

Exploiting Potential Citation Papers in Scholarly Paper Recommendation

Kazunari Sugiyama
National University of Singapore
Computing 1, 13 Computing Drive,
Singapore 117417
sugiyama@comp.nus.edu.sg

Min-Yen Kan^{1,2}
¹National University of Singapore
²NUS Interactive and Digital Media Institute
Computing 1, 13 Computing Drive,
Singapore 117417
kanmy@comp.nus.edu.sg

ABSTRACT

To help generate relevant suggestions for researchers, recommendation systems have started to leverage the latent interests in the publication profiles of the researchers themselves. While using such a publication citation network has been shown to enhance performance, the network is often sparse, making recommendation difficult. To alleviate this sparsity, we identify “potential citation papers” through the use of collaborative filtering. Also, as different logical sections of a paper have different significance, as a secondary contribution, we investigate which sections of papers can be leveraged to represent papers effectively.

On a scholarly paper recommendation dataset, we show that recommendation accuracy significantly outperforms state-of-the-art recommendation baselines as measured by nDCG and MRR, when we discover potential citation papers using imputed similarities via collaborative filtering and represent candidate papers using both the full text and assigning more weight to the conclusion sections.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: Information filtering, Search process; H.3.7 [Digital Libraries]: Systems issues, User issues

General Terms

Algorithms, Experimentation, Performance

Keywords

Digital library, Information retrieval, Recommendation, Citation analysis, Collaborative filtering

1. INTRODUCTION

Newly discovered knowledge is now largely captured in digital form and archived throughout the world. Archival materials are also being digitized and are increasingly becoming more accessible online. The modern researcher has unprecedented level of access

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

JCDL'13, July 22–26, 2013, Indianapolis, Indiana, USA.

Copyright 2013 ACM 978-1-4503-2077-1/13/07 ...\$15.00.

to the sum total of human knowledge. While certainly advantageous, this creates a problem of over abundance, commonly known as “information overload”: where researchers find an overwhelming number of matches to their search queries, but for which the majority are largely irrelevant to their latent information needs.

Work in recommendation systems is one promising approach to address the information overload. In digital library studies, this approach has been employed to obtain and refine search results to satisfy each user’s information needs [10, 18, 3, 23, 13]. However, these approaches do not fully leverage the user’s context, largely relying on the idea of session-as-context. This legacy is ported from research in Web search, where session click-through data are used to form the context. To address this problem, in our previous work, we observed that the scholarly context allows us to leverage the role of the searcher-as-author [16]. We modeled a searcher’s context in the form of a profile by capturing previous research interests embodied in their past publications, and showed elevated success at scholarly paper recommendation. Our approach in [16] also took advantage of the explicit citation network of publications as a source of knowledge to improve recommendation accuracy. The contents of papers that cite an author’s papers as well as the contents of the works referenced in the papers provide supplementary evidence used in modeling the author’s research interests.

In this paper, we propose two extensions of [16] that further mine additional signals from the full text and citation network – using (1) potentially cited papers and (2) their fragments. Citation papers are papers that explicitly cite previous work and often contain a summary of its salient points. Such citation papers may be viewed as an endorsement of the cited paper, and they may help model the target paper more accurately. In addition, fragments are parts of a paper such as abstract, introduction, conclusion, and so on.

Authors of papers also may not cite certain relevant papers in their publications, either purposefully (*e.g.*, to save space) or not (*e.g.*, were unaware of the specific relevant work). If we enhance the citation network with such potentially citable papers (hereafter, *pc*), we hypothesize that we can model the target papers to recommend more accurately to achieve better recommendation performance. In our work, we apply collaborative filtering (CF) to find such potential citation papers. While CF is often used to recommend items to users directly, we apply CF to discover potential citation papers that help in representing target papers to recommend.

Through a series of experiments on a scholarly paper recommendation dataset, we show that proper modeling of potential citation papers – as well as properly representing papers with both their full text and assigning more weight to the conclusion – improve recommendation accuracy significantly ($p < 0.05$ or better) as judged by both mean reciprocal rank (MRR) and normalized discounted

cumulative gain (nDCG). We also show that our approach can outperform state-of-the-art scholarly paper recommendation systems.

2. RELATED WORK

As the field of recommendation systems is large, we focus our literature review on systems for scholarly paper recommendation. In addition, as finding potential papers can be viewed as a type of link discovery, we also briefly review two types of work, citation link prediction and content link detection.

2.1 Scholarly Paper Recommendation

With respect to scholarly paper recommendation, McNee *et al.* [10] proposed an approach to recommending citations. Their approach applied collaborative filtering (CF) to social networks to create a graph formed by the citations between research papers. This data can be mapped into a framework of CF and used to overcome the cold-start problem. Expanding this approach, Torres *et al.* [18] proposed a method for recommending research papers by combining CF and content-based filtering (CBF). However, a single final ranking obtainable by merging the output from both CF and CBF is purposefully not done, as the authors claim that pure recommendation algorithms are not designed to receive input from another recommendation algorithm. Gori and Pucci [3] devised a PageRank-based method for recommending research papers. But in their approach, a user must prepare an initial set of relevant articles to obtain better recommendations, and the damping factor d that affects the score of PageRank is not optimized. Yang *et al.* [23] presented a scholarly paper recommendation system using a ranking-oriented CF. Although their system overcomes the cold-start problem by utilizing implicit behaviors extracted from a user’s access logs, the predefined settings for parameters used to select effective data are not justified nor investigated in detail. In recent work, Nascimento *et al.* [13] developed a scholarly paper recommendation system, in which they use the title to construct user profiles, and the title and abstract to generate feature vectors of candidate papers to recommend. However, we feel that such a small span of text does not effectively represent a user’s interest and candidate papers.

Scholarly paper recommendation studies are also emerging in data mining. Wang and Blei [20] proposed collaborative topic regression model which combines ideas from CF and content analysis based on probabilistic topic modeling. They used the abstract and title of the paper to model a user and characterize candidate papers to recommend, which occasionally results in irrelevant recommendations, similar to [13]. El-Arini and Guestrin [2] proposed a method for discovering a small set of scholarly papers that are relevant to a query yet diversified. They defined “influence” to capture the transfer of ideas as individual concepts among papers in the query. Their approach then returns papers related to these concepts. However, users need to prepare trusted papers in advance to discover relevant and diverse papers.

2.2 Link Discovery

Citation Link Prediction. Researchers can benefit from a citation recommendation system because searching for the right paper to cite is a laborious task. Strohman *et al.* [17] experimented with a citation recommendation system where the relevance between two documents is measured by a linear combination of text features and citation graph features. They concluded that similarity between bibliographies of documents and Katz distance [9] are the most important features. He *et al.* [5] developed a citation recommendation system based on a non-parametric probabilistic model. Their system requires a user to prepare query manuscript without a bibliography that indicates locations where citations are

needed, resulting in additional burden for the user. In their subsequent work, they solved this problem by automatically analyzing the query manuscript to suggest locations where citations are needed [4].

Content Link Detection. Content link detection aims to discover similar content across different input and make such links explicit. In Wikipedia link detection, Milne and Witten [12] created explanatory links to all documents using supervised machine learning. They observed that decision tree generator gives better results than other learning techniques. West *et al.* [21] addressed the same task using unsupervised learning through principal component analysis. Following these studies, Kaptein *et al.* [8] proposed finding links from Wikipedia pages to external Web pages by using a language modeling approach. In story link detection, Nomoto [14] proposed a two-tier model of similarity, at both the document and collection levels. His similarity model adapted the idea of relevance feedback to link detection, where stories are measured for similarity not merely based on the document, but on a collection of relevant documents. Finally, by combining two algorithms proposed in [12] and [21], West *et al.* [22] created a hybrid algorithm that suggests topics to authors of text documents.

3. PROPOSED METHOD

Our work tackles the core problem of matching users to candidate papers. Unlike existing scholarly paper recommendation systems which focused on user profile construction [3, 16, 13], our work leverages the scholarly papers more effectively, through the modeling of potential citation papers and their fragments, and enhancing the citation network with automatically identified potential citation papers. Unlike citation link prediction, we apply collaborative filtering to discover potential citation papers that help model target papers to recommend. And unlike previous work in content link detection which mainly focused on finding an effective learning framework, we focus on how to best use the scholarly corpora available to us.

3.1 Baseline System [16]

Our method starts with our former scholarly paper recommendation system [16], and as such it is instructive to first describe our system and its basis. It consists of three steps:

- Step 1: Construct a user profile \mathbf{P}_{user} from a researcher’s list of published papers;
- Step 2: Compute feature vectors \mathbf{F}^{pj} ($j = 1, \dots, t$) for each of the papers in its scholarly paper knowledge base;
- Step 3: Compute the cosine similarity $Sim(\mathbf{P}_{user}, \mathbf{F}^{pj})$ between \mathbf{P}_{user} and \mathbf{F}^{pj} ($j = 1, \dots, t$), and recommend papers with high similarity to the target user.

A candidate paper to recommend (p) is represented as a feature vector \mathbf{f}^p . We employ TF and TF-IDF schemes in Steps 1 and 2, respectively. Both \mathbf{P}_{user} and \mathbf{F}^{pj} are constructed as the combination of \mathbf{f}^p as defined by Equation (1). As such, our method views both user profiles and candidate papers to recommend as vectors of terms with specific, per-term tuned weights. As CBF relies on the item’s content to provide its recommendations, it is important to represent an item’s contents faithfully. A key innovative step in [16] is to model a target paper of interest based on not merely its own textual content but also an appropriately weighted inclusion of the text from its context as defined by the neighborhood of scholarly works it referenced, as well as those works that cite it (see Figure 1 (a)). In this past work, we found that when the text of such

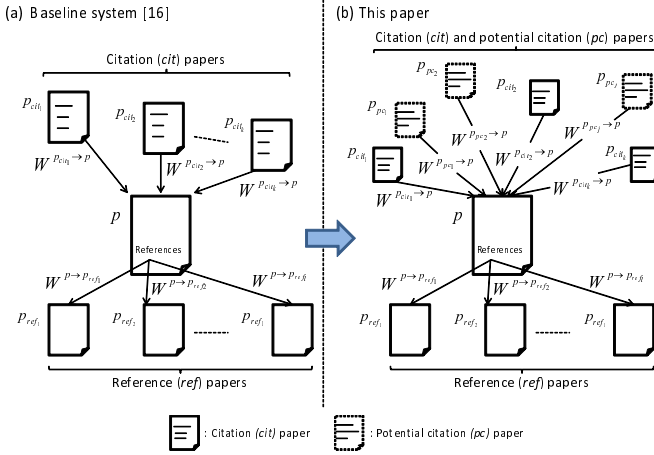


Figure 1: Comparison of paper representations between our former work [16] and the current paper (notations simplified from [16]). Our work in this paper leverages additional potential citation papers to enrich the description of a target paper.

contextual papers is added to the original target paper weighted by cosine similarity to the target paper, recommendation accuracy is improved the most among other alternatives explored.

In the current paper, we further enhance Step 2 above, by both **enlarging** what is meant by context through the discovery of potential citation papers, as well as **refining** its use in specific, well-linked parts of the contextual documents through the specific modeling of potential citation papers and their fragments (Figure 1(b)).

To facilitate our discussion later, we show the formula for Step 2 defined in [16] to compute the feature vector for each paper p :

$$\mathbf{F}^p = \mathbf{f}^p + \sum_{x=1}^k W^{p_{cit_x} \rightarrow p} \mathbf{f}^{p_{cit_x}} + \sum_{y=1}^l W^{p \rightarrow p_{ref_y}} \mathbf{f}^{p_{ref_y}}, \quad (1)$$

where p_{cit_x} ($x = 1, \dots, k$) and p_{ref_y} ($y = 1, \dots, l$) denote papers that cite p and papers that p refers to, respectively. In addition, $W^{p_{cit_x} \rightarrow p}$ and $W^{p \rightarrow p_{ref_y}}$ are weights for the citation papers and weights for the reference papers, respectively. We can define these weights in a more general form as follows: Let $W^{u \rightarrow v}$ be the coefficient used to compute the weight for between target v and its source u . In addition, let \mathbf{f}^u and \mathbf{f}^v be the feature vectors of the source u and target v papers, respectively. Then cosine similarity $sim(\mathbf{f}^u, \mathbf{f}^v)$ between the two vectors is used as $W^{u \rightarrow v}$.

Generally speaking, a target paper’s feature vector comprises of three parts: words from its own body, words from papers that cite it, and words from papers that it refers to (first, second and third terms in Equation (1), respectively).

3.2 Leveraging Potential Citation Papers

While a rich source of information, a citation network is subject to certain limitations that blunt its effectiveness in modeling target papers. We note that the citation network is constantly expanding; with every new publication, new citation links are added to older work. In studies depending solely on the citation network, cutting-edge work is marginalized as they do not have any citations yet; this is a kind of “cold-start problem” in scholarly recommendation systems that is analogous to the same problem in recommendation systems in general.

Also, because references and citations in a paper are static and never change, newer relevant papers to older ones have the “respon-

sibility” of creating a citation link between them. The static nature of the citation network exacerbates missing and noisy citations.

Finally, the citation network is an artifact of the physical scholarly paper. In many cases, listing all relevant work would be infeasible, as the reference list may grow too long. Many venues have space limitations (including this one), to help encourage authors to use their editorial powers to choose the most relevant references to include. However, this can also cause authors to prune potentially citable references from their bibliographies. Especially, when authors save the space, the surplus may be used to expand the description of their own approaches or experiments.

The above factors led us to believe that the observable, explicit citation network – while certainly of high-quality – is just “the tip of the iceberg”; where *iceberg* refers to the implicit set of relevant works for a target paper. We term papers in this implicit set *potential citation (pc)* papers. If we can predict these implicitly relevant papers, we obtain more content for representing a scholarly paper, which in turn, will improve recommendation performance.

In our approach, we discover such potential citation papers by applying collaborative filtering (CF). CF is usually used to recommend items directly to users. However, we employ it indirectly, by using it to discover potential citation papers, which are then used to represent papers to recommend. This discovery process is needed to better represent papers, which in turn enhances recommendation accuracy. Importantly, our use of CF operates on the paper-citation matrix, and is markedly different from its traditional one-step use in the user-item matrix; in contrast, we employ the citation network twice: both in directly represent target papers through citations and references as well as in finding potential citation papers. The details of our approach also break down into discovery of potential citation papers ((A1) using CF and (A2) imputation-based CF) and (B) feature vector construction for target papers using the discovered potential citation papers.

(A1) Discovery of Potential Citation (pc) Papers with CF

We apply the neighborhood-based algorithm [6] in CF for use in discovering potential citation papers, by substituting papers for users and items for citations. At a high level, we can think of papers as actors that can recommend citations to each other, where CF lets papers that are more similar to a target paper (from a citation perspective) recommend citations with more weight. The algorithm has the following steps analogous to neighborhood-based CF:

- A1.1: Weight all papers with respect to similarity to a target paper. As with the original CF algorithm, paper similarity is measured using the Pearson correlation coefficient between their citation vectors.
- A1.2: Select n papers that have the highest similarity with the target paper. These papers form the *neighborhood* for the target.
- A1.3: Compute a prediction from a weighted combination of the neighbor’s values using a suitable similarity score.

In Step A1.1, the similarity between target paper p_{tgt} and other citation papers p_{cit_u} ($u = 1, \dots, N$), denoted as $S_{tgt,u}$ is computed using the Pearson correlation coefficient:

$$S_{tgt,u} = \frac{\sum_{i=1}^N (r_{tgt,i} - \bar{r}_{tgt}) \times (r_{cit_u,i} - \bar{r}_{cit_u})}{\sqrt{\sum_{i=1}^N (r_{tgt,i} - \bar{r}_{tgt})^2 \times \sum_{i=1}^N (r_{cit_u,i} - \bar{r}_{cit_u})^2}}, \quad (2)$$

where $r_{tgt,i}$ is the score given to citation paper p_{cit_i} by paper p_{tgt} , and \bar{r}_{tgt} is the mean score given by paper p_{tgt} , and N is the total number of papers in the dataset.

In Step A1.2, a subset of appropriate papers is chosen based on their similarity to the target paper and a weighted aggregate of

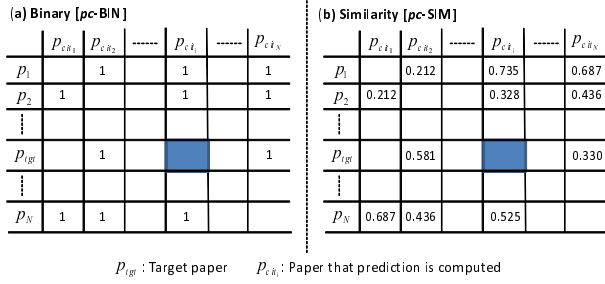


Figure 2: Paper-citation matrix for our adapted collaborative filtering using (a) binary [pc-BIN] and (b) similarity [pc-SIM] weighting.

their ratings is used to generate predictions for the target paper in Step A1.3.

In Step A1.3, predictions are computed as the weighted average of deviations from the neighbor’s mean, shown in Equation (3):

$$p_{tgt,i} = \bar{r}_{tgt} + \frac{\sum_{u=1}^n (r_{u,i} - \bar{r}_u) \times S_{tgt,u}}{\sum_{u=1}^n S_{tgt,u}}, \quad (3)$$

where $p_{tgt,i}$ is the prediction for a target paper p_{tgt} for a citation paper p_{cit_i} . n is the number of papers in the neighborhood.

We explored two possible methods for calculating $S_{tgt,u}$ and $p_{tgt,i}$ above; a binary notion of citation (Figure 2 (a)) as well as a fine-grained similarity version of citation (Figure 2 (b)).

The binary scheme is illustrated in Figure 2 (a), which shows a paper-citation matrix with binary incidence values [pc-BIN]. Entries with a ‘1’ indicate citations by the paper identified by the column to the target paper identified by the row (e.g., paper p_{tgt} is only cited in papers p_{cit_2} and p_{cit_N}).

It is generally agreed that citations have different functions. A key reference that acts as the foundation for the current work is likely more of a positive endorsement than a citation within a list of examples of applications of a particular model. We choose to use cosine similarity between papers as a simple means to model endorsement strength. Figure 2 (b) shows a corresponding paper-citation matrix with similarity values [pc-SIM]. For example, in Figure 2 (b), the similarity between the target paper p_{tgt} and p_{cit_2} , and the target paper p_{tgt} and p_{cit_N} is 0.581 and 0.330, respectively.

To be clear, in both models, multiple citations to the same target paper within a paper are not represented.

(A2) Discovery of Potential Citation (pc) Papers with Imputation-based CF

In Figure 2, the matrices are sparse because each paper can only make a limited number of citations (see Section 4 about how sparse our dataset is). This affects the process of finding relevant potential papers. However, when the corpus of publications is large, we can utilize the fact that there are many other similar papers that potentially could have been cited but were not.

To leverage this opportunity and address sparseness, we employ imputation (hereafter, [pc-IMP]) as we can directly compute similarity between papers and citation papers, unlike the case of the user-item matrix based CF which requires manual ratings. This is a variant of [pc-SIM] and consists of three steps:

A2.1: Impute similarities between all papers, recording them into an intermediate imputed paper-citation matrix (Figure 3).

A2.2: For the target paper, find the n most similar papers from the “(a) original matrix” in Figure 3:

- Weight all papers with respect to similarity to the target

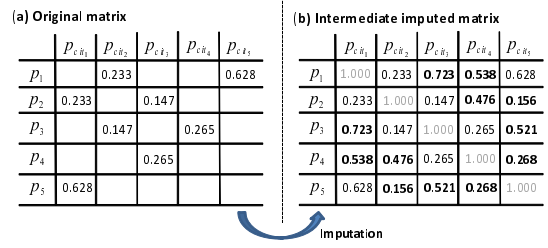


Figure 3: Similarity imputation: (a) original matrix and (b) intermediate imputed matrix (imputed values are bolded).

[Target paper p_1 and corresponding imputed similarities of neighborhood (p_2 , p_4 , and p_5) from “(b) imputed matrix” in Figure 3]

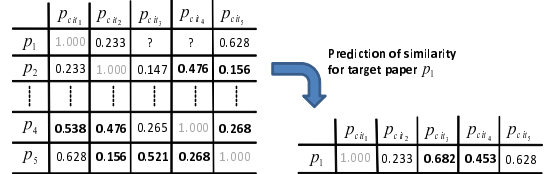


Figure 4: Predictions computed for the target paper p_1 using corresponding neighbors, p_2 , p_4 , and p_5 with similarities from the intermediate imputed matrix.

paper (e.g., p_1). This similarity between papers is measured using the Pearson correlation coefficient between the papers’ citation vectors,

- Select n papers that have the highest similarity with the target paper. These papers form the n -neighborhood for the target paper. In the left of Figure 4, p_2 , p_4 , and p_5 are determined to be the 3-neighborhood for p_1 .

A2.3: Compute a prediction from a weighted combination of the neighbor’s similarity (Figure 4, right). We use Figure 3’s “(b) intermediate imputed matrix” for the n -neighborhood.

(B) Feature Vector Construction for Target Papers

With the discovery and weightage of our discovered potential papers, we can now build the feature vector for target papers. Let F^p be the feature vector for a paper to recommend p . We then define F^p as follows:

$$F^p = f^p + \sum_{x=1}^j W^{p_{pc_x} \rightarrow p} f^{p_{pc_x}} + \sum_{y=1}^k W^{p_{cit_y} \rightarrow p} f^{p_{cit_y}} + \sum_{z=1}^l W^{p \rightarrow p_{ref_z}} f^{p_{ref_z}}, \quad (4)$$

where p_{pc_x} ($x = 1, \dots, j$), p_{cit_y} ($y = 1, \dots, k$), and p_{ref_z} ($z = 1, \dots, l$) denote potential citation papers, papers that cite p , and papers that p refers to, respectively. We employ cosine similarity weight for $W^{p_{pc_x} \rightarrow p}$, $W^{p_{cit_y} \rightarrow p}$, and $W^{p \rightarrow p_{ref_z}}$ as it was found effective in our previous work [16].

3.3 Leveraging Fragments in Potential Citation Papers

In the above, we have artificially enriched the citation network to combat sparsity. We now also consider refining and improving the quality of information in the existing citation network. As scholars, we often acknowledge the importance of others’ previous work

by citation, an explicit reference to previous work which is accompanied by a bibliographic reference to help others locate and trace the prior work. The in-text citation often clearly and succinctly describes a key point of the target paper of import to the current paper, as illustrated below:

Resnik (1999) addressed the issue of language identification for finding Web pages in the languages of interest.

Citation sentences have been used to build summaries of target papers [11, 15] as well as for supporting scientific literature search [1]. However, to the best of our knowledge, they have not been used as an evidence source for recommendation.

Since citation sentences often present a clear representation of a target paper, we hypothesize that careful weighting of citation sentences improves recommendation accuracy. On the other hand, citation sentences are very small text fragments in citation papers. Larger text fragments of the (potential) citation papers may be more useful than using just single citation sentences. Thus, we also experiment with other larger fragments of the source paper: its abstract, introduction, and conclusion sections. We also examine the impact of using other short texts as evidence: keywords (1 to 10 words with the highest TF-IDF score), in place of citation sentences to model candidate papers to recommend.

We note that Mei and Zhai [11] proposed generating summaries using a paper and its citation context (hereafter, [CC]), rather than just using the bare citation sentence. Their approach fixed the citation context to two sentences before and after the citing sentence. For these reasons, we also explore varying the context, the number of sentences before and after the citing sentence, N_{cs} ($1 \leq N_{cs} \leq 3$).

Given these possible (potential) citation paper fragments, we tried the following two different schemes to weight the fragments' words in constructing the target candidate paper's feature vector F^p .

1. [frg-SIM]: Fragments with cosine similarity weighting. In this approach, we add an additional vector obtained from the fragment in the actual or potential citation paper. This approach effectively allows the tunable weights to assign customized weights to the words that appear in the associated fragments, modifying the feature vector F^p to:

$$\begin{aligned}
F^p &= \sum_{x=1}^j W_{(frg)}^{ppc_x \rightarrow p} f_{(frg)}^{ppc_x} + \sum_{y=1}^k W_{(frg)}^{pcity \rightarrow p} f_{(frg)}^{pcity} \\
&+ f^p + \sum_{x=1}^j W^{ppc_x \rightarrow p} f^{ppc_x} \\
&+ \sum_{y=1}^k W^{pcity \rightarrow p} f^{pcity} \\
&+ \sum_{z=1}^l W^{p \rightarrow p_{ref_z}} f^{p_{ref_z}}, \quad (5)
\end{aligned}$$

where the first row are two added terms to Equation (4) that account for evidence from the fragments in potential and explicit citation papers, respectively. As in previous sections, we use cosine similarity as the weighting scheme for both coefficients.

2. [frg-TW]: [frg-SIM] with tunable weight. In this variation, we further augment the feature vector obtained from a fragment with tunable constant weight α ($0 \leq \alpha \leq 1$), which changes the feature vector calculation to:

$$\begin{aligned}
F^p &= \alpha \left(\sum_{x=1}^j W_{(frg)}^{ppc_x \rightarrow p} f_{(frg)}^{ppc_x} + \sum_{y=1}^k W_{(frg)}^{pcity \rightarrow p} f_{(frg)}^{pcity} \right) \\
&+ (1 - \alpha) \left(f^p + \sum_{x=1}^j W^{ppc_x \rightarrow p} f^{ppc_x} \right. \\
&\quad + \sum_{y=1}^k W^{pcity \rightarrow p} f^{pcity} \\
&\quad \left. + \sum_{z=1}^l W^{p \rightarrow p_{ref_z}} f^{p_{ref_z}} \right), \quad (6)
\end{aligned}$$

where α represents the balance between the contribution from the full text and the fragments, and allows our model a bit more expressiveness by finding optimal parameters.

To be clear, in both the [frg-SIM] and [frg-TW] methods, only contribution of terms in a fragment is changed; *i.e.*, Equations (5) and (6) only differ from Equation (4) in the first row, dealing with the contribution of the fragments.

4. EXPERIMENTS

We use the publication lists of 50 researchers who have been engaged in various fields in computer science such as databases, embedded systems, graphics, information retrieval, networks, operating systems, programming languages, software engineering, security, user interface. The researchers also have publication lists in DBLP¹. As DBLP lists many important venues in computer science, we assume here that a researcher's DBLP list is representative of their main interests.

We construct the user profile for each researcher using their respective publication list in DBLP. All of 50 researchers' names are no ambiguity in the list.

The candidate papers to recommend is constructed from proceedings in the ACM Digital Library² (ACM DL). Among them, we collected 100,351 papers published in English, in conferences, symposiums, and workshops held more than three times. We also manually collected citation and reference papers for each paper. In collecting citation and reference papers, we used information on the "Cited By" tab attached in each paper in ACM DL, and those in the references section of each paper. Then, we construct feature vectors for these papers as described in Section 3. Stop words³ were eliminated from each user's publication list and from the candidate papers to recommend. Stemming was performed using the Porter Stemmer⁴. We manually compiled the gold-standard results, by asking each researcher to mark papers relevant to their recent research interest. We divided these datasets into a training set (for parameter tuning) and a test set (for evaluation). Table 1 shows some statistics about our experimental data. In the paper-citation matrix in Figure 2, only 17.2% of all cells are filled, demonstrating that the paper-citation matrix for our dataset is sparse. Our dataset is publicly available⁵.

4.1 Evaluation Measure

As in standard information retrieval (IR), top ranked documents are the most important, since users often scan just the first ranks. As such, we adopt ranked IR evaluation measures, specifically: (1) normalized discounted cumulative gain (nDCG) [7], and (2) mean

¹<http://www.informatik.uni-trier.de/~ley/db/>

²<http://dl.acm.org/>

³<ftp://ftp.cs.cornell.edu/pub/smart/english.stop>

⁴<http://www.tartarus.org/~martin/PorterStemmer/>

⁵<http://www.comp.nus.edu.sg/~sugiyama/SchPaperRecData.html>

Table 1: Some statistics on our scholarly paper dataset.

| (a) Researchers | | |
|--|-----------------|-----------------|
| | Training set | Test set |
| Number of researchers | 25 | 25 |
| Average number of DBLP papers | 10.4 | 9.6 |
| Average number of relevant papers in our dataset | 76.3 | 74.5 |
| Average number of citation papers | 15.3 (max. 169) | 14.4 (max. 145) |
| Average number of reference papers | 15.8 (max. 47) | 14.2 (max. 58) |
| (b) Candidate papers to recommend | | |
| | Training set | Test set |
| Number of papers | 50,176 | 50,175 |
| Average number of citation papers | 19.4 (max. 175) | 16.5 (max. 158) |
| Average number of reference papers | 15.7 (max. 45) | 15.4 (max. 53) |

reciprocal rank (MRR) [19].

(1) Normalized Discounted Cumulative Gain (nDCG). NDCG is well suited to evaluation of recommendation system, as it rewards relevant items in the top ranked results more heavily than those ranked lower. For a given user profile P_{user_i} , the ranked results are examined top-down, where nDCG is computed as:

$$nDCG_i = Z_i \sum_{j=1}^R \frac{2^{r(j)} - 1}{\log(1 + j)},$$

where Z_i is a normalization constant calculated so that a perfect ordering would obtain nDCG of 1; and each $r(j)$ is an integer relevance level (for our case, $r(j) = 1$ and $r(j) = 0$ for relevant and irrelevant recommendations, respectively) of result returned at the rank j ($j = 1, \dots, R$). Then, $nDCG_i$ is averaged over all our target researchers. As a typical recommendation system will just recommend a few items, we are only concerned about whether the top ranked results are relevant or not. Therefore, in this work, we use $nDCG@R$ ($R = \{5, 10\}$) for evaluation where R is the number of top- R papers recommended by our proposed approaches.

(2) Mean Reciprocal Rank (MRR). MRR indicates where in the ranking the first relevant item is returned by the system, averaged over all researchers. This measure provides insight in the ability of the system to return a relevant paper at the top of the ranking. Let r_i be the rank of the highest ranking relevant paper for a target researcher i , then MRR is just the reciprocal rank, averaged over all target researchers, N_{tr} :

$$MRR = \frac{1}{N_{tr}} \sum_{i=1}^{N_{tr}} \frac{1}{r_i}.$$

4.2 Experimental Results

We first optimize our method’s parameters using the training set, and then show experimental results after applying the optimal parameters to the test set. Since there are a few parameters to tune in our approach, we divide the tuning into two halves, where the first half (Phase 1) determines optimal parameters to discover potential citation papers, used in the two independent phases in the second half (Phases 2A and 2B) to leverage fragments.

For simplicity, we only show the best results obtained by using optimal tunable weight α in [pc-BIN] as the improvement compared with the baseline system (see Section 3.1) is marginal and we observe the same trends as [pc-SIM] and [pc-IMP]. In addition, in [CC] and “keywords,” we only show the best result, namely, $N_{cs} = 1$ and 5 keywords, respectively. The remaining parameters (“Weight SIM,” Th , γ , and d) that are inherited from the previous framework, are optimized here, following the methodology in [16].

We also compare our proposed approach with state-of-the-art scholarly paper recommendation systems [13, 20]. We apply their

Table 2: Tuning to address paper-citation matrix sparsity: Recommendation accuracy in [TUNE:pc] when modeling candidate papers using (potential) citation papers under [pc-BIN].

| pc-BIN | nDCG@5 | nDCG@10 | MRR |
|---|--------------|--------------|--------------|
| $n = 2, N_{pc} = 5$ | 0.541 | 0.506 | 0.765 |
| $n = 4, N_{pc} = 5$ | 0.547 | 0.514 | 0.768 |
| $n = 8, N_{pc} = 5$ | 0.530 | 0.500 | 0.758 |
| $n = 10, N_{pc} = 5$ | 0.525 | 0.496 | 0.755 |
| Baseline [16] | 0.519 | 0.486 | 0.748 |
| (Weight “SIM,” $Th = 0.4, \gamma = 0.23, d = 3$) | | | |

optimal settings to our experiments. For our implementation of [13], we construct user profile using the title, and construct the feature vector of candidate papers to recommend using the title and abstract, both of which are generated by bigram frequency. In [20], we apply their “in-matrix prediction” to predict the score of paper-citation matrix in Figure 2 to discover potential citation papers, as their experimental setting is similar to ours. Note that collaborative filtering is *indirectly used* to discover potential citation papers and expand citation network – our method’s use of collaborative filtering is not a direct application, and thus we cannot compare our content-based filtering with collaborative filtering to recommend papers (*i.e.*, we do not have user ratings for papers).

Phase 1 – Parameter tuning to discover potential citation papers [TUNE:pc]. We first optimize parameters for finding the potential citation papers, namely the number of neighborhood n and the number of potential citation papers N_{pc} . We optimize these parameters by using [pc-BIN], [pc-SIM], and [pc-IMP] approaches as described in Section 3.2. Table 2, Figure 5 (a) to (c), and Figure 5 (d) to (f) show experimental results obtained by using [pc-BIN], [pc-SIM], and [pc-IMP], respectively.

From Table 2 and Figure 5, we observe that the optimal parameters that give the best recommendation accuracy are ($n = 4, N_{pc} = 5$) in [pc-BIN] and [pc-SIM], and ($n = 4, N_{pc} = 6$) in [pc-IMP]. These n and N_{pc} values are held constant in Phase 2.

Phase 2A – Tuning fragments in frg -SIM [TUNE:frg-SIM]. After obtaining the optimized parameters n and N_{pc} to find potential citation papers, we further explore which fragments in the citation and potential citation papers give the best recommendation accuracy using Equation (5). Table 3 shows the results.

Phase 2B – Tuning α and fragments in frg -TW [TUNE:frg-TW]. In the other second phase experiment, we optimize the weight for α in Equation (6), and fragments in citation and potential citation papers that give the best recommendation accuracy. Table 4 (a) and Figure 6, Table 4 (b) and Figure 7 show recommendation accuracy obtained by using “only fragments” and “both full text and fragments” in citation and potential citation papers, respectively.

Finally, after applying the optimized parameters on the test set, we arrive at the final test recommendation accuracies. These results are shown in Table 5 and discussed in full later.

4.3 Discussion

In the experiment in [TUNE:pc], according to Figure 5, when the number of neighbors n is too small ($n = 2, 3$) or too large ($n \geq 10$), we obtain poor predictions which result in selecting irrelevant potential citation papers. On the other hand, when the number of neighbors n is 4, our method can select relevant potential citation papers (N_{pc}), resulting in higher recommendation accuracy. We also observe that N_{pc} remains stable – 5 in [pc-BIN] (Table 2) and [pc-SIM] (Figure 5 (a) to (c)), and 6 in [pc-IMP] (Figure 5 (d) to (f)). This indicates that [pc-IMP] finds potential citation papers more effectively than [pc-BIN] and [pc-SIM]. We believe that its

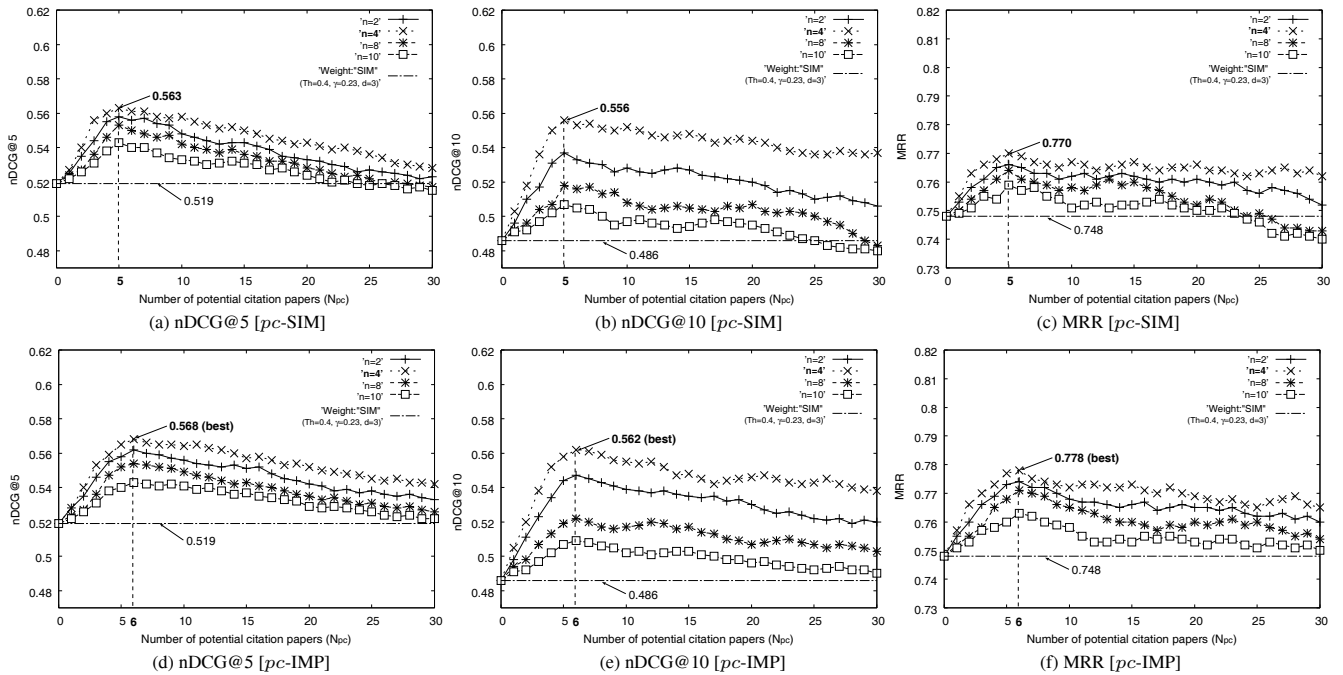


Figure 5: Tuning to address paper-citation matrix sparsity: Recommendation accuracy [TUNE:pc] when using a variable number of citation and potential citation papers in [pc-SIM] ((a) to (c)) and [pc-IMP] ((d) to (f)).

effectiveness is due to its use of the additional n -neighborhood of context papers used to impute the missing values for a target paper.

According to the experimental results in [TUNE:frg-SIM] (Table 3), we obtain the best recommendation accuracy (“Full text + Conclusion”: nDCG@5 of 0.574, nDCG@10 of 0.570, MRR of 0.789) in [pc-IMP]. We again observe that [pc-IMP] generally gives better results compared with [pc-BIN] and [pc-SIM]. Across the board, we see that fragments by themselves perform less well; but that they have a meaningful positive effect if used in conjunction with the full text. The length of the individual fragments may be an important consideration; as keywords and citation sentences are generally quite short. The conclusion may also serve as a factual summary of a paper, not needing to include introductory or motivating material common in abstracts and introduction fragments.

Table 4, Figures 6 and 7 show experimental results in [TUNE:frg-TW]. In this approach, we obtain the best recommendation accuracy (“ $\alpha = 0.4$, Full text + Conclusion”: nDCG@5 of 0.577, nDCG@10 of 0.580, MRR of 0.790) when we employ [pc-IMP] (see Figure 7 (d) to (f)). When we employ “only fragments,” according to Table 4 (a) and Figure 6, we observe that “Conclusion” gives better results than other fragments. Other fragments also yield better results when appropriately tuned. For example, “citation context (CC ($N_{cs} = 1$))” in nDCG@5 gives the recommendation accuracy of 0.533, 0.546, and 0.552 at $\alpha = 0.9$ in [pc-BIN] (Table 4 (a)), [pc-SIM] (Figure 6 (a)), and [pc-IMP] (Figure 6 (d)), respectively. The same trends are also observed in nDCG@10 (Figure 6 (b) and (e)) and MRR (Figure 6 (c) and (f)). However, recommendation accuracy obtained by fragments, “Abstract,” “Introduction,” “Citation Context (CC),” and “5 keywords” generally underperform “Conclusion.” The same trend is also observed with nDCG@10 (Figure 6 (b) and (e)) and MRR (Figure 6 (c) and (f)).

On the other hand, when we employ “both full text and fragments” (Table 4 (b) and Figure 7), the recommendation accuracy generally outperforms that obtained using “only fragments.” For example, “Full text + Conclusion” in nDCG@5 gives the best rec-

ommendation accuracy of 0.559, 0.573, and 0.577 at $\alpha = 0.4$ in [pc-BIN] (Table 4 (b)), [pc-SIM] (Figure 7 (a)), and [pc-IMP] (Figure 7 (d)), respectively. As well as experimental results obtained using “only fragments” described above, other fragments do not give the better recommendation accuracy.

According to Table 4, and Figures 6 and 7, we observe the followings about the value of α : From Table 4, we observe that the value of α that gives the best recommendation accuracy is different between the approach that uses “(a) only the fragment” ($\alpha = [0.7 - 0.9]$) and the approach that uses “(b) both full text and fragments” ($\alpha = [0.4 - 0.6]$). This indicates that in order to characterize candidate papers better, fragments (which contain relatively less text) need to be given larger weight compared to when they are used in conjunction with the full text. Interestingly, in Figures 6 and 7, we see the same trends graphically, but further observe that fixing a particular α value leads to different fragments being more important: in Figure 7, an α value of 0.6 yields “Citation Contexts (CC)” as most useful, 0.5 yields the “Abstract” as being most useful and 0.4 yields the “Conclusion” as best. This indicates that the tunable weight α is an important factor to set if optimal results are desired. We do note that the “Conclusion” fragment’s performance is noticeably better and hence more stable than other fragments’ performance levels, so we recommend this setting.

Table 5 shows recommendation accuracy obtained by applying optimal parameters and fragments to the test set. In the test set, we observe the same trends as in training: [pc-IMP] outperforms both [pc-BIN] and [pc-SIM]. In particular, [pc-IMP] ($n = 4$, $N_{pc} = 6$) with [frg-TW] ($\alpha = 0.4$, “Full text + Conclusion”) gives the best recommendation accuracy, similar to the best results in the training settings. This shows that our dataset keeps a useful balance between the training and test sets.

Furthermore, we observe that our baseline system [16] outperforms others ([13] and [20]) and our approach proposed in this paper ([pc-IMP] + [frg-TW]) gives the best recommendation accuracy. As described in Section 2.1, in [13], user profiles are con-

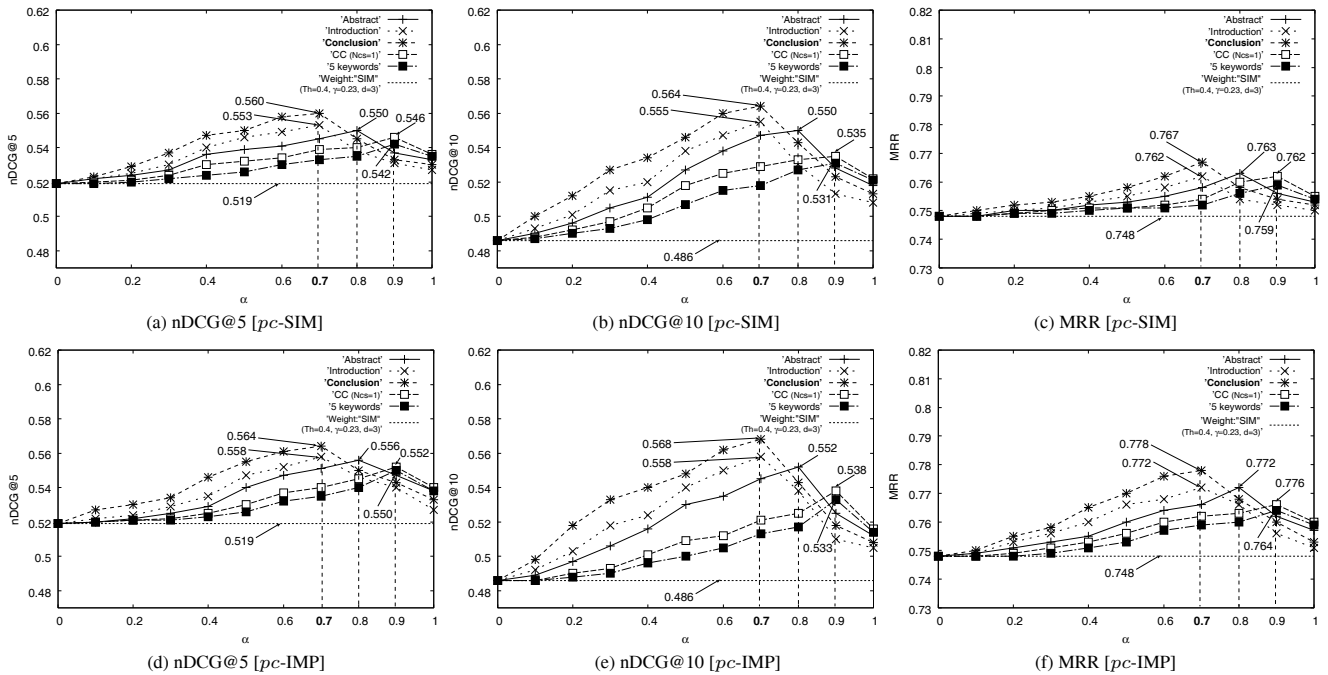


Figure 6: Tuning (potential) citation papers' text weight: Recommendation accuracy in [TUNE: frag-TW] obtained by modeling candidate papers using "only fragments" in (potential) citation papers discovered by [pc-SIM] ((a) to (c)) and [pc-IMP] ((d) to (f)).

structured from title, and feature vectors of candidate papers are generated from title and abstract, resulting in poor recommendation accuracy. This indicates that better representation of users and papers cannot be achieved by using short fragments such as title and abstract only. As described in Section 4.2, in [20], user-paper matrix with binary value, which is similar to Figure 2 (a), is applied to predict missing values to discover potential citation papers. These missing values are computed based on a probabilistic topic model generated from words in the abstract and title. These fragments are too short and uninformative, resulting in discovery of ineffective potential citation papers and irrelevant recommendation of scholarly papers. In light of these observations, we believe that our approach that uses full text and effective fragment (conclusion) in potential citation papers with appropriate tuning contributes to characterize candidate papers to recommend much better, resulting in much more relevant scholarly paper recommendation.

Microscopic Analyses. Diving into individual results, we observed that one of the researchers who works on computer graphics, received relevant recommendations in our optimized system, where the first relevant recommendation was at the first rank. In the baseline system, he could not obtain any relevant recommendation in the top-10 results, and the first relevant result was ranked 63rd. Another researcher, who works on mobile computing, also was provided a relevant recommendation in the first rank, while his most relevant recommendation is ranked 52nd using the baseline.

Inspecting the topics of the papers identified as potential citations also helps to further understand the reach of our algorithm. We observed that our approach discovers relevant potential citation papers to characterize candidate papers to recommend. For example, when the topic of candidate paper to recommend is "access control of business documents based on role mining," often cited by security papers, our approach identifies papers about data mining in databases and papers about information theory related to security as potential citation papers. In another example, given the candidate paper topic of "real world gesture analysis," often cited

in human-computer interaction, our imputation approach discovers potential citation papers whose topics are biomechanics, computer-based music conducting systems, and machine learning. These examples indicate how our approach can characterize papers better than the baseline. This is then reflected in more relevant recommendations and higher accuracy.

Limitations. We also observed that if the topic of the target paper is interdisciplinary, our proposed approach may perform erratically. Further analysis shows that the imputation approach discovers "skewed" potential citation papers. In one instance, where the topic of candidate paper was "understanding mobile user's behavior patterns" that is equally embodied by mobile technology, user search behavior and clustering, our approach discovers potential citation papers that only addressed mobile technology. Examining further, we found that this skewed discovery of potential citation papers was caused by the fact that the selected n -neighborhood of papers consists almost exclusively of the mobile technology topic. Therefore, we need to develop a method for selecting a balanced neighborhood relevant to the target paper in future work, akin to work on addressing data skew in other fields.

Finally, our approach is statistically significant compared with the baseline system [16], improving recommendation accuracy by almost 10% when measured by nDCG@10 ($p < 0.01$), and by over 4% when measured by MRR ($p < 0.05$). We believe that our proposed approach is effective in characterizing candidate papers to recommend to obtain much higher recommendation accuracy.

5. CONCLUSION

We have explored two significant approaches to improve the state-of-the-art in scholarly paper recommendation. In particular, we examine (1) how to alleviate data sparsity using collaborative filtering to find potential citation papers, and (2) how to refine the use of citing papers in characterizing a target candidate paper using fragments in the citation and potential citation papers.

Our results show that, in discovery of potential citation papers,

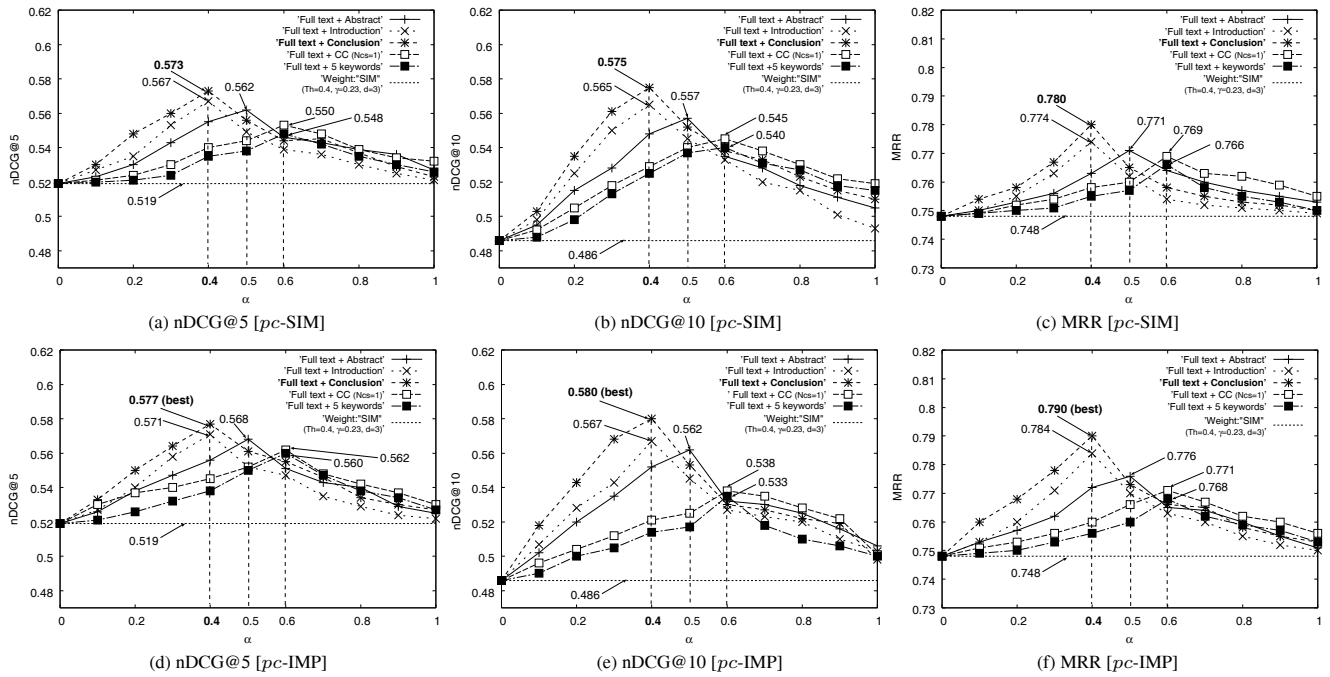


Figure 7: Tuning (potential) citation papers' text weight: Recommendation accuracy in [TUNE: *frag*-TW] obtained by modeling candidate papers using “both full text and fragments” in (potential) citation papers discovered by [pc-SIM] ((a) to (c)) and [pc-IMP] ((d) to (f)).

imputation-based CF is more effective than CF with binary or similarity values. Additionally, the potential citation paper approach, when appropriately tuned, can improve recommendation accuracy significantly. Especially, when we characterize candidate papers to recommend using “full text and conclusion” in both citation and potential citation papers, we achieve the best accuracy and outperform state-of-the-art scholarly paper recommendation systems.

We believe that our approach can be applied more generally. The notion of enriching a network with potential items can be applied to any network that feature asymmetric directional links, such as social networks, patent documents and email dialogues. The notion of using potential citation papers can be applied wherever textual evidence is associated with the links, such as in patent documents and customer testimonials.

To discover much more relevant potential citation papers, in future work, we plan to develop a method for selecting balanced neighborhood for interdisciplinary target papers in collaborative filtering. Moreover, small-sized text fragments in papers might be better characterized by modifying the classical term weighting schemes appropriately. Our future work direction also aims to produce such a novel term weighting scheme suitable for short text.

6. REFERENCES

- [1] S. Bethard and D. Jurafsky. Who Should I Cite? Learning Literature Search Models from Citation Behavior. In *Proc. of the 19th International Conference on Information and Knowledge Management (CIKM'10)*, pages 609–618, 2010.
- [2] K. El-Arini and C. Guestrin. Beyond Keyword Search: Discovering Relevant Scientific Literature. In *Proc. of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, pages 439–447, 2011.
- [3] M. Gori and A. Pucci. Research Paper Recommender Systems: A Random-Walk Based Approach. In *Proc. of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006)*, pages 778–781, 2006.
- [4] Q. He, D. Kifer, J. Pei, P. Mitra, and C. L. Giles. Citation Recommendation without Author Supervision. In *Proc. of the 4th International Conference on Web Search and Data Mining (WSDM'11)*, pages 15–24, 2011.
- [5] Q. He, J. Pei, D. Kifer, P. Mitra, and C. L. Giles. Context-aware Citation Recommendation. In *Proc. of the 19th International World Wide Web Conference (WWW2010)*, pages 421–430, 2010.
- [6] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl. An Algorithmic Framework for Performing Collaborative Filtering. In *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '99)*, pages 230–237, 1999.
- [7] K. Järvelin and J. Kekäläinen. IR Evaluation Methods for Retrieving Highly Relevant Documents. In *Proc. of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2000)*, pages 41–48, 2000.
- [8] R. Kaptein, P. Serdyukov, and J. Kamps. Linking Wikipedia to the Web. In *Proc. of the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'10)*, pages 839–840, 2010.
- [9] L. Katz. A New Status Index Derived from Sociometric Analysis. *Psychometrika*, 18(1):39–43, 1953.
- [10] S. M. McNee, I. Albert, D. Cosley, S. L. P. Gopalkrishnan, A. M. Rashid, J. S. Konstan, and J. Riedl. Predicting User Interests from Contextual Information. In *Proc. of the 2002 ACM Conference on Computer Supported Cooperative Work (CSCW '02)*, pages 116–125, 2002.
- [11] Q. Mei and C. Zhai. Generating Impact-Based Summaries for Scientific Literature. In *Proc. of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, pages 816–824, 2008.

Table 3: Recommendation accuracy (nDCG@5, nDCG@10, and MRR) in [TUNE:frg-SIM] obtained by fragments in potential citation papers. The parameters n and N_{pc} are optimized ones in [pc-BIN], [pc-SIM], and [pc-IMP].

| pc-BIN ($n = 4, N_{pc} = 5$) | nDCG@5 | nDCG@10 | MRR |
|---------------------------------|--------------|--------------|--------------|
| Abstract | 0.520 | 0.468 | 0.746 |
| Introduction | 0.522 | 0.472 | 0.745 |
| Conclusion | 0.525 | 0.476 | 0.748 |
| CC ($N_{cs} = 1$) | 0.518 | 0.471 | 0.738 |
| 5 keywords | 0.514 | 0.466 | 0.727 |
| Full text | 0.547 | 0.514 | 0.768 |
| Full text + Abstract | 0.551 | 0.518 | 0.771 |
| Full text + Introduction | 0.554 | 0.523 | 0.772 |
| Full text + Conclusion | 0.558 | 0.528 | 0.773 |
| Full text + CC ($N_{cs} = 1$) | 0.552 | 0.524 | 0.771 |
| Full text + 5 keywords | 0.550 | 0.521 | 0.772 |

| pc-SIM ($n = 4, N_{pc} = 5$) | nDCG@5 | nDCG@10 | MRR |
|---------------------------------|--------------|--------------|--------------|
| Abstract | 0.524 | 0.476 | 0.747 |
| Introduction | 0.527 | 0.483 | 0.745 |
| Conclusion | 0.531 | 0.488 | 0.758 |
| CC ($N_{cs} = 1$) | 0.522 | 0.473 | 0.740 |
| 5 keywords | 0.518 | 0.469 | 0.733 |
| Full text | 0.563 | 0.556 | 0.770 |
| Full text + Abstract | 0.560 | 0.556 | 0.772 |
| Full text + Introduction | 0.565 | 0.562 | 0.773 |
| Full text + Conclusion | 0.569 | 0.566 | 0.776 |
| Full text + CC ($N_{cs} = 1$) | 0.560 | 0.555 | 0.772 |
| Full text + 5 keywords | 0.559 | 0.554 | 0.773 |

| pc-IMP ($n = 4, N_{pc} = 6$) | nDCG@5 | nDCG@10 | MRR |
|---------------------------------|--------------|--------------|--------------|
| Abstract | 0.535 | 0.498 | 0.757 |
| Introduction | 0.538 | 0.507 | 0.755 |
| Conclusion | 0.543 | 0.514 | 0.762 |
| CC ($N_{cs} = 1$) | 0.535 | 0.510 | 0.754 |
| 5 keywords | 0.532 | 0.505 | 0.745 |
| Full text | 0.568 | 0.562 | 0.778 |
| Full text + Abstract | 0.570 | 0.565 | 0.788 |
| Full text + Introduction | 0.569 | 0.563 | 0.787 |
| Full text + Conclusion | 0.574 | 0.570 | 0.789 |
| Full text + CC ($N_{cs} = 1$) | 0.569 | 0.564 | 0.788 |
| Full text + 5 keywords | 0.568 | 0.563 | 0.787 |

- [12] D. Milne and I. H. Witten. Learning to Link Wikipedia. In *Proc. of the 17th International Conference on Information and Knowledge Management (CIKM'08)*, pages 509–518, 2008.
- [13] C. Nascimento, A. H. F. Laender, A. S. da Silva, and M. A. Gonçalves. A Source Independent Framework for Research Paper Recommendation. In *Proc. of the 11th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2011)*, pages 297–306, 2011.
- [14] T. Nomoto. Two-Tier Similarity Model for Story Link Detection. In *Proc. of the 19th International Conference on Information and Knowledge Management (CIKM'10)*, pages 789–798, 2010.
- [15] V. Qazvinian and D. R. Radev. Scientific Paper Summarization Using Citation Summary Networks. In *Proc. of the 22nd International Conference on Computational Linguistics (Coling'08)*, pages 689–696, 2008.
- [16] K. Sugiyama and M.-Y. Kan. Scholarly Paper Recommendation via User's Recent Research Interests. In *Proc. of the 10th ACM/IEEE Joint Conference on Digital Libraries (JCDL '10)*, pages 29–38, 2010.
- [17] T. Strohman and W. B. Croft and D. Jensen. Recommending Citations for Academic Papers. In *Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, pages 705–706, 2007.
- [18] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl. Enhancing Digital Libraries with TechLens. In

Table 4: Tuning results: Recommendation accuracy in [TUNE:frg-TW] obtained by modeling candidate papers to recommend using “(a) only fragments” and “(b) both full text and fragments” in citation and potential citation papers discovered by [pc-BIN].

| (a) Only fragments | | | |
|---|--------------|--------------|--------------|
| pc-BIN ($n = 4, N_{pc} = 5$) | nDCG@5 | nDCG@10 | MRR |
| Abstract ($\alpha = 0.8$) | 0.537 | 0.515 | 0.760 |
| Introduction ($\alpha = 0.7$) | 0.539 | 0.518 | 0.761 |
| Conclusion ($\alpha = 0.7$) | 0.546 | 0.525 | 0.764 |
| CC ($N_{cs} = 1$) ($\alpha = 0.9$) | 0.533 | 0.509 | 0.758 |
| 5 keywords ($\alpha = 0.9$) | 0.529 | 0.506 | 0.756 |
| Baseline system [16] | 0.519 | 0.486 | 0.748 |
| (Weight “SIM,” $Th = 0.4, \gamma = 0.23, d = 3$) | | | |

| (b) Both full text and fragments | | | |
|---|--------------|--------------|--------------|
| pc-BIN ($n = 4, N_{pc} = 5$) | nDCG@5 | nDCG@10 | MRR |
| Full text + Abstract ($\alpha = 0.5$) | 0.548 | 0.520 | 0.768 |
| Full text + Introduction ($\alpha = 0.4$) | 0.552 | 0.525 | 0.772 |
| Full text + Conclusion ($\alpha = 0.4$) | 0.559 | 0.533 | 0.778 |
| Full text + CC ($N_{cs} = 1$) ($\alpha = 0.6$) | 0.543 | 0.517 | 0.766 |
| Full text + 5 keywords ($\alpha = 0.6$) | 0.540 | 0.514 | 0.764 |
| Baseline system [16] | 0.519 | 0.486 | 0.748 |
| (Weight “SIM,” $Th = 0.4, \gamma = 0.23, d = 3$) | | | |

Table 5: Recommendation accuracy obtained by applying optimal parameters and fragments to the test set. “*” and “” denote the difference between the best results in the baseline system [16] (underlined scores) and the each result in this paper is significant for $p < 0.01$ and $p < 0.05$, respectively.**

| | nDCG@5 | nDCG@10 | MRR |
|---|----------------|----------------|---------------|
| pc-BIN ($n = 4, N_{pc} = 5$) | | | |
| frg-SIM (Full text + Conclusion) | 0.556* | 0.535* | 0.766* |
| frg-TW ($\alpha = 0.4$, Full text + Conclusion) | 0.561* | 0.540* | 0.773* |
| pc-SIM ($n = 4, N_{pc} = 5$) | | | |
| frg-SIM (Full text + Conclusion) | 0.566* | 0.563** | 0.770* |
| frg-TW ($\alpha = 0.4$, Full text + Conclusion) | 0.572** | 0.570** | 0.781* |
| pc-IMP ($n = 4, N_{pc} = 6$) | | | |
| frg-SIM (Full text + Conclusion) | 0.572** | 0.568** | 0.787* |
| frg-TW ($\alpha = 0.4$, Full text + Conclusion) | 0.579** | 0.577** | 0.793* |
| Baseline system [16] | <u>0.525</u> | <u>0.479</u> | <u>0.751</u> |
| (Weight “SIM,” $Th = 0.4, \gamma = 0.23, d = 3$) | | | |
| Nascimento <i>et al.</i> [13] | 0.336 | 0.308 | 0.438 |
| (“Frequency of bi-gram” obtained from title and abstract) | | | |
| Wang and Blei [20] | 0.393 | 0.371 | 0.495 |
| (“In-matrix prediction” in collaborative topic regression) | | | |

- Proc. of the 4th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2004)*, pages 228–236, 2004.
- [19] E. M. Voorhees. The TREC-8 Question Answering Track Report. In *Proc. of the 8th Text REtrieval Conference (TREC-8)*, pages 77–82, 1999.
- [20] C. Wang and D. M. Blei. Collaborative Topic Modeling for Recommending Scientific Articles. In *Proc. of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*, pages 448–456, 2011.
- [21] R. West, D. Precup, and J. Pineau. Completing Wikipedia’s Hyperlink Structure through Dimensionality Reduction. In *Proc. of the 18th International Conference on Information and Knowledge Management (CIKM'09)*, pages 1097–1106, 2009.
- [22] R. West, D. Precup, and J. Pineau. Automatically Suggesting Topics for Augmenting Text Documents. In *Proc. of the 19th International Conference on Information and Knowledge Management (CIKM'10)*, pages 929–938, 2010.
- [23] D. Yang, B. Wei, J. Wu, Y. Zhang, and L. Zhang. CARES: A Ranking-Oriented CADAL Recommender System. In *Proc. of the 9th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2009)*, pages 203–211, 2009.