

令和元年6月4日現在

機関番号：12601

研究種目：基盤研究(B) (一般)

研究期間：2016～2018

課題番号：16H02905

研究課題名(和文) 実世界情報を用いたマルチソーシャルメディア解析の高度化

研究課題名(英文) Enhancing Multi Social Media Analysis Utilizing Real World Data

研究代表者

豊田 正史 (Toyoda, Masashi)

東京大学・生産技術研究所・教授

研究者番号：60447349

交付決定額(研究期間全体)：(直接経費) 13,700,000円

研究成果の概要(和文)：本研究の目的はソーシャルメディアとセンサデータを相互に解析および可視化することにより、新たな実世界応用アプリケーションを創出することにある。ソーシャルメディアからの実世界情報抽出のための、新規固有表現発見手法の高度化、人々の行動に影響を与える情報の検索手法の提案を行うとともに、ソーシャルメディアおよびセンサデータの統合可視化システムの構築を行った。台風や豪雨等の災害を事例とした実証実験により手法の有効性を確認した。

研究成果の学術的意義や社会的意義

本研究において提案した手法により、センサデータから得られる豪雨等の実世界イベントに対するソーシャルメディア上の言及を統合的に可視化することが可能となった。提案した手法は、災害時には被害状況の把握や復旧支援等に役立つとともに、平時においても人々の移動を支援するナビゲーションシステムにおいて大きなリスクを避けるために必要な情報を提供できるものであり、社会的な意義は大きい。

研究成果の概要(英文)：This research aims to develop new real world applications by mutually analyzing and visualizing social media and sensor data. For extracting real world information from social media, we improved our emerging entity discovery methods, and proposed a method for retrieving information that affect people's behavior. Based on these methods, we constructed a visualization system for integrating social media and sensor data. The effectiveness of our methods was confirmed by case studies on disastrous situations such as typhoons and heavy rains.

研究分野：情報学

キーワード：ソーシャルウェブ 実世界情報

1. 研究開始当初の背景

World Wide Web (以下 Web と略記) におけるソーシャルメディアのユーザは、Statista 社の調査によると 2012 年に 14 億人を超え、2015 年には Facebook のみでも月間アクティブユーザ数が 14 億人を超えるに至っている。特に、ユーザが友人との情報共有を行えるソーシャルネットワークサービスや、友人間で短いメッセージを公開・共有できるマイクロブログサービスなどの隆盛が顕著である。ソーシャルメディアの利用時間は既に検索エンジンや電子メールサービスの利用時間を上回っており、日常的に利用される主要な情報源の一つとなっている。一方で、米国 Cisco 社の調査によるとインターネットに接続されたデバイスの数は 2010 年以前に既に世界人口を超えており、今後のデジタルデータの増加は、IoT (Internet of Things) に拠るところが大きいと見込まれている。スマートフォン等のモバイルデバイスのみならず、自動車に設置されたカーナビやドライブレコーダ、工場や発電所等の機器に設置されたセンサなどを活用した様々なソリューションが模索されているところである。

人とモノから創出される所謂ビッグデータは、これまでは困難であったリアルタイムかつ詳細な社会事象観測を可能にするドライバとなり得る。ソーシャルメディアからは人が目撃した事件や事故や災害等のイベント情報が得られ、気象・交通等のセンサからは、気温、降雨、渋滞等の環境情報が得られる。これらを統合することにより、個々のデータのみでは識別不可能だった事象を明らかにできる可能性がある。ソーシャルメディアの解析は従来からその信憑性に対する疑問が付き物であったが、センサデータからはその証拠となる情報を取得できる。ソーシャルメディアとセンサデータを相互に解析しあうことにより、こうしたイベント解析の精度および信頼性向上、さらには新たな付帯状況の獲得等が期待できる。

2. 研究の目的

本研究は、ソーシャルメディアとセンサデータを相互に解析および可視化することにより、新たな実世界応用アプリケーションを創出することを目的として、以下のテーマを実施する。

- (1) ソーシャルメディアからの実世界イベントに関する情報収集手法の研究開発
- (2) ソーシャルメディアおよび実世界に設置されたセンサから得られた情報を可視化し、インタラクティブに探索可能とするシステムの構築

3. 研究の方法

主な分析対象として、これまでに収集してきたマイクロブログ (Twitter) のデータを用いた。Twitter に関しては、延べ 200 万以上のユーザの発言を、2011 年より継続して約 500 億程度収集しており、幅広い実世界イベントに関する言及を分析するのに十分なデータが整っている。この大規模な Twitter データから、実世界におけるユーザの行動、新たに出現した固有表現等のイベントに関連する情報収集を行う手法の研究開発を実施した。

実世界センサデータとしては、主に XRAIN 等の雨量データを用いた。上記のソーシャルメディアからの実世界イベント情報収集手法を用いて、豪雨等の実世界イベントに関する Twitter ユーザからの情報を収集し、雨量データと統合可視化するシステムの構築を行った。

4. 研究成果

(1) ソーシャルメディアからの実世界イベントに関する情報収集手法

現実世界では日々新しい事物や概念が誕生しており、それらの情報を網羅したデータベースは文書理解や情報検索、ウェブの監視など様々な場面で必要不可欠なものとなる。しかし、これら全てを手で発見し補完することはもはや不可能であるため、計算機により自動でかつできる限り迅速に収集することが望ましい。これまで我々は、事物のデータベース構築の支援のため、リアルタイムに次々と新しい情報が発信される Twitter を対象とし、そこから新しい事物や概念を指す新固有表現を精度良く発見する手法を提案してきた。本研究においては、この新固有表現発見手法の精度向上を行い、大規模な Twitter データセットを用いた実証実験により出現頻度の少ない新固有表現であっても幅広く獲得可能であることを示した (雑誌論文)。

マイクロブログに投稿される画像には、実世界で起きているイベントに関するものも多く含まれているため、与えられたイベントに関連する画像を、指定された文脈に依じて効率的に獲得する手法も提案した (学会発表)。本手法は、画像の特徴とつづやきに含まれる言語特徴を考慮し、指定された文脈に依じて画像獲得の際に用いる特徴を調整する。本手法を用いると、例えば熊本地震といった災害イベントに対して、被害の状況を写した写真、被災者の役に立つ画像などの文脈に応じた画像を獲得することが可能となる。

さらにソーシャルメディア自体が実世界に与える影響を分析する研究も行った。ソーシャルメディア上での他者との対話や投稿の閲覧は、オンラインだけでなく実世界にも影響を与えるため、どのような情報が実世界での人々の行動を変化させ得るかを分析することが実世界イベントの分析において重要となる。実世界での主要な人間行動の1つひとつである「購買」を対象として、ソーシャルメディアから人々の購買行動の選択に影響を与える投稿を検索する手法を提案した(雑誌論文)。提案手法は、行動の種類を指定する所与のクエリに対して、クエリに関連する行動の実行に影響を与えたソーシャルメディア上の投稿集合を検出する。次に、得られた投稿集合を適合性・影響力・網羅性の観点からランキングすることで行動選択の判断に有用なものを上位に配置する。1年間のTwitterデータを用いた評価実験によって、提案手法が影響力のある多様な投稿集合を検索可能であることを確認した。

(2) ソーシャルメディアおよび実世界センサ情報の統合可視化システム

ソーシャルメディアから得られる実世界イベント情報を用いて、日々発生する実世界イベントの影響や時空間的な広がりを視覚的に閲覧可能とする3次元可視化手法の研究開発を行った(雑誌論文)。Twitterのストリーム中の位置参照表現に着目し、位置座標情報が付加されていないつぶやきについても、地名や施設名などの位置参照表現に基づいて位置に関連付け、局所的なイベント、広範囲イベントを認識して3次元空間に地理空間ワードクラウドとして多層的に可視化し、その時間変化をアニメーション可能とする可視化システムを実現した。図1(上)はJRの運行に大きな影響を与えた火災発生時の状況を可視化した事例を本システムで可視化したものであり、線路付近で発生した火災が新幹線の運行に影響を与えたことが地図上にオーバーレイされたキーワードから読み取れる。また、図1(下)は、東京マラソンを対象として可視化を行ったものであり、マラソンに関して言及される場所が、スタート地点からゴール地点へと移動していく様子を見て取ることができる。

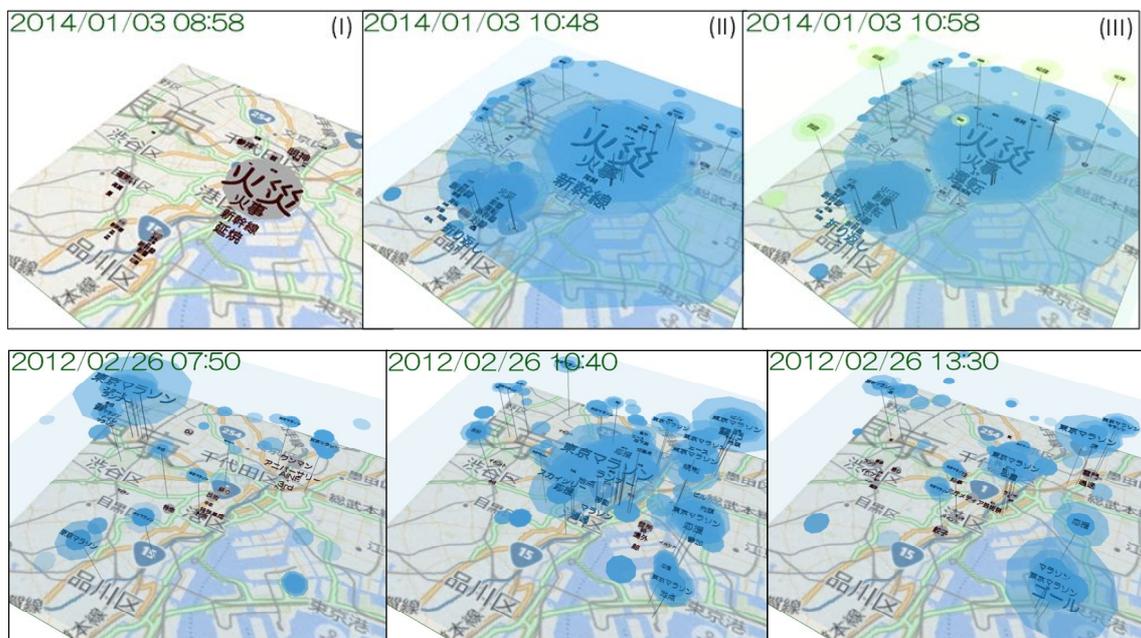


図1 時空間イベントの3次元統合可視化手法

本手法に基づき、火災やゲリラ豪雨などの災害及び事故発生時の状況可視化を目的として、雨量情報や物理センサと連携可能にする統合フレームワークを提案した(学会発表)。物理世界とバーチャル世界を相補的に探索することで、個別のセンサデータからでは分からない、イベントの要因、関係性などを発見することが可能となる。情報通信研究機構ビッグデータ活用研究室と連携し、ソーシャルメディアおよび実世界データからゲリラ豪雨の発生に伴う交通や人々の反応の変化を抽出及び可視化することで豪雨リスクをよりの確に把握するための技術を開発した(学会発表)。ソーシャルメディアから得られた豪雨による影響情報の地理空間ワードクラウドと、PANDAレーダから得た豪雨の警戒円及びXRAINから得た実際の降雨状況の可視化を統合可視化し、実際に台風やゲリラ豪雨が発生した場所日時を事例として実証実験を行った。

5 . 主な発表論文等

〔雑誌論文〕(計 14 件)

Satoshi Akasaki, Naoki Yoshinaga, Masashi Toyoda, Early Discovery of Emerging Entities in Microblogs. Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI2019), 査読有, 採録決定, 2019.

Taro Takaguchi, Takanori Maehara, Ken-ichi Kawarabayashi, Masashi Toyoda, Existence of outsiders as a characteristic of online communication networks. Network Science, 査読有, Vol. 6, pp. 431-447, 2018.

DOI: <https://doi.org/10.1017/nws.2018.17>

梅本和俊, 豊田正史, 実世界での行動に影響を与える情報のソーシャルメディアからの発見. 日本データベース学会和文論文誌, 査読有, Vol. 16-J, pp 15:1-15:8, 2018.

Hongshan Jin, Masashi Toyoda, Naoki Yoshinaga, Can Cross-Lingual Information Cascades Be Predicted on Twitter? Social Informatics. SocInfo 2017. Lecture Notes in Computer Science, 査読有, Vol. 10539, pp. 457-472, 2017.

DOI: https://doi.org/10.1007/978-3-319-67217-5_28

豊田正史, 大規模な階層的グラフのインタラクティブ可視化システム及びそのWebメディア分析への応用. 可視化情報学会誌, 査読無(招待), Vol. 36, No. 141, pp. 63-67, 2016.

川本貴史, 豊田正史, 吉永直樹, マイクロブログからの社会的影響力を持つ情報カスケードの検知手法. 情報処理学会論文誌データベース(TOD), 査読有, Vol. 9, No. 2, pp. 23-33, 2016.

石渡祥之佑, 鍛冶伸裕, 吉永直樹, 豊田正史, 喜連川優, 文脈語間の対訳関係を用いた単語の意味ベクトルの翻訳. 人工知能学会誌, 査読有, Vol. 31, No. 6, pp. AI30-A, 1-10, 2016.

DOI: <https://doi.org/10.1527/tjsai.AI30-A>

Ryosuke Nishi, Taro Takaguchi, Keigo Oka, Takanori Maehara, Masashi Toyoda, Ken-ichi Kawarabayashi, Naoki Masuda, Reply trees in Twitter: data analysis and branching process models. Social Network Analysis and Mining, 査読有, Vol. 6, No. 1, pp. 26:1-26:13, 2016.

DOI: <https://doi.org/10.1007/s13278-016-0334-0>

Tatsuya Iwanari, Naoki Yoshinaga, Nobuhiro Kaji, Toshiharu, Nishina, Masashi Toyoda, Masaru Kitsuregawa, Ordering Concepts Based on Common Attribute Intensity, Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI-16), 査読有, pp. 3747-3753, 2016.

Masahiko Itoh, Naoki Yoshinaga, Masashi Toyoda, Word-Clouds in the Sky: Multi-layer Spatio-temporal Event Visualization from a Geo-parsed Microblog Stream. Proceedings of the 20th International Conference on Information Visualization (IV2016), 査読有, pp. 282-289, 2016.

DOI: <https://doi.org/10.1109/IV.2016.62>

〔学会発表〕(計 36 件)

Masahiko Itoh, Takeshi Sagara, Ukyo Suzuki, Kazutoshi Umemoto, Naoki Yoshinaga, Masashi Toyoda, Koji Zettsu, Yutaka Kidawara, Information Integrated Visualization System for Heavy Rainfall Risk Analysis. The 2018 IEEE Pacific Visualization Symposium (PacificVis 2018), Poster, 2018.

Hongshan Jin, Detection and Characterization of Influential Cross-lingual Information Diffusion on Social Networks. The 23rd international conference on World Wide Web (WWW2017), PhD Symposium, 2017.

Shoetsu Sato, Naoki Yoshinaga, Masashi Toyoda, Masaru Kitsuregawa, Modeling Situations in Neural Chat Bots, Annual Meeting of the Association for Computational Linguistics (ACL 2017), Student Research Workshop, 2017.

小泉実加, 吉永直樹, 豊田正史, マイクロブログにおける画像に関する投稿文脈の獲得. 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(DEIM2017), 2017.

Masahiko Itoh, Masashi Toyoda, EventStacks: Integration of Event Visualizations for Physical and Social Sensor Data. IEEE VIS 2016 Workshop on The Event Event: Temporal & Sequential Event Analysis, 2016.

6 . 研究組織

(1)研究協力者

研究協力者氏名：吉永 直樹
ローマ字氏名： (Naoki Yoshinaga)

研究協力者氏名：伊藤 正彦
ローマ字氏名： (Masahiko Itoh)

研究協力者氏名：梅本 和俊
ローマ字氏名： (Kazutoshi Umemoto)

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。