

科学研究費助成事業 研究成果報告書

令和 4 年 6 月 2 日現在

機関番号：12601

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20413

研究課題名（和文）投稿群・投稿者群を単位するソーシャルメディア上の情報拡散のモデル化に関する研究

研究課題名（英文）Research on modeling of information diffusion on social media based on post groups and user groups

研究代表者

榊 剛史（Sakaki, Takeshi）

東京大学・未来ビジョン研究センター・客員研究員

研究者番号：00735805

交付決定額（研究期間全体）：（直接経費） 3,000,000円

研究成果の概要（和文）：本研究では、SNS上での大規模な情報拡散を対象として、それらには実際にはどのような投稿者や話題が含まれているかを俯瞰的に捉えるための分析手法を確立すると共に、拡散に寄与する投稿者や話題の特徴を明らかにした。特に、拡散に寄与する投稿者同士のネットワークが持つ特徴・言語的な特徴が情報拡散の規模に影響を与えることを明らかにした。一方で、投稿者のネットワーク形成においては、共通の話題が影響することが明らかになった。これにより、最初は共通の話題についてコミュニケーションを行う中で投稿者のネットワークが形成された後、そのネットワークの特徴によって情報拡散の規模がある程度規定される、という仮説が得られた。

研究成果の学術的意義や社会的意義

SNS上での炎上やインフォデミックの発生の発生に伴い、SNS上の情報拡散が社会全体に与える影響が大きなものとなっている。そのような中で、本研究課題では、情報量の多さ故に人間が捉えることが難しい大規模な情報拡散について、それらを俯瞰的に理解可能な形で可視化できる分析手法を提供した。このような分析手法により現象を可視化することで、人々の適切な意思決定を支援することができ、ひいては大規模な情報拡散から生じる社会の不利益を低減できることができると考えている。また、そのような情報拡散を引き起こす要因を明らかにしたことにより、誤った情報や不適切な情報の拡散を低減することにも繋がると考えている。

研究成果の概要（英文）：In this study, I established an analytical method to overview the posters and topics included in large-scale information diffusion on SNS, and clarified the characteristics of the posters and topics that contribute to the size of information diffusion. In particular, I found that linguistic features of posters and the network structure of poster affect the scale of information diffusion. Additionally, I found that talking common topic influences the formation of the network structure of posters. This leads to the hypothesis that after the network of contributors is initially formed through communication on a common topic, the scale of information diffusion is to some extent determined by the characteristics of that network.

研究分野：ウェブマイニング

キーワード：SNS インフォデミック 情報拡散分析 社会ネットワーク分析 ユーザ行動モデリング 計算社会科学

1. 研究開始当初の背景

コンピュータ・ウェブ・電子デバイスの発達に伴い、コンピュータを媒介としたコミュニケーション (Computer-mediated Communication, 以下 CMC) の割合はますます増加している。それに伴い個人の行動やコミュニケーションのデータが大規模に蓄積されるようになった。その結果、社会学においてメゾスコピック¹な領域を扱える素地ができあがってきた。社会学においては、マイクロマクロリンクと呼ばれるように個人のミクロな行動が蓄積されて全体としてマクロな社会現象を引き起こすという考え方がある。これまでは、大規模な人数に対して粒度の粗い調査を行うことで世の中全体の傾向や特徴を明らかにするマクロなアプローチ、ごく小規模な人数に対して粒度の細かい調査を行うことで個人の行動の傾向や特徴を明らかにするミクロなアプローチが行われてきたが、その中間、つまりメゾスコピックな領域に関する研究は、シミュレーションを除いては、あまり行われてきてこなかった。これには様々な要因が考えられるが、数万人～数十万人について詳細な行動やコミュニケーションのデータを蓄積することの困難さが大きな要因の一つだろう。しかし、個人の行動やコミュニケーションに関する詳細なデータが大規模に蓄積されるようになった現在、ミクロな個人行動とマクロな社会現象の間をつなぐようなメゾスコピックな領域について、そのメカニズムを解明する素地が整ったと言える。一方、CMCの増加により、人同士のコミュニケーションのあり方が変化している。それに伴い今までは問題とならなかった現象が発生している。特にソーシャルメディアの登場により、1対多のコミュニケーションのあり方は大きく変化した。以前は、1対多のコミュニケーションはマスメディアなど一部の対象だけに限られていたが、ソーシャルメディアにより小規模な組織や個人でも多数に向けて情報を発信することが可能となった。それによりコミュニケーションに起因する問題が発生し、社会全体に負の影響をもたらしている。例えば、これまでは個人のたかだか数十人～百人程度に向けたメッセージが数万～数十万規模に拡散されることで、異なる価値観の衝突 (炎上) や極性化現象 (エコーチェンバー現象) などが観測されるようになった。また、情報発信のコストが下がることで、広告費の獲得や意見の誘導を目的とした情報発信 (フェイクニュース・悪質なキュレーションサイト) が出現しているが、これらのサイトはユーザ全体に影響を与えることを目的としておらず、あくまで一定規模のユーザを集めることを目的としている。これら2つの事例は、ソーシャルメディア上の情報拡散により、マクロとミクロの中間領域 (ここれでは数万～数十万規模のユーザ) において問題のある現象を生み出していると言える。このような中でミクロとマクロを繋ぐようなメゾスコピックな領域における情報拡散をモデル化することの重要性が高まっている。

2. 研究の目的

本研究の目的は、a.投稿情報 (投稿内容や属性)、b.投稿者情報、c.投稿者と投稿の関係性という情報拡散の構成する3つの異種情報に着目し、投稿群・投稿者群を単位として、ソーシャルメディア上の情報拡散をモデル化することである。

3. 研究の方法

SNS上の情報拡散分析について、投稿者群 (コミュニティ) / 投稿群 (トピック) という単位で情報拡散を分析することで、認識可能な形で可視化することに成功した。ここでのコミュニティとは、同一の属性を持つユーザ群を意味する。また、トピックとは、類似した内容を扱った投稿群を意味する。

(1) コミュニティとトピックの抽出

投稿を拡散した投稿者の類似性 (cの情報) を用いて投稿をクラスタリングし、それをトピックとした。また日常的な投稿者間の社会ネットワーク (bの情報) を用いて投稿者をクラスタリングしてコミュニティとした。そして、各トピックをコミュニティ単位の分布により表現することで、その特定の情報拡散の事例が、どの程度幅広く語られているかを定量的評価する。またその際に、トピックの妥当性を評価するために a. 投稿内容の情報を用いる

(2) コミュニティ・トピックを用いた情報拡散の分析

コミュニティ・トピックを単位として、情報拡散の分析を行ういくつかの手法を提案した。

¹ メゾスコピックとは、物理学においてマクロとミクロの間に位置する対象を扱う学術領域を意味する。

① 情報拡散の偏りと多用性に関する評価

構築したコミュニティを用いて情報拡散の偏りと多用性を評価する手法を提案した。

偏りの評価指標については、特定の情報拡散について、元々ユーザ規模に基づくコミュニティの分布と、分析対象に含まれるユーザが所属するコミュニティの分布の違いに着目し、それらの違いが大きければ大きいほど、特定のコミュニティに偏った話題であると考えた。具体的には2つのコミュニティ分布の KL Divergence を算出し、それを情報拡散の偏りを表す指標とした。

多用性の指標については、トピックを拡散したユーザが所属するコミュニティがどの程度多様化か？を情報量の考え方から求めるトピックエントロピーという指標を提案した。

② 次数分布によるコミュニティの外向性の評価

複数のコミュニティについて、それらの外向性を評価する手法を提案した。具体的には分析対象のコミュニティが持つネットワークに置いて、コミュニティ内に向けたリンクの次数分布とコミュニティ外に向けたリンクの次数分布を比較することで、複数のコミュニティの外向性(内向性)を評価する手法を提案した。

③ 言語的特徴に基づくコミュニティ間の類似性の評価

複数のコミュニティについて、それらの持つ言語的特徴の類似性を比較する手法を提案した。具体的には、各コミュニティがよく利用する形容詞を抽出し、それらの意味的類似度の平均値を作出することにより、コミュニティの言語的特徴の類似性を評価する手法を提案した。意味的類似度としては、事前学習済の単語分散表現を用いた。

4. 研究成果

提案した手法を用いて新型コロナ禍、様々な社会課題、政治的なトピックに関する情報拡散分析を行った。

(1) 新型コロナ禍に関する分析

新型コロナ禍の初期段階において、新型コロナに関する話題がどのように拡散していったかの過程を分析した。

図1より、最初期の2020年1月は偏りが大きく、まだ一部の話題であったことが分かる。1月31日に日本の感染者が10名を超えた段階で、一旦極小値を取るが、またしばらくは偏りが上昇する。2月13日に日本初の死者が出たあたりから、再び偏りが減少し、幅広いユーザ群により語られるようになった。その後は偏りは上がらなかったことから、一般的な話題として定着したことが見て取れる。

このように新型コロナに関する拡散の偏りを可視化することで、どのタイミングで SNS 全体に危機感が広がり、話題として注目されるようになったかを俯瞰的に理解することができる。



図1 KL-Divergence による偏りの評価

(2) 社会課題の分類

研究課題 19K20412「SNS データを利用した早期の社会課題検出のための基盤技術開発」と連携し、SNS から早期検知された社会課題について、その拡散度合いの可視化を行った。

図2は縦軸が社会課題ごとの SNS 上での言及数、横軸がそれらのトピックエントロピー (多用性指標) を表した散布図である。このように言及数と多用性の2つの軸で表すことで、それらを4種類の社会課題に分けることができる。右上の象限のような言及数・多用性共に高い「1. 大きな社会課題」、右下のような言及数は少ないが多用性は高い「2. 社会課題の芽」、左上のような言及数は多いが多用性は低い「3. 一部の人のための重要な社会課題」、左下のような言及数・多用性共に低い「4. 一部の人の

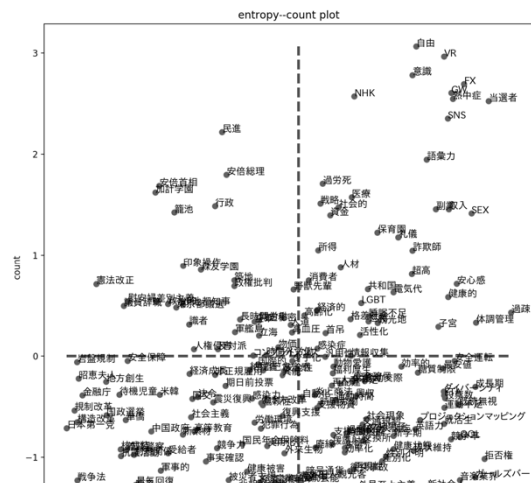


図2 社会課題の言及数と多用性

とっての社会課題の芽」という4種類である。「1. 大きな社会課題」として熱中症や過労死, また, 「2. 社会課題の芽」として QOL(Quality Of Life:生活の質) や既読無視や英語力といった言葉が得られた. このように多用性と言及数を組み合わせて表現することで, 社会課題を4つに分類した上で, 潜在的な社会課題と思われる「2. 社会課題の芽」を選択的に抽出することが可能になる.

(3) 政治的なトピックの分析

大規模な情報拡散が頻繁に起きているトピックとして政治に着目し, 情報拡散を誘発する要因について分析を行った. 具体的には, 政治トピックの拡散に関わる SNS ユーザ群を政治的イデオロギー(保守・リベラル・中間層)で分類した後, それらの SNS 全体への波及効果を調査した. さらに, どのような特徴が波及効果の大きさに影響を及ぼしているかを分析した.

具体的には, 保守ユーザ層の投稿の方がリベラル層の投稿よりも中間層から拡散されやすいことが明らかになった. そこでそれら2つのコミュニティの情報拡散について, その違いを生み出す要因の分析を行った.

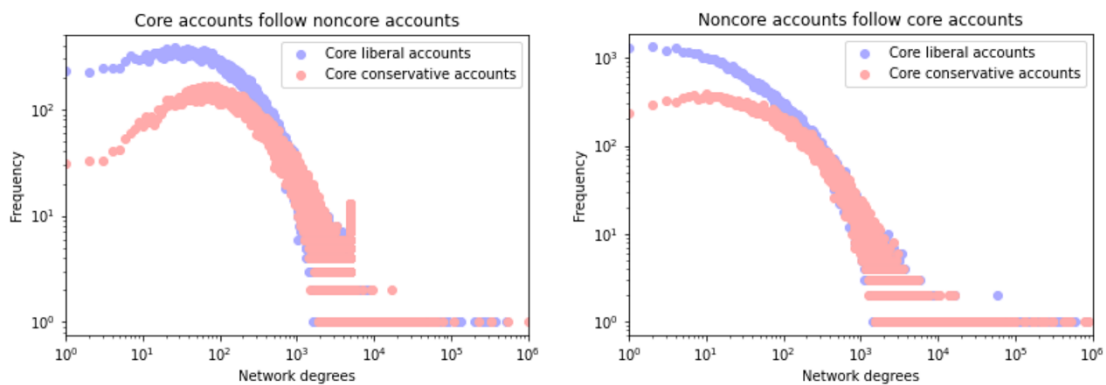


図3 保守・リベラルアカウントの外向性の比較

図3は, 保守アカウント・リベラルアカウントの外向性を表したグラフである. 図3の左図は, 保守アカウントから中間層アカウント・リベラルアカウントから中間層アカウントへの有向エッジの次数分布である. 保守アカウントの方がより多数の中間層アカウントに向けて有効エッジを張っている事がわかる. 図3の右図は中間層アカウントから保守アカウント・中間層アカウントからリベラルアカウントへの有向エッジの次数分布である. 保守アカウントの方がより多数の中間層アカウントから有効エッジを張られている事がわかる. これより, リベラルアカウントと比べて, 保守アカウントの方が中間層アカウントとより密なネットワークを構築していることがわかる.

図4は, 保守アカウントとリベラルアカウント・中間層アカウントについて, それぞれの言語的な類似度を表したグラフである. 上3つは名詞の類似性, 下3つは形容詞の類似性である. 使用している名詞の上では保守アカウントとリベラルアカウントは類似している. つまり類似したトピックに言及していると考えられる. 他方, 使用している形容詞の上では, リベラルアカウントと比べて保守アカウントの方が中間層アカウントと類似している. つまり, 言葉遣いの上では, 保守アカウントとリベラルアカウントで違いがあり, 保守アカウントの方がより中間層アカウントに近いことがわかる.

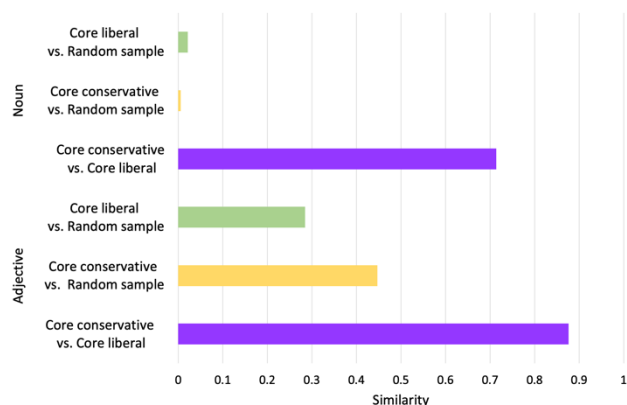


図4 保守・リベラルアカウントの言語的類似性

これらの2つの分析より, 保守アカウント・リベラルアカウントの投稿に対する中間層アカウントによる情報拡散の違いは, それらの持つネットワーク構造と言語的類似性が影響することが示唆された.

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 3件）

1. 著者名 Toriumi Fujio, Sakaki Takeshi, Yoshida Mitsuo	4. 巻 35
2. 論文標題 Social Emotions Under the Spread of COVID-19 Using Social Media	5. 発行年 2020年
3. 雑誌名 Transactions of the Japanese Society for Artificial Intelligence	6. 最初と最後の頁 F~K45_1-7
掲載論文のDOI (デジタルオブジェクト識別子) 10.1527/tjsai.F-K45	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 Chou Jen Shiau, Masanao Ochi, Takeshi Sakaki, Ken Nagahama, Kanji Sakai, Junichiro Mori, Ichiro Sakata	4. 巻 2020
2. 論文標題 Constructive Approach for Early Extraction of Viral Spreading Social Issues from Twitter	5. 発行年 2020年
3. 雑誌名 The Proceedings of 12th ACM Conference on Web Science	6. 最初と最後の頁 96,105
掲載論文のDOI (デジタルオブジェクト識別子) 10.1145/3394231.3397899	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 Asatani Kimitaka, Yamano Hiroko, Sakaki Takeshi, Sakata Ichiro	4. 巻 11
2. 論文標題 Dense and influential core promotion of daily viral information spread in political echo chambers	5. 発行年 2021年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 7491
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-021-86750-w	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 Uchida Kazuki, Toriumi Fujio, Sakaki Takeshi	4. 巻 17
2. 論文標題 Comparative evaluation of two approaches for retweet clustering: A text-based method and graph-based method	5. 発行年 2019年
3. 雑誌名 Web Intelligence	6. 最初と最後の頁 271~284
掲載論文のDOI (デジタルオブジェクト識別子) 10.3233/WEB-190418	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6件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 鳥海 不二夫, 榊 剛史
2. 発表標題 コミュニティの偏りを用いた炎上の一般性評価:再燃焼系炎上における分析
3. 学会等名 人工知能学会全国大会論文集 2020
4. 発表年 2020年

1. 発表者名 榊 剛史, 鳥海 不二夫
2. 発表標題 公開 ソーシャルボルトノ仮説に基づくメディア企業SNSアカウントの扇動性に関する分析
3. 学会等名 人工知能学会全国大会論文集 2020
4. 発表年 2020年

1. 発表者名 榊 剛史, 大坪 雄平, 鳥海 不二夫
2. 発表標題 SNSデータを用いたセキュリティインシデント早期検知に関する実現可能性検証
3. 学会等名 コンピュータセキュリティシンポジウム2019論文集
4. 発表年 2019年

1. 発表者名 榊 剛史, 鳥海 不二夫, 大知 正直
2. 発表標題 ソーシャルメディア上の大規模情報拡散に関する俯瞰的可視化手法の提案
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 小山 耕平、浅谷 公威、榊 剛史、坂田 一郎
2. 発表標題 ネット炎上におけるユーザーの共振構造
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 榊 剛史、鳥海 不二夫
2. 発表標題 ユーザ反応時間に基づくメディア企業SNSアカウントの傾向分析
3. 学会等名 電子情報通信学会 言語理解とコミュニケーション研究会 2019年度2月研究会
4. 発表年 2020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
---------	---------