

Mention ツイートからのバースト期間の抽出

Extracting Burst Period from Mention Tweets

小出 明弘†

高橋 友美†

斉藤 和巳†

風間 一洋‡

鳥海 不二夫§

Akihiro Koide

Tomomi Takahashi

Kazumi Saito

Kazuhiro Kazama

Fujio Toriumi

1. まえがき

近年Twitterのようなソーシャルメディアを対象にした研究が盛んに行われている[1]. その中でも、我々はソーシャルメディア上でのユーザの行動の特徴を把握し、現実のユーザ行動を反映したモデルを構築することに関心がある.

本論文では、ソーシャルメディア上のユーザ行動のモデル化のための事前調査として、社会影響を受けた際のTwitter上でのユーザ行動について分析する. 具体的には、東日本大震災が発生した前後の期間での、TwitterのMention ツイートとHashtag ツイートを対象にし、それらの各トピックにおけるツイートのバースト期間を抽出することにより、これらの機能を利用したツイートがどのように変化したのかを分析する. 実験により、Mention ツイート、Hashtag ツイートともに震災の影響を大きく受けたトピックのバースト現象が抽出される一方で、Hashtag ツイートでは震災に左右されないバースト現象が抽出されることを示す.

2. バースト期間抽出法

本章では、本分析で利用するツイートのバースト期間抽出法[4]について述べる. あるトピックに対するツイート時刻集合 $\mathcal{D} = \{t_0, t_1, \dots, t_N\}$ が観測されたとき、トピックのツイート頻度における変化点を T_j とする. ここで、 $t_0 < T_j < t_N$ である. この時、ツイート間隔が指数分布 $p(t_n - t_{n-1}; r) = \text{rexp}(-r(t_n - t_{n-1}))$ に従うと仮定し、変化点 T_j の直前までは r_j 、変化点 T_j 直後では r_{j+1} というパラメータの指数分布に従うとする. いま、 J 個の変化点集合を $S_J = \{T_1, \dots, T_J\}$ とし、パラメータベクトルを $\mathbf{r}_{J+1} = (r_1, \dots, r_{J+1})$ とすると、ツイート時刻集合 $\mathcal{D} = \{t_0, t_1, \dots, t_N\}$ に対するバースト期間抽出は、以下の式を最大化する変化点集合を求める問題となる.

$$L(\mathcal{D}; \mathbf{r}_{J+1}, S_J) = -N - \sum_{j=1}^{J+1} |\mathcal{D}_j| \log \left(\frac{1}{|\mathcal{D}_j|} \sum_{t_n \in \mathcal{D}_j} (t_n - t_{n-1}) \right). \quad (1)$$

実際には、変化点が J 個存在するときと、変化点が存在しない時の尤度比の対数（以下バースト度と呼ぶ）を最大化する変化点集合 S_J を求める問題を考え、さらに[4]で提案されている解品質を向上させる手法を利用するが、詳細は割愛する.

3. 実験

本章では、本分析で使用するMention ツイートとHashtag ツイートに関する詳細を述べ、本分析の実験設定について説明する.

3.1 使用データと実験設定

一つ目は、Mention ツイートデータ[3]である. 2011 年 3 月 7 日から 3 月 23 日までの@User がついたツイート（Mention ツイート）174,794,582 件から、そのユーザごとの時系列データをツイート時刻集合として用いた. 具体的には、ツイート数上位 1000 ユーザのツイート時刻集合を用いた. 二つ目は、Hashtag ツイートデータ[4]である. Mention ツイートと同様の期間内において、記号“#”の直後に文字列がついたツイート（Hashtag ツイート）22,482,446 件から、ハッシュタグごとの時系列データをツイート時刻集合として利用する. 具体的には、ツイート数上位 1000 件のツイート時刻集合を用いた.

これらのツイート時刻集合に対して、2 章で述べた手法を用いてバースト期間を抽出する. 本分析では、 $J = 2$ とし、バーストした 1 区間を求める.

3.2 実験結果

図1、図2にMentionツイート、Hashtagツイートの累積ツイート分布とバースト期間抽出結果を示す. 本稿では、バースト度が上位4位となったものを示しており、横軸は時刻、縦軸は累積ツイート数を表す. グラフ中の縦軸に平行な2本の線で挟まれた区間が検出したバースト期間となる.

図1のバースト度1位～3位では、バースト期間の始点が震災直後になっている. さらに、1位のNHK_PRと3位のnhk_seikatsuは、ほぼ同様の区間がバースト区間として検出される. これらのアカウントはNHKの公式アカウントとして日常的にある程度拡散されるような情報を提供しており、震災後もユーザにとって有用な情報を提供していたことから、細かなバースト現象が起きたと考えられる. 一方、バースト度2位のkazhime1977は個人アカウントであり、震災前には拡散を生むようなツイートはみられなかったが、震災直後に阪神淡路大震災の際の経験談をツイートしたことにより、そのツイートが極めて短時間に急激に拡散していた. このことから、大きな社会現象により、緊急性の高い内容を含んだツイートであれば発信源に関わらず急激な拡散が発生すると考えられる.

また、バースト度4位のitsumonoTLは震災数日後からツイートが始まり、その後一日単位でほぼ同様の推移をしている. このアカウントは震災後の自粛的な状況で、それまで日常的に発言するような内容をツイートしにくくなっ

†静岡県立大学, University of Shizuoka

‡NTT 未来ねっと研究所, NTT

§東京大学, University of Tokyo

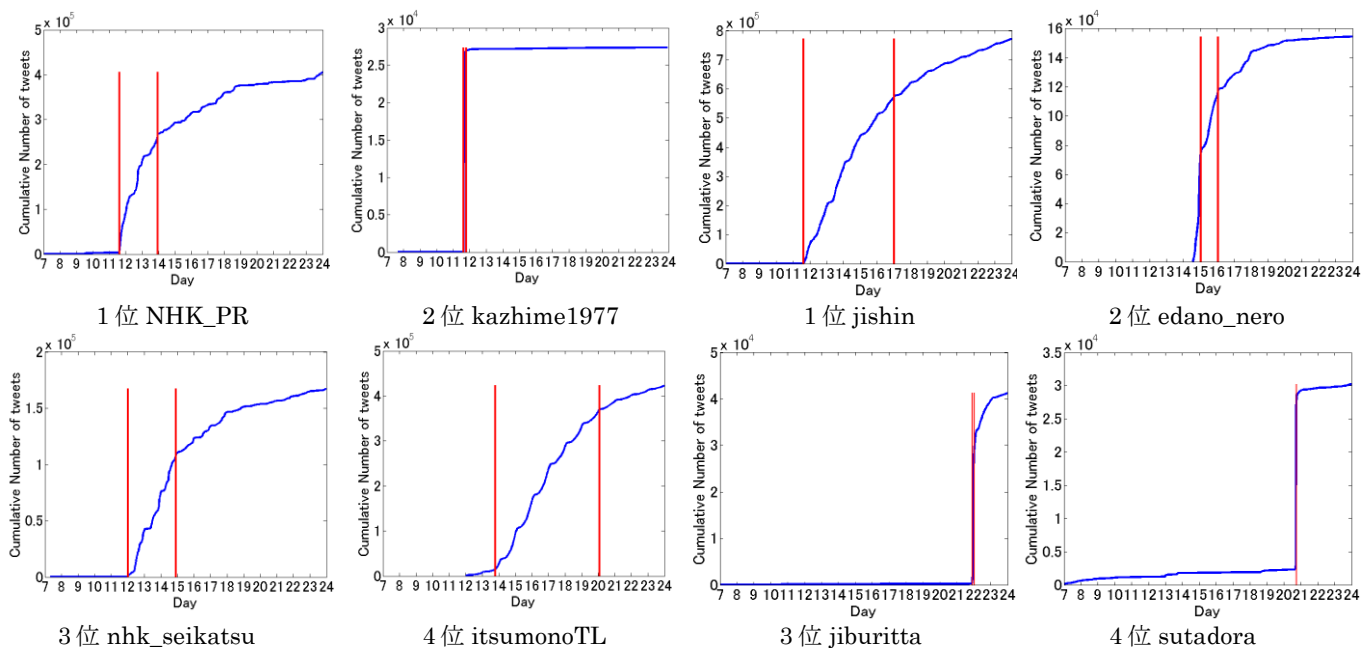


図 1 Mention ツイートのバースト期間と累積ツイート数

たユーザが、このアカウントに **Mention** することで日常のツイートをつぶやけるように配慮して作られたものであり、震災の影響を受けて新たに生まれたものである。これらの結果から、バースト度の高い **Mention** ツイートは、期間に違いはあるものの震災の影響を多分に受けたものが検出されている。

図 2 の **Hashtag** ツイートでは、**Mention** ツイートとバースト期間の大きく異なるものがみられる。まず、バースト度 1 位の **jishin** は **Mention** ツイートの **NHK** 関連と同様に震災直後からバーストが始まり、その後は **itsumonoTL** と同様に一日単位で一定の増加傾向がある。このような傾向が見られた理由として、ハッシュタグ **jishin** では拡散を求めるような緊急性のツイートと情報共有のツイートが混合していることが原因である。

一方、2 位の **edano_nero** はこれまでにないような変化をしている。この **Hashtag** は震災後に作られたもので、枝野官房長官に対するねぎらいの意味を込めた造語であった。なお、このハッシュタグは 0 の状態から急激にバースト現象が起こったことにより、本稿で利用したバースト期間検出法では、発生直後の急激な増加部分を検出することができていない。さらに、3 位の **jiburitta** と 4 位の **sutadora** は、バーストした期間が震災後 10 日以上経過したところで起こっている。**jiburitta** は、診断メーカー (<http://shindanmaker.com/>) と呼ばれるサービスによって作成されたもので、**sutadora** は特定のアニメに関するものであった。これらの **Hashtag** は震災との直接的な関係はないものであるが、ソーシャルメディア内外での流行や習慣によって急激なバースト現象が見られるものであった。

Mention ツイートと **Hashtag** ツイートのバースト傾向を比較すると、震災という大きな社会影響を受けたことにより、両ツイートともに影響を反映した形でユーザ行動が大きく変化していることがわかる。また、それまで注目されなかったようなトピックが急激にバーストする現象が観測される点に関しても、共通的な特徴である。その一方で、

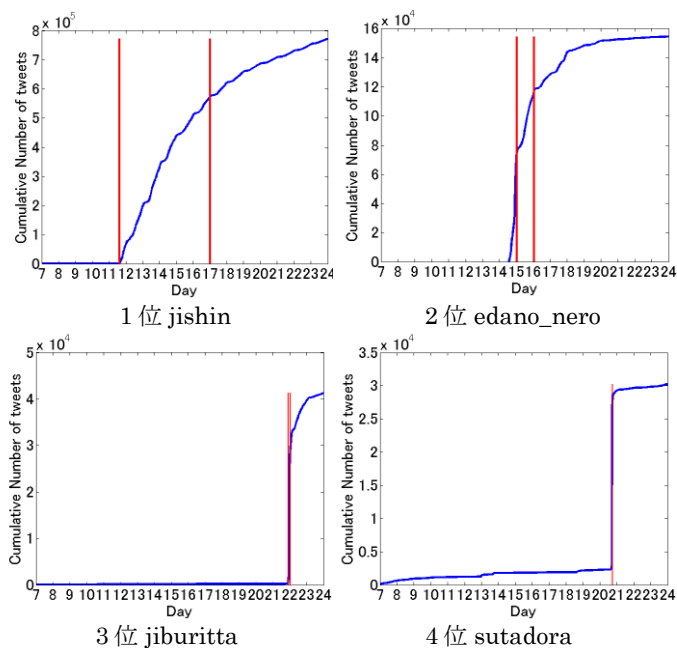


図 2 Hashtag ツイートのバースト期間と累積ツイート数

Hashtag ツイートでは大きな社会影響に左右されず、メディア内の流行や習慣などの比較的小規模な社会影響によりバーストするようなトピックが存在する。このことから、社会影響を反映した変化が得られる点において両者は共通であるが、**Hashtag** は **Mention** 以上に小さな社会影響に対しても敏感に反応してユーザ行動に変化をもたらすことが推測される。

4. おわりに

本稿では、社会影響による **Twitter** 上でのユーザ行動の変化をツイートのバースト期間抽出によって調査し、さらにその類似性について分析した。その結果、社会影響に対するユーザの行動のいくつかの知見が得られた。

今後の課題として、本稿で得られた結果を反映したユーザ行動のモデルの作成、並びに評価を行う予定である。

謝辞

本研究は **NTT** との共同研究および、科研費(22500133)の助成を受けた。

参考文献

- [1] H.Kwak, C.Lee, H.Park, and S.Moon, "What is Twitter, a social network or a news media?", In Proceedings of the 19th international conference on World wide web, pp.591-600. ACM,2010.
- [2] 大原剛三, 齊藤和巳, 木村昌弘, 元田浩 "情報拡散モデルに基づくツイート系列からのバースト期間抽出", 第 3 回ソーシャルコンピューティングシンポジウム,2012.
- [3] 鳥海不二夫, 篠田孝祐, 栗原聡, 榎剛史, 風間一洋, 野田五十樹, "震災がもたらしたソーシャルメディアの変化", 第 7 回ネットワークが創発する知能研究会, 2011.
- [4] 高橋友美, 小出明弘, 伏見卓恭, 齊藤和巳, "Twitter のHashtag に着目した社会影響分析", 第 4 回データ工学と情報マネジメントに関するフォーラム, 2012.