

Original Research Article

## Research on theme mining of tourists' perception in intangible cultural heritage homestay scene based on LDA model

Wanlian Li, Zhangjie Zhu

*School of Business Administration, Anhui University of Finance and Economics, Bengbu, Anhui, 233000, China*

**Abstract:** The integration of intangible cultural heritage and the homestay industry is conducive to the inheritance, protection and practice of intangible cultural heritage, and can also promote destination culture and enhance tourists' cultural identity. Based on the scene theory, this study constructs a scene perception framework from three dimensions: material, society, and culture. It uses Python web scraping, LDA topic analysis, and SnowNLP sentiment analysis to extract themes and conduct sentiment analysis on online reviews of intangible cultural heritage homestays. The results show that tourists' perception is concentrated on 6 themes: environment, cultural experience, service, customer group, facilities and comfort. Their emotions are generally positive, and negative emotions are mostly related to facilities. Based on this, it is proposed to improve the scene of intangible cultural heritage homestays from aspects such as facility optimization, service enhancement, and highlighting the characteristics of intangible cultural heritage.

**Keywords:** intangible cultural heritage homestay; scene; visitor perception; LDA models; sentiment analysis

### 1. Introduction

Intangible cultural heritage is an important carrier of China's fine traditional culture, embodying historical memory and national spirit. It is a key element for cultural and tourism destinations to build cultural core competitiveness (Chen, 2020). In recent years, China has introduced policies to encourage the in-depth integration of intangible cultural heritage resources with tourism homestays, providing policy support for the cultural transformation and upgrading of tourism homestays. Against this backdrop, intangible cultural heritage homestays have rapidly risen as a new form of cultural and tourism integration, and many places have formed regional characteristic clusters. However, industry practice has highlighted problems such as fierce homogenized competition, insufficient exploration of cultural connotations, and formalistic integration of intangible cultural heritage elements. Therefore, breaking through the superficial predicament of the integration of intangible cultural heritage and homestays and transforming intangible cultural heritage resources into scene-based elements that can be experienced, consumed and disseminated has become a key issue for promoting high-quality development in this field. Academic research on tourism homestays mainly focuses on aspects such as local cultural integration (Wu & Tian, 2022), community economic synergy (Li & Li, 2019), tourists' perceived experience (Yu & Lyu, 2024), cultural identity and emotional value experience (Huang, et al., 2021). Research on intangible cultural heritage tourism focuses on issues such as resource protection and development (Qiao, 2017), comprehensive development effects (Xiao, et al., 2024), and co-creation of tourists' perception and experience (Song & Wang, 2024). The scene theory was proposed by the New Chicago School. With authenticity, legitimacy and drama as the core dimensions, it is dedicated to explaining the interaction mechanism between cultural elements and spatial consumption (Silver & Clark, 2019). At present, this theory has gradually expanded from urban research to fields such as rural cultural tourism and intangible cultural heritage tourism, demonstrating strong theoretical applicability. Based on this theory, intangible cultural heritage homestays can be defined as a composite experience scene that carries the content of intangible cultural heritage, covering three core dimensions: physical space, social behavior and cultural symbols. However, at present, the academic circle lacks in-depth exploration of a series of issues regarding the development of intangible cultural heritage homestays, especially the exploration of their scene value and the perception mechanism of tourists have not been given sufficient attention. In view of this, this study takes the scene of intangible cultural heritage homestays as the

object, builds a framework based on scene theory, and uses the LDA topic model to mine the theme of tourists' perception of comments. This research not only makes up for the insufficiency of specialized studies on intangible cultural heritage homestays, but also provides theoretical and practical guidance for optimizing scene design, deepening cultural integration and enhancing the tourist experience, which is of academic and practical significance.

## **2. Research methods and data sources**

### **2.1. Data acquisition and preprocessing**

This study selected online reviews from Ctrip and Tongcheng as the data collection subjects because they are leading in the domestic online accommodation booking market, with a large volume and wide coverage of review data, which can reflect tourists' feedback on the accommodation experience of intangible cultural heritage homestays. Using a data capture program developed in Python, review information was collected on both platforms with intangible cultural heritage homestays as the keyword. The collection deadline was April 11, 2025, and the content included user nicknames, review texts, and review times. To ensure data reliability, preprocess the collected data: (1) Eliminate worthless data such as duplicate comments, invalid characters, and emojis; (2) Use Python's jieba word segmentation tool to remove stop words and segment them. Finally filter out 4,860 valid comments.

### **2.2. LDA topic model**

The latent Dirichlet distribution (LDA) is a probability-based topic model used to reveal the latent topic structure of text. The basic assumption is that each document is the probability distribution of the topic, and each topic is the probability distribution of the vocabulary. LDA excels in identifying large-scale document topics and is widely used in areas such as user review analysis and hotspot tracking. This study extracted key topics from online reviews of visitors to intangible cultural heritage homestays through steps such as data collection, text preprocessing, construction of LDA models, and determination of the optimal number of topics. Specifically, using Python's jieba word segmentation tool to segment and remove stop words, invoking the gensim library to build the LDA model, calculating perplexity and cohenrence metrics to determine the optimal number of topics, and achieving effective classification of text data.

### **2.3. SnowNLP sentiment analysis**

SnowNLP is an open-source Chinese text processing library for Python. Its main functions include Chinese word segmentation, sentiment analysis, and text classification. Its core sentiment analysis capabilities, based on Naive Bayes principles and built-in training sets, are superior in performance and suitable for analyzing the sentiment tendencies of user comments. This study uses SnowNLP to analyze online reviews of visitors to intangible cultural heritage homestays by splitting text into sentences, giving sentiment ratings, and assigning positive, neutral, or negative labels to identify overall sentiment tendencies.

## **3. Research findings**

### **3.1. LDA topic model analysis**

#### **3.1.1. Determine the optimal number of topics**

In this study, the LDA topic model was used to conduct topic clustering analysis on preprocessed intangible cultural heritage homestay review data to reveal related topics and characteristic words. In the LDA model, determining the optimal number of topics  $K$  is key to ensuring the quality of topic extraction. In this study, the optimal number of topics was determined by calculating perplexity and coherence metrics. Perplexity reflects the degree to which the model fits the dataset, and the lower the value, the better the clustering effect; Coherence is used to assess the correlation of feature words within a topic, and the higher the value, the better the correlation of feature words within the topic. The perplexity curve (Figure 1) shows that as the number of topics increases, perplexity decreases and model fitting improves. However, when the number of topics reaches 6, the downward trend flattens, and further increasing the number of topics has limited effect on model performance improvement. The coherence curve (Figure 2) shows that when the number of topics is 6, the consistency score is better,

followed by large fluctuations and reduced stability. Taking into account the trends of the two indicators, it is reasonable to choose 6 as the optimal number of topics, which ensures the model's discriminatory ability while maintaining good interpretability.

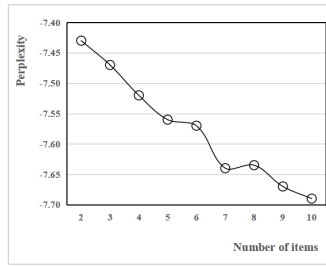


Figure 1. Changes in perplexity.

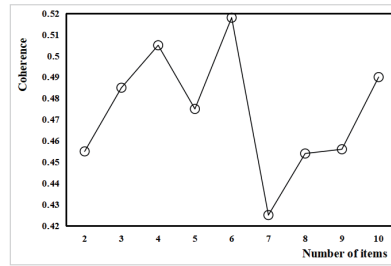


Figure 2. Changes in coherence.

3.1.2. Analysis of topic outcomes

Based on the results of the "theme-word" clustering analysis, this study summarized and identified the theme of the intangible cultural heritage homestay scene, and explored the composition of the scene element theme from the three dimensions of physical, social, and cultural scenes. The specific analysis results are shown in Table 1.

Table1. Core dimensions of intangible cultural heritage homestays' scene and their thematic composition.

| Scene dimensions         | Include topics      | Topic feature words  |
|--------------------------|---------------------|--|
| Physical scene dimension | Environment         | Room, clean, scenic area, Confucius Temple, hygienic, location               |
|                          | Facilities          | Toilet, things, supplies, very small, bed sheets, bedroom                    |
|                          | Comfort experience  | Sleeping, disadvantages, not so good, merits, imperfections                  |
| Social scene dimensions  | Customer group      | Kids, friends, nice, couples, outings, parent-child                          |
|                          | service             | Little sister, help, detail, intangible cultural heritage, time travel, feel |
| Cultural scene dimension | Cultural experience | Ancient town, evening, special, Satisfying, Luoyi, Cross Street              |

3.1.2.1. Thematic analysis of the physical scene dimension

The physical scene dimension of the intangible cultural heritage homestay explored in this study mainly focuses on the objective environmental space of the homestay, covering three themes: the environment, facilities and comfort experience of the intangible cultural heritage homestay. First, the thematic feature words for the environment of the intangible cultural heritage homestay include "room", "clean", "scenic area", "Confucius Temple", "location", "hygienic", etc. Among them, "room", "clean" and "hygienic" emphasize tourists' emphasis on the cleanliness and quality of the internal environment of the homestay, constituting the core material scene inside the intangible cultural heritage homestay. "Scenic area", "Confucius Temple" and "Location" show the importance of the surrounding scenic area and geographical location as external material scenes. Comments such as "The room is very clean and very close to the scenic spot" and "the environment is nice and convenient in the Confucius Temple scenic area" reflect that the quality material scene of the intangible cultural heritage homestay has a significant impact on tourists' living experience. Secondly, in terms of the facilities of the intangible cultural heritage homestay, facilities such as "toilets", "supplies", "bed sheets", "king-sized beds" are key components, and their quality affects the overall evaluation of tourists. Comments such as "The facilities are complete, the hygiene is very clean, disposable toiletries are provided for free, and disposable bath towels are also provided for free" "The bed is very comfortable, the toilet is very clean" "The bed sheets and utensils are very clean, no odor or stains" indicate that complete facilities and good hygiene conditions can create a comfortable material scene. But there are also complaints from visitors about problems such as "the battery doesn't charge, the room is a bit small, and the Windows are very small", indicating that inadequate infrastructure can affect visitor satisfaction. More spacious rooms and better facilities would better meet the needs of tourists. Finally, in terms of comfort experience, tourists' comfort experience is influenced by a combination of the physical environment. The frequent appearance of words like "disadvantages", "merits", "not so good", "imperfections" indicates that while tourists recognize the intangible cultural heritage homestay, they also point out the areas for improvement, which is the result of the combined effect of various elements, such as "the sound insulation is not very good, 9 points recommendation" "merits: all video VIP on TV, very cool; It's not chaotic and quite quiet. Disadvantage: The

room height is too low; There is a wall in front of the bathtub, making the room seem small. Overall impression: Good, worth recommending". Comments such as "Not very good experience" suggest that the lack of physical scenes can affect the guest's stay experience and accommodation satisfaction.

### **3.1.2.2. Thematic analysis of the social scene dimension**

In this study, the core element of the social scene dimension of intangible cultural heritage homestays is the social space provided for tourists, covering two themes: intangible cultural heritage homestay services and customer groups. First, from the comments on the service theme of the intangible cultural heritage homestay, high-frequency words such as "lady", "help", and "detail" were extracted to reveal the positive interaction between the service staff and the tourists, and to construct a warm and friendly social interaction scene. Such as "Excellent service attitude, actively helping to store luggage, and giving a woodcarving experience when leaving" and "satisfactory service, and additional watermarked fans for tourists to stamp themselves. The front desk staff kindly returned the Piggy doll that the child had carelessly left behind", highlighting the role of the service staff. They provided thoughtful service and gave away souvenirs with traditional cultural elements, which enhanced guests' satisfaction with their stay. Words like "intangible cultural heritage" and "time travel" indicate that the services of intangible cultural heritage activities are well-received by tourists, such as "the homestay arranges iron flower making, Huizhou ink lessons and mooncake making activities in the evening" "The homestay offers a good experience of intangible cultural heritage handicrafts and afternoon tea" "There are Huizhou ink lessons at the homestay after 8 p.m., free, experience ink making, it's a magical feeling to be so close to intangible cultural heritage". Tourists interact and get in touch with the local culture during their participation in intangible cultural heritage activities, build social scenes with cultural connotations, and the quality of the activity experience affects the image and influence of the homestay. Secondly, the theme words of the intangible cultural heritage homestay guests include "children", "friends", "couples", "parent-child", etc., reflecting the diversity of the tourist groups and the differences in their demands. Parent-child families hope that their children will be exposed to traditional culture, such as "This homestay is really family-friendly, there are so many activities, basically the kids go to play with their own friends" "The traditional culture lessons in the homestay can plant tiny seeds in the young hearts of children, hoping that through the enlightenment of the cultural lessons they can actively feel, understand and inherit Chinese traditional culture"; Couples and friends value the unique decoration style and atmosphere, such as "The hotel decoration design has its own unique style and is very suitable for couples to travel or to stay with friends on vacation". Therefore, intangible cultural heritage homestays need to create diverse and targeted social interaction scenarios based on the needs of different customer groups to meet the expectations of tourists.

### **3.1.2.3. Thematic analysis of the cultural scene dimension**

In this study, the cultural scene dimension of intangible cultural heritage homestays is dramatically associated with scene theory, referring to the experience process of tourists in the cultural scene of intangible cultural heritage homestays. It can be summarized as the theme of "cultural experience", with words like "ancient city" and "night" appearing frequently. Comments show that tourists choose intangible cultural heritage homestays not only for rest and accommodation, but also for the experience of traditional culture, as some tourists commented: "Self-driving to Luoyang, I chose this homestay to experience the culture of the Sui and Tang Dynasties" "In the ancient city of Luoyi, the view at night is very nice and you can experience Hanfu" "Experiencing the culture of Huizhou ink is a very special experience" "The folk display in the store is particularly interesting and the store decoration is full of exotic colors, it is an unforgettable cultural experience". Intangible cultural heritage homestays are accommodation places that carry intangible cultural heritage. The space construction incorporates folk characteristics and traditional cultural elements, such as ancient architecture, traditional handicraft decorations, etc., which constitute the core elements of the cultural scene. In the intangible cultural heritage scene, tourists gain an intuitive understanding of different cultural connotations through visits, deepen their knowledge of intangible cultural heritage, and experience local folk customs.

## **3.2. Affective analysis research**

### **3.2.1. Analysis of overall emotional tendencies**

Emotional analysis of online reviews of intangible cultural heritage homestays can help understand tourists' emotional tendencies. With the help of the SnowNLP sentiment analyzer in Python to calculate the sentiment

value of the comments, the result is an interval value of [0,1]. The higher the value, the more positive the sentiment tendency; the lower the value, the more negative the sentiment tendency. In this study, the distribution of emotional tendencies of tourists visiting intangible cultural heritage homestays is as follows: positive emotions account for the highest proportion, reaching 55.5%, indicating that the overall attitude of visitors is positive; neutral emotions accounted for 39.8%, indicating that some tourists' evaluations were objective and neutral; negative emotions accounted for 4.7%, indicating that a small number of tourists held negative views. As shown in Figure 3, the emotional value is around 0.6, with positive evaluations being the most frequent; extreme positive comments with an emotion value greater than 0.8 followed, there are fewer extremely negative comments with a sentiment value of less than 0.2. This reveals that the majority of visitors are positive, but not extremely enthusiastic. That is, intangible cultural heritage homestays are widely recognized, but some aspects need improvement and do not fully meet the highest expectations of tourists.

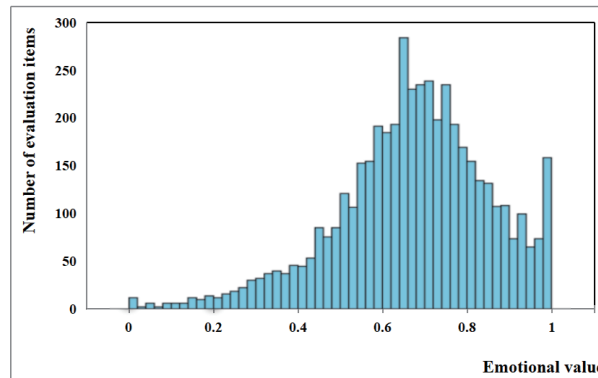


Figure 3. Distribution of emotional values.

### 3.2.2. Multidimensional analysis of affective tendencies

Tourists' positive emotional tendencies towards intangible cultural heritage homestays scene from three dimensions: the homestay environment in the material scene, homestay services in the social scene, and cultural experience elements in the cultural scene. Specifically, as follows:

#### 3.2.2.1. Cultural experience

Comments showed that tourists were highly satisfied with the folk experience at the intangible cultural heritage homestay, such as saying, "Very satisfied. The owner is a native of Xinjiang, and it can be seen that the people of Xinjiang have a beautiful and magnificent aesthetic" "The lobby is a wooden structure with ethnic decorations, and there are many intangible cultural heritage elements, demonstrating respect and integration for the local culture" "The homestay is very ethnic." The intangible cultural heritage homestay incorporates elements of ethnic and regional intangible cultural heritage, builds a unique cultural scene, showcases the local unique history, culture and folk customs, and promotes tourists' understanding and inheritance of local intangible cultural heritage. In addition, the cultural scenes incorporate local traditional architectural styles and cultural elements, providing tourists with rich cultural experience Spaces.

#### 3.2.2.2. Homestay services

Visitors in the comments expressed satisfaction with the service, such as "The service was also very good, all questions were answered, and I learned about the intangible cultural heritage of ink-making" "Intangible Cultural Heritage Experience - the Hui Ink gilding course was great, and I also watched the butler make iron flowers with his life" "The special activities of Hui ink gilding and iron flower making were great, experiencing two intangible cultural heritages at one time" "Woodcut watermark experience: very good". In the social setting of the intangible cultural heritage homestay, through course teaching and practical experience, tourists' understanding of intangible cultural heritage is deepened, which is conducive to the learning and dissemination of intangible cultural heritage traditional culture.

#### 3.2.2.3. Homestay environment

In the comments, tourists said, "The hotel has convenient transportation. It's very convenient to take a taxi. It's inside the Confucius Temple. It's a 5-minute walk to the Qinhuai River Wharf, Nanjing Examination Hall, Confucius Temple, Wuyi Lane. It's very convenient to take a night tour of the Qinhuai River" The superior

physical scene of the intangible cultural heritage homestay attracts a large number of tourists, especially when it is located near the scenic area or in a convenient transportation location, it is more favored by tourists, enhancing its popularity and appeal. At the same time, some tourists said "The environment is comfortable, clean and hygienic", indicating that the comfort of the physical scene of the homestay has been recognized. The positive feedback from tourists reflects the achievements and efforts of intangible cultural heritage homestays in cultural inheritance and service provision. Negative comments, on the other hand, reveal operational deficiencies. Text analysis shows that the negative comments mainly focus on the facilities of the material scene. The poor infrastructure has led to a decline in the visitor experience, such as "There's no hot water to take a bath in the middle of the night after a day of hiking, it's really frustrating" "The lack is no air conditioning, it's too hot to sleep at night" "There's no parking lot, you have to walk after getting off the vehicle" "The wooden room has poor sound insulation", which can cause dissatisfaction among visitors.

## **4. Research conclusions and countermeasures**

### **4.1. Research conclusions**

Based on visitor reviews of intangible cultural heritage homestays on Ctrip and Tongcheng Travel, this study used jieba word segmentation and LDA topic clustering analysis to reveal scene types and sentiment analysis through SnowNLP. The study found that, first, the analysis of high-frequency words showed that tourists' focus was concentrated on the accommodation experience, service, location, cultural characteristics, decoration style and infrastructure of the intangible cultural heritage homestay. Second, thematic modeling identified three core dimensions of the intangible cultural heritage intangible cultural heritage homestay scene, namely the physical scene, the social scene, and the cultural scene, covering six major thematic elements: intangible cultural heritage homestay environment, cultural experience, service, customer group, facilities, and comfort experience. Finally, sentiment analysis shows that tourists have a positive overall sentiment towards intangible cultural heritage homestays. By dimension, tourists have a high proportion of positive emotions towards the homestay environment in the material scene, homestay services in the social scene, and cultural experience in the cultural scene, while negative emotions towards the facility conditions in the material scene are obvious.

### **4.2. Enhancement strategies for the construction of intangible cultural heritage homestay scene**

#### **4.2.1. Physical scene dimension: Precision upgrading and intelligent transformation of facilities**

Build intelligent shared parking systems, develop off-peak shared parking lots, integrate reservation applications and intelligent navigation to alleviate the shortage of parking Spaces; Optimize the acoustic environment of guest rooms by using double-pane glass, soundproof cotton and sound-absorbing carpets to reduce negative noise reviews; Establish a standardized system for facilities, set parameters for bathrooms, hot water supply, and intelligent environmental control, and implement modular design optimization.

#### **4.2.2. Social scene dimension: Service quality improvement and precise customer group operation**

Establish a talent cultivation system for the integration of intangible cultural heritage inheritance and services, integrate intangible cultural heritage skills courses, and add cultural interpretation specialists; Implement customer demand-oriented service strategies and develop parent-child intangible cultural heritage workshops, cultural master workshops, and business scenario service packages; Build a digital service response matrix, deploy a scanning evaluation system and a fault diagnosis cloud platform, and set up a 24-hour service hotline.

#### **4.2.3. Cultural scene dimension: Deepening of intangible cultural heritage experience and construction of memory symbols**

Create a "space-activity-product" experience system, set up intangible cultural heritage performance areas and night events, and develop experiential cultural and creative derivatives; Improve the narrative interpretation system of intangible cultural heritage through digital archives, dynamic demonstrations and cultural decoding to achieve the dissemination at the meaning level; Establish a "homestay - inheritor - community" collaborative mechanism, innovate the inheritor residency system, develop seasonal intangible cultural heritage projects, and build community ecological nodes.

## Fundings

This study was financially supported by the Anhui Social Science Innovation and Development Research Project, "Research on the Construction and Innovative Development of Anhui Intangible Cultural Heritage Homestays from a Scene Theory Perspective" (Project No. 2023CX524).

## References

- [1] Y. C. CHEN. Protection of Intangible Cultural Heritage and Local Development Path in Hunan Province, *Economic Geography*, vol. 40 (2020), 227-232.
- [2] Y. X. WU, Y. TIAN. Research on the Path of Rural Revitalization and Development Promoted by "Homestay Inns+" Model — Taking Baihuashan Homestay as Examples, *Chinese Landscape Architecture*, vol. 38 (2022), 13-17.
- [3] J. J. LI, Y. C. LI. Reflections on the High-Quality Development of Bed&Breakfast Industry in Ethnic Minority Areas of China, *Journal of Yunnan Minzu University (Social Sciences)*, vol. 36 (2019), 70-75.
- [4] Z. Y. YU, W. Q. LYU. Research on Tourists' Symbol Perception and Experience of Rural Residential Consumption Space — Based on the Investigation and Analysis of Several Villages in Chengdu, *Price: Theory & Practice*, (2024), 102-107.
- [5] H. P. HUANG, Y. LIU, W. ZHONG, et al. Scene Cognition and Conation Tendency: Mediation of Emotional Identity and Moderation of Perceived Value — Take Shanghai High-End Homestay as an Example, *Tourism Science*, vol. 39 (2025), 113-133.
- [6] Y. QIAO. Study on Interaction of Intangible Cultural Heritage Protection and Tourism Development Based on Urbanization: A Case of Contiguous Area of Wuling Mountain, *Guizhou Ethnic Studies*, vol. 38 (2017), 168-171.
- [7] J. H. XIAO, W. Z. LI, F. WANG, et al. A Study of Non-Market Value of Tourism Resources Based on Choice Experiment and Stated Attribute Non-Attendance: A Case Study of Changdao Fishermen's Work Songs as a National Intangible Cultural Heritage, *Tourism Science*, vol. 38 (2024), 62-81.
- [8] X. SONG, S. H. WANG. Experience Co-creation in Intangible Cultural Heritage Tourism: Dimension, Measurement and Empirical Research, *Journal of Sichuan Normal University (Social Sciences Edition)*, vol. 51 (2024), 39-49.
- [9] D. A. Silver, T. N. Clark. *Scenescapes: How Qualities of Place Shape Social Life*(Social Sciences Literature Press, Beijing, 2019), p.46 - 66. (In Chinese).