

拡張 SIR モデルによる Twitter でのデマ拡散過程の解析

False Rumor Diffusion Analysis based on The SIR-Extended Information Diffusion Model

岡田 佳之*1
Y. Okada

榊 剛史*2
T. Sakaki

鳥海 不二夫*2
T. Toriumi

篠田 孝祐*3
K. Shinoda

風間 一洋*4
K. Kazuhiro

野田 五十樹*5
I. Noda

沼尾 正行*1
M. Numao

栗原 聡*6
S. Kurihara

*1 大阪大学
Osaka Univ.

*2 東京大学
Tokyo Univ.

*3 理研
RIKEN

*4 和歌山大学
Wakayama Univ.

*5 産総研
AIST

*6 電通大
UEC

Twitter is a famous social networking service and has received attention recently. Twitter user have increased rapidly, and many users exchange information. When 2011 Tohoku earthquake and tsunami happened, people were able to obtain information from social networking service. Though Twitter played the important role, one of the problem of Twitter, a false rumor diffusion, was pointed out. In this research, we focus on a false rumor diffusion. We propose a information diffusion model based on SIR model, classify the way of diffusion in four categories, and reappear the real diffusion by using this new model.

1. はじめに

現在、インターネットの発展に伴い、Facebook*1や Twitter*2といったソーシャル・ネットワーキング・サービス (SNS) と呼ばれるサービスが話題を呼んでいる。しかしながら、SNS にはいくつかの問題点も挙げられる。そのひとつが、デマ情報の拡散である。実際に、東日本大震災時に緊急の情報を発信する手段として活躍した Twitter であるが、大震災直後の混乱した状況とも相まって、非常に多くのデマ情報が広がってしまうこととなった。そこで、本研究では、東日本大震災時において Twitter 上で実施に拡散した数例のデマに対し、拡散過程の解析、デマ拡散のモデル化、そしてそのモデルの検証を行う。

2. 関連研究

Weng らは情報拡散に寄与する Twitter ユーザーの識別方法として、従来の重要度を測るアルゴリズムである PageRank を拡張した、Twitter ネットワークにてしばしば見られる同類性を考慮した TwitterRank を提案している [7]。Balshy らは口コミなどの商業的観点から URL を含むツイートの拡散に着目し、拡散の起点となったユーザーの特徴やツイートに含まれる URL のリンク先の内容と、情報が拡散する範囲の関係などについて調べている [2]。

また、Twitter 上を流れる情報の信頼性についても研究が行われている。Castillo らは、投稿されたツイートの長さや主題、ツイートを投稿した各種ユーザーの特徴から、ツイートの内容の信頼度を算出する手法を提案している [3]。梅島らはデマツイートに見られる傾向を把握するため、東日本大震災時に多くのユーザーにリツイートされたツイートを分析し、デマツイートが持つ特徴を調べている [6]。Marcelo らは、流れてきた情報に対する信頼度を図るべく、チリ地震時のツイートを調べ、

デマと真実では伝播の仕方が異なることを発見した [8]。しかしながらデマツイートの拡散に着目すると、Castillo らや梅島らによれば完全にデマかどうかを判別することができないことに加え、情報拡散の研究は行われているものの、情報の収束を対象とした研究は行われていない。

3. デマ情報・訂正情報拡散モデル

3.1 SIR モデル

SIR モデルは Kermack らによって提案された、伝染病拡散の数理モデルのひとつである [4]。

SIR モデルではある集団に属する人を、健康者 (S)、感染者 (I)、回復者 (R) の 3 種類に分類する。このモデルでは、 S の人は I の人と接触することにより、感染率 $\rho_{(S \rightarrow I)}$ に従って病気に感染し、 I に変化する。 I の人は治癒速度 $\rho_{(I \rightarrow R)}$ に従い、病気を治癒し免疫を獲得し、 R に変化する。時間 t における S の人数を $S(t)$ 、 I の人数を $I(t)$ 、 R の人数を $R(t)$ とすると、時間の経過による S 、 I 、 R の人数の変化は式 1 で表される。

$$\begin{cases} \frac{dS(t)}{dt} = -\rho_{(S \rightarrow I)} I(t) S(t) \\ \frac{dI(t)}{dt} = \rho_{(S \rightarrow I)} I(t) S(t) - \rho_{(I \rightarrow R)} I(t) \\ \frac{dR(t)}{dt} = \rho_{(I \rightarrow R)} I(t) \end{cases} \quad (1)$$

集団の全人数 N は、 $N = S(t) + I(t) + R(t)$ で一定である。

3.2 病気と情報の伝播の違い

SIR モデルを情報拡散に適用するにあたり、病気と情報の伝播について以下の点が挙げられる。

[R になる条件]

デマ情報の場合、訂正情報を受け取ることで、ようやくデマとして認識されるため、病気の場合のような時間経過のみによる I から R への変化は無い。

[訂正情報の拡散]

デマ情報だけでなく訂正情報も SIR モデルにおける病気であるとみなすことができ、デマ情報と訂正情報の両方が拡散する様子を記述する必要がある。

連絡先: 岡田 佳之, 大阪大学産業科学研究所,
〒567-0047 大阪府茨木市美穂ヶ丘 8-1,
Tel:06-6879-8426, Fax:06-6879-8428,
E-mail:okada@ai.sanken.osaka-u.ac.jp

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

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

$[S, I, R$ 以外の状態]

デマおよび訂正情報を受け取ったが、その情報を拡散させなかった人についても考慮する。これは後に訂正情報が拡散した場合、デマ拡散の事実を知らなかった人 (S)、デマ拡散の事実を知っていたが拡散させなかった人 (I_{get} とする)、デマを拡散させた人 (I) では、再びこれらの情報を取得した場合の振る舞いが異なると考えられるためである。

3.2.1 デマ情報・訂正情報拡散モデルの構築

以上の違いを踏まえ、ユーザーの感染状態を、次のように定める。

- S : デマ情報, 訂正情報の両方を見たことがない状態。
- I_{get} : デマ情報のみを見たことがある状態。
- I : デマ情報のみを投稿した状態。
- R_{get} : 訂正情報を見たことがある状態。
- R : 訂正情報を投稿した状態。

R_{get} や R のユーザーは、 S, I_{get}, I に変化することは無いものとする。各感染状態における感染率を $\rho_{(S \rightarrow I)}$ (S のユーザーがデマ情報を見たときに、 I となる確率)、 $\rho_{(I_{get} \rightarrow I)}$ 、 $\rho_{(S \rightarrow R)}$ 、 $\rho_{(I_{get} \rightarrow R)}$ 、 $\rho_{(I \rightarrow R)}$ 、 $\rho_{(R_{get} \rightarrow R)}$ とする。全ユーザー数を N 、全ユーザーの友人 (フォロワー) の平均人数を F とする。また、ある時間 t における $S, I_{get}, I, R_{get}, R$ の人数を $S(t), I_{get}(t), I(t), R_{get}(t), R(t)$ とする。さらに、ユーザーが投稿した情報を、全ての友人が見たと仮定する。SNS でのデマ拡散の場合も、デマ情報の感染速度は SIR モデル同様 S の人数に比例する。しかし SNS では、あるユーザーが投稿した文章は、基本的にそのユーザーの友人しか見ることはない。このため、デマの感染速度は、各ユーザーの I 以外の状態である友人の数に比例する。以上から、訂正情報がまだ一度も投稿されていない状況 (式 2)、デマ情報と訂正情報の両方が拡散している状況 (式 3) を表す。

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\frac{F}{N} I(t) S(t) \\ \frac{dI_{get}(t)}{dt} = (1 - \rho_{(S \rightarrow I)}) \frac{F}{N} I(t) S(t) \\ \quad - \rho_{(I_{get} \rightarrow I)} \frac{F}{N} I_{get}(t) I(t) \\ \frac{dI(t)}{dt} = \rho_{(S \rightarrow I)} \frac{F}{N} I(t) S(t) \\ \quad + \rho_{(I_{get} \rightarrow I)} \frac{F}{N} I_{get}(t) I(t) \end{array} \right. \quad (2)$$

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\frac{F}{N} I(t) S(t) - \frac{F}{N} R(t) S(t) \\ \frac{dI_{get}(t)}{dt} = (1 - \rho_{(S \rightarrow I)}) \frac{F}{N} I(t) S(t) - \rho_{(I_{get} \rightarrow I)} \frac{F}{N} I_{get}(t) I(t) \\ \quad - \frac{F}{N} I_{get}(t) R(t) \\ \frac{dI(t)}{dt} = \rho_{(S \rightarrow I)} \frac{F}{N} I(t) S(t) + \rho_{(I_{get} \rightarrow I)} \frac{F}{N} I_{get}(t) I(t) \\ \quad - \frac{F}{N} I(t) R(t) \\ \frac{dR_{get}(t)}{dt} = (1 - \rho_{(S \rightarrow R)}) \frac{F}{N} R(t) S(t) \\ \quad + (1 - \rho_{(I_{get} \rightarrow R)}) \frac{F}{N} I_{get}(t) R(t) \\ \quad + (1 - \rho_{(I \rightarrow R)}) \frac{F}{N} I(t) R(t) \\ \quad - \rho_{(R_{get} \rightarrow R)} \frac{F}{N} R_{get}(t) R(t) \\ \frac{dR(t)}{dt} = \rho_{(S \rightarrow R)} \frac{F}{N} R(t) S(t) + \rho_{(I_{get} \rightarrow R)} \frac{F}{N} I_{get}(t) R(t) \\ \quad + \rho_{(I \rightarrow R)} \frac{F}{N} I(t) R(t) + \rho_{(R_{get} \rightarrow R)} \frac{F}{N} R_{get}(t) R(t) \end{array} \right. \quad (3)$$

4. ツイートの収集と分析

4.1 ツイート収集手順

本研究では、提案モデルによる現実の拡散状況の再現性を図るべく、SNS のうち、特に拡散速度が速いといわれている Twitter に焦点を当てた。研究対象として、鳥海らの東日本大震災前後における Twitter ネットワークの変化に関する研究 [5] にて収集されたツイートと 2011 年 1 月 30 日時点でのユーザーのフォローネットワークを用いた。その中から、まず、あるデマに言及しているデマツイートおよび訂正ツイートを抽出し、分析を行った。

デマツイートおよび訂正ツイートの抽出は表 1 に示す手順で行った。

表 1: デマツイートおよび訂正ツイートの抽出手順

- Step1: 実際に拡散したデマをひとつ選択する。
- Step2: デマの内容から、当該デマ関連のツイート収集のためのキーワードを設定する。
- Step3: Step2 で設定したキーワードを含むツイートを全ツイートデータから抽出する。
- Step4: Step3 で抽出したツイートの中で、RT 数が多いデマおよび訂正ツイートを複数選択する。
- Step5: 選択したデマおよび訂正ツイートから、当該デマに関するデマおよび訂正ツイートを抽出するためのキーワードをそれぞれ設定する。
- Step6: Step5 で設定したキーワードを含むデマおよび訂正ツイートを Step3 で抽出したツイートの中からそれぞれ抽出する。

4.2 選択したデマ

本研究で研究対象として選択したデマを図 2 に示す。

表 2: 選択したデマ

| ラベル名 | デマ内容 |
|-------|--|
| コスモ石油 | コスモ石油の爆発により有害物質が雲などに付着し、雨などといっしょに降る |
| 節電 | 関東地区に電力の融通を行うため、他の地域でも節電をするのがよい |
| ヨウ素 | 放射性ヨウ素の被害を防ぐため、うがい薬やとるる昆布などに含まれるヨウ素が良い |
| トルコ | トルコが日本に 100 億円寄付 |
| 台湾 | 日本政府が台湾からの義援金を拒否した |
| ポケモン | 人気ゲームソフトであるポケットモンスターのクリエイターが津波の犠牲になり、亡くなった |
| フジテレビ | フジテレビへ募金するとそのお金は日本ユニセフへ向かう |

4.3 収集結果

本研究で選択した 7 つのデマに関するツイート数の時間推移と、ユーザーの感染状態の変化を以下の図に示す。なお、表 1 に示した各キーワードについての例を表 3 に示す。

表 3: 各種キーワード (コスモ石油)

| | |
|---------------|---|
| 関連ツイート収集キーワード | コスモ石油, 有害物質 |
| デマツイート収集キーワード | 傘, カッパ, レインコート |
| 訂正ツイート収集キーワード | デマ, ガセ, 嘘, 誤, 偽, 否定, 無害, チェーンメール, チェンメ, 事実, ない, ありません |

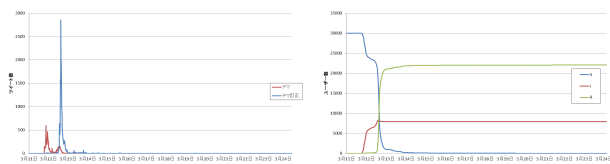


図 1: コスモ石油に関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

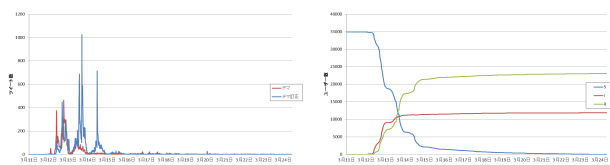


図 2: 節電に関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

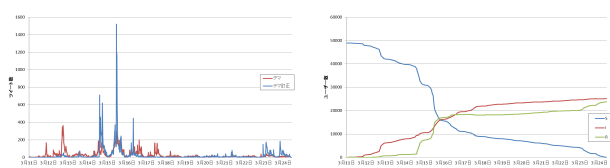


図 3: ヨウ素に関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

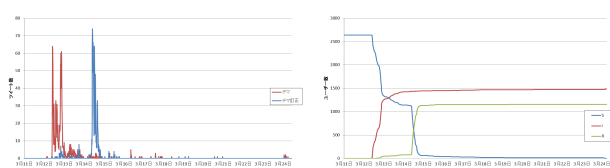


図 4: トルコに関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

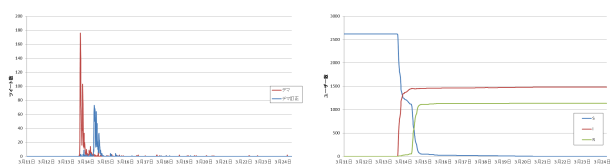


図 5: 台湾に関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

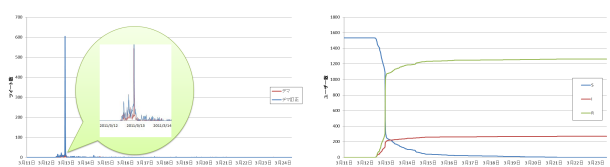


図 6: ポケモンに関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

図 1~7 の、各デマテーマにおけるデマツイート数およびデマ訂正ツイート数の時系列データのグラフを基に、デマツイートおよびデマ訂正ツイートの、拡散期間と拡散時期という 2 つの観点から、それらのツイートの拡散傾向を 4 つに分類した。

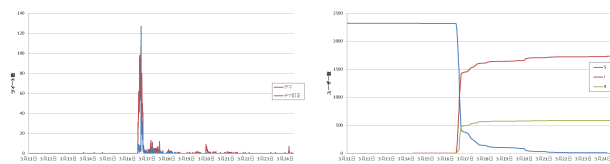


図 7: フジテレビに関するデマ: ツイート数の変化 (右) とユーザーの状態変化 (左)

その分類とは、一発非同時型 (コスモ石油, トルコ, 台湾), 一発同時型 (ポケモン, フジテレビ), 複非数同時型 (ヨウ素), 複数同時型 (節電) である。なお、一発型は爆発, 寄付, 死亡といった事件性を持つ話題であり, 複数型はそれがデマであると気づくには専門的な知識を要する話題となっている。

5. シミュレーションによる検証

5.1 シミュレータ

本研究では, Twitter ネットワークにおけるデマツイートおよびデマ訂正ツイートの拡散モデルの検証を行うため, シミュレータを作成した。シミュレーションで実行される処理を表 4 に示す。

表 4: シミュレーション 1 ステップにおけるプロセス

- Process1: 感染状態の状態遷移を行う。前ステップの結果に従い, 各ノードの状態を決定する。
- Process2: Process1 にて状態が I または R に変化したノードのフォロワーについては, 各状態における感染率に従い, 次ステップ開始時に遷移する感染状態を決定する。

5.2 シミュレーション環境の設定

今回用いたネットワークのパラメータは表 5 の通りであり, 全シミュレーションで同一ネットワークを用いた。

表 5: シミュレーション: ネットワークのパラメータ

| | |
|----------|---------------------------------|
| ノード数 | 50,000 |
| リンク数の期待値 | 上限 =30.00, 下限 =10, パレート指数 =0.5 |
| リンクされやすさ | 上限 =15.0, 下限 =0.05, パレート指数 =0.5 |

本研究では一発非同時型と一発同時型についてシミュレーションを行う。この両一発型のシミュレーションを行うにあたり, 簡単のため, 各ノードの情報拡散の行動について, 以下の条件を設定する。

- 各ノードは, デマおよび訂正ツイートをそれぞれ 1 回ずつのみ投稿することができる。
- 各ノードは, デマツイートを始めて受け取った際にそれを拡散させなかった場合 (I_{get}), 他のユーザーから再びデマツイートを受け取ったとしても拡散は行わない (I に変化しない)。訂正ツイートについても同様である。

これは, 一発型のデマの場合は拡散期間が短期であることから, 複数のデマ情報を見たノードは少ないという推測に基づく。また, 実データからはツイートを見ただけのユーザーの数を正確に把握できないということもあり, 今回は, I_{get} , R_{get} は考慮せずにシミュレーションを行う。

5.3 シミュレーション（一発非同時型）

提案モデルの検証のため、拡散傾向が一発非同時型であると定義した、コスモ石油、トルコ、台湾の関するデマケースに対して、シミュレーションを行う。それぞれのデマケースについて第 3.2.1 項にて定義した拡散率についてのパラメータを設定し、シミュレーションを表 4 の手順で行った。

5.3.1 シミュレーション結果

シミュレーションによるノードの感染状態別の数を図 8 に示す。実データとの比較のために、図 8 にはそれぞれ図 1, 図 4, 図 5 の一部も併記する。なお、各デマの解析結果を基にした感染率の設定を表 6 に示す。

表 6: 感染率の設定

| 感染率 | コスモ石油 | トルコ | 台湾 |
|-------------------------------|-------|--------|--------|
| $\rho(S \rightarrow I)$ | 0.05 | 0.05 | 0.06 |
| $\rho(I_{get} \rightarrow I)$ | 0 | 0 | 0 |
| $\rho(S \rightarrow R)$ | 0.1 | 0.0375 | 0.0072 |
| $\rho(I_{get} \rightarrow R)$ | 0.1 | 0.0375 | 0.0072 |
| $\rho(I \rightarrow R)$ | 0.147 | 0.016 | 0.026 |
| $\rho(R_{get} \rightarrow R)$ | 0 | 0 | 0 |

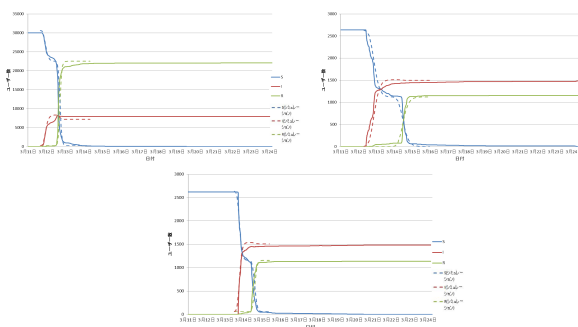


図 8: 各ノードの感染状態の変化のシミュレーション:コスモ石油 (左上), トルコ (右上), 台湾 (下)

5.4 シミュレーション（一発同時型）

第 5.3 節と同様に、一発同時型に属するポケモン、フジテレビに関するデマについてのシミュレーションを行う。なお、今回はデマと訂正の発生ステップを同一に設定した。

5.4.1 シミュレーション結果

シミュレーションによるノードの感染状態別の数を図 9 に示す。実データとの比較のために、図 9 にはそれぞれ図 6, 図 7 の一部も併記する。なお、各デマの解析結果を基にした感染率の設定を表 7 に示す。

表 7: 感染率の設定

| 感染率 | ポケモン | フジテレビ |
|-------------------------------|-------|-------|
| $\rho(S \rightarrow I)$ | 0.21 | 0.05 |
| $\rho(I_{get} \rightarrow I)$ | 0 | 0 |
| $\rho(S \rightarrow R)$ | 0.357 | 0.034 |
| $\rho(I_{get} \rightarrow R)$ | 0.357 | 0.034 |
| $\rho(I \rightarrow R)$ | 0.088 | 0.032 |
| $\rho(R_{get} \rightarrow R)$ | 0 | 0 |

6. まとめ

本研究では、まず SIR モデルを基礎とした、Twitter 上での情報拡散モデルを構築した。次に、東日本大震災直後に拡散したデマを解析し、デマの拡散の様子を調べた。その結果を基に、それぞれのデマにおけるデマおよび訂正情報の拡散過程に着目し、それらの情報の拡散方式を一発非同時型、一発同

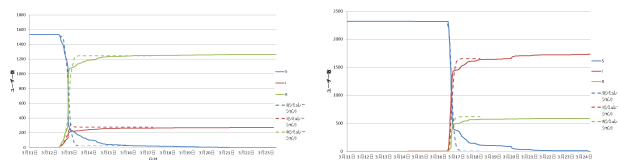


図 9: 各ノードの感染状態の変化のシミュレーション:ポケモン (左), フジテレビ (右)

時型、複数非同時型、複数同時型の四つに分類した。そして、一発型について、解析結果と提案モデルによるシミュレーションを行い、現実のデマ拡散の様子の再現に成功した。今後は、複数型においても提案モデルの検証を図る必要がある。さらに、Twitter 以外の SNS における情報拡散についても考慮すべきと考える。

謝辞

本研究を遂行するに当たり、Twitter 検索のログデータを提供して頂いたクックパッド株式会社の兼山元太氏、また、本研究を共同で行った、現在 NTT ネットワークサービスシステム研究所所属の白井高士氏に深く感謝の意を表す。

参考文献

- [1] Yasuyoshi Aosaki, Taro Sugihara, Katsuhiko Umemoto: Examining the Trend toward a Service Economy in Information Media through Changes to Technology: Influence of Twitter on Media Companies, *Proceedings of Technology Management for Global Economic Growth (PICMET) '10*, pp.1-5, 2010.
- [2] Eytan Bakshy, Jake M. Hofman, Winter A. Mason, Duncan J. Watts: Everyone's an Influencer: Quantifying Influence on Twitter, *Proceedings of the fourth ACM international conference on Web search and data mining*, pp.65-74, 2011.
- [3] Carlos Castillo, Marcelo Mendoza, Barbara Poblete: Information Credibility on Twitter, *Proceedings of the 20th international conference on World wide web*, pp.675-684, 2011.
- [4] W. O. Kermack, A. G. McKendrick: A Contribution to the Mathematical Theory of Epidemics, *Proceedings of the Royal Society 115A*, pp.700-721, 1927.
- [5] 鳥海不二夫, 篠田孝祐, 栗原聡, 榎剛史, 風間一洋, 野田五十樹: 震災がもたらしたソーシャルメディアの変化, *JWEIN11*, pp.41-46, 2011.
- [6] 梅島彩奈, 宮部真衣, 荒牧英治, 瀬本明代: 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, Vol.2011-IFAT-103, No.4*, pp.1-6, 2011.
- [7] Jianshu Weng, Ee-Peng Lim, Jing Jiang, Qi He: Twitter-Rank: Finding Topic-sensitive Influential Twitterers, *Proceedings of the third ACM international conference on Web search and data mining*, pp.261-270, 2010.
- [8] Marcelo Mendoza, Barbara Poblete, Carlos Castillo: Twitter Under Crisis: Can we trust what we RT?, *1st Workshop on Social Media Analytics (SOMA '10)*, 2010.