

論文

東京都議会議員選挙におけるツイッターの解析

室谷 心

Analyses of the Tweets on the Tokyo Metropolitan Assembly Election

MUROYA Shin

要 旨

2017年7月2日に行われた東京都議会議員選挙は、小池百合子都知事率いる「都民ファーストの会」が圧勝し、自民党は歴史的な大敗を喫した。本論文は、投票日前後での都議選に関するキーワードを含むツイートを集め、その内容を解析したものである。研究の興味としては、ツイート数の時系列での変化、ツイートに含まれる単語の使用頻度分布、ツイート者数分布、リツイート数、被リツイート数分布などから、代表的SNSであるツイッターシステム上に流れる、都議会議員選挙関連情報の特徴を明らかにすることである。

今回の解析の結果、都議会議員選挙関連のツイートの特徴として、リツイートの占める割合が非常に高いこと、異常に饒舌な参加者はいないことが明らかとなった。また、頻出単語を見る限り、“豊洲移転”や“受動喫煙”といった争点の話題は目立たず、ネガティブキャンペーンの単語の方が頻出していたという結果であった。

キーワード

SNS ツイッター ビッグデータ 都議会議員選挙

目 次

- I. はじめに
- II. 都議会議員選挙のまとめ
- III. 技術的設定
- IV. 結果
- V. まとめ

文献

I. はじめに

インターネットが普及し、だれでも自由に一般大衆に対して意見表明できる状態をティム・オライリーがweb2.0と名付けたのは2000年代中盤であった¹⁾。その後スマートフォンの普及とともに、ブログ、ミクシー、フェースブック、ツイッターといったインターネット上のコミュニティーサービスが広まり、多くのユーザーが存在している。対面ではない多対多のコミュニケーションであり、誰でも情報を発信できて、場合によってはテレビ視聴者数をはるかに超える大勢が、そのメッセージを読んだり聞いたりすることが容易に起きる時代となった。

日本ではインターネットの普及により、平成25年からインターネットでの選挙活動が解禁され²⁾、ほとんどの政党も候補者もフェースブックやツイッターで意見表明し、動画サイトで選挙演説を行っている³⁾。当然各政党は、SNS選挙の分析と対策を行っている⁴⁻⁶⁾。

本研究では、2017年7月2日に行われた東京都議会議員選挙に関連するツイッターを収集し、その特徴の分析を行った。SNSへの書き込みと選挙結果の関係に関しては、しばしば話題に上り、特に全国規模での選挙において、SNS上での予想が非常に正確であることが報告されている^{7, 8)}。したがって、SNSの代表格であるツイッターには世論がよく反映されていることが期待される。

ネット情報から世論への影響としては、選挙に限らずネットでの書き込みの特徴として、「炎上」現象のような集団的先鋭化が指摘されており、ネット上での議論の盛り上がりによって世論が極端なものへと引きずられる可能性が指摘されている^{9, 10)}。しかしその一方で、ネット上では異なった意見を容易に見つけられることから、逆に、バランスの取れた見解を持ちやすいという報告もあり¹¹⁾、ネットでの議論によって選挙結果が大きく左右されるのかどうか、ネット上での書き込みと選挙結果の関係は未だ明白なわけではない。

選挙に関連したツイッターの解析は2013年参議院選挙における毎日新聞の分析¹²⁾以来いろいろ行われていて、例えば今回の都議会議員選挙に対しては、都議会ドットコムというサイトが投票までの一週間のツイートを選挙の争点ごとに分類した分析を行っている¹³⁾。

本研究は、ツイートとしてネット上に流れる情報の特徴を明らかにすることを目指して、ツイート数の時系列での変化、ツイートに含まれる単語の使用頻度分布、リツイート数、ツイート者数分布、被リツイート数分布などを調べた。

II. 都議会議員選挙のまとめ

2017年の都議会議員選挙は、小池都知事を代表とする「都民ファーストの会」が公明党などの支持勢力と合わせて過半数を確保できるかが焦点であった¹⁴⁾。また、森友学園問題、加計学園問題、稲田防衛大臣を始めとする閣僚の失言などで逆風の続く政府与党自民党と、都民ファーストの会のいずれが第一党となるかも注目されていた。

選挙結果は表1のとおりで、新聞のタイトルは「自民惨敗 小池系圧勝」¹⁵⁾、「自民惨敗 過去最低」¹⁶⁾というものであった。都民ファーストの会が第一党となり、公明党など支持勢力が議会過半数

表1 東京都議選党派別当選者数¹⁶⁾

党派名	計	現有議席
自民	23	57
公明	23	22
共産	19	17
民進	5	7
都民ファーストの会	49	6
東京・生活者ネットワーク	1	3
維新	1	1
社民	0	0
諸派	0	0
無所属(都民ファーストの会推薦)	6	9
無所属	0	4
合計	127	126

となった。自民党は過去最低だった38議席を下回り、現有勢力から半減するという結果となり、首相の求心力低下が指摘され1ヵ月後の8月3日内閣改造に至った。

この自民党大敗の背景としては、安倍首相夫妻にかかわる加計学園と森友学園の2つの学校法人問題と、稲田防衛大臣を始めとする複数の大臣の失言が指摘されており、さらに30日夕方には、秋葉原で行われた安倍首相による本都議選初めての立会演説会で、群衆から「帰れ」コールが起こり、そしてそれに対する首相の「こんな人たち」発言があった¹⁷⁾。さらにその場には、森友学園の籠池氏も訪れたことがテレビや新聞紙上で頻繁に報道されていた。これらの文言が、ツイッターではどのように表れているかが、本研究の問題意識である。

Ⅲ. 技術的設定

ツイートの収集と分析については鳥海による総合的な解説¹⁸⁾がある。今回の解析は、PythonとツイッターのAPIを組み合わせてデータの収集を行った¹⁹⁾。ツイッターが提供するAPIにはREST APIとStream APIの2種類がある。本研究では、大量のツイートを効率よく捕まえるために、リアルタイムでツイートを集めるStream APIを利用した。

Stream APIはログオンしてサーバーとのコネクションを確立したままの状態を維持し、新しいデータ(ツイート)が生まれるたびにサーバーからデータが送られてくるpush型のサービスで、Public streamsをfilterモードで利用することによって、つぶやかれた全ツイートを対象として特定の語を含むツイートを選び出し記録することができる。

PythonからのTwitter API利用にあたっては、ライブラリーのTweepyを利用し、通信エラーなどでサーバーとのコネクションが切れた場合には240秒後に再接続するようなシェルスクリプトを設定した²⁰⁾。

ツイートには本文に加えてハンドル名やスクリーン

名、位置情報などいろいろなフィールドの情報が含まれている。今回のデータ収集にあたっては、名前、スクリーン名、本文、時刻、場所、フォロワー数を記録した。必ずしも全てのツイートが全データフィールドを持っている訳ではなく、例えば位置情報は持っていないツイートの方が多かった。Stream APIはツイートがつぶやかれた瞬間に逐次記録していくので、当然ながらそのツイートが何回リツイートされたかというデータは含まれてはいない。

APIで集めたツイートから単語を取り出すには、日本語の文法に則って単語に分けてやる必要がある。今回は形態素解析エンジンとしてMeCab²¹⁾を用いた。また、辞書はmecab-ipadic-NEologd (version 102)²²⁾を用いた。

次節で見るように、今回の解析では「小池」と「小池百合子」、「安倍」と「安倍首相」を別々に数えている。これは「小池百合子」や「安倍首相」を固有名詞として登録してある辞書登録の結果である。したがって数え方や語数の詳細は使用する辞書に依存する。また、ハッシュタグとして#のついた単語、例えば「#都議選終了」は、MeCabでは

#	記号、一般、*、*、*、*
都議選	名詞、固有名詞、一般、*、*、*、都議選、トギセン、トギセン
終了	名詞、サ変接続、*、*、*、*、終了、シュウリョウ、シューリョー

と分解され、URLである<http://www.yahoo.co.jp/>は

http://	名詞、固有名詞、一般、*、*、*、 http:// 、エイチティーティーピー、エイチティーティーピー
www	名詞、固有名詞、一般、*、*、*、WWW、スリーダブリュー、スリーダブリュー
.	記号、一般、*、*、*、*
yahoo	名詞、固有名詞、一般、*、*、*、Yahoo、ヤフー、ヤフー
.	記号、一般、*、*、*、*
co.jp	名詞、固有名詞、一般、*、*、*、co.jp、シーオージェイピー、シーオージェイピー

通過してしまう。従って、ツイートの中身を見て判断をしていないことから混入する、雑音がデータにあると考えるべきであろう。

ツイートの収集システムは2017年6月30日17時34分51秒に起動して2017年7月18日10時45分23秒に停止させた。上記5つのキーワードのいずれかを含む368,562ツイートを保存した。収集データ全体のヒストグラムから判断し、本解析の主要な部分は投票日の前日から2日後までとした。

図1が今回収集したツイートの一日当たりの分布である。左右の端の6月30日と7月18日は時間が短く一日分になっていない。一日当たりのツイート数で最も多かったのは、投票前日の7月1日、ついで投票と即日開票が行われた7月2日で、いずれも一日で約10万ツイートであった。同じデータを1時間ごとの頻度に直したものが図2である。一日の中により細かい構造が見られ、選挙前日の7月1日の夕方と7月2日の夜に大きなピークが現れている。2日のピーク以降ツイート数は急速に減少し、5日以降大きな動き

はほとんど無いが、7月10日に小さなピークが見られる。

図2のグラフで6月30日から7月3日までを拡大したものが図3で、昼間増え深夜に減少するという生活リズムの振動に加えて、7月1日の夕方と7月2日の夜に大きなピークが見られる。

図3で白抜き棒グラフが図2と同じ全データで、ドットパターンの棒グラフは本文中に「秋葉原」という語を含むツイートである。また、斜線パターンの棒グラフは本文中に「小池」を含むツイートの分布である。

投票前日の秋葉原での安倍首相の立会演説会が、7月1日午後4時に大きな話題となり、徐々に関連ツイートが減少していく様子が、ドットパターンの棒グラフの変化だと考えられる。一方で「小池」の話題は全体を通じて500ツイート/時以下であったが、7月2日の夜9時に5,000ツイート/時以上に増加する。これは「都民ファーストの会」が圧勝した開票速報の結果に対する反応と考えることができる。

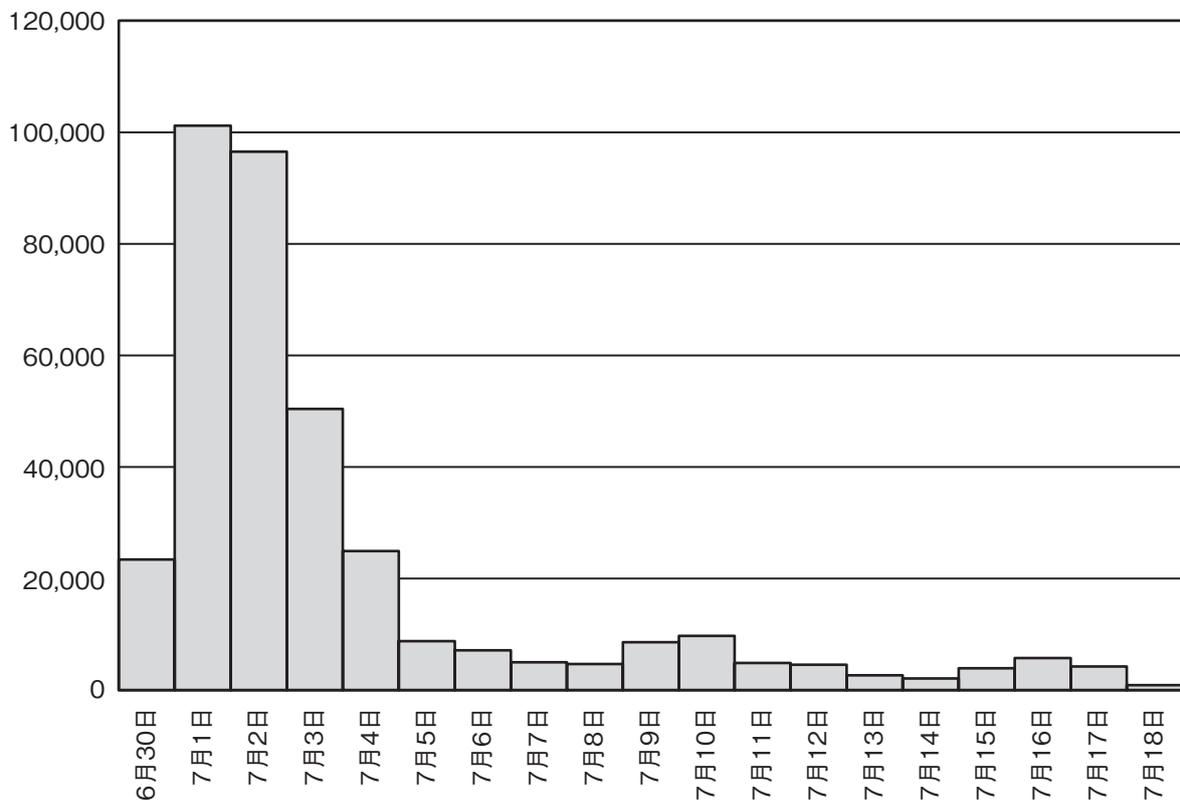


図1. 収集した都議選関連ツイート数分布 (1日毎)

同じ期間のツイートを、「安倍」または「自民党」を含むツイート（以下データ1とする）と「小池」または「都民ファースト」を含むツイート（以下データ2とする）に分けてプロットしたものが図4である。データ1、データ2の選択はPythonによる文字列処

理なので、「安倍首相」は“「安倍」を含む”という条件に含まれ、「都民ファーストの会」は“「都民ファースト」を含む”という条件に含まれる。さらに上記の2つの条件は排他的にはなっていないので、データ1とデータ2の両方にカウントされるツイート

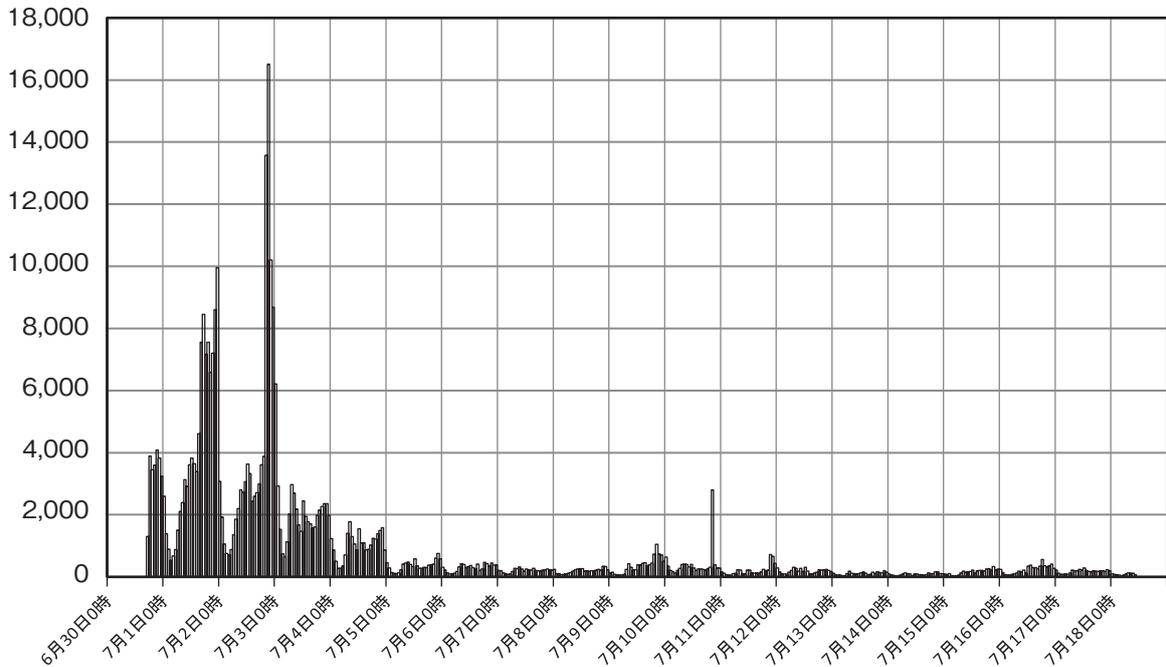


図2. 今回収集した都議選関連ツイート数分布 (1時間毎)

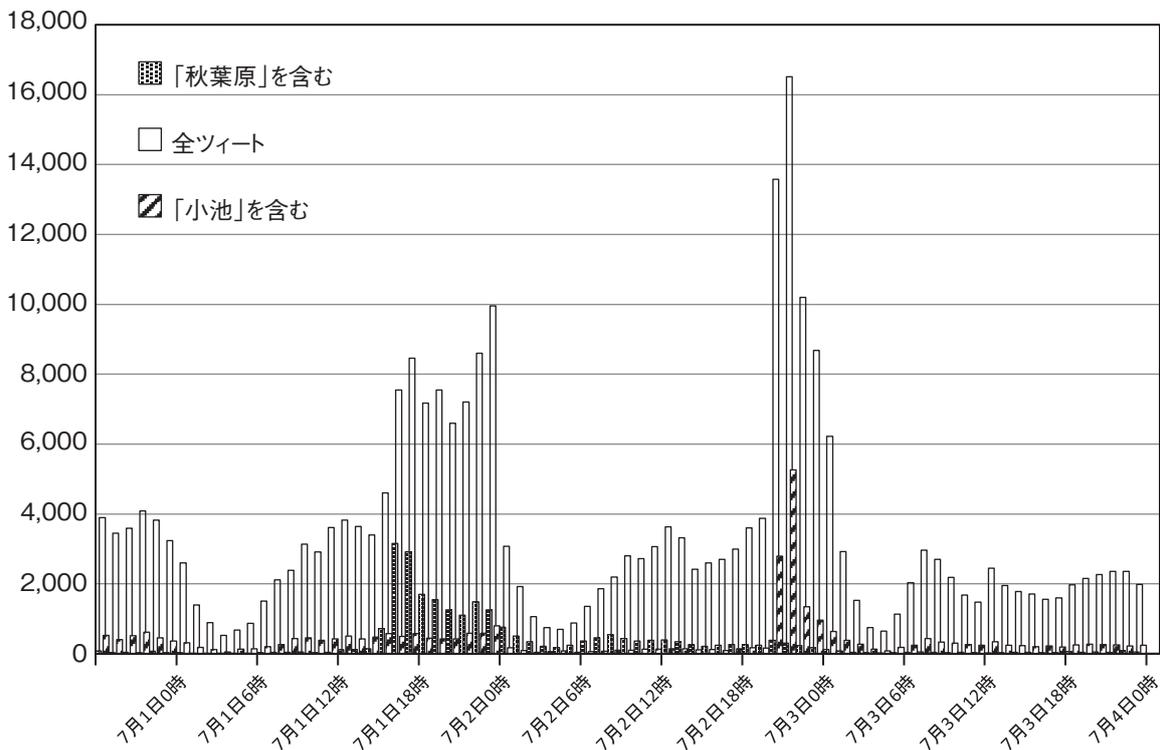


図3. 6月30日から7月3日までのツイートの時間ごとの分布

も存在する。Pythonによる文字列処理とMeCabによる単語頻度の数え方では違いがあることに注意が必要である。

投票前日のピークではデータ1の自民党関連ツイートが多く、開票速報の時間では自民党関連のツイートも急増しているが、データ2の「都民ファースト」関連のツイートの方がはるかに多く、データ1の2倍程度になっている。1日夕方のピークでのデータ1も2日夜間のピークでのデータ2も、いずれも全関連ツイートの半分程度である。この2つのピークの時期以外では、「安倍」「自民党」を含むツイートと「小池」「都民ファースト」を含むツイートは同程度で全体の2~3割程度でしかない。

図4の全ツイートについて、MeCabで単語に分け、固有名詞と一般名詞の出現頻度を数え得た結果が表2の一番左欄（全ツイート）である。MeCabによる分類で、URLの一部分であるhttpやt、coも名詞として扱われている。前節でも述べたが、今回使用した辞書は新しい辞書で「都民ファースト」や「都民ファーストの会」が登録しており、ただの「都民」

とは区別して扱われている。本文中の単語使用頻度なので、一つのツイート中に複数回出てくればその数だけ数え、ツイート数よりも多くなることがある。

ここでは個々のツイートの中身を確認してはいないが、9位の「秋葉原」、24位の「人」、29位の「コール」、30位の「街頭演説」は、投票前日の秋葉原での安倍首相の街頭演説での「帰れコール」とそれに対する首相の「あんな人」発言についてのツイートと考えられる。安倍首相の秋葉原での街頭演説は予定されていたものなので、事前予告もツイートされているが、図4で見る限り、実際に演説が行われた直後のツイート数が圧倒している。表2の期間のツイートでは森友も加計も稲田も上位30位にはランクインしていない。23位のtokyonewsroomは後述するが、ツイッターのスクリーン名で、リツイートの際に使われたものと考えられる。データ1とデータ2について単語の頻度を調べたものが表2のデータ1、データ2の欄である。データ1では、「秋葉原」「コール」「街頭」「聴衆」といった単語が並び、秋葉原での

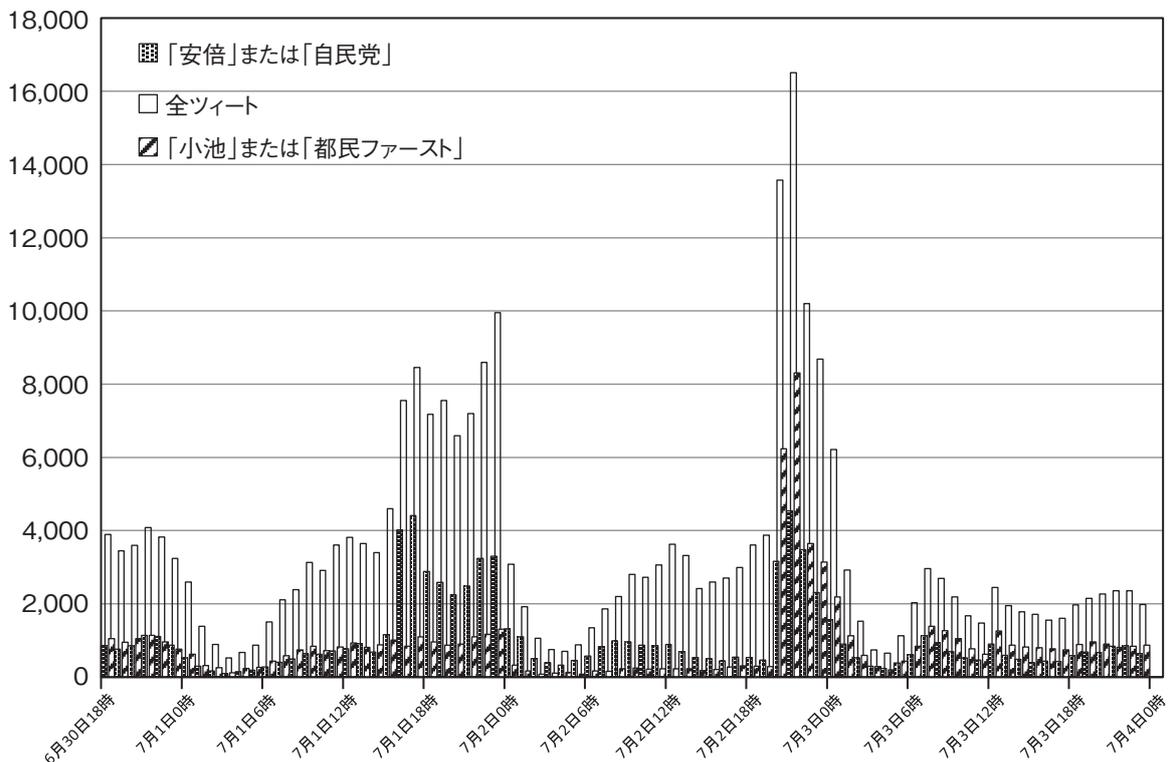


図4. 6月30日から7月3日までのツイートの時間ごとの分布（その2）

立会演説会の話題が多い。一方データ2では「平慶翔」「秘書」「時代」「PC」「下村博文」が表に見られるが、これは特定の候補者に対するネガティブキャンペーンの文言で、このような頻度順の表に現れるくらい大量のネガティブツイートが行われていたことになる。

表2では文献9) に倣って、固有名詞と一般名詞を選び出して数えているが、参考のためデータ1につ

いて形容詞を数えてみると表3のようになる。形容詞の使用頻度は名詞と比べて1桁小さく、頻度順にすれば表2の順位表には載らない。本解析では固有名詞と一般名詞に限定して頻度計数作業を行った。

つぶやいた人の登録名であるスクリーン名ごとのツイート数の回数が表4である。全27万ツイート中で最多ツイートは424件でしかない。したがって、

表2 6月30日18時から7月3日14時までのツイートでの高頻度単語

	全ツイート	270,144	データ1	81,960	データ2	70,549
順位	語	使用頻度	語	使用頻度	語	使用頻度
1	都議選	208,999	都議選	62,560	都民ファースト	45,950
2	https	165,440	自民党	52,192	都議選	40,375
3	t	160,924	https	45,502	https	34,100
4	co	155,190	t	44,290	t	33,472
5	自民党	52,192	co	43,049	co	32,922
6	都民	46,192	安倍	24,356	都民	28,557
7	都民ファースト	45,950	安倍晋三	19,802	都民ファーストの会	19,066
8	安倍	24,356	秋葉原	18,955	小池	16,220
9	秋葉原	23,388	安倍首相	15,309	自民党	15,122
10	候補	20,316	議席	12,011	小池百合子	13,642
11	安倍晋三	19,802	応援演説	11,233	議席	10,471
12	都民ファーストの会	19,066	都民ファースト	9,857	平慶翔	8,078
13	東京	18,751	コール	9,174	自民	6,804
14	議席	17,696	街頭	8,329	民進党	6,384
15	小池	16,220	民進党	8,057	秘書	6,216
16	自民	15,375	めろ	7,711	ki	5,953
17	安倍首相	15,309	安倍総理	6,940	3g	5,943
18	民進党	14,738	共産党	6,906	wa	5,934
19	共産党	13,972	アキバ	6,847	bu	5,929
20	小池百合子	13,642	内閣総理大臣	6,679	TR	5,928
21	応援演説	13,237	人たち	6,319	aDeJM	5,927
22	東京都議会議員選挙	12,821	公明党	6,262	off	5,824
23	tokyonewsroom	11,842	聴衆	6,084	take	5,823
24	人	11,417	安倍政権	5,322	dress	5,822
25	候補者	10,552	kmokmos	5,318	知事	5,735
26	政治	10,493	自民	5,068	PC	5,661
27	公明党	10,268	tokyonewsroom	5,003	時代	5,495
28	都議会	10,090	政治	4,986	下村博文	5,254
29	コール	10,055	街頭演説	4,828	公明党	4,867
30	街頭演説	9,995	最終日	4,648	候補者	4,718

ツイート全体を支配するような、非常に多数のツイートを流す特異なユーザーは存在せず、ツイッターには実際に大勢の人のツイートが流れていると考えてよい。自民党関連のデータ1での突出して多いアカウントからのツイートであっても、関連全ツイート8万件に対して3百件でおよそ300分の1でし

がなく、支配的なほど多いとは言えない。つまり、表2のデータ2で見られた特定の候補者に対する大量のネガティブツイートも安倍首相の秋葉原での対応に対するツイートも、特定の個人ではなく多くのユーザーから出されたツイートの結果ということになる。

表3 データ1に使われている形容詞の頻度

順位	データ1 語	形容詞 使用頻度
1	ない	2,791
2	いい	1,432
3	凄い	1,372
4	激しい	1,148
5	凄まじ	946
6	すごい	919
7	良い	809
8	厳しい	745
9	悪い	668
10	強い	639
11	おおきい	600
12	ほしい	573
13	多い	569
14	なく	545
15	少なかっ	525
16	もの凄く	508
17	大きく	472
18	面白い	424
19	無い	409
20	おかしい	408

ツイッターは、「リツイート」と呼ばれる他人のツイートの引用が容易にできるシステムになっている。リツイートの場合引用元を「RT@スクリーン名」という形で表示するので、テキスト部分の先頭がRT@スクリーン名で始まるデータを数えたものが図5である。白抜きグラフが全270,144ツイートの分布で、ドットハッチ部分218,854ツイートが、本文がRTで始まるリツイートで、全体の81%であった。ここでは、本文の先頭に来るものだけを数えたので、非公式リツイートは含まれていない。

表5は図5のドットハッチ部分のリツイートを集計した被リツイート数ランキングである。表5のトップ20に対するリツイートで、全ツイートの20%を占め、トップ30に対するリツイートで全ツイートの30%となる。表4に見たように、ツイート自体は多くの個人がそれぞれつぶやいていて、特にツイート数の多いスクリーンネームは見られないが、全ツイートの8割はリツイートで、システムの中では、同じ話題の情報が何度も何度も繰り返し流れていたことになる。

表4 スクリーン名の分布

順位	全ツイート	総数 270,144	データ1	総数 81,960	データ2	総数 70,549
	スクリーン名	回数	スクリーン名	回数	スクリーン名	回数
1	@mititosi	424	@aizawanatuko	305	@7thclouds	180
2	@togisencom	309	@brewingbear	189	@kkawaister	167
3	@aizawanatuko	306	@tomo20336301	180	@lovechuprimo	155
4	@gFMEeazPQhaLuYQ	302	@Maeshiro_2011	156	@ikukotaryu	152
5	@yumiyam1	294	@Maeshiro2011	155	@yamakawatanbo	132
6	@Blueyellow_81	292	@mititosi	154	@nonkisaburo	126
7	@beats_groove	274	@masa457028	142	@suzuran211	124
8	@lAx8CrWnqaVGcTw	269	@Study__Japan	138	@masa457028	116
9	@7thclouds	266	@yab887	125	@Amuayumi	109
10	@cell_tm	258	@kmokmos	120	@heisei280622	96

図3 (図4、図5) で大量にツイートが流れた時間帯とそれ以外とを比較するために、7月1日8時から16時、7月1日16時から24時、7月2日12時から20時、7月2日20時から7月3日4時の8時間の時間帯4つを取り出して、単語の使用頻度を比較したものが表6である。ランキングの1位から4位までは全ての時間で共通で、都議選関連のネット情報であることを意味する単語であった。全体を見たところ、政策論争に関わるような争点を示す単語は、特に見当たらない。

時間帯から考えての予想通り、7月1日16時から24時までの間のツイートでは、秋葉原での立会演説会を連想させる単語が多く見られる。籠池氏も秋葉原に表れたと報道されており、籠池という単語も立会演説会関連のツイートと考えられる。7月2日20時から3日4時までのツイートでは、議席や池上、当確といった選挙速報関連の単語が並んでいる。美波や櫛坂46はアイドルグループの櫛坂46の小池美波のことで、小池という名前であるため、filterを通

表5 被リツイート数のトップ20

順位	リツイート	数
1	RT@tokyonewsroom	11,688
2	RT@take_off_dress	8,220
3	RT@wa_bu_ki	5,933
4	RT@kmokmos	5,479
5	RT@togisencom	4,733
6	RT@ozuemura	3,894
7	RT@TMasuda31500754	3,690
8	RT@camomillem	3,034
9	RT@nhk_burakoshi	2,748
10	RT@bilderberg54	2,596
11	RT@sugoi_hi	2,534
12	RT@tocho_senkyo	2,491
13	RT@ulaken	2,116
14	RT@iwakamiyasumi	1,808
15	RT@hiromakikousiki	1,792
16	RT@ISOKO_MOCHIZUKI	1,745
17	RT@HuffPostJapan	1,604
18	RT@umekichkun	1,535
19	RT@thewaytomyplace	1,489
20	RT@TVTOKYO_senkyo	1,460

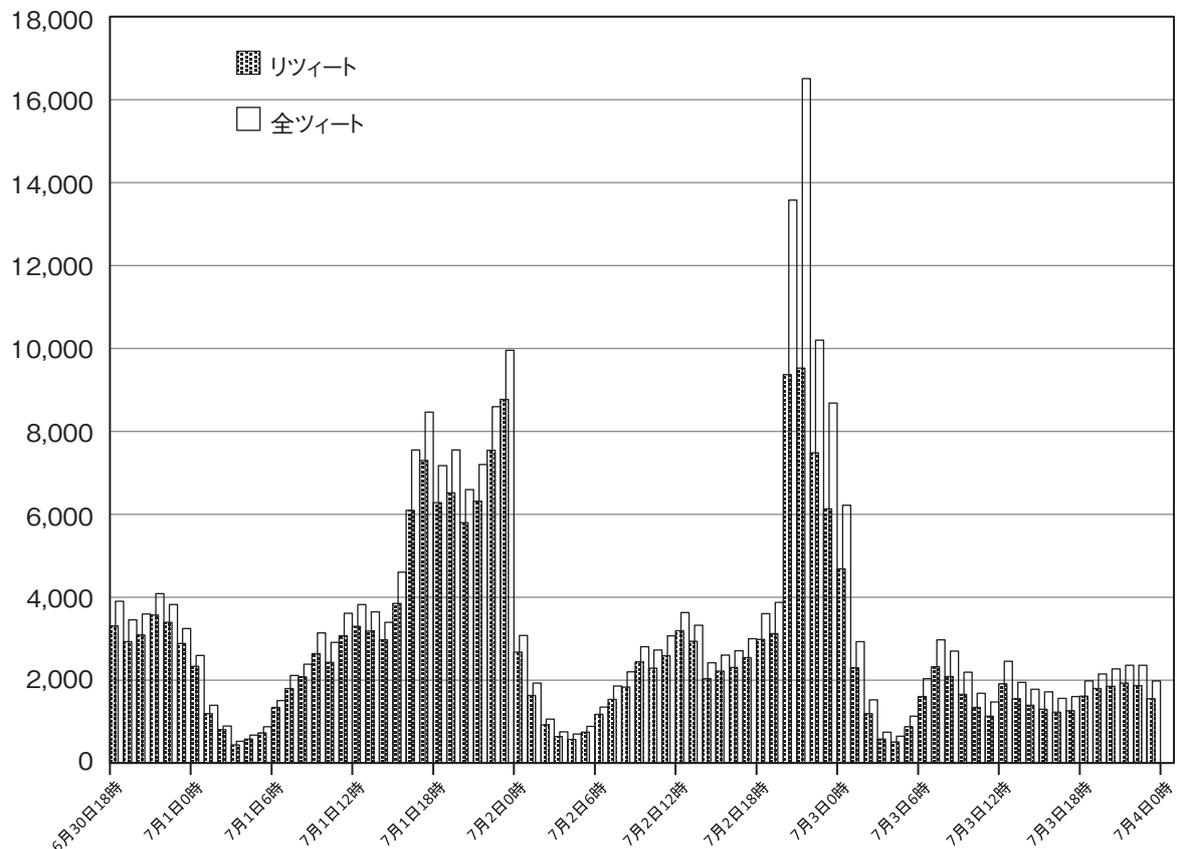


図5. 全ツイート中のリツイート分布

過しデータに入って来たものである。しかしながら、表6にはあるが表2には表れていないので、この時間に特に大量の“櫻坂46小池美波”関連のツイートが紛れ込んだことになる。池上の名前も表にあるが、テレビ東京の池上彰の選挙速報番組の時間も、もともとあった櫻坂46のテレビ東京でのレギュラー

番組の時間とぶつかったことが大量の“小池美波”の原因かもしれない。

リツイートについても表6と同様に4つの区間に分けて数えたものが表7である。表7の被リツイート順位で見ると、「7月1日8時から16時」「7月1日16時から24時」「7月2日12時から20時」については、比較

表6 ツイート分布のピーク直前とピーク領域の8時間での単語頻度

順位	7月1日8時から16時		7月1日16時から24時		7月2日12時から20時		7月2日20時から3日4時	
	tweet数	27,504	tweet	63,082	tweet	25,151	tweet	60,380
1	都議選	21,054	都議選	53,253	都議選	23,709	都議選	48,154
2	https	17,045	https	44,767	https	16,261	https	32,301
3	t	16,295	t	43,213	t	15,760	t	31,977
4	co	15,657	co	42,058	co	14,398	co	31,267
5	自民党	4,608	秋葉原	14,062	自分	3,275	都民	19,274
6	候補	4,134	自民党	11,045	投票率	3,251	都民ファースト	16,052
7	都民ファースト	3,788	安倍	10,656	nhk	2,976	自民党	12,730
8	都民ファーストの会	3,387	安倍晋三	8,822	投票所	2,908	小池	8,649
9	東京	2,168	候補	6,772	前回	2,566	自民	8,462
10	平慶翔	1,845	コール	5,061	tokyonewsroom	2,514	議席	6,294
11	共産党	1,844	安倍総理	5,034	東京	2,375	池上	4,551
12	最終日	1,805	東京	5,017	安倍首相	2,340	安倍	4,111
13	小池百合子	1,708	都民ファースト	4,604	東京都議会議員選挙	2,270	美波	3,570
14	東京都議会議員選挙	1,659	安倍首相	4,548	動画	2,027	櫻坂46	3,437
15	桜井	1,572	街頭演説	4,390	街頭	1,927	当確	3,412
16	候補者	1,568	応援演説	4,383	秋葉原	1,913	配信中	3,345
17	世田谷区	1,567	最後	4,355	自民党	1,908	nhk	3,044
18	築地	1,541	めろ	4,297	7月2日	1,877	Z	2,980
19	都議会	1,536	共産党	4,287	投票日	1,797	tokyonewsroom	2,967
20	小池	1,505	世田谷区	4,161	応援演説	1,765	安倍晋三	2,932
21	街頭演説	1,489	都民ファーストの会	4,110	未来	1,711	都民ファーストの会	2,903
22	都民	1,468	街宣	4,073	都民	1,429	東京	2,854
23	人	1,415	内閣総理大臣	3,818	burakoshi	1,404	小池百合子	2,695
24	純子	1,410	桜井	3,639	ulaken	1,396	NHK	2,555
25	松下	1,394	東京都議会議員選挙	3,610	安倍	1,388	過半数	2,468
26	秘書	1,377	純子	3,504	使い道	1,383	公明党	2,436
27	玲子	1,372	人	3,436	ジワジワ	1,377	ZZ	2,432
28	政治	1,311	籠池	3,311	お小遣い	1,377	開票速報	2,410
29	議員	1,300	政治	3,220	決め	1,377	池上彰	2,368
30	take	1,278	都政	2,937	ネット	1,250	国政	2,269

表7 ツイート数のピーク前とピーク領域の8時間での被リツイート頻度

	7月1日8時から16時		7月1日16時から24時		7月2日12時から20時		7月2日20時から3日4時	
	tweet 数 = 27,504 RT 23,519		tweet 数 = 63,082 retweet 54,948		tweet 数 = 25,151 RT 21,328		tweet 数 = 60,380 RT 41,237	
1	RT@take_off_dress	1,260	RT@kmokmos	2,672	RT@tokyonevnewsroom	2,485	RT@tokyonevnewsroom	2,941
2	RT@tokyonevnewsroom	819	RT@togisencom	1,547	RT@ulaken	1,396	RT@wa_bu_ki	2,123
3	RT@ozuemura	715	RT@take_off_dress	1,529	RT@nhk_burakoshi	1,380	RT@sugoi_hi	1,799
4	RT@togisencom	712	RT@bilderberg54	1,502	RT@ISOKO_MOCHIZUKI	695	RT@mizuya55	1,256
5	RT@support_reiko	507	RT@tokyonevnewsroom	1,421	RT@togisencom	610	RT@nhk_burakoshi	909
6	RT@camomillem	489	RT@ozuemura	1,164	RT@take_off_dress	465	RT@TVTOKYO_senkyo	893
7	RT@OgawaToshioMP	446	RT@Los_HKO	960	RT@ozuemura	418	RT@kmokmos	725
8	RT@Cieletdleau	428	RT@taro_koho	705	RT@ishiitakaaki	418	RT@nikkeilive	720
9	RT@ikukotaryu	392	RT@tanikiri17	699	RT@mynamenamekamikaze	386	RT@nikkei	625
10	RT@akahigepapa	373	RT@camomillem	683	RT@JPN_1st	382	RT@saitohisanori	585

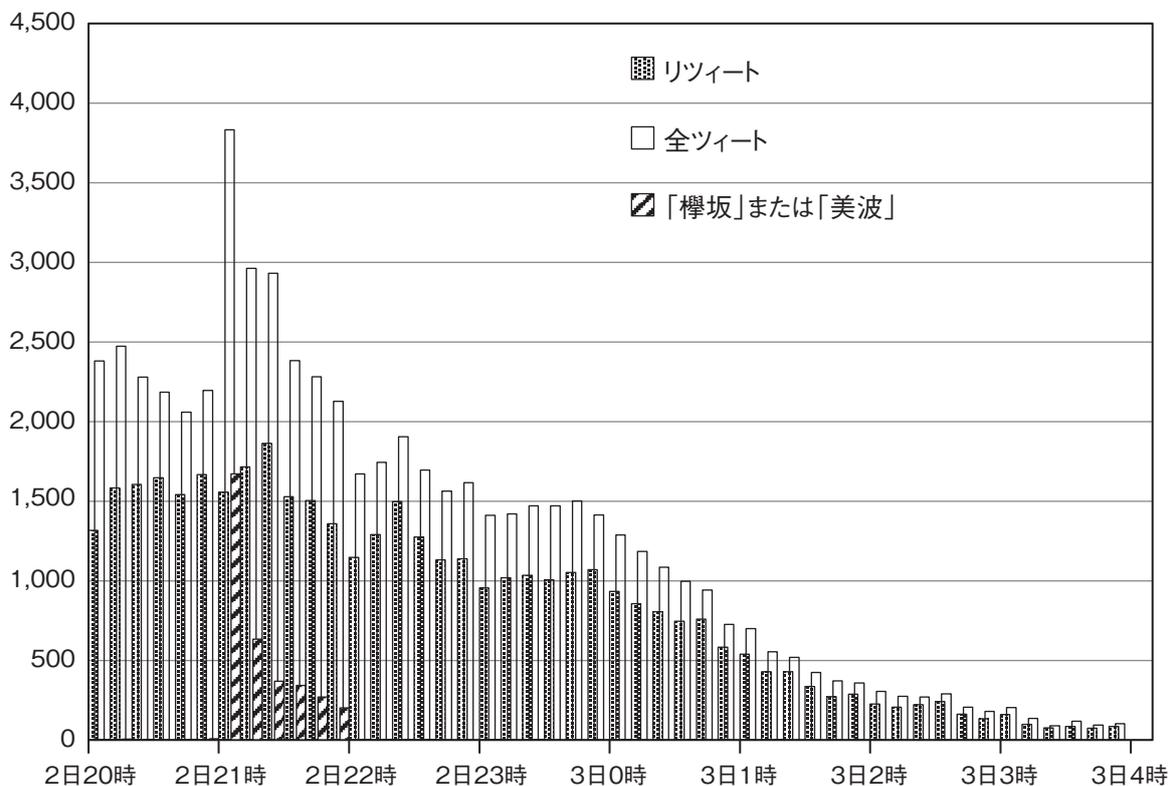


図6. 7月2日20時から3日4時のツイートの分布

的共通のスクリーンネームが見られる。RT@tokyonewsroom、RT@kmokmos、RT@take_off_dress、RT@camomillem、RT@ozuemura、RT@togisencomが複数の時間帯でトップテンに表れており、被リツイート of 常連である。一方で、「7月2日20時から3日4時」の時間帯では、RT@tokyonewsroomとRT@kmokmosは他の時間帯同様にリツイートされているが、そのほかの常連スクリーンネームのリツイートは見られず、RT@TVTOKYO_senkyo、RT@nikkeilive、RT@nikkeiといった速報番組系のリツイートが表れてい

る。この時間帯は開票の最中なので、都議選関連の話題としては、意見や事件等よりも開票速報の事実が一番のリツイート話題であったということであろう。

図5に示したように、今回の都議選関連で集めたツイートはリツイートが多く、全体の80%以上がリツイートである。その中で、「7月2日20時から3日4時」の区間だけは、リツイートの割合が下がるという特徴が見られる。表6に見るように、この時間だけ「櫻坂48」「美波」「配信中」という単語が含まれている。実際にこの時間のツイートデータを見てみ

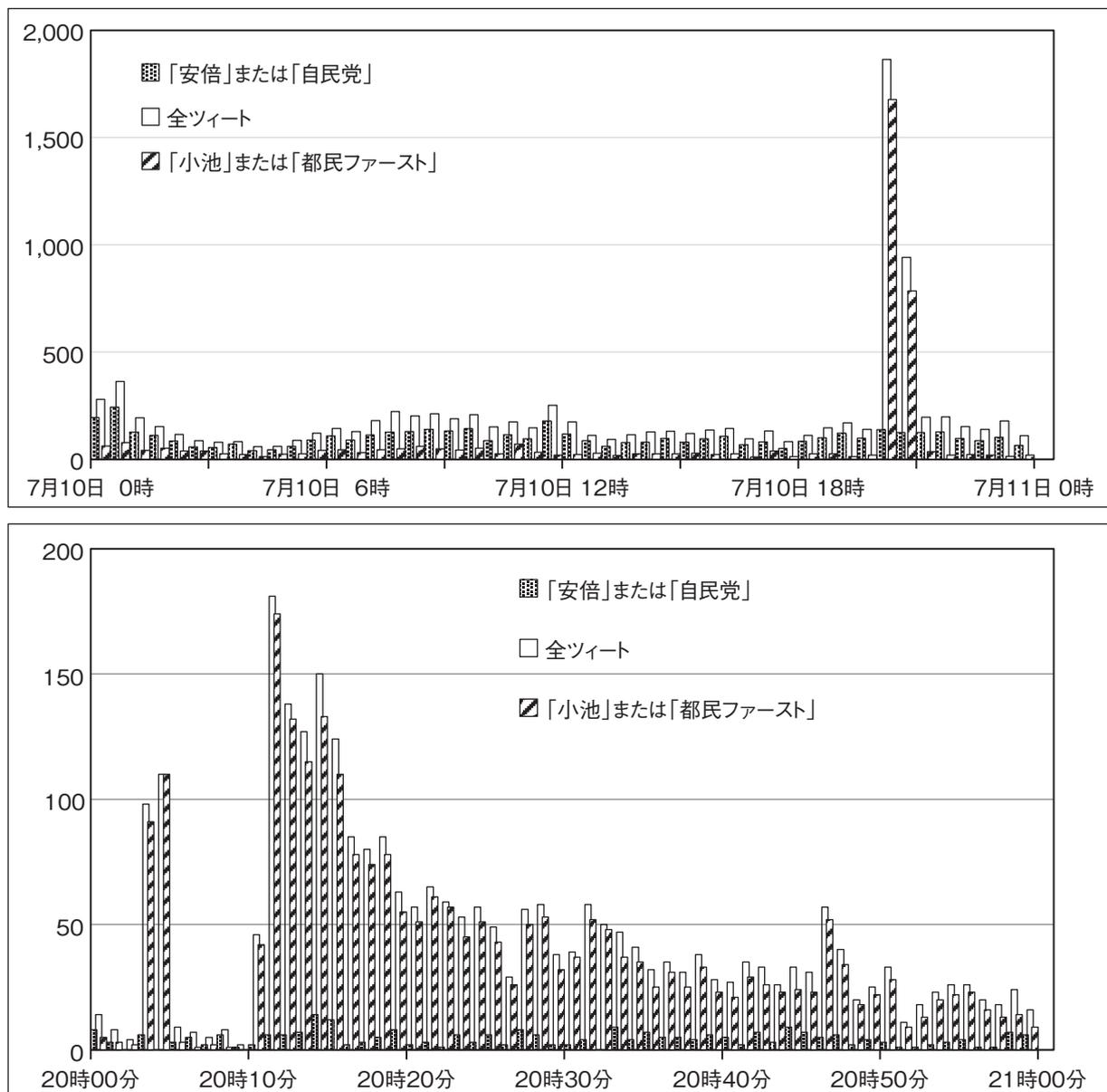


図7. 7月10日のツイート分布。一日のデータ(上)と20時台1時間のデータ(下)

ると、「櫻坂48 小池美波 配信中」というRT@スクリーン名のない、つまりリツイートではないツイートがたくさん見られる。試みに、この時間のリツイートではないツイートのうち、「櫻坂」もしくは「美波」を含むツイートを選び出して数えてみたところ、図6の斜線ハッチ部の分布になった。白抜きが全体のツイートであり、ドットハッチ部がRT@スクリーンネームがついたリツイートの分布である。この時間帯で、特に21時付近にリツイートではないツイートがたくさん起きているが、この特異な増加分は“RTのないツイートで、「櫻坂」もしくは「美波」を含む”ツイートの分布（図中斜線ハッチ部）と一致し、“櫻坂48の小池美波の配信に関するツイート”という選挙関連とは異質のツイートが雑音として混入したためだと考えられる。

実は同様のことが、7月10日にも起こっており、図2で小さいが鋭いピークが7月10日に見られたが、この部分を拡大したものが図7である。図4と同様に、ツイートを、「安倍」または「自民党」を含むツイート（以下データ1とする）と「小池」または「都民ファースト」を含むツイート（以下データ2とする）に

分けて頻度を求めた。白抜きは全ツイート、ドットハッチが自民党系のデータ1であり、斜線ハッチが都民ファースト系のデータ2である。7月10日のツイートは、図7上で見られるように、自民党系で200ツイート/30分程度、都民ファースト系で50ツイート/30分でほぼ一定であるが、20時台に突然都民ファースト系のツイートが急増し、1時間かけて減少していくという現象が起きている。

7図下の20時から21時までの1時間のツイートについて、単語使用頻度を調べると表8の結果となった。表8を見ると、この時間帯のデータでは、「小池美波」「櫻坂46」「配信中」というツイートが支配的であることが明らかである。

従って、7月10日20時台のピークも都議選関連のツイートではなく、「小池」でfilterを通り抜けた櫻坂46小池美波関連のツイートであった。

また、図5同様に、この時間内の全リツイート（白抜き）とリツイート（ドットハッチ）をプロットしたものが、図8である。

これを見ると、この時間帯にたくさんつぶやかれた「櫻坂46 小池美波」関連のツイートにはリツイ

表8 7月10日20時から21時までの1時間のツイートの単語使用頻度

	すべて		データ1		データ2	
	tweet数	2,806	tweet数	263	tweet数	2,463
1	https	2,619	自民党	245	https	2,447
2	t	2,611	https	117	小池	2,441
3	co	2,596	t	107	t	2,440
4	小池	2,441	co	107	co	2,436
5	美波	2,408	塚本幼稚園	82	美波	2,408
6	櫻坂46	2,314	安倍	81	櫻坂46	2,314
7	配信中	2,283	安倍内閣	80	配信中	2,283
8	自民党	245	議員	68	X	83
9	マスコミ	85	内閣支持率	60	Broadcasting	78
10	X	83	ikukotaryu	59	YybLjI	53
11	塚本幼稚園	82	マスコミ	58	U	28
12	安倍	81	世論調査	51	o	27
13	世論調査	81	安倍総理	48	q	23
14	安倍内閣	80	参院	47	i	23
15	ikukotaryu	79	青山繁晴	43	I	23

トが少ないことが分かる。図5と図8とを比較すると、都議選関連ツイートのリツイートの割合の高さが際立つ。

V. まとめ

本研究では、2017年東京都議会議員選挙関連のツイートをストリーム形式で収集し、時間系列分布や単語頻度、リツイートなどについての解析を行った。本研究での主なリサーチクエッションは、下記の3点であった。

1. ツイッター数の時系列での変化。
2. 頻出単語から見た特徴的な話題の存在の有無。
3. 書き込み参加者の動向と、特徴的な書き込み者の存在の有無。

1の時系列については、投票前日1日の夕方と投票日2日の夜にピークがみられた。投票前日7月1日のピークは、自民党関連のツイートで、安倍首相が秋葉原で行った立会演説会において起こった「帰れ」コールとそれに対する首相の「こんな人」発言に関連したツイートであった。図3、図4のドットハッ

チの棒グラフの変化から、立会演説会のあった16時からツイートが集まり、深夜にかけて終焉した。自民党関連と呼んだが、「自民党」「安倍」というキーワードを含むツイートグループに対するラベルであって、自民党支持というわけではない。話題のなり方から言って、批判的なツイートと見て良いであろう。

7月2日のピークは20時で、開票速報と合わせたピークが出るという点では、文献24)で示されている2014年12月の衆議院選挙と同様であったが、我々のデータでは、21時台に突然巨大のピークが現れた。これは、その後の解析で、「櫻坂46小池美波配信中」というツイートが大量に流れ、「小池」というキーワードのために本解析のデータに混入したためであった。図6に見られるように、“リツイートでなく、かつ「櫻坂」か「美波」を含む”ツイートを選り出した、斜線ハッチの分布がなければ、7月2日夜のツイートの分布は、20時に立ち上がり、深夜に向けて滑らかに減衰していく分布になる²⁴⁾。

2に関しては、頻出単語から見られる特徴的な話題は、基本的には投票前日の秋葉原での立会演説会の話題と、開票による都民ファーストの圧勝が中

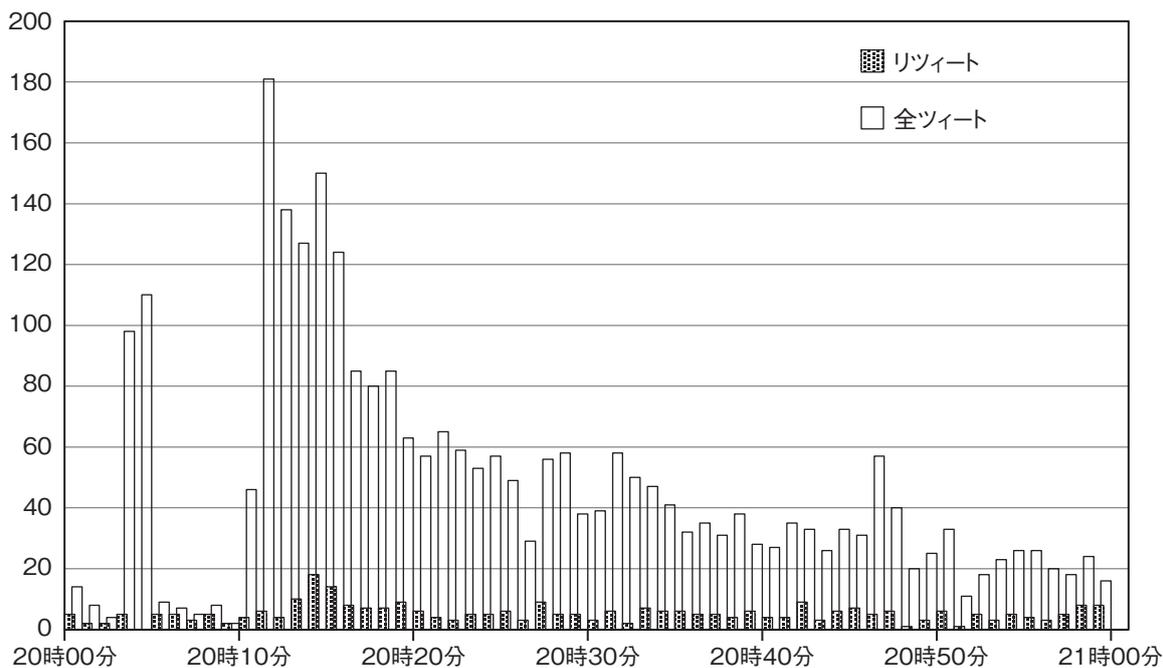


図8. 7月10日20時から21時の間のツイートとそのうちのリツイート

心であった。今回の選挙の争点は何であったのか、文献13)のサイトでは争点に着目したツイート分析を行っているが、本研究の解析では、豊洲移転問題や受動喫煙の話題は頻出単語には見られなかった、自民党関連のデータ1で首相の秋葉原の話題、そして都民ファースト関連のデータ2で「秘書」「時代」「PC」といった特定候補者に対するネガティブキャンペーンの単語が表2にみられる。圧勝した都民ファーストの会の特にポジティブな面は浮かび上がってはこない。表6で見たようにピーク時とそれ以外の時間でデータを分け、語の頻度を比較するとことにより、ピークを形成するツイートの中身が同定できる²⁵⁾。

3の特徴的な書き込み手であるが、スクリーン名ごとの書き込み回数分布で見ると、異常に饒舌な参加者はいない。総ツイート数と合わせて考慮すれば、ツイートというのは本当に大勢の人が参加してつぶやいているシステムであるといえる。その一方で、今回の選挙関連で集めたツイートでは、リツイートが占める割合が非常に多く、図5の期間のツイートのうち81%がリツイートであった。リツイートということはだれかがツイートした話題を後の時刻で再ツイートしていくことなので、同じ意見が鳴り続ける共鳴箱効果があると考えて良いであろう。7月2日21時のピークの解析から分かるように、リツイートが多いというのは、ツイッター全体の特徴ではない。実際リツイートの多い都議選関連ツイートからリツイートではないことをカギに樺坂関連の大量のツイートを取り出すことができた。リツイートが多かったのが選挙という行動に関係したものなのかどうか、リツイート率の違いの要因分析は今後の課題である。また、リツイートされることの多いつぶやき手を表7でリストアップしたが、自身のツイート数が少なくても、リツイートが多ければ社会に与える影響は大きい。

今回は実際に流れたツイートに焦点を当てたので、つぶやき手に属する性質であるフォロワー数は解析していない。つぶやき手のフォロワー数を重み

にして統計を取ると、単純なツイート数の集計とは違った様相がみられるかもしれない。フォロワー数と被リツイート数の関係や影響力の違いなどは興味深い問題である。

本論文で対象としたのはあくまでツイートであり、実際の投票行動ではない。ツイートにはネガティブなものやポジティブなもの、そしてただのアナウンスのニュートラルなものがある。選挙の結果と関係づけるには、極性辞書を組み合わせて感情分析を行う必要がある。きのこvsたけのこ論争²⁶⁾のように簡単ではない、これも今後の課題としたい。

謝辞

コンピュータシステムの設置にあたり、松本大学情報センターの支援に感謝したい。

文献

- 1) 「WEB2.0への道」インターネットマガジン別冊, インプレスR&D (2006).
- 2) 総務省インターネット選挙運の解禁に関する情報, http://www.soumu.go.jp/senkyo/senkyo_s/naruhodo/naruhodo10.html (2017年9月4日閲覧).
- 3) dwangoプレスリリース, 参院選2016におけるniconicoの取り組み, <http://dwango.co.jp/pins/2016/0620/> (2017年4月5日閲覧).
- 4) 「各政党の「つぶやき分析体制」が判明」, ITpro, 『日経コンピュータ』Report2013年7月16日号 p14, <http://itpro.nikkeibp.co.jp/article/COLUMN/20130705/489455/> (2017年9月4日閲覧).
- 5) 「民主党が参院選に向け「LINE」などでネット選挙活動、ネットの「民意」解析ツールも運用」, ITpro 『日経コンピュータ』ニュース2013年6月28日, <http://itpro.nikkeibp.co.jp/article/NEWS/20130628/488529/> (2017年9月4日閲覧).
- 6) 「自民党がネット選挙の専任チームを発足、口コミ動向は候補者に毎日レポート」, ITpro, 『日経コンピュータ』Report2013年6月19日号, <http://itpro.nikkeibp.co.jp/article/NEWS/20130619/486383/> (2017年9月4日閲覧).
- 7) 【全475議席中、452議席の当確予測が的中】ニコニコアンケートもとに衆院選の「当確予測」を発表(12月15日12時更新), <http://news.nicovideo.jp/watch/nw1364374> (2017年9月4日閲覧).
- 8) 【参院選2016】ニコニコ独自当確予測の的中率発表, <http://blog.nicovideo.jp/niconews/ni062070.html> (2017年9月4日閲覧).
- 9) 田中辰雄, 山口真一, 「ネット炎上の研究」勁草書房 (2016年4月25日).
- 10) W.クアトロチョッキ, 「陰謀論を増幅 ネットの共鳴箱効果」, 日経サイエンス2017年7月号 pp54-58.
- 11) 「ネットは「偏った意見」を強化するか:研究結果」『WIRED』NEWS 2010.05.13 THU 21:00, <https://wired.jp/2010/05/13/%E3%83%8D%E3%83%83%E3%83%88%E3%81%AF%E3%80%8C%E5%81%8F%E3%81%A3%E3%81%9F%E6%84%8F%E8%A6%8B%E3%80%8D%E3%82%92%E5%BC%B7%E5%8C%96%E3%81%99%E3%82%8B%E3%81%8B%EF%BC%9A%E7%A0%94%E7%A9%B6%E7%B5%90%E6%9E%9C/> (2016年4月25日閲覧).
- 12) ネット選挙 ツイッター分析—毎日新聞・立命館大共同研究, <http://senkyo.mainichi.jp/2013san/analyze/20130731.html> (2017年9月4日閲覧).
- 13) 2017年 東京都議会議員選挙, <http://tokyogo2senkyo.com/> (2017年9月4日閲覧).
- 14) 毎日新聞7月2日, 第50869号長野県版朝刊, p1.
- 15) 毎日新聞7月3日, 第50870号長野県版朝刊, p1.
- 16) 信濃毎日新聞7月3日朝刊, p1.
- 17) 毎日新聞7月2日, 第50869号長野県版朝刊, 社会面続12版, p23.
- 18) 鳥海不二夫, 「Twitter上のビッグデータ収集と分析」『組織科学』Vol.48, No.4, pp47-59 (2015).
- 19) 加藤耕太, 「Pythonクロウリング&スクレイピング—データ収集・解析のための実践開発ガイド」技術評論社 (2016年12月16日).
- 20) Streaming APIで大量のつぶやきをリアルタイムに保存する方法(Python編), <http://blog.unfindable.net/archives/4257> (2017年9月4日閲覧).
- 21) MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/> (2017年9月4日閲覧).
- 22) neologd/mecab-ipadic-neologd, <https://github.com/neologd/mecab-ipadic-neologd/wiki/Home.ja> (2017年9月4日閲覧).
mecab-ipadic-NEologd: Neologism dictionary for MeCab, <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md> (2017年9月4日閲覧).
- 23) PythonでのMeCabを速くするtips, <http://qiita.com/yukinoi/items/81a707c1317c97f5fdf9> (2017年9月4日閲覧).
- 24) せっかくだから俺は衆院選挙のツイートを調べず, http://qiita.com/mima_ita/items/a30d8809c428f73fd360 (2017年9月4日閲覧).
- 25) スタバのTwitterデータをpythonで大量に取得し、データ分析を試みる, <http://qiita.com/kenmatsu4/items/02034e5688cc186f224b> (2017年9月4日閲覧).
- 26) 「きのこの山VSたけのこの里」論争ついに終結か ツイッター分析ツールでわかるネットの本音, Jcastニュース、ビジネス&メディアウォッチ (2014/3/14 10:46), <https://www.j-cast.com/2014/03/14198430.html?p=all> (2017年9月4日閲覧).