

一様採用付き Voter Model の提案

Proposing the Voter Model with Uniform Adoption

伏見 卓恭[†] 斉藤 和巳[†] 木村昌弘[‡] 大原剛三[¶] 元田浩[§]

Takayasu Fushimi Kazumi Saito Masahiro Kimura Kouzou Ohara Hiroshi Motoda

1. はじめに

社会ネットワーク上でのクチコミや広告宣伝, 志向や意見 (オピニオン) など, ある種の情報が拡散する. これらを最も広く拡散させるような高影響度のノードを特定することはバイラルマーケティングなどへの応用において重要である. よって意見や志向が形成されていく過程で, 期待影響度を推定する問題は非常に重要な研究課題である.

近年, ブログや知識共有サイトに代表される Web の技術革新が著しい. 商品などの購買者が商品のクチコミやレビューを投稿できるような Web サイトが増えている. こうした情報を閲覧したユーザがレビューに触発され, 商品に対する関心を持ち, その商品を購入する. そして, そのユーザも商品に関するレビューを投稿し, 他のユーザがそのレビューを閲覧するといった普及の連鎖が生じる.

このような現象をモデル化したものに Voter Model[1] (以下 VM) がある. VM では, 自己ループ付きの無向ネットワークを仮定し, 期待影響度を計算する. VM は基本的な確率モデルで, 隣接するノードのオピニオンの変化によって自身のオピニオンが決まる. 隣接するノードの影響を受けて状態が決まる点で線形閾値モデル[3]と同様の特徴を有している.

一方 Web ページの重要度を評価する指標として有名な Google の PageRank[2] (以下 PR) がある. PR では Web のハイパーリンク構造に基づき各 Web ページのスコアを計算し, 近年の Web ランキングシステムにおいて非常に重要な役割を果たしている. 出次数から得られる推移確率行列を用いて, ベキ乗法により PR スコアを計算する.

そこで我々は既存の VM を拡張する. すなわち, 対象とするデータを有向ネットワークに一般化し, オピニオンの一様採用 (Uniform Adoption, 以下 UA) を導入する. UA を導入した VM と PR は無向ネットワークにおいて等価な計算となり, 拡張版 VM による期待影響度のランキングと PR スコアのランキングは一致する.

本論文では, 提案する拡張版 VM における UA 確率 α の効果を実験により確かめ, さらに対象を有向ネットワークデータに一般化し提案モデルの評価を行う. 詳細には, 提案モデルと PR の関係を明確化し, 実有向ネットワークデータを用いて, 提案モデルによるランキングと PR のランキングの関係, コミュニティ構造の異なる張替ネットワークを生成し, コミュニティ構造と α の関係に注目した分析を行う.

2. 既存モデル

本論文では既存モデルである VM[1]を拡張したモデルを提案する. VM は各ノードが自己リンクを持つ無向グラフ $G=(V, E)$ で定義される. この章では元となる既存モデルである VM と PR の説明をする.

2.1 Voter Model

オピニオンダイナミクスの適切なモデルの一つに VM が挙げられる. VM は基本的な確率モデルで, 隣接するノードのオピニオンの変化によって自身のオピニオンが決まるように, 隣接するノードの影響を受けて状態が決まる点で線形閾値モデル[3]と同様の特徴を有している.

基本的な VM は無向ネットワークで定義され, 各ノードは K オピニオンのうちの 1 つのオピニオンを有するように初期設定する. そして, 各離散時間のタイムステップでランダムに隣接するノードのオピニオンを選ぶ. 我々の扱う基本的な VM では, ノード $v \in V$ の隣接ノード集合 $\Gamma(v) = \{u \in V; (u, v) \in E\}$ と定義し, 各ノードは 2 つのオピニオン (オピニオン 1, オピニオン 2) のどちらかを持つ. そして離散時間タイムステップの中で, ノード v は $\Gamma(v)$ からランダムに親ノード u を選び, u が時刻 $t-1$ で持っていたオピニオンを得る.

より形式的には, $f_t: V \rightarrow \{1, 2\}$ を時刻 t でのオピニオン分布とし, $f_t(v)$ を時刻 t でのノード v の持つオピニオンとする. ここで $f_0: V \rightarrow \{1, 2\}$ を初期オピニオン分布とし, $f_t: V \rightarrow \{1, 2\}$ 以下のように定義する. 任意の $v \in V$ に対して,

$$f_t(v) = \begin{cases} 1, & \text{with probability } \frac{n_1(t-1, v)}{n_1(t-1, v) + n_2(t-1, v)} \\ 2, & \text{with probability } \frac{n_2(t-1, v)}{n_1(t-1, v) + n_2(t-1, v)} \end{cases}$$

ここで, $n_k(t, v)$ は時刻 t において k オピニオンを持つノード v の隣接ノードの数を表す.

2.2 PageRank

有向ネットワーク G が与えられ, G のノード集合 S は $S = \{1, \dots, N\}$ と定義する. そして, G の隣接行列を A とする. A の (i, j) 成分は a_{ij} と表し, ノード i からノード j へリンクが存在すれば $a_{ij}=1$, 存在しなければ $a_{ij}=0$ となる. $d_{out}(i)$ をノード i の出次数とし, 推移確率行列 P を考える. 行列 P の各要素は, 次のように定義される.

$$p_{i,j} = \begin{cases} a_{i,j} / d_{out}(i) & (d_{out}(i) > 0) \\ z_j & (d_{out}(i) = 0) \end{cases} \quad (1)$$

ここで z は, $z_i \geq 0$ かつ $\sum_{i=1}^N z_i = 1$ となるような確率である. これは, ぶら下がり (dangling) ノード i ($d_{out}(i)=0$) か

[†] 静岡県立大学 University of Shizuoka

[‡] 龍谷大学 Ryukoku University

[¶] 青山学院大学 Aoyama Gakuin University

[§] 大阪大学 Osaka University

ら任意のノード j へは推移確率 z_j で一様ジャンプすることを意味する. この z の値はユーザ自身の好みで決まるため, ベクトル z はパーソナライズ・ベクトルと呼ばれる.

さらに, PageRank スコアベクトル y は次の反復処理の極限により定義される.

$$y_{k+1}^T = y_k^T ((1-\beta)P + \beta \cdot ze^T) = (1-\beta)y_k^T P + \beta \cdot e^T \quad (2)$$

ここで, β は一様ジャンプ行列 ze^T との混合比すなわち, 一様ジャンプ確率を表し, $e^T = \{1, \dots, 1\}$ である. この式は, ユーザが β の割合で他のページへ確率 z に従いランダムサーファージャンプすることを意味する. そして初期ベクトル $y_0 = (1/N)e$ から反復処理を収束するまで繰り返す. この反復処理の収束性能は以下の式で定義される.

$$\|y_{k+1} - y_k\| \equiv \sum_{i=1}^N |y_{k+1,i} - y_{k,i}| \quad (3)$$

式(3)の値がほぼ 0 になれば, 式(2)は唯一の解をもつことが保証される.

3. 提案モデル

本論文では, 前述した VM を拡張して, UA (Uniform Adoption) 付き VM を提案する. 従来の基本的な VM は, 隣接するノードのオピニオンに影響を受けて自身のオピニオンを決定するような現象を表すモデルであるが, UA (一様採用) を導入することによって, 隣接するノードのオピニオンのみでなく, 隣接関係のないノードのオピニオンの影響を受けるモデルである. これは, PR のランダム一様ジャンプに対応している.

PR のスコアベクトル y を求めるためには, (2) 式の推移確率行列 P を分解する. ネットワーク G の各ノードに関して出次数行列 D_{out} を考える.

$$D_{out} = \begin{cases} d_{out}^{-1}(i), & (i = j) \\ 0, & (i \neq j) \end{cases}$$

ここで $d_{out}(i)$ は, ノード i の出次数を表し, 自己ループ付きネットワークでは $d_{out}(i) > 0$ である. D_{out} を用いて, 式 (2) を書き換えると,

$$y_{k+1}^T = y_k^T ((1-\beta)D_{out}A + \beta \cdot ze^T) = (1-\beta)y_k^T D_{out}A + \beta \cdot e \quad (4)$$

と変形できる.

提案する拡張版の VM による期待影響度ベクトル計算を文献[1]に類似した考え方で定式化すると,

$$x_{k+1} = ((1-\alpha)AD_{in} + \alpha \cdot ze^T)x_k = (1-\alpha)AD_{in}x_k + \alpha \cdot e \quad (5)$$

となる. ここで, α は UA を選ぶ確率すなわち UA 確率のことである. 初期ベクトルは $x_0 = e$ である. さらに入次数行列 D_{in} は以下のように定義される.

$$D_{in} = \begin{cases} d_{in}^{-1}(i), & (i = j) \\ 0, & (i \neq j) \end{cases}$$

ここで $d_{in}(i)$ は, ノード i の入次数を表している.

式(4)と式(5)を見比べるとわかるように, PR と提案モデルでは, 同様の計算構造になっている. このことより, 無向ネットワークを対象に計算を行うと, 式中の A が対称行

列となり, さらに出次数行列 D_{out} と入次数行列 D_{in} は同一な行列となるため, PR によるスコアのランキングと提案モデルによる期待影響度のランキングは結果が一致する.

本論文では, 対象とするネットワークデータを有向ネットワークに一般化することにより, 提案モデルの UA 確率 α の効果や PR との関係を実験により示す.

4. 分析手法

本論文では, 大きく分けて3つのステップで実験する.

1 目目のステップでは, 提案モデルにおける UA 確率 α の効果を示すために, 各 α の値による期待影響度どうしの相関係数を計算し, 各 α ごとの相関係数をプロットする (図 1). さらに, 提案モデルによる期待影響度と PR によるスコアの相関係数を α ごとに計算しプロットする (図 2).

2 目目のステップとして, コミュニティ構造の異なるネットワークにおける提案モデルの挙動を分析するために, オリジナルネットワークと異なるコミュニティ構造をもつ張替ネットワークを生成する. ここでオリジナルネットワークとは, 我々の所有している実験に用いるネットワークデータのことを指す (詳細は後述).

張替ネットワークは, コミュニティ構造の違う同規模のネットワークによる情報拡散の期待影響度の違いを分析するために, 張替確率 ($p=2^{(k)}$; $0 \leq k \leq 10$) に従いオリジナルのネットワークの任意のリンク同士を張り替える作業を行う. コミュニティ構造の違いだけに注目するために, 各ノードの次数やネットワーク全体のノード数を不変でリンクを張り替える. このリンク張り替え時には, 自分から自分へリンク (self-link) は固定したまま, また同じ相手ノードに対し複数のリンク (multiple-link) を張らないように注意しなければならない.

具体的な張替ネットワークの生成方法は, 以下の通りである. オリジナルのネットワークからリンクリスト $L_E = (e_1, \dots, e_{|E|})$ を準備する. それぞれのリンクは, 例えば $e = (u, v)$ のように順序対になっており, *from-part* と *to-part* から成っている. これから, 2つのノードリスト L_{from} と L_{to} を作り出す. そして注意事項である, self-link と multiple-link に気をつけ, L_{to} をランダムに並び替える. その後 L_{from} と L_{to} を結合させ, 新しいリンクリストを作り, これに基づきオリジナルネットワークとは構造の異なる張替ネットワークを作成する.

3 目目のステップとして, コミュニティ構造の異なる張替ネットワークを用いて, 提案モデルにおける張替確率 p の効果を確認する. ステップ 2 で生成した 11 個の張替ネットワークに対して, 提案モデルによる期待影響度ベクトル x を求める. 具体的には, 提案モデルの各張替確率での張替ネットワークの期待影響度とオリジナルネットワークとの相関係数を計算しプロットする (図 3). また, 提案モデルと PR の関係を明確にするために, 各張替確率での提案モデルと PR の相関係数も同様に求める (図 4).

5. 評価実験と考察

提案モデルにおける UA 確率 α の効果を実験により分析するために, 上述した 4 つの相関分析を行う. 評価実験には 2 つの有向ネットワークデータを用いる.

5.1 ネットワークデータ

1つ目のネットワークは、ブログのトラックバックネットワーク(以下ブログネットワーク)である。ブログネットワークは、「goo ブログ」(<http://blog.goo.ne.jp/usertheme/>)の「JR 福知山線脱線事故」というテーマからトラックバックを10段辿ることにより、2005年5月に収集したものである。このネットワークは、12,047ノードと79,920リンクをもつ有向ネットワークである。多くの大規模なネットワークと同じように、自分に向かうリンク次数の分布(入次数分布)も自分から向かうリンク次数の分布(出次数分布)もべき則分布に従うという特徴を有す。入次数と出次数の相関係数は0.743771である。

2つ目のネットワークは、化粧品の口コミサイト“@cosme”(<http://www.cosme.net/>)から2009年12月に取得したデータである。サイトに登録しているあるユーザーXがユーザーYをお気に入り登録をすると、ユーザーXからユーザーYに対してリンクが張られる。このようにして生成されたネットワークは社会ネットワークとしてみなすことができる。このネットワークはランダムに選択したユーザから10段辿って収集したもので、ノード数は45,024でリンク数は546,930である。入次数と出次数の相関係数は0.519404である。

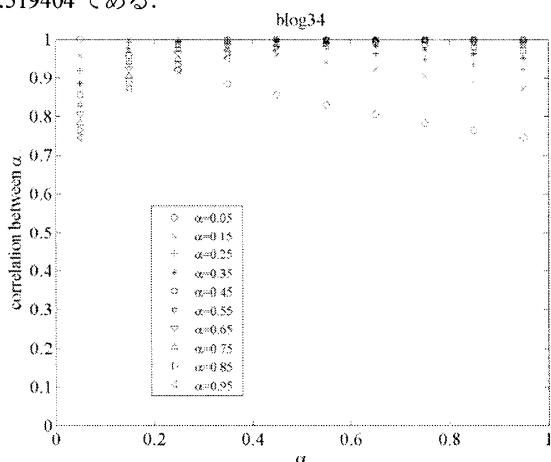
5.2 実験結果

5.2.1 UA 確率 α の効果

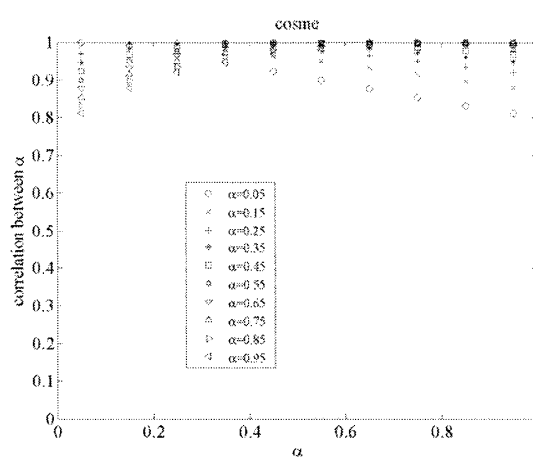
提案モデルにおける UA 確率 α の効果を実験により分析するために、各 α での相関係数をプロットする。

図1(a)はブログネットワークでの結果である。 $\alpha=0.25$ のとき、どの α との相関係数も高く、最も安定している。(b)はコスメネットワークでの結果である。 $\alpha=0.35$ のとき、どの α との相関係数も高く、最も安定している。一方 PR では、一様ジャンプ確率 $\beta=0.15$ 程度のときが安定していて、広く使われている[2]。このことから、どの構造のネットワークでも $0.15 < \alpha < 0.35$ の範囲で提案モデルの期待影響度は安定することが示唆される。また α が小さいときは、 α の差に対する相関係数の値は、ほぼ線形に減少するが、 α が大きいときは α が小さいときと比較して、上に凸な関数のように減少する。

さらに、図2では、任意の α での提案モデルの期待影響度と任意の β での PR のスコアの相関係数をプロットしたものである。どちらのネットワークでも、 α や β の値が大きくなるほど相関係数の値は高くなる。特に $\alpha=0.95$ では、入次数と出次数の相関係数程度の値を示している。

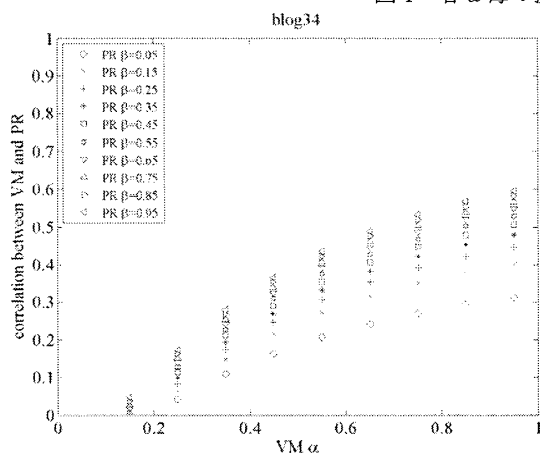


(a) ブログネットワーク

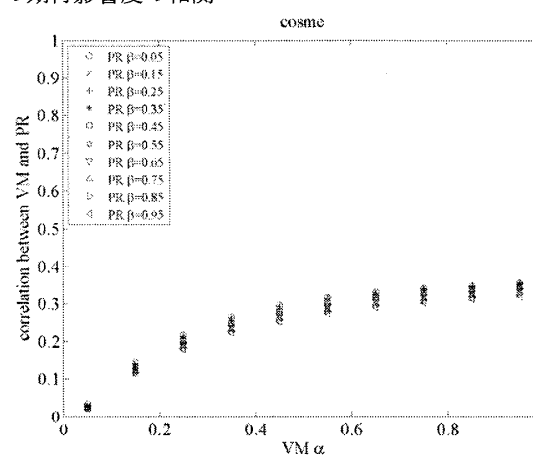


(b) コスメネットワーク

図1 各 α 毎の提案モデルの期待影響度の相関

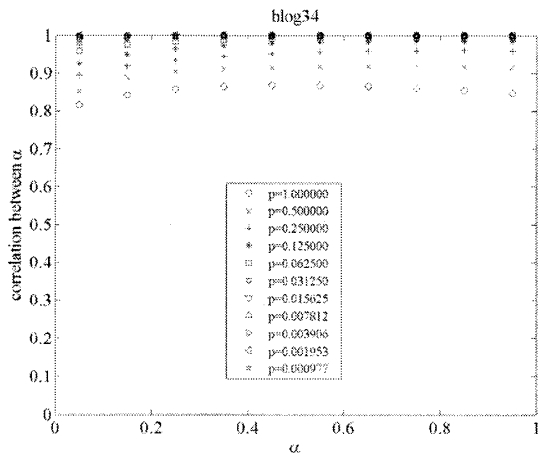


(a) ブログネットワーク

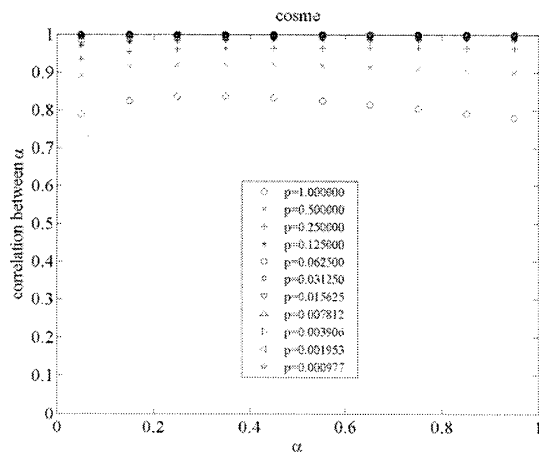


(b) コスメネットワーク

図2 各 α 毎の提案モデルの期待影響度と PR スコアの相関

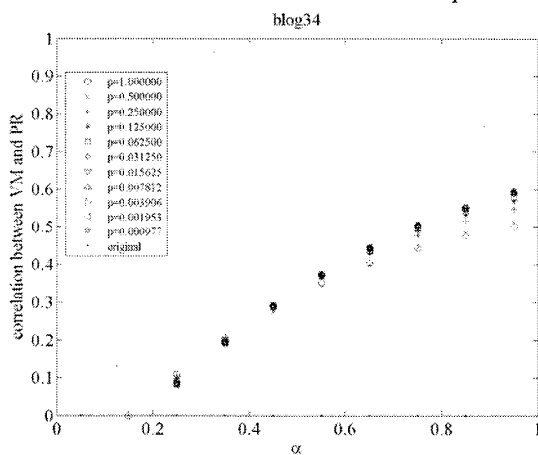


(a) ブログネットワーク

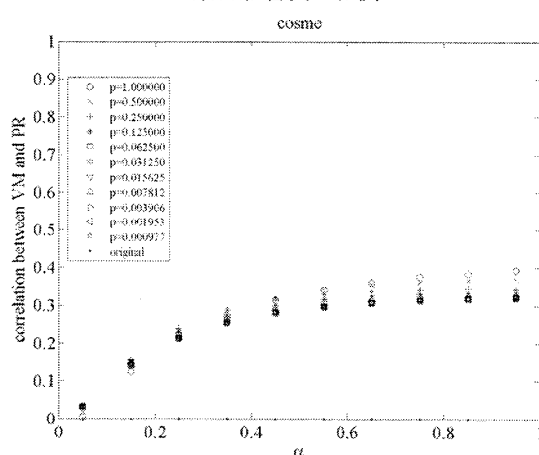


(b) コスメネットワーク

図3 提案モデルの各 p 毎の張替 NW とオリジナル NW の期待影響度の相関



(a) ブログネットワーク



(b) コスメネットワーク

図4 各 p 毎の張替 NW での提案モデルの期待影響度と PR スコアの相関

5.2.2 コミュニティ構造の効果

提案モデルにおける張替確率 p の効果を実験により分析するために、各 α での相関係数をプロットする。

図3の結果からわかるように、 α の値に関わらずコミュニティ構造が変化（張替確率を大きく）しても、依然として大きい相関係数を維持している。さらに図4の結果からも、曲線の分散が小さいため、コミュニティ構造による提案モデルの期待影響度への顕著な効果はない事が示唆される。また図4から、張替ネットワークにおいても α と β の値が大きくなるほど提案モデルと PR の相関係数は高くなる事がわかる。

以上の分析の結果をまとめると、1) 提案モデルにおける α の効果として、 $0.15 < \alpha < 0.35$ で期待影響度は安定する、2) 提案モデルと PR の関係として、 α の値が大きくなるほど相関が高くなり、特に $\alpha = 0.95$ では入次数と出次数の相関係数程度の値を示している、3) コミュニティ構造と α の値は関係がないことが示唆される。

上記評価実験から得られた提案モデルの性質から、我々の提案する UA 付き VM は従来の重要ノード抽出のためのランキングや中心性（例えば PR）とは異なった固有の価

値を持つと考えられる。その点でも提案モデルは新規性がある有用なモデルであると期待できる。

6. おわりに

本論文では、VM に UA を導入することにより拡張し、対象データを有向ネットワークに一般化することにより、提案モデルに評価実験を行った。その結果提案モデルにおける α の値は、コミュニティ構造とは関係ないが α の値が大きくなると PR と相関関係が高くなる事がわかった。また PR と同様に $0.15 < \alpha < 0.35$ で期待影響度が安定することもわかった。これらの性質から、重要ノード抽出のための新たなランキング手法となりうる事が示唆された。

参考文献

- [1] Eyal Even-Dar, Asaf Shapira, "A Note on Maximizing the Spread of Influence in Social Networks", Lecture Notes in Computer Science, Vol.4858, (2007).
- [2] Amy N. Langville, Carl D. Meyer, "Deeper Inside PageRank", Internet Mathematics, Vol.1(3) (2004).
- [3] Kempe D., Kleinberg J., Tardos E., "Maximizing the spread of influence through a social network", ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2003).