

科学研究費助成事業 研究成果報告書

令和 3 年 6 月 4 日現在

機関番号：12601

研究種目：基盤研究(C)（一般）

研究期間：2017～2020

課題番号：17K00427

研究課題名（和文）大規模学術文献情報からの異種ネットワークの分散表現学習手法に関する研究

研究課題名（英文）Representation learning from multiple heterogeneous graph using scholarly big data

研究代表者

森 純一郎（Mori, Junichiro）

東京大学・大学院情報理工学系研究科・准教授

研究者番号：30508924

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究では、大規模な学術文献情報からの有用な知識の抽出と発見を支援するために、複数の異種ネットワークデータから分散表現を学習するための手法の研究を行った。その上で、実際に大規模論文データ分析のシステム構築を行った。特に、COVID-19に関する大規模な学術文献データセットに対して研究成果である学術産業技術俯瞰システムを用いて引用ネットワーク解析を行い、COVID-19に関する科学的エビデンスや重要な技術などの情報を抽出した。COVID-19に対するエビデンスベースの取り組みを支援するため、その解析結果を広く一般に公開した。これらの研究成果を複数の国際学会や論文誌で発表した。

研究成果の学術的意義や社会的意義

本研究では、大規模な論文データから生成される複数の異種ネットワークから適切な分散表現学習をする手法の知見を明らかにした。また、学習されたネットワーク分散表現を論文データ分析における複数のタスクに適用しその有効性を明らかにした。その上で、実際に大規模論文データ分析のシステム構築を行い、政策立案者、研究者、データベースプロバイタなど科学技術の複数のステークホルダの視点から、ネットワークデータを大規模な学術文献情報からの知識発見に利活用するための知見を明らかにした。

研究成果の概要（英文）：In this study, we propose the method for representation learning from multiple heterogeneous network data in order to support the extraction and discovery of useful knowledge from large-scale scholarly data. We develop the system for large-scale scholarly data analysis. In particular, we conducted citation network analysis on a large dataset of academic literature on COVID-19 using the Academic Industry and Technology Overview System, which is one of our research results, and extracted information on scientific evidence and important technologies related to COVID-19. The results of the analysis have been widely available to the public in order to support evidence-based approaches to COVID-19. The results of these studies have been published in several international conferences and journals.

研究分野：人工知能

キーワード：学術文献データ 引用ネットワーク 表現学習

1. 研究開始当初の背景

学問領域の細分化と科学技術知識の深化に伴い、学術文献情報が大規模に蓄積されて来おり、発表される学術論文の数は、年々爆発的に増加している。論文とその引用関係の集合は膨大なテキストとネットワークのデータであり、それらの大規模な論文データから情報技術を用いて科学技術政策、研究開発、社会問題の解決に有用な高次の知識を横断的に抽出しようという研究の取り組みが、国内外で近年特に盛んになされて来ている。特に、論文の引用ネットワークは、古くから計量書誌分野を中心に研究が幅広く行われて来たが、引用ネットワークのような大規模なネットワークデータをどのように先に挙げたような知識発見に活用するかについては、国内外の研究動向を見ても、その方法論の研究開発は途上にある段階である。特に、研究開発が先んじて現在活発になされている膨大な論文のテキストデータ分析と併せて、ネットワークデータをどのように分析し統合するかを明らかにすることは、大規模な学術文献情報の利活用において重要な課題である。

2. 研究の目的

本研究では、大規模な学術文献情報からの有用な知識の抽出と発見を支援するために、大規模な論文データが構成するネットワークの観点から、ネットワークデータを知識発見に活用するための基本的な方法論の研究を行う。具体的には以下の項目について研究を行う。

大規模な論文データから生成される、引用ネットワーク、共著ネットワーク、などの複数の異種ネットワークから論文や著者などのエンティティの分散表現を適用タスクに応じた低次元のベクトルとして学習する技術

ネットワーク分散表現を用いて論文や著者のクラスタリングやインパクト予測などのタスクを行うことで、政策立案者、研究者、データベースプロバイダなど科学技術の複数のステークホルダのために、大規模な学術文献情報からの有用な知識の抽出と発見を支援する技術

3. 研究の方法

初年度は主に、大規模な論文データから生成される、引用ネットワークや共著ネットワークなどの複数の異種ネットワークから論文や著者などのエンティティの分散表現を学習する技術の実現のために、関連する論文データを網羅的に収集した上で「ネットワークデータからの構造的な内容的な特徴量の抽出手法」および「異種ネットワークデータからの分散表現学習手法」について研究を進める。2年目以降は「異種ネットワーク分散表現を用いた論文や著者のクラスタリング手法やインパクト予測手法」について研究を進めた上で、最終的には研究開発した手法を応用したシステムを構築し、実証実験により手法およびシステムの評価・改善を行う。

4. 研究成果

本研究では、大規模な学術文献情報からの有用な知識の抽出と発見を支援するために、複数の異種ネットワークデータから分散表現を学習するための手法の研究を行った。その上で、実際に大規模論文データ分析のシステム構築を行った。特に、COVID-19に関する大規模な学術文献データセットに対して研究成果である学術産業技術俯瞰システムを用いて引用ネットワーク解析を行い、COVID-19に関する科学的エビデンスや重要な技術などの情報を抽出した。COVID-19に対するエビデンススペースの取り組みを支援するため、その解析結果を広く一般に公開した。これらの研究成果を複数の国際学会や論文誌で発表した。

・ 大規模引用ネットワーク解析基盤の研究開発

代表的な文献データベースである Web of Science、Scopus や PubMed などから論文の大規模な書誌情報ならび引用情報の収集を行い、分析用の高速なデータベースの構築を行った。さらに、収集した書誌情報ならびに引用情報を用いて、論文間の複数の大規模な引用ネットワークを構築し、プロジェクト全体において大規模文献情報を利活用できる基盤を構築した。これらの基盤を元に、大規模文献情報をユーザ自らがインタラクティブに分析可能なウェブシステムである「学術産業技術俯瞰システム」を開発し、公開・運用を進めた。同システムでは、大規模な引用ネットワークに対して逐次的なグラフパーティショニングによるクラスタリングを行い、数百万の論文の関係性を分析することができる。これにより、大規模文献データに基づく学術領域の動向の抽出や可視化が可能にした。

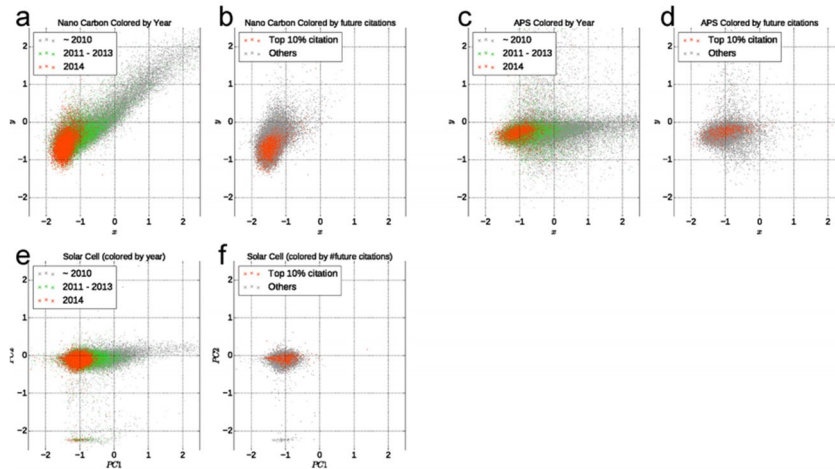
2020 年度の COVID-19 のパンデミックの発生以降において、その学術研究は急速に増加しており、米国の Allen AI 研究所ではそれらの COVID-19 に関連する学術文献データセットを公開しており、その数は 2020 年 8 月時点で 20 万を超える文献の書誌情報が含まれている。同データセットに対して学術産業技術俯瞰システムを用いて引用ネットワーク解析を行い、COVID-19 に関する科学的エビデンスや重要な技術などの情報を抽出した。これにより、研究者のみならず技術者、政策決定者などさまざまなステークホルダーに対して、COVID-19 に対するエビデンスベースの取り組みを支援するため、その解析結果を広く一般に公開した。



学術産業技術俯瞰システムによる COVID-19 関連文献引用ネットワークの解析

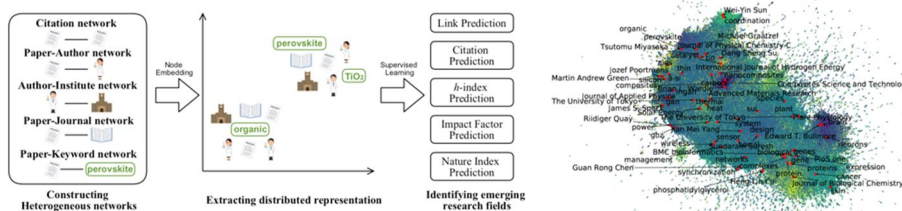
・ 学術領域のトレンド予測手法の研究開発

大規模引用ネットワークに対して、それらの特徴量化するための表現学習手法の検討を行い、論文の引用ネットワーク構造から学習された表現ベクトルを用いて論文の将来のインパクトを予測する基礎的な方法論を構築した。引用ネットワークから学習したネットワーク特徴量を用いることで、学術領域の動向や領域の融合・離散を可視化できることを明らかにした。



学術領域の動向や領域の融合・離散の可視化

さらに、論文、研究者、ジャーナル、研究機関など異なるエンティティから構成される異種の多層ネットワークに対して、各エンティティを特徴量化するための表現学習手法の研究開発を行った。これらの新たな特徴量を元に、対象分野の研究者の将来の H-Index、論文誌の Impact Factor、研究機関の Nature Index などの指標の予測を行い、従来手法を上回る精度で予測ができることを明らかにした。



多層ネットワークの表現学習に基づく研究指標の予測

これらの研究成果は学術産業技術俯瞰システムと併せて国内の科学技術政策関連機関や複数の企業の研究開発部門において、技術のトレンド分析や未来予測に活用されているに至った。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 1件）

1. 著者名 Masaru Isonuma, Junichiro Mori, and Ichiro Sakata	4. 巻 1
2. 論文標題 Unsupervised Neural Single-Document Summarization of Reviews via Learning Latent Discourse Structure and its Ranking	5. 発行年 2019年
3. 雑誌名 Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL2019)	6. 最初と最後の頁 2142-2152
掲載論文のDOI（デジタルオブジェクト識別子） 10.18653/v1/P19-1206	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Masaru Isonuma, Junichiro Mori, and Ichiro Sakata	4. 巻 1128
2. 論文標題 Unsupervised Joint Learning for Headline Generation and Discourse Structure of Reviews	5. 発行年 2019年
3. 雑誌名 Advances in Artificial Intelligence: Selected Papers from the Annual Conference of Japanese Society of Artificial Intelligence (JSAI 2019)	6. 最初と最後の頁 139-149
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-39878-1_13	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Johannes De Smedt, Junichiro Mori, and Masanao Ochi	4. 巻 1128
2. 論文標題 Using Sequence Constraints for Modelling Network Interactions	5. 発行年 2019年
3. 雑誌名 Advances in Artificial Intelligence: Selected Papers from the Annual Conference of Japanese Society of Artificial Intelligence (JSAI 2019)	6. 最初と最後の頁 3-13
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-39878-1_1	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Kimitaka Asatani, Junichiro Mori, Masanao Ochi, Ichiro Sakata	4. 巻 13
2. 論文標題 Detecting trends in academic research from a citation network using network representation learning	5. 発行年 2018年
3. 雑誌名 PloS ONE	6. 最初と最後の頁 197260
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0197260	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo and Ichiro Sakata	4. 巻 1
2. 論文標題 Extractive Summarization Using Multi-Task Learning with Document Classification	5. 発行年 2017年
3. 雑誌名 Proceeding of EMNLP	6. 最初と最後の頁 pp.2101-2110
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計13件 (うち招待講演 0件 / うち国際学会 2件)

1. 発表者名 大知正直
2. 発表標題 異種ネットワーク上のノードエンベディング法による萌芽的研究分野特定のための分散表現抽出
3. 学会等名 第33回人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 田爪聡
2. 発表標題 勾配降下法を用いたニューラルネットワークにおける局所受容野の最適化手法の提案
3. 学会等名 第33回人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 今給黎成彬
2. 発表標題 訓練済み深層生成モデルの潜在変数間相互条件付きサンプリングによるマルチモーダル双方向生成モデル
3. 学会等名 第33回人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 磯沼大
2. 発表標題 木構造ニューラルトピックモデル
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2019年

1. 発表者名 蕭高仁
2. 発表標題 社会課題の早期推定に向けたクラウドソーシングによる教師データの作成とその検討
3. 学会等名 情報処理学会知能システム研究会
4. 発表年 2019年

1. 発表者名 Kimitaka Asatani
2. 発表標題 Keyword extraction using citation network: Verification of network-based method
3. 学会等名 Second International Workshop on SCientific DOCument Analysis (SCIDOCA 2018) (国際学会)
4. 発表年 2018年

1. 発表者名 Kimitaka Asatani
2. 発表標題 Interpersonal relationship in large-scale mobility data
3. 学会等名 "10th International Conference on Complex Networks COMPLENET'19" (国際学会)
4. 発表年 2019年

1. 発表者名 田中 和哉
2. 発表標題 学際性評価のための分野を超えた研究インパクト評価の標準化
3. 学会等名 第32回人工知能学会全国大会
4. 発表年 2018年

1. 発表者名 久保田 修平
2. 発表標題 文脈情報を用いたソーシャルメディアからの社会課題抽出に関する研究
3. 学会等名 第13回Webインテリジェンスとインタラクション研究会
4. 発表年 2018年

1. 発表者名 Kimitaka Asatani
2. 発表標題 Predicting future citation from the temporal information of citation network
3. 学会等名 Predicting future citation from the temporal information of citation network
4. 発表年 2017年

1. 発表者名 磯沼 大
2. 発表標題 文書分類とのマルチタスク学習による重要文抽出
3. 学会等名 第31回人工知能学会全国大会
4. 発表年 2017年

1. 発表者名 浅谷 公威
2. 発表標題 ネットワークからの時系列情報の抽出と可視化
3. 学会等名 第31回人工知能学会全国大会
4. 発表年 2017年

1. 発表者名 田中 和哉
2. 発表標題 深層学習を用いた論文書誌情報による研究専門分野ラベリング
3. 学会等名 第31回人工知能学会全国大会
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関