

doi:10.19920/j.cnki.jmsc.2025.05.005

# 生鲜电商用户在线选择行为研究<sup>①</sup>

## ——考虑评论的影响效用

何 勇<sup>1</sup>, 陈 静<sup>1</sup>, 李姗姗<sup>2</sup>, 陈旭辉<sup>1</sup>

(1. 东南大学经济管理学院, 南京 210096; 2. 南京审计大学金融学院, 南京 211815)

**摘要:** 生鲜电商要实现精益经营, 需要及时准确地挖掘用户需求. 因此, 面对新零售和互联网大数据背景下的生鲜农产品线上消费场景, 本研究对在线异质用户的隐性需求及偏好进行了挖掘和研究. 首先通过自然语言处理技术挖掘评论中的用户需求特征, 其次结合用户属性及生鲜农产品属性(产地、重量、果径等), 考虑评论隐反馈信息对用户购买决策行为的影响, 构建了融合评论影响效用的 Mixed Logit 模型, 并对用户在线选择行为进行了分析. 最后, 本研究抓取了生鲜电商平台评论数据进行算例分析. 研究发现: 生鲜农产品属性变量(显性需求)不能完全充分地捕捉用户需求, 部分隐性需求隐含在其他用户的评论中, 影响用户的选择行为. 与不考虑评论影响效用的模型对比, 纳入评论隐反馈信息所建立的考虑评论影响社会效用的生鲜电商用户选择模型具有更好的解释性, 验证了本研究模型对于研究在线用户选择行为的实际价值与有效性. 为生鲜电商实现精准营销、经营策略调整、产品优化等决策提供思路.

**关键词:** Mixed Logit 模型; 在线评论; 生鲜电商; 用户选择行为

**中图分类号:** F272.3      **文献标识码:** A      **文章编号:** 1007-9807(2025)05-0070-14

## 0 引 言

随着互联网的发展和用户消费习惯的逐渐转变, 电商平台成为一个不可或缺的生鲜农产品购买渠道. 中国生鲜电商市场发展迅速, 拥有广阔的市场前景<sup>[1]</sup>. 从供给端看, 农产品生产的集约化和标准化、冷链技术的逐步提高, 均为生鲜电商平台发展提供了支撑, 如盒马鲜生、京东生鲜、叮咚买菜等头部商家, 与处在行业末端的小型商家并存<sup>[2]</sup>. 从需求端看, 用户在线购买生鲜农产品的消费习惯正在逐步养成, 生鲜电商用户数量和行业渗透率不断提升.

然而, 由于生鲜农产品存在易腐易变质的特性, 使得生鲜电商的可持续运营与精准营销存在许多亟待解决的问题. 例如, 低价不稳定品质策略

下的用户黏性过低、无法准确把握消费者需求引起的库存管理和高损耗问题等, 导致生鲜电商在与传统农产品竞争的局面下仍处于劣势, 并且更易受到复杂供应链结构的影响<sup>[3]</sup>. 在市场需求导向及数据驱动的大背景下, 用户需求直接影响着生鲜电商的产品、服务类型、运营绩效等. 如何准确挖掘用户需求及偏好, 是生鲜电商提升运营绩效的重要目标之一.

互联网环境下, 社交网络和移动终端技术允许用户在购买产品后发布有关产品信息(价格, 品质, 品牌, 包装等)或服务信息(物流, 售后, 服务等)的评论. 随后, 此类评论逐渐演变成用户生成内容的一种重要形式. 一方面, 用户评论是对产品口碑和服务水平的重要信息披露, 对产品的销量具有重要影响作用<sup>[4, 5]</sup>. 另一方面, 评论中隐含

① 收稿日期: 2020-12-29; 修订日期: 2023-07-21.

基金项目: 国家自然科学基金资助项目(72171047; 71771053; 72001113); 江苏自然科学基金资助项目(BK20201144); 江苏重点研发计划资助(现代农业)项目(BE2018385).

作者简介: 何 勇(1975—), 男, 安徽合肥人, 博士, 教授, 博士生导师. Email: hy@seu.edu.cn

着用户对评价对象的隐反馈信息,往往体现了用户的隐性需求.具体而言,评论中不仅包含了用户关于产品和服务的实际反馈,也隐含着用户的购买体验和品牌认知,成为潜在用户一种有效的购前信息获取渠道,辅助他们进行品质判断进而影响其购买决策.但由于1)生鲜农产品易损耗、易变质等特性,2)食品安全问题,3)生鲜农产品运输损耗导致的退换货等问题,生鲜电商平台的用户相对更加注重生鲜农产品售前信息的可查询性,而评论会对用户的购买决策产生影响,即用户在 $t-1$ 时刻的评论和打分会影响其他用户在 $t$ 时刻的购买决策,如果用户选择购买,同样会进行评论和打分,进而影响其他用户在 $t+1$ 时刻的决策.从生鲜电商平台运营商的角度看,其经过多年的运营积累了广泛的用户群体,拥有大量的产品订单及评论数据,但这些数据的价值尚未得到充分利用.此外在线用户的选择是离散的,现有文献通常基于离散选择模型来模拟用户购买决策,以研究用户在不同的产品/服务属性水平和价格条件下如何做出选择,从而预测其购买行为<sup>[6]</sup>.多项Logit模型是应用最广泛的离散选择模型之一,但在应用时的选择项必须符合独立不相关假设

(IIA).然而,在线购物环境下,用户偏好的异质性不容忽视,产品效用往往难以满足IIA假设.

综上,本研究在新零售和互联网背景下,从定量的角度探讨在线评论对用户选择行为的影响,创新性地将个体离散选择行为理论与自然语言处理(NLP)技术有机地结合起来,通过从评论中挖掘用户的隐反馈需求信息,研究用户对不同生鲜农产品属性和隐反馈需求特征的异质性偏好,丰富了传统离散选择模型的应用领域.考虑到评论中的隐反馈信息对用户购买决策的影响,本研究将在线评论纳入对效用的影响因子中,同时为了消除IIA假设和固定偏好特征带来的影响<sup>[7]</sup>,本研究使用Mixed Logit模型来模拟用户在线消费场景.因此,本研究可以回答将在线评论纳入效用影响后是否能够挖掘出评论透露的某些隐性需求,以及Mixed Logit模型在本研究建立的考虑评论影响效用的在线选择行为模型上是否具有较好的适应性两个主要的问题.

本研究对生鲜电商实现精准营销、个性化产品推荐、数据驱动运营等方面具备一定指导意义,体现了党的二十大关于推动现代服务业同现代农业深度融合的发展理念.研究问题与框架如图1所示.

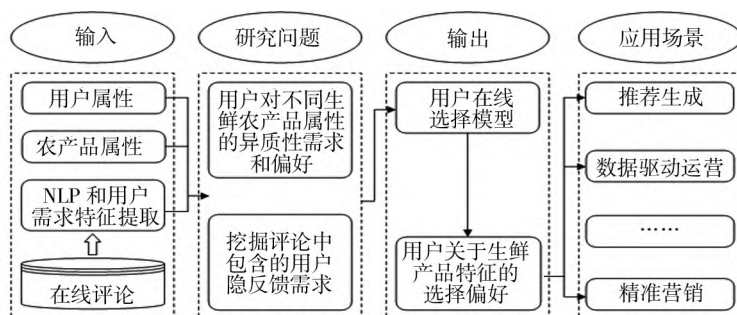


图1 研究问题及其应用场景

Fig. 1 Research questions and its application

## 1 文献述评

本章着重从评论特征抽取、离散选择模型及其应用两个方面对国内外研究现状进行综述分析,进而指出本研究的创新之处.

### 1.1 评论特征抽取

评论对商品的销量能产生很大影响,发表评论是用户线上消费行为中的一种,通过对评论内容的深入分析,企业可以准确地了解用户对各种

产品或服务的诉求.为了更深入地获取用户需求,近年来从商品评论文本中提取相关的特征属性成为了一项研究热点.Xie等指出大部分评论意见主要针对商品和服务具体的特征属性,但从评论中提取商品或服务的特征会受到文本过长、情绪判断模糊、词汇多样性以及有效信息匮乏等因素的影响<sup>[8]</sup>,并且常用的自然语言处理工具需要人工干涉算法中的参数,从而可能导致特征提取结果受到限制.针对此局限性,Henderson等构建了联合生成模型并应用增量Sigmoid置信网络(IS-

BN)架构训练其同步推导的联合生成模型,一定程度上克服了人工干涉参数的问题<sup>[9]</sup>,使算法所提取特征的客观性与随机性得到更高层次的保证.除以上两种方法外,Zhao等考虑了频率指标和六个网络指标研究了网络结构特征对共词网络(co-word network)排名主题的影响,并通过Web of Science收集了三个学科的文献数据,根据优先附件假设生成了40个模拟网络对所提出方法进行了验证<sup>[10]</sup>,能够更精准地挖掘出有效的隐性与关联特征.在提取特征后,面对不同属性的特征词则需要进一步对其进行情感分析.Yan等和王伟等学者使用“特征-观点对”的方法对特征观点划分了不同的重要程度,并基于PageRank算法进行隐式特征抽取<sup>[11,12]</sup>,但该方法的局限性在于需要对选定的商品进行持续跟踪,形成较长维度的时间序列数据,因此不适用于生命周期较短的商品.情感分析不仅能够对用户发表的评论进行情感值计算,还能够应用到其他多方面的研究中,例如Kim和Kang提出了一种用于在评论文本中提取和分析竞争商品间歧视性特征的方法,用以比较竞争性产品的优势和劣势<sup>[13]</sup>,而金晓玲等则从中心路径(信息质量)和边缘路径(来源可信度和情绪因素)出发研究了在线用户点赞与评论行为产生之间的差异,进一步丰富了在线信息互动行为的相关理论<sup>[14]</sup>.

综上可知商品评论的研究核心为评论数据的获取、特征词的提取和情感值的计算,即基于评论文本挖掘评价对象实体的特征属性,并对其进行情感分析.在语料粒度基础上对语句进行划分并加权计算情感极性,以及使用卷积神经网络和模糊数学与情感计算结合等诸多方法均可以进一步优化文本情感值的计算<sup>[15-17]</sup>.本研究使用Python对生鲜产品评论进行抓取,其次基于TF-IDF算法对关键词进行抽取,并采用Google word2vec工具对关键词进行同近义词聚类,最后使用情感词典进行评论情感值的计算.

## 1.2 离散选择模型及其应用

另一种分析用户需求的方式是从个体选择行为的角度进行建模.在用户购买决策中,离散选择模型可用于获知用户如何在不同产品/服务属性水平和价格条件下进行选择,从而来预测其购买行为.但是用户个人属性及异质性因素使其面对

不同商品特征时的偏好与选择性差异可能较为显著,此外异质性偏好在不同消费种类选择上也可能体现出较大的差异.例如徐璞等考虑常数-阶跃和常数-线性的出行效用偏好假设,建立了可以描述早高峰地铁通勤用户到达时刻离散选择行为的多项Logit、混合Logit和嵌套Logit模型,并讨论了用户通勤特征的异质性对出行效用偏好的影响<sup>[18]</sup>,而Duan等则针对铁路货运服务,分别建立了随机效用最大化、随机遗憾最小化及混合RUM-RPM模型,对铁路货运客户的选择偏好异质性和服务价值进行估计<sup>[19]</sup>.针对客户异质性偏好的进一步划分,杜慧滨等从异质行为视角出发,探讨了信息传播范围和传播渠道对新能源汽车技术扩散的影响机制,以及对异质性消费者购买决策的影响<sup>[20]</sup>,而Cirillo等则将用户的异质性偏好运用到机票购买情境中,提出了一种假设用户具有前瞻性视角的旅客退票和换票的跨时间选择模型,以预测退票和改签的时间<sup>[21]</sup>.郭鹏建立了考虑顾客策略行为的网络型非参数离散选择模型,采用EM算法对顾客到达率和概率质量函数进行联合估计,能够更准确地反映产品间网络替代效应对顾客选择行为影响<sup>[22]</sup>.可见异质性偏好在交通出行和在线消费等诸多方面均存在一定程度的影响,此外目前普遍采用基于Logit模型的数据研究方法来研究用户的异质性偏好,例如Yan等基于期望效用理论和离散选择模型,建立了多项Logit模型,从手机制造商的角度探讨了用户个人属性如何影响其对手机类型的偏好<sup>[23]</sup>.相比传统多项式Logit模型,Mixed Logit模型由于其可以设定模型系数服从随机分布,以捕捉个体间的异质性,并释放传统Logit模型的IIA假设,在近年来得到更广泛的应用<sup>[24]</sup>.唐立等基于Mixed Logit模型,对网约车模式的选择行为进行了研究,并分析了网约车费用的边际效应<sup>[25]</sup>.除在交通领域的应用外,Mixed Logit模型也常应用于以大规模数据为基础的个体决策行为研究中,因为该模型相对于其他Logit模型而言可以更好地适应个体偏好异质性和用户选择的非独立性<sup>[26]</sup>.

综上可知在现有研究中,大量研究分析如何从评论中提取信息,而在考虑评论对用户选择行为方面的影响则相对较匮乏.线上购买环境下,评论对用户的购买决策具有不可忽略的社会影响效

用,将离散选择模型运用到互联网大数据背景下,需考虑用户选择的独立性和用户偏好的异质性问题,这两类问题在目前文献中鲜有涉及.因此,本研究将个体选择行为理论与自然语言处理技术有机结合,将评论带来的社会影响效用纳入到离散选择模型中,从定量的角度探寻评论中包含的隐反馈信息对用户选择行为产生的影响,进一步挖掘生鲜电商平台用户的异质性需求偏好.

## 2 用户在线选择行为建模

为了深入剖析评论中隐含的用户需求以及引起的社会影响效用,本节建立研究模型来分析用户的在线选择行为.

### 2.1 随机效用刻画

效用理论是决策者选择行为的基础,同时也是需求理论的基础.随机效用理论最早由 McFadden 建立<sup>[27]</sup>,用户  $n$  可以在选择集  $C$  中选择  $i (i \in C)$ ,并获得效用  $U_{ni}$ .效用  $U_{ni}$  包含两部分:可观测效用  $V_{ni}$  和随机效用  $\varepsilon_{ni}$ ,  $V_{ni}$  为效用的可观测部分,通常包含用户属性和产品特征,  $\varepsilon_{ni}$  为不可观测效用,包含难以观测到的效用和观测误差带来的影响.

$$U_{ni} = V_{ni} + \varepsilon_{ni}, i \in C \quad (1)$$

决策者选择行为最基本的假设就是决策者追求效用的最大化.根据效用最大化理论,理性的用户会选择使效用最大的选项,用户  $n$  选择选项  $i$  的概率为

$$\begin{aligned} P_{ni} &= \text{Prob}(U_{ni} > U_{nj}, \forall i \neq j) \\ &= \text{Prob}(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \forall i \neq j) \quad (2) \\ &= \text{Prob}(V_{ni} - V_{nj} > \varepsilon_{nj} - \varepsilon_{ni}, \forall i \neq j) \end{aligned}$$

概率  $P_{ni}$  服从累积分布

$$P_{ni} = \int_{\varepsilon} I(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}, \forall j \neq i) f(\varepsilon_n) d\varepsilon_n \quad (3)$$

其中  $I(\cdot)$  表示指示性函数,当括号内的内容为真时为 1,否则为 0,如下式所示.

$$I = \begin{cases} 1, & \varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}, \forall j \neq i \\ 0, & \varepsilon_{nj} - \varepsilon_{ni} \geq V_{ni} - V_{nj}, \forall j \neq i \end{cases} \quad (4)$$

### 2.2 离散选择构建

离散选择模型作为一种用以描述决策者在备

选项有限的情形下的选择决策问题,被广泛地运用于交通方式选择、经济布局、用户购买决策、智能决策等经济决策领域的研究中<sup>[28]</sup>.在用户购买决策中,它可用于获知用户在不同产品特征属性水平和价格条件下如何进行选择,从而来预测其购买行为.其中, Mixed Logit 模型由于其摆脱了 IIA 假设和偏好固定的特性,具有更广泛的应用场景<sup>[29]</sup>.在 Mixed Logit 模型中,个体选择备选项所获得的效用能够被分解成可观测效用、随机效用及误差部分,可以表示为

$$U_{ni} = V_{ni} + \varepsilon_{ni} + \epsilon_{ni} \quad (5)$$

与多项 Logit 模型及嵌套 Logit 模型相比, Mixed Logit 模型的效用函数增加了误差部分,误差项  $\epsilon_{ni}$  允许选择项之间存在相关性(解决了 IIA 问题). Mixed Logit 模型概率函数为多项 Logit 模型在其

密度函数  $f\left(\frac{\beta}{\theta}\right)$  上的积分形式,可以表示为

$$P_{ni} = \int L_{ni}(\beta) f\left(\frac{\beta}{\theta}\right) d\beta \quad (6)$$

$$L_{ni}(\beta) = \frac{e^{V_{ni}}}{\sum_{j \in C} e^{V_{nj}}} \quad (7)$$

另外  $f\left(\frac{\beta}{\theta}\right)$  为参数  $\beta$  所服从分布的概率密度.

将式(7)代入式(6),得到 Mixed Logit 的概率为

$$P_{ni} = \int \frac{e^{V_{ni}}}{\sum_{j \in C} e^{V_{nj}}} f\left(\frac{\beta}{\theta}\right) d\beta \quad (8)$$

用户的选择偏好反映了在一定支付能力下的选择过程中所表现出来的偏好.选择偏好具有同质性和异质性之分.同质性偏好是指不同用户对于同一属性的偏好具有明显的相似性,进而可以使用统一的偏好来表示该类用户群体,异质性偏好则指不同用户对同一属性的选择偏好具有明显的差异性<sup>[30]</sup>.为了便于研究,许多研究假设不同用户的偏好是固定的,即将调查群体进行分类,再针对不同的群体进行针对性研究.然而电商平台的用户数量十分庞大,用户群体偏好无法再使用固定值来表示,因此偏好  $\beta$  不再假定为常数,而作为随机变量以刻画个体的异质性偏好.同时网络平台的用户生成内容,如评论信息,作为一种有效的信息获取渠道,能够辅助用户进行产品品质判断进而影响购买决策,故选择项之间具有很强的相关性,很难符合 IIA 假定.

效用最大化过程可用于模拟生鲜平台用户选择购买哪种生鲜农产品的决策过程. 用户在选择购买哪种生鲜农产品时, 一般会受到其个人偏好与经济属性、商品的特征属性以及其他不可观测的特征影响, 因此用户会从备选商品中, 选择一个自我感知效用最大的进行购买.

### 2.3 考虑评论影响效用的在线选择行为建模

假设用户在购买时遵循效用最大化的原则, 在  $N$  个备选的生鲜农产品中选择感知效用最大的进行购买. 在互联网环境中, 用户不再固定为某一特定群体, 具有异质性偏好, 因此参数作为随机变量以反应用户的异质性偏好. 由于线上购物的用户无法对生鲜农产品进行实地挑选, 因此在线上购买生鲜农产品的用户通常会依赖评论来获取更多有关产品的信息. 假设评论信息会对在线用户的购买决策产生社会影响效用, 具体来说, 已购买用户在  $t-1$  发表的评论和打分将会在  $t$  时刻影响其他用户的购买决策, 在  $t$  时刻购买的用户也有可能在  $t+1$  时刻发表评论, 进而在  $t+1$  时刻影响其他用户的购买决策, 具体过程如图 2 所示.

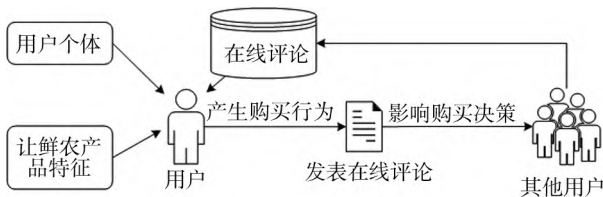


图 2 考虑评论影响效用的在线选择行为模式

Fig. 2 The selection behavior considering the impact of reviews on utility

因此在构建的生鲜电商用户在线选择行为模型中, 纳入评论影响共同构成了可观测效用函数的一部分, 如式(9)所示

$$V_{ni} = \sum_{0 < k < t} \lambda_{t-k} \sum_{n' \in N^{t-k}} \frac{1}{\eta} R_{t-k}^{in'} + \beta_n X_n + \beta_i X_i \quad (9)$$

上式的第一部分为评论信息引起的社会影响效用, 源于用户浏览商品评论时基于评论中包含的信息产生的判断.  $X_n$  表示用户的特征变量,  $X_i$  为产品的特征变量, 由于涉及用户隐私问题, 许多用户特征无法从电商平台中获取, 因此从平台上获取的可观测用户特征变量是有限的.  $\beta_n$  和  $\beta_i$  为待估参数, 可将  $\beta_n$  和  $\beta_i$  假设为服从期望为  $\mu_\beta$ , 标准差为  $\sigma_\beta$  的正态分布随机变量, 以刻画用户的异质性偏好, 其中  $\mu_\beta$  表示用户偏好期望(反映平均

偏好),  $\sigma_\beta$  表示用户偏好标准差(反映偏好异质性). 此外,  $R_{t-k}^{in'}$  表示评论的社会影响效用, 即  $t-k$  时刻用户  $n'$  对于产品  $i$  的评论对于待购买用户  $n$  的影响.  $\lambda_{t-k}$  代表  $t-k$  时刻的评论的时效性, 越接近当前时刻的评论  $\lambda$  值越大,  $\lambda_{t-k} \in [0, 1]$ .  $\eta$  为该商品的评论数.

为更贴合实际, 对  $R_{t-k}^{in'}$  中存在的不可观测误差做出如下说明: 1) 首先, 评论对用户的影响不仅来源于文本, 也来源于用户上传的图片和视频, 用户会根据评论中的实际图片判断描述是否与实物相符; 2) 本研究假设社会影响来源于线上评论, 但社会影响也很可能来源于线下的实际生活中, 例如朋友或家人的口头评价; 3) 此外, 自然语言处理技术具有一定局限性, 它无法识别评论文本中的一些特殊语气, 例如反讽等; 4) 用户  $n'$  的偏好异质性也会影响其所发布评论的内容, 称为社会影响异质性. 基于以上分析, 考虑到这些不可观测的误差, 为了使研究更具实际意义, 在  $R_{t-k}^{in'}$  中引入  $\gamma_w N(\mu_{\gamma_w}, \sigma_{\gamma_w})$ , 其中  $\mu_{\gamma_w}$  代表评论社会影响的期望,  $\sigma_{\gamma_w}$  反映评论社会影响在其均值周围的偏差程度, 也就是社会影响的不可观测部分, 具体表示为

$$R_{t-k}^{in'} = \gamma_w s_{t-k,w}^{n',i} \quad (10)$$

代入总效用函数(5)获得用户的总效用函数, 如式(11)所示

$$U_{ni} = \sum_{w \in S_i} \gamma_w \sum_{\substack{0 < k < t \\ n' \in N^{t-k}}} \lambda_{t-k} \frac{1}{\eta} s_{t-k,w}^{n',i} + \beta_i X_i + \beta_n X_n + \varepsilon_{ni} + \epsilon_{ni} \quad (11)$$

其中  $s_{t-k,w}^{n',i}$  表示用户  $n'$  在  $t-k$  时刻关于产品  $i$  评论中所提取的关键词  $w \in S_i$  的情感值,  $S_i$  为所提取的与购买产品  $i$  质量和体验服务相关关键词的集合, 关键词及其情感值将通过自然语言处理技术获得.

最后, 基于 Mixed Logit 模型的生鲜农产品选择概率函数如式(12)及式(13)所示, 可看作多项 Logit 模型的概率函数在参数的概率密度函数上的积分. 由于上述 Mixed Logit 模型的非封闭解不存在, 须通过计算机模拟实验过程来实现, 需要采用蒙特卡洛(Monte Carlo)模拟方法进行求解<sup>[31]</sup>.

$$P_{nit} = \iint \frac{e^{V_{nit}}}{\sum_{j \in C} e^{V_{njt}}} f\left(\frac{\beta}{\theta}\right) f\left(\frac{\gamma}{\theta}\right) d\beta d\gamma$$
$$\beta \sim N(\theta_\beta), \gamma_w \sim N(\theta_{\gamma_w}) \tag{12}$$
$$V_{nit} = \sum_{w \in S_i} \gamma_w \sum_{\substack{0 < k < t \\ n' \in N^{t-k}}} \lambda_{t-k} \frac{1}{\eta} s_{t-k,w}^{n'} + \beta_i X_i + \beta_n X_n \tag{13}$$

3 模型参数估计

本节基于京东生鲜电商平台的用户在线评论和生鲜农产品数据,建立并分析生鲜农产品与用户购买决策间的因果关系,探究各类特征属性对用户在线选择行为的影响,并对不考虑评论社会影响和考虑评论社会影响两种模型进行对比分析,探讨评论对发掘用户潜在需求的作用。

3.1 数据收集及变量选择

根据京东生鲜电商平台的数据特性,将因变量为备选的生鲜农产品(水果),自变量分为三类:1)用户的个人属性,包括是否为会员、评论日期、购买月份(生鲜产品可能受到季节因素影响)和是否大促活动期间下单(主要为年货节、38 女王节、618、双十一、双十二等);2)商品属性,包括品牌、价格、好评率、产地、果径与重量等;3)评论的用户需求属性,根据评论文本内容提取出关键词并量化其情感强度,即评论文本中提取的

关键特征指标,数据类别如表 1 所示。

表 1 数据类别与解释

Table 1 Categories and explanations of data

属性及类别		解释
用户属性	User ID	用户唯一标识
	VIP	是否会员
	Review Date	评论日期
	Promotion	是否大促活动期间下单
	Month	购买月份
生鲜农产品属性	Brand( dummy )	品牌名称
	Price( continuous )	价格
	Rating( continuous )	好评率
	Origin( dummy )	产地
	Fruit Trail( dummy )	果径
	Weight( continuous )	重量
	Marketing time	上市时间
	Others	……
评论用户需求属性	Key Attributes ( continuous )	评论关键词,通过情感强度量化,区间[ -1,1 ]

本研究以京东生鲜电商平台中售卖的苹果作为实际研究案例,收集了京东生鲜电商平台 2021 年四个季度约十八万条苹果商品的有效在线评论,涵盖了三十种不同品牌的商品,研究用户对其的选择行为.表 2 列举了生鲜农产品的特征属性,包括品牌、产地、果径、重量、评论数量和好评率.其中大多数品牌对苹果的果径进行了细分,剩余未细分果径的品牌根据其销售苹果的单品重量参考同重量其他品牌进行推算。

表 2 生鲜产品(苹果)的特征属性

Table 2 Attributes of the fresh products ( Apple )

编号	品牌	产地	重量/kg	果径/mm	价格/元	评论总数量	好评率
1	京鲜生红富士#19	山东	5.79	75 ~ 80	79.9	40 万 +	98 %
2	京鲜生红富士#02	山东	2.10	80 ~ 85	49.9	53 万 +	98 %
3	京鲜生红富士#64	山东	3.10	80 ~ 90	69.9	54 万 +	98 %
4	京鲜生红富士#26	山东	5.75	85 ~ 90	119.0	32 万 +	94 %
5	京鲜生红富士#26	山东	2.50	75 ~ 85	54.0	10 万 +	94 %
6	洛川苹果#67	陕西	3.50	80 ~ 85	50.8	3 万 +	91 %
7	洛川苹果#48	陕西	5.50	75 ~ 80	69.9	2 万 +	98 %
8	洛川苹果#22	陕西	4.50	85 ~ 90	78.9	10 万 +	95 %
9	洛川苹果#99	陕西	2.20	75 ~ 85	36.9	20 万 +	98 %
10	洛川苹果#09	陕西	4.50	70 ~ 75	64.8	2 万 +	98 %
11	青沃阿克苏苹果#98	新疆	4.50	80 ~ 85	39.9	10 万 +	90 %
12	青沃阿克苏苹果#05	新疆	4.50	80 ~ 90	49.9	5 万 +	90 %
13	青沃阿克苏苹果#04	新疆	4.50	75 ~ 85	38.8	5 万 +	90 %
14	京鲜生栖霞苹果#34	山东	3.50	80 ~ 85	69.9	30 万 +	98 %
15	栖霞苹果#70	山东	5.00	80 ~ 90	46.0	10 万 +	94 %

续表 2

Table 2 Continues

编号	品牌	产地	重量/kg	果径/mm	价格/元	评论总数量	好评率
16	栖霞苹果#79	山东	2.50	75~80	21.0	5 万+	94%
17	荣鼎京愿红富士#32	陕西	4.50	80~90	38.8	15 万+	93%
18	荣鼎京愿红富士#03	陕西	4.50	80~85	28.8	2 万+	93%
18	荣鼎京愿红富士#66	陕西	2.50	75~80	19.8	2 万+	93%
19	陇原红静宁苹果#85	甘肃	5.00	70~75	59.9	2 万+	97%
20	陇原红静宁苹果#84	甘肃	2.50	80~85	57.9	5 000+	96%
21	陇原红静宁苹果#89	甘肃	5.00	90~95	138.0	2 000+	97%
22	绿美红富士	甘肃	5.00	80~85	84.9	2 万+	96%
23	佳农红富士#35	山东	5.00	75~80	88.0	5 万+	96%
24	佳农红富士#49	山东	3.00	80~85	73.8	10 万+	96%
25	源外阿克苏苹果#94	新疆	2.50	75~80	29.9	1 万+	90%
26	源外阿克苏苹果#93	新疆	5.00	80~85	68.9	4 万+	92%
27	源外阿克苏苹果#74	新疆	5.00	80~85	75.9	1 万+	92%
28	朝道洛川红富士#55	陕西	4.50	75~80	54.8	1 万+	94%
29	朝道洛川红富士#56	陕西	4.50	80~85	59.8	4 万+	94%
30	金皓红富士	山西	1.50	75~80	12.0	5 万+	91%

### 3.2 生鲜农产品评论文本处理

#### 3.2.1 评论文本处理步骤

本节采用自然语言处理技术对评论文本进行处理,以获取各种生鲜农产品评论中词频最高的关键特征及其情感值.常用的关键词识别的方法主要包括 TF-IDF(term frequency-inverse document frequency)和 LDA(latent dirichlet allocation)等,其中 TF-IDF 更适合处理评论类的短文本.情感分析又称“观点挖掘”,分析的对象一般是带有个人情绪的主观性评论文本(称为情感文本).在线用户评论情感语义分析是对非结构化的网络评论短文本进行提取、分析、处理、归纳和推理,识别用户的观点、情感、态度和立场等重要信息.因此本研究采用 TF-IDF 方法进行关键词特征识别,并基于情感词典进行情感值计算.评论文本处理的具体步骤如下.

1)首先通过 Python 的 jieba 分词模块对非结构化文本进行预处理,包括分词与去除停用词,处理逻辑如图 3 所示.

2)其次基于 TF-IDF 算法进行关键词抽取及词频统计,再使用卡方统计的方法,选择排名前 200 的关键词,TF-IDF 公式为

$$TF_i = N(t_i, d) \quad (14)$$

$$IDF_i = \log \frac{N}{dt_i} \quad (15)$$

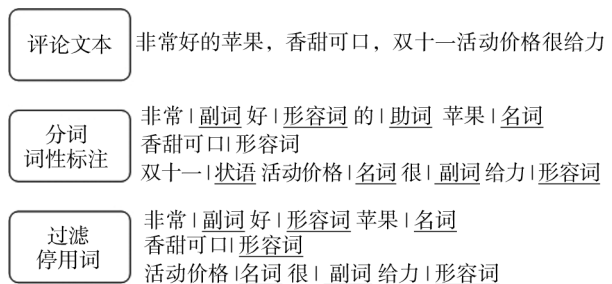


图3 评论文本分词与词性标注示意图

Fig. 3 Text segmentation and part-of-speech tagging

其中  $N$  是语料库中的文档总数,  $dt_i$  表示特征项  $t_i$  的 DF 值.

3)利用 Google word2vec 工具对关键词进行同义词聚类.

4)在提取的关键词中量化观点的强度,基于情感词典进行情感值计算.

#### 3.2.2 评论文本关键词提取结果

根据上节内容,本节使用 Python3.9 编写程序获取评论文本的关键词提取与聚类结果.表 3 列举了其中一种水果的好评文本分析结果,共统计为 14 类关键词.其中,第四列为根据同义词聚类生成的同义词/近义词.结果显示“质量”和“价

格”及其近义词的词频相对较高,在评论中被提及的次数较多,其次是“口感”、“速度”和“物流”。“价格”、“质量”和“速度”及其近义词具有

相对较高的情感强度,说明用户对该水果的价格、质量及物流速度评价较为正面,情感强度较高的关键词其次为“口感”、“服务”等等.

表 3 评论关键词词频统计及情感强度计算

Table 3 Comment keyword frequency statistics and emotional intensity calculation

关键词类别	总计词频	综合情感强度	同义词/近义词
质量	16 431	0.664 8	新鲜,优质,品质,高质
物流	8 525	0.600 3	快递,配送,送货,到货,运送,收货,上门
速度	10 458	0.643 4	立马,时间,快,直达,直送
包装	4 342	0.348 9	纸箱,破损,包装袋,礼盒,包裹,结实,外箱
尺寸	6 957	0.501 7	大,正常,饱满
价格	14 775	0.795 2	划算,实惠,半价,价钱,性价比,价位,促销,优惠券,便宜
功能	2 515	0.315 2	营养,榨汁,果汁,维生素
品相	2 006	0.437 7	颜色,粉,白,红,光滑,色泽
服务	3 379	0.546 9	态度,客服,体验
口感	13 651	0.614 3	汁水,好吃,味道,水分,甜度,脆,鲜甜,爽口
产地	4 963	0.296 7	洛川,陕西
售后	3 351	0.336 4	保障,保质期,保存,保鲜期,保证,可靠
品牌	3 978	0.254 8	信赖,京东,平台,正品
重量	1 142	0.250 7	分量,秤量,称重

3.3 参数估计结果

本节根据 2.3 节所建立的考虑评论影响效用的 Mixed Logit 模型和算例中的变量,对模型进行参数估计与统计量检验.

3.3.1 模型展开与描述

效用函数(13)在该算例中可展开为

$$V_{nit} = \left[ \sum_{w \in S_i} \gamma_w \sum_{0 < k < t} \lambda_{t-k} \sum_{n' \in N^{t-k}} \frac{1}{\eta} s_{t-k,w}^{n'i} \right] +$$
$$\beta_{n, \text{ vip }} I_{n, \text{ vip }} + \beta_{n, \text{ prom }} I_{n, \text{ prom }} +$$
$$\beta_{n, \text{ month }} X_{n, \text{ month }} + ASC_i + \beta_i X_i \quad (16)$$

其中,是否会员及是否大促活动期间下单为二分类变量可表示为  $I_{n, \text{ vip }} = \begin{cases} 1, \text{消费者 } n \text{ 是会员} \\ 0, \text{消费者 } n \text{ 不是会员} \end{cases}$ ,是否大促活动期间下单  $I_{n, \text{ prom }}$  同理.  $ASC_i$  为替代常数项(alternative specific constants),反映了各个商品的固有效用,  $X_i$  为商品的各项特征. 假设  $\eta = 50$ ,即用户在购买生鲜农产品时仅浏览前50 条商品评论,来自前 50 条评论的关键词的时效程度取值为 1,其他时效度取值为 0,则  $\gamma_w \sum_{\substack{0 < k < t \\ n' \in N^{t-k}}} \lambda_{t-k} \times$

$\frac{1}{\eta} s_{t-k,w}^{n'i}$  可看作是前 50 条评论的关键词  $w$  的平均情感值乘以一定的权重  $\gamma_w$ ,上式第一项即为该商品的评论社会影响效用.

3.3.2 参数标定结果

针对该部分内容使用 Stata 软件的 Asmilogit 模块来构建 Mixed Logit 模型并进行参数估计,根据表 1 所列举的变量,模型的参数估计结果如表 4 所示. 参数均值的大小和正负可以判别变量影响选择项的大小和方向,参数系数的标准差可以判断该参数系数在总体中是否存在变异,也就是说,个体对该参数的偏好是否存在较大的感知差异,即是否存在偏好异质性.  $p$  值用于检验变量的影响是否显著,当  $p$  值小于 0.05 时对效用具有统计学意义上的影响. 使用 McFadden 似然率指标检验模型的效果,McFadden 似然率指标介于 0 到 1 之间,值越大说明模型的解释力越强.

$$R_M^2 = 1 - \frac{LL(\theta)}{LL(0)}$$

(17)

表 4 为参数的标定结果,过滤了结果不显著的估计项,具体分析如下.



表 4 参数估计结果

Table 4 Estimation results of the parameters

类别	变量名	符号	含义	随机参数值	标准差	<i>p</i> value
用户属性	用户为会员	$\beta_{n, vip}$	京鲜生红富士#64	2.03	0.344	0.02
			京鲜生栖霞苹果#34	1.89	0.406	0.02
	大促活动期间下单	$\beta_{n, prom}$	京鲜生红富士#02	1.16	0.175	0.00
			京鲜生红富士#64	2.65	0.146	0.01
			洛川苹果#99	2.19	0.137	0.00
			京鲜生栖霞苹果#34	1.67	0.120	0.03
			荣鼎京愿红富士#32	0.75	0.084	0.03
			栖霞苹果#70	1.53	0.091	0.01
	购买月份	$\beta_{n, month}$	金皓红富士 × 九月	1.19	0.312	0.04
			绿美红富士 × 十一月	0.81	0.379	0.04
生鲜农产品属性	替代常数项	$\beta_{ASC}$	京鲜生红富士#64	7.41	0.874	0.03
			京鲜生红富士#02	4.33	0.603	0.02
			洛川苹果#22	3.85	1.210	0.03
			佳农红富士#49	-1.04	0.452	0.00
			陇原红静宁苹果#89	-1.79	0.606	0.00
			源外阿克苏苹果#94	-2.44	0.433	0.02
			金皓红富士	-3.57	0.941	0.01
	农产品属性	$\beta_{price}$	京鲜生红富士#26	-0.22	1.450	0.01
			洛川苹果#09	-0.36	0.831	0.00
			陇原红静宁苹果#89	-0.80	0.455	0.00
			绿美红富士	-0.73	2.014	0.01
			佳农红富士#49	-0.51	0.908	0.00
			价格 × 会员	-0.41	0.345	0.00
			价格 × 非会员	-0.55	0.274	0.00
		$\beta_{i, diam}$	果径	0.33	0.171	0.03
		$\beta_{i, weig}$	重量	2.45	0.894	0.01
	产地	$\beta_i(\mu)$	山东	5.76	0.701	0.00
			陕西	3.34	0.733	0.00
			新疆	2.58	0.425	0.00
		$\beta_i(\sigma)$	山东	4.13	1.458	0.00
评论关键词	评论情感值	$\gamma_w(\mu)$	口感	22.46	2.95	0.01
			质量	17.83	3.36	0.00
			速度	10.11	2.10	0.03
			价格	10.02	5.78	0.24
			服务	3.27	1.32	0.02
			包装	4.64	1.77	0.13
			好评率	0.51	0.12	0.01
		$\gamma_w(\sigma)$	价格	9.37	5.57	0.01
拟合优度指标	McFadden 似然率 $R_M^2$			0.274		

1) 用户属性. 首先, 在显著项中京东 PLUS 会员更倾向于购买京鲜生品牌的苹果, 且多为较大果径(80 mm ~ 90 mm), 因为京鲜生为京东自营

品牌, 因此可表明京东 PLUS 会员用户更偏向于在京东自营生鲜店铺进行购买. 此外, 从该商品的商品评论词频统计及情感性分析中可以发现其更

加突出地体现了“自营”、“保证”和“可靠”的特征,这说明会员用户对平台自营店铺品牌的依赖性与认可度相对更高。其次可以看出选择京鲜生、洛川苹果、荣鼎和栖霞四个品牌的用户在大促活动期间下单更为显著,从表2可知除京鲜生为平台自营外,其余三个品牌的对应单品累计评论均相对高于其他苹果,说明该三种单品的销量相对较高,且商家通常会选择销量较高的商品来参加平台的大促活动。最后金皓和绿美两个品牌在季节性因素上更加显著,可能由于是小众品牌缘故或产量问题等导致用户对其的选择呈现出更为显著的季节性差异。

2) 生鲜农产品属性。①替代常数项(ASC)反映了各商品的固有效用,ASC值最大的是京鲜生#64,其次是京鲜生#02,最低的是金皓红富士(-3.57),该品牌苹果的好评率为90%,且价格与重量均最低,评价综合显示其品质相对其他品牌而言较差。②京东自营品牌的苹果受价格影响程度要小于商家品牌,此外销量高的苹果受价格影响也相对较小。为了分析价格在不同用户属性下是否有较大的差异,特别是考虑价格和用户是否为会员情况的交叉效应,虽然非会员的价格参数系数值更小,但相比会员的价格系数不够显著。其他商品特征例如上市时间,对用户的在线购买没有影响,分析原因发现该类信息的披露位置在网页或移动端页面中不明显,而用户对于生鲜类产品的这类信息关注度也相对较低。③果径和重量的参数估计系数为正,说明果径和苹果重量对总效用具有正向的影响,即消费者除考虑价格因素外,普遍更偏好尺寸和重量大的苹果。④由于选择项中水果的产地来源只有五种且较为集中,结果显示山东、陕西和新疆三个产地的水果较显著,产地为山东的标准差显著且参数值较大(4.13),表明山东生产的苹果的偏好需求分布较分散,也可能由于京东生鲜自营的苹果产地来自于山东烟台,而自营又占据较高的比重所以使用户偏好需求较分散。

3) 评论。表4的第三部分为评论文本中挖掘的特征情感值的估计情况。可以看出,尽管某些用

户需求特征,如“速度”、“服务”、“口感”等,无法直接通过系数 $\beta$ 进行估计,但通过评论关键词情感强度进行商品效用的估计结果是显著的,说明从评论中挖掘用户隐含的需求具有可行性。随机参数估计的最大值是“口感”,它与水果的甜度、新鲜度、水分等性质相关。其次是“质量”,反映了生鲜农产品本身的质量、新鲜度及配送到用户手中的完好程度。“速度”、“包装”、“服务”这些属性与平台的服务质量相关,反映了物流配送、售后等服务属性在生鲜农产品网购中的关注度。“好评率”的影响显著且系数为正,说明线上渠道的购买决策会受到其他用户好评率的正面影响。此外,尽管评论中关于“价格”这一属性的 $p$ 值很大(无显著影响),但标准差显著,反映了社会影响的异质性,以及可能存在的不可观测部分,其可能是由于自然语言处理无法捕捉到的反讽语气等所导致。以上关键词情感强度的显著性估计结果反映了用户的需求不能完全被生鲜农产品属性变量(显性需求,直接影响用户行为)捕获,部分隐性需求隐含在其他用户的评论中,这些评论可能影响用户对产品的选择行为。

4) 拟合优度指标。模型的拟合优度指标 McFadden's Pseudo  $R^2$  为0.274 大于0.2,说明该模型具有相对较好的拟合优度和解释力。

### 3.4 模型对比

为了证明评论发掘用户的潜在需求的作用,本节将考虑评论影响效用和不考虑评论影响效用的模型进行对比。

对比模型为不考虑评论影响效用的可观测效用函数,如下式所示

$$V_{int} = \beta_{n, vip} I_{n, vip} + \beta_{n, prom} I_{n, prom} + \beta_{n, month} X_{n, month} + ASC_i + \beta_i X_i \quad (18)$$

利用式(19)中的公式计算 $\rho^{-2}$ 值,作为模型性能的评估指标。其中, $LL(\theta)$ 为包含所有变量的对数似然率, $LL(0)$ 为参数值都为0时的对数似然率, $N$ 表示估计的参数个数。

$$\rho^{-2} = 1 - (LL(\theta) - N)/LL(0) \quad (19)$$

相比于 McFadden 似然率指标,该指标考虑

到模型的参数个数,可排除参数个数对模型拟合效果的影响,可用于对比不同参数个数的模型效果.一般情况下, $\rho^{-2}$  越大则模型的解释性越强,能更好地解释用户关于产品的行为和偏好,结果如表 5 所示.

根据模型对比结果,模型 2 所提出的考虑评论影响效用的在线选择模型具有更高的  $\rho^{-2}$  值,说明其具有更好的拟合效果,证明纳入在线评论中的隐反馈信息对研究在线用户的选择行为具有实际价值.

表 5 两模型对比结果

Table 5 Comparison results of two models

模型对比	$LL(0)$	$LL(\theta)$	$N$	$\rho^{-2}$
模型 1	-3 952.6	-3 492.0	10	0.114
模型 2	-3 952.6	-2 889.1	24	0.263

注:模型 1——不含评论信息;模型 2——考虑评论信息.

## 4 结束语

本研究面向生鲜农产品线上消费场景,研究了生鲜电商平台用户的购买决策.首先,利用自然语言处理技术挖掘用户评论中的生鲜农产品特征属性,并量化用户对产品特征属性的情感强度.随后,构建融合评论隐反馈信息的 Mixed Logit 模型,深入挖掘大规模用户的异质性偏好,分析用户评论信息的社会影响效用和潜在用户需求.研究丰富了个体选择理论和离散选择模型的应用领域,对实现生鲜平台数据驱动下的精准营销具有重要意义.

本研究具有下列重要发现.第一,评论文本中的某些特征指标的标准差显著,反映该参数在均值周围的异质性,进一步证实了社会影响无法全部通过在线评论数据进行捕捉,存在社会影响的不可观测部分,其可能是由于自然语言处理无法捕捉到的反讽语气等所导致.第二,研究发现不同品牌生鲜产品本身的固有效用不同,可通过这一结论更加系统地评价不同生鲜农产品对用户的效用,进而在实际运营时及时更新不同效用值的产品广告投放量和推广度,例如得益于相对较高的

会员黏性,平台可减少对自营商品的部分广告投入.第三,商家在面对大促活动时需要衡量售卖商品与平台自营商品的竞争情况,因为大促期间自营商品活动力度较大,普遍销量较高,小众品牌商家可考虑在大促期间控制曝光与推广成本,而在非大促期间可适当增加推广力度,避免成本与收益的失衡.第四,纳入在线评论中的隐反馈信息所建立的考虑评论社会影响效用的生鲜平台用户选择模型具有较好的解释性,相比不考虑评论信息影响效用的模型而言具有更高的拟合优度,对于研究在线用户的选择行为更具有实际价值,更适用于生鲜电商的线上消费场景.

通过上文模型与案例分析,具体能够得出以下管理启示.首先,生鲜电商平台与商家可以利用该方法分析在线用户偏好,更系统地评估各种生鲜农产品对用户的效用,自营商品可以利用会员黏性的优势适当减少广告与曝光率的投入,此外小众品牌则建议控制大促活动期间的推广力度,以避免与平台自营商品大促期间的高力度补贴相冲突,导致成本投入与收益的失衡.其次用户的会员属性对价格的敏感性有差异但不明显,从评论相关的文本情感分析中发现产品质量和物流速度是苹果消费者着重考虑的因素,而产品价格可能存在较强难以观测的因素使其参数值并不显著但又有相对更显著的标准差,这可能是由于自然语言处理无法捕捉到的反讽语气等所导致,可能也是由于苹果的四季性、普遍性与相对较低的价格所导致的.而针对某些具有较强时效性、季节性与较高价值的生鲜产品,价格因素可能需要更加严格的衡量与控制.最后,本研究通过模型与案例研究证明了用户的需求不能完全被生鲜产品属性变量所捕获,通过考虑评论影响效用能够为在线用户选择提供更加实际的价值参考.

研究不足与展望:1)由于平台对用户隐私的保密导致用户的个人属性种类较少,未来可结合问卷、实验对用户样本进行扩充;2)本研究主要考虑了评论信息的社会影响作用,模拟了生鲜电商用户的在线购买行为,以发掘用户的潜在需求和异质性偏好,未来的研究可进一步扩展到个性化产品推荐等决策问题上.

## 参 考 文 献:

- [1] Yang C, Feng Y, Whinston A. Dynamic pricing and information disclosure for fresh produce: An artificial intelligence approach[J]. *Production and Operations Management*, 2022, 31(1): 155 – 171.
- [2] 邵腾伟, 吕秀梅. 生鲜农产品电商分布式业务流程再造[J]. *系统工程理论与实践*, 2016, 36(7): 1753 – 1759.  
Shao Tengwei, Lü Xiumei. Process reengineering on e-commerce of fresh food by distributing[J]. *Systems Engineering: Theory & Practice*, 2016, 36(7): 1753 – 1759. (in Chinese)
- [3] 范体军, 郑 琪, 蔡 路. 考虑权力结构及供应商竞争的生鲜供应链决策[J]. *管理科学学报*, 2022, 25(1): 23 – 38.  
Fan Tijun, Zheng Qi, Cai Lu. Supply chain decisions for fresh products with competitive suppliers under different power structures[J]. *Journal of Management Sciences in China*, 2022, 25(1): 23 – 38. (in Chinese)
- [4] Yang C, Zhang H, Jiang B, et al. Aspect-based sentiment analysis with alternating coattention networks[J]. *Information Processing & Management*, 2019, 56(3): 463 – 478.
- [5] Ren G, Hong T. Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews[J]. *Information Processing & Management*, 2019, 56(4): 1425 – 1438.
- [6] Hackbarth A, Madlener R. Consumer preferences for alternative fuel vehicles: A discrete choice analysis[J]. *Transportation Research Part D: Transport and Environment*, 2013, 25: 5 – 17.
- [7] Fiebig D G, Keane M P, Louviere J, et al. The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity[J]. *Marketing Science*, 2010, 29(3): 393 – 421.
- [8] Xie J, Liu X, Zeng D D, et al. Understanding medication nonadherence from social media: A sentiment-enriched deep learning approach[J]. *MIS Quarterly*, 2022, 46(1): 341 – 372.
- [9] Henderson J, Merlo P, Titov I, et al. Multilingual joint parsing of syntactic and semantic dependencies with a latent variable model[J]. *Computational Linguistics*, 2013, 39(4): 949 – 998.
- [10] Zhao W, Mao J, Lu K. Ranking themes on co-word networks: Exploring the relationships among different metrics[J]. *Information Processing & Management*, 2018, 54(2): 203 – 218.
- [11] Yan Z, Xing M, Zhang D, et al. Exprs: An extended pagerank method for product feature extraction from online consumer reviews[J]. *Information & Management*, 2015, 52(7): 850 – 858.
- [12] 王 伟, 王洪伟. 特征观点对购买意愿的影响: 在线评论的情感分析方法[J]. *系统工程理论与实践*, 2016, 36(1): 63 – 76.  
Wang Wei, Wang Hongwei. The influence of aspect-based opinions on user's purchase intention using sentiment analysis of online reviews[J]. *Systems Engineering: Theory & Practice*, 2016, 36(1): 63 – 76. (in Chinese)
- [13] Kim S G, Kang J. Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews[J]. *Information Processing & Management*, 2018, 54(6): 938 – 957.
- [14] 金晓玲, 周中允, 尹梦杰, 等. 在线用户点赞与评论行为的产生机理差异研究——以医疗健康类企业微信公众号为例[J]. *管理科学学报*, 2021, 24(4): 54 – 68.  
Jin Xiaoling, Zhou Zhongyun, Yin Mengjie, et al. Understanding antecedent differences across online users' like and comment behaviors: The case of healthcare enterprise WeChat public platform[J]. *Journal of Management Sciences in China*, 2021, 24(4): 54 – 68. (in Chinese)
- [15] 王洪伟, 郑丽娟, 尹 裴, 等. 基于句子级情感的中文网络评论的情感极性分类[J]. *管理科学学报*, 2013, 16(9): 64 – 74.  
Wang Hongwei, Zheng Lijuan, Yin Pei, et al. Classification of sentimental polarity for Chinese online reviews based on

- sentence level sentiment[J]. *Journal of Management Sciences in China*, 2013, 16(9): 64–74. (in Chinese)
- [16] Hyun D, Park C, Yang M, et al. Target-aware convolutional neural network for target-level sentiment analysis[J]. *Information Sciences*, 2019, 491: 166–178.
- [17] Phan H T, Nguyen N T, Hwang D. Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis[J]. *Information Sciences*, 2022, 589: 416–439.
- [18] 徐 璞, 刘天亮, 四兵锋. 基于智能卡数据的早高峰地铁通勤用户出行偏好建模与实证[J]. *系统工程理论与实践*, 2023, 43(5): 1484–1498.
- Xu Pu, Liu Tianliang, Si Bingfeng. Modeling and empirical study on travel preference of morning rail transit commuters using smart card data[J]. *Systems Engineering: Theory & Practice*, 2023, 43(5): 1484–1498. (in Chinese)
- [19] Duan L, Rezaei J, Tavasszy L, et al. Heterogeneous valuation of quality dimensions of railway freight service by Chinese shippers: Choice-based conjoint analysis[J]. *Transportation Research Record*, 2016, 2546(1): 9–16.
- [20] 杜慧滨, 邹宏阳, 张永杰, 等. 异质行为下新能源汽车的技术采纳与扩散[J]. *管理科学学报*, 2021, 24(12): 62–76.
- Du Huibin, Zou Hongyang, Zhang Yongjie, et al. Technology adoption and diffusion of new energy vehicle (NEV) under heterogeneous behaviors[J]. *Journal of Management Sciences in China*, 2021, 24(12): 62–76. (in Chinese)
- [21] Cirillo C, Bastin F, Hetrakul P. Dynamic discrete choice model for railway ticket cancellation and exchange decisions[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2018, 110: 137–146.
- [22] 郭 鹏. 航线网络中需求无约束估计的非参数选择模型[J]. *管理科学学报*, 2020, 23(12): 30–51.
- Guo Peng. A nonparametric discrete choice model for unconstrained demand estimation in airline network[J]. *Journal of Management Sciences in China*, 2020, 23(12): 30–51. (in Chinese)
- [23] Yan Y, Sun X, Liu Z. Factors affecting decision on mobile phone types for consumer purchase based on discrete choice model[M]. Singapore: Springer Singapore, 2019: 638–650.
- [24] 李 磊, 张曙阳, 李 彤. 基于嵌套 Logit 模型的竞争性选址问题研究[J]. *系统工程学报*, 2021, 36(4): 524–538.
- Li Lei, Zhang Shuyang, Li Tong. A new competitive facility location model based on nested Logit model[J]. *Journal of Systems Engineering*, 2021, 36(4): 524–538. (in Chinese)
- [25] 唐 立, 邹 彤, 罗 霞, 等. 基于混合 Logit 模型的网约车选择行为研究[J]. *交通运输系统工程与信息*, 2018, 18(1): 108–114.
- Tang Li, Zou Tong, Luo Xia, et al. Choice behavior of taxi-hailing based on Mixed-Logit model[J]. *Journal of Transportation Systems Engineering and Information Technology*, 2018, 18(1): 108–114. (in Chinese)
- [26] Li H, Webster S, Mason N, et al. Product-line pricing under discrete mixed multinomial logit demand[J]. *Manufacturing & Service Operations Management*, 2019, 21(1): 14–28.
- [27] Mcfadden D. Conditional Logit Analysis of Qualitative Choice Behavior[M]. New York: Academic Press, 1973.
- [28] Flores A, Berbeglia G, Van Hentenryck P. Assortment optimization under the sequential multinomial logit model[J]. *European Journal of Operational Research*, 2019, 273(3): 1052–1064.
- [29] Sen A, Atamturk A, Kaminsky P. Technical note: A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model[J]. *Operations Research*, 2018, 66(4): 994–1003.
- [30] 胡望斌, 张玉利, 杨 俊. 同质性还是异质性: 创业导向对技术创业团队与新企业绩效关系的调节作用研究[J]. *管理世界*, 2014, (6): 92–109.
- Hu Wangbin, Zhang Yuli, Yang Jun. Is homogeneity or heterogeneity? A study on the adjustment function of the relationship between the technical entrepreneurial team and new ventures' performance[J]. *Management World*, 2014, (6): 92–109. (in Chinese)

- [31] Munger D L, Ecuyer P, Bastin F, et al. Estimation of the Mixed Logit likelihood function by randomized quasi-monte carlo [J]. Transportation Research Part B: Methodological, 2012, 46(2): 305–320.

## Online choice behavior for fresh e-commerce customers: Considering the social influence of reviews

HE Yong<sup>1</sup>, CHEN Jing<sup>1</sup>, LI Shan-shan<sup>2</sup>, CHEN Xu-hui<sup>1</sup>

1. School of Economics and Management, Southeast University, Nanjing 210096, China;

2. School of Finance, Nanjing Audit University, Nanjing 211815, China

**Abstract:** To achieve a lean operation, fresh e-commerce enterprises need to grasp users' demands timely and accurately. This research aims to find consumers' heterogeneous preferences and implicit demands from reviews of fresh products on e-commerce platforms. Considering the social influence of implicit feedback information from reviews on customers' purchasing decisions, this study designs a Mixed Logit model incorporating the impact of reviews on customers' utility to simulate the different purchase behaviors and capture customers' implicit requirements and heterogeneity effectively. Firstly, this research collects the online reviews of fresh products and then adopts natural language processing (NLP) technology to extract the customers' implicit requirements from the reviews. Secondly, the implicit requirements are combined with customer attributes and fresh product attributes to construct the utility function and establish the customers' choice behavior model. Finally, an example analysis is conducted based on the actual data from a fresh product e-commerce platform. The results show that the explicit attributes of fresh products cannot fully capture customers' demands, and more invisible demands are hidden in other customers' comments, which would affect the purchase decisions of potential customers. By incorporating the implicit feedback information in online reviews, the discrete choice model which considers the social impact of reviews is well interpreted and suitable for purchasing scenarios of fresh product e-commerce platform. This model shows a higher degree of goodness of fit compared with the model without considering the social influence of reviews. Thus, the implicit feedback in online reviews has been demonstrated to present more practical value for influencing customers' choice behavior. The proposed method could provide valuable insights for fresh product e-commerce enterprises on precision marketing, business strategy adjustment, and product optimization.

**Key words:** Mixed Logit model; online reviews; fresh e-commerce; customers' choice behavior