

一种基于在线评论确定产品属性权重的模型

徐子鸣, 钟 正

江南大学商学院, 江苏 无锡

收稿日期: 2023年3月23日; 录用日期: 2023年4月20日; 发布日期: 2023年4月27日

摘 要

产品属性权重表示消费者对属性的重视程度, 是消费者选择产品和商家进行产品优化的依据, 但目前基于在线评论确定属性权重的方法存在主观性较大和结果易偏差的弊端, 因此文章提出了一种融合模型, 首先基于改进LDA主题模型挖掘出产品属性, 然后基于属性情感分析对评论中属性满意度进行标注。之后先对满意度标注样本进行N折交叉处理, 然后计算所有不同样本下各属性的信息增益值, 以此对各属性重要性值分布进行参数估计。然后, 计算两两分布的距离因子, 最后运用改进AHP方法确定属性权重。以肉类生鲜产品为对象进行实证分析, 结果表明, 所提出的融合模型能够准确计算出产品各属性权重。

关键词

评论文本, 属性权重, 融合模型, 文本挖掘

A Model for Determining Product Attribute Weight Based on Online Review

Ziming Xu, Zheng Zhong

Business School, Jiangnan University, Wuxi Jiangsu

Received: Mar. 23rd, 2023; accepted: Apr. 20th, 2023; published: Apr. 27th, 2023

Abstract

Product attribute weight indicates the importance consumers attach to attributes, and is the basis for consumers to select products and businesses to optimize products. However, the current method of determining attribute weight based on online reviews has the disadvantages of subjectivity and easy deviation of results. Therefore, this paper proposes a fusion model. First, the product attributes are mined based on the improved LDA topic model, and then the attribute satisfaction in the reviews is marked based on attribute emotion analysis. After that, the satisfaction labeling samples are processed with N-fold crossover, and then the information gain value of each attribute

under all different samples is calculated, so as to estimate the distribution of the importance value of each attribute. Then, the distance factor of pairwise distribution is calculated, and the attribute weight is determined by using the improved AHP method. The empirical analysis of fresh meat products shows that the proposed fusion model can accurately calculate the weight of each product attribute.

Keywords

Comment Text, Attribute Weight, Fusion Model, Text Mining

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

产品属性权重是衡量消费者对产品属性重视程度的量, 产品属性权重越高则说明消费者对该属性的重视程度越高。而现实生活中, 消费者对同类产品不同属性的重视程度往往是不同的, 比如, 以汽车产品为例, 相比于动力和舒适性, 油耗和性价比可能更受消费者重视[1]。因此, 确定属性权重并识别出关键属性在商品排序与选择研究[2] [3]、产品和服务优化研究[4]、消费者满意度研究等领域都是必须且重要的步骤。由于在线评论是消费者对所购买产品的评价, 具有较高客观性和价值密度, 且对消费者的购买决策影响显著[5], 所以在线评论可以作为挖掘消费者对属性重视程度, 即确定产品属性权重的有效数据。但目前基于在线评论确定属性权重的研究较少, 根据在线评论的利用程度将主要相关研究分为两类: 一类是仅利用在线评论挖掘产品属性, 而属性权重确定与在线评论关系不大。陈卓群和王忠义通过词频统计和机器学习对在线评论进行处理, 获得产品属性体系, 然后通过传统专家打分的层次分析法确定属性权重[6]。另一类是产品属性挖掘和属性权重确定都基于在线评论, 对在线评论的利用更充分、更完整。该类方法中最简单的就是用特征描述文本长度确定特征的权重[7] [8], 该方法操作简单容易理解但不够严谨。另外一种方法和思路是基于评论中评价分值或情感分值的分布函数进行属性权重确定, 比如习扬和樊治平用离散型概率分布函数来表示在线评价信息, 然后以加权累积分布函数决策矩阵与理想累积分布向量的距离为目标函数, 构建确定属性权重的优化模型[1]; 冯坤等基于 LDA 主题模型提取出在线评论中的属性及属性情感值分布函数, 然后根据随机占优准则和 PROMETHEE-II 方法给出不同商品类别中属性的排序结果[9]。该方法容易受到评论样本质量的影响, 尤其在样本不大的情况下, 特征分布会与大样本分布存在较大偏差。除此之外, 李保珍和封胜杰构建以属性值和产品总体评价分别为自变量和因变量的 Logistic 回归模型, 以求解得到的自变量系数作为权重系数[10]; 陶玲玲和尤天慧基于消费者对酒店各属性的评价矩阵, 运用熵权法确定属性权重值[11]。

如何从挖掘出的属性中识别出关键属性也是重要的研究方向, 其中降维算法是较普遍使用的研究方法, 典型的线性降维算法有 PCA (主成分分析)、LDA (线性判别分析) [12]、ICA (独立成分分析) [13]等; 非线性降维算法主要包括 KLDA (核线性判别分析) [14]、KPCA (核主成分分析) [15]、HE (海森特征谱方法) [16]等。但基于在线评论文本对属性值的描述是定性的且目标属性的确定存在一定难度, 所以传统的降维算法在基于在线评论识别关键属性的研究中存在一定局限性。所以, 较多改进算法被提出来, 比如熊熙等提出了基于模糊选项关系的关键属性提取方法, 并分别与传统降维算法 PCA 和 LDA 相结合[17]。而产品关键属性的一般定义是指受到消费者关注度高的属性集合。所以, 识别产品关键属性不完全是一

个降维问题, 而是如何量化消费者对属性关注度的问题, 比如刘宪立和赵坤运用模糊集理论与 DEMATEL 方法, 得到影响在线评论有用性的属性权重和中心度, 并以此对属性进行排序确定关键属性[18]; 钟嘉佑等依照数据目标对用户的在线产品评论进行了有效性判断, 选择并提取出了表达用户需求的产品关键设计属性[19]。

综上所述, 基于在线评论确定属性权重的本质是多属性决策问题, 但与传统属性权重研究不同, 此类研究需要运用文本挖掘技术, 如何表达在线评论信息是研究关键与难点。本文将文本挖掘技术与属性权重确定方法相结合, 提出了一种融合模型, 该模型包括改进 LDA 主题模型、情感分析词典法、信息增益和改进层次分析法, 用于对评论文本数据进行深度处理, 从而确定产品各属性权重。

2. 模型过程

2.1. 基于改进 LDA 模型进行产品属性挖掘

Word2Vec 词向量法可以把对文本内容的处理简化为向量空间中的向量运算, 计算出向量空间上的相似度, 表示文本语义上的相似度, 从而识别出近似词。采用融合 Word2Vec 的改进 LDA 主题模型能够有效识别特征词中的近似词, 从而提高主题间的区分度。

1) 评论文本预处理。首先对评论文本进行去重、去除评论内容为“此用户未作出评论!”, 然后运用 Jieba 分词对评论文本进行分词, 通过基知网停用词表去除停用词和表情图案。最后, 在剔除<名词, 形容词>词语结构中的形容词后, 将包含名词和形容词的评论提取出来作为主题挖掘文本数据。

2) 候选特征词词集构建。对预处理后的文本数据进行词频统计, 设置词频的阈值, 将高于阈值的词汇提取出高频词, 并人工剔除无法明确体现特征的词汇, 构建初始特征词词集。

3) 构建相似特征词词典。以评论语料作为输入语料, 借助 Gensim 库的 Word2vec 函数训练生成词向量文件, 利用相似度函数公式(2.1), 计算得出候选特征词的相似特征词, 构建相似特征词词典{候选特征词 L_n : 相似词 1, 相似词 2, ……}。

$$Similarity = \frac{u \cdot v}{\|u\| \cdot \|v\|} \quad (1)$$

其中 u 和 v 分别表示两个词汇的词向量。

4) 近似词整合与 LDA 主题挖掘。根据相似特征词词典对预处理后的主题挖掘语料库进行相似词整合, 然后通过一致性得分确定主题数 k , 再进行 LDA 主题挖掘, 得到属性 - 特征词集合, $W_k = \{w_{k1}, \dots, w_{kj}, \dots, w_{kJ_k}\}$, 其中 w_{kj} 表示属性 s_k 的第 j 个特征词, J_k 表示第 k 个主题属性所包含的特征词数量。

2.2. 基于情感分析进行满意度标注

首先基于 HowNet 领域情感词词典对各属性进行情感分析, 计算评论中所含特征词的情感值, 然后将情感值转换为满意度并对评论进行标注:

1) 提取含有总体评价特征词的评论。先对原始文本数据进行词频分析, 提取构建反映总体评价的标志词集合 s_0 。然后根据集合提取出含有总体评价的评论, 并剔除仅含由总体评价不含有产品属性特征的无效评论;

2) 分离文本单元。先将单个评论文本按照标点符号和连接词分成多个文本单元, 使每个文本单元都含有一个特征词, 则得到评论 - 文本单元集合, $C_m = \{C_{m1}, \dots, C_{mn}, \dots, C_{m\tau_m}\}$, 其中 C_{mn} 表示第 m 条评论的第 n 个文本单元, $n = 1, 2, \dots, \tau_m$, 其中 τ_m 表示第 m 条评论所包含的文本单元数量。

3) 标记出特征词及其所属属性。将文本单元分词结果与主题属性 - 特征词集合相匹配, 标注其包含的特征词, 并剔除不含特征词的文本单元, 则得到评论 - 特征词集合, $W'_m = \{W'_{m1}, \dots, W'_{mn}, \dots, W'_{m\theta_m}\}$, 其中 W'_{mn} 表示第 m 条评论的第 n 个特征词, θ_m 表示第 m 条评论所包含的特征词数量。将“评论 - 特征词”集合和“属性 - 特征词”集合进行匹配, 可以得到“评论 - 属性”集合, $S_m = \{s_{m1}, \dots, s_{mn}, \dots, s_{m\theta_m}\}$ 。

4) 计算文本单元情感值。将文本单元的分词结果与完善后的情感词典进行匹配, 标记每个文本单元中情感词的位置及其情感权重, 正面情感词得分+1, 消极情感得分-1。同时考虑否定词和程度副词, 程度副词具体维度见表 1, 若情感词前有奇数个否定词, 则情感倾向改变; 若短句中有偶数个否定词, 则情感倾向不变。而程度副词能够反映情感程度大小, 如“口感非常好”和“口感还可以”, 两者都是对“口感”这一特征的正面评论, 但程度副词的不同使二者存在较大的情感程度差异。参照 HowNet 程度副词词典建立程度副词词典, 并赋予相应的权重, 确保计算所得情感值的准确性。

Table 1. List of adverbs of degree

表 1. 程度副词词表

程度副词	程度级别	权值大小
完全、十分、很、非常、特别	2	2
比较、挺、还	1	1.5

依据文本单元中否定词词表和程度副词词典, 对各个文本单元情感得分进行修改, 则文本单元 C_{mn} 准确的情感得分:

$$Emotion(c_{mn}) = Q_{mn} = (-1)^g PD \quad (2)$$

其中 g 表示文本单元 C_{mn} 中否定词数量, D 和 P 分别表示文本单元 C_{mn} 中程度副词的权重值和情感词得分。因为一个文本单元仅含有一个特征词, 所以文本单元的情感值等价于特征词的情感值, $score(c_{mn}) = score(w_{mn}) = Q_{mn}$ 。

5) 情感值转换与满意度标注。消费者对产品的满意度和情感值的区别在于: 满意度是和购买前期望值的比较量, 而情感值是描述消费者情感方向与程度的标量, 二者不能等价。所以对消费者满意度进行分析需要经过转换关系将情感值转换为满意度, 根据李克特五分量表法设置转换关系, 如表 2 所示。其中作出相关假设并作为标注规则:

- 对于评论中未提到的属性, 假设其基本符合消费者期望, 以“基本满意”进行标注;
- 对于评论中提到的属性, 假设其受消费者重视程度高于未被提到的属性, 以“比较满意”或“非常满意”进行标注;
- 对于单个评论中存在的多个相同属性下的特征词, 取其中情感值最低的特征词;

Table 2. Emotional value and satisfaction conversion scale

表 2. 情感值与满意度转换关系量表

情感值	满意度 U	与期望值 E 关系	满意度分值
-2	非常不满意	$U \ll E$	1
-1, -1.5	比较不满意	$U < E$	2
0	基本满意	$U = E$	3
1, 1.5	比较满意	$U > E$	4
2	非常满意	$U \gg E$	5

按照以上满意度标注规则对每条评论进行标注, 得到评论 - 满意度集合, $U_m = \{U_{m1}, \dots, U_{mk}, \dots, U_{mK}\}$, 其中 U_{mk} 表示第 m 条评论中属性 s_k 的满意度水平。整个基于情感分析进行评论满意度标注过程如图整个基于情感分析进行评论满意度标注过程如图 1 所示。

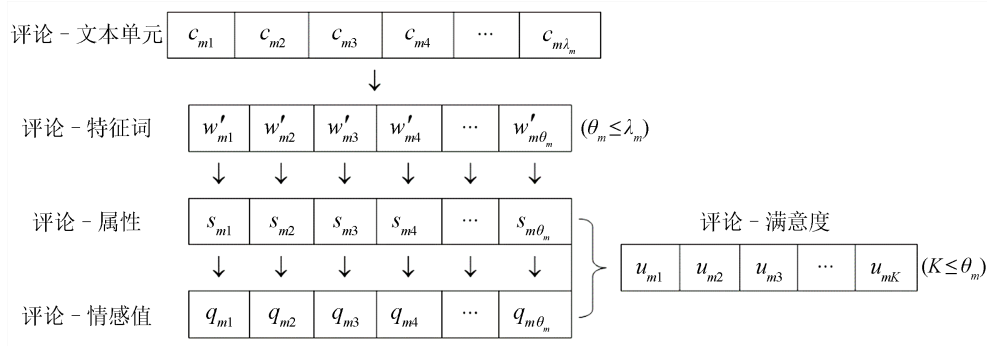


Figure 1. The entire process of satisfaction labeling based on emotional analysis
图 1. 基于情感分析的满意度标注全过程

2.3. 基于信息增益值进行属性重要性确定

以满意度标注结果为样本数据, 计算各属性的信息增益值用于衡量其重要性。考虑到消费者对购买的同一产品的属性偏好具有一致性, 且不同群体消费者对同一属性的偏好程度具有差异性[17], 因此假设属性的重要性值服从正态分布: $I \sim N(\mu, \sigma^2)$ 。同时为了降低样本质量对计算结果的影响, 需对信息增益值进行多次计算。

- 1) 对样本数据进行 N 折交叉处理, 随机选择其中 n ($n < N$) 折, 得到样本集合,

$$Sample = \{sample_1, sample_2, \dots, sample_t\} \quad t = C_N^n = \frac{N!}{n!(N-n)!};$$

- 2) 分别计算 t 个样本中目标属性 D 的信息熵值:

假设目标属性 D 的概率分布为 $P(U(D) = i) = p_i, i = 1, 2, 3, 4, 5$, 则

$$H(D) = -\sum_{i=1}^n p_i \log p_i \tag{3}$$

- 3) 分别计算 t 个样本中属性 s_k 下的条件熵值:

假设特征属性 s_k 的概率分布为 $P(U(s_k) = i) = p_i^k, i = 1, 2, 3, 4, 5$, 则

$$H(D|s_k) = -\sum_{i=1}^5 p_i^k H(D|u_{mk} = i) \tag{4}$$

- 4) 分别计算 t 个属性 s_k 的信息增益值:

$$IG_K = H(D) - H(D|s_k) \tag{5}$$

- 5) 重要性值的正态分布表示: 根据以上步骤得到的 $k \times m$ 个信息增益值结果, 计算各属性总体正态分布参数估计值。

$$\mu_k = \frac{\sum IG_t^k}{t}, \quad \sigma_k = \sqrt{\frac{\sum (IG_t^k - \mu_k)^2}{t-1}} \tag{6}$$

则属性 s_k 的重要性服从以下正态分布: $I \sim N\left(\frac{\sum IG_t^k}{t}, \sqrt{\frac{\sum (IG_t^k - \mu_k)^2}{t-1}}\right)$ 。

2.4. 基于改进层次分析法进行属性权重确定

1) 构建判断矩阵。不同于通过专家对指标进行 9 标度打分的传统主观方法, 本文依据各属性重要性值分布间的距离大小进行 4 标度, 构建判断矩阵 A :

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KK} \end{bmatrix} \quad (7)$$

矩阵中, a_{ij} 为判断矩阵中因素 i 与 j 比较值, 且满足 $a_{ij} = \frac{1}{a_{ji}}$ 。

$$a_{ij} = \begin{cases} 1 & \text{因素 } i \text{ 和因素 } j \text{ 一样重要} \\ 2 & \text{因素 } i \text{ 比因素 } j \text{ 相对重要} \\ 3 & \text{因素 } i \text{ 比因素 } j \text{ 明显重要} \\ 4 & \text{因素 } i \text{ 比因素 } j \text{ 绝对重要} \end{cases} \quad (8)$$

对于都服从正态分布的两个变量 i 和 j , 依据正态分布的累积概率密度对 δ 进行层级划分与标度 a_{ij} 设置: 当 $\delta = 0$ 时, 即 $\mu_i = \mu_j$, 则 $a_{ij} = 1$; 当 $0 < \delta \leq 1$ 且 $\mu_i > \mu_j$, 则 $a_{ij} = 2$; 当 $1 < \delta \leq 2.58$ 且 $\mu_i > \mu_j$, 则 $a_{ij} = 3$; 当 $\delta > 2.58$ 且 $\mu_i > \mu_j$, 则 $a_{ij} = 4$ 。

2) 构建优化矩阵。为了满足一致性要求, 通过转化公式对判断矩阵元素 a_{ij} 进行转换, 得到优化矩阵 B :

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1K} \\ b_{21} & b_{22} & \cdots & b_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K1} & b_{K2} & \cdots & b_{KK} \end{bmatrix} \quad (9)$$

$$b_{ij} = \frac{\sqrt[K]{\prod_{l=1}^K a_{il}}}{\sqrt[K]{\sum_{l=1}^K a_{jl}}} \quad (10)$$

3) 指标权重计算。计算各属性指标的单一权重值:

$$w_k = \sqrt[k]{\prod_{j=1}^K b_{kj}} \quad (11)$$

然后作归一化处理:

$$w'_k = \frac{w_k}{\sum_{k=1}^K w_k} \quad (12)$$

3. 实证研究

3.1. 数据获取与预处理

以同一品牌的肉类生鲜产品为研究对象, 采用 Python 编写爬虫程序从天猫平台共爬取 25,560 条该类产品在线评论。首先进行评论去重, 然后结合人工词典, 运用 Jieba 分词对评论文本进行分词, 通过知网停用词表去除停用词和表情图案。然后将能够表现产品属性及特征的名词和形容词提取出来作为主题挖掘文本语料。但其中<名词, 形容词>词语搭配中的形容词往往无法体现产品的具体属性, 比如形容词

“高”和“快”等，所以将该类形容词从语料中剔除。

3.2. 产品属性挖掘

采用 LDA 主题模型进行主题挖掘，首先根据一致性得分(Coherence Score)确定最优主题数为 6，然后运用 sklearn 中的 LDA 模型进行主题挖掘。对于主题挖掘结果，首先只保留各个主题下权重大于 2% 的特征词，然后由五组研究人员分别确定主题名称，最后依据投票原则对主题名称进行统一确定。同时对于存在多个主题下的特征词，由五组研究人员重新将该特征词归于一个主题中，最后依据投票原则确定该特征词所属主题。最后挖掘出该产品具有六大主题，包括“价格”、“包装”、“物流”、“服务”、“品质”和“份量”，具体结果见表 3。

Table 3. Product attribute mining results

表 3. 产品属性挖掘结果

属性主题	标签	特征词
价格	s_1	价格、实惠、划算、便宜、性价比、物美价廉
包装	s_2	包装、冰袋、保温袋、冰、冰块
物流	s_3	物流、快递、发货、运输、送货、速度
服务	s_4	服务、服务态度、客服、商家、售后服务、品牌
品质	s_5	品质、新鲜、干净、品相、味道、肉质、口感
份量	s_6	份量、分量、斤两、大小、个头、重量

3.3. 情感分析与满意度标注

对预处理后的评论文本数据进行词频分析，从词频结果挖掘并构建出反映总体评价的标志词集合 s_0 ，{“总之”，“总体”，“总的来说”，“总得来说”，“体验”，“购物”，“购物体验”，“回购”，“推荐”，“好评”，“差评”}。然后提取出含有该类评价特征词的评论，共有 16,034 条，并剔除其中不含有属性特征词的评论，如“总的来说很好，下次还来”，还有 10,651 条有效评论作为情感分析与满意度标注的文本数据。

基于 HowNet 领域情感词词典对各属性进行情感分析，通过公式(2)计算评论中所含特征词的情感值，如表 4 所示，然后根据表 2 转换关系量表将情感值转换为满意度，并按照标注规则进行评论满意度标注，如表 5 所示，整个情感分析和满意度标注过程见图 2。

Table 4. Results of emotional value analysis of comments

表 4. 评论情感值分析结果

ID	s_1	s_2	s_3	s_4	s_5	s_6	s_0
1	1	2	1	0	0	0	1
2	0	2	2	0	0	0	2
3	0	0	0	0	0	0	0
4	2	1	2	2	1	2	2
5	1	0	0	0	0	-1	1
□	□	□	□	□	□	□	□
10,651	1	0	0	0	-1	0	1

Table 5. Comment satisfaction labeling results**表 5.** 评论满意度标注结果

ID	s_1	s_2	s_3	s_4	s_5	s_6	s_0
1	4	5	4	3	3	3	4
2	3	5	5	3	3	3	5
3	3	0	3	3	3	3	3
4	5	4	5	5	4	5	5
5	4	3	3	3	3	2	4
□	□	□	□	□	□	□	□
10651	4	3	3	3	2	3	4

虽然价格<价格>比实体店贵<-1 1>//但物流<物流>迅速<1 4>//客服

服务态度<服务>好<1 4>//很不错<2 5>的一次购物体验<总体>

情感值	评论ID	价格	包装	物流	服务	品质	份量	总体
	51	-1	0	1	1	0	0	2

满意度	评论ID	价格	包装	物流	服务	品质	份量	总体
	51	1	2	4	4	2	2	5

Figure 2. Example of emotional analysis and satisfaction labeling**图 2.** 情感分析和满意度标注实例

3.4. 属性信息增益值计算与总体分布估计

对样本数据进行十折交叉处理, 随机选择其中 6 折, 得到包含 210 个不同样本的样本集合。通过公式(3)、(4)和(5)分别依据 210 个不同数据样本计算各属性的信息增益值, 结果见表 6。然后通过公式(6), 估计出各属性总体正态分布参数, 参数估计结果见表 7。“物流”属性为例, 该属性重要性值服从正态分布 $I(s_3) \sim N(0.6718, 0.0200^2)$ 。

Table 6. Calculation results of gain values for each attribute information**表 6.** 各属性信息增益值计算结果

ID	s_1	s_2	s_3	s_4	s_5	s_6
1	0.8119	0.7639	0.6604	0.9132	0.9042	0.7625
2	0.7989	0.7370	0.6635	0.8805	0.9061	0.7617
3	0.8086	0.7556	0.6584	0.9104	0.9128	0.7664
4	0.8345	0.7384	0.6826	0.8739	0.9224	0.7352
5	0.8182	0.7733	0.6700	0.8991	0.9121	0.7767
□	□	□	□	□	□	□
210	0.7842	0.7561	0.6765	0.8970	0.9296	0.7585

Table 7. Estimated values of overall distribution parameters for each attribute
表 7. 各属性总体分布参数估计值

参数	s_1	s_2	s_3	s_4	s_5	s_6
μ_k	0.8178	0.7463	0.6718	0.8858	0.9200	0.7599
σ_k	0.0156	0.0153	0.0200	0.0145	0.0103	0.0162

3.5. 属性权重确定

首先计算所有属性重要性值分布两两之间的距离大小 δ ，此次为依据确定标度 a_{ij} ，得到判断矩阵 A ：

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{bmatrix} = \begin{bmatrix} 1 & 3 & 4 & 1/3 & 1/4 & 3 \\ 1/3 & 1 & 3 & 1/4 & 1/4 & 1/2 \\ 1/4 & 1/3 & 1 & 1/4 & 1/4 & 1/3 \\ 3 & 4 & 4 & 1 & 1/3 & 4 \\ 4 & 4 & 4 & 3 & 1 & 4 \\ 1/3 & 2 & 3 & 1/4 & 1/4 & 1 \end{bmatrix}$$

如“价格”和“包装”两属性重要性分布间距离 $\delta_{31} = \frac{|0.6718 - 0.8178|}{|0.0200 + 0.0156|} = 4.1101$ ，因为 $\delta > 2.58$ 且

$\mu_1 > \mu_3$ ，则 $a_{13} = 4$ ，同时 $a_{31} = 1/4$ ；

然后对通过转化公式(2.10)对判断矩阵元素 a_{ij} 进行转换，得到优化矩阵 B ：

$$B = \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\ b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} \\ b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66} \end{bmatrix} = \begin{bmatrix} 1 & 2.14 & 3.46 & 0.60 & 0.40 & 1.70 \\ 0.47 & 1 & 1.62 & 0.28 & 0.19 & 0.79 \\ 0.29 & 0.62 & 1 & 0.17 & 0.11 & 0.49 \\ 1.67 & 3.56 & 5.77 & 1 & 0.66 & 2.83 \\ 2.52 & 5.39 & 8.73 & 1.51 & 1 & 4.28 \\ 0.59 & 1.26 & 2.04 & 0.35 & 0.23 & 1 \end{bmatrix}$$

$$\text{如 } b_{12} = \frac{\sqrt[6]{1 \times 3 \times 4 \times \frac{1}{3} \times \frac{1}{4} \times 3}}{\sqrt[6]{\frac{1}{3} \times 1 \times 3 \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{2}}} = 2.14。$$

由得到的优化矩阵通过式(7)和式(8)得到归一权重值 w_k ：

$$w_k = [0.0156, 0.077, 0.048, 0.246, 0.378, 0.095]$$

所以，“价格”、“包装”、“物流”、“服务”、“品质”和“份量”六个产品属性权重分别为 0.18、0.17、0.06、0.25、0.35 和 0.09，因此权重最大属性是产品品质，其次是服务质量、产品价格，而产品份量、包装和物流的权重较低，说明对于生鲜产品，消费者最看重生鲜产品的新鲜度、是否干净和口感好不好等品质问题，其次是商家服务质量和态度好不好，以及产品价格是否实惠。

4. 结束与展望

产品属性权重的确定与关键属性的识别具有重要现实意义，既能够帮助消费者进行商品比较与选择，也为公司进行产品和服务优化决策提供切实依据。考虑到在线评论作为消费者对商品的售后评论具有较高客观性和价值密度，因此，文章以在线评论为研究数据，将文本挖掘技术与属性权重确定方法相结合，

提出了 LDA-IGR-IAHP 的方法模型, 用于对评论文本数据进行深度处理, 从而确定产品各属性权重并识别出关键属性。首先基于改进 LDA 主题模型挖掘出产品属性, 然后基于属性情感分析对评论中属性满意度进行标注。之后先对满意度标注样本进行 N 折交叉处理, 然后计算所有不同样本下各属性的信息增益值, 以此对各属性重要性值分布进行参数估计。然后, 计算两两分布的距离因子, 最后运用改进 AHP 方法确定属性权重。以肉类生鲜产品的在线评论为数据进行实证分析, 挖掘出肉类生鲜产品的六个属性, 并计算得到各属性的重要性权重值。结果表明, 所提出的融合模型能够准确计算出产品各属性权重。

文章提出的方法模型既克服了传统权重确定方法的主观性弊端, 同时也尽可能避免了样本质量对结果的干扰。本研究不足之处在于只构建了一级产品属性体系, 而部分商品的属性特征往往比较复杂。

参考文献

- [1] 习扬, 樊治平. 使用在线评价信息的属性权重确定及方案排序方法[J]. 控制与决策, 2016, 31(11): 1998-2004.
- [2] Najmi, E., Hashmi, K., Malik, Z., et al. (2015) CAPRA: A Comprehensive Approach to Product Ranking Using Customer Reviews. *Computing: Archives for Informatics and Numerical Computation*, **97**, 843-867. <https://doi.org/10.1007/s00607-015-0439-8>
- [3] Liu, Y., Bi, J.W. and Fan, Z.P. (2017) Ranking Products through Online Reviews: A Method Based on Sentiment Analysis Technique and Intuitionistic Fuzzy Set Theory. *Information Fusion*, **36**, 149-161. <https://doi.org/10.1016/j.inffus.2016.11.012>
- [4] Kim, S.W., Lee, K., Sohn, J.S., et al. (2020) Product Development Using Online Customer Reviews: A Case Study of the South Korean Subcompact Sport Utility Vehicles Market. *Applied Sciences*, **10**, 6918. <https://doi.org/10.3390/app10196918>
- [5] Lee, Y.J., Hosanagar, K. and Tan, Y. (2015) Do I Follow My Friends or the Crowd? Information Cascades in Online Movie Ratings. *Management Science*, **61**, 2241-2258. <https://doi.org/10.1287/mnsc.2014.2082>
- [6] 陈卓群, 王忠义. 意见挖掘中特征权重的计算方法研究[J]. 情报科学, 2012, 30(5): 759-763.
- [7] 王安宁, 张强, 彭张林, 倪鑫. 基于在线评论的区域需求偏好识别方法[J]. 中国管理科学, 2019, 27(7): 167-176.
- [8] 熊回香, 李晓敏, 李跃艳. 基于图书评论属性挖掘的群组推荐研究[J]. 数据分析与知识发现, 2020, 4(Z1): 214-222.
- [9] 冯坤, 杨强, 常馨怡, 李延来. 基于在线评论和随机占优准则的生鲜电商顾客满意度测评[J]. 中国管理科学, 2021, 29(2): 205-216.
- [10] 李保珍, 封胜杰. 基于用户选择的产品属性权重确定及对 TOPSIS 法的改进[J]. 运筹与管理, 2017, 26(3): 54-62.
- [11] 陶玲玲, 尤天慧. 基于在线评价且考虑消费者类型的酒店排序方法[J]. 运筹与管理, 2021, 30(5): 122-128.
- [12] Duda, R.O., Hart, P.E. and Stork, D.G. (2000) Pattern Classification. 2nd Edition, John Wiley & Sons, Inc., Hoboken.
- [13] Comon, P. (1994) Independent Component Analysis, a New Concept. *Signal Processing*, **36**, 287-314. [https://doi.org/10.1016/0165-1684\(94\)90029-9](https://doi.org/10.1016/0165-1684(94)90029-9)
- [14] Bernhard Schouml, et al. (1998) Nonlinear Component Analysis as a Kernel Eigenvalue Problem. *Neural Computation*, **10**, 1299-1319. <https://doi.org/10.1162/089976698300017467>
- [15] Kamel, S., Nicola, L.C., et al. (2007) Optimally Regularised Kernel Fisher Discriminant Classification. *Neural Networks*, **20**, 832-841. <https://doi.org/10.1016/j.neunet.2007.05.005>
- [16] Donoho, D.L. and Grimes, C. (2003) Hessian Eigenmaps: Locally Linear Embedding Techniques for High-Dimensional Data. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5591-5596. <https://doi.org/10.1073/pnas.1031596100>
- [17] 熊熙, 乔少杰, 韩楠, 元昌安, 张海清, 李斌勇. 一种基于模糊选项关系的关键属性提取方法[J]. 计算机学报, 2019, 42(1): 190-202.
- [18] 刘宪立, 赵昆. 在线评论有用性关键影响因素识别研究[J]. 现代情报, 2017, 37(1): 94-99+105.
- [19] 钟嘉佳, 胡磊, 孙琦宗, 华尔天. 基于用户分类的参与式交互方法研究[J]. 机电工程, 2018, 35(6): 572-576.