

SNSの中で“つくられる真実” と“対立する正しさ”

東京工業大学環境・社会理工学院 准教授

笹原和俊（さはら かずとし）



Profile—

2005年、東京大学大学院総合文化研究科修了。博士（学術）。名古屋大学大学院情報学研究科講師等を経て、2020年より現職。専門は計算社会科学。主著に『フェイクニュースを科学する：拡散するデマ、陰謀論、プロパガンダのしくみ』（単著、化学同人）がある。

偽物が本物に、本物が偽物になる世界

今ツイッターやフェイスブックなどのSNS（交流サイト）で目にしているニュースは真実なのか？ 単なる誤報か、誰かの嘘か、あるいは、意図的に操作された情報なのか？ ミシガン大学のアビーブ・オバディアは、虚偽が溢れ、あらゆる情報が信じられなくなる未来をインフォカリプス（情報の終焉）と表現した¹。その避けたい未来はいくつかの段階を経て、杞憂とは言い切れない状況になってきている。

2016年の米国大統領選では、根拠不明の偽情報がSNSを中心として拡散し、社会を混乱させた。偽情報の中には、「ローマ法王がトランプ氏を支持」のような、広告収入が目当てのクリックベイト（ユーザの興味を引くような記事で閲覧数を増やす手法）もあれば、選挙ビジネスや国家による情報操作を目的とした偽情報もあった。トランプ大統領は選挙後も不正確な情報をSNSで発信し、不都合な事実を報道するメディアを「フェイクニュース！」と糾弾した。フェイク（偽）ニュースやポスト真実という言葉が生まれ、正しさを求める人々の心が揺らぎ出したのはこの頃である。2020年に選挙で敗れた後も、トランプ大統領を英雄に祭り上げるQアノン陰謀論は消滅せず、「メディアが伝えない真実」なるものを追いつけている²。

2020年は新型コロナウイルス感染症（COVID-19）のパンデミックの発生にともない偽情報が氾濫した。日本では「お湯で新型コロナが死滅する」という偽の予防法が出回った。世界では「ビル・ゲイツが新型コロナのワクチ

ンを人々に接種させ、監視用のマイクロチップを埋め込もうとしている」という陰謀論が登場し、ワクチン忌避や人々の不安を煽った。こうした不確かな情報が人から人へと伝染する状態を世界保健機関（WHO）はインフォデミックと呼び、警鐘を鳴らしている。

そして、COVID-19のパンデミックもインフォデミックも収束していない2022年2月24日、ロシアによるウクライナへの軍事侵攻が始まった。SNSは現地の様子を伝え、遠隔の人々をつなぐ役割を果たしているが、偽ニュースやプロパガンダを増幅する役割もしている。同年3月16日には、ゼレンスキー大統領がウクライナ国民に降伏を呼びかける偽動画がSNSに投稿された。これはディープフェイクと呼ばれる技術で合成された動画だが、不自然に体の動きが少なく、声が本人よりも低いなど、すぐに偽物と見破られる低品質のものだった。しかし、これがもっと精度が高いディープフェイクだったら、と思うと背筋が凍る。

意見の凝集の反発を促進する

エコーチェンバー

SNSにおいて自分と似た価値観や興味関心をもつ人とばかりつながり、同じような情報ばかりが流通する閉鎖的な情報環境ができてしまう。このような同じ「声」がこだまする環境をエコーチェンバーという。エコーチェンバーは様々なSNSでその存在が確認されており³、偽ニュースの温床となったり、ヘイトを増幅させたりし、意見の対立や社会的分断を加速させる

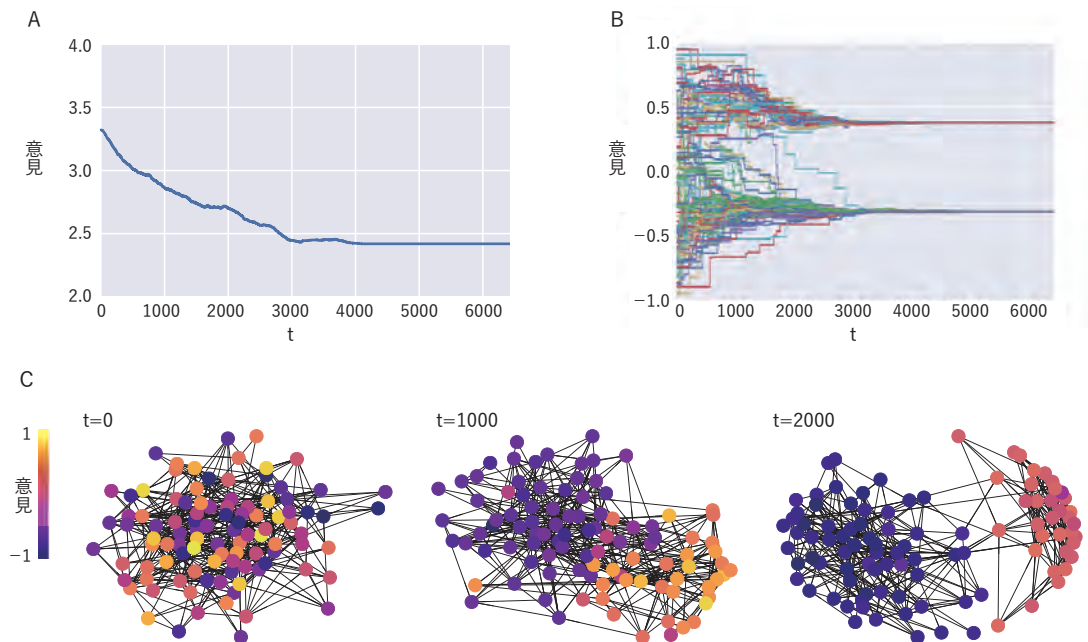


図1 エコーチェンバーの計算モデルにおける意見の分極と社会的ネットワークの分断

可能性がある。

私の研究グループは、エコーチェンバーの発生仕組みを知るために、SNSでつながった人々の意見形成をモデル化し、その挙動を分析した⁴。このモデルに登場するユーザは各自の意見を持ち、意見は-1から1までの実数で表現される（-1はリベラル、0は中道、1は保守）。そして、各ユーザは意見の許容範囲を持ち、その範囲にある他者の意見は「自分と似ている」と判断する。さらに、各ユーザは別のユーザをフォローすることができ、フォローしたユーザ（友人）の投稿（意見）のみ見ることができる。毎回、各ユーザは自分と似た意見の友人たちから影響を受けて、自分の意見を少し変更し、自分と似ていない意見を投稿した友人をアンフォローして、別のユーザを新しくフォローする。その後、自分の意見を投稿するか、自分に似た友人の意見を選んで共有する（ツイッターのリツイートやフェイスブックのシェアに相当）。このような、SNSで日常的に行われている過程を繰り返すことで、ネット環境を模した意見形成が進行する（簡易版のデモが公開されている⁵）。

このモデルを使ってシミュレーションした結

果が図1である。初期状態において多様な意見やつながりがあったとしても、時間が経つにつれて、(A) ユーザが自分の環境で目にする意見の多様性が減少し（同じような情報ばかりに接触するようになり）、(B) ユーザ全体の意見の二極化が生じた。そして、(C) ユーザどうしのフォロー関係からなる社会的ネットワークは、異なる意見をもった2つのグループに分断されてしまった。この結果は、SNSを使っているだけでエコーチェンバーが自然に発生してしまうことを示唆している。特に、社会的影響が強く、アンフォローの頻度が高い条件でエコーチェンバー化が加速されることがわかった。つまり、「似た他者から社会的影響を受ける」、「似ていない他者を社会的に切断する」というわれわれの行動傾向や、これらの行動を容易にするSNSの存在がかけ合わさることで、エコーチェンバー化が加速している可能性がある。

アルゴリズムが生み出すフィルターバブル

情報環境で生じるもうひとつの問題がフィルターバブルである。フィルターバブルとは、ユーザの個人情報や学習したアルゴリズムに

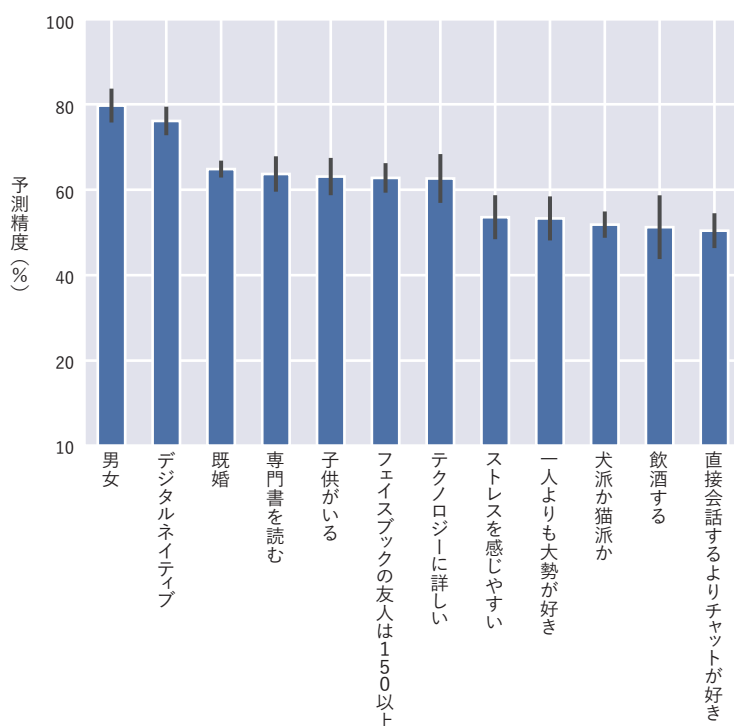


図2 ツイートからの個人情報の推定

よって、その人にとって興味関心がありそうな情報ばかりがやってくる環境のことである⁶。検索、オンライン広告、SNSのニュースフィードなど、多くのオンラインサービスにこのような情報のフィルターが埋め込まれている。このようなフィルターは、自分と異なる意見や異なる価値観の情報を隠してしまうだけでなく、ユーザごとにパーソナライズされてしまうので、みんなが共通の事実に接しづらい分断された状況（バブル）が生じてしまう。

私の研究グループでは、ツイッターに投稿された日本語のデータを対象として、投稿に含まれるテキストのみから個人情報をどのくらい推測可能なのかを研究した。調査会社を通じて約600人の調査対象者を集め、性別や年齢層などの基本属性だけでなく、「ストレスを感じやすいか」「犬派か猫派か」などの質問に回答してもらい、許可を得たうえで参加者のツイートを収集した。そして、収集したツイートを一定数まとめたものを自然言語処理の方法でベクトル化し、それを入力、参加者の回答を正解として、

機械学習のモデルをトレーニングした。そして、12種類の個人情報のそれぞれを二択で予測できるかを調査した⁷。

その結果が図2である。性別や年齢層（デジタルネイティブかどうか）などの7種類の個人情報には60%以上の確率で予測できることがわかった。「フェイスブックの友人は150人以上いるか」という交友関係に関する情報も、ツイッターのある程度の書き込みから予測できるというのは興味深い。友人の多寡に関する情報が自発的な投稿内容に表れているのである。また、性別や年齢層などの個人情報は高い精度で予測ができ、「専門書を読む」などの特徴もある程度予測することができた。何気なくしているツイートには個人情報がコード化され、フィルターバブルが容易に作られる状況にあることがわかる。今後ますます機械学習の精度が上がる事が予想されるため、推定された個人情報が情報操作などに悪用される危険性が高まる。

インフォカリプスを回避するためのヒント

生成の仕組みは異なるが、エコーチェンバーもフィルターバブルも、自分が見たいと思う情報だけが流通する閉じた情報環境であるという点は共通している。前者は「あなたが誰とつながるか」ということによって、後者は「フィルターがあなたを学習する」ことによって、そのような閉じた情報環境ができあがる。この環境にフェイクが紛れ込んだならば、あるいは、真実なのにフェイクだと伝えられたならば、瞬く間に共有の連鎖が生じてそれが増幅されてしまう。そして、フェイクを訂正したり、非難したりする声は届かないか、打ち消されてしまう。一方にとっての真実は、他方にとってのフェイク。そのような状況が生じてしまう。

これらに通底するのは指数関数的に増え続ける情報量の問題である。2000年代以降、インターネットが発展し、SNSが普及して、PCやスマホのユーザー一人ひとりが情報の発信者になり、生成される情報量は飛躍的に増えた。しかし、増え続ける情報量に対して人間のアテンション（注意力）という認知的資源は相変わらず有限なままである。このような情報過多の状況では、注意力不足から誤った情報や不正確な情報を適切に処理できないという状態に陥る。興味関心を同じにする友人知人の影響を重要視するのも、アルゴリズムによる情報のフィルターが必要なのも、そのためである。加えて、剥き出しの感情や憎悪に満ちた中傷など本能を刺激する偽ニュースや偽動画は、理性を飛び越えて人々の心の中に侵入し、集団極性化や社会的分断を引き起こす一因となる。現にロシアのウクライナ侵略では、SNSが武器として使われ、情報大戦の様相を呈している。国家によるプロパガンダもスマホの自撮りも渾然一体となって、タイムラインに溢れている。

また、情報技術の発展は事実とフィクションの境界をどんどん曖昧にしている。GPT-3などの最新の自然言語処理を搭載したAIは、人が書いたのと変わらないレベルの文章を生成することができる⁸。また、高度なディープフェイクを用いれば、真偽の見分けが困難な偽動画

を合成することも可能になる。フェイクをつくるテクノロジーが高度化している。早晩これらの情報技術はコモディティ化し、誰でも高度な偽記事や偽動画を簡単に作れる時代がやってくる。そうなった時に、経済的動機や政治的意図でこうした情報技術が悪用され、インフォカリプス的な未来に近づく可能性がある。偽情報の高度化に対しては対抗技術だけでなく、レジリエントな社会の仕組みを探究する必要がある。そして何よりもSNSを使うわれわれ自身が、これらの環境問題を理解し、賢いユーザとなる必要がある。

文 献

- Schick, N. (2020) *Deep fakes and the infocalypse*. London: Hachette.
- Xu, W., & Sasahara, K. (accepted) A network-based approach to QAnon user dynamics and topic diversity during the COVID-19 infodemic. *APSIPA Transactions on Signal and Information Processing*.
- Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrocioni, W., & Starnini, M. (2021) The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118, e2023301118.
- Sasahara, K., Chen, W., Peng, H., Ciampaglia, G. L., Flammini, A., & Menczer, F. (2021) Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science*, 4, 381–402.
- Hao, P., Menczer, F., & Sasahara, K. (2017) *EchoDemo*. <https://osome.iu.edu/demos/echo/>
- Pariser, E. (2012) *The filter bubble: What the internet is hiding from you*. London: Penguin Books. (イーライ・パリサー、井口耕二訳 (2016) 『フィルターバブル：インターネットが隠していること』 早川書房)
- 笹原和俊 (2021) 『フェイクニュースを科学する：拡散するデマ、陰謀論、プロパガンダのしくみ』 化学同人
- Brown, T. B. et al. (2020) *Language models are few-shot learners*. arXiv:2005.14165 [cs.CL].

*COI：本記事に関連して開示すべき利益相反はない。