Technology in Society 71 (2022) 102092

Contents lists available at ScienceDirect

Technology in Society

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

Changes in service quality of sharing accommodation: Evidence from airbnb

Wenming Zuo [a,c], Weijing Bai [a,*], Wenfeng Zhu [a], Xinming He [b], Xinxin Qiu [a]

[a] Department of Electronic Business, South China University of Technology, No. 382, Waihuan East Road, Panyu District, Guangzhou, Guangdong, China
[b] Durham University Business School, Durham University, Mill Hill Lane, Durham, DH1 3LB, UK
[c] Pazhou Laboratory, No. 248, Pazhou Qiaotou Street, Haizhu District, Guangzhou, Guangdong, China

ARTICLE INFO

Keywords:
Online reviews
Weakly supervised topic model
Sharing accommodation
Service quality
Sharing economy

ABSTRACT

As consumers' perceptions of quality change over time, service providers should track the dynamic changes in service quality and adjust their services to adapt to these changes. In the context of sharing accommodation, although many studies have focused on evaluating service quality, the dynamic changes in service quality over time remain unexplored. This study aims to explore the dynamic changes in the service quality of sharing accommodation. We propose a novel framework that uses deep learning and SERVQUAL to analyze online reviews of sharing accommodation services, thus contributing to the literature from a dynamic perspective. Using a 10-year-span longitudinal dataset of Airbnb's online reviews of San Francisco's listings ($n_1 = 366{,}643$), we construct a weakly supervised topic model that extracts service quality topics from online reviews and then classifies reviews into irrelevant-topics and relevant-topics. Each relevant-topic review is mapped to one of the SERVQUAL dimensions, combined with its sentiment analysis score, which constitutes the output of text mining. We then analyze the dynamic changes in the service quality. The results show that both the overall service quality and that in each dimension of SERVQUAL exhibit a slight downward trend. We obtain the similar results for the Beijing longitudinal dataset ($n_2 = 251{,}081$), which confirms that the downward trend in service quality is not unique to San Francisco. We discuss the reasons for this trend and provide managerial guidance for the platform and its hosts.

1. Introduction

As a disruptive innovation to traditional accommodation [2], sharing accommodation platforms, exemplified by Airbnb, have been booming globally for more than a decade [3,4]. Thus, interesting questions have emerged. What changes have occurred in customers' perceived service quality of sharing accommodation? Understanding the changes in the service quality of sharing accommodation in the past can provide important guidance to current players, policy makers, and business leaders regarding the future development of the industry.

The service quality evaluation of sharing accommodation is not only an important practical challenge, but also a critical research topic [5,6]. Previous studies have recommended measurable service quality scales [7], and through empirical research identified important factors that affect perceived service quality in sharing accommodation, including convenience, assurance, flexibility, efficiency, reliability, privacy, and security [8,9].

These studies have enriched our understanding of service quality in the context of sharing economy and accommodation. Although several studies on sharing accommodation focus on evaluating service quality, the changes in service quality over time remain unexplored. Zeithaml [10]. indicated that consumers' perceptions of quality change over time, and service providers should track these dynamics and adjust their strategies accordingly [10]. In this study, we define the changes in service quality over time as the dynamic nature of service quality. Related studies have adopted questionnaire surveys (e.g., Refs. [11–14]) and expert opinions (e.g. Refs. [15–17]), to collect data, which generate small quantities of data and are static in nature. Other studies using large datasets with time attributes, such as online reviews, also failed to consider dynamic changes in service quality (e.g., Refs. [18,19]).

To fill this void, we propose a model and method that takes advantage of big data technology's development and the accumulation of

* Corresponding author.
E-mail addresses: wmzuo@scut.edu.cn (W. Zuo), bai.weijing@qq.com (W. Bai), wenfengst@qq.com (W. Zhu), xinming.he@durham.ac.uk (X. He), qiuxin_n@qq.com (X. Qiu).

https://doi.org/10.1016/j.techsoc.2022.102092
Received 23 April 2022; Received in revised form 11 July 2022; Accepted 14 August 2022
Available online 22 August 2022
0160-791X/

online reviews, which have been valued as the genuine and available "voice of the customer" [20] and mined to extract useful service quality information. This is a helpful method for gathering useful data and has been applied in areas such as e-commerce [21], food-delivery services [22], and the airline industry [23,24]. Online reviews have the potential to generate a large amount of data and are a natural continuity in time, making it easy to track dynamic changes in the focal phenomenon. However, customers' online comments often appear as vague natural language representations in the form of unstructured data [25]. Machine learning appears to be a helpful tool for transforming these unstructured qualitative natural language data into structured quantitative data [26]. For instance, Ref. [27] proposed a seed-guided topic model for dataless text filtering and classification (DFC) as a weakly supervised classifier, which achieved good text classification effects on two datasets: Reuters-10 and 20-Newsgroup. Given a collection of unlabeled documents, and for each specified category, a small set of seed words that are relevant to the semantic meaning of the category, DFC filters out irrelevant documents and classifies the relevant documents into corresponding categories by topic influence [27].

Following this analysis, we collect 366,643 reviews of San Francisco's listings on Airbnb spanning 10 years. Subsequently, we employ the DFC weakly supervised topic model and deeply integrate SERVQUAL to extract the service quality information. Moreover, to quantify the level of service quality, we conduct sentiment analysis on reviews and measure the service quality using sentiment scores. This enables us to visualize long-term dynamic changes in service quality. The results show that both the overall service quality and that in each dimension exhibit a slight downward trend. We use the Auto Regressive Integrated Moving Average (ARIMA) model to forecast overall service quality, further validating the downward trend in overall service quality. To test whether it is unique to San Francisco, we further use a 10-year-span longitudinal dataset of Beijing (n2 = 251,081), a city that is geographically, linguistically, and culturally very different from San Francisco, and obtain the similar results. This study contributes to the literature on service quality in the sharing accommodation sector.

2. Literature review

2.1. Service quality evaluation

Research on service quality evaluation has a wide range of applications and the evaluation methods applied are diverse. We systematically summarize relevant studies on service quality evaluation over the last six years (from 2016 to 2021), including service quality evaluation methods, data sources, evaluation objects, and whether dynamic factors are considered, as shown in Table 1.

As shown in Table 1, most studies have adopted questionnaire surveys (*e.g.* Refs. [11–14,37–42]) and expert opinions (e.g. Refs. [15–17]) to collect data. Furthermore, some studies have adopted big data but did not consider dynamic changes (e.g., Refs. [18,21]). Other studies considered dynamic changes but did not adopt longitudinal data (e.g. Ref. [31]), or failed to deeply integrate service quality evaluation models (e.g., Ref. [24]).

2.2. Sharing accommodation service quality

A recent literature review [7] of 2211 related articles on service quality in hospitality and tourism from 1984 to 2014 identified 17 research themes, including important service quality scales [8], [9], [92] and the consequences of service quality [43,44]. For sharing

Table 1
Literature on service quality evaluation in the last six years (from 2016 to 2021).

Literature	Method	Data source	Object	Whether to consider dynamics
Berezina et al. [18]	PASW Modeler, CATPAC	Online reviews	Hotel service quality	No
Bezerra and Gomes [11]	Confirmatory factor analysis	Survey	Airport service quality	No
de Oña et al. [28]	Classification, cluster analysis	Customer satisfaction surveys	Transit service quality	No
Guirao et al. [12]	Factorial analysis, multiple regression analysis and Multiple Indicators Multiple Causes (MIMIC) models	Customer satisfaction surveys	Public transportation service quality	No
Rahman et al. [29]	Empirical models	Survey	Paratransit service quality	No
Blut [30]	Meta-analysis	Survey	E-service quality	No
Aydin [31]	Statistical analysis, fuzzy trapezoidal numbers and TOPSIS	Customer satisfaction surveys	Rail transit systems' service quality	Yes
Li et al. [13]	Fuzzy AHP and 2-tuple fuzzy linguistic method	Survey	In-flight service quality	No
Gupta [15]	Best-worst method, VIKOR methodology	Expert opinion	Airline service quality	No
Palese and Usai [21]	Text mining	Online reviews	E-commerce service quality	No
Miranda et al. [32]	Fuzzy set QCA approach	Survey	Railway service quality	No
Pham and Yeo [16]	Consistent Fuzzy Preference Relation (CFPR) method	Expert opinions	Shipping companies' service quality	No
Theodosiou et al. [33]	Structural Model Estimation	Survey	E-service quality	No
Ocampo et al. [34]	AHP-TOPSIS method	Survey	Public service quality	No
Tuzkaya et al. [35]	IVIF-TOPSIS method	Survey	Hospital service quality	No
Lupo and Bellomo [17]	Multi-Criteria-Decision-Analysis (MCDA)-based approach	Expert opinions	Restaurant service quality	No
Mirzaei et al. [36]	Factor analysis	Survey	Community pharmacies' service quality	No
Korfiatis et al. [24]	Structural Topic Models (STM)	Online reviews	Airline service quality	Yes
Atalay et al. [14]	FIPIA with information entropy method	Survey	Airline service quality	No
Brochado et al. [23]	Mixed content analysis	Online reviews	Airline service quality	No
Fei et al. [37]	Evidential best-worst method	Survey	Hospital service quality	No
Ming et al. [38]	Best-worst method	Survey	Medical services' quality	No
Barrios-Ipenza et al. [39]	Kano model	Survey	Health services' quality	No
Wang et al. [40]	TOPSIS method	Survey	Public transport service quality	No
Carvalho and Medeiros [41]	Cluster Analysis and Structural Equation Modeling	Survey	Airline service quality	No
Li et al. [42]	Structural Equation Modeling	questionnaire surveys	Bank services	No

accommodation, previous studies have recommended evaluation metrics for service quality and through empirical research found important factors that affect perceived service quality [8]. For instance, Sutherland and Kiatkawsin [25] analyzed topics of interest that promote customer experience and satisfaction in sharing accommodation by extracting topics from 1,086,800 Airbnb reviews of New York City.

Furthermore, studies have been conducted on the relationships between service quality, customer satisfaction and customer loyalty in sharing accommodation. For instance, Ju et al. [43] studied Airbnb's service quality attributes and their impact on customer satisfaction by analyzing 16,430 Airbnb online reviews and 322 online surveys. Another recent study [45] investigated the association between peer-to-peer interactions and several outcome variables (encounter satisfaction, word-of-mouth intention, and continuous intention to use) by analyzing 503 responses from an online research panel.

Customers' perception of service quality is changing [10], leading to variations in customer loyalty and satisfaction [46]. Therefore, tracking changes in service quality is critical to maintain customer relationships and studying dynamic changes in service quality should be an important topic. However, to the best of our knowledge, no study has yet investigated the dynamic changes of service quality in the context of sharing accommodation.

To better understand the service quality, we employ SERVQUAL [47] as the instrument, which has been widely applied in the service sector [48–50]. The traditional accommodation industry evaluates service quality using SERVQUAL or SERVQUAL-based service quality instruments specifically for the hospitality sector, such as LODGSERV [51], LODGQUAL [52], and the "Lodging Quality Index" (LQI) [53].

SERVQUAL measures service quality based on the gap between customers' perceived and expected service levels, including five dimensions. Each dimension is defined as follows [47]:

- *Assurance: Knowledge and courtesy of employees and their ability to inspire trust and confidence.*
- *Empathy: Caring, individualized attention the firm provides its customers.*
- *Reliability: Ability to perform the promised service dependably and accurately.*
- *Responsiveness: Willingness to help customers and provide prompt service.*
- *Tangibles: Physical facilities, equipment, and appearance of personnel.*

With the development of Internet information technology, network services have become an important form of modern service, and their service quality evaluation has emerged as an important research topic. Udo et al. [54] assessed the quality of e-learning by using a modified SERVQUAL instrument. Parasuraman et al. [55] constructed a multi-item scale (E-S-QUAL) based on SERVQUAL to measure the service quality of online shops.

Guttentag [2] suggests that compared to traditional accommodation, sharing accommodation is a disruptive innovation and that has its own disadvantages (such as lack of security and trust due to the inability to engage in face-to-face transactions) and advantages (such as convenience, cost saving, and the potential for more authentic local experiences). This is verified by Ref. [56] who extracted topics from 1,026,988 Airbnb online reviews in seven U.S. cities and found that compared with the topics extracted from traditional hotel reviews [57], those from Airbnb have some unique topics, namely, "late check-in", "patio and deck view", "food in kitchen", "help from host", "door lock/key", "sleep/bed condition", and "host response". These unique topics can also be classified into the five dimensions of SERVQUAL. For example, "late check-in" can be classified into the reliability dimension; "patio and deck view", "door lock/key", and "sleep/bed condition" can be classified into the tangibles dimension; "help from host" can be classified into the empathy dimension; and "host response" can be classified into the responsiveness dimension.

Over time, the security of online booking services has increased and the issue of Internet trust has gradually been resolved. Additionally, as the performance level of innovative products increases, most consumers become satisfied with the performance, thus reducing market heterogeneity [1]. This makes the characteristics that previously distinguished traditional and sharing accommodation decreasingly relevant.

2.3. Online reviews

The development of big data technology and the accumulation of online review data has made comprehensive analysis of online reviews possible, which are valuable source to reflect the genuine "voice of the customer" [20] and can help researchers more precisely understand consumer preferences and demands [57].

Online customer reviews are particularly useful for service firms. In several cases, customers can post feedback on received services online. These reviews can further affect other customers' purchase intentions [59–65], thus impacting company performance [66–70]. Therefore, businesses should take customer reviews seriously, extract and analyze the key elements of online feedback on service quality [25,26,57,71–75], evaluate and improve their service quality [23,24,76] and customer satisfaction [43].

Over time, online reviews accumulate information from customers' feedback, which reflects the variations in customers' service evaluations and reveals the changes in service quality. Although many studies exist on mining service quality information from online reviews, we surprisingly find that no study has focused on the dynamic changes in service quality.

2.4. Research framework

The discussion above clearly identifies the gap in the dynamic nature of existing research evaluating the service quality of Airbnb's sharing accommodation. To address these issues, in combination with the SERVQUAL model, we propose a service quality evaluation method based on a weakly supervised topic model to extract information from online reviews and evaluate the service quality of sharing accommodation from a dynamic perspective. It integrates both qualitative and quantitative methods to analyze the quality of online accommodation services. The research framework is presented in Fig. 1.

In Section 3, we consider San Francisco as an example to present the procedures and methods of service quality mining for sharing accommodation. In Section 4, by combining the mining results of Section 3 with the SERVQUAL model for an in-depth analysis, we obtain the dynamic changes in the service quality of San Francisco. To test the generality of the results, in Section 5, we repeat the text mining process as in Section 3 with the Beijing dataset and obtain the similar results.

3. Online review mining

3.1. Case selection and data collection

Airbnb is a global sharing accommodation platform. We use San Francisco as an example and choose Beijing to test the results, as two representatives from the eastern and western hemispheres. They are both popular and vibrant destinations with many reviews on the Airbnb platform. We generate datasets of Airbnb users' online reviews of San Francisco and Beijing venues through Inside Airbnb (insideAirbnb.com) to evaluate the service quality of the Airbnb platform. The San Francisco dataset includes 366,643 reviews from July 2009 to October 2019, mainly in English. The Beijing dataset includes 251,081 reviews from August 2010 to October 2019, primarily in Chinese. Each item of the review data covers listing_id, review_id, date, reviewer_no, and comments.

The online review text mining process used in this study is shown in Fig. 2. First, we preprocess the dataset to obtain the document-word matrix. Second, taking this matrix as the input of the LDA (Latent

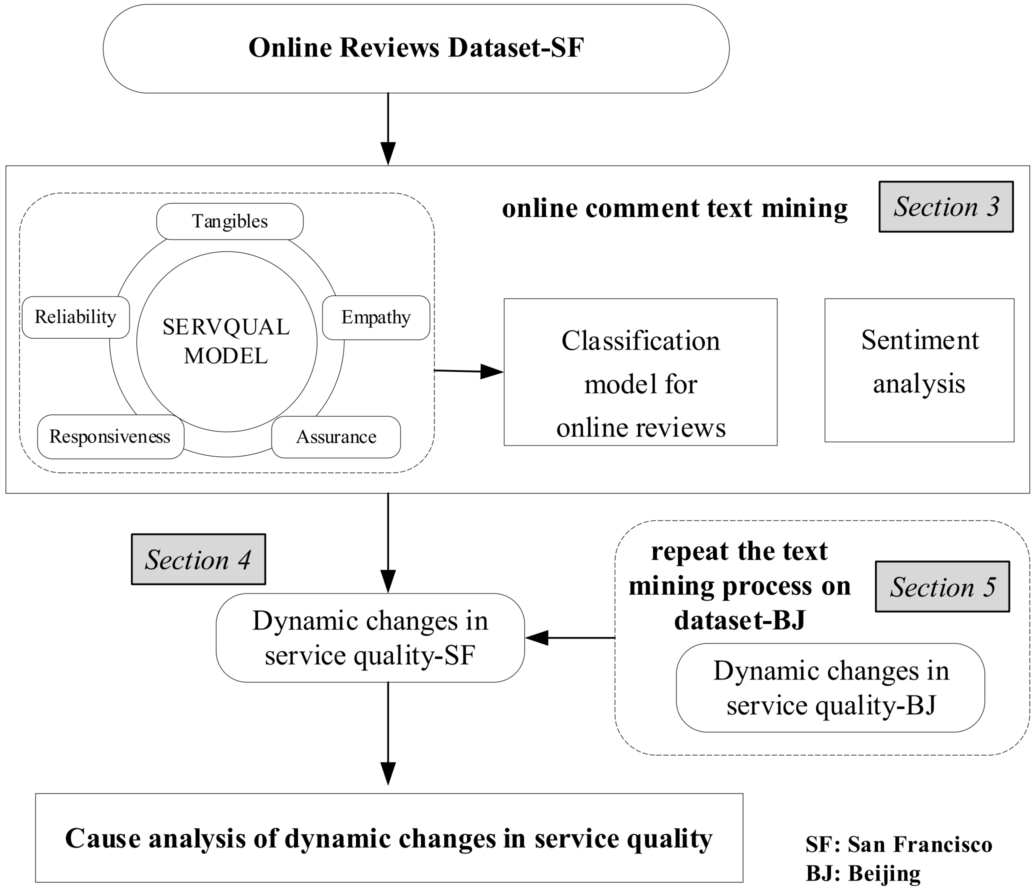

Fig. 1. Research framework.

Dirichlet Allocation) topic model, we determine the optimal number of topics by calculating the perplexity of the number of topics. In addition, we extract hidden topics from the document-word matrix with the optimal number of topics to obtain the LDA output result. Third, based on SERVQUAL dimensions and definitions of measurement items, we preset the initial seed words of the weakly supervised topic model and extend them to obtain the seed word set using word embedding. Fourth, we calculate the distance between each LDA topic and all the seed words and take the top-10 topical words under each of the least relevant LDA hidden topics as pseudo seed words for the irrelevant-topics in the DFC. Finally, taking the document-word matrix, seed word set, and pseudo seed word set as the input of the weakly supervised topic model (DFC), we extract the service quality information in the text and perform sentiment tendency analysis on the text, which serves as the basis of service quality evaluation. The output of the DFC model includes document_id (review_id), topic_id (the SERVQUAL dimension that the review maps), and the sentiment score of this review.

3.2. Topic extraction by LDA

Latent Dirichlet Allocation (LDA) is a topic model proposed by Ref. [77] that is used to infer the topic distribution of documents and extract hidden topics in the documents through unsupervised learning methods.

The key to the extraction of LDA topics is determining the optimal number of topics. We use *perplexity* to evaluate the effect of topic extraction and calculate the perplexity value of the model under different number of topics. Experiments on the Beijing and San Francisco datasets indicate that as the number of topics increases, the perplexity values of these two models generally show a downward trend, which means that the more topics there are, the better the LDA's topic-modeling ability. However, a large number of topics significantly increases the training cost. The experimental data showed that after the number of topics exceeded 200, the downward trend of the models' perplexity values tended to be flat and became no longer obvious. Considering the perplexity value and training cost, the number of LDA topics in both the datasets is set to 200.

The two output results of LDA are used as the inputs for the weakly supervised topic model:

(1) Dictionary of Online Reviews: *wordmap.txt.* It is used as the input to set seed words (see Section *3.3.1, Presetting and extending seed words* for details).
(2) LDA topic-extraction results. It is used as the input of the weakly supervised topic model to calculate the pseudo-seed words information of irrelevant-topics (see Section *3.3.1, Setting pseudo seed words* for details).

3.3. Weakly supervised topic extraction (DFC)

Li et al. [27] proposed a seed-guided topic model for dataless text filtering and classification (DFC), which achieved good text classification effects on two datasets: Reuters-10 and 20-Newsgroup. The authors evaluated the filtering and classification performance of DFC against other state-of-the-art dataless text classifiers and supervised learning methods, including GE-FL, DescLDA, SNB-EM, SVM, SSVM, sLDA, and MedLDA. DFC outperforms all of these competitors and is very robust to parameter settings [27].

The DFC model does not need to label text in advance. Given a set of unlabeled documents and a small group of semantically related seed

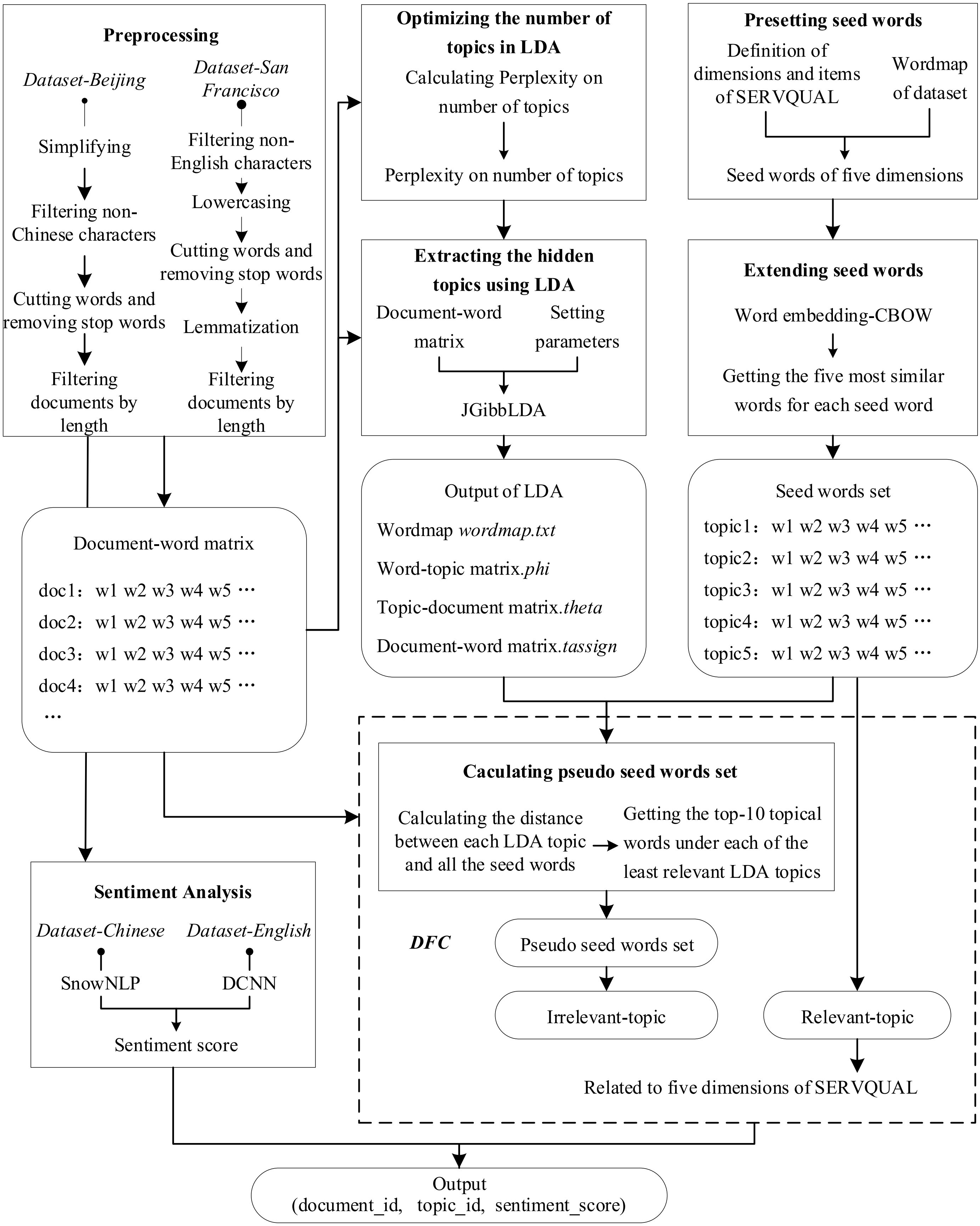

Fig. 2. Text mining of online reviews.

words set for each specified category, DFC can infer which category of concern the document belongs to (i.e., relevant-topics) or does not belong to any category of concern (i.e., irrelevant-topics). According to DFC, a category has a one-to-one correspondence with its topic.

In our study, the five categories corresponding to SERVQUAL's five dimensions are of interest, and their topics are collectively referred to relevant-topics. The remaining topics that are irrelevant to SERVQUAL belong to irrelevant-topics.

3.3.1. Setting seed words and pseudo seed words

3.3.1.1. Presetting seed words. The seed words are preset in five dimensions for the Chinese and English datasets, and the number of preset seed words in each dimension is limited to approximately 10, as shown in Table 2. The seed words are selected from the high-frequency keywords in the dictionary generated by the review dataset. The selection criteria are as follows: (1) keywords appearing in the SERVQUAL dimension definition and measurement items [47]; (2) words that conform to the meaning of the SERVQUAL dimension and measurement item definitions. In Table 2, the underlined words are the seed words selected according to criterion (1) and the unlined words are the seed words selected according to criterion (2).

3.3.1.2. Extending seed words. Artificially set seed words are unable to cover all semantic details of the SERVQUAL dimensions. Therefore, we enrich the seed words by using Word2vec to train the word vector to obtain the most similar words in the corpus to extend the seed words in use. Aided by the open-source toolkit gensim, a CBOW model is constructed for the preprocessed comment corpus, and the word vector of each word in the comment is trained. We calculate the similarity between each word and the seed word, and take the five words with the highest similarity to each preset seed word in each service dimension as the augmented seed word. After merging and deduplication, we generate a full list of augmented seed words for each service dimension. Table 3 presents an example of the empathy dimension.

3.3.1.3. Setting pseudo seed words. Our study uses the topic extraction result of LDA as the input to the DFC model to generate pseudo-seed words of irrelevant-topics. The detailed process is as follows:

Step 1: We use LDA to extract the hidden topics from the reviews.
Step 2: We calculate the distance between each LDA hidden topic and all seed words provided for the relevant categories.

Table 2
Initially preset seed words of SERVQUAL's five dimensions- SF.

Dimension	Definition	Dataset- San Francisco
Assurance	Knowledge and courtesy of employees and their ability to inspire trust and confidence.	<u>knowledge safe adequate</u> nice recommend welcome lovely friendly kind recommendation
Empathy	Caring, individualized attention the firm provides its customers.	<u>care personal need interest heart convenient</u> comfortable love enjoy cozy comfy decorate
Reliability	Ability to perform the promised service dependably and accurately.	<u>accurately problem</u> thank helpful help describe picture appreciate issue
Responsiveness	Willingness to help customers and provide prompt service.	<u>prompt busy respond request</u> quick responsive communication question communicate
Tangibles	Physical facilities, equipment, and appearance of personnel.	<u>facility equipment neat</u> clean beautiful bathroom kitchen spacious amenity bedroom garden door equip

Table 3
Augmented seed words of SERVQUAL's empathy dimension-SF.

Dimension	Dataset- San Francisco
Empathy	culture restore awesome soft require close lovely decorate convenient maintain comfortable cultural need border decorative plush well-kept enjoyable thoughtful history cozy love stylish heavenly adorable pride special personalize initiative personal necessary ideal care design doorstep appoint consideration want necessity fascinate comfy near furnish diverse aback middle essential edge enjoy personalize heart luxurious interest easy walk-in cosy central effort

Step 3: We take the top-10 topical words under each of the least relevant LDA hidden topics as pseudo seed words for the irrelevant-topics in DFC.

Table 4 presents an example of pseudo-seed words for irrelevant-topics.

3.3.2. Topic extraction and text classification

After determining the seed words of each SERVQUAL dimension and the pseudo seed words of irrelevant-topics, we use the open-source tool DFC to model the review texts of the San Francisco dataset and calculate the probability that a review belongs to a certain category of topic. According to the probability results, each review is mapped to a category's topic. If a review is mapped to a topic corresponding to SERVQUAL's dimension, it is highly relevant to the dimension. Otherwise, a review is mapped to an irrelevant-topic, which means that it is irrelevant to the five dimensions of SERVQUAL.

Table 5 shows examples of reviews classified into the corresponding SERVQUAL dimensions. The second column shows the output results of preprocessing the sentences of online reviews. The preprocessing steps include filtering non-English characters, converting uppercase letters to lowercase, cutting words, removing stop words, lemmatization, and filtering documents by length.

The first review concerns the host's warm and friendly evaluation, which matches the assurance dimension. The second review is regarding the host's thoughtful and careful evaluation, which matches the empathy dimension. The third review is about the host's evaluation of helping to solve difficulties, which is matched to the reliability dimension. The fourth review concerns the host's quick response to the email, and is matched to the responsiveness dimension. The fifth review focuses on the evaluation of the house's bathroom, bedroom, and other facilities, which are matched to the tangibles dimension.

3.4. Sentiment analysis

Following recent research (e.g. Refs. [43,78], we conduct sentiment analysis on review data matched to the five dimensions of SERVQUAL with DCNN[1] for the English dataset.

The output result of the model is the probability that the sentiment polarity of the text is positive, with a value between 0 and 1. When the output result is less than or equal to 0.5, the sentiment tendency is negative. When the output result is greater than 0.5, the sentiment

Table 4
Pseudo-seed words of irrelevant-topic -SF.

Dimension	Dataset- San Francisco
irrelevant-topic-1	couple away half line polk fillmore deli number strip major pick street intersection block blow park foot tuck direction side bakery ness mile right market cafe stop cora step grab

[1] https://github.com/xiaohan2012/twitter-sent-dnn.

Table 5
Examples of reviews matching each SERVQUAL dimension of the San Francisco dataset.

Online review	After preprocessing	SERVQUAL dimension
I just relocated to SF from Florida and Aaron helped me out so much! He is an awesome host; very friendly, knowledgeable, punctual, and flexible. The room I stayed in was fantastic with a great view and the neighborhood it is located in is very safe and quiet. I would definitely recommend staying at A Friendly Hotel to anyone and everyone.	relocate florida aaron help much awesome host friendly knowledgeable punctual flexible room stay fantastic great view neighborhood locate safe quiet definitely recommend stay friendly hotel	Assurance
Monika and Kevin have been very awesome hosts. My 2 friends (French and Czech) and I spent 2 nights in their house on a hill, very well located next to Castro district. We had a very warm welcome and everything one needs to have an enjoyable stay. We really recommend them!	awesome host friend french czech spend night house hill well locate next castro district warm welcome need enjoyable stay really recommend	Empathy
Gae's apartment is exactly as described - its an artful place that an art lover would appreciate. Gae was also a very helpful host, and I enjoyed talking with her. Her knowledge of napa valley was also very impressive. Overall, a very positive experience!	apartment exactly describe artful place lover appreciate also helpful host enjoy talk knowledge napa valley also impressive overall positive experience	Reliability
The place was beautiful and clean. We really enjoyed staying here. I never met the owner however we communicated with her via email and she was quick to respond. The only issue we had was the parking however she allowed us to use her spot after we let her know we were having an issue. We appreciate that. Thanks	place beautiful clean really enjoy stay never meet owner however communicate email quick respond issue park however allow spot know issue appreciate thank	Responsiveness
The room was quiet, comfortable, clean and attractive. The bathroom was superb, but I would have preferred an entrance from the bedroom. Overall, it was a very good value. And the location is great. I can recommend it.	room quiet comfortable clean attractive bathroom superb prefer entrance bedroom overall good value location great recommend	Tangibles

Note: The underlined words are the seed words of the corresponding dimension.

tendency is positive. To evaluate service quality more intuitively, we assign sentiment scores to each review based on the output results, and then convert the probability value of 0–1 into sentiment scores of 1–5, as shown in Table 6.

Table 6
Mapping from probability values to sentiment scores.

probability value (pv)	sentiment score
$0< pv\leq 0.2$	1
$0.2 < pv \leq 0.4$	2
$0.4 < pv\leq 0.6$	3
$0.6 < pv\leq 0.8$	4
$0.8 < pv\leq 1$	5

The frequency of reviews corresponding to each sentiment score of San Francisco in the five dimensions is shown in Table 10 (see Section 6.1 Chi-square test for details).

4. Dynamic changes in service quality of San Francisco

We calculate the mean values of the five dimensions and overall sentiment scores on a quarterly basis to observe the dynamic trends of overall service quality and service quality in each dimension over time. As the review data published before 2012 in the San Francisco dataset have a small data volume and obvious data fluctuations, we employ the review data published from the first quarter of 2012 to the third quarter of 2019. The dynamic trend of the mean sentiment score over time is shown in Fig. 3.

Both the overall service quality and that in each dimension experienced fluctuations in the initial stage and then gradually stabilized. In terms of the dimensions of assurance and empathy, the quality of service is relatively higher, more stable, and even significantly higher than the other three dimensions. Service quality in these two dimensions shows a more stable trend of dynamic changes over time, with the overall decline being smaller than that in the other three dimensions.

To forecast the service quality level in the coming quarters and thus verify the downward trend of overall service quality, we construct an ARIMA model to fit and predict the quarterly overall sentiment score sequence of San Francisco to find the trend of dynamic changes in service quality. ARIMA model is a popular and widely used statistical method for time series forecasting. In an "ARIMA (p, d, q)" model, where: p (AR) is the number of autoregressive terms, d (I) is the number of nonseasonal differences needed for stationarity, and q (MA) is the number of lagged forecast errors in the prediction equation. The values of p and q can be obtained using a (partial) autocorrelation graph, and d can be obtained using the Augmented Dickey-Fuller (ADF) test. We fit the best ARIMA model for the overall service quality time series in San Francisco and predict the next six periods.

ARIMA (1, 1, 0) was fitted as the best ARIMA model to determine the overall sentiment score of San Francisco, and the model formula is shown in Equation (1). The model fitting parameters are listed in Table 7. The AIC value is the minimum value for various possible models; therefore, the fitted model is the best model. The value of Q6 is 0.006, and the corresponding p-value is greater than 0.05. This shows that the null hypothesis cannot be rejected at the significance level of 0.05; that is, the first 6-order autocorrelation coefficients of the residuals meet the white noise. The final fitted model is the AR model, and the $\alpha 1$ coefficient is −0.386.

Fig. 4 shows a trend graph of the true, fitted, and predicted values, which also shows a downward trend.

$$y(t) = 0.013 - 0.386*y(t-1) \tag{1}$$

5. Investigating the dynamic change: evaluation of service quality in beijing

We chose Beijing to test the dynamic change results in San Francisco, because the two cities are from different hemispheres and have different cultures and languages. They are both popular and vibrant destinations with numerous reviews on the Airbnb platform. The Beijing dataset includes 251,081 Airbnb reviews from August 2010 to October 2019, mainly in Chinese. We mine the Beijing dataset using the similar data-mining process as in San Francisco.

5.1. Online review mining of the beijing dataset

Step1. Setting seed words and pseudo seed words of the Beijing dataset

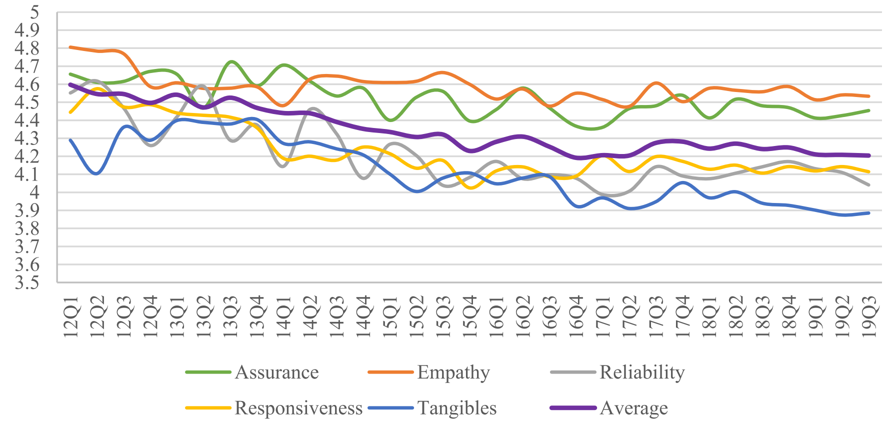

Fig. 3. Dynamic trend of service quality in San Francisco.

Table 7
Time series regression results of overall service quality in San Francisco.

Parameters of ARIMA (1, 1, 0)		
Item	Symbol	Value
Constant	c	−0.013
MA	α1	−0.386
Q	Q6(p-value)	0.006(0.939)
	Q12(p-value)	1.132(0.980)
	Q18(p-value)	3.633(0.989)
	Q24(p-value)	6.927(0.991)
	Q30(p-value)	10.440(0.992)
Information criterion	AIC	−107.23
	BIC	−103.026

*p < 0.05 **p < 0.01

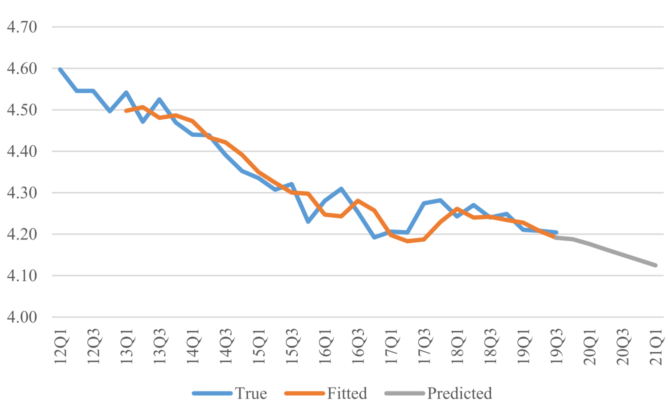

Fig. 4. Fitting and predicting of overall service quality in San Francisco.

The selection process and criteria of the seed words and the pseudo-seed words in the Beijing dataset are the same as those in the San Francisco dataset (see Section *3.3.1 Setting seed words and pseudo seed words* for details).

Notably, as all seed words are selected from the dictionary generated by the review dataset, the preset seed words corresponding to different datasets are not the same. As shown in Table 8, the preset seed words for the Beijing dataset are different from those of San Francisco (Table 2). Moreover, to extract as much service quality information as possible, we use a relatively loose caliber when setting artificial seed words. The full list of augmented seed words of the empathy dimension after the extension of seed words is shown in Table 9.

Step2. Topic extraction and text classification of the Beijing dataset

Table 8
Initially preset seed words of SERVQUAL's five dimensions-BJ.

Dimension	Definition	Dataset- Beijing
Assurance	Knowledge and courtesy of employees and their ability to inspire trust and confidence.	保证(assurance) 放心(reassurance) 礼貌(polite) 热情(enthusiastic) 热心(warmhearted) 推荐(recommendation) 周到(thoughtful) 热情好客(hospitable) 担心(concerned) 亲切(kind) 友好(friendly)
Empathy	Caring, individualized attention the firm provides its customers.	关心(care) 需求(need) 舒服(comfortable) 温馨(cozy) 性价比(economical) 耐心(patient) 用心(care) 细心(careful) 照顾(care for) 特色(characteristic)
Reliability	Ability to perform the promised service dependably and accurately.	准确(accurate) 困难(problem) 感谢(thank) 解决(solve) 帮(help) 谢谢(thank) 帮忙(helpful) 描述(describe) 非常感谢(grateful) 解答(answer)
Responsiveness	Willingness to help customers and provide prompt service.	响应(respond) 回复(reply) 沟通(communicate) 很快(quick) 问(question) 信息(message) 订(book) 预定(reserve) 回应(respond) 有问必答(responsive)
Tangibles	Physical facilities, equipment, and appearance of personnel.	设施(facility) 设备(equipment) 现代化(modern) 整洁(neat) 卫生间(bathroom) 床(bed) 厨房(kitchen) 客厅(sitting room) 卧室(bedroom) 冰箱(refrigerator) 沙发(sofa) 电视(television) 床单(sheet)

Table 9
Augmented seed words of SERVQUAL's empathy dimension.

Dimension	Dataset- Beijing
Empathy	软和(tender) 雅致(elegant) 用心(care) 款待(entertain) 疑惑(wonder) 品味(tasteful) 热情接待(passionate) 心思(thoughtful) 花心思(thoughtful) 无微不至(meticulous) 需求(need) 提出(propose) 舒适(comfy) 温暖(warm) 耐心(patient) 事无巨细(meticulous) 问候(greeting) 精致(delicate) 有心(thoughtful) 特色(characteristic) 别致(fancy) 性价比(economical) 比价(competitive price) 关心(care) 想得(want) 安全系数(safety) 清爽(refreshing) 疑问(wonder) 尽力(best effort) 挺舒服(comfortable) 细心(careful) 舒服(comfortable) 精心(meticulous) 温馨(cozy) 情调(emotional) 格调(stylish) 效率(efficiency) 给予(offer) 不厌其烦(patient) 协助(assist) 照顾(care for) 招待(serve) 利用率(availability) 在线(online) 软软(cozy) 很暖心(warm) 关照(care) 软(soft) 贴心(thoughtful) 体贴(considerate)

Chinese reviews are classified into the corresponding SERVQUAL dimensions according to the probability that a review belongs to a certain category of topic. For example, the following review focuses on the evaluation of the house's room, bed, and kitchen facilities, being matched to the tangibles dimension.

房子是酒店式公寓，房间很大，床也很大很舒服，不过稍微有点软睡着腰不是很舒服。厨房也非常好，厨具和餐具齐全。另外，Helen人超好！借给我们葱蒜油盐酱醋糖，有什么需要都很快帮我们实现。非常好的住宿体验！

(The house is a hotel-style apartment. The room is very large and the bed is very large and comfortable, but it is a little soft to sleep and the waist is not very comfortable. The kitchen is also very good, with complete kitchenware and tableware. Besides, Helen is very nice! Lend us scallions, garlic, oil, salt, soy sauce, vinegar and sugar, and help us quickly whenever we need. Very good accommodation experience!)

Step3. Sentiment analysis of the Beijing dataset

We conduct sentiment analysis on review data matched to the five dimensions of SERVQUAL with SnowNLP[2] for the Chinese dataset and DCNN for the English dataset (see Section 3.4 Sentiment analysis for details). Both SnowNLP and DCNN classify the sentiment polarity of text using a pre-trained model. Moreover, the training corpuses of these two models are from Twitter, which can minimize the deviation.

To evaluate service quality more intuitively, we also assign sentiment scores to each review based on the output results, and then convert the probability value of 0–1 into sentiment scores of 1–5 (see Section 3.4 Sentiment analysis for details).

The frequency of reviews corresponding to each sentiment score of Beijing in the five dimensions is shown in Table 10 (see Section 6.1 Chi-square test for details).

5.2. Dynamic changes in service quality of beijing

We calculate the mean values of the five dimensions and overall sentiment scores on a quarterly basis to observe the dynamic trends of overall service quality and service quality in each dimension over time. As the review data published before 2015 in the Beijing dataset have a small data volume and obvious data fluctuations, we employ the review data published from the first quarter of 2015 to the third quarter of 2019 in Beijing. The dynamic trend of the mean sentiment score over time is shown in Fig. 5.

Beijing shows the similar dynamic trend in service quality as San Francisco. Both the overall service quality and that in each dimension experienced fluctuations in the initial stage and then gradually stabilized.

We also construct an ARIMA model to fit and predict the quarterly overall sentiment score sequence of Beijing to determine the trend of dynamic changes in service quality. Following the same approach as San Francisco, ARIMA (0, 1, 1) was fitted as the best ARIMA model to determine the overall sentiment score of Beijing, and the model formula is shown in Equation (2). The final fitted model is the MA model, and the $\beta 1$ coefficient is -1.000.

Fig. 6 shows a trend graph of the true, fitted, and predicted values, which also shows a downward trend.

$$y(t) = 0.019 - 1.000 * \varepsilon(t-1) \quad (2)$$

6. Analysis of results

6.1. Changes in servqual dimensions

As shown in Figs. 3 and 5, San Francisco and Beijing have a high degree of consistency in the dynamic trend of service quality. Both the

Table 10
Chi square test results of sentiment score in two cities.

Dimension	City	Sentiment Score	Frequency	Chi-square	p
Assurance	Beijing	1	251	155.361	$p < 0.001$**
	Beijing	2	85		
	Beijing	3	106		
	Beijing	4	175		
	Beijing	5	4662		
	San Francisco	1	2646		
	San Francisco	2	1012		
	San Francisco	3	1054		
	San Francisco	4	1479		
	San Francisco	5	27,075		
Empathy	Beijing	1	338	823.75	$p < 0.001$**
	Beijing	2	150		
	Beijing	3	148		
	Beijing	4	278		
	Beijing	5	14,778		
	San Francisco	1	868		
	San Francisco	2	368		
	San Francisco	3	384		
	San Francisco	4	535		
	San Francisco	5	11,177		
Reliability	Beijing	1	1034	66.488	$p < 0.001$**
	Beijing	2	264		
	Beijing	3	243		
	Beijing	4	402		
	Beijing	5	5785		
	San Francisco	1	2394		
	San Francisco	2	730		
	San Francisco	3	694		
	San Francisco	4	1009		
	San Francisco	5	11,328		
Responsiveness	Beijing	1	3360	1098.957	$p < 0.001$**
	Beijing	2	468		
	Beijing	3	394		
	Beijing	4	567		
	Beijing	5	7299		
	San Francisco	1	4190		
	San Francisco	2	1382		
	San Francisco	3	1350		
	San Francisco	4	1744		
	San Francisco	5	21,057		
Tangibles	Beijing	1	5067	1978.286	$p < 0.001$**
	Beijing	2	593		
	Beijing	3	552		
	Beijing	4	659		
	Beijing	5	6089		
	San Francisco	1	3574		
	San Francisco	2	954		

(continued on next page)

[2] https://github.com/isnowfy/snownlp.

Table 10 (*continued*)

Dimension	City	Sentiment Score	Frequency	Chi-square	p
	San Francisco	3	885		
	San Francisco	4	1251		
	San Francisco	5	13,682		

*$p < 0.05$ **$p < 0.01$.

overall service quality and that in each dimension experienced fluctuations in the initial stage and then gradually stabilized. To analyze whether there were differences in the sentiment scores of the two cities in each dimension, we performed a chi-square test.

As shown in Table 10, at the 99% confidence level, the sentiment scores of Beijing and San Francisco show significant differences across all dimensions. In evaluating service quality, there are significant differences in the sentiments of customers in Beijing and San Francisco, and there are significant differences in all dimensions of SERVQUAL.

Fig. 7 illustrates the matched reviews frequency of sharing accommodation services in Beijing and San Francisco under the five dimensions of SERVQUAL. Beijing accounted for the highest in the empathy dimension (29.20%), while San Francisco accounted for the highest in the assurance dimension (29.49%).

As shown in Figs. 3, 5 and 7, the changes in service quality of San Francisco and Beijing demonstrate the following characteristics.

First, either in the overall service quality or in each dimension, fluctuations can be observed in the initial stages and then became gradually stabilized, indicating that the service capacity of sharing accommodation was not perfect in the early periods of development. However, with the accumulation of service experiences by hosts and sharing platforms, the service facilities and capacities have been improved, with the service quality levels getting gradually stabilized. Since the beginning of development, both the overall service quality and specific dimensions as perceived by customers have been showcased a slight downward trend.

Second, the service quality is obviously better in the dimensions of assurance and empathy than in the other three dimensions, and is more stable with a lower overall decline. This shows that hosts have the knowledge, courtesy and ability to inspire trust and confidence in their customers and provide them with caring, individualized attention. Moreover, from the perspective of dynamic changes, such service advantages have maintained for a long time in the two cities.

However, there are some differences in the advantages of the two cities. In Beijing, for instance, the empathy dimension gets the highest number of reviews, indicating that customers are strongly satisfied with the performance in this dimension, with many positive reviews posted. In contrast, in San Francisco, the dimension of assurance gets the highest number of positive reviews. Beijing provide customers Caring, individualized attention. San Francisco has an advantage in professionalism, so customers trust it more.

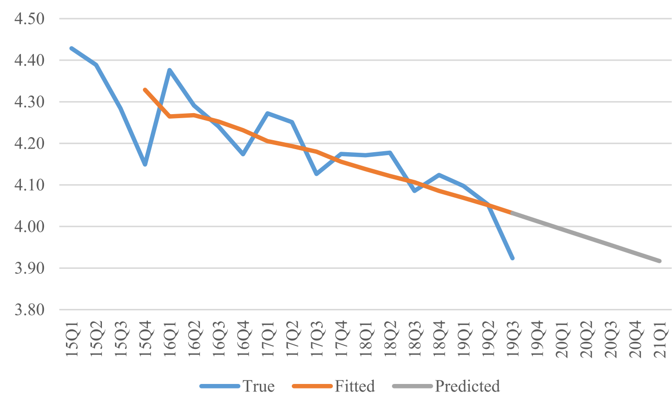

Fig. 6. Fitting and predicting of overall service quality in Beijing.

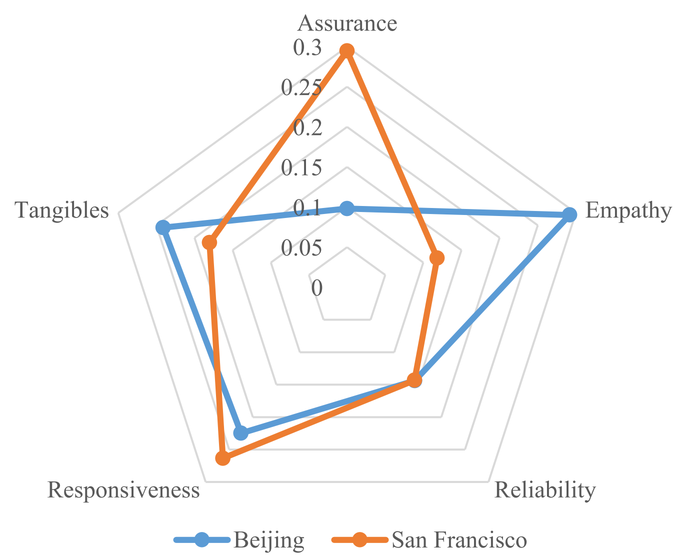

Fig. 7. The frequency of reviews in five dimensions for two cities.

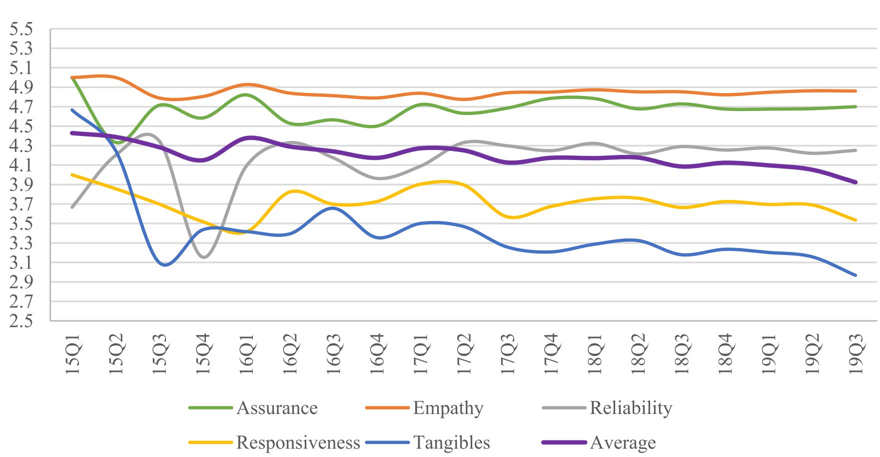

Fig. 5. Dynamic trend of service quality in Beijing.

Third, in both of the cities, the service quality in tangibles dimension is the lowest, showing a serious downward trend, indicating very poor service quality in this dimension, which is getting still worse. With word frequency analyses, we have identified the top five keywords with the highest frequency in tangibles dimension. For San Francisco, the top five keywords include *clean*, *room*, *beautiful*, *bathroom*, and *kitchen*. Those for Beijing are *设施(facility)*, *整洁(neat)*, *卫生间(bathroom)*, *床(bed)*, and *厨房(kitchen)*.

6.2. Causes of the dynamic changes

As shown in Figs. 3 and 5, both the overall service quality and that in each dimension experienced fluctuations in the initial stage and then gradually stabilized, indicating that the service capacity of sharing accommodation is not perfect at the initial stage of development. However, with the accumulation of the service experience of hosts and sharing platforms, the service facilities and capacity improve, and the service quality level gradually stabilizes.

The overall service quality and that in each dimension perceived by customers shows a slight downward trend from the beginning of development. SERVQUAL measures service quality based on the gap between customers' perceived and expected service levels [47]. Customers' perceptions and expectations are influenced by many factors.

At the beginning of the Airbnb establishment, as a disruptive innovation to traditional accommodation [2], according to Christensen's research of disruptive technologies [79], Airbnb's initial service quality was inferior to traditional accommodation. Therefore, customers chose Airbnb not due to the quality of service, but the economic benefits (i.e., less spending and cheaper prices) [8,80–82] and more interaction with locals [8,78,81,83–85], which led to lower expectations of customers from Airbnb services [80,82]. In addition, interpersonal contact with local hosts creates a sense of belonging, providing customers with a "feeling of home" [80,81,83,85], which makes customers perceive their service providers in a more indulgent way. For example, consumers have better evaluations of services [86], do not easily give negative evaluations even if they encounter service failures [80], and still maintain a higher loyalty [86]. These are clues that customers have a good perception of service.

As Airbnb advertises the uniqueness of its listings and some positive eWOM spreads across the web, customers' expectations are rising. However, with the sharp increase in the number of hosts, the number of underserved hosts and various service problems, such as personal injury [87] and sexual assault [88], have also increased, which has led to a decline in customers' perceptions.

With the explosion of Airbnb, disruptive innovation begins to dominate the market and becomes mainstream [89], and market heterogeneity reduces [1]. The increase in multi-listing, institutional, and remotely operated hosts has led to higher housing prices on Airbnb and lower levels of cultural exchange and interpersonal contact [90], which may lead to an increase in service expectations. Customers' expectations rising to flat or even higher than traditional accommodation [91]. At the same time, higher housing prices and lower levels of cultural exchange and interpersonal contact have also contributed to a reduction in customer tolerance for service and a decline in customers' service perceptions.

Moreover, customers expect Airbnb as an intermediary platform, not only to provide system services, but also to manage and regulate the services of hosts and to mediate conflicts and protect the interests of customers when necessary [91], which puts forward higher service expectations for Airbnb. While direct negative interactions between customers and the platform or between customers and hosts may result in lower service perceptions [91].

In conclusion, as Airbnb continues to expand its market share [3,4], the composition of hosts has changed from individual owners with idle houses to diversification [90], and users have changed from niche segments to mainstream customers [89], resulting in users' service expectations rising from lower than traditional accommodation [80,82] to flat or even higher (Huang and Jin, 2020) and a decline in customers' service perceptions. The combination of these factors has resulted in a slight downward trend in the overall quality of customers' perceived services.

7. Discussion

7.1. Theoretical implications

This study contributes to the literature in two ways. First, we present a novel framework for evaluating service quality based on online review text mining, which combines big data, machine learning, and the SERVQUAL model. In the context of sharing accommodation, we implemented the framework in large longitudinal datasets of two international hubs. Our study generates helpful comprehensive results, which prove the effectiveness of the method. Moreover, this framework can be applied to a wider range of scenarios of services.

Second, responding to the lack of dynamics in the prior service quality evaluation literature (e.g. Refs. [18,21,31]), our research reveals the changing trend of Airbnb service quality using large longitudinal datasets of two international hubs: (1) The service quality has experienced fluctuations in the initial stage and then gradually stabilized. (2) The service quality in the dimensions of assurance and empathy is relatively higher and more stable than that in the other three dimensions of SERVQUAL; the tangible dimension has the lowest quality of service and the most severe downward trend. (3) The overall service quality and that in each dimension perceived by customers show a slight downward trend, which is expected to continue.

7.2. Management implications

Consumers' perceptions of quality change over time, and tracking these dynamics is a key issue for service providers [10]. Our study generates helpful results on dynamic changes in the service quality of sharing accommodations. The feedback from customers in both regions reveals a slight downward trend in overall service quality and in all five dimensions. This trend is expected to continue and should draw attention from both the platform and hosts.

In the process of Airbnb's explosive development [3,4], customers and hosts have gradually become mainstream in the market from niche segments [89,90]. Rising customer service expectations [91] and the loss of the host's service advantages [90] are the main reasons for the downward trend. Retaining and developing service advantages in the mainstream market is a key direction for platforms and hosts to improve their service quality and reputation. Differentiated operations and unique local services are solutions that exploit Airbnb's greatest strengths.

Airbnb, a global sharing accommodation platform, faces varied service disadvantages in different regions. The platform can use big data and artificial intelligence technology to identify service problems in different regions from eWOM; and then, it can provide differentiated management and practice solutions. Take San Francisco and Beijing for examples: their poor service quality in the tangibles dimension is still getting worse. They present three same negative high-frequency topics: clean, bathroom, and kitchen. In addition, hosts in Beijing need to pay attention to "设施(facility)" and "床(bed)", while those in San Francisco shall make improvements in "room" and "(not) beautiful". These negative high-frequency keywords reflect customers' dissatisfaction with housing facilities. Therefore, it gives cues for the platform to formulate certain criteria for the facilities of the listings. The listings that fail to meet the criteria should be tagged on the platform, which helps customers understand the real situation before making a reservation. Such information serves an effective reference for customers' decision-making, while driving hosts to address these service issues.

In addition, customers have put forward higher requirements for the

Airbnb platform as a rule maker and conflict mediator. Complaints from customers that the platform favors hosts can often be seen on social media. Balancing the interests of customers and hosts is a major challenge for Airbnb.

8. Conclusion

This is the first study, to the best of our knowledge, to focus on the dynamic changes in service quality of sharing accommodation. We found a small downward trend in overall customers' perception of service quality, and this trend is expected to continue. As a disruptive innovation to traditional accommodation, retaining and developing service advantages in the mainstream market is a key direction for platforms and hosts to improve their service quality and reputation. Additionally, balancing the interests of customers and hosts is a major challenge for platforms.

However, there is a limitation in this study. Due to the complexity and difficulty of multilingual text processing and time constraints, this study only considers data from San Francisco and Beijing. In the future, with the improvement of multilingual natural language processing technology, we can test the model in more regions with different language backgrounds to explore the dynamics of service quality.

Author statement

Wenming Zuo: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing - Original Draft, Funding Acquisition, Resources, Supervision; **Weijing Bai:** Conceptualization, Software, Validation, Visualization, Investigation, Formal Analysis, Writing - Original Draft, Writing - Review & Editing; **Wenfeng Zhu:** Conceptualization, Data Curation, Software, Validation, Investigation, Writing - Original Draft; **Xinming He:** Writing - Review & Editing; **Xinxin Qiu:** Writing - Review & Editing;

Acknowledgments

The authors are very grateful to editors and referees for their valuable comments and suggestions that helped to improve the quality of the original work. This work was supported by the National Office for Philosophy and Social Sciences [grant numbers 18BGL110 and 20FGLB034], the Natural Science Foundation of Guangdong Province [grant number 2020A1515010830], and the Pazhou Lab [grant number PZL2021KF0027].

References

[1] R. Adner, When are technologies disruptive? A demand-based view of the emergence of competition, Strat. Manag. J. 23 (8) (2002) 667–688.
[2] D. Guttentag, Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector, Curr. Issues Tourism 18 (12) (2015) 1192–1217.
[3] Y. Heo, Sharing economy and prospects in tourism research, Ann. Tourism Res. 58 (2016) 166–170.
[4] B. Richard, S. Cleveland, The future of hotel chains: branded marketplaces driven by the sharing economy, J. Vacat. Mark. 22 (3) (2016) 239–248.
[5] B.B. Dedeoğlu, H. Demİrer, Differences in service quality perceptions of stakeholders in the hotel industry, Int. J. Contemp. Hospit. Manag. 27 (1) (2015) 130–146.
[6] H. Wilkins, B. Merrilees, C. Herington, Towards an understanding of total service quality in hotels, Int. J. Hospit. Manag. 26 (4) (2007) 840–853.
[7] I.K.W. Lai, M. Hitchcock, Literature review on service quality in hospitality and tourism (1984-2014): future directions and trends, Int. J. Contemp. Hospit. Manag. 30 (1) (2018) 114–159.
[8] C.V. Priporas, N. Stylos, R. Rahimi, L.N. Vedanthachari, Unraveling the diverse nature of service quality in a sharing economy: a social exchange theory perspective of Airbnb accommodation, Int. J. Contemp. Hospit. Manag. 29 (9) (2017) 2279–2301.
[9] E. Cristobal-Fransi, F. Hernández-Soriano, B. Ferrer-Rosell, N. Daries, Exploring service quality among online sharing economy platforms from an online media perspective, Sustainability 11 (13) (2019) 3690.
[10] V.A. Zeithaml, Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence, J. Market. 52 (3) (1988) 2–22.
[11] G.C.L. Bezerra, C.F. Gomes, Measuring airport service quality: a multidimensional approach, J. Air Transport. Manag. 53 (2016) 85–93.
[12] B. Guirao, A. García-Pastor, M.E. López-Lambas, The importance of service quality attributes in public transportation: narrowing the gap between scientific research and practitioners' needs, Transport Pol. 49 (2016) 68–77.
[13] W. Li, S. Yu, H. Pei, C. Zhao, B. Tian, A hybrid approach based on fuzzy AHP and 2-tuple fuzzy linguistic method for evaluation in-flight service quality, J. Air Transport. Manag. 60 (2017) 49–64.
[14] K.D. Atalay, B. Atalay, F.B. Isin, FIPIA with information entropy: a new hybrid method to assess airline service quality, J. Air Transport. Manag. 76 (2019) 67–77.
[15] H. Gupta, Evaluating service quality of airline industry using hybrid best worst method and VIKOR, J. Air Transport. Manag. 68 (2018) 35–47.
[16] T.Y. Pham, G.T. Yeo, Evaluation of transshipment container terminals' service quality in Vietnam: from the shipping companies' perspective, Sustainability 11 (5) (2019) 1503.
[17] T. Lupo, E. Bellomo, DINESERV along with fuzzy hierarchical TOPSIS to support the best practices observation and service quality improvement in the restaurant context, Comput. Ind. Eng. 137 (2019), 106046.
[18] K. Berezina, A. Bilgihan, C. Cobanoglu, F. Okumus, Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews, J. Hospit. Market. Manag. 25 (1) (2016) 1–24.
[19] M. Lee, M. Jeong, J. Lee, Roles of negative emotions in customers' perceived helpfulness of hotel reviews on a user-generated review website: a text mining approach, Int. J. Contemp. Hospit. Manag. 29 (2) (2017) 762–783.
[20] A. Griffin, J.R. Hauser, The voice of the customer, Market. Sci. 12 (1) (1993) 1–27.
[21] B. Palese, A. Usai, The relative importance of service quality dimensions in E-commerce experiences, Int. J. Inf. Manag. 40 (2018) 132–140.
[22] S. Melián-González, Gig economy delivery services versus professional service companies: consumers' perceptions of food-delivery services, Technol. Soc. 69 (2022), 101969.
[23] A. Brochado, P. Rita, C. Oliveira, F. Brochado, Airline passengers' perceptions of service quality: themes in online reviews, Int. J. Contemp. Hospit. Manag. 31 (2) (2019) 855–873.
[24] N. Korfiatis, P. Stamolampros, P. Kourouthanassis, V. Sagiadinos, Measuring service quality from unstructured data: a topic modeling application on airline passengers' online reviews, Expert Syst. Appl. 116 (2019) 472–486.
[25] I. Sutherland, K. Kiatkawsin, Determinants of guest experience in Airbnb: a topic modeling approach using LDA, Sustainability 12 (8) (2020) 3402.
[26] J. Büschken, G.M. Allenby, Sentence-based text analysis for customer reviews, Market. Sci. 35 (6) (2016) 953–975.
[27] C. Li, S. Chen, J. Xing, A. Sun, Z. Ma, Seed-guided topic model for document filtering and classification, ACM Trans. Inf. Syst. 37 (1) (2018) 1–37.
[28] J. de Oña, R. de Oña, G. López, Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation, Transportation 43 (5) (2016) 725–747.
[29] F. Rahman, T. Das, M. Hadiuzzaman, S. Hossain, Perceived service quality of paratransit in developing countries: a structural equation approach, Transport. Res. Pol. Pract. 93 (2016) 23–38.
[30] M. Blut, E-service quality: development of a hierarchical model, J. Retailing 92 (4) (2016) 500–517.
[31] N. Aydin, A fuzzy-based multi-dimensional and multi-period service quality evaluation outline for rail transit systems, Transport Pol. 55 (2017) 87–98.
[32] S. Miranda, P. Tavares, R. Queiró, Perceived service quality and customer satisfaction: a fuzzy set QCA approach in the railway sector, J. Bus. Res. 89 (2018) 371–377.
[33] M. Theodosiou, E. Katsikea, S. Samiee, K. Makri, A comparison of formative versus reflective approaches for the measurement of electronic service quality, J. Interact. Market. 47 (2019) 53–67.
[34] L. Ocampo, J. Alinsub, R.A. Casul, G. Enquig, M. Luar, N. Panuncillon, C. O. Ocampo, Public service quality evaluation with SERVQUAL and AHP-TOPSIS: a case of Philippine government agencies, Soc. Econ. Plann. Sci. 68 (2019), 100604.
[35] G. Tuzkaya, B. Sennaroglu, Z.T. Kalender, M. Mutlu, Hospital service quality evaluation with IVIF-PROMETHEE and a case study, Soc. Econ. Plann. Sci. 68 (2019), 100705.
[36] A. Mirzaei, S.R. Carter, J.Y. Chen, C. Rittsteuer, C.R. Schneider, Development of a questionnaire to measure consumers' perceptions of service quality in community pharmacies, Res. Soc. Adm. Pharm. 15 (4) (2019) 346–357.
[37] L. Fei, J. Lu, Y. Feng, An extended best-worst multi-criteria decision-making method by belief functions and its applications in hospital service evaluation, Comput. Ind. Eng. 142 (2020), 106355.
[38] Y. Ming, L. Luo, X. Wu, H. Liao, B. Lev, L. Jiang, Managing patient satisfaction in a blood-collection room by the probabilistic linguistic gained and lost dominance score method integrated with the best-worst method, Comput. Ind. Eng. 145 (2020), 106547.
[39] F. Barrios-Ipenza, A. Calvo-Mora, F. Criado-García, W.H. Curioso, Quality evaluation of health services using the kano model in two hospitals in Peru, Int. J. Environ. Res. Publ. Health 18 (11) (2021) 6159.
[40] H. Wang, J. Yao, X. Zhang, Y. Zhang, An area similarity measure for trapezoidal interval type-2 fuzzy sets and its application to service quality evaluation, Int. J. Fuzzy Syst. (2021) 1–18.
[41] R.C.D. Carvalho, D.D.D. Medeiros, Assessing quality of air transport service: a comparative analysis of two evaluation models, Curr. Issues Tourism 24 (8) (2021) 1123–1138.
[42] F. Li, H. Lu, M. Hou, K. Cui, M. Darbandi, Customer satisfaction with bank services: the role of cloud services, security, e-learning and service quality, Technol. Soc. 64 (2021) 101487.

[43] Y. Ju, K.J. Back, Y. Choi, J.S. Lee, Exploring Airbnb service quality attributes and their asymmetric effects on customer satisfaction, Int. J. Hospit. Manag. 77 (2019) 342–352.
[44] C.V. Priporas, N. Stylos, L.N. Vedanthachari, P. Santiwatana, Service quality, satisfaction, and customer loyalty in Airbnb accommodation in Thailand, Int. J. Tourism Res. 19 (6) (2017) 693–704.
[45] H. Moon, L. Miao, L. Hanks, N.D. Line, Peer-to-peer interactions: perspectives of Airbnb guests and hosts, Int. J. Hospit. Manag. 77 (2019) 405–414.
[46] F.R. Dwyer, P.H. Schurr, S. Oh, Developing buyer-seller relationships, J. Market. 51 (2) (1987) 11–27.
[47] A. Parasuraman, V.A. Zeithaml, L. Berry, SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality, vol. 64, 1988, pp. 12–40, 1988 1.
[48] M.S. Rosenbaum, I.A. Wong, Modeling customer equity, SERVQUAL, and ethnocentrism: a Vietnamese case study, J. Serv. Manag. 20 (5) (2009) 544–560.
[49] H. Dinçer, S. Yüksel, L. Martinez, Analysis of balanced scorecard-based SERVQUAL criteria based on hesitant decision-making approaches, Comput. Ind. Eng. 131 (2019) 1–12.
[50] Z. Shao, X. Li, Y. Guo, L. Zhang, Influence of service quality in sharing economy: understanding customers' continuance intention of bicycle sharing, Electron. Commer. Res. Appl. 40 (2020), 100944.
[51] B. Knutson, P. Stevens, C. Wullaert, M. Patton, F. Yokoyama, LODGSERV: a service quality index for the lodging industry, Hospitality Research Journal 14 (2) (1990) 277–284.
[52] J.M. Getty, K.N. Thompson, A procedure for scaling perceptions of lodging quality, Hospitality Research Journal 18 (2) (1994) 75–96.
[53] J.M. Getty, R.L. Getty, Lodging quality index (LQI): assessing customers' perceptions of quality delivery, Int. J. Contemp. Hospit. Manag. 15 (2) (2003) 94–104.
[54] G.J. Udo, K.K. Bagchi, P.J. Kirs, Using SERVQUAL to assess the quality of e-learning experience, Comput. Hum. Behav. 27 (3) (2011) 1272–1283.
[55] A. Parasuraman, V.A. Zeithaml, A. Malhotra, ES-QUAL: a multiple-item scale for assessing electronic service quality, J. Serv. Res. 7 (3) (2005) 213–233.
[56] J. Zhang, What's yours is mine: exploring customer voice on Airbnb using text-mining approaches, J. Consum. Market. 36 (5) (2019) 655–665.
[57] Y. Guo, S.J. Barnes, Q. Jia, Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation, Tourism Manag. 59 (2017) 467–483.
[59] A.M. Evans, O. Stavrova, H. Rosenbusch, Expressions of doubt and trust in online user reviews, Comput. Hum. Behav. 114 (2021), 106556.
[60] E. Sidnam-Mauch, L. Bighash, How controversy leads to commitment: predecisional distortion in reactions to premarket products through online review systems, Comput. Hum. Behav. 124 (2021), 106902.
[61] R. Panda, S. Verma, B. Mehta, Emergence and acceptance of sharing economy in India: understanding through the case of Airbnb, Int. J. Online Market. 5 (3) (2015) 1–17.
[62] W.C. Tsao, M.T. Hsieh, L.W. Shih, T.M. Lin, Compliance with eWOM: the influence of hotel reviews on booking intention from the perspective of consumer conformity, Int. J. Hospit. Manag. 46 (2015) 99–111.
[63] B.M. Noone, K.A. McGuire, Effects of price and user-generated content on consumers' prepurchase evaluations of variably priced services, J. Hospit. Tourism Res. 38 (4) (2014) 562–581.
[64] A.G. Mauri, R. Minazzi, Web reviews influence on expectations and purchasing intentions of hotel potential customers, Int. J. Hospit. Manag. 34 (2013) 99–107.
[65] B.A. Sparks, V. Browning, The impact of online reviews on hotel booking intentions and perception of trust, Tourism Manag. 32 (6) (2011) 1310–1323.
[66] J. Zhang, Listening to the consumer: exploring review topics on Airbnb and their impact on listing performance, J. Market. Theor. Pract. 27 (4) (2019) 371–389.
[67] A. Babić Rosario, F. Sotgiu, K. De Valck, T.H. Bijmolt, The effect of electronic word of mouth on sales: a meta-analytic review of platform, product, and metric factors, J. Market. Res. 53 (3) (2016) 297–318.
[68] C.K. Anderson, S. Han, Hotel Performance Impact of Socially Engaging with Consumers, 2016.
[69] K. Floyd, R. Freling, S. Alhoqail, H.Y. Cho, T. Freling, How online product reviews affect retail sales: a meta-analysis, J. Retailing 90 (2) (2014) 217–232.
[70] Q. Ye, R. Law, B. Gu, W. Chen, The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings, Comput. Hum. Behav. 27 (2) (2011) 634–639.
[71] Z. Alrawadieh, R. Law, Determinants of hotel guests' satisfaction from the perspective of online hotel reviewers, Int. J. Cult. Tourism Hospit. Res. 13 (1) (2019) 84–87.
[72] N. Hu, T. Zhang, B. Gao, I. Bose, What do hotel customers complain about? Text analysis using structural topic model, Tourism Manag. 72 (2019) 417–426.
[73] B. Kim, S.S. Kim, C.Y. Heo, Analysis of satisfiers and dissatisfiers in online hotel reviews on social media, Int. J. Contemp. Hospit. Manag. 28 (9) (2016) 1915–1936.
[74] L. Zhou, S. Ye, P.L. Pearce, M.Y. Wu, Refreshing hotel satisfaction studies by reconfiguring customer review data, Int. J. Hospit. Manag. 38 (2014) 1–10.
[75] H. Li, Q. Ye, R. Law, Determinants of customer satisfaction in the hotel industry: an application of online review analysis, Asia Pac. J. Tourism Res. 18 (7) (2013) 784–802.
[76] P.C.B. Smyth, G. Wu, D. Greene, Does TripAdvisor Makes Hotels Better, Derek Greene School of Computer Science & Informatics, University College Dublin Belfield, 2010.
[77] D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
[78] Y. Zhu, M. Cheng, J. Wang, L. Ma, R. Jiang, The construction of home feeling by Airbnb guests in the sharing economy: a semantics perspective, Ann. Tourism Res. 75 (2019) 308–321.
[79] C.M. Christensen, The Innovator's Dilemma: when New Technologies Cause Great Firms to Fail, Harvard Business Review Press, 2013.
[80] R. Pera, G. Viglia, L. Grazzini, D. Dalli, When empathy prevents negative reviewing behavior, Ann. Tourism Res. 75 (2019) 265–278.
[81] D. Guttentag, S. Smith, L. Potwarka, M. Havitz, Why tourists choose Airbnb: a motivation-based segmentation study, J. Trav. Res. 57 (3) (2018) 342–359.
[82] J. Hamari, M. Sjöklint, A. Ukkonen, The sharing economy: why people participate in collaborative consumption, Journal of the association for information science and technology 67 (9) (2016) 2047–2059.
[83] E. Ert, A. Fleischer, N. Magen, Trust and reputation in the sharing economy: the role of personal photos in Airbnb, Tourism Manag. 55 (2016) 62–73.
[84] I.P. Tussyadiah, J. Pesonen, Impacts of peer-to-peer accommodation use on travel patterns, J. Trav. Res. 55 (8) (2016) 1022–1040.
[85] S.Q. Liu, A.S. Mattila, Airbnb: online targeted advertising, sense of power, and consumer decisions, Int. J. Hospit. Manag. 60 (2017) 33–41.
[86] S. Shuqair, D.C. Pinto, A.S. Mattila, Benefits of authenticity: post-failure loyalty in the sharing economy, Ann. Tourism Res. 78 (2019), 102741.
[87] R. Lieber, Questions about Airbnb's responsibility after attack by dog, N. Y. (2015). Retrieved, https://www.nytimes.com/2015/04/11/your-money/questions-about-airbnbs-responsibility-after-vicious-attack-by-dog.html. (Accessed 3 January 2016).
[88] R. Lieber, Airbnb horror story points to need for precautions, N. Y. (2015). Retrieved from, https://www.nytimes.com/2015/08/15/your-money/airbnb-horror-story-points-to-need-for-precautions.html.
[89] A. Carr, What Hotel Operators Really Think of Airbnb, 2014 retrieved 02.02.22, https://www.fastcompany.com/3027976/what-hotel-operators-really-think-of-airbnb.
[90] D. Wang, J.L. Nicolau, Price determinants of sharing economy based accommodation rental: a study of listings from 33 cities on Airbnb. com, Int. J. Hospit. Manag. 62 (2017) 120–131.
[91] D. Huang, A. Coghlan, X. Jin, Understanding the drivers of Airbnb discontinuance, Ann. Tourism Res. 80 (2020), 102798.
[92] J. Ban, R.R. Ramsaran, An exploratory examination of service quality attributes in the ecotourism industry, J. Travel Tour. Mark. Res. 34 (1) (2017) 132–148.