

Twitterにおける位置と時刻の関係の可視化

Visualization of Relationships between Locations and Timings in Twitter

飯島 彩輝

Saiki Iijima

法政大学情報科学部デジタルメディア学科

E-mail: saiki.ijima.7i@stu.hosei.ac.jp

Abstract

With the spread of smart phones and social networking services such as micro-blogs, an increasing number of people record their actions in micro-blogs. A widely used micro-blog service Twitter provides messages called "tweets" that include 140-character texts, user identifications, timings, and locations. There is a possibility that more useful information can be obtained from relations between locations and timings of tweets. Information visualization is considered as a useful way for analyzing such relations between locations and timings. However, the location information obtained from Twitter consists of latitudes and longitudes. Therefore, when they are merged with timings, it is necessary to visualize at least three attributes at the same time, which is difficult for usual two-dimensional display. This paper proposes a method for visually showing these three attributes in two dimensions. It enables users to observe the geographic movement of tweets with similar contents. It also supports animation and keyword search to help users' analysis. The paper presents the experimental results of applying the method to tweets obtained from Twitter.

1. はじめに

近年、マイクロブログや SNS、スマートフォンの普及が進み、個人が行動を記録する機会が増えている。特に日本ではマイクロブログである Twitter [1]が流行しており、大量の投稿がされている。Twitter への投稿には、つぶやきと呼ばれる 140 文字以内のテキストや、ユーザ ID、時刻、位置の情報などが付与されている。それらを集集、分析することにより様々な情報を得ることが可能である。例えば、白井らの ChronoView [2]は、投稿されたイベントと時刻の関係を可視化する。また、梶原のシステム [3]は、単語と位置情報の関係を扱う。

位置と時刻の関係をを用いることにより有用な情報が得られる可能性がある。位置情報と時間の関係を分析する際に役立つ方法として可視化がある。しかし、Twitter から得られる位置情報は緯度、経度それぞれの値をもっており、時刻と合わせると少なくとも三つの属性を一度に表さなければならないため、単純に二次元で可視化することは難しい。

本研究では位置と時刻の関係を視覚的に分かりやすく表現する手法とアプリケーションを提案する。この手法ではつぶやきの密度を色の違いで表現し、時刻をユーザが調整可能なタイムスライダで表現する。また、分析の助けとなるよう、変遷をアニメーション的に表現できる機能や、単語による絞り込み機能なども実装した。このアプリケーションを利用してつぶやきの移動を可視化することにより、対象とするイベントがどのように移動したのかを視覚的に観測出来るようにする。

2. 関連研究

値が動的に変化するデータを可視化し、多角的な分析を可能とする手法に、伊藤らによる三次元可視化フレームワーク [4]がある。これは直接操作が可能な時間軸コンポーネントであるタイムスライダと、可視化キャンパスを組み合わせて、三次元で可視化する手法である。

ChronoView [2]はイベントのもつ時刻の集合を二次元平面上の位置で視覚的な表現する。アナログ時計の文字盤のような円を表示し、イベントの時刻の集合をその中に配置することで、データに記録されているイベントと時刻の関係性のおおよその分布を視覚的に提示する。

3. 提案手法

本研究では、マイクロブログへの投稿から得られる位置と時刻の情報を対象として、位置が時間とともにどのように変化するかを分析できる手法を提案し、アプリケーションを作成する。このアプリケーションはデータの取得、データの整理、そして可視化の三つから構成される。

3.1. 対象とするデータ

本研究が対象とする Twitter への投稿はつぶやきと呼ばれ、テキストや、ユーザ ID、時刻、位置の情報などが付与されている。本研究では公開されている API [5]を利用して、日本国内のつぶやきに限定した位置や時刻等の情報を 8 月から 12 月までの約四ヶ月間継続的に取得したデータを用いる。

3.2. 可視化

本研究では緯度、経度、時刻の三つ属性を可視化する。

3.2.1. 時刻の可視化

時刻に関しては、伊藤らの三次元可視化フレームワークからヒントを得たタイムスライダで表現する。画面下部に、時刻に対応したタイムスライダを表示する。ユー

はタイムスライダを動かすことにより自分の参照したい時刻の情報を直観的に表示することが出来る。また、連続的に動かすことにより時刻による位置情報の変遷をアニメーション的に閲覧することも可能となる。アニメーション的な閲覧のため、自動的に時刻を進め続けるボタンと自動的に時刻を戻し続けるボタンも設置している。

3.2.2. 位置情報の可視化

位置情報に関しては、地図のように緯度、経度に対応した座標に表示すると直観的に分かりやすい。ユーザがタイムスライダによって指定した時刻にされた各つぶやきの位置をヒートマップ [6]として視覚化する。具体的には、まず通常の地図のように上を北、下を南とする平面上を 200×200 のマス目にわけ、次に各マスに対応する緯度、経度の投稿の個数を計測し密度として格納する。そして、各マスの密度を色で表現する。

3.3. 絞り込み

取得したデータの中から観測したいデータのみを表示できるように、単語の絞り込み機能を実装する。例えば台風に関するつぶやきと時刻の関係を分析したいときに、取得されたデータから「台風」が含まれるもののみを検索し表示することが出来る。

3.4. 表示

これらの手法を使うことにより、緯度、経度、時刻を二次元で可視化することが可能となる。図 1 に提案手法による表示例を示す。

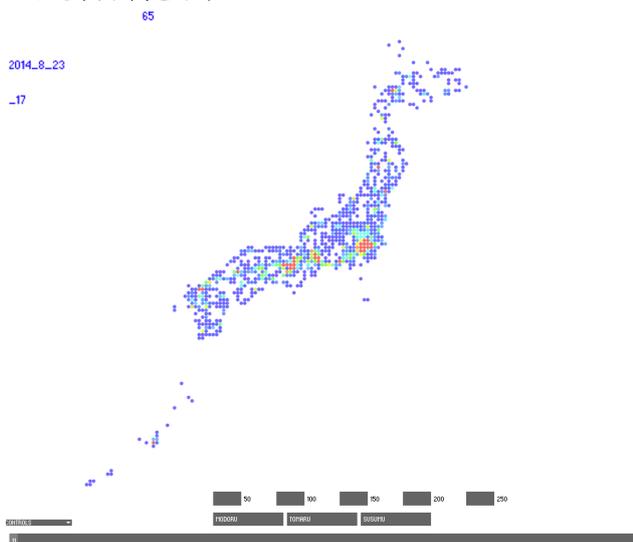


図 1 提案手法の表示例

4. 実装

前節の提案手法の実装について述べる。

4.1. データの取得

データの取得には、Java で簡単に Twitter API を利用できるライブラリ Twitter4J [7]を利用した。Twitter API 中の現在投稿されているつぶやきをリアルタイムに取得できる streaming API を利用して、日本国内のつぶやきを

取得し CSV ファイルとして保存し続けるプログラムを作成し 2014 年 8 月 21 日から動かし続けた。CSV ファイルは一日を 1 ファイルとしファイル名を日付とした。streaming API は投稿されたつぶやきのありとあらゆる情報を取得できるが、その中でも本研究に利用することを想定して、緯度、経度、時、分、秒、曜日、週番号、ユーザ ID、地区、県、内容に限定して取得した。

4.2. データの整理

時刻操作に伴う表示の変更の際に毎回、データの読み込み、計算、表示を行っている、計算自体が単純だとしてもデータの量が莫大なため、処理に非常に時間がかかり、実用に耐えることが出来ない。その問題を解決するため、初回起動時に CSV ファイルとして保存してあるデータを読み込み必要な処理を行い、その結果を各日付の時刻ごとにデータベースに格納するようにした。これにより初回以降のシステム起動時にはデータの整理に関する処理を実行する必要がなくなり、また、時刻操作の際のデータの読み込み、計算の処理をする必要も無く、データベースに格納された結果を表示するだけで済むようになる。例えば、2014 年 8 月 21 日 12 時を指定した場合、データベーステーブル 2014_8_21_12 を読み込み表示するだけで良い。結果として速度に関してはストレスを感じる事が無く利用できるようになった。

4.3. 可視化

可視化システムの実装にビジュアルデザインのためのプログラミング言語である Processing を使用した。Processing は Java を単純化し、グラフィック機能に特化した言語であるため、シンプルなコードで視覚的なアプリケーションを作成することが可能である。

4.3.1. タイムスライダ

表示時刻をユーザが指定するために画面下部にタイムスライダを実装した。実装には Processing の UI 実装を助けるライブラリ controlP5 を利用した。これを利用することでタイムスライダやボタン、チェックボックス等を簡単に実装することが出来る。タイムスライダの左から右を時間の流れとし、左から右に 1 ピクセルずらすと時刻は一時間進む。表示サイズの関係からスライダの最大値を 24 時間×61 日の二ヶ月分とし、そのうえに配置したチェックボックスによりどの二ヶ月を表示するか選択できるようにした。

4.3.2. 表示

データ整理にて作られたデータベースから指定されている時刻のものを読み出し画面に表示する。上を北、下を南とする平面上を 200×200 のエリアにわけ、その各エリアの総数に従って各エリアに着色する。エリアの着色にはヒートマップの要領で色相を利用する。エリアの総数が大きいほど赤に近く少ないほど青に近い色がつけられる。

4.3.3. 絞り込み機能

可視化システムによる分析を行う際の助けになるよう、つぶやきの中から指定したい単語が含まれているもののみを抽出し表示する機能を実装した。希望の単語を指定

することによりその単語が含まれているデータのみを対象としたデータベーステーブルを作成することが出来る。これにより、その単語が含まれたつぶやきが時刻とともにどのように遷移していくかを見ることが出来る。例えば「雨」という単語を指定すれば、「雨」を含むつぶやきが実際の雨と同じように動くのかどうかなどを分析することが出来る。

5. 実験

本可視化システムを利用した3つの例を示す。

5.1. 地震

地震の発生とつぶやきの分布が関連するという仮説をたて、地震に関連するつぶやきの可視化と分析を行った。具体的には、8月26日11時に栃木県北部で起きた地震、8月26日3時に大阪湾で起きた地震、11月22日22時に長野県北部で起きた地震を対象とした。



図2 8月26日11時の可視化結果

図3 8月26日3時の可視化結果

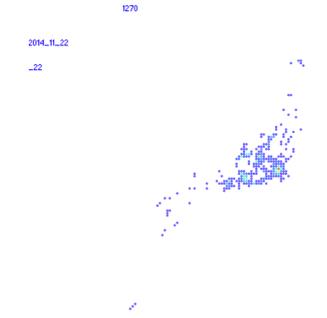


図4 11月22日22時の可視化結果

8月26日11時の栃木県北部の地震(図2)と8月26日3時の大阪湾の地震(図3)については地震の発生位置とつぶやきの分布は一致し、仮説通りとなった。しかし11月22日22時の長野県北部の地震(図4)に関しては、地震の発生位置とつぶやきの分布は重ならず、人口の多い関東・関西地域でのつぶやきが多かった。これは地震の規模と関係していると考えられる。前者二つは共に最大震度は3、後者の最大震度は5である。震度が大きいと揺れの伝わる範囲も広がるため、つぶやきの分布も広がったと考えられる。結果として実際に揺れた地域とつぶやきの分布は一致しているものと考えられる。この結果よりどの地域までの人々が揺れを体感したのかを観測することが可能である。

5.2. 台風

2014年10月13日から14日にかけて日本に上陸した台風19号を対象として、「台風」を含むつぶやきに絞り込み、13日から14日まで表示し分析を行った。台風19号の経路図を図5に、可視化結果を図6、図7に示す。

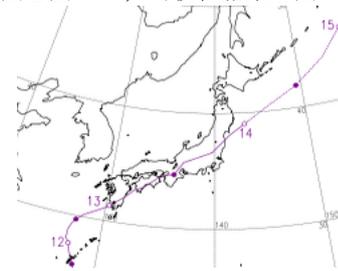


図5 台風19号の経路図(数字は日付)

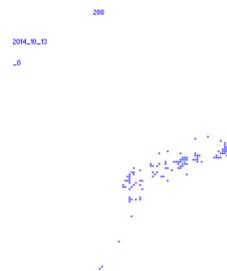


図6 10月13日0時の可視化結果

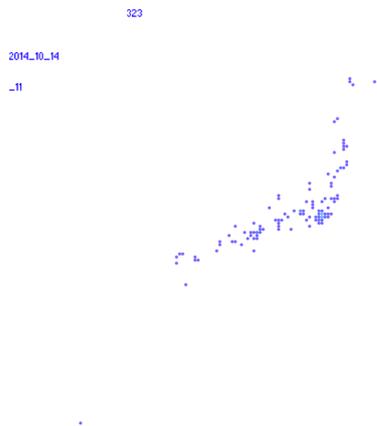


図7 10月14日11時の可視化結果

当初、「台風」を含むつぶやきは実際の台風の動きにそって動くという仮説をたてた。しかしながら実際に経路図と比較すると、台風が九州を直撃している13日0時でもつぶやきが九州に集中すること無く、関西・関東でもつぶやかかれており、北海道を通り過ぎつつあった14日11時でも同様に関西・関東で多くつぶやかれていた。すなわち「台風」を含むつぶやきは実際の台風の位置と連動して移動しない。これは台風に関する情報がその台風が観測された時点でテレビニュース等に伝えられるため、実際に自身の地域に台風が到達する、しないに関係なく、つぶやくことが多いからだと思われる。台風に関するつぶやきは台風がどこにあるかではなく、台風に関する情報を得たかどうかと、その地域の人口に起因することが分かった。

5.3. 起床

起床に関するつぶやきを可視化し分析することにより地域ごとの起床時刻の違いや生活リズムの違いが観測できるのではないかと仮説をたて実験を行った。今回は起床に関するつぶやきとして「おはよう」という単語を含むつぶやきを検索した。

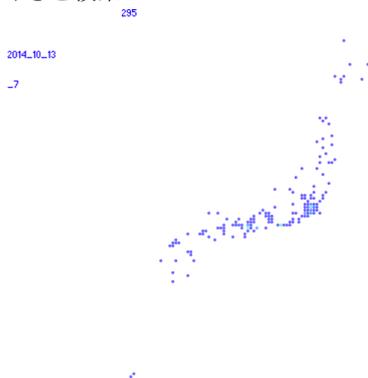


図8 起床に関する7時のつぶやきを可視化

この実験の結果、地域ごとによる差異を観測するまでには至らなかった(図8)。これは時刻の表示が1時間ごとと大きいことと、やはりつぶやきの密集度は人口の多さに起因しているためだと考えられる。

6. 議論

本研究で開発した可視化システムを利用することにより様々な事柄の移動していく様子をつぶやきの移動として観測することが可能になると当初考えていたが、実験の結果、必ずしも上手くはいかなかった。「台風」、「地震」の他、「雨」、「雪」、「ゲリラ豪雨」、「起床」、「就寝」等に関するつぶやきでの分析も行ったが、地震以外のいずれも事柄の移動を観測するにいたらず、大抵、人口の多い関東・関西地域で多くのつぶやきが観測される結果となった。これらの結果から、地震のような突発的なケース以外では、その事柄がどこで発生しているかよりも、その事柄がニュース等でどれだけ話題になっているかと、Twitterを使っている人口がどれだけいるかという点につぶやきの量が依存していることが分かった。従って、単純に単語のみでの絞り込みで事柄の発生位置、移動等を観測するのは難しいと考えられる。

しかし、本可視化システムに更に形態素解析や文法解析を含めることで改善できる可能性がある。また、単純ではあるが、表示時刻の単位を現在の1時間から30分や10分に縮めるだけでも改善できる可能性がある。

7. おわりに

本研究ではTwitterから得られる緯度、経度、時刻の三つの属性を同時に可視化する手法を提案し、可視化システムを開発した後、つぶやきの位置と時刻の関係を観測することを目的として実験と分析を行った。しかし、十分な結果は得られなかったため、時刻の細分化、形態素解析、文法解析を行うなどの改善が必要である。

文献

- [1] Twitter, [Online]. Available: <https://twitter.com>.
- [2] 白井智子, "大量の時刻情報付きデータの可視化手法の開発," 筑波大学システム情報工学研究科, 2013.
- [3] 梶原浩紀, "マイクロブログを用いた キーワードと地理的位置の対応付けシステム," 慶應義塾大学環境情報学部, 2010.
- [4] 伊藤正彦, 豊田正史, 喜連川優, "構造を持つ時系列情報に関する3次元可視化フレームワーク," 第19回インタラクティブシステムとソフトウェアに関するワークショップ(WISS2011), 2011, pp. 42-47.
- [5] Twitter, "The Streaming APIs," [Online]. Available: <https://dev.twitter.com/streaming/overview>.
- [6] LabEscape, "About Heat Maps," [Online]. Available: <http://labescape.com/info/about-heat-maps>.
- [7] Twitter4J, [Online]. Available: <http://twitter4j.org/>.