WILEY | Hindawi

Research Article

Optimization of Tourism Information Analysis System Based on Big Data Algorithm

Jing Yang,[1] **Bing Zheng** (ID)**,**[1] **and Zhenghua Chen**[2]

[1]*School of Information Engineering, Hainan Vocational University of Science and Technology, Haikou 571126, China*
[2]*School of of Finance and Economics, Hainan Vocational University of Science and Technology, Haikou 571126, China*

Correspondence should be addressed to Bing Zheng; 21227082@zju.edu.cn

On the basis of ecological footprint theory and tourism ecological footprint theory, the sustainable development indexes such as ecological footprint, ecological carrying capacity, ecological deficit, and ecological surplus of the research area were calculated and the long-term change pattern of each index was analyzed. This paper shows that the ecological footprint of the research area increases year by year, but the ecological footprint is always smaller than the ecological carrying capacity, indicating that the area is still in the state of sustainable development. However, the per capita ecological surplus shows a decreasing trend year by year, indicating that the sustainable development of the region is getting worse. This paper proposes a reordering method of tourist attractions based on heterogeneous information fusion, and realizes the retrieval and reordering of tourist attractions based on user query and fusion of heterogeneous information, so as to help users make travel decisions. In view of the shortage of tourism commercial websites to passively provide scenic spot information, this paper puts forward a scenic spot retrieval method based on query words to enable users to obtain scenic spot information according to their needs, and constructs a tourist consumer data analysis system. The preprocessing methods and methods adopted by the data preprocessing module are analyzed in detail, and the algorithms used in the travel route analysis and consumer spending ability analysis are described in detail. The data of tourism consumers are analyzed by this system, and the results are evaluated.

1. Introduction

Tourism information is the basic data guarantee of tourism informatization, so it is particularly important to obtain high-quality tourism data resources. Tourism information not only exists in a large number of government tourism departments and tourism companies, but also with the rapid development of social media in recent years, and there are also a large number of tourism information resources available on the major social media websites. However, it is difficult for people to accurately obtain data from the vast amount of travel information, especially for the diversity of travel destinations or dazzling many travel routes. The current tourism information service simply displays the information of tourist attractions on the Internet, and its disadvantages are mainly reflected in the following aspects: (1) the information is not comprehensive enough, and the rich tourism information shared on social media is not fully utilized; (2) information is only passively presented to users and cannot be searched and filtered according to users' needs; (3) ignore user needs, fail to consider user context information, and fail to achieve customized and personalized services for users. Without intelligent consideration of the comprehensiveness and preferences of the information, it is often impossible to recommend satisfactory results for users. These disadvantages have seriously restricted the acquisition of high-quality tourism information, the recommendation of destinations in line with the needs and personalized planning, and customization of a series of tourism behaviors. Therefore, in addition to mastering huge tourism information resources, tourism informatization also needs to deal with mine and analyze these massive data professionally.

No matter at home or abroad, tourism route planning is an important part of tourism planning and design [1, 2] and the core of travel agency management. With the continuous development and innovation of tourism, many new tourism modes have emerged, such as self-help travel, self-driving travel, and experiential tourism. Tourists can choose scenic spots and routes according to their preferences, and their tour preferences are more personalized. However, tourist routes planned by travel agencies are mostly for the purpose of making profits, and the design of relevant tourist routes is mainly based on factors such as time, traffic, and scenic spots, and other factors in specific profit space in specific areas [3, 4]. At present, tourism route planning mostly focuses on regional tourism planning [5, 6], which is aimed at a large tourist scenic spot. Within the scenic spot, there are a set of scenic spots located in different spatial positions. Different sightseeing modes can be obtained through tour sequence design of these scenic spots. However, the route planning method for scenic spots is not applicable to the planning of a large area such as scenic spots in the city [7]. In the field of tourism, users usually share their experiences and comments after a trip, forming a large number of user-generated content including user comments, photos, travel notes, and other information. These data provide great convenience for trip planning [8, 9]. While there may be noise or bias in a single comment or travel note, incorporating a large amount of user-generated content as a whole can effectively capture the essence of a site. Therefore, in an increasing number of studies, spatial analysis and data mining technologies are used to analyze these contents [10, 11], to obtain users' relevant preferences and historical track information, and to find the similarity between tourists, so as to realize the recommendation of tourist routes [12, 13]. With the advent of Web 2.0 era, online multimedia sharing websites have become popular. The information uploaded by users contains a large number of travel-related contents, which can be widely used in the tourism system. Therefore, in recent years, many intelligent tourism systems have been established to achieve accurate retrieval or personalized recommendation functions through the analysis and mining of tourism multimedia information, thus making travel more convenient and faster [14, 15]. Wikitravel is an early tourism information system before the advent of Web 2.0 era, providing users with open, complete, real-time, and trusted tourism information [16, 17]. It provides online travel services, devotes itself to mining high-quality travel photos from image sharing websites, and designs a user interface with search function and map positioning function, which can provide help for users to plan their path and travel [18, 19]. By analyzing more than 110,000 images with geographical marks on Flickr, the visual perspective of scenic spots is generated by using the images, and the diversity of scenic spot search results is satisfied [20, 21]. Combined with text, geotagged images, and video, the scenic spot summary is generated, and then personalized scenic spot summary is recommended to users by means of query [22, 23]. Another type of system focuses on the retrieval and recommendation of tourism multimedia information. Through the analysis of photos and text information, as well as the relevant knowledge in Yahoo travel, the popular scenic spots are recommended for users, and the summary information of scenic spots is returned to users [24]. Through the acquired images and travel notes, the routes within and between scenic spots are mined, so as to provide users with travel route planning [25, 26]. A mobile travel search framework is proposed, which can display multiple perspectives of scenic spots based on image information to users through compression transmission technology [27, 28]. Low resolution query images are processed by remote server, and then scenic spots are identified and searched, and the corresponding scenic spots are reconstructed from the 3d perspective based on the photo set [29]. In order to improve the performance of the system and overcome the limitations of traditional recommendation algorithms, a hybrid recommendation algorithm has gradually become popular. It uses two or more recommendation algorithms by mixing, weighting, switching, cascading, and feature combination to make full use of the advantages of each recommendation algorithm to obtain higher performance. The most common examples of hybrid recommendation algorithms are the hybrid collaborative filtering algorithm and other recommendation algorithms to alleviate cold start and data sparsity problems.

A reordering method of tourist attractions based on heterogeneous information fusion is proposed. For analysis of the current attraction retrieval method and the deficiency of the need to solve the problem, this chapter then introduces the algorithm implementation of the block diagram; then, it is introduced based on the content and the heterogeneous information mining method based on the score, as well as to the scenic spots, based on the query of the initial sorting method, based on the content of the resort sorting method, and based on the score of the reorder adjustment method. Finally, the objective and subjective experiments verify that the proposed method of scenic spot reordering based on heterogeneous information fusion can efficiently obtain scenic spot information based on user query. Data stored in the database may reflect noise, anomalies, or incomplete data objects. These objects may have some adverse effects on the analysis process, resulting in the over-adaptation of data to the constructed knowledge model or the failure of mining analysis. As a result, the patterns found can be very inaccurate. Data cleaning methods and data analysis methods to deal with data noise and outlier mining methods to find and analyze abnormal situations are required.

2. Optimization Method and Framework of Tourism Information Big Data Analysis

The ecological carrying capacity and sustainability of an area are affected by many aspects, such as meteorology, hydrology, geology, environment, social economy, and other industries and fields. The change of ecological carrying capacity and sustainability is not only related to the state of each single factor, but also to the result of the interaction of all factors. This chapter will introduce the monitoring

methods and principles of all aspects of the scenic and historic interest area and its surrounding areas in detail from three aspects of ecological carrying capacity and sustainability assessment. Among them, the three aspects mainly include the mining and analysis of the laws of the long time series of tourism ecological environment elements, the evaluation of ecological carrying capacity and sustainable is based on the ecological footprint theory, and the trend prediction and early warning based on the big data technology. The specific frame system is shown in Figure 1. First of all, the long time series multisource heterogeneous original data are collected and sorted. On this basis, the typical tourism ecological environment elements are extracted based on the extraction principles and methods of thematic information. Secondly, for extraction of tourism ecological environment factors of single factor, long time series, the trend of the discussion, and analysis, the main tourism ecological environment factors, including land use cover, vegetation coverage, biodiversity, landscape vulnerability, climate comfort level, the level of economic development, the industrial support ability, tourist reception capacity, swim in proportion, and attractive tourism resources, lay the foundation for subsequent driving force of the ecological footprint analysis. Thirdly, the ecological footprint theory, which is widely adopted and highly recognized in the world, is used to discuss the long-term changes of sustainable development indicators such as ecological footprint, ecological carrying capacity, and ecological deficit or ecological surplus in the research area. Finally, the big data analysis method is adopted and two kinds of time series prediction models (the ARIMA model and LSTM model) are used to predict and warn the future sustainable development status of the research area.

2.1. Analysis of Tourism Ecological Environment Elements.

Vegetation coverage is the ratio of the vertical projection of stems, leaves, branches, and other vegetation onto the ground to the total area of the statistical area, which is usually expressed by percentage. Like the normalized difference vegetation index, it is one of the important indicators to measure the growth status of surface vegetation, and it is of great significance to the regional ecological environment assessment. The calculation methods of vegetation coverage mainly include pixel dichotomy model, regression model, and vegetation index method. Among them, the vegetation coverage calculated based on binary pixel model has been widely used by researchers. The assumptions in the binary model divide each pixel into two parts, one with vegetation coverage and the other with no vegetation coverage. The electromagnetic spectrum information observed by the remote sensor is calculated by the linear weighted sum of the two parts, and the weight of each part is related to the area proportion of the part in the whole pixel, then the vegetation coverage of the pixel is equal to the percentage of the surface covered by vegetation in the whole pixel area.

Because the change of land use cover type directly affects the change of biodiversity, the land use cover map obtained by remote sensing can reflect the difference of regional biodiversity to a certain extent. The weights of various surface feature types are shown in Table 1, including cultivated land, construction land, woodland, water area wetland, grassland, and unused land.

The ratio of tourists to residents refers to the ratio between the total number of tourists to the scenic spot and the total number of local residents, which is an indicator reflecting the psychological carrying capacity of local residents. The development of tourism influences the social culture of local residents to a certain extent, which has positive and negative effects. On the positive side, more tourists will promote the local economic development, increase the employment rate of local residents, and greatly improve people's life. The negative aspect is that the increase of tourists will put pressure on the local ecological environment resources and make the living environment of local residents worse than before. In general, the greater is the density of tourists, the stronger the impact. From the perspective of tourism development, the most essential value of tourism resources lies in its attraction to the whole tourism market, that is, its ability to attract tourists. The core competitiveness of tourism resources development is the attraction of tourism resources, which is affected by many factors, including the quality and richness of tourist attractions, tourist traffic conditions, reception facilities, services, and accommodation. The quantitative expression of the only objective standard to measure this value is the number of visitors the tourist resource can attract.

In order to make up for the limitation of search engine using text to retrieve images, visual features are used to reorder images to make up for the semantic gap between text and image. In addition, due to the different sensitivity of images to different visual features, multivisual features are combined to generate mixed features for reordering. At the same time, in order to ensure the correlation between query words and reordered images, a reordering framework based on the graph model is adopted to complete the reordering of image search results, so as to help users get the most relevant images from a large number of search engine returned images. The proposed image search reordering method based on the mixed feature graph model is to mix the visual features and then use the reordering framework based on the graph model to complete the reordering. Figure 2 shows the block diagram of the image search reordering method based on the hybrid feature graph model, which is mainly divided into two parts: learning hybrid feature and graph-based reordering. In the offline learning stage, all images, after visual features are extracted, use potential semantic analysis to learn the mixed features based on visual feature fusion. In the online sorting stage, a query word is given, and the initial sorting result of the returned image is obtained after matching with the text information of the image. Then, the similarity between the images based on the mixed features is calculated under the reordering framework of the graph model to complete the construction of the graph model and finally the reordering result is given.

In order to verify the proposed method and ensure the higher correlation performance of the image in front, the image number is selected here. In different cases, find its
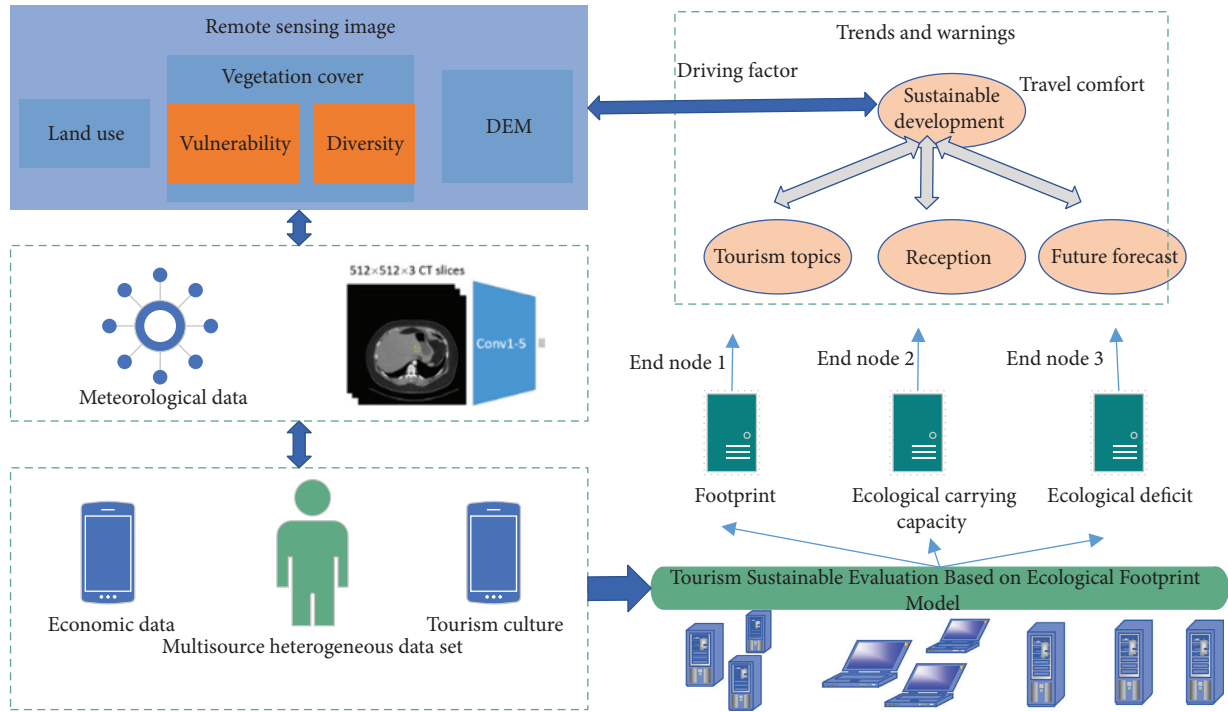
FIGURE 1: Frame diagram of big data analysis of tourism information system.

TABLE 1: Biodiversity weight table.

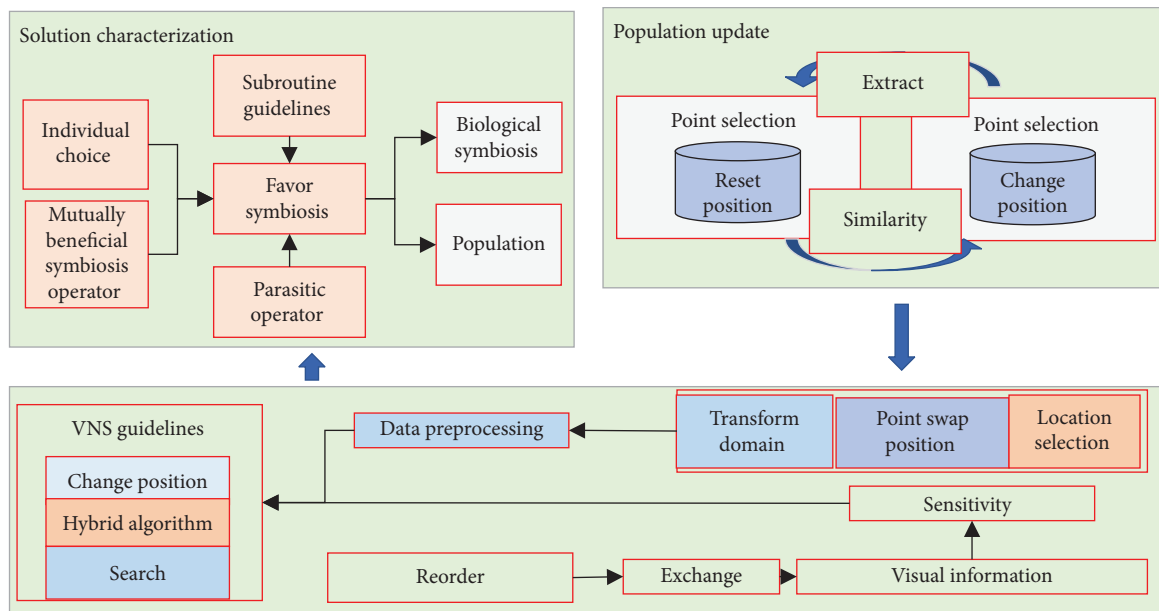| Types of objects	Woodland	Grass	Waters	Arable land	Construction land	Unused
Weight	0.32	0.24	0.26	0.13	0.03	0.02

FIGURE 2: Block diagram of image search reordering based on the hybrid feature graph model.

NDCG value. The change in the NDCG value is shown in Figure 3.

The NDCG value is the average of the NDCG values for the 20 query words used in the experiment. It can be seen from the figure that the effectiveness of the proposed method is still higher than that of the comparison method, which again indicates that the proposed method can ensure a strong correlation between the reordering images and the
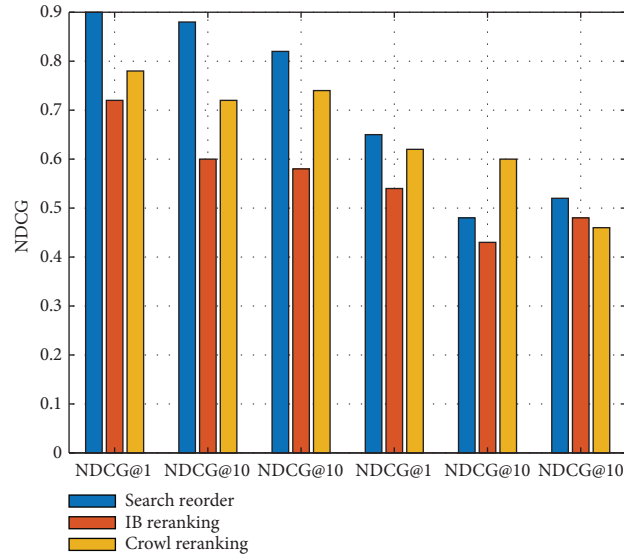
Figure 3: Comparison of reordering results.

previous ones. At the same time, it can be seen that as the value increases, the NDCG value decreases. This is because after reordering, images with low correlation are ordered in the back, thus affecting the NDCG value, which also meets the requirements of reordering.

2.2. Research on Big Data Analysis Algorithm of Tourism Route Planning.
In order to carry out short-time travel route planning, this paper uses the basic idea map solution such as short-time travel route planning algorithm. The detailed solution steps are as follows:

(1) First, the user enters six input values according to his/her own time arrangement and travel preference: departure date, duration of visit, departure location, must-see scenic spots, category of scenic spots of interest, and transportation mode

(2) Secondly, use the input values in the previous step and combine with the machine learning regression algorithm to carry out the training of short-time travel route scoring model

(3) Then, according to the input value of tourists, the routes in the short-time tour route library are conditionally screened and the routes in the screening results are scored using the short-time tour route scoring model

(4) Finally, sort the scores in the previous step and output the top 10 tourist routes with the highest scores

In view of the shortcomings of ID3 algorithm, our solution idea is to record the number of records satisfying the condition of the path from the node to the root node while generating the node of the decision tree, so as to solve the predictive ability problem of the decision tree in the case of missing attribute values.

The data processing capacity of the ID3 algorithm is reduced by introducing minimum support and minimum confidence. In fact, the training sample data set have many rules, but not all of the rules has practical application value, therefore, through the association rules in the analysis of the concept of minimum support and minimum confidence, introducing part will not often appear data filtering for the rest of the operating data of ID3 algorithm.

Rules extracted by the ID3 algorithm can solve the type of data object attribute values are given identification problem, but if not all, of the given attribute values, then the rule is obtained by decision tree will not be able to give the judgment result, and we are through at the node of decision tree, at the same time record the content from the node to the path of the root node of the condition number of records, can effectively solve the missing attribute value of the decision tree's ability to predict:

(1) The support degree of decision attribute to category identification attribute is used to reduce the scale of training set processed by the ID3 algorithm

In the process of data processing, the results often have a lot of redundancy. The most important manifestation is that in the generated decision tree, the decision tree has too many branches, which makes the tree too large and too cumbersome. In this way, the decision information obtained will be too complex and complicated, and many unnecessary rules will be generated. The minimum support degree is used to effectively control the number of possible branches of decision nodes layer by layer.

When the support degree is less than the specified minimum support decision makers in the tuple filter, its basic idea is the decision tree branch. If the current layer corresponds to a subset of a tuple of the category attribute support small, which is the next layer of support smaller, the corresponding tuples in

this group do not need to generate a new branch. It can be seen that in the process of decision tree generation, the pruning of branches that may be generated but have less practical value is carried out according to the value of each decision attribute. Since the size of the child value to some extent represents the amount of effective information contained in the attribute, our processing work is to a large extent based on the amount of information contained in the decision attribute. Since the information contained in the discarded data is relatively small, the above processing ensures the information content of the data to a certain extent.

(2) The redundant branches of the decision tree generated by the minimum confidence level scavenging D3 algorithm

By using the minimum support degree, we reduce the amount of data processing in the process of decision tree generation. However, in the construction of decision tree, the knowledge contained in some branches of the decision tree is too unreliable to have application value. We use the minimum confidence to cut off the branches with low confidence in the decision tree generated by the ID3 algorithm, so as to reduce the scale of the decision tree and make the generated decision tree more practical.

We combine the above two improvements on the ID3 algorithm in one processing process to obtain the following ID3 improved algorithm, as shown in Algorithm 1.

3. Big Data Analysis and Forecast of Tourism Information

Since the per capita ecological footprint reflects the amount of resources consumed by a single individual in an area, the per capita ecological carrying capacity represents the sum of the amount of resources that the area can provide to each individual and the amount of resources needed to deal with the waste it produces. Therefore, from the perspective of regional development, per capita ecological deficit and per capita ecological surplus appear to be very critical and important. In this paper, the time series prediction of the calculated per capita is carried out by using the ARIMA model to explore the regional sustainable development capacity allocated to each individual in the context of rapid regional development and increasingly strong tourism trend. In order to enhance the accuracy, science and rationality of the ARIMA prediction model, input data sets were divided into the training set (80%) and test set (20%).

The ADF test results are shown in Table 2, where the ADF test result is -0.656688, which is greater than the threshold value of significance level from 1% to 10% given. Therefore, the null hypothesis of the existence of unit root is accepted, which further verifies the instability of 1nXt sequence. Therefore, a first-order difference is made for 1nXt to check its trend characteristics, as shown in Figure 4. It can be seen from Figure 4 that the first-order difference series

has no trend characteristics and the time trend is basically eliminated, so it can be considered as a stationary series. The unit root test was carried out on D (1nXt), and the results showed that the null hypothesis of unit root was rejected at the significance level of 0.01, indicating that the first-order difference sequence was stationary series. The test results are shown in Table 3. Therefore, $d = 1$.

Figure 5 shows a block diagram of a tourist attraction reordering method based on social media heterogeneous information fusion. First, heterogeneous travel information was taken from social media, building a database by scraping user reviews and ratings from Tripadvisor, images from Flickr, and site descriptions from Wikitravel. Then, data preprocessing is carried out, denoising heterogeneous information and structural analysis. Secondly, when positioning city is given, users input query words according to their own needs, and the initial ranking results will be given according to text information matching. Then, by analyzing the heterogeneous information of tourism in social media, the reordering framework based on the graph model is used to reorder the features of text and image fusion of scenic spots, and finally, the final ranking of scenic spots is carried out based on the numerical information.

4. Results Analysis

First, experiments are designed to discuss the influence of content-based operations and vision-based operations on the construction of video traceability relationship. Among them, content-based operation detector can be used to find the video of content change, while the visual-based operation detector can be used to find the video of visual perception change from the perspective of human visual perception. For the time-scale transformation detector, the duration statistics are introduced to pseudoclassify similar videos, which can improve the detection speed and accuracy. Figure 6 shows the use of duration in time-scale transformation detection. Known videos (a) and (b) are divided into a group according to the duration of the video, and the detection results show that the video (b) has additional shots. Since videos (a) and (b) have the same grouping in a group of similar videos, they are considered to be the same version, so there is no need to compare them with other groups of videos to save detection time. And this will not detect a slightly longer video (b) as an added shot. Therefore, without the reference of original video, it is very necessary to detect its preclassification.

In Figure 7, the recall rate and precision of the content-based detector are plotted. Here, five thresholds are given for each detector to observe the change of recall rate and precision. Similarly, it can be seen that recall rate and precision rate tend to change with the change of threshold, and in the actual detection of video traceability relationship, the threshold value can be set according to the demand for recall rate or precision rate. There is still room for improvement in the recall and precision of spatial information coverage detection and time scale transformation detection. In the time-scale transformation detection, the matching algorithm of similar video frames can be improved and the

```
Input: R:a set of noncategorical attributes
    D: the categorical
    T: training set
Output: a decision tree
Begin
    If T is null
        Return empty flags or single data point flags
If the records in T all have the same classification mark
        Return the classification value with a single node flag
Assign threshold and confidence
For all attribute X in D
        Calculate the obtained value of (x, T)
        Let g be the maximum value obtained by (x, T)
            Let x = Xi
    Let W be the attribute with the greatest gain
        Calculate the percentage of each class in the data set and the subset of different values of the decision attribute W;
    If the percent value < x then
        Not counted in the queue
    End if
End
```

ALGORITHM 1: D3 NEW AND ID3 IMPROVED ALGORITHM.

TABLE 2: Results of the 1nXt unit root test.

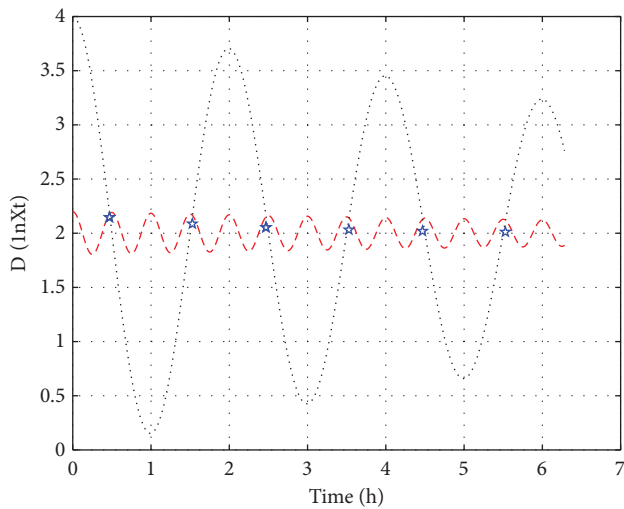| ADF test statistic	−0.656677	1.5% critical value	−3.808548
		5.5% critical value	−3.020683
		10.5% critical value	−2.650414
		One-sided p values	

FIGURE 4: First-order difference sequence diagram.

matching accuracy can be improved. At the same time, in the spatial information coverage detection algorithm, when the coverage information block is very small, it is filtered as noise, resulting in the failure to detect the information coverage.

Table 4 shows the accuracy of some video detectors. For the spatial information superposition detector, it is necessary to find the appropriate threshold. Here, according to empirical values, five different thresholds are selected to draw the recall and precision of operational behavior detection.

The audio conversion detector also gives a good performance value, because most audio conversions in video are full audio replacement, which can be easily detected using existing techniques. For visual based detection, including detection of spatial color change, spatial scale transformation, and visual quality detection, the detector can easily detect the change of spatial scale and spatial color, which has a good accuracy. Moreover, the detector can easily detect the change of spatial scale and spatial color, so it has a good accuracy.

The test uses cross validation to avoid the overfitting of the model. The test data set is divided into 10 parts, one part of which is taken out as the test set and the other nine parts are the training set. The average of the ten test scores is finally output as the final score of the model.

As can be seen from Figure 8, due to the different styles of scenic spots in different cities, the precision and recall rates of the same query words are different. The results of all reordering methods are better than that of the reference method TF/IDF. The results of reordering based on multiple hidden features and reordering based on heterogeneous information fusion are better than that of single modal information, which indicates that the reordering result of multimodal fusion is better than that of single modal information in the reordering problem setting in this chapter. The results of reordering based on the characteristics of multiple hidden topics are good because recall and precision are mainly used to examine content-based relevance. At the same time, the reordering result of multihidden theme features is superior to the reordering result of single mode text or image and their mean value, which means that multihidden theme features not only integrate the important information of text and visual features, but also excavate the potential information between the two kinds of information. Heterogeneous information fusion based on social media sites reordering mainly combines the hidden theme features
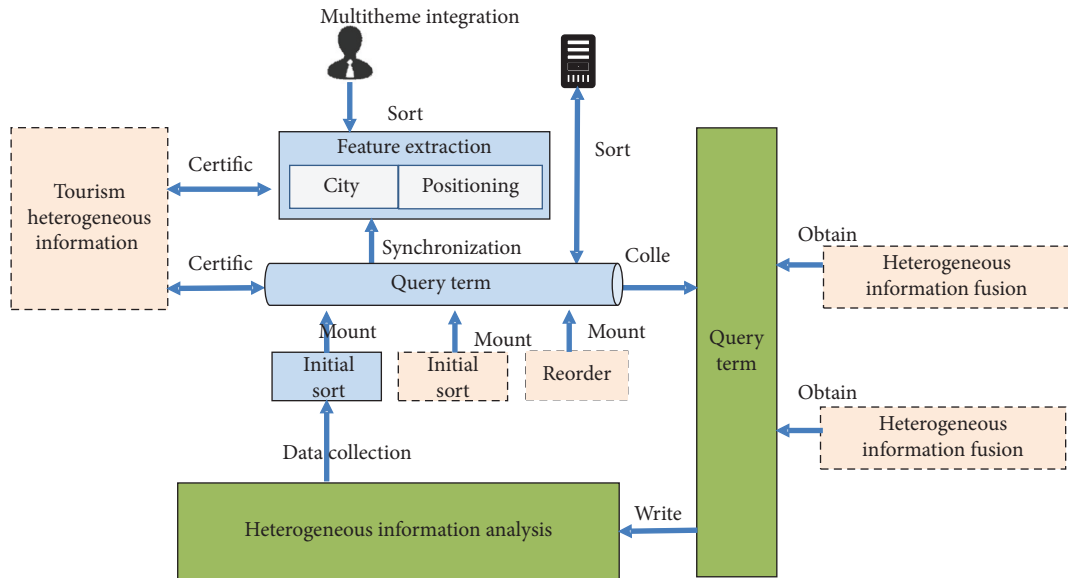
FIGURE 5: Reordering block diagram of tourist attractions based on heterogeneous information fusion.
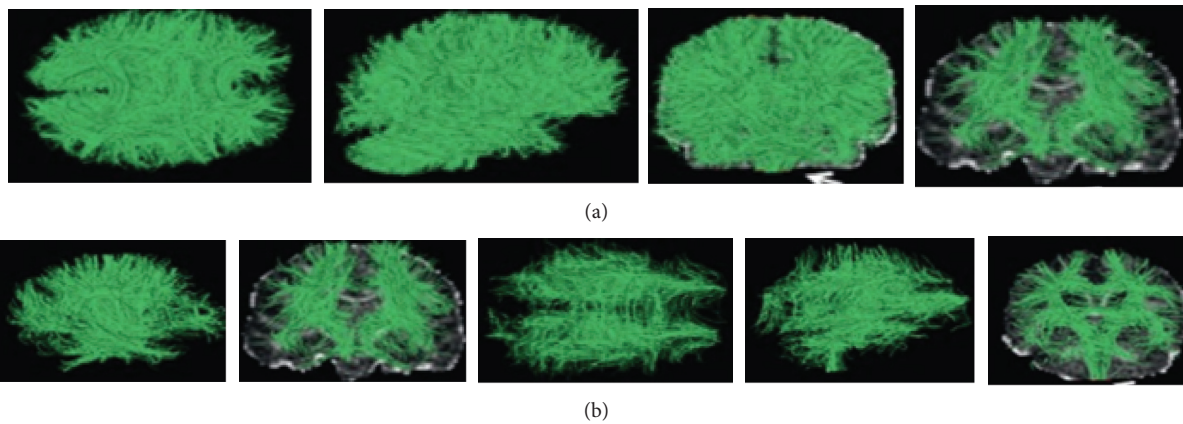

(a)

(b)

FIGURE 6: Time-scale transformation detection: (a) duration of 65 seconds; (b) duration of 155 seconds.
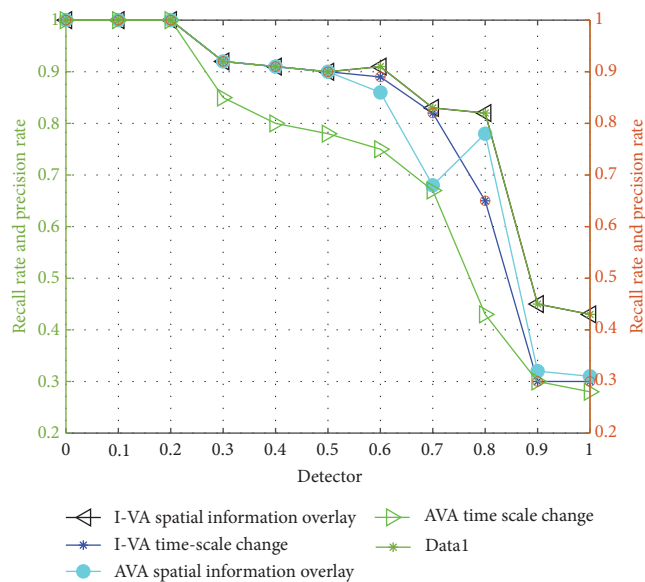

FIGURE 7: Graph of recall and precision of content-based detectors.

Table 3: Unit root test results.

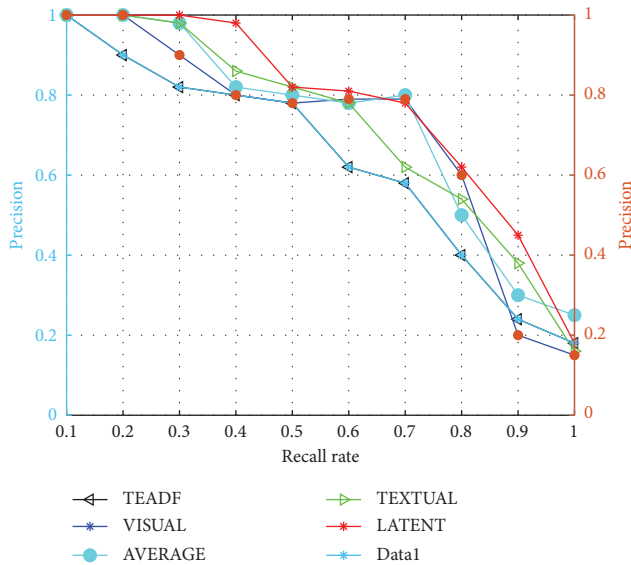| ADF test statistic	−0.686677	1.5% critical value	−3.82832
		5.5% critical value	−3.029683
		10.5% critical value	−2.655418
		One-sided p values	

Figure 8: Precision and recall curves of all methods.

Table 4: The recall rate and precision rate of the detection subset of user operation behavior.

	Audio conversion (%)	Space scale (%)	Visual quality (%)
Precision	99.2	78.4	99.2
Recall	97.3	99.5	95.6

Table 5: Statistics of travel data of real and reliable users in different cities.

City	Number of users	
	Total	Number of tourist attractions over 5
Beijing	4255	326
London	16953	1258
New York	16172	1303
In Paris	14466	1392
Singapore	11578	503
Xi'an	3457	162

and user ratings reorder, which can be seen from the diagram. Sometimes, more implicit characteristics of reorder results are better than the result of the heterogeneous information fusion. Because the recall and precision of this experiment are related to the content, the result of reordering based on multiple hidden features is better than the final reordering based on heterogeneous information fusion.

Table 5 gives a summary of the real and reliable user travel data obtained in different cities. In this experiment, 80 users were selected from each city to test the proposed model. For each user, that is, the user has been to the scenic spots and has not been to the scenic spots. In the scenic spots that the user has been to, the scenic spots are classified into two categories, one as marked scenic spots, and the other as unmarked scenic spots mixed into the scenic spots that have not been to. Similarly, in order to verify accuracy, the comparison methods are PR and U-CF to compare with our PAS model. Here, U-CF helps users find people who have been to the same scenic spots as them, and similar users who have been to other scenic spots are considered as the scenic spots that this user may go to, and are recommended to him. The percentage of the recommended data that contains the sites the user has visited is used as the accuracy of the real data authentication. So, precision N is the proportion of positive samples are evaluated in the first N data. The evaluation criterion is the proportion of the first N recommended results containing samples visited by users, and the value of PrecisionL0 is calculated, where the number of marked positive samples varies from 1 to 4.

5. Conclusion

For the first time, the collective wisdom that exists in travel information has been harnessed to personalize recommendations of attractions, as social media travel information contains a wealth of experiences for travelers. In order to make full use of the collective wisdom in the Internet, structured knowledge can be extracted from the collective wisdom. In order to solve the problem of sparsity and diversity of tourism data, the preferred scenic spots in the current city are obtained by explicit feedback from users. Combining with the attractions of collective wisdom and user feedback, similar classification problem to solve the problem of recommended is established, which can adaptively adjust the weights of multimodal information, get similar sites, and interact with the user. Then, combined with the user positioning information in one of the user contexts, the candidate recommended scenic spots are screened again, and the personalized scenic spots are recommended to the user finally, so as to complete the personalized recommendation of tourist scenic spots based on collective wisdom. The long-time series early warning and prediction method of big data are used to evaluate the sustainable development ability of the research area in the future. The ARIMA model was used to predict the per capita ecological surplus. The results show that the per capita ecological surplus of the research area will decrease year by year in the next 10 years; that is, the local natural ecological resources are more and more difficult to meet the needs of human society and economic development. Aiming at the task of classification and discovery of big data, the ID3 algorithm, the basic method of the travel recommendation decision tree algorithm, is studied, and it has been improved from two aspects to improve the effectiveness of the algorithm, reduce the amount of data processing, and enhance the prediction ability. The practical application proves the effectiveness of the algorithm. The preprocessing methods and methods adopted by the data preprocessing module are analyzed in detail, and the algorithms used in the travel route analysis and consumer spending ability analysis are described in

detail. The data of tourism consumers are analyzed by this system, and the results are evaluated.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] G. N. S. Navneet, "A novel algorithm for big data classification based on lion optimization," *Journal of Theoretical and Applied Information Technology*, vol. 95, no. 7, pp. 1525–1532, 2017.

[2] W. Daopeng, F. Jifei, and F. Hanliang, "Research on optimization of big data construction engineering quality management based on RNN-LSTM," *Complexity*, vol. 2018, Article ID 9691868, 16 pages, 2018.

[3] S. Zhai, J. Duan, and X. Ai, "Research on the third party logistics system and economic performance optimization based on big data analysis," *Boletin Tecnico/Technical Bulletin*, vol. 55, no. 11, pp. 301–308, 2017.

[4] S. Chakraborti and S. Dey, "Analysis of competitor intelligence in the era of big data: an integrated system using text summarization based on global optimization," *Business & Information Systems Engineering*, vol. 61, no. 3, pp. 345–355, 2019.

[5] D. Li, Z. L. Deng, and Z. Cai, "Statistical analysis of tourist flow in tourist spots based on big data platform and DA-HKRVM algorithms," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 87–101, 2020.

[6] G. Stefanos, L. Moritz, G. Tais, S. Vanhuysse, B. Johnson, and E. Wolff, "Normalization in unsupervised segmentation parameter optimization: a solution based on local regression trend analysis," *Remote Sensing*, vol. 10, no. 2, pp. 222–234, 2018.

[7] J.-Y. Park, G. Kim, and H.-J. Oh, "A study on tourism resource strategy of film location using social bigdata based on SNS trend analysis of jeonju area," *The Journal of the Korea Contents Association*, vol. 16, no. 11, pp. 477–487, 2016.

[8] X. Gao and J. Cai, "Optimization analysis of urban function regional planning based on big data and GIS technology," *Boletin Tecnico/technical Bulletin*, vol. 55, no. 11, pp. 344–351, 2017.

[9] Y. Wu, T. Z. Xia, R. YuYi, and J. Wang, "Characteristics and optimization of core local network: big data analysis of

[10] football matches," *Chaos, Solitons & Fractals*, vol. 138, pp. 110136–110145, 2020.

[10] H. Zuo, "Development and robust testing of big data analysis system based on optimized Java technique and parallel oracle database," *Revista de la Facultad de Ingenieria*, vol. 32, no. 2, pp. 180–189, 2017.

[11] Y. Chen, X. J. Luo, and X. Yao, "A medical big data analysis algorithm based on access control system," *International Journal of Reasoning-Based Intelligent Systems*, vol. 10, no. 1, pp. 51–63, 2018.

[12] C. Lu and F. Wen, "GIS information feature estimation algorithm based on big data," *International Journal of Information and Communication Technology*, vol. 15, no. 2, pp. 198–213, 2019.

[13] X. Xia, "Fast search of art culture resources based on big data and cuckoo algorithm," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 127–138, 2020.

[14] H. Shin, "Analysis of subway passenger flow for a smarter city: knowledge extraction from seoul metro's "untraceable" big data," *IEEE Access*, vol. 8, pp. 69296–69310, 2020.

[15] J. Al-Sawwa and S. A. Ludwig, "Parallel particle swarm optimization classification algorithm variant implemented with Apache Spark," *Concurrency Practice & Experience*, vol. 32, no. 2, 2020.

[16] M. Tang and Z. Wu, "Research on the mechanisms of big data on consumer behavior using the models of C2C e-commerce and countermeasures," *African Journal of Business Management*, vol. 9, no. 1, pp. 18–34, 2015.

[17] W. Wang, "Deployment and optimization of wireless network node deployment and optimization in smart cities," *Computer Communications*, vol. 155, no. 4, pp. 117–124, 2020.

[18] R. Xu, L. Z. Wen, B. LiLu, and X. Wang, "Ensemble with estimation: seeking for optimization in class noisy data," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 2, pp. 231–248, 2020.

[19] X. Gao and Y. Wang, "Optimized integration of traditional folk culture based on DSOM-FCM," *Personal and Ubiquitous Computing*, vol. 24, no. 2, pp. 273–286, 2020.

[20] H. Guo, L. J. Liu, and L. Zhao, "Big data acquisition under failures in FiWi enhanced smart grid," *IEEE Transactions on Emerging Topics in Computing*, vol. 7, no. 3, pp. 420–432, 2019.

[21] B. D. Rouhani, A. Mirhoseini, E. M. Songhori et al., "Automated real-time analysis of streaming big and dense data on reconfigurable platforms," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 10, no. 1, pp. 8.1–8.22, 2016.

[22] Y. Uematsu, K. Y. Fan, and W. LinLv, "SOFAR: large-scale Association network learning," *IEEE Transactions on Information Theory*, vol. 65, no. 8, pp. 4924–4939, 2019.

[23] J. Fan, K. H. Song, F. YanLiu, and W. Lian, "Real-time manifold regularized context-aware correlation tracking," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 334–348, 2020.

[24] S. Sharma, J. LuBmann, and J. So, "Controller independent software-in-the-loop approach to evaluate rule-based traffic signal retiming strategy by utilizing floating car data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 9, pp. 3585–3594, 2019.

[25] Y. Li, Y. J. Bao, Z. GongWu, and Y. Zheng, "Mining the most influential -location set from massive trajectories," *IEEE Transactions on Big Data*, vol. 4, no. 4, pp. 556–570, 2018.

[26] X. Xu, F. R. Gu, and S. WanQi, "Multi-objective computation offloading for Internet of Vehicles in cloud-edge computing," *Wireless Networks*, vol. 26, no. 3, pp. 1611–1629, 2020.

[27] T. Kim, W. Li, A. Behm et al., "Similarity query support in big data management systems," *Information Systems*, vol. 88, no. Feb., pp. 101455.1–101455.23, 2020.

[28] S. Knoch, S. Herbig, F. Kosmalla et al., "Sensor-based human-process interaction in discrete manufacturing," *Journal on Data Semantics*, vol. 9, no. 1, pp. 21–37, 2020.

[29] X. Zhao, W. Meng, and J. Su, "Parameter optimization of single sample virtually expanded method," *The International Arab Journal of Information Technology*, vol. 16, no. 6, pp. 988–994, 2019.