Research on Sentiment Analysis of Tourism Review in The West Lake Scenic Area

Shuyi Jin

ABSTRACT:
With the advent of the digital tourism era, the convergence of information technology and tourism resources has become pervasive, with tourists relying on Internet-based tools as an essential part of their travel experience. In this study, we select the tourism reviews of The West Lake Natural Scenic Area as our research sample and conduct an in-depth analysis of its current situation through the sentiment analysis. A total of 3,376 data points related to reviews on Hangzhou's West Lake Scenic Area was extractedtrip website using the "Octopus" web crawler software. After preprocessing, this study employed word frequency analysis, generation of word clouds, and other data visualization charts to analyze ROST EA and micro. Additionally, the LDA thematic clustering algorithm and SnowNLP sentiment analysis were utilized. The findings revealed that overall, tourists predominantly expressed positive emotions towards The West Lake Scenic Area compared to negative emotions. Positive sentiments primarily stemmed from factors such as the natural beauty of The West Lake, management practices, service quality, and charges. Furthermore, different tourists exhibited varying emotional tendencies.
KEYWORDS: Sentiment Analysis; LDA; SnowNLP sentiment analysis; Tourism Review; The West Lake scenic area

1. Introduction

The rapid advancement of Internet technology has significantly influenced people's lifestyles and consumption patterns. With the advent of the Internet and smart tourism, coupled with the surge in online tourism users, there has been a notable increase in tourism reviews as a means of expressing one's travel experiences. For managers, these vast data resources enable them to gain a comprehensive understanding of visitors' overall travel experiences at specific attractions or even entire cities. For travelers, this facilitates obtaining real-time insights into the true situation of a city or tourist destination, aiding in their trip planning endeavors. Consequently, effectively extracting useful and essential information from these reviews quickly poses as quite challenging. In response to the issues, researchers have shown keen interest in text sentiment analysis, also referred to as opinion mining.

The marketing model of The West Lake in recent years has also been remarkably distinctive, achieving certain accomplishments. However, with the development of regional tourism and intelligent tourism, the encountered challenges in The West Lake's development have become increasingly prominent, including industrial structure, resource allocation, publicity and marketing strategies. Therefore, this paper aims at analyze tourists' perceptions of the image of The West Lake tourist landscape to identify current shortcomings and provide managers with enhanced marketing ideas and optimization proposals.

2. Literature Review

The recognition of the significance of emotions in relation to tourism experience and marketing has gradually increased at the end of the 20th century. Foreign scholars have primarily focused on four aspects regarding tourists' emotions; Investigating the correlation between emotional analysis and outcome variables such as customer satisfaction and loyalty [1]; Exploring specific their antecedents or outcomes[2]; Examining factors influencing tourists' evaluation of emotions[3]; Employing affective measurement techniques[4]

The research on text sentiment analysis by domestic scholars primarily focuses on subjectivity and objectivity, sentiment tendency, and sentiment intensity. Mao Chaoqun developed an online tourism text sentiment classification model based on sentiment analysis theory and conducted sentiment analysis on tourists' online comments[5].Wang Wei-Ch utilized analytical categories and ROST CM6.0 software for analyzing perceived image and emotionalimage of touristsinMingyue Mountain tourismarea[6].Additionally, some scholars have focused onnegativeemotion perception tourists as astarting point to explorereasonsfor such emotions as well asimprovementmeasures. Huang Shengnan

analyzed comprehensive perceptions of tourists regarding Huangshan Scenic Area, subsequently providing suggestions for improvements based on identified issues reflected by negative tourist perceptions [7]. Furthermore, scholars have conducted research on the overall tourism image of scenic areas to establish a positioning strategy for reference. Gao Fen employed network text analysis to examine the perceived image of Laoshan Scenic Area various perspectives and compared it with the official destination's projected image[8].

3. Research Methodology

This study aims at conduct a comprehensive and systematic analysis of online reviews related to tourist attractions. The research will utilize literature review, semantic network analysis, SnowNLP sentiment analysis, and LDA topic model analysis methods in order to obtain valuable content and results. The semantic network approach should be employed for preliminary organization, analysis, and evaluation of acquired data. In order to delve deeper into tourists' concerns and emotional tendencies, we employ ROST EA to analyze the sentiments expressed in all textual data to obtain an overall assessment. Additionally, constructing the LDA topic model and utilizing the SnowNLP algorithm for emotion analysis can help derive sentiment analysis findings in this study, thereby proposing suggestions for enhancement

3.1 LDA topic model

The LDA topic model is derived from the pLSA (Probabilistic Latent Semantic Analysis) model. Comparison to the pLSA model, the LDA topic model extracts both topic distribution and word probability distribution by initially extracting them from Dirichlet prior distributions. The model is highly scalable for large-scale document collections, as illustrated in the figure. Through employing the LDA model, it becomes possible to discover topic distributions from unstructured online comment texts, where each comment possesses its own unique
distribution of topics.

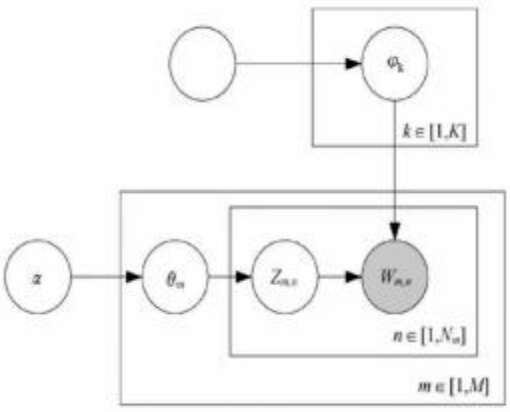

Figure Schematic diagram of the LDA model

α Prior distribution of topics (Dirichlet distribution)

β Prior discussions of words (Dirichlet distribution)

M The number of documents in the sample set

K The number of topics implied in the sample set

$Z_{m,n}$ The NTH topic in the article m document

$W_{m,n}$ The NTH word in the article m document

N_m The length of the document in article m, the number of words

θ_m The topic probability distribution of the Article m document

φ_k Term probability distribution in topic k

This paper employs the LDA topic model to extract topics from preprocessed travel reviews through four distinct steps:1. Identify the number of themes present. 2. Extract and analyze the topics at hand. 3. Allocate appropriate names to each identified topic. 4. Compute and record the comment count for every individual topic.

3.2 SnowNLP sentiment analysis

SnowNLP sentiment analysis is a Python library specifically designed for processing Chinese text content, addressing the lack of natural language processing libraries focused on Chinese. Inspired by TextBlob, SnowNLP sentiment analysis offers a convenient solution without relying on NLTK. SnowNLP, also known as Opinion Mining, refers to the use of computer technology to mine and analyze the subjectivity, viewpoint, emotion, and polarity of a given text in order to determine its emotional inclination. It utilizes the Simple Bayes Theorem for sentiment classification. When calculating the positive or negative sentiment of each utterance, the prior probability is multiplied by the conditional probability associated with each attribute feature word respectively to obtain the probability value of sentiment. The polarity with higher magnitude indicates overall sentiment expressed in an utterance.

4. Data analysis

4.1 Data source and collection

To thoroughly examine the West Lake Scenic Area, this

research heavily relies on user-generated travel reviews from various platforms in light of the rapid and efficient growth of online tourism. The selection criteria considered factors such as data precision, website reliability, and ease of access. Specifically, we analyzed publicly available reviews retrieved from two reputable platforms, Ctrip (https://hotels.ctrip.com/) and MaBeeWoo Travel (https://www.mafengwo.cn/), spanning from January to August 2023.

The Octopus data collection system is built on an internally developed distributed cloud computing platform, which serves as its fundamental infrastructure. By eliminating the need for manual searching and gathering of data, it significantly reduces costs associated with retrieving information while simultaneously improving overall operational efficiency[9]. The Octopus web crawler tool was utilized to retrieve online reviews and travelogues from two prominent domestic travel websites, Ctrip and MaHoneycomb. The keyword "West Lake Scenic Area" was used for the search. Considering data timeliness and limitations of Ctrip.com, online data from January 23 to August 23 were crawled, resulting in a total of 3376 pieces of data. Additionally, 50 travelogues were extracted from visible web pages on MaHuWo.com. Useful information (travelogue contents) was parsed and stored in a local Excel file named 'travelogue table', as illustrated in the sample chart below.

Through statistical collation of online comments, it was found that the scoring system used by online tourism sites are generally based on a five-point scale. Scores above 3 indicate good reviews while scores below or equal to 3 indicate bad reviews. The number of positive comments amounted to 29,768 (98.79%), whereas there were only 361 negative reviews (1.21%). Two Excel files named 'good review table' and 'bad review table' were created accordingly. The former contains fields such as rating time, commenters' names/comments/evaluation scores etc., while the latter only stores the content itself. Examples can be seen in Figures b and c.

Figure a. travelogue table

Figure b.good review table

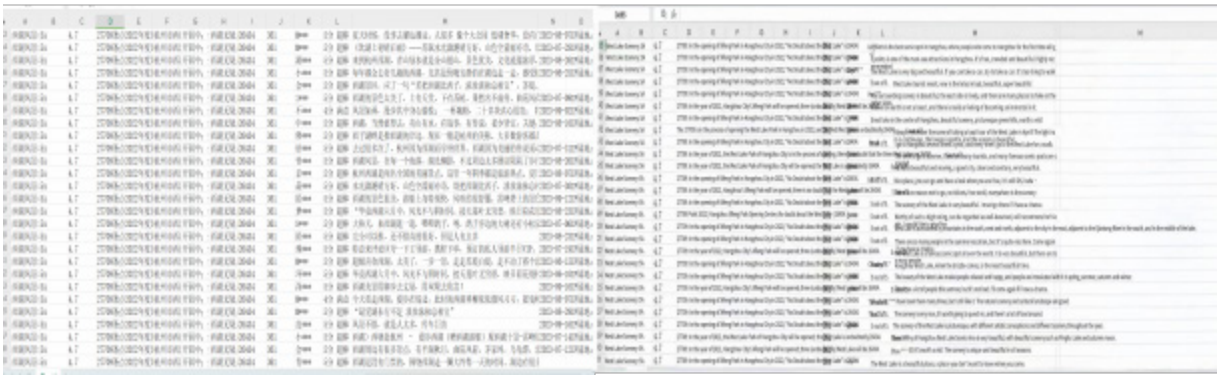

Figure c. bad review table

4.2 Data Pre-processing

4.2.1 Data cleaning

Due to the diverse and disorganized nature of the collected data obtained through big data collection software tools, it is impractical to differentiate or filter individual travelogue contents. As a result, inherent flaws may exist in the acquired dataset, such as repetitions and deviations from thematic relevance, which hinder further analytical procedures. In order to ensure both validity and representativeness of our findings while maintaining methodological rigor in this study's results, we undertake a meticulous manual browsing and screening process.

(1) The crucial step in the process involves removing irrelevant travelogues and reviews, as well as those with low relevance.

(2) Content processing. Travelogue writing is characterized by its unrestricted nature, resulting in the inclusion of a significant amount of content unrelated to the research topic. This includes: 1) advertising 2) personal anecdotes or stories that are not pertinent to the research destination, along with local news events or general knowledge dissemination; 3) content related to tourist destinations other than West Lake being included within the same travelogue.

(3) Unified processing. To ensure accurate and convenient data analysis, it is crucial to standardize the processing of tourist comments. Additionally, addressing typographical errors in visitors' comments becomes imperative.

(4) Processing of blank content and content format is crucial.

4.2.2 Chinese text parsing

We first perform the comment clause, in order to prepare for the later comment clause traditional sentiment analysis is based on the emotional characteristics of the sentence and the degree of adverbs of positive and negative for the weight of the score plus or minus, and ultimately get the sentiment score is the average result of the weight algorithm in the long comments. In order to get more accurate and real sentiment analysis results and to reduce the complex work of manual checking, the above text data are integrated and saved in a new Excel sheet, using the clause splitting function of Python software to identify the "," ". " " !" "..." and other symbols as clause markers, and the visitor comments were processed as clauses, and a new Excel with 35665 comment data was obtained for the next research.

The writing conventions differ significantly between Chinese and English languages. While words in English are naturally separated by spaces, this is not the case in Chinese texts where explicit word boundaries are absent. CWS involves dividing a continuous sequence of Chinese characters into individual words or phrases based on specific rules. These segmented units serve as features for sentiment classification purposes. The accuracy of subsequent experimental classifications heavily depends on the quality of word performed during preprocessing stage [10]. Chinese participle technology has emerged to effectively meet individuals' demands for natural language processing and address specific human-computer communication. However, Chinese participle technology faces various complexities related to word segmentation specification, disambiguation cutting, and recognition of neologisms.

The online travel reviews and travelogues pertaining to the West Lake in Hangzhou, utilized in this study, are partial texts and suffer from the aforementioned issues. Therefore, the data stored in the travelogue table, positive review table, and negative review table are merged. Subsequently, a new wordlist is generated using the novel word mining function within Micro Word software (as illustrated in Figure d). This comprehensive list encompasses details such as newly discovered words their properties, frequency of occurrence, cohesion level, degree of freedom, and probability for each individual word. Moreover, Hangzhou Ten Scenes, Hangzhou New

Ten Scenes, and Samping Ten Scenes are incorporated to tailor the dictionaries accordingly. Ultimately, these freshly created dictionaries are exported to Micro Word Cloud for Chinese splitting functionality. To further enhance accuracy and of results obtained from Chinese participle process (as depicted in Figure e), synonyms like "bad" and "have been included.

New words	Word frequency	Part of speech	Number of words	Coherence	Degrees of freedom	Probability of new words
Glass fish tank	3	Noun, noun	2	10.8403	1.791759	1
Literati poet	3	Noun, Verb	2	10.8403	1.386294	1
A photography enthusiast	3	Noun, noun	2	11.2414	0.693147	1
Gradually light up	3	Adverb, verb	2	11.2442	0.693147	1
Tingting Yuli	3	Names, Verbs	2	11.2457	0.693147	1
Asian tea culture	3	Place name, noun	2	11.2425	0.693147	1
Zan Zan Zan Ling	3	2 string, character	3	11.2425	0.693147	1
Zan Ling Ling Ling	3	2 string, character	2	11.2425	0.693147	1
Shake and roll	3	Verb, Noun	2	10.8359	0.693147	1
Prince Bay	4	Nouns, status words	2	10.551	0.636514	1
Lotus leaves are endless	6	nouns, adjectives	2	10.8967	0.627741	1
Vinegar fish	348	Noun, noun	2	6.0803	0.693147	0.999522
Leifeng is in the sunset	9	Names, Nouns	2	9.633	0.636514	0.982143
Surrounded by mountains on three sides	7	Numeral, Noun	2	10.694	0.966434	0.954545
Will be back again	3	strings, adverbs	2	10.5491	1.273028	0.95
Punch a card	3	String, Noun	2	10.5526	1.098612	0.95
Gongdi	34	Noun, noun	2	10.2917	0.980105	0.943478
Yoji	12	adverbs, adjectives	2	10.8836	0.766197	0.930556
Cultural Relics Protection Unit	15	Nouns	2	9.6935	2.159469	0.928571
Ice drink	2	Noun, noun	2	10.8403	0.693147	0.928571
Like a picture	2	Adverb	2	10.8368	0.693147	0.928571
Ticket 35	2	Nouns, numerals	2	10.4303	0.693147	0.928571
Mountain scenery, sky and rain	3	Nouns, personal names	2	10.5526	0.636514	0.9
Fried shrimps	4	Noun, String	2	11.9373	2.079442	0.875
This city	2	String, Noun	2	10.5478	0.693147	0.875
A number of water surfaces		Numeral, Noun	2	10.5493	0.693147	0.875

Figure d. a part of the new words list

Single word	All	quantity	Number of re	tf-idf
dependency	noun	787	787	0.0007044804
right	adjective	406	342	0.0005897724
very	adverb	389	322	0.0005807416
satisfaction	verb	377	375	0.0005244344
place	noun	286	246	0.0004783961
Beautiful view	noun	287	271	0.0004615224
Be worth	verb	287	283	0.0004532177
unknown	verb	250	250	0.0004154866
exactly	adverb	230	208	0.0004104769
view	noun	236	227	0.0004074216
veritable	adverb	209	183	0.0003908449
Zhejiang (Province)	toponym	173	170	0.0003320185
well	adjective	102	33	0.0003061881

Figure e. Apart of results from Chinese participle process

3.4 Textual analysis of online reviews and travelogues

3.4. 1 The analysis of word frequency

The word cloud is an emerging method of information dissemination on the Internet, where it filters a vast amount of textual data and highlights keywords with higher frequencies using different colors and font sizes. By visually representing these keywords, viewers can quickly grasp the main idea of the text.

In this study, the previously saved local excel files "good reviews", "bad reviews" and "travelogue" are importedinto the micro-word cloud webpage, and the results are as follows figures A ,figure B,figure C

Figure A of Word cloud diagrams for a collection of good reviews

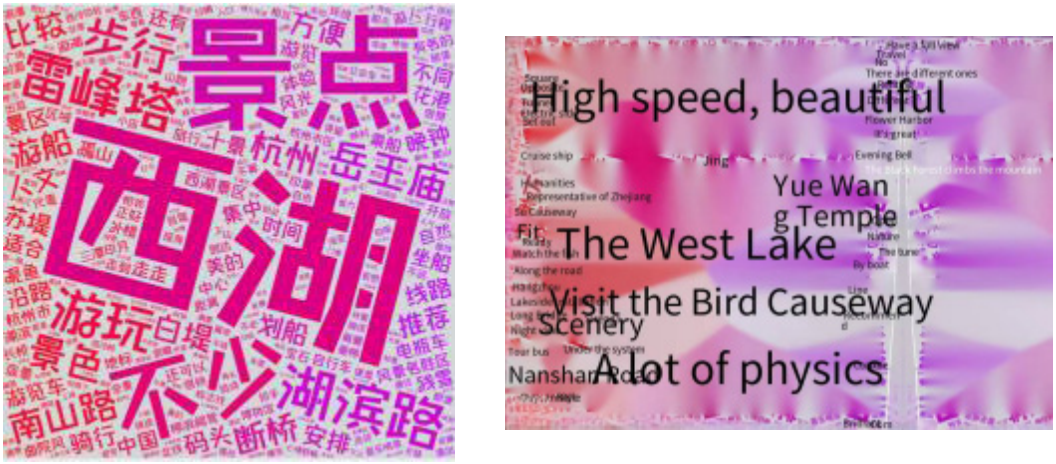

Figure B of Word cloud diagrams for a collection of bad reviews

Figure C of word cloud diagrams for the travelogues

The word cloud map formed by the combination of the Hornet's Nest in the travelogue set and Ctrip reveals numerous popular attractions at West Lake. The high frequency usage of these words indicates tourists' strong interest in these characteristic tourist spots from around the world. Furthermore, it highlights key elements contributing to tourism development within this scenic area. Tourists place great emphasis on quality experiences, ease of access, and affordability when visiting West Lake Scenic Area. Moreover, the overall feedback reflects a higher level of recognition among network tourists towards Hangzhou's West Lake Scenic Area. The word cloud generated from positive feedback data incorporates transportation-related terms such as 'cruise,' 'ticket,' and 'boat' to enhance accessibility to West Lake and establish seamless connections with various scenic spots, including Longxiangqiao subway station and bus routes. However, the word cloud generated from the dataset of negative reviews presents a perplexing combination of terms related to West Lake visits mentioned earlier and expressions associated with dissatisfaction. Notably, confusion is frequently mentioned along with both elderly individuals and children. The vocabulary related to accessing the West Lake and connecting its different attractions includes negative terms like overcrowded and noisy, which aptly depict the prevailing traffic problems in the West Lake area, especially on weekends and holidays.

3.4.2Analysis of semantic networks

(1) Principles of Semantic Network Analysis

The principle of semantic network analysis is to extract high-frequency words in the text and use them as nodes to draw the way of thinking of the semantic network of the text by using the directional relationship between nodes. The connection of the node question is close, and the more directional connections there are, the more obvious the lines are (which can be expressed with the help of power). Generally speaking, semantic network analysis is like a directed network diagram, which is used for the concept of language and the expression of relationships.

The analysis of text inevitably involves disambiguation processing, which can disrupt the original utterance's overall structure and introduce chaos into word relationships. To effectively interpret textual comments, it is crucial to accurately reconstruct disrupted word relationships. Semantic network analysis offers a highly effective solution by establishing meaningful connections between nodes and integrating confused words within the network. This method significantly contributes to constructing causal relationships among nodes and enables clear restoration of meaning conveyed in text comments. For example, extracted tourist concerns can be visually demonstrated through a semantic network analysis graph if most tourists are primarily interested in the correlation between scenic tickets and

cost-effectiveness.

(2) Building a Common Line Diagram of Social Semantic Networks

The ROST CM6 software was employed to construct the social and semantic network covariates for analyzing word frequency in the aforementioned tourists' comments. The social and semantic network analysis function automatically extracts high-frequency words from the text content, along with their associations (e.g., "queuing - very slow" indicates that tourists expressing concerns about queuing pay the attention to its speed). The direction of line segments represents the association's direction between high-frequency words ducks and mandarin ducks, implying observation of both types of ducks; a sunrise at the West Lake). Furthermore, stronger connections among these characteristics are depicted by heavier commonality through connecting lines. Consequently, we obtain a socio semantic covariance network graph illustrating online comments on the Lake Scenic Area.

The network relationships are plotted for the set of travelogues and the set of positive reviews. The results are as follows

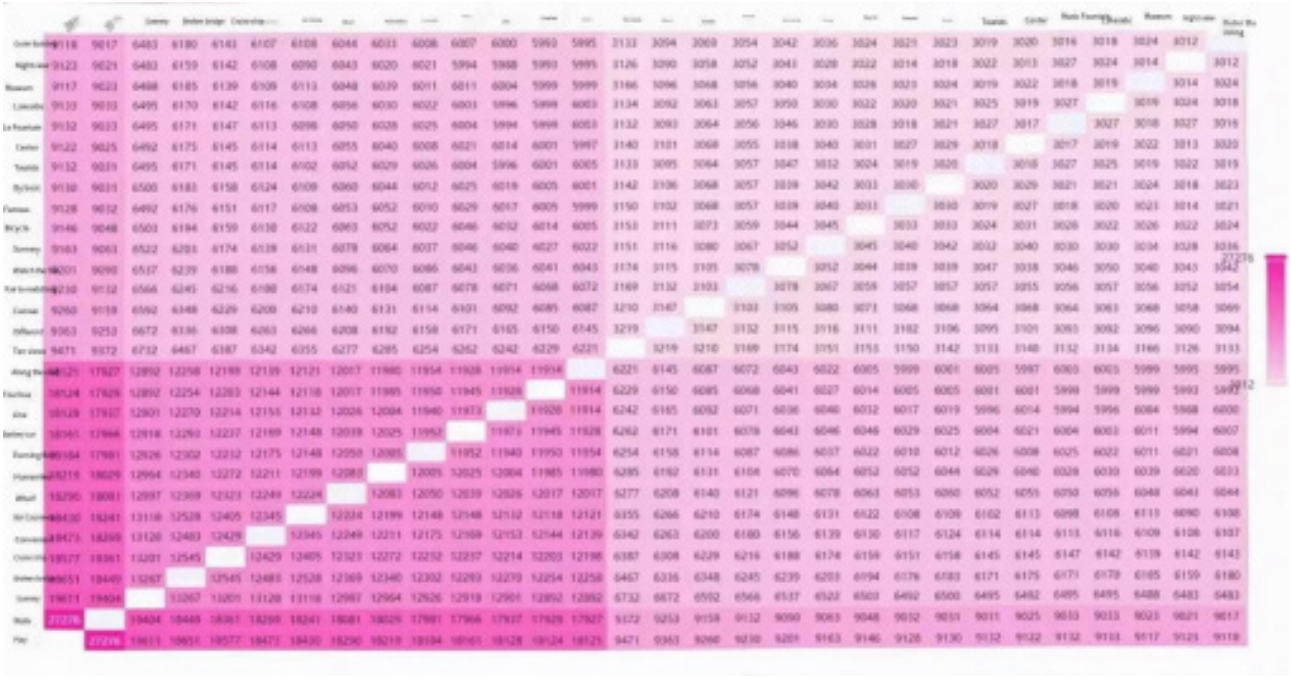

Figure 1.the network relationship for the set of positive reviews and the set of travelogues

As illustrated in the diagram, "Play," "Walk," "Scenery," "Broken Bridge," and "Cruise" serve as pivotal nodes within the entire relational network. The relationship chains of "Playing a Su Causeway-Walking" and "Scenery-Playing" hold utmost significance. This demonstrates that tourists' primary concern and enduring memory of West Lake lie in its awe-inspiring scenery, which they prefer to experience through leisurely walks. To enhance clarity, the depiction excludes West Lake itself due to its vast expanse, rendering the diagram less comprehensible. However, by incorporating the logo Lake, it becomes evident that it is the most frequently mentioned core term within the network. Additionally, words such as 'tickets,' 'remnant snow,' and 'beautiful scenery' closely associate with 'West recognition of its status aszhou. The proximity of,' 'scenic spot,' 'scenery,' and 'wind lotus' near nodes representing Leifeng Pagoda and Su Causeway indicates their significant impression on tourists.

3.4.3 Grounded analysis

The ROST semanticanalysistool is utilized to obtainhigh-frequencycharacteristicwordsforcoding. It extracts60 conceptualelementssuchasWestLake,HangzhouBridge,LeifengTower,time,appreciation, storiesassociated withtheenchanting beautyof snow-covered bridgesat night onthelakesurfaceor breathtaking hotelviewsetc. However,theseelementscrucial practical meanings in actual semantic environmentsthatrequirefurthertreatment Thefocusofthistopicisonthesemanticenvironmentin development coding, refining atotalof 11 corecategories includingnatural landscape,specialtyfood, humanistic landscape, accommodation level,and environmental hygiene.

(1) Natural landscape. The natural landscape of West Lake has made a profound impression on tourists. Some travelogues and comments highlight

issues with the scenic spots including similarities between attraction experiences,content independence among attractions,and low association between them,resulting in excessive time spent observing similar attractions

(2) Scenic spot. Hangzhou's West Lake Scenic Spot, known for its legends and cultural treasures such as Leifeng Pagoda, Lingyin Temple, Wusong's Tomb, Museum of Fine Arts, BaoDo Pagoda, and former residences of celebrities.

(3) Hotel accommodation. From the specific content of travel notes, it can be observed that tourists' choices regarding accommodation are closely intertwined with factors such as their income level, travel routes, seasonality, and duration of stay. These diverse hotel different price ranges cater to varied consumer preferences and enhance the inclusivity of the West Lake Scenic Area for motivated tourists.

(4) Special food. The perception of West Lake's special food among tourists is divided. Surprisingly, the same dish receives both praise and criticism from different individuals.

(5) Price level. The price level is a crucial factor assessed by tourists in the West Lake scenic area. In terms of the components of the scenic area, tourists have expressed diverse opinions, with some perceiving it as costly while others find it to be cost-effective. Specifically, accommodation and catering are regarded as expensive, whereas transportation and tickets are considered affordable.

(6) Tour guide service. Given the abundance of tourism resources, as evidenced by the 299 available at West Lake, the provision of tour guide services is essential. The high level of satisfaction among tourists with the tour guide service at West Lake is exemplified by features such as the one-click smart tour activity tool found at Hangzhou East Railway Station - an introduction to the wonders of West Lake. Comprehensive tour guide services encompassing tourism, gastronomy, lodging, transportation, shopping, and entertainment are provided for visitors exploring West Lake.

(7) Climate and weather. The Hangzhou West Lake boasts a consistently pleasant climate, pristine environment, and abundant vegetation coverage throughout the year. Different visitors are drawn to the West Lake for its diverse weather conditions.

(8) Programme activities. Programme activities include various events for visitors, such as the Hangzhou Marathon, Kite Festival, Low Carbon Ride, International Expo, Silk Expo and more. Tourists' comments suggest that many of these activities are ornamental in nature and may contribute to improving visitor loyalty.

(9) Supporting facilities. Some travel notes mention the scenic spot logo, announcement toilet and other basic supporting facilities. However, in general, tourists' overall perception of the supporting facilities of the West Lake Scenic Area is relatively weak, and the frequency of mention is not high, and it is mostly a summary.

(10) Transportation. In terms of transportation, the majority of tourists expressed convenience in accessing the West Lake. Based on the travelogue's content, public transport emerges as the preferred choice among tourists due to the well-connected network within the West Lake scenic area. Walking, cycling, and cruising are three common modes of exploration in this scenic area.

(11) Tourism evaluation. Many people have shared their views in their travel notes, or praise, or complain. Among them, most tourists praise them. However, some tourists gave their children a negative perception, saying.

By integrating, refining, and comparing the aforementioned main axis codes for natural resources, humanistic landscapes, and public transport while combining textual content recounted in travelogues; tourism resources; tourism facilities; tourism environment; tourism consumption and service; as well as tourism attitude, an accurate understanding of tourists' overall perceptions of West Lake as a tourist destination can be achieved by analyzing and evaluating these five main categories along with summarizing the frequency of mentions for each main axis code.

3.4.4 LDA topic model analysis and Snownlp Emotion Analysis Model

(1)Snownlp Emotion Analysis Model

Firstly, instead of utilizing the sentiment scores derived, this study employs labeled comments for training in order to avoid introducing significant errors when evaluating the emotional inclination of comments. The original Aira corpus is replaced with comments that have scores of 3 and below to represent negative emotions, while comments with scores of 4-5 are considered indicative of positive emotions. Subsequently, SnowNLP is used to train the substituted corpus, and if the posterior positive emotion surpasses that of negative emotion, it is classified as positive emotion; otherwise, it is categorized as negative emotion..Furthermore, by comparing recall, precision, and F1 value between using the original corpus and using the replaced corpus for training purposes, it was observed that replacing the corpus significantly sentiment discrimination achieved by model. Finally, all valid comments undergo sentiment analysis using our trained model: a score less than 0.3; a score greater than 0.7 suggests positive sentiment; any other score represents neutral sentiment..The results demonstrate a higher

proportion (96.6%) of positive emotional tendencies compared to negative ones in relation to people's tourism experiences at West Lake Scenic Area (96.6% exhibit a tendency), highlighting areas for improvement in enhancing tourism experiences.

(2) LDA topic model

Word	Whole	Quantity	Emotions	Number of front bars	Number of	Number of neutral bars	Total	Total score	Average score
Ride	Verb	22		3	2	4	9	0.00025138122	354	39.3333
Orioles are singing in the [illegible]	[illegible]	22		3	1	6	10	0.00023859905	374	37.4
Mainly	Distinguishing word	22		4	0	6	10	0.00023859905	480	48
Xiling Bridge	Noun	22		0	1	2	3	0.00049001503	95	31.6667
Leifeng is in the sunset	Custom words	21		4	1	4	9	0.00024222753	367	40.7778
Homestay	Noun	21		4	1	4	9	0.00024222753	307	34.1111
Find	Verb	21		1	4	11	16	0.00016613822	647	40.4375
Architecture	Noun	21		4	0	5	9	0.00024222753	399	44.3333
Opposite	Noun	20		3	4	8	15	0.00016249993	464	30.9333
Walk	Noun	20		4	2	8	14	0.00017186696	550	39.2857
Come	Verb	20		1	6	6	13	0.00018188865	410	31.5385
The top of the mountain	[illegible]	20		2	0	7	9	0.00023073383	410	45.5556
Should	Verb	20		3	2	5	10	0.00021690004	407	40.7
Gen flowing clouds	Custom words	20		0	1	5	6	0.00028252314	210	35
Rowing	Noun	20		1	0	0	1	0.00046440099	50	50
Only	Conjunction	19		3	1	9	13	0.00017277922	563	43.3077
And	Conjunction	19		5	2	9	16	0.00014601125	648	40.5

Conducting LDA topic analysis offers supplementary insights to inform subsequent recommendations in a scholarly and rigorous manner, aligning with the standards expected by Nature journal.

5. Conclusion

Drawing upon grounded theory and online travel narratives, this article integrates essential notions extracted from tourists' perspectives into five discernible dimensions: tourism attitudes, environmental influences, consumption behaviors and services, tourism assets, and infrastructure advancement.

As a fundamental component of cognitive perception, tourism resources account for up to 85% of overall cognition, with tourists possessing the most profound understanding of them. Both cultural and natural landscapes at West Lake attract tourists, among which Lei Feng Pagoda stands out as the most prominent attraction in the former category. In contrast, tourist programs within the scenic area are perceived at a relatively lower level, with only "Impression • West Lake" and musical fountains receiving attention.

The level of tourists' awareness regarding infrastructure ranks second only to that of tourist resources at a significant rate of 65%. Among these aspects, transportation comprehension emerges as the most one, closely followed by tour guide services; meanwhile, public toilets are ranked third in terms of importance. However, it seems that tourists have relatively weak perceptions towards supporting facilities. When visiting scenic areas in other provinces or regions, most visitors prefer utilizing public transportation options; additionally, cycling and walking also stand out as popular choices for exploring attractions while boating remains another highly recognized method among tourists.

The perception of tourism services and consumption by tourists is moderately satisfactory, with a rating of only 50%. During their travel experiences, tourists demonstrate a strong awareness of ticket prices and often emphasize the availability of free admission. The evaluation of service levels by tourists varies, indicating

room for improvement in the attitude displayed by staff within the scenic area. There is no significant disparity in tourists' understanding of shopping. However, they do highlight that overall consumption levels in the scenic area are relatively high but not widely accepted, particularly due to negative perceptions towards tourist shopping and hotel accommodation.

By integrating online travelogues and reviews, it becomes evident that tourists' negative evaluations

predominantly revolve around the subsequent aspects: firstly, exorbitant dining expenses coupled with substandard service attitude in the vicinity; secondly, excessive crowds and overcrowding at scenic spots; thirdly, dispersed attractions leading to fatigue from sightseeing; fourthly, opaque charges for services within the scenic area.

To optimize employee management, regular training and assessment are conducted to uphold service standards. This includes providing employees with the necessary skills and knowledge to perform their duties effectively, as well as evaluating their performance on a regular basis to identify areas for improvement. In addition, incentives such as bonuses or promotions may be offered to encourage employees to excel in their roles.

Government intervention and the involvement of relevant departments ensure price transparency and regulate market fluctuations. This involves setting regulations that govern pricing practices within the industry, monitoring prices regularly to ensure compliance with these regulations, and taking action against any companies found to be engaging in unfair pricing practices.

Three sets of traffic management plans are proposed for different occasions: on workdays (approximately 220 days), the focus is on integrating urban development with scenic areas while accommodating transit traffic crossing the city's third quadrant, allowing limited parking within designated areas; on weekends and general holidays (approximately 130 days), measures include implementing dedicated bus lanes around the lake area and enforcing one-way traffic for public and private vehicles within the scenic zone, as well as temporary control of certain road sections based on traffic flow along specific roads; during approximately 15 days a reservation-based access system is implemented in measures taken during weekends and general holidays, with an emphasis on promoting park-and-ride facilities. These plans aim not only at reducing congestion but also at improving safety by ensuring that all modes of transportation can coexist harmoniously.

Overall, these efforts demonstrate a commitment towards creating a sustainable environment where businesses can thrive while maintaining high levels of service quality. By investing in employee training programs, regulating market practice.

References

[1] Lee Y J. (2016) The Relationships among emotional experience, cognition, and behavioral intention in battlefield tourism .Asia Pacific Journal of Tourism Research 21(6):697-715.

[2] Prayag G, Hosany S, Muskat B, et al.(2017) Understanding the relationships between tourists' emotional experiences, perceived overall image, satisfaction, and intention to recommend , Journal of Travel Research 56(1):41-54.

[3] Knobloch U, Robertson K, Aitken R, et al. (2017)Experience, emotion, and eudaimonia: A consideration of tourist experiences and well-being , Journal of Travel Research, 56(5):651-662.

[4] Kim J J, Fesenmaier D R. (2015)Measuring emotions in real time: Implications for tourism experience design [J]. Journal of Travel Research, 54(4):419-429.

[5] Mao Chaoqun. (2018) Research on the emotional classification of online travel comment texts based on improved emotional dictionaries Hangzhou: Zhejiang University of Technology and Industry.

[6] Wang Weiqing. (2019)Research on image perception of tourist destinations based on online text analysis . Nanchang: Jiangxi University of Finance and Economics

[7] Huang Shengnan.(2014) Research on the image perception of tourist destinations based on the network text of this analysis method . Hefei: Anhui University

[8] Gao Fen.(2020)Research on the local image and co-improvement strategy of Laoshan Scenic Area Tourism Day based on the network text analysis method .Qingdao University

[9] Han Yunbo, (2015)Jiang Dengke. References National Standard CB/T7714--2015 Revision Features and Detailed Rules .Journal of Southwest University (Social Science Edition), 41 :157

[10] Fan Xiaojin. (2014)Research on Chinese text card classification and emotion classification of basic machine learning. Nanjing University of Technology