公開

研 究 終 了 報 告 書

「異なる価値観を融合する検索基盤の創成」

研究期間: 2020 年 11 月～2023 年 3 月
研 究 者: 吉田 壮

1. 研究のねらい

現代の情報化社会において、ソーシャルネットワークが情報と出会う場としての基幹インフラの位置を占めている。膨大なデータ群を高速に探索して望む候補を見つけ出すために、AI 技術が利活用されている。ただ、検索から得た知識は自分好みの偏ったものとなりやすい上、その偏りが自覚されにくい。偏った情報発信を行った場合、その情報が即座に拡散され、非難・炎上するリスクもある。こうしたリスクを回避するには、多様な価値観に触れ、それらを正しく理解して判断することが必要である。

こうした問題に対して、本研究の狙いは、多様な価値観が存在するソーシャルネットワークから正確な情報とその全体像を解釈可能な形で公平に提供する検索を実現することである。これまで利用者のリテラシーに委ねられた正確性の判断と多様な意見の収集・整理を代行する情報科学技術の開発を目指す。主要な研究項目の概要を以下に示す。本研究課題の全体像に関しては図 1 を参照。

【研究項目 1】キーワードの関連度とコンテンツの信頼度の両者に基づくランキング技術：虚偽情報特有のネットワーク情報伝播を学習して、ランキングに虚偽情報が含まれるか否かを数値化した信頼度を推定する。その信頼度を提案者の関連度に基づくランキング最適化手法へ統合することで、関連度と信頼度の両者を考慮する。

【研究項目 2】異なる価値観のコンテンツを広範囲に探索してまとめあげるマイニング技術：ソーシャルメディアから「同じ価値観を有する集合」をマイニングし、複数の集合から等しくコンテンツを抽出して表示することで、多様性に富む検索結果を公平に利用者へ提供する。

【研究項目 3】ブラックボックスモデルの説明可能性技術：【項目 1】における虚偽情報の検出理由および【項目 2】におけるコンテンツの推薦理由を入力変数から説得力のある説明を生成する手法を構築する。

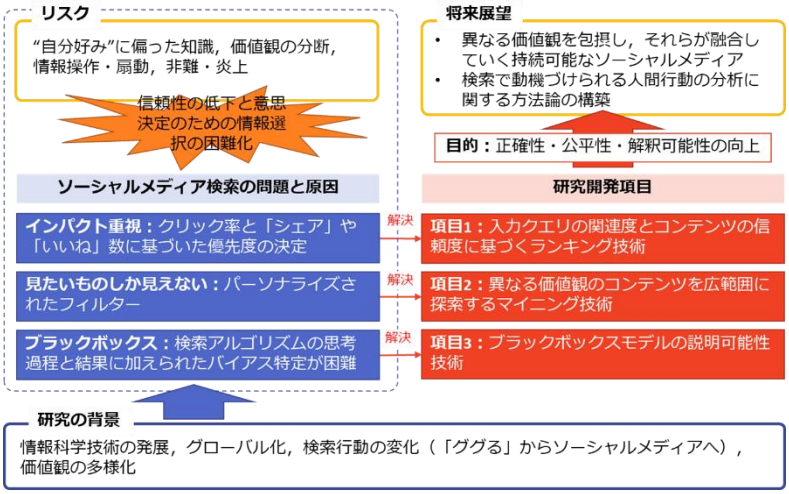

図 1 研究の全体像

公開

2. 研究成果

（1）概要

本研究では、正確で多用な視点の情報を解釈可能な形で検索する手段の構築を目的とする。2022 年度までの研究期間の中で、研究項目 1 キーワードの関連度とコンテンツの信頼度の両者に基づくランキング技術の開発、研究項目 2 異なる価値観のコンテンツを広範囲に探索してまとめあげるマイニング技術、研究項目 3 ブラックボックスモデルの説明可能性技術の開発を実施した。以下にその概要を示す。

まず、情報拡散特性を学習するグラフ深層学習を用いて情報の信頼度を定義し、信頼度と検索意図との適合度の両者でランキングを最適化する技術を構築した。正確な情報をくみ上げ、虚偽情報を目に付かない位置に低ランク付けすることで、複雑で新しい虚偽に対しても柔軟に対処可能とした。このように、真偽の判断をランキング問題とみなし、検索意図と適合する情報検索と同時に実現する点が独自である。

次に、ソーシャルネットワークから同じ意見を有する集合をマイニングするアルゴリズムおよび多様な視点の情報をユーザへ推薦するアルゴリズムの開発を行った。まず、ソーシャルネットワークから同じ価値観を有する集合を抽出する手段を検討した。さらに、既存の推薦システムが生成する推薦候補とコミュニティとの関連の分析を実施することで、フィルターバブルの存在を確認した。この結果を踏まえて、ユーザの閲覧履歴との関連性を最大化する既存の推薦システムとは異なる、推薦対象ユーザが属するコミュニティとは異なる複数のコミュニティのユーザの投稿を推薦する手法を構築した。本研究遂行に必要なソーシャルメディアデータは、議論や情報拡散の中心的な手段となっている Twitter から収集した。

上記研究項目 1・2 における虚偽情報の検出モデルおよび推薦モデルは、グラフニューラルネットワーク（GNN）に基づいて構築されている。最後に、GNN に対する予測結果の説明機構を構築した。ここでは、予測結果に重要な役割を果たす部分グラフと特徴量を特定する方式を提案した。この成果に基づいた実データを用いた可視化システムを開発、研究協力者を通じた実験により有効性を評価した。

（2）詳細

<u>研究項目 1 キーワードの関連度とコンテンツの信頼度の両者に基づくランキング技術</u>

① 虚偽情報の拡散特性に基づく信頼度の推定

ソーシャルネットワーク上の情報が虚偽に関与しているか否かを数値化した信頼度を推定する手法を提案した。ここでは、真偽で異なる情報が投稿されてから情報拡散特性を GNN（Graph Neural Network)に基づいて、真偽の 2 クラス識別モデルを学習およびそこから出力される値を信頼度に利用する方式を構築した。この基礎アルゴリズムには代表者の技術が用いられる[5-(1)-1]。ここでは、ソーシャルネットワークの相互作用（Twitter ではリツイートを指す)による情報拡散をグラフでモデル化した。投稿内容の真偽を表すラベルを、政治にまつわる発言・声明の信憑性をチェックするサイト PolitiFact を参照して付与することで、ネットワークの訓練に必要なトレーニングデータを作成した。406,333 名のユーザ（内、虚偽情報の拡散に関与したユーザは 268,513 名）から計 471,723 件（真：338,866 件、偽：132,857 件）のツイートデータと 583,516 件のリツイートデータを取得した。実験の結果、GNN アルゴリズムを適用して情報拡散特性を利用した結果が従来手

法よりも上回っていることが確認できた。情報拡散特性を考慮した特徴ベクトルへ更新することで満足な特徴量が得られ、大幅な精度向上に貢献した。したがって、GNN を用いて伝播特性を得た特徴ベクトルを用いると、フェイクニュースを拡散するユーザの検出精度が向上することが確認できた[5-(3)-4]。本研究を進展させた成果は、2021 年時点で最新の手法を上回る精度を達成した[5-(1)-2]。また、信頼度を推定するアイデアは、一般の Deep Neural Networks (DNN)へ拡張し、ノイズラベルを含む訓練データを用いた場合でも頑健な DNN の学習法を提案した[5-(1)-1, 5-(3)-2]

② ランキング技術の構築

推定した信頼度を代表者のランキング最適化へ導入することで、検索意図との適合度と信頼度の両方が高い情報を並べたランキングを生成し、虚偽情報の順位を下げる手法を提案した。本手法は、ランキングの近傍順位に存在する文書特徴量の類似度を最大とすることを尤度とした目的関数(Yoshida et al., 2018)を基に、ランキングを並べ替えることで検索意図との整合性を取る。ここでは、キーワードと合致する文書の候補集合から単語分散表現法の BERT 特徴量(Devlin et al., 2018)を抽出し、①で推定した信頼度によって重み付けられた類似度の算出法を構築した。手法の性能は、Twitter データセットに対して 5 つのクエリを含むセット: 1) 2018 us election result、 2) north korea nuclear issue、 3) president donald trump re-election、 4) national anthem protests、 5) las vegas shooting をそれぞれ入力して検索を行った際に、検索上位にどの程度の虚偽情報が含まれる割合に基づいた比較実験により評価した。実験結果から、提案手法を適用することで、全ての順位で精度向上を確認でき、具体的な数値では、検索上位 10 件で 43%の精度向上を実現した。

<u>研究項目 2 異なる価値観のコンテンツを広範囲に探索してまとめあげるマイニング技術</u>

③ ソーシャルネットワークコミュニティの同質性評価

まずソーシャルネットワークから同じ価値観を有する集合を抽出する手段を検討した。本研究では、ネットワークの中で親しく、強く結びついているユーザのクラスタをコミュニティと呼び、コミュニティが同じ価値観の意見が醸成する場となると仮定する。この仮定を実証するため、Twitter ユーザのリプライ・リツイートネットワークから抽出されるコミュニティに属するユーザの同質性を分析した。実験では、2020 年 1 月から 6 月の期間に Black Lives Matter とその対抗運動である AllLivesMatter および BlueLivesMatter に反応したツイートとそれらを支持するリプライ・リツイートを収集することで、異なる 3 種類の考え方が含まれるデータセットを作成した。本研究では、Leiden クラスタリング法(Traag et al., 2019)を用いて、ツイート・リプライ・リツイートの構造からユーザをクラスタに分類することで、コミュニティを検出した。続けて、ユーザの属性をそのユーザが支持する運動に割り当てることで、コミュニティの同質性を定量的に分析した。その結果、同一のコミュニティに属するユーザが持つ意見は、高い同質性を持つことを明らかにした。このように、ネットワークをコミュニティに分割することは、同じ価値観を有する集合を検出するのに有効である。

次に、フィルターバブルのコミュニティとの関連について分析を行った。既存の推薦システムは、ユーザの行動履歴に基づいて推薦候補ツイートを生成するため、リプライ・リツイート構造がコミュニティの中で作られている場合、システムは同コミュニティに属する投稿ツイートを優先的に選択し、バブルを生み出す。ここでは、データセットから 330 名のユーザをランダムに抽出し、各ユ

公開

ーザがリプライ・リツイートしたツイートの投稿者が属するコミュニティを調査した。図 2 では、Gini 係数を用いて、ユーザがリツイートした先のコミュニティの偏りの大きさを示した。調査の結果、69%のユーザが 5 未満のコミュニティからリツイートしていること、22%のユーザは 1 つのコミュニティからリツイートしていることが明らかになった。以上の結果は、ユーザは少数のコミュニティから発信された情報を閲覧すること、これらの履歴を基礎に推薦候補を生成するシステムはコミュニティの偏りを助長することを示唆する。

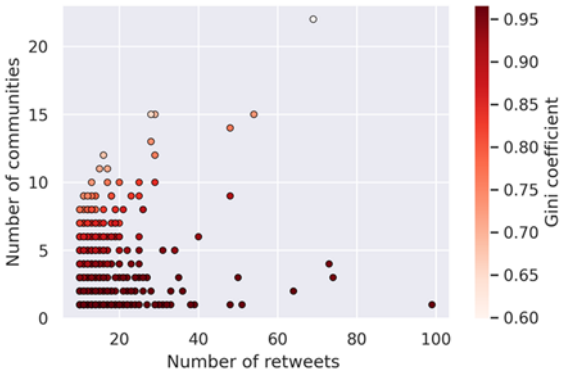

図 2 ユーザのリツイート数と元ツイートの投稿者が属するコミュニティの関係

④ 多様な視点の情報をユーザへ推薦するアルゴリズム開発

多様な視点の情報を推薦するアルゴリズムを提案した。提案モデルは、推薦候補が属するコミュニティを検出し、推薦候補のリストのコミュニティ網羅性を最大化することで、ユーザが探索可能な意見の多様性を拡張する。ここでは、推薦候補を順位付けして並べたものをリストと呼ぶ。具体的に、まずリストのコミュニティ網羅性と非冗長性に基づいて多様性を評価する定量指標を定義した。非冗長性とは、リストに同じコミュニティからの情報が連続しないことを指す。次に、この指標に基づいて、ユーザの行動履歴との関連性を維持しつつも、リストのコミュニティ冗長性を排除するように推薦候補の順位を並べ換える手法を構築した。本手法は、行動履歴に基づく推薦候補の生成及び多様性に基づく推薦順位の決定を end-to-end で可能、関連性と多様性のパラメータ λ を用いて、リストの多様性を制御可能である。手法の性能は、ユーザの閲覧履歴との関連性を最大化する既存の推薦手法との比較実験により検証した。実験では、③と同様に Twitter データセットから 330 名のユーザを抽出し、各ユーザに対して提案手法と比較手法が生成する推薦候補の上位 200 件のコミュニティ多様性を、ジニ係数を用いて評価した。また、Social Bayesian Personalized Ranking（Zhao et al., 2014）と Neural Graph Collaborative Filtering（Wang et al., 2019）をベンチマークとし、提案手法に加えて、ランダムにリストを並べ替えた結果と比較した（図 3）。提案手法は、ランダムを上回るように推薦結果の多様性を拡張できることを確認した。③の結果を踏まえて、コミュニティの偏りが無い情報をユーザに提供することで、異なる価値観の意見に触れる機会の増大に提案手法は寄与すると考えている。

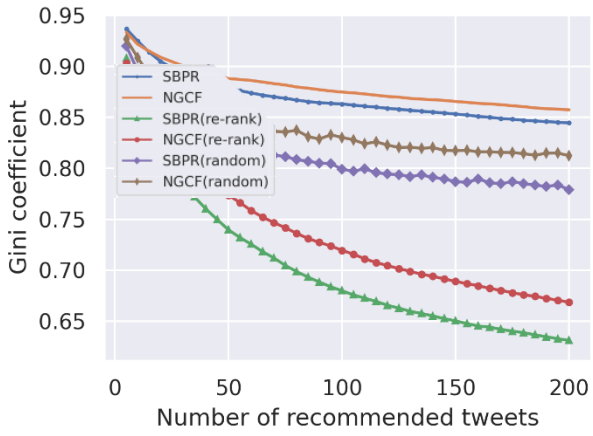

図 3 推薦順位上位 5 件から 200 件までの Gini 係数の比較（ λ =0.5）

公開

研究項目 3　ブラックボックスモデルの説明可能性技術

⑤　予測結果に影響を与えた説明変数の可視化法の開発

グラフニューラルネットワークに基づいたグラフ分類モデルの予測結果を説明可能な特徴を抽出する手法について開発を行った[5-(1)-3]。研究項目 1 および研究項目 2 において GNN が基盤アルゴリズムであることを踏まえ、GNN を用いたグラフデータの分類モデルを構築し、各グラフと検出モデルを説明モデルに入力し説得力のある説明を生成する。具体的に、GNN の予測結果と可能な部分グラフの分布との間の相互情報量を最大にする条件を抽出する方式を導入した。ここで抽出される条件が説明性の最も高い条件となる。提案手法の有効性は、上記①②で整備したデータセットに対して適用することで確認した。以上のように、グラフ分類モデルの説明可能性技術を開発した一方で、コンテンツの推薦理由を説明するためにはリンク予測問題への拡張が必要であり、さらなる検討が必要と考えている。

⑥　可視化システムの開発

上記⑤で有効性が確認された予測結果の可視化法を実装した可視化システムを構築した（図 4）。本システムは、GNN モデルを入力、特定された予測結果に重要な役割を果たす部分グラフと特徴量を可視化する。また、提案システムには異なる予測結果の比較機能が実装されている（図 5）。異なる予測結果を示した 2 つの入力を並べ、それらの間で有意差が存在する特徴を可視化する。すなわち、本システムは可視化による説明可能性を付与するのみでなく、使用者自身のコンテンツの信頼度を見極める能力を養うことができる。システムの有効性は、実験協力者を対象に SUS(System Usability Scale)等の指標を基に評価した。しかし、SUS は妥当性評価最低基準付近である 56 を示し満足な結果を得ることができなかった。同時に実施したアンケート調査では、学習のしやすさという点では有効であった一方で、生成した説明の理解のために専門的な知識を有することが指摘された。

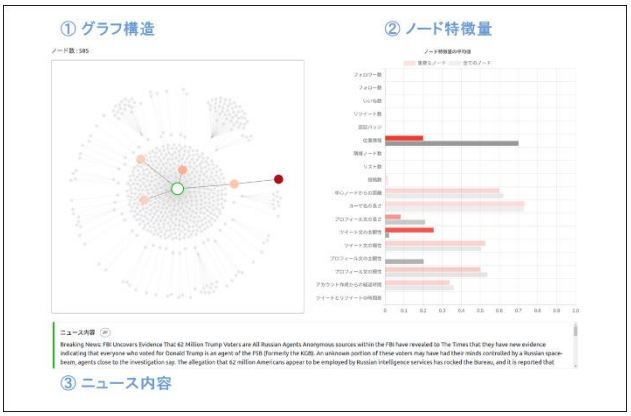

図 4　可視化システムのユーザ・インターフェース

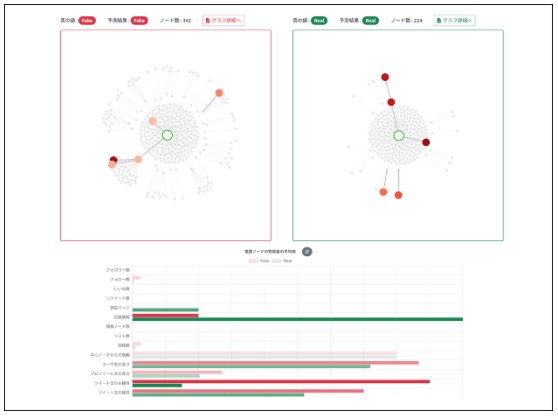
図 5　異なる予測結果の比較機能の実装

3.　今後の展開

本研究課題の実施計画に基づき、多様な価値観が存在するソーシャルネットワークから正確な情報を公平に提供する検索基盤を開発できたと考えている。しかし、その全体像を解釈可能とする技術の開発、すなわち【研究項目 3】の進展が今後の目標と考えている。予測結

果を可視化するシステムのデザインはできているため、アルゴリズム開発を行えば、運用できると考えている。

現状の成果には課題がある。研究成果の項で述べた結果は、本研究で作成したデータセットが支持する範囲である点が、本研究のリミテーションである。今後は、データセット規模を拡張することで、本研究の学術的な側面の発表を進展させることを考えている。

また、ACT-X での研究の社会実装に向けて、科研費・若手研究に応募し採択された。ここでは、リアルタイム性の高い社会問題をテーマに取り上げ、本研究項目を進展させ、有効性の追試等を行うことで、本研究課題の遂行が社会問題解決のために応用・展開できるかの検討を行うことを考えている。

4. 自己評価

研究目的の達成状況

本研究では、2022 年度までの研究期間の中で 3 つの具体的な研究項目の達成を目指した。いずれもアルゴリズム構築および実社会で生成されたデータを用いたシミュレーションと定量評価までを実施できたことから、基盤構築については多くの進捗を得られたと考えている。手法面については特に、【研究項目 1】で開発した情報の正確性を判定する GNN に基づく虚偽情報検出法は、2021 年に公開された最新の比較手法を上回る精度を達成することに成功した。【研究項目 2】については、代表者が知る限り、システムが推薦する意見の多様性を広げフィルターバブルを軽減するために、コミュニティの網羅性を拡張する推薦の再順位付け法を提案する最初のアプローチであった。また、ソーシャルネットワークネットワークのコミュニティを単位にバブルが形成されることを実証するため、既存の推薦システムの振る舞いに関する実験的な分析を実施できた点が自己評価できる。しかし、【研究項目 3】については、GNN モデルの分類問題への説明生成は達成できたものの、推薦理由を述べるリンク予測問題への説明生成には至らなかった。一方で、継続されるプロジェクト内で粘り強く取り組み続けることで、1 年内で手法の開発を進める。改善点として、今後の展開の項で述べたように、学会発表等を通じて研究成果の周知を図り社会への波及に努めたい。

研究の進め方

本研究課題を通じて、研究代表者が研究を遂行できるように、GPU 搭載したワークステーションや RAID 付きストレージなど各種機器を揃えた。これにより、深層学習に必要な並列演算およびソーシャルネットワークから収集されるデータセットの拡充が可能となり、研究に専念することができた。したがって、個人型研究を遂行するために必要な周辺環境に対して、適切に研究費を執行できた。

研究実施体制については、当初計画では学生などの研究補助員の雇用を考えていたが、適切な人材を見つけることができず、採用を見送った。そのため、研究代表者が単独でデータ整理やアノテーション作業を行った。これらの補助業務を研究補助員が担当することで、研究代表者が基盤アルゴリズムの開発に注力できたことが反省点である。

研究成果の科学技術及び社会・経済への波及効果

今日、偏った意思決定と情報発信は非難の的として世界に拡散される。ソーシャルネットワークは世論形成の影響力が大きいため虚偽や扇動が発生し、情報操作のためのツール

公開

化される危険性がある。このような問題の解決へ導く技術の実現は、一層求められている状況にある。これまで利用者のリテラシーに委ねられた正確性の判断と多様な意見の収集・整理を代行する情報科学技術を支える基礎アルゴリズムを開発し、社会実装することは、この要求に答えることができるため、社会的にインパクトがある成果を生むことができると考える。

5. 主な研究成果リスト

（1）代表的な論文（原著論文）発表

（2）研究期間累積件数：7件

1. H. Takeda, S. Yoshida and M. Muneyasu, Training Robust Deep Neural Networks on Noisy Labels Using Adaptive Sample Selection with Disagreement, IEEE Access, vol. 9, pp. 141131-141143, 2021.

本論文では、ノイズラベルを含む訓練データを用いた場合でも頑健な Deep Neural Networks (DNN)の学習法を提案した。提案手法では、DNN の Memorization Effect に基づく小損失基準を用いること、つまり、学習中のネットワークの損失を観察することで、クリーンなサンプルとノイズの多いサンプルを識別している。提案手法のノイズの多いラベルに対する頑健性を検証するために、一般的に用いられる 5 つのベンチマーク、MNIST, CIFAR-10, CIFAR-100, NEWS, T-ImageNet を用いて実験を行った。DNN の損失値に基づいて、サンプルの信頼度を推定するアイデアは研究項目 1 で導入された。

2. H. Matsumoto, S. Yoshida and M. Muneyasu, Propagation-Based Fake News Detection Using Graph Neural Networks with Transformer, Proc. 2021 IEEE 10th Global Conference on Consumer Electronics（GCCE2021）, pp.19-20, 2021.

本論文では、Graph Transformer Network (GTN) を用いたフェイクニュース検出手法を提案した。最近の研究では、フェイクニュースとリアルニュースはソーシャルメディア上で異なる拡散構造を持つことが報告されている。そこで、ユーザをノード、ニュース共有チェーンをエッジとするグラフを構築し、グラフニューラルネットワーク（GNN）を用いて伝播パターンとユーザの嗜好を同時に学習する伝播型検出手法が注目されている。しかし、グラフからユーザの嗜好を抽出するためには、未接続のノード間の関係を学習することが課題である。GTN は、元のグラフのノード間の有用な接続を特定しながら、効率的にノード表現を学習することができる。提案手法の有効性は、Twitter データからなるデータセットを用いた比較実験により示した。

3. H. Matsumoto, S. Yoshida and M. Muneyasu, Flexible Framework to Provide Explainability for Fake News Detection Methods on Social Media, Proc. 2022 IEEE 11th Global Conference on Consumer Electronics（GCCE2022）, pp.421-422, 2022.

本論文では、グラフニューラルネットワークに基づく分類モデルに対し、柔軟な方法で説明性を付加できるフレームワークを提案した。提案手法では、GNN を用いたグラフデータの分類モデルを構築し、各グラフと検出モデルを説明モデルに入力し説得力のある説明を生成する。Twitter から抽出した実データセットに対する実験により、提案する説明の有効性を定量的に評価した。

公開

（3）特許出願

研究期間累積件数：0 件

（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）

1. K. Soga, S. Yoshida and M. Muneyasu, Propagation-Based Fake News Detection Using a Combination of Different Content Features, Proc. 2022 IEEE 11th Global Conference on Consumer Electronics（GCCE2022）, pp.411-412, 2022. *(IEEE GCCE 2022 Excellent Poster Award Silver Prize)*
2. R. Higashimoto, S. Yoshida and M. Muneyasu, A Robust Learning Framework Using Self-Supervised Learning for Learning With Noisy Labels, Proc. 2022 IEEE 11th Global Conference on Consumer Electronics（GCCE2022）, pp.409-410, 2022.
3. 東本 良太, 吉田 壮, 棟安 実治, 学習初期の正則化と加重損失を用いたラベルノイズに頑健な半教師あり学習, SIS2022-16, 2022.
4. 吉田 壮, 松本 勇人, 棟安 実治, グラフニューラルネットワークを用いたフェイクニュースを拡散するユーザ検出, SIS2020-48, 2021.