

ビックデータ VS 世論調査

～Twitter を多角的に検証する～

中川 純一（株式会社もしもしホットライン）



1. はじめに

最近、ビックデータの活用が話題となっている。調査分野においても「新興勢力 VS 既成勢力」という構図のように、「ビックデータ VS 世論調査」と位置付けられている。世論調査と対比される「ビックデータ」とは、ソーシャルメディア（以下、SNS）のことをさしている。主なメディアは「Twitter」や「ブログ」などであり、これらの特徴は、PCやスマートフォンなど、いつでもどこからでも無料で、個人が発信できることにある。

SNS が世論調査と対比される理由は、(1)「ビックデータは、データが大量であるため、多くの人の意見が収集できる」、(2)「ビックデータは、データが動的であるため、常に最新の意見が収集できる」、(3)「若者の利用率が高いため、若年層の意見が収集できる」など、広く意見収集するツールとして期待されており、世論調査を代替する有力候補という議論にもなっている。

大手マスコミにおいても、SNS を活用して世論の把握を試みている。例えば、本年の参院選期間には、朝日新聞社が SNS の分析をおこなっている。同社の取組みは、一般の SNS 利用者の書込みテキストの分析以外に、立候補者サイドの分析を実施し、ネット選挙の特性の把握を実施している。また、毎日新聞社では、電話世論調査（RDD）と Twitter の比較を実施しており、両者で取り上げられる話題の相違について検証している。

ここで、SNS データと世論調査データの特徴を整理してみた。まず、世論調査の特徴は、データの質が科学的に担保されていることにある。調べたい対象（以下、母集団）の全数から無作為に対象者を選定し、母集団を縮図にした集団（サンプル）をつくる。そして、その集団（サンプル）に調査をする。このように、世論調査は、再現可能なプロセスを経てデータを収集している。よって、必ずしも大量データである必要がない。一方、SNS データは、科学的なプロセスを経てデータが収集されていない。また、収集される書込みテキストは、SNS の利用者に偏っているなど、データの品質を評価する手段がない。このことから、SNS は世論調査の代替手段ではないことがわかる〔図表 1〕。

しかし、昨今の世論調査の実査環境の悪化により、世論調査の科学的なプロセスの維持が困難になり、新たな調査方法の模索が叫ばれている。本稿では、SNS と世論調査を対決軸でとらえるのではなく、SNS と世論調査との間での相互補完関係構築の可能性を模索した。

図表 1 SNS と世論調査の比較

	データ名	SNS	世論調査
1	データ量	大規模	小規模
2	データ更新	動的	静的
3	データ偏り	偏りあり	偏りなし
4	データ内容	形式のないフリーコメント	質問票への回答 クローズドアンサー

2. SNS を世論調査で活用する

SNS の世論調査での活用を考える前に、先行活用しているビジネスでの事例を参考としてゆきたい。ビジネスでの活用は「商品・サービスのプロモーション効果測定」、「商品・サービスの風評リスク検知」などが一般的である。

たとえば、ある企業のコンタクトセンターに入ってくる苦情が、二ヶ月も前に SNS で話題になっていたことや、あるスマートフォンメーカーが、どこよりも早く、初期不良の検知のために SNS を活用していることなど。これらの事例は、ネガティブな局面で SNS を活用した事例になる。

今回の検証では、SNS のネガティブな局面での活用事例、すなわち、リスク管理ツールとしての側面に焦点を絞って、SNS が世論調査を補完するツールとなり得るのか検証してみた。

3. 検証方法

1) 検証対象

本年実施された、共同通信社による「第 23 回参議院選挙に関する電話調査（全国トレンド調査）」（2013 年 6 月 22 日～同年 7 月 20 日、全 5 回）を参考対象とする。同調査では、投票日直前の 5 回目の調査で内閣支持率が急落した（4 回目：65.3%→5 回目：55.6%）、同不支持率（4 回目：24.7%→5 回目：32.4%）。この支持率下落（不支持率上昇）に着目をした。

次に、同電話（RDD）調査と同じ期間で、SNS データを収集する。「大量」且つ「動的」な SNS として、Twitter を素材とした。Twitter データの収集には、KDDI 研究所の協力で同社の開発した「ソーシャルメディア・ビジュアライザー（以下、SMV）」を使用した。

2) データ収集と処理方法

安倍内閣に関連するキーワード「安倍内閣」「安倍総理」「安倍首相」をインターネットよりクローリングをしてデータを収集した。これらには、RT（リツイート）が含まれてい

る。RT は、意見の拡散を目的として他者のコメントを引用している。この拡散の特徴を利用し、宣伝広告のような効果を期待して、組織的に RT を活用しているケースがあるため、今回は、RT を集計対象外とした。

結果、205,236 件の書込みテキストデータを収集した〔図表 2〕。集計に際しては、テキストマイニングツールに辞書設定 (126 件)、同義語設定 (54 件)、ポジティブ・ネガティブ単語設定 (331 件)、不要テキストの削除などテキスト集計の準備作業を行った。また、整理されたテキストデータを時系列比較するために、投票日を起点に直後 1 週～直前 5 週までデータを合計 6 週間で区切って週単位でデータセットした。

次に、205,236 件のテキスト内容を定量的に把握するために、「話題分類」により全ての書込みテキストの話題をカテゴリーに分けて集計を行った。「話題分類」で最も多いのは、「選挙の話題」が約 21%で、次いで「マスコミ報道の話題」が約 19%、3 位が「外交の話題」が 17%となり、上位 3 つの話題で 60%弱が占めている〔図表 3〕。

また、205,236 件のテキストでポジティブ・ネガティブ分析を実施したところ、ポジティブが約 10%、ネガティブが 11%、どちらでもない 79%となった〔図表 4〕。

図表 2 Twitter データセット

	データ名	対象期間	件数
1	直前5週	6月16日～6月22日	30,156
2	直前4週	6月23日～6月29日	30,402
3	直前3週	6月30日～7月6日	31,109
4	直前2週	7月7日～7月13日	30,437
5	直前1週	7月14日～7月20日	45,657
6	直後1週	7月21日～7月27日	37,475

図表 3 話題分類

(N=205,236)

	データ名	件数(MA)
1	選挙の話題	42,756
2	マスコミ報道の話題	39,310
3	外交の話題	33,864
4	原発の話題	27,477
5	経済の話題	19,905
6	憲法の話	13,418
7	SNSの話題	10,143
8	国会の話題	5,904
9	世論調査の話題	4,823
10	閣僚の話題	971
11	その他	39,225

図表 4 ポジティブ・ネガティブ分類一覧

(N=205,236)

	データ名	件数(MA)
1	ポジティブ	19,699
2	ネガティブ	23,188
3	どちらでもない	164,857

また、今回の Twitter に書き込みをした人の属性を調べてみた。KDDI 研究所の「SMV」による年代推計モデルを使って性年代比を集計してみたところ、205,236 件のテキストに対して性別では男性 89.1%、女性 6.0%、属性不明 4.9%、と圧倒的に男性の書き込みが多いことが分かる。年代推計では、40 代以上が 53.9%、30 代が 20.0%、20 代が 15.6%、10 代が 5.5%、年代不明は 5.0%と若い世代のデータが収集されている。

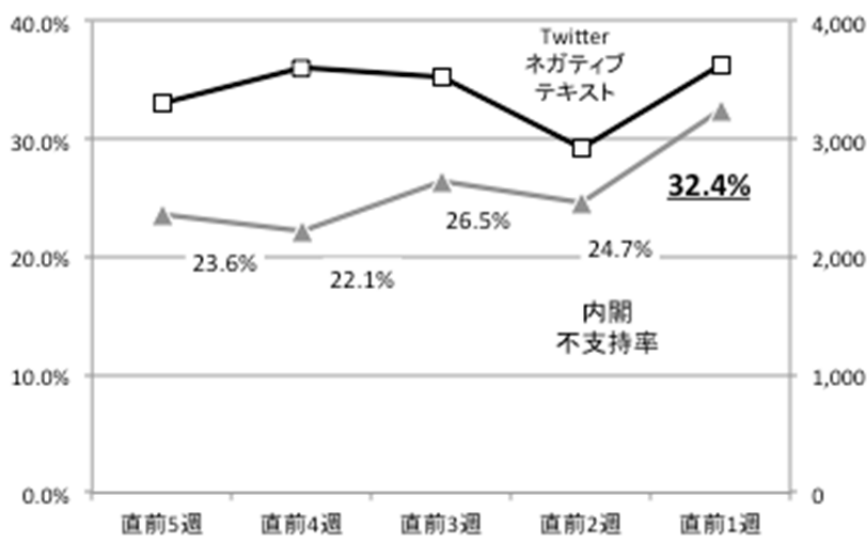
4. 検証結果

1) 内閣支持率と Twitter の定量比較

電話世論調査の安倍内閣の不支持率 (%) と Twitter のネガティブテキストの件数の時系列変化をみたところ、内閣支持率と Twitter の件数には明確な相関がみられなかった (相関係数 0.34)。ただ、直前 2 週間から直前 1 週間にかけての内閣不支持率とネガティブなテキスト件数が伴って増加している [図表 5]。その詳細を以下、詳しく述べてゆく。

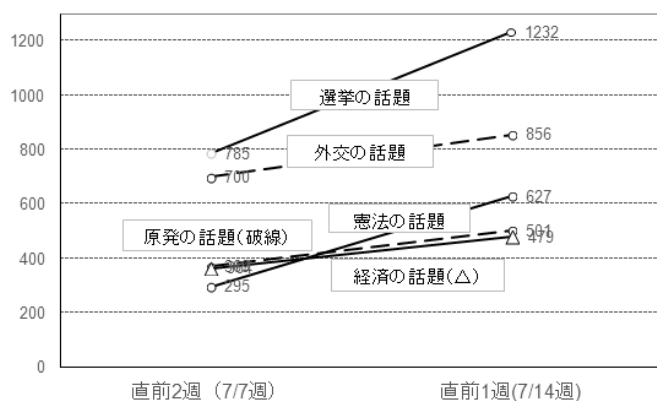
図表 5 Twitter と RDD の時系列変化

(Twitter は件数)



直前1週に Twitter でどのようなことが語られていたのかを直前2週との比較を行う。ネガティブテキストの、直前1週間と直前2週間の話題分類の比較をしてみると、「選挙」「外交」「憲法」の話題が増えている一方、「原発」「経済」の話題は横這いであったことがわかる〔図表6〕。以下、直前2週と直前1週での伸張率の高い「選挙の話題」と「憲法の話題」について詳しくみてゆく。

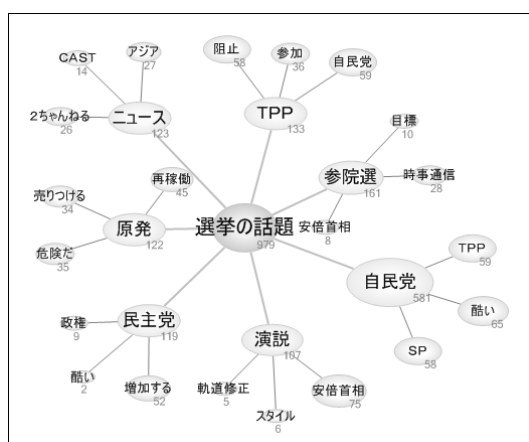
図表6 直前1週と直前2週の話題比較（件数）



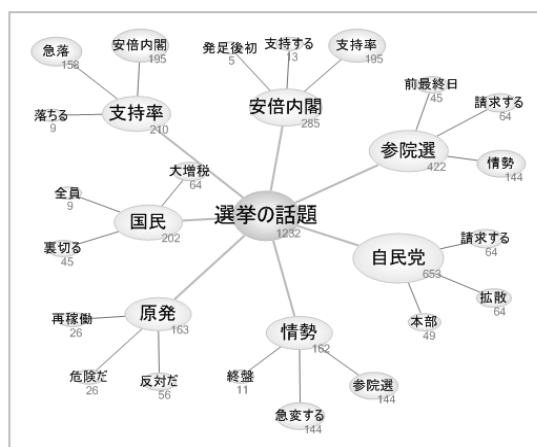
1) 直前1週の「選挙の話題」

直前1週での特徴は、マスコミが報道した参院選情勢調査の話題が語られている。同様に直前2週では、「政策」「政党」「選挙活動」についての話題が中心であり、直前1週となって話題の中心がマスコミの情勢調査になっている〔図表7〕。

図表7 「選挙の話題」マッピング図
(直前2週)



(直前1週)

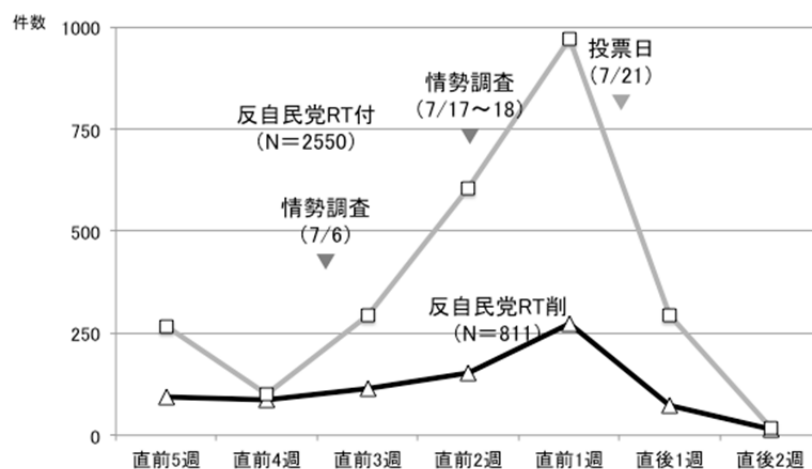


この参院選情勢調査の報道が、投票者の投票行動に影響したのかを検証してみた。同調査は参院選期間中2回ほど行われている。最初の調査は、2013年7月6日に報道された。内容は、「自公 過半数は確実」（読売新聞）、「自公、過半数は確実」（朝日新聞）、2回目の報道は、同年7月17日「衆参ねじれ解消確実」（読売新聞）や、同18日「与党 過半数は確実」（朝日新聞）などの代表例に挙げた通り、マスコミ各社において自民党圧勝の報道がされた。

この報道に対抗する「反自民党」の人びとの動きを、Twitter上で捕捉した。「反自民党」のテキストを集計するにあたり、キーワード（「反自民」）や、係り受け（「自民党-投票しない」「自民党-落とす」）を集計ツールに設定しデータを作成した（直前5週から直後2週まで7週間を集計）、また、拡散を目的としたRT（リツイート）の動向をとらえた〔図表8〕。

情勢調査が報道されている週と、報道されていない週とでは「反自民党」テキストの件数が大きく違うことが分かる。自民党圧勝の報道が「反自民党」の人びとに影響した可能性があると推察される。また、RTを含めたテキスト（3361件）に含まれる単語ランキング（動詞）上位3位は、「落とす」（1284件）、「負ける」（484件）、「投票しない」（431件）となっている。

図表8 参院情勢調査の影響（件数）

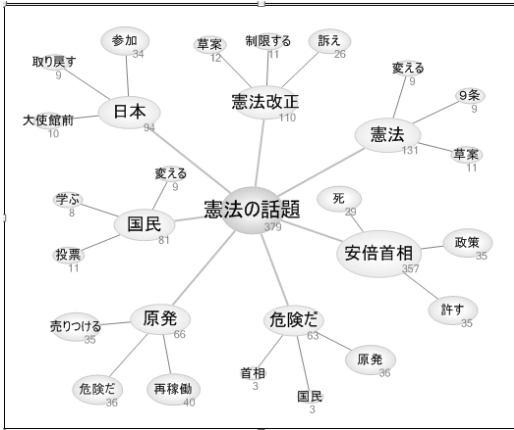


2) 直前1週の「憲法の話題」

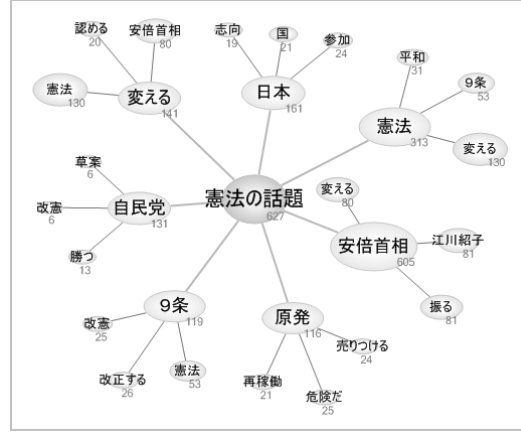
直前1週の「憲法の話題」は、直前2週に比較して「9条」というキーワードが目立つ。これは、安倍首相が7月15日の長崎国際テレビで9条改正を明言した放送に反応してのテキストが多く投稿されている。Twitter上での9条というキーワードへの反応は非常に敏感であることが伺える〔図表9〕。

図表9 「憲法の話題」マッピング図

(直前2週)

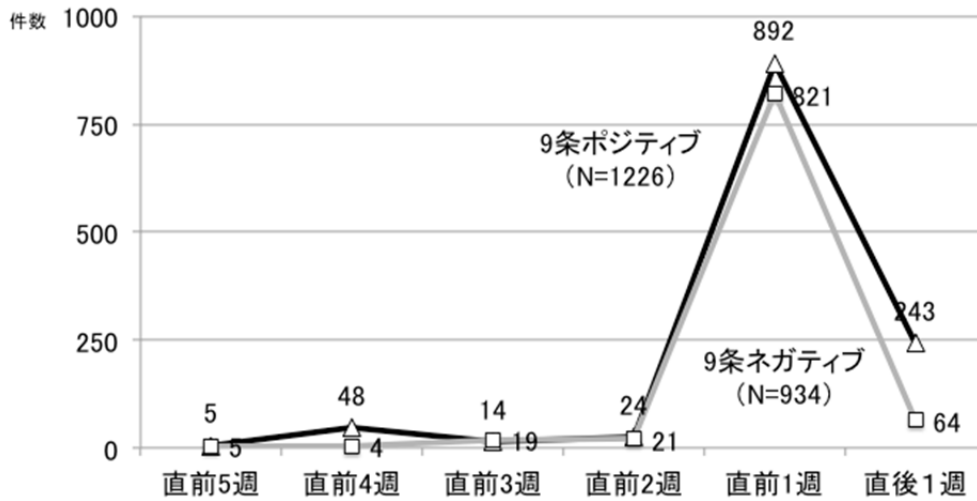


(直前1週)



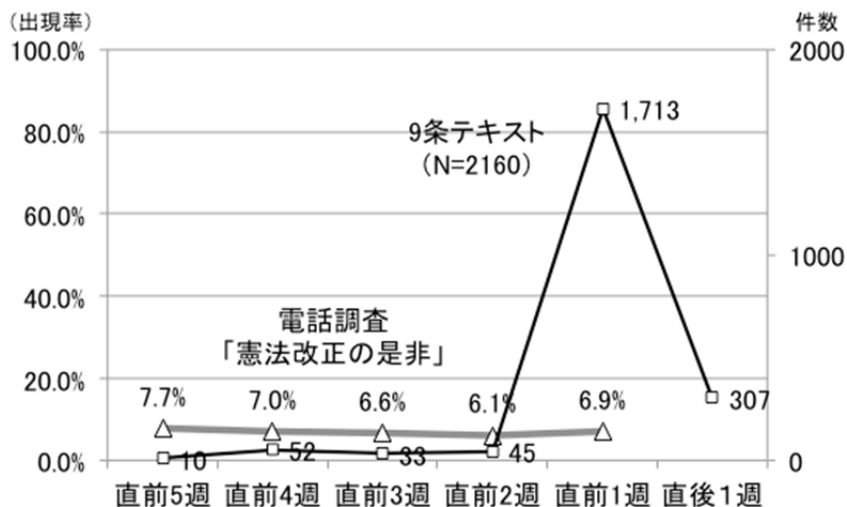
9条に関する「9条-ポジ・ネガ」の動向では、ポジティブワードとネガティブワードが拮抗しており、意見が二分しているといえる[図表10]。

図表10 「9条」を含むテキスト件数動向 (ポジ・ネガ)



一方、RDD 調査における投票先の判断基準としての「憲法に改正の是非」の推移と、Twitter での 9 条の話題の推移を重ねてみたところ、RDD 調査では安定した世論であるが、Twitter では 9 条に関しては不安定な動きをしている。参院選ではあえて憲法改正を争点としなかったために、Twitter でちょっとした刺激(長崎国際テレビ)があると敏感に反応しており、9 条に関して議論が渴望されていると推測できる[図表 11]。

図表 11 「9条」を含むテキスト件数動向と RDD 調査の動向



5. まとめ

本検証で、RDD 調査での投票日直前の内閣支持率下落の原因を、Twitter のテキストから探る試みを行ったところ、マスコミの参議院選挙情勢調査で自民圧勝が報道されると、「反自民」のテキストが急増していることが判明した。しかし、「反自民」テキストが量的に拡散された形跡は確認できなかった。この「反自民」テキストが量的に拡散しなかったからといって無視してよいものだろうか。Twitter は、世論の一部分を確実に反映しており、本検証の場合は情勢調査の報道がどう受け止められているかを観測し、それが拡散するかをモニタリングすることで世論の広がりや推計することに活用できるのではないだろうか。

また、動的に話題が変化してゆく Twitter に着目して参院選直前5週から動きをみてきた。その結果、件数が多く動きのある話題の発信源はマスコミ情報であった。更に、その話題を追跡してゆくと、1日ごとに話題がマスコミより提供され、Twitter 上で急激に盛り上がりを見せるが、翌日には何もなかったように消費されて忘れ去られるような動きであった。

一方で、Twitter には、マスコミ報道がなくとも恒常的な話題があった。たとえば、外交では「中韓関係」や「尖閣問題」、原発では「福島原発」や「再稼働問題」、経済では「TPP」や「アベノミクス」、憲法では「改憲」や「9条」などである。Twitter には、機敏に反応する部分（動的）と恒常的な部分（静的）の二重構造が存在している。

この二重構造を世論調査に活用する場面として、マスコミ各社が実施している定例電話世論調査（RDD）がある。定例電話世論調査（RDD）は、通常1ヶ月に1回、調査を実施しており、次の調査までの間が約1ヶ月間空いている。この月1回での定例電話世論調査（RDD）では、世論の把握が切断される状態になっている。この点と点を結ぶ世論の把握を、SNS の動的・静的な話題でつなげることで、世論調査を補完する関係を構築できる可能性がある。そのためには、実用的な SNS の知見蓄積の場が必要となってくる。