


Exploiting Potential Citation Papers in Scholarly Paper Recommendation

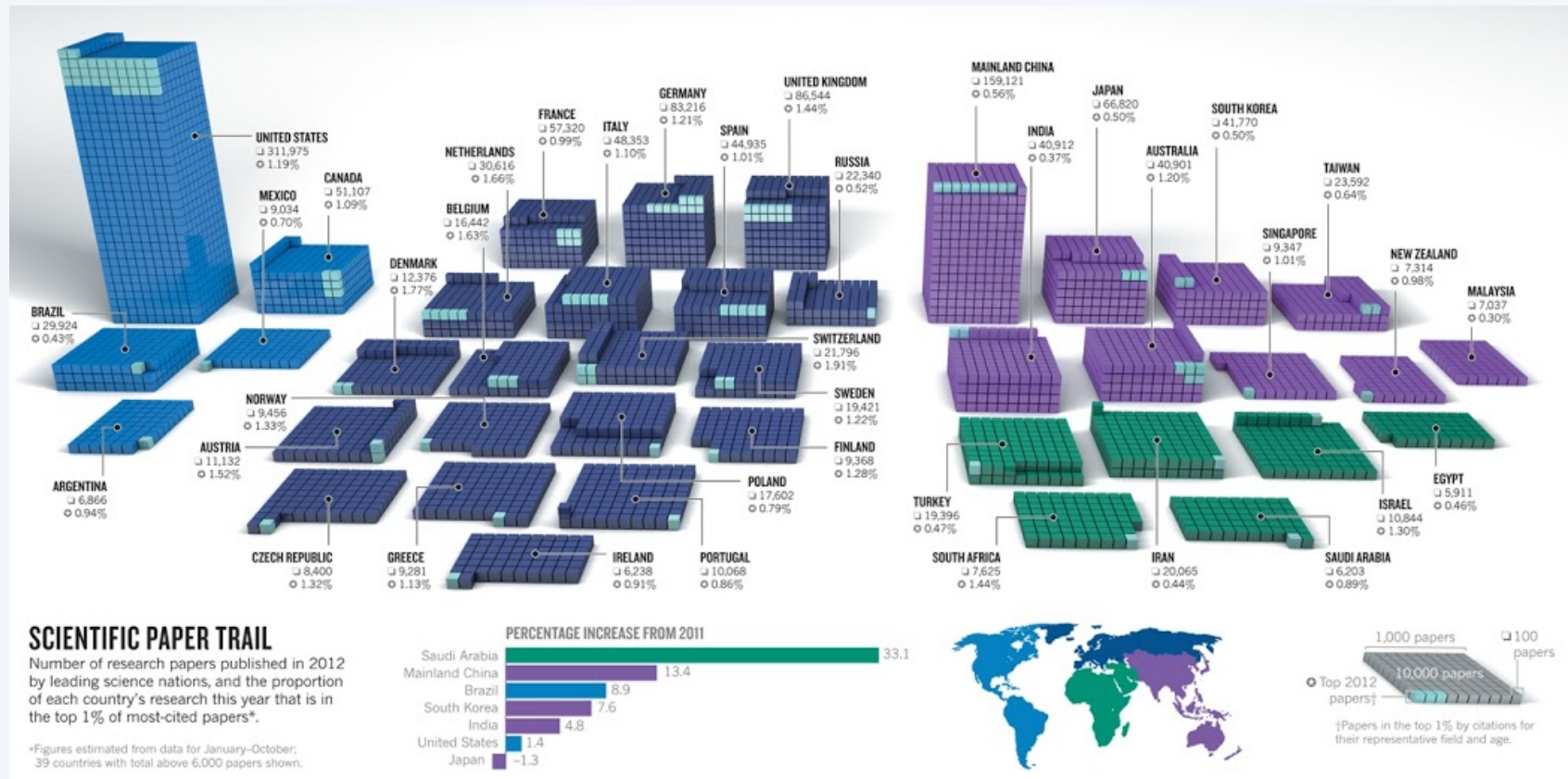
Kazunari Sugiyama, 

Min-Yen Kan, 

National University of Singapore 

Introduction

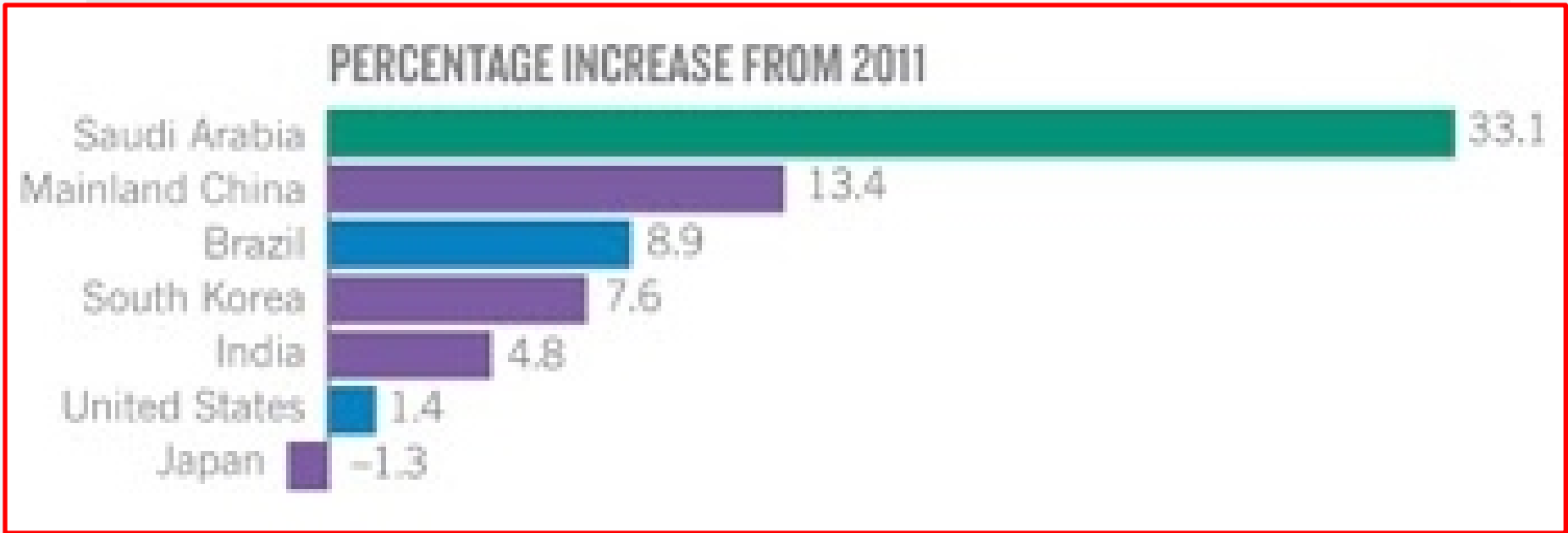
How many papers are published in 2012?



Source: Originally pinned from Nature (<http://www.nature.com/news/366-days-2012-in-review-1.12042>), with data from Thomson Reuters/Essential Science Indicators.

Introduction

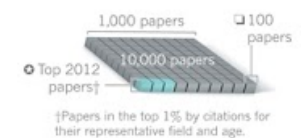
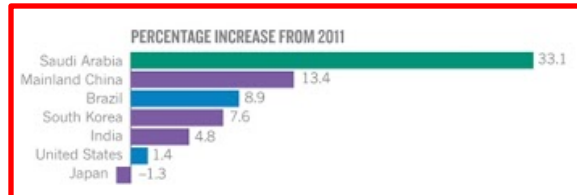
How many papers are published in 2012?



SCIENTIFIC PAPER TRAIL

Number of research papers published in 2012 by leading science nations, and the proportion of each country's research this year that is in the top 1% of most-cited papers*.

*Figures estimated from data for January–October; 39 countries with total above 6,000 papers shown.



Source: Originally pinned from Nature (<http://www.nature.com/news/366-days-2012-in-review-1.12042>), with data from Thomson Reuters/Essential Science Indicators.

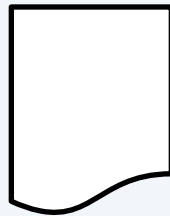
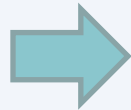
Introduction

Recommendation System

(Especially, content-based system)



Researcher

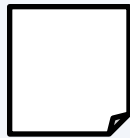
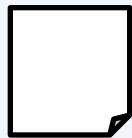
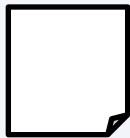


User profile U



Candidate papers to recommend

F_1, F_2, \dots, F_P



Recommended papers

Introduction

Feature vector construction for candidate papers

Fragments

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.

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

Full text

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, to 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 '11, July 22–26, 2011, Indianapolis, Indiana, USA.
Copyright 2011 ACM 978-1-4503-2071-1/11/07...\$15.00.

model. User systems require a user to prepare query manually without a bibliography that indicates locations where citations are located. User systems require a user to prepare query manually without a bibliography that indicates locations where citations are located. User systems require a user to prepare query manually without a bibliography that indicates locations where citations are located.

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

Figure 1(a). In this past work, we found that when the test of each

ir sub-
analyz-
ers are
Full text
discov-

to dis-
links
(2) en-
machine
a letter
dread
of com-
8) pro-
page
section,
with the
and the
restate-
on a
to sign-
hybrid

1) condi-
tion sys-
[3], our
sign the
and en-
mental
suble-
model
content
s lemm-
corpora

0.871*
0.779*
0.815*
0.811
0.818
0.826

restate-
the or-
0.826

list of
k of the
task
between
ms with
ing for
W/ACM
covery
and 2,
arbitra-
pedia's
vectors
cies on
sation
3-106,
we skip
merely
gnoring
the 19th
edge

RES. A
Proc.
of the 9th ACM/IEEE Joint Conference on Digital Libraries
(JCDL 2009), pages 203–211, 2009.

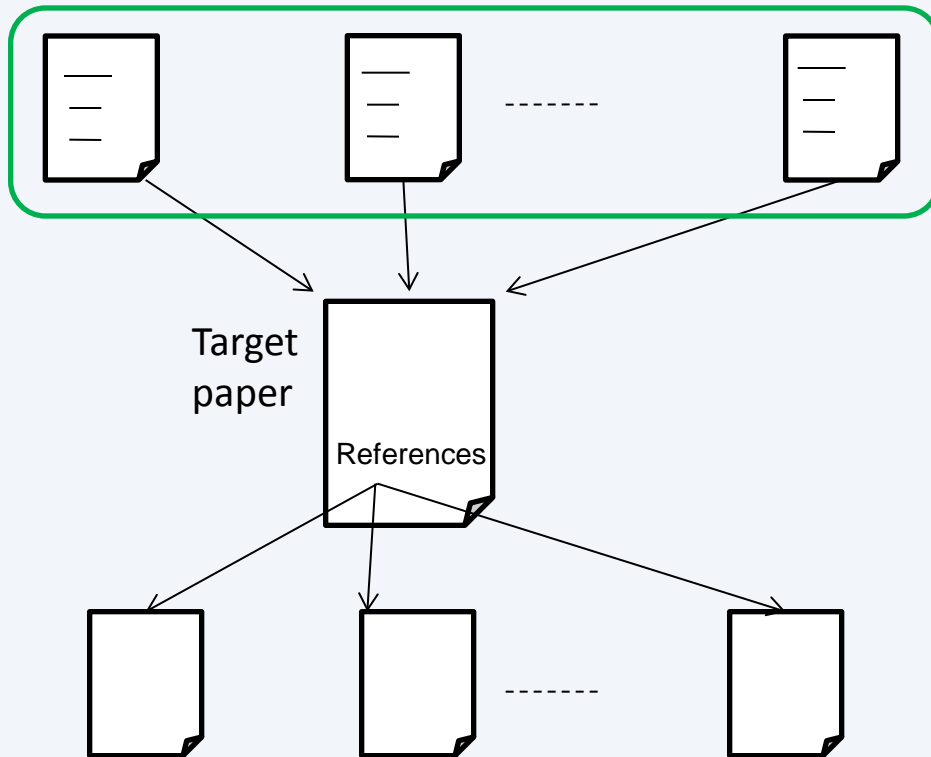
[18] R. Torres, S. M. McInerney, M. Abel, J. A. Konstan, and J. Riedl. Enhancing Digital Libraries with TechLms. In

Ranking-Oriented CADL Recommender System. In Proc. of the 9th ACM/IEEE Joint Conference on Digital Libraries (JCDL 2009), pages 203–211, 2009.

Introduction

Feature vector construction for candidate papers

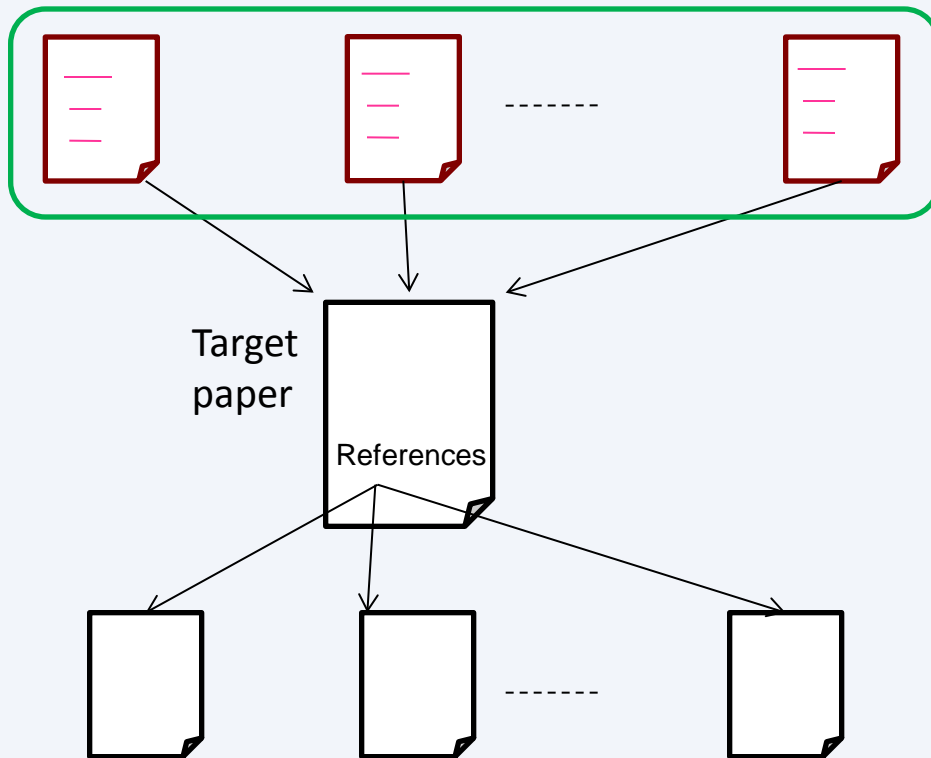
Citation and Reference Papers



Introduction

Feature vector construction for candidate papers

Citation and Reference Papers



Citation papers:

Endorsement of the target paper

Full text

Fragments

“abstract,” “introduction,”

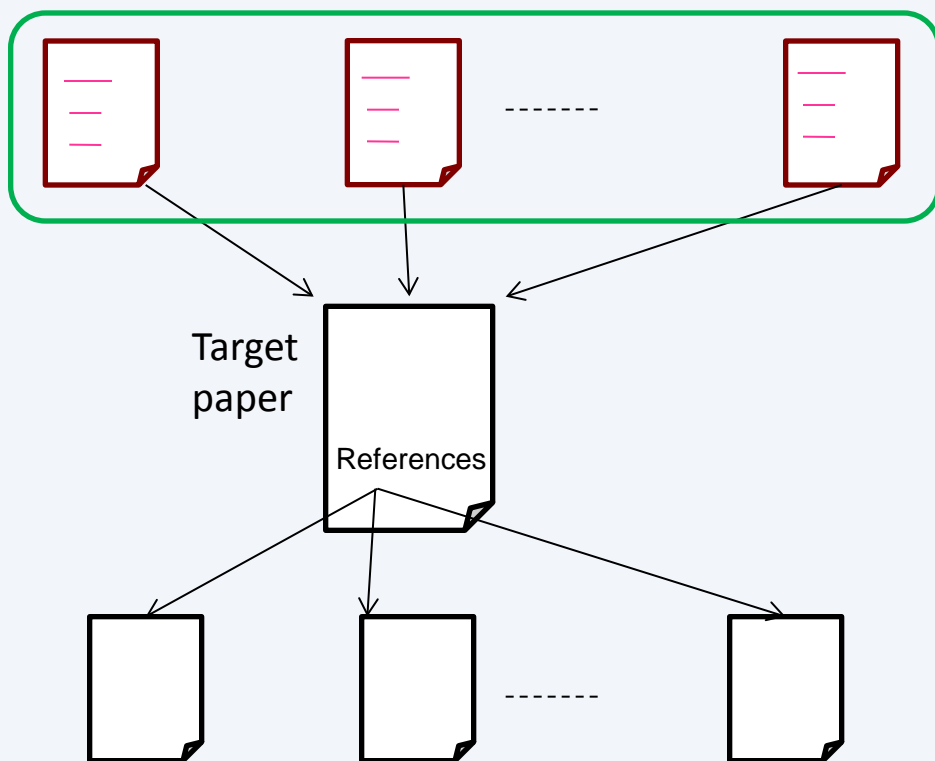
“conclusion,” ...



Introduction

Feature vector construction for candidate papers

Citation and Reference Papers



Citation papers:

Endorsement of the target paper

Full text

Fragments

“abstract,” “introduction,”
“conclusion,” ...

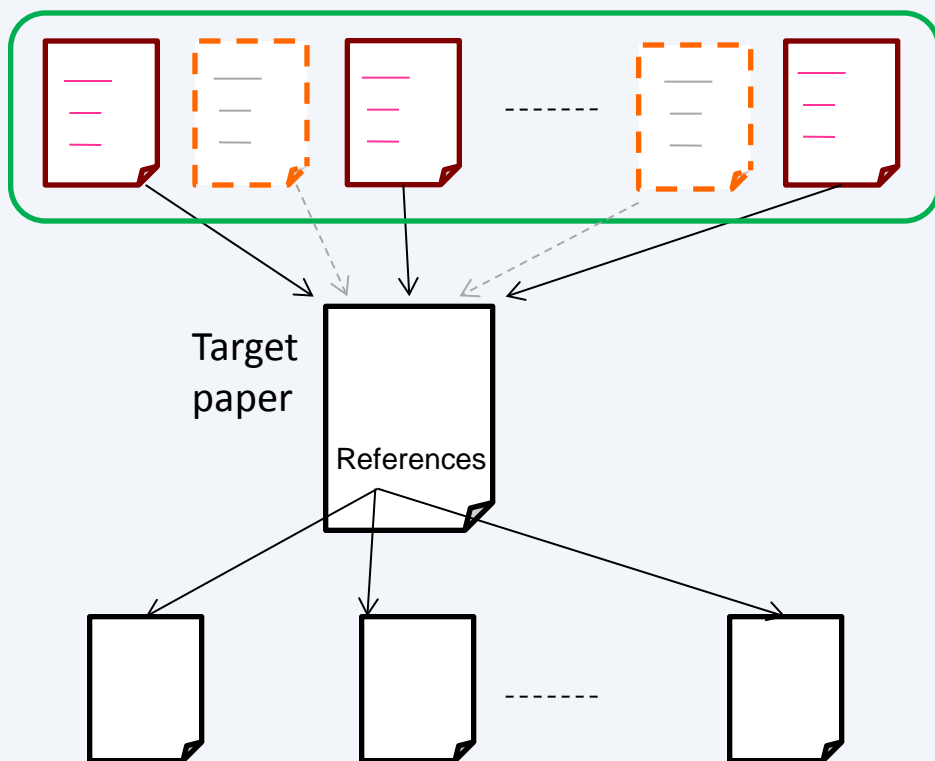
Authors of citation papers

- May not cite relevant papers due to space limit
- Unaware of the relevant papers

Introduction

Feature vector construction for candidate papers

Citation and Reference Papers



Citation papers:

Endorsement of the target paper

Full text

Fragments

“abstract,” “introduction,”
“conclusion,” ...

Authors of citation papers

- May not cite relevant papers due to space limit
- Unaware of the relevant papers



Potential citation papers

Our Goal

- To find potential citation papers to model candidate papers to recommend much better for better recommendation
- To refine the use of citation papers in characterizing candidate papers to recommend using fragments



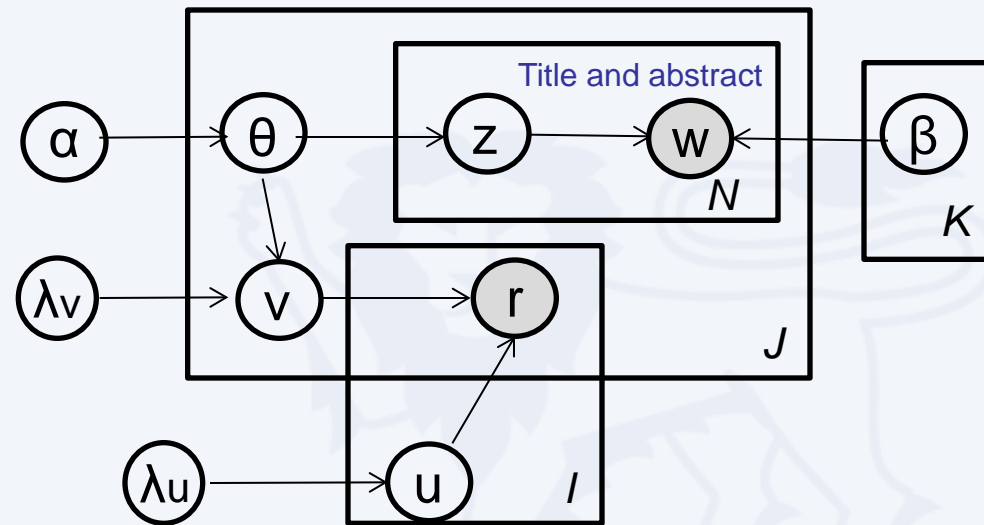
Related Work

Scholarly Paper Recommendation

[Wang and Blei., KDD'11]: Collaborative topic regression
 Combines ideas from collaborative filtering and content analysis
 based on probabilistic topic modeling

	p_1	p_2	-----	p_j	-----	p_N
u_1				0		
u_2		1		1		1
⋮						
u_i	1	1		r_{ij}		0
⋮						
u_U		0		1		

$r_{ij} \in \{0,1\}$ whether user u_i includes paper p_j
 In the user's preference



Outline of Baseline System

[Sugiyama and Kan, JCDL'10]

(1) Construct user profile
from each researcher's
past papers



Researcher

\mathbf{P}_{user}

(2) Compute similarity between

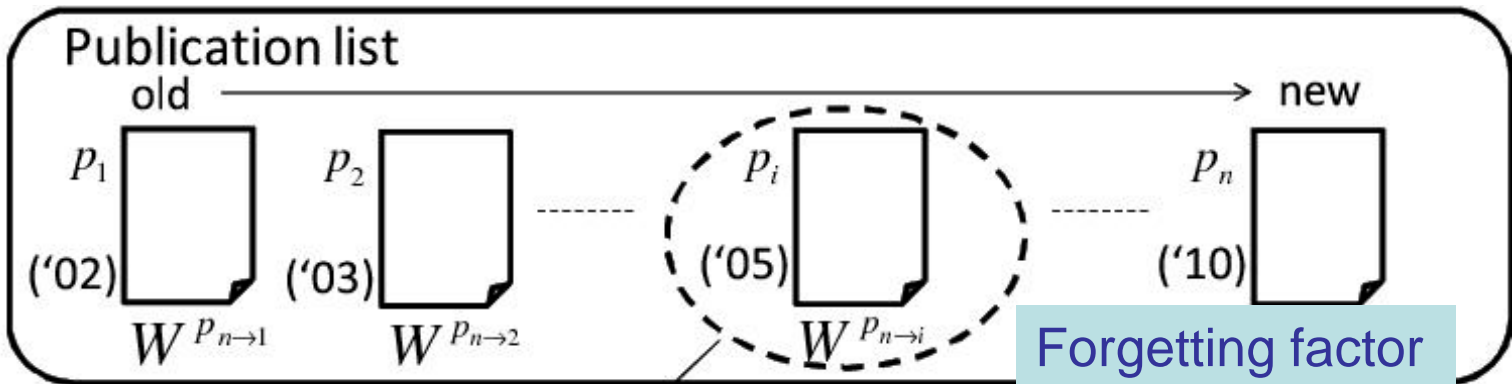
\mathbf{P}_{user} and $\mathbf{F}^{P_{rec_j}}$ ($j = 1, \dots, t$)



Candidate papers to recommend

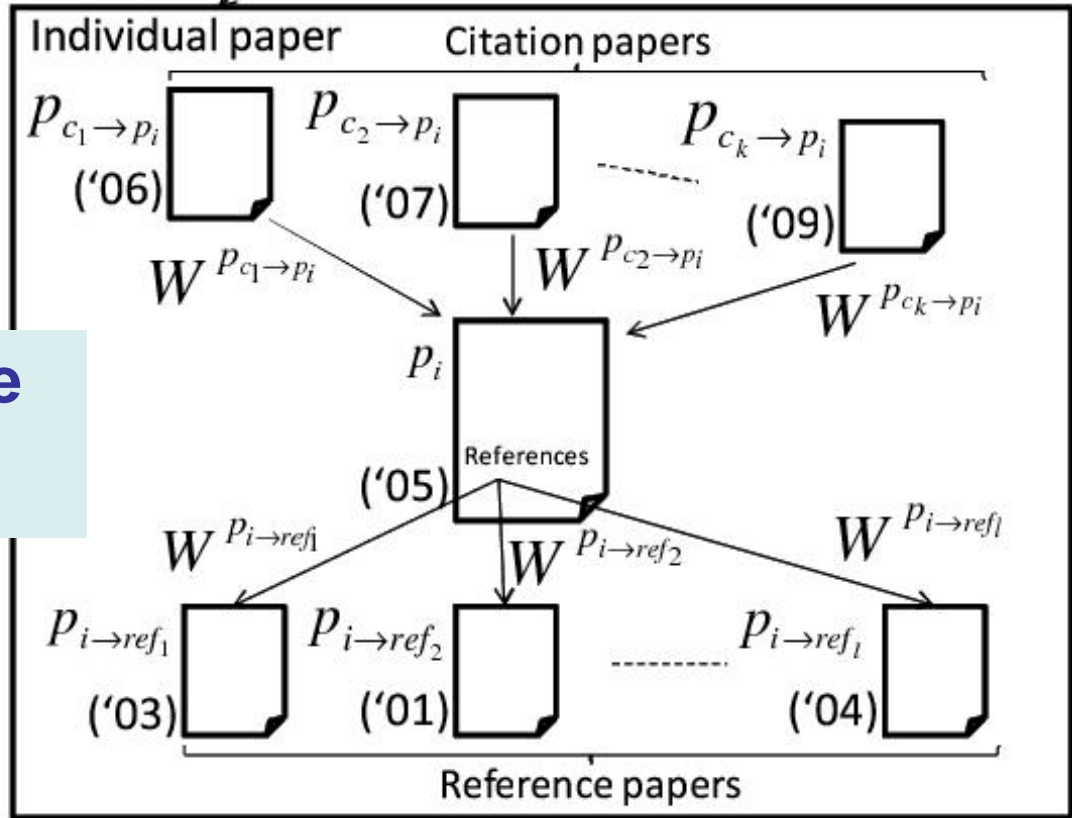
$\mathbf{F}^{P_{rec_1}}$ to $\mathbf{F}^{P_{rec_t}}$

(3) Recommend papers
with high similarity



Relation between citation or reference papers and P_i

Weighting scheme
Cosine similarity



Outline of Baseline System

[Sugiyama and Kan, JCDL'10]

- (1) Construct user profile
from each researcher's
past papers

\mathbf{P}_{user}



Researcher

- (2) Compute similarity between

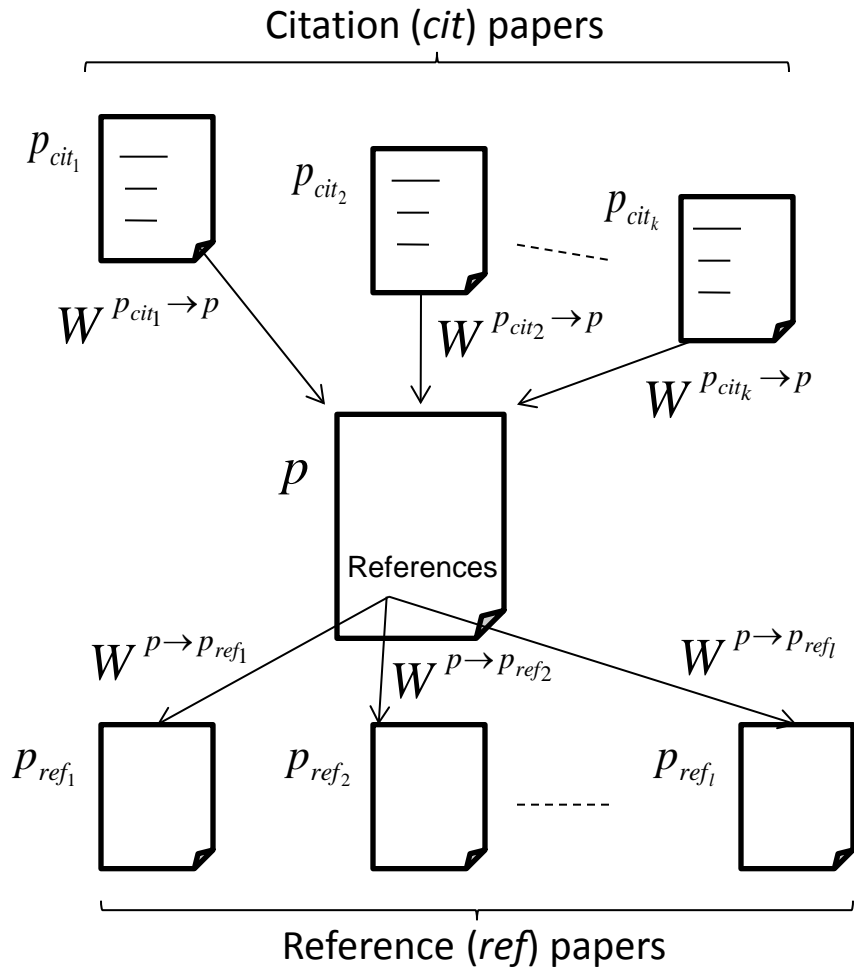
\mathbf{P}_{user} and $\mathbf{F}^{P_{rec_j}}$ ($j = 1, \dots, t$)




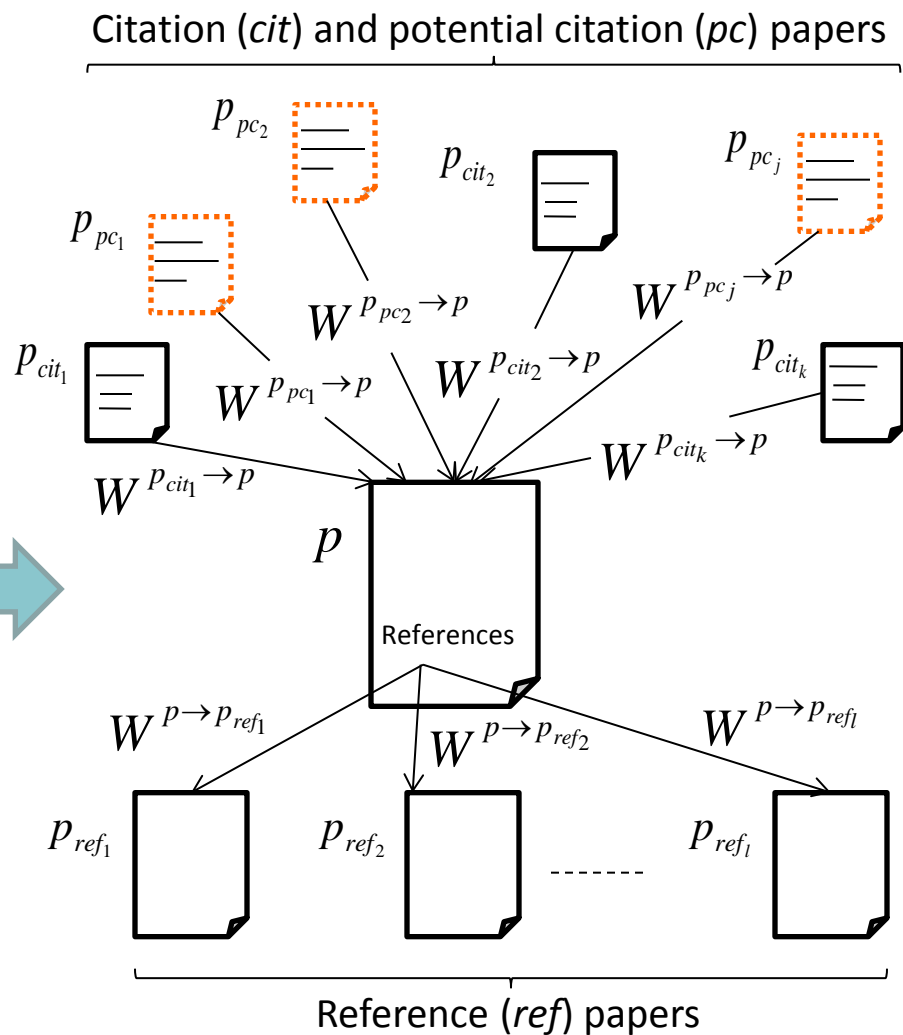
Candidate papers to recommend

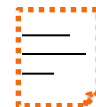
$\mathbf{F}^{P_{rec_1}}$ to $\mathbf{F}^{P_{rec_t}}$

- (3) Recommend papers
with high similarity



 : Citation (*cit*) paper



 : Potential citation (*pc*) paper

Proposed Method

(1) Leveraging Potential Citation Papers

(2) Leveraging Fragments in Potential Citation Papers



Proposed Method

(1) Leveraging Potential Citation Papers

How are potential citation papers discovered?



Proposed Method

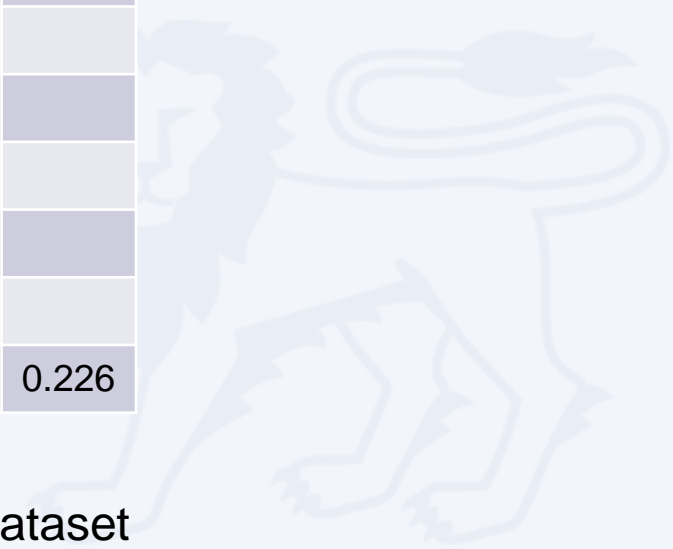
(1) Leveraging Potential Citation Papers

How are potential citation papers discovered?

	p_{cit_1}	p_{cit_2}	p_{cit_3}	-----	p_{cit_j}	-----	$p_{cit_{N-2}}$	$p_{cit_{N-1}}$	p_{cit_N}
p_1		0.212			0.735		0.687		
p_2	0.656				0.328		0.436		
p_3		0.764					0.527		0.385
⋮									
p_{tgt}		0.581					0.330		
⋮									
p_{N-1}	0.383				0.248			0.176	
p_{N-2}		0.654					0.525		
p_N		0.265					0.430		0.226

p_i ($i = 1, 2, \dots, N$) : All papers in dataset

p_{cit_j} ($j = 1, 2, \dots, N$): Papers as citation papers in dataset

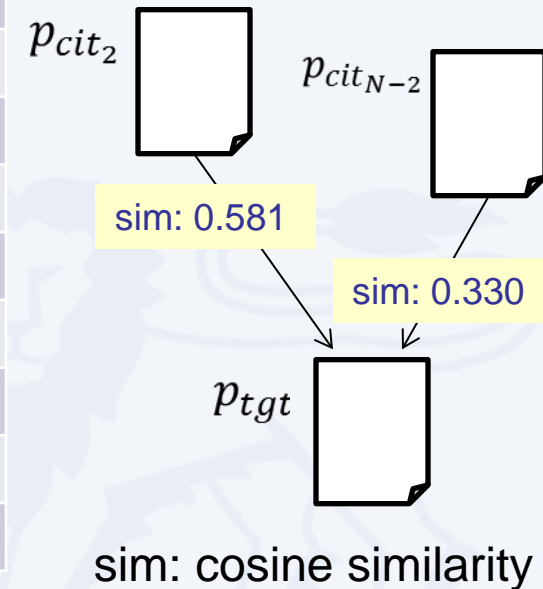


Proposed Method

(1) Leveraging Potential Citation Papers

How are potential citation papers discovered?

	p_{cit_1}	p_{cit_2}	p_{cit_3}	-----	p_{cit_j}	-----	$p_{cit_{N-2}}$	$p_{cit_{N-1}}$	p_{cit_N}
p_1		0.212			0.735		0.687		
p_2	0.656				0.328		0.436		
p_3		0.764					0.527		0.385
⋮									
p_{tgt}		0.581					0.330		
⋮									
p_{N-1}	0.383				0.248			0.176	
p_{N-2}		0.654					0.525		
p_N		0.265					0.430		0.226



p_i ($i = 1, 2, \dots, N$) : All papers in dataset

p_{cit_j} ($j = 1, 2, \dots, N$): Papers as citation papers in dataset

Proposed Method

(1) Leveraging Potential Citation Papers

How are potential citation papers discovered?

	p_{cit_1}	p_{cit_2}	p_{cit_3}	-----	p_{cit_j}	-----	$p_{cit_{N-2}}$	$p_{cit_{N-1}}$	p_{cit_N}
p_1		0.212			0.735		0.687		
p_2	0.656				0.328		0.436		
p_3		0.764					0.527		0.385
⋮									
p_{tgt}		0.581					0.330		
⋮									
p_{N-1}	0.383				0.248			0.176	
p_{N-2}		0.654					0.525		
p_N		0.265					0.430		0.226

Pearson correlation

0.538

0.216

0.475

0.304

0.513

0.487

p_i ($i = 1, 2, \dots, N$) : All papers in dataset

p_{cit_j} ($j = 1, 2, \dots, N$): Papers as citation papers in dataset



Neighborhood of the target paper (e.g., set to 4)

Proposed Method

(1) Leveraging Potential Citation Papers

How are potential citation papers discovered?

	p_{cit_1}	p_{cit_2}	p_{cit_3}	-----	p_{cit_j}	-----	$p_{cit_{N-2}}$	$p_{cit_{N-1}}$	p_{cit_N}
p_1		0.212			0.735		0.687		
p_2	0.656				0.328		0.436		
p_3		0.764	0.152				0.527		0.385
⋮									
p_{tgt}		0.581					0.330		
⋮									
p_{N-1}	0.383				0.248			0.176	
p_{N-2}		0.654					0.525		
p_N		0.265					0.430		0.226

Pearson correlation

0.538

0.216

0.475

0.304

0.513

0.487

p_i ($i = 1, 2, \dots, N$) : All papers in dataset

p_{cit_j} ($j = 1, 2, \dots, N$): Papers as citation papers in dataset



Neighborhood of the target paper (e.g., set to 4)

Proposed Method

(1) Leveraging Potential Citation Papers

How are potential citation papers discovered?

	p_{cit_1}	p_{cit_2}	p_{cit_3}	-----	p_{cit_j}	-----	$p_{cit_{N-2}}$	$p_{cit_{N-1}}$	p_{cit_N}
p_1		0.212			0.735		0.687		
p_2	0.656				0.328		0.436		
p_3		0.764	0.152				0.527		0.385
⋮									
p_{tgt}	0.435	0.581	0.536	-----	0.211	-----	0.330	0.472	0.368
⋮									
p_{N-1}	0.383				0.248			0.176	
p_{N-2}		0.654					0.525		
p_N		0.265					0.430		0.226

Pearson correlation

0.538

0.216

0.475

0.304

0.513

0.487

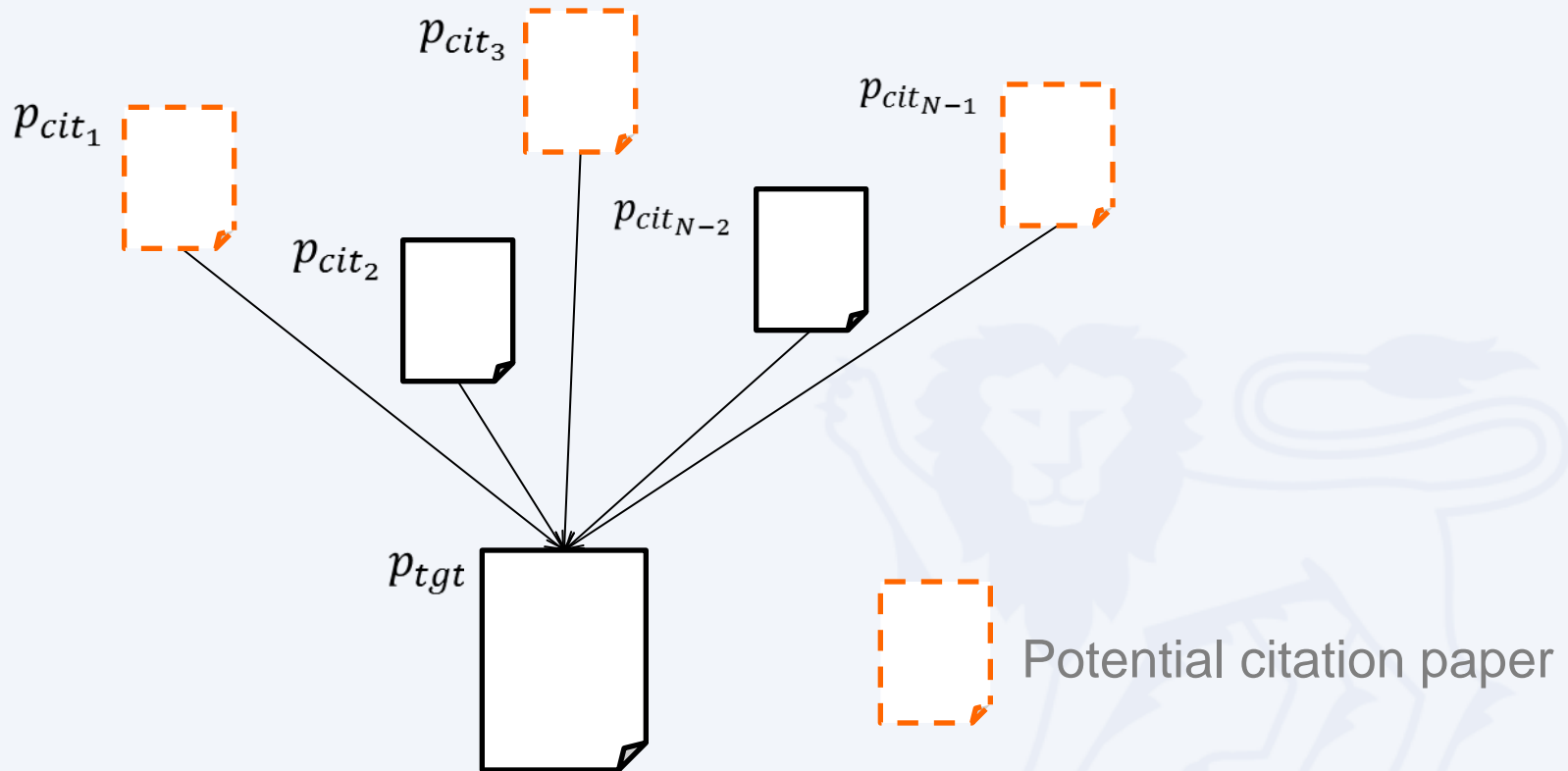
“potential citation papers”
(e.g., set to 3)

p_i ($i = 1, 2, \dots, N$) : All papers in dataset

p_{cit_j} ($j = 1, 2, \dots, N$): Papers as citation papers in dataset

Proposed Method

Identified Potential Citation Papers



Proposed Method

(1) Leveraging Potential Citation Papers

How is the sparsity of matrix is solved?



Proposed Method

(1) Leveraging Potential Citation Papers

How is the sparsity of matrix is solved?

Original matrix

	p_{cit_1}	p_{cit_2}	p_{cit_3}	p_{cit_4}	p_{cit_5}
p_1		0.233			0.628
p_2	0.233		0.147		
p_3		0.147		0.265	
p_4			0.265		
p_5	0.628				

Imputed matrix

	p_{cit_1}	p_{cit_2}	p_{cit_3}	p_{cit_4}	p_{cit_5}
p_1	1.000	0.233	0.723	0.538	0.628
p_2	0.233	1.000	0.147	0.476	0.156
p_3	0.723	0.147	1.000	0.265	0.521
p_4	0.538	0.476	0.265	1.000	0.268
p_5	0.628	0.156	0.521	0.268	1.000

The values in the cell:
Cosine similarity between papers

Imputation



Proposed Method

(1) Leveraging Potential Citation Papers

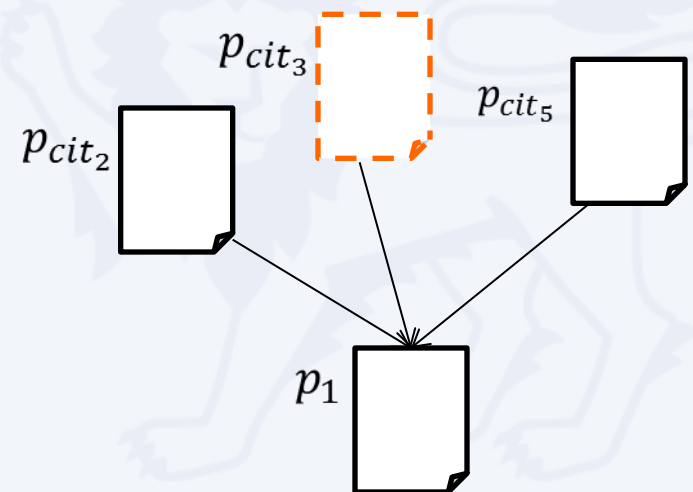
How is the sparsity of matrix is solved?

	p_{cit_1}	p_{cit_2}	p_{cit_3}	p_{cit_4}	p_{cit_5}
p_1	1.000	0.233	?	?	0.628
p_2	0.233	1.000	0.147	0.476	0.156
⋮	⋮	⋮	⋮	⋮	⋮
p_4	0.538	0.476	0.265	1.000	0.268
p_5	0.628	0.156	0.521	0.268	1.000

Target paper (p_1) and corresponding imputed similarities of neighborhood ($p_2, p_4,$ and p_5) from “Imputed matrix”

	p_{cit_1}	p_{cit_2}	p_{cit_3}	p_{cit_4}	p_{cit_5}
p_1	1.000	0.233	0.682	0.453	0.628

○ “potential citation papers” (e.g., set to 1)

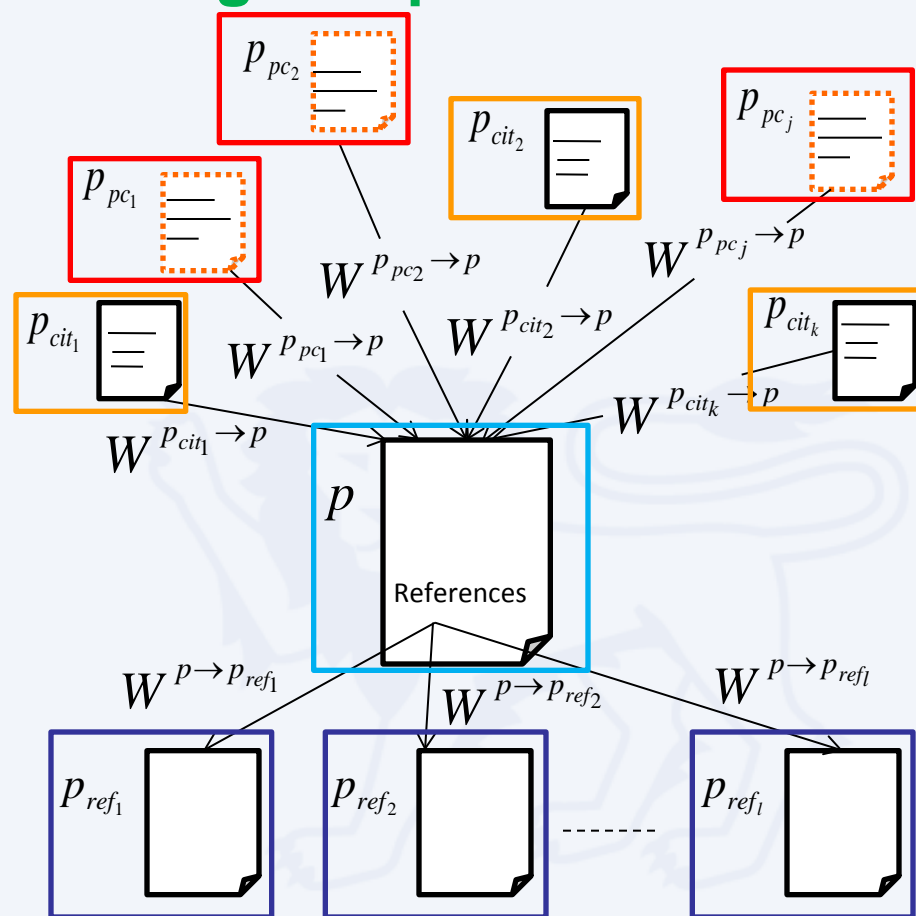


Proposed Method

(1) Leveraging Potential Citation Papers

Feature Vector Construction for Target Papers

$$\begin{aligned}
 \mathbf{F}^P &= \boxed{f^P} + \sum_{x=1}^j W^{P_{pc_x} \rightarrow P} \boxed{f^{P_{pc_x}}} \\
 &+ \sum_{y=1}^k W^{P_{cit_y} \rightarrow P} \boxed{f^{P_{cit_y}}} \\
 &+ \sum_{z=1}^l W^{P \rightarrow P_{ref_z}} \boxed{f^{P_{ref_z}}}
 \end{aligned}$$




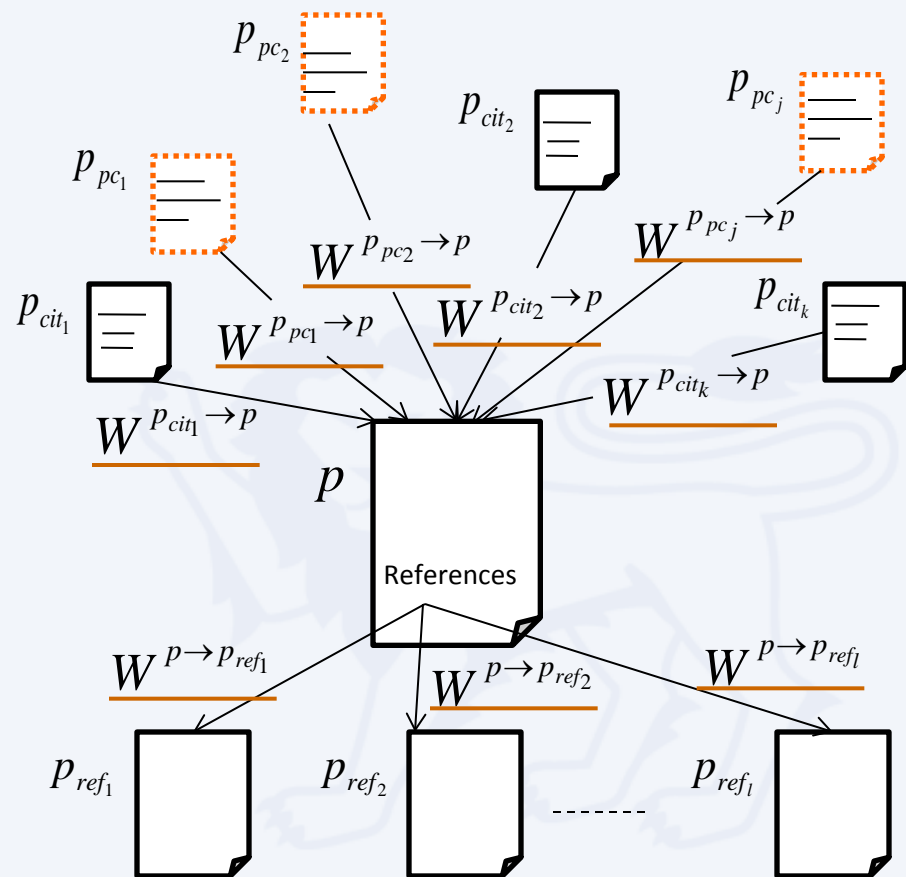
Proposed Method

(1) Leveraging Potential Citation Papers

Feature Vector Construction for Target Papers

$$\begin{aligned}
 \mathbf{F}^P &= f^P + \sum_{x=1}^j \underline{W^{P_{pc_x} \rightarrow P}} f^{P_{pc_x}} \\
 &+ \sum_{y=1}^k \underline{W^{P_{cit_y} \rightarrow P}} f^{P_{cit_y}} \\
 &+ \sum_{z=1}^l \underline{W^{P \rightarrow P_{ref_z}}} f^{P_{ref_z}}
 \end{aligned}$$


 cosine similarity



Proposed Method

(2) Leveraging Fragments in Potential Citation Papers

- [*frg*-SIM]: Fragments with cosine similarity weighting
- [*frg*-TW]: [*frg*-SIM] with tunable weight



Proposed Method

(2) Leveraging Fragments in Potential Citation Papers [frg-SIM]: Fragments with cosine similarity weighting

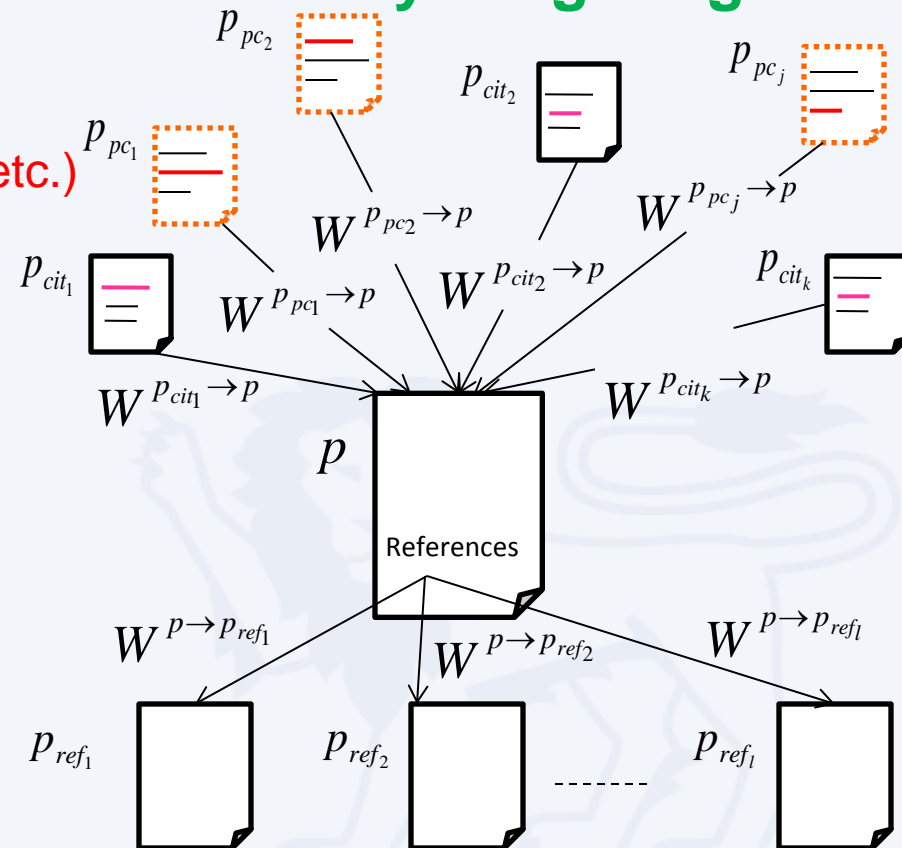
Fragments

("abstract," "introduction," "conclusion," etc.)

$$\mathbf{F}^P = \sum_{x=1}^j W_{(frg)}^{P_{pc_x} \rightarrow P} f_{(frg)}^{P_{pc_x}} + \sum_{y=1}^k W_{(frg)}^{P_{cit_y} \rightarrow P} f_{(frg)}^{P_{cit_y}}$$

$$+ f^P + \sum_{x=1}^j W^{P_{pc_x} \rightarrow P} f^{P_{pc_x}} + \sum_{y=1}^k W^{P_{cit_y} \rightarrow P} f^{P_{cit_y}} + \sum_{z=1}^l W^{P \rightarrow P_{ref_z}} f^{P_{ref_z}}$$

Full text



Proposed Method

(2) Leveraging Fragments in Potential Citation Papers

[*frg*-TW]: [*frg*-SIM] with tunable weight

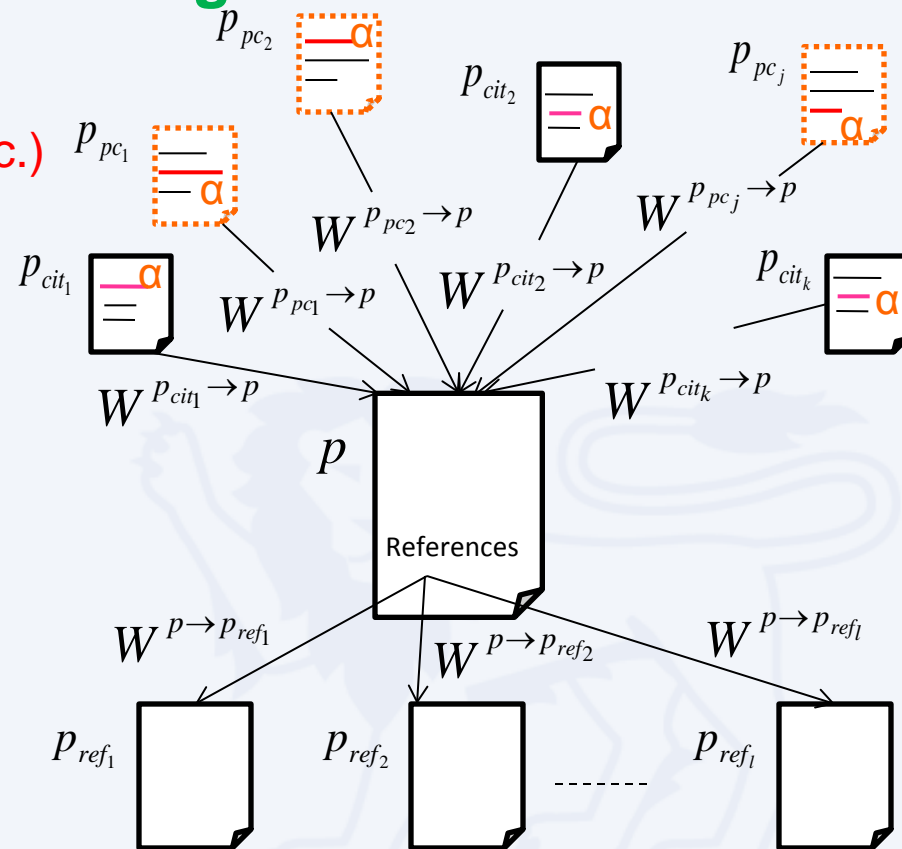
Fragments

("abstract," "introduction," "conclusion," etc.)

$$\mathbf{F}^p = \alpha \left(\sum_{x=1}^j W_{(frg)}^{p_{pc_x} \rightarrow p} f_{(frg)}^{p_{pc_x}} + \sum_{y=1}^k W_{(frg)}^{p_{cit_y} \rightarrow p} f_{(frg)}^{p_{cit_y}} \right)$$

$$+ (1-\alpha) \left(f^p + \sum_{x=1}^j W^{p_{pc_x} \rightarrow p} f^{p_{pc_x}} + \sum_{y=1}^k W^{p_{cit_y} \rightarrow p} f^{p_{cit_y}} + \sum_{z=1}^l W^{p \rightarrow p_{ref_z}} f^{p_{ref_z}} \right)$$

Full text



Experiments

Experimental Data

(to be released soon from

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

(a) Researchers (they have publication lists in DBLP)

	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 citations	15.3 (max. 169)	14.4 (max. 145)
Average number of references	15.8 (max. 47)	14.2 (max. 58)

(b) Candidate papers to recommend (constructed from ACM Digital Library)

	Training set	Test set
Number of papers	50,176	50,175
Average number of citations	19.4 (max. 175)	16.5 (max. 158)
Average number of references	15.7 (max. 45)	15.4 (max. 53)

Experiments

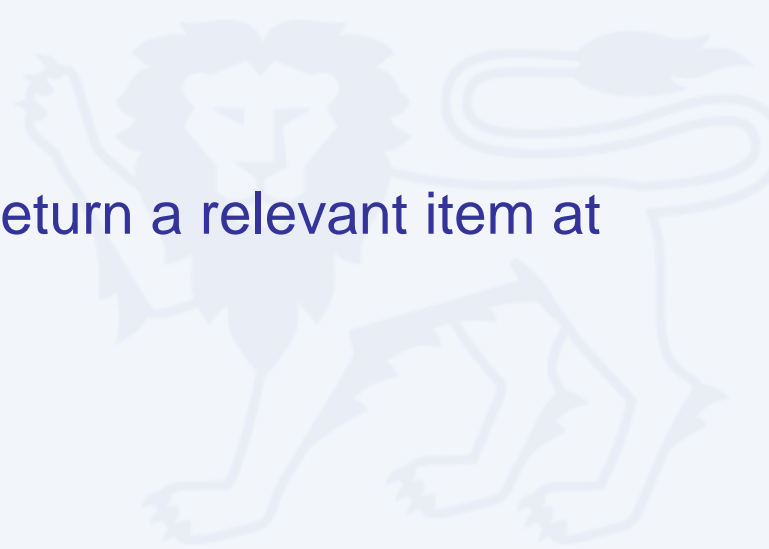
Evaluation Measure

- **NDCG@5, 10 [Järvelin and Kekäläinen, SIGIR'00]**

- Gives more weight to highly ranked items
- Incorporates different relevance levels through different gain values
 - 1: Relevant search results
 - 0: Irrelevant search results

- **MRR [Voorhees, TREC-8, '99]**

- Provides insight in the ability to return a relevant item at the top of the ranking



Experiments

Experimental Results

(1) Leveraging potential citation papers

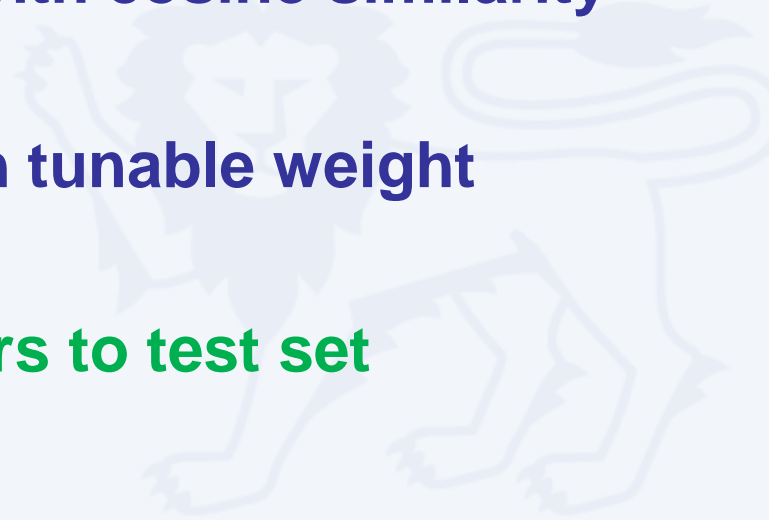
[Tune:*pc*] Parameter tuning to discover potential citation papers

(2) Leveraging fragments in potential citation papers

[Tune:*frg-SIM*] Fragments with cosine similarity weighting

[Tune:*frg-TW*] [*frg-SIM*] with tunable weight

(3) Applying optimized parameters to test set

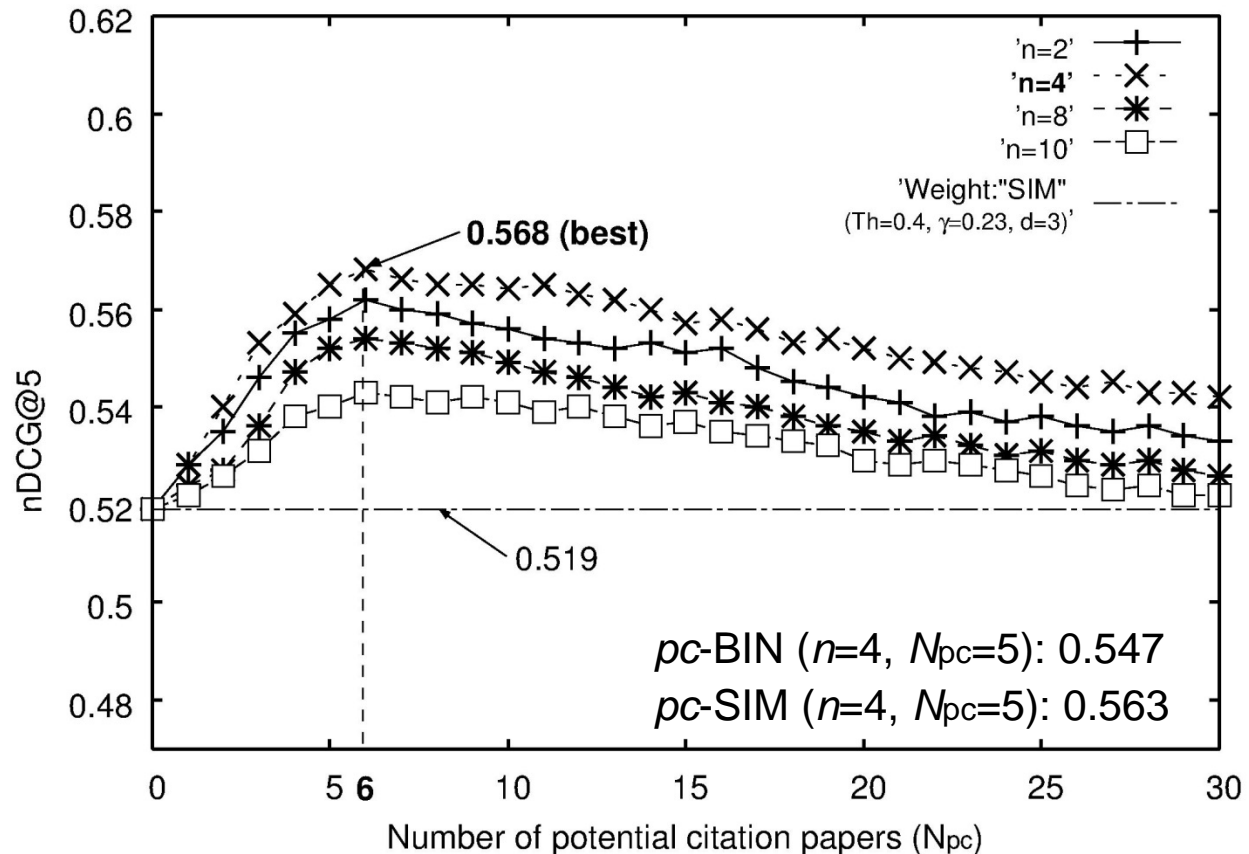


(1) Leveraging Potential Citation Papers

[Tune:pc]

Parameter tuning to discover potential citation papers

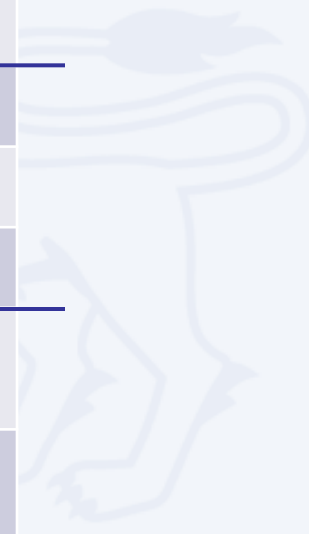
nDCG@5 [pc-IMP]



(2) Leveraging Fragments in Potential Citation Papers [Tune: *frg-SIM*]

Fragments with cosine similarity weighting

<i>pc-IMP</i> ($n=4, N_{pc}=6$)	<i>nDCG@5</i>
Abstract	0.535
Introduction	0.538
Conclusion	0.543
Full text	0.568
Full text + Abstract	0.570
Full text + Introduction	0.569
Full text + Conclusion	0.574
<i>pc-BIN</i> ($n=4, N_{pc}=5$), Full text + Conclusion	0.558
<i>pc-SIM</i> ($n=4, N_{pc}=5$), Full text + Conclusion	0.569

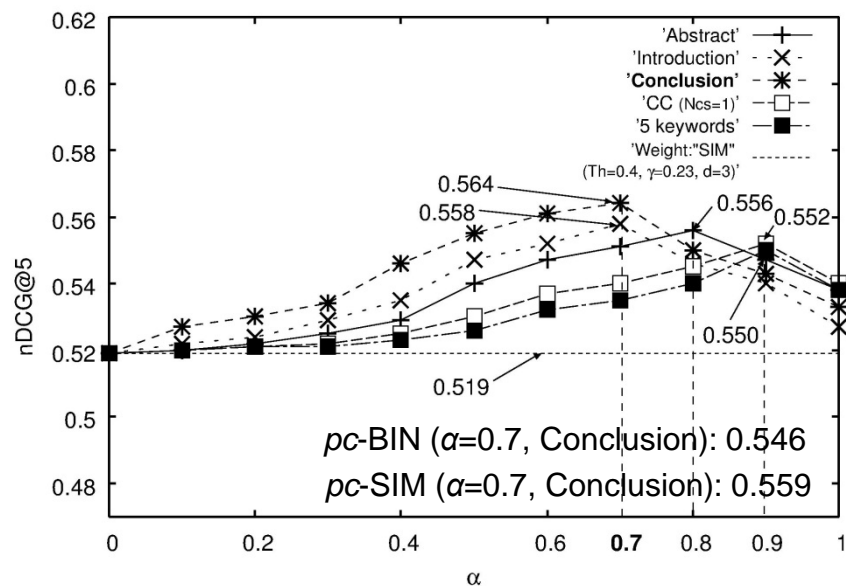


(2) Leveraging Fragments in Potential Citation Papers

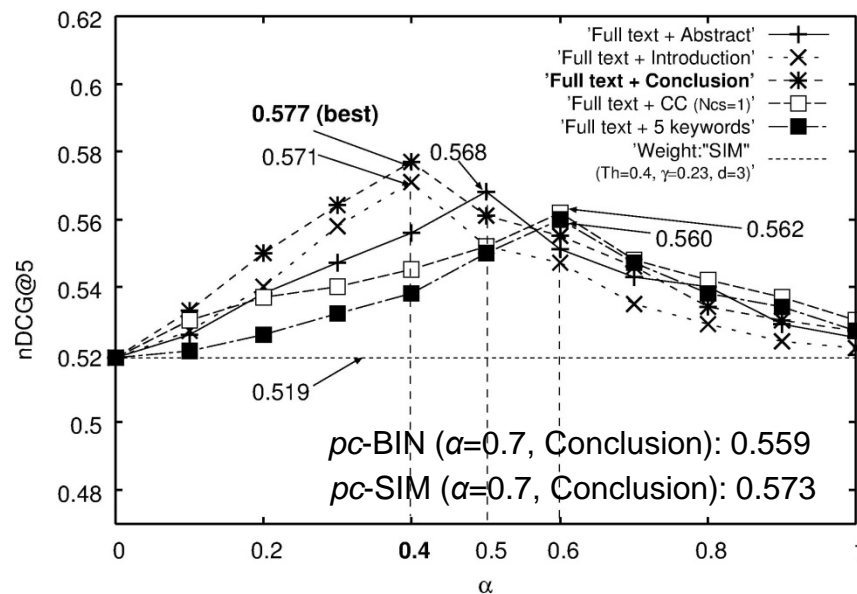
[Tune: *frg-TW*]

[*frg-SIM*] with tunable weight

nDCG@5 [*pc-IMP*]
 (“Only fragments”)

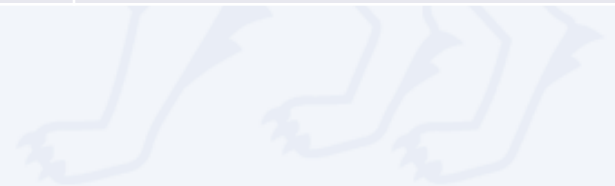


nDCG@5 [*pc-IMP*]
 (“Full text and fragments”)



(3) Applying Optimized Parameters to Test Set

	<i>n</i> DCG@5	MRR
pc-IMP ($n=4, N_{pc}=6$) <i>frg</i> -SIM (Full text + Conclusion) <i>frg</i>-TW ($\alpha=0.4$, Full text + Conclusion)	0.572 0.579	0.787 0.793
Baseline system [Sugiyama and Kan, JCDL'10] (Weight "SIM," Th=0.4, $\gamma=0.23, d=3$)	0.525	0.751
[Nascimento et al., JCDL'11] ("Frequency of bi-gram" obtained from title and abstract)	0.336	0.438
[Wang and Blei., KDD'11] ("In-matrix prediction" in collaborative topic regression)	0.393	0.495



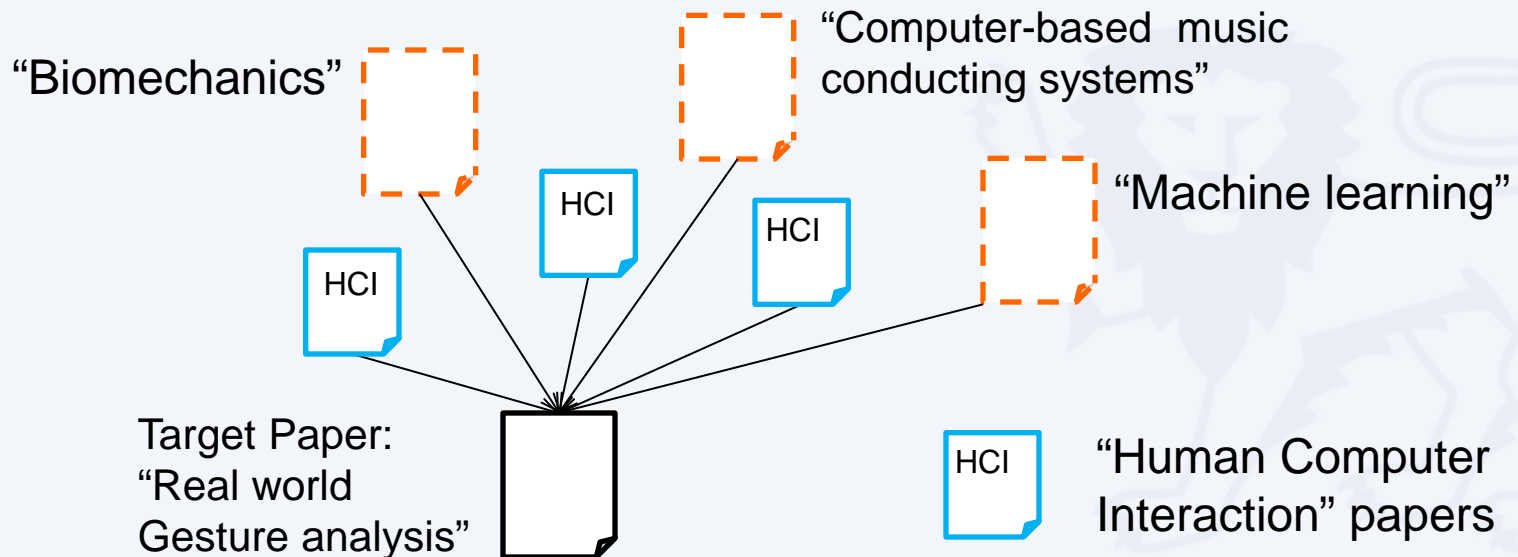
Microscopic Analysis

- **1st Relevant Result in Recommendation List for a “Mobile Computing” Researcher**

[Sugiyama and Kan, JCDL'10]: 52nd

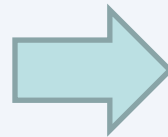
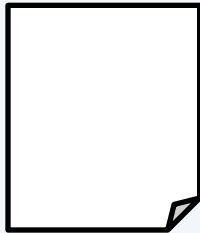
[Sugiyama and Kan, JCDL'13]: 1st

- **Example of Identified Potential Citation Papers**



Limitations

Target
paper



Identified Potential Citation Papers

“Understanding mobile
user’s behavior”

- Mobile technology
- User search behavior
- Clustering

Interdisciplinary paper

“Mobile
Technology”



“Mobile
Technology”



“Mobile
Technology”



Target paper:
“Understanding mobile
user’s behavior”

Conclusion

To recommend scholarly papers much more relevant to a researcher's interests, we have proposed:

- (1) How to identify potential citation papers
- (2) How to leverage fragments in citation papers

to characterize candidate papers to recommend.

(1) → Collaborative filtering with imputed matrix

(2) → “Full text + Conclusion” with tunable weight

Future Work

We plan to develop methods for:

- Selecting balanced neighborhood for interdisciplinary target papers in collaborative filtering
- Developing a novel term weighting scheme suitable for small-sized text fragments in papers

Thank you very much!

