



Service quality evaluation and service improvement using online reviews: A framework combining deep learning with a hierarchical service quality model

Xin-Xin Liu^{a,b}, Zhen-Yu Chen^{a,*}

^a Department of Information Management and Decision Sciences, School of Business Administration, Northeastern University, Shenyang 110169, China

^b School of Management, Harbin Institute of Technology, Harbin 150001, China

ARTICLE INFO

Keywords:

Service quality evaluation
Service improvement
Deep learning
Online reviews
Service quality model

ABSTRACT

In the era of big data, service quality evaluation using online reviews has become a popular topic. However, very few studies focus simultaneously on service quality evaluation and service improvement. In this study, a research framework for service quality evaluation and service improvement is proposed, sentiment analysis is used to extract the temporal scores of the service attributes of each subdimension of the service quality model from online reviews, and a long short-term memory network is used to predict the scores for the service quality provider. Furthermore, a long short-term memory network-based sensitivity analysis, in conjunction with improvement costs, is used to rank the subdimensions in the service quality model. Then, service improvement strategies are determined according to the rankings of the service attributes. Hotels' online reviews were used to investigate the effectiveness of the proposed framework. A series of service improvement strategies for the specific service attributes are provided.

1. Introduction

Compared with a physical product, a service has the characteristics of intangibility, heterogeneity and inseparability, which make it difficult to evaluate service quality (Parasuraman et al., 1985). Comprehensive evaluation methods are usually applied to aggregate the scores for each dimension of a service quality model to obtain the overall service quality evaluation result. Representative comprehensive evaluation methods include the analytic hierarchy process (Yucesan and Gul, 2020), fuzzy comprehensive evaluation (Wei et al. 2015), the technique of order preference by similarity to the ideal solution (TOPSIS) (Yucesan and Gul, 2020), the decisions making trial and evaluation laboratory (DEMATEL) (Tseng, 2009), and data envelopment analysis (DEA) (Lee and Kim, 2014). These methods of service quality evaluation have the shortcoming of relying on experts' subjective scoring data and often lack sufficient samples (Wei et al. 2015). In addition, the Kano model (Hsu et al., 2018; Bi et al., 2019; Qi et al., 2016) and importance-performance analysis (IPA) (Deng et al., 2008) are often used for service quality evaluation. In addition to comprehensive evaluation methods, multivariate regression is applied to rank the importance of service quality dimensions (Palese and Usai, 2018). However, nonlinear statistical

methods and machine learning methods are rarely used. In this study, the time-varying characteristics of service quality are considered, and long short-term memory network (LSTM), an advanced machine learning method, is used to effectively deal with time series data.

With the rise of the Internet and E-commerce, it is becoming more common for users to share their service experiences in online reviews. As a hot topic in the last several years, the researchers used text mining technologies for service satisfaction analysis and customer satisfaction prediction (Mejia et al., 2021; Xu et al., 2021; Zhao et al., 2019; Zhao et al., 2022). Related studies have extracted service attributes from online reviews using word frequencies (Xiang et al., 2015) and topic models (Guo et al., 2017), evaluated service satisfaction using sentiment analysis (Deng et al., 2021; Geetha et al., 2017), and mined comparative relationships to determine service competitiveness (Gao et al., 2018). Based on the service attributes extracted from online reviews, machine learning methods such as artificial neural networks (ANNs) (Leong et al., 2015; Zeinalizadeh et al., 2015) and Bayesian methods (Nourikhah, and Akbari, 2016) have been used for customer satisfaction prediction.

Unlike the studies on customer satisfaction evaluation and prediction mentioned above, relatively few studies focus on text mining of online reviews to perform overall service quality evaluation at the enterprise

* Corresponding author.

E-mail address: zychen@mail.neu.edu.cn (Z.-Y. Chen).

<https://doi.org/10.1016/j.elerap.2022.101174>

Received 15 January 2022; Received in revised form 20 June 2022; Accepted 6 July 2022

Available online 14 July 2022

1567-4223/© 2022 Elsevier B.V. All rights reserved.

level. In Saha and Theingi (2009), satisfaction is viewed as the antecedent of the perceived service quality, while in Parasuraman et al. (1985, 1988), the perceived service quality is viewed as the antecedent of satisfaction. This study adopts the opinion that service quality is the antecedent of satisfaction and uses online reviews as a reflection of customer satisfaction to directly evaluate the service quality. Existing studies cannot evaluate the overall service quality using the extracted service attributes and cannot guarantee the completeness of the extracted service attributes (James et al., 2017). This shortcoming can be addressed by combining service quality models with text mining techniques (Palese and Usai 2018). Moreover, very few studies simultaneously focus on service quality evaluation and service improvement. Service quality evaluation and service improvement can be viewed as one sequentially connected system, although they are usually studied separately. For example, text mining is used to screen and identify service attributes from online reviews for service quality improvement (Palese and Usai 2018).

With the rise of data analysis technologies, service quality evaluation and service improvement based on text mining and machine learning approaches are potential research directions and have not received widespread attention. To the best of our knowledge, deep learning approaches have not been applied to the study of service quality evaluation and service improvement.

In this study, a research framework for service quality evaluation and service improvement that combines the deep learning approach and the hierarchical service quality model is proposed. In this framework, the temporal sentiment scores of the service attributes of each subdimension in the service quality model are obtained by a lexicon-based sentiment analysis of online reviews. The comprehensive evaluation results of service quality using the temporal sentiment scores of service attributes are obtained with a LSTM network. Based on the trained LSTM, a sensitivity analysis is conducted to obtain the importance of each subdimension in the service quality model. Combined with the interest degree of the service attributes extracted from online reviews, the improvement scores of the service quality subdimensions and the reasons for dissatisfaction mined from online reviews, service improvement strategies can be provided.

The contributions of this study are summarized as follows: (1) This study takes advantage of both service quality models and text mining techniques to extract the performance values of service quality subdimensions from online reviews. (2) To the best of our knowledge, this is the first study to evaluate service quality by using LSTM to consider the time-varying characteristics of service quality and the first to rank the service quality subdimensions by combining LSTM with a sensitivity analysis. (3) This study considers service quality evaluation and service improvement as a system. To the best of our knowledge, this is the first study to develop service quality improvement strategies by simultaneously considering the importance degrees of service quality subdimensions obtained by sensitivity analysis, the interest degrees of service attributes obtained by text mining, and the improvement cost of service attributes.

The remainder of this paper is organized as follows: A literature review of the relevant research in this field is provided in Section 2. The research framework and methods are introduced in Section 3. The data description and sentiment analysis are given in Section 4. The experimental design and results and the management implications are provided in Section 5. Section 6 summarizes the weaknesses and proposes future research directions.

2. Literature review

2.1. Service quality and service quality model

In the first stage of research on service quality, the concept of perceived service quality, in which the perceived service is compared with the expected service, was put forward as a landmark work

(Grönroos, 1984). In the opinion of Cronin and Taylor (1992), service quality can be measured only by the service perceived by customers, regardless of the expected service. In the second stage of service quality research, studies focused on the service quality model (SERVQUAL) (Parasuraman et al. 1985) and the performance-based measure of service quality (SERVPERF) (Leong et al., 2015). Questionnaires, focus group interviews and executive interviews are usually used to obtain the required information in this service quality model (Parasuraman et al. 1985; Hsu et al., 2018).

Parasuraman et al. (1988) built the SERVQUAL model, which considers the gap between the expected service and perceived service to be the service quality. In the SERVQUAL model, a 22-item instrument is used to assess customer perceptions of service quality, and five dimensions that influence service quality—tangibles, reliability, responsiveness, assurances and empathy—are developed. The SERVQUAL model has the limitation of being unable to capture the technical quality. Cronin and Taylor (1992) proposed the SERVPERF model, which directly measures the perceived service quality, although it does not consider the expected service quality. In the SERVPERF model, the five dimensions and 22 items of SERVQUAL are completely retained. Dabholkar et al. (1996) proposed a hierarchical service quality model with three levels: the customers' overall perceptions of service quality, the primary dimensions and the subdimensions. Brady and Cronin (2001) used the idea of a hierarchical service quality model to sort and analyze the influence of the 22 items in the five dimensions of the SERVQUAL model and proposed a new multilevel, multidimensional service quality model.

Service quality evaluation for hotel industries usually uses the models described above or modifications of them, such as LODGSERV (Knutson et al., 1990), GLSERV (Lee and Cheng, 2018). Among them, the LODGSERV scale has been adopted by many scholars, but it is not as popular among hospitality and tourism researchers as the SERVQUAL (Carrasco et al., 2017). In this study, the service quality model proposed by Brady and Cronin (2001) is adopted. The model proposed by Brady and Cronin (2001) combines multiple service quality conceptualizations into a comprehensive, hierarchical and multidimensional framework, that contains three dimensions directly determine service quality and nine subdimensions. The combination of all the dimensions and subdimensions constitutes a customer's overall perception of the service quality (Brady and Cronin, 2001).

2.2. Product and service satisfaction based on online reviews

With the development of e-commerce, a large number of online reviews for products and services have become available. Compared with digital ratings, online reviews can reflect satisfaction levels with more detailed information. Consequently, studies of product/service satisfaction using online reviews have become popular (Bi et al. 2019). The existing studies of product/service satisfaction mainly focus on customer satisfaction at the individual level (Leong et al. 2015; Li et al., 2020; Zeinalizadeh et al. 2015; Zhang et al., 2021; Zhao et al. 2019). There are a few studies on service satisfaction at the aggregate level. In these studies, the average of all customers' satisfaction degrees is used as the overall service satisfaction degree of a hotel (Kim et al., 2015; Xie et al., 2014; Ye et al., 2011). However, it is very possible that the relationship between customer satisfaction and overall service satisfaction is not linear because of the incomplete sampling of customers in online reviews.

The research on customer satisfaction based on online reviews can be divided into the extraction of product/service dimensions or attributes and the impacts of product/service dimensions or attributes on customer satisfaction (Bi et al. 2019). Studying the linguistic attributes of online reviews and employing text mining techniques are two ways to extract product/service dimensions or attributes. The linguistic attributes of online reviews include subjectivity, diversity, readability, length and breadth (Palese and Usai 2018; Zhao et al. 2019). Commonly used text

mining techniques include frequency analysis (Xiang et al. 2015), sentiment analysis (Deng et al., 2021; Geetha et al. 2017; Hong et al., 2020; Zhang et al., 2021; Zuo et al., 2019) and topic modeling (Palese and Usai 2018; Chung et al., 2022). Research on the impacts of linguistic attributes on customer satisfaction is not helpful for hotel service improvement. Moreover, the validity and completeness of product/service dimensions extracted by text mining techniques cannot be verified. This limitation can be overcome by combining text mining techniques with service quality models. As far as we know, Palese and Usai (2018) is the only study that uses the combination method to determine service quality dimensions. In Palese and Usai (2018), a service quality model was combined with a topic model; the number of sentences describing the same topic was used as the measure of the service quality dimensions, and multivariate regression was used to evaluate the impacts of different service quality dimensions on the service quality. In contrast to Palese and Usai (2018), our study combines sentiment analysis with the hierarchical service quality model to more accurately determine the measure of the service attributes in each service quality dimension, and LSTM-based sensitivity analysis is used for service quality evaluation and service improvement.

2.3. Service quality improvement

For service quality improvement, it is necessary to understand the importance of service quality dimensions or service attributes and then target the important service quality dimensions or service attributes to improve. Traditional methods adopt subjective judgments for service quality improvement (Farrington et al., 2018). The scores of service evaluation indexes are obtained from experts or customers, and low scores are considered to indicate service quality defects to be improved (Farrington et al. 2018; Wei et al. 2015). For IPA, the order of service attributes to be improved is determined by the quadrant of the service attribute (Deng et al., 2008). In the Kano model, service attributes are divided into five categories— attractive, one-dimensional, must-be, indifference and reverse—and the order of the service attributes to be improved is determined according to the different influences of the categories on customer satisfaction (Bi et al., 2019; Hsu et al., 2018). For the Kano model, when the performance of service attributes in the same category is similar, their importance cannot be sorted. For IPA, a few strong assumptions are made, and these assumptions are seldom satisfied.

Unlike the above qualitative methods, some researchers have developed quantitative methods to obtain the relative importance of service attributes. Li et al., (2010) used graph-based methods to obtain the weights of service aspects as the ranking values. The weights of the input layer in an ANN were used as the rankings of the input factors (Zeinalizadeh et al. 2015). Nourikhah, and Akbari (2016) proposed a Bayesian model to calculate the relative importance of each service attribute. The absolute values of the standardized coefficients in a multivariate regression were taken as the relative importance of the linguistic attributes of online reviews (Zhao et al., 2019). Zuo et al. (2019) used some data mining techniques to identify service issues, and used a visualized tool, *i.e.*, process chain network, to make service quality management and optimization.

2.4. Related methods

Three methods, including LSTM, sentiment analysis, and sensitivity analysis, are involved in this study. LSTM, as a type of recurrent neural networks (RNN), is suitable for time series data over a long period and can handle vanishing and exploding gradient problems in RNN (Liang and Cai, 2020). LSTM has been extensively used for a lot of domains such as financial market (Liang and Cai, 2020; Fischer and Krauss, 2018), natural language processing (NLP) (Ghasemi and Momtazi, 2021) and sale prediction (He et al., 2022). Fischer and Krauss (2018) used LSTM to predict out-of-sample directional movement for the

constituent stocks and achieved better performances than random forest, deep neural network, and logistic regression. In Liang and Cai (2020), LSTM was adopted to forecasting peer-to-peer platform default rate. Ghasemi and Momtazi (2021) proposed a LSTM based model to improve recommender system performances by considering the text similarity of online reviews. He et al. (2022) combined LSTM with the particle swarm optimization for sale forecasting in E-commerce companies. Chen et al. (2020) proposing a novel LSTM to model the mouse interaction data for search satisfaction evaluation. To the best of our knowledge, besides Chen et al. (2020), there are no studies for service quality evaluation by using LSTM.

Sentiment analysis is a natural language processing (NLP) technique that can quantitatively extracts sentiment scores from online reviews (Deng et al., 2021). Qi et al., (2016) used sentiment analysis to determine the weights of the product attributes. Deng et al. (2021) classified the polarity of a restaurant review into positive, negative, or neutral by conducting sentiment analysis. Sentiment analysis can also be used to examine the relationship between information embedded in a text and the downstream outcomes such as customer satisfaction (rating). Geetha et al. (2017) established a relationship between customer sentiments in online reviews and customer satisfaction for hotels.

Sensitivity analysis can measure how an output of neural networks is influenced by its input perturbations (Yeung et al., 2010). Leong et al. (2015) used ANN-based sensitivity analysis to calculate the value of the sensitivity for each input variable and then took this value as an indicator of the importance of the predictors. In this study, the LSTM-based sensitivity analysis is used to calculate the importance of each service quality subdimension for service quality improvement.

3. Research framework and methodology

The research framework consists of three major steps: service attribute extraction, LSTM-based service quality evaluation and service quality improvement. The research framework and the hierarchical service quality model are shown in Fig. 1 and Fig. 2, respectively.

3.1. Service attribute extraction

The idea of SERVPERF is used, and the process of service quality evaluation is viewed as the process of determining service satisfaction. Therefore, the customer experience described in online reviews is adopted to build the service quality dimensions of the customers' perceived service quality. The research framework adopts hierarchical service quality model (Brady and Cronin 2001), which divides the service dimensions in a detailed way and integrates the technical quality, that is related to what a customer receives, functional quality, that is related to how a customer receives it, and the dimensions of SERVQUAL. Three primary dimensions of service quality are the interaction quality, physical environment quality and outcome quality. Each of these dimensions contains three subdimensions. As shown in Fig. 2, a modification is made by moving the valence subdimension to the "social factors" subdimension and putting emerging factors into the "others" subdimension.

Service attribute extraction follows the idea of classifying service attribute words in a predefined service attribute lexicon according to the subdimensions of service quality (Palese and Usai 2018). Furthermore, instead of using the number of topics as in Palese and Usai (2018), a lexicon-based sentiment analysis is adopted to calculate the sum of the sentiment scores of all service attributes of each subdimension, and then the sum of the sentiment scores is used as the performance of that subdimension. Based on the sentiment lexicon (Qi et al., 2016), the sentiment scores of service attributes are calculated as follows:

$$x_{it} = \sum_{j=1}^{J_i} \sum_{h=1}^H adv_{jih} \times s_{jih} \times (-1)^{jkh} \quad (1)$$

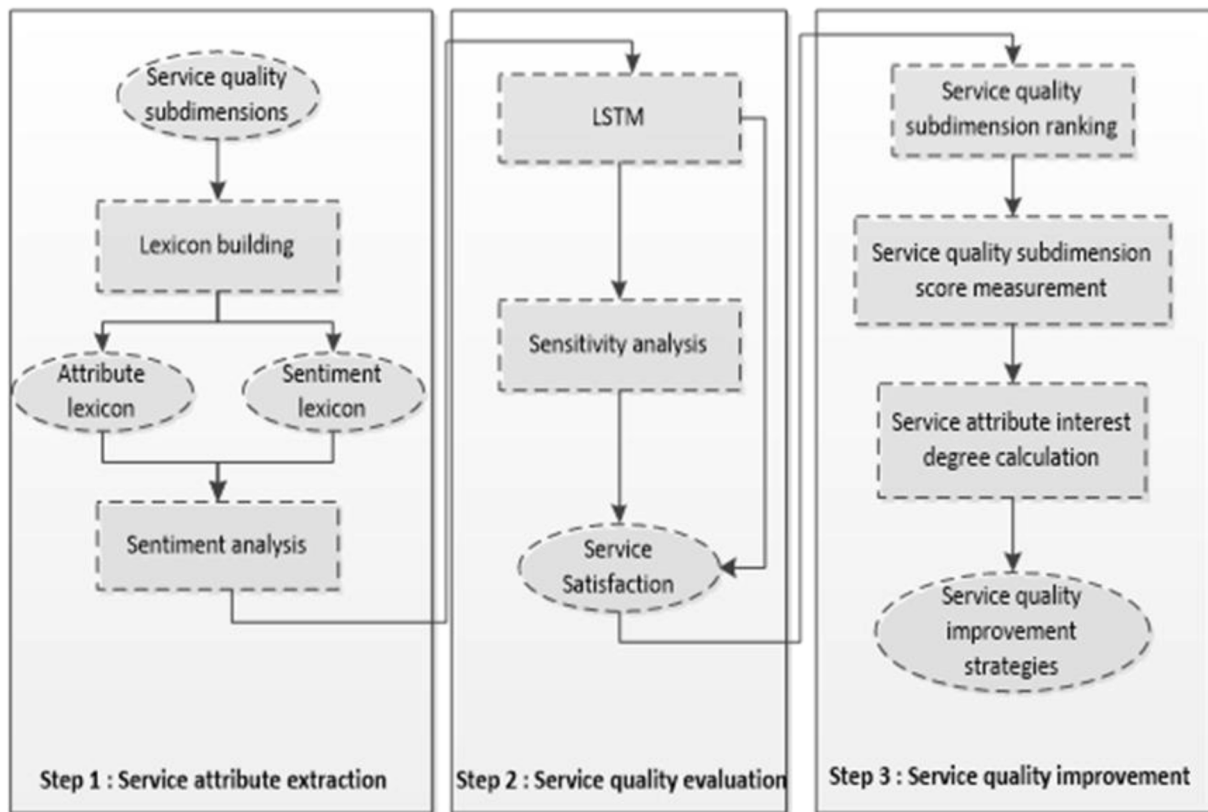


Fig. 1. The research framework.

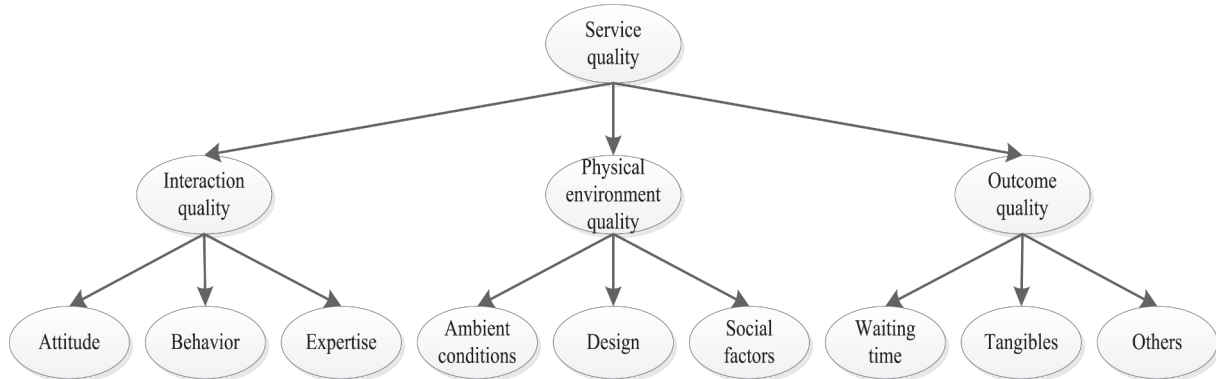


Fig. 2. Hierarchical service quality model.

where x_{it} represents the sentiment score of all service attributes in the i th subdimension of service quality in period t , J_i is the set of all service attribute words in the i th subdimension, h represents the word frequencies of service attribute word j , s_{jh} represents the value of the sentiment strength when service attribute word j appears h times, adv_{jh} is the sentiment strength of s_{jh} , and $(-1)^{jhk}$ represents that negative words appear k times when their related service attribute word j appears h times.

3.2. LSTM-based service quality evaluation

Even for the same hotel, customers may perceive different service quality across times. To overcome the problem of being unable to consider the dynamics of the service attributes in service quality evaluation by using artificial neural network (ANN) (Leong et al., 2015), this

study adopts LSTM, which is suitable for time series data. A recurrent neural network (RNN) is a traditional deep learning method for time series data. The RNN is composed of an input layer, hidden layer (recurrent layer) and output layer. Liang and Cai (2020) provide the architecture of an RNN in details. The hidden layer state s_t is calculated according to the input vector x_t and its lagged states s_{t-1} , and the output vector y_t is calculated according to the hidden layer state s_t by iterating Eqs. (2) and (3) over $t = 1, 2, \dots, T$ as follows:

$$s_t = \sigma(W^{ih}x_t + W^{hh}s_{t-1} + b_h) \tag{2}$$

$$y_t = W^{ho}s_t + b_o \tag{3}$$

where W^{ih} , W^{hh} and W^{ho} represent an input-hidden weight matrix, a hidden-hidden weight matrix, and a hidden-output weight matrix, respectively; b_h and b_o represent the hidden and output bias; σ is the activation function; $x_t = (x_{1t}, x_{2t}, \dots, x_{9t})$ represents the sentiment scores

of service attribute words in the nine subdimensions of service quality at time t , and $s_{t-1} = \mathbf{0}$. The sentiment scores of service attribute words are extracted from online reviews according to Eq. (1) and described as nine-dimensional monthly time series.

For the RNN, vanishing and exploding gradients may occur; thus, the RNN cannot handle time series data over a long period (Liang and Cai, 2020). For LSTM, the number of neurons in the input layer is equal to the number of subdimensions in the service quality model, and the neurons in the output layer represent the service quality score. The hotels' actual overall score is used as the label of LSTM.

LSTM differs from RNN in terms of the design of hidden layers. For LSTM, each hidden layer is composed of the memory cell with a gating mechanism regulating the flow of information. LSTM can effectively avoid the problems of vanishing and exploding gradients. Each memory cell is controlled by an input gate, output gate and forget gate to add, output and remove information, respectively. The structure of a memory cell is shown in Fig. A1 of Online Appendix. The calculation process of LSTM (Liang and Cai, 2020) is described as follows:

$$f_t = \sigma(W^{xf}x_t + W^{hf}h_{t-1} + b_f) \quad (4)$$

$$\bar{c}_t = \tanh(W^{xc}x_t + W^{hc}h_{t-1} + b_c) \quad (5)$$

$$i_t = \sigma(W^{xi}x_t + W^{hi}h_{t-1} + b_i) \quad (6)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \bar{c}_t \quad (7)$$

$$o_t = \sigma(W^{xo}x_t + W^{ho}h_{t-1} + b_o) \quad (8)$$

$$h_t = o_t \otimes \tanh(c_t) \quad (9)$$

where h_t is the output vector of the hidden layer at time t ; W^{xf} , W^{xi} and W^{xo} are the weight matrices that connect x_t with the forget gate f_t , the input gate i_t and the output gate o_t , respectively; W^{hf} , W^{hi} and W^{ho} are the weight matrices that connect h_{t-1} with the forget gate f_t , the input gate i_t and the output gate o_t , respectively; c_t and \bar{c}_t represent the memory cell states and candidate values; W^{xc} and W^{hc} are the weight matrices that connect \bar{c}_t with x_t and the output of memory cell h_{t-1} at time $(t-1)$, respectively; and b_f, b_i, b_o and b_c are the biases. The sigmoid function is used as the activation function $\sigma(\cdot)$ for all three gates in Eqs. (4), (6) and (8). The sigmoid function can transform the inputs of the three gates to the interval 0 to 1. Thus, the memory cells can selectively accept information from the inputs x_t and h_{t-1} to make the LSTM able to learn from time series data with a long period. When $t = 1$, the initial value of cell state h_0 and c_0 are 0, the output of the hidden layer h_1 can be calculated and the cell state c_1 can be updated according to Eqs. (3)-(8). When $t \geq 2$, the calculation process for the memory cells is as follows:

(1) The memory cell receives h_{t-1} from the cell in the previous step, and the forget gate of the memory cell operates first, using the inputs x_t and h_{t-1} according to Eq. (4).

(2) The hidden layer determines how much information in the memory cell is updated. The candidate values \bar{c}_t of the memory cell are calculated according to Eq. (5), and the input gate i_t is calculated according to Eq. (6).

(3) The current state c_t of the memory cells is calculated according to Eq. (7).

(4) The output gate o_t is calculated according to Eq. (8), and the output h_t of the hidden layer is calculated according to Eq. (9).

When the outputs of the hidden layer h_t are used to replace the hidden layer state s_t in the RNN, the predicted value y_t of the final service quality score is obtained according to Eq. (3).

For LSTM, the number of neurons in the input layer equal to the number of subdimensions in the service quality model, and the neurons in the output layer represent the predicted service quality scores. The actual scores are used as the labels of LSTM.

3.3. Service quality improvement

Based on the results of a service quality evaluation, the strategies for service quality improvement are obtained by a quantified method. The method includes the following three steps: (1) service quality subdimension ranking, (2) service quality subdimension score measurement, and (3) service attribute interest degree calculation.

Service quality subdimension ranking. The greater score changes in the service quality due to the change in the service quality subdimension score, the more important the service quality subdimension is. In other words, the service quality score is sensitive to this subdimension, and thus, this subdimension can improve service quality score effectively. Unlike the text mining methods used in the existing studies, LSTM-based sensitivity analysis is used to rank the service quality subdimensions in this study. The sensitivity analysis can be combined with the machine learning methods to rank the features (Leong et al., 2015). In this study, the service quality subdimensions are used as the features of LSTM, and thus the sensitivity analysis can be naturally extended to rank the service quality subdimensions. The basic idea of LSTM-based sensitivity analysis is that the perturbation of the input values of each service quality subdimension at each time point is performed in the trained LSTM. Specifically, the perturbation analysis (Yeung et al., 2010) is carried out by increasing the values of the input $x_t = (x_{1t}, x_{2t}, \dots, x_{9t})$ by 10 %.

$$x_{it} = x_{it} \times (1 + 10\%) \quad (10)$$

$$y_{it}^{sen} = |y_t(x_{it}) - y_t(x_{it})| / 10\% \quad (11)$$

$$\bar{y}_i^{sen} = 1/m \sum_{m=1}^m y_{it}^{sen} \quad (12)$$

where x_{it} is the revised input after the input perturbation, $y_t(x)$ is the output of the trained LSTM, y_{it}^{sen} is the sensitivity value of the i th subdimension of the service quality model at time t , m is the number of months, and \bar{y}_i^{sen} is the monthly average of the sensitivity value of y_{it}^{sen} . Then, the monthly average value \bar{y}_i^{sen} is used to rank the service quality subdimensions.

Service quality subdimension score measurement. The service quality subdimensions that have a high sensitivity value have great impacts on the service quality. After the ranking of the subdimensions in the service quality model, the improvement cost is determined considering the improvement costs of the subdimensions. The sensitivity value and the improvement costs of the subdimensions are integrated to obtain the order of the subdimensions to be improved.

$$C_i = \sum_j \beta_j \quad (13)$$

$$O_i = \bar{y}_i^{sen} - \alpha C_i \quad (14)$$

where C_i is the total improvement cost in the i th subdimension, β_j is the improvement cost of attribute j in the i th subdimension, which is determined by leveraging the domain knowledge, α is the cost coefficient, and O_i is the score measure of the i th subdimension. To maintain the competitiveness of a hotel, the subdimensions with high score measures should be improved.

Service attribute interest degree calculation. After determining the score measure of the subdimensions according to Eqs. (10)-(14), it is necessary to further sort the service attributes in the service quality subdimensions. Because each subdimension contains many service attributes, the ranking of the service attributes rather than that of the subdimensions can help manager design specific service improvement strategies. To rank the service attributes, the keywords of each service attribute appearing in online reviews of a firm or a segmentation of firms are extracted and summarized based on text mining. The word

frequency is used as the interest degree of each service attribute for the specific firm or the segmentation of firms.

4. Data description and preprocessing

4.1. Data description

In this study, 728 hotels' online reviews and the corresponding overall ratings, that are used as the actual scores (labels) in LSTM, were crawled from Ctrip (<https://www.ctrip.com>), one of the largest hotel reservation platforms in China. All hotels crawled from the website are located in Shenyang and Dalian in China. Among the hotels, there are 54 five-star hotels, 155 four-star hotels, 112 three-star hotels, and 518 two-star and below hotels. The website [Ctrip.com](https://www.ctrip.com) stipulates that only customers who have stayed at a hotel can post reviews, and thus, the validity and authenticity of online reviews are guaranteed. In this study, online reviews from March to August 2019 were selected. In accordance with the service perception theory, the time of the stay is adopted, while the review time is adopted when the time of the stay is not provided.

4.2. Sentiment analysis

Python 3.7, with the packages pandas, numpy and jieba, is used for data preprocessing. There are more than 20,000 online reviews in the corpus. The word frequency of service attributes was calculated in constructing the service attribute lexicon. We first recruit two research assistants to manually classify the service attribute words according to SERVQUAL, respectively, however, two assistants cannot agree with each other on many attribute words. Then, the hierarchical service quality model was chosen, and the classification results are shown in Table A1 of Online Appendix.

In this study, the sentiment lexicon in HowNet is used to calculate the sentiment scores of the service attributes. Before the calculation, the non-Chinese text and the stop words were removed, the negative words were collected manually, and a sentiment strength lexicon with six levels was built. The strength and the corresponding scores s_{jh} in the sentiment strength lexicon are shown in Table A2 of Online Appendix. For each hotel, the sum of the sentiment scores of the service attributes in each subdimension is calculated according to Eq. (1). Hotels with zero sentiment scores and with many missing comments were excluded, and a total of 630 hotels were kept. The sentiment scores of nine service quality subdimensions of 630 hotels for 6 months were obtained, and then the sentiment scores were standardized according to the z-score.

5. Experiments

5.1. Parameter design

The LSTM, RNN and ANN methods were implemented using keras in TensorFlow. The mean squared error (MSE) was used as the loss function, and the mean absolute error (MAE) and MSE were used as the evaluation indexes to calculate the differences between the predicted and actual scores.

To address overfitting, the 5-fold cross-validation was adopted. The L_1 regularized loss function was selected, the regularized parameter was set to 0.001, and the dropout value was set to 0.2. A dropout layer that randomly discards neurons in the hidden layers was added between the hidden and the full connection layer. RMSprop was selected as the optimizer, and its default learning rate was used. The number of neurons in the hidden layers was set to 16, and the sigmoid function was used as the activation function in the hidden layer.

5.2. Service quality evaluation results

The RNN and ANN were used as the benchmarks of the LSTM. The input data and the parameter settings of the RNN were the same as those

of the LSTM. The ANN could not consider temporal effects. Thus, the online reviews were summarized along the time dimension, and the sum of the sentiment scores of the service attributes in each subdimension was calculated and used as the input of the ANN.

The results of the service quality evaluation using the LSTM, RNN, and ANN are provided in Table 1. As shown in this table, the average MAE and MSE of the LSTM are slightly better than those of the RNN and better than those of the ANN.

5.3. Service attribute ranking and service improvement strategies

The sensitivity values of each month of the subdimensions were calculated according to Eqs. (10)-(12), and the results are shown in Table 2. The "others" subdimension is the most important. The "design" and "waiting" subdimensions follow. The monthly sensitivity values and the average normalized sensitivity values of the primary dimension are shown in Table 3.

The improvement costs of the subdimensions were set, and then the improvement scores of the subdimensions could be obtained according to Eq. (14) (see Table 4). As shown in Table 4, the "others" subdimension has the highest improvement score, and the "waiting time" and "behavior" subdimensions have high improvement scores.

After obtaining the improvement scores of the subdimensions, the next problem is to determine the ranking of the service attributes in each subdimension. For example, the "others" subdimension, which related to Wi-Fi, catering and payment methods, has a high improvement score, but which service attributes should be improved? Wi-Fi or food? To address this problem, the interest degrees of service attributes were calculated. For each subdimension, the keywords and their weights were collected from the online reviews of 630 hotels. General nouns such as "hotel" and meaningless words were removed, and the interest degrees of similar service attributes such as shower and bath water were merged. According to the interest degrees of service attributes, the service attributes in each subdimension with a high interest degree were targeted to be improved. Moreover, according to the reasons for dissatisfaction mined from online reviews, the service improvement strategies for the service attributes with a high interest degree can be obtained. When all the online reviews of the hotels are used, the service attribute interest degrees for the corresponding attribute words in a subdimension are reported in Tables 5 and 6 and the improvement strategies for the corresponding attribute words in a subdimension are reported in Tables A3-A7 of Online Appendix. The service attributes in the subdimensions of "waiting time", "ambient conditions", "behavior" and "attitude" had high similarities, and thus, their service improvement strategies are only discussed below and not included in these tables. When the online reviews of a hotel or a segmentation of hotels are selected, the improvement strategies for the hotel(s) can be provided.

In the "others" subdimension, improvement strategies include replacing Wi-Fi facilities and providing high-quality catering and multiple methods of payment (see Table A3 of Online Appendix). In the "waiting time" subdimension, office staff and electronic equipment can be added to improve efficiency, simplify customer handling procedures and reduce the waiting time. It is not easy to improve the hardware of the hotels included in the "design" subdimension. The decoration style can be changed to introduce more themed rooms and to enhance the user experience (see Table A4 of Online Appendix). In the "social factors" subdimension, prices can be moderately reduced, and rooms with different prices and coupons can be offered. If no supermarkets or convenience stores are near a hotel, the hotel can start its own

Table 1
Service quantity evaluation results using LSTM, RNN and ANN.

Measures	LSTM	RNN	ANN
MAE	0.1650	0.1670	0.1790
MSE	0.0252	0.0267	0.0285

Table 2
Monthly sensitivity values (SV) and the normalized average sensitivity values (NASV) of the subdimensions.

Subdimensions	SV						NASV
	Mar.	Apr.	May.	Jun.	Jul.	Aug.	
Attitude	0.0142	0.0196	0.0007	0.0122	0.0146	0.0254	67.72 %
Behavior	0.0005	0.0303	0.0146	0.0193	0.0248	0.0054	74.15 %
Expertise	0.0092	0.0207	0.0082	0.0029	0.0049	0.0219	53.12 %
Ambient conditions	0.0107	0.015	0.0097	0.0147	0.0245	0.0219	75.48 %
Design	0.0014	0.0243	0.0008	0.0223	0.002	0.0605	87.15 %
Social factors	0.0036	0.0033	0.0456	0.0102	0.0269	0.0129	80.19 %
Waiting time	0.0018	0.0083	0.0212	0.019	0.0533	0.0058	85.61 %
Tangibles	0.0194	0.0068	0.0078	0.0026	0.004	0.0288	54.24 %
Others	0.0244	0.0167	0.0238	0.0084	0.036	0.0186	100.00 %

Table 3
The original and normalized sensitivity values of the primary dimensions.

Measures	Interaction quality	Physical environment quality	Outcome quality
Sensitivity values	0.0415	0.0517	0.0511
Normalized sensitivity values	80.30 %	100.00 %	98.78 %

Table 4
Improvement costs and improvement scores of the subdimensions.

Subdimensions	Improvement costs (Y)	Improvement scores
Attitude	300,000	0.0114
Behavior	400,000	0.0118
Expertise	500,000	0.0063
Ambient conditions	900,000	0.0071
Design	700,000	0.0116
Social factors	900,000	0.0081
Waiting time	600,000	0.0122
Tangibles	1,000,000	0.0016
Others	800,000	0.0133

supermarket (see Table A5 of Online Appendix). The “ambient conditions” subdimension is also not easy to improve. To address a traffic problem, providing pick-up and drop-off services can be considered; for the problem of surrounding noise, the persons in charge of the

Table 5
Service attributes interest degrees for the top five attribute words in the subdimensions of others, waiting time, design and ambient conditions.

Others		Waiting time		Design		Social factors		Ambient conditions	
Attribute words	Interest degrees	Attribute words	Interest degrees	Attribute words	Interest degrees	Attribute words	Interest degrees	Attribute words	Interest degrees
Catering	29.66 %	Check in	7.19 %	Fitment	2.66 %	Cost performance	4.81 %	Location	13.47 %
Dining room/hall	1.23 %	Check out	1.42 %	Sound insulation	2.17 %	Transportation	4.28 %	Environment	6.59 %
Wi-Fi signal	1.00 %	Transaction	0.65 %	Style	0.80 %	Trip mode	2.87 %	Surrounding	3.49 %
Take-out	0.32 %	Room reservation	0.20 %	Layout	0.21 %	Price	2.80 %	Noise	0.24 %
Payment method	0.54 %	Settle accounts	0.04 %	Design	0.08 %	Shopping	1.50 %	Outside	0.19 %

Table 6
Service attributes interest degrees for the top five attribute words in the subdimensions of behavior, attitude, expertise and tangible.

Behavior		Attitude		Expertise		Tangible	
Attribute words	Interest degrees	Attribute words	Interest degrees	Attribute words	Interest degrees	Attribute words	Interest degrees
Service	8.31 %	Reception	7.32 %	Hygiene	6.26 %	Facilities	5.77 %
Front desk service	2.50 %	Attitude	7.24 %	Experience	1.66 %	Parking lot	3.16 %
Waiter	2.00 %	Security staff	0.22 %	Professional	0.47 %	Toilet	1.77 %
Staff	0.24 %	Protocol	0.18 %	Politeness	0.45 %	Windows	1.17 %
Server	0.11 %	concierge	0.09 %	Humanization	0.43 %	Lobby	0.98 %

surrounding environment can be consulted. An improvement in the subdimensions of “behavior” and “attitude” can be obtained by providing regular training for the hotel’s service personnel, ensuring stable service outcomes, and providing regular assessment. The implementation of hotel standards and investments in professional team management can be strengthened to gain service improvements in the “expertise” subdimension (see Table A6 of Online Appendix). In the “tangibles” subdimension, the hardware facilities of the hotels can be improved if funds are abundant. In terms of the interest degree of service attributes, customers show more interest in parking lots, toilets, windows and beds, which provides guidance for the design of hotels (see Table A7 of Online Appendix).

5.4. Managerial implications

Different subdimensions of the service quality model have different sensitivity values and improvement scores. As shown in Table 2, the normalized average sensitivity value \bar{y}_i^{sen} ranges from 53.12 % to 100 %. As shown in Table 4, the subdimensions of “others” and “waiting time” have the highest improvement scores, and thus, the service attributes in these subdimensions can be chosen as candidates for improvement. For the service attributes in the “others” subdimension, keywords related to the Internet access, catering and methods of payment are involved. This indicates that Chinese consumers pay attention to wireless networks and multiple methods of payment, such as cash, WeChat and Alipay. The “design” subdimension reflects the requirements of taste in living. The

“waiting time” subdimension represents the requirements of convenience in a fast-paced society.

The method proposed in this paper provides hotel managers with a new way to understand the different impacts of the primary dimension, the subdimensions and the service attributes on service quality. For example, the results in Table 3 show that the physical environment quality is the most important dimension among the primary dimensions. In particular, the introduction of the interest degree of service attributes can help hotels improve their daily operations and solve the problems indicated by customers in online reviews in a timely manner. It is beneficial for hotels to allocate their limited resources to the important service attributes.

6. Conclusions

In this study, a research framework for service quality evaluation and service improvement using online reviews and the hierarchical service quality model is developed. The hierarchical service quality model is combined with online reviews and text mining technology to effectively obtain the scores of the subdimensions in the service quality model. LSTM is used for service quality evaluation, and LSTM-based sensitivity analysis is used to rank the subdimensions in the service quality model. Then, service improvement strategies are obtained by considering the score measures of subdimensions in the service quality model, the interest degree of the service attributes in each subdimension and the reasons for dissatisfaction mined from online reviews. The results of the online reviews of hotels show that LSTM obtained better prediction results than the RNN and ANN. Moreover, the rankings of the primary dimensions and the subdimensions in the service quality model and the service attributes extracted from the online reviews were reported, and a series of service improvement strategies for the specific service attributes were provided. From the results, emerging service attributes such as the Wi-Fi, food and methods of payment are the most important attributes to be improved, and the sub-dimensions of “design” and “waiting time” reflect the customer’s requirements and should also be improved; in contrast, the “tangibles” subdimension has a low improvement score.

There are several limitations to be further studied. First, sentiment lexicon-based coarse-grained sentiment analysis is used. In future work, a fine-grained sentiment analysis can be conducted. Second, in addition to online reviews, the ratings, pictures and videos over a longer time period can be considered to further improve the performance of the LSTM. Finally, for hotels with different numbers of stars, different rankings of the different dimensions, subdimensions and service attributes in the service quality model can be obtained to generate specialized service improvement strategies.

7. Compliance with ethical standards

Informed consent: Informed consent was obtained from all individual participants included in the study.

CRedit authorship contribution statement

Xin-Xin Liu: Data curation, Formal analysis, Investigation, Software, Writing – original draft. **Zhen-Yu Chen:** Conceptualization, Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.elerap.2022.101174>.

References

- Bi, J.W., Liu, Y., Fan, Z.P., Cambria, E., 2019. Modelling customer satisfaction from online reviews using ensemble neural network and effect-based Kano model. *Int. J. Prod. Res.* 57 (22), 7068–7088.
- Brady, M.K., Cronin Jr., J.J., 2001. Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. *J. Market.* 65 (3), 34–49.
- Carrasco, R.A., Sánchez-Fernández, J., Muñoz-Leiva, F., Francisca Blasco, M., Herrera-Viedma, E., 2017. Evaluation of the hotels e-services quality under the user’s experience. *Soft. Comput.* 21 (4), 995–1011.
- Chen, L., Fan, A., Shi, H., Chen, G., 2020. Region-action LSTM for mouse interaction sequence based search satisfaction evaluation. *Inf. Process. Manage.* 57 (6), 102349.
- Chung, J., Lee, J., Yoon, J., 2022. Understanding music streaming services via text mining of online customer reviews. *Electron. Commer. Res. Appl.* 53, 101145.
- Cronin Jr, J.J., Taylor, S.A., 1992. Measuring service quality: a reexamination and extension. *J. Market.* 56 (3), 55–68.
- Dabholkar, P.A., Thorpe, D.I., Rentz, J.O., 1996. A measure of service quality for retail stores: scale development and validation. *J. Acad. Mark. Sci.* 24 (1), 3–16.
- Deng, W.J., Chen, W.C., Pei, W., 2008. Back-propagation neural network based importance–performance analysis for determining critical service attributes. *Expert Syst. Appl.* 34 (2), 1115–1125.
- Deng, Y., Zheng, J., Khern-am-nuai, W., Kannan, K.N., 2021. More than the quantity: the value of editorial reviews for a UGC platform. *Manage. Sci.* <https://doi.org/10.1287/mnsc.2021.4238>.
- Farrington, T., Antony, J., O’Gorman, K.D., 2018. Continuous improvement methodologies and practices in hospitality and tourism. *Int. J. Contemp. Hospitality Manage.* 30 (1), 581–600.
- Fischer, T., Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. *Eur. J. Oper. Res.* 270 (2), 654–669.
- Gao, S., Tang, O., Wang, H., Yin, P., 2018. Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *Int. J. Hospitality Manage.* 71, 19–32.
- Geetha, M., Singha, P., Sinha, S., 2017. Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. *Tourism Manage.* 61, 43–54.
- Ghasemi, N., Momtazi, S., 2021. Neural text similarity of user reviews for improving collaborative filtering recommender systems. *Electron. Commer. Res. Appl.* 45, 101019.
- Grönroos, C., 1984. A service quality model and its marketing implications. *Eur. J. Mark.* 18 (4), 36–44.
- Guo, Y., Barnes, S.J., Jia, Q., 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Manage.* 59, 467–483.
- He, Q.-Q., Wu, C., Si, Y.-W., 2022. LSTM with particle swarm optimization for sales forecasting. *Electron. Commer. Res. Appl.* 51, 101118.
- Hong, W., Yu, Z., Wu, L., Pu, X., 2020. Influencing factors of the persuasiveness of online reviews considering persuasion methods. *Electron. Commer. Res. Appl.* 39, 100912.
- Hsu, S.W., Qing, F., Wang, C.C., Hsieh, H.L., 2018. Evaluation of service quality in facebook-based group-buying. *Electron. Commer. Res. Appl.* 28, 30–36.
- James, T.L., Calderon, E.D.V., Cook, D.F., 2017. Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Syst. Appl.* 71, 479–492.
- Kim, W.G., Lim, H., Brymer, R.A., 2015. The effectiveness of managing social media on hotel performance. *Int. J. Hospitality Manage.* 44, 165–171.
- Knutson, B., Stevens, P., Wullaert, C., Patton, M., Yokoyama, F., 1990. LODGSERV: A service quality index for the lodging industry. *Hospitality Res. J.* 14 (2), 277–284.
- Lee, W.H., Cheng, C.C., 2018. Less is more: A new insight for measuring service quality of green hotels. *Int. J. Hospitality Manage.* 68, 32–40.
- Lee, H., Kim, C., 2014. Benchmarking of service quality with data envelopment analysis. *Expert Syst. Appl.* 41 (8), 3761–3768.
- Leong, L.Y., Hew, T.S., Lee, V.H., Ooi, K.B., 2015. An SEM–artificial-neural-network analysis of the relationships between SERVPERF, customer satisfaction and loyalty among low-cost and full-service airline. *Expert Syst. Appl.* 42 (19), 6620–6634.
- Li, S., Hao, J., Chen, Z., 2010. Graph-based service quality evaluation through mining Web reviews. In *Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE-2010)* (pp. 1–8). IEEE.
- Li, H., Liu, Y., Tan, C.W., Hu, F., 2020. Comprehending customer satisfaction with hotels: data analysis of consumer-generated reviews. *Int. J. Contemporary Hospitality Manage.* 32 (5), 1713–1735.
- Liang, L., Cai, X., 2020. Forecasting peer-to-peer platform default rate with LSTM neural network. *Electron. Commer. Res. Appl.* 43, 100997.
- Mejia, J., Mankad, S., Gopal, A., 2021. Service quality using text mining: Measurement and consequences. *Manuf. Service Oper. Manage.* 23 (6), 1354–1372.
- Nourikhab, H., Akbari, M.K., 2016. Impact of service quality on user satisfaction: Modeling and estimating distribution of quality of experience using Bayesian data analysis. *Electron. Commer. Res. Appl.* 17, 112–122.
- Palese, B., Usai, A., 2018. The relative importance of service quality dimensions in E-commerce experiences. *Int. J. Inf. Manage.* 40, 132–140.

- Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1985. A conceptual model of service quality and its implications for future research. *J. Market.* 49 (4), 41–50.
- Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1988. SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *J. Retail.* 64 (1), 12–40.
- Qi, J., Zhang, Z., Jeon, S., Zhou, Y., 2016. Mining customer requirements from online reviews: A product improvement perspective. *Inf. Manage.* 53 (8), 951–963.
- Saha, G.C., Theingi, 2009. Service quality, satisfaction, and behavioural intentions: A study of low-cost airline carriers in Thailand. *J. Service Theory Practice* 19 (3), 350–370.
- Tseng, M.L., 2009. A causal and effect decision making model of service quality expectation using grey-fuzzy DEMATEL approach. *Expert Syst. Appl.* 36 (4), 7738–7748.
- Wei, X., Luo, X., Li, Q., Zhang, J., Xu, Z., 2015. Online comment-based hotel quality automatic assessment using improved fuzzy comprehensive evaluation and fuzzy cognitive map. *IEEE Trans. Fuzzy Syst.* 23 (1), 72–84.
- Xiang, Z., Schwartz, Z., Gerdes Jr, J.H., Uysal, M., 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? *Int. J. Hospitality Manage.* 44, 120–130.
- Xie, K.L., Zhang, Z., Zhang, Z., 2014. The business value of online consumer reviews and management response to hotel performance. *Int. J. Hospitality Manage.* 43, 1–12.
- Xu, Y., Armony, M., Ghose, A., 2021. The interplay between online reviews and physician demand: An empirical investigation. *Manage. Sci.* 67 (12), 7344–7361.
- Ye, Q., Law, R., Gu, B., Chen, W., 2011. 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), 634–639.
- Yeung, D.S., Cloete, I., Shi, D., Ng, W., 2010. *Sensitivity Analysis for Neural Networks*. Springer.
- Yucesan, M., Gul, M., 2020. Hospital service quality evaluation: an integrated model based on Pythagorean fuzzy AHP and fuzzy TOPSIS. *Soft. Comput.* 24 (5), 3237–3255.
- Zeinalizadeh, N., Shojaie, A.A., Shariatmadari, M., 2015. Modeling and analysis of bank customer satisfaction using neural networks approach. *Int. J. Bank Market.* 33 (6), 717–732.
- Zhang, J., Lu, X., Liu, D., 2021. Deriving customer preferences for hotels based on aspect-level sentiment analysis of online reviews. *Electron. Commer. Res. Appl.* 49, 101094.
- Zhao, W., Liu, Q.B., Guo, X., Wu, T., Kumar, S., 2022. Quid pro quo in online medical consultation? investigating the effects of small monetary gifts from patients. *Prod. Oper. Manage.* 31 (4), 1698–1718.
- Zhao, Y., Xu, X., Wang, M., 2019. Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *Int. J. Hospitality Manage.* 76, 111–121.
- Zuo, W., Zhu, W., Chen, S., He, X., 2019. Service quality management of online car-hailing based on PCN in the sharing economy. *Electron. Commer. Res. Appl.* 34, 100827.