

流言情報クラウド： 人間の発信した訂正情報の抽出による流言収集

宮部 真衣¹梅島 彩奈²灘本 明代²荒牧 英治^{1,3}¹ 東京大学知の構造化センター² 甲南大学知能情報学部³ 科学技術振興機構 さきがけ

mai.miyabe@gmail.com

1 はじめに

近年、Facebook¹やTwitter²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, 4]。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た[5]。流言は適切な情報共有を阻害し、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。

ここで、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、悪意をもって他者に伝えることは少ない[6]。つまり、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。そこで、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。

そこで本研究では、流言拡散を防ぐための仕組みとして、流言情報クラウドを提案する。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、流言拡散を防止する。本研究のポイントは、**流言が含まれる発言ではなく、流言の不確かさについて言及している発言（以下、訂正情報と呼ぶ）を抽出することにより、自動的な流言情報の収集を実現しようとしている点である**。本稿では、流言情報クラウドの概要を述べた後、流言情報を自動的に収集するための訂正情報分類器について述べる。

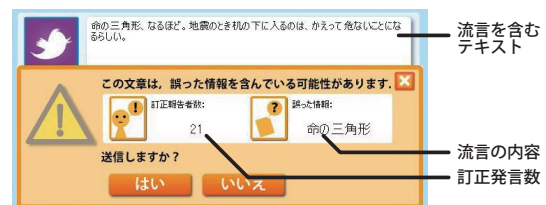


図1: 流言情報クラウドを用いたアプリケーションの例

2 流言情報クラウド

本章では、流言情報クラウドの概要および流言情報の収集手法について述べる。

2.1 概要

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。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^{3,4}。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、本研究では訂正情報に基づき、自動的に流言を収集・提供する仕組みとして、流言情報クラウドを提案する。流言情報クラウドを用いたアプリケーションのイメージを図1に示す。流言情報クラウドによって自動的に収集した情報は、図1のように、アプリケーション上でのユーザに対する警告などに利用できるようにする。

流言情報クラウドにおける流言情報蓄積の流れを図2に示す。流言情報クラウドでは、あらかじめ訂正情報からリアルタイムに流言情報を蓄積する。基本的

¹<http://www.facebook.com/>²<http://twitter.com/>³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>⁴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://tsuinomumika.iku4.com/Entry/67/>

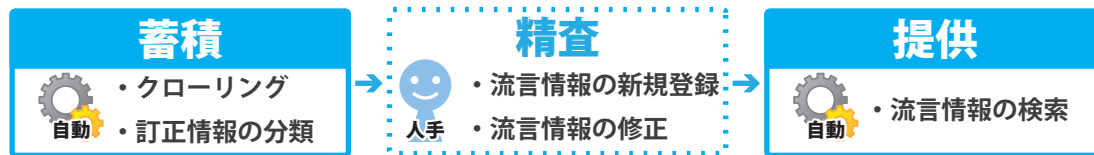


図 2: 流言情報クラウドにおける流言情報蓄積の流れ

表 1: 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、デマだろうとしても、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、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。
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://...

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可能であるが、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。また、蓄積した情報をユーザに提供することにより (図 1)、流言の拡散を防ぐ。

2.2 流言情報の収集手法

本節では、自動的に流言情報を収集する手法について述べる。

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。

一方、情報に誤りがある場合、人間によって誤りを指摘する発言 (訂正情報) が発信されていることが多い。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、比較的判定が容易であると考えられる。

そこで、本研究では、流言自体を抽出するのではなく、訂正情報を抽出することにより、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 1 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。

我々はこれまでに、東日本大震災における Twitter の流言に関する分析を行った [7]。分析の結果、訂正情報には、「デマ」や「間違い」といった、訂正していることを明示する用語 (以降、**流言マーカー**と呼ぶ) が含まれる場合が多いことがわかった。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?」という発言は、「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、以下の手順により流言情報を収集する。

1. 流言マーカーを含む発言を収集する。
2. 3章で述べる訂正情報分類器を用いて、収集した発言から訂正情報を抽出する。

3 訂正情報分類器の構築

本章では、流言情報クラウドの構成機能の 1 つである訂正情報分類器について述べる。

3.1 コーパス

本研究では、訂正情報を収集するためのリソースとして、Twitter を用いる。今回は、以下の 2 種類のデータを用いてコーパスを構築した。

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 [4]

今回は、流言マーカーを「デマ」とし、訂正情報の抽出を試みる。まず、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 件抽出した。次に、2.2 節で述べた判定条件に基づき訂正情報かどうかを人手で判定し、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。

3.2 訂正情報分類器と学習アルゴリズム

3.1 節で述べたコーパスを用いて、発言内容が訂正情報であるかを判定する分類器を構築した。

表 2: 訂正情報コーパス

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。拡散しないようにご注意ください。→【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～複雑だけどよかった
+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。

3.2.1 素性

今回は、素性として以下の項目を用いた。

- 流言マーカー「デマ」の周辺文脈
- 形態素数
- URLの有無
- 引用 (RT @)の有無

素性とする流言マーカーの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、流言マーカーの両側の周辺文脈のウィンドウサイズを1~3(形態素数⁵)とした場合⁶に、比較的精度のよい結果が得られた。そこで、これ以降はウィンドウサイズを1~3とし、精度検証を行う。

3.2.2 学習アルゴリズム

分類器構築に用いる学習アルゴリズムを選択するために、学習アルゴリズムの精度を比較した。今回は、図3に示す5種類のアロリズムを比較した⁷。

検証結果を図3に示す。図3より、SVMが最も高い精度を示した。そこで、以降の検証においては、SVMを用いることとする。

4 実験

構築した訂正情報分類器の精度を検証するため、3.1節で述べたデータを用いて2種類の実験を行う。それぞれの実験において、以下の内容を検証する。

実験1: 訂正情報分類器によって、訂正情報を判定できるか？

実験2: 平常時のデータを用いた分類器は、災害時にも性能を発揮できるか？

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。

⁵形態素解析にはJUMANを用いた。

⁶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。

⁷検証においては、TinySVMおよびデータマイニングツールであるWEKAを利用した、それぞれ、パラメータはデフォルト値を用いた。

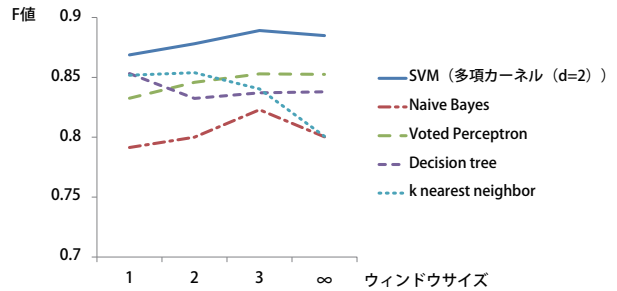


図 3: 各学習アルゴリズムによる判定精度

5 実験結果と考察

5.1 分類器による判定精度

訂正情報分類器によって、訂正情報を判定できるかどうか(実験1)を検証するために、各データを用いた10分割交差検定を行った。結果を図4に示す。図4より、各データにおいて概ね良好な結果が得られた。

次に、平常時のデータを用いた分類器が、災害時にも性能を発揮できるかどうか(実験2)を検証するために、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確認した。結果を図5に示す。図5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。

5.2 流言情報の判定精度との比較

本節では、既知の流言情報を用いた分類器による、異なる流言情報の判定精度を検証し、提案手法(訂正情報の判定)との比較を行う。流言情報として、東日本大震災発生後にTwitter上で拡散された6種類の流言情報(表3)を用いて、流言情報の分類器を構築した。なお、それぞれの流言情報に含まれる内容は異なるため、学習時の素性として、ツイートに含まれる

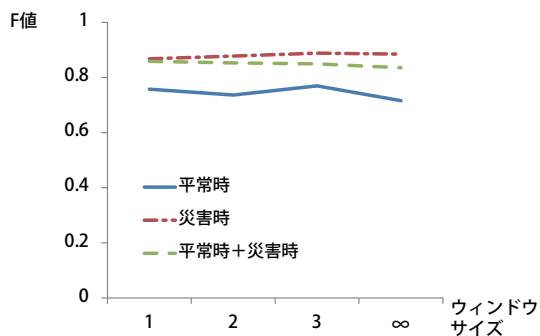


図 4: 各データの 10 分割交差検定結果

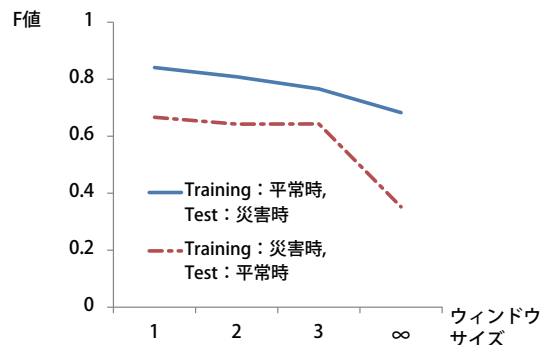


図 5: 各データによる判定精度

表 3: 流言情報の例

流言の種類	ツイート例	ツイート数
1	地震が起きた時、社内サーバールームにいたのだが、ラックが倒壊した。腹部を潰され、血が流れている。	21
2	命の三角形、ためになりました。地震では机の下にすぐ入る事を考えがちだけど、机等のすぐ横のほうがつぶされない三角形の空間になり、助かる可能性が高いんだね。	84
3	【関西地区の皆さん】関西電力が電力の提供を始めたようなので、コチラで節電すれば立派な支援になります。出来るかぎり節電を心がけましょう。	86
4	ワンピースの作者 尾田栄一郎氏、地震の被害者救済に 15 億円を寄付 「自分が幸せになったということは、世の中から受けたひとつの借りだ」	5
5	「菅さんはとてもラッキーな人」民主党の仙谷良人前・官房長官が東北大地震を「ラッキー」と表現。	15
6	近畿のプレートが小さくなっている模様。これが元に戻ろうとすれば次は近畿に大きな地震が起きる可能性が非常に大きいので明日、明後日は注意してください。	5

全ての形態素および 3.2.1 項で述べた素性（流言マーカの周辺文脈は除く）を用いた。

6 種類の流言情報のうち、5 種類を学習データとし、残りの流言を判定できるかどうかを検証した結果、F 値の平均は 0.283 となり、訂正情報を用いた場合の精度を大きく下回った。したがって、訂正情報は流言情報よりも高精度に抽出可能であると考えられる。

6 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、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、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。また、流言情報自体の判定精度と比較して、訂正情報の判定精度が高いことを示した。訂正情報には流言情報が含まれている場合が多く、提案手法は、流言情報の抽出において有効であると考えられる。

今後は、流言情報クラウドを実現するために、流言情報を提供する機能など、訂正判定分類器以外の機能を構築する。

謝辞

本研究は、JST 戦略的創造研究推進事業による。

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] 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，グループウェアとネットワークサービス研究会，Vol.2011-GN-81, No.17, pp.1-7 (2011).
- [5] 荻上チキ：検証 東日本大震災の流言・デマ，光文社新書 (2011).
- [6] 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).
- [7] 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).