Tracking Trends: Incorporating Term Volume into Temporal Topic Models

Liangjie Hong*, Dawei Yin*, Jian Guo§, Brian D. Davison*

Dept. of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA *
Dept. of Statistics, University of Michigan, Ann Arbor, MI, USA §

Outline

- motivation
- related work
- our model
- experiments
- conclusion & future work

Motivation

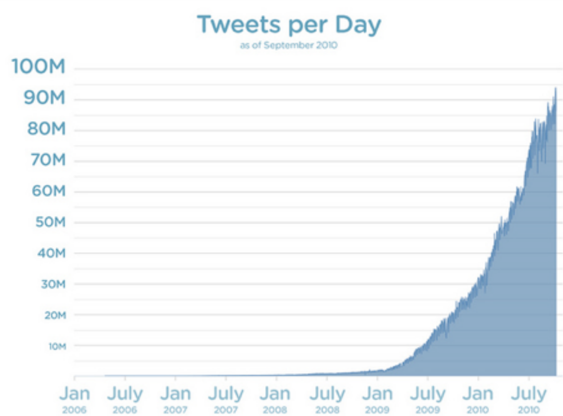
- ever-growing datasets
- “topics” or “interests” drift over time

YAHOO! ANSWERS

	U.S.	World
Users:	~~21~~ 25 million	~~95~~ 135 million
Answers:	~~160~~ 237 million	~~350~~ 500 million

Black = August 2007

RED = MARCH 2008

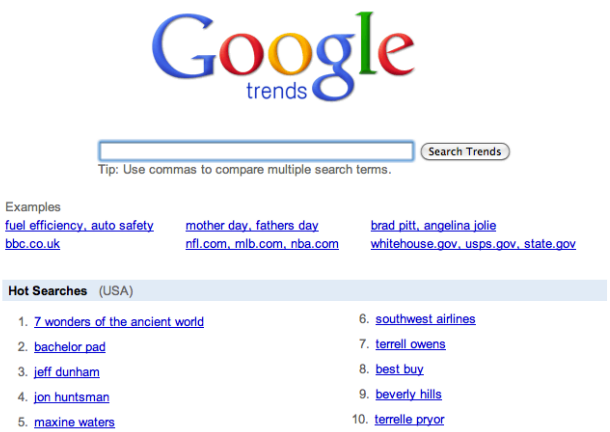


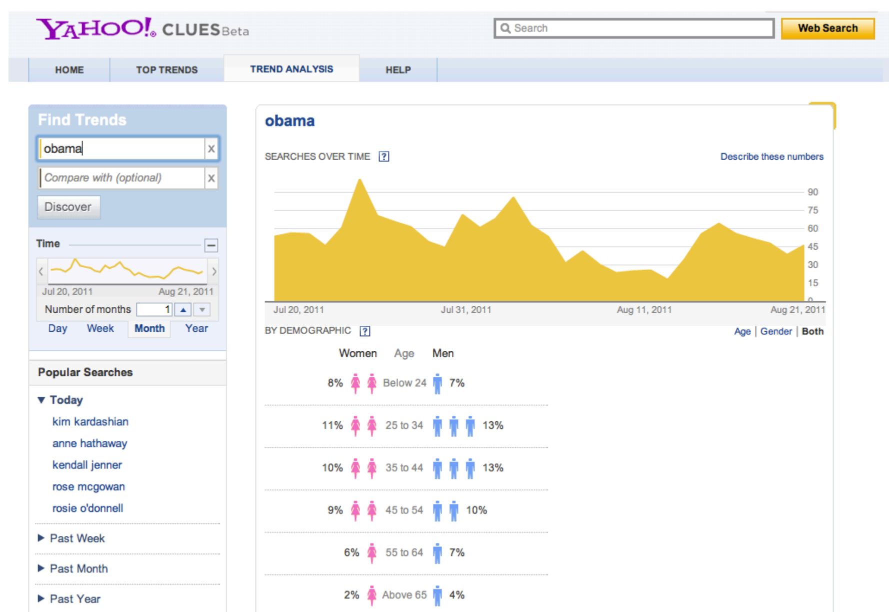
Motivation

- track the popularity of “topics”
- understand correlations between terms

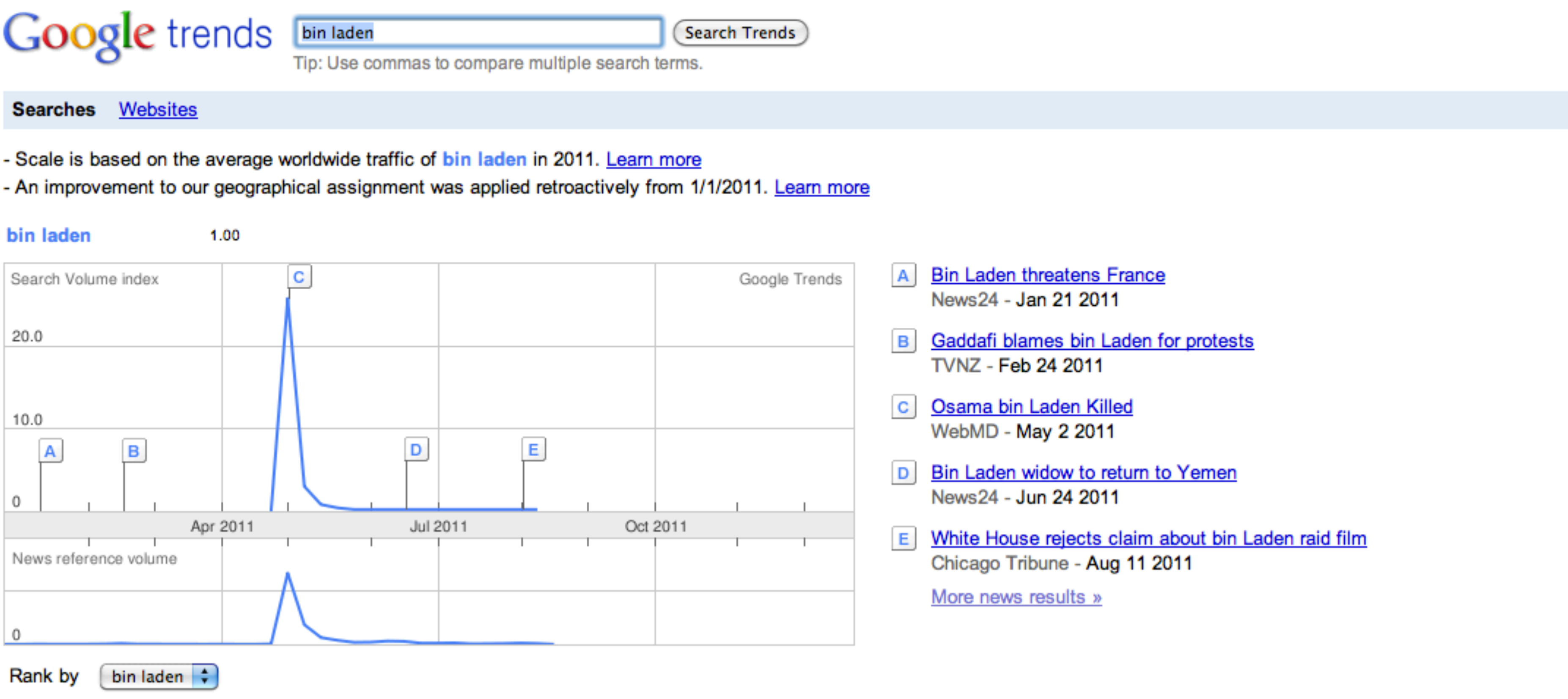
Motivation

- track the popularity of “topics”
- understand correlations between terms





tracking might be difficult …



tracking might be difficult …

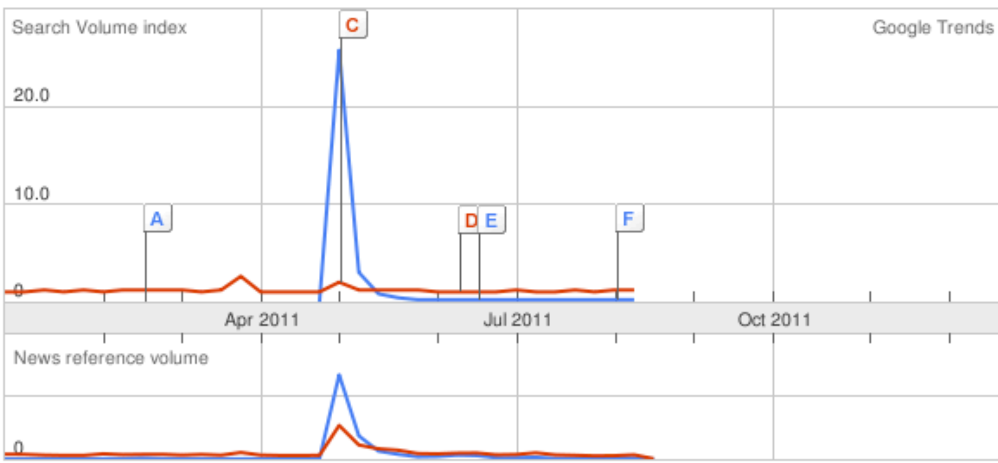
Google trends [bin laden, pakistan] Search Trends

Tip: Use commas to compare multiple search terms.

Searches Websites

- Scale is based on the average worldwide traffic of **bin laden** in 2011. Learn more
- An improvement to our geographical assignment was applied retroactively from 1/1/2011. Learn more

bin laden 1.00 **pakistan** 1.20



Rank by [bin laden]

A Gaddafi blames bin Laden for protests
TVNZ - Feb 24 2011

B Osama bin Laden Killed
WebMD - May 2 2011

C Osama bin Laden killed in Pakistan
Washington Post - May 2 2011

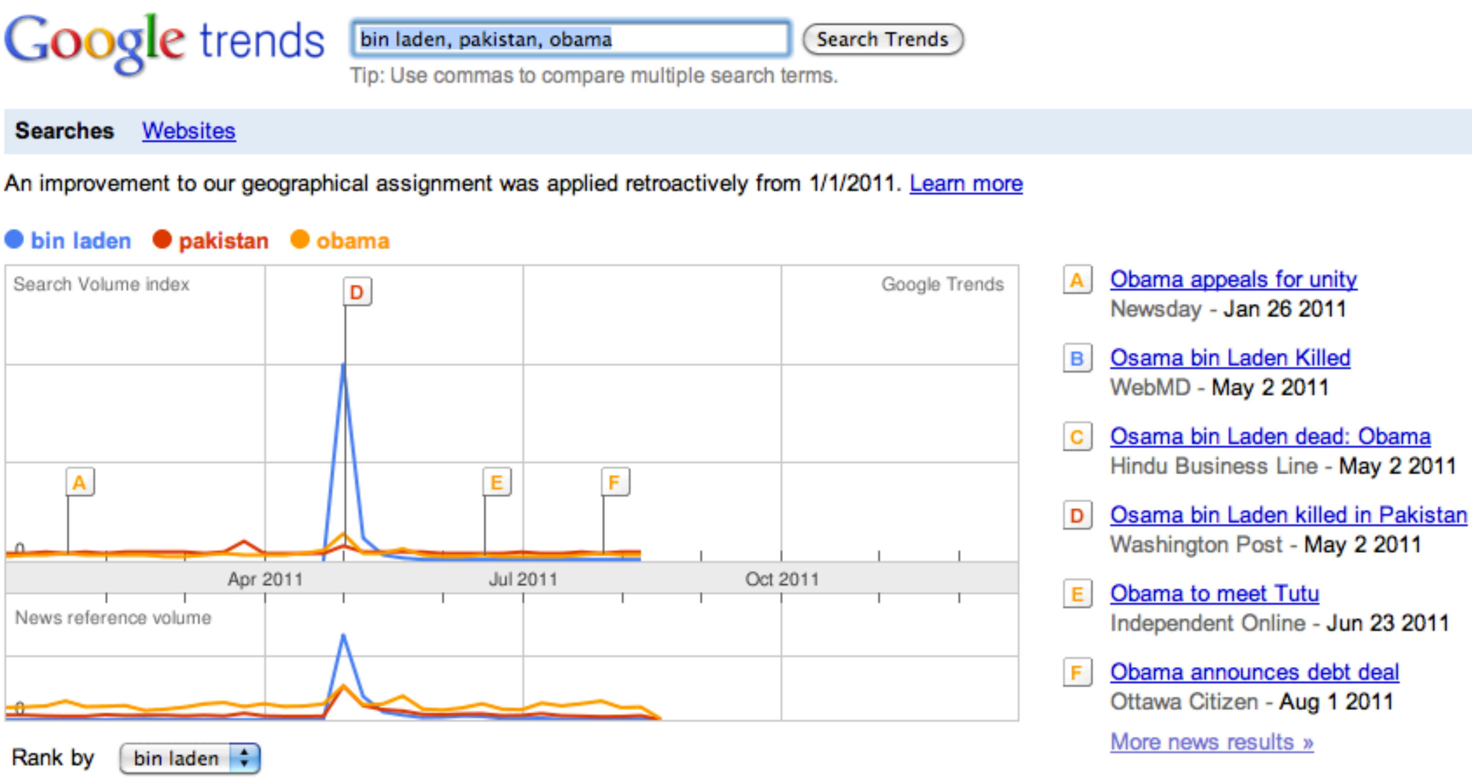
D Pakistan arrests CIA informants
Straits Times - Jun 15 2011

E Bin Laden widow to return to Yemen
News24 - Jun 24 2011

F White House rejects claim about bin Laden raid film
Chicago Tribune - Aug 11 2011

More news results »

tracking might be difficult …



Problems

- need to know all the keywords beforehand
- whether they are thematically related or not?

Problems

- need to know all the keywords beforehand
- whether they are thematically related or not?

- based on historical data, is it possible to do prediction?

Related Work

- temporal topic models

Related Work

- temporal topic models
  - general-purpose models
  - hard to evaluate

Related Work
evaluation of temporal topic models

Temporal Perplexity	[Blei & Lafferty, 2006] [Nallapati et al., 2007] [Wang et al., 2008] [Wang et al., 2009] [Wang et al., 2010] [Ahmed et al., 2010] [Iwata et al., 2010] [N. Kawamae et al., 2010]
Timestamp Prediction	[Wang et al., 2006] [Wang et al., 2008] [N. Kawamae et al., 2010]
Classification/Clustering	[Zhang et al., 2010]
Ad-hoc	[Wang et al., 2006] [Wang et al., 2009] [Zhang et al., 2010]

Problem Definition

Input:

- text documents, segmented into time epochs

Output:

- cluster terms
- track term volumes

Our Approach

two sub-tasks:

- cluster terms – temporal topic models
- tracking volumes – linear regression

Our Approach

two sub-tasks:

- cluster terms – temporal topic models
- tracking volumes – linear regression

- combine two tasks
- reinforce with each other

Input:

- text documents, segmented into time epochs

Output:

- cluster terms – topics (distribution over terms)
- track term volumes – volume regression

Input:

- text documents, segmented into time epochs

Output:

- cluster terms – topics (distribution over terms)
- track term volumes – volume regression

Term Volume Tracking

- topics as latent features
- term volumes as response

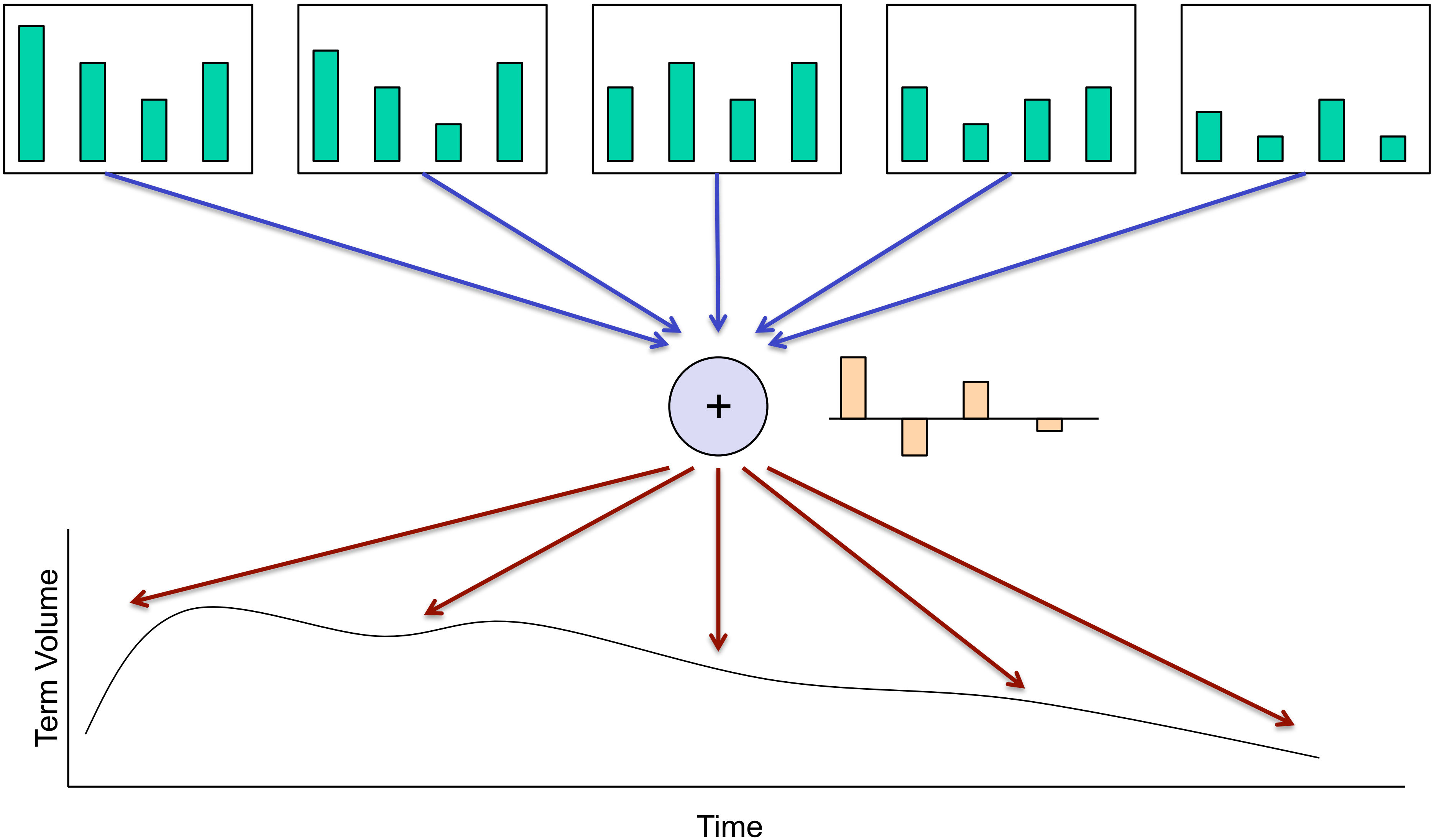
$$Y_v^{(t)} = \sum_{k=0}^{K} \pi_{(v,k)} \beta_{(k,v)}^{(t)} + \epsilon_v$$

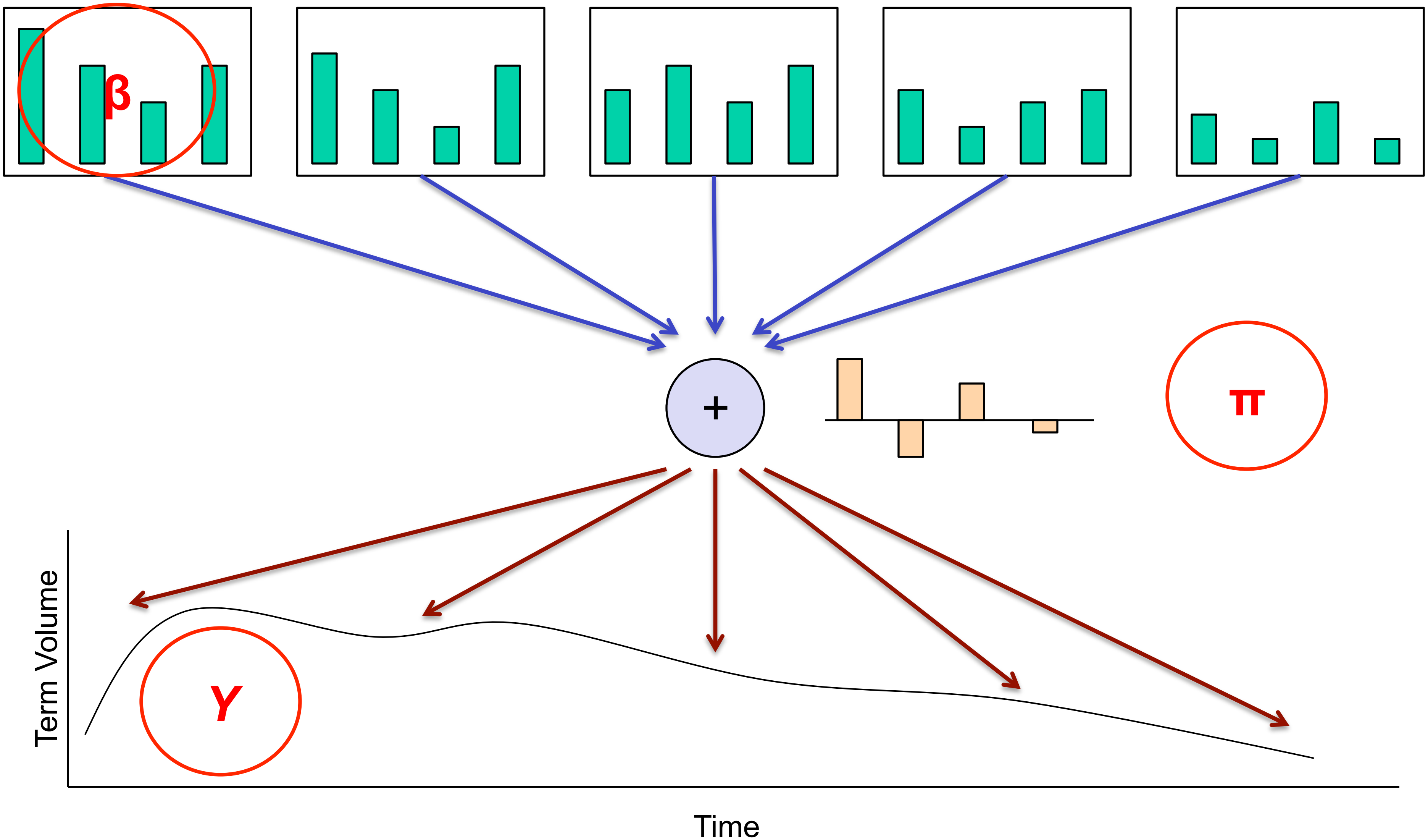
Term Volume Tracking

- topics as latent features
- term volumes as response

$$Y_v^{(t)} = \sum_{k=0}^{K} \pi_{(v,k)} \beta_{(k,v)}^{(t)} + \epsilon_v$$

- independence assumption


Term Volume
Time


β
+
π
Y
Term Volume
Time

Problem

- How to obtain β over time?

Problem

- How to obtain β over time?
  - state-space model
  - special proposed functions
  - linear combination

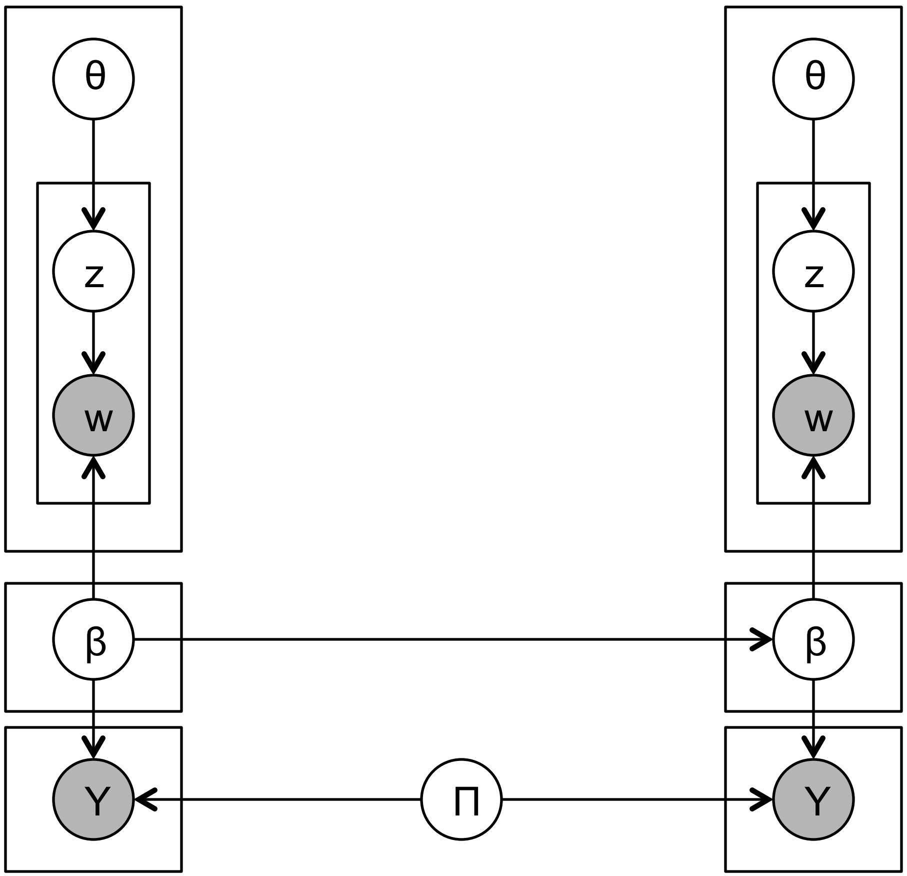
  …

Problem

- How to obtain β over time?
  - state-space model
  - special proposed functions
  - linear combination

  …

Incorporate Term Volumes with LDA



Incorporate Term Volumes with LDA

1. For each topic k in K:
   Draw topics $\beta_k^{(t)} \mid \beta_k^{(t-1)} \sim \mathcal{N}\left(\beta_k^{(t-1)}, \delta^2 I\right)$.
2. For each term v in V:
   Draw term volume $Y_v^{(t)} \sim \mathcal{N}\left(\pi_v^T \beta_{(*,v)}^{(t)}, \sigma^2\right)$.
3. For each document d in time epoch t:
   (a) Draw $\theta_d \sim \mathrm{Dir}(\alpha)$
   (b) For each word n:
      i. Draw $z_{(d,n)} \sim \mathrm{Multi}(\theta)$.
      ii. Draw $w_{(d,n)} \sim \mathrm{Multi}\left(\pi(\beta_z^{(t)})\right)$

Approximate Inference

Obtain

- per document: θ
- per word: z

- per time epoch, per topic: β
- per word: π

Approximate Inference

Obtain

- per document: θ
- per word: z

- per time epoch, per topic: β
- per word: π

- Gibbs Sampling
- Variational Inference

Approximate Inference

Obtain

- per document: θ
- per word: z

- per time epoch, per topic: β
- per word: π

- Gibbs Sampling
- Variational Inference

Variational Inference

Initialize β randomly.
while relative improvement in $L > 0.00001$ do
"E step":
for $t = 1$ to T do
    for $i = 1$ to D do
        Update λ_d
        Update ϕ_d
"M Step":
for $v = 1$ to V do
        Update π_v
        Update σ_v
for $t = 1$ to T do
    Update β_t by using Conjugate Gradient

Prediction:
state-space model's common practice

Experiments

– NIPS dataset:

4,360 papers with 38,029 distinct terms, 24 years.

– ACL dataset:

14,590 papers with 74,189 distinct terms, 37 years.

– Metric:

$$\mathrm{RMSE}_t = \sqrt{\frac{1}{V}\sum_v \left(\hat{Y}_v^{(t)} - Y_v^{(t)}\right)^2}$$

- Baselines
  - Univariate Autoregressive Model *AR(p)*:

$$X_t = w + \sum_{k=1}^{p} \pi_k X_{t-k}$$

  - Multivariate Autoregressive Model *MAR(p)*:

$$\mathbf{X}_t = \mathbf{w} + \sum_{k=1}^{p} \mathbf{A}_k \mathbf{X}_{t-k}$$

  - LDA
  - Dynamic Topic Model
    [Blei & Lafferty, 2006]

- Baselines
  - Univariate Autoregressive Model *AR(p)*:

$$X_t = w + \sum_{k=1}^{p} \pi_k X_{t-k}$$

  - Multivariate Autoregressive Model *MAR(p)*:

$$\mathbf{X}_t = \mathbf{w} + \sum_{k=1}^{p} \mathbf{A}_k \mathbf{X}_{t-k}$$

  - LDA
  - Dynamic Topic Model
    [Blei & Lafferty, 2006]

- Baselines

Table 2: AR model on NIPS dataset

p	2007	2008	2009	Avg.
1	98.57	90.51	99.42	96.17
2	101.72	83.20	**91.06**	92.00
3	**97.66**	77.31	97.00	**90.39**
4	112.83	**75.62**	95.98	94.81
5	118.10	91.64	108.33	106.03
6	118.65	99.00	108.34	108.66
7	118.76	98.99	117.50	111.75
8	122.73	95.93	116.72	111.79
9	122.55	96.23	115.85	111.54
10	143.17	100.71	124.40	122.76

- Baselines

Table 3: AR model on ACL dataset

p	2005	2006	2007	2008	2009	Avg.
1	**131.85**	524.04	**39.57**	592.91	**126.29**	282.93
2	210.74	316.38	106.31	434.15	181.98	249.91
3	247.73	248.17	104.72	**381.84**	140.87	**224.65**
4	258.74	246.58	114.23	447.71	166.09	246.67
5	244.41	**223.99**	53.12	428.17	185.00	226.94
6	250.49	297.98	42.74	385.26	209.24	237.14
7	169.25	328.75	51.14	345.98	262.54	231.53
8	168.54	332.20	51.58	396.08	291.13	247.90
9	155.96	326.73	47.11	400.96	291.60	244.47
10	156.59	355.13	49.15	399.28	310.65	254.16

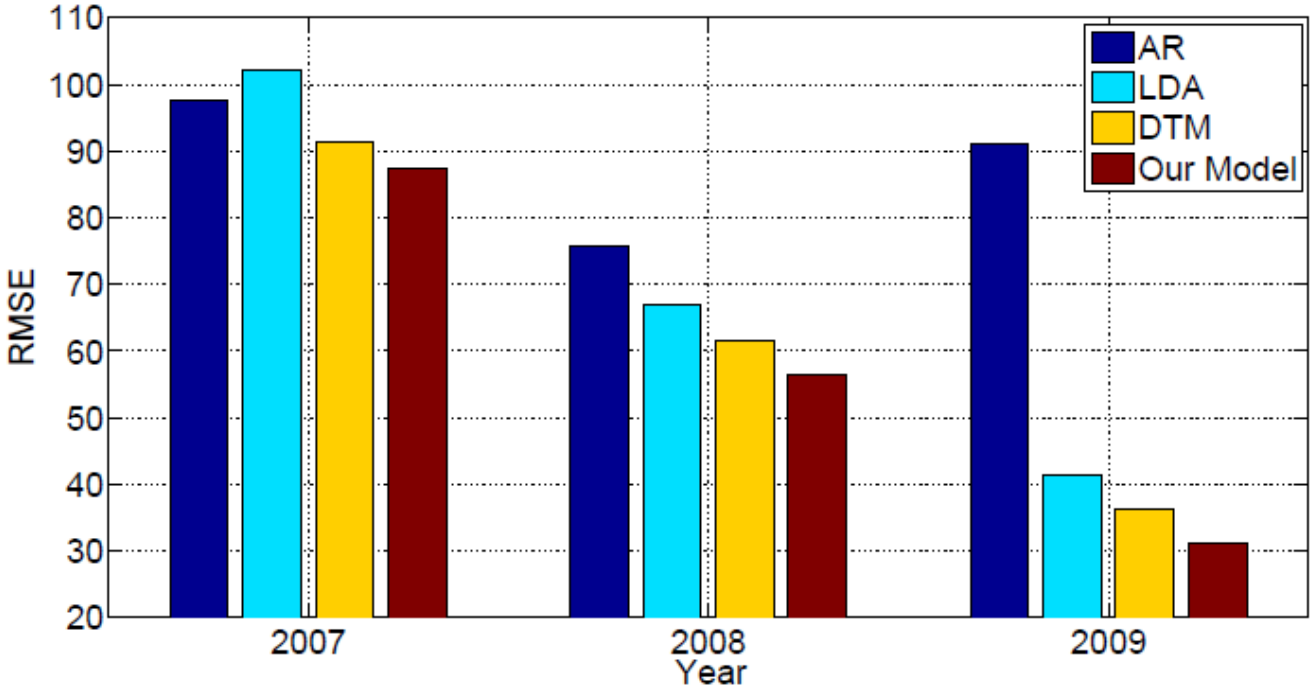


Figure 2: Performance comparison on NIPS dataset. The best RMSE values achieved by each model are shown for the last three years.

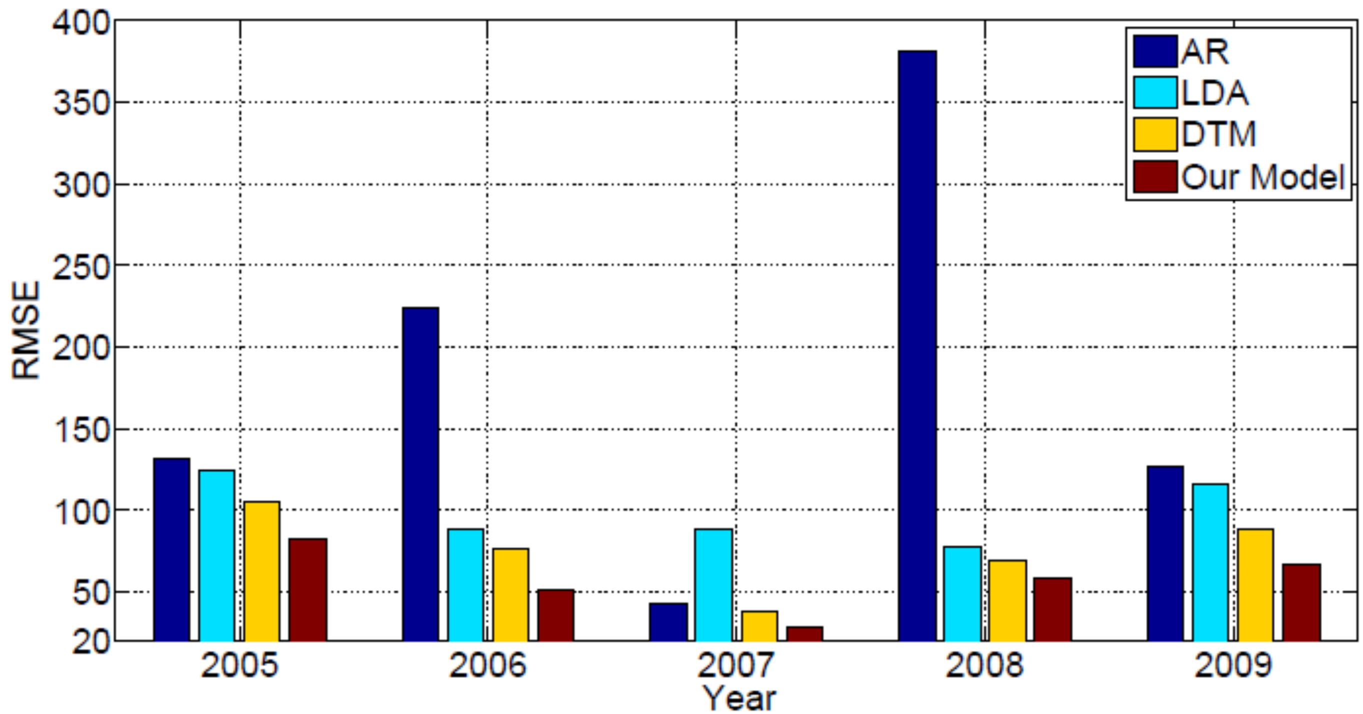


Figure 4: Performance comparison on ACL dataset. The best RMSE values achieved by each model are shown for the last five years.

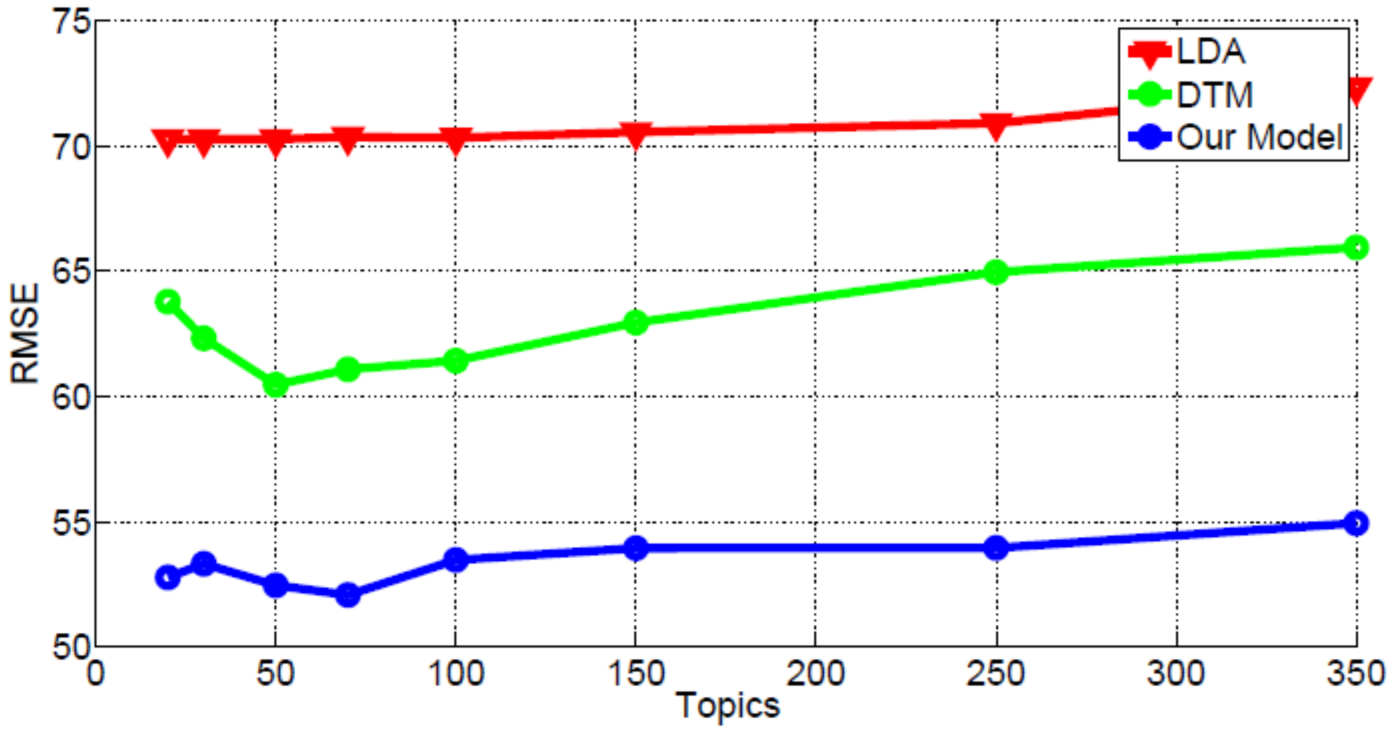


Figure 3: Performance comparison by varying the number of topics K on NIPS dataset.

Conclusion & Future work

- clustering terms + tracking terms
- topic modeling + state-space model
  + supervised learning
- latent features help prediction

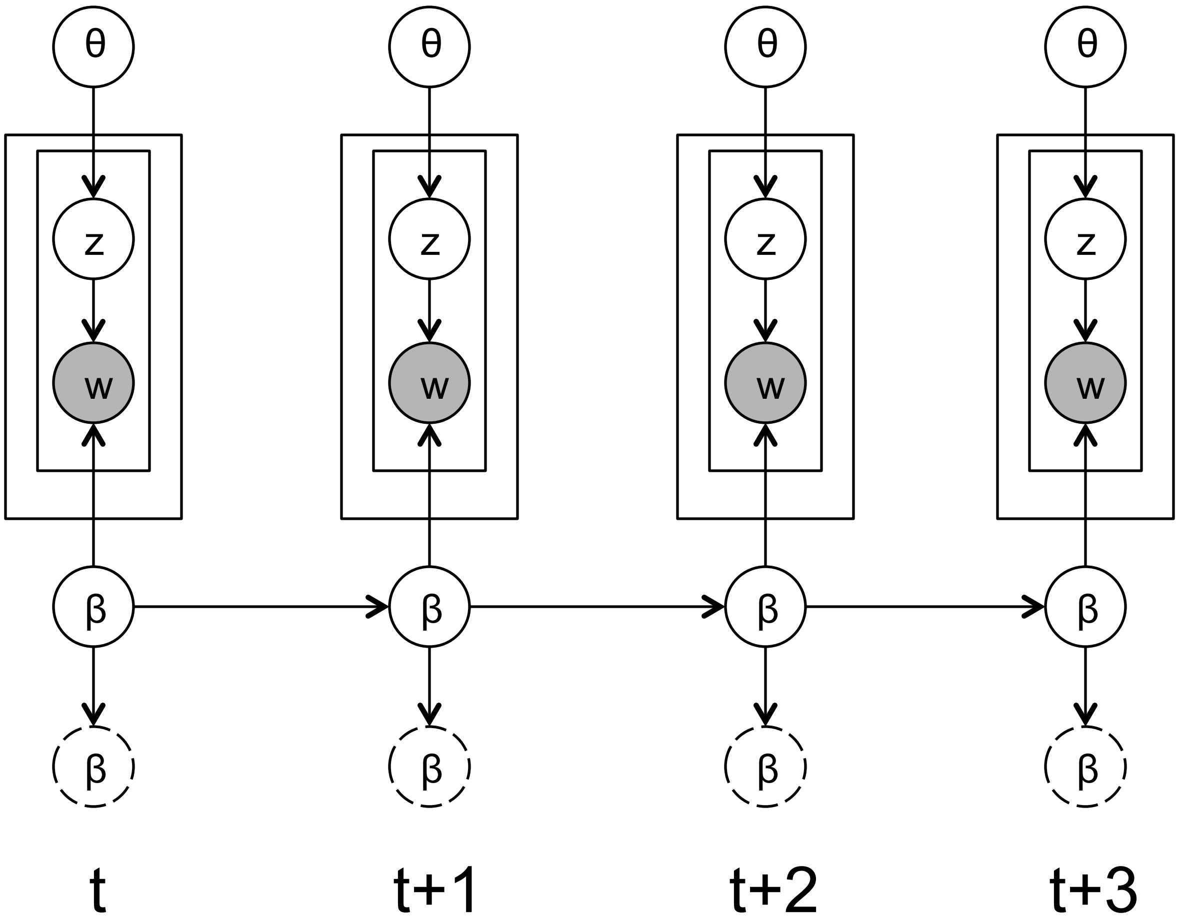
- explore other temporal models
  (see [Hong et al. 2011])
- capture correlations between terms
- explore more efficient inference algorithms

Thank you!

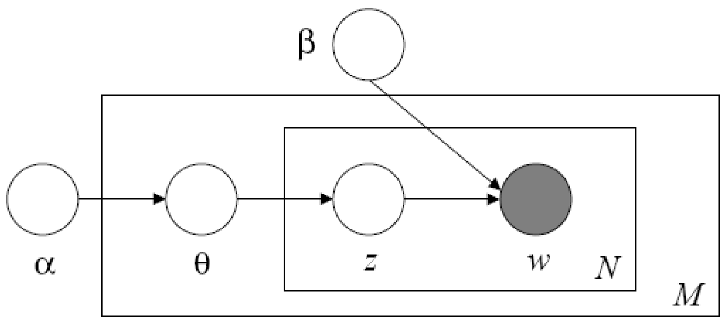
Contact Info:

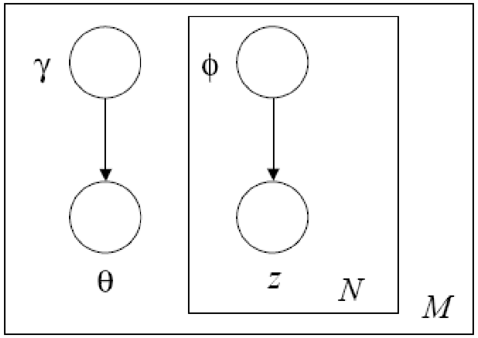
Liangjie Hong
hongliangjie@lehigh.edu
WUME Laboratory
Computer Science and Engineering
Lehigh University
Bethlehem, Pennsylvania, USA

Variational Inference with Kalman Filter



Variational Inference





In variational inference, we consider a simplified graphical model with variational parameters γ, ϕ and minimize the KL Divergence between the variational and posterior distributions.

$$(\gamma^*, \phi^*) = \arg\min_{(\gamma,\phi)} KL(q(\theta, z|\gamma, \phi)||p(\theta, z|w, \alpha, \beta))$$

- Variational Inference with Kalman Filter
  - State-space Model

$$
\begin{aligned}
\beta_k^{(t)} \mid \beta_k^{(t-1)} &\sim \mathcal{N}\left(\beta_k^{(t-1)}, \delta^2 I\right) \\
\hat{\beta}_k^{(t)} \mid \beta_k^{(t)} &\sim \mathcal{N}\left(\beta_k^{t}, \hat{\delta}_t^2 I\right)
\end{aligned}
$$

  - Two basic operations:
    - Smoothing

$$p(\mathbf{x}_t|\mathbf{y}_1, \cdots, \mathbf{y}_T)$$

    - Filtering

$$p(\mathbf{x}_t|\mathbf{y}_1, \cdots, \mathbf{y}_t)$$

- Variational Inference with Kalman Filter

$$q(\beta_{1:T}, \theta, \mathbf{Z} | \hat{\beta}_{1:T}, \lambda, \Phi) =$$
$$\prod_{k=1}^{K} q(\beta_k^1, \cdots, \beta_k^T | \hat{\beta}_k^1, \cdots, \hat{\beta}_k^T) \times$$
$$\prod_{t=1}^{T} \Big(\prod_{d=1}^{D_t} q(\theta_d | \lambda_d) \prod_{n=1}^{N_d} q(z_{(d,n)} | \phi_{(d,n)}) \Big)$$

The variational parameters are:

- Dirichlet λ_d
- Multinomial ϕ
- “Observations” for Kalman Filter $\hat{\beta}$