

投稿場所に着目したソーシャルメディア上の情報拡散の分析

落合 涼†

伊與田 光宏‡

千葉工業大学大学院情報科学研究科

1. はじめに

コンピュータやスマートフォンなどのデジタルデバイスが普及し、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っている。それに伴って Twitter や Facebook をはじめとするソーシャルメディアが急成長を遂げている。中でも Twitter はツイートと呼ばれる半角 240 文字以内のテキストと画像や動画を投稿するだけと言った手軽さから、日本国内での月間アクティブユーザ数 (MAU: Monthly Active Users) が 4,000 万を超え[1], 若者を中心に広く普及している。

Twitter はソーシャルメディアの中でも特にリアルタイムなコミュニケーションツールとして、ユーザは日常的に起きる出来事を時間や場所に関係なく 1 日に何度も発信し他のユーザと共有しあっている。それゆえにソーシャルメディアに投稿される大量の情報は世の中で起きていることの集合知として捉えることができ、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と期待されている。

ソーシャルメディアが普及し、誰でも簡単に情報が発信し共有できる時代では、発信された情報が自らが想像もしないような規模で拡散されることがある。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の場所が含まれている情報は時間が経過すると共に広い地域に拡散されていくことが示されている[2]。この情報拡散の規模と拡散速度を分析することで、出来事の重大性を測ることができるのではないかと推測することができる。

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、投稿場所に着目して投稿された情報がどのように他の地域に拡散され、人々に伝達されるのかを分析する。出来事の発生からの情報拡散の様子を分析することで、その出来事の種類や重大性の違いによるソーシャルメディアの地理的な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クサービス (SNS) を集合知として捉え、世の中で起きている出来事を把握しようとする研究は盛んに行われている。

榊らは Twitter ユーザをセンサとして捉え、彼らが発信する地震や台風などの自然災害の位置情報付きツイートを観測することで、発生した自然災害をリアルタイムに検出するシステムを開発した。また、岩渕らは 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して電車の混雑を予測するシステムを構築した。

ソーシャルメディアやソーシャルネットワークサービス (SNS) を集合知として捉え、世の中で起きている出来事を把握しようとする研究は盛んに行われている。榊ら[3]は Twitter ユーザをセンサとして捉え、彼らが発信する地震や台風などの自然災害の位置情報付きツイートを観測することで、発生した自然災害をリアルタイムに検出するシステムを開発した。また、岩渕ら[4]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して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に着目して他の地域のユーザに情報がどのように伝播するのかを分析する。

4. 分析手法

Twitter に投稿されたツイートを収集し分析を行うためシステムを構築する。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について示す。

4.1. ツイート収集

分析対象となる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[5] 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。

本研究では Streaming API の中に含まれる Sample API と Filter API を用いて Twitter からのツイート取得を行う。Sample API は投稿された全てのツイートの一部をランダムに取得できる。また Filter API は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。分析対象とするツイートは日本語のツイートのみとする。

4.2. 形態素解析

ツイートの内容を把握するためにツイート本文に対して形態素解析を行い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辞書には mecab-ipadic-NEologd[7] を使用する。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定期的に追加し更新を続けている辞書で、標準の IPA 辞書ではうまく形態素解析できない新語や固有表現を解析することが可能になる。Twitter ではユーザ動詞が日常的会話に近い表現でツイートが投稿されることが多いため、標準の辞書より mecab-ipadic-NEologd の方が最適な形態素解析を行えると考えた。

4.3. 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、日本においてはプライバシーなどの観点から位置情報を付与しているユーザは非常に少ない。

本研究ではツイートが投稿された場所が重要になるため、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必要がある。これを実現するためにユーザ情報に含まれる場所の情報とツイート本文に含まれる場所の情報の2つを組み合わせて投稿場所を推定し、ジオコーディングを行うことで位置情報を取得する。

4.4. ツイート分類

取得したツイートがどのような内容なのかを分類器を用いて分類を行う。ツイートの特徴を抽出するために4.2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) の要領でベクトル化を行い、いくつかのグループに分類する。予め用意したツイートを手作業で分類し、学習を行った分類器を用いて実際に収集したツイートを分類する。

4.5. 分析

投稿場所を抽出し、内容ごとに分類されたツイートを時系列ごとに集計し分析を行う。全文検索エンジンであるElasticsearch[9]に処理したツイートを蓄積し、分析フレームワークであるKibana[10]を用いて分析を行う。

5. 分析結果

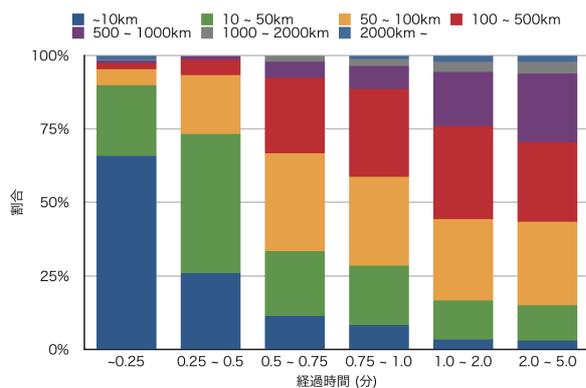


図2. 地震時からの震源からの距離ごとのツイート数の比率

図2は地震に関するツイートを対象に分析を行った結果である。一般的なTwitterユーザは揺れを感じたタイミングで地震に関するツイートを発信する。主要動と呼ばれる地震の大きな揺れはS波によって引き起こされる。S波は岩盤中で3~4km/sで進むため、揺れが発生する地域もその速度で広がっていく。地震発生時から15秒ごとに集計されたツイートを見るとS波とほぼ同じ速度で地震の情報が広がっていくことが分かる。

また、図3に示した地震発生時からのツイート種別の比率を見ると、地震発生直後は通常ツイートが大部分を占めているのに対して、地震発生からある程度の時間が経過するとリツイート(引用リツイートを含む)の比率が増えて

いる事が分かる。地震の場合、ユーザは発生のタイミングで地震に関する投稿をし、その後気象庁の情報を元に投稿される地震の詳細情報に関するツイートをリツイートし情報拡散を行っていることが読み取れる。

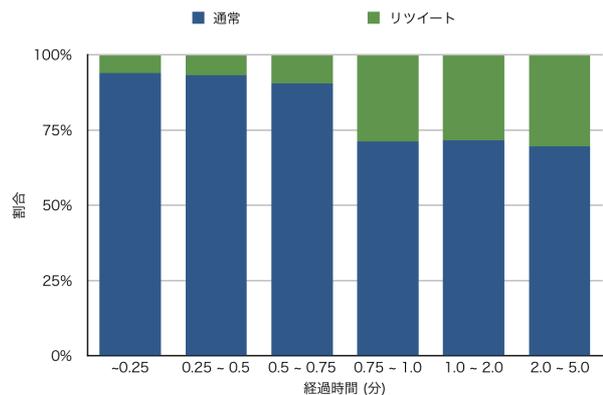


図3. 地震時からの震源からの距離ごとのツイート種別の比率

地震発生から2分後以降は実際に揺れが発生していない地方でもリツイートなどが行われ情報拡散が行われる。地震の場合は発生直後は揺れた地域を中心に、揺れが収まって各メディアから情報が公開されると日本中に情報が拡散されていくことが分かった。

6. 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分析した。分析により情報伝播は時間の経過と共に進み、その速度や地域的な広がりや情報の種類によって特徴が出ることを示した。

参考文献

- [1] Twitter Japan (@TwitterJP), <https://twitter.com/TwitterJP/status/793649186935742465> (2016).
- [2] 落合 涼, 伊與田 光宏, "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", FIT2017 第16回情報科学技術フォーラム (2017).
- [3] Takeshi Sakaki, Makoto Okazaki and Yutaka Matsuo, "Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors", WWW2010, pp.851-860 (2010).
- [4] 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).
- [5] Twitter Developers, <https://dev.twitter.com/> (2017).
- [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] 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).
- [9] Elasticsearch, <https://www.elastic.co/jp/products/elasticsearch> (2017).
- [10] Kibana, <https://www.elastic.co/jp/products/kibana> (2017).