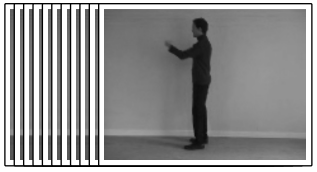
Supervised Hierarchical Dirichlet Processes with Variational Inference

C. Zhang, C. H. Ek, X. Gratal, F.T. Pokorny and H. Kjellström

CVAP/CAS,KTH Royal Institute of Technology

About Classification


Boxing


Hand-clapping


Hand-waving

Action Classification


Headphone


Water Lilly


Lamp

Object Classification

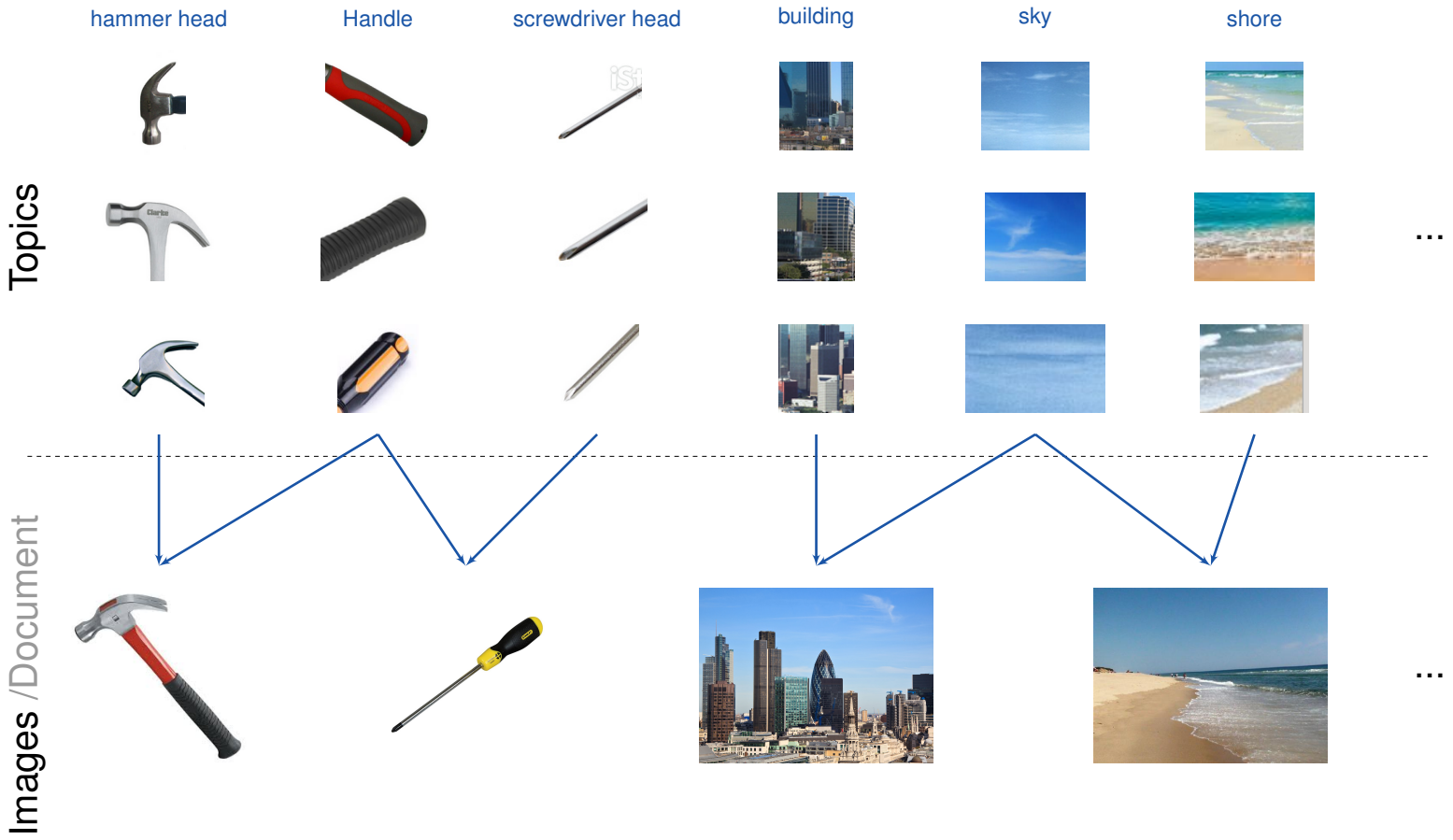

Mountain


Forest

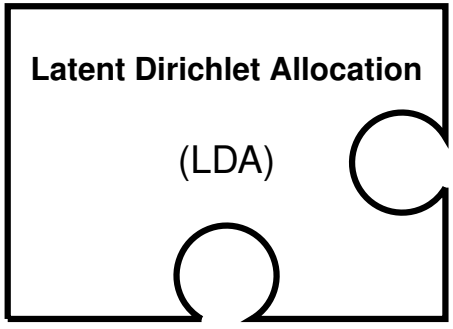

Coast

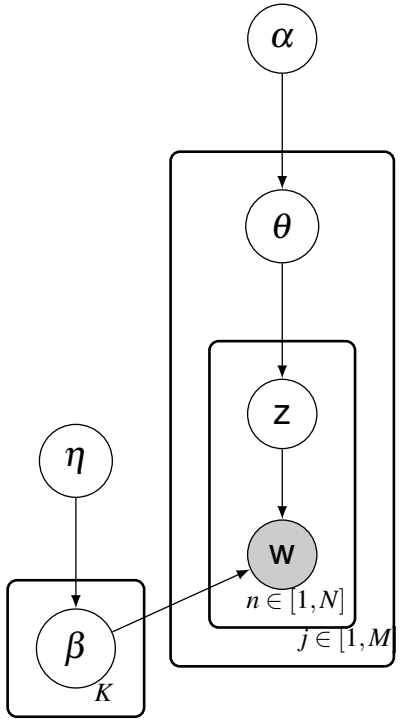
Scene Classification

Introduction

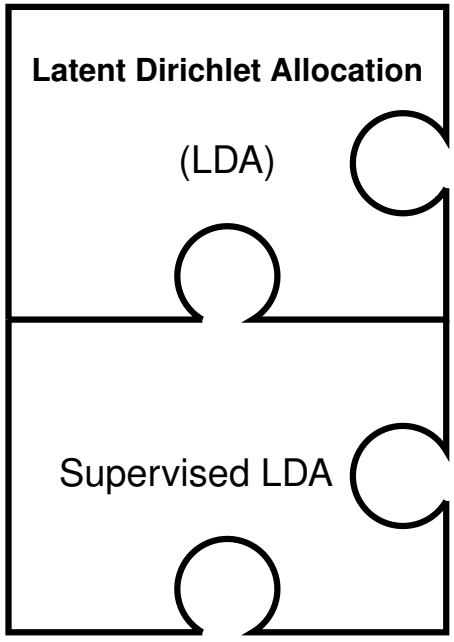


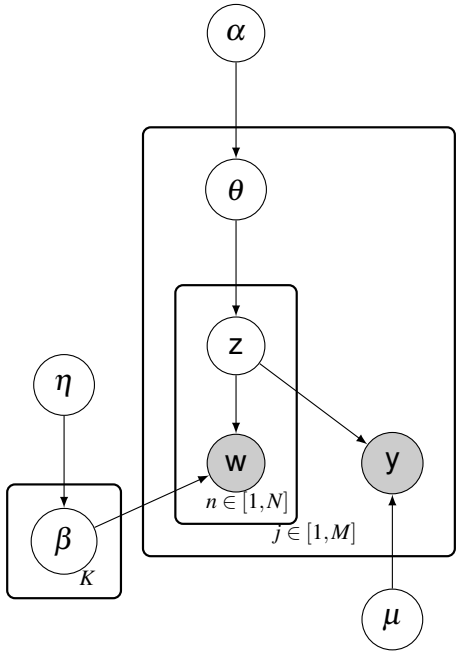
Topic Modeling





Supervised Topic Model

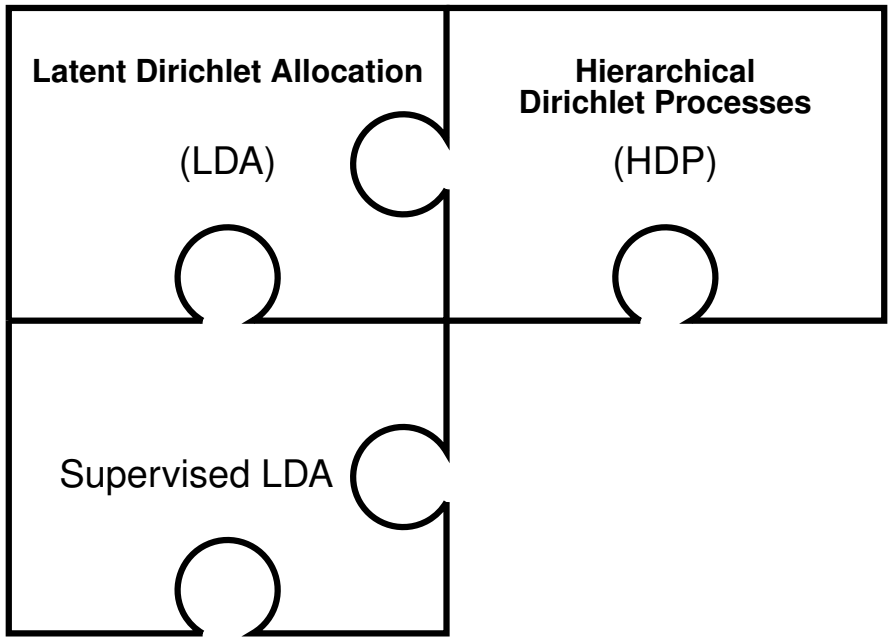




1

[1] D. M. Blei and J. D. McAuliffe. Supervised topic models, arxiv:1003.0783, 2010.

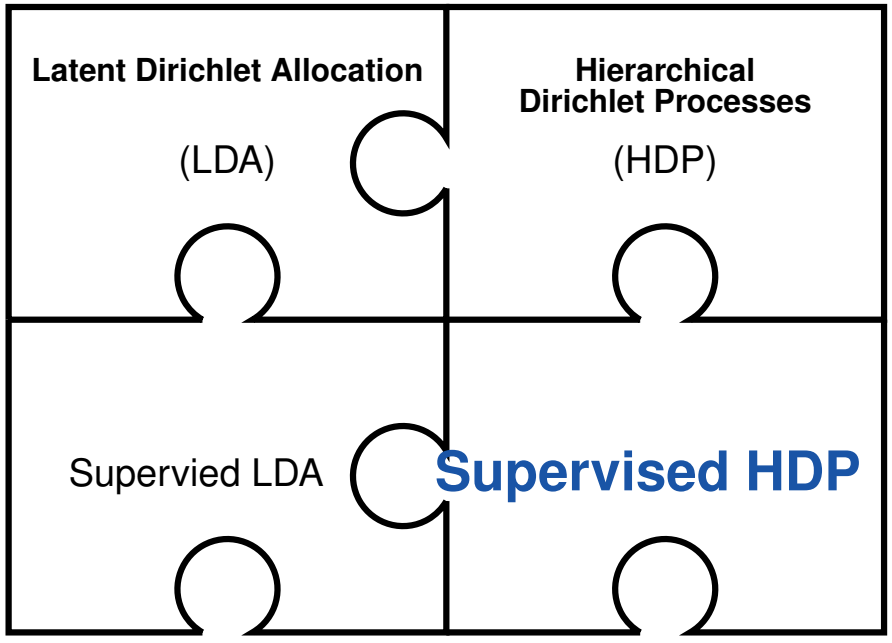
Nonparametric Topic Model



2

[2] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. Hierarchical Dirichlet processes. Journal of the American Statistical Association, 101(476):1566-1581, 2006.

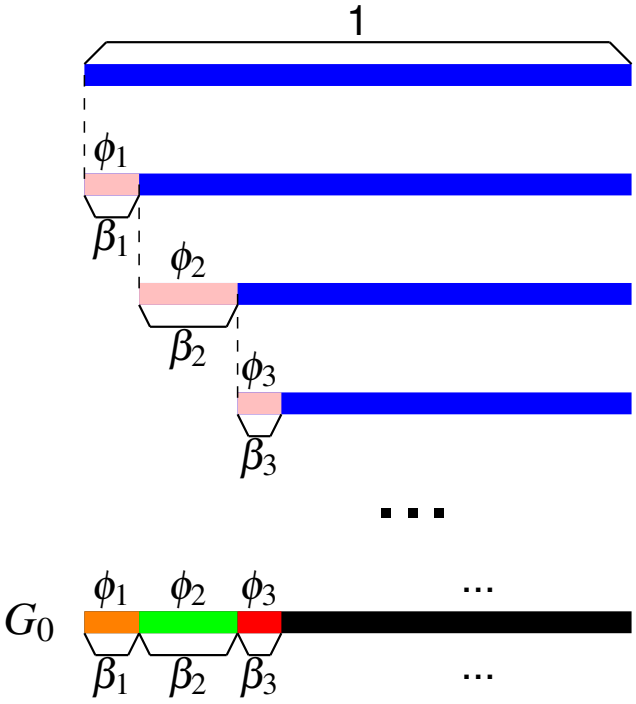
Supervised Nonparametric Topic Model



Outline

1. Dirichlet Process
2. Hierarchical Dirichlet Processes
3. Supervised HDP
4. Variational Inference for SHDP
5. Experiment
    a) Natural Scene Classification
    b) Action Classification

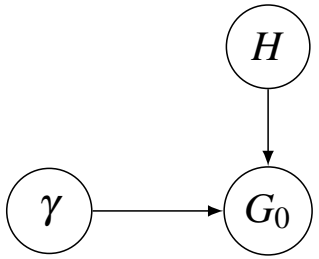
Dirichlet Process– Stick Breaking Construction



$$G_0 \sim DP(H, \gamma)$$

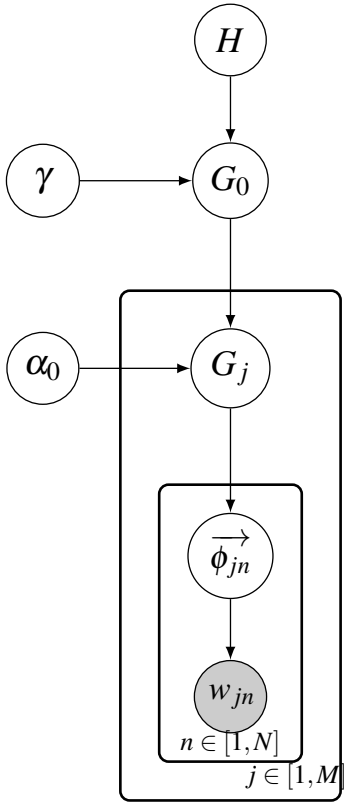
$$\beta'_k \sim Beta(1, \gamma),$$
$$\beta_k = \beta'_k \prod_{l=1}^{k-1}(1 - \beta'_l),$$
$$\phi_k \sim H,$$
$$G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k}.$$

β:s are general decreasing but not strictly



[3] J. Sethuraman. "A constructive definition of Dirichlet priors", Statistica Sinica, 4, pp.639-650, 1994

Hierarchical Dirichlet Processes (HDP)



M: number of documents/images/videos

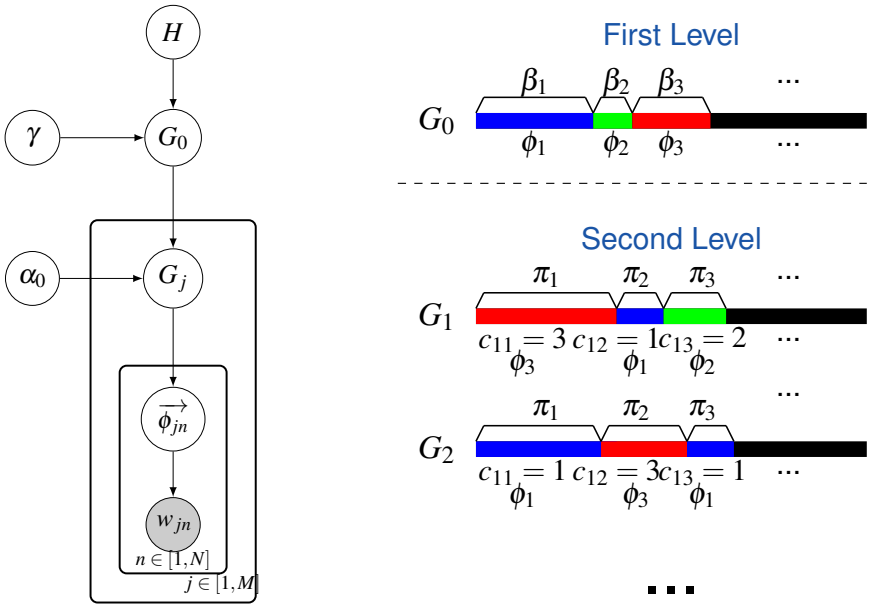
N: number of words/visual words

in each document/image/video

$$G_0 \sim DP(H, \gamma)$$

$$G_j \sim DP(G_0, \alpha_0)$$

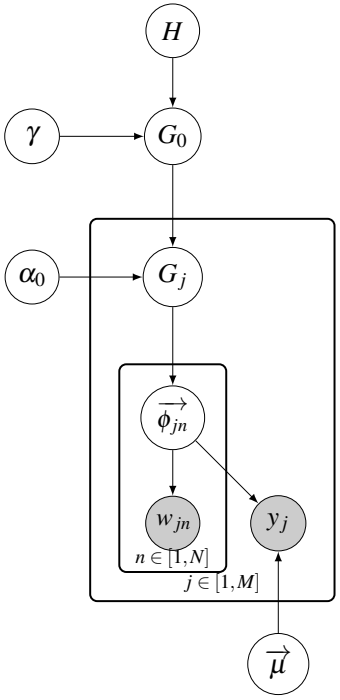
HDP – Stick Breaking Construction

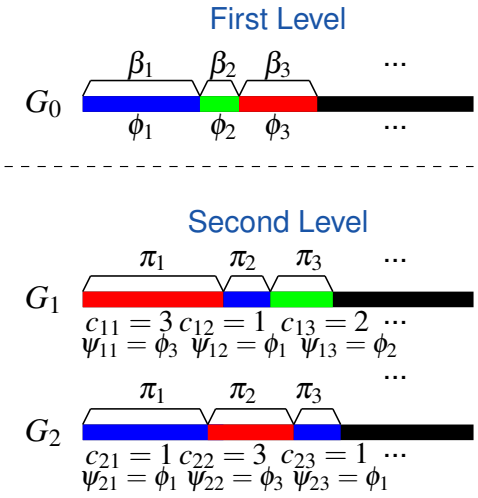


4

[4] C. Wang, J. Paisley, and D. Blei. Online variational inference for the Hierarchical Dirichlet Process. In AISTATS, 2011.

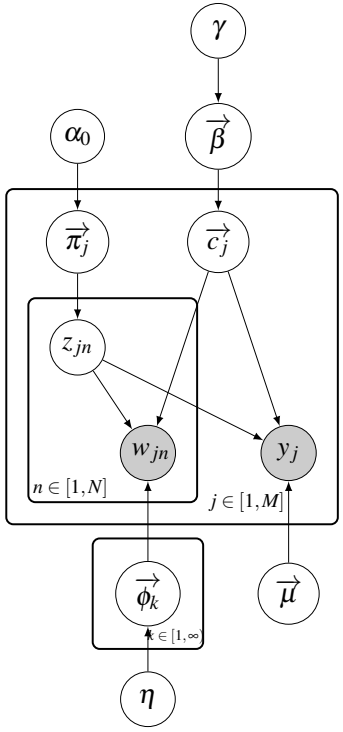
SHDP – Stick Breaking Construction



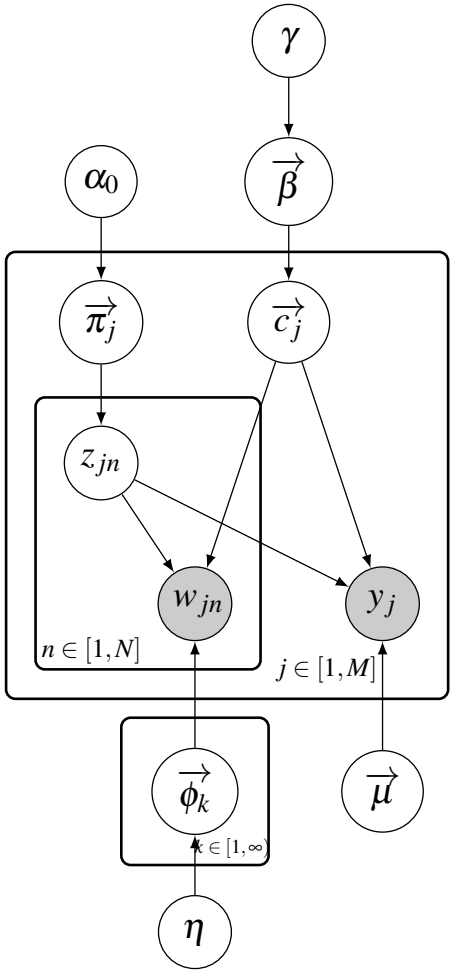


...

$$c_{jt} \sim Mult(\beta)$$



Supervised HDP



Softmax Function

$$p(y_j|\bar{\theta}_j,\mu) = \frac{\exp(\mu_{y_j}^T \bar{\theta}_j)}{\sum_{l=1}^{C} \exp(\mu_l^T \bar{\theta}_j)}$$

$$\bar{\theta}_j = \frac{1}{N}\sum_{n=1}^{N} \Theta_{jn} = \frac{1}{N}\sum_{n=1}^{N} \mathfrak{c}_{jz_{jn}}$$

Classification

$$\hat{y}_j = argmax_{y_j \in \{1,\ldots,C\}} \mathbb{E}[\mu_{y_j}^T \bar{\theta}_j]$$

Variational Inference for SHDP

Use a fully factorized variational distribution.

$$q(\beta', \pi', \mathbf{c}, \mathbf{z}, \phi) = q(\beta')q(\pi')q(\mathbf{c})q(\mathbf{z})q(\phi),$$

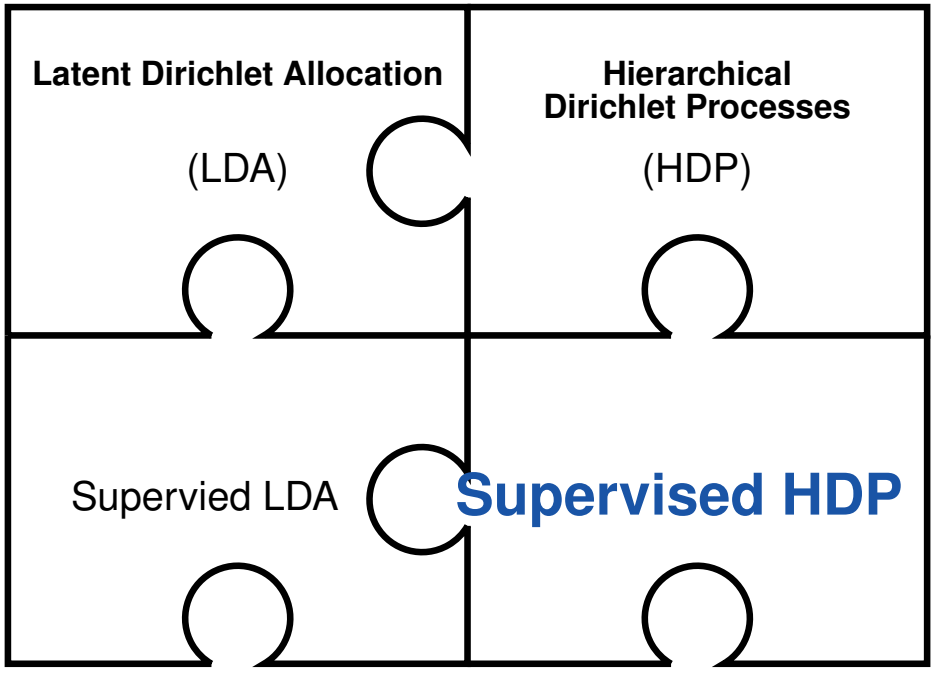
The supervision term in the lower bound:

$$\mathbb{E}_q[\log p(y_j|\mathbf{z}, \mathbf{c}, \mu)] \geq \mu_{y_j}^T (\frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{T} \rho_{jt} \zeta_{jnt})$$
$$- \log \left(\sum_{l=1}^{C} \prod_{n=1}^{N} \left(\sum_{e=1}^{K} \sum_{i=1}^{T} \rho_{jie} \zeta_{jni} \exp(\frac{1}{N} \mu_{le}) \right) \right).$$

Nonlinear function with ρ and μ.

Online: $\mathscr{L} = \sum_j^M \mathscr{L}_j = \mathbb{E}_j[M\mathscr{L}_j]$.

SHDP Summary



Supervised Model

It can be used to classification tasks.

Non-parametric Model

Data determine the number of topics.

Hierarchical Structure

It provides a better topic presentation.

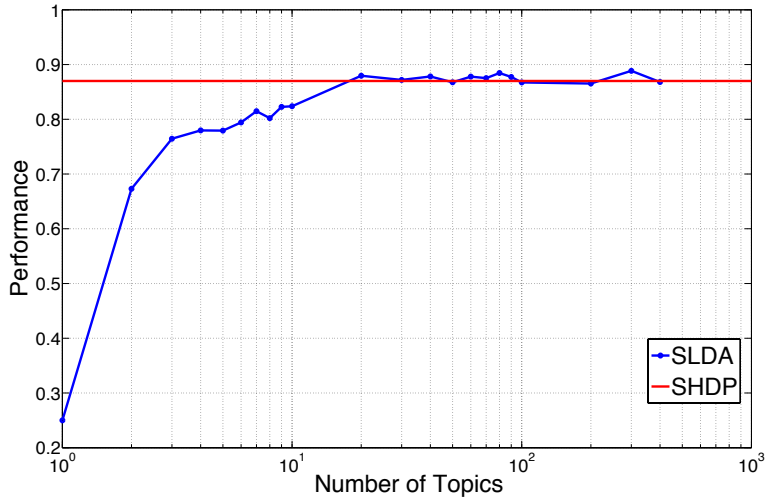
Variational Inference

It can be used with large dataset.

Experiment–Natural Scene Classification

	coast	forest	mountain	opencountry
coast	0.78	0	0	0.22
forest	0	0.95	0.03	0.02
mountain	0	0.01	0.89	0.09
opencountry	0.1	0.01	0.05	0.84

(a) Confusion Matrix 86.68%



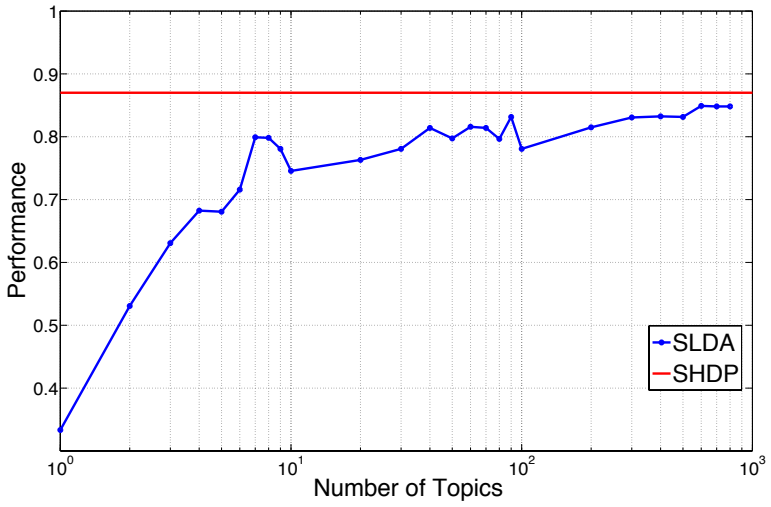
(b) Scene Classification performance

Data: 4 classes from MIT LabelMe Natural Scene
80% for training; 20% for testing.

Experiment– Action Classification

	boxing	clapping	waving
boxing	0.85	0.1	0.05
clapping	0	1	0
waving	0.1	0.15	0.75

(c) Confusion Matrix 86.67%



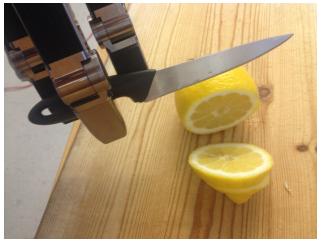
(d) Action Classification performance

Data: 3 classes from KTH Action Dataset
80% for training; 20% for testing.

Future Work

1. Even more efficient algorithm
2. Factorized Models [5]

3. Contexual Modelling [6]



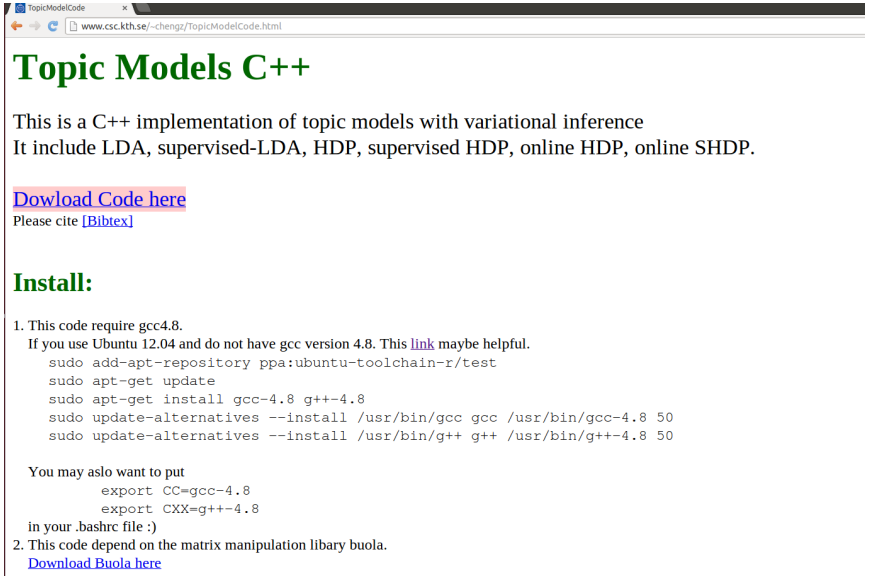
[5] C. Zhang, C. H. Ek, A. Damianou, and H. Kjellström, "Factorized topic models," in International Conference on Learning Representations, 2013.

[6] C. Zhang, D. Song, and H. Kjellström. Contextual modeling with Labeled Multi-LDA. In IROS, 2013.

Thanks & Questions

Code Available:

http://www.csc.kth.se/~chengz/TopicModelCode.html



Contact: chengz@kth.se

Welcome to our poster.

This research has been supported by the EU through TOMSY, IST-FP7- Collaborative Project-270436, and the Swedish Research Council (VR).