

# ネットワーク上の情報に対する 拡散現象と潜在的トピックの同時学習

Joint Learning of Social Propagation and Latent Topics for Information on Social Networks

吉川 友也\*<sup>1</sup> 岩田 具治\*<sup>2</sup> 澤田 宏\*<sup>2</sup>  
Yuya Yoshikawa Tomoharu Iwata Hiroshi Sawada

\*<sup>1</sup>奈良先端科学技術大学院大学 Nara Institute of Science and Technology  
\*<sup>2</sup>NTT コミュニケーション科学基礎研究所 NTT Communication Science Laboratories

Pieces of information transmitted in the Internet diffuse through the user networks of microblogs and social networking services, and are shared by many individuals. Such information have an impact on the decision makings of the individuals on the real world. For the prediction of the trends, it is important to quantitatively understand both influences of information contents and individuals. In this paper, we propose a probabilistic model for simultaneously estimating the influences of information contents and the strength of relationships between individuals from social propagation phenomena. For the proposed model, we present a learning procedure based on the variational bayes method, which is robust compared to the EM algorithm. In experiments, we evaluate the effectiveness of the proposed model using a real microblog dataset.

## 1. はじめに

Web上で発信された情報は、マイクロブログやSNSのユーザネットワークを介して拡散し、共有される。多くの人によって共有された情報は、そのネットワークでトレンドとなり、結果的にWebから離れた現実の世界に影響を与える。例えば、Web上のユーザの発信した情報から、映画の興行収入や株価を予測できるという報告がある [Reddy 12, Zhang 11]。Web上のトレンドを予測する技術開発は、現実社会のトレンドを予測する上で重要である。

トレンドを予測する上で、情報それ自体の広がりやすさと情報を発信する人の影響力を定量的に知ることは重要である。本論文では、ソーシャルネットワークにおける情報拡散現象から、これらを同時に推定するための確率モデルを提案する。提案モデルは、情報の内容に依存して拡散の起こりやすさ（感染力）が変化するものと仮定する。各情報は複数のトピックから構成され、感染力は情報の持つトピックの分布とすべての情報で一貫したトピックの重みから決定される。加えて、個人間のリンクには繋がりの強さがあると仮定する。これは情報の内容には依存せず決まり、信頼出来る人からは影響されやすく、そうでない人からは影響されにくいことを反映する。提案モデルはこの仮定に基づき、各個人はネットワークを介して情報を受け取り、共有するとする。

提案モデルは、情報拡散の履歴と情報の内容（テキスト）を観測データとして使用する。情報拡散の履歴には、誰がどの情報を受け取ったのかが記録されている。これは、個人が誰かの発言に影響され、その情報を知人へと伝える行動に相当する。

提案モデルの学習は、変分ベイズ法に基づいて行う。変分ベイズ法はベイズ確率モデルに対する近似推論法の一つであり、類似の学習法であるEMアルゴリズムと比べ、頑健な学習が行える。提案モデルにおいては、情報のトピックの事後分布を計算が簡単な変分事後分布で近似し、真の事後分布と変分事後分布が近づくように個人間の繋がりの強さや感染力に対するトピックの重みを学習する。

実験では、代表的なマイクロブログサービスのTwitter\*<sup>1</sup>から連絡先: 吉川友也, 奈良先端科学技術大学院大学,

yuya-y@is.naist.jp

\*<sup>1</sup> <https://twitter.com>

ら取得したデータを使い、提案モデルの有効性を評価する。

## 2. 関連研究

SNSやマイクロブログが一般的に利用されるようになったため、これらの上での情報拡散の効果や予測に関する研究が、近年盛んに行われている [Richardson 02, Leskovec 07, Jansen 09]。最も良く利用される情報拡散の仕組みを表現する確率モデルの一つはIndependent Cascade Model (ICM)である [Kempe 03]。提案モデルは、ICMと同様、個人間の繋がりの強さによって情報が伝わるとするが、情報それ自体が持つ感染力も考慮できる点が異なる。

提案モデルにおける情報の生成は、Latent Dirichlet Allocation (LDA) [Blei 03] と呼ばれる潜在的なトピックを考慮した文書生成モデルに基づく。LDAでは、各文書は複数のトピックによって構成され、単語はトピック毎の単語確率分布から生成されるとする。LDAは情報検索や協調フィルタリングといった幅広い分野で応用され、良い結果を示している。提案モデルは、LDAと情報拡散モデルを組み合わせたモデルである。両者を結合することにより、同じように拡散に影響を与える単語を同じトピックになるように学習できる。

情報の内容を使った情報拡散モデルは既に存在する。[吉川 12, Barbieri 12] は提案モデルと同様、情報の内容を考慮する情報拡散モデルであるが、これらは1つの情報は1つのトピックのみを持つと仮定するのに対し、提案モデルは1つの情報が複数のトピックを持つと仮定する点が異なる。[Blei 03] が複数トピックの方を仮定することにより単語の予測性能が高くなることを示したように、提案モデルにおいても、新しい情報に対して拡散予測性能の向上が期待できる。

## 3. 提案モデル

個人の集合  $V$  と個人間の関係  $E \subseteq V \times V$  から成る有向ネットワーク  $G = (V, E)$  を考える。なお、ノード  $v \in V$  の親ノード集合を  $B(v) = \{u \mid (u, v) \in E\}$ 、子ノード集合を  $F(v) = \{w \mid (v, w) \in E\}$  とする。

ネットワーク  $G$  上では、個人の発言などの情報がリンクを介して共有される。1つの発言を1つの情報  $i$  とみなし、

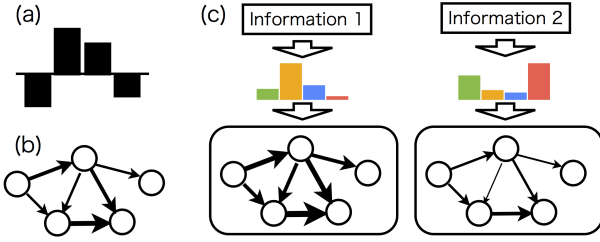


図 1: 提案モデルのパラメータ。(a) 感染力に対する各トピックの重み。(b) 情報に依存しない各リンクの繋がりの強さ。リンクの太さが繋がりの強さを表す。(c) 情報が与えられると、そのトピック分布と(a),(b)のパラメータから各リンクの情報の伝わりやすさが決まる。

情報の集合を  $I$  とする。各情報  $i \in I$  の内容は、通常、テキストで表現される。bag-of-words 表現を使うことで、情報  $i$  の単語列  $\mathbf{w}_i = \{w_{i,j}\}_{j=1}^{N_i}$  が観測される。ここで、 $w_{i,j}$  は情報  $i$  で  $j$  番目に出現した単語を表す。加えて、情報  $i$  を発信・共有したノード集合を観測する。今後、情報によって影響されたノードをアクティブノードと呼ぶ。各情報  $i$  に対して、複数ノードがアクティブになるので、実際には拡散系列  $\mathbf{d}_i = \{v \mid v \in V, v \text{ は情報 } i \text{ でアクティブ}\}$  が得られる。また、ノード  $v$  がアクティブであるとき、ノード  $v$  は親ノード  $u \in B(v)$  のいずれかから影響を受け、我々はそのノード  $u$  を観測できるとする。すなわち、各情報  $i$  について、

$$\mathbf{r}_i = \{(u, v) \mid (u, v) \in E, v \text{ は } u \text{ から影響を受けた}\}$$

となる経路  $\mathbf{r}_i$  を観測する。

提案モデルは、各情報の単語は複数の潜在トピックから生成されるとする。すなわち、情報  $i$  のトピック分布  $\theta_i$  と、 $j$  番目のトピック割り当て  $z_{i,j}$  と単語  $w_{i,j}$  は、以下の分布から生成される。

$$\theta_i \sim \text{Dirichlet}(\boldsymbol{\alpha}), \quad (1)$$

$$z_{i,j} \sim \text{Multinomial}(\theta_i), \quad (2)$$

$$w_{i,j} \sim \text{Multinomial}(\beta_{z_{i,j}}), \quad (3)$$

ここで、 $\alpha > 0$  は  $K$  次元のディリクレ事前分布パラメータ、 $\beta_{z_{i,j}} > 0$  はトピック  $z_{i,j} \in \{1, \dots, K\}$  の単語分布を表す。この部分は LDA [Blei 03] と同じ仮定である。また、情報  $i$  の経験トピック分布を、 $\bar{z}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta(z_{i,j})$  と定義する。ただし、 $\delta(k)$  は  $k$  番目の要素のみ 1 で他は 0 の  $K$  次元ベクトルである。

次に、拡散系列  $\mathbf{d}_i$  の生成について説明する。提案モデルでは、情報  $i$  において、ノード  $u \in V$  がアクティブであるときに子ノード  $v \in F(u)$  がアクティブになる確率は、式 (9) で示すように、繋がりの強さ  $\kappa_{u,v}$  と感染力  $\pi_i$  の積で定義される。なお、情報  $i$  の感染力  $\pi_i$  は、

$$\pi_i = \sigma(\bar{z}_i^\top \boldsymbol{\eta}) = \left(1 + \exp(-\bar{z}_i^\top \boldsymbol{\eta})\right)^{-1} \quad (4)$$

と定義する。ここで、 $\boldsymbol{\eta} = [\eta_k]_{k=1}^K$ 、 $\eta_k \in \mathbb{R}$  は感染力に対するトピックの重みである。したがって、 $\eta_k > 0$  ならば、トピック  $k$  は感染力を押し上げ、 $\eta_k < 0$  ならば、感染力を押し下げる効果がある。また、式 (4) はシグモイド関数なので、 $0 \leq \sigma(x) \leq 1$ 、 $x \in \mathbb{R}$  である。図 1 は提案モデルのパラメータを示す。

## 4. 推論アルゴリズム

この節では、3 節の提案モデルのパラメータを変分ベイズ法によって推定する方法を紹介する。

観測データ  $\mathbf{W} = \{\mathbf{w}_i\}_{i \in I}$ 、 $\mathbf{R} = \{\mathbf{r}_i\}_{i \in I}$  に対する提案モデルの周辺尤度は以下のように書ける。

$$\begin{aligned} P(\mathbf{W}, \mathbf{R} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\kappa}) &= \prod_{i \in I} \int \sum_{\mathbf{z}_i} \left[ P(\mathbf{w}_i | \mathbf{z}_i, \boldsymbol{\beta}) P(\mathbf{z}_i | \boldsymbol{\theta}_i) P(\boldsymbol{\theta}_i | \boldsymbol{\alpha}) \right. \\ &\quad \left. P(\mathbf{r}_i | \mathbf{z}_i, \boldsymbol{\eta}, \boldsymbol{\kappa}) \right] d\boldsymbol{\theta}_i \end{aligned} \quad (5)$$

最初の 3 つの因子は式 (1),(2),(3) の仮定に基づき、以下のよように定式化される。

$$P(\mathbf{w}_i | \mathbf{z}_i, \boldsymbol{\beta}) = \prod_{j=1}^{N_i} \beta_{z_{i,j}, w_{i,j}}, \quad (6)$$

$$P(\mathbf{z}_i | \boldsymbol{\theta}_i) = \prod_{j=1}^{N_i} \theta_{i, z_{i,j}}, \quad (7)$$

$$P(\boldsymbol{\theta}_i | \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{i,k}^{\alpha_k - 1}, \quad (8)$$

ここで、 $\Gamma(\cdot)$  はガンマ関数である。また、最後の因子は経路  $\mathbf{r}_i$  の尤度であり、

$$P(\mathbf{r}_i | \mathbf{z}_i, \boldsymbol{\eta}, \boldsymbol{\kappa}) = \prod_{(u,v) \in \mathbf{r}_i} \pi_i \kappa_{u,v} \prod_{(u,v) \in \mathbf{r}_i^-} (1 - \pi_i \kappa_{u,v}). \quad (9)$$

なお、 $\mathbf{r}_i^-$  は、情報  $i$  において拡散に失敗したリンク集合で、

$$\mathbf{r}_i^- = \{(u, v) \mid (u, v) \in E \setminus \mathbf{r}_i, u \in \mathbf{d}_i\}$$

である。ここで我々の目的は、周辺尤度を最大にする  $\beta, \eta, \kappa$  を推定することである。より形式的には、以下のように書ける。

$$\hat{\beta}, \hat{\eta}, \hat{\kappa} = \arg \max_{\beta, \eta, \kappa} P(\mathbf{W}, \mathbf{R} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\kappa}). \quad (10)$$

しかしながら、式 (10) は計算量の問題で直接行えないため、代わりに、対数周辺尤度に対する以下の変分下限を最大化する。

$$\begin{aligned} \log P(\mathbf{W}, \mathbf{R} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\kappa}) &\geq \sum_{i \in I} \left( E_{q_i} [\log P(\mathbf{w}_i | \mathbf{z}_i, \boldsymbol{\beta})] + E_{q_i} [\log P(\mathbf{z}_i | \boldsymbol{\theta}_i)] \right. \\ &\quad \left. + E_{q_i} [\log P(\boldsymbol{\theta}_i | \boldsymbol{\alpha})] + E_{q_i} [\log P(\mathbf{r}_i | \mathbf{z}_i, \boldsymbol{\eta}, \boldsymbol{\kappa})] \right. \\ &\quad \left. + E_{q_i} [\log q(\mathbf{z}_i, \boldsymbol{\theta}_i | \boldsymbol{\lambda}_i, \boldsymbol{\phi}_i)] \right) \\ &= \mathcal{L}. \end{aligned} \quad (11)$$

ここで、 $E_{q_i}[f]$  は変分事後分布  $q(\mathbf{z}_i, \boldsymbol{\theta}_i | \boldsymbol{\phi}_i, \boldsymbol{\lambda}_i) = q(\mathbf{z}_i | \boldsymbol{\phi}_i) q(\boldsymbol{\theta}_i | \boldsymbol{\lambda}_i)$  による期待値で、

$$E_{q_i}[f] = \int \sum_{\mathbf{z}_i} q(\mathbf{z}_i | \boldsymbol{\phi}_i) q(\boldsymbol{\theta}_i | \boldsymbol{\lambda}_i) f(\mathbf{z}_i, \boldsymbol{\theta}_i) d\boldsymbol{\theta}_i, \quad (12)$$

ただし、 $q(\mathbf{z}_i | \boldsymbol{\phi}_i)$  は多項分布、 $q(\boldsymbol{\theta}_i | \boldsymbol{\lambda}_i)$  はディリクレ分布をそれぞれ仮定する。したがって、

$$q(\mathbf{z}_i | \boldsymbol{\phi}_i) = \prod_{j=1}^{N_i} \phi_{i,j, z_{i,j}}, \quad (13)$$

$$q(\boldsymbol{\theta}_i | \boldsymbol{\lambda}_i) = \frac{\Gamma(\sum_{k=1}^K \lambda_{i,k})}{\prod_{k=1}^K \Gamma(\lambda_{i,k})} \prod_{k=1}^K \theta_{i,k}^{\lambda_{i,k} - 1} \quad (14)$$

である。周辺尤度の最大化は、この変分事後分布と  $\mathbf{z}_i, \boldsymbol{\theta}_i$  の真の事後分布の KL 距離の最小化と等価であるため、最終的には、変分事後分布は真の事後分布の近似になる。

式 (11) を最大化するために、変分ベイズ法では、変分パラメータ  $\lambda_i, \phi_i$  を計算する変分 E ステップと、提案モデルのパラメータを計算する変分 M ステップを交互に繰り返す。

変分 E ステップ: 現在のパラメータ値のもと、式 (11) を最大化するように変分パラメータ  $\lambda_i, \phi_i$  を更新する。変分ディリクレパラメータ  $\lambda_i$  は経路集合  $R$  の生成に寄与しないため、 $\lambda_i$  の更新式は LDA と同じである。

$$\lambda_i \leftarrow \alpha + \sum_{j=1}^{N_i} \phi_{i,j}. \quad (15)$$

変分多項パラメータ  $\phi$  の更新式は、式 (11) を  $\phi_{i,j,k}$  で偏微分し 0 となる解を求めることで、以下のように得られる。

$$\phi_{i,j,k} \propto \beta_{k,w_{i,j}} \exp \left( \Psi(\lambda_{i,k}) - \Psi \left( \sum_{k'=1}^K \lambda_{i,k'} \right) + \frac{\partial \mathcal{L}}{\partial \phi_{i,j,k}} \right) \quad (16)$$

ただし、

$$\frac{\partial \mathcal{L}}{\partial \phi_{i,j,k}} = \frac{1 - E[\pi_i]}{N_i} \eta_k \left( |r_i| + \sum_{(u,v) \in r_i^-} \frac{E[\pi_i] \kappa_{u,v}}{1 - E[\pi_i] \kappa_{u,v}} \right) \quad (17)$$

である。ここで、 $\Psi(\cdot)$  はディガンマ関数である。また、 $E[\pi_i]$  は情報  $i$  の感染力の期待値、 $E[\bar{z}_i]$  は情報  $i$  における経験トピック分布の期待値で、

$$E[\pi_i] = E_{q_i} [\sigma(\bar{z}_i^\top \boldsymbol{\eta})] \approx \sigma \left( \frac{1}{N_i} \sum_{j=1}^{N_i} \phi_{i,j}^\top \boldsymbol{\eta} \right), \quad (18)$$

$$E[\bar{z}_i] = E_{q_i} [\bar{z}_i] = \frac{1}{N_i} \sum_{j=1}^{N_i} \phi_{i,j} \quad (19)$$

である。ここで、式 (18) はテーラー展開による一次近似である [Braun 10]。

変分 M ステップ: 更新した変分パラメータ  $\lambda^{\text{new}}, \phi^{\text{new}}$  のもとで、モデルパラメータを更新する。各  $(u, v) \in E$  について、繋がりの強さ  $\kappa_{uv}$  は以下のように更新する。

$$\kappa_{u,v} \leftarrow \frac{\sum_{i \in I} \mathbb{1}((u, v) \in r_i)}{\sum_{i \in I} \mathbb{1}((u, v) \in r_i \cup r_i^-) E[\pi_i]}, \quad (20)$$

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

感染力に対する重み  $\boldsymbol{\eta}$  の更新式は、他の  $\eta_k$  の値に依存するため、閉形式では求められない。したがって、以下の一階偏微分に基づき、準ニュートン法などの最適化法によって求める。

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\eta}} = \sum_{i \in I} (1 - E[\pi_i]) E[\bar{z}_i] \left( |r_i| + \sum_{(u,v) \in r_i^-} \frac{E[\pi_i] \kappa_{u,v}}{1 - E[\pi_i] \kappa_{u,v}} \right). \quad (21)$$

トピック  $k$  における単語  $n$  の出現確率  $\beta_{k,n}$  は、以下のように更新する。

$$\beta_{k,n} \propto \sum_{i \in I} \sum_{j=1}^{N_i} \mathbb{1}(w_{i,j} = n) \phi_{i,j,k}. \quad (22)$$

なお、 $\beta_{k,n}$  は  $\sum_{n \in U} \beta_{k,n} = 1$  になるように正規化する。

## 5. 感染力の予測

学習済みのパラメータ  $\hat{\beta}, \hat{\boldsymbol{\eta}}$  のもとで新しい情報を観測したとき、感染力  $\pi_{\text{new}}$  を推定したいとする。この推定は式 (18) に

基づいて、

$$\pi_{\text{new}} = \sigma \left( \frac{1}{N_{\text{new}}} \sum_{j=1}^{N_{\text{new}}} \phi_{\text{new},j}^\top \hat{\boldsymbol{\eta}} \right) \quad (23)$$

で行われる。添字の new は新しい情報を指し示すために使う。ただし、 $\phi_{\text{new}}$  は新たに計算する必要があるため、変分 E ステップと類似の以下の計算を交互に行う。

$$\lambda_{\text{new}} \leftarrow \alpha + \sum_{j=1}^{N_{\text{new}}} \phi_{\text{new},j}, \quad (24)$$

$$\phi_{\text{new},j,k} \propto \beta_{k,w_{\text{new},j}} \exp \left( \Psi(\lambda_{\text{new},k}) - \Psi \left( \sum_{k'=1}^K \lambda_{\text{new},k'} \right) \right). \quad (25)$$

## 6. 実験

我々は提案モデルの有効性を示すために、二つの実験を行った。その際、過学習を避けるため、各リンクの繋がりの強さは親ノード依存とした。つまり、 $\kappa_{u,v} = \kappa_u$  である。また、ディリクレパラメータ  $\alpha_k, k \in \{1 \dots K\}$  は 0.5 に設定した。

### 6.1 データ

本研究では、Twitter から取得した日本語のツイートデータを使用した。2013 年 2 月 1 日から 7 日に 100 回以上発言した 31,113 ユーザを抽出し、このユーザ集合の有向ネットワークを 2012 年 12 月から 2013 年 1 月の @付きツイートから構築した。ここで、@付きツイートは特定のユーザへの発言を表す。例えば、ユーザ A からユーザ B へ @付きツイートが送られたとき、A は B が発信する情報を見れるとして、B から A へリンクを張る。また、そのユーザ集合がリツイートした 4,324 ツイートを取得し、それぞれの拡散系列を作成した。実験では、拡散系列の長さが 10 以上の情報のみを使用した。拡散経路は、Twitter から得られるデータから明示的に知ることは出来ない。しかし、Twitter の仕様上、同じ情報が同時期にユーザに到達することはなく、ユーザに影響を与えた可能性のある親ユーザが複数いる場合には、時刻の早い方から影響を受けたとして、拡散経路を決定した。各情報の単語列は、ツイート中の名詞、形容詞、動詞の単語と URL とハッシュタグ\*2 から構成した。表 1 は、Twitter データの統計量を示す。

### 6.2 アクティブノードの予測性能

新しい情報が与えられたとき、この情報によって影響を受けるノードを推定することは応用上有用である。この実験では、5 分割交差検定によって、各ノードがアクティブになるかどうかの予測性能を評価する。なお、トレーニングは各訓練セットに対してモデルパラメータの初期値を変えて 10 回行い、最も周辺尤度 (11) の値が大きい結果をテストで使用する。評価尺度として、式 (9) に基づく以下の対数尤度を使う。

$$\mathcal{L}^{\text{test}} = \sum_{i \in I^{\text{test}}} \log P(r_i | z_i, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\kappa}}). \quad (26)$$

新しい拡散系列と拡散経路に対してモデルが適切ならば、この対数尤度の値は大きくなる。ここで、新しい情報の感染力は式 (23) によって求める。

モデルに対する感染力の効果を評価するため、情報の内容を考慮しない情報拡散モデルの Independent Cascade Model (ICM) [Kempe 03] を対抗モデルとする。なお、提案モデルは  $\pi_i = 1$  で固定すると ICM と等価になるため、ICM の場合は  $\pi_i = 1$  として評価する。

\*2 #から始まるツイートの内容を表すタグ

表 1: Twitter データの統計量

ノード数	リンク数	平均次数	情報数	語彙数	平均単語数	平均系列長
31,113	1,330,761	43.14	4,324	97,674	17.56	12.57

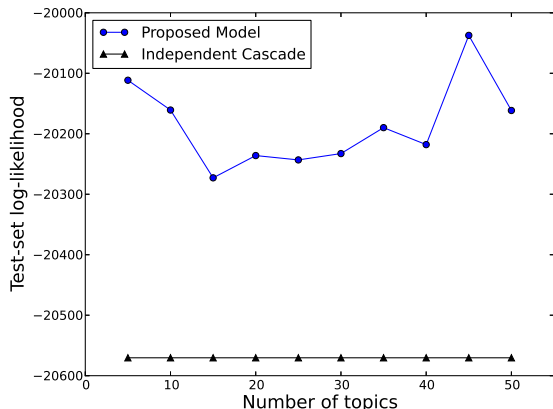


図 2: アクティブノードのテスト対数尤度

図 2 はトピック数毎のアクティブノードのテスト対数尤度の平均値を示す．なお、ICM の結果はトピック数に依存しないため、トピック数に対して定数である．図から分かるように、提案モデルは ICM よりも良い性能を示した．

### 6.3 提案モデルによる分析

表 2 は、トピック数  $K = 10$  で学習した感染力の重みの上位 3 トピックを示す．なお、各トピック  $k$  において、 $\xi_{k,n} = \beta_{k,n} / \sum_{k'=1}^K \beta_{k',n}$  の値が大きい単語  $n$  を一部表示している．1 位と 2 位のトピックでは、2 月 1 日から 7 日の間で起きたイベントを反映する単語がよく含まれていた．例えば、「南南東豆 鬼 まき」は節分の話題である．また、URL が感染力を押し上げる効果があることは、URL がつくことによって情報の信頼性が増したと考えられ、我々の直感に合致している．3 位のトピックは、ツイートを自動投稿するボットによる発言が多く含まれていた．これが大きい重みを持ったのは、ボット間でツイートを自動的に共有するようになっていたためだと考えられる．

## 7. おわりに

本論文では、ソーシャルネットワークにおける情報拡散現象を表現するために、情報拡散の履歴と情報の内容を使った情報拡散モデルを提案し、変分ベイズ法による学習アルゴリズムを示した．実験では、Twitter データを使って、提案モデルはアクティブノードの予測性能が既存モデルよりも優れていることを示した．

提案モデルでは、情報拡散の経路は既知であると仮定した．しかし、より一般的なモデル化のためには、経路は未知としてモデル化の方が望ましい．今後は、経路も潜在変数としたモデルを構築し、様々なデータセットにおけるモデルの評価を行う予定である．

## 参考文献

[Barbieri 12] Barbieri, N., Bonchi, F., and Manco, G.: Topic-Aware Social Influence Propagation Models, *IEEE*

表 2: 感染力の重み

順位	重み	単語
1	1.789	#usagipic 奇跡 トレンド #AKB48 チロルチョコキーボード
2	1.732	[URL] 南南東豆 鬼 まき
3	1.702	#IFOLLOWALL must #autofollow #RETWEET #SOUGOFOLLOW #FOLLOW4FOLLOW

12th International Conference on Data Mining, No. iii, pp. 81–90 (2012)

[Blei 03] Blei, D. M., Ng, A. Y., and Jordan, M.: Latent dirichlet allocation, *The Journal of Machine Learning Research*, Vol. 3, No. 4-5, pp. 993–1022 (2003)

[Braun 10] Braun, M. and McAuliffe, J.: Variational inference for large-scale models of discrete choice, *Journal of the American Statistical Association*, Vol. 105, No. 489, pp. 324–335 (2010)

[Jansen 09] Jansen, B. J., Zhang, M., Sobel, K., and Chowdury, A.: Twitter power: Tweets as electronic word of mouth, *Journal of the American society for information science and technology*, Vol. 60, No. 11, pp. 2169–2188 (2009)

[Kempe 03] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the spread of influence through a social network, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, p. 137 (2003)

[Leskovec 07] Leskovec, J., Adamic, L. A., and Huberman, B. A.: The dynamics of viral marketing, *ACM Transactions on the Web*, Vol. 1, No. 1, p. 5 (2007)

[Reddy 12] 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)

[Richardson 02] Richardson, M. and Domingos, P.: Mining knowledge-sharing sites for viral marketing, *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 61–70 (2002)

[Zhang 11] 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)

[吉川 12] 吉川友也, 岩田具治, 澤田宏: 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, 第 26 回人工知能学会全国大会 (2012)