Using an explicit query and a topic model for scientific article recommendation

Boussaadi Smail[1] [ORCID] **· Hassina Aliane[2] · Ouahabi Abdeldjalil[3]**

Abstract

The search for relevant scientific articles is a crucial step in any research project. However, the vast number of articles published and available online in digital databases (Google Scholar, Semantic Scholar, etc.) can make this task tedious and negatively impact a researcher's productivity. This article proposes a new method of recommending scientific articles that takes advantage of content-based filtering. The challenge is to target relevant information that meets a researcher's needs, regardless of their research domain. Our recommendation method is based on semantic exploration using latent factors. Our goal is to achieve an optimal topic model that will serve as the basis for the recommendation process. Our experiences confirm our performance expectations, showing relevance and objectivity in the results.

Keywords Topic modeling · Latent Dirichlet Allocation · Non-negative Matrix Factorization · Scientific recommendation · Scientific article

1 Introduction

Recommender systems (RSs) of scientific articles are applications designed with modern engineering methods and artificial intelligence algorithms. Their role is to filter enormous scientific publications in online scientific databases (such

---

✉ Boussaadi Smail
sboussaadi@gmail.com

Hassina Aliane
haliane@hotmail.com

Ouahabi Abdeldjalil
ouahabi@univ-tours.fr

[1] DTISI, Research Center on Scientific and Technical Information Cerist, Algiers, Algeria

[2] Director of Information Sciences R&D Laboratory Head of Natural Language Processing and Digital Content Team Cerist, Algiers, Algeria

[3] 1.Polytech Tours, Imaging anBrain, University of Tours, INSERM U930 Tours, France

as Google Scholar and Semantic Scholar). Indeed, one of the obstacles that researchers face is the problem of cognitive overload, and scientific recommendation systems effectively solve this problem Adomavicius and Tuzhilin (2005). In the literature, we distinguish four approaches for recommending scientific articles. The classic approaches are based on collaborative filtering (CF) Adomavicius and Tuzhilin (2005) and content-based filtering (CBF) Adomavicius and Tuzhilin (2005). Recently appeared approaches based on graphs (Bai et al., 2020) and approaches based on topic modeling (Beel et al., 2016). Hybrid approaches combine two or more techniques from previous approaches to improve the quality of recommendations (Sakib et al., 2022). Content-based RS is similar to machine learning classifiers, the underlying logic relies on the knowledge of the researcher's interactions, which causes the cold start problem and the overspecialization problem that prevents the discovery of new publications (Bai et al., 2020). The RS based on collaborative or social filtering relies on social opinions. Although they can theoretically adapt to any domain, they suffer from the double problem of cold start (new researcher / new article). Indeed, the recommendation process cannot identify similar researchers for a new researcher without a history of evaluations or very few ratings; the same problem occurs with articles (Sakib et al., 2022). Graph-based approaches mainly focus on building a graph to model a citation network (Citation Graph) or a social network; nodes and links, respectively, model researchers and articles. A major limitation of this approach is that the recommendation process does not consider the articles' content and the researchers' topic interests (Bai et al., 2020).

Topic model-based RS use latent factors to model the topic interests of researchers or identify communities of researchers in the recommendation process (Amami et al., 2016).

In this article, we propose a new method that combines latent factors and content-based filtering, to minimize the limitations of the first-mentioned approaches, by translating the explicit requirements of the researcher into topics. Combining latent factors with content-based filtering can reduce the amount of data needed to make accurate recommendations. While content-based filtering typically requires a large amount of data to provide accurate recommendations, the use of latent factors can help reduce this requirement. As a result, our method allows for more efficient and accurate recommendations with fewer data.

Our work aims to improve suggestion accuracy and relevance, helping researchers stay up-to-date on the latest work in their domains and discover pertinent new research that they might not have otherwise encountered. The proposed method's key contributions are summarized as follows:

- Firstly, the proposed method does not require an a priori user profile.
- Secondly, this study evaluates and compares the performance of two popular topic modeling algorithms, Latent Dirichlet Allocation (LDA)(Blei et al., 2000) and Non-Negative Matrix Factorization (NMF)(Lee & Seung, 1998), which can assist researchers in determining the optimal technique for a specific task in topic modeling.

2 Related work

This section provides a brief review of the scientific literature related to the field of topic model-based scientific article recommendations. The most commonly utilized topic model algorithms in the design of scientific article recommendation systems are LDA. Wang & Blei (2011) introduced the Collaborative Topic Regression (CTR) recommendation model, which integrates matrix factorization and LDA into a single generative process. In the CTR model, latent factors of articles are obtained by adding an offset latent variable to the topic proportion vector. In another study, Amami et al. (2016) proposed a hybrid approach that combines latent factors generated by LDA and relevance-based language modeling. The study utilized the clustering of searchers based on latent topics of interest as reliable sources to produce recommendations. However, this approach suffered from consistency issues and the interpretability of inferred topics.In contrast, Rossetti et al. (2017) proposed a topic model based on researcher interactions, rather than article content, to extract consistent and interpretable topics.

Recent research has incorporated topic modeling techniques into content-based filtering and collaborative filtering to develop recommender systems. For instance, Subathra and Kumar (2019) proposed a recommendation algorithm that utilizes a double iteration of LDA in combination with collaborative filtering to generate recommendations. Nonetheless, LDA is a non-deterministic algorithm, resulting in different topics being generated each time the algorithm is run. This can lead to confusion and a reduction in the efficiency of content exploration, leading to a decrease in the relevance of the recommended articles.

In another study, Boussaadi et al. (2020) integrated collaborative filtering and latent topics to detect community researchers based on the dominant topic in an academic social network. The proposed approach significantly enhances the relevance of recommendations compared to existing methods. However, the recommendation process relied on a single source of information to identify the topic interests of the researchers, which reduces the accuracy and interpretability of the researchers' profiles.

Bagul and Barve (2021) proposed a method for scientific literature recommendations that combines content-based filtering with LDA. For similarity computation, the authors used Jensen-Shannon distance (JSD), a static metric that cannot establish the semantic link between probability vectors.

In their most recent study, Hadhiatma et al. (2023) introduced a novel approach to scientific paper recommendation that combines community detection and a modified PageRank algorithm. Specifically, the authors utilized Latent Dirichlet Allocation to identify the main topics in scientific papers and to form multi-topic communities of papers. Within each community, the PageRank algorithm was employed to rank the papers based on their relevance to the research topic. Although PageRank has been widely used for ranking web pages, its efficacy in ranking scientific papers is not yet fully established. Furthermore, the proposed modification may not always be optimal or appropriate for ranking scientific papers, particularly in dynamic and rapidly evolving research domains.

The major challenge with previous work is that the inferred topics suffer from consistency and interpretability issues, limiting the precision in identifying the researcher's thematic interests and reducing the quality of the recommendations. This paper

proposes a new method based on explicit query formalism combined with latent topics. The proposed method does not suffer from the dual researcher-article cold start problem or the overspecialization problem.

2.1 Problematic of topic modeling technique

The major challenge and ambiguity involved in topic modeling is model validation. Both modeling techniques, LDA and NMF, have been the subject of several research studies to demonstrate the performance of each approach. In (Stevens et al., 2012), the authors demonstrated that NMF obtains better results in the classification task, but LDA produces topics with better coherence. According to the study's authors, it is more likely to work well for humans because of the stable coherence of LDA.

The authors (Berhil et al., 2020) used a corpus of 13,000 citations on the subject of COVID-19 to compare the consistency of the two techniques and found that, for applications in which a human end-user will interact with learned topics, the flexibility of LDA and the coherence advantages of LDA warrant strong consideration. They conclude that the LDA model is more relevant than the NMF model in the case of the large corpus. In (Jooneghani, 2021), the authors explored human concerns towards the COVID-19 vaccine using Twitter data and concluded that NMF performed better than LDA in the coherence score.

LDA tends to produce more coherent and human-interpretable topics than NMF. However, NMF tends to give better classification potential regardless of data size. Both techniques have compelling arguments depending on the context of their use. These arguments are a basis for comparing the results of the two techniques in our study and exploiting the better-performing algorithm for the recommender process.

3 Concept related to our research

3.1 Topic model (TM)

Topic models (TM) are an essential machine learning technology that used in different fields such as text analysis (Albalawi et al., 2020), recommendation systems (Almonte et al., 2022), image analysis (Jelodar et al., 2019), medical sciences (Leung and Khalvati 2022). There are many articles in this domain, and we definitely cannot mention all of them. TM is used to discover hidden topics in a text collection, such as documents, short texts, chats, Twitter and Facebook posts, blogs, and emails.

3.2 Topic coherence measures

The Coherence Score is a widely used performance metric to evaluate topic modeling methods. It gives a realistic way to figure out how many different topics there are in a document. Each generated topic has a list of words, such as a cluster of words. This measure looks at the average pairwise word similarity scores of the

words that are linked to the subject. The topic model with high Coherence Measure value is considered a suitable topic model.

Let's take a quick look at the most popular coherence measures and how they are calculated:

- C_v is a measure based on a sliding window, one-set segmentation of the top words, and an indirect confirmation measure that uses normalized pointwise mutual information (NPMI) and the cosine similarity.
- C_umass is based on document co-occurrence counts, a one-preceding segmentation, and a logarithmic conditional probability as confirmation measure.

3.3 Finding semantic similarity articles

The vectors representing the articles are probability vector representations of topic distributions. We select for a modified Jensen-Shannon divergence (JSD) measure (Levene, 2022). Indeed, when employed for similarity computation, JS Distance cannot determine the semantic link between subjects. To address this issue, (Guo et al., 2021) proposed a novel technique for assessing similarity that considers the semantic correlation of subjects from the perspective of word co-occurrence and augments the original JS measurement method by computing the semantic correlation of topic feature words.

Assume z_i is the topic of the article d_i, word set $W = \{w_{i1}, w_{i2}, ..., w_{in}\}$ is the feature word of the topic. According to the co-occurrence Eq. (1).

$$p(w_{im},\ w_{in}) = p(w_{im}|z_i) \times p(w_{in}|z_i) \tag{1}$$

where $p_{11}, p_{12}, \ldots, p_{nn}$ is the co-occurrence probability of feature word. The similarity computing equation of w_{im} and w_{jn} is as shown in Eq. (2).

$$Correlation(w_{im},\ w_{jn}) = \frac{p_{mn}}{p_{im} + p_{jn} - p_{mn}} \tag{2}$$

Hence, the semantic similarity between two articles d_i and d_j is shown in Eq. (3).

$$Similarity(d_i,d_j) = \lambda D_{js}(d_i,d_j) + (1-\lambda)\left[\frac{\sum_{m,n=1}^{v_d}\left(1 - correlation(w_{im}, w_{jn})\right)}{\left(v_d(v_d - 1)\right)}\right] \tag{3}$$

v_d denotes the number of feature words of this selected article. $\lambda \in [0,\ 1]$ denotes a correlation coefficient assigned to this article. D_{js} is the Jensen-Shannon distance.

4 The proposed approach

This section presents our method for scientific article recommendations to solve the limitations mentioned in Sect. 2. Our proposal includes first a selection step of the algorithm that generates the most coherent topic model, which will serve as a basis

for the recommendation process, the purpose of this step is a reference matrix of articles/topics; to achieve this goal, we use, the most popular topic modeling algorithms, namely LDA and NMF.

In the second step, the recommendation process performs a semantic similarity calculation to generate a list of relevant articles, which will present to the target researcher Fig. 1

4.1 Notation and approach

We denote by $A = \{A_1, A_2, \ldots, A_n\}$ the set of target researchers, and by A_i a generic researcher in A, and by $D = \{d_1, d_2, \ldots, d_m\}$, our corpus, that contains the articles that could be potentially interesting to our generic researcher.

The target article d_t is represented by a topic distribution associated with the predominant topic (obtained by applying the best-performing algorithm between LDA and NMF).

Our recommendation algorithm aims to present to a target researcher (A_i), a list of the most relevant and similar articles to the target article. The proposed approach has the following two steps.

Step 1:

- Application and evaluation of the LDA and NMF algorithms on the experimentation corpus (with different combinations of hyperparameters), the goal is to select the best performing algorithm, which we call algorithm_1.
- Referencing the recommendation corpus, by applying algorithm_1. each article will be represented by its predominant topic.
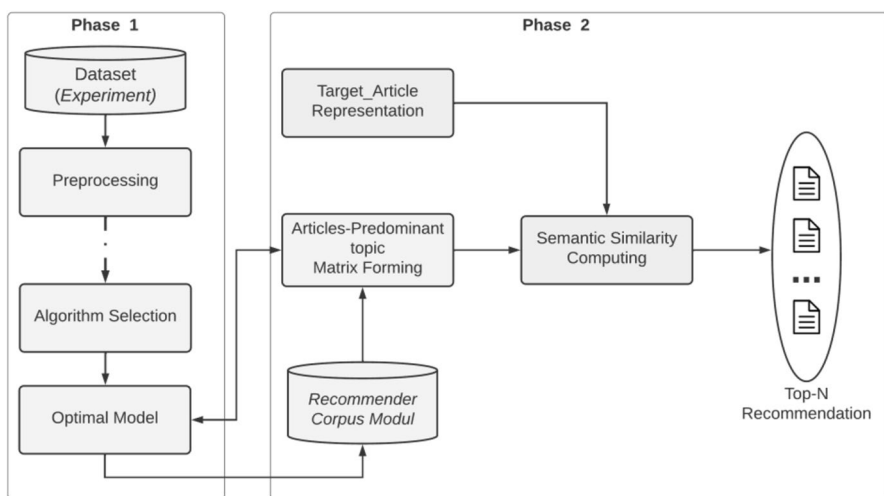

Fig. 1 Flowchart of proposed recommendation approach

Step 2:

We accept the researcher's query to identify the target article (designated d_j).

- Retrieval all the meta data set of the target article d_j (from google scholar to be precise) and we apply algorithm_1. it is assumed that the target article d_j is referenced by its predominant topic designated by topic_d.
- For each article $d_i (d_i \in D, d_i \neq d_j)$ and all articles referenced by topic_d do:

  Computing $Similarity(d_i, d_j)$ using Eq. (3).

- Ranking of the similar articles in descending order of relevance, and recommendation to the generic researcher A_i the top-ranked articles.

5 Experimental design

5.1 Experimental dataset and tools

We used the dataset of Elsevier (Kershaw & Koeling, 2020); This is a corpus of 40 k (40,091) open access (OA) CC-BY articles from across Elsevier's journals representing a large scale, cross-discipline set of research data to support NLP and ML research. The dataset contains research articles across various fields, including but not limited to, biology, chemistry, engineering, health sciences, mathematics, and physics. It includes articles published from 2014 to 2020 and is categorized in 27 Mid-Level ASJC (All Science Journal Classification) codes. ASJC codes represent the scientific discipline of the journals in which the article was published.

We split the dataset into two smaller datasets; the first is designated DM, and we use it for the selection of the best-performing algorithm. The second dataset is designated by DR; we use it for the recommendation process; the statistical details of our datasets are as follows in Table 1.

5.2 Data preprocessing

Text document preprocessing is an essential task for text exploring semantics. Data preprocessing is used to extract information into a structured format to analyze the patterns (hidden) within the data. The preprocessing steps are supported in Stanford's NLTK Library (Albalawi et al., 2020) and contain the following patterns:

Table 1 The statistics of the dataset

Dataset	Number of Articles	Vocabulary size	Training set size	Training set test
DM	1600	400k	70%	30%
DR	2400	675k	80%	20%

- Convert Text Data into Lower Case: Text datasets are converted into lower case for preventing the various words differences.
- Punctuation: punctuation such as (".", ",", "−", "!" etc.) are eliminated from the text datasets.
- Stop-word elimination: removal of the most common words in a language such as "and", "are", "this" etc., that is not helpful and in general unusable in text mining and that do not contain applicable information for the study.
- Stemming: the conversion of words into their root, using stemming algorithms such as Snowball Stemmer.
- Tokenization: Text datasets are converted into tokens (words). The tokenization identifies the meaningful keywords from the text data. The outcome of tokenization is a sequence of tokens.

5.3 Topic model algorithm

The NMF implementation was done by scikit-learn's NMF (version Optimized for the Frobenius Norm) functionality. For LDA, we use two different implementations as follows:

- The Gensim algorithm was implemented with the Gensim library, a Python library for topic modeling; the algorithm utilizes the inference technique online variational Bayes (Lee & Seung, 1998).
- The Mallet topic model package incorporates a rapid and highly scalable implementation of Gibbs sampling.

The reason to choose these different implementations was how they differ in inference technique. There is no standard best solution for making inferences since it heavily depends on the dataset. Since the inference method most suitable for the scientific text was unknown, both implementations were reasonable to evaluate. For both implementations, the output of the models is in the same format as in NMF.

5.4 NMF and LDA algorithm selection

Figure 2, shows the process applied to the DM dataset and evaluated by the quantitative metrics (Sect. 2.3). This procedure aimsure is to select the algorithm that meets our objective. Figures 3 and 4 clearly show that the LDA Mallet algorithm provides a better graphical quantitative evaluation for several subjects $k = 50$, compared to the other algorithms. for the rest of our work, we opt for the subject model produced by the LDA Mallet version.
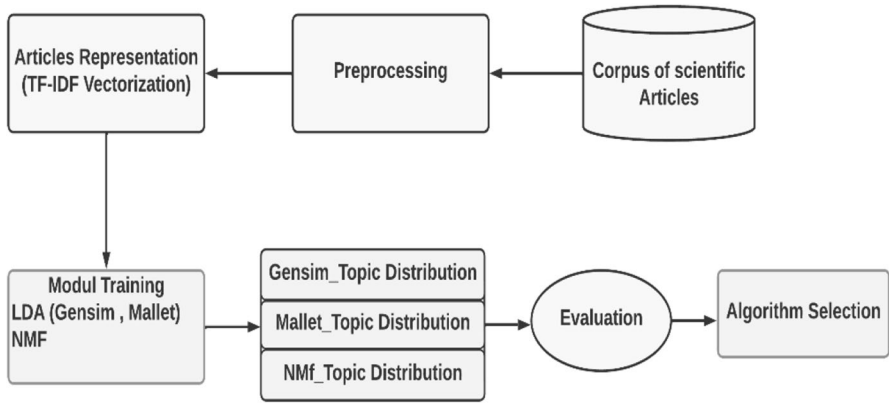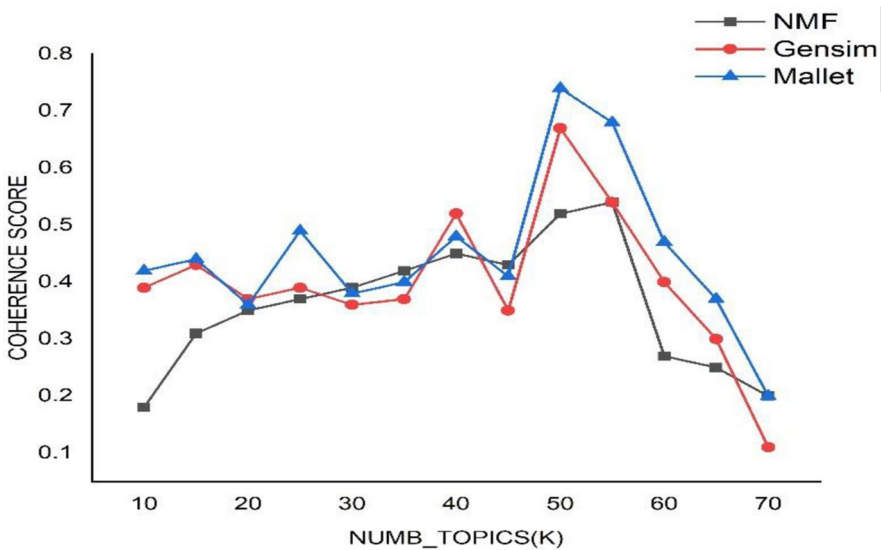
Fig. 2 Block Diagram for Algorithm Selection

Fig. 3 Quantitative (C_v score) results for LDA and NMF

5.5 Baseline methods

In assessing the effectiveness of our proposed method, we compare the recommendation results with two methods representing baseline CBF methods. In the first method (denoted by LDA_ASPR), the authors (Amami et al., 2016) have exploited the topics related to the researcher's scientific production (authored articles) to define her/his profile formally; in particular, they employed the topic modeling to represent the user profile formally, and language modeling to formally represent each unseen paper. The recommendation technique they proposed relies on assessing
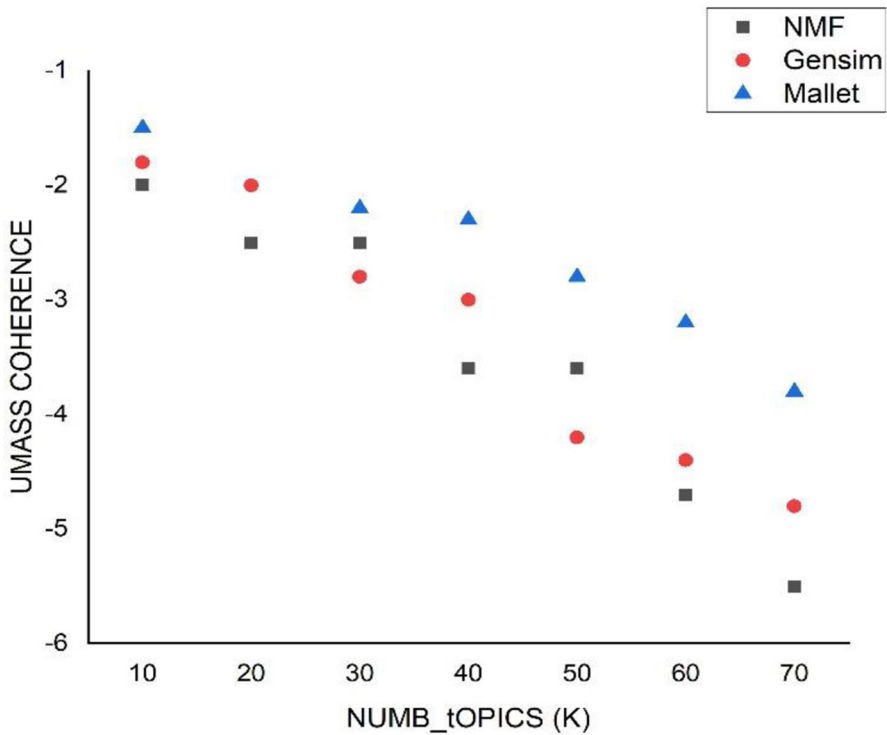
Fig. 4 Quantitative (UMass score) results for LDA and NMF

the closeness of the language used in the researcher's papers and the one employed in the unseen papers. The authors proposed the PRPRS (Personalized Research Papers Recommender System) in the second method, which extensively designed and implemented a user profile-based algorithm to extract keyword by keyword and keyword inference (Hong et al., 2013).

5.6 Evaluation metrics

We evaluate the general performance of our method using the two most commonly used evaluation metrics in recommender systems: Precision and recall. Precision or true.positive accuracy is calculated as the ratio of recommended articles relevant to the total number of recommended articles; Precision has given by Eq. (4).

$$Precision = \frac{\sum (relevant\_articles) \cap \sum (retrieved\_articles)}{\sum (retrieved\_articles)} \quad (4)$$

Recall or true positive rate is calculated as the ratio of recommended articles that are relevant to the total number of relevant articles, recall given by Eq. (5),

$$Recall = \frac{\sum (relevant\_articles) \cap \sum (retrieved\_articles)}{\sum (retrieved\_articles)} \tag{5}$$

6 Results and discussions

Our method is designed to favor precision which has more influence on the satisfaction of the target researcher than recall. Indeed, precision reflects the performance of the recommendation system in satisfying the target researcher's need for valuable articles. The different experiments we have conducted show that the scientific article recommendation method we have proposed provides the target researcher with scientific articles of high relevance compared to the state-of-the-art methods that rely on a single topic modeling technique. Our method first selects the best performing algorithm considering the textual nature (scientific text), which provides an optimal topic model, which serves as a basis for the recommendation process hence the accuracy obtained as illustrated in Fig. 5. Indeed, the accuracy of our method surpasses that of other methods, even if it is significantly close to the LDA_ASPR method for the values of N = 20 (relevant recommended papers up to position N). However, when the value of N increases beyond the value N = 25, it seems clear that the accuracy increases and surpasses that of the LDA_ASPR and PRPRS methods.

Figure 6 illustrates the comparison based on the recall; as we can see, the difference in performance between our method and LDA_ASPR is very insignificant. Indeed, the LDA_ASPR method performs slightly for N = 15 and N = 20; however, our method shows a significant difference as the number N increases, especially when N's value is greater than 25. The low recall-based performance of our method is the result of the qualification constraints of the candidate items. Thus, our method selects only those articles whose content is semantically very close to the content of the target article and thus leaves many other less relevant articles unrecalled. These improvements are mainly due to the strictness in

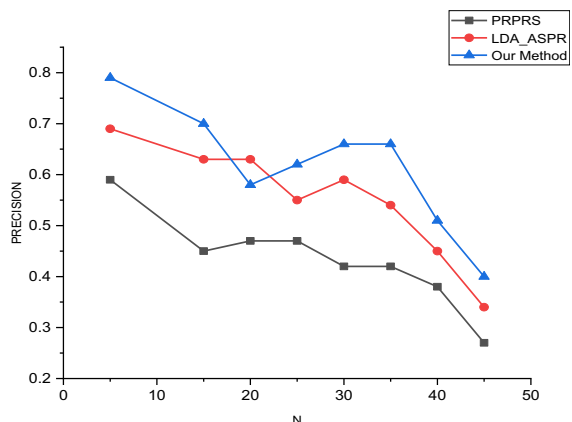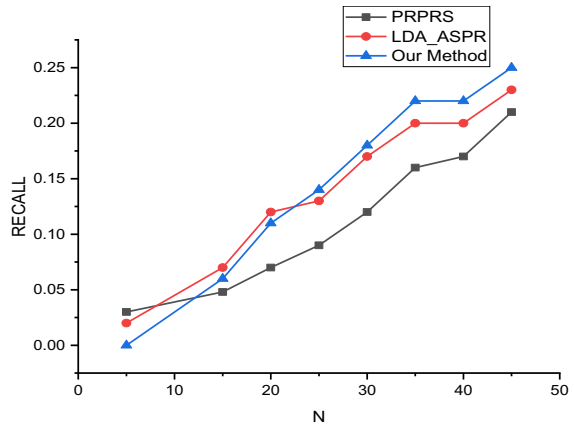**Fig. 5** Precision performance on the dataset

Fig. 6 Recall performance on the dataset

qualifying a candidate paper which removed less relevant papers to the target paper. This, therefore, increases the system's ability to return relevant and practical recommendations at the top of the recommendation list.

7 Conclusion and future work

In this article, we introduce a novel method for scientific article recommendation that leverages the power of both content-based filtering (CBF) and latent factors. Our method involves analyzing the two most commonly used topic modeling techniques to determine the most effective topic model for recommendation purposes. Given the absence of a consensus in the literature regarding the best modeling approach, our strategy is grounded in this knowledge. The solution presented in the article outperforms current reference methods on commonly used performance criteria, as demonstrated through experiments on publicly available datasets. Additionally, the solution can be utilized as an API to provide real-time recommendation services to suggest the top N articles to a target researcher.

In future work, we plan to investigate the integration of neural topic modeling based on transformer architecture and citation factors to produce highly accurate recommendations. Specifically, citation factors can offer valuable insights into the impact and influence of specific articles or authors, thereby serving as a complementary source of information for the recommendation process.

Declarations

Conflict of interest None.

References

Adomavicius, G. & Tuzhilin, A. (2005). Toward the next generation of recommender systems : a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, *17*(6). https://doi.org/10.1109/TKDE.2005.99

Albalawi, R., Yeap, T. H. & Benyoucef, M. (2020). Using Topic Modeling Methods for Short-Text Data : A Comparative Analysis. *Frontiers in artificial intelligence*, *3*. https://doi.org/10.3389/frai.2020.00042

Almonte, L., Guerra, E., Cantador, I., & de Lara, J. (2022). Recommender systems in model-driven engineering. *Software and Systems Modeling, 21*, 249–280. https://doi.org/10.1007/s10270-021-00905-x

Amami, M., Pasi, G., Stella, F., & Faiz, R. (2016). An LDA-based approach to scientific paper recommendation. In Springer eBooks (pp. 200–210). Springer Nature. https://doi.org/10.1007/978-3-319-41754-7_17

Bagul, D., & Barve, S. (2021). A novel content-based recommendation approach based on LDA topic modeling for literature recommendation. *International Conference on Inventive Computation Technologies*. https://doi.org/10.1109/icict50816.2021.9358561

Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., & Xai, F. (2020). Scientific Paper Recommendation : A Survey. *IEEE Access, 7*, 9324–9339.

Beel, J., Gipp, B., & Langer, S. (2016). Research-paper recommender systems : A literature survey. *International Journal on Digital Libraries, 17*, 305–338. https://doi.org/10.1007/s00799-015-0156-0

Berhil, S., Benlahmar, H., & Labani, N. (2020). A review paper on artificial intelligence at the service of human resources management. *Indonesian Journal of Electrical Engineering and Computer Science, 18*, 32–40.

Blei, D., & M., Ng, A. & Jordan, M., I. (2000). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*, 993–1022.

Boussaadi, S., Aliane, H., Abdeldjalil, O., Houari, D., & Djoumagh, M. (2020). Recommender systems based on detection community in academic social network. *2020 International Multi-Conference on : "Organization of Knowledge and Advanced* Technologies*" (OCTA)*. https://doi.org/10.1109/octa49274.2020.9151729

Guo, C., Lu, M., & Wei, W. (2021). An Improved LDA Topic Modeling Method Based on Partition for Medium and Long Texts. *Annals of Data Science, 8*, 331–344. https://doi.org/10.1007/s40745-019-00218-3

Hadhiatma, A., Azhari, A., & Suyanto, Y. (2023). A Scientific Paper Recommendation Framework Based on Multi-Topic Communities and Modified PageRank. *IEEE Access.*, 1–1. https://doi.org/10.1109/ACCESS.2023.3251189

Hong, K., Jeon, H. & Jeon, C. (2013). Personalized Research Paper Recommendation System using Keyword Extraction Based on UserProfile. *Journal of Convergence Information Technology*, 106–116

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling : Models, applications, a survey. *Multimedia Tools and Applications, 78*(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jooneghani, Z. K. N. (2021). Comparison of topic modeling methods for analyzing tweets on Covid-19 vaccine.

Kershaw, D., & Koeling, R. (2020). Elsevier OA CC-BY corpus. *Elsevier Data Repository*, V3. https://doi.org/10.17632/zm33cdndxs.3

Lee, D. & Seung, H. (1998). Learning the parts of objects by non-negative matrix factorization. *Nature*, 788–791. https://doi.org/10.1038/44565

Leung, Y. T., & Khalvati, F. (2022). Exploring COVID-19–Related stressors : Topic modeling study. *Journal of Medical Internet Research, 24*(7), e37142. https://doi.org/10.2196/37142

Levene, M. (2022). A Skew Logistic Distribution for Modelling COVID-19 Waves and Its Evaluation Using the Empirical Survival Jensen-Shannon Divergence. *Entropy, 24*(5), 600. https://doi.org/10.48550/arXiv.2201.13257

Rossetti, M., Pettit, B., Vargas, S., Kershaw, D., Jack, K., Magatti, D. & Hristakeva, M. (2017). Effectively identifying users' research interests for scholarly reference management and discovery. *the 1st Workshop on Scholarly Web Mining*, 17–24. https://doi.org/10.1145/3057148.3057151

Sakib, N., Ahmad, R., B. & Haruna, K. (2022). A collaborative approach toward scientific paper recommendation using citation context. IEEE Access, *8*, 51246–51255.

Stevens, K., Kegelmeyer, P., Andrzejewski, D. & Buttler, D. (2012). Exploring topic coherence over many models and many topics. *conference on empirical methods in natural language processing and computational natural language learning*, 952–961

Subathra, P., & Kumar, P. (2019). Recommending Research Article Based on User Queries Using Latent Dirichlet Allocation. *Springer Singapore eBooks*, 163–175. https://doi.org/10.1007/978-981-15-2475-2_15

Wang, C. & Blei, D., M. (2011). Collaborative topic modeling for recommending scientific articles. *Knowledge discovery and data mining*, 448–456