

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

柿本大輔*¹ 宮部真衣*² 荒牧英治*³ 吉野孝*⁴

Construction of Information-credibility Verification-behavior Facilitation System for Preventing False Rumors Spreading

Daisuke Kakimoto*¹, Mai Miyabe*², Eiji Aramaki*³ and Takashi Yoshino*⁴

Abstract – As the widespread of microblogs, we are now capable of obtaining and conveying various information. Microblogs can send out information easily and in real-time, on the other hand, they may widely spread uncertain and false rumors. One of the factors that false rumors are spread is the lack of confirmation about the credibility of the information before retweeting. In this study, we have developed a system to make aware of existence of rumor in web browsing. We found that the proposed system can make aware of existence of rumor and does not disturb web browsing. This system can provide the awareness about rumors to users in web browsing, and can prevent spreading uncertain rumors.

Keywords : rumor, microblog, twitter

1. はじめに

近年、SNS の普及により、ネットワーク上でのコミュニケーションが活性化している。特に、SNS の一つであるマイクロブログサービスでは、ユーザがリアルタイムに多種多様な情報を取得・発信することが可能であり、その普及も著しい。代表的なマイクロブログサービスの一つに、Twitter¹がある。Twitterでは、情報発信時の入力可能文字数が140字に制限されていることなどから、ユーザの情報発信に対するハードルが大きく下がっている^[1]。2011年3月の東日本大震災の際には、Twitterは速報性の高さを発揮し、重要な情報インフラとして活用されていた^{[2][3][4][5]}。

一方、情報の取得・発信時に、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た^[6]。東日本大震災の際、Twitterにおいて多く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、後に独立行政法人放射線医学総合研究所から訂正と注意喚起がなされた^[6]。この

ような流言は、専門知識を持たないユーザにとっては真偽の判断が難しく、その情報を鵜呑みにして行動に至ると、人間の身体に有害な影響を及ぼす可能性がある。特に災害時などでは、流言は発生しやすく^[7]、流言の伝播によって深刻な問題が引き起こされる可能性がある。そのため、流言の拡散を防止する仕組みが必要である。

我々は、流言が拡散する原因の一つとして、情報発信者の発信前の真偽確認不足があると考えた。前述のように、Twitterなどのマイクロブログは手軽に利用可能であり、特に流言が拡散しやすいと考えられる。そこで我々はまず、インターネットにおける流言拡散の防止について検討を行った^[8]。検討においては、マイクロブログの中でも特に利用者の多いTwitterを対象とし、情報確認行動に関する調査を行った。本研究ではこの調査の結果をもとに、流言拡散防止のための情報確認行動促進システムを構築する。

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・拡散されており、Twitterにおける流言に関する研究は盛んである。Twitterにおける流言拡散防止に関する研究としては、流言の検出および信頼性の自動推定に着目したものがある。Takahashiら^[9]は、Twitter上の流言の出現自体を抑えることは困難であるが、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が出来ると述べている。そこで、訂正ツイートの早期発信に役立てるため、ツイートの固有表現、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る、Twitterにおける流言検出手法を提

*1: 和歌山大学大学院 システム工学研究科

*2: 諏訪東京理科大学 経営情報学部

*3: 奈良先端科学技術大学院大学 研究推進機構

*4: 和歌山大学 システム工学部

*1: Graduate School of Systems Engineering, Wakayama University

*2: Faculty of Business Administration and Information, Tokyo University of Science, Suwa

*3: Center for Frontier Science and Technology, Nara Institute of Science and Technology

*4: Faculty of Systems Engineering, Wakayama University
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（悪意をもった捏造か自然発生か）は問わないものとする。

案した。Castillo ら^[10]は、ソーシャルメディアの中には、ユーザが情報の信頼性判断をすることができるシグナルがあると考え、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。Yang ら^[11]は、Sina Weibo³を対象に、Castillo ら^[10]と同様の手法を用いて、情報の信頼性を自動評価する手法を提案した。

また、Twitter におけるリツイート機能⁴は、その手軽さから Twitter 上での情報の伝播に影響を与えやすいと考えられ、特にリツイート機能に着目して、情報の信頼性推定を行った研究がある。中原ら^[12]は、Twitter におけるリツイートのうち、非公式リツイート（引用リツイート）に付加されたコメント、およびリツイートされた回数から、あるツイートに対し訂正ツイートと危険度という2つの情報を取得し、ユーザに提示するシステムを提案した。

このように、流言自体の検出および信頼性の自動推定を試みる研究は多数行われており、流言として検出あるいは信頼性が低いと推定された情報は、流言の拡散防止の一助となる可能性がある。しかし、これらの研究はあくまでも流言と思われる情報を検出することに主眼を置いたものであり、検出した情報をユーザに提供する方法や、ユーザの不用意な拡散を踏みとどまらせる方法など、インタフェース面での検討は行われていない。検出した情報を人々に適切かつ効果的に提供する方法が実現されることで、流言の拡散防止につながると考えられる。そのため、情報を閲覧している段階で、情報確認行動を促すような仕組みおよびインタフェースが必要である。そこで本研究では、ユーザに対し流言に関する気づきを提供し、情報の真偽確認行動を促進することで、流言拡散防止を目指すシステムを検討する。さらに、Twitter 上の拡散防止にとどまらず、より広い場面においてユーザの流言拡散を防ぐためのシステムを目指す。

また、気づきの提供および情報確認行動の促進においては、これまでに我々が提案した流言情報蓄積システム^[13]により収集した流言情報を用いて、ユーザに情報提供を行う。

3. Twitter における情報拡散行動に関する調査

流言が拡散する原因の一つとして、情報発信前の確認不足が考えられる。そこで、Twitter においてリツイート機能を利用する際、ツイート内容の真偽をユーザが確認しているかどうかをアンケートにより調査し

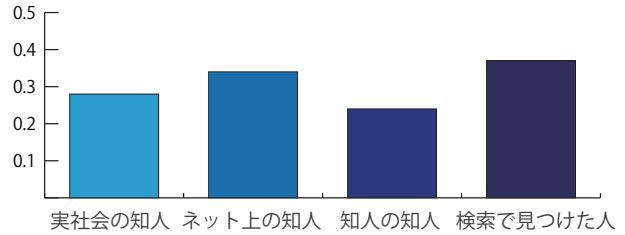


図1 情報の真偽確認を行っている人の割合
Fig.1 Percentage of users confirming the information credibility.

た。調査対象は、Twitter ユーザ 108 人 (10 代: 19 人, 20 代: 25 人, 30 代: 24 人, 40 代: 23 人, 50 代: 17 人) である。また、情報の確認を行うかどうかは、その情報の発信者を知っているかどうか、さらに実社会で会ったことのある人かどうかという、発信者との親しさが関係する可能性がある。そこで、ツイートの発信者を以下の4つに分類し、それぞれの発信者のツイートに対するリツイート時の行動を調査した^[8]。

- ・フォロユーザ (実社会の知人)
- ・フォロユーザ (ネット上の知人)
- ・フォロユーザのフォロユーザ (知人の知人)
- ・非フォロユーザ (検索で見つけた人)

なおアンケートは、ユーザ属性、Twitter の利用方法および情報拡散行動に関する質問 (計 44 問) からなる。

図1に、リツイート利用経験者 (実社会の知人: 108 人中 86 人, ネット上の知人: 108 人中 93 人, 知人の知人: 108 人中 76 人, 検索で見つけた人: 108 人中 59 人) のうち、リツイート前に情報の真偽を確認していると答えた人の割合を示す。情報の真偽を確認していると答えた人の割合は、4 種類のツイート全てにおいて、0.4 を下回った。このことから、Twitter における情報拡散時、情報の真偽を確認している人の割合が低いことがわかった。

4. 情報確認行動促進システム

3 章で述べたように、リツイート対象とするツイート内容の真偽を意識せずにリツイート機能を利用している Twitter 利用者が多く、流言が拡散されやすい可能性があることがわかった。これまでに、Twitter 上で流言の拡散防止を支援する手法およびシステムは提案されている^{[9][10][12][13]}が、情報の真偽に対する関心の低さを考慮すると、情報取得時および拡散前に、これらのシステムを利用しその情報の真偽を確認するという、能動的な行動を行う人は少ないと考えられる。流言の拡散を防止するためには、このような人々に情報の真偽確認行動を促すことが重要であると考えられる。

そこで本研究では、「閲覧している情報は流言である

3: <http://weibo.com>

4: 他のユーザが発信したツイートを再発信することができる機能。リツイートには、情報を元の状態のまま再発信する公式リツイートとコメントを付加して再発信する非公式 (引用) リツイートがある。

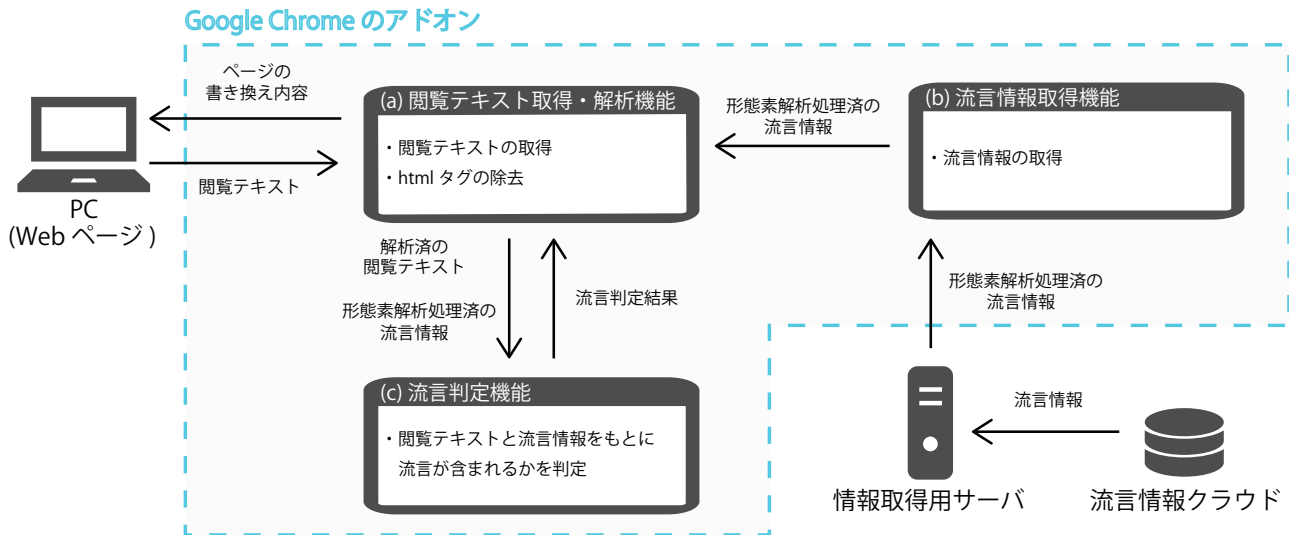


図 2 システムの構成
Fig. 2 System configuration.

可能性がある」ということをユーザに気づかせ、情報の真偽確認行動を促進することで、流言の拡散防止を支援するシステムを構築する。

4.1 設計方針

Twitter ユーザは情報の真偽を確認することなくリツイートすることが多いことを考慮すると、システムの構築においては、ユーザによる真偽確認負担を軽減することが重要であると考えられる。また、流言は Twitter だけでなく、その他の SNS や Web サイト上にも掲載される可能性があるため、Twitter に限定せず、インターネット上の幅広い情報閲覧場面において、流言拡散を防止できることが望ましい。そこで、流言が含まれる可能性がある情報を閲覧している際、ユーザに気づきを提供することで、流言の拡散防止を試みるシステムを構築する。

システムの設計方針を以下に示す。

1. 真偽確認作業の負担を軽減する
ユーザが目にする全ての情報に関して、ユーザ自身が能動的に真偽確認することは難しい。そこで、ユーザが閲覧中のページに含まれる情報に、流言が含まれるかどうかをシステムが判定し、流言が含まれる場合、ユーザが受動的に情報を受け取ることができる仕組みを構築する。
2. 幅広い場面での対応を可能にする
流言は Twitter だけでなく、その他の SNS や Web ページなどにも掲載される可能性がある。そこで、Web ブラウザのアドオン機能として実装することにより、ブラウザ上での様々なページ閲覧時に、システムから情報が提供されるようにする。
3. 最新の流言情報に基づいて情報を提供する
ユーザに注意を促す際には、新しく発生した流言

情報などにも対応できることが望ましい。そこで、2012 年から運用している流言情報クラウド^[13]と連携し、最新の流言情報を提供可能にする。

4.2 システムの構成

本システムは、Google 社⁵が提供する Web ブラウザである、Google Chrome⁶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:58.0)の全世界シェアは、2017 年 7 月 10 日時点で 44.5%であり、1 位となっている⁷。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開発を行うことにした。

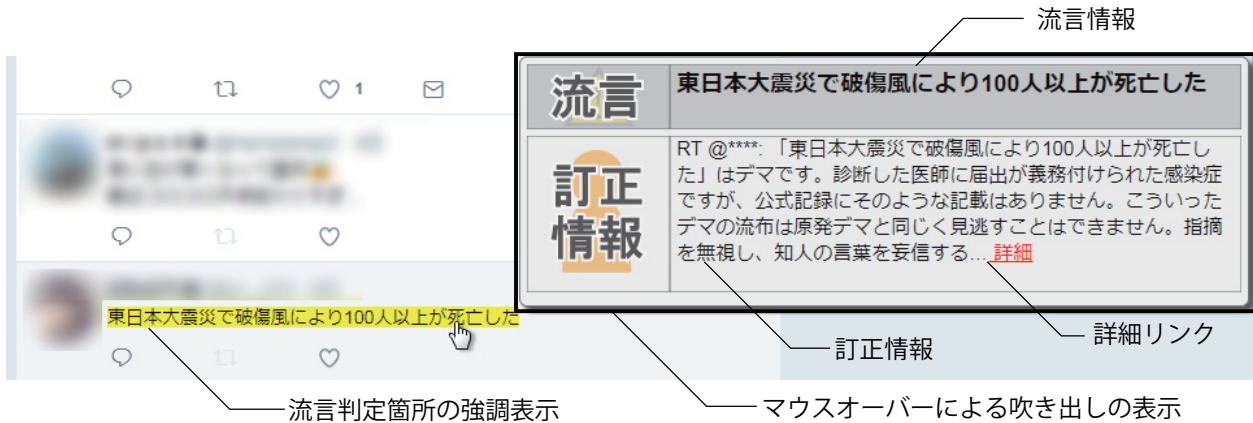
システムの構成を図 2 に示す。まず、(a) 閲覧テキスト取得・解析機能により、Web ブラウザで閲覧しているコンテンツのテキスト(閲覧テキスト)を取得する。同時に、(b) 流言情報取得機能により、データ取得用サーバを介して、流言情報を取得する。次に、取得した閲覧テキストおよび流言情報をもとに、(c) 流言判定機能により、流言判定を行う。最後に、その結果をもとに閲覧中 Web ページ上でのテキストの強調表示と、該当部分のマウスオーバーによる吹き出しの表示という形式でページの書き換えを行い、ユーザに流言に関する気づきを提供する。

システムの動作画面例を図 3 に示す。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。テキストを強調する手法としては、下線や

5: <http://www.google.co.jp/>

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

7: https://webrage.jp/techblog/pc_browser_share/



流言判定箇所の強調表示：流言が含まれる可能性のある部分を判定し，強調表示。
 マウスオーバーによる吹き出しの表示：強調表示部分のマウスオーバーで，吹き出しを表示。
 流言情報：判定された流言情報。
 訂正情報：流言情報に対する訂正ツイート。
 詳細リンク：これまでの訂正数および直近一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。

図3 システムの動作画面例
 Fig.3 Example of the system interface.

太字などによる強調があるが，これらは Web ページにおける URL の記載や，そのページ内の本来の強調部分と混同してしまう可能性がある．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，本システムが行うテキストの強調表示については，背景色を変更するという形をとった．また，吹き出しの表示についても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），詳細リンク（これまでの訂正数および直近一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）が含まれる．

4.3 各機能の実装

4.2節で述べた (a) 閲覧テキスト取得・解析機能，(b) 流言情報取得機能および (c) 流言判定機能の実装手法を以下に示す．

(a) 閲覧テキスト取得・解析機能

アドオンを起動している場合，Web ページに挿込まれるスクリプトファイルにより，ユーザが閲覧しているテキストを取得し，不要な html タグの除去を行う．また，流言判定の結果をもとに，閲覧ページの書き換えを行う．

(b) 流言情報取得機能

バックグラウンドで常時動作しているスクリプトファイルにより，情報取得用サーバとの通信を介して，流言情報クラウドから最新の流言情報の取得を行う．また，情報取得用サーバでは定期的に流言情報クラウドから流言情報を取得し，Juman^[14]による形態素解析を行う．

(c) 流言判定機能

取得した閲覧テキストと形態素解析済みの流言情報をもとに，閲覧中のテキストに流言が含まれているかどうかを判定する．この際，閲覧テキストを一行ごと⁸に分割し，流言の形態素解析結果のうち，助詞以外が全て含まれていれば，そのテキストに流言が含まれているとみなす．

5. 評価実験

本研究で構築したシステムは，閲覧中の情報に流言が含まれる可能性があることに対する気づきをユーザに提供し，真偽確認行動を促進することで，流言拡散防止を目指すものである．つまり，従来の情報閲覧時には存在しない，本来は要求されたものではない追加の情報を自動的⁹に提供するインタフェースである．もし，システムの利用が，ユーザの情報閲覧を妨げてしまうと，情報の真偽に対する関心の低いユーザは，本システムを利用しなくなる可能性がある．そのため，本研究で構築したシステムにおいては，ユーザの受容性の高さが重要であると考えられる．そこで，システムが情報閲覧の妨げにならず，ユーザに流言に関する気づきを提供することができるかどうかを検証するために，評価実験を行う．

本システムは，Web ブラウザによるページ閲覧における幅広い場面での利用を可能にするために，Web プ

8: 改行コードを行の区切りとみなし，改行コードが出現した時点で分割する．そのため，一行内に複数の文が含まれる場合もある．

9: アドオンはユーザの意思で動作・停止することが可能であり，アドオンの動作中は，流言判定箇所の強調表示については自動的に行われる．なお，吹き出し表示は，ユーザが強調表示箇所をマウスオーバーした際に表示されるため，自動的なインタフェースではない．

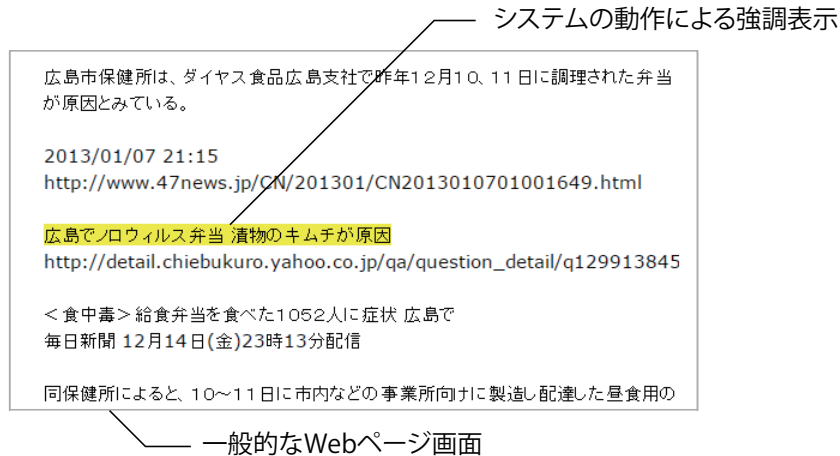


図4 評価対象とした Web ページの例
 Fig. 4 Example of a web page used for evaluation.

ブラウザのアドオンとして実装した。そこで、まず、一般的な Web ページ閲覧時を対象とした評価実験を行う(実験1)。また、流言は Twitter のような、様々なユーザが発信した情報が入り交ざった環境において拡散されることが多い。そこで、Twitter のタイムライン¹⁰ 閲覧時を対象とした評価実験を行う(実験2)。

5.1 検証仮説

本研究で構築したシステムは、ユーザの Web ページ閲覧の妨げにならずに、流言に関する気づきを提供することができる必要がある。そこで本実験では、以下の仮説を立て、検証を行う。

仮説1: システムの動作は、ユーザのページ閲覧の妨げにならない

本システムは、Web ページ上で流言判定部分を強調表示し、吹き出し表示による流言情報の提示を行う。システムの動作がユーザのページ閲覧の妨げになってしまうと、システム自体を使ってもらえなくなる可能性がある。そのため、システムの利用が、ページ閲覧の妨げにならないかどうかを評価する。

仮説2: システムは、流言に関する気づきをユーザに提供することができる

本システムは、Web ページ閲覧時、ユーザに流言に関する気づきを提供し、流言の拡散防止を支援する。そのため、ページ閲覧時に、システムによる強調表示および吹き出し表示から、ユーザが流言情報を取得し、ページ閲覧の役に立ったかどうかを検証する。

5.2 実験1の概要

実験1では、流言情報が含まれる Web ページ 15 件¹¹ を対象に、各ページ閲覧における本システムの利用に関する評価を行う。実験協力者は大学生および大学院生の男女 10 名(男性: 8 名, 女性: 2 名)である。Web ページは、見出しや同じ段落、近くの段落などに同じ内容の記述が複数回出現することがある。流言内容が多数出現し、多数の強調表示が行われると、ユーザのページ閲覧を妨げる可能性がある。そこで、強調表示される箇所数を 3 種類に分けて実験を行った。Web ページは、流言判定が 1 箇所～2 箇所, 3 箇所～5 箇所, 6 箇所以上行われるページの 3 種類について、それぞれ 5 件ずつ収集した(計 15 ページ)。各ページに含まれる流言の内容は、15 件のページで全て異なる。また本実験は、システムが Web ページ閲覧の妨げにならず、流言に関する気づきをユーザに与えられるかどうかを検証することが目的であるため、各 Web ページ内の流言に関するテキストが正しく強調表示されるよう、流言の形態素解析結果には著者が修正を加えている。さらに、評価の際は、Web ページの評価順による実験結果への影響を避けるため、協力者ごとに Web ページの評価順をランダムに並び替えた。実験に用いた Web ページの例を図 4 に示す。

協力者には、各ページ閲覧終了時¹²に、「Q1. 強調

10: フォローしているユーザの発信したツイートが一覧で表示されるページ

11: システムが動作した際、ユーザの操作に問題がないかどうかを評価するため、流言の含まれる Web ページを用いる必要がある。そのため事前に流言情報が含まれるかどうかを確認し、流言情報が含まれる Web ページを実験対象とした。

12: 各ページの閲覧時間に制限は設けず、実験協力者が「表示された Web ページの内容をある程度把握できた」と判断した段階で終了とした。なお、実験協力者には強調表示および吹き出し表示機能についての説明を行ったが、強調表示された箇所のマウスオーバーによる吹き出し表示を行うかどうかについては指示せず、自由に閲覧してもらった形式をとった。

表1 Q1 (強調表示は、ページ閲覧の妨げにならなかった) に対する結果
Table 1 Result of Q1 (The highlighting did not disturb the page browsing) .

流言判定箇所	評価の分布					中央値	最頻値	平均値
	1 (人)	2 (人)	3 (人)	4 (人)	5 (人)			
少: 1~2箇所	0	0	1	3	6	5	5	4.5
中: 3~5箇所	0	0	2	5	3	4	4	4.1
多: 6箇所以上	0	2	1	3	4	4	5	3.9

評価の分布は、「1. 強く同意しない」、「2. 同意しない」、「3. どちらともいえない」、「4. 同意する」、「5. 強く同意する」である。なお、流言判定箇所別のページ3種類における、それぞれ5件の評価結果の中央値を各協力者の代表値としている。

表2 Q2 (吹き出し表示は、ページ閲覧の妨げにならなかった) に対する結果
Table 2 Result of Q2 (The speech balloon did not disturb the page browsing) .

流言判定箇所	評価の分布					中央値	最頻値	平均値
	1 (人)	2 (人)	3 (人)	4 (人)	5 (人)			
少: 1~2箇所	0	0	0	4	6	5	5	4.6
中: 3~5箇所	0	1	0	5	4	4	4	4.2
多: 6箇所以上	0	1	2	4	3	4	4	3.9

評価の分布は、「1. 強く同意しない」、「2. 同意しない」、「3. どちらともいえない」、「4. 同意する」、「5. 強く同意する」である。なお、流言判定箇所別のページ3種類における、それぞれ5件の評価結果の中央値を各協力者の代表値としている。

表3 システムの利用に関するアンケート結果
Table 3 Result of the questionnaire about the use of the system.

	質問項目	評価の分布					中央値	最頻値	平均値
		1 (人)	2 (人)	3 (人)	4 (人)	5 (人)			
Q3	「流言情報」はページ閲覧をする上で理解の助けになった。	0	0	5	4	1	3.5	3	3.6
Q4	「訂正情報」はページ閲覧をする上で理解の助けになった。	1	2	2	5	0	3.5	4	3.1
Q5	「詳細リンク」はページ閲覧をする上で理解の助けになった。	2	2	3	3	0	3	3,4	2.7

評価の分布は、「1. 強く同意しない」、「2. 同意しない」、「3. どちらともいえない」、「4. 同意する」、「5. 強く同意する」である。

表示はページ閲覧の妨げにならなかった」および「Q2. 吹き出し表示は、ページ閲覧の妨げにならなかった」の2つの評価項目に、5段階のリッカート尺度(1. 強く同意しない, 2. 同意しない, 3. どちらともいえない, 4. 同意する, 5. 強く同意する)で回答および理由の記述をしてもらった。また、15件全てのWebページの閲覧およびQ1, Q2の評価終了後には、システムの利用に関するアンケート(Q3(「流言情報」はページ閲覧をする上で理解の助けになった), Q4(「訂正情報」はページ閲覧をする上で理解の助けになった), Q5(「詳細リンク」はページ閲覧をする上で理解の助けになった))に、5段階のリッカート尺度(1. 強く同意しない, 2. 同意しない, 3. どちらともいえない, 4. 同意する, 5. 強く同意する)で回答および理由の記述をしてもらった。

なお本論文では、Q1~Q5の評価尺度として5段階のリッカート尺度を用いた。本論文で用いた評価尺度

(リッカート尺度)については、順序尺度として扱う¹³。評価結果については、参考値として平均値を示し、結果の考察については中央値および最頻値に基づいて行うこととする¹⁴。

5.3 実験1の結果

Q1(強調表示は、ページ閲覧の妨げにならなかった)に対する評価結果を表1に示す。また、Q2(吹き出し表示は、ページ閲覧の妨げにならなかった)に対する評価結果を表2に示す。実験1における、Q1(強調表示は、ページ閲覧の妨げにならなかった)およびQ2(吹き出し表示は、ページ閲覧の妨げにならなかった)に対し、協力者にはページごとの評価を行っ

13: 評価尺度(リッカート尺度)の等間隔性を確保するためのワーディング等の検討・改善については行っておらず、順序尺度を間隔尺度として安定して使用できるかどうかについては保証されていない^[16]ためである。

14: 少数のデータにおいては、平均値は外れ値の影響が強く現れるためである。

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

表4 システムの良かった点・悪かった点・要望 (実験1)

Table 4 Good points/Bad points/Request about the system (experiment 1).

良かった点	<ul style="list-style-type: none"> ・直感的に操作できる. ・流言に気づきやすくなる. ・情報を鵜呑みにすることが無くなる.
悪かった点	<ul style="list-style-type: none"> ・提示される訂正情報自体の信頼性がわからない. ・同じページ内に、ハイライトが複数あると読みづらい.
要望	<ul style="list-style-type: none"> ・文字の色に合わせて強調表示の色を変えて欲しい. ・吹き出しから、直接オリジナルの流言情報にアクセスしたい.

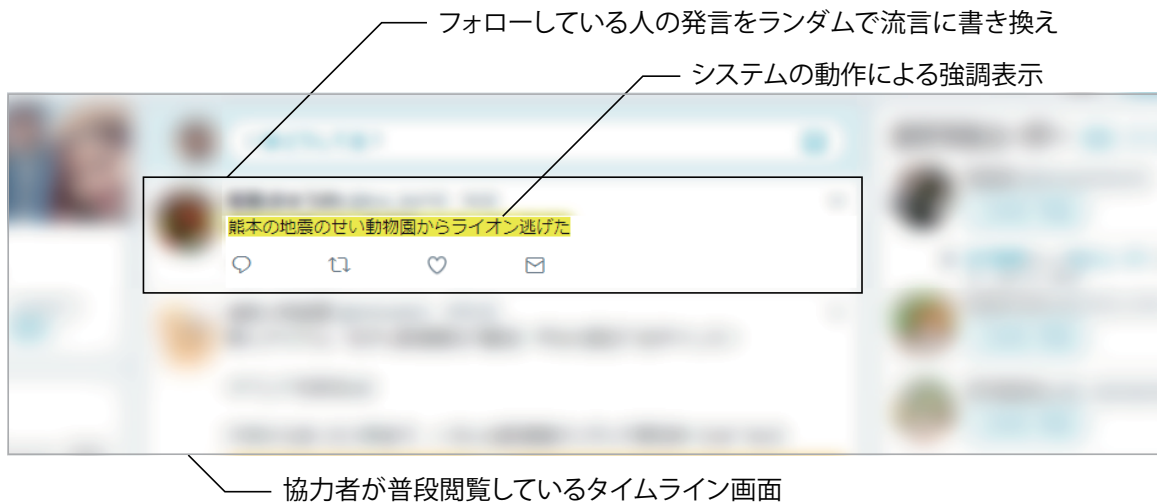


図5 評価対象とした Twitter タイムラインの例
Fig.5 Example of a timeline on twitter used for evaluation.

てもらった。各協力者において、流言判定箇所数別のページ5件での評価結果に大きなばらつきはなかったため、これらの結果については、流言判定箇所数別のページ5件の評価結果の中央値を代表値とし、各協力者の評価結果として扱う。表1より、Q1(強調表示は、ページ閲覧の妨げにならなかった)に対する評価結果における、実験協力者間での中央値および最頻値は、流言判定箇所数に関わらず、全て4以上となった。表2より、Q2(吹き出し表示は、ページ閲覧の妨げにならなかった)に対する評価結果においても同様に、実験協力者間での中央値および最頻値が、流言判定箇所数に関わらず、全て4以上となった。これらの結果より、本システムはユーザのWebページ閲覧を妨げることなく利用可能であることがわかった。しかし、流言判定箇所が中程度(3~5箇所)および多い(6箇所以上)のページにおいては、評価結果が2となった協力者がいるため、流言判定箇所数の多いページに対しては、システムによる強調表示および吹き出し表示がページ閲覧の妨げになる可能性がある。

システムの利用に関するアンケート(Q3(「流言情報」はページ閲覧をする上で理解の助けになった)、Q4(「訂正情報」はページ閲覧をする上で理解の助け

になった)。Q5(「詳細リンク」はページ閲覧をする上で理解の助けになった))の結果を表3に示す。表3より、Q5(「詳細リンク」はページ閲覧をする上で理解の助けになった)では、1および2の結果となった協力者が4名おり、高評価とはならなかったものの、Q3(「流言情報」はページ閲覧をする上で理解の助けになった)では、全ての協力者において3以上の結果となり、高評価が得られた。また、Q4(「訂正情報」はページ閲覧をする上で理解の助けになった)では、1および2の結果となった協力者が3名いるものの、最頻値が4となっており、概ね高評価が得られた。これらの結果より、吹き出しに提示する情報のうち、流言情報および訂正情報はユーザのページ閲覧の助けになる可能性があることがわかった。

システムの良かった点、悪かった点および要望に関する回答結果の一部を表4に示す。システムの良かった点に関する回答結果として、「流言に気づきやすくなる」「情報を鵜呑みにすることがなくなる」などが得られ、「ユーザに対し閲覧中の情報に流言が含まれる可能性があることを気づかせ、真偽確認行動を促進する」という、システムの目的は満たしていると考えられる。

表5 アンケート結果 (実験2)
Table 5 Result of the questionnaire (experiment 2).

	質問項目	評価の分布					中央値	最頻値	平均値
		1 (人)	2 (人)	3 (人)	4 (人)	5 (人)			
Q1	強調表示は、タイムラインの閲覧の妨げにならなかった。	0	0	0	6	4	4	4	4.4
Q2	吹き出し表示は、タイムラインの閲覧の妨げにならなかった。	0	0	2	3	5	4.5	5	4.3
Q3	「流言情報」はタイムラインの閲覧をする上で理解の助けになった。	0	1	1	4	4	4	5	4.1
Q4	「訂正情報」はタイムラインの閲覧をする上で理解の助けになった。	0	1	2	6	1	4	4	3.7
Q5	「詳細リンク」はタイムラインの閲覧をする上で理解の助けになった。	0	4	3	2	1	3	2	3.0

評価の分布は、「1. 強く同意しない」、「2. 同意しない」、「3. どちらともいえない」、「4. 同意する」、「5. 強く同意する」である。

表6 システムの良かった点・悪かった点・要望 (実験2)
Table 6 Good points/Bad points/Request about the system (experiment 2).

良かった点	<ul style="list-style-type: none"> ・流言に対して関心を持つようになった。 ・詳しく見たいと思った情報だけを見ることができた。 ・普段タイムラインを閲覧している際と変わりなくシステムを利用できた。
悪かった点	<ul style="list-style-type: none"> ・強調表示以外の部分を見落とす可能性がある。 ・詳細リンクが必要ないと思った。
要望	<ul style="list-style-type: none"> ・情報提示方法をユーザが制御できると良いと思った。 ・オリジナルの流言をたどれたり、訂正情報以外のソースがあればいいと思った。

表7 強調表示情報の選択状況
Table 7 Selection results of the highlighted information.

	協力者									
	1	2	3	4	5	6	7	8	9	10
A. 強調表示されており、選択した (件)	9	10	8	7	8	9	8	10	9	6
B. 強調表示されていたが、選択しなかった (件)	1	0	2	3	2	1	2	0	1	4
C. 強調表示されていなかったが、選択した (件)	0	0	0	0	0	0	0	0	0	0
D. 強調表示されておらず、選択しなかった (件)	10	10	10	10	10	10	10	10	10	10
正解率	0.95	1.00	0.90	0.85	0.90	0.95	0.90	1.00	0.95	0.80
吹き出しを表示させた (件)	6	9	6	3	8	0	0	10	10	9

正解率は、(A+D) / (A+B+C+D) である。

5.4 実験2の概要

実験2では、WebブラウザでTwitterのタイムラインを閲覧する際の本システムの利用に関する評価を行う。実験協力者は大学生および大学院生の男女10名(男性:9名, 女性:1名)である。各協力者がフォローしているユーザのツイート100件が表示されたTwitterのタイムラインページを協力者ごとに取得し、それぞれのページについてその中の10件のツイートを流言に書き換えた¹⁵。書き換えに用いた流言は全

て異なる内容のものとし、書き換えるツイートはランダムで選定することとした。また、実験1で対象とした一般的なWebページに対し、Twitterタイムラインにおいては、ページ内に同内容の記述が複数重複する可能性が低いと考えられるため、本実験では流言判定箇所数別の評価は行わなかった。なお、実験2における評価対象ページの流言判定箇所数については、実験1における流言判定箇所数6箇所以上の条件に相当する。ただし、ページ全体のテキスト量に対する強調表示箇所の量を考慮すると、実験時の状況としては、実験1における流言判定箇所数6箇所未満に相当すると考えられる。実験に用いたTwitterタイムラインの例

15: 実験の際、一時的な書き換えを行った。協力者には実験終了後、その旨を伝え、本来の情報発信者とは一切関係がないということを理解してもらった。

を図 5 に示す。

実験協力者には、Twitter タイムラインの閲覧終了時¹⁶、システムの利用に関するアンケートに、5段階のリッカート尺度(1.強く同意しない, 2.同意しない, 3.どちらともいえない, 4.同意する, 5.強く同意する)で回答および理由の記述をしてもらった。また、テキストの強調表示は、強調した情報の再生に有効であることがわかっている^[15]が、Twitter タイムラインのような様々な情報が混在する状況では、強調表示を見落とし、流言の存在に気づかない場合があると考えられる。そのため、タイムライン上に含まれる流言に関して、ユーザに気づきを与えられたかどうかを検証する。協力者には、20件の流言(書き換えに用いた流言10件と、実験では用いていない1流言10件)から、実験において強調表示がされていた情報がどれかを選択してもらった。その際、選択する数については教示せず、協力者には当てはまると思う情報を全て選択してもらうこととした。

5.5 実験2の結果

システムの動作に関する評価結果およびアンケート結果を表5に示す。Q1(強調表示は、タイムラインの閲覧の妨げにならなかった)に対する評価結果については、全ての協力者において4以上の評価結果が得られた。また、Q2(吹き出し表示は、タイムラインの閲覧の妨げにならなかった)に関する評価結果については、8人の協力者において4以上の評価結果が得られた。評価結果が3となった2人の協力者に関しては、実験時、吹き出し表示を一度も行わなかった協力者であったため、「3:どちらともいえない」と評価したと考えられる。つまり、吹き出し表示を行った協力者は全員「妨げにならなかった」と回答しており、本システムはユーザのTwitterのタイムライン閲覧を妨げることなく利用可能であることがわかった。

また、Q3~Q5に対する評価結果については、Q5(「詳細リンク」はタイムラインの閲覧をする上で理解の助けになった)に対する結果が低評価である傾向が見られるものの、Q3(「流言情報」はタイムラインの閲覧をする上で理解の助けになった)およびQ4(「訂正情報」はタイムラインの閲覧をする上で理解の助けになった)に対する結果は高評価が得られた。これらの結果より、実験1と同様に、吹き出しに提示する情報のうち、流言情報および訂正情報はユーザのタイムライン閲覧の助けになる可能性があることがわかった。

システムの良かった点、悪かった点および要望に関する回答結果の一部を表6に示す。システムの良かった点に関する回答結果として、「流言に対して関心を持つようになった」などが得られ、実験1の結果と同様に、「ユーザに対し閲覧中の情報に流言が含まれる可能性があることを気づかせ、真偽確認行動を促進する」という、システムの目的は満たしていると考えられる。

実験において強調表示がされていた情報がどれか選択してもらった結果を表7に示す。強調表示されていた10件の情報のうち、8件以上の情報を正しく選択することができた協力者は、8名であった。強調表示されていなかった情報を誤って選択した協力者は0名であった。強調表示されていたいくつかの情報について、正しく選択することができなかった協力者は、8名であった。また、正解率は、全ての協力者において0.80以上であった。これらの結果より、システムによる強調表示は、ユーザに流言に関する気づきを提供することができる可能性があると考えられる。さらに、8名の協力者が、吹き出し表示を行い、そのうち7名の協力者は6回以上吹き出し表示を行ったことより、本システムによる強調表示が、情報確認行動を促進することができる可能性がある。

6. 考察

5章で述べた実験の結果から、「仮説1:システムの動作は、ユーザのページ閲覧の妨げにならない」、「仮説2:システムは、流言に関する気づきをユーザに提供することができる」が成り立つかどうかについて考察する。

実験1および実験2において、Q1(強調表示は、タイムラインの閲覧の妨げにならなかった)およびQ2(吹き出し表示は、タイムラインの閲覧の妨げにならなかった)に対する回答結果から、どちらも高評価が得られた。しかし実験1における、Q1(強調表示は、タイムラインの閲覧の妨げにならなかった)では、流言判定箇所が6箇所以上のページでの評価結果が2である協力者が2人となった(表1)。これらの協力者が、評価値を2以下とした際の評価理由には、「強調表示部分が多すぎる」「強調表示部分を注視しすぎてしまう」「ページデザイン上、強調された文字が読みにくくなっていった」などがあつた。このことから、流言判定数の多いページにおいては、ページ閲覧の妨げになると感じるユーザも存在することがわかった。一方、実験2においては、Q1(強調表示は、タイムラインの閲覧の妨げにならなかった)およびQ2(吹き出し表示は、タイムラインの閲覧の妨げにならなかった)に対し2以下の回答をした協力者はいなかった。Twitterのタイムライン(実験2)と流言判定箇所数数が6箇所以

16: Twitter タイムラインの閲覧時間に制限は設けず、実験協力者が「表示されたタイムラインの内容をある程度把握できた」と判断した段階で終了とした。なお、実験協力者には強調表示および吹き出し表示機能についての説明を行ったが、強調表示された箇所のマウスオーバーによる吹き出し表示を行うかどうかについては指示せず、自由に閲覧してもらう形式をとった。

上のページ（実験1）は、強調表示される箇所数は同一区分であるが、流言判定箇所数が6箇所以上のページの評価が低くなった。この理由としては、以下が考えられる。

- ・実験1で対象とした一般的なWebページ閲覧時では、同じ内容の情報が複数強調表示される。

- ・Twitterのタイムライン閲覧時と比較してページ全体のテキスト量が少なく、強調表示箇所が密集する。

以上より、流言判定箇所数が6箇所未満のWebページおよびTwitterタイムライン閲覧場面では、本システムによる強調表示および吹き出し表示はユーザのページ閲覧の妨げにならず、「仮説1：システムの動作は、ユーザのページ閲覧の妨げにならない」が成り立つと考えられる。しかし、流言判定箇所数が6箇所以上のWebページについては、ユーザのページ閲覧の妨げになる可能性がある。そのため、ページ内全体のテキスト量における、強調表示の対象となるテキスト量の割合を考慮した、ページごとの強調色および流言提示方法について、検討する必要があると考えられる。また、本研究における評価実験では、強調表示および吹き出しの表示による影響のみしか考慮していないため、今後は、実行時間など、その他の要素への影響も考慮し、検証する。

実験2において、強調表示されていた流言を選択してもらった結果では、全ての協力者の正解率が0.80以上となった。また、システムの悪かった点に対する回答結果（表4,6）では、システムが提供する情報自体の信頼性を疑問視する回答もあったものの、システムの良かった点に対する回答結果では、「情報を鵜呑みにすることはなくなると思う」というような回答が得られた。しかし、吹き出し表示により提供される情報について、Q5（「詳細リンク」はページ閲覧をする上で理解の助けになった）では、評価結果が2である協力者が4人となった。この理由としては、ユーザは閲覧ページから新たにページ遷移を行ってまで情報の真偽確認を行わない、詳細リンクへの遷移により閲覧可能である、その情報のこれまでの訂正数などの情報が真偽確認において役立たない可能性があるなどの理由が考えられる。このことから、吹き出し表示により提供する情報については、「流言情報」および「訂正情報」以外の情報を検討する必要があると考えられる。以上より、ユーザの情報拡散行動前に、流言に関する気づきを提供するという、本システムの目的については満たしており、「仮説2：システムは、流言に関する気づきをユーザに提供することができる」については、立証されたと考えられる。しかし、強調表示されていた情報を選択させる実験では、その情報の内容を覚えているかどうかについての影響を考慮していない。そのた

め、今後は、流言を強調表示した場合およびしなかった場合の比較実験を行う。また、吹き出し表示により提供する情報としては、ユーザの真偽確認負荷をより軽減する情報の提供および真偽確認の助けになる情報の提供について、検討する必要がある。

7. おわりに

流言は、ユーザ間の適切な情報共有を阻害し、災害時などでは特に深刻な問題を引き起こす可能性がある。

我々は、インターネットにおける流言拡散防止を検討するにあたり、まず、Twitterにおける情報拡散行動に関する調査を行った。調査の結果、リツイート前にその情報の真偽について確認を行うユーザは少ないことが明らかとなった。そこで本研究では、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の構築を行った。システムの設計方針として、(1) 真偽確認作業負担を軽減する、(2) 幅広い場面での対応を可能にする、(3) 最新の流言情報に基づいて情報を提供する、という3点を立て、WebブラウザであるGoogle Chromeのアドオンとしてシステムの構築を行った。本システムは、流言である可能性のある情報をユーザが閲覧している際、該当する部分を強調表示することにより、ユーザに情報の信頼性に関する気づきを提供する。また、強調表示部分をマウスオーバーすることで、流言情報、訂正情報および詳細リンクが吹き出し表示されるため、ユーザは受動的にこれらの情報を取得することができる。

システムの評価実験の結果、一般的なWebページ閲覧およびTwitterのタイムライン閲覧時、「強調表示は、ページおよびタイムラインの閲覧の妨げにならなかった」、「吹き出し表示は、ページおよびタイムラインの閲覧の妨げにならなかった」という質問項目に対し、5段階評による中央値および最頻値が全て4以上となり、システムはユーザのページ閲覧を妨げていないことがわかった。また、Twitterのタイムライン閲覧時、強調表示されていた流言を選択してもらった結果で、全ての協力者の正解率が0.80以上となり、システムは、流言に関する気づきを提供することができることがわかった。

今後は、Webページに対する流言判定の精度検証と改善を行う。また、流言に関する気づきの提供および情報確認行動の促進により、流言の拡散防止を支援することができるかどうかを、評価実験により検証する。

謝辞

本研究は、JSPS 科研費 15H05317 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] 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).
- [3] インプレス R&D インターネットメディア総合研究所: インターネット白書 2011; インプレスジャパン (2011).
- [4] 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).
- [5] 立入勝義: 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).
- [6] 荻上チキ: 検証 東日本大震災の流言・デマ; 光文社, pp. 27-28 (2011).
- [7] 田中淳, 吉井博明: 災害情報入門 (シリーズ災害と社会7); 弘文堂 (2008).
- [8] 柿本大輔, 宮部真衣, 荒牧英治: 流言拡散防止のための情報確認行動促進システムの提案; 第15回情報科学技術フォーラム講演論文集, No. 2, pp.107-108, (2016).
- [9] Takahashi, T., 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, pp.452-457 (2012).
- [10] Castillo, C., Mendoza, M., B, Poblete.: Information Credibility on Twitter; Proceedings of the 20th international conference on World wide web, pp.675-684 (2011).
- [11] Yang, F., Liu, Y., Yu, X., Yang, M.: Automatic detection of rumor on Sina Weibo; Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, No. 13, pp.1-7 (2012).
- [12] 中原英美, 富永一成, 牛尼剛聡: ツイット構造を用いたデマ拡散防止支援手法; DEIM Forum 2012, F2-3, pp.1-7 (2012).
- [13] 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).
- [14] Daisuke, K., Kurohashi, S.: A Fully-Lexicalized Probabilistic Model for Japanese Syntactic and Case Structure Analysis; In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL2006), pp.176-183 (2006).
- [15] 関友作: テキストの内容把握に対する箇条書とキーワード強調の影響; 日本教育工学会誌/日本教育工学雑誌, Vol.21, No.SuppL, pp.17-20(1997).
- [16] 井上信次: 項目反応理論に基づく順序尺度の等間隔性-質問紙調査の回答選択肢 (3-5 件法) の等間隔性と回答のしやすさ-; 川崎医療福祉学会誌, Vol.25, No.1, pp.23-35(2005).

(2017年8月7日受付, 10月30日再受付)

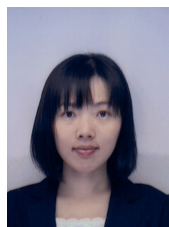
著者紹介

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。現在、同大学システム工学研究科システム工学専攻博士前期課程在学中。流言の拡散防止に関する研究に従事。

宮部 真衣



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システム工学専攻博士前期課程修了。2011年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後期課程修了。博士(工学)。現在、諏訪東京理科大学経営情報学部講師。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



2000年京都大学総合人間学部卒業。2002年同大学大学院情報学研究所修士課程修了。2005年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特任准教授。医療情報学, 自然言語処理の研究に従事。

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、和歌山大学システム工学部教授。CSCW, HCIの研究に従事。

