

# Twitter上で多数の訂正情報が発信される 流言の特徴分析

草竹 大暉<sup>1,a)</sup> 平林(宮部) 真衣<sup>2,b)</sup> 吉野 孝<sup>3,c)</sup>

受付日 2022年4月11日, 採録日 2022年10月4日

**概要:** マイクロブログサービスの1つである Twitter では、正しい情報だけでなく、流言を含む情報も発信されることがある。流言には社会的混乱を引き起こすような内容のものが多く、流言が広がりにくい仕組みを作る必要がある。そこで本研究では、真偽が不明、または真偽が人々に疑われている情報を流言と定義し、Twitter 上で多数の訂正情報が発信される流言の特徴を分析する。分析の結果、Twitter 上で多数の訂正情報が発信される流言の特徴として以下の3点を明らかにした。(1) 訂正ツイートの多さは、流言ツイートの多さを必ずしも反映しない。(2) 流言ツイートと訂正ツイートのバーストのタイミングには、差がほとんどみられず、両者ともに短期的に収束する傾向がある。(3) 今回分析した8つの流言については、いずれも流言ツイートの発生は単発的かつ短期収束傾向がある。一方、訂正ツイートの推移に関しては、その対象となる流言ツイートの Twitter 上での拡散数により、傾向が変わる可能性がある。

**キーワード:** 訂正情報, リツイート, 流言, ツイッター

## Feature Analysis of Rumors in which a Large Amount of Correction Information is Transmitted on Twitter

HIROKI KUSATAKE<sup>1,a)</sup> MAI MIYABE HIRABAYASHI<sup>2,b)</sup> TAKASHI YOSHINO<sup>3,c)</sup>

Received: April 11, 2022, Accepted: October 4, 2022

**Abstract:** Twitter, a microblogging service, transmits not only correct information, but also sometimes transmits information that contains rumors. Some of these rumors may cause social confusion. Hence, it is necessary that a system is created to prevent the spread of such rumors. In this study, the meaning of “rumor” is defined as information of which truth or falsity is unknown or suspicious. Further, the characteristics of the “rumors” being transmitted extensively on Twitter are analyzed. The results reveal three features of such “rumors” in which a large amount of correction information is transmitted on Twitter: (1) The number of correction tweets does not necessarily reflect the number of rumor tweets. (2) There is almost no difference in the burst timing between the rumor tweet and the correction tweet, and both tend to converge in the short term. (3) In all of the eight rumors analyzed in this study, the occurrence of the rumor tweet tended to be one-shot and converged in a short period of time. On the other hand, regarding the trend of the correction tweet, the trend can change depending on the number of the rumor tweet spread on Twitter.

**Keywords:** correction information, retweet, rumor, Twitter

<sup>1</sup> 和歌山大学大学院システム工学研究科  
Graduate School of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan

<sup>2</sup> 東京大学大学院医学系研究科  
Graduate School of Medicine, The University of Tokyo, Bunkyo, Tokyo 113-0033, Japan

<sup>3</sup> 和歌山大学システム工学部  
Faculty of Systems Engineering, Wakayama University, Wakayama 640-8510, Japan

## 1. はじめに

日本国内における SNS の利用者（アクティブユーザ）は年々増加傾向にあり、2022 年末には利用者数は 8,241 万

a) kusatake.hiroki@g.wakayama-u.jp

b) hirabayasi@m.u-tokyo.ac.jp

c) yoshino@wakayama-u.ac.jp

人、ネットユーザ全体に占める利用率は 82.3%に達する見通しである [1]。マイクロブログサービスの 1つである Twitter\*<sup>1</sup>は、情報発信時の入力文字数が最大 140 文字に制限されていることなどから、ユーザは情報発信を手軽に行うことができ [2]、情報が広範囲かつ急速に拡散される特徴を持つ [3]。2011 年 3 月 11 日に発生した東日本大震災時においては、地震と津波の発生後、電話やメールがつかない状況の中で、多くの人が Twitter を利用し災害情報の収集を行い、家族や知人の消息を知った事例が報告されている [4]。

しかし、SNS の普及によって、正しい情報だけではなく、誤情報や虚偽情報を含む情報も広範囲に、そして急速に拡散するようになった [5]。その伝播力は正しい情報よりも高く [6]、2019 年末に発生した新型コロナウイルスの感染が拡大する中、多数の誤情報が拡散された。それにとともに、米 Twitter が 2020 年 12 月 16 日 (現地時間) に、「新型コロナワクチンは人々を操る目的や危害を加える目的で使われる」「新型コロナは深刻な問題ではなく、予防接種を受ける必要はない」といった新型コロナウイルスのワクチン接種に関する誤った情報を含むツイートを削除する方針を発表した\*<sup>2</sup>。このように、流言\*<sup>3</sup>には社会的混乱を引き起こすような内容のものもあり、流言が広がりにくい仕組みを作る必要がある。

人々がある情報を他者に伝える場合、その情報が正しいと思って伝えていることが多く、本人がでたらめだと思つ話を悪意を持って他者に伝えることは少ない [7]。つまり、マイクロブログ上の流言の伝達は、主にユーザが流言であることを認識していないことに起因すると考えられる。もし、そうであるならば、流言の訂正情報\*<sup>4</sup>は、ユーザに対して流言であることに気づきを与えるという、流言拡散防止において重要な役割を果たすと考えられる。

そこで本研究では、真偽が不明、または真偽が人々に疑われている情報を流言と定義し、Twitter 上で多数の訂正情報が発信される流言の特徴を分析する。多数の訂正情報が発信される流言は、多くのユーザから注目を受け、閲覧されている可能性があり、通常の流言と比べて、流言拡散防止のための知見が現れやすいと考えられる。

本研究では、実際に多数の訂正情報が発信された流言に関して、過去のツイートを大規模に収集・分析することで、流言拡散防止のための仕組みの構築に役立つ知見を得ることを目指す。

\*<sup>1</sup> <https://twitter.com>

\*<sup>2</sup> ITmedia NEWS 「Twitter、新型コロナワクチン関連の偽情報ツイートを削除へ」<https://www.itmedia.co.jp/news/articles/2012/17/news087.html>

\*<sup>3</sup> 本研究では、十分な根拠がなく、その真偽が不明、または真偽が人々に疑われている情報を流言と定義し、その発生過程で悪意を持つかどうかは問わないものとする。

\*<sup>4</sup> マイクロブログ上では、流言に対して訂正情報が発信されることがある。

## 2. 関連研究

### 2.1 流言の定義や伝達に関する研究

流言の古典的な分類としては、ナップによる第 2 次世界大戦時の流言の分類がある [8]。ナップは、流言を「恐怖流言 (不安や恐れへの投影としての流言)」「願望流言 (願望への投影としての流言)」「分裂流言 (憎しみや反感への投影としての流言)」の 3 つに分類した。これらの流言がどの程度の割合で流通するかは、社会状況によって決まると述べられており、ナップの第 2 次世界大戦中の流言の研究では分裂流言 (66%) が圧倒的に多かったとされている。社会状況は流言を伝達させる要因の 1 つであり、たとえば震災の直後など、社会状況が多くなると不安を感じさせる状況は、流言の発生や伝達に関係する。

また、流言の伝達には、不安、曖昧さ、信用度の 3 つの指標が強く関係することが示されている [7]。オルポートとポストマンは、流言の流布量を、 $R \sim i \times a$  のように定式化し、「流言の流布量 (R) は、当事者に対する問題の重要さ (i) と、その議題についての証拠の曖昧さ (a) の積に比例する」と述べている [9]。

このように、古くから流言に関する研究が行われてきたが、これらの流言に関する先行研究は、現実社会の中での口伝えでの流言の伝達について行われたものであり、インターネット上での流言の伝達に関する分析は行われていない。本研究では、流言が広範囲かつ急速に拡散するマイクロブログの 1 つである Twitter を対象とする点でこれらの先行研究と異なる。

### 2.2 情報の信頼性判定に関する研究

Qazvinian らは、ナイーブベイズ分類器を構築し、Twitter 上の流言を自動的に検出する手法を提案した [10]。学習データとして、評価者がアノテーション付けを行った 10,000 件以上のツイートをを用いた。学習時には、ツイートの内容、ユーザ情報、Twitter 独自の情報を特徴量として使用し、交差検証を行ったところ、提案手法は MAP\*<sup>5</sup>で 0.95 以上を達成した。

Castillo らは、報道価値のある情報の信頼性を自動的に評価する手法を提案した [11]。Twitter の投稿の信頼性について、評価者がアノテーション付けを行ったツイートを様々なモデルを用いて学習させた結果、メッセージの伝播方法に違いがあり、70%から 80%の範囲の適合率と再現率で、メッセージを信頼できるものと信頼できないものとして自動的に分類することに成功した。

鳥海らは、Twitter 上の流言に対して注意喚起を行っているツイートを抽出することで、ツイートが流言であるかを 5 段階で評価するシステムを提案した [12]。システムは、

\*<sup>5</sup> Mean Average Precision

判定対象となる文が入力されると、訂正情報から独自のアルゴリズムにより算出したスコアを付与し、スコアに応じて「流言の可能性が非常に高い」から「流言だという情報はない」まで5段階の評価を与える。流言を含むツイートおよび正しい情報を含むツイートについて流言かどうかの判定を行ったところ、提案システムは、適合率（流言と判定されたツイートが流言だった割合）が80.6%、再現率（流言ツイートのうち流言と判定された割合）が82.9%であり、最終的な判定精度は81.7%であった。

インターネット上の膨大な情報について、すべての真偽確認を手で行うのは困難であるからこそ、情報の信頼性を自動判定するシステムの開発は、流言の拡散を防ぐうえで重要な役割を果たすと考えられる。本研究では、これらの情報信頼性の判定に関する研究に活用できる知見を得ることが目的である。

### 2.3 災害時のソーシャルメディアに関する研究

Mendozaらは、2010年のチリ地震におけるTwitterの研究を行っている[13]。この研究では、それぞれ7種類の正しい情報と流言に関するツイートを、「支持」「否定」「疑問」「不明」に分類し、正しい情報と流言との違いを分析している。分析結果として、正しい情報を否定するツイートの割合は0.3%と非常に低いが、流言を否定するツイートの割合は約50%にのぼることを示した。

Quらは、四川大地震および青海地震において中国のオンラインフォーム（BBS）がどのように利用されたのかについて分析している[14]、[15]。四川大地震に関する研究では、災害後の復旧段階における緊急対応システムの設計に焦点を当てた分析を行った[14]。オンラインフォーム上での災害関連の情報を分類し、オンラインフォーラムが、特に災害関連の情報とユーザの意見の表明の役割をサポートするうえで災害に迅速かつ効果的に対応していたことが分かった。また、地震後に発生する流言の真偽を明確にすることが重要であるため、信頼できる情報源からの情報は、モデレーターによって異なる色で強調表示されていたとも述べている。青海地震に関する研究では、大災害時にマイクロブログは、災害状況、意見、感情、行動の呼びかけに関する情報の発信に利用されていたことが分かった[15]。

災害時のソーシャルメディアの利用方法について分析した研究としては、LonguevilleらやViewegらの研究がある[16]、[17]。Longuevilleらは、2009年にフランスで発生した森林火災に関してTwitterに発信されたツイートの分析を行っている[16]。ツイートの発信者の分類では、既存の情報を引用するユーザの割合は31%の割合であった。あるメディアによって発表された情報が誤りであった場合、訂正情報が同じ経路をたどってユーザに伝達するとは限らない。そのため、情報元とそれを引用する情報を識別する必要があることを示した。Viewegらは、2009年のオクラ

ホマの家事やレッドリバーでの洪水におけるTwitterの利用方法を調査している[17]。この研究では発信された内容を分類し、情報の発信の方法（情報発信か返信か）や、その位置関係について議論し、緊急時に生成されたツイートがユーザにどのような影響度を明らかにするフレームワークを提案した。しかしこれらは、情報が流言かどうかといった観点からの分析は行われていない。

平林らは、新型コロナウイルス感染症流行時（2019年12月1日から2020年5月26日）において拡散された流言を「非常事態関連の流言」として分析した[18]。収集された流言全体において、非常事態関連の流言が占める割合は少ないが、期間中の各日において、非常事態関連の流言の最大の割合は61%であることを示すなど、Twitter上で関連キーワードの増加を注視するなどの対策の検討に役立つ分析を行っている。

本研究では、「災害時」や「非常事態」のような特定の期間に限定せず、ある期間に訂正された流言について、過去のツイートを大規模に収集・分析することで、流言拡散防止の仕組みの検討材料となりうる流言の特徴を明らかにすることを旨とする。

## 3. Twitter上で多数の訂正情報が発信される流言の特徴分析

### 3.1 分析の概要

ある情報が流言であった場合、ユーザはそれを認識せずに拡散しているならば、流言の訂正情報は、ユーザに対して流言であることに気づきを与えるという重要な役割を果たすと考えられる。本研究では、流言の拡散防止において重要であると考えられる「訂正」に着目し、訂正情報が発信される流言の中でも特に、多数の訂正情報が発信される流言に着目した分析を行う。多数の訂正情報が発信される流言は、多くのユーザから注目を受けた可能性があり、流言拡散防止のための知見が現れやすいと考えられる。

本研究では、以下の項目を検証することで、流言拡散防止のための仕組みの構築に役立つ知見を明らかにする。

#### 検証項目1 流言ツイート数と訂正ツイート数の比較

Twitter上に流布している流言ツイートがユーザの目にとどまる機会が多いほど、多数の訂正情報が発信される可能性がある。そこで、流言ツイート数と訂正ツイート数の比較を行い、「訂正ツイート数が多いとTwitter上の流言ツイート数も多い」という仮説を検証する。

#### 検証項目2 流言ツイートと訂正ツイートの時系列における関係性

Twitterにおいて流言が発生してから収束するまでにどの程度の期間がかかるのか、また、収束には訂正が効果的であったのかを知ることは、Twitter上の流言対策を講じるうえでの重要な知見となりうる。そこで、流言ツイートと訂正ツイートの時系列における関係性を分析

する。検証項目2では、「流言ツイートと訂正ツイートの発生と収束のタイミング」と「流言ツイートと訂正ツイートが発生してから収束するまでの期間」を調査する。

### 3.2 流言情報クラウド

本節では、流言情報クラウド\*6について概説する。

“流言情報クラウド”は、「○○○はデマです」のような、流言を訂正しようとする表現を含む情報（以降、訂正情報と呼ぶ）を抽出することにより、その訂正情報に含まれる流言情報をTwitterから自動的に収集するサービスであり、2012年6月から継続的に情報を蓄積している[19]。このサービスは、「デマ」や「嘘」といった訂正に関わるキーワードを含むツイートを収集したうえで、それらが本当に流言を訂正しているツイート（訂正ツイート）かどうかを自動判別し、訂正ツイート中の流言内容をパターンマッチングにより抽出・蓄積している。本研究では、流言情報クラウドが収集した流言情報を分析対象の選定材料として用いている。

### 3.3 分析対象データの収集

本研究では、分析対象のデータを以下の手順で収集する。

(1) Twitter上で多数の訂正情報が発信される流言の抽出  
まず、流言情報クラウドで2020年9月1日から2020年9月30日\*7に収集された訂正ツイート数（訂正数）を集計する。次に、訂正数が上位10件の流言を「多数の訂正情報が発信される流言」として抽出\*8した。抽出した流言とその訂正数を表1に示す。

(2) 分析対象とする流言の選別

手順(1)で抽出した流言を確認し、同一内容を含む流言の統合および内容が不明確な流言の除外を行い\*9、抽出した10種類の流言のうち、8種類の流言を本研究の分析対象として選定した。

(3) Twitter APIによる分析対象データの収集

Twitter APIを用いて、手順(2)で分析対象となった8種類の流言（以降、AからHと表記する）に関するツイートを収集する。各流言を表すうえで最低限含まれる必要があると考えられる語句を検索クエリとし、

表1 流言情報クラウドにおいて訂正数が上位10件の流言  
Table 1 Rumors of the top 10 corrections in the rumor information cloud.

順位	流言名	訂正数
1	新型コロナには正露丸が効く	2,959
2	X知事リコールのための署名をすると署名した人の個人情報が漏洩する	1,905
3	潰瘍性大腸炎なのに焼肉食いまくってた	1,470
4	出産費用は高級などこじゃない限り実質無料でしょ	1,056
5	STAP細胞が再現された	1,040
6	朝鮮人が井戸に毒を入れたり、放火・略奪をし、人を襲っている	983
7	AはBだ	401
8	「品川に津波」「首相暗殺」と並んで「朝鮮人が井戸に毒を投げ入れた」	349
9	日本で育つと誰でも性的暴行を受ける	311
10	社会学者が『小説Y』を書き換えろと言っている	278

\*2位および10位の流言中のXとYは個人名と作品名のため伏せ字としている。7位の流言に含まれているAとBは伏せ字ではなく、原文そのままを掲載している。

2006年3月21日\*10から2021年4月30日までの期間のツイートを収集した。流言AからHに関するツイートの収集結果を表2に示す。表2には、流言AからHに関するツイートの収集数と収集するにあたって用いた検索クエリを示している。

### 3.4 分析対象データの分類

本節では、ツイートの分類方法について概説する。

#### 3.4.1 分類対象データ

Twitter上で拡散（リツイート）された数は、多くのユーザに閲覧されたことを表す指標として用いることができる。そこで本研究では、Twitter APIで収集したツイートのうち、50件以上リツイートされたツイートを分類対象とする。ツイート数の集計時には、分類対象となったツイートに紐付いているリツイートもまとめて集計する。

#### 3.4.2 分類方法

本研究では、Mendozaら[13]がTwitter上の流言に関するツイートの分類において用いたカテゴリをもとに、以下の4つの分類カテゴリを定義し、人手で分類を行う。

- 流言ツイート
  - － 「流言に関する内容を事実のように述べている表現」を含むツイート
  - － 「流言に関する内容を事実のように述べている表現＋客観性がある根拠」を含むツイート（例：○○というのは、科学的知見に基づいて立証されている。）
  - － 「流言に関する内容を事実のように述べている表現＋その流言に対して支持的な表現」を含むツイート（例：

\*10 Twitter サービスが始まった日付。

\*6 <http://mednlp.jp/~miyabe/rumorCloud/rumorlist.cgi>

\*7 2020年9月1日から2020年9月30日の期間は、流言情報クラウドのデータ収集において欠損のない期間であり、ユーザの流言への訂正パターンを分析するうえで、1カ月の期間は最低限の期間であると考えられる。

\*8 訂正数が上位10件の流言の訂正数の合計は10,752件で、2020年9月1日から2020年9月30日に流言情報クラウドで収集された16,739件の訂正ツイート数のうちの64.2%であり、訂正数が上位10件の流言に対する訂正ツイートのみで訂正ツイート全体の半数以上を占めている。

\*9 表1の8位の流言には6位の流言と同じ内容が含まれるため、「朝鮮人は井戸に毒を入れた」という1つの流言として同等に扱う。また、表1の7位の「AはBだ」という流言は、不明確な内容であるため、分析の対象外とした。

表 2 流言 A から H に関するツイートの収集結果  
Table 2 Collection results of tweets about rumor A to H.

	流言	検索クエリ	収集数
A	新型コロナには正露丸が効く	コロナ & 正露丸 & 効く	19,505
B	X 知事リコールのための署名をすると署名した人の個人情報が漏洩する	(X   愛知) & リコール & 署名 & 漏洩	30,581
C	潰瘍性大腸炎なのに焼肉食いまくってた	潰瘍性大腸炎 & 焼肉	12,276
D	出産費用は高級などこじゃない限り実質無料でしょ	出産費用 & 高級 & 無料	1,679
E	STAP 細胞が再現された	STAP 細胞 & 再現された	1,783
F	朝鮮人が井戸に毒を入れた	朝鮮人 & 井戸 & 毒	116,264
G	日本で育つと誰でも性的暴行を受ける	日本 & 育つ & 性的暴行	85,490
H	社会学者が「小説 Y」を書き換えると言っている	社会学者 & 小説 Y & 書き換えろ	3,916

※流言 B および流言 H 中の X と Y は個人名と作品名のため伏せ字とした。

〇〇という説は、否定され続けているが、私は真実だと思う。）

- 「流言に関する内容を事実のように述べている表現 + 流言に関連しない内容の羅列」を含むツイート（例：最近のニュースまとめ「〇〇という説」「△△という説」「□□という説」）
  - 「流言に関する内容を事実のように述べている表現 + その流言を交えた小話」を含むツイート（例：〇〇が起こつたらしい。実際、私の祖母にも〇〇のようなことが起こつた。）
  - 訂正ツイート
    - 流言であることを指摘するツイート
  - 疑問ツイート
    - 流言に対して、疑問を表すツイート（例：〇〇は本当なの？）
  - その他ツイート
    - 流言とは関連のないツイート
    - 上記のカテゴリのいずれにもあてはまらないツイート
- 次に、先行研究における噂への対処方法の定義 [7] に基づいた分類を試みる。先行研究では、噂への対処戦略として、以下の 3 種類があると述べられている。
- 否定戦略
    - 流言の内容を明確に否定する。
  - 対抗戦略
    - 流言について否定しない。流言自体とは異なるイメージを流す。
  - 無視戦略
    - 流言に対して反応せず、流言が流れるままにしておく。
- これらの対処戦略のうち、否定戦略・対抗戦略は流言に対して何らかの反応をする、という戦略である。つまり、Twitter 上において、流言に対して何らかのツイートが流れることになる。たとえば、「出産費用は高級などこじゃない限り実質無料」のような内容の流言ツイート（なお、ここで提示するツイート例は実際に Twitter で投稿されたそのままの情報ではなく、事例を分かりやすくするために著者らが一部修正・作成した内容である）が流れた場合を

考える。「出産費用が実質無料というのはデマです」のように、流言を否定する表現を含むようなツイートは否定戦略と考えられる。また、「出産費用は病院や入院する部屋の種類、分娩方法によって額は異なってきますが、**実際正常分娩では 30~70 万円です**」のように、流言の内容を直接的に否定する表現は含まないものの、異なる情報（ここでは、流言で述べられている「無料」に対し、「30~70 万円」という情報を述べる）を含むようなツイートは、対抗戦略と考えることができる。

一方、無視戦略は、流言に対して反応しない、という戦略である。つまり、Twitter 上においては、閲覧した流言ツイートに対して反応しないことが無視戦略と考えられる。しかし、閲覧したうえで無視したのか、見逃したために反応がない（無視した）のかを、取得したツイートのログから判断することは難しく、無視戦略が行われたかどうかを判断することは困難である。そこで、本研究では訂正ツイートを、無視戦略を除いた、否定戦略と対抗戦略に分類する。

また人手での分類作業については、評価者 1 名（評価者 A）がすべてのツイートについて行った。しかし、評価者 A の評価結果が客観的かどうかは不明確であったため、以下の手順により、評価者 A の評価結果が一般的なものであるかどうかを検証した。

- (1) 評価者 A が分類した、否定戦略・対抗戦略・流言・疑問・その他に該当するツイートから、A から H の各流言について、それぞれ最大 100 件ずつ（100 件未満のものはすべて）ランダムサンプリングする。
- (2) 2 名の評価者に、ランダムサンプリングしたツイートの分類を依頼する。
- (3) 結果が一致するかを検証する。

3 人の Kappa 値はそれぞれ 0.38, 0.286, 0.323 であり、Kappa 値の解釈方法<sup>\*11</sup>の 1 つである Landis らの目安 [20] においては「Fair (ある程度の一致)」という結果となった。

<sup>\*11</sup> Kappa 値は測定の一貫性を示す代表的な指標である。1 は完全な一致であるが、0~1 未満の値の解釈の目安は複数存在し、本研究では、1977 年に Landis らが示した目安 [20] を用いる。

表 3 ツイートの分類結果  
Table 3 Tweet classification results.

	流言	訂正		流言	疑問	その他	合計
		否定戦略	対抗戦略				
A	新型コロナには正露丸が効く	7,497 (51.1%)	6,643 (45.2%)	545 (3.7%)	0 (0%)	0 (0%)	14,685 (100%)
B	X 知事リコールのための署名をすると署名した人の個人情報が漏洩する	28,483 (96.8%)	227 (0.8%)	451 (1.5%)	220 (0.7%)	54 (0.2%)	29,435 (100%)
C	潰瘍性大腸炎なのに焼肉食いまくってた	6,592 (60.9%)	1,019 (9.4%)	3,216 (29.7%)	0 (0%)	0 (0%)	10,827 (100%)
D	出産費用は高級なとこじゃない限り実質無料でしょ	1,019 (61.1%)	0 (0%)	650 (38.9%)	0 (0%)	0 (0%)	1,669 (100%)
E	STAP 細胞が再現された	1,079 (93.2%)	79 (6.8%)	0 (0%)	0 (0%)	0 (0%)	1,158 (100%)
F	朝鮮人が井戸に毒を入れた	50,821 (63.0%)	29,129 (36.1%)	292 (0.4%)	303 (0.4%)	62 (0.1%)	80,607 (100%)
G	日本で育つと誰でも性的暴行を受ける	42,749 (57.8%)	30,374 (41.1%)	546 (0.7%)	241 (0.3%)	0 (0%)	73,910 (100%)
H	社会学者が『小説 Y』を書き換えろと言っている	133 (3.8%)	76 (2.2%)	3,258 (94.0%)	0 (0%)	0 (0%)	3,467 (100%)

※百分率は、「合計」列のツイート数に対するそれぞれの分類カテゴリのツイート数の割合を表している。

### 3.4.3 分類結果

3.4.2 項の手順で分類した結果を表 3 に示す。表 3 はそれぞれの分類カテゴリにおけるツイート数を表しており、「合計」列は、それぞれの分類カテゴリにおけるツイート数の合計 (3.4.1 項で分類対象データとした流言 A から H の「リツイートが 50 件以上なされたツイート」と「そのリツイート」の総数) を表している。

## 4. 分析結果

### 4.1 検証項目 1: 流言ツイート数と訂正ツイート数の比較

本研究では、訂正ツイートを否定戦略と対抗戦略に分類した。一般的には「訂正」とは誤りを正しく直すという意味で使用され、流言の内容を明確に否定すると定義をした否定戦略の方が本来の訂正の意味に近い。そこで、本節では否定戦略に分類されたツイート (以降、否定戦略ツイートと呼ぶ) を訂正ツイートと見なし、流言ツイートと否定戦略ツイートの数の比較により検証項目 1 を検証する。

表 3 より、否定戦略ツイートの割合は、流言 H は 3.8% と低いが、流言 H 以外の流言については、50% 以上であり、分類を行ったツイートの大部分を占めている。それに対し、流言ツイートの割合は、流言 H は 94.0% と高いが、流言 H 以外の流言は 40% 未満 (流言 A, B, E, F, G においては 5% 未満) となっていることから、今回分析対象とした流言については、3.1 節で立てた「訂正ツイート数が多いと Twitter 上の流言ツイート数も多い」という仮説は成り立たなかった。

Twitter において、訂正ツイートの多さは、流言ツイートの多さを必ずしも反映するものではなく、訂正ツイート

のみが拡散している場合があるため、流言拡散防止の仕組みを構築する指標として「流言への訂正の多さ」を用いることには注意を払う必要があると考えられる。

### 4.2 検証項目 2: 流言ツイートと訂正ツイートの時系列における関係性

本節では、表 3 において、訂正対象となる流言ツイートが一定頻度確認された流言 (流言ツイートの割合が 2% 以上) A, C, D, H と、訂正ツイートの数と比較してほとんど流言ツイートが流れていない流言 (流言ツイートの割合が 2% 未満) B, E, F, G に大別し、流言ツイートと訂正ツイートの時系列における関係性を分析する。

流言 A, C, D, H の流言ツイートと否定戦略ツイートの時系列の推移を図 1 に、流言 B, E, F, G の流言ツイートと否定戦略ツイートの時系列の推移を図 2 に示す。図 1 と図 2 は、横軸を 1 日区切りとし、縦軸は 1 日あたりのツイート数を表す。また、青色の棒グラフが流言ツイート数、赤色の棒グラフが否定戦略ツイート数を表す。ただし図 2 の流言 F は、流言ツイートと否定戦略ツイートの出現期間が長く、すべての期間の流言ツイートと否定戦略ツイートに同じ傾向の時系列の推移がみられたため、2018 年 1 月 1 日から 2021 年 4 月 30 日までの期間に絞って時系列の推移を示している。

#### 4.2.1 訂正対象となる流言ツイートが一定頻度確認された流言

流言ツイートが一定頻度 (割合が 2% 以上) 確認された流言 A, C, D, H における流言ツイートと訂正ツイートの発生と収束に関する特徴を述べる。

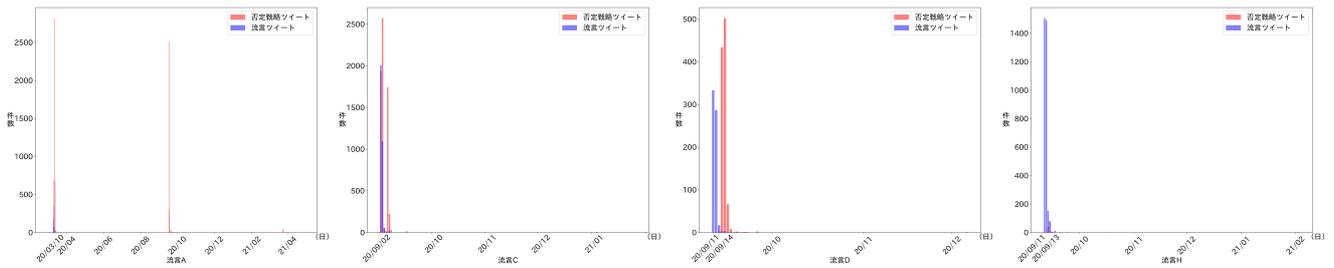


図 1 流言 A, C, D, H の流言ツイートと否定戦略ツイートの推移

Fig. 1 Changes in rumor tweets and negative strategy tweets regarding rumor A, C, D, and H.

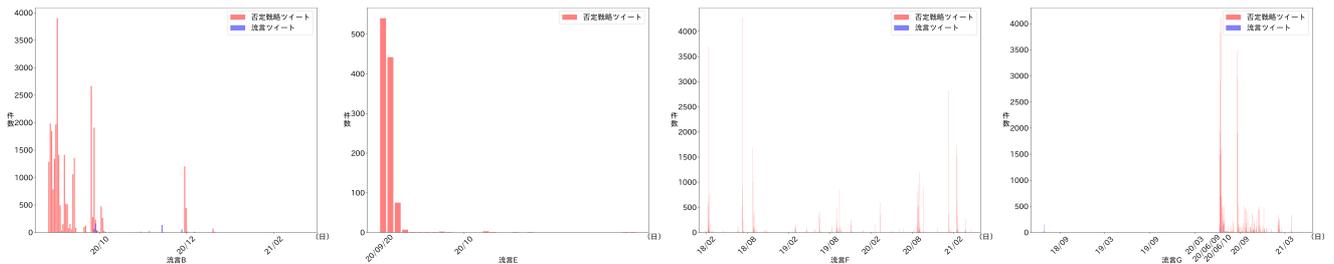


図 2 流言 B, E, F, G の流言ツイートと否定戦略ツイートの推移

Fig. 2 Changes in rumor tweets and negative strategy tweets regarding rumor B, E, F, and G.

**流言 A** 図 1 から、2020 年 3 月 10 日に流言ツイートと否定戦略ツイートが突発的に発生し、その数日後、両者ともに収束している。

**流言 C** 図 1 から、2020 年 9 月 2 日に流言ツイートと否定戦略ツイートが突発的に発生し、その数日後、両者ともに収束している。

**流言 D** 図 1 から、2020 年 9 月 11 日に流言ツイートが、2020 年 9 月 14 日に否定戦略ツイートが突発的に発生し、その後、両者ともに収束している。

**流言 H** 図 1 から、2020 年 9 月 11 日に流言ツイートが、2020 年 9 月 13 日に否定戦略ツイートが突発的に発生し、その後、両者ともに収束している。

以上のことから、これらの流言は流言ツイートの発生から否定戦略ツイートが発生するまでの日数の差はほとんどなく、両者ともに突発的に発生し収束する傾向がみられた。すなわち、これらの流言は流言ツイートのバースト<sup>\*12</sup>と訂正ツイート（否定戦略ツイート）のバーストの発生のタイミングには、差がほとんどみられず、両者ともに短期的に収束する傾向があることが分かった。

Twitter 上の流言は、バースト的に発生し、短期間で収束する場合があります。その話題を信じたユーザによるさらなる流言の拡散を防ぐために、リアルタイムにユーザへ注意喚起を促す必要がある。

#### 4.2.2 訂正ツイートの数と比較してほとんど流言ツイートが流れていない流言

訂正ツイートの数と比較してほとんど流言ツイートが流れていない（割合が2%未満）流言 B, E, F, G における

流言ツイートと訂正ツイートの発生と収束に関する特徴を述べる。

**流言 B** 図 2 から、否定戦略ツイートが1カ月以上わたって長期的に発生し、その後収束している。

**流言 E** 図 2 から、2020 年 9 月 20 日に否定戦略ツイートのみが突発的に発生していることが分かる。

**流言 F** 図 2 から、数年にわたって否定戦略ツイートが長期的に発生し、その後収束する傾向がみられない。

**流言 G** 図 2 から、2020 年 6 月 9 日に否定戦略ツイートが発生し、その1日後の2020 年 6 月 10 日から2020 年 6 月 12 日に流言ツイートの発生が確認できるが、それ以降は流言ツイートの発生は確認できず、否定戦略ツイートによる流言の訂正が続いている。

以上のことから、流言 B, F, G は、Twitter 上の流言ツイートが拡散しているかどうかにかかわらず、1 月から数年にわたって訂正ツイートが発生する場合があった。

Twitter においては、訂正ツイートが拡散されていても、流言ツイートが拡散されていない場合がある。

## 5. おわりに

Twitter 上で多数の訂正情報が発信される流言を「1 カ月間における訂正数が上位 10 件の流言」と定義し、それらの流言に関するツイートを収集し、以下の項目を検証した。  
**検証項目 1** 流言ツイート数と訂正ツイート数の比較  
**検証項目 2** 流言ツイートと訂正ツイートの時系列における関係性

検証の結果、Twitter 上で多数の訂正情報が発信される流言に関して、以下の特徴がみられた。

<sup>\*12</sup> 本研究では、ツイート数が平常時と比べて急激に増加することをバーストと定義する。

**知見 1** 訂正ツイートの多さは、流言ツイートの多さを必ずしも反映しない。

**知見 2** 流言ツイートのバーストと訂正ツイートのバーストの発生のタイミングには、差がほとんどみられず、両者ともに短期的に収束する傾向がある。

**知見 3** 今回分析した8つの流言については、いずれも流言ツイートの発生は単発的かつ短期収束傾向がある。一方、訂正ツイートの推移に関しては、(A) Twitter上で訂正対象となる流言ツイートが一定頻度確認された流言（流言ツイートの割合が2%以上）と、(B) 訂正ツイートの数と比較してほとんど流言ツイートが流れていない流言（流言ツイートの割合が2%未満）とで違いがみられた。(A)に該当する4つの流言については、流言ツイートと同様に、訂正ツイートも単発的発生かつ短期収束傾向であったが、(B)に該当する4つの流言のうち3つについては、1カ月から数年にわたって訂正ツイートが発生する場合があった。

これらの知見は、訂正数が上位の流言に限定したものであり、Twitter上のすべての流言に適用できるわけではない。今後、訂正数が中位や下位の流言に対しても同じことがいえるのか調査する必要がある。

**謝辞** 本研究の一部は、JSPS 科研費 19H04221 の助成による。

#### 参考文献

- [1] ICT 総研：2016 年度 SNS 利用動向に関する調査，入手先 (<https://ictr.co.jp/report/20200729.html>).
- [2] 西谷智弘：I 見聞録：Twitter 研究会，情報処理学会誌，Vol.51, No.6, pp.719-724 (2010).
- [3] 風間一洋：Twitter における情報伝搬，人工知能学会誌，Vol.27, No.1, pp.35-42 (2012).
- [4] 山本雅人，小笠原寛弥，鈴木育男，古川正志：東日本大震災時の Twitter における情報伝搬ネットワーク，情報処理学会誌，Vol.53, No.11, pp.1184-1191 (2012).
- [5] 福長秀彦：流言・デマ・フェイクニュースとマスメディアの打ち消し報道，放送研究と調査，Vol.68, No.11, pp.84-103 (2018).
- [6] Vosoughi, S., Roy, D. and Aral, S.: The Spread of true and false news online, *Science*, Vol.359, No.6380, pp.1146-1151 (2018).
- [7] 川上善郎：うわさが走る 情報伝搬の社会心理，pp.32-47, サイエンス社 (1977).
- [8] Knapp, R.H.: A Psychology of Rumor, *Public Opinion Quarterly*, Vol.8, No.1, pp.22-37 (1944).
- [9] G.W. オルポート，ポストマン：デマの心理学，pp.41-60, 岩波書店 (2008).
- [10] Qazvinian, V., Rosengren, E., Radev, D.R. and Mei, Q.: Rumor has it: Identifying misinformation in microblogs, *Proc. Conference on Empirical Methods in Natural Language Processing*, pp.1589-1599, ACM (2011).
- [11] Castillo, C., Mendoza, M. and Poblete, B.: Information Credibility on Twitter, *Proc. 20th International Conference on World Wide Web*, pp.675-684 (2011).
- [12] 鳥海不二夫，篠田孝祐，兼山元太：ソーシャルメディアを用いたデマ判定システムの判定精度評価，情報処理学会，

- デジタルプラクティス，Vol.3, No.3, pp.201-208 (2012).
- [13] Mendoza, M., Poblete, B. and Castillo, C.: Twitter under crisis: can we trust what we RT?, *Proc. 1st Workshop on Social Media Analytics (SOMA '10)*, pp.71-79 (2010).
  - [14] Qu, Y., Wu, P.F. and Wang, X.: Online Community Response to Major Disaster: A Study of Tianya Forum in the 2008 Sichuan Earthquake, *42nd Hawaii International Conference on Systems Science (HICSS-42 2009)*, pp.1-11 (2009).
  - [15] Qu, Y., Huang, C., Zhang, P., et al.: Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake, *Proc. ACM 2011 Conference on Computer Supported Cooperative Work (CSCW '11)*, pp.25-34 (2011).
  - [16] Longueville, B.D., Smith, R.S. and Luraschi, G.: "OMG, from here, I can see the flames!": A use case of mining location based social networks to acquire spatiotemporal data on forest fires, *Proc. 2009 International Workshop on Location Based Social Networks (LBSN '09)*, pp.73-80 (2009).
  - [17] Vieweg, S., Hughes, A.L., Starbird, K., et al.: Microblogging during two natural hazards events: What twitter may contribute to situational awareness, *Proc. 28th international conference on Human factors in computing systems (CHI '10)*, pp.1079-1088 (2010).
  - [18] 平林(宮部)真衣，吉野 孝，河添悦昌：新型コロナウイルス感染症流行時における Twitter 上の流言訂正情報に関する分析，情報処理学会論文誌，Vol.63, No.1, pp.29-44 (2022).
  - [19] 宮部真衣，灘本明代，荒牧英治：人間による訂正情報に着目した流言拡散防止サービスの構築，情報処理学会論文誌，Vol.55, No.1, pp.563-573 (2014).
  - [20] Landis, J.R. and Koch, G.G.: The measurement of abserver agreement for categorical data, *International Biometric Society*, Vol.33, No.1, pp.159-174 (1977).



草竹 大暉 (学生会員)

2022 年和歌山大学システム工学部システム工学科卒業。現在，同大学大学院システム工学研究科システム工学専攻博士前期課程に在学中。流言の拡散防止に関する研究に従事。



平林(宮部) 真衣 (正会員)

2006 年和歌山大学システム工学部デザイン情報学科中退。2008 年同大学大学院システム工学研究科博士前期課程修了。2011 年同大学院システム工学研究科博士後期課程修了。博士(工学)。現在，東京大学大学院医学系研究科特任助教。コミュニケーション支援，ソーシャルメディア分析に関する研究に従事。



吉野 孝 (正会員)

1992年鹿児島大学工学部電子工学科卒業。1994年同大学大学院工学研究科電気工学専攻修士課程修了。博士(情報科学)。現在、和歌山大学教授。コミュニケーション支援およびビッグデータの社会的応用に関する研究に興味を持つ。本会シニア会員。