

令和 4 年 5 月 27 日現在

機関番号：14603

研究種目：若手研究

研究期間：2019～2021

課題番号：19K20279

研究課題名（和文）健康ソーシャルデータの高信頼化とナビゲーションに関する研究

研究課題名（英文）A Research on High Reliability and Navigation of Health-related Social Media Data

研究代表者

若宮 翔子（Wakamiya, Shoko）

奈良先端科学技術大学院大学・先端科学技術研究科・准教授

研究者番号：60727220

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）：本研究課題では、Twitterのようなソーシャルメディアデータを解析し、信頼性の高い健康ビッグデータを抽出し、疾患やユーザを考慮したデータの可視化や行動指針を提示する情報ナビゲーションの実現を目的として取り組んだ。信頼性の高い健康ビッグデータの抽出のために、ソーシャルボットやフェイクニュースの検出技術の開発を行い、COVID-19のパンデミックをケーススタディとして、ソーシャルメディアデータやウェブ検索クエリデータからの情報分析や可視化に取り組んだ。本研究成果は、健康分野への応用に向けたソーシャル・コンピューティング基盤の構築に向けての一助になると期待される。

研究成果の学術的意義や社会的意義

ソーシャルメディアには、大量のユーザ生成コンテンツが共有・蓄積されており、健康分野における応用に向けた期待は高まっている。ソーシャルメディアデータを健康ビッグデータとして効果的に活用するために、信頼性の高いユーザやコンテンツの解析だけでなく、得られた結果をどのように可視化し、行動させるかについてまでを含めて、実用化レベルでの健康分野への応用を検討した点で学術的意義や社会的意義があると考えている。

研究成果の概要（英文）：In this project, we attempted to analyze social media data such as tweets on Twitter, extract reliable health big data, and visualize the data by considering diseases and users with the aim of realizing information navigation that provides action guidelines. Specifically, to extract reliable health big data, we examined approaches that use multiple media to supplement missing data and a social reporting application. We also developed social bot and fake news detection techniques for extracting reliable social sensor and content. In addition, taking the COVID-19 pandemic as a case study, we analyzed and visualized information from social media data and web search query data.

研究分野：ソーシャルコンピューティング

キーワード：ソーシャルメディア 健康ビッグデータ ソーシャル・コンピューティング

1. 研究開始当初の背景

ソーシャルメディアには、大量のユーザ生成コンテンツが共有・蓄積されており、健康分野における応用に向けた期待は大きい。特に、疾患となるまでの未病の状態の詳細など、医療機関では把握しきれない情報が、ソーシャルメディアには多く投稿されている。オンライン上での情報伝搬や情報監視に関する Infodemiology や Infoveillance と呼ばれる新興分野 [Eysenbach2009] や医療情報学の分野においても、ソーシャルメディアの健康・医療分野への応用に関する研究が進められている。このように、ソーシャルメディアデータはかつてない健康ビッグデータとなりうる。この健康ビッグデータを活用できれば、国内で前例のない感染症の発生・流行の有効な対策となる可能性もあると期待される。これは、東京オリンピックをはじめとする国際的な大規模イベントで渡航者の増加する日本にとって必須の技術となりうる。

しかし、ソーシャルメディア解析には、データや投稿者の信頼性が問題となる。さらに、信頼性の高い健康ビッグデータが得られたとして、それをどのように活用するか、例えば、どのように情報を適切に集約してどのように可視化するかなどといった、行動を提示するビッグデータ・ナビゲーション技術についての知見は、国内外を問わず未だ多いとはいえない。

Eysenbach: Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res. 2009;11(1):e11.



図1 高信頼の花粉症の症状報告を収集するために用いたスマホアプリ

2. 研究の目的

本研究課題は、健康分野への応用に向けたソーシャル・コンピューティング基盤を構築に向けて、ソーシャルメディアデータ解析による信頼性の高い健康ビッグデータ抽出から、疾患やユーザを考慮したデータ可視化や行動指針を提示する情報ナビゲーションまでを実現することを目的としている。

3. 研究の方法

本研究課題では、主に次の3つの研究項目に取り組んだ。

(1) 欠損データを補完する複数メディア統合とソーシャルレポート

信頼性の高い健康ビッグデータの抽出のためには2つのアプローチがある。(A) 複数メディアを用いて欠損データを補完するアプローチと(B) スモールデータであるが、高い信頼性のあるメディアを扱うアプローチである。具体的には、(A) では、Twitter のツイートや Web 上での人々の振舞ログデータ(検索エンジンクエリなど)などを活用する。(B) では、高信頼データ取得のために、人々の自発的な健康データの発信を促進するためのソーシャルレポート基盤を構築する。そのために、ユーザやその家族が疾患の症状の有無をワンクリックで報告すると、位置情報をもとに周辺ユーザの症状の報告を確認できるスマホアプリ(図1)を開発し、これを通して得られるデータを収集する。

(2) 信頼できるソーシャルセンサとコンテンツの抽出

ソーシャルメディアやウェブによって拡散されるフェイクニュースは大きな社会問題となっている。きちんと調べれば明らかにデマであると分かるものもあるが、中には人々が誤った情報を信じて実際に行動してしまう(もしくは行動をやめてしまう)ような、影響力のあるものや深刻度の高いものも存在する。そのために、どのソーシャルメディアユーザ(ソーシャルセンサ)のどのコンテンツがどの時期にどれくらい信頼できるのかを評価することができれば、健康データとしての質は大きく高まる。ソーシャルメディアにおけるソーシャルボットやフェイクニュース検出手法を研究開発し、ソーシャルセンサやコンテンツの信頼性の評価への展開について検討した。

(3) ビッグデータ・ナビゲーション

高い信頼性で情報を得られた場合、次に問題となるのは、どうやって疾患を避けるかといった行動の指針である。これを研究するためには、健康ビッグデータを疾患やユーザに応じて解釈可能な形式に変換するデータ可視化技術だけでなく、行動の指針となるようなナビゲーションの生成技術も必要となる。当初は、疾患が流行している場所や時間を避ける経路の推薦といった実空

間ナビゲーションや、疾患に対する典型的な人々の反応をパターン化して現在のフェーズ(疾患の流行前・流行中・流行収束期などのフェーズ)を示すことで意思決定支援のための情報ナビゲーションなどを開発し、研究期間中盤にあたる東京オリンピック開催期間に合わせ、これらの検証を行い、さらにこの結果や知見をもとに、2021 年度には国際的にも展開させる計画であった。しかし、COVID-19 のパンデミックの影響により、COVID-19 のパンデミックをケーススタディとしたソーシャルメディアデータやウェブ検索クエリデータからの情報分析や可視化に取り組むことに計画を変更し、パンデミック時の情報ナビゲーションについて検討を行った。

4. 研究成果

3 つの研究項目について研究成果をまとめる。

(1) 欠損データを補完する複数メディア統合とソーシャルレポート

信頼性の高い健康ビッグデータの抽出のために、(A) 複数メディアを用いて欠損データを補完するアプローチでは、Twitter に投稿されているツイートや Yahoo JAPAN との共同研究により検索クエリデータを利用した分析を行い、これらの統合について検討した。(B) スモールデータであるが、高い信頼性のあるメディアを扱うアプローチでは、ソーシャルレポートのスマートフォンアプリケーション(図 1)を開発し、データを収集した。しかし、アプリケーションを実際に広めて利用してもらうことは難しく、欠損データを補完するまでのデータを十分に集めることはできなかった。今後の課題として、近年注目を集めているテキストの自動生成などのアプローチによるデータの補完についても検討していく必要がある。

(2) 信頼できるソーシャルセンサとコンテンツの抽出

Twitter に投稿されているツイートを主な対象に、信頼性の高いユーザやコンテンツの抽出について取り組んだ。信頼性の高いユーザの抽出のために、ツイートのテキストに基づく特徴を活用したボット(人間でない)アカウントと人間によるアカウントを分類する手法の研究開発を行った。また、信頼性の高いコンテンツの抽出のために、ツイート内容がフェイクニュース・デマであるか事実であるかの判定や暴言など不適切な発言の検出などについて取り組んだ。これらの研究成果は、PLOS ONE などの英文論文誌に採録された。

(3) ビッグデータ・ナビゲーション

COVID-19 のパンデミックの影響を受け、研究開始当初の計画を変更し、COVID-19 のパンデミックをケーススタディとしたソーシャルメディアデータやウェブ検索クエリデータからの情報分析や可視化に取り組んだ。具体的には、ウェブ検索クエリデータの発信位置の可視化によるクラスタ早期検知についての検討(図 2)や、フェイクニュースやデマに関するツイートの収集や話題の推移を可視化するシステムを構築した。後者の可視化システムでは、デマであると指摘された話題を集約して、定量的・定性的にモニタリング可能とすることで、市民に向けて注意喚起すべきフェイクニュースの判定などの一助にもなると期待される。また、COVID-19 やワクチンに対する市民の反応のように、一般的なアンケートではその本音を把握することが難しい反応を推定しモニタリングするために、検索クエリに基づく指標の提案を行った。さらに、COVID-19 やワクチンに対する市民の反応や影響を把握するために、ツイートを分析し、話題の推移の可視化を行った。

研究期間を通して、健康ビッグデータを疾患やユーザに応じて解釈容易な形式に変換するデータ可視化技術を中心に検討を行ったが、行動の指針となるようなナビゲーションの生成技術については、今後さらに検討していく必要がある。例えば、話題の継続性や深刻度を考慮したフェイクニュースにどのように対処すべきかといったユーザへの行動指針の提供や、従来では人々の反応のモニタリング結果のフィードバックの高度化などについて検討していく必要がある。

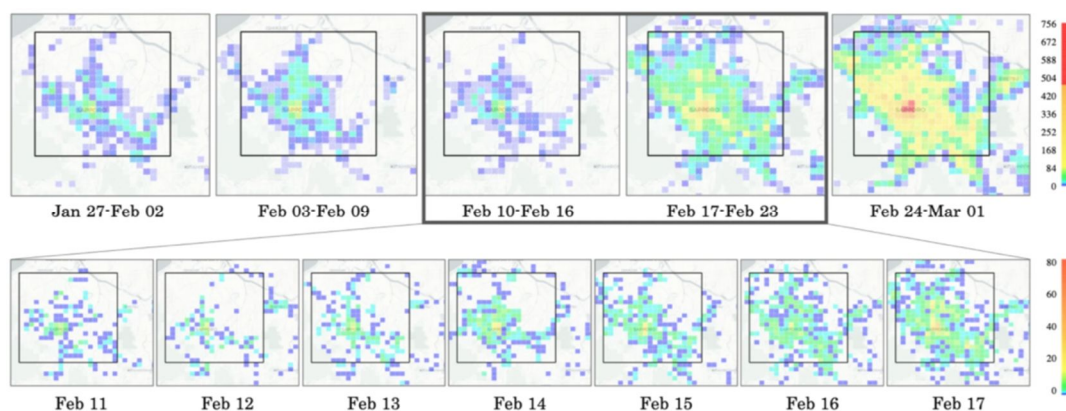


図2 COVID-19のパンデミック初期におけるソーシャルデータ可視化結果

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 4件）

1. 著者名 Hisada Shohei, Murayama Taichi, Tsubouchi Kota, Fujita Sumio, Yada Shuntaro, Wakamiya Shoko, Aramaki Eiji	4. 巻 10
2. 論文標題 Surveillance of early stage COVID-19 clusters using search query logs and mobile device-based location information	5. 発行年 2020年
3. 雑誌名 Scientific Reports	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-020-75771-6	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 Murayama Taichi, Shimizu Nobuyuki, Fujita Sumio, Wakamiya Shoko, Aramaki Eiji	4. 巻 15
2. 論文標題 Robust two-stage influenza prediction model considering regular and irregular trends	5. 発行年 2020年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0233126
掲載論文のDOI (デジタルオブジェクト識別子) 10.1371/journal.pone.0233126	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 Siriaraya Panote, Wang Yuanyuan, Zhang Yihong, Wakamiya Shoko, Jeszenszky Peter, Kawai Yukiko, Jatowt Adam	4. 巻 8
2. 論文標題 Beyond the Shortest Route: A Survey on Quality-Aware Route Navigation for Pedestrians	5. 発行年 2020年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 135569 ~ 135590
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ACCESS.2020.3011924	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 Sumaila Nigo, Shoko Wakamiya, Eiji Aramaki	4. 巻 18
2. 論文標題 Making Twitter Safer: Uncovering Social-Bot on Twitter through User's Content Approach	5. 発行年 2020年
3. 雑誌名 DBSJ Journal	6. 最初と最後の頁 1-7
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 オープンアクセスとしている (また、その予定である)	国際共著 -

[学会発表] 計13件(うち招待講演 1件/うち国際学会 8件)

1. 発表者名 Zhiwei Gao, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki
2. 発表標題 Offensive Language Detection on Video Live Streaming Chat
3. 学会等名 The 28th International Conference on Computational Linguistics (COLING 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Taichi Murayama, Shoko Wakamiya, Eiji Aramaki
2. 発表標題 Fake News Detection using Temporal Features Extracted via Point Process
3. 学会等名 The International Workshop on Cyber Social Threats (CySoc2020) (国際学会)
4. 発表年 2020年

1. 発表者名 村山太一, 若宮翔子, 荒牧英治
2. 発表標題 フェイクニュース検出データセットにおける通時的バイアス
3. 学会等名 言語処理学会第27回年次大会 (NLP2021)
4. 発表年 2021年

1. 発表者名 村山太一, 若宮翔子, 荒牧英治, 小林亮太
2. 発表標題 Twitterにおけるフェイクニュース拡散モデルの提案
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム (DEIM2021)
4. 発表年 2021年

1. 発表者名 村山太一, 若宮翔子, 荒牧英治
2. 発表標題 検索クエリを活用した複数国におけるインフルエンザ流行予測
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム (DEIM2021)
4. 発表年 2021年

1. 発表者名 安西崇, 村山太一, 矢田峻太郎, 若宮翔子, 荒牧英治
2. 発表標題 ならまちハッピーマップ: クラウドソーシングと深層学習による街の雰囲気に基づく経路探索手法の提案
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム (DEIM2021)
4. 発表年 2021年

1. 発表者名 Zhiwei Gao, Kongmeng Liew, Taichi Murayama, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki, Nobuyuki Shimizu, Sumio Fujita
2. 発表標題 Public Concern About COVID-19 Through Search Queries
3. 学会等名 第13回データ工学と情報マネジメントに関するフォーラム (DEIM2021)
4. 発表年 2021年

1. 発表者名 Mir'atul Khusna Mufida, Wenang Anurogo, Muhammad Zainuddin Lubis, Shoko Wakamiya, Eiji Aramaki
2. 発表標題 Disease Outbreaks related to Natural Disaster Analysis based on Twitter Data
3. 学会等名 The 2nd International Conference on Applied Engineering (ICAE 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Wannita Takerngsaksiri, Shoko Wakamiya, Eiji Aramaki
2. 発表標題 City Link: Finding Similar Areas in Two Cities using Twitter Data
3. 学会等名 International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Shoko Wakamiya
2. 発表標題 Towards Industry 4.0: Connecting Industry 4.0 to the information system
3. 学会等名 The International Conference on Applied Engineering (ICAE 2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Shoko Wakamiya
2. 発表標題 Developing Medical/Health Applications Exploiting Social Media Data
3. 学会等名 Korea-Japan Database Workshop 2019 (KJDB2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Patrick Ramos, Kiki Ferawati, Kongmeng Liew, Eiji Aramaki, Shoko Wakamiya
2. 発表標題 Emotion Analysis of Writers and Readers of Japanese Tweets on Vaccinations
3. 学会等名 The 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Shu Anzai, Taichi Murayama, Shuntaro Yada, Shoko Wakamiya, Eiji Aramaki
2. 発表標題 Finding “Retro” Places in Japan: Crowd-sourced Urban Ambience Estimation
3. 学会等名 The 5th ACM SIGSPATIAL International Workshop on Location-Based Recommendations, Geosocial Networks, and Geoadvertising (LocalRec 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