

Scholarly Paper Recommendation via User's Recent Research Interests*

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

Min-Yen Kan
National University of Singapore
Computing 1, 13 Computing Drive,
Singapore 117417
kanmy@comp.nus.edu.sg

ABSTRACT

We examine the effect of modeling a researcher's past works in recommending scholarly papers to the researcher. Our hypothesis is that an author's published works constitute a clean signal of the latent interests of a researcher. A key part of our model is to enhance the profile derived directly from past works with information coming from the past works' referenced papers as well as papers that cite the work. In our experiments, we differentiate between junior researchers that have only published one paper and senior researchers that have multiple publications. We show that filtering these sources of information is advantageous – when we additionally prune noisy citations, referenced papers and publication history, we achieve statistically significant higher levels of recommendation accuracy.

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, Human factors, Performance

Keywords

Digital library, Information retrieval, Recommendation, User modeling

1. INTRODUCTION

Digital libraries (DLs) are entering a golden age: today, much of the world's new knowledge is now largely captured in digital form and archived within a digital library system. However, these trends also lead to information overload, where users find an overwhelming number of publications that match their search queries but are largely irrelevant to their latent information needs.

To alleviate these problems, past researchers have focused their attention on finding better ranking algorithms for paper search. In particular, the PageRank algorithm [20] has been employed [33, 16,

*This work was partially supported by a National Research Foundation grant "Interactive Media Search" (grant # R-252-000-325-279)

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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

JCDL'10 June 21–25, 2010, Gold Coast, Queensland, Australia.

Copyright 2010 ACM 978-1-4503-0085-8/10/06 ...\$10.00.

27] to induce a better global ranking of search results. A problem with this approach is that it does not induce better rankings that are personalized for the specific interests of the user.

To address this issue, digital libraries such as Elsevier¹, PubMed², SpringerLink³ all have systems that can send out email alerts or provide RSS feeds on paper recommendations that match user interests. These systems make the DL more proactive, sending out matched articles in a timely fashion. Unfortunately, these require the user to state their interests explicitly, either in terms of categories or as saved searches, and take up valuable time on the part of the user to set up.

We aim to address this problem by providing recommendation results by using latent information about the user's research interests that exists in their publication list. A researcher's publication list both a historical and current list of research interests and requires little to no effort on the part of the user to provide. The aim of our work in this paper is to study and assess the effectiveness of different models in representing this information in their user profile. Our main contribution in this work is in developing the whole model which accounts for information contained not only in the papers that are published by an author, but also in papers that are referenced by or that cite the author's work. We extend this paradigm in modeling candidate papers to recommend, enriching their representation to also include their referenced work and works that cite them. We show that modeling these contexts is crucial for obtaining higher recommendation accuracy.

In performing analysis and evaluation, we differentiate between junior researchers whom we define as only having published a single work, and senior researchers that have multiple publications that span several years. We discuss these details in Section 3.1. As junior researchers only have published a single work, we believe that recommendation for this class of researchers is more difficult, due to the limited amount of data available to construct their user profile.

Other works have also differentiated their analysis of scholarly paper recommendation by user experience. For example, Torres *et al.* [36] investigated recommendation satisfaction according to two types of researcher's level, students (masters and PhD students) and professionals (researchers and professors). However, their recommendation system modeled each researcher's interest using only one paper that the researcher must manually choose.

In our analysis of our basic model, we noted that certain references and cited works are not highly relevant in describing a target paper. The content of such tangential papers causes the representation of the target paper to drift from its core topic. To further improve recommendation accuracy, we tested methods to remove the influence of such papers via pruning.

Similarly, we hypothesize that the set of topics represented by *all* of a senior researcher papers is not representative of a researcher's

¹http://www.elsevier.com/wps/find/homepage.cws_home

²<http://www.ncbi.nlm.nih.gov/pubmed/>

³<http://www.springerlink.com/home/main.mpx>

current interests. To further improve recommendation accuracy, we also tested methods for placing a larger emphasis on recent publications in a researcher's profile.

This paper is organized as follows: In Section 2, we review related work on how to provide relevant search results in digital library systems. In Section 3, we detail our approach to recommendation using a user's most recent publication. In Section 4, we present an exhaustive analysis of different parameterizations of our model and dissect the evaluation results in detail. Finally, we conclude the paper with a summary and directions for future work in Section 5.

2. RELATED WORK

Recommendation can be viewed as periodic searching of a digital library. As such, we first review work on document ranking in DL search, with a special emphasis on the use of the PageRank algorithm. We then review work on recommendation systems in the environment of scholarly DLs, and conclude our review with a discussion on the representations that have been used to construct a robust user profile for use in recommendation.

2.1 Improving Ranking in Digital Libraries

The PageRank algorithm [20] simulates a user navigating the Web at random, by choosing between jumping to a random page with a certain probability (referred to as the *damping factor* d), and following a random hyperlink. While this algorithm has been most famously applied to improve ranking of Web search results, it has also been applied to the digital library field in two ways: (1) in improving the ranking of search results, and (2) in measuring the importance of scholarly papers.

2.1.1 Ranking Search Results

Since the beginning of bibliometric analysis, scholars have been measuring the count of other publications that refer to a particular author or work. This notion of *citation count* is widely used in evaluating the importance of a paper because it has been shown to strongly correlate with academic document impact [19]. The Thomson Scientific ISI impact factor (ISI IF) is the representative approach using citation count [9], which factors citation counts with a moving window to calculate the impact of certain journals. The advantages of citation count are (1) its simplicity of computation; and (2) that it is a proven method which has been used for many years in scientometrics. However, citation counting has well-known limitations: Citing papers with high impact and ones with low impact are treated equally in standard citation counting.

In order to overcome the above shortcomings of impact factor, Sun and Giles [33] noted that conference venues in computer science are a prime vehicle for impact calculation that are neglected by the ISI impact factor. They proposed to remedy this problem by incorporating the popularity factor to consider venue as an information cue and to reflect the influence of a publication venue. This popularity factor is defined based on citation analysis of publication venues and the PageRank algorithm.

When PageRank is introduced to scholarly papers, the rank of a paper can decrease if the paper contains a large quantity of outgoing links. In some ways, this is counter-intuitive, as a well-referenced paper may better contextualize its contributions with respect to existing work, and would thus be a mark of higher quality work. Krapivin and Marchese [16] proposed "Focused PageRank" algorithm to alleviate this problem. In their approach, a reader of an article (referred to as a "focused surfer") may follow the references with different probabilities, so their random surfer model becomes focused on some of the references in the article.

While PageRank can estimate authority of the article, one of its problems is that it ranks articles based on the prior popularity (number of citations) or prior prestige (PageRank score). Therefore, recent articles always obtain lower scores. However, it is important for researchers to be able to find such recent articles because they can discover new research directions, solutions and approaches, and digest new work that is relevant to their current interests. As

such, Sayyadi and Getoor [27] proposed "FutureRank" which computes the expected future PageRank, focusing on citation network of scholarly papers.

2.1.2 Measuring the Importance of Scholarly Papers

PageRank has also been applied to measure the importance of scholarly papers. Unlike the studies in Section 2.1.1, these works define the importance of a paper among a certain given set.

The ISI Impact Factor is also flawed in that its rankings are biased towards popularity. In order to overcome this problem, Bollen *et al.* [4] compared the rankings of journals obtained by the following approaches: (1) ISI Impact Factor, (2) weighted PageRank, and (3) their contribution called Y -factor, that is a product of (1) and (2). Journal ranking obtained by their Y -factor showed that the top ranked journal closely matched personal perception of importance.

Chen *et al.* [7] applied the PageRank algorithm to the scientific citation networks. They found that some classical articles in physics domain have a small quantity of citations but also a very high PageRank. They called these papers *scientific gems* and concluded that existence of such gems is caused by the PageRank model, which captures not only the total citation count but also the rank of each of the citing papers.

2.2 Recommendation in Digital Libraries

Recommendation systems provide a promising approach to ranking scholarly papers according to a user's interests. Such systems are classified by their underlying method of recommendation.

Collaborative filtering [10, 24, 15] is one of the most successful recommendation approaches that works by recommending items to target users based on what other similar users have previously preferred. This method has been used in e-commerce site such as Amazon.com⁴, Ebay⁵ and so on. However, it suffers from "cold-start problem," in which it cannot generate accurate recommendations without enough initial ratings from users. Recent works alleviate this problem by introducing pseudo users that rate items [21] and imputing estimated rating data using some imputation technique [39].

Content-based filtering [5, 2, 26] is also widely used in recommender systems. This approach provides recommendations by comparing candidate item's content representation with the target user's interest representation. This method has been applied mostly in textual domains such as news recommendation [8] and hybrid approaches with collaborative filtering [1, 2, 18].

We now examine recommendation systems in the field of scholarly digital libraries. McNee *et al.* [17] proposed an approach to recommending citations using collaborative filtering. Their approach extended Referral Web [13] by exploring ways to directly apply collaborative filtering to social networks that they term as the "Citation Web," a graph formed by the citations between research papers. This data can be mapped into a framework of collaborative filtering and used to overcome the cold-start problem. Expanding this approach, Torres *et al.* [36] proposed a method for recommending research papers by combining collaborative filtering and content-based filtering. However, the final ranking scheme obtained by merging the output from collaborative filtering and content-based filtering is not performed as the authors claim that pure recommendation algorithms are not designed to receive input from another recommender algorithm. Gori and Pucci [11] devised a PageRank-based method for recommending research papers. But in their approach, a user have to prepare initial set of relevant articles to get better recommendation, and the damping factor d that affects the score of PageRank is not optimized. Yang *et al.* [40] presented a recommendation system for scholarly papers that used a ranking-oriented collaborative filtering approach. Although their system overcomes the cold-start problem by utilizing implicit behaviors extracted from a user's access logs, Web usage data are noisy and not reliable generally as pointed out in [28]. In addition,

⁴<http://www.amazon.com/>

⁵<http://www.ebay.com/>

their predefined criteria and parameters to select effective data are not investigated in detail.

2.3 Robust User Profile Construction in Recommendation Systems

We believe that for recommending scholarly papers, content-based filtering is more appropriate, as scholarly papers are textual data and provide a wealth of data to base recommendations on. Content-based filtering is widely used in the field of Web information retrieval and focuses on constructing a robust user profile. Teevan et al. [35] proposed an approach to personalizing Web search results based on implicit representations of each user’s long term and short term interests, similar to [31]. However, in order to construct user profile, they utilize broad spectrum of sources such as the user’s previously visited Web pages and email history. Since this approach is holistic, it is difficult to capture each user’s interests from such broad sources. Kim et al. [14] proposed a method for constructing a robust user profile by using frequent patterns obtained from the user’s click-history, in addition to a term weighting scheme. However, frequent patterns are fixed features; as such, this approach is not flexible enough to represent user interests. In the context of content-based filtering, Chu and Park [8] proposed a recommendation method for dynamic content (such as news) by constructing user profiles using metadata captured from the user: preferences regarding the user’s interest, demographics, and implicit interaction data.

Moreover, there are several works that represent user’s information need from long-term search history to improve retrieval accuracy. While Shen et al. [29] and Tan et al. [34] employed statistical language modeling, in particular, the Kullback-Leibler divergence retrieval model to represent a user’s information need, White et al. [38] constructed user profiles with a small number of Web pages preceding the current browsing page to recommend Web sites. As well as [14] and [8] described above, these works used Web usage data such as click-through history.

3. PROPOSED METHOD

To sum up, systems employing PageRank have demonstrated improved ranking of search results. However, since PageRank is a global ranking scheme, the user’s research interests are not considered in the ranking and the generated ranking is thus not customized to the user. When we examined the commonalities of the recommendation systems described in Section 2.2, we note that they consider a user’s interests in only a limited sense, by virtue of using metadata or collaborative filtering. Building a user profile derived directly from user content is thus most relevant to our scenario. However, as described in Section 2.3, the existing methods for constructing a robust user profile consider only click-through data for Web page recommendation. We believe this method is possibly too ephemeral for research interests, which are more long term in general. For scholarly paper recommendation, we should utilize the textual nature of the papers themselves.

To address these shortcomings, we propose recommending papers based on an individual’s recent research interests as modeled by a profile derived from their publication list. We hypothesize that this will result in high recommendation accuracy as we believe that a user’s research interests are reflected in their prior publications.

We first construct each researcher’s profile using their list of previous publications, and then recommend papers by comparing the profiles with the contents of candidate papers. Unlike research studies described in the previous section, our approach is novel because it directly addresses each user’s research interest using their publication history. A key aspect of our approach is that we include contextual evidence about each paper in the form of its *neighboring papers*: the papers that cite the target paper (we term these *citation papers*) and papers referenced by the target paper (*reference papers*). While the use of contextual information from neighbors is not new – it has also been successfully applied to the problems of Web page representation [32], Web page classification [23], spam detection [6] – to the best of our knowledge, it has not been utilized

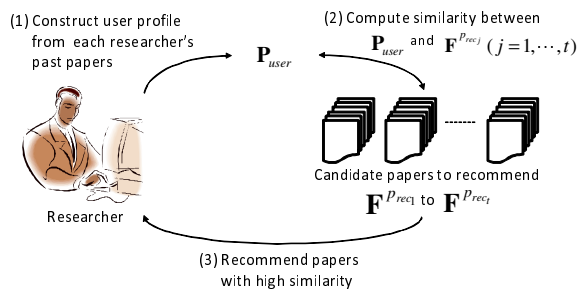


Figure 1: System overview.

in scholarly paper recommendation. An additional desirable property of our approach is that it is domain-independent. Its simple requirement is that contextual information from such neighboring publications needs to be accessible.

Figure 1 shows an overview of our approach. (1) We first construct a user profile P_{user} from a researcher’s list of published papers; (2) then construct feature vectors $F^{P_{rec_j}}$ ($j = 1, \dots, t$) for candidate papers to recommend; (3) compute cosine similarity $Sim(P_{user}, F^{P_{rec_j}})$ between P_{user} and $F^{P_{rec_j}}$ ($j = 1, \dots, t$), and recommend papers with high similarity. In the following, we describe how to construct the user profile P_{user} and feature vectors $F^{P_{rec_j}}$ ($j = 1, \dots, t$) used in the first two steps.

3.1 User Profile Construction

We first divide researchers into (i) junior researchers, and (ii) senior researchers. This is because the two types of researchers’ publication lists exhibit different properties. We define junior researchers as having only one recently published paper, which has yet to attract any citations (i.e., no citation papers). Senior researchers differ in having multiple past publications, where their past publications may have attracted citations. This is shown graphically in Figures 2 and 3.

Our representations of the user profile are based on foundation of a paper represented as a feature vector. For each paper p on the publication list of a researcher, we transform p into a feature vector f^p as follows:

$$f^p = (w_{t_1}^p, w_{t_2}^p, \dots, w_{t_m}^p), \quad (1)$$

where m is the number of distinct terms in the paper, and t_k ($k = 1, 2, \dots, m$) denotes each term. Using term frequency (TF), we also define each element $w_{t_k}^p$ of f^p in Equation (1) as follows:

$$w_{t_k}^p = \frac{tf(t_k, p)}{\sum_{s=1}^m tf(t_s, p)},$$

where $tf(t_k, p)$ is the frequency of term t_k in a paper p . We prefer TF rather than adopting the standard TF-IDF scheme [25], as the small number of papers in a researcher’s publication list may adversely affect the IDF score calculation.

Based on the set of feature vectors f^p , we can then construct the user profiles for junior researchers, and senior researchers by using the relations between each researcher’s published paper and its citation and reference papers.

In our framework, we assign weights to modify the influence of citation and reference papers. Let $W^{p_u \rightarrow v}$ be the multiplicative coefficient used to integrate the target paper v with a source paper u . We explore the following three different weighting schemes for this framework.

(W1) Linear Combination (LC)

This baseline weighting scheme simply combines papers u and v . In other words, we define $W^{p_u \rightarrow v}$ as follows:

$$W^{p_u \rightarrow v} = 1. \quad (2)$$

This method treats each neighboring paper v on a parity with the researcher’s own paper u .

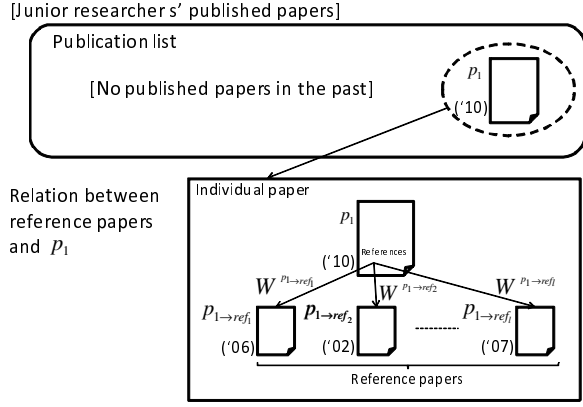


Figure 2: Publication list by junior researchers.

(W2) Cosine Similarity (SIM)

Here we employ the cosine similarity between papers u and v as the weighting scheme. Applying Equation (1), let \mathbf{f}^u and \mathbf{f}^v be feature vector of papers u and v , respectively. Then the similarity $\text{sim}(\mathbf{f}^u, \mathbf{f}^v)$ between these two feature vectors is computed by Equation (3), and we use $\text{sim}(\mathbf{f}^u, \mathbf{f}^v)$ as $W^{p_u \rightarrow v}$.

$$W^{p_u \rightarrow v} = \text{sim}(\mathbf{f}^u, \mathbf{f}^v) = \frac{\mathbf{f}^u \cdot \mathbf{f}^v}{|\mathbf{f}^u| \cdot |\mathbf{f}^v|}. \quad (3)$$

This approach strengthens the signal from a researcher's paper u by emphasizing papers that are more similar among its citation and reference papers.

(W3) Reciprocal of the Difference Between Published Years (RPY)

We focus on the publication year of papers u and v , and use the reciprocal of the difference between these two years. Let the published year of papers u and v be Y_u and Y_v ($Y_u \geq Y_v$), respectively. Then, the weight $W^{p_u \rightarrow v}$ is defined as follows:

$$W^{p_u \rightarrow v} = \frac{1}{(Y_u - Y_v) + c}, \quad (4)$$

where c is a positive constant to prevent the value of $W^{p_u \rightarrow v}$ from being extremely large when Y_u and Y_v are the same published year. We set the value of c to 0.9. This scheme assigns larger weights to newer papers and smaller weights to older papers, to favor the representation of papers published close together temporally.

With weighting schemes now defined, we can construct user profiles for the two classes of researchers.

User Profile for Junior Researchers

As shown in Figure 2, junior researchers have only one paper p_1 that is most recently published (e.g., '10). The paper p_1 has reference papers $p_{1 \rightarrow \text{ref}_y}$ ($y = 1, \dots, l$) (published older than '10). However, there are no papers that cite the paper p_1 because p_1 is just published recently. Therefore, the weights are only applicable to reference papers. In terms of that, user profile \mathbf{P}_{user} is defined as follows:

$$\mathbf{P}_{user} = \mathbf{f}^{p_1} + \sum_{y=1}^l W^{p_1 \rightarrow \text{ref}_y} \mathbf{f}^{p_{1 \rightarrow \text{ref}_y}}, \quad (5)$$

where $W^{p_1 \rightarrow \text{ref}_y}$ ($y = 1, \dots, l$) denotes each weight assigned to paper $p_{1 \rightarrow \text{ref}_y}$ computed on the basis of paper p_1 , defined by a choice among (W1) to (W3).

User Profile for Senior Researchers

As shown in Figure 3, senior researchers have several published papers p_i ($i = 1, \dots, n-1$) in the past (e.g., '02, '03, ...) as well as the most recently published paper p_n (e.g., '10). With the exception

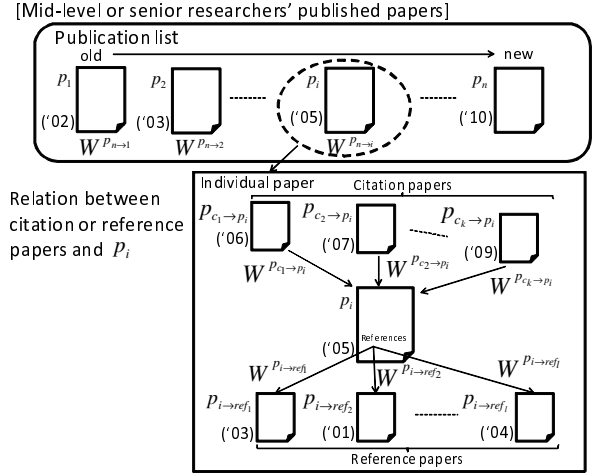


Figure 3: Publication lists by senior researchers.

of the most recent paper p_n , each past published paper p_i may be cited by other papers $p_{c_x \rightarrow p_i}$ ($x = 1, \dots, k$). In addition, each paper p_i has reference papers. Therefore, we first construct the feature vector \mathbf{F}^{p_i} for each paper p_i , using its feature vectors for citation papers $\mathbf{f}^{p_{c_x \rightarrow p_i}}$ and their weights $W^{p_{c_x \rightarrow p_i}}$, and its feature vectors for reference papers $\mathbf{f}^{p_{i \rightarrow \text{ref}_y}}$ and their weights $W^{p_{i \rightarrow \text{ref}_y}}$, as follows:

$$\begin{aligned} \mathbf{F}^{p_i} = \mathbf{f}^{p_i} &+ \sum_{x=1}^k W^{p_{c_x \rightarrow p_i}} \mathbf{f}^{p_{c_x \rightarrow p_i}} \\ &+ \sum_{y=1}^l W^{p_{i \rightarrow \text{ref}_y}} \mathbf{f}^{p_{i \rightarrow \text{ref}_y}}, \end{aligned} \quad (6)$$

Then, using Equation (6), the user profile \mathbf{P}^{user} for a senior researcher is defined as follows:

$$\begin{aligned} \mathbf{P}_{user} &= W^{p_{n \rightarrow 1}} \mathbf{F}^{p_1} + W^{p_{n \rightarrow 2}} \mathbf{F}^{p_2} \\ &+ \dots + W^{p_{n \rightarrow n-1}} \mathbf{F}^{p_{n-1}} + \mathbf{F}^{p_n} \\ &= \sum_{z=1}^{n-1} W^{p_n \rightarrow z} \mathbf{F}^{p_z} + \mathbf{F}^{p_n}, \end{aligned} \quad (7)$$

where $W^{p_n \rightarrow z}$ ($z = 1, \dots, n-1$) denotes each weight assigned to paper $p_{n \rightarrow z}$ computed on the basis of the most recent paper p_n , defined by a choice among (W1) to (W3).

As these researchers have multiple prior publications, we also employ an additional forgetting factor (W4) that gives a larger weight (close to 1) to more recent papers and smaller weight (close to 0) to older papers, under the assumption that a user's research interest gradually decays as years pass.

(W4) Forgetting Factor (FF)

Let γ ($0 \leq \gamma \leq 1$) and d be defined as a forgetting coefficient and the difference between the published year of the most recent paper and previously published papers, respectively. Then, the weight $W^{p_n \rightarrow z}$ in this scheme is defined as follows:

$$W^{p_n \rightarrow z} = e^{-\gamma \cdot d}. \quad (8)$$

3.2 Feature Vector Construction for Candidate Papers

Unlike the TF representation of papers used in the user profile, we employ TF-IDF [25] for the calculation of the feature vector $\mathbf{f}^{p_{rec}}$ of a candidate paper p_{rec} to be considered for recommendation. Identical to Equation (1), we first define the feature vector $\mathbf{f}^{p_{rec}}$ of p_{rec} as follows:

$$\mathbf{f}^{p_{rec}} = (w_{t_1}^{p_{rec}}, w_{t_2}^{p_{rec}}, \dots, w_{t_m}^{p_{rec}}), \quad (9)$$

where m is the number of distinct terms in the paper, and t_k ($k = 1, 2, \dots, m$) denotes each term. Using TF-IDF, each element $w_{t_k}^{p_{rec}}$ of $\mathbf{f}^{p_{rec}}$ in Equation (9) is defined as follows:

$$w_{t_k}^{p_{rec}} = \frac{tf(t_k, p_{rec})}{\sum_{s=1}^m tf(t_s, p_{rec})} \cdot \log \frac{N}{df(t_k)},$$

where $tf(t_k, p_{rec})$ is the frequency of term t_k in the target paper p , N is the total number of papers to recommend in the collection, and $df(t_k)$ is the number of papers in which term t_k appears. We favor TF-IDF here rather than pure TF for candidate papers, as the pool for candidate papers is usually much larger. In our experiments as we describe later in Section 4.1, our candidate paper base consists of several hundreds of papers, making IDF more reliable and consistent. Critically, our dataset also contains clean citation information that allows us to construct correct citation and reference papers. Therefore, we also use this information to characterize a candidate paper better and obtain high recommendation accuracy: Let $\mathbf{F}^{p_{rec}}$ be the feature vector for paper to recommend, as well as Equation (6), this is denoted as follows:

$$\mathbf{F}^{p_{rec}} = \mathbf{f}^{p_{rec}} + \sum_{x=1}^k W^{p_{c_x} \rightarrow p_{rec}} \mathbf{f}^{p_{c_x} \rightarrow p_{rec}} + \sum_{y=1}^l W^{p_{rec} \rightarrow p_{ref_y}} \mathbf{f}^{p_{rec} \rightarrow p_{ref_y}}. \quad (10)$$

where $p_{c_x \rightarrow p_{rec}}$ ($x = 1, \dots, k$) and $p_{rec \rightarrow p_{ref_y}}$ ($y = 1, \dots, l$) denote papers that cite p_{rec} and papers that p_{rec} refers, respectively.

3.3 Recommendation of Papers

Using the user profile defined by Equation (5) or (7), and feature vector for the candidate paper to recommend defined by Equation (10), our system computes similarity $sim(\mathbf{P}_{user}, \mathbf{F}^{p_{rec}})$ between \mathbf{P}_{user} and $\mathbf{F}^{p_{rec}}$ by Equation (11):

$$sim(\mathbf{P}_{user}, \mathbf{F}^{p_{rec}}) = \frac{\mathbf{P}_{user} \cdot \mathbf{F}^{p_{rec}}}{|\mathbf{P}_{user}| \cdot |\mathbf{F}^{p_{rec}}|}, \quad (11)$$

and ranks the set of candidate papers in order of decreasing similarity. While technically a ranking function, we can consider the top n papers as recommended to the user, where n could be set to 5 or 10.

4. EXPERIMENTS

4.1 Experimental Data

We use the single publication of 15 junior researchers, and publication lists of 13 senior researchers who have been engaged in natural language processing and information retrieval, and have publication lists in DBLP⁶. As DBLP lists many important venues in computer science, we believe that a researcher’s DBLP list is representative of their main interests.

Junior researchers published their papers in 2008 or 2009. We construct the user profile for each researcher using their respective publication list in DBLP. Table 1 shows the statistics about these researchers. Since we focus on recommending scientific paper only, we removed references to Web sites, books, and other URLs for our experiments.

The candidate papers to recommend is the ACL Anthology Reference Corpus (ACL ARC)⁷ [3]. The ACL ARC is constructed from a significant subset of the ACL Anthology⁸, a digital archive of conference and journal papers in natural language processing and computational linguistics. The ACL ARC consists of 10,921 articles from the February 2007 snapshot of the ACL Anthology.

⁶<http://www.informatik.uni-trier.de/~ley/db/index.html>

⁷Version 20080325, <http://acl-arc.comp.nus.edu.sg/>

⁸<http://aclweb.org/anthology-new/>

Table 1: Some statistics about junior and senior researchers.

	Junior researchers	Senior researchers
Number of subjects	15	13
Average number of DBLP papers	1.3	9.5
Average number of relevant papers in ACL’00-’06	28.6	38.7
Average number of citation papers	0	10.5 (max. 199)
Average number of reference papers	18.7 (max. 29)	19.4 (max. 79)

Among them, we selected 597 full papers published in ACL 2000-2006. We asked each of junior and senior researchers to mark papers relevant to their recent research interest. This corpus features information about citation and reference papers for each paper. We use this information to construct feature vectors for these papers as described in Section 3.2. Stop words⁹ are eliminated from each user’s publication list and from the candidate papers to recommend. Stemming was performed using Porter Stemmer for English¹⁰ [22].

4.2 Evaluation Measure

As in standard information retrieval (IR), the top ranked documents are the most important to get correct, since users check these ranks more often. To properly account for this effect, we employ IR evaluation measures: (1) normalized discounted cumulative gain (NDCG) [12], and (2) mean reciprocal rank (MRR) [37].

4.2.1 Normalized Discounted Cumulative Gain (NDCG)

Discounted cumulative gain (DCG) is a measure that gives more weight to highly ranked documents and incorporates different relevance levels (relevant, and irrelevant) through different gain values.

$$DCG(i) = \begin{cases} G(1) & \text{if } i = 1 \\ DCG(i-1) + \frac{G(i)}{\log(i)} & \text{otherwise,} \end{cases}$$

where i denotes the i^{th} ranked position. In our work, the relevance level depends on just a binary notion of relevance: whether recommended papers are relevant or not to the user. We use $G(i) = 1$ for relevant search results and $G(i) = 0$ for irrelevant search results. The average normalized DCG over all users is selected to show the accuracy of recommendation. 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 $NDCG@N$ ($N = 5, 10$) for evaluation where N is the number of top- N papers recommended by our proposed approaches.

4.2.2 Mean Reciprocal Rank (MRR)

Mean reciprocal rank (MRR) indicates where in the ranking the first relevant item is returned by the system, averaged over all users. This measure provides insight in the ability of the system to return a relevant item at the top of the ranking.

4.3 Experimental Results

In this section, we show our experimental results for junior researchers and senior researchers. In the following, we denote citation and reference papers as “C” and “R,” respectively. In our experiments, we construct feature vectors for the candidate papers to recommend using the target paper only (ACL ARC paper, denoted as AP), the target paper and its citation papers (denoted as AP+C), the target paper and its reference papers (AP+R), and the target paper and both citation and reference papers (AP+C+R) defined by Equation (10). We also compare our proposed approach with ranking results obtained by the standard PageRank algorithm, which computes a global notion of importance of a scholarly paper.

⁹<ftp://ftp.cs.cornell.edu/pub/smart/english.stop>

¹⁰<http://www.tartarus.org/~martin/PorterStemmer/>

We evaluate the recommendation accuracy of our approach using these feature vectors for candidate papers and user profile described in the following.

4.3.1 Junior Researchers (*J*)

(J1) Using Only the Most Recent Paper

In this experiment (J1), we first verify recommendation accuracy using a user profile constructed by the single (thus the most recent) paper that a junior researcher has published, as shown in Figure 2. In this experiment, the user profile is defined by Equation (5). Note that, in this case, the single paper has no citing papers although it does possess reference papers, as shown in Figure 2. We can thus compare recommendation accuracy obtained by using the user profile constructed only by using the paper alone (Most recent paper, MP) versus a profile created from the paper and its reference papers (MP+R). Table 2 shows the recommendation accuracy evaluated with NDCG@5, 10 and MRR.

According to these results, recommendation accuracy obtained by user profile with references (MP+R) outperforms the accuracy obtained by most recent paper alone (MP), in most cases. In particular, the higher accuracy is obtained when we use the “SIM” weighting scheme. This is because “SIM” gives fine-grained score, but the other schemes do not. In this case, we achieve NDCG@5 of 0.457, NDCG@10 of 0.422, and MRR of 0.568. Four of 15 junior researchers achieved the recommendation accuracy, MRR of 1.0. This results show that we can recommend papers ranked at the top even if we construct user profile using only the most recent paper. As the “SIM” weighting scheme performs well, we employ “SIM” to weight citation and reference papers in the candidate papers to recommend, and reference papers in the most recent paper in user profile, for all of our subsequent experiments.

(J2) Using the Most Recent Paper with Pruning Its Reference Papers

In the previous experiment (J1), we used all of reference papers in the most recent paper to construct user profile, and all of citation and reference papers to construct feature vector for the candidate papers to recommend. However, some of the reference papers are, in fact, not closely related to the contents of the target paper, and may not be effective in characterizing the target paper. As such, we explore the pruning of references and citation papers that have low similarity to each target paper (both in the construction of the user profile and in candidate papers to recommend) by defining the threshold Th_j . In our preliminary study, we found that the cosine similarity between a target paper and its citation or reference papers reaches a maximum of 0.7. Therefore, we vary Th_j from 0 to 0.6. For example, if we set Th_j to 0.1, we prune the reference or citation papers of a target paper that have a similarity less than 0.1 and use the remaining papers to construct user profiles and feature vectors for candidate papers to recommend. After pruning, the weights described in (W1) to (W3) are used to construct the user profile defined by Equation (5). Figure 4 (a), (b) and (c) show the recommendation accuracy evaluated with NDCG@5, 10 and MRR, respectively. To facilitate direct comparison with the previous experiment, we show the best results obtained from (J1) as point (a0-J1), (b0-J1) and (c0-J1) in (a), (b) and (c), respectively, in Figure 4.

We obtain the best recommendation accuracy (NDCG@5:0.521, NDCG@10:0.459, MRR:0.624) when we set the threshold of similarity Th_j to 0.2, and assign “SIM” after pruning reference papers. Weighting schemes “LC” and “RPY” continue to underperform. As described in (J1), “SIM” gives fine-grained score compared with other schemes, which we believe holds as the explanation in this experiment as well.

Figure 4 shows concave down curves for all three evaluation metrics, as we vary the similarity threshold. This indicates that pruning can yield higher accuracy, but that there is a limit on its effectiveness. Overzealous pruning (when we set Th_j to greater threshold than 0.4) results in lower accuracy. We believe that overpruning results in sparse data, which damages our capability to represent

papers’ content effectively. When we explored the detailed results, we achieved very low recommendation accuracy for one of the junior researchers, in other words, NDCG@5 of 0.132, NDCG@10 of 0.215 and MRR of 0.200. But after pruning reference papers with $Th_j=0.2$, the accuracy is dramatically improved, NDCG@5 of 0.868, NDCG@10 of 0.721 and MRR of 1.00.

(J3) Results of Statistical Test

In order to verify whether the results are statistically significant or not, we perform paired *t*-tests for recommendation accuracy. We obtain accuracy by averaging results over each researcher. In such case, *t*-test is appropriate for testing the difference between means [30]. As shown in Table 2, in each evaluation measure, NDCG@5, 10 and MRR, the differences between recommendation accuracy obtained by user profile constructed using the most recent paper only (MP) and user profile (MP+R) constructed using “Weight:SIM” with pruning ($Th_j=0.2$) are statistically significant (*p*-value < 0.05).

4.3.2 Senior Researchers (*S*)

(S1) Using Only the Most Recent Paper

We start the set of experiments for senior researchers in the same way. We first verify recommendation accuracy using user profile constructed using only the most recent paper. In this experiment, the user profile is defined by Equation (7). Unlike junior researchers, senior researchers have several published papers in the past as shown in Figure 3. Therefore, we compare recommendation accuracy obtained by user profile constructed by the most recent paper only (MP), the most recent paper and its citation papers (MP+C), the most recent paper and its reference papers (MP+R), and the most recent paper and both citation and reference papers (MP+C+R). Table 3 shows the recommendation accuracy evaluated with NDCG@5, 10, and MRR.

According to these results, regarding user profile, the recommendation accuracy obtained by user profile (MP+C+R) outperforms that obtained by user profile (MP), in most cases. In particular, better accuracy is obtained when we use the “SIM” weighting scheme, achieving an NDCG@5:0.421, NDCG@10:0.382, and MRR:0.739. These results mirror the trend found in the case of junior researchers. However, the overall accuracy is low compared with the results of junior researchers. We hypothesize that this is because senior researchers have more diverse research interests. Therefore, it is difficult to construct user profile using only most recent paper. In the following experiments, we employ “SIM” to weight citation and reference papers in both the candidate papers and citation and reference papers in the user profile.

(S2) Using the Most Recent Paper with Pruning Its Citation and Reference Papers

Based on the same reason described in (J2), we introduce references and citation paper pruning based on the threshold of similarity Th_{ms} . Similar to the experiments in (J2), we vary the threshold of similarity from 0 to 0.6. After pruning, the weight defined in (W1) to (W3) is introduced to construct user profile defined by Equation (6). Figure 5 (a), (b) and (c) show the recommendation accuracy evaluated with NDCG@5, 10, and MRR, respectively. For direct comparison, we again show the best performance from (S1) as point (a0-S1), (b0-S1) and (c0-S1) in (a), (b) and (c), respectively, in Figure 5.

According to these results, we obtain the best recommendation accuracy when we set the similarity threshold Th_{ms} to 0.4 and assign “SIM” after pruning reference and citation papers (NDCG@5:0.478, NDCG@10:0.422, MRR:0.764). With regard to other weighting schemes, we observe the same tendency as experiment in (J2). In other words, “LC” and “RPY” do not bring very good results. These results also indicate that it is effective in pruning reference papers to construct user profiles and pruning both citation and reference papers to construct feature vector of the candidate papers to recommend. In our detailed analysis, we saw that for some researchers, the NDCG@5 obtained by pruning with $Th_{ms}=0.4$, is improved by more than 0.2 compared with that obtained in (S1).

Table 2: Recommendation accuracy for junior researchers evaluated with NDCG@5, NDCG@10 and MRR [“C”, “R”, and “C+R” denotes citation papers, reference papers, and both citation and reference papers, respectively]. “*” denotes difference between the results obtained by the most recent paper only (AP,MP) and [(AP+C+R),(MP+R, Weight:“SIM,” $Th_j = 0.2$)] is significant for $p < 0.05$.

NDCG@5		The Most recent Paper in user profile (MP)						
		Weight:“LC”		Weight:“SIM”			Weight:“RPY”	
		MP	MP+R	MP	MP+R	MP+R ($Th_j=0.2$)	MP	MP+R
ACL Papers to recommend (AP)	AP	<u>0.382</u>	0.442	0.382	0.443	-	0.382	0.431
	AP+C	0.388	0.429	0.390	0.435	-	0.389	0.438
	AP+R	0.402	0.405	0.427	0.451	-	0.404	0.440
	AP+C+R	0.418	0.445	0.435	0.457	0.521*	0.423	0.452

NDCG@10		The Most recent Paper in user profile (MP)						
		Weight:“LC”		Weight:“SIM”			Weight:“RPY”	
		MP	MP+R	MP	MP+R	MP+R ($Th_j=0.2$)	MP	MP+R
ACL Papers to recommend (AP)	AP	0.392	0.401	0.392	0.406	-	0.392	0.403
	AP+C	0.401	0.406	0.402	0.409	-	0.394	0.400
	AP+R	0.403	0.399	0.408	0.417	-	0.402	0.408
	AP+C+R	0.407	0.403	0.410	0.422	0.459*	0.408	0.415

MRR		The Most recent Paper in user profile (MP)						
		Weight:“LC”		Weight:“SIM”			Weight:“RPY”	
		MP	MP+R	MP	MP+R	MP+R ($Th_j=0.2$)	MP	MP+R
ACL Papers to recommend (AP)	AP	0.455	0.505	0.455	0.522	-	0.455	0.520
	AP+C	0.450	0.477	0.452	0.525	-	0.448	0.489
	AP+R	0.453	0.494	0.490	0.524	-	0.469	0.492
	AP+C+R	0.472	0.538	0.521	0.568	0.624*	0.515	0.526

(S3) Using Past Published Papers

Senior researchers differ from juniors in that they have a solid publication record. Focusing on this point, we have taken their publication history into account when generating recommendations. We also expect that more recent publications reflect their current interests more accurately. In this experiment (S3), we assess our methods for utilizing the past publication history, and introduce the forgetting factor (“Weight:FF”) defined earlier by Equation (8), to weight older papers less.

Figure 6 (a), (b), and (c) show the recommendation accuracy evaluated using NDCG@5, 10, and MRR, respectively. For comparison with our previous experiment, we denote the best results obtained from (S2) as point (a0-S2), (b0-S2), and (c0-S2) in (a), (b), and (c), respectively, in Figure 6. According to these results, we obtain the best results when the difference of published year d , and the value of γ are set to 3, and 0.2, respectively. In addition, when the value of γ is small, lower scores are assigned to the terms in older papers. However, in the case where γ is too large, the term scores of salient words in recent papers that are shared with older papers also become lower, causing reduced improvements in recommendation accuracy. Furthermore, when we construct user profile using recently published papers within three years (i.e., $d \leq 3$), recommendation accuracy tends to be improved. On the other hand, when we construct user profile using older papers (i.e., $d \geq 4$), recommendation accuracy is not improved so much. This is based on the characteristics of forgetting factor. In other words, the larger the value becomes, the more smaller the forgetting factor becomes because d is an integer. Thus, most of term scores in a user profile become small, and it results in low recommendation accuracy. Compared with the best results obtained in (S2), we observed that NDCG@10 is improved a lot for most senior researchers.

(S4) Results of Statistical Test

In order to verify whether the obtained results are statistically significant or not, we perform paired t -tests for recommendation accuracy obtained by the following user profile in (S2): (i) most recent paper only, and (ii) Weight:“SIM” ($Th_{ms} = 0.4$, the best accuracy) (underlined in Figure 5 (a), (b), and (c)). In addition, we also perform paired t -tests for recommendation accuracy obtained by the following user profile in (S3): (i) Weight:“SIM” ($Th_{ms} = 0.4$, the best accuracy in (S2)), and (ii) Weight:“FF” ($\gamma = 0.2$, $d = 3$, the best accuracy in (S3)) (underlined in Figure 6 (a), (b), and (c)).

As shown in Table 4, in each evaluation measure, NDCG@5, 10, and MRR, the differences between recommendation accuracy obtained by user profile constructed using the most recent paper only and user profile constructed using weight “SIM” with pruning

($Th_{ms} = 0.4$) are statistically significant (p -value < 0.01 in NDCG@5 and MRR, p -value < 0.05 in NDCG@10). In addition, the differences between recommendation accuracy obtained when using a user profile constructed using weight “SIM” with pruning ($Th_{ms} = 0.4$) and user profile constructed using past papers (Weight:“FF,” $\gamma = 0.2$, $d = 3$) are also statistically significant (p -value < 0.05).

4.3.3 Comparison of PageRank and Our Approach

In Figures 4 to 6, for both junior researchers and senior researchers, we also show the results of using PageRank for recommendation (shown as a single horizontal line). In all cases, the results obtained by PageRank score give much lower accuracy than that of our proposed approaches. As older papers tend to be cited more often than newer papers, PageRank scores tend to favor older papers. Even among papers of the same age, more popular topics are accorded higher PageRank. Finally, PageRank yields the same generic score for all researchers, and cannot customize results for a particular researcher. All of these facts contribute to the lower recommendation accuracy when using PageRank.

4.3.4 Summary of Obtained Results

Our key result is that incorporating the context of a paper, in the form of references and, when available, citations, results in improved recommendation accuracy. Furthermore, pruning the context by similarity weighting is helpful in filtering noise from the boosted signal. An important factor that we discovered is that the pruning is sensitive; the threshold must be appropriately set to balance the benefit of noise filtering against the side effect of data sparsity. For senior researchers, we show that while the modeling the most recent paper alone can generate good recommendations, better accuracy is obtained by using their publication list appropriately. We show that the appropriate weighting of papers of up to three years from the most recent paper is the most effective. Modeling papers beyond this threshold is more computationally expensive and does not yield better performance.

5. CONCLUSION

We have proposed a generic model towards recommending scholarly papers relevant to a researcher’s interests by capturing their research interests through their past publications. The key contribution of our model is in its representation of the user profile as not only past publications but also incorporating its neighboring papers such as citation and reference papers as context. In our analysis, we have verified the effectiveness of our approach for two classes of researchers: junior researchers, and senior researchers. We evaluated recommendation accuracy using NDCG and MRR, and achieve consistent results using both metrics. We

Table 3: Recommendation accuracy for senior researchers evaluated with NDCG@5, NDCG@10, and MRR [“C”, “R”, and “C+R” denotes citation papers, reference papers, and both citation and reference papers, respectively].

		NDCG@5											
		The Most recent Paper in user profile (MP)											
		Weight:“LC”				Weight:“SIM”				Weight:“RPY”			
		MP	MP+C	MP+R	MP+C+R	MP	MP+C	MP+R	MP+C+R	MP	MP+C	MP+R	MP+C+R
ACL Papers to recommended (AP)	AP	0.325	0.334	0.390	0.401	0.325	0.351	0.406	0.401	0.325	0.338	0.395	0.401
	AP + C	0.332	0.341	0.378	0.384	0.335	0.383	0.399	0.406	0.334	0.381	0.401	0.404
	AP + R	0.345	0.408	0.353	0.410	0.374	0.373	0.416	0.418	0.348	0.393	0.402	0.408
	AP + C+R	0.367	0.390	0.390	0.417	0.384	0.402	0.419	0.421	0.374	0.415	0.413	0.418
		NDCG@10											
		The Most recent Paper in user profile (MP)											
		Weight:“LC”				Weight:“SIM”				Weight:“RPY”			
		MP	MP+C	MP+R	MP+C+R	MP	MP+C	MP+R	MP+C+R	MP	MP+C	MP+R	MP+C+R
ACL Papers to recommended (AP)	AP	0.305	0.323	0.351	0.362	0.305	0.346	0.365	0.362	0.305	0.318	0.351	0.362
	AP + C	0.325	0.353	0.376	0.369	0.335	0.363	0.368	0.367	0.327	0.353	0.348	0.371
	AP + R	0.327	0.362	0.368	0.374	0.353	0.356	0.367	0.373	0.335	0.365	0.366	0.376
	AP + C+R	0.343	0.375	0.372	0.377	0.367	0.372	0.376	0.382	0.353	0.377	0.373	0.379
		MRR											
		The Most recent Paper in user profile (MP)											
		Weight:“LC”				Weight:“SIM”				Weight:“RPY”			
		MP	MP+C	MP+R	MP+C+R	MP	MP+C	MP+R	MP+C+R	MP	MP+C	MP+R	MP+C+R
ACL Papers to recommended (AP)	AP	0.621	0.657	0.670	0.709	0.621	0.696	0.688	0.709	0.621	0.696	0.688	0.709
	AP + C	0.615	0.696	0.688	0.696	0.621	0.696	0.692	0.727	0.615	0.696	0.656	0.696
	AP + R	0.618	0.651	0.659	0.696	0.658	0.657	0.648	0.697	0.637	0.657	0.661	0.657
	AP + C+R	0.637	0.709	0.709	0.710	0.689	0.696	0.728	0.739	0.681	0.688	0.696	0.709

Table 4: Results of paired *t*-test for senior researchers (“ \ll ” and “ \ll ” denote significance levels of $p < 0.01$ and $p < 0.05$, respectively).

NDCG@5	(AP, MP) \ll (AP+C+R, MP+C+R [Weight:“SIM”, $T h_{m_s}=0.4$]) \ll (AP+C+R, MP+C+R [Weight:“SIM”, $T h_{m_s}=0.4, \gamma = 0.2, d = 3$])
NDCG@10	(AP, MP) \ll (AP+C+R, MP+C+R [Weight:“SIM”, $T h_{m_s}=0.4$]) \ll (AP+C+R, MP+C+R [Weight:“SIM”, $T h_{m_s}=0.4, \gamma = 0.2, d = 3$])
MRR	(AP, MP) \ll (AP+C+R, MP+C+R [Weight:“SIM”, $T h_{m_s}=0.4$]) \ll (AP+C+R, MP+C+R [Weight:“SIM”, $T h_{m_s}=0.4, \gamma = 0.2, d = 3$])

observe higher recommendation accuracy when our model prunes neighboring papers with low similarity, effectively enhancing the signal of the original topic of the paper.

Torres *et al.* [36] pointed out the importance of recommendations in the digital libraries based on the experience level of the individual target researcher. Our work is an implemented content-based solution that addresses this challenge. In future work, we plan to compare other approaches to user profile construction reviewed in Section 2.3. In addition, for senior researchers, we plan to develop methods for helping recommend interdisciplinary papers that could encourage a push to new frontiers. For junior researchers, we plan to develop methods for recommending papers that are easier to understand to quickly acquire knowledge about their intended research.

6. ACKNOWLEDGMENTS

We thank the many researchers who volunteered their time to mark relevant papers for this study. Without their help, this work would not have been possible.

7. REFERENCES

- [1] M. Balabanovic and Y. Shoham. Fab: Content-Based, Collaborative Recommendation. *Communications of the ACM*, 40(3):66–72, 1997.
- [2] C. Basu, H. Hirsh, and W. Cohen. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. In *Proc. of the 15th National Conference on Artificial Intelligence (AAAI '98)*, pages 714–720, 1998.
- [3] S. Bird, R. Dale, B. J. Dorr, B. Gibson, M. T. Joseph, M.-Y. Kan, D. Lee, B. Powley, D. R. Radev, and Y. F. Tan. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In *Proc. of the 6th International Conference on Language Resources and Evaluation Conference (LREC'08)*, pages 1755–1759, 2008.
- [4] J. Bollen, M. A. Rodriguez, and H. V. D. Sompel. Journal Status. *Scientometrics*, 69(3):669–687, 2006.
- [5] J. S. Breese, D. Heckerman, and C. Kadie. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In *Proc. of the 14th Conference on Uncertainty in Artificial Intelligence (UAI '98)*, pages 43–52, 1998.
- [6] C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri. Know Your Neighbors: Web Spam Detection Using Web Topology. In *Proc. of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2007)*, pages 423–430, 2007.
- [7] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding Scientific Gems with Google’s PageRank Algorithm. *Journal of Informetrics*, 1(1):8–15, 2007.
- [8] W. Chu and S.-T. Park. Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models. In *Proc. of the 18th International World Wide Web Conference (WWW2009)*, 2009, 691–700.
- [9] E. Garfield. Citation Indexing: Its Theory and Application in Science, Technology, and Humanities. *New York: John Wiley and Sons*, 1979.
- [10] D. Goldberg, D. Nichols, B. M. Oki, and D. B. Terry. Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35(12):61–70, 1992.
- [11] 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.
- [12] 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.
- [13] H. Kautz, B. Selman, and M. Shah. Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM*, 40(3):63–65, 1997.
- [14] H.-N. Kim, I. Ha, S.-H. Lee, and G.-S. Jo. A Collaborative Approach to User Modeling for Personalized Content Recommendations. In *Proc. of the 11th International Conference on Asian Digital Libraries (ICADL2008)*, Lecture Notes in Computer Science (LNCS), Vol. 5362, pages 215–224. Springer-Verlag, 2008.
- [15] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, 40(3):77–87, 1997.
- [16] M. Krapivin and M. Marchese. Focused PageRank in Scientific Papers Ranking. In *Proc. of the 11th International Conference on Asian Digital Libraries (ICADL 2008)*, Lecture Notes in Computer Science (LNCS), Vol. 5362, pages 144–153, 2008.
- [17] 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.
- [18] P. Melville, R. J. Mooney, and R. Nagarajan. Content-Boosted Collaborative Filtering for Improved Recommendations. In *Proc. of the 18th National Conference on Artificial Intelligence (AAAI2002)*, pages 187–192, 2002.
- [19] F. Narin. Evaluative Bibliometrics: The Use of Publication and

- Citation Analysis in the Evaluation of Scientific Activity. *Cherry Hill, N.J.: Computer Horizons*, 1976.
- [20] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report SIDL-WP-1999-0120, Stanford Digital Library Technologies Project, 1998.
- [21] S.-T. Park, D. Pennock, O. Madani, N. Good, and D. DeCoste. Naïve Filterbots for Robust Cold-Start Recommendations. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, pages 699–705, 2006.
- [22] M. F. Porter. An Algorithm for Suffix Stripping. *Program*, 14(3):pages 130–137, 1980.
- [23] X. Qi and B. D. Davison. Classifiers Without Borders: Incorporating Fielded Text From Neighboring Web Pages. In *Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008)*, pages 643–650, 2008.
- [24] P. Resnick, N. Iacovou, M. Suchak, and J. R. P. Bergstrom. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In *Proc. of the ACM 1994 Conference on Computer Supported Cooperative Work (CSCW '94)*, pages 175–186, 1994.
- [25] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [26] B. M. Sarwar, G. Karypis, and J. A. Konstan. Analysis of Recommendation Algorithms for E-commerce. In *Proc. of the 2nd ACM Conference on Electronic Commerce (EC '00)*, pages 158–167, 2000.
- [27] H. Sayyadi and L. Getoor. FutureRank: Ranking Scientific Articles by Predicting their Future PageRank. In *Proc. of the 9th SIAM International Conference on Data Mining*, pages 533–544, 2009.
- [28] C. Shahabi and Y.-S. Chen. An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback. *Distributed and Parallel Databases*, 14(3):173–192, 2003.
- [29] X. Shen, B. Tan, and C. Zhai. Context-Sensitive Information Retrieval Using Implicit Feedback. In *Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005)*, pages 43–50, 2005.
- [30] M. D. Smucker, J. Allan, and B. Carterette. A Comparison of Statistical Significance Tests for Information Retrieval. In *Proc. of the 16th International Conference on Information and Knowledge Management (CIKM'07)*, 2007. 623-632.
- [31] K. Sugiyama, K. Hatano, and M. Yoshikawa. Adaptive Web Search Based on User Profile Constructed without Any Effort from Users. In *Proc. of the 13th International World Wide Web Conference (WWW2004)*, pages 675–684, 2004.
- [32] K. Sugiyama, K. Hatano, M. Yoshikawa, and S. Uemura. Refinement of TF-IDF Schemes for Web Pages Using their Hyperlinked Neighboring Pages. In *Proc. of the 14th ACM Conference on Hypertext and Hypermedia (HT '03)*, pages 198–207, 2003.
- [33] Y. Sun and C. Giles. Popularity Weighted Ranking for Academic Digital Libraries. In *Proc. of the 29th European Conference on Information Retrieval (ECIR 2007)*, pages 605–612, 2007.
- [34] B. Tan, X. Shen, and C. Zhai. Mining Long-Term Search History to Improve Search History. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'06)*, pages 718–723, 2006.
- [35] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing Search via Automated Analysis of Interests and Activities. In *Proc. of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2005)*, pages 449–456, 2005.
- [36] R. Torres, S. M. McNee, M. Abel, J. A. Konstan, and J. Riedl. Enhancing Digital Libraries with TechLens. In *Proc. of the 4th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2004)*, pages 228–236, 2004.
- [37] E. M. Voorhees. The TREC-8 Question Answering Track Report. In *Proc. of the 8th Text REtrieval Conference (TREC-8)*, pages 77–82, 1999.
- [38] R. W. White, P. Bailey, and L. Chen. Predicting User Interests from Contextual Information. In *Proc. of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2009)*, pages 363–370, 2009.
- [39] X. Su and T. M. Khoshgoftaar and R. Greiner. Imputed Neighborhood Based Collaborative Filtering. In *Proc. of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2008)*, pages 633–639, 2008.
- [40] 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.

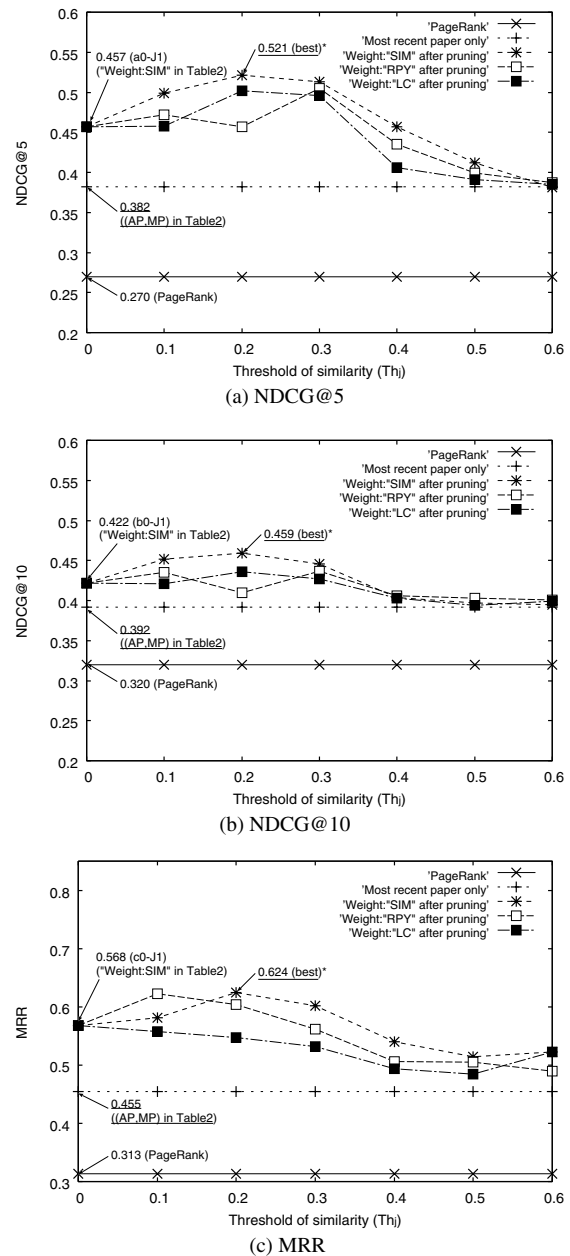
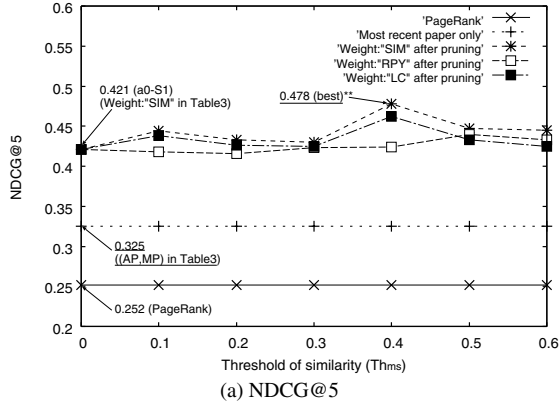
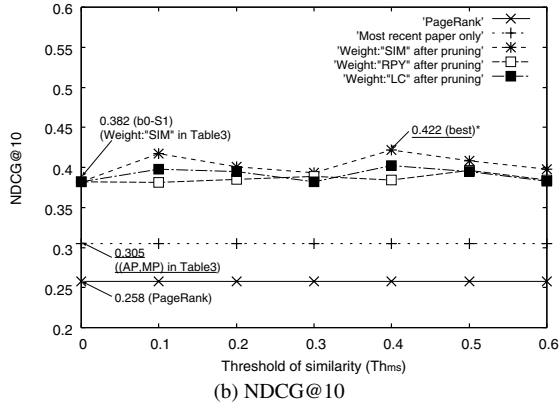


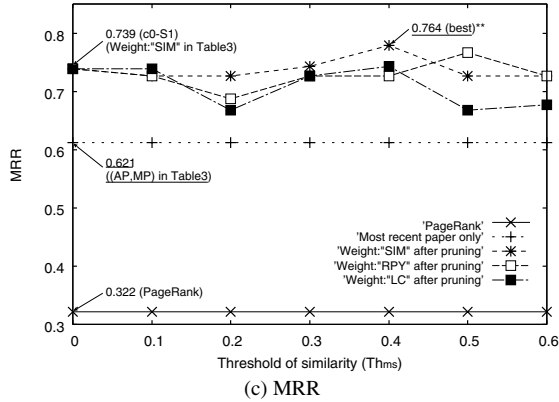
Figure 4: Recommendation accuracy for junior researchers obtained by user profile constructed using the most recent paper with reference paper pruning [(a) NDCG@5, (b) NDCG@10, and (c) MRR]. As in Table 2, “*” denotes difference between ‘Most recent paper only’ and the best value of ‘Weight: “SIM” ($Th_j=0.2$)’ is significant for $p < 0.05$.



(a) NDCG@5

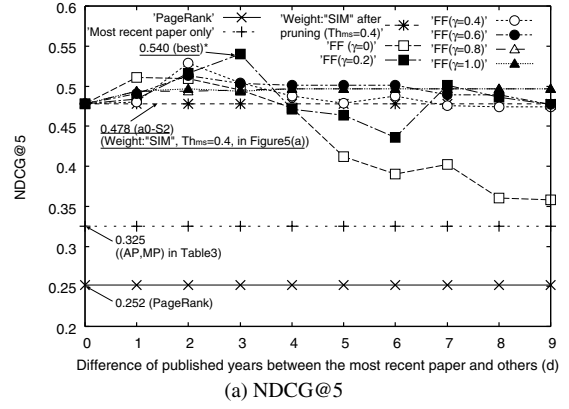


(b) NDCG@10

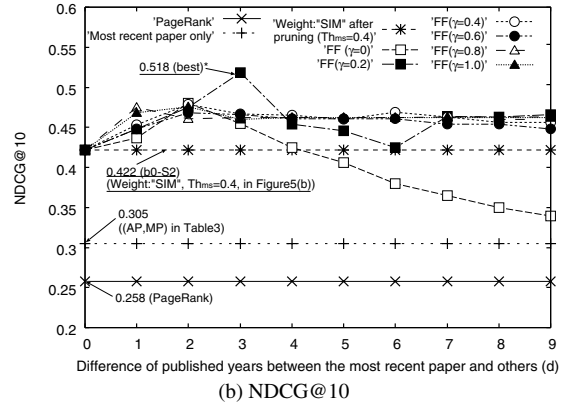


(c) MRR

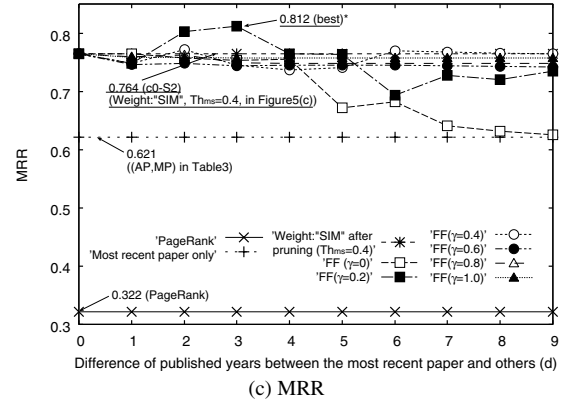
Figure 5: Recommendation accuracy for senior researchers obtained by user profile constructed using the most recent paper with citation and reference paper pruning [(a) NDCG@5, (b) NDCG@10, and (c) MRR]. As in Table 4, “**” and “*” denote the difference between ‘Most recent paper only’ and the best value of ‘Weight:SIM’ ($Th_{ms}=0.4$) is significant for $p < 0.01$ and $p < 0.05$, respectively.



(a) NDCG@5



(b) NDCG@10



(c) MRR

Figure 6: Recommendation accuracy for senior researchers obtained by user profile constructed using past published papers [(a) NDCG@5, (b) NDCG@10, and (c) MRR]. As in Table 4, “*” denotes difference between the best value of ‘Weight:SIM’ ($Th_{ms}=0.4$) and the best value of ‘Weight:FF’ ($\gamma = 0.2, d = 3$) is significant for $p < 0.05$.