

令和元年6月5日現在

機関番号：14603

研究種目：若手研究(B)

研究期間：2016～2018

課題番号：16K16057

研究課題名(和文) 時空間コンテキストを考慮した大規模ソーシャルメディアデータ解析に関する研究

研究課題名(英文) Research on Large Scale Social Media Data Analysis Considering Spatio-Temporal Context

研究代表者

若宮 翔子(WAKAMIYA, Shoko)

奈良先端科学技術大学院大学・研究推進機構・特任助教

研究者番号：60727220

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)：本研究課題では、代表的なソーシャルメディアの一つであるTwitterのツイートを対象に、[A] 時空間・意味解析による大規模ソーシャルメディアデータ分析基盤の構築と[B] ソーシャルメディアデータの時空間コンテキストを考慮した地域情報システムの開発・評価について取り組んだ。具体的には、(1) 位置情報が欠落したソーシャルメディアデータの位置推定手法の開発、(2) ソーシャルメディアデータを活用したランドマーク検出と経路推薦システムの開発、そして(3) ソーシャルメディアデータを活用した地域の対応付け手法の開発を通して、ソーシャルメディアデータの地域情報システムへの応用の有用性を示した。

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データは、多様なコンテキストにおける多様な人々の意思決定や行動支援のための重要な情報源になっている。社会学や医療・健康といった多様な研究分野においても、その活用は期待されている。一方で、データ欠損の影響が大きく、いまだ十分に活用されていない。さらに、ユーザらがコミュニケーションの一端として発信するデータは表現の揺らぎが大きく曖昧性が高い。本研究課題では、ソーシャルメディアデータの欠点であるデータの信頼性や曖昧性を分析・解消する方法について研究し一定の成果を挙げた。さらに、時空間コンテキストを考慮した地域情報システムの開発し、その有用性を示した点で意義がある。

研究成果の概要(英文)：This research project utilizing tweets over Twitter, which represents social media services, attempted that [A] the construction of a framework for analyzing large scale social media data by considering spatio-temporal and semantic features and [b] the development and evaluation of regional information systems by considering spatio-temporal contexts of social media. We showed the applicability of social media data for regional information systems through the following three themes: (1) to develop a tweet location estimation method, (2) to create a landmark-based route suggestion system using social media data and geographic data, and (3) to determine city area relationships in terms of human activities.

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 位置ベースSNS 時空間データ 地域情報 ランドマーク抽出 位置推定 地域特徴 ナビゲーション

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

急速に普及・発達しているスマートフォンのような携帯端末を用いて、大量のユーザ発信コンテンツ(マイクロブログ、画像、動画など)がソーシャルメディア上(Twitter, Foursquare, Flickr, Instagram など)でリアルタイムに共有されている。これらのデータにはGPSベースの位置情報をはじめ様々なメタデータが付与されており、実世界における群衆の経験を観察可能である。ソーシャルメディアにおけるユーザをセンサとみなし、ユーザらが発信するデータを群衆センシングデータとして分析・活用したイベント検出が盛んに研究(文献)されている。このことから、ソーシャルメディアデータの活用・分析は、多様なコンテキストにおける多様な人々の意思決定や行動支援のための重要な研究分野の一つとなっている。

このようにソーシャルメディアデータの利用可能性は指摘されている一方で、データ欠損の影響が大きくいまだ十分に活用されていない。さらに、ユーザらがコミュニケーションの一環として発信するデータは表現の揺らぎが大きく曖昧性が高い。そのため、群衆センシングデータの時空間コンテキストを考慮した意味解析を行い、高い信頼性でのデータ抽出や補完が重要である。このようなソーシャルメディアのビッグデータ分析は、当該研究分野だけでなく、社会学や医療(文献)といった多様な研究分野からも注目されており重要で意義がある。

<引用文献>

Atefeh et al.: A Survey of Techniques for Event Detection in Twitter, Comput. Intell. (2015)

Paul et al.: Twitter Improves Influenza Forecasting, PLOS Currents Outbreaks (2014)

2. 研究の目的

本研究課題は、ソーシャルメディア上で共有・蓄積されている大量のユーザ発信コンテンツを活用するために、ユーザ発信コンテンツの欠点であるデータの信頼性や曖昧性を分析・解消する方法について研究する。さらに、ソーシャルメディアデータを活用し、地域情報案内などの実用的なシステムを構築することを目指す。研究開始前までに、群衆センシングデータのGPS位置情報に基づく群衆移動の定量的分析による地域特徴抽出に関する研究、地域間の関連性分析に関する研究やGPS位置情報とツイートで言及された場所との差異に基づく群衆の集合的空間関心の可視化に関する研究などに取り組んできた。本研究課題では、これらを発展させ、時空間・意味解析による大規模ソーシャルメディアデータ分析基盤を構築し、ソーシャルメディアデータの時空間コンテキストを考慮した地域情報システムの開発・評価に取り組んだ。

3. 研究の方法

本研究課題では、代表的なソーシャルメディアの一つであるTwitterのツイート(以降、発言とも記す)を対象に、[a] 時空間・意味解析による大規模ソーシャルメディアデータ分析基盤の構築と[b] ソーシャルメディアデータの時空間コンテキストを考慮した地域情報システムの開発・評価について取り組んだ。3つの具体的な研究内容を示す。

(1) ソーシャルメディアデータの位置推定

本研究では、位置情報が欠損しているソーシャルメディアデータに位置情報を補完するために、自然言語処理により発言位置(以降、単に位置と記す)を推定する手法について研究した。特に、位置匿名化の観点から、発言内容から特定の地点を推定できる(全く位置匿名化がなされていない)レベル、発言内容から特定の地点は推定できないが複数の地点あるいは特定のエリアを推定できる(やや位置匿名化がなされている)レベル、発言内容から位置を全く推定できない(位置匿名化がなされている)レベルといった、位置推定の実現可能性のレベルについて検討した。具体的には、位置情報付きツイートに対してロジスティック回帰を適用し、位置推定モデルを構築した。さらに、機械と人間のそれぞれについても異なる位置推定の実現可能性のレベルが存在すると考え、それぞれに対応した位置推定の分類器を構築した。このため、ツイート内容から発言位置を推定できるかを人間が2値分類するクラウドソーシングを行ない、閾値以上の参加者が推定可能と回答したものを位置推定の実現可能性のあるツイート、それ以外を実現可能性のないツイートに分類し、それぞれの特徴を考察した。また、ツイートを入力すると、地図上に推定位置をその実現可能性レベルとともに可視化するアプリケーションを開発した。

(2) ソーシャルメディアデータを活用したランドマーク検出と経路推薦

本研究では、TwitterやFlickrなどのSNSデータと地理データを組み合わせ活用し、ナビゲーションなどの際に有用となるランドマークを発見する手法について研究した。具体的には、点・線・面の3種類のランドマークを定義し、これらのランドマークからなる経路を検索するシステムを開発し評価を行なった。このために、SNSデータと地理データから、場所ごとに3つの指標:訪問人気度(人々の訪問ログから場所の人気を示す尺度)、間接的な可視率(場所を直接視認することができなくとも、場所の方角の認識可能性を示す尺度)、直接的な可視率(場所を直接視認することができる視認可能エリアの大きさを示す尺度)を求め、点・線・面のランドマークを抽出した。次に、点・線・面のランドマークを用いて、実際の道路ネットワークとは独立した新たな経路グラフを生成した。最後に、2地点が与えられると、より経路長が短く、利用

するランドマークの数がより少ない経路を経路グラフ上で探索するアルゴリズムを実装した。提案手法を評価するために、実在する都市における SNS データと地理データ(実験データ 1(サンフランシスコ市) と実験データ 2(鹿児島市)) を用いた提案手法による経路と、既存手法(直接的な可視率のみ) による経路、ならびに、既存の道路ネットワーク上での Google Directions による経路を被験者実験により比較した。具体的には、各経路を仮想空間 (Google Street View) で辿るシミュレーション評価と実空間で辿る現地評価を行い、出発地から目的地までの経過時間、自己位置の参照回数、提示された案内経路の参照回数を評価項目として比較した。

(3) ソーシャルメディアデータを活用した地域間の対応付け

本研究では、Twitter のツイート进行分析することで、異なる 2 つの都市における人々の話題や行動の類似性に基づき、類似機能を持つ地域を発見し、対応付ける手法について研究する。これにより、例えば、「バンコクにおける、京都の河原町のような場所」を検索したり、「京都がバンコクだったら」という観点での観光情報マップを作成したりすることが可能になる。具体的には、2 つの都市の位置情報付きツイートを収集して共通の言語 (英語など) に翻訳し、グリッド分割したエリア内のツイートを 1 つの文書とみなしてベクトル化 (Doc2Vec) した。この文書ベクトルを用いて、それぞれの都市内で類似度が高い隣接グリッドエリアをマージしたマージエリアを設定し、2 つの都市間で類似度が高いマージエリアを対応付けた。さらに、地図上で対応関係を可視化した。京都市とタイのバンコク市を対象に行なったケーススタディでは、それぞれの都市のランドマークベースの対応関係情報をクラウドソーシングにより収集し、提案手法により対応付けられたエリアの妥当性を考察するために用いた。

4. 研究成果

3. 研究方法に挙げた 3 つの具体的な研究内容について、それぞれの研究成果を示す。

(1) ソーシャルメディアデータの位置推定

京都市の中心エリアを対象としたケーススタディにより、位置推定の実現可能性のレベルによって異なる特徴があることが示された。例えば、店の営業に関する発言など、同一の発言が 1 箇所から投稿されることから、ツイート内容によって位置推定が可能となる。一方、特定のエリアに偏在していない普通名詞が店の名前として用いられている場合など、ツイート内容によって位置推定の実現可能性レベルが低くなる。さらに、単語の組み合わせにより位置の推定が可能であるという仮説に基づく位置匿名化アルゴリズムへの応用 (例えば、「京都駅」は「京都」と「駅」という単語の組み合わせであり、いずれかを削除することにより、位置推定の実現可能性のレベルが大きく下がる) についても検討し、結果の一部について目的通りの匿名化処理が可能であることが確認された。

今後の展望としては、各ツイートの文脈を考慮することや、時間帯や季節などを考慮したモデルの構築が挙げられる。

本研究内容に関連する成果は、雑誌論文 [\[1\]](#)、学会発表 [\[2\]](#) にて発表した。

(2) ソーシャルメディアデータを活用したランドマーク検出と経路推薦

提案手法を実装し、出発地と目的地を指定すると経路を探索するシステムを開発した。評価実験の結果、提案手法により、自己位置や案内経路を頻繁に参照することなく、かつ、短時間で目的地に到達可能な、記憶に残る経路推薦を実現できることを確認した。このことから、二輪車による移動や歩行中、GPS を利用できない状況、道に迷ってしまったときや災害などで目的地に向かっている間に自己位置や案内地図を確認することが困難な場合などに特に有効であると考えられる。

今後の展望としては、建物密度や地理的形状が異なる (あるいは類似する) 複数の都市を対象を拡張してデータを取得・整備し、提案手法の妥当性を検討する。さらに、提案手法により抽出されるランドマークを経路推薦以外のアプリケーションの活用についても検討することが望まれる。

本研究内容の成果は、雑誌論文 [\[3\]](#)、学会発表 [\[4\]](#) ¹² にて発表した。本研究で用いた要素技術は、雑誌論文 [\[5\]](#)、学会発表 [\[6\]](#) ¹³ にて発表し、これを医療・健康分野に応用した成果は、雑誌論文 [\[7\]](#)、学会発表 [\[8\]](#) ¹¹ にて発表した。

(3) ソーシャルメディアデータを活用した地域間の対応付け

京都市とタイのバンコク市を対象に行なったケーススタディにより、提案手法により、ランドマークベースで妥当な対応付けがなされることを確認した。例として、京都御所と Grand Palace や京都大学と Chulalongkorn University の対応などが挙げられる。

今後の展望としては、複数の都市を対象を拡張してデータを取得・整備することである。また、複数の機能を持つ (人々の行動に多様性がある) 地域を考慮することや、時間帯や季節などを考慮した対応付けについて検討する。さらに、抽出した対応付け関係を活用したアプリケーションについても検討することが望まれる。

本研究内容の成果は、学会発表 [\[9\]](#) にて発表した。また、本研究内容に関連する技術は、学会発表 [\[10\]](#) で発表した。

5 . 主な発表論文等

〔雑誌論文〕(計 5 件)

Shoko Wakamiya, Yukiko Kawai, Eiji Aramaki, Twitter-Based Influenza Detection After Flu Peak via Tweets With Indirect Information: Text Mining Study, JMIR Public Health and Surveillance, 査読有, Vol. 4, No. 3, e65, 2018,

DOI: 10.2196/publichealth.8627

Hayate Iso, Shoko Wakamiya, Eiji Aramaki, Conditional Density Estimation of Tweet Location: a Feature-dependent Approach, MEDINFO 2017: eHealth-enabled Health, 査読有, Volume 245, pp. 408-411, 2017,

DOI: 10.3233/978-1-61499-830-3-408

若宮 翔子, 森永寛紀, 岡山愛, 脇海晟, Adam JATOWT, 河合由起子, 秋山豊和, 川崎洋, LandmarkNavi: マイクロブログを用いた効果的なランドマーク発見, 日本データベース学会和文論文誌, 査読有, Vol.15-J, Article No.1, pp. 1-8, 2017,

http://dbsj.org/journal/dbsj_journal_j/dbsj_journal_vol_15_01/

若宮 翔子, ヤトフト アダム, 河合 由起子, 秋山 豊和, 荒牧 英治, Twitter における空間差異に基づく群衆の関心分析のための可視化システム構築, 情報処理学会論文誌 データベース (TOD), 査読有, Vol. 9, No. 3, pp. 14-24, 2016,

<http://id.nii.ac.jp/1001/00174765/>

森永 寛紀, 若宮 翔子, 谷山 友規, 赤木 康宏, 小野 智司, 河合 由起子, 川崎 洋, 点と線と面のランドマークによる道に迷いにくいナビゲーション・システムとその評価, 情報処理学会論文誌, 査読有, Vol. 57, No. 4, pp. 1227-1238, 2016,

<http://id.nii.ac.jp/1001/00159024/>

〔学会発表〕(計 13 件)

Wannita Takerngsaksiri, Shoko Wakamiya, Eiji Aramaki, City Link: Finding Similar Areas in Two Cities using Twitter Data, In Proc. of International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2019), LNCS Vol. 11474, pp. 13-27, 2019.

Nigo Sumaila, Shoko Wakamiya, Eiji Aramaki, User's Content-Oriented Social-Bot Discovery on Twitter, DEIM Forum 2019, P1-139, 2019.

Shoko Wakamiya, Panote Siriaraya, Yihong Zhang, Yukiko Kawai, Eiji Aramaki, Adam Jatowt, Pleasant Route Suggestion based on Color and Object Rates. In Proc. of the 12th ACM International Conference on Web Search and Data Mining (WSDM '19), pp. 786-789, 2019.

Yihong Zhang, Panote Siriaraya, Yuanyuan Wang, Shoko Wakamiya, Yukiko Kawai, Adam Jatowt, Walking down a Different Path: Route Recommendation based on Visual and Facility based Diversity. In Companion Proc. of the Web Conference 2018 (WWW '18), pp. 171-174, 2018.

村山 太一, 若宮 翔子, 荒牧 英治, WORD GINI: 使用頻度の偏りを捉える指標の提案とその応用, 言語処理学会 第 24 回年次大会, 2018.

村山 太一, 若宮 翔子, 荒牧 英治, ソーシャルメディアを用いた依存症者の発言分類とその空間分析, 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017),

2017.

田口 勝弥, 若宮 翔子, 荒牧 英治, 位置推定とその実現可能性を考慮した新しい匿名化の提案, 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017), 2017.

井上 翔, 河合 由起子, 秋山 豊和, 川崎 洋, 若宮 翔子, 荒牧 英治, 安井 豪基: 地下街歩行者ナビのための人流情報提示システムの提案, 情報処理学会 第 79 回全国大会論文集, 5K-09, 2017.

Hayate Iso, Shoko Wakamiya, Eiji Aramaki, Conditional density estimation for the origin of social media geolocation, 第 9 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 Forum 2017) 論文集, H8-4, 2017.

柴田 大作, 若宮 翔子, 宮部 真衣, 大西 正輝, 山下 倫央, 野田 五十樹, 荒牧 英治, Twitter による群衆密度の推定 - 第 29 回関門海峡花火大会での実証実験 - , 第 9 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 Forum 2017) 論文集, D6-3, 2017.

Shoko Wakamiya, Yukiko Kawai, Eiji Aramaki, After the Boom No One Tweets: Microblog-based Influenza Detection Incorporating Indirect Information, In Proc. of the 6th International Conference on Emerging Databases (EDB 2016), pp. 98-106, 2016.

Shoko Wakamiya, Hiroshi Kawasaki, Yukiko Kawai, Adam Jatowt, Eiji Aramaki, Toyokazu Akiyama, Lets Not Stare at Smartphones while Walking: Memorable Route Recommendation by Detecting Effective Landmarks, In Proc. of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2016), pp. 1136-1146, 2016.

Shoko Wakamiya, Adam Jatowt, Yukiko Kawai, Toyokazu Akiyama, Analyzing Global and Pairwise Collective Spatial Attention for Geo-social Event Detection in Microblogs, In Proc. of the 25th International Conference Companion on World Wide Web (WWW '16 Companion), pp. 263-266, 2016.

[その他]

奈良先端科学技術大学院大学 研究推進機構 ソーシャル・コンピューティング研究室 個人ホームページ

http://sociocom.jp/~wakamiya/study_Japanese.html

本研究課題に関連して企画・主催した国際ワークショップ User Interfaces for Spatial and Temporal Data Analysis (UISTDA)

UISTDA 2018: <http://sociocom.jp/~event/uistda2018/>

UISTDA2019: <http://sociocom.jp/~event/uistda2019/>

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。