

令和 4 年 6 月 17 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：19K20412

研究課題名（和文）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

研究課題名（英文）Development of Fundamental Technology for Early Detection of Social Issues Using SNS Data

研究代表者

大知 正直（Ochi, Masanao）

東京大学・大学院工学系研究科（工学部）・特任研究員

研究者番号：20805527

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：SNSデータの解析により社会課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？本予算では、この困難な研究課題に対して必要な基本技術の開発を行なった。本研究は、社会に潜む多様な声を拾い上げ、多数の人々に届けるしくみづくりのきっかけとなることを目指している。本予算の成果は情報拡散型の社会問題を提案し、人手によるデータ作成を行なった。次に、SNSデータから情報拡散型の社会問題を推定する手法を開発した。これによって、マスメディアに報道されるよりも早く一部の社会課題を検出できることを示した。

研究成果の学術的意義や社会的意義

SNSデータから、クラウドソーシングを利用して比較的安価かつ大量に悩みや社会課題について投稿している内容を判定する手段の開発を行なった。また、収集した投稿を用いて、自動的に情報拡散型の社会課題を見出す方法の開発を行なった。これによって、マスメディアに報道されるよりも早く一部の社会課題を検出できることを示した。本成果によって、社会課題を早期発見し、問題を重症化させず解決に導ける可能性を示した。これによって、社会に潜む多様な声を拾い上げ、多数の人々に届けるしくみづくりの最も基礎的な部分の開発ができた。

研究成果の概要（英文）：Is it possible to detect social problems early by analyzing SNS data, thereby leading to solutions without making the situations more serious? I have used this budget to develop the fundamental technologies necessary to address this problematic research question. This research aims to provide a catalyst for creating a system that picks up the various voices latent in society and delivers them to a large number of people. The outcome of this budget is a proposal for an information-diffusion-type social problem, and I made the dataset using cloud sourcing. Next, we developed a method for estimating information-diffusion-type social problems from SNS data. I showed that this method detected some social issues earlier than the mass media.

研究分野：ネットワーク分析

キーワード：社会課題 社会ネットワーク分析 ネットワーク機械学習

1. 研究開始当初の背景

IMF の報告によるとここ 5 年の日本の経済成長率は約 1%前後で推移しており、停滞している。この要因は様々に議論されているが、本研究では、社会の成長を阻害する要因を“社会課題”と呼ぶ。政府は、この社会課題を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中長期的に成長を実現する“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。しかし、一方でこのような社会課題をいかに早く発見し、解決のための技術イノベーションを起こしていくか、ということについての議論は見られない。

また、一方で大規模なネットワーク分析のための機械学習の発展が挙げられる。特にネットワークの分散表現を獲得する技術は大きな発展を見せており、今後さらに大きな発展が期待されている分野である。申請者自身もこれまでマルチレイヤーネットワークの分散表現獲得に関する研究を進めてきた。こうした分散表現の獲得は、ノード数を N とし、分散表現の次元数を D とした場合、ネットワークの表現を従来の隣接行列による表現 ($N \times N$) から低次元の分散表現行列 ($N \times D$) にすることができる。大規模なネットワークの場合ノード数 N は次元数 D と比較して 10^5 倍以上の違いが想定される。つまり、この技術の進展によって、これまでよりもより大規模なネットワーク分析を行えるようになることが期待できる。

さらに、ソーシャルメディア上のコミュニティ抽出、情報抽出に関する研究もこれまで様々に行われてきた。申請者自身もメディアを介した情報伝搬に関する研究をこれまで行ってきた。こうした技術的背景、蓄積によって SNS データを利用した社会イノベーション課題発見のための人工知能の開発が可能になる。

2. 研究の目的

本研究の目的は、“SNS データを利用した早期の社会イノベーション課題発見のための人工知能開発”である。具体的には大規模に収集した SNS データを用い、ユーザ、メディアを元にネットワークを構成し、それを元にコミュニティ抽出、社会イノベーション課題の発見を行う。これによって、早期の社会イノベーション課題を発見し、提示することを目指す。

3. 研究の方法

本研究では社会課題は特定のコミュニティの中で発見し、時には社会全体で議論される大きなものとなると考える。そこで、課題そのもののいかに検出するか、議論が行われるコミュニティをどのように検出するかこの 2 点についての開発が必要となる。その上で、本研究で開発する手法について概要を図 1 に示す。本研究では、主に 2 つの機能を持つ AI の開発を目指す。

AI1. 社会イノベーション課題推定 AI

AI2. SNS 上のコミュニティの推定 AI

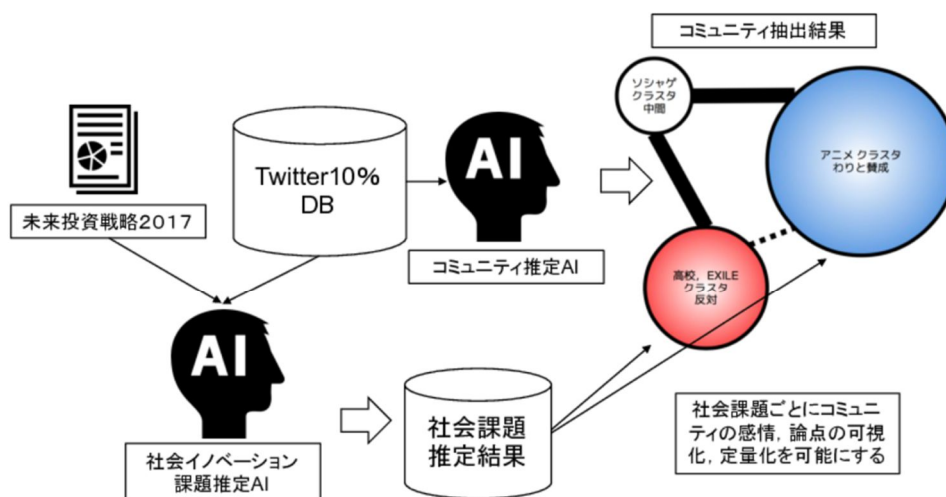


図 1: 開発する SNS データを利用した早期の社会イノベーション課題発見のための人工知能。

まず、AI1 では、2 つのデータを入力とし、社会イノベーション課題の推定のための手法開発を行う。この 2 つのデータは、第一に社会課題と解決を期待される技術イノベーションに関するデータ、第二に SNS の膨大なデータである。第一のデータは、社会イノベーション課題について AI に学習させるための教師データとして利用する。図中では、未来投資戦略 2017 としている

が、他に社会イノベーション課題について述べられている資料があれば積極的に活用していく。例えばそれは、国連開発計画 (UNDP) が策定した SDGs の一部等が挙げられる。第二のデータは SNS の膨大なデータである。図中では比較的入手が容易な Twitter のデータを想定しているが、他にも利用可能なデータがあれば積極的に利用していきたい。このデータを用いて、第一のデータを教師として訓練を行い、新しい社会イノベーション課題を推定するモデル開発を行う。次に、AI2 では、SNS の膨大なデータを入力とし、その SNS 上でのコミュニティ推定をする手法開発を行う。ここでは、人々は SNS 上でのコミュニケーションや接触するメディアによってコミュニティを形成しており、それによって社会イノベーション課題に対する意見も異なる、という仮説をおいている。このようなコミュニティにおける意見の偏りは、エコチェンバー現象として知られている。ここで開発する AI は、このような仮説の元、コミュニティを抽出し、そのコミュニティの特徴(主な関心、年代、性別等のプロフィール情報)を明らかにすることを目的とする。

4. 研究成果

・AI1. 社会イノベーション課題推定 AI に関する成果

まず、構築主義的アプローチによる社会課題推定手法の開発を行なった。概要を図1に示す。社会課題推定は2つのプロセスを経て推定を行う。1つは投稿者本人の悩みに関する投稿であるかどうかを判定するモデル(クレーム申し立て投稿の判定)もう一つは検出した用語が社会課題性を持つ用語であるかどうかを判定するモデルの2つである。

クレーム申し立て投稿の判定モデルの開発は以下のように行なった。はじめに教師データの作成をクラウドソーシングを用いて行なった。クラウドソーシングでは個別の投稿を複数のユーザにクレーム投稿であるかどうか判定をしてもらった。これにより 29,988 件の投稿に関してアノテーションを行うことができた。

次に社会課題の判定モデルは通時的分散表現抽出法を組み合わせた判定モデルの開発を行なった。具体的には通時的分散表現抽出法を用いて抽出した日毎の各用語の分散表現を用い、それと悩みなどのネガティブな用語群との距離変化の測定を行うことで実現した。検出手法の概要を図2に示す。図2では検出したクレーム投稿群に含まれる語 w のある時点での分散表現を著したものである。この語がこの後数日して、「悪い」語との距離が小さくなった場合に社会問題語として検出を行った。結果図3のような語が社会課題語として抽出された。結果として抽出時期にニュースとなっていた語が多くとれ、社会課題語として十分に機能していることがわかった。またこれらの語のうちマスメディアで報道されたものについてはマスメディアが報道するより2週間ほど早く検出できることを示した。

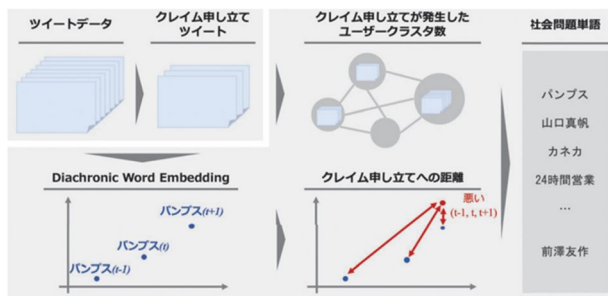


図1 情報拡散型社会課題の推定の概要

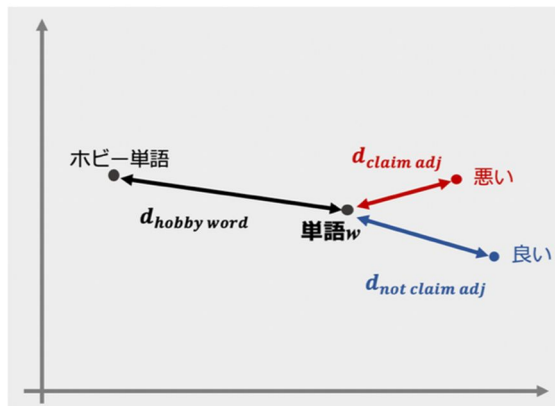


図2 社会課題語の検出法

提案手法
暴風雨, ハンドル,
ソウル市, 区議,
ラマダン, 薄給,
ロボトミー, 景気動向指数,
日本車, 不注意,
被害者遺族, プロレスラー
中国本土, 精神状態,
柴山, 渋沢,
親族, 年寄り
カネカ, 祝辞,

図3 抽出された社会課題語の例

特徴量	M&A 企業の組	非 M&A 企業の組
買収元の媒介中心性	2.84e-2	1.11e-2
買収元の近接中心性	2.10e-1	0.05e-1
買収元の PageRank	1.55e-1	-3.05e-1
買収元の対数(次数)	3.88e-1	-0.43e-1
被買収の媒介中心性	3.81e-2	0.90e-2
被買収の近接中心性	-2.24e-1	-0.43e-1
被買収の PageRank	-2.25e-1	0.31e-1
被買収の対数(次数)	-4.32e-1	0.11e-1

図4 M&A とネットワーク中心性の関係

・ AI2. SNS 上のコミュニティの推定 AI に関する成果

本 AI の構築は、指標や手法の持つ特性を適切に評価するために SNS 以外のデータも利用して開発を進めた。まず、企業間の取引データと M&A の関係について分析を行なった。本分析ではネットワーク特徴量に注目し、M&A とそれぞれのネットワーク特徴量について分析を行なった。この結果、M&A の被買収企業は、次数や PageRank、近接中心性等は買収元企業と比較して小さいが、媒介中心性は高い傾向があることがわかった（図 4）。つまり被買収企業はネットワーク上の直接の繋がりには低い企業が多いが、一方その企業と繋がらなければ、取引関係にならない企業を多く持っている可能性があることがわかった。

次に、学術文献データを用いて、ネットワーク上に存在するさまざまなノードをエンベディングする手法の開発を行なった。図 5 に示す通り、本手法は論文、著者等を異種ネットワークとしてノードとして取り扱い、それぞれを同一空間内でネットワーク上の関係性を保つように配置する手法である。本手法を用いることで、ノード間のリンク関係の予測や将来的な被引用数予測に有効であることを示した。

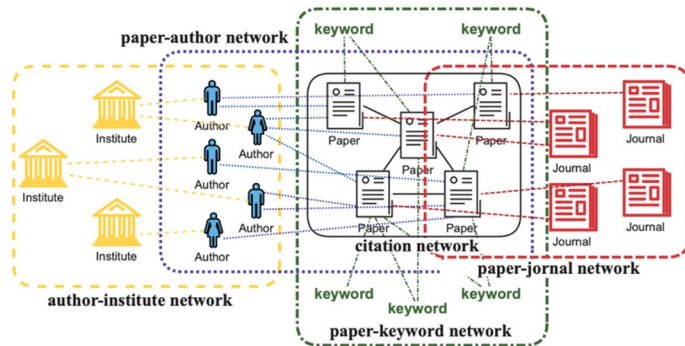


図 5 異種ネットワークエンベディング法

これら 2 つの手法は課題である社会イノベーション課題発見のための人工知能開発において基礎となる技術である。SNS は言うまでもなくネットワークであるため、その中で日々議論される課題、情報拡散の行動もネットワーク的であるからである。

5. 主な発表論文等

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

1. 著者名 大知正直	4. 巻 2021.09月号
2. 論文標題 情報拡散型社会問題の推定 サイレントマジョリティの 意見形成に向けて	5. 発行年 2021年
3. 雑誌名 統計と情報の専門誌「エストレーラ」	6. 最初と最後の頁 8-13
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 無
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 長野 希美,池田 修己,三輪 誠,坂田 一郎,浅谷 公威,大知 正直,市瀬 龍太郎	4. 巻 34
2. 論文標題 人工知能による科学技術研究の加速	5. 発行年 2019年
3. 雑誌名 人工知能：人工知能学会誌	6. 最初と最後の頁 783-789
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 無
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 川村 秀憲,大知 正直,清 雄一,福田 直樹,横山 想一郎	4. 巻 61
2. 論文標題 2050年の情報処理：20. 2050年の知能システム	5. 発行年 2020年
3. 雑誌名 情報処理	6. 最初と最後の頁 482-483
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 無
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5件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 Shiau Chou Jen
2. 発表標題 Constructive Approach for Early Extraction of Viral Spreading Social Issues from Twitter
3. 学会等名 WebSci '20: 12th ACM Conference on Web Science (国際学会)
4. 発表年 2020年

1. 発表者名 蕭 喬仁
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 第196回ICS研究発表会
4. 発表年 2019年

1. 発表者名 大知 正直
2. 発表標題 ネットワーク中心性を用いたM&A類型化への試み
3. 学会等名 第203回ICS研究発表会
4. 発表年 2021年

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 2019年度人工知能学会全国大会（第33回）
4. 発表年 2019年

1. 発表者名 Masanao Ochi
2. 発表標題 Which Is More Helpful in Finding Scientific Papers to Be Top-cited in the Future: Content or Citations? Case Analysis in the Field of Solar Cells 2009.
3. 学会等名 WEBIST 2021（国際学会）
4. 発表年 2021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
---------	---------