CSC2535 Spring 2013 Advanced Machine Learning

## Models of Text & Documents

**Richard Zemel** 

## Outline

Models of words and documents

- Simple document models
- Probabilistic document models
  - Aspect model
  - Latent Dirichlet Allocation
- Extensions of topic models
  - Author-recipient topic model
  - Dynamic topic model
  - Hierarchical topic models
- Topic models and vision

### **Topic Models**

- Have been applied to many types of data
  - Text
  - Images
  - Biological data
  - Relational data
  - Videos
  - and more ....









## **Document Modeling**

- automated analysis, visualization of text documents: crucial to effective use of large text archives (news stories, email collections, web)
- information retrieval: one of largest application areas of ML, growing steadily
- for example, the next generation of web searching will likely rely on automated summarization; paper-reviewer matching example
- today: statistical models of documents and text; examples of influential/interesting models

## **Representations of Documents**

# standard document representation: count occurrences of each word stem (bag-of-words)





## **Representations of Documents**

### D documents; V distinct words $\rightarrow$

Does high value of  $f_{wd}$  indicate an important word?

One transform: tf-idf (term frequency-inverse document frequency) →

G = VxD matrix of tf-idf values = tf \* idf

$$tf_{vd} = P(v | d) = f_{vd} / \sum_{v'} f_{v'd} \quad idf_v = D / \sum_d [f_{vd} > 0]$$
  
d to represent search query: sum of tf-idf of

Used to represent search query: sum of tf each query words

## **Topic Modeling**

Aim: Find low-dimensional description of highdimensional text

From ML viewpoint - just a latent variable problem!

- Topic models facilitate:
  - Summarization: find concise restatements
  - Similarity: evaluate distance between texts

Great case study of simple, extendable graphical model: test-bed for approximate inference, nonparametric variants

## Latent Semantic Analysis/Indexing

Reduced representation of F: apply SVD



- reduced representation of word i: row of AD -- can describe semantic relationships
- relationships between words described by cosine of angle between respective vectors

applications:

- train on 2K pages of English text, achieved average score on synonym portion of TOEFL
- train on introductory psychology textbook, achieved passing score on multiple-choice exam (Deerwester et al, 1990)

## Plates for Graphical Models

Probabilistic representations of documents: start with plate notation

Example: coin with unknown bias

- $\theta$  = probability of heads (parameter)
- X = coin toss outcome

N observations (repetitions)



H

$$P(X = H \mid \theta) = \theta$$

Observe: TTHTTTHT ML:  $\theta = 1/4$ 

## Simple Probabilistic Topic Models

Unigram model - each word with its own probability of appearing in document of length N



Problem: does not represent document containing a set of topics

Mixture of unigrams (single topic per document)



## Probabilistic LSI

Topic (aspect) model (Hoffman, 1999): probabilistic model of word production

$$P(w,d) = \sum_{k} P(w \mid z_{k}) P(z_{k} \mid d) P(d)$$

Generative model:

- select document d with probability P(d)
- select latent topic z with probability P(z|d): Mult $(z_k | \theta_k^{d_i})$
- generate word with probability P(w|z):

 $\mathrm{Mult}(w_j \,|\, \phi_j^k)$ 

Problems

- Lots of parameters mixture parameters for each document
- Does not generalize well



## **Conjugate Distributions**

To improve generative model, need to understand conjugate distributions

X = coin toss outcome (Bernoulli) N observations (repetitions) Prob of observing n heads:  $P(n | \theta, N) \propto \theta^n (1 - \theta)^{N-n}$ 



Prior over θ: Beta(α,β) [think of a as count of heads; β as count of tails] θ = probability of heads (variable) Key property: posterior is same form as prior

## **Conjugate Distributions**

Prior: pseudo-observations of heads/tails: Beta( $\alpha, \beta$ ):  $P(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}$ 

After n heads and N-n tails, posterior another Beta distribution, with a change in parameters:

$$P(\theta \mid n, N, \alpha, \beta) = \frac{P(n, N \mid \theta) P(\theta \mid \alpha, \beta)}{\int P(n, N \mid \theta') P(\theta' \mid \alpha, \beta) d\theta'}$$

$$\propto \left[\theta^{n}(1-\theta)^{N-n}\right]\left[\theta^{\alpha-1}(1-\theta)^{\beta-1}\right]$$
$$\propto \theta^{n+\alpha-1}(1-\theta)^{N-n+\beta-1}$$

## **Conjugate Distributions**

Prior  $P(\theta)$  is conjugate to class of likelihood if resulting posterior is in same family as  $P(\theta)$ 

$$P(\theta \mid X) = \frac{P(X \mid \theta) P(\theta)}{\int P(X \mid \theta') P(\theta') d\theta'}$$

Important because it avoids integration required to calculate posterior

Other conjugate distributions (all exponential family distributions have conjugate priors), e.g.,

[Likelihood-Prior-Posterior]: Gaussian-Gaussian-Gaussian; Poisson-Gamma-Gamma; Multinomial-Dirichlet-Dirichlet

(Dirichlet generalizes Beta to K alternatives)

## **Dirichlet Distribution**

Exponential family distribution over simplex of positive vectors that add up to 1

$$\operatorname{Dir}(\alpha_1,\ldots,\alpha_K): P(\theta \mid \alpha) = \prod_{k=1}^{K} (\theta_k)^{\alpha_k - 1} \frac{1}{B(\alpha)}$$

Used as a distribution over discrete distributions Symmetric Dirichlet: all  $a_k$  equal Concentration param. a controls shape, peakiness of  $\theta$ grows from 1: more uniform shrinks from 1 to 0: sparse

### Latent Dirichlet Allocation



- K number of latent topics.
- D number of documents
- $N_d$  Number of words in document d.
- V Number of words in vocabulary
- $\beta$  Dirichlet prior on  $\Phi_k$  (V-dim)

 $\Phi_k$  – distribution of words generated from topic k

 $\alpha$  – Dirichlet prior on  $\theta_d$  (K-dim)

 $\theta_d$  – distribution of topics in document d (K-dim)

z - latent topic (per-word)

w – observed word

(Blei et al., 2003)

### Latent Dirichlet Allocation



Generative process per doc: Choose  $\theta_d \sim Dir(\alpha)$ For each of  $N_d$  words w: Choose topic  $z_{dn} \sim Mult(\theta_d)$ Choose word  $w_{dn} \sim Mult(\Phi z_{dn})$ 

## Intuition into Representation

#### **Topics**

0.04 aene 0.02 dna genetic 0.01 . , , life 0.02 evolve 0.01 organism 0.01 . , , brain 0.04 0.02 neuron 0.01 nerve data 0.02 0.02 number computer 0.01 - 7 7

**Documents** 

Haemophilus genome 1703 genes

### Topic proportions and assignments

#### Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive! Last week at the genome meeting here,\* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The

other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

\* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

SCIENCE • VOL. 272 • 24 MAY 1996

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational mo-

lecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

Redundant and

parasite-spec

Genes In common 233 genes 422 genes 429 genes 420 genes

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

Genes

## Rough Summary of LDA

- Two aims
  - For each document, allocate words to just a few topics
  - For each topic, assign high probability to just a few terms
- But:
  - If one topic in document, all words must have high prob under that topic
  - If few terms in topic, cannot cover document's words

### Latent Dirichlet Allocation



Generative process per doc: **Choose**  $\theta_d \sim \text{Dir}(\alpha)$ For each of N<sub>d</sub> words w: Choose topic  $z_{dn} \sim Mult(\theta_d)$ Choose word  $w_{dn} \sim Mult(\Phi z_{dn})$ Mixed membership model: generalized mixture, each doc exhibits multiple topics  $\prod_{k=1}^{D} \left[ P(\theta_d \mid \alpha) \prod_{k=1}^{N} \left( P(z_{d,n} \mid \theta_d) P(w_{d,n} \mid z_{d,n}, \phi_{1:K}) \right) \right]$ 

## Inference in LDA

- Aim is to infer from a collection of documents
  - Per-corpus topic distributions  $\Phi_k$
  - Per-document topic proportions  $\boldsymbol{\theta}_d$
  - Per word topic assignments  $z_{dn}$
- Tricky to compute posterior over hidden variables given a document:  $P(z \neq A, w \mid \alpha, \beta)$

$$P(z,\phi,\theta \mid w,\alpha,\beta) = \frac{P(z,\phi,\theta,w \mid \alpha,\beta)}{P(w \mid \alpha,\beta)}$$

- Numerator tractable (conjugacy), but denominator not tractable, since it involves summing over all z
- note that document represented as continuous mixture:  $P(w \mid \alpha, \beta) = \int P(\theta \mid \alpha) \left( \prod_{n=1}^{N} P(w_n \mid \theta, \beta) \right) d\theta$

## Variational Inference for LDA

- Coordinate ascent in objective
- Each update closely related to true posterior
- For each topic k, term v

$$\lambda_{kv} = \beta_{kv} + \sum_{d} \sum_{n} I[w_{dn} = v]\varphi_{dnk}$$

For each document d

$$\gamma_{dk} = \alpha_k + \sum_n \varphi_{dnk}$$

- For each word n

 $\varphi_{dnk} \propto \exp\{E_q[\log(\theta_{dk}) + \log(\phi_{kw_{dn}})]\}$ 

### Collapsed Gibbs Sampling for LDA

- The latent topics, z, are sampled
- The Dir. distributions  $\boldsymbol{\theta}$  and  $\boldsymbol{\Phi}$  are integrated out
- Closed form sampling equations

$$\Pr(z_{nd} = k \mid \mathbf{Z}_{-(nd)}, \mathbf{w}) \propto (N_{kd} + \alpha) \frac{(N_{wk} + \beta)}{\sum_{w'} (N_{w'k} + \beta)}$$

- Each iteration requires O(K\*corpus size) ops.
- EXPENSIVE when counts are large
- Overall slower, more accurate than variational inference

### Model visualization

Can compute expectations of key terms based on posterior e.g., probability of term v in topic k:

$$\hat{\phi}_{kv} = E[\phi_{kv} \mid w_{1:D,1:N}]$$

 $\hat{\phi}_{kv} = \frac{\lambda_{kv}}{\sum \lambda_{kv'}}$ 

Given variational parameters these are easy to compute Describe topic:

term probabilities:

term-score (like tf-idf):

 $\hat{\phi}_{kv} \log \left( \frac{\hat{\phi}_{kv}}{\left( \prod_{k'} \hat{\phi}_{k'v} \right)^{1/K}} \right)$ Describe document: topic proportions

$$\hat{\theta}_{dk} = \frac{\gamma_{dk}}{\sum_{k'} \gamma_{dk'}}$$

### Example learned topics & doc model

Train on 160K documents; use variational EM, 100 topics, compute topic proportions and word assignments for test document

NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give \$1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be \$200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive \$400,000 each. The Juilliard School, where music and the performing arts are taught, will get \$250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual \$100,000 donation, too.

### **Example of learned topics**

Learned topics reveal hidden, implicit semantic categories in corpus

In many cases, can represent documents with 10<sup>2</sup> topics instead of 10<sup>5</sup> words

Especially important for short documents, e.g., emails - topics overlap when words don't

FIELD	SCIENCE	BALL	JOB
MAGNETIC	STUDY	GAME	WORK
MAGNET	SCIENTISTS	TEAM	JOBS
WIRE	SCIENTIFIC	FOOTBALL	CAREER
NEEDLE	KNOWLEDGE	BASEBALL	EXPERIENCE
CURRENT	WORK	PLAYERS	EMPLOYMENT
COIL	RESEARCH	PLAY	<b>OPPORTUNITIES</b>
POLES	CHEMISTRY	FIELD	WORKING
IRON	TECHNOLOGY	PLAYER	TRAINING
COMPASS	MANY E	BASKETBALL	SKILLS
LINES	MATHEMATICS	COACH	CAREERS
CORE	BIOLOGY	PLAYED	POSITIONS
ELECTRIC	FIELD	PLAYING	FIND
DIRECTION	PHYSICS	HIT	POSITION
FORCE	LABORATORY	TENNIS	FIELD
MAGNETS	STUDIES	TEAMS	OCCUPATIONS
BE	WORLD	GAMES	REQUIRE
MAGNETISM	[ SCIENTIST	SPORTS	OPPORTUNITY
POLE	STUDYING	BAT	EARN
INDUCED	SCIENCES	TERRY	ABLE

### Learned topics: term-scores

4	10	3	13
tax	labor	women	contract
income	workers	sexual	liability
taxation	employees	men	parties
taxes	union	sex	contracts
revenue	employer	child	party
estate	employers	family	creditors
subsidies	employment	children	agreement
exemption	work	gender	breach
organizations	employee	woman	contractual
year	iob	marriage	terms
treasury	bargaining	discrimination	bargaining
consumption	unions	male	contracting
taxpayers	worker	social	debt
earnings	collective	female	exchange
funds	industrial	parents	limited
6	15	1	16
6	15 sneech	1 firms	16 constitutional
6 jury trial	15 speech	1 firms price	16 constitutional
6 jury trial crime	15 speech free amendment	1 firms price corporate	16 constitutional political constitution
6 jury trial crime defendant	15 speech free amendment freedom	1 firms price corporate firm	16 constitutional political constitution
6 jury trial crime defendant defendants	15 speech free amendment freedom expression	1 firms price corporate firm value	16 constitutional political constitution government iustice
6 jury trial crime defendant defendants sentencing	15 speech free amendment freedom expression protected	1 firms price corporate firm value market	16 constitutional political constitution government justice amendment
6 jury trial crime defendant defendants sentencing iudges	15 speech free amendment freedom expression protected culture	1 firms price corporate firm value market cost	16 constitutional political constitution government justice amendment history
6 jury trial crime defendant defendants sentencing judges punishment	15 speech free amendment freedom expression protected culture context	1 firms price corporate firm value market cost canital	16 constitutional political constitution government justice amendment history people
6 jury trial crime defendant defendants sentencing judges punishment iudge	15 speech free amendment freedom expression protected culture context equality	1 firms price corporate firm value market cost capital shareholders	16 constitutional political constitution government justice amendment history people legislative
6 jury trial crime defendant defendants sentencing judges punishment judge crimes	15 speech free amendment freedom expression protected culture context equality values	1 firms price corporate firm value market cost capital shareholders stock	16 constitutional political constitution government justice amendment history people legislative oninion
6 jury trial crime defendant defendants sentencing judges punishment judge crimes evidence	15 speech free amendment freedom expression protected culture context equality values conduct	1 firms price corporate firm value market cost capital shareholders stock insurance	16 constitutional political constitution government justice amendment history people legislative opinion fourteenth
6 jury trial crime defendant defendants sentencing judges punishment judge crimes evidence sentence	15 speech free amendment freedom expression protected culture context equality values conduct ideas	1 firms price corporate firm value market cost capital shareholders stock insurance efficient	16 constitutional political constitution government justice amendment history people legislative opinion fourteenth article
6 jury trial crime defendant defendants sentencing judges punishment judge crimes evidence sentence jurors	15 speech free amendment freedom expression protected culture context equality values conduct ideas information	1 firms price corporate firm value market cost capital shareholders stock insurance efficient assets	16 constitutional political constitution government justice amendment history people legislative opinion fourteenth article majority
6 jury trial crime defendant defendants sentencing judges punishment judge crimes evidence sentence jurors offense	15 speech free amendment freedom expression protected culture context equality values conduct ideas information protect	1 firms price corporate firm value market cost capital shareholders stock insurance efficient assets offer	16 constitutional political constitution government justice amendment history people legislative opinion fourteenth article majority citizens

### Model evaluation

Standard topic model results entail showing some suggestive groupings of words into topics; quantitative evaluation hard



Or document classification: represent document based on its posterior topic proportions  $\rightarrow$  win when small proportion of dataset labeled

#### Author-Recipient-Topic model (McCallum et al., 2007)

extend LDA: analyze roles and relationships between people by analyzing email words wrt topic distributions

Latent Dirichlet Allocation

(LDA)

[Blei, Ng, Jordan, 2003]

**Author-Recipient Topic** 

(ART)

[McCallum, Corrada, Wang, 2004]

Ν

D





#### **Inference in Author-Recipient-Topic model**

models message content, and directed social network in which messages are sent

generative process, for each message d:

- 1. observe author  $a_d$  and set of recipients  $\mathbf{r}_d$
- 2. for each word in message d
  - (a) pick recipient r from  $\mathbf{r}_d$
  - (b) pick topic from author-recipient pair-specific multinomial  $\theta_{a_d,r}$
  - (c) pick word w from topic-specific multinomial  $\phi_z$

Aim: calculate posterior distribution of topic and recipient assignments given words –  $P(\mathbf{z}, \mathbf{r}|\mathbf{w}) = P(\mathbf{w}, \mathbf{z}, \mathbf{r}) / \sum_{\mathbf{z}, \mathbf{r}} P(\mathbf{w}, \mathbf{z}, \mathbf{r})$ 

can compute joint, by integrating out unknown  $\phi$  and  $\theta$  distributions (taking advantage of conjugate Dirichlet priors), but denominator cannot be calculated directly

instead use Gibbs sampling (see tutorial)

#### **Enron email corpus**

250K email messages, 147 people, 23K unique words

```
Date: Wed, 11 Apr 2001 06:56:00 -0700 (PDT)
From: debra.perlingiere@enron.com
To: steve.hooser@enron.com
Subject: Enron/TransAltaContract dated Jan 1, 2001
Please see below. Katalin Kiss of TransAlta has requested an
electronic copy of our final draft? Are you OK with this? If
so, the only version I have is the original draft without
revisions.
DP
Debra Perlingiere
Enron North America Corp.
Legal Department
1400 Smith Street, EB 3885
Houston, Texas 77002
dperlin@enron.com
```

### **Topics and prominent sender/receivers**

-	Topic	17	Торіс	27	Topic 45				
	"Document	Review"	"Time Sche	duling"	"Sports Pool"				
<b>T</b> eo 10	attached	0.0742	day	0.0419	game	0.0170			
lop words	agreement	0.0493	friday	0.0418	draft	0.0156			
within tonic :	review	0.0340	morning	0.0369	week	0.0135			
within topic .	questions	0.0257	monday	0.0282	team	0.0135			
	draft	0.0245	office	0.0282	eric	0.0130			
	letter	0.0239	wednesday	0.0267	make	0.0125			
	comments	0.0207	tuesday	0.0261	free	0.0107			
	сору	0.0165	time	0.0218	year	0.0106			
	revised	0.0161	good	0.0214	pick	0.0097			
2019a	document	0.0156	thursday	0.0191	phillip	0.0095			
Тор	G.Nemec	0.0737	J.Dasovich	0.0340	E.Bass	0.3050			
outhor rooinionto	B.Tycholiz		R.Shapiro		M.Lenhart				
author-recipients -	G.Nemec	0.0551	J.Dasovich	0.0289	E.Bass	0.0780			
exhibiting this	M.Whitt		J.Steffes		P.Love				
extine tang tine -	B.Tycholiz	0.0325	C.Clair	0.0175	M.Motley	0.0522			
topic	G.Nemec		M.Taylor		M.Grigsby				

#### **ART: Learns roles**

#### ART implicitly finds roles of individuals

Topic	34	Topic	37	Topic 4	41	Topic 42				
"Operat	ions"	"Power M	arket"	"Government l	Relations"	"Wirele	ss"			
operations	0.0321	market	0.0567	state	0.0404	blackberry	0.0726			
team	0.0234	power	0.0563	california	0.0367	net	0.0557			
office	0.0173	price	0.0280	power	0.0337	www	0.0409			
list	0.0144	system	0.0206	energy	0.0239	website	0.0375			
bob	0.0129	prices	0.0182	electricity	0.0203	report	0.0373			
open	0.0126	high	0.0124	davis	0.0183	wireless	0.0364			
meeting	0.0107	based	0.0120	utilities	0.0158	handheld	0.0362			
gas	0.0107	buy	0.0117	commission	0.0136	stan	0.0282			
business	0.0106	customers	0.0110	governor	0.0132	fyi	0.0271			
houston	0.0099	costs	0.0106	prices	0.0089	named	0.0260			
S.Beck	0.2158	J.Dasovich	0.1231	J.Dasovich	0.3338	R.Haylett	0.1432			
L.Kitchen		J.Steffes		R.Shapiro		T.Geaccone				
S.Beck	0.0826	J.Dasovich	0.1133	J.Dasovich	0.2440	T.Geaccone	0.0737			
J.Lavorato		R.Shapiro		J.Steffes		R.Haylett				
S.Beck	0.0530	M.Taylor	0.0218	J.Dasovich	0.1394	R.Haylett	0.0420			
S.White		E.Sager		R.Sanders		D.Fossum				

#### Beck = "Chief Operations Officer"

Dasovich = "Government Relations Executive" Shapiro = "Vice Presidence of Regulatory Affairs" Steffes = "Vice President of Government Affairs"

#### **Discovering role similarity**



#### connection strength (A,B) =

Similarity in recipients they sent email to Similarity in authored topics, conditioned on recipient

reflects jobs: Blair ('gas pipeline logistics')  $\approx$  Watson ('pipeline facility planning'); Geaconne ('executive assistant') vs. McCarty ('vice-president')

#### Dynamic Topic Models (Blei & Lafferty, 2006)

imagine topics evolve over time, so order of documents important; assume data divided by time-slice (e.g., year)

both Dirichlet distributions (over document topic proportions, and topic word proportions) replaced by simple dynamic model

1. Draw topics

$$\beta_t | \beta_{t-1} \sim \mathcal{N}(\beta_{t-1}, \sigma^2 I)$$

2. 
$$\alpha_t | \alpha_{t-1} \sim \mathcal{N}(\alpha_{t-1}, \gamma^2 I)$$

- 3. for each document
  - (a)  $\theta \sim \mathcal{N}(\alpha_t, a^2 I)$
  - (b) for each word
    - i.  $Z \sim Mult(softmax(\theta))$ ii.  $W_{tdn} \sim$

 $Mult(softmax(\beta_{tz}))$ 



#### **Dynamic Topic Models: Results**

*Science* corpus: 30K articles, 1881-1999, 250/yr; 16K vocabulary; 20-topic dynamic model; trained using Kalman filter variational approximation

1881		1890	)	1900	)	1910		1920	)	1930	)	1940	)	1950	)	1960	)	1970	)	1980	)	1990	1	2000	1
force		motion		magnet		force		atom		ray		energy		energy		radiat		electron		electron		electron		state	
energy		force		electric		magnet		theory		measure		measure		radiat		energy		energy		energy		atom		energy	
motion		magnet		measure		theory		electron		energy		electron		ray		electron		atom		particle		energy		electron	
differ		energy		force		electric		energy		theory		light		electron		measure		measure		field		structur		magnet	
light	٠	measure	►	theory	►	atom	•	measure	┝╸	light	┝	atom	↦	measure	┝╸	ray	┢╸	radiat	-	radiat	┢►	field	↦	field	
measure		differ		system		system		ray		wave		particle		atom		atom		field		model		model		atom	
magnet		direct		motion		measure		electr		radiat		ray		particle		field		ray		atom		state		system	
direct		line		line		line		line		atom		radiat		two		two		model		two		two		two	
matter		result		point		energy		force		electric		point		light		particle		particle		ray		magnet		quantum	
result		light		differ		body		value	I .	value		theory		absorpt		observe		magnet		measure	L .	rav		physic	L .



1900 On Kathode Rays and Some Related Phenomena 1917 ``Keep Your Eye on the Ball"

1920 The Arrangement of Atoms in Some Common Metals

1933 Studies in Nuclear Physics

1943 Aristotle, Newton, Einstein, II

1950 Instrumentation for Radioactivity

1881 On Matter as a form of Energy 1892 Non-Euclidean Geometry

- 1965 Lasers
- 1975 Particle Physics: Evidence for Magnetic Monopole Obtained

1985 Fermilab Tests its Antiproton Factory

1999 Quantum Computing with Electrons Floating on Liquid Helium

1881	٦	1890	ſ	1900	)	1910	)	1920	1	1930		1940	1	1950	)	1960		1970	)	1980	11	1990	1	2000
brain		movement		brain		movement		movement		stimulate		record		respons		response		respons		Cell		cell		neuron
moveme	it	eye		eye		brain		sound		muscle		nerve		record		stimulate		cell		neuron	1	channel		active
action		right		movement		sound		muscle		sound		stimulate		stimulate		record		potential		response	1	neuron		brain
right		hand		right		nerve		active		movement		response		nerve		condition		stimul		active		ca2		cell
eye	⊦	brain	►	left	┝	active	┢	nerve	Þ	response	┝	muscle	┝╸	muscle	┝╸	active	┢	neuron	┝╸	brain	-	active	↦	fig
hand		left		hand		muscle		stimulate		nerve		electrode		active		potential		active		stimul	11	brain		response
left		action		nerve		left		fiber		frequency		active		frequency		stimulus		nerve		muscle		receptor		channel
muscle		muscle		vision		eye		reaction		fiber		brain		electrode		nerve		eye		system		muscle		receptor
nerve		sound		sound		right		brain		active		fiber		potential		subject		record		nerve	11	respons		synapse
sound	J	experiment	J	muscle	J.	nervous	J	response	J.	brain		potential	J	study	J	eye		abstract	J.	receptor		current	Ļ	signal

"Neuroscience"



1887 Mental Science
1900 Hemianopsia in Migraine
1912 A Defence of the ``New Phrenology''
1921 The Synchronal Flashing of Fireflies
1932 Myoesthesis and Imageless Thought
1943 Acetylcholine and the Physiology of the Nervous System
1952 Brain Waves and Unit Discharge in Cerebral Cortex
1963 Errorless Discrimination Learning in the Pigeon
1974 Temporal Summation of Light by a Vertebrate Visual Receptor
1983 Hysteresis in the Force-Calcium Relation in Muscle

1993 GABA-Activated Chloride Channels in Secretory Nerve Endings

### **Infinite Topic Models**

So far all the topic models require specification of the number of topics

now consider infinite version, where the number of topics is potentially infinite

non-intuitive, yet fundamental idea underlying nonparametric Bayesian statistics

represent only as many topics as needed for a given dataset

examples of infinite models: Gaussian processes, Dirichlet process mixture models