

位置情報に関連する投稿を可視化するつぶやきヒートマップ

末吉 佑一^{†1,a)} 岩井 将行^{†2,†1,1,b)}

概要: Twitter を代表とするリアルタイムストリーミングを特徴とする web サービスでは、単語から位置情報を推測、あるいは直接位置情報を埋め込んでいる投稿が存在する。これらの情報を収集し、第三者のユーザがある一定の範囲、又はキーワードを指定してリアルタイムでイベント検知する web サービスを構築した。以下のリンクでつぶやきヒートマップのソースを公開している。 <https://github.com/SUXEEE/JPTweetD3> またこれら以外にもユーザ参加型センシングやセンサ群から得られた情報を、包括的に web 上で表示する為に soxfire を使用しダッシュボードの構築も行う。

Proposal and Implementation of tweet heat map to visualize the post related to the position information

YUUCHI SUEYOSHI^{†1,a)} MASAYUKI IWAI^{†2,†1,1,b)}

Abstract: In the web services that features a real-time streaming to a representative of the Twitter, I guess the position information from the word, or posts that are embedded directly position information is present. Collect this information, we constructed a web service that events sensed in real time by specifying a certain range to the third party's user, or keywords. It has been opened to the public the source of the tweet heat map in the following link. <https://github.com/SUXEEE/JPTweetD3> In addition also performs construction of the dashboard using soxfire to display the information obtained from the user participatory sensing and sensor group in addition to these, comprehensively on the web.

1. はじめに

IoT デバイスや SNS データのリアルタイムストリーミングは現在耳目を集めている分野である。そしてその膨大なデータを扱う際に分散処理や機械学習といった技術を用いて効率的に収集する手法が代表的である。本論文では SNS の一つである Twitter を対象として、これらの手法を用い、位置情報に関する投稿をリアルタイムで抽出する。投稿の本文に対しては形態素解析を行い、その中でも頻出する場所・名詞などを地図上に表示して、イベント検知を目的とした web サービスを構築した。

1.1 関連研究

Song ら [1] の研究ではリアルタイムでの Twitter のトレンドマイニングを行い社会問題やその変化を分析しているが、投稿の付加情報として位置情報を可視化などはしていない。しかし社会問題と地理的な要因は切り離すことは出来ないとも考える。Kuna ら [2] の研究では Android 端末を用いてセンサデータを収集し XMPP サーバへその情報を格納し、環境モニタリングを行っている。スマートフォンや、そのサーバとの通信に XMPP を用いている点では共通しているが、それを web 上で視覚的に確認する方法は提供していない。他の通信プロトコルの CoAP は無線環境での通信や、輻輳のおこりやすいネットワークでのメリットがある。

Shriramoju[3], Hunkeler[4] らは XMPP とは別のプロトコルである MQTT や CoAP を用いたスマートフォンをセンサノードとして用いた情報収集システムの構築や実験を行っている。Davis ら [5] の研究では Pub/Sub 通信プロト

¹ 情報処理学会

IPSJ, Chiyoda, Tokyo 101-0062, Japan

^{†1} 現在, 東京電機大学大学院

Presently with Graduate School of Tokyo Denki University

^{†2} 現在, 東京電機大学

Presently with Tokyo Denki University

a) sue@cps.im.dendai.ac.jp

b) iwai@im.dendai.ac.jp

コルの比較として MQTT-S と CoAP を上げておりパケット損失や信頼性を担保するシステムの構築の際に一助となる研究を行っているが、XMPP については考慮していない。

Tang ら [6] の研究ではスマートフォンでユーザ参加型センシングを行う際のエネルギー効率をテーマにしている。サーバサイドやクライアントでの省エネルギーアルゴリズムの実装に機械学習を用いて最小化を目指しており、本研究への一助となった。

Meng ら [7] は分散処理のプラットフォームである Apache Spark で使用できる機械学習のライブラリについて述べている。本研究では SNS である Twitter のリアルタイムストリーミング解析に Apache Spark を用いており、ツイートごとの重み付けなどに、述べられている MLlib を使用した。

Yasumoto[8] らの研究では IoT センシングでいずれ頭打ちになると予測されているリアルタイムでビッグデータ処理に対する新しいフレームワークを提案している。リアルタイムストリーミングを SNS 解析で行っている本研究も分散処理のストリームキューレーションを将来的には応用する必要がある。

Ueda[9] らは高齢者介護などに応用が期待されている監視や生活認識に関する研究を行っている。これは私の研究に比べてプライバシーへの配慮が高く、またその為に信頼性が低くなってしまいうデータについても独自の機械学習から得られたアルゴリズムを用いることでそれを高めている。

2. システム構成



図 1 システム概要

2.1 Apache Spark

Twitter のようなリアルタイムで大規模データを処理する際に Apache Spark の分散処理機能を利用した。この時にメモリ上でストリームからデータを取得、RDD(Resilient Distributed Dataset) に格納し Scala で kuromoji の形態素解析や、単位時間あたりでの県ごとの頻出地名・単語をカウントした。

2.2 D3.js

D3.js とはデータに基づいて図形やグラフ・ドキュメントを JavaScript で表現出来るライブラリである。今回 Twitter からリアルタイムストリーミングで位置情報と単語を取得し、ブラウザ上の日本地図に表示するというデータ駆動型 DOM と視覚化をこれによって実現した。

2.3 Soxfire

慶應義塾大学徳田研究室が公開している、Soxfire と Java で通信を行う Java SOX Library を使用した。この Library を利用することで Android スマートフォンから Soxfire サーバへの通信が簡便化された。XMPP は他の通信プロトコル (MQTT, HTTP+JSON) などと比較してグループを構築しやすく、メッセージフォーマットが厳格であるため採用した。Soxfire で収集したデータを管理するメリットとして、他者の Soxfire と連結した際に容易にデータの Pub/Sub が可能である。

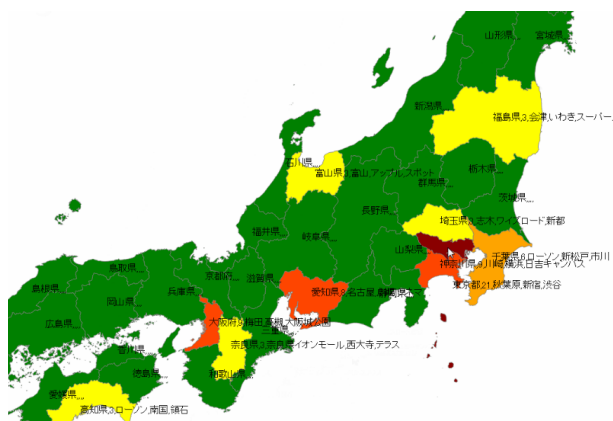


図 2 動作中のつばやきヒートマップ

3. おわりに

謝辞 本研究は NICT ソーシャル・ビッグデータ活用・基盤技術の研究開発「オープン・スマートシティを実現するソーシャル・ビッグデータ活用・還流基盤」の支援および、SIP (戦略的イノベーション創造プログラム) 課題「レジリエントな防災・減災機能の強化」,(地域連携による地域災害対応アプリケーション技術の研究開発, 巨大都市・大規模ターミナル駅周辺地域における複合災害への対応支援アプリケーションの開発) の支援を受けている。慶應義塾大学 SFC 特任講師米澤拓郎氏, 慶應義塾大学 SFC 研究員伊藤友隆氏, 慶應義塾大学徳田・高汐・中澤研究室の皆様には本研究において Soxfire と Sensor-Over-XMPP ライブラリ (Sox ライブラリ) のご提供と多大な技術的なご支援と研究に関するアドバイスをいただきました。ここに深く感謝いたします。

参考文献

- [1] Song, Min, and Meen Chul Kim. "RT²M: Real-Time Twitter Trend Mining System." *Social Intelligence and Technology (SOCIETY)*, 2013 International Conference on. IEEE, 2013.
- [2] Kuna, Matko, et al. "Android/OSGi-based Machine-to-Machine context-aware system." *Telecommunications (ConTEL)*, Proceedings of the 2011 11th International Conference on. IEEE, 2011: 95-102.
- [3] Shriramoju, Suman Kumar, Jagadeeshwar Madiraju, and Adloori Ramesh Babu. "AN APPROACH TOWARDS PUBLISH/SUBSCRIBE SYSTEM FOR WIRELESS NETWORKS." *IJCER* 2.4 (2013): 502-508.
- [4] Hunkeler, Urs, Hong Linh Truong, and Andy Stanford-Clark. "MQTT-S—A publish/subscribe protocol for Wireless Sensor Networks." *Communication systems software and middleware and workshops*, 2008. comsware 2008. 3rd international conference on. IEEE, 2008: 791-798.
- [5] Davis, Ernesto Garca, Anna Calveras, and Ilker Demirkol. "Improving packet delivery performance of publish/subscribe protocols in wireless sensor networks." *Sensors* 13.1 (2013): 648-680.
- [6] Tang, Zaiyang, et al. "Energy-Efficient Transmission Scheduling in Mobile Phones using Machine Learning and Participatory Sensing." *Vehicular Technology, IEEE Transactions on* 64.7 (2015): 3167-3176.
- [7] Meng, Xiangrui, et al. "Mllib: Machine learning in apache spark." *arXiv preprint arXiv:1505.06807* (2015): 1-7.
- [8] Yasumoto Keiichi, Hirozumi Yamaguchi, and Hiroshi Shigeno. "Survey of Real-time Processing Technologies of IoT Data Streams." *Journal of Information Processing* 24.2 (2016): 195-202.
- [9] Ueda Kenki, et al. "Exploring Accuracy-Cost Trade-off in In-Home Living Activity Recognition based on Power Consumptions and User Positions." *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM)*, 2015 IEEE International Conference on. IEEE, 2015: 1130-1137.