

Web ページに含まれる収集された流言情報の抽出手法の検討

Discussion of Extraction Method of Collected False Rumor Contained on Web Page

梅本 美月[†] 吉野 孝[†] 宮部 真衣[‡]
Mizuki Umemoto Takashi Yoshino Mai Miyabe

1. はじめに

近年, SNS をはじめとしたソーシャルメディアの普及により, 誰もが手軽に情報を取得, 発信可能になっている [1]. その中でも特に, Twitter は情報の投稿容易性, 即時性, 伝搬性が高く, 国内のアクティブユーザが 4000 万人を超える, 大規模なマイクロブログサービスである [2]. その即時性の高さから, 2011 年 3 月の東日本大震災では重要な情報インフラとしても利用された [3].

一方, 投稿が容易であり, 誰でも情報を発信できるという特性上, 常に正しい情報のみが発信されるわけではなく, 流言¹が伝播されるという問題がある. 東日本大震災で実際に拡散された流言の一つに, 「放射性物質にはうがい薬が効く」という内容のものがあり, この流言に対しては, 放射線医学総合研究所から訂正と注意喚起がなされた [4]. 流言の中には, 専門知識を持たないユーザーにとって真偽の判断が困難なものも存在し, その情報を鵜呑みにすると, 身体に有害な影響を及ぼしたり, 不安を煽られ混乱が悪化したりする可能性がある.

そこで我々は, 流言の拡散を防止する仕組みが必要であると考え, Web 上に存在する流言情報を検出してユーザーに提示するシステムを開発した [5]. このシステムは, Web 上に存在する流言情報を検出し, 該当部分に強調表示と補助説明表示を行うことで, ユーザーに気づきを与え, 情報の真偽確認を促進することを目的としている. しかし, 現在のシステムでは, ユーザーの長期継続利用を妨げる大きな原因の一つである誤検出や, 実際の流言情報部分以外の文章に強調表現されるなどといった問題がある. 本システムは, ユーザーの普段の検索行動の中で Web ページ上に補助的に表示を行うものであり, ユーザーのタスクを妨げないために, 不要な表示や過剰な表現は避けるべきであると考えられる. 本稿では, Web テキスト内の流言部分の抽出手法を改善し, 適切な強調表示を行うための検討をする.

2. 関連研究

Takahashi ら [6] は, Twitter において流言の発生自体を減らすことは困難であると考え, 訂正ツイートが発信されることで流言の拡散を抑制することを目指した. そこで, ツイートの固有表現やリツイート比率, キーワードにもとづくフィルタリングにより, Twitter における流言検出手法を提案した. また Castillo ら [7] は, ソーシャルメディアの中に情報の信頼性判断のもとになる要素があると考え, ソーシャルメディア上に存在する情報のみを用いて, 情報の信頼性を自動評価する手法を提案した.

[†] 和歌山大学, Wakayama University

[‡] 公立諏訪東京理科大学, Suwa University of Science

¹ 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, 発生過程での悪意の有無は問わないものとする.

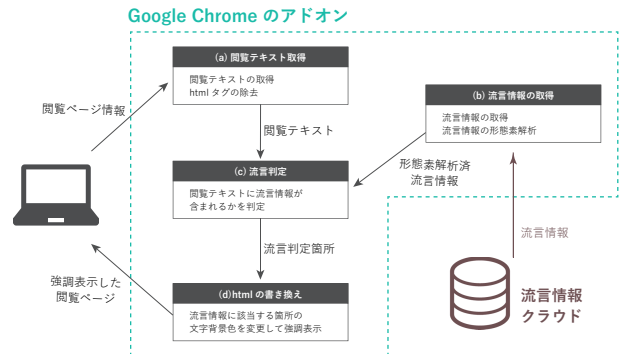


図 1: システム構成



図 2: システムの動作画面例

しかし, これらの研究はソーシャルメディアの中の文章が流言か否かを判定することに焦点を当てた研究である. 本システムは, 既に流言であると判定され収集された流言情報を利用し, 同様の内容が Web ページに含まれる場合, ユーザーに情報を提示する. 検出され, 収集された流言情報を適切に提供する方法が実現されることで初めて, 流言拡散防止に繋がると考えられる. 本稿では, 収集された流言情報と同様の内容を検出し, 該当部分を適切に抽出する方法について検討する.

3. 流言の確認行動促進システム

3.1 概要

本システムの構成を図 1 に, 本システムの動作画面例を図 2 に示す. 本システムは Google Chrome の拡張機能として実装している. システムはまず, テキスト情報を取得する (図 1-(a)). 次に, リアルタイムに流言情報を収集している流言情報クラウド [9] から, 情報取得用のサーバを用いて流言情報を取得し (図 1-(b)), (a) で取得したテキスト内に (b) で取得した流言情報が含まれるかを判定する (図 1-(c)).

表 1: 従来の手法で判定した強調表示部分の例

	流言情報	流言情報が含まれる Web テキスト	理想的な強調表示部分	従来手法での強調表示部分	表示の成否
(ア)	韓国人や在日の方々は NHK 受信料を全員免除	NHK 関係者在日韓国人受信料を全員免除暴露し大炎上	NHK 関係者在日韓国人受信料を全員免除	NHK 関係者在日韓国人受信料を全員免除暴露し大炎上	成功
(イ)	蛇口に青い印があれば飲み水、そうでなければ飲めない水	手で回すタイプの蛇口は「ハンドルの中心に青い印があれば飲み水、色がないものは飲めない水が出る」——そんなツイートが 5 万回以上 RT されて拡散中です	手で回すタイプの蛇口は「ハンドルの中心に青い印があれば飲み水、色がないものは飲めない水が出る」	手で回すタイプの蛇口は「ハンドルの中心に青い印があれば飲み水、色がないものは飲めない水が出る」——そんなツイートが 5 万回以上 RT されて拡散中です。結論から言うと、印の有無で飲み水か否かを判断することはできません。	成功（ただし多すぎる）
(ウ)	ニコニコと Abema と地上波で全部内容が違う	(なし)	(なし)	昼間に放送されている地上波テレビの内容が「貧乏人に許される娯楽は憎しみと怒りだけ」という感じでとてもつらい	失敗（誤判定）
(エ)	飯塚幸三の息子は安倍首相の秘書	池袋暴走「息子は安倍首相の秘書」は誤情報 国会議員もブログを拡散、謝罪	息子は安倍首相の秘書	(なし)	失敗（判定漏れ）

(a) のテキストに (b) の流言情報が含まれていた場合、閲覧ページにおける (a) に該当する部分の背景色を変更する強調表現を行い (図 1-(d)), 該当した流言の情報をトースト表示²で画面の右上に表示する。また、強調表示された部分にユーザがマウスカーソルを合わせると吹き出しを表示し、吹き出し表示にマウスカーソルを合わせると吹き出し表示内容を補足する、吹き出し詳細表示を行う。

3.2 従来の流言部分の判定方法

本システムは、流言と思われる情報に対し「誤った情報が含まれる」とユーザに伝えることが目的ではなく、「真偽が不明瞭な情報が含まれる」という気づきを与えることが目的である。本システムの流言判定においては、適合率 (流言判定が正確かどうか) の高さよりも、再現率 (流言が網羅できているかどうか) の高さを重視している。よって、流言情報の形態素解析結果の名詞が、テキスト (閲覧ページの html 要素のうち、最もネストが深い箇所) 内に全て含まれる場合に流言情報として判定する、という基準で検出している。

しかし、名詞が含まれるかどうかのみの判定を行うと、特に流言情報が短文の場合に誤検出が頻発する。再現率を重視するとはいえ、流言に関係のない箇所が多数強調表示されてしまうと、ユーザにとって不要な表示が増えてしまい、本来のタスクの妨げになってしまう可能性が高い。そこで、名詞が含まれるかどうかの判定に加え、レーベンシュタイン距離 [8] を用いた判定基準を併せて用いている。本システムでは、テキストの長さを考慮した判定を行うために、2つのテキストのレーベンシュタイン距離を、長いほうのテキストの文字数で割った「標準化レーベンシュタイン距離」を用いる。

3.3 従来の流言部分の判定例

表 1 に従来の手法で判定した強調表示部分の例を示す。表 1-(ア) は、流言情報に対して適切な長さの強調表示がされた例である。これは、強調表示された部分が Web ページの見出しであり、html 要素 (この場合は < h1 > 要素) 内が簡潔であったため、適切な長さの表示になっている。一方、表 1-(イ) は、流言情報に対して過剰な範囲が強調表示された例である。これは、強調表示された部分が Web ページの記事本文の一部であり、html 要素内 (この場合は < p > 要素) が長文であったためである。

表 1-(ウ) は、流言情報と強調表示された部分の内容が異なる例である。これは、流言情報を形態素解析して名詞を抜き出した結果が「地上」「波」「内容」となり、この3つが全て判定部に含まれたため、該当すると判定されている。表 1-(エ) は、流言情報と同様の内容が含まれるのに対して、システムによる強調表示がされなかった例である。これは、流言情報を形態素解析して名詞を抜き出した結果が「飯塚」「幸三」「息子」「安部」「首相」「秘書」となり、判定部に「飯塚」「幸三」が含まれなかったため、強調表示されていない。

これらの結果から、従来の「名詞が全て含まれる」かつ「標準化レーベンシュタイン距離が一定以下」という判定基準で「html 要素の最もネストが深い部分を強調表示」という手法は、適切な長さで流言部分を強調表示するのに対して不十分であると考えられる。

3.4 提案手法

3.3 節で述べた問題点より、流言情報部分を適切に抽出する手法を提案する。この手法は、流言情報に含まれる名詞以外の単語も利用し、同じ単語が近接して存在する部分を抜き出す。ここで利用する単語は、自立語の中でも、単語として意味を持つと考えられる「動詞」「形容詞」「形容動詞」「連体詞」「副詞」「名詞」である。また、Web 上の文章を形態素解析する場合、形態素解析の辞書に登録されていない新語や造語が「未定義語」として分類される。流言

² 本論文では、システム上の機能を指す。閲覧ページ上の決まった位置に一定時間表示されるポップアップ表示のこと。

表 2: 従来手法と比較した提案手法の適用結果例

	流言情報	従来手法の抽出部分	理想的な抽出部分	提案手法の抽出部分	従来手法との比較
(a)	韓国人や在日の方々はNHK 受信料を全員免除	NHK 関係者が在日韓国人受信料を全員免除暴露し大炎上	NHK 関係者が在日韓国人受信料を全員免除	NHK 関係者が在日韓国人受信料を全員免除	同程度の抽出範囲
(b)	ローソンは三菱商事の子会社で役員に安倍の兄がいる。ゆえにローソンだけが自衛隊の輸送支援を受けた	反アベ「ローソンだけ三菱商事の子会社で役員に安倍の兄がいるから自衛隊の輸送支援を受けた！」→自衛隊の輸送支援を受けた企業一覧が投下されるwww	ローソンだけ三菱商事の子会社で役員に安倍の兄がいるから自衛隊の輸送支援を受けた	ローソンだけ三菱商事の子会社で役員に安倍の兄がいるから自衛隊の輸送支援を受けた	抽出範囲の精度の向上
(c)	iphone を水に浸すと音質が良くなる	「iPhone5 を水に浸すとスピーカーの音質が良くなる」というデマが流れている、とネットで話題になると、あまりにも分かりやすいデマだと笑いが起こり「塩水がより効果的」「おれサラダ油派」「ロックがオーケストラに聞こえるぞ」などとネットの掲示板やブログには大量の悪ノリの書き込みが登場した。	iPhone5 を水に浸すとスピーカーの音質が良くなる	水に浸すとスピーカーの音質が良く	抽出範囲の不足
(d)	ソニーが FGO を全世界の地域でサービスを開始することを検討	『Fate/Grand Order』の成功は、ソニーミュージックの利益の 23% を占めており、ソニーは、このゲームを全世界の地域でサービスを開始することを検討している。	このゲームを全世界の地域でサービスを開始することを検討	(なし)	判定漏れ
(e)	ニコニコと Abema と地上波で全部内容が違う	昼間に放送されている地上波テレビの内容が「貧乏人に許される娯楽は憎しみと怒りだけ」という感じでもつまらない	(なし)	(なし)	誤判定の改善

はリアルタイムに発生し、最新の情報が含まれる可能性が高いため、流言情報に含まれる「未定義語」も意味のある単語（新しく生まれた重要な単語）として本手法では利用する。以下、本手法で利用する品詞を「有効単語」と定義する。

抽出の詳細な手順を以下に示す。

- (1) 流言情報を形態素解析し有効単語を標準形で抽出する
- (2) 判定するテキストを形態素解析し、全単語を標準形で抽出する
- (3) (2) の抽出単語が (1) の流言情報の有効単語と一致するかどうか順番に確かめる
- (4) 後述する判定条件に基づき、流言情報部分を抽出する。

流言情報部分を判定するとき、有効単語同士が連続して存在している箇所を流言情報部分とみなす。このとき、(3) の判定で連続して有効単語に該当した場合だけでなく、有効単語同士の間在一定数³の別単語が含まれていても連続していると見なす。

4. 提案手法を適用したときのパターン

提案手法の有効性を検証するために、過去に流言情報クラウドで収集された流言情報のうち、流言の訂正数の総数上位 100 件を利用して、Web 上のテキストに提案手法を適用した。このとき、流言情報の文章をそのまま検索欄に入

³この数は文章によって調整が必要である。

力した検索結果の 1 ページ目のうち、流言情報部分に該当すると人手で判定した部分を利用した。また、100 件のうち、不適切な内容⁴だと判断したものや、内容が曖昧で検索が困難なものを除いている。

今回は適用時のパターンを明らかにするため、暫定的に有効単語の欠如を 2 語まで許容し、有効単語同士の間に含まれる別の単語を 2 語まで許容するよう設定した⁵。適用の結果、従来手法と比べて以下の 6 つのパターンが存在することが分かった。

- (a) 従来手法と同等の抽出ができています
- (b) 従来手法よりも適切に流言部分を抽出できている
- (c) 従来手法よりも流言情報部分を絞れているが、流言情報の一部が不足している
- (d) 従来手法では抽出できていた部分を抽出できていない (判定漏れ)
- (e) 従来手法では誤って抽出されていた部分が、抽出されない (誤判定を改善)
- (f) 従来手法では抽出できていなかった部分が、正しく抽出された (判定漏れを改善)

⁴過激または下品な内容など、公序良俗に反する内容もの

⁵この許容数はそれぞれ、適用結果によって今後検討していく必要がある。

表 3: 従来手法で判定できなかった文章への提案手法の適用結果例

	流言情報	流言情報が含まれる Web テキスト	理想的な抽出部分	提案手法の抽出部分	従来手法との比較
(f)	福島県立医科大学病院で多指症手術 100 人待ち	「今、多指症の子どもの手術がいっぱい、100 人近い順番待ちなんです」福島県立医科大学／白血病は寒気がして咳が出るなど風邪と同じ症状になり免疫力が弱って死んでいく。	今、多指症の子どもの手術がいっぱい、100 人近い順番待ちなんです」福島県立医科大学	多指症の子どもの手術がいっぱい、100 人近い順番待ちなんです」福島県立医科大学	判定漏れの改善

表 2 に、提案手法の適用結果のうち (a)~(e) の例を示す。なお、この例は、流言情報の文章と Web テキストの流言情報部分の文章が完全に一致していた場合を除いている⁴⁶。

表 2-(a) は、従来手法と同等の抽出ができた例である。これは、従来手法での抽出部分がもともと適切な長さだったため、結果的にどちらも適切な長さになった。表 2-(b) は、従来手法よりも適切な範囲を抽出できた例である。これは、有効単語の間にあるそれ以外の単語が少なく、提案手法が適切に適用されたためである。表 2-(c) は、従来手法の抽出よりもより流言情報に近い部分のみを抽出できたが、前後の単語がいくつか欠落している例である。これは、2 つの文章の iPhone の部分が、流言情報では「iphone」という未定義語が有効単語として抽出され、Web テキストでは「iPhone5」が未定義語として抽出されたため、抽出されなかったことが原因である。また、「良くなる」の「なる」は「接尾辞」として定義されており、今回の有効単語の定義に当てはまらなかったため、抽出されなかった。表 2-(d) は、従来手法では抽出できていた部分が、抽出できなかった例である。これは、流言情報に略語が含まれていたのに対して、Web テキストでは正式名称で記述されていたことと、有効単語同士の間には許容数を超えた別の単語が含まれていたためである。表 2-(e) は、従来手法では誤って抽出されていたが、提案手法では抽出されなかった例である。これは、従来手法では単語の名詞のみが全て含まれているかを見ていたため「地上」「波」「内容」が含まれる文章を抽出していたが、提案手法では名詞以外も参照し、誤判定を防いだためである。

また、表 3 に、提案手法の適用結果のうち (f) の例を示す。これは、流言情報に含まれる「大学」という名詞が、Web テキストの方には含まれなかったため、従来手法では抽出されなかった例である。提案手法では、単語の不足も該当しない単語が含まれる数も許容数以内だったため、問題なく抽出された。

以上のことから、提案手法は従来手法に比べて抽出精度を上げられる可能性があるが、単語の許容数を調整し、接頭辞や接尾辞など、より幅広い表記に対応する必要があることが分かった。

5. おわりに

本研究では、Web 上に存在する流言情報を検出し、ユーザに対して適切に提示するシステムの構築を目指している。本稿では、Web 上に存在する流言情報部分を適切に抽出する手法を検討し、適用した。適用した結果、提案手法は抽

出精度を上げられる可能性があることが分かったが、調整が必要な部分も明らかになった。今後は、より幅広い表記に対応できるように許容値や抽出の設定を追加し、精度向上に向けて詳細な分析を行う。

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 (2019).
- [2] 西谷智広: I 見聞録: Twitter 研究会, 情報処理学会誌, Vol.51, No.6, pp.719-724 (2010).
- [3] 三浦麻子, 鳥海不二夫, 小森政嗣, 松村真宏, 平石界: ソーシャルメディアにおける災害情報の伝播と感情: 東日本大震災に際する事例, 人工知能学会論文誌, Vol.31, No.1, p.NFC-A-1-9 (2016).
- [4] 萩上チキ: 検証 東日本大震災 そのときソーシャルメディアは何を伝えたか?, ディスカヴァー・トゥエンティワン (2011).
- [5] 柿本大輔, 宮部真衣, 荒牧英治, 吉野孝: 流言拡散防止のための情報確認行動促進システムの構築, ヒューマンインタフェース学会論文誌, Vol.20, No.1, pp.1-11 (2018).
- [6] Takahashi, T. and Igata, N.: *Rumor detection on twitter*, Soft Computing and Intelligent Systems(SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on, IEEE, pp.452-457 (2012).
- [7] Castillo, C., Mendoza, M. and Poblete, B.: *Information Credibility on Twitter*, Proceedings of the 20th international conference on World Wide Web, pp.675-684 (2011).
- [8] Levenshtein, V. I.: Binary codes capable of correcting deletions, insertions and reversals, Soviet Physics Doklady, Vol.10, No.8, pp.707-710 (1966).
- [9] 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).

⁴⁶完全一致した場合必ず抽出が成功するので、検証の必要がない