

# SNS由来の情報拡散型社会問題の発見・分析システムの開発

大知 正直<sup>1,a)</sup> 蕭 喬仁<sup>2</sup> 山中 志一<sup>3</sup> 榎 剛史<sup>4</sup> 森 純一郎<sup>1</sup> 坂田 一郎<sup>1</sup>

**概要:** 本稿では、SNS由来の情報拡散型社会問題の発見・分析システムの開発について説明する。まず、筆者らがこれまで開発してきた SNS 上の投稿内容を投稿者の抱く不平・不満であるかを判定するモデル、投稿者群を興味・関心に応じて分類する手法について説明する。次にそれらを統合し、投稿データから自動的に判定、分析を行い、可視化するシステムについてこれまで開発した内容の説明を行う。最後に特定のケースについて Twitter 上から投稿データを収集し、可視化した結果の有効性の評価を行う。本システムの実用化によって、SNS 上の個人では認識の難しい極めて多様化した社会問題を俯瞰的、網羅的に可視化することが可能になり、(立法によって解決できる議員のような) 解決の可能性のある人々に広く認知してもらい仕組みを構築できる可能性がある。

## 1. はじめに

政府は、技術のイノベーションによって得られた成果を次々に取り入れることによって社会問題を解決し、中長期的に経済的成長を実現する「Society5.0」と呼ばれる改革を行うことを「未来投資戦略2017」としてまとめた。これは、少子高齢化のようなすでに表出している大きな問題をこれまでのシーケンシャルな解決方法の中で情報の収集・統合や分析を効率化することで解決しようとするもので、大きな問題となってから人々への認知が進み、解決へ取り組む、という解決への手順自体は変わらない。一方で、本稿では政府や公的機関が調査・分析結果を発表することで表面化する社会問題（少子高齢化、地球温暖化等）を情報公開型社会問題と分類し、個人の悩み・不満に関する投稿の集積、共感から発現する社会問題（校則厳格化、ハイヒール強要（#KuToo 運動）、育休取得社員への報復人事等）を情報拡散型社会問題と分類し、区別する。近年、この情報拡散型社会問題が人々の生活に大きな影響を及ぼすようになってきており、多様化した産業構造や社会生活のために、大小様々な多数の情報拡散型社会問題が次々と発現するようになってきている [7]。このような背景の中で、

従来どおりに公的機関が解決のためのリソースを負うモデルは、次々と発現する情報拡散型社会問題を解決することには効果的ではない。そして、最近では多様なステークホルダーがビジネスモデルを構築し、社会問題の解決に参画していくニーズが高まっている。特定のコミュニティにおける社会問題について分析する取り組み [6] はこれまで行われてきているが、社会問題を早期に発見し、解決に協力できる技術、ノウハウを持った企業・団体に周知し、解決に向けて協力し新たなイノベーションに繋げる試みは皆無である。

そこで、本研究の最終的な目的は、人々の悩みや共感を社会問題として検出し、可視化、解決へつなげる仕組みを構築することである。急速に高度化・複雑化する現代社会において、人々の悩みも多様化している。そして人々は自身に近い関係者の悩み、社会全体の問題のようなごく一部の問題しか関知できなくなっている。過去に筆者らが行った国会議員に対する調査においても、約 63% の議員が社会問題を認知するための情報源不足を感じているという結果を得ている。申請者はこれまで様々な社会ネットワークデータに対する独自の分析技術を開発してきた経験と共に、SNS には私的な悩みが数多く投稿されていること、それが社会に大きく注目され社会問題化することがあることを見出した。

そこで本稿ではこのような SNS 上の悩みの投稿とそれに対する共感の広がりによって顕在化する問題を「情報拡散型社会問題」として定義し、自動的に検出するモデルを

<sup>1</sup> 東京大学  
University of Tokyo, Bunkyo, Tokyo, 113-8654  
<sup>2</sup> ヨモギテクノロジーズ  
<sup>3</sup> パイオニア  
<sup>4</sup> ホットリンク  
a) masanao.oochi@gmail.com

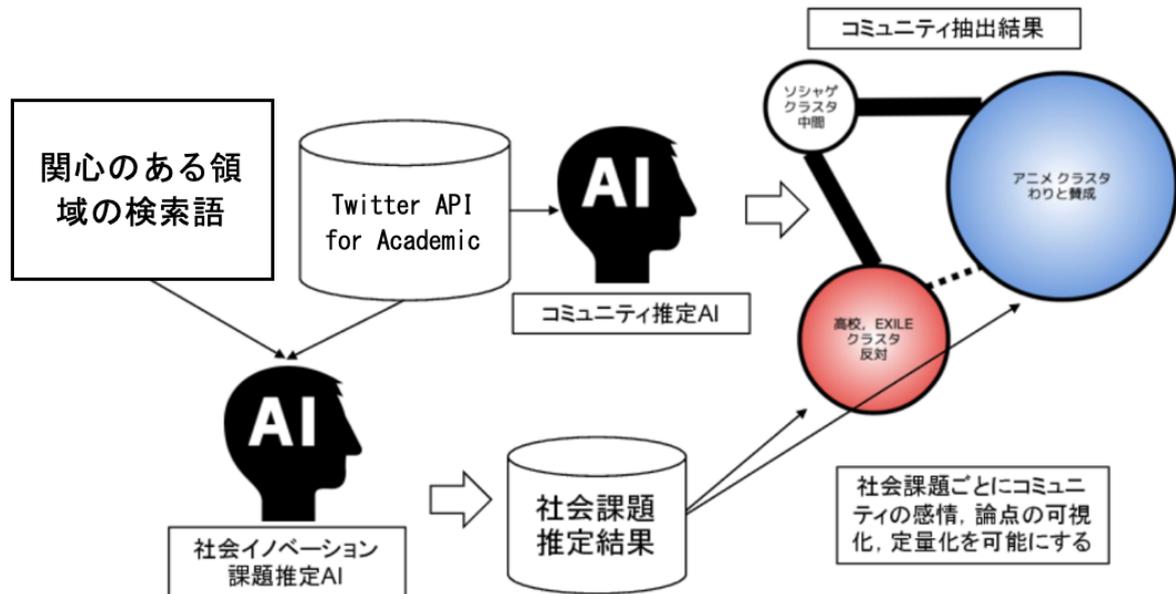


図 1 開発したシステムの処理の概要。

利用し、分析することを可能にする“SNS データを利用した情報拡散型社会問題検出システム”の開発について述べる。本システムは、個人では認識の難しい極めて多様化した社会問題を俯瞰的、網羅的に可視化することで、(立法によって解決できる議員のような)解決の可能性がある人々に広く認知してもらう仕組みを構築する基礎となる可能性がある。

本稿は以下のように構成される。まず、第 2 章で関連する研究について述べる。第 3 章で開発したシステムの説明を行う。その後、本システムを用いた分析結果を第 4 章で示し、議論を行う。最後に refconclusion 章で結論と今後の方向性を示す。

## 2. 関連研究

SNS データの解析により情報拡散型社会問題を早期発見することで問題を重症化させず解決に導くことは可能だろうか？ここで情報拡散型社会問題とは、一般の人々が SNS 等を利用して特定の問題について発信し、次々と共感していくことで拡散する社会問題とする。近年、SNS が爆発的に普及したことで、「アラブの春<sup>\*1</sup>」に代表されるような情報拡散型社会問題が人々の社会生活に大きな影響を及ぼすようになってきている。本研究は、重要な社会問題を検出すること自体を目的とするのではなく、今後拡散され社会問題となり得るような課題を早期に網羅的に発見し、多くのステークホルダーに向けて提示することで同時に並行的に多数の問題への解決へ着手できるようにすることを目

標とする。

社会問題という言葉は多様な種類の問題を含んでおり、その対象とする現象の範囲は幅広い。本研究では、「地球温暖化」、「化石燃料の枯渇」のような全人類共通で専門家が指摘することで発現するような情報公開型社会問題ではなく、私的な悩み・不満が共感・集積し、やがて大きく注目されるようになるような情報拡散型社会問題を検出対象とする。例えば、「高校生の校則厳格化」や「#kutoo (女性への職場でのヒール・パンプスの強制)」が代表例である。これらの問題は、高校生や女性など当事者の人々にとっては重要な問題だが、当事者ではない人々にとっては大々的に報道されることが無ければ知ることもない問題である。そして、学校で頭髪検査を受けたくない、ヒールを履いて仕事をしたくない、という気持ちは個人的な問題として認識されやすく、苦痛を受けている当人も社会でどの程度の人が同じ苦痛を受けているのか、そもそもそれが個人的な問題なのではなく、社会で共有・共感されるべき問題なのか認識することはできない。本研究では、このような情報拡散型社会問題を「構築主義」と呼ばれるアプローチによって検出し、網羅的・俯瞰的に分析・提示することを目指す。構築主義における社会問題とは、「ある状態が存在し、それが問題であると認識する人々による活動」と定義される [5]。そして、ある状態を問題であると認識する人が他の人へ向けに行う活動をクレーム申立て活動として定義し、その活動の発生・集積・持続によって社会問題が発現するとしている。本研究でもこの社会問題に対するアプローチを取り、SNS 上に投稿される私的な悩み・不満をクレーム申立て活動として検出し、その集積・共感の状況を分析することで、情報拡散型の社会問題として検出・分

\*1 “Civil Movements: The Impact of Facebook and Twitter”:Dubai School of Government トピック「アラブの春」とソーシャルメディア:総務省 情報通信白書 平成 24 年度 第一部 第一章第一節



図 2 開発したシステムの画面。

析・可視化することを目指す。

これまで社会問題に関する分析は、個別のケースに関する分析が中心であった [6]。一方でビジョンを共有することで、社会イノベーションを促進する枠組みについて提案・検証する取り組みが報告されている [7]。これら従来の研究から、広く早期に社会問題を検出し、多くのステークホルダーに提示することができれば、社会イノベーションを促進できるだろう。これまで、社会問題の推定には固有表現抽出の手法を用い、新聞記事から推定されていたが [3]、ニューラルネットワークを用いた新たな推定手法を開発した [4]。また、一方で最近ネットワーク上での分散表現抽出を行う深層学習を利用した取り組みが盛んに行われている [1]。開発したモデルを利用し、コミュニティ間の情報伝搬を考慮した分析 [2] を行うことで社会問題を網羅的、俯瞰的に分析・可視化することを可能にするだろう。本研究は、従来の個別の社会問題を分析する研究に対し、新たな問題を検出し数多く提示するという独創性があり、多様なステークホルダーの参加を促進し早期の社会問題解決を行うパラダイムシフトを促す可能性を備えている。

### 3. 情報拡散型社会問題の発見・分析システム

システムの概要を図 1 に示す。まず Twitter API for Academic よりデータを取得するため、分析の対象となる適切なクエリ（検索語、期間、Retweet 等の要不要）を入力する。取得したデータからクレーム投稿の同定と社会問題語の抽出を行うための“社会イノベーション課題 AI”に入力する。クレーム投稿の同定には、Shiau らの手法 [4] を用いる。次に同定されたクレーム投稿から社会問題語の抽出を行う。抽出には、語の頻度、Tweet ごとの平均 TFIDF、投稿に割り振られたクレーム投稿スコアを平均化したものの 3つの方法を用いる。コミュニティ推定 AI は、予め収

集したユーザのプロフィール文をもとにクラスタリングを行なったものである。ここでは同定したクレーム投稿とそれに含まれる語のコミュニティの分布の可視化、計算を行う。具体的にはコミュニティごとの投稿数の分布を計算し、エントロピーの算出を行う。

分析結果を可視化するシステムの画面を図 2 に示す。分析内容は主として 5つある。まず、社会問題の散布図である。これは抽出した社会問題語の投稿数とエントロピーを軸としてプロットしたものである。2つ目はコミュニティごとで投稿されている社会問題の割合を円グラフとして表示する機能である。3つ目は、社会問題語の時系列的な頻度の推移をグラフ化するものである。4つ目は、社会問題語の議論されているコミュニティを可視化するものである。最後は、その社会問題語を投稿している実際の Tweet の内容を表示するものである。

表 1 社会問題語のエントロピーと日経テレコンの検索数の比較。

社会問題語	エントロピー	頻度
国民	2.553	35
発言	2.260	53
野郎	1.926	59
被害者	2.744	31
従業員	2.742	30
税金	2.837	31
政府	2.415	40
苦情	2.521	39
オリンピック	2.506	35
イラ	2.737	31

### 4. システムを用いた分析結果

システムの有効性を評価するため、特定の場合の分

頻度			平均TFIDF		平均クレイム度	
ranking	word	value	word	value	word	value
1	楽天	6704	楽天	50	国民	0.887
2	楽天モバイル	1898	chibalotte	41	発言	0.875
3	楽天カード	426	rakuteneagles	33	野郎	0.869
4	回線	294	sbhawks	27	被害者	0.857
5	ファン	275	楽天モバイル	23	従業員	0.857
6	楽天ペイ	271	lovefighters	23	税金	0.856
7	球団	264	seibulions	22	政府	0.855
8	問題	253	野郎	13	苦情	0.839
9	三木谷	248	hanshin	7	オリンピック	0.837
10	電波	245	楽天カード	7	イラ	0.836

図 3 社会問題語推定手法による抽出語の違い。

析を行う。ここでは検索クエリとして“楽天 lang:ja - is:retweet place.country:JP”と設定した場合の分析結果を示す。Twitter API for Academic からこのクエリのデータを取得すると、169,623 件の投稿が取得できた。期間は 2018 年 1 月 6 日～2022 年 6 月 2 日であった。

クレイム申し立て投稿と判定された投稿から社会問題語を抽出する方法の比較結果を図 3 に示す。

抽出された社会問題語とコミュニティでの議論の集中度を評価するため、投稿数とエントロピーとの比較の結果を表 1 に示す。そして、それぞれの値同士のピアソン相関係数は、 $-0.624^{**}$  であった。 $^{**}$  相関係数は 1% 水準で有意(両側)であることを示す。これらの結果はエントロピーが小さいほど、投稿数が増えることを示している。これは特定のコミュニティにおいて、特定の話題に関する議論が非常に活発になる傾向があり、そうでない話題は広いコミュニティで時々議論されるような状態になっていることを示している。特定のコミュニティでの集中的な議論はエコチェンバー効果と関係があると考えられ、興味深い。

## 5. 結論

本稿では、開発した SNS 由来の情報拡散型社会問題の発見・分析システムの説明を行なった。まず、これまで開発してきた SNS 上の投稿内容を投稿者の抱く不平・不満であるかを判定するモデル、投稿者群を興味・関心に応じて分類する手法について説明した。次にそれらを統合し、投稿データから自動的に判定、分析を行い、可視化するシステムについてこれまで開発した内容の説明を行った。最後に“楽天”のケースについて Twitter 上から投稿データを収集し、可視化した結果の有効性の評価を行った。社会問題語は平均クレイム度の高い語を提示した場合、特定のコミュニティで議論されている語ほど、投稿数が増えるこ

とを示した。本システムの実用化によって、SNS 上の個人では認識の難しい極めて多様化した社会問題を俯瞰的、網羅的に可視化することが可能になり、(立法によって解決できる議員のような)解決の可能性のある人々に広く認知してもらう仕組みを構築できる可能性がある。

**謝辞** 本研究は JSPS 科研費 21K17860 の助成を受けたものです。

## 参考文献

- [1] Kipf, T. N., et al.: Semi-Supervised Classification with Graph Convolutional Networks. *ICLR*, (2016)
- [2] Lin S., Hu Q., et al.: Understanding Community Effects on Information Diffusion. *PAKDD*, (2015)
- [3] Peters, M. E., et al.: Deep contextualized word representations. *NAACL*, p. 2227-2237(2018)
- [4] Chou Jen Shiau, et al.: Constructive Approach for Early Extraction of Viral Spreading Social Issues from Twitter. *Proceeding of the 12th ACM Conference on Web Science(WebSci'20)*, Southampton, United Kingdom, pp. 96-105(2020)
- [5] Spector, Malcolm and Kitsuse, John I, (joint author.): Constructing social problems. *Cummings Pub. Co, Menlo Park, Calif*, (1977)
- [6] Yawar, S. A., et al.: Management of social issues in supply chains: a literature review exploring social issues, actions and performance outcomes. *Journal of Business Ethics*, 141(3), 621-643(2017)
- [7] 岡安 英俊, et al.: ビジョン主導型の社会システム変革に関するプログラム・フレームワークの提案, 国際 P2M 学会誌, 6 巻, 1 号, p.1-13(2011)