2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Scientific Literature Retrieval and Recommendation Model Based on RoBERTa and SASRec

Yuhui Zhang[1][✉]

[1] School of Cyberspace Security and Computer Science, Hebei University,
Baoding 071002, China,
z1129264936@163.com

Abstract. With the rapid growth in the quantity and variety of scientific literature, efficiently retrieving and recommending relevant documents for researchers has become a challenge. This paper proposes a scientific literature retrieval and recommendation model integrating Robustly Optimized BERT Pretraining Approach (RoBERTa) and Self-Attentive Sequential Recommendation (SASRec). By incorporating semantic feature information extracted by the RoBERTa model and domain category information predicted by the FastBERT model and combining traditional self-attention sequence recommendation models with proxy attention mechanisms and learnable filtering encoders, the model effectively captures the long-term dependencies of user behavior. This enhances the accuracy of scientific literature retrieval and recommendation. Experimental results demonstrate that the proposed model outperforms traditional methods regarding retrieval and recommendation accuracy, personalization, and efficiency.

Keywords: Literature Retrieval and Recommendation; RoBERTa; FastBERT; SASRec; Sequential Recommendation

1 Introduction

Scientific literature is a core medium for technological advancement and academic communication, carrying important research findings and technological innovations. Most existing scientific literature retrieval systems rely on text or keyword matching, often overlooking the significance of mathematical expressions and their contextual semantics. In many disciplines, mathematical expressions play a crucial role in scientific literature—representing complex theories and models and driving technological innovation and disciplinary development. Therefore, if the importance of mathematical expressions [15,23] is overlooked during retrieval, the accuracy of the search may fails to meet researchers' needs. Moreover, mainstream scientific literature recommendation methods primarily rely on collaborative filtering and content-based [16] approaches. Collaborative filtering methods often depend on the similarity between user behaviors, neglecting researchers' personalized research interests and directions. While content-based recommendation methods leverage the content of scientific literature, they typically lack a deep understanding of the underlying semantics and fail to capture the dynamic evolution of user behaviors and needs, overlooking the temporal changes in user interests. Researchers' reading behaviors and research interests continuously evolve as time progresses, and their research directions may shift. Therefore, an effective recommendation system must adapt to these dynamic changes to provide more accurate and personalized literature recommendations.

This paper addresses the aforementioned issues by integrating mathematical expressions with contextual information and incorporating domain-specific features to provide researchers with more accurate scientific literature retrieval and recommendation sequences. The contributions of this paper are as follows:

1. The semantic features extracted by RoBERTa [18] and the domain category information predicted by FastBERT are integrated to improve the accuracy of scientific literature retrieval.

2. This paper introduces Hesitant Fuzzy Sets (HFS) and Formula Description Structure (FDS) in mathematical formula parsing and retrieval to assist the SimCSE method in addressing formula retrieval's ambiguity and accuracy issues, thereby improving retrieval precision and robustness.

3. This paper incorporates traditional self-attention-based sequential recommendation methods [12]. It refines their model structure to better capture both short-term and long-term dependencies, enhancing recommendations' accuracy and efficiency.

2 Related Work

Schubotz et al. [22] proposed the VMEXT model, which directly converts MathML-formatted expressions into visual expression trees, highlighting similar or identical parts between two expressions and computing their similarity. Additionally, they researched formula encoding format conversion and proposed a method that incorporates the textual context of formulas [21] to reduce mathematical format conversion errors. Zhong et al. [30] explored the retrieval of mathematical expression substructures. They proposed a dynamic pruning algorithm based on inverted indexing, representing mathematical expressions as Operator Trees (OPTs) to improve retrieval efficiency. Dadure et al. [3] introduced a BERT-based embedding model, which generates vector representations of expressions and computes cosine similarity to determine the final ranking of mathematical expressions to enhance semantic representation in mathematical expression embeddings.Gao et al. [5] proposed the Simple Contrastive Sentence Embeddings (SimCSE) method, a contrastive learning approach widely used in unsupervised and semi-supervised learning, which learns effective data representations by comparing similarities and differences between data points. Liu et al. [17] introduced FastBERT, a novel acceleration method that integrates adaptive inference and knowledge distillation. Traditional adaptive inference methods operate either at the token level or the patch level, where they either add iterative steps to individual tokens [7] or dynamically adjust the number of layers executed within discrete image regions [4,25]. On the other hand, knowledge distillation transfers knowledge from a large, complex teacher model to a lightweight student model. Several distillation-based models have been developed: PKD-BERT [24] employs an incremental extraction process, allowing the student model to learn generalized knowledge from intermediate layers of the teacher model. TinyBERT [11] adopts a two-stage learning process, involving general-domain pretraining followed by task-specific fine-tuning. DistilBERT [20] further enhances knowledge transfer by introducing a triple loss function, leveraging inductive biases from large-scale models.

Sequential recommendation models aim to integrate user behavior personalization (based on historical activities) with the contextual information from users' most recent actions. Markov Chains (MCs) serve as a classic example, assuming that the next action depends only on the previous one (or a few previous ones), and they have been successfully applied to describe short-range item transitions in recommendation systems. Additionally, a large body of research has explored more advanced neural network architectures, such as memory networks [1,10] and self-attention mechanisms [14], to improve the ability to effectively capture dynamic user preferences. Zhou et al. [31] observed that self-attention-based recommendation models tend to perform worse when recorded sequential data contains noise. To address this issue, they proposed a novel filter-enhanced MLP (multi-layer perceptron) recommendation method, which removes the self-attention components from Transformers and adopts a fully MLP-based stacked block structure.

3 Experimental Methodology

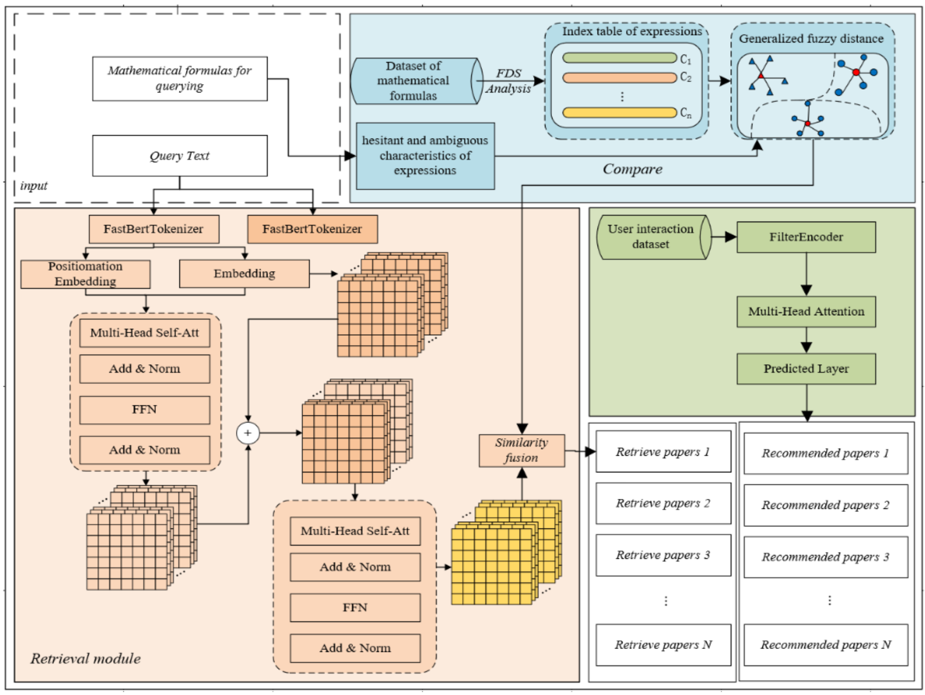

Fig.1. Framework of the Scientific Literature Retrieval and Recommendation Model Based on RoBERTa and SASRec

3.1 Mathematical Expression Matching Module

The structure of the mathematical expression matching module is shown in Figure 1. This module is designed for retrieving mathematical expressions in scientific literature. It consists of key steps such as expression preprocessing, symbol normalization, sub-expression feature extraction, similarity matching, and retrieval result output. The module employs the FDS [26,29] algorithm to parse expressions and define the symbol membership function. Additionally, it utilizes HFS [6,27] to measure the similarity between user input and mathematical expressions in scientific literature.

3.2 Contextual Semantic Similarity

The FastBERT [17] method predicts the domain category of the input text, while SimCSE [5] maps sentences into a high-dimensional semantic space, further improving the accuracy of similarity computation.

Literature Category Prediction Module.

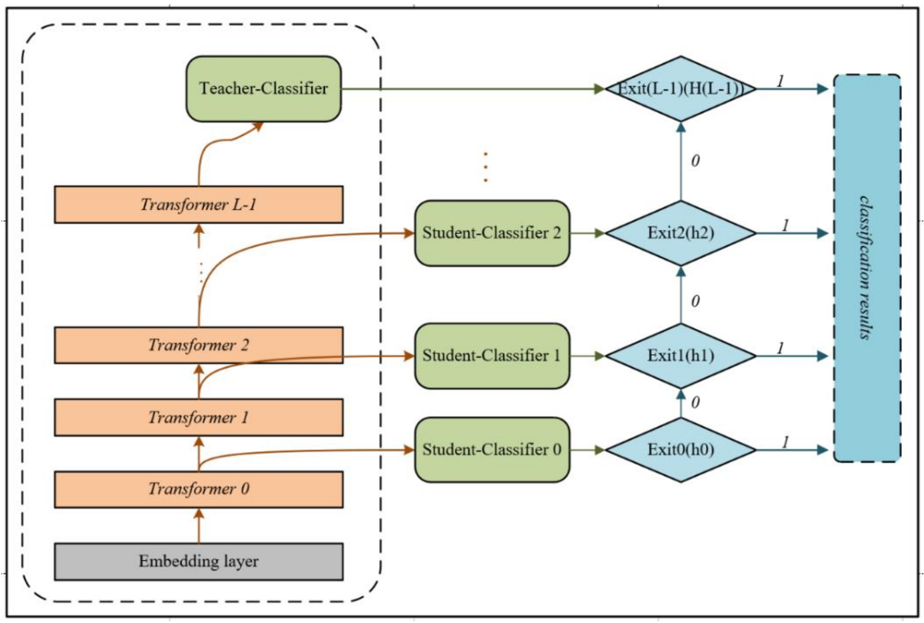

Fig.2. FastBERT Model Diagram

As shown in Figure 2, the FastBERT module is used for document category prediction. Given an input sentence S, it is first tokenized and encoded into a vector sequence $X = [x_1, x_2, \ldots, x_n]$, where n denotes the number of tokens, and $x_i \in R^d$ represents the vector embedding of each token.

$$X = Embedding\,(S) \tag{1}$$

$$hi = \mathrm{Transformer}\,i\,(hi - 1) \tag{2}$$

Here, h_i (i = -1, 0, …, L-1) represents the feature output of the i-th Transformer layer, where h_{-1} = X, and L denotes the total number of Transformer layers.

The model incorporates an adaptive exit mechanism, allowing the binary classifier to dynamically determine whether to exit early at a particular layer and output the prediction result based on the complexity of the input text.

$$Exiti\,(hi) = \begin{cases} 1 \text{ If the certainty is high enough, terminate execution.} \\ 0 \text{ Otherwise, proceed to the next layer.} \end{cases} \tag{3}$$

Since FastBERT has L-1 student classifiers, the loss of the whole model is then composed of the sum of the KL scatter of these L-1 student classifiers:

$$\mathrm{loss}(ps_0, \ldots, ps_{L-2}, pt) = \sum\nolimits_{i\,=\,1}^{L\,-\,2} (\sum\nolimits_{j\,=\,1}^{L} ps(i) \boxtimes log\,\frac{ps(i)}{pt(j)}) \tag{4}$$

Where p_s is the student classifier prediction and p_t is the teacher soft label.

Text Embedding Module.

This paper adopts unsupervised SimCSE as the sentence representation learning method to obtain high-quality sentence vectors for similarity computation. The key aspect of SimCSE is its use of a contrastive learning strategy, which minimizes the distance between different dropout variations of the same input sentence while maximizing the distance between different sentences. This approach enhances the discriminative power of sentence embeddings. Specifically, the same sentence is fed into the model twice, generating two embeddings, Z and Z', with different dropout masks. These two embeddings form a positive pair, while other texts in the same batch serve as "negative pairs." The loss function is then computed using in-batch negative sampling cross-entropy. The training objective for unsupervised SimCSE is as follows:

$$\iota i = -\,log\,\frac{e^{sim(h_i^{z_i}, h_i^{z'_i})/\tau}}{\sum_{j=1}^{N} e^{sim(h_i^{z_i}, h_i^{z'_i})/\tau}} \tag{5}$$

Where i denotes the i-th text within the batch; $\mathrm{sim}(h_1, h_2) = \frac{h_1^T h_2}{\|h_1\| \|h_2\|}$ denotes the cosine similarity of the two vectors, τ is a temperature hyperparameter used to adjust the smoothness of the similarity distribution, N is the number of samples in a mini-batch; h_i represents an independent Dropout sampling.

Feature Fusion.
The category features predicted by FastBERT and the query statements are input into SimCSE together to get the semantic features and category features FS after fusion, and the cosine similarity is used to compute the FS and the data features in the whole semantic space to get the final scientific and technological literature retrieval results.

3.3 Literature Recommendation Module

In this paper, we modify the structure of the traditional SASRec model [12], as shown in Fig. 3. The core idea is to use the Self-Attentive mechanism in the Transformer to capture the user's behaviors and locations in the history sequence, to predict the user's next most likely choice or item of interest.

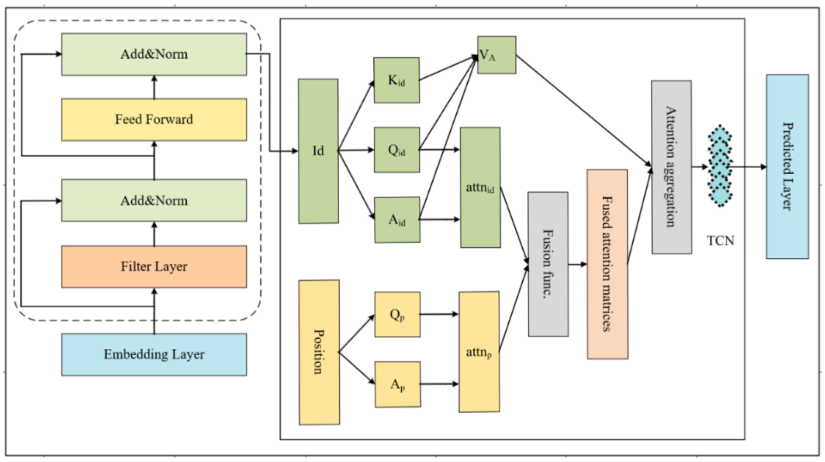

Fig.3. Improved SASRec Module

FilterEncoder.
Inspired by FMLP [31], this paper adopts a lightweight FilterEncoder as the core encoder, replacing the original Transformer encoder in SASRec to efficiently model both local and global features of user behavior sequences.

The Filter Layer module is the fundamental component of FilterEncoder, utilizing frequency domain transformation to perform global modeling of the input sequence features. The input sequence X is mapped to the frequency domain X_{freq} via the Fast Fourier Transform (FFT). X_{freq} is a complex-valued tensor representing the spectrum of X, which is then modulated by multiplying it with the complex-weight parameters $W_{complex} \in C^{1* (\frac{T}{2}+1) *H}$ of a learnable filter:

$$X_{filtered} = X_{freq} \odot W_{complex} \tag{6}$$

$\odot$is element-wise multiplied, and the filter W is referred to as a learnable filter because it can be optimized using Stochastic Gradient Descent (SGD), allowing it to represent arbitrary filters in the frequency domain adaptively.

Finally, $X_{filtered}$ is transformed back to the time domain as X_{time} via the Inverse Fast Fourier Transform (IFFT). To mitigate gradient vanishing and unstable training, residual connections, dropout, and normalization are applied. Additionally, the feedforward layer further captures nonlinear features.

Multi-Head Attention.

Inspired by DIF-SR [28], this paper introduces a positional attention mechanism for positional encoding, addressing the issue of weakened positional relationships in traditional self-attention mechanisms. Meanwhile, the conventional dot-product attention is replaced with Agent Attention [8]. The positional attention generates positional relationship features as follows:

$$Q_p = XW_Q，K_p = XW_k，V_p = XW_V \tag{7}$$

$$V_A = \mathrm{Attn}(A, K, V) = soft\ max(\frac{AK^T}{\sqrt{d_K}})V \tag{8}$$

$$P = \mathrm{Attn}(Q_P, A_P, V_A) \tag{9}$$

Here, $W_{Q\backslash K\backslash V} \in \mathbb{R}^{c*d}$ represents the projection matrices, where C and d denote the channel dimensions of the model and each attention head, respectively. A simple pooling operation is used to obtain the agent vector A, which, together with the filtered user history sequence, is fed into the next stage.

To further model the temporal dependencies in the attention results and to uncover the implicit features in the attention output, a Temporal Convolutional Network (TCN) [9] is added after the attention mechanism output. The dilated convolution property of TCN allows it to capture the variation patterns of the attention results across multiple time scales, thereby extracting both long-term and short-term dependencies and temporal dynamic features.

The prediction layer employs a matrix factorization method to predict the relevance of the next item. A higher interaction score indicates higher relevance; thus, recommendations can be generated by ranking the scores.

The model combines the efficient feature extraction of FilterEncoder, the fine-grained sequence modeling of positional attention and Agent Attention, and the multi-scale temporal feature extraction of TCN from the attention results. This enhances the recommendation accuracy, ultimately forming a hierarchical sequence recommendation architecture.

4 Experiments and Results Analysis

4.1 Experimental Setup

The dataset in this study is based on the publicly available metadata dataset from arXiv [2,13], which contains approximately 1.7 million English scientific papers. To facilitate further research, we selected 12,000 papers from this dataset to construct the experimental dataset. To ensure the completeness of the experiment, we introduced an extended Chinese dataset. We used the ArXiv dataset to simulate a scientific literature recommendation dataset along with the Amazon-beauty dataset [19]. For all datasets, user interaction records were grouped, sorted in ascending order based on timestamps, and filtered to remove unpopular items and inactive users with fewer than five interactions. We selected four commonly used evaluation metrics: MAP (Mean Average Precision), NDCG (Normalized Discounted Cumulative Gain), Recall, and MRR (Mean Reciprocal Rank).

4.2 Ablation Experiment

An ablation study was conducted to analyze the contribution of each key component in the model to the final performance. By gradually removing different modules and observing model performance changes, we verified each module's impact on scientific literature retrieval and recommendation results.

We remove the following modules separately and record their impact on the final results:(1) Removing SimCSE and FastBERT: Without semantic features, the model relies solely on traditional mathematical expression retrieval. (2) Removing FDS: Without using mathematical expression information, the model only relies on traditional textual content for retrieval. (3) Removing the Learnable Filter Encoder: The learnable filter encoder is replaced with a standard fixed filter layer. (4) Removing the Agent Attention Mechanism: The agent attention mechanism is removed to observe its effect on the recommendation performance.

Table 1. Performance Comparison of Different Methods on the Arxiv Dataset and Chinese Scientific Dataset. The best-performing and second-best-performing methods are highlighted in bold and underlined fonts, respectively.

	MAP@5		MAP@10		NDCG@5		NDCG@10	
	En	Ch	En	Ch	En	Ch	En	Ch
Math	0.880	0.887	0.825	0.859	0.846	0.813	0.760	0.836
Text	0.879	0.868	**0.847**	0.833	0.803	0.708	0.782	0.741
Ours	**0.931**	**0.929**	0.832	**0.866**	**0.892**	**0.872**	**0.831**	**0.861**

Table 2. Performance Comparison of Different Methods on the Amazon Dataset. The best-performing and second-best-performing methods are highlighted in bold and underlined fonts, respectively.

Model	Recall@50	Recall@100	NDCG@50	NDCG@100
SASRec	0.1780	0.2368	0.0571	0.0666
SASRec+ FilterEncoder	0.1908	0.2510	0.0607	0.0705
Ours	**0.1943**	**0.2535**	**0.0622**	**0.0718**

The experimental results are shown in Table 1 and Table 2. Removing each component leads to a decline in model performance, validating the importance of these modules in the scientific literature retrieval and recommendation tasks. Specifically, after removing the SimCSE and FastBERT modules, the model's NDCG@5 metric dropped by approximately 5.9%. After removing the FDS module, the NDCG@5 metric decreased by about 17%, indicating that semantic and mathematical expression information contributes to improved retrieval performance. Removing the learnable filter encoder and agent attention modules also led to a performance decline, demonstrating that these components effectively capture long-term user behavior dependencies.

4.3 Comparative Experiment

To evaluate the proposed fusion model's effectiveness, comparative experiments were conducted with the following mainstream methods: (1) The CLFE model learns the potential structure and content information of formulas through contrastive learning methods, and generates formula embeddings for formula retrieval. (2) ColBERT is a post interaction based ranking model that encodes text word vectors and calculates the correlation between query text and documents through post interaction methods. (3) Tangent CFT combines SLT tree and OPT tree, and uses FastText embedding model to convert formulas into vector representations for retrieval. (4) SASRec is a unidirectional sequential recommendation model based on Transformer; (5) FMLP is a sequential recommendation model based on multi-layer perceptron. These methods represent different retrieval and recommendation strategies, covering a range of techniques from shallow to deep models.

Table 3. Performance Comparison of Different Methods on the Arxiv Metadata Dataset. The best-performing and second-best-performing methods are highlighted in bold and underlined fonts, respectively.

	Map@5	Map@10	NDCG@5	NDCG@10
Clfe	0.843	0.812	0.874	**0.859**
ColBert	0.868	0.826	0.869	0.858
Tangent-CFT	0.816	0.762	0.848	0.827
Ours	**0.931**	**0.832**	**0.892**	0.831

Table 4. Performance Comparison of Different Methods on the Amazon Dataset. The best-performing and second-best-performing methods are highlighted in bold and underlined fonts, respectively.

	Recall@50	Recall@100	NDCG @50	NDCG @100	MRR@50	MRR@100
SASRec	0.1780	0.2368	0.0571	0.0666	0.0270	0.0278
FMLP	0.1908	0.2510	0.0607	0.0705	0.0284	0.0293
Ours	**0.1943**	**0.2535**	**0.0622**	**0.0718**	**0.0293**	**0.0302**

The experimental results in Tables 3 and 4 demonstrate that the proposed integrated model outperforms other comparison methods on most evaluation metrics.

4.4 Results Analysis

The integrated model effectively combines the advantages of SimCSE and FastBERT, making full use of the semantic information and domain category features of the literature, thereby improving the accuracy of literature retrieval. In addition, the traditional self-attention sequence recommendation model has been modified, enabling it to better capture users' long-term behavioral dependencies, further enhancing the quality of recommendations.

5 CONCLUSION

This paper proposes a scientific literature retrieval and recommendation model based on RoBERTa and SASRec to address the challenges of insufficient semantic understanding, poor personalized recommendation performance, and difficulties in modeling the temporal evolution of research interests in current literature retrieval and recommendation models.

The model first leverages RoBERTa's powerful semantic representation capabilities to deeply understand the textual content of scientific literature. It integrates the HFS and FDS modules to disambiguate the fuzziness of mathematical formulas effectively. By introducing a proxy attention mechanism, the accuracy of user interest modeling is further enhanced. Additionally, a learnable filtering encoder is employed to capture the dynamic evolution of user interests over time, enabling more precise and personalized literature retrieval and recommendation. Notably, the model demonstrates strong robustness and generalization ability, particularly in handling complex literature structures and dynamically evolving research interests.

6 REFERENCES

1. Chen, X., Xu, H., Zhang, Y., Tang, J., Cao, Y., Qin, Z., Zha, H.: Sequential recommendation with user memory networks. In: Proceedings of the eleventh ACM international conference on web search and data mining. pp. 108–116 (2018)
2. Clement, C.B., Bierbaum, M., O'Keeffe, K.P., Alemi, A.A.: On the use of arXiv as a dataset (2019), https://arxiv.org/abs/1905.00075
3. Dadure, P., Pakray, P., Bandyopadhyay, S.: Bert-based embedding model for formula retrieval. In: CLEF (working notes). pp. 36–46 (2021)
4. Figurnov, M., Collins, M.D., Zhu, Y., Zhang, L., Huang, J., Vetrov, D., Salakhutdinov, R.: Spatially adaptive computation time for residual networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1039–1048 (2017)
5. Gao, T., Yao, X., Chen, D.: Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821 (2021)
6. Goguen, J.A.: La zadeh. fuzzy sets. information and control, vol. 8 (1965), pp. 338–353.-la zadeh. similarity relations and fuzzy orderings. information sciences, vol. 3 (1971), pp. 177–200. The Journal of Symbolic Logic 38(4), 656–657 (1973)
7. Graves, A.: Adaptive computation time for recurrent neural networks. arXiv preprint arXiv:1603.08983 (2016)
8. Han, D., Ye, T., Han, Y., Xia, Z., Pan, S., Wan, P., Song, S., Huang, G.: Agent attention: On the integration of softmax and linear attention. In: European Conference on Computer Vision. pp. 124–140. Springer (2024)
9. Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., Liu, Y.: Temporal convolutional neural (tcn) network for an effective weather forecasting using time-series data from the local weather station. Soft Computing 24, 16453–16482 (2020)
10. Huang, J., Zhao, W.X., Dou, H., Wen, J.R., Chang, E.Y.: Improving sequential recommendation with knowledge-enhanced memory networks. In: The 41st international ACM SIGIR conference on research & development in information retrieval. pp. 505–514 (2018)
11. Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., Liu, Q.: Tinybert: Distilling bert for natural language understanding. arXiv preprint arXiv:1909.10351 (2019)
12. Kang, W.C., McAuley, J.: Self-attentive sequential recommendation. In: 2018 IEEE international conference on data mining (ICDM). pp. 197–206. IEEE (2018)
13. Library, C.U.: arXiv dataset (2020), https://www.kaggle.com/Cornell-University/arxiv
14. Li, J., Wang, Y., McAuley, J.: Time interval aware self-attention for sequential recommendation. In: Proceedings of the 13th international conference on web search and data mining. pp. 322–330 (2020)
15. Li, R., Wang, J., Tian, X.: A multi-modal retrieval model for mathematical expressions based on convnext and hesitant fuzzy set. Electronics 12(20), 4363 (2023)
16. Lixiao, G., Gaojie, J., Yahan, L., Weizhong, S., Shihao, M.: Personalized recommendation research of university library literature resources based on improved content filtering algorithm. Library and Information Service 62(21), 112–117 (2018)
17. Liu, W., Zhou, P., Zhao, Z., Wang, Z., Deng, H., Ju, Q.: Fastbert: a self-distilling bert with adaptive inference time. arXiv preprint arXiv:2004.02178 (2020)
18. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
19. McAuley, J., Targett, C., Shi, Q., Van Den Hengel, A.: Image-based recommendations on styles and substitutes. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. pp. 43–52 (2015)

20. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019)
21. Schubotz, M., Greiner-Petter, A., Scharpf, P., Meuschke, N., Cohl, H.S., Gipp, B.: Improving the representation and conversion of mathematical formulae by considering their textual context. In: Proceedings of the 18th ACM/IEEE on joint conference on digital libraries. pp. 233–242 (2018)
22. Schubotz, M., Meuschke, N., Hepp, T., Cohl, H.S., Gipp, B.: Vmext: A visualization tool for mathematical expression trees. In: Intelligent Computer Mathematics: 10th International Conference, CICM 2017, Edinburgh, UK, July 17-21, 2017, Proceedings 10. pp. 340–355. Springer (2017)
23. Siwen, D., Xinfu, L., Fang, Y., Xuedong, T.: Mathematical expression query expansion method based on dependency parsing. Computer Applications and Software 41(04), 251–255, 290 (2024)
24. Sun, S., Cheng, Y., Gan, Z., Liu, J.: Patient knowledge distillation for bert model compression. arXiv preprint arXiv:1908.09355 (2019)
25. Teerapittayanon, S., McDanel, B., Kung, H.T.: Branchynet: Fast inference via early exiting from deep neural networks. In: 2016 23rd international conference on pattern recognition (ICPR). pp. 2464–2469. IEEE (2016)
26. Tian, X., Yang, S., Li, X., Yang, F.: An indexing method of mathematical expression retrieval. In: Proceedings of 2013 3rd International Conference on Computer Science and Network Technology. pp. 574–578. IEEE (2013)
27. Torra, V.: Hesitant fuzzy sets. International journal of intelligent systems 25(6), 529–539 (2010)
28. Xie, Y., Zhou, P., Kim, S.: Decoupled side information fusion for sequential recommendation. In: Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. pp. 1611–1621 (2022)
29. Yang, S.Q., Tian, X.D.: A maintenance algorithm of fds based mathematical expression index. In: 2014 International Conference on Machine Learning and Cybernetics. vol. 2, pp. 888–892. IEEE (2014)
30. Zhong, W., Rohatgi, S., Wu, J., Giles, C.L., Zanibbi, R.: Accelerating substructure similarity search for formula retrieval. In: Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part I 42. pp. 714–727. Springer (2020)
31. Zhou, K., Yu, H., Zhao, W.X., Wen, J.R.: Filter-enhanced mlp is all you need for sequential recommendation. In: Proceedings of the ACM web conference 2022. pp. 2388–2399 (2022)