Research on Customer Satisfaction of Budget Hotels Based on Revised IPA and Online Reviews

Xue Liu, Ning Zhang*

Department of Management Science and Engineering, Business School of Qingdao University, Qingdao, China

Email address:
1481571639@qq.com (Xue Liu), shiny548984@163.com (Ning Zhang)
*Corresponding author

Abstract: Online reviews are the emotional expressions of customers after product or service experience. Compared with survey questionnaires, they can more truly reflect customers' perception of product or service. Therefore, combined with online reviews and importance-performance analysis (IPA), managers can make corresponding corporate strategic according to the priority of features. This research uses the Meituan.com hotel online reviews of budget hotels as an example. First, we uses natural language processing technology to preprocess online reviews, and uses K-means to build a feature lexicon. Second, based on the sentiment dictionary, we perform fine-grained sentiment analysis on "feature-view pairs" to obtain feature satisfaction scores. Third, combined with the revised IPA, we obtain implicitly derived importance, and then the priority of each feature improvement is determined. The conclusions show that (1) service, location, and price are the advantages of budget hotels. Managers should maintain a competitive advantage and ensure the supply of resources. (2) Catering and room facilities are the main disadvantages of budget hotels. Managers should improve these two features to improve customer satisfaction as soon as possible. This study implements the method of managing IPA through online reviews, which replaces the previous questionnaire method. At the same time, revised IPA provides more realistic and concrete reference for hotel managers when making decisions.

Keywords: Budget Hotel, Customer Satisfaction, Online Review, Revised IPA

1. Introduction

With the rapid development and popularity of the Web 2.0 era, people are more inclined to post their ideas and interact on social platforms. Previous studies have shown that online reviews are an important form of social interaction [1], so online customer reviews (OCRs) become a major source of information for consumers and industry managers. Compared with the objective descriptions published by merchants, OCRs are derived from the emotional expression of consumers after the experience, so they are more trustworthy and persuasive, especially for experiential products. Hotels are one of the most typical experiential products, OCRs will have a significant impact on the consumer behavior of potential customers.

Marketing and economic theory [2] believe that products and services have multi-dimensional features, and consumers' preferences for each feature are different, so the degree of the emotional expression of each feature displayed in OCRs will be different. Recognizing the importance of multi-dimensional features can be more accurately improved according to consumer needs. At present, most consumer satisfaction surveys are in the form of questionnaires. Some researches set a large number of items in the questionnaire to investigate the comprehensiveness and accuracy of the information, such as the classic service quality model (SERVQUAL) [3], and its theory is relatively strong, there are certain limitations in both the size of the data and the richness of the content. Summarizing the existing research, it is found that the researches on hotel satisfaction have been more in-depth, but most of them use questionnaires or directly use website numerical scores, and rarely analyze the text of OCRs.

At present, the most widely used hotel satisfaction measurement model is importance-performance analysis (IPA), which was proposed by Martilla and James in 1977 [4], and has been widely used in many fields since then. IPA mainly determines the improvement strategy to achieve the optimal use of resources by comparing the satisfaction and importance values of each feature. Although IPA is widely

used, it also has certain defects [5]. When mapping each feature to a matrix figure, it is required that the two coordinates, that is, importance and satisfaction are independent of each other. However, due to the similarity between importance and satisfaction, the interviewees could not distinguish the two well, so that the final quadrant distribution results were also biased. Based on the above problems, many scholars have revised the IPA analysis method. Deng has summarized the research results of other scholars and optimized the IPA analysis method statistically [6]. The revised model has achieved good results.

This research uses the reviews of Meituan.com to conduct research to explore consumer preferences and attention to hotel features. First, combined with natural language processing, machine learning and other methods, we process review texts. Second, we transform the crawled hotel reviews into word vectors through Word2Vec, and use the K-means method to cluster the word vectors. Third, the sentiment dictionary method is used to assign features to obtain features satisfaction. Finally, according to revised IPA to acquire importance, the hotel features are judged through the IPA matrix figure, so as to provide more accurate and reasonable suggestions to the hotel industry.

2. Related Research

2.1. Features of Budget Hotels and Consumer Satisfaction

The features of the hotel are divided into many categories, such as intangible features and tangible features; practical features and experience features. The perception of hotel is the evaluation of the importance and satisfaction of all the features by consumers during the process of staying in the hotel. Callan and Bowman [7] conducted a survey of 38 features of the hotel. They found that the features that consumers consider important include employee service attitude, service efficiency, health environment, etc. At the same time, many respondents said that the experiential features are relative important. Rhee and Yang [8] researched the literature on hotel features and summarized the six categories that consumers are more concerned about. At the same time, many tourism websites also used these six categories of hotel features to gauge consumer evaluation. Consumers' perception of the importance and satisfaction of hotel features is uncertain. Some features may have high satisfaction but may not be very important to consumers. Some features are important but their existence and optimization do not improve satisfaction. The KANO model proposed by Kano [9] explains the above features. KANO model is classified according to the provision of product attribute and consumer satisfaction, it mainly includes: basic attributes, performance attributes, and excitement attributes, where the basic attributes and the excitement attributes are non-linearly related. Combining and summarizing the existing literature, we found that there are more domestic researches on the tourism market, but less research on the features of budget hotels. Although there are literature survey on domestic consumers' needs and satisfaction in hotels, they only studies specific features. Gao et

al. [10] found that hotel location is an important feature that consumers will pay special attention to when making decisions. Shan et al. [11] researched the four features of hotel type and room size, etc. when portraying user portraits of online reviews for hotels. In general, the analysis of all features of budget hotels lacks systematic and in-depth research. Therefore, in the domestic hotel industry research, it is necessary to further explore consumer perceptions of budget hotel features.

2.2. Traditional IPA and Revised IPA

The IPA method uses the satisfaction and importance of features as a combination evaluation of various factors, and analyzes each factor based on the quadrant distribution in the matrix, thereby to find out the factors that need urgent improvement. The IPA method is relatively widely used at home and abroad. Wu [12] studied typical domestic budget hotels and found that the most important thing for consumers is still the hotel location. Kuo et al. [13] explored the factors affecting Hong Kong consumers' choice of hotel accommodation, and used IPA analysis to study 26 hotel service features, and found that the hotel's location, environment, catering and other aspects urgently need to be improved and optimized.

There are two necessary prerequisites for the use of IPA analysis. First, the two coordinate axes in the matrix diagram are required to be independent of each other; second, the relationship between the satisfaction of each feature and overall satisfaction must be linearly related and symmetrical. In reality, because of the similarity between importance and satisfaction, consumers cannot distinguish them well. Many scholars have revised and expanded the IPA analysis method. Deng summarized previous research and revised the problems existing in traditional IPA. He proposed to replace the self-reported importance with the implicitly derived importance, that is, to analyze the correlation logarithm of the satisfaction evaluation of each factor, and then conduct a partial correlation analysis of overall satisfaction and the satisfaction of each feature. The obtained partial correlation coefficient is used as the implicitly derived importance [6]. The partial correlation coefficient excludes the impact of other satisfaction variables on the correlation between the specified variable and the overall satisfaction, and can reflect the true situation of each feature [14].

2.3. Sentiment Analysis of Online Reviews

Sentiment analysis, also is known as opinion mining or comment extraction. It is an objective and subjective analysis of unstructured text data, and classifies an emotional polarity of the extracted subjective sentences. The purpose of sentiment analysis is to obtain praise or disapproval opinions on a certain commodity, and to provide a basis for decision-making [15]. Sentiment analysis is divided into coarse-grained sentiment analysis and fine-grained sentiment analysis. Coarse-grained sentiment analysis includes chapter-level and sentence-level sentiment analysis. Fine-grained sentiment analysis is based on the analysis of

feature level. Medhat et al. [16] believe that the main steps in fine-grained analysis of online reviews are: emotion recognition, feature extraction, emotion classification, and emotion polarity recognition. Zhao [17] used the DBSCAN-based text clustering process to mine the hotel's reputation dimension. The results show that consumers are more concerned about the features include hardware, service, environment, diet and value. This research draws on the previous method based on sentiment dictionary, and calculates sentiment scores based on sentiment polarity and intensity to obtain satisfaction which are used to analyze the revised IPA.

2.4. K-means Features Clustering

K-means clustering is an unsupervised machine learning algorithm. Because of its simple principle, easy implementation, and good results, it is a more commonly used clustering method. It calculates the distance between objects and judges the similarity between them, and then classifies them. The main purpose of clustering is to divide the data set into K classes, so that the data within the class is smallest, and the inter-class clustering is the largest.

The main process of K-means clustering includes four steps. First, determining the number of clusters K, and then it can randomly select K initial points as the cluster centers. Second, calculating the distance between each point in the data set and K initial points, they can assign the points to the cluster center with the smallest distance. Third, after classifying all points in the data set, it should recalculate the cluster center. Finally, repeating second and third steps until the cluster center does not change, then it is considered that the optimal clustering has been reached.

3. Research Design

The overall research framework of this paper mainly includes text preprocessing, Word2Vec, K-means clustering, fine-grained sentiment analysis, and partial correlation analysis. The concrete process is shown in Figure 1. The stage1 main includes data processing and features extraction. The stage2 main obtain satisfaction and implicitly derived importance of features. In stage3, we construct the IPA figure to analysis the features of hotel and provide suggestions.
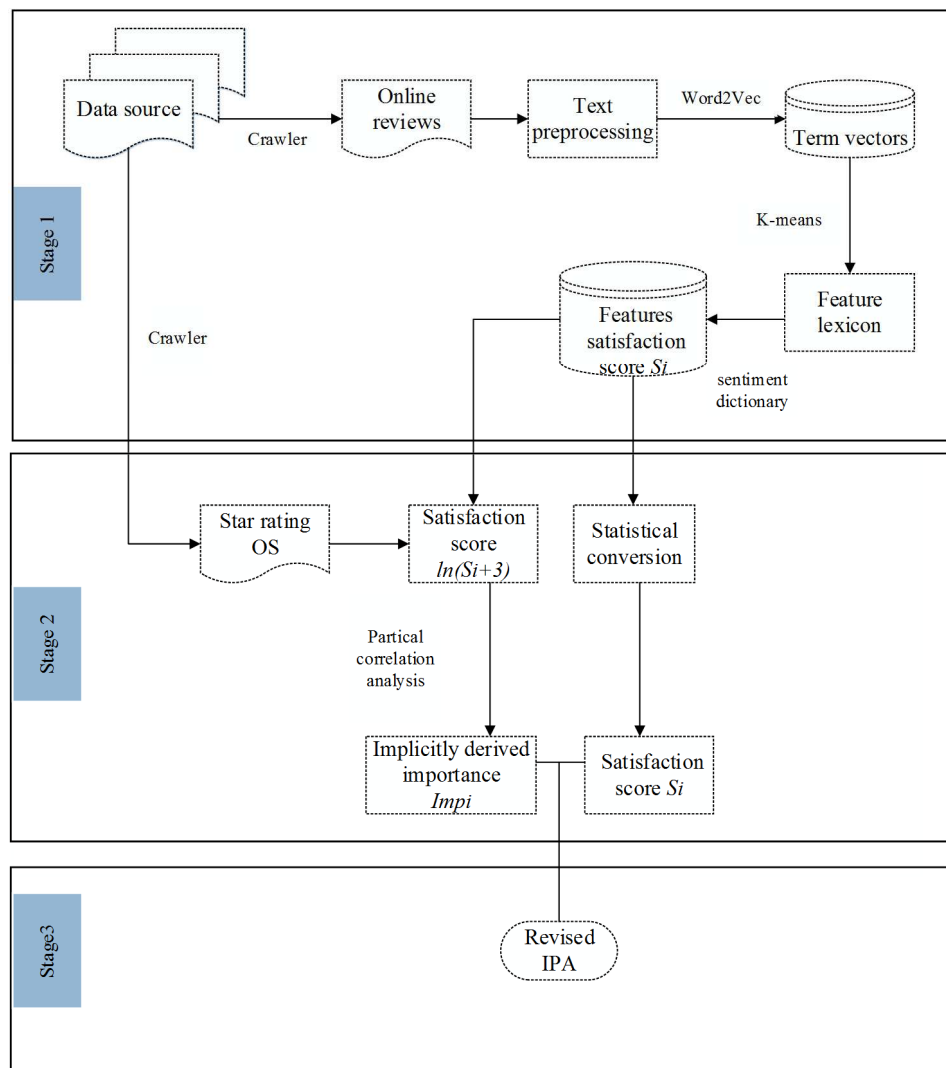

Figure 1. Research framework.

3.1. Data Sources

The representative brands of Chinese budget hotels including Home Inn, Hanting Inn and 7Days Inn, etc. Therefore, this study chooses 21 budget hotels from Qingdao including Home Inn, Hanting Inn and 7Days Inn, etc. By writing a Python program for the web crawler, a total of 35,398 online reviews were obtained, including text reviews, star scores, and review usefulness scores. Through clearing all data and deleting invalid reviews, this research finally gets 30,877 reviews.

3.2. Hotel Features Clustering Based on K-means

1. Online reviews are segmented form a Word2Vector corpus, and then we use the Word2Vec technology trains the segmented corpus (window = 10, vector = 300), finally, the words map to the K-dimensional vector space to form the corresponding word vector.
2. K-means algorithm is used clusters nouns in the dictionary. First we need to determine the number of clusters of K-means. There are two best clustering methods: the elbow method and the contour method. The core idea of the elbow method is that as the number of clusters increases, the sample segmentation will be more accurate, and the sum of the squared errors will gradually become smaller. When the number of true clusters is reached, the degree of aggregation obtained by increasing the number of clusters will rapidly decrease, so the decrease in the sum of squared errors will decrease sharply. Then it becomes flat as the number of clusters increases, that is, the elbow corresponds to the true number of clusters in the data. The core index of the contour method is the Silhouette Coefficient. The larger the average contour coefficient, the better the clustering effect. Therefore, the maximum average contour coefficient is the optimal number of clusters.

Because the number determined by the contour method is not necessarily the optimal, it is sometimes necessary to judge by SSE, so the elbow method is selected to determine the optimal number of clusters. The number of clusters is set to 1 to 15 for repeated prediction. The final results show that when the number of clusters is 13, the number of clusters is optimal. Therefore, this paper divides features into 13 categories for research.

The process of K-means clustering is as follows: (1) Segmenting the reviews, the nouns in the reviews are selected as candidate words for feature clustering. And keep the nouns that appear at least 10 times to form a noun dictionary. (2) After the words in the dictionary are trained by Word2Vec, they are transformed into word vectors as a clustered corpus. (3) The K-means algorithm is used to cluster the word vectors. According to the optimal number of clusters, the number of clusters is 13 and judged based on the distance formula. At the same time, referring to previous related research and manual classification, the final hotel feature is classified. The concrete classification is shown in Table 1.

Table 1. Hotel Features and representative feature words.

Code	Feature	Representative feature words
1	Catering	Breakfast, yogurt...
2	Bedding	Sheets, quilts...
3	Service	Attitude, politeness...
4	Insulation	Voice, movement...
5	Environment	Construction, environment...
6	Price	Money, price...
7	Infrastructure	Elevators, corridors...
8	Room Facility	Heating, TV...
9	Cleanliness	Toilet, toilet paper...
10	Brand	Reputation, brand...
11	Network	Network, wireless...
12	Bathroom	Bathroom, sprinkler...
13	Location	Squares, scenic spots...

3.3. Sentiment Analysis Based on Sentiment Dictionary

Consumers usually express their opinions on specific features when they post reviews, so "feature-view pairs" can be extracted as evaluation units. The central idea of sentiment analysis is to judge sentiment through adjectives and adverbs closest to feature nouns. After extracting the "feature-view pairs" of each sentence, we will classify the features by sentiment dictionary. In order to make the classification results more accurate, the words related to hotels are added in this research to form a special dictionary for hotels. The concrete construction principles are as follows.

HowNet2007 [18] dictionary is used as the basic dictionary. The positive basic sentiment dictionary is a combination of positive evaluation words and positive emotional words. The negative basic sentiment dictionary is a combination of negative evaluation words and negative emotional words. In the end, 7020 positive emotion words and 5949 negative emotion words are obtained. In order to improve the classification accuracy of the hotel-specific sentiment dictionary, manual processing is required: all valid reviews are processed by word segmentation, etc., and then word frequency statistics are performed on the adjectives, and the sentiment polarity of the selected adjectives is judged to form a special sentiment dictionary.

Except for nouns and adjectives, consumers have different emotional strengths when they post reviews. Therefore, this research determines the quantitative standard of emotional polarity based on the emotional level of degree adverbs in HowNet2007, as shown in Table 2. The sentiment polarity score of the feature is equal to the degree adverb multiplied by the sentiment polarity value, if there is no degree adverb, the score is 1 or -1. The extracted "feature-view pairs" are calculated, and then calculate the average of all features in a category to get the sentiment score of every category of the hotel.

Table 2. Sentiment polarity quantification criteria.

Degree adverb	Degree value
Very, absolutely...	2
Especially, really...	1.5
More, even more...	1
Slightly, a little...	0.5

Degree adverb	Degree value
Positive emotion words without degree adverbs	1
Negative emotion words without degree adverbs	-1

3.4. Implicitly Derived Importance and Satisfaction of Features

Through the above experiments, 13 categories of hotel features are obtained. The sentiment score represents the satisfaction score of each feature. The correlation test is performed on the extracted 13 categories of features. There are a total of 78 correlation coefficients, of which there are only 8 features that are not related to each other, including: catering-infrastructure, catering-room facilities, catering-network, insulation-network, environment-network, brand-bathroom, brand-location and network-location.

In order to resolve the mutual influence between the features, the existing data was revised according to Deng's conversion method, and the implicitly derived importance is used to replace the self-stated importance. The conversion method mainly has two steps: First, the natural logarithm $\ln(Si)$ is taken for each feature. Since the sentiment score in this paper is calculated, the minimum positive score is -2, therefore, $\ln(Si + 3)$ is used as the independent variable of each attribute to make it linearly distributed. Secondly, using $\ln(Si + 3)$ and overall satisfaction OS, we perform partial correlation analysis. The overall satisfaction score is obtained by the crawler for the hotel star rating, with an integer ranging from 1 to 5, where 1 is very dissatisfied and 5

is very satisfied. Partial correlation coefficients obtained from partial correlation analysis are made to be implicitly derived importance. The implicitly derived importance and satisfaction scores of hotel features are shown in Table 3.

Table 3. Implicitly derived importance and satisfaction score.

Code	Feature	Implicitly derived importance	Satisfaction
1	Catering	0.127	-0.022
2	Bedding	0.036	0.084
3	Service	0.162	0.464
4	Insulation	0.003	0.046
5	Environment	0.069	0.269
6	Price	0.083	0.175
7	Infrastructure	0.060	0.102
8	Room Facility	0.073	0.105
9	Cleanliness	0.056	-0.026
10	Brand	0.069	0.123
11	Network	0.011	0.022
12	Bathroom	0.054	-0.010
13	Location	0.111	0.386

3.5. Revised IPA

The satisfaction is taken as the horizontal axis and implicitly derived importance is taken as the vertical axis. The mean value of all feature satisfaction and implicitly derived importance are used as the center points to divide the matrix into four quadrants. The 13 features are mapped in the matrix quadrant according to the scores of each feature in Table 3. The results are shown in Figure 2.
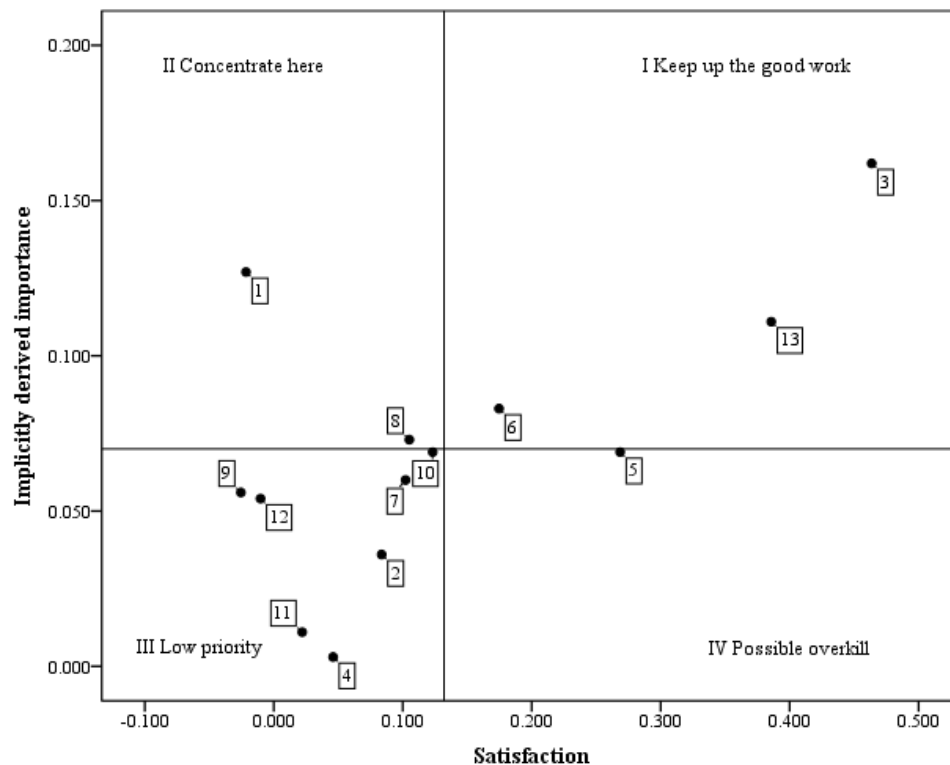

Figure 2. IPA analysis of implicitly derived importance and satisfaction.

According to the IPA analysis figure, the satisfaction and importance of the quadrant I are high, including three features:

service (3), location (13), and price (6). Among them, the satisfaction and importance of service are the highest,

reflecting the customer's attention to service personnel and service attitude. With the development of society and the improvement of civilized quality, customers pay more attention to the requirements of intangible features and more emphasis on the consumption of things. As a typical experience product, hotels should pay more attention to the improvement of service. The second is the hotel location. The hotel location has always been used as an important reference feature. In the past research on hotel features, the "hotel location" was also researched separately. Therefore, location is very important. It is more inclined to choose hotels with convenient transportation, around the scenic area or surrounding facilities. Finally, for the price, the rapid development of budget hotels in recent years has benefited from its price positioning. Compared with other types of hotels, such as four-star and high-end hotels, budget hotels have obvious price advantages. Since most consumers' economic conditions are relatively ordinary, and they are still sensitive to price issues. Therefore, budget hotels should continue to give play to their price advantages, and strive to optimize based on the market share they have already occupied. Because the features of the quadrant I represent the competitive advantage of the company, the strategy adopted is "keep up the good work."

The quadrant II is low satisfaction and high importance, and includes two features: catering (1) and room facilities (8). Catering is essential for people to travel. The initial service mode of budget hotels is accommodation and breakfast. With the modern people's pursuit of healthy living, breakfast has become the standard of daily life. Therefore, customers are paying more attention to the free food and beverage provided by hotels. Because the features of quadrant II represent the aspects that enterprises urgently need to improve, if they are ignored, it will pose a serious threat to the development of the enterprise. People's demand for hotel catering is not harsh, and does not require the variety and richness of catering types, therefore, it is necessary for hotel managers to provide a relatively simple diet and improve it. For room facilities, such as air conditioners, kettles, and other basic room types, they are the basis for the consumer experience and the most basic service of the hotel. They directly affect consumers' overall perception and satisfaction of the hotel. The strategy adopted by this quadrant is "concentrate here."

The quadrant III, that is, satisfaction and importance are relatively low, including seven features: brand (10), infrastructure (7), cleanliness (9), bathroom (12), bedding (2), network (11), and insulation (4). Infrastructure such as parking lots, bathrooms such as bathing facilities, beddings such as quilts, and insulation of rooms, are the basic equipment of the hotel, so they do not attract much attention. Customers are accustomed to these essential features, so they are of low importance and have a low impact on overall satisfaction. Regarding brand feature, this research finds that a certain amount of online reviews will mention the hotel name, such as "Hanting Inn", "Home Inn", etc. Therefore, customers will be affected by the brand to a certain extent when making consumer choices, but most customers prioritize other factors

before considering the brand. The cleanliness feature is also an important basic condition. Under the condition that other basic features are provided, the cleanliness of the hotel should reach a certain standard. Network as an indispensable factor for modern cities, especially tourism cities, it can already reach a high coverage rate, and it is no longer an additional feature that customers pay attention to. Therefore, the strategy adopted for the features of the quadrant III is "low priority", and it can be improved under the condition of sufficient hotel resources.

The quadrant IV is high satisfaction and low importance, and includes one feature: environment (5). The environment may be noisy in places with convenient transportation or around the scenic area, and the environment may be relatively quiet in relatively remote places, however, there is also less convenient transportation. Because location is a feature that customers pay great attention to when consuming, and the convenience of location contradicts the quietness of the environment to a certain extent. Therefore, the general customer will give priority to the hotel location and ignore the quiet environment. It is considered to be "possible overkill". When the resources of the hotel are limited, the features of the other quadrants are prioritized, and the excessive resources are transferred to other aspects to increase the overall satisfaction of the hotel.

4. Conclusion

The IPA analysis method is widely used because of its simplicity and ease of operation. However, due to the limitations of the IPA method and the complexity of traditional questionnaires, the use of IPA has also been restricted. Therefore, from the perspective of management, the focus of this article is the empirical application of Bi et al.'s [19] method of introducing online reviews into IPA. Compared with the questionnaire, online reviews are more authentic and easier to obtain. Based on a large amount of data, more objective and accurate results can be obtained. Hotel managers determine more accurate and reasonable improvement strategies based on the objective and precise results, while effectively reducing the company's human, material, and time costs. Secondly, we use the existing principles and methods to revise IPA to meet the assumptions of the use of IPA, so as to make the conclusions more accurate. From the perspective of practice, it is possible to understand the current development status of budget hotels in Qingdao more realistically and effectively. Based on a more trustworthy conclusion, it is proposed the improvement strategy. The strategy can obtain higher satisfaction through meeting customer needs as much as possible and provide reference for the improvement of hotels for Qingdao to build a national central city.

The research in this paper also has certain limitations. Firstly, the selected hotel sample is only targeted at the economy of Qingdao, and cannot represent the overall development of the hotel industry in Qingdao. Follow-up research and analysis will be conducted for different types of

hotels. Secondly, although the revised IPA referenced in this article includes the three-factor theory of asymmetry, it cannot accurately reflect the asymmetry between the satisfaction and importance of each feature. Therefore, we will carry out more in-depth research on the revision method of IPA.

References

[1] Godes, D., Mayzlin, D., Chen, Y., Das, S., Dellarocas, C., & Pfeiffer, B., et al. (2005). The Firm's Management of Social Interactions. *Marketing Letters*, 16, 415-428.

[2] Jerdee, T. H., & Rosen, B. (1974). Effects of Opportunity to Communicate and Visibility of Individual Decisions on Behavior in the Common Interest. *Journal of Applied Psychology*, 59 (6), 712-716.

[3] Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality. *Journal of Retailing*, 64 (1), 12-40.

[4] Martilla J A, James J C. Importance-Performance Analysis. (1977). *Journal of Marketing*, 41 (1): 77-79.

[5] Sever, I. (2015). Importance-Performance Analysis: A Valid Management Tool?. *Tourism Management*, 48, 43-53.

[6] Deng W J. (2007) Using a Revised Importance-Performance Analysis Approach: The Case of Taiwanese Hot Springs Tourism. *Tourism Management*, 28 (5), 1274-1284.

[7] Callan, R. J., & Bowman, L. (2000). Selecting a Hotel and Determining Salient Quality Attributes: A Preliminary Study of Mature British Travelers. *International Journal of Tourism Research*, 2 (2), 97-118.

[8] Rhee, H. T., Yang, S. B. (2015). Does Hotel Attribute Importance Differ By Hotel? Focusing on Hotel Star-Classifications and Customers' Overall Ratings. *Computers in Human Behavior*, 50, 576-587.

[9] Kano N, Seraku N, Takahashi F, et al. Attractive Quality and Must-Be Quality.(1984). *The Journal of Japanese Society for Quality Control*, 14: 39-48.

[10] Gao, B. J., Sun, H. L., and Wang, H. N. Influence of Online Reviews on Hotels' Full-Occupancy Rates. (2016). *Tourism Tribune*, 31 (4): 109-117.

[11] Shan, X. H., Zhang, X. Y., and Liu, X. Y. (2018). Research on User Portrait Based on Online Review: Taking Ctrip Hotel as an Example. *Information Studies: Theory & Application*, 41 (4).

[12] Wu, J. F. (2007). IPA on the Economy Hotel Service Quality in China——Based on the Survey of Home Inn &Jinjiang Inn. *East China Economic Management*, 21 (11): 95-98.

[13] Kuo, C. M., Chen, L. C., & Lin, S. Y. (2010). Exploring the Relationship Between Hotel-Based Service Attribute Importance and Customer Satisfaction at International Conferences in Taiwan. *Journal of Convention & Event Tourism*, 11 (4), 293-313.

[14] Chen, X. (2013). The Modified Importance-performance Analysis Method and its Application in Tourist Satisfaction Research. *Tourism Tribune*, 28 (11): 59-66.

[15] Tang, X. B. and Liu, G. C. (2017). Research Review on Fine-Grained Sentiment Analysis. *Library and Information Service*, 5: 132-140.

[16] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 5 (4), 1093-1113.

[17] Zhao, X. F., Tang, Q., Zhang, R., and Li, Y. (2012). Exploration of Dimensions of the Online Hotel Reputation Based on Customers' Text Comments and Corpus. *Library and Information Service*, 56 (12): 124-129.

[18] Dong, Z. D. CNKI How Net. http://www.keenage.com.

[19] Bi J W, Liu Y, Fan Z P, et al. (2019). Wisdom of Crowds: Conducting Importance-Performance Analysis (IPA) Through Online Reviews. *Tourism Management*, 70, 460-478.