

図 2: パープレキシティの比較

$\alpha_{t,u}$  である以下のディリクレ事前分布に従って生成される。

$$P(\phi_{t,u} | \hat{\phi}_{t-1,u}, \alpha_{t,u}) \propto \prod_k \phi_{t,u,k}^{\alpha_{t,u} \hat{\phi}_{t-1,u,k} - 1} \quad (1)$$

時刻  $t$  におけるトピック  $k$  の単語分布  $\theta_{t,k}$  についても同様に、平均が一時刻前の単語分布の推定値  $\hat{\theta}_{t-1,k}$ 、精度が  $\beta_{t,k}$  である以下のディリクレ事前分布から生成される。

$$P(\theta_{t,k} | \hat{\theta}_{t-1,k}, \beta_{t,k}) \propto \prod_v \theta_{t,k,v}^{\beta_{t,k} \hat{\theta}_{t-1,k,v} - 1} \quad (2)$$

本稿では、以上に述べた TTM の機構を、2 節で述べた改良モデルに取り入れたモデルを提案する。提案モデルの生成過程とグラフィカルモデルをそれぞれ図 3, 図 4 に示す。提案モデルにより、Twitter の特徴を考慮した上でユーザの興味およびトピックの流行の変化を逐次推定することが可能となる。

### 5. おわりに

本稿では、Twitter に適したオンライン学習が可能なトピックモデルを提案した。初めに、Twitter 向けに LDA を改良したトピックモデルである Twitter-LDA に対して、一般語とトピック語の割合はユーザごとに異なるという仮定を新たに加えた改良モデルを考案した。実験の結果から、改良モデルは Twitter-LDA や LDA と比較してパープレキシティが改善されることを示した。本稿ではさらに、この改良モデルに対して、時間発展を考慮したトピックモデルである TTM の機構を取り入れた新しいモデルを提案した。

今後は、提案モデルを Twitter におけるリアルタイムでの話題の検出、時間発展の追跡などに適用し、既存のモデルとの比較評価を行っていく予定である。

### 参考文献

- [1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan: Latent dirichlet allocation, Machine Learning Research, Vol. 3, pp. 993-1022, 2003
- [2] Marco Pennacchiotti and Siva Gurumurthy: Investigating topic models for social media user recommendation, In WWW2011, pp. 101-102, 2011
- [3] Jianshu Weng, Ee-Peng Lim, Jing Jiang, and Qi He: Twitterrank: finding topic-sensitive influential twitterers, In WSDM 2010, 2010
- [4] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li: Comparing twitter and traditional media using topic models,

1. Draw  $\theta_{t,B} \sim \text{Dirichlet}(\lambda)$
2. For each topic  $k = 1, \dots, K$ ,
  - (a) draw  $\theta_{t,k} \sim \text{Dirichlet}(\beta_{t,k} \hat{\theta}_{t-1,k})$
3. For each user  $u = 1, \dots, U$ ,
  - (a) draw  $\phi_{t,u} \sim \text{Dirichlet}(\alpha_{t,u} \hat{\phi}_{t-1,u})$
  - (b) draw  $\pi_{t,u} \sim \text{Beta}(\gamma)$
  - (c) for each tweet  $s = 1, \dots, N_u$ 
    - i. draw  $z_{t,u,s} \sim \text{Multinomial}(\phi_{t,u})$
    - ii. for each word  $v = 1, \dots, N_{u,s}$ 
      - A. draw  $y_{t,u,s,v} \sim \text{Bernoulli}(\pi_{t,u})$
      - B. draw  $w_{t,u,s,v} \sim \text{Multinomial}(\theta_{t,B})$  if  $y_{t,u,s,v} = 0$  or  $\text{Multinomial}(\theta_{t,z_{t,u,s}})$  if  $y_{t,u,s,v} = 1$

図 3: 提案モデルにおけるツイートの生成過程

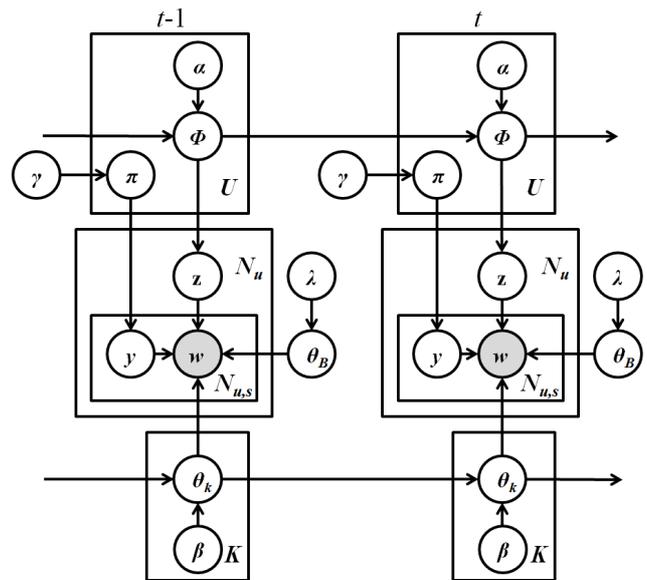


図 4: 提案モデル

In Proceedings of the 33rd European conference on Advances in information retrieval, pp. 338-349, 2011

- [5] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda: Topic tracking model for analyzing consumer purchase behavior, in Proc. IJCAI, pp. 1427-1432, 2009