Article

Urban Tourism Destination Image Perception Based on LDA Integrating Social Network and Emotion Analysis: The Example of Wuhan

Xu Chen, Jun Li , Wenxin Han and Shudong Liu

School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan 430037, China; chenxu@whu.edu.cn (X.C.); lijunlg@foxmail.com (J.L.); hanwenxinwork@163.com (W.H.)
* Correspondence: liushudong@zuel.edu.cn

Abstract: Tourism destination image perception aims to depict the urban tourism image from the perspective of the perception of tourists, which, therefore, sheds new light on the advancement and innovation of urban tourism. The model proposed in this study can effectively describe the image perception of a tourism destination, with its research conclusions providing a vital referential basis for the sustainable development of urban tourism. Combined with LDA, we construct the research framework of tourism destination image perception and then take the online comments of popular scenic spots in Wuhan on Ctrip Travel as an example. The results show that four aspects are included in tourists' perception of the city image of Wuhan: experience, history culture, leisure service, and tourist destination. Among them, the social network of the experience dimension is most closely related. In addition, emotion analysis illustrates that tourists' emotional tendencies tend to be positive under the four perceptual dimensions.

Keywords: urban tourism; latent dirichlet allocation (LDA); social network analysis (SNA); emotion analysis; image perception

Citation: Chen, X.; Li, J.; Han, W.; Liu, S. Urban Tourism Destination Image Perception Based on LDA Integrating Social Network and Emotion Analysis: The Example of Wuhan. *Sustainability* **2022**, *14*, 12. <https://doi.org/10.3390/su14010012>

Academic Editor: Mark A. Bonn

Received: 25 November 2021

Accepted: 19 December 2021

Published: 21 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recent days have witnessed increasingly fierce competition in the urban tourism market. Hoping to win potential tourists and achieve sustainability in tourism industry development, a growing number of cities are seeking to outline a better urban image, hence, making tourism destination image research an important topic in the field of today's urban tourism research.

Tourism destination image (TDI), a concept first proposed by scholar John Hunt, refers to the overall cognition, evaluation, and impression of tourists and potential tourists on a tourism destination [1,2]. After the 1970s, some scholars set their sights on the connotation mechanism [3], influencing factors [4], and formation factors [5] of tourism destination image. For instance, Gunn divides the image of a tourism destination into two types: the original image formed through social media and the induced image formed through commercial information sources [6]. Tourism destination image affects tourists' travel decisions to a certain extent [7]; a distinctive tourism image is instrumental in building local tourism brands, attracting potential travelers, and developing urban tourism industry in a more sustainable way.

With the advent of the Web 2.0 era, the wider application of Internet technology in the tourism industry enables more and more tourists to post personal experiences and evaluations of their trips online on tourism platforms. Huge amounts of real, independent, and valuable online comment data thus emerge, together with a new direction: researchers use these user comment texts for the study of tourism destination image. However, the large scale of data source also brings about great difficulty for manual analysis. That is why the development of natural language processing technology becomes essential; this

new tool makes it possible to analyze large-scale data, which, therefore, provides technical support for research of the image perception of urban tourism destinations.

Combined with LDA (latent Dirichlet allocation), this study constructs the framework of tourism destination image perception by carrying out social network analysis as well as emotion analysis and looks into the practical case of the tourism destination image perception in Wuhan. Firstly, Octopus software is used to crawl tourists' comments on popular scenic spots in Wuhan that are published on Ctrip Travel. Then, the main research framework is proposed. Through data preprocessing, high frequency words in tourist comment texts are counted; at the same time, the LDA theme model is established to identify tourist perception dimensions. Further, social network analysis and emotion analysis are employed to explore the correlation of feature words and tourists' sentiment orientations under each perceptual dimension. The research results can help pave a new road for future sustainable development and the innovation of tourism in Wuhan and other cities.

2. Related Works

2.1. Tourism Destination Image Perception

Research on the image perception of an urban tourism destination has always been a hot spot in the tourism field. At present, the existing research methods on urban tourism image perception mostly focus on questionnaire surveying and web text mining.

As a traditional research method, questionnaire surveying enjoys the advantage that its results, obtained by questionnaires and interviews, are easy for quantification, statistics, and analysis. Martens et al. study the image perception of German tourists on two tourist destinations in Abu Dhabi and Dubai through field investigations [8]. Cassia et al. collect questionnaires from residents and tourists in the Italian city of Verona to compare the differences in the perception of urban tourism image between local residents and tourists [9]. Kim et al. conduct an empirical study on the structural equation model of urban personality, urban image perception, and willingness to revisit, based on the data from 302 interviewees participating in an online survey [10]. However, when applying this research method, it is difficult for researchers to design a reasonable and comprehensive questionnaire, as a result of which the results are highly likely to be biased due to subjective impacts.

Another research method, web text mining, explores online comments made by tourists on relevant tourist destinations, which has gradually become a mainstream method now. In recent years, the rapid prevalence of online tourism platforms has encouraged a mass of tourists to comment and share notes online, providing a new source of information for research on tourism destination image. Zaid Alrawadieh et al. study the tourism image of Istanbul based on blogs published by western travel bloggers [11]. Hillary Clarke and Ahmed Hassanien use content analysis to evaluate the image of Toronto by digging into contents generated by tourists on Twitter [12]. Gao Yin et al. analyze tourists' perception of Chongqing Garden Expo Park from the aspects of seasonal influences, emotional scores, and attention factors by using Chongqing Garden Expo Park Network comment data [13]. Dong Shuang et al. take online reviews of the National Mine Park as a sample and find that tourists' perceptions of the park are mainly reflected by their cognition of functional objects (such as attraction to scenic spots, service quality, tourism destinations, etc.) [14]. However, one problem exists that, in general, these studies rarely emphasize the characteristics and differences of each perception dimension from a micro perspective. Even fewer refined studies on social network analysis and emotion analysis take into account tourists' perception dimensions.

2.2. Research on Topic Mining

The topic mining method is an effective tool to identify text topics and delve into users' online opinions [15]. Currently, there are two types of topic mining: the traditional topic clustering model that depends on text similarity [16,17] and the probabilistic topic model, such as the LDA topic model [18,19]. Studies have shown that the LDA model

can extract negative comments online about e-commerce [20] and identify the subject distribution within data on social media [21]. In addition, the LDA model can be used to detect real tourism hot spots and the key information needs of tourists in historical tourism experiences, and then, efficiently, find the theme features of short texts [22]. Thus, this study adopts the LDA theme model to identify the theme of tourist comments.

First proposed by Blei et al. in 2003, the LDA (Late Dirichlet Allocation) topic model is the most representative among text topic mining methods [23] and is also known as three-layer Bayesian probability model. As a document topic generation model, the multiple distribution of topics in each paragraph of text is generated by Dirichlet prior related to a corpus. There is a set of documents, denoting the number of documents with D and the number of topics with K ; each document D is a sequence of N words, where $w_{d,n}$ is the n^{th} word of document D , and $z_{d,n}$ is the n^{th} topic of document D . $\theta_d \sim \text{Dir}(\alpha)$ obeys Dirichlet probability distribution, so the hyperparameter is α . $\beta_k \sim \text{Dir}(\mu)$ obeys Dirichlet probability distribution, so the hyperparameter is μ . The joint probability distribution function of the LDA model is shown in Formula (1) [24]:

$$p(\beta, \theta, z, \omega) = \left(\prod_{d=1}^D p(\theta_d | \alpha) \prod_{n=1}^N p(z_{d,n} | \theta_d) \prod_{d=1}^D p(z_{d,n} | \beta_{i,k}, z_{d,n}) \right) \cdot \left(\prod_{i=1}^k p(\beta_i | \mu) \right) \quad (1)$$

Although we can effectively grab topic information using the probabilistic topic model, it is still difficult to find the relationship between topics, as well as respondents' sentiment orientations, by relying solely on topic mining. Therefore, besides the application of a probabilistic topic model, this study further constructs a social network with the help of social network analysis; meanwhile, emotion analysis is introduced to quantify customers' emotional tendencies. Social network analysis (SNA) can generate microscopic results that indicate the association among feature words under each topic, according to the dependence and cooperation of each network node [25]. Emotion analysis, using natural language processing technology, identifies and assesses subjects' views and attitudes towards physical objects (such as products, events, and themes) [26].

At present, SNA based on LDA is widely used in many research fields. Take the example of the field of science and technology: the related literature shows how the LDA model combined with the social network method measures the patent technology fusion of joint robots and, thus, predicts the trend of technology convergence [27]. When it comes to the medical field, SNA and LDA realize the comparison with regard to the adverse medical event reports of medical staff [28]. In terms of a mining news hot spot, an online news text hot spot mining model is proposed based on LDA and SNA; such application proves strong portability of this model, too [29]. As for urban planning, some studies use the LDA and network analysis methods for reference when exploring the evolution path and power of urban portraits [30].

Moreover, emotion analysis based on LDA gradually proves an effective way for scholars to identify users' emotional tendencies. He Yue et al. capture the sentiments hidden behind original microblogs under the "smog" topic and use the Gephi platform to analyze the social network of those comments [31]. Ma Guizhen et al. take the review data of five-star hotels in Beijing on the Tripadvisor website as the research object to conduct a cross study of theme mining and emotion analysis [32]. Wafa Shafqat proposes a scenic spot recommendation mechanism using theme modeling and emotion analysis, which finds that the combined features of LDA, SVM, ratings, and cross mapping help improve the performance of the recommendation system [33].

Despite the current success of the LDA model, tourism destination research still goes beyond its capacity without the assistance of other complementary tools. One reason is the large categorical range of tourist attractions: according only to the classification of tourism resources, there are 8 main categories, 23 subcategories, and 110 basic categories included [34]. Moreover, the attention of tourists to different types of tourist destinations varies a lot. Hence, this study introduces social network analysis and emotion analysis

on the basis of identifying topics, so that the research work can better explore tourists' perception dimensions and their emotional tendencies to the destination.

3. Research Framework Design

Referring to the past literature about tourism destination image perception at home and abroad, this study constructs a framework of tourism destination image perception in Wuhan; the technical methods employed are natural language processing, topic mining, social network analysis, and emotion analysis. The research framework is shown in Figure 1:

Figure 1. Research framework.

3.1. Data Acquisition and Preprocessing

This paper selects the tourist comments on Ctrip Travel (the largest online travel platform in China) as the data source. With the help of Octopus software 11,826 user comments are crawled from 195 popular scenic spots (i.e., Yellow Crane Tower and East Lake) in Wuhan. After that, 2208 duplicate or priceless data (comments with empty content, advertisements, etc.) are removed. Finally, we obtain 9618 valid tourist comments for follow-up research.

Data preprocessing mainly consists of Chinese word segmentation, part of speech tagging, removal of stop words, and integration of synonyms. To start with, for Chinese word segmentation and part of speech tagging, we make use of the Lexical Analysis of Chinese, an open source technology of Baidu natural language processing. Its overall accuracy of word segmentation, part of speech, and proper name recognition is 95.5%, which is sufficient to accomplish Chinese analysis and part of speech tagging. In the second step, we form the initial stop words list in accordance with the Baidu stop words list and the Chinese stop words list. Then, with the word frequency analysis results, those meaningless high frequency words identified in the perception dimension are added to the final stop words list. The last step is to conclude words with similar meanings in the thesaurus, so as to complete the semantic deduplication and integration. After the above data preprocessing, noise data interfering with perceptual dimension recognition is reduced.

3.2. Establishing LDA Topic Model

Recognition of the tourism destination perception dimension depends on the results of LDA subject classification. Since the number of LDA topics cannot be obtained directly, our study refers to the method for determining the optimal number of LDA topics based on topic similarity and confusion [35] from the relevant literature. The value range of subject number K is set as [1, 10], and four groups of test sets are extracted to calculate the Perplexity index.

$$\text{Perplexity}(D_{\text{test}}) = \exp \left\{ \frac{-\sum_{d=1}^M \log(p(w_d))}{\sum_{d=1}^M N_d} \right\} \quad (2)$$

M is the size of the test corpus and N_d is the text size (number of words) of article D.

$$p(w_d) = \sum_k p(k)p(w|z, \gamma) \quad (3)$$

K is the topic, W is the document, and gamma is the text topic distribution trained by the training set. The stronger the model generation ability is, the smaller the property value is.

Now use Python's data visualization package Matplotlib to draw a line graph of confusion and the number of topics, observe the inflection point of the line graph, and obtain the value to be selected for the optimal number of topics. Then, substitute the optimal number of topics to be determined into the LDA model for testing. To ensure that there is no overlap between various topics, the LDA is visually analyzed in advance by the gensim package in pyLDAvis; the visualization results determine the number of topics with better classification effect, and the final LDA topic model is now established. With this LDA model, we obtain the probability distribution of topic feature words and can, therefore, analyze the semantic characteristics of feature words under each topic. The topic names are manually summarized, that is, the perception dimension of the tourism destination image.

3.3. Social Network Analysis

The social network of feature words is a fine description of the image of Wuhan city as a tourist destination. Characteristic words are sorted under each perception dimension, after which the top 10 feature words in each perception dimension form a feature word dictionary. Finally, the co-occurrence matrix of feature words and feature words can be calculated.

The number of common occurrences between two phrases in the text is counted by this co-occurrence matrix, so as to describe the intimacy between the two. The construction process of the co-occurrence matrix is as follows:

- (1) Select the top 10 feature words under each perception dimension to form a feature word dictionary.
- (2) Store all feature words in one list and their corresponding word frequency in another list.
- (3) Construct a two-dimensional matrix whose vertical and horizontal correspond to each word in the total feature words.
- (4) Assign an element on the diagonal of the co-word matrix the number of times it appears in all text.
- (5) Loop through the feature word list to construct the combination between all two words and then traverse the word segmentation results of each article. If two words appear in the same article, the weight of them increases by one; at the same time, the two words are stored in the corresponding positions of the common word matrix.

Finally, we use the social network analysis tools UCINET and Netdraw to visually illustrate the social network relationship of feature words under each perceptual dimension of Wuhan's tourism destination image.

3.4. Emotion Analysis

With the results of LDA topic classification, the perception dimension corresponding to each comment can be calculated using the Senta model by paddlehub. Then, based on Bi-LSTM (Bi-directional Long Short-Term Memory) [36] emotion classification, we analyze the emotional tendency of each tourist comment.

The emotion classification model based on Bi-LSTM includes three layers: word semantic layer, sentence semantic layer, and output layer. The model is shown in Figure 2.

Figure 2. Emotion classification model based on Bi-LSTM.

- (1) The word semantic layer mainly converts each word in the input text into a continuous semantic vector representation, or, that is, the embedding of words.
- (2) In the sentence semantic layer, the sequence of word semantics is transformed into the semantic representation of the whole sentence through the Bi-LSTM network structure.
- (3) The output layer calculates the probability of emotional tendency based on sentence semantics.

There are three types of emotion: positive, neutral, and negative. Calculated according to its confidence level, the emotion score is distributed in the $(-1, 1)$ interval, where -1 represents negative, 0 represents neutral, and 1 represents positive. We count the emotional scores of tourists' comments in each perception dimension; calculate the average score of positive and negative emotions, respectively; and, lastly, display it intuitively by charts or tables. The emotion analysis process is shown in Figure 3.

Figure 3. Emotion analysis flow.

4. Results and Analysis

4.1. Data Processing

In this study, first, Octopus software is used to collect data of tourist comments on Wuhan from Ctrip Travel, which are stored in the CSV text file named tourist comments as the data set. The tourist comments extracted come from the top 195 tourist attractions in Wuhan, including both cultural attractions such as Yellow Crane Tower and Hubei Provincial Museum as well as natural landscapes such as East Lake and Mulan Tianchi, with a total number of 9618 comments in the final data set.

Next, the Chinese part of speech analysis method *lac* is applied to segment the comment text, and then the Baidu stop word list and the Chinese stop word list are used to remove stop words. During the above process, we enrich the stop word list by adding to it words with high frequency (shown by the word frequency statistics results) but little meaning for the recognition of perceptual dimensions (such as you and yes); also useless adverbs, such as when and whole, are removed. Then, words with high frequency and the same meanings are added to the synonym list to combine the word items, such as Wuhan and Wuhan City. The final analysis results display the top 52 feature words according to word frequency from high to low, as shown in Table 1:

Table 1. High-frequency characteristic words of Wuhan tourists' online comments.

Characteristic Words	Frequency	Characteristic Words	Frequency	Characteristic Words	Frequency	Characteristic Words	Frequency
Wuhan	2941	park	609	look	478	tourists	296
good	2306	scenic area	599	East lake	457	air	291
not bad	1938	Yellow Crane Tower	585	greatest	434	take a picture	85
landscape:	1232	be located	577	Cherry blossoms	406	convenient	278
local	1095	very good	555	traffic	401	area	276
worth	1000	environment	548	see	394	night	262
architecture	960	special	547	like	381	project	253
scenery	770	interesting	523	visit	370	Yangtze River	252
history	739	cost performance	522	city	336	go in	251
fit	694	tickets	520	China	333	weekend	246
scenic spot	675	culture	505	museum	330	recommend	245
amusing	651	time	499	big	326	Qingchuan Pavilion	241
feel	632	total	490	kid	322	beautiful	229

Word frequency statistics show that the word Wuhan, as a tourist destination, appears the most frequently, far more than other words. In terms of tourist experience, good, worthy, and fun appear more frequently than others, indicating that tourists are highly satisfied with their travels in Wuhan. When it comes to tourist attractions, Yellow Crane Tower, Qingchuan Pavilion, and East Lake enjoy greater popularity. In addition, the Yangtze River and the cherry blossoms of Wuhan University are popular attractions for tourists as well. With regard to the tourism environment, words such as history, park, and museum appear frequently, which corresponds to the strong historical, cultural, and humanistic atmosphere of Wuhan. As for tourism services, cost performance and tickets reflect existing issues that tourists are more concerned about.

4.2. LDA Topic Mining Results

To establish the LDA model, we first need to determine the optimal number of topics. Referring to the experience of the related literature [3], this study sets the value range of the number of topics K as [1, 10], calculates the perplexity index under the different number of topics, and then uses Python's data visualization package Matplotlib to draw a line graph of confusion related to the number of topics (as is shown in Figure 4).

In the line chart of the four test sets, the number of topics of inflection points is 4 and 8. $K = 4$ and $K = 8$ are, therefore, substituted into the LDA model, respectively. Then, the LDA model is visually analyzed using the gensim package in pyLDAvis. When $K = 8$, as is shown in Figure 5 below, theme 6 and theme 8 overlap in a large area; also, theme 2 and theme 4 overlap in part, too. Since there are many same feature words in each theme, it is difficult to summarize the theme.

When $K = 4$, as is shown in Figure 6 below, each topic has clear boundaries with no overlap and good independence. So, after comparison, it is determined that the optimal number of topics is 4. Thus, the LDA topic model with $K = 4$ is established.

Figure 4. Perplexity: a line graph of the number of topics.

Figure 5. LDA topic model with K = 4. (a) Topic 1, (b) Topic 2, (c) Topic 3, and (d) Topic 4.

The top 10 feature words of each topic are selected as representatives. We then analyze the semantic features of feature words, manually identifying and summarizing the four perceptual dimensions. Table 2 below shows the obtained perception dimensions of image perception of tourism destination in Wuhan, as well as the probability distribution of characteristic words in each dimension, respectively.

Figure 6. The results of social network visualization: (a) experience, (b) history and culture, (c) leisure service, and (d) tourist destination.

Table 2. Perception dimension of tourism destination image and probability distribution of characteristic words in Wuhan.

Topic 1 Experience		Topic 2 History Culture		Topic 3 Leisure Service		Topic 4 Tourist Destination	
not bad	0.028	museum	0.009	local	0.019	Yellow Crane Tower	0.018
good	0.023	campus	0.007	ticket	0.015	Qingchuan Pavilion	0.016
worth	0.012	history	0.005	traffic	0.008	Hankou	0.011
fit	0.012	exhibition hall	0.005	scenic spot	0.006	temple	0.01
fun	0.011	art	0.005	street food	0.006	church	0.009
very good	0.01	architecture	0.005	lively	0.006	Yangtze River	0.006
cost performance	0.008	culture	0.004	play	0.006	Iron Gate Pass	0.005
interesting	0.008	Wuhan University	0.004	take a picture	0.005	Guishan	0.005
environment	0.008	library	0.003	convenient	0.005	famous	0.004
greatest	0.007	time	0.003	facility	0.004	Wuchang	0.004

The results of the LDA model tell us that the four perception dimensions of image perception of tourism destination in Wuhan are experience, history and culture, leisure service, and tourism destination. The characteristic words with high probability in the experience dimension are good, can, worth, suitable, and fun, which represent the overall feeling of tourists towards tourism in Wuhan. As Wuhan is widely known for its many places of interest and its long history, the historical and cultural dimension is characterized by museums and campuses as well as history and exhibition halls. Tourists' attention to the dimension of leisure service mainly lies in tickets, transportation, snacks, play, and so on, indicating that urban leisure service plays a vital role in tourists' travel satisfaction. Besides, Hankou and Wuchang prove to be the areas where more tourists choose to travel in the three towns of Wuhan.

4.3. Social Network Analysis Results

Based on the findings of the LDA perceptual dimension recognition, the feature word co-occurrence matrix is constructed for feature words under each perceptual dimension. With the matrix, the social network relationship of those feature words under each percep-

tual dimension is visually presented by use of UCINET and Netdraw software. The results of social network visualization are shown in Figure 6.

With the results visualized, we further analyze the social network relationship between feature words in each perceptual dimension. The thickness of feature word lines reflects the co-occurrence degree between corresponding feature words. The thicker the lines, the higher the co-occurrence frequency of the two feature words.

Figure 6a represents the dimension of experience and feeling, where the feature words are generally closely related. Among them, the co-occurrence of words such as good, excellent, suitable, and can witnesses the highest frequency, and this implies that in most cases, tourists hold positive perceptions and evaluations of Wuhan's tourism destination image.

In Figure 6b, for the historical and cultural dimension, the thickness of the connecting lines between characteristic words varies greatly. The high common occurrence frequency of campus, Wuhan University, culture, and library embodies the heavy cultural atmosphere of Wuhan. The co-occurrence of architecture and museum is high too, which is in line with the architectural attribute of the museum. What is more, art and exhibition hall also experience frequent co-occurrence, indicating that an exhibition hall might be the main platform for displaying artistic works.

In Figure 6c, most of the feature words in the dimension of leisure service are connected independently, which means that the relevance of these feature words is probably low. For example, taking pictures, bustle, and snacks are closer to leisure, while transportation and convenience are more related to scenic services. However, the relevance of tickets and scenic spots is exceptionally high; this could be because most scenic spots require tickets for entry.

Figure 6d shows the co-occurrence degree of characteristic words in the dimension of a tourism destination. This actually reflects the geospatial characteristics of scenic spots in Wuhan. For instance, the high co-occurrence degree between Yellow Crane Tower and Wuchang indicates that the Yellow Crane Tower is located in Wuchang District, in terms of geospatial space. Another noticeable point is that the Yangtze River lies at the center of this social network, related to all other characteristic words, and this possibly explains the laudatory name for Wuhan: river city.

4.4. Emotion Analysis Results

In this part, the Senta model provided by paddlehub is employed to analyze the emotional tendency of tourist comments. The emotional score of each tourist comment is calculated, as is the mean and variance of positive emotion and negative emotion under each perception dimension. The values we obtain are listed in Table 3:

Table 3. Average emotion value.

Topic	Experience		History Culture		Leisure Service		Tourist Destination	
Emotion Tendency	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
percent/%	95	5	94.2	5.8	89.1	10.9	93.5	6.5
mean value	0.87	−0.46	0.89	−0.55	0.85	−0.51	0.85	−0.4
variance	0.03	0.09	0.03	0.09	0.05	0.1	0.04	0.09

It can be seen from the table that the proportion of positive emotion under the four perceptual dimensions is more than 90%, indicating that most tourists tend to show positive attitudes towards the city image of Wuhan as a tourism destination. The value range of emotional score of positive emotion is [0, 1]: the closer the score is to 1, the higher satisfaction the tourist expresses. Similarly, the value range of negative emotion score is [−1, 0], and the closer the score is to −1, the stronger the negative emotion of the tourist holds. When comparing the four perceptual dimensions, there seems to be little difference. To be specific, the mean scores of positive emotion are 0.87, 0.89, 0.85, and 0.85, respectively,

and the variance is 0.03, 0.03, 0.05, and 0.04, respectively; the mean scores of negative emotion were -0.46 , -0.55 , -0.51 , -0.4 , respectively, and the variances were 0.09, 0.09, 0.1, and 0.09, respectively. What we can reasonably infer is that the negative emotion scores experience greater dispersion than the positive ones.

According to the proportion of positive and negative emotions under each perception dimension, we calculate the average emotion value of each perception dimension. The results are shown in Figure 7.

Figure 7. Mean emotional score.

As is shown above, among the four dimensions, the positive emotion score of history and culture accounts for the largest proportion, proving tourists' appreciation of Wuhan's beautiful cultural landscape and attractive historical heritage. In addition, the negative emotion score of the leisure service dimension is relatively high. Analysis of tourists' comments with low emotion score indicates that factors causing low tourist satisfaction mainly consist of poor notification management, polluted scenic environment, annoying entrance services, and so on.

5. Discussion about the Case

Combined with natural language processing technology, this study carries out social network analysis and emotion analysis based on the LDA topic model, and, therefore, constructs a research framework of tourism destination image perception. By collecting online comments by tourists on the Ctrip Travel platform, we employ the constructed framework to look into the image perception of tourism destination for Wuhan, with the main findings listed as follows:

Through LDA theme recognition, it is found that the four dimensions of Wuhan's image perception as a tourism destination are experience, history and culture, leisure service, and tourist destination. The high frequency characteristic words of the dimension of experience are good, can, worthy, suitable, etc., which indicate that most tourists choose Wuhan as a tourism destination wishing for a better tourism experience. For the historical and cultural dimension, key words are museum, campus, history, exhibition hall, etc., reflecting the deep historical accumulation and profound humanistic atmosphere of Wuhan. In terms of the leisure service dimension, words such as tickets, transportation, snacks, play, and so on clearly show the keen attention tourists attach to food, transportation, consumption, etc., when traveling to Wuhan. As to the tourist destination dimension, Yellow Crane Tower, Qingchuan Pavilion, church, temple, etc., are mentioned most frequently, which point out those most popular scenic spots in Wuhan.

Based on the LDA model, the social network analysis shows the relationship and intimacy degree of feature words under each perception dimension. For the experience dimension, those feature words are generally closely related, and tourists share a similar sentiment of feeling good and interested. Under the historical cultural dimension, the relevance of different characteristic words varies greatly, among which the common occurrence frequency of campus, culture, and library is higher, so is that of architecture and museum as well as that of art and exhibition hall. The dimension of leisure service sees little relevance among feature words, apart from the exceptionally close relationship between ticket and scenic spot. With regard to the dimension of tourism destination, the co-occurrence degree between Yellow Crane Tower and Wuchang is very high; also, the Yangtze River is at the center of the social network, which to some extent expresses the regional distribution characteristics of tourist attractions in Wuhan. The map of Wuhan city is shown in Figure 8.

Figure 8. The map of Wuhan city.

With the classification results of the LDA model, we analyze the emotion of tourists' comments under each perceptual dimension, coming to the conclusion that the vast majority of tourists' sentiment tendencies are positive: these happy feelings mainly come from the charming natural scenery and cultural buildings, a variety of delicious snacks, a warm and cheerful urban atmosphere, etc. Most of the existing negative emotions, however, grow out of the following issues: firstly, some of the ticket prices are unreasonably high, while the viewing experience and services of the scenic spots do not deserve such expensive tickets; secondly, poor environment, outdated facilities, and lack of a unique attraction are irritating to most tourists; thirdly, visitors suffer from the excessive flow of people at peak times, together with terrible congestion and endless queuing times, which adversely impact travel experiences to a large extent; finally, over-commercialization is highly likely to destroy the original cultural connotation of many scenic spots.

As is shown by the recognition of the image perception of Wuhan as a tourism destination, tourists mainly pay attention to tourism attractions, such as travel destination, food and services, while their perception of other tourism elements, such as hotel accommodation, shopping consumption, entertainment and leisure, is relatively limited. So based on

the research conclusion, several suggestions are put forward to promote sustainability in the development of tourism industry in Wuhan:

First, optimize the management and service quality of scenic spots. Second, strengthen the construction of facilities in scenic areas. Third, create a characteristic image for city tourism (The image of Wuhan city is shown in Figure 9). Fourth, regulate marketing publicity.

Figure 9. The image of Wuhan city.

6. Conclusions

Based on the recognition of the perceived dimensions of a tourism destination, and the analysis and research on the correlation degree of feature words and tourists' emotional tendency under each perceived dimension, the primary contributions of this paper are as follows:

First of all, this study introduces a new research perspective, with an applicable comprehensive urban tourism research model constructed accordingly. This makes up for the defect of previous studies, where only a few aspects concerned are looked into and relevance is often ignored. For example, the study on the evaluation of tourism service quality in Wuhan constructs a tourism service quality index system to measure the perception of tourists [37]. Other research on the image perception of Wuhan's tourist destinations concentrates on participants' evaluation, from the two aspects of cognition and emotion [38]. Our research, however, builds a social network between feature words in the perceptual dimension, which can reflect the relevance of topic words under each topic in more detail. The co-occurrence degree of feature words under the tourist destination dimension enables us to find the geographical spatial characteristics of Wuhan's scenic spots. One example is the high degree of co-occurrence between Yellow Crane Tower and Wuchang, which implies that the Yellow Crane Tower is located in Wuchang District; or another example, that Yangtze River is at the center of the social network and is related to other characteristic words, explaining the special name Jiangcheng of Wuhan.

Secondly, the factors studied in this research are expected to help optimize the level of urban tourism management, so as to realize sustainable development goals to the fullest. The tourism industry in cities could improve the overall service quality in a targeted manner, and the promotion of tourist spots could be more efficiently strategized.

Last but not least, the results of this study provide valuable and practical information for tourism stakeholders as a reference. Surrounding businesses could make specific and timely adjustments to their services, hence better meeting the preferences of their target audience and enhancing customer satisfaction. Moreover, tourists can benefit from conveniently obtaining the information they need when selecting a tourism destination.

Author Contributions: Conceptualization, funding acquisition, project administration, supervision, writing—review and editing, X.C.; data curation, visualization, writing—original draft, J.L.; investigation, methodology, writing—original draft, W.H.; formal analysis, validation, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Funds of Social Science, grant number 21BXW076, the Natural Science Foundation of China project, grant number 61602518, the Philosophy and Social Science Research Project of Hubei Provincial Department of Education, grant number 20G026 and the Innovation Research of Young Teachers of Central Universities in 2021, grant number 2722021BZ040.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hunt, J.D. Image as a factor in tourism development. *J. Travel Res.* **1975**, *13*, 1–7. [\[CrossRef\]](#)
- Crompton, J.L. An Assessment of the Image of Mexico as a Vacation Destination and the Influence of Geographical Location Upon That Image. *J. Travel Res.* **1979**, *17*, 18–23. [\[CrossRef\]](#)
- Steve, P. Destination image analysis—a review of 142 papers from 1973 to 2000—ScienceDirect. *Tour. Manag.* **2002**, *23*, 541–549.
- Schroeder, H.W.; Louviere, J. Stated Choice Models for Predicting The Impact Of User Fees at Public Recreation Sites. *J. Leis. Res.* **1999**, *31*, 300–324. [\[CrossRef\]](#)
- Howell, R. Vacationscape: Designing tourist regions. *Ann. Tour. Res.* **1989**, *16*, 445–447. [\[CrossRef\]](#)
- Clare, A.G. Book Review: Vacationscape: Designing Tourist Regions. *J. Travel Res.* **1973**, *11*, 24.
- Hong, S.; Kim, J.; Jang, H.; Lee, S. The roles of categorization, affective image and constraints on destination choice: An application of the NMNL model. *Tour. Manag.* **2005**, *27*, 750–761. [\[CrossRef\]](#)
- Martens, H.M.; Reiser, D. Analysing the image of Abu Dhabi and Dubai as tourism destinations—The perception of first-time visitors from Germany. *Tour. Hosp. Res.* **2019**, *19*, 54–64. [\[CrossRef\]](#)
- Cassia, F.; Vigolo, V.; Ugolini, M.M.; Baratta, R. Exploring city image: Residents’ versus tourists’ perceptions. *TQM J.* **2018**, *30*, 476–489. [\[CrossRef\]](#)
- Kim, H.; Lee, S. Impacts of city personality and image on revisit intention. *Int. J. Tour. Cities* **2015**, *1*, 50–69. [\[CrossRef\]](#)
- Zaid, A.; Mithat, Z.D.; Fusun, I.D.; Parvin, M. Understanding destination image from the perspective of Western travel bloggers: The case of Istanbul. *In. J. Cult. Tour. Hosp. Res.* **2018**, *12*, 198–212.
- Clarke, H.; Hassanien, A. An Evaluation of Toronto’s Destination Image Through Tourist Generated Content on Twitter. *Int. J. Cust. Relat. Mark. Manag. (IJCRMM)* **2020**, *11*, 1–16. [\[CrossRef\]](#)
- Yin, G.; Jianlin, Z. Research on tourists’s perception of landscape image of Chongqing Expo Park Based on network data. *J. Southwest Normal Univ. (Nat. Sci. Ed.)* **2021**, *46*, 132–141.
- Shuang, D.; Qiuju, W. Identification of tourist perception dimension based on LDA: Research framework and empirical research—Taking National Mine Park as an example. *J. Beijing Union Univ. (Human. Soc. Sci. Ed.)* **2019**, *17*, 42–49.
- Chi, M.; Pan, M.; Wang, W. Cross platform comparative study of shared accommodation and hotel user comments: Thematic social network and emotion analysis based on LDA. *Library Inf. Work* **2021**, *65*, 107–116.
- Hui, L.; Liting, W. Research on microblog hot topic discovery based on word item heat. *Inf. Sci.* **2018**, *36*, 45–50.
- Huosong, X.; Baoguo, L.; Pei, Y. Topic extraction of online news comments based on improved k-means clustering. *J. Inf.* **2016**, *35*, 55–65.
- Xiaobo, T.; Xin, Q. Research on topic oriented high quality comment mining model. *Modern Library Inf. Technol.* **2015**, *7*, 104–112.
- Haili, T.; Xiaobo, T.; Li, X. Research on user requirements mining model based on online comments. *J. Inf. Technol.* **2015**, *34*, 1088–1097.
- Yinhao, W.; Narissa; Hui, L. Representative negative comment extraction based on Gaussian LDA and spectral clustering. *Inf. Sci.* **2021**, *39*, 136–142.
- You, M.; Kun, Y.; Zichen, Z.; Xiaoyi, W.; Jianbin, G. Social media data extraction based on knowledge map and LDA model. *J. East China Normal Univ. (Nat. Sci. J.)* **2018**, *5*, 183–194.
- Linlu, L.; Yating, L. Research on tourism information recommendation based on theme mining and expression of travel notes. *Modern Intell.* **2017**, *37*, 61–67.
- Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res. Arch.* **2003**, *3*, 993–1022.
- Yanan, H.; Jianwei, L.; Xionglin, L. Overview of probabilistic topic models. *J. Comp. Sci.* **2021**, *44*, 1095–1139.
- Xiwei, W.; Yunfei, X.; Dan, Z.; Jiaying, L. The Study of Network Public Opinion Dissemination with Social Network Analysis Under the Mobile Environment: A Case of “Haze” in Sina Micro-blog. *Library Inf. Work* **2015**, *59*, 14–22.
- Liu, B. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies. Synth. Lect. Human Lang. Technol.* **2012**, *5*, 1–167. [\[CrossRef\]](#)
- Kai, L.; Xiaodong, Y. Research on patent technology integration trend based on LDA theme model and social network—Taking joint robot as an example. *J. Inf.* **2021**, *40*, 89–97.
- Lin, Z.; Iris, R.; Roger, M. Extension is combined SNA and LDA methods to understand adverse medical events: Doctor and nurse perspectives. *Int. J. Risk Saf. Med.* **2020**, *31*, 221–246.

29. Hongwei, W.; Song, G.; Kai, L. Research on online news hotspot identification based on LDA and SNA. *J. Inf. Technol.* **2016**, *35*, 10.
30. Guanghui, Y.; Tong, X. Dynamic City Profile Based on Evolutionary Analysis. *Data Anal. Knowl. Disc.* **2020**, *4*, 100–110.
31. Yue, H.; Tingting, Z. Research on haze public opinion based on microblog emotion analysis and social network analysis. *Inf. Sci.* **2018**, *36*, 91–97.
32. Guizhen, M.; Xia, P. Topic mining and emotion analysis of tourist reviews based on big data of hotel reviews—Taking five-star hotels in Beijing as an example. *J. Beijing Union Univ.* **2021**, *35*, 58–68.
33. Shafqat, W.; Byun, Y.C. A Recommendation Mechanism for Under-Emphasized Tourist Spots Using Topic Modeling and Sentiment Analysis. *Sustainability* **2019**, *12*, 320. [[CrossRef](#)]
34. Kai, B.; Xin, W. Update, review and research prospect of the national standard classification, investigation and evaluation of tourism resources. *J. Nat. Resour.* **2020**, *35*, 1525–1540.
35. Peng, G.; Yuefen, W. Research on the method of determining the optimal topic number of LDA topic model in scientific and technological information analysis. *Modern Library Inf. Technol.* **2016**, *32*, 42–50.
36. Huang, Z.; Xu, W.; Yu, K. Bidirectional LSTM-CRF Models for Sequence Tagging. *arXiv* **2015**, arXiv:1508.01991.
37. Lingkang, X. *Research on Evaluation of Wuhan Tourism Service Quality*; South-Central University for Nationalities: Wuhan, China, 2016.
38. Yu, F.; Xiang, H. Research on Image Perception of Tourist Destinations in Wuhan—Based on Web Text Analysis Method. *J. Cent. China Normal Univ. (Nat. Sci. Ed.)* **2017**, *51*, 709–714. [[CrossRef](#)]