

# Twitterにおける想起現象と忘却 —俳優の結婚ニュースに着目して—

五十嵐 直輝<sup>\*1</sup>

岡田 幸彦<sup>\*2</sup>

佐野 幸恵<sup>\*3</sup>

**Abstract** — 本研究では、ニュースなど外的要因によって記憶が思い出される想起現象とその忘却について着目し、日本の俳優（戸田恵梨香）の結婚ニュースで起きた想起現象と忘却について、Twitter における 1 時間単位の時系列データを用いて数理モデル化検証を行った。結果として、べき指数が 1 程度のべき関数でモデル化されることが明らかになった。

**Keywords** : Twitter, 集合的記憶, 想起現象, 社会経済物理

## 1 はじめに

### 1.1 ソーシャルメディアの普及

近年のビッグデータ時代の到来により、SNS は人々の生活に密着したツールとして確立されている。例えばソーシャルメディアの 1 つである Twitter は、2020 年のデイリーアクティブユーザ数が 1 億 8700 万人であると報告されている [1]。このようにソーシャルメディアが生活に密接になったことにより、多くの投稿データが、日々蓄積され続けている。さらにソーシャルメディアの特性として政治、経済、時事ニュースなど、社会現象に沿った情報がリアルタイムに書き込まれるため、そのデータ集合には、社会の興味や関心、世の中の情勢が強く反映されていることが期待できる。このデータの特性を生かし、ソーシャルメディアが研究対象としても注目されている。

佐野ら [2] は 18 億を超えるブログデータを分析し、ニュースなど外的要因によって急速に話題になった単語の投稿回数の増加や減少が、べき関数でモデル化されることを明らかにしている。さらに Zhao ら [3] は、デマニュースと正しいニュースの Twitter 上での拡散の違いに着目し、最初の投稿から 5 時間の段階での拡散の性質の相違を明らかにしている。これらの研究分野は社会経済物理学と呼ばれ、物理法則に従う粒子を、アナロジー的に人間個人に置き換えることにより、社会に潜むさまざまな社会現象を理解しようとする比較的新しい分野である [4]。

### 1.2 集合的記憶の想起について

集合的記憶とは、国や家族など社会性のある特定のグループが共有する、過去の出来事に対する記憶である [5]。例えば、日本人の多くが東日本大震災という過

去の記憶を共有している。これは日本人というグループが、東日本大震災という過去の記憶を共有しているため集合的記憶と言える。また集合的記憶は、外的要因によってふたたび想起（思い出）されることがある。

Kanhabua ら [6] は自然災害や、事故、テロ行為などの人的災害が発生した際に、過去に発生した同様の事例が想起されることを Wikipedia のページビューを用いて明らかにし、その想起スコア（想起の大きさ）について、想起イベントからの時間経過が短く、発生地点間の距離が近いほどその値が大きくなるという傾向を明らかにしている。渡辺は、英語の Wikipedia ページへのアクセスログにおいて、ある俳優の訃報が入った際に、その俳優と共演履歴のある俳優の Wikipedia ページビューが大きく増加し、その後、1 日あたりのページビュー数がべき関数的に減少していくことを示した [7]。

この集合的記憶の想起に関しては、まだ明らかにされていない部分が多くある。その 1 つが忘却のメカニズムである。佐野ら [2] は、日本語のブログを用いて突発的なニュース後の書き込み数の推移がべき関数に従うことを示したが、1 日単位のスケールであり、より細かな時間スケールでもべき関数に従っているのかは、明らかになっていない。また 2010 年以降、存在感を増している Twitter のデータにおいても同様の忘却現象が起きているのかは、まだ確認されていない。そこで本研究では、想起現象が起きた後の忘却について、Twitter における 1 時間単位の時系列データを用いて数理モデル化検証を行い、想起現象の忘却メカニズムについて考察を行う。

<sup>\*1</sup>筑波大学 理工情報生命学術院 システム情報工学研究科  
E-mail: naokigarashi@gmail.com

<sup>\*2</sup>筑波大学 システム情報系

<sup>\*3</sup>筑波大学 システム情報系

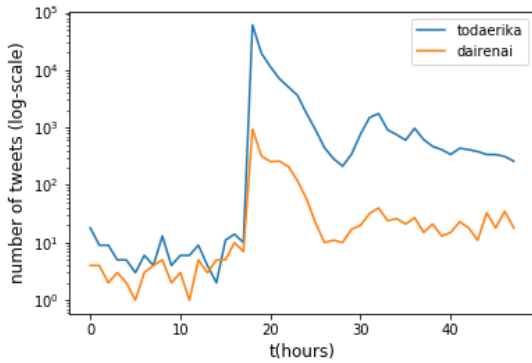


図1 「戸田恵梨香」と「大恋愛」のツイート数

## 2 分析方法

### 2.1 対象データ

本研究では、想起現象を起こす外的要因の例として、一人の俳優（戸田恵梨香）の結婚報道を用いることとした。この報道は世間でも大きな反響を生み、結婚相手である俳優（松坂桃李）の報告ツイートに対しては、ツイートから10分あまりで3.1万リツイート、7万近くの「いいね」が付いたと、Yahoo!ニュースの記事としても紹介されている [8]。このニュースは、人気の俳優同士の結婚だったこともあり、関連するドラマ名や映画名などにいくつかの想起現象が見られた。図1は、結婚ニュースが報道された2020年12月10日～11日の「戸田恵梨香」を含むツイートと、過去に戸田恵梨香が主演し、人気のあったドラマ名「大恋愛」を含むツイートの時系列データである（どちらもRTを除く）。 $t = 0$  は12月10日の0時に対応しており、そこから1時間ごとのツイート数を表示している。この図からもわかるように、「戸田恵梨香」と「大恋愛」がおおよそ同じようなふるまいをしている。すなわち結婚報道により、まず俳優（戸田恵梨香）に注目が集まり、出演していたドラマである「大恋愛」も想起されたことが示唆できる。本研究では、想起現象が示唆された「大恋愛」と同様に、過去に出演したドラマや映画の名前を含むツイートを分析対象とした。ドラマ名や映画名は、当該俳優（戸田恵梨香）のWikipediaページを参考に取得し、ツイートの収集は、Twitter APIを用いて行った。Twitter APIとはTwitter社が提供しているサービスであり [7]、期間やキーワードを指定してツイートを取得することができる。今回は着目する俳優の出演履歴がある「デスノート」「라이어ゲーム」「流星の絆」「エイプリルフールズ」「スカーレット」をキーワードとして指定し、想起されるであろうツイートの収集を行った。また、データ取得期間は結婚報道があった日と翌日（2020年12月10日 2020年12月11日）とし、得られたツイートデータを1時間ごとの時系列

データに変換した。

### 2.2 前処理

モデル化を行うにあたり、1日の周期性をふまえた前処理を行った。1日の中では、ツイート行動が行われやすい時間帯とそうでない時間帯がある。1時間ごとの時系列データでは、この1日の中で変化する周期的な性質が影響を及ぼし、人々のツイート行動がうまくモデル化できなくなる可能性がある。そこで外的要因による影響をほぼ受けない接続詞をランダムに1つ指定し（「それで」を形態素解析の辞書から取得）、その1時間ごとの時系列データを用いて前処理を行った。具体的には、接続詞の時系列データの平均が1となるように相対化し、その逆数を取得したすべての時系列データに乗算した。これにより、ツイート回数自体が少ない時間帯のツイートには重みをもたせ、1日の周期性に影響を受けない時系列データへと変換した。

### 2.3 フィッティング

今回、想起の忘却に関して、以下のべき関数（式 (1)）と指数関数（式 (2)）でフィッティングを行った。

$$W(t) = At^{-\alpha} \quad (1)$$

$$W(t) = Be^{-\beta t} \quad (2)$$

$t$  は結婚ニュースからの時間経過（1時間ごと）を表し、今回は想起が起きた時間を  $t = 1$  とし、その時間は、2日分の時系列データで最大値をとった時間とした。今回はすべてのデータで2020年12月10日の18:00:00からの1時間にツイート数が最大値を取ったことから、1時間分アップサンプリングした12月10日の18:00を、 $t = 1$  としてフィッティングに使う時系列データ  $W(t)$  とした。「エイプリルフールズ」に関しては、1時間ごとに時系列データを取ると、ツイート数が0になる部分が存在してしまった。これについては、集団的性質が見えにくくなってしまうという理由から分析対象から外すことにした。

フィッティング期間については、 $t = 1$  のツイート数  $W(1)$  の  $1/4$  以上のピーク（ $\frac{W(1)}{4}$ ）を10時間以降に持った場合には、その1時間前までをフィッティング期間とした。これはツイートを確認したところ、最初のピークからの忘却とは異なる話題が新たに発生していたためである。この処理を行った結果、フィッティング期間の中央値は30時間となった（表1）。

フィッティング方法は、べき関数と指数関数の特性を利用して行った。べき関数は  $x$  軸と  $y$  軸を対数変換すると線形になり、指数関数は  $y$  軸のみを対数変換すると線形になる。この特性を活かして変換したデータを用いて線形回帰モデルを構築し、最小二乗法によって係数を求めた。さらにモデルの当てはまりの良さを

表1 フィッティングを行なった期間 (単位は時間)

検索語	フィッティング期間
戸田恵梨香	30
大恋愛	30
デスノート	30
라이어ゲーム	30
流星の絆	13
スカーレット	12

表2 フィッティング結果

検索語	$R^2$ (べき)	$R^2$ (指数)	べき指数 $\alpha$
戸田恵梨香	0.88	0.79	1.4
大恋愛	0.86	0.71	1.1
デスノート	0.93	0.81	0.8
라이어ゲーム	0.77	0.62	1.1
流星の絆	0.7	0.58	1.1
スカーレット	0.73	0.64	0.8

決定係数  $R^2$  で評価した。

### 3 分析結果

フィッティング結果は、表2のようになった。

すべての単語についてべき関数における決定係数  $R^2$  が比較的大きくなっている。また、指数関数の  $R^2$  と比較してもべき関数の方がモデルの当てはまりは良いという結果となった。また、べき指数については  $\alpha = 0.8$  から  $\alpha = 1.4$  というある一定の値の範囲で収まるという結果が得られた。このべき指数の値は、日本語のブログや [2] や英語の Wikipedia におけるページビューにおける1日単位の忘却過程で観測できた値 [7] とも近い。図2, 図3に「戸田恵梨香」「大恋愛」それぞれの減衰曲線と、フィッティングしたべき関数をそれぞれ両対数でプロットした。おおよそ直線的に減衰していることから、得られたデータのツイート数がべき関数に従って減衰していることが確認できている。

### 4 まとめ

本研究では、ある俳優(戸田恵梨香)の結婚ニュースで起きた想起現象の忘却について、1時間単位のツイート時系列を用いた分析を行なった。結婚のニュースが報道されると、その俳優の名前を含むツイートが大きく増加するだけでなく、その俳優が過去に出演したドラマや映画の名前も同時に想起され、ツイート数が増加することを確認した。これらのツイート数が増加した後、減衰する時系列のモデル化検証を行い、べき関数でモデル化されることを明らかにした。

これまで、想起現象の定量化には、日次単位のブログや Wikipedia のページビューを用いられてきており、

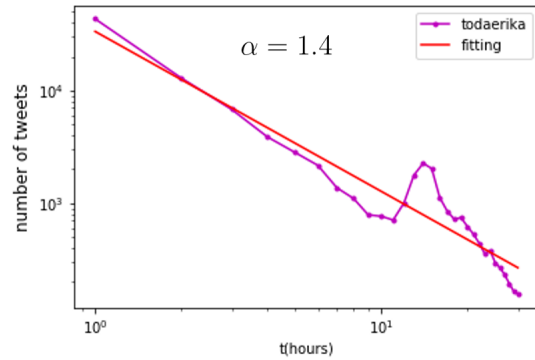


図2 「戸田恵梨香」の減衰曲線とフィッティングしたべき関数 (1)

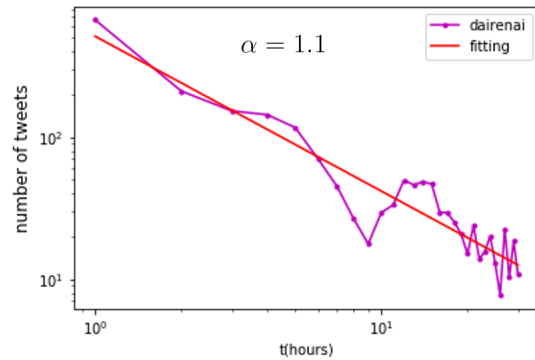


図3 「大恋愛」の減衰曲線とフィッティングしたべき関数 (1)

べき指数がおおよそ1となるべき関数的な推移が報告されている。本研究では、Twitter というこれまでとは異なるソーシャルメディア媒体、さらに異なる時間スケールでも同じ現象を確認することができた。

### 5 今後の展望

今回は、データのフィッティング時に対数変換した時系列に対して、最小二乗法を適用するという簡易的な方法で行なった。今後は、最尤法や非線形最小二乗法を用いてより精緻なフィッティングによる検証を行う必要がある。またフィッティング期間についても慎重に検討する必要がある。

またの本研究成果の一般化に向けて、サンプル数を増やした同様の調査を行う必要がある。今回は一つの事例の想起現象の検証であったため、災害発生時や訃報ニュースなど、さまざまな外的要因による想起現象の忘却についてはまだ検証できていない。今後は総合的なデータ収集から同様の調査を行い、想起現象の忘却の一般的な性質を明らかにする必要がある。さらに、今回収集したツイートデータに含まれるノイズに関しても検討が必要である。キーワードを用いたツイート

収集を行ったが, そのキーワードを含むツイートには, スパム的に入るものなど想起に関するトピックを含まない可能性が存在する. すなわち Twitter から想起に関する話題のみを抽出する技術などを用いて同様の調査をすることで, さらに妥当性の高い研究結果が得られることが期待できる.

#### 参考文献

- [1] Q3 2020 Letter to Shareholders, October 29,2020 @TwitterIR
- [2] Sano, Y., Yamada, K., Watanabe, H., Takayasu, H., & Takayasu, M. (2013). Empirical analysis of collective human behavior for extraordinary events in the blogosphere. *Physical Review E*, 87(1), 012805.
- [3] Zhao, Z., Zhao, J., Sano, Y., et al., (2020). Fake news propagates differently from real news even at early stages of spreading. *EPJ Data Science*, 9(1), 7.
- [4] Buchanan, M., (2008). *The social atom: Why the rich get richer, cheaters get caught, and your neighbor usually looks like you*. Bloomsbury Publishing USA.
- [5] Coser, L. A. (1992). *On collective memory*.
- [6] Kanhabua, N., Nguyen, T. N., & Niedere, C. (2014, September). What triggers human remembering of events? A large-scale analysis of catalysts for collective memory in Wikipedia. In *IEEE/ACM Joint Conference on Digital Libraries* (pp. 341-350). IEEE.
- [7] 渡辺岳人, 「オンラインメディアにおけるトレンド持続期間の解析」, 筑波大学システム情報工学研究科 修士論文 (2020 年)
- [8] <https://news.yahoo.co.jp/articles/5abda538f01771a7139a304d069845dac9a6d721>  
(参照日時:2021 年 2 月 1 日)
- [9] <https://developer.twitter.com/ja/docs>  
(参照日時:2021 年 2 月 1 日)