A. Khadka, I. Cantador and M. Fernandez, "Capturing and Exploiting Citation Knowledge for Recommending Recently Published Papers," *2020 IEEE 29th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)*, 2020, pp. 239-244

DOI: http://doi.org/10.1109/WETICE49692.2020.00054

Capturing and Exploiting Citation Knowledge for Recommending Recently Published Papers

Anita Khadka
Knowledge Media Institute
The Open University
Milton Keynes, United Kingdom
anita.khadka@open.ac.uk

Iván Cantador
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Madrid, Spain
ivan.cantador@uam.es

Miriam Fernandez
Knowledge Media Institute
The Open University
Milton Keynes, United Kingdom
miriam.fernandez@open.ac.uk

***Abstract*—With the continuous growth of scientific literature, discovering relevant academic papers for a researcher has become a challenging task, especially when looking for the latest, most recent papers. In this case, traditional collaborative filtering systems are ineffective, since they are unable to recommend items not previously seen, rated or cited. This is known as the item cold-start problem. In this paper, we explore the potential of exploiting citation knowledge to provide a given user with relevant suggestions about recent scientific publications. A novel hybrid recommendation method that encapsulates such citation knowledge is proposed. Experimental results show improvements over baseline methods, evidencing benefits of using citation knowledge to recommend recently published papers in a personalised way. Moreover, as a result of our work, we also provide a unique dataset that, differently to previous corpora, contains detailed paper citation information.**

***Index Terms*—Scientific publications, recommender systems, citations, dataset.**

I. Introduction

With the continuous and increasing growth of scientific literature, it is becoming more and more challenging for researchers to keep up to date with the latest papers of interest. A recent report by the International Association of Scientific, Technical and Medical Publishers[1] claims the existence of 33,100 active scholarly peer-reviewed English-language journals in mid-2018, collectively publishing over 3 million papers a year. The report also states that the production of scientific publications is steadily increasing at a 4% yearly rate.

Given the scale of the available information, a wide range of efforts have been invested in the last decade to discover, analyse and exploit scholarly content. Among these efforts, we can highlight the development of academic search engines like Google Scholar[2], review management platforms [1], scientometrics systems [2], and Recommender Systems (RS) for research papers, books, patents, among others [3].

Regarding the recommendation of academic papers, several tasks have been addressed in the literature, including recommending relevant papers for varied targets such as a user [4], [5], a paper [6], [7], a particular snapshot of content (title, abstract, etc.) [8], a particular collection of papers [9], and a manuscript (i.e. a paper yet to be published) [10], [11].

[1]https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf
[2]https://scholar.google.co.uk/

While most of the existing solutions for recommending relevant papers are performed independently of the time regardless of their publication date, we focus on the real-world problem of recommending recently published (new) papers [12]. Addressing this problem, we investigate how citation knowledge could be captured and exploited to support users towards the discovery of recent and relevant scientific publications. On doing so, we propose a novel recommendation method that explores the users' publication history (–their authored and cited papers) to build their profiles, and citation knowledge including the citations between papers (i.e. citation graph), the sections where citations appear (i.e. citation section), and texts that surround citations within the papers (i.e. citation context) to provide personalised recommendations. In this context, it has to be noted that, as best we know, public datasets containing such rich information do not exist (see Table I).

Existing datasets used for academic recommendation tasks do not provide either the entire user publication history or full text of papers, but just their metadata (e.g., title, abstract etc.). As part of our work, we have built and made publicly available a novel dataset to enable the evaluation of RS for the particular setting of recommending recently published papers to users. In this sense, we provide the following contributions:

- An in-depth analysis of existing state of the art methods and datasets for the recommendation of scientific papers.
- A novel hybrid recommendation method that exploits citation knowledge in order to suggest novel and relevant items to users in a personalised way.
- A rigorous evaluation of the proposed method against multiple baselines following a time-based data split.
- A unique dataset that includes the publication history of users as well as the textual content of scientific papers.

The remainder of the manuscript is structured as follows. Section II surveys the state of the art on academic papers RS, including both existing methods and datasets. Section III describes the building of our dataset, and how citation knowledge has been captured from it. Section IV presents the proposed recommendation method and baselines, and Section V reports conducted experiments and achieved results. Lastly, Section VI concludes the work and discusses main findings, limitations and open research issues.

TABLE I
PUBLICLY AVAILABLE DATASETS FOR ACADEMIC RECOMMENDER SYSTEMS. PDF_{av} STANDS FOR PDF DOCUMENT AVAILABILITY, AND UPH_{av} STANDS FOR AUTHORS' PUBLICATION HISTORY AVAILABILITY

Dataset	Description	Users	Items	Ratings	PDF_{av}	UPH_{av}
AMiner [13]	AMiner contains a series of datasets capturing relations among citations, academic social networks, topics, etc. We report data here about the citations dataset V11	Not specified	4M	No	No	No
Open Citations [14]	Open repository of scholarly citation data	Not specified	7.5M	No	No	No
Open Academic Graph [15]	Large knowledge graph combining Microsoft Academic Graph and AMiner	253M	381M	No	No	No
ArXiv [16]	Open access e-prints papers in different fields such as physics, mathematics etc.	Not specified	1,5M	No	Yes	No
CORE [17]	Dataset of open access research publications published up to 2018	No	9.8M	No	Yes	No
CiteULike [18]	Dataset of users' selected bookmarks to academic papers	5,551	16,980	No	No	No
Mendeley [19]	Dataset shared by Mendely for a recommender system challenge	50,000	4.8M	Yes	No	No
ACL anthology [20]	Corpus of scholarly publications about Computational Linguistics	Not specified	22,878	No	Yes	No
SPD 1 [21]	ACL anthology based papers published between 2000-2006	28	597	Yes	Yes	No
SPD 2 [22]	ACM proceedings based papers published between 2000-2010	50	100,531	Yes	No	No

II. RELATED WORK

Approaches that have dealt with the problem of recommending scientific publications (also referred as research papers) can be categorised based on how user preferences are modelled, how item features are captured, and which recommendation methods are applied considering both the user's preferences and the items' features.

User preferences can be captured by considering explicit and implicit feedback. Approaches based on explicit feedback collect explicit preferences from the user (e.g., ratings) in order to build user profiles. On the contrary, approaches based on implicit feedback capture implicit information (e.g., from browse, click) to model a user's profile. It is important to note that when limited information exists, such as authors who have published few papers or do not have many logged activities within the system, user profiles may be incomplete and inadequate to provide accurate recommendations [23]. Hence, the use of citation knowledge may be helpful to create more complete user profiles for the recommendations.

Item features can be captured by considering metadata such as title, abstract, publication year, bibliography (i.e. the list of publications that are cited in a paper) etc. and also the textual content of a paper, including *citation-context* [10], [23]. Due to the inaccessibility of the full content of papers, very few works up to date have exploited the notion of citation context to provide recommendations (see Section III).

Different user and item representations –e.g., vectors, matrices, and knowledge bases– are built to gather and exploit user preferences and item features, and **recommendation methods** are designed based on such representations. Among the popular approaches, we can highlight content-based (**CB**), collaborative filtering (**CF**), hybrid (**H**), and, graph-based (**GB**). Content-based approaches recommend a user (author), items (papers) that are *similar* to those they *liked* in the past. Collaborative Filtering approaches recommend a user, items that are preferred by like-minded users. Hybrid approaches jointly exploit multiple approaches. Finally, graph-based approaches utilise the relations that exist between authors, publications, venues, etc. The reader is referred to [3], [24], which present recent surveys for the all these types of approaches.

In this work, we focus on capturing information about authors and papers, and on exploring the use of citation knowledge to recommend recent and relevant publications to authors. Hence, publication-time awareness and citation-knowledge are two key aspects of the literature to consider.

Regarding the **publication time awareness**, while the concept of time has previously used in RS to better define and delimit long-term vs. short-term user preferences [23], and to suggest papers to users with no previous activity (i.e. new users) [25]. To the best of our knowledge, only [12] has addressed the problem of recommending the most recently published scientific work (i.e. new papers). It proposes a graph-based Belief Propagation approach that recommends a list of new papers for a target user. An undirected graph is built to capture relations among papers based on citations; the graph does not distinguish between 'cites' and 'being cited' inverse relations. The authors experimented with a dataset which is not available and details on where and how it could be collected were not given. [26] worked on the related problem of 'out-of-matrix prediction. They proposed collabborative topic regression (CTR) model, which combines CF with topic modelling. CTR is compared against LDA and Matrix Factorisation (MF), where CTR and LDA achieve a relatively lower recall, and MF is unable to provide recommendations. Their evaluation was conducted on a selection of CiteULike data which is also not available.

Regarding the **citation knowledge awareness**, existing approaches that capture and exploit citation knowledge have focused on recommending relevant papers for a target paper [7], [27], or a manuscript [10]. Fewer works, in contrast, have focused on recommending relevant papers for a user [5], [23]. To the best of our knowledge, none of them have addressed the use case of recommending recently published papers.

Sugiyama et al. [23] explored the content of the publications citing the user's work, including: (i) citation context, since it may be viewed as an endorsement of the work and, (ii) textual content from other sections to complement citation context, for enriching users' profiles. [5] explored the use of the citation graph from CiteSeer data to provide recommendations.

While these works show how citation knowledge can help enhancing recommendation performance, they do not explore the use of citation knowledge on the real-world scenario of recommending the latest scientific publications to users. In addition to exploring the use of citation knowledge in this scenario, our work proposes a more comprehensive view

of citation knowledge including: the citation graph, citation context and citation section.

III. Dataset building

Multiple datasets are available to evaluate RS for academic papers. We provide a comprehensive list of publicly available datasets in Table I. The table shows a brief description of the type of data, the number of users, items and ratings, and availability of the full text of papers PDF_{av} and publication history of the users UPH_{av}. However, these datasets have several limitations such as unavailability of full texts and knowledge about the authors and their publication histories. To address such limitations, we built a new dataset that includes the textual content of papers and the authors' publication histories and is publicly available [28]. Next, we describe our dataset building process.

A. Collecting Data

We aimed to build a new dataset that provides the textual content of papers from which fine-grained citation knowledge could be extracted. Since we are interested in exploring the usage of citation knowledge for recommendations, we needed to ensure that there are sufficient papers cited by other papers within the dataset. Following this requirement, we gathered the publication history of authors working on the same field (e.g., publishing in the same conference), since they are likely to cite each other's publications. Specifically, we selected the ACM Conference Series on Recommender Systems (RecSys) and collected data for 1,931 authors who have published in the conference between 2007 and 2018. The complete publication histories of the authors were collected from the bibliography data provider DBLP[3]. Note that, the publication history of an author contains not only their RecSys papers but also papers published in other venues (journals, conferences, etc.) and collectively 1,931 authors have published 80,808 papers.

While initiatives like open access enabled full access to many scientific publications, many of them are still hidden behind pay-walls and thus are not publicly accessible[4]. As a consequence, we only obtained textual content for 35,473 out of the 80,808 papers. To ensure that we had sufficient historical data to capture user preferences, we discarded authors for which we obtained less than 4 publications, keeping a total of 1,336 authors.

We then divided the dataset into training and test sets by observing the publication time distribution, and selected the 1st of January 2018 as split date (see Figure 1). All papers published before that date were considered part of the training set and after that date were considered the test set. Lastly, we kept those authors having at least 60% of the data in the training set, and at least 10% in the test set. The final dataset consists of 547 authors and 15,174 papers, from which 12,641 belong to the training set and 2,533 represent the test set.

[3]https://dblp.uni-trier.de/

[4]https://www.theguardian.com/higher-education-network/2018/may/21/scientists-access-journals-researcher-article

B. Modelling Papers, Citations and Author Preferences

Figure 2 shows the different features captured for authors, publications and citations, as well as their relations. For each author, we capture information such as name, affiliation, identifiers etc. For each paper, we capture metadata and identifiers (– DBLP URL, Google Scholar URL and internal identifier within the dataset), and the Google Scholar URL is used to download the PDF of publications, if available.

We then parsed the available PDF files using the GROBID parser[5] and extract citation knowledge. From each publication, we have extracted: (i) the bibliography, (ii) the sections within the publication where citations appear (introduction, state of the art, conclusion and other sections) and (iii) the citation context. We consider citation context as three sentences: the one where the citation appears, and the ones before and after, when available. The reference lists are then matched against the 15,174 publications of the dataset to identify the citation-based relations and generate the citation graph. A series of heuristics are adopted to minimise errors including applying lower case, matching at least one author, and computing the Levenshtein distance between the title of the publication and the title of the reference where an 85% minimum threshold was empirically selected. These heuristics are needed to discard the references containing errors or incomplete information. In total, we identified 1,806 distinct referenced publications cited 4,358 times in the introduction sections, 3,999 in the related work sections, 82 in the conclusion sections, and 12,213 in other sections within our dataset.

When the publication history of an author (user) is sparse, the data may be insufficient to build a reliable profile for personalised recommendation. Then, relying on citation information could help enrich their profiles. Hence, we distinguish between two main ways of capturing user preferences, where we consider that an author has a preference for all of their authored as well as their cited publications, since those publications can encapsulate research that the author considers relevant in relation to their works. Figure 1 illustrates these two preference models where the left part shows a rating matrix R_P relating authors (rows) and papers (columns) where a cell has a value 1 if the corresponding author authored the associated paper, and 0 otherwise, and the right part shows a rating matrix R_{PC} where a cell has a value 1 if the user authored or cited the paper, and 0 otherwise. In addition, we also consider an enriched version of R_{PC}, R_{PCX}, where X stands for context (i.e. text around a citation while citing). The above three matrices are split into training and test sets according to a target time, 01/01/2018, producing the following data splits:

- $R_P^{training}$: 547 users, $12,641$ items and $14,555$ ratings
- R_P^{test} : 547 users, $2,533$ items and $3,082$ ratings
- $R_{PC}^{training}$547 users, $12,641$ items and $20,756$ ratings
- R_{PC}^{test} : 547 users, $2,533$ items and $3,233$ ratings
- $R_{PCX}^{training}$: 547 users, $12,641$ items, $20,756$ ratings
- R_{PCX}^{test} : 547 users, $2,533$ items and $3,233$ ratings

[5]https://github.com/kermitt2/grobid

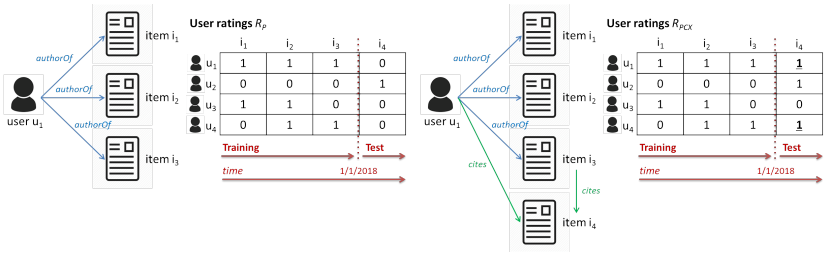

Fig. 1. Modelling user preferences

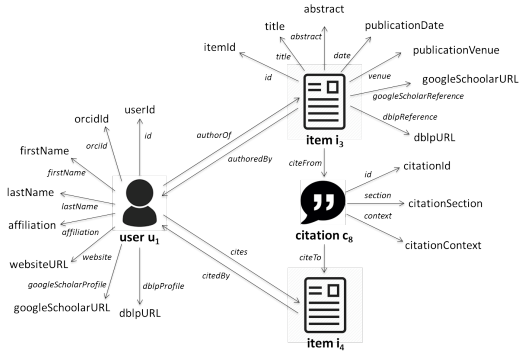

Fig. 2. Capturing citation knowledge

IV. Recommendation Methods

This section describes our proposed hybrid recommendation approach (see Section IV-B), and the baselines used for the comparison. We considered several baselines including content-based (see Section IV-A), graph-based (including PageRank [29] and ItemRank [30]), MF [31], Factorisation Machine (FM) [32] and the Random method from the RankSys framework. Since content-based methods achieved significantly better performance results than the above mentioned baselines, in our analysis, we focus on content-based methods(see Table II). We note that collaborative filtering methods are not reported, since they are unable to provide recommendations in the addressed use case, where new (i.e. not previously seen/rated) papers are the ones to be recommended.

A. Content-based Recommender Systems

Content-based filtering methods recommend items (papers) to a user that are 'similar' to those they positively rated (i.e. authored or cited). The similarity between users and items is computed based on profiles built from textual information. The recommendation score of an item for a target user is then computed as the cosine similarity between the profiles of user and item. We refer this method as **cb**. The features used to model users' profiles for **cb** varies according to the available citation knowledge (see Section III-B). For R_P, a user's profile is built by considering the titles of the papers they authored. For R_{PC}, a user's profile is built by using the titles of the papers they authored and also cited. Lastly, for R_{PCX}, a user's profile is built by considering the titles of the papers they authored and cited, as well as the citation context.

B. Hybrid Recommender Systems

Next, we present our hybrid recommendation approach, which jointly exploits the content of the papers and the user-item ratings to provide personalised recommendations. Hybrid methods [33] aim to mitigate the disadvantages of individual approaches by combining the strengths of various methods. Here, we aim to mitigate the ineffectiveness of CF when recommending the latest scientific publications by combining it with **cb** and exploiting the captured citation knowledge.

The proposed approach, **hyb**, is based on the item-based nearest neighbour CF heuristic[6] where content features are used to compute item similarities. In item-based CF [34], similarities between items are used to estimate scores for a (user, item) pair. In our case, item profiles are generated based on textual features, where the features vary with respect to the available citation knowledge (see Section III-B): for R_P, R_{PC} and R_{PCX}. We formulate our **hyb** method in Equation (1):

$$\hat{r}_{u,i} = \frac{\sum_{i \epsilon N(i')} Sim(i,i').r_{u,i'}}{\sum_{i \epsilon N(i')} |Sim(i,i')|} \quad (1)$$

where $\hat{r}_{u,i}$ is the preference score to be predicted for the target user u and item i, $Sim(i,i')$ is the similarity between the interacted item i' and an item i from the neighbourhood $N(i')$ of item i'. Cosine similarity is used to measure the similarity between items. Finally, $r_{u,i'}$ is the preference (rating) given by user u to the item i'. We also use different sizes of neighbours, specifically 5, 10, 15 and 20.

To investigate the relevance of citation section, we further modified our hybrid method i.e. Equation (1) by incorporating a weight, $w_{u,i}$, that reflects the strength of an item i for a user u based on the different sections where u cites i in their publications. We refer it as **hybSec** and formulate in Equations (2) and (3).

$$\hat{r}_{u,i} = \frac{\sum_{i \epsilon N(i')} Sim(i,i').r_{u,i'}.w_{u,i'}}{\sum_{i \epsilon N(i')} |Sim(i,i')|} \quad (2)$$

where the strength (weight) $w_{u,i'}$ is computed by considering all the instances where i' is cited by u; Note that an item i' may be cited by u in several publications, and in different sections of the same publication. Then, the weight is normalised by the total number of instances. More formally, the strength $w_{u,i'}$ of item i' for user u is calculated as:

$$w_{u,i'} = \frac{\sum_{j=1}^{n}(w_{j_{int}} + w_{j_{relWork}} + w_{j_{concl}} + w_{j_{others}})}{n_{u,i'}} \quad (3)$$

[6]We also tested the user-based CF heuristic, but discarded it due to its non competitive performance results

where $n_{u,i'}$ is the number of times i' is cited by u in the user's papers, and $w_{j_{int}}$, $w_{j_{relWork}}$, $w_{j_{concl}}$ and $w_{j_{others}}$ reflect the number of times i' is cited in the introduction, related work, conclusion or other sections respectively.

V. Experiments

This section reports the experiments conducted to evaluate our proposal to citation knowledge exploitation for personalised recommendation of recently published papers. In Section V-A, we discuss the evaluation methodology and metrics used in our experiments. We then summarise the settings of the proposed methods and baselines in Section V-B. Lastly, Section V-C discusses the obtained performance results referring to Table II, which shows the results for all rating matrices, recommendation methods and evaluation metrics.

A. Evaluation Methodology and Metrics

As explained in Section III-B, we model user preferences and capture citation knowledge through the rating matrices R_P, R_{PC} and R_{PCX}. Moreover, we perform a time-based split of these matrices to generate training and test sets to evaluate our proposed recommendation approach. The selected date for the split, 1/1/2018, has been allocated according to the publication time distribution of the papers in the dataset, and ensuring that, for each user, at least 60% of the data falls into training and 10% falls in to test set.

We focus on ranking-based metrics, in particular, we compute precision, recall, F1 measure, Mean Average Precision (MAP) and Normalised Discounted Cumulative Gain (nDCG), which favour the accuracy of the first items within the recommendation list. For each metric, we consider the top 5 and 10 cutoffs, addressing the scenario where the target user is recommended with a limited list of items. All the metrics were computed using the RiVal evaluation framework[7].

B. Recommendation Methods

The recommendation methods, including the proposed hybrids (**hyb** and **hybSec**), and the content-based (**cb**) baselines (see Section IV-A). For our methods, **hyb** and **hybSec**, we have tested different neighbourhood sizes (5, 10, 15 and 20). Due to the space limitation, we only report results of **hybSec** using neighbourhoods of size 5 with R_{PCX}. We also experimented with 10, 15 and 20 neighbours, but obtained worse results. Moreover, we evaluated a large number of configurations of **hybSec** in terms of the weights $W_{int}, W_{relWork}, W_{concl}$, W_{others} associated to the sections where citations appear, namely introduction, related work, conclusion, and other sections respectively. We applied a grid search for the optimal values of such weights. For clarity purposes, we only show representative configurations in Table II.

C. Results

Table II summarises results obtained in our experiments. The first conclusion, we derive is the fact that incorporating knowledge from the *citation graph* into the original author-publication matrix R_P entails an improvement of the generated recommendations for hybrid methods for all metrics, but only when the number of neighbours is higher than 10. The less rating sparsity of the R_{PC} matrix allows finding valuable item similarities and relations that are effectively exploited when more than 10 neighbours are considered.

When incorporating citation contexts knowledge, captured in the R_{PCX} matrix, we achieve further improvements on our **hyb** approach with all neighbourhood sizes (i.e. 5, 10, 15 and 20), for all metrics over both the matrices R_P and the R_{PC}. The best results for all metrics is obtained when considering a size of 5 neighbours. This indicates that citation context is a prominent feature to enhance recommendations in our given setting. However, when adding *citation section* knowledge on the R_{PCX} matrix, our **hybSec** approach does not outperform its **hyb** counterpart. This indicates that the section of the citation within the paper may not be a relevant feature to enhance recommendations in the studied scenario.

We are targeting a particular difficult scenario (– new item recommendations) where items in the test set do not have any connections to items in the training set, hence Collaborative Filtering (CF) methods do not work, and some of the studied baselines (see Section IV) performed very poorly. It is, however, promising to observe how, in this scenario, the use of citation knowledge, and more particularly the use of the citation-graph and citation context, can help providing more accurate recommendations to users.

VI. Discussion and Conclusions

We have addressed the problem of providing personalised recommendations of recently published papers. For this problem, CF approaches cannot be used since they are unable to establish rating-based similarities and patterns between new items. Motivated by this fact, in addition to content-based features, we advocate for the exploitation of citations and its related knowledge as a bridge to connect related papers.

Note that, while **citation knowledge** has been explored in the literature to provide paper recommendations in different scenarios, our work brings two key novelties with respect to previous works: i) a real-world and challenging scenario, where new papers are to be recommended and, ii) an exploration of a wider notion of citation knowledge, which includes the citation graph, citation context and citation section.

In particular, we have presented a hybrid approach that makes use of the citation graph to enrich the rating matrix, while exploring the use of the citation context and citation section for recommendations. Our experimental results show that incorporating citation knowledge in terms of the citation graph and citation context (–hyb) allows for effective new paper recommendations, while the incorporation of citation section (–hybSec) could not outperform our hyb method in this particular setting. It is also important to note that, while we have implemented multiple baselines (including PageRank, ItemRank, MF, FM, and the Random method) to compare against our proposed hybrid methods, the results obtained with these baselines were significantly worst than the ones achieved by content-based methods. Our hypothesis is that the targeted

[7]http://rival.recommenders.net/

TABLE II

EXPERIMENT RESULTS OF THE BASELINES AND PROPOSED HYBRID RECOMMENDATION METHODS. A GRAY SCALE IS USED TO HIGHLIGHT BETTER (DARK GRAY) AND WORST (WHITE) VALUES FOR EACH METRIC (COLUMN). BEST VALUES ARE IN BOLD FOR EACH METRIC

matrix	method	p@5	p@10	r@5	r@10	F1@5	F1@10	MAP@5	MAP@10	nDCG@5	nDCG@10
R_P	cb	0.054	0.039	0.063	0.088	0.058	0.054	0.044	0.048	0.081	0.084
	hyb5	0.056	0.040	0.071	0.090	0.062	0.055	0.044	0.047	0.080	0.083
	hyb10	0.059	0.039	0.072	0.093	0.065	0.055	0.044	0.048	0.084	0.084
	hyb15	0.055	0.039	0.068	0.089	0.061	0.054	0.041	0.045	0.078	0.081
	hyb20	0.050	0.038	0.064	0.086	0.056	0.052	0.041	0.046	0.076	0.081
R_{PC}	cb	0.052	0.041	0.062	0.091	0.056	0.056	0.042	0.047	0.077	0.084
	hyb5	0.052	0.040	0.066	0.095	0.058	0.056	0.043	0.049	0.078	0.085
	hyb10	0.056	0.038	0.068	0.091	0.062	0.053	0.044	0.048	0.081	0.083
	hyb15	0.055	0.040	0.068	0.090	0.061	0.055	0.044	0.048	0.081	0.085
	hyb20	0.055	0.041	0.067	0.092	0.060	0.057	0.044	0.049	0.081	0.086
R_{PCX}	cb	0.029	0.023	0.041	0.061	0.034	0.033	0.025	0.029	0.043	0.050
	hyb5	**0.065**	**0.048**	0.076	**0.108**	**0.070**	**0.066**	**0.053**	**0.060**	**0.095**	**0.102**
	hyb10	0.062	0.047	**0.077**	0.106	0.069	0.065	0.052	0.058	0.093	0.098
	hyb15	0.062	0.045	0.076	0.105	0.068	0.063	0.051	0.056	0.091	0.096
	hyb20	0.060	0.044	0.073	0.104	0.066	0.062	0.049	0.055	0.088	0.095
	hybSec5 (1, 0, 0, 0)	0.059	0.044	0.074	0.101	0.065	0.061	0.050	0.055	0.086	0.093
	hybSec5 (0, 1, 0, 0)	0.058	0.044	0.072	0.103	0.064	0.062	0.050	0.056	0.086	0.094
	hybSec5 (0, 0, 1, 0)	0.058	0.043	0.070	0.098	0.063	0.060	0.049	0.054	0.085	0.091
	hybSec5 (0, 0, 0, 1)	0.059	0.045	0.073	0.103	0.065	0.063	0.050	0.056	0.086	0.095
	hybSec5 (0.5, 0.5, 0, 0)	0.057	0.043	0.069	0.097	0.062	0.059	0.048	0.054	0.083	0.090
	hybSec5 (0, 0.5, 0.5, 0)	0.057	0.043	0.069	0.097	0.063	0.059	0.048	0.054	0.083	0.090
	hybSec5 (0, 0, 0.5, 0.5)	0.056	0.044	0.068	0.100	0.061	0.061	0.048	0.054	0.082	0.091
	hybSec5 (0.5, 0, 0, 0.5)	0.055	0.043	0.069	0.100	0.061	0.060	0.048	0.054	0.082	0.091
	hybSec5 (0.25, 0.25, 0.25, 0.25)	0.054	0.041	0.066	0.094	0.059	0.057	0.047	0.052	0.080	0.086

scenario poses significant challenges for these methods, since there is no connections between items in the test and training sets. This problem was also highlighted by [26].

Moreover, while existing services, such as Google Scholar, do have their own RS to provide paper recommendations, the methods behind these systems are not public and hence it has not been possible for us to replicate them as baselines. Comparisons against these systems could be conducted by means of user studies, which is one of our future work.

Understanding the semantics of the text surrounding the citations (e.g., whether a user is criticising or praising within a citation) and capturing the experience and current goals of the target user (e.g., a senior researcher vs. a PhD student) are also part of our future research work.

REFERENCES

[1] F. Wang, N. Shi, and B. Chen, "A comprehensive survey of the reviewer assignment problem," *ITDM*, vol. 9, no. 04, pp. 645–668, 2010.

[2] J. E. Hirsch, "An index to quantify an individual's scientific research output," *National academy of Sciences*, pp. 16 569–16 572, 2005.

[3] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recommender systems:a literature survey," *IJDLS*, vol. 17, pp. 305–338, 2016.

[4] S. E. Middleton, D. C. De Roure, and N. R. Shadbolt, "Capturing knowledge of user preferences: Ontologies in recommender systems," in *1st K-CAP*, 2001, pp. 100–107.

[5] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl, "Enhancing digital libraries with techlens+," in *4th JCDL*, 2004, pp. 228–236.

[6] Y. Liang, Q. Li, and T. Qian, "Finding relevant papers based on citation relations," in *WAIM*, 2011.

[7] A. Khadka and P. Knoth, "Using citation-context to reduce topic drifting on pure citation-based recommendation," in *12th RecSys*, 2018, pp. 362–366.

[8] S. Bethard and D. Jurafsky, "Who should i cite:learning literature search models from citation behavior," in *19th CIKM*, 2010, pp. 609–618.

[9] M. D. Ekstrand, P. Kannan, J. A. Stemper, J. T. Butler, J. A. Konstan, and J. T. Riedl, "Automatically building research reading lists," in *4th ACM Conference on Recommender Systems*, 2010, pp. 159–166.

[10] Q. He, J. Pei, K. Daniel, M. Prasenjit, and L. Giles, "Context-aware citation recommendation," 2010, pp. 421–430.

[11] T. Strohman, W. B. Croft, and D. Jensen, "Recommending citations for academic papers," in *30th SIGIR*, 2007, pp. 705–706.

[12] J. Ha, S.-H. Kwon, S.-W. Kim, and D. Lee, "Recommendation of newly published research papers using belief propagation," in *2014 Conference on Research in Adaptive and Convergent Systems*, 2014, pp. 77–81.

[13] "Aminer," https://www.aminer.cn/aminer_data, 2019.

[14] "Open citations," https://opencitations.net/corpus, 2019.

[15] "Open academic graph," https://www.openacademic.ai/oag/, 2019.

[16] "Arxiv," https://arxiv.org/help/bulk_data, 2019.

[17] "Core," https://core.ac.uk/services/dataset/, 2019.

[18] "Citeulike," https://old.datahub.io/dataset/citeulike, 2019.

[19] K. Jack, J. Hammerton, D. Harvey, J. J. Hoyt, J. Reichelt, and V. Henning, "Mendeleys reply to the datatel challenge," *Procedia Computer Science*, vol. 1, pp. 1–3, 2010.

[20] "Acl anthology," https://acl-arc.comp.nus.edu.sg/, 2019.

[21] "Spd 1," https://acl-arc.comp.nus.edu.sg/, 2019.

[22] "Spd 2," ttps://www.comp.nus.edu.sg/~sugiyama/Dataset2.html, 2019.

[23] K. Sugiyama and M. Y. Kan, "Exploiting potential citation papers in scholarly paper recommendation," in *13th JCDL*, 2013, pp. 153–162.

[24] Z. Y. X. K. F. X. Wei Wang, Jiaying Liu, "Scientific paper recommendation: A survey," *IEEE Access*, vol. 7, pp. 9324–9339, 2019.

[25] M. Hristakeva, D. Kershaw, M. Rossetti, P. Knoth, B. Pettit, S. Vargas, and K. Jack, "Building recommender systems for scholarly information," in *1st Workshop on Scholarly Web Mining*, 2017, pp. 25–32.

[26] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *17th SIGKDD*, 2011, pp. 448–456.

[27] S. M. McNee, I. Albert, D. Cosley, P. Gopalkrishnan, S. K. Lam, A. M. Rashid, J. A. Konstan, and J. Riedl, "On the recommending of citations for research papers," in *CSWC*, 2002.

[28] A. Khadka, I. Cantador, and M. Fernandez, "Citation Knowledge with Section and Context Dataset," 5 2020. [Online]. Available: https://ordo.open.ac.uk/articles/Citation_Knowledge_with_Section_and_Context/11346848

[29] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," in *Proceedings of the 7th WWW*, 1998, pp. 107–117.

[30] M. Gori and A. Pucci, "Itemrank:a random-walk based scoring algorithm for recommender engines," in *20th IJCAI*, 2007, pp. 2766–2771.

[31] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *2008 IEEE ICDM*, 2008, pp. 263–272.

[32] S. Rendle, "Factorization machines," in *Proceedings of the 2010 ICDM*, ser. ICDM '10. IEEE Computer Society, 2010, pp. 995–1000.

[33] R. Burke, "Hybrid recommender systems: Survey and experiments," *UMUAI*, vol. 12, no. 4, pp. 331–370, 2002.

[34] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *10th WWW*, 2001, pp. 285–295.