PAPER Special Issue on Data Engineering and Information Management

Interdisciplinary Collaborator Recommendation Based on Research Content Similarity

Masataka ARAKI[†a)], *Nonmember*, **Marie KATSURAI**[†b)], *Member*, **Ikki OHMUKAI**[††c)], *Nonmember*, *and* **Hideaki TAKEDA**[††d)], *Member*

SUMMARY Most existing methods on research collaborator recommendation focus on promoting collaboration within a specific discipline and exploit a network structure derived from co-authorship or co-citation information. To find collaboration opportunities outside researchers' own fields of expertise and beyond their social network, we present an interdisciplinary collaborator recommendation method based on research content similarity. In the proposed method, we calculate textual features that reflect a researcher's interests using a research grant database. To find the most relevant researchers who work in other fields, we compare constructing a pairwise similarity matrix in a feature space and exploiting existing social networks with content-based similarity. We present a case study at the Graduate University for Advanced Studies in Japan in which actual collaborations across departments are used as ground truth. The results indicate that our content-based approach can accurately predict interdisciplinary collaboration compared with the conventional collaboration network-based approaches.

key words: *interdisciplinary research, collaborator recommendation, academic database analysis*

1. Introduction

Research collaboration can have a positive impact on scientific productivity [1, 2]. Interdisciplinary collaboration among researchers with different areas of expertise can potentially produce innovative ideas, solutions, and technologies beyond existing frameworks [3, 4]. However, due to a lack of understanding of other fields and communication opportunities across fields, it is generally difficult for researchers to find new collaborators outside their own areas of expertise. To encourage collaboration, an automatic method for recommending relevant researchers is necessary.

Most state-of-the-art methods for collaborator recommendation use the bibliographic data of papers for finding relevant researchers [5–7]. For example, co-authorship networks are derived from author lists of papers to which network analysis methods are applied to find prospective collaborators for a target researcher [6, 7]. Such an existing network-based approach is called a *friend on friends* approach, which is an application of link prediction in social networks [8]. Although conventional methods have been evaluated using sub-domains within a specific discipline, their effectiveness for recommending collaborators across different fields has not been adequately investigated. In addition, depending on existing collaboration relationships might result in losing a new collaboration opportunity. To bridge the gap between different fields beyond the social networks, we need identify potential collaborators who work in different fields but have similar research interests.

This paper presents an interdisciplinary collaborator recommendation method based on research content similarity. Our method consists of two steps: calculating researchers' feature vectors and computing the relevance between researchers. In the first step, we characterize the interests of each researcher using textual documents in a research database. To cover all research fields, we use research reports in the Database of Grants-in-Aid for Scientific Research, named KAKEN[†], which includes information about research grants provided from the Japan Society for the Promotion of Science (JSPS). From each report, a title, abstract, and a set of research keywords are used to form our method's textual features. Textual features of all reports associated with a researcher are then summarized to a feature vector of that researcher. In the second step, given a target researcher, we find the most relevant researchers from other fields. However, it is challenging to identify a researcher's field without self-assessment; thus, we assume that the department to which each researcher belongs represents his/her research field and consider cross-department recommendation as an alternative. To recommend relevant researchers who work in other departments, we provide two types of strategies: pairwise similarity calculation in a feature space and similarity propagation in an existing collaboration network.

To evaluate the performance of interdisciplinary collaborator recommendation, we focus on collaboration examples across different departments in the Graduate University for Advanced Studies (SOKENDAI) in Japan. SOKENDAI has various departments corresponding to research institutes with active researchers who have obtained funding, and thus it is suitable as a case study. Results of the experiments show that our content-based approach achieved better recommendation results than the conventional methods.

The main contributions of this study are twofold. First, we present a new attempt to analyze interdisciplinary col-

Manuscript received June 29, 2016.

[†]The authors are with Doshisha University, Kyotanabe-shi, 610-0394 Japan.

[††]The authors are with the National Institute of Infomatics, Tokyo, 101-8430 Japan.

a) E-mail: araki@mm.doshisha.ac.jp

b) E-mail: katsurai@mm.doshisha.ac.jp

c) E-mail: i2k@nii.ac.jp

d) E-mail: takeda@nii.ac.jp

DOI: 10.1587/transinf.E100.D.1

[†]https://kaken.nii.ac.jp/

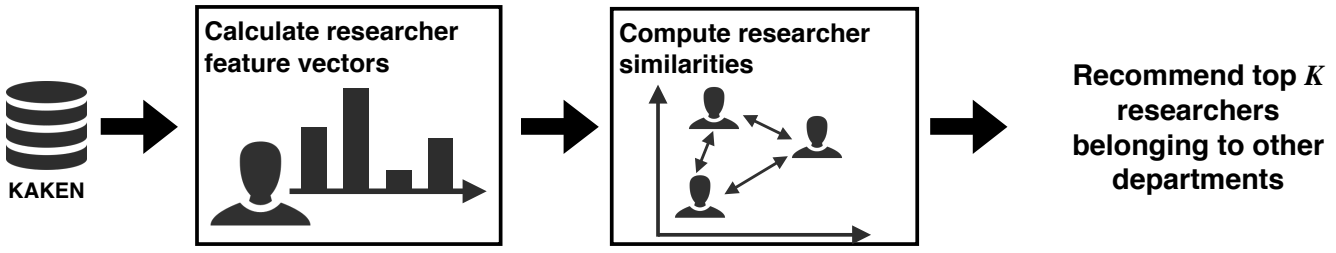


Fig. 1 An overview of interdisciplinary collaborator recommendation using a research database.

laboration at a university and a computational approach to promote it using a research grant database. Second, in the interdisciplinary collaborator recommendation task for SOKENDAI, we present a comprehensive comparison between the content-based approach and the conventional collaboration network-based approach with available evaluation metrics. Results of the experiments provide us new insights about the advantages of the content-based approach for facilitating interdisciplinary research. The remainder of this paper is organized as follows: Section 2 provides a brief review of related studies. Section 3 describes an interdisciplinary collaborator recommendation method using research content similarity. Section 4 presents the analysis of the current collaboration in SOKENDAI and the experiments to verify the effectiveness of the proposed method. Finally, Section 5 provides conclusions and suggests possible directions for future work.

2. Related Work

To facilitate research activities, automatic collaborator recommendation methods have been presented [5–7,9,10]. Gollapallis et al. [9] presented a similar researcher recommendation method based on researcher profiles calculated from academic homepages. Tang et al. [10] constructed a researcher network using a topic model that learns a set of topics from collaboration examples and evaluated its performance within five sub-domains related to computer science. Since the computational cost of the method [10] depends on the number of combinations of research fields, it is difficult to perform when multiple disciplines are targeted. Collaborator recommendation methods have also been developed as an application of social network analysis techniques [5–7]. This approach first derives a researcher network from the bibliographic data of papers and then calculates relevant scores of researchers against a target researcher. To calculate the relevance scores of researchers in the network, Random Walk with Restart (RWR) [11] is often exploited. Specifically, Huynh et al. [5] used several features, including a citation network-based rating of researcher importance. Li et al. [6] constructed a co-authorship network and weighted an edge between two researchers with their co-authorship frequency. Guo et al. [7] combined a co-citation network with a co-authorship network to recommend collaborators. These conventional methods have been evaluated within specific fields, such as computer science-related sub-domains whose papers are included in DBLP [12].

The importance of interdisciplinary collaboration has been emphasized by many institutions. Because universities and research institutes have been divided into departments according to research content, lowering barriers to cross boundaries among the departments is necessary to facilitate interdisciplinary research [13, 14]. Thus, it is reasonable that our work considers relevant researchers who belong to other departments to establish interdisciplinary collaboration. To the best of our knowledge, this paper is the first to focus on collaboration across departments of a university in Japan. To analyze interdisciplinary collaboration patterns, a large-scale research database that covers researchers in all fields is required. Nichols [15] used grant proposals and awards in the National Science Foundation (NSF) database[†] to quantify the interdisciplinarity. NSF is the only federal funding agency that supports all of the basic sciences, and examining it allows for an assessment across all disciplines. In Japan, KAKEN plays a similar role, which includes research reports of projects awarded by MEXT/JSPS[††]. Thus, our study uses KAKEN as an information source to model the researchers in several departments of SOKENDAI and analyze their collaborations.

3. Interdisciplinary Collaborator Recommendation Based on Research Content Similarity

This section presents an interdisciplinary collaborator recommendation method based on research content similarity. Figure 1 presents an overview of the proposed method. As shown, we first calculate a feature vector that reflects the interest of each researcher using textual documents in the research grant database (see Section 3.1). Then, we calculate the similarity between researchers using the features and recommend interdisciplinary collaborators (see Section 3.2).

3.1 Researcher feature vectors and similarity measures

To characterize a researcher's interests, a set of textual documents authored by the researcher should be obtained. It is usually difficult to prepare a common information source for

[†]http://www.nsf.gov/
[††]https://www.jsps.go.jp/english/e-grants/

modeling researchers because research cultures, such as citations and publications are different across disciplines [16]. To solve this problem, we propose to use a research grant database that offers research reports in a common format for all research fields. In KAKEN, each researcher is assigned a unique researcher number and linked to research projects [17]. Each project is also assigned a unique project number and associated with all project members (i.e., researchers), title, research period, research field, and keywords. A principal researcher of each project normally submits an annual research report, and only the latest report becomes available as an abstract of the project. The research keywords are manually identified by the principal researcher, which gives a short description of the project. Note that our method can be applied to an arbitrary database that (i) covers all research fields and (ii) is equipped with author disambiguation.

In our framework, features of researcher expertise can be modeled using various representations, e.g., a word frequency vector and its low-dimensional vector. Below, we present these features and their similarity measures.

Bag-of-Words (BoW) representation. BoW extracts a set of words from all research projects associated with a target researcher. There are various measures to compute the similarity between two sets of words. First, we use a Jaccard similarity to compute the word set-based similarity between two researchers i and j as follows:

$$sim_{jac}(i,j) = \frac{|W_i \cap W_j|}{|W_i \cup W_j|}, \tag{1}$$

where W_i represents a set of unique words for researcher i and $|W|$ is the cardinality of W.

It is well known that applying term frequency-inverse document frequency (TF-IDF) [18] to a set of words is effective for considering the importance of each word in a target corpus. Let V represent a vocabulary of unique words in a dataset, and represent the TF-IDF vector for researcher i by $\boldsymbol{x}_i = [x_i^1, x_i^2, \cdots, x_i^{|V|}]$. Then, the weighted Jaccard similarity [19] between two vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ can be computed as follows:

$$sim_{wjac}(i,j) = \frac{\sum_{k=1}^{|V|} \max(x_i^k, x_j^k)}{\sum_{k=1}^{|V|} \min(x_i^k, x_j^k)}. \tag{2}$$

We compare the above measures with the cosine similarity between two TF-IDF vectors

$$sim_{cos}(i,j) = \frac{\boldsymbol{x}_i \cdot \boldsymbol{x}_j}{\|\boldsymbol{x}_i\| \|\boldsymbol{x}_j\|}. \tag{3}$$

Topic representation. Recent studies exploit probabilistic topic models, such as Latent Dirichlet Allocation (LDA) to represent the researchers' interests [15, 20–22]. The calculation of topic distributions corresponds to reducing the dimensionality of the textual feature space and finding semantic word clusters. Topic representation has been effectively used for collaboration analysis [20], research community detection [21], and author disambiguation [22]. We investigate its effectiveness for the interdisciplinary collaborator recommendation task, in which the same feature extraction approach in [22] is exploited. We apply LDA to a training dataset to produce word distributions for each topic. Using the learned LDA, we represent projects of each researcher by topic distributions and then average them to construct a feature vector of that researcher. The similarity between topic representations of researchers is also computed using cosine similarity as defined in Eq. (3).

In our case study, we investigated the performance of each combination of feature vectors and similarity measures.

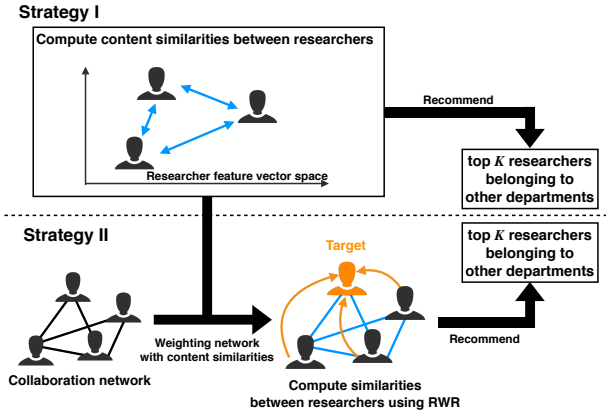


Fig. 2 Two strategies of interdisciplinary collaborator recommendation using a collaboration network.

3.2 Relevant researcher recommendation for interdisciplinary research

Recent methods exploit social networks around researchers to find future collaborators without expertise modeling [6]. In contrast, our assumption is that for promoting interdisciplinary research, different fields should be bridged using content-based potential relevance. In our method, as shown in Fig. 2, we compare the following two recommendation strategies that exploit content-based similarity between researchers.

Strategy I: Calculating pairwise similarities in a feature space. This strategy recommends interdisciplinary collaborators based on research content only. Pairwise similarities between researchers are computed using one of the measures described in Section 3.1. For a target researcher, top K researchers who have high similarities and work in other departments are recommended.

Strategy II: Weighting existing collaboration network with research content similarity. This strategy recommends interdisciplinary collaborators based on both social relationships and research content similarities. Similar to the conventional method [6], a collaboration network is constructed in which each node corresponds to a researcher, and an edge is depicted if two researchers share at least one project. Then, each edge weight is computed as a similarity between the corresponding researchers using one of

Table 1 List of departments in SOKENDAI and their numbers of researchers found in KAKEN (numbers of researchers refer to those who also belong to other departments).

Department	# of researchers
National Museum of Ethnology (Minpaku)	45
International Research Center for Japanese Studies (Nichibunken)	20
National Museum of Japanese History (Rekihaku)	32
The Open University of Japan Center for Open Distance Education (OUJC)	13
National Institute of Japanese Literature (NJIL)	25
Institute for Molecular Science (IMS)	60
National Astronomical Observatory (NAOJ)	96(1)
National Institute for Fusion Science (NIFS)	59
Institute of Space and Astronautical Science (ISAS)	74
Accelerator Laboratory (KEK)	134(1)
Institute of Materials Structure Science (IMSS)	55
Institute of Particle and Nuclear Studies (IPNS)	101
Institute of Statistical Mathematics (ISM)	45(1)
National Institute of Polar Research (NIPR)	50
National Institute of Informatics (NII)	65
National Institute of Genetics (NIG)	60
National Institute for Basic Biology (NIBB)	57(1)
National Institute for Physiological (NIPS)	62(1)
School of Advanced Sciences (SAC)	16(6)
The Center for the Promotion of Integrated Sciences (CPIS)	4(6)
Total	1,073(17)

the measures described in Section 3.1. Given a target researcher, relevant researchers are determined using RWR in the constructed network. Let $\mathbf{S}$ be a row-normalized state transition matrix whose (i, j)-th entry denotes the normalized edge weight between researchers i and j. In RWR, the relevance scores of the researchers are updated at the t-th iteration using the following equation:

$$\boldsymbol{r}^{(t+1)} = \alpha \mathbf{S}\boldsymbol{r}^{(t)} + (1-\alpha)\boldsymbol{q}, \tag{4}$$

where $\boldsymbol{r}^{(t)}$ is a vector whose i-th entry denotes the relevance score of researcher i; $\boldsymbol{q}$ is the initial vector, which sets a score of the target researcher to one and scores of other researchers to zero; $\alpha \in [0, 1]$ is a parameter that balances the contribution of the initial vector and the researcher relationships. By sorting the scores, the top K researchers belonging to other departments are presented to the target researcher as interdisciplinary collaborators.

Comparing Strategies I and II, we can investigate the influence of the existing collaboration network on finding opportunities for interdisciplinary research.

4. Evaluation and Analysis

In this section, we present a case study at SOKENDAI to verify the effectiveness of our content-based approach for interdisciplinary collaborator recommendation. Each faculty member of SOKENDAI belongs to at least one department, and a department usually corresponds to a specific subject or discipline. Table 1 shows a list of all departments and their numbers of researchers that are registered in KAKEN. As of April 2015, 1,227 faculty members belonged to SOKENDAI, among which 1,081 researchers (17 of these belong to more than one department) had unique researhcer numbers in KAKEN. Using each researcher number as a query, we extracted 7,698 research projects from KAKEN. If two researchers share at least one project, then we regarded them as having a collaborative relationship. Consequently, 1,331 projects were collaboration projects by SOKENDAI researchers. The "interdisciplinary collaboration relationship" was established as ground truth only when the two researchers belonged to different departments. The constructed dataset associated with project titles, members, research keywords, and abstracts is denoted as the KAKEN dataset.

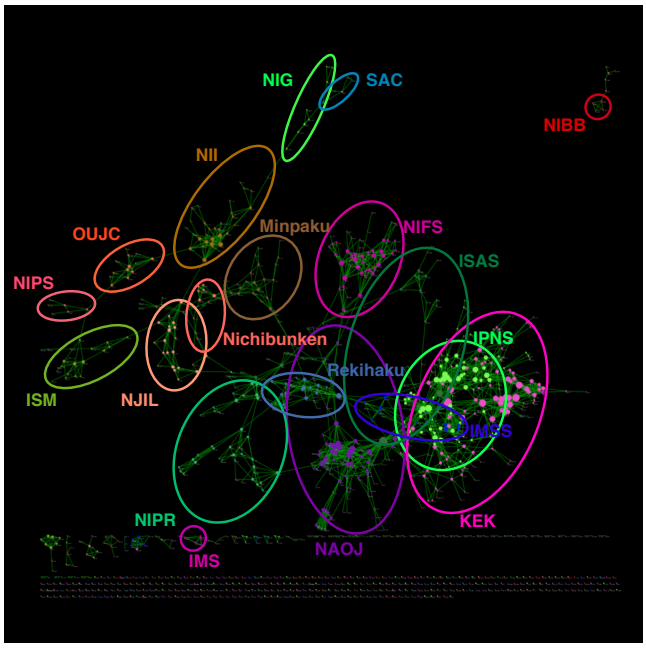


Fig. 3 Visualization of a collaboration network constructed from the KAKEN dataset. The node size is proportional to the number of projects associated with the researcher, and the node color corresponds to the researcher's department. The number of nodes is 1,081, and the number of edges is 2,107.

In Section 4.1, we first analyze the current situation of collaboration across departments at SOKENDAI by visualizing the collaboration relationships. In Section 4.2, we investigate the effectiveness of each feature extraction method in the proposed method. In Section 4.3 we compare performance between the content-based approach and the collaboration network-based approach.

4.1 Current situation of interdisciplinary collaboration at SOKENDAI

First, we analyzed a current situation of collaboration across SOKENDAI departments using the KAKEN dataset. Figure 3 shows a collaboration network depicted using the force-directed graph-drawing algorithm in which a node corresponds to a researcher and an edge represents a collaboration relationship. The node size is proportional to the number of projects associated with the researcher, and the

node color corresponds to the researcher's department. If a researcher belongs to more than one department, then the researcher was associated with the department with fewer members. In this network, 577 nodes out of 1,081 nodes belong to the largest graph, while there are 288 isolated nodes. For easier viewing, we manually drew a circle with an abbreviated department name to indicate a cluster of the department members. As shown, most collaborations were within department. There were only 222 interdisciplinary collaboration projects; this corresponds to about 2.9% of the total number of projects in the KAKEN dataset. This visualization indicates that the connection between the departments is not dense. Link prediction based on the existing collaboration network might suffer from the small number of positive examples.

4.2 Evaluation of feature vectors

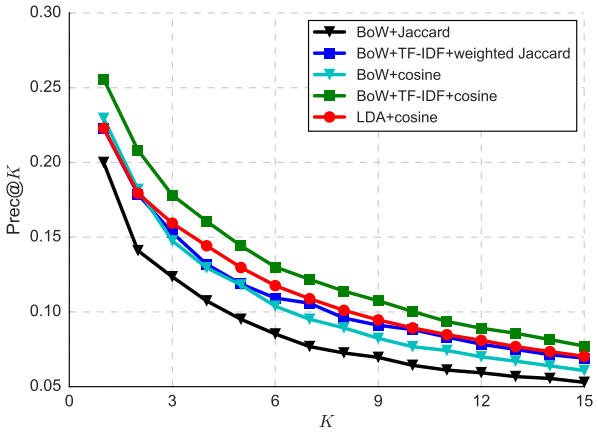
This subsection evaluates the performance of each type of researcher feature vector and similarity measure. In this experiment, we used 222 interdisciplinary collaboration projects as a testing dataset and the remaining 7,476 projects as a training dataset. Before feature extraction, because texts in KAKEN have been registered using OCR from prints of reports, we corrected any OCR-based minor errors using KAKEN-specific normalization[†]. In addition, we used preprocessing techniques, such as NFKC normalization for Japanese UTF-8 texts, converting all uppercase characters to lowercase characters and removing URL strings. Then we applied morphological analysis to the project titles and abstracts, using MeCab[††]; in this analyzer, we introduced a list of all research keywords in the dataset and a list of page names in the Japanese Wikipedia[†††] to consider named entities. Finally, we used only nouns (including named entities) from the title and abstract of each project and its research keywords as word features of the project. For topic representation (i.e., LDA-based feature vectors), we followed the same setting in [22]; the number of topics was empirically set to 500, and LDA hyperparameters α and β were set to 0.1 and 0.01, respectively. The words that appeared less than three times in all texts or appeared in more than 50% of the documents were removed in LDA training.

Each type of researcher feature vector was calculated using the training dataset, and relevant researchers were recommended to each researcher in the testing dataset using Strategy I that exploits pairwise similarities. If the recommended researcher was an actual collaborator, then we regarded it as a true positive. As evaluation metrics, we used an averaged precision of the top K recommended results for each researcher (denoted as Prec@K) and an averaged recall of the top K recommended researchers for each researcher (denoted as Recall@K).

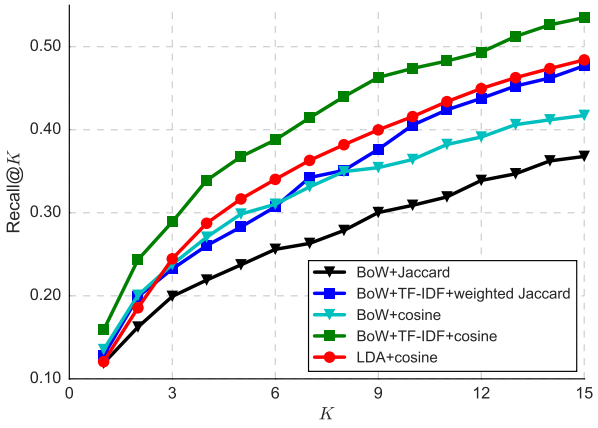
[†]For example, some long vowel symbols were registered as hyphens and only whitespaces were sometimes included as research keywords. These errors were fixed before feature extraction.

[††]http://taku910.github.io/mecab/

[†††]https://ja.wikipedia.org/



(a)



(b)

Fig. 4 Averaged precision and recall of the top K recommendations for each researcher using each type of researcher feature vector: (a) Prec@K and (b) Recall@K.

Figure 4 shows the recommendation results of each combination of researcher feature vector and similarity measure. As shown, the weighted Jaccard similarity gave better results than the Jaccard similarity, which indicates the effectiveness of TF-IDF weighting. TF-IDF vectors with cosine similarity achieved the best Prec@K and Recall@K for every K. Although LDA-based representation has been useful for several applications, such as author disambiguation [22], it could not outperform the TF-IDF features in this setting. We suggest this is because a slight overlap between researchers' interests was missed due to the dimension reduction of word features to topics. Thus, we should be careful about using dimension reduction methods for finding the potential relevance of different disciplines. This observation will be useful for developing a sophisticated recommendation model for interdisciplinary collaboration.

4.3 Performance comparison among content-based and collaboration network-based approaches

Next, we present a performance comparison to investigate whether the existing collaboration network is beneficial for interdisciplinary collaborator recommendation. Similar to conventional studies [6, 10], this experiment was designed to predict future collaborations from historic collaboration examples. We divided the KAKEN dataset into two parts: 4,103 projects before 2004 as a training dataset and 3,595 projects since 2005 as a testing dataset. The numbers of collaboration examples in the training and testing datasets were 119 and 103, respectively. If an interdisciplinary collaborator was accurately recommended to a testing researcher using the training dataset, then we regarded the recommendation as a true positive. Note that the evaluation was performed for only 73 researchers who had at least one interdisciplinary collaborator in the testing dataset and also had a path to the collaborator in the training dataset. We used Prec@K and Recall@K as evaluation metrics and compared the following listed methods.

Adamic/Adar [23]. This method has been utilized for link prediction in co-authorship networks [8]. It computes a similarity between two researchers i and j based on their common collaborators as follows:

$$Adamic/Adar(i,j) = \sum_{z \in \Gamma(i) \bigcap \Gamma(j)} \frac{1}{\log |\Gamma(z)|}, \qquad (5)$$

where $\Gamma(i)$ is a set of researchers adjacent to the researcher i in the collaboration network. This method corresponds to exploiting the local structure of an original collaboration network without edge weights.

ACRec_mod. The collaborator recommendation method presented in [6], called ACRec, weights an edge between two researchers using the bibliographic data of their co-authored publications, such as the co-author order, latest collaboration time point, and the total times of collaboration. Because KAKEN does not include the data of co-author order for a project, we omitted the co-author order from the original ACRec and denote it by ACRec_mod. The relevant scores of researchers are calculated using RWR in the weighted network.

Content. This method uses Strategy I, which recommends interdisciplinary collaborators using pairwise similarities among researchers in a feature space.

Network+Content. This method uses Strategy II, which corresponds to weighting the existing collaboration network with research content similarity. The relevant scores of researchers are calculated using RWR in the weighted network.

Note that in Content and Network+Content, we used TF-IDF vectors and cosine similarity, which achieved the best performance in the previous experiment. Prior to comparison, we had to investigate the influence of parameter α for RWR in ACRec_mod and Network+Content. Figure 5 shows Prec@K and Recall@K ($K = 5, 10$) for these two methods when α changes in the RWR. Although the value of α is generally set to 0.2 in social network analysis, a larger α tends to produce slightly better results, which implies the speciality of this collaboration network. We chose $\alpha = 0.9$ for performance comparison.

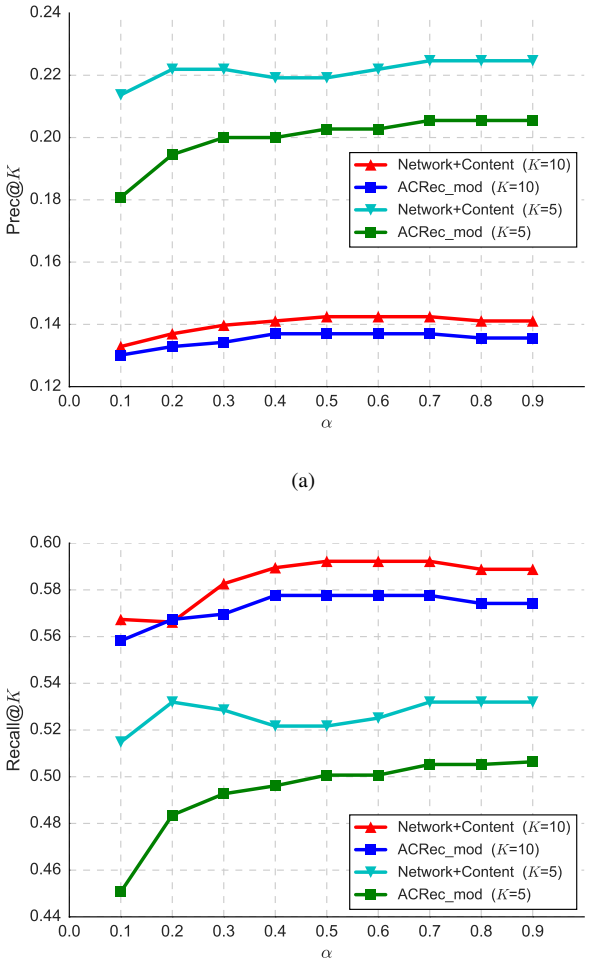
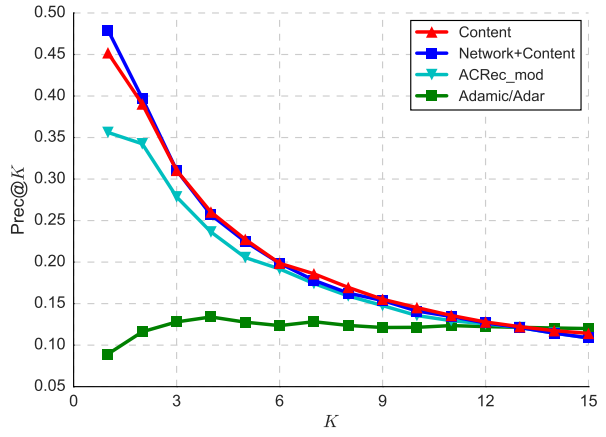
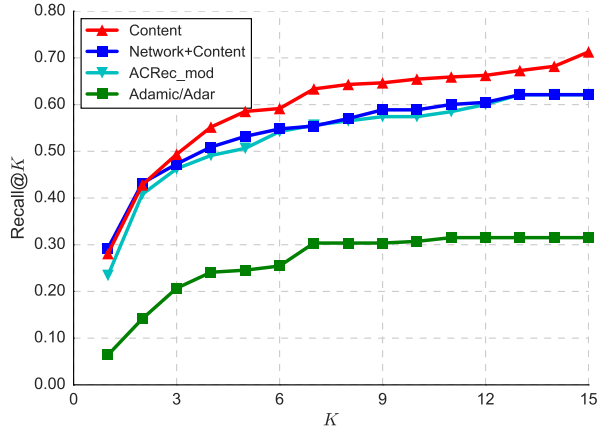


Fig. 5 Averaged precision and recall of the top K ($K = 5, 10$) recommendations for Network+Content and ACRec_mod when changing α in the RWR: (a) Prec@K and (b) Recall@K.

Figure 6 shows Prec@K and Recall@K for all methods. Note that because the ground truth was constructed using the actual collaborations, there is a possibility that the recommended collaboration is fundamentally correct and Prec@K should be higher in such a case. Thus, Recall@K can be more reliable in our experiments. Specifically, Content often produced better Recall@K than Network+Content, which means Strategy I can find relevant researchers beyond the existing network. As shown, our content-based methods, i.e., Content and Network+Content, often achieved better results than collaboration network-based methods, which highlights the importance of researcher expertise modeling. From the



(a)



(b)

Fig. 6 Averaged precision and recall of the top K recommendations for each researcher using each method: (a) Prec@K and (b) Recall@K.

results, we consider that focusing on research content similarity is more beneficial for the interdisciplinary collaborator recommendation than depending on the existing collaboration network. This matches well with the assumption that interdisciplinary collaborators cannot always be found in historical social networks. These findings from our case study will help develop computational approaches to promote interdisciplinary research.

5. Conclusions and Future Work

In this paper, we present an interdisciplinary collaborator recommendation method based on research content similarity. In the proposed method, we first calculated researcher feature vectors using research reports in KAKEN. Then, to find relevant researchers who work in other departments, we presented two types of strategies that exploit content-based similarity: Strategy I based on pairwise similarities and Strategy II based on the collaboration network with research content similarities. A case study at SOKENDAI revealed fewer collaborations across departments compared to collaborations within departments.

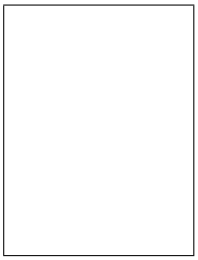
In our content-based approach, TF-IDF vectors produced better results than LDA-based topic representation for interdisciplinary collaborator recommendation. This suggests a difficulty in dimension reduction for finding a slight overlap between different disciplines, which is useful knowledge for developing a sophisticated recommendation model in future work. Comparing Strategy II to ACRec_mod (a modified version of the conventional method [6]), we observed that research content similarities are more effective as edge weights in the collaboration network than the collaboration frequency. Strategy I based on pairwise similarities achieved better recommendation performance than the collaboration network-based approaches. From the results, we conclude that the research content similarity has a significant influence on interdisciplinary collaborator recommendation beyond the existing social relationship.

Our work in this paper is the first to focus on how to bridge individual departments for promoting interdisciplinary research. Further room for investigation and improvement exists in our study; for example, how to characterize the researchers' interests should be studied for better recommendations. Based on the results obtained, we will investigate additional information sources and develop a sophisticated model for effective recommendation. Due to the small number of positive examples, failure recommendations by our method might be collaborations that should actually be established but have not been performed yet. A user study with the cooperation of faculty members of universities is effective for providing the ground truth. According to [24], the interdiscipline of a research project can be evaluated by three metrics, i.e., variety, balance, and disparity. Such an analysis for each project can contribute to characterize a researcher's interdiscipline using historical projects. Our future work also includes recommending interdisciplinary research groups and research themes.

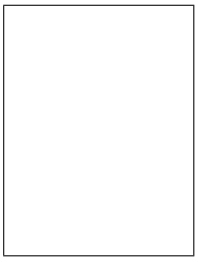
References

[1] G. Abramo, C.A. D'Angelo, and F. Di Costa, "Research collaboration and productivity: Is there correlation?," Higher Education, vol.57, no.2, pp.155–171, 2009.
[2] S. Lee and B. Bozeman, "The impact of research collaboration on scientific productivity," Social Studies of Science, vol.35, no.5, pp.673–702, 2005.
[3] C. Schmickl and A. Kieser, "How much do specialists have to learn from each other when they jointly develop radical product innovations?," Research Policy, vol.37, no.3, pp.473–491, 2008.
[4] F.J. Van Rijnsoever and L.K. Hessels, "Factors associated with disciplinary and interdisciplinary research collaboration," Research Policy, vol.40, no.3, pp.463–472, 2011.
[5] T. Huynh, A. Takasu, T. Masada, and K. Hoang, "Collaborator recommendation for isolated researchers," Proceedings of the 28th International Conference on Advanced Information Networking and Applications Workshops (WAINA), pp.639–644, 2014.
[6] J. Li, F. Xia, W. Wang, Z. Chen, N.Y. Asabere, and H. Jiang, "ACRec: A co-authorship based random walk model for academic collaboration recommendation," Proceedings of the 23rd International Con-

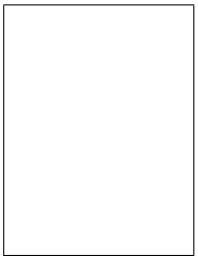
ference on World Wide Web Companion, pp.1209–1214, 2014.
[7] Y. Guo and X. Chen, "Cross-domain scientific collaborations prediction with citation information," IEEE 38th International Computer Software and Applications Conference Workshops (COMPSACW), pp.229–233, 2014.
[8] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," Journal of the American Society for Information Science and Technology, vol.58, no.7, pp.1019–1031, 2007.
[9] S.D. Gollapalli, P. Mitra, and C.L. Giles, "Similar researcher search in academic environments," Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, pp.167–170, 2012.
[10] J. Tang, S. Wu, J. Sun, and H. Su, "Cross-domain collaboration recommendation," Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1285–1293, 2012.
[11] H. Tong, C. Faloutsos, and J.Y. Pan, "Fast random walk with restart and its applications," Proceedings of the 6th International Conference on Data Mining, pp.613–622, 2006.
[12] M. Ley, "DBLP: Some lessons learned," Proceedings of the VLDB Endowment, vol.2, no.2, pp.1493–1500, 2009.
[13] National Academy of Sciences, National Academy of Engineering & Institute of Medicine, Facilitating Interdisciplinary Research, The National Academies Press, 2004.
[14] H. Ledford, "Team science," Nature, vol.525, no.7569, pp.308–311, 2015.
[15] L.G. Nichols, "A topic model approach to measuring interdisciplinarity at the national science foundation," Scientometrics, vol.100, no.3, pp.741–754, 2014.
[16] J.C. Shin and W.K. Cummings, "Multilevel analysis of academic publishing across disciplines: research preference, collaboration, and time on research," Scientometrics, vol.85, no.2, pp.581–594, 2010.
[17] K. Kurakawa, H. Takeda, M. Takaku, A. Aizawa, R. Shiozaki, S. Morimoto, and H. Uchijima, "Researcher name resolver: identifier management system for japanese researchers," International Journal on Digital Libraries, vol.14, no.1, pp.39–58, 2014.
[18] R. Baeza-Yates and B. Ribeiro-Neto, Modern information retrieval, ACM press New York, 1999.
[19] D. Jurafsky and J.H. Martin, Speech and Language Processing: International Version: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Pearson, 2008.
[20] N. Johri, D. Ramage, D.A. McFarland, and D. Jurafsky, "A study of academic collaboration in computational linguistics with latent mixtures of authors," Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, pp.124–132, 2011.
[21] E. Yan, Y. Ding, S. Milojević, and C.R. Sugimoto, "Topics in dynamic research communities: An exploratory study for the field of information retrieval," Journal of Informetrics, vol.6, pp.140–153, 2012.
[22] M. Katsurai, I. Ohmukai, and H. Takeda, "Topic representation of researchers' interests in a large-scale academic database and its application to author disambiguation," IEICE Transactions on Information and Systems, vol.E99-D, no.4, pp.1010–1018, 2016.
[23] L.A. Adamic and E. Adar, "Friends and neighbors on the web," Social Networks, vol.25, no.3, pp.211–230, 2003.
[24] A. Stirling, "A general framework for analysing diversity in science, technology and society," Journal of the Royal Society Interface, vol.4, no.15, pp.707–719, 2007.



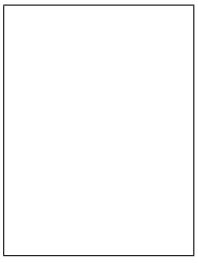
Masataka ARAKI received the B.S. degree in Engineering from Doshisha University in 2016. He is currently a student in the Graduate School of Science and Engineering, Doshisha University. His research interests include academic database analysis.



Marie KATSURAI received the B.S. degree in Engineering, the M.S. degree and the Ph.D. degree in Information Science and Technology from Hokkaido University in Sapporo, Japan in 2010, 2012, and 2014, respectively. She was a Research Fellow of the Japan Society for the Promotion of Science from 2013 to 2015, and she was with the National Institute of Informatics from 2014 to 2015. She is currently an Assistant Professor in the Department of Information Systems Design, Doshisha University. Her research interests include multimedia information retrieval and data mining. She is a member of the IEICE, IEEE, and ACM.



Ikki OHMUKAI received his Ph.D. degree in informatics from the Graduate University for Advanced Studies in 2005. He joined National Institute of Informatics in 2005 and has been an Associate Professor since 2009. His research interests are the semantic web and social media. He is a member of IPSJ and JSAI.



Hideaki TAKEDA is a professor at National Institute of Informatics (NII) Japan, and a professor at the Graduate University for Advanced Studies (SOKENDAI). He received B. Eng., M. Eng. and Dr. Eng. degrees in Precision Machinery Engineering from the University of Tokyo, Japan, in 1986, 1988 and 1991, respectively. He worked at the Norwegian Institute of Technology and the Nara Institute of Technology prior to joining the current institution. He has been the Sumitomo Endowed Professor in the University of Tokyo between 2005 and 2010. His research interests include the semantic Web, knowledge sharing systems and design theory.