

研究終了報告書

「異なる学術領域の共通問題を発見する時系列ナレッジグラフ基盤の創出」

研究期間：2019年10月～2022年3月

研究者：桂井 麻里衣

1. 研究のねらい

複数の学術領域による連携と融合は、既存の分野の枠組みをこえた発想や技術を生み出すといわれており、協同促進に向けて国内外で様々な政策的努力がなされている。しかし、研究分野は細分化し続け、分野内に閉じた形で急速な発展を遂げており、他分野の知識体系の正確な把握は困難である。

馴染みのない研究分野を渡り歩くには、各分野の知識構造を表す二次元マップの作成が重要といわれている。これまで、単語、論文、研究者などをノードとし、それらに関係がある場合にエッジを描画するネットワーク表現（以降、ナレッジグラフ）の有効性が示されてきた。対象とする分野の情報源を時間区間ごとに分割すると、ナレッジグラフの時系列が得られ、知識の潮流や発展の分析に役立つ。一方、学際・融合研究を推進するには、互いに関連する分野を特定し、分野間で概念や理論の共通基盤を見出す必要があるといわれている。これに着想を得ると、ナレッジグラフから分野間の共通点を発見することで、単一分野では発想困難であった革新的テーマの着手支援が可能になると考えられる。

現実の研究活動には「問題解決のための理論の転用」や「提案技術の新たなアプリケーション」など、より複雑な関係が存在する。ゆえに、学際的・融合的なテーマ発見へつなげるためには、(1)グラフ構造で表現される共通コミュニティを発見し、(2)理論・技術(シーズ)と適用先の課題(ニーズ)のマッチングを考慮する必要がある。そこで本研究課題では、この二点を包括した、学術領域間の共通問題を発見する時系列ナレッジグラフ基盤の創出を目指す。

2. 研究成果

(1) 概要

本研究による成果は、「理論・技術と適用先課題の関係を表すナレッジグラフの構築」と、「異分野の知識を効率的に獲得するための推薦システムの開発」に大別される。

はじめに、ウェブ上で取得可能な論文情報(タイトル・アブストラクト)から理論・技術および適用先課題の情報を抽出するための言語処理ツールを構築した。具体的には、文中でテクニカルタームと共起するフレーズ集を精査し、抽出対象となる語を発見するために効果的と考えられる手がかり語(前置詞など)のリストと、略語抽出のためのルールを定義した。これにより各分野の用語辞書を自動構築するとともに、タイトル・アブストラクト内で2つのクラスに該当する語をハイライトする手法を提案した。辞書内容と語抽出ルールが共に明確なため、書誌情報分析においてツールのブラックボックス性が排除される。データベース分野と図書館情報学分野の関連性分析に応用した結果と、デジタルヒューマニティーズ分野のトレンド解析に応用した結果は、それぞれジャーナル論文として公表した。

次に、本研究で目指す異分野融合支援を具現化すべく、推薦システムのインタフェース設計に着手した。まず、異分野の関連論文誌検索システム Relevance Venue Finder を提案した。ユーザがクエリ論文誌名を入力すると、その論文誌と関連性の高い分野から別の論文誌をシステムが推薦する。そこから任意の論文誌名を選択すると、クエリおよび選択論文誌に対し、年ごとの研究トレンド情報とそれらの共通点を可視化する。以上のシステムの内容は国際会議 JCDL2021 で発表した。

さらに、特定の研究目的に対して多様な解決策を提供する論文を推薦するシステム SolutionTailor を構築した。具体的には、ニューラル言語モデルを用いて、アブストラクト文から背景/目的、手法の説明に対応する文を抽出し、それらの文埋め込みベクトルを用いて論文間の類似度指標を定義した。ユーザは対象とする研究分野を指定したあと、自身の興味に相当する研究要旨を入力するだけで、その研究と同じ課題で異なるアプローチを提供する論文を指定分野から検索できる。これらの成果は国際会議 ECIR2022 で発表予定である。

以上のように ACT-X 研究期間を通じて、異分野間の共通問題を発見するためのナレッジグラフ構築と、推薦システムへの応用という二点を達成できた。

(2) 詳細

研究テーマ A「理論・技術と適用先課題の自動抽出に基づくナレッジグラフの構築」

各論文が用いている理論・技術名(シーズ)と適用先ドメイン名(ニーズ)をタイトルとアブストラクトから自動抽出する手法を構築した。一般に論文タイトルは提案内容を簡潔に表すよう考案される。そこで、タイトル内で理論・技術名または適用先ドメイン名を発見するのに適した手がかり語を考案した(例:「Using A」ならば A が理論・技術名, 「For B」ならば B が適用先ドメイン名)。手がかり語に基づき、対象分野の論文タイトル集合から 2 クラス(理論・技術または適用先)それぞれの辞書を構築した。タイトルに両クラスの語が存在しない場合、論文アブストラクトが第二の情報源となりえる。そこで論文アブストラクトの文集合のうち背景/目的、手法の説明に対応する文をニューラル言語モデルで選択し、さらに手がかり語と辞書内の語との組合せを発見することで、各論文の理論・技術または適用先の情報を抽出した。

2019 年度には「データマイニング分野」と「図書館情報学分野」のみを対象としたケーススタディを行い、提案手法で抽出した用語に基づき各分野の流行トピックを分析した。これらの成果は学術論文誌 Journal of Library and Information Studies に採録された(主な研究成果リスト[2])。この分析経験を生かして 2020 年度からはデジタルヒューマニティーズ分野のトレンド分析を行い、学術論文誌 Journal of Documentation に採録された。

論文の利用技術と適用先がわかると、それらの関係性をグラフで表すことで、論文間の関係や分野間の関係を発見できる可能性がある。そこで 2021 年度は論文や用語をノードにもつ異種情報グラフの開発を進めた。図 1 は、論文の理論・技術名および適用先の関係を表すグラフである。このグラフで DTPD というメタパスを辿る場合、異なる分野の適用先課題(左右二種類の D)において共通の技術(T)が用いられていることを発見できる。これらの成果は 2022 年 1 月以降に取りまとめて発表する予定である。

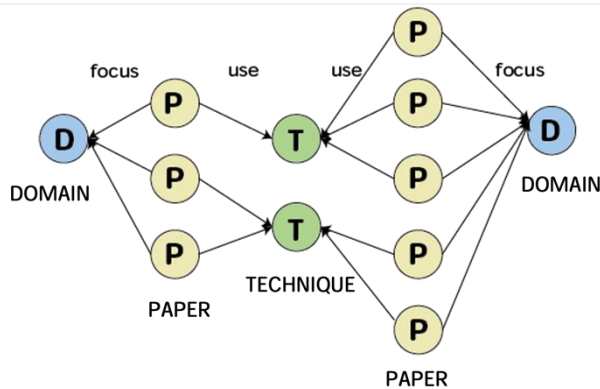


図 1:本研究で新たに構築したナレッジグラフ。

力関係をグラフ化する手法を提案し、国際会議 ISSI2021 で発表した。

研究テーマ B「異なる分野の関連論文検索システムの開発」

分野横断的に関連研究を調査するための関連論文誌検索システム Relevance Venue Finder を構築した。まず、Semantic Scholar コーパスを用いて論文誌(国際会議予稿集合む)間の引用関係ネットワークを構築し、コミュニティ検出手法の適用により論文誌クラスタを計 88 個得た。次に、引用関係ネットワークのエッジ重みに基づき各コミュニティがもつ学際性を定量化した。各論文誌の収録タイトルから共起語を抽出し、[Katsurai&Ono, *Scientometrics* (2019)]で提案した手法を用いて年ごとのトレンド情報を抽出した。Relevance Venue Finder は、クエリ論文誌名が与えられたとき、その論文誌が属するコミュニティの隣接コミュニティのうち、学際性の高いコミュニティから優先的に複数の論文誌を推薦する。任意の論文誌を選択すると、クエリ論文誌と選択論文誌を左右に並べ、それらのトレンド情報および共通する語を可視化する。以上の成果を取りまとめたデモ論文は国際会議 JCDL2021 に採録された(主な研究成果リスト[3])。デモ動画は <https://mm.doshisha.ac.jp/sci2/FindingRelevance.html> に

て参照可能である。発表後はシステムデザインの改良を重ね、共通トレンドとなる語を一覧表示し、その語をクエリとしてさらに関連国際会議を検索可能な機能を追加した。

研究者にとって、関連論文誌を知ることに加え、対象とする課題を取り扱う論文を網羅的に調査することは重要である。そこで本研究では、特定の研究目的に対して多様な解決策を提供する論文を推薦するシステム SolutionTailor を構築した(図 2)。このシステムのユーザは対象とする研

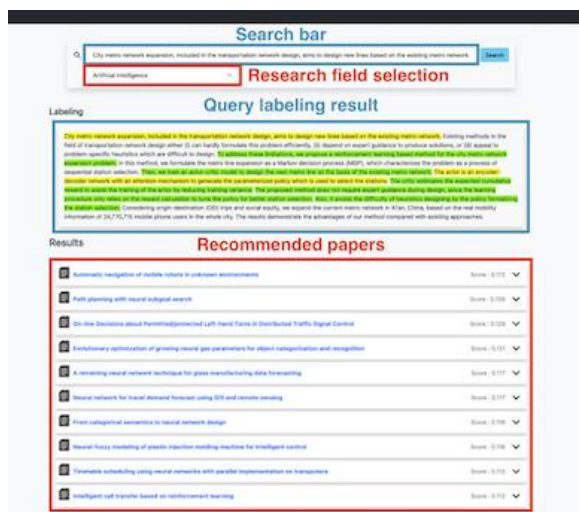


図 2: SolutionTailor のスクリーンショット。

究分野を指定したあと、自身の興味に相当する研究要旨を入力するだけで、その研究と同じ課題で異なるアプローチを提供する論文を指定分野から検索できる。ユーザの研究履歴(過去の引用情報や共著者情報など)を必要としない手軽さが一つの利点である。システム内では、ニューラル言語モデルを用いて、アブストラクト文から背景/目的、手法の説明に対応する文を抽出し、それらの文埋め込みベクトルを用いて論文間の類似度指標を定義した。人工知能関連の約3万5千本の論文コーパスを用いた実験では、全文の埋め込み間で類似度を算出するベールラインシステムに比べ、提案システムの方がクエリ論文の引用情報を超えて特定の目的の文献を推薦できることを示した。以上の成果を取りまとめたデモ論文は国際会議 ECIR2022 に採録決定した(主な研究成果リスト [1])。デモ動画は <https://mm.doshisha.ac.jp/sci2/SolutionTailor.htm> にて参照可能である。

3. 今後の展開

研究期間内では主に科学文献を対象とした共通性発見技術を開発してきたが、個人の発表文献を集約することで研究者間の共通性発見へと拡張可能であると考えている。今後は、国際会議 ICADL2020 にて発表した研究者検索システムに対し、研究トレンド情報や分野間の共通課題などの多様な情報を提示する機能を追加する予定である。

4. 自己評価

【研究目的の達成状況】

異なる学術領域の共通問題に基づく推薦システムを実装した点と、システムのさらなる改良につながる技術要素であるナレッジグラフ構築まで完了できた点から、研究目的は概ね達成できたといえる。一方、時系列グラフの評価は遂行できなかったため、その研究開発は今後の課題としたい。

異分野理解の推進を具現化するデモシステムの構築を重点的に進めてきたが、システムの技術要素には改良の余地があるため、それらはジャーナル論文として今後取りまとめたいと考えている。

研究計画時には想定できなかった点として新型コロナウイルスの発生があり、入構制限などの対処で実装や成果発表に遅れが生じてしまったため、ACT-X 研究期間外も引き続き本課題のアイデアを発展させていきたい。

【研究の進め方:研究実施体制】

本研究は計7名の学生・研究補助者(毎年度更新)の協力の下で進め、大規模データでの実験やデモシステム開発を計画に沿って進めることができた。その他、必要な専門内容に応じて外部の共同研究者(東京大学, 国立情報学研究所, University of Kentucky)と連携した。

【研究の進め方:研究費執行状況】

物品費は主にデータ分析用パソコンの購入に充てた。新型コロナウイルスにより旅費の

支出予定がなくなったため、オンライン環境における研究活動を効率化するために、研究補助業務に必要な物品購入と謝礼に振り替えた。

2021年度には成果展開シーズ・企業提携研究費用サポートに申請する機会をいただき、研究費の追加支援をいただいた。この予算をもとに研究補助者のプログラミング業務を追加し、実験用機器と謝礼の予算を増額した。

【研究成果の科学技術及び社会・経済への波及効果】

2020年度はJST 共済 Beyond ミーティングやJST サイエンスインパクトラボに参加し、多様な業種の方々に対し学術データ分析の意義と将来目標をプレゼンテーションした。2021年度からはJST 産連事業広聴会に出席し、大学URAや知財部門の方々とのディスカッションから得られた知見を今後の実システム開発に生かす予定である。また、本研究で構築したデモシステムをもとに複数の企業と議論し、社会実装への検討を進めた。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 8件

1. Tetsuya Takahashi and Marie Katsurai, "SolutionTailor: Scientific paper recommendation based on fine-grained abstract analysis," in Proceedings of the 44th European Conference on Information Retrieval (ECIR2022), 5 pages, in press.

特定の研究目的に対して多様な解決策を提供する論文を推薦するシステム SolutionTailor を構築した。検索対象とする研究分野を指定したあと、自身の興味に相当する研究要旨を入力すると、同一目的下で異なるアプローチを提供する論文リストを取得できる。実験では、全文の埋め込み間で類似度を算出するベールラインシステムに比べ、提案システムの方がクエリ論文の引用情報を超えて特定の目的の文献を推薦できることを示した。

2. Marie Katsurai and Soohyung Joo, "Adoption of Data Mining Methods in the Discipline of Library and Information Science," Journal of Library and Information Studies, pp. 1-17, 2021.

図書館情報学分野におけるデータマイニング手法利用の動向を調査した論文である。両分野の文献情報を Scopus データベースから収集し、データマイニング手法の用語辞書を自動構築したあと、それらの語の図書館情報学分野における出現頻度を分析した。その結果、ビッグデータ、機械学習、テキストマイニング、情報検索、次元削減など様々な観点のトレンドを明らかとした。特に近年では機械学習技術の応用が活発であることを示した。

3. Tomoya Nishide and Marie Katsurai, "Finding the Relevance Between Publication Venues Based on Research Trend Similarity and Citation Relationships," in Proceedings of 2021 ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL2021), pp. 310-311, 2021.

異分野間での共同研究機会を促進するために、論文誌間の関連性を可視化する新しいツールを提案した。任意の論文誌名を入力すると、引用ネットワーク分析の結果を用いて、関

連する論文誌名を複数表示する。その中から二つ目の論文誌を選択すると、二つの論文誌のトレンド情報を年ごとにグラフ表示し、それらの間の共通キーワードを提示する。

(2) 特許出願

なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- 桂井麻里衣, 「学術データの分析と可視化」, 電子情報通信学会信号処理研究会, 2021年8月。(招待講演)
- 桂井麻里衣, 第11章「異分野融合の促進に向けたサイエンスマッピング」(pp. 217-236), 『デジタル信号処理におけるシステム最適化技術 —基礎技術から音声・音響信号処理, 産業応用と情報システムまで—』, 第11章を分担, オーム社, 2021年10月.
- 桂井麻里衣, 「研究者の知見を産学官で活用するためのマッチングシステム構築」, JSTサイエンスインパクトラボ, 2021年5月.
<https://www.jst.go.jp/sis/scenario/list/2021/05/202105-02.html>