

令和元年6月25日現在

機関番号：12612

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K00149

研究課題名(和文) 情報の内容と空間構造の対応付けを考慮したジオ・ソーシャルメディア活用基盤

研究課題名(英文) Application platform for geo-social media focusing on the mapping of contents of information to the spatial structures

研究代表者

藤田 秀之(Fujita, Hideyuki)

電気通信大学・大学院情報理工学研究科・准教授

研究者番号：90431840

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：GPS等の測位機能で取得した緯度経度を持たず、本文に地名語を含むTwitterのデータを対象として、テキスト中の地名が、投稿者の現在位置(発信地)なのか、そうではないのか(関心地)を自動分類する手法を提案した。緯度経度と地名語の両者を持つデータを対象に、両者の距離に着目した分析を行い、教師データの作成を行った。実データを用いて有効性を評価した。加えて、再現率と精度のトレードオフという一般的な結果に加え、発信地と関心地とでこれらの値が逆転することや、広域スケールでは、発信地と比較して関心地が集中し、詳細スケールでは、関心地と比較して発信地が集中する、といった新たな知見を得た。

研究成果の学術的意義や社会的意義

現在盛んに研究されている、ソーシャルメディアのビッグデータ処理において、データと地理空間の対応付けという極めて基礎的な知見を提供し得るという学術的意義を持つ。また、地理空間座標に基づくジオ・ソーシャルメディアのデータ・コンテンツ処理において、実世界の場所についての情報抽出の実用性を高める一手法としての社会的意義を持つ。

研究成果の概要(英文)：Mobile social media represented by Twitter are expected to be a suitable source of data for analyzing human behaviour and statuses of locations. This research considers in particular problems that concern the relationship between data meaning and their spatial structures. With regard to Twitter, in general, the location from which a tweet is posted is attached to a geotagged tweet. For example, the location coordinates attached to the geotagged tweet "Heavy rain in Miura Peninsula" by NHK are not those of the Miura Peninsula, but of Shibuya in Tokyo (where NHK is located). Therefore, the tweet is not found by a spatial search around the Miura Peninsula or even Kanagawa Prefecture. To resolve such problems, we propose a framework that distinguishes locations of interest and locations of activity. We propose a method for automatically classifying such locations and develop a data collection, classification, and visualization system based on this method.

研究分野：地理情報システム

キーワード：地理情報システム コンテンツの空間構造 ジオコーディング 情報可視化 情報抽出 地図 空間データベース ジオ・ソーシャルメディア

様式 C-19、F-19-1、Z-19、CK-19 (共通)

1. 研究開始当初の背景

写真やメッセージの共有サービス、スマートフォンによるナビゲーションから参加型のセンシングを含め、コミュニケーションや行動支援を目的とする多くのサービスにより、実世界の位置や場所に関するデータ(空間データ)が大量に生み出されている。こうした空間データは、建物や施設、道路等の幾何形状を中心とし、地図描画を目的とする従来の空間データとは異なる価値を持ち、新しいデータ処理手法による、新しいデータの活用が期待されている。

2. 研究の目的

ソーシャルメディアの大規模データの基本的な活用方法のひとつに、データを位置情報で絞り込み、場所に応じた情報提供を行うことがあげられるが、データの内容と空間構造の対応付けに起因する実用上の問題が放置されている。一般に、データの投稿時に付与されるのは、スマートフォンの測位機能などで得られた位置情報であり、「横浜は雨」と渋谷で投稿すると、データは渋谷にマッピングされ、神奈川県を空間検索してもヒットしない。本研究では、発信地・関心地を始めとする空間構造を考慮することでこの問題を解決し、さらに、空間関係に基づく、より高度で実用的な情報抽出を行う手法を提案する。

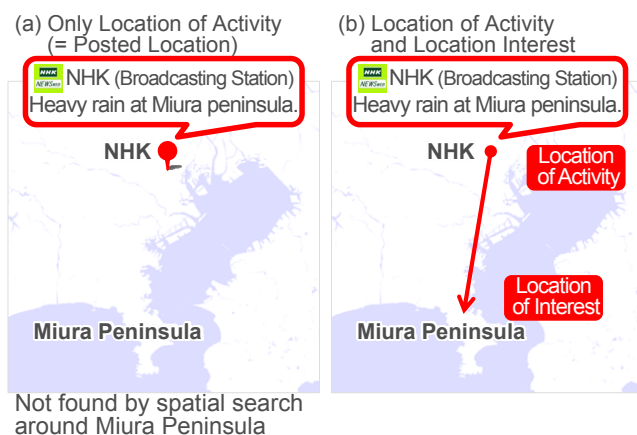


図1. Location of Activity and Location of Interest

3. 研究の方法

主に以下の3つのテーマに分けて研究を遂行した。

(A) Twitter データ収集・可視化システム

位置情報付きのTwitter データを大規模に分散収集するシステムを開発する

(B) 場所情報抽出エンジンの開発

Twitter を対象として、(a) スマートフォン等の測位機能を利用して付与された緯度経度、(b) 本文中の地名、(c) いくつかの場所情報共有サービスとの同時投稿に着目し、これら複数の手掛かりからデータの位置情報を抽出するシステムを開発した。日本語・1週間分のデータ約800万件に適用し、それぞれの手掛かりから抽出される位置情報の特徴をまとめる

(C) 発信地・関心地の自動分類と可視化

GPS等の測位機能で取得した緯度経度を持たず、本文に地名語を含むTwitterのデータを対象として、テキスト中の地名が、投稿者の現在位置(発信地)なのか、そうではないのか(関心地)を自動分類する手法を提案する。緯度経度と地名語の両者を持つデータを対象に、両者の距離に着目した分析を行い、機械学習に用いる教師データの作成を行う。

4. 研究成果

前述の3テーマについて、それぞれ以下の成果があった。

(A) Twitter データ収集・可視化システム

開発したシステムにおいて、同一の期間・領域に対し、一般的な手法の数倍の量のデータ収集に成功した。また収集した大規模なデータを地図上に可視化するシステムを開発した。収集されているデータ量は、1年間で約1億件程度である。事例として、東京都心20km四方、2週間分の約350万件、21万ユーザの日別・時間別のデータ数の変化を地図上に可視化し、時間空間的なデータ量の変化により、実世界における広範囲にわたるイベント(ゲリラ豪雨など)と、局所的なイベント(ある場所でおきたデモ)を区別して可視化できることを示した。

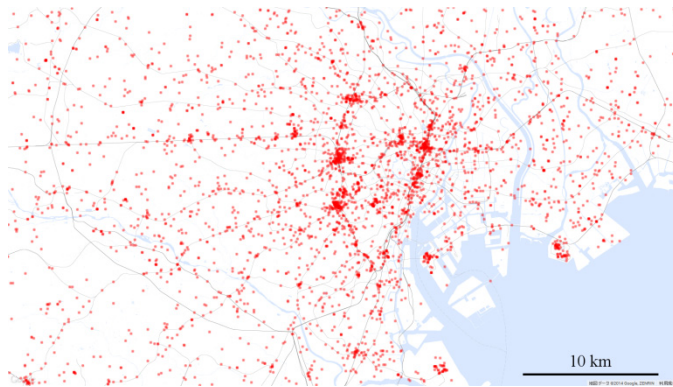
(B) 場所情報抽出エンジンの開発

Twitter を対象として、(a) スマートフォン等の測位機能を利用して付与された緯度経度、(b) 本文中の地名、(c) いくつかの場所情報共有サービスとの同時投稿に着目し、これら複数の手掛かりからデータの位置情報を抽出するシステムを開発した。日本語・1週間分のデータ約800万件に適用し、それぞれの手掛かりから抽出される位置情報の特徴をまとめた。

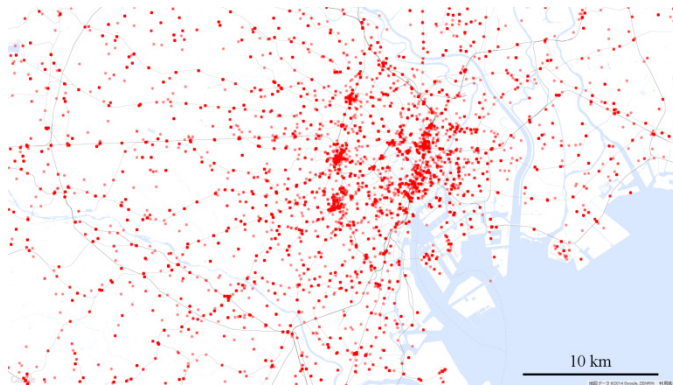
(C) 発信地・関心地の自動分類と可視化

約3万件の実データを用いて提案した自動分類手法の有効性を確認した。加えて、再現率と

精度のトレードオフという一般的な結果に加え、発信地と関心地とでこれらの値が逆転することや、広域スケールでは、発信地と比較して関心地が集中し、詳細スケールでは、関心地と比較して発信地が集中する、といった新たな知見を得た。



(a) LOA tweets



(b) LOI tweets

図2. Tweet distribution (one week, Greater Tokyo Metropolitan Area)

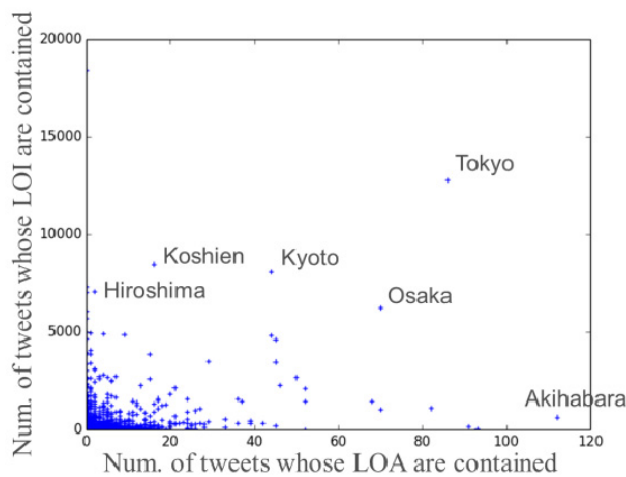


図3. Distribution of grid cells by number of LOI tweets and number of LOA tweets

5. 主な発表論文等

〔雑誌論文〕 (計6件)

- ① 査読有り, [Hideyuki Fujita](#), Shota Sagara, Tadashi Ohmori, Takahiko Shintani, Finding hotspots by location-direction-enabled photographs focusing on intersection of lines of sight, Proceedings of the International Cartographic Association, vol. 2, 5pages, 2019 (発表予定)
- ② 査読有り, [Hideyuki Fujita](#), Shota Sagara, Tadashi Ohmori, Takahiko Shintani, Route Network Construction with Location-Direction-Enabled Photographs, Proceedings of the 28th International Cartographic Conference (ICC 2017), 4pages, 2017
- ③ 査読有り, Yuichi S. Hayakawa, [Hideyuki Fujita](#), Sohee Lee, Takeshi Sagara, Developing a data-sharing system for geospatial research: A case study on the Joint Research

Assist System (JoRAS), International Journal of Spatial Data Infrastructures Research, Vol.12, pp.141-160, 2017

- ④ 査読有り, 藤田秀之, 早川裕弼, 李召熙, 相良毅, 空間データ基盤を活用する共同研究支援システムの構築, GIS-理論と応用, Vol.25, No.2, pp.7-14, 2017
- ⑤ 査読有り, Hideyuki Fujita, Information Extraction and Visualization from Twitter Considering Spatial Structure, International Journal for Geographic Information and Geovisualization (Cartographica), Vol.52, No.2, pp.178-193, 2017
- ⑥ 査読有り, Hideyuki Fujita, Information Extraction from Twitter Considering Spatial Structure, Proceedings of the 27th International Cartographic Conference (ICC 2015), 15pages, 2015

〔学会発表〕(計5件)

- ① 査読なし, 秋庭武, 藤田秀之, 大森匡, 新谷隆彦, 語の一般性と空間領域との関係に着目した位置情報付き SNS からの地域特徴語抽出, 情報処理学会 第 81 回全国大会講演論文集, 6P-04, 2019, 2019/3/16
- ② 査読なし, 藤田秀之, 相楽翔太, 大森匡, 新谷隆彦, 位置・方向情報付き写真群によるホットスポット抽出のための視線のクラスタリング手法, 地理情報システム学会 第 26 回研究大会 講演論文集, 2017, 2017/10/28
- ③ 査読なし, 相楽翔太, 藤田秀之, 大森匡, 新谷隆彦, 位置・方向情報付写真群を用いた空間を要約する経路の生成, 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM 2016) 論文集, H5-4, 8pages, 2016
- ④ 査読なし, 相楽翔太, 藤田秀之, 大森匡, 新谷隆彦, 位置・方向情報付写真群を用いた Data storytelling に向けた Storyline 構成手法, 第 14 回情報科学技術フォーラム (FIT 2015), D-029, 4pages, 2015
- ⑤ 査読なし, 藤田秀之, 地図とインフォグラフィックス: データストーリーテリング, 日本地図学会 平成 27 年度定期大会発表論文・資料集, Ss-1, 2015

〔図書〕(計0件)

〔産業財産権〕

- 出願状況 (計0件)
- 取得状況 (計0件)

6. 研究組織

- (1) 研究分担者
なし
- (2) 研究協力者
なし

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。