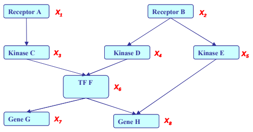
School of Computer Science
Carnegie Mellon

Applications in IR
— Probabilistic Topic Models

Probabilistic Graphical Models (10-708)

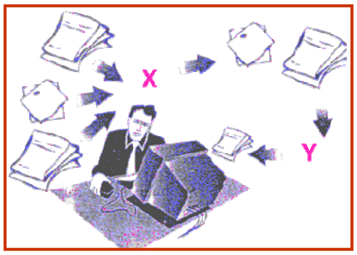
Lecture 22, Dec 3, 2007



Eric Xing

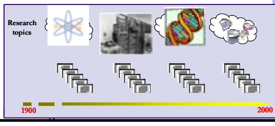
Reading:

NLP and Data Mining



We want:

- Semantic-based search
- infer topics and categorize documents
- Multimedia inference
- Automatic translation
- Predict how topics evolve
- …

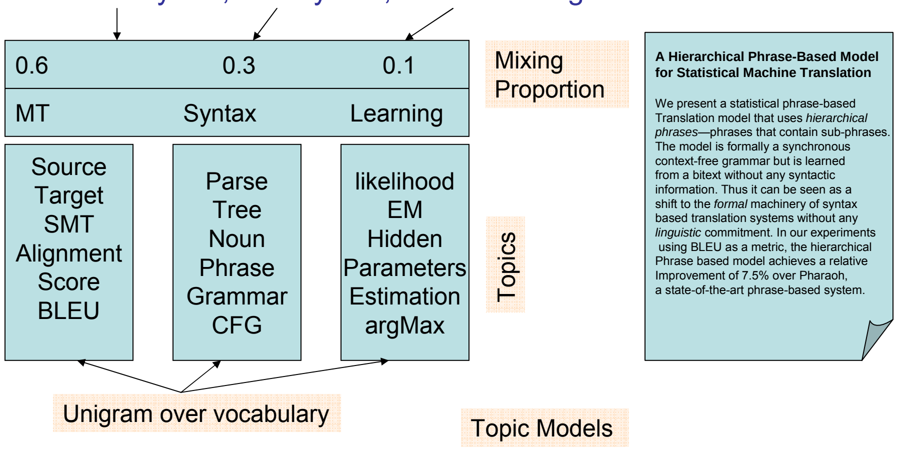


This Talk

- A graphical model primer
- Two families of probabilistic topics models and approximate inference
  - Bayesian admixture models
  - Random models
- Three applications
  - Topic evolution
  - Machine translation
  - Multimedia inference

How to Model Semantic?

- Q: What is it about?
- A: Mainly MT, with syntax, some learning



0.6
0.3
0.1
MT
Syntax
Learning
Mixing Proportion
Source Target SMT Alignment Score BLEU
Parse Tree Noun Phrase Grammar CFG
likelihood EM Hidden Parameters Estimation argMax
Topics
Unigram over vocabulary
Topic Models

A Hierarchical Phrase-Based Model for Statistical Machine Translation

We present a statistical phrase-based Translation model that uses *hierarchical phrases*—phrases that contain sub-phrases. The model is formally a synchronous context-free grammar but is learned from a bitext without any syntactic information. Thus it can be seen as a shift to the *formal* machinery of syntax based translation systems without any *linguistic* commitment. In our experiments using BLEU as a metric, the hierarchical Phrase based model achieves a relative Improvement of 7.5% over Pharaoh, a state-of-the-art phrase-based system.

Why this is Useful?

- Q: What is it about?
- A: Mainly MT, with syntax, some learning

0.6	0.3	0.1
MT	Syntax	Learning

Mixing Proportion

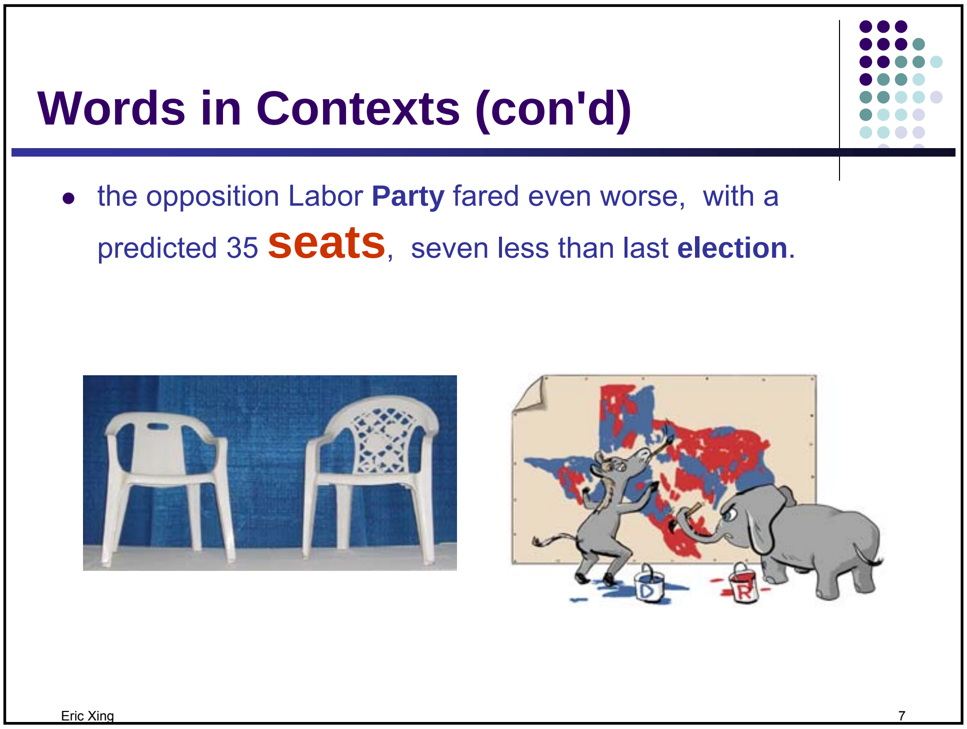
A Hierarchical Phrase-Based Model for Statistical Machine Translation

We present a statistical phrase-based Translation model that uses *hierarchical phrases*—phrases that contain sub-phrases. The model is formally a synchronous context-free grammar but is learned from a bitext without any syntactic information. Thus it can be seen as a shift to the *formal* machinery of syntax based translation systems without any *linguistic* commitment. In our experiments using BLEU as a metric, the hierarchical Phrase based model achieves a relative Improvement of 7.5% over Pharaoh, a state-of-the-art phrase-based system.

- Q: give me similar document?
  - Structured way of browsing the collection
- Other tasks
  - Dimensionality reduction
    - TF-IDF vs. topic mixing proportion
    - Classification, clustering, and more …

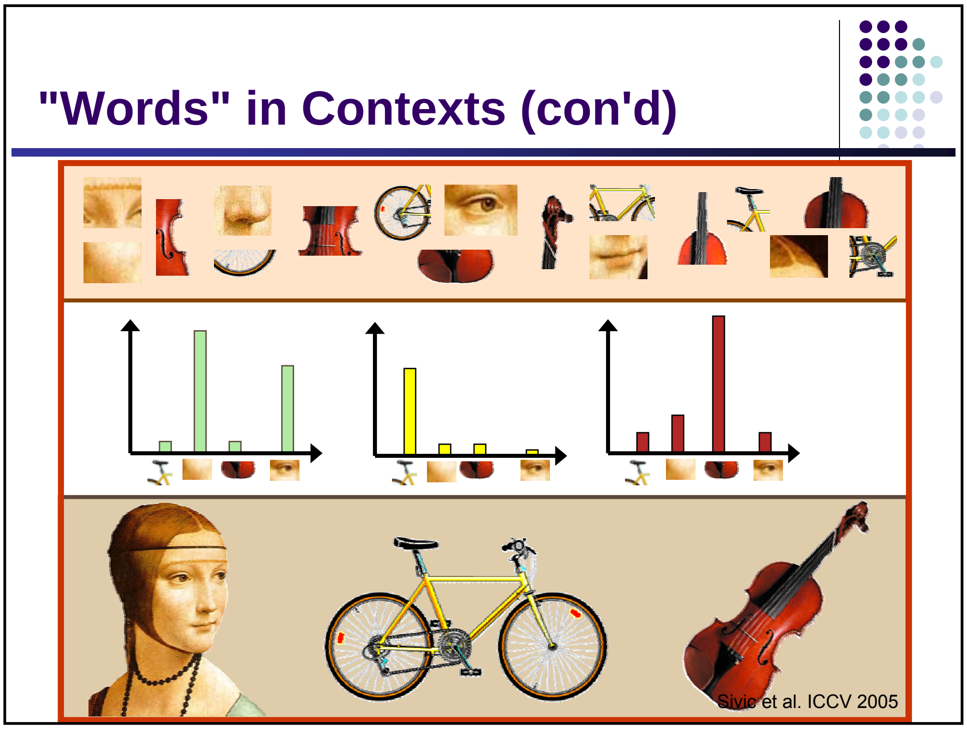
Words in Contexts

- “It was a nice **shot**. ”



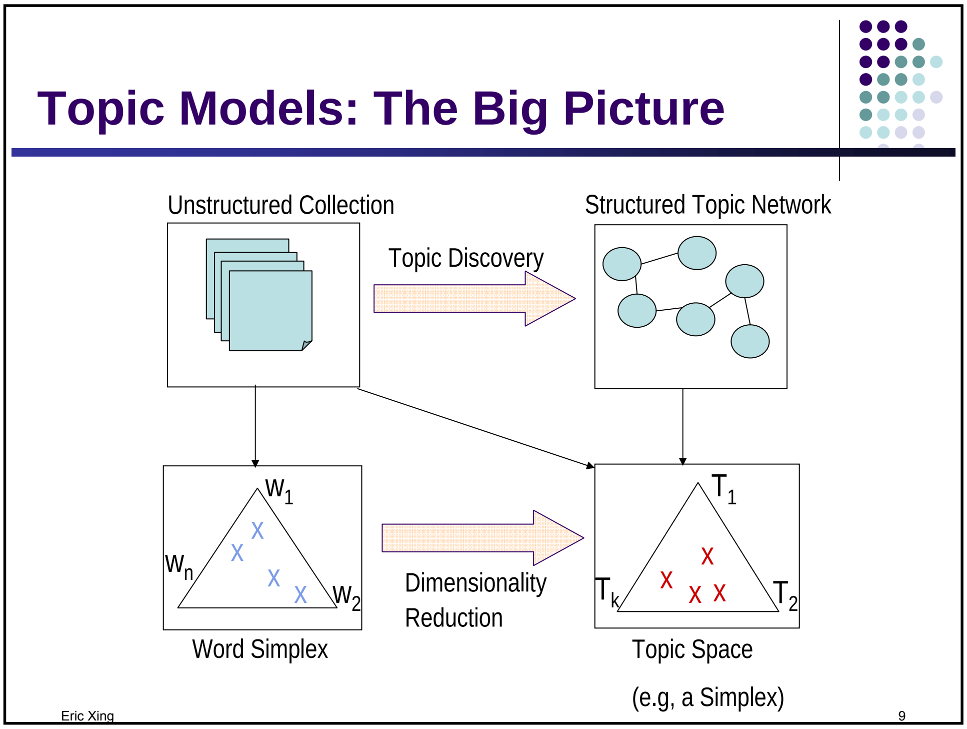
Words in Contexts (con'd)

- the opposition Labor **Party** fared even worse, with a predicted 35 **seats**, seven less than last **election**.



"Words" in Contexts (con'd)

Sivic et al. ICCV 2005



Topic Models: The Big Picture

Unstructured Collection

Topic Discovery

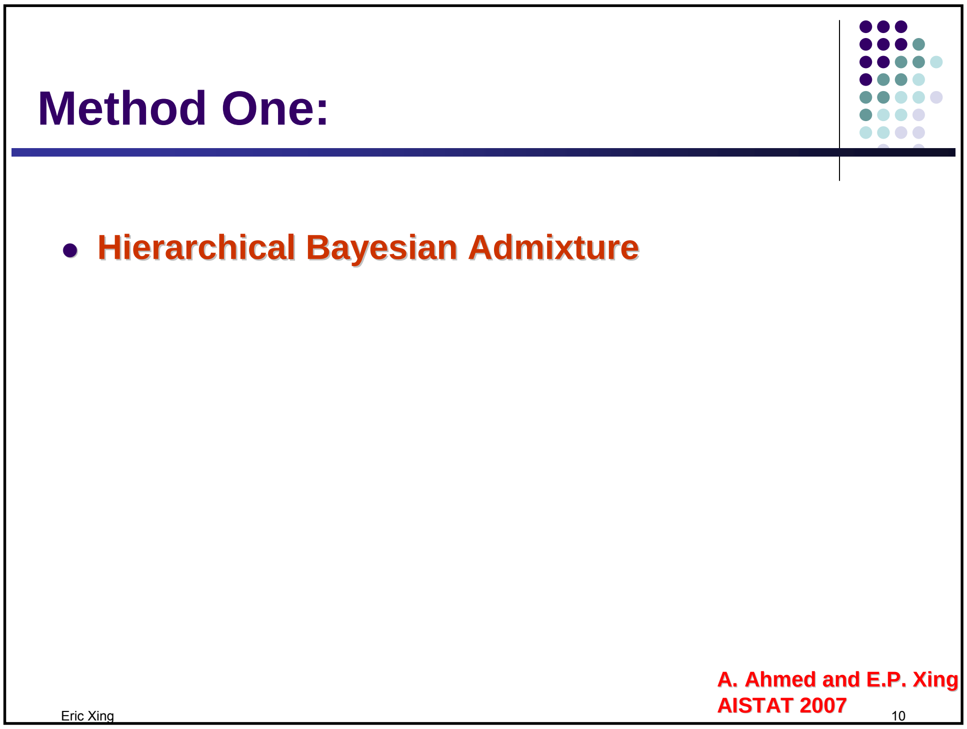
Structured Topic Network

Dimensionality Reduction

Word Simplex

Topic Space

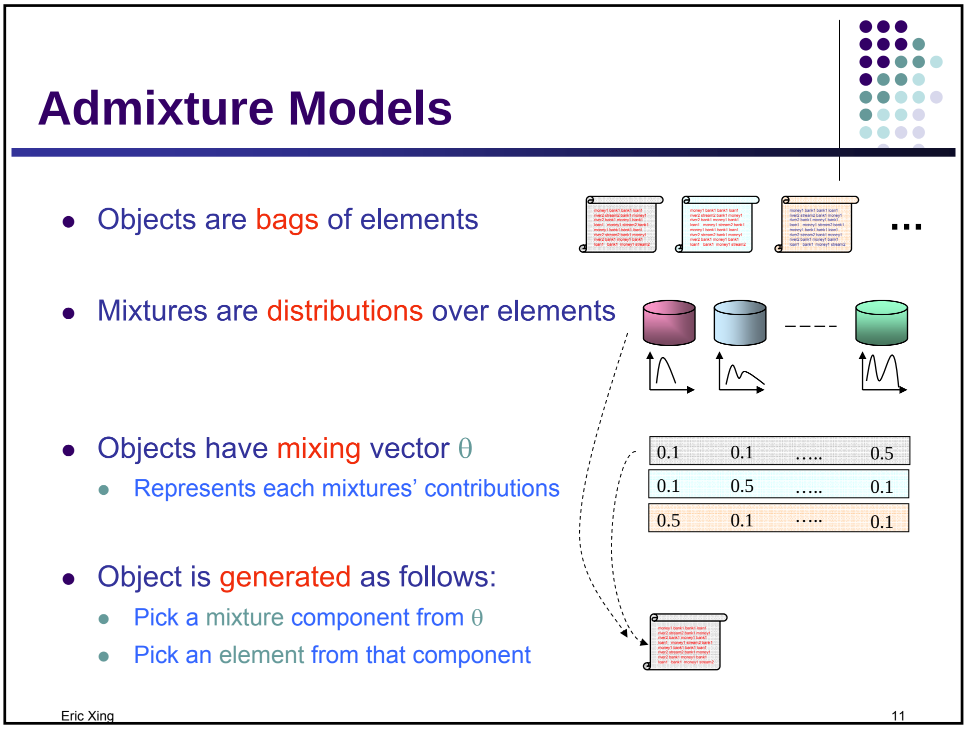
(e.g, a Simplex)



Method One:

- **Hierarchical Bayesian Admixture**

A. Ahmed and E.P. Xing
AISTAT 2007

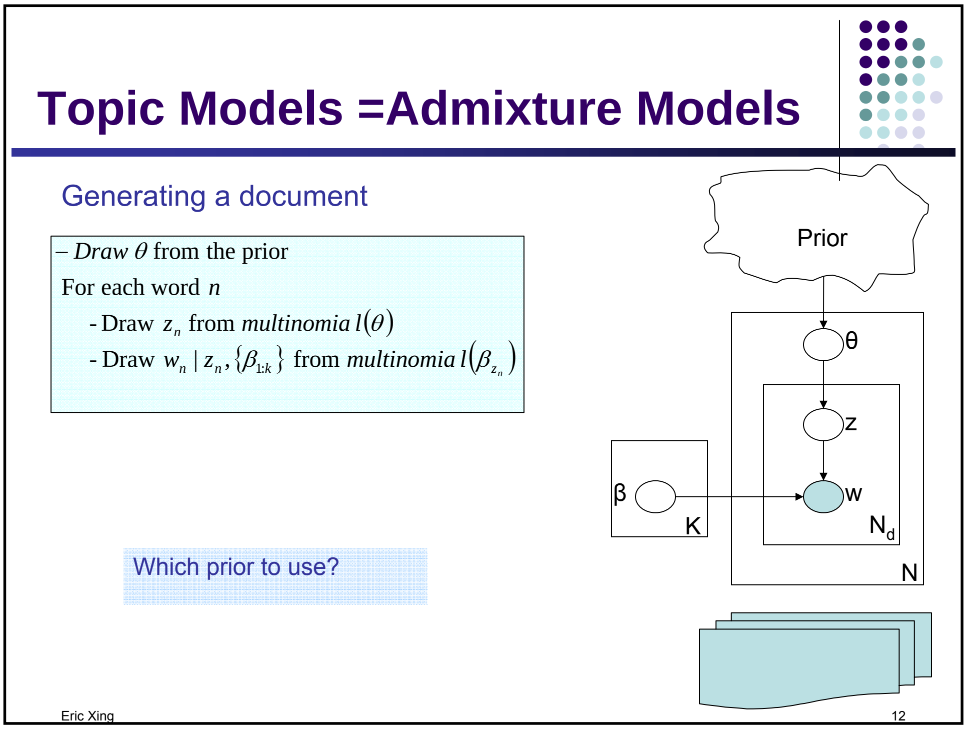


Admixture Models

- Objects are bags of elements …
- Mixtures are distributions over elements
- Objects have mixing vector θ
  - Represents each mixtures' contributions
- Object is generated as follows:
  - Pick a mixture component from θ
  - Pick an element from that component

0.1	0.1	…..	0.5
0.1	0.5	…..	0.1
0.5	0.1	…..	0.1

Eric Xing 11



Topic Models =Admixture Models

Generating a document

– *Draw θ* from the prior

For each word n

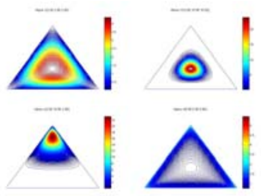
- Draw z_n from *multinomial* (θ)
- Draw $w_n \mid z_n, \{\beta_{1:k}\}$ from *multinomial* (β_{z_n})

Which prior to use?

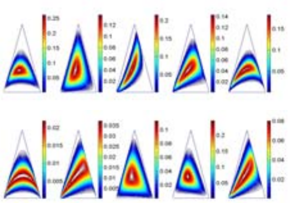
Eric Xing 12

Prior Comparison

- Dirichlet (LDA) (Blei et al. 2003)
  - Conjugate prior means efficient inference
  - Can only capture variations in each topic's intensity independently

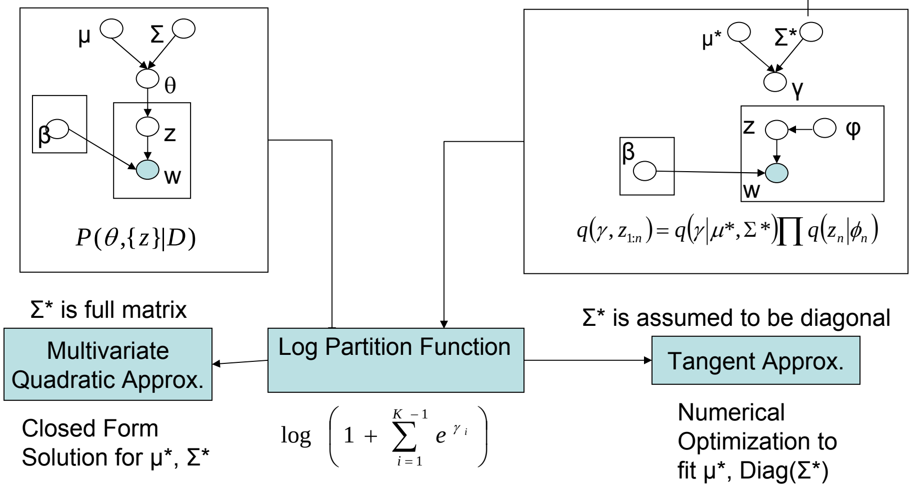


- Logistic Normal (CTM=LoNTAM) (Blei & Lafferty 2005, Ahmed & Xing 2006)
  - Capture the intuition that some topics are highly correlated and can rise up in intensity together
  - Not a conjugate prior implies hard inference



Approximate Inference

(e.g., MF, Jordan et al 1999, GMF, Xing et al 2004)



$P(\theta,\{z\}|D)$

$q(\gamma, z_{1:n}) = q(\gamma|\mu^*,\Sigma^*)\prod q(z_n|\phi_n)$

Σ* is full matrix

Multivariate Quadratic Approx.

Log Partition Function

Σ* is assumed to be diagonal

Tangent Approx.

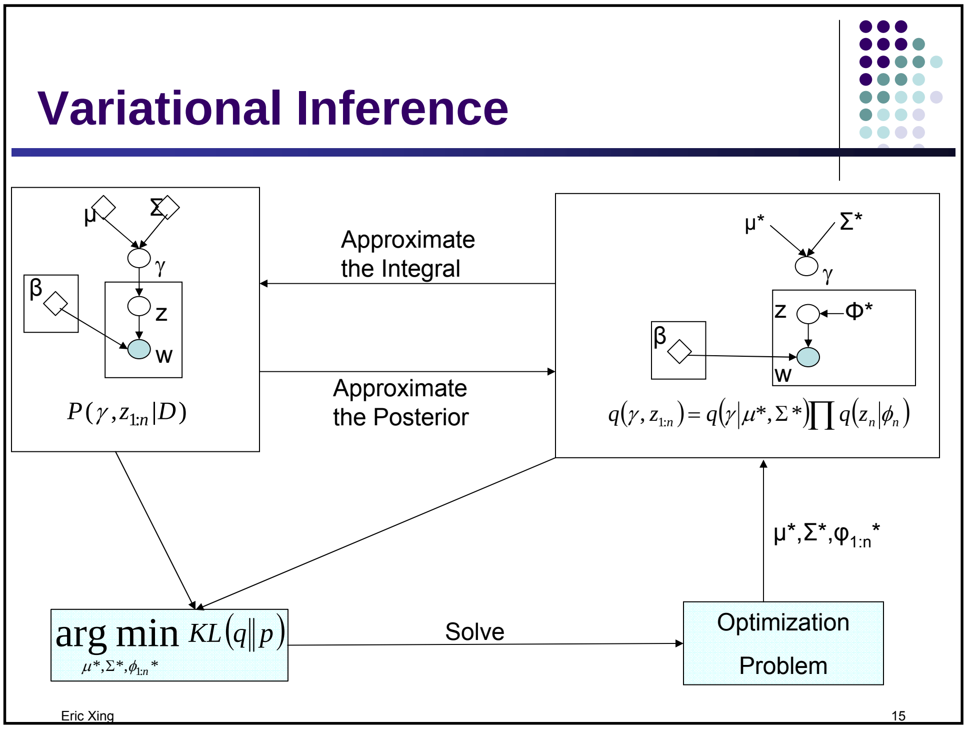
Closed Form Solution for μ*, Σ*

$\log\left(1+\sum_{i=1}^{K-1} e^{\gamma_i}\right)$

Numerical Optimization to fit μ*, Diag(Σ*)

Ahmed&Xing

Blei&Lafferty



Variational Inference

$P(\gamma, z_{1:n}|D)$

Approximate the Integral

Approximate the Posterior

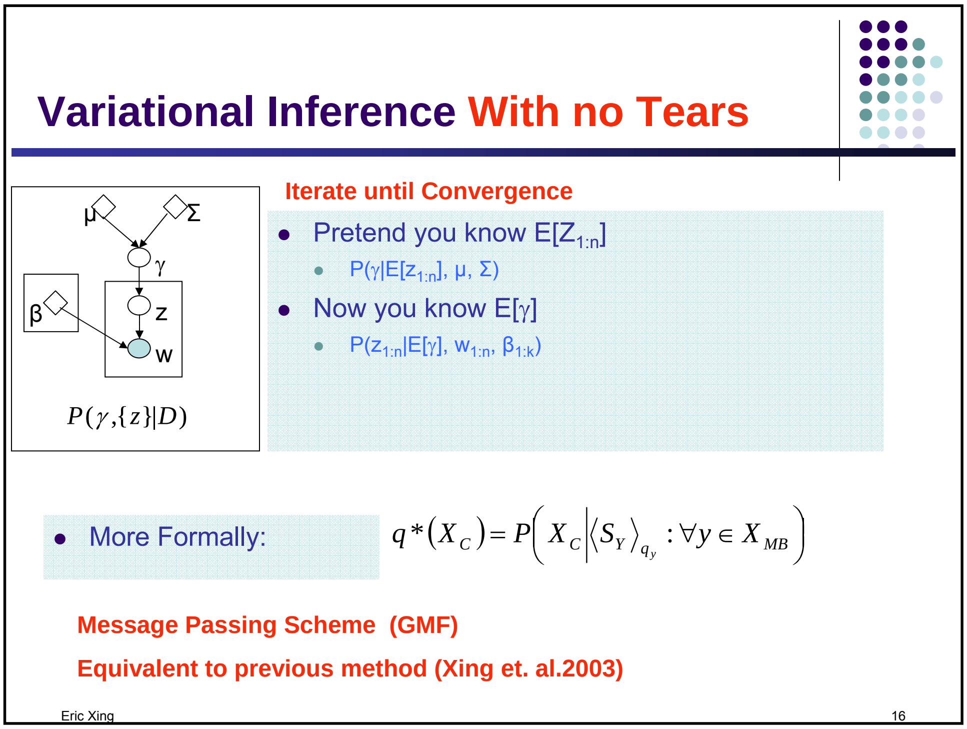
$q(\gamma, z_{1:n}) = q(\gamma|\mu^*, \Sigma^*)\prod q(z_n|\phi_n)$

$\mu^*, \Sigma^*, \varphi_{1:n}^*$

$\arg\min_{\mu^*, \Sigma^*, \phi_{1:n}^*} KL(q\|p)$

Solve

Optimization Problem



Variational Inference With no Tears

$P(\gamma, \{z\}|D)$

Iterate until Convergence

- Pretend you know E[$Z_{1:n}$]
  - P(γ|E[$z_{1:n}$], μ, Σ)
- Now you know E[γ]
  - P($z_{1:n}$|E[γ], $w_{1:n}$, $\beta_{1:k}$)

- More Formally: $q^*(X_C) = P\left(X_C \middle| \langle S_Y \rangle_{q_y} : \forall y \in X_{MB}\right)$

Message Passing Scheme (GMF)

Equivalent to previous method (Xing et. al.2003)

LoNTAM Variations Inference

- Fully Factored Distribution

$$q(\gamma, z_{1:n}) = q(\gamma)\prod q(z_n)$$

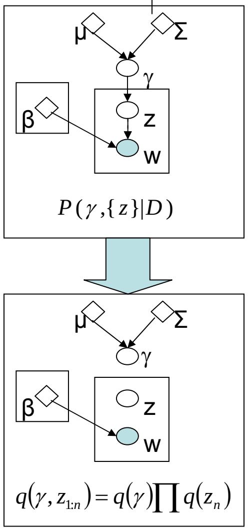
- Two clusters: λ and $Z_{1:n}$

$$q^*(X_C) = P\left(X_C \middle| \langle S_Y \rangle_{q_y} : \forall y \in X_{MB} \right)$$

- Fixed Point Equations

$$q_\gamma {}^*(\gamma) = P\left(\gamma \middle| \langle S_z \rangle_{q_z}, \mu, \Sigma\right)$$

$$q_z {}^*(z) = P\left(z \middle| \langle S_\gamma \rangle_{q\gamma}, \beta_{1:k} \right)$$



Variational γ

$$q_\lambda {}^*(\gamma) = P\left(\gamma \middle| \langle S_z \rangle_{q_z}, \mu, \Sigma\right)$$

$$\propto P(\gamma \mid \mu, \Sigma) P\left(\langle S_z \rangle_{q_z} | \gamma\right)$$

Now what is $\langle S_z \rangle_{q_z}$?

$$S_z = m = \left[\sum_n I(z_n = 1), \ldots, \sum_n I(z_n = k) \right]$$

$$\propto N(\gamma \mid \mu, \Sigma) \exp\left\{ \langle m \rangle_{q_z} \gamma - N \times C(\gamma) \right\}$$

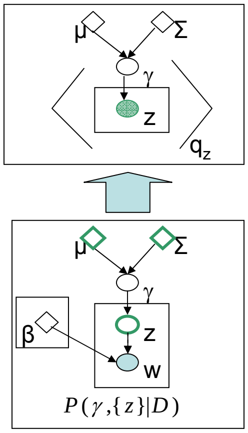
$$\propto \exp\left\{ \underbrace{-\frac{1}{2}\gamma' \Sigma^{-1} \gamma + \gamma \Sigma^{-1} \mu + \langle m \rangle_{q_z}}\gamma - N \times \underbrace{\boxed{C(\gamma)}} \right\}$$

$$\boxed{C(\gamma) = C(\gamma_\wedge) + g'_\lambda (\gamma - \gamma_\wedge) + .5\,(\lambda - \gamma_\wedge)' H\,(\gamma - \gamma_\wedge)}$$

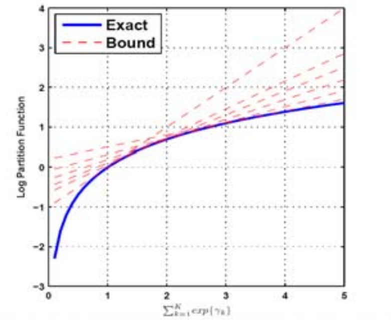
$$q_\lambda^*(\gamma) = N(\mu_\gamma, \Sigma_\gamma)$$

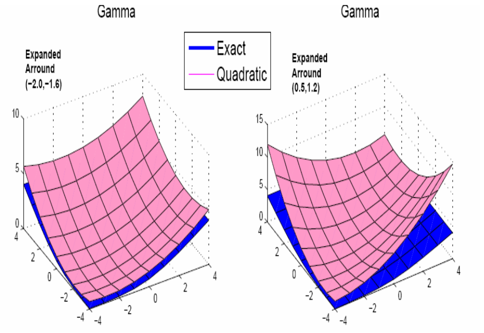
$$\Sigma_\gamma = inv\,(\Sigma^{-1} + NH)$$

$$\mu_\gamma = \Sigma_\gamma\left(\Sigma^{-1}\mu + NH\gamma_\wedge + \langle m \rangle - Ng\right)$$

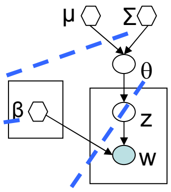


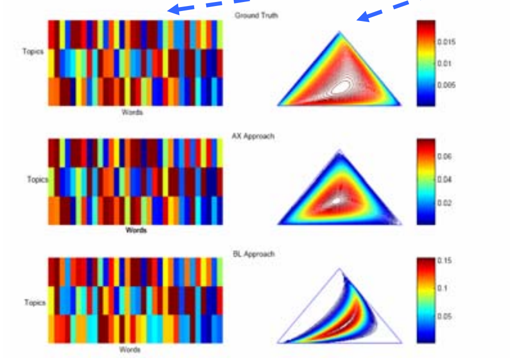
Tangent Approximation

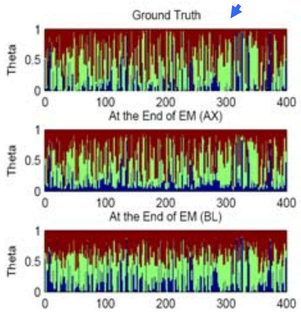




Test on Synthetic Text



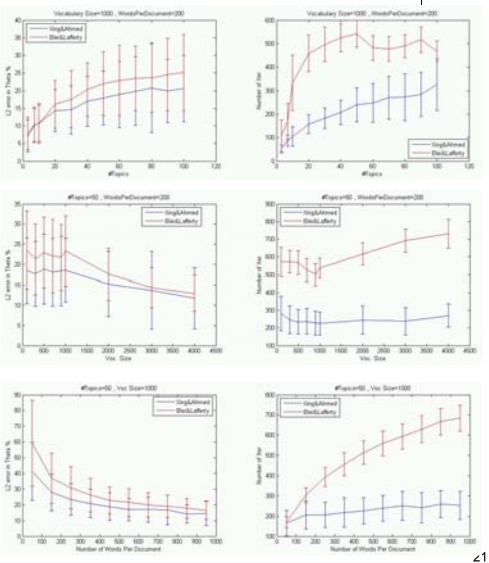




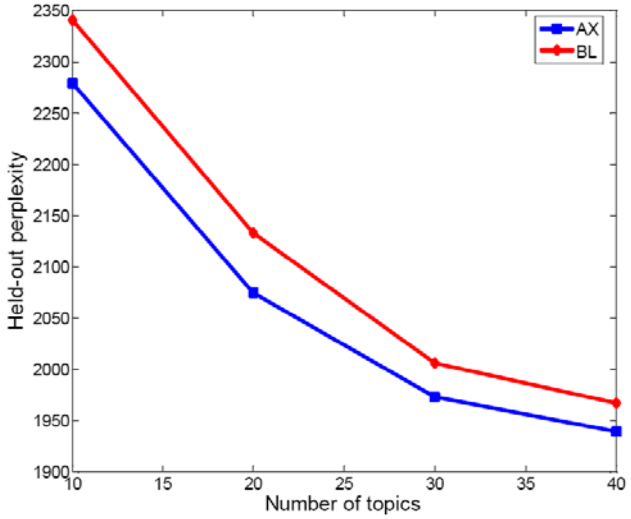
Comparison: accuracy and speed

L2 error in topic vector est. and # of iterations

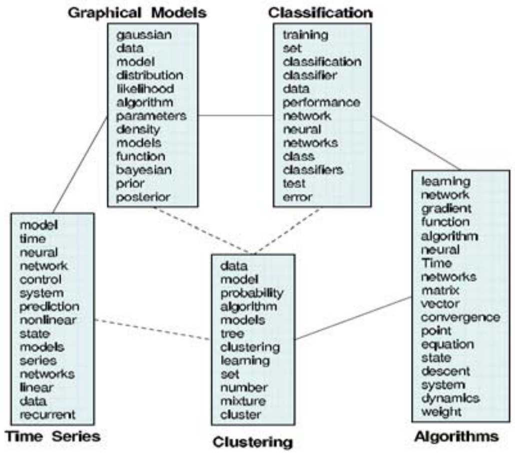
- Varying Num. of Topics
- Varying Voc. Size
- Varying Num. Words Per Document



Comparison: perplexity



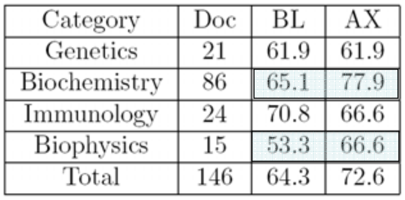
Topics and topic graphs



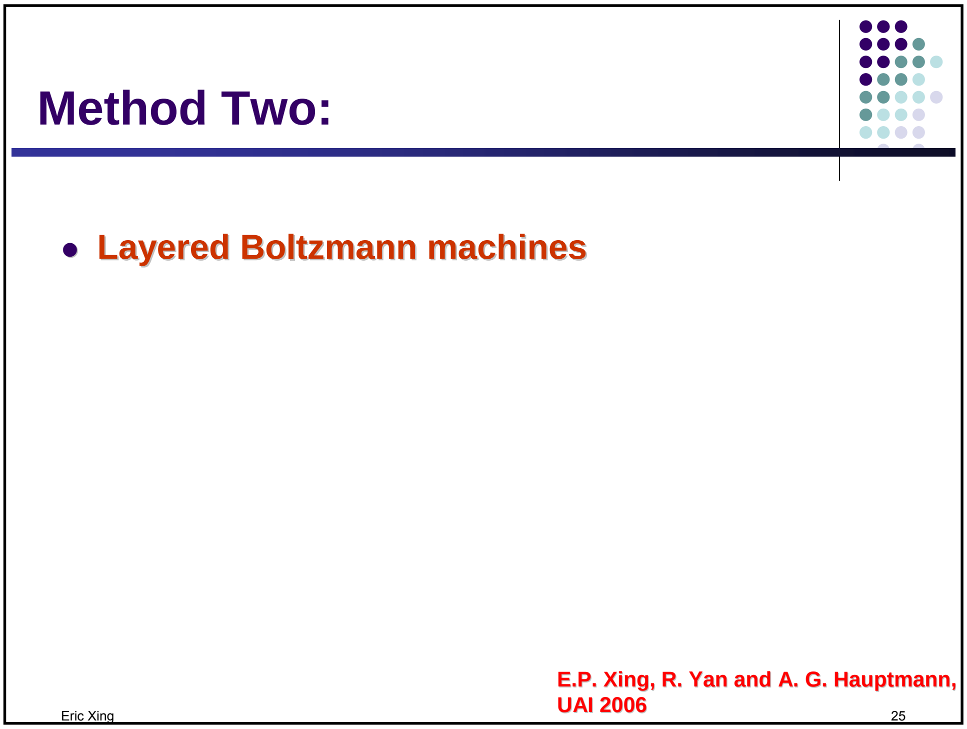
Result on PNAS collection

- PNAS abstracts from 1997-2002
  - 2500 documents
  - Average of 170 words per document
- Fitted 40-topics model using both approaches
- Use low dimensional representation to predict the abstract category
  - Use SVM classifier
  - 85% for training and 15% for testing

Classification Accuracy



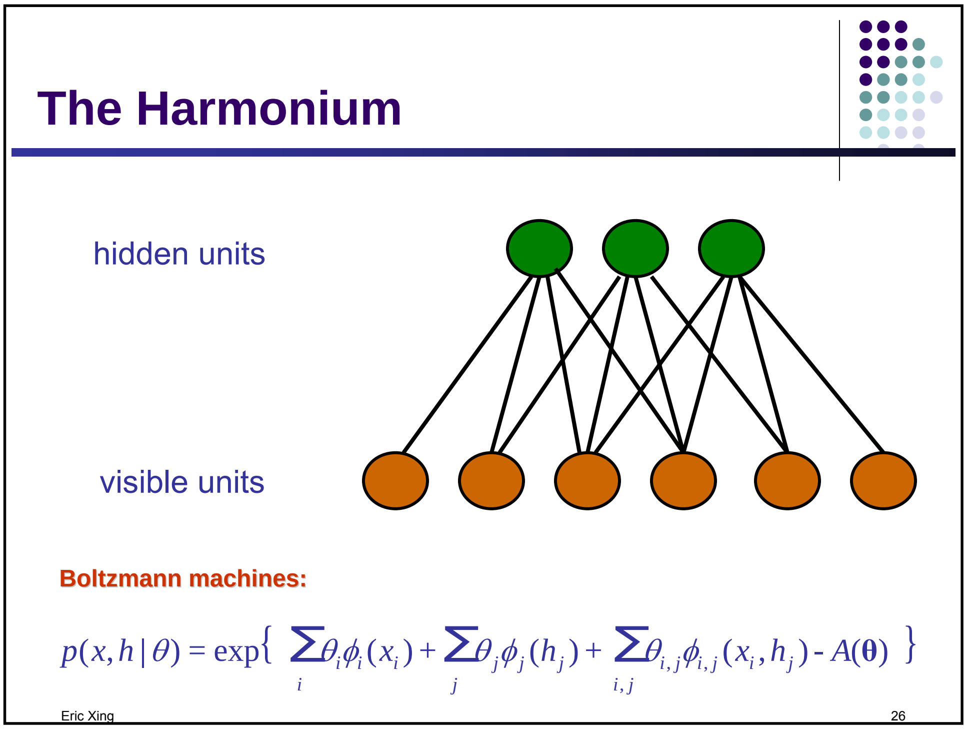
Category	Doc	BL	AX
Genetics	21	61.9	61.9
Biochemistry	86	65.1	77.9
Immunology	24	70.8	66.6
Biophysics	15	53.3	66.6
Total	146	64.3	72.6



Method Two:

- Layered Boltzmann machines

E.P. Xing, R. Yan and A. G. Hauptmann, UAI 2006



The Harmonium

hidden units

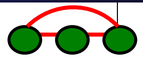
visible units

Boltzmann machines:

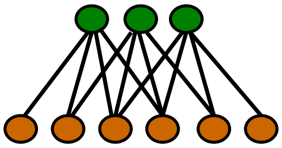
$$p(x,h\,|\,\theta) = \exp\Big\{ \sum_i \theta_i \phi_i(x_i) + \sum_j \theta_j \phi_j(h_j) + \sum_{i,j} \theta_{i,j} \phi_{i,j}(x_i, h_j) - A(\boldsymbol{\theta}) \Big\}$$

Properties of Harmoniums

- Factors are marginally *dependent*.

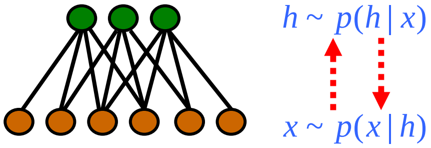


- Factors are conditionally *independent* given observations on the visible nodes.

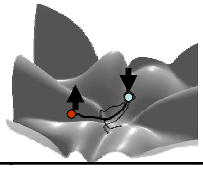


$$P(\ell | \mathbf{w}) = \prod_i P(\ell_i | \mathbf{w})$$

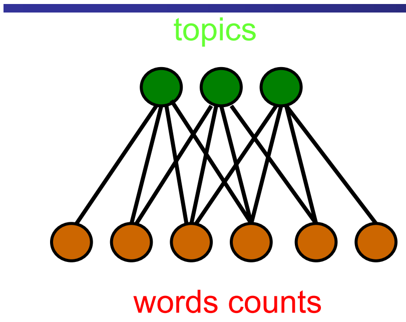
- Iterative Gibbs sampling.



- Learning with contrastive divergence



A Binomial Word-count Model



h_j = 3: *topic j has strength 3*

$h_j \in \mathbf{R}$, $\langle h_j \rangle = \sum_i W_{i,j} x_i$

x_i = n: *word i has count n*

$x_i \in \mathbf{I}$

$$\text{Bi}_{x_i}[N, p] = C_{x_i}^N p^{x_i} (1-p)^{N-x_i} = C_{x_i}^N \left(\tfrac{p}{1-p}\right)^{x_i} (1-p)^N$$

Let $p = \frac{\exp(\alpha_j + \sum_j W_{ij} h_j)}{1+\exp(\alpha_j + \sum_j W_{ij} h_j)}$,

$$\text{Bi}_{x_i}[N, p] = C_{x_i}^N \frac{\left(\exp(\alpha_i + \sum_j W_{ij} h_j)\right)^{x_i}}{\left(1+\exp(\alpha_i + \sum_j W_{ij} h_j)\right)^N}$$

$$\propto C_{x_i}^N \exp\left\{\left(\alpha_i + \sum_j W_{ij} h_j\right) x_i + A_i\right\}$$

Reduce to softmax when N=1 !

$$p(\mathbf{h} | \mathbf{x}) = \prod_j \text{Normal}_{h_j} \left[\sum_i \vec{W}_{ij} \vec{x}_i, 1 \right]$$

$$p(\mathbf{x} | \mathbf{h}) = \prod_i \text{Bi}_{x_i} \left[N, \frac{\exp(\alpha_j + \sum_j W_{ij} h_j)}{1+\exp(\alpha_j + \sum_j W_{ij} h_j)} \right]$$

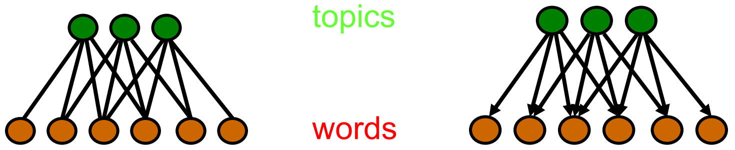
$$\Rightarrow p(\mathbf{x}) \propto \exp\left\{\left(\sum_i \alpha_i x_i - \log\Gamma(x_i) - \log\Gamma(N - x_i)\right) + \tfrac{1}{2} \sum_j \left(\sum_i W_{i,j} x_i\right)^2\right\}$$

The Computational Trade-off

Undirected model: Learning is hard, inference is easy.

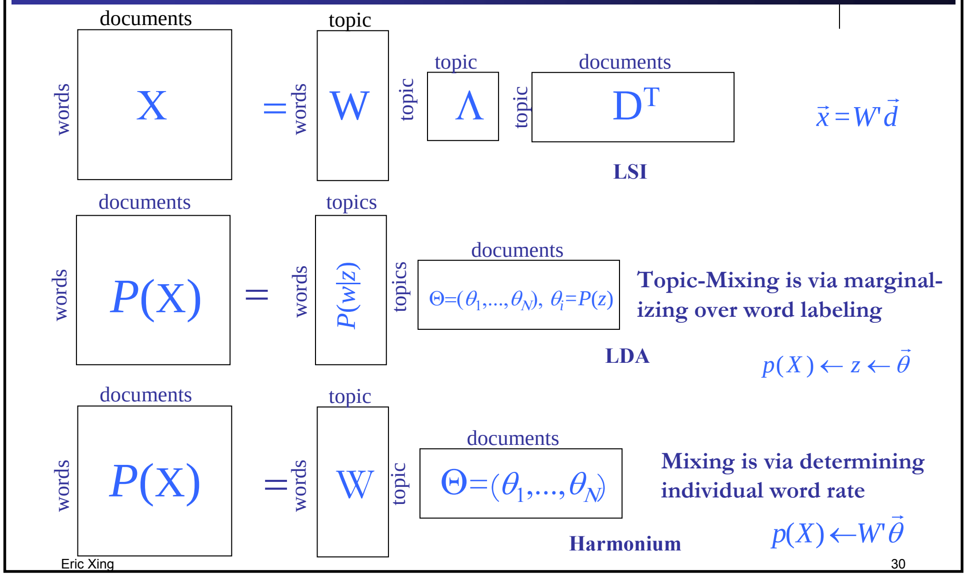
Directed Model: Learning is "easier", inference is hard.

Example: Document Retrieval.



Retrieval is based on comparing (posterior) topic distributions of documents.
- directed models: inference is slow. Learning is relatively “easy”.
- undirected model: inference is fast. Learning is slow but can be done offline.

Comparison of model semantics



$\vec{x} = W'\vec{d}$

LSI

Topic-Mixing is via marginalizing over word labeling

LDA

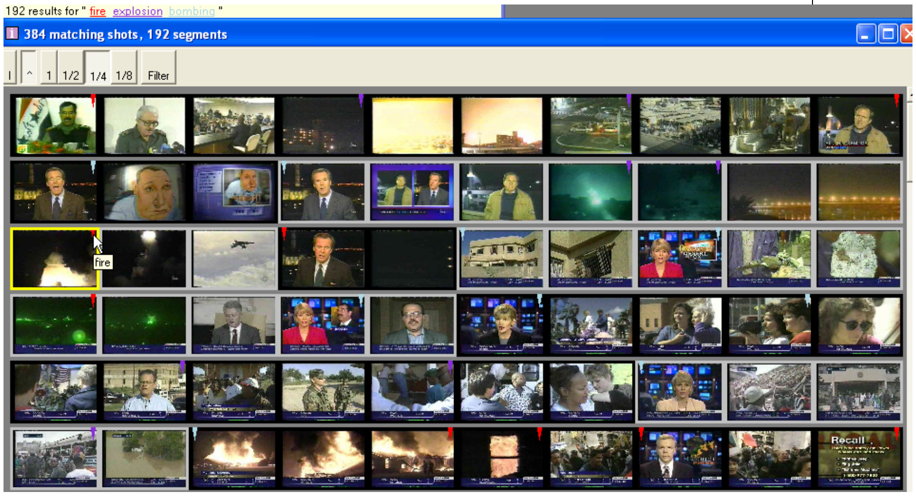
$p(X) \leftarrow z \leftarrow \vec{\theta}$

Mixing is via determining individual word rate

Harmonium

$p(X) \leftarrow W'\vec{\theta}$

Multi-Source Data

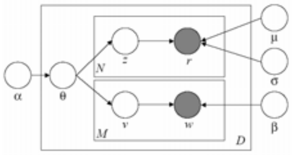


192 results for " fire explosion bombing "

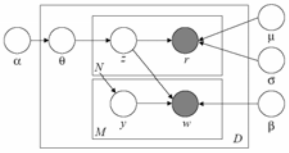
384 matching shots, 192 segments

TRECVID 2004 Example Images

Inter-Source Associations

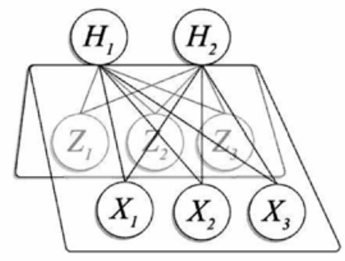


GM LDA ⇓

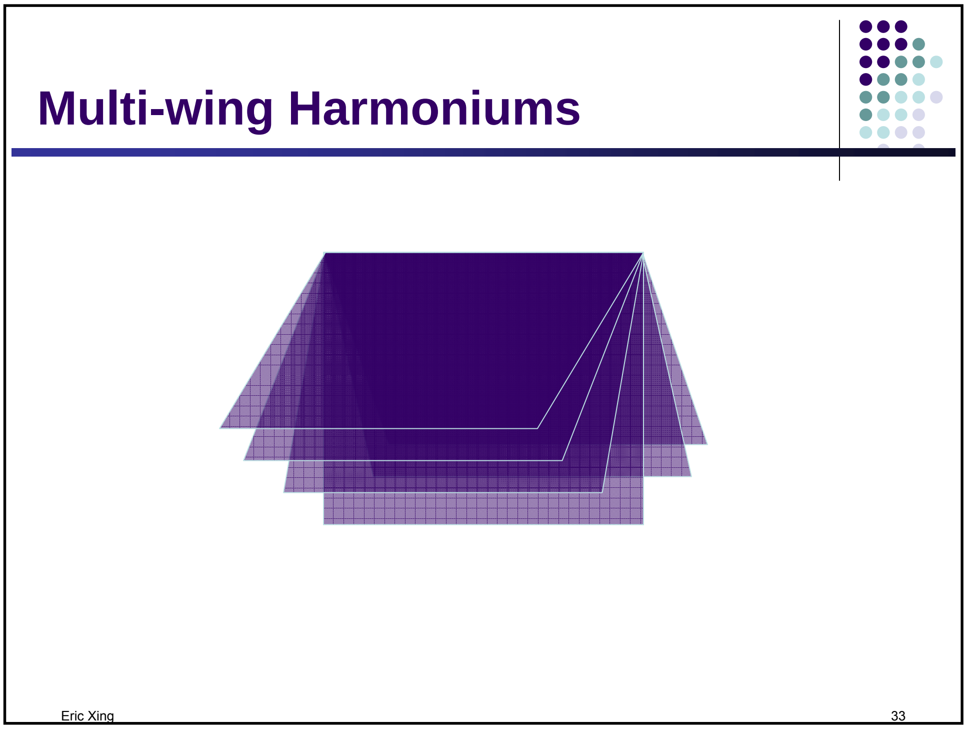


Co-LDA

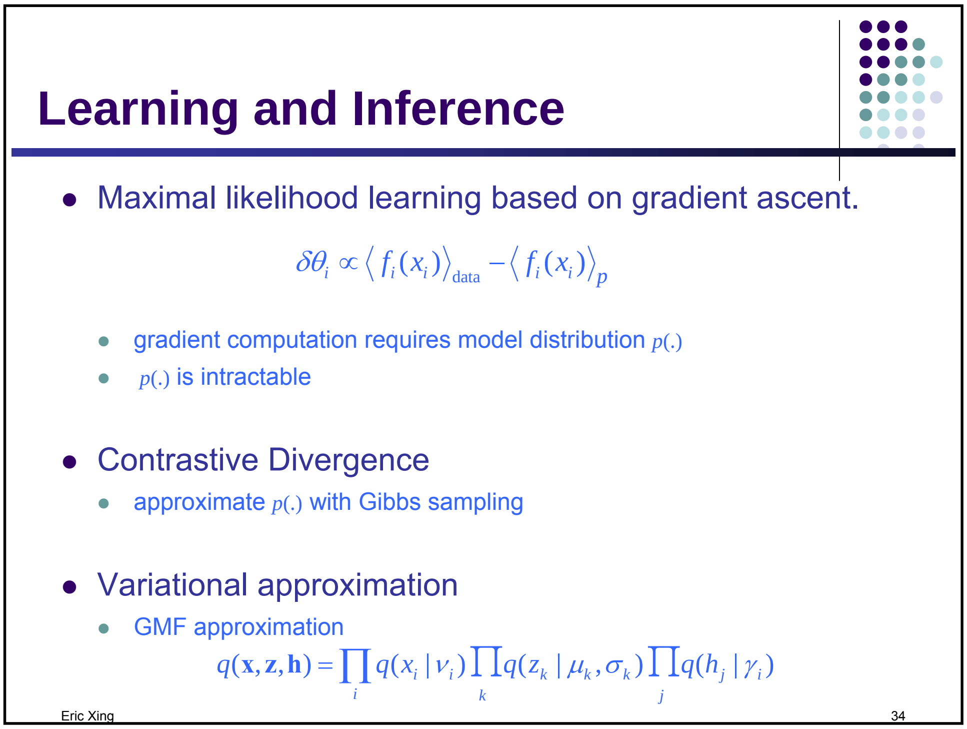
DWH



Z and *X* are marginally dependent (same as GM-LDA)



Multi-wing Harmoniums



Learning and Inference

- Maximal likelihood learning based on gradient ascent.

$$\delta\theta_i \propto \left\langle f_i(x_i)\right\rangle_{\text{data}} - \left\langle f_i(x_i)\right\rangle_p$$

  - gradient computation requires model distribution $p(.)$
  - $p(.)$ is intractable
- Contrastive Divergence
  - approximate $p(.)$ with Gibbs sampling
- Variational approximation
  - GMF approximation

$$q(\mathbf{x},\mathbf{z},\mathbf{h}) = \prod_i q(x_i \mid \nu_i)\prod_k q(z_k \mid \mu_k,\sigma_k)\prod_j q(h_j \mid \gamma_i)$$

Inter-source Inference

- GMF approximation to DWH

$$q(\mathbf{x},\mathbf{z},\mathbf{h}) = \prod_i q(x_i \mid N,\nu_i)\prod_k q(z_k \mid \mu_k,\sigma_k)\prod_j q(z_k \mid \mu_k,\sigma_k)$$

  - Expected mean value of topic strength:

$$\gamma_j = \sum_i W_{i,j}\nu_i + \sum_k U_{k,j}\mu_k$$

  - Expected mean value of image-feature :

$$\mu_k = \sigma_k^2\left(\beta_k + \sum_j {}_j U_{k,j}\gamma_j\right)$$

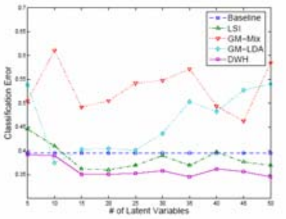
  - Expected mean count

$$N\nu_i = N\frac{\exp(\alpha_j + \sum_j W_{ij}\gamma_j)}{1+\exp(\alpha_j + \sum_j W_{ij}\gamma_j)}$$

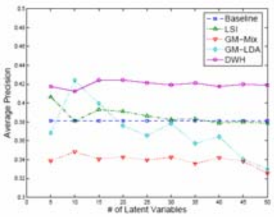
Examples of Latent Topics

T_1	storms gulf hawaii low forecast southeast showers
T_2	rebounds 14 shouting tests guard cut hawks
T_3	engine flying craft asteroid say hour aerodynamic
T_4	safe cross red sure dry providing services
T_5	losing jersey sixth antonio david york orlando

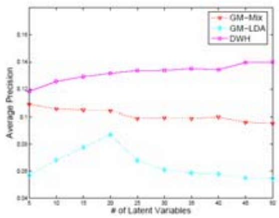
Performance



Classification



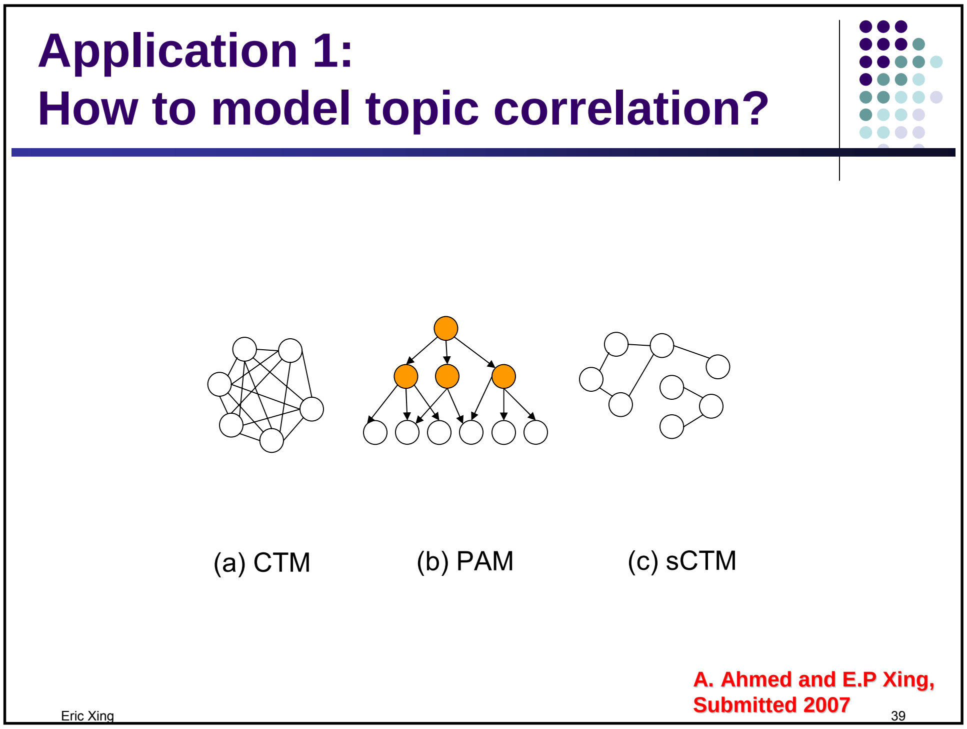
Retrieval



Annotation

This Talk

- A graphical model primer
- Two families of probabilistic topics models and approximate inference
  - Bayesian admixture models
  - Random models
- Three applications
  - Learning topic graphs and topic evolution
  - Machine translation
  - Multimedia inference



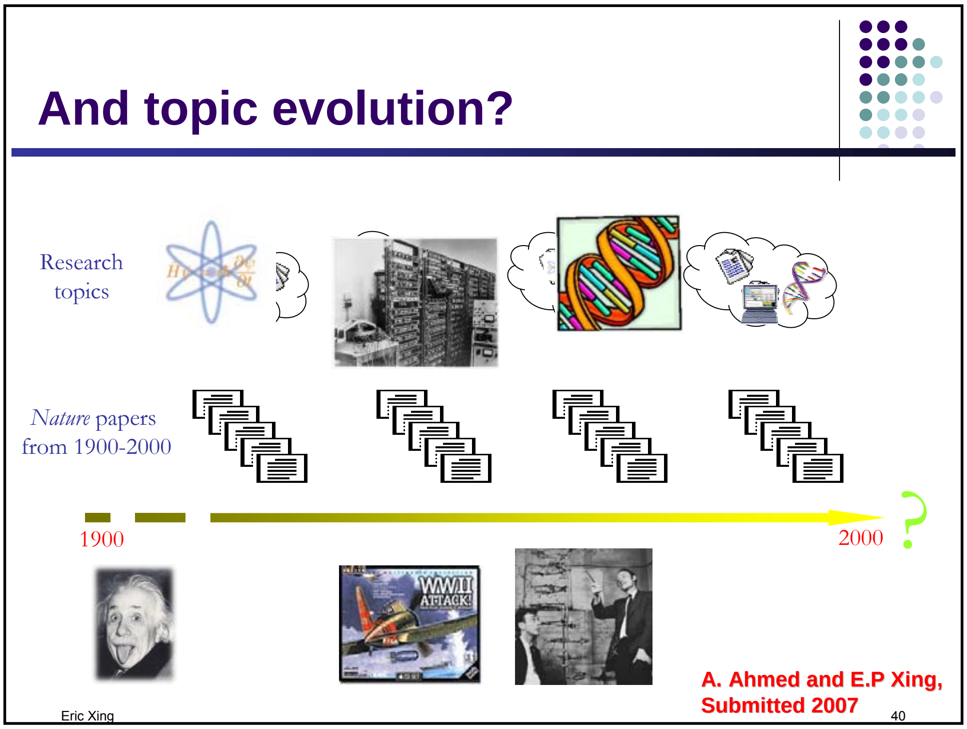
Application 1: How to model topic correlation?

(a) CTM

(b) PAM

(c) sCTM

A. Ahmed and E.P Xing, Submitted 2007



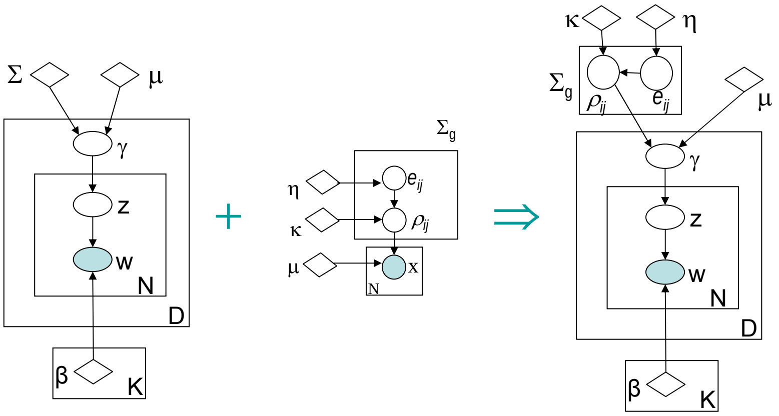
And topic evolution?

Research topics

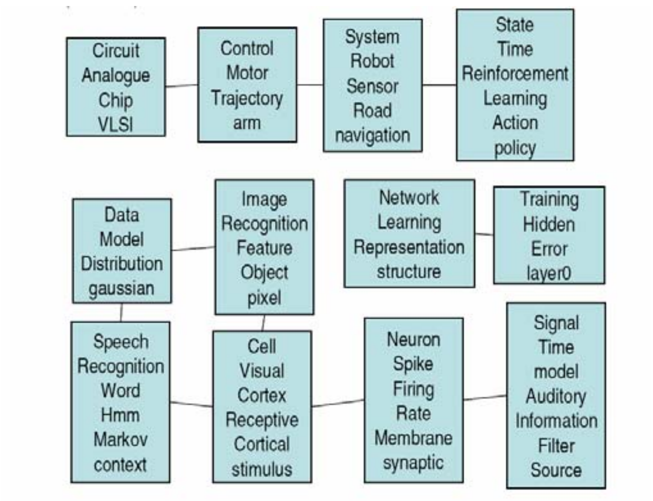
Nature papers from 1900-2000

A. Ahmed and E.P Xing, Submitted 2007

Sparse Correlated Model (SCTM)



NIPS: Example Network



Circuit Analogue Chip VLSI

Control Motor Trajectory arm

System Robot Sensor Road navigation

State Time Reinforcement Learning Action policy

Data Model Distribution gaussian

Image Recognition Feature Object pixel

Network Learning Representation structure

Training Hidden Error layer0

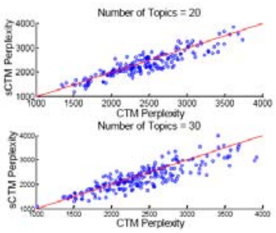
Speech Recognition Word Hmm Markov context

Cell Visual Cortex Receptive Cortical stimulus

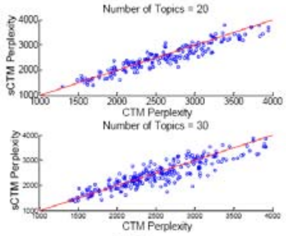
Neuron Spike Firing Rate Membrane synaptic

Signal Time model Auditory Information Filter Source

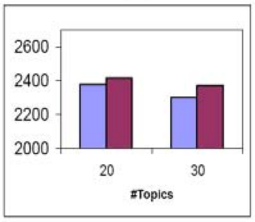
NIPS: Held-out Perplexity

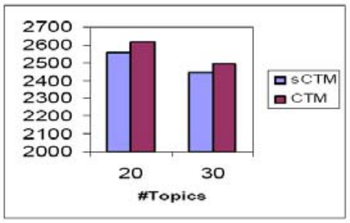


(c)



(d)





How to Model Topic Evolution

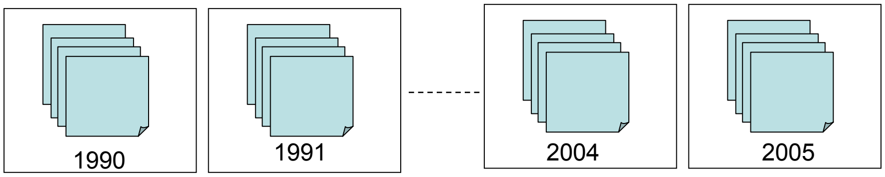
Topic Trends

Topic Keywords

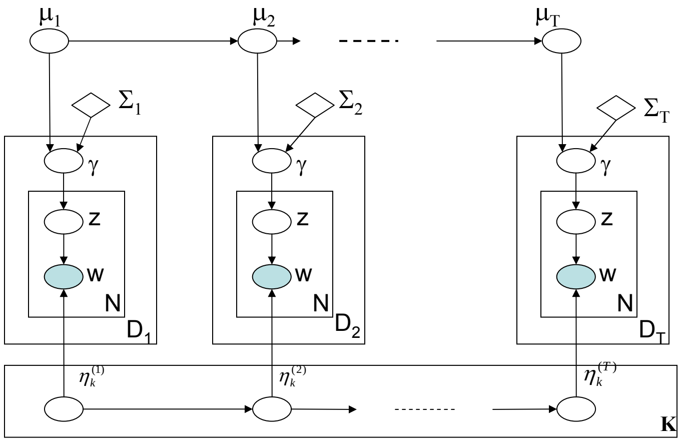
Topic correlations

~~Number of topics~~

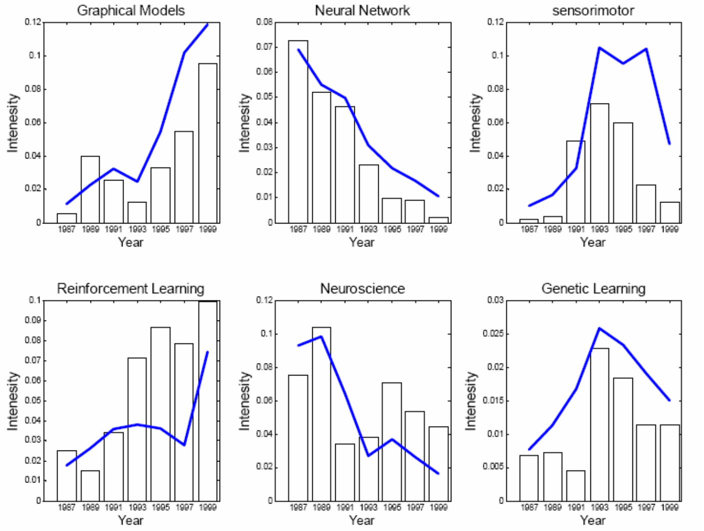
The Dynamic Correlated Topic model

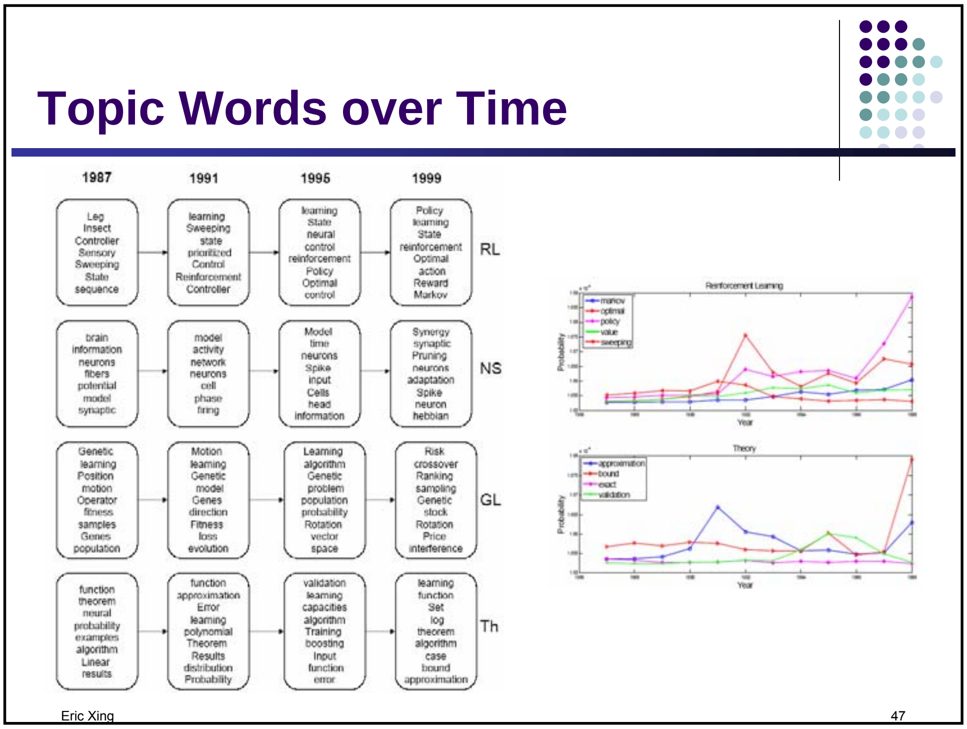


The Dynamic CTM



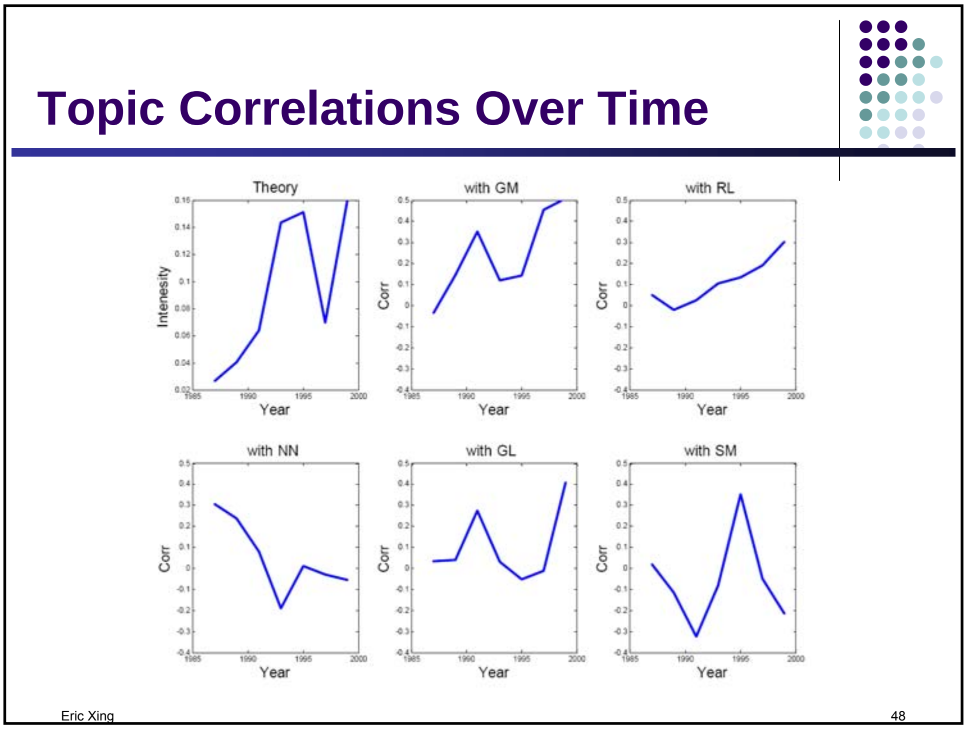
Topic Trends



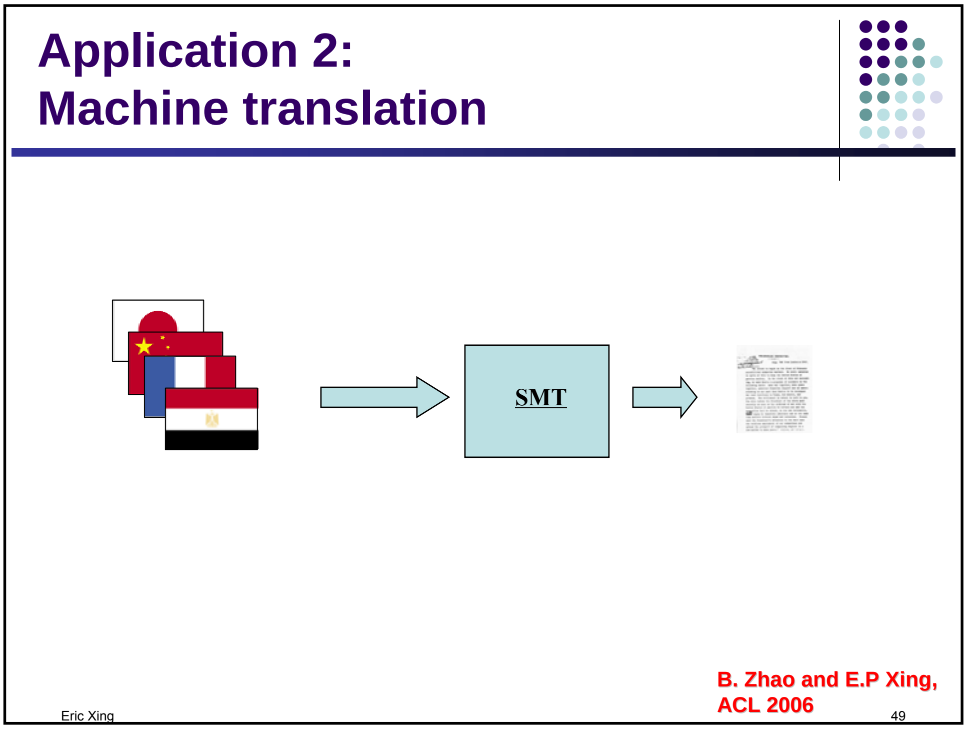


Topic Words over Time

1987	1991	1995	1999	
Leg Insect Controller Sensory Sweeping State sequence	learning Sweeping state prioritized Control Reinforcement Controller	learning State neural control reinforcement Policy Optimal control	Policy learning State reinforcement Optimal action Reward Markov	RL
brain information neurons fibers potential model synaptic	model activity network neurons cell phase firing	Model time neurons Spike input Cells head information	Synergy synaptic Pruning neurons adaptation Spike neuron hebbian	NS
Genetic learning Position motion Operator fitness samples Genes population	Motion learning Genetic model Genes direction Fitness loss evolution	Learning algorithm Genetic problem population probability Rotation vector space	Risk crossover Ranking sampling Genetic stock Rotation Price interference	GL
function theorem neural probability examples algorithm Linear results	function approximation Error learning polynomial Theorem Results distribution Probability	validation learning capacities algorithm Training boosting Input function error	learning function Set log theorem algorithm case bound approximation	Th

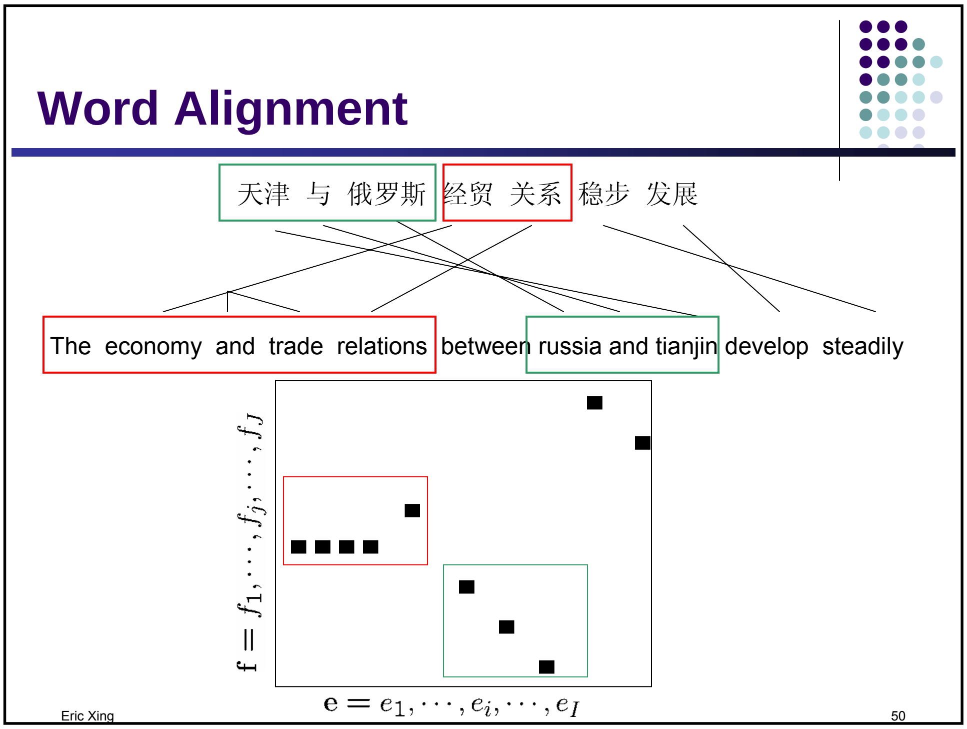


Topic Correlations Over Time



Application 2: Machine translation

B. Zhao and E.P Xing, ACL 2006

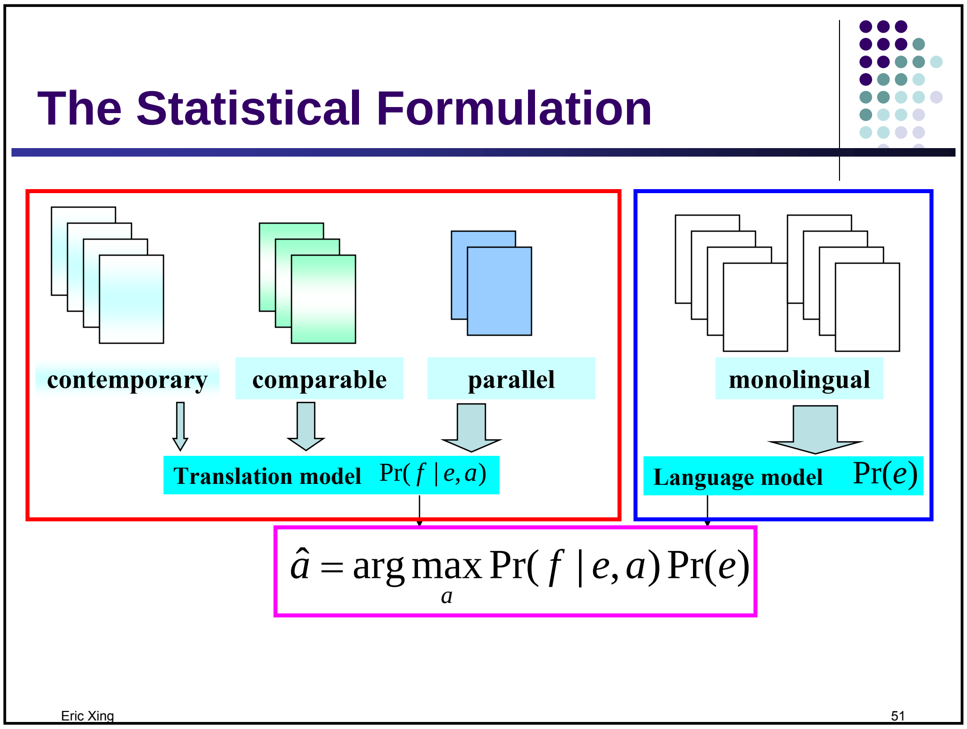


Word Alignment

天津 与 俄罗斯 经贸 关系 稳步 发展
The economy and trade relations between russia and tianjin develop steadily

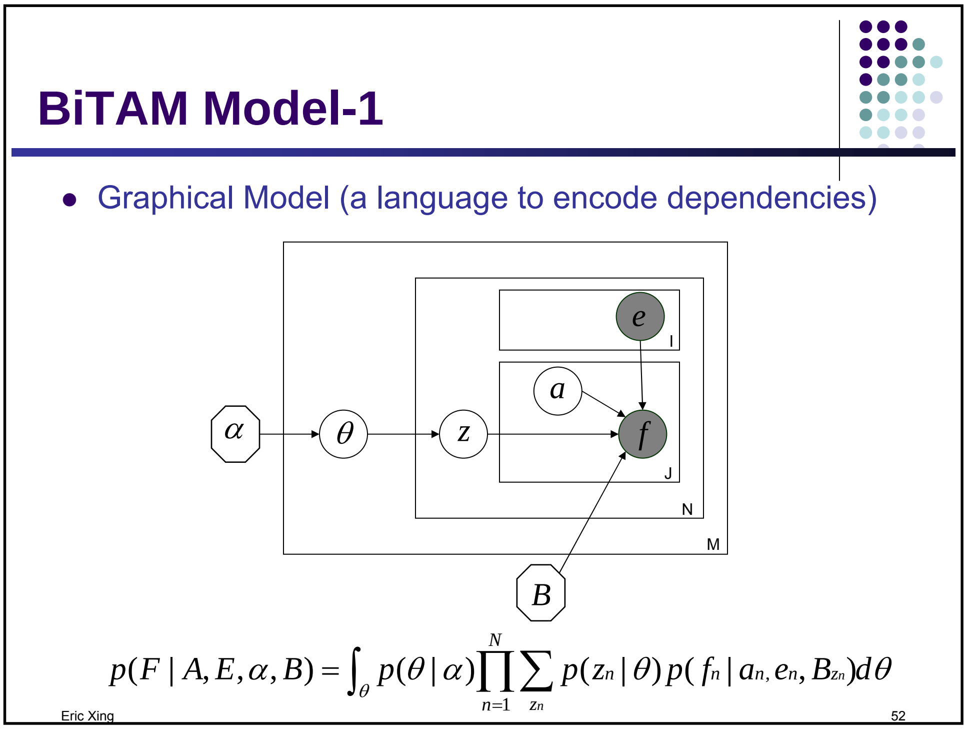
$$\mathbf{f} = f_1, \cdots, f_j, \cdots, f_J$$

$$\mathbf{e} = e_1, \cdots, e_i, \cdots, e_I$$



The Statistical Formulation

$$\hat{a}=\arg\max_{a}\Pr(f\,|\,e,a)\Pr(e)$$

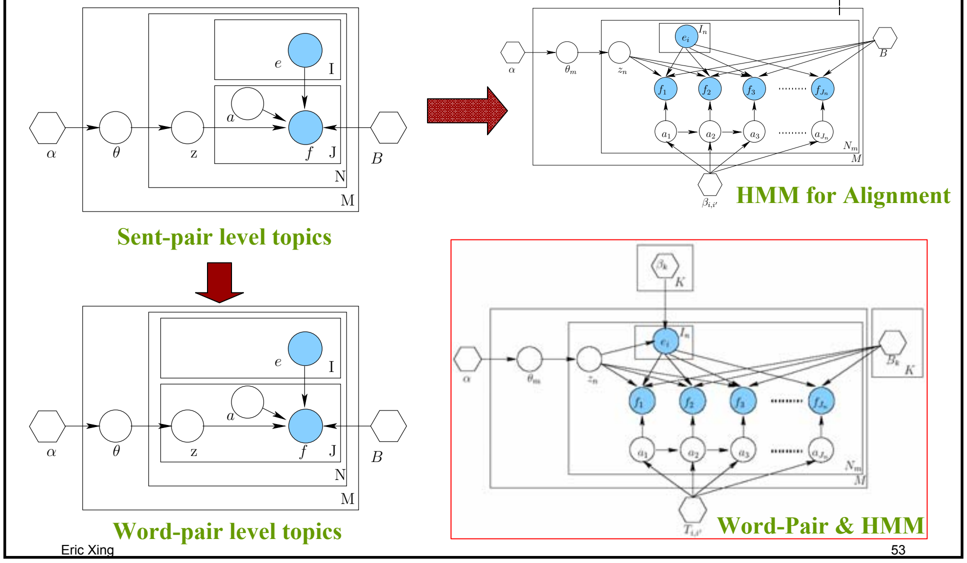


BiTAM Model-1

- Graphical Model (a language to encode dependencies)

$$p(F\,|\,A,E,\alpha,B)=\int_{\theta}p(\theta\,|\,\alpha)\prod_{n=1}^{N}\sum_{z_n}p(z_n\,|\,\theta)\,p(f_n\,|\,a_n,e_n,B_{z_n})d\theta$$

An upgrade path for BiTAMs



Sent-pair level topics

HMM for Alignment

Word-pair level topics

Word-Pair & HMM

Experiments

- Training data
  - Small: Treebank 316 doc-pairs (133K English words)
  - Large: FBIS-Beijing, Sinorama, XinHuaNews, (15M English words).

Train	#Doc.	#Sent.	#Tokens	
			English	Chinese
Treebank	316	4172	133K	105K
FBIS.BJ	6,111	105K	4.18M	3.54M
Sinorama	2,373	103K	3.81M	3.60M
XinHua	19,140	115K	3.85M	3.93M
FOUO	15,478	368K	13.14M	11.93M
Test	95	627	25,500	19,726

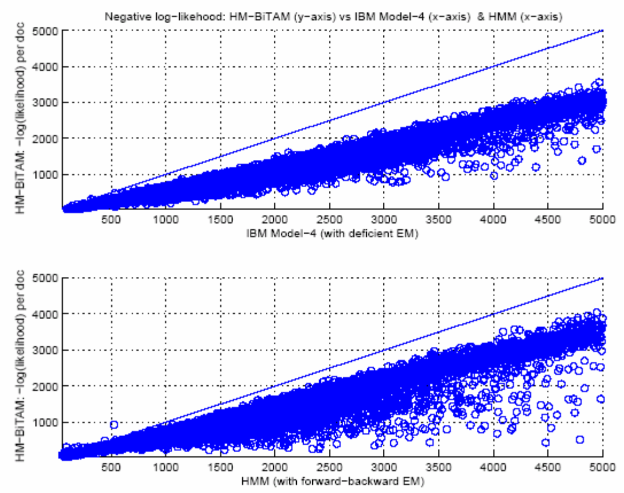
- Word Alignment Accuracy & Translation Quality
  - F-measure
  - BLEU

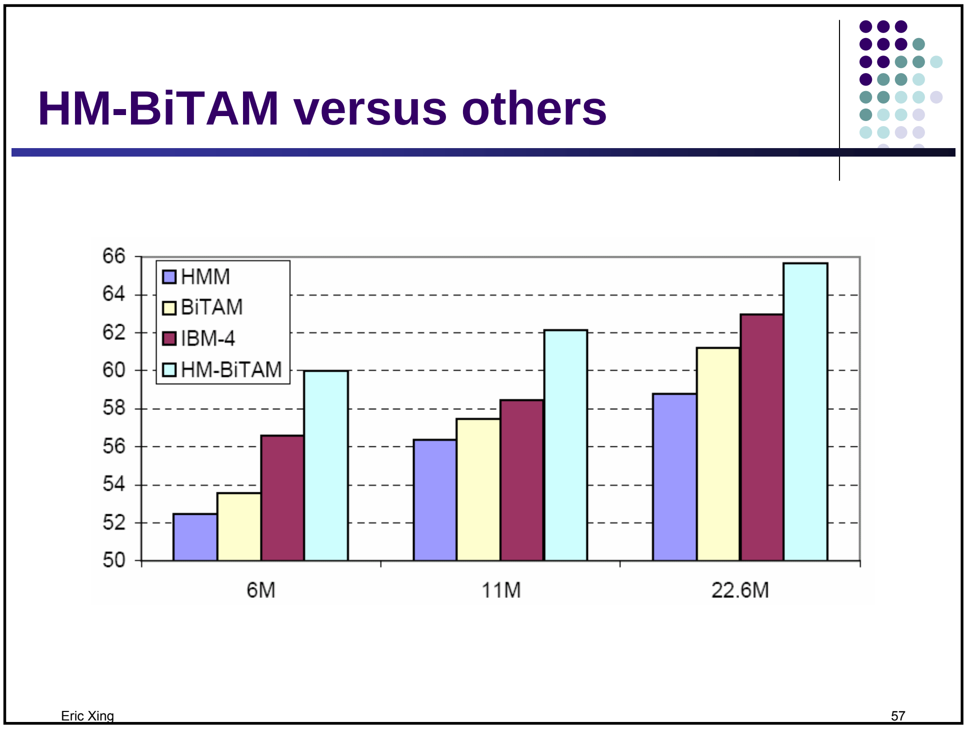
Topics

T1	Teams, sports, disabled, games members, people, cause, water, national, handicapped
T2	Shenzhen, singapore, hongkong, stock, national, investment, yuan, options, million, dollar
T3	Chongqing, company, takeover, shenzhen, tianjin, city, national, government, project, companies
T4	Hongkong, trade, export, import, foreign, tech., high, 1998, year, technology
T5	House, construction, government, employee, living, provinces, macau, anhui, yuan
T6	Gas, company, energy, usa, russia, france, chongqing, resource, china, economy, oil

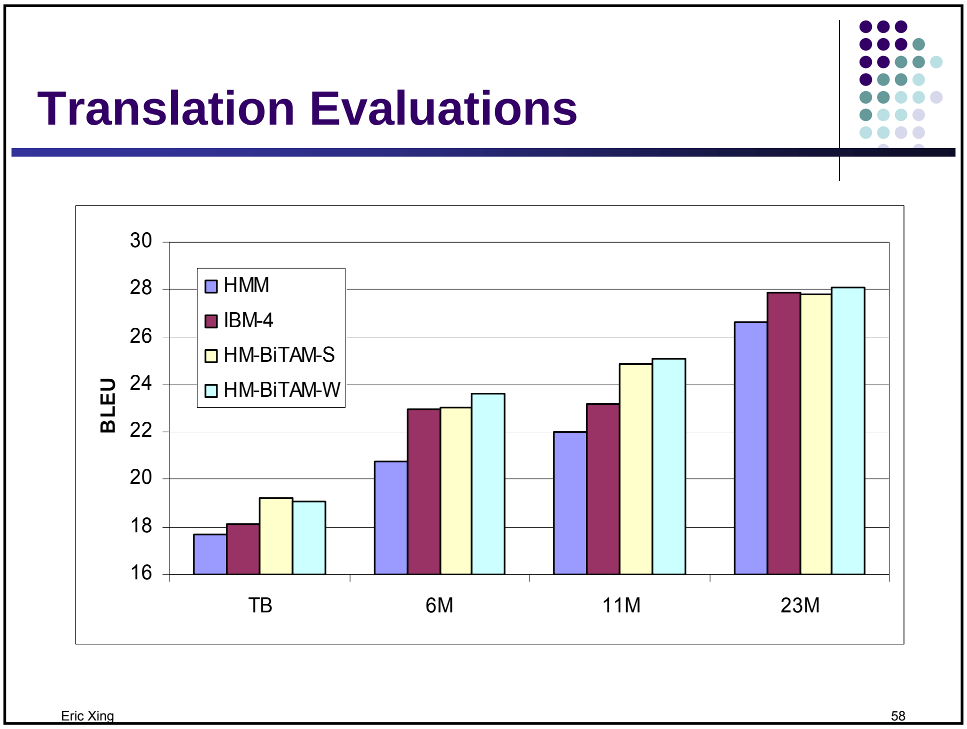
T1	人, 残疾, 体育, 事业, 水, 世界, 区, 新华社, 队员, 记者
T2	深圳, 深, 新, 元, 有, 股, 香港, 国有, 外资, 新华社
T3	国家, 重庆, 市, 区, 厂 , 天津, 政府, 项目, 国, 深圳
T4	香港, 贸易, 出口, 外资, 合作, 今年, 项目, 利用, 新, 技术
T5	住房, 房, 九江, 建设, 澳门, 元, 职工, 目前, 国家, 占, 省
T6	公司, 天然气, 两, 国, 美国, 记者, 关系, 俄, 法, 重庆

Comparison



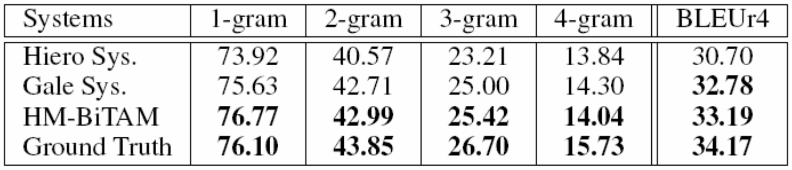


HM-BiTAM versus others



Translation Evaluations

Translation Evaluations



Systems	1-gram	2-gram	3-gram	4-gram	BLEUr4
Hiero Sys.	73.92	40.57	23.21	13.84	30.70
Gale Sys.	75.63	42.71	25.00	14.30	**32.78**
HM-BiTAM	**76.77**	**42.99**	**25.42**	**14.04**	**33.19**
Ground Truth	**76.10**	**43.85**	**26.70**	**15.73**	**34.17**

Application 3: video representation/classification

- Video: a complex, multi-modal data type for representation and classification
  - Image, text (closed-captions, speech transcript), audio
- Goal: classify video segments called video shots into semantic categories

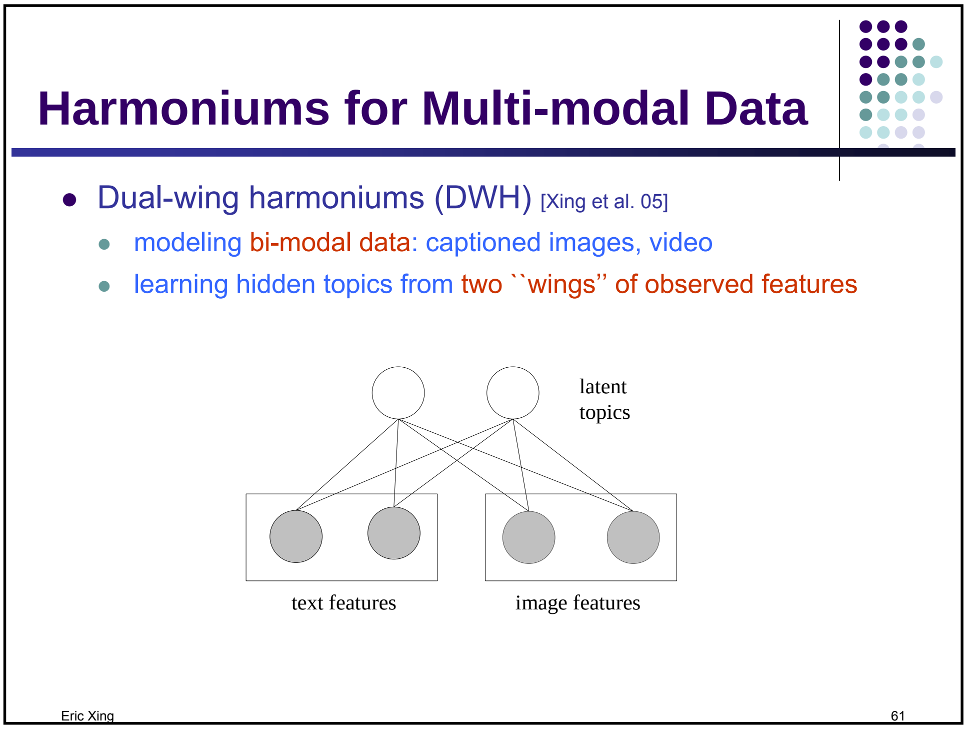
anchor

building

meeting

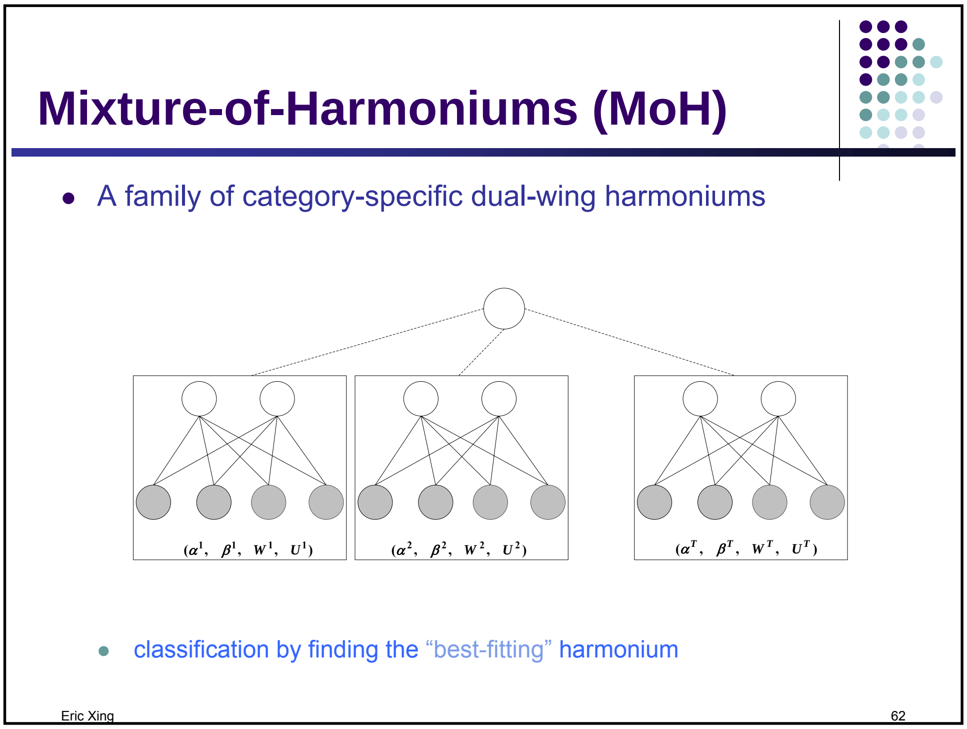
speech

J. Yang, Y. Liu, E. P. Xing and A. Hauptmann, SDM 2007, **BEST PAPER Award**



Harmoniums for Multi-modal Data

- Dual-wing harmoniums (DWH) [Xing et al. 05]
  - modeling bi-modal data: captioned images, video
  - learning hidden topics from two ``wings'' of observed features



Mixture-of-Harmoniums (MoH)

- A family of category-specific dual-wing harmoniums

$(\alpha^1, \beta^1, W^1, U^1)$
$(\alpha^2, \beta^2, W^2, U^2)$
$(\alpha^T, \beta^T, W^T, U^T)$

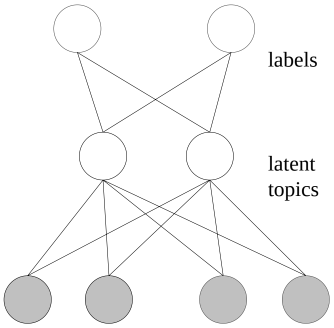
  - classification by finding the "best-fitting" harmonium

$H_1 \quad \cdots$

$\cdots \quad X_N$

Hierarchical Harmonium (HH)

- Incorporate category labels as a layer of hidden nodes on top of latent topic nodes



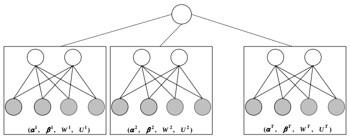
- classification by inference of label nodes

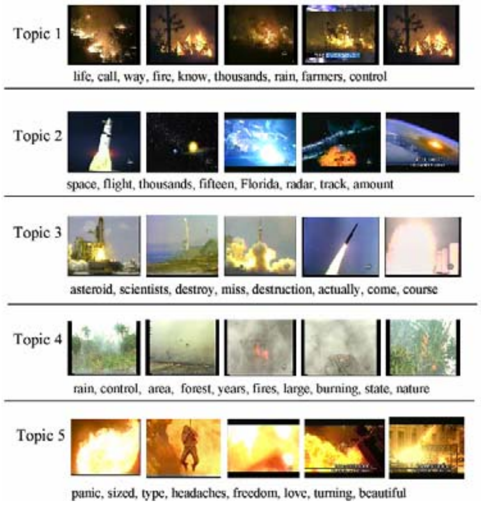
Y_1 · · ·

Semantic Topics by FoH

- Revealing "sub-topics'' of each category

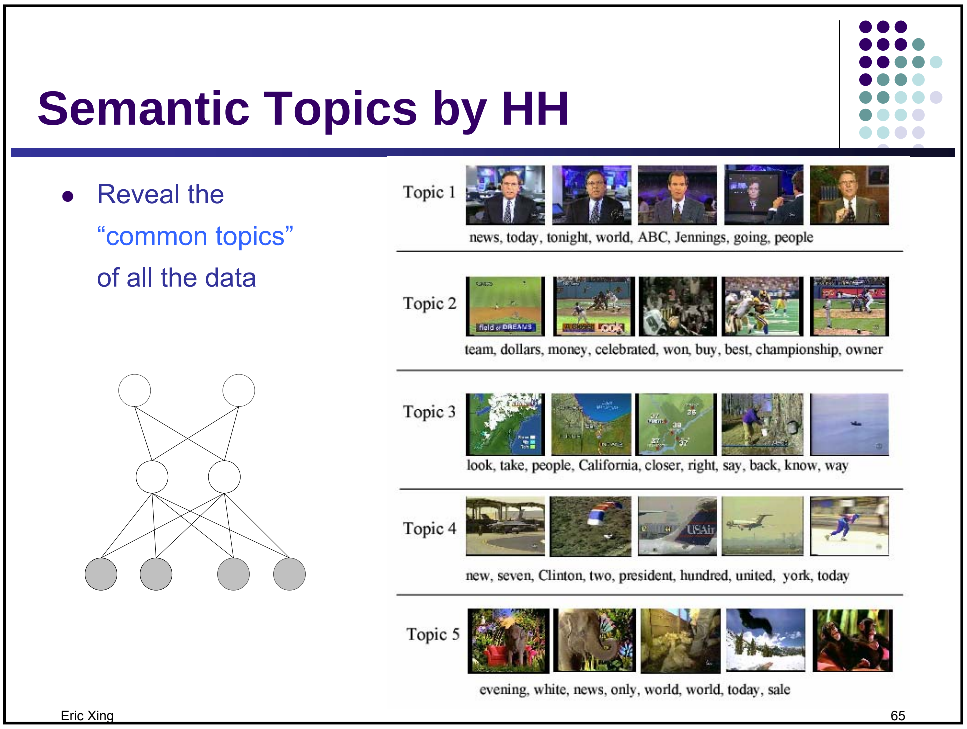
- Co-clusters of both text and image features


$(\alpha^1, \beta^1, W^1, U^1)$
$(\alpha^2, \beta^2, W^2, U^2)$
$(\alpha^T, \beta^T, W^T, U^T)$


Topic 1
life, call, way, fire, know, thousands, rain, farmers, control
Topic 2
space, flight, thousands, fifteen, Florida, radar, track, amount
Topic 3
asteroid, scientists, destroy, miss, destruction, actually, come, course
Topic 4
rain, control, area, forest, years, fires, large, burning, state, nature
Topic 5
panic, sized, type, headaches, freedom, love, turning, beautiful

H_1 · ·

X_1 · · · X_N



Semantic Topics by HH

- Reveal the "common topics" of all the data

Topic 1

news, today, tonight, world, ABC, Jennings, going, people

Topic 2

team, dollars, money, celebrated, won, buy, best, championship, owner

Topic 3

look, take, people, California, closer, right, say, back, know, way

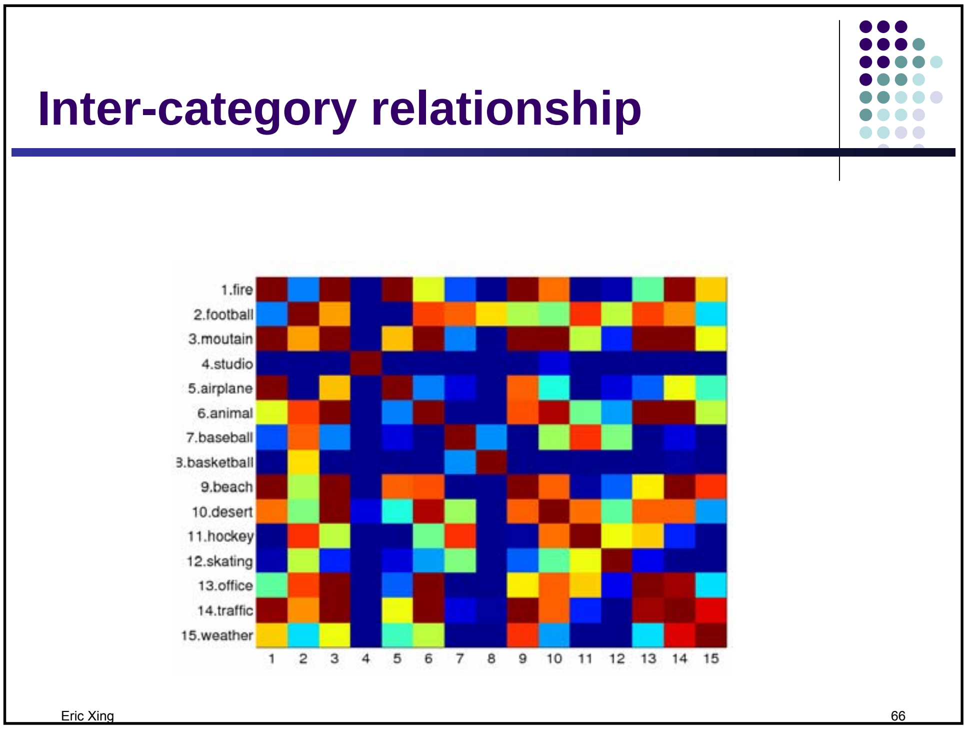
Topic 4

new, seven, Clinton, two, president, hundred, united, york, today

Topic 5

evening, white, news, only, world, world, today, sale

Eric Xing

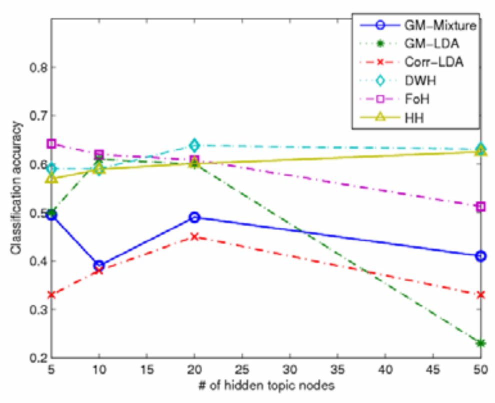


Inter-category relationship

Eric Xing

Classification Accuracy

- Harmonium models outperform directed models (e.g., LDA)



Conclusion

- GM-based topic models are cool
  - Flexible
  - Modular
  - Interactive
- There are many ways of implementing topic models
  - Directed
  - Undirected
- Efficient Inference/learning algorithms
  - GMF, with Laplace approx. for non-conjugate dist.
  - MCMC
- Many applications
  - …
  - Word-sense disambiguation
  - Word-net
  - Network inference