

# Twitter 上で多数の訂正がなされる流言の特徴分析

## Feature Analysis of Rumors that Frequently Corrected on Twitter

草竹 大暉<sup>†</sup> 平林 (宮部) 真衣<sup>‡</sup> 吉野 孝<sup>†</sup>  
 Hiroki Kusatake Mai Miyabe Hirabayashi Takashi Yoshino

### 1. はじめに

近年, SNS の普及によりインターネット上での個人の情報発信が盛んに行われている [1]. 特に Twitter<sup>1</sup> では, ユーザは気軽に情報を発信・拡散し, 必要な情報をリアルタイムで取得することができる. また, Twitter のようなソーシャルメディアは, 実世界のリアルタイムな状態を観測するために活用されている [2]. 東日本大震災時には, 電話やメールが繋がらない状況の中で, 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も報告されている [3].

しかし, Twitter などの SNS の普及によって誤情報がかつてなく広範囲に, そして急速に拡散するようになった [4]. 誤情報は正確な情報よりも拡散されやすいとされている [5]. 実際に, 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, 多数の流言<sup>2</sup>が拡散した. 米 Twitter が 2020 年 12 月 16 日 (現地時間), 「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>3</sup> というニュースも存在する. このように, 流言には社会的混乱を引き起こすような内容のものもあり, ユーザの情報に対する注意力を高める仕組みが必要である.

Twitter では, 発信・拡散された流言に対し, 訂正情報が発信されることがある. 訂正情報はユーザへの注意喚起のために有用である. そこで我々は, 訂正がなされる流言の中でも多くのユーザから注目を受け, 閲覧されたと考えられる多数の訂正がなされる流言に着目する. 本稿では, 多数の訂正がなされる流言の特徴を明らかにすることで, 流言に対する効果的な注意喚起の仕組みの構築に役立つ知見を得ることを目指す.

### 2. 関連研究

Mendoza らは, 2010 年のチリ地震における Twitter の研究を行っている [6]. 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

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

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 愛知県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』を書き換えると言っている	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た. 7 位の流言に含まれている A と B は伏せ字ではなく, 原文そのまま掲載している.

合は約 50 % にのぼることを示した. また, 疑問を示すツイートの割合は, 正しい情報より流言に関する情報の方が多いということを示した. 本研究では, 災害時における流言の分析ではなく, Twitter において訂正数が多い流言に着目した分析を行う.

宮部らは, 平常時のデータおよび災害時のデータから抽出した, 8 種類の流言を用いた研究を行った [7]. この研究では, 流言に関するツイートを 3 種類 (流言ツイート, 疑問ツイート, 訂正ツイート) に分類し, マイクロブログ上における流言の特徴として以下の 3 点を明らかにした.

- 流言の発信は, 不特定多数に向けたリツイートをインフラとしており, 口伝えでの流言の拡散と比較して, 流言内容の変容が起こりにくい可能性がある.
- 流言に関するツイートを複数回投稿するユーザは少なく, また流言情報を投稿したユーザが, 自身の発言を訂正することも少ない.
- 流言に関する訂正ツイートが発信されても, 流言ツイートはすぐには収束しない.

宮部らの研究では, 分析対象とする流言の抽出指標として, 平常時と災害時の流言に着目し, それぞれ抽出を行った. 1 章でも述べたように, 多数の訂正がなされる流言は, 多くのユーザから注目・閲覧されるものであり, Twitter における流言の特徴が現れる可能性があると考えている. そのため本研究では, 多数の訂正がなされる流言を分析対象とし, 分析を行う.

<sup>†</sup> 和歌山大学 システム工学部, Faculty of Systems Engineering, Wakayama University

<sup>‡</sup> 東京大学大学院 医学系研究科, Graduate School of Medicine and Faculty of Medicine, The University of Tokyo

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

<sup>2</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>3</sup> 「Twitter, 新型コロナウイルスワクチン関連の偽情報ツイートを削除へ」ITmedia NEWS, 入手先: <https://www.itmedia.co.jp/news/articles/2012/17/news087.html>

表 2: Twitter API の収集結果

	流言	検索クエリ	総ツイート数	50 件以上リツイートされたツイート数
A	新型コロナには正露丸が効く	コロナ & 正露丸 & 効く	19,505	14,685 (75.3 %)
B	X 愛知県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	29,435 (96.3 %)
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	10,827 (88.2 %)
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	1,669 (99.4 %)
E	STAP 細胞が再現された	STAP 細胞 & 再現された	1,783	1,158 (64.9 %)
F	朝鮮人が井戸に毒を入れた	朝鮮人 & 井戸 & 毒	116,264	80,607 (69.3 %)
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	73,910 (86.5 %)
H	社会学者が『小説 Y』を書き換えろと言っている	社会学者 & 小説 Y & 書き換えろ	3,916	3,467 (88.5 %)

※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。

表 3: ツイートの分類結果

流言	訂正		流言	疑問	不明	合計
	否定戦略	対抗戦略				
A	7,497 (51.1 %)	6,643 (45.2 %)	545 (3.7 %)	0 (0 %)	0 (0 %)	14,685 (100 %)
B	28,483 (96.8 %)	227 (0.8 %)	451 (1.5 %)	220 (0.7 %)	54 (0.2 %)	29,435 (100 %)
C	6,592 (60.9 %)	1,019 (9.4 %)	3,216 (29.7 %)	0 (0 %)	0 (0 %)	10,827 (100 %)
D	1,019 (61.1 %)	0 (0 %)	650 (38.9 %)	0 (0 %)	0 (0 %)	1,669 (100 %)
E	1,079 (93.2 %)	79 (6.8 %)	0 (0 %)	0 (0 %)	0 (0 %)	1,158 (100 %)
F	50,821 (63.0 %)	29,129 (36.1 %)	292 (0.4 %)	303 (0.4 %)	62 (0.1 %)	80,607 (100 %)
G	42,749 (57.8 %)	30,374 (41.1 %)	546 (0.7 %)	241 (0.3 %)	0 (0 %)	73,910 (100 %)
H	133 (3.8 %)	76 (2.2 %)	3,258 (94.0 %)	0 (0 %)	0 (0 %)	3,467 (100 %)

### 3. 分析の概要

#### 3.1 分析対象データ

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 上で多数の訂正がなされた流言の抽出  
流言情報クラウド<sup>1)</sup>で 2020 年 9 月 1 日から 2020 年 9 月 30 日に収集された訂正数が上位 10 件の流言を抽出した。抽出され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  
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、抽象的な内容であるため、分析の対象外とし、表 1 に示した 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用いる。

<sup>1)</sup> 流言情報クラウドでは、流言情報自体を直接収集するのではなく、訂正情報を抽出することにより間接的に流言を収集している。

#### (3) Twitter API<sup>2)</sup>で対象データの収集

手順 (2) で分析対象となった流言に関するツイートを収集する。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、Twitter API で 2006 年 3 月 21 日<sup>3)</sup>から 2021 年 4 月 30 日の間のツイートを収集した。検索クエリと収集したツイート数を表 2 に示す。

#### 3.2 分析方法

本研究では、流言の特徴を明らかにするために、ツイートの分類を行う。まず、以下の 4 種類にツイートを分類する。

- (1) 訂正ツイート：流言であることを指摘するツイート
- (2) 流言ツイート：流言に関するツイート
- (3) 疑問ツイート：流言に対して、疑問を表すツイート (例：○○は本当なの?)

<sup>2)</sup> <https://developer.twitter.com/en/docs/twitter-api>

<sup>3)</sup> Twitter サービスが始まった日付

(4) 不明ツイート：上記の分類指標のどれにも当てはまらない、分析対象の流言とは関連のないツイート

次に、先行研究における定義に基づき、訂正ツイートをさらに分類する。先行研究では、噂への対処方法として、以下の3種類があると述べられている [8].

**否定戦略**：噂の内容を明確に否定する。(例：○○はデマです.)

**対抗戦略**：噂について否定しない。噂自体とは異なるイメージを流す。(例：○○は、どこかの国では□□と言われているらしい.)

**無視戦略**：噂に対して反応せず、噂が流れるままにしておく。

これらの対処方法のうち、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ため、本研究では、否定戦略と対抗戦略に分類する。

多くのユーザによって拡散されたツイートは、他の人に閲覧され、また興味を持たれる内容を含んでいると考えられ、注意喚起のための特徴を持つ可能性がある。そこで本研究では、API で収集したツイートのうち、一定数の拡散がされたツイート（リツイートが 50 件以上なされたツイート）を分類することとした。

### 3.3 検証項目

1 章で述べたように、多数の訂正がなされた流言は、多くの人々が目にした可能性があり、その特徴は注意喚起方法の検討などに利用できる可能性がある。そこで本研究では、「Twitter 上で多数の訂正がなされる流言」について分析する。本研究では、以下の2つのリサーチクエスチョン (RQ) を立てる。

**RQ1**：多数の訂正がなされる流言は、その流言ツイートも多いのか？

**RQ2**：多数の訂正がなされる流言と訂正の時間的推移はどうなっているのか？

まず、RQ1 を検証する理由として、多数の訂正がなされる流言は、その流言自体の数が多いため、多くの注意喚起を促すユーザの目に留まり、多数の訂正がなされると考えられるためである。

次に、RQ2 を検証する理由として、多数の訂正がなされる流言の訂正のなされ方を分析することで、流言を特徴づける知見が得られると我々は考えたためである。訂正のなされ方を分析するために、RQ2 では、流言ツイートと否定戦略ツイートの時間的推移に着目した分析を行う。

## 4. 分析結果

ツイートの分類結果を表3に示す。本研究では、ユーザ自身が作成したと考えられるツイート（ツイート、引用ツイート、リプライ）と拡散のみが目的と考えられるツイート（リツイート）を同等のツイートとして扱うため、表3の分類結果の表し方として、拡散されたツイートとそのリツイートを合わせた総数で表している。

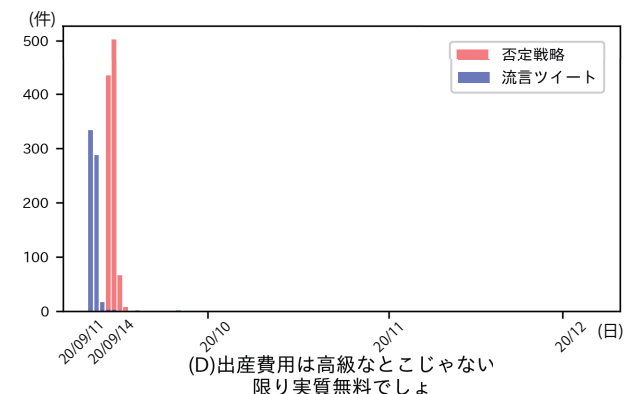
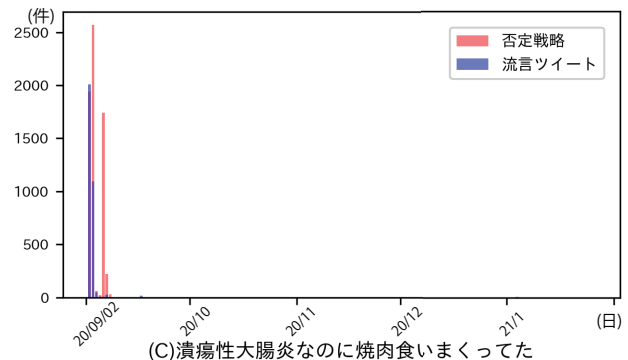
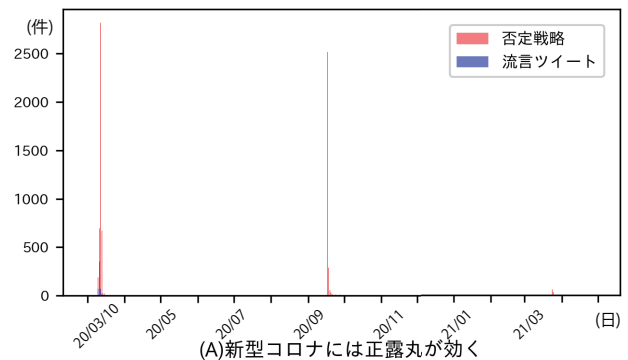


図 1: 流言ツイートと否定戦略ツイートの推移

### 4.1 「RQ1：多数の訂正がなされる流言は、その流言ツイートも多いのか？」の検証

表3では、訂正ツイートを否定戦略と対抗戦略に分類を行った。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、噂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、流言ツイートと否定戦略ツイートの数の比較により、RQ1 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 % と低いですが、流言 H 以外のすべての流言で、50 % 以上であり、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 以外のすべての流言は、40 % 未満（流言 A, B, E, F, G においては 5 % 未満）となっている。また、流言 A~H の流言ツイート数と否定戦略ツイート数の相関係数は -0.41 であり、弱い負の相関がみられるのみであるため、流言数と訂正数の間には、ほとんど関係性がないことがわかる。

以上をまとめると、訂正ツイートの多さは、流言ツイートの多さには結びつかない可能性がある。

#### 4.2 「RQ2:多数の訂正がなされる流言と訂正の時間的推移はどうなっているのか?」の検証

流言ツイートと否定戦略ツイートの数のバランスが比較的とれた流言 A, C, D を分析対象とし、流言ツイートと訂正ツイートの時間的推移における関係性を分析する。

流言 A, C, D の流言ツイートと否定戦略ツイートの推移を図 1 に示す。図 1 に示すように、流言ツイート、否定戦略ツイートのいずれも、バースト<sup>4</sup>後にツイートの発信がほとんど見られないことから、今回分析した流言については、流言の発生期間と訂正期間は短いことが分かった。

また図 1 に示すように、流言 A では、2020 年 3 月 10 日に、流言 C では、2020 年 9 月 2 日に流言ツイートと否定戦略ツイートのバーストが同じ日に発生した。さらに、流言 D については、2020 年 9 月 11 日に流言ツイートのバーストが発生した後、2020 年 9 月 14 日に否定戦略ツイートのバーストが発生し、流言ツイートのバーストが発生してから、否定戦略ツイートのバーストが発生するまでの日数の差は 3 日であった。以上のことから、流言のバーストと訂正のバーストが行われるまでの日数の差はほとんどなく、流言のバーストと訂正のバーストの発生のタイミングには、差がほとんど見られない可能性がある。

### 5. おわりに

本研究では、Twitter 上で多数の訂正がなされる流言の特徴を明らかにするために分析を行った。分析対象として、流言情報クラウドにおいて 2020 年 9 月に訂正数が上位の流言に関するツイートを Twitter API で収集した。また、流言に関するツイートを分類し、流言ツイートと否定戦略ツイートに着目した分析を行った。

分析の結果、Twitter 上で多数の訂正がなされる流言には、以下の傾向があることを明らかにした。

- 訂正ツイートの多さは、流言ツイートの多さを反映するものではない。
- 多くの人が閲覧した可能性のある、多数の訂正がなされた流言については、流言と訂正がバースト的に発生し、短期間で収束する傾向がある。
- 流言のバーストと訂正のバーストの発生のタイミングには、差がほとんど見られない可能性がある。

今後、注意喚起の仕組みの構築におけるこれらの知見の活用方法について検討を行う。

### 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] 垂水浩幸:実世界インタフェースの新たな展開:4. ソーシャルメディアと実世界, 情報処理学会誌, Vol.51, No.7, pp.782-788 (2010).

- [2] 榊剛史, 松尾豊:ソーシャルセンサとしての Twitter: ソーシャルセンサは物理センサを凌駕するのか?, 人口知能学会誌, Vol.27, No.1, pp.67-74 (2012).
- [3] 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).
- [4] 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).
- [5] Soroush Vosoughi, Deb Roy, Sinan Aral: The Spread of true and false news online, Science, Vol.359, No.6380, pp.1146-1151 (2018).
- [6] 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).
- [7] 宮部真衣, 梅島彩奈, 灘本明代, 荒牧英治:マイクロブログにおける流言の特徴分析, 情報処理学会誌, Vol.56, No.1, pp.223-236 (2013).
- [8] 川上善郎:うわさが走る-情報伝搬の社会心理学, サイエンス社 (1997).

<sup>4</sup>本研究では、ツイート数が通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