有害な流言は訂正されうるか？

－Twitterから収集した1年間の流言訂正情報の分析－

宮部　真衣†　　灘本　明代††　　荒牧　英治†,†††

† 京都大学学際融合教育研究推進センターデザイン学ユニット　〒600–8815 京都府京都市下京区中堂寺粟田町93 京都リサーチパーク6号館202号室

†† 甲南大学知能情報学部　〒658–8501 兵庫県神戸市東灘区岡本8-9-1

††† 科学技術振興機構さきがけ　〒102–0076 東京都千代田区五番町7　K's 五番町

E-mail: †{mai.miyabe,eiji.aramaki}@gmail.com, ††nadamoto@konan-u.ac.jp

あらまし　Twitterなどのマイクロブログの普及により，ユーザは様々な情報を瞬時に取得・拡散することができるようになった．一方，マイクロブログでは流言も拡散されやすい．流言は適切な情報共有を阻害し，場合によっては深刻な問題を引き起こす．特に災害時は，流言が救援活動などに悪影響を及ぼす可能性が高いため，流言の広がりにくい環境を作る必要がある．我々はこれまでに，流言拡散防止を目的とし，人間の発信した訂正情報に基づき流言情報を収集・提供するシステムを構築してきた．本稿では，1年間収集した流言情報の分析を行い，訂正情報をもとに収集可能な流言の特徴および流言の有害性と訂正情報の発信との関連について述べる．

キーワード　マイクロブログ，Twitter，流言，デマ，自然言語処理

1. はじめに

近年，Twitter(注1)などのマイクロブログが急速に普及している．主に自身の状況や雑記などを短い文章で投稿するマイクロブログは，ユーザの情報発信への敷居が低く，現在，マイクロブログを用いた情報発信が活発に行われている．2011年3月11日に発生した東日本大震災においては，緊急速報や救援物資要請など，リアルタイムに様々な情報を伝える重要な情報インフラの1つとして活用された [1–4]．マイクロブログは，重要な情報インフラとなっている一方で，情報漏洩や流言の拡散などの問題も抱えている．実際に，東日本大震災においても，様々な流言が拡散された [5]．

本稿では，マイクロブログの問題の1つである，流言に着目する．流言については，これまでに多くの研究が多方面からなされている [6–8]．流言と関連した概念として噂，風評，デマといった概念がある．これらの定義の違いについては諸説あり，文献毎にゆれているのが実情である．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．よって，最終的に正しい情報であっても，発言した当時に，十分な根拠がない場合は，流言とみなす．流言は適切な情報共有を阻害する．特に災害時には，流言が情報受信者を誤った行動に導き，様々な損失を与える場合がある．そのため，マイクロブログ上での流言の拡散への対策を検討していく必要があると考えられる．

では，なぜ人間は流言を拡散させるのであろうか．一般に，人々がある情報を他者に伝える場合，その情報が正しいと思って伝えていることが多く，本人がでたらめだと思う話を，悪意をもって他者に伝えることは少ない [8]．また，流言とは，曖昧な状況に巻き込まれた人々が，自分たちの知識や情報を寄せ集めることにより，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もある [9]．つまり，流言は何らかの役に立ち得る（有用性のある）情報を含み，それを共有するために善意で拡散されている可能性がある．流言の伝達は，主に伝達している情報が流言であることを認識していないことに起因すると考えられる．もしそうであるならば，人々に流言情報を提供することにより，流言の拡散を防止できる可能性があると考えられる．

これまでに我々は，人間によって投稿された流言の訂正（以降，訂正情報と呼ぶ）をもとに，自動的に流言情報を収集するサービス“ 流言情報クラウド ”を構築・運用してきた [10]．流言情報クラウドは，リアルタイムに訂正情報を収集し，そこから流言情報を抽出・提供することにより，流言拡散防止を支援するサービスである．

本稿では，流言情報クラウドにより1年間かけて収集してきた流言情報をもとに，次の2点について分析を行い，その結果を報告する．

（1） **訂正情報の発信傾向**：1年間にわたり，どの程度訂正情報が発信されるのかを分析する．

（2） **訂正される流言の有害性**：人間によって訂正される流言は，有害性の高いものなのかを分析する．

以下，2章において関連研究について述べる．3章では流言情報クラウドの概要を述べる．4章において，1年間収集した流言情報および分析手法について述べ，5章でそれらの分析結果

(注1)：http://twitter.com/

について考察する．最後に 6 章で本稿の結論についてまとめる．

2. 関 連 研 究

本章では，まず，流言に関するこれまでの定義について述べた後，ソーシャルメディアにおける異常状態の検出に関する研究について述べる．

2.1 流言の定義と流言の伝達

流言の分類としては，ナップによる第 2 次世界大戦時の流言の分類がある [6]．ナップは，流言を「恐怖流言（不安や恐れの投影）」「願望流言（願望の投影）」「分裂流言（憎しみや反感の投影）」の 3 つに分類している．また，これらの流言がどの程度の割合で流通するかは社会状況によって決まると述べられている．社会状況は流言を伝達させる要因の 1 つであり，例えば震災の直後など，社会状況が多くの人々に不安を感じさせる状況は，流言の発生や伝達に関係する．

また，流言の伝達には，曖昧さ，重要さ，不安という 3 つの要因が強く関係することが示されている [8]．オルポートとポストマンは，流言の流布量を，$R \sim i \times a$ のように定式化し，「流言の流布量（R）は，重要さ（i）と曖昧さ（a）の積に比例する」と述べている [7]．これらの流言に関する先行研究は，現実社会の中での口伝えでの流言の伝達について行われたものである．

近年では，ソーシャルメディア上の流言を扱った研究も行われている．Mendoza らは，2010 年のチリ地震における Twitter ユーザの行動について分析を行っている [11]．この研究では，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，支持ツイート，否定ツイートの数について，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は少ないが（0.3%），流言を否定するツイートは約 50%に上ることを示している．

本研究では，マイクロブログを対象とし，これらの先行研究での知見とマイクロブログの特徴に基づき，流言拡散防止サービスを構築する．

2.2 異常状態の検出

流言の拡散を防止するためには，ある時点で拡散されている流言を検出する必要がある．流言が拡散されている場合，ある流言を含む情報の頻度が急激に高まる可能性があり，流言の拡散されている状態は，異常状態の 1 種と見なせる．そこで，本節では，異常状態の検出に関する研究について述べる．

Twitter をセンサとして捉え，災害などの異常事態の検出を試みた研究がある．Sakaki らは，Twitter を用いた地震や台風の位置の推定に関する研究を行っている [12]．Abel らは，緊急放送システムをモニタリングしておき，災害の発生を確認した後，Twitter から災害に関連するツイートを収集し，有益な情報をユーザに提供するシステムの開発を行っている [13]．Aramaki らや Paul らは，Twitter を用いてインフルエンザの把握を行っている [14, 15]．これらは，平常時からソーシャルメディアなどを監視しておくことで，異常事態発生時にいち早くその情報を伝えるという警告型のサービスである．

また，流言の検出を試みた研究も行われている．Qazvinian らは，マイクロブログ（Twitter）における特定の流言に関する情報を網羅的に取得することを目的とし，流言に関連するツイートを識別する手法を提案している [16]．実験の結果，ある流言に関連するツイートを高精度に識別可能であることを示しているが，課題として新しく発生した流言の検出が挙げられている．また，Rattanaxay らは，「らしい」といった，流言に含まれる曖昧な表現に着目した流言情報の検出手法を提案している [17]．しかし，Rattanaxay らの手法では，曖昧な表現を含む情報はすべて流言と見なしており，たとえ正確な情報であっても，曖昧な表現を含むものは流言として検出してしまう．

本研究では，誤った情報および根拠がない情報を検出するために，人間によって発信される訂正情報に着目した，流言拡散防止サービスを提案する．

3. 流言情報クラウド：訂正情報に基づく流言拡散防止サービス

1 章で述べたように，情報が誤っていることをユーザに提示できれば，誤った情報の拡散を防ぐことができる可能性がある．2011 年の東日本大震災発生後には，Web 上で広がった流言について，ブログなどでまとめ記事が作成されるなど，流言の拡散を防止するための活動が行われた(注2)(注3)．しかし，これらのブログ上の情報は人手によりまとめられており，発生した流言をリアルタイムに反映することは容易ではない．

そこで，本研究では流言拡散防止を支援するサービスとして，流言情報クラウドを構築した．以降の節において，本研究における流言情報収集のアプローチ，および流言情報クラウドのシステム構成について述べる．

3.1 本研究における流言情報収集のアプローチ

本節では，流言情報を収集するためのアプローチについて述べる．

流言情報を蓄積するためには，ある情報に流言が含まれているかを判定する必要がある．しかし，人間が信じてしまうような流言を自動的に流言だと判定することは極めて難しい．また，その時点では情報の真偽を判断できず，後になって真偽がわかることも多い．さらに，流言の内容は多様であり，既知の流言情報を用いて判定しても，正しく抽出することは容易ではないと考えられる．

これまでにマイクロブログ上の流言を分析してきた結果 [18]，マイクロブログ上では，話題によって発信数は異なるものの，流言に対して訂正ツイートや疑問ツイートが発信されており，さらに，流言への対処として，「デマ」「ガセ」といったある種の固定表現（本稿では以下，**流言マーカー**と呼ぶ）を含む形で訂正情報が発信されていることがわかっている．つまり，内容が多様であり，情報の形式に統一性のない流言自体を特定するよりも，流言を否定している訂正情報を特定する方が容易である可能性がある．

(注2)：荻上式 BLOG「東北地方太平洋沖地震，ネット上でのデマまとめ」：http://d.hatena.ne.jp/seijotcp/20110312/p1

(注3)：ついのすみか「東北地方太平洋沖地震のデマ情報まとめ」：http://tsuinosumika.iku4.com/Entry/67/

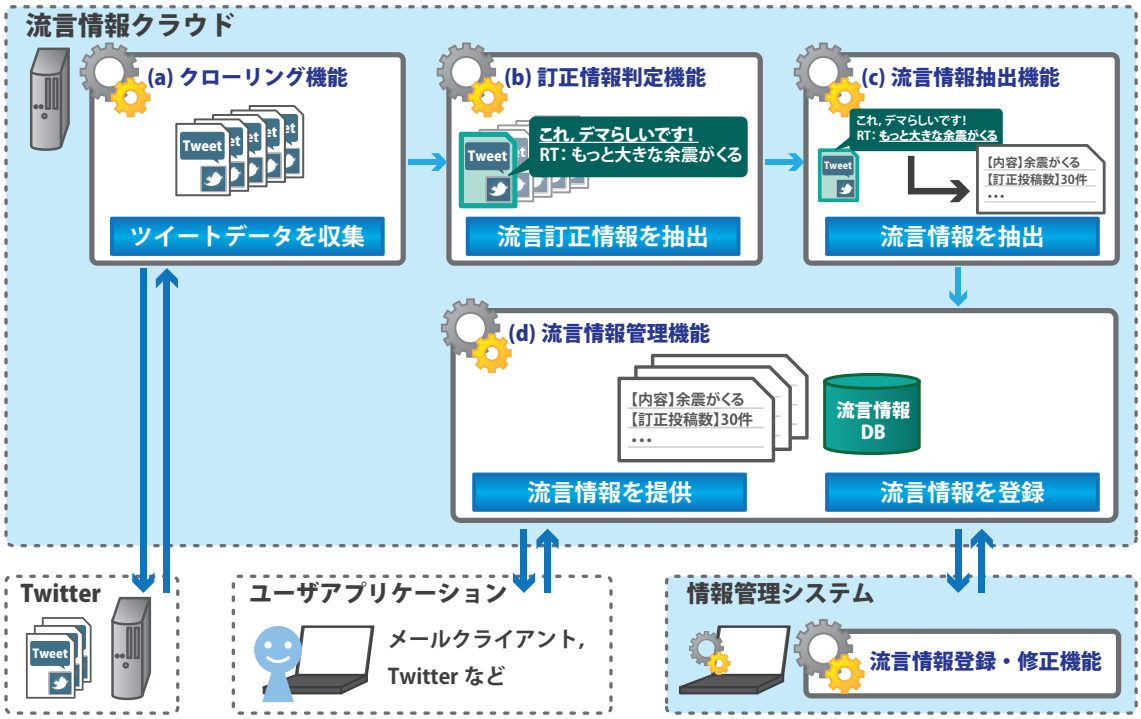


図 1　流言情報クラウドのサービス構成

表 1　訂正キーワード

デマ，嘘，ツリ，釣り，偽情報，ガセ，ネタ，誤報，都市伝説，ウソ，狂言，迷信，間違い，いたずら，チェーンメール

そこで，本研究では，流言情報を直接収集（ある情報が流言かどうかを判定）するのではなく，**訂正情報を抽出することにより，間接的に流言情報を収集する．**

3.2　流言情報クラウドの構成

流言情報クラウドのサービス構成を図 1 に示す．流言情報クラウドは，以下の 4 つの機能により構成される．

(a)　クローリング機能：Twitter から訂正キーワード（表 1）を含むテキストを収集する

(b)　訂正情報判定機能：SVM を用いた訂正情報分類器を用いて，テキストが訂正情報かどうかを判定する

(c)　流言情報抽出機能：訂正情報の中に含まれる流言情報を抽出する

(d)　流言情報管理機能：流言情報データベースの管理（検索，登録，修正）を行う

情報蓄積には機能 (a)〜(d) を，拡散防止には機能 (b)〜(d) を利用する．次節において，情報蓄積における流れを詳細に説明する．

3.3　情報蓄積の流れ

基本的には自動で流言情報を蓄積し，人手を介すことなく情報提供を可能にする．また，人手による精査も可能とすることにより，提供する情報の信頼性を向上できるようにする．

情報蓄積のフェーズでは，流言マーカーをもとに訂正情報を抽出することにより，間接的に流言情報を収集する．本研究における訂正情報の定義を表 2 に示す．不正確さを含む記述が含まれていた場合，不正確さに関する確信度の高さに関わらず，訂正情報と判定することとする．

一方，流言マーカーを含むものが，必ずしも訂正情報であるとは限らない．例えば，「デマゴギーって何？デマの省略前の言葉？」というツイートには，「デマ」という流言マーカーが含まれるが，流言の訂正情報ではない．

そこで，以下の手順により訂正情報の判別，流言情報の抽出を行い，流言情報を蓄積する．

（1）　流言マーカーを含むツイート群を収集する．

（2）　訂正情報分類器を用いて，収集したツイート群から訂正情報を抽出する．

（3）　手順 (2) で抽出した訂正情報から，“「〜」というデマ”のような，流言部分が明示的な訂正情報については，パターンマッチング(注4)により訂正情報に含まれる流言部分を抽出し，「登録日時」「ID」「流言訂正情報」「流言内容」「訂正ツイート数」を流言情報管理機能により蓄積する．本システムで使用しているパターンの例を表 3 に示す．

（4）　手順 (3) で流言部分を特定できなかった場合は，抽出した訂正情報をクラスタリングツール(注5)を用いて分類し，各分類結果について，最も所属度の高い訂正情報を代表として「登録日時」「ID」「流言訂正情報」「訂正ツイート数」を流言情報管理機能により蓄積する．なお，手順 (4) では「流言内容」は登録しない．

一方，上記の流れで自動的に蓄積された流言訂正情報には，誤って訂正情報と判定されたものや，正しく訂正情報と判定されていないものが含まれる可能性がある．そこで，人手による

(注4)：パターンマッチングでは，“「〜」というデマ”，“『〜』っていうデマ”など，かぎ括弧や特定するためのフレーズを組み合わせた 27 種類のパターンを利用している．

(注5)：http://code.google.com/p/bayon/

表 2　訂正情報の定義とその例

判定条件	該当例
a) ある情報に関する不正確さの記述が主題である	このツイートはデマです。 RT xxx: ○○○ **（確信度：高）** ○○○は本当なの？デマじゃないの？ **（確信度：中）** ○○○が、デマだとしても、備えあれば憂いなし。 **（確信度：低）**
b) ある情報に関する不正確さの記述が含まれるが，主題ではない	○○○というデマを広げた人間がいるみたいだね。
c) 流言に関してまとめたサイトを紹介している	地震に関するデマ http:// …

表 3　流言抽出のパターン

	PTN1	PTN2
<PTN1> などという*<PTN2>	「*」 『*』 “ * ”	訂正キーワード（表 1）
<PTN1> という*<PTN2>		
<PTN1>っていう*<PTN2>		
<PTN1>って*<PTN2>		
<PTN1>っていう話は <PTN2>		
<PTN1> は*<PTN2>		
<PTN1> という根拠のない		
<PTN1> という根拠薄弱な		
<PTN1> といった根拠なし		

<PTN1> <PTN2> には該当する文字列が入る．
＊はワイルドカードを示す．

流言情報の精査を可能にし，提供する情報の信頼性を向上できるようにする．人手で登録または修正が行われた流言情報については，流言情報としての信頼度を高く設定することにより，精査が行われていない情報との区別ができるようにしている．

4.　1 年間収集した流言訂正データの分析

本章では，収集した流言訂正データの分析方法について述べる．

流言の拡散防止において特に重要となるのは，情報受信者にとって有害性のある情報に対する拡散防止である．流言情報クラウドでは，人間によって訂正された流言情報を収集する．では，人間によって訂正される流言情報とは，有害性の高いものなのであろうか？本稿では，流言情報クラウドによって収集された訂正情報に関する傾向（どの程度発信されるのか，発信数はどのように推移するのか）と併せて，それらの訂正情報から抽出された流言情報の有害性についても分析を行い，今後の流言拡散防止における指針を検討する．

以降の節において，分析対象となるデータ，および流言内容の有害性評価手法について述べる．

4.1　対象データセット

本分析では，流言情報クラウド上で流言情報の収集を開始した 2012 年 6 月 22 日から 1 年間のデータを用いることとした(注6)．

まず，2012 年 6 月 22 日〜2013 年 6 月 21 日までに訂正キーワードをもとに収集・抽出した訂正情報（本システムのデータ収集源は Twitter であるため，以降「訂正ツイート」と表記する）から，3.3 節で述べた手順により，人間によって訂正された流言情報を抽出した．なお，この手順では同じ流言の異なる表現のバリエーション（表 4）も抽出されうる．同一の流言を指し示すと考えられる異なる表現の流言は，1 日毎あるいは 1 年全体で 1 つにまとめて分析を行う．

(注6)：なお，収集期間中には Twitter API の仕様変更によって，データの収集ができなかった日が存在する．

表 4　同じ流言に対する表現のバリエーションの例

- iPhone を電子レンジでチンすると直ぐに充電ができる
- iPhone を電子レンジでチンすると急速充電できる
- iPhone を電子レンジでチンすると充電できる
- iPhone が電子レンジで充電できる

4.2　流言内容の有害性に関する主観評価

災害時の流言拡散において，実際的に問題となるのは，その流言が実際に流言（虚偽の情報）であった場合，どれくらい有害であるかという点である．

そこで，流言情報の有害性に関して，以下の 2 項目の主観評価を実施した．

有害性 A　この情報が間違っている場合，自分 にとって有害である．

有害性 B　この情報が間違っている場合，自分以外の他者 にとって有害である．

各項目の評価には，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）を用いることとし，5 名の評価者が評価作業を行った．また，評価者が流言情報を見ても，その文だけでは意味が理解できないと判断した場合，評価不能（-1）と評価してもらうこととした．

5.　分析結果と考察

本章では，1 年間収集した流言訂正ツイートの分布と推移および流言の有害性評価結果を示した後，有害性の高さと訂正ツイートの発信との関連について議論する．

5.1　流言の訂正ツイート数

まず，1 年全体で訂正された流言の種類数を調査した．調査の結果，本収集期間においては，2953 種類の流言に対する訂正ツイート（計 39922 件）が投稿されていた．各流言に対し，それぞれどの程度の訂正ツイートが投稿されたのかの分布を図 2 に示す．

図 2 より，訂正ツイートが 1 件のみしか投稿されない流言が多く，全体の 62%を占めている．一方，10 件以上訂正ツイートが投稿された流言は全体の 10%程度（288 種類）であった．

次に，1 日あたりの訂正ツイート数および訂正された流言の種類数を調査した．表 5 に 1 日あたりの訂正ツイート数およ

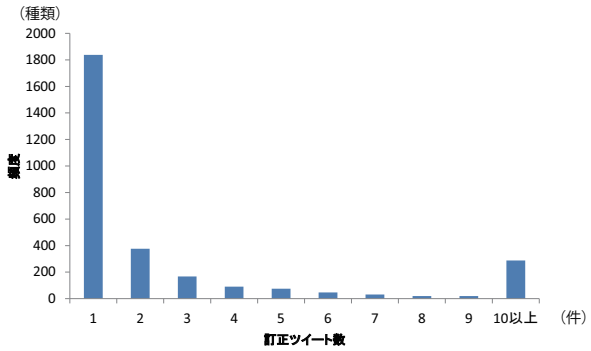
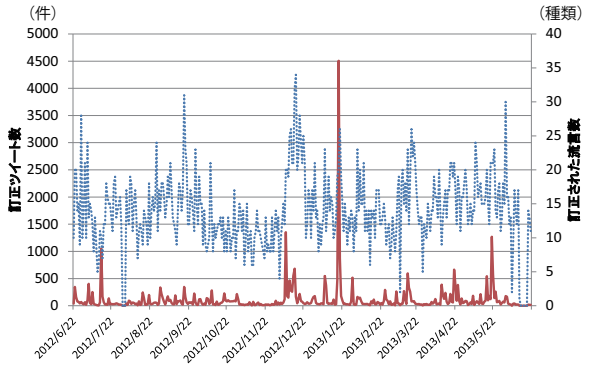


図 2　訂正ツイート数の分布

表 5　1 日あたりの訂正ツイート数，訂正された流言の種類数

	訂正ツイート数（件）	訂正された流言の種類数（種類）
平均	109.4	14.0
標準偏差	278.9	5.3
最大	4507	34
最小	0	0



実線が訂正ツイート数，点線が訂正された流言の種類数の推移を示す．

図 3　訂正ツイート数，訂正された流言の種類数の推移

び訂正された流言の種類数を示す．また，図 3 に訂正ツイート数，訂正された流言の種類数の 1 年間の推移を示す．1 日あたりに訂正される流言の種類数は，平均 14 種類，最大 34 種類であった．一方，1 日あたりの訂正ツイート数は平均 109 件，最大 4507 件であった．

図 3 より訂正ツイート数の急激な増加がみられる部分があるが，この際，必ずしも訂正される流言の種類数の増加は見られない．これは，多数のユーザによる特定の流言に対する訂正ツイートが投稿されたことにより，訂正ツイート数のみが急激に増加している．例えば，図 3 の 2013 年 1 月 19 日を見ると，流言の種類数自体は 20 種類と大きな増加は見られないが，「iPhone を電子レンジでチンすると直ぐに充電ができる」という特定の流言に対する訂正 4477 件が発生したことにより，訂正ツイート数の急増が見られた．他にも，訂正ツイート数が 1000 件を超えている日が 3 日（2012 年 7 月 14 日，12 月 8 日，2013 年 5 月 21 日）確認できた．これらはいずれも特定の流言(注7)に

(注7)：2012 年 7 月 14 日には「トトロのメイちゃんは『死んでいる』」という流言，12 月 8 日には「地震でガレキの中に閉じ込められた、助けて」という流言，2013 年 5 月 21 日には「iPhone を水に浸すと音質が良くなる」という流言がそれぞれ増加している．なお，7 月 14 日，12 月 8 日の流言訂正の急増には，それぞれ前日の出来事（7 月 13 日：テレビでの「となりのトトロ」の放送，12 月 7 日：震度 5 弱の地震の発生）が影響したと考えられる．

表 6　自分／他者にとっての有害性評価結果

		有害性 B					
		1	2	3	4	5	計
有害性 A	1	196	72	60	53	0	381
	2	0	49	67	188	12	316
	3	0	0	8	54	7	69
	4	0	0	0	65	28	93
	5	0	0	0	0	15	15
	計	196	121	135	360	62	874

有害性 A：自分 にとっての有害性
有害性 B：自分以外の他者 にとっての有害性

対する訂正が急激に増加したため，流言の種類数の急増を伴わずに訂正ツイート数のみが増加している．

5.2　流言の有害性

本稿では，4.2 節で述べた有害性評価の結果，評価者 5 名の内 1 名でも評価不能と判断したものは，有害性の分析対象から除外することとした．確認の結果，1 年全体で訂正された流言 2953 種類のうち，2079 種類が評価不能と判断された(注8)ため，それらを除外した 874 種類の流言情報を分析対象とする．なお，有害性の評価結果については，5 名の評価者による評価結果の中央値を用いて分析する．

表 6 に有害性評価の結果を示す．自分にとっての有害性（有害性 A）の評価結果を見ると，有害性が低い（評価値 1，2）と判断されたものが全体の 79.7%（697 種類）を占め，有害性が高い（評価値 4，5）と判断されたものは 12.4%（108 種類）にとどまっている．一方，他者にとっての有害性（有害性 B）の評価結果を見ると，有害性が低いと判断されたものが全体の 36.3%（317 種類），有害性が高いと判断されたものが 48.3%（422 種類）となっており，有害性が低いと判断された流言の割合を上回っている．表 6 を見ると，他者にとっての有害性（有害性 B）が，自分にとっての有害性（有害性 A）よりも低く評価される（例えば，有害性 A の評価値が 4 の場合に，有害性 B の評価値が 4 未満になる）ケースがないことがわかる．つまり，流言の有害性を評価する際，人間は，他者にとっての有害性を自分にとっての有害性よりも高く見積もる可能性があることが示唆される．

5.3　有害性と訂正ツイートの発信の関連

1 章で述べたように，人がある情報を他者に伝える場合，悪意をもって他者に伝えることは少なく，善意で情報が伝搬されている可能性がある．そうであるならば，有害性の高いものほど，善意に基づき訂正数が増えるのではないだろうか？そこで本節では，有害性と訂正ツイートの発信数との関連について述

(注8)：評価不能と判断された流言の例としては，「日本による組織的な強制連行」という流言情報がある．この情報のみでは誰が連行されたのか不明であり，有害性が評価できないため，評価不能と判断されている．このような情報の一部分が不明確な流言や，理解には専門知識を要する語句が含まれる流言が評価不能と判断されている．

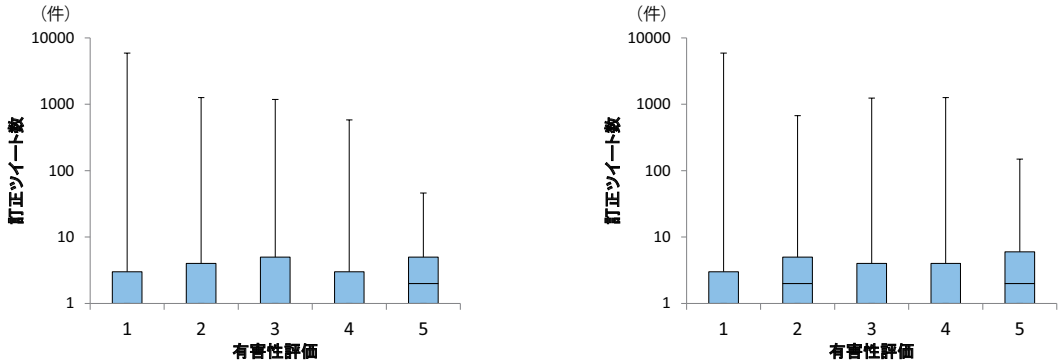


図 4　有害性評価結果と訂正ツイート数の分布

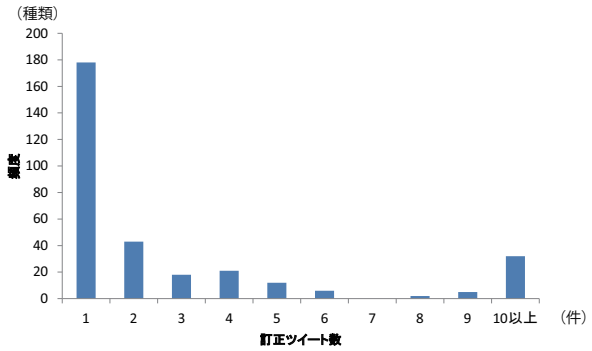


図 5　有害性 B が低い（評価値 1，2）流言の訂正ツイート数の分布

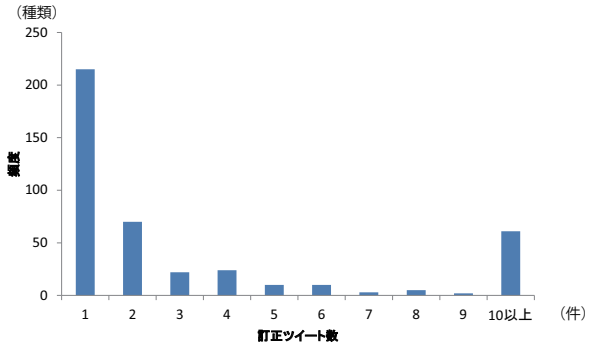


図 6　有害性 B が高い（評価値 4，5）流言の訂正ツイート数の分布

べる．

有害性評価結果と訂正ツイート数の分布を図 4 に示す．図 4 より，自分にとっての有害性，他者にとっての有害性のいずれについても，各有害性評価値における訂正ツイート数は分散していることがわかる．2 種類の有害性評価結果（自分にとっての有害性および他者にとっての有害性）と，訂正ツイート数との間の順位相関係数を確認した結果，訂正ツイート数と自分にとっての有害性の順位相関係数は 0.047，訂正ツイート数と他者にとっての有害性の順位相関係数は 0.074 であり，いずれも相関は見られなかった．

図 5 および図 6 に，有害性 B（他者にとっての有害性）が低い流言および高い流言の訂正ツイート数の分布を示す．図 5，6 を見ると，いずれも図 2 と同様に，訂正ツイート数の少ない流言が大部分を占めたロングテールとなっている．つまり，訂正ツイートの多さは，必ずしも有害性の高さを表すものではないと考えられる．

表 7 に一部の流言に関する訂正ツイート数と有害性評価結果の事例を示す．表 7 の 1〜6 は他者にとっての有害性（有害性 B）が高いと評価された流言である．「ほくろの毛をいじると癌化する」や「コントレックスは被曝を防げる」は，いずれも人間の行動に影響を与えうる情報であり，有害性が高いと判断されているものの，後者と比較して前者の訂正ツイート数は少ない．Twitter のようなマイクロブログにおいては，ユーザが受信する情報はフォロー関係に基づいており，拡散力のないユーザが訂正情報を発信した場合，そのユーザと繋がった，限られたソーシャルネットワーク内のみに伝搬するだけで，その訂正情報は十分に拡散されない可能性がある．また，表 7 の 7〜10 は有害性 B が低いと評価された流言であるが，「mixi 利用者が 97%から 2%に激減」という流言は，有害性が低いと評価されているにもかかわらず，多くの訂正ツイートが発信されている．有害性が高くなくとも，情報受信者の興味を引く内容であれば拡散される可能性がある．また，前述したように，その情報が拡散されているソーシャルネットワークが影響している可能性もある．

これらのことから，訂正ツイートの多さは，その流言の有害性だけでなく，情報受信者にとっての内容の興味深さや，その情報が伝搬するネットワークなども影響すると考えられる．そのため，有害性の判断指標として，訂正ツイートの多さを単純に利用することは困難である．

5.4　今後の課題

流言に対する注意を人に促す際には，有害性の高いものを抽出し，優先して提示する仕組みが必要となる．これまでの研究において，修辞ユニット分析を用いた分析の結果，「〜してください」のような，情報受信者の行動に影響を与えうる表現を含む情報は，震災時に高い有用性と有害性を持つことが明らかになっている [19]．今回の分析では，訂正ツイート数と有害性には相関が見られていないことから，単純に訂正ツイートの多さを有害性の判断指標として用いることはできないと考えられる．今後，有害性の高い流言を抽出する際には，訂正情報の投稿の多さではなく，流言内容の解析を利用した有害性判定を実装す

表 7　流言の訂正ツイート数と有害性評価結果の一部

	流言	訂正ツイート数	有害性 A	有害性 B
1	デキるビジネスパーソンは敬称に「うじ」を使う!	1	1	4
2	ほくろの毛をいじると癌化する	1	4	4
3	mixi が有料会員制をはじめるらしいぞ!	1	4	5
4	コントレックスは被曝を防げる	30	3	5
5	福島の子は子供を産めない	49	4	4
6	『機動戦士ガンダム THE ORIGIN』2014 年 TV アニメ化	583	4	4
7	バンクーバーには日本人がたくさんいる	1	1	1
8	PC の画面を割ると中から二次元のキャラが飛び出してきます	1	2	2
9	緊張を解くためには手のひらに『人』という字を 3 回書いて舐めるといい	415	1	1
10	mixi 利用者が 97%から 2%に激減	675	2	2

有害性 A：自分 にとっての有害性
有害性 B：自分以外の他者 にとっての有害性

る必要がある．

6. おわりに

本稿では，人間による流言訂正情報に基づき自動的に流言情報を収集するサービス“ 流言情報クラウド ”によって 1 年間かけて流言情報を収集し，Twitter 上で訂正される流言情報の推移や有害性に関する分析を行った．

1 年分の訂正情報および流言情報の分析により，以下のことを明らかにした．

流言訂正のロングテール　1 年間に訂正された流言の大部分は少数のユーザのみが訂正したものであり，10 件以上の訂正ツイートが投稿された流言は 10%程度であることが観察された．

訂正数増加要因の多様性　多くの人間によって訂正ツイートの発信される流言が，必ずしも有害性の高いものであるとは言えない．訂正ツイート数の増加には，流言内容の有害性だけでなく，その内容の興味深さや発信者の情報拡散能力など，多様な要素が関わっていると考えられる．

ただし，本稿で示した結果は，流言部分が明示的に示された訂正情報の分析により得られたものである．マイクロブログ上では，流言部分が明示的でない訂正情報も投稿されているため，今後，それらも併せた分析を行う必要があると考えられる．また，訂正ツイート数と実際に拡散された流言ツイート数の関係については未だ明らかではないため，今後は，流言情報自体の投稿傾向と併せて分析することにより，効率的な流言拡散防止手法の検討を行う．

謝　辞

本研究は，JST 戦略的創造研究推進事業の助成を受けた．

文　献

[1] 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.
[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] 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.
[4] 宮部真衣, 荒牧英治, 三浦麻子. 東日本大震災における twitter の利用傾向の分析. 情報処理学会研究報告. GN, [グループウェアとネットワークサービス], Vol. 2011, No. 17, pp. 1–7, 2011.
[5] 荻上チキ. 検証東日本大震災の流言・デマ. 光文社, 2011.
[6] Robert Knapp. Rumor clinic. Technical report, 1944.
[7] G.W. オルポート, L. ポストマン. デマの心理学. 岩波書店, 2008.
[8] うわさが走る: 情報伝播の社会心理. うわさが走る: 情報伝播の社会心理. サイエンス社, 1997.
[9] 佐藤健二. 関東大震災後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.
[10] 宮部真衣, 梅島彩奈, 荒牧英治, 灘本明代. 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. マルチメディア, 分散, 協調とモバイル（DICOMO2012）シンポジウム, pp. 1442–1449, 2012.
[11] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. Twitter under crisis: can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pp. 71–79. ACM, 2010.
[12] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pp. 851–860, New York, NY, USA, 2010. ACM.
[13] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. Twitcident: Fighting Fire with Information from Social Web Stream. In *International Conference on Hypertext and Social Media, Milwaukee, USA*. ACM, 2012.
[14] Eiji Aramaki, Sachiko Maskawa, and Mizuki Morita. Twitter catches the flu: Detecting influenza epidemics using twitter. In *EMNLP*, pp. 1568–1576, 2011.
[15] Michael J. Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In Lada A. Adamic, Ricardo A. Baeza-Yates, and Scott Counts, editors, *ICWSM*. The AAAI Press, 2011.
[16] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. Rumor has it: identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pp. 1589–1599, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
[17] Keothammavong Rattanaxay, 相田慎, 青野雅樹. ツィッターのデマ率の推定. 情報処理学会第 74 回全国大会，第 2 分冊, pp. 523–524, 2011.
[18] 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. 情報処理学会論文誌, Vol. 54, No. 1, pp. 223–236, jan 2013.
[19] 宮部真衣, 田中弥生, 西畑祥, 灘本明代, 荒牧英治. マイクロブログにおける流言の影響の分析. 自然言語処理, Vol. 20, No. 3, pp. 485–511, 2013.