Vol.2012-ICS-168 No.3
2012/9/19

東日本大震災時におけるリツイートの分析

鳥海 不二夫[1,a)]　篠田 孝祐[2]　榊 剛史[1]　風間 一洋[3]　栗原 聡[4]　野田 五十樹[5]

概要：本論文では，東日本大震災時の前後に Twitter に投稿された約 4 億のツイートを用いて，震災が Twitter を用いた情報共有行動に与えた影響を分析した．その結果，震災直後からリツイートの利用が増加し，また情報源となるユーザが変化し，単発の情報提供者が増加したことが明らかとなった．また，リツイートの時系列を混合正規分布を用いてモデル化し，震災直後にはリツイートが行われるタイミングが短くなり多くの情報が素早く大勢のユーザに共有されたことを明らかにした．以上の結果より，東日本大震災後の Twitter には集合知を用いた情報共有ツールとしての役割が与えられたと考えられることが明らかとなった．

キーワード：ソーシャルメディア, Twitter, 東日本大震災, 時系列分析，混合対数正規分布

Analysis of Retweet under the Great East Japan Earthquake

Abstract: In this paper, we analyzed the 400 millions of Tweet data which posted around the Great East Japan Earthquake to find how the twitter used and how the Twitter was influenced by the disaster. We modeled the time series data of Retweet by Log Normal Mixture Model. By analyzing the model, we found that the peak times of the retweets are become shorter, and there are few long range retweets after the disaster. As a result, we can say that the role of the Twitter was changed from communication tools to information sharing tools since the Great East Japan Earthquake were occurred.

Keywords: Social Media, twitter, the Great East Japan Earthquake, Time Series Analysis, Log-Normal Mixture Model

1. はじめに

近年ソーシャルメディアと呼ばれる WEB 上のサービスが増加している．その中でも，Twitter を始めとするマイクロブログは近況をつぶやくというこれまでにない情報共有の形を示しており，新しいコミュニケーションツールとして注目されている．このような中，Twitter などマイクロブログに関する研究が盛んに行われている．Java ら [3] は Twitter のソーシャルネットワークを分析し，Twitter の利用目的は日常的な会話と情報の共有であることを明らかにしている．Kwak ら [4] は 4000 万人分のユーザデータと 14.7 億の社会的関係性に基づいて，Twitter における社会的ネットワークの分析を行い，その特徴を明らかにするとともに，リツイートの構造を分析しリツイートが広まっていく様子を分析している．また，社会的にインパクトの強いイベントが発生した際に Twitter がどのように利用されたかについても研究が行われている．震災など緊急時の Twitter 利用に関する研究としては，2010 年チリで発生した地震の際にどのように Twitter が使われたかを分析した Mendoza らの研究 [5] などがある．また，Heverin ら [1] は，2009 年にワシントン州シアトルで発生した警官 4 人殺人事件の際に Twitter がどのように利用されたのかを分析している．

このような中，2011 年 3 月 11 日に発生した東日本大震災はソーシャルメディアがさまざまな目的で広く活用され

[1] 東京大学
the University of Tokyo
[2] 理化学研究所
RIKEN
[3] NTT 未来ねっと研究所
NTT Network Innovation Laboratories
[4] 大阪大学
Osaka University
[5] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
[a)] tori@sys.t.u-tokyo.ac.jp

た．特に，Twitter は震災時に大きくクローズアップされ，情報の流通に大きな影響を与えたといわれている [8]．その一方で，デマ情報なども多く飛び交い，その信頼性は必ずしも高くは無かった [6] とも言われている．また，災害の大きい地域では直接的なコミュニケーションが増加した一方で，そのほかの地域では情報の拡散が積極的に行われ [7]，短縮 URL システムを用いた情報共有も数多く行われ [2]，Twitter が情報共有ツールとして使われていた可能性が示唆されている．そこで，本研究では震災前後の期間におけるツイート情報を大量に取得することによって，情報共有という視点から Twitter の変化を分析する．特に，ユーザのツイートやリツイートなどの行動に関する統計的情報の時系列変化に着目し，震災前後で Twitter 上でのユーザの情報共有行動がどのように変化したかを明らかにする．

本論文では，まず第 2 章で研究に用いたデータの収集方法について述べ，3 章でリツイート行動の基本情報の分析結果を述べる．次に 4 章でリツイートの情報源となるユーザがどのように変化したかを分析する．第 5 章では，リツイートの時間遅れのモデル化を行い，震災の前後でリツイートのタイミングがどのように変化したかを示す．第 6 章で本論文をまとめる．

2. データ収集

本研究では，3 月 5 日〜3 月 24 日までの日本語で投稿されたツイートの収集を行った．収集には TwitterAPI を用いた．収集の方法は以下のとおりである．

(1) 当該期間までに 200 件以上ツイートを行ったユーザを列挙する．
(2) 各ユーザについて 200 件ずつツイートを収集する
(3) 全ユーザの収集が終了した時点で，はじめのユーザに戻り改めて未収集のツイートを最大 200 件収集する

これによって，対象となるユーザのツイートに関しては，概ね網羅的に収集が可能となる．ただし，リストが一周する間に 200 件以上ツイートしているヘビーユーザについては全ツイートを収集できてはいない．また，東日本大震災直後の 3 月 12 日以降，計画停電などの影響により一部データの収集に失敗している．

そこで，それらのデータについては後日当該期間にツイートが収集できていないことが明らかとなったユーザに関して，再収集を試みた．ただし，TwitterAPI の制限により最大で 3200 ツイートまでしかさかのぼることができないため，震災時から収集時までにそれ以上のツイートを行ったユーザについては一部データが欠落している．また，3 月 5,6,24 日についてはデータが完全ではないことが分かっている．そこで，本論文では収集した 3 月 7 日 00:00:00 から 3 月 23 日 23:59:59 までの計 362,435,649 ツイートを利用し分析を行う．

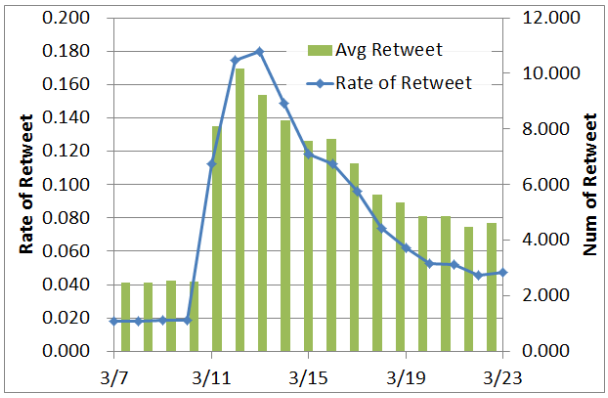

図 1 リツイート率と，平均リツイート数

3. リツイートの基本情報

3.1 リツイート利用率の変化

本論文で扱うリツイートは，ツイッターに含まれる機能の一つであり，他のユーザが投稿したツイートを自分のフォロワーに伝達するための手段である．リツイートには，公式のリツイート機能を利用したものと，他のユーザの投稿をコピーすることでリツイートと同様の記述を行う非公式リツイートが存在する．本論文中では，公式リツイートをリツイートとして扱うこととする．ただし，今回用いたデータには「どのツイートに対するリツイートか」という情報が含まれていないため，得られたリツイートと思われるツイートについて，過去のツイート群から元となるツイートを推定し，リツイート関係があるものと判断した．

東日本大震災時にリツイートがどのように行われたのか，その基本的な情報について述べる．まず，全体のツイートに占めるリツイートの割合および，リツイートされたツイートにおける，1 ツイートあたりの平均リツイート数を図 1 に示す．Rate of Retweet は全ツイートに対するリツイートの割合である．これより，震災直後からリツイートの割合が，1.8% 程度から 18% 程度にまで増加したことが分かる．一方，Avg. Retweet は 1 ツイートあたりの平均リツイート数を示している．震災直後から，リツイート率が増加するとともに，1 ツイートがリツイートされる数も増加していることが分かる．

次に，図 2 に，各日に最も多くリツイートされたツイートのリツイート数を示す．これより，日によって違いはあるものの，震災前のリツイート数は最大でも 5000 程度なのに対し，震災後は 10000 リツイート以上されるツイートが存在している．ここからも，震災後は震災前と比べ上方の拡散力が増加し，より幅広い人に情報を伝えようという力が働いたと考えられる．

3.2 公式リツイート利用の呼びかけの効果

リツイート数が増加した原因の一つに，Twitter 上で公

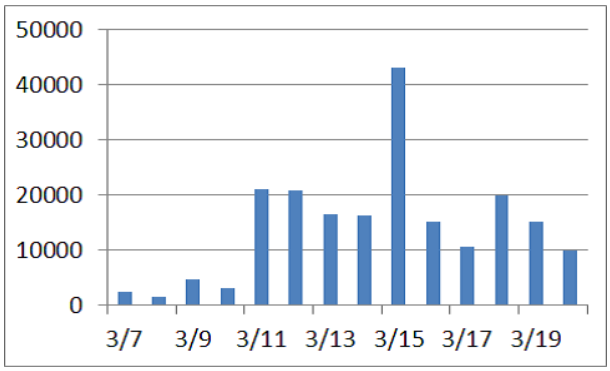

図 2 日ごとの最大リツイート数

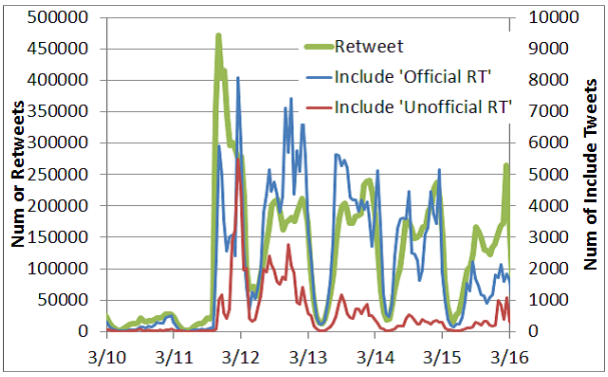

図 3 リツイート数と，「公式リツイート」「非公式リツイート」を含むツイート数

式 RT を利用するよう呼びかける動きがあった[*1]ことの影響が考えられる．すなわち，それまでほとんどリツイートが行われていなかったのは，リツイートの存在をユーザが知らなかったためであり，震災をきっかけにリツイートの存在を知り，リツイートを積極的に利用するようになった可能性がある．

そこで，Twitter 上で行われたこのような呼びかけが，リツイートの増加に効果的であったかどうかを確認するため，分析を行った．ここでは，ツイート内に「公式 RT」または「非公式 RT」が含まれるツイートは呼びかけに使われていた可能性が高いと考え，これらの単語を含むツイート数と公式 RT 数の関係を分析する．「公式 RT」または「非公式 RT」が含まれるツイート数とリツイート数とを一時間単位でプロットしたものを図 3 に示す．これを見ると，公式 RT，非公式 RT を含むツイートのピークは 3 月 11 日 23 時に存在している．一方，リツイートのピークは 3 月 11 日 16 時である．すなわち，震災直後からのリツイート増加はリツイートの利用が呼びかけが原因とはいえない．リツイートが増加した理由は呼びかけによるものではなく，自然発生的なものである可能性が高いことがこの分析結果から明らかとなった．

[*1] ITmedia ニュース http://www.itmedia.co.jp/news/articles/1103/12/news013.html など

4. 情報源ユーザの分析

ここでは，どのようなユーザによって投稿されたツイートがリツイートされたのかを分析する．リツイートが情報の伝播であると考えると，情報源となるユーザはどのようなユーザだったのだろうか．もともと，情報源として信頼されていたユーザが震災後も情報源として活用されたのか，必ずしも情報源として利用されていなかったユーザの情報が広く伝播するようになったのかを明らかにする．

ここで，震震災前 (7-10 日)，震災直後 (12-15 日)，震災後 (17-20 日) において，のべ 100 人以上にリツイートが多くなされたツイートを投稿したユーザについて分析を行った．ここでは，最大リツイート数を総リツイート数で割ったものを最大リツイート占有率と定義し分析を行う．最大リツイート占有率が高いユーザは，最も多くリツイートされた一回のツイートで，当該ユーザの受けたリツイートがほとんど占有されてしまうため，一回のツイートが大勢のユーザにリツイートされたユーザであり，単発の情報提供者であると言える．一方，最大リツイート占有率が低いユーザは，最も多くリツイートされたツイートが，当該ユーザの受けたリツイートの大半を占めるわけではないので，複数のツイートが大勢のユーザにリツイートされた定常的な情報提供者であると考えられる．

各ユーザについて最大リツイート占有率を求めたものが図 4 である．

これより，震災前は最大リツイート占有率は 0.12 と 0.95 付近にピークが存在していることが分かる．すなわち，当該ユーザのリツイートされたツイートの中で，最も多くリツイートされたものが，

(1) 総リツイート数の 10%前後であり，平均的にリツイートされているユーザ
(2) 総リツイート数の 95%であり，一回のリツイートがリツイートされたほぼすべてであるユーザ

の二つのパターンに大きく分かれているといえる．

二極化の傾向は震災直後，震災後でも大きく変わってはいないが，最大リツイート占有率の低いユーザの割合が減少し，最大リツイート占有率 0.95 付近のユーザが増加したことが分かる．したがって，一回のツイートが大勢にリツイートされた，単発の情報提供者が増加したといえる．

5. 混合対数正規モデルによるリツイートのモデル化

5.1 ツイート数間隔による時間分布の分析

一般に，リツイートは元となるツイートが投稿されてからそのツイートをみたユーザが興味深いと思ったとき「リツイートボタン」を押すことで行われる．そこで，リツイートがどのようなタイミングで行われたかを分析するこ

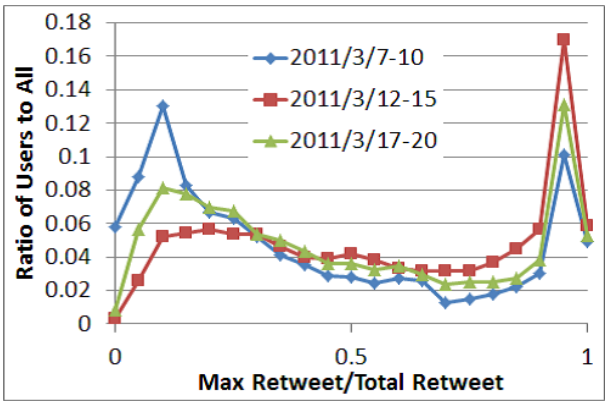

図 4 最大リツイート占有率

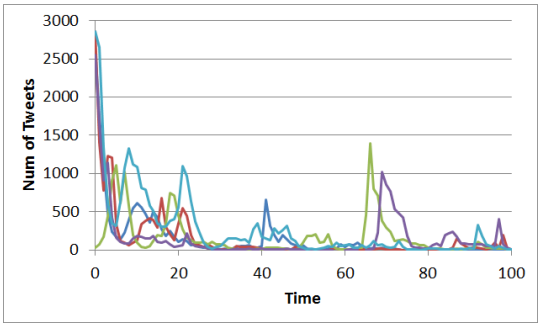

図 5 リツイートの時間分布

とで，リツイートのされ方にどのような特徴があるのかを明らかにし，その特徴が震災の前後でどのように変化したかを確認する．これによって，震災前後での情報伝播のタイミングがどのように異なるのかが明らかになる．

図 5 に，分析期間でリツイート回数が多かった上位 5 ツイートについて，リツイートのタイミングを示している．これは，横軸に時間，縦軸に各間隔内に行われたツイート数を示したものである．ここから，リツイートはツイートが行われた直後に最も多く行われるが，その後も単調減少するのではなく時々多くのリツイートが行われるタイミングが存在することが分かる．

5.2 混合対数正規分布によるモデル化

震災前後で，ツイートおよびリツイートのタイミングがどのような分布になっているかをパターン化するため，各リツイートデータに対し，混合正規分布 (Gausian Mixture Model) によるモデル化を行った．ここでは，観測値 x をオリジナルのツイートからリツイートが行われるまでの時間とし，その分布は対数正規分布，

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \tag{1}$$

によって決定されるものとする．

実際に行われるリツイートは，いくつかの分布が組み合わさってできる混合モデルによって出現すると考え，あるリツイートが行われる時間 x は，

$$p(x) = \sum_{k=1}^{K} w_k P_k(x_i|\mu_k, \sigma_k^2) \tag{2}$$

によって表される混合対数正規モデルによって得られると仮定する．ただし，w_k は各対数正規分布に掛かる重みを，μ_k, σ_k^2 は対数正規分布のパラメータである．このようなリツイートの分布について，EM アルゴリズムを用いて混合対数正規分布を推定することで，各リツイートの時間分布をモデル化することが可能である．

あるツイートに対するリツイートの時間分布を $\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$ としたとき，混合対数正規分布における EM アルゴリズムは以下の通りである．

E-Step:パラメータ μ, σ 及びある分布の寄与度 ω を用いて，あるツイート x_i が k 番目の対数正規分布に属する確率を，

$$z_{ik} = \frac{\omega_k f(x_i|\mu_k, \sigma_k^2)}{\sum_{l=1}^{K} \omega_l f(x_i|\mu_l, \sigma_l^2)} \tag{3}$$

とすると，対数尤度比 $Q(\mathbf{x})$ は，

$$Q(\mathbf{x}) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{ik}^{(t)} \log \omega_k f(x_i|\mu_k, \sigma_k^2) \tag{4}$$

となる．

M-Step:パラメータ $\mu_k, \sigma_k, \omega_k$ は，以下のように更新される．

$$\mu' = \frac{\sum_{i=1}^{N} z_{ik} \ln x_i}{\sum_{i=1}^{N} z_{ik}} \tag{5}$$

$$\sigma' = \frac{\sum_{i=1}^{N} z_{ik} (\ln x_i - \mu_k)^2}{\sum_{i=1}^{N} z_{ik}} \tag{6}$$

$$\omega' = \frac{1}{N}\sum_{i=1}^{N} z_{ik} \tag{7}$$

以上の E-Step と M-Step を交互に，μ, σ, ω の値が十分収束するまで繰り返す．

このようにして推定されたモデルを分析することで，震災前後でリツイートのされ方がどのように変化したかを確認する．なお，分析対象となるツイートはリツイートが 100 回以上行われたものに限定して行う．混合モデル数 K は赤池情報基準量 (AIC) を用いて決定した．

図 6 に混合正規分布モデルによってモデル化を行った例を示す．この図では，もっともリツイートが多かったツイートについて混合正規分布モデルを推定し，モデルに基づいて点をプロットしたものを元のデータと比較している．これより，概ね実データを表現できるモデルが構築されていることが分かる．

5.3 混合対数正規モデルの分析

リツイート数が 100 以上であった 34852 ツイートについて，推定されたモデルの分析を行った．混合モデル数の最

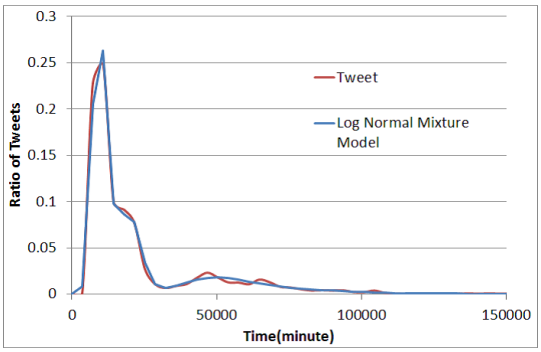

図 6 混合対数正規分布によるモデル化

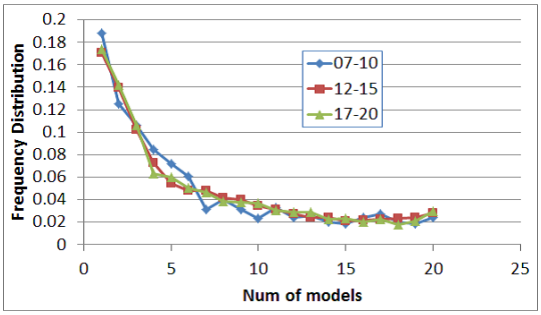

図 7 混合モデル数分布

大を 20 とし，その中で AIC 最小のものをモデルとして採用した．

まず，図 7 に各リツイートについて，混合モデル数の分布がどのようになっていたかを，震災前 (3/7-10)，震災直後 (3/12-15)，震災後 (3/17-20) でそれぞれ示す．これより，多くのリツイートが混合モデル数 1 で表現されており，一回のピークを持ったリツイートであることが分かる．この傾向は震災前後での変化は見られず，半分以上のリツイートについて，その分布がモデル数 5 以下で表現可能であることが分かる．

5.4 最大ピークの分析

リツイートの時間分布における最大のピークの分析を行う．最大ピークとは，混合対数正規分布によるモデル化における ω が最大のものを指す．これによって当該リツイートにおいて最も多くの人が反応したタイミングについて分析することが可能である．

まず，ピークの最頻値及び標準偏差を分析する．ここで，最頻値はピークの最大値が現れる時刻を示しており，最初のツイートからピークが現れるまでの時間遅れを示していると言える．次に，標準偏差はピークの幅を示しており，標準偏差が小さければ多くのリツイートがほぼ同時に起きたことを示し，標準偏差が大きければ長い時間をかけてリツイートが広まっていったことを示す．

図??に，最大ピークにおける最頻値と標準偏差の一日ごとの変化を示す．これより，震災前は最頻値の発生がツイート後 10000 秒以上後にあり 3〜5 時間後にリツイート

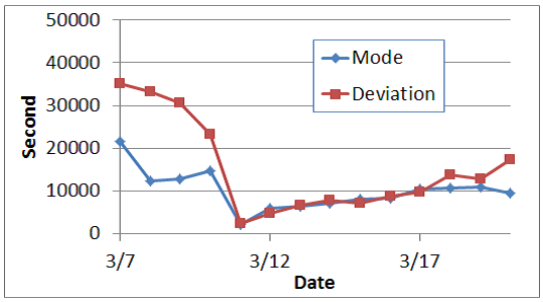

図 8 最大ピークにおける最頻値と標準偏差の変化

のピークが発生していた．一方，震災直後は 2000〜8000 秒後にピークが存在し，30〜2 時間半後にはピークが現れている．特に，3 月 11 日はピークの最頻値が現れるまでの時間が 2032 秒と非常に短い．標準偏差についても震災前は 20000〜30000 だったのに対し，震災後は 10000 以下となり，ピークがより鋭くなっていることが分かる．

以上より，震災発生により，リツイートは短期間に短いスパンで行われるようになったことが明らかとなった．これより，ツイッターが情報共有ツールとして積極的に利用されていたと推測できる．特に，震災直後は多くの情報がすぐに多くのユーザに伝播され，かつ情報自体の寿命は非常に短かったと言える．

6. 終わりに

本論文では，東日本大震災前後の Twitter 上に日本語で投稿されたツイートを取得し，Twitter 上の情報共有行動であるリツイートの分析を行った．平常時には情報源となるユーザは一部の影響力の強いユーザに限られていたのに対し，震災後は多数のユーザからの情報が共有され，それぞれのユーザ自身が情報源となりうる構造を持つようになった．また，リツイートのタイミングを分析した結果，震災後は 1〜2 時間以内にでリツイートされることがほとんどであり，長期にわたって情報として共有され続けるものは減少したことが明らかとなった．震災の発生前後で Twitter の利用方法が「コミュニケーション」から「情報共有」に大きく変化し，特にリツイートを利用した情報共有が活発に行われたことが明らかとなった．特に，有益な情報が多くのユーザの手によって瞬時に共有されていき，集合知的な情報共有が実現されていたとことが明らかとなった．今後，このような Twitter をはじめとするソーシャルメディアの性質を利用し，新しい災害救助支援方法を構築していくことが重要な課題である．

謝辞

本研究は科研費 (24300064) の助成，および NTT 未来ねっと研究所との共同研究による助成を受けて行われたものである．

参考文献

[1] Thomas Heverin and Lisl Zach. Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in Seattle-Tacoma, Washington Area. In *Proceedings of the 7th International ISCRAM Conference*, Seatle, Washington, 2010.

[2] Takeru Inoue, Fujio Toriumi, Yasuyuki Shirai, and Shinichi Minato. Great east japan earthquake viewed from a url shortener. In *Proceedings of the Special Workshop on Internet and Disasters*, SWID '11, pp. 8:1–8:8, New York, NY, USA, 2011. ACM.

[3] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*, pp. 56–65. ACM, 2007.

[4] 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.

[5] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. Twitter under crisis: can we trust what we RT? In *Proceedings of the First Workshop on Social Media Analytics - SOMA '10*, pp. 71–79, New York, New York, USA, July 2010. ACM Press.

[6] 梅島彩奈, 宮部真衣, 荒牧英治, 灘本明代. 災害時 Twitter におけるデマとデマ訂正 RT の傾向. 第 152 回 データベースシステム・第 103 回 情報基礎とアクセス技術合同研究発表会, 2011.

[7] 宮部真衣, 荒牧英治, 三浦麻子. 東日本大震災における Twitter の利用傾向の分析. 第 148 回マルチメディア通信と分散処理・第 81 回グループウェアとネットワークサービス・第 53 回電子化知的財産・社会基盤合同研究発表会, 2011.

[8] 総務省情報通信国際戦略局情報通信政策課情報通信経済室. 平成 23 年版情報通信白書の概要. *CIAJ journal*, Vol. 51, No. 10, pp. 10–15, 2011-10.