

関連研究者と所属位置情報の検索・可視化システム

丹後 綱也[†] 西澤 浩之^{††} 近澤 悠登^{††} 桂井麻里衣[†]

[†] 同志社大学工学部 〒610-0394 京都府京田辺市多々羅都谷 1-3

^{††} 同志社大学大学院理工学研究科 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: †{tango,nishizawa,chikazawa,katsurai}@mm.doshisha.ac.jp

あらまし 研究課題が複雑化するにつれて、高い専門性や広範囲の知識が必要となる。難解な研究課題を円滑に進めるには、複数の研究者が共同で研究に取り組むことが有効といわれ、共同研究の重要性は増している。そのため、専門内容や分野など、研究成果情報から得られる特徴に基づき自動で共同研究者候補を推薦する手法が種々提案されている。一方、推薦手法に基づきユーザが関連研究者を発見できるシステムを実際に構築した例は少ない。本研究では、地理的に近い関連研究者を効率的に発見可能な検索・可視化システムを提案する。提案システムは、研究者や研究機関、キーワードをクエリとして、関連研究者を日本地図上にマッピングする。研究者の詳細画面では、専門内容や過去の科研費採択課題、氏名と所属に基づくウェブ画像を表示する。システムの提示する関連研究者は、既存の共同研究関係から導かれる関連性との重複が少なく、新たな共同研究者候補の発見支援が可能となる。

キーワード 学術情報検索, 共同研究者推薦, 研究支援, ネットワーク可視化

1 はじめに

複雑化した研究課題を円滑に進めるには、複数の研究者が共同で研究に取り組むことが有効といわれる。共同研究と生産性の関係については多数の議論がある。例として、Lee と Bozeman [1] は、共同研究者数がジャーナル論文の出版数に与える影響を調査した。Abramo ら [2] は、共同研究と生産性の関係を実証するために、異分野共同研究や外部との共同研究、産学連携とその研究成果の相関を分析した。Lopes ら [3] は研究成果のランク付け手法を提案し、高いランクの研究者らはより共同で研究する傾向があることを示した。以上の共同研究の重要性から、これまでに共同研究者候補の推薦手法が種々提案されている [4-12]。従来手法の多くは、既存の人間関係の近さや研究分野の関連性など、研究成果情報から得られる特徴に基づき研究者間の類似度を定義する。しかし、これらの推薦手法に基づきユーザが関連研究者を発見できるシステムを実際に構築した例は少ない。加えて、2019年に発表された手法では所属機関の国ラベルを導入することで共同研究者候補の推薦性能が向上したと報告されており [12]、より詳細な地理情報を導入することで国内の共同研究を促進できると考えられる。

そこで本研究では、関連研究者と所属位置情報の検索・可視化システムを提案する。提案システムは、研究者や研究機関、キーワードをクエリとして、それに関連する研究者のピンを日本地図上にマッピングする。研究者間の関連性は、既存の人間関係の近さと、専門内容の近さの二種類から選択可能とする。検索結果から任意のピンをクリックすると、研究者の詳細画面が開かれ、専門内容や過去の科研費採択課題、氏名と所属に基づくウェブ画像を表示する。これにより、ユーザは近隣地域から所望の内容を専門とする研究者を次々と発見することが可能となる。地理的に近い場所に存在する研究者との共同研究や意

見交換を促進するために、各研究者の所属位置情報を地図上にマッピングする。これにより、地方機関のコラボレーション支援につながると考えられる。科学研究費助成事業データベース KAKEN¹の採択課題でシステムを構築した結果、二種類の関連度算出が提示する研究者集合には重複が少なく、様々な観点から共同研究者候補を発見支援できることを示した。

2 関連研究

2.1 共同研究者推薦に関する研究

共同研究の重要性から、共同研究者候補の推薦手法が研究されてきた [5, 7-12]。Chen ら [5] は、共著ネットワークと研究者の専門分野に基づく関連研究者推薦モデルを構築した。Tang ら [7] は、共同研究課題集合においてトピックモデルを学習し、専門内容に基づく推薦性能を向上させた。Li ら [9] は、論文の著者名の順序や最後に共同研究した時期、共同研究の頻度に基づきエッジ重みを算出したネットワークにおける Random Walk with Restart を提案した [13]。Guo と Chen [8] は論文の共引用関係に基づき専門内容の関連性を発見した、文献 [10] では、研究者の論文トピック間の類似度に基づく推薦と、共同研究関係のみに基づく推薦を比較し、研究内容の類似度が重要であると述べた。Kong ら [11] は、研究内容と引用ネットワーク、さらに時間による研究者の研究関心の嗜好の変化を考慮し推薦した。Zhang ら [12] は、Kong ら [11] と同様に研究内容の時間変化を考慮し、同じ国の研究者ほど共同研究に至りやすいことから、地理的な情報として国ラベルを導入した。本論文では、上記の先行研究で論じられてきた多様な特徴（既存の共同研究関係、地理的情報、専門分野）を自由に組み合わせて関連研究者を発見できるシステムを提案する。

1: <https://kaken.nii.ac.jp/>

2.2 学術分野における推薦システムに関する研究

前述のように、共同研究者推薦に関する研究は活発に行われてきたが、推薦結果をユーザが容易に確認できるシステムは報告が少ない。Guerraら [14] は論文トピック抽出に基づき共同研究者検索を支援するシステムを医療分野において構築した。Sousaら [15] は、共著ネットワークを用いた手法 [4, 16, 17] による推薦結果を比較できるシステムを提案した。提案されたシステムはクエリ研究者と類似度尺度を自由に選択して推薦結果を確認できる有用性をもつ。江上ら [18] は、論文に加えて研究プロジェクトの意味的な文書ベクトルを算出することで、通常引用関係をもたないプロジェクトや、最新の会議論文を分析できるシステムを構築した。具体的には、文書データにクラスタリングを適用し、引用関係、特徴語、時系列変化等の各種統計情報による検索等の機能を同一画面で可視化した。本研究においても、関連研究者をもたらす二種類の検索機能を実装し、多角的な視点から共同研究者候補を発見できるシステムを構築する。さらに、所属位置情報を地図上にマッピングした新たな可視化を提案する。

3 提案システム

本章では、共同研究関係と所属位置情報に基づく研究者検索システムを提案する。まず3.1で提案システムの概要を説明し、3.2で検索用データセット構築方法を説明する。3.3では共同研究関係ネットワークの構築と可視化、3.4では研究者間の関連度算出と可視化をそれぞれ提案する。

3.1 概要

提案システムのアーキテクチャを図1に示す。システムで表示する研究者間の関連性は、ローカルの計算機で定期的的分析する。得られた結果はクラウド上のデータベースに格納する。サーバはクライアントからのAPIリクエストに従って、データベースに格納された大きなネットワークから小さなサブネットワークを切り出し、APIレスポンスを返す。サーバからのレスポンスをクライアントが処理することで、ユーザがインタラクティブに結果を閲覧できるようになる。

クライアント側のホーム画面を図2に示す。この画面で右下の検索ボタンをクリックすると、図3のように検索クエリを与えることができる。画面内に円で表示された研究者ピンをクリックすると、図4の詳細画面に遷移する。

提案システムは地図上に研究者ピンをマッピングするが、地図は図3から図4までの航空写真に限らず、図5の市街地図へも変更可能である。これにより、関連研究者の所属の最寄り駅のようなローカル情報も瞬時に確認できる。

提案システムには、専門内容の類似した研究者を表示するモードと、研究者間の実際の共同研究関係を表示するモードの2種類を実装する。これらのモードは画面下部のスイッチを通じてユーザが自由に切り替え可能である。

3.2 データセットの整備と所属位置情報の取得

本研究では、様々な分野の研究者を検索可能なシステムの構築

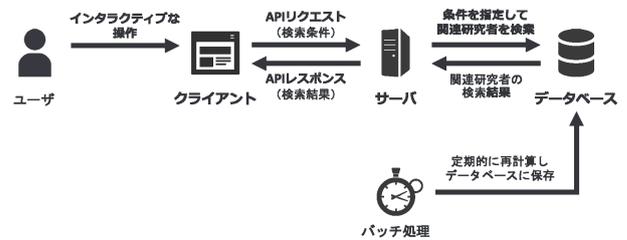


図1 提案システムのアーキテクチャ。

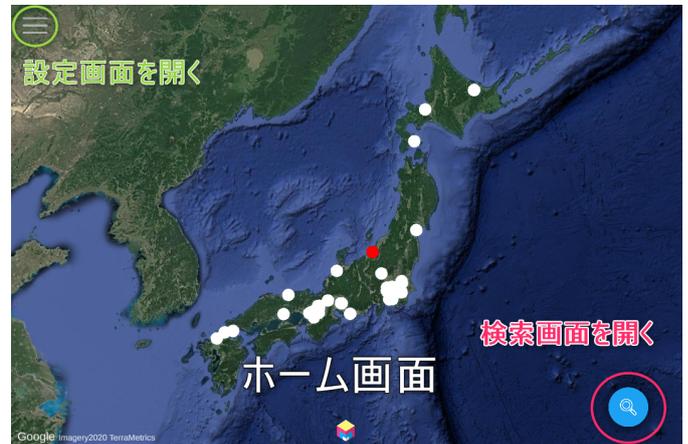


図2 提案システムのホーム画面。

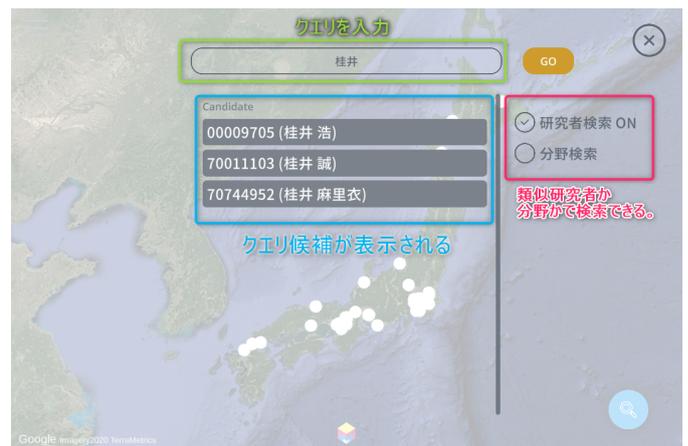


図3 提案システムの検索インターフェース。

を目的とし、科研費データベース KAKEN を用いる。KAKEN は、文部科学省および日本学術振興会が全ての学問領域にわたって幅広く交付する科学研究費助成事業に採択された研究課題の情報を、国立情報学研究所の協力によって公開しているデータベースである。そのため、他の論文データベースに比べ分野ごとの偏りが少なく、かつ全てのデータが同一形式で得られる利点がある。

KAKEN 上の研究者には、研究者番号が一意に割り当てられており、各研究者番号には過去に携わった研究課題が正確に紐付けられている。研究課題には一意な研究課題番号が付与されているとともに、研究課題名、研究代表者、研究分担者、研究機関、研究分野、研究キーワード、要約、課題に参画した研究者らによる成果情報が登録されている。



図 4 研究者の詳細画面.

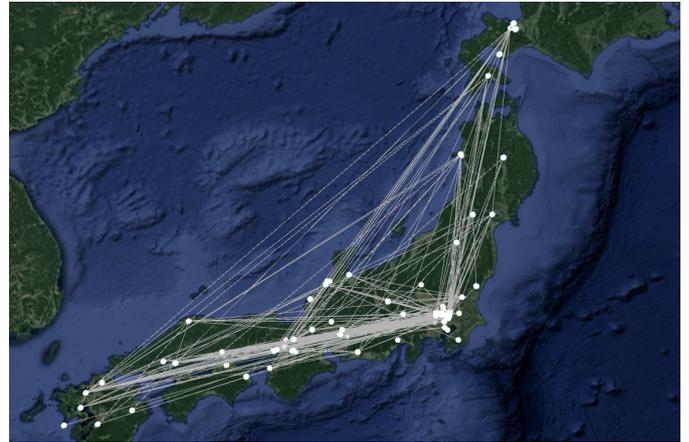


図 6 クエリ研究者に関する共同研究ネットワークの可視化.



図 5 市街地図に変更した場合の可視化.

表 1 所属情報の修正の一例.

正規化前	正規化後
埼玉大学	埼玉大学
愛教大学	愛知教育大学
明学大	明治学院大学
東瀧大学	新潟大学
アジ研	ジェトロ・アジア経済研究所
愛知県立女子	愛知県立大学
JSPS	日本学術振興会
北海総医療大学	北海道医療大学
同志社大学	同志社大学
北海道大学	北海道大学

本研究では、911,724 件の研究課題と、研究者番号をもつ 259,509 名の研究者からなるデータセットを構築した。また、提案システムでは研究者ごとに最新の所属情報を表示することを目的に、開始年が最も直近となる研究課題の所属メタデータを用いる。

所属位置情報（緯度・経度）は、Google Maps API²を用いて取得する。KAKEN の所属情報には表記揺れが存在したため、著者が手動で修正した。修正の例を表 1) に示す。

さらに、提案システムでは研究者詳細画面において各人の顔

画像を表示するため、各研究者の氏名と所属文字列をクエリとし、Google 画像検索結果を収集した。具体的には、顔画像のみを収集するよう絞り込み検索し、検索結果一番目の画像を採用した。

3.3 共同研究ネットワークの構築と可視化

本節では、提案システムにおける共同研究ネットワークの構築と可視化方法について述べる。まず、研究者をノードとし、研究者間の科研費採択課題分担関係をエッジで表すネットワークを構築する。研究者を検索クエリとして入力した場合は、その研究者の共同研究関係を地図上に表示する。具体的には、検索画面における入力文字列から部分一致する研究者リストを表示し、ユーザの選択した研究者の ID に紐付いたサブネットワークをサーバから取得する。図 6 のように取得結果を描画することで、クエリ研究者の人間関係が国内でどのように広がっているかを地図上で確認できる。分野や研究機関をクエリにした場合は、それぞれに該当する研究者らの共同研究関係を表示する。この機能により、任意の分野に強い研究機関を地図上で発見することが可能となる。

3.4 研究者間の関連度算出と可視化

まず、研究者間の人間関係の近さを算出する方法を述べる。研究者 r_1 , r_2 の共同研究者集合をそれぞれ R_{r_1} , R_{r_2} としたとき、二人の人間関係の近さを次式の Jaccard 係数で表す。

$$\frac{|R_{r_1} \cap R_{r_2}|}{|R_{r_1} \cup R_{r_2}|} \quad (1)$$

共同研究者の重複を全くもたない研究者間の関連度は 0 となる。

次に、専門内容に基づく研究者間の関連度算出と可視化方法を述べる。各研究課題はこれまでの研究活動の成果に基づいて提案されており、研究者の専門内容を反映したテキストデータとみなせる。そこで、各研究課題のタイトル、成果登録された学術論文のタイトル、研究キーワードおよび要約から、図 7 に示すように TF-IDF ベクトルを算出する。具体的な手順を次に示す。

文書のベクトル化には、文献 [18] のように単語埋め込みモデルを用いることも考えられるが、提案システムは単語ベース

2 : <https://cloud.google.com/maps-platform/>

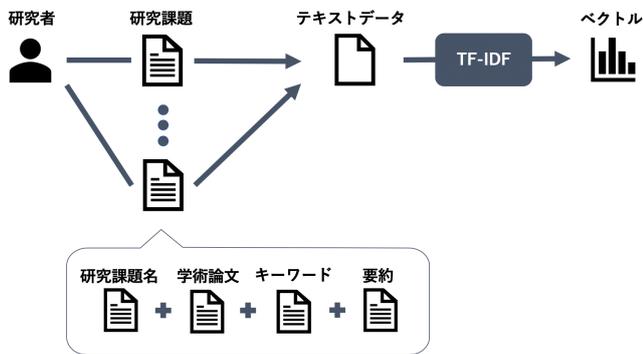


図7 研究内容を表すベクトル生成の概要図。

のマッチングにより研究興味の一貫性を明確に捉えるねらいである。

最後に、研究者の TF-IDF ベクトル間の類似度を用いて、クエリ研究者の関連研究者を探索する。10 万を超えるデータの類似度を総当たりで出すことは計算量が大きく、現実的ではない。そこで本手法では、Faiss [19] を用いて高速に最近傍探索する。Faiss は直積量子化 [20, 21] によりベクトルを次元圧縮して保持し近似的に最近傍探索する。Faiss では検索速度を向上させるために、事前に検索対象のベクトル集合をボロノイ分割し、検索範囲とするボロノイ領域の数を指定する。本研究ではボロノイ領域分割数を 100 とし、検索範囲数を 10 に設定した。以上により各研究者に対し 500 名の共同研究者候補を算出し、データベースに保存する。

4 評価実験

3.4 で提案した二種類の関連度算出がそれぞれユーザに新規の関連研究者をもたらすことを示すために、研究内容に基づく検索結果に含まれる新規の関連研究者数を調査する。具体的には、次式で表される検索結果の重複率 *overlap* を算出する。

$$overlap(sim1, sim2) = \frac{|S_{sim1} \cap S_{sim2}|}{|S_{sim1}|} \quad (2)$$

ここで、 S_{sim} は関連度 sim に基づき提示した研究者集合である。 $overlap(sim1, sim2)$ が低いほど、関連度 $sim2$ に基づき提示した研究者集合に含まれない新規研究者が類似度 $sim1$ によって発見できることを意味する。本実験では全研究者集合からランダムに 100 名の研究者を選択し、それらの全てのペアについて重複率を算出した。その結果、 $overlap$ (研究内容に基づく類似度, 共同研究関係に基づく類似度) の平均値は 0.199 となった。値が小さいことから、構築システムは既存の人間関係からは知り得なかった新規の関連研究者を提示可能だと示された。

5 おわりに

本研究では、関連研究者と所属位置情報の検索・可視化システムを提案した。提案システムは研究者や研究機関、キーワー

ドをクエリに許容し、ユーザは国内研究者を地図上で自由に探索できる。具体的には、研究者間の共同研究ネットワークを可視化するモードと、関連研究者を提示するモードを実装した。後者のモードでは、人間関係の近さに基づく提示と、研究内容の関連度に基づく提示の二種類から選択可能とした。提案システムにより、地理的に近い研究者との共同研究や意見交換の促進が期待できる。今後は、システムの有用性を検証するための定性的評価を行うとともに、より多くの関連度算出方法を検討する予定である。

謝辞 本研究の一部は、JSPS 科研費 (課題番号:17K12794)、JST ACT-I (課題番号:JPMJPR18UC)、JST ACT-X (課題番号:JPMJAX1909) の助成によって行われた。

文 献

- [1] S. Lee, B. Bozeman “The Impact of Research Collaboration on Scientific Productivity” *Social Studies of Science* **35**, 673–702 (2005).
- [2] G. Abramo, C. A. D’ Angelo, F. Di Costa “Research collaboration and productivity: is there correlation?” *Higher Education* **57**, 155–171 (2009).
- [3] G. Lopes, M. Moro, R. Silva, E. Barbosa, J. Palazzo Moreira de Oliveira “Ranking Strategy for Graduate Programs Evaluation” (2011).
- [4] G. R. Lopes, M. M. Moro, L. K. Wives, J. P. M. de Oliveira “Collaboration Recommendation on Academic Social Networks” in *Advances in Conceptual Modeling – Applications and Challenges* pages 190–199 Springer Berlin Heidelberg (2010).
- [5] H.-H. Chen, L. Gou, X. Zhang, C. L. Giles “CollabSeer: A Search Engine for Collaboration Discovery” in *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries* pages 231–240 ACM (2011).
- [6] D. H. Lee, P. Brusilovsky, T. Schleyer “Recommending Collaborators using Social Features and MeSH Terms” *Proceedings of the Association for Information Science and Technology* **48**, 1–10 (2011).
- [7] J. Tang, S. Wu, J. Sun, H. Su “Cross-domain Collaboration Recommendation” in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* pages 1285–1293 ACM (2012).
- [8] Y. Guo, X. Chen “Cross-domain Scientific Collaborations Prediction with Citation Information” in *Computer Software and Applications Conference Workshops (COMP-SACW), 2014 IEEE 38th International* pages 229–233 IEEE (2014).
- [9] J. Li, F. Xia, W. Wang, Z. Chen, N. Y. Asabere, H. Jiang “ACRec: A Co-authorship based Random Walk Model for Academic Collaboration Recommendation” in *Proceedings of the 23rd International Conference on World Wide Web* pages 1209–1214 ACM (2014).
- [10] M. Araki, M. Katsurai, I. Ohmukai, H. Takeda “Interdisciplinary Collaborator Recommendation Based on Research Content Similarity” *IEICE Transactions on Information and Systems* **100**, 785–792 (2017).
- [11] X. Kong, H. Jiang, W. Wang, T. M. Bekele, Z. Xu, M. Wang “Exploring dynamic research interest and academic influence for scientific collaborator recommendation” *Scientometrics* **113**, 369–385 (2017).
- [12] Q. Zhang, R. Mao, R. Li “Spatial-temporal restricted supervised learning for collaboration recommendation” *Scien-*

tometrics **119**, 1497–1517 (2019).

- [13] H. Tong, C. Faloutsos, J.-Y. Pan “Fast Random Walk with Restart and Its Applications” in *Proceedings of the Sixth International Conference on Data Mining* pages 613–622 IEEE (2006).
- [14] J. Guerra, W. Quan, K. Li, L. Ahumada, F. Winston, B. Desai “SCOSY: A Biomedical Collaboration Recommendation System” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* pages 3987–3990 IEEE (2018).
- [15] G. A. de Sousa, M. A. Diniz, M. A. Brandão, M. M. Moro “CNARE: Co-authorship Networks Analysis and Recommendations” in *Proceedings of the 9th ACM Conference on Recommender Systems* pages 329–330 ACM (2015).
- [16] M. A. Brandão, M. Moro, G. Lopes, J. Palazzo Moreira de Oliveira “Using Link Semantics to Recommend Collaborations in Academic Social Networks” (2013).
- [17] F. Xia, Z. Chen, W. Wang, J. Li, L. T. Yang “MVCWalker: Random Walk-Based Most Valuable Collaborators Recommendation Exploiting Academic Factors” *IEEE Transactions on Emerging Topics in Computing* **2**, 364–375 (2014).
- [18] 江上周作, 渡邊勝太郎, 川村隆浩 “Mapping science - ナレッジグラフに基づく科学技術マップの高度検索と対話的操作の実現” *人工知能学会研究会資料* **47**, 1–10 (2019).
- [19] J. Johnson, M. Douze, H. Jégou “Billion-scale similarity search with GPUs” *arXiv preprint arXiv:1702.08734* (2017).
- [20] H. Jegou, M. Douze, C. Schmid “Product quantization for nearest neighbor search” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **33**, 117–128 (2010).
- [21] T. Ge, K. He, Q. Ke, J. Sun “Optimized Product Quantization for Approximate Nearest Neighbor Search” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pages 2946–2953 (2013).