Intelligent Retrieval Method of Massive Music Information Resources Based on Deep Learning

Yi Liao[1(✉)] and Lin Han[2]

[1] Music Department of Xi'an , Shiyou University, 710065, China
228212611@qq.com
[2] Nanchang JiaoTong Institute, Nanchang 330100, China

Abstract. With the increase of data storage capacity and the development of transmission technology, the number of music shows an unprecedented growth. However, this explosive growth makes it more and more difficult to find interesting music clips in such a huge library of music information resources, so it is of great significance to study the intelligent retrieval methods of massive music information resources. Use the word segmentation tool to segment and label the music information resources. According to the word segmentation and part of speech tagging, build a feature extractor based on deep learning, obtain the features of massive music information resources and label the resources using approximate matching method. Combined with the result of resource feature annotation, the size of resource buffer is determined to build an intelligent hierarchical index. For different hierarchical indexes, sparse reconstruction technology is used to preprocess information resources, and pseudo-correlation feedback is used to expand the cache query. Based on the cache query expansion, match the query content entered by the user with the extension words with the similar meaning of the keyword to obtain the matching cost. According to this cost, the result of matching cost less than the set threshold is fed back to the user through feedforward neural network to obtain the retrieval results of massive music information resources. The experimental results show that the maximum expansion coefficient of this method can reach 0.99, the retrieval speed is fast, and the maximum error between the retrieval recall and the actual data is 0.30 kB.

Keywords: Deep Learning · Music Information Resources · Intelligent Retrieval

1 Introduction

With the rapid development of the Internet and the information society, the characteristics of intelligence, globalization and non-group information widely exist in life, and the information age enables users to transmit resources in a more convenient way. Text search has become more and more mature, but the sensory information expressed in other forms is difficult to describe clearly. If the characteristics of a certain type of information can be found for matching and retrieval, this can be regarded as the best way to improve efficiency and save resources. Sound information is an important medium in life, accounting

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 532, pp. 408–423, 2024.
https://doi.org/10.1007/978-3-031-50571-3_30

for about 25% of the total amount of information. Music is one of the most common sound media. With the development of the Internet era, people require music information retrieval to be faster, more accurate, more convenient, and more efficient. However, music has its own many characteristics, such as tone, melody, timbre, and sound intensity. It is difficult to specify the characteristics of music itself, unlike traditional literary instinct, which is described in an intuitive way. As audio text retrieval is common in life, it can be retrieved by inputting song names, lyrics, etc. As far as music is concerned, the first message is its own melody, and the second message is descriptive text. Everyone has had this experience. When a song appears in their mind, they can only remember the melody, but can't remember the song name. People often want to use the song melody they remember to get the song they want, as is the case with music information retrieval. It has become an inevitable trend that users should adopt better information technology means to process music signals, and now it has attracted people's attention. How to quickly and accurately extract the information you need from a large amount of information has a very important application prospect. Traditional retrieval technology often matches similar resources in certain modalities, and the obtained resources are mostly homomorphic resources with high similarity. Therefore, it is necessary to retrieve massive music information resources. Some literatures propose a retrieval method driven by approximate resource matching, which constructs an artificial convolutional neural network model. Using approximate resource matching results to extract music information features, measure multi-tag similarity according to the concept of conditional entropy, thereby realizing massive music information resource retrieval; some literatures propose a retrieval method based on the combination of on-chain and off-chain. Blockchain and distributed storage technology are combined to achieve the purpose of decentralized data storage. It also provides data retrieval interfaces to managers to achieve the integrity retrieval of massive music information resources.

Due to the correlation between different music information resource data, although the above two methods can effectively retrieve data, it is difficult to determine the characteristics of different massive music information resources and the size of the resource buffer, resulting in the inability to build relevant indexes and the decline of the intelligent retrieval effect of massive music information resources. In order to solve this problem effectively, this paper introduces deep learning and proposes a new intelligent retrieval of massive music information resources. The proposal and application of the concept of deep learning is the biggest breakthrough in the field of speech recognition in the 21st century. In 2011, Microsoft first used deep neural network, which has achieved remarkable results in the task of speech recognition. Deep neural network has attracted more and more attention in speech recognition. Considering that deep learning has the advantages of strong learning ability, wide coverage, strong adaptability and good portability, it can effectively determine the characteristics of different massive music information resources and the size of the resource buffer, so as to realize the intelligent retrieval of massive music information resources, which has high practical application value.

2 Feature Extraction of Massive Music Information Resources Based on Deep Learning

For feature extraction of massive music information resources, text features should be extracted first, and redundant function words should be eliminated at the same time. Because the information objects in the massive music information resource library are described by the same multimodal information resource specification, in different cases, the internal nodes of the multimodal information resource standard tree of each information object are part of the multimodal information resource specification, and the difference is the element value on the leaf node. Comparing the query tree with the standard tree, it is found that matching the query tree with the standard tree data is unnecessary to match the metadata of the object with the query tree data [1, 2]. When the query tree matches the metadata tree, since there is no node on the subtree that can match the node in the query tree, there is no need to consider the matching situation of the subtree with the node as the root node. Let A and A' be two unordered label trees, and the edit distance between them is:

$$d\left(A, A'\right) = \min\left\{f \mid A \xrightarrow{f} A'\right\} \tag{1}$$

In formula (1), f is the edit sequence mapping. Therefore, before querying, first match the query tree with the multimodal data standard scheme tree of the resource target database, and then record the matching information (i.e. preprocessing information) of the associated node [3]. By analyzing the obtained information, we can avoid a large number of non associated nodes matching in the future matching between the query tree and the multimodal data standard tree, and avoid unnecessary duplication [4].

The tagging method is used to extract the word segmentation features of music information resources marked with word segmentation tools, and the weighting method of (word frequency - inverse word frequency) is used to calculate the feature weight of music information resources. The formula is:

$$w = \eta \times \log_2 \frac{\vartheta}{n} \tag{2}$$

In formula (2), η represents the number of times the information to be retrieved appears in the retrieval document. Generally speaking, the most important information in music information resources is the information with the highest frequency in all resources, and the resource word frequency is usually used to measure the characteristics of similar texts. The lower the frequency of resource occurrence, the better the identification of resource classification [5].

In a standard restricted Boltzmann machine all observed variables are related to different parameters in the hidden layer. Understand the principle of the model from the perspective of the image, which will make the explanation and understanding more convenient [6, 7]. As the dimension of the image increases, the number of connection weights in deep learning will also become quite large, which will make the operation more complicated and slower in the process of training and updating. In fact, only a few parameters are needed to describe spatial local features, and these parameters also play an important role in extracting information resource features [8]. Therefore, in order to solve these problems, a deep learning-based feature extractor is constructed, as shown in Fig. 1.

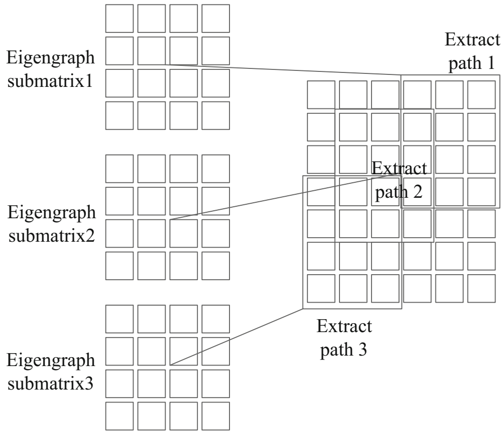

Fig. 1. Feature extractor based on deep learning

As can be seen from Fig. 1, Each feature graph sub-matrix of the feature extractor based on deep learning corresponds to a feature extraction path. The deep learning feature extractor is responsible for connecting the visible layer unit and the hidden layer unit, that is, K 3 × 3 feature extractors. The hidden layer units are divided into K, which are called sub-matrices of the feature map. The specific features extracted in each visible layer 3 × 3 neighborhood unit are represented by each hidden layer unit. Identical features in different places in visible cells are represented by cells in a feature map [9]. The preprocessing part, sentence class hypothesis and detection, and semantic block composition are used as processing links, input natural language sentences, and then output the corresponding sentence class representation and word description. The specific extraction steps are as follows: first, do preprocessing, then make assumptions about possible information resources based on the concepts neutralized by all information in the resources, and judge the type of information resources based on the conceptual knowledge contained; And according to the content of the third part, use semantic blocks to judge resources. It is assumed that failure to pass the above steps during the extraction process will lead to traceability. At this time, "hypothesis" and "detection" need to be carried out again before the successful test is completed. Query extraction technology is the key technology of semantic retrieval. By adding words or concepts related to the query semantics of the original query language, the query time is longer than the original query, thus improving the efficiency, recall and accuracy of document retrieval [10]. The semantic information related to the formation of professional knowledge is extracted through the comparison of relevant content and user needs; For the phrases not found in the resource library, use the universal semantic dictionary to expand their cross language semantics, and present them to users in the form of tables for self recognition. The query string is expanded into a search engine query, and the query results are clustered and presented to the user.

3 Resource Annotation Based on Approximate Matching

In the process of retrieving massive music information resources, the principle of tree matching is introduced, and a multi-level and multi-modal approximate matching model is established. In this model, the principle of tree matching is applied, which is to realize the modal matching of different data through constraint mapping by mapping the nodes between two trees. Combined with the structural characteristics of massive music information resources, the concepts of structure-based search and semantic search are proposed, and the affinity constraint principle is introduced into this concept, and a retrieval framework based on approximate matching model as shown in Fig. 2 is constructed.

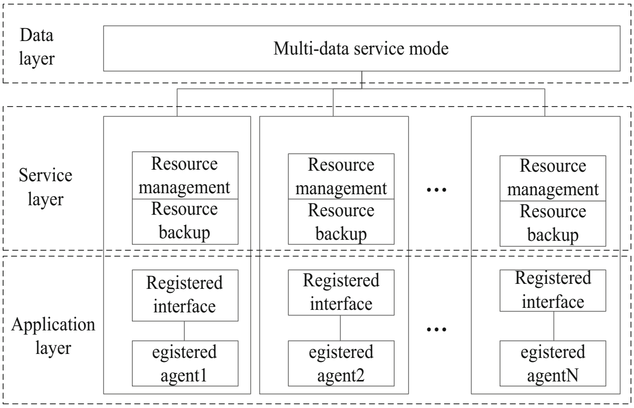

Fig. 2. Retrieval framework based on approximate matching model

Massive music information resource retrieval is divided into three levels according to the content in Fig. 2, structure and function. The first layer is the application layer, which provides an accessible interface for users to query information resources. The second layer is the service layer. During the retrieval of information resources, multiple servers are used to define user access rights, ensure network security, set identity registration interfaces, and provide data backup and management functions. The third layer is the data layer, which includes information of multiple data brokers. The expression of information resources is diversified. Under the information resource retrieval framework, it uses the approximate matching principle to calculate the connection value between query tree resources and standard tree information resources through data association. The calculation formula is as follows:

$$\lambda = \frac{\vartheta}{\sqrt{a^2 + b^2}}\theta \tag{3}$$

In formula (3), a and b represent the data values in the horizontal and vertical directions of the query tree respectively; ϑ stands for massive music information resources; θ represents the deflection angle between the query tree and the massive music information resources. According to the actual demand of information resources, the standardized description method is adopted to analyze the standard tree of massive music information resources, set the standard threshold range, uniformly classify massive music information resources, and determine the relationship between query tree data and multimodal data of the standard tree.

Using deep learning technology to detect the matched resources, it can adapt to scanning, photographing, etc. in various occasions, and has high recognition accuracy. Since the network structure of the deep learning algorithm is relatively simple, overfitting to specific resources can be avoided. According to the feature map of the backbone network, a block is generated, and after a series of subsequent processing, a resource area is finally formed, and the effective resources in this area are marked. The detailed marking process is as follows:

Step 1: Obtain feature resource databases of different sizes by sampling input resources at different levels, and construct a feature pyramid according to the feature resource database to obtain a feature resource database of uniform size;

Pre process the marked candidate resources according to the weight of the index feature components, and determine the features of the feature resource library:

$$R = \frac{\omega_1 x_1 + \omega_2 x_2 + \omega_3 x_3}{\omega_1 + \omega_2 + \omega_3} \tag{4}$$

In formula (4), ω_1, ω_2 and ω_3 respectively represent the weights of the input layer, hidden layer and output layer in the deep learning network; x_1, x_2, x_3 respectively represent the input features of the input layer, hidden layer and output layer in the deep learning network.

Step 2: Use the feature resource database under the same scale for training, and the training results can effectively distinguish between effective resources and invalid resources, reducing the overlap between information resources;

Step 3: The boundary supervision is introduced into the threshold resource database and used as the threshold of the unified scaling feature resource database;

Step 4: Annotate the resource library based on in-depth learning of the target annotation model;

Step 5: Obtain the specific content of multiple resource databases;

Step 6: In each contract section, determine the feature vectors of the paragraphs respectively, and combine them in a certain order to generate the corresponding paragraph feature vectors;

Step 7: Adopt the paragraph feature vector sequence as the input value of the paragraph labeling mode, so that the labeling model outputs a label sequence consistent with the prediction result, and the prediction structure of each resource segment is determined according to the order of the music information resources;

Step 8: In the resource library, the length of the resource is equal to the length of the predicted structure label, and the label includes the title, the content of the clause and the content of the resource.

4 Intelligent Hierarchical Index Construction

In order to obtain massive music information resources, it is necessary to filter out abnormal data, and the abnormal behavior of massive music information resources is resource anomaly. Therefore, a hierarchical index construction method based on deep learning is proposed. Abnormal resources are data generated by abnormal indexing behavior. The research focus of abnormal index tracking is to find the abnormal index path and locate the abnormal resources. When the resource index is running stably, resource exceptions occur within time. For this situation, the time interval of resource sampling should be calculated to locate the abnormal resource location in time. The size of the resource buffer obtained based on the deep learning method determines the sampling interval, which is the time of index construction.

If the length of the buffer area is set to be L, the probability that the resources in the buffer area are not full within the time t can be expressed as:

$$P_2 = \frac{\sum_{n=1}^{m} (p_1 t)^t}{n} \tag{5}$$

In formula (5), n represents the resource statistics results; p_1 indicates the index success rate after user request; m indicates the number of indexes.

The sampling interval of the buffer length record of the entire buffer area can be expressed as:

$$T = e^{-L^n/l} \tag{6}$$

In formula (6), l represents the unit record length. According to the above formula, the sampling interval can be obtained, that is, the index construction time based on deep learning.

If the index of a certain period is classified, the information resources of each stage can be directly obtained in the query process, which saves a lot of time for query and evaluation. In this way, the index information can be obtained effectively at any time without over-expanding the index, and the phenomenon of over-expansion of the index size will not occur. Based on this, an inverted index based on time domain is constructed, as shown in Fig. 3.

It can be seen from Fig. 3 that the inverted index structure based on time domain actually contains three kinds of data, which are composed of inverted table (kn, tn), term dictionary (kn) and term index (tn). When querying music information resources for a period of time, we can obtain an original music information resource by using common query techniques according to the term dictionary and term index. Then merge the two inverted tables to keep the order of the first table, so as to obtain the files related to the time period.

In order to shorten the construction time of hierarchical index, the sampling interval is divided into k sub time period $t_1, t_2, \cdots t_k$, and the dynamic statistical rules are applied to all sub time periods to obtain k statistical results, thus forming the temporal variation relationship of music information resources. For the division of sub time periods, when the dynamic statistical time is the same, the statistical results of each sub time period

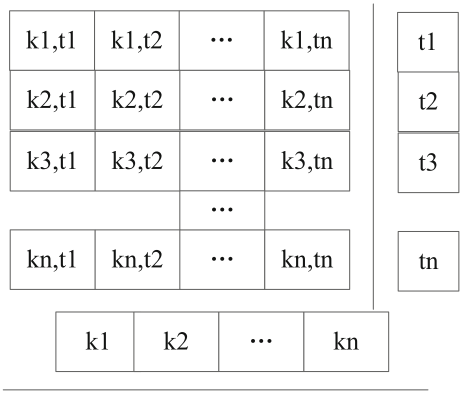

Fig. 3. Inverted index structure based on time domain

are consistent, so that the division results can be quickly obtained, the index grading time can be shortened, and the purpose of quickly building a hierarchical index can be achieved.

5 Intelligent Retrieval Based on Deep Learning

5.1 Information Resource Preprocessing

Using sparse reconstruction technology to preprocess information resources, taking the sparse representation of information resources as a prerequisite, reasonably constructing the measurement matrix and using the signal reconstruction algorithm are the keys to achieve accurate estimation. In the wireless positioning mode, the signal can be represented in a sparse format by selecting an appropriate sparse representation matrix, which can be directly constructed in a sparse manner. Divide the entire time domain into m equal parts according to the delay. Assuming that any division of the delay corresponds to a potential path, the resulting sparse gain vector can be expressed as:

$$z = \left[z_0 e^{-ucg_0}, z_1 e^{-ucg_1}, \cdots z_m e^{-ucg_m}\right] \tag{7}$$

In formula (7), $[z_0, z_1, \cdots z_m]$ represents sparse vector; u represents the number of divisions; c represents carrier frequency; g_m represents the time domain division result. Each sparse vector corresponds to a potential path. In order to reflect the sparsity of information resources in the time range, the number of potential multipath required is more than the number of actual multipath required. Therefore, the gain sparse vector is a sparse information resource gain coefficient, thus completing the information resource preprocessing.

5.2 Pseudo-Correlated Feedback

In order to obtain more query results, based on the initial query content input by the user, expansion words with similar meanings to the keywords can be selected, and combined with the input content to form expanded query terms, so as to improve the richness of the query content, accuracy and completeness. The pseudo-relevant feedback technology can effectively realize query expansion, and the expansion words obtained through this technology are the results of maximizing the expansion of the query content. Extract expansion words to further enrich the query content. It can be seen that the number and accuracy of expansion words are determined by this part of online music information.

The pseudo correlation feedback technology has an assumption that the sorted online music information in the query results obtained according to the original query content of the user is indeed related to the keyword sentences entered by the user. The process of initial sorting of query results based on this is as follows:

Step 1: Sort the query results for the first time according to the original content entered by the user, and select the network music information with the highest rank according to the relevance of the content.

Step 2: Extract content keywords from the above k pieces of online music information, and use the top two words with the most occurrences as expansion words related to the user input.

Step 3: A second query is performed according to a new query phrase composed of a combination of keywords input by the user and an expanded word, and a new query result is obtained.

Step 4: Use the query likelihood model of the search engine to initially sort the above results, and the sorting basis, that is, the correlation, is calculated by the following formula:

$$\log \sigma(S) = \sum_{i=1}^{m} \log \frac{\chi_x + \nu p_i}{|S| + \nu} \tag{8}$$

In formula (8), χ_x represents word frequency; ν represents the smoothing parameter required by Dirichlet; p_i represents the distribution probability of key information in the data set; S represents the music similarity of the initial sorting.

5.3 Intelligent Search

With the support of the approximate matching process of massive music information resources, the design of the retrieval process of massive music information resources is as follows: Firstly, the user specifies the matching type in the query tree. After preprocessing, using the required preprocessing information, the corresponding matching algorithm is called to find the cost that matches the standard tree. According to the cost, the result of the matching cost less than the set threshold is fed back to the user, and the threshold can be set as:

$$\varepsilon = \sum_{i \in v} \varphi \cdot \omega_i \tag{9}$$

In formula (9), φ represents the approximate matching cost; ω_i represents the node weight; i represents the number of nodes; v represents the label value. This threshold is equivalent to one half of the cost of removing the entire query tree. It is a preset threshold. When no specific type is specified, the resource target metadata specification scheme tree is first used to preprocess the query tree. On this basis, based on the matching preprocessing information obtained, the retrieval tree and the standard tree are approximately matched, compared with the query tree embedding results, and the results are fed back to the user.

Due to erroneous data in the intelligent retrieval process of resources, it needs to be corrected using deep learning methods. The attention model is input into the feedforward neural network for training. The structure of the feedforward neural network based on deep learning is shown in Fig. 4.

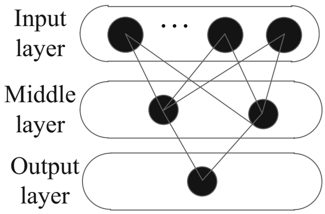

Fig. 4. Correction structure based on deep learning

It can be seen from Fig. 4 that the structure of feedforward neural network based on deep learning is composed of input layer, hidden layer and input layer, a characteristic matrix can be obtained from the structure training results, thus completing the information correction of the input part. The information resource set is taken as the input sample set, and the information resource dictionary is obtained through the same coding processing steps. The dictionary is used to encode the sample set of information resources, and the processed results are trained for word embedding to obtain the tag embedding matrix with location coding. In this part, we need to use two attention models, stack the two models and input them into the feedforward neural network, and take the output of the network as the input value of softmax function to obtain a probability. The prediction result with the highest probability is selected. By comparing with the dictionary, the corresponding information resources can be obtained and the output part of the correction model is completed.

Code the selected sample to obtain the required sample; By embedding the input samples, the marked embedding matrix is obtained; The input is an input sample set, while the output is an embedded matrix; A modified model is obtained through training steps; The data processing method based on input model is adopted, the speech to be corrected is vectorized, and it is input into the trained correction model to obtain the corresponding correction information resources, so as to obtain the intelligent correction results.

In a given set of massive music information resources, according to the principle of maximum likelihood estimation, the following log-likelihood function can be obtained:

$$f(x) = \arg\max_{i} \sum_{i=1}^{j} \log \gamma \left(X_j \mid x, \theta_i \right) \tag{10}$$

In formula (10), X_j represents the modal data under j training times; γ semantic concept prior distribution. The optimal estimation result of the prior parameters can be obtained through the maximization formula (10). Since the set of information resources follows the polynomial distribution of the prior parameters, the prior parameter estimates can be obtained according to the Lagrangian operator. In order to make the multimodal resource generation process can be estimated efficiently, it is necessary to follow the Gaussian distribution when generating multimodal resources from semantic vectors. In all resource sets, semantic concepts follow the Gaussian distribution, and these characteristic covariance matrices are consistent with the set covariance matrix, thus ensuring that the retrieval process has an optimal solution.

In the multimodal joint retrieval, the acquired resources are all multimodal. Compared with traditional retrieval methods, intelligent retrieval of music information resources has higher retrieval efficiency. Suppose the query resource set is x_n composed of n data, the resource to be retrieved is y_k composed of k resources, and the similarity between x_n and y_k can be calculated by the following formula:

$$Sim(x_n, y_k) = \frac{p(x_n, y_k, x_k, y_n)}{p(x_n, y_k)} \tag{11}$$

In formula (11), x_k represents k resources; y_n indicates that it is composed of n resources. When the relationship between the target retrieval resources and the query resources is obtained, the resources are sorted according to the order of similarity from large to small. The resources that do not have repeatability and are ranked in the top few items are the retrieval results.

6 Experiments

6.1 Experimental Setup

Converting the query interface into a user program calling interface can generate indexes independently, reduce the impact of users on the system, and enable the system to better focus on indicators. Based on this, the structure of the retrieval system is constructed, as shown in Fig. 5.

It can be seen from Fig. 5 that the system is mainly composed of index server and Web query interface. The index server is an independent system that allows users to set index type, index data and index record. The index server will index the relevant data according to the way set by the user without affecting other users; The Web query interface can be seen as an interface that can be directly embedded into the user's application program without modification, mainly including the program page, including index query, query application, and data query page. The main function is to index the Web query, index desktop application query, and database index query. Different program interfaces are different for different users, and users can make an index mark, It allows users to write their own query statements according to the index marks, and also make some small changes to the index marks and embed them in their own systems.

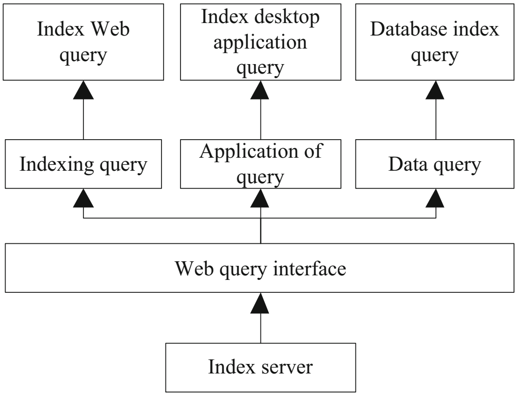

Fig. 5. Retrieval system structure

6.2 Experimental Dataset

Select the music query results of the HIFIVE network platform to establish an experimental sample set. The platform database saves 2.8 million pieces of music data. Each piece of data contains rich content information and various social information added by users. These information can be divided into 2 types Tag domain, one is music content information such as < publisher > < title >, and the other is social information such as < similarproducts > < tags >. From 2017 to 2020, a certain number of query records are screened from the platform database to build the training set and test set of the system, and their respective data compositions are shown in Table 1.

Table 1. Experimental dataset

Group	Training set	Test set
A	250 queries (2017)	100 queries (2018)
B	250 queries 100 queries (2017–2018)	400 queries (2018)
C	250 queries 100 queries 700 queries (2017–2018)	700 queries (2019)

6.3 Importing Information Sources

Since the information source is external data, the basic parameters of the information source must be introduced into the research. Figure 6 shows the input process of the information source.

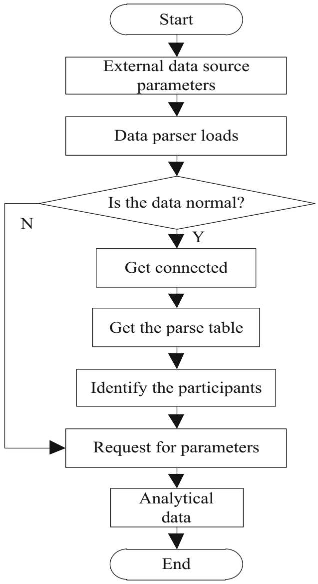

Fig. 6. Information source import implementation process

It can be seen from Fig. 6 that the user fills in the parameter information of the information source of the HIFIVE network platform into the view layer, and transmits it to the model layer in the form of URL. This method obtains the URL by calling the Controller function, analyzes it, and passes the analysis result to the model layer as the returned value. The model layer analyzes the information through the appropriate loading amount according to the judgment result of the returned value.

6.4 Experimental Indicators

For massive music information resources, the more index resources are, the greater the expansion coefficient is, the faster the retrieval speed is. If no index retrieval is created for all massive music information resources, the expansion coefficient is 0; If the index resource is as large as all resources, the expansion coefficient is 1. The index expansion coefficient formula can be expressed as:

$$\alpha = \frac{M}{N} \tag{12}$$

In formula (12), M represents the size of index information resources; N represents the size of all information resources.

The evaluation criteria for intelligent resource retrieval are the classic indicators in data retrieval, namely recall and retrieval error. The formula is:

$$P_N = \frac{N_a}{N} \times 100\% \tag{13}$$

$$e = hx - hx' \tag{14}$$

In formula (13) and (14), N_a represents the required data retrieved; hx represents the number of retrieved data; hx' represents the amount of data retrieved.

6.5 Experimental Results and Analysis

Using the retrieval method based on approximate resource matching, the retrieval method based on the combination of on-chain and off-chain, and the retrieval method based on deep learning, the retrieval speed is compared and analyzed, and the comparison results are shown in Fig. 7.

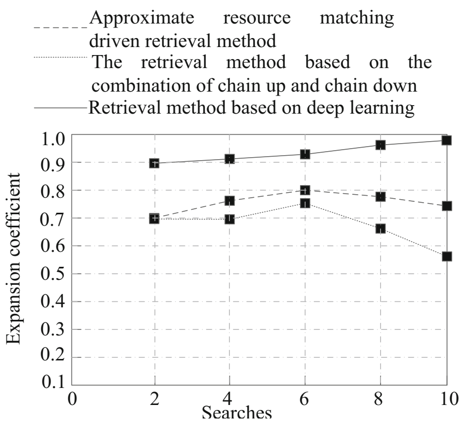

Fig. 7. Comparative analysis of retrieval speed of different methods

It can be seen from Fig. 7 that using the retrieval method driven by approximate resource matching, the expansion coefficient is the largest at the sixth retrieval number, which can reach 0.80, and the smallest at the second retrieval number, which is 0.70; Based on the retrieval method of combining on chain and off chain, the expansion coefficient is the largest at the sixth retrieval number, which can reach 0.75, and the smallest at the tenth retrieval number, which is 0.55; Using the retrieval method based on depth learning, the expansion coefficient is the largest in the 10th retrieval number, which can reach 0.99, and the smallest in the 2nd retrieval number, which is 0.90. It has a fast retrieval effect.

For retrieval integrity verification, three methods were used to compare retrieval recall, and the results are shown in Fig. 8.

It can be seen from Fig. 8 that the retrieval results of the approximate resource matching driven retrieval method and the combination of on chain and off chain retrieval method are inconsistent with the actual data, and the maximum difference between the retrieval results and the actual data is 5.0 kB in the fourth retrieval. The retrieval results using the method studied are basically consistent with the actual data, and the maximum error with the actual data is 0.30 kB only in the 10th retrieval.

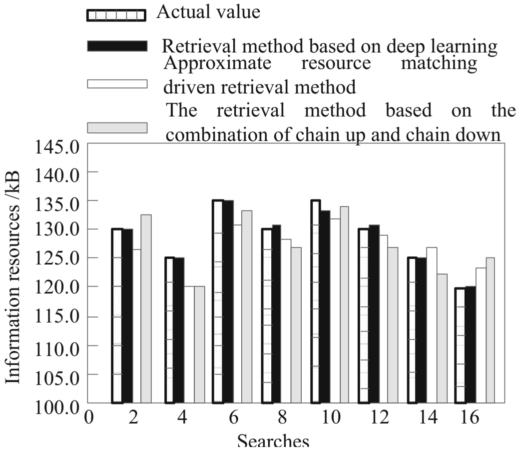

Fig. 8. Comparative analysis of retrieval recall rate of different methods

The retrieval error is analyzed by three methods respectively, and the results are shown in Fig. 9.

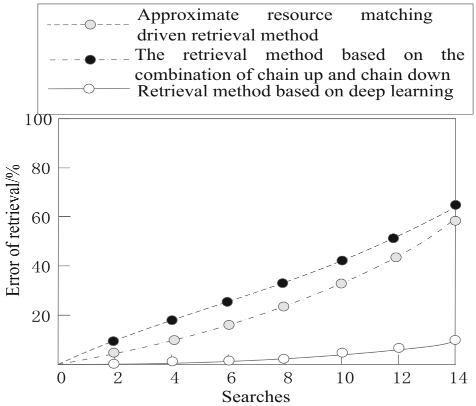

Fig. 9. Comparative analysis of retrieval errors of different methods

As can be seen from Fig. 9, using the retrieval method driven by approximate resource matching, the maximum retrieval error is 60%; Using the retrieval method based on the combination of on-chain and off-chain, the maximum retrieval error is 65%; Using the retrieval method based on deep learning, the maximum retrieval error is 10%.

7 Conclusion

In order to effectively improve the quality of online music query results, an intelligent retrieval method of massive music information resources based on deep learning is proposed and designed. The pseudo correlation feedback technology is used to initially expand the search scope of online music and obtain the initial sorting results. The depth learning theory is introduced into this matching process, and then the convolution depth confidence network model in depth learning is studied in depth. The retrieval efficiency based on the depth learning model is also compared with the experiment results. The results show that this method greatly improves the accuracy of network music information query, and provides users with a good experience. In the future, we need to think about how to recommend the search results to music lovers, so as to improve the efficiency of users' use of online information resources, realize the full use of online music information resources, and promote the further development of music information resources retrieval.

References

1. Chen, L., Liu, Y.: Design of interactive information retrieval algorithm based on user demand mining. Comput. Simulat. **39**(5), 418–422 (2022)
2. Wang, H., Liu, L., Lin, M., et al.: Music personalized recommendation algorithm based on k-means clustering algorithm. J. Jilin Univ. Eng. Technol. Edn. **51**(5), 1845–1850 (2021)
3. Li, W., Zhang, Z., Wang, J.: Storage and retrieval method of massive small files based on EHDFS. Comput. Eng. Design **43**(2), 376–383 (2022)
4. Chen, Z., Zhang, M., Guo, M., et al.: Music lyrics-melody intelligent evaluation algorithm based on sequence model. CAAI Trans. Intell. Syst. **15**(1), 67–73 (2020)
5. Jin, S., Qin, J.: Music visualization intelligent design based on image style migration. Packaging Eng. **41**(16), 193–198 (2020)
6. Li, Y., Guo, J., Yu, Z., et al.: Improve low-resource cross-language information retrieval by constructing pseudo query sentences based on word mapping . J. Shanxi Univ. (Nat. Sci. Edn.) **45**(2), 322–331 (2022)
7. Liang, S., Zhu, H., Wu, D.: Cross-lingual information retrieval based on named entity recognition and translation of public digital cultural resources. Library Develop. **1**, 87–95 (2022)
8. Lu, W., Tao, C., Li, H., et al.: A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. Remote Sens. Environ. **270**(1), 112830–112842 (2022)
9. Wang, W., Sohail, M.: Research on music style classification based on deep learning. Comput. Math. Methods Med. **22**(10), 1–8 (2022)
10. Furner, M., Islam, M.Z., Li, C.T.: Knowledge discovery and visualisation framework using machine learning for music information retrieval from broadcast radio data. Expert Syst. with Appli. **182**, 115236.1–115236.11 (2021)