

Twitter ネットワークにおける集団注意の創発ダイナミクス

笹原和俊* 平田祥人† 豊田正史† 合原一幸†
Kazutoshi Sasahara Yoshito Hirata Masashi Toyoda Kazuyuki Aihara

Abstract: Quantifying online social data is essential to explore collective social dynamics. Here we propose a simple method for quantifying collective attention on Twitter network. The difference between regular and irregular states of tweet stream is measured by Jensen-Shannon divergence, which is associated with the intensity of collective attention. We then associate the detected incidents with corresponding events, to which a large amount of people pay attention, on the basis of the popularity and the popularity enhancement of terms in tweets. We demonstrate this method to be effective in a large dataset of Twitter, with a discovery of the emergence of various collective attentions.

Keywords: collective attention, emergence, Jensen-Shannon divergence, Twitter

1 導入

現在、人々は SNS を利用して情報を発信・共有し、実世界とは違うかたちのコミュニケーションを行っている。そしてそれは、実世界に対しても大きな影響力を持つ。本研究では、ポピュラーな SNS の 1 つである Twitter に注目する。Twitter は、「世の中の今を伝え合う」ツールで、ユーザーは今どうしているのかを 140 文字以内でつぶやき、別のユーザーがつぶやきで反応し、その連鎖によってユーザーネットワーク上を瞬く間に情報が伝搬する。このようなリアルタイム性、ネットワーク性が高い集団社会現象の本質に迫るためには、ソーシャルデータの特性を考慮した解析手法が必要となる。そこで、ツイートストリームにおける定常と非定常の差に着目して、Twitter 上で生じる集団注意 (Collective Attention) を定量的に捉えるための手法を提案し、集団注意の創発現象を分析する。

2 方法

Twitter REST API¹を利用してスノーボールサンプリングを行い、約 40 万人のユーザーから約 5 億ツイートを収集し、データベースを構築した。各データには、ツイートのテキストの他に、ユーザープロフィール、タ

イムスタンプや位置情報などのメタデータも含まれている。本研究では 2010 年と 2011 年のデータを解析対象とし、ツイートのテキストとタイムスタンプのみを解析に用いた。

通常、ツイートストリームは三相の概日リズムを示すが、実世界において大きなイベントが生じると、ツイートのバースト的な増加や不安定な振動が生じる (図 1)。この観測事実に基づき、ツイートストリームの定常状態と非定常状態の差を Jensen-Shannon ダイバージェンス (JS) で定量化して、ユーザーが大きく反応したイベント、すなわち集団注意を検出する。ここで JS の大きさは集団注意の強度と解釈される。 JS は Kullback-Leibler ダイバージェンス (KL) を対称化したもので、確率分布 $P = \{p_i\}$ と $Q = \{q_i\}$ の差異を測るのに用いられる [1]。 JS は非負の値をとり、 KL と違って常に有界のため、実データの解析に応用するのに適している。

$$JS(P, Q) = \frac{1}{2} \left(KL(P, \frac{P+Q}{2}) + KL(Q, \frac{P+Q}{2}) \right)$$

$$KL(P, Q) = \sum_i p_i \log_2 \frac{p_i}{q_i}$$

P は各日毎に求めたツイートの確率分布を、 Q は年平均を用いる。

次に、ツイートのテキストを形態素解析して得られたトークン (名詞) の頻度と頻度増加率 (当日と前日の同時間帯における頻度の比) に着目して、検出された事象と対応するイベントを同定する。形態素解析には

*名古屋大学大学院情報科学研究科, 464-8601 愛知県名古屋市千種区不老町, sasahara@is.nagoya-u.ac.jp, Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Naogyu, Aichi 464-8601, Japan

†東京大学生産技術研究所, 153-8505 東京都目黒区駒場 4-6-1 Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

¹<https://dev.twitter.com/docs/api/>

MeCab²とNAIST-jdic³を用いた。

3 結果と議論

解析の結果、 JS が 0.005 を超える事象が、2010 年は 34 件、2011 は 26 件が検出された (図 2)。 JS が大きな値を示したところは、何らかのイベントが実世界で生じ、それが人々のツイートをアフォードして、集団注意が創発したことを示している。 JS が平均以上の値を示した時間帯に投稿されたツイートを形態素解析し、名詞のみに着目して頻度と頻度増大率を求めた。この結果に基づき検出された事象を分類したところ、自然災害、スポーツ、文化、年間行事などに分類された。

特に大きな JS を示したのが、自然災害とスポーツイベントに関するものだった。例えば、大震災の当日は、2011 年で最大の集団注意が生じ、この日から 4 日間連続で JS は 0.005 を超えた。そして、「地震」、「津波」、「停電」などの、日常ではほとんど使われないような言葉がツイートの多くに含まれていた。このように高強度の集団注意が連続して観測されたのは、この時のみである。人々の注意の移ろいやすさを考えると、日本人がいかに大震災から大きく影響を受けたのが定量的に分かる。また、日本女子サッカーの世界カップ優勝やバンクーバーオリンピックのフィギュアスケートなど、国際的なスポーツイベントにおいても大きな集団注意が生じ、ゲームの進行と同調して、感嘆や応援の言葉がツイートされた。一方、中程度の JS に目を向けてみると、文化や科学や政治など、興味深いタイプの集団注意が見られた。例えば、はやぶさの帰還や皆既月食、選挙速報、中にはアニメのクライマックスシーンと同期した集団注意も見られた。最後の例は、日本独特の文化や慣習を反映していて興味深い。また、正月や大晦日などの年間行事は、一日を通して全体的にツイート投稿のパターンが通常と異なり、非同期的な集団注意の存在も確認された。

このように、ツイートストリームの定常からの逸脱とその程度を見積もることで、集団注意の創発を検出できることを実証し、2010 年と 2011 年に生じた全事象を同定することができた [2]。今回、日本語のツイートに解析を限ったが、この方法は他の言語にも応用可能である。ただしその場合は、データは国もしくは地域を限定した方が検出力が上がると予想される。ソーシャルメディアの登場によって、「行動の化石」がデジタルに蓄積されるようになり、そのようなデータからの知識発見の手法は、これからますます重要になる。本手法は、予測能力

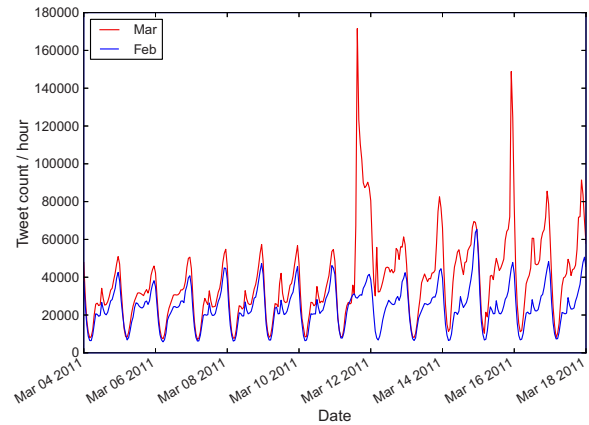


図 1: ツイートストリームの例

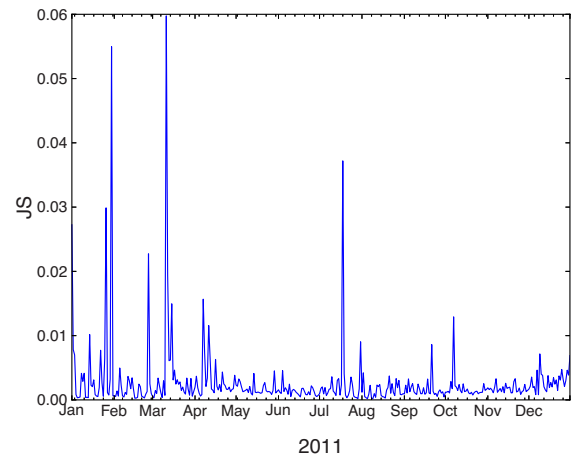


図 2: 集団注意の強度

こそ持たないものの、集団注意を簡便に定量化することが出来るという点において有効である。そこから得られる結果は、集団的社会現象を探求するための重要な基礎データとなる。

謝辞

本研究は、総合科学技術会議により制度設計された最先端研究開発支援プログラム (FIRST 合原最先端数理モデルプロジェクト) により、日本学術振興会を通して助成されたものです。

参考文献

- [1] J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, Vol. 37, No. 1, pp. 145–151, 1991.
- [2] K. Sasahara, Y. Hirata, Y. Toyoda, M. Kitsuregawa, and K. Aihara. Quantifying Collective Attention From Tweet Stream (in prep.).

²<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

³<http://sourceforge.jp/projects/naist-jdic/>