ISSN: 2029-4913 / eISSN: 2029-4921

VILNIUS TECH
Vilnius Gediminas Technical University

TECHNOLOGICAL and ECONOMIC DEVELOPMENT of ECONOMY

2025
Volume 31
Issue 5
Pages 1291–1319
https://doi.org/10.3846/tede.2025.23838

EXPLORING SERVICE IMPROVEMENT THROUGH IMPORTANCE-PERFORMANCE ANALYSIS CONSIDERING THE RELIABILITY OF MULTIPLE ONLINE PLATFORMS

Shanshan YANG [1], Huchang LIAO [2]✉, Chonghui ZHANG [2]

[1] *Business School, Sichuan University, Chengdu, China*
[2] *College of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China*

Article History:
- received 03 July 2024
- accepted 17 March 2025

Abstract. Service improvement has emerged as a pivotal task for hoteliers to ensure competitive advantage. This study proposes a service improvement method based on online reviews from multiple platforms considering the reliability of online platforms and different evaluation modes, where the reliability of an online platform is defined based on the number of online reviews on that platform and the degree of review helpfulness. In our method, Latent Dirichlet Allocation model is utilized to extract keywords, and lexicon-based sentiment analysis methods are employed to analyze the sentiment of online reviews on each platform considering different evaluation modes. The importance of attributes on each platform is measured by the TextRank method. A multi-platform-oriented importance-performance analysis model is constructed based on the integrated performance and the importance of attributes, so as to classify attributes and formulate service improvement strategies. A case study about hotel service improvement is implemented to illustrate the effectiveness of the method. Results show that the attributes classification results considering the reliability of multiple platforms is more reasonable compared to the results based on a single platform, providing more effective service improvement strategy and clearer view of attribute status on various platforms for hoteliers.

Keywords: service improvement strategies, multiple online platforms, importance-performance analysis, sentiment analysis, probabilistic linguistic term set.

JEL Classification: C10, C60, C88, M31, Z30.

✉Corresponding author. E-mail: ***liaohuchang@163.com***

1. Introduction

Service improvement strategies are critical to enhance enterprise competitiveness in a challenging market environment (Zhang & Xu, 2024). Importance-performance analysis (IPA) is a useful tool to provide insights into making strategic decisions (Salimi, 2021; Boley & Jordan, 2023). With the development of social media, online reviews are favored by both consumers and hoteliers owing to their characteristic of low-cost availability and real-time updates. They serve as an indispensable reference for consumers to make purchase decisions, as well as a promising data source for hoteliers to optimize services and improve customer satisfaction (Lu et al., 2023; Zhang et al., 2023; Zhao et al., 2024). By collecting and analyzing evaluation information from online platforms, hoteliers can obtain comprehensive user feedback, thereby accurately grasping the advantages and disadvantages of products or services, and providing a strong basis for the formulation of service improvement strategies. Due to the

competitive market environment, hoteliers often provide hotel ordering channels on multiple platforms to increase customer groups and avoid peer competition, generating a large amount of review data on different platforms. Service improvement based on the IPA model under the online environment has been researched a lot (Bi et al., 2019; Albayrak et al., 2021; Wu et al., 2023b), while few research focused on service improvement based on multiple platforms. Actually, service improvement based on a single platform may lack data comprehensiveness and result in sampling bias, limiting the rationalization of findings. In this sense, service improvement based on multiple platforms needs further investigation.

The reliability of platforms is a critical factor under the online environment inundated with fake reviews (Verma et al., 2023). Platform reliability refers to the extent to which customers perceive reviews on a platform as actual and credible. Platforms with low reliability attracts fewer customers, and the reliability of knowledge extracted from these platforms is correspondingly low. Differences in reliability across platforms arise from factors such as the numbers of reviews, evaluation mechanisms, and target customers. A straightforward approach to assess platform reliability involves using features extracting from online reviews, such as the number of online reviews. Wu et al. (2024b) employed cohesion degree of platforms and the percentage of reviews containing features to the total number of reviews to determine the weights of platforms. Zhao et al. (2021) determined platform weight using the number of evaluators. However, these approaches fail to account for the critical aspect of review helpfulness, which reflect customers' approval of reviews (Chen et al., 2024) and plays a significant role in shaping a platform's reputation. Neglecting review helpfulness in reliability assessments can result in incomplete evaluations and potentially undermine the accuracy of the results. Thus, *how to quantitatively measure the reliability of different platforms considering review helpfulness* is the first research question of this study.

Detailed reviews and *pros/cons* reviews are two common evaluation modes on online platforms. The former provide an overall evaluation of a product/service, including consumer's experience and both positive and negative opinions across various attributes. In contrast, the latter separates positive and negative evaluations into distinct sections. Existing studies on multiple platforms (Kou et al., 2021; Wu et al., 2024b) focused mainly on the mode of *detailed* reviews, the *pros/cons* reviews have been largely overlooked. It is necessary to consider *pros/cons* review mode as it reflects the difference and diversity of online platforms. Differences in evaluation modes lead to different strategies for analyzing the sentiment of online reviews. *Pros/cons* evaluation mode is more likely to include phrases and individual words, such as "positive: food, location". Although "food" and "location" are not linked to sentiment words, they inherit positive sentiment from being listed in the positive category (Mirtalaie et al., 2018). Thus, *pros/cons* reviews require a separate analysis for positive or negative evaluations with a distinct method compared with the *detailed* reviews. Furthermore, owning to the ambiguity associated with consumer evaluation, recent studies (Darko et al., 2023; Liu et al., 2023b) have utilized fuzzy sets to portray complex sentiment information. To develop effective service improvement strategies, it is necessary to aggregate fuzzy sentiment information mining from multiple platforms. Existing studies rarely address the challenges of aggregating sentiment information characterized by fuzzy sets across platforms with varying evaluation modes. Thus, the second research question of this study is, *how to analyze the sentiments of*

online reviews considering different evaluation modes and aggregate the sentiment information characterized by fuzzy sets mining from multiple platforms.

Compared to questionnaires, online reviews facilitate the data collection process for hoteliers to analyze costumer demands. Different online platforms may gather reviews from diverse user groups with varying backgrounds and needs. Considering online reviews from multiple platforms ensures a broader range of feedback, which further improves the robustness of service improvement strategies. IPA (Martilla & James, 1977) is a business analysis tool to identify attributes that require more attention or consume excessive resources based on the performance and importance of attributes. Its low-cost and easily-understanding make it a critical instrument for providing insights in developing service improvement strategies. Existing IPA models (Glaveli et al., 2023; Wu et al., 2023b) mainly rely on online reviews from a single platform to mine consumer preferences and classify attributes. Focusing on a single platform limits the generalizability of the results, as it overlooks the diversity of consumer opinions and feedback available across multiple platforms. Although Kou et al. (2021) proposed a market structure analysis method based on multiple platforms for product-attribute improvement, which determined the priority of attribute improvement by computing the relative distance between each attribute and the ideal attribute. However, this method can only provide attribute ordering but cannot identify the property of attributes. The IPA model classifies attributes into four categories, with each category adopting distinct strategies for service improvement. Overall, constructing IPA model based on reviews from multiple platforms remains unexplored. Thus, the third research question is *how to construct an IPA model based on the performance and importance of attributes mining from online reviewers in multiple platforms*.

For the first research question, the reliability of online platforms is measured based on the number of online reviews and the degree of review helpfulness. A critical factor influencing review helpfulness is the length of online reviews (Ganguly et al., 2024). Review helpfulness generally increases with review length within a certain range, as longer reviews provide consumers with more detailed information about a product/service. However, once the length of a review exceeds a certain point, it causes information overload and review helpfulness decreases as the length of the review increases. An inverted U-shaped relationship exists between the review length and review helpfulness (Li & Huang, 2020). In this paper, an inverted U-shape function is proposed to estimate the effect of review length on review helpfulness. Additionally, customers are prone to trust a review that is consistent with other reviews on the same platform (Jia et al., 2022). In other words, the consistency between reviews or ratings also reflects review helpfulness. This study determines the consistency of reviews by calculating the semantic similarity of reviews to other reviews, and evaluates the consistency of ratings by measuring the distance of ratings to the average rating. Overall, the review helpfulness is measured by the consistency and the length of reviews, and is combined with the number of reviews by an aggregation formula to calculate the reliability of platforms.

For the second challenge, a framework that integrates the LDA model for keyword extraction, lexicon-based methods for sentiment analysis tailored to different review modes, and probabilistic linguistic term sets (PLTSs; Pang et al., 2016) for characterizing sentiment analysis results is developed to obtain the overall performance of attributes from multiple platforms.

For *detailed* reviews, the Vader method is utilized to analyze the sentiment polarity of sentence containing attributes, and then the degree adverbs are extracted to determine the sentiment strength. For *pros/cons* reviews, the sentiment polarity is identified according to the *pros/cons* categories to which the reviews belong, and the sentiment strength is determined in the same way as *detailed* reviews. In this study, sentiment analysis results are represented in PLTSs[1], where linguistic terms represent sentiment polarities of attributes and probabilities correspond to the percentage of occurrence of different sentiment polarities. The expectation function of PLTSs (Wu & Liao, 2019) is employed to transform PLTSs to crisp scores. Then, the scores corresponding to different platforms are aggregated using the reliability of platforms to obtain the overall performance of attributes.

With respect to the third challenge, a multi-platform-oriented IPA model is constructed to classify attributes based on the overall performance and the overall importance of attributes. The overall performance of attributes is obtained by the aforementioned procedure. The importance of attributes on each platform is derived by the TextRank method, and then integrated with the reliability of platforms to obtain the overall importance of attributes. The boundary points are determined by calculating the separate averages of the overall performance and the overall importance of all attributes. According to the boundary points, the attributes are classified into four categories including "keep up the good work", "concentrate here", "low priority", and "possible overkill". Different from the IPA model based on a single platform, this approach integrates performance and importance across multiple platforms, offering a more comprehensive and robust classification.

In summary, this study dedicates to proposing a service improvement method based on the multi-platform-oriented IPA model considering different evaluation modes in multiple platforms and the reliability of different platforms. The contributions of the paper include:

(1) A reliability determination method is proposed to aggregate the performance and importance of attributes from different online platforms. Unlike existing approaches, this method incorporates both the number of online reviews and a comprehensive measure of review helpfulness, which accounts for the inverted U-shaped relationship between review length and helpfulness, as well as the consistency among reviews and the consistency among rating. This approach enhances the depth and comprehensiveness of reliability assessment, providing a nuanced understanding of platform reliability.

(2) A framework is proposed to analyze and aggregate sentiments from platforms with different evaluation modes. The framework incorporates both detailed reviews and *pros/cons* reviews by applying tailored sentiment analysis methods for each mode. It aggregates sentiments represented by PLTSs with the degrees of platform reliability. This comprehensive framework provides a unique solution for evaluating the overall performance of attributes across multiple platforms with varying evaluation modes.

(3) A multi-platform-oriented IPA model is constructed to classify attributes according to their overall performance and importance. Unlike previous IPA models that rely on data from a single platform, this model integrates data from multiple platforms,

[1] PLTS is a representation model for complex linguistic information, which can reflect ambiguity and hesitation in evaluation information.

considering the structural diversity of platforms and their reliability. The proposed model offers a more reasonable and realistic basis for formulating service improvement strategies than existing ones.

The rest of this paper is organized as follows: Section 2 reviews relevant studies, including the reliability measurement of different platforms, information aggregation from multiple platforms, and service improvement based on online reviews. Section 3 demonstrates the proposed method. Section 4 validates the effectiveness of the proposed method and provides insights. The last section concludes this study.

2. Literature review

2.1. Service improvement based on online reviews

Service improvement plays a significant role in enhancing customer satisfaction, improving service quality, and optimizing resource allocation. Service improvement methods, such as the IPA and Kano model, are essential for hoteliers to develop service improvement strategies in a competitive environment. Table 1 displays related studies about service improvement based on online reviews.

Keyword extraction and sentiment analysis are two key steps to achieve service improvement. The Latent Dirichlet Allocation (LDA) model is a popular method to extract topics from online reviews (Shin et al., 2024). Sentiment analysis methods can be categorized into machine learning methods and lexicon-based methods. Lexicon-based sentiment analysis methods focus on determining the sentiment of texts using a predefined set of words and their associated sentiment values. Machine learning methods involve training models on labeled

Table 1. Service improvement based on online reviews

Reference	Keyword extraction method	Sentiment analysis method	Service evaluation method	Data source
Luo et al. (2021)	LDA	Support vector machine	IPA model	Baidu travel, Ctrip travel, Tongcheng travel, Qunar
Zhang et al. (2022a)	LDA	Convolutional neural network	Improved Kano model	Amazon, TMall
Liu et al. (2023a)	LDA	Sentiment lexicon	Kano model	Amazon
Zhang et al. (2023)	LDA	HowNet	Importance-Kano model	Tmall.com
Wu and Yang (2023)	K-means clustering	PaddlePaddle NLP tool	IPA model	Ctrip
Ma et al. (2024)	BERTopic	Bert-base-multilingual-uncased-sentiment model	IPA model	Google Play
Wu et al. (2024a)	LDA	Dictionary-based sentiment analysis model, convolutional neural network-long short-term memory model, and large language model	IPA model	Ctrip

datasets to classify text sentiment. Compared to machine learning methods, lexicon-based methods have the advantages of simplicity and interpretability, and no need for labeled data.

The IPA model classifies attributes by quantitatively estimating the performance and importance of attributes. Due to its intuition and simplicity, the IPA model has been applied in the hotel industry (Quan et al., 2022), restaurant domain (Mejia et al., 2022) and finance industry (Ban et al., 2022). This study utilizes the IPA model to identify the priority of attributes for service improvement. Although many studies (Luo et al., 2021; Zhang et al., 2023; Wu & Yang, 2023) have been carried out on service improvement through online reviews, most were centered on a single platform and failed to account for multiple platforms and platform reliability.

2.2. Reliability measurement of different platforms

Some online platforms offer limited information such as reviews and ratings, while others provide additional details, such as evaluator-specific information. Service improvement based on multiple online platforms considers the requirements of broader customers on different platforms and minimizes the risk of overlooking critical feedback, thus generating a more effective service improvement strategy compared to service improvement based on a single platform. Given the inconsistent information provided by different platforms, the point of measuring the reliability of different online platforms lies in discovering common factors across platforms. The reliability of an online platform serves as a weight to integrate customer sentiments from different platforms.

Table 2 lists studies about reliability measurement of multiple platforms. As can be seen in Table 2, factors such as the number of evaluators (Zhao et al., 2021), the cohesion of platforms (Wu et al., 2024b), and the information entropy of platforms (Wu et al., 2023a) have been applied to measure platform reliability. The sentiments extracted from multiple platforms is also a source to determine the reliability of platforms (Kou et al., 2021). How-

Table 2. Studies related to reliability measurement of multiple platforms

	Determining platform weights	Application
Kou et al. (2021)	Take each platform as an expert and utilize the projection method to determine the weight of each platform based on the decision matrix obtained from the sentiment of attributes	Provide a market structure analysis method across multiple platforms using multi-attribute group decision-making methods
Zhao et al. (2021)	Measure the weights of platforms using the number of evaluators	Select hotels based on online reviews and ratings from multiple platforms
Wu et al. (2023a)	Information entropy and the number of participants involved in the platform	Rank large-scale alternatives from multiple platforms
Wu et al. (2024b)	Percentage of reviews containing features and the cohesion degree of platforms	Rank products on multiple platforms based on large-scale group decision-making methods
Yang et al. (2024)	Calculate the weights of platforms by analytic hierarchy process method with the evaluation of users and experts	Propose a cross-platform online ratings aggregation approach for decision making.

ever, these methods primarily focused on the quantity or distribution of platform data while overlooking the quality of the reviews. By accounting for factors including review length and consistency in this paper, review helpfulness offers a comprehensive evaluation of the content quality of a platform, thereby making the calculation of platform weights more scientific and reasonable. In addition, most of the existing studies on multiple platforms (Zhao et al., 2021; Wu et al., 2023a; Wu et al., 2024b) focused on selecting or ranking alternatives from consumer's perspective, rather than developing service improvement strategies from hotelier's perspective. This study fills this gap by exploring service improvement based on multiple platforms considering the reliability of online platforms.

2.3. Aggregating information from multiple platforms

Integrating the evaluation information from multiple online platforms is beneficial for obtaining an overall understanding of consumer preferences, which provides support for service improvement based on multiple platforms.

Table 3 lists relevant studies about information aggregation from multiple platforms. As can be seen in Table 3, some studies utilized fuzzy sets to depict the sentiment of attributes and employed aggregation operators (Yang et al., 2020), Choquet integral (Zhang et al., 2022b) or optimization models (Liang, 2024) to aggregate sentiments from multiple platforms, so as to provide support for decision making. The applications are mainly concentrated on product ranking (Yang et al., 2020) and evaluation (Zhang et al., 2022b). However, they seldom considered different evaluation modes of multiple platforms. This paper utilizes PLTSs to represent the sentiments of attributes on each platform, employs the expectation function of PLTSs (Wu & Liao, 2019) to transform PLTSs to precise scores, and then aggregates the scores with the reliability of different platforms. Different evaluation modes including *detailed* and *pros/cons* modes are considered in integrating the sentiment of attributes.

Table 3. Information aggregation from multiple platforms

	Aggregation method	Application
Yang et al. (2020)	Utilize an aggregation operator to aggregate sentiments from multiple platforms	Rank mobile phones by a score function of fuzzy sets to provide references for consumers' purchase
Zhang et al. (2022b)	Integrate information from multiple websites by the probabilistic linguistic Choquet integral	Develop a hotel evaluation model
Yang et al. (2024)	Aggregate cross-platform distributed ratings using the basic uncertain linguistic information-aggregation function	Propose a cross-platform online ratings aggregation approach for ranking vehicles
Liang (2024)	Aggregate utility values from multiple platforms using a weighted average maximization model	Propose an incentive mechanism to address the collaboration across multiple platforms in crowdsourcing task allocation

3. An importance-performance analysis model for service improvement based on online reviews from multiple platforms

This study dedicates to proposing a service improvement method based on multiple online platforms from hotelier's perspective. For hoteliers, leveraging online reviews from various platforms facilitates comprehensive coverage of customer needs, thereby formulating effective service improvement strategies. The quality of online reviews from different platforms varies due to platform differences, thus necessitating the measurement of platform reliability to distinguish data utilization. Given the diverse evaluation models across platforms, different sentiment analysis methods should be employed when analyzing sentiments of reviews from different platforms. In addition, integrating sentiments from different platforms and using service evaluation models to classify attributes based on the integrated sentiments is a key step for service improvement based on multiple platforms.

Suppose that hoteliers provide consumers with hotel ordering channels on multiple platforms, denoted as $\{A_1, A_2, \ldots, A_m\}$. Consumers evaluate hotels with respect to a set of attributes $\{C_1, C_2, \ldots, C_n\}$. The reliability of online platforms is represented as $\{W_1, W_2, \ldots, W_m\}$. The research questions of this study are summarized as:

(1) How to measure the reliability of different online platforms?

(2) How to integrate the attribute sentiments of multiple platforms using the reliability of platforms considering different evaluation modes?

(3) How to construct a multi-platform-oriented IPA model to classify attributes and develop service improvement strategies.

To address these issues, this Section proposes a multi-platform-oriented IPA model considering different evaluation modes in multiple platforms and the reliability of different platforms. Specifically, Section 3.1. calculates the review helpfulness and combines it with the number of online reviews to determine the reliability of platforms. Section 3.2. computes attribute sentiments on each platform using lexicon-based sentiment analysis methods considering different evaluation modes of platforms and aggregates the attribute sentiments from multiple platforms to obtain the overall performance of attributes. Section 3.3. constructs a multi-platform-oriented IPA model to classifying attributes and develops service improvement strategies. Section 3.4. presents a summary of the proposed model. The framework of the proposed model is shown in Figure 1.

3.1. Determine the reliability of each platform

Determining the reliability of multiple platforms plays a key role in aggregating the sentiment information of multiple platforms. To measure the reliability of different platforms, the number of online reviews and reviews helpfulness are taken into account.

- *The number of online reviews*. The more a platform has reviews, the higher the weight should be given to the platform, as it reflects the wide popularity of the product and the high level of attention from consumers. Let Q_i denote the number of online reviews on platform A_i.
- *Reviews helpfulness*. Helpful reviews provide potential consumers with comprehensive and authentic information about a product or service, thereby helping them make wiser

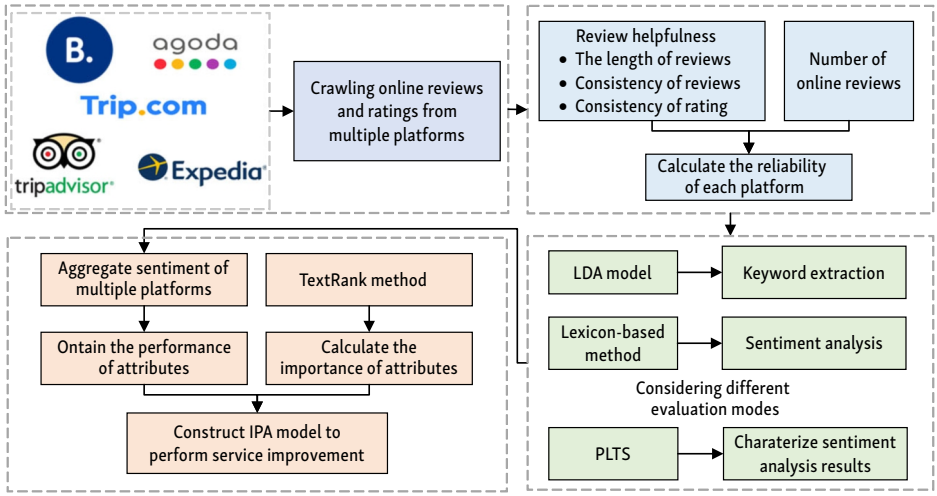

Figure 1. The framework of the proposed model

purchasing decisions. Useful reviews contribute to increasing the reliability of platforms and attract more consumers as consumers are more willing to transact on a trustworthy platform. Therefore, review helpfulness is an important aspect for measuring the reliability of platforms.

The length of reviews reflects the amount of information reviews incorporate and is a key factor that influences review helpfulness (Li & Huang, 2020). Based on the information processing theory, short reviews are not enough to eliminate consumers' uncertainty about a product, while overly long reviews require more information processing resources, ultimately causing consumers to resist or abandon reading reviews (Feng et al., 2023). Thus, an inverted U-shape exists between the length of reviews and review helpfulness (Lutz et al., 2022). In this study, the review helpfulness is measured from the perspective of the length of review by an inverted U-shape function. Since the length of reviews has extreme values, logarithmic normalization is used to scale the review length to [0,1]. The normalized length of each review, NT_{iq}, is obtained by Eq. (1).

$$NT_{iq} = \frac{\log(T_{iq})}{\log(\underset{1\le q\le Q,1\le i\le m}{Max}(T_{iq}))}, 0 \le NT_{iq} \le 1, \tag{1}$$

where T_{iq} is the length of the q*th* review on platform A_i. Then, NT_{iq}, ranging from 0 to 1, is mapped to review helpfulness U_{iq} of the q*th* review on platform A_i by an inverted U-shape function shown as Eq. (2).

$$U_{iq} = -\frac{1}{(NT^*)^2}(NT_{iq} - NT^*)^2 + 1, NT^*_{iq} \ge 0.5, 0 \le U_{iq} \le 1, \tag{2}$$

where NT^* is the normalized value of the review length that maximizes the review helpfulness, that is, the overload point for transforming the helpfulness from increasing to decreasing. The average helpfulness AU_i of all reviews on platform A_i can be obtained by Eq. (3), where Q_i is the number of reviews on platform A_i.

$$AU_i = \frac{\sum_{q=1}^{Q_i} U_{iq}}{Q_i}. \tag{3}$$

Reviews or ratings with high consistency usually reflect a high degree of consumer consensus on the product, enhancing the reliability of the platform. Reviews helpfulness increases as its rating aligns more closely with other ratings of products (Jia et al., 2022). Inspired by this idea, we measure the review helpfulness from the viewpoint of the consistency of online reviews and ratings. Since customer opinions from reviews and ratings may be inconsistent owing to the discreteness of ratings and ambiguity in online reviews, their consistency is evaluated separately. Let RA_{iq} be the rating of the q*th* review collected from platform A_i. Owning to the different evaluation scales of different platforms, the maximum and minimum normalization is performed separately on ratings from platforms with different evaluation scales using Eq. (4) and the normalized rating NRA_{iq} can be obtained. Then, the consistency of ratings CO_i^1 on platform A_i is calculated by Eq. (5).

$$NRA_{iq} = \frac{RA_{iq} - \underset{1\le q\le Q}{Min}(RA_{iq})}{\underset{1\le q\le Q}{Max}(RA_{iq}) - \underset{1\le q\le Q}{Min}(RA_{iq})}, 0 \le NRA_{iq} \le 1; \tag{4}$$

$$CO_i^1 = 1 - \frac{1}{Q_i}\sum_{q=1}^{Q_i}\left(NRA_{iq} - \frac{1}{Q_i}\sum_{q=1}^{Q_i} NRA_{iq}\right)^2. \tag{5}$$

The consistency of reviews is determined by comparing the semantic similarity between reviews based on a pre-trained Glove model[2]. Eq. (6) shows the average semantic similarity Sim_{iq} of the q*th* review on platform A_i to other reviews.

$$Sim_{iq} = \frac{1}{Q_i - 1}\sum_{\theta=1,q\ne\theta}^{Q_i} \frac{\beta_\theta \beta_q}{\left|\beta_\theta\right|\left|\beta_q\right|}, \theta = 1,2,\ldots,Q_i, q = 1,2,\ldots,Q_i, \tag{6}$$

where β_θ and β_q are the word vector of the θ*th* and q*th* review, respectively. $\beta_q\beta_\theta \big/ \left|\beta_q\right|\left|\beta_\theta\right|$ is the cosine similarity between β_θ and β_q. The consistency of reviews CO_i^2 on platform A_i is obtained by Eq. (7), which is measured by the average semantic similarity of all reviews.

$$CO_i^2 = \frac{\sum_{q=1}^{Q_i} Sim_{iq}}{Q_i}. \tag{7}$$

The review helpfulness and the number of online reviews can be aggregated by Eq. (8), where the front part of the equation represents the review helpfulness resulted from the length of reviews and the consistency of reviews and ratings, and the latter indicates the influence of the number of reviews to the reliability of platforms. W_i represents the reliability of platform A_i.

$$W_i = \frac{1}{2} * \left(\frac{AU_i + CO_i^1 + CO_i^2}{3} + \frac{Q_i}{\sqrt{\sum_{i=1}^{m} Q_i^2}}\right). \tag{8}$$

[2] The Glove model is a word embedding model that leverages the global co-occurrence statistics and local context information of words to capture semantic relationships (Pimpalkar & Jeberson Retnaraj, 2022).

3.2. Aggregating sentiments from multiple platforms

Aggregating the performance and importance of attributes across different platforms facilitates a comprehensive understanding of consumers' requirements. The sentiments of attributes are regarded as attribute performance in service improvement. Keyword extraction and sentiment analysis are two crucial steps to obtain the sentiments of attributes. This Section aims to analyze and aggregate the sentiments of attributes based on online reviews from multiple platforms.

Keyword extraction refers to extracting keywords with similar meanings related to the attributes concerned by customers. The LDA model is an unsupervised topic extraction method that has been widely used in identifying attributes from online reviews. Given its wide applicability and effectiveness, the LDA model is utilized in this study to extract attributes from online reviews collected from multiple platforms. These attributes represent the common focus areas across platforms. Data preprocessing is the primary step in implementing LDA, which includes eliminating special characters, converting uppercase to lowercase, and removing stop words. Results generated by the LDA model may include incorrect categorization or noise. Thus, manual filtering is required to ensure the correctness of the results. Specifically, words with the same meaning are fused into a category, irrelevant words are removed from the category, and each category is tagged and regarded as an attribute C_j. The keywords belonging to attribute C_j are denoted as $\left\{K_{j1}, K_{j2}, ..., K_{ja}\right\}$, where a is the number of keywords belonging to C_j. It is worth noting that manual filtering is a necessary supplement to the results of LDA models. Although this process may introduce subjectivity, manual filtering guided by domain knowledge is a standard practice in current LDA applications (Bi et al., 2019; Zhang et al., 2021), which can significantly improve the accuracy and practicality of results.

Sentiment analysis aims to identify the sentiment or attitude of customers to attributes. Lexicon-based sentiment analysis methods are employed to perform sentiment analysis as they do not require feature construction and manually labeled data, saving a lot of time and labor compared with machine learning methods. Vader is a lexicon-based sentiment analysis instrument that has been widely used to analyze the sentiments of online reviews on social media owing to its simplicity, speed, and effectiveness. It excels in handling context-dependent sentiment classification tasks, providing a solid foundation for analyzing attribute performance. Figure 2 displays a *detailed* review in Expedia and a *pros/cons* review in Booking. A *detailed* review expresses consumer's experience and sentiments to attributes, while a *pros/cons* review conveys positive and negative sentiments separately even without sentiment words. For different evaluation models, we adopt different strategies for sentiment analysis.

For *detailed* reviews, we divide each review into sentences $DS = \left\{DS_1, DS_2, ..., DS_t\right\}$ based on punctuation to minimize the number of attributes contained in a sentence so that the extracted degree adverbs are basically related to the attribute. Then, find out the sentences from DS containing C_j, which are denoted as $DS^j = \{DS_1^j, DS_2^j, ..., DS_L^j\}, l = 1, 2, ..., L$, and L is the number of sentences including C_j. Given that containing an attribute in a sentence essentially involves mentioning the keywords that belong to the attribute, the sentences containing C_j are actually the sentences mentioning the keywords $\{K_{j1}, K_{j2}, ..., K_{ja}\}$ that belong to attribute C_j. Next, analyze the sentiment of DS_l^j with the Vader method and obtain the sentiment score of sentence DS_l^j. Based on Eq. (9), the sentiment score of attribute C_j can be determined

Figure 2. Different evaluation modes on different online platforms

according to the sentiment of sentence DS_l^j. SC_j^l is the sentiment score of attribute C_j in sentence DS_l^j. cp is the compound score of the sentence DS_l^j calculated by the Vader method. In order to identify the sentiment intensity of C_j in current sentence, degree adverbs d_j^l, which play a role in deepening or weakening sentiments, are extracted from DS_l^j. A part of speech tagging is implemented to achieve it. Degree adverbs d_j^l are categorized into different levels to portray the difference in intensity. The sentiment intensity is determined by Eq. (10), where D_j^l is the sentiment intensity of C_j in sentence DS_l^j, DC_1, DC_2, and DC_3 represent the collection of degree adverbs with different intensities, separately. If a sentence includes more than one degree adverb, then the highest adverbial degree is utilized. The sentiment polarity of C_j in sentence DS_l^j considering sentiment intensity is calculated by Eq. (11).

$$SC_j^l = \begin{cases} 1, cp > 0 \\ 0, cp = 0 \\ -1, cp < 0 \end{cases}; \tag{9}$$

$$D_j^l = \begin{cases} 1, d_j^l \in DC_1 \\ 2, d_j^l \in DC_2 \\ 3, d_j^l \in DC_3 \end{cases}; \tag{10}$$

$$S_j^l = SC_j^l \times D_j^l, S_j^l \in \{-3, -2, -1, 0, 1, 2, 3\}. \tag{11}$$

For *pros/cons* reviews, the reviews are sliced into sentences $PS = \{PS_1, PS_2, ..., PS_z\}$ for the same purpose as *detailed* reviews. Similarly, search for the reviews from PS containing C_j, and represent the reviews as $PS^j = \{PS_1^j PS_2^j, ..., PS_B^j\}, b = 1, 2, ..., B$, where B is the number of sentences including C_j. Since the *pros/cons* reviews have embedded positive and negative sentiments, there is no need to judge the sentiment polarity by lexicon-based methods. Eq. (12) illustrates the sentiment score SC_j^b of attribute C_j in sentence PS_b^j, PO represents a collection of reviews posted under positive items, and NE is a set of reviews posted under

negative items. The sentiment intensity of C_j in sentence PS_b^j is obtained by Eq. (13). The sentiment polarity of C_j in review PS_b^j considering sentiment intensity is calculated by Eq. (14).

$$SC_j^b = \begin{cases} 1, PS_b^j \in PO \\ -1, PS_b^j \in NE \end{cases}; \tag{12}$$

$$D_j^b = \begin{cases} 1, d_j^b \in DC_1 \\ 2, d_j^b \in DC_2 \\ 3, d_j^b \in DC_3 \end{cases}; \tag{13}$$

$$S_j^b = SC_j^b \times D_j^b, S_j^b \in \{-3,-2,-1,0,1,2,3\}. \tag{14}$$

Considering the ambiguity of human evaluations and the incomplete accuracy of sentiment analysis methods, PLTS, an effective tool for portraying complex linguistic information in the context of uncertainty, is utilized to characterize the probability distribution of each attribute under each sentiment polarity. Let $\{s_{-3} = \text{very negative}, s_{-2} = \text{quite negative}, s_{-1} = \text{negative}, s_0 = \text{neutral}, s_1 = \text{positive}, s_2 = \text{quite positive}, s_3 = \text{very positive}\}$ characterizes the sentiment polarity S_j^l and S_j^b, where the value of S_j^l and S_j^b corresponds to the subscript of s_α, respectively. The probability distribution of sentiment polarity of attribute C_j on platform A_i is denoted as $h_s^{ij}(p) = \left\{ s_\alpha^{ij(r)}(p^{ij(r)}) \mid r = 1,2,\ldots R, \sum_{r=1}^{R} p^{ij(r)} \le 1 \right\}, i = 1,2,...,m; j = 1,2,...,n$, where $s_\alpha^{ij(r)}$ represents the linguistic term of sentiment category r under attribute C_j on platform A_i, and $\alpha = \{-3,-2,-1,0,1,2,3\}$. $p^{ij(r)}$ is the probability of $s_\alpha^{ij(r)}$ and is calculated by Eq. (15), where N_i^{jr} denotes the number of occurrences of sentiment category r regarding attribute C_j on platform A_i. $h_s^{ij}(p)$ is the PLTSs of attribute C_j on platform A_i. A score function (Wu & Liao, 2019) is utilized to calculate the expectation value of PLTSs, where a linguistic scale function $f(s_\alpha)$ (Wu & Liao, 2019) is introduced (Eq. (17)). The expectation value $E(h_s^{ij}(p))$ is regarded as attribute performance S_j^i of attribute C_j on platform A_i. $1_{\{\alpha\in[-\tau,0]\}}$ is an indicative function, υ and μ are two parameters indicating risk preferences. Based on $W = \{w_1, w_2, \ldots, w_m\}$ and S_j^i, the attribute performance S_j of multiple platforms is obtained by Eq. (18).

$$p^{ij(r)} = \frac{N_i^{jr}}{\sum_{r=1}^{R} N_i^{jr}}; \tag{15}$$

$$S_j^i = E(h_s^{ij}(p)) = \sum_{r=1}^{R} (f(s_\alpha^{ij(r)}) \cdot p^{ij(r)} / \sum_{r=1}^{R} p^{ij(r)}); \tag{16}$$

$$f(s_\alpha) = \frac{\upsilon^\tau - \upsilon^{-\alpha}}{2\upsilon^\tau - 2} \times 1_{\{\alpha\in[-\tau,0]\}} + \frac{\mu^\tau + \mu^\alpha - 2}{2\mu^\tau - 2} \times 1_{\{\alpha\in(0,\tau]\}}, \alpha \in [-\tau,\tau]; \tag{17}$$

$$S_j = \sum_{i=1}^{m} W_i S_j^i. \tag{18}$$

3.3. Classify attributes based on the IPA model

This section extends the IPA model to the multi-platform environments considering the reliability across platforms. The performance and importance of attributes are important aspects in constructing the IPA model. The TextRank method is a word graph-based keyword extrac-

tion method and can capture the importance of words in the context. We take the online reviews from each platform as a document. The document is modeled by a graph network G = (*V*, *E*), where *V* is the node set and *E* is cooccurrence between two words located in the same sliding window. The importance of words is calculated by an iterative formula Eq. (19), where $I(\upsilon_p)$ is the weight of υ_p, υ_p and υ_q represent related words, CW_{qp} is the connection weight between υ_p and υ_q. $In(\upsilon_p)$ the set of nodes pointing to υ_p, while $Out(\upsilon_q)$ is the set of nodes that υ_q points to. η is a damping coefficient. The equation shows that the weight of υ_p relies on the connection weight from υ_q to υ_p and the sum of connection weights from υ_q to the nodes that υ_q points to. The higher the score is, the more important the word is. Since keywords are a fine-grained description of an attribute, the importance of attributes can be obtained based on the importance of keywords belonging to the attribute. The importance of attribute C_j on platform A_i is calculated by Eq. (20), where $I^i(K_{jg})$ is the importance of the q*th* keyword under attribute C_j on platform A_i. The importance of attribute C_j is obtained by Eq. (21).

$$I(\upsilon_p) = (1-\eta) + \eta * \sum_{\upsilon_q \in In(\upsilon_p)} \frac{CW_{qp}}{\sum_{\upsilon_o \in Out(\upsilon_q)} CW_{qo}} I(\upsilon_q); \tag{19}$$

$$I_j^i = \frac{\sum_{g=1}^{a} I^i(K_{jg})}{a}; \tag{20}$$

$$I_j = \sum_{i=1}^{m} W_j I_j^i. \tag{21}$$

Take average performance $AS = \frac{1}{n}\sum_{j=1}^{n} S_j$ and average importance $AI = \frac{1}{n}\sum_{j=1}^{n} I_j$ of attributes as boundary points to classify attributes. According to Bi et al. (2019), attributes can be classified into the following categories based on *AS* and *AI*:

(1) keep up the good work (Q_1). Attributes exhibit both high importance and performance $\left(S_j \geq AS, I_j \geq AI\right)$, which indicates the attributes are crucial strengths and advantages of the product/service. These attributes should receive sustained attention and investment to maintain competitiveness.

(2) concentrate here (Q_2). This refers to those attributes that have high importance but low performance $\left(S_j < AS, I_j \geq AI\right)$.

Managers are suggested to focus on improving these attributes as they are critical but do not perform well, so as to meet customer's needs and expectations.

(3) low priority (Q_3). The importance and performance of attributes in this category are below average $\left(S_j < AS, I_j < AI\right)$. Attributes belonging to this category have low resource allocation priority.

(4) possible overkill (Q_4). This category refers to attributes that display low importance and high performance $\left(S_j \geq AS, I_j < AI\right)$. It implies that managers should appropriately reduce investment in such attributes to optimize resource utilization.

3.4. Summary of the proposed method

The proposed method for service improvement considering the reliability of multiple online platforms is concluded in the following steps.

Step 1: Crawl online reviews and ratings about hotels from multiple online platforms.

Step 2: Calculate the number of online reviews and review helpfulness of each platform, and aggregate two indicators to obtain the reliability of each platform.

Step 3: Perform data preprocessing and LDA model to determine attributes and corresponding keywords.

Step 4: Analyze the sentiments of online reviews on each platform by lexicon-based sentiment analysis methods based on the rules for *detailed* reviews and *pros/cons* reviews in Section 3.2. Express the probability distribution of attribute sentiments on each platform with PLTSs. Aggregate the sentiments of attributes from different platforms and obtain the overall performance of attributes.

Step 5: Calculate the importance of attributes on each platform by the TextRank method. Aggregate the importance of attributes from different platforms. Construct an IPA model to develop service improvement strategies. Specifically, determine the boundary points based on the overall performance and overall importance of attributes. Classify the attributes into four categories based on the boundary points, where each category adopts different strategies for service improvement.

4. Case study

This section aims to develop hotel service improvement strategies based on the proposed model. Sensitivity analysis and comparative analysis are given to illustrate the stability and effectiveness of the proposed model.

4.1. Case description

In the fiercely competitive hotel market, exploring consumer preferences and satisfaction to improve service quality has become an important task for hoteliers. Consumers can book hotels through various channels and post their opinions and experiences on different platforms. It is necessary for hoteliers to analyze the evaluation information of consumers across all platforms to prevent the strategies from being biased due to different evaluation mechanisms and insufficient data, so as to formulate comprehensive and effective service improvement strategies.

Five online platforms including TripAdvisor, Booking, Agoda, Expedia, and Trip in the hospitality domain are selected for hotel service improvement, these platforms occupy a major position in the online hotel market. 2756 online reviews of Artyzen Grand Lapa Macau hotel in Macao, China are collected from five platforms. The data span March 1, 2005 to May 24, 2024. The number of online reviews on platforms TripAdvisor(A_1), Booking(A_2), Agoda(A_3), Expedia(A_4), and Trip(A_5) is 840, 207, 1099, 259, and 351 respectively. The evaluation mode of Booking is *pros/cons* type, while the rest are *detailed* types. To ensure the reliability and valid-

ity of the data used in this study, we carefully examined the data collection and preprocessing processes, as well as potential external influences that could impact the dataset. During the data collection phase, platforms providing the data have well-established review mechanisms. Most platforms, such as Booking, Expedia, and Agoda, only allow users who have made a purchase to leave reviews. Additionally, these platform's review verification mechanisms and reporting features effectively prevent fake or manipulated reviews, enhancing the reliability of the data. During data preprocessing, reviews with suspiciously high similarity or repetitive content are removed as they may indicate human manipulation. We also conducted a comprehensive review of potential event-driven anomalies during the period covered by the data and confirmed that no significant events occurred that could distort the review data.

4.2. Resolving process

This Section applies the proposed method to classify attributes and develop service improvement strategies. First, the length of reviews is calculated by counting the words of the review and is normalized by Eq. (1). The review helpfulness resulting from the length of reviews is calculated by Eq. (2), where NT^* is set as 0.6, which corresponds to the upper quartile of the length of all reviews. By averaging the review helpfulness of all reviews on each platform using Eq. (3), AU_i is calculated and the results are displayed in the second column of Table 4. Since the evaluation scale of Booking, Agoda, and Expedia ranges from 0 to 10, while TripAdvisor and Trip ranges from 0 to 5. We normalize the ratings on platforms with different evaluation scales separately by Eq. (4). The consistency of ratings on each platform is calculated by Eq. (5). and CO_i^1 for each platform is obtained, with the results shown in the third column of Table 4. In terms of the consistency of reviews, the Glove model trained on a Twitter dataset with 100 dimensions vectors is utilized to transform the reviews into word vectors and calculate the semantic similarity between reviews. The average similarity of each review to other reviews is computed by Eq. (6). Then, we aggregate the average similarity of each review by Eq. (7), and the corresponding CO_i^2 of each platform are presented in the fourth column of Table 4. The review helpfulness and the number of reviews are aggregated by Eq. (8), and the reliability W_i for each platform is determined as shown in Table 4.

Table 4. Results of review helpfulness, consistency, and reliability metrics for each platform

Platform	AU_i	CO_i^1	CO_i^2	W_i
A_1	0.970	0.954	0.914	0.760
A_2	0.860	0.976	0.708	0.495
A_3	0.851	0.976	0.807	0.814
A_4	0.820	0.967	0.796	0.519
A_5	0.833	0.974	0.811	0.556

The Python gensim toolkit (https://github.com/piskvorky/gensim) is imported to conduct the LDA model. The number of topics is set as eight. By a manual adjustment to the results of the LDA model in Section 3.2, the attributes and keywords are determined as shown in Table 5.

Table 5. Attributes and keywords

Attributes	Keywords
Room	bed, water, clean, door, shower, floor, comfortable, smell
Location	shuttle, bus, ferry, terminal, distance, walking, taxi, center, nearby, located, near
Environment	square, view, resort, wharf, pier, city, attraction
Facility	pool, wifi, tv, internet, amenity, spa, swimming, gym, iron
Food	breakfast, buffet, coffee, restaurant, menu
Service	staff, check, guest, front, friendly, helpful, polite
Style	modern, interior, design, glitz, area, old
Value	money, free, price, quality, cost

It is difficult to analyze the sentiments of attributes when a review contains multiple attributes. To overcome this issue, each review is split into sentences to make a sentence contain fewer attributes so that the extracted degree adverbs are basically associated with the attribute. 10986 sentences are obtained from 2756 online reviews using Python nltk (https://github.com/nltk/nltk) toolkit. Then, adverbs are extracted from 10986 sentences and 11478 adverbs are found in total utilizing Python spaCy (https://github.com/explosion/spaCy) toolkit. After removing repeated adverbs, 675 adverbs are obtained, including time adverbs, place adverbs, manner adverbs, frequency adverbs, and degree adverbs. Only degree adverbs are considered as they influence sentiment intensity. To quantify the strength of degree adverbs, the degree adverbs are divided into three classifications, the results are displayed in Table 6.

Table 6. Different strengths of degree adverbs

Degree	Degree adverbs
3	amazingly, completely, entirely, exceptionally, extremely, highly, incredibly, immensely, remarkably, significantly, strongly, superbly, tremendously, utterly, very, absolutely, definitely, greatly, ideally, perfectly, vastly, severely, surprisingly, terribly, thoroughly, totally, appallingly, horribly, fully, overly, unbelievably, seriously, heavily, deeply, too
2	fairly, pretty, quite, really, rather, especially, particularly, strangely, unfortunately, hardly, badly, less, disappointingly, scarcely, truly
1	generally, substantially, consistently, probably, solely, barely, just, merely, slightly, somewhat, enough, marginally, temporarily, nearly, approximately

For the platform with *detailed* evaluation mode, Vader is utilized to identify the sentiment polarity of sentences by Eq. (9). For the platform with *pros/cons* evaluation mode, the sentiment polarity of sentences is assigned using Eq. (12) according to the category they belong to. In addition, the degree adverbs contained in sentences are extracted and assigned a degree score by Eq. (10) and Eq. (13). By multiplying the sentiment score and degree using Eq. (11) and Eq. (14), the sentiment polarity of the sentence can be obtained. Table 7 shows a few cases about determining the sentiment polarity of sentences.

Table 7. The cases about determining sentiment polarity

	Review	Attribute	Sentiment score	Adverb degree	Sentiment polarity
Tripadvisor/Agoda/Expedia/Trip	The sound separations between room to room and room to corridor are terribly poor.	Keyword: room Attribute: room	–0.7717→[–1]	terribly (3)	–3
	Breakfast is the real pain, the breakfast restaurant is too small so if you're not there by 7 am you will queue for ages to get a table!	Keywords: breakfast, restaurant Attribute: food	–0.5562→[–1]	too (3)	–3
	If you are traveling to and from hk, the ferry can somewhat be easily accessed since it is very close.	Keyword: ferry Attribute: location	0.2782→[1]	somewhat (1), very (3)	3
	The casino staff and especially dealers were really rude and mean	Keyword: staff Attribute: service	–0.5095→[–1]	especially (2), really (2)	–2
	This hotel's location is really good, especially for families who avoid crowded places and just want to relax.	Keyword: location Attribute: location	0.6361→[1]	really (2), especially (2), just (1)	2
Booking	Positive: the room	Keyword: room Attribute: room	[1]	–	1
	Positive: the room was very comfortable and the bed extremely comfortable.	Keywords: room, bed Attribute: room	[1]	very (3), extremely (3)	3
	Negative: I liked everything except that the pool was just too crowded with too many people.	Keyword: pool Attribute: facility	[–1]	just (1), too (3)	–3

The sentiment polarity distribution of attributes on each platform is calculated via Eq. (15). The results are shown as follows:

$$A_1=\begin{cases} C_1: s_{-3}(0.013), s_{-2}(0.010), s_{-1}(0.066), s_0(0.140), s_1(0.459), s_2(0.068), s_3(0.244) \\ C_2: s_{-3}(0.015), s_{-2}(0.004), s_{-1}(0.045), s_0(0.246), s_1(0.464), s_2(0.032), s_3(0.194) \\ C_3: s_{-3}(0.011), s_{-2}(0.008), s_{-1}(0.036), s_0(0.131), s_1(0.544), s_2(0.080), s_3(0.190) \\ C_4: s_{-3}(0.011), s_{-2}(0.017), s_{-1}(0.052), s_0(0.105), s_1(0.555), s_2(0.046), s_3(0.214) \\ C_5: s_{-3}(0.015), s_{-2}(0.009), s_{-1}(0.038), s_0(0.143), s_1(0.523), s_2(0.064), s_3(0.208) \\ C_6: s_{-3}(0.018), s_{-2}(0.009), s_{-1}(0.047), s_0(0.097), s_1(0.471), s_2(0.050), s_3(0.308) \\ C_7: s_{-3}(0.022), s_{-2}(0.012), s_{-1}(0.045), s_0(0.190), s_1(0.455), s_2(0.056), s_3(0.220) \\ C_8: s_{-3}(0.007), s_{-2}(0.020), s_{-1}(0.044), s_0(0.058), s_1(0.633), s_2(0.058), s_3(0.180) \end{cases}'$$

$$A_2=\begin{cases} C_1: s_{-3}(0.032), s_{-2}(0.017), s_{-1}(0.266), s_1(0.468), s_2(0.040), s_3(0.177) \\ C_2: s_{-3}(0.091), s_{-1}(0.303), s_1(0.455), s_2(0.015), s_3(0.136) \\ C_3: s_{-3}(0.035), s_{-1}(0.103), s_1(0.448), s_2(0.069), s_3(0.345) \\ C_4: s_{-3}(0.019), s_{-2}(0.048), s_{-1}(0.171), s_1(0.571), s_2(0.048), s_3(0.143) \\ C_5: s_{-3}(0.043), s_{-2}(0.012), s_{-1}(0.288), s_1(0.500), s_2(0.054), s_3(0.163) \\ C_6: s_{-3}(0.038), s_{-1}(0.185), s_1(0.523), s_2(0.039), s_3(0.215) \\ C_7: s_{-3}(0.029), s_{-2}(0.030), s_{-1}(0.235), s_1(0.412), s_2(0.059), s_3(0.235) \\ C_8: s_{-3}(0.042), s_{-1}(0.292), s_1(0.541), s_2(0.042), s_3(0.083) \end{cases}'$$

$$A_3 = \begin{cases} C_1 : s_{-3}(0.019), s_{-2}(0.010), s_{-1}(0.066), s_0(0.183), s_1(0.496), s_2(0.045), s_3(0.181) \\ C_2 : s_{-3}(0.002), s_{-2}(0.005), s_{-1}(0.065), s_0(0.270), s_1(0.479), s_2(0.040), s_3(0.139) \\ C_3 : s_{-1}(0.042), s_0(0.211), s_1(0.542), s_2(0.048), s_3(0.157) \\ C_4 : s_{-3}(0.006), s_{-2}(0.011), s_1(0.073), s_0(0.148), s_1(0.546), s_2(0.056), s_3(0.160) \\ C_5 : s_{-3}(0.004), s_{-1}(0.087), s_0(0.115), s_1(0.559), s_2(0.056), s_3(0.179) \\ C_6 : s_{-3}(0.011), s_{-2}(0.002), s_{-1}(0.032), s_0(0.113), s_1(0.507), s_2(0.048), s_3(0.287) \\ C_7 : s_{-3}(0.016), s_{-2}(0.016), s_{-1}(0.053), s_0(0.266), s_1(0.457), s_2(0.064), s_3(0.128) \\ C_8 : s_{-2}(0.004), s_{-1}(0.069), s_0(0.069), s_1(0.659), s_2(0.047), s_3(0.151) \end{cases},$$

$$A_4 = \begin{cases} C_1 : s_{-1}(0.068), s_0(0.263), s_1(0.491), s_2(0.042), s_3(0.136) \\ C_2 : s_{-3}(0.011), s_{-2}(0.011), s_{-1}(0.069), s_0(0.276), s_1(0.529), s_2(0.012), s_3(0.092) \\ C_3 : s_{-1}(0.067), s_0(0.300), s_1(0.566), s_3(0.067) \\ C_4 : s_{-1}(0.042), s_0(0.104), s_1(0.645), s_2(0.042), s_3(0.167) \\ C_5 : s_{-1}(0.044), s_0(0.174), s_1(0.500), s_2(0.043), s_3(0.239) \\ C_6 : s_{-3}(0.009), s_{-2}(0.008), s_{-1}(0.047), s_0(0.118), s_1(0.614), s_2(0.031), s_3(0.173) \\ C_7 : s_{-3}(0.024), s_{-2}(0.024), s_{-1}(0.048), s_0(0.238), s_1(0.381), s_2(0.048), s_3(0.237) \\ C_8 : s_{-1}(0.137), s_0(0.227), s_1(0.500), s_3(0.136) \end{cases},$$

$$A_5 = \begin{cases} C_1 : s_{-3}(0.027), s_{-2}(0.007), s_{-1}(0.027), s_0(0.215), s_1(0.469), s_2(0.040), s_3(0.215) \\ C_2 : s_{-2}(0.015), s_{-1}(0.072), s_0(0.261), s_1(0.507), s_2(0.029), s_3(0.116) \\ C_3 : s_{-1}(0.053), s_0(0.132), s_1(0.579), s_2(0.079), s_3(0.157) \\ C_4 : s_{-1}(0.053), s_0(0.132), s_1(0.578), s_2(0.044), s_3(0.193) \\ C_5 : s_{-1}(0.076), s_0(0.227), s_1(0.561), s_2(0.015), s_3(0.021) \\ C_6 : s_{-3}(0.011), s_{-1}(0.046), s_0(0.051), s_1(0.538), s_2(0.036), s_3(0.318) \\ C_7 : s_0(0.241), s_1(0.552), s_2(0.069), s_3(0.138) \\ C_8 : s_{-1}(0.022), s_0(0.067), s_1(0.822), s_3(0.089) \end{cases}.$$

The expectation score of PLTSs is obtained by Eqs. (16)–(17) with the parameters υ = μ = 1.5. The score is regarded as the performance of attributes and is displayed in Table 7. Then, we regard the reviews of each platform as a document and utilize the TextRank method to calculate the importance of keywords using Eq. (18). The importance of attributes can be obtained by Eqs. (19)–(20). The results are also shown in Table 8. By aggregating the performance and importance of each platform using Eq. (21), the comprehensive performance and importance of multiple platforms are obtained as shown in Table 9.

Table 8. The performance and importance of each attribute on each platform

Attribute	TripAdvisor		Booking		Agoda		Expedia		Trip	
	S_1	I_1	S_2	I_2	S_3	I_3	S_4	I_4	S_5	I_5
C_1	0.672	0.075	0.600	0.093	0.635	0.069	0.624	0.088	0.649	0.102
C_2	0.641	0.045	0.543	0.071	0.621	0.046	0.589	0.057	0.608	0.054
C_3	0.6620	0.034	0.710	0.038	0.644	0.035	0.586	0.037	0.655	0.047
C_4	0.662	0.053	0.604	0.096	0.639	0.056	0.658	0.067	0.663	0.073
C_5	0.662	0.063	0.600	0.088	0.652	0.064	0.679	0.062	0.616	0.060
C_6	0.700	0.092	0.634	0.116	0.700	0.095	0.649	0.107	0.715	0.109
C_7	0.654	0.052	0.629	0.048	0.611	0.044	0.649	0.059	0.645	0.049
C_8	0.658	0.041	0.558	0.045	0.649	0.040	0.607	0.035	0.629	0.044

Table 9. The performance and importance of each attribute of multiple platforms

Attribute	Performance	Importance
C_1	2.009	0.261
C_2	1.904	0.166
C_3	2.046	0.118
C_4	2.032	0.209
C_5	2.024	0.209
C_6	2.150	0.321
C_7	2.001	0.156
C_8	1.969	0.128

The boundary points are determined by average performance and importance. According to the classification rules, an IPA model is constructed based on the performance and importance of attributes. As shown in Figure 3, "service", "food", and "facility" are classified as "keep up good work", which indicates that these attributes are valued by consumers and consumers are satisfied with the attributes, hoteliers should continue to keep investing in these attributes in the future because they have a direct impact on customer satisfaction. The attribute "room" is located in the area of "concentrate here", which shows that customers attach importance to the attribute "room", but the performance on this attribute fails to meet the expectations of customers. "Room" is suggested as a priority item for improvement. The attributes "location", "value", and "style" belong to "low priority", which indicates the importance of these attributes is low and the performance is also less than satisfactory. Hoteliers can prioritize them low when formulating service improvement strategies. "Environment" is categorized as "possible overkill". It is advisable for hoteliers to consider appropriate reductions or resource reallocations in the attribute and concentrate on upgrading other critical attributes.

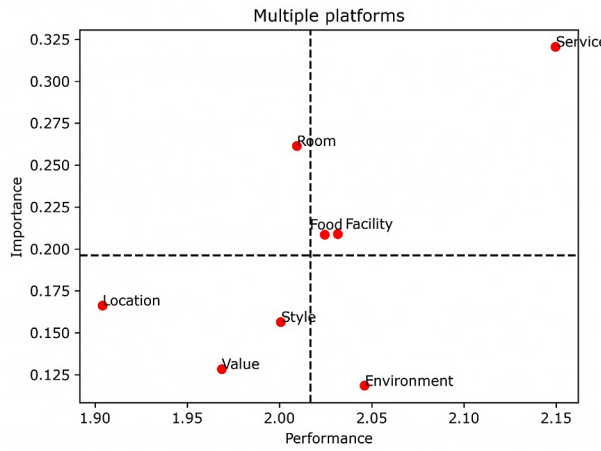

Figure 3. The classification results of the IPA model

4.3. Sensitivity analysis

The determination of the overload point NT^* has no standardized criterion and influences the reliability of platforms. Thus, we take $NT^* \in [0.5,1]$ with a step size of 0.05 to test the influence of changing NT^* on the classification results. Figure 4 shows different inverted U-shape curves with different vertex determined by NT^*. The vertical axis indicates the review helpfulness determined by the length of reviews. With the increase of NT^*, the length of reviews whose review helpfulness reverses from growth to decline tends to be longer. Table 10 shows the average review helpfulness resulting from the review length of each platform. By aggregating with other indicators, the reliability of platforms under different NT^*can be obtained, as shown in Table 10.

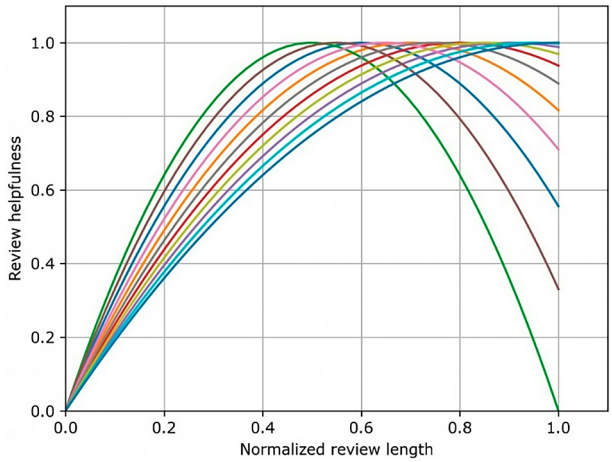

Figure 4. Different functions for mapping normalized review length to review helpfulness

Table 10. The reliability of platforms determined by different overload points

	TripAdvisor		Booking		Agoda		Expedia		Trip	
	AU_1	W_1	AU_2	W_2	AU_3	W_3	AU_4	W_4	AU_5	W_5
NT^* = 0.50	0.900	0.748	0.884	0.499	0.869	0.817	0.858	0.525	0.866	0.562
NT^* = 0.55	0.949	0.756	0.877	0.497	0.866	0.817	0.842	0.523	0.854	0.560
NT^* = 0.60	0.970	0.760	0.860	0.495	0.851	0.814	0.820	0.519	0.833	0.556
NT^* = 0.65	0.973	0.760	0.837	0.491	0.831	0.811	0.794	0.515	0.808	0.552
NT^* = 0.70	0.966	0.759	0.812	0.487	0.807	0.807	0.767	0.510	0.782	0.548
NT^* = 0.75	0.952	0.757	0.786	0.482	0.782	0.803	0.740	0.506	0.755	0.543
NT^* = 0.80	0.933	0.753	0.760	0.478	0.757	0.798	0.713	0.501	0.729	0.539
NT^* = 0.85	0.913	0.750	0.735	0.474	0.732	0.794	0.688	0.497	0.703	0.534
NT^* = 0.90	0.891	0.746	0.710	0.470	0.708	0.790	0.663	0.493	0.679	0.530
NT^* = 0.95	0.868	0.743	0.686	0.466	0.685	0.786	0.640	0.489	0.655	0.526
NT^* = 1.00	0.846	0.739	0.664	0.462	0.663	0.783	0.618	0.485	0.633	0.523

Although the reliability of platforms varies with the different overload points, there is no effect on the classification results of the attributes. The attribute classification results under different overload points are all: $C_1 \in Q_2$, $C_4, C_5, C_6 \in Q_1$, $C_2, C_7, C_8 \in Q_3$, and $C_3 \in Q_4$, which proves the stability of the model.

4.4. Comparative analysis

To illustrate the characteristics and strengths of the proposed method, a comparative analysis with existing studies concerning product ranking across multiple platforms is performed. The comparison results are shown in Table 11.

Previous approaches primarily focused on factors such as the number of evaluators (Zhao et al., 2021) or information entropy (Wu et al., 2023a) to determine platform weights. They rarely considered the quality of reviews, which is a critical factor influencing platform reliability. This study introduces review helpfulness into platform weight measurement, enhancing the comprehensiveness and rationality of the assessment process. In addition, the proposed model leverages PLTSs to represent sentiment analysis results, addressing the inherent ambiguity and uncertainty in online reviews. Using linguistic terms combined with probabilities increases the flexibility and comprehensiveness of sentiment expression, making PLTSs superior to other fuzzy sets such as 2-tuple linguistic model (Jin et al., 2024) and hesitant fuzzy linguistic term sets (Gai et al., 2024) in modeling complex linguistic information under uncertainty. Unlike other methods that failed to address varying review evaluation modes, the proposed model accounts for both detailed reviews and pros/cons reviews. Moreover, existing studies rarely explored service improvement strategies based on multi-platform data. This model addresses the gap by prioritizing service improvements derived from multi-platform insights, offering practitioners an effective and actionable framework to optimize attributes.

In order to further illustrate the effectiveness of the service improvement strategy based on multiple platforms, the proposed method is compared with the IPA model based on a single platform. The attribute classification results based on different platforms are shown in Figure 5a–e.

Table 11. Comparisons with existing studies concerning product ranking across multiple platforms

Method	Platform weighting	Review quality	Sentiment expression	Different evaluation modes	Service improvement
Zhao et al. (2021)	The number of evaluators	Not considered	PLTSs	Not considered	×
Wu et al. (2023a)	Information entropy and the number of participants	Not considered	–	Not considered	×
Yang et al. (2024)	Analytic hierarchy process	Not considered	Basic uncertain linguistic information	Not considered	×
The proposed model	The number of reviews and review helpfulness	Considered	PLTS	Considered	√

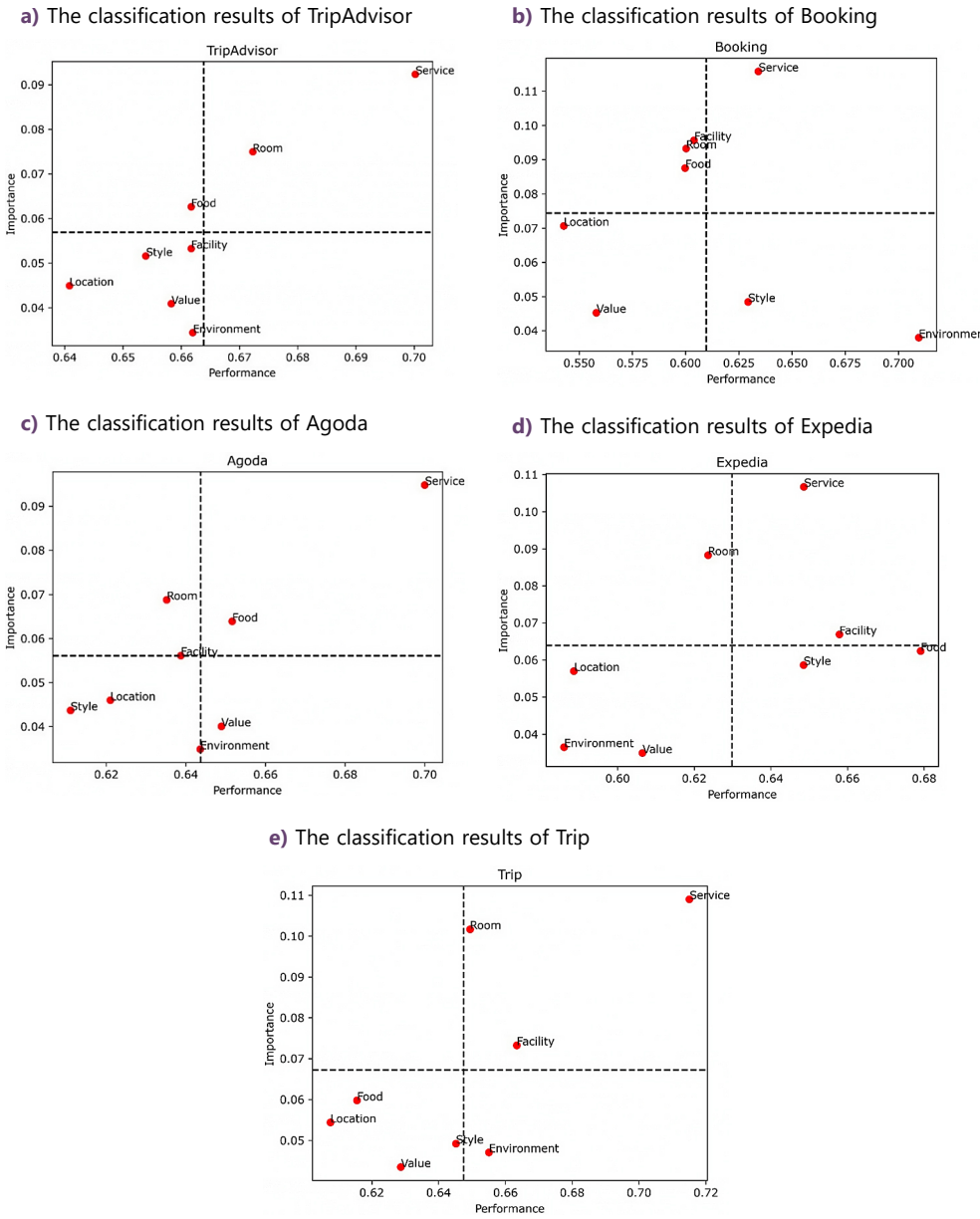

Figure 5. The classification results of each platform

As can be seen from Figure 5a–e, the attributes "service" and "location" belong to "keep up the good work" and "low priority" on each platform, and are the same as the results of multiple platforms. However, other attributes present different categorizations on different platforms. "Room" is located in the area of "concentrate here" on platforms Booking, Agoda, and Expedia, while in "keep up the good work" on platforms Tripadvisor and Trip. The comprehensive result of multiple platforms is "concentrate here", which is identical to most plat-

forms. "Environment" falls in the "low priority" region on platforms Tripadvisor, Agoda, and Expedia, while "possible overkill" in Booking and Trip. "Facility" is categorized in three ways on different platforms, where "keep up the good work" on Expedia and Trip, "concentrate here" on platform Booking and Agoda, and "low priority" on Tripadvisor. The result of multiple platforms is consistent with Expedia and Trip. The categorization of "food" varies significantly across platforms. It belongs to "keep up the good work" on Agoda, "concentrate here" on Tripadvisor and Booking, "low priority" on Trip, and "possible overkill" on Expedia. "Style" belongs to "low priority" on platforms Tripadvisor, Agoda, and Trip, while "possible overkill" on Booking and Expedia. "Value" is located in the area of "low priority" in most cases, while in "possible overkill" area on Agoda. In summary, there are differences in the classification results of attributes among different platforms due to the target customer group, comment mechanism, platform interactivity and atmosphere. Attribute classification based on a single platform cannot incorporate the evaluation information of all target customers, which may lead to a sampling bias and skewed understanding of the market structure, limiting the generalizability of the service improvement strategy. This paper integrates the online reviews from multiple platforms considering different evaluation modes and introduces the reliability of platforms to differentiate the usefulness of knowledge extracted from platforms, which is more reasonable for performing service improvement.

4.5. Managerial implications

By applying the proposed model to hotel service improvement, some interesting insights are uncovered to inform hotel practitioners.

(1) This study evaluates the reliability of platforms from the perspective of the number of reviews and review helpfulness, which provides hotelier insights into reliability measurement. Understanding different reliability of platforms enables hoteliers to allocate management resources to highly reliable platforms. In this paper, the platform with highest reliability is Agoda. The hoteliers should focus on Agoda platform as the platform is considered more reliable by consumers, which could potentially attract more customers.

(2) This study provides the probability distribution of the sentiment polarity of each attribute on each platform, which gives a view of the positive and negative attitudes of consumers towards each attribute on different platforms. This is beneficial for hoteliers to grasp the operational status of the hotel on various aspects and platforms. From the probability distribution in Section 4.2, the attribute "location" of Booking platform has the highest probability of "very negative", implying the attribute "location" performs poorly in meeting consumers' expectation on Booking. In terms of the probability of "very positive", the attribute "environment" on Booking platform is the largest, which indicates a relatively large number of consumers on booking are very satisfied with "environment".

(3) This study integrates evaluation information from various platforms and considers reliability, providing hoteliers with a comprehensive and effective service improvement strategy. Based on the classification results, hoteliers are suggested to perform service improvement according to the prioritization: $C_1 \succ C_4 \sim C_5 \sim C_6 \succ C_2 \sim C_7 \sim C_8 \succ C_3$.

Specifically, hoteliers should prioritize improving the "room" attribute, as it holds high importance but demonstrates poor performance, requiring urgent enhancement to meet customer expectations. Following this, "service", "food", and "facility" should be the next focus areas, as they are critical to maintaining customer satisfaction and require consistent investment to sustain their positive impact. Attributes like "location", "value" and "style" should be considered for gradual improvements, as they are less critical to customers but still show room for enhancement. "Environment" has the lowest priority. Hoteliers can minimize further investment in this attribute and redirect resources toward more critical areas.

5. Conclusions

This paper proposed a service improvement method to formulate service improvement strategies considering the reliability of multiple platforms and different evaluation modes. The proposed method is structured as follows. Firstly, the reliability of platforms was measured by the number of online reviews and review helpfulness, where review helpfulness was evaluated from the perspective of the length of reviews and the consistency of ratings and reviews. Secondly, the sentiments of attributes on each platform were analyzed by the lexicon-based sentiment analysis method considering different evaluation modes, and the PLTSs was utilized to characterize the sentiment analysis results of attributes. The PLTSs were transformed into a precise score by an expectation function and the scores were aggregated with the reliability of platforms to obtain the performance of attributes. After that, the importance of attributes on each platform was derived by the TextRank method and also aggregated using the reliability to obtain the importance of attributes. Last, an IPA model was constructed to classify attributes based on the performance and importance of attributes. To validate the stability of the proposed model, sensitivity analysis was conducted to test the variation of overload points to the attribute classification. To verify the model's effectiveness, the proposed model was compared with the IPA model based on a single platform. In general, the proposed model has the following contributions:

(1) This paper proposed a method for determining platform weights. By incorporating a novel perspective for measuring platform weights, i.e., review helpfulness, which considers the inverted U-shaped relationship between review length and helpfulness as well as the consistency among reviews and ratings, this study enhanced the measurement of platform reliability.

(2) The proposed framework uniquely integrated sentiments from multiple platforms with varying evaluation modes. By considering both detailed reviews and pros/cons reviews and representing sentiment results through PLTSs, the framework provided an effective evaluation of attribute performance across platforms while accounting for diverse consumer feedback.

(3) This study constructed a multi-platform-oriented IPA model. By classifying attributes according to their overall performance and importance, the model overcame the limitations of traditional IPA methods that rely solely on single-platform data. This advancement allowed hoteliers to develop practical and targeted service improvement strategies that effectively account for the structural diversity of platforms and their varying levels of reliability.

This study offers valuable managerial insights for hotel practitioners. It underscores platform reliability as a key factor, guiding hoteliers to allocate resources toward highly reliable platforms like Agoda to attract more customers. Through an analysis of sentiment polarity distributions, the study identifies specific operational weaknesses, such as the underperformance of the "location" attribute on Booking, providing actionable insights for targeted improvements. Additionally, the improvement prioritization recommends focusing on critical attributes like "room" as a top priority, followed by "service", "food", and "facility", while reducing investments in lower-priority attributes such as "environment".

There are some issues worth investigating. This paper only integrated the sentiments extracted from online reviews on different platforms but did not consider ratings of multiple platforms. It is necessary to consider the integration of reviews and ratings from multiple platforms for service improvement in the future. In addition, this paper measured the reliability of the platform based on the number of reviews and review helpfulness, ignoring the impact of false reviews on platform reliability. It is worth studying incorporating the authenticity of reviews in reliability measurement.

Funding

The work was supported by the National Natural Science Foundation of China under Grant 72171158, 72371173, and the Fundamental Research Funds for the Central Universities under Grant No. 2023ZY-SX001.

Author contributions

Shanshang Yang and Huchang Liao proposed the original idea and conceived the study. Shanshang Yang and Huchang Liao were responsible for developing the method, collecting and analyzing the data. Shanshang Yang and Huchang Liao wrote the first draft of the article. Chonghui Zhang revised the paper.

Disclosure statement

The authors have no competing financial, professional, or personal interests from other parties that are related to this paper.

References

Albayrak, T., Cengizci, A. D., Caber, M., & Fong, L. H. N. (2021). Big data use in determining competitive position: The case of theme parks in Hong Kong. *Journal of Destination Marketing & Management*, *22*, Article 100668. https://doi.org/10.1016/j.jdmm.2021.100668

Ban, O. I., Droj, L., Tuşe, D. A., & Botezat, E. (2022). Operationalization of importance-performance analysis with nine categories and tested for green practices and financial evaluation. *Technological and Economic Development of Economy*, *28*(6), 1711–1738. https://doi.org/10.3846/tede.2022.17653

Bi, J.-W., Liu, Y., Fan, Z.-P., & Zhang, J. (2019). Wisdom of crowds: Conducting importance-performance analysis (IPA) through online reviews. *Tourism Management*, *70*, 460–478. https://doi.org/10.1016/j.tourman.2018.09.010

Boley, B. B., & Jordan, E. (2023). Leveraging IPA gap scores to predict intent to travel. *Journal of Hospitality and Tourism Management*, *57*, 97–101. https://doi.org/10.1016/j.jhtm.2023.09.006

Chen, K., Tsai, C.-F., Hu, Y.-H., & Hu, C.-W. (2024). The effect of review visibility and diagnosticity on review helpfulness – An accessibility-diagnosticity theory perspective. *Decision Support Systems*, *178*, Article 114145. https://doi.org/10.1016/j.dss.2023.114145

Darko, A. P., Liang, D., Xu, Z., Agbodah, K., & Obiora, S. (2023). A novel multi-attribute decision-making for ranking mobile payment services using online consumer reviews. *Expert Systems with Applications*, *213*, Article 119262. https://doi.org/10.1016/j.eswa.2022.119262

Feng, Y., Yin, Y., Wang, D., Dhamotharan, L., Ignatius, J., & Kumar, A. (2023). Diabetic patient review helpfulness: Unpacking online drug treatment reviews by text analytics and design science approach. *Annals of Operations Research*, *328*, 387–418. https://doi.org/10.1007/s10479-022-05121-4

Gai, T., Wu, J., Liang, C., Cao, M., & Zhang, Z. (2024). A quality function deployment model by social network and group decision making: Application to product design of e-commerce platforms. *Engineering Applications of Artificial Intelligence*, *133*, Article 108509. https://doi.org/10.1016/j.engappai.2024.108509

Ganguly, B., Sengupta, P., & Biswas, B. (2024). What are the significant determinants of helpfulness of online review? An exploration across product-types. *Journal of Retailing and Consumer Services*, *78*, Article 103748. https://doi.org/10.1016/j.jretconser.2024.103748

Glaveli, N., Manolitzas, P., Palamas, S., Grigoroudis, E., & Zopounidis, C. (2023). Developing effective strategic decision-making in the areas of hotel quality management and customer satisfaction from online ratings. *Current Issues in Tourism*, *26*(6), 1003–1021. https://doi.org/10.1080/13683500.2022.2048805

Jia, H., Shin, S., & Jiao, J. (2022). Does the length of a review matter in perceived helpfulness? The moderating role of product experience. *Journal of Research in Interactive Marketing*, *16*(2), 221–236. https://doi.org/10.1108/JRIM-04-2020-0086

Jin, W., Gai, T., Cao, M., Zhou, M., & Wu, J. (2024). A personalized bidirectional feedback mechanism by combining cooperation and trust to improve group consensus in social network. *Computers & Industrial Engineering*, *188*, Article 109888. https://doi.org/10.1016/j.cie.2024.109888

Kou, G., Yang, P., Peng, Y., Xiao, H., Xiao, F., Chen, Y., & Alsaadi, F. E. (2021). A cross-platform market structure analysis method using online product reviews. *Technological and Economic Development of Economy*, *27*(5), 992–1018. https://doi.org/10.3846/tede.2021.12005

Li, M., & Huang, P. (2020). Assessing the product review helpfulness: Affective-cognitive evaluation and the moderating effect of feedback mechanism. *Information & Management*, *57*(7), Article 103359. https://doi.org/10.1016/j.im.2020.103359

Liang, Y. (2024). Crowdsourcing incentive mechanisms for cross-platform tasks: A weighted average maximization approach. *Engineering Applications of Artificial Intelligence*, *133*, Article 108008. https://doi.org/10.1016/j.engappai.2024.108008

Liu, H., Wu, S., Zhong, C., & Liu, Y. (2023a). The effects of customer online reviews on sales performance: The role of mobile phone's quality characteristics. *Electronic Commerce Research and Applications*, *57*, Article 101229. https://doi.org/10.1016/j.elerap.2022.101229

Liu, Z., Liao, H., Li, M., Yang, Q., & Meng, F. (2023b). A deep learning-based sentiment analysis approach for online product ranking with probabilistic linguistic term sets. *IEEE Transactions on Engineering Management*, *71*, 6677–6694. https://doi.org/10.1109/TEM.2023.3271597

Lu, L., Xu, P., Wang, Y. Y., & Wang, Y. (2023). Measuring service quality with text analytics: Considering both importance and performance of consumer opinions on social and non-social online platforms. *Journal of Business Research*, *169*, Article 114298. https://doi.org/10.1016/j.jbusres.2023.114298

Luo, Y., He, J., Mou, Y., Wang, J., & Liu, T. (2021). Exploring China's 5A global geoparks through online tourism reviews: A mining model based on machine learning approach. *Tourism Management Perspectives*, *37*, Article 100769. https://doi.org/10.1016/j.tmp.2020.100769

Lutz, B., Pröllochs, N., & Neumann, D. (2022). Are longer reviews always more helpful? Disentangling the interplay between review length and line of argumentation. *Journal of Business Research*, *144*, 888–901. https://doi.org/10.1016/j.jbusres.2022.02.010

Ma, B., Wong, Y. D., Teo, C.-C., & Wang, Z. (2024). Enhance understandings of Online Food Delivery's service quality with online reviews. *Journal of Retailing and Consumer Services*, *76*, Article 103588. https://doi.org/10.1016/j.jretconser.2023.103588

Martilla, J. A., & James, J. C. (1977). Importance-performance analysis. *Journal of Marketing*, *41*(1), 77–79. https://doi.org/10.1177/002224297704100112

Mejia, C., Bąk, M., Zientara, P., & Orlowski, M. (2022). Importance-performance analysis of socially sustainable practices in US restaurants: A consumer perspective in the quasi-post-pandemic context. *International Journal of Hospitality Management*, *103*, Article 103209. https://doi.org/10.1016/j.ijhm.2022.103209

Mirtalaie, M. A., Hussain, O. K., Chang, E., & Hussain, F. K. (2018). Extracting sentiment knowledge from pros/cons product reviews: Discovering features along with the polarity strength of their associated opinions. *Expert Systems with Applications*, *114*, 267–288. https://doi.org/10.1016/j.eswa.2018.07.046

Pang, Q., Wang, H., & Xu, Z. (2016). Probabilistic linguistic term sets in multi-attribute group decision making. *Information Sciences*, *369*, 128–143. https://doi.org/10.1016/j.ins.2016.06.021

Pimpalkar, A., & Jeberson Retnaraj, R. (2022). MBiLSTMGloVe: Embedding GloVe knowledge into the corpus using multi-layer BiLSTM deep learning model for social media sentiment analysis. *Expert Systems with Applications*, *203*, Article 117581. https://doi.org/10.1016/j.eswa.2022.117581

Quan, L. J., Kim, J. J., & Han, H. (2022). Customer views on comprehensive green hotel selection attributes and analysis of importance-performance. *Journal of Travel & Tourism Marketing*, *39*(6), 535–554. https://doi.org/10.1080/10548408.2022.2162657

Verma, D., Dewani, P. P., Behl, A., Pereira, V., Dwivedi, Y., & Del Giudice, M. (2023). A meta-analysis of antecedents and consequences of eWOM credibility: Investigation of moderating role of culture and platform type. *Journal of Business Research*, *154*, Article 113292. https://doi.org/10.1016/j.jbusres.2022.08.056

Wu, X., & Liao, H. (2019). A consensus-based probabilistic linguistic gained and lost dominance score method. *European Journal of Operational Research*, *272*(3), 1017–1027. https://doi.org/10.1016/j.ejor.2018.07.044

Wu, J., & Yang, T. (2023). Service attributes for sustainable rural tourism from online comments: Tourist satisfaction perspective. *Journal of Destination Marketing & Management*, *30*, Article 100822. https://doi.org/10.1016/j.jdmm.2023.100822

Wu, X., Liao, H., & Tang, M. (2023a). Decision making towards large-scale alternatives from multiple online platforms by a multivariate time-series-based method. *Expert Systems with Applications*, *212*, Article 118838. https://doi.org/10.1016/j.eswa.2022.118838

Wu, X., Liao, H., & Zhang, C. (2023b). Importance-performance analysis to develop product/service improvement strategies through online reviews with reliability. *Annals of Operations Research*, *342*, 1905–1924. https://doi.org/10.1007/s10479-023-05594-x

Wu, D. C., Cao, C., Wu, J., & Hu, M. (2024a). Wine tourism experiences of Chinese tourists: A tourist-centric perspective. *International Journal of Contemporary Hospitality Management*, *36*(8), 2601–2631. https://doi.org/10.1108/IJCHM-07-2023-1003

Wu, X., Liao, H., & Tang, M. (2024b). Product ranking through fusing the wisdom of consumers extracted from online reviews on multiple platforms. *Knowledge-Based Systems*, *284*, Article 111275. https://doi.org/10.1016/j.knosys.2023.111275

Salimi, N. (2021). Opportunity recognition for entrepreneurs based on a business model for sustainability: A systematic approach and its application in the Dutch dairy farming sector. *IEEE Transactions on Engineering Management*, *70*(11), 3728–3744. https://doi.org/10.1109/TEM.2021.3082872

Shin, J., Joung, J., & Lim, C. (2024). Determining directions of service quality management using online review mining with interpretable machine learning. *International Journal of Hospitality Management*, *118*, Article 103684. https://doi.org/10.1016/j.ijhm.2023.103684

Yang, Z., Ouyang, T., Fu, X., & Peng, X. (2020). A decision-making algorithm for online shopping using deep-learning-based opinion pairs mining and q-rung orthopair fuzzy interaction Heronian mean operators. *International Journal of Intelligent Systems*, *35*(5), 783–825. https://doi.org/10.1002/int.22225

Yang, Y., Xia, D.-X., Pedrycz, W., Deveci, M., & Chen, Z.-S. (2024). Cross-platform distributed product online ratings aggregation approach for decision making with basic uncertain linguistic information. *International Journal of Fuzzy Systems*, *26*, 1936–1957. https://doi.org/10.1007/s40815-023-01646-3

Zhang, C., Xu, Z., Gou, X., & Chen, S. (2021). An online reviews-driven method for the prioritization of improvements in hotel services. *Tourism Management*, *87*, Article 104382. https://doi.org/10.1016/j.tourman.2021.104382

Zhang, M., Sun, L., Wang, G. A., Li, Y., & He, S. (2022a). Using neutral sentiment reviews to improve customer requirement identification and product design strategies. *International Journal of Production Economics*, *254*, Article 108641. https://doi.org/10.1016/j.ijpe.2022.108641

Zhang, Y., Liang, D., & Xu, Z. (2022b). Cross-platform hotel evaluation by aggregating multi-website consumer reviews with probabilistic linguistic term set and Choquet integral. *Annals of Operations Research*, 1–35. https://doi.org/10.1007/s10479-022-05075-7

Zhang, D. F., Shen, Z. F., & Li, Y. (2023). Requirement analysis and service optimization of multiple category fresh products in online retailing using importance-Kano analysis. *Journal of Retailing and Consumer Services*, *72*, Article 103253. https://doi.org/10.1016/j.jretconser.2022.103253

Zhang, C. X, & Xu, Z. S. (2024). Gaining insights for service improvement through unstructured text from online reviews. *Journal of Retailing and Consumer Services*, *80*, Article 103898. https://doi.org/10.1016/j.jretconser.2024.103898

Zhao, M., Li, L., & Xu, Z. (2021). Study on hotel selection method based on integrating online ratings and reviews from multi-websites. *Information Sciences*, *572*, 460–481. https://doi.org/10.1016/j.ins.2021.05.042

Zhao, M., Liu, M., Xu, C., & Zhang, C. (2024). Classifying travellers' requirements from online reviews: An improved Kano model. *International Journal of Contemporary Hospitality Management*, *36*(1), 91–112. https://doi.org/10.1108/IJCHM-06-2022-0726