Journal of Engineering Science and Technology Review 16 (6) (2023) 195 - 201

Research Article

JOURNAL OF
Engineering Science and Technology Review

www.jestr.org

Intelligent Retrieval Method for Multimedia Digital Audio Based on Deep Learning

Shuixin Zhang[1], Yanyong Lin[1], Lian Chen[2,*] and Cheng Jiang[3]

[1]*Wuhan Vocational College of Software and Engineering, Wuhan, 430000, China*
[2]*Huanggang Normal University, Huanggang, 438000, China*
[3] *Dongseo University, Busan 47011, Republic of Korea*

Received 23 June 2023; Accepted 30 November 2023

Abstract

The amount of multimedia digital audio in networks has increased sharply with the rapid development of multimedia technology. The data carried by multimedia digital audio is valuable, and how to retrieve useful audio data in the environment of massive multimedia digital audio big data become a key research focus. An intelligent retrieval method for multimedia digital audio based on deep learning was proposed to achieve accurate and fast retrieval for multimedia digital audio. Bayesian and K-SVD (singular value decomposition) algorithms were used to enhance the preprocessing of multimedia digital audio, the preprocessed multimedia digital audio was taken as input, and the convolutional neural network in the deep learning algorithm was used to extract the multimedia digital audio feature vectors. After the similarity of multimedia digital audio feature vectors was calculated by using the concept tree method, the multimedia digital audio corresponding to the highest similarity was selected as the retrieval result. Experimental results show that, after this method preprocesses multimedia digital audio, the maximum peak signal-to-noise ratio of the audio can reach 78.4 dB, and the accuracy of extracting multimedia digital audio features is as high as 100%. The similarity values when retrieving media digital audio are higher than 0.95. Compared with regression theory, incomplete weight retrieval methods, and the deep hashing algorithm multimedia retrieval method, the method obtained from this study has stronger multimedia digital audio intelligent retrieval capabilities and a more significant application effect. The proposed method can retrieve useful audio data in a massive multimedia digital audio big data environment, achieve accurate and fast multimedia digital audio retrieval, and provide strong support for the development and application of the field of audio information retrieval.

Keywords: Deep learning, Multimedia, Digital audio, Intelligent retrieval method, Convolutional neural network, Feature extraction

1. Introduction

With the rapid development of multimedia technology, a large amount of multimedia digital audio is generated in networks, with varying storage methods. As a main component of data interaction in the network, multimedia digital audio not only carries a large amount of valuable data but is also closely related to people's lives and work. The large amount of information in multimedia digital audio and its diverse characteristics have made the retrieval of multimedia digital audio more difficult, which has therefore attracted increasing attention. People have higher expectations for locating, searching, and recommending audio content such as music, radio programs, and sound effects.

Traditional keyword-based text retrieval methods cannot meet this demand. First, traditional audio retrieval methods have difficulty accurately extracting and representing the semantic information of audio. They often face difficulties when processing diverse and complex audio, and have difficulty accurately matching and retrieving the results required by users [1]. Second, traditional methods have difficulty understanding and analyzing the semantic content and emotional characteristics of audio [2]. For example, the emotion, sentiment, or implicit semantic information in audio cannot be accurately identified, thus limiting the accuracy and personalization of retrieval and recommendation [3]. In addition, current audio retrieval methods are faced with challenges in obtaining and annotating audio data [4], which leads to data sparsity problems during model training and matching, which in turn affects the retrieval and recommendation performance of the system.

Therefore, this study designs an intelligent retrieval method for multimedia digital audio based on deep learning. When the multimedia digital audio retrieval framework starts to run, enhanced preprocessing is performed on the multimedia digital audio dataset and retrieval audio target based on Bayesian and K-SVD (singular value decomposition) algorithms. The preprocessed multimedia digital audio is used as the input of the Convolutional Neural Network (CNN) of deep learning. After the CNN model in the deep learning algorithm is used to extract the multimedia digital audio features, the concept tree algorithm is used to calculate the digital audio vector similarity. On the basis of a similarity design audio retrieval program, after the similarity between the retrieval target audio and the audio features extracted from the multimedia digital audio dataset is measured, the results are sorted in descending order to obtain the multimedia digital audio-based retrieval results, improving the intelligent retrieval for multimedia digital audio.

The rest of this study is structured as follows: Section 2 reviews the current research status, and then the overall process of the deep learning multimedia digital audio

*E-mail address: fiona-lian@hgnu.edu.cn
ISSN: 1791-2377
doi:10.25103/jestr.166.24

intelligent retrieval method is given to address the problem of intelligent retrieval for multimedia digital audio. In Section 3, on the basis of Bayesian and K-SVD algorithms for preprocessing multimedia digital audio, the multimedia digital audio feature vectors of deep learning are extracted. In Section 4, the concept tree algorithm is used to calculate the similarity of digital audio vectors and an audio retrieval program is designed based on the similarity to obtain the retrieval results. The practical application effect of this method is verified through experiments. Section 5 is the conclusion

2. Literature Review

Previous retrieval methods for multimedia digital audio manually transmitted multimedia digital audio to the multimedia database and used manual methods to annotate multimedia digital audio. However, this method increases the time cost and will lead to an increase in labor costs. At the same time, due to manual annotation, the accuracy of labeling multimedia digital audio will be low, affecting the efficiency of the retrieval method for multimedia digital audio. Many scholars are studying audio retrieval algorithms. For example, Furner et al. [5] proposed a music information retrieval method within radio station data, which uses a machine learning algorithm within the ZeitMetric framework to retrieve music information and automatically mark multimedia digital audio and then uses a self-organizing map to present the multimedia digital audio search results. This method takes a long time to retrieve multimedia digital audio information in practical applications.

Many scholars have also studied music information retrieval methods. For example, Zali et al. [6] proposed an application method for music information retrieval, which is based on the idea of harmonic impact separation within multimedia digital audio signals and retrieves multimedia by decomposing independent spectrograms of digital audio data. However, this method is affected by interference noise in the multimedia digital audio signal, and the output results are not accurate enough and the application effect is not good. Kale et al. [7] proposed an audio information retrieval method using autoencoders, which implements the retrieval method for multimedia digital audio through retrieval encoding after deep stacking sparse automatic encoding of multimedia digital audio in the database. However, when this method encodes audio content with high similarity, such content can be ignored easily, thus resulting in a poor application effect. Steltner et al. [8] proposed an all-sky multimedia digital audio search method in public data, which divides multimedia digital audio search into several stages. The staged search method can ensure the accuracy of multimedia digital audio search. However, this method has a high time complexity and a poor application effect. Cercevik et al. [9] proposed a metaheuristic search method. When this method is applied to search for multimedia music, it will set constraints on the search process and optimize the search. However, this method is not effective when used to search for multimedia music. Andic et al. [10] proposed the application method of a robust crow search algorithm in audio retrieval. This method regards multimedia digital audio as the original vector and then designs the optimal retrieval number and uses the crow search method to retrieve multimedia digital audio. However, this method has poor operation stability and thus has a poor application effect.

In terms of retrieval methods for multimedia digital audio, Perepu et al. [11] proposed a retrieval method for multimedia digital audio based on the longest common subsequence algorithm. This method divides the multimedia digital audio into sequences and obtains the results based on the similarity of each subsequence. However, this method involves human subjectivity when dividing multimedia digital audio subsequences, resulting in a smaller final search range. Preininger et al. [12] proposed an agent dialogue retrieval method for multimedia digital audio, which sets up agent dialogue retrieval based on the keywords and title navigation of multimedia digital audio. Intelligent retrieval of multimedia digital audio is realized after the mode and retrieval frequency are selected. However, this method has a small search range and is not widely used. Guerra et al. [13] proposed a new multimedia digital audio intelligent search method based on the feasible interior point method to solve the constrained optimization problem of multimedia digital audio search. The method has a high retrieval accuracy for multimedia digital audio, but its search range is small due to the large amount of audio data. Zhang et al. [14] proposed an evolutionary search method guided by a graph neural network. This method establishes a graph neural network model and obtains retrieval results of multimedia digital audio through model iteration. However, this model is prone to falling into local extremes when iterating, resulting in a smaller application scope.

In the above-mentioned studies, some scholars calculated the incomplete weight of multimedia audio and then designed a retrieval program based on regression theory to obtain retrieval results of multimedia digital audio [15-16]. However, these methods still have the problem of long time consumption when retrieving multimedia digital audio information. To improve the accuracy and reliability of retrieval methods for multimedia digital audio, some studies introduced machine learning algorithms. The metaheuristic search method is used to set constraints on the search process and optimize the search. Multimedia digital audio is treated as the original vector, the optimal retrieval quantity is designed, and the crow search method is used to obtain the retrieval results of multimedia digital audio. After the digital audio is divided into sequences, the retrieval results of multimedia digital audio are obtained based on the similarity of each subsequence. These technologies shorten the time consumed in retrieving multimedia digital audio information, making audio retrieval more accurate and feasible.

In sum, the retrieval method for multimedia digital audio is a challenging field, and scholars continue to explore this field through different methods and technologies to improve the application effect and accuracy of audio retrieval. However, current research retrieval methods for multimedia digital audio still require further research to provide more comprehensive retrieval results, extract emotional features in audio, and provide users with a more personalized experience.

3. Methodology

3.1 Overall process of deep learning multimedia digital audio intelligent retrieval method

In this study, a multimedia digital audio intelligent retrieval method based on deep learning is proposed. The overall process of the method is shown in Fig.1. When the retrieval framework for multimedia digital audio starts to run, it first performs enhanced preprocessing on the multimedia digital audio dataset and the retrieval audio target. After the CNN model in the deep learning algorithm is used to extract the multimedia digital audio features, the retrieval multimedia audio target is extracted. After similarity measurement is performed on the audio features extracted from the digital audio dataset, the results are sorted to obtain the retrieval results of multimedia digital audio.

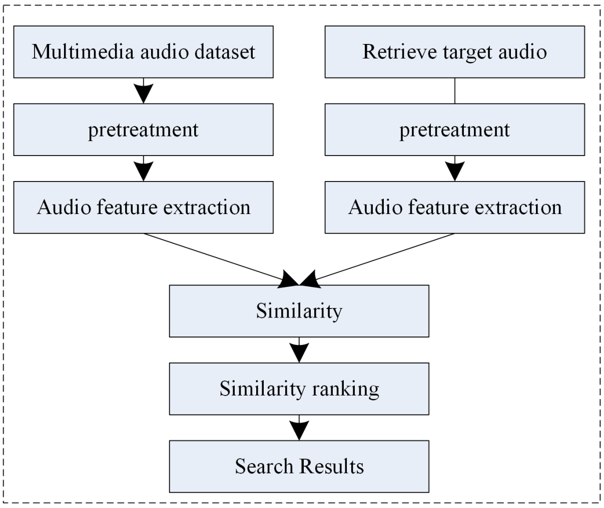


Fig.1. Overall Process of Multimedia Digital Audio Intelligent Retrieval Method Based on Deep Learning

3.2 Multimedia digital audio preprocessing based on Bayesian and K-SVD

All the multimedia digital audio in the network is used to establish a dataset. The multimedia digital audio signal is represented by $x_i \in R^n$ in the dataset, where R^n is the multimedia digital audio dataset, and n is the number of audio in the dataset. The sparse representation dictionary algorithm (K-SVD) is applied to multimedia digital audio in the dataset, and then each multimedia digital audio in the dataset has an over-complete dictionary $D \in R^{n \times K}$, where K is the number of over-complete dictionaries. The sparse representation of multimedia digital audio can be denoted as (ε, L, D), and ε and L represent the allowable error and the number of dictionary atoms, respectively. Let S represent the multimedia digital audio signal with interference noise. The interference-free signal is X, the noise is N, and $S = X + N$. On the basis of the maximum posterior estimate in the Bayesian algorithm, the observation function of the multimedia digital audio is obtained using the following formula:

$$\hat{a} = \arg\min_{a_i} \left\| \|Da_i - s_i\|_2^2 + \zeta \|a_i\|_0 \right] \tag{1}$$

Where $\hat{a}_i$ is the observed sparse coefficient of multimedia digital audio; Da_i is the original observed sparse coefficient; ζ represents the adjustable variable; and S_i is the i^{th} multimedia digital audio signal.

The sparse representation of the multimedia digital audio signal can be obtained by using formula (1).

$\sqrt{N} \times \sqrt{N}$ represents the size of the multimedia digital audio signal X after sparse representation, where each signal block belongs to the sparse domain[15]. Then, the maximum a posteriori estimation formula of the multimedia signal is

$$\min_{X,D} = \left\{ \|X - S\| + \sum_{ij} \zeta_{ij} \|a_{ij}\|_0 + \sum_{ij} \|Da_{ij} - R_{ij}S\|_2^2 \right\} \tag{2}$$

where R_{ij} is the signal block matrix; ζ_{ij} is the adjustment variable of signal blocks i and j; and a_{ij} is the sparse coefficient of signal blocks i and j.

On the basis of the above process, the enhanced preprocessing process for multimedia digital audio signals is as follows:

1) Initialization: Let $X = S$. The complete dictionary is initialized, and the number of iterations is set to 1.

2) Sparse coding is performed on multimedia digital audio to obtain the sparse coefficient a_{ij} of the audio signal.

3) Each column in the complete dictionary is updated, and the corresponding sparse coefficient is stored in the data subset ω_i.

4) The residual of each multimedia digital audio signal is calculated in ω_i [16], and the residual is regarded as each column of the matrix E_i. After matrix E_i is decomposed, the multimedia digital audio signal X without noise can be obtained.

After the above steps, multimedia digital audio enhancement preprocessing is achieved.

3.3 Deep learning multimedia digital audio feature vector extraction

The preprocessed multimedia digital audio is used as the input of the CNN of deep learning, and the CNN is used to extract the media digital audio feature vectors to improve the accuracy of the media digital audio feature vector extraction and lays the foundation for the intelligent retrieval of the multimedia digital audio. The starting layer of CNN is the input layer, the ending layer is the output layer, and the middle layer of CNN includes the pooling layer, fully connected layer, and convolution layer [17]. The multimedia digital audio feature vector extraction process of deep learning is as follows:

(1) The preprocessed multimedia digital audio vector is input into the CNN input layer.

(2) The convolution layer is used to implement convolution operations to extract multimedia digital audio feature vectors [18].

The convolutional layer consists of multiple convolution kernels [19], which are used to implement convolution transformation to extract multimedia digital audio feature vectors. The operation process of the convolution layer is shown in formulas (3) and (4)

$$s^{l+1} = b^l + w^l x^{l+1} \tag{3}$$

$$x^{l+1} = f(s^l) \tag{4}$$

where s^{l+1} is the input of layer $l+1$; b^l is the bias of layer l; l is the number of convolutional layers; w^l is the weight of the convolution kernel of layer l; x^{l+1} is the output of layer $l+1$; and f is the activation function.

After the convolution operation, the activation function can be used to improve the robustness of CNN. The ReLU function with fast convergence speed is selected as the CNN activation function, and its formula is as follows:

$$f(x) = MAX(0, x) \tag{5}$$

(3) After pooling is implemented through the pooling layer, the local features of the multimedia digital audio vector extracted by the statistical convolution layer are aggregated to avoid overfitting.

The pooling process of the pooling layer is expressed as follows:

$$Y_g^{l+1} = f(b^{l+1} + w^{l+1} \times down(U_g)) \tag{6}$$

where b^{l+1} is the bias of layer $l+1$; $down(U_g)$ is the pooling function; and w^{l+1} is the weight of layer $l+1$.

(4) Multimedia digital audio vector classification is implemented through the classification function softmax in the fully connected layer [20], and the output layer outputs the multimedia digital audio feature vector extraction results.

The linear prediction expression of the n^{th} category is

$$z_n = b_n + w_n^T x \tag{7}$$

where w^l is the bias, and w is the weight.

The expression of the softmax function is as follows:

$$\delta(z) = \frac{\exp(z_n)}{\sum_{j=1}^{m} \exp(z_j)}, j = 1, 2, \cdots m \tag{8}$$

With the use of the function of formula (8), the multimedia digital audio feature vector can be output.

After the above steps, the characteristics of the multimedia digital audio vector are obtained.

3.4 Audio vector similarity calculation and audio retrieval method

On the basis of the obtained digital audio feature vectors, the concept tree algorithm is used to calculate the digital audio vector similarity and an audio retrieval program is designed based on the similarity. The detailed process is as follows:

Multimedia digital audio feature vectors are used to establish a concept tree and determine different weight factors of the concept tree.

(1) Density factor

After multimedia audio feature vectors are used to establish a concept tree, the densities of different areas in the concept network are different. Let p represent the parent node of the multimedia digital audio feature vector concept tree, then the out-degree and in-degree of this node and node c can be determined by directed edges. $p \to c$ represents, then the weight calculation formula of the directed edge $p \to c$ is as follows:

$$weight\_\deg(c, p) = \frac{in\deg ree(c) + out\deg ree(p)}{in\deg ree(H) + out\deg ree(H)} \tag{9}$$

Where $weight\_\deg(c, p)$ represents the weight of directed edge $p \to c$; $in\deg ree(c)$ and $out\deg ree(p)$ represent the in-degree and out-degree of concept nodes c and p, respectively; H represents the ontology of the multimedia digital audio feature vector concept tree; $in\deg ree(H)$ and $out\deg ree(H)$ represent the in-degree and out-degrees of the ontology, respectively.

(2) Depth factor

In the multimedia digital audio feature vector concept tree, the depth factor is the refinement of each layer's concept of the previous layer. Below the directed edge $p \to c$, its depth factor weight $weight\_len(c, p)$ is calculated as follows:

$$weight\_len(c, p) = \sum_{n=1}^{depth(p)} \frac{1}{2^n} \tag{10}$$

where $depth(p)$ represents the depth of the node concept in the hierarchical tree, and n is the depth level.

On the basis of the above two directed edge weight factors, they are integrated into

$$weight(c, p) = h(weight\_\deg(c, p), weight\_len(c, p)) \tag{11}$$

where *C* represents the comprehensive weight value of the directed edge *D*, and *E* represents the integration function.

On the basis of the results of formula (11), the similarity $Dis(c, p)$ of multimedia digital audio feature vector concept tree nodes c and p is calculated by using the following formula:

$$Dis(c, p) = \frac{1}{weight(c, p)} \tag{12}$$

The similarity of multimedia digital audio feature vectors can be obtained using the formula (12). The results of this formula are arranged in descending order, and the audio in the multimedia database corresponding to the maximum similarity value is the search result.

4. Results Analysis

With NetEase Cloud multimedia music software taken as the experimental object, the proposed method is used to retrieve the multimedia digital audio in the multimedia music software to verify its practical application effect.

First, 10 audio signal segments were intercepted from the music software, and the peak signal-to-noise ratio before and after preprocessing was used as a measurement indicator to verify the enhanced preprocessing capabilities of the proposed method for multimedia digital audio signals. The test results are shown in Table 1. Table 1 shows that the maximum peak signal-to-noise ratio of the 10 multimedia digital audio clips before enhanced preprocessing was only 60.4 dB. After the proposed method was applied for enhanced preprocessing, the maximum peak signal-to-noise ratio was 78.4 dB, and the minimum peak signal-to-noise ratio was as high as 60.5 dB, indicating a greatly improved peak signal-to-noise ratio of the multimedia digital audio clips after enhanced preprocessing. Thus, the proposed method can effectively enhance the preprocessing of multimedia digital audio.

Extracting multimedia digital audio features is the basis for retrieval. The proposed method was used to extract different types of features from multimedia digital audio, and its ability to extract multimedia digital audio features was analyzed. The test results are shown in Fig.2. An analysis of Fig.2 shows that the proposed method achieved a maximum feature extraction accuracy of 100% and a minimum accuracy of 99.4% when used to extract different multimedia audio feature vectors. Therefore, the proposed method has a strong ability to extract multimedia digital audio features and retrieve multimedia digital audio.

Table 1. Peak Signal to Noise Ratio (dB) of Multimedia Digital Audio

Multimedia digital audio clips	Before pretreatment	After pre-treatment
Fragment 1	33.6	60.5
Fragment 2	40.2	71.4
Fragment 3	51.9	75.9
Fragment 4	36.8	66.8
Fragment 5	42.6	76.4
Fragment 6	43.7	70.3
Fragment 7	50.6	78.4
Fragment 8	56.4	77.6
Fragment 9	58.3	70.4
Fragment 10	60.4	74.2

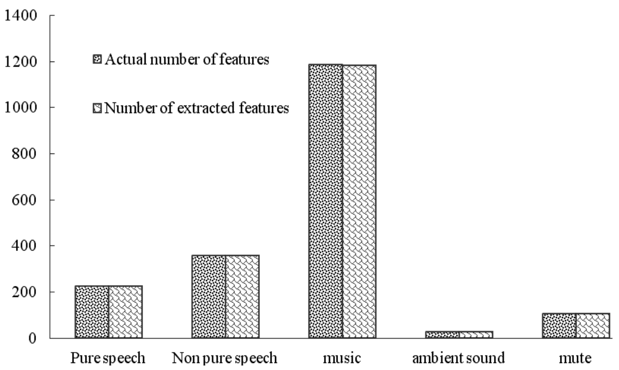


Fig. 2. Multimedia Digital Audio Feature Extraction Results

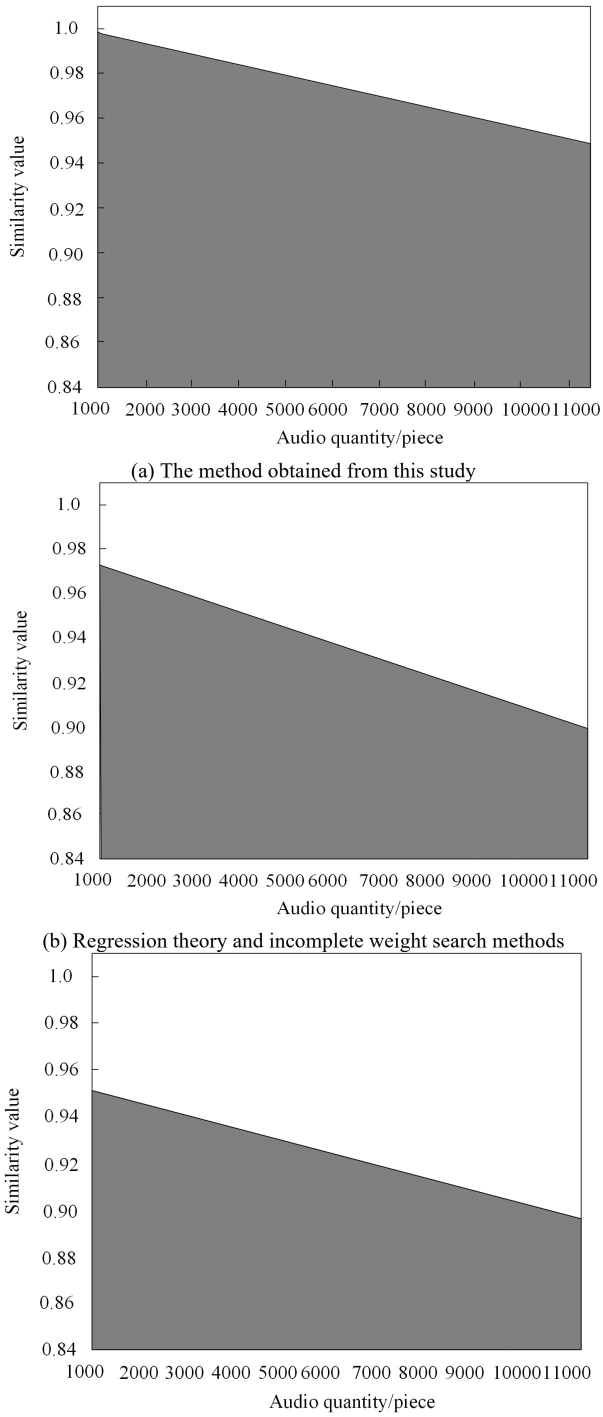
The multimedia audio retrieval capability of the proposed method was verified by using eight audio clips as retrieval targets, and the proposed method was used to retrieve multimedia digital audio in the music software. The retrieval results are shown in Table 2.

Table 2. Retrieval results of multimedia digital audio

Audio retrieval clip	Search Results	Similarity
Fragment 1	Whale incarnating as an island	0.98
Fragment 2	As boundless as the sea and sky	0.99
Fragment 3	Mudanjiang River	0.97
Fragment 4	Fragrance of rice	0.95
Fragment 5	High in the clouds	0.96
Fragment 6	Daytime cat	0.98
Fragment 7	Unpredictable	0.97
Fragment 8	My Boo	0.98

Table 2 shows that the proposed method can effectively retrieve the corresponding multimedia digital audio from the database of the music software based on the multimedia audio clips, and the similarity value between the audios is at least 0.95. This numerical value shows that the multimedia digital audio retrieved by this method has higher accuracy and thus, the method has better application effects.

The retrieval effect of multimedia digital audio of this method is verified further by using the retrieval similarity as a measurement index, consistent with regression theory and incomplete weight retrieval method (reference [3]) and the multimedia retrieval method of deep hashing algorithm (reference [4]), to compare and test the retrieval capabilities of the three methods when the database contains a large amount of multimedia digital audio. The test results are shown in Fig.3. Fig.3 shows that when the proposed method retrieves multimedia digital audio, the similarity value decreases with the increase in the number of audios, but the decrease is smaller. When the number of audio is 11,000, the similarity value is still as high as around 0.95. However, the retrieval methods of regression theory and incomplete weight and the multimedia retrieval method of the deep hashing algorithm achieve a greater reduction in similarity value and a lower value. This result shows that when retrieving multimedia digital audio, the proposed method is less affected by the number of audio in the database and has a strong retrieval ability.



(a) The method obtained from this study
(b) Regression theory and incomplete weight search methods
c. Multimedia retrieval method using deep hashing algorithm
Fig. 3. Similarity of retrieval methods for multimedia digital audio

5. Conclusion

This study explores an intelligent retrieval method for multimedia digital audio based on deep learning. On the basis of the maximum posterior estimation within the Bayesian algorithm, the observation function of multimedia digital audio is obtained to achieve enhanced preprocessing of multimedia digital audio. The CNN model in the deep learning algorithm is used to extract multimedia digital audio feature vectors. These multimedia digital audio feature vectors are then used to establish a concept tree, and the different weight factors of the concept tree are determined. Multimedia digital audio is then retrieved by calculating the similarities and sorting. The following conclusions were obtained:

(1) After the proposed method is applied to enhance the preprocessing of multimedia digital audio signals, the maximum peak signal-to-noise ratio is 78.4 dB, and the minimum peak signal-to-noise ratio is as high as 60.5 dB. Thus, the peak signal-to-noise ratio of the preprocessed multimedia digital audio clips is greatly enhanced, indicating that the proposed method can effectively enhance the preprocessing of multimedia digital audio.

(2) The proposed method has a maximum feature extraction accuracy of 100% and a minimum accuracy of 99.4% when used to extract multimedia audio feature vectors for different types of audio features. This result shows that the proposed method has a strong ability to extract multimedia digital audio features and retrieve multimedia digital audio.

(3) The proposed method can effectively retrieve the corresponding multimedia digital audio from the music software database based on the multimedia audio clip, and the similarity value between the audios is at least 0.95. Thus, the multimedia digital audio retrieved by this method has high accuracy and a good application effect.

(4) When the proposed method retrieves multimedia digital audio, the similarity value decreases with the increase in the number of audios, but the decrease is smaller. When the number of audios is 11,000, the similarity value of the proposed method when retrieving multimedia digital audio is still the same, reaching about 0.95, thereby showing that when retrieving multimedia digital audio, the proposed method is less affected by the number of audios in the database and has strong retrieval ability.

Deep learning-based audio retrieval methods can combine other media information to provide more comprehensive search results. However, the fusion and effective utilization of multimodal information remains a challenge. Future research can explore better methods to achieve cross-modal feature learning and fusion and provide richer and more accurate retrieval results. The problem of data scarcity must be solved, and the generalization ability of the model can be improved through methods such as semi-supervised or unsupervised learning. The interpretability and credibility of the model also need to be improved, and corresponding evaluation indicators and methods should be designed so that users can understand and trust the results of the model. On the premise of protecting user privacy, a safe and reliable audio retrieval system needs to be designed, and corresponding privacy protection policies and technologies must be formulated.

References

[1] X. Bai and J. Li. "Applied research of knowledge in the field of artificial intelligence in the intelligent retrieval of teaching resources," *Sci. Program*., vol. 2021, pp. 1-11, Jun. 2021.
[2] Z. Wang. "Analysis of user personalized retrieval of multimedia digital archives dependent on BP neural network algorithm," *Adv. Multimedia.*, vol. 2021, Dec. 2021, Art. no.2630254.
[3] K. Zhang, Y. M. Wang, and J. Zheng. "Regret theory-based case-retrieval method with multiple heterogeneous attributes and incomplete weight information," *Int. J. Comput. Intell. Syst.*, vol. 14, no. 1, pp. 1022-1033. Mar. 2021.
[4] Z. Li, Y. Zhou, and H. Wang. "Social media image classification and retrieval method based on deep hash algothm," *Int. J. Web. Based. Communities.*, vol. 18, no. 3-4, pp. 276-287, Sep. 2022.
[5] M. Furner, M. Z. Islam, and C. T. Li. "Knowledge discovery and visualisation framework using machine learning for music information retrieval from broadcast radio data," *Expert Syst. Appl.*, vol. 182, Nov. 2021, Art. no.115236.
[6] Z. Zali, M. Ohrnberger, F. Scherbaum, F. Cotton, and E. P. Eibl. "Volcanic tremor extraction and earthquake detection using music information retrieval algorithms," *Seismol. Res. Lett.*, vol. 92, no. 6, pp. 3668-3681, Nov.2021.
[7] M. Kale and S. Mukhopadhyay. "Efficient color image retrieval method using deep stacked sparse autoencoder," *J. Electron. Imaging.*, vol. 31, no. 2, Art. no.023003, Mar.2022.
[8] B. Steltner, *et al.* "Einstein@ Home all-sky search for continuous gravitational waves in LIGO O2 public data," *Astrophys. J.*, vol. 909, no.1, Mar. 2021. Art. no.79
[9] A. E. Çerçevik, Ö. Avşar, and A. Dilsiz. "Optimal placement of viscous wall dampers in RC moment resisting frames using metaheuristic search methods," *Eng. Struct.*, vol. 249, Dec. 2021, Art. no.113108
[10] C. Andic, A. Ozturk, and B. Turkay. "Power system state estimation using a robust crow search algorithm based on PMUs with limited number of channels," *Electr. Power Syst. Res.*, vol. 217, Apr. 2023, Art. no.109126.
[11] P. Kumar Perepu. "OpenMP implementation of parallel longest common subsequence algorithm for mathematical expression retrieval," *Parallel. Process. Lett.*, vol. 31, no.2, May. 2021, Art. no.2150007.
[12] A. M. Preininger, *et al.* "Differences in information accessed in a pharmacologic knowledge base using a conversational agent vs traditional search methods," *Int. J. Med. Inform*., vol. 153, Sep. 2021, Art. no.104530.
[13] L. Guerra. "A class of new search directions for full-NT step feasible interior point method in semidefinite optimization," *RAIRO-OPER. Res.*, vol. 56, no. 6, pp. 3955-3971. Nov.2022.
[14] J. Zhang, A. Koneru, S. K. Sankaranarayanan, and C. M. Lilley. "Graph neural network guided evolutionary search of grain boundaries in 2D materials," *ACS Appl. Mater. Interfaces.*, vol. 15, no. 16, pp. 20520-20530, Apr. 2023.
[15] M. Zakariah, M. K. Khan, and H. Malik. Digital multimedia audio forensics: past, present and future. *Multimed. Tools. Appl.*, vol. 77, pp. 1009-1040, Jan. 2018.
[16]S. Pouyanfar, Y.Yang, S. C. Chen, M. L. Shyu, S. S. Iyengar. "Multimedia big data analytics: A survey." *ACM. Comput. Surv.*, vol.51, no.1, Jan. 2018. Art.no. 10.
[17] T. Hou, J. Wang, Y. Liu, X. Sun, A. Li, and B. Ai. "A joint design for STAR-RIS enhanced NOMA-CoMP networks: A simultaneous-signal-enhancement-and-cancellation-based (SSECB) design," *IEEE Trans. Veh. Technol*., vol. 71, no. 1, pp. 1043-1048, Nov. 2021.
[18] A. Garg and O. P. Sahu. "Deep convolutional neural network-based speech signal enhancement using extensive speech features," *Int. J. Comput. Methods.*, vol. 19, no. 8, Dec. 2022, Art.no. 2142005.

[19] N. R. Aljohani, A. Fayoumi, and S. U. Hassan. "A novel focal-loss and class-weight-aware convolutional neural network for the classification of in-text citations," *J. Inf. Sci*., vol. 49, no. 1, pp. 79-92, Mar. 2021

[20] T. Wu, J. Lu, J. Zou, N. Chen, and L. Yang. "Accurate prediction of salmon freshness under temperature fluctuations using the convolutional neural network long short-term memory model," *J. Food. Eng*., vol. 334, Dec. 2022, Art.no. 111171.

[21] D. Ai, F. Mo, Y. Han, and J. Wen. "Automated identification of compressive stress and damage in concrete specimen using convolutional neural network learned electromechanical admittance," *Eng. Struct.*, vol. 259, May.2022, Art.no. 114176.

[22] M. Li, L. Wu, G. Xu, F. Duan, and C. Zhu. "A robust 3D-convolutional neural network-based electroencephalogram decoding model for the intra-individual difference," *Int. J. Neural. Syst*, vol. 32, no. 7, May.2022, Art.no. 2250034.