

基于在线品牌社区意见领袖的用户关键需求挖掘

申彦¹, 刘春华²

(1. 江苏大学 信息管理与信息系统系, 江苏 镇江 212013;
2. 江苏大学 计算机科学与通信工程学院, 江苏 镇江 212013)

摘要: 随着社会的快速发展以及技术的不断进步, 人们生活节奏不断加快, 对产品的需求也在快速发生着变化。在线评论是目前用户需求表达的重要渠道。为克服不加区分挖掘所有评论的传统用户需求挖掘方法耗时过长, 难以聚焦用户关键需求的问题, 从用户关键性与需求关键性的双关键性角度出发, 研发了一种基于在线品牌社区意见领袖的用户关键需求挖掘方法以快速获取用户重要需求, 简称 KEY-DEMANDS-OL。该方法依据帕累托法则, 依托意见领袖评论大数据, 采用优化的情感程度及初始改进率, 结合 KANO 模型对用户关键需求进行挖掘。该方法不仅考虑了程度副词的语义信息, 提高了情感分析的准确率, 而且能够完成意见领袖的生成内容与 KANO 模型的自动整合, 实现用户关键需求的获取与分类。研究表明, 与挖掘所有评论的传统方法相比, KEY-DEMANDS-OL 可以快速获取用户的关键需求, 为企业制定产品优化方案提供辅助决策支持。

关键词: 用户需求挖掘; 在线品牌社区; KANO 模型; 用户评论; 意见领袖

中图分类号: TP391

文献标识码: A

文章编号: 1673-629X(2024)02-0023-09

doi: 10.3969/j.issn.1673-629X.2024.02.004

Users' Key Demands Mining Based on Opinion Leaders in Online Brand Community

SHEN Yan¹, LIU Chun-hua²

(1. Department of Information Management and Information System, Jiangsu University, Zhenjiang 212013, China;
2. School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang 212013, China)

Abstract: With the rapid development of society and the continuous progress of technology, the pace of people's lives is constantly accelerating. At the same time, people's demand for products is also rapidly changing. Currently, the online reviews are the important channels through which the users can express their demands. To overcome the problem that the traditional user demand mining method of mining all comments without discriminating takes too long and is difficult to focus on the key needs of users, starting from both the key users and the key demands, a method for users' key demands mining based on the opinion leaders in online brand community is developed to quickly obtain the users' important demands, which can be called KEY-DEMANDS-OL for short. It discovers the users' key demands integrated with KANO model based on the reviews big data of opinion leaders according to the Pareto principle, which also improves the initial improvement rates and the emotional degree. This novel method increased the accuracy of emotional analysis by taking the semantics of degree adverbs into account. In addition, it can integrate the UGC of opinion leaders with the KANO model automatically and obtain the users' key demands and the category of the demands. The research showed that comparing to the traditional method that mining all the uses' reviews, the KEY-DEMANDS-OL can obtain the key demands of users quickly and provide auxiliary decision-making support for enterprises to develop product optimization plans.

Key words: users' demands mining; online brand community; KANO model; users' reviews; opinion leaders

0 引言

随着社会的快速发展与技术的不断进步, 产品迭代更新周期越来越短, 用户需求也在快速发生变化。

企业要想在不断变化的市场中占据优势, 就必须深刻理解和把握用户的需求, 并将其作为产品改进和提升的依据, 快速响应动态变化的用户需求。

收稿日期: 2023-03-02

修回日期: 2023-07-05

基金项目: 国家社会科学基金项目(21BTQ070); 国家自然科学基金项目(61702229); 江苏省产学研合作项目(BY2021075)

作者简介: 申彦(1982-), 男, 副教授, 硕导, 博士, CCF 会员(O4282M), 通信作者, 研究方向为社会计算、大数据挖掘; 刘春华(1987-), 女, 硕士生, 研究方向为大数据挖掘、舆情分析。

互联网时代的到来,催生了大量的社交平台。为了聚焦品牌,维护客户关系管理,企业纷纷创建了自身品牌专属的社交平台即在线品牌社区。从在线品牌社区中的海量用户及其评论信息中快速捕获用户关键需求成为了企业新的挑战与机遇。

KANO 模型是根据马斯洛需求层次理论,研究需求与满意度的经典模型,然而其需求挖掘过程融入了过多调查人员的主观性^[1-2]。在线品牌社区中的海量评论是用户关切的重要表达,已成为用户关键需求获取的重要渠道,弥补了 KANO 模型主观性强的不足。因此,该文从“双关键性”即用户的关键性和需求的关键性双角度出发,提出了一种基于在线品牌社区意见领袖的用户关键需求挖掘方法(Users' Key Demands Mining Based on the Opinion Leaders in Online Band Community),简称 KEY-DEMANDS-OL,快速捕获用户关键需求,其主要贡献如下:

(1)克服 KANO 模型主观性较强的不足,利用客观评论大数据获取用户需求。

(2)构建了在线品牌社区意见领袖识别指标体系,通过意见领袖评论大数据获取用户关键需求。

(3)在贝叶斯情感极性分类时考虑了程度副词的语义,进一步精确了用户评论情感极性的分类。

1 相关研究工作

现有用户需求挖掘方法可依据是否考虑了需求类别划分为两类^[3]。

未考虑需求类别的用户需求挖掘方法一般会将关注度、情感程度低的产品特征直接作为用户需求。李贺等^[4]对在线手机评论进行评论主题及产品特征挖掘,有效识别了用户需求要素。吴东胜等^[5]结合观点值和关注值构建了产品的需求挖掘模型。纪雪等^[6]则根据产品属性的平均满意程度来辅助确定新产品的开发需求。邓昭等^[7]依据关键词的权重排序来辅助汽车仪表盘的设计。此外,张国方等^[8]使用 BERT 模型对汽车之家在线口碑进行了情感分析,挖掘出了用户的需求。

考虑了需求类别的用户需求挖掘方法是将各产品特征的情感值与 KANO 模型中的各需求类型进行匹配,分类别挖掘用户需求。白涛等^[9]提出了一种利用模糊 KANO 模型来进行用户需求分类和重要性统计的方法,并验证了方法有效性。Wu 等^[10]利用改进的模糊 KANO 问卷对用户需求进行了优先级排序。胡东方等^[11]提出了基于 KANO 的工程特征映射模型和基于人工免疫系统的产品设计模型,设计满足顾客需求的产品方案。徐海丽等^[12]利用文本挖掘,建立属性情感值与 KANO 模型的联系,获得用户的需求类别。

Shwetank 等^[13]将 KANO 模型和 QFD 方法整合,对用户需求进行了分类。

通过梳理可发现,大多数研究仍采用的是主观性较强的 KANO 问卷,亟需研究出一种能够自动整合用户生成内容和 KANO 模型进行需求分类的高效用户需求挖掘方法,为企业提供决策支持。

2 KEY-DEMANDS-OL 的具体步骤

在数据源方面,KEY-DEMANDS-OL 没有使用主观性较强的 KANO 问卷,而选择了在线品牌社区的海量用户评论;在初始改进率方面,KEY-DEMANDS-OL 在考虑情感极性的同时,增加了情感程度,丰富了需求分类语义,具体步骤如下所示。

2.1 在线品牌社区意见领袖的识别

2.1.1 识别指标体系构建

借鉴原欣伟等^[14]、Li Y Y 等^[15]、祝琳琳等^[16]的研究,经过对当前在线品牌社区特征的分析,从用户的行为特征和内容特征两个角度,KEY-DEMANDS-OL 最终确立了 5 项一级指标和 14 项二级指标,具体如表 1 所示。

2.1.2 意见领袖识别方法

(1)基于熵权法的指标权重确定。

假设被评价对象如用户共 M 个,其中每个被评价对象的评价指标共有 N 个,则指标数据集可以用矩阵 L 表示,其中 $L = (L_1, L_2, \dots, L_m, \dots, L_M)$,每个被评价对象可以表示为 $L_m = (l_{m_1}, l_{m_2}, \dots, l_{m_n}, \dots, l_{m_N})$ 。其中 $m = 1, 2, \dots, M, n = 1, 2, \dots, N$ 。

(a)对数据进行标准化处理,如式 1。

$$l'_{m_n} = \frac{l_{m_n} - \min_n l_{m_n}}{\max_n l_{m_n} - \min_n l_{m_n}} \quad (1)$$

其中, $\min_n l_{m_n}$ 表示 n 个指标值中最小的一个值, $\max_n l_{m_n}$ 表示 n 个指标中最大的一个值。

(b)分别对各指标下各个被评价主体指标值的占比 p_{m_n} 进行计算,如式 2 所示。

$$p_{m_n} = \frac{l'_{m_n}}{\sum_{m=1}^M l'_{m_n}} \quad (2)$$

利用式 3 和式 4 计算第 n 个指标的熵值 H_n 。 p 表示指标的信息熵系数。

$$p = \frac{1}{\ln(M)} \quad (3)$$

$$H_n = -p \sum_{m=1}^M \ln(p_{m_n}) \quad (4)$$

(c)计算得出第 n 个指标的熵权值,如式 5。

$$\omega(n) = \frac{1 - H_n}{N - \sum_{n=1}^N H_n} \quad (5)$$

其中, $\omega(n)$ 表示第 n 个指标的权重系数。 n 个指标的权重满足式 6。

$$\sum_{n=1}^N \omega(n) = 1 \tag{6}$$

表 1 意见领袖识别指标

指标来源	一级指标	二级指标	指标含义
内容特征	内容信息	详尽度	用户发帖中的中文字符数
		支持度	回帖中与用户发帖情感一致所占的比例
		可读性	$4.71 * (\text{总字符数}/\text{总字数}) + 0.50 * (\text{总字数}/\text{总句数})$
	内容表达	丰富度	用户发帖中的多媒体类型数量
		积分数	用户参与社区活动获得的积分数
		发帖数	用户发布帖子的数目
行为特征	参与度	回帖数	用户回复帖子的数目
		累计签到天数	用户在社区中签到的总天数
		人气值	社区对用户受欢迎程度的肯定
	号召力	威望值	社区对用户发帖内容质量的肯定
		浏览数	用户发帖的查看数目
		回复数	用户发帖的回复数目
关系建立		好友数	用户在社区中的好友数目
		评论指数	评论指数 = 评论数 / 浏览数

(2) 基于灰色关联分析的意见领袖影响力计算。

设 X_0 表示意见领袖代表的理想指标序列, 为参考序列, X_i 表示用户的实际指标数列:

(a) 对参考序列 X_0 和比较序列 X_i 进行确定。设原始数据集 X 如式 7 所示。

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \tag{7}$$

其中, m 为用户总数, n 为指标总数。参考序列为 $X_0 = (X_{01}, X_{02}, \dots, X_{0j}, \dots, X_{0n})$, 比较序列为 $X_i = (X_{i1}, X_{i2}, \dots, X_{ij}, \dots, X_{in})$, $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ 。

(b) 原始数据集进行无量纲标准化处理, 如式 1 所示。得到的标准化后的数据集 X' 如式 8 所示。

$$X' = \begin{bmatrix} X'_{11} & X'_{12} & \cdots & X'_{1n} \\ X'_{21} & X'_{22} & \cdots & X'_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ X'_{m1} & X'_{m2} & \cdots & X'_{mn} \end{bmatrix} \tag{8}$$

(c) 根据标准化后的指标集合和参考序列 X_0 进行绝对差值计算, 如式 9 所示。

$$\varphi = |x_{0j} - x'_{ij}| = \begin{bmatrix} |x_{01} - x'_{11}| & |x_{02} - x'_{12}| & \cdots & |x_{0n} - x'_{1n}| \\ |x_{01} - x'_{21}| & |x_{02} - x'_{22}| & \cdots & |x_{0n} - x'_{2n}| \\ \cdots & \cdots & \cdots & \cdots \\ |x_{01} - x'_{m1}| & |x_{02} - x'_{m2}| & \cdots & |x_{0n} - x'_{mn}| \end{bmatrix} \tag{9}$$

其中, X_{0j} 表示参考序列中第 j 个评价指标的指标值, X'_{ij} 表示无量纲标准化处理后的评价指标矩阵中第 i 个用户的第 j 个评价指标的指标值。

(d) 对灰色关联系数进行计算, 如式 10 所示。

$$\xi_{ij} = \frac{\min_i \min_j \varphi + \rho \max_i \max_j \varphi}{\varphi + \rho \max_i \max_j \varphi} \tag{10}$$

其中, $\min_j \varphi$ 表示在 $j = 1, 2, \dots, n$ 中两个序列之间的一级最小绝对差, $\min_i \min_j \varphi$ 表示在 $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ 中两个序列之间的二级最小绝对差, $\max_j \varphi$ 表示在 $j = 1, 2, \dots, n$ 中两个序列之间的一级最大绝对差, $\max_i \max_j \varphi$ 表示在 $i = 1, 2, \dots, m, j = 1, 2, \dots, n$ 中两个序列之间的二级最大绝对差。 ρ 为分辨系数, 主要用来控制灰色关联分析法的分辨率, 通常 ρ 取 $[0, 1]$ 。KEY-DEMANDS-OL 方法参考陈志霞等^[17]、杨柳等^[18]的研究, 设定 $\rho = 0.5$ 。

(e) 计算各个指标的加权灰色关联度。用到灰色关联度将分散的灰色关联系数 ξ_{ij} 综合到一个数值上以便比较。具体计算如式 11 所示。

$$Y_i = \sum_{j=1}^n \omega(j) \times \xi_{ij} \tag{11}$$

$$\sum_{j=1}^n \omega(j) = 1 \tag{12}$$

其中, $\omega(j)$ 表示第 j 个指标的权重, 指标的权重和满足式 12。

根据熵权法和灰色关联分析法计算每个用户的灰色关联度, 确定用户影响力, 识别出意见领袖。

2.2 用户关键需求挖掘

2.2.1 用户需求特征提取

评论中词语的词性和词语的频率决定了该词能否成为关键词。此外,名词和动词更能表达句子的核心含义,应作为核心关键词;副词表示程度,可用于情感程度的计算。因此,词频统计以及词性识别是根据用户评论挖掘用户需求的关键。

在识别在线品牌社区意见领袖的基础上,对其发表的评论进行词频的统计与词性的识别。保留其中的名词 n 、动词 v 以及副词 adv 并且统计其词频,代表该评论的核心关键,作为意见领袖代表的需求特征。经过处理的意见领袖的评论输入情感分析模块,进行基于贝叶斯的需求满足度计算。

2.2.2 基于贝叶斯的需求满足度计算

用户的评论实际上反映了用户对产品需求的满足

$$P_{r_j}^{f_k} = P(\text{pos} | \omega_1, \omega_2, \dots, \omega_n) = \frac{P(\omega_1, \omega_2, \dots, \omega_n | \text{pos}) \cdot P(\text{pos})}{P(\omega_1, \omega_2, \dots, \omega_n | \text{pos}) \cdot P(\text{pos}) + P(\omega_1, \omega_2, \dots, \omega_n | \text{neg}) \cdot P(\text{neg})}$$

根据式 13 考虑取值范围对概率 $P_{r_j}^{f_k}$ 进行判断即可判断出文本的情感极性 $S_{r_j}^{f_k}$ 。

$$S_{r_j}^{f_k} = \begin{cases} -1, & P_{r_j}^{f_k} \leq 0.4 \\ 1, & P_{r_j}^{f_k} \geq 0.6 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

其中, f_k 表示用户的第 k 个需求。 $F = \{f_1, f_2, \dots, f_k, \dots, f_K\}$, F 表示所有的用户需求, K 表示需求总个数。 r_{ij} 表示第 i 条评论中的第 j 条子句。 $P_{r_j}^{f_k}$ 表示第 i 条评论中的第 j 条子句中,用户对第 k 个需求 f_k 的情感极性概率值。当 $P_{r_j}^{f_k} \geq 0.6$ 时,表示该评论子句中用户对需求 f_k 的满足度是正向的;当 $P_{r_j}^{f_k} \leq 0.4$ 时,表示该评论子句中用户对需求 f_k 的满足度是负向的;否则该评论子句中用户对需求 f_k 无明显的满足倾向,为中性。

(2) 情感程度计算。

采用哈工大情感程度词表,依据副词表达的情感强烈程度,将情感程度划分成 5 个等级,详细内容如表 2 所示。在对意见领袖评论的情感程度进行计算时,依据其需求特征提取时保留的副词所对应的等级对满足度进行加权,详细计算过程如式 14 所示。

表 2 情感程度量化

程度副词	程度等级值
极其、百分之百、十分……	2.0
非常、很、分外、特别……	1.5
更为、较为、比较、还……	1.25
略微、稍微、有点、有些……	0.5
不大、半点、丝毫、不怎么……	0.25

程度,因此,用户的实际需求可以通过对用户评论的分析来挖掘,用户的需求满足度也可通过对评论的情感极性判断和情感程度的计算获得。详细计算过程如下:

(1) 基于贝叶斯的需求极性计算。

采用贝叶斯模型判断极性,具体的计算方法如下:

$T_{r_j}^{f_k}$ 表示第 i 条评论中第 j 条子句包含需求 f_k ,其经分词后的表示为 $T_{r_j}^{f_k} = \{\omega_1, \omega_2, \dots, \omega_n\}$ 。 $T_{r_j}^{f_k}$ 若为正向情感,则属于 pos 类,负向情感,则属于 neg 类。

根据贝叶斯定理可得 $T_{r_j}^{f_k}$ 的正向情感概率 $P_{r_j}^{f_k} = P(\text{pos} | \omega_1, \omega_2, \dots, \omega_n)$:

$$P_{r_j}^{f_k} = P(\text{pos} | \omega_1, \omega_2, \dots, \omega_n) = \frac{P(\omega_1, \omega_2, \dots, \omega_n | \text{pos}) \cdot P(\text{pos})}{P(\omega_1, \omega_2, \dots, \omega_n)}$$

根据全概率公式进一步化简上式可得:

$$SW_{r_j}^{f_k} = S_{r_j}^{f_k} \times W_{r_j}^{f_k} \quad (14)$$

其中, $W_{r_j}^{f_k}$ 表示第 i 条评论中的第 j 条子句中有关用户需求 f_k 的程度等级, $W_{r_j}^{f_k} = \{2.0, 1.5, 1.25, 0.5, 0.25\}$ 。 $SW_{r_j}^{f_k}$ 表示第 i 条评论中的第 j 条子句中有关用户需求 f_k 的满足程度,根据式 15 可得出各个需求的满足程度。

$$SW_{f_k}^* = \{SW_{f_k}^+, SW_{f_k}^-\} = \sum_{i,j} SW_{r_j}^{f_k} \quad (15)$$

其中, $SW_{f_k}^*$ 表示用户需求 f_k 的满足程度, $SW_{f_k}^+$ 表示正向极性下用户需求 f_k 的满足程度, $SW_{f_k}^-$ 表示负向极性下用户需求 f_k 的满足程度。

2.2.3 需求分类

KANO 模型将需求分为了期望型、魅力型、基本型、无关型和反向型五类。当前在实际的生产运营过程中通常只考虑用户的期望型需求、魅力型需求、基本型需求和无关型需求四种^[19-20]。

所提 KEY-DEMANDS-OL 方法在传统用户需求分类挖掘方法的基础上进行了改进,没有利用主观设计的调查问卷,而是利用在线品牌社区中海量的用户评论数据,识别代表性用户即意见领袖,利用算法自动整合意见领袖的评论和 KANO 模型。此外,修正了用户需求的初始改进率,增加考虑了情感程度,对用户关键需求进行自动化的分类挖掘,详细的计算过程如下:

首先,确定需求初始改进率 $IR_0^*(f_k)$,如式 16:

$$IR_0^*(f_k) = \frac{(m_{f_k}^+ + m_{f_k}^-) * \max(W_{r_j}^{f_k})}{SW_{f_k}^*} \quad (16)$$

许多相似文献等的研究中^[19-20],初始改进率 $IR_0(f_k)$ 由用户需求 f_k 中所有有效问卷数 n_{f_k} 和用户满

意见问卷数 $n_{f_i}^+$ 的比值确定,如式 17。

$$IR_0(f_k) = \frac{n_{f_i}}{n_{f_i}^+} \quad (17)$$

为了降低主观性,同时提高用户满意度测度的准确程度,修正了初始改进率,将 $IR_0(f_k)$ 改为 $IR_0^*(f_k)$, 作为新的初始改进率,如式(16)所示。 $IR_0^*(f_k)$ 由具有情感倾向的评论数与最大情感程度值的乘积和现有情感值的比值确定。 $m_{f_i}^+$ 表示正向情感倾向下含有用户需求 f_k 的评论数, $m_{f_i}^-$ 表示负向情感倾向下含有用户需求 f_k 的评论数, $\max(W_{r_e}^{f_i})$ 表示最大情感程度值,这里取 $2^{[19-20]}$ 。

其次,计算不同用户需求特征的平均不满意度 $\overline{X_{f_i}}$ 和平均满意度 $\overline{Y_{f_i}}$ 。计算过程如式 18 和式 19 所示。

$$\overline{X_{f_i}} = \frac{SW_{f_i}^-}{m_{f_i}^+ + m_{f_i}^-} \quad (18)$$

$$\overline{Y_{f_i}} = \frac{SW_{f_i}^+}{m_{f_i}^+ + m_{f_i}^-} \quad (19)$$

然后,根据不同用户需求特征的平均不满意度 $\overline{X_{f_i}}$ 和平均满意度 $\overline{Y_{f_i}}$,确定 KANO 满意度指数 α_{f_i} 和重要度指数 r_{f_i} ,详细计算公式如式 20 和式 21 所示。

$$\alpha_{f_i} = \tan^{-1} \frac{\overline{Y_{f_i}}}{\overline{X_{f_i}}} \quad (20)$$

$$r_{f_i} = |\vec{r}_{f_i}| = \sqrt{\overline{X_{f_i}}^2 + \overline{Y_{f_i}}^2} \quad (21)$$

其中, $0 \leq \overline{X_{f_i}} \leq 1, 0 \leq \overline{Y_{f_i}} \leq 1, 0 \leq \alpha_{f_i} \leq \frac{\pi}{2}$ 。此时,用户需求特征 f_k 就可以使用向量 $\vec{r}_{f_i} = (\alpha_{f_i}, r_{f_i})$ 来表示,以向量 \vec{r}_{f_i} 的模表示 KANO 重要度指数, α_{f_i} 表示

KANO 满意度指数。借助 KANO 满意度指数 α_{f_i} 和重要度指数 r_{f_i} 两个维度,可以将二维坐标图分割为四个象限,然后分别照应 KANO 需求模型的四种类别,并针对 KANO 需求模型的四种类别对产品功能属性提出具体的改进策略。

接着,对初始改进率 $IR_0^*(f_k)$ 进行近似变换,以消除用户满意度与产品特征的满足程度之间的线性关系,满足 KANO 模型的要求。改进公式如下:

$$IR_{adj}^*(f_k) = (IR_0^*(f_k))^{1/\tan\alpha_{f_i}} \quad (22)$$

最后,计算用户需求特征 f_k 的重要性 I_{f_i} ,详细的计算公式如下:

$$I_{f_i} = r_{f_i} * \frac{1}{IR_{adj}^*(f_k)} \quad (23)$$

至此,可按计算所得的用户需求特征的重要性排序 I_{f_i} 获得用户关键需求。

3 实验与结果分析

3.1 实验数据的获取与预处理

采用八爪鱼对华为在线品牌社区——花粉俱乐部中的华为 Mate40 板块进行数据采集。经过去除无效信息、去除重复信息等预处理后,最终得到包含 1 000 个用户的 22 572 条评论与讨论信息。KEY-DEMANDS-OL 算法采用 Python 实现。

3.2 KEY-DEMANDS-OL 结果分析

3.2.1 意见领袖识别结果

指标权重系数、关联系数和意见领袖识别结果如表 3~表 5 所示。实验中,参考文献[19-20],并依据帕累托法则,将影响力排名前 20% 的用户作为意见领袖。

表 3 指标权重系数

信息 详尽度	支持度	可读性	丰富度	积分数	发帖数	回帖数	累计签 到天数	人气值	威望值	浏览数	回复数	好友数	评论 指数
0.071 4	0.071 4	0.071 4	0.071 6	0.071 4	0.071 5	0.071 5	0.071 5	0.071 4	0.071 4	0.071 4	0.071 4	0.071 4	0.071 4

表 4 指标相关系数

用户名	信息 详尽度	支持度	可读性	丰富度	积分数	发帖数	回帖数	累计签 到天数	人气值	威望值	浏览数	回复数	好友数	评论 指数
时光如玉	0.997	0.942	0.996	0.333	0.999	0.988	1.000	0.977	0.994	1.000	0.998	1.000	0.990	0.926
一曲琵琶 倾城梦	0.965	0.998	0.968	0.796	0.970	0.972	0.895	0.545	0.935	0.992	0.984	0.984	0.872	0.953
荣耀小八 s	0.999	1.000	0.970	0.938	0.997	0.747	0.998	0.784	0.992	0.998	0.576	0.794	0.987	0.867
黑白之间 哀而不伤	0.548	0.997	0.995	0.972	0.993	0.974	0.995	0.824	0.986	0.995	0.999	0.999	0.395	0.932
花粉九袋	0.900	0.980	0.524	0.972	0.971	0.694	0.955	0.709	0.970	0.976	0.963	0.967	0.987	0.976
sincere99	0.982	0.979	0.946	0.847	0.934	0.980	0.985	0.600	0.930	0.937	0.941	0.847	0.987	0.626
小邓先森	0.361	1.000	0.971	0.719	0.961	0.978	0.975	0.924	0.956	0.965	0.950	0.765	0.990	0.995

续表 4

用户名	信息详尽度	支持度	可读性	丰富度	积分	发帖数	回帖数	累计签到天数	人气值	威望值	浏览数	回复数	好友数	评论指数
huafans 01277623342	0.697	0.994	0.962	0.972	0.999	0.999	0.999	0.922	0.999	0.999	0.987	0.962	0.333	0.635
瞬間流星	0.708	0.948	0.850	0.719	0.973	0.975	0.968	0.730	0.965	0.978	0.978	0.951	0.985	0.710
希小天	0.809	0.973	0.645	0.641	0.885	0.969	0.973	0.861	0.887	0.888	0.972	0.973	0.987	0.957

表 5 意见领袖识别结果 (Top10)

排名	用户名	影响力	排名	用户名	影响力
1	时光如玉	0.938 21	6	sincere99	0.894 38
2	一曲琵琶倾城梦	0.916 21	7	小邓先森	0.893 36
3	荣耀小八 s	0.903 33	8	huafans01277623342	0.889 88
4	黑白之间哀而不伤	0.900 41	9	瞬間流星	0.888 54
5	花粉九袋	0.896 09	10	希小天	0.887 52

3.2.2 用户关键需求分析

该步骤的数据来自于意见领袖的评论和讨论信息。经过词频统计、词性分析与词语相似度检验,结合在线品牌社区评论组织结构及分类得到的用户需求特

征如表 6 所示。

通过对情感极性以及基于贝叶斯的需求满足程度的计算,结合 KANO 模型确定的用户需求类别和重要性排列顺序如图 1 和表 7 所示。

表 6 用户需求特征

序号	用户需求特征	包含的特征词
J_1	系统	操作系统,智能,智慧,主题,EMUI,EMUI 11.0,应用
J_2	内存	RAM,随机存储,随机动态存储,ROM,机身容量
J_3	外观	款式,设计,外貌,比例,高级感,色彩,外表,外形,颜值
J_4	拍照	拍出来,照片,镜头,前置摄像头,后置摄像头,超广角
J_5	电池	续航,充电,充电速度,超级快充,无线快充,无线充电
J_6	屏幕	曲面屏,周冬雨屏,瀑布屏,刘海屏,全面屏,非环幕屏
J_7	价格	加价,物超所值,掉价,保值,涨价,差价,划算,性价比
J_8	安全	密码锁,支付,刷脸,指纹,指纹解锁,人脸识别,密码
J_9	网络通讯	通话,SIM 卡,下载,网络,无线网络,5G,WIFI,掉线
J_{10}	娱乐	王者,王者荣耀,吃鸡,电影,音量,外放,刷刷,和平精英
J_{11}	配件	充电器,数据线,卡针,卡套,耳机,手机壳,外壳,钢化膜

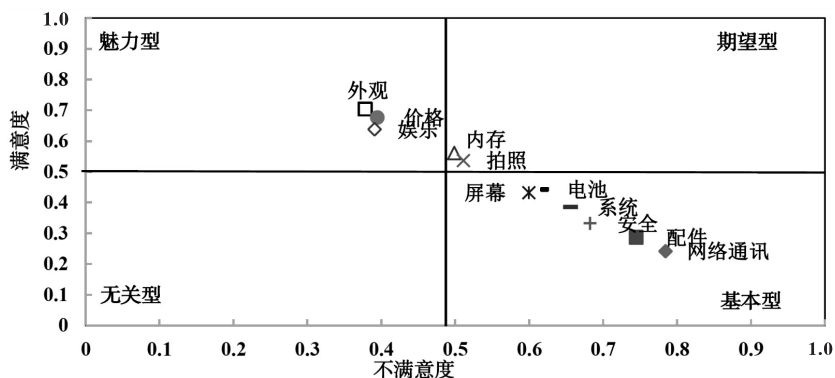


图 1 关键需求分类结果

从表 7 可以看出,分析得到的需求排序结果为:“外观”“价格”“娱乐”“内存”“拍照”“电池”“屏幕”“系统”“安全”“配件”“网络通讯”。结合 KANO 模型,从图 1 可知,在 11 个用户需求项中,魅力型需求一共有 3 个,包括“外观”“价格”“娱乐”;期望型需求 2 个,包括“拍照”以及“内存”;基本型需求一共有 6 个,

依次为“电池”“屏幕”“系统”“安全”“配件”和“网络通讯”。根据表 7 中的用户需求重要性排序及 KANO 模型可知,在进行产品改进时的先后顺序依次应为:“电池”“屏幕”“系统”“安全”“配件”“网络通讯”“拍照”“内存”“外观”“价格”“娱乐”。

表 7 用户关键需求排序结果

用户需求特征	\bar{X}_{f_k}	\bar{Y}_{f_k}	r_{f_k}	α_{f_k}	$\tan\alpha_{f_k}$	$IR_0^*(f_k)$	$IR_{adj}^*(f_k)$	I_{f_k}
外观	0.378 31	0.703 46	0.798 73	1.077 38	1.859 48	2.843 10	1.754 05	0.455 36
价格	0.394 77	0.676 02	0.782 85	1.042 25	1.712 44	2.958 49	1.884 03	0.415 52
娱乐	0.391 28	0.637 43	0.747 94	1.020 27	1.629 12	3.137 59	2.017 56	0.370 72
内存	0.499 16	0.560 34	0.750 43	0.843 07	1.122 55	3.569 29	3.106 39	0.241 58
拍照	0.511 19	0.536 40	0.740 97	0.809 46	1.049 32	3.728 57	3.504 91	0.211 41
电池	0.616 67	0.442 34	0.758 91	0.622 25	0.717 31	4.521 39	8.194 26	0.092 62
屏幕	0.599 76	0.431 67	0.738 95	0.623 85	0.719 74	4.633 16	8.417 05	0.087 79
系统	0.656 09	0.385 36	0.760 89	0.531 08	0.587 37	5.189 92	16.503 47	0.046 11
安全	0.682 53	0.332 18	0.759 07	0.452 95	0.486 69	6.020 83	39.989 93	0.018 98
配件	0.745 01	0.286 07	0.798 04	0.366 63	0.383 99	6.991 27	158.288 8	0.005 04
网络通讯	0.784 57	0.241 30	0.820 84	0.298 38	0.307 56	8.288 43	969.038 8	0.000 85

KEY-DEMANDS-OL 捕获的用户关键需求对企业优化产品具有重要的启示与指导。华为在推出 Mate40 时,提出了该款手机具有六大优势即非凡性能、非凡设计、非凡影像、非凡快充、非凡体验和非凡安全。针对此产品的基于在线品牌社区意见领袖的用户关键需求挖掘结果的启示及建议如下。

(1)魅力型需求包含了“外观”“价格”和“娱乐”。这说明用户对该款手机在这些方面的表现感到非常满意,符合 Mate40 手机预设。魅力型需求是进一步吸引用户,提升用户满意度和忠诚度的关键,企业在资源允许的条件下,应全力满足用户的魅力型需求,提高品牌竞争力。

(2)用户的期望型需求包含了“拍照”和“内存”。这说明了“拍照”尽管是该款手机的卖点之一,但很多用户在拍照时对相应的拍摄功能并不了解,没有拍摄出应有的出色效果,造成了很多用户的满意度较低,使得该特征成为了用户的期望型需求。此外,用户期待非凡的体验,需要手机大容量内存的支持,但当前产品内存提供仍显得不足,“内存”亦是期望型需求之一。期望型需求是用户需求的“痒处”,企业下一步急需对“拍照”这一功能加入新手指导模块或者在拍摄过程中给与更多快捷方式的支持。同时,相比竞品,进一步增大手机内存,促进其转化为魅力型需求,提升用户的满意度。

(3)用户的基本型需求包含了“电池”“屏幕”“系统”“安全”“配件”和“网络通讯”。这说明了对于该

款手机提出的非凡安全和非凡快充两大亮点所对应的“安全”和“电池”两个特征,并没有让用户满意,成为了基本型需求。其主要原因是该款手机的安全模式虽功能强大,但使用繁琐。电池虽支持快充,性能优异但充电时发热严重。此外,该款手机虽然采用了好看的瀑布屏设计、启用最新的 EMUI 系统、赠品较多、支持 5G 信号,但也存在着屏幕绿屏严重、系统更新慢、赠品领取扣费和 WIFI 信号不稳定的问题,使得“屏幕”“系统”“配件”和“网络通讯”四个特征成为了用户的基本型需求。基本型需求是用户需求的“痛点”,当需求被满足时,用户不会感到满意,但当不被满足时,用户会很不满意。因此,企业在追求极致功能的同时,也要时刻关注用户的基本型需求,针对其不满意点不断进行产品优化和改进,促使其向期望型需求进行转化。

3.3 对比分析

对比 KEY-DEMANDS-OL 和挖掘所有用户信息的传统方法的运行效率,记录下两种方法从载入数据到分析完成的时间,共 10 次,计算其平均值。因算法执行的绝对时间受计算机 CPU 及内存等性能参数影响,不同设备执行时间并不一致,为了更好地体现所提方法优势,同时给出了本次实验不同方法运行时间的相对比值以及绝对值,如图 2 所示。

为验证基于在线品牌社区意见领袖的用户关键需求挖掘方法(KEY-DEMANDS-OL)的有效性,与挖掘所有用户评论信息的传统方法进行了对比分析,如表 8 所示。

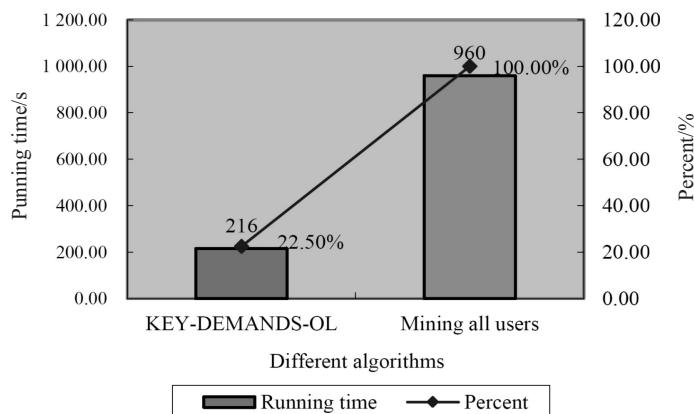


图 2 运行时间对比

表 8 与传统方法的对比分析

用户需求特征	意见领袖需求重要性排名(文中方法)	所有用户需求重要性排名(传统方法)
系统	8	8
内存	4	5
外观	1	1
拍照	5	4
电池	6	6
屏幕	7	7
价格	2	3
安全	9	9
网络通讯	11	11
娱乐	3	2
配件	10	10
...	-	...
首发处理器	-	48
多用户	-	49
奢华	-	50

从表 8 中可以看出,在基于所有用户评论挖掘的用户需求重要性排名中,仅有“内存”“拍照”“娱乐”和“价格”4 个需求与基于意见领袖挖掘的用户需求重要性排序有所差异,但相差不超过 1。而其余 7 个需求的重要性排名与基于意见领袖挖掘的需求重要性排名相同。实例分析的结果再次验证了帕累托法则,即约 20% 的意见领袖代表了约 80% 的用户关键需求,可通过意见领袖的识别,捕获用户关键需求。更为重要的是,KEY-DEMANDS-OL 方法在进行需求挖掘时所需处理的数据量大幅降低,提高了用户需求挖掘的时效性,且可以帮助企业从纷繁复杂,无代表性的海量用户需求中解脱出来,通过意见领袖的代表性,聚焦用户的关键需求,以精准优化产品。例如,在挖掘所有用户评论获取的前 50 大需求中,有诸如“首发处理器”“多用户”“奢华”等长尾小众需求。当然,小众需求并不是不重要,而是主流关键需求须首先满足。在这个前

提之下,再在后续产品或者特色产品中考虑小众需求。而 KEY-DEMANDS-OL 方法仅挖掘获取意见领袖代表的主流关键需求,是当下产品要尽快解决的问题以及须改进的地方,因此能更加聚焦,快速发现用户关键需求。

由实验结果可知,KEY-DEMANDS-OL 方法仅需分析约 20% 的意见领袖的信息,便可发现当下用户的主流需求,大幅提高了用户关键需求挖掘的执行效率,更加适合在当前产品迭代更新飞快、用户需求多变的情况下对用户关键需求进行快速追踪。

4 结束语

从意见领袖和用户关键需求着手,采用熵权法和灰色关联分析法对意见领袖进行了识别,进而采用文本挖掘、情感分析以及 KANO 模型对用户需求进行分类和排序,挖掘出代表性用户的关键需求。针对实际案例对提出的方法进行了验证,表明了所提 KEY-DEMANDS-OL 方法的实用性与有效性。该文为相关研究提供了一个用户关键需求快速挖掘的方法,对产品的研发与改进提供了重要的定量参考与决策支持。此外,随着在线品牌社区的更迭,在 KEY-DEMANDS-OL 方法的后续使用中,主体框架可保持不变,意见领袖识别指标体系可以进行调整,以适应不同的业务领域。诚然,KEY-DEMANDS-OL 方法取得了一定的进展,但仍存不足,如在进行海量用户评论处理时没有很好地利用分布式计算框架,评论处理效率有待进一步提升。此外,在进行评论分析时上下文关联语义考虑不足。笔者将在后续研究中持续关注,力求有新突破。

参考文献:

[1] 王雪,董庆兴,张斌.面向在线评论的用户需求分析框架与实证研究——基于 KANO 模型[J].情报理论与实践,2022,45(2):160-167.
 [2] ZOU F,QIAN Y,ZHANG Z,et al. A data mining approach for analyzing dynamic user needs on UGC platform[C]//

- IEEE international conference on industrial engineering and engineering management (IEEM). Piscataway: IEEE, 2021: 1067-1071.
- [3] WANG T, ZHOU M. Integrating rough set theory with customer satisfaction to construct a novel approach for mining product design rules[J]. *Journal of Intelligent and Fuzzy Systems*, 2021, 41(1): 331-353.
- [4] 李贺, 曹阳, 沈旺, 等. 基于LDA主题识别与Kano模型分析的用户需求研究[J]. *情报科学*, 2021, 39(8): 3-11.
- [5] 吴东胜, 王忠群, 蒋胜, 等. 基于评论的商品改进需求获取方法研究[J]. *长江大学学报: 自科版*, 2017, 14(5): 44-48.
- [6] 纪雪, 高琦, 李先飞, 等. 考虑产品属性层次性的评论挖掘及需求获取方法[J]. *计算机集成制造系统*, 2020, 26(3): 747-759.
- [7] 邓昭, 张璇. 基于产品语义分析的汽车仪表盘设计方法研究[J]. *机械设计*, 2020, 37(7): 132-137.
- [8] 张国方, 寇姣姣, 陈令华. 网络评论文本驱动的汽车设计规划方法[J]. *机械设计*, 2021, 38(2): 139-144.
- [9] 白涛, 李中凯. 基于模糊Kano模型的顾客需求重要度计算方法[J]. *中国机械工程*, 2012, 23(8): 975-980.
- [10] WU M, WANG L. A continuous fuzzy Kano's S model for customer requirements analysis in product development[J]. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 2012, 226(3): 535-546.
- [11] 胡东方, 李奕辰, 李彦兵. 基于卡诺和人工免疫系统的顾客需求产品设计[J]. *计算机集成制造系统*, 2018, 24(10): 164-174.
- [12] 徐海丽, 唐晓波, 谢力. 基于在线评论的用户需求挖掘模型研究[J]. *情报学报*, 2015, 34(10): 1088-1097.
- [13] SHWETANK A, ROHIT S, RASHMI R. Qfd and fuzzy Kano model based approach for classification of aesthetic attributes of suv car profile[J]. *Journal of Intelligent Manufacturing*, 2020, 31(2): 271-284.
- [14] 原欣伟, 杨少华, 王超超, 等. 基于用户特征抽取和随机森林分类的用户创新社区领先用户识别研究[J]. *数据分析与知识发现*, 2017(11): 62-74.
- [15] LI Y Y, MA S Q, ZHANG Y H, et al. An improved mix framework for opinion leader identification in online learning communities[J]. *Knowledge-Based Systems*, 2013, 43: 43-51.
- [16] 祝琳琳. 在线评论信息质量感知研究[D]. 长春: 吉林大学, 2020.
- [17] 陈志霞, 徐杰. 基于TOPSIS与灰色关联分析的城市幸福指数评价[J]. *统计与决策*, 2021, 37(9): 59-62.
- [18] 杨柳, 罗文倩, 邓春林, 等. 基于灰色关联分析的舆情分级与预警模型研究[J]. *情报科学*, 2020, 38(8): 28-34.
- [19] 刘志明, 刘鲁. 微博网络舆情中的意见领袖识别及分析[J]. *系统工程*, 2011, 29(6): 8-16.
- [20] 沈卓, 李艳. 基于PreLM-FT细粒度情感分析的餐饮业用户评论挖掘[J]. *数据分析与知识发现*, 2020, 40(4): 67-75.