Check for updates

Construction of Intelligent Recommendation Retrieval Model of FuJian Intangible Cultural Heritage Digital Archives Resources

Xueqing Liao*

Quanzhou University of Information Engineering, School for Creative Studies, Quanzhou, Fujian, 362000, China
*Corresponding Author: Xueqing Liao. Email: biliao489447@163.com

Abstract: In order to improve the consistency between the recommended retrieval results and user needs, improve the recommendation efficiency, and reduce the average absolute deviation of resource retrieval, a design method of intelligent recommendation retrieval model for Fujian intangible cultural heritage digital archive resources based on knowledge atlas is proposed. The TG-LDA (Tag-granularity LDA) model is proposed on the basis of the standard LDA (Linear Discriminant Analysis) model. The model is used to mine archive resource topics. The Pearson correlation coefficient is used to measure the relevance between topics. Based on the measurement results, the FastText deep learning model is used to achieve archive resource classification. According to the classification results, TF-IDF (term frequency–inverse document frequency) algorithm is used to calculate the weight of resource retrieval keywords to achieve resource retrieval, and a recommendation model of intangible cultural heritage digital archives resources is built through the knowledge map to achieve comprehensive and personalized recommendation of resources. The experimental results show that the recommendation and retrieval results of the proposed method are more in line with users' needs, can provide users with personalized digital archive resources, and the average absolute deviation of resource retrieval is low, the recommendation efficiency is high, and the utilization effect of archive resources is effectively improved.

Keywords: Knowledge map; intangible cultural heritage digital archives; intelligent recommendation; search; TG-LDA model; fasttext model

1 Introduction

With the arrival of the digital media era, the intangible cultural heritage industry has gained new development opportunities. Whether it is non genetic traditional handicrafts or intangible cultural heritage digital resources, they are constantly integrated with digital media, which also proves that digital media provides new opportunities and channels for the development of traditional intangible cultural heritage industry [1]. The intangible cultural heritage involves more than ten categories such as art, drama and music, and its digital archive resources include text, image, audio, video and other

forms [2]. Due to different methods of resource collection and processing, different data structures or types may be generated. In the context of the growing popularity of digital media, in order to effectively protect and correctly use the rich and complex intangible cultural heritage archive resources, it is necessary to correctly understand the importance of resource collection, storage, analysis, retrieval and recommendation, so as to improve the application effect of intangible cultural heritage digital archives resources [3,4].

In the above context, in order to improve the resource utilization effect, relevant scholars have carried out a lot of research. Literature [5] proposed a resource recommendation algorithm based on deep learning. First, extract the user and resource features of the recommendation system, and select the sum of their feature difference values as the target function of the recommendation system. Then a generalized regression neural network resource recommendation model is constructed. Considering the strong dependence of GRNN (generalized regression neural network) training effect on smoothing factor and kernel function center, differential evolution algorithm is introduced to optimize the smoothing factor and kernel function center shift factor of GRNN. The minimum feature difference value solving function is selected as the fitness function of DE (Differential Evolution) algorithm, and the optimal smoothing factor and offset factor are obtained through multiple crossover, mutation and selection operations of DE algorithm. Finally, the optimized smoothing factor and offset factor are used for resource recommendation. The experimental results show that the algorithm has a low RMSE (Root Mean Square Error) value and can obtain better resource recommendation results, but the method has the problem of long recommendation time, and the recommended retrieval results have low compliance with user needs. Literature [6] proposed a design method of digital resources personalized retrieval system based on Web knowledge discovery. The system uses Web knowledge discovery, intelligent agent, data mining and other technologies to design user login module, user interest generation module, optimization of search results and other modules. Through each module, can obtain the impact of user behavior on interest, update the personalized model, and improve the retrieval quality of Web digital resources. The experimental results show that this method can provide personalized resource retrieval services and provide users with some valuable reference materials, but this method has the problem of large average absolute deviation of resource retrieval. Literature [7] proposed a resource retrieval method based on RFID (Radio Frequency Identification) technology, which uses open source RFID middleware to establish a key logical storage structure, and then combines the main classification function to complete resource classification and processing. On this basis, a new resource naming rule is defined, the necessary execution process of resource retrieval is improved according to the known mapping model, and the design and application of the resource retrieval method based on RFID technology are realized. The experimental results show that, with the support of RFID technology, the maximum level of the resource query threshold index of this method is relatively low, which meets the practical application requirements of error free retrieval of resource information, but the recommended retrieval results are less consistent with the user's needs. A point of interest recommendation algorithm based on the combination of location category and social network, CSRS, first obtains the user's location category preference from the user's historical check-in record, and then considers the difference of category preference among friends with tomato as the research object, the weighing method is used to measure the real-time transpiration of crops, and the greenhouse microclimate data is obtained in real time through the deployment of sensors, including air temperature (AT), relative humidity (RH), light intensity (LI) as the model's microclimate environment input, and the canopy relative leaf area index (RLAI) as the model's crop growth input. On this basis, A prediction model of tomato transpiration based on Long short term memory (LSTM) was proposed Aiming at the challenges of sparsity of sign in data and dynamic user interest, a time

aware interest point recommendation algorithm based on LBSN dynamic heterogeneous network is proposed. Session node type is added to LBSN heterogeneous network model Through dynamic meta path, time information, location information and social information are effectively integrated into the semantic relationship between users and interest points The set of dynamic element paths between users and interest points is set, and a method to calculate the preference degree of dynamic path instances is proposed. Matrix decomposition model is used to decompose different dynamic preference matrices According to the user feature matrix and interest point feature matrix of different dynamic meta paths, the recommended list of interest points visited by users at target time is obtained.

Aiming at the above problems, this paper focuses on the research of Fujian intangible cultural heritage digital archives resources, and proposes a design method of resource intelligent recommendation retrieval model based on knowledge atlas. The difference between this method and the traditional method is that it realizes the classification of resources through topic mining and relevance measurement, which improves the comprehensiveness of resource retrieval. Through the construction of digital archive resource knowledge map, the personalized recommendation of resources is realized.

2 Design of an Intelligent Recommendation Retrieval Model for Digital Archive Resources

Under the support of digital media and modern information technology, build the intangible cultural heritage digital archive resource information management center, unify and standardize the relevant archive data resources, give full play to the advantages of the big data platform of joint management, sharing and supervision of data between different institutions and departments, which is conducive to the optimal management and maintenance of archive data, and provide good technical support for the protection, preservation, standardized management and legal use of intangible cultural heritage digital archives.

### 2.1 Classification of Digital Archive Resources

Due to the huge type and quantity of resources in the platform, how to effectively classify these resources is the key to affecting resource recommendation and retrieval. Therefore, in order to improve the efficiency of retrieval and recommendation, the digital archive resources should be classified first. Generally, recommendation is to find the most interesting items on the basis of classification results. Recommendation is implemented on the basis of classification. Therefore, the recommendation results are optimized and improved through resource classification.

2.1.1 Archive Resource Theme Mining Based on TG-LDA Model

The LDA (Latent Dirichlet Allocation) model is a three-layer Bayesian probability model. The LDA model overcomes the shortcomings of the traditional Vector Space Model (VSM) based modeling, such as too high and extremely sparse text dimensions, and ignoring text semantic information. In this model, archive resources are viewed as texts, different themes are formed through different texts, and texts are viewed as a mixture of potential themes, which are probability distributions in word space. By viewing potential themes as soft clusters of word features, the content of digital resources is summarized from a more abstract level.

However, when the standard LDA model is applied to the classification of cross type digital archive resources, it is unable to effectively mine and distinguish the sharing and unique topics between different fields. Therefore, this paper proposes the TG-LDA model based on the standard LDA model. This model is a model for mining training data and test data sharing and unique potential topics by expanding the parameters of the LDA model. It realizes the classification of cross type archive

resources by mining and sharing potential topics as a bridge for text classification knowledge transfer. TG-LDA model is shown in Fig. 1.
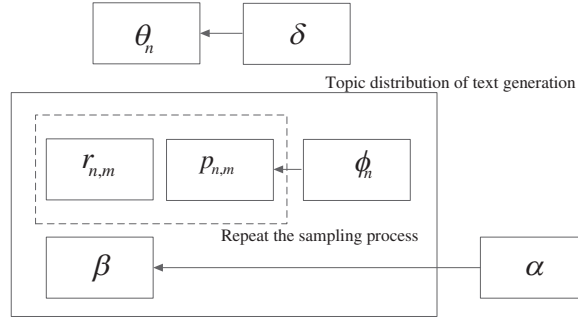
Figure 1: Schematic diagram of TG-LDA model

In Fig. 1, the inner rectangle represents the repeated sampling process, that is, word packets, and the outer rectangle represents the topic distribution ϕ generated from n text. According to the topic distribution, m feature words are generated from the text one by one. Parameter α obeys Bernoulli distribution and is used to control the value of decision random variable β. By introducing new decision random variables into LDA model and jointly modeling training and test data, two types of potential topics can be effectively introduced into the data generation process. The generation process is transformed into a probability model, and the following joint probability distribution is obtained, as shown in Formula (1):

$$f(\phi, r, p | \alpha, \beta) = \prod_{i=1}^{n} f(r_i | \phi)^{\alpha} f(p_i | \phi)^{\beta} \tag{1}$$

In the TG-LDA model, the co-occurrence relationship between text s and feature word c is associated with potential topic T, and T is sampled from the shared topic set and the unique topic set of training data and test data. The sampling process is controlled by the introduced decision random variable $\beta = \{0, 1\}$, and the decision variable β controls the type of potential topic from which the current word is generated.

The derivation of super parameters of TG-LDA model is mainly realized by Gibbs Sampling method. The sample closest to the probability distribution value is extracted by constructing a Markov chain [8] that converges to the target probability distribution. The posterior estimates of the text topic distribution ϕ_s and topic feature distribution δ_c in the final model are obtained through derivation:

$$\phi_s = \frac{m_{s,i} + \alpha}{m_s + Z\alpha} \tag{2}$$

$$\delta_c = \frac{m_{c,i} + \beta}{m_c + Z\beta} \tag{3}$$

where, $m_{s,i}$ represents the number of words allocated to topic i in text s; m_s means all words of the text s assigned to the subject i; $m_{c,i}$ represents the frequency of the word c assigned to topic i; m_c represents all the words assigned to topic i.

2.1.2 Measurement of Inter Topic Relevance

After mining the sharing and unique potential themes of training data and test data through the TG-LDA model, it is necessary to calculate the correlation between the two types of unique potential

themes and the shared theme, so that the shared theme can be used as a bridge for knowledge migration to achieve the classification of test data. This is because there is no co occurrence relationship between the unique themes of training data and test data, and the similarity between them cannot be directly measured by similarity. However, if the unique topics of two data are associated with many of the same shared topics, there may be semantic correlation between them. Therefore, the relationship between the two types of unique topics can be established through the correlation between unique topics and shared topics, and then the test data can be classified based on the unique topics of shared topics and related training data.

For the correlation between unique potential topics and shared topics, this paper uses the Jensen Shannon distance [9] (Jensen Shannon Divergence, JSD) between the distribution of shared topics on the text and the distribution of unique topics on the text as a similarity measure between shared topics and unique topics. JSD is widely used to calculate the similarity between two probability distributions. Specifically, for two discrete probability distributions P_1 and P_2, Jensen Shannon between them is defined as:

$$JSD\left(P_1, P_2\right) = \left(KLP_1 + KLP_2\right)/M \tag{4}$$

where, KL represents the Kullback Leibler distance between two probability distributions. By using Jensen Shannon distance as the similarity measurement, the similarity $Sim\left(I_k, I_u\right)$ between shared theme I_k and unique theme I_u is defined as:

$$Sim\left(I_k, I_u\right) = \frac{1}{JSD\left(P_1, P_2\right) + KL\left(I_k, I_u\right)} \tag{5}$$

According to Formula (5), the larger the value of $Sim\left(I_k, I_u\right)$, the smaller the JSD of shared theme I_k and unique theme I_u, and the greater the correlation between them.

After obtaining the correlation between the unique theme and all shared themes, it is necessary to infer the relationship between the unique theme of training data and the unique theme of test data according to the calculated results. This paper mainly uses Pearson Correlation Coefficient (PCC) to calculate the correlation between them. Pearson correlation coefficient is often used to measure the linear correlation between two random variables. By using Pearson correlation coefficient, the correlation $Sim\left(I_i^k, I_i^u\right)$ between the unique theme I_i^k of training data and the unique theme I_i^u of test data can be obtained. The calculation formula is defined as:

$$Sim\left(I_i^k, I_i^u\right) = \frac{\overline{S_k} - \overline{S_u}}{\sqrt{\left(I_i^k + I_i^u\right)\left(I_i^k - I_i^u\right)}} \tag{6}$$

where, $\overline{S_k}$ and $\overline{S_u}$ represent the average similarity between I_i^k and I_i^u of training data and test data and all shared topics respectively. According to the nature of Pearson correlation coefficient, if the $Sim\left(I_i^k, I_i^u\right)$ value is positive, it means that the unique themes of training data and test data have high or low similarity with many shared themes at the same time, while the $Sim\left(I_i^k, I_i^u\right)$ value is negative, which means that the unique themes of one data have high similarity with some shared themes, while the unique themes of another data have low similarity with these shared themes. Through the above calculation, the unique theme set of training data with the highest correlation coefficient with the unique theme of test data can be obtained, thus realizing the measurement of the correlation between the themes of digital archive resources.

2.1.3 Realization of Digital Archive Resource Classification

Based on the above calculation results, this paper proposes a fast classification method of digital archive resources based on the FastText deep learning model [10]. This model is a text classification depth learning model based on the CBOW (Continuous Bag-of-Word Model)framework of word2vec. As shown in Fig. 2, FastText model is divided into three layers of training graph structure: input layer, hidden layer and output layer. The input layer is the initialization vector, and n-gram features are added to the vector to enhance the integrity of resources. The average value of each resource vector is obtained through the hidden layer, and the weight parameters are updated according to the optimizer and gradient descent algorithm. Finally, the loss function and the corresponding classification category are calculated.
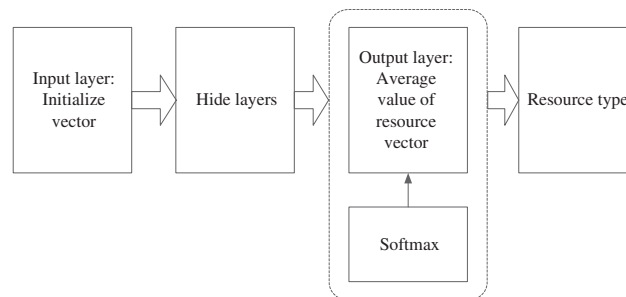
Figure 2: Schematic diagram of FastText model

FastText model uses a hierarchical classifier (rather than a flat architecture), and different categories are integrated into the tree structure. In text classification tasks with many categories, the calculation of linear grade is very complex. In order to improve the running time, the FastText model uses Softmax layering technology. This technology is based on Huffman coding, and is mainly used to encode text data labels, which can effectively shorten the training time. The training process of FastText is shown below.

Input: text s and characteristic word c; Select loss function g_{loss}; Set learning rate h

Output: output digital archive resource category

1. Preprocess the text s, and add a label at the end of each line;
2. Set loss function g_{loss} and learning rate h, and select category output;
3. Train according to gradient descent algorithm, and update the weight of text s and feature word c;
4. After the training, the classification model is obtained;
5. Realize the classification of digital archive resources.

2.2 Intelligent Retrieval of Digital Archive Resources

Based on the classification results of digital archive resources, further calculate the keyword weight of resource retrieval, and use TF-IDF (term frequency–inverse document frequency) algorithm to complete the keyword weight calculation of digital archive resources retrieval. This algorithm can evaluate the importance of keywords. The more keywords appear, the more important they are; The fewer keywords appear, the less important they are. If want to indicate that a keyword has strong category discrimination, it appears more often in a digital archive resource type, but less often in other digital archive resource types.

TF is used to measure the number of times a keyword is presented in a digital archive resource type. Set the specified digital archive resource type to be described by I_j; The importance of any keyword k_j in it is described by formula (7):

$$\eta\left(k_j\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} I_j\left(t_{j,k} \times N_{j,k}\right) \tag{7}$$

where, $t_{j,k}$ represents the frequency of keyword presentation; $N_{j,k}$ represents the total number of all keyword presentation frequencies.

IDF is used to measure the universal importance of a keyword. The weight of the keyword can be obtained by the total number of digital archive resource types. The specific calculation formula is:

$$W\left(k_j\right) = \frac{\log_{10}\left(N_d - N_k\right) \times \eta\left(k_j\right)}{N} \tag{8}$$

where, N_d represents the total number of digital archive resource types; N_k represents the number of digital archive resources with keywords. If the keyword does not exist in the digital archive resource, the denominator is 0.

The formula for calculating the keyword weight of digital archive resource retrieval is as follows:

$$W_{jk} = \sum_{j,k=1}^{n} W\left(k_j\right)\left(1 - \theta\right) + \sigma w_t \tag{9}$$

where, σ represents the scale parameter; w_t represents the total number of keywords presented in a digital archive resource type; θ represents the overall importance of keywords.

The flowchart of intelligent retrieval of digital archive resources is shown in Fig. 3.
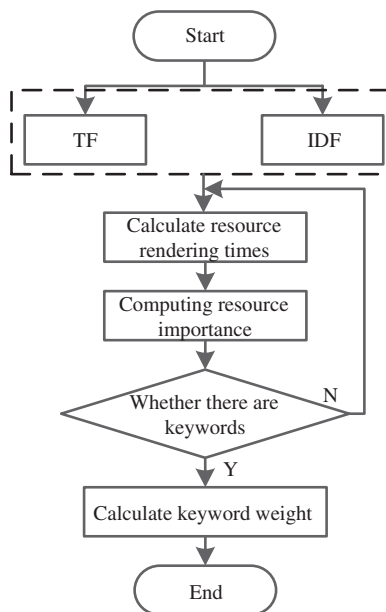
Figure 3: Flow chart of intelligent retrieval of digital archive resources

2.3 Personalized Recommendation of Digital Archive Resources
2.3.1 Knowledge Map Construction

Knowledge map [11] is a technology that uses computers to store, manage and present concepts and relationships between concepts. It can be divided into semantic knowledge maps based on RDF (Resource Description Framework) storage (that is, associated data) and generalized knowledge maps based on graph databases. The semantic knowledge map (associated data) focuses on the release and link of knowledge, while the generalized knowledge map focuses on the mining and calculation of knowledge. The associated data is the continuation and development of Google's knowledge map. The generalized knowledge map studies the combination of graph operations and associated data. In the field of map, information and digital humanities, semantic knowledge atlas (associated data) is proposed. In recent years, knowledge atlas has been favored by researchers in various fields for its advantages such as extensive data collection, comprehensive knowledge coverage, quantitative analysis of macro, and visualization of data graphs. The research on the industrialization of non heritage protection has a long history, a large number and many achievements. Using the method of knowledge atlas to sort out the research results of industrialization of non heritage protection has strong practical value and theoretical significance for objectively and comprehensively displaying the historical context of industrialization of non heritage protection research, focusing on research hotspots, exploring future development trends and optimizing the protection path.

The knowledge map constructed in this study is constructed in a top-down manner. There are three data sources of the knowledge map, namely, local resources, intangible cultural heritage malls and resources in digital museums. The three resources are structured data, which is very helpful to form the knowledge map of intangible cultural heritage digital archive resources. The flowchart of the knowledge diagram construction is shown in Fig. 4.
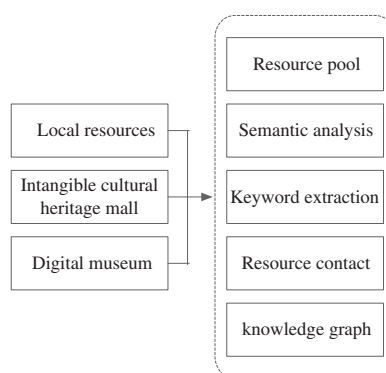
**Figure 4:** Flow chart of knowledge map construction

Taking local upload as an example, the local uploaded intangible cultural heritage digital archive resources require filling in the title, description and content. Resource content can be in the form of video or text to form a complete resource description library. After semantic analysis of the resource description database through semantic analysis tools, keyword extraction is carried out. Keyword extraction is calculated according to the frequency of keywords appearing in digital archive resources and the different weights of their positions. Keyword weights are sorted from large to small as topics, descriptions and contents. The co word analysis method is used to find the relationship between resources, and the newly added resource needs are compared with all resources to find related resources to form a new knowledge map. The content in the resource package is a structured xml file, so it can

be uploaded as a whole to form a knowledge map. However, due to the large content of the resource package and the large number of keyword matches, it takes a long time. In order to solve this problem, resources are uploaded in batches, that is, resources are uploaded according to the classification results of digital archives. In this way, multiple resource networks or resource databases will be formed, and the resource database associated with resources is a knowledge map, which will be finally displayed to users through web page programming language.

2.3.2 Recommendation Model of Intangible Cultural Heritage Digital Archive Resources Based on Knowledge Map

The previous section introduced the construction of the knowledge map of digital archive resources, which mainly introduced the design process of the recommendation model based on the knowledge map. The recommended method is designed completely based on the relationship between digital resources. The knowledge map gathers intangible cultural heritage digital archives resources into a knowledge network. However, the process of drawing the knowledge map in this study is to match keywords through semantic analysis [12] to form the relationship between resources. Due to the limitations of current semantic analysis research, there may be no common keywords between some knowledge points, In addition, considering the particularity of the knowledge map, the recommended model of intangible cultural heritage digital archive resources based on the knowledge map is developed as shown in Fig. 5 .
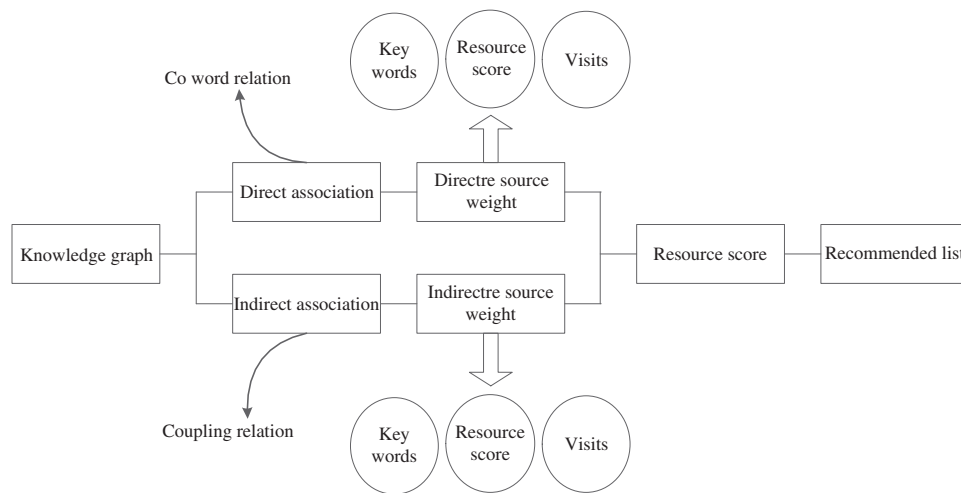
Figure 5: Schematic diagram of recommendation model based on knowledge map

It can be seen from Fig. 5 that the resources in the knowledge map are divided into two types: direct association and indirect association. The resources that are directly associated are two resources with common keywords. There are direct connections on the knowledge map, which are called co word relationships; Indirect association means that two resources are directly related to the third resource, so there is no direct connection between the two resources in the knowledge map, which is called coupling relationship. Each resource may have multiple resources that are directly and indirectly related to it. The recommendation process is mainly the process of arranging and displaying these resources. Similar to search engines, users try to recommend the most relevant web pages to the top when searching for problems, and rank the resources. The scoring principle is to take a weighted approach to the impact factors.

In this study, the keyword weight of intangible cultural heritage digital archive resource recommendation process is slightly different from the resources of co word relationship and coupling relationship:

(1) The influencing factors of co word relationship include the number of key words, the score of digital resources, and the number of visits. The number of key words is the number of common keywords contained in the two resources of the co word relationship. The more common keywords are, the stronger the correlation is; The resource score is the score of the directly connected resource itself, and the score of the resource is calculated by the same method. The higher the score, the higher the value of the resource, and it is recommended in the front position; The number of visits refers to the number of times the resource is accessed by users. The number of visits represents the degree of interest in the resource.

(2) The resources in the coupling relationship are not connected directly, so there is no key word number in the influence factor [13–15]. The score of coupling resources is added. Coupling resources are the link between the two resources, so the score of coupling resources has a strong impact on the two resources. In this way, the sum of the product of the score of the direct resource impact factor and its weight, and the sum of the product of the score of the indirect resource impact factor and its weight, will be multiplied by the sum of the weights of the direct resources and the coupling resources to obtain the score of each type of resources. Rank the resources according to the final score, and push the resources to the corresponding page according to the high and low order to achieve personalized recommendation of intangible cultural heritage digital archive resources [16,17].

3 Experimental Analysis

In order to verify the effectiveness of the design method of intelligent recommendation retrieval model for Fujian intangible cultural heritage digital archives resources based on knowledge atlas, the data set uses the digital archives resources data of a library in Fujian Province, and this paper conducts experimental research with the conformance of retrieval results with user needs, recommended list generation time, and average absolute deviation as test indicators.

3.1 Experimental Environment Design

First, the test environment for intelligent recommendation retrieval of digital archive resources is constructed, including the software environment, the test end and the server. The specific experimental parameters are shown in Table 1.

Table 1: Experimental parameter setting

Parameter	Specifications
Test framework	GOJAND1.0
Testing environment	CharmPy
Memory	32G
Hard disk	RDD3
Middleware	Mqrabbit
Adapter	IGENC
Operating system	Windows 10

The constructed test environment is used to carry out the intelligent recommendation retrieval experiment of digital archive resources. In order to ensure the contrast of the experimental results, the traditional literature [5] method is compared with the literature [6] method to compare the performance of the three intelligent recommendation retrieval methods of digital archive resources.

3.2 Result Analysis

The judgment of intelligent recommendation retrieval performance is mainly based on the conformity between the intelligent recommendation retrieval results of digital archive resources and user needs. The higher the compliance, the better the performance of intelligent recommendation retrieval. The methods of the literature [5], the literature [6] and the proposed methods are used to test respectively. Three different methods are used to recommend digital archive resources for users, and the conformity of the recommended retrieval results of the three methods with the user's needs is compared. The test results are shown in Fig. 6.
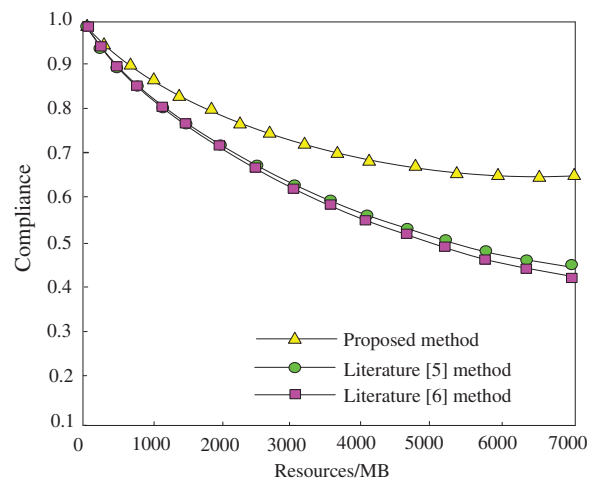

Figure 6: Comparison of the compliance between recommended search results and user needs

It can be seen from Fig. 6 that with the increase of digital archive resources, the conformity of recommended retrieval results of the three methods with user needs shows a gradual downward trend. Among them, the conformity of literature [5] method and literature [6] method is relatively close, while the conformity of the proposed method is higher, and its minimum value remains above 0.65. Through comparison, it can be seen that the recommendation and retrieval results of the proposed method are more in line with user needs and can provide users with personalized digital archive resources. This is because this method proposes the TG-LDA model based on the standard LDA model. This model is used to mine archive resource topics. Pearson correlation coefficient is used to measure the correlation between topics.

As digital archive resources are increasing at an alarming rate, how to quickly search and recommend massive resources is a key indicator to measure the application performance of different methods. Comparing the recommended list generation time of the literature [5] method, the literature [6] method and the proposed method, the results are shown in Fig. 7.

It can be seen from the analysis of Fig. 7 that when the proposed method is used to recommend digital archive resources for users, the time taken to generate the recommended list is within 20 s; When using the literature [5] method to recommend digital archive resources for users, the longest time to

generate a recommended list is up to 42.5 s; When using the literature [6] method to recommend digital archive resources for users, the longest time to generate a recommended list is 34.2 s. By comparing the test results of three different methods, it can be seen that the proposed method can generate a recommendation list for users in a short time, because the method classifies resources through the TG-LDA model before recommendation, which shortens the time used to generate the list and verifies the high recommendation efficiency of the method.
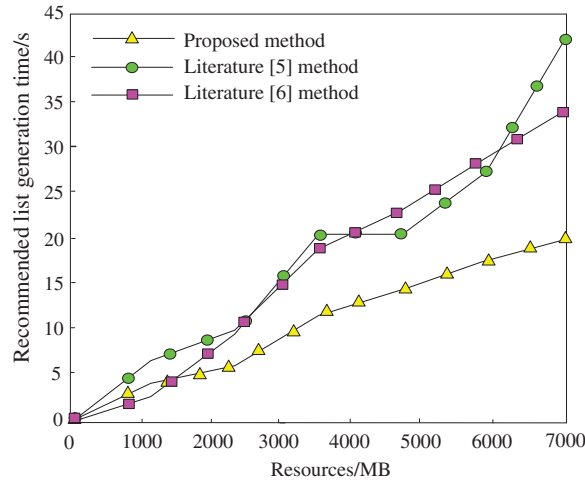
Figure 7: Comparison results of recommendation list generation time

The retrieval accuracy is a key indicator to evaluate the quality of resource retrieval. The average absolute deviation existing in resource retrieval is taken as the measurement indicator. The average absolute deviation can more intuitively display the quality of resource retrieval, so as to measure the retrieval quality of different methods. The larger the average absolute deviation is, the lower the quality of the retrieval results. The calculation formula is as follows:

$$MAE = \frac{\sum_{i=1}^{N} |x_i - y_i|}{N} \tag{10}$$

where, x_i represents the actual scores of N users; y_i represents the predicted score of N users.

By comparing the retrieval quality of literature [5] method, literature [6] method and the proposed method with the average absolute deviation, the results are shown in Fig. 8.

It can be seen from the analysis of Fig. 8 that the average absolute deviation of the proposed method is far less than the average absolute deviation of the methods in literature [5] and literature [6], and the average absolute deviation always does not exceed 0.2. The smaller the average absolute deviation is, the higher the retrieval quality of the representative method is. Therefore, the retrieval quality of the digital archive resources of the proposed method is higher. This method measures the number of times a keyword is presented in a digital archive resource type through TF, and measures the universal importance of a keyword through IDF to obtain the weight of the digital archive resource keywords. In the retrieval process, the resource list is constantly updated and improved, reducing the average absolute deviation.This is because this method uses FastText deep learning model to achieve archive resource classification. According to the classification results, TF-IDF algorithm is used to calculate the weight of resource retrieval keywords to achieve resource retrieval. A recommendation

model of intangible cultural heritage digital archives is established through knowledge map to achieve comprehensive and personalized recommendation of resources.
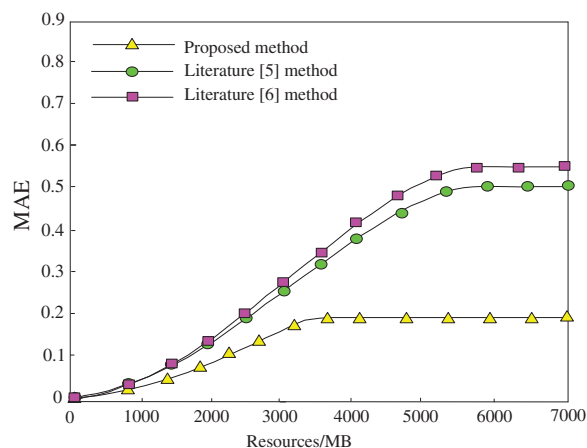
Figure 8: Comparison results of average absolute deviation of different methods

4 Conclusion

In order to improve the consistency between recommended retrieval results and user needs, improve the efficiency of recommendation, and reduce the average absolute deviation of resource retrieval, this paper proposes a design method of intelligent recommendation retrieval model for Fujian intangible cultural heritage digital archives resources based on knowledge maps. The main research contents of this method are as follows:

(1) The TG-LDA model is used to mine archive resource topics, Pearson correlation coefficient is used to measure the relevance between topics, and FastText deep learning model is used to achieve archive resource classification, reducing the complexity of resource retrieval recommendation.

(2) TF-IDF algorithm is used to calculate the weight of resource retrieval keywords to achieve resource retrieval.

(3) The recommendation model of intangible cultural heritage digital archives resources is constructed through the knowledge map to achieve comprehensive and personalized recommendation of resources and improve the compliance of recommended retrieval results with user needs.

(4) The experimental results show that the recommendation and retrieval results of the proposed method are more in line with users' needs, can provide users with personalized digital archive resources, and the average absolute deviation of resource retrieval is low. The time used to generate the recommendation list is within 20 s, and the recommendation efficiency is high, which effectively improves the utilization effect of archive resources.

Availability of Data and Materials: 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 to report regarding the present study.

References

[1] S. H. Chen, "Design of VR-based digital management system for intangible cultural heritage resources," *Modern Electronics Technique*, vol. 46, no. 16, pp. 89–91, 2020.

[2] B. P. Yan, Q. Y. Huang and C. Yuan, "The digital design of intangible cultural heritage under the playfulness visual domain," *Packaging Engineering*, vol. 42, no. 22, pp. 40–46, 2021.

[3] L. H. Yao and G. Z. Yu, "Relational database information resource retrieval result classification method simulation," *Computer Simulation*, vol. 36, no. 1, pp. 445–448, 2019.

[4] L. Wenige and J. Ruhland, "Similarity-based knowledge graph queries for recommendation retrieval," *Semantic Web*, vol. 10, no. 6, pp. 1007–1037, 2019.

[5] F. F. Song, D. Sui and X. Z. Zhou, "Intelligence learning resource recommendation algorithm based on deep learning," *Journal of Nanjing University of Science and Technology*, vol. 46, no. 2, pp. 185–191, 2022.

[6] X. G. Huang, "Personalized retrieval system of digital book resources based on web knowledge discovery," *Computer Systems & Applications*, vol. 30, no. 8, pp. 111–117, 2021.

[7] X. J. Zhai, "E-book resource retrieval method based on RFID technology," *Modern Electronics Technique*, vol. 45, no. 11, pp. 99–103, 2022.

[8] A. Gellert, "Web usage mining by neural hybrid prediction with markov chain components," *Journal of Web Engineering*, vol. 20, no. 5, pp. 1341–1358, 2021.

[9] V. Asha, "Automatic fabric inspection using GLCM-based Jensen-Shannon divergence," *Informatica: An International Journal of Computing and Informatics*, vol. 46, no. 1, pp. 19–25, 2022.

[10] S. Kuutti, R. Bowden, Y. Jin, P. Barber and S. Fallah, "A survey of deep learning applications to autonomous vehicle control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 712–733, 2021.

[11] S. Singh and M. Siwach, "Handling heterogeneous data in knowledge graphs: A survey," *Journal of Web Engineering*, vol. 21, no. 4, pp. 1145–1186, 2022.

[12] C. Schwarz, "Isemantica: A command for text similarity based on latent semantic analysis," *The Stata Journal*, vol. 19, no. 1, pp. 129–142, 2019.

[13] Y. W. Wu, Q. T. Cai, Z. Liu and Y. Z. Deng, "Digital resource recommendation based on multi-source data and scene similarity calculation," *Data Analysis and Knowledge Discovery*, vol. 5, no. 11, pp. 114–123, 2021.

[14] Y. M. Chen, "A recommendation algorithm for digital resources based on python," *Library Research*, vol. 49, no. 1, pp. 105–109, 2019.

[15] X. Liang and J. Yin, "Recommendation algorithm for equilibrium of teaching resources in physical education network based on trust relationship," *Journal of Internet Technology*, vol. 23, no. 1, pp. 133–141, 2022.

[16] M. A. Albaom, F. Sidi, M. A. Jabar, R. Abdullah, I. Ishak *et al.,* "The moderating role of personal innovativeness in tourists' intention to use web 3.0 based on updated information systems success model," *Sustainability*, vol. 14, no. 21, pp. 13935, 2022.

[17] M. A. Albaom, F. Sidi, M. A. Jabar, R. U. S. L. I. Abdullah, I. S. K. A. N. D. A. R. Ishak *et al.,* "The impact of tourist's intention to use web 3.0: A conceptual integrated model based on TAM & DMISM," *Journal of Theoretical and Applied Information Technology*, vol. 99, no. 24, pp. 6222–6238, 2021.