

Twitter におけるバーストの生起要因と類型化に関する分析

Analysis of the occurrence factor and classification of bursty status on Twitter of Infosociomics

水沼 友宏 (Yuhiro MIZUNUMA)¹・池内 淳 (Atsushi IKEUCHI)²

山本 修平 (Shuhei YAMAMOTO)¹・山口 裕太郎 (Yutaro YAMAGUCHI)¹

佐藤 哲司 (Tetsuji SATOH)³・島田 諭 (Satoshi SHIMADA)⁴

¹筑波大学図書館情報メディア研究科 博士前期課程

²筑波大学図書館情報メディア系 筑波大学知的コミュニティ基盤センター 准教授

³筑波大学図書館情報メディア系 教授

⁴法政大学マイクロ・ナノテクノロジー研究センターP・D

[Abstract]

Today, there are many different social communication platforms (eg. Facebook, Twitter, LinkedIn, Google+, mixi and so forth). Internet users are spending more time on social media. Twitter which is a kind of microblog launched in October 2006, has become one of the most popular social communication tool in Japan. Because of Twitter's characteristic features such as immediacy, easiness and bi-directionality, Twitter's timeline instantly reflects the real world. Once a major event happens, the number of tweets rapidly increases. In this article, this phenomenon is defined as 'burst'. The authors gathered Japanese tweets on public timeline from Twitter API over a period of four months starting from March 14, 2012 to July 13, 2012. We collected a total of 1,379,844,343 posts from 58,866,923 distinct users. This study shows that tweets in bursty term are more retweet, less reply and less characters than usual. We have detected four types of bursts by cluster analysis. And the Japanese scale of earthquake and the distance from epicentre to Tokyo have significantly affected occurrence of burst on Twitter.

[キーワード]

マイクロブログ、Twitter、クラスタリング、ロジスティック回帰

1. はじめに

近年、数多くのソーシャルメディア・プラットフォームが公開され、そのユーザ数や接触時間は増加の一途をたどっており[1]、2012年現在、アクティブユーザ数は世界で10億人を越えたことが報告されている。Cosenza[2]によれば、欧米諸国の多くにおいて、Facebook^{*1}が最もトラフィックの多いソーシャルメディアであると推計されているが、日本では、Facebookよりも、マイクロブログの一種であるTwitter^{*2}の利用者数が多い[3]。ユーザの国別トラフィック比率を見ても、日本は米国について第二位となっている。また、2011年の東日本大震災以降、単なるソーシャルメディアとしてだけでなく、災害時のライフラインの一つとしても認知されるようになった。Twitterの即時性、簡便性、双方向性といった特性から、公衆の耳目を集める社会現象が生じた際、しばしば、ツイート数が平常時と比較して、大きく増加する現象が観測される。本研究ではこうした現象を「バースト(burst)」と定義し、ネットワーク上のマイクロブログユーザが種々の社会的事象をどのように捉え、伝達し、報じている

*1<http://www.facebook.com>

*2<https://twitter.com>

のか、さらには、他の情報メディアとどのような関係にあるのかについて定量的および定性的な分析を行う。これによって、Twitter というメディアの特性を描写することを目的とする。

Twitter におけるバーストの分析は Twitter のマーケティングリソースとしての利用やサーバの高負荷予測に関する基礎的知見を与えると考えられる。また、通常 Twitter はフォローしたユーザのツイートがタイムライン上に表示されるため、すべてのユーザが同じ情報を受け取ることは考えにくい。しかし、バースト時は個々のユーザが同じイベントに関してのツイートを行うため、ユーザの多くが類似した情報を受け取る可能性が高く、バースト分析は情報拡散ツールとしての Twitter 活用可能性検討の一助となる。

ここでは、2012年3月14日から7月13日までの4ヶ月間に収集された日本語ツイート計1,379,844,343件(ユニークユーザ数:58,866,923)を対象として、(1)バーストの検出、(2)バースト時のツイート傾向の分析、(3)バーストの類型化、(4)地震とバーストの関係の分析を行う。

本稿の構成は次の通りである。まず、第2章で関連研究を概観し、第3章でデータ収集法と収集データの基本統計、ならびに、調査手法について述べる。第4章では、バーストの類型化と地震とバーストの関係に関する調査結果について詳述し、第5章において全体のまとめと今後の課題について論じる。

2. 関連研究

Twitter の普及にともない、それがどのようなメディアであり、ユーザがどのような行動をとっているのかに着目した研究が多く行われるようになってきている。その嚆矢として、Java ら[4]は、2007年4月1日から5月30日までの2ヶ月間に、76,177のユニークユーザが投稿したパブリック・タイムライン上のツイート(1,348,543件)を対象に、Twitter の構造やユーザの投稿動機などを調査している。その結果、(1)フォロー関係に基づくネットワーク・グラフの直径が6であること、(2)ユニークユーザのうち37,183ユーザ(48.8%)がプロフィールに地理情報を記載しており、うち56.6%にあたる21,640が北米地域のアカウントであること、(3)最も多いツイート内容は「いま、自分が何をしているか」であること、(4)全ツイートの21%が@をともなう会話ツイートであること、(5)同じく13%はURLが含まれた情報共有を意図したツイートであること、(6)リンク構造の解析から、ネットワークのハブであり多数のフォロワーを持つ情報源アカウントや多数のアカウントをフォローするものの、自らはほとんどツイートしない情報収集アカウントなどの存在が識別されたことを明らかにしている。Krishnamurthy ら[5]は、パブリック・タイムラインだけではなく、二つの異なるアルゴリズムを用いてユーザ情報とツイートをクロウリングし、Java らと同様の分析を行うとともに、データセットごとの相違点について言及している。また、Poblete[6]は、246カ国の4,736,629ユーザによる5,270,609,213のツイートを収集し、Twitter のネットワーク構造の国際的な比較を行っている。

ユーザ間の関係や情報伝播といった観点から Twitter の実態を明らかにしようとした研究も少なくない。たとえば、Kwak ら[7]は、2009年6月6日から6月31日にかけて、4170万ユーザのプロファイルと1,470,000,000のフォロー・フォロワー関係情報を抽出している。その結果、77.9%は一方方向のフォローであり、相互フォローは22.1%に過ぎなかった。これは他のソーシャル・ネットワーク・サービスには見られない特徴である。さらに、67.6%のアカウントはいずれのフォロワーからもフォローされていなかった。以上のことから、Twitter はコミュニケーション・ツールというよりもむしろ情報源としての性質の強いことが示唆された。また、フォロワー数とPageRankに基づくユーザのランキングには類似性が見られたものの、被リツイート(RT)に基づくランキングは、それらとは異なる傾向があったことが述べられている。情報伝播やユーザの影響力に関するものとしては、ほかにフォロワー数とリツイートとメンションの影響力と関係を見た Cha ら[8]、ハッシュタグ(#)に着目した Romero ら[9]、リスト機能を用いた Wu ら[10]の研究が挙げられる。

一方、Twitter におけるバーストに関する研究として、ツイート本文に出現する単語からバーストを検出する研究は、幾らか行われている。例えば、Diao ら[11]は、シンガポールのユーザを中心に収集した2011年9月1日から11月30日までのツイート3,967,927ツイートに対しLDA(latent Dirichlet allocation)、および、二つのLDA改良アルゴリズム(UserLDA, TimeLDA)を用いて、バーストする単語を抽出し、抽出した単語からトピックの自動検出実験を行っている。比較の結果、改良アルゴリズムは、より精緻にユニークなトピックを検出することができたと報告されている。また、白木原ら[12]は buzztter^{*3} から流行語を取得し、Kleinberg の提案するアルゴリズムを用いて、流行語を含むツイートが急増する時間帯を検出している。

しかしながら、ツイート数の増減に着目した調査はほとんど行われていない。Twitter 社では[13]、一秒間の

*3Twitter のタイムラインから頻出する語句を表示するウェブサイト <http://buzztter.com>

ツイート数である TPS (tweets per second) を測定し、公式ブログ等においてそのランキングを公表している。また、乾ら[14]は3月11日の東日本大震災発生時付近のツイート 179, 286, 297 ツイートを対象とした分析を行っている。分析の結果、震災後1週間で1分当たりのツイート数が最も多かったのは、3月15日の静岡県東部の最大震度6強の地震のときであり、次に多かったのは同日の三陸沖の地震発生直後であったことを報告している。

3. 方法

調査にあたって、まずバーストの閾値を設定し、閾値にもとづきバーストを検出する。次に、バースト時と非バースト時を比較し、バースト時の特徴を明らかにする。また、特徴量がバースト生起に関わるイベントの性質によって変わるのではないかという仮説から個々のバーストを特徴量に応じて類型化する。さらに、地震によるバーストの生起要因をロジスティック回帰によって明らかにする。

3. 1 分析対象

分析に使用するツイートは、Twitter の Search API^{*4}を用いて収集した。日本語で記述されたツイートを収集するため、言語に“ja”(日本語)と、日本全域をカバーする位置情報^{*5}とを検索条件として指定した。

ツイートに付与される位置情報には、ユーザのプロフィールに自由記述する「location」情報と投稿時に GPS 等の値を自動的に付与する「geocode」情報の2種類がある。位置情報を検索条件とすることで「geocode」が指定した範囲内にあるツイートが収集できる。「geocode」が付与されていないツイートでは、「location」に記入された情報が参照される。Search API では、「location」が実際の地名等と一致しない場合は、デフォルトで東京とみなしていると思われる。実際にツイートを収集してみると、

「geocode」が付与されたツイートは僅かであり、また、「location」情報として「あっち」や「このへん」、「地方」、「ひみつ」といった曖昧な内容が書かれていることもある。このようなユーザのツイートでも Search API に上述の検索条件を指定することで収集できている。分析対象は2012年3月14日から7月13日に収集したデータとし、各々のツイートごとにツイート ID、投稿時刻、ツイート本文の内容等の情報を収集している。ユーザのフォローに関する情報、お気に入り、およびツイートを非公開に設定しているユーザのツイートは、収集にかかる制限が大きいことから、分析に用いるデータとして網羅的な収集は行っていない。

収集したデータの基本統計は表1に一覧する。なお、休日は土曜、日曜、国民の祝日とする。また、リツイート (RT) はツイート本文の先頭に「RT」という文字が見られる場合とし、被コメントを付加した非公式リツイートや QT (引用) によるリツイートは含まない。同様に、@で始まる場合のみリプライ (@) とし、ツイート本文の文頭以外に@ユーザ名が入っている場合は@の比率には含まない。文字数の平均値は45.65文字であるが、最頻値は21文字であった。

表-1 収集データの基本統計

	全体	平日	休日
データ数 (ツイート数)	1, 379, 844, 343	975, 793, 195	404, 051, 148
ユニークユーザ数	5, 866, 923	5, 503, 254	4, 548, 077
本文平均文字数 (文字)	45.65	46.32	44.79
RT 比率 (%)	8.37	8.48	8.09
@比率 (%)	39.27	39.66	38.35

3. 2 バースト閾値の設定

上述のように本研究では、ある要因によって平常時と比較してツイート数が大きく増加する現象を「バースト (burst)」と定義する。バースト要因は必ずしもひとつだけではなく混在して生起する場合がある、という点には留意すべきだが、巨視的な観点から調査を行うにあたり、収集したツイートについて本文を全て確認することは

*⁴<http://search.twitter.com/search.json>

*⁵兵庫県西脇市を中心とする半径2,000km 圏内

困難である。そこで一定の閾値を定め、閾値を越えた場合をバーストとする。閾値を越えたものに関しては適宜ツイート本文を確認し、バースト要因を特定した。また、ツイート数は時間によって増減が見られるため、全ての時間に同じ閾値を適用することは望ましくない。そのため、バースト検出のための閾値は次式のように設定した。

$$N_{nt}(t) = \bar{N}(t) + 3\sigma(t) \tag{1}$$

ある時間(t)における閾値を $N_{nt}(t)$ とする。 $\bar{N}(t)$ はある時間(t)における一日当たりの平均ツイート数であり、 $\sigma(t)$ はある時間(t)における標準偏差である。ある時間(t)における最大値、最小値ははずれ値として閾値算出のデータセットから除外する。なお、ツイートの投稿時刻は秒単位まで取得しているものの、分析にあたっては分単位で処理を行う。

平日と休日では投稿数の分布に差異が見られる。図1は全期間における平日と休日の平均をグラフにしたものだが、これを見て分かるように平日は8時前後と12時台にツイート数が増加し、その後減少している。8時台は朝から情報を確認していることや出勤時間にTwitterを確認していること、12時台は昼食時間にTwitterを利用することによりこのような増加が見られると推測される。しかし、休日は平日のような増減が見られず朝から夜にかけて徐々にツイート数が増加している。したがってバースト閾値は平日と休日を分けて算出する。本調査期間中に検出されたバーストは平日1,692回、休日524回であった。

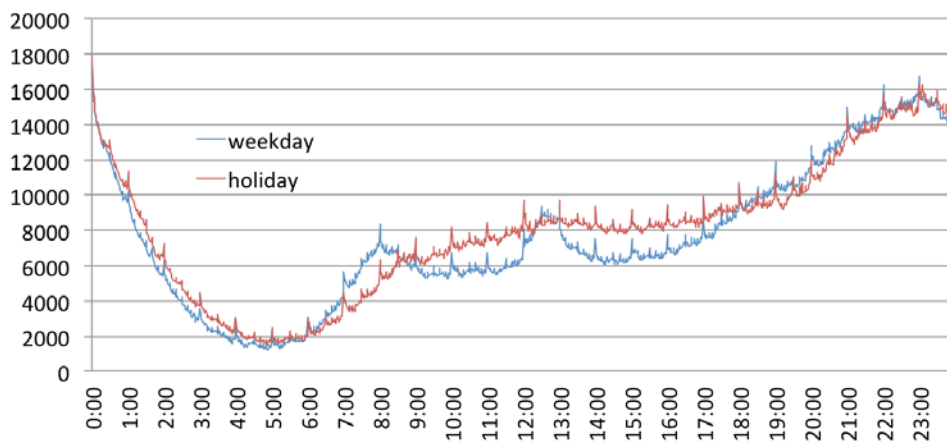


図-1 平日・休日のツイート数の分布

検出されたバーストについて、主なバースト要因について纏めたものが表2である。「徹子の部屋」や「ルパン三世カリオストロの城」、「スマイルプリキュア」などのようなテレビ番組によるバーストが複数回見られる。これに加え、「サッカーオマーン戦」、「旭天鵬優勝」、「フェデラー優勝」といったスポーツ関連のバーストはすべてテレビで放映されていたこと、さらに高橋容疑者逮捕によるバーストはNHKニュース速報の3分後に生起していることからバースト生起へのテレビの影響が窺える。加えて、「アニソン三昧Z」はラジオで放送されていたことからバースト生起と他のメディア、特に速報性の高いメディアとの関連が予想される。一方、Twitter サーバダウンに代表されるように、他のメディアに拠らないTwitter固有のバーストも見られる。災害との関連性も深く、地震、竜巻、ゲリラ豪雨などによってもバーストが生起した。また、ジャスティン・ビーバーのTV出演と東京における集中豪雨によるバーストなど、バースト要因として複数の要因が絡んでいると考えられるバーストも見つかっている。

表-2 対象期間中の主なバースト要因

3月	4月	5月	6月	7月
茨城県震度 5 強の地震(3/14)	エイプリルフール(4/1)	竜巻(5/6)	AKB 総選挙(6/14)	うるう秒(7/1)
徹子の部屋ローラ出演(3/27)	新年度(4/1)	サッカーオマーン戦(5/14)	高橋容疑者逮捕(6/15)	神奈川県震度 3 の地震(7/3)
プロ野球横浜 DeNA 対阪神戦(3/30)	爆弾低気圧(4/3)	旭天鵬優勝(5/20)	アニソン三昧 Z(6/16)	千と千尋の神隠し(7/6)
ルパン三世カリオストロの城(3/30)	仮面ライダーフォーゼ(4/29)	金環日食(5/21)	台風(6/19)	スマイルプリキュア(7/8)
アナログ放送終了(3/31)	千葉県震度 5 弱の地震(4/29)	ゲリラ豪雨(5/28)	Twitter サーバダウン(6/22)	フェデラー優勝(7/9)

4. 結果と考察

4. 1 バースト時と非バースト時の比較

バースト時のユーザの投稿特性を把握するために (1) ユニークユーザ比率、(2) ツイートの平均文字数、(3) ツイートに占めるリツイート(RT)の比率、(4) ツイート数に占めるリプライ(@)の比率を算出し、バーストが検出されたツイートと検出されなかったツイートとを比較する。ユニークユーザ比率は各々のバースト時におけるツイートに占めるユニークユーザの比率の平均であり、バーストの全時間を通したユニークユーザとは異なる指標である。バースト時と非バースト時、全体の基本統計をそれぞれ表3に示す。

バースト時は非バースト時に比べ平均文字数が少ないことが分かる。これは、バースト時は時間とツイート内容の関係性が強く、即時性が重要度を増すことによると推測される。つまり、バースト時は、ユーザは猶予なく情報の伝達を試み、文字数を最小限にとどめるため、本文が短いという特徴が生じる。例えば、地震時は、「ゆれた」、「jisin」、「こわい」、「キヤー」といった短いツイートが随所で確認された。また、リプライ(@)の割合がバースト時は減少する一方で、リツイート(RT)の割合は増加する傾向にある。リツイート(RT)の比率が高いのはイベントに関する情報を不特定多数へ拡散するというユーザの意図に起因すると考えられる。最もリツイート(RT)の比率が高かったのは3月14日の地震によるバーストで、リツイート(RT)の比率は20.9%にのぼる。この時間のツイート本文を確認すると“RT @zishin3255_2: ■■緊急地震速報(第12報)■■ 三陸沖で地震 最大震度 3 (推定) [詳細] 2012/3/14 18:08:29 発生 M7.0 深さ 10km 東京到達時刻: 18:11:26 (あと約177秒) #緊急地震速報”や、“RT @NHK_PR: 青森県太平洋沿岸 岩手県に津波注意報が出ています”など公共機関の防災情報のリツイート(RT)が多数確認された。このようにバースト時はユーザが情報の拡散を意図してツイートを投稿するためリツイート(RT)の比率が高く、逆に特定のユーザに向けたツイートは減少し、リプライ(@)の比率が減少すると考えられる。Twitter がコミュニケーション・ツールというよりは情報源としての性質の強いことは既にKwak[7]により指摘されているが、バースト時は特に、情報入手、拡散ツールとしての特性が強くなることが分かる。

表-3 バースト時と非バースト時の基本統計

	バースト時	非バースト時	全体
データ数・ツイート数	23,970,301 (1.74%)	1,355,874,042 (98.26%)	1,379,844,343
(1)ユニークユーザ (%)	91.79	91.86	91.86
(2)ツイート平均文字数(文字)	41.36	46.57	46.37
(3)RT比率 (%)	9.27	8.35	8.37
(4)@比率 (%)	31.78	39.40	39.27

4. 2 バーストの特徴と類型化

4.1 節ではバースト時は文字数が短く、リツイート (RT) 比率が高く、リプライ (@) の比率が低いという特徴が示された。しかし、これらの特徴量はバーストを生起させたイベントによって変わるのではないかとこの仮説が立てられる。そこで特徴量に応じて、検出されたバーストの類型化を試みる。

類型化に用いる特徴量としては、まず、4.1 節で明らかにしたバースト時の特徴量であるリプライ (@) の比率、リツイート (RT) の比率、文字数が挙げられる。これに加え、バーストの形体もイベントの性質を反映していると考えられる。例えば、図2を見て分かるように地震など予測不可能な現象が起こった際はイベント直後にツイート数が急激に伸び、短期間でイベント前のツイート数まで収束する。一方、金環日食のようにイベント中に最も盛り上がる時間が分かっている場合はその頂点に向かってツイート数が徐々に増加し、頂点を過ぎると緩やかに減少していく (図3)。図4はサッカー日本対豪州戦が開催された6月12日のツイート数の推移を示したものである。試合開始時間である19時より前から通常時に比べるとツイート数が多く、試合終了時間の20時50分以降もしばらくは通常と比較してツイート数が多い。さらにその中でもゴール時や試合終了時はツイート数が急激に増加していることが分かる。特に、日本対オーストラリア戦において本田選手がフリーキックを蹴る前にホイッスルが鳴らされたことによる不満が、試合終了時のツイート数の増加に拍車をかけたことが予測される。最後に、図5に示した爆弾低気圧によるバーストは平均との差はそれほど大きくないものの、長時間バーストが続いている。このように文字数、RT率、@率に加え、ツイート数の増減傾向、バーストの高さや長さがバーストを生起させたイベントに対するTwitterユーザーの認識を表現していると考えられる。

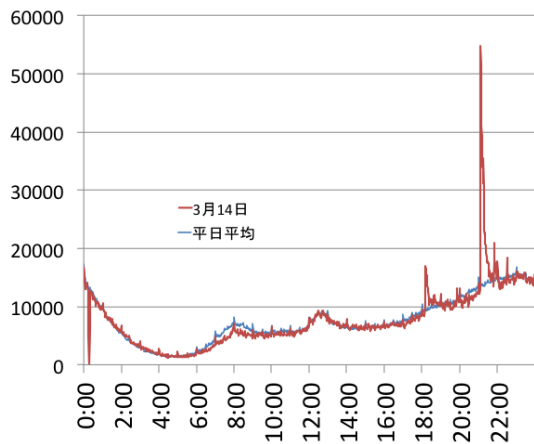


図-2 地震によるバースト

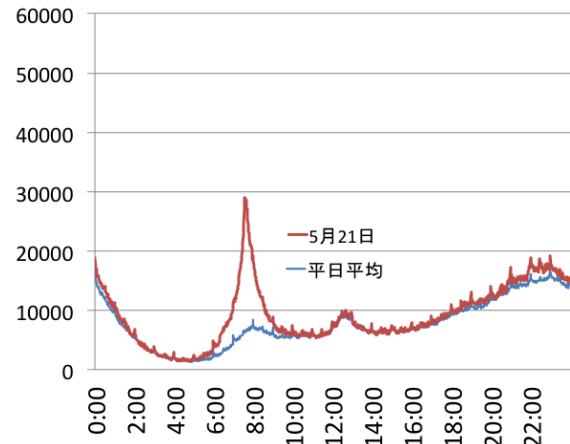


図-2 金環日食によるバースト

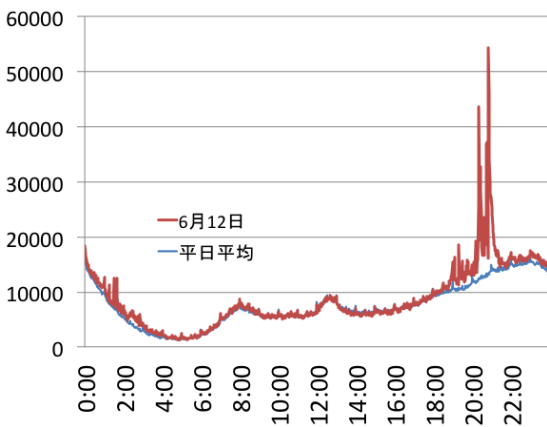


図-3 サッカー日本-豪州戦によるバースト

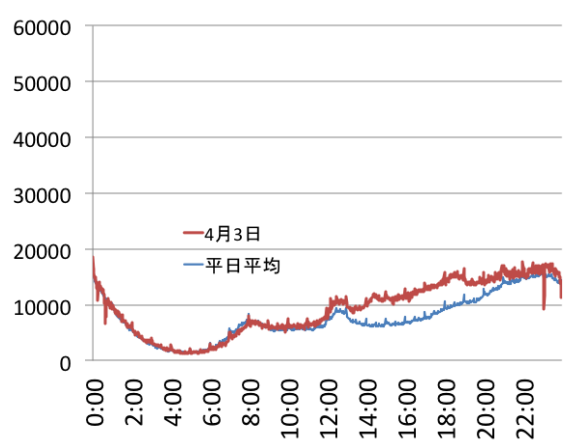


図-4 爆弾低気圧によるバースト

以上のことからバーストの形質がイベントの性質を表しているという仮説が導かれる。そこでバーストを特徴量により類型化し、各々の類型についてイベント性質を把握することを試みる。特徴量はバーストの形体を表す(1)バーストの継続時間、(2)閾値との差、さらに、バースト時の特徴的な傾向である(3)リプライ(@)の比率、(4)リツイート(RT)の比率、(5)ツイート本文の文字数とする。バーストの継続時間とはツイートがバースト閾値を連続して越えた場合の最初に越えた時刻からバースト閾値を下回るまでの合計時間を指す。連続していない場合は、継続期間は1分間である。バーストが継続している場合、その継続期間全体を一つのバースト現象として算出すると、調査期間中405回のバースト現象が観測された。この405回のバーストについて基本統計をまとめたものが表4である。「高さ(ave.)」はバースト期間中の閾値との差の平均値であり、「高さ(max)」はバースト期間中の閾値との差の最大値である。最も継続時間が長かったのは4月3日の331分間で、爆弾低気圧によるバーストと考えられる。

表-4 バーストの基本統計 (継続するバーストを一つのバーストとした場合)

	tweet	@	RT	継続分数	高さ(ave.)	高さ(max)	@比率	RT比率
平均	159,801	43738	24659	11	5,055.5	7,498.1	0.28	0.15
標準偏差	253,663	89,387.5	26793.48	20.51	6,223.2	6,951.1	0.06	0.03
最大値	4,091,191	1,559,183	433,859	331	37,686.2	41,589.2	0.43	0.16
中央値	14,392	4,647	1,091	2	3,860.8	3,594.8	0.30	0.07
最小値	1,896	410	77	1	491.2	491.2	0.14	0.02

次に、特徴量に応じてバーストのクラスタリングを試みた。ここで、継続しているバーストを一つのバーストとすると、同じ要因によるバーストであっても1度閾値を下回った場合は別のバーストとして扱われることが懸念される。また、同じ要因によるバーストでも時間によって特徴量が変わっていくことがある。例えば、災害バーストにおいて、災害直後は被災者が現状をツイートすることによってバーストが生じるが、時間が経過するに従って、災害情報や他のユーザのツイートのリツイートによってバーストが起こる、ということが考えられる。したがって特徴量によって分類を試みる場合、継続期間をまとめて分析するのは不適當だと考えられる。そこで、継続時間については特徴量の一つとして分析することと、クラスタに分類する際は1分ごとのバースト(23,970,301件)の特徴量から類型化する。類似度にはユークリッド距離を、結合アルゴリズムはWard法を用いた。クラスタ分析の結果を表5に示す。

表-5 クラスタ分析の結果

	継続時間	閾値との差	@比率 (%)	RT比率 (%)	文字数
第一クラスタ	149.28	4,839.47	32.38	8.73	40.99
第二クラスタ	33.56	1,729.27	30.80	8.98	43.43
第三クラスタ	41.36	11,624.22	30.83	9.24	39.80
第四クラスタ	40.05	20,372.03	28.79	8.76	36.60

クラスタごとにイベントの特徴を推測すると、第一クラスタは、バースト継続時間が長く、閾値との差がそれほど高くないといった特徴から、爆弾低気圧や大雨、雷雨など断続的イベント型と推測される。第二クラスタは継続時間、閾値との差、リツイート(RT)比率が低く、リプライ(@)比率が高く、文字が長いなど非バースト時と似た傾向を持つ。そのため、イベントを認識したユーザが少ない、小さなバーストと推測される。第三クラスタは継続時間が比較的長く、閾値との差も大きい。これらの特徴から、既知のイベントのバーストがここに含まれると推測される。例えば、金環日食はテレビや雑誌などTwitter以外の多くのメディアがイベントを報道していた。そのためユーザはイベントの実態をイベント前から知っており、図3に示したような特徴的な分布が示される。第四クラスタは閾値との差が最も大きく、リプライの割合が低く、文字数が短いことから突発的イベント発生時のバーストと推測される。地震によるバーストや、サッカーで点が入ることに起因するバースト、AKB総選挙の順位発表によるバーストはこのクラスタに含まれる。

4. 3 地震バーストに影響を与える要因

Twitter は即時性といった特性をもち、140 文字という字数制限もあいまってユーザがツイートした直後に他のユーザのタイムラインにツイートを表示することができる。災害時に於いても例外ではなく他のメディアに比べ情報がいち早く伝達すると考えられる。そのため、Twitter と災害の関連性を明らかにすることは災害情報の迅速な提供に資すると言える。Sakaki ら[15]は位置情報とツイート内容から災害状況の推定、提供を試みた。しかし、バーストと災害の関係と言った観点からの分析はこれまでほとんど行われていない。

災害時はバーストが起りやすく、調査期間中も台風、豪雨、地震等災害に起因するバーストは複数回観測された。なかでも地震が原因と考えられるバーストは 28 回に上る。そこで、地震によるバーストとそれに影響を与える災害の要素を検討する。

地震直後のバーストは揺れに反応したユーザがツイートすることによって発生するため、バースト有無に影響を与えられる要因としてまず震度が挙げられる。これに加え、都心からの距離もバースト発生の有無に関わる要因の一つと考えられる。これは都心に近いほど Twitter ユーザも多く、ユーザ数が多い場所はそれに比例してツイート量も増加するためである。

図 6 は震度 5 以上の地震が起きた日のツイート分布に最大震度を記録した都道府県名を記載したものだが、同じ震度でもツイート数に差があることが分かる。また、5 月 24 日に青森県で最大震度 5 強を記録した地震より、5 月 29 日に東京都渋谷区で最大震度 4 の地震を記録した場合の方が、ツイート数が多く、これらは都心からの距離によると推測される。また、後述するように、本研究では都心の代表地点を東京都庁とするが、東京都庁から最も近い地震観測点である東京都新宿区歌舞伎町観測点では、調査期間中、震度 1 の地震が 8 回、震度 2 の地震が 9 回観測された。このうちバーストが検出されたものは 16 回にのぼり、都心が揺れている場合は高い確率でバーストが検出されることが分かる。

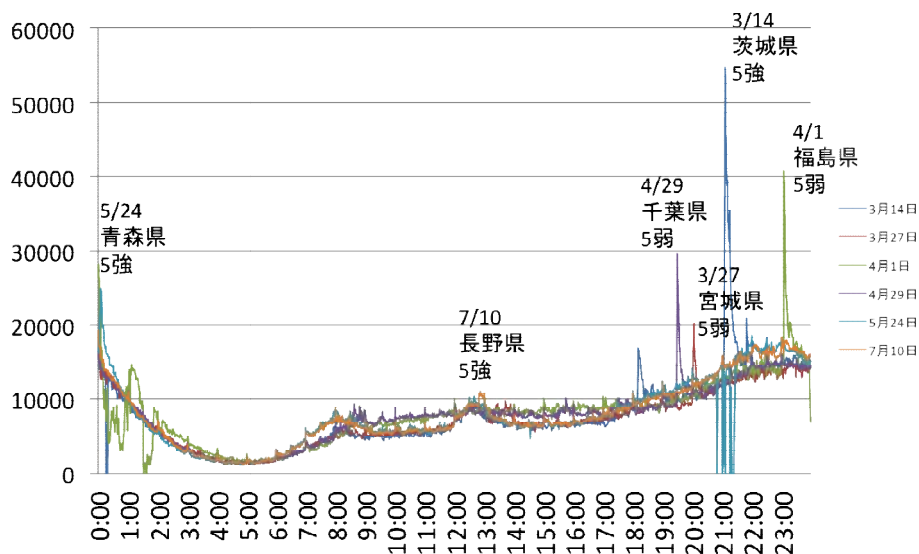


図-6 震度 5 以上の地震が起きた日のツイート分布

そこで、震度、都心からの距離がバーストの生起に影響を与えるという仮説を検証する。調査対象期間中に日本国内で最大震度 3 以上を記録した 114 回の地震について、(1) 発生時刻、(2) 震源地、(3) マグニチュード、(4) 最大震度、(5) 最大震度を記録した市町村、(6) 最大震度を記録した市町村から都心までの距離についてそれぞれデータを収集した。調査対象期間中に発生した地震とその震度、最大震度を記録した市町村は日本気象協会の提供する地震情報^{*6}を参照する。本調査では都市部の代表地点を東京都庁とし、都庁から最大震度を記録した市町村の市町村役場までの距離を算出する。距離の算出には Google Maps API^{*7}を使用した。この際、最大震度

*⁶<http://bousai.tenki.jp/bousai/earthquake/>

*⁷<https://developers.google.com/maps/>

を記録した場所が複数ある場合は都心から最も近い市町村との距離を算出している。なお、代表的な大都市は東京だけではなく、名古屋や大阪なども含まれるが、本研究期間中大阪や名古屋付近での地震が極端に少なかったことから、本研究では都心を東京に限定する。114 回の地震のうち、Twitter 上でバーストが検出されたのは計 32 回であったが、うち、4 回は地震発生以前から他の要因によってバースト状態にあった。そこでそれらを除き、110 回の地震、および 28 回のバーストについて都心からの距離と震度ごとにバースト検知率を纏めたものが表 6 である。

表-6 震度、都心からの距離によるバースト検知率

都心からの距離	震度 3	震度 4	震度 5 以上
100km 以内	54.5% (6/11)	83.3% (5/6)	100.0% (2/2)
100-200km	14.3% (4/28)	50.0% (2/4)	0.0% (0/1)
200-300km	8.3% (1/12)	50.0% (1/2)	100.0% (1/1)
300km 以上	3.1% (1/32)	33.3% (3/9)	100.0% (2/2)

震度 3、震度 4 の場合は都心からの距離が近いほどバースト検知率が高く、都心からの距離とバースト生起の有無との関連が窺える。他方、震度 5 を超えると場所に拘らず高い確率でバーストが検出されていることが分かる。これは、震度とバースト生起の関係性を示唆する。

そこで「最大震度」と「最大震度を記録した市町村から都心までの距離の逆数」を独立変数とし、バースト生起の有無を目的変数としたロジスティック回帰分析を行った。また、分析には R2. 15. 1 の glm 関数を用いた。

Cox-Snell の R² は 0. 481、Nagelkerke の R² は 0. 641^{*}といずれも高い値を示している。また、推定された回帰式に基づく判別率は 86. 1%に上った。各々の独立変数の偏回帰係数はいずれも有意水準 1%で有意であり、Wald 統計量の値などから最大震度よりも都心から被災地までの距離の短さがバースト生起により強い影響を及ぼすことが明らかになった。なお、独立変数間の相関係数は 0. 1 未満であり、各々の独立変数の分散拡大要因 (VIF) はいずれも 2 を下回っていることから、多重共線性の存在も認められなかった。しかしながらこのことは人口の少ない地域における災害をメディアでは矮小化してしまうという危険性を示唆しているとも言える。

表-7 ロジスティック回帰の結果

	-2 対数極度	McFadden's ρ	Cox-Sness' sR ²	Negelkerke' sR ²	判別率
モデルの適合度	78. 99	0. 472	0. 481	0. 641	0. 86
	偏回帰係数	標準誤差	Wald	有意確率	オッズ比
最大震度	1. 820	0. 419	18. 84	1. 419E-0. 5**	6. 171
都心からの距離の逆数	1. 419	0. 283	25. 21	5. 133E-0. 7**	4. 133

5. おわりに

本研究では、Twitter 上でユーザが種々の社会的事象をどのように捉え、伝達しているのか、また、他の情報メディアとどのような関係にあるのか、について Twitter のメディア特性を描写することを目的としてバーストの分析を行った。具体的には、バースト時と非バースト時のユーザのツイート傾向の比較、バーストの類型化、地震バーストに影響を与える要因の算出を行った。

まず、通常時とバースト時のユーザのツイート傾向の比較により、バースト時は非バースト時と比較するとツイートの文字数が短く、リツイート (RT) の比率が高く、リプライ (@) の比率が低いといった傾向が明らかとなった。このことからバースト時には特に情報入手、拡散ツールとしての側面が強くなることが示された。また、バーストの類型化を行い、各々のクラスごとにイベントの特徴を推測した。さらに、最大震度と都心から被災地までの距離の短さがバースト生起の有無に影響を及ぼすこと、特に都心から被災地までの距離の短さがバース

*Nagelkerke の R² は Cox-Snell の R² を 0-1 の値をとるように正規化したもの

ト生起により強い影響を及ぼすことが明らかとなった。

今後は、ユーザに着目したバースト分析を試みる事が課題として挙げられる。地震バーストに影響を与える要因を考えるにあたっては、「揺れに反応するユーザ」と「地震の情報に反応するユーザ」の違いを考慮し、位置情報やツイート時間などを加味した分析を試みる必要がある。また、ユーザ推薦システムに資することを企図してユーザをネットワーク構造やプロフィール情報により類型化し、バースト生起に関する属性ごとの差異等の観点から分析を行う予定である。加えて、より精緻な分析を行うためにバースト時にツイートに含まれる文字列からバーストの要因の特定を行うことも課題としたい。

[参考文献]

- [1] インターネットメディア総合研究所編. ソーシャルメディア調査報告書2011. インプレスR&D, 2011, 156p. (インプレス R&D インターネットメディア総合研究所新産業調査レポートシリーズ)
- [2] Vincenzo Cosenza. World Map of Social Networks. <http://vincos.it/world-map-of-social-networks/>, (accessed 2012-09-30).
- [3] トライバルメディアハウス, クロス・マーケティング編著. ソーシャルメディア白書. 翔泳社, 2012, 251p.
- [4] Java, Akshay., Song, Xiaodan., Finin, Tim., Tseng, Belle. Why We Twitter: Understanding Microblogging Usage and Communities. In Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007. August 12, 2007.
- [5] Krishnamurthy, Balachander., Gill, Phillipa., Arlitt, Martin. A Few Chirps About Twitter. In Proceedings of the First Workshop on Online Social Networks. pp. 19-24, 2008.
- [6] Poblete, Barbara., Garcia, Ruth., Mendoza, Marcelo., Jaimes, Alejandro. Do All Birds Tweet the Same? Characterizing Twitter Around the World. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management. pp.1025-1030, 2011.
- [7] HaewoonKwak, Changhyun Lee, Hosung Park, and Sue 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, 2010.
- [8] Cha, Meeyoung., Haddadiy, Hamed., Benevenutoz, Fabr'icio., Gummadi, Krishna P. Measuring User Influence in Twitter: The Million Follower Fallacy. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. pp.10-17, 2010.
- [9] Romero, Daniel M., Meeder, Brendan., Kleinberg, Jon. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter. In Proceedings of the 20th International Conference on World Wide Web. pp.695-704, 2011.
- [10] Wu, Shaome., Hofman, Jake M., Mason, Winter A., Watts, Duncan J. Who Says What to Whom on Twitter. In Proceedings of the 20th International Conference on World Wide Web. pp.705-714, 2011.
- [11] Diao, Qiming., Jiang., Jing, Zhu, Feida, Ee-Peng Lim Finding Bursty Topics from Microblogs. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. pp.536-544, 2012.
- [12] 白木原渉, 大石哲也, 長谷川隆三, 藤原博, 越村三幸. Twitter における流行語先取り発言者の検出システムの開発. 情報処理学会研究報告 情報学基礎研究会報告. 2010-IFAT-99(2), p. 1-8, 2010.
- [13] 日本語版 Twitter 公式アカウント(@twj)によれば、日本テレビ系列で宮崎駿監督の「天空の城ラピュタ」放映中の 2011 年 12 月 9 日(水)に、TPS が 25,088 の世界最高記録を更新したことが報告されている <https://twitter.com/twj/status/146751303584980992>, (accessed 2012-09-30).
- [14] “FrontPage/Project311/トレンド分析”. 乾・岡崎研究室. 2012-10-13. <http://www.cl.ecei.tohoku.ac.jp/index.php?Project%E3%83%88%E3%83%AC%E3%83%B3%E3%83%89%E5%88%86%E6%9E%9>. (参照 2012-02-20)
- [15] Takeshi,Sakaki., Makoto, Okazaki., Yutaka, Matsuo. EarthquakeShakes Twitter Users: Real-Time Event Detection by Social Sensors. in Proceedings of International Conference on World Wide Web. pp. 851-860, 2010.

(2013年2月24日受理)