

Analysis of Health and Wellness Tourism in Shangluo City Based on Text Mining

Siming Ma, Yangwei Wei, Xinyu Yang, Wenbo Wang

School of Mathematics and Computer Applications, Shangluo University, Shangluo, Shaanxi, China

Abstract: To fully leverage the tourism resources and geographical advantages of Shangluo City and to propel the high-quality development of its health and wellness tourism industry, this study focuses on the health and wellness tourism market of Shangluo City as its primary research subject. By gathering relevant reviews from prominent domestic travel platforms such as Ctrip and utilizing advanced big data technologies, including semantic network analysis and data analysis methods, the research conducts a thorough examination and investigation into the health and wellness tourism market and image perception evaluations in Shangluo City. The Jinsi Grand Canyon scenic area in Shangluo City serves as a case study, where the ROSTCM6.0 analysis tool is employed to evaluate the image perception of the tourism market through semantic network analysis, high-frequency feature words, and sentiment analysis. The findings indicate that while the natural beauty of the scenic area is highly praised by tourists, there are significant shortcomings in sanitation and personnel management within the area. In conclusion, this study proposes several recommendations for the advancement of wellness tourism, including improving the quality of tourism services and providing personalized travel experiences.

Keywords: Jinsi Grand Canyon Scenic Area; Health and Wellness Tourism; ROSTCM6.0; Tourists' Cognitive Images; Tourists' Emotional Image

1. Introduction

The "China Domestic Tourism Development Report (2023-2024)" issued by the China Tourism Research Institute in collaboration with Global Travel News reveals a notable upswing in both tourist numbers and tourism revenue during major holidays in 2023,

compared to the previous year. This growth trajectory indicates a robust start and a steady expansion throughout the year. A significant highlight was the 8-day combined "Mid-Autumn Festival and National Day" holiday, during which both the volume of domestic travelers and the corresponding tourism revenue not only rebounded to pre-pandemic 2019 levels but also experienced a marginal increase. Specifically, the number of travelers during this period rose by 4.1% compared to 2019, while tourism revenue saw a 1.5% increase from the same benchmark year [1,2].

Since the Shangluo Municipal Government unveiled its strategic initiative to establish the city as the "Capital of Health and Wellness in China" in July 2021, Shangluo has been diligently refining its brand identity under the banner of "22°C Shangluo · Capital of Health and Wellness in China." This initiative aims to position Shangluo as a premier hub for the health and wellness industry in China and a sought-after destination for health and wellness tourism in the Qinling Mountains.

To realize these ambitious goals, this study leverages a wealth of data, including online reviews, to assess the allure of Shangluo's health and wellness tourism market. By employing big data analytics, the research identifies the primary tourism activities and distinctive features of the region. It seeks to delineate the core competitive advantages and target market positioning for Shangluo's health and wellness tourism development. Furthermore, the study focuses on establishing a core brand for Shangluo's health and wellness tourism, optimizing the allocation of developmental resources, and enhancing the supply of tourism products. By transforming digital innovation into a new driving force for the tourism sector, this paper aims to catalyze the advancement and transformation of Shangluo's health and wellness tourism industry.

2. Materials and Methods

2.1 Introduction to Jinsi Grand Canyon Scenic Area

Jinsi Canyon National Forest Park, commonly referred to as "Jinsi Canyon," is a prestigious national AAAAA-level tourist attraction, recognized as both a national forest park and a national geological park. Esteemed as the "City of Canyon," "China's Marvelous Canyon," "China's Most Beautiful Grand Canyon," and "the most stunning Grand Canyon in China," this scenic area is situated in Shangnan County, the southeastern gateway of Shaanxi Province. The park is a botanical crossroads where northern and southern flora converge, featuring a diverse array of temperate and warm temperate deciduous broad-leaved species, as well as subtropical evergreen and deciduous broad-leaved species, with an impressive forest coverage rate of 98%. In August 2009, it was designated a national geological park by the Ministry of Land and Resources, marking it as the second such park in Shaanxi Province, China.

Nestled at the southern foot of the Qinling Mountains and adjacent to the northern slope of the Bashan Mountains, Jinsi Canyon lies within the middle reaches of the Danjiang River system in the Yangtze River Basin. The landscape is characterized by its undulating terrain, with a harmonious alternation of valleys and ridges, predominantly composed of low mountains and hilly landforms. It is celebrated as a "canyon wonder and ecological kingdom." The park achieved provincial-level forest park status on July 1, 2002, and was elevated to a national forest park in December of the same year [3].

2.2 Data Preprocessing

The Octopus Collector was employed to gather comments and information pertinent to health and wellness tourism in Shangluo from prominent Chinese travel platforms and forums, including Ctrip. To enhance the precision of the data sampling, a manual curation process was undertaken during the extraction of subpages related to Jinsi Canyon and other scenic spots within Shangluo City. This involved the removal of certain user comments that were deemed irrelevant, such as those containing nonsensical symbols or content unrelated to the

scenic areas. Additionally, specific terms like "Jinsi Canyon" and "Heilong Canyon" were incorporated into the segmentation lexicon for reanalysis. The collected data underwent processing through ROSTCM6, which facilitated segmentation, the construction of a semantic network, high-frequency word analysis, and sentiment analysis. The outcome of this process yielded a refined dataset, comprising a total of 2072 entries for comments specifically pertaining to Jinsi Canyon [4-6].

2.3 Research Methods and Ideas

Content analysis [7,8] is a research methodology designed to measure variables and systematically, objectively, and quantitatively describe explicit content. It serves as a technique for the objective analysis of such content. ROSTCM6.0 is a content mining system software primarily utilized for analyzing key modules including word frequency analysis, semantic network analysis, and sentiment analysis, thereby facilitating text mining and analysis. In this study, ROSTCM6.0 software is employed to examine the Jinsi Canyon scenic area through its word frequency, semantic network, and sentiment analysis capabilities. Initially, ROSTCM6.0 is used to extract high-frequency feature words and their corresponding frequencies from the sample data related to the Jinsi Canyon Scenic Area. This is followed by a semantic network analysis based on the word frequency results to glean insights into tourists' perceptions of the Jinsi Canyon Scenic Area's image. Subsequently, the sentiment analysis function within ROSTCM6.0 is applied to the samples to identify the core experiential sentiments of tourists. Finally, by synthesizing the analysis results, an overall perception of the Jinsi Canyon scenic area as experienced by tourists is derived.

3. Result Analysis

3.1 Analysis of Tourists' Cognitive Images

3.1.1 Analysis of high-frequency characteristic words

The analysis of high-frequency feature words to some extent reflects various fields of cognitive image of tourist destinations. ROSTCM6.0 was used to partially remove text from 2072 comments extracted from the Jinsi

Canyon comment section on Ctrip website, resulting in high-frequency vocabulary for image perception (as shown in Table 1). From Table 1, it can be seen that "White Dragon Canyon" and "Black Dragon Canyon", as scenic spots in Jinsi Canyon Scenic Area, have received widespread attention from tourists. These attractions leave a deep impression in the minds of tourists and have become the attractions of Jinsi Canyon Scenic Area. "Kilometers", "hours", "convenience", and "highways" reflect tourists' emphasis on

distance planning for travel. The frequency of words such as "park", "nature", and "ecology" reflects tourists' affirmation of the natural scenery of Jinsi Canyon, while the frequency of words such as "student" and "child" is relatively low, reflecting the lack of attractiveness of Jinsi Canyon for young people's travel. The high frequency of words such as "admission ticket" indicates that tourists in the scenic area have more suggestions for admission tickets.

Table 1. High frequency Vocabulary of Tourist Image Perception in Online Text of Jinsi Canyon Scenic Area

Serial number	high-frequency word	part of speech	word frequency	Serial number	high-frequency word	part of speech	word frequency	Serial number	high-frequency word	part of speech	word frequency
1	Jinsi Canyon	noun	144	20	Shangnan County	noun	17	39	feel	verb	11
2	hours	noun	91	21	whole journey	noun	16	40	White Dragon Canyon	noun	11
3	landscape	noun	81	22	go up	verb	15	41	pay attention to	verb	11
4	Kilometers	noun	65	23	beautiful	adjective	15	42	landscape	noun	10
5	place	noun	50	24	Scenery	noun	15	43	relaxed	adjective	10
6	staff	noun	47	25	Play Things	noun	14	44	next time	colloquialism	9
7	suggestion	verb	44	26	take photos	verb	14	45	fare	noun	9
8	country	noun	39	27	probably	colloquialism	14	46	mood	noun	9
9	convenience	adjective	38	28	export	noun	14	47	afternoon	noun	9
10	go up the mountain	verb	34	29	walk	verb	13	48	super	adjective	9
11	park	noun	31	30	downstream	noun	12	49	entrance	noun	8
12	nature	noun	31	31	online	noun	12	50	characteristic	noun	8
13	admission fee	noun	30	32	eat	verb	12	51	up and down	adjective	8
14	air	noun	29	33	organic	noun	12	52	southeast	adjective	8
15	second	noun	20	34	minute	noun	12	53	extremely	adjective	8
16	highways	noun	20	35	area	noun	12	54	guide	noun	8
17	Black Dragon Canyon	noun	18	36	return	verb	12	55	whole	adjective	8
18	ecology	noun	18	37	child	noun	12	56	Water Show	adjective	8
19	set out	verb	17	38	entire	adjective	11	57	student	noun	8

3.1.2 Part of speech analysis of high-frequency feature words

The 57 high-frequency words extracted have parts of speech including nouns, adjectives, verbs, adverbs, and phrases (Table 1). Among them, nouns account for about 63.15%, verbs account for 15.78%, and adjectives account for about 17.5%. Nouns mainly refer to scenery, facilities, etc. Adjectives mainly refer to tourists' personal experiences while playing, such as convenience, relaxation, beauty, etc. Verbs mainly refer to tourists' travel behaviors and activities, with the most representative being photography, walking, etc. Adverbs and phrases mainly refer to tourists' willingness to revisit, such as "next time", etc. From the analysis of part of speech, it can be concluded that the main carriers of image perception in

Jinsi Canyon Scenic Area are scenic spots, scenery, infrastructure, and other major aspects.

3.1.3 Semantic network analysis

Although high-frequency vocabulary provides a certain reflection of the primary aspects of tourism image perception, it falls short of intuitively revealing the intricate connections and underlying structural relationships among the terms. Semantic network diagrams, on the other hand, offer a clearer visualization of the hierarchical relationships between entries, enabling a more effective identification of the correlations among diverse words [9,10].

The semantic network analysis of visitor comments on Jinsi Canyon Scenic Area reveals a distinctive three-tiered structure characterized by "core -> sub-core -> periphery" relationships. This structural configuration

effectively illustrates the varying degrees of attractiveness among different elements within the tourist destination, offering valuable insights for the strategic planning and comprehensive management of the scenic area. Beyond the predominant terms frequently mentioned by online commentators, secondary-level descriptors such as "country", "nature", "hour", and "scenery" emerge with significant frequency. The strong semantic correlations among these terms underscore Jinsi Canyon's status as a national 5A-level scenic area, which generates substantial online engagement and is renowned for its exceptional natural landscapes. The secondary lexical cluster, including terms like "hour", "mountain climbing", "entire journey", and "highway", demonstrates meaningful semantic connections that reflect tourists' overall satisfaction with their travel experiences. Furthermore, the extensive network connections among terms such as "nature", "ecology", "scenery", and

"park" strongly indicate the scenic area's predominant orientation towards natural ecological characteristics, highlighting its distinctive environmental features and visitor perceptions.

3.2 Analysis of Tourists' Emotional Image in Jinsi Canyon Scenic Area

According to the sentiment analysis of 512 randomly selected tourist comments on Jinsi Canyon Scenic Area, positive emotions constituted 77.15%, neutral emotions made up 1.17%, and negative emotions represented 21.68%. It suggests that the majority of visitors hold a favorable view of the Jinsi Canyon scenic area.

Based on the outcomes of sentiment analysis, a further segmentation of emotions was conducted. The statistical breakdown of positive emotions is presented in Table 2, while the segmented statistical results of negative emotions are displayed in Table 3.

Table 2. The Statistical Results of Positive Emotions

Level	Scope	Number of Comments	Proportion
General	(5,15]	121	23.63%
Moderate	(15,25]	84	16.41%
Height	(25,+∞)	190	37.11%

Table 3. The Statistical Results of Negative Emotions

Level	Scope	Number of Comments	Proportion
General	[-15,5)	32	6.25%
Moderate	[-25, -15)	7	1.37%
Height	(-∞, -25)	15	2.93%

From further analysis results, it can be concluded that 37.11% of positive reviews from tourists are highly positive, while only 2.93% of negative reviews are highly negative. For negative evaluations, they mostly focus on aspects such as "hygiene", "diet", "admission tickets", "management", and "money laundering". Positive evaluations focus on aspects such as "natural landscapes" and "scenery". It reflects that although Jinsi Canyon Scenic Area has beautiful scenery, it lacks personnel management and scenic area development, and its attractiveness to young tourists and students is insufficient.

4. Conclusions and Recommendations

4.1 Conclusion

(1) From the analysis of high-frequency words, it can be concluded that tourists have more opinions on ticket prices, mostly have a

positive attitude towards natural scenery, and have more evaluations on travel modes and times, as well as the duration of sightseeing in scenic spots.

(2) From the analysis of semantic network diagrams, it can be concluded that natural landscapes are more likely to attract tourists' attention in tourism landscapes. After visiting, many tourists will give suggestions to the scenic area, which are mostly suggestions for the process of climbing up the mountain and overall suggestions for the "Jinsi Canyon" scenic area. Jinsi Canyon is the center of the semantic network, with the highest frequency of mention and correlation, followed by words such as "country", "nature", "hour", and "scenery", indicating that people's evaluations of Jinsi Canyon mainly focus on the natural landscape, reflecting the significant attraction of natural landscapes to tourists.

(3) From emotional analysis, it can be

concluded that tourists have more positive evaluations of the Jinsi Canyon scenic area, mainly reflected in the natural scenery, while negative evaluations are mainly reflected in aspects such as "tickets", "personnel management", and "hygiene".

4.2 Suggestion

Through a series of analyses, it is concluded that the natural scenery of Jinsi Canyon Scenic Area is excellent, but there are many problems in personnel management, which leads to a high number of hygiene issues. Based on the content of the paper, the following suggestions are provided:

- (1) It is suggested that scenic spots should strengthen personnel management, handle the problem of dirty and disorderly hygiene in scenic spots, and improve the service level for tourists, continuously meeting personalized user needs.
- (2) Enhance tourism resources by intensifying the promotion of scenic spots and incorporating more tourism elements that appeal to students and young tourists, complementing the natural scenery.
- (3) Scenic spots should continuously open up channels for tourist feedback and adopt more suggestions from netizens to thoroughly solve the repeated user feedback and negative reviews.

Acknowledgements

This study was supported by the National College Students' Innovation and Entrepreneurship Training Program (Grant No. 202311396008), the Science and Technology Plan Project of Shangluo City (Grant No. 2022-J-0009), and the Natural Science Research Project of Shangluo University (Grant No. 24KYZX07).

References

- [1] Zhu Q, Wang Z, Umer M. The Suitability of Developing Ecotourism in the Shanxi Area of Taihangshan National Park, a Candidate Area for National Parks in China. *Sustainability*, 2025, 17(3): 841-841.
- [2] Shufeng C, Lee C, Rahim A S A. Does urban tourism development impact urban housing prices?. *International Journal of Housing Markets and Analysis*, 2025, 18(1): 5-24.
- [3] Yang H X. An Analysis of the Development of Tourism in the Mountainous Area and the Ecological Environment Protection – Take Jinsixia Park as an Example. *Applied Mechanics and Materials*, 2013, 2301(295-298): 2233-2237.
- [4] Peter A D, Decui L, Yinrunjie Z, et al. Service quality in football tourism: an evaluation model based on online reviews and data envelopment analysis with linguistic distribution assessments. *Annals of operations research*, 2022, 325(1): 31-34.
- [5] Ramadhani P D, Alamsyah A, Febrianta Y M, et al. Large-Scale Cross-Cultural Tourism Analytics: Integrating Transformer-Based Text Mining and Network Analysis. *Computers*, 2025, 14(1): 27-27.
- [6] Tang R, Moon J, Heo R G, et al. Exploring the knowledge structure and potential research areas of sustainable tourism in sustainable development: Based on text mining and semantic network analysis. *Sustainable Development*, 2023, 32(4): 3037-3054.
- [7] Vila D T, Darcy S. Beyond technical website compliance: Identifying and assessing accessible tourism value chain information content on national tourism organisation websites. *Tourism Management Perspectives*, 2025, 55101332-101332.
- [8] Guo X, Chen S, Guo Y. Advancing multimodal teaching: a bibliometric and content analysis of trends, influences, and future directions. *Humanities and Social Sciences Communications*, 2024, 11(1): 1718-1718.
- [9] Lebakula V, Porter B, Richardson S M, et al. A structural equation modeling approach to leveraging the power of extant sentiment analysis tools. *Journal of Computational Social Science*, 2024, 8(1): 3-3.
- [10] Kumar S A. Developing and testing a tool to classify sentiment analysis. *International Journal of Data Science and Big Data Analytics*, 2021, 1(2): 23-30.