Non-parametric Estimation of Probabilistic Topic Hierarchies

Elias Zavitsanos†, **Georgios Paliouras**†, George A. Vouros‡
izavits@iit.demokritos.gr, paliourg@iit.demokritos.gr, georgev@aegean.gr

†Institute of Informatics and Telecommunications, NCSR "Demokritos"
‡Dpt. of Information and Communication Systems Engineering, Univ. of Aegean, Samos

Learning Topic Hierarchies

- Document indexing and classification
- Document modeling
- Reflect relations between concepts or topics
- Crucial step for ontology learning

Hierarchical Clustering

- Hard clustering techniques and decision trees are employed
- A document is assigned to a single topic
- Hinders the efficient retrieval of documents

Our Aim is..

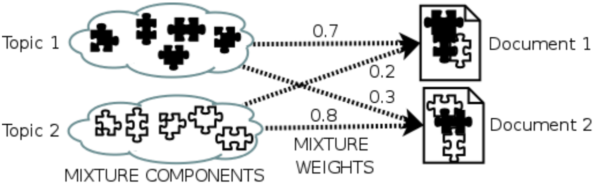
.. to learn topic hierarchies where:

- nodes reflect the shared terminology between documents
- nodes reflect the intended meaning of documents
- nodes high in the hierarchy reflect abstract notions
- predict unseen documents
- have a non-parametric nature
- learning is language and domain independent

Outline

Probabilistic Topic Models

- Generative models for documents ([SG07])
- Based on the “bag-of-words” theorem ([Fin31])
- Differ in their generative process



Categories

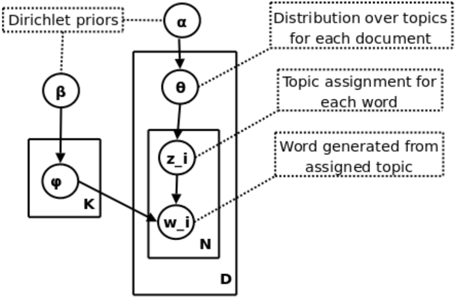
Flat modeling:

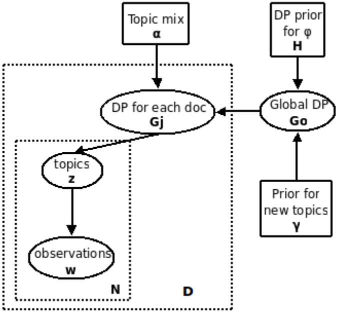
- Probabilistic Latent Semantic Analysis (PLSA) ([Hof99])
- Latent Dirichlet Allocation (LDA) ([BNJ03])
- Hierarchical Dirichlet Processes (HDP) ([TJBB06])

Hierarchical modeling:

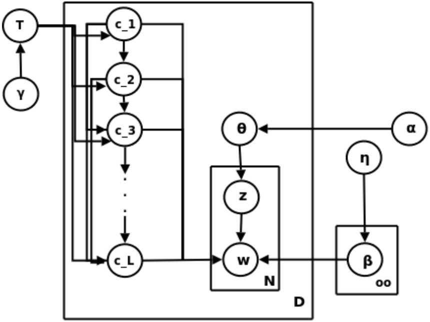
- Hierarchical Probabilistic Latent Semantic Analysis (HPLSA) ([GGPC02])
- Hierarchical Latent Dirichlet Allocation (hLDA) ([BGJT04])
- PAM - HPAM - NPPAM ([LM06], [MLM07], [LBM07])

Flat Generative Process





Hierarchical Generative Process



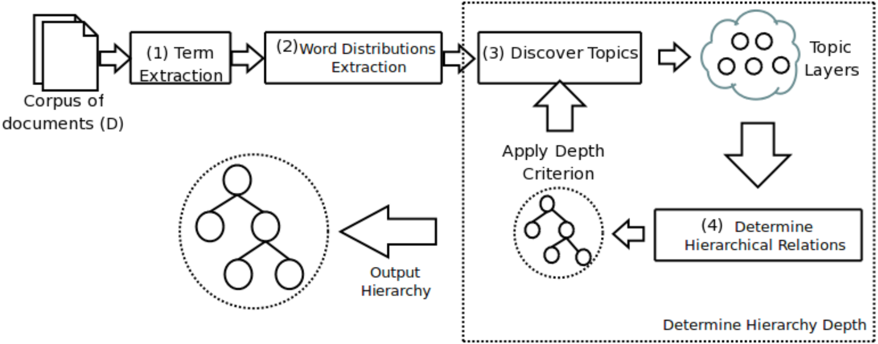
Model Estimation

- Observations: words in documents
- Task: learn the latent hierarchy

Usual a priori requirements:

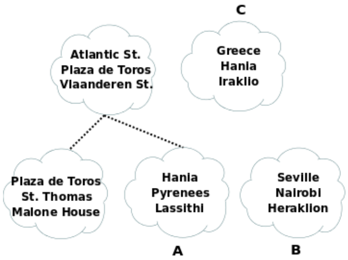
- Number of topics
- Number of hierarchy levels

Proposed Method



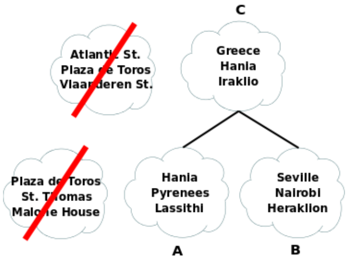
Hierarchy Construction

- Find a topic C of level L given which, topics A and B of level $L+1$ are conditionally independent
- $\mid \hat{P}(A \cap B \mid C) - \hat{P}(A \mid C)\hat{P}(B \mid C) \mid \leq th$
- Topic C is broader than A and B, and contains at least the mutual information of A and B



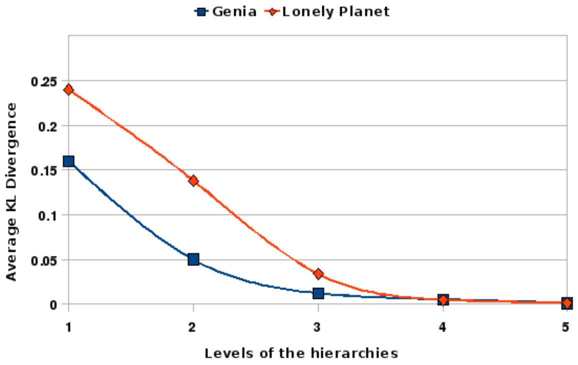
Hierarchy Construction

- Find a topic C of level L given which, topics A and B of level $L{+}1$ are conditionally independent
- $| \hat{P}(A \cap B \mid C) - \hat{P}(A \mid C)\hat{P}(B \mid C) | \leq th$
- Topic C is broader than A and B, and contains at least the mutual information of A and B

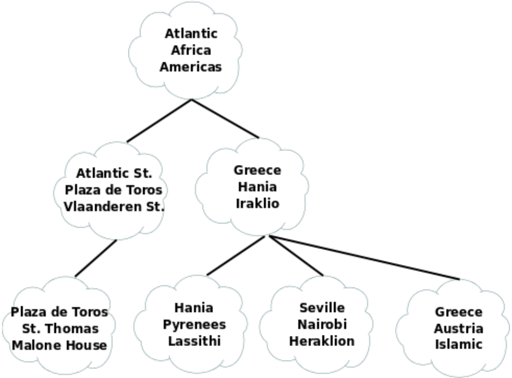


Hierarchy Depth

Iterate until the latent topics are as “specific” as possible



Example of a Learned Hierarchy



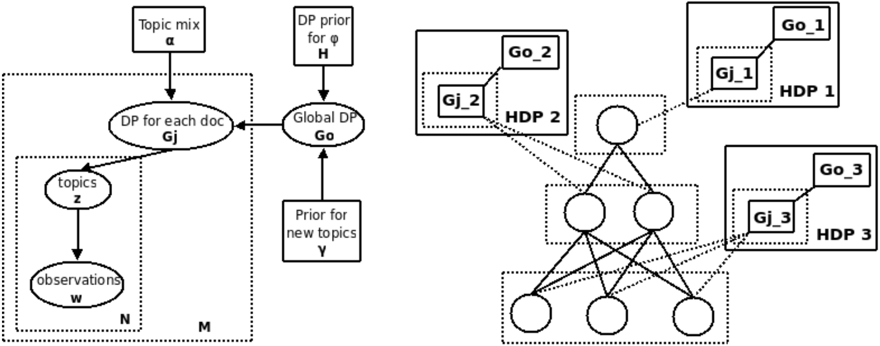
Summing Up

- Statistical method
- Language and domain independence
- Calculation of hierarchy depth and branching factor
- Naive definition of number of topics

Motivations

- Represent all topics as distributions over words
- Allow subtopics to be shared among supertopics
- Allow topics to be shared among documents
- Infer the size of the hierarchy automatically
- Predict unseen documents
- Represent probabilities over relations

The hHDP model



(a)

(b)

Generative Process

1. Choose $N \sim \text{Poisson}(\xi)$
2. For each of the N words:
3. For each level λ, $0 < \lambda < \Lambda$:
   1. Choose global probability measure $G_{0\lambda} \sim DP(\gamma, H)$
   2. Choose probability measure $G_{i\lambda} \sim DP(\alpha, G_{0\lambda})$
   3. Choose a topic $\theta_{\lambda j} \sim P(\cdot \mid G_{i\lambda}, \theta_{\lambda-1})$
4. For the level Λ:
   1. Choose global probability measure $G_0 \sim DP(\gamma, H)$
   2. Choose probability measure $G_\Lambda \sim DP(\alpha, G_0)$
   3. Choose a topic $\theta_{\Lambda j} \sim P(\cdot \mid G_\Lambda, \theta_{\lambda-1})$
   4. Choose a word $w_{\Lambda j} \sim P(\cdot \mid \theta_{\Lambda j})$

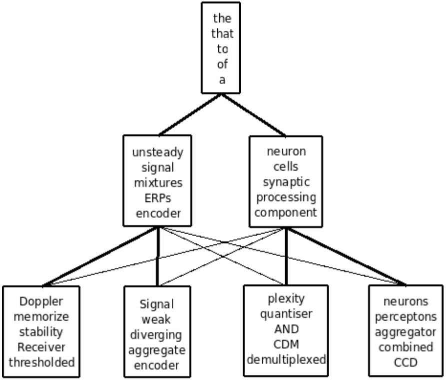
Model Estimation from Data

Data: Term - Document matrix of frequencies
Result: Estimated topic hierarchy
set M=number of documents
set V=vocabulary size
estimate leaf topics K
set $T = K$
while $\mid T \mid > 1$ **do**
    // **transform document space**
    set $M = K$
    set input=MxV matrix of frequencies
    estimate topics K of next level up
    set $T = K$
end

Model Estimation from Data (Coarse estimation)

Data: Term - Document matrix of frequencies
Result: Estimated coarse topic hierarchy
set M=number of documents
set V=vocabulary size
estimate leaf topics K
set $T = K$
while $\mid T \mid > 1$ **do**
    // **transform term space**
    set $V = K$
    set input=MxV matrix of frequencies
    estimate topics K of next level up
    set $T = K$
end

Example of Learned Hierarchy

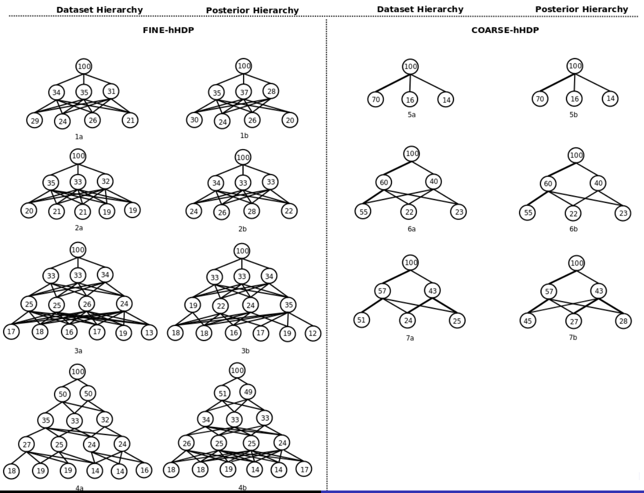


Experiments

Application to three tasks:

1. Analysis of artificial data
2. Ontology Learning
3. Document modeling

Analysis of Artificial Data



Numeric Results

Precision		Recall		Experiment
Topics	Edges	Topics	Edges	Case
1.0	1.0	1.0	0.93	1a-b
1.0	1.0	0.88	0.83	2a-b
1.0	1.0	1.0	0.71	3a-b
1.0	0.72	1.0	1.0	4a-b
1.0	1.0	1.0	1.0	5a-b
1.0	1.0	1.0	0.88	6a-b
1.0	0.88	1.0	1.0	7a-b

Ontology Learning

- Genia and Lonely Planet datasets
- Genia documents: #2000
- LonelyPlanet documents: #300
- Genia and Lonely Planet ontologies as *Gold Standard*
- Evaluation using the method of [ZPV08]

Numeric Results

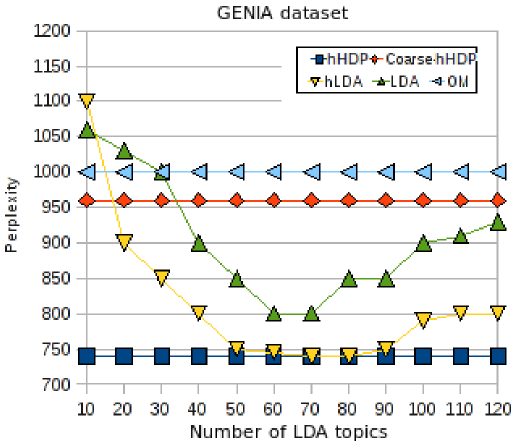
	Genia			LonelyPlanet		
Model	P	R	F	P	R	F
hHDP	0.65	0.60	0.624	0.22	0.15	0.17
hHDP-pruned	0.88	0.80	**0.838**	0.35	0.23	0.27
hLDA	0.62	0.55	0.58	0.07	0.01	0.017
OL LDA-based	0.89	0.70	0.78	0.42	0.31	**0.35**

Document Modeling

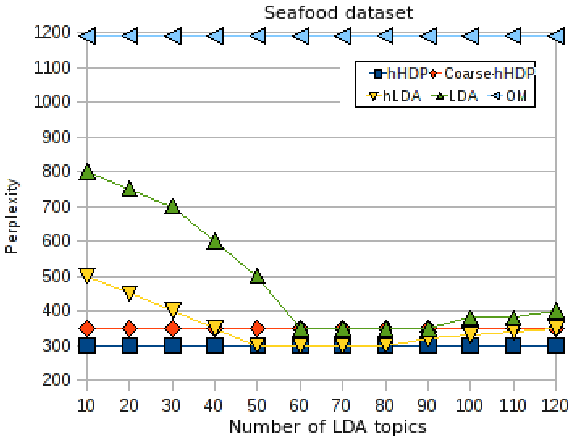
- Comparison with: LDA, hLDA, OM, MEM
- Evaluation with the measure of *Perplexity*

  $Perplexity(D) = exp\{-\sum_{i=1}^{N} \frac{1}{N} \log p(w_i)\}$
- Evaluation in five datasets: Genia, LP, Seafood, Elegance, NIPS
- Perform 10-fold cross validation and provide mean values

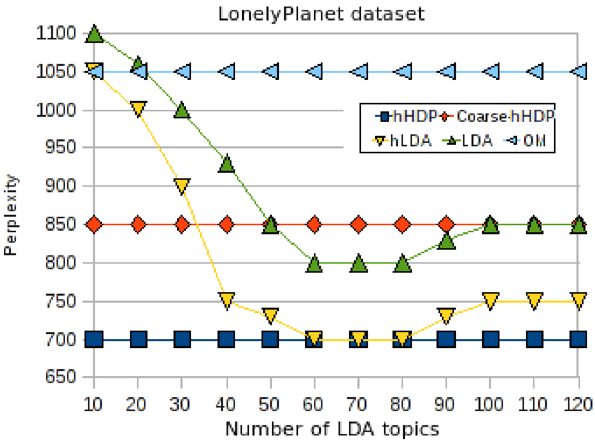
Mean Perplexity on Genia Dataset



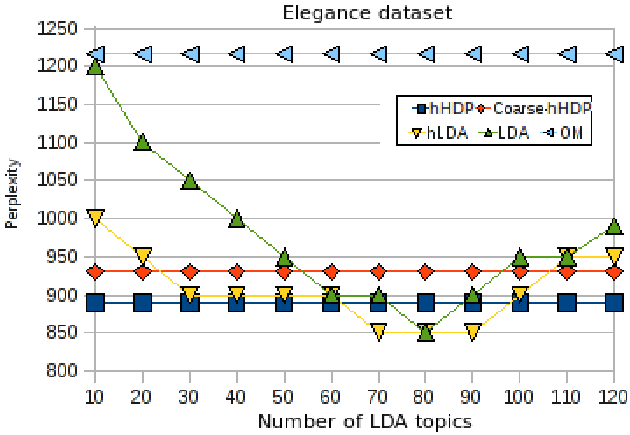
Mean Perplexity on Seafood Dataset



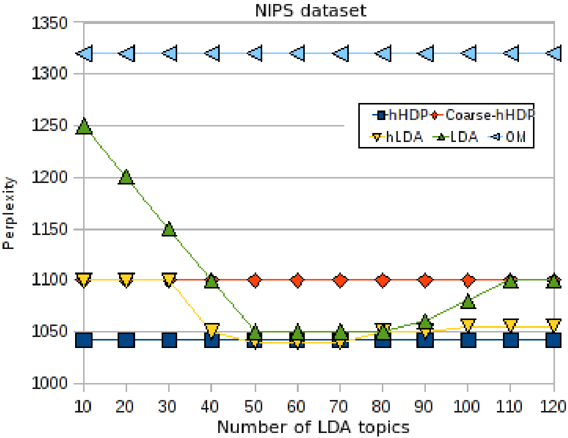
Mean Perplexity on Lonely Planet Dataset



Mean Perplexity on Elegance Dataset



Mean Perplexity on NIPS Dataset



Conclusions

- Statistical Methods
- Language and Domain independence
- No need for user parameters
- Infer the size of the hierarchy
- Represent all nodes as distributions over words
- Suitable for Ontology Learning and Document Modeling
- Promising results

Future Directions

- Study word burstiness in topic models
- Adaptive Gibbs sampler in the hHDP model
- Semantics of Hierarchical Probabilistic Topic Models
- Use different priors on HPTMs
- Evaluation in different types of dataset (e.g. images)
- Use the model for Folksonomy learning

Thank you!

References I

D.M. Blei, T.L. Griffiths, M.I. Jordan, and J.B. Tenenbaum.
Hierarchical topic models and the nested chinese restaurant process.
In *Advances in Neural Information Processing Systems 16*, 2004.

D.M. Blei, A.Y. Ng, and M.I. Jordan.
Latent dirichlet allocation.
Journal of Machine Learning Research, 3:993–1022, 2003.

B.D. Finetti.
Atti della R. Academia Nazionale dei Lincei, Serie 6. Memorie, Classe di Scienze Fisiche, Mathematice e Naturale, chapter Funzione Caratteristica di un Fenomeno Aleatorio, pages 251–299.
1931.

E. Gaussier, C. Goutte, K. Popat, and F. Chen.
A hierarchical model for clustering and categorising documents.
In *Advances in Information Retrieval - Proceedings of the 24th BCS-IRSG European Colloquium on IR Research*, pages 229–247. Springer, 2002.

T. Hofmann.
Probabilistic latent semantic indexing.
In *Proceedings of the Twenty-Second Annual International SIGIR Conference on Research and Development in Information Retrieval*, 1999.

References II

W. Li, D. Blei, and A. McCallum.
Nonparametric bayes pachinko allocation.
In *Uncertainty in Artificial Intelligence*, 2007.

W. Li and A. McCallum.
Pachinko allocation: Dag-structured mixture models of topic correlations.
In *Proceedings of the 23rd Internationcal Conference on Machine Learning*, pages 577–584, 2006.

D. Mimno, W. Li, and A. McCallum.
Mixtures of hierarchical topics with pachinko allocation.
In *Proceedings of the 24th International Conference on Machine Learning*, pages 633–640, 2007.

M. Steyvers and T. Griffiths.
Handbook of Latent Semantic Analysis, chapter Probabilistic Topic Models.
Hillsdale, NJ: Erlbaum, 2007.

Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei.
Hierarchical Dirichlet Processes.
Journal of the American Statistical Association, 2006.

E. Zavitsanos, G. Paliouras, and G.A. Vouros.
A distributional approach to evaluating ontology learning methods using a gold standard.
In *Proceedings of the ECAI 2008 Workshop on Ontology Learning and Population*, 2008.