

Article

High-Performance Artificial Intelligence Recommendation of Quality Research Papers Using Effective Collaborative Approach

Vinoth Kumar Venkatesan ¹, Mahesh Thyluru Ramakrishna ^{2,*}, Anatoliy Batyuk ³, Andrii Barna ⁴
and Bohdana Havrysh ⁵

¹ School of Information Technology and Engineering, VIT University, Vellore 632014, Tamil Nadu, India

² Department of Computer Science and Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Bangalore 562112, Karnataka, India

³ Department of Automated Control Systems, Lviv Polytechnic National University, 79013 Lviv, Ukraine

⁴ Department of Artificial Intelligence, Lviv Polytechnic National University, 79013 Lviv, Ukraine

⁵ Department of Publishing Information Technologies, Lviv Polytechnic National University, 79013 Lviv, Ukraine

* Correspondence: t.mahesh@jainuniversity.ac.in

Abstract: The Artificial Intelligence Recommender System has emerged as a significant research interest. It aims at helping users find things online by offering recommendations that closely fit their interests. Recommenders for research papers have appeared over the last decade to make it easier to find publications associated with the field of researchers' interests. However, due to several issues, such as copyright constraints, these methodologies assume that the recommended articles' contents are entirely openly accessible, which is not necessarily the case. This work demonstrates an efficient model, known as RPRSCA: Research Paper Recommendation System Using Effective Collaborative Approach, to address these uncertain systems for the recommendation of quality research papers. We make use of contextual metadata that are publicly available to gather hidden relationships between research papers in order to personalize recommendations by exploiting the advantages of collaborative filtering. The proposed system, RPRSCA, is unique and gives personalized recommendations irrespective of the research subject. Thus, a novel collaborative approach is proposed that provides better performance. Using a publicly available dataset, we found that our proposed method outperformed previous uncertain methods in terms of overall performance and the capacity to return relevant, valuable, and quality publications at the top of the recommendation list. Furthermore, our proposed strategy includes personalized suggestions and customer expertise, in addition to addressing multi-disciplinary concerns.

Keywords: recommender system; quality; artificial intelligence; publications; research paper; collaborative approach; accuracy; precision; recall



Citation: Venkatesan, V.K.; Ramakrishna, M.T.; Batyuk, A.; Barna, A.; Havrysh, B. High-Performance Artificial Intelligence Recommendation of Quality Research Papers Using Effective Collaborative Approach. *Systems* **2023**, *11*, 81. <https://doi.org/10.3390/systems11020081>

Academic Editor: Fernando De la Prieta Pintado

Received: 13 January 2023

Revised: 31 January 2023

Accepted: 31 January 2023

Published: 4 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A recommendation system is a type of artificial intelligence that uses Big Data to advise or promote more products to customers. It is typically coupled with machine learning. Systems that make recommendations to users based on various parameters are known as recommender systems. These systems forecast the products that customers will be most interested in and likely to buy. Recommender systems are used by businesses such as Netflix, Amazon, and others to assist their customers with finding the ideal product or movie for them. The recommender system handles the abundance of information by filtering the most crucial information based on the information provided by a user, as well as other criteria that consider the user's choice and interest. It determines whether a user and an item are compatible and then assumes that they are similar in order to make recommendations.

The overabundance of data accessible through the internet makes it difficult to find information. As a result, researchers find it difficult to access and monitor their interest in the most important and promising research papers [1]. Sending a request message to the site to provide the relevant details is the simplest and most popular method for searching for similar publications [2]. However, the outcomes of this technique rely primarily on how well the user can fine-tune the request message and the inability to individualize the result of the search. Following the references list from documents they already own is another practical approach that many researchers use [3]. While this method can be very effective in a few cases, it does not ensure complete coverage of all the papers related to research and recommendations. Articles written after the possessed paper cannot be tracked. Furthermore, the list of sources may not be publicly accessible; hence, it may be difficult to access the researchers.

The research paper recommendation systems suggest related papers automatically to researchers, which is an alternate solution presented in the literature [4,5] based on initial user knowledge that is more elaborate than a few keywords. Recommendation systems integrate users' contexts and the consumed material's potential contextual details to provide more detailed and specific recommendations [6]. Different researchers have suggested using details given by another person, such as using a citation list [7], an author's [8] articles list, a single paper [9], etc. In these methods, from this initial information, a user profile is built to reflect the users' preferences and to search the system for things or other profiles which are similar to another to generate the recommendations. Thus, the biggest problem was not to deliver excellent suggestions to the researchers or users at any point, irrespective of any place, but to give the right publication in the correct way to the intended researcher [10–12].

The proposed approach's key contributions are summarized as follows:

- The benefits of publicly accessible contextual metadata are used to recommend an independent research paper that does not include an a priori user profile.
- The proposed approach presents tailored suggestions for the recommendation of quality research papers, independent of any field of research or customer expertise.

Unlike existing works, this paper proposes a novel method based on collaborative filtering, which uses publicly accessible contextual metadata to personalize suggestions based on the secret ties between research papers. Not only does our proposed approach include personalized recommendations irrespective of the study area, regardless of consumer knowledge, but it also tackles multi-disciplinary issues.

The remaining portion of the paper is structured as follows: The second section provides a comprehensive survey of the related work being carried out. The third section details the proposed methodology. The fourth section details the implementation of the suggested strategy and a comparison to existing approaches to demonstrate the improved performance. The final section concludes the paper and describes future enhancements.

2. Related Works

The critical drawback of the current uncertain systems is that their expectation is for the entire content of the suggested papers to be available for free. This might be invalid sometimes due to reasons such as limitations on copyright. The authors [13] developed the idea of a collaborative method for mining the secret connections between specific papers and, of course, their references to resolve this issue and present unique and valuable research lists of papers as recommendations. The intention was to remove the secret ties between a specific target and its sources. In this article, we also include as information the hidden relations among the citations of particular research contained in the target paper. Our work is not to create a clear connection among paper–citation relationships, since a researcher who is directly or indirectly involved in the requirement of a specific paper, has the privilege to access its restricted references as well as the citations.

Over the past decade, research paper recommenders have emerged to assist researchers in seemingly seeking jobs of their interest in the so-called cyber ocean of knowledge, offer-

ing the best recommendations for all alternative uncertain systems. This article presents the effective collaborative filtering method (CF) used for recommendation systems. Based on what was previously favored by other, similar users [14–16], it is a method that suggests products to target users. It has been used to present films [17], audio compact disks [18], electronic commerce [19], and music [20], among others, in various applications. Employing this technique to suggest scholarly papers has been questioned by some researchers. According to the authors mentioned in [21–26], particularly in the research field where the count of users requesting suggestions is more significant than the total number of things to be nominated or recommended, collaborative filtering is proper; such few domains include films [27], music [28], etc. However, the point is that the researchers are not prepared to invest their precious time in explicitly providing their own ratings. In addition, a tangible count of scores is required for a consumer to obtain valuable recommendations.

Nevertheless, many papers can be traced amid the aforementioned problems. These recommend relevant articles dependent upon collaborative filtering via association mining among scholarly papers. These relations or associations are obtained either directly, by considering paper citations as ranking ratings, or by indirectly tracking the behavior of researchers [27]. Citation analysis, such as bibliographic coupling and the co-citation analysis mentioned in [28], was also used to classify papers identical to the target research paper.

In the survey conducted in [27], the relationships between research papers were classified into indirect and direct relations using uncertain systems. In this paper, three approaches were established based on the viewpoint of paper sources for detecting the relationships between articles. It is, therefore, only appropriate for identifying similarity relations via occasional papers. Citation analysis, on the other hand, can generate further connections between research articles, but it is unable to produce semantic text relationships. A context-based collaborative framework (CCF) was suggested in [3] that uses easily accessed citation associations as data sources. The framework uses an association-mining method to acquire a paper representation in the context of paper citation. A pairwise comparison was performed to compute the degree of similarities among papers. The use of collaborative filtering was also discussed in the paper [7]. This method uses the citation web to build a ranking matrix between scholarly articles. The goal was to propose a few additional references to the paper taken as the input, using the paper–citation relationship. While doing so, the authors explored using another six different citation selection algorithms. They discovered a significant difference in the accuracy returned by each of the six algorithms using offline evaluation.

In [6], the authors hypothesized that their previous publications had caused the researcher to develop a latent interest. One of the critical components of their proposed approach was to refine the profile of the user with the data. This data came straight away from references to the forgoing works of the researcher and the papers that cited them. This method, however, raises the issue of sparsity. They also rectified the citing of papers to use fragments in the citation, as well as using possible citation papers to describe a target candidate document. Although this method works very well for some researchers with a single discipline, it produces weak results for multi-disciplinary researchers.

With the previous study, the critical problem was that most of the historical knowledge from the suggested, referenced, and cited articles must be entirely available for recommenders. However, this information is not always freely accessible because of copyright constraints. Their reliance (reliability) is a more significant challenge with the current research paper recommendation systems. A major setback in constructing the new recommendation framework is the user's prior profile, which allows the system to function well if and only if it has a number of registered users. This is because suggested papers are stored beforehand and restricted to a specific study area, so the system does not accurately check the complete databases to discover links between articles. In addition, many of the current research paper systems are meant to operate exclusively in a single discipline and cannot be used to solve multi-disciplinary scholars' problems.

3. Proposed Methodology

Centered on the viewpoint of paper–citation relationships, our objective is to recognize the latent relationships or associations among the research papers of uncertain systems. A candidate research paper is valid to be considered in [3] only if it cites some of the credentials of the specific target paper. A candidate paper is eligible for inclusion in our proposed approach if it mentions any certificates or references to the particular target paper. Then, the degree of similarity among specific targets as well as candidate papers which qualify is determined and weighed. The most comparable top-N research papers are then proposed based on the presumption that some degree of correlation occurs between them if there is significant co-occurrence among the specific target paper as well as all eligible candidate papers.

The creation of candidates is the initial phase of a recommendation. The system produces a list of pertinent candidates in response to a query. Two popular methods for candidate generation are shown in Table 1, as follows.

Table 1. Different filtering techniques.

Type	Definition
content-based filtering	Uses similarities between products to suggest products comparable to the user’s preferences.
collaborative filtering	Provides recommendations by concurrently utilizing similarities between the products and the query.

Each item and each query (or context) is mapped to an embedding vector in a shared embedding space, $E = R^d$, by content-based and collaborative filtering, respectively. The embedding space typically captures some latent structure of the item or query set and is low-dimensional (i.e., substantially less than the size of the corpus). Similar objects are grouped together in the embedding area, such as YouTube videos that the same individual typically watches. A similarity measure establishes the concept of “closeness”.

A function $s : E * E \rightarrow R$ called a “similarity measure” takes two embeddings and outputs a scalar measuring how similar the two are. The embeddings can be used for candidate generation as follows: given a query embedding $q \in E$, the system looks for item embeddings $x \in E$ that are close to q , that is, embeddings with high similarity $s(q, x)$. Most recommendation algorithms use cosine, dot, or Euclidean distance to calculate the degree of similarity. The angle between the two vectors is simply the cosine of this angle, as shown in Equation (1).

$$s(q, x) = \cos(s, x) \tag{1}$$

The dot product of two vectors is given by Equation (2).

$$s(q, x) = \langle q, x \rangle = \sum_{i=1}^d q_i x_i \tag{2}$$

It is also provided by the cosine of the angle multiplied by the product of norms, as shown in Equation (3).

$$s(q, x) = \|x\| \|q\| \cdot \cos(q, x) \tag{3}$$

Thus, the dot product and cosine coincide if the embeddings are normalized. Hence, Euclidian distance is the usual distance in Euclidean space, as shown in Equation (4).

$$s(q, x) = \|q - x\| = \left[\sum_{i=1}^d (q_i - x_i)^2 \right]^{\frac{1}{2}} \tag{4}$$

Higher similarity results from a lesser distance. Note that the squared Euclidean distance and the dot product (and cosine) coincide up to a constant when the embeddings are normalized, as shown in Equation (5).

$$\frac{1}{2} \|q - x\|^2 = 1 - \langle q, x \rangle \quad (5)$$

The dot product similarity is more sensitive to the embedding norm than the cosine is. In other words, for products with an acute angle, the higher the resemblance, and the more likely it is that the item will be recommended. Items that appear in the training set quite frequently (for instance, well-known YouTube videos) typically contain embedding with high norms. If gathering information about popularity is essential, the dot product is the better option. If not careful, the most well-liked things could take over the recommendations. Rarely updated items might not be updated regularly during training. As a result, if they have a high initial norm, the algorithm might favor rare things over more pertinent ones. Using the proper regularization with caution when embedding initialization is necessary to prevent this issue. Other similarity measure variations that place less emphasis on the item's norm can be used in real life, as shown in Equation (6).

$$s(q, x) = \|q\|^\alpha \|x\|^\alpha \cos(q, x) \text{ for some } \alpha \in (0, 1) \quad (6)$$

The idea that individuals with similar preferences would rate items equally is used through collaborative filtering. Content-based filtering uses the hypothesis that goes with identical objective characteristics that would be similarly scored. For instance, if you like words such as "tomato sauce" on a web page, you would also like another "tomato sauce" web page. The difficulty of the issue is appropriately extracting the characteristics of the most predictive products. Then, to generate a profile of users of features from the items rated by the user, the user's profile is compared with new profiles of the entities whose characteristics are extracted [7]. The inappropriateness of the collaborative recommended approach or technique to any research paper is expressed in the researchers' deficiency of research paper ratings [13]. We have mined the rating scores of the researchers as well as research papers based upon the citation relationships of the paper to provide a solution to this issue.

Let P_{ij} be the citation count of research papers, and j be the cited paper from matrix "P", called the citation matrix. When a research paper "i" is cited by another paper "j", then $P_{ij} = 1$; otherwise, $P_{ij} = 0$.

The methodology of our proposed approach RPRSCA starts by first converting all the recommended papers of our database into a matrix of paper-citation relationships where the rows and columns, respectively, represent the recommended papers and their citations. Our approach was aimed at dealing with uncertain systems with the following scenarios, in which:

- A researcher who, after some initial study, finds an interesting article and wants to have more associated quality papers comparable to that.
- A reviewer is interested in exploring further based on a paper that discusses a topic in which he may not be a specialist.
- A student has obtained a paper from his supervisor to study the subject area covered by that paper.
- Researchers who would like to study more from their previous publications.

In all situations, we will consider that the sources, as well as processed paper citations showing the user's interests, are available to the public. The RPRSCA is shown in Algorithm 1.

Algorithm 1. RPRSCA

Input required: Research Target Paper

Output Expected: Top-N Recommendations

Step_1: Let Research Target Paper Query be as (RPi), then:

(a) Retrieve Target paper(RPi)'s references(Refi) and Citations (CoCi).

(b) For each reference, extract all other papers (OtherRefi) that are also cited by any of the Target Paper (RPi)'s references (Refi).

(c) Furthermore, extract all the references to the target paper's citations, and these are referred to as nearest neighbors of the target paper (RPi).

Step_2: For every neighbouring paper, qualified candidate papers are co-cited with RPi and have also been referenced by a minimum of one of the target papers' references.

Step_3: Then, by using Jaccard's similarity measure 'J', the degree of similitude between the qualified candidate papers as well as the target paper (RPi) is measured.

Step_4: Top-N quality research papers are recommended to the users.

Based on the defined research target, the Top-N method was applied. The mentioned artificial intelligence algorithm selects the references and citations of all the target papers. It removes all other articles from the site for each reference that also quoted some of the concerns from those target articles. In addition, for every citation of the target, it also extracts the remaining paper from the web, which cites a few of the target citations. Extracted papers were referred to as the nearest neighbors' target research papers. For any of the papers (neighboring), we then classify candidate papers, which are also co-cited. This is carried out along with the targeted article and has been referenced via some citations in the target papers. The degree of similarity between these eligible candidate papers as well as the target is then determined by collaborative similarity measuring using the similarity measure provided by Jaccard, which was mentioned in Equation (7). Then, we can suggest to the researcher the most similar top-N articles. In addition, to calculate the degree of similarity between the target and each of the eligible candidate articles, Jaccard's similarity is also used to measure the deviations. The Jaccard similarity coefficient J is given as

$$J = W^{RPi \rightarrow Pc} = \frac{Y_{11}}{Y_{01} + Y_{10} + Y_{11}} \quad (7)$$

where

A and B are research papers whose attributes are 0 or 1. $RPi \rightarrow$ target paper

$Pc \rightarrow$ All candidate papers which contain the target references as well $Y_{11} \rightarrow$ Total No. of attributes A = 1 and B = 1

$Y_{01} \rightarrow$ Total No of attributes A = 0 and B = 1 $Y_{10} \rightarrow$ Total No of attributes A = 1 and B = 0.

Consider the scenario shown in Figure 1. RPi represents the target paper, Ref1 to Refn are the references, and CoC1 to CoCn are the citations. Each of the concerns of RPi has other sources from any of OtherRef1 to OtherRefn or /and from CoC1 to CoCn other than RPi. Our method qualifies papers (OtherRef1 to OtherRefn) that have been referenced by at least one of the target papers' references and are co-cited with RPi. CoC1 cites OtherRef1 and OtherRef2 with the target publication. OtherRef2, on the other hand, has no ties to any of the target paper's references and is, thus, disqualified by our suggested algorithm. OtherRef1, on the other hand, is not only co-cited with the target paper by CoC1 but is also referenced by one of the target publications. Only OtherRef1 and OtherREF3 have qualified candidate papers, as shown in Figure 1.

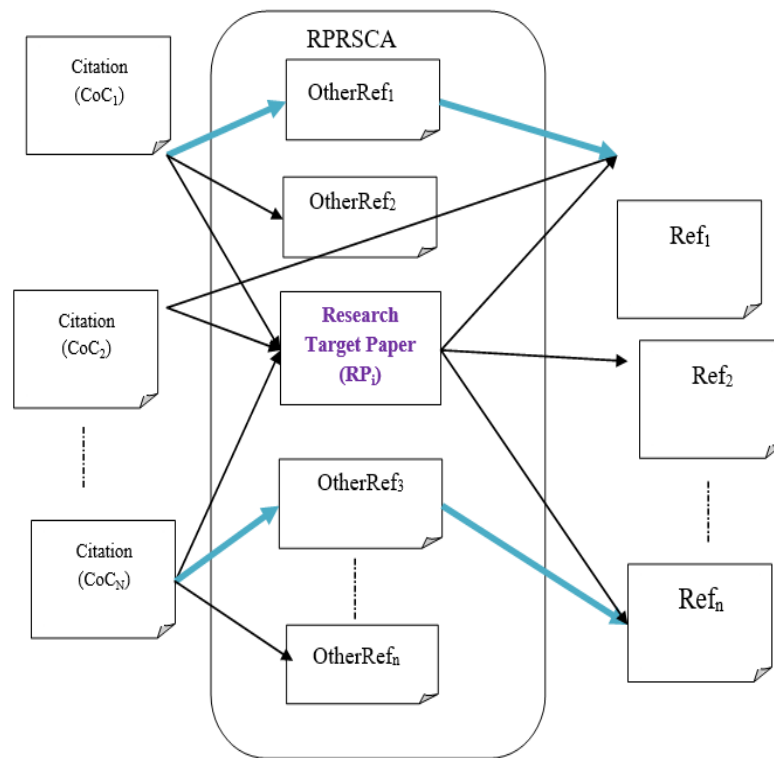


Figure 1. Proposed Scenario of Recommendation.

3.1. Experimental Setup

We used the dataset provided in [2], which is publicly accessible. The dataset included a list of researchers (around 50) whose interests in research range across several domains. Their references, as well as citations, were collected and extracted via Google Scholar. Some statistics of the utilized dataset are presented in Table 2.

Table 2. Statistics of the dataset.

Total Number of Researchers	50
The average number of researchers' publications	10
The average number of citations of each researcher's publications	14.8 (max. 169)
The average number of references to each researcher's publications	15.0 (max. 58)
Total number of recommending papers	100,351
The average number of citations of the recommended papers	17.9 (max. 175)
The average number of references to the recommended papers	15.5 (max. 53)

3.2. Metrics for Evaluation

To determine our methodology's accuracy, we validated it by choosing 20% as a test sample. By using the three most widely used assessment metrics in information retrieval systems, we tested the general performance: accuracy, recall, and F1. Precision is a metric that measures how many correct optimistic forecasts have been made. As a result, precision estimates the accuracy of the minority class. The precision shown in Equation (8) computes the system's ability to claim back as many applicable papers as possible in response to the target paper.

$$precision = \frac{\Sigma(relevant_papers) \cap \Sigma(received_papers)}{\Sigma(retrieved_papers)} \quad (8)$$

Recall is a metric that measures how many correct optimistic predictions were produced out of all possible positive predictions. Unlike precision, which only considers

the accurate positive predictions out of all positive predictions, recall weighs the positive predictions that were missed. The recall given by Equation (9) tests the system's ability to reclaim or claim back as few irrelevant papers as possible in response to the target paper.

$$recall = \frac{\Sigma(relevant_papers) \cap \Sigma(retrieved_papers)}{\Sigma(relevant_papers)} \quad (9)$$

F1-score is a better measure than accuracy, since it is the harmonic mean of precision and recall. The harmonic mean is given by Equation (10).

$$F1 = \frac{2 * precision * recall}{precision + recall} \quad (10)$$

Average accuracy (AP) is the average of the accuracy values of all scores of related papers, and the average of all APs is the mean average precision (MAP) given by the Equation mentioned in (11).

$$MAP = \frac{1}{I} \sum_{i \in I} * \frac{1}{ni} \sum_{k=1}^N * P(R_{ik}) \quad (11)$$

$P(R_{ik})$ indicates precision of returned papers; recommendation list length is represented by N ; the total number of related papers in the recommendations list is represented by ni , and the collection of papers by I .

Equation (12) gives the mean reciprocal rank (MRR), the ranking level at which the system has returned the first relevant research paper averaged across all researchers. It assesses the system's ability to produce an appropriate research paper at the top of the list of recommendations.

$$MRR = \frac{1}{N_p} \sum_{i \in I} * \frac{1}{rank(i)} \quad (12)$$

$Rank(i)$ indicates the highest rank where i is the first related paper; N_p denotes the sum of target papers.

3.3. Results and Discussions

This section discusses the results of different techniques against the proposed one with respect to several performance metrics.

3.3.1. Results

To be precise, the total averages of all 50 researchers in our dataset reflect each metric evaluation's results. We begin by comparing the assessment of the available output of the proposed method. Methods are focused on the three most widely used measurement criteria for information retrieval. Table 3 depicts the values of recall, precision, and F1-score of three approaches: co-citation; context-based collaborative filtering; and the proposed method, RPRSCA.

The graphical comparisons, which are based on performance, recall, and assessment measure F1, are shown in Figures 2–4, respectively. The accurate outcomes of our proposed model, RPRSCA, have outperformed the other methods, such as context-based collaborative filtering (ContextCF) and the co-citation process, in terms of retrieving the appropriate methods, as seen in Figure 2. Our strategy will objectively exclude recommended papers less connected to the target article.

Figure 2 shows the comparison based on the recall. The output gap between our proposed model and the ContextCF is not very significant, as can be observed in the figure. Sometimes, the ContextCF approach is marginally equal to or better than our proposed technique at specific points. However, our suggested solution began to demonstrate the critical difference between them when N , the number of research papers, exceeded 20.

Table 3. Values of performance metrics.

No of Papers→		5	10	15	20	25	30
Precision	RPRSCA	0.5	0.55	0.6	0.55	0.56	0.56
	ContextCF	0.3	0.4	0.35	0.3	0.35	0.37
	Co-Citation	0.35	0.33	0.32	0.3	0.27	0.25
Recall	RPRSCA	0.03	0.07	0.12	0.13	0.14	0.15
	ContextCF	0.04	0.06	0.11	0.12	0.13	0.14
	Co-Citation	0.02	0.03	0.04	0.05	0.06	0.07
F1-Score	RPRSCA	0.04	0.13	0.16	0.17	0.19	0.24
	ContextCF	0.05	0.12	0.15	0.16	0.17	0.19
	Co-Citation	0.03	0.08	0.1	0.12	0.14	0.17

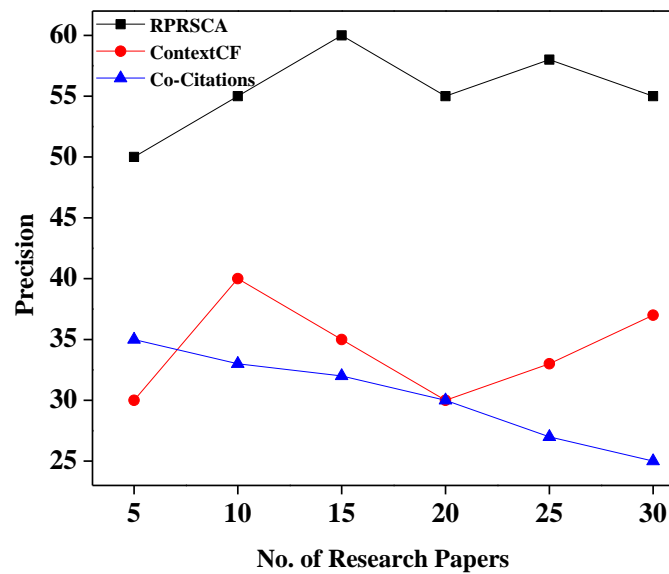


Figure 2. Performance of precision.

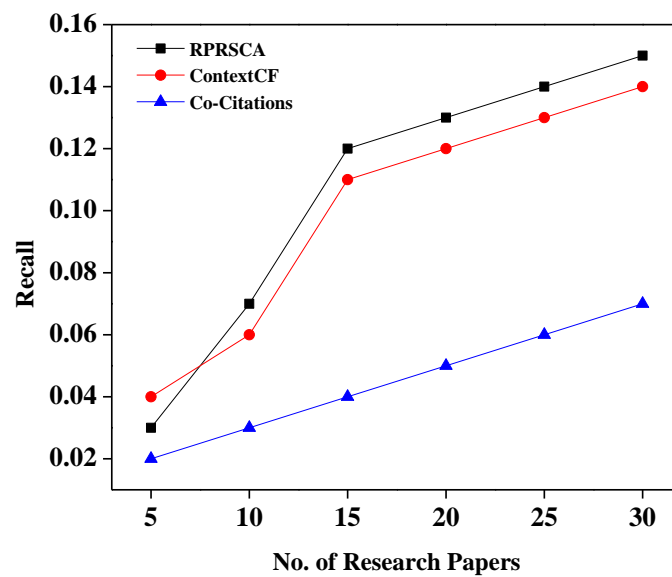


Figure 3. Performance of recall.

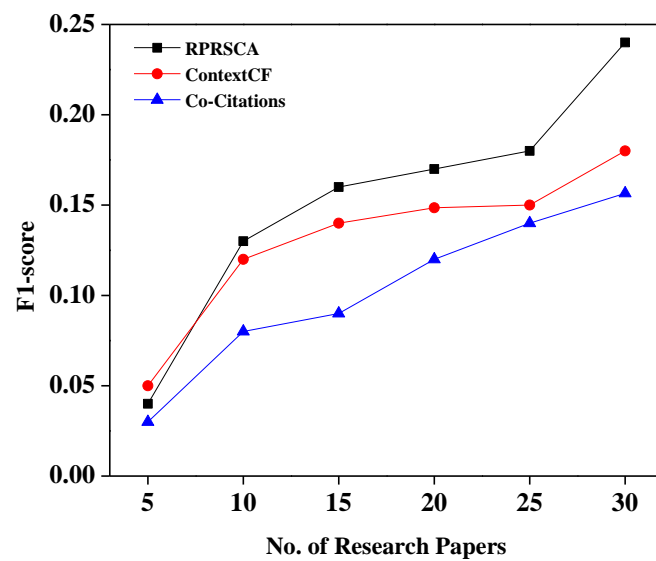


Figure 4. Performance of F1-score.

The harmonic mean among precision, as well as recall (F1 measurement), is depicted in Figure 4; also, it can be observed from the graph that the difference in output between the proposed technique, RPRSCA, and the ContextCF approach is also negligible for N values less than or equal to 20. The proposed approach, however, showed an outstanding improvement compared to ContextCF, especially when the value of N exceeded 20.

3.3.2. Discussions

Table 4 compares all three algorithms, concerning their precision, recall, and F1 measure when implemented on a data set of 50 research papers.

Table 4. Comparison of the algorithms concerning performance metrics.

Performance Metrics	Co-Citation	Context-Based Collaborative-Filtering (CCF) Approach	RPRSCA
Precision	Performance is very low compared to other approaches	Performance is better compared to co-citation but lower when compared to the proposed approach	Performance is better compared to the CCF approach
Recall	Performance is very low compared to other approaches	It is slightly better when N (No of research papers) is 5, but when N increases, there is a slight reduction in the Performance observed	The performance increases when N is above 20 because the proposed method only recalls the most significant
F1 measure	Performance is very low compared to other approaches	It is insignificant when the value of N is below 20; however, when N increases from 20, the performance slightly increases compared to the proposed method	The proposed method is insignificant when the value of N is below 20; however, when N increases from 20, the performance is better compared to CCF

The average precision ranges from 0.5 for balanced data to 1.0 for positive examples (perfect model). As mentioned earlier, the mean average accuracy (MAP), often known as AP, is a widely used metric for assessing the performance of models performing document/information retrieval and object detection tasks. MAP considers whether all of the relevant items tend to be ranked highly. The abbreviation MRR stands for mean reciprocal rank. It is also referred to as the average reciprocal hit ratio (ARHR). MRR is solely interested in the highest-ranking relevant item. MRR is involved if the system returns a pertinent article in the third-highest position. It makes no difference if the other relevant

items are ranked 4th or 10th. The MAP and MRR values of RPRSCA, co-citation, and context-based collaborative filtering are depicted in Table 5.

Table 5. MAP and MRR comparison with the proposed approach.

No. of Papers→		5	10	15	20	25	30
MAP	Proposed Method	0.5	0.6	0.58	0.57	0.56	0.55
	ContextCF	0.47	0.4	0.45	0.44	0.43	0.42
	Co-Citation	0.35	0.38	0.38	0.37	0.37	0.36
MRR	Proposed Method	0.73	0.75	0.8	0.8	0.8	0.8
	ContextCF	0.7	0.72	0.72	0.72	0.72	0.72
	Co-Citation	0.5	0.48	0.48	0.48	0.48	0.48

As depicted in Figure 5, the proposed technique, RPRSCA, has significantly exceeded the other methods in all cases based on the Mean Average Precision (MAP), while also returning the related recommendations to the top of the list. In addition, as shown in Figure 6, the comparison based on mean reciprocal rank (MRR) has demonstrated that our proposed approach, RPRSCA, shows exceptional results compared to the baseline methods in all scenarios. As previously stated, all of the improvements are hugely attributable to the strictness with which candidate articles were qualified and the removal of less relevant studies from the target paper. As a result, the system's ability to return relevant and valuable recommendations near the top of the list improves.

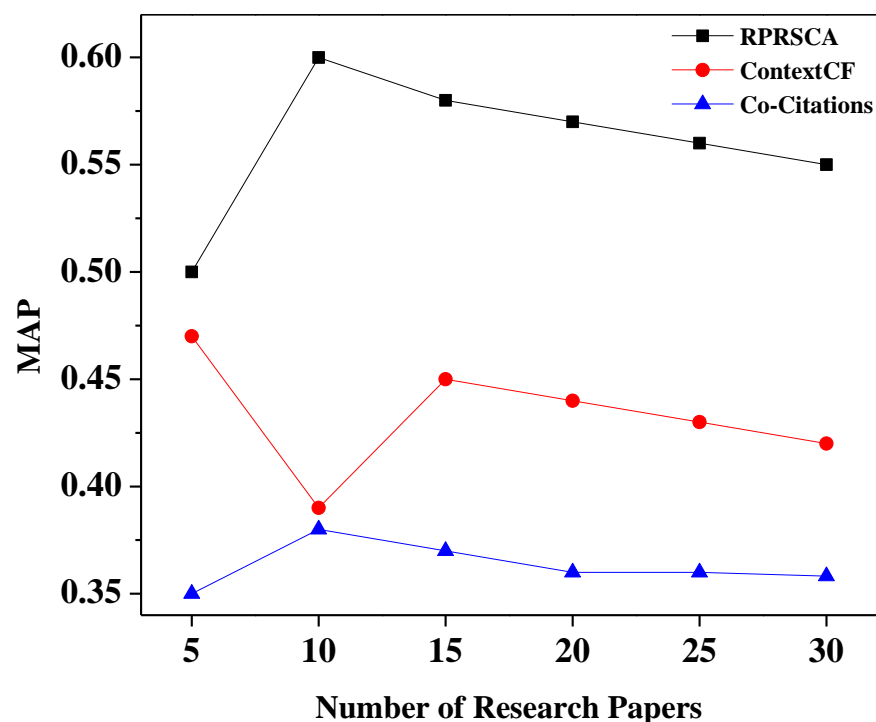


Figure 5. Performance of MAP (mean average precision).

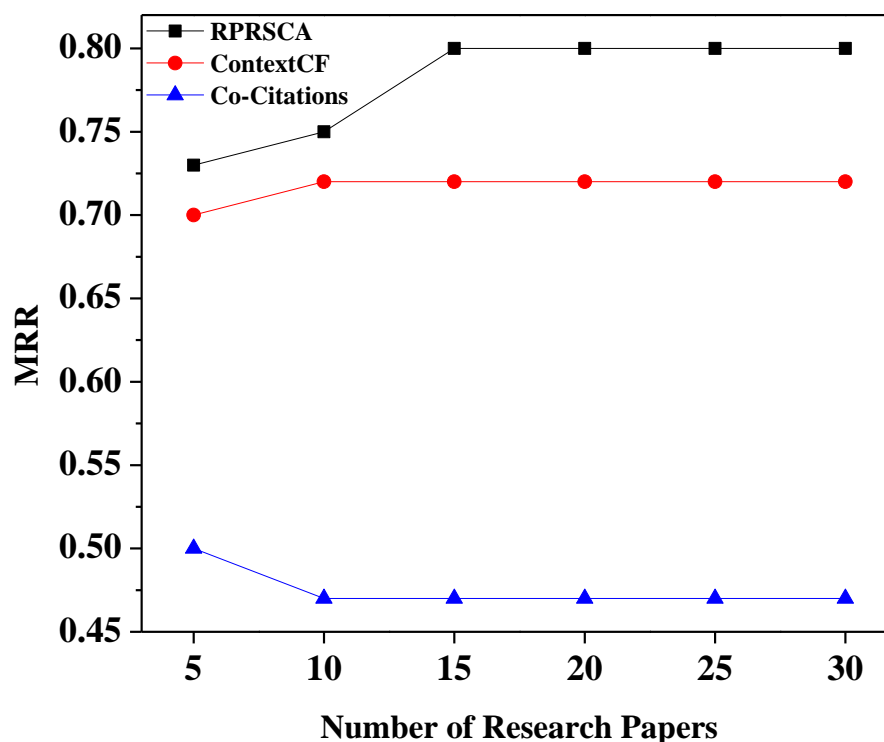


Figure 6. Performance of MRR (mean reciprocal rank).

4. Conclusions

In this paper, for the purpose of recommending a collection of quality, related papers to an intended researcher depending upon paper–citation relationships, we have used publicly accessible contextual metadata to exploit the benefits of a collaborative filtering method. Using paper–citation relations, the RPRSCA technique influenced the latent linkages between a research paper, its references, and citations of uncertain systems. The logic behind the strategy is elementary: if two or more papers co-occur substantially with same referencing paper(s), they must be identical to some degree. Our suggested methodology, RPRSCA, has considerably improved the three most widely used metrics of information retrieval systems: accuracy, recall, and F1 measurement. Our suggested process for presenting appropriate and valuable recommendations is based upon mean average precision (MAP) and mean reciprocal rank (MRR), and we reported substantial improvements over the baseline methods. The suggested system focuses on the customers’ or users’ interests and offers complete satisfaction when searching for a specific research paper. For all values of N (the total number of research papers), the overall accomplishments of our suggested technique outperformed the existing baseline methods in terms of precision, recall, and F1 measure. In our following line of research, aside from examining collaborative relationships among quality research articles, we will also consider public contextual information, such as paper titles and abstracts, to improve the model’s performance.

Author Contributions: Conceptualization, V.K.V. and M.T.R.; methodology, M.T.R.; software, M.T.R. and A.B. (Anatoliy Batyuk); validation, V.K.V. and A.B. (Andrii Barna); formal analysis, A.B. (Anatoliy Batyuk); investigation, M.T.R. and A.B. (Andrii Barna); resources, A.B. (Andrii Barna); data curation, B.H.; writing—original draft preparation, M.T.R. and V.K.V.; writing—review and editing, M.T.R. and B.H.; visualization, V.K.V. and M.T.R.; supervision, B.H.; project administration, V.K.V.; funding acquisition, B.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset used for the findings is included in the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haruna, K.; Ismail, M.A. An Ontological Framework for Research Paper recommendation. *Int. J. Soft Comput.* **2016**, *11*, 96–99. [CrossRef]
2. Sugiyama, K.; Kan, M.-Y. A comprehensive evaluation of scholarly paper recommendation using potential citation papers. *Int. J. Digit. Libr.* **2015**, *16*, 91–109. [CrossRef]
3. Liu, H.; Kong, X.; Bai, X.; Wang, W.; Bekele, T.M.; Xia, F. Context-Based Collaborative Filtering for Citation Recommendation. *IEEE Access* **2015**, *3*, 1695–1703. [CrossRef]
4. Roopashree, S.; Anitha, J.; Mahesh, T.; Kumar, V.V.; Viriyasitavat, W.; Kaur, A. An IoT based authentication system for therapeutic herbs measured by local descriptors using machine learning approach. *Measurement* **2022**, *200*, 111484. [CrossRef]
5. Beel, J.; Langer, S.; Genzmehr, M.; Nürnberger, A. Introducing Docear’s research paper recommender system. In Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, Indianapolis, IN, USA, 22–26 July 2013; pp. 459–460. [CrossRef]
6. Ramakrishna, M.T.; Venkatesan, V.K.; Izonin, I.; Havryliuk, M.; Bhat, C.R. Homogeneous Adaboost Ensemble Machine Learning Algorithms with Reduced Entropy on Balanced Data. *Entropy* **2023**, *25*, 245. [CrossRef]
7. McNee, S.M.; Albert, I.; Cosley, D.; Gopalkrishnan, P.; Lam, S.K.; Rashid, A.M.; Konstan, J.A.; Riedl, J. On the recommending of citations for research papers. In Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, New Orleans, LA, USA, 16–20 November 2002; pp. 116–125. [CrossRef]
8. Vivek, V.; Mahesh, T.R.; Saravanan, C.; Kumar, K.V. *A Novel Technique for User Decision Prediction and Assistance Using Machine Learning and NLP: A Mod-el to Transform the E-commerce System*. In *Big Data Management in Sensing*; River Publishers: Gistrup, Denmark, 2022; pp. 61–76.
9. Nascimento, C.; Laender, A.H.; da Silva, A.S.; Gonçalves, M.A. A source independent framework for research paper recommendation. In Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries, Ottawa, ON, Canada, 13–17 June 2011; pp. 297–306. [CrossRef]
10. Richa; Bedi, P. Trust and Distrust based Cross-domain Recommender System. *Appl. Artif. Intell.* **2021**, *35*, 326–351. [CrossRef]
11. Haruna, K.; Ismail, M.A.; Damiasih, D.; Sutopo, J.; Herawan, T. A collaborative approach for research paper recommender system. *PLoS ONE* **2017**, *12*, e0184516. [CrossRef] [PubMed]
12. Singh, P.K.; Pramanik, P.K.D.; Dey, A.K.; Choudhury, P. Recommender Systems: An Over-view, Research Trends, and Future Directions. *International J. Bus. Syst. Res.* **2021**, *15*, 14–52. [CrossRef]
13. Sakib, N.; Ahmad, R.B.; Haruna, K. A Collaborative Approach Toward Scientific Paper Recommendation Using Citation Context. *IEEE Access* **2020**, *8*, 51246–51255. [CrossRef]
14. Pawlicka, A.; Pawlicki, M.; Kozik, R.; Choraś, R. A Systematic Review of Recommender Systems and Their Applications in Cybersecurity. *Sensors* **2021**, *21*, 5248. [CrossRef] [PubMed]
15. Mahesh, T.R.; Kumar, V.V.; Vivek, V.; Raghunath, K.M.K.; Madhuri, G.S. Early predictive model for breast cancer classification using blended ensemble learning. *Int. J. Syst. Assur. Eng. Manag.* **2022**, *13*, 1–10. [CrossRef]
16. Li, L.; Zhang, Z.; Zhang, S. Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation. *Sci. Program.* **2021**, *2021*, 7427409. [CrossRef]
17. Devarajan, D.; Alex, D.S.; Mahesh, T.R.; Kumar, V.V.; Aluvalu, R.; Maheswari, V.U.; Shitharth, S. Cervical Cancer Diagnosis Using Intelligent Living Behavior of Artificial Jellyfish Optimized With Artificial Neural Network. *IEEE Access* **2022**, *10*, 126957–126968. [CrossRef]
18. Mahesh, T.R.; Kumar, V.V.; Muthukumaran, V.; Shashikala, H.K.; Swapna, B.; Guluwadi, S. Performance Analysis of XGBoost Ensemble Methods for Survivability with the Classification of Breast Cancer. *J. Sensors* **2022**, *2022*, 1–8. [CrossRef]
19. Shao, B.; Li, X.; Bian, G. A survey of research hotspots and frontier trends of recommendation systems from the perspective of knowledge graph. *Expert Syst. Appl.* **2021**, *165*, 113764. [CrossRef]
20. Thomas, B.; John, A.K. Machine Learning Techniques for Recommender Systems—A Comparative Case Analysis. *IOP Conf. Series: Mater. Sci. Eng.* **2021**, *1085*, 012011. [CrossRef]
21. Beltagy, I.; Lo, K.; Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. *arXiv* **2019**, arXiv:1903.10676.
22. Osman, N.A.; Noah, S.A.M.; Darwich, M.; Mohd, M. Integrating contextual sentiment analysis in collaborative recommender systems. *PLoS ONE* **2021**, *16*, e0248695. [CrossRef] [PubMed]
23. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv* **2019**, arXiv:1910.01108.
24. Indraneel, A.; Sai Pranav, K.; Mamatha, H.R. Hybrid Recommendation System for Scientific Literature. Lecture Notes on Data Engineering and Communications Technologies. 2021, pp. 725–735. Available online: <https://www.springerprofessional.de/en/hybrid-recommendation-system-for-scientific-literature/18874338> (accessed on 30 January 2023). [CrossRef]
25. Ghosh, P.; Mahesh, T.R. A privacy preserving mutual authentication protocol for RFID based automated toll collection system. In Proceedings of the 2016 International Conference on ICT in Business Industry & Government (ICTBIG), Indore, India, 18–19 November 2016; pp. 1–5. [CrossRef]

26. Haruna, K.; Ismail, M.A.; Bichi, A.B.; Chang, V.; Wibawa, S.; Herawan, T. A Citation-Based Recommender System for Scholarly Paper Recommendation. In Proceedings of the International Conference on Computational Science and Its Applications—ICSA 2018, Melbourne, VIC, Australia, 2–5 July 2018; pp. 514–525. [[CrossRef](#)]
27. Kim, J.; Lee, J.-H. A novel recommendation approach based on chronological cohesive units in content consuming logs. *Inf. Sci.* **2019**, *470*, 141–155. [[CrossRef](#)]
28. Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; Tan, T. Session-Based Recommendation with Graph Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 346–353. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.