The Author-Topic Model

by M Rosen-Zvi, et al. (UAI 2004)

Presented by:
Darshan Ramakant Bhat
Freeze Francis
Mohammed Haroon D

Overview

- Motivation
- Model Formulation
  - Topic Model (i.e LDA)
  - Author Model
  - Author Topic Model
- Parameter Estimation
  - Gibbs Sampling Algorithms
- Results and Evaluation
- Applications

Motivation

- joint author-topic modeling had received little attention.
- author modeling has tended to focus on problem of predicting an author given the document.
- modelling the interests of authors is a fundamental problem raised by a large corpus (eg: NIPS dataset).
- this paper introduced a generative model that simultaneously models the content of documents and the interests of authors.

Notations

- Vocabulary set size : **V**
- Total unique authors : **A**
- Document **d** is identified as $(\mathbf{w}_d, \mathbf{a}_d)$
- $\mathbf{N}_d$: number of words in Document **d**.
- $\mathbf{w}_d$: Vector of $\mathbf{N}_d$ words (subset of vocabulary)

$$\mathbf{w}_d = [\mathbf{w}_{1d}\ \mathbf{w}_{2d}\ \ \ldots\ldots\ \ \mathbf{w}_{N_d d}]$$

- $\mathbf{w}_{id}$: $\mathbf{i}^{th}$ word in document **d**
- $\mathbf{A}_d$: number of authors of Document d
- $\mathbf{a}_d$: Vector of $\mathbf{A}_d$ authors (subset of authors)

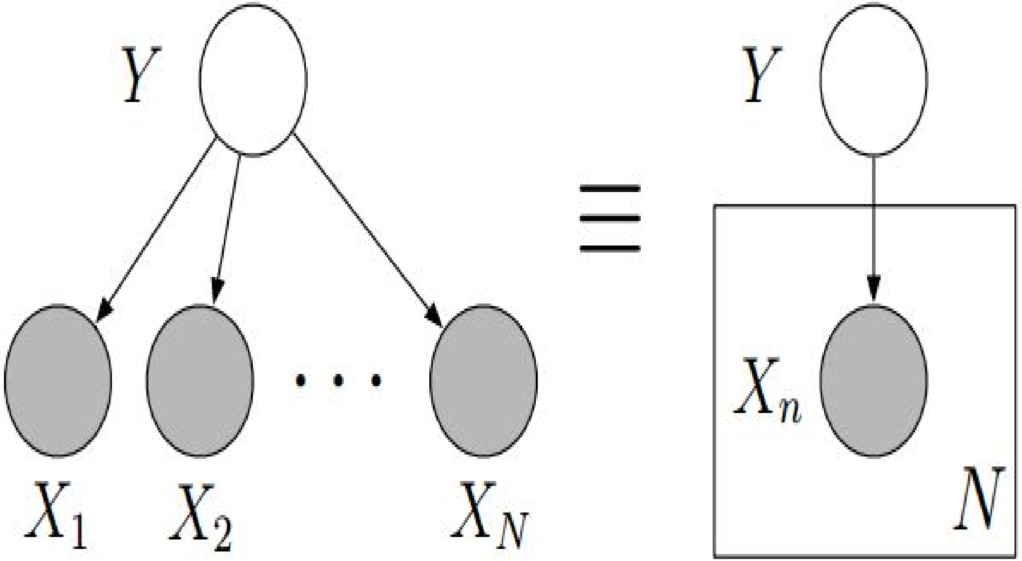
$$\mathbf{a}_d = [\mathbf{a}_{1d}\ \mathbf{a}_{2d}\ \ \ldots\ldots\ \ \mathbf{a}_{A_d d}]$$

- Corpus $\mathfrak{D}$ **:** Collection **D** documents

$$\mathfrak{D} = \{ (\mathbf{w}_1, \mathbf{a}_1), (\mathbf{w}_2, \mathbf{a}_2) \ldots\ldots (\mathbf{w}_D, \mathbf{a}_D) \}$$

Plate Notation for Graphical models

- Nodes are random variables
- Edges denote possible dependence
- Observed variables are shaded
- Plates denote replicated structure

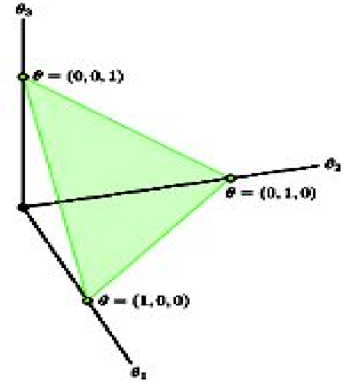


Dirichlet distribution

- generalization of beta distribution into multiple dimensions

$$f(x_1,\cdots,x_K;\alpha_1,\cdots,\alpha_K)=\frac{1}{\mathrm{B}(\boldsymbol{\alpha})}\prod_{i=1}^{K}x_i^{\alpha_i-1}$$

- Domain : $\mathbf{X} = (\mathbf{x_1, x_2 \ldots\ldots x_k})$
  - $\mathbf{x_i \in (0\ ,\ 1)}$
  - $\mathbf{\Sigma_k\, x_i = 1}$
  - **k**-dimensional multinomial distributions
- "***distribution over multinomial distributions***"
- Parameter : $\boldsymbol{\alpha} = (\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2} \ldots\ldots \boldsymbol{\alpha_k})$
  - $\boldsymbol{\alpha_i} > 0\ ,\ \forall\, i$
  - determines how the probability mass is distributed
  - Symmetric : $\boldsymbol{\alpha_1} = \boldsymbol{\alpha_2} = \ldots\ldots \boldsymbol{\alpha_k}$
- Each sample from a dirichlet is a multinomial distribution.
  - **α ↑ :** denser samples
  - **α ↓ :** sparse samples



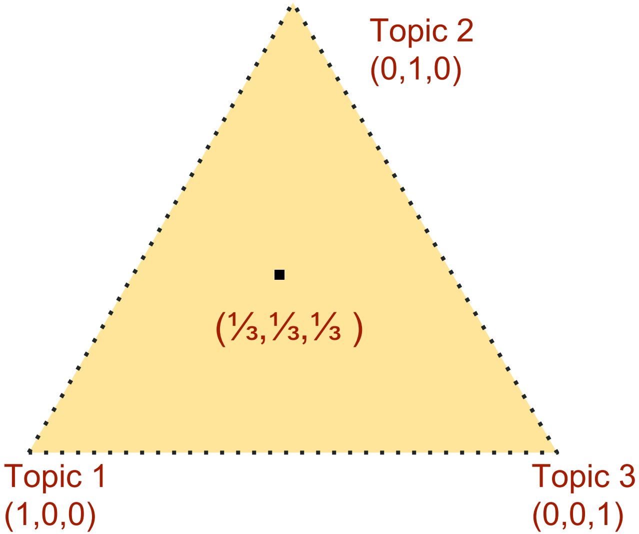


Fig : Domain of Dirichlet (k=3)

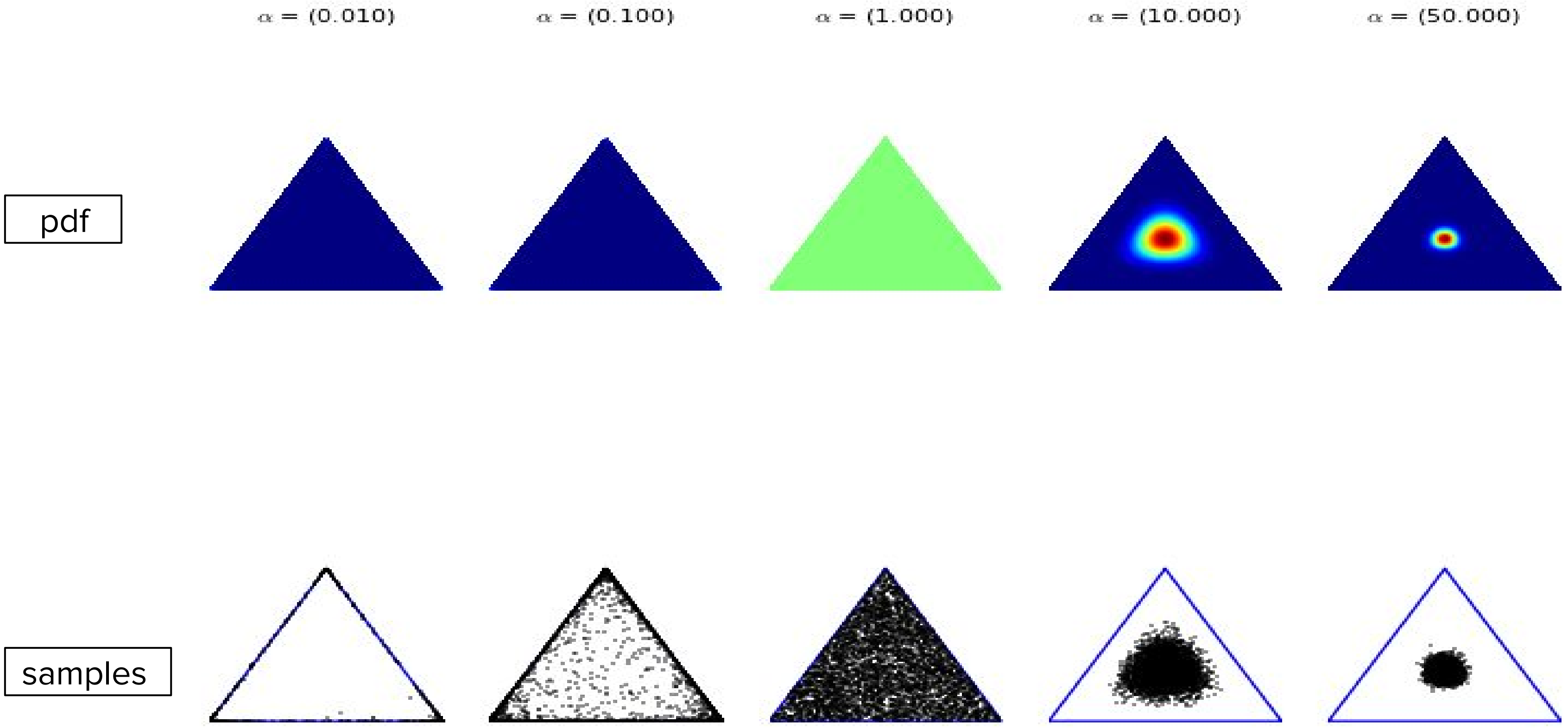


Fig : Symmetric Dirichlet distributions for k=3

Model Formulation

Topic Model (LDA)

Model topics as distribution over words
Model documents as distribution over topics

Author Model

Model author as distribution over words

Author-Topic Model

Probabilistic model for both author and topics
Model topics as distribution over words
Model authors as distribution over topics

Topic Model : LDA

- Generative model
- Bag of word
- Mixed membership
- Each word in a document has a topic assignment.
- Each document is associated with a distribution over topics.
- Each topic is associated with a distribution over words.
- **α** : Dirichlet prior parameter on the per-document topic distributions.
- **β** : Dirichlet prior parameter on the per-topic word distribution.
- **w** : observed word in document
- **z :** is the topic assigned for **w**

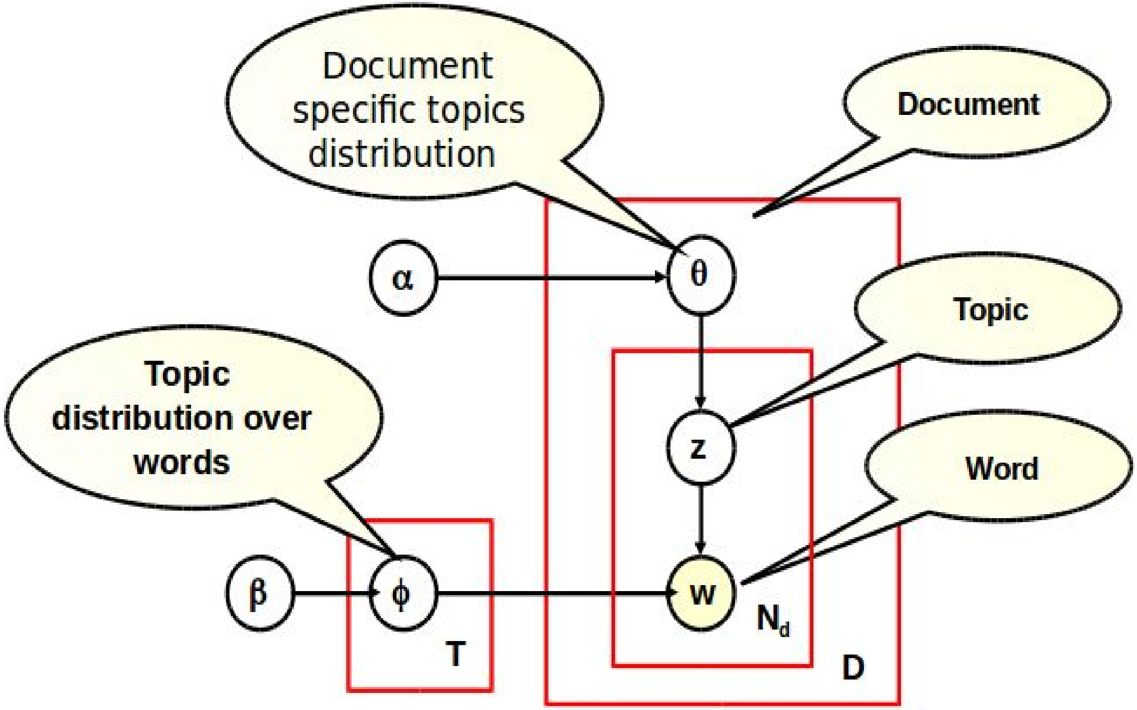


Fig : LDA model

Author Model

- Simple generative model
- Each author is associated with a distribution over words.
- Each word has a author assignment.
- **β** : Dirichlet prior parameter on the per-author word distributions.
- **w :** Observed word
- **x :** is the author chosen for word **w**

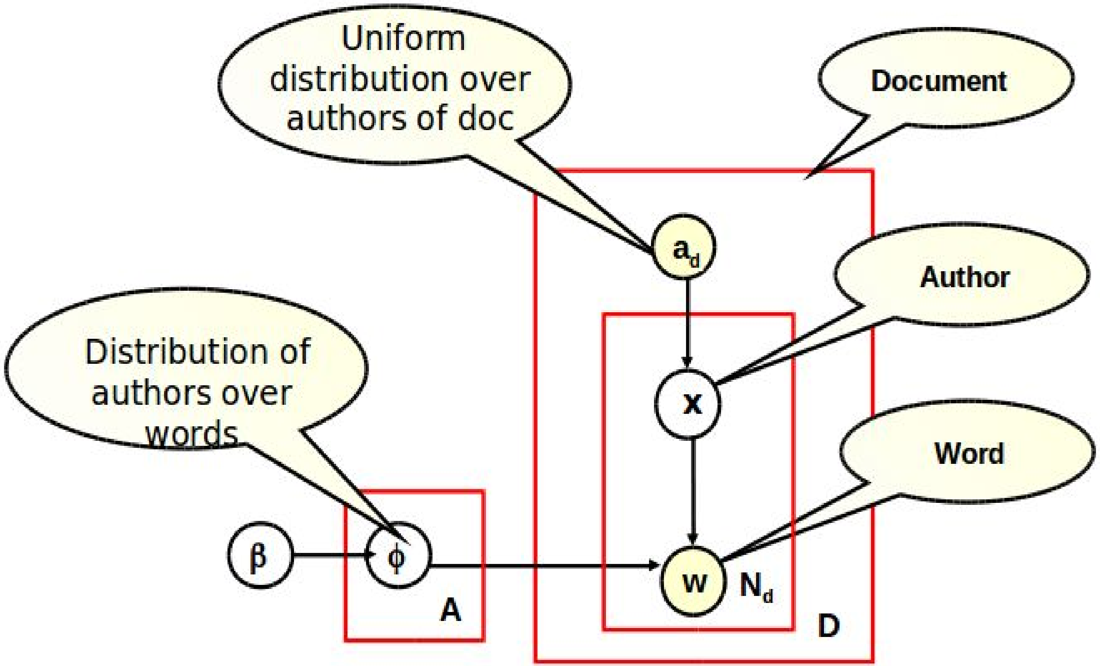
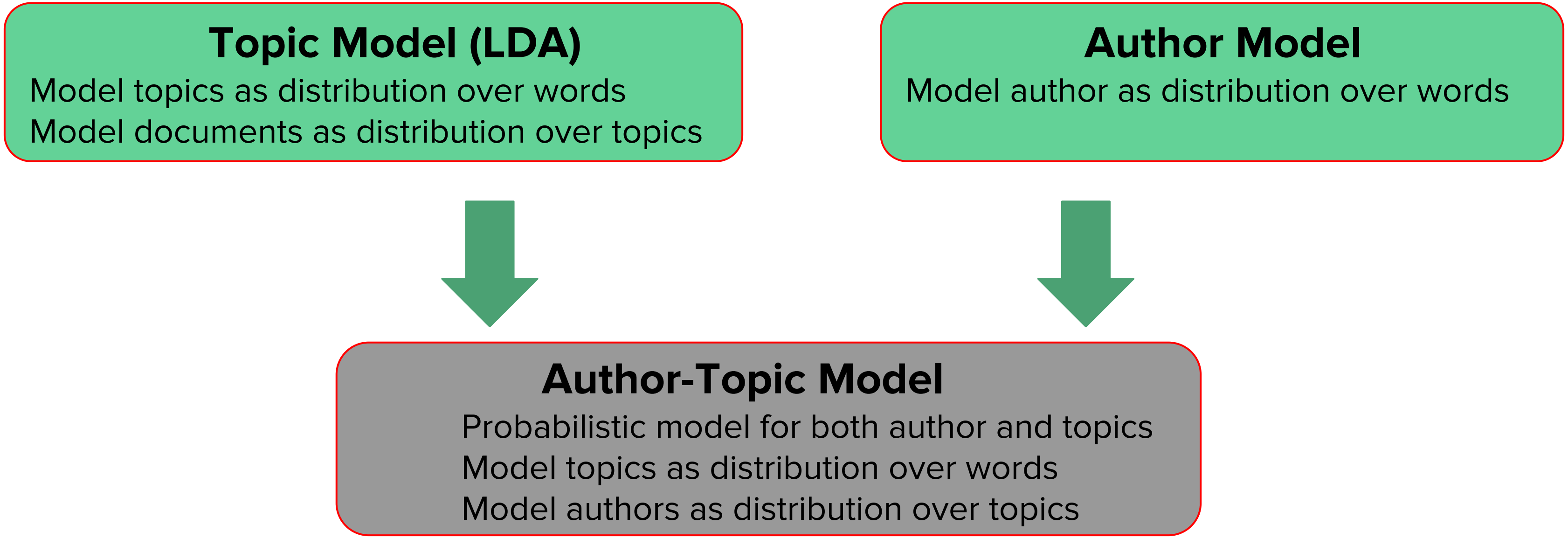


Fig : Author Model

Model Formulation



Author-Topic Model

- Each topic is associated with a distribution over words.
- Each author is associated with a distribution over topics.
- Each word in a document has an author and a topic assignment.
- Topic distribution of document is a mixture of its authors topic distribution
- **α** : Dirichlet prior parameter on the per-author topic distributions.
- **x** : author chosen for word **w**
- **z :** is the topic assigned for **w** from author **x** topic distribution
- **w** : observed word in document

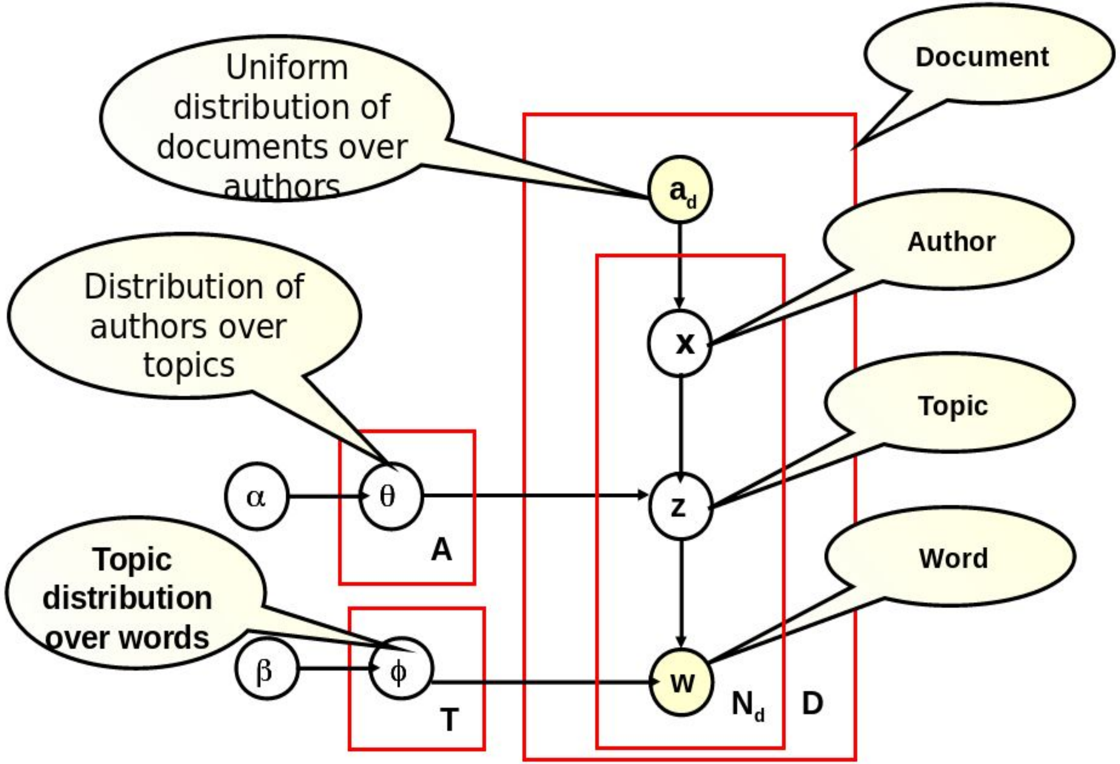


Fig : Author Topic Model plate notation

Author-Topic Model

	1	2	3	.
epilepsy	0.4	0.01	0.1	.
dynamic	0.1	0.18	0.3	.
bayesian	0.03	0.37	0.25	.
EEG	0.4	0.01	0.15	.
model	0.07	0.43	0.2	.
.	.	.	.	.

ϕ_T **(words X topics)**

	1	2	3	4
author1	0.02	0.69	0.08	.
author2	0.58	0.02	0.18	.
author3	0.23	0.12	0.45	.
author4	0.35	0.4	0.2	.
author5	0.15	0.2	0.1	.
author6	.	.	.	.

θ_{AT} **(author X topic)**

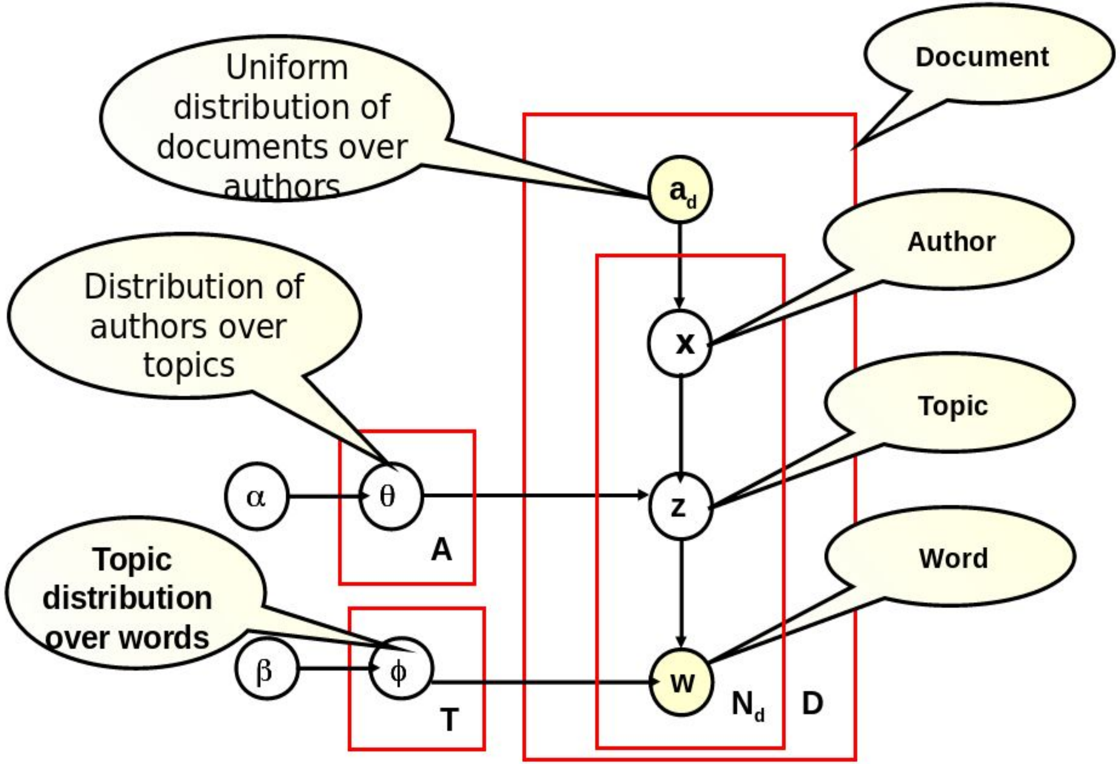


Fig : Author Topic model

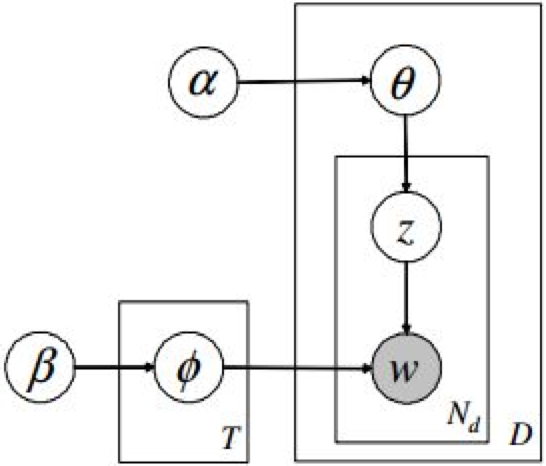
Special Cases of ATM

Topic Model (LDA):

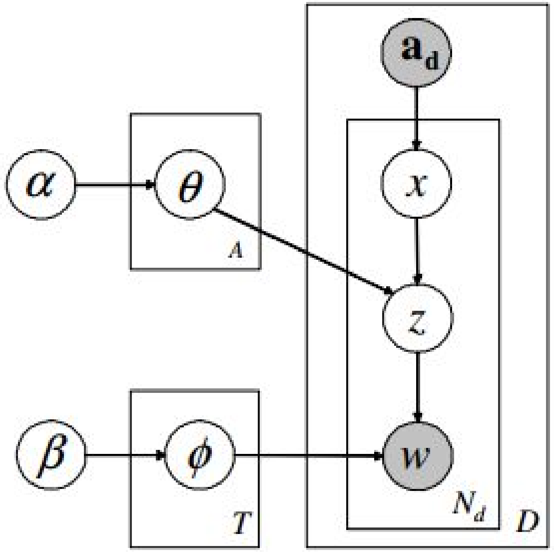
Each document is written by a unique author

- the document determines the author.
- author assignment becomes trivial.
- author's topic distribution can be viewed as document's topic distribution.

Topic (LDA)



Author-Topic

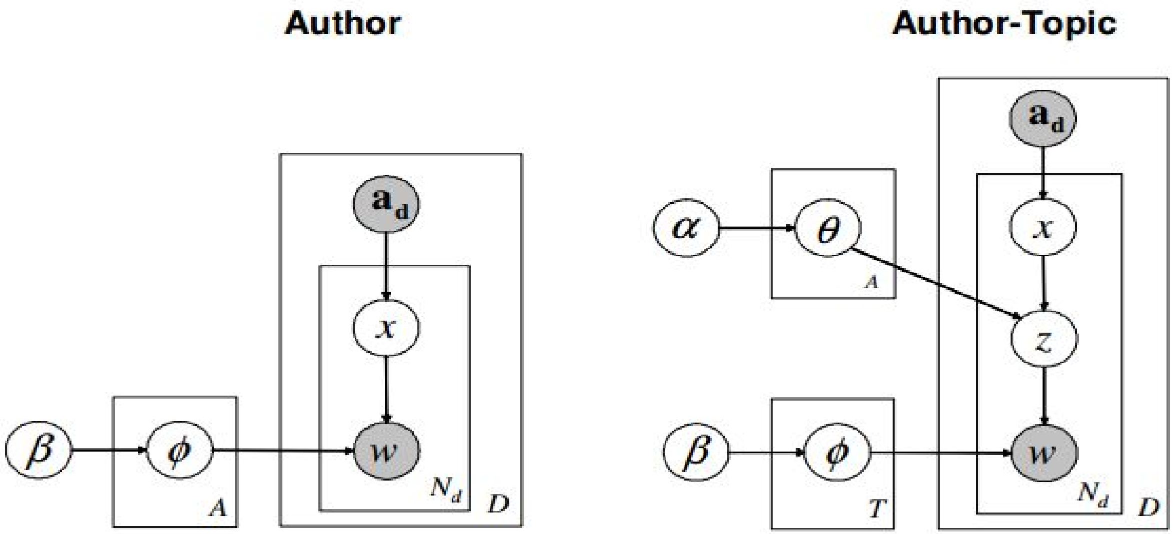


Special Cases of ATM

Author Model:

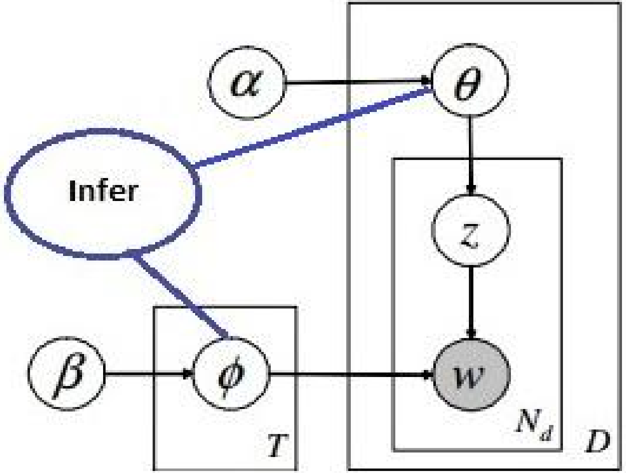
each author writes about a unique topic

- the author determines the topic.
- topic assignment becomes trivial.
- topic's word distribution can be viewed as author's word distribution.



Variables to be inferred

LDA Topic Model



	1	2	3
epilepsy	0.4	0.01	0.1
dynamic	0.1	0.18	0.3
bayesian	0.03	0.37	0.25
EEG	0.4	0.01	0.15
model	0.07	0.43	0.2

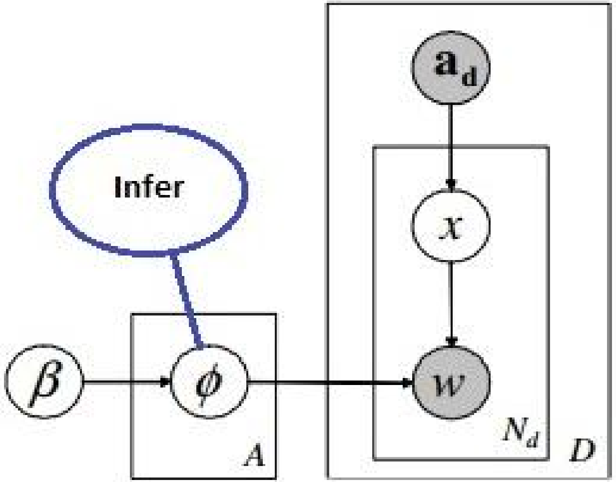
ϕ_T **(words X topics)**

	1	2	3
1	0.6	0.1	0.3
2	0.69	0.01	0.3
3	0.1	0.89	0.01
4	0.01	0.9	0.09
5	.	.	.
6	.	.	.
7	.	.	.
8	.	.	.
9	.	.	.
10	0.1	0.7	0.2

θ_T **(documents X topics)**

Variables to be inferred

Author Model

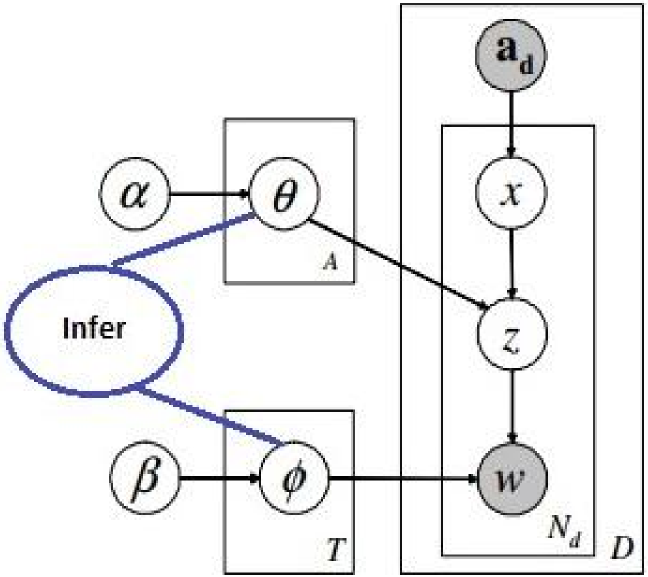


	epilepsy	dynamic	bayesian	EEG	model	...
author1	0.2	0.4	0.08	0.2	0.04	.
author2	0.04	0.2	0.4	0.02	0.3	.
author3	0.2	0.03	0.25	0.04	0.12	.
author4	.	.	.	.	.	.

ϕ_A**(author X words)**

Variables to be inferred

Author - Topic Model



	1	2	3	.
epilepsy	0.4	0.01	0.1	.
dynamic	0.1	0.18	0.3	.
bayesian	0.03	0.37	0.25	.
EEG	0.4	0.01	0.15	.
model	0.07	0.43	0.2	.
.	.	.	.	.

ϕ_T **(words X topics)**

	1	2	3	4
author1	0.02	0.69	0.08	.
author2	0.58	0.02	0.18	.
author3	0.23	0.12	0.45	.
author4	0.35	0.4	0.2	.
author5	0.15	0.2	0.1	.
author6	.	.	.	.

θ_{AT}**(author X topic)**

Collapsed Gibbs Sampling - LDA

- The assignment variables contain information about ϕ and θ.
- Use this information to directly update the Topic assignment variable with the next sample. This is Collapsed Gibbs Sampling.

$$P(z_i = j | w_i = m, \mathbf{z}_{-i}, \mathbf{w}_{-i}) \propto \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{m'j}^{WT} + V\beta} \frac{C_{dj}^{DT} + \alpha}{\sum_{j'} C_{dj'}^{DT} + T\alpha}$$

$$\phi_{mj} = \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{m'j}^{WT} + V\beta}$$

times word 'm' assigned to doc 'j' by
total #words assigned to doc 'j'

$$\theta_{dj} = \frac{C_{dj}^{DT} + \alpha}{\sum_{j'} C_{dj'}^{DT} + T\alpha}$$

times topic 'j' assigned in doc 'd' by
total #words in the doc 'd'

Collapsed Gibbs Sampling - Author Model

- Need to infer Author - word distribution ϕ_A
- WIth Collapsing GIbbs sampling, directly infer author word assignments

$$P\left(x_i = k | w_i = m, \mathbf{x}_{-i}, \mathbf{w}_{-i}, \mathbf{a}_d\right) \propto \frac{C_{mk}^{WA} + \beta}{\sum_{m'} C_{m'k}^{WA} + V\beta}$$

$$\phi_{mk} = \frac{C_{mk}^{WA} + \beta}{\sum_{m'} C_{m'k}^{WA} + V\beta}$$

#times word 'm' assigned to author 'k' by
total #words assigned to author 'k'

Author Topic Model - Collapsed Gibbs Sampling

- Need to infer ϕ_T and Θ_A
- Instead through collapsed Gibbs Sampling directly infer Topic and Author assignments
- By finding the author - topic joint distribution conditioned on all other variables

$$P\left(z_i = j, x_i = k \middle| w_i = m, \mathbf{z}_{-i}, \mathbf{x}_{-i}, \mathbf{w}_{-i}, \mathbf{a}_d\right) \propto \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{m'j}^{WT} + V\beta} \frac{C_{kj}^{AT} + \alpha}{\sum_{j'} C_{kj'}^{AT} + T\alpha}$$

$$\phi_{mj} = \frac{C_{mj}^{WT} + \beta}{\sum_{m'} C_{m'j}^{WT} + V\beta}$$

# of times word 'm' assigned to topic 'j' by total # of assignments of topic 'j'

$$\theta_{kj} = \frac{C_{kj}^{AT} + \alpha}{\sum_{j'} C_{kj'}^{AT} + T\alpha}$$

# of times author 'k' assigned topic 'j' by total # of assignments of author 'k'

Example

AUTHOR	1	2	2	1	1
TOPIC	3	2	1	3	1
WORD	epilepsy	dynamic	bayesian	EEG	model

	TOPIC-1	TOPIC-2	TOPIC-3
epilepsy	1	0	35
dynamic	18	20	1
bayesian	42	15	0
EEG	0	5	20
model	10	8	1
...			

	TOPIC-1	TOPIC-2	TOPIC-3
AUTHOR-1	20	14	2
AUTHOR-2	20	21	5
AUTHOR-3	11	8	15
AUTHOR-4	5	15	14
...			

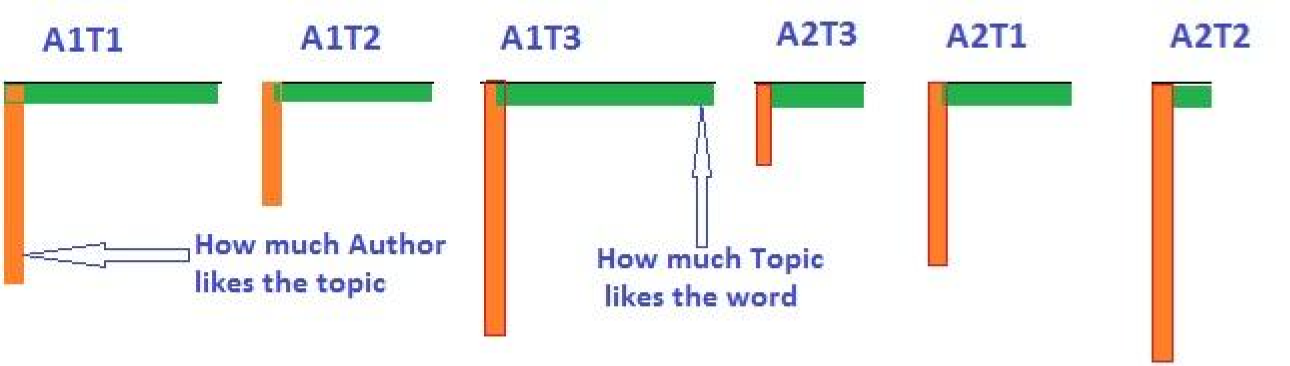
Example

AUTHOR	1	2	?	1	1
TOPIC	3	2	?	3	1
WORD	epilepsy	dynamic	bayesian	EEG	model

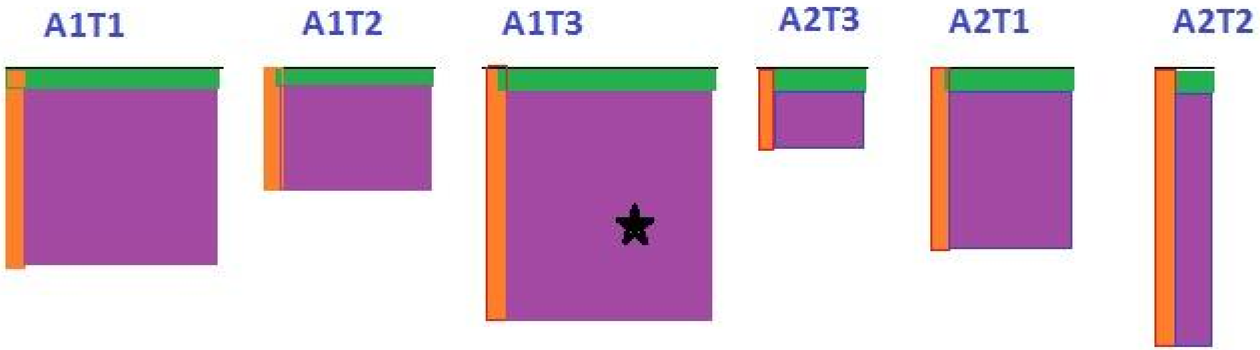
	TOPIC-1	TOPIC-2	TOPIC-3
epilepsy	1	0	35
dynamic	18	20	1
bayesian	42(41)	15	0
EEG	0	5	20
model	10	8	1
...			

	TOPIC-1	TOPIC-2	TOPIC-3
AUTHOR-1	20	14	2
AUTHOR-2	20(19)	21	5
AUTHOR-3	11	8	15
AUTHOR-4	5	15	14
...			

Author-Topic	How much topic ‘j’ likes the word ‘bayesian’	How much author ‘k’ likes the topic
A1T1	41/97	20/36
A2T1	41/97	19/45
A1T2	15/105	14/36
A2T2	...	...
A2T3	...	...
A2T3	...	...



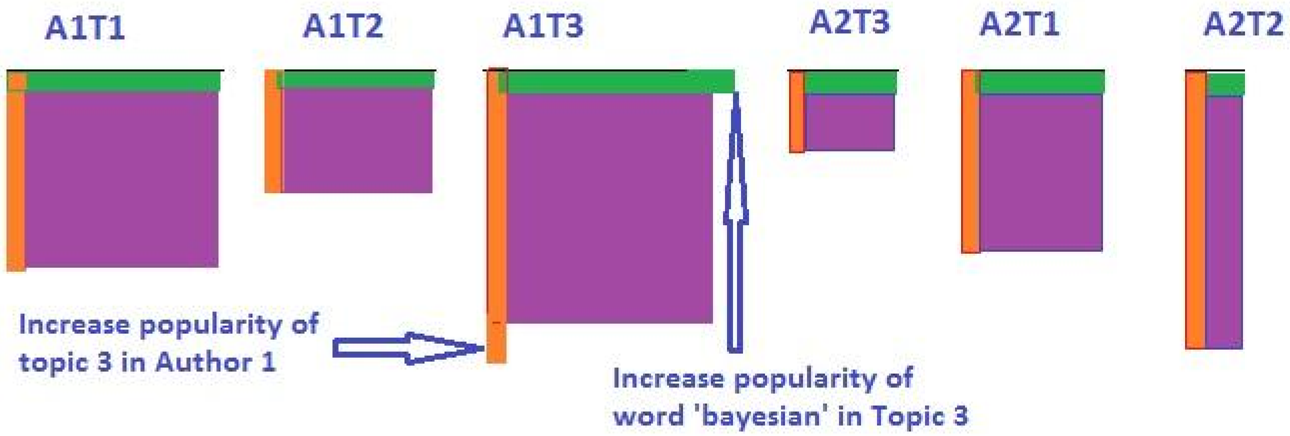
- To draw new Author-Topic assignment (equivalently)
  - Roll **K X a**$_d$ sided die with these probabilities
  - Assign the Author-Topic tuple to the word



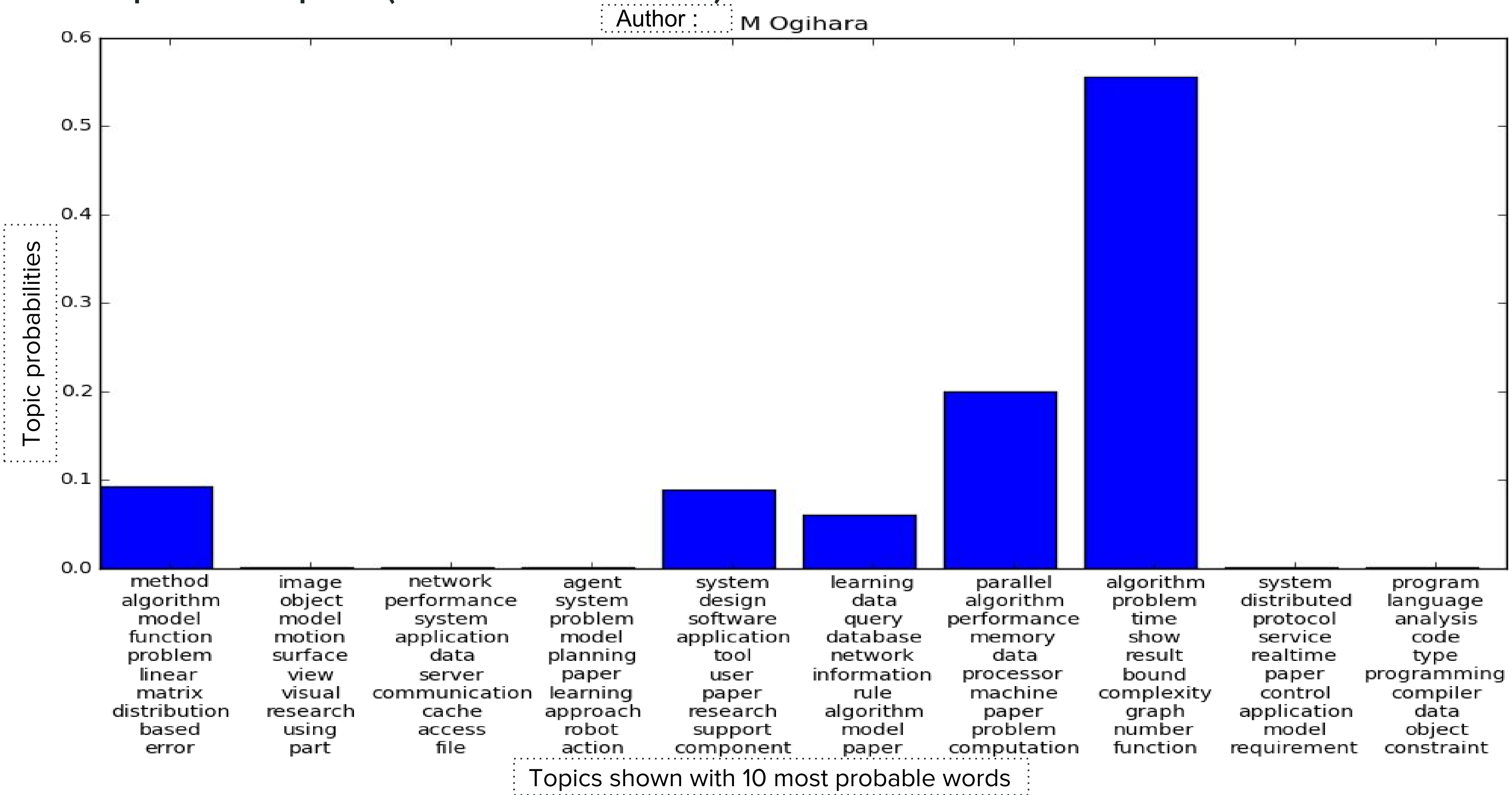
AUTHOR	1	2	1	1	1
TOPIC	3	2	3	3	1
WORD	epilepsy	dynamic	bayesian	EEG	model

Example (cont..)

- Update the probability (ϕ_T and θ_A) tables with new values
- This implies increasing the popularity of topic in author and word in topic



Sample Output (CORA dataset)



Experimental Results

- Experiment is done on two types of dataset, papers are chosen from NIPS and CiteSeer database
- Extremely common words are removed from the corpus, leading V=13,649 unique words in NIPS and V=30,799 in CiteSeer, 2037 authors in NIPS and 85,465 in CiteSeer
- NIPS contains many documents which are closely related to learning theory and machine learning
- CiteSeer contains variety of topics from User Interface to Solar astrophysics

Examples of topic author distribution

TOPIC 31	
WORD	**PROB.**
SPEECH	0.0823
RECOGNITION	0.0497
HMM	0.0234
SPEAKER	0.0226
CONTEXT	0.0224
WORD	0.0166
SYSTEM	0.0151
ACOUSTIC	0.0134
PHONEME	0.0131
CONTINUOUS	0.0129
AUTHOR	**PROB.**
Waibel_A	0.0936
Makhoul_J	0.0238
De-Mori_R	0.0225
Bourlard_H	0.0216
Cole_R	0.0200
Rigoll_G	0.0191
Hochberg_M	0.0176
Franco_H	0.0163
Abrash_V	0.0157
Movellan_J	0.0149

TOPIC 61	
WORD	**PROB.**
BAYESIAN	0.0450
GAUSSIAN	0.0364
POSTERIOR	0.0355
PRIOR	0.0345
DISTRIBUTION	0.0259
PARAMETERS	0.0199
EVIDENCE	0.0127
SAMPLING	0.0117
COVARIANCE	0.0117
LOG	0.0112
AUTHOR	**PROB.**
Bishop_C	0.0563
Williams_C	0.0497
Barber_D	0.0368
MacKay_D	0.0323
Tipping_M	0.0216
Rasmussen_C	0.0215
Opper_M	0.0204
Attias_H	0.0155
Sollich_P	0.0143
Schottky_B	0.0128

TOPIC 71	
WORD	**PROB.**
MODEL	0.4963
MODELS	0.1445
MODELING	0.0218
PARAMETERS	0.0205
BASED	0.0116
PROPOSED	0.0103
OBSERVED	0.0100
SIMILAR	0.0083
ACCOUNT	0.0069
PARAMETER	0.0068
AUTHOR	**PROB.**
Omohundro_S	0.0088
Zemel_R	0.0084
Ghahramani_Z	0.0076
Jordan_M	0.0075
Sejnowski_T	0.0071
Atkeson_C	0.0070
Bower_J	0.0066
Bengio_Y	0.0062
Revow_M	0.0059
Williams_C	0.0054

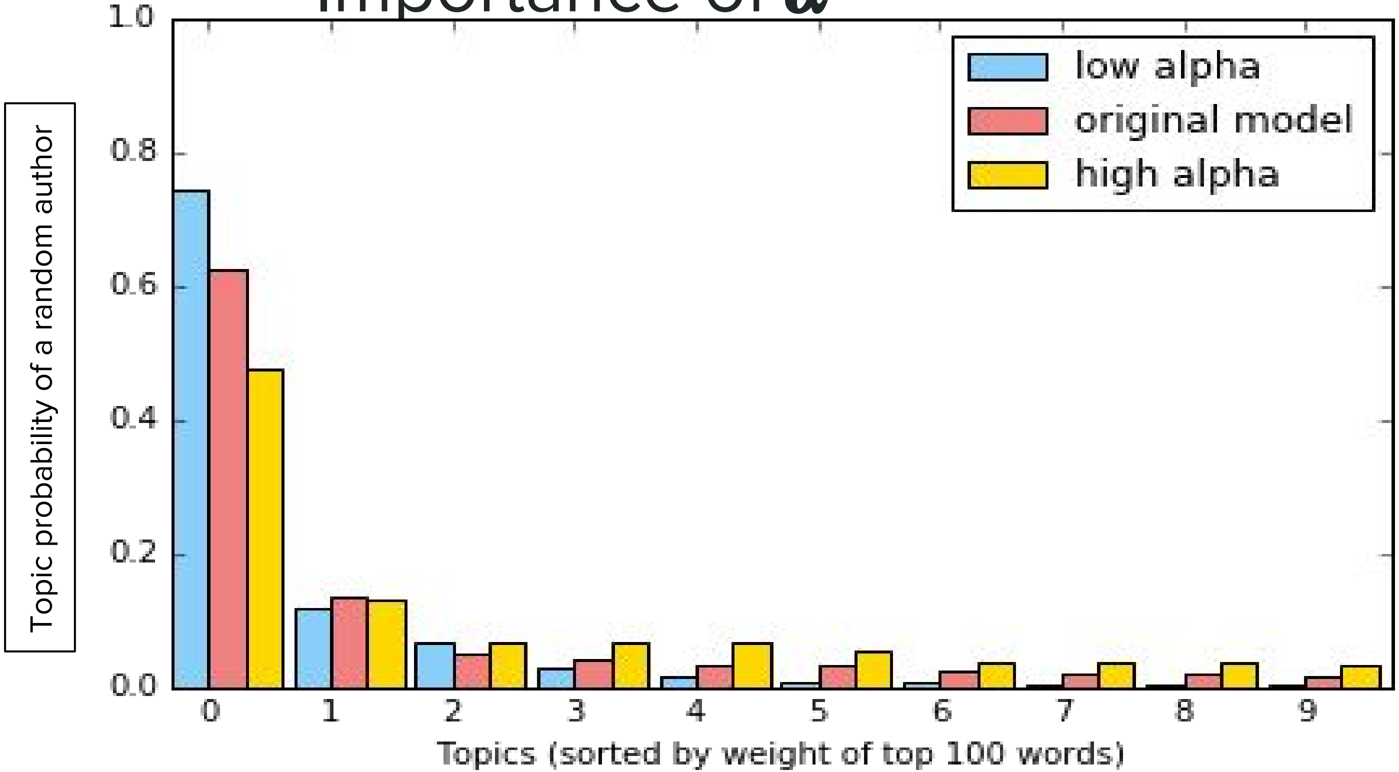
TOPIC 10	
WORD	**PROB.**
SPEECH	0.1134
RECOGNITION	0.0349
WORD	0.0295
SPEAKER	0.0227
ACOUSTIC	0.0205
RATE	0.0134
SPOKEN	0.0132
SOUND	0.0127
TRAINING	0.0104
MUSIC	0.0102
AUTHOR	**PROB.**
Waibel_A	0.0156
Gauvain_J	0.0133
Lamel_L	0.0128
Woodland_P	0.0124
Ney_H	0.0080
Hansen_J	0.0078
Renals_S	0.0072
Noth_E	0.0071
Boves_L	0.0070
Young_S	0.0069

TOPIC 209	
WORD	**PROB.**
PROBABILISTIC	0.0778
BAYESIAN	0.0671
PROBABILITY	0.0532
CARLO	0.0309
MONTE	0.0308
DISTRIBUTION	0.0257
INFERENCE	0.0253
PROBABILITIES	0.0253
CONDITIONAL	0.0229
PRIOR	0.0219
AUTHOR	**PROB.**
Friedman_N	0.0094
Heckerman_D	0.0067
Ghahramani_Z	0.0062
Koller_D	0.0062
Jordan_M	0.0059
Neal_R	0.0055
Raftery_A	0.0054
Lukasiewicz_T	0.0053
Halpern_J	0.0052
Muller_P	0.0048

TOPIC 87	
WORD	**PROB.**
USER	0.2541
INTERFACE	0.1080
USERS	0.0788
INTERFACES	0.0433
GRAPHICAL	0.0392
INTERACTIVE	0.0354
INTERACTION	0.0261
VISUAL	0.0203
DISPLAY	0.0128
MANIPULATION	0.0099
AUTHOR	**PROB.**
Shneiderman_B	0.0060
Rauterberg_M	0.0031
Lavana_H	0.0024
Pentland_A	0.0021
Myers_B	0.0021
Minas_M	0.0021
Burnett_M	0.0021
Winiwarter_W	0.0020
Chang_S	0.0019
Korvemaker_B	0.0019

TOPIC 20	
WORD	**PROB.**
STARS	0.0164
OBSERVATIONS	0.0150
SOLAR	0.0150
MAGNETIC	0.0145
RAY	0.0144
EMISSION	0.0134
GALAXIES	0.0124
OBSERVED	0.0108
SUBJECT	0.0101
STAR	0.0087
AUTHOR	**PROB.**
Linsky_J	0.0143
Falcke_H	0.0131
Mursula_K	0.0089
Butler_R	0.0083
Bjorkman_K	0.0078
Knapp_G	0.0067
Kundu_M	0.0063
Christensen-J	0.0059
Cranmer_S	0.0055
Nagar_N	0.0050

Importance of α



Evaluating the predictive power

- Predictive power of the model is measured quantitatively using perplexity.
- Perplexity is ability to predict the words on a new unseen documents
- Perplexity of set of d words in a document (w_d,a_d) is

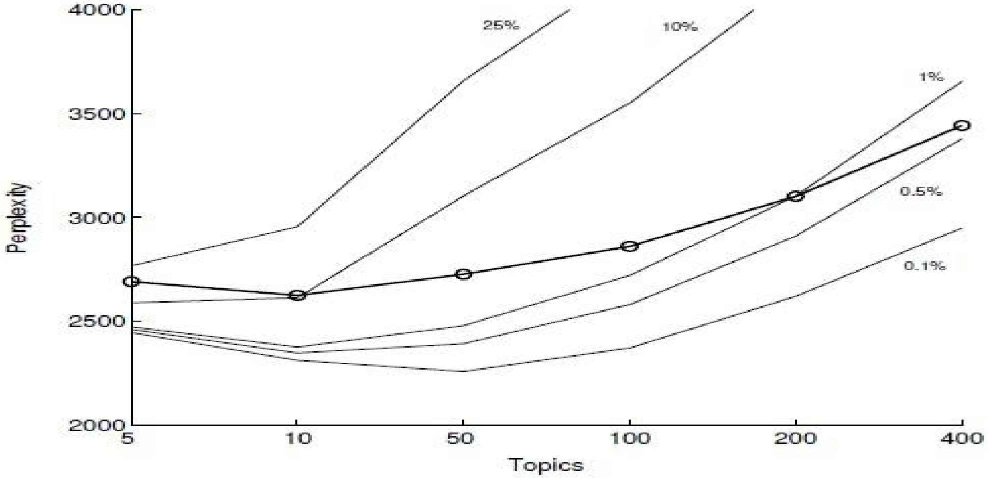
$$\text{perplexity}(\mathbf{w}_d|\mathbf{a}_d) = \exp\left[-\frac{\ln p(\mathbf{w}_d|\mathbf{a}_d)}{N_d}\right]$$

- When $p(w_d|a_d)$ is high perplexity ≈ 1, otherwise it will be a high positive quantity.

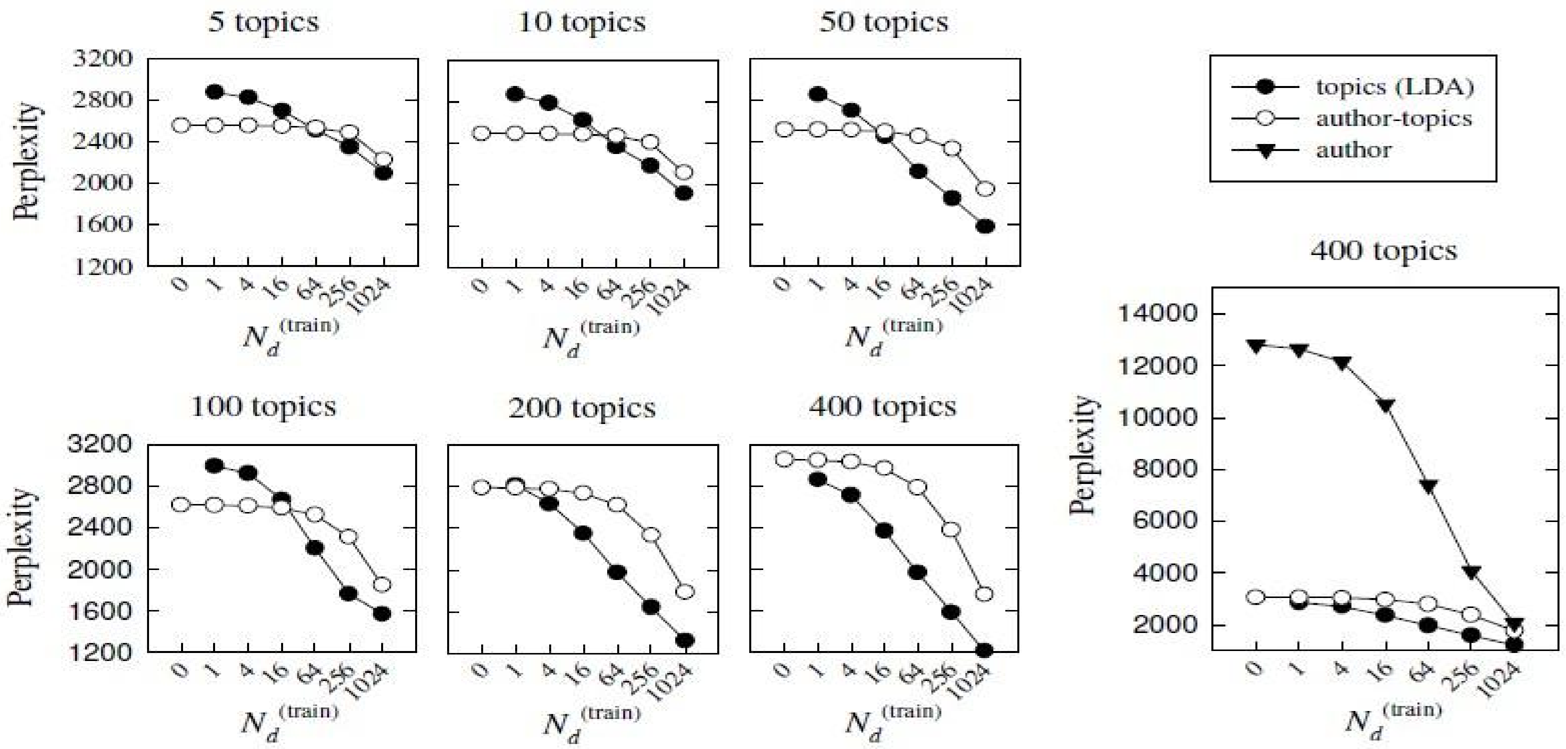
Continued....

- Approximate equation to calculate the joint probability

$$p(\mathbf{w}_d|a_d) = \frac{1}{S}\sum_{s=1}^{S}\prod_{m=1}^{N_d}\left[\sum_j \theta^s_{a_d j}\phi^s_{w_m j}\right]$$



Continued....


5 topics
10 topics
50 topics
100 topics
200 topics
400 topics
Perplexity
3200
2800
2400
2000
1600
1200
0 1 4 16 64 256 1024
$N_d^{(train)}$
topics (LDA)
author-topics
author
14000
12000
10000
8000
6000
4000
2000

Applications

- It can be used in automatic reviewer recommendation for a paper review.
- Given an abstract of a paper and a list of the authors plus their known past collaborators, generate a list of other highly likely authors for this abstract who might serve as good reviewers.

$$sKL(i,j) = \sum_{t=1}^{T} \left[\theta_{it} \log \frac{\theta_{it}}{\theta_{jt}} + \theta_{jt} \log \frac{\theta_{jt}}{\theta_{it}} \right].$$

Authors	n	T=400	T=200	T=100
Bartlett_P (8) Shawe-Taylor_J (8)	-	2.52	1.58	0.90
Barto_A (11) Singh_S (17)	2	3.34	2.18	1.25
Amari_S (9) Yang_H (5)	3	3.44	2.48	1.57
Singh_S (17) Sutton_R (7)	2	3.69	2.33	1.35
Moore_A (11) Sutton_R (7)	-	4.25	2.89	1.87
MEDIAN	-	5.52	4.01	3.33
MAXIMUM	-	16.61	14.91	13.32

Note: n is number of common papers in NIPS dataset.

References

- Rosen-Zvi, Michal, et al. "The author-topic model for authors and documents." *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004.
- D. M. Blei, A. Y. Ng, and M. I. Jordan (2003). “Latent Dirichlet Allocation”. Journal of Machine Learning Research, 3 : 993–1022.
- T. L. Griffiths, and M. Steyvers (2004). “Finding scientific topics”. Proceedings of the National Academy of Sciences, 101 (suppl. 1), 5228–5235.
- https://github.com/arongdari/python-topic-model
- https://www.coursera.org/learn/ml-clustering-and-retrieval/home/welcome


Questions?
Thank you!