

# Twitterにおける誤情報の拡散収束過程の可視化

渡邊研斗<sup>†</sup> 鍋島啓太<sup>†</sup> 水野淳太<sup>†</sup> 岡崎直観<sup>†‡</sup> 乾健太郎<sup>†</sup>  
 東北大学<sup>†</sup> 科学技術振興機構 さきがけ<sup>‡</sup>

{kento.w, nabeshima, junta-m, okazaki, inui}@ecei.tohoku.ac.jp

## 1 はじめに

東日本大震災では、電話やメールよりも大量の情報を即座に得られる Twitter が情報提供・収集に大きく貢献した。しかし、必要以上の情報が行き交い「コスモ石油の爆発で有害な雨が降る」などの誤情報が拡散し、人々の混乱の要因となった [1, 2]。これまで震災時の誤情報を収集する研究 [3, 4] が報告されているが、誤情報の拡散を防ぐための方策にまで踏み込んだ研究は少ない。そこで、本研究では以下の3つの課題に取り組む。

- 誤情報の発生拡散収束のケース・スタディを行い、誤情報を早く収束させる要因を分析する。
- 誤情報の発生から訂正情報が流れるまでの時間や、誤情報の拡散を食い止めるまでの時間を計測し、集合知に基づく誤情報の自動検出手法の実用性を探る。
- 誤情報の自動検出手法として、訂正表現を用いたシンプルな手法を提案し、その性能を評価する。

## 2 誤情報はどのように拡散し、収束したのか？

まず、東日本大震災で拡散した誤情報の発生、拡散、収束のケース・スタディを行った。分析では、東日本大震災ビッグデータワークショップで Twitter Japan より配布された震災直後一週間分の全ツイートを対象に、鍋島ら [3] の手法で獲得した 14 件の誤情報を用いた。各誤情報（例えば「コスモ石油の爆発で有害な雨が降る」）に対し、適切に検索クエリ（例えば「コスモ石油 AND 雨」）を選ぶことにより、誤情報を拡散するツイート、訂正するツイートの両方を区別せずに収集した。それらのツイートを誤情報（誤情報を拡散・支持する情報）、訂正情報（誤情報を訂正・阻止する情報）、その他（誤情報に言及していない情報）に手作業で分類した（表 1）。全部で 5195 件のツイートを対象とし、2462 件の誤情報ツイート、2376 件の訂正情報ツイート、357 件のその他のツイートを同定した。

このように用意したコーパスを用いて、各誤情報の発生から訂正情報が出始めるまでの時間 (TTC: Time To Correction) や、誤情報の数が減り始めるまでの時間 (TTS: Time To Suppress) を計測した。表 1 のコスモ石油の例のように、長い時間拡散し続ける誤情報は社会に大きな損害をもたらす。しかし、誤情報の発生から最初の訂正情報が出始めるまでの時間は、概ね数時間である。よって、訂正情報をうまく捉えることで、誤情報の検出と注意喚起を行うことは十分可能である。

更に、誤情報の個別のケースを詳しく調べるため、誤情報の「拡散」ツイートと「訂正」ツイートの数を、それぞれ一定時間おきに折れ線グラフにプロットし、誤情報の拡散状況を可視化するシステムを開発した。このシステムでは、各時点でどのようなツイートが拡散してい

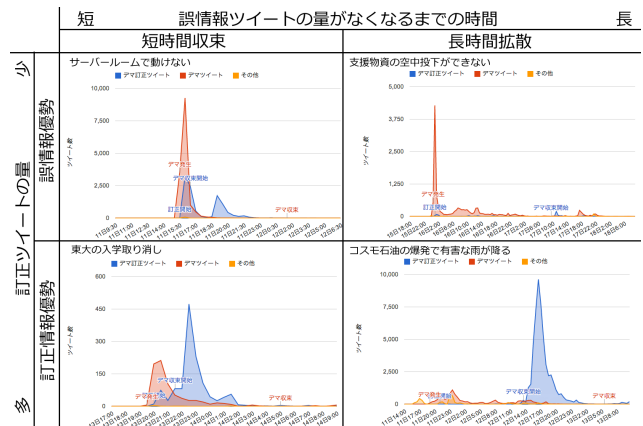


図 1: 4 種類に分けられる拡散状況

たのか、ツイート本文を閲覧できるようになっている。なお、グラフにプロットするツイートの数はリツイート数も考慮し、ツイート空間上での情報の拡散状況を表している。14 件の誤情報に対して、本システムで可視化したグラフを観察すると、誤情報の拡散状況は、主に訂正ツイートの量と収束までの時間で特徴づけられることが分かった（図 1）。

これらの 2 つの要素の組み合わせにより、大きく 4 種類の拡散収束状況に分類できる。例えば、誤情報ツイート数が訂正ツイート数を上回り、かつ誤ツイート量が 0 になるまでの時間が 1 日未満なら、訂正情報劣勢・短時間収束型である。以降では、訂正情報の数や収束までの時間を決定づける要因について考察する。

**訂正情報の量を決定する要因:** 誤情報より訂正情報の量が少ない場合、訂正情報の信憑性・影響力が小さいことが考えられる。例えば「支援物資の空中投下は法律で認められていない」という誤情報において、「許可があれば可能」という訂正情報が流れたが、決定的な証拠や公式発表がなかった。このため、危機意識に駆られた人々が誤情報をどんどん拡散し、訂正情報が浸透しなかった。

逆に訂正情報の量が誤情報より多い場合、訂正情報の信憑性・影響力が強いことが考えられる。例えば「被災者の合格者が期限までに書類を提出できないと東大の入学が取り消される」という誤情報に対し、東大がウェブサイト上で「合格者本人の意志を確認せずに入学の資格を取り消すようなことはありません」と発表した。人々の不安を取り除くに十分な訂正情報だったため、誤情報よりも訂正情報の量が多くなり、誤情報を効果的に抑制することができた。

**収束までの時間を決定する要因:** 誤情報が収束するまでの時間が短いということは、有効な訂正情報が早期

表 1: 誤情報の拡散収束過程の分析 (TTC: 訂正情報が出始めるまでの時間, TTS: 誤情報の数が減り始めるまでの時間)

誤情報	誤情報拡散からの時間 (h)		正解データの比率			自動分類の評価		
	TTC	TTS	誤情報	訂正情報	その他	再現率	適合率	F1
トルコが 100 億円支援	1.5	3.5	100	43	17	0.914	1.000	0.955
ONEPEACE 作者尾田栄一郎が 15 億円寄付	2	1	170	134	7	0.902	1.000	0.949
コスモ石油の爆発で有害な雨が降る	2	16	382	499	98	0.871	0.995	0.929
阪神大震災では三時間後に最大の揺れが来る	0.5	2	506	84	20	0.797	1.000	0.887
阪神大震災でレイプが多発した	2	1.5	69	82	1	0.841	0.932	0.884
ポケモンクリエーターの田尻智が死去	0.5	0	2	36	3	0.750	1.000	0.857
支援物資の空中投下が認められていない	0.5	32.5	38	58	69	0.741	0.977	0.843
サーバーラックが倒れて動けない	0	1	742	401	12	0.678	0.996	0.807
フジテレビの募金は日本ユニセフに行く	0	4	82	64	7	0.578	0.948	0.718
放射線対策にイソジン(うがい薬)が利く	9.5	32.5	162	700	63	0.490	0.985	0.654
東大が合格者の入学取り消し	0.5	1.5	140	81	28	0.419	0.918	0.576
天皇陛下が京都御所へ避難	12	0	25	129	17	0.108	0.875	0.193
福島第一原発が核爆発の恐れ	0	0.5	16	45	13	0.111	0.625	0.188
辻本補佐官が米軍の救助活動に抗議	2	0	28	16	2	1.000	0.380	0.551
平均	2.3	6.9				0.657	0.902	0.714

に現れたということである。例えば「サーバルームで身動きが取れない」という誤情報では、この情報の発信者がジョークとしてつぶやいていたことが早期に発覚したため、短期間で収束した。逆に誤情報が収束するまでの時間が長い場合、有効な訂正情報の出現が遅いことが考えられる。例えば「コスモ石油の爆発で有害な雨が降る」という誤情報は、コスモ石油本社からの「そのような事実はない」という発表が遅れたため、長時間にわたり拡散した。

このように、真偽の検証に必要な情報の信憑性・入手性により、誤情報の拡散と訂正の過程が変化する。

### 3 誤情報の拡散・訂正ツイートの自動分類

前節では、ツイートを誤情報、訂正情報、その他に人手で分類した。しかし、情報の混雑が発生している際に、大量のツイートを手作業で分類することは現実的ではない。そこで、誤情報に関するツイート群を、誤情報の「拡散」もしくは「訂正」に自動的に分類した。本研究では「デマ」や「風説」などの訂正表現を含むツイートを「訂正情報」とし、含まないものを「訂正情報ではない」ツイートとした。訂正表現は震災時のツイートを読みながら、121 個を用意した。

表 1 に、本手法が訂正情報を認識する精度(再現率・適合率・F1 スコア)を示した。この評価では、リツイートは削除し、オリジナルのツイートのみ用いている。表 1 によると、訂正表現に対するマッチングという単純な手法でも、高い適合率が得られることが分かる。つまり誤情報に関するツイートにおいて「デマ」などの訂正表現を含むものは、かなりの確度で訂正情報と見なせる。「デマ」という語を伴って誤情報の拡散を行うことは、通常では考えにくいので、これは直感的に理解できる。

しかし、量は少ないものの、訂正表現を含む誤情報拡散ツイートも見受けられた。

万が一原発から放射能が漏れ出した際、被爆しない為にイソジンを 15 cc 飲んでおいて下さい! 原液です! **ガセ**ネタではありません。お医者さんからの情報です。これは RT ではないので信じてください!

このツイートでは「ガセ」という訂正表現を含んでいるが「ガセ」をさらに否定しているので、二重否定により誤情報の拡散ツイートと解釈できる。さらに、訂正表現を用いずに誤情報を否定するツイートも存在する。

厚生労働省です不特定多数の方に送信されている、コスモ石油千葉製油所における火災関連のメールについては、厚生労働省からの発表情報ではありませんのでご留意願います

このツイートでは「デマ」「嘘」などの訂正表現は一切使われておらず、また誤情報の内容(「コスモ石油の火災により有害物質の雨が降る」)も説明していないが、内容から誤情報を訂正するツイートであると判断できる。このようなツイートを訂正ツイートと認識するためには、深い言語解析(例えば「火災関連のメール」を「火災により有害物質の雨が降るというチェーンメール」と解釈する)や、ツイートやユーザ間の関係(例えば、厚生労働省はこの誤情報に関連して別のツイートで訂正表現を用いて打ち消しを行った、等の手がかり)を用いる必要がある。

### 4 おわりに

本研究では東日本大震災時の内容を誤情報、訂正情報、その他に手作業で分類し、誤情報の発生・拡散・収束の過程を調査した。また、訂正表現に基づく誤情報の自動検出の可能性を、実用面、技術面から検証した。さらに、訂正表現のマッチングにより、誤情報と訂正情報の自動分類を行う手法を提案し、評価を行った。この手法は非常にシンプルであるが、高い精度を達成することができた。今後は誤情報・訂正情報の分類のリアルタイム化や自動分類の精度向上などに取り組む予定である。

### 謝辞

本研究は、文部科学省科研費(23700159)、およびJST 戦略的創造研究推進事業さきがけ、および総務省・情報通信ネットワークの耐災害性強化のための研究開発事業の一環として行われた。貴重なデータを提供して頂いた Twitter Japan 株式会社感謝いたします。

### 参考文献

- [1] 萩上チキ. 検証 東日本大震災の流言・デマ. 光文社, 2011.
- [2] 情報支援プロボノ・プラットフォーム(iSPP). 3.11 被災地の証言—東日本大震災 情報行動調査で検証するデジタル大国・日本の盲点—. インプレスジャパン, 2012.
- [3] 鍋島啓太, 水野淳太, 岡崎直観, 乾健太郎. 訂正パターンに基づく誤情報の抽出と集約. 情報処理学会 第 75 回全国大会 発表論文集, 2013.
- [4] 宮部真衣, 梅島彩奈, 灘本明代, 荒牧英治. 流言情報クラウド: 人間の発信した訂正情報の抽出による流言収集. 言語処理学会 第 18 回年次大会 発表論文集, pp. 891-894, 2012.