

ユーザの潜在特徴を考慮した ソーシャルネットワーク上の情報拡散モデル

吉川 友也^{1,a)} 岩田 具治^{2,b)} 澤田 宏^{3,c)}

受付日 2013年6月21日, 採録日 2013年8月12日

概要: オンラインコミュニティ上では, 流行するトピックに対する意見や購買行動, 噂などの情報が友人関係などを通じて広がり, 共有されている. このようなソーシャルネットワーク上の情報拡散をモデル化することにより, 将来流行する話題の予測や, 影響度の大きいユーザの抽出などが可能になる. 本稿では, 各ユーザに潜在する特徴をとらえることによって, ユーザ間の情報の伝わりやすさを表す伝播確率をより正確に推定できる情報拡散モデルを提案する. 具体的には, ユーザは情報送信者としての潜在特徴と情報受信者としての潜在特徴を持つと仮定し, 伝播確率は送受信両ユーザの潜在特徴に依存して決まるとする. 実験により, 提案モデルは従来モデルと比較して, 少ない観測データの場合でもパラメータを正確に推定できることを示す. また, 実データを使った実験により, 提案モデルを使うことによって, 未知の拡散現象に対する予測性能が向上することを示す.

キーワード: 情報拡散モデル, 潜在特徴モデル, ソーシャルネットワーク, 独立カスケードモデル

Information Diffusion Model with Latent Features of Users on Social Networks

YUYA YOSHIKAWA^{1,a)} TOMOHARU IWATA^{2,b)} HIROSHI SAWADA^{3,c)}

Received: June 21, 2013, Accepted: August 12, 2013

Abstract: People share various types of information including opinions on hot topics, bookmarking activities and rumors via online communities. To make it possible to predict future trends in online communities, it is important to develop a model of information diffusion through social networks and a method for estimating its parameters. In this paper, we present an information diffusion model, which can effectively estimate diffusion probabilities by capturing the latent features of users. In particular, we assume each user has features as information sender and receiver, and diffusion probabilities depend on the features of both the sender and receiver nodes. We demonstrate experimentally that the proposed model can estimate the diffusion probabilities more accurately than commonly used methods. In the experiments using real datasets, we show the proposed model can improve the predictive performance for unknown diffusion phenomena.

Keywords: information diffusion model, latent feature model, social network, independent cascade model

¹ 奈良先端科学技術大学院大学
Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan

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, Seika, Kyoto 619-0237, Japan

³ 日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, Yokosuka, Kanagawa 239-0847, Japan

a) yuya-y@is.naist.jp

b) iwata.tomoharu@lab.ntt.co.jp

c) sawada.hiroshi@lab.ntt.co.jp

1. はじめに

マイクロプロダクトサービスやソーシャルネットワーキングサービス (SNS) は, 多くのユーザにとって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. このようなソーシャルネットワーク上では, ユーザは情報 (たとえば, 噂, 新商品のレビュー, デマなど) を発信・共有し, その情報はユーザ間のリンクを介して広がる. 最近では, ソーシャルネットワークでのユーザの言動から映画

の興行収入 [15] や株価 [20], 選挙の得票数 [2] を予測する研究が一定の成果をあげており, ソーシャルネットワーク上の情報拡散現象から将来の流行を予測するための研究がさかんに行われている [3], [6], [11], [13].

情報拡散研究でよく利用される基本的な確率モデルは, 独立カスケードモデル (ICM) と線形閾値モデル (LTM) である [8]. ICM は, リンクに割り当てられた伝播確率に従って, 情報を持つユーザからそのリンク先のユーザへ他とは独立に情報が伝わるとする. LTM は, 受け取った重みの合計がそのユーザに割り当てられた閾値を超えたとき情報を受け取るようにモデル化されている. また最近では, 意見の移り変わりをモデル化した投票者モデル [9], [18] や, 特定のソーシャルネットワーク以外からの影響を考慮したモデル [14] が存在する. なお, 本研究では, 最も基本的かつ情報拡散過程としてよく利用されているという理由から, ICM に基づく情報拡散モデルを扱う.

ICM に基づく研究では, 情報拡散による影響を最大 (最小) 化させるノード集合の発見 [8], [10] や, 時間ごとのノードの影響度推定 [21] などが行われてきた. これらの研究はシミュレーションベースであり, ICM の挙動を制御するパラメータを事前に決定する必要がある. しかし, 実際にはパラメータは直接的に観測できない. したがって, ある情報を誰がいつ受け取ったのかを記録したログデータから, ICM のパラメータを推定する必要がある.

ICM のパラメータを推定する方法はいくつか存在する [1], [5], [16], [17]. これらの方法は, リンクに付与されるパラメータを最尤推定によって直接推定している. しかし, 現実のソーシャルネットワークのようなリンク数が膨大なネットワークでは, 推定に十分なデータを得ることが難しく, 正確にパラメータを推定することが困難である. そのため, 全リンク共通のパラメータを持つように ICM を単純化する場合 [21] もあるが, これはモデルの柔軟性を著しく損う欠点がある.

本稿では, 各ユーザに潜在する特徴をとらえることによって, ICM におけるユーザ間の情報の伝わりやすさを表す伝播確率を正確に推定できる情報拡散モデルを提案する. 具体的には, ユーザは情報送信者としての特徴と情報受信者としての特徴を持つと仮定し, 伝播確率は送受信両ユーザの特徴に依存して間接的に決まるとする. また, EM アルゴリズムを用いた事後確率最大化によるパラメータ推定法を導出する. 実験により, 提案モデルは従来モデルと比較して, 少ない観測データからでもパラメータを正確に推定できることを示す. また, 実データを使った実験では, 提案モデルを使うことによって, 未知の拡散現象に対する予測性能が向上することを示す.

以下の本文では, まず 2 章で ICM が仮定する情報拡散過程について述べ, 3 章では ICM に基づく関連研究を紹介する. 続く 4 章では提案法について述べ, 5 章では提案

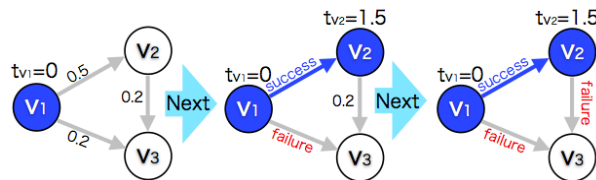


図 1 独立カスケードモデルの拡散過程の例. (左) 初期状態. 各リンクには事前に伝播確率が割り当てられている. 情報源ノード v_1 から拡散が開始する. (中央) 拡散試行の結果, ノード v_2 がノード v_1 から影響を受けて時刻 t_{v_2} でアクティブになった. ノード v_3 への拡散は失敗した. (右) ノード v_2 はノード v_3 へと情報伝播を試みるが, 失敗した

Fig. 1 An example of Independent Cascade Model. (left) An initial state. A diffusion probability for each link is assigned in advance. An information diffusion starts from information source node v_1 . (center) As a result of the diffusion trial, node v_2 became active at time t_{v_2} by influenced from node v_1 , but node v_3 did not. (right) Then node v_2 tried to affect node v_3 , but the trial failed.

法と従来法を定量的に比較する. 最後に 6 章で, 結論と今後の課題を述べる.

2. 情報拡散過程

ユーザを表すノード集合 \mathcal{V} と, ユーザ間の関係を表すリンク集合 $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ からなる有向グラフ $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 上で情報拡散が行われるとする. 本研究で使用する情報拡散過程 ICM は, 各リンク $(u, v) \in \mathcal{E}$ の伝播確率 κ_{uv} と時間遅れパラメータ r_{uv} の 2 種類のパラメータを持つ. 簡便のため, $\kappa = \{\kappa_{uv}\}_{(u,v) \in \mathcal{E}}, r = \{r_{uv}\}_{(u,v) \in \mathcal{E}}$ とする. ここで, ノードが情報を持っている状態をアクティブ, 情報を持っていない状態を非アクティブと呼ぶことにする. また, ICM では 1 度アクティブになったノードは 2 度と非アクティブには戻らない.

今, パラメータは既知とする. ICM では, 最初に情報源ノード集合 $S \subseteq \mathcal{V}$ を設定する. これらのノードから情報拡散が開始され, 次の 2 つのステップが繰り返し行われる.

- ノード u がアクティブのとき, u は 1 度だけその子ノード v へ情報を伝える試行を行う. ただし, その試行は伝播確率 κ_{uv} の確率で成功する.
- ノード v への情報拡散試行が成功したとき, v がアクティブになる時刻は $t_u + \delta$ である. ただし, δ は r_{uv} をパラメータとして持つ指数分布から生成される.

新しくアクティブになるノードがなくなったとき, ICM は終了する. 図 1 は, ICM の拡散過程の具体例を示す. ICM の詳細な手続きを Algorithm 1 で示す.

3. 関連研究

この章では, ICM に関する研究を簡単に紹介し, 本研究の立ち位置を明確にしておく. 2001 年, Goldenberg ら [4]

は、マーケットにおける新製品支持の速度とネットワークのつながり方の関係について研究した。その際使われたモデルは、確率的セル・オートマトンとして定式化された。これは、現在我々が使用している ICM と類似した定式化であるため、ICM の起源と考えられる。初めて ICM が詳細に定義されたのは、2003 年の Kempe ら [8] の研究である。その研究では、ICM の伝播確率は既知として、影響を最大化させるノード集合を発見するために ICM が使われた。反対に、好ましくない情報の広がりを最小化する研究 [10] も存在する。

実際には、ICM のパラメータである伝播確率を観測することはできない。したがって、ICM でシミュレーションを行うためや、パラメータから情報拡散の特性を知るために、ICM のパラメータ推定の研究も行われている。Gruhl ら [5] は、ICM のパラメータを推定する方法を初めて提案した。Saito ら [16] は、ICM の尤度関数を厳密に定式化し、最尤推定を行うことによるパラメータ推定法を提案した。その後、パラメータ推定精度の向上のために、Saito ら [17] はノードの属性の類似度によって伝播確率が決まる情報拡散モデルを提案している。最近では、Dickens ら [1] が伝播確率に対する事前知識を組み込んだモデルを提案している。

我々の研究もまた、ICM のパラメータ推定の方法を構築するものである。Saito ら [17] と Dickens ら [1] は、どちらもパラメータ推定精度の向上が目的という点で我々の目的と類似する。しかし、我々とこれらの研究は以下の部分が異なる。Saito ら [17] との違いは、ノードの性別や年齢などの属性情報を必要としない点である。これは我々の提案モデルが追加の観測データを必要としないことを意味する。また、Dickens ら [1] は、各リンクの伝播確率に対して 1 つのベータ分布を仮定して精度向上を試みるが、我々はノードの隣接ノード間での振舞いを潜在特徴ベクトルを介して考慮することで、精度向上を試みる点が異なる。

Algorithm 1 ICMGENERATOR

Require : ネットワーク \mathcal{G} , 情報源ノード集合 \mathcal{S} , 時間遅れパラメータ r , 伝播確率 κ

```

1:  $\mathbf{d}_m \leftarrow \{(u, 0) | u \in \mathcal{S}\}$ 
2: repeat
3:    $(u, t_{mu}) \leftarrow \min_{t_{mx}} \{(x, t_{mx}) | (x, t_{mx}) \in \mathbf{d}_m, x \in \mathcal{S}\}$ 
4:   for  $v \in \{u \text{ の子ノードで, 非アクティブなノード}\}$  do
5:     if  $u$  が  $\kappa_{uv}$  の確率で  $v$  へ拡散成功 then
6:        $t_{mv} \leftarrow t_{mu} + \delta$ , where  $\delta \sim \text{Exponential}(\delta | r_{uv})$ 
7:        $\mathbf{d}_m \leftarrow \mathbf{d}_m \cup \{(v, t_{mv})\}$ 
8:        $\mathcal{S} \leftarrow \mathcal{S} \cup \{v\}$ 
9:     end if
10:  end for
11:   $\mathcal{S} \leftarrow \mathcal{S} \setminus \{u\}$ 
12: until  $\mathcal{S} = \emptyset$ 
13: return  $\mathbf{d}_m$ 

```

4. 提案法

4.1 生成モデル

M 個の異なる情報が有向ネットワーク $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ 上で拡散する現象を考える。ノード $v \in \mathcal{V}$ の親ノード集合 $B(v)$ と子ノード集合 $F(v)$ を以下のように定義する。

$$B(v) = \{u | (u, v) \in \mathcal{E}\}, \quad F(v) = \{w | (v, w) \in \mathcal{E}\}.$$

表 1 は提案モデルで使用する記号の一覧である。

観測データとして、情報 $m \in \{1, \dots, M\}$ を発信したノード $v \in \mathcal{V}$ と発信した時刻 $t_{mv} > 0$ の組からなる拡散系列 $\mathbf{d}_m = \{(v_{m1}, t_{mv_{m1}}), (v_{m2}, t_{mv_{m2}}), \dots\}$ が得られる。なお、最初に情報を発信したノードを除き、各ノードは親ノードの 1 つからのみ影響を受けて情報を発信すると仮定し、影響を与えた親ノードは観測できないとする。簡便のため、 $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_M\}$ と表記する。

提案モデルでは、ノード $u \in \mathcal{V}$ は 2 種類の潜在特徴ベクトル $\mathbf{x}_u, \mathbf{y}_u \in \mathbb{R}^K$ を持つと仮定する。 \mathbf{x}_u はノード u の情報の送信者としての特徴ベクトル、 \mathbf{y}_u はノード u の情報の受信者としての特徴ベクトルを表す。これらはそれぞれ、以下の K 次元の正規分布に従うとする。

$$\mathbf{x}_u \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathbf{I}), \tag{1}$$

$$\mathbf{y}_u \sim \mathcal{N}(\mathbf{0}, \sigma_Y^2 \mathbf{I}). \tag{2}$$

ここで、 $\mathbf{0}$ は K 次元の零ベクトル、 \mathbf{I} は $K \times K$ の単位行列である。簡便のため、 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{|\mathcal{V}|}]$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{|\mathcal{V}|}]$ と定義する。そして、各リンク $(u, v) \in \mathcal{E}$ に付与される ICM の伝播確率 κ_{uv} は、以下の式で計算される。

$$\kappa_{uv} = f(\mathbf{x}_u, \mathbf{y}_v, \gamma) = (1 + \exp(-\mathbf{x}_u^\top \mathbf{y}_v - \gamma))^{-1}. \tag{3}$$

ここで、 γ は伝播確率の値を調整するバイアスパラメータである。また、この関数はシグモイド関数であり、 $0 \leq \kappa_{uv} \leq 1$ が保証されていることに留意されたい。関数 f は単調増加関数であり、 $\mathbf{x}_u^\top \mathbf{y}_v$ の値が大きいとき、 κ_{uv} の値もまた大

表 1 表記法

Table 1 Notation.

記号	説明
K	特徴ベクトルの長さ (次元数)
M	拡散系列の数
\mathbf{x}_u	ノード u の情報の送信者としての特徴ベクトル
\mathbf{y}_u	ノード u の情報の受信者としての特徴ベクトル
\mathbf{d}_m	m 番目の拡散系列
κ_{uv}	リンク (u, v) の伝播確率
r_{uv}	リンク (u, v) の時間遅れパラメータ
γ	伝播確率を制御するバイアスパラメータ
σ_X	\mathbf{X} の標準偏差を表すハイパーパラメータ
σ_Y	\mathbf{Y} の標準偏差を表すハイパーパラメータ

きくなる。

上記の仮定に基づいて生成された伝播確率 $\kappa = \{\kappa_{uv}\}_{(u,v) \in \mathcal{E}}$ と、時間遅れパラメータ $\mathbf{r} = \{r_{uv}\}_{(u,v) \in \mathcal{E}}$ と情報源ノード集合 \mathcal{S} が与えられたとき、拡散系列 \mathbf{d}_m は 2 章で説明した情報拡散過程 ICM に従って生成される。

$$\mathbf{d}_m \sim \text{ICM}(\mathcal{G}, \mathcal{S}, \mathbf{r}, \kappa). \quad (4)$$

潜在特徴ベクトル \mathbf{X}, \mathbf{Y} 、時間遅れパラメータ \mathbf{r} の下で、拡散系列 \mathbf{D} の尤度を考える。 $\Delta_{uv}^m = t_{mv} - t_{mu}$ と定義する。また、情報 m におけるアクティブノード集合 C_m と時刻 t までのアクティブ集合 $C_m(t)$ を以下のように定義する。

$$C_m = \{u | (u, t) \in \mathbf{d}_m\},$$

$$C_m(t) = \{u | (u, t') \in \mathbf{d}_m, t' < t\}.$$

情報 m の拡散系列において、リンク $(u, v) \in \mathcal{E}$ の両端がともにアクティブであり、かつ $\Delta_{uv}^m > 0$ のとき、ノード u がノード v をアクティブにさせる確率密度を、

$$p_{u \rightarrow v}(\Delta_{uv}^m) = \kappa_{uv} r_{uv} \exp(-r_{uv} \Delta_{uv}^m) \quad (5)$$

と定義する。ICM では、ノードは親ノードの 1 つのみから影響を受けアクティブになると仮定している。したがって、リンク $(u, v) \in \mathcal{E}$ の両端がともにアクティブでも、ノード u からノード v への情報伝播に失敗する可能性を考える必要がある。式 (5) を使って、アクティブなノード v が親ノード $u \in B(v) \cap C_m(t_{mv})$ によってアクティブにされなかった確率は、

$$p_{u \rightarrow v}(\Delta_{uv}^m) = 1 - \int_0^{t_{mv} - t_{mu}} p_{u \rightarrow v}(t) dt$$

$$= \kappa_{uv} \exp(-r_{uv} \Delta_{uv}^m) + (1 - \kappa_{uv}) \quad (6)$$

と書くことができる。以上より、情報 m におけるアクティブなノード v の尤度は、親ノード中の 1 つから情報伝播が成功し、その他の親ノードからは失敗する確率をすべての組合せで足し合わせることで求められる。

$$L_{mv}^+ = \sum_{u \in B(v) \cap C_m(t_{mv})} p_{u \rightarrow v}(\Delta_{uv}^m) \prod_{\substack{x: x \neq u, \\ x \in B(v) \cap C_m(t_{mv})}} p_{u \rightarrow v}(\Delta_{uv}^m). \quad (7)$$

一方で、ノード v が親ノードのいずれから影響を受けずアクティブにならなかった場合の尤度は、拡散に失敗する確率の総乗で表現できる。

$$L_{mv}^- = \prod_{u \in B(v) \cap C_m} p_{u \rightarrow v}(T_m - t_{mu}). \quad (8)$$

ここで、 T_m は情報 m における観測の打ち切り時刻と定義する。打ち切り時刻 T_m が十分に大きいならば、 $T_m \rightarrow \infty$ と見なせる。その結果、式 (7) の定義より式 (8) は

$$L_{mv}^- = \prod_{u \in B(v) \cap C_m} 1 - \kappa_{uv} \quad (9)$$

と書き換えることができる。

拡散系列 \mathbf{d}_m の尤度は、拡散の成功の尤度 (7) と失敗の尤度 (9) を掛け合わせて計算される。したがって、拡散系列集合 \mathbf{D} の尤度は以下の式となる。

$$P(\mathbf{D} | \mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}, \mathcal{G}) = \prod_{m=1}^M \left[\prod_{v \in C_m} L_{mv}^+ \times \prod_{v \notin C_m} L_{mv}^- \right]. \quad (10)$$

また、式 (1), (2) の仮定から、潜在特徴ベクトル \mathbf{X}, \mathbf{Y} の事前確率は、

$$P(\mathbf{X} | \sigma_X) = \prod_{u=1}^{|\mathcal{V}|} \frac{1}{\sqrt{2\pi}\sigma_X^K} \exp\left(-\frac{\mathbf{x}_u^\top \mathbf{x}_u}{2\sigma_X^2}\right), \quad (11)$$

$$P(\mathbf{Y} | \sigma_Y) = \prod_{u=1}^{|\mathcal{V}|} \frac{1}{\sqrt{2\pi}\sigma_Y^K} \exp\left(-\frac{\mathbf{y}_u^\top \mathbf{y}_u}{2\sigma_Y^2}\right) \quad (12)$$

である。

最終的に、拡散系列集合 \mathbf{D} が与えられたときの潜在特徴ベクトル \mathbf{X}, \mathbf{Y} の事後確率は、式 (10), (11), (12) を使用した以下の式となる。

$$P(\mathbf{X}, \mathbf{Y} | \mathbf{D}, \gamma, \mathbf{r}, \sigma_X, \sigma_Y, \mathcal{G}) \propto P(\mathbf{D} | \mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}, \mathcal{G}) P(\mathbf{X} | \sigma_X) P(\mathbf{Y} | \sigma_Y). \quad (13)$$

我々の目的は、この事後確率を最大化するパラメータ $\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}$ を推定することである。

4.2 パラメータ推定

事後確率式 (13) を最大化するパラメータ $\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}$ を最大事後確率 (MAP) 推定によって求める。提案モデルにおける MAP 推定は、Expectation Maximization (EM) アルゴリズムによって可能である。提案モデルの EM アルゴリズムでは、E ステップでアクティブノードに影響を与えた親ノードを推定することにより情報拡散の経路を特定し、M ステップでは特定した経路に基づいてパラメータを推定する。なお、混乱を避けるため、 $\bar{\mathbf{x}}$ と書くときは、パラメータ \mathbf{x} の現在の推定値を表すこととする。

E ステップ: 文献 [16] と同様の方法を適用することによって、事後確率式 (13) から完全データの事後確率の期待値関数 Q を導出する。

$$Q(\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}; \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\gamma}, \bar{\mathbf{r}}) = \sum_{m=1}^M \sum_{v \in C_m} \left[\sum_{u \in B(v) \cap C_m(t_{mv})} \left(\bar{\xi}_{muv} \log \kappa_{uv} + \bar{q}_{muv} \log r_{uv} + (1 - \bar{\xi}_{muv}) \log(1 - \kappa_{uv}) + \bar{\xi}_{muv} r_{uv} \Delta_{uv}^m \right) + \sum_{w \in F(v) \setminus C_m} \log(1 - \kappa_{vw}) \right] + \log P(\mathbf{X} | \sigma_X) + \log P(\mathbf{Y} | \sigma_Y). \quad (14)$$

ここで、

$$q_{muv} = \frac{p_{u \rightarrow v}(\Delta_{uv}^m) p_{u \leftarrow v}(\Delta_{uv}^m)^{-1}}{\sum_{u' \in B(v) \cap C_m(t_{mv})} p_{u' \rightarrow v}(\Delta_{u'v}^m) p_{u' \leftarrow v}(\Delta_{u'v}^m)^{-1}}, \quad (15)$$

$$\eta_{muv} = \frac{\kappa_{uv} \exp(-r_{uv} \Delta_{uv}^m)}{\kappa_{uv} \exp(-r_{uv} \Delta_{uv}^m) + (1 - \kappa_{uv})}, \quad (16)$$

$$\xi_{muv} = q_{muv} + (1 - q_{muv}) \eta_{muv} \quad (17)$$

である。ここで、関数 Q は事後確率式 (13) の下限であることが保証されている。すなわち、

$$P(\mathbf{X}, \mathbf{Y} | \mathbf{D}, \gamma, \mathbf{r}, \sigma_X, \sigma_Y, \mathcal{G}) \geq Q(\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}; \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\gamma}, \bar{\mathbf{r}})$$

である。また、 q_{muv} は m 番目の拡散系列において、ノード v がノード u から影響を受けてアクティブになった確率を表す。したがって、Eステップでは、拡散系列から情報拡散の経路を推定していると見なすことができる。次のMステップでは、関数 Q を最大化するようなパラメータ \mathbf{X} , \mathbf{Y} , γ , \mathbf{r} を求める。

Mステップ: 現在の推定値 $\bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\gamma}, \bar{\mathbf{r}}, \bar{q}_{muv}, \bar{\eta}_{muv}, \bar{\xi}_{muv}$ に基づいて、パラメータ $\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}$ を推定する。シグモイド関数の非線形性により、パラメータ $\mathbf{X}, \mathbf{Y}, \gamma$ に対する閉形式の解は存在しない。したがって、これらのパラメータの推定には、パラメータに関する勾配に基づく最適化法を使う必要がある。本研究では、一階微分のみを必要とする準ニュートン法 (L-BFGS) を用いる。ノード $u \in \mathcal{V}$ に対するパラメータ $\mathbf{x}_u, \mathbf{y}_u$ と、パラメータ γ に関する一階微分は以下の式である。

$$\frac{\partial Q(\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}; \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\gamma}, \bar{\mathbf{r}})}{\partial \mathbf{x}_u} \quad (18) = \sum_{m=1}^M \left(\sum_{v \in F(u) \cap C_m} (\bar{\xi}_{muv} - f(\mathbf{x}_u, \bar{\mathbf{y}}_v, \bar{\gamma})) \bar{\mathbf{y}}_v - \sum_{w \in F(u) \setminus C_m} f(\mathbf{x}_u, \bar{\mathbf{y}}_w, \bar{\gamma}) \bar{\mathbf{y}}_w \right) - \frac{1}{\sigma_X^2} \mathbf{x}_u,$$

$$\frac{\partial Q(\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}; \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\gamma}, \bar{\mathbf{r}})}{\partial \mathbf{y}_v} \quad (19) = \sum_{m=1}^M \left(\sum_{u \in B(v) \cap C_m(t_{mv})} (\bar{\xi}_{muv} - f(\bar{\mathbf{x}}_u, \mathbf{y}_v, \bar{\gamma})) \bar{\mathbf{x}}_u - \sum_{s \in B(v) \cap C_m, v \notin C_m} f(\bar{\mathbf{x}}_s, \mathbf{y}_v, \bar{\gamma}) \bar{\mathbf{x}}_s \right) - \frac{1}{\sigma_Y^2} \mathbf{y}_v,$$

$$\frac{\partial Q(\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}; \bar{\mathbf{X}}, \bar{\mathbf{Y}}, \bar{\gamma}, \bar{\mathbf{r}})}{\partial \gamma} \quad (20) = \sum_{m=1}^M \left(\sum_{v \in C_m} \sum_{u \in B(v) \cap C_m(t_{mv})} (\bar{\xi}_{muv} - f(\bar{\mathbf{x}}_u, \bar{\mathbf{y}}_v, \gamma)) - \sum_{v \in C_m} \sum_{s \in F(v) \setminus C_m} f(\bar{\mathbf{x}}_s, \bar{\mathbf{y}}_v, \gamma) \right).$$

Algorithm 2 LFICMESTIMATOR

Require: ネットワーク \mathcal{G} , 拡散系列 \mathbf{D} , 特徴ベクトルの長さ K , ハイパーパラメータ σ_X, σ_Y
 1: $\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}$ を初期化
 2: **repeat**
 3: **E**ステップ: 式 (15), (16), (17) を使って, $q_{muv}, \eta_{muv}, \xi_{muv}$ を更新
 4: **M**ステップ: 式 (18), (19), (20) を使った準ニュートン法によって $\mathbf{X}, \mathbf{Y}, \gamma$ を更新. 式 (21) を使って \mathbf{r} を更新.
 5: **until** 事後確率式 (13) の値の上昇が収束
 6: **return** $\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}$

時間遅れパラメータ \mathbf{r} は閉形式の解を持つので、各リンク $(u, v) \in \mathcal{E}$ について以下の式で更新する。

$$r_{uv} = \frac{\sum_{m=1}^M \mathbf{1}(v \in C_m, u \in C_m(t_{mv})) \bar{q}_{muv}}{\sum_{m=1}^M \mathbf{1}(v \in C_m, u \in C_m(t_{mv})) \bar{\xi}_{muv} \Delta_{uv}^m}. \quad (21)$$

ここで、 $\mathbf{1}(A)$ は条件 A が真のとき 1, そうでなければ 0 を返す関数である。

EM アルゴリズムは、上記で説明した E ステップと M ステップを交互に繰り返し計算する。パラメータ $\mathbf{X}, \mathbf{Y}, \gamma, \mathbf{r}$ は、この計算を事後確率式 (13) の値の上昇が収束するまで続けることによって推定される。Algorithm 2 は、パラメータ推定の手続きを示す。

4.3 潜在特徴ベクトルの導入による利点

本提案法では、新たに 2 種類の潜在特徴ベクトル $\mathbf{x}_u, \mathbf{y}_u$ を導入した。これによるモデルの学習上の利点を示す。伝播確率のみをパラメータとして持つ ICM を考える。この場合、ICM は各リンクで 1 つの伝播確率が割り当てられるので、Saito ら [16] の方法における自由パラメータの数は $\mathcal{O}(|\mathcal{V}|^2)$ となる。一方で、提案法は伝播確率を直接推定せず、各ノードの潜在特徴ベクトルを推定することによって間接的に伝播確率を計算する。したがって、自由パラメータの数は $\mathcal{O}(|\mathcal{V}|K)$ である。一般的に、 $|\mathcal{V}| \gg K$ なので、提案法は従来法よりも少ないパラメータの推定となる。その結果、特にパラメータ推定に十分な拡散系列が得られないとき、より正確な伝播確率の計算につながる。

5. 実験

5.1 人工データによるパラメータ推定精度の評価

実ネットワークデータと人工で生成した拡散系列を使用して、パラメータ推定精度の評価を行った。

ネットワーク構造: 3 種類の実ネットワーク構造を使用した。KARATE は、1970 年代のアメリカの大学の空手クラブのメンバー間の友人関係のデータセットである。このデータセットは、最初に文献 [19] で使われ、それ以降もネットワークコミュニティのモデル化で頻繁に使われている。BLOG は、ブログのトラックバックネットワークのデータセットである。このデータは、2005 年 5 月の goo ブログ内のトラックバックを追跡して取得したものである。MIXI

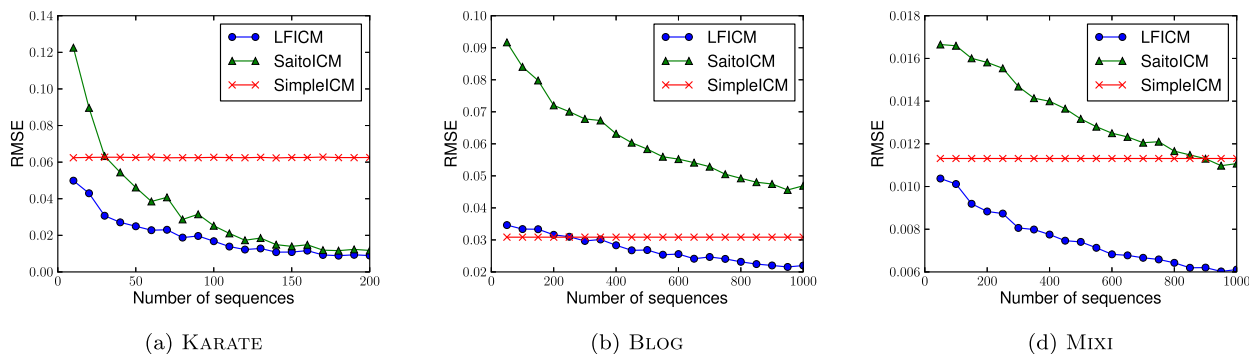


図 2 真の伝播確率と推定した伝播確率の平均 RMSE

Fig. 2 Average RMSE between true and estimated diffusion probabilities.

表 2 ネットワーク統計量と人工データのパラメータ

Table 2 Network statistics and parameters for synthetic data.

	ノード数	リンク数	平均 出次数	K	γ	α
KARATE	34	156	4.59	3	-1.5	0.1
BLOG	12,047	79,920	6.63	5	-3.5	0.1
MIXI	80,608	571,136	7.09	7	-6.0	0.01

は、2010 年におけるミクシィ内のマイミク関係の一部を収集したネットワークのデータセットである。マイミクはユーザ間の合意によって作られる関係のため、すべて双方向リンクとなっている。表 2 は各データセットの統計量を示す。

パラメータと拡散系列の生成：各ノード $u \in \mathcal{V}$ について、潜在特徴ベクトル $\mathbf{x}_u, \mathbf{y}_u$ は、平均が K 次元零ベクトル $\mathbf{0}$ 、共分散行列が $K \times K$ の単位行列 \mathbf{I} の K 次元正規分布 $\mathcal{N}(\mathbf{0}, \mathbf{I})$ から生成した。その後、各リンク $(u, v) \in \mathcal{E}$ について、式 (3) を使って、真の伝播確率 $\kappa_{uv}^* = f(\mathbf{x}_u, \mathbf{y}_v, \gamma) + \epsilon$ を計算した。ここで、 ϵ は平均 0、分散 α の正規分布に従うノイズを表す。もし、生成した伝播確率が $0 < \kappa_{uv} < 1$ でなければ、 ϵ を再生成する。各ネットワークでの K, γ, α の値は表 2 に示す。拡散系列 \mathbf{d}_m は、一様ランダムに 1 つの情報源ノードを選び、Algorithm 1 を適用し生成した。なお、時間遅れパラメータ \mathbf{r} は各リンク共通で 1.0 に設定した。拡散系列の数 M は、KARATE は $M \in \{10, 20, \dots, 200\}$ 、BLOG, MIXI は $M \in \{50, 100, \dots, 1000\}$ とし、それぞれの M において、5 セットのデータを生成した。

対抗手法：2 種類のパラメータ推定法を対抗手法とする。1 つ目は、文献 [16] で提案された各リンクの伝播確率を最尤推定によって直接推定する方法である。この方法による結果を SaitoICM と表記する。2 つ目は、すべてのリンクの伝播確率を同一の値として推定する方法である。SaitoICM と比較してモデルの表現能力は乏しい一方で、頑健なパラメータ推定が可能であることから、ICM を使った応用で使用されている [21], [22]。この方法による結果を SimpleICM と表記する。

推定した伝播確率の評価：各リンク $(u, v) \in \mathcal{E}$ に対して、真の伝播確率 κ_{uv}^* と推定した伝播確率 $\hat{\kappa}_{uv}$ が与えられたとき、その誤差は以下の RMSE を使って評価する。

$$RMSE = \sqrt{\frac{1}{|\mathcal{E}|} \sum_{(u,v) \in \mathcal{E}} (\kappa_{uv}^* - \hat{\kappa}_{uv})^2}$$

なお、LFICM は各 $K \in \{1, \dots, 8\}$ で 5 回パラメータ推定を行い、最も対数尤度の値が大きくなった試行を採用した。SaitoICM と SimpleICM も同様である。

図 2 は、KARATE, BLOG, MIXI における伝播確率の平均推定誤差を示す。横軸は推定に使用した拡散系列の数を表し、縦軸は真の伝播確率と推定した伝播確率の RMSE を表す。まず、すべてのネットワークにおいて、提案モデル LFICM は従来手法の SaitoICM よりもつねに小さい RMSE となった。伝播確率にはノイズを乗せているため、提案モデルが簡単にとらえられるような規則性は失われていると想像できるが、その状況でも提案モデルは精度良く伝播確率を推定できることを示している。また、SimpleICM は拡散系列が少なくても、ある程度精度良く伝播確率を推定できている。これは、このモデルがネットワーク全体で 1 つの伝播確率しか持たないため、少ない拡散系列でも過学習を引き起こさないからである。しかし、モデルが十分に複雑ではないために、拡散系列の数が増やしても推定誤差は小さくならない。提案モデル LFICM は、SaitoICM よりもきわめて少ない拡散系列の段階で SimpleICM よりも良い推定精度に到達していることから、少ない観測データからでも精度良く伝播確率を推定できることが分かる。

5.2 次元数 K の選択

観測した拡散系列を十分に表現できる潜在特徴ベクトルの次元数 K を選択することは、パラメータ数の削減につながり、モデルの過学習を防ぐ効果がある。ここでは、人工的に生成した拡散系列から真の潜在特徴ベクトルの次元数 K^* を選択できるかを確認する。

この実験では、KARATE ネットワークを使い、以下の手順で拡散系列を生成した。各 $K^* \in \{2, 3, 4\}$ において、平均

表 3 実データセットの統計量
Table 3 Statistics of real datasets.

	ノード数	リンク数	平均次数	拡散系列数	平均系列長	最大系列長	最小系列長
DIGG	11,832	167,244	14.13	1,000	180.25	3,563	15
FLIXSTER	20,254	124,088	6.12	3,000	49.27	6,444	1

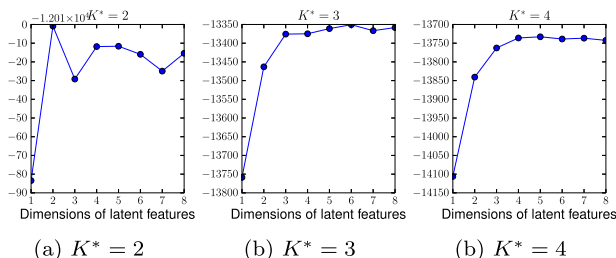


図 3 KARATE データにおける開発用データに対する対数尤度
Fig. 3 Log-likelihood of development data in KARATE dataset.

0, 分散 1 の K^* 次元正規分布から \mathbf{X}, \mathbf{Y} を生成し, 式 (3) を使って各リンク (u, v) の伝播確率 $\kappa_{uv} = f(\mathbf{x}_u, \mathbf{y}_v, -1.5)$ を計算する. この伝播確率の下で, 訓練用データとして $M = 200$ の拡散系列 5 セット, 開発用データとして $M = 500$ の拡散系列を 1 セット作成した.

図 3 は, それぞれの K^* における開発用データの対数尤度の値を示す. 縦軸は開発用データの対数尤度, 横軸は推定時に設定した K の値である. 対数尤度の値は, 式 (10) の対数である. $K^* = 2$ の場合では, $K = 2$ で推定したモデルの対数尤度が最も大きくなった. $K^* = 3$ と $K^* = 4$ の場合では, 真の次元数のときに対数尤度が最大にならないが, その K 以降で対数尤度の上昇が収束していることが見て取れる. この結果から, 上記のように開発用データを使うことによって, データ生成の背後にある次元数 K^* が選択できる可能性が示唆された.

5.3 実データによる提案モデルの評価

データセット: 現実の Web サービスから取得された DIGG データと FLIXSTER データを使用した. DIGG は, ソーシャルニュースサイト Digg^{*1} から抽出されたデータセットである [12]. Digg では, ユーザは Web ページに投票を行うことができ, 投票行動はソーシャルネットワークを介して共有される. 実験では, 1,000 の Web ページを選択し, 各ページの投票ユーザと投票時刻から拡散系列を作成した. また, その Web ページに投票したユーザからなるネットワークを抽出した. FLIXSTER は, ソーシャル映画サイト Flixster^{*2} から抽出されたデータセットである [7]. Flixster では, ユーザは映画に評点を付けることができる. 実験では, 3,000 本の映画を選択し, 点数にかかわらず各映画に評点したユーザと評点した時刻から拡散系列を作成した.

*1 <http://digg.com/>

*2 <http://www.flixster.com/>

また, 映画に評点したユーザからなるネットワークを抽出した. 表 3 は, DIGG と FLIXSTER の統計量を示す.

テストセット対数尤度による評価: パラメータ推定に使用する拡散系列の数に対するテストセットの対数尤度を評価する. 学習したモデルとテストセットが与えられたとき, テストセット対数尤度の平均は式 (10) を使って以下のように定義する.

$$\mathcal{L}_{test} = \frac{1}{|D_{test}|} \log P(D_{test} | \hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\gamma}, \hat{r}, \mathcal{G}) \quad (22)$$

ここで, $\hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\gamma}, \hat{r}$ はそれぞれ訓練セットによる推定値を表す. この指標は, ノードがアクティブになるかどうかの予測性能を測るために使用できる. したがって, 最大のテストセット対数尤度を持つモデルは, 未知の拡散現象を最も正しく予測するモデルであるといえる. SaitoICM や SimpleICM を評価する場合には, $\hat{\mathbf{X}}, \hat{\mathbf{Y}}, \hat{\gamma}$ を介さずに推定した伝播確率を直接使用する. なお, テストセットの拡散系列として, DIGG は 200 系列, FLIXSTER は 500 系列抽出した.

図 4 は, 学習で使用する拡散系列の数を変化させたときのテストセット対数尤度の平均を示す. 縦軸はテストセット対数尤度の平均, 横軸は訓練セットの拡散系列の数である. 図から分かるように, 両データセットにおいて, LFICM は K の値によらず最も大きい対数尤度を達成した. DIGG の結果では, 最も少ない拡散系列で学習した LFICM が最も多い拡散系列で学習した SaitoICM や SimpleICM よりも大きい対数尤度の値になった. また, FLIXSTER においては, SaitoICM が 2,500 系列使って学習した場合と LFICM が 500 系列を使って学習した場合で同等の対数尤度の値となった. 人工データを使った実験 (図 2) で見たように, LFICM が少ないデータからでも伝播確率を精度良く推定できていることが示唆され, その結果, テストセットの対数尤度も向上したと考えられる.

期待影響度による評価: ノード u の期待影響度とは, ノード u が情報源ノードとなったとき, そのノードが最終的に影響を与えるノード数の平均値である. 期待影響度では, 情報源ノードから直接影響を受けたノードと他のノードを介して間接的に影響を受けたノードの両方が数え上げられる. したがって, テストセット対数尤度は隣接ノードの状態のみを考慮したが, 期待影響度は 2 歩以上先のノードの状態まで考慮するため, ネットワーク構造にも依存する評価尺度である. この実験では, 学習した ICM のパラメー

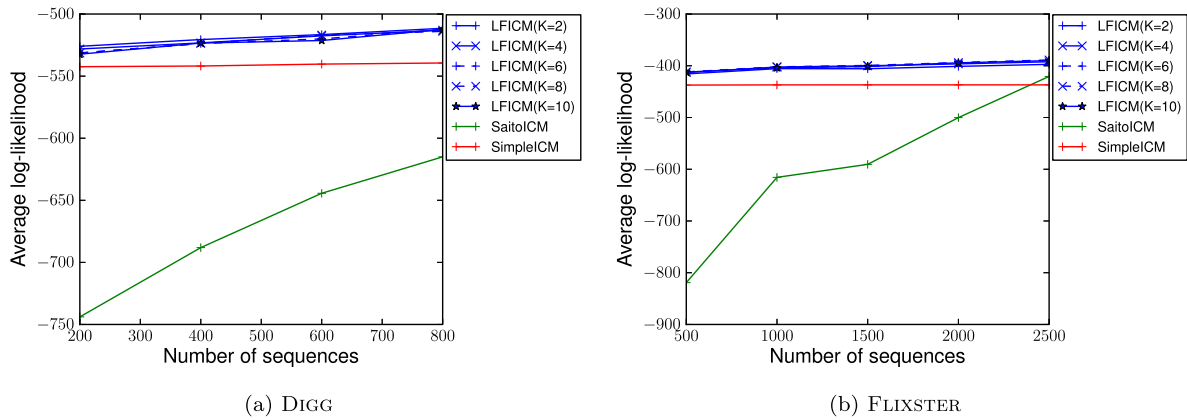


図 4 学習で使用する拡散系列の数を変えたときのテストセットに対する平均対数尤度

Fig. 4 Average log-likelihood of test data at varying numbers of diffusion sequences for learning.

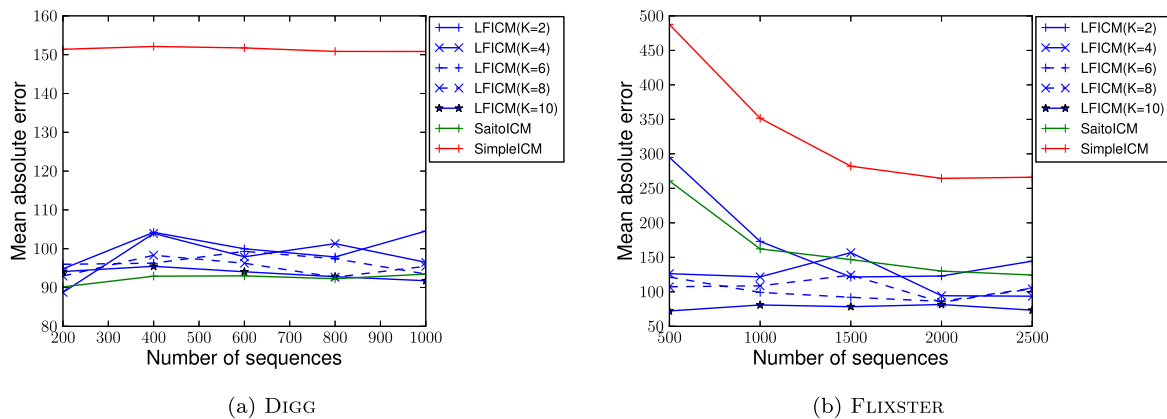


図 5 学習で使用する拡散系列の数を変えたときの期待影響度の推定誤差

Fig. 5 Estimation error of the expected degree of influence at varying numbers of diffusion sequences for learning.

Algorithm 3 INFLUENCEPREDICTOR

Require : ネットワーク \mathcal{G} , 情報源ノード集合 \mathcal{S} , 時間遅れパラメータ r , 伝播確率 $\{\kappa_{uv}\}_{(u,v) \in \mathcal{E}}$, 試行回数 T

- 1: $influence \leftarrow 0$
- 2: **for** $m \leftarrow 1, 2, \dots, T$ **do**
- 3: $\mathbf{d}_m \leftarrow \text{ICMGENERATOR}(\mathcal{G}, \mathcal{S}, r, \{\kappa_{uv}\}_{(u,v) \in \mathcal{E}})$ // Alg. 1
- 4: $influence \leftarrow influence + \frac{1}{T} |\mathbf{d}_m|$
- 5: **end for**
- 6: **return** $influence$

タを使ってシミュレーションを行うことによって期待影響度を推定する。Algorithm 3 は、期待影響度推定の手順を示す。なお、試行回数 T は 50 に設定した。

以下の手順で実験を行う。まず、DIGG と FLIXSTER のそれぞれにおいて、すべての拡散系列を使って各ノードの真の期待影響度を計算する。なお、1 度も情報源ノードとして出現しなかったノードは真の期待影響度が計算できないため、評価では使用しない。次に、一部の拡散系列を使って伝播確率と時間遅れパラメータを推定し、推定したパラメータの下で、Algorithm 3 を使って各ノードの期待影響度を計算する。

図 5 は、真の期待影響度と推定した期待影響度の誤差を示す。縦軸は誤差、横軸は学習で使った拡散系列の数である。なお、誤差は平均絶対誤差 (MAE) で計測した。両データセットにおいて、DIGG では、SimpleICM と比較して、LFICM は平均して 50 程度小さい MAE を達成した。しかし、SaitoICM と比較すると、LFICM は同等かそれ以下の結果になった。FLIXSTER では、拡散系列の数によらず、 $K = 6, 8, 10$ の LFICM が SaitoICM と SimpleICM よりも良い精度を達成した。また、観測する拡散系列の数が少ない場合で LFICM は他の手法より優れており、FLIXSTER では 500 系列を学習で使用する時、LFICM ($K = 10$) は SaitoICM の約 1/4、SimpleICM の約 1/7 の誤差で推定した。

パラメータの計算時間：図 6 は、FLIXSTER のパラメータを LFICM ($K = 10$) で推定した際の計算時間を示す。縦軸は計算時間 (分)、横軸は学習で使った拡散系列の数である。また、実験はクロック周波数 2.93 GHz の計算機で行った。LFICM の時間計算量は、リンク数 $|\mathcal{E}|$ 、拡散系列の数 M 、潜在特徴ベクトルの次元数 K に依存する。したがっ

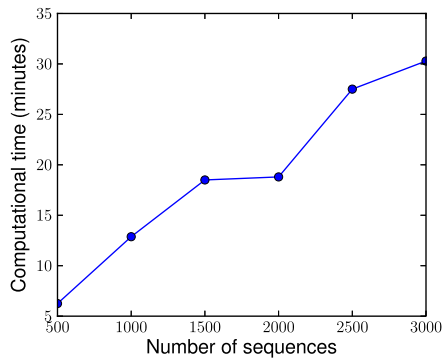


図 6 FLIXSTER のパラメータ計算時間

Fig. 6 Computation time for parameter estimation on FLIXSTER.

て、本稿の FLIXSTER による実験では $K = 10$, $M = 3000$ のとき最も計算時間がかかるが、その計算時間は平均 30 分である。

6. おわりに

本稿では、情報拡散現象のモデル化や情報拡散の予測などで一般的に使われる独立カスケードモデル (ICM) の伝播確率をより正確に推定するための新しい確率モデルを提案した。具体的には、各ノードが情報の伝えやすさを表す潜在ベクトルと、情報の受け取りやすさを表す潜在ベクトルを持つようにモデル化し、各リンクの伝播確率はリンクの両端ノードの潜在特徴ベクトルに基づいて決まるように設計した。また、EM アルゴリズムを用いた事後確率最大化によるパラメータ推定法を示した。実験では、現実のネットワークデータと人工で作成した拡散系列を使って、提案モデルが既存の方法よりも正確にパラメータを推定できることを示した。また、実データによる実験では、提案モデルで推定したパラメータを使うことによって、未知の拡散系列に対する予測性能を高めることができることを示した。

ここで提案したモデルは、同じ趣味や所属のユーザ間では情報が伝わりやすい可能性があったり、異なる考えを持つユーザでは情報が伝わりにくいといったように、ユーザの属性がユーザ間の伝播確率に影響を与えるという直感的なアイデアに基づいている。提案モデルはこのようなユーザの特徴を追加の属性情報なしで推定し、伝播確率を効率的に推定することができる。

最後に、今後の研究課題をまとめる。DIGG データセットを使った影響度推定の実験では、伝播確率の推定精度が良くても影響度推定の精度は必ずしも良くならないことが示唆された。この原因ははっきりしないが、ネットワーク構造の性質と関係がある可能性がある。また、ICM の情報拡散過程がこのデータセットに合っていない可能性も考えられる。今後は、線形閾値モデル [8] などの ICM 以外の情報拡散モデルにおいてもノードの潜在特徴を考慮し、モ

デル間の比較を行っていく。また、提案モデルでは、標準偏差 σ_X , σ_Y は推定しないパラメータとして扱った。しかし、現実の問題設定では、ユーザ間で影響力の大きさや影響の受けやすさが多様であることが容易に想像され、その多様さは未知である。したがって、特徴ベクトルの標準偏差 σ_X , σ_Y も推定できるように、適切な事前分布を用いたベイズモデルへの拡張を行うことを試みる。

謝辞 本研究は JSPS 特別研究員奨励費の助成を受けたものです。

参考文献

- [1] Dickens, L., Molloy, I. and Lobo, J.: Learning Stochastic Models of Information Flow, *2012 IEEE 28th International Conference on Data Engineering (ICDE)* (2012).
- [2] Franch, F.: 2010 UK Election Prediction with Social Media, *Journal of Information Technology & Politics*, Vol.10, No.1, pp.57-71 (2013).
- [3] Galuba, W. and Aberer, K.: Outtweeting the Twitterers-Predicting Information Cascades in Microblogs, *Proc. 3rd Conference on Online Social Networks* (2010).
- [4] Goldenberg, J. and Libai, B.: Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-mouth, *Marketing Letters*, Vol.12, No.3, pp.211-223 (2001).
- [5] Gruhl, D., Guha, R., Liben-Nowell, D. and Tomkins, A.: Information Diffusion through Blogspace, *Proc. 13th International Conference on World Wide Web*, pp.491-501 (2004).
- [6] Iribarren, J. and Moro, E.: Impact of Human Activity Patterns on the Dynamics of Information Diffusion, *Physical Review Letters*, Vol.103, No.3, pp.8-11 (2009).
- [7] Jamali, M. and Ester, M.: A Matrix Factorization Technique with Trust Propagation for Recommendation in Social Networks, *Proc. 4th ACM Conference on Recommender Systems*, pp.135-142 (2010).
- [8] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the Spread of Influence through a Social Network, *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, p.137 (2003).
- [9] Kimura, M., Saito, K., Ohara, K. and Motoda, H.: Learning to Predict Opinion Share in Social Networks, *Proc. 24th AAAI Conference on Artificial Intelligence*, pp.1364-1370 (2010).
- [10] Kimura, M., Saito, K. and Motoda, H.: Blocking Links to Minimize Contamination Spread in a Social Network, *ACM Trans. Knowledge Discovery from Data*, Vol.3, No.2, pp.1-23 (2009).
- [11] Kuo, T.-T., Hung, S.-C., Lin, W.-S., Lin, S.-D., Peng, T.-C. and Shih, C.-C.: Assessing the Quality of Diffusion Models Using Real-World Social Network Data, *2011 International Conference on Technologies and Applications of Artificial Intelligence*, pp.200-205 (2011).
- [12] Lerman, K. and Ghosh, R.: Information Contagion: An Empirical Study of Spread of News on Digg and Twitter Social Networks, *Proc. 4th International Conference on Weblogs and Social Media (ICWSM)* (2010).
- [13] Matsubara, Y., Sakurai, Y., Prakash, B.A., Li, L. and Faloutsos, C.: Rise and Fall Patterns of Information Diffusion: Model and Implications, *Proc. 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.6-14, ACM (2012).

- [14] Myers, S., Zhu, C. and Leskovec, J.: Information Diffusion and External Influence in Networks, *Proc. 18th ACM SIGKDD* (2012).
- [15] Reddy, A.S.S., Siva, A., Kasat, P. and Jain, A.: Box-Office Opening Prediction of Movies based on Hype Analysis through Data Mining, *International Journal of Computer Applications*, Vol.56, No.1, pp.1-5 (2012).
- [16] Saito, K., Kimura, M., Ohara, K. and Motoda, H.: Learning Continuous-Time Information Diffusion Model for Social Behavioral Data Analysis, *1st Asian Conference on Machine Learning*, pp.322-337 (2009).
- [17] Saito, K., Ohara, K., Yamagishi, Y. and Kimura, M.: Learning Diffusion Probability Based on Node Attributes in Social Networks, *19th International Symposium, ISMIS 2011*, pp.153-162 (2011).
- [18] Sood, V. and Redner, S.: Voter Model on Heterogeneous Graphs, *Physical Review Letters*, Vol.94, No.17, p.178701 (2005).
- [19] Zachary, W.: An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research*, Vol.33, No.4, pp.452-473 (1977).
- [20] Zhang, X., Fuehres, H. and Gloor, P.A.: Predicting Stock Market Indicators Through Twitter “I hope It Is Not as Bad as I Fear”, *The 2nd Collaborative Innovation Networks Conference*, Vol.26, pp.55-62 (2011).
- [21] 吉川友也, 齊藤和巳, 元田 浩, 大原剛三, 木村昌弘: 情報拡散モデルに基づくソーシャルネットワーク上でのノードの期待影響度曲線推定法, *電子情報通信学会論文誌 D, 情報・システム*, Vol.94, No.11, pp.1899-1908 (2011).
- [22] 木村昌弘, 齊藤和巳, 中野良平, 元田 浩: 社会ネットワークにおける有力ノード抽出のための情報拡散モデルの学習, *人工知能学会論文誌*, Vol.25, No.1, pp.215-223 (2010).



吉川 友也

平成 23 年静岡県立大学経営情報学部卒業。平成 25 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在、同大学博士後期課程在籍中。平成 25 年より日本学術振興会特別研究員 (DC1)。データマイニング、

機械学習の研究に従事。



岩田 具治 (正会員)

平成 13 年慶應義塾大学環境情報学部卒業。平成 15 年東京大学大学院総合文化研究科修士課程修了。同年 NTT 入社。平成 20 年京都大学大学院情報学研究科博士課程修了。博士 (情報学)。平成 24 年より 1 年間 Cambridge 大学客員研究員。現在、NTT コミュニケーション科学基礎研究所研究主任。機械学習、データマイニング、情報可視化の研究に従事。平成 16 年船井ベストペーパー賞、平成 21 年情報処理学会論文誌論文賞等受賞。電子情報通信学会会員。



澤田 宏

1991 年京都大学工学部情報工学科卒業。1993 年同大学大学院修士課程修了。同年日本電信電話株式会社 (NTT) 入社。以来、同社コミュニケーション科学基礎研究所にて、VLSI 向け CAD および計算機アーキテクチャの研究に従事。2000 年より信号処理、特にブラインド音源分離の研究に従事。2009 年より知能創発環境研究グループグループリーダー。2013 年よりサービスエボリューション研究所ヒューマンアナリシスプロジェクトグループリーダー、現在に至る。2001 年京都大学博士 (情報学)。電子情報通信学会、日本音響学会、IEEE 各会員。

(担当編集委員 平手 勇宇)