

# Twitter ユーザの時間・天気による行動傾向の可視化

## Visualization of Twitter Users' Action Tendencies by Using Timings and Weathers

田渕 翔太

Shota Tabuchi

法政大学情報科学部コンピュータ科学科

E-mail: shota.tabuchi.8d@cis.k.hosei.ac.jp

### Abstract

*An increasing number of Twitter users post information about their actions. If we can find tendencies of their actions from Twitter data, we will be able to effectively use it to promote local economies. This paper presents a method for visualizing tendencies of Twitter users' actions by using information about timings, weathers, and locations. The method uses circles to visualize users' actions on a blank map, where the locations and sizes of the circles express the barycenters and frequencies of the same actions respectively, and their colors express whether the actions depend on sunny or rainy weather. Therefore, it is possible to intuitively understand the results and find the tendencies of actions at a glance. The method uses Dynamic Queries to enable a simple graphical user interface for database queries that restrict the amount of visualized data. The Dynamic Queries interface provides buttons and sliders that allow users to pay attention to timings and weathers. The paper also provides the results of an experiment on the use of the Dynamic Queries interface for finding Twitter users' action tendencies by using information about timings and weathers.*

### 1. はじめに

近年、Twitter [1]のユーザ数が増加しており、Twitterに自身の行動を記録することが増えてきた。Twitter上のデータからユーザの情報、特に行動に関する情報を読み取る事ができれば、地域経済の活性化に有効利用することが可能になる。しかし、情報量が膨大なため、表示する情報量のフィルタリングと可視化の手法が必要となる。

本研究では Twitter ユーザの行動の可視化を行う。Twitter ユーザが自身の行動を決定する際の情報として、晴れているか、雨が降っているかなどの天気に関する情報、今いる場所が都心か山村地帯かなどの位置に関する情報、今が朝か夜か、夏か冬かなどの時間に関する情報が挙げられる。そこで、本研究では時間・天気・位置の情報に着目して可視化を行う。特に行動を確実に左右すると考えられる時間に重点を置く。

可視化では、行動を表す円を白地図上に配置して位置情報を表し、その円の上に行動名を記す事でその円の行動を表す。配置した円の大きさでその行動の情報量を表し、円の色で行動の天気情報を表す。

可視化する行動の決定に動的検索 [2]を用いる。動的検索によって、ユーザは SQL 文などの難しい文法のクエリを入力する事なく、ボタンやスライダーなどの簡単なグラフィカルユーザインタフェース(GUI)の操作でクエリを実行出来る。このため、エラーや文法ミスがなくストレスフリーな可視化が行える。時間や天気、曜日など様々な種類のボタンを用意することで、ユーザが任意の情報について重点的な可視化を行う事を可能にする。

### 2. 関連研究

本研究の先行研究として、原口の可視化の研究と Shneiderman らの動的検索が挙げられる。原口 [3]は、本研究と同様に時間・天気・位置に着目し、マイクロブログユーザの行動傾向を可視化した。

Shneiderman らの動的検索 [2]では、スライダーとボタンなどの簡単な GUI によるデータベース検索の支援を行う。動的検索はクエリの作成において SQL 文を入力するのに比べて、ユーザが文法を理解する必要がなく、GUI を操作するだけで直感的にクエリを作成することができる。さらに、結果の出力が動的に行われるという特徴を持っている。本研究では、可視化する情報のフィルタリングに動的検索を用いる。

### 3. 提案手法

本研究では取得したツイートについて、時間・天気・位置に着目してユーザの行動がどのように変化しているのかを白地図上に可視化する手法を提案する。可視化手法として動的検索を採用する。ボタンやスライダーなどの GUI を利用しクエリを作成することによって、初心者でも動的な可視化を可能にする。

#### 3.1. 抽出する行動の辞書の作成

行動を抽出する前に、何を行動とするかを決める必要がある。そのため、行動とする単語の一覧を辞書として作成している。辞書に登録する単語としては、野球やサッカーなどのスポーツに関する単語や、昼食やラーメンなどの食事に関する単語、駅やカラオケなどの場所に関する単語などを主に登録している。

### 3.2. 可視化手法

本研究ではツイートを元にユーザの行動を時間・天気・位置に着目して白地図上に可視化する。行動の多さを円の大きさで、行動の重心の位置を円の位置で、行動の天気への依存度を円の色で表す。可視化する情報は、動的検索インタフェースのボタンやスライダーを操作し作成されたクエリでフィルタリングする。

#### 3.2.1. 円の配置と大きさ

行動を表す円を白地図上に可視化する。円の位置は行動ごとのツイートの位置 $(x_i, y_i)$ とその総数(num)を使い、各都県での重心の位置 $(px_i, py_i)$ に配置される。

$$px_i = \frac{\sum_{i=1}^{\text{num}} x_i}{\text{num}}, \quad py_i = \frac{\sum_{i=1}^{\text{num}} y_i}{\text{num}}$$

円の大きさ(size)は、その円が表す行動の多さ(num)を表す。ただし、各行動のツイート数は均等に分布していないため、一部の突出したツイートの円が大きくなりすぎるのを防ぐために一定数以上の場合にはツイート数の平方根を取る。

$$\text{size} = \begin{cases} 5 + \text{num} & \text{num} < 10 \\ 15 + \sqrt{\text{num} - 10} & \text{num} \geq 10 \end{cases}$$

#### 3.2.2. 天気の可視化

天気は行動を示す円の色で表現する。晴れの時は 1、曇りのときは 0、雨のときは-1 と数値化し、その合計値をその行動の数(num)で割る事で都県ごとにその行動の天気の平均(a1)を取る。

$$a1 = \frac{\sum_{i=1}^{\text{num}} a_i}{\text{num}}, \quad a_i = \begin{cases} 1 & \text{晴れ} \\ 0 & \text{曇り} \\ -1 & \text{雨} \end{cases}$$

さらに、各県の全行動の平均(a2)を同様の算出方法で出し、各行動の平均(a1)を全行動(num)の平均(a2)で割る事で、平均と比べて各行動が雨に多いか晴れに多いかを表す重み(w1)を得る。

$$a2 = \frac{\sum_{i=1}^{\text{num}} a_i}{\text{num}}, \quad w1 = \frac{a1}{a2}$$

w1は地域の平均と各行動の平均を比較し、前者よりも後者が大きい場合に 1 よりも大きくなり、行動が晴れの時に多くツイートされている。前者が後者よりも大きい場合に 1 より小さくなり、行動が雨の時に多くツイートされていることがわかる。

わかりやすさのために w1 から 1 を減らし 0 を中心とした重み(w2)とし、w2が正の場合に晴れをイメージしやすい赤で、負の場合に雨をイメージしやすい青で円を描写することで天気を可視化する。その際の色濃さはその行動が天気へ強く依存するほど濃くなるようにする。具体的には、天気への依存度を表すw2に 128 を掛け、その結果(w3)を 128 に足したものを RED, 128 から引いたものを BLUE, RED と BLUE の内低い方を GREEN とした RGB 値を色とした。これにより、天気への依存度が低い行動は灰色に近い色になり、依存度が高い行動は純粋な赤、青に近く可視化される。

#### 3.2.3. 動的検索

本研究では、ユーザが任意の情報に重点を置いた可視化と、表示する情報のフィルタリングを行うためにクエ

リの作成に動的検索を用いる。本研究のようなクエリ間の変化が重要になる場合には SQL 文を何度も入力するの比べ、GUI を繰り返し操作するだけで変化に着目しやすいため、動的検索が向いていると考えたためである。

#### 3.2.4. 詳細表示

白地図上に可視化した行動の元となったツイートをすることで、さらなる情報を得る事が出来る。このために白地図上の行動を左クリックすることでその行動の元のツイート本文を表示する。また、各ツイートの座標に点を打つことによって実際の分布を見ることができるようになる。その際に件数が 1 などの円は小さく、円を正確にクリックすることは難しいため、クリック位置と各行動の円の位置までの距離を計算し、最短距離の行動についての詳細表示を行うようにする。

## 4. 実装

### 4.1. ツイートの取得

本研究では、Java を用いてプログラムを実装した。ツイートの取得には Twitter4J [4] という Java 用の Twitter API ライブラリを使用し、実際に投稿されたツイートの約 1% を収集した。データの取得期間は 2014 年 7 月 20 日から 2014 年 10 月 20 日までの 4 ヶ月間である。

### 4.2. データの加工

収集したツイートの情報のうち、本研究で利用する情報である日付、時刻、位置、本文のみを対象に以下のような加工を行いデータベースとして登録した。

#### 4.2.1. 地域の限定

取得したツイートのうち関東地方でのツイートを抜き出した。ツイートにある位置情報は緯度経度の情報のみなので、逆ジオコーディング [5] を行い、緯度経度の情報から都道府県の情報を読み取り、神奈川県・東京都・埼玉県・群馬県・茨城県・栃木県・千葉県に該当するツイートのみ絞り込んだ。

#### 4.2.2. 天気情報の付与

気象庁 [6] は過去の天気情報が各県の官署ごとに CSV ファイルで公開している。この天気情報と、取得した日付、時刻と逆ジオコーディングから得られた県情報を組み合わせることでツイートの情報に天気の情報を関連付けた。

#### 4.2.3. 行動の抽出

スポーツや食べ物、場所などのユーザの行動が読み取れるのではないかと予想された単語を 200 単語登録した行動の辞書と部分一致検索を行い、一致した単語をそのツイートの行動として抽出した。さらにこれらの単語は文章によって漢字、カナなどの表記のゆれが生じる事が考えられるため、テキスト解析を行い単語の読みで判断することで抽出漏れを減らした。

### 4.3. データベースの作成

加工したツイート情報を SQLite の JDBC [7] を利用して Eclipse 上からデータベースの作成、クエリの送信をできるようにした。加工したツイートの持つ情報である年、月、日、曜日、時、分、秒、緯度、経度、県、天気、行

動、本文をカラムとして持つデータベースを作成し、これまでに加工した全ツイートを登録した。

#### 4.4. 可視化プログラム

可視化プログラムには GUI として、日付、時間、件数のスライダーを持ち、天気、月、曜日、県のボタンを持つ動的検索インターフェースを実装したクラスに白地図の描写領域と詳細表示のためのテキストエリアを組み合わせた。図 1 に実際のプログラムの画面を示す。画面の左から順に白地図、詳細表示のためのテキストエリア、動的検索インターフェースとなっている。動的検索インターフェースのスライダー、ボタンを操作しデータベースへのクエリを作成することによって、画面左側の白地図上にクエリの結果が行動の円として可視化される。

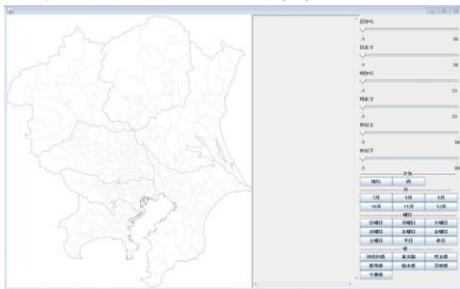


図 1 動的検索インターフェース

日付は 1 日から 31 日までの範囲のスライダーを 2 つ用意し、20 日から 3 日までのように月を跨ぐ検索にも対応した。時間に関しては 1 時から 24 時まで 1 時間刻みで実装した。件数は 1 件から 101 件まで実装し、2 つのスライダーで以上と以下を表すことでフィルタリングを行った。その際に 101 件以下しか表示できなくなるのを避けるために、以下を表すスライダーは 101 件にセットした場合には上限なしと設定した。天気は晴れ、雨の 2 種、月は 7 月から 12 月までの 6 種、曜日は 7 種、県は神奈川県、東京都、埼玉県、群馬県、茨城県、栃木県、千葉県 の 7 種を用意して ON・OFF のトグルで表現した。

### 5. 実験

実装した可視化プログラムの動的検索インターフェースのボタンやスライダーを操作して白地図上に動的に出力される行動から、場所・天気・時間についてユーザの行動傾向が読み取れるかどうかを確認した。さらに行動傾向が読み取れた行動については詳細表示を確認することで、その原因についての考察を行った。

#### 5.1. 場所による行動傾向

場所に依存する行動としては、ディズニーシーや鴨川シーワールドなどのレジャー施設などでは、白地図上の円の位置と実際の施設の位置が重なっており、そのような場所で今どこにいるのかツイートで報告している人が多い事がわかった。図 2 に動的検索インターフェースを操作し、34 件以上 68 件以下のツイート数の行動を白地図上に可視化した結果を示す。ディズニーランドの円がある場所が実際の位置とほぼ重なっている。

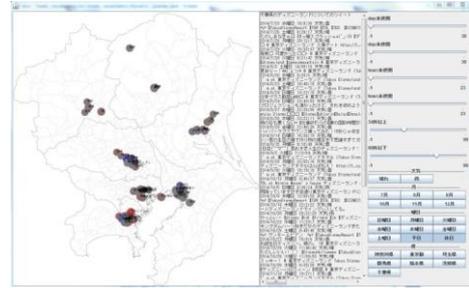


図 2 ディズニー

駅についてのツイートが突出して多かったため、その理由を詳細表示で確認した。図 3 に示す通り Foursquare [8] というアプリケーションの「I am at ○○駅」という現在位置をツイートする機能のため、駅が多く抽出される結果となっていることが詳細表示から明らかになった。

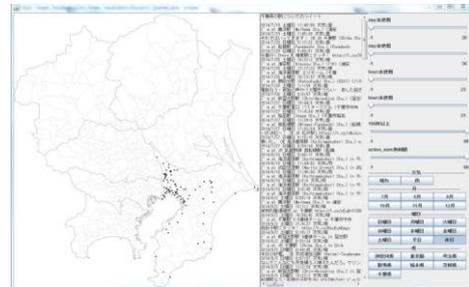


図 3 Foursquare

#### 5.2. 天気による行動傾向

雨に依存している行動としては、台風が見受けられた。ツイート数 10 件以上で雨に依存した行動を可視化した結果を図 4 に示す。半分以上の地域において台風という行動が表れている。しかし色が完全な青ではないので詳細表示で確認したところ、台風が来る前に心配するツイートや、台風が過ぎた後に安堵するようなツイートが晴れや曇りの日に行われていることがわかった。

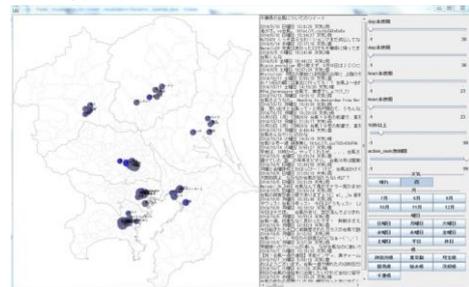


図 4 台風

晴れに依存している行動の例として、3 件以上のツイートで晴れに依存している行動を図 5 に示す。ゴルフやサッカーなどのアウトドアスポーツがわずかに見られたが、雨の場合と比べて天気に大きく依存している結果は見られず多くが灰色であった。

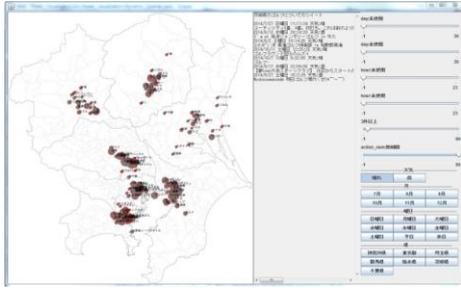


図5 アウトドアスポーツ

### 5.3. 時間による行動傾向

図6に平日の朝6時~8時でツイート数が24件以下の行動の可視化結果を示す。朝ごはんや起床、起きたなど朝の行動が多く可視化されているのがわかる。

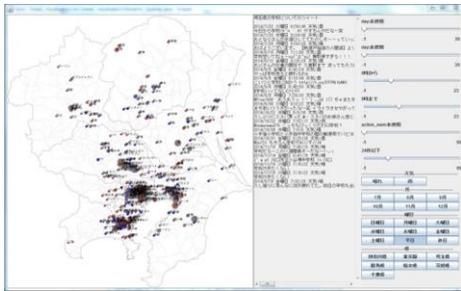


図6 平日朝

また、10時から14時でツイート数が10件以上50件以下の行動を可視化した結果を図7に示す。ランチやラーメン、うどん、カレーなどの人気の高い食事が多く見られ、昼食時にツイートする人が多いことがわかる。

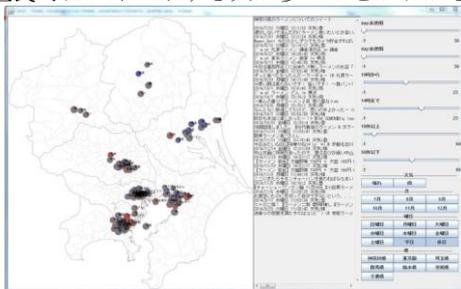


図7 昼

### 6. 議論

天気による行動の変化はあまり見られなく、ほとんどが灰色になった。行動の抽出を単純な一致検索で行っているため、現在、過去、未来を考慮することができず、天気と行動の結びつきが弱くなる事や、係り受けを考えていないためにその行動に対してのユーザの考えがネガティブなのかポジティブなのか判断できず、精度が下がってしまった事が考えられる。

リツイートでは情報が意味をなさない問題も見受けられた。有名人の目撃情報などが場所の行動として可視化された場合に、他のユーザによって場所や時間に関係な

くリツイートが行われるために情報の意味がなくなってしまうと考えられる。

行動の辞書は駅などの名詞が多く、動詞などは活用のため少なくなってしまった。カタカナや漢字などの表記の揺れを考慮したが、形態素解析も行い動詞の活用にも対応する必要があると考えられる。さらに表記の揺れが考えられない単語は読みによる検索をやめる事で誤検知を減らすなどの精度の向上が必要であると考えられる。

都市部にツイートが集中していて可視性が低く、逆に他の部分にツイートが存在せずに無駄な領域が多い事がわかった。その対策として動的検索インタフェースを微調整し、目的の行動のみを可視化するという方法は取れるが、どうしても読み取れない場合が存在した。これについては3次元の可視化が有効だと考えられる。

### 7. おわりに

本研究の目的であったTwitterユーザの時間・天気による行動傾向の可視化は、いくつかのユーザの行動傾向を読み取ることに成功した。しかし、テニスなどのスポーツは晴れに依存し、起床や朝食などは朝に集中するなどといった明確な結果は得られなかった。

動的検索インタフェースを採用する事で簡単にクエリを高速に生成することができた。しかし、今回扱ったデータベースでは要求する処理によっては動的と言えないような数秒程度のレスポンスとなることがあった。

データベースの精度の向上と、行動が一箇所に集中して可視性が下がる点の改善は今後の課題としたい。

### 文 献

- [1] "About Twitter," <https://about.twitter.com/>.
- [2] C. Williamson and B. Shneiderman, "The dynamic HomeFinder: evaluating dynamic queries in a real-estate information exploration system," in *Proc. 15th ACM SIGIR conference on research and development in information retrieval (SIGIR '92)*, pp. 338-346, 1992.
- [3] 原口義経, "マイクロブログユーザの行動傾向の可視化 (卒業論文)," 法政大学情報科学部コンピュータ科学科, 2014.
- [4] "Twitter4J - A Java library for the Twitter API," <http://twitter4j.org/ja/index.html>.
- [5] 農研機構, "簡易逆ジオコーディングサービス / Finds.jp Web サービス," <http://www.finds.jp/wdocs/rgeocode/index.html.ja>.
- [6] "気象庁 Japan Meteorological Agency," <http://www.jma.go.jp/jma/index.html>.
- [7] T. L. Saito, "xerial / sqlite-jdbc -- Bitbucket," <https://bitbucket.org/xerial/sqlite-jdbc>.
- [8] Team Foursquare, "Foursquare について," <https://ja.foursquare.com/about/>.