Efficient Collapsed Gibbs Sampling for Latent Dirichlet Allocation

Han Xiao, Thomas Stibor
{xiaoh, stibor}@in.tum.de

Department of Informatics
Technical University of Munich, GERMANY

Outline

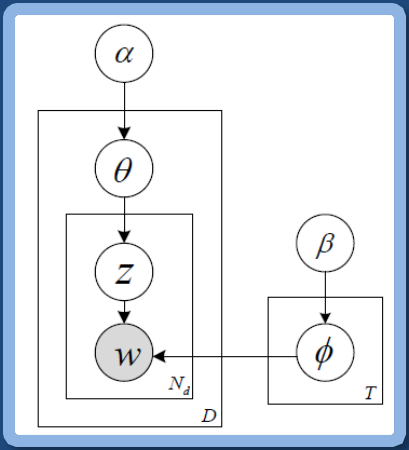
- Introduction
- Related work
- Shortcut sampling & Dynamic sampling
- Experimental results
- Summary

Introduction

- Latent Dirichlet allocation (LDA)
  - Variational EM (Blei et al. 2003)
  - Expectation propagation (Minka and Lafferty, 2002)
  - Collapsed Gibbs sampling (Griffiths and Steyvers, 2004)
- Collapsed Gibbs sampling (CGS)
  - ☺ Relatively simple
  - ☺ Highly extendable
  - ☹ BUT, low efficiency

Recap: LDA and CGS

- LDA model



$$
\begin{aligned}
\phi &\sim Dirichlet(\beta) \\
\theta &\sim Dirichlet(\alpha) \\
z_{di}|\theta_d &\sim Multinomial(\theta_d) \\
w_{di}|z_{di}, \phi_{z_{di}} &\sim Multinomial(\phi_{z_{di}})
\end{aligned}
$$

- CGS: sampling from posterior probability

$$P(z_{di} = k \mid w_{di} = v, \mathbf{W}_{\neg w_{di}}, \mathbf{Z}_{\neg z_{di}}, \alpha, \beta) \propto (C_{dk} + \alpha) \frac{C_{vk} + \beta}{\sum_{v'} C_{v'k} + V\beta}.$$

Pseudo-code of CGS

```
1  foreach d ∈ {1, ··· , D}
2      foreach i ∈ {1, ··· , N_d}
3          v ← w_di, I_di ← N_di
4          foreach j ∈ {1, ··· , I_di}
5              k̂ ← z_dij
6              C_dk̂ ← C_dk̂ − 1, C_vk̂ ← C_vk̂ − 1
7              for k = 1 to K do
8                  ρ_k ← ρ_{k−1} + (C_dk + α) × (C_kv + β)/(Σ_{v'} C_{v'k} + Vβ)
9              x ~ Uniform(0, ρ_K)
10             k̂ ← BinarySearch(k̂ : ρ_{k̂−1} < x < ρ_k̂)
11             C_dk̂ ← C_dk̂ + 1, C_vk̂ ← C_vk̂ + 1
12             z_dij ← k̂
```

Complexity: O(KW)

Related work

1 **foreach** $d \in \{1, \cdots, D\}$
2 **foreach** $i \in \{1, \cdots, N_d\}$
3 $v \leftarrow w_{di}, \mathcal{I}_{di} \leftarrow N_{di}$
4 **foreach** $j \in \{1, \cdots, \mathcal{I}_{di}\}$
5 $\hat{k} \leftarrow z_{dij}$
6 $C_{d\hat{k}} \leftarrow C_{d\hat{k}} - 1, C_{v\hat{k}} \leftarrow C_{v\hat{k}} - 1$
7 **for** $k = 1$ **to** K **do**
8 $\rho_k \leftarrow \rho_{k-1} + (C_{dk} + \alpha) \times (C_{kv} + \beta)/(\sum_{v'} C_{v'k} + V\beta)$
9 $x \sim \texttt{Uniform}(0, \rho_K)$
10 $\hat{k} \leftarrow \texttt{BinarySearch}(\hat{k} : \rho_{\hat{k}-1} < x < \rho_{\hat{k}})$
11 $C_{d\hat{k}} \leftarrow C_{d\hat{k}} + 1, C_{v\hat{k}} \leftarrow C_{v\hat{k}} + 1$
12 $z_{dij} \leftarrow \hat{k}$

PLDA (Wang et al. 2009)
LDA-GPU (Yan et al. 2009)

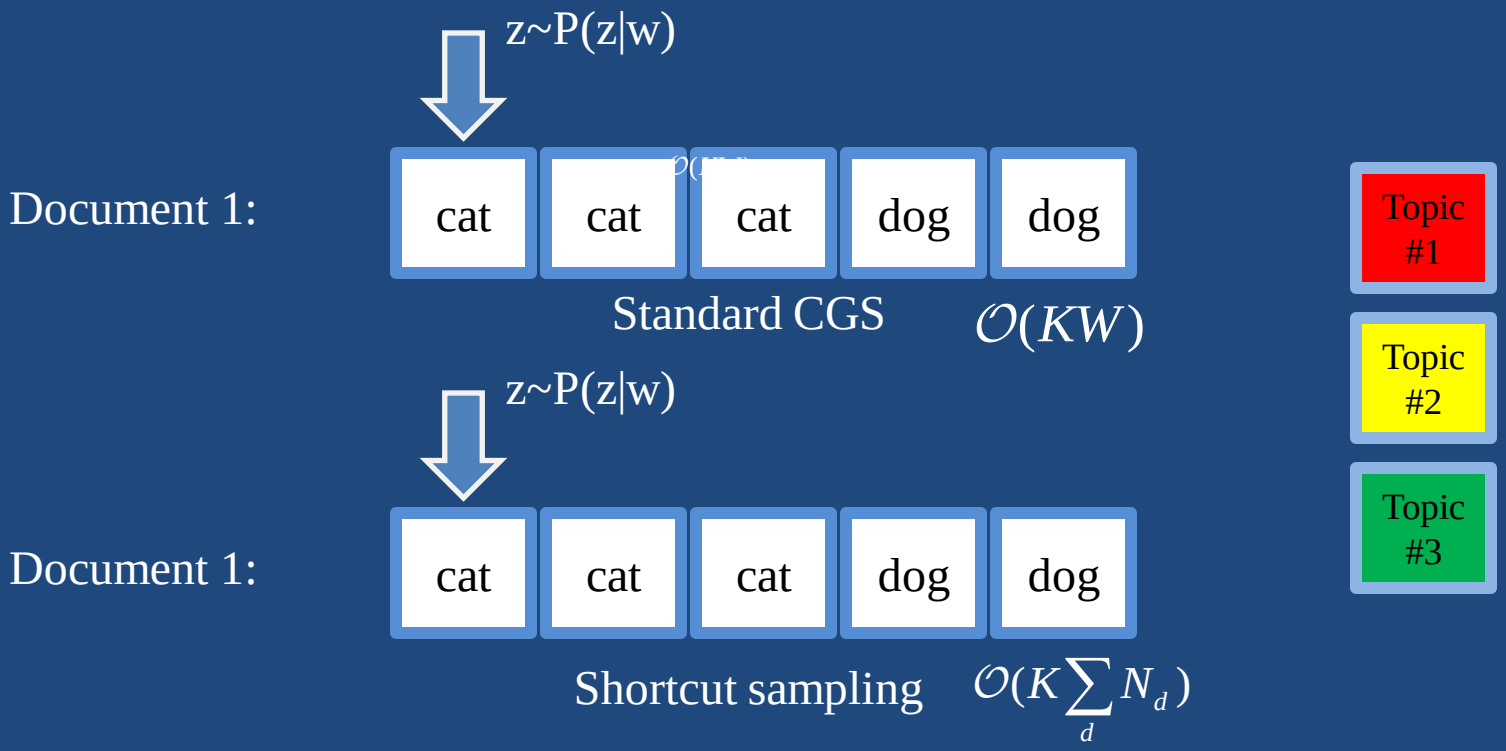
FastLDA (Porteous et al. 2008)
SparseLDA (Yao et al. 2009)

Our work

- Motivated by the sparseness of P(z|w)
- Reducing total sampling times
- A clean, simple but effective sampling strategy

Shortcut sampling

- Only sampling for different word types in document.



Sampling rate

- Sampling rate for a word type:

$$\mathcal{S}_{di}^{t} = \frac{\mathcal{I}_{di}}{N_{di}}, \text{ where } \mathcal{S}_{di}^{t} \in [0,1].$$

- Average sampling rate for a collection:

$$\overline{\mathcal{S}}^{t} = \frac{\sum_{d=1}^{D}\sum_{i=1}^{N_d}\mathcal{S}_{di}^{t}N_{di}}{W}, \text{ where } \overline{\mathcal{S}}^{t} \in [0,1].$$

High sampling rate, slow speed

	S_{cat}	S_{dog}	Avg. S
Standard	1.00	1.00	1.00
Shortcut	0.33	0.50	0.40

- Standard CGS:
  - Maximum sampling rate
  - Slowest
  - Optimality
- Shortcut sampling:
  - Minimum sampling rate
  - Fastest
  - Only sub-optimum

Dynamic sampling

- Dynamic sampling rate
- Fast
- Retaining optimality

Dynamic sampling

- Gradually decrease the sampling rate

$$1 \geq \overline{S}^1 \geq \overline{S}^2 \geq \cdots \geq \overline{S}^t$$

- Modeling sampling rate of each type as a random variable
- An auxiliary multinomial in Gibbs sampling procedure

Dynamic sampling

1 **foreach** $d \in \{1, \cdots, D\}$
2 **foreach** $i \in \{1, \cdots, N_d\}$
3 $v \leftarrow w_{di}$, $\mathcal{M} \leftarrow \emptyset$
4 $\mathcal{I}_{di} \sim \texttt{multinomial}(P(\mathcal{I}_{di}|v, \Gamma_{di}))$
5 **foreach** $j \in \{1, \cdots, \mathcal{I}_{di}\}$
6 $\hat{k} \leftarrow z_{dij}$
7 $C_{d\hat{k}} \leftarrow C_{d\hat{k}} - 1$, $C_{v\hat{k}} \leftarrow C_{v\hat{k}} - 1$
8 $\hat{k} \sim \texttt{multinomial}(P(z|w_{di}))$
9 $C_{d\hat{k}} \leftarrow C_{d\hat{k}} + 1$, $C_{v\hat{k}} \leftarrow C_{v\hat{k}} + 1$
10 $z_{dij} \leftarrow \hat{k}$
11 $\mathcal{M} \leftarrow \mathcal{M} \cup \{\hat{k}\}$
12 $u \leftarrow |\mathcal{M}|$, $\Gamma_{diu} \leftarrow \Gamma_{diu} + 1$

$\Gamma_{cat} = [0, 0, 0, 1]$ $\mathcal{I}_{cat} = 1$ $\Gamma_{dog} = [0, 1]$ $\mathcal{I}_{dog} = 2$

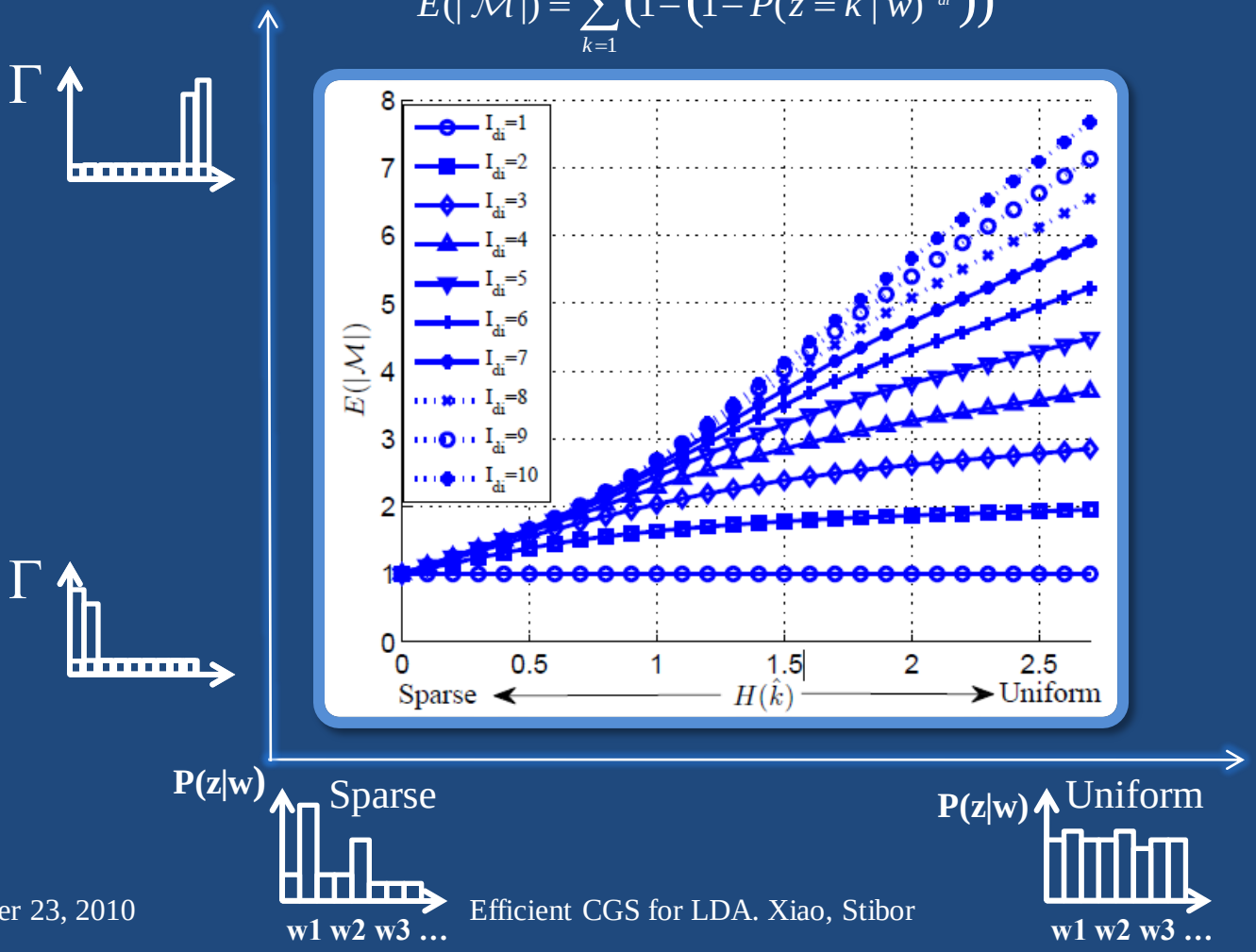
cat	cat	cat	cat	dog	dog

$\Gamma_{cat} = [0, 0, 1, 1]$ $\mathcal{I}_{cat} = 3$ $\Gamma_{dog} = [0, 2]$ $\mathcal{I}_{dog} = 2$

Sparseness of P(z|w)

- Expected number of different topics in I_{di} times sampling:

$$E(|\mathcal{M}|) = \sum_{k=1}^{K} \left(1 - \left(1 - P(z = k \mid w)^{\mathcal{I}_{di}}\right)\right)$$



Initialization of Γ

Standard CGS ← Dynamic sampling → Shortcut sampling

$\Gamma_{cat} = [0, \cdots, 0, \infty]$

Max sampling rate

$\mathcal{O}(KW) >$

$\Gamma_{cat} = [0, \cdots, 0, \gamma]$,
where γ is a small integer

Random sampling rate

$\Gamma_{cat} = [1, 0, \cdots, 0]$

Min sampling rate

$> \mathcal{O}(K \sum_d N_d)$

Experimental results

- Three data sets: KOS, NIPS and NYTimes
  - http://archive.ics.uci.edu/ml/datasets/Bag+of+Words

Name	#Doc	#Token	#Vocabulary
KOS	3,430	0.4×10^6	6,906
NIPS	1,500	6.4×10^6	12,419
NYTimes	300,000	1.0×10^8	102,660

- Baselines:
  - GibbsLDA (http://sourceforge.net/gibbslda/)
  - FastLDA (Porteous et al. 2008)
  - ECGS-Shortcut
  - ECGS-Dynamic

- Parameters:

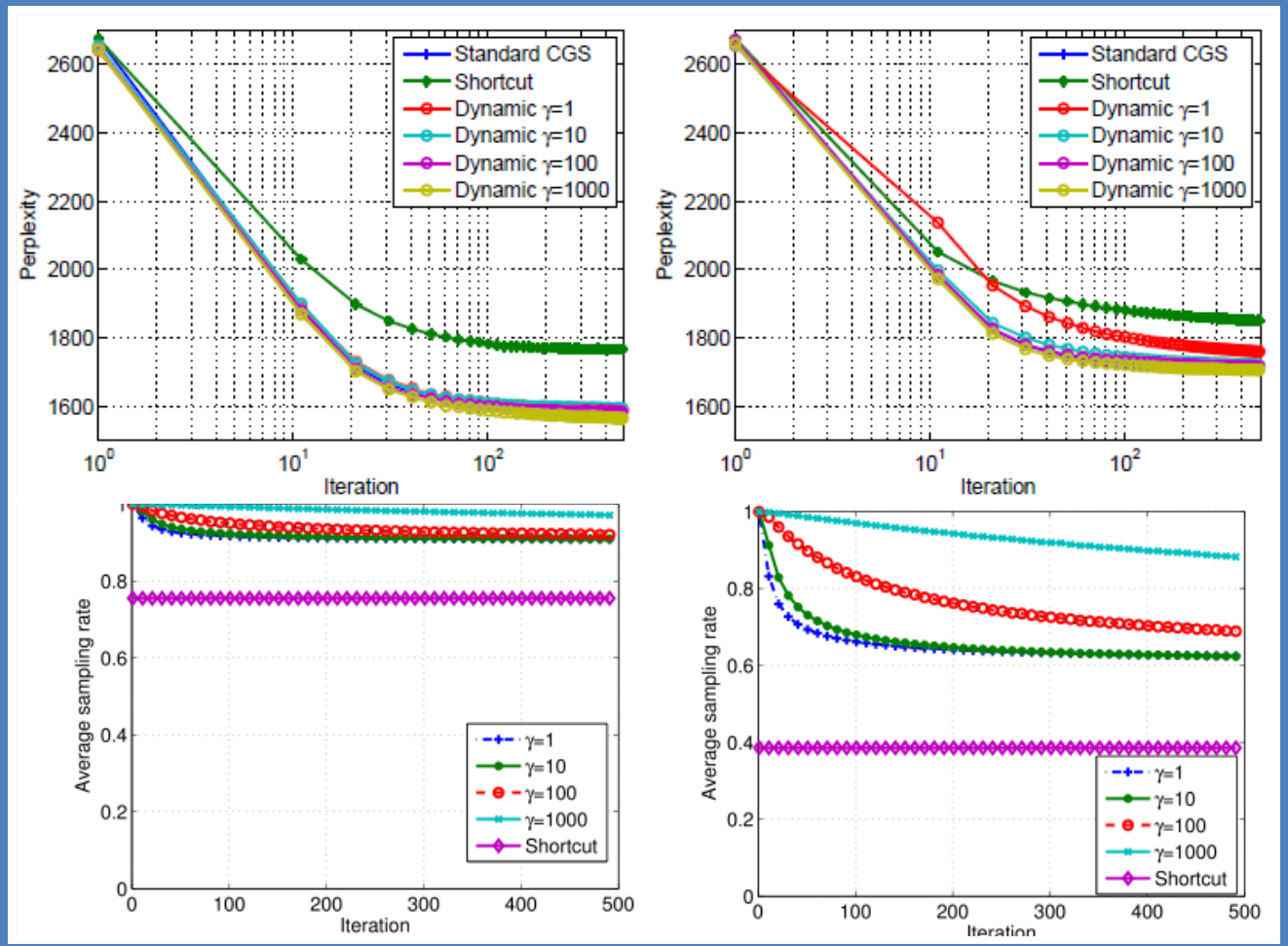
  -alpha 50/#topic
  -beta 0.02
  -4 fold cross validation
  -random initialization
  -compile using gcc –o3

Convergence Analysis

KOS NIPS

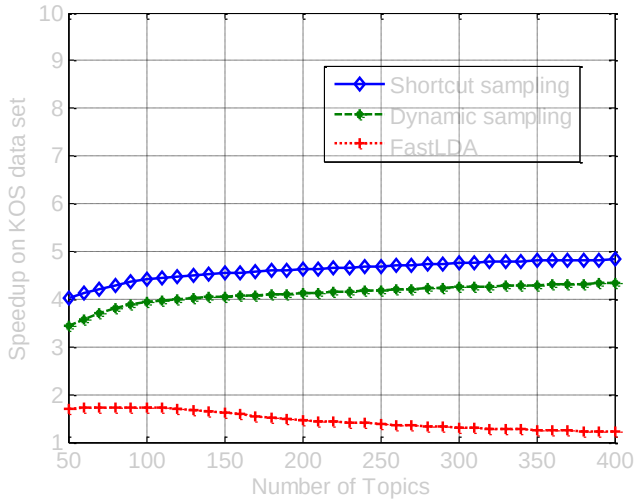
Perplexity

Avg. sampling rate

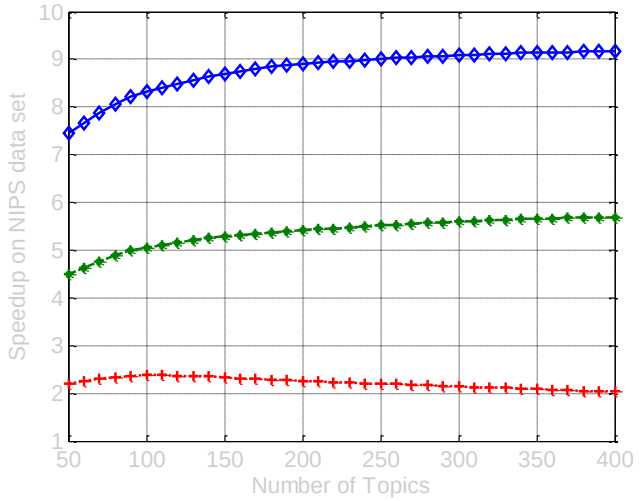


Speedup Results

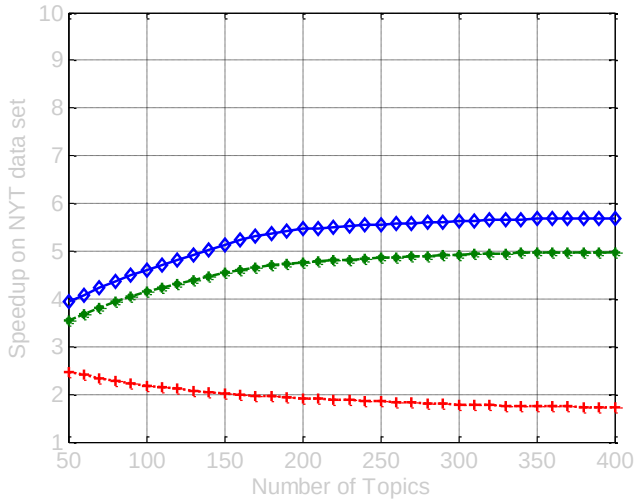
KOS



NIPS



NYTimes



Data set	$\sum_d N_d / W$
KOS	0.755
NIPS	0.385
NYTimes	0.700

Summary

- A simple but effective sampling strategy
  - Using auxiliary multinomial distribution
- Retains optimality guarantees ✓
- Unsensitive to the parameter
- Provides significant speedup
  - 3-6x over GibbsLDA
  - 2-3x over FastLDA
- Highly extendable (parallelization)
- http://home.in.tum.de/~xiaoh/pub/ecgs.html

Thanks!

Q&A