

個人の特性と話題を考慮した情報伝播モデルの研究

千葉 隆雄 † 鈴木 育男 † 山本 雅人 † 古川 正志 †

†北海道大学大学院情報科学研究科

1 はじめに

インターネットの普及に伴い、個人のウェブサイトや掲示板・ブログといった個人の情報発信ツールが活発に利用されるようになってきた。そのため、情報が伝播する際の振る舞いが変化しつつあると考えられる。

従来の情報伝播の研究では、パーコレーション理論 [1] や SIS モデル・SIR モデルに代表される伝染病の感染モデル [2] が利用されてきた。しかし、これらのモデルでは、ある情報がネットワーク構造によってどのように伝播するかを解析できる意味で有用なものであるが、人に対応するノードの条件が均一に定義されているため、情報の伝播に重要であると考えられるノードの属性(興味)が考慮されていない点で不十分であるといえる。

そこで、本研究では、各人の興味に基づいて話題が伝播するネットワーク上の情報伝播モデルを提案する。また、ウェブ上の画像データの伝播情報を用いて、本研究の有効性について考察する。

2 提案手法

本研究では各ノード及び話題の属性を表す属性ベクトルを提案し、属性ベクトルに応じてネットワーク上で情報を伝播するモデルを作成した。

2.1 ネットワークの構成

提案モデルでは、人をノード、人間関係をリンクとした人間関係ネットワークを想定する。さらに、情報伝播の経路をリンクとした伝播経路ネットワークを作成する。

ネットワークは以下のように定義される

$$N = \langle V, E \rangle, V = \{v_i\}, E = \{(v_i, v_j) | v_i, v_j \in V\}$$

2.2 話題のカテゴリ

話題はその内容によっていくつかのカテゴリに属している。そこで、本研究では話題 T_i が属するカテゴリ

A Study of A Information Propagation Model Related with Topic and Person Characteristics

†Takao CHIBA †Ikuo SUZUKI †Masahito YAMAMOTO
†Masashi FURUKAWA

†Graduate School of Information Science and Technology, Hokkaido University

takao@complex.eng.hokudai.ac.jp

を表す話題ベクトル $t^{(i)}$ を以下のように定義する。

$$t^{(i)} = [t_{c_1}^{(i)} t_{c_2}^{(i)} \dots t_{c_n}^{(i)}]$$

$$t_{c_k}^{(i)} = \begin{cases} 1 & \text{情報 } T_i \text{ がカテゴリ } c_k \text{ に属している} \\ 0 & \text{情報 } T_i \text{ がカテゴリ } c_k \text{ に属していない} \end{cases} \quad (1)$$

2.3 ノードの興味

現実世界での人々は、様々なカテゴリに対して興味を持っており、どのカテゴリに興味を持っているかは人によって異なる。そこで、本研究では人を表すノード v_i が、どのカテゴリに興味を持っているかを、ノードの属性である興味ベクトル $a^{(i)}$ で以下のように定義する。

$$a^{(i)} = [a_{c_1}^{(i)} a_{c_2}^{(i)} \dots a_{c_n}^{(i)}]$$

$$a_{c_k}^{(i)} = \begin{cases} 1 & \text{ノード } v_i \text{ がカテゴリ } c_k \text{ に興味がある} \\ 0 & \text{ノード } v_i \text{ がカテゴリ } c_k \text{ に興味がない} \end{cases} \quad (2)$$

2.4 話題の伝播

また、話題の伝播はノード v_i, v_j 両方が話題の持つカテゴリに対して興味を持っている場合のみ起こるものとする。

これまで定式化してきた話題ベクトル t_i 、興味ベクトル a_i 、重み α を用いてノード v_i, v_j 間の伝播確率 $p_{v_i v_j}$ を以下に定義する。

$$p_{v_i v_j} = \alpha \sum_k a_{c_k}^{(i)} a_{c_k}^{(j)} t_{c_k} \quad (3)$$

2.5 伝播アルゴリズム

情報伝播で用いるノードの状態は SIR モデルを用いる。すなわち、情報を受け取っていない「未受信状態」(S)、情報を近隣ノードに伝える「発信状態」(I)、情報を伝えなくなった「非発信状態」(R) の 3 つの状態を用いる。発信状態のノード v_i は、1 ステップごとにノード v_j に対して話題の伝播確率 $p_{v_i v_j}$ を基にした確率で話題の伝播を試み、その結果に基づいてを行う。

実際の伝播は以下の手順に従って行われる

- 最初に 1 つのノードに情報を受信させ発信状態にする。
- 発信状態のノード v_i は人間関係ネットワーク上で隣接関係にある未受信状態のノード v_j に対して、

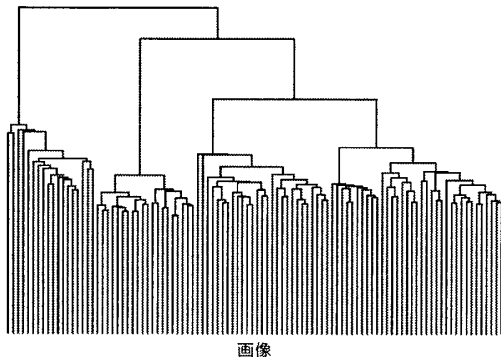


図 1: 伝播した画像のクラスタリング樹形図

情報の伝播確率 p_{v_i, v_j} に応じて情報を発信する。情報を受信したノードは発信状態へと遷移する。

3. 発信状態のノードは確率 β で非発信状態に遷移する。
4. 終了条件を満たすまで、手順 2,3 を繰り返す。

以上が話題の伝播アルゴリズムである。次章では、このモデルを用いて実験を行う。

3 実験

本研究では、ウェブ上の画像データの伝播情報を用いて、提案したモデルに関するシミュレーション実験を行い、提案モデルの有効性などについて議論する。

3.1 データ

本研究では Tumblr (<http://www.tumblr.com/>) から取得した画像データの伝播情報を用いる。取得したデータの画像は 92,802 種類、ユーザ数は 15,742 である。

以下の実験では、取得したデータの中で伝播規模の大きさ上位 100 個の伝播情報を用いる。この上位 100 個の伝播情報を用いて、ユーザをノード、情報の伝播経路をノード間の有向辺としたネットワークを作成した。その結果作成されたネットワークのノード数は 4,973 個、有向辺は 13,617 本である。

3.2 実験条件

本実験では、実データから作成したネットワーク上で提案モデルによるシミュレーション実験を行い、実際の情報の伝播規模と、提案モデルの伝播規模との比較を行う。

実験を行うために、情報のベクトル、ノードの興味ベクトルを以下の方法で設定する。伝播したノードをベクトルとして、ネットワークの生成に用いた 100 個の情報を階層的クラスタリングによってクラスタリングし、その結果を図 1 に示す。

α	0.05	0.08	0.1
伝播規模の比	0.594	1.557	2.130

図 2: シミュレーションと実データの伝播規模の比

この結果から、話題を 4 つのカテゴリに分割し、話題 T_i が属するカテゴリが 1、属していないカテゴリが 0 となるように T_i の話題ベクトルを設定する。また、ノード v_i の興味ベクトルを、 v_i が一度でも受信したことのあるカテゴリを 1、それ以外のカテゴリを 0 と設定する。

シミュレーション実験は各情報につき 10 回、合計 1000 回の試行を行った。

3.3 結果と考察

シミュレーション実験から得られた伝播規模の平均と実データの伝播規模の比とその時の α の値を表 2 に示す。また、 $\beta = 0.5$ である。

結果から、パラメータによって伝播規模が大きく変化するため、伝播規模を実際のデータに近づけることは可能である。しかし、情報によって、伝播規模の比には大きくばらつきがあり、情報の伝わったノードの構成がシミュレーションと実データの間で大きく異なっているものも多い。これは、実験設定において、一度でも受信したことのあるカテゴリであれば、何度も受信しているカテゴリと同様の重みがノードの興味ベクトルに与えられているためと考えられる。これは、あるカテゴリに対する興味の重みがノードによって異なる可能性があることを示唆している。

4 まとめ

本論文では、ノードの興味と情報のカテゴリを考慮した情報伝播モデルを提案した。また、提案したモデルを用いて、ウェブ上のデータとの比較実験を行い、ノードの興味に重みなどを設定する必要があることがわかった。

今後、このモデルの改良を行う。また、実データの解析、モデルによるシミュレーション結果との比較を行っていきたい。

参考文献

- [1] G.Deutscher, R.Zallen, J.Adler. Percolation structures and processes, 1983.
- [2] Herbert W.Hethcote. Qualitative analyses of communicable disease models, 1976.