Expert Systems With Applications 245 (2024) 123015

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Product improvement in a big data environment: A novel method based on text mining and large group decision making

Fang Zhang [a], Wenyan Song [b,c,*]

[a] School of Economics, Capital University of Economics and Business, Beijing, 100070, China
[b] School of Economics and Management, Beihang University, Beijing, 100191, China
[c] Key Laboratory of Complex System Analysis, Management and Decision, Beihang University, Ministry of Education, Beijing, 100191, China

ARTICLE INFO

Keywords:
Product improvement
Consumer big data
Large group decision making
Text mining
Word2Vec

ABSTRACT

Product improvement has become a multifaceted and uncertain endeavour for manufacturers in an increasingly competitive business environment. Online platforms have emerged to solicit consumer opinions and product feedback. However, product improvement requires a critical yet complex decision-making approach for manufacturers. Managers face the challenge of identifying the most effective decision-making methodology for product improvement, especially in a big data environment. In this research, we comprehensively evaluate different product improvement decision-making methodologies through a series of experimental investigations. Specifically, three different experiments are conducted, including: i) an initial selection guided by intuitive perception, ii) expert decision-making, and iii) a hybrid method that incorporates consumer big data and large-scale group decision-making. Product criteria sets are categorized using the Latent Dirichlet Allocation (LDA) method, while the importance of these criteria is determined by applying the TextRank and Word2Vec algorithms. Our empirical results show that the mixed method, which utilizes text-mining techniques in conjunction with large-group decision-making, provides a more reliable and effective approach to facilitating product improvement.

1. Introduction

The development of online business platforms has led to an enormous influx of consumer review data. These reviews reflect consumers' experiences with products and the overall shopping process (Zhang et al., 2019). The wealth of consumer review big data is a valuable resource for new consumers seeking product information and manufacturers aiming to identify consumers' needs (Beugelsdijk et al., 2017; Raghupathi et al., 2015). Such reviews contain insights into consumers' potential needs, preferences, and shopping experiences, making them crucial for guiding product improvement decisions. However, extracting meaningful information from consumer reviews can be challenging, as real opinions are often embedded within natural language expressions (Zhou et al., 2010).

The decision-making process for product improvement by manufacturers is both complex and scientific. It involves two key aspects. Firstly, expert decisions are vital in advancing products from a professional perspective, contributing to functional design. Secondly, leveraging user behavior decision models based on consumers' behavior is also beneficial for product improvement. In this context, online reviews, particularly those with high volume and sincerity, are essential for informing product improvement decisions. Nevertheless, online review data's sheer quantity and unstructured nature make manual analysis impractical. An automated, computer-enabled solution is required to effectively extract consumers' preferences and needs from these reviews.

Various natural language processing techniques are utilized to summarize consumers' online comments (Ravi & Ravi, 2015). Most of these approaches are feature-based, extracting words and terms as key factors to express consumers' opinions and product characteristics (Asadabadi et al., 2023; Burnap & Williams, 2016; Chan et al., 2016; Chen et al., 2023; Horvat et al., 2019; Tuarob & Tucker, 2015; Xu et al., 2016). However, these frameworks extract the phrases as pivotal factors, considering consumers' sentiments to the product features. Although the extraction is valuable, the holistic essences of user opinions are not captured, which are vital for product improvement. In addition, the decision-making made by experts, who rank product designs based on their preferences, prior knowledge, and experience is also necessary for

* Corresponding author.
E-mail address: songwenyan@buaa.edu.cn (W. Song).

https://doi.org/10.1016/j.eswa.2023.123015
Received 27 July 2023; Received in revised form 2 December 2023; Accepted 20 December 2023
Available online 27 December 2023

product improvement. Group decision-making has garnered significant attention in academia. Multi-criteria decision-making (MCDM) models are widely employed to determine the priority of alternatives for group decision-making issues. Previous research has typically treated the decision-making process for product improvement as a separate entity for designers and users. Designers often focus solely on professional product improvement, neglecting the needs and opinions of end users (Sinha & Anand, 2018). In fact, the experts' opinion and users' preferences are both significant. There is little literature on product improvement, considering the holistic essences of the consumers' reviews and the large group decision-making.

To address the gap in current research, this study introduces a scientific method that concurrently considers consumer preferences and expert decision-making in product improvement. By synthesizing consumers' online reviews and experts' evaluation, we aim to propose a more rigorous and holistic methodology for product enhancement. Further, an experimental design is implemented to corroborate its effectiveness. Firstly, we conduct a comprehensive review of the existing literature on product improvement, emphasizing the significance of considering both designers' opinions and user preferences and requirements in criteria selection. Secondly, an experimental study is performed to validate the efficacy of consumer-involved product improvement. Thirdly, the study employs online reviews analysis and group decision methods to achieve its objectives.

This paper contributes to the existing research from the following aspects: Firstly, we introduce a nuanced strategy that uniquely synthesizes end-user online reviews with traditional expert decision-making, forming a robust, holistic approach for product improvement. This method addresses the limitations of traditional processes, particularly the extensive time and potential inconsistency, by incorporating feedback from end-users. This integration provides decision-makers with a comprehensive, practical, and more immediate basis for strategy formulation, enhancing both efficiency and relevance of the improvements. Secondly, our model's effectiveness is rigorously tested through three thoughtfully designed experiments, comparing decisions influenced by intuitive perception, those using Multi-Criteria Decision Making (MCDM) factoring in group preferences and online review data, and decisions by larger groups. This approach underscores our model's precision and versatility, demonstrating its superiority in enabling more strategic, informed decision-making processes. Finally, departing from standard review analytics, our study employs Latent Dirichlet Allocation (LDA) for deeper, more refined data extraction, revealing nuanced thematic structures in user feedback. This method is complemented by machine learning algorithms that impartially handle weight distribution, enhancing analytical objectivity and robustness. Our approach offers manufacturers an innovative perspective, potentially revolutionizing product improvement strategies.

The following is the structure description of the paper. Section 2 reviews the existing research on consumer-involved product improvement and the approaches for product improvement. Section 3 presents the framework and experiment process of the study. Section 4 is a case study to illustrate the experiments. Section 5 evaluates the experiment results and discusses the weights obtained and final scores. Section 6 wraps up with a general conclusion.

2. Related work

2.1. Online product reviews and new product development

The rise of the internet and information technology has significantly transformed the role of social media and e-commerce platforms, establishing them as essential channels for sharing and exchanging product information. This information typically includes crucial details about product features and pricing, all readily accessible online. Furthermore, these platforms empower users to post product evaluations online to express their experiences and thoughts, providing valuable insights into product performance, quality, and associated services. In this context, the impact of online reviews is particularly noteworthy, as they play a crucial role in influencing consumer decision-making, especially when it comes to product purchases, thus shaping the landscape of retail marketing (Ventre & Kolbe, 2020; Xu et al., 2022).

Consumers' online feedback is increasingly acknowledged as a vital source of information for product improvement. Such feedback often captures consumers' preferences and requirements, offering invaluable insights for refining products (Chan et al., 2016). Tang et al. (2019) emphasized that online reviews not only convey consumers' shopping experiences but also provide qualitative assessments. Therefore, a thorough examination of these reviews can unveil latent consumer needs, enabling manufacturers to optimize their offerings and enhance consumer satisfaction (Ireland & Liu, 2018; Liu et al., 2020).

Product improvement stands as a pivotal aspect of the product lifecycle, bridging the gap between consumer demand and production expertise. Jeong et al. (2019) underscored that the essence of product improvement lies in the ability to translate the concepts of product designers into commercially viable products that align with the real needs of consumers. Consequently, there has been considerable academic interest in investigating the key success factors driving product improvement (González & Palacios, 2002).

In the realm of product improvement, the role of consumer feedback is increasingly central. The insights gleaned from understanding consumer preferences and requirements are invaluable for refining products (Chan et al., 2016). Tan et al. (2018) elucidated that consumer online product reviews are rich data sources, encompassing purchasing experiences and qualitative consumer assessments. For manufacturers, delving into these reviews is crucial to uncover potential consumer needs and preferences, thereby guiding the enhancement of new products to better meet consumer satisfaction (Ireland & Liu, 2018; Liu et al., 2020).

The utilization of online reviews actively promotes product improvement. Therefore, manufacturers should not overlook the importance and value of online review data in the product improvement. Existing research consistently demonstrates the impact of online reviews on product improvement. For instance, Ding et al. (2022) analyzed hotel reviews to understand the growth of independent and branded hotels. Jain et al. (2021) conducted consumer sentiment analysis using online reviews in various fields such as healthcare, finance, and travel.

While there is a growing body of literature on the influence of online reviews in areas such as hotel, healthcare and finance service, there remains a scarcity of research specifically focused on improving products in the manufacturing industry. Chan et al. (2016) used the probability weighting function to evaluate the criteria weights deriving from the reviews on mobile phones on social media. The existing research considered the unbiased unstructured data extracting from consumers' online reviews into product development (Goldberg & Abrahams, 2022; Huang et al., 2022). Asadabadi et al. (2023) involved text mining on online product reviews using sentiment analysis. Chen et al. (2023) focus on product configuration designs, which revolve around predetermined product modules tailored to cater to varying consumer needs. Sentiment analysis, while effective, is not conducive to innovative product development that introduces new product attributes. In contrast, genuine product improvement relies on a product's holistic data.

It is crucial to find out a scientific and effective method to analyze and understand reviews data, and respond to the critical information in the product improvement process. Thus, this study extracts the consumers' reviews to support the product improvement process using a scientific computing model to identify the critical factors of the consumer-generated data.

2.2. Text mining of online product reviews

Content analysis is widely used to explore the qualitative data online and quantify the end users' reviews. The online review process contains topic identification and context analysis. The latent Dirichlet Allocation (LDA) method is preferred for extracting the topic distribution from the documents. The LDA method is first introduced briefly, and then the topic criteria extracted by the LDA are processed. As a member of the unsupervised machine learning method, LDA, first proposed by Blei et al. (2003), belongs to the generative probabilistic model of a corpus. As documents essentially consist of distribution over words, LDA shows that documents can be regarded as an alternative form of random combinations for latent topics. Meanwhile, LDA assumes a common Dirichlet exists prior to topic distribution of every document. We assume each latent topic in the LDA model is an alternative form of a probabilistic distribution over words. In the same way, a common Dirichlet prior exists for topic distribution of every word. Generally, we can assume that each online review can be regarded as a mixed probability distribution consisting of several relevant topics. Each topic is provided with a probability distribution over several terms. As Blei et al. (2003) proposed, the novel approach to dealing with online reviews is determining the topics and then choosing the terms according to the topics with the probability distribution. Bastani et al. (2019) adopted LDA to analyze the Consumer Financial Protection Bureau consumer complaints. Guo et al. (2017) and Jeong et al. (2019) used the LDA model to identify consumers' voices.

As to the context analyzing algorithms, two kinds of machine learning models are population for identifying the dominant factors and the importance of the criteria from the online reviews, including term frequency-inverse document frequency (TF-IDF) algorithm and TextRank algorithm. Razzaghnoori et al. (2018) pointed out that the TF-IDF algorithm is a popular textual information analysis method for acquiring key factors and weights. The importance of a word will depend upon the frequency of its occurrence in the whole text. Wang et al. (2020) used TF-IDF and Word2Vec to analyse travelers' concerns and opinions regarding hotel selection. Unfortunately, the TF-IDF algorithm will result in a curse of dimensionality (Razzaghnoori et al., 2018). Additionally, TF-IDF ignores the relationship between terms.

TextRank algorithm is proposed to overcome these problems. TextRank algorithm originates from the PageRank algorithm (Mihalcea & Tarau, 2004), measuring the importance for each web page shown in Google. Words will be considered as "nodes on the network graph" and the significance of each word will be calculated based on the co-occurrence relationship between words. Qin and Zeng (2022) extracted evaluation attributes and determined weights from online reviews for product ranking using a novel method based on the TextRank algorithm. Irma Patricia Delgado-Solano et al. (2018) extracted keywords from users' written requirements by applying the TextRank algorithm and inverse frequency approach. However, the model is assumed to ignore the overall distribution between words. This suggests that while the TextRank algorithm is effective in isolating key attributes and keywords, it might not fully account for the nuanced interplay and distribution of words within the textual data.

For similarity analysis, Word2Vec is the first choice to measure the amount of similarity between two textual comments. Mikolov et al. (2013) proposed Word2Vec by using TensorFlow, the deep learning framework in Google. Word2Vec assumes that high probability occurs when the words are similar for semantic and syntactic. Word2Vec algorithm is also an effective way to compensate for dimensional disaster by obtaining the resemblance between key factors and criteria. Word2Vec, a widely recognized tool, has seen extensive use in various studies for its ability to extract meaningful criteria from online reviews. Wu et al. (2022) have notably applied this methodology. The technique was further employed by Zuo et al. (2017) specifically to calculate similarity matrices among words, enhancing the understanding of lexical relationships. Jatnika et al. (2019) and Zhang et al. (2020) expanded upon this usage by extracting semantic information between words from external documents using Word2Vec. Their research validated the algorithm's effectiveness in information extraction. Moreover, the capability of Word2Vec in sentiment prediction for products has been substantiated by Balakrishnan et al. (2022), who verified its accuracy in this domain. Joung and Kim (2023) also innovatively utilized Word2Vec to identify and group product feature words in reviews. Their approach involved a systematic process of clustering, filtering, and refining these words, showcasing the versatility of Word2Vec in processing and analyzing text data. However, in previous literature, the Word2Vec algorithm has seldom been applied to alleviate the curse of dimensionality.

Considering the features of the algorithms reviewed above, our research employs a scientific computing model to analyze the consumer review data. In this paper, we apply the Latent Dirichlet Allocation (LDA) model to extract criteria for product improvement from online reviews and combine TextRank and Word2Vec to extract key factors and calculate the weights of these criteria. This innovative approach addresses the shortcomings of each method, such as the dimensional complexity in Word2Vec and the potential oversight in word distribution by TextRank, thereby offering a more comprehensive and efficient solution for data analysis in product improvement contexts.

2.3. Decision-making methods for product improvement

The MCDM model is confirmed as a scientific approach for product improvement. Ishizaka and Siraj (2018) conducted experiments with prior ranking, MCDM ranking and posterior ranking to evaluate the effectiveness and usefulness of the MCDM. Hotel selection is conducted with a novel model based on online reviews on a tourism website (Yu et al., 2018). A model based on sentiment analysis of online reviews and TODIM is proposed to rank products (Zhang et al., 2020).

MCDM has been adopted to help manufacturers decide appropriate product designs for manufacturers. MCDM aims to select the best one among the alternatives in the presence of multiple, conflicting, decision criteria. Existing studies have concentrated on introducing different MCDMs for product improvement analysis. PROMETHEE-based MCDM approach is applied to rate alternatives premediating consumers' online reviews. The analytic hierarchy process (AHP) is adopted to rank renewable options for the Algerian electricity system (Haddad et al., 2017). Işıklar and Büyüközkan (2007) applied AHP-TOPSIS to assess the mobile phone alternatives concerning the consumers' preferences. The AHP method requires comparing every two pairs with more complex calculations. To solve the massive computing problem, Keršuliene et al. (2010) put forward the Step-wise Weight Assessment Ratio Analysis (SWARA) model as the group decision-making technique provides a comprise agreeing on the topic or combining different opinions.

In addition, regarding the prevalence of network media, large-scale group decision-making (LSGDM) makes a significant amount of decision-maker's decision results feasible. The LSGDM shows the following characteristics. Firstly, the number and geographical distance of decision-makers are unrestricted. Dozens to hundreds of decision-makers can express their idea at different places. Secondly, group decision-makers are from various fields. Two standards require the number of decision-makers to be 11 and 20 (Liu et al., 2014; Xu et al., 2015). LSGDM has been employed in various fields involving supply chain management, engineering management, hotel selection, etc. Ding et al. (2018) evaluated three alternatives for a highway construction project with 30 decision-makers. Wang et al. (2018) proposed a unanimous-based approach for LSGDM to establish a new project management system for a Chinese construction engineering company. The application of LSGDM for industrial product is few. Chen et al. (2020) crawled passengers' reviews online to extract their demand and evaluate the high-speed rail system using LSGDM methods. Ji et al. (2023) proposed a LSGDM to evaluate the peer-to-peer accommodation in sharing economy.

After reviewing the above studies, we realized the superiority and significance of consumers' online review and decision-making. We thus extend the large-scale decision-making approach to product improvement by constructing three experiments, including previous ranking, traditional decision making and combining online reviews and decision-making, to verify the scientific decision-making approach.

2.4. Knowlegdge gap

Although the field of product improvement through online reviews and expert decision-making has been substantial research, several gaps persist that this study intends to bridge:

(1) **Integrated Synergy between End-User Reviews and Expert Decision-making**: Existing literature has highlighted the significance of both online reviews and expert decision-making separately. However, there is a clear gap in research that integrates these two vital components into a unified, sophisticated approach for product improvement. The necessity for a strategy that transcends the time constraints of traditional expert decision-making processes is clear. Such a strategy should enhance efficiency and provide a more holistic viewpoint by amalgamating insights from end-users and industry experts. This integrated approach promises a more thorough and nuanced understanding of the market landscape, ensuring that product development decisions are timely, relevant, and reflective of diverse perspectives.
(2) **Empirical Validation of Combined Decision-making Approaches**: The existing research mainly proposes theoretical methodologies, but there's a noticeable dearth of literature providing experimental validation, especially those combining online reviews with large-scale decision-making. The comparative effectiveness of intuitive perception-based decision making, MCDM infused with online review data, and decisions by large-scale decision groups remain room for exploration.
(3) **Holistic Semantic Analysis to Product Improvement:** Although extant literature has involved online review analytics, they mainly focus on product configuration design. The essence of product improvement, which is grounded in a product's comprehensive data, is required. The profound exploration of customer requirements is necessary for product improvement.

3. Methodology

This paper provides an integrated decision framework to explore the key criteria of product improvement and the decision-making results regarding different decision methods. In this study, we conduct three experiments on product improvement decision-making, including cognitive decision-making, group decision-making, and the combination of online consumer reviews and large-scale decision-making.

This research supplies a feasible guideline for scholars to identify the key factors and the scores of the criteria. This study gives an insight of accurate product development for manufacturers to cater consumers' needs. The process of this study can be summarized as follows. Firstly, this study identifies the key factors representing consumers' concerns on product improvement, extracting from online reviews using the TextRank algorithm. Secondly, this study explores the criteria for product improvement from the key factors. The product improvement criteria are summarized from the existing research and extended by the key factors from the online reviews. Word2Vec algorithm is applied to identify the criteria importance by analyzing the key factors contents. Thirdly, this study conducts three experiments to explore the scientific decision-making methods. The experiments examine the ascendency of the significance of large-scale consumers' participation in product improvement.

The present study's research architecture is delineated in Fig. 1, encompassing three distinct experimental procedures. First, prior decision-making is conducted. Participants are prompted to rank alternatives based on their associated background information, measuring their subjective perceptions. Second, group decision-making using

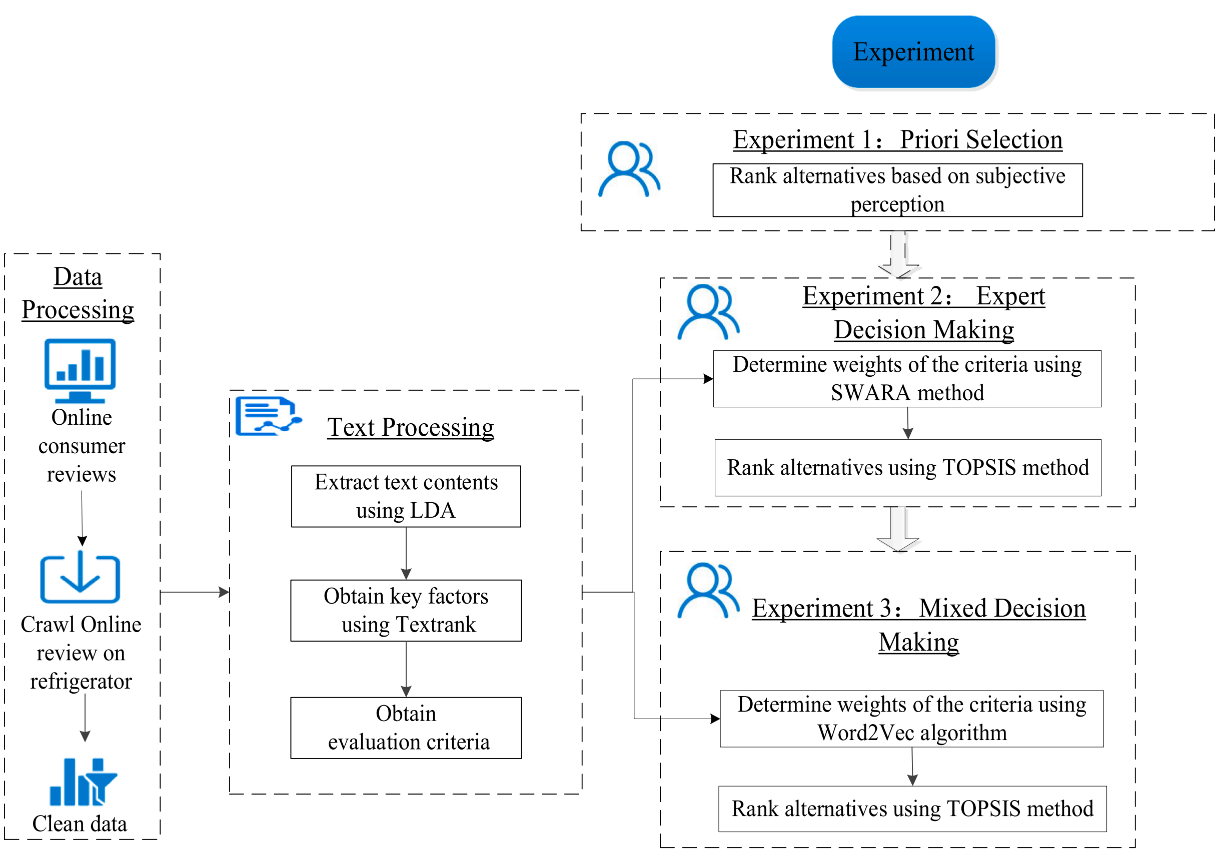

Fig. 1. The framework of the data-driven product improvement decision making.

MCDM is undertaken. Third, decision-making informed by online reviews and a large decision-making group is conducted. Online consumer feedback is sourced from an e-commerce platform, with rigorous text analysis techniques deployed to identify pivotal factors. Further, a multi-criteria method is implemented to rank the alternatives.

Online consumer feedback serves as a critical source of information for product improvement from the client side, determining the criteria for both Experiment 2 and Experiment 3. The pre-processing of online reviews encompasses two procedures: firstly, data processing, and secondly, text processing. The online reviews contain terms relating to product improvement and some irrelevant words. Initially, these reviews undergo a data processing phase. The online reviews filter and delete the stop words, sentiment, and degree terms. Subsequently, text processing is implemented. A “Review-Word” matrix is derived by calculating the recurrence rate of each conceptual term. LDA model is applied to train the data and obtain the “Review-Topic” matrix, “Topic-Word matrix,” and the topic list. To acquire more accurate results, we sieve the noisy words devoid of practical relevance. Topics with similar connotations are amalgamated, and labels are assigned based on the semantic value of the encompassed terms. The key factors and their weights are obtained using the TextRank algorithm. Upon acquiring the key factors, key factors can be grouped into eight topics. Referring to the criteria from existing literature on product improvement and combing the eight topics, we craft a comprehensive evaluation criteria framework. Evaluation criteria are categorized into four primary domains: price, product quality, product performance, and service, which are further divided into eight sub-criteria: price, easy to use, appearance design, cooling effect, noise, space, energy consumption and service. The extracted topics and corresponding terms from online reviews are represented in Table 1.

3.1. Description for experiment 1: priori selection

Ranking the alternatives based on decision-makers’ intuitive cognition. In our experiment, the background information of the four refrigerator alternatives is provided to the participants. The background information includes eight criteria shorted: “price”, “easy to use”, “performance design”, “cooling”, “noise”, “space”, “energy consumption” and “services”. Participants are asked to read the background information and briefly understand the alternatives.

Then, each participant is asked to rank the four alternatives according to their knowledge and personal perception. The participants are twenty experts, six new product designers for refrigerators, six refrigerator product sales managers, and eight experts from academia. The expert panel has rich product design and development expertise.

3.2. Description for experiment 2: group decision-making

Ranking the alternatives based on multi-criteria decision-making approaches. This paper applies the SWARA-TOPSIS approach to analyze the ranking results for the decision-making group. Twenty participants are invited to score their preferences. Eight criteria are mentioned in Section 3.1. The expert group is built to provide the judgment for the criteria to mensurate criteria’s weights and give the preference of the criteria to assess the significance for product improvement. The experts in Experiment 2 are the same as the expert participants in Experiment 1.

Based on the algorithm principle, participants were asked to provide different information. Subsequently, the SWARA-TOPSIS approach is applied to calculate the orders of the alternatives. The specific description of the approach is presented as follows.

(1) Obtain the weight coefficient using SWARA.

Some studies applied SWARA to define the subjective weight of the criterion in the MCDM process, which requires personal knowledge and experience. Using a technique based on group decision-making, the study aims to evaluate various judgments and priorities. The geometric mean method combines the different preferences and obtains a compromise. The geometric mean method is widely used to integrate the group’s decisions (Saaty & Shang, 2007). The specific procedures are shown as follows.

Step 1: Decision-makers are required to rate the criteria by importance. The top one is assumed to be the most crucial criterion. The others are sorted in descending order based on their significance. The last one is normally regarded as the least significant criterion.

Step 2: Decision-makers are required to compare the criteria. The comparison results are given after determining the ranking of all the criteria referring to the method developed by Zolfani and Saparauskas (2013). The importance value is given as “comparative advantage of the

Table 1
Evaluation criteria for product improvement.

Criteria	Sub-criteria	Reference source	Term and its occurrence frequency
Price C1	Price C11	(Chiu & Kremer, 2013; Dinerstein et al., 2018)	Discount (66), Pricerite (512), cost-effective (15), price (395), economy (122), cheap price (51), cheap (150), cost performance (334), deserved (439), attractive and reasonable price (100), excellent quality and affordable price (67), valuable (26)
Product quality C2	Easy to use C21	(Sinha and Anand; 2018)	Quality (529), problem (191), product quality (53), easy to use (109), nature of the product (138), effect (195), satisfaction (665), utility (177)
Product performance C3	Appearance design C31	(Chien et al., 2016; Yang et al., 2016)	Good looking (84), product appearance (166), level up (55)
	Cooling effect C32	(Hmida et al., 2019; Rahman et al., 2020)	Refrigeration (214)
	Noise C33	(Tao et al., 2019)	Mute (67), voice (238), noise (238)
	Space C34	(Tao et al., 2019)	Space (152), convenience (74), capacity (61)
	Energy consumption C35	(Tao et al., 2019)	Electricity consumption (108)
Service C4	Services C41	(Liu, 2011; Tao et al., 2018)	Attitude (233), provide good service (97), enthusiasm (62), patience (59), thoughtful service (157), service attitude (329), customer service (88), Delivery (1211), logistics (667), dispatching (50), home delivery (105), shipping (161), speed (394), ultrafast (45), in time (302), pretty fast (81), very fast (295), Installation (285), service (687), after-sales service (75), after sale (170), technician (622), installation personnel (106)

mean value (q_i)". q_i denotes the comparative importance of i th criterion to $i+1$ th criterion.

Step 3: The parameters t_i can be computed after comparison as Stanujkic et al.'s work (2015). Using the Eq. (1), we can compute the t_i coefficient.

$$t_i = \begin{cases} 1, i = 1 \\ q_i + 1, i > 1 \end{cases} \tag{1}$$

Step 4: Eq. (2) can determine the variable m_i.

$$m_i = \begin{cases} 1, i = 1 \\ \frac{t_{i-1}}{t_i}, i > 1 \end{cases} \tag{2}$$

Step 5: By using Eq. (3), we can acquire the criteria's weights.

$$w_i = \frac{m_i}{\sum_{i=1}^{n} m_i} \tag{3}$$

Where w_i represents the weight of criterion i. Then by evaluating the criteria for each weight, the priority vector is calculated by the evaluation.

(2) Rank the alternatives adopting TOPSIS.

Step 1. Obtain the crisp importance matrix of the decision scores as Eq. (4).

$$p = \begin{bmatrix} p_{11}\ p_{12}\cdots p_{1n} \\ p_{21}\ p_{22}\cdots p_{2n} \\ \cdots\ \cdots\ \cdots\ \cdots \\ p_{m1}\ p_{m2}\cdots p_{mn} \end{bmatrix} \tag{4}$$

Step 2. Normalize the decision matrix and build the weighted normalized decision matrix.

Using Eq. (5) and Eq. (6), we can convert criteria into normalization form.

$$[p_{ij}^*] = [\frac{p_{ij}}{\max_{i=1}^{m}(p_{ij})}] \tag{5}$$

$$[u_{ij}] = [w_j \times p_{ij}^*] \tag{6}$$

Where u_{ij} is called as the weighted normalized decision matrix; w_j denotes the j th criterion's weight.

Step 3. Define the positive ideal solutions (PIS) and negative ideal solutions (NIS) as Eq. (7):

$$\begin{aligned} u^+(j) &= \{ (\max_{i=1}^{m} u_{ij} | j \in B), (\min_{i=1}^{m} u_{ij} | j \in c) \} \\ u^-(j) &= \{ (\min_{i=1}^{m} u_{ij} | j \in B), (\max_{i=1}^{m} u_{ij} | j \in c) \} \end{aligned} \tag{7}$$

where $u^+(j)$ and $u^-(j)$ denotes the PIS and NIS of the criterion j, respectively. B and C are set with benefit and cost criteria, respectively.

Step 4. Compute the relative closeness of the alternative to PIS and NIS.

Applying the definition of the Euclidean distance, we can calculate the similarity for each refrigerator alternative referring to Eq. (8) and Eq. (9) from PIS and NIS.

$$d_i^+ = \sqrt{\sum_{j=1}^{n} (u_{ij} - u_j^+)^2} \tag{8}$$

$$d_i^- = \sqrt{\sum_{j=1}^{n} (u_{ij} - u_j^-)^2} \tag{9}$$

The relative closeness of the alternatives is computed using Eq. (10):

$$c_i = \frac{d_i^-}{d_i^- + d_i^+} \tag{10}$$

where $i = 1, 2, \ldots, m$

3.3. Description for experiment 3: mixed-method based on text-mining and large-scale group decision making

Ranking the alternatives based on the mixed method. The mixed-method is proposed using the online reviews to obtain weights and to apply MCDM to rank the alternatives. Participants were asked to score the criterion preference after seeing the weights acquiring by the crawled online reviews. The MCDM computed the ranking orders. Firstly, the online review data of the refrigerator is crawled. Secondly, the TextRank algorithm is applied to obtain key factors and criteria for product design decisions. Word2Vec algorithm is used to acquire the similarity degree between the key factors and the criterion. Thirdly, the multi-criterion decision methods are employed to rank the alternatives.

The data selected in this paper is from a certain e-commerce sales platform in China, one of the most popular electric appliance e-commerce platforms. Comparing with other e-commerce platforms, the selected platform focuses on electric appliance sales. The online review data includes exhaustive purchasing time, review time, scoring and detailed comments (i.e., space, service, energy consumption). Consumer reviews were mining from the comments associated with four distinct refrigerator models. These models, representing a range of configurations including single-door, double-door, and triple-door designs, are among the top-performing units on the online retail platform. The consumer reviews ranges from November 12, 2013 to November 19, 2018 were downloaded for the analysis with 32,085 comments. We pre-processed the text by deleting the invalid comments. Invalid comments include: (1) Comments such as "The consumer did not make any review in time, and the system defaults to praise!" (2) Single word is constantly repeated, such as "Good, good, good, good". This type of comment has less information but will increase the frequency of these words in the corpus. (3) The ultra-short comment is deleted. This type of comment is ultra-short and has a small amount of information. The ultra-short comments (less than 5 characters) can not reflect user experience information effectively. For instance, "good", "very good", etc. are deleted. In addition, a type of comment is also be deleted, such as "Good!!!" with more than 5 characters but less than 5 valid characters. After data pre-processing, the data length is 10191.

(1) Text processing

The text processing part aims at obtaining the key factors and criteria for the product improvement. The specific steps are as follows:

Step 1. Pre-process text.

As the text comments are not clean after being crawled, we pre-process the data as the following procedures: Firstly, each text comment was divided into several constituent units (words and sentences.). Secondly, eliminating stop words is used to clean data. There is no meaningful contribution for stop words (e.g., a, an, the, of, for,i.e) in the whole text, which must be filtered out before processing the text. Besides the regular stop words, this study also includes brand names and refrigerator names in the list.

Step 2. Use TextRank algorithm to acquire key factors.

Key factors represent the high consumers' attention on product reviews, showing high frequency and probability for consumers' comments. This paper utilizes the TextRank algorithm to extract key factors. TextRank model is originated from the PageRank algorithm. The words are assumed as "nodes on the network" for the model. The significance of each term can be calculated based on the co-occurrence relationship among each word. Eq. (11) shows the typical computational model for the TextRank algorithm.

$$TR(M_i) = (1 - t) + t \times \sum_{M_j \in In(M_i)} \frac{w_{ji}}{\sum_{M_k \in Out(M_j)} w_{jk}} TR(M_j) \tag{11}$$

Where $TR(M_i)$ is the TextRank value of node i; $In(M_i)$ represents the entry node-set pointing to node i; $Out(M_j)$ is regarded as the chain-out node-set pointed out by node j; w_{ji} is set as the weight of the edge between node j and node i, the damping factor t is selected from the range 0 to 1, and the general value is set as 0.85. For the TextRank model, every node in the graph will jump randomly with a probability of 1-t.

Step 3. Use Word2Vec to achieve the similarity degrees for key factors and each criterion.

Due to the negligence of the semantic similarity between words, the TextRank model will lead to dimensional inaccuracy and disorder. Referring the existing literature and combing with the extracted key factors, we craft the evaluation criteria framework with eight criteria (i.e., price, easy to use, appearance design, cooling effect, noise, space, energy consumption and service), which are assumed as key factors that influence product improvement. Word2Vec algorithm is applied to calculate the similarity of vectors representing each word and the selected criterion based on cosine similarity. This paper set the vector length of 200 as the layer size for the Word2Vec algorithm.

Step 4. Calculate the criteria's weights.

The equation for computing the criterion's weight is as Eq. (12):

$$w_j = \sum_{t=1}^{l} w_t \times s_{tj} \tag{12}$$

where w_t is the weights of the key factors obtained using the TextRank algorithm, and s_{tj} denotes the similarity degrees among key factors and each criterion.

(2) Rank the alternatives using the TOPSIS method, which is mentioned in Section 3.2.

4. Results

This section applies a case study on refrigerator product development to verify the proposed method that considers the online consumer reviews and group decision-making on product improvement. The group decision-making considers product designers' insights for product improvement from the professional perspective. Online reviews reveal consumers' experience and needs on the product. It is necessary to consider both these two sides.

4.1. Key factors and weights obtained by TextRank

This paper uses the TextRank algorithm to obtain key factors and weights. This study selects 30 key factors for subsequent analysis and criterion importance evaluation. The weights of the key factors are listed in Table 1. From Table 2, we can conclude that the weights of delivery service, logistics, and satisfaction are greater than other factors with 1, 0.903, and 0.842, respectively. Appearance designs such as "concise appearance" and "beauty" rank relatively low places, weighing 0.138 and 0.108, respectively.

Table 2
Key factors and weights obtained by TextRank.

key factor	weight	key factor	weight
delivery service	1	appearance	0.249
logistics	0.903	noise	0.23
satisfaction	0.842	attitude	0.213
quality	0.544	after sale	0.212
speed	0.491	install	0.21
product	0.394	cheap	0.178
price worthy	0.377	shipment	0.165
brand	0.315	practicability	0.152
cooling	0.313	delivery	0.149
trust	0.305	space	0.143
price	0.304	concise appearance	0.138
cost performance	0.303	support	0.124
service attitude	0.283	consumer service	0.123
voice	0.261	believe in	0.12
performance	0.255	beauty	0.108

4.2. Similarity degrees between key factors and each criterion obtained by Word2Vec

From the key factors results above, we classify key factors "delivery" and "logistics" into logistics criteria. The key factors results indicate that consumers prefer to employ different words for the same criterion. Thus, this paper adopts Word2Vec to calculate the similarity degrees between the obtained key factors and criteria. Table 3 presents the similarity degrees between the top 30 key factors and the eight criteria. The results show that price, cooling effect, noise, space, and service are highly related to the key factors.

4.3. Experiment processing and results

The analysis contains three experiments. A large-scale decision group consisting of 20 experts is invited to the experiments. Following the description, three experiments are processed. In the first experiment, the decision-makers are asked to rank the four alternatives with instructive perception. The expert group is asked to evaluate the criteria preference in the second experiment. The SWARA and TOPSIS methods are applied to evaluate the weights of the criteria and the ranking of the four refrigerators. In the third experiment, consumers' reviews are crawled online. After data cleaning and text processing, the weights of the selected eight criteria are computed. The results are summarized in Fig. 2. Fig. 2 indicates the refrigerator product design criterion preference and their ranks. The top three criteria online reviews determine are service, space, and price. However, cooling, price and noise are the top three criteria group decision-makers obtained.

The results of the three experiments for refrigerator alternative selection are different. Table 4 summarizes the results. From the results, we can conclude that the ranking results for Experiment 1 and Experiment 2 are similar, except Alternative A ranks fourth place and Alternative B ranks third place in Experiment 1 with intuitive decision; meanwhile, they rank third and fourth place respectively in Experiment 2. Comparing with the results between Experiment 2 and Experiment 3, the ranking results are similar, except Alternative C ranks first place and Alternative D ranks second place in Experiment 2 under the decision-making group; however, they rank second and first place in Experiment 3, which considering consumers' comments.

4.4. Deviation analysis

In this part, a deviation analysis is conducted to examine the merits of combining online reviews and group decision-making. To illustrate the efficiency of our proposed methodology, we use the deviation degree as a quantitative metric for assessing the three experiment results. The deviation degree is calculated as Eq. (13).

$$\frac{\max\left[\zeta(p_j)\right] - \zeta(p_j)}{\max\left[\zeta(p_j)\right]} \tag{13}$$

where $\zeta\left(p_j\right)$ represents the standard relevance score of the j^{th} Experiment result, and $\max\left[\zeta\left(p_j\right)\right]$ denotes the standard dominance degree of the j^{th} Experiment result.

Fig. 3 Depicts the deviations between each alternative and the reference alternative, which has the highest close efficiency for Experiments 2 and 3. Firstly, Fig. 3 shows that the deviation fluctuation trends are similar for Experiment 2 and Experiment 3. There is an increasing trend from Alternative A and Alternative B for both two experiments. The difference in the results is that Alternative C ranks the first order among the alternatives in Experiment 2 and ranks the second order in

Table 3
Similarity degree of the top 30 key factors and the criteria.

	Price	Easy to use	Appearance design	Cooling effect	Noise	Space	Energy consumption	Service
delivery service	0.758	0.325	0.397	0.703	0.718	0.729	0.266	0.769
logistics	0.755	0.306	0.425	0.681	0.723	0.764	0.281	0.785
satisfaction	0.734	0.337	0.468	0.673	0.694	0.729	0.201	0.753
quality	0.776	0.415	0.468	0.655	0.714	0.772	0.321	0.792
speed	0.747	0.333	0.420	0.669	0.677	0.742	0.279	0.784
product	0.746	0.340	0.452	0.676	0.685	0.742	0.270	0.786
price worthy	0.707	0.331	0.456	0.623	0.651	0.664	0.216	0.730
brand	0.735	0.328	0.473	0.642	0.697	0.703	0.223	0.754
cooling	0.682	0.286	0.367	1.000	0.664	0.645	0.364	0.672
trust	0.758	0.335	0.444	0.622	0.706	0.698	0.213	0.741
price	1.000	0.326	0.429	0.682	0.739	0.752	0.280	0.737
cost performance	0.783	0.413	0.483	0.672	0.743	0.748	0.332	0.745
service attitude	0.725	0.296	0.373	0.619	0.657	0.701	0.193	0.762
voice	0.720	0.292	0.407	0.581	0.780	0.671	0.295	0.719
performance	0.715	0.353	0.433	0.701	0.725	0.763	0.316	0.768
appearance	0.709	0.340	0.568	0.631	0.658	0.721	0.306	0.721
noise	0.739	0.368	0.453	0.664	1.000	0.719	0.406	0.706
attitude	0.723	0.293	0.384	0.657	0.666	0.727	0.218	0.768
after sale	0.718	0.321	0.449	0.672	0.677	0.657	0.282	0.726
install	0.722	0.328	0.465	0.691	0.694	0.703	0.298	0.701
cheap	0.721	0.292	0.437	0.635	0.707	0.682	0.249	0.717
shipment	0.701	0.283	0.377	0.580	0.632	0.673	0.195	0.682
practicability	0.677	0.336	0.404	0.659	0.687	0.720	0.250	0.710
delivery	0.603	0.245	0.355	0.551	0.595	0.651	0.208	0.624
space	0.752	0.417	0.523	0.645	0.719	1.000	0.334	0.743
concise appearance	0.671	0.310	0.502	0.605	0.647	0.712	0.277	0.657
support	0.705	0.282	0.393	0.644	0.657	0.706	0.166	0.717
consumer service	0.581	0.190	0.285	0.512	0.479	0.512	0.151	0.547
believe in	0.720	0.295	0.385	0.646	0.704	0.692	0.233	0.705
beauty	0.757	0.327	0.501	0.676	0.700	0.725	0.282	0.702

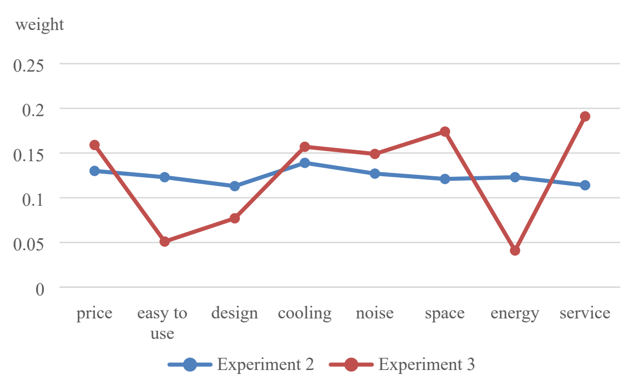

Fig. 2. Weight results of each criterion.

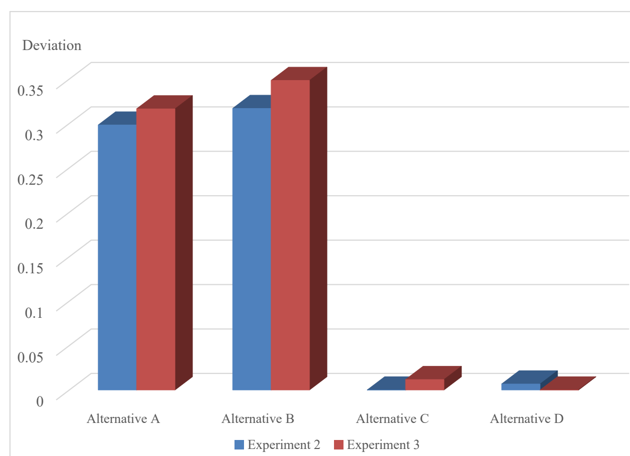

Fig. 3. Deviation analysis results.

Table 4
Results of the three Experiments.

	Experiment1	Experiment2	Score	Experiment3	Score
Alternative A	4	3	0.414	3	0.406
Alternative B	3	4	0.403	4	0.387
Alternative C	1	1	0.590	2	0.587
Alternative D	2	2	0.586	1	0.594

Experiment 3. Correspondingly, Alternative D ranks second among the alternatives in Experiment 2 and ranks first in Experiment 3. Secondly, the deviation line of Experiment 3 is higher than Experiment 2. The average deviation of Experiment 3 is 0.3165 and the average deviation of Experiment 2 is 0.2983. Thus, Experiment 3 has higher discrimination than other methods with higher average deviations. The product development process can comprehensively explore the consumers' need and expert panel's professional ideas.

Furthermore, to quantitatively underscore the advantages of our proposed methodology, we employ the Spearman's rank correlation coefficient. This coefficient is pivotal in assessing the consistency and reliability of ranking outcomes across various methodologies. The Spearman's correlation can be calculated using Eq. (14):

$$r = \frac{\mathrm{cov}(R(i), R(j))}{\sigma(R(i))^{*}\sigma(R(j))} \tag{14}$$

where $R(i)$ and $R(j)$ is the rank variable of Experiment i and Experiment j, respectively. $\mathrm{cov}(R(i), R(j))$ is the covariance of the rank variables. $\sigma(R(i))$ is the standard deviation of the rank variable.

As shown in Fig. 4, the correlation coefficient between Experiment 2 and both Experiment 1 and Experiment 3 is 0.8. The high correlation between Experiment 2 and Experiment 1 indicates that when experts rely on the Multi-Criteria Decision Making (MCDM) method for evaluation and rely on intuitive scoring, their decision outcomes show a

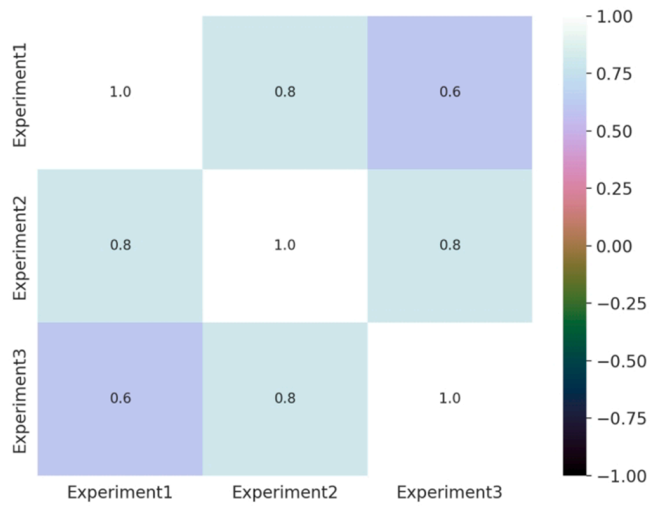

Fig. 4. The correlation coefficient values between different experiment.

certain degree of similarity, further validating the consistency of expert decision-making. The high correlation between Experiment 2 and Experiment 3 suggests a certain level of consistency in large-scale decision-making processes. However, the correlation coefficient between Experiment 1 and Experiment 3 is relatively low, at only 0.6, indicating some inconsistency between the product improvement index weights derived from customer reviews and those based solely on expert intuitive scoring. In summary, these findings suggest that results considering both customer reviews and large-scale expert decision-making are more robust.

5. Discussion

5.1. Differences in criterion weight determined by different experiments

In Experiment 2, the experts group scores the criterion preference based on alternative information background and cognition. The SWARA method is adopted to determine the criterion weight. Unlike Experiment 2, Experiment 3 uses machine learning to mine online comments and conduct textual analysis to determine criterion weight.

Table 5 shows that criteria weights are different for experts and consumers. Cooling, price, and noise are more important criteria in experts' opinion. Price takes relatively high weight among all the criteria, taking second place in Experiment 2 and third place in Experiment 3. Price is significant both for experts and consumers (Dinerstein et al., 2018). The reason for the preference is that price is closely related to sales volume. Customers prefer the product with good quality, and they are eager to spend less money to buy the products with the best property. As to the manufacturers, cost affects the final selling price and the sales volume.

Regarding the criteria measuring product performance, they take different places in these two Experiments. For example, "cooling" emerges as the paramount criterion identified by experts, yet it only secures fourth place in consumer comments. "Noise" holds third place with a weight of 0.127 and falls to fifth place with a weight of 0.149. "Space" is in the sixth position with a weight of 0.121 in Experiment 2, while it climbs to second place with a weight of 0.174 in Experiment 3. "Energy consumption" finds itself in fourth place in Experiment 2 but drops to the eighth position in Experiment 3. Regarding "service," it occupies the seventh spot with a weight of 0.114 but remarkably ascends to first place with a weight of 0.191 in the subsequent assessment. These fluctuations underscore the diverse priorities and perceptions between experts and consumers, highlighting the importance of a multifaceted approach in evaluating criteria for product development.

The results indicate that experts pay more attention to product performance and price property. Consumers pay more attention to service, space and price. Additionally, the deviation of the criteria weight obtained by experts' decision-making is smaller than weights extracted by online reviews.

Additionally, to thoroughly compare the effectiveness of various methods, we evaluated criterion weights from expert decision-making alongside weights derived from online reviews through TF-IDF and Word2Vec, as detailed in Table 5. There is a notable difference in the weights obtained via TextRank paired with Word2Vec and those acquired through the combined use of TF-IDF and Word2Vec. For instance, particularly considering the 'Price' criterion, the weight deduced by the TextRank and Word2Vec collaboration is 0.159, ranking the third position, in contrast to the 0.1571 wt from the TF-IDF and Word2Vec combination, which ranks the fifth position. This difference may be attributed to the advantages of TextRank in analyzing the contexts.

From the algorithmic perspective, TextRank excels in capturing the intricate relationships between words, thus offering a deeper insight into textual semantics. Firstly, as a graph-based sorting algorithm, TextRank visualizes text elements (like words or phrases) as interconnected nodes, with their importance gauged by the network of these connections, encompassing aspects such as co-occurrence and semantic similarity. Furthermore, TextRank's context-sensitive approach to determining word significance enables a more accurate depiction of crucial information and sentiment in text. This method's independence from predefined vocabularies or corpora also adds to its adaptability in diverse textual analyses (Bordoloi et al., 2020; Chen et al., 2020; Fakhrezi et al., 2021; Wang et al., 2020). Conversely, the TF-IDF algorithm has limitations, particularly in analyzing semantic relationships. This algorithm

Table 5
Weight of each criterion from different decision methods.

C riterion	Sub-criterion	Weights obtained from experts' decision making					
(Experiment 2)	Weights extracted by online reviews using TextRank and Word2Vec						
(Experiment 3)	Weights extracted by online reviews using TF-IDF and Word2Vec						
		Weight	Rank	Weight	Rank	Weight	Rank
Price	price	0.130	2	0.159	3	0.1571	5
Product quality	easy to use	0.123	5	0.051	7	0.0370	7
Product performance	design	0.113	8	0.077	6	0.0655	6
	cooling effect	0.139	1	0.157	4	0.1698	3
	noise	0.127	3	0.149	5	0.1702	2
	space	0.121	6	0.174	2	0.1685	4
	energy	0.123	4	0.041	8	0.0397	8
Service	service	0.114	7	0.191	1	0.1922	1

emphasizes word frequency within a set of documents, potentially neglecting the broader linguistic context. It fails to consider the interplay and sequence of words, which can lead to overlooking nuanced semantic information (Nasar et al., 2019; Yahav et al., 2018).

Regarding the application in consumer review analysis, discerning context and semantic connections are essential in consumer review analysis. TextRank demonstrates its superiority. Its ability to capture critical information makes it particularly adept at handling reviews with longer sentences, where context plays a significant role (Ramadhan et al., 2020; Zhou et al., 2022). In contrast, the TF-IDF algorithm is proficient in keyword extraction and key term identification for shorter contexts (Rehman et al., 2019). However, it falls short in grasping the full scope of consumer review. In summary, the preferential performance of TextRank in customer review analysis is substantiated by its methodological strengths over TF-IDF.

5.2. Differences in the alternative ranking for different experiments

As shown in Table 4, Alternative C is the best refrigerator selection for the first two experiments, which means experts prefer more expensive and better product performance and service. The decision-making results evaluated by MCDM are greatly consistent with experts' intuitive perception. Alternative A is the worst refrigerator for Experiment 1, and ranks the 3rd order in Experiment 2 and Experiment 3. Alternative B is the third selection rank refrigerator for Experiment 1, and ranks the 4th order in Experiment 2 and Experiment 3. Alternative D is the 2nd selection rank refrigerator for Experiment 1 and Experiment 2, and ranks the 4th order in Experiment 3.

Ranking results are different among the three experiments. Experiment 1, based on the overall intuition of experts, can not figure out and improve the specific criteria of product improvement. Remarkably, the results of Experiment 2 and Experiment 3 are different. A comparative analysis was implemented by adopting the deviation analysis to test the accuracy of decision-making results. The comparative analysis shows that the results of Experiment 3 have a significant deviation. Thus, the results made by Experiment 3 are more reasonable and authentic. Consistent with the previous literature, the experimental method was applied to determine whether consumers were interested in participating in co-creation of product improvement (Khrystoforova & Siemeniako; 2019). The main reasons for the differences probably are the weights variation between these two decision-making methods.

Table 4 indicates that the expert group in Experiment 2 sets cooling effect, price and noise as more significant elements in product improvement. However, from the weights shown in Table 4, we can conclude that consumers pay more attention to service, space and price from their online reviews. The different preference demonstrates that expert group consumers' focus points differ. Experts pay more attention to function issues, while consumers pay more attention to user experience.

6. Conclusion

This paper has investigated the effectiveness of different decision-making methodologies for product improvement in a big data environment. We have conducted three experiments to compare the performance of intuitive perception, expert decision-making, and a hybrid method that combines consumer big data and large-group decision-making. We have used text-mining techniques like LDA, TextRank, and Word2Vec to extract and analyze the product criteria from online reviews. We have found that the hybrid method outperforms the other two methods regarding accuracy and reliability. The hybrid method can capture consumers' diverse and dynamic preferences and opinions, and incorporate them into the product improvement process. The hybrid method can also reduce the uncertainty and complexity of decision-making, and provide more scientific and objective results. The paper contributes to the literature on product improvement by proposing a novel mixed model that integrates online reviews and experts' opinions. The paper also provides practical implications for manufacturers who want to improve their products based on consumer feedback. The paper suggests manufacturers should leverage big data and large-group decision-making to enhance product quality and competitiveness.

6.1. Implications

This paper has explored the impact of consumer involvement on product improvement and decision-making. It has conducted a series of experiments to compare the performance of different decision-making methodologies, using refrigerators as a case study. It has also applied text-mining techniques like LDA, TextRank, and Word2Vec to extract and analyze product criteria from online reviews. The paper has several implications for both theory and practice.

First, this paper contributes to the literature on product improvement by proposing a novel mixed model that integrates online reviews and experts' opinions. The paper demonstrates that online reviews are a valuable source of information for understanding consumers' needs and preferences (Wamba et al., 2015). The paper also shows that the product criteria should cover the important issues raised by both consumers and experts (Hou et al., 2019).

Second, this paper provides practical guidance for manufacturers who want to improve their products based on consumer feedback. The paper suggests that manufacturers should leverage big data and large-group decision-making to enhance product quality and competitiveness. The paper shows that the hybrid method involving consumers' opinions leads to more accurate and reliable outcomes than intuitive perception or expert decision-making.

Third, this paper advances the methodology of product improvement by using text-mining techniques to extract and analyze the product criteria from online reviews. The paper adopts the LDA method to categorize the topics of online reviews, and combines them with the criteria derived from the existing literature (Chan et al., 2016; Hou et al., 2019). The paper also uses TextRank and Word2Vec algorithms to determine the importance and correlation of product criteria. The paper demonstrates that these techniques can reduce the complexity and subjectivity of decision-making, and provide more objective and accurate results.

6.2. Limitations and future directions

This paper has contributed to the research on product improvement and decision-making methods, but it also has some limitations that can be addressed in future studies. First, this paper does not consider the different characteristics of consumers (e.g., family size, and age, etc.). Consumers with different characteristics may have diverse preferences for product improvement. It is important to tailor the product development to the target consumers. Satisfying different consumers' needs can enhance the competitiveness of enterprises. Second, the natural language processing methods used in this paper require a large amount of online reviews. The practicality and effectiveness of the approach need to be further verified when online reviews are scarce. Third, the consumer sentiments expressed in online reviews should be taken into account in future studies.

CRediT authorship contribution statement

Fang Zhang: Data curation, Methodology, Writing – original draft. **Wenyan Song:** Conceptualization, Methodology, Resources, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This research has received funding from the National Natural Science Foundation of China (71971012) and the Fundamental Research Funds for the Central Universities (YWF-23-JT-103).This research was also supported in part by the Key project of Aeroengine and Gas Turbine Basic Science Center (P2023-B-I-005-001)

References

Asadabadi, M. R., Saberi, M., Sadghiani, N. S., Zwikael, O., & Chang, E. (2023). Enhancing the analysis of online product reviews to support product improvement: Integrating text mining with quality function deployment. *Journal of Enterprise Information Management, 36*(1), 275–302.

Balakrishnan, V., Shi, Z., Law, C. L., Lim, R., Teh, L. L., & Fan, Y. (2022). A deep learning approach in predicting products' sentiment ratings: A comparative analysis. *The Journal of Supercomputing, 78*(5), 7206–7226.

Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications, 127*, 256–271.

Beugelsdijk, S., Nell, P. C., & Ambos, B. (2017). When do distance effects become empirically observable? An investigation in the context of headquarters value creation for subsidiaries. *Journal of International Management, 23*(3), 255–267.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research, 3*(Jan), 993–1022.

Bordoloi, M., Chatterjee, P. C., Biswas, S. K., & Purkayastha, B. (2020). Keyword extraction using supervised cumulative TextRank. *Multimedia Tools and Applications, 79*(41–42), 31467–31496.

Burnap, P., & Williams, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science, 5*(1), 11.

Chan, H. K., Wang, X., Lacka, E., & Zhang, M. (2016). A mixed-method approach to extracting the value of social media data. *Production and Operations Management, 25*(3), 568–583.

Chen, Z.-S., Liu, X.-L., Chin, K.-S., Pedrycz, W., Tsui, K.-L., & Skibniewski, M. J. (2020). Online-review analysis based large-scale group decision-making for determining passenger demands and evaluating passenger satisfaction: Case study of high-speed rail system in China. *Information Fusion.*

Chen, Z. Y., Liu, X. L., & Yin, L. P. (2023). Data-driven product configuration improvement and product line restructuring with text mining and multitask learning. *Journal of Intelligent Manufacturing, 34*(4), 2043–2059.

Chien, C.-F., Kerh, R., Lin, K.-Y., & Yu, A.-P.-I. (2016). Data-driven innovation to capture user-experience product design: An empirical study for notebook visual aesthetics design. *Computers & Industrial Engineering, 99*, 162–173.

Chiu, M.-C., & Kremer, G. E. O. (2013). An investigation on centralized and decentralized supply chain scenarios at the product design stage to increase performance. *IEEE Transactions on Engineering Management, 61*(1), 114–128.

Delgado-Solano, I. P., Núñez-Varela, A. S., & Pérez-González, G. H. (2018, October). Keyword Extraction From Users' Requirements Using TextRank and Frequency Analysis, and Their Classification into ISO/IEC 25000 Quality Categories. In 2018 6th International Conference in Software Engineering Research and Innovation (CONISOFT) (pp. 88-92). IEEE.

Dinerstein, M., Einav, L., Levin, J., & Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review, 108*(7), 1820–1859.

Ding, X., Gao, B., & Liu, S. (2022). Understanding the interplay between online reviews and growth of independent and branded hotels. *Decision Support Systems, 152*, Article 113649.

Ding, R.-X., Wang, X., Shang, K., Liu, B., & Herrera, F. (2018). Sparse representation-based intuitionistic fuzzy clustering approach to find the group intra-relations and group leaders for large-scale decision making. *IEEE Transactions on Fuzzy Systems, 27*(3), 559–573.

Fakhrezi, M. F., Bijaksana, M. A., & Huda, A. F. (2021). Implementation of automatic text summarization with TextRank method in the development of Al-qur'an vocabulary encyclopedia. *Procedia Computer Science, 179*, 391–398.

Goldberg, D. M., & Abrahams, A. S. (2022). Sourcing product innovation intelligence from online review. *Decision Support Systems, 157*, Article 113751.

González, F. J. M., & Palacios, T. M. B. (2002). The effect of new product development techniques on new product success in Spanish firms. *Industrial Marketing Management, 31*(3), 261–271.

Guo, Y., Barnes, S. J., & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management, 59*, 467–483.

Haddad, B., Liazid, A., & Ferreira, P. (2017). A multi-criteria approach to rank renewables for the Algerian electricity system. *Renewable Energy, 107*, 462–472.

Hmida, A., Chekir, N., Laafer, A., Slimani, M. E. A., & Brahim, A. B. (2019). Modeling of cold room driven by an absorption refrigerator in the south of Tunisia: A detailed energy and thermodynamic analysis. *Journal of Cleaner Production, 211*, 1239–1249.

Horvat, A., Granato, G., Fogliano, V., & Luning, P. A. (2019). Understanding consumer data use in new product development and the product life cycle in European food firms–An empirical study. *Food Quality and Preference, 76*, 20–32.

Hou, T., Yannou, B., Leroy, Y., & Poirson, E. (2019). Mining changes in user expectation over time from online reviews. *Journal of Mechanical Design, 141*(9), Article 091102.

Huang, S., Zhang, J., Yang, C., et al. (2022). The interval grey QFD method for new product development: Integrate with LDA topic model to analyze online reviews. *Engineering Applications of Artificial Intelligence, 114*, Article 105213.

Ireland, R., & Liu, A. (2018). Application of data analytics for product design: Sentiment analysis of online product reviews. *CIRP Journal of Manufacturing Science and Technology, 23*, 128–144.

Ishizaka, A., & Siraj, S. (2018). Are multi-criteria decision-making tools useful? An experimental comparative study of three methods. *European Journal of Operational Research, 264*(2), 462–471.

Işıklar, G., & Büyüközkan, G. (2007). Using a multi-criteria decision making approach to evaluate mobile phone alternatives. *Computer Standards & Interfaces, 29*(2), 265–274.

Jain, P. K., Pamula, R., & Srivastava, G. (2021). A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews. *Computer Science Review, 41*, Article 100413.

Jatnika, D., Bijaksana, M. A., & Suryani, A. A. (2019). Word2vec model analysis for semantic similarities in English words. *Procedia Computer Science, 157*, 160–167.

Jeong, B., Yoon, J., & Lee, J. M. (2019). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management, 48*, 280–290.

Ji, F., Cao, Q., Li, H., et al. (2023). An online reviews-driven large-scale group decision making approach for evaluating user satisfaction of sharing accommodation. *Expert Systems with Applications, 213*, Article 118875.

Joung, J., & Kim, H. (2023). Interpretable machine learning-based approach for customer segmentation for new product development from online product reviews. *International Journal of Information Management, 70*, Article 102641.

Keršuliene, V., Zavadskas, E. K., & Turskis, Z. (2010). Selection of rational dispute resolution method by applying new step-wise weight assessment ratio analysis (SWARA). *Journal of Business Economics and Management, 11*(2), 243–258.

Khrystoforova, V., & Siemeniako, D. (2019). Internet-based consumer co-creation experience of the new product development process. *Engineering Management in Production and Services, 11*(3), 60–68.

Liu, H.-T. (2011). Product design and selection using fuzzy QFD and fuzzy MCDM approaches. *Applied Mathematical Modelling, 35*(1), 482–496.

Liu, B., Shen, Y., Chen, X., Chen, Y., & Wang, X. (2014). A partial binary tree DEA-DA cyclic classification model for decision makers in complex multi-attribute large-group interval-valued intuitionistic fuzzy decision-making problems. *Information Fusion, 18*, 119–130.

Liu, Y., Soroka, A., Han, L., Jian, J., & Tang, M. (2020). Cloud-based big data analytics for customer insight-driven design innovation in SMEs. *International Journal of Information Management, 51*, Article 102034.

Mihalcea, R., & Tarau, P. (2004). Textrank: Bringing order into text. Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 404–411.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems, 26*, 3111–3119.

Nasar, Z., Jaffry, S. W., & Malik, M. K. (2019). Textual keyword extraction and summarization: State-of-the-art. *Information Processing & Management, 56*(6), Article 102088.

Qin, J., & Zeng, M. (2022). An integrated method for product ranking through online reviews based on evidential reasoning theory and stochastic dominance. *Information Sciences, 612*, 37–61.

Raghupathi, D., Yannou, B., Farel, R., & Poirson, E. (2015). Customer sentiment appraisal from user-generated product reviews: A domain independent heuristic algorithm. *International Journal on Interactive Design and Manufacturing (IJIDeM), 9*(3), 201–211.

Rahman, S. M. A., Hachicha, A. A., Ghenai, C., Saidur, R., & Said, Z. (2020). Performance and life cycle analysis of a novel portable solar thermoelectric refrigerator. *Case Studies in Thermal Engineering, 19*, Article 100599.

Ramadhan, M. R., Endah, S. N., & Mantau, A. B. J. (2020). Implementation of textrank algorithm in product review summarization. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 1–5). IEEE.

Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowledge-Based Systems, 89*, 14–46.

Razzaghnoori, M., Sajedi, H., & Jazani, I. K. (2018). Question classification in Persian using word vectors and frequencies. *Cognitive Systems Research, 47*, 16–27.

Rehman, A. U., Malik, A. K., Raza, B., & Ali, W. (2019). A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications, 78*, 26597–26613.

Saaty, T. L., & Shang, J. S. (2007). Group decision-making: Head-count versus intensity of preference. *Socio-Economic Planning Sciences, 41*(1), 22–37.

Sinha, A. K., & Anand, A. (2018). Development of sustainable supplier selection index for new product development using multi criteria decision making. *Journal of Cleaner Production, 197*, 1587–1596.

Stanujkic, D., Karabasevic, D., & Zavadskas, E. K. (2015). A framework for the selection of a packaging design based on the SWARA method. *Engineering Economics, 26*(2), 181–187.

Tan, H., Lv, X., Liu, X., & Gursoy, D. (2018). Evaluation nudge: Effect of evaluation mode of online customer reviews on consumers' preferences. *Tourism Management, 65*, 29–40.

Tang, F., Fu, L., Yao, B., & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences, 488*, 190–204.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., & Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology, 94*(9–12), 3563–3576.

Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., et al. (2019). Digital twin-driven product design framework. *International Journal of Production Research, 57*(12), 3935–3953.

Tuarob, S., & Tucker, C. S. (2015). Quantifying product favorability and extracting notable product features using large scale social media data. *Journal of Computing and Information Science in Engineering, 15*(3).

Ventre, I., & Kolbe, D. (2020). The impact of perceived usefulness of online reviews, trust and perceived risk on online purchase intention in emerging markets: A Mexican perspective. *Journal of International Consumer Marketing, 32*(4), 287–299.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165*, 234–246.

Wang, L., Wang, X., Peng, J., & Wang, J. (2020). The differences in hotel selection among various types of travellers: A comparative analysis with a useful bounded rationality behavioural decision support model. *Tourism Management, 76*, Article 103961.

Wang, P., Xu, X., Huang, S., & Cai, C. (2018). A linguistic large group decision making method based on the cloud model. *IEEE Transactions on Fuzzy Systems, 26*(6), 3314–3326.

Wu, J., Liu, C., Wu, Y., Cao, M., & Liu, Y. (2022). A novel hotel selection decision support model based on the online reviews from opinion leaders by best worst method. *International Journal of Computational Intelligence Systems, 15*(1), 19.

Xu, X., Du, Z., & Chen, X. (2015). Consensus model for multi-criteria large-group emergency decision making considering non-cooperative behaviors and minority opinions. *Decision Support Systems, 79*, 150–160.

Xu, Z., Frankwick, G. L., & Ramirez, E. (2016). Effects of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research, 69*(5), 1562–1566.

Xu, C., Park, J., & Lee, J. C. (2022). The effect of shopping channel (online vs offline) on consumer decision process and firm's marketing strategy. *Internet Research, 32*(3), 971–987.

Yahav, I., Shehory, O., & Schwartz, D. (2018). Comments mining with TF-IDF: The inherent bias and its removal. *IEEE Transactions on Knowledge and Data Engineering, 31*(3), 437–450.

Yang, Y., Fu, C., Chen, Y.-W., Xu, D.-L., & Yang, S.-L. (2016). A belief rule based expert system for predicting consumer preference in new product development. *Knowledge-Based Systems, 94*, 105–113.

Yu, S., Wang, J., Wang, J., & Li, L. (2018). A multi-criteria decision-making model for hotel selection with linguistic distribution assessments. *Applied Soft Computing, 67*, 741–755.

Zhang, Y., Chen, F., Zhang, W., Zuo, H., & Yu, F. (2020). April). Keywords extraction based on word2vec and textrank. In *Proceedings of the 2020 The 3rd International Conference on Big Data and Education* (pp. 37–42).

Zhang, L., Chu, X., & Xue, D. (2019). Identification of the to-be-improved product features based on online reviews for product redesign. *International Journal of Production Research, 57*(8), 2464–2479.

Zhou, N., Shi, W., Liang, R., & Zhong, N. (2022). Textrank keyword extraction algorithm using word vector clustering based on rough data-deduction. Computational Intelligence and Neuroscience, 2022.

Zhou, F., Jiao, J. R., Schaefer, D., & Chen, S. (2010). Hybrid association mining and refinement for affective mapping in emotional design. *Journal of Computing and Information Science in Engineering, 10*(3).

Zolfani, S. H., & Saparauskas, J. (2013). New application of SWARA method in prioritizing sustainability assessment indicators of energy system. *Engineering Economics, 24*(5), 408–414.

Zuo, X., Zhang, S., & Xia, J. (2017). The enhancement of TextRank algorithm by using word2vec and its application on topic extraction. *Journal of Physics: Conference Series, 887*, 12028.