Enhancing Scholarly Paper Recommendation by Modelling Diversity of Research Interests

Xueli Pan[1] [0000−0002−3736−7047], Shuai Wang[1] [0000−0002−1261−9930], Ting Liu[1] [0000−0002−8208−3430], Jacco van Ossenbruggen[1] [0000−0002−7748−4715], Victor de Boer[1] [0000−0001−9079−039X], and Zhisheng Huang[1] [0000−0003−3794−9829]

Department of Computer Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, Amsterdam, Netherlands

{x.pan2, shuai.wang, t.liu, jacco.van.ossenbruggen, v.de.boer, z.huang}@vu.nl

Abstract. Recommender systems help researchers identify relevant papers in scientific document collections. A precise user interest model is crucial for content-based scholarly paper recommendation. Arguably, past publications play an important role in modelling researchers' interests. However, not all publications account for the interest model equally. Existing approaches introduce weighting schemes to emphasize the impact of recent articles published by each researcher. However, these weighting schemes fail to explain the content-wise relationship (e.g. diversity) among their publications. In this paper, we introduce a new feature to capture the diversity of research interests derived from each researcher's publications, which can be combined with such weighting schemes. We further employ this feature in two weighting schemes to model research interests for each researcher. We investigate the effect of the new feature with two text representation models to represent papers and compare the effectiveness of four weighting schemes to model user interest. We conduct experiments on a public dataset of 50 researchers. Results show that although the accuracy obtained with our proposed weighting schemes is not stable with different parameter settings, our methods in optimal settings reveal an increase in accuracy measured by NDCG@10 and P@10, compared to other existing weighting schemes.

Keywords: Recommender system · User modelling · Scientific recommendation.

1 Introduction

The scientific recommender system, a common approach to address the problem of information overload in academic recommendation, has received considerable attention from both industry and academia. Content-based filtering(CBF) and collaborative filtering(CF) are two popular approaches in scholarly recommender systems. The CBF approach mainly recommends papers similar to papers that users have published in the past, while the CF approach recommends papers based on the interaction on papers of neighbor users whose profiles are similar to the target user. In a literature survey of scholarly recommender systems [1], 55% of the reviewed approaches applied content-based filtering and only 18% applied collaborative filtering. The reason for the widespread use of content-based filtering in this scenario is three-fold. First, researchers' publication records are publicly available via Google Scholar or other scholarly engines. Second, researchers' publications contain rich contextual information about their research interests. Third, historical interaction data is not required in content-based recommender systems, thus avoiding the cold start problem. Therefore, we investigate applying a content-based filtering to enhance scholarly paper recommendation.

Research interest modelling is crucial in the content-based scholarly paper recommendation. Studies have been conducted in academia to model researchers' interests according to their publications [2, 9, 12, 13]. Most existing works in this direction focus on utilizing side information, such as referenced papers, cited papers [13] and co-author's publications [9] to enrich the context of researchers' publications, thus enhancing the accuracy in scholarly paper recommendation. Although introducing additional information could enrich the context of papers, this will bring other drawbacks, such as being more expensive in computational memory. Therefore, in this paper, we investigate how researchers' interest models can be improved by taking full advantage of their publications.

Existing work [12] indicates that researchers' publications constitute a clear signal of their latent interests. However, not all publications contribute equally to the modelling of research interests. Some studies investigate the influence of publication recency in modelling researchers' interests by introducing weighting schemes to emphasize the impact of recent articles published by each researcher. Nevertheless, the recency characteristic does not explain the content-wise relationship among a researcher's publications, e.g. diversity.

Problem Statement: For researchers, how can we take into account the diversity of research interests derived from their publications to improve user interest modelling, thus enhancing the accuracy of scholarly paper recommendation? By analyzing the distribution of cosine similarity among each researcher's publications, we identify a new feature, namely the standard deviation of cosine similarity among all papers that a researcher has published (denoted as std), to capture the diversity of research interests for each researcher (details in Section 3.2 and Section 5.3).

To address the above-mentioned problem, we study the following research question: What is the effect of this new feature that takes into account diversity on user interest modelling in a content-based scholarly paper recommender system? We investigate the research question with one dataset, two frequently used text representation models to represent the content of papers with full text, and four weighting schemes to model researchers' interests based on their past publications. We first study the effectiveness of this new feature with two different text representation models, Term Frequency-Inverse Document Frequency (TF-IDF) and Word2Vec when representing scientific papers using their full text in scholarly paper recommendation. Simultaneously, we compare the effectiveness of four weighting schemes on user interest modelling, including our proposed weighting schemes with the new feature. We conduct experiments on a public scholarly paper recommendation dataset [11] of 50 researchers in the domain of computer science and the accuracy of recommendation is measured by normalized discounted cumulative gain (NDCG), precision (P) and mean reciprocal rank (MRR). The experimental results show that although the accuracy obtained by our methods is not stable with different parameter setting, in optimal settings, our proposed simple yet effective weighting schemes with the new feature show potential to capture the diversity of research interests, thus improving user interest modelling to enhance the accuracy.

This paper makes the following research contributions:

- We identify a new feature that captures the diversity of research interests among each researcher's publications, namely the standard deviation of cosine similarity among papers on a researcher's publication list.
- We propose two weighting schemes to study the effectiveness of this new feature on user interest modelling, compared with two existing weighting schemes.
- We conduct experiments on a real-world dataset of 50 researchers and experimental results show how the recommendation accuracy could be improved by simple yet effective weighting schemes with appropriate parameter setting.

The remainder of the paper is organized as follows. In Section 2, we review the related literature. In Section 3, we present our methods. In Section 4, we evaluate the performance of our proposed methods. In Section 5, we discuss the results. In Section 6 we summarize our main contributions, outline the limitations and future research directions.

2 Related Work

User modelling is an important component in content-based scholarly paper recommendation, but it is often ignored [1]. The papers that researchers have read or have published in the past have been used to infer their research interests in scholarly paper recommendation. Most current studies focus on using the recency characteristic of papers to model researchers' interests. Sugiyama et al. [12] investigated the use of the most recent published article to model research preferences for researchers. In a subsequent work, they proposed to use all published papers with a forgetting factor to assign more weights to newer publications [10]. Jin Zhang et al. [16] took into account the year difference between the oldest paper and the others on researchers' publication list to generate the embedding of users' research interests.

Document representation is another important component in scholarly paper recommendation. Current studies focus more on what kind of data could be used to represent the content of papers and what kind of techniques could be applied to represent the content of papers. Title, keywords and abstract are often used to represent the content of papers because they are publicly available with little copyright concern [3, 4, 5, 8, 14, 15]. Text representation models such as TF-IDF and Word2Vec are often used in scholarly paper recommendation. TF-IDF has become one of the most popular term-based document representation models in content-based paper recommendations. In a survey [1], 83% of text-based recommender systems in digital libraries use TF-IDF due to its interpretability and easy implementation. Although TF-IDF is easy to implement, it only considers term frequency and lacks semantics about words. Word2Vec, a word embedding technique, has been proposed to enrich words with semantics [7]. Bulut et al. [2] and Meijer et al. [6] showed that Word2Vec outperformed TF-IDF in representing papers with short text such as titles, keywords and abstracts. In our work, we investigate the effect of the new feature with these two text representation models to represent papers using their full text.

3 Methodology

Our work aims to improve the accuracy of scholarly paper recommendation by improving the modelling of user interest. Unlike existing approaches which focus on the impact of publication recency in modelling researchers' interests, our work identifies a new feature that could capture the diversity of research interests for each researcher. We propose two new weighting schemes with the new feature for user interest modelling. We further study the effect of this new feature in two aspects. First, we study the performance of the new feature with two text representation models TF-IDF and Word2Vec; Second, we compare the effectiveness of four weighting schemes in user interest modelling, including our proposed weighting schemes with the new feature and two other existing weighting schemes.

3.1 Scholarly Paper Recommendation Pipelines

The scholarly paper recommendation task is defined as follows: For each researcher, we first model their interests as vectors using the full text of their publications, and then recommend papers relevant to their interests according to the cosine similarity between vectors of their interests and vectors of candidate papers. With the task defined, we construct the pipeline for the task from the perspective of a classic content-based recommendation that consists of the following three steps:

1. For each researcher $r \in R$, we model user interests U_r derived from full text of their publications. For each paper p on a researcher's publication list, we denote f_p as the feature vector of its content. In our work, the feature vector of each paper f_p is constructed using TF-IDF or Word2Vec.
2. For each paper $c \in C$ in candidate papers to be recommended, a feature vector of the paper f_c is built using TF-IDF or Word2Vec.
3. For each researcher r, given the feature vector of their research interests U_r and feature vector of each candidate paper f_c in our dataset, we rank all candidate papers according to the cosine similarity between U_r and f_c to make recommendations.

3.2 User Interest Modelling

To investigate the effect of the new feature, we propose two weighting schemes $\mathcal{W}_\lambda$ and $\mathcal{W}_{log}$ to model researchers' interests. Two other existing weighting schemes $\mathcal{W}_{mr}$ and $\mathcal{W}_{all}$ are considered as baselines.

1. $\mathcal{W}_{mr}$: **using only the most recent publication**. The first weighting scheme takes into account only the most recent published paper, which is also used by Sugiyama et al. [12]. For each researcher, the vector of their research interests is simply the vector f_{p1} of the most recent published paper p_1:

$$U_r = f_{p1} \tag{1}$$

2. $\mathcal{W}_{all}$: **Using all publications with equal weights**. The second weighting scheme takes into account all papers on each researcher's publication list with equal weights. N denotes the number of publications and px denotes the x^{th} paper of researcher r.

$$U_r = \frac{1}{|N|} \sum_{x=1}^{n} f_{px} \tag{2}$$

3. $\mathcal{W}_{\lambda}$: **Using different weighting schemes for different researchers**. Based on our observation (details in Section 5.3) on the distribution of cosine similarity among papers on each researcher's publication list[1], we propose the following hypotheses. First, a low standard deviation with a high average cosine similarity indicates that most of the publications of a researcher are highly relevant to a topic. Second, a high standard deviation with a low average cosine similarity indicates that the topics derived from the researcher's publications are diverse. We later identify a new feature to capture the diversity of interest for each researcher, namely the standard deviation of cosine similarity among papers on their publication list (denoted as σ in Equation 3 and Equation 4). Based on the hypotheses, we propose to employ different weighting schemes to different researchers according to this new feature. For researchers of the first hypothesis, we assume that taking into account all publications with equal weights could capture all topic information about a researcher's preference because they have a narrow range of topics. For researchers of the second hypothesis, we assume that utilizing only the most recent published article could capture their current interests because they have a broad range of topics.

$$U_r = \begin{cases} f_{p1}, & (\sigma \geq \lambda) \\ \frac{1}{|N|} \sum_{x=1}^{n} f_{px}, & (\sigma < \lambda) \end{cases} \tag{3}$$

Here λ is a tunable constant ranging from 0 to 1.

4. $\mathcal{W}_{log}$: **Integrating the new feature into a weighting function**. In this variation, we further explore how user interest modelling derived from their publications could be affected by the combination of this feature and the recency characteristic. According to our observation, the larger the σ, the greater the difference in cosine similarity between publications, indicating the degree of diversity among the publications of each researcher. Therefore, newer publications are closer to the current research interests of each researcher, and larger weights should be assigned to newer publications and smaller weights to older publications. The weight W_{px} of the x^{th} paper on the publication list of researcher r and the corresponding user interest model U_r are defined as follows:

$$W_{px} = -1 * \log_{x+\beta}(\sigma + \alpha) \tag{4}$$

$$U_r = \sum_{x=1}^{n} \left(f_{px} * \frac{W_{px}}{\sum_{x=1}^{n} W_{px}} \right) \tag{5}$$

Here α and β are tunable positive constants.

[1] Details examination in supplementary material.

4 Experiments

4.1 Dataset

We conduct experiments on a Scholarly Paper Recommendation Dataset [2] created by Sugiyama et al. [11]. This dataset is composed of three parts. (1) Publication lists of 50 researchers in the domain of computer science in DBLP[3]: For all researchers, the minimum number of publications is 2 and the maximum number of publications is 31; (2) Candidate papers to be recommended: The dataset contains 100,531 papers published in various kinds of ACM proceedings between 2000 and 2010 in the domain of computer science; (3) Relevant papers of each researcher: Each researcher was asked to manually mark papers relevant to their recent research interest, making them a gold-standard results for evaluation. The average number of relevant papers of all researchers is 74.56. Apart from containing a gold standard, this dataset contains the full text information of each paper, which makes it an ideal dataset for our work, since we leverage full text to represent the content of papers.

4.2 Data Pre-processing

For each paper, we reuse the TF-IDF vector in the dataset and build the Word2Vec vector using its full text provided by the creator of the dataset. Since all papers are in PDF format, we first use a Python package pdfminer.six[4] to extract text information from PDF documents. Then, stop words are eliminated and stemming is performed using Porter Stemmer for English in nltk[5]. After the text cleaning, we obtain a corpus containing the clean text of each paper. We pretrain a Word2Vec model based on the corpus obtained instead of using the general Word2Vec model trained on Google News. In our experiments, the vector of a paper is the sum of the vectors of words that appear in the document after text cleaning.

4.3 Implementation

All experiments were carried out on a laptop with an 8-core CPU (1.4GHz). We used Python 3.9 with libraries including pdfminer.six[4], nltk[5] and gensim[6]. The code and data are available in the GitHub repository [7] for reproducibility.

4.4 Evaluation Metrics

We evaluate the accuracy of recommendation with three widely used metrics in the evaluation of ranked information retrieval, specifically: (1) normalized discounted cumulative gain (NDCG), (2) precision (P) and (3) mean reciprocal rank (MRR). As users

[2] https://doi.org/10.25540/BBCH-QTT8
[3] https://dblp.org/
[4] https://pdfminersix.readthedocs.io/en/latest/#
[5] https://www.nltk.org/
[6] https://pypi.org/project/gensim/
[7] https://github.com/SherryPan0/user-interest-modeling.git

often scan only the first few ranks, we only take into account the top 10 ranking documents. Therefore, in this work, we use NDCG@10, P@10 and MRR for evaluation.

For the following metrics, r denotes a researcher from the set of users U.

- **Normalized Discounted Cumulative Gain (NDCG)** is a measure of ranking quality taking into account both relevancy and the rank position of each item on the recommendation list. The Cumulative Gain(CG) is the sum of the graded relevance values of all results on a recommendation list. In our work, the relevance level depends on a binary notion of relevance: Whether recommended papers are relevant or not for a researcher. We use $G(i) = 1$ for relevant documents and $G(i) = 0$ for irrelevant documents. As such, CG is the number of relevant papers on a recommendation list. Discounted Cumulative Gain (DCG) is a refinement of CG that takes into account the rank position of each paper on the recommendation list. In Equation 6, i denotes the ranking position of a specific paper on the recommendation list. In our work, the maximum possible DCG through position i, is called the ideal DCG (IDCG) through that position. The average normalized DCG(NDCG) for 50 researchers is used to measure the overall accuracy in our experiments.

$$DCG_i = \begin{cases} G(1), & (i=1) \\ \sum_{1}^{i-1} DCG(i-1) + \frac{G_i}{\log(i+1)}, & (i>1) \end{cases} \tag{6}$$

$$NDCG@10 = \frac{1}{50} \sum_{r \in U} \frac{DCG_{10}}{IDCG_{10}} \tag{7}$$

- **Precision** is the fraction of relevant documents on the recommendation list. Here, we take the average precision for all users.

$$P@10 = \frac{1}{50} \sum_{r \in U} \frac{n_r}{10} \tag{8}$$

  Where n_r refers to the total number of relevant documents on the recommendation list for user r.
- **Mean Reciprocal Rank (MRR)** is the average of the reciprocal ranks of the results for all users U. It only takes into account the position of the first relevant document on the ranking list. The MRR is defined as follows:

$$MRR = \frac{1}{50} \sum_{r \in U} \frac{1}{i_r} \tag{9}$$

  Where i_r refers to the rank position of the first relevant document for researcher r.

5 Experimental Results and Discussion

5.1 Using TF-IDF and Word2Vec to Represent Papers

We study the performance of recommendation with two text representation models, TF-IDF and Word2Vec. Table 1 shows the accuracy of recommendation evaluated with

NDCG@10, P@10 and MRR. When comparing TF-IDF with Word2Vec, Word2Vec outperforms TF-IDF in our experiments with respect to NDCG@10 and P@10, regardless of the weighting schemes. TF-IDF outperforms Word2Vec but the margin is not significant when the accuracy is measured by MRR. More specifically, compared to the fusion of TF-IDF and different weighting schemes, recommendation accuracy measured by NDCG@10 is improved by 30.25%, 23.51%, 25.15% and 27.02%, respectively, when combining Word2Vec with different weighting schemes. When it comes to P@10, accuracy is improved by 39.58%, 18.75%, 35.85% and 8.33%, respectively. In contrast, when the accuracy is measured by MRR, TF-IDF performs slightly better than Word2Vec with an improvement of 3.65%, 5.80%, 17.43% and 2.08% respectively, combining with different weighting schemes. In general, Word2Vec outperforms TF-IDF in capturing the content of scientific papers with full text in our experiments. Although this result is not surprising, to the best of our knowledge, we are the first to empirically investigate the performance of TF-IDF and Word2Vec in representing research papers with full text in this scenario.

Table 1: Accuracy of recommendation with different text representation models and different weighting schemes

		$\mathcal{W}_{mr}(\%)$	$\mathcal{W}_{all}(\%)$	$\mathcal{W}_{\lambda}(\%)$	$\mathcal{W}_{log}(\%)$ $\alpha = 0.1, \beta = 1$	$\mathcal{W}_{log}(\%)$ $\alpha = 0.1, \beta = 10$
NDCG@10	TF-IDF	28.10	30.20	31.80 ($\lambda = 0.24$)	29.90	30.50
	Word2Vec	36.60	37.30	**39.80** ($\lambda = 0.08$)	37.98	36.15
P@10	TF-IDF	9.60	9.60	10.60 ($\lambda = 0.24$)	9.60	9.80
	Word2Vec	13.40	11.40	**14.40** ($\lambda = 0.08$)	10.40	11.40
MRR	TF-IDF	28.40	34.70	**35.70** ($\lambda = 0.24$)	33.89	34.55
	Word2Vec	27.40	32.80	30.40 ($\lambda = 0.08$)	33.20	32.61

5.2 Using Different Weighting Schemes to Model User Interests

We study the effectiveness of this new feature in user interest modelling by using different weighting schemes to construct research interest for each researcher. Table 1 shows the overall accuracy with respect to NDCG@10, P@10 and MRR. As shown in Table 1, combining $\mathcal{W}_{\lambda}$ with the Word2Vec model, we obtain the best NDCG@10 of 39.8% and the best P@10 of 14.4% when λ is 0.08. Consistent with what we have mentioned in Section 5.1, the best MRR(35.7%) is obtained when we combine the TF-IDF model with $\mathcal{W}_{\lambda}$. As we have mentioned in Section 3.2, we use the new feature to capture the diversity of research topics in each researcher's publications. The improvement in accuracy when employing $\mathcal{W}_{\lambda}$ shows the effectiveness of the new feature to differentiate researchers with a single research topic and those with more diverse topics.

In Figure 1, we further study the effect of this new feature when using $\mathcal{W}_{\lambda}$ to model researchers' interests and using Word2Vec to represent papers with full text. As shown in this figure, the accuracy is affected by different λ. When λ is greater than 0.08, our proposed weighting scheme $\mathcal{W}_\lambda^i$ outperforms the other two weighting schemes in NDCG@10. When λ is less than 0.08, $\mathcal{W}_\lambda^i$ performs worse than $\mathcal{W}_{mr}$ or $\mathcal{W}_{all}$ or both. The accuracy measured by P@10 and MRR also shows that the weighting scheme $\mathcal{W}_\lambda^i$ outperforms others only with an appropriate λ. If λ is not set properly, the accuracy of the recommendation could be reduced.

We also attempt to integrate the new feature with the recency of publications into a unified weighting function $\mathcal{W}_{log}^i$. As shown in Table 1, compared to the combination of Word2Vec and the other two weighting schemes $\mathcal{W}_{mr}$ and $\mathcal{W}_{all}$, combining $\mathcal{W}_{log}^i$ with Word2Vec performs slightly better in NDCG@10 (37.98%) and MRR (33.20%) when $\alpha = 0.1$ and $\beta = 1$. However, with another setting of $\alpha = 0.1$ and $\beta = 10$, the accuracy is reduced.

These results show that the accuracy of recommendation is affected by employing different weighting schemes with the new feature to model user interests. On the one hand, these indicate the effectiveness of the new feature in modelling researchers' interests. On the other hand, setting parameters is of great importance. Appropriate parameters could increase accuracy, while inappropriate parameters could reduce accuracy.

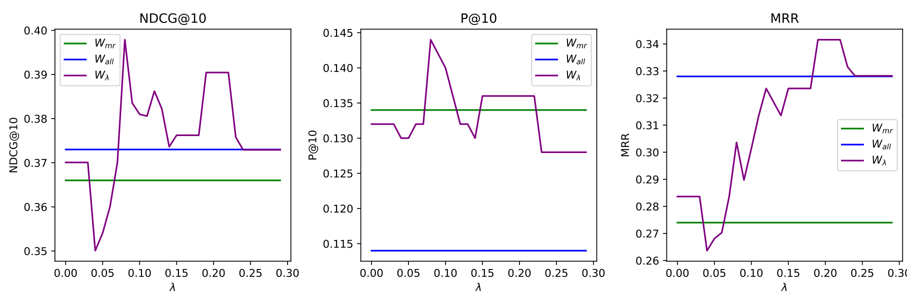


Fig. 1: The effect of different λ on accuracy wrt NCDG@10, P@10 and MRR when combining $\mathcal{W}_\lambda^i$ and Word2Vec

5.3 Discussion

We investigate the effect of the new feature with two text representation models to represent the content of papers and four weighting schemes to model researchers' interests based on their publications. The experimental results show that the combination of $\mathcal{W}_\lambda^i$ and Word2Vec outperforms others with respect to NDCG@10 and P@10 when λ is 0.08. To illustrate the discovery of this new feature and the rationale behind its corresponding weighing scheme $\mathcal{W}_\lambda^i$, we analyze the distribution of cosine similarity among papers on each researcher's publication list to capture the diversity of research interests. We take two representative researchers from the dataset to illustrate our observation. They have a similar number of publications but show a very different pattern regarding diversity of research interests derived from their publications.

Table 2 and Table 3 show the cosine similarity among the publications of researchers A and B, respectively. The colors of cosine similarity of 1, 0.5, and 0 are set as brown, pink and light grey, respectively. The darker the color, the higher the value of cosine similarity. As shown in Table 2, researcher A has four published papers in the database, where a_1 denotes the latest publication and a_4 denotes the oldest. (a_1, a_2) and (a_3, a_4) have much higher cosine similarity scores(≥ 0.50), while the rest of the pairs have much lower cosine similarity scores(≤ 0.15). A wide range of cosine similarity results in a low average cosine similarity of 0.25 and a relatively high standard deviation of 0.291. As shown in Table 3, researcher B has five published papers. All publications are highly similar to each other, resulting in a high average cosine similarity of 0.79 and a relatively low standard deviation of 0.07.

These tables illustrate that, although A and B share a similar number of publications, the distributions of cosine similarity among their publications show notable dissimilarities. This indicates that the implicit association among their publications is different in terms of content similarity. A one-size-fits-all weighting scheme could not work well for all researchers. Therefore, we propose to employ different weighting schemes for different types of researchers based on our hypotheses mentioned in Section 3.2.

For researchers similar to A, we assume that using only the most recent paper could better capture their recent research interest. For researchers similar to B, we assume that all publications could be considered in modelling their interests since they are highly relevant in content similarity. Based on these analyses, we identify the new feature to capture the diversity of research interests and employ it in two weighting schemes $\mathcal{W}_\lambda$ and $\mathcal{W}_{log}$ to model research interest for each researcher. The experimental results show that when Word2Vec is combined with our proposed weighting scheme $\mathcal{W}_\lambda$ with $\lambda = 0.08$, the accuracy measured by NDCG@10 is improved by 6.7% in our experiments. Using the same dataset, the state-of-the-art study [10] improved the accuracy of recommendation by almost 10% in NDCG@10 by enriching the representation of scientific papers. Unlike the state-of-the-art study, our methods focus on user interest modelling. Therefore, our methods could be combined with theirs to achieve better performance.

Despite the improvement in accuracy we obtain with our proposed weighting schemes, determining the value of λ in $\mathcal{W}_\lambda$ remains a challenge. As shown in Table 1 and Figure 1, the best λ may be subject to different methods and different metrics. Therefore, further investigation of the parameter configuration needs to be carried out.

6 Conclusion and Future Work

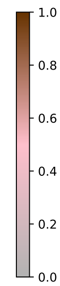
Based on an analysis of the distribution of cosine similarity among publications for each researcher, we identify a new feature to capture the diversity of research interests for each researcher, namely the standard deviation of cosine similarity among papers on a researcher's publication list. To study the effect of this feature in a content-based scholarly paper recommender system, we propose to employ this new feature in two weighting schemes to model research interests for each researcher. On the one hand, we investigate the effect of the new feature with two frequently used text representation models, TF-IDF and Word2Vec. On the other hand, we compare the performance of

Table 2: Cosine similarity among A's publications

	a_1	a_2	a_3	a_4
a_1	-	0.50	0.09	0.05
a_2	0.50	-	0.14	0.10
a_3	0.09	0.14	-	0.60
a_4	0.05	0.10	0.60	-

Table 3: Cosine similarity among B's publications

	b_1	b_2	b_3	b_4	b_5
b_1	-	0.78	0.79	0.83	0.82
b_2	0.78	-	0.87	0.71	0.76
b_3	0.79	0.87	-	0.71	0.75
b_4	0.83	0.71	0.71	-	0.85
b_5	0.82	0.76	0.75	0.85	-



our proposed weighting schemes with two existing weighting schemes for user interest modelling.

Our experimental results on a public scholarly paper recommendation dataset of 50 researchers in the domain of computer science show the effectiveness of our proposed weighting schemes both for researchers with a single research topic and for those with more diverse topics. Although the accuracy of the method is dependent on parameter settings, which need to be experimentally established for a specific recommendation task, in optimal settings, our methods show an increase in accuracy when compared with the baselines.

Compared to the state of the art, the main value of our user interest modelling approach is that it can be integrated with other scientific recommendation methods in the future. More research on time efficiency and computational resources could be conducted when employing our weighting schemes to more complex methods.

Acknowledgements The authors thank the Scholarly Paper Recommendation Datasets group for sharing the raw PDF files with us.

References

[1] Joeran Beel et al. "Paper recommender systems: a literature survey". In: *International Journal on Digital Libraries* 17 (2016), pp. 305–338.

[2] Betül Bulut et al. "A Paper Recommendation System Based on User's Research Interests". In: *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ASONAM '18. Barcelona, Spain: IEEE Press, 2020, pp. 911–915. ISBN: 9781538660515.

[3] Betül Bulut et al. "User's Research Interests Based Paper Recommendation System: A Deep Learning Approach". In: 2019. URL: `https://api.semanticscholar.org/CorpusID:213248698`.

[4] Guibing Guo et al. "Leveraging Title-Abstract Attentive Semantics for Paper Recommendation". In: *AAAI Conference on Artificial Intelligence*. 2020. URL: `https://api.semanticscholar.org/CorpusID:213269997`.

[5] Xiao Ma and Ranran Wang. “Personalized Scientific Paper Recommendation Based on Heterogeneous Graph Representation”. In: *IEEE Access* 7 (2019), pp. 79887–79894. URL: https://api.semanticscholar.org/CorpusID:195775088.

[6] HJ Meijer, J Truong, and R Karimi. “Document embedding for scientific articles: Efficacy of word embeddings vs TFIDF”. In: *arXiv preprint arXiv:2107.05151* (2021).

[7] Tomas Mikolov et al. “Distributed representations of words and phrases and their compositionality”. In: *Advances in neural information processing systems*. 2013, pp. 3111–3119.

[8] Julia A. Sterling and Matthew M. Montemore. “Combining Citation Network Information and Text Similarity for Research Article Recommender Systems”. In: *IEEE Access* 10 (2022), pp. 16–23. DOI: 10.1109/ACCESS.2021.3137960.

[9] Kazunari Sugiyama and Min-Yen Kan. “A comprehensive evaluation of scholarly paper recommendation using potential citation papers”. In: *International Journal on Digital Libraries* 16.2 (2015), pp. 91–109.

[10] Kazunari Sugiyama and Min-Yen Kan. “Exploiting potential citation papers in scholarly paper recommendation”. In: *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*. 2013, pp. 153–162.

[11] Kazunari Sugiyama and Min-Yen Kan. “Scholarly Paper Recommendation Datasets”. In: (2013). DOI: https://doi.org/10.25540/BBCH-QTT8.

[12] Kazunari Sugiyama and Min-Yen Kan. “Scholarly paper recommendation via user’s recent research interests”. In: *Proceedings of the 10th annual joint conference on Digital libraries*. JCDL ’10. New York, NY, USA: Association for Computing Machinery, June 21, 2010, pp. 29–38. ISBN: 978-1-4503-0085-8. DOI: 10.1145/1816123.1816129. URL: https://doi.org/10.1145/1816123.1816129 (visited on 05/30/2023).

[13] Kazunari Sugiyama and Min-Yen Kan. “Towards higher relevance and serendipity in scholarly paper recommendation”. In: *ACM SIGWEB Newsletter* Winter (2015), pp. 1–16.

[14] Hao Tang, Baisong Liu, and Jiangbo Qian. “Content-based and knowledge graph-based paper recommendation: Exploring user preferences with the knowledge graphs for scientific paper recommendation”. In: *Concurrency and Computation: Practice and Experience* 33 (2021). URL: https://api.semanticscholar.org/CorpusID:233926837.

[15] Jie Yu, Junchen He, and Lingyu Xu. “Hypergraph-Based Academic Paper Recommendation”. In: *Knowledge Science, Engineering and Management*. 2022. URL: https://api.semanticscholar.org/CorpusID:250926147.

[16] Jin Zhang et al. “Personalized scientific and technological literature resources recommendation based on deep learning”. In: *Journal of Intelligent & Fuzzy Systems* 41.2 (2021), pp. 2981–2996.