

Optimization of the Agricultural, Cultural, and Tourism Integration Pathway in Yijun County Based on Big Data Analysis

Meng Qin^{1,*}

¹ Business School of China University of Political Science and Law, Beijing, 100000, China

Corresponding authors: (e-mail: qinmengbs@126.com).

Abstract Yijun County is located in Tongchuan City, Shaanxi Province, China, and despite sitting on rich agricultural, cultural and tourism resources, its tourism industry is currently still facing the dilemma of insufficient competitiveness of natural landscape and ineffective combination of agriculture and tourism. In this paper, data analysis is used as a research tool, and based on knowledge mapping technology, the integration strategy of culture and tourism knowledge is proposed to construct the correlation relationship between attraction scenery and culture. Under this theoretical framework, the improved SO-PMI algorithm is selected as the screening method of emotion seed words in the text of attraction reviews to design an exclusive emotion dictionary in the field of tourism. Meanwhile, the social network analysis method is used as a research tool for the characterization of the spatial structure and influencing factors of regional cultural and tourism industry integration. Using the designed research method to analyze and summarize the characteristics of the review text of online tourism websites in Yijun County, the distribution of “landscape” theme feature words is more and accounts for the highest proportion, and the weight interval of its first five feature words is [0.021,0.051]. It shows that the construction process of the integration path of agriculture, culture and tourism industry needs to take the local unique attractions and landscapes as the core content and develop around its own characteristics.

Index Terms SO-PMI algorithm, sentiment lexicon; Yijun County, agriculture, culture and tourism industry integration, knowledge graph, social network analysis

I. Introduction

Industrial integration refers to the dynamic development process in which different industries or different sectors of the same industry interpenetrate and intersect with each other, and eventually merge into one and gradually form a new industry [1]. Nowadays, the deep integration of culture and tourism has become the key to enhance the core competitiveness of the industry and improve the economic structure [2]. The integration of agriculture, culture and tourism is a new mode of integration and development of three industries based on ecological agriculture, with culture and tourism as the new mode, and with the goal of upgrading the rural industry [3]. Taking Yijun County agriculture, culture and tourism industry as an example, it can not only promote the integration and development of rural resources and culture and tourism industry, but also extend the agricultural industry chain, so as to promote the sustainable development of the economy of Yijun County [4]-[6]. However, the integration of agriculture, culture and tourism industry also exists problems such as shallow integration, the quality of product supply needs to be improved, the lack of technical talents, and the unsound guarantee mechanism, and its industrial development needs to be optimized and innovated [7]-[9].

In the context of the new quality productivity, the integration of the Internet, big data and other cutting-edge digital technology with agricultural and rural development has opened up a brand new development path for the integrated development of agriculture, culture and tourism [10], [11]. Digital technology plays a crucial role in this integration process, and is gradually becoming a key way and innovative catalyst to promote the deep integration of culture and tourism, reshaping the future pattern of the industry [12]-[14]. The cutting-edge development of digital technology provides solid technical support and strong impetus for the transformation and high-tech integration of the agriculture, culture and tourism industry, accelerates its transformation process in the direction of high-tech integration, and promotes the deep integration of agriculture, culture and tourism at a deeper level, a wider range, and a higher level to provide important support.

This paper stands in the perspective of culture and tourism integration, and elaborates the integration strategy of culture and tourism knowledge under the knowledge graph approach as the theoretical foundation of research and

analysis. Then the improved SO-PMI algorithm is introduced to elaborate the construction process of the sentiment dictionary of tourism exclusive domain based on the improved SO-PMI. Subsequently, the calculation and analysis steps of the spatial structure characteristics and influencing factors of the integration of culture and tourism industry under the four perspectives of overall network density, centrality, cohesive subgroups, and QAP analysis in the social network analysis method are respectively explained. Finally, collect the text of tourists' comments in the online website of Yijun County, sort out the word frequency and analyze the emotional characteristics using the designed emotional dictionary. Based on the social network analysis method, the semantic network is established and the hot topics of comments are summarized. Combined with the analysis, the optimization path of tourism development in Yijun County under the integration of agriculture, culture and tourism industry is explored.

II. Data analysis of the integration of agricultural, cultural and tourism industries

II. A. Strategies for integrating culture and tourism knowledge

The fusion of cultural knowledge and tourism knowledge is built on the basis of the expansion and updating of the knowledge graph, and its main goal is to integrate and update the cultural and tourism knowledge graph by updating the historical and cultural knowledge obtained from network big data to the tourism knowledge graph after knowledge assessment, based on the representation learning, and the reasoning operation with the existing ternary group in the tourism knowledge graph, which is assessed to be the correct knowledge. This section focuses on the method of integrating network historical and cultural knowledge into tourism knowledge graph in the process of knowledge expansion.

In this paper, a three-layer cultural and tourism knowledge fusion framework is constructed from the perspective of cultural and tourism fusion. The modeling framework consists of a knowledge division and iteration layer, a knowledge acquisition layer, and a knowledge fusion layer, with each layer based on the knowledge graph as well as subgraph division.

The knowledge division and iteration layer divides the tourism knowledge graph into subgraphs, each subgraph obtains corresponding entities and relationships based on different perspectives, and the division of subgraphs reduces the expenditure of the whole knowledge graph training, and the subgraphs can be updated in a chronological order. After acquiring structured and semi-structured data such as travelogue text data and encyclopedia website data, the knowledge acquisition layer extracts ternary groups through information extraction and organization. The knowledge fusion layer performs knowledge validation and knowledge fusion on the ternary groups based on representation learning, and fuses and updates the cultural and tourism knowledge graph.

Based on the three-layer knowledge fusion, the cultural and tourism knowledge map can be based on the original tourism knowledge map, and make full use of the multi-source cultural big data on the Internet to organize massive multi-source data for the realization of the intelligent cultural and tourism platform, and make use of the sub-diagram division, representation learning and other methods to continuously fuse and update the knowledge map, and provide a basis for the downstream application based on the cultural and tourism knowledge map to realize the sharing and utilization of cultural and tourism knowledge. The knowledge fusion strategy is shown in Figure 1.

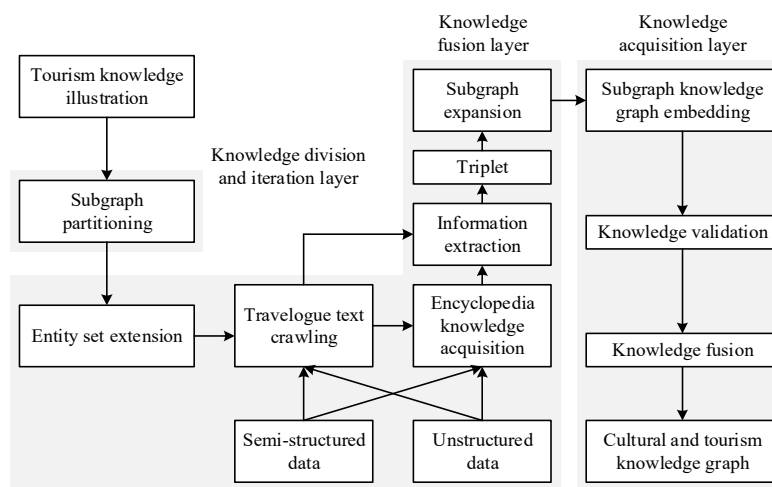


Figure 1: Framework diagram of the Integration of cultural and Tourism knowledge

II. B. Emotion dictionary construction in tourism domain based on improved SO-PMI

Sentiment lexicon is a key thing for conducting experiments of sentiment tendency analysis, but due to the increasingly prosperous development of the Internet, there are more and more emerging terms, and the sentiment words in different domains often contribute different sentiment tendencies to Chinese texts, and the domains in which some fixed sentiment lexicons can be used are also limited. Therefore, there are many emotion dictionaries specialized in a certain domain have been constructed, such as the inert emotion dictionary in the domain of product review, the exclusive emotion dictionary in the domain of social media text and the exclusive dictionary in the domain of public opinion in emergencies, etc. In this paper, we refer to the research of related scholars to establish an emotion dictionary belonging to the exclusive domain of tourism. There are two methods to judge the sentiment tendency of a word, one is the semantic similarity algorithm of KnowledgeNet, and the other is the sentiment word polarity algorithm of SO-PMI.

In this paper, we use the emotional tendency point mutual information algorithm SO-PMI algorithm to carry out the establishment of tourism exclusive domain emotion dictionary, this algorithm is based on information theory, which is able to calculate the co-occurrence probability between words and words in the text. Where the probability of co-occurrence between words and words, i.e. PMI calculation formula is shown in equation (1).

$$PMI(h_i, h_j) = \log_2 \left[\frac{p(h_i, h_j)}{P(h_i) * P(h_j)} \right] \quad (1)$$

In Equation (1), h_i and h_j denote the feature words of two texts, $P(h_i)$ and $P(h_j)$ denote the probability of the occurrence of the two feature words h_i and h_j in the comment text, and $p(h_i, h_j)$ denotes the probability that both h_i and h_j feature words appear in the comment text. The higher the probability that h_i and h_j occur simultaneously in a small range in the comment text, the higher the association of the two feature words, and vice versa. Where $\frac{p(h_i, h_j)}{P(h_i) * P(h_j)}$ the value of this formula is the independence measure of the feature words h_i and h_j , then the value of $PMI(h_i, h_j)$ have three cases as in equation (2):

$$PMI(h_i, h_j) = \begin{cases} > 0 & \text{At this point, } h_i \text{ and } h_j \text{ are correlated.} \\ = 0 & \text{At this point, } h_i \text{ and } h_j \text{ are independent} \\ < 0 & \text{At this point, } h_i \text{ and } h_j \text{ are mutually exclusive} \end{cases} \quad (2)$$

On the basis of PMI (point mutual information), the emotional information is integrated to calculate the emotional tendency of the characteristic words, the algorithm is called the sentiment tendency point mutual information algorithm SO-PMI algorithm, which is used to combine the seed word set, the general sentiment dictionary and the establishment of the tourist comment text prediction library to establish the emotion dictionary in the exclusive field of tourism, the principle of the SO-PMI algorithm is: under the standard of positive emotion seed words T_1 and negative emotion seed words T_2 , The difference between the PMI value of the words in the feature words h_i and T_1 and the PMI values of the words in the feature words h_i and T_2 indicates the emotional tendency of the feature words h_i , as shown in Eq. (3).

$$SO-PMI(h_i) = \sum_{T_{1i} \in \text{SeedDic}} PMI(h_i, T_{1i}) - \sum_{T_{2i} \in \text{SeedDic}} PMI(h_i, T_{2i}) \quad (3)$$

There are three cases of $SO-PMI(h_i)$ as in equation (4):

$$SO-PMI(h_i) = \begin{cases} > 0 & \text{Then the feature word } h_i \text{ is a positive emotional word} \\ = 0 & \text{Then the feature word } h_i \text{ is a neutral sentiment word} \\ < 0 & \text{Then the feature word } h_i \text{ is a negative emotion word} \end{cases} \quad (4)$$

However, the SO-PMI algorithm also has significant drawbacks: one is that the use of the algorithm must be based on the corpus in order to determine the sentiment tendency of a feature word, but in practice the possibility of a feature word and the sentiment seed word obtained in the previous section appearing in a comment text is unpredictable, or even will not be co-occurring, and the value of the SO-PMI in this case is very likely to be 0, which will make the experiment suffer that leads to large errors in the sentiment analysis results. Secondly, the size of the co-occurrence window of two feature words is usually set manually, in a larger corpus, the length of a sentence can

be set as the size of the window, while for some short texts such as microblog copy or the visitor's comment text studied in this paper, whose textual data is more loosely organized, the size of a comment text file can usually be set as a co-occurrence window. And the size of the window will affect the content of the sentiment dictionary, which in turn affects the final results of the sentiment analysis experiment.

Therefore, on the basis of SO-PMI's algorithm, this paper adds the smoothing technique.

The formulas of $p(h_i, h_j)$, $P(h_i)$ and $p(h_j)$ in Equation (1) are shown in (5), (6) and (7).

$$p(h_i, h_j) = \frac{f(h_i, h_j)}{n} \quad (5)$$

$$P(h_i) = \frac{f(h_i)}{n} \quad (6)$$

$$p(h_j) = \frac{f(h_j)}{n} \quad (7)$$

where $f(h_i, h_j)$ denotes the number of texts in which the feature words h_i and h_j appear simultaneously in the comment text, $f(h_i)$ and $f(h_j)$ denote the number of texts in which the h_i and h_j appear in the comment text, and n denotes the total number of texts in the visitor's comment text. Thus the $PM(h_i, h_j)$ formula can be organized as Equation (8).

$$PMI(h_i, h_j) = \log_2 \frac{n * f(h_i, h_j)}{f(h_i) * f(h_j)} \quad (8)$$

In the comment text h_i and h_j may not appear at the same time, so in this paper, we will use the Laplace plus one technique for Eqs. (5), (6), and (7).

Equation (8) can be organized into equation (9).

$$PMI(h_i, h_j) = \log_2 \frac{n^2 * f(h_i, h_j)}{(n+1) * f(h_i) * f(h_j)} \quad (9)$$

Equation (3) can be transformed and organized into equation (10).

$$SO-PMI(h_i) = \sum_{\substack{T_{1i} \in E_{SeedDic} \\ T_{2i} \in E_{SeedDic}}} \log_2 \frac{[f(h_i, T_{1i}) + 1] * f(T_{2i})}{[f(h_i, T_{2i}) + 1] * f(T_{1i})} \quad (10)$$

In this paper, we construct a domain-specific sentiment lexicon on the basis of SeedDic and the improved SO-PMI with the following steps:

Step 1: The collected comment text is partitioned using jieba, and after removing the deactivated words, the words with word frequency greater than 100 after partitioning are deposited into the text word set called candidate word set as in equation (11):

$$BTSet = \{w_1, w_2, \dots, w_n\} \quad (11)$$

The candidate sentiment word set is obtained by taking the intersection of the word set BTSet with the three publicly available sentiment dictionaries.

Step 2: Iterate the candidate word set BTSet with SeedDic and deposit the words belonging to the positive sentiment seed words of the candidate word set into the set B_1 . The words belonging to the negative sentiment seed words of the candidate word set are deposited into B_2 . The words that are not matched are added to the word set BTSet2.

Step 3: Calculate the values of SO-PMI of words in BTSet2 with sentiment seed word sets E_1 and E_2 using the improved SO-PMI algorithm in this paper.

Step 4: Words with SO-PMI values greater than zero and their corresponding SO-PMI values are deposited into the positive sentiment lexicon B_1 together, and words with less than zero and their corresponding SO-PMI values are deposited into the negative sentiment lexicon B_2 together, and the lexicon B_1 together with B_2 constitutes the exclusive domain sentiment lexicon.

The process of constructing the tourism exclusive domain sentiment dictionary is shown in Fig. 2.

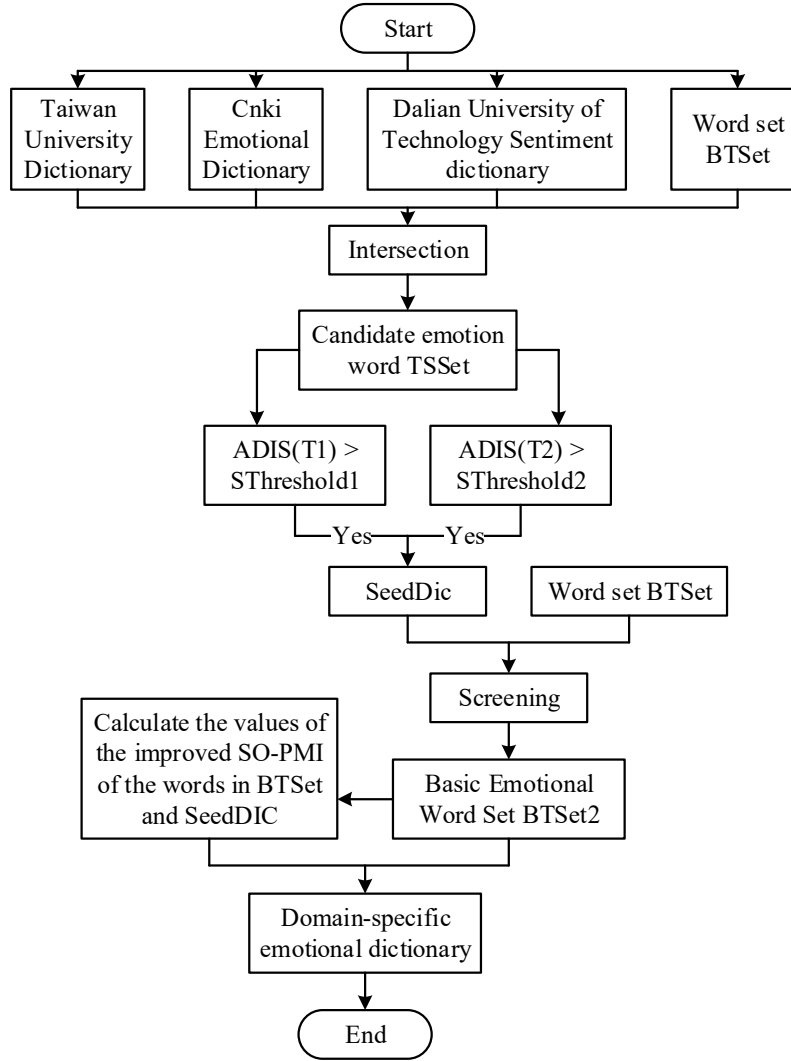


Figure 2: Flowchart for the construction of domain-specific emotional dictionaries

II. C. Social network analysis

Social network analysis is a systematic analysis method to assess the structural characteristics of complex networks through relational data. Many scholars calculate the gravitational strength through the gravitational model and introduce the social analysis method to characterize the network structure in the fields of economy, technology diffusion, and inter-city cooperation. This paper focuses on the spatial structural characteristics and influencing factors of cultural and tourism industry integration from the perspectives of overall network density, centrality, cohesive subgroups, and QAP analysis.

The overall network density mainly measures the degree of connectivity of the whole network and portrays the degree of connection between nodes. The greater the overall network density, the closer the spatial correlation of the quality of cultural and tourism industry integration between the whole cities, as shown in equation (12):

$$D_e = \frac{n}{m \times (m-1)} \quad (12)$$

In Eq. (12) D_e is the overall network density, n is the actual number of joint edges of the directed network, and m is the theoretical maximum number of joint edges of the directed network.

Centrality is used to measure the characteristics of individual nodes in the network, mainly reflecting the importance of the point in the network. It is divided into three metrics in practical calculations: point degree centrality, mediator centrality, and proximity centrality.

Point degree centrality measures the degree of direct connection between a node and other nodes, and the quantitative index is the number of joint edges between this node and other nodes. The larger the point degree

centrality is, the greater the direct correlation between this city and other cities in the integrated development of culture and tourism industry, which is a kind of direct embodiment of influence. As shown in equation (13):

$$C_D(n_i) = \sum_{j=1}^m L_{ij} \quad (13)$$

In Eq. (13), $C_D(n_i)$ is the point centrality degree of the i th node, and L_{ij} is the number of direct edges between node i and node j . In directed networks, the point centrality degree is divided into the in-degree and the out-degree centrality degree, where the out-degree stands for the ability of influencing other cities, and the in-degree centrality degree stands for the influence of being influenced by other cities. Degree of centrality.

The intermediary centrality mainly portrays the degree of control of the node over other nodes, the larger the intermediary centrality of the node the more times the node appears in the shortest path of other point pairs, indicating that the city plays a role as a bridge in the propagation of the influence of industrial integration in other cities, and even has the ability to control and influence the integrated development of cultural and tourism industries in other cities. As shown in equation (14):

$$C_B(n_i) = \frac{\sum_{j < k} g_{jk}(n_i)}{g_{jk}} \quad (14)$$

In Eq. (14) $C_B(n_i)$ is the mediator centrality of node n_i , $g_{jk}(n_i)$ is the number of times node n_i is in the shortest path of node j and node k , and g_{jk} is the number of node j and node k the number of shortest paths.

Proximity centrality describes the degree to which a node is not controlled by other nodes, i.e., when the proximity centrality is high, the point can be connected to other nodes through shorter paths. The higher the degree of proximity to the center, the more the city is at the center of gravity of the network, has more autonomy and is less disturbed by other cities. As shown in equation (15):

$$C_C(n_i) = \sum d_{ij} \quad (15)$$

In equation (15), $C_C(n_i)$ is the proximity centrality of node i , and d_{ij} is the shortest distance from node i to node j , which is divided into the in-degree proximity centrality and out-degree proximity centrality when the network is a directed graph.

Cohesive subgroup analysis, also called small group analysis, refers to when certain nodes in the network are particularly close, forming a small group, such a small group is called a cohesive subgroup in social network analysis. In the process of intercity cultural and tourism industry integration and development, due to the proximity of intercity areas and the similarity of economic levels, the density of relationships between certain members is higher than others, so it is also necessary to study these group phenomena.

QAP analysis is a method to quantitatively analyze the relationship between matrices. Traditional regression analysis may suffer from the problem of multicollinearity, which can be effectively avoided by QAP analysis. The basic step is to compare the similarity of the individual grid values of two square matrices to give the correlation coefficients between the two matrices, and at the same time the coefficients are subjected to nonparametric tests, so that the two matrices reach a statistically significant correlation when the significance level is between 0.01 and 0.1. When the regression relationship between multiple matrices and one matrix is studied, then QAP regression is used.

III. Construction and Optimization of Tourism Development Path in Yijun County

In this paper, a total of 22,000 comment texts of Yijun County online tourism website during the 5-year period of 2015-2019 are selected as the research samples, and a total of 21,926 comment texts are obtained after completing the pre-processing of the data through the operations of filtering, culling, cleaning, etc., and the proposed dictionary of sentiment analysis of the tourism field is used in this chapter to conduct the word frequency analysis and the sentiment analysis. Combined with the results of the sentiment analysis, the semantic network of the feature words and the hot topics of the comments are analyzed sequentially using social network analysis. Synthesizing the analysis content, the proposal of tourism development path in Yijun County under the integration development strategy of agriculture, culture and tourism is put forward.

III. A. Emotional characterization of review texts

III. A. 1) Word frequency analysis

The word frequency of the text can effectively reflect the strength of the subjective perception of tourists on different elements, i.e. the “ease of perception” of different elements or the “depth of impression” of tourists. We analyze the word frequency of the processed data, summarize the synonyms and near-synonyms, and count the top 300 words. The top 50 words are listed in Table 1.

Table 1: High-frequency words in tourist comments

Entry	Frequency	Entry	Frequency
Yijun County	3978	Guishan park	231
Avoid summer heat	1668	Good-looking	221
Revolution	1020	Walnut	222
Take a picture	902	Delicious food	216
Sala	845	Commercialize	216
Snack	726	Free of charge	205
culture	696	Public transportation	203
The homeCounty of Peng Zu	686	Street	203
The grottoes	675	Park	181
Yangshao culture	647	Inexpensive	177
Feature	624	Lively	165
Architecture	623	To city	153
Weekend	603	Out slice	153
Deserve	592	Han Chinese Clothing	144
The Great Wall	583	On the street	105
Dryland terraced fields	499	Old street	103
History	452	Tomato steamed bun	72
Cartoon	432	Kuquan Terraced Fields	70
Tickets	420	Yunmeng Mountai	69
Traffic	395	parking lots	53
Make an appointment	346	Fudi Lake	45
Huaxi valley	333	Scenery	42
Shaanxi	328	Holidays and Festivals	42
Folk painting	242	Wowo Noodles	42
Jiangnu Tear Spring	239	Ear covers	39

III. A. 2) Sentiment analysis

Sentiment analysis yielded positive, neutral, and negative percentages of comments, of which the number of positive sentiments was 13,897, accounting for 63.38%, far exceeding half of the total number of comments. Neutral sentiment 5645 articles, accounting for 25.75%. Negative sentiment 2384 articles, accounting for 10.87%, indicating that tourists are generally satisfied with the tourism experience in Yijun County. The sentiment analysis of visitors' comments in Yijun County is shown in Table 2, and in order to further identify the real emotions embedded in visitors' comments, the following grading is further made in the three qualitative comments: high, moderate, and general.

Combined with the visitor perception were extracted spatial perception words of the text, through the sentiment dictionary for sentiment analysis to derive the range [0-1] between the sentiment value, the results of the closer to 1 indicates that the comment text is more positive, the closer to 0 indicates the more negative, all the text of the sentiment value of the average value, listed in the rankings of the top 10 words of the results in Table 3.

As can be seen from Table 3, the top 10 entries with mean emotion are: Waiquan Terraces, Huaxi Valley, Tomato Steamed Bun, Grottoes, Great Wall, Food, Scenery, Yunmeng Mountain, Wowo Noodles, and Jiang Nu Tear Spring. Among them, the average sentiment analysis of "Waiquan Terraces" was the highest (0.88), followed by "Huaxi Valley" (0.83), and "Tomato Steamed Bun" ranked third with 0.82. Observing other entries, it can be found that most of them are famous scenic spots or special delicacies in Yijun County. According to this, local scenic spots and delicacies can positively affect tourists' perception of tourism space in Yijun County.

Table 2: Emotional Analysis of Tourists' Comments on Yijun County

Comments on emotions	Number of texts	Proportion (%)	Evaluation of classification	Text quantity	Proportion (%)
Positive	13897	63.38	High	7542	54.27
			Moderate	4411	31.74
			Average	1944	13.99
Neutral	5645	25.75	High	2903	51.43
			Moderate	1817	32.19
			Average	925	16.39
Negative	2384	10.87	High	745	31.25
			Moderate	1068	44.80
			Average	571	23.95

Table 3: Analysis of Spatial Perception Words in Yijun County

Perception entry of Yijun County	Frequency	Average value of sentiment analysis
Kuquan Terraced Fields	70	0.88
huaxi valley	333	0.83
Tomato steamed bun	72	0.82
The grottoes	675	0.8
The Great Wall	583	0.79
delicious food	216	0.77
scenery	42	0.75
Yunmeng Mountai	69	0.73
Wowo Noodles	42	0.71
Jiangnu Tear Spring	239	0.7

III. B. Spatial structure characteristics and influencing factors

III. B. 1) Semantic network analysis

Multiple key factors that tourists pay attention to can be obtained through high-frequency word analysis, but the relationship between these factors is not directly expressed, so it is necessary to identify the relationship between different factors through semantic network analysis, so as to reveal the hidden correlation between factors. Based on the statistics in Table 1, the co-occurrence matrix between words is constructed. Taking the top 10 words with the highest frequency (Yijun County, summer vacation, revolution, photography, Shorea tree, snack, culture, Pengzu's hometown, grottoes, Yangshao culture) as an example, the co-occurrence matrix between the 10 high-frequency words is constructed by numbering the words in the order of C1-C10, as shown in Table 4. Since the top 10 words are mainly tourist attractions, the co-occurrence frequency varies greatly among different words. The co-occurrence frequency of (C1) Yijun County and the other 9 words and phrases are in the following order: 9253, 8741, 9986, 8686, 9027, 8412, 9635, 9216, and 9084, which indicates that the 9 words and phrases are the main reasons for tourists to choose to travel to Yijun County.

Table 4: The co-occurrence matrix of high-frequency words

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
C1	1	9253	8741	9986	8686	9027	8412	9635	9216	9084
C2	9253	1	86	14	742	68	3396	107	3879	113
C3	8741	86	1	6803	27	25	8649	9079	7453	1514
C4	9986	14	6803	1	7785	33	8403	8024	8501	8009
C5	8686	742	27	7785	1	11	6307	16	76	21
C6	9027	68	25	33	11	1	2144	61	28	16
C7	8412	3396	8649	8403	6307	2144	1	9472	9382	9178
C8	9635	107	9079	8024	16	61	9472	1	562	6756
C9	9216	3879	7453	8501	76	28	9382	562	1	8648
C10	9084	113	1514	8009	21	16	9178	6756	8648	1

Comprehensively analyzing the feature associations of the semantic network, it can be found that the factors affecting tourists' visiting experience in Yijun County scenic spots are roughly the following three aspects: landscape, experience, and food, which together constitute the core elements of the tourism process, and provide a solid foundation for the overall experience of tourists.

III. B. 2) Analysis of Hot Topics in Commentary

Through high-frequency words and semantic analysis, the factors affecting tourists' experience and their basic relationships can be identified, but the semantic mining of online reviews is not yet in-depth. In order to further reveal the hidden theme information in online reviews, theme analysis of tourists' online reviews is conducted through LDA model. For better theme identification, after several experiments, it is determined that the number of themes is set to 3: (Theme I) Landscape, (Theme II) Experience, and (Theme III) Diet, and the feature words are set to 5. After the resulting theme data are organized and analyzed, the distribution of the themes and the theme words obtained is shown in Table 5.

Table 5: The distribution of online comment topics and key words in Yijun County

Theme One:Landscape		Theme Two:Experience		Theme Third: Diet	
Feature word	Weight	Feature word	Weight	Feature word	Weight
The hometown of Peng Zu	0.051	Take a picture	0.042	Snack	0.045
The grottoes	0.038	Traffic	0.027	Walnut	0.038
Dryland terraced fields	0.032	Make an appointment	0.026	Tomato steamed bun	0.035
Huaxi valley	0.023	Commercialize	0.017	Wowo Noodles	0.025
Jiangnu Tear Spring	0.021	Park	0.011	Ear covers	0.019

Overall, (T1) landscape theme feature words are more distributed, and its first five feature words have the highest weight interval of [0.021,0.051]. The overall distribution of (T2) experience theme feature words and (T3) diet feature words, on the other hand, is relatively less, with [0.011,0.042] and [0.019,0.045], respectively.

In theme 1, the words "Peng Zu's hometown", "grottoes", "dry farming terraces", "Huaxi Valley", and "Jiang Nu Tear Spring" are all expressions of the landscape of the scenic spot, so theme 1 is named "landscape". The five characteristic words reflect that the unique revolutionary and historical and cultural landscape of Yijun County is the main content of Yijun County's scenic spots to tourists, and it is also the main purpose of tourists to visit Yijun County. Tourists have a unique viewing experience by viewing the unique revolutionary sites and historical and cultural attractions in Yijun County, and are deeply impressed by the visiting experience, so as to better understand the revolution and history and culture.

The five characteristic words in Theme 2 are: taking pictures, transportation, reservation, commercialization, and parking, and the five characteristic words all reflect the tourists' travel experience in Yijun County, so Theme 2 is named "Experience". The five characteristic words reflect that tourists' main concerns about the tourism experience of Yijun County are the souvenir value, transportation convenience and service quality of scenic spots.

The five characteristic words "snacks", "walnuts", "tomato steamed buns", "wowo noodles" and "ear covers" in theme 3 are all descriptions of the local specialties of Yijun County, so theme 3 is named "food". The results show that the local specialty food in Yijun County can provide tourists with a different taste bud experience, which not only has a certain attraction to tourists, but also helps to improve tourists' tourism satisfaction.

III. C. Tourism development path in Yijun County

Combining the above analysis with the actual development of Yijun County, it can be seen that Yijun County is a typical county-level area with rich agricultural, cultural and tourism resources. In agriculture, it has unique terraced landscape and rich agricultural resources. In the cultural and tourism resources, its territory has Xuanzang hand-planted Shorea tree, Peng Zu's hometown, Northern Wei grottoes, Jiangnu Tears Spring, Yunmeng Mountain, the Great Wall of Wei in the Warring States period, Fudi Lake, Anime Flower Valley, etc., not only is one of the birthplaces of Yangshao culture, but also the Yanmen detachment of the active old red revolution area. It can be seen that Yijun County does not lack of hard conditions for the development of tourism industry. Therefore, its tourism development path should be cut from the perspective of "integration development", through the integration of many different aspects of the existing resources, the formation of unique tourism characteristics. In this regard, this paper proposes the following major integration development paths:

(1) Integration of local characteristics, promote the construction of agriculture, culture and tourism integration development path. Take the dry terraces as an example, although it has unique natural scenery and rich

agricultural resources, but in the current tourism market compared to other mature terraced scenic spots, Yijun terraces purely based on the traditional sightseeing mode does not have a clear competitive advantage. Therefore, by dividing the experience area, setting up the experience guide position, explaining the formation history and construction principle of terraces, and providing tourists with the opportunity of in-depth participation and personalized experience, tourists can further understand the role of terraces in the process of first-hand experience. The unique tourism experience will effectively differentiate the same type of scenic spots in other areas, and serve as a brand new development direction and path for Yijun Terraces. By allowing tourists to participate in farming and experience the joy of farming, it can not only help Yijun County Terraces realize the transformation from single sightseeing to multi-functional direction of sightseeing, experience, culture, etc., but also provide local farmers with more jobs and opportunities to promote income generation.

(2) Create a cultural brand and carry out cultural activities. In the above proposed “agriculture, culture, tourism” tourism industry integration development path, culture is the key link between agriculture and tourism. In the design of cultural activities, in addition to the need to create a unique cultural brand based on the characteristics of existing tourism resources, but also need to focus on the diversity of cultural activities, fun and knowledge. After the formation of a unique cultural brand, you can also make full use of modern network media platforms to publish and publicize the relevant cultural content, help Yijun County tourism industry and cultural brand awareness, attraction, and promote the prosperous development of the tourism industry.

IV. Conclusion

(1) Based on the knowledge mapping technology, this paper proposes the integration of culture and tourism knowledge into the strategy as the theoretical basis of the research. Design the emotion dictionary of tourism exclusive domain based on improved SO-PMI to obtain the emotion characteristics of online review texts of tourism websites. And social network analysis is used to analyze the spatial characteristics and influencing factors of attractions. As a result, the analysis method of attractions' spatial characteristics based on the emotional distribution characteristics of tourists' comments as well as the influencing factors under the perspective of integration of culture and tourism is comprehensively established.

(2) A total of 21,926 comment texts from Yijun County tourism website were combed as research samples, and the designed emotion dictionary was used to analyze the word frequency distribution and emotion characteristics. Among them, the number of positive emotions is 13897, accounting for 63.38%, far exceeding half of the total number of comments. The number of neutral emotions is 5645, accounting for 25.75%. Negative emotions 2384 articles, accounting for 10.87%, indicating that overall tourists have a good experience in Yijun County. Based on the results of comment text feature analysis to construct the co-occurrence matrix of high-frequency words, and summarize the landscape, experience, food and drink three hot topics of the comments, of which the “landscape” theme feature word distribution is more, and the weight of its first five feature words is [0.021,0.051], which accounts for the highest proportion. The overall distribution of “experience” and “food” feature words is relatively small, respectively [0.011,0.041] and [0.019,0.045].

(3) For the current development path of agricultural, cultural and tourism industry integration in Yijun County, this paper suggests: 1) Integrate the unique natural and cultural resources of Yijun County to promote the common development of agricultural, cultural and tourism industries; 2) Take culture as the key link for the integration of the three, and improve the recognition and attractiveness by creating unique cultural brands and carrying out cultural activities, so as to effectively help the development of agricultural, cultural and tourism industry integration.

References

- [1] Sick, N., Preschitschek, N., Leker, J., & Broering, S. (2019). A new framework to assess industry convergence in high technology environments. *Technovation*, 84, 48-58.
- [2] Tang, M., & Xu, H. (2023). Cultural integration and rural tourism development: A scoping literature review. *Tourism and Hospitality*, 4(1), 75-90.
- [3] Liang, T. (2025). In-depth Reflections on the Deep Integration of Agriculture, Culture, and Tourism in Support of Comprehensive Rural Revitalization. *Artificial Intelligence and Academic Research*, 1(1).
- [4] Jingran, L. O. N. G., & Shanshan, D. U. (2021). Mechanism and model of rural vitalization guided by culture-tourism integration. *Economic geography*, 41(7), 222-230.
- [5] Zhang, Y., Zheng, Q., Tang, C., Liu, H., & Cui, M. (2024). Spatial characteristics and restructuring model of the agro-cultural heritage site in the context of culture and tourism integration. *Heliyon*, 10(9).
- [6] Yang, R. (2022). Development of tourism for culture and innovation based on convergence of data in the perspective of industrial integration. *Mobile Information Systems*, 2022(1), 4174050.
- [7] Fang, L., Liu, Z., & Jin, C. (2023). How does the integration of cultural tourism industry affect rural revitalization? The mediating effect of new urbanization. *Sustainability*, 15(14), 10824.
- [8] Liu, C., Dou, X., Li, J., & Cai, L. A. (2020). Analyzing government role in rural tourism development: An empirical investigation from China. *Journal of Rural Studies*, 79, 177-188.

- [9] Luo, W., & Liu, J. (2024). From Tradition to Innovation: The Role of Culture Tourism in Transforming Chinese Agriculture. *Agriculture*, 14(11), 2042.
- [10] Lu, J. (2023). The integration of cultural tourism based on the Internet of Things and big data analysis. *Journal of Computational Methods in Science and Engineering*, 23(1), 325-334.
- [11] Li, B. (2023). The integration of cultural tourism and rural revitalization in big data and the internet of things. *Journal of Computational Methods in Science and Engineering*, 23(1), 51-62.
- [12] Wang, Y., Zhao, Z., Xu, M., Tan, Z., Han, J., Zhang, L., & Chen, S. (2023). Agriculture–tourism integration's impact on agricultural green productivity in China. *Agriculture*, 13(10), 1941.
- [13] Kachniewska, M. A. (2015). Tourism development as a determinant of quality of life in rural areas. *Worldwide Hospitality and Tourism Themes*, 7(5), 500-515.
- [14] Ponsignon, F., & Derbaix, M. (2020). The impact of interactive technologies on the social experience: An empirical study in a cultural tourism context. *Tourism Management Perspectives*, 35, 100723.