Hindawi
Scientific Programming
Volume 2022, Article ID 5373531, 11 pages
https://doi.org/10.1155/2022/5373531

Research Article

A Multimodal Retrieval and Ranking Method for Scientific Documents Based on HFS and XLNet

Meichao Yan,[1,2] Yu Wen,[1] Qingxuan Shi,[1,2,3] and Xuedong Tian [1,2,3]

[1]*School of Cyber Security and Computer, Hebei University, Baoding 071002, China*
[2]*Institute of Intelligent Image and Document Information Processing, Hebei University, Baoding 071002, China*
[3]*Hebei Machine Vision Engineering Research Center, Hebei University, Baoding 071002, China*

Correspondence should be addressed to Xuedong Tian; xuedong_tian@126.com

Received 17 September 2021; Revised 12 November 2021; Accepted 29 November 2021; Published 4 January 2022

Academic Editor: Liang Zhao

Copyright © 2022 Meichao Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aiming at the defects of traditional full-text retrieval models in dealing with mathematical expressions, which are special objects different from ordinary texts, a multimodal retrieval and ranking method for scientific documents based on hesitant fuzzy sets (HFS) and XLNet is proposed. This method integrates multimodal information, such as mathematical expression images and context text, as keywords to realize the retrieval of scientific documents. In the image modal, the images of mathematical expressions are recognized, and the hesitancy fuzzy set theory is introduced to calculate the hesitancy fuzzy similarity between mathematical query expressions and the mathematical expressions in candidate scientific documents. Meanwhile, in the text mode, XLNet is used to generate word vectors of the mathematical expression context to obtain the similarity between the query text and the mathematical expression context of the candidate scientific documents. Finally, the multimodal evaluation is integrated, and the hesitation fuzzy set is constructed at the document level to obtain the final scores of the scientific documents and corresponding ranked output. The experimental results show that the recall and precision of this method are 0.774 and 0.663 on the NTCIR dataset, respectively, and the average normalized discounted cumulative gain (NDCG) value of the top-10 ranking results is 0.880 on the Chinese scientific document (CSD) dataset.

1. Introduction

Scientific literature retrieval and ranking is an important way for workers to obtain scientific and technological information. As an important part of scientific documents, mathematical expressions and contextual texts with mathematical semantics are the primary basis for scientific document retrieval and ranking. However, the traditional full-text retrieval model for one-dimensional is not effective when facing the special two-dimensional pattern retrieval of mathematical expressions. At present, research studies on mathematical expression retrieval and ranking have been carried out with some progress, and methods and prototype systems [1–6] with mathematical retrieval functions have been proposed.

In terms of mathematical expression retrieval, WikiMirs3.0 [7] constructed a hybrid index composed of the formulas index and the context index to enable more comprehensive use of mathematical information. In addition, the importance of formulas in the document is calculated for distinguishment. Zhang and Youssef [8] proposed a multidimensional similarity index based on a vector model to determine and evaluate five factors: system distance, data type level, matching depth, query coverage, and whether it is a formula. According to these five factors, the similarity between the query expression and the matching expression parsed by MATHML can be calculated.

In the research of mathematical expression retrieval and ranking that fuses mathematical expressions with textual information, MIaS [9] used the LRO (Leave Rightmost Out) method to split the original query generated by the combination of keywords and mathematical expressions into subqueries and merged the results using appropriate weighting to obtain more relevant results to the original

topic. Zai and Tian [10] used FDS [11, 12] to parse the formulas and retrieved relevant documents using obtained operators. The cosine distance between the input word vectors and the keyword vectors in the documents after the word embedding model is calculated to obtain the similarity between the two, which enables a more reasonable and comprehensive retrieval and ranking. The textual information of a mathematical expression is usually contained in the context of the expressions. Kristianto [13] proposed the concept of mathematical expression dependency, using rich semantic information to obtain better accuracy and improve the retrieval results of the mathematical search system.

Multimodality refers to any combination of two or more modalities. Piergiovanni and Ryoo [14] proposed a joint multimodal representation space method, using adversarial formulas for unmatched text and video data to improve the joint embedding space. Frome et al. [15] proposed a deep visual semantic embedding model based on the semantic information in the labeled image data and unlabeled text to identify visual objects. Jin et al. [16] proposed a generalized deep multimodal hashing framework for scalable image-text and video-text retrieval that explored feature representation learning, intermodality similarity preserving, intramodality semantic label preserving, and hash function learning with different types of loss functions simultaneously. Shen et al. [17] proposed a novel unsupervised hashing method (multiview discrete hashing) to learn compact hash codes from multiview data. The proposed method jointly learned the hash codes and cluster labels via factorization techniques and spectral analysis. And they developed an efficient alternating algorithm to optimize the proposed model. The generated hash codes not only could reflect the underlying semantics from multiple views but also enjoy high discrimination. Lu et al. [18] proposed an Online Multimodal Hashing with Dynamic Query-adaption (OMHDQ) method in a novel fashion that was designed to adaptively preserve the multimodal feature information into hash codes. Moreover, the online module was parameter-free. It could avoid time-consuming and inaccurate parameter adjustment in the unsupervised query hashing process.

In the image recognition of mathematical expressions, the mathematical document INFTY system [19] utilized the optical character recognition techniques to analyze the structure of mathematical expressions and recognized printed mathematical expressions into LaTeX and XML markup formats. Deng et al. [20] explored an image-text generation technology, applied them to mathematical expression recognition, used a convolutional neural network (CNN) to extract image features, and employed a recurrent neural network (RNN) for encoding and decoding.

The abovementioned research on the recognition and retrieval of mathematical expressions has achieved certain results. However, the single-modal retrieval model has great limitations because mathematical expressions in scientific documents often exist in multiple forms, such as embedding descriptions and images. Based on this, this study proposes a multimodal retrieval method for scientific documents based on HFS [21, 22] and XLNet [23]. This method integrates the functions of mathematical expression images and contextual text to improve the accuracy of retrieval results. In this study, the input form of mathematical expressions is no longer limited, and the information of mathematical expressions in images and text format can be input, which increases the flexibility and practicability of retrieval. In addition, the context of mathematical expression is closely related to the mathematical expression itself in scientific documents, and the combination between mathematical expression and context makes the retrieval and sorting of scientific documents more reasonable.

The contributions of this study can be summarized as follows:

(1) Multimodal retrieval is introduced into the retrieval task of scientific documents, and the complementarity between image mode and text mode is utilized to retrieve scientific documents.

(2) Mathematical expressions and their context are combined to retrieval and ranking, and XLNet is used to generate word vector, so that a richer semantic representation of mathematical expression context can be obtained.

(3) The hesitancy fuzzy set is used to calculate the hesitancy fuzzy measure of scientific documents. The hesitancy fuzzy set considers the attributes of the documents. In addition, Chinese scientific documents (CSD) were added to the retrieved dataset.

2. Model Framework

The multimodal retrieval and ranking process of scientific documents based on HFS and XLNet is shown in Figure 1.

First, in the query module, mathematical expression images and text keywords are inputted.

The processing module of the image model is used to calculate the similarity between mathematical expression in images and in candidate technical documents. The LaTeX forms of the input mathematical expressions are obtained by recognizing the images of the input mathematical expression, and FDS is used to analyze the recognition result. Then, the hesitant fuzzy set theory is introduced to calculate the similarity between the mathematical expressions and the results are returned to the document processing module.

The processing module of text modal is used to calculate the similarity between the mathematical expression context. The text in the context of mathematical expressions in the dataset is extracted and used to pretrain XLNet. XLNet is used to calculate the similarity between the query text and the mathematical expression context of the candidate scientific documents.

The document processing module is used to output documents in order. The document attributes are designed, the scores of the documents are calculated by hesitation fuzzy set, and the ranking results are output in descending order of similarity.

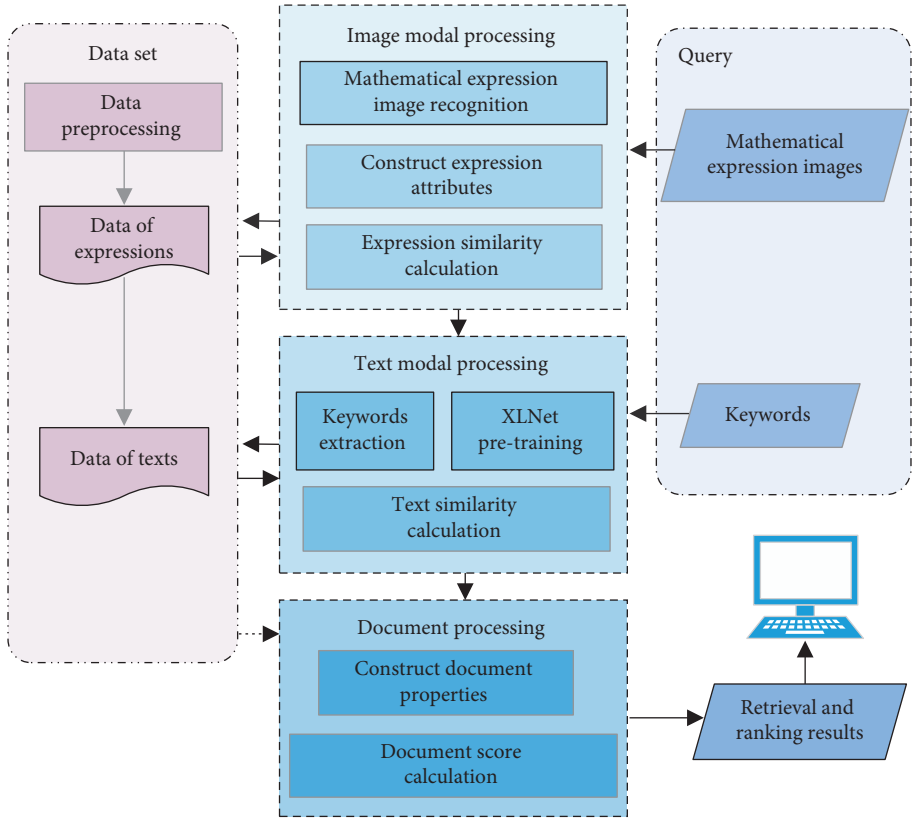

Figure 1: The flowchart of the multimodal retrieval and ranking for scientific documents.

3. Similarity Measure of Multimodal Mathematical Expressions

3.1. Mathematical Expression Image Model's Similarity Measure

3.1.1. Mathematical Expression Image Recognition. The ViT and transformer models proposed in the literature [24–26] for processing sequence problems and image tasks are shown in Figure 2.

The model consists of a ViT [24] encoder with a deep residual network (ResNet) [25] backbone and a Transformer [26] decoder. The encoder is used for feature extraction, and the decoder is used to convert the mathematical expression information in the image into the LaTeX form. The experimental results show that the accuracy of Bilingual Evaluation Understudy (BLEU) is 0.88.

3.1.2. Mathematical Expression Image Similarity. The hesitant fuzzy set proposed by Torra [21, 22] is used to measure the similarity between query expressions and candidate expressions. The value of membership in the hesitant fuzzy set is a value set containing several possible membership degrees. Therefore, the results can be evaluated from multiple aspects. This approach avoids the errors due to a single phenomenon. The degree of hesitation of people in the process of transaction processing can be more objectively reflected.

Definition 1 (hesitating fuzzy set). Let X be a nonempty set, and the definition of the hesitation fuzzy set is

$$E = \{ < x, h_E(x) > | x \in X \}, \tag{1}$$

where $h_E(x)$ represents the set of possible membership degrees for $x \in X$, which is a subset of the interval [0, 1] [21, 22]. Among them, $h_E(x)$ means evaluation attributes, which may be one or more. Each group of evaluation attributes contains multiple evaluation indicators.

The similarity of the analytical mathematical expression of FDS [11, 12] is calculated by the hesitant fuzzy set. The evaluation attribute of the mathematical expression is defined as a triple (h_S, h_O, h_N) [27], where h_S is the structural attribute of the expression, h_O is the operator attribute of the expression, and h_N is the operand attribute of the expression. The structure and operator characteristics of the expression are evaluated, respectively. Each evaluation attribute contains several evaluation indicators. By setting the membership function for each indicator, the query expression E_q and the hesitant membership degree of each result expression E_D for each attribute are evaluated.

In conclusion, the set of hesitating fuzzy evaluation attributes (h_S, h_O, h_N) and the set of hesitating fuzzy elements $h_{E(x)} = \{u_{h_S}, u_{h_O}, u_{h_N}\}$ are constructed based on the above attributes. $u_{h_S}, u_{h_O}, u_{h_N}$ are the corresponding hesitant fuzzy membership functions of each evaluation attribute.

(1) Structural Attribute h_S

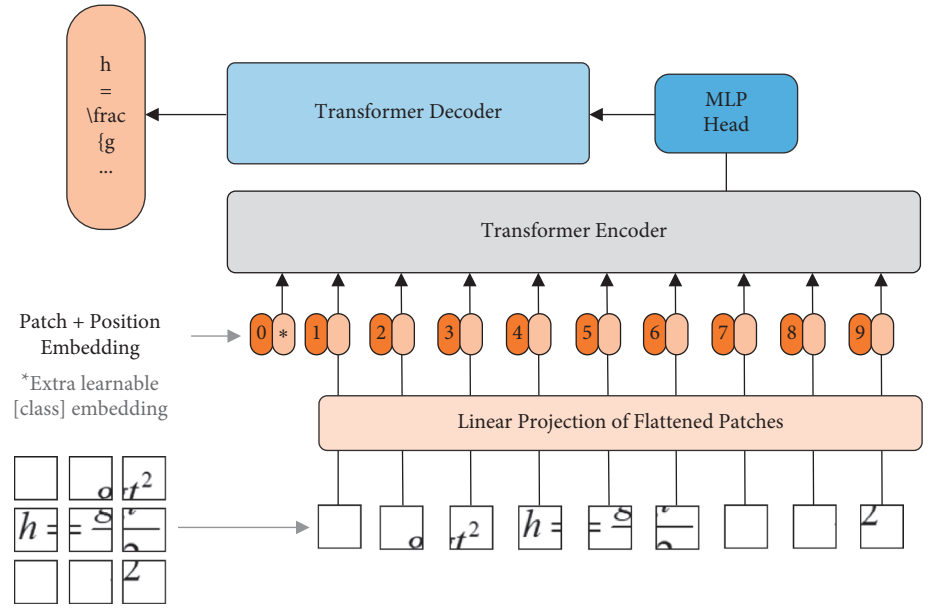

Figure 2: Image recognition of mathematical expressions.

Definition 2 The subformula weight distribution method [28] in the traditional tree index structure is referred, and the flag, length, and operator level in the subexpression are used to replace the structural complexity, length, and depth of nodes in the traditional method.

$$u_{h_S} = \frac{\sqrt{u_l \cdot u_f}}{\sqrt[3]{u_{\text{level}}}}, \tag{2}$$

where

$$\begin{aligned} u_l\left(E_q, E_D\right) &= \frac{l_{E_q}}{l_{E_D}}, \\ u_f\left(E_q, E_D\right) &= \frac{f_{E_q}}{f_{E_D}}, \\ u_{\text{level}}\left(E_q, E_D\right) &= \frac{1}{n}\left(\text{level}_1 + \text{level}_2 + \cdots + \text{level}_n\right). \end{aligned} \tag{3}$$

Here, f_{E_q} is the lowest form of the flag bit of the current subexpression, f_{E_D} is the flag of the subexpression in the expression, l_{E_q} is the length of the subexpression, l_{E_D} is the length of the entire expression, and level is the level of operators in the subexpression. When the subexpression appears several times in the query results, the average is taken as its level attribute value.

(2) Operator Attribute h_O. Here, the BM25 algorithm is referenced as the membership function of the operator index:

$$u_{h_O} = \sum_{i \in \Theta} \log \frac{N}{N_{E_m}} \frac{(k1+1) f_{E_{qi}}}{k + f_{E_{qi}}} \frac{(k2+1) fq}{k2 + fq}. \tag{4}$$

The formula can be disassembled into three components. The first component N represents the total number of expressions in the database, N_{E_m} represents the total number of expressions, which contains E_q. The second component is the weight of the query word in the database, where f_{E_q} represents the frequency of the operator in the database, and $k1$ and k are empirical parameters. The third component is the weight of the query operator itself, where fq represents the word frequency of the query operators in the user's queries, which is usually set to 1 for shorter queries. $k2$ is an empirical parameter.

The evaluation of operand attribute h_N is similar to the operator attribute h_O, so the description will not be repeated.

(3) Similarity Calculation

$$\text{sim}\left(E_q, E_{mi}\right) = 1 - \left[\frac{1}{3}\sum\left(\frac{1}{l_{x_i}}\sum_{j=1}^{l_{x_i}} \left|h_M^{\sigma(j)}(x_i) - h_N^{\sigma(j)}(x_i)\right|^{\lambda}\right)\right]^{1/\lambda}, \tag{5}$$

where l_{x_i} is the number of evaluation values and $h_M^{\sigma(j)}(x_i)$ and $h_N^{\sigma(j)}(x_i)$ represent the j-th element in $h_M(x)$ and $h_N(x)$, respectively.

Let E_q be $a^2 + b^2$, and some of the retrieval results and the corresponding hesitant fuzzy sets are shown in Table 1.

Definition 3. Let the set of mathematical expressions corresponding to the document set $D = \{D_n | N \in R\}_{n=1}^{N}$ be $E = \{E_m | M \in R\}_{m=1}^{M}$.

The mathematical expression similarity calculation algorithm is as follows:

3.2. Mathematical Expression Context Similarity Measure. XLNet [23] is a generalized autoregressive pretraining model. The text in the documents is extracted, one-third of which is annotated to train XLNet, so that a richer semantic representation of the mathematical expression text can be obtained. The main structure is shown in Figure 3 (assuming the factorization order is $3 \longrightarrow 2 \longrightarrow 4 \longrightarrow 1$).

Table 1: Some of the retrieval results of $a^2 + b^2$ and the corresponding hesitant fuzzy sets.

No.	Expression	$\{u_{h_S}, u_{h_O}, u_{h_N}\}$
1	$a^2 + b^2$	{1.000, 0.174, 0.089}
2	$a^2 + b^2 = c^2$	{0.622, 0.089, 0.061}
3	$\sqrt{a^2 + b^2}$	{0.532, 0.065, 0.061}
4	$c(a^2 + b^2)$	{0.460, 0.065, 0.061}
5	$(a^2 - b^2)/(a^2 + b^2)$	{0.091, 0.051, 0.048}

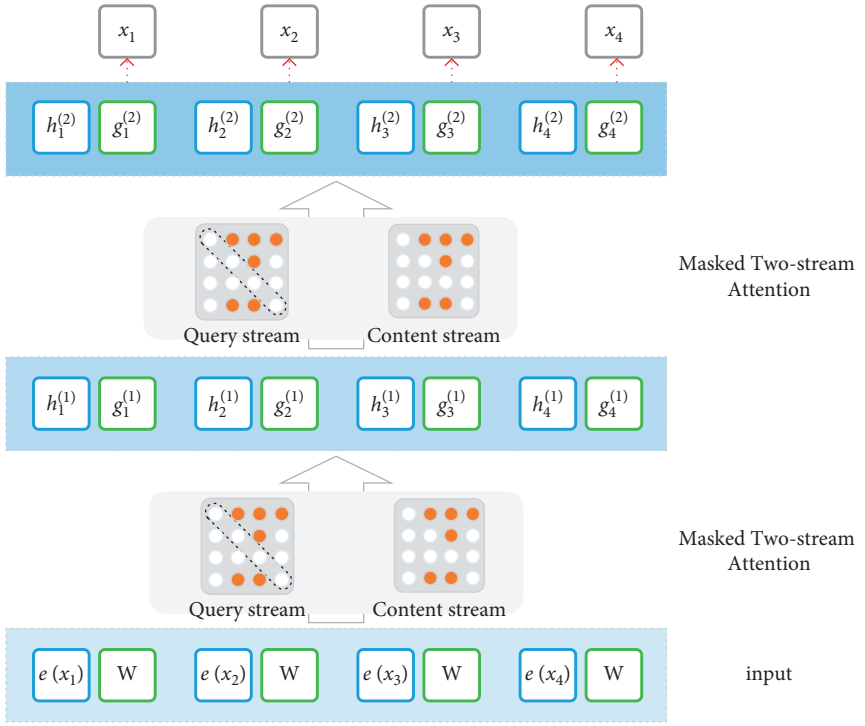

Figure 3: The schematic diagram of the XLNet model.

The same keyword may have different meanings in different contexts, and textual information that explains a mathematical expression often appears around the expression. The example is in the document "Parasitic capacitance.html." The expression in this document is $i = Cdv/dt$, and its contexts are "When two conductors at different potentials are close to one another, they are affected by each others' electric field and store opposite electric charges like a capacitor" and "where C is the capacitance between the conductors." The meanings of "potentials," "electric charges," and "capacitance" may have different meanings in other contexts, and the constructed vectors are also different.

This study introduces the XLNet [23] language model to generate word vectors to be rich in semantics. XLNet solves the problem that BERT did not consider the relationship between the words that are shielded and the words that are not shielded during the training process; that is, the independence between words was not taken into account. The XLNet model implements a new bidirectional coding based on autoregressive (AR) language model. When calculating the text similarity, XLNet will fully consider the semantic information of word vectors, and therefore, the accuracy of calculating text similarity is improved.

The TF-IDF algorithm is used to extract keywords and their weights in the context of mathematical expressions. By analyzing a large number of scientific literature studies, the context of the mathematical expressions is used to analyze mathematical expressions and explain symbols. It can be seen that the context of expressions is closely related to mathematical expressions, so it is very important to extract the context of mathematical expressions for the retrieval of mathematical expressions. The context and keywords corresponding to two mathematical expressions are selected as shown in Table 2.

4. Calculation of the Similarity of Scientific Documents

Retrieval and ranking of scientific and technological documents is a comprehensive measurement with multiple attributes including mathematical expressions and keywords. Different scientific documents have different meanings, even if they contain the same formula. Therefore, hesitating fuzzy sets are used to evaluate scientific documents in an all-around way to achieve the final sorting in this study.

Input: a LaTeX form of the recognized mathematical expression E_q
Output: a set of mathematical expressions similar to E_q
(1) //Initialize the feature vector database $E_D(id, \text{expstring})$
(2) FDS_{E_q} //parsed by FDS
(3) $\text{FDS}_{E_{mi}}$
(4) **while** (FDS_{E_q}) do
(5) **for** FDS_{E_q} in $\text{FDS}_{E_{mi}}$:
(6) $u_{h_{Si}} = \{u_{li}, u_{fi}, u_{\text{level}i}\}$ //Structural attribute membership value
(7) $u_{h_{Oi}}, u_{h_{Ni}}$//Operator and operand attribute membership values
(8) $h_{E_{mi}}(x) = \{u_{h_{Si}}, u_{h_{Oi}}, u_{h_{Ni}}\}$ //Hesitant fuzzy set of E_{mi}
(9) $\text{sim}(E_q, E_{mi}) = \text{sim}(\text{HFS}_{E_q}, \text{HFS}_{E_{mi}})$ //The similarity between expressions E_q and E_{mi} is transformed into the similarity between the hesitant fuzzy set HFS_{E_q} and $\text{HFS}_{E_{mi}}$
(10) Add to table simexp $(\text{id}, \text{expstring}, \text{sim}(E_q, E_{mi}))$
(11) **end for**
(12) **end while**
(13) return simexp
(14) **END**

Algorithm 1: Mathematical expression similarity calculation.

Table 2: The context of two expressions and the corresponding keywords.

No.	Filename	Expression	Preceding paragraphs	Following paragraphs	Keywords
1	Quadratic formula	$ax^2 + bx + c = 0$	The general quadratic equation is	One can verify that the quadratic formula satisfies the quadratic equation by inserting the former into the latter.	Quadratic formula, quadratic equation
2	Gaussian function	$f(x) = ae^{-(x-b)^2/2c^2}$	In mathematics, a Gaussian function, often simply referred to as a Gaussian, is a function of the form:	For arbitrary real constants, and it is named after the mathematician Carl Friedrich Gauss.	Gaussian function, Gaussian

Define the attribute of the scientific document as a five-tuple $(H_{\text{exp}}, H_{\text{word}}, H_{\text{loc}}, H_{ef}, H_{wf})$, where H_{exp} is the similarity attribute of the mathematical expression, H_{word} is the keyword similarity attribute, H_{loc} is the relative position attribute of the expression, H_{ef} is the frequency attribute of the expression, and H_{wf} is the frequency attribute of the keyword. The mathematical expressions and keywords of scientific documents are evaluated.

Definition 4. $W = \{W_{E_m}\}_{m=1}^{M}$ is the keyword set corresponding to $\{E_m\}_{m=1}^{M}$; $\{\text{VEC}_{E_m}\}_{m=1}^{M}$, and $\{\text{VEC}_q\}_{q=1}^{Q}$ are the word vectors corresponding to $\{W_{E_m}\}_{m=1}^{M}$ and the query keyword W_q.

Definition 5. The function $U_{\text{exp}}(E_q, E_{mi})$ is used to calculate the similarity between the query expression and the expression in the candidate document.

$$U_{\text{exp}}(E_q, E_{mi}) = \text{sim}(E_q, E_{mi}), \tag{6}$$

where $\text{sim}(E_q, E_{m_i})$ represents the similarity between the mathematical expression E_q of the query and the mathematical expression E_{mi} in the candidate scientific document.

Definition 6. The function $U_{\text{word}}(W_q, W_{E_{mi}})$ is used to express the similarity between the query keyword W_q and the keyword $W_{E_{m_i}}$ in the context.

$$U_{\text{word}}(W_q, W_{E_{mi}}) = \text{sim}(\text{VEC}_q, \text{VEC}_{E_{mi}}), \tag{7}$$

where $W_{\text{E}_{m_i}}$ represents the keyword in the document retrieved in the candidate scientific document.

Definition 7. The function $U_{\text{loc}}(E_{mi}, D_{N_i})$ is used to express the position of the expression E_{mi} in the document D_{N_i}.

$$U_{\text{loc}}(E_{mi}, D_{N_i}) = 1 - \frac{\text{loc}_{\text{exp}}}{\text{num}}, \tag{8}$$

where loc_{exp} is the position where the query expression E_q appears for the first time in the document D_{N_i}, and num represents the total number of characters contained in the document D_{N_i}.

Definition 8. The function $U_{ef}(E_q, D_{N_i})$ is used to express the frequency of the query expression E_q in the document D_{N_i}.

$$U_{ef}(E_q, D_{N_i}) = 1 - e^{-\alpha(k_{\text{exp}}/k_{\text{sum}_e})}, \tag{9}$$

where α is the feature weight coefficient of the number of mathematical expressions in the document, which is obtained by counting the number of expressions in all documents in the database. $k_{\exp}$ represents the number of expressions in the document D_{N_i} that matches the query expression E_q, and k_{sum} represents the total number of expressions contained in the document D_{N_i}.

Definition 9. The function $U_{wf}(W_q, D_{N_i})$ is used to express the frequency of the query keyword W_q in the document D_{N_i}.

$$U_{wf}\left(W_q, D_{N_i}\right) = 1 - e^{-\alpha\left(t_{W_q}/t_{\text{sum}_w}\right)}, \tag{10}$$

where α is the feature weight coefficient of the number of keywords in the document, which is obtained by counting the number of keywords in all documents in the database. t_{W_q} represents the number of keywords in the document D_{N_i} that matches the query keyword W_q, and t_{sum_w} represents the total number of keywords contained in the document D_{N_i}.

Definition 10. The function $S(D_{N_i})$ is used to calculate the score of scientific document retrieval results.

$$S\left(D_{N_i}\right) = 1 - \left[\frac{1}{5}\sum\left(\frac{1}{l_{h_i}}\sum_{j=1}^{l_{h_i}}\left|H_{EW_q}^{\sigma(j)}(h_i) - H_{D_{Ni}}^{\sigma(j)}(h_i)\right|^{\lambda}\right)\right]^{1/\lambda}, \tag{11}$$

where $S(D_{N_i})$ is the scoring function of the result document D_{N_i} when querying the input expressions and keywords, and $h_i (i = \exp, \text{word}, \text{loc}, ef, wf)$ is the five evaluation attributes of the document. $H_{EW_q}^{\sigma(j)}(h_i)$ and $H_{D_{N_i}}^{\sigma(j)}(h_i)$ are the j-th largest elements in $H_{EW_q}(h_i)$ and $H_{D_{N_i}}(h_i)$, respectively. l_{h_i} is the number of evaluation values included in the evaluation attribute h_i. The attributes of the document are shown in Table 3.

The sorting algorithm of retrieval result documents is as follows:

5. Experimental Process and Result Analysis

5.1. Experimental Data. For the image recognition part of mathematical expressions, we use the IM2LATEX-100K dataset for training and testing. The IM2LATEX-100K dataset contains 103,556 images of different mathematical expressions. The label data consist of the LaTeX format of mathematical expressions.

For the scientific document retrieval and ranking part, the public dataset Ntcir-MathIR-Wikipedia-Corpus (NTCIR) is used, and 31,742 documents are extracted, which contains 518,929 mathematical expressions. In addition, Chinese scientific documents (CSD) are added to expand the dataset, which contains 10,372 documents and 121,495 mathematical expressions.

5.2. System Experiments

5.2.1. Image Recognition of Mathematical Expressions. The image recognition algorithm model [24–26] is used to recognize mathematical expression images and conducts a lot of experiments on different types of mathematical expression images in this study. According to the BLEU evaluation standard, the model result reaches 0.88.

For this recognition algorithm, five different types of mathematical expression images are selected for recognition and display in this study, and the recognition results are shown in Table 4 (the content of the image here is expressed in text).

5.2.2. Ablation Study. Ten groups of formulas and keywords are selected in Table 5 as queries for retrieval. The proposed method includes three main parts, and the performance is continuously improved by gradually increasing the functions of each part. The baseline experiment was image expression retrieval. The final reordering of our has the best performance. The average recall rates of this study are 77.4% and 77.8%. And the average precision rates are 66.3% and 69.2%. All of them are shown in Table 6.

5.2.3. Performance on NTCIR Dataset. In this section, the method in this article is compared with some traditional methods and current existing methods using the NTCIR dataset. FDS + Word Embedding [10] combines the FDS and Word Embedding to retrieve scientific documents: FDS is used to parse expressions, and Word Embedding is used to generate the word vectors of keywords in scientific documents, hereinafter referred to as Method 1. And SearchOnMath [29] is a mathematical formula retrieval tool that aims at accurately matching mathematical expressions, However, SearchOnMath implements pure mathematical expression retrieval and does not consider the important information of the scientific document itself, hereinafter referred to as Method 2. MIaS [4] is based on the full-text search engine Apache Lucene. MIaS processes text and math separately. The text is tokenized and stemmed to unify inflected word forms, hereinafter referred to as Method 3.

In this study, NDCG is used to evaluate the ranking results, which is the search result after the normalization of DCG (discount cumulative gain). The calculation method is as follows:

$$n\text{DCG}_l = \frac{\text{DCG}_l}{\text{IDCG}_l}, \tag{12}$$

where

$$\begin{aligned}\text{DCG}_l &= \sum_{i=1}^{l}\frac{r_i}{\log_2(i+1)},\\ \text{IDCG}_l &= \sum_{i=1}^{|\text{REL}|}\frac{r_i}{\log_2(i+1)},\end{aligned} \tag{13}$$

where l is the number of search results, r_i is the relevance score, IDCG_l is the ideal DCG value, and |REL| indicates that the search results are all related to the query expression.

Input: document collection of the retrieval results of scientific documents
Output: the ranking sequence of the documents
(1) **while** (Result) do:
(2) $U_{\exp}(E_q, E_{m_i}) = \mathrm{sim}(E_q, E_{m_i})$ //Mathematical expression similarity degree of membership value
(3) $U_{\mathrm{word}}(W_q, W_{E_{mi}}) = \mathrm{sim}(\mathrm{VEC}_q, \mathrm{VEC}_{E_{mi}})$
(4) location (E_{m_i}, D_{N_i}) //The position of the mathematical expression in the document
(5) $U_{\mathrm{loc}}(E_{m_i}, D_{N_i})$
(6) $U_{ef}(E_q, D_{N_i})$, $U_{wf}(W_q, D_{N_i})$
(7) $(H_{\exp}, H_{\mathrm{word}}, H_{\mathrm{loc}}, H_{ef}, H_{wf})$ //hesitant fuzzy set of E_{Di}
(8) simText = $\mathrm{sim}((H_{\exp}, H_{\mathrm{word}}, H_{\mathrm{loc}}, H_{ef}, H_{wf}), (1,1,1,1,1))$ //The similarity between documents is transformed into the similarity between the hesitant fuzzy set $(H_{\exp}, H_{\mathrm{word}}, H_{\mathrm{loc}}, H_{ef}, H_{wf})$
(9) return simText
(10) **end while**
(11) return simText DESC //Sort in descending order, return results
(12) **END**

Algorithm 2: Ranking of the retrieval results of scientific documents.

Table 3: Document attributes.

File	ID of file	Five-tuple
D_{N_1}	1	$(H_{\exp_1}, H_{\mathrm{word}_1}, H_{\mathrm{loc}_1}, H_{ef_1}, H_{wf_1})$
D_{N_2}	2	$(H_{\exp_2}, H_{\mathrm{word}_2}, H_{\mathrm{loc}_2}, H_{ef_2}, H_{wf_2})$
...	...	...
D_{N_n}	n	$(H_{\exp_n}, H_{\mathrm{word}_n}, H_{\mathrm{loc}_n}, H_{ef_n}, H_{wf_n})$

Table 4: Image recognition results of five mathematical expressions.

No.	Images containing the mathematical expression	Recognition result
1	$a + b$	a + b
2	$S = \pi r^2$	\[S = \pi {r^2}\]
3	$x = (-b \pm \sqrt{b^2 - 4ac})/2a$	x = \frac{{ - b \pm \sqrt {{b^2} - 4ac} }}{{2a}}
4	$\cos\alpha + \cos\beta = 2\cos(1/2)(\alpha+\beta)\cos(1/2)(\alpha-\beta)$	\cos \alpha + \cos \beta = 2\cos \frac{1}{2}\left({\alpha+\beta }\right)\cos\frac{1}{2}\left({\alpha - \beta } \right)
5	$\sqrt{ab} \le (a+b)/2$	\sqrt {ab} \le \frac{{a + b}}{2}

Table 5: The list of documents.

No.	Exp	Keyword
1	$f(x) = a^x$	Exponential function (指数函数)
2	$S = \pi r^2$	Circular area (圆形面积)
3	$x = (-b \pm \sqrt{b^2 - 4ac})/2a$	Quadratic formula (求根公式)
4	$\lim_{n\longrightarrow\infty} (1 + 1/n)^n$	Limit theorem (极限)
5	$\sqrt{ab} \le (a+b)/2$	Inequality (不等式)
6	$\log_a x$	Logarithm (对数)
7	$a^2 - b^2$	Square variance (方差)
8	$f'(x)$	Differential coefficient (微分)
9	$\tan\theta = \sin\theta/\cos\theta$	Trigonometric function (三角函数)
10	$P(X = k) = \lambda^k/k!e^\lambda$	Poisson (泊松)

The query formula and keywords in Table 5 are taken as the query, and the method in this study and other methods top-10 experts ranking results are shown in Figure 4. Method 2 starts with a higher value than the method in this article, but as the number of expression retrievals increases, the method in this article is all higher than Method 2. The average NDCG of this method is higher than the other three methods. And the average value of NDCG ($n = 10$) is 0.865 on the NTCIR dataset in this study. The experimental results show that the ranking performance of the proposed method is better and the retrieval result is more reasonable.

5.2.4. Performance on CSD Dataset. In this section, the method in this article is compared with Method 1 using the NTCIR dataset. Chinese scientific documents (CSD) are

Table 6: Results of ablation assessment on NTCIR and CSD data.

	Recall		Precision		F1	
Dataset	NTCIR	CSD	NTCIR	CSD	NTCIR	CSD
Baseline	0.753	0.751	0.629	0.635	0.685	0.688
Baseline + XLNet	0.761	0.772	0.636	0.649	0.693	0.705
Our	**0.774**	**0.778**	**0.663**	**0.692**	**0.714**	**0.732**

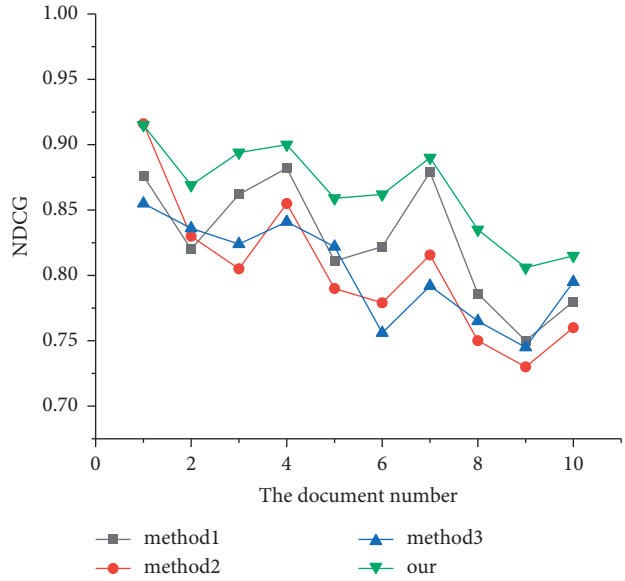

Figure 4: Comparison of the NDCG between the proposed method and other methods.

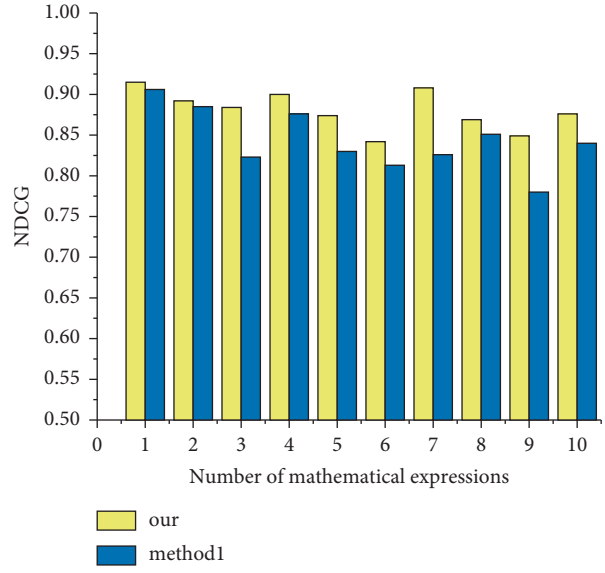

Figure 5: Comparison between the proposed method and Method 1.

added to expand the dataset, which contains 10,372 documents and 121,495 mathematical expressions. The experimental results are shown in Figure 5.

It can be seen that the NDCG of the method in this study is higher than the comparison method. The average value of NDCG ($n = 10$) is 0.88 on the CSD dataset in this study. So the results of the method in this study are more reasonable, and the retrieval and ranking performance is improved.

5.2.5. Retrieval System. A large number of experiments are conducted for different expressions. The first ten search results are selected for display in this study. When the

Table 7: Partial retrieval and ranking results of scientific documents based on HFS and XLNet.

No.	FileName	Score
1	Poisson distribution	0.92
2	Variance	0.88
3	Poisson games	0.83
4	Zero-truncated Poisson distribution	0.79
5	Geometric Poisson distribution	0.76
6	Displaced Poisson distribution	0.73
7	Poisson process	0.72
8	Poisson limit theorem	0.70
9	Inhomogeneous Poisson process	0.68
10	Poisson point process	0.67

input formula image is "$P(X = k) = \lambda^k/k!e^\lambda$" and the keyword is "Poisson," some of the search results are shown in Table 7.

First of all, the method in this study identifies the LaTeX form of the formula as "P\left({X = k} \right) = \frac {{{\lambda^k}}}{{k!}}{e^{-\lambda }}," finding out a collection of documents similar to the formula, and the XLNet model is used to obtain the word vector of "Poisson" and document expressions context keywords, and the similarity between them is calculated. Finally, according to the keywords and formula information, the similarity calculation of the documents is performed again using the hesitant fuzzy set so as to sort and output. FileName is the name of the document where the expression is located, and Score is the document score in Table 7.

6. Conclusion

Based on the retrieval and ranking mode of combining mathematical expression image and text, this study proposes a multimodal retrieval and ranking method for scientific documents based on HFS and XLNet. This method obtains the LaTeX structure information of mathematical expressions through image recognition algorithms and solves the single-modal problem of scientific document retrieval. The similarity between mathematical expressions is obtained by the evaluation of hesitant fuzzy sets, which solves the problem of the unity of evaluation of traditional mathematical expressions. In combination with the context of mathematical expression, the words with similar query keywords are obtained according to XLNet, which enriches the singleness problem of mathematical expression retrieval. Finally, the similarity between of attributes of mathematical expressions and the keywords in the documents is calculated through the hesitation fuzzy set, which makes the ranking of the retrieval results of scientific documents more reasonable.

This experimental method also has some shortcomings. In the future, the following points will be considered for improvement:

(1) Only the mathematical expressions whose recognition results are in LaTeX form are analyzed, and different forms of mathematical expressions (such as MathML) will be analyzed

(2) The evaluation attributes of documents will be further improved, and the evaluation attributes of document similarity will be increased

(3) Only images and texts are analyzed, and an attempt will be made to expand the multimodality more widely and apply voice or video to retrieval

Data Availability

Our data still need to be studied in the next stage, so it is not convenient to provide it directly. The data can be made available upon request via e-mail to the corresponding author.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work is supported by Hebei Natural Science Foundation, China (No. F2019201329), and the Science and Technology Project of Hebei Education Department (No. QN2018214).

References

[1] M. Schubotz, N. Meuschke, T. Hepp, H. S. Cohl, and B. Gipp, "A visualization tool for mathematical expression trees," in *Proceedings of the International Conference on Intelligent Computer Mathematics*, pp. 340–355, Springer, Edinburgh, UK, June 2017.

[2] K. Yamada and H. Murakami, "Mathematical expression retrieval in PDFs from the Web using mathematical term queries," in *Proceedings of the International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pp. 155–161, Springer, Kitakyushu, Japan, September 2020.

[3] P. Libbrecht and E. Melis, "Methods to access and retrieve mathematical content in activemath," in *Proceedings of the International Congress on Mathematical Software*, pp. 331–342, Springer, Berlin, Heidelberg, Germany, September2006.

[4] P. Sojka, M. Růžička, and V. Novotný, "MIaS: math-aware retrieval in digital mathematical libraries," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 1923–1926, New York, USA, August 2018.

[5] W. Zhong, S. Rohatgi, J. Wu, L. C. Giles, and R. Zanibbi, "Accelerating substructure similarity search for formula retrieval," *Advances in Information Retrieval*, vol. 12035, pp. 714–727, 2020.

[6] W. Guo, W. Su, L. Li, N. An, and L. Cui, "MQL: a mathematical formula query language for mathematical search," in *Proceedings of the 14th IEEE International Conference on Computational Science and Engineering*, pp. 245–250, Dalian, China, August2011.

[7] Y. H. Wang, L. C. Gao, S. M. Wang, Z. Tang, X. Liu, and K. Yuan, "WikiMirs 3.0: a hybrid MIR system based on the context, structure and importance of formulae in a document," in *Proceedings of the 15th ACM/IEEE-CS joint conference on digital libraries*, pp. 173–182, New York, USA, June2015.

[8] Q. Zhang and A. Youssef, "An approach to math-similarity search," in *Proceedings of the International Conference on Intelligent Computer Mathematics*, pp. 404–418, Springer, Coimbra, Portugal, July2014.

[9] M. Líška, P. Sojka, and M. Růžička, "Combining text and formula queries in math information retrieval: evaluation of query results merging strategies," in *Proceedings of the First International Workshop on Novel Web Search Interfaces and Systems*, pp. 7–9, New York, USA, October 2015.

[10] X. Y. Zai and X. D. Tian, "Retrieving scientific documents with formula description structure and word embedding," *Data Analysis and Knowledge Discovery*, vol. 4, pp. 131–138, 2020.

[11] X. D. Tian, S. Q. Yang, X. F. Li, and F. Yang, "An indexing method of mathematical expression retrieval," in *Proceedings of the 2013 3rd International Conference on Computer Science and Network Technology*, pp. 574–578, Dalian, China, October 2013.

[12] S. Q. Yang and X. D. Tian, "A maintenance algorithm of FDS based mathematical expression index," in *Proceedings of the 2014 International Conference on Machine Learning and Cybernetics*, vol. 2, pp. 888–892, Lanzhou, China, July 2014.

[13] G. Y. Kristianto, G. Topić, and A. Aizawa, "Utilizing dependency relationships between math expressions in math IR," *Information Retrieval Journal*, vol. 20, no. 2, pp. 132–167, 2017.

[14] A. J. Piergiovanni and M. Ryoo, "Learning multimodal representations for unseen activities," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 517–526, Colorado, USA, March 2020.

[15] A. Frome, G. Corrado, J. Shlens et al., "Devise: a deep visual-semantic embedding model," in *Proceedings of the NIPS*, pp. 2121–2129, NV, USA, January2013.

[16] L. Jin, Z. H. Li, and J. H. Tang, "Deep Semantic Multimodal Hashing Network for Scalable Image-Text and Video-Text Retrievals," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1–14, 2020.

[17] X. Shen, F. Shen, L. Liu, Y. H. Yuan, W. Liu, and Q. S. Sun, "Multiview Discrete hashing for scalable multimedia search," *ACM Transactions on Intelligent Systems and Technology*, vol. 9, no. 5, pp. 1–21, 2018.

[18] X. Lu, L. Zhu, Z. Y. Cheng, L. Q. Nie, and H. X. Zhang, "Online multi-modal hashing with dynamic query-adaption," in *Proceedings of the 42nd International ACM SIGIR Conference On Research And Development In Information Retrieval*, pp. 715–724, N Y, U S.A, April 2019.

[19] M. Suzuki, F. Tamari, R. Fukuda, and S. Uchida, T. Kanahori, "Infty: an integrated ocr system for mathematical documents," in *Proceedings of the 2003 ACM Symposium on Document Engineering*, pp. 95–104, NY, USA, January 2003.

[20] Y. T. Deng, A. Kanervisto, J. Ling, and A. M Rush, "Image-to-markup generation with coarse-to-fine attention," in *Proceedings of the International Conference On Machine Learning. PMLR*, pp. 980–989, Sydney, Australia, August 2017.

[21] V. Torra, "Hesitant fuzzy sets," *International Journal Of Intelligent Systems*, vol. 25, no. 6, pp. 529–539, 2010.

[22] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 1, pp. 338–353, 1965.

[23] Z. L. Yang, Z. H. Dai, Y. M. Yang, J. Carbonell, R Salakhutdinov, and Q. V Le, "Xlnet: generalized autoregressive pretraining for language understanding," in *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, vol. 32, pp. 5754–5764, Vancouver, Canada, 2019.

[24] A. Dosovitskiy, L. Beyer, A. Kolesnikov et al., "An image is worth 16x16 words: transformers for image recognition at scale," 2020, https://arxiv.org/abs/2010.11929.

[25] K. M. He, X. Y. Zhang, S. Q. Ren, and J Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern recognition (CVPR)*, pp. 770–778, Las Vegas, NV,USA, June 2016.

[26] A. Vaswani, N. Shazeer, N. Parmar et al., "Attention is all you need," in *Proceedings of the Advances In Neural Information Processing Systems*, pp. 5998–6008, Long Beach, USA, December, 2017.

[27] X. D. Tian, K. G. Zhang, N. Zhou et al., "A relevance ranking algorithm of mathematical expression retrieval results," *Computer Engineering*, vol. 43, no. 3, pp. 204–212, 2017.

[28] Y. X. Xu, W. Su, M. Cheng, Z. Y. Qu, and H Li, "N-gram index Structure Study for Semantic Based Mathematical Formula," in *Proceedings of the 2014 Tenth International Conference On Computational Intelligence And Security*, pp. 293–298, IEEE, Kunming, China, November 2014.

[29] R. M. Oliveira, F. B. Gonzaga, V. C. Barbosa, and G. B. Xexéo, "A distributed system for SearchOnMath based on the Microsoft BizSpark program," in *Proceedings of the 33rd Brazilian Symposium On Databases*, pp. 289–294, Uberlândia, Brazilian, November 2017.