Complex & Intelligent Systems (2023) 9:2571–2583
https://doi.org/10.1007/s40747-020-00223-7

ORIGINAL ARTICLE

Voice of urban park visitors: exploring destination attributes influencing behavioural intentions through online review mining

Ke Ma[1] · **Beibei Jiang**[1]

Received: 26 July 2020 / Accepted: 19 October 2020 / Published online: 7 November 2020

Abstract
In this paper, we will identify the destination attributes of a popular urban park and investigate their specific roles in forming visitors' behavioural intentions using text mining approaches. The principles of natural language processing and psychometric procedure were combined to achieve the objectives of the research. Initially, park visitors' online reviews were collected and analysed to identify possible latent dimensions for questionnaire design. Then, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were used for crucial factor selection and verification. Lastly, a structural equation model (SEM) was constructed to investigate the impacts of these park attributes on the behavioural intention of visitors.

Keywords Destination attributes · Behavioural intention · Natural language processing · Structural equation modelling

Introduction

As the guardian of the city's natural environment, urban parks are indispensable public resources for the sustainable development of urban ecosystems [1]. However, there is always a lack of adequate understanding of visitors' experiences by park managers or planners. Post-occupancy evaluation (POE), as a bridge between them, forms a feedback loop to connect the two parties. The user feedback can inform the functional comfort and environmental stress of a built environment [2], which has recently become the focus in POE research [3]. Nowadays, the massively user-generated content (UGC) on travel websites can be regarded as passionate, insightful and spontaneous review by visitors [4]. It can be taken as a rich source of online user feedback for helping understand user's real preferences or needs [5, 6].

Moreover, in empirical research, a pilot survey is crucial for scale design. It requires to identify the core dimensions of influential factors comprehensively and concisely [7]. However, traditional methods of pilot survey often rely solely on on-site interviews or observations. Due to the limited sample size, it is challenging to locate the overall visitors' concerns accurately, and some other essential dimensions may be ignored unnecessarily. Furthermore, most original items were adapted from prior literature of similar disciplines, which may lead to inconsistencies with current research scenarios [8]. Therefore, in the pilot survey, we suggest identifying the necessary attribute factors from the rich source of UGC.

Evaluating park usage and comprehending its driving elements are very significant for expanding park usage and subsequently human prosperity. Past examinations researched the impacts of various physic and culture variables on park use utilizing guest reviews and direct perceptions of park clients, which are typically site explicit and tedious. Many researches measured and analysed visitors' behavioural intentions for various kinds of parks utilizing uninhibitedly accessible geotagged registration information from web-based media. Some authors explored how park aspects, area, setting and transport influenced the visitors' behavioural intentions, utilizing different straight relapses. Regardless of likely inclinations in the utilization of web-based media information, utilizing a recreation center typology, visitors' behavioural intentions is fundamentally varied between various kinds of parks. Although social relics parks and enormous metropolitan parks had specific visitors' behavioural intentions, neighborhood parks had higher appearance rates per unit of region. Park size and extra charges were related with adequate visitors' behavioural intentions for a wide range of parks. For parks that mostly serve neighborhood inhabi-

✉ Beibei Jiang
jiangbeibei@tom.com

[1] College of Landscape Architecture, Sichuan Agricultural University, Chengdu 611130, Sichuan, China

tants, the separation from metropolitan focus fundamentally influenced park usage [5–8].

In this study, we attempted to explore the key attributes of an urban park that may influence the behavioural intentions of visitors and what is the relative impact weight of each park attribute on behavioural intention. Moreover, the means–end theory was applied in this research context. Natural language processing (NLP) and psychometric procedure were combinedly used in the research process.

Lots of prior studies have explored the role of a broad range of factors affecting park use behaviours, such as environmental quality, income, accessibility, demographic characteristics and individual preferences [7]. Nevertheless, few of those studies were based on established cognitive or behaviour theories. After that, Zhang and Tan researched determinant park-use behavioural factors based on the Theory of Planned Behavior (TPB) [9]. Han explored the relation between halal-friendly destination attributes and revisit intention based on the complexity theory [9]. Thus far, to the best of our knowledge, little research is known about the relationship between destination attributes and behavioural intention based on the means-end theory in park POE.

Destination attributes are recognized as an amalgam of different elements of the destination [10]. Visitors can perceive a variety of natural or artificial destination attributes [11], which aids in the forming of a visitor's on-site experience. In many prior studies, destination attributes are viewed as the primary and critical antecedent of behavioral intention [9, 12, 13].

Behavioral intention is assumed to be the likelihood of taking specific actions based on one's subjective tendencies [14]. Revisit and recommend intention are both essential components of behavioural intention [15]. In many studies, the behavioural intention was always the last factor in the perception chain. It is a crucial indicator for evaluating visitors' loyalty of a destination [9, 12, 13]. Favourable behavioural intentions usually represent the conative loyalty of visitors. It is critical for a destination's long-term viability and sustainability. Loyal visitors are more likely to return to the destination and give favourable word-of-mouth (WOM) to their relatives, friends or other potential visitors [16, 17]. In practice, actual loyal visitors' behaviours are often difficult to measure; thus, most studies employed behavioural intention as a compromise of it [18].

The means–end theory assumes that people are goal-oriented, and they need to achieve individual values by purchasing particular attributes of a product or service [19]. It provides a practical perspective to explore the impacts of different park attributes on a visitor's behaviour. For example, a visitor may be especially satisfied due to a unique set of features provided by the facilities or services of a park. Therefore, this study tries to extend the means–end theory in the context of park POE to explore the attributes that may significantly influence visitor's behaviours.

The rest of the paper is organized as follows: In the second section, we discuss the methodology behid the proposed method, as well as the entire research process. In Sect. "Numerical results", we present the numerical results and compare them to some well-known algorithms, with specific attention given to the demographic characteristics in Sect. "Demographic characteristics" and the exploratory factor analysis and the confirmatory factor analysis in Sects. "Exploratory factor analysis" and Confirmatory factor analysis", respectively. Structural equation modelling analysis is discussed in Sect. "Confirmatory factor analysis". Finally, after the algorithm discussion in Sect. "Discussion", the final section provides the conclusion and future research remarks.

Methodology

In this research, Wangjiang Pavilion Park was taken as a case study, which is next to the south bank of Jinjiang River and Sichuan University, Chengdu. Wangjiang Pavilion Park was built to commemorate the poet XueTao in Tang Dynasty, who cherished bamboo all her life and praised its personified noble quality. So far, more than 200 species of bamboo have been planted in the park, and the 39-m-high Wangjiang Pavilion is a landmark of that area [20]. It has now become a popular place for travelling and recreational purposes. According to official information, the daily visit has grown steadily, which has exceeded an average of 3000 times/day and reached over 7000 times/day on weekends.

The entire research process includes two significant steps:

1. Collect online reviews and use text analysis to identify the latent dimensions, specifically the following:
   - text pre-processing,
   - word frequency analysis,
   - word co-occurrence network analysis,
   - sentiment analysis and
   - latent Dirichlet allocation (LDA) topic analysis.
2. Conduct the questionnaire survey.
   - Perform exploratory factor analysis and confirmatory factor analysis.
   - Build a structural equation model to evaluate the impacts of park destination attributes on visitors' behaviour intentions.
   - Derive the overall research flow chart, as shown in Fig. 1.

Text analysis and latent dimension identification

Text analysis is a sort of natural language processing (NLP) technology. NLP has gone through an array of technologies such as naive Bayes, TF/IDF, word2vec, LDA, LSTM, fast-

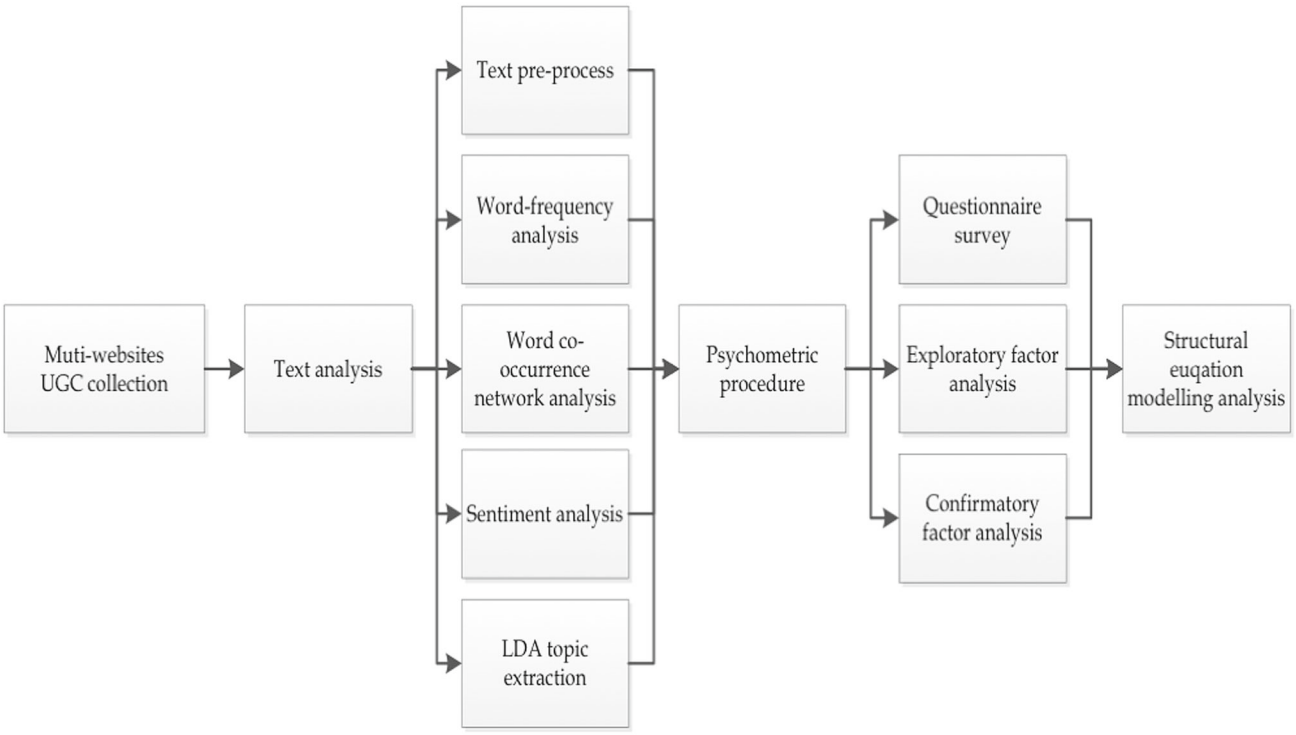


Fig. 1 Research flow chart

text, BERT and even the latest ALBERT [21, 22], which has vastly improved the text analysis quality and efficiency. The analysis process of recent researches on visitors' behaviours can be classified into two categories. The first category was based entirely on online review analysis, which extracted features only from the text for investigation [6, 23, 24]. However, such processes are:

Less pertinent than methods using questionnaires.

For unstructured text such as reviews or blogs, user ratings are unavailable.

Self-selection bias may be included in the analyses [5].

The second category relied solely on questionnaires [7, 25, 26], which could not make full use of big textual data to obtain the most representative features of the population; therefore, we proposed a two-step approach in this research. Step one is to use text analysis to identify the latent dimensions of online reviews in the pilot survey. Step two is to survey the target population and build a structural equation model (SEM) for empirical analysis. In this way, we can take full advantages of both methods.

In this study, we employed multiple text analysis methods, including word frequency statistics, word co-occurrence network, sentiment analysis and latent Dirichlet allocation (LDA). The target population are park visitors (including tourists and residents). We collected 2,435 reviews from nine Chinese travel websites or blogs by searching using the keyword "Wangjiang Pavilion Park". All the nine websites are ranked at the top positions of Alexa (https://www.alexa.com) or China Travel website (https://top.chinaz.com/hangye/index_jiaotonglvyou_lvyou.html). Multi-threaded web crawlers were used to collect text data between 2017.6 and 2019.6. The list of UGC collection source websites with appropriate rankings and reviews are presented in Table 1.

Table 1 UGC collection source websites

Website	Alexa ranking	China travel website ranking	Reviews
www.weibo.com	18	–	881
www.ctrip.com	705	1	562
www.tripadvisor.cn	223	19	33
ly.baidu.com	3	–	406
www.mafengwo.cn	2490	3	105
www.dianping.com	615	–	121
www.qunar.com	2895	4	303
www.ly.com	28,261	–	21
www.tuniu.com	72,345	5	3
Total			2435

We used Baidu lexical analysis of Chinese (LAC) for word segmentation. LAC is a lexical analysis model using a recurrent neural network (RNN). Its accuracy on the test set can reach 95.5% [27] and can complete tasks such as word

Table 2 Keyword occurrence

Order	Keyword	Occurrence	Order	Keyword	Occurrence
1	Wangjiang Pavilion Park	1361	20	Like	104
2	Chong Li Building	568	21	Pleasantly cool	101
3	XueTao	544	22	Culture	98
4	bamboo	495	23	Cosplay	98
5	Sichuan University	267	24	NineEye Bridge	94
6	Quiet	265	25	Worth	90
7	Entrance ticket	239	26	Fresh air	87
8	Bamboo forest	224	27	Aged people	83
9	Poet	201	28	Unfree ticket	76
10	Tang dynasty	177	29	XueTao Well	76
11	Commemorate	168	30	History	72
12	Environment	147	31	Scenery	67
13	Ancient buildings	142	32	Scenic spot	67
14	Jinjiang River	142	33	Cultural relics zone	64
15	Drink tea	136	34	Urban center	63
16	Bamboo species	136	35	Green bamboo	60
17	Mahjong	135	36	River flow	57
18	Leisure	134	37	Stroll	57
19	Ticket free	133	38	Zhuojin Building	52

cutting, part-of-speech (POS) labelling and word stemming. Jieba (a word segmentation package on Python 3.7) has much better performance than LAC, but its accuracy is lower than that of LAC. Thus we first used LAC to generate a lexicon and then used Jieba for sentence-by-sentence word segmentation. Among the extracted keywords, we performed manual inspections by two students and removed the stop and meaningless words [28]. We also merged synonyms to improve the word segmentation performance. Some frequently used keywords are presented in Table 2, along with their respective occurrence.

We used Python 3.7 to calculate the word co-occurrence matrix and record the number of keywords that appeared simultaneously in each review. Then, we used the netdraw module in the UCINET software to draw the word co-occurrence network graph and observe its clusterings and correlations [29]. The depiction of a network of keyword co-occurrence is presented in Fig. 2.

As a result, we identified five node groups, where semantically related nodes were labelled with the same colours:

1. Group one: nodes connected to the bamboo forest landscape (green);
2. Group two: nodes related to culture and heritage (purple);
3. Group three: nodes associated with people's activities in the park (yellow);
4. Group four: nodes related to the adjacent Sichuan University (orange);
5. Group five: nodes associated with the entrance ticket (red).

The most connected nodes were keywords such as "bamboo", "bamboo forest", "XueTao", "Poet" and "Tang Dynasty".

Since visitors will only give specific comments on the topics that they are most concerned,, the keyword distribution showed a "long tail" pattern [30]. Some low-frequency words may still contain valuable information. Therefore, we inspected them and found some essential keywords like "older people", "cosplay", "pleasantly cool" and "fresh air".

Also, we used the Bi-LSTM model of the Baidu Senta platform for sentiment analysis. Senta has been pre-trained on large corpora and can give sentiment scores ranging from 0 to 1 on specific texts [31]. After the sentiment analysis, we found that the keyword "ticket fee" was with strong emotion score and included "ticket charge" as an item in the questionnaire.

Finally, we used the LDA algorithm to extract topics from online reviews. The LDA is an unsupervised Bayesian algorithm [32], which automatically extracts text topics without manual labelling. Therefore, it is suitable for identifying latent dimensions from a large volume of unstructured texts [4, 33]. We performed this analysis using the LDA module from the Gensim package. However, the user needs to set the topic numbers in advance. Therefore, we used the coherence model to get the optimal topic numbers [34]. Through calculating the coherence values of topics from 2 to 100,

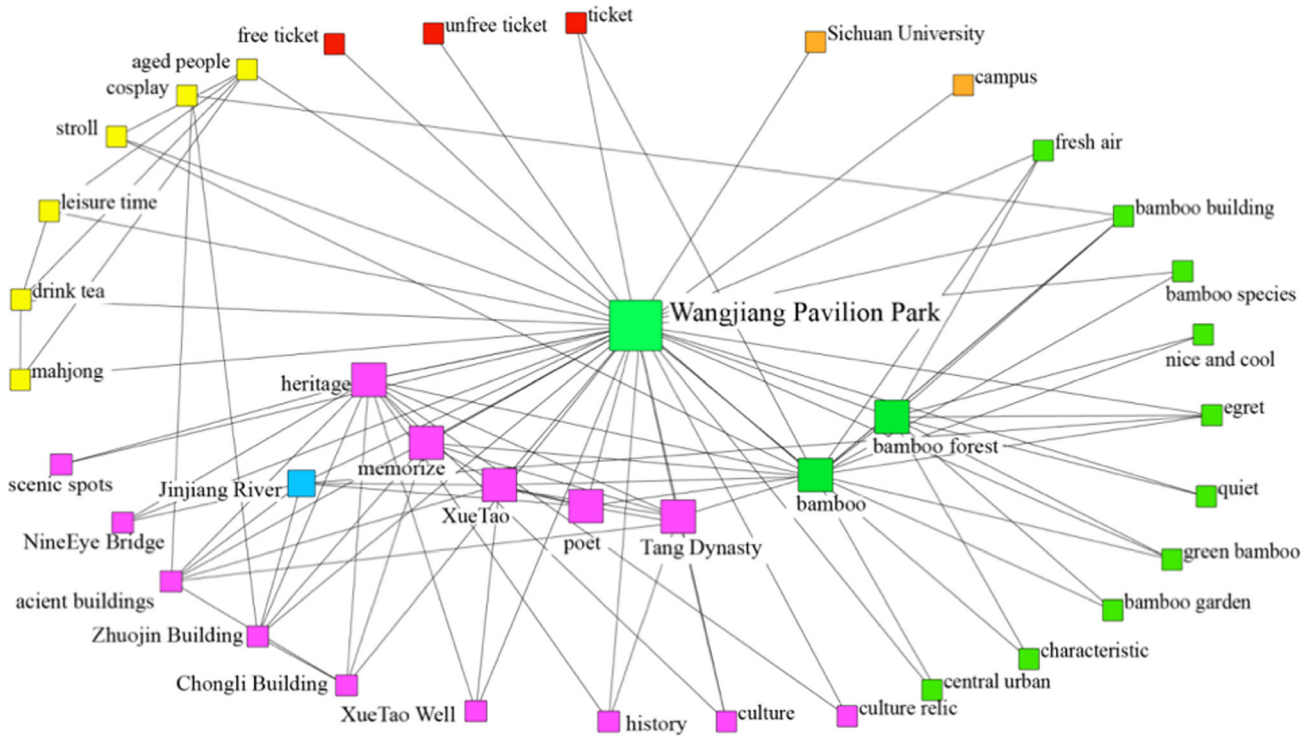


Fig. 2 Keyword co-occurrence network

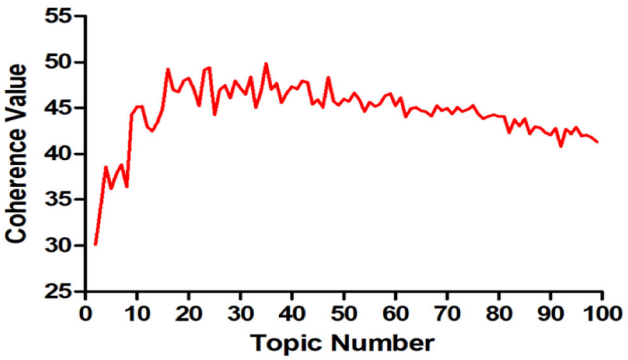


Fig. 3 The topic number and coherence value

the maximum coherence value was obtained when the topic number equalled 35, as can be seen in Fig. 3. The LDA topics are listed in Table 3.

Questionnaire design and survey

The questionnaire was built based on the pilot survey results. A 5-point Likert scale was employed for measuring the model items. The initial version of the questionnaire was written in English. Then, it was translated into Chinese using the back-translation method, which was then reviewed by a native Chinese speaker to ensure that the meaning was clearly delivered. There are three parts of the questionnaire:The first part briefly presented the survey guidance on the privacy and anonymity protection of the respondent. The second part contained survey questions for measuring the items in the model. Among them, the behavioural intention containing three items was adapted from the researches [9, 12]. The elements of the park destination attribute referenced the results of the pilot survey and used operational definitions. The final part is a demographic survey of participants, including age, gender and education.

This study adopted the method of convenience sampling. A total of 312 visitors were invited to the survey, and 306 of them completed the questionnaire. The investigator first gave a brief introduction to the research and then surveyed the respondent under his permission. The data collection process lasted for 3 weeks, from March to April 2019. Then, through screening, a total of 299 valid cases were retained. According to the rule of 10:1 observations per each indicator [35], the 299 samples met the relevant criteria.

Numerical results

Demographic characteristics

As shown in Table 4, among the samples, 52.8% were males (158 cases), and 47.2% were females (141 cases). People

Table 3 LDA topics

Topic	Keyword01	P01	Keyword02	P02	Keyword03	P03	Keyword04	P04	Keyword05	P05
1	Cannot forget	0.055	Egret	0.033	distinguished bearing	0.027	Special	0.027	Wind sound	0.027
2	Wangjiang Pavilion	0.078	Scenery	0.030	Environment	0.028	Free admission	0.027	conquer	0.026
3	Bamboo	0.055	bamboo forest	0.042	Worth	0.039	Form	0.037	Taste tea	0.030
4	Drink tea	0.048	Play mahjong	0.042	entrance ticket	0.039	Friend	0.037	Bosom friend	0.036
5	Close	0.039	Traffic jam	0.038	Neck	0.038	Master	0.037	Ordinary	0.037
6	Bamboo	0.088	Wangjiang Pavilion	0.028	come in and go out	0.027	Egret	0.026	Verdant	0.023
7	Bamboo	0.045	Drink tea	0.038	Wangjiang Pavilion	0.031	Gorgeous	0.030	Appearance	0.030
8	Retain	0.034	Exuberant	0.031	Scenery	0.030	Humanity	0.029	Take pictures	0.029
9	Culture	0.060	Wangjiang pavilion	0.033	Plant	0.030	Environment	0.026	Poet	0.025
10	Artistic conception	0.020	Leisure	0.018	Influence	0.016	Official	0.016	river flow	0.016
11	Literati	0.096	environment	0.042	Use	0.036	Play	0.036	Quiet and Secluded	0.033
12	Visitor	0.035	Sigh	0.031	Bumpy	0.031	The four seasons	0.030	Enclosing	0.026
13	Water lilies	0.032	River water	0.032	Culture	0.024	tradition	0.023	wide varieties	0.021
14	Worth	0.076	Fragmentary	0.037	Egret	0.030	Open area	0.023	Scenery	0.022
15	Statue	0.043	Wangjiang Pavilion	0.026	Beautiful	0.024	Bamboo forest	0.023	Report	0.022
16	Noble and unsullied	0.054	Environment	0.047	Association	0.041	Bamboo	0.035	Bend one's head	0.027
17	Wangjiang Pavilion	0.053	Bamboo	0.030	Bamboo forest	0.028	Characteristic	0.026	Mahjong	0.026
18	Egret	.121	Urban	0.035	Bamboo forest	0.024	Sentiment	0.022	Bamboo	0.020
19	The surface of the water	0.052	Bamboo	0.047	mosquito	0.031	Wangjiang Pavilion	0.029	dinner	0.026
20	Poet	0.033	Worth	0.032	Artistic conception	0.031	Architecture	0.026	Commemorate	0.024
21	Bamboo	0.038	Answer	0.037	magnificent	0.029	Stroll	0.025	Aesthetic taste	0.022
22	Poet	0.056	Commemorate	0.030	Intense aroma	0.029	Drink tea	0.026	Play mahjong	0.022
23	Fierce	0.038	Sincere	0.038	Life	0.021	Existing	0.021	Accomplish	0.009
24	Bamboo	0.031	bamboo forest	0.030	House	0.027	Human world	0.024	Master	0.023
25	Thought	0.081	Life	0.049	Yibin	0.028	Foster	0.028	Try to please	0.028
26	Poet	0.030	Wangjiang pavilion	0.025	natural and unrestrained	0.023	Commemorate	0.023	bamboo	0.021
27	Wangjiang Pavilion	0.037	Gifted scholar	0.029	Upper couplet	0.028	Stroll around	0.022	Body and mind	0.022
28	Egret	0.071	Bamboo forest	0.058	Sentiment	0.025	Leisure time	0.021	Worth	0.019
29	Leisure time	0.090	Vegetation	0.032	Comfort	0.023	Elegant	0.023	Restaurant	0.022
30	Straight and upright	0.040	Arrive	0.040	Painting	0.040	Literati	0.039	Writing poetry	0.021
31	Bamboo	0.067	bamboo forest	0.058	Temperament and interest	0.042	Poetic flavour	0.037	Charm	0.027
32	Wangjiang pavilion	0.066	Bamboo	0.044	Bamboo forest	0.041	Poet	0.029	Environment	0.024
33	Environment	0.058	Ancient buildings	0.040	Charge	0.038	Bamboo	0.025	Free admission	0.021
34	Literature	0.059	Quiet	0.036	Entrance ticket	0.035	Pleasant scenery	0.031	Worth seeing	0.030
35	Environment	0.033	Pleasant scenery	0.032	Motion	0.031	Taxi fare	0.031	Cordial	0.031

Table 4 Demographic information

	Count		Percentage
Gender	Male	158	52.8
	Female	141	47.2
Age	≤18	61	20.4
	18–25	119	39.8
	25–35	74	24.7
	35–45	28	9.4
	>45	17	5.7
Education	High school education or below	45	15.1
	Bachelor	170	56.9
	Master	62	20.7
	Doctor	22	7.4

aged between 18 and 25 accounted for 39.8% (119 cases), which was the most substantial proportion among respondents. The educational level of the participants was that 15.1% (45 cases) of them had a high school education or below, 56.9% (170 cases) of them had a Bachelor's degree, 20.70% (62 cases) of them had a Master's degree and 7.4% (22 cases) of them had a doctoral education.

Exploratory factor analysis

We used IBM SPSS 17.0 to perform exploratory factor analysis (EFA) on park destination attributes [36] and used principal component analysis (PCA) and rotated component matrix to determine their underlying factors and groups. The results indicated that the kaiser–meyer–olkin (KMO) value was 0.815>0.8, and Bartlett's sphericity test was significant (p <0.000), which validated the adequacy of EFA [37]. After removing items with cross-loading or low factor loading (<0.40) [38], 14 attribute items were retained (Table 5). The eigenvalues of these four factors were all greater than 1, accounting for approximately 78.601% of the total variance. These factor items were grouped and named according to the rotated component matrix. As shown in Table 5, factor on is denoted as "Ecological Environment" which consisted of four items, accounting for 18.021% of the total variance. Factor two is expressed as "Culture and Heritage" which consisted of four items, accounting for 17.167% of the total variance. Factor three is named as "Service Facilities" which consisted of three items, accounting for 14.847% of the total variance. Factor four is designed as "Bamboo Forest Landscape" which consisted of three items, accounting for 14.183% of the variance. The factor loading of 14 items all exceeded the threshold of 0.50 [39]. The Cronbach's is used to check the internal consistency of the items in the corresponding group. They are $\alpha_1 = 0.889$, $\alpha_2 = 0.872$, $\alpha_3 = 0.897$ and $\alpha_4 = 0.870$, which were all greater than 0.70, indicating that they met the relevant criteria [40].

Therefore, we proposed the following four hypotheses:

H1 Bamboo forest landscape attributes impacts on behavioural intention significantly and positively.

H2 Cultural and heritage attributes impacts on behavioural intention significantly and positively.

H3 Service facilities attributes impacts on behavioural intention significantly and positively.

H4 Ecological environment attributes impacts on behavioural intention significantly and positively.

Confirmatory factor analysis

We used IBM AMOS 22.0 for confirmatory factor analysis (CFA) to verify the reliability, discriminant and convergent validity of the measuring model [41]. Table 6 gave the results of the CFA. The standard factor loading of each item was greater than 0.70 and less than 0.95, which met relevant standards [42]. The goodness-of-fit statistics indicators of the measuring model was: $\chi^2 = 105.266$, $df = 71$, $\chi^2/df = 1.483$, $p = 0.005$, RMSEA = 0.04, CFI = 0.985, GFI = 0.952, NFI = 0.956. It indicated that the proposed measuring model fitted with the data appropriately [43].

Moreover, the composite reliability (CR) values reached over the required cutoff rate of 0.70, which indicated that these items were internally consistent and reliable [39]. The average variance extract (AVE) values all exceeded the 0.50 threshold, which met the relevant criteria [39]. As shown in Table 7, each correlation coefficient value was found below the AVE's square root [39]. The measurement model is given in Fig. 4.

Structural equation modelling analysis

To verify the proposed hypotheses, we conducted SEM analysis using IBM AMOS 22.0. The goodness-of-fit statistics indicators were: $\chi^2 = 145.485$, $df = 109$, $\chi^2/df = 1.335$, RMSEA = 0.033, SRMR = 0.0387, $p = 0.011$, CFI = 0.988, IFI = 0.988, TLI = 0.985, AGFI = 0.926, indicating that the structural model fitted well with the data.

As shown in Table 8 and Fig. 5, the bamboo forest landscape attributes are positively and significantly correlated with behavioural intention (H1: $\beta_1 = 0.261$, $t = 4.114$, p <0.001); the cultural and heritage attributes is positively and significantly correlated with behavioural intention (H2: $\beta_2 = 0.252$, $t = 4.165$, p <0.001); the service facilities' attributes are positively and significantly correlated with behavioural intention (H3: $\beta_3 = 0.157$, $t = 2.798$, $p = 0.005$); the ecological environment attributes is positively and significantly correlated with behavioural intention (H3: $\beta_4 = 0.308$, $t = 5.007$, p <0.001).

Table 5 Exploratory factor analysis results

Factors	Loadings	Eigen-values	Explained variance (%)	Cronbach's α
Factor1: Ecological Environment		4.646	18.021	.889
ECO1: Bird Habitat	.890			
ECO2: Acoustic environment	.866			
ECO3: Micro-climate	.830			
ECO4: Air quality	.820			
Factor2 Culture and Heritage		3.014	17.167	.872
CUL1: Park history	.881			
CUL2: Leisure culture	.836			
CUL3: Cultural figures	.830			
CUL4: Ancient Architectures	.809			
Factor3 Service facilities		2.623	14.847	.897
FAC1: Activity platform for older people	.917			
FAC2: Activity platform for the youth	.912			
FAC3: Ticket charge	.894			
Factor4 Bamboo forest landscape		1.373	14.183	.870
BAM1:Bamboo variety	.870			
BAM2: Bamboo building	.868			
BAM3: Bamboo garden	.853			
Sig = 0.000; KMO and Bartlett's test = 0.815,		Total: 78.601		

The outcomes of the SEM analysis indicated that the park attributes impacted on behavioural intention significantly. Among them, the most significant proportion is the ecological environment attributes ($\beta_4 = 0.308$). Its sub-items include micro-climate environment, acoustic environment, air quality and bird habitat. These are related to the rich ecological functions of the bamboo forest [44]. The second is the bamboo forest landscape attributes ($\beta_1 = 0.261$). Its sub-items are bamboo varieties, bamboo garden bamboo architecture. The third is the cultural and heritage attributes ($\beta_2 = 0.252$). Its sub-items include park history, cultural figures, ancient architectures and leisure culture. Among them, the cultural figure mainly refers to the poet XueTao. The fourth is the service facilities attributes ($\beta_3 = 0.157$). Its sub-items are ticket fees, activity platform for older people and activity platform for young people. According to the on-site observation, young people mainly like to engage in cosplay activities in the park, and middle-aged or older people prefer to do various exercise activities in the park such as strolling, Taichi, Yoga, dancing, Yoyo. The R^2 of behavioural intention is 0.32. To sum up, the proposed structural equation model has a good ability to explain the correlations among the variables.

Discussion

By incorporating NLP technologies into the pilot survey, we extracted the latent dimensions from online reviews. The outcomes of the SEM analysis confirmed the validity of the text analysis methods. Sentiment analysis was found quite useful for locating key concerns of visitors. Although some keywords were distributed in the low-frequency zone, their emotions were very intense and could be recognized by sentiment analysis. Although manual analysis of online reviews may be more accurate than using machine learning methods [5], it is almost impossible to accomplish the work when faced with the massive amount of online user-generated content. Therefore, this study used a pre-trained Baidu Senta Bi-LSTM model to perform this task [31].

Furthermore, many studies relied on users' online ratings as a primary quantitative indicator [5, 24, 45]. Nonetheless, with the soaring number of unstructured online texts, their ratings are no longer available. Hence, unsupervised topic models have been increasingly adopted [4, 6]. However, the analysis granularity of LDA was still relatively coarse in this study, and it was impossible to identify more subtle items. Therefore, we used both word frequency analysis and word co-occurrence networks [29] to extract latent dimensions and found that it was effective.

Consequently, it is necessary to incorporate multiple text analysis methods into one research. Also, we performed the

Table 6 Confirmatory factor analysis results

Variable	Scale item		Unstd	S.E	*T* value	*p*	Std	SMC	C.R	AVE
Bamboo forest landscape									.873	.696
	BAM1:	I think the bamboo variety in the park is rich	1.000				0.875	0.766		
	BAM2:	I like the bamboo forest garden in the park	1.037	0.067	15.464	***	0.811	0.658		
	BAM3:	I like the design and decoration of the bamboo buildings in the park	0.971	0.063	15.514	***	0.814	0.663		
Culture and heritage									0.873	0.632
	CUL1:	I think the park is rich in history and culture	1.000				0.872	0.760		
	CUL2:	I like the cultural character poet XueTao of the park	0.802	0.055	14.696	***	0.760	0.578		
	CUL3:	I like the ancient architectures in the park	0.798	0.054	14.746	***	0.762	0.581		
	CUL4:	I like the leisure culture in the park	0.797	0.052	15.222	***	0.781	0.610		
Service facilities									0.898	0.746
	FAC1:	I think the ticket charge of the park is reasonable	1.000				0.840	0.706		
	FAC2:	I think the park provides a rich platform for the older people	1.151	0.065	17.674	***	0.869	0.755		
	FAC3:	I think young people like to do various activities in the park	1.159	0.065	17.873	***	0.881	0.776		
Ecology environment									0.891	0.671
	ECO1:	I think the climate in the park is comfortable	1.000				0.783	0.613		
	ECO2:	I think the sound environment in the park is comfortable	1.082	0.071	15.340	***	0.838	0.702		
	ECO3:	I think the air in the park is fresh	0.870	0.062	13.999	***	0.775	0.601		
	ECO4:	I think the park environment is suitable for bird inhabitation	1.106	0.069	16.039	***	0.877	0.769		
Behavioural intention									0.898	0.747
	INT1:	I ams willing to recommend the park to my relatives or friends	1.000				0.880	0.774		
	INT2:	I am willing to visit this park again	1.026	0.053	19.313	***	0.902	0.814		
	INT3:	I will share the reviews or photos about the park to my blog or circle of friends	0.932	0.054	17.156	***	0.808	0.653		

Table 7 Discriminant and convergent validity

	AVE	Ecological environment	Service facilities	Culture and heritage	Behavioural intention	Bamboo forest landscape
Ecological environment	0.671	0.819				
Service facilities	0.746	− 0.066	0.864			
Culture and heritage	0.632	− 0.029	− 0.043	0.795		
Behavioural intention	0.747	0.366	0.129	0.303	0.864	
Bamboo forest landscape	0.696	0.289	0.014	0.252	0.416	0.834

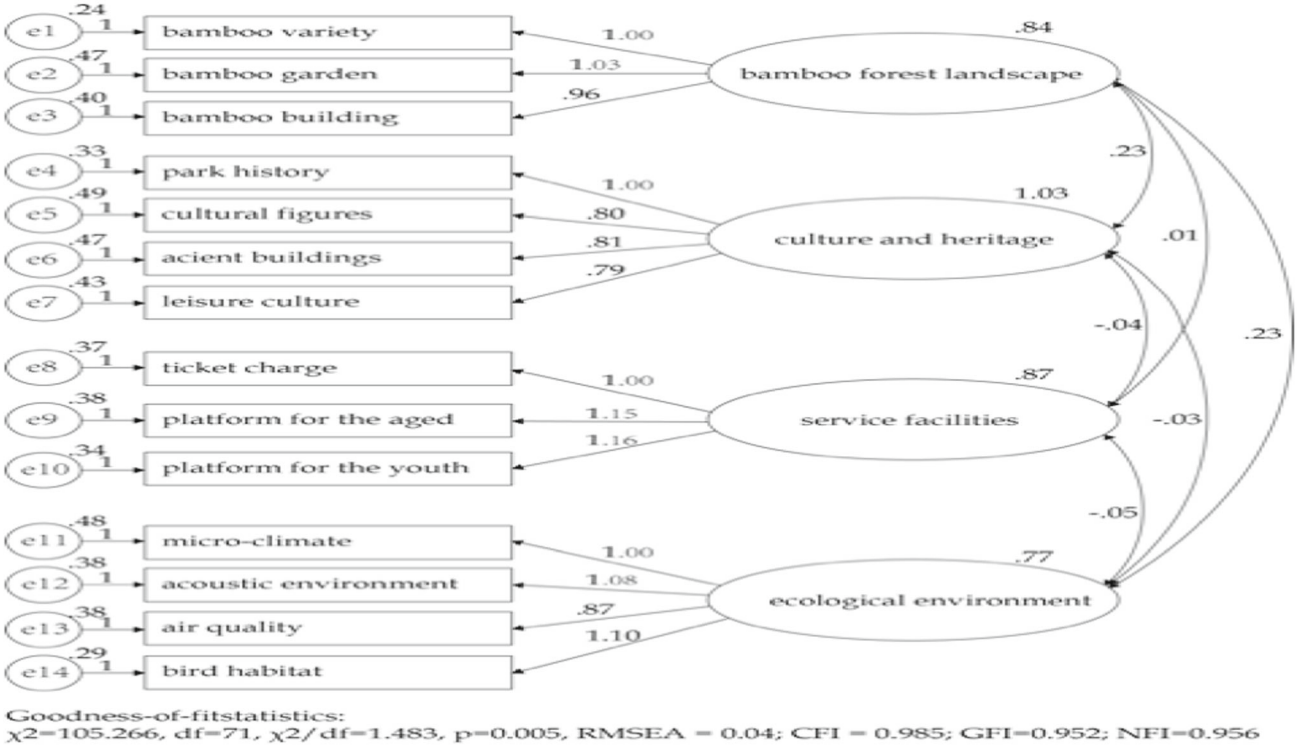


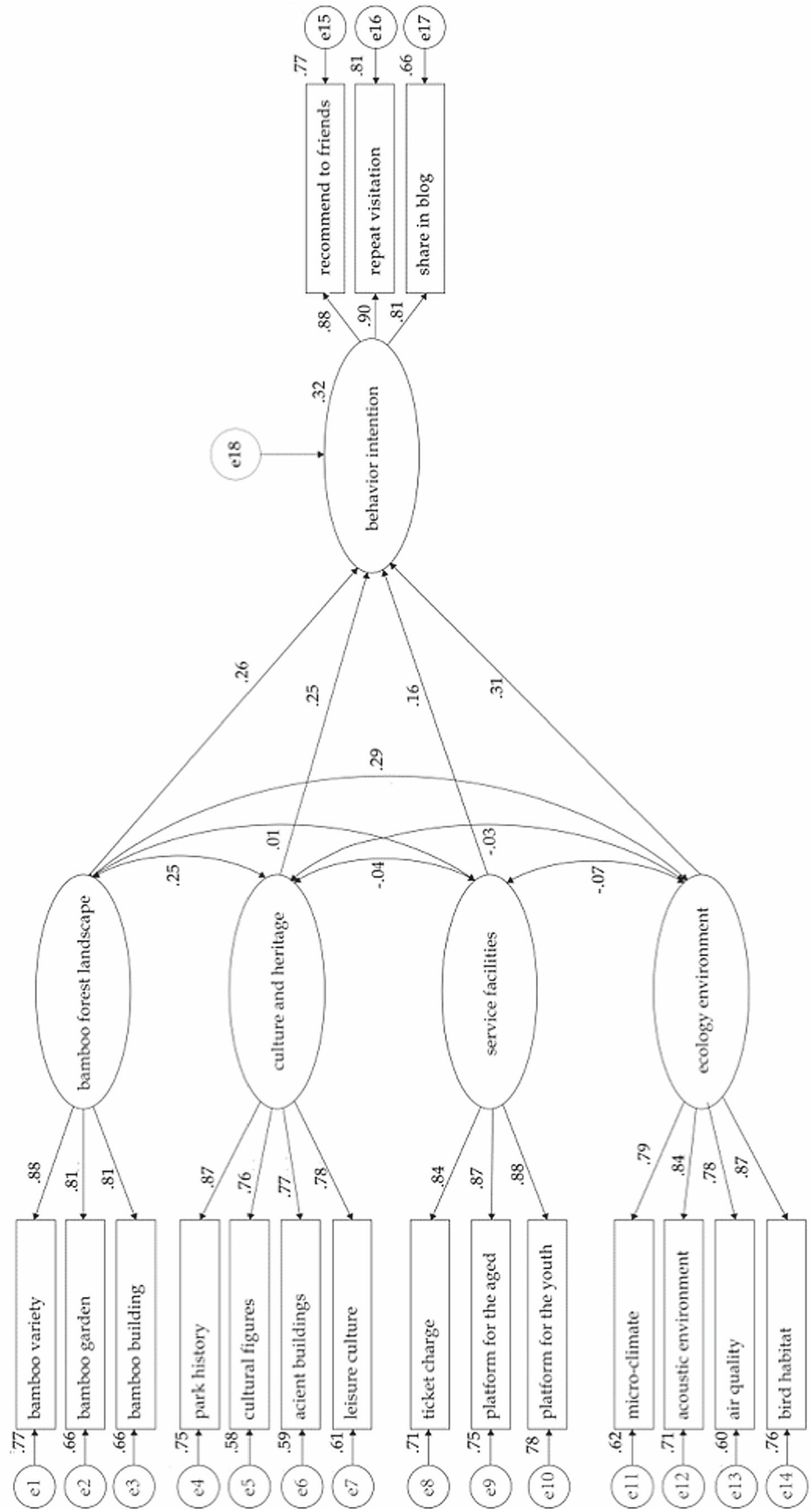
Fig. 4 Measurement model

Table 8 Hypotheses test results ($N = 299$)

			Estimate	Std.Estimate	S.E	*T* value	*p*
Behavioural intention	←	Bamboo forest landscape	0.267	0.261	0.065	4.114	***
Behavioural intention	←	Culture and heritage	0.277	0.252	0.066	4.165	***
Behavioural intention	←	Service facilities	0.13	0.157	0.046	2.798	0.005
Behavioural intention	←	Ecological environment	0.361	0.308	0.072	5.007	***

manual inspection on keywords and merged some synonyms to improve the expression of the word frequency distribution. Therefore, to increase efficiency, the introduction of the synonym merging mechanism may be worthy of attention in future studies.

The trends in POE research indicated that end-user feedback was critical for improving the quality of the built



Goodness-of-fit statistics:
$\chi 2=145.485$, df=109, $\chi 2/df=1.335$, RMSEA=0.033, SRMR=0.0387, p=0.011, CFI=0.988, IFI=0.988, TLI=0.985, AGFI=0.926

Fig. 5 The diagram of the SEM analysis

environment [3]. Tveit argued that people's perception of the landscape was at the heart of the European Landscape Convention [46]. The outcomes of this study indicated a significant association between the inner behavioural intentions of visitors and the outer park attributes. From the visitors' overall perception, we can trace the attributes that may significantly affect their behaviours. Meanwhile, the relative impact weight of each destination attribute on behavioural intention was revealed through the SEM analysis. Moreover, It is also possible for eliciting further exploration of more specific attributes in details.

Furthermore, the market competition in scenic spots is becoming increasingly fierce. It is imperative to enhance the differentiated competitiveness of the destination. The common sustainable indicators of urban parks have been thoroughly investigated [47], but research on how to explore the unique attributes of a typical park destination is relatively limited. The two-step analysis approach proposed in this research can help to explore the key attributes of a successful case and to provide references for future designs.

Conclusion

This study attempted to explore the significant park attributes influencing visitors' behavioural intentions in the park POE. Wangjiang Pavilion Park was taken as a case study. We combined both the natural language processing (NLP) technology and psychometric test procedure into the research process, which allowed us to listen to the voice of visitors more effectively. Text analysis was initially conducted on online reviews to identify the latent park attributes. The extracted attributes were further examined and grouped by exploratory factor analysis (EFA). Then, the reliability, discriminant and convergent validity of these factors were verified by confirmatory factor analysis (CFA). Last, a structural equation model was constructed to include the selected variables and calculate the relations and impacts among them. The overall outcomes revealed four main dimensions: bamboo forest landscape, ecological environment, culture and heritage and service facilities. The SEM analysis indicated that the impacts of these four dimensions on behavioural intention were positive and significant, which is favourable for increasing visitors' repeat visitation and recommendation behaviours.

In the future research, researchers can deal with incorporating multiple text analysis methods into one research. Since we performed the manual inspection on keywords and merged some synonyms to improve the expression of the word frequency distribution, and the synonym merging mechanism can be explored in order to increase efficiency.

Also, in future designs, the key attributes of a successful case can be explored using the two-step analysis approach proposed in this research.

Acknowledgements This research is financed by the National Natural Science Foundation of China (Grant Number 31800601).

Author contributions BJ designed and funded the study. KM wrote the manuscript, collected and analyzed the data. BJ revised the final version of manuscript.

Funding This Research is funded by the National Natural Science Foundation of China (Grant Number 31800601).

Availability of data and materialö All the data and materials are mentioned in the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest in this research.

Code availability Software application and custom code are available from corresponding author upon reasonable request.

References

1. Liu H, Hu Y, Li F, Yuan L (2018) Associations of multiple ecosystem services and disservices of urban park ecological infrastructure and the linkages with socioeconomic factors. J Clean Prod 174:868–879
2. Vischer JC (2008) Towards a user-centred theory of the built environment. Build Res Inf 36(3):231–240
3. Li P, Froese TM, Brager G (2018) Post-occupancy evaluation: State-of-the-art analysis and state-of-the-practice review. Build Environ 133:187–202
4. Guo Y, Barnes SJ, Jia Q (2017) Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. Tour Manag 59:467–483
5. Gerdt SO, Wagner E, Schewe G (2019) The relationship between sustainability and customer satisfaction in hospitality: an explorative investigation using eWOM as a data source. Tour Manag 74:155–172
6. Hu N, Zhang T, Gao B, Bose I (2019) What do hotel customers complain about? Text analysis using structural topic model. Tour Manag 72:417–426
7. Zhang J, Tan PY (2019) Demand for parks and perceived accessibility as key determinants of urban park use behaviour. Urban For. Urban Green 126420

8. Danaher PJ, Haddrell V (1996) A comparison of question scales used for measuring customer satisfaction. Int J Serv Ind Manag 7(4):4–26
9. Han H, Al-Ansi A, Olya HGT, Kim W (2019) Exploring halal-friendly destination attributes in South Korea: perceptions and behaviors of Muslim travelers toward a non-Muslim destination. Tour Manag 71:151–164
10. Lew AA (1987) A framework of tourist attraction research. Ann Tour Res 14(4):553–575
11. Yangzhou Hu, Ritchie JRB (1993) Measuring destination attractiveness: a contextual approach. J Travel Res 32(2):25–34
12. Moon H, Han H (2018) Destination attributes influencing Chinese travelers' perceptions of experience quality and intentions for island tourism: a case of Jeju Island. Tour Manag Perspect 28:71–82
13. Chen CF, Chen FS (2010) Experience quality, perceived value, satisfaction and behavioral intentions for heritage tourists. Tour Manag 31(1):29–35
14. Fishbein M, Ajzen I (1975) Belief, attitude. Intent Behav Introd Theory Res
15. Oliver Richard L (1997) Satisfaction: a behavioral perspective on the consumer. Irwin-McGraw-Hill, New York
16. Meleddu M, Paci R, Pulina M (2015) Repeated behaviour and destination loyalty. Tour Manag 50:159–171
17. Hosany S, Prayag G, Van Der Veen R, Huang S, Deesilatham S (2017) Mediating effects of place attachment and satisfaction on the relationship between Tourists' emotions and intention to recommend. J Travel Res 56(8):1079–1093
18. Yang Z, Peterson RT (2004) Customer perceived value, satisfaction, and loyalty: the role of switching costs. Psychol Mark 21(10):799–822
19. Zeithaml VA (1988) Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. J Mark 52(3):2–22
20. "Introduction of Wangjiang Pavilion Park," 2019. https://www.wangjianglou.com/about.html.
21. Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations," pp. 1–16, 2019.
22. Li Q, Li S, Zhang S, Hu J, Hu J (2019) A review of text corpus-based tourism big data mining. Appl Sci 9(16):3300
23. Park E (2019) Motivations for customer revisit behavior in online review comments: analyzing the role of user experience using big data approaches. J Retail Consum Serv 51:14–18
24. Li X, Wu C, Mai F (2019) The effect of online reviews on product sales: A joint sentiment-topic analysis, vol 56, no 2. Elsevier B.V, Amsterdam
25. López-Mosquera N, Sánchez M (2014) Cognitive and affective determinants of satisfaction, willingness to pay, and loyalty in suburban parks. Urban For Urban Green 13(2):375–384
26. Barnes SJ, Mattsson J, Sørensen F (2016) Remembered experiences and revisit intentions: a longitudinal study of safari park visitors. Tour Manag 57:286–294
27. Jiao Z, Sun S, Sun K (2018) Chinese lexical analysis with deep Bi-GRU-CRF network, pp 1–10
28. Anandarajan M, Hill C, Nolan T (2019) Text preprocessing, pp 45–59
29. van Duijn MHM (2005) Software for social network analysis. Model methods Soc Netw Anal 28
30. Anderson C (2006) The long tail: why the future of business is selling less of more. Hachette Books
31. Chen Y, Wang T, Zhu S, Lian P (2020) Will you miss me if I am leaving? unexpected market withdrawal of Norwegian Joy and customer satisfaction. Tour Manag 76:103951
32. Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
33. Taecharungroj V, Mathayomchan B (2019) Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand. Tour Manag 75(June):550–568
34. Altaweel M, Bone C, Abrams J (2019) Documents as data: a content analysis and topic modeling approach for analyzing responses to ecological disturbances. Ecol Inform 51:82–95
35. Chin WW (1998) The partial least squares approach to structural equation modeling. Mod methods Bus Res 295(2):295–336
36. Churchill GA Jr (1979) A paradigm for developing better measures of marketing constructs. J Mark Res 16(1):64–73
37. Darren G, Mallery P (1999) SPSS for windows step by step: a simple guide and reference. In: Spss B (eds) Wind. Step by step a simple guid. Ref. Needham Height. MA Allyn Bacon
38. Li M, Cai LA (2012) The effects of personal values on travel motivation and behavioral intention. J Travel Res 51(4):473–487
39. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2006) Multivariate data analysis (Vol. 6). Pearson Prentice Hall, Upper Saddle River
40. Nunally JC, Bernstein IH (1978) Psychometric theory. McGraw-Hill, New York
41. Anderson JC, Gerbing DW (1988) Structural equation modeling in practice: a review and recommended two-step approach. Psychol Bull 103(3):411
42. Jöreskog KG (1999) How large can a standardized coefficient be
43. Ford JK, MacCallum RC, Tait M (1986) The application of exploratory factor analysis in applied psychology: a critical review and analysis. Pers Psychol 39(2):291–314
44. Ben-zhi Z, Mao-yi F, Jin-zhong X, Xiao-sheng Y, Zheng-cai L (2005) Ecological functions of bamboo forest: research and application. J For Res 16(2):143–147
45. Xie H, Miao L, Kuo P-J, Lee B-Y (2011) Consumers' responses to ambivalent online hotel reviews: the role of perceived source credibility and pre-decisional disposition. Int J Hosp Manag 30(1):178–183
46. Tveit MS, Sang AO, Hagerhall CM (2012) Scenic beauty: Visual landscape assessment and human landscape perception. Environ Psychol An Introd, 37–46
47. Chan CS, Si FH, Marafa LM (2018) Indicator development for sustainable urban park management in Hong Kong. Urban For Urban Green 31:1–14