

# SNSの共有行動を用いたユーザーの興味モデル化に対する考察 -2013年参議院議員選挙を題材として-

## A Consideration of User Interest Modeling using Re-sharing Activities in SNS

村上 明子\*<sup>1</sup> 榎 美紀\*<sup>1</sup> 斎藤 翔太\*<sup>2</sup>  
Akiko Murakami Miki Enoki Shota Saito

\*<sup>1</sup>日本アイ・ビー・エム株式会社 東京基礎研究所  
IBM Research - Tokyo

\*<sup>2</sup>東京大学大学院情報理工学系研究科  
Graduate School of Information Science and Technology, The University of Tokyo

In social media, resharing contents is one of the important ways for users to propagate information to the other users. In this paper, we presuppose that retweeting behavior of Twitter users shows expected interests of users towards the originator. Based on this assumption, we compare tweets posted by four twitter users during Japanese House of Councillors election in 2013 with during the whole period. We show that tweets including the expected interest of the users tend to be larger than the others.

### 1. はじめに

2013年の参議院議員普通選挙より、ネット上での選挙活動(以下、ネット選挙)が認められるようになった。このネット選挙の解禁により、候補者がネット上で直接有権者の質問に答えたり、ネット上での討論会などを自由に開催できるようになり、より有権者に対して候補者の主義主張を伝えることができるものと期待できる。

選挙活動では、候補者や応援する人(サポーター)の発言をより多くの有権者に届けることが有効となる。FacebookやTwitterといったソーシャルネットワークサービスでは、シェアやリツイートといった発言の再共有をする仕組みがあり、候補者やサポーターのフォロワーに再共有してもらうことで、普段これらの人の発言を読まない人にまで発言を広く知らしめることが可能となる。

Twitterで多くのフォロワーを抱えるユーザーはその影響力の大きさから、ネット選挙などでも重要な情報拡散源、つまり、インフルエンサーになると考えられてきた。情報拡散源としての効果を知るためには、そのユーザーが発信した情報がどれだけ再共有されたか、つまり「被リツイート数」がひとつの指標として考えられ、その数はフォロワー数に相関があると考えられてきた。しかし、実際にリツイートされた数をツイートごとに見てみると発言によって大きく異なり、候補者を宣伝する発言が正しくフォロワー数に比例した効果を発揮しているとは限らない。

例えば、2013年の参議院議員選挙では、フォロワー数90万人を超えるある女性歌手が候補者の一人である伊藤洋介に対して応援するツイートを3件行った。これらのツイートの被リツイート数は一番多いもので878件であるが、公示期間中に被リツイート数が一番多いツイートは彼女の本業である音楽に関するツイート(被リツイート数1629件)である。この被リツイート数の違いは、他のユーザーからの彼女の発言の内容に関する期待の違いとして捉えることができる。このツイートをリツイートした多くが彼女のフォロワーであると仮定すると、彼女のフォロワーは選挙の話ではなく、音楽の話を楽しんでいると考えることができる。

フォロワーの数が多くの人に発言をしてもらったとしても、選挙の話題に興味のないフォロワーしかいない場合は正しく反応が得られない可能性がある。そのため、本稿では、フォロワーが発言者に期待している話題を、過去のリツイートの反応から調べることができないか検討する。また、フォロワーの発言者への内容の期待が被リツイート数などの指標に表れるかどうか、2013年参議院選挙に関わったTwitterユーザーのデータを用いて検証する。

### 2. 関連研究

Twitter上での情報の拡散の分析については多くの研究がある。返信、Retweetなどによる関連でのグループ化に基づいたTweetの情報の流れを可視化した風間ら[風間2010]の研究や、情報伝播の速度や広がりや大きさを内容の類似度から推測する研究[Yang2010]、ハッシュタグに着目し情報の拡散経路を見た研究[Rattananitnont2011]などがある。

公式リツイートの情報はオリジナルの発言とそれを再共有したユーザーしか知ることができないが、榎ら[榎2014]の研究では実際に再共有する元になった直近の情報源が誰かを、Twitterのフォロー・フォロワーネットワークから推測し、真の拡散経路を求めることをしている。

リツイートがユーザーの興味の表れであることを仮定した研究もいくつかある。太田ら[太田2011]は、リツイートしている対象をその人の好む話題とみなし、ユーザーの嗜好を推定している。また、ユーザーによってリツイートされるかどうかを、過去のリツイートのデータを用いて確率的協調フィルタリングモデルで推定するZamanらの研究[Zaman2010b]や、自分の発言と内容の近さやオリジナルの発言者などの指標から判定するYangらの研究[Yang2010a]もある。また筆者らは、どの発言をリツイートするのかによってユーザーの興味をモデル化できるとし、将来のリツイート行動を予測する確率モデルを提案した[村上2012]。

表 1: 2013 年参議院議員公示期間におけるデータ詳細

アカウント名	役割	フォロワー数	全期間			選挙公示期間		
			発言数	最大被 RT	平均被 RT	発言数	最大被 RT	平均被 RT
AbeShinzo	党首	286,048	344	18,310	387.27	96	904	141.75
shiikazuo	党首	31,720	894	2,782	252.37	157	1428	205.45
hmikitani	サポーター	907,112	606	1,064	30.53	168	301	17.92
susumu_fujita	サポーター	159,068	331	215	16.32	18	28	8.95

表 2: 全期間と選挙期間の平均発言数の違い

	全期間平均	選挙期間平均	選挙/全体
AbeShinzo	0.80	5.65	7.03
shiikazuo	2.91	9.24	3.17
hmikitani	1.40	9.88	7.06
susumu_fujita	0.79	1.06	1.34

### 3. 参議院選挙関連ユーザーによる検証

#### 3.1 検証対象のユーザーならびに取得データ

他のユーザーによるリツイートについて調べるため、2013 年参議院選挙でそれぞれの立場で選挙活動した Twitter アカウントを 4 つ選出した。4 名の内訳は政党党首 2 名、候補者のサポーター 2 名である。それぞれのアカウントについて、2014 年 3 月から過去に向かって取得し、3200 件以上ツイートがある場合は 3200 件まで取得した。この中にはそのアカウントによりリツイートされた他のユーザーの発言も混在しているが、今回はこのアカウントからの発言の影響を知りたいため、これらの発言は取り除いた。各ツイートはデータ取得時点での公式リツイート回数の情報を含んでいる。

全アカウントについて 2013 年 6 月以前からのデータが取得できたため、2013 年参議院議員選挙の公示期間 (2013 年 7 月 4 日～2013 年 7 月 20 日) を含んでいる。公示期間中と通常期間をその他を比較するため、全体と公示期間中における発言ツイート数、最大被リツイート数、ならびに平均被リツイート数を調べた。2014 年 3 月時点でのユーザー情報 (役割、フォロワー数) と共に表 1、全期間と選挙期間の平均発言数を 2 に示す。

党首の 2 アカウントを比較すると、AbeShinzo のほうが shiikazuo よりもフォロワー数も多い (約 9 倍)。全体のデータにおいては最大被リツイート数も平均被リツイート数も多いが、フォロワー数ほどの差は見られない。AbeShinzo の全体での最大被リツイート数であったツイートは東京オリンピック決定のツイートで、2 位のツイートの 4 倍近い数であった。

選挙期間中は、両アカウントとも全期間より多く発言しているが、AbeShinzo のほうが選挙期間中積極的に発言している。選挙公示期間は両アカウントとも全体より平均被リツイート数は減っているが、shiikazuo のほうが最大被リツイートも平均被リツイート数も多く、平均被リツイート数の減少も穏やかである。このことから、選挙期間中の shiikazuo の発言は AbeShinzo よりもフォロワーの期待する内容により近かったのではないかと期待できる。

サポーターの 2 アカウントは、選挙期間中に発言が増加し

た hmikitani と比較し、susumu\_fujita は、ほぼ変化がなかった。しかし、両アカウントとも選挙公示期間には最大被リツイート数も平均被リツイート数も大幅に減少している。2 アカウントとも選挙公示期間中は多くのツイートを選挙関連のツイートで占められていた為、本来フォロワーが期待していた内容と異なってしまったのではないかと推察できる。

#### 3.2 被リツイート数上位ツイートの特徴

フォロワーが興味を持った話題をリツイートすると仮定すると、被リツイート数の多いツイートに含まれる話題は、フォロワーが発言者に期待している話題であると考えられる。被リツイート数上位ツイートに含まれる話題を知るために、形態素解析を用いてツイート中の一般名詞を抽出し、全体と被リツイート数上位ツイートそれぞれの頻度の高い名詞上位 10 位を抽出した。全期間の頻出名詞を表 3 に、選挙期間中の頻出名詞を表 4 に示す。被リツイート数上位ツイートとは、発言を被リツイート数でソートして上位から取得したものであり、それぞれ件数で全期間は上位 5%、選挙期間は上位 10% のツイートとした。また、被リツイート数上位ツイートから取り出した頻出名詞のうち、全体の発言に頻出名詞として現れていない名詞を太字で示した。なお、頻度が 1 以下のものについては 1 つのツイートにしか現れていないと考え、表から除外した。選挙期間中の susumu\_fujita に関しては件数が 20 件未満であったため調査対象外としている。

全期間での全体発言とリツイート上位発言の比較を見ると、AbeShinzo は「祈り」や「神社」といった単語から靖国神社参拝に関するツイートが多くリツイートされていたことが分かる。shiikazuo は「反対」「自民党」「秘密」などが特徴的に現れているが、これは自民党が提出した「秘密保護法」への反対に関するツイートが多くリツイートされた事により頻出語となっている。このように特定のトピックが見て取れるが、内容としては政治や選挙に関連する話題であるといえよう。

サポーターの 2 アカウントのうち、susumu\_fujita はもともとリツイートされる数も多くなく、被リツイート数上位ツイートでも「経済」「リーダー」という語が目立つものの他は全体と比べて相違がない。一方で hmikitani は本人の所有する野球の球団に対する話題に関してフォロワーがリツイートすることが多いことが見て取れる。

選挙期間中を見ると、AbeShinzo が街頭演説や会場といった選挙関連のイベントの情報の話題が多いのに対し、shiikazuo は他の党との比較の話題や自身の主張に関する発言への話題が被リツイート数上位ツイートに多いということが分かる。

サポーターの 2 人は選挙期間中は選挙に関する話題が多く、hmikitani はリツイート上位の発言に関しても、特に野球についての話題は見当たらない。susumu\_fujita については選挙期間中も日々のツイート数はさほど変わっておらず、自身の普段の話題が少なくなっていたのではと考えられる。

#### 4. 考察

両期間とも、AbeShinzoの全体での上位に「Facebook」と「写真」が頻出語としてみることが出来る。これはFacebookへ記事を投稿したときに、この記事へのリンクのURLと共にTwitterに発言されているものである。これらの発言は「新しい写真をFacebookに投稿しました [URL]」の形で投稿され、内容をTwitterから知ることはできない。そのため、他の発言に比べリツイートされることは少なく、AbeShinzoの発言でもこれらの発言は被リツイート数では下位に集中している。fujita\_susumuの発言の上位には「ブログ」という語があるが、これはsusumu.fujita ブログサイトへブログが投稿されたときにURLと共にTwitterで発言されたものである。ところが、こちらは被リツイート数上位の話題にも見受けられる。Facebookと異なる点は、「ブログを更新しました。『[タイトル] [URL]』という形でツイートされており、ブログの内容をある程度推測することができる。ブログサイトが発言者の所有するブログサイトであること、シェアした相手に内容を伝えることができることから、リツイートをされるのが比較的多いのではないかと推測できる。つまり、リツイートのされやすさ、またはされにくさ、は話題だけではなくリンクの有無や情報量の大きさにも左右されると考えられる。今後はブログや外部サイト、画像といったリンク先の種別についての調査も必要であろう。

全体では野球に関する話題が多くリツイートされているhmikitaniであるが、選挙公示期間の被リツイート上位ツイートにはその話題が見当たらないことは先に述べた。全期間の発言の中で球団名である「イーグルス」が出てくる発言は23、そのうち9件が上位5%に入っている一方で、選挙公示期間中の球団名が出てくる発言の4件のうち上位10%には1件、5%には0件である。これは全期間で球団名を含むツイートで多くリツイートされたものはリーグ優勝、あるいはリーグ優勝がかかった試合に関するツイートであり、選挙公示期間中はまだそれほど野球が盛り上がっていなかったからだと推測することができる。このため、話題の時期的な盛り上がりなどにも影響も考えなくてはならないだろう。

選挙期間中もツイート数に変化のなかったsusumu.fujitaは、選挙関連の発言は多いものの、支持している候補者名を出している発言は2つしかなく、平均的な被リツイート数である。これは、彼の支持活動がほぼ他の人の発言の自身によるリツイートであり、彼の影響を正しく測るためには、彼がリツイートをを行った発言を読んだフォロワーが、リツイートをした数について調べるべきである。

#### 5. まとめと今後の課題

本稿では、SNS上での再共有の行動はそのユーザーの興味を示していると仮定し、被リツイート数が多い発言を調べることで、あるユーザーがフォロワーから何の話題を期待されているか判明するのではないかと仮説を述べた。2013年参議院議員選挙で候補者の支援活動を行った4ユーザーに対して、何らかの期待される話題が見えるかどうか検討した。被リツイート数上位ツイートと全体ツイートに出現する高頻度な名詞を比べることで、ある程度のリツイートされやすい話題を同定することは可能であると言えた。

本稿では名詞について同義語や異表記のゆれなどについては扱っていない。また「選挙」と「投票」が関連が深いといった意味的な関係も何も用いていないため、表層が異なった語や、意味的に近い語などを鑑みて、人手で話題となっていることを

同定した。今後は、同義語辞書や単語の関係などの外部リソースを用いてこのような問題を解決していく必要がある。

また、AbeShinzoのオリンピック、hmikitaniの野球などのように、一時的に話題となって注目されたものに関しては、通常ずっとフォロワーが期待しているかどうか疑わしい。このような時系列的な話題の変化も考慮に入れなくてはならない。

今回、被リツイート数を得るときに、リツイートをしたユーザーがその発言者のフォロワーか否かは考慮しなかった。発言が多くリツイートされ多くの人に届くためには、まずはフォロワーがリツイートという行動をしなくてはならず、フォロワーの期待の話題かどうかを真に判定する必要がある。そのため、フォロワーのみからの被リツイート数を用いたほうが適切である可能性もある。

今回は語レベルでの発言の再共有への反応について調べたが、意味まで含んだ話題レベルの反応ならびに時系列的な変化を追うことで、今後は再共有、つまり被リツイート数の予測などのモデルの構築を考えていきたい。

#### 参考文献

- [榎 2014] 榎美紀, 村上明子, レイモンドルディー, 小口正人. ソーシャルメディア上の情報拡散分析. In *DEIM*, 2014.
  - [風間 2010] 風間一洋, 今田美幸, 柏木啓一郎. Twitterの情報伝播ネットワークの分析. 第24回JSAI全国大会, 2010.
  - [村上 2012] 村上明子, 鈴木秀幸. Twitterでのretweet情報を利用した情報拡散予測. 第26回JSAI全国大会, 2012.
  - [太田 2011] 太田侑介, 寺田実, 丸山一貴. Twitterにおけるリツイート経路の重ね合わせによるユーザ発信支援. 第10回情報科学技術フォーラム (FIT2011), 2011.
  - [Rattananritnont 2011] Geerajit Rattananritnont, Masashi Toyoda, and Masaru Kitsuregawa. A study on characteristics of topic-specific information cascade in twitter. In *Forum on Data Engineering (DE2011)*, 2011.
  - [Yang 2010] Jiang Yang and Scott Counts. Predicting the speed, scale, and range of information diffusion in twitter. In *ICWSM*, 2010.
  - [Yang 2010a] Zi Yang, Jingyi Guo, Keke Cai, Jie Tang, Juanzi Li, Li Zhang, and Zhong Su. Understanding retweeting behaviors in social networks. In *Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10*, 2010.
  - [Zaman 2010b] T. R. Zaman, R. Herbrich, J. Van Gael, and D. Stern. Predicting information spreading in twitter. In *Proceedings of Computational Social Science and the Wisdom of Crowds Workshop*, 2010.
- \*1 「日本共産党」と「共産党」は同じとみなすこともできるが、表層の違いとして太字としている。

表 3: 全体および被リツイート数上位ツイートにおける頻出名詞（全期間）

	AbeShinzo	shiikazuo	hmikitani	susumu_fujita
全体	Facebook	秘密保護	楽天	ブログ
	写真	国民	日本	社長
	日本	法案	鈴木	藤田
	今日	首相	本日	サイバーエージェント
	秘書	安倍	三木谷	自分
	安倍	日本	寛	今日
	本日	今日	今	起業家
	東京	原発	討論会	話
	国	政府	ネット	堀江
	総裁	問題	経済	今
上位 5%	日本	秘密保護	楽天	ブログ
	祈り	法案	イーグルス	自分
	本日	国民	日本	藤田
	国	首相	ファン	堀江
	心	反対	優勝	経済
	秘書	問題	選手	リーダー
	安倍	自民党	応援	晋
	人々	国会	政治	社長
	今年	声	日本一	条件
	神社	秘密	東北	サイバーエージェント

表 4: 全体および被リツイート数上位ツイートにおける頻出名詞（選挙公示期間）

	AbeShinzo	shiikazuo	hmikitani	susumu_fujita
全体	Facebook	躍進	鈴木寛	ブログ
	写真	演説	日本	今
	秘書	比例	討論会	ネット
	安倍	街頭演説	すずきかん	朝
	候補	日本共産党	投票	選挙
	選挙区	党首討論会	本日	記事
	総裁	明日	参院選	一昨日
	日本	安倍	経済	ボタン
	自民党	選挙区	候補者	明日
	今日	訴え	三木谷	応援
上位 10%	安倍	安倍	日本	-
	日本	自民党	投票	-
	秘書	共産党 <sup>*1</sup>	鈴木寛	-
	闘い	国民	参院選	-
	晋三	政治	選挙	-
	総裁	党首討論会	民主党	-
	街頭演説	問題	三木谷	-
	会場	自民	本日	-
	-	暴走	応援	-
	-	原発	自民党	-