

RumorFinder：情報真偽確認促進システムの 一般公開とその利用評価

柿本 大輔¹ 平林 (宮部) 真衣^{2,a)} 荒牧 英治³ 吉野 孝^{1,b)}

受付日 2020年4月2日, 採録日 2020年10月6日

概要：インターネットにおける流言拡散防止のためには、情報閲覧時のユーザに対し、真偽確認を促進することで、不用意な情報拡散を防ぐことが重要である。我々はこれまでに、ユーザに対し、流言の存在に関する気づきを提供し、真偽確認を促進するシステム「RumorFinder」の構築を行ってきた。本研究では、効果的な真偽確認促進および役に立つ情報の提供について検討するため、システムの一般公開を行ったうえで、利用データの収集および分析を行った。分析の結果、以下の点を明らかにした。(1) ユーザは吹き出し表示による真偽確認を行っており、本システムはユーザに対し、真偽確認を促進することができる可能性がある。(2) 情報の真偽に関心を持ちうるユーザについては、現状のシステム仕様であっても長期利用による注意の怠りは生じず、流言の内容などに応じてシステムが活用される可能性がある。(3) 一定期間後にシステムを利用しなくなった可能性があるユーザが見られており、多数のページにおいて流言検出・強調表示がされたことが、煩わしさや不快感を与え、その一因となった可能性がある。(4) 真偽確認行動に遷移した流言とそうでないものを比較した結果、何らかの真偽確認行動に遷移した流言は、ある程度内容の把握が可能であるが、不足情報は一部にとどまる。吹き出し表示をしても何らかの真偽確認行動に移らない流言は、複数の事項が不明確であり、内容の特定に複数回の検索が必要になる可能性がある。

キーワード：流言, 気づき, 真偽確認支援, 拡散防止, インタフェース

RumorFinder: Release and Its Evaluation of Information-truth Confirmation Prompting System

DAISUKE KAKIMOTO¹ MAI MIYABE HIRABAYASHI^{2,a)} EIJI ARAMAKI³ TAKASHI YOSHINO^{1,b)}

Received: April 2, 2020, Accepted: October 6, 2020

Abstract: In order to prevent the spread of groundless rumor on the Internet, it is important to prevent the inadvertent spread of information by encouraging users to confirm the authenticity of the information. We have developed a system called “RumorFinder” to provide users with awareness of the existence of groundless rumors and to prompt them confirm their authenticity. In this study, we collected and analyzed the usage data of the system after releasing it to the public in order to study the effective promotion of verification and provision of useful information. The results of the analysis showed the following points: (1) The user performs a veracity check with a callout display, and this system may be able to facilitate the user’s veracity check. (2) Users who are interested in the authenticity of the information can use the system without neglecting to pay attention to rumors due to long-term use. (3) We observed some users who might have stopped using the system after a period of time. The reason for this is that the detection and highlighting of rumors on many pages may have caused annoyance and discomfort. (4) The features of the rumors that users confirmed were missing some information about its content. The features of the rumors that users not confirmed were missing much information about its content.

Keywords: rumor, awareness, confirmation support, anti-diffusion, interface

1. はじめに

近年、ソーシャルネットワーキングサービス (SNS) の普及により、ユーザの手軽な情報取得・発信が可能になっている。代表的な SNS の 1 つである Twitter^{*1}においては、発言および議論される話題の約 85% が実社会における話題に関連するという報告があり [1], Twitter は、従来のテレビや新聞などに代わる新たなニュースメディアとしての役割を持っていると考えられる。

しかし、Twitter などの SNS においては、情報発信者の匿名性や情報の多様性から、誰もが情報の信頼性を正しく判断することができるとは限らず、流言^{*2}が伝播されるという問題も起こっている [2]。流言は、ユーザ間の適切な情報共有を阻害し、流言の伝播によって深刻な問題を引き起こされる可能性がある。特に、2011 年 3 月に発生した東日本大震災や 2019 年 11 月に発生が確認された 2019 新型コロナウイルス感染症 (COVID-19) のパンデミックなど、不安定な社会情勢においては多くの人が不安をかかえた状況となり、流言が発生・拡散しやすい。そのような状況下では、伝搬した流言をきっかけに、食料品や生活用品の買い占め、健康被害につながりうる誤った情報に基づく行動が起こるなど、生活に影響を与える場合がある。そのため、個々人の情報に対する危機意識を高め、流言の拡散を防止する仕組みが必要である。

流言が拡散する要因は様々であると考えられるが、悪意を持ったユーザによる情報の発信が、流言拡散の要因になるとは限らない。流言とは、曖昧な状況に巻き込まれた人々が知識や情報を寄せ集め、その状況について意味のある解釈を行おうとするコミュニケーションであるという考察 [3] もあり、有用な情報を他者と共有しようとする行動により、善意で拡散される場合もあると考えられる。つまり、流言の拡散は、ユーザが情報を発信する際、情報が流言であることを認識していないことや、情報の真偽を確認するという行動が欠如していることなどに起因する可能性がある。特に、インターネットにおける流言の拡散は、情報の取得・発信が手軽である SNS の利用者増加により、これらの点が要因となっている可能性がある。

そこで、これまでに我々は、Web ページ閲覧時のユーザに対し、閲覧ページ内に存在する流言への気づきを提供し、真偽確認を促進するシステム「RumorFinder (ルーマーファインダー)」の開発を行ってきた [4]。本システム

は、ユーザが閲覧している Web ページ内に流言が含まれる可能性がある場合、該当箇所の強調表示により、流言に関する気づきを提供する。さらに、強調箇所をマウスオーバーした際に、吹き出しを表示させ、関連する流言についての情報を提供する。また、評価実験の結果、強調表示や吹き出し表示は通常の Web ページ閲覧行動を阻害しないこと、また流言に関する気づきを提供可能であることを示した。一方、インターネットにおける流言拡散防止のためには、ユーザに対し、閲覧している情報が流言であることを認識させるだけでなく、その情報に関する真偽確認を促進することが重要である。また、利用者から継続的かつ積極的に利用され、利用者に対して情報 (流言) への注意喚起が定常的にできることが望ましい。しかし、システムを利用し始めた段階では、目新しさなどから利用頻度が高くなると想定されるが、長期間利用しているうちに、システムへの慣れや飽きが生じ、提供される流言情報への注意の怠りが起こる可能性があると考えられる。

そこで本論文では、本システムを一般公開したうえで、一定期間の利用ログを蓄積し、利用者の利用動向を分析することにより、本システムが与える効果や継続利用への要件について議論する。

2. 関連研究

Twitter では情報発信の手軽さゆえに、流言が多数発信・拡散されている。そのため、Twitter における流言拡散防止に関する研究はさかんである。Qazvinian らは、Twitter における情報のみ (ツイートの内容、ユーザ情報、Twitter 独自の情報) を特徴量とした、バイズ分類器による情報の信頼性自動推定手法を提案した [5]。Gupta らは、ランク学習手法の 1 つである RankSVM と、情報検索における精度改善手法の 1 つである適合性フィードバック手法を用い、ツイートの信頼性をランキングする手法を提案した [6]。Lim らは、ツイート内に含まれる語をクエリとした Web 検索を行い、検索結果の Web ページ特定のキーワード (fake, doubt など) が含まれているかどうかをもとに、ツイートの信頼性を自動推定するシステムを提案した [7]。

また、Mendoza ら [8] の、2010 年のチリ地震における、Twitter ユーザの行動についての分析によると、誤情報に関連するツイートのうち約 50% が、その内容の否定に関するツイートであったと述べられており、インターネットおよび Twitter から、流言に関する情報の自動検出・収集ができれば、流言の早期訂正および拡散防止に役立つ可能性がある。そこで我々はこれまでに、あるツイートを訂正する内容のツイート (訂正ツイート) を用いることで、間接

¹ 和歌山大学
Wakayama University, Wakayama 640-8510, Japan

² 東京大学
The University of Tokyo, Hongo 113-8654, Japan

³ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Nara 630-0192, Japan

a) mai.miyabe@gmail.com

b) yoshino@wakayama-u.ac.jp

^{*1} <http://twitter.com>

^{*2} 本研究では、十分な根拠がなく、その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 (悪意を持った捏造か自然発生か) は問わないものとする。

的に流言に関する情報を自動収集するシステム「流言情報クラウド」の構築を行ってきた [9].

情報の信頼性の自動推定を試みる研究は多数行われており、インターネットにおける膨大な情報の信頼性を自動で推定することができれば、流言の拡散防止の一助となる可能性がある。しかし、インターネットにおける流言拡散防止のためには、信頼性の自動推定手法の検討のみではなく、推定結果や流言として検出された情報を適切にユーザに提供し、ユーザ自身の情報の真偽確認行動を促すような仕組みが必要である。そこで我々は、あらかじめ検出・収集された流言情報を用いることで、情報閲覧時のユーザに対し、流言の存在に関する気づきを与え、情報の真偽確認行動を促進するシステムの構築を行う。

閲覧中のコンテンツになんらかの情報を付与することで、ユーザの真偽確認行動を促進するインタラクションを提案した研究はいくつか行われている。Rob Ennals ら [10] は、DisputeFinder を開発した。DisputeFinder は、インターネット上で議論になっている内容を閲覧中の Web ページに含まれている場合に、議論になっている事項について、ハイライトして、賛成意見と反対意見を閲覧できる。Schwarz ら [11] は、Web ページ全体の信憑性の判断支援のための可視化手法を提案した。PageRank、閲覧者の専門家の割合、閲覧数の履歴などを Web ページと一緒に提示している。Saito ら [12] は、Web ページに存在する言葉を濁した表現をハイライトすることで、注意深いウェブ探索を促進する手法を提案している。特に、DisputeFinder と本研究は、真偽確認行動の促進を目指している点、ハイライトなどでマークする点、関連する情報を閲覧できる点などが類似している。これらの研究と本研究との違いは、Twitter 上の流言を対象としており、対象とするコンテンツの性質が異なる点である。また、本論文では特に、ユーザの真偽確認行動を促進するインタラクションの支援に対する、ユーザの長期的な行動に焦点を当てている点が大きく異なる。

3. RumorFinder

これまでに我々は、Web ページ閲覧時のユーザに対して流言に関する気づきを提供し、情報の真偽確認を促進するシステム「RumorFinder (ルーマーファインダー)」の構築を行ってきた [4]^{*3}。本システムの動作画面例を図 1 に示す。本システムは、ユーザが閲覧中の Web ページにおいて、流言である可能性のある情報が含まれる場合、テキストの強調表示を行う。なお、流言かどうかの判定は、間接的に流言に関する情報を自動収集するシステム「流言情報クラウド」[9] で収集された流言情報を用いて、単語単位の文字列マッチングで判定している。また、訂正情報や訂正数の情報も、「流言情報クラウド」で収集された結果を利

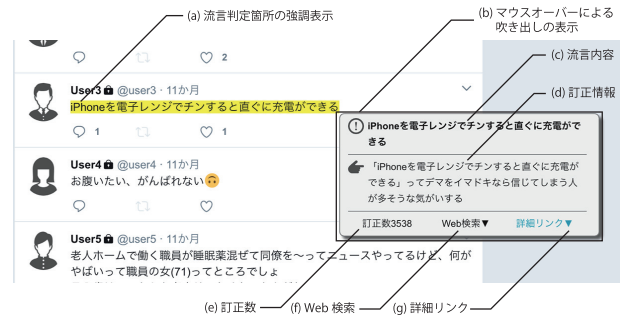


図 1 システムの動作画面例
Fig. 1 Example screenshot of the system.

用して、提示している^{*4}。

本システムは、様々な Web ページにおいて動作するシステムであり、Web ページによっては、テキストの下線や着色・太字による強調がページ内のハイパーリンク部分や、本来の強調部分と混同してしまう可能性がある。そこで、システムの利用がユーザのページ閲覧の妨げにならないよう、本システムによるテキストの強調表示については、背景色を変更する^{*5}という形をとった。

Web ページ上に新たなオブジェクトを発生させ、ユーザに対し気づきを提供する手法として、通知ダイアログおよび吹き出しの表示などがあるが、吹き出しの表示は、通知ダイアログの表示などに比べ、利用者の作業を阻害する可能性が低い [13]。そのため、本システムにおいては、吹き出しの表示により、流言に関する情報の提示および真偽確認の促進を行う。また、吹き出しについては、ユーザのページ閲覧の妨げにならないよう、強調表示箇所のマウスオーバーにより、ユーザが任意で表示させられるようにした。なお、吹き出しには、流言情報、訂正情報（流言情報に対する訂正ツイート）、Web 検索^{*6}、詳細リンク^{*7}が含まれる。

また、テキストを強調表示する際、テキスト量の多い Web ページや、Twitter のタイムラインのような環境でも、ユーザが流言に関する情報を見落とさないよう、トースト通知機能の追加を行った。トースト通知とは、画面端に数秒表示される通知であり、ユーザに操作を要求する通知ではないため、操作を妨げることがないという特徴がある。トースト通知機能の動作例を図 2 に示す。さらに、トースト通知と同時に、検出した流言の種類数を Web ブラウ

^{*4} 流言情報クラウドは、毎日、流言情報を更新している。評価実験期間最終日 (2018 年 12 月 29 日時点) の蓄積流言数 (ユニーク) は、14,827 件である。ただし、用いるデータは前日までの 1 週間に検出された流言であり、毎日、変化する。2018 年 7 月 26 日 (8 月 2 日の 1 週間前) から 12 月 29 日までの流言数 (ユニーク) は、1,665 件であった。

^{*5} 閲覧する Web ページによっては、背景色の変更がユーザのページ閲覧を妨げる可能性がある。そこで、背景色の変更においては、ユーザが任意で色を選択できるようにした。

^{*6} Web 検索を押すと、流言情報の抽出語句を Google 検索の検索語として自動入力し、その結果のページを表示する。

^{*7} 詳細リンクとは、流言情報に対するこれまでの訂正数および直近 1 週間の訂正数の推移を表すグラフを確認できるページへのリンクである。

^{*3} RumorFinder の紹介ページ https://www2.yoslab.net/~kakimoto/addon_intro/index.html

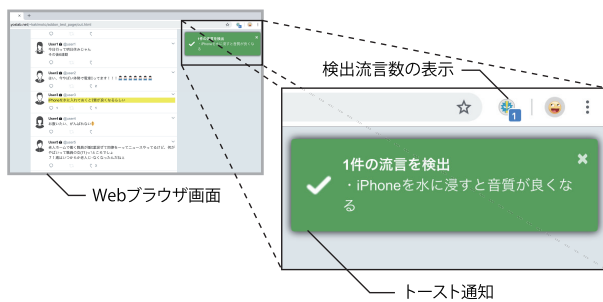


図 2 トースト通知機能の動作例

Fig. 2 Example screenshot of a toast notification.

ザ右上のアイコン (図 2) に表示させるようにした。

なお、本システムは、Google 社^{*8}が提供する Web ブラウザである、Google Chrome^{*9}のアドオンとして実装した。現在、様々なブラウザが普及しているが、Google Chrome (バージョン: 70.0) の全世界シェアは、2018 年 11 月時点で 48.8%であり、1 位となっている^{*10}。また、アドオンなど、機能の拡張がしやすいという特徴もある。そこで、今回は Google Chrome のアドオンとしてシステムの構築を行った。

4. システムの一般公開と利用ログの分析

これまでの研究 [4] で行った評価実験においては、強調表示や吹き出し表示は通常の Web ページ閲覧行動を阻害しないこと、また流言に関する気づきを提供可能であることを示した。しかし、この評価実験は、実験用に用意した Web ページを用いた実験となっており、日常的なユーザの利用行動に基づく評価は実施できていない。また、一時的な利用のみであり、継続利用がなされるかどうかといった観点での評価は行っていない。

そこで本論文では、システムを一般公開し、利用ログの収集を行った。収集した利用ログの分析により、日常的な利用における行動における効果や、継続利用に関する効果について明らかにし、情報の真偽確認促進において効果的な仕組みおよびインタフェースを検討する。

4.1 検証仮説

本システムは、利用者から継続的かつ積極的に利用され、利用者に対して情報 (流言) への注意喚起が定常的にできることが望ましい。しかし、前述したように、システムを利用し始めた段階では、目新しさなどから利用頻度やシステムへの着目度合いが高くなると想定されるが、長期間利用しているうちに、システムへの慣れや飽きが生じ、提供される流言情報への注意の怠りが起こる可能性がある。

システムで流言情報が検出・提示された際、利用者がと

*8 <http://www.google.co.jp/>

*9 <http://www.google.co.jp/chrome/browser/desktop/index.html>

*10 https://webrage.jp/techblog/pc_browser_share

りうる行動としては、まず、(1) ページ上の強調表示部分にマウスオーバーし、吹き出しに表示される情報を確認し、その後、(2) 吹き出し上に表示された情報から、詳細リンクや Web 検索リンクへの遷移を行う^{*11}、という流れになる。しかし、利用者が必ずしも (1), (2) の流れをたどるとは限らず、(1) のみを行い、(2) は行わない利用者や、(1) さえ行わない利用者が存在する可能性もある。特に (2) については、そのとき見ていた Web ページから離脱することとなるため、行動の敷居が高い可能性がある。また、表示される流言の内容は、情報の真偽確認行動を起こすかどうかに影響を与える可能性がある。なお、本システムでは強調表示・吹き出し表示以外に、トースト通知機能も備えているが、トースト通知については利用ログの取得ができないため、分析の対象外としている。

そこで本研究では、一般公開後の利用ログを対象として、以下の点について検証する。

検証項目 1: 利用者は、流言が検出された場合に、吹き出しの確認を行うか?

検証項目 2: 利用者は、真偽確認行動 (詳細リンクや Web 検索リンクへの遷移) に至るか?

検証項目 3: システム利用時に検出された流言にはどのようなものがあるか?

4.2 利用状況

本研究で構築したシステムは、Chrome ウェブストアを介して、2018 年 7 月 4 日より一般公開している。システムの一般公開後、本システムがどのように利用されるのかを分析するために、25 人の大学生および大学院生にシステムを利用してもらうよう呼びかけを行った。ただし、利用に際し、利用ログの送信を許可してもらうよう伝え、その他の指示は与えず、自由にってもらうこととしている^{*12}。

利用ログについては、閲覧したページ、検出された流言、強調表示箇所のマウスオーバー、詳細リンクへの遷移および Web 検索リンクへの遷移を取得することとし、2018 年 8 月 2 日より継続して取得している。本システムのユーザ数の推移を図 3 に示す。2018 年 12 月 29 日時点でのシステムのユーザ数は 54 人であり、期間中に利用ログの送信を 1 度でも行ったユーザ数は 71 人であった^{*13}。

*11 本研究では、この行動をとったことをもって、情報の真偽確認行動を行ったと見なすこととする。

*12 本評価実験ではユーザに対して、システムの使い方について説明していない。ただし、Chrome の拡張機能のインストールを促すため、機能の説明用の動画や Web ページなどは見られるため、ある程度の使い方は理解していると考えられる。

*13 利用ログは端末ごとに取得しているため、同一ユーザによる複数端末におけるシステムの利用および複数ユーザによる同一端末におけるシステムの利用が考えられる。本研究では、利用ログを取得した端末 1 台を 1 ユーザと見なす。

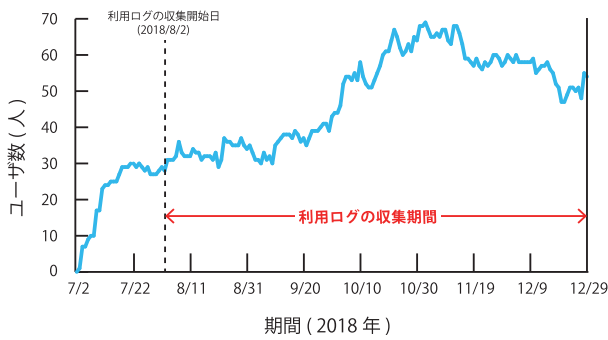


図 3 システムのユーザー数の推移

Fig. 3 Number of users in the system.

4.3 利用ログの分析概要

本研究では、2018年8月2日～2018年12月29日の150日間の利用ログを分析の対象とする。

本システムは、1度インストールされると、ユーザーに対して流言情報の検出が自動通知される仕組みをとっており、システムからの通知に対し、確認行動をとるかどうかはユーザーの自由である。ただし、本システムが流言の拡散防止に貢献するためには、インストール直後に起こりがちな一過性の利用ではなく、積極的な確認行動をユーザーが継続できるようにすることが求められる。

継続利用を実現するためには、(1)一過性の利用にとどまった原因を明らかにすること、(2)継続利用につながったユーザーの特徴を明らかにすること、が有用な知見となると考えられる。しかし、本システムは実験ではなく、あくまでも一般公開という形式をとっており、システムを誰でも自由にダウンロードできるようになっている。そのため、一過性の利用のみで、その後利用が見られないユーザーに対して、なぜ利用しないのかなどを直接調査すること(調査(1))が難しい。一方、継続した利用が見られるユーザーについては、その利用状況を確認することで、継続利用につながる特徴を確認(調査(2))できる可能性がある。そこで、本研究では、利用率の低い一過性利用ユーザーは分析対象外とする。

利用者のログを確認したところ、利用ログの送信を行った71人のうち、28人は、閲覧ページ数の合計が10件未満であった。そこで、本研究では、まずこれらの28人を分析対象から除外する。一定の継続利用をしていると考えられる残りの43人を分析対象とする。

次に、分析対象外のユーザーを除外した43人の利用ログを確認したところ、150日間の閲覧ページ数の合計は475,872件であった。このうち、1人の閲覧ページの合計が、258,975件となっており、そのうち約97.35%(252,120件)が同一のWebページへのアクセスであった^{*14}。ユーザーはこのWebページの内容を、ページ読み込みごとに閲覧していたわけ

^{*14} このWebページは、一定時間ごとにページ読み込みを繰り返す仕組みであり、ページ読み込みごとにページ閲覧に関するログの送信が行われる。

ではないと考えられる。そのため、分析においては、このWebページ^{*15}に関するログを除外することとした。除外後の対象閲覧ページ数の合計は223,712件である。

また、分析対象外ユーザーを除外した43人それぞれの閲覧ページ数を確認したところ、閲覧ページ数上位20人の合計が194,114件であり、全体の86.78%を占めていた。つまり、これら20人のユーザーは、本システムのヘビーユーザーであると考えられる。そこで、本研究では、20人のヘビーユーザーを分析対象とし、その利用状況を分析することで、継続利用につながる特徴があるかどうかを検証する。なお、本システムは、開始当初は、大学生および大学院生25人に利用を呼びかけたが、一般公開しているため、参加者の詳細については不明である。

5. 分析結果

5.1 期間全体の利用傾向

閲覧ページ数が多かった上位20人のユーザーのシステムの利用状況(閲覧ページ数、流言検出ページ数、検出割合および吹き出し表示回数)を、表1(ユーザーID1~10)および表2(ユーザーID11~20)に示す。なお、表1、表2では、システムの概要を紹介するWebページおよび流言情報クラウドのWebページにおける動作ログは除外している。

表1、表2より、(F)流言検出割合については、ユーザーID1~20において、10%未満となり、平均は2.77%であった。この結果より、ユーザーがWebページ閲覧時、本システムにより流言の検出が通知される頻度は低いことが分かった。この理由としては、(1)そもそも、災害時などを除いた平常時においては、ユーザーが流言情報を目にする機会がさほど多くないこと、(2)システムが流言として検出すべき情報を、正しく検出できていない可能性があることが考えられる。

(G)吹き出し表示割合については、ユーザーID1~20において、平均16.23%であり、流言が検出されたとき、吹き出しの表示は一定の割合で行われていると考えられる。しかし、吹き出しの表示割合が10%未満のユーザーが6人おり、そのうち1人のユーザーは吹き出し表示をまったく行わなかったことが分かった。

以上より、期間全体の利用ログに関する分析結果を総括すると、まず、一定の利用を行っているヘビーユーザーについては、システムが強調表示を行うと、吹き出し表示を行っていた。これまでに実施した評価実験[4]などの利用者だけでなく、一般公開によりシステムをインストールしたユーザーに対しても、吹き出し上の流言情報への気づきを与えられる可能性があることが分かった。ただし、吹き出しの表示割合が10%未満のユーザーや、吹き出し表示をまったく行わないユーザーもいることから、たとえば、トースト

^{*15} なお、このWebページにおいて、流言の検出は1度も行われていなかった。

表 1 システムの利用状況 (ユーザ ID1~10)
Table 1 Usage status of the system (user id 1 to 10).

	ユーザ ID									
	1	2	3	4	5	6	7	8	9	10
(A) 閲覧ページ数 (件)	22,824	17,186	16,342	16,158	13,083	11,671	10,112	9,896	8,435	8,425
(B) 流言検出ページ数 (件)	186	182	1,484	154	146	124	160	672	174	200
(C) 吹き出し表示回数 (件)	41	53	111	16	11	12	42	22	0	24
(D) 詳細リンク遷移回数 (件)	4	5	0	1	0	0	2	0	0	0
(E) Web 検索リンク遷移回数 (件)	3	9	0	0	0	0	1	0	0	0
(F) 流言検出割合 (%)	0.81	1.06	9.08	0.95	1.12	1.06	1.58	6.79	2.06	2.37
(G) 吹き出し表示割合 (%)	22.04	29.12	7.48	10.39	7.53	9.68	26.25	3.27	0.00	12.00
(H) 詳細リンク遷移割合 (%)	9.76	9.43	0.00	6.25	0.00	0.00	4.76	0.00	-	0.00
(I) Web 検索リンク遷移割合 (%)	7.32	16.98	0.00	0.00	0.00	0.00	2.38	0.00	-	0.00

※「吹き出し表示回数」は、ハイライト箇所を 2 回マウスオーバーした場合には、2 回とカウントされる。

表 2 システムの利用状況 (ユーザ ID11~20)
Table 2 Usage status of the system (user id 11 to 20).

	ユーザ ID									
	11	12	13	14	15	16	17	18	19	20
(A) 閲覧ページ数 (件)	8,137	7,929	7,307	6,855	6,637	6,024	5,229	4,710	3,678	3,476
(B) 流言検出ページ数 (件)	227	203	85	230	61	520	0	60	189	92
(C) 吹き出し表示回数 (件)	42	65	10	54	29	56	0	7	37	5
(D) 詳細リンク遷移回数 (件)	1	2	0	0	1	1	0	0	0	0
(E) Web 検索リンク遷移回数 (件)	0	0	0	0	1	1	0	0	0	0
(F) 流言検出割合 (%)	2.79	2.56	1.16	3.36	0.92	8.63	0.00	1.27	5.14	2.65
(G) 吹き出し表示割合 (%)	18.50	32.02	11.76	23.48	47.54	10.77	-	11.67	19.58	5.43
(H) 詳細リンク遷移割合 (%)	2.38	3.08	0.00	0.00	3.45	1.79	-	0.00	0.00	0.00
(I) Web 検索リンク遷移割合 (%)	0.00	0.00	0.00	0.00	3.45	1.79	-	0.00	0.00	0.00

※「吹き出し表示回数」は、ハイライト箇所を 2 回マウスオーバーした場合には、2 回とカウントされる。

- (A) 閲覧ページ数において、同一の Web ページの重複に関する考慮は行っていない。
- (B) 流言検出ページ数は、1 件以上の流言情報が検出された Web ページの数である。
- (F) 流言検出割合は、((B) 流言検出ページ数/(A) 閲覧ページ数)*100 の値である。
- (G) 吹き出し表示割合は、((C) 吹き出し表示回数/(B) 流言検出ページ数)*100 の値である。
- (H) 詳細リンク遷移割合は、((D) 詳細リンク遷移回数/(C) 吹き出し表示回数)*100 の値である。
- (I) Web 検索リンク遷移割合は、((E) Web 検索リンク遷移回数/(C) 吹き出し表示回数)*100 の値である。

通知機能およびカスタマイズ機能のような、テキストの強調表示および吹き出し表示以外の情報提供方法を追加・変更することも検討する余地があると考えられる。

次に、ヘビーユーザであっても、情報の真偽確認行動(詳細リンクおよび Web 検索リンクへの遷移)を行うユーザは少ない可能性があることが分かった。一般公開に基づく利用ログ分析を行っており、利用ユーザを特定して直接アンケートをとることが難しいことから、この原因について詳細に分析することは難しいが、確認行動に至らない原因としては、(1) そもそも表示された流言内容に興味が無い、(2) 吹き出し上で表示している内容や項目に問題があり、確認行動を起こすに至らない、(3) 吹き出し上の情報提示方法(インタフェース)に問題がある、といったことが考えられる。そのため、システム側で対処可能な原因(2)、(3)については、今後再検討する必要がある。

5.2 検出された流言

2018 年 8 月 2 日~2018 年 12 月 29 日の 150 日間に、システムが検出した流言情報(検出数上位 15 件)を表 3 に示す。

表 3 より、システムは同期間中に、流言情報(5)「携帯で 1, 1, 0 を押して電話を掛ける」および流言情報(12)「#BTS がユニセフに原爆看板を持ち込み」のような、実際にインターネット上で拡散された流言情報を検出した。しかしその一方で、流言情報(1)「あ、これ 1 人抜けたよね。他の 4 人にいじめられたんだってさ。」のような、議論の対象が不明瞭なものや、流言情報(4)「この人は虫としゃべれる」のような、そもそも真偽を確認する必要のないような流言情報も検出された。

5.1 節で述べたように、詳細リンクおよび Web 検索リンクのような他ページへの遷移を行い、情報の真偽確認行動を行うユーザ自体がそもそも少ない可能性はあるものの、

表 3 検出された流言 (検出数上位 15 件)
Table 3 Detected groundless rumors. (Top 15).

ID	流言情報	検出数
(1)	あ、これ 1 人抜けたよね。他の 4 人にいじめられたんだってさ。	2,121
(2)	個人情報が抜き取られた	1,902
(3)	その情報どこから出たん!?	1,303
(4)	この人は虫としゃべれる	83
(5)	携帯で 1、1、0 を押して電話を掛ける	83
(6)	NYTIMES は韓国を批判している	77
(7)	中国人が.....	70
(8)	できてすらいもない本を手売りしてたらしい	53
(9)	学校では決して教えてくれない情報	48
(10)	沖縄は基地で食っている	45
(11)	この花火大会見たらわかれまーす!	45
(12)	#BTS がユニセフに原爆看板を持ち込み	36
(13)	お祝いで店のものタダだって!	27
(14)	24 日はクリスマスイブ	23
(15)	アプリメーカー 乗っ取り	22

表 3 の (1) や (4) のようなユーザが吹き出し表示以外の確認行動を行う必要がない流言情報もシステム上で検出されていた。

そのため、流言情報の重要度や拡散時の深刻度などについても検討し、それらの指標に基づき、検出された流言情報に応じた情報確認行動の促進方法を変更するなどの改善が必要である可能性がある。

6. 考察

5.1 節で述べたように、期間全体の利用ログの分析の結果、分析対象としたヘビーユーザ 20 人の吹き出し表示割合は平均 16.23%であった。一方、詳細リンク遷移割合および Web 検索リンク遷移割合はそれぞれ平均 2.27%、平均 1.77%と低く、まったく使用していなかったユーザも存在していた。今回分析対象とした利用ログは一般公開による利用ログであり、システムを使用しなかったユーザを特定して、なぜ使用しなかったのかなどを事後調査することは難しい。

そこで本章では、本研究で分析対象としたヘビーユーザ 20 人のうち、真偽確認行動 (詳細リンクまたは Web 検索リンクの利用) が見られたユーザ (以降、「確認行動ユーザ」と表記する) 8 人 (表 1, 表 2 におけるユーザ ID1, 2, 4, 7, 11, 12, 15, 16) の利用傾向を考察し、何らかの確認行動をとったユーザの特徴について考察する。

6.1 時間経過にともなう変化の有無

確認行動ユーザ 8 人に関する、流言検出ページ数と吹き出し表示率 (吹き出し表示ページ数/流言検出ページ数) の 150 日間の推移を図 4 に示す。なお、各グラフにおいて、棒グラフが流言検出ページ数 (左軸)、マーカー付き折れ線グ

表 4 システムの利用開始日と最終利用日
Table 4 Start and end date of system use.

ユーザ ID	(a) 利用開始日	(b) 最終利用日	利用日数 (日) (b) - (a)
1	8 月 2 日	12 月 12 日	133
2	8 月 2 日	11 月 6 日	97
4	8 月 2 日	12 月 27 日	148
7	8 月 2 日	12 月 28 日	149
11	11 月 9 日	12 月 26 日	48
12	8 月 3 日	12 月 29 日	149
15	8 月 2 日	12 月 28 日	149
16	11 月 6 日	12 月 14 日	39

ラフが吹き出し表示率 (右軸) となっている。

図 4 の各ユーザの棒グラフを見ると、どのユーザについても、流言検出がされない日は多い。ここでは、閲覧ページ中に流言が検出された際、吹き出し表示をしたのかを表す吹き出し表示率について、時間経過とともに変化があったのかを確認する。もし、システムを利用し始めてから時間が経つにともない、慣れや飽きなどにより真偽確認を行わなくなった場合、折れ線グラフは右下がりの傾向を示すと考えられる。図 4 において吹き出し表示率を表すマーカー付き折れ線グラフの推移を見ると、いずれのユーザも、グラフの右側に向かっていっても、吹き出し表示率が 0 でない日があることが確認できる。すなわち、今回分析対象とした 8 人の、150 日間の利用ログの中では、時間経過にともなう真偽確認行動の割合が低下する傾向は見られなかった。すなわち、情報の真偽に関心を持ちうるユーザについては、現状のシステム仕様であっても長期利用による注意の怠りは生じず、流言の内容などに応じてシステムが活用される可能性があると考えられる。

6.2 利用継続の可否

図 4 を見ると、一部のユーザは、一定期間ののちグラフの途切れ (ログデータの無い日) が見られる。そこで、確認行動ユーザ 8 人について、利用ログが取得され始めた日 (利用開始日) と、最後にログが確認できた日 (最終利用日) を確認した。表 4 に利用開始日と最終利用日を示す。

今回分析対象としたログ期間は 12 月 29 日までである。12 月 29 日に利用が確認できたのはユーザ 12 のみであるが、ユーザ 4, 7, 11, 15 については、いずれも 12 月 26~28 日にシステムの利用ログが確認できている^{*16}。

一方、ユーザ 1, 2, 16 については、15 日以上利用ログが確認できない日が続いており、これらのユーザはシステムを利用しなくなった可能性がある。システムを利用しな

^{*16} 図 4 を見ると、ユーザ 4, 7, 11, 15 のグラフは 12 月 26~28 日までのデータがなく、それ以前での途切れが見られる。これは、図 4 に示しているのが流言の検出および吹き出し表示率のみであるためであり、流言の検出されないページ閲覧ログがユーザ 4, 7, 11, 15 については表 4 に示した日まで確認できている。

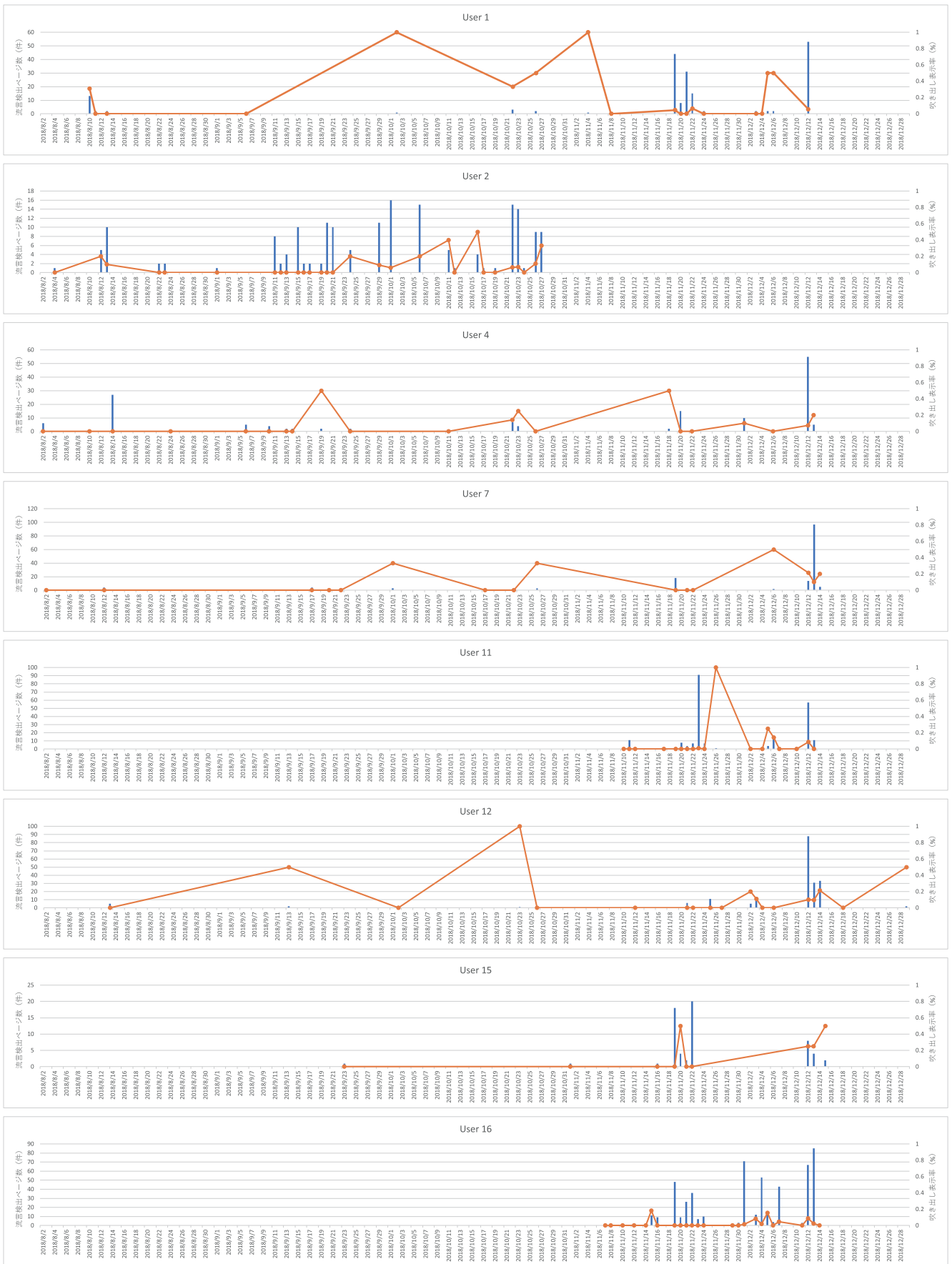


図 4 流言検出ページ数と吹き出し表示率の推移

Fig. 4 Transition in the number of pages with a groundless rumor and the rate of callout displays.

くなった理由については、システムに飽きた、不具合が生じた、ユーザの需要に対応しきれなかったなど、システム自体が問題をかかえている可能性はあるものの、使用するブラウザ自体を変えた、使用する機器を変えた、長期間使用しない状況になったなど、システム以外の要因で利用しなくなった可能性など、様々な原因が考えられる。しかし、図4のユーザ1, 2, 16のグラフを見ると、グラフが途切れる日付近で、流言検出ページ数(棒グラフ)が比較的多くなっている傾向がある。多数のページで流言が検出され、強調表示されたことにより、煩わしさや不快感を与えた可能性も考えられる。ただし、一定の利用が見られたユーザ自体が少なく、またその中でも利用しなくなった可能性があるユーザも少ないため、今後継続して一般公開中の利用ログを収集し、分析を進める必要がある。

6.3 真偽確認行動がとられた流言

本節では、確認行動ユーザが、何らかの確認行動(詳細リンクやWeb検索リンクへの遷移)を行った流言の内容には特徴があるのかどうかを考察する。確認行動ユーザが吹き出しを表示させた流言20件の内容を表5に示す。

表5においては、吹き出し表示がされた流言を分類A, B, C, Dの4種類に分類している。分類Aは、詳細リンク遷移を行ったユーザと、Web検索リンク遷移を行ったユーザのどちらも存在する流言である。分類Bは、詳細リンク遷移を行ったユーザは存在するが、Web検索リンク遷移を行ったユーザは存在しない流言である。分類Cは、詳細リンク遷移を行ったユーザは存在しないが、Web検索リンク遷移を行ったユーザは存在する流言である。分類Dは、吹き出し表示はされたが、詳細リンク遷移、Web検索リンク遷移は行われなかった流言である。

表5に示した、システムで検出され、確認行動ユーザが吹き出し表示した流言の内容は多岐にわたっている。流言の内容に関して真偽確認行動をとるかどうかは、ユーザがその内容に対して興味があるかといったことが影響する可能性が高いが、今回、一般公開する際には、ユーザ登録などによる興味情報の登録などはせずに利用できるようにしており、興味度合の影響を分析することが難しい。そこで、ここでは、それ以外に真偽確認行動の有無に影響しうる特徴があるかどうかを考察する。

表5の分類A~Dを比較すると、分類Aに含まれる8件の流言は、比較的具体的に内容が記載されており、理解できるが、一部不足している情報がある流言が多く含まれている傾向がある。たとえば、「できてすらいな本を手売りしてたらしい」という流言は、「誰が」手売りをしてたのかという情報が不足している。また、「携帯で1, 1, 0を押して電話を掛ける」という流言は、「どこ」に電話がかかるのかが分からない。「沖縄は基地で食っている」という流言については、特に不足している情報はないように

も判断できるが、その他の7件については、何らかの情報が不足している傾向がある。

分類Bに含まれる3件の流言は、件数が少なく、明確な特徴は見られないが、内容の抽象度が高いもの(「その情報どこから出たん!?!」)や、調べる必要性が低い可能性があるもの(「あと3日で2019年」)が含まれていた。

分類Cに含まれる3件の流言についても、件数が少なく、明確な特徴は見られなかった。

分類Dに含まれる6件の流言は、分類Aと比較すると抽象度が高く、記載内容のみでは理解が難しく、かつ、不足している情報が多い傾向がある。たとえば、「中国人が.....」という流言は、中国人が「何を」「どこで」「どうしたのか」といった詳細な情報が欠落している。また、「学校では決して教えてくれない情報」という流言は、「どの」学校なのか、「どのような」情報なのかといった複数の事項が不明確である。このような情報については、内容を特定して理解するためには、複数回の検索行動などが必要になる可能性があるため、内容に対する興味がない場合、確認行動に遷移しにくい可能性がある。

以上の結果から、今回確認された真偽確認行動の遷移に影響しうる傾向をまとめる。吹き出し表示された後、何らかの真偽確認行動に遷移した流言は、ある程度内容の把握が可能であるが、一部情報が不足している可能性がある傾向が見られた。一方、吹き出し表示をしても何らかの真偽確認行動に移らない流言は、複数の事項が不明確であり、内容の特定に複数回の検索が必要になる可能性がある傾向が見られた。

これらの傾向は、システムからユーザに対して情報提示する際、不足している事項が多い場合は提示の優先度を下げること、通知の煩雑さを低減させるなど、今後のシステム開発における検討指針として用いることができる可能性がある。

ただし、今回分析対象とした利用ログのうち、8人の確認行動ユーザが確認行動をとった流言は20件であり、事例としては少ないため、今後継続して利用ログを収集し、同様の傾向が見られるかどうかを検証していく必要がある。また、前述のように、真偽確認行動をとるかかどうかには、内容に対する興味なども影響しうるため、今後はそれらをふまえた検証方法についても検討していく必要がある。

7. おわりに

インターネットにおける流言拡散は、「ユーザが情報を発信する際、情報が流言であることを認識していないこと」「情報の真偽を確認するという行動が欠如していること」などに起因する可能性がある。そこで、我々はこれまでに、ユーザがWebページを閲覧している際、流言である可能性がある情報が含まれることに対する気づきを提供し、情報の真偽確認行動を促進するシステム「RumorFinder」の

表 5 真偽確認行動がとられた流言

Table 5 Groundless rumors of a confirmation action being taken.

分類	流言内容	流言検出 ユーザ数	吹き出し表示 ユーザ数	詳細リンク 遷移ユーザ数	Web 検索リンク 遷移ユーザ数
A	できてすらいな本を手売りしてたらしい	5	4	3	2
	沖縄は基地で食っている	4	3	2	2
	携帯で 1、1、0 を押して電話を掛ける	5	2	2	1
	あ、これ 1 人抜けたよね。他の 4 人にいじめられたんだってさ。	7	7	1	1
	個人情報が抜き取られた	7	2	1	1
	アプリメーカー 乗っ取り	2	1	1	1
	他国に押し付けられた憲法	2	1	1	1
	分煙もきっちりされていて	1	1	1	1
B	その情報どこから出たん!?	8	5	1	0
	人という字は人と人が支え合って...	2	1	1	0
	あと 3 日で 2019 年	1	1	1	0
C	この人は虫としゃべれる	6	4	0	1
	今はマンガ売れても儲からない	2	1	0	1
	24 日はクリスマスイブ	1	1	0	1
D	中国人が.....	6	2	0	0
	この花火大会見たらわかれまーす!	4	1	0	0
	学校では決して教えてくれない情報	3	1	0	0
	*****に追い詰められて自殺した人がいる	2	1	0	0
	俺***と*****	1	1	0	0
	民事裁判は 2 回目以降は開かれていない。だからこの事件は終わった	1	1	0	0

※人名や地名などの固有名詞および公序良俗に反する内容を含む部分は伏字で提示している。

構築を行ってきた。

本研究では、開発したシステムを一般公開し、その利用ログの分析を行った。まず、期間全体の利用傾向の分析により、以下の点を明らかにした。

- (1) 実験による統制を行っていない一般利用者であっても、強調表示がされた場合、吹き出しを表示して情報を閲覧している利用者が見られており、吹き出し上の流言情報への気づきを与えられる可能性がある。
- (2) 吹き出し上に、詳細リンクおよび Web 検索リンクの提示しても、ユーザはそれらを利用しての情報確認までは行わず、吹き出しの表示のみにとどまる場合が多い。
- (3) 一般公開後、システムが検出し利用者に表示された流言情報の中には、確認行動を行う必要がない流言情報もあったため、流言情報の重要度や拡散時の深刻度などについても検討し、検出された流言情報に応じた情報確認行動の促進方法を変更するなどの工夫が必要である。

また、真偽確認行動（詳細リンクまたは Web 検索リンクの利用）が見られたユーザ 8 人を対象とした、150 日間の利用傾向の分析により、以下の点を明らかにした。

- (1) 分析対象とした 8 人の 150 日間の利用ログでは、時間経過にともなって真偽確認行動の割合が低下する傾向は見られなかった。情報の真偽に関心を持ちうるユーザについては、現状のシステム仕様であっても長期利

用による注意の怠りは生じず、流言の内容などに応じてシステムが活用される可能性がある。

- (2) 8 人中 3 人のユーザは、一定期間後にシステムを利用しなくなった可能性がある。システムを利用しなくなった理由としては、システム自体の問題（システムへの飽き、需要と供給の不一致）やシステム外の問題（使用するブラウザや使用機器の変更、長期間使用しない状況の発生）など様々な原因が考えられるが、利用ログにおいては、利用ログが取得されなくなった日付近で、多数のページにおいて流言検出・強調表示がされた可能性があり、煩わしさや不快感を与えた可能性も考えられるため、検出・提示回数については検討していく必要がある。
 - (3) 分析対象とした 8 人が、吹き出し表示後に真偽確認行動に遷移した流言とそうでないものを比較した結果、何らかの真偽確認行動に遷移した流言は、ある程度内容の把握が可能であるが、一部情報が不足している可能性がある傾向が見られた。一方、吹き出し表示をしても何らかの真偽確認行動に移らない流言は、複数の事項が不明確であり、内容の特定に複数回の検索が必要になる可能性がある傾向が見られた。これらの傾向は、今後のシステム開発における検討指針として用いることができる可能性がある。
- ただし、これらの知見は、ある程度システムを利用した

ことがあるという条件のもとで確認された知見であり，今回分析対象とした利用ログにおいては，一定の利用が見られたユーザ自体が少ないため，今後継続して一般公開中の利用ログを収集し，分析を進める必要がある。

今後は，本システムの吹き出しが提供する，直接的に取得可能な情報の閲覧以外の，Web 検索など他ページへの遷移による真偽確認の促進について検討する。

謝辞 本研究は，JSPS 科研費 15H05317，19H04221 の助成による。

参考文献

- [1] Kwak, H., Lee, C., Park, H. and Moon, S.: What is Twitter, a social network or a news media?, *Proc. 19th International Conference on World Wide Web*, pp.591–600, ACM (2010).
- [2] 小林啓倫：災害とソーシャルメディア—混乱，そして再生へと導く人々の「つながり」，pp.140–187, 毎日コミュニケーションズ (2011).
- [3] 佐藤健二：関東大震災における社会の変容，立命館大学・神奈川大学 21 世紀 COE プログラムジョイントワークショップ報告書『歴史災害と都市—京都・東京を中心に』，pp.81–89 (2007).
- [4] 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).
- [5] 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).
- [6] Gupta, A. and Kumaraguru, P.: Credibility Ranking of Tweets during High Impact Events, *Proc. 1st Workshop on Privacy and Security in Online Social Media*, No.2, pp.1–8, ACM (2012).
- [7] Lim, W.Y., Lee, M.L. and Wynne, H.: iFACT: An Interactive Framework to Assess Claims from Tweets, *Proc. 2017 ACM on Conference on Information and Knowledge Management*, pp.787–796, ACM (2017).
- [8] Mendoza, M., Poblete, B. and Castillo, C.: Twitter Under Crisis: Can we trust what we RT?, *Proc. 1st Workshop on Social Media Analytics*, pp.71–79, ACM (2010).
- [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).
- [10] Rob Ennals, B.T. and Agosta, J.M.: Highlighting disputed claims on the web, *Proc. 19th International Conference on World Wide Web (WWW '10)*, pp.341–350, ACM (2010).
- [11] Schwarz, J. and Morris, M.: Augmenting web pages and search results to support credibility assessment, *Proc. SIGCHI Conference on Human Factors in Computing Systems (CHI'11)*, pp.1245–1254, ACM (2011).
- [12] Saito, Y.S.F. and Yamamoto, Y.: Highlighting Weasel Sentences for Promoting Critical Information Seeking on the Web, *Proc. 20th International Conference on Web Information Systems Engineering (WISE 2019)*, pp.1245–1254, ACM (2011).
- [13] 吉田 悠，池上輝哉，福住伸一：オブジェクトの速度変化とユーザの気づきに関する考察—データ数増減の可視化手法の開発とユーザ実験による検証，*情報処理学会/研究報告ヒューマンコンピュータインタラクション (HCI)*，Vol.2012-HCI-148, No.21, pp.1–8 (2012).



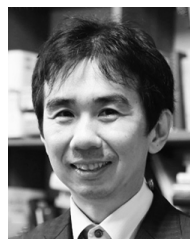
柿本 大輔

2017 年和歌山大学システム工学部デザイン情報学科卒業。2019 年同大学大学院システム工学研究科博士前期課程修了。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。



荒牧 英治 (正会員)

京都大学総合人間学部基礎科学科卒業。東京大学大学院情報理工学系研究科博士課程修了。博士 (情報理工学)。現在，奈良先端科学技術大学院大学教授。自然言語処理の医療応用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。CSCW, HCI の研究に従事。