Using Both Latent and Supervised Shared Topics for Multitask Learning

Ayan Acharya, Aditya Rawal, Raymond J. Mooney, Eduardo R. Hruschka

UT Austin, Dept. of ECE

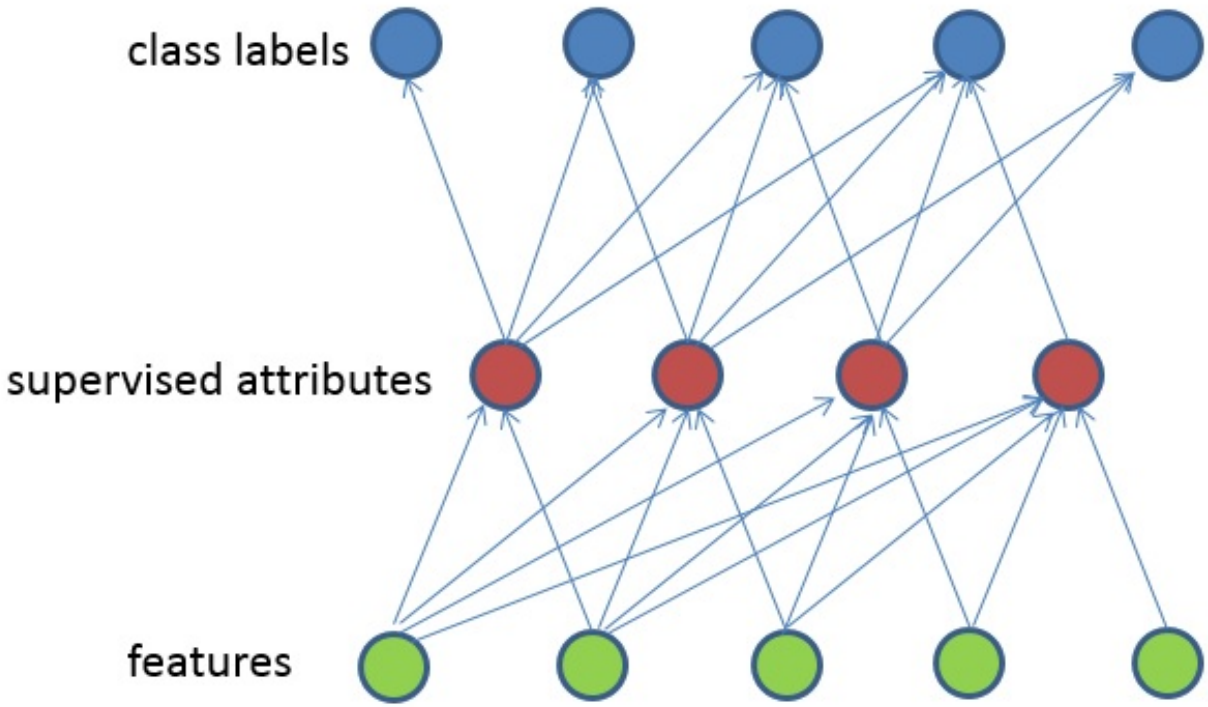
September 21, 2013

Problem Definition

- An MTL framework that can use both attributes and class labels
- In training corpus each document belongs to a different class and has a set of attributes ("supervised topics").
- Objective: Train a model using the words, supervised topics and class labels, and classify completely unlabeled test data (no supervised topic or class label)

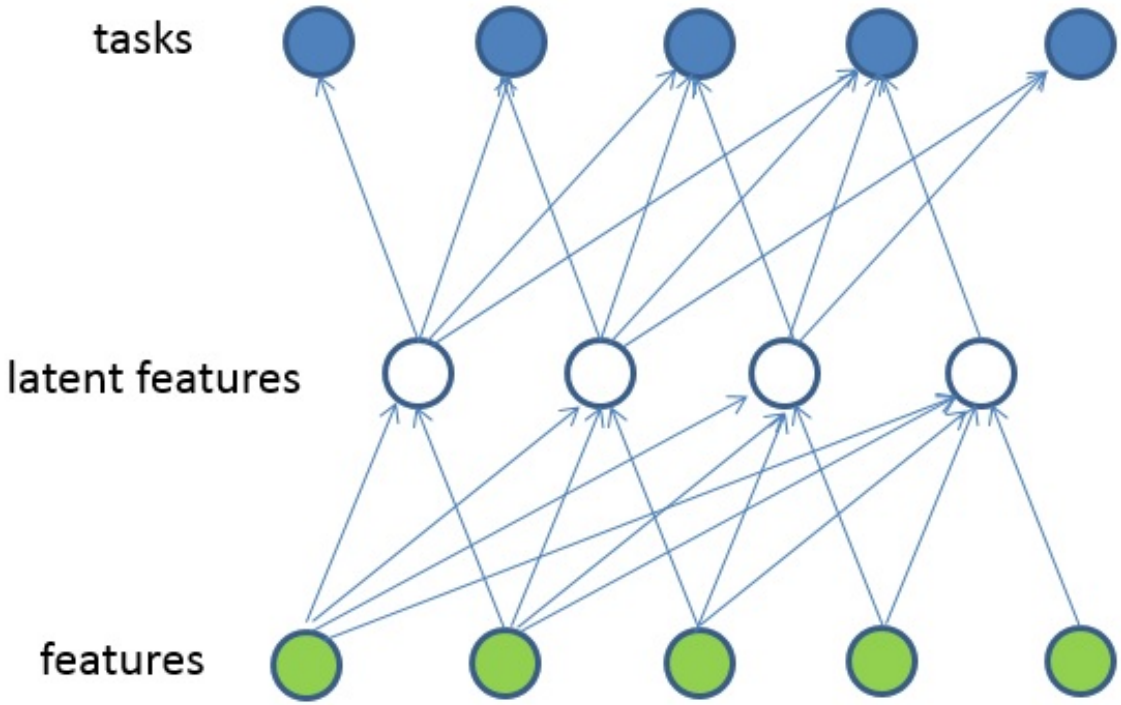
Attributes: "is 3d Boxy?", "has torso?", "has wheels?" etc.

Transfer with Supervised Shared Attributes



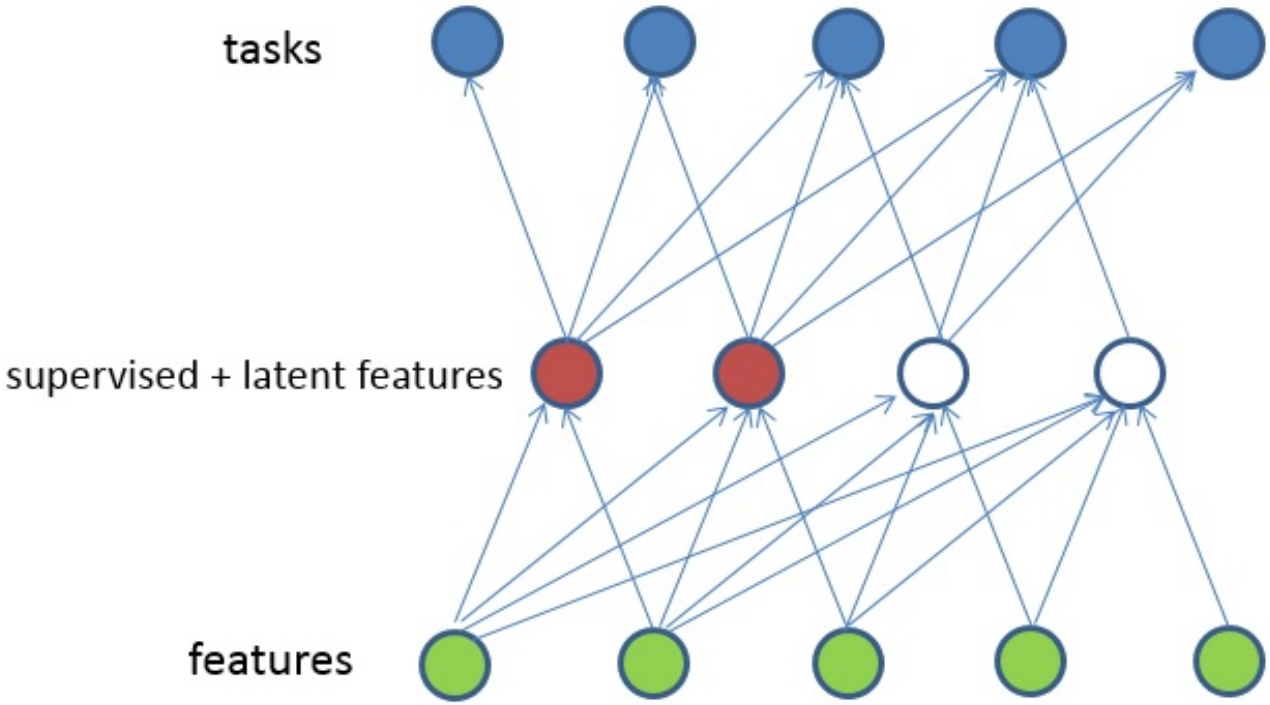
- Train to infer attributes from visual features
- Train to infer categories from attributes (Lampert *et al.*, CVPR 2009)

Multitask Learning with Shared Latent Attributes



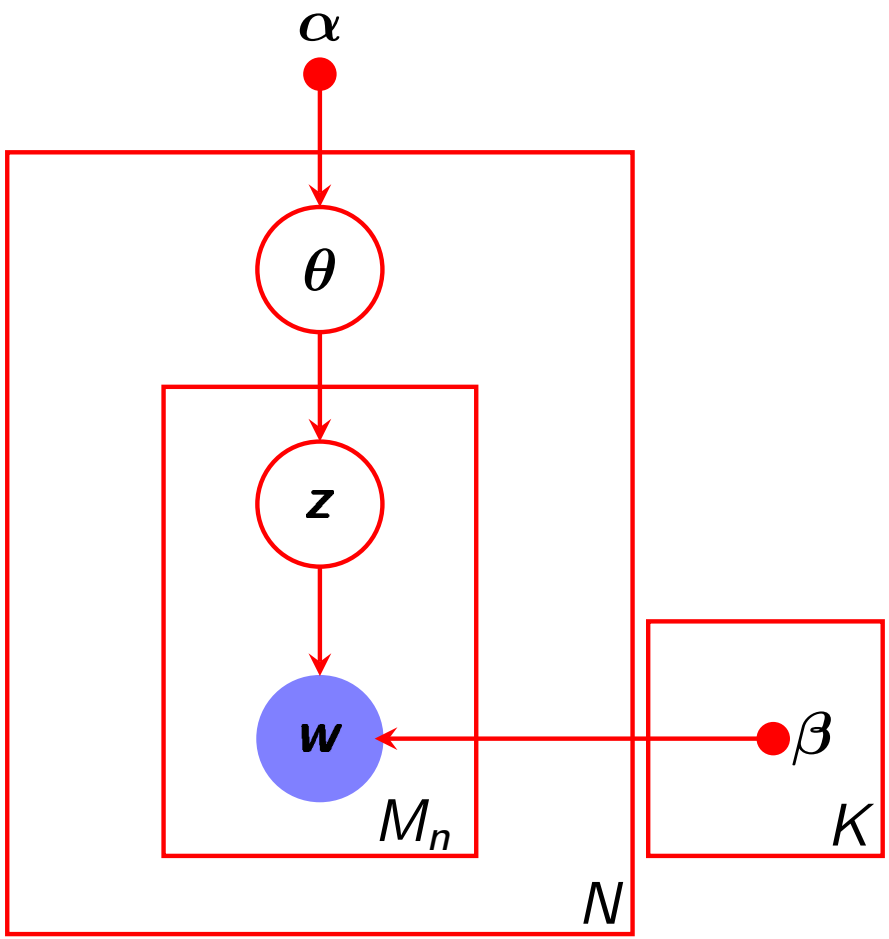
- work on multitask learning by R. Caruana (Machine Learning, 1997)

Transfer with Shared Latent and Supervised Attributes



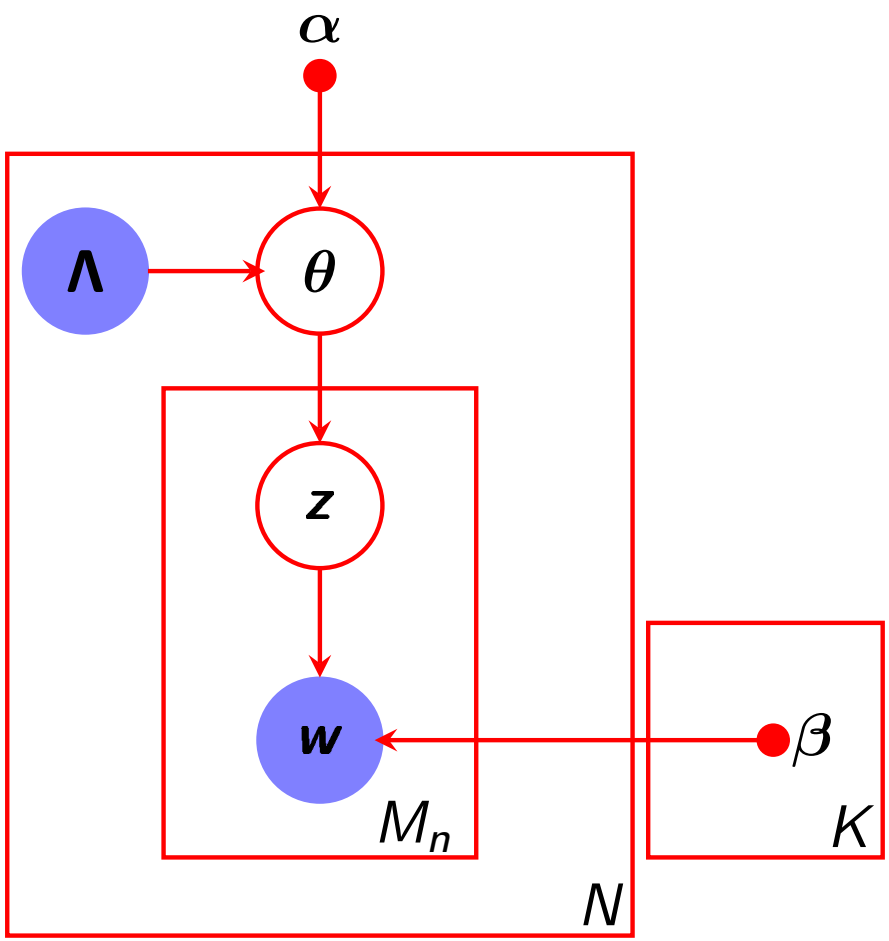
Latent Dirichlet Allocation (LDA)

Reference: Blei *et al.*, JMLR, 2003



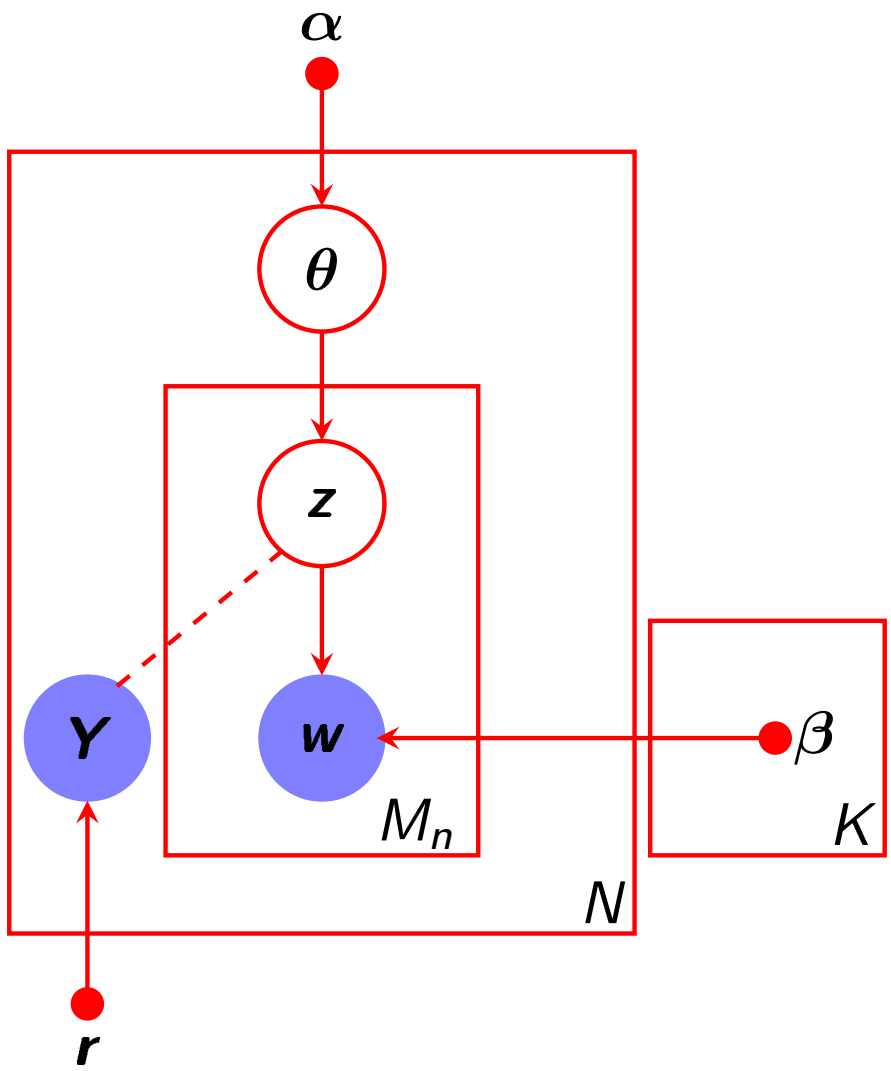
Labeled LDA (LLDA)

Reference: Ramage *et al.*, EMNLP, 2009

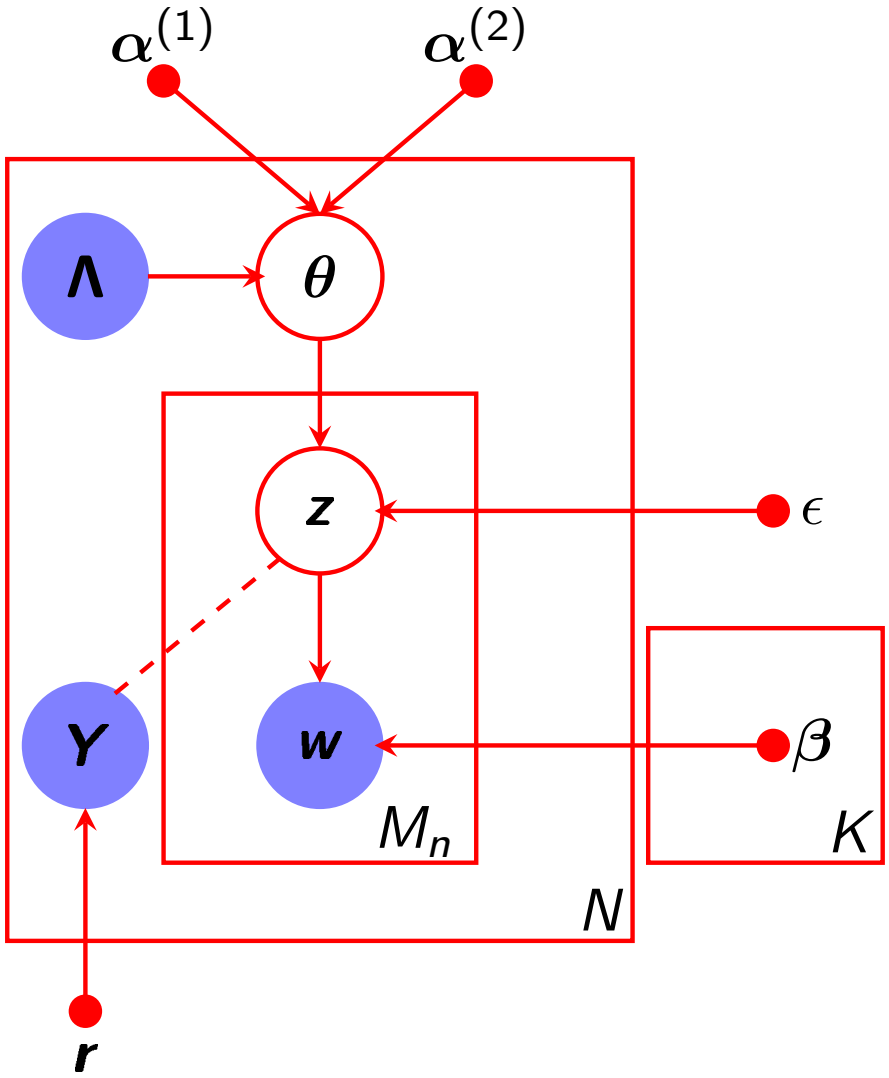


Maximum Entropy Discriminant LDA (MedLDA)

Reference: Zhu *et al.*, ICML, 2009



Doubly Supervised LDA (DSLDA)



Objective Function in DSLDA

$$\min_{q,\boldsymbol{\kappa}_0,\{\xi_n\}} \frac{1}{2}||\boldsymbol{r}||^2 - \mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{\kappa}_0) + C\sum_{n=1}^{N}\xi_n,$$

$$\text{s.t. } \forall n, y \neq Y_n : \mathbb{E}[\boldsymbol{r}^T \Delta f_n(y)] \geq 1 - \xi_n; \xi_n \geq 0.$$

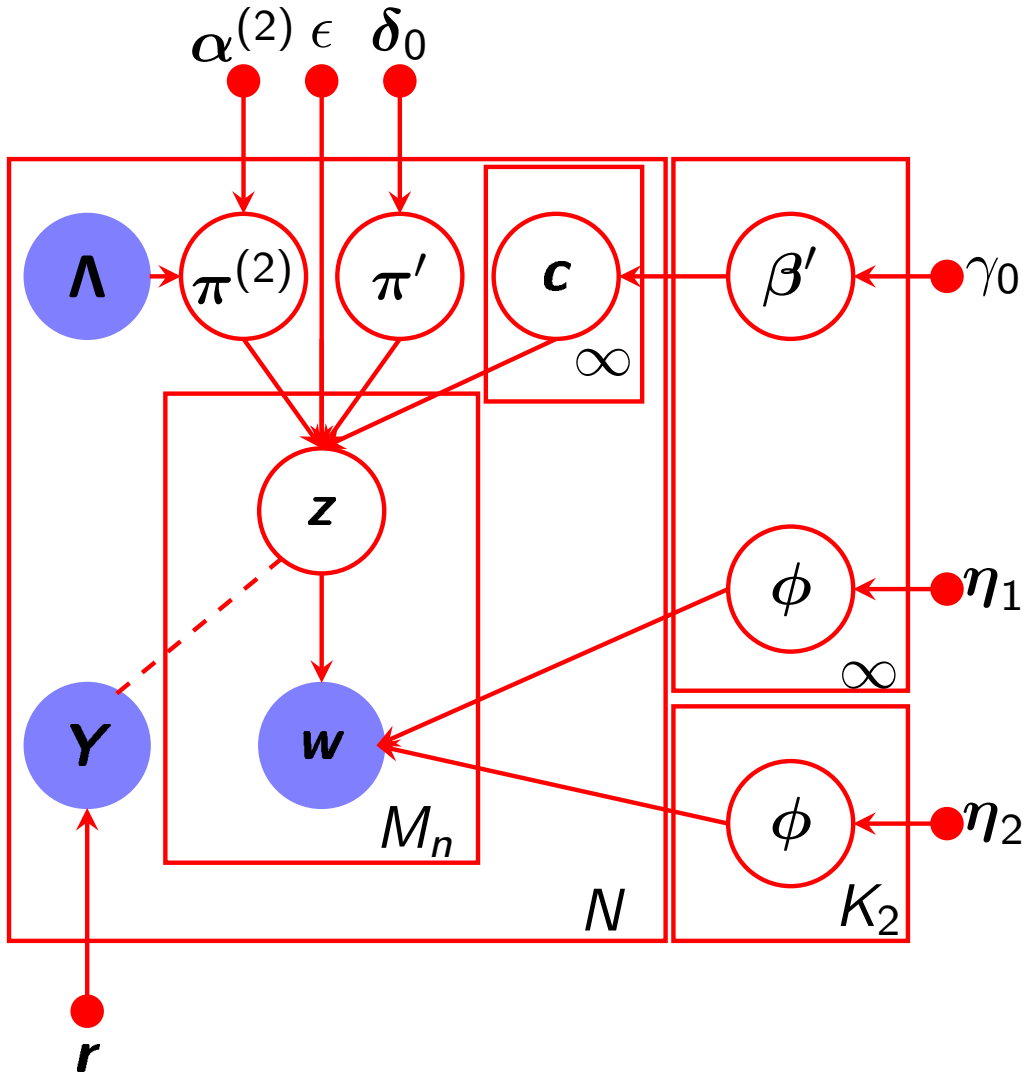
Objective Function in DSLDA

$$\min_{q,\boldsymbol{\kappa}_0,\{\xi_n\}} \frac{1}{2}||\boldsymbol{r}||^2 - \mathcal{L}(q(\boldsymbol{Z}),\boldsymbol{\kappa}_0) + C\sum_{n=1}^{N}\xi_n,$$

$$\text{s.t. } \forall n, y \neq Y_n : \mathbb{E}[\boldsymbol{r}^T \Delta f_n(y)] \geq 1 - \xi_n; \xi_n \geq 0.$$

- $\boldsymbol{\kappa}_0$: set of model parameters
- $\Delta f_n(y) = f(Y_n, \bar{\boldsymbol{z}}_n) - f(y, \bar{\boldsymbol{z}}_n)$
- $f(y, \bar{\boldsymbol{z}}_n)$: zero padded feature vector
- $\mathcal{L}(q(\boldsymbol{Z}))$: lower bound from variational approximation $q(\boldsymbol{Z})$

Non-parametric Doubly Supervised LDA (NPDSLDA)



Baseline Models

1. MedLDA with **o**ne-**v**s-**a**ll classification (MedLDA-OVA)
2. MedLDA with **m**ulti**t**ask **l**earning (MedLDA-MTL)
3. DSLDA with **o**nly **s**hared **s**upervised **t**opics (DSLDA-OSST)
4. DSLDA with **n**o **s**hared **l**atent **t**opics (DSLDA-NSLT)
5. **M**ajority **c**lass **m**ethod (MCM)

Model	Supervised Topics	Latent Topics
MedLDA-OVA	absent	not shared
MedLDA-MTL	absent	shared
DSLDA-OSST	present	absent
DSLDA-NSLT	present	not shared
MCM	absent	absent

Description of Dataset: aYahoo

- **Classes:** carriage, centaur, bag, building, donkey, goat, jetski, monkey, mug, statue, wolf, and zebra
- **Supervised topics:** "has head", "has wheel", "has torso" and 61 others

Description of Dataset: ACM Conference

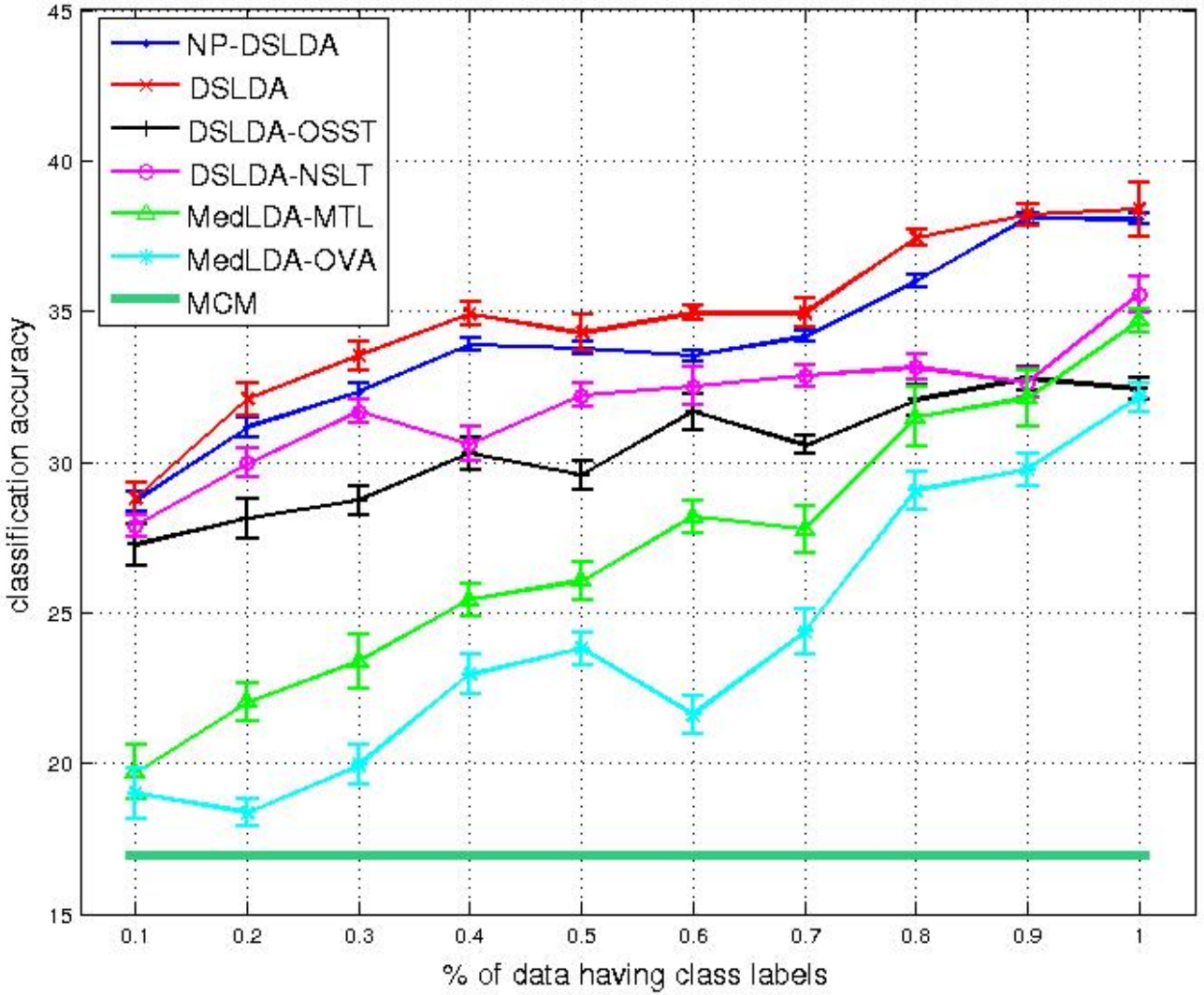
- **Classes:** First group – WWW, SIGIR, KDD, ICML; Second group – ISPD, DAC; abstracts of papers are treated as documents
- **Supervised topics:** keywords provided by the authors

Experimental Methodology

- Multitask training that evaluates benefits of sharing information between classes on the predictive accuracy of all classes
- Varied both fraction of training data that contains supervised topic labels and the fraction that contains supervised class labels

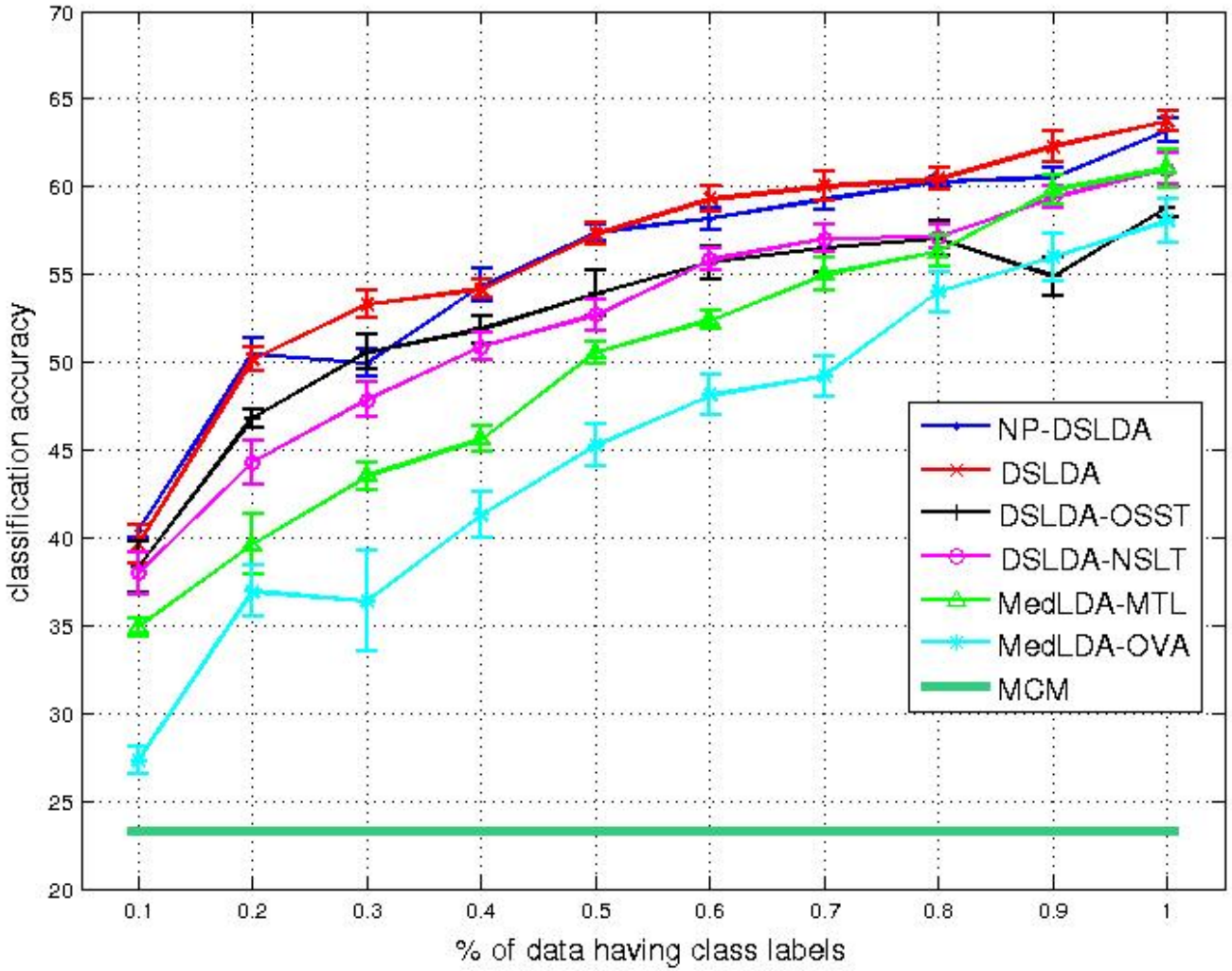
Results from aYahoo Data

- 50% training with supervised topic labels



Results from Text Data

- 50% training with supervised topic labels



Future Work

- Active learning for efficient query over both supervised topics and class labels
- Online training to update the model parameters
- The general idea of "double supervision" could be applied to many other models, for example, in multi-layer perceptrons, latent SVMs or in deep belief networks.

Questions?

References:

1. Multitask Learning, R. Caruana, Machine Learning, 1997. [Link].
2. Learning to detect unseen object classes by between class attribute transfer, CVPR 2009, Lampert *et al.* [Link].
3. Actively Selecting Annotations Among Objects and Attributes, ICCV 2011, Kovashka *et al.* [Link].
4. MedLDA: Maximum Margin Supervised Topic Models for Regression and Classification, ICML 2009, Zhu *et al.* [Link].
5. Online Variational Inference for the Hierarchical Dirichlet Process, AISTATS 2011, Wang *et al.* [Link].