

流言訂正情報に基づいた流言情報クラウドの提案

宮部 真衣^{†1} 梅島 彩奈^{†2}
灘本 明代^{†2} 荒牧 英治^{†1,†3}

近年、マイクロブログの普及に伴い、マイクロブログを用いた個人での情報発信が増加している。2011年3月11日に発生した東日本大震災においては、Twitterに多くの情報が投稿された。重要な情報が伝搬された一方で、様々な流言の拡散も多数行われた。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。本研究では、流言拡散を防ぐための仕組みとして、流言訂正情報に基づいた流言情報クラウドを提案する。

Proposal of Rumor Information Cloud based on Rumor-Correction Information

MAI MIYABE,^{†1} AYANA UMEJIMA,^{†2} AKIYO NADAMOTO^{†2}
and EIJI ARAMAKI^{†1,†3}

Recently, conveying information from individuals has increased due to the development of microblog. After the Great Eastern Japan Earthquake in Japan 2011, numerous tweets were exchanged on Twitter. Twitter provided a lot of important information after the earthquake. On the other hand, various false rumors were spread on Twitter. False rumors negatively affect important activities such as relief activities. Therefore, it is required to structure the environment that prevents people from spreading false-rumors. In this paper, we propose rumor information cloud based on rumor-correction information.

1. はじめに

近年、Facebook^{*1}やTwitter^{*2}などのマイクロブログが急速に普及し、ユーザによるマイクロブログを用いた情報発信が活発化している。特にTwitterは、140文字という制限によりユーザの情報発信への敷居が大きく下がっており¹⁾、2011年3月11日に発生した東日本大震災においては、リアルタイムに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^{2),3)}。しかし、安否情報などの重要な情報の共有・伝搬が行われた一方で、多くの流言も拡散されるという問題も生じた⁴⁾。流言は適切な情報共有を阻害する要因となり、悪影響を及ぼす。特に災害発生時には、流言が救命のための機会損失を生む場合もあるため、流言の広がりにくい環境を作る必要がある⁴⁾。

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ふ話を他者に伝えることは少ない⁵⁾。つまり、流言の伝達過程において、人々は伝達している情報が流言であると認識していない場合が多いと考えられる。人々に流言情報を提供し、他者に伝達しようとしている情報が流言であることを認識させることにより、流言の拡散を防止できる可能性があると考えられる。

流言情報を提供するためには、実世界で伝達されている情報の中から流言情報を検出し、蓄積しておく必要がある。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。そこで、本研究では、流言の判定は人間にまかせ、流言を流言だと判定できた人間の声を自動的に抽出することを目的とする。実際に東日本大震災時に広まった流言については、疑問を投げかけたり、内容が誤りであることを指摘したりするユーザも存在した。本研究では、このような流言の不正確さに関する発言（以下、流言訂正情報と呼ぶ）に着目する。

そこで我々は、流言拡散を防ぐための仕組みとして、流言訂正情報に基づいた流言情報クラウドを提案する。

^{†1} 東京大学知の構造化センター
Center for Knowledge Structuring, The University of Tokyo

^{†2} 甲南大学知能情報学部
Faculty of Intelligence and Informatics, Konan University

^{†3} 科学技術振興機構 さきがけ
PRESTO, Japan Science and Technology Agency(JST)

*1 <http://www.facebook.com/>

*2 <http://twitter.com/>

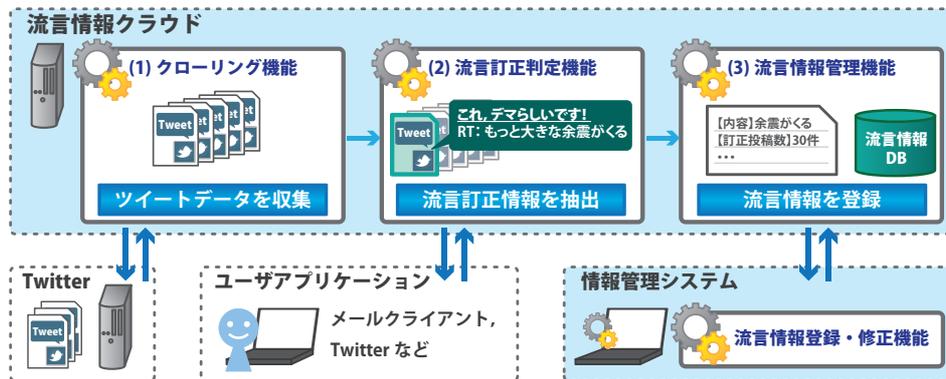


図 1 流言情報クラウドのサービス構成
Fig. 1 Service configuration of rumor information cloud.

流言情報クラウドは、リアルタイムに流言情報を蓄積し、その情報を提供することにより、ユーザの流言拡散を防止する。本稿では、流言情報クラウドの概要および流言情報クラウドを構成する機能の1つである流言訂正情報分類器について述べる。

2. 流言訂正情報に基づく流言情報クラウド

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。実際に、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、流言の拡散防止のための活動が行われた^{*1*2*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と考えられる。

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、流言の内容は多様であり、判定は容易ではない。一方、情報に誤りがある場合、流言訂正情報が発信されていることが多い。流言訂正情報には、訂正に関わるキーワードが含まれる場合が多く、比較的検出が容易であると考えられる。流言訂正情報を検出可能にすることにより、様々な流言情報の自動蓄積ができる可能性がある。

そこで、本研究では流言訂正情報に基づいた流言情報クラウドを提案する。流言情報クラウドの機能は大別して2つある。まず、あらかじめ流言訂正情報からリアルタイムに流言情報を蓄積する(情報蓄積フェー

ズ)。次に、蓄積した情報をユーザに提供し、流言の拡散を防ぐ(拡散防止フェーズ)。

流言情報クラウドのサービス構成を図1に示す。提案サービスは、以下の3つの機能により構成される。

- (1) クローリング機能 Twitterからテキストを収集する
- (2) 流言訂正判定機能 我々の提案する流言訂正情報分類器を用いて、テキストが流言訂正情報かどうかを判定する
- (3) 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

2.1 情報蓄積フェーズ

基本的には自動で流言情報を蓄積し、人手を介さず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。

これまでに、東日本大震災におけるTwitterの流言に関する分析を行った⁶⁾。分析の結果、流言訂正情報には、「デマ」「嘘」などの訂正に関わるキーワードが含まれる傾向がみられた。そこで、キーワードをもとにTwitterからツイートを自動収集し、流言情報を蓄積する。本研究では、検索キーワードとして「デマ」を用いる。

流言情報の自動蓄積は、以下の流れで行う。

- (i) クローリング機能により、Twitterから「デマ」を含むツイートを収集する。
- (ii) 流言訂正判定機能により、収集したツイートが流言訂正情報かどうかを判定する。流言訂正情報である場合、流言情報を登録する。

一方、上記の流れで自動的に蓄積された流言情報には、誤って流言と判定されたものや、正しく流言が抽出されていないものも含まれる可能性もある。そこで、人手による流言情報の精査を可能にし、提供する情報

*1 松永英明 公式サイト 絵文録ことのは「震災後のデマ 80件を分類整理して見えてきたパニック時の社会心理」:
<http://www.kotono8.com/2011/04/08dema.html>

*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>

*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://tsuinsumika.iku4.com/Entry/67/>

表 1 流言訂正情報の定義とその例
Table 1 Examples of rumor-correction information.

判定条件	該当例
(1) 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、デマだろうとしても、備えあれば憂いなし。 (確信度: 低)
(2) ある情報に関する不正確さの記述が含まれるが、それが主題ではない	というデマを広げた人間がいるみたいだね。
(3) 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://...

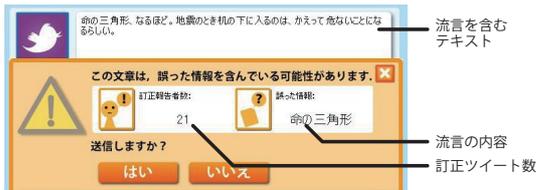


図 2 流言拡散防止のイメージ
Fig. 2 Image of prevention of rumor-spreading.

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する。

2.2 拡散防止フェーズ

蓄積した流言情報を用いて、Twitter や電子メールなどでの誤った情報の拡散を防止できるようにする。流言情報クラウドは、ユーザアプリケーションから受け取った入力テキストに関連する流言情報を提供する。情報提供の流れを以下に示す。

- (i) ユーザアプリケーションからテキストを受け取る。
- (ii) 流言訂正判定機能により、受け取ったテキストが流言訂正情報かどうかを判定する。
- (iii) 流言情報管理機能により、テキストに流言情報が含まれるかどうかを調べる。
- (iv) 流言訂正情報ではなく、流言情報が含まれる場合、ユーザアプリケーション上でユーザに警告を行う(図 2)。

3. 流言訂正情報分類器の構築

提案する流言情報クラウドにおいてその情報が流言訂正かどうかを判定するために、SVM を用いた流言訂正判定機能を提案する。

3.1 流言訂正情報の定義

本研究では、ある情報に対する不正確さを示す記述が含まれる発言を流言訂正情報として扱う。流言訂正情報と判断する条件を表 1 に示す。

今回の流言訂正情報分類器の構築においては、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、流言訂正情報と判定することとする。

3.2 コーパス

本研究では、以下の 2 種類のデータを用いる。
平常時データ: 2010 年 3 月のツイート
災害時データ: 2011 年 3 月の「地震」というキーワードを含むツイート

各データから「デマ」という表現を含むツイートを無作為にそれぞれ 1000 件抽出し、3.1 節で述べた判定条件に基づき流言訂正情報かどうかを手で判定し、コーパスとした。コーパスの一部を表 2 に示す。ツイートが流言訂正情報である場合は正例、そうでない場合は負例とした。平常時データにおける正例の数は 1000 件中 187 件、災害時データにおける正例の数は 1000 件中 602 件である。

3.3 流言訂正情報分類器

3.2 節で述べたコーパスを用いて、発言内容が流言訂正情報であるかを判定する分類器を構築した。今回は「デマ」の周辺文脈を素性とすることとした。SVM による学習には、多項カーネル ($d=2$) を用い、パラメータはデフォルト値を用いた^{*1}。

4. 実験

流言情報クラウドを実現するためには、まず流言訂正情報を適切に収集しなければならない。そこで、提案サービスの精度評価の第 1 段階として、本稿では流言訂正情報分類器の精度について評価する。

4.1 実験概要

流言訂正情報分類器の精度を検証するため、3.2 節で述べたデータを用いて以下の 2 つの実験を行う。

実験 1: 各データの 10 分割交差検定

実験 2: 一方のデータをトレーニングデータにした精度検証

なお、実験 1 では、2 種類のデータ(平常時、災害時)と併せて、2 つを統合したデータ(平常時+災害時)も用いた。

素性とする周辺文脈の適切な大きさを調査するために予備実験を行ったところ「デマ」という語の両側の周辺文脈のウィンドウサイズを 1~3 (形態素数^{*2}) と

*1 <http://chasen.org/~taku/software/TinySVM/>

*2 形態素解析には JUMAN を用いた。

表 2 流言訂正情報コーパス

Table 2 Examples of rumor-correction information corpus.

正例(+1) / 負例(-1)	ツイート
+1	千葉のコスモ石油、有害な雨が...の件、デマ確定です。拡散しないようご注意ください。【東北地方太平洋沖地震】コスモ石油、「有害物質が降る」メールに注意呼びかけ
+1	近畿の地震デマだったんだ～複雑けどよかった
+1	千葉の有害雨もプレート型による深夜の地震もデマか
-1	デマゴギーって何？デマの省略前の言葉？
-1	なにかデマ騒動があったのかな？
-1	明らかなデマであったなら、論外だけど、そうでないんだから、頭使えよ！、ってかんじだよな。

表 3 各データの 10 分割交差検定結果

Table 3 Result of experiments using 10-fold cross validation.

データ	ウィンドウサイズ	F 値
平常時	1	0.757
	2	0.736
	3	0.770
		0.716
災害時	1	0.868
	2	0.877
	3	0.888
		0.885
平常時+災害時	1	0.859
	2	0.853
	3	0.849
		0.835

表 4 各データによる判定精度

Table 4 Classification accuracy using another data for training data.

トレーニングデータ	テストデータ	ウィンドウサイズ	F 値
平常時	災害時	1	0.841
		2	0.808
		3	0.766
			0.682
災害時	平常時	1	0.667
		2	0.642
		3	0.643
			0.352

した場合^{*1}に、比較的精度のよい結果が得られた。そこで、本実験ではウィンドウサイズを 1~3 とした場合の精度について検証する。

4.2 実験結果

10 分割交差検定の結果を表 3 に示す。表 3 より、各データにおいて概ね良好な結果が得られた。

一方のデータをトレーニングデータにした場合の精度を表 4 に示す。表 4 より、平常時データをトレーニングデータとした災害時の流言訂正情報の判定は比較的高精度にできることがわかる。一方、災害時のデータに基づく平常時の流言訂正情報の判定精度は、平常

時のデータに基づく判定精度を下回った。

これらの結果から、災害時の流言訂正情報は平常時のデータをもとに判定可能であり、平常時を含めたデータを用いることで、平常時・災害時のどちらでも高精度な判定が期待できる。

5. おわりに

本研究では、流言拡散を防ぐための仕組みとして、流言訂正情報に基づいた流言情報クラウドを提案した。また、流言情報クラウドを構成する機能の 1 つである流言訂正情報分類器を構築した。平常時および災害時のデータを用いてコーパスを作成し、分類器の精度について検証を行った結果、災害時の流言訂正情報は平常時のデータをもとに判定可能であり、平常時を含めたデータを用いることで、平常時・災害時のどちらでも高精度な判定が期待できることを示した。

今後は、流言情報クラウドを実現するために、流言訂正判定機能以外の機能を構築する。また、実際に運用を行い、蓄積される流言情報の検証や流言情報の提供によるユーザへの影響、課題などを明らかにする。

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) 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) 荻上チキ：検証 東日本大震災の流言・デマ、光文社新書 (2011).
- 5) 川上善郎：うわさが走る 情報伝搬の社会心理、サイエンス社 (1997).
- 6) 梅島彩奈、宮部真衣、荒牧英治、灘本明代：災害時 Twitter におけるデマとデマ訂正 RT の傾向、情報処理学会研究報告、データベースシステム研究会, Vol.2011-DBS-152, No.4, pp.1-6 (2011).

*1 ツイート中に 2 回以上「デマ」が出現する場合、2 回目以降の周辺文脈は素性として用いないこととした。