进而实现用户画像,根据画像结果粗、细粒度挖掘产品特征级信息,有效地刻画用户对产品特征的态度,为企业进行产品迭代提供一种新的参考模式。

2 用户画像模型构建

本文的用户画像模型构建框架如图 1 所示,主要包括如下 3 个步骤。

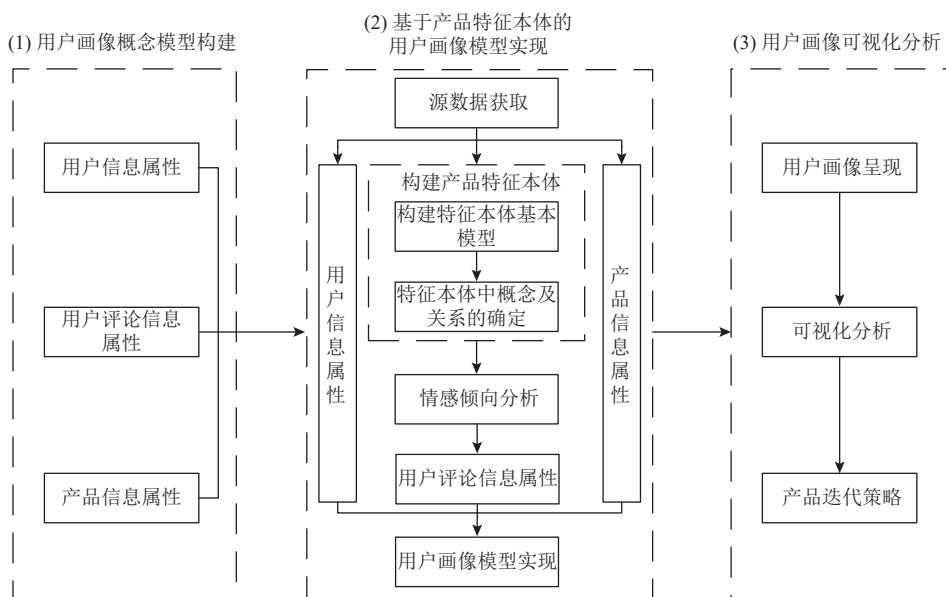


图 1 用户画像模型构建框架

Figure 1 Framework for constructing user profile models

1) 用户画像概念模型构建。从用户信息属性(包含用户名、用户等级、性别、年龄、教育程度、

地域等信息)、用户评论信息属性(包含用户评分、评论内容、评论时间、点赞数、追评内容、追评时

间等信息)、产品信息属性(包含产品名称、价格、适用场景、零部件组成、技术参数等信息)3个维度构建用户画像概念模型。

2) 基于产品特征本体的用户画像模型实现。主要包括源数据获取、提炼用户信息属性和产品信息属性以及基于产品特征本体的用户评论信息属性、用户画像模型实现。

3) 用户画像可视化分析。借助文字描述、统计图表、词云图等方式呈现画像内容, 并对画像结果进行分析, 最终得到产品迭代策略。

3 基于产品特征本体的用户画像模型实现

3.1 数据获取与处理

本文通过爬虫技术爬取网购平台的用户数据、在线用户评论数据, 并基于网购平台获取产品信息所属字段, 分别用来实现用户画像模型的3个属性。

对得到的用户评论语料剔除无用评论, 进而得到有效评论文本。为加强所提取产品特征的规范性, 以产品规格说明书为基础将相关产品特征加入自定义词典, 然后利用 Python 编程对得到的有效评论文本进行 jieba 分词、去停用词及词性标注等预处理工作。

3.2 产品特征本体构建

本体是共享概念模型的形式化规范说明, 具有从抽象到具体的结构优势, 能从不同粒度去刻画用户画像。但是用本体进行画像建模时无需使用本体中蕴含的所有信息, 一般只需领域内概念及概念之间的关系即可。

本文利用产品在线用户评论构建产品特征本体, 基于产品特征本体对评论文本实施情感极性分析, 进而实现用户画像建模。这种方法既能挖掘用户高关注度的产品特征信息, 又能精简特征本体构建过程, 加快用户画像建模, 聚焦产品特征优化方向。产品特征本体构建流程如图2所示。

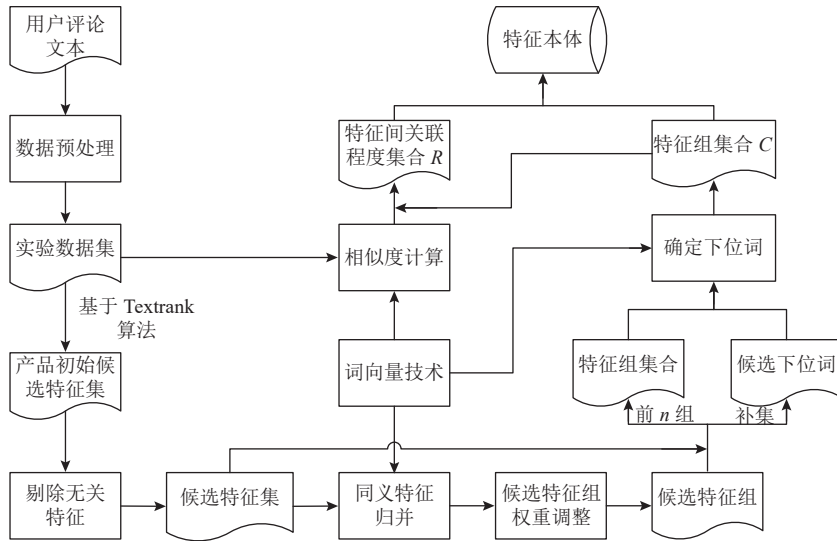


图2 特征本体构建流程

Figure 2 Process of feature ontology construction

3.2.1 特征本体基本模型

由于在线用户评论具有自由性和复杂性的特点, 词语之间可能是同义关系也可能是上下义关系。基于在线用户评论构建产品特征本体, 明确了产品特征及其之间的关系, 有助于将非结构化的产品评论转化为结构化的信息表达。产品特征本体由若干特征组组成, 特征本体的结构可由一个二元组 $\{F, R\}$ 来表示。其中, F 表示特征组集合, R 表示

特征组内相关联特征之间的关联程度集合。 $F = \{I, C, S, CH, SH\}$, I 表示特征组的唯一标识信息, C 为概念特征, 是所属该特征组的代表, $S = \{S_1, S_2, \dots, S_n\}$ 是 C 的同义词集, CH 表示 C 的下位词集, $SH = \{S_1H, S_2H, \dots, S_nH\}$ 为 S 的下位词集。 $R = \{D, R(v_i, v_j)\}$, D 表示的是 $R(v_i, v_j)$ 中特征词 v_i 与 v_j 之间的相似度。

以相机评论为例, 部分特征本体为

$$F_1 = \{01, \text{画质}, \text{防抖}, \text{像素}, \emptyset\};$$

$F_2 = \{02, \text{拍摄}, S(\text{拍照}, \text{录像}), \emptyset, SH[\emptyset, S_2H$
(续航, 屏幕)];

$R = \{0.78, (\text{画质}, \text{像素})\}$ 。

3.2.2 特征本体中概念及概念间关系的确定

本文使用 Word2vec 技术确定产品特征本体中概念及概念之间的关系。Word2vec 是在神经概率语言模型的基础上进行简化和改进后的词语语义计算工具, 利用 Word2vec 技术对产品在线评论进行词向量训练可以得到产品特征词之间的语义关系。Word2vec 由两大模型构成, 分别是 CBOW 模型和 Skip-gram 模型, 本文选用 Skip-gram 模型。Skip-gram 模型能够快速地完成对数 10 亿词的大规模数据训练, 可以根据当前词预测其上下文词及上下文词出现的概率, 且对低频词汇处理具有优势。如图 3 所示, Skip-gram 模型由 3 层网络模型构成, 包括输入层、投影层、输出层。

模型训练完成后, 将每个词语映射为一个词向量, 通过向量运算可以计算两个特征词之间的语义

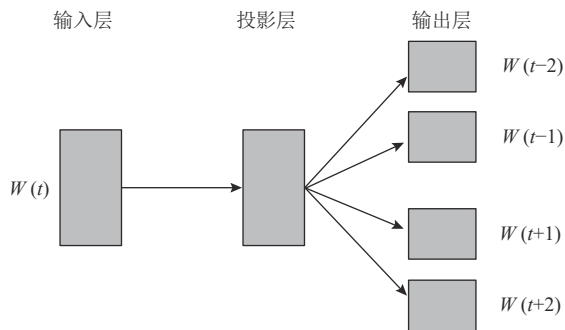


图 3 Skip-gram 模型

Figure 3 Skip-gram model

相似度, 判别词间关系。语义相似度计算方法如式 (1) 所示。

$$\text{sim}(e, f) = \frac{e \cdot f}{\|e\| \cdot \|f\|} \in [0, 1]. \quad (1)$$

其中, e 和 f 是两个特征词的词向量, $\|e\|$ 和 $\|f\|$ 分别表示 e 和 f 的模, $\text{sim}(e, f)$ 代表两个词之间的语义相似度, 取值为 $[0, 1]$ 。

确定特征本体中概念及概念间关系的流程如图 4 所示。本节所使用的主要符号如表 2 所示。

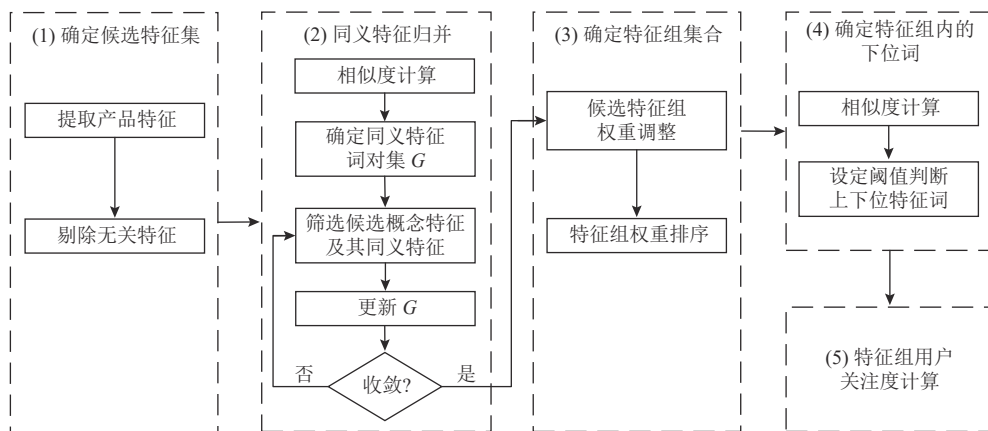


图 4 概念及概念间关系确定流程图

Figure 4 Flowchart for determining concepts and their relationships

1) 确定候选特征集。Textrank 算法是一种由 Pagerank 算法改进而来的图模型排序算法, 不需要提前利用多篇文档进行训练, 其通过建立图模型抽取关键词, 重点研究文本内部词语结构关系, 具有简单而高效的特点。本文利用 Textrank 算法抽取预处理后评论文本中的特征词, 取 T_r 权重值较大的名词 (n) 和动词 (v) 作为产品初始候选特征集, 剔除和产品特征无关的词后得 N 个高关注度特征词作为候选特征集。

2) 同义特征归并。首先, 利用 Word2vec 中的

Skip-gram 模型对语料文本进行训练, 对训练学习后的词语进行词向量转换, 运用式 (1) 计算 N 个候选特征词两两之间的语义相似度。

然后, 设置阈值 θ , 将相似度大于等于 θ 的候选特征词对定义为同义特征词对, 从而得到同义特征词对集 G 。 G 可表示为一个图 $G = (V, E)$, V 为节点代表同义特征词集, E 为边代表同义特征词对集。在 G 中先取出相似度最大的同义特征词对 (v_i, v_j) , 比较特征词 v_i 和 v_j 之间的 T_r 权重值, 将 T_r 值更大的特征词 v_i (这里假设 v_i 的 T_r 值大于 v_j 的 T_r 值) 作为候选概念

表2 主要符号说明

Table 2 Descriptions of main notations

符号	含义
v_i	特征词 i
v_j	特征词 j
T_r	特征词的初始权重值
N	高关注度特征词的个数
V	同义特征词集(结点)
E	同义特征词对集(边)
G	同义特征词对集
θ	同义特征归并时的阈值
v_{ip}	与 v_i 相邻的节点, $p = 1, 2, \dots, n$
v_{ipq}	与 v_{ip} 相邻的节点, $q = 1, 2, \dots, n$
T_r'	候选特征组的权重
v_{ij}	v_i 的同义特征, $j = 1, 2, \dots, l$
l	同义特征词的个数
m	特征本体中特征组的个数
θ_1	确定特征组内的下位词时的阈值
T_r''	调整后的特征组权重
v_{ih}	概念特征 v_i 的下位词
k	概念特征 v_i 的下位词的个数
v_{ijh}	特征 v_{ij} 的下位词
r	特征 v_{ij} 的下位词的个数
W	特征组用户关注度

特征, 特征词 v_j 定义为 v_i 的同义特征词。

接着, 寻找与 v_i 相邻的节点 v_{ip} ($p = 1, 2, \dots, n$)以及 v_{ip} 相邻的节点 v_{ipq} ($q = 1, 2, \dots, n$)比较 $\text{sim}(v_i, v_{ip})$ 与 $\max\{\text{sim}(v_{ip}, v_{ip1}), \text{sim}(v_{ip}, v_{ip2}), \dots, \text{sim}(v_{ip}, v_{ipn})\}$ 若 $\text{sim}(v_i, v_{ip}) > \max\{\text{sim}(v_{ip}, v_{ip1}), \text{sim}(v_{ip}, v_{ip2}), \dots, \text{sim}(v_{ip}, v_{ipn})\}$, 则把特征词 v_{ip} 定义为候选概念特征 v_i 的同义特征词。

最后, 在 G 中将候选概念特征 v_i 及其同义特征剔除, 从而得到新的同义特征词对集, 重复上述步骤, 直至再无同义特征词对时收敛结束。最终得到一个或多个包含同义特征词对的部分候选特征组, 其组内包含唯一的候选概念特征及一个或多个同义特征。

3) 确定特征组集合。对候选概念特征的定义是: 可能成为特征本体中概念特征的特征。对候选特征组的定义是: 可能成为特征本体中特征组集合一部分的特征组, 其内包含候选概念特征。为确定特征组集合, 需要先按照公式(2)计算步骤1得到候选特征组的权重。以候选概念特征 v_i 为例, 其所在候选特征组的权重为

$$T_r'(v_i) = T_r(v_i) + \sum_{j=1}^l [\text{sim}(v_i, v_{ij}) \cdot T_r(v_{ij})]。 \quad (2)$$

式中, $T_r'(v_i)$ 表示特征 v_i 所在的特征组权重, $T_r(v_i)$ 为特征 v_i 的原始权重, $\text{sim}(v_i, v_{ij})$ 为特征 v_i 与特征 v_{ij} 之间的相似度。

然后, 把未在候选特征组内的每个候选特征分别表示为一个单独的候选特征组, 每个候选特征是其组内的候选概念特征, 其所在特征组的权重即为该候选特征的原始权重。对所有候选特征组的权重进行降序排序, 取前 m 个作为特征本体的特征组集合, 此时特征组内只包含概念特征与其同义特征或只包含概念特征。

4) 确定特征组内的下位词。定义 m 个特征组在候选特征集中的补集为候选下位词集, 计算每个候选下位词与 m 个特征组内特征的相似度。设定阈值 θ_1 , 若候选下位词与若干特征组内特征的相似度大于等于 θ_1 , 则取与该候选下位词相似度最大的特征作为其上位词; 若候选下位词与 m 个特征组内所有特征之间的相似度均小于 θ_1 , 则舍弃该候选下位词。

5) 特征组用户关注度计算。经过以上步骤确定了特征本体的特征组集合及每个特征组组内特征之间的关系。为衡量各个特征组受用户关注的程度, 本文按照公式(3)重新调整各特征组的权重, 以概念特征 v_i 为例, 其所在特征组调整后的权重为

$$T_r''(v_i) = T_r'(v_i) + \sum_{h=1}^k [\text{sim}(v_i, v_{ih}) \cdot T_r(v_{ih})] + \sum_{j=1}^l \sum_{h=1}^r [\text{sim}(v_i, v_{ij}) \cdot \text{sim}(v_{ij}, v_{ijh}) \cdot T_r(v_{ijh})]。 \quad (3)$$

式中, $\text{sim}(v_i, v_{ih})$ 为概念特征 v_i 与其下位特征 v_{ih} 之间的相似度, $T_r(v_{ih})$ 为特征 v_{ih} 的原始权重, $\text{sim}(v_i, v_{ij})$ 为概念特征 v_i 与其同义特征 v_{ij} 之间的相似度, $\text{sim}(v_{ij}, v_{ijh})$ 为特征 v_{ij} 与其下位特征 v_{ijh} 之间的相似度, $T_r(v_{ijh})$ 为特征 v_{ijh} 的原始权重。

最后, 对调整后的权重 $T_r''(v_i)$ 归一化处理, 从而得到各特征组的用户关注度。以概念特征 v_i 为例, 其所在特征组的用户关注度 $W(v_i)$ 为

$$W(v_i) = \frac{T_r''(v_i)}{\sum T_r''(v_i)}。 \quad (4)$$

3.3 特征—评论情感分析

情感分析又称观点挖掘, 是指对输入的文本数

据进行自然语言处理,分析文本所蕴含的情绪。通过情感分析可以挖掘文本作者所要表达的态度,一般分为积极、消极和中性 3 种情感态度。目前,情感分析被广泛应用于在线评论分析,对产品特征的评价实施情感分析可以挖掘用户对不同产品特征属性的情感倾向。

在产品评论中,同一用户对不同产品特征可能持不同的情感态度。例如,某用户对相机的评论:“拍摄非常清晰,但续航很差”,可以看出该用户对特征“拍摄”的评价态度是积极的,而对特征“续航”的评价态度是消极的。简单地用整条评论的情感分析结果表示同一用户对不同特征的态度显然不合适。为了更准确地获取用户对不同特征的态度,本研究首先根据构建的产品特征本体从原始评论中筛选出包含所有本体特征的用户评论,把用户评论切分为短句,使得每个短句只蕴含用户对一个产品特征的评价,例如把评论“拍摄非常清晰,但续航很差”切分为两个短句:“拍摄非常清晰”、“续航很差”;然后,对切分后的短句按照产品特征本体中的特征进行分类,使得每类短句集只包含用户对一个本体特征的评价;最后,对分类后的短句集实施情感分析,得到每个短句的情感极性。

借助百度 AI 开放平台提供的情感倾向分析接口对特征—评论文本进行细粒度情感极性分类。百度情感倾向分析是结合了机器学习和情感词典的一种高效率情感分析模型,其训练成功和通用准确率高于 92%,支持多种开发语言进行调用,可用性及准确率都非常高。利用百度情感倾向分析模块判断评论文本情感极性(积极、消极、中性)的步骤如下:首先,在百度 AI 中创建情感倾向分析模块;然后,利用 Python 编程语言获取 Access Token;接着在 Python 中调用百度 AI 接口对输入的评论文本进行情感分析;最后输出评论文本的情感极性。

根据输出的情感极性结果计算用户对产品特征组及细粒度产品特征的质量满意度(Quality Satisfaction, QS)。对质量满意度的定义是:情感极性为积极的评论短句数量与总评论短句数量之比,可表示用户对产品特征组或细粒度产品特征的满意程度, QS 越大意味着用户对相应产品特征组或产品特征越满意。QS 的计算方法如式(5)所示。

$$QS = \frac{\text{积极评论短句数量}}{\text{总评论短句数量}} \quad (5)$$

3.4 用户画像模型实现

本文基于产品特征本体对评论文本实施情感分析,实现用户评论信息属性,得到用户关注度画像与粗、细粒度特征质量满意度画像,结合用户信息属性及产品信息属性呈现整体用户画像。为识别产品核心优势特征与亟待优化特征,本文根据 W 和 QS 计算特征组优选度,特征组优选度越大,在进行产品迭代时越优先考虑,优选度的计算方法见式(6)。

$$\text{优选度} = \beta W + (1 - \beta)(1 - QS) \quad (6)$$

对于优选度较大的特征组,比较其质量满意度与产品总体质量满意度,进一步判断核心优势特征组与亟待优化特征组。利用文字描述、词云、图表等方式呈现产品细粒度特征画像,分析产品优势,聚焦亟待优化特征,对标同类型产品利用 K 值法制定产品优化策略,为商家进行产品迭代提供参考方案。

4 实证分析

在线评论的数据量会影响数据挖掘的质量,数据量太小不仅不具有代表性且不适合使用大数据挖掘技术挖掘用户需求,所以应用本文方法的产品应具有一定的在线评论数据。本文以大疆 Pocket2 相机为研究对象,从京东爬取其在线用户评论语料,评论截止时间为 2022 年 8 月 7 日,经数据清洗后得到 2 681 条有效评论数据,以此为数据源提取用户高关注度产品特征并构建产品特征本体。基于产品特征本体对评论文本实施情感分析,并从用户关注度和质量满意度两个维度实现用户画像。最后,本文通过整体用户画像与产品细粒度特征画像,聚焦产品核心优势特征与待优化特征,采用 K 值法制定产品优化策略,为企业提供较优产品迭代方案。

4.1 产品特征本体构建

本实验首先对获得的有效评论数据进行 jieba 分词、去停用词、词性标注等预处理工作;然后基于 TextRank 算法选取权重前 100 的名词(n)和动词(v)作为产品初始候选特征集,剔除和产品特征无关的词后得到高关注度候选特征集;接着利用 Word2vec 技术确定产品特征本体中的概念及概念之间的关系;最后得到的产品特征本体包含 6 组特征,结果如表 3 所示,按照式(4)计算各特征组的用户关注度(W),结果如图 5 所示。

表 3 产品特征组集合

Table 3 Product feature groups

I	C	CH	S	SH
“画质”组	画质	像素	防抖	
			照片	画面
			拍照	
“拍摄”组	拍摄		视频	
			录像	续航、电池、屏幕
			剪辑	
“便携性”组	便携性		操作	
“云台”组	云台			
“防水”组	防水			
“功能”组	功能		镜头	连接

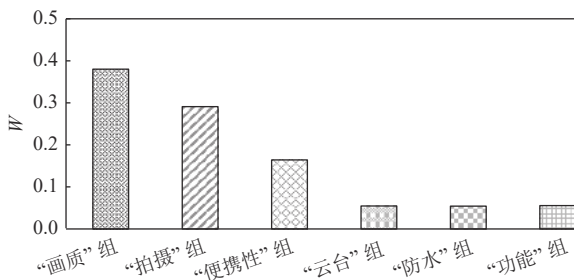


图 5 特征组用户关注度画像

Figure 5 User attention profiles of feature groups

4.2 特征—评论情感极性分类

本实验面向表 3 中的各个产品特征把评论语料切分为短句, 使每个短句只蕴含对一个本体特征的情感评价, 共得到 2 830 条评论短句。在百度 AI 中创建情感倾向分析模块得到 AppID、API Key、Secret Key 等信息, 利用 Python 编程获取 Access Token, 上传不同特征对应的评论短句集并返回情感极性数据。按照式 (5) 计算产品特征组及组内各个产品特征的质量满意度, 其中各产品特征组的质量满意度画像如表 4 所示, 组内各特征的质量满意度画像如图 6 所示。

4.3 整体用户画像呈现

结合图 7, 从用户信息属性、产品信息属性、在线用户评论信息属性 3 个维度呈现整体用户画像。

可以看出, 购买此相机的用户中京东 PLUS 会员用户的占比达 67.7%, 说明该产品的消费群体具有一定的经济实力, 购买力较强。相机的价格范围是 2 499 ~ 3 818 元, 其中“经典黑全能套装+随心换两年版”的价格最高为 3 818 元, “经典黑标准版+官方标配”(只含单个相机)的价格最低为 2 499 元, 且购买“经典黑标准版+官方标配”的人数最多。可以看

表 4 特征组质量满意度画像

Table 4 Quality satisfaction profiles of feature groups

特征组	总评论数	积极评论数	QS
“画质”组	794	742	0.934 5
“拍摄”组	995	805	0.809 0
“便携性”组	514	457	0.889 1
“云台”组	126	113	0.896 8
“防水”组	108	82	0.759 3
“功能”组	293	246	0.839 6
总计	2 830	2 445	0.864 0

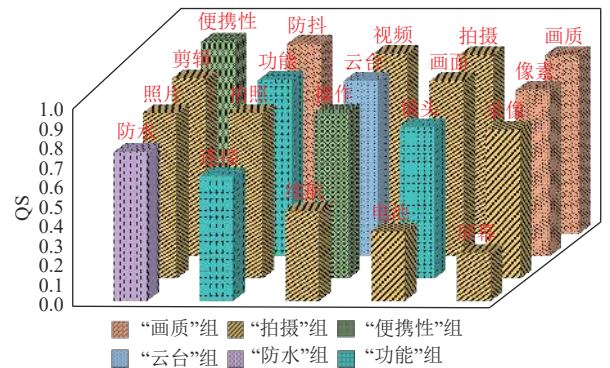


图 6 组内各特征质量满意度画像

Figure 6 Quality satisfaction profiles of each feature in groups

出, 多数用户更愿意以最低价格购买单个相机, 所以厂商应把注意力更多地放在相机价格和质量两个方面。相机的适用场景为 Vlog、慢动作录制, 可以在旅游、聚会等场合用来记录生活。相机的特色功能包括电子防抖、延时摄影、自动美颜、机械增稳、智能跟随、一键剪辑、混合对焦等, 可以看出相机的功能较为强大。

对图 5 和表 4 进行分析可以看出, 产品总体质量满意度较高, 整体来说, 用户对该相机较为满意。用户最关注的特征组是“画质”组且该特征组的质量满意度最高, “拍摄”组的用户关注度次之但其质量满意度低于总体满意度, “防水”组、“功能”组的质量满意度同样低于总体满意度, 其不失为产品优化的候选对象。按照式 (6) 计算各特征组的优选度 (取 $\beta=0.5$), 结果如图 8 所示, 可以看出, 综合考虑用户对产品特征的关注程度及用户对各特征组的态度之后, 商家在进行产品迭代时有理由优先考虑“拍摄”组、“画质”组、“防水”组所属特征。为确定产品迭代策略, 本实验聚焦产品“画质”组、“拍摄”组和“防水”组进行细粒度特征画像分析。

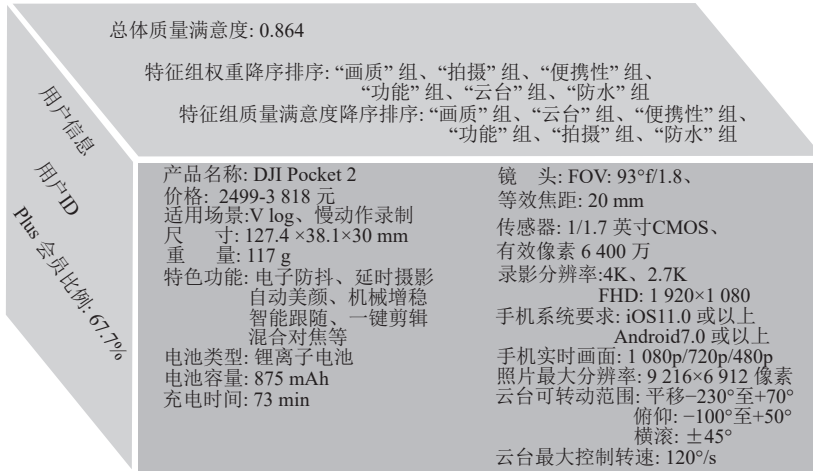


图 7 大疆 Pocket 2 整体画像

Figure 7 Overall profile of DJ Pocket 2

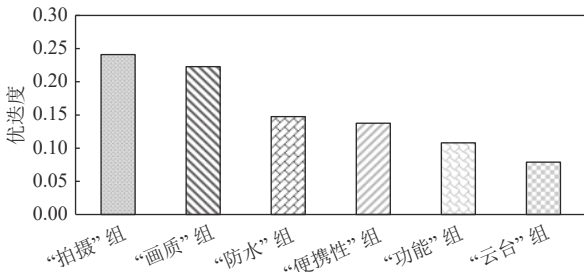


图 8 特征组优选度

Figure 8 Feature group priority

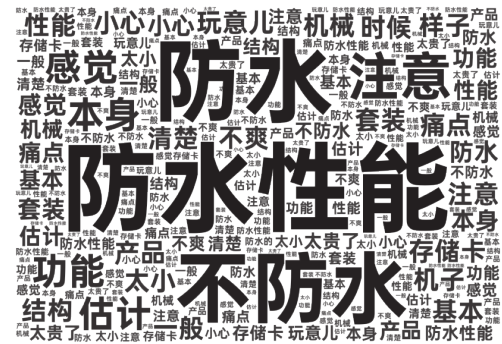
4.4 细粒度特征画像呈现

从图 8 可以看出,“画质”组的优选度仅次于“拍摄”组,在进行产品迭代时,其是重点考虑的对象。表 4 显示,“画质”组的质量满意度最高,高于总体质量满意度,结合产品信息和在线用户评论信息可以发现,该相机具有电子防抖功能、4K/60P 视频拍摄能力,搭载 6 400 万像素 CMOS 传感器,评论中有诸如“画质效果:很清晰”“主要是像素很高,总体来说值得购买”“防抖性能优秀”等好评。该产品在画质方面得到用户很高的认可度,且从图 5 可以看出该特征组是用户最关注的特征组,因此商家要继续保持该方面的优势,可以将其作为产品营销的亮点。

除“画质”组外,“拍摄”组和“防水”组是优选度最大的两个特征组,但其质量满意度低于总体满意度,亟待从中挖掘待优化特征。分别筛选出“拍摄”组和“防水”组包含的消极评论,然后对每组消极评论绘制词云图,识别亟待优化特征,结果如图 9 所示。可以看出,“拍摄”组的问题集中于续航、电



(a) “拍摄”组



(b) “防水”组

图 9 亟待优化特征词云图

Figure 9 Word cloud of urgently optimizable features

池、屏幕 3 个特征,具体表现在电池续航时间短,电池易发热,屏幕太小等,针对特征可以看到如“续航时间特别短,根本没法用”、“用了不到五分钟电池就开始发热了”、“屏幕小,左右滑动不太灵敏”等评论;“防水”组的问题是相机不防水,可以看到如“没有任何防水性能,防水得买防水盒”的评论。

由以上分析可知,在亟待优化的特征组中,续航、电池、屏幕、防水这 4 个特征的问题最为突

出, 属于亟待优化特征。基于以上4个特征, 本实验进行细粒度特征情感分析, 计算每个特征对应短句集的情感极性, 得到各特征的质量满意度如图10所示。结果显示这几个特征的QS值相对偏小, 说明用户对该相机的这几个特征不太满意, 尤其是特征屏幕, 其QS值仅有0.25, 屏幕太小导致了用户的使用体验很差, 相机屏幕急需改进。

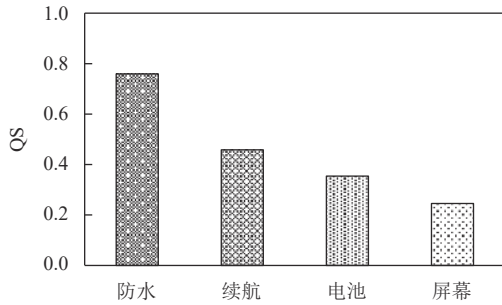


图10 大疆 Pocket 2 特征质量满意度

Figure 10 Feature quality satisfaction of DJ Pocket 2

4.5 差异化产品迭代设计

本文依据以上4个亟待优化特征进行差异化产品迭代设计, 最终得到多种产品优化设计方案。分别将关注每个亟待优化特征的用户交叉分类, 在优化某个亟待优化特征的同时, 确定是否优化其余亟

待优化特征, 以及具体的优化策略, 从而实现产品差异化迭代设计。该过程主要分为以下4个步骤。

1) 分别筛选出含各亟待优化特征的用户评论, 本实验得到四组评论集, 即屏幕组、电池组、续航组和防水组评论集。

2) 在各组评论集中, 将涉及其他亟待优化特征的评论语句挑选出来, 按照本文方法对评论语句进行分类、切分和情感分析, 得到组内其余亟待优化特征的组内质量满意度 QS' 。

3) 确定每组各特征是否优化。在某个亟待优化特征的评论集中, 若该评论集对应的用户群体对其其他的亟待优化特征的质量满意度相等或更低, 则认为该用户群体同时希望优化其他特征。优化某个亟待优化的特征时, 比较组内其他亟待优化特征的组内质量满意度 QS' 与该亟待优化特征的质量满意度 QS (见图10), 优化组内质量满意度 $QS' \leq QS$ (见 图 10), 优化组内质量满意度 QS' 小于等于该亟待优化特征质量满意度 QS 的其他特征 (例如, 特征续航的质量满意度为 $QS_{续航}$, 在续航组中, 特征屏幕的组内质量满意度 $QS' \leq QS_{续航}$, 则同时优化特征屏幕)。本实验得到的各组组内特征优化结果如表5所示。

表5 各组组长内特征优化结果

Table 5 Optimization results of intra-group features in each group

特征	屏幕组		电池组		续航组		防水组	
	$QS' \leq QS_{续航}$	是否优化	$QS' \leq QS_{电池}$	是否优化	$QS' \leq QS_{续航}$	是否优化	$QS' \leq QS_{防水}$	是否优化
屏幕	√	是	√	是	√	是	√	是
电池		否	√	是		否	√	是
续航		否		否	√	是	√	是
防水		否		否	√	是	√	是

4) 对标同类型产品确定每组的产品优化策略。顾客购买意愿主要受到产品质量和价格的影响^[20], 顾客总是倾向于以较低价格买到质量较高的产品。但是, 产品升级优化一般以价格增加为代价, 因此需要探究产品增价与产品质量的关系, 进而以此确定产品优化策略。本实验对标同类型产品并利用K值法制定每组产品优化策略, 具体方法如下:

对标同类型 Vlog、口袋运动相机: Insta360 ONE X3。相比大疆 Pocket 2, Insta360 ONE X3 具有 10 m 防水功能, 电池容量更大, 发热不明显, 且拥有较大的屏幕, 其标配版售价比大疆 Pocket 2

高 500 元。本文从京东和天猫平台爬取 1 658 条 Insta360 ONE X3 的在线用户评论, 筛选出含以上 4 个特征的评论语句, 按照本文方法对评论语句进行切分、分类和情感分析, 得到 Insta360 ONE X3 4 个特征的质量满意度如图 11 所示, 结果显示其各特征的 QS 值均高于大疆 Pocket 2。

一般来说, 优化的产品特征数量越多, 产品增价越高。在本文的对标实验中, 优化以上 4 个特征, 相机总增价为 500 元, 本实验假设每优化其中一个特征相机增价 125 元。每组优化的产品特征数量不同, 则相应迭代产品的增价也不同。结合表 5

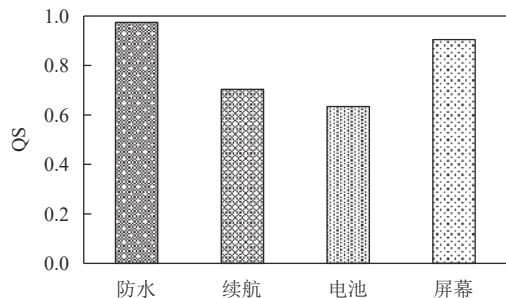


图 11 Insta360 ONE X3 特征质量满意度

Figure 11 Feature quality satisfaction of Insta360 ONE X3

可以得到每组迭代产品的产品增价, 结果如表 6 所示。

得到了各组迭代产品的产品增价后, 本文利用 K 值法, 以两个点拟合的曲线来衡量每组大疆 Pocket 2 相机的价格增价与每组待优化特征的质量满意度 QS 或组内质量满意度 QS' 之间的关系。第一个点的横坐标为 0, 表示大疆 Pocket 2 当前价格下

无增价, 纵坐标为大疆 Pocket 2 当前价格下各组组内待优化特征的质量满意度 QS 或组内质量满意度 QS' ; 第二个点的横坐标为每组的产品增价, 纵坐标为预期大疆 Pocket 2 优化后达到 Insta360 ONE X3 对应特征的质量满意度 QS 。各组拟合结果如图 12 所示, 图中每条曲线都拥有一个斜率 K 值。

以每条曲线的斜率 K 值衡量组内每个亟待优化特征的边际质量满意度, 在进行产品优化时对不同的优化特征投入相同的资金, K 值越大的特征其提升的质量满意度越高, 应优先对其进行优化。因此可以依据 K 值大小确定组内各特征的优化优先级, 以此分别制定每组的产品优化策略。本实验最终的产品差异化迭代设计方案如表 6 所示, 共得到 4 种迭代产品, 每种迭代产品的优化特征、优化策略不同, 同时该迭代方案实现了产品差异化定价, 为消费者提供了多样的购买选择, 提升了产品的竞争力。

表 6 差异化产品迭代设计方案

Table 6 Iterative designing schemes of different products

迭代产品	优化特征	特征优化优先级				预期优化效果	相机增价/元
		(1)	(2)	(3)	(4)		
A (屏幕组)	屏幕	屏幕				屏幕尺寸增大	125
B (电池组)	屏幕、电池	屏幕	电池			屏幕尺寸增大、改善电池发热	250
C (续航组)	屏幕、续航、防水	屏幕	防水	续航		屏幕尺寸增大、增加防水功能、续航能力增强	375
D (防水组)	屏幕、电池、续航、防水	续航	屏幕	防水	电池	续航能力增强、屏幕尺寸增大、增加防水功能、改善电池发热	500

5 结束语

互联网时代, 产品在线用户评论内容日益真实且丰富, 数据驱动下的用户需求挖掘具有重要的理论意义和应用价值, 为克服传统产品设计耗时耗力、效率较低等困难提供了新的解决思路。基于此, 本文提出了面向产品迭代设计的用户画像及建模研究, 构建用户画像概念模型, 基于产品特征本体, 使用 Word2vec 技术和情感分析方法得到产品特征用户关注度和质量满意度; 结合用户信息属性及产品信息属性呈现整体用户画像, 根据优选度判断优势特征组与亟待优化特征组, 并采用文字描述、词云、图表等方式呈现产品粗、细粒度特征画像; 最后利用 K 值法制定产品优化策略, 得到差异化产品迭代设计方案。以某品牌相机为例进行实验, 结果表明该方法能够有效挖掘用户高关注度产品特征及用户对特征的满意度, 识别用户的迫切诉

求, 为商家进行产品迭代设计提供较优参考方案, 同时拓展了用户画像的应用领域。

本研究的主要贡献如下。其一, 数据驱动下基于在线用户评论挖掘用户高关注度特征, 并利用情感分析方法识别用户的态度, 能够帮助商家深入了解用户需求; 其二, 挖掘产品核心优势特征与亟待优化特征, 在此基础上为商家提供了一种具体、可行的产品迭代设计方案, 有助于商家进行产品改进和市场营销; 其三, 用户评论影响产品口碑及消费者购买决策, 基于用户负面评论的产品优化有利于减少产品负面评论的数量, 增强顾客对产品的认可度, 提高用户购买转化率, 增加商家的经济收益。

未来值得深入研究的问题有: 1) 由于在线购物平台的用户个人隐私保护机制, 本文在用户画像方面较少涉及用户个人信息, 未来可以考虑从其他方面丰富用户信息属性; 2) 本文方法属于静态分析, 未来研究可以尝试引入时间序列动态预测用户需

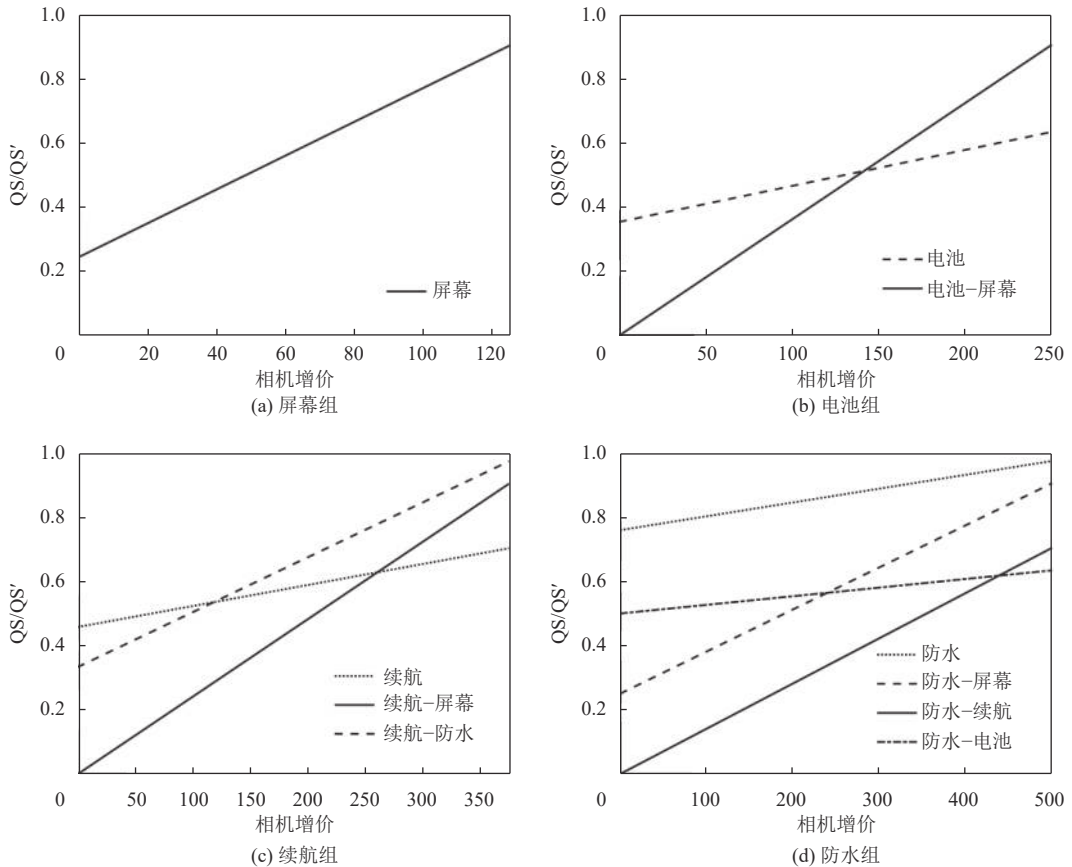


图 12 4组特征的K值图

Figure 12 K-values of features in four groups

求, 以此进行产品迭代设计; 3) 本文仅针对单一产品进行了实验分析, 未来研究可以尝试将本文方法拓展到产品族的迭代设计当中。

参考文献:

- [1] 中国互联网络信息中心. 第52次中国互联网络发展状况统计报告[R/OL]. (2023-8-28)[2023-12-05]. <https://www.cnnic.net.cn/n4/2023/0828/c88-10829.html>.
- [2] 王军, 王栋, 刘勇, 等. 基于用户参与思维的产品设计需求获取方法[J]. 机械设计, 2022, 39(10): 126-132.
WANG Jun, WANG Dong, LIU Yong, et al. Product design requirements acquisition method based on user engagement thinking[J]. Journal of Machine Design, 2022, 39(10): 126-132.
- [3] 贾丹萍, 靳健, 耿骞, 等. 感性工学视角下的用户需求挖掘研究[J]. 情报学报, 2020, 39(3): 308-316.
JIA Danping, JIN Jian, GENG Qian, et al. A kansei engineering integrated approach for customer-needs mining from online product reviews[J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(3): 308-316.
- [4] DU Y, LIU D, DUAN H. A textual data-driven method to identify and prioritise user preferences based on regret/rejoicing perception for smart and connected products[J]. International Journal of Production Research, 2022, 60(13): 4176-4196.
- [5] 黄晟青, 张杰, 李明, 等. 分离视角下产品迭代设计的决策信

息挖掘方法[J]. 机械设计, 2021, 38(5): 138-144.

HUANG Shengqing, ZHANG Jie, LI Ming, et al. Decision information mining method of product iteration design from the perspective of separation[J]. Journal of Machine Design, 2021, 38(5): 138-144.

- [6] YANG C, WU L, TAN K, et al. Online user review analysis for product evaluation and improvement[J]. Journal of Theoretical and Applied Electronic Commerce Research, 2021, 16(5): 1598-1611.
- [7] 沈超, 王安宁, 陆效农, 等. 基于在线评论的客户偏好趋势挖掘[J]. 系统工程学报, 2021, 36(3): 289-301.
SHEN Chao, WANG Anning, LU Xiaonong, et al. Customer preference trend mining based on online reviews[J]. Journal of Systems Engineering, 2021, 36(3): 289-301.
- [8] JOUNG J, KIM H. Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews[J]. International Journal of Information Management, 2023, 70: 102641.
- [9] 王克勤, 刘朝明. 基于在线评论的重要度绩效竞争对手分析的产品设计改进方法[J]. 计算机集成制造系统, 2022, 28(5): 1496-1506.
WANG Keqin, LIU Chaoming. Product design improvement based on importance performance competitor analysis of online review[J]. Computer Integrated Manufacturing Systems, 2022, 28(5): 1496-1506.

- [10] 吴运明, 王令村, 魏子栋, 等. 基于 Canopy-Kmeans 的移动端商务用户需求聚合挖掘及分析研究[J]. 情报科学, 2022, 40(10): 97-106.
WU Yunming, WANG Lingcun, WEI Zidong, et al. Mobile Commerce User Demand Aggregation and Mining Based on Canopy-Kmeans algorithm[J]. Information Science, 2022, 40(10): 97-106.
- [11] WANG P, GONG Y, XIE H, et al. Applying CBR to machine tool product configuration design oriented to customer requirements[J]. Chinese Journal of Mechanical Engineering, 2017, 30: 60-76.
- [12] SON Y, KIM W. Development of methodology for classification of user experience (UX) in online customer review[J]. Journal of Retailing and Consumer Services, 2023, 71: 103210.
- [13] GOLDBERG D M, ABRAHAMS A S. Sourcing product innovation intelligence from online reviews[J]. Decision Support Systems, 2022, 157: 113751.
- [14] XU Y, WANG Q, ROXAS R E O, et al. Online shopping user portrait model considering psychological and emotional factors[J]. International Journal of Neuropsychopharmacology, 2022, 25(Supplement_1): A21-A21.
- [15] 陈添源, 吴锦辉, 杨思洛. 数据驱动的高校图书馆用户画像构建研究[J]. 国家图书馆学报, 2023, 32(3): 64-75.
CHEN Tianyuan, WU Jinhui, Yang Siluo. Research on the Construction of University Library User Profile Based on Data-Driven[J]. Journal of the National Library of China, 2023, 32(3): 64-75.
- [16] JIANG H, YIN X. Association between community psychological label and user portrait model based on multimodal neural network[J]. Frontiers in Psychology, 2022, 13: 918274.
- [17] 滕春娥, 何春雨. 在线医疗社区用户画像构建与应用[J]. 图书情报工作, 2021, 65(12): 147-154.
TENG Chun'e, HE Chunyu. Construction and application of user persona in online health community[J]. Library and Information Service, 2021, 65(12): 147-154.
- [18] HU M, LI H, SONG H, et al. Tourism demand forecasting using tourist-generated online review data[J]. Tourism Management, 2022, 90: 104490.
- [19] DOU R, LI W, NAN G, et al. How can manufacturers make decisions on product appearance design? A research on optimal design based on customers' emotional satisfaction[J]. Journal of Management Science and Engineering, 2021, 6(2): 177-196.
- [20] 马东升, 宋华明, 古晓宇, 等. 基于战略顾客行为的质量差异化产品定价策略[J]. 管理科学学报, 2021, 24(6): 76-87.
MA Dongsheng, SONG Huaming, GU Xiaoyu, et al. Behavior-based pricing strategy of quality-differentiated products[J]. Journal of Management Sciences in China, 2021, 24(6): 76-87.

(责任编辑: 孟晓燕)

(上接第 56 页)

- [7] 宫勇, 张梦哲, 侯冠华, 等. 颜色分组对图标视觉搜索效率的影响[J]. 计算机辅助设计与图形学学报, 2023, 35(4): 599-608.
GONG Yong, ZHANG Mengzhe, HOU Guanhua, et al. Effect of color grouping on icon visual search efficiency[J]. Journal of Computer-Aided Design & Computer Graphics, 2023, 35(4): 599-608.
- [8] MARSH J, MILLER M E, GEISELMAN E E, et al. Display size and distractor complexity effects on visual search[J]. Displays, 2021, 66: 101988.
- [9] CHANG T, CHEN N, YU P. The effects of presentation method and information density on visual search ability and working memory load[J]. Computers & Education, 2012, 58(2): 721-731.
- [10] HSIAO C, WANG M, LIU Y, et al. Effects of Chinese character size, number of characters per line, and number of menu items on visual search task on tablet computer displays for different age groups[J]. International Journal of Industrial Ergonomics, 2019, 72: 61-70.
- [11] XIE X, SONG F, LIU Y, et al. Study on the effects of display color mode and luminance contrast on visual fatigue[J]. IEEE Access, 2021, 9: 35915-35923.
- [12] SHEN Z, ZHANG L, Li R, et al. The effects of color combinations, luminance contrast, and area ratio on icon visual search performance[J]. Displays, 2021, 67: 101999.
- [13] 吴晓莉, 许盼盼, 江晓曼. 基于注意捕获的工业监控界面色彩层级编码研究[J]. 工业工程与管理, 2022, 27(4): 32-38.
WU Xiaoli, XU Panpan, JIANG Xiaoman. Color Level Coding on Industrial Monitoring Interface Based on Attention Capture[J]. Industrial Engineering and Management, 2022, 27(4): 32-38.
- [14] WELLER D. The effects of contrast and density on visual web search[EB/OL]. [2023-10-16]. <https://api.semanticscholar.org/CorpusID:61529908>.
- [15] WANG A, HWANG S, KUO H. Effects of text / background color combination, ambient illuminance, and display type on discriminating performance for young and elderly users[J]. Journal of the Society for Information Display, 2012, 20(2): 87-93.
- [16] 吴晓莉, 陈艳, 吴新兵. 工业控制系统出错因子—视觉感知—界面布局关联效应[J]. 工业工程与管理, 2020, 25(6): 9-14.
WU Xiaoli, CHEN Yan, WU Xinbing, et al. Correlative effect among error factors-visual perception-interface layout in industrial control system[J]. Industrial Engineering and Management, 2020, 25(6): 9-14.
- [17] FU X, FENG D, JIANG X, et al. The Effect of correlated color temperature and illumination level of led lighting on visual comfort during sustained attention activities[J]. Sustainability, 2023, 15(4): 3826.
- [18] WANG Y, HUANG H, CHEN G. Effects of lighting on ECG, visual performance and psychology of the elderly[J]. Optik, 2020, 203: 164063.
- [19] NIELSEN J. Usability laboratories[J]. Behaviour & Information Technology, 1994, 13(1-2): 3-8.
- [20] LI Y, HUANG Y, LI X, et al. The influence of brightness combinations and background colour on legibility and subjective preference under negative polarity[J]. Ergonomics, 2022, 65(8): 1046-1056.

(责任编辑: 孟晓燕)