

トピックを考慮したソーシャルネットワーク上の情報拡散モデル

Topic-dependent Information Diffusion Model on Social Networks

吉川 友也*¹ 岩田 具治*² 澤田 宏*²
 Yuya Yoshikawa Tomoharu Iwata Hiroshi Sawada

*¹奈良先端科学技術大学院大学
 Nara Institute of Science and Technology

*²日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
 NTT Communication Science Laboratories, NTT Corporation

Modeling the phenomenon which information spread from person to person on social networks precisely is important to understand the information diffusion and available for the applications using the information diffusion models. The conventional models assume that information spreads based on the model's parameters which are independent from the contents of information. However, in general, interesting topics spread widely and boring topics do not diffuse. We thus develop Topic-dependent Information Diffusion Model, a general framework of information diffusion model which can describe logs of diffusion and content of the diffused information jointly. We thus derive the method of model's parameter estimation based on EM algorithm.

1. はじめに

マイクロブログサービスやソーシャルネットワーキングサービス (SNS) は、最近多くのユーザを獲得し、コミュニケーションツールとして生活に欠かせないものになっている。これらのサービス上では、ユーザはお気に入りのユーザを見つけ、そのユーザが発信する情報を受け取るために「フォロー」や「友達申請」という手段によって関係を構築する。ここで、ユーザをノード、ユーザ間の関係をリンクとみなせば、これはソーシャルネットワークとして見る事ができる。

ソーシャルネットワーク上では、各ユーザが発信する情報がユーザ間のリンクを介して広がっていく。情報の種類は様々で、ニュース、アイデア、時としてデマが流れる。また、情報の内容によっては、多くのユーザに情報が伝わったり伝わらなかつたりする場合があります、また、伝わるスピードも速かつたり遅かつたりと、情報拡散の仕方は、情報の内容・ネットワーク構造などに依存し多様に変化する。

情報拡散モデルによるソーシャルネットワーク上の情報拡散に関する研究はいくつか存在する。この種の研究でよく利用される基本的な確率モデルは、**独立カスケード (IC) モデル** [Kempe 03] である。IC モデルは、病気の広がり伝播モデルである *SIR* (*susceptible/infected/recovered*) モデルと同等とみなすことができる。IC モデルでは、あるノード u が情報を受け取ったとき、ノード u の子ノード v に対して拡散確率 $\kappa_{u,v}$ で情報が伝わるとする。ここで、IC モデルでは情報の内容については考慮しない。したがって、情報の内容に関わらずいつも同じ拡散確率によって情報が伝播することになり、現実の情報拡散現象を上手く捉えるのは難しいと考える。

本論文では、情報の内容を考慮するために、トピックを考慮した情報拡散モデルの枠組みを提案する。また、この枠組みを IC モデルに適用したトピック依存独立カスケード (TIC) モデルを提案し、EM アルゴリズムに基づいた TIC モデルのパラメータ推定アルゴリズムを導出する。提案モデルを評価するために、二種類の実験を行う。一つ目の実験では、我々が与えた

表 1: 表記法

記号	説明
M	情報拡散が行われた回数
W	重複なしの単語集合
N_m	m 番目の情報拡散の情報の単語数
S_m	m 番目の情報拡散で情報を受け取ったノード数
K	トピック数
$w_{m,n}$	m 番目の情報拡散の n 番目の単語
$d_{m,s}$	m 番目の情報拡散で s 番目に情報を受け取ったノードとその時刻のペア
z_m	m 番目の情報拡散のトピックで、 $z_m \in \{1, \dots, K\}$
$B(v)$	ノード v の親ノード集合
$F(v)$	ノード v の子ノード集合
$C_m(t)$	m 番目の情報拡散における時刻 t までにアクティブになったノード集合
C_m	m 番目にアクティブになった全てのノード集合

パラメータを使って TIC モデルの生成過程によって人工的に生成した拡散系列から、パラメータを復元できることを確かめる。二つ目の実験では、Twitter リツイートデータから TIC モデルのパラメータを推定し可視化することで、構築されたクラスタが定性的に有意義であることを示す。

2. トピック依存情報拡散モデルの枠組み

トピック依存情報拡散モデルは、情報の内容を考慮したソーシャルネットワーク上での情報拡散の生成モデルである。情報拡散が行われた場合、我々は情報の内容と、誰がいつ情報を受け取ったかを観測する。ソーシャルネットワーク上で拡散する情報は、一般的に自然言語である。我々はこの自然言語を *bag-of-words* で表現し、これを $\mathbf{w}_m = \{w_1, w_2, \dots, w_{N_m}\}$ とする。誰がいつ情報を受け取ったかを表す観測データを **拡散系列** と呼ぶ。拡散系列はある情報 \mathbf{w}_m を受け取ったノードと受け取った時刻のペアの集合 $\mathbf{d}_m = \{(u, t_{m,u}), (v, t_{m,v}), \dots\}$ で表す。

連絡先: 吉川友也, 奈良先端科学技術大学院大学, yuyaya@is.naist.jp

次に、トピック依存情報拡散モデルがどのように情報 \mathbf{w}_m と拡散系列 \mathbf{d}_m を生成するのかを示す。Procedure 1 は、トピック依存情報拡散モデルの生成過程である。

Procedure 1 トピック依存情報拡散モデルの生成過程

- 1: **for** $m = 1$ to M **do**
 - 2: トピックを選択: $z_m \sim \text{Multinomial}(\boldsymbol{\alpha})$
 - 3: 拡散系列を生成: $\mathbf{d}_m \sim \text{InfDif}(\boldsymbol{\pi}_{z_m})$
 - 4: **for** $n = 1$ to N_m **do**
 - 5: 単語を生成: $w_{m,n} \sim \text{Multinomial}(\boldsymbol{\theta}_{z_m})$
 - 6: **end for**
 - 7: **end for**
-

情報 \mathbf{w}_m や拡散系列 \mathbf{d}_m は隠れ変数であるトピック z_m によって、 z_m に依存するパラメータから生成される。よって、この生成過程は、混合ユニグラムモデル [Nigam 99] の拡張とみなせる。ここで、 $\boldsymbol{\alpha}$ は K 次元の多項分布パラメータ、 $\boldsymbol{\theta}_{z_m}$ は $|W|$ 次元の多項分布パラメータ、 $\text{InfDif}(\boldsymbol{\pi}_{z_m})$ はパラメータ集合 $\boldsymbol{\pi}_{z_m}$ を持つ任意の情報拡散モデルを表す。また、 $\Theta = \{\boldsymbol{\theta}_k\}_{k=1}^K$ 、 $\Pi = \{\boldsymbol{\pi}_k\}_{k=1}^K$ とする。

以上より、 M 回の情報拡散が行われるとき、つまり、 $\mathcal{W} = \{\mathbf{w}_m\}_{m=1}^M$ と $\mathcal{D} = \{\mathbf{d}_m\}_{m=1}^M$ を観測するとき、情報 \mathcal{W} と拡散系列 \mathcal{D} に関する尤度は式 (1) となる。

$$P(\mathcal{W}, \mathcal{D} | \boldsymbol{\alpha}, \Pi, \Theta) = \prod_{m=1}^M \sum_{z_m} P(z_m | \boldsymbol{\alpha}) P(\mathbf{d}_m | \boldsymbol{\pi}_{z_m}) P(\mathbf{w}_m | \boldsymbol{\theta}_{z_m}) \quad (1)$$

ここで、式 (1) 中の因子はそれぞれ以下のようになる。

$$P(z_m | \boldsymbol{\alpha}) = \alpha_{z_m}, \quad P(\mathbf{w}_m | \boldsymbol{\theta}_{z_m}) = \prod_{w \in \mathbf{w}_m} \theta_{z_m, w}$$

ただし、 $\sum_{z_m} \alpha_{z_m} = 1$ 、 $\sum_{w \in W} \theta_{z_m, w} = 1$ である。 $P(\mathbf{d}_m | \boldsymbol{\pi}_{z_m})$ は採用する情報拡散モデルに依存する式となる。

2.1 パラメータ推定法

トピック依存情報拡散モデルのパラメータ $\boldsymbol{\alpha}, \Theta$ を EM アルゴリズムに基づくアルゴリズムによって推定する。最初に、式 (1) の対数尤度を、イエンセンの不等式から以下のように書き直す。

$$\begin{aligned} \log P(\mathcal{W}, \mathcal{D} | \boldsymbol{\alpha}, \Pi, \Theta) &= \sum_{m=1}^M \log \sum_{z_m} P(z_m, \mathbf{d}_m, \mathbf{w}_m) \\ &\geq \sum_{m=1}^M \sum_{z_m} q(z_m | \mathbf{d}_m, \mathbf{w}_m) \log \frac{P(z_m, \mathbf{d}_m, \mathbf{w}_m)}{q(z_m | \mathbf{d}_m, \mathbf{w}_m)} \\ &\propto \sum_{m=1}^M \sum_{z_m} q(z_m | \mathbf{d}_m, \mathbf{w}_m) \log P(z_m, \mathbf{d}_m, \mathbf{w}_m) \\ &= Q(\boldsymbol{\alpha}, \Pi, \Theta | \bar{\boldsymbol{\alpha}}, \bar{\Pi}, \bar{\Theta}) \end{aligned} \quad (2)$$

ここで、

$$P(z_m, \mathbf{d}_m, \mathbf{w}_m) = P(z_m | \boldsymbol{\alpha}) P(\mathbf{d}_m | \boldsymbol{\pi}_{z_m}) P(\mathbf{w}_m | \boldsymbol{\theta}_{z_m}), \quad (3)$$

$$q(z_m | \mathbf{d}_m, \mathbf{w}_m) = \frac{P(z_m | \boldsymbol{\alpha}) P(\mathbf{d}_m | \boldsymbol{\pi}_{z_m}) P(\mathbf{w}_m | \boldsymbol{\theta}_{z_m})}{\sum_{z'_m} P(z'_m | \boldsymbol{\alpha}) P(\mathbf{d}_m | \boldsymbol{\pi}_{z'_m}) P(\mathbf{w}_m | \boldsymbol{\theta}_{z'_m})} \quad (4)$$

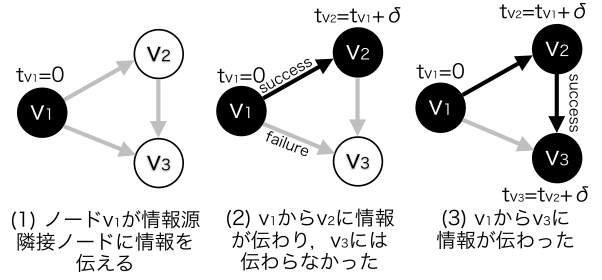


図 1: IC モデルの情報拡散の流れ

である。 $P(z_m, \mathbf{d}_m, \mathbf{w}_m)$ と $q(z_m | \mathbf{d}_m, \mathbf{w}_m)$ はパラメータに依存する式だが、表記が長くなるため省略した。

パラメータ推定は、式 (2) をそれぞれのパラメータについて最適化することにより行う。結果として、以下のパラメータ更新式が導出できる。

$$\alpha_{z_m} = \frac{1}{M} \sum_{m=1}^M q(z_m | \mathbf{d}_m, \mathbf{w}_m) \quad (5)$$

$$\theta_{z_m, w} = \frac{\sum_{m=1}^M \sum_{x \in \mathbf{w}_m, x=w} q(z_m | \mathbf{d}_m, \mathbf{w}_m)}{\sum_{m=1}^M \sum_{w' \in \mathbf{w}_m} q(z_m | \mathbf{d}_m, \mathbf{w}_m)} \quad (6)$$

式 (2), (5), (6) を対数尤度 $\log P(\mathcal{W}, \mathcal{D} | \boldsymbol{\alpha}, \Pi, \Theta)$ の上昇が収束するまで順番に繰り返し計算を行う。情報拡散モデルのパラメータ Π の更新式はモデルに依存する形になるため、ここでは書かない。次節以降では、具体的に、 $P(\mathbf{d}_m | \boldsymbol{\pi}_{z_m})$ が独立カスケードモデルの尤度となる場合を考え、このモデルパラメータの更新式の導出をする。

3. トピック依存独立カスケードモデル

前節で説明したトピック依存情報拡散モデルの枠組みを独立カスケードモデルに適用する。これによって構築したモデルを、トピック依存独立カスケード (TIC) モデルと呼ぶ。以下では、TIC モデルによる情報拡散過程を説明した後、TIC モデルのパラメータ推定法について言及する。

3.1 情報拡散過程

ノード V とリンク $E (\subset V \times V)$ で構成される自己リンクなしのネットワーク $G = (V, E)$ 上で情報拡散が起きるとする。ノードが情報を持っている状態のことをアクティブ、情報を持っていない状態のことを非アクティブと呼ぶ。

まず、図 1 を使って、独立カスケードモデルの情報拡散過程を説明する。IC モデルでは、事前に、各リンク $(u, v) \in E$ に対して拡散確率 $\kappa_{u,v,z} (0 \leq \kappa_{u,v,z} \leq 1)$ と、時間遅れパラメータ $r_{u,v} (> 0)$ を割り当てる。情報源ノードが v_1 のとき、 v_1 は子ノード v_2, v_3 へとリンクごとの拡散確率に従って情報を伝える試行を行う (図 1(1))。その結果、 v_2 に情報が伝わったとする。このとき、 v_2 がアクティブになる時刻は、 r_{v_1, v_2} をパラメータとする指数分布からサンプリングされて決まる (図 1(2))。 v_2 がアクティブになる時刻になると、 v_2 は子ノード v_3 へ情報拡散試行を行う。その結果、 v_3 へと情報を伝えることに成功し、指数分布から決定された時刻でアクティブになる (図 1(3))。最終的に、次の時刻で新しくアクティブになるノードがいなくなったら、情報拡散は終了する。

TIC モデルの場合では、事前に各リンク $(u, v) \in E$ に対してトピック z ごとに異なる拡散確率 $\kappa_{u,v,z}$ ($0 \leq \kappa_{u,v,z} \leq 1$) と、時間遅れパラメータ $r_{u,v,z}$ (> 0) を割り当てる。この2種類のパラメータは π_{z_m} に相当する。TIC モデルと IC モデルの違いは、TIC モデルでは情報拡散の初めに、どのトピック z の情報を発信するかを決める点である。 z が決定したら、以後は、 z に対応するパラメータ $\kappa_{u,v,z}, r_{u,v,z}$ に従って IC モデルと同様の方法で情報拡散が行われる。

3.2 尤度の定式化とパラメータ推定

独立カスケードモデルの尤度 $P(\mathbf{d}_m | \pi_{z_m})$ について簡単に説明する。本研究では、[Saito 09] での非同期独立カスケードモデルの定式化に準じているため、詳細な説明はここではしない。

TIC モデルにおいて、リンク $(u, v) \in E$ の両端がともにアクティブで、 $t_{m,u} < t_{m,v}$ のとき、トピック z の情報が u が v をアクティブにさせる確率密度は、

$$\Phi_{u,v,z}^{(m)} = \kappa_{u,v,z} r_{u,v,z} \exp(-r_{u,v,z}(t_{m,v} - t_{m,u})). \quad (7)$$

また、ノード v が時刻区間 $[t_{m,u}, t_{m,v}]$ 内で親ノード $u \in B(v) \cap C_m(t_{m,v})$ によってアクティブにされなかった確率は、

$$\Psi_{u,v,z}^{(m)} = \kappa_{u,v,z} \exp(-r_{u,v,z}(t_{m,v} - t_{m,u})) + (1 - \kappa_{u,v,z}). \quad (8)$$

そして、拡散系列 \mathbf{d}_m の独立カスケードモデルの尤度は以下の式で表すことが出来る。

$$\begin{aligned} P(\mathbf{d}_m | \pi_{z_m}) &= \prod_{v \in C_m} \sum_{u \in B(v) \cap C_m(t_{m,v})} \Phi_{u,v,z}^{(m)} \prod_{x \in B(v) \cap C_m(t_{m,v}) \setminus \{u\}} \Psi_{x,v,z}^{(m)} \\ &\times \prod_{v \in C_m} \prod_{w \in F(v) \setminus C_m} (1 - \kappa_{v,w,z}) \end{aligned} \quad (9)$$

式 (2) に式 (9) を代入し、TIC モデルのパラメータ α, Θ, Π を推定する。ただし、 α, Θ はそれぞれ式 (5), (6) で計算可能なため、ここでは特に、拡散確率 $\kappa_{u,v,z}$ と時間遅れパラメータ $r_{u,v,z}$ の更新式を説明する。

式 (5), (6) の導出と同様に、 $\kappa_{u,v,z}$ と $r_{u,v,z}$ で最適化することにより、以下の更新式を導出する。

$$\kappa_{u,v,z} = \frac{\sum_{m=1}^M q(z_m | \mathbf{d}_m, \mathbf{w}_m) \left\{ \lambda_{u,v,z}^{(m)} + (1 - \lambda_{u,v,z}^{(m)}) \beta_{u,v,z}^{(m)} \right\}}{|\{m \in \{1, \dots, M\}; u \in C_m\}|} \quad (10)$$

$$r_{u,v,z} = \frac{\sum_{m=1}^M q(z_m | \mathbf{d}_m, \mathbf{w}_m) \lambda_{u,v,z}^{(m)}}{\sum_{m=1}^M q(z_m | \mathbf{d}_m, \mathbf{w}_m) (\lambda_{u,v,z}^{(m)} + (1 - \lambda_{u,v,z}^{(m)}) \beta_{u,v,z}^{(m)}) (t_{m,v} - t_{m,u})} \quad (11)$$

ここで、 $\lambda_{u,v,z}^{(m)}$ と $\beta_{u,v,z}^{(m)}$ はそれぞれ、

$$\lambda_{u,v,z}^{(m)} = \frac{\Phi_{u,v,z}^{(m)} (\Psi_{u,v,z}^{(m)})^{-1}}{\sum_{u' \in B(v) \cap C_m(t_{m,v})} \Phi_{u',v,z}^{(m)} (\Psi_{u',v,z}^{(m)})^{-1}}, \quad (12)$$

$$\beta_{u,v,z}^{(m)} = \frac{\kappa_{u,v,z} \exp(-r_{u,v,z}(t_{m,v} - t_{m,u}))}{\kappa_{u,v,z} \exp(-r_{u,v,z}(t_{m,v} - t_{m,u})) + (1 - \kappa_{u,v,z})}. \quad (13)$$

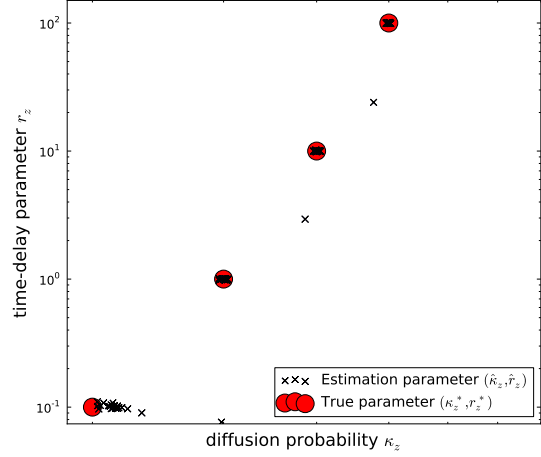


図 2: 人工データのパラメータ κ_z, r_z 推定結果

4. 実験

TIC モデルを評価するために、人工データと Twitter のリツイートデータを使った実験を行う。実験 1 では、我々が与えたパラメータを使って TIC モデルの生成過程によって生成した拡散系列から、パラメータを復元できることを確かめる。実験 2 では、Twitter リツイートデータから TIC モデルのパラメータを推定し可視化することで、構築されたクラスタが定性的に有意義であることを示す。なお、本実験では、各リンクのパラメータをトピックにのみ依存するようにする。つまり、

$$\forall (u, v) \in E, \forall z \in \{1, \dots, K\}, \kappa_{u,v,z} = \kappa_z, r_{u,v,z} = r_z$$

である。この単純化は、パラメータを頑健に推定する上で不可欠で、他の研究でも取り入れられている [Saito 09]。

4.1 実験データ

実験では、人工データと Twitter リツイートデータの 2 種類のデータを使用する。人工データのネットワーク G は BA モデル [Barabási 99] によって生成し、ノード数は 10,000、リンク数は 199,800 である。Twitter リツイートデータでは、ネットワーク G は、Twitter 上でユーザ A からユーザ B へと Reply が一回以上あった場合、A から B へとリンクを張って構築した Reply ネットワークである。Reply ネットワークは、ノード数 2,366,252、リンク数 5,546,320 で構成される。なお、Twitter 上で情報が流れる向きは Reply ネットワークのリンクの逆向きであるため、実験に使用する際は、Reply ネットワークは逆向きに変換する。拡散系列は、#c80, #tpp, #anti.tpp, #seiji, #news, #AKB48, #earthquake の 7 種類のハッシュタグが付けられた情報に関して、10 回以上リツイートされた 858 系列を使用する*1。

4.2 実験 1: 人工データによるパラメータ復元

$M = 100$ の拡散系列 \mathcal{D} を 20 個生成した後、各拡散系列から情報拡散パラメータ κ_z, r_z を推定し、その精度を評価する。その際、拡散系列 \mathcal{D} は、図 2 の True parameter で示す κ_z^*, r_z^* のパラメータを持つトピック数 $K = 4$ の TIC モデルから生成した。また、情報 \mathcal{W} は全ての要素が 0.5 のパラメータを持つディリクレ分布からサンプリングした θ_z によって自動的に生

*1 一つの情報に複数のハッシュタグが付くケースもある。また、情報 \mathcal{W} はハッシュタグを示す文字列を含まない。

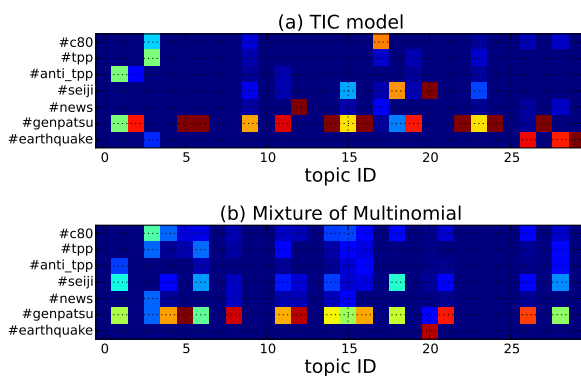


図 3: ハッシュタグの各トピックへの寄与度の可視化結果

成し、語彙数 $|W| = 1000$ 、情報に含まれる単語数 $N_m = 100$ とした。

図 2 を見て分かるように、各トピックのパラメータ推定値 $\hat{\kappa}_z, \hat{r}_z$ は、真のパラメータ κ_z^*, r_z^* の近傍に分布し、トピック間のパラメータの違いをはっきりと示すことができる。しかし、パラメータの初期値によっては局所解に落ちることがあり、この場合は、正しく推定できていない。

4.3 実験 2: Twitter リツイートデータのパラメータ可視化

Twitter リツイートデータに対して、TIC モデルによってパラメータ推定を行った。その後、各拡散系列のトピックに対する寄与度を式 (4) で計算し、各トピックに含まれるハッシュタグの割合を図 3 で示した。図 3 上部は TIC モデルの結果、下部は比較のために、TIC モデルと同様のことを混合多項分布で行った結果である。混合多項分布は、式 (1) の $P(\mathbf{d}_m | \boldsymbol{\pi}_{z_m})$ を取り除いたものである。

混合多項分布の場合、すべてのトピックで様々なハッシュタグが混在している。これは、Twitter の性質上、少数の単語列からトピックの割り当てを決めなければならず、正確にトピックを分けることが難しいためだと考えられる。また、地震 (#earthquake) のトピックは一つのトピックに集中している。これは、このハッシュタグが付いたツイートは、定型文の地震情報が多いためである。一方で、提案法である TIC モデルでは、各トピックは 1,2 個のハッシュタグで構成されている。これは、情報の単語列のみだけでなく拡散系列をパラメータ推定に使っているため、拡散系列の情報がトピックを決める補助的な役割を果たしているからと考えられる。また、地震のトピックは 3 つのトピックに分かれた。これは、情報拡散パラメータの違いが表れたためで、詳細に調べてみると、各トピックで情報発信者が異なっていた。したがって、情報発信者の影響力の違いによって、トピックが分かれたと見ることができる。また、TIC モデルでは新たに TPP に関する話題 (#tpp) と TPP 反対派の話題 (#anti.tpp)、そして、コミックマーケット 80 の話題*2 (#c80) のトピックを発見できている。

総合的に見て、TIC モデルではトピックが明確に分かれる傾向があるといえる。これによって、新たに拡散系列とその情報の内容が与えられたとき、それがどのトピックに属してどのような伝わり方をするのか判断しやすいという長所がある。また、応用として情報拡散シミュレーションによって情報が何人

に伝わるか (影響度) を予測する問題においては、情報の内容だけではどのトピックか判断できず正しい情報拡散パラメータでシミュレーションできない可能性があることが示唆された一方で、拡散早期の拡散系列を使うことで、適切なトピックへと誘導できることが期待できる。

5. おわりに

本研究では、情報の内容を考慮するトピック依存情報拡散モデルの枠組みを提案した。そして、この枠組みを独立カスケードモデルに適用し、新たに構築したトピック依存独立カスケード (TIC) モデルのパラメータ推定法を EM アルゴリズムに基づいて導出した。実験では、TIC モデルの生成過程で人工的に生成したデータから、TIC モデルのパラメータ推定法によって、元のパラメータを復元できることを示した。また、Twitter リツイートデータに対して TIC モデルのパラメータ推定法を適用し、混合多項分布の場合と比べ、意味あるトピックを抽出できることを示した。今後は、線形閾値モデル [Saito 10] や投票者モデル [Kimura 10] といった情報拡散モデルに対してもトピック依存情報拡散モデルが適用可能であることを実証していくとともに、情報拡散モデルに基づく影響度予測 [Yoshikawa 10] にトピック依存情報拡散モデルを適用して、現実のデータで影響度が予測可能か調べていく。

参考文献

- [Barabási 99] Barabási, A. and Albert, R.: Emergence of scaling in random networks, *science*, Vol. 286, No. 5439, pp. 509–512 (1999)
- [Kempe 03] Kempe, D., Kleinberg, J., and Tardos, É.: Maximizing the spread of influence through a social network, in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 137–146 ACM (2003)
- [Kimura 10] Kimura, M., Saito, K., Ohara, K., and Motoda, H.: Learning to predict opinion share in social networks, in *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI 2010)*, pp. 1364–1370 (2010)
- [Nigam 99] Nigam, K., McCallum, A. K., Thrun, S., and Mitchell, T.: Text Classification from Labeled and Unlabeled Documents using EM, in *Machine Learning*, pp. 103–134 (1999)
- [Saito 09] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Learning continuous-time information diffusion model for social behavioral data analysis, *Advances in Machine Learning*, pp. 322–337 (2009)
- [Saito 10] Saito, K., Kimura, M., Ohara, K., and Motoda, H.: Behavioral analyses of information diffusion models by observed data of social network, *Advances in Social Computing*, pp. 149–158 (2010)
- [Yoshikawa 10] Yoshikawa, Y., Saito, K., Motoda, H., Ohara, K., and Kimura, M.: Acquiring expected influence curve from single diffusion sequence, *Knowledge Management and Acquisition for Smart Systems and Services*, pp. 273–287 (2010)

*2 <http://www.comiket.co.jp/>