

投稿場所に注目したソーシャルメディア上の情報拡散の可視化と分析 Visualization and Analysis of Information Diffusion Focusing on Posted Place in Social Media

落合 涼¹

Ryo Ochiai

伊與田 光宏¹

Mitsuhiro Iyoda

1. はじめに

インターネットを介したコミュニケーションが盛んとなる中で、Twitter や Facebook をはじめとするソーシャルメディアが急成長している。中でも Twitter は 140 文字以内のテキストと画像や動画を投稿するという手軽さから、日本国内の月間アクティブユーザ数 (MAU) が 4,000 万を超え[1], 若者を中心に広く普及している。

ソーシャルメディアはリアルタイムなコミュニケーションツールとして、人々は日常的に起きる出来事を場所や時間に捕われず発信し共有しあっている。それゆえに、ソーシャルメディアに発信された大量の情報を集合知と捉えることで、世の中で起きている出来事や問題を抽出することができること期待され、様々な分野でソーシャルメディアに対する研究が行われている。

ソーシャルメディアが普及し、誰でも簡単に情報が発信できる時代では、発信した情報が自らが想像しないような規模で拡散されることがある。その中でも地震や豪雨といった自然災害や事故、電車の運転見合わせや遅延など投稿された場所が重要視される情報が他の場所にいるユーザの元へどのように分析することで、その出来事の重大性を測ることができると推測できる。

2. 目的

本研究では代表的なソーシャルメディアの 1 つである Twitter を対象とし、投稿場所に注目してソーシャルメディアに投稿される情報がどのようにして拡散され、他の場所にいる人々に伝わるのかを時系列ごとに可視化して分析する。自然災害や事件などの出来事が発生してからの情報の伝播を長期的に観測することで、出来事が発生してからの時間軸ごとの情報伝播のスピード、出来事の種類の重なりや重大性によるソーシャルメディアの情報伝達能力の差異を評価する。

3. 関連研究

ソーシャルメディアやソーシャルネットワーキングサービス (SNS) を対象として世の中で起きている出来事を把握しようとしている研究は盛んに行われている。Sakaki ら[2]は Twitter に投稿される地震や台風などの自然災害を位置情報付きツイートを観測することでリアルタイムに発生した災害を検出するシステムを開発した。

また岩渕ら[3]は Twitter に投稿されたイベントについて言及しているツイートを元に電車の混雑を予測するシステムを構築した。

本研究ではツイートの投稿場所を元にして情報の伝播の様子を地図上にヒートマップにして可視化することで、ソーシャルメディアに発信された情報がどのように他の場所にいるユーザの元へ届くのかを分析する手法を提案する。

4. 分析手法

Twitter に投稿されたツイートを収集・分析するために、図 1 のようなシステムを構築する。本システムは大きく 5 つの処理に分かれており、以下にそれぞれの処理の概要を示す。

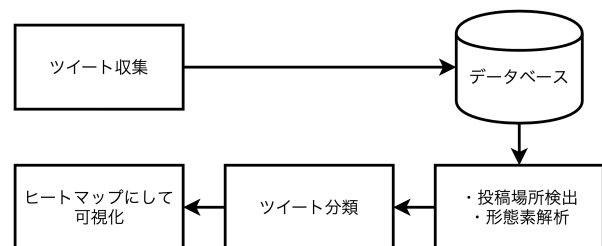


図 1 システム構成図

4.1. ツイート収集

Twitter からツイートを収集するために、Twitter 社が提供している API の 1 つである Streaming API[4] を利用する。Streaming API は Twitter に投稿されたツイートをリアルタイムに取得することができる API である。本研究は Streaming API の中の Sample メソッドと Filter メソッドを用いる。Sample メソッドは投稿された全てのツイートの一部を取得することができる。また Filter メソッドを予め用意したフィルタに合致したツイートを取得することができる。Filter メソッドで用いるフィルタは後述するツイートを分類する際のグループに関連した単語を著者の主観で選び使用する。

本研究では日本からのツイートのみを対象とするため、取得したツイートから日本語のツイートのみをフィルタし、データベースに格納する。ツイートデータのうちデータベースに格納する情報は、ツイート ID、ツイート本文、位置情報、ユーザ情報、投稿時間の 5 つである。

¹ 千葉工業大学大学院 Chiba Institute of Technology

4.2. 投稿場所抽出

ツイートに含まれる位置情報を投稿場所として使用する。しかし、日本において位置情報を付与してツイートをしているユーザは非常に少なく、橋本ら[5]の研究によると日本から発信されるツイートのうち約 0.18% しか位置情報が付与されているツイートが存在しないと分析している。

本研究ではツイートに含まれる位置情報以外から投稿場所を抽出するために、ユーザ情報に含まれる『場所』の情報を元に投稿場所の推定を行う。多くのユーザはこの『場所』に自分が住んでいる地域や地名を設定しているため、これをツイートの投稿場所であると、ジオコーディングすることで位置情報を取得する。

4.3. 形態素解析

ツイートがどのような内容のものであるのかを判定するためにツイート本文に対して形態素解析を行い、名詞や動詞など品詞ごとに文章を分割する。

本研究では形態素解析器に Java 製の Kuromoji[6] を使用し、形態素解析で用いるシステム辞書に mecab-ipadic-NEologd[7] を使用する。辞書で用いる mecab-ipadic-NEologd は Web 上の大量のデータから得られた新語や固有表現を追加することでカスタマイズされた mecab 用の辞書で、標準の IPA 辞書ではうまく形態素解析することができない新語や固有表現が多く含まれる文章に対して形態素解析をするのに有効であるとされている。Twitter にはユーザの日常的な会話に近い文章や流行に沿った内容の文章が投稿されるため、新語や固有表現が多く含まれると予想できる。したがって mecab-ipadic-NEologd を使用してより最適な形態素解析を行う。

4.4. ツイート分類

ツイートがどのような内容なのかを単純ベイズ分類器を用いて分類する。ツイートの特徴を抽出するために前節で行った形態素解析の結果から名詞と動詞の原型を抜き出し、BoW (Bag of Words) の要領で単語の集合で表し、以下のグループに分類する。

- 地震
- 豪雨、強風などの地震以外の自然災害
- 交通機関の運転見合わせや遅延
- 事件や事故

名詞の他に動詞の原型を抜き出す理由として、地震などが発生した時に発信されるツイートに「揺れた」など動詞のみ投稿されるケースが多く存在するためである。

予め用意したツイートを手作業で上記のグループに分類し、学習を行った分類器を使用して実際に収集したツイートを分類する。分類を行うことでそれぞれの種類のツイートの情報拡散の違いについて観測することが可能となる。

4.5. 可視化

投稿場所を抽出し、内容ごとに分類されたツイートを地図上にヒートマップにして可視化する。本研究では、上記の処理を行った後のツイートを、全文検索エンジンである Elasticsearch[8] に蓄積する。Elasticsearch に蓄積されたデータを高速にフィルタリングでき、特に時系列データの取扱いが強い分析フレームワークである Kibana を用いて蓄積されたデータ可視化を行う。Kibana を用いて時系列ごとに可視化を行うことで情報拡散の様子を観測することが可能となる。

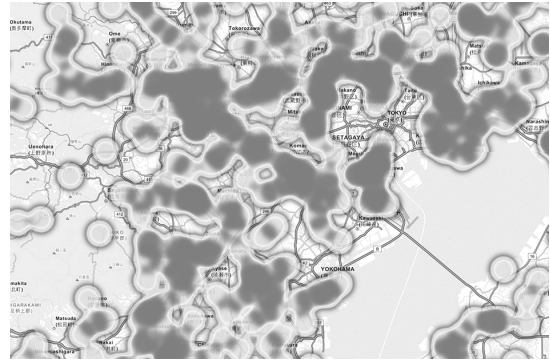


図2 ヒートマップによる可視化の例

5. まとめ

本研究では代表的なソーシャルメディアの1つである Twitter を対象に、投稿場所に注目して情報がどのように拡散されていくのかを可視化し分析する方法を提案した。今後は実際に収集されたツイートに対して処理を行うための実装を行い、投稿場所に注目した情報伝播の可視化・分析を進めていく予定である。

参考文献

- [1] Twitter Japan (@TwitterJP), <https://twitter.com/TwitterJP/status/793649186935742465> (2016).
- [2] Takeshi Sasaki, Makoto Okazaki and Yutaka Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors", WWW2010, pp.851-860 (2010).
- [3] 岩淵 和紀, 萩原 威志, "Twitter でのイベント検出を用いた電車混雑の予測システム", FIT2016 第15回情報科学技術フォーラム (2016).
- [4] Twitter Developers, <https://dev.twitter.com/> (2017).
- [5] 橋本 康弘, 岡 瑞起, "都市におけるジオタグ付きツイートの統計", 人工知能学会誌 27 (4), pp.424-431 (2012).
- [6] Kuromoji, <https://www.atilika.com/ja/products/kuromoji.html> (2017).
- [7] GitHub neologd/mecab-ipadic-neologd, <https://github.com/neologd/mecab-ipadic-neologd> (2017).
- [8] Elasticsearch, <https://www.elastic.co/jp/products/elasticsearch> (2017).
- [9] Kibana, <https://www.elastic.co/jp/products/kibana> (2017).