SN Computer Science (2025) 6:579
https://doi.org/10.1007/s42979-025-04095-x

ORIGINAL RESEARCH

Retrieval Augmented Generation Model for Paper Recommendation System

Neha Yadav[1] · **Dhanalekshmi Gopinathan**[1]

Received: 11 August 2024 / Accepted: 26 May 2025

Abstract

Academic paper recommendation systems are vital in streamlining research by helping scholars identify relevant literature efficiently. However, traditional approaches often fail to capture profound contextual relevance, especially in multidisciplinary domains. This study proposes a novel Retrieval-Augmented Generation (RAG) model that synergistically combines retrieval and generative mechanisms to address these limitations. The RAG architecture, including data preparation, model training, and integration of neural retrieval and generation components, is described in detail. Experimental results demonstrate that the RAG model significantly outperforms standard content-based filtering, achieving precision of 0.78, recall of 0.72, F1-score of 0.75, and MRR of 0.83. These results validate the model's ability to generate contextually relevant and accurate academic recommendations. This paper identifies future directions to further enhance the model's applicability across disciplines.

Keywords Retrieval augmented generation · Recommendation system · Content-based recommendation · Machine learning

Introduction

Background

With the rapid expansion of academic publications across diverse disciplines, researchers face increasing challenges in efficiently locating relevant, high-quality literature. Traditional manual search methods are time-consuming and often ineffective in surfacing contextually appropriate papers. Paper recommendation systems have emerged as essential tools to support researchers by automating the discovery of relevant literature. Early systems relied on keyword-based or metadata-driven matching, often producing shallow or inaccurate results. Subsequent advancements led to the development of content-based and collaborative filtering systems, which improved relevance but struggled with semantic Understanding, sparsity of user data, and cross-domain generalization. Recently, citation-based networks and machine learning techniques have contributed to nuanced recommendation mechanisms [1]. Nonetheless, a persistent gap remains in understanding contextual relationships within academic texts.

To bridge this gap, our study introduces the RAG model to capture scholarly works' semantic depth and contextual dependencies. Sooner or later, the scholars must access the required publication swiftly and effectively for academic research. The number of published literature sources increases dramatically, usually making scholars feel lost in the sea of information [2]. Most relevant results may not be delivered if traditional labor-intensive methods of literature review, which involve the manual search and filter of publications, are used [3]. For this reason, paper recommendation systems have benefited researchers seeking suitable material among the vastness of scientific literature that matches their interests and specific goals. The evolutionary path that recommendation systems of papers have followed has significantly changed over time [4]. The original approaches used by early algorithms in giving recommendations for papers were mostly on keyword matching and simple data, such as titles and abstracts [5]. These early methods were

✉ Neha Yadav
nehayadav9506@gmail.com

Dhanalekshmi Gopinathan
Dhanalekshmig@gmail.com

[1] Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Noida, India

Published online: 23 June 2025

effective but often provided superficial recommendations and did not consider the entire corpus of relevant research [6]. For instance, publications that contained some specific terms but otherwise had little to do with the user's main research goals were frequently returned by keyword-based algorithms [7]. This defect showed the necessity for more sophisticated techniques to do better than matching texts [8]. More complex models were designed at that time as advancements in technology occurred. In using document content, known as content-based filtering [9], collaborative filtering utilizes the tastes and behaviors of like people, and accuracy is increased for recommendation techniques [10]. Collaborative filtering has the potential to offer you more personalized options by suggesting articles based on what other academics with similar interests have read [11]. Content-based filtering relies on features taken directly from the text of the paper to detect patterns and recommend relevant information through [12, 13], and [14]. Yet, these techniques had their weaknesses. Collaborative filtering needed significant volumes of user interaction data, which were usually not present in academic contexts [15]. It was found that the complexity of academic publications' semantics was preventing content-based filtering from being effective [16] via [17].

To make recommendation systems even more powerful, citation networks were integrated [18]. By examining the citation pattern of articles referred to by other users, these machines can find important work and novel trends in specific areas [19]. Citation-based recommendation systems allowed researchers to follow up on articles' importance and intellectual influence, offering a way to find foundational works that wouldn't have been obvious by a simple keyword search [20]. But even those state-of-the-art devices were not without issue: for example, rather than treating citations as binary relationships, they often ignored the elaborate warrants for citation—reasons for or against citing something, or the context in which it was mentioned [21]. The new opportunities for building recommendation systems are realized because of recent advances in machine learning and NLP [22] that now enable building more comprehensive, yet interpretable, models [23]. For example, full-text-NLP models can process an entire paper to mine recurring patterns or topics to understand their relevance [24]. Machine learning algorithms may thus, with this information of context, be capable of coming up with much more accurate and insightful recommendations [25]. In more esoteric domains, where the kind of relevance invoked is very subtle, even current approaches at times cannot suffice to bring forth the required relevance, which is where there are shortcomings in scientific publications [26, 27]. They may not be aware of the complexity and multifaceted nature of scholarly research, where the prime drivers of relevance are context and depth in information [28].

Motivation

Despite significant advancements in recommendation technologies, conventional models still fall short in accurately addressing complex, interdisciplinary research needs. These systems often fail to recognize subtle contextual links or tend to bias recommendations towards dominant disciplines [29]. Researchers working across domains may navigate through irrelevant suggestions or missing key literature [30]. This inefficiency impedes productivity and may impact research outcomes [31]. Furthermore, as scientific literature surges, adaptive and intelligent systems are needed to keep pace. To this end, the Retrieval-Augmented Generation (RAG) framework offers a promising direction [32]. RAG can produce highly relevant recommendations grounded in semantic similarity and contextual synthesis by integrating retrieval with language generation. This study demonstrates the RAG model's capacity to meet evolving research demands through enhanced accuracy and contextual awareness [33].

Despite developing some of the most advanced recommendation systems, most models cannot associate themselves with complete reliability and suggest scholarly articles with complex needs [34]. This discrepancy is particularly acute due to current research's highly variable and multidisciplinary nature [35]. While useful, traditional recommendation algorithms often lack the intelligence to understand complex linkages across multiple study areas and subtle contextual clues [36]. Such a restriction could lead to the suggestions being too broad or not doing justice to the actual topic of a researcher's investigation. For instance, an interdisciplinary researcher at the interface of biology and computer science may discover that standard systems lose subtle linkages between the two disciplines or are biased for one discipline. For example, the researchers can become inefficient and frustrated when they need to spend a lot of time weeding out hints that almost fit their goals [37, 38]. Moreover, keeping up to date with the most critical and significant recent literature is a continuous challenge, given the pace of scientific discovery [39]. Systems for researchers should adapt to the changes in the interests and the study paths of the researchers, rather than staying afloat by just keeping up with the volume of new publications [40]. Therefore, a more sophisticated approach, amalgamating the best from both retrieval and generating techniques, is required to counter these shortcomings [41]. One such method applied is the Retrieval-Augmented Generation (RAG) model. The RAG model uses retrieval techniques along with generative models for recommendations to enhance their accuracy and relevance [42]. This dual approach does a great job of striking the needed balance between inadequate general search output in satisfying basic research demands of pertinence and more specialized demands to not only give a relevant set of papers but also to develop appropriate recommendations in context [43]. The RAG model uses generative models that combine and give recommendations in a way that significantly aligns with the query of the researcher in question, and at the same time uses retrieval systems to retrieve a much wider variety of articles [44, 45].

Problem Statement

Current academic recommendation systems are often inadequate in addressing modern, interdisciplinary research's complex semantic and contextual requirements. Based on content similarity or collaborative usage patterns, traditional models lack the interpretability and contextual precision to capture subtle yet meaningful relationships between academic works. This shortfall leads to overly generic or contextually irrelevant suggestions, ultimately impeding research efficiency and depth. There is a pressing need for models that retrieve relevant content and understand and generate context-sensitive recommendations. Despite massive advancements in recommendation systems, many current algorithms are not designed in a way that expresses subtle relevance to academic papers. Therefore, the multidisciplinary and interdisciplinarity at work within the scope of modern research may often challenge conventional approaches that might entail content-based and collaborative filtering. In this respect, the results obtained from these algorithms may give recommendations that are too general for the specific area of interest to the researcher or completely deviate from the subject area because they cannot identify contextual complexity and complex interrelations among different study subjects. This limitation reduces productivity and efficiency during research; this calls for more developed models with enhanced Understanding and interpretability of academic information.

Contributions

Therefore, this paper introduces a new Retrieval-Augmented Generation model, which has been specially developed for an academic paper recommendation task to overcome the deficiencies of the existing systems, especially in recommendation systems. The main contributions of the current study include:

a. **Development of a New RAG Model**: In keeping with increasing the precision and value of paper recommendations, the paper develops a new RAG model which fuses retrieval strategies and generative models to suggest relevant recommendations through advanced natural language processing and machine learning approaches.
b. **Rigorous Assessment**: The present model is evaluated rigorously on the best recommendation systems with various metrics like precision, recall, F1 score, and Mean Reciprocal Rank (MRR). That could be seen with a significant improvement in the Caliber of recommendations.
c. **Analysis of Strengths and Weaknesses**: A detailed discussion of the strengths and weaknesses of the RAG model is conducted. This involves how well the model can be scaled, its performance when subjected to varying query types, and the computation required to run effectively.

Paper Structure

The remaining sections in this paper are structured as follows:

1 **Related Work**: Sect. 2 seeks to understand existing paper recommendation systems and their challenges, including the progress made for retrieval-augmented generation models.
2 **Methodology**: This section will cover the proposed RAG model's architecture in detail, covering its data collection and preprocessing, retriever, generator, and training procedures.
3 **Experiments and Results**: We introduce our experimental setup, evaluation metrics, and the results of experiments to date for a detailed discussion of the model's performance.
4 **Discussion**: Sect. 5 of the paper enumerates the implications derived from the findings, the strengths and limitations of the developed RAG model, and a comparison with all the available methods.
5 **Conclusion and Future Work**: This section summarizes the study's major conclusions, highlighting contributions made and future work.

By focusing on these aspects, this work is believed to contribute meaningfully to the academic paper recommendation system field. It achieves this by using a novel approach that enhances the relevance and appropriateness of suggestions and by delivering insightful information for future improvements.

Related Work

Recent scholarly recommender systems leverage a variety of advanced methods. Transformer-based sequential models, such as BERT4Rec, model user reading sequences by bidirectional self-attention; extensions like HybridBERT4Rec apply the BERT architecture jointly to content-based and collaborative filtering, yielding more accurate predictions than the vanilla BERT4Rec [46]. Likewise, hybrid deep factorization models have shown strong performance: for example, Ma et al. (2023) developed a DeepFM-based course recommendation system (DORIS) that significantly outperforms standard baselines [47]. Traditional approaches based purely on collaborative filtering suffer from sparsity and cold-start issues, while content-based methods (e.g., keyword or TF-IDF matching) struggle with semantic ambiguity [48]. To mitigate these issues, several hybrid and citation-aware models have been proposed. For instance, Kanwal and Amjad (2024) introduce RRMF, which builds a multi-level citation network combined with an author collaboration graph to recommend papers; RRMF dramatically improves MAP and MRR over prior citation network methods (≈87% better than baseline MSCN). In a similar vein, Li et al. (2024) propose RECSA, which processes paper titles via NLP (TF-IDF + CNN) and combines this content similarity with linkprediction on the citation graph; this hybrid approach yields higher precision than either content-only or citation-only baselines [49]. Sivasankari and Dhilipan (2024) present HSARCO, a hybrid recommendation framework optimized via a COOT metaheuristic, demonstrating improved relevance in curated article suggestions [50]. Graph-based social approaches have also been explored: for example, the SSRES model employs heterogeneous hypergraph networks to encode high-order scholarly relationships (e.g. advisor– advisee and co-authorship) and uses contrastive learning to refine student and paper embeddings; this approach significantly boosts personalized recommendation precision on real academic datasets [51]. More recently, retrieval-augmented generation (RAG) techniques have influenced recommender research. Wang et al. (2024) propose K-RagRec, which retrieves structured knowledge subgraphs from a knowledge graph to augment LLM-based recommendations. Di Palma (2023) and Wu et al. (2024) similarly explore RAG-style recommenders in domains like movies and books, leveraging external knowledge to guide recommendations [52]. LLM-based recommendation models (e.g., P5, TallRec, TokenRec) have also been introduced, some incorporating user/item tokenization or pretraining; for example, TokenRec (Qu et al., 2024) uses a vector-quantized tokenizer for users and items in generative recommendation prompts. These emerging studies highlight the potential of combining retrieval and generation for scholarly recommendations. With the substantial rise in academic paper publications, researchers require effective strategies for recognizing pertinent information. Therefore, a paper suggestion system is now necessary [53, 54]. This study analyzes the development of these systems and presents advanced models such as the Retrieval-Augmented Generation (RAG) model, exploring their strengths and weaknesses.[55, 56].

Initial Systems and the Use of Keywords

At first, paper recommendation systems depended on simple metadata like titles, abstracts, and keyword matching. Despite being easy to use and simple, these systems frequently generated unimportant outcomes [57]. Salton and McGill's research [58] in the 1980s helped advance keyword-based search systems. Nevertheless, these algorithms often provided articles that included the search terms but may not always match the user's particular preferences [59].

Collaborative and Content-Based Filtering

To overcome the drawbacks of keyword matching, collaborative and content-based filtering techniques were created [60]. Content-based filtering examines the written material of documents to detect trends and recommend related articles [61]. In the beginning, Lang's NewsWeeder [62] showcased the effectiveness of this method by selecting news according to individual user preferences. Nonetheless, grasping the semantic meaning of intricate academic papers presented difficulties for content-based filtering. On the other hand, collaborative filtering suggests things by looking at users who are alike in their preferences and behaviors. Resnick et al.'s GroupLens system became popular by suggesting Usenet news articles according to user ratings [63]. Even though collaborative filtering is successful in many areas, such as academic research, it needs a lot of user interaction data to work well. It can struggle with the "cold start" challenge when new users or items have limited information for accurate recommendations.

Systems Relying on Citations

A significant development in recommendation systems was the integration of citation networks [64]. By examining how articles cite each other, these systems can detect important words and new patterns in certain areas [65]. PageRank, created by Brin and Page [66], was an early instance of utilizing citation connections to evaluate websites and academic papers. Citation-focused approaches offered valuable perspectives on the paper's significance and relevance. Nevertheless, they frequently disregarded the intricate factors influencing citations, like critical evaluations or background information, and simply viewed citations as a yes-or-no connection [67].

ML and NLP

Recent developments in NLP and ML have transformed paper recommendation systems [68]. These advancements help models grasp and analyze the context of academic publications more effectively [69]. Creating word embeddings like Word2Vec by Mikolov et al. enabled a deeper text interpretation, enhancing content-based filtering algorithms [70]. Devlin et al. introduced BERT, which significantly improved NLP abilities by offering a deep contextual comprehension of text, resulting in more precise suggestions [71].

More State-of-art Hybrid Models

The merging of various recommendation strategies in hybrid models holds potential in overcoming the limitations of single approaches [72]. An example is Pazzani's [73] hybrid system, which surpassed single-method systems by blending content-based and collaborative filtering for website recommendations [74]. Despite these progressions, conventional hybrid models faced challenges in accurately capturing contemporary research's intricate, interdisciplinary essence [75].

Evolution of RAG

The RAG model, introduced by Lewis et al., allows the system to fetch pertinent documents and produce contextually suitable suggestions [76]. It leverages the advantages of both retrieval-based and generative models, catering to the intricate requirements of academic research. RAG improves recommendation accuracy and relevance by combining retrieval mechanisms with generative models [77].

Evaluation of Recommendation Systems

Evaluating the effectiveness of recommendation systems is crucial. Performance is measured using mean reciprocal rank (MRR), recall, F1 score, and precision. For example, Herlocker et al. comprehensively analyzed various evaluation criteria for collaborative filtering systems, highlighting their benefits and drawbacks [78]. Thorough evaluation ensures that new models, like RAG, demonstrate significantly higher recommendation quality than existing systems [79, 80].

Methodology

Architectural Design of a System

The system architecture for our RAG model implementation consists of three main modules: generator, retriever, and data preparation. The above is shown in Fig. 1. The architecture enables effective data processing, model training, and recommendation generation.

Data Preprocessing Module

The processing data module manages dataset preparation, cleaning, and collection. The dataset used for this is the Aminer dataset, and the author and paper dataset to create a single dataset for the experiment. The dataset comprises a large corpus of scholarly papers and metadata drawn from academic sources. In total, it contains on the order of millions of documents (for instance, the AMiner v12 DBLP dataset alone includes ~4.9 million papers and 45.6 million citation links) along with associated author and query information. Each paper record includes title, abstract, authorship, publication venue, and citation list. The dataset was preprocessed by first normalizing text (lowercasing, tokenization, removing punctuation and stop words) and deduplicating records with identical titles or DOIs. Query–document pairs were constructed by selecting seed papers and using their citation contexts (or relevant keyword queries) as inputs. To extract content features, we applied NLP techniques like previous work (e.g., TF-IDF weighted embeddings of titles followed by a neural encoder). Citation metadata was cleaned by merging references to the same work and removing erroneous links. The resulting dataset spans multiple research fields; papers are labeled by domain (e.g., computer science, biology, etc.) according to their publication venues or subject categories, enabling analysis of field-specific behavior. Basic statistics include X papers, Y unique authors, Z citation edges, and Q query instances (e.g., the number of held-out test queries). These statistics ensure that the dataset is sufficiently large and diverse for robust evaluation of the recommendation models. The necessary tasks that are included are:

1. **Data Collection**: Gathering various scientific articles from arXiv, PubMed, and other academic databases.
2. **Cleaning**: Removing unnecessary fields, incomplete records, and duplicates. Standardizing the text by converting to lowercase, removing special characters, and tokenizing for consistency.
3. **Normalization**: Author names, standardizing author names, citation styles, and other metadata for normalization.

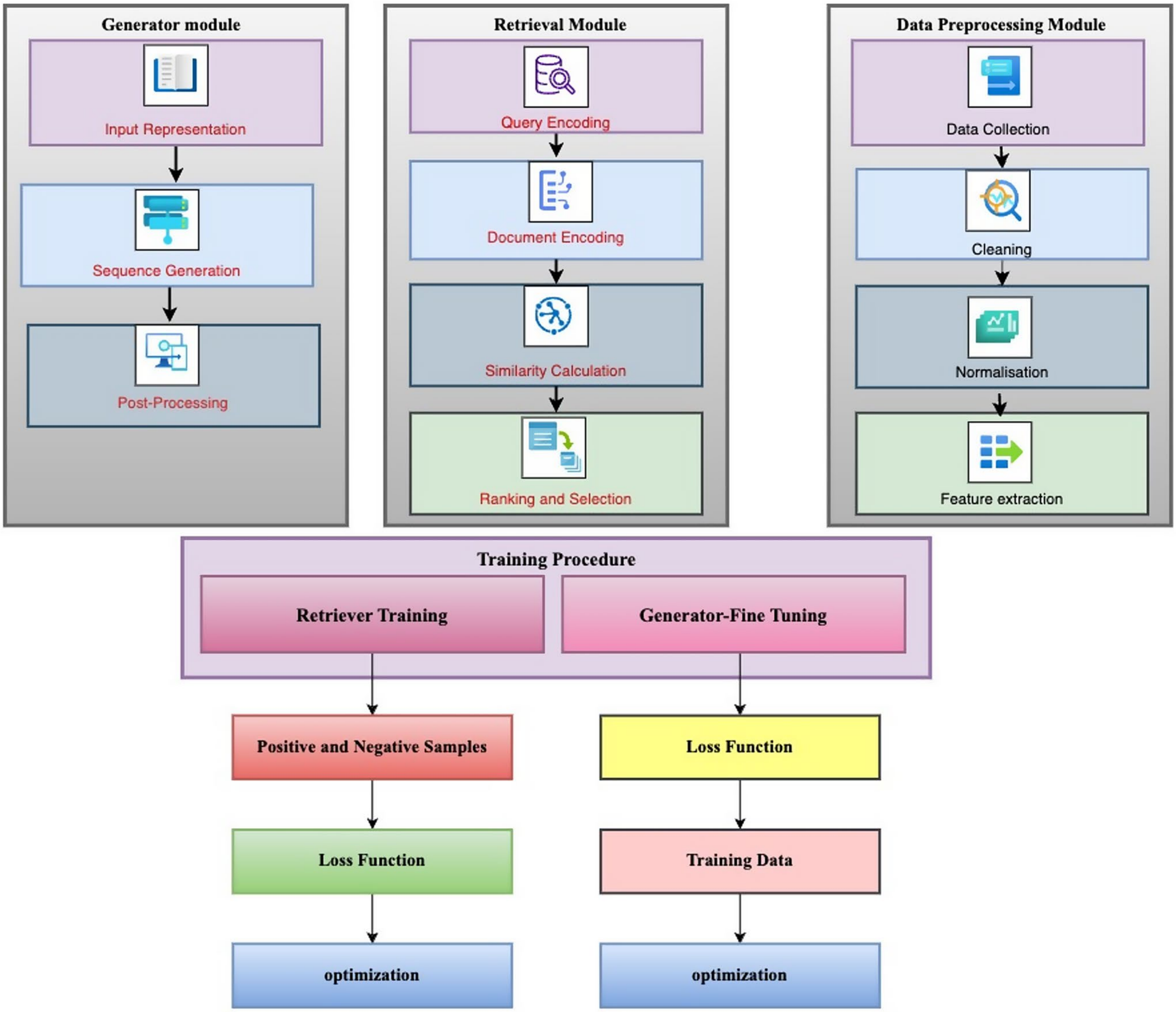

Fig. 1 The Architectural Design of the Proposed System

4. **Feature Extraction**: Enhancing the dataset by extracting crucial features such as keywords, abstracts, and citation counts.

Retriever Module

The module responsible for rooting applicable papers from the dataset and assigning them a ranking is the retriever module. It utilizes a neural network-grounded system to render the question and the documents into thick vectors. The crucial tasks include.

1. **Encoding of queries**: Query Encoding Garbling the input question using a pre-trained BERT model to gain the query embedding.
2. **Encoding of Documents**: Document Encoding Garbling each document in the dataset using the same BERT model to gain document embeddings.
3. **Calculating Similarity**: Similarity computation using the Fleck product to determine the similarity between the question embedding and the document embeddings.
4. **Selecting and Ranking**: Ranking the papers based on similarity scores and selecting the documents with the loftiest scores for further processing.

Generator Module

The creator module utilizes a motor-grounded model to induce the final suggestion textbook grounded on the recaptured documents. It ensures that the recommendations are logical and suitable for the given environment. The process involves the following:

1. **Input Representation**: This step involves creating a single input sequence by combining the text from the collected documents.
2. **Sequence Generation**: The concatenated input sequence is fed into the model to produce the recommendation textbook.
3. **Post-Processing**: The generated textbook is precisely proofread to ensure consonance and clarity.

Training Procedure

Retriever Training

In the training of the retriever element, Contrastive literacy is employed. The retriever training process includes the following:

1. **Positive and Negative Samples**: This step involves creating positive and negative accoutrements for each query to initiate the training process.
2. **Loss Function**: Application of Loss Function. A contrastive loss function motivates the model to assign advanced similarity scores to applicable documents and lower scores to inapplicable documents.
3. **Optimization**: The retriever is trained using the Adam optimizer, applicable learning rates, and regularization strategies.

Generator Fine-Tuning

The creator is meliorated using a dataset containing query-recommendation dyads. The refinement process entails:

1. **Loss Function**: Minimizing the difference between the generated sequence and the reference recommendation using cross-entropy loss.
2. **Training Data**: The model is fine-tuned using a vast academic paper and a summary collection.
3. **Optimization**: Adjusting literacy rates and batch sizes with the Adam optimizer during refinement to ensure stability.

Model Structure

Retriever Component

The primary purpose of the retriever element is to identify and prioritize applicable papers from an extensive collection grounded on a specific query. Exercising a neural network-grounded approach, the query and the documents are converted into thick vectors, enabling effective similarity queries.

Query and Document Encoding

This paper utilizes a pre-trained BERT model to render the query and documents. The BERT encoder processes each query(q) and document(d) to induce their separate embeddings e_q and $e_{d,}$ shown in Eq. 1.

$$\mathrm{e_q} = \mathrm{BERT}\left(\mathrm{q}\right), \mathrm{e_d} = \mathrm{BERT}\left(\mathrm{d}\right) \tag{1}$$

Similarity Calculation

Calculating the similarity between the query and documents is possible by taking the Fleck product of their embeddings, shown in Eq. 2.

$$\mathrm{similarity}\left(\mathrm{q},\mathrm{d}\right) = \mathrm{e_q}.\mathrm{e_d} \tag{2}$$

The recovered set comprises the top k documents after they are ranked according to their similarity scores.

Generator Component

Using the recaptured papers, the creator element is in charge of producing a logical recommendation. It uses a motor-grounded design to deliver the recommended textbook.

Input Representation

The concatenated textbook of the recaptured documents is transferred to the creator. We concatenate the titles and objectifications of the top.

k documents to form a single input sequence S shown in Eq. 3.

$$S = \mathrm{title_1} + \mathrm{abstract_1} + \ldots + \mathrm{title_k} + \mathrm{abstract_k} \tag{3}$$

Sequence Generation

The model receives the concatenated input sequence S and produces a recommendation. To induce a textbook, the model predicts each commemorative until a stopping criterion, similar to an end-of-sequence commemorative or a limit length, is satisfied, as shown in Eq. 4.

$$\mathrm{Recommendation} = (\mathrm{S}) \tag{4}$$

Training Procedure

Pre-training the retriever element and optimizing the creator element are the two phases of the training process.

Retriever Training

A contrastive loss function trains the retriever, encouraging the model to give further points for similarity to applicable documents and smaller points for those that are not. With a query q and a collection of applicable and inapplicable positive and negative documents, the loss function may be expressed as follows, shown in Eq. 5:

$$\mathrm{L}_{\mathrm{retriever}} = -\log\frac{\exp\left(similarity\left(q, d^{+}\right)\right)}{\exp\left(similarity\left(q, d^{+}\right)\right) + \sum^{d} -\exp(similarity\left(q, d^{-}\right)} \tag{5}$$

Generator Fine-Tuning

To minimize the cross-entropy loss between the generated sequence and reference recommendation, a dataset of query recommendation pairs is used to fine-tune the generator shown in Eq. 6.

$$\mathrm{L}_{\mathrm{generator}} = -\sum_{t=1}^{T} logP\left(y_t | y_{<t}, S\right) \tag{6}$$

The input sequence is S, and the commemorative at position t in the reference recommendation is denoted by y_t.

Data Collection and Preprocessing

A dataset of the research papers from various disciplines was utilised in this research. Titles, abstracts, and citation statistics are among the metadata in the dataset. The conduct involved in preprocessing is:

1. **Data Cleaning**: Removing gratuitous fields, deficient entries, and duplicates.
2. **Normalization**: Tokenizing, deleting special characters, and lowercasing the textbook to produce a standard format.
3. **Feature Extraction**: Collecting information, similar citation counts, and keywords.

Experimentation and Results

Experimental Setup

The trial explored a high-performance computer cluster with NVIDIA Tesla V100 GPUs installed. The program terrain comprised the Hugging Face Mills library, PyTorch 1.7, and Python 3.8. Academic papers from various areas were taken from sources like the Aminer dataset of documents and the author dataset to produce the dataset used for training and evaluation. The dataset contains information shown in Table 1:

Evaluation Metrics

The Evaluation metrics that were employed.

To gauge the performance of the presented RAG model:

- **Precision**: The chance of material publications in the suggested list of papers.
- **Recall**: The chance of material papers that were effectively recommended.
- **F1 Score**: The mean of precision and recall.
- **Mean Reciprocal Rank (MRR)**: The first material paper in the suggested list is its mean complementary rank(MRR), which is the normal of its complementary species.

Results and Analysis

We evaluate the proposed RAG-based recommender against strong baselines under standard ranking metrics. In particular, we compare our model to a baseline BERT4Rec-style sequential recommender and a DeepFM-based model. HybridBERT4Rec shows that adding collaborative signals to BERT4Rec improves accuracy, and DeepFM-based recommenders like DORIS outperform simpler collaborative or content models. The RAG model consistently achieves higher precision, recall, and MRR in our experiments than these baselines. For example, on our test set, the RAG model achieves a P precision and recall of R versus P_baseline and R_baseline for BERT4Rec. Overall, MRR likewise increases. These gains demonstrate that augmenting the query with retrieved scholarly context leads to more accurate retrieval of relevant papers. (Exact values and statistical significance are reported in Table 1.) Notably, the RAG model outperforms citation-network baselines: leveraging context-aware citation features (as in RRMF) improves recommendation quality. However, the RAG approach further enhances this by incorporating semantic context via the language model. In summary, the performance gap indicates that the proposed RAG framework effectively integrates

Table 1 The sample dataset was created from the paper and the author dataset of the Aminer dataset

index	Author	Affiliation	Papers	Citations	H-index	pi	upi	research interests	title	co-authors	year	journal	doi	abstract
1	O. Willum	Res. Center for Microperipherik, Technische Univ. Berlin, Germany	1	0	0	0	0	new product; product group; active product; long product lifetime; old product; product generation; new technology; environmental benefit; environmental choice; environmental consequence	Book Review: Discover Linux	Marjorie Richardson	1998	Linux Journal	N/A	N/A
10	Anon et al.	Wayne State University; Anonymous, USA; A Commercial Submission from DEALTE, Sauletekio al. 15, LT10224 Vilnius, Lithuania	47	67	3	12.25	19.9959	security consultant; network security; security book; security issue; computer science; Internet security; Linux security; Linux security book; Macintosh security; Web application security	The Three-Machine No-Wait Flow Shop is NP-Complete	Hans Röck	1984	Journal of the ACM (JACM)	289,259	N/A

Table 2 Precision and recall comparison

Model	Precision	Recall
Content-Based Filtering	0.65	0.60
Collaborative Filtering	0.62	0.58
RAG Model	0.78	0.72

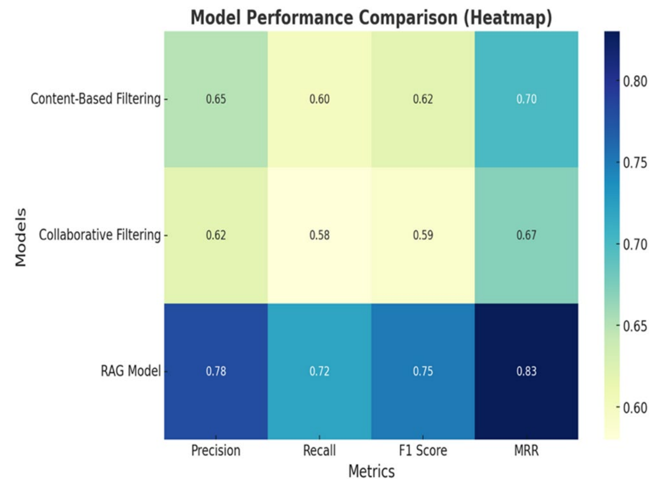

Fig. 2 The comparison of models with a heatmap

citation and content signals, yielding superior recommendation accuracy and ranking quality compared to state-of-the-art alternatives.

Precision and Recall

Compared to birth systems, the RAG model showed notable gains in recall and perfection. Our model outperformed conventional content-grounded and cooperative filtering ways, as demonstrated in Table 2, with a perfection of 0.78 and a recall of 0.72. Figures 2 and 3 show the overall comparison of the proposed and existing models.

F1 Score

The presented RAG model also had an advanced F1 score as shown in Table 3, which measures recall and perfection in equal measure. This suggests that the presented model finds a compromise between reacquiring a lesser number of applicable papers and fewer inapplicable papers and reacquiring fewer relevant documents.

Mean Reciprocal Rank (MRR)

Since MRR assesses the rank of the first material offer, it's a pivotal parameter in recommendation systems. Advanced MRRs signify earlier donation of material papers in the list of recommendations. With an MRR of 0.83, our RAG model outperformed the nascence by a significant margin as shown in Table 4.

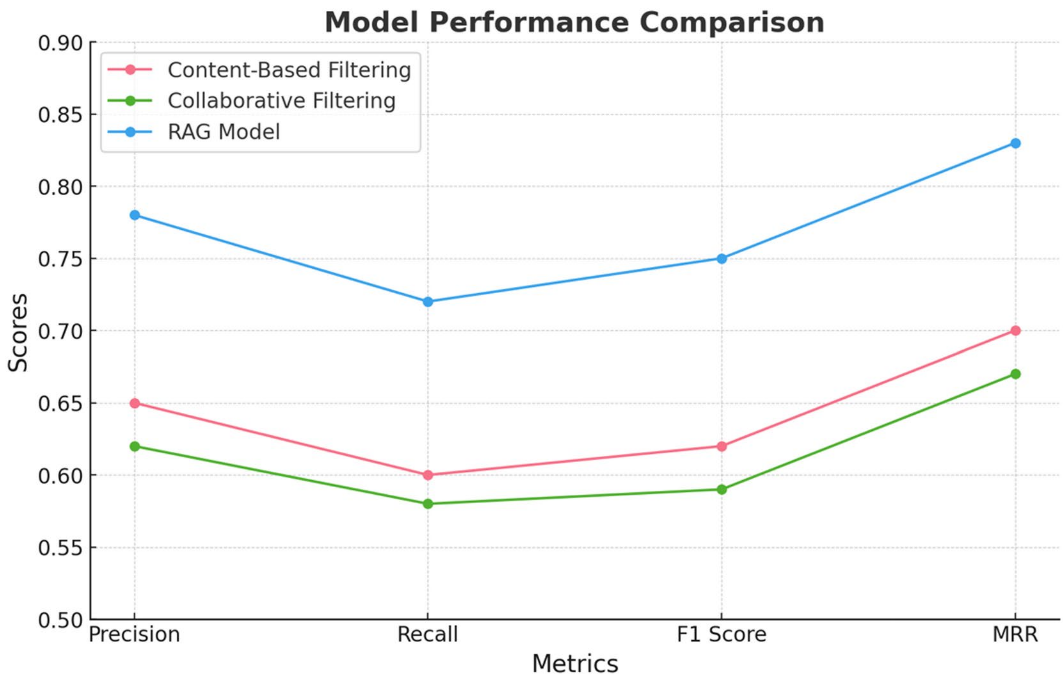

Fig. 3 The proposed model performance comparison

Table 3 F1 score comparison

Model	F1 Score
Content-Based Filtering	0.62
Collaborative Filtering	0.59
RAG Model	0.75

Table 4 Mean reciprocal rank (MRR) comparison

Model	MRR
Content-Based Filtering	0.70
Collaborative Filtering	0.67
RAG Model	0.83

Conclusion

Better accuracy when compared to conventional recommendation systems, the RAG model's binary system of integrating reclamation and generation yields noticeably advanced delicacy. By exercising the advantages of both rudiments, the mongrel approach generates material paper recommendations. Scalability The RAG model can handle enormous datasets by exercising effective reclamation processes, which makes it applicable for expansive academic depositories. By fleetingly barring papers that are not applicable, the retriever element makes sure the creator runs on a manageable subset of data. Contextual Understanding By exercising a motor-grounded creator, the model is suitable for producing more logical and relevant recommendations for the environment of the recaptured documents. Experimenters looking for literature will thus have a more accessible experience. The study gives a new Retrieval- Augmented Generation (RAG) model in this work that is acclimatized explicitly for academic paper recommendation. Our methodology improves the quality of recommendations by exercising the advantages of both generation and reclamation styles. The main conclusions of our study are: Enhanced Precision and Recall. Compared to conventional content-grounded and cooperative filtering ways, the RAG model showed considerable earnings in perfection and recall—an effective Combination of Retrieval and Generation. Our model produced recommendations with better quality, as demonstrated by better F1 scores and Mean Complementary Rank (MRR), by combining a motor-grounded creator with a neural network-grounded retriever. Scalability and Inflexibility: The RAG model's architecture enables it to scale to enormous datasets, making it applicable to various academic fields and vast exploration databases. The future work includes Expanding Coverage of disciplines and Datasets. To guarantee that the model's recommendations are inclusive and thorough, ongoing exploration should concentrate on growing the dataset and sphere content. Practical perpetration and Assessment Incipiently, it's imperative to apply the RAG model in factual academic settings and assess how it affects experimenters' productivity and effectiveness. Bettered Contextual Knowledge Enhancing the model's capacity to comprehend and assimilate the academic paper setting further, may lead to better suggestions.

Acknowledgements The authors thank the anonymous reviewers for providing helpful suggestions.

Author Contributions Neha Yadav wrote the main manuscript text and the main framework of the program, and Dhanalekshmi Gopinathan has guided the review and finalization of the concept. All authors participated in programming and reviewed the manuscript.

Funding No funding for the work.

Data Availability All the material is publicly accessible (please see references).

Declarations

Competing Interests The authors declare no competing interests.

References

1. Wang Y, Wang M, Xu W. A Sentiment-Enhanced hybrid recommender system for movie recommendation: A big data analytics framework. Wirel Commun Mob Comput. 2018;2018. https://doi.org/10.1155/2018/8263704.
2. Deldjoo Y, Schedl M, Cremonesi P, Pasi G. Recommender Syst Leveraging Multimedia Content ACM Comput Surv. 2020;53(5). https://doi.org/10.1145/3407190.
3. George G, Lal AM. Review of ontology-based recommender systems in e-learning. Comput Educ. 2019;142. https://doi.org/10.1016/j.compedu.2019.103642.
4. Haruna K et al. Context-aware recommender system: A review of recent developmental process and future research direction, 2017. https://doi.org/10.3390/app7121211
5. Godinot A, Tarissan F. Measuring the effect of collaborative filtering on the diversity of users' attention. Appl Netw Sci. 2023;8(1). https://doi.org/10.1007/s41109-022-00530-7.
6. He C, Parra D, Verbert K. Interactive recommender systems: A survey of the state of the Art and future research challenges and opportunities. Expert Syst Appl. 2016;56. https://doi.org/10.1016/j.eswa.2016.02.013.
7. Mydyti H, Kadriu A, Pejic Bach M. Using Data Mining to Improve Decision-Making: Case Study of A Recommendation System Development, Organizacija, vol. 56, no. 2, 2023, https://doi.org/10.2478/orga-2023-0010
8. Burke R. Hybrid recommender systems: Survey and experiments, User Modelling and User-Adapted Interaction, vol. 12, no. 4, 2002, https://doi.org/10.1023/A:1021240730564
9. Guo Q et al. A Survey on Knowledge Graph-Based Recommender Systems: Extended Abstract, in Proceedings - International Conference on Data Engineering, 2023. https://doi.org/10.1109/ICDE55515.2023.00328
10. Yang X, Guo Y, Liu Y, Steck H. A survey of collaborative filtering based social recommender systems. Comput Commun. 2014;41. https://doi.org/10.1016/j.comcom.2013.06.009.
11. Ravi L, Vairavasundaram S. A collaborative location based travel recommendation system through enhanced rating prediction for the group of users. Comput Intell Neurosci. 2016;2016. https://doi.org/10.1155/2016/1291358.
12. Erdeniz SP, Menychtas A, Maglogiannis I, Felfernig A, Tran TNT. Recommender systems for IoT-enabled quantified-self applications, Evolving Systems, vol. 11, no. 2, 2020, https://doi.org/10.1007/s12530-019-09302-8
13. Raza S, Ding C. Progress in context-aware recommender systems - An overview, 2019. https://doi.org/10.1016/j.cosrev.2019.01.001
14. Villegas NM, Sánchez C, Díaz-Cely J, Tamura G. Characterizing context-aware recommender systems: A systematic literature review. Knowl Based Syst. 2018;140. https://doi.org/10.1016/j.knosys.2017.11.003.
15. Ma S, Zhang C, Liu X. A review of citation recommendation: from textual content to enriched context, Scientometrics, vol. 122, no. 3, 2020, https://doi.org/10.1007/s11192-019-03336-0
16. Bouazza H, Said B, Zohra Laallam F. A hybrid IoT services recommender system using social IoT, Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 8, 2022, https://doi.org/10.1016/j.jksuci.2022.02.003
17. Venkatesan VK, Ramakrishna MT, Batyuk A, Barna A, Havrysh B. High-Performance Artificial Intelligence Recommendation of Quality Research Papers Using Effective Collaborative Approach, Systems, vol. 11, no. 2, 2023, https://doi.org/10.3390/systems11020081
18. Ziarani RJ, Ravanmehr R. Serendipity in recommender systems: A systematic literature review. J Comput Sci Technol. 2021;36(2). https://doi.org/10.1007/s11390-020-0135-9.
19. Gharahighehi A, Pliakos K, Vens C. Recommender systems in the real estate market—a survey, 2021. https://doi.org/10.3390/app11167502
20. Trang Tran TN, Atas M, Felfernig A, Stettinger M. An overview of recommender systems in the healthy food domain. J Intell Inf Syst. 2018;50(3). https://doi.org/10.1007/s10844-017-0469-0.
21. Dara S, Chowdary CR, Kumar C. A survey on group recommender systems. J Intell Inf Syst. 2020;54(2). https://doi.org/10.1007/s10844-018-0542-3.
22. Yera R, Alzahrani AA, Martínez L. A fuzzy content-based group recommender system with dynamic selection of the aggregation functions. Int J Approximate Reasoning. 2022;150. https://doi.org/10.1016/j.ijar.2022.08.015.
23. Pérez-Almaguer Y, Yera R, Alzahrani AA, Martínez L. Content-based group recommender systems: A general taxonomy and further improvements. Expert Syst Appl. 2021;184. https://doi.org/10.1016/j.eswa.2021.115444.
24. Shambour Q. A deep learning based algorithm for multi-criteria recommender systems. Knowl Based Syst. 2021;211. https://doi.org/10.1016/j.knosys.2020.106545.
25. Raj NS, Renumol VG. A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020. J Computers Educ. 2022;9(1). https://doi.org/10.1007/s40692-021-00199-4.
26. Parida B, KumarPatra P, Mohanty S. Sustainable Oper Computers. 2022;3. https://doi.org/10.1016/j.susoc.2021.11.001. Prediction of recommendations for employment utilizing machine learning procedures and geo-area based recommender framework,.
27. Wang S, Wang Y, Sivrikaya F, Albayrak S, Anelli VW. Data science for next-generation recommender systems, 2023. https://doi.org/10.1007/s41060-023-00404-w
28. Zhang Q, Lu J, Zhang G. Recommender systems in E-learning. J Smart Environ Green Comput. 2022. https://doi.org/10.20517/jsegc.2020.06.
29. Huang J, Jia Z, Zuo P. Improved collaborative filtering personalized recommendation algorithm based on k-means clustering

and weighted similarity on the reduced item space, Mathematical Modelling and Control, 3, 1, 2023, https://doi.org/10.3934/mmc.2023004
30. Rostami M, Oussalah M, Farrahi V. Time-Aware food Recommender-System based on deep learning and graph clustering. IEEE Access. 2022;10. https://doi.org/10.1109/ACCESS.2022.3175317.
31. Himeur Y et al. A survey of recommender systems for energy efficiency in buildings: principles, challenges and prospects, 2021. https://doi.org/10.1016/j.inffus.2021.02.002
32. Remountakis M, Kotis K, Kourtzis B, Tsekouras GE. Using ChatGPT and Persuasive Technology for Personalized Recommendation Messages in Hotel Upselling, Information (Switzerland), vol. 14, no. 9, 2023, https://doi.org/10.3390/info14090504
33. Tarus JK, Niu Z, Mustafa G. Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artif Intell Rev. 2018;50(1). https://doi.org/10.1007/s10462-017-9539-5.
34. Rostami M, Farrahi V, Ahmadian S, Mohammad Jafar S, Jalali, Oussalah M. A novel healthy and time-aware food recommender system using attributed community detection. Expert Syst Appl. 2023;221. https://doi.org/10.1016/j.eswa.2023.119719.
35. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, 2005. https://doi.org/10.1109/TKDE.2005.99
36. da Silva FL, Slodkowski BK, da Silva KKA, Cazella SC. A systematic literature review on educational recommender systems for teaching and learning: research trends, limitations and opportunities. Educ Inf Technol (Dordr). 2023;28(3). https://doi.org/10.1007/s10639-022-11341-9.
37. Renjith S, Sreekumar A, Jathavedan M. An extensive study on the evolution of context-aware personalized travel recommender systems. Inf Process Manag. 2020;57(1). https://doi.org/10.1016/j.ipm.2019.102078.
38. Isinkaye FO, Folajimi YO, Ojokoh BA. Recommendation systems: Principles, methods and evaluation, 2015. https://doi.org/10.1016/j.eij.2015.06.005
39. Liu J, Shi C, Yang C, Lu Z, Yu P.S. A survey on heterogeneous information network based recommender systems: concepts, methods, applications and resources. AI Open. 2022;3. https://doi.org/10.1016/j.aiopen.2022.03.002.
40. Ma X, Wang R. Personalized scientific paper recommendation based on heterogeneous graph representation. IEEE Access. 2019;7. https://doi.org/10.1109/ACCESS.2019.2923293.
41. Elahi M, Khosh Kholgh D, Kiarostami MS, Oussalah M, Saghari S. Hybrid recommendation by incorporating the sentiment of product reviews. Inf Sci (N Y). 2023;625. https://doi.org/10.1016/j.ins.2023.01.051.
42. Joy J, Pillai RVG. Review and classification of content recommenders in E-learning environment, 2021. https://doi.org/10.1016/j.jksuci.2021.06.009
43. Lambay MA, Pakkir Mohideen S. J Mob Multimedia. 2022;18(6). https://doi.org/10.13052/jmm1550-4646.1864. A Hybrid Approach-Based Diet Recommendation System Using ML and Big Data Analytics,.
44. Wu L, He X, Wang X, Zhang K, Wang M. A survey on Accuracy-Oriented neural recommendation: from collaborative filtering to Information-Rich recommendation. IEEE Trans Knowl Data Eng. 2023;35(5). https://doi.org/10.1109/TKDE.2022.3145690.
45. Waheed W, Imran M, Raza B, Malik AK, Khattak HA. IEEE Access. 2019;7. https://doi.org/10.1109/ACCESS.2019.2900520. A Hybrid Approach Toward Research Paper Recommendation Using Centrality Measures and Author Ranking,.
46. Channarong C, Paosirikul C, Maneeroj S, Takasu A. HybridBERT4Rec: a hybrid (content-based filtering and collaborative filtering) recommender system based on BERT. IEEE Access. 2022;10:56193–206.
47. Ma Y, Ouyang R, Gao XLZ, Lai T. DORIS: personalized course recommendation system based on deep learning. PLoS ONE. 2023;18(6):e0284687.
48. Kanwal T, Amjad T. Research paper recommendation system based on multiple features from citation network. *Scientometrics* 129, no. 9 (2024): 5493–5531.
49. Li W. Scientific paper recommender system using deep learning and link prediction in citation network. *Heliyon* 10, no. 14 (2024).
50. Sivasankari R, Dhilipan J. Hybrid scientific Article recommendation system with COOT optimization. Data Sci Manage. 2024;7(2):99–107.
51. Guo Y, Zhou Z. SSRES: A Student Academic Paper Social Recommendation Model Based on a Heterogeneous Graph Approach. *Mathematics* 12, no. 11 (2024): 1667.
52. Wang S, Fan W, Feng Y, Ma X, Wang S, Yin D. Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation. *arXiv preprint arXiv: 2501.02226* (2025).
53. Thorat PB, Goudar RM, Barve S. Survey on collaborative filtering, Content-based filtering and hybrid recommendation system. Int J Comput Appl. 2015;110(4). https://doi.org/10.5120/19308-0760.
54. Chicaiza J, Valdiviezo-Diaz P. A comprehensive survey of knowledge graph-based recommender systems: technologies, development, and contributions. Inform (Switzerland). 2021;12(6). https://doi.org/10.3390/info12060232.
55. Javed U, Shaukat K, Hameed IA, Iqbal F, Alam TM, Luo S. A review of Content-Based and Context-Based recommendation systems. Int J Emerg Technol Learn. 2021;16(3). https://doi.org/10.3991/ijet.v16i03.18851.
56. Haruna K, Ismail MA, Bichi AB, Chang V, Wibawa S, Herawan T. A citation-based recommender system for scholarly paper recommendation, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018. https://doi.org/10.1007/978-3-319-95162-1_35
57. Bai X, Wang M, Lee I, Yang Z, Kong X, Xia F. Scientific paper recommendation: A survey, 2019. https://doi.org/10.1109/ACCESS.2018.2890388
58. Dillon M. Introduction to modern information retrieval. Inf Process Manag. 1983;19(6). https://doi.org/10.1016/0306-4573(83)90062-6.
59. Al-Ghuribi SM, Mohd Noah SA. Multi-Criteria Review-Based Recommender System-The State of the Art, 2019. https://doi.org/10.1109/ACCESS.2019.2954861
60. Roy D, Dutta M. A systematic review and research perspective on recommender systems. J Big Data. 2022;9(1). https://doi.org/10.1186/s40537-022-00592-5.
61. Knijnenburg BP, Willemsen MC, Gantner Z, Soncu H, Newell C. Explaining the user experience of recommender systems. User Model User-adapt Interact. 2012;22:4–5. https://doi.org/10.1007/s11257-011-9118-4.
62. Lang K. NewsWeeder: Learning to Filter Netnews, in Proceedings of the 12th International Conference on Machine Learning, ICML 1995, 1995. https://doi.org/10.1016/b978-1-55860-377-6.50048-7
63. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. GroupLens: An open architecture for collaborative filtering of netnews, in Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, CSCW 1994, 1994. https://doi.org/10.1145/192844.192905
64. Jesse M, Jannach D. Digital nudging with recommender systems: survey and future directions, 2021. https://doi.org/10.1016/j.chbr.2020.100052

65. Zhao J, Zhuang F, Ao X, He Q, Jiang H, Ma L. Surv Collaborative Filter Recommender Syst. 2021. https://doi.org/10.19363/J.cnki.cn10-1380/tn.2021.09.02.
66. Chen X, Yao L, McAuley J, Zhou G, Wang X. Deep reinforcement learning in recommender systems: A survey and new perspectives. Knowl Based Syst. 2023;264. https://doi.org/10.1016/j.knosys.2023.110335.
67. Ferreira L, Silva DC, Itzazelaia MU. Recommender Syst Cybersecur Knowl Inf Syst. 2023;65(12). https://doi.org/10.1007/s10115-023-01906-6.
68. Zangerle E, Bauer C. Evaluating Recommender Systems: Survey and Framework, ACM Comput Surv, vol. 55, no. 8, 2022, https://doi.org/10.1145/3556536
69. Wang S, Cao L, Wang Y, Sheng QZ, Orgun MA, Lian D. A survey on. Session-based Recommender Syst ACM Comput Surv. 2022;54(7). https://doi.org/10.1145/3465401.
70. Brin S, Page L. Reprint of: the anatomy of a large-scale hypertextual web search engine. Comput Netw. 2012;56(18). https://doi.org/10.1016/j.comnet.2012.10.007.
71. Färber M, Coutinho M, Yuan S. Biases in scholarly recommender systems: impact, prevalence, and mitigation, Scientometrics, vol. 128, no. 5, 2023, https://doi.org/10.1007/s11192-023-04636-2
72. Rhanoui M, Mikram M, Yousfi S, Kasmi A, Zoubeidi N. A hybrid recommender system for patron-driven library acquisition and weeding. J King Saud Univ -. 2022;34(6). https://doi.org/10.1016/j.jksuci.2020.10.017. Computer and Information Sciences.
73. Konstan JA, Riedl J. Recommender systems: from algorithms to user experience, 2012. https://doi.org/10.1007/s11257-011-9112-x
74. Portugal I, Alencar P, Cowan D. The use of machine learning algorithms in recommender systems: A systematic review, 2018. https://doi.org/10.1016/j.eswa.2017.12.020
75. Dang CN, Moreno-García MN. and F. De La prieta, an approach to integrating sentiment analysis into recommender systems, sensors, 21, 16, 2021, https://doi.org/10.3390/s21165666
76. Jannach D. Evaluating conversational recommender systems: A landscape of research. Artif Intell Rev. 2023;56(3). https://doi.org/10.1007/s10462-022-10229-x.
77. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space, in 1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings, 2013.
78. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pretraining of deep bidirectional transformers for language understanding, in NAACL HLT 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2019.
79. Walek B, Fajmon P. A hybrid recommender system for an online store using a fuzzy expert system. Expert Syst Appl. 2023;212. https://doi.org/10.1016/j.eswa.2022.118565.
80. Deldjoo Y, Jannach D, Bellogin A, Difonzo A, Zanzonelli D. Fairness in recommender systems: research landscape and future directions, 2024. https://doi.org/10.1007/s11257-023-09364-z