# Research and Analysis on the Intentional Behaviour of Tourists in Scenic Spots Based on the Background of Computerized Big Data

Shangshang Huang *, Lewei Li[a]

East China Jiaotong University,Nanchang,Jiangxi

* Corresponding author: 861964033@qq.com, [a]2272751505@qq.com

**Abstract :**This paper is based on the background of big data research combined with certain tourism scenic development system, the amount of the overall tourist intention of the scenic area has been studied, in the direction of new media research also has a certain integration and analysis. From the scenic area of the new media approach to the study. This paper has made certain research and analysis on the basis of big data combined with the visual impact of landscape design, and has made new research directions and breakthroughs in the combination of historical landscape analysis and computers. The research results of this paper can not only further improve the theoretical system of neighbourhood tourism development and raise the level of evaluation and analysis of such tourism projects, but also provide a more comprehensive and accurate analysis and grasp of the relevant influencing factors, so as to formulate sustainable development strategies and improve its comprehensive development level on the basis of clear scientific development trends. The new media landscape creativity research in this paper meets the higher needs of people in the new era on also does great groundwork for the new media research field.

**Keywords:** Big data algorithms; The historic city; Landscape research; Sustainable development

## 1. INTRODUCTION

In the process of the inheritance and development of human civilization, the historical and cultural district is a typical representative and concentrated embodiment of historical and cultural achievements, and an important carrier of urban memory, a high concentration of material and spiritual culture [1]. The value of the city's cultural symbols is not only reflected in the physical aspect of the buildings and structures, but also in the immaterial aspect of its historical heritage. Historic districts are also important representatives of cultural heritage, so their conservation and preservation will become a major part of today's cultural work [2]. However, in the face of a changing social environment, the preservation and protection of traditional culture is becoming a challenge, and ensuring that historical heritage keeps pace with the times and that traditional culture remains vibrant is therefore becoming a key issue. This study selects the representative historical and cultural district of Tanhualin in Wuhan as the research object, and discusses and analyses the scientific development model of humanistic tourism to seek a scientific way for the sustainable development of the historical and cultural district, so as to provide scientific reference and reference for the inheritance and protection of

ICPDI 2023, September 01-03, Chongqing, People's Republic of China

DOI 10.4108/eai.1-9-2023.2338698

historical and cultural achievements [3]. Humans have always had a close relationship with nature, and the natural landscape has always influenced us all, both physically and spiritually.

The natural landscape has always influenced each and every one of us, both physically and spiritually. In recent years, with the rapid development of the city, people have gradually grown tired of the noise and bustle of the city while conquering nature.With the rapid development of cities in recent years, people have grown tired of the noise and bustle of the city and are eager to 'return to nature, back to the basics'. But nowadays, due to various institutional and design problems in urban constructionHowever, due to various institutional and design problems, the industrial industry has flourished and has encroached on the original green space of the city.The natural visual patches made up of vegetation in the mountain parks and the artificial patches made up of buildings and structures in the urban landscape.The proportion of visual area between the natural visual patches made up of vegetation in mountain parks and the artificial patches made up of buildings and structures in the urban landscape is seriously out of proportion.and the lack of guidance and control of urban colour in the urban planning, the visual patches of the city are seriously mixed up with each other and polluted, resulting in a poor visual coordination of the overall urban.The planning and design of the mountain park ignores the upper level of urban planning, resulting in a fragmentation of the visual landscape connection with the surrounding environment. The planning and design of the mountain parks ignore the upper level of urban planning, resulting in a fragmentation of the visual landscape connection with the surrounding environment and a weak urban landscape integrity. The poor level of designers and the strong ecological sensitivity of some areas have led to poor planting in the mountain parks themselves, with uncoordinated combinations of form and colour between plant patches. The uncoordinated combination of form and colour between plant patches has led to a situation where the park is at odds with the ecological environment and incompatible with the landscape environment, depriving people of the opportunity to enjoy the landscape while living and working. This deprives people of the right to appreciate the beauty of the landscape while they live and work. These various factors, both natural and man-made, have exacerbated.

## 2. RESEARCH METHODS AND BIG DATA ALGORITHMS

### 2.1 Research Methodology

It should be roughly divided into three segments: (1) Obtaining images of landscape resources through fieldwork; (2) the use of questionnaires to obtain quantitative data on the visual perception and behavioural intentions of visitors; (3) using SPSS software to test the questionnaire data for normal distribution, reliability and validity; (4) Regression analysis was performed on the data to obtain the corresponding regression models to analyse which aspects (variables) of visual perception were significant and used [4].

### 2.2 Recommendation algorithms

In order to handle recommendations in complex business scenarios and big data scenarios, the business process of industrial grade recommendation systems is generally divided into three phases: recall, sorting and re-ranking [5]. Recall is the process of removing items that may be of interest to users from a full pool of candidate items. Due to the huge pool of item candidates, the recall models used are often simple and efficient, resulting in fast recall; the ranking layer is responsible for carefully sorting the recalled items according to the probability that a user is likely to click on them, often using complex deep learning models to fully exploit user interests in order to improve accuracy; and re-ranking is mainly the adjustment of the final recommendation list according to some complementary policies and business metrics [6]. The three stages in tandem form the pipeline architecture of the recommendation algorithm.


**Figure 1.** Recommendation algorithm pipeline architecture

### 2.3 Classical recall layer model

(1) Family of collaborative filtering algorithms

The collaborative filtering algorithm is recognised as the most classical recommendation algorithm in the field of recommender systems, dating back as far as the Xerox Research Centre's mail screening system in 1992 [7]. The recommendation principle can be divided into two main categories: user-based and item-based.

Collaborative filtering relies entirely on the behavioural relationships between users and items to make recommendations, the idea of which can be summarised as "things come in groups, people come in groups" [8]. The principle of item-based collaborative filtering is that "things are clustered together", i.e. the items that are most similar to the items that the user has acted upon are recommended to the user. The preference of user u for an item v, $sim(u, v)$, can be expressed as:

$$sim(u, v) = \sum_{v_i \in V} score(u, v_i) \times sim(v_i, v) \tag{1}$$

where V is the set of items for which the user has generated behaviour, $score(u, v_i)$ is the degree of preference of user u for item , $v_i$, and $sim(v_i, v)$ is the degree of similarity between item , $v_i$ and item v.

The principle of user-based collaborative filtering is that "people are divided by groups", i.e., the items most similar to the user are recommended to the user, and the preference of user u for an item v, $sim(u, v)$, is expressed as:

$$sim(u, v) = \sum_{u_i \in U} sim(u, u_i) \times score(u_i, v) \tag{2}$$

where U is the set of similar users of user u, $sim(u, u_i)$ is the degree of similarity between user u and user $u_i$, and $score(u_i, v)$ has the same meaning as equation 1.

The above two equations need to calculate the similarity between users or between items, in the co-occurrence matrix users and items are reflected in the form of row vectors and column vectors respectively (as shown in Figure 2), the similarity between vectors can be calculated using the cosine similarity method of equation 3:

$$sim(i, j) = \cos(i, j) = \frac{i \cdot j}{\|i\| \cdot \|j\|} \tag{3}$$

Collaborative filtering is simple, intuitive and interpretable, but weak in generalisation. Top items are easily associated with a large number of items and thus similar due to the high number of reviews, leading to a significant head effect, while long-tail items have few reviews and a sparse vector, making them difficult to associate with other items and leading to few recommendations [9].

The matrix decomposition algorithm generates a hidden vector for each user and item so that users and items can represent similarity in a space of hidden vectors of the same dimensionality, which is essentially a decomposition of the co-occurrence matrix into the form of a product of hidden vectors [10]. The hidden vectors are denser and therefore enhance the ability to handle sparse matrices.

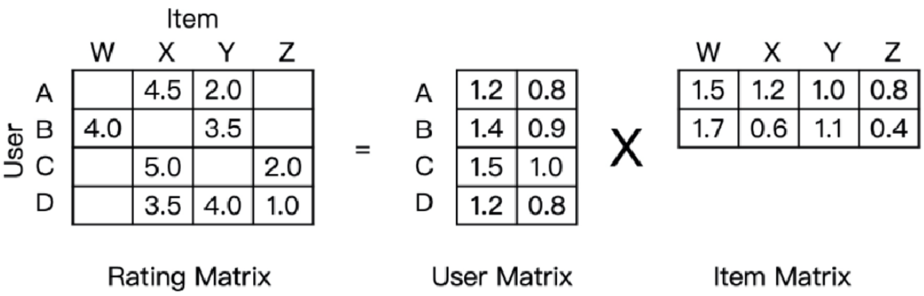


**Figure 2.** Schematic diagram of matrix decomposition

Eq. 4 shows the decomposition of the mxn-dimensional co-occurrence matrix R into the product of the user matrix $U_{m\times k}$ item matrix $V_{k\times n}$. where m is the number of users, n is the number of items and k is the hidden vector dimension. The dimension size of k determines the expressiveness of the hidden vector, and once obtained the hidden vector can be recommended using the similarity method of equations 1 and 2.

$$\mathcal{R}_{m\times n} = U_{m\times k} * V_{k\times n} \tag{4}$$

Spark MLlib uses Alternating Least Squares (ALS) to solve for matrix decomposition hidden vectors, and we describe its basic principles.

Let user $u$'s prediction score for an item $v$ be $\hat{r}_{uv} = p_u * q_v^T, \Delta r = r_{uv} - \hat{r}_{uv}$ indicating the error between the true value and the predicted value, the smaller the $||\Delta r||$, the more accurate the prediction. Transforming the above problem into an optimization problem to find the minimum of $||\Delta r||$ by adding the regularization term, the objective function obtained is shown below:

$$\min_{p^*,q^*} \sum_{(u,v)\in A}(r_{uv} - p_u * q_v^T)^2 + \lambda(||p_u||^2 + ||q_v||^2) \tag{5}$$

ALS finds the minimum value by alternating optimization as follows:

I. Generate random initialization values for $p_u$, $q_v$.

II. Fix $p_u$ and solve for the minimum value of $q_v$ by gradient descent.

III. Fix the current $q_v$ and solve for the minimum value of $p_u$ by gradient descent.

IV. Repeat steps II and III until the objective function converges or the maximum number of iterations is reached.

(2) YouTube DNN

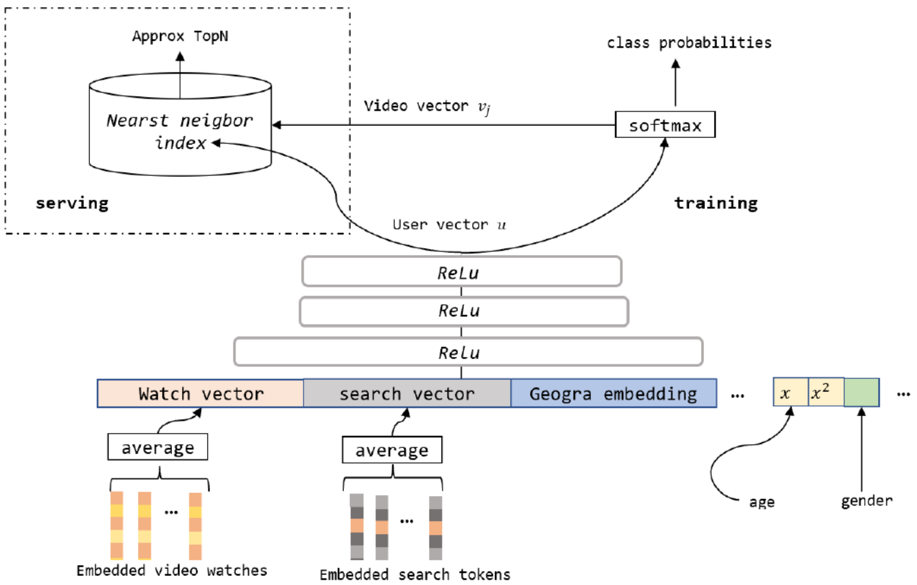


**Figure 3.** YouTube DNN network structure

As show in figure 3.YouTube is the world's largest video sharing site, with the vast majority of its content coming from UGC. YouTube DNN serves personalised recall tasks across YouTube's million-strong video base.

The overall structure of the YouTube DNN recall model is relatively simple, the input layer consists of user viewing sequences, average embedding vectors of search word sequences, where the embedding vectors of both video and search sequences are generated by the Word2vec method; in addition, features such as geolocation embedding vectors, age and gender are also input; all features are connected and input into the Relu fully connected layer for training, and then the user vector u is obtained; finally, it is then passed through the softmax layer to obtain the viewing probability of each video.

The YouTube DNN separates offline TRAINING from online SERVING, which is the essence of this recall model. The YouTube DNN converts the recommendation problem into a multi-classification problem by softmax. When offline TRAINING, the user vector u is inner-producted with the embedding vector $v_j$ of each video, and the probability of possible viewing of each video by the user is obtained after softmax, and then TopN is taken for recall as shown in Equation 6.

$$p(w_t = i|U, C) = \frac{e^{v_i u}}{\sum_{j \subset V} e^{v_i u}} \tag{6}$$

If 100W videos are available at training, then the recall model becomes a 100W classification problem, so negative sampling is engineered to speed up training.

## 3. IMPACT STUDY RESULTS

### 3.1 Data collection

In accordance with the research design described above, two aspects were clarified in turn: firstly, how many landscape resources had an impact, expressed in a graph; secondly, how many people were selected as subjects, the proportion of men and women, the age structure, the number of questionnaires, the return rate, the effectiveness rate, etc., and the data needed to be given.

### 3.2 Normal distribution test

The visual perception factors (independent variables) and the behavioural intentions of visitors (response variables) were tested for normal distribution for each aspect of the different landscape scenes (types). A test of normal distribution for all data is required to select the appropriate correlation coefficient for the correlation analysis. After implanting the data into SPSS Statistics 26, a one-sample Komogorov- Smimov (K-S) test was used. Use Pearson's Correlation Coefficient if normally distributed, otherwise use Spearman Correlation Coefficient.

### 3.3 Reliability and validity tests

As this is a questionnaire scale, it is important to test for reliability and validity. Firstly, the internal consistency of the scale was tested for reliability using Cronbach's alpha coefficient, Spearman-Brown discounted half confidence coefficient, and McDonalds omega coefficient. Secondly, the structural validity of the section was tested using exploratory factor analysis, including KMO, Bartlett's sphericity, variance explained and commonality, and the difference was considered statistically significant at $P < 0.05$. statistically significant. Finally, validation factor analysis tests were carried out for each factor and each analysis term, including the AVE, CR and AVE square root.

### 3.4 Principles for the selection of indicators

(1) Completeness and objectivity.

Completeness refers to the selection of indicators that can summarize the evaluation object as a whole and reflect its overall characteristics; while objectivity refers to the reflection of the characteristics of the evaluation object without subjective bias as far as possible, to avoid being influenced by the indicator system

(1) Objectivity means that,as far as possible, there is no subjective bias in reflecting the characteristics of the evaluation object and avoiding being influenced by the subjective will of the system maker;

(2) Principal component and independence.

It is required that the selected indicators can both summarise the content to the greatest extent and have the least correlation among them, i.e. the indicators are comprehensive and independent of each other;

(3) Accessibility and measurability.

(4) Sensitivity and accuracy.

(5) Dynamicity and stability.

According to the knowledge of spatio-temporal weighted neural networks (STWNN), the spatio-temporal weight matrix is defined as shown in the following equation:

$$W(s_i, t_i) = STWNN({Mp_{veci}}^{ST}) \tag{7}$$

where ${Mp_{veci}}^{ST}$ is a vector of spatio-temporal proximity-specific expressions for a location point i generated by the spatio-temporal proximity deep neural network.

In this paper, we ignore the time factor and consider only the space, so we obtain the expression for the spatial weight matrix, defined as follows:

$$W(s_i) = STWNN({Mp_{veci}}^{s}) \tag{8}$$

Where SWNN refers to Spatial Weighted Neural Network (SWNN); ${Mp_{veci}}^{s}$ refers to a specific expression vector of spatial proximity relations for a location point i, which is generated by SPNN.

In addition to studying whether the GNNWR model is applicable to the modelling of geospatial relationships, this paper also focuses on whether the spatial weight kernel function is reasonably effective for the construction of neural networks and the feasibility analysis of carrying out spatially weighted neural networks [8]. Therefore, to better compare the spatially weighted neural network with the spatial weight kernel function for the analysis, for the input of the spatial proximity relationship in equation (9), this paper uses the traditional Euclidean spatial distance, which is the same as the GWR model, i.e:

$$W(s_i) = SWNN([d_{i1}^s, d_{i2}^s, \dots d_{in}^s]) \tag{9}$$

In summary, the procedure for calculating the fitted value $\hat{y}_i$ for the ith sample point of the GNNWR model is shown in Figure 1, where the dashed boxed part is the spatially

## 4. CONCLUSION

The purpose of the construction of historical and cultural districts in the context of cultural and tourism integration is to explore cultural tourism resources and provide comprehensive assistance to the city in preserving its history and culture, as well as being a witness to the development of the city and the crystallisation of human civilisation, and reflecting the cultural changes of the city in its long history in both material and immaterial forms. As the history and direction of change varies from city to city, many cities retain historic and cultural districts with a unique and special charm, and preserving this part of the cultural heritage is a way of preserving the culture of the city and the country. In the context of contemporary tourism development processes, the importance of historical and cultural districts is becoming increasingly evident, and the effects and influences generated are of great research value to urban tourism development. Therefore, in the study of the historical and cultural district of Tanhualin, the current situation of its development and the problems associated with the process are summarised, and the comprehensive effect of tourism development is assessed through the introduction of the concept of sustainable development, so as to evaluate the comprehensive effect of the historical and cultural district of Tanhualin in terms of tourism development, and the results are more intuitive and comprehensive.

## REFERENCES


[1] Wu Xinrui, Dai Xiangyi. Significance and construction of cultural landscape symbols in historical districts under tourism utilization: the case of Qianmen Street in Beijing[J]. China Ecotourism,2023,13(01):111-127.

[2] Wang Juan, Zhao Jie, Feng Jiejie. Study on the correlation between street structure and spatial distribution of tourism elements in the historic district of Qingdao[J]. Journal of Northwest Normal University (Natural Science Edition),2022,58(06):61-69.DOI:10.16783/j.cnki.nwnuz.2022.06.009.

[3] Zheng Ruihong, Ma Yan. Evaluation of the appropriateness of tourism value development in urban historical districts[J]. Economic Forum, 2022(04):55-65.

[4] Zhang Yufen,Wang Jinqiao,Wang Qiang. A study on the perceptions and attitudes of residents in historical districts towards tourism development: A case study of the ancient commercial port of Chikan, Zhanjiang[J]. Business and Management,2022(08):187-192.DOI:10.16517/j.cnki.cn12-1034/f.20220328.002.

[5] Cai, Wina, Guan, Xiaoqi. Research on the effect of tourism development in historical and cultural districts based on network texts--a case study of Zhongshan Fourth Road in Chongqing[J]. China Business, 2022(05):48-52.DOI:10.19699/j.cnki.issn2096-0298.2022.05.048.

[6] Jiang Nai Bin. Classification of historical and cultural districts in Shandong Province and feasibility study of cultural tourism [J]. Urban architecture,2021,18(30):64-67. doi:10.19892/j.cnki.csjz.2021.30.19.

[7] Li Chengjun,Luo Yixin. The development of cultural tourism in historical districts from the perspective of experience economy--Liwan District, Guangzhou City as an example[J]. Modern Business,2021(23):88-91.DOI:10.14097/j.cnki.5392/2021.23.027.

[8] Ren Yanyi. Design strategies for the spirit of place in historical and cultural districts in tourism development: the example of the old street of Hohhot City on the Seaside[J]. Architecture and Culture,2021(03):103-104.DOI:10.19875/j.cnki.jzywh.2021.03.036.

[9] Mazza davidd , Chen Lei, Zhou Jun Gang. A genealogical perspective on the renovation of the historical style and appearance of the neighbourhood, taking Nanchang Greenland Xiangnan Centre Jingxiang cultural tourism neighbourhood as an example[J]. DOI:10.13717/j.cnki.ta.2021.01.024.

[10] Mohanmod riza. Cultural communication-oriented tourism bag design for historical districts [J]. Design,2020,33(21):119-121.