The Frontiers of Society, Science and Technology
ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

A Picture-Based Approach to Tourism Recommendation System

Yuan Zhang*, Shiqi Liu

Shenzhen Campus, Jinan University, Shenzhen, 518052, China
**Corresponding author: qyuan17@stu2020.jnu.edu.com*

Abstract: *Due to the subjective nature of the experience, it is challenging to classify travel products according to uniform criteria as physical products, which poses a significant challenge for recommendation techniques. In practical application scenarios, this study introduces an alternative image-based approach as an implicit elicitation of user preferences for travel products to accurately recommend products to travellers with different tastes. We developed a model based on exploratory factor analysis. First, based on exploratory factor analysis, we compared the pictures tapped by users to five types of travel preferences. Then, we use the Euclidean algorithm to project the images and scenic spot features into a five-dimensional space. It calculates the distance, thus quantifying the relationship between the travel and the user's favourite pictures. The final result shows that the model effectively improves the user's travel matching and effectively contributes to the field of travel destination recommendation.*

Keywords: *Travel recommendation algorithm, Travel personality, Factor analysis*

1. Introduction

Over the past few decades, the Web has become integral to travel and tourism. It had an even more significant impact on these industries than predicted when travel e-commerce emerged in the 1990s [1-3]. The Web has changed the travel industry and how consumers search for information about their upcoming trips.

Nearly two-thirds of people use this medium online to search for travel-related information. However, about one-third of users cannot clearly express their travel needs and expectations [4]. Usually, this is at an early stage of deciding on their "next travel destination". In keyword-based search engines, if preferences cannot be explicitly worded, this makes a prerequisite for making travel recommendations unfulfilled, namely that the user already knows what term they are searching for. Therefore many options are excluded, which is a limitation imposed by a traditional web search for travel search and recommendations.

New recommendation systems can overcome this problem, but first, make sure that the system is modelled rigorously enough to allow users to express their preferences. As mentioned above, users may need to be made aware of their intentions in many cases. In addition, travel recommendations are uniquely challenging because travel products are very complex and "emotional". Therefore, the problem is twofold: not only is it difficult for the system to provide appropriate product recommendations in the travel domain, but users may need help to tell the system precisely what they want. Therefore, more sophisticated techniques are required to support eliciting user preferences and may be helpful.

This paper presents an approach. This study suggests a model in which a user's preferences are represented by five factors summarized by a scale adapted from the Big Five Personality Inventory. The values of these factors are determined by the set of pictures selected by that user. The model of this study allows us to quantify these factors. With a highly personalized description of the user's travel needs, it can be deployed in the next step to provide travel-related suggestions. In other words, this study's approach maps the user, the pictures and the travel items into a 5-dimensional space. The main advantages of this study's approach are shown below.

(1) Considering the characteristics of travel activities, we designed a model that focuses on travellers' needs rather than product characteristics and effectively elicits travellers' preferences by selecting images.

ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

(2) When capturing traveller preferences, users do not have to answer a long and tedious questionnaire but choose from a set of attractive predefined images.

(3) The present study shows that the five dimensions generated by relying on the Big Five personality are sufficient to capture the diversity of travel behaviour. It suggests that calculations using the methods in this study are more straightforward and cheaper than before and can be applied on a large scale.

Therefore, assigning preferences with the help of pictures is extremely meaningful in tourism.

The rest of the paper is organized as follows. In Section 2, this study examines recommendation methods in the travel domain and gives an overview of the literature on travel behaviour patterns. In Section 3, this study describes the research design, i.e. how the scale analysis was used to capture different travel preferences and the experiments to link pictures, preference factors, POIs and users. It lays the foundation for this study's picture-based recommendation system. In Section IV, this study evaluates the results of the model and gives an overview of the system architecture. In Section 5, this study summarizes the modelling process of this new recommendation algorithm and provides suggestions and outlooks for the improvement of the algorithm.

2. Literature Review

2.1 Travel Preference Model

Only a few applications on today's online travel or review sites offer recommendations based on previously identified types of travel. One example is the TUI website [4]. The TUI website recommends hotels for its users based on seven exclusive categories: classic, beach, premium, significant, lifestyle, nature and scene. Users are advised to choose one or more hotel categories only after answering three questions and selecting up to eight priorities (e.g., nature, Wi-Fi, sports, and entertainment). Travel guide publisher Lonely Planet serves American Express travel activities. This guide has a traveller's type of trip determined by a test consisting of four questions, each with three possible answers. These answers point to five types of travel: Global Gourmet, Shopaholic, Classic Globetrotter, Thrill Seeker and Trailblazer. And for each class, the guide offers only a few destination suggestions. No theoretical framework is introduced in any of these business cases, and thus the rigour of the travel preference model may not be sufficient.

The main goal of this study is to facilitate the elicitation of users' needs and preferences through a method [5] which is only later applied to specific recommendation techniques. However, these systems usually focus on the process of user self-representation. This study argues that recommendation results should be iteratively refined. Users do not have to specify their preferred alternative content from the beginning. In the example of this study, the images reflect the preferences of individual users. The approach in this study has some similarities to preference construction as a model of human decision-making [6]. That is because this study assumes that tourists do not clearly understand their preferences in the early destination selection stage. The elicitation of preferences is thus usually a process constructed through sequential interaction and decision-making. For example, the users in this study had to experiment with a step-by-step selection of a set of pictures.

2.2 Travel Recommendation Algorithm

The goods recommended by traditional recommendation systems are usually referred to as books, movies or music clips, but they can also refer to more complex products or services. The technologies currently used in travel recommendation systems on the market can be summarized in the following categories: content-based travel recommendation, which builds interest descriptions for these goods based on the goods the user has purchased, obtains a user profile, finds goods with similar interests to the user and makes recommendations.

Content-based recommendation algorithms can model user interests and obtain higher recommendation accuracy by increasing the dimension of item attributes. However, when applied to the travel field, few novel travel products are recommended to users, and information about users' previous travel experiences and evaluations is challenging to be captured by the algorithm.

Collaborative filtering-based travel recommendation uses the historical behaviour of other users on products to guide target users, recommending products that are similar to their interests and that they like. This recommendation algorithm can handle complex unstructured objects without domain-specific

ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

knowledge but suffers from (1) data sparsity problem (few user rating data), (2) cold start problem (including item cold start and user cold start), (3) algorithm scalability and (4) ignores users' implicit interests and influencing factors.

The demographic-based travel recommendation discovers the correlation between users based on their basic information, classifies users according to their attributes, and then recommends other products that are similar to users' favourites to the current user. Although there is no need for users to rate products and no cold start problem, the collected user profile information violates users' privacy to some extent. And the managed user profile information may be very rough, and thus, the recommendation accuracy is not enough. Geolocation-based travel recommendation [7-9] exploits the fact that users make different choices in different geographic environments and uses the location context to obtain user preferences more accurately and provide personalized recommendations. This algorithm simplifies the information search and filtering process with the user's location information, making the recommendation results more accurate. Still, it also involves the privacy protection issue of collecting user profiles.

Knowledge-based travel recommendation obtains the attributes of products and user profiles, models the products and users comprehensively, and then uses the knowledge of users and products to derive the products that best meet users' needs. The main reason why the advantage of this algorithm is limited is that it is difficult to obtain enough knowledge to build a knowledge base about users and products.

This study combines existing technologies and practical research scenarios to design a content- and knowledge-based travel recommendation algorithm and validate it through empirical research.

3. Processing of questionnaires and scales

3.1 Questionnaire design

Traditional tests of Big Five personality traits have generally been based on non-specific reference contexts, emphasizing the measurement of general personality traits across environmental contexts. However, in recent years, researchers have gradually questioned this approach, arguing that contextualization-based measures are more meaningful and better take into account the uniqueness of personality factors that may be influenced by different environmental contexts, leading to the emergence of many multi-domain scales based on Big Five personality theory. For example, Pathki et al. (2021) developed the Organizational Big Five Personality Inventory (ORG-B5) for use in organizational scenarios [10]. They confirmed through a series of studies that these scales have better validity in specific contexts than non-contextualized personality tests and are short, efficient, and easily accessible. Based on these findings, this study developed a preference scale in tourism scenarios for better application in the context of travel recommendations.

3.2 Data collection and analysis

Based on the principle of random sampling, the study posted links to questionnaires on WeChat and Weibo, the social networking services with the highest number of users in China, and invited respondents who had travel experiences in the past year to participate in an "anonymous" manner. "The questionnaire is voluntary. In the questionnaire, firstly, all the questions were set as compulsory questions to prevent respondents from missing answers. The second is to set the answering device monitoring, that is, the same computer or cell phone can only answer once, to avoid the situation of duplicate answers. To further ensure the data quality, the researcher removed the questionnaires with short response time and obvious answer patterns in the questionnaire system. A total of 290 questionnaires were finally collected, with 282 valid questionnaires collected from May 25, 2022, to June 8, 2022.

3.3 Scale reliability analysis

3.3.1 Reliability analysis

Table 1: Results of confidence analysis

Reliability statistics	
Cronbach Alpha	Number of items
. 893	25

In this paper, using SPSS 26.0, Cronbach's α coefficient was used to judge the validity of the tourism preference scale. The coefficient value of Cronbach's α reliability of all variables of the 5-level Likert

ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

scale consisting of 25 question items was 0.893's (as shown in Table 1), indicating that the scale has good internal consistency and high quality of data reliability, which can be used for further analysis.

3.3.2 Validity analysis

Table 2: Results of validity analysis

KMO and Bartlett's test		
KMO The number of sample suitability measures.		.811
Bartlett sphericity test	Approximate cardinality	2198.734
	Degree of freedom	300
	Significance	.000

As shown in Table 2, the questionnaire KMO value was 0.811 with a significance of 0.000, indicating that information can be effectively extracted from the data.

3.4 Results of the questionnaire analysis

3.4.1 Descriptive statistical analysis

The demographic variables of the questionnaire data were analyzed descriptively using SPSS 26.0 to understand the basic information of the sample. In terms of the gender composition of the sample, there were 115 males (40.96%) and 167 females (59.04%), and the overall number of differences was not significant.

Regarding age composition, 91 people aged 25-28 were in the sample, accounting for 32.45% of the total sample, followed by the group aged 17-20, accounting for 30.32% of the whole sample. The main population of the sample is the group of college students aged 17-20, and the group of newcomers to the workplace aged 25-28.

In terms of education level, the largest number of people is undergraduate, with 246 people, accounting for 87.23% of the total sample; college, master's and doctoral groups account for 11.17%, 9.57% and 1.06% of the whole sample, respectively. The overall education level of the sample population is high, concentrated in the undergraduate population.

In terms of marital status, the main body of the sample was unmarried people, accounting for 87.23% of the total sample; while only 36 people were married or cohabiting, accounting for 12.77%.

3.4.2 Factor analysis

Table 3: Component matrix after rotation

		Ingredients				
	Title item	1	2	3	4	5
1	I often have endless dreams about the destinations I will visit.	.908				
2	I have a spirit of adventure that others do not have.	.818				
3	I'm eager to try something new, even if they're related to.	.811				
4	I'm a risk taker and out of the box person.	.783				
5	My mind is often filled with vivid images of travel.	.759				
6	Despite the mixed reviews of the destination, I am confident that he will not disappoint me.		.902			
7	I don't really care if people have bad encounters in their travels.		.784			
8	I was bored with the harmony and ease of the tour.		.754			
9	I was stimulated by the collision and clash of cultures reflected in the destination.		.753			
10	I think that attractions that lots of people have been to are more reliable.*		-.751			
11	I was often able to have fun in the lively spots.			.821		
12	I enjoy social and recreational gatherings during my travels.			.820		
13	Most people think I am a warm and friendly person.			.773		
14	I want to avoid the crowded and noisy environment as much as possible during the tour.*			-.773		
15	I get bored with social situations with lots of people.*			-.719		
16	When making travel plans, I am more concerned with logic and organization.				.807	
17	I often make travel decisions only after careful consideration.				.803	
18	Others think I am a prudent person.				.768	
19	I will do my best to take into account all the circumstances of the trip.				.753	
20	I decide on the trip very casually.*				-.736	
21	I often worry that something terrible will happen on this trip.					.852
22	If this tour makes me feel stressed, I will break down.					.788
23	I am often afraid of the potential dangers of travel.					.785
24	Sometimes I worry that this trip will be unproductive.					.780
25	I seldom feel depressed when I travel.*					-.507

ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

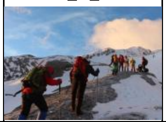
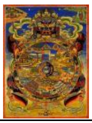
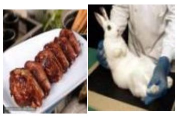
The study extracted five factors by principal component analysis, and the variance explained values of the five factors were 28.960%, 12.276%, 10.379%, 9.337%, and 7.120%, respectively, and the cumulative variance explained after rotation was 68.071% > 50%. It means that the information content of the study items can be effectively extracted. Finally, the factor loading coefficients of all question items in Table.3 are greater than 0.6, which is consistent with expectations and can be further analyzed. Considering the actual meaning of the items, the study summarized the five factors as "stress resistance," "imagination," "newness," "rigour," and "extroversion"and "extroversion".

4. Experimental design

4.1 Establishing connections between POIs, images and visitors

The analysis results of the questionnaire showed that the travel preference scale designed by the study was successful and could clearly express people's travel preferences through the five factors. In the first step of the experiment, this study combined with the opinions of experts from the Shenzhen Institute of Tourism of Jinan University, 25 pictures containing five factors of 5 POIs in random disorder were given, and the results of the picture coding are listed in Table.4.

Table 4: Experimental POI image coding

	Pressure resistance	Imagination	Forced Newness	Rigor	Extroversion
Haba Snow Mountain	1-1 	1-2 	1-3 	1-4 	1-5 
Yarlung Tsangpo Great Canyon	2-1 	2-2 	2-3 	2-4 	2-5 
Sedar	3-1 	3-2 	3-3 	3-4 	3-5 
Nanjing Museum	4-1 	4-2 	4-3 	4-4 	4-5 
Chongqing	5-1 	5-2 	5-3 	5-4 	5-5 

The panel of experts (15 people) was also asked to rank the 5 images of each POI by the degree of association, and the results are shown in Table.5, with the data represented in the table. For example,

$$F_{Sedar} = 0.08*factor1+0.25*factor2+0.33*factor3+0.11*factor4+0.23*factor5 \quad (1)$$

Table 5: POI and preference factor association scoring

	i*Factor1	i*Factor2	i*Factor3	i*Factor4	i*Factor5
Haba Snow Mountain	0.46	0.07	0.19	0.21	0.07
Yarlung Tsangpo	0.28	0.55	0.05	0.06	0.06
Sedar	0.08	0.25	0.33	0.11	0.23
Nanjing Museum	0.03	0.27	0.14	0.41	0.15
Chongqing	0.04	0.14	0.31	0.12	0.39

In the second step of the experiment, a questionnaire was randomly given to 20 people to make the

ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

subjects choose five pictures from a gallery of 25 pictures, where the ith of them (i is the i-th of the five selected by the user) picture was ranked kth. And asking by name only made them choose the most wanted one of the 5 POIs (to validate the model).

In the third step of the experiment, the questionnaires sent to the subjects were recovered. Let $X_i = (5+1-k)/5$ and $X_i \in (0,1)$ express the value assigned by the subject to each selected image. For example, the first image selected by subject A is assigned a value of 1, the second image is assigned a value of 0.8, and so on. To convert the assignment of images to the weight of each factor selected by the subject and to associate the subject in the same order of magnitude as the POI with the factor, all the scores are reassignments to 1/3 of the original values so that the sum of the selection weights of each subject for each of the five factors should be equal to 1.

4.2 Quantifying the relationship between POI and visitors using images as a vehicle

After obtaining the experimental results from the relevant steps in 4.1, the study applied the Euclidean metric to calculate five preference factors simultaneously representing the distance, i.e., the POI and the visitor, and the two were projected into a 5-dimensional space. The POI with the shortest Euclidean distance from the 20 subjects is calculated as the recommended POI.

The distance between Fu and Fp is expressed as

$$d_{(Fu,Fp)}=\sqrt{\sum_{m=1}^{5}\left(f_m^u - f_m^p\right)^2} \tag{2}$$

4.3 Validating the model

The model strengths and weaknesses were verified by checking whether the model recommendation results matched with the subjects' self-selected intention POI results, and the model matching results are shown in Table.6. Due to space limitation, only a part of the samples are shown below.

Table 6: Model matching results

Subjects	POI					Model matching results	Self-selected results	Whether Consistent
	Haba	Yarlung Tsangpo	Sedar	Nanjing Museums	Chongqing			
Zhang*	0.61	**0.11**	0.58	0.69	0.71	Yarlung Tsangpo	Yarlung Tsangpo	Yes
Liu*	0.56	0.33	**0.32**	0.59	0.52	Sedar	Yalu Yarlung Tsangpo	No
Zhou*	0.68	**0.32**	0.46	0.64	0.52	Yarlung Tsangpo	Yarlung Tsangpo	Yes
Wang*	0.46	**0.37**	0.68	0.72	0.8	Yarlung Tsangpo	Yarlung Tsangpo	Yes
Yin*ni	0.44	**0.41**	0.48	0.58	0.51	Yarlung Tsangpo	Yarlung Tsangpo	Yes
Li*	0.83	**0.28**	0.7	0.82	0.85	Yarlung Tsangpo	Yarlung Tsangpo	Yes
Hu*	0.47	0.35	**0.23**	0.51	0.32	Sedar	Sedar	Yes
Lu*yin	0.63	**0.2**	0.47	0.66	0.64	Yarlung Tsangpo	Haba	No
Zhao*	**0.38**	0.48	0.57	0.63	0.61	Haba	Haba	Yes

The model-recommended POI of 6 subjects was not consistent with the self-selected intention POI, and the model-recommended POI of the remaining 14 subjects was consistent with the self-selected intention POI, resulting in a model consistency of 0.7.

5. Conclusion

This paper presents an image-based POI user preference elicitation and recommendation method with the following significant findings.

(1) This study develops a computational model of tourist preferences considering tourist behaviour and personality traits, where tourist preferences are captured in a complex way as a mixture of several

ISSN 2616-7433 Vol. 5, Issue 5: 124-130, DOI: 10.25236/FSST.2023.050520

underlying factors. This approach is a content- and knowledge-based recommendation method.

(2) Tourists in the early destination selection stage are very emotional. Therefore, this study enables tourists to express their preferences with the help of pictures, providing travel decision support for tourists unable to express their needs. It yields a new method of preference elicitation that makes it unnecessary for tourists to verbally represent their interests and opinions or to translate them into product attributes, nor does it require a long interaction with the system.

A limitation of this study's methodology is that the image structure of a picture may trigger different emotional responses and associations for different people, especially in different cultural contexts. In addition, mapping POIs to these five factors currently relies strongly on expert judgment and is quite time-consuming. It makes the research method not currently scaled. In this issue, the study plans to use text mining methods in user reviews to improve the way POIs are configured. In addition, the study intends to extend to other sentiment objects, such as audio samples and video clips, or to use the method to support the group decision-making process.

References

[1] Tang, G., & Zeng, H. (2021). Evaluation of Tourism E-Commerce User Satisfaction. Journal of Organizational and End User Computing (JOEUC), 33(5), 25-41.

[2] Cristobal-Fransi E, Daries N, Martin-Fuentes E, Montegut-Salla Y. Industrial Heritage 2.0: Internet Presence and Development of the Electronic Commerce of Industrial Tourism. Commerce of Industrial Tourism. Sustainability. 2020

[3] Heqing Zhang, Tingting Guo, Xiaobo Su, "Application of Big Data Technology in the Impact of Tourism E-Commerce on Tourism Planning", Complexity, vol. 2021, Article ID 9925260, 10 pages, 2021

[4] Neidhardt J, Seyfang L, Schuster R, et al. A picture-based approach to recommender systems [J]. Information Technology & Tourism, 2015, 15(1): 49-69.

[5] Sertkan, Mete et al. "What is the "Personality" of a tourism destination?" Information Technology & Tourism 21 (2018): 105-133.

[6] Chen L, De Gemmis M, Felfernig A, et al. Human decision making and recommender systems [J]. ACM Transactions on Interactive Intelligent Systems (TiiS), 2013, 3(3): 1-7.

[7] Paüli Agustí D. Characterizing the location of tourist images in cities. Differences in user-generated images (Instagram), official tourist brochures and travel guides [J]. Annals of Tourism Research, 2018, vol. 73, num. November 2018, p. 103-115, 2018.

[8] Kang Y, Cho N, Yoon J, et al. Transfer learning of a deep learning model for exploring tourists' urban image using geotagged photos [J]. ISPRS International Journal of Geo-Information, 2021, 10(3): 137.

[9] Ai Jingchao. Research on personalized travel route recommendation based on improved collaborative filtering technology [J]. Modern Electronic Technology,2019,42(23):182-186.

[10] Pathki C S, Kluemper D H, Meuser J D, et al. The org-B5: development of a short work frame-of-reference measure of the big five [J]. Journal of Management, 2022, 48(5): 1299-1337.