Dynamic Topic Model for Topic Split & Merge

Jing Chen, Mengtian Li, Lanxiao Xu

April 29

Table of Contents

Modeling Time-varying Topics

- Automatically mining topic evolution from time-varying text corpora is important in exploratory text analytics
- Users are often curious about how topics of a certain corpora evolve from time to time

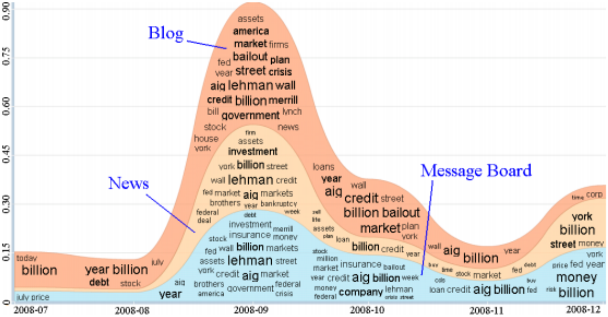


Figure: Zhang et al. (2010). Evolutionary Hierarchical Dirichlet Processes for Multiple Correlated Time-varying Corpora

Modeling Time-varying Topics

- In reality, new topics often emerge from old topics through split and merge behavior.
- Uncovering topic split and merge can be very beneficial in trend understanding, especially for news or academic literature corpora.

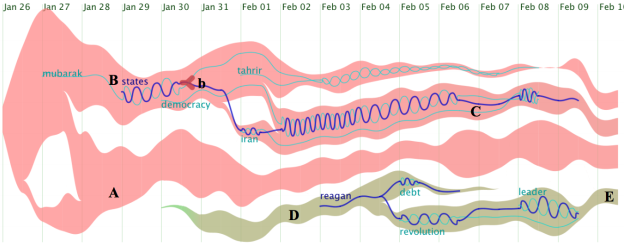


Figure: Cui et al. TextFlow: Towards Better Understanding of Evolving Topics in Text

Dirichlet Process (DP)

- Dirichlet process provides a random distribution over distributions over infinite sample spaces.
- $G \sim \text{DP}(\xi, H)$ if

$$(G(A_1), ..., G(A_r)) \sim \text{Dir}(\xi H(A_1), ..., \xi H(A_r)))$$

- The posterior of G is still a DP: Suppose $\theta_1, ..., \theta_i \sim G$. Define $\phi_1, ..., \phi_K$ to be the distinct values taken on by $\theta_1, ..., \theta_{i-1}$, and let m_k be the number of values $\theta_{i'}$ that are equal to ϕ_k for $1 \leq i' < i$. We have

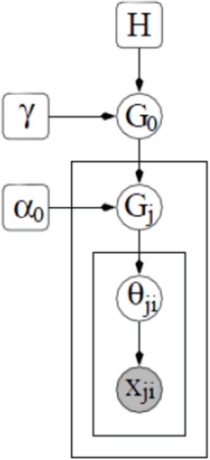
$$G|\theta_1, ..., \theta_{i-1}, \xi, H \sim \text{DP}\left(\xi + i - 1, \frac{\sum_{k=1}^{K} m_k \delta_{\phi_k} + \xi H}{\xi + i - 1}\right)$$

Hierarchical Dirichlet Process (HDP)

- Constructed by a set of DPs, and provide a nonparametric prior as the base measure of the DPs.

$$
\begin{aligned}
G_0|\gamma, H &\sim \text{DP}(\gamma, H) \\
G_j|\alpha_0, G_0 &\sim \text{DP}(\alpha_0, G_0) \\
\theta_{ji}|G_j &\sim G_j \\
x_{ji}|\theta_{ji} &\sim F(\theta_{ji})
\end{aligned}
$$

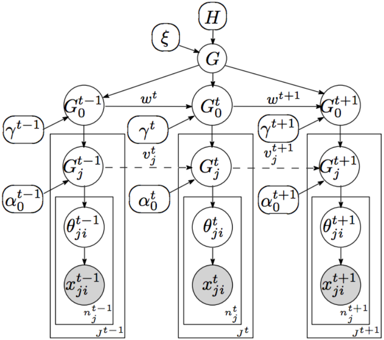
where F can be customized according to different problems.



Evolutional Hierarchical Dirichlet Process (EvoHDP)

- One more layer than the HDP model to incorporate time-varying HDPs.
- For each timestamp t, the local measure is the weighted average of the local measure of the previous timestamp and the global bookkeeping measure.

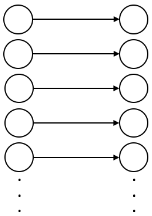
$$G_0^t \sim \text{DP}(\gamma^t, w^t G_0^{t-1} + (1 - w^t) G)$$
$$G_j^t \sim \text{DP}(\alpha_0^t, v_j^t G_j^{t-1} + (1 - v_j^t) G_0^t)$$



Topic Transitions

EvoHDP: one-to-one Transition

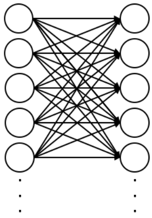
- $G^t \sim \text{DP}(\gamma^t, w^t G^{t-1} + (1 - w^t)G)$



v.s.

Split/Merge: Mixed Transition

- $G^t \sim \text{DP}(\gamma^t, w^t \Lambda(G^{t-1}) + (1 - w^t)G)$

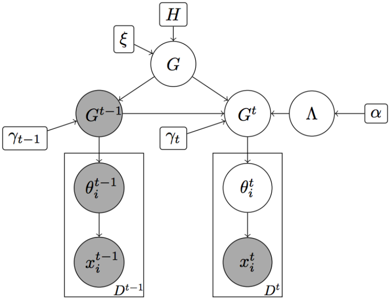


- If we represent $G^{t-1} = \sum_{k=1}^{\infty} \beta_k^{t-1} \delta_k$ as an infinite dimensional vector, Λ can be regarded as a probabilistic transitional matrix $\Lambda = [\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, ...]^T$, which encodes the evolutionary behavior of topic transition

Model: Original Representation

- Online dynamic topic modeling, always assume everything in the previous timestamp is fixed, and thus treated observed.

$$
\begin{aligned}
G &\sim \textit{DP}(\xi, H),\\
G^t &\sim \textit{DP}(\gamma^t, \hat{G}^t)\\
\hat{G}^t &= w^t \Lambda(G^{t-1}) + (1 - w^t) G\\
\theta_i^t &\sim G^t\\
x_i^t &\sim F(x|\theta_i^t)
\end{aligned}
$$



Model: Stick-Breaking Representation

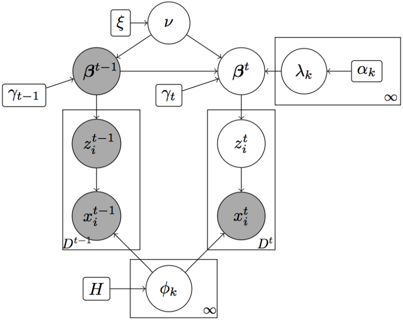
- According to the stick-breaking construction of a DP, we can write the explicit form of G and G^t.

$$G = \sum_{k=1} \nu_k \delta_{\phi_k}, \boldsymbol{\nu} \sim \text{GEM}(\xi)$$

where GEM refers to the stick-breaking process.

$$G^t = \sum_{k=1}^{\infty} \beta_k^t \delta_{\phi_k}, \boldsymbol{\beta}^t \sim \text{DP}(\gamma^t, \hat{\boldsymbol{\beta}}^t)$$

where $\hat{\beta}_k^t = \dfrac{\boldsymbol{\lambda}_k}{\sum_j \lambda_{kj}} \cdot \boldsymbol{\beta}^{t-1}$. And $\boldsymbol{\lambda}_k / \sum_j \lambda_{kj} \sim \text{Dir}(\alpha_k)$.



Sampling $\boldsymbol{\beta}^t$

- Theoretically there are infinite number of topics. But in actual implementation only a finitely number of topics are associated with the data.
- The augmented representation Teh et al. (2006) reformulates original infinite vector $\boldsymbol{\beta}^t$ to an equivalent finite one with length $K+1$.

$$G^t|\gamma^t, \hat{\boldsymbol{\beta}}^t, \{N_k^t\} \sim \mathsf{DP}\left(\gamma^t + \sum_k N_k^t, \frac{\gamma^t \hat{\boldsymbol{\beta}}^t + \sum_{k=1}^K N_k^t \delta_{\phi_k}}{\gamma^t + \sum_k N_k^t}\right)$$

$$\Rightarrow G^t = \sum_{k=1}^{K} \beta_k^t \delta_{\phi_k} + \beta_u^t G_u, G_u \sim \mathsf{DP}(\gamma^t, \hat{\boldsymbol{\beta}}^t)$$

$$\Rightarrow (\beta_1^t, ..., \beta_K^t, \beta_u^t) \sim \mathsf{Dir}(\tilde{\beta}_1^t, ..., \tilde{\beta}_K^t, \tilde{\beta}_u^t)$$

where $\tilde{\beta}_k^t = N_k^t + w^t \gamma^t \frac{\lambda_k \cdot \boldsymbol{\beta}^{t-1}}{\sum_{j=1}^{K'} \lambda_{kj}} + (1-w^t)\nu_k$ and $\tilde{\beta}_k^t = w^t \gamma^t \frac{\lambda_u \cdot \boldsymbol{\beta}^{t-1}}{\sum_{j=1}^{K'} \lambda_{uj}} + (1-w^t)\nu_u$

Sampling λ_{kj}

$$p(\lambda_{kj}|\text{rest}) \propto p(\lambda_k|\alpha_k)p(\beta^t|\lambda_k, \boldsymbol{\beta}^{t-1}, \gamma^t)$$

$$p(\lambda_k|\alpha_k) \propto \prod_{j=1}^{K'} \lambda_{kj}^{\alpha_k - 1}$$

$$p(\beta^t|\boldsymbol{\lambda}, \boldsymbol{\beta}^{t-1}, \gamma^t) \propto \left(\prod_{k=1}^{K} {\beta_k^t}^{\gamma^t \sum_{j=1}^{K'} \beta_j^{t-1} \lambda_{kj} / \sum_j \lambda_{kj}} \right) {\beta_u^t}^{\gamma^t \sum_{j=1}^{K'} \beta_j^{t-1} \lambda_{uj} / \sum_j \lambda_{kj}}$$

$$\propto \prod_{j=1}^{K'} \left(\left(\prod_{k=1}^{K} {\beta_k^t}^{\gamma^t \beta_j^{t-1} \lambda_{kj} / \sum_j \lambda_{kj}} \right) {\beta_u^t}^{\gamma^t \beta_j^{t-1} \lambda_{uj} / \sum_j \lambda_{kj}} \right)$$

Thus,

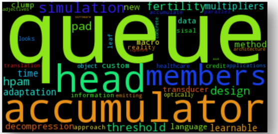
$$p(\lambda_{kj}|\text{rest}) \propto \lambda_{kj}^{\alpha_k - 1} {\beta_k^t}^{\gamma^t \beta_j^{t-1} \lambda_{kj} / \sum_j \lambda_{kj}}$$

where $k \in \{1, ..., K, u\}$.

Dataset

- DBLP Citations V2 (`https://aminer.org/DBLP_Citation`)
- A dataset Tang et al. (2008) that consists of papers from DBLP.
- Contains information of **year**, **publication venue**, **abstract** and **citations**.
- We take a subset of the dataset with year ranging from 1996 - 2010. And we treat publication venue as an additional word.
- $F(x|\theta)$ is set to be a multinomial distribution representing the generation of a paper in a topic.

Topic Modeling

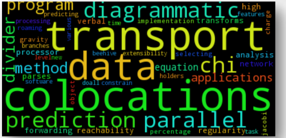


A

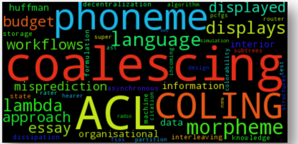
D

C

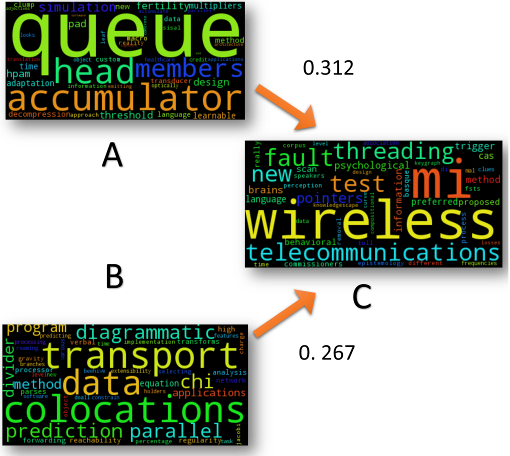
B



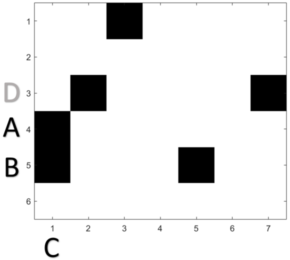
E



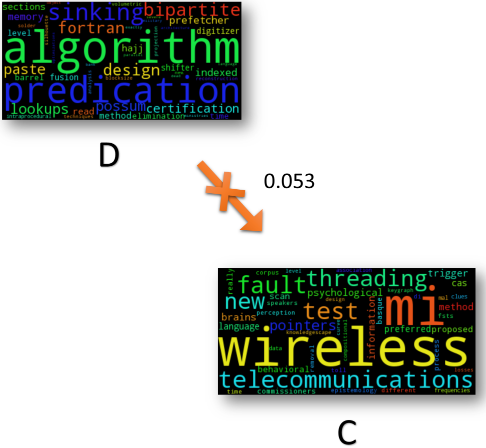
Topic Transition - Positive Example



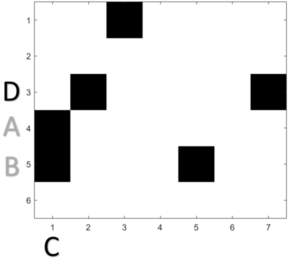
Each column of the matrix shows the **transition strength** of topics in year 1997 to that in 1998.



Topic Transition - Negative Example



Each column of the matrix shows the **transition strength** of topics in year 1997 to that in 1998.



Conclusion and Future Works

- Conclusion:
  - We have designed a novel non-parametric graphical model to characterize the transition of topics directly in a time-varying corpus.
  - Our experimental result on the DBLP dataset shows the evolution of research topics across different years.
- Future Works:
  - Modify our model to enforce sparsity for the Λ matrix.
  - Generalize our implementation to model topic evolution in multiple corpora.
  - Design more effective visualization of our results.
  - Evaluate our algorithm with other metrics.

Tang, Jie, Zhang, Jing, Yao, Limin, Li, Juanzi, Zhang, Li, and Su, Zhong. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 990–998. ACM, 2008.

Teh, Yee Whye, Jordan, Michael I, Beal, Matthew J, and Blei, David M. Hierarchical dirichlet processes. *Journal of the american statistical association*, 2006.

Zhang, Jianwen, Song, Yangqiu, Zhang, Changshui, and Liu, Shixia. Evolutionary hierarchical dirichlet processes for multiple correlated time-varying corpora. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1079–1088. ACM, 2010.