Dirichlet Process with Mixed Random Measure

A Nonparametric Topic Model for Labeled Data

Dongwoo Kim
Suin Kim
Alice Oh

users and information
http://uilab.kaist.ac.kr

Motivation

- L-LDA (Labeled LDA) improves topic interpretation by defining single topic over each observed label
- But document is only generated by the set of topics of observed labels
- How to keep interpretability and relax the restriction of L-LDA?

Contribution

- Propose DPMRM which defines a random measure for each label
- Resulting model infers an unbounded number of topics for each label
- Compare the performance on single/multi labeled dataset with MedLDA, LDA-SVM, and L-LDA
- Modeling multi-labeled images for image segmentation and object labeling

DPMRM

Model Description

- DPMRM is a hierarchical construction of Dirichlet processes(DP)
- First level DP

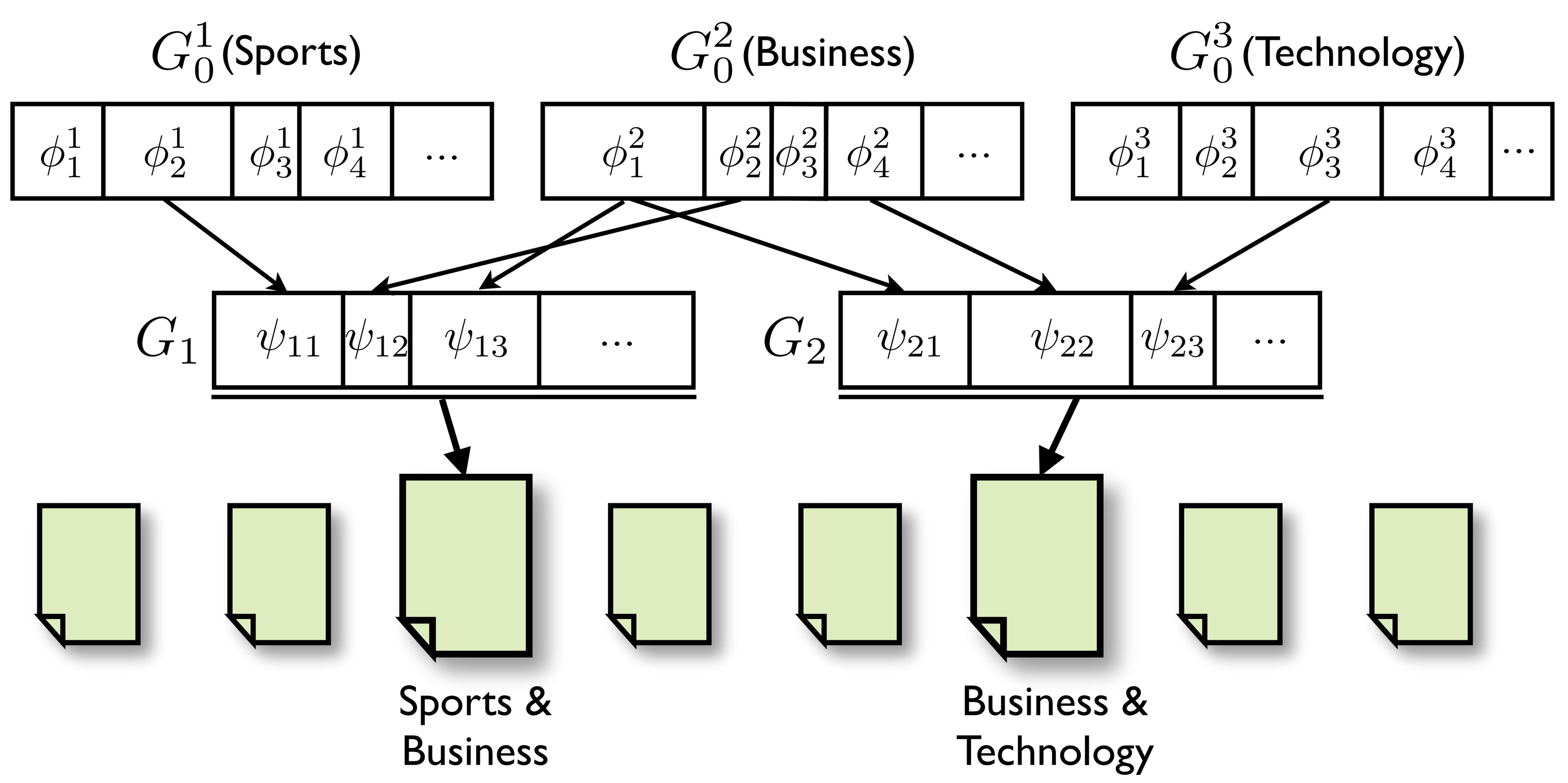
$$G_0^k|\gamma_k, H \sim \mathrm{DP}(\gamma_k, H) \qquad G_0^k = \sum_{l=0}^{\infty} \pi_l^k \delta_{\phi_l^k}$$

- Second level DP

$$G_j|label(\cdot), \alpha, \lambda_j \sim \mathrm{DP}(\alpha, \sum_{k \in label(j)} \lambda_{jk} G_0^k) \qquad G_0^j = \sum_{t=0}^{\infty} \pi_{jt} \delta_{\psi_{jt}}$$

second level DP is a mixture of first level random measures!
HDP can be viewed as a specialized instance of our model where K=1

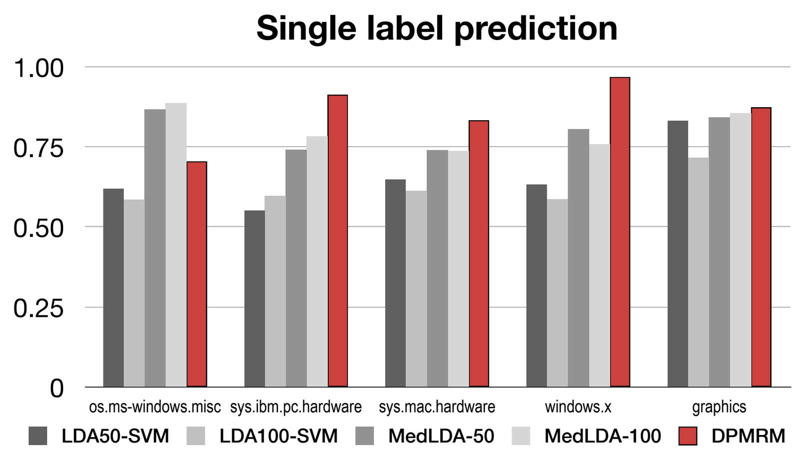
Graphical Illustration of DPMRM



with Document Corpus

Single Label Prediction

- Predict a label of document with trained models
- Compare the performance with LDA-SVM and MedLDA
- Dataset : sub-category (computer) of 20ng corpus



Accuracies of DPMRM MedLDA, and LDA-SVM on classification of 20NG. DPMRM outperforms LDA-SVM for all five labels and MedLDA for four labels

Multi Label Prediction

- Predict labels of document with trained model
- Compare the performance with L-LDA
- Dataset: RCV(news corpus), Ohumed(medical corpus)

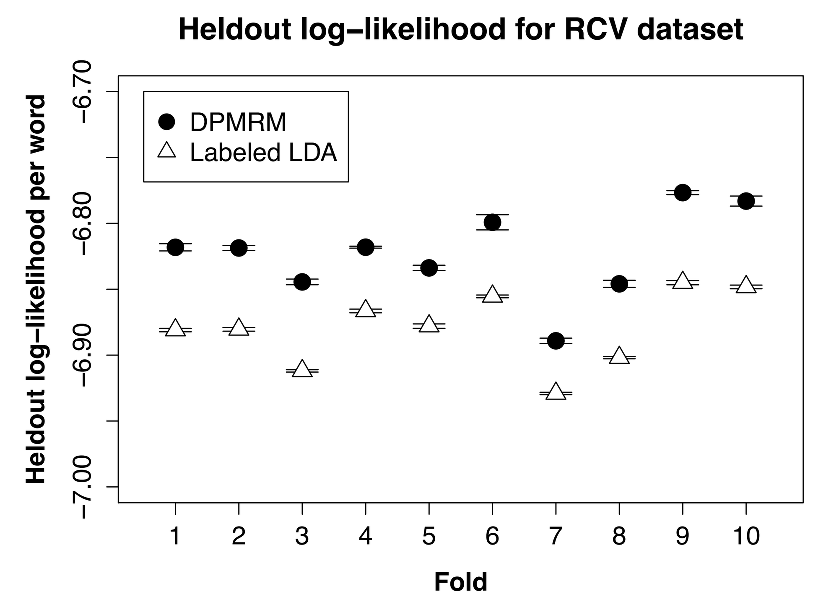
RCV	DPMRM	L-LDA
MicroF1	0.520	0.473
MacroF1	0.266	0.331

Ohsumed	DPMRM	L-LDA
MicroF1	0.392	0.382
MacroF1	0.223	0.263

- Result : DPMRM consistently performs better than L-LDA for microF1, but not for macroF1

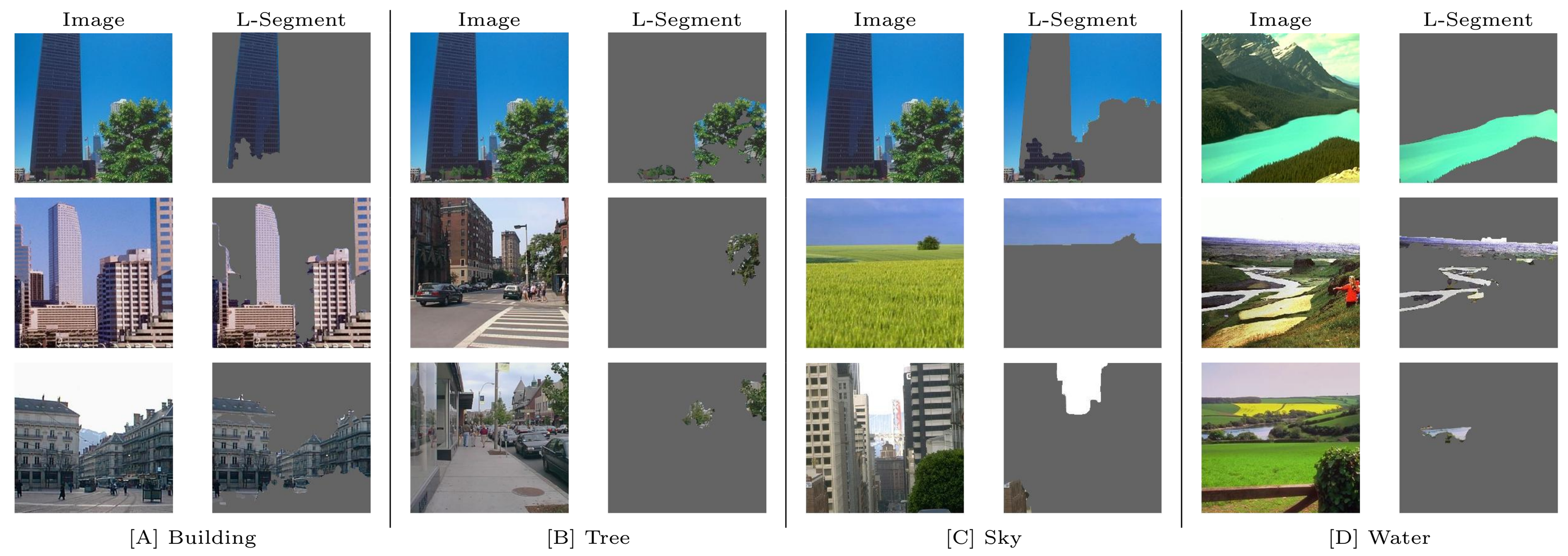
Predictive Performance

- Compare the heldout-likelihood of DPMRM with L-LDA.
- DPMRM consistently outperform L-LDA over ten fold datasets

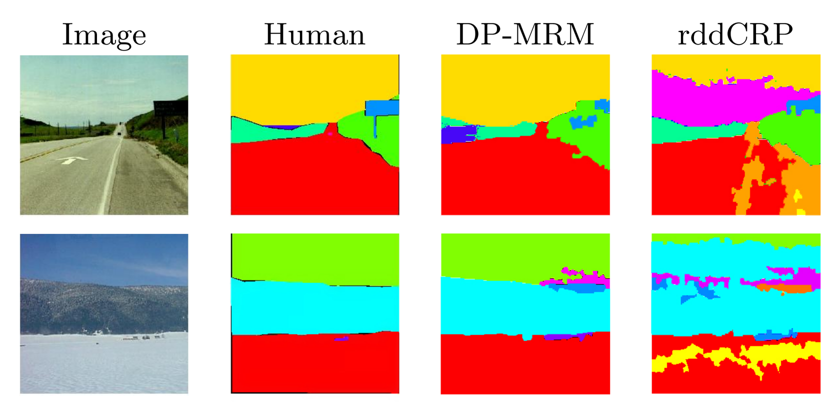


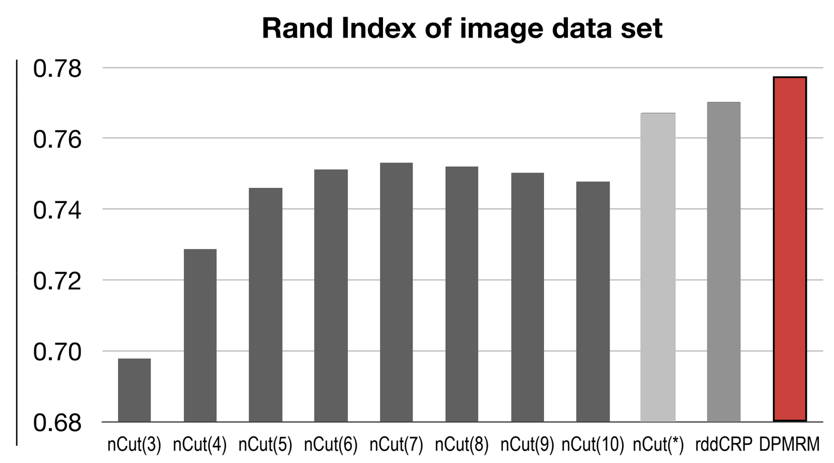
Infant				Corporate/Industrial						
L-LDA	DP-MRM			L-LDA	DP-MRM					
children	children	colon	tumor	compan	million	oil	shar	ton	airlin
infect	infect	aeruginosa	patient	million	profit	pow	compan	million	air
month	infant	express	leukemia	percent	percent	ga	bank	percent	carg
patient	month	gene	cell	market	half	compan	percent	produc	flight
ag	ag	type	chemotherapi	produc	expect	produc	million	export	servic
infant	antibodi	dna	dose	stat	compan	plant	invest	crop	airport
studi	hiv	mutat	therapi	bank	billion	operat	stock	wheat	carri
vaccin	vaccin	ha-ra	receiv	invest	result	refin	market	grain	plan
viru	viru	excret	treatment	plan	market	unit	stat	juli	operat
antibodi	test	urinari	remiss	billion	shar	million	plan	sugar	aircraft

with Labeled Images



<Labeled Segments>





<Image segmentation & Performance (Rand Index)>

DPMRM + ddCRP

- Incorporate ddCRP(distance-dependent Chinese restaurant process) into DPMRM
- To account the spatial dependencies within an image
- Segment images into corresponding labels without pixel-level supervision

Updated Polya Urn Scheme

$$\theta_{ji}|\theta_{j1}, ..., \theta_{ji-1}, \alpha, \eta, G_0^1, ..., G_0^K$$

$$\sim \sum_{i'}^{i-1} \frac{f(d_{ii'})}{f_{sum}^i + \alpha} \delta_{\theta_{ji'}} + \frac{\alpha}{f_{sum}^i + \alpha} \sum_k \frac{m_{jk\cdot} + r_{jk}\eta}{m_{j\cdot\cdot} + |\mathbf{r}_j|\eta} G_0^k$$

Consider distances between pixel i and i' <f:decay function>

Experiments

- Dataset: LabelMe human annotation data
- Decay function: window decay of size 1 (neighborhood pixels)
- Measure the segmentation performance with Rand index
- By using DPMRM+ddCRP we can simultaneously segment & label images without pixel-level supervision

Dirichlet Process with Mixed Random Measure

A Nonparametric Topic Model for Labeled Data

Dongwoo Kim
Suin Kim
Alice Oh

users and information
http://uilab.kaist.ac.kr

Motivation

- L-LDA (Labeled LDA) improves topic interpretation by defining single topic over each observed label
- Document is only generated by the set of topics of observed labels
- How to keep interpretability and relax the restriction of L-LDA?

Contribution

- Propose DPMRM which defines a random measure for each label
- Resulting model infers an unbounded number of topics for each label
- Compare the performance on single/multi labeled dataset with MedLDA, LDA-SVM, and L-LDA
- Modeling multi-labeled images for image segmentation and object labeling

DPMRM

Model Description

- DPMRM is a hierarchical construction of Dirichlet processes(DP)
- First level DP

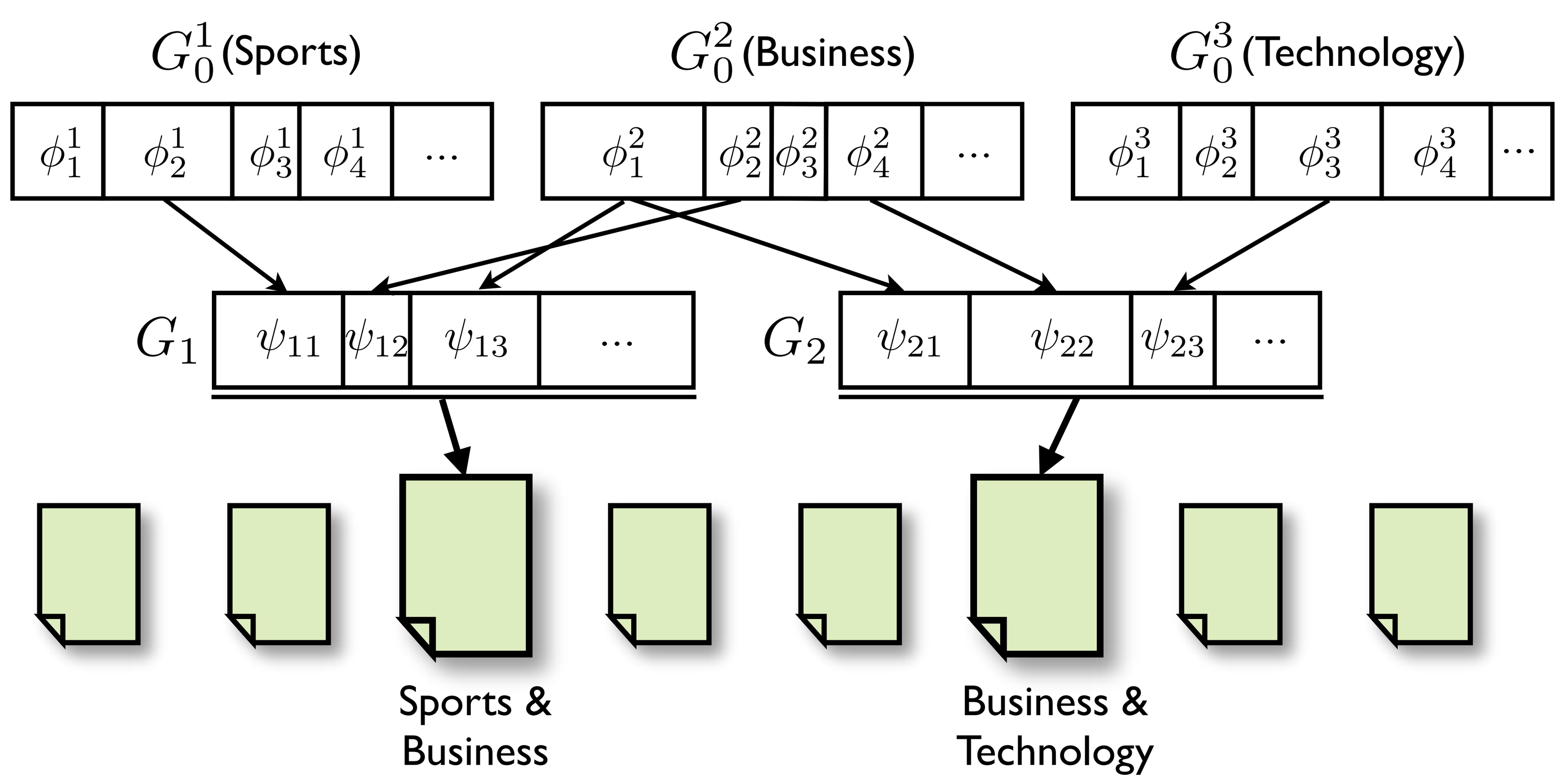
$$G_0^k | \gamma_k, H \sim \mathrm{DP}(\gamma_k, H) \qquad G_0^k = \sum_{l=0}^{\infty} \pi_l^k \delta_{\phi_l^k}$$

- Second level DP

$$G_j | label(\cdot), \alpha, \lambda_j \sim \mathrm{DP}(\alpha, \sum_{k \in label(j)} \lambda_{jk} G_0^k) \qquad G_0^j = \sum_{t=0}^{\infty} \pi_{jt} \delta_{\psi_{jt}}$$

second level DP is a mixture of first level random measures!
HDP can be viewed as a specialized instance of our model where K=1

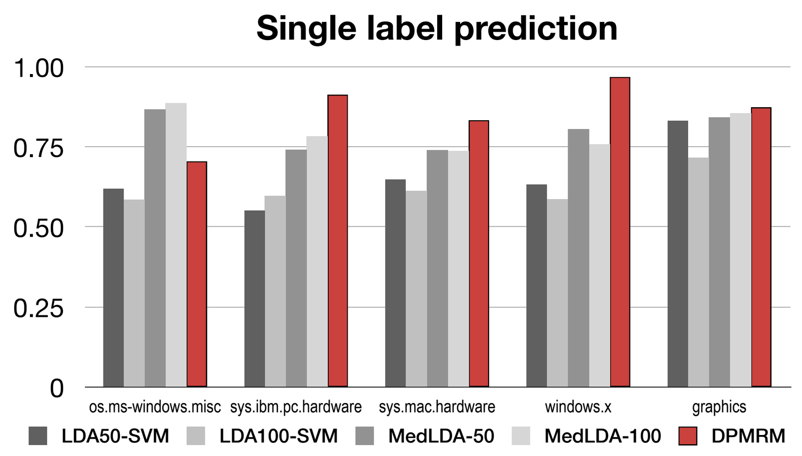
Graphical Illustration of DPMRM



with Document Corpus

Single Label Prediction

- Predict a label of document with trained models
- Compare the performance with LDA-SVM and MedLDA
- Dataset : sub-category (computer) of 20ng corpus



Accuracies of DPMRM MedLDA, and LDA-SVM on classification of 20NG. DPMRM outperforms LDA-SVM for all five labels and MedLDA for four labels

Multi Label Prediction

- Predict labels of document with trained model
- Compare the performance with L-LDA
- Dataset: RCV(news corpus), Ohumed(medical corpus)

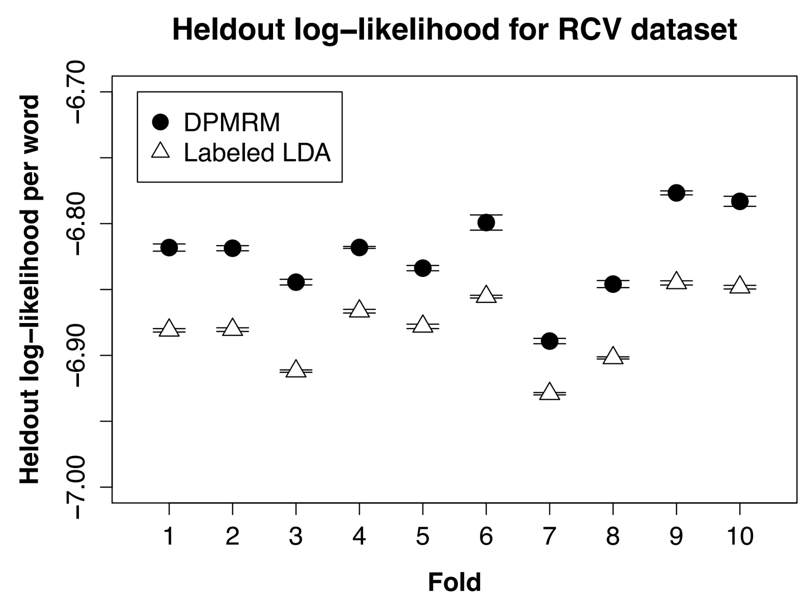
RCV	DPMRM	L-LDA
MicroF1	0.520	0.473
MacroF1	0.266	0.331

Ohsumed	DPMRM	L-LDA
MicroF1	0.392	0.382
MacroF1	0.223	0.263

- Result : DPMRM consistently performs better than L-LDA for microF1, but not for macroF1

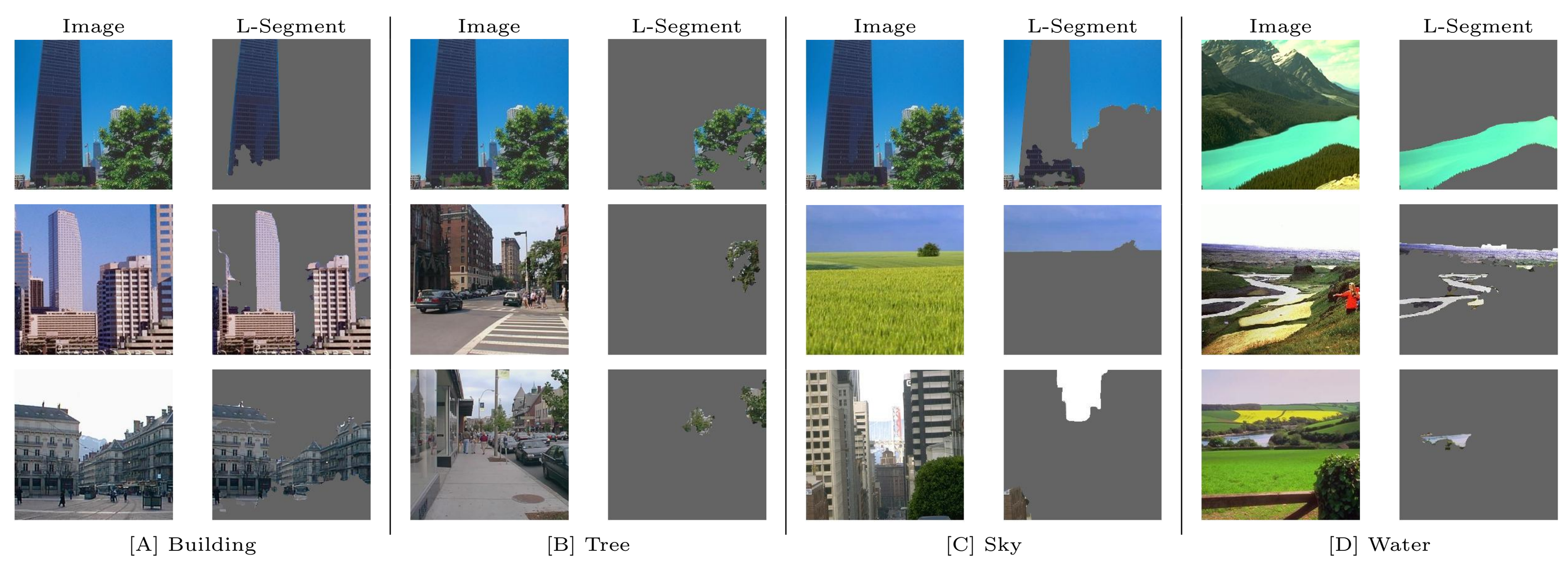
Predictive Performance

- Compare the heldout-likelihood of DPMRM with L-LDA.
- DPMRM consistently outperform L-LDA over ten fold datasets

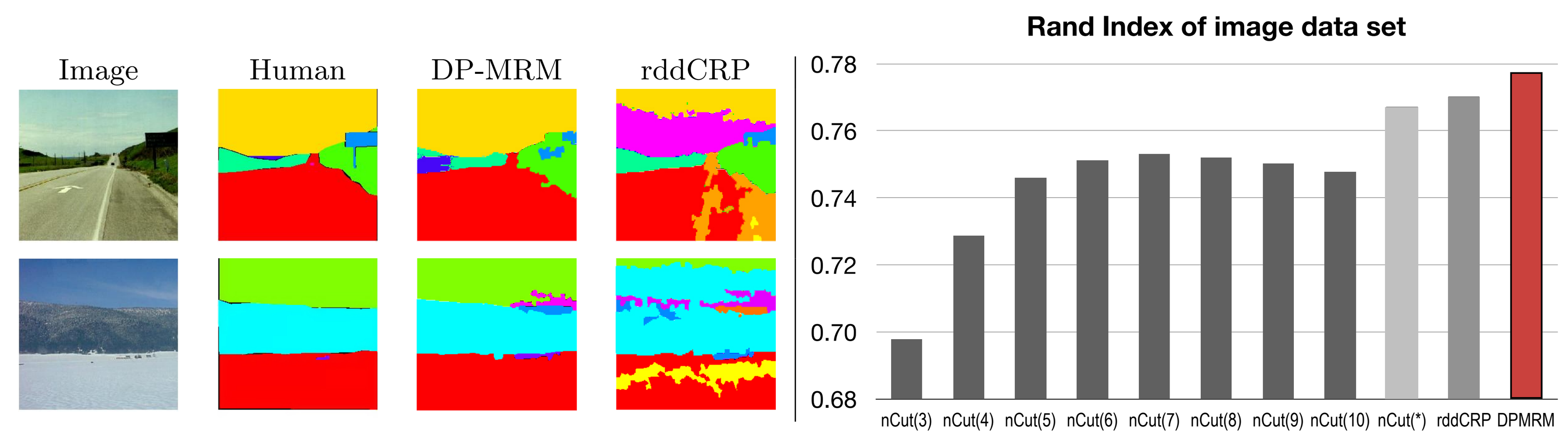


Infant				Corporate/Industrial					
L-LDA	DP-MRM			L-LDA	DP-MRM				
children	children	colon	tumor	compan	million	oil	shar	ton	airlin
infect	infect	aeruginosa	patient	million	profit	pow	compan	million	air
month	infant	express	leukemia	percent	percent	ga	bank	percent	carg
patient	month	gene	cell	market	half	compan	percent	produc	flight
ag	ag	type	chemotherapi	produc	expect	produc	million	export	servic
infant	antibodi	dna	dose	stat	compan	plant	invest	crop	airport
studi	hiv	mutat	therapi	bank	billion	operat	stock	wheat	carri
vaccin	vaccin	ha-ra	receiv	invest	result	refin	market	grain	plan
viru	viru	excret	treatment	plan	market	unit	stat	juli	operat
antibodi	test	urinari	remiss	billion	shar	million	plan	sugar	aircraft

with Labeled Images



<Labeled Segments>



<Image segmentation & Performance (Rand Index)>

DPMRM + ddCRP

- Incorporate ddCRP(distance-dependent Chinese restaurant process) into DPMRM
- To account the spatial dependencies within an image
- Segment images into corresponding labels without pixel-level supervision

Updated Polya Urn Scheme

$$\theta_{ji} | \theta_{j1}, ..., \theta_{ji-1}, \alpha, \eta, G_0^1, ..., G_0^K$$

$$\sim \sum_{i'}^{i-1} \frac{f(d_{ii'})}{f_{sum}^i + \alpha} \delta_{\theta_{ji'}} + \frac{\alpha}{f_{sum}^i + \alpha} \sum_k \frac{m_{jk\cdot} + r_{jk}\eta}{m_{j\cdot\cdot} + |\mathbf{r}_j|\eta} G_0^k$$

Consider distances between pixel i and i' <f:decay function>

Experiments

- Dataset: LabelMe human annotation data
- Decay function: window decay of size 1 (neighborhood pixels)
- Measure the segmentation performance with Rand index
- By using DPMRM+ddCRP, we can simultaneously segment & label images without pixel-level supervision