Topic Models

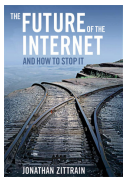
Brandon Malone

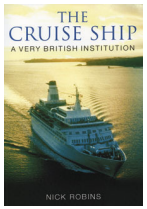
Much of this material is adapted from Blei 2003.
Many of the images were taken from the Internet

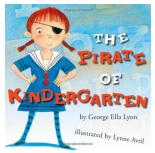
February 20, 2014

Topic Models

Suppose we have a large number of books. Each is about several unknown topics.



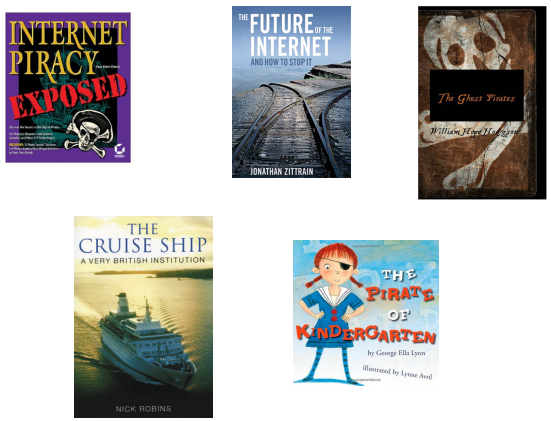




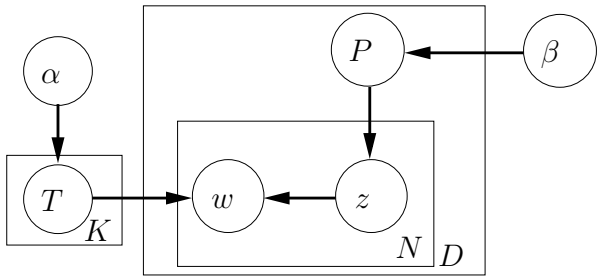
How can we tell which books are about similar (sets of) topics?

Topic Models

Suppose we have a large number of books. Each is about several unknown topics.



$\Rightarrow$



How can we tell which books are about similar (sets of) topics?

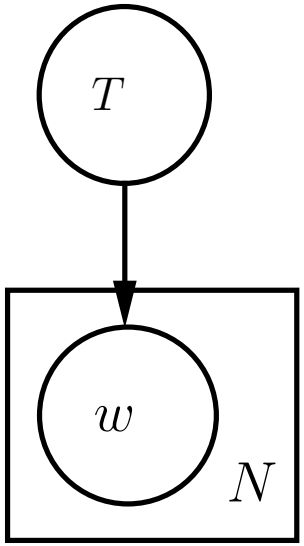
This can be described by a **topic model**, in particular, the latent Dirichlet allocation model.

Topic Model Notation

We will discuss everything in terms of text classification. Often, other domains can be mapped onto these concepts.

- T. The (multinomial) distribution over words for a particular topic
- w. A word selected to be in the document of interest
- N. The number of words in the document of interest
- K. The number of topics
- z. The topic for the word of interest
- P. The (multinomial) distribution over topics for a particular document
- D. The number of documents
- α. The parameter to a Dirichlet distribution which is a prior for T
- β. The parameter to a Dirichlet distribution which is a prior for P

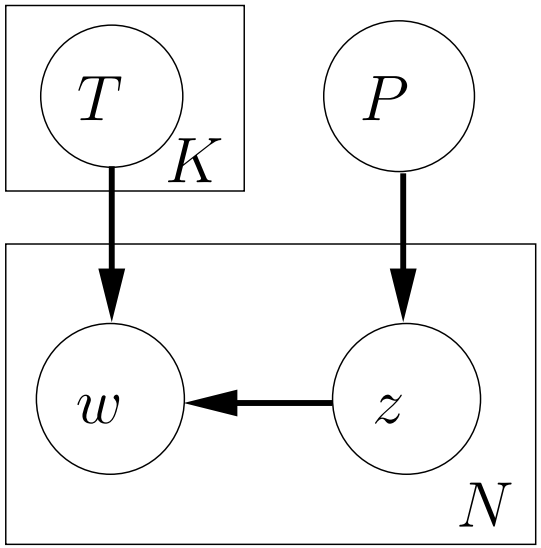
Text classification with naive Bayes



We used the naive Bayes model for a document about a single topic.

Why shouldn't we use this for a document about multiple topics?

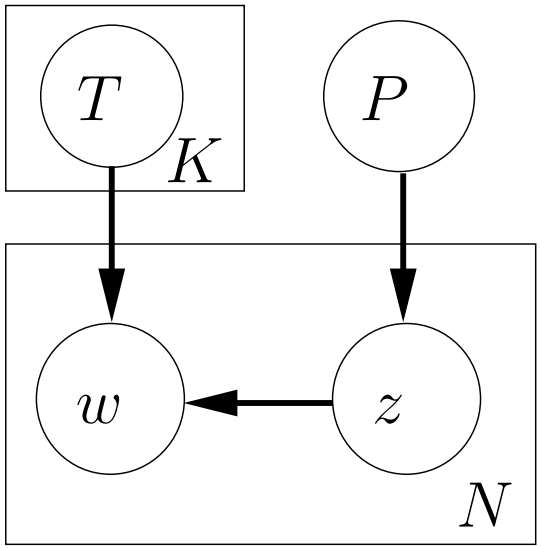
Mixtures of naive Bayes



Each topic has its own distribution over the words.

Look familiar?

Mixtures of naive Bayes

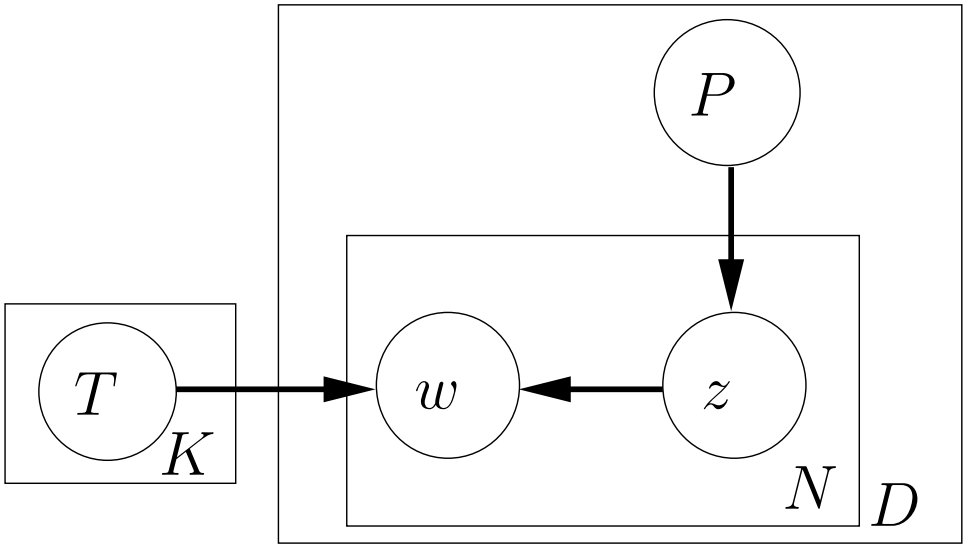


Each topic has its own distribution over the words.

Look familiar?

All mixture models look basically the same.

Extending to multiple documents

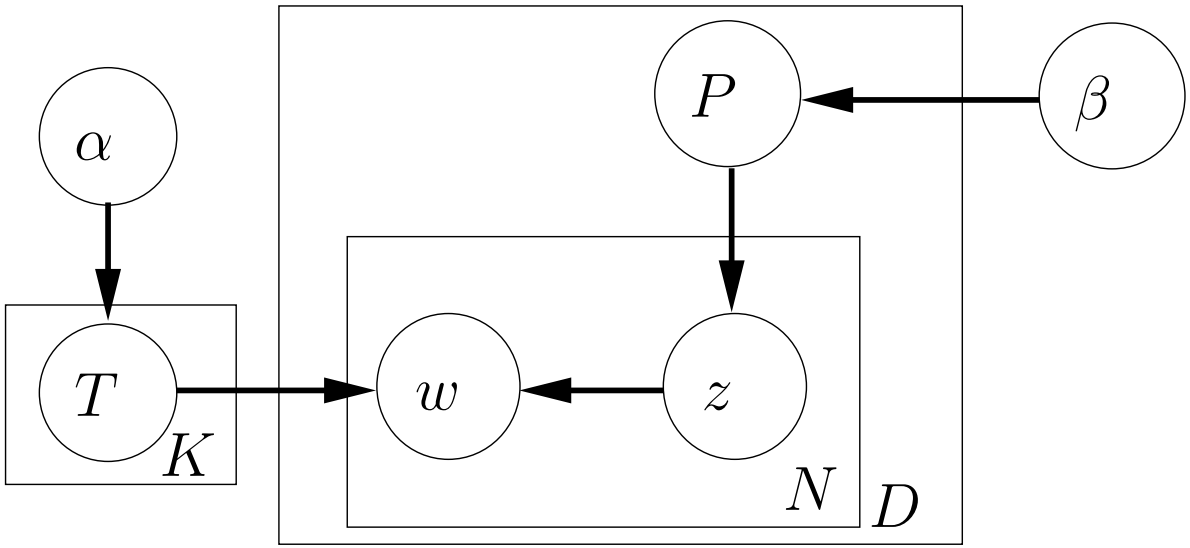


Each document now has its own distribution over topics.

We are now interested in the **joint probability** of topics and words over **all** documents.

Contrast this with how we trained naive Bayes.

Moving away from maximum likelihood



We previously saw problems with MLE parameters. They still exist.

We can add Dirichlet priors to the multinomial distributions.

This is called the **latent Dirichlet allocation** model (LDA).

Typical goals in LDA

We are often interested in the posterior distributions related to z.

- High values of $P(w|z)$, which show the words highly associated with each topic
- High values of $P(z|P)$, which shows the topics with which each document is associated

Parameter inference in LDA

Estimating the parameters in LDA (and similar models) is non-trivial.

- **Variational inference**. The network is simplified by ignoring some edges and an EM-like algorithm is used to optimize the parameters.
- **Gibbs sampling**. We randomly draw samples from the model and based on the samples we see, we update the parameters.

Extensions

- **Correlated topic models**. Some topics are more likely to occur together than others.
- **Dynamic topic models**. The word distributions may change over time.
- **Discriminitive topic models**. We may want to make class membership predictions.
- **Nonparametric models**. We may not *a priori* know how many topics there are.
- Many, many others...

Bayesian networks in environment analysis

Bayesian networks have been used to model willow seed dispersion (as a proxy for species invasion) in the Florida wetlands.

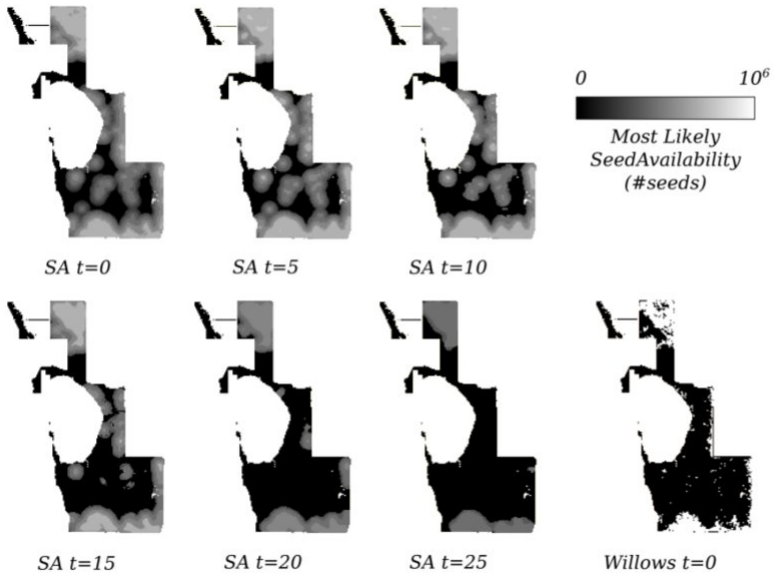


Figure 7: Seed availability predicted by the Willows ST-OODBN at $t = 0, 5, 10, 15, 20, 25$ years, across the Blue Cypress Marsh Conservation Area (138 x 208 cells). Adult willow occupancy at $t = 0$ is shown in the bottom right panel; black indicates absence, grey presence.

These results are used to help decide how much logging is allowed.

Wilkinson *et al.*, 2013

LDA in consumer behavior

LDA has been used to find groups of product relevant to particular consumers, as well as "infuencers" who impact other consumers.

Product	Prob.	Category
Chicken pasta (cream sauce)	0.0197	*Pasta*
Chicken pasta (pesto sauce)	0.0193	*Pasta*
Pork pasta (tomato sauce)	0.0167	*Pasta*
Pork steak	0.0155	*Meat*
Bacon pasta (cream sauce)	0.0153	*Pasta*
Spicy pasta (tomato sauce)	0.0149	*Pasta*
Clam garlic linguine	0.0146	*Pasta*
Tomato sauce pasta	0.0142	*Pasta*
German sausage sauce	0.0132	*Pasta*
Italian pasta (cooked)	0.0129	*Pasta*

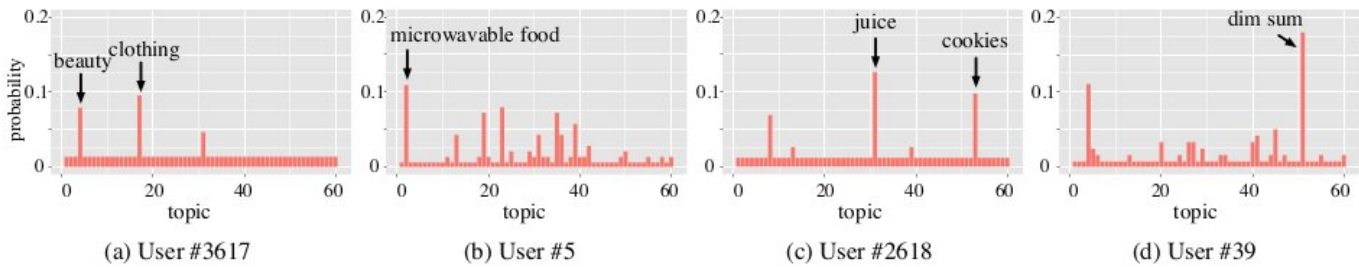
(a) Topic 1, "pasta"

Product	Prob.	Category
Ham sandwich	0.0101	*Bread*
Cheese sandwich	0.0089	*Bread*
Milk bar cookie	0.0080	*Cookie*
Cherry chocolate tart	0.0078	*Cake*
Cheese roll	0.0077	*Cake*
Cheese almond tart	0.0074	*Cake*
Taro toast	0.0073	*Bread*
Creme Brulee	0.0073	*Cake*
Raisin toast	0.0071	*Bread*
Wheat ham sandwich	0.0070	*Bread*

(b) Topic 18, "bread and cakes"

Product	Prob.	Category
Knit Hat	0.0169	*Accessory*
Knit Scarf	0.0165	*Accessory*
Legging	0.0133	*Clothing*
Wool scarf	0.0120	*Accessory*
Long Pant	0.0111	*Clothing*
Cotton Socks	0.0099	*Accessory*
Wool Gloves	0.0097	*Accessory*
Facial Masks	0.0090	*Body Care*
Wool socks	0.0088	*Accessory*
Brown knit scarf	0.0081	*Accessory*

(c) Topic 53, "women accessories"



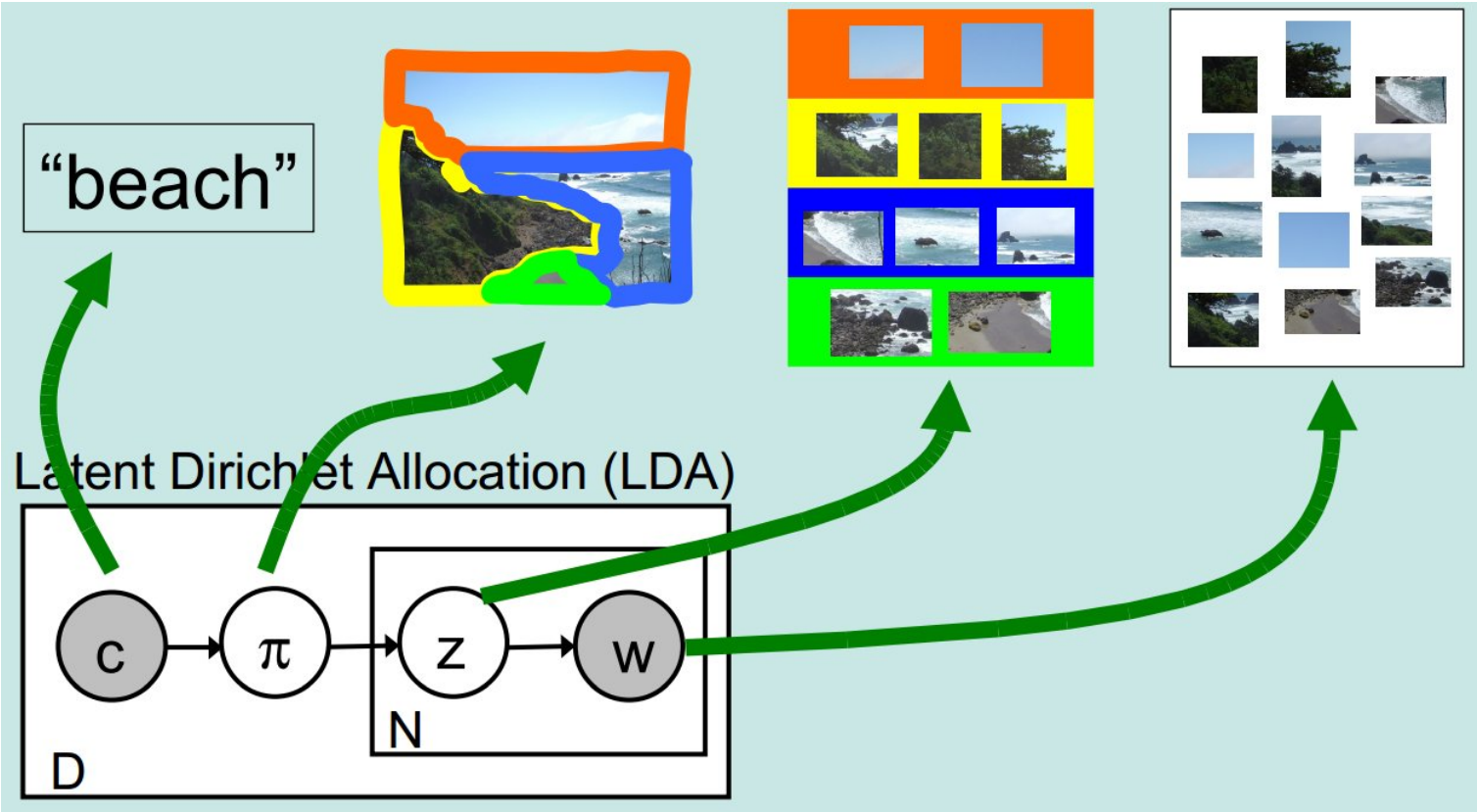
(a) User #3617 (b) User #5 (c) User #2618 (d) User #39

These results are used to direct advertising for a group-deal website.

Sun *et al.*, 2013

Discriminitive LDA in image recognition

Discriminitive LDA has been used to effectively classify images based on small parts of the image ("words").



Fei-Fei *et al.*, 2005

Correlated topic models for exploring *Science*

Correlated topics models have been used to identify related articles in *Science*.

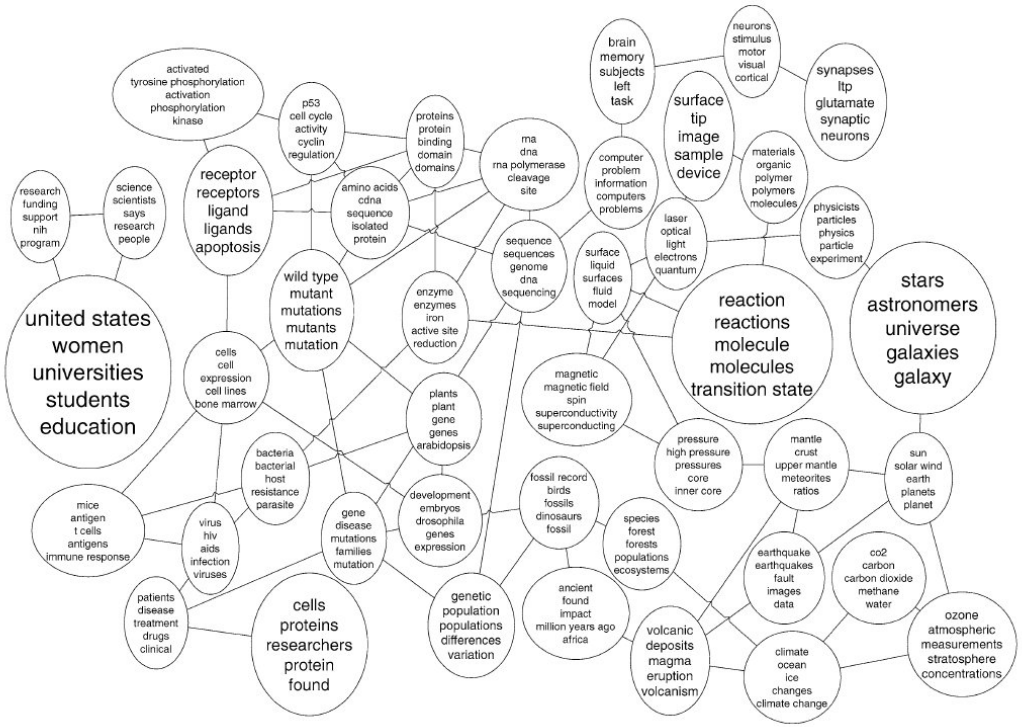


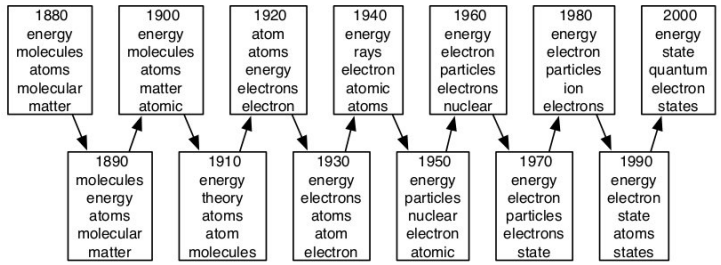
FIG. 2. *A portion of the topic graph learned from* 16,351 *OCR articles from Science* (1990–1999). *Each topic node is labeled with its five most probable phrases and has font proportional to its popularity in the corpus. (Phrases are found by permutation test.) The full model can be found in* http://www.cs.cmu.edu/~lemur/science/ *and on STATLIB.*

Blei and Lafferty, 2007

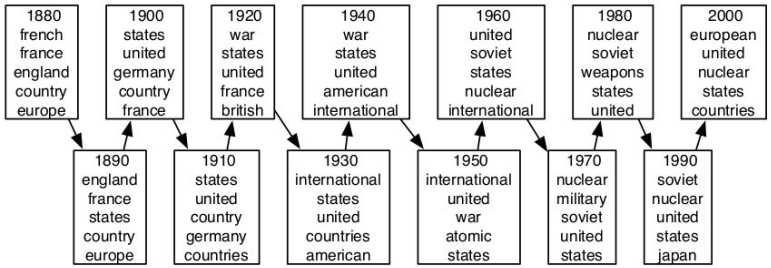
Dynamic topic models for exploring *Science*

Dynamic topics models have been used to analyze how the trends in *Science* have changed over time.

Topic 1:



Topic 2:



Blei and Lafferty, 2009

Recap

During this part of the course, we have discussed:

- Topic models as a method for uncovering underlying structure of objects
- Successful applications of graphical models "in the wild"

Final exam

The final exam is **Monday, February 24, 16:00 - 18:30 in B123**.

You can use a **simple** (non-graphing, not a cell phone, etc.) calculator and one A4 **handwritten** sheet of notes (front and back is okay).

The exercises were worth 25 points. The exam is worth 35 points. You must score at least 20 points on the exam to receive a grade of 1 or more. See the syllabus for the grading scale.

The renewal exam is **April 15**. The exercise points will count towards this exam.

Conclusions

Thanks for participating in the course.

I hope you have enjoyed it and will remember the problem-solving strategies we discussed when dealing with future challenges.

Feel free to contact me if you have any questions about probabilistic models in the future.