

# 事業者の行動とユーザーのリテラシーを考慮した SNS 上のデマ拡散に関する分析

○関澤佑（早稲田大学 高橋研究室）

## Analysis of hoax spread on SNS considering operator behavior and user literacy

\*Y.Sekizawa (Waseda University)

**概要**— 近年、SNS 上で問題になっているデマ情報の拡散について、感染症のウイルスの拡散モデルである SIR モデルを拡張して情報拡散のモデルとして分析する。その際、先行研究では考えられていなかったネットリテラシーや運営事業者の行動などを考慮して、より現在の SNS に近いモデルを作成し、SNS 事業者がどの程度介入すれば悪質なユーザー、果てはデマ情報の拡散が収束するかを確認する。今回 SNS 事業者の行動としてユーザーの凍結を考えている。デマ情報を流しているユーザーのうち悪質なものを事業者が直接アカウント凍結を行い、稼働不能にするものである。  
**キーワード:** SIR モデル 事業者 ネットリテラシー

### 1 研究背景

近年、Twitter や Facebook, Instagram などの SNS が発達することで人々は様々な情報をインターネット上で手に入れることができるようになった。特に、クチコミという拡散モデルが生まれ、より実体験に近い情報を SNS を通じて得られるようになった。

その一方で、デマ情報(根拠のない、間違った情報や不利益・混乱をもたらす情報をふくむもの)が問題となっている。大震災の時は緊急の情報を発信する手段として Twitter が活躍したが、震災直後の混乱や悪意あるユーザーなども相まって、非常に多くのデマ情報が広がってしまった。<sup>1)</sup>

それ以降様々な研究がなされたが、いまだにデマ情報が広がっている状況を改善する方策は出されておらず、熊本地震や西日本豪雨の際にもデマ情報が流れた。またある事件に無関係の人や会社が、名前が似ているというだけで犯人と関係があるとされ、クレームの電話などの営業妨害などが起きている現状もある。その原因はデマ情報が収束することなく燃え上がり、拡散されてしまうことにある。

### 2 従来研究・研究目的

白井らは、伝染病が広まっていく様子を記述する SIR モデルを情報拡散に用い、SNS におけるデマ情報・訂正情報の拡散のモデル化を行った。<sup>1)</sup>

池田らは、白井らのモデルをベースとして、エージェントの多様性と情報価値の変化を考慮した「マルチエージェント型拡張 SIR モデル」を提案した。<sup>2)</sup>

この研究により、ユーザーの好みや感度をモデルに反映することができた。

また、本研究と従来研究との違いや新規性を Table.1 に示す。

本研究では、SIR モデルを用いて実データ(東日本大震災のデマ情報)などを再現している従来研究とは違い、SNS の構造そのものをモデル化することに重きを置いており、その際に従来研究では考慮されていなか

った運営事業者の行動や、ユーザーのネットリテラシーをモデルに組み込んでいる。

Table.1 各研究の立ち位置

	白井ら,2012	池田ら,2014	本研究
状態遷移	確率モデル	エージェントベース	エージェントベース
実データ	比較, 再現	再現	×
他要素	×	興味度 感度 影響度	事業者の行動, 興味度(ネットリテラシーを組み込む)
状態	SIR	SIR	SIRF

本研究の目的は従来の SIR モデルで考えられていない運営事業者の動き(悪質なユーザーを凍結させる)や、ユーザーのリテラシーの向上を踏まえた拡張型 SIR モデルをエージェントベースでモデル化し、どの程度厳しく、どの程度介入すればデマの拡散が抑制されるか確かめることにある。

### 3 モデル化

#### 3.1 モデルの全体構造

本研究のモデルは大まかに以下の状態に分かれている。

##### ①初期状態生成

・・・SNS のネットワーク構造の作成や、リテラシーのあるユーザーにの感度を調整する

##### ②ユーザーのデマ拡散および回復(SIRF モデル上での状態遷移)のフェーズ

③運営事業者が悪質なユーザーを凍結させるフェーズ(I 状態→F 状態)のフェーズ

以上の3段階に分かれている、初期状態を生成した後は②と③を繰り返す。

### 3.2 ②-1. S 状態→I 状態

本モデルでは SIR モデルの状態遷移をエージェントベースで考えており、ユーザーには以下の3つのパラメータを付与する。

**Sensitivity** : 感度 を表しており、情報の信じやすさ、流されやすさを表す。感度が高いほどデマ情報を拡散しやすい。初期状態でリテラシーのあるユーザーは感度を下げた状態でシミュレーションを行う。

**Interest** : 興味度 を表しており、そのデマ情報のトピックに対する興味を表している。  $-1 \leq \text{Interest} \leq 1$  の範囲で動かす予定である。

**Influence** : 影響度 そのユーザーがフォローされているユーザーに与える影響を表している。フォローとフォロワーの比で定義し、フォローよりフォロワーの方が多い(多くの人にフォローされている)ユーザーは影響度が高いとする。

以上の3変数を用いて、S 状態→I 状態の繊維に用いる変数  $\text{StoI}$  を以下のように定義する。

$$\text{StoI}_{user \cdot t} = \text{StoI}_{user \cdot t-1} + \text{Interest}_{user} \text{Sensitivity}_{user} \sum_n \text{Influence}_n$$

この値と、各ユーザーの閾値を比べ、状態遷移するかどうかを決める。

### 3.3 ②-2 I 状態→S 状態

I 状態から R 状態への繊維は SIR モデルでは訂正情報や自然治癒による回復、または死を表しているが、本モデルではデマだと気づいて拡散しなくなる状態として定義している。その際従来の SIR モデルの研究と同じように確率遷移で状態を変化させる予定である。

### 3.4 運営事業者の行動(I 状態→F 状態)

SNS 上運営事業者ができる行動としては悪質なユーザーを凍結させる、というものがある。

本研究ではこの行動をモデルに組み込み、SNS のモデルを作るのが目的であり、凍結された状態を表す F(Freeze) という状態を新たに加え、SIRF モデルとしている。I 状態のユーザーの悪質さ(どれだけデマの拡散に関与しているか)を表すパラメータとして、Badness というパラメータを与え、事業者が決めた閾値を超えた場合に事業者はそのユーザーを凍結させる。

Badness の与え方は、ある時間ステップで S 状態から I 状態へ変わったユーザーがフォローしていた I 状態のユーザー全員に、そのユーザーのフォロワー数(今後デマ情報を与える人数)の値を均等に割り振る。また、Badness は単調に増加し、減ることはなく累積で決める。具体的な図を Fig.1 に示す

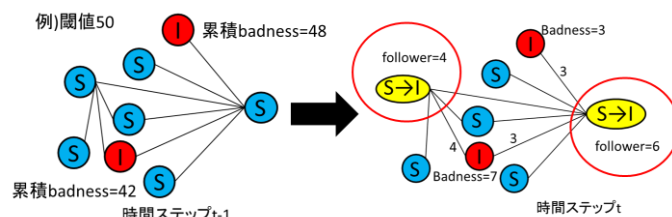


Fig.1

事業者が決めた閾値を 50 と仮定して、ある時間ステップ  $t$  でのふるまいを考える。時間ステップ  $t-1$  での各ユーザーのつながりと状態は左のようになっているものとし、 $t-1$  までの累積 Badness はそれぞれ 48 と 42 であるとする。

時間ステップ  $t$  において、3.2 節の式を参照してユーザーが 2 人 I 状態へ変わった(デマ情報に騙されて拡散)とすると、右図 2 人のユーザーのフォロワー数とフォローしている I 状態のユーザーの数から、上の赤いユーザーには Badness が 3、下の赤いユーザーには 7 がこの時間ステップ  $t$  で割り振られる。

次に事業者は累積 Badness を更新し、上の赤いユーザーは 51、下の赤いユーザーは 49 となる。

閾値 50 を超えた上の赤いユーザーが凍結対象となり、時間ステップ  $t+1$  以降、SNS ネットワークから除外される。

## 4 今後の予定

まだモデルの全体像を組んでいる段階のため、より細かいパラメータや閾値の設定方法などを決めていき、同時に状況に合わせて事業者の閾値(厳しさ)やネットリテラシーのあるユーザーの割合などをシナリオとして、作成していく予定である。

また、妥当性を持たせるために新たに作成した F 状態や事業者の凍結などを SIR モデルの状況との内的整合性を合わせていくことも課題である。

## 参考文献

- 1) 白井嵩士, 榊剛史, 鳥海不二夫, 篠田孝祐, 風間一洋, 野田五十樹, 沼尾正行, 栗原聡, 「Twitter ネットワークにおけるデマ拡散とデマ拡散防止モデルの推定」, 人工知能学会, 2012
- 2) 池田圭佑, 岡田佳之, 榊剛史, 鳥海不二夫, 篠田孝祐, 風間一洋, 野田五十樹, 諏訪博彦, 栗原聡 「マルチエージェント型拡散 SIR モデルを用いた情報拡散シミュレーションの評価」, 情報処理学会, 2014
- 3) <https://this.kiji.is/457461473188529249>, latest access 20190809