

## テキストマイニングによる Twitter 世論の定量評価 ～SNS 投稿から多数派の意見を汲み取ることは可能か？～

昨今、SNS 上のムーブメントが政策決定への影響を強めている。SNS での政治運動は民意をある程度反映している面もあるが、SNS 上の意見を政策にどの程度織り込むべきかについては深い議論がされていない。一方でテキストマイニングなどを用いた定量的な分析・研究の進展によって、SNS の投稿から民意を抽出し、政策運営に活用する道筋が見えつつある。

### 1. 強まる SNS による政治への影響力

SNS（ソーシャル・ネットワーキング・サービス）上の発信が大きな世論のうねりとなって、政治に影響を与える動きが近年散見されるようになった。

日本では、2020 年の国会に検察官の定年延長を可能にする「検察庁法改正案」が提出されたが、成立が見送りとなった。その最大の要因として、Twitter で著名人から始まった反対運動が爆発的に盛り上がった動きが指摘されている<sup>1</sup>。Twitter 上で「検察庁法改正案に抗議します」というハッシュタグを含む数百万の投稿が観測され、いわゆる「Twitter 世論」が政治を動かした実例となった。

SNS が政治に大きな影響を与えた例として最も有名なものは、2010 年代初頭にアラブ諸国で広まった民主化運動「アラブの春」であろう。チュニジアに端を発した民主化運動は各国に波及し、同国のみならずエジプト、リビアなどでも長期独裁政権を崩壊させるに至った<sup>2</sup>。一連の民主化運動において、SNS は一般市民の有力な情報源となり、抗議活動を呼びかける媒介としての機能を果たした。また、世界に対して緊迫した情勢が SNS を通じてリアルタイムに発信され、国際世論を動かしたと評されている<sup>3</sup>。

「アラブの春」から 10 年近く経ち、SNS は各国で政治への影響力を増大させており、市民が政治的意見を表明する場となっている。

### 2. テキストマイニングによる SNS から民意を抽出する試み

SNS の投稿を政治にどの程度反映するべきかという問題は議論の途上にあり、定量的な分析が試みられている。本章ではその概要を紹介する。

#### (1) SNS 世論の有用性と歪み

SNS によるインターネット世論を重視すべきだとする見解は、「集合知」という概念に立脚している。集合知とは、多様な集団が到達した結論は 1 人の優秀な専門家の意見に勝るという説<sup>4</sup>である。集合知が有用であるのは、個人の意見を足し合わせることで各々の意見に含まれる誤差が相殺されて真の値に近づくという確率論的な背景がある。したがって、集団の多様性、独立性などが前提となる。

一方で、SNS 上の世論にはバイアス（一方向への意見の偏り）が多く含まれていることも明らかになっている。ユーザーは関心のある話題や、同じ考えを持つ者をフォローする傾向があり、集合知の前提である集団の多様性や独立性は SNS 上で必ずしも担保されていない。類似の政治志向を持つ SNS ユーザーが集まるコミュニティでは、同調圧力によって集団の意に即した思想や言動が実勢以上に増幅される傾向<sup>5</sup>にあり、投稿の多寡のみで実社会の民意を推し量ることは難しい。

## (2) テキストマイニングの隆盛と Twitter への適用

SNS の投稿が持つバイアスを解消して、SNS から世論を定量的に評価しようとする試みが近年活発に行われている。

テキストを文法などに即して定性的に解釈する試みはかなり昔から行われていたが、コンピュータによって大量のテキストの定量的な分析を行う「テキストマイニング」は、文書の電子化が飛躍的に進展した 20 世紀末から急速に普及した<sup>6</sup>。テキストマイニングで得られる典型的な知見は、ともに使われる頻度が高い単語の組み合わせの発見、テキストに基づき経済や政治など文章が対象としているカテゴリーへの分類、ある表現が肯定的か否定的かの判定などが挙げられる（図表 1）。実務面では、既往の膨大な医学論文を対象に症例と物質の組み合わせ発見する分析<sup>7</sup>や、コールセンターでの顧客とのやり取りを分析するといった研究<sup>8</sup>が進められてきた。

〈図表 1〉 テキストマイニング分析で得られる主な知見

	概要	活用事例
共起表現の発見	ある単語と同じ文や文書の中で使われる単語や表現を発見する	翻訳、大量の論文の中で登場する単語の組み合わせ（例：症例と物質）の発見、コールセンターのやり取りから商品ごとの問い合わせ傾向の分析
カテゴリー分類	文章を「政治」や「経済」などの予め定めたカテゴリーに分類する	ユーザーへのレコメンド広告、スパム投稿の検知
ポジティブ/ネガティブ分析	文章の内容をポジティブかネガティブか判定する	経済のセンチメント分析、口コミの定量評価

（出典）SOMPO 未来研究所作成

21 世紀に入り、スマートフォンの普及とともに SNS の利用拡大が進み、Twitter をはじめとする SNS 上の膨大なテキストデータをテキストマイニングの対象に拡大する機運が高まった。

グローバルに利用されている SNS として、Facebook、Twitter、Instagram などが挙げられるが、研究対象として最も注目されているのが Twitter である。Twitter はテキストを中心とした投稿が一般に公開されているのに対し、Facebook はクローズドな空間での投稿が多く、Instagram は画像データが中心に投稿されており、分析の難易度が高いことが要因である<sup>9</sup>。したがって、以下では Twitter を対象としたテキストデータ分析を中心に紹介する。

Twitter ではユーザーが 140 文字以内で気軽に投稿を行うため、Twitter 上のユーザーは社会情勢を反映するセンサーであり、投稿はセンサーからの出力であるとする「ソーシャルセンサー」という仮説も提唱された<sup>10</sup>。Twitter のリアルタイム性を帯びた一般ユーザーの投稿からなるビッグデータは、企業活動において金銭的価値を持ち得るほか、複雑な社会現象を理解する知見をもたらすのではとの期待も膨らんだ<sup>11</sup>。実際に 2010 年に発表された論文では、約 1,000 万の Twitter への投稿から感情分析を用いて、ダウ平均株価の上昇／下落を 86.7%の精度で予測することに成功し<sup>12</sup>、後続の研究に大きなインパクトを与えた。

一方で、SNS を対象としたテキストマイニングでは、①誤字脱字や文法誤り、ボット（自動で情報を投稿するプログラム）やスパム（なりすましや同一内容を投稿し続ける等の迷惑行為）を含むノイズ、②データの入手可能性と解析結果の再現性、③データのバイアス、などの課題も生じている（図表 2）。

〈図表 2〉 SNS 分析の主な問題点

ノイズが多い	公文書やビジネス文書に比べて誤字脱字、文法誤りが多く、スラングも使用される。ボットやスパムによる投稿も混在する。
データの入手可能性と解析結果の再現性	SNSデータのアクセス頻度には運営事業者によって制限が課されており、別の人が同じ投稿を収集できるとは限らない。また、SNSデータを研究者同士で共有することにも制限が存在する。
SNSデータが含むバイアス	SNS利用者は社会全体を母集団とした際に偏ったサンプルであることを認識し、バイアスを含むことを前提に分析する必要がある。

（出典）SOMPO 未来研究所作成

### （3）政治分野での研究事例

Twitter 分析を政治分野に活用しようという研究も活発に行われている。

代表的なものが、有権者に焦点を当て、選挙結果の予測を行う研究である。こうした研究は多数存在し、ドイツ議会選挙、米国大統領選挙・議会選挙、日本の参議院選挙などを舞台に分析が行われているが、予測可能性について疑問が呈されている研究も多く、今後の進展が期待されている<sup>13</sup>。

また、Twitter ユーザーの政治的投稿に着目して、帰属する集団を分類する「ネットワーク分析」も有望視されている。2017 年に行われた衆議院議員総選挙を対象にした分析では、選挙前 2 週間に政党名を含む投稿から約 23,000 人のユーザーを複数のコミュニティに分類した<sup>14</sup>。各コミュニティに所属するユーザー同士は同じ政党や情報源をフォローし、投稿する政治的なテーマも類似していることが確認された。同様の調査が 2020 年の東京都知事選でも行われており、「保守・右派」や、「リベラル・左派」などの各コミュニティ内で関心に沿った投稿が拡散されている<sup>15</sup>。ネットワーク分析を活用すると、SNS の投稿が市民全体の民意なのか、はたまた一部のコミュニティで盛り上がっている言論であるかが判断可能である。

冒頭に述べた検察庁法改正案についての興味深い分析・考察もすでに行われている。東京大学の鳥海不二夫准教授は「検察庁法改正案に抗議します」というタグを含む 473 万件の投稿について、ユーザーごとの投稿回数、拡散に寄与したユーザーの過去の投稿数などの観点から分析を行い、スパムやボットによる投稿は多くないと結論付けている<sup>16</sup>。さらに、上記の投稿についてネットワーク分析を試みたところ、保守やリベラルなどで二分されずにユーザーが満遍なく法改正案に反対していることを示した<sup>17</sup>。この結果は、検察庁法改正案については Twitter の投稿が一部のコミュニティに偏ることなく民意を反映したものであった可能性を示唆している。

### （4）SNS 分析と世論調査

民意を測る伝統的な手法は世論調査であり、SNS 分析と目的は共通しているが方法は大きく異なる。世論調査では、抽出された市民に対し特定のトピックについて質問を投げかける積極的な関わり方をするのに対し、SNS 分析はインターネット上の投稿を収集し観察する点で受動的である。したがって、世論調査では対象者を無作為に抽出できるが、SNS 分析においてはユーザーの属性情報について信頼性が低く、偏りがある可能性を排除できない。加えて、SNS の投稿は運営会社による検閲や恣意的な削除などの可能性を否定できない点、敵対国などによる情報操作の懸念が指摘されている<sup>18</sup>。

一方で、SNS 分析は膨大なデータをタイムリーに取得できるという利点もある。マスコミ各社が毎月行う定例電話世論調査では期間が 1 ヶ月空いてしまうが、SNS では常に最新のデータの観察が可能である<sup>19</sup>。

さらに、世論調査では調査員の目を意識して、社会規範に即した回答をする可能性があるのに対し、匿名性を持つ Twitter などの SNS はより真意を発信しやすい側面もある。

こうしたメリットとデメリットを勘案して、研究者の間では真の民意を把握するためには世論調査か SNS 分析を択一的に使用するのではなく、両者を補完的に用いるべきだとする向きが多い<sup>20,21</sup>。

### 3. 今後の展望

SNS をテキストマイニングで分析して民意を把握するという試みは途上にあるが、多くの研究で有用性が検証されている。現在、SNS の動向を政策決定に活かす枠組みは用意されていないが、検察庁法改正案の抗議運動のように SNS 上の世論が政治を動かした事例は世界でも散見される。今後は分析手法の発展によって SNS 世論の定量評価が可能になれば、政策運営に当たり SNS 世論を活かす道も可能になるかもしれない。市民が選挙以外に政治参加を行う機会を得ることで、かねてより指摘されている政治的無関心の払拭や、熟議を経ない強行採決などに代表されるクローズド（閉鎖的）な政治過程の変容から、一層民主的な社会の形成も展望できる。SNS の投稿を政策運営に活用するにあたっては慎重な検討が必要になるであろうが、今後の研究の深化に注目したい。

【副主任研究員 菊武省造】

- 1 谷卓生「"ツイッターデモ"の歴史的高まりで、「検察庁法改正案」国会での成立見送り」（放送研究と調査 2020年7月号）
- 2 外務省「わかる！国際情勢 Vol.87 「アラブの春」と中東・北アフリカ情勢」（visited Sept. 3, 2020）  
<<https://www.mofa.go.jp/mofaj/press/pr/wakaru/topics/vol87/>>
- 3 遠藤薫「ソーシャルメディアは政治・選挙を変えるか」（Voters、No.12、2013年）
- 4 ジェームズ・スロウィッキー「みんなの意見は案外正しい」（角川文庫、2009年）
- 5 キャス・サンスティーン「#リパブリック：インターネットは民主主義になにをもたらすのか」（勁草書房、2018年）
- 6 小木しのぶ「テキストマイニングの技術と動向」（計算機統計学、第28巻・第1号、2015年）
- 7 Hearst, M. “*Untangling Text Data Mining*”, in the Proceedings of ACL'99: the 37th Annual Meeting of the Association for Computational Linguistics, June 1999
- 8 那須川哲哉「コールセンターにおけるテキストマイニング」（人工知能学会誌、Vol.16 No.2、2001年）
- 9 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、2015年）
- 10 T. Sakaki, M. Okazaki, Y. Matsuo “*Earthquake Shake Twitter Users : Real-time Event Detection by Social Sensors*”, Proc. 18th Int. World Wide Web Conference, 2010
- 11 津川翔「SNSに蓄積された情報の蓄積」（電子情報通信学会通信ソサイエティマガジン、2020年13巻4号、2020年）
- 12 J. Bollen, H. Mao, X. Zeng, “*Twitter mood predicts the stock market*”, Journal of Computational Science, Vol.2, Issue 1, Mar. 2011
- 13 那須野薫、松尾豊「Twitterにおける候補者の情報拡散に着目した国政選挙当選者予測」（2014年度人工知能学会全国大会、2014年）
- 14 ロバート・ファーヒ「SNSデータから有権者の本音を推論できるか」（日本世論調査協会報「よろん」123巻、2019年）
- 15 日本経済新聞「ネット民主主義、誰が動かす「専門家村」に潜む悪意」（2020年8月3日）
- 16 鳥海不二夫「#検察庁法改正案に抗議した人は本当はどのくらいいたのか」（visited Sept. 4, 2020）  
<<https://note.com/torix/n/n5074423f17cd>>
- 17 同上
- 18 前掲注14
- 19 中川純一「ビッグデータ vs 世論調査：ツイッターを多角的に検証する」（埼玉大学社会調査研究センター第3回世論・選挙調査大会プログラム、2013年）
- 20 前掲注14
- 21 前掲注19