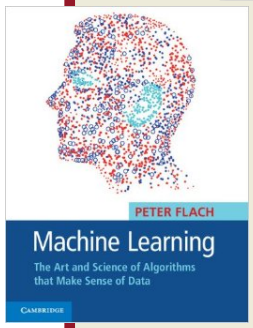
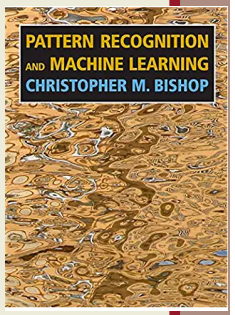
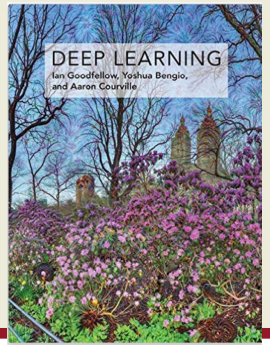
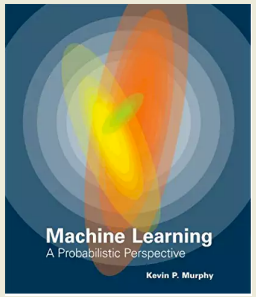


Machine Learning for SA





Lyle Ungar
Computer and information Science
University of Pennsylvania



Tutorial outline

- **Core sentiment analysis (SA) methods**
  - **Simple:** using lexica (dictionaries)
  - **Aspect-based:** using information extraction
- **Machine learning for SA**
  - Unsupervised: open language SA
  - Supervised: regression and deep learning
- **SA extensions**
  - Post, person and community
  - Multi-media

SA Goals and methods

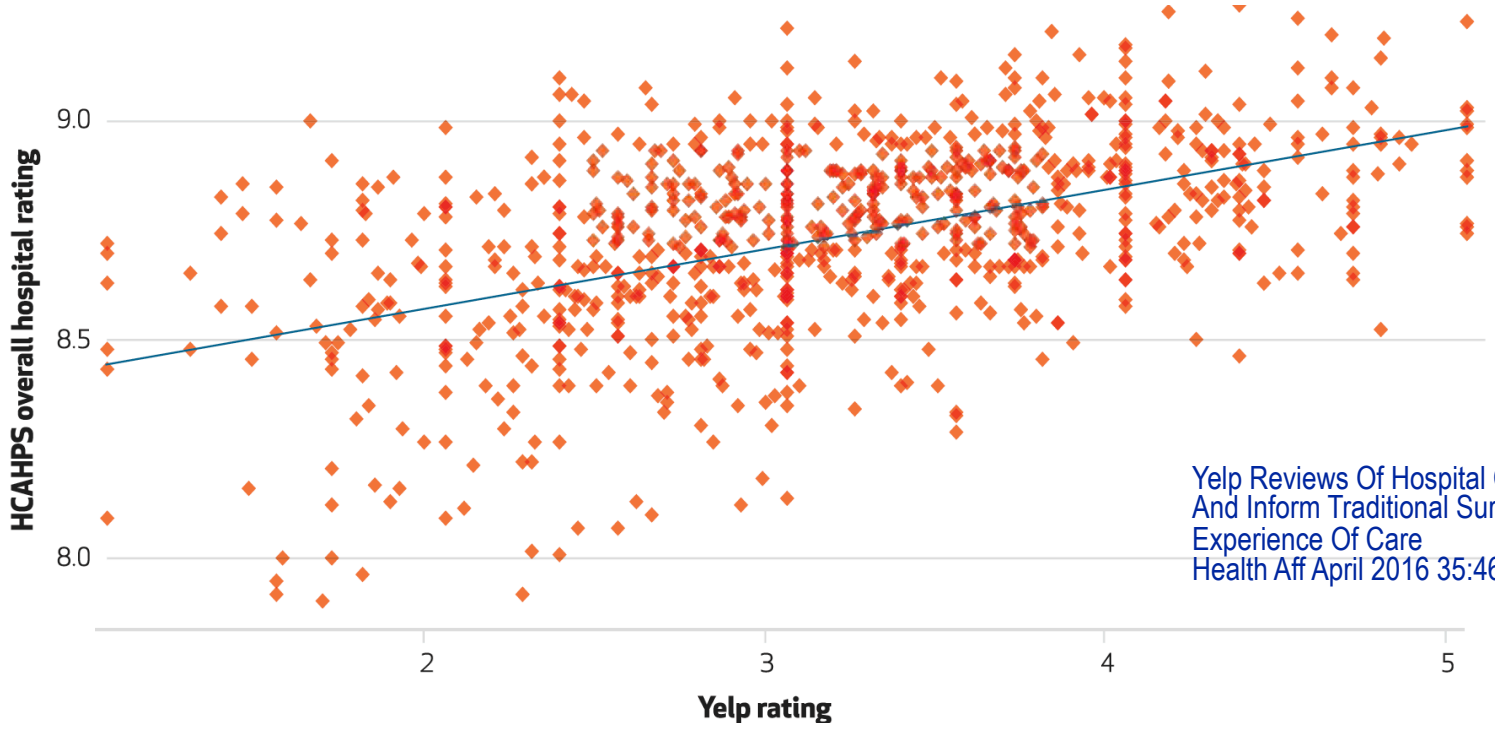
- **Prediction vs. Insight**
  - **Most SA is for insight**
    - Not “how well did people like this product?”
    - But “what did they like or dislike about it?”
- **Closed vs. Open vocabulary**
  - Closed: lexica selected *a priori*
  - Open: look to see what words correlate

Outline

- **Unsupervised learning: Open language SA**
  - Differential language analysis
  - LDA
  - Visualization
- **Supervised learning**
  - **Regression and regularization**
    - Semi-supervised learning
  - **Deep learning**
    - CNN, GRNN, LSTM
    - SA deep learning example: Socher

SA for hospitals from Yelp

- US hospitals are rated by formal surveys (HCAHPS) and on Yelp



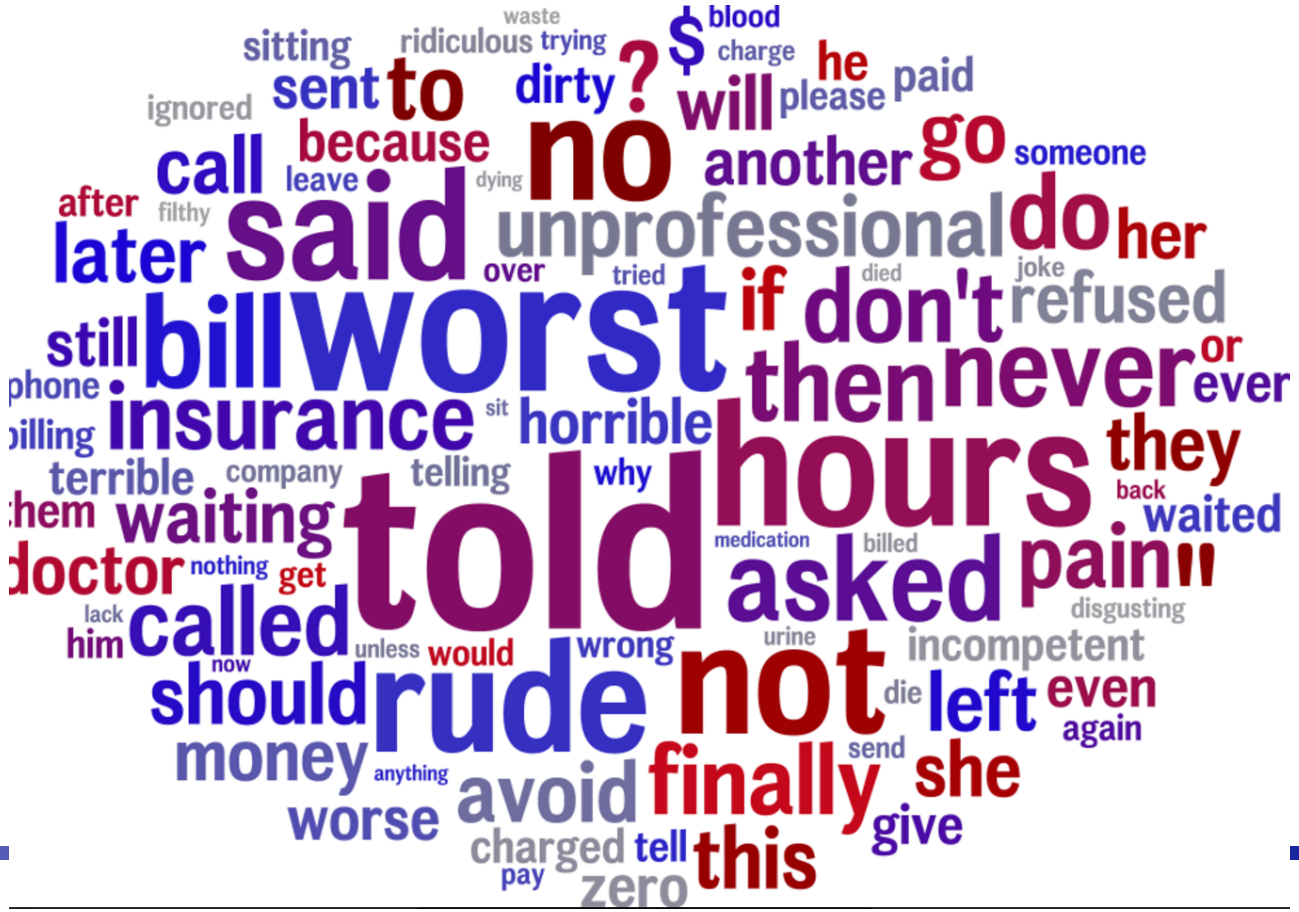
Yelp Reviews Of Hospital Care Can Supplement And Inform Traditional Surveys Of The Patient Experience Of Care
Health Aff April 2016 35:4697-705;

Differential Language Analysis

- **To get insight into what drives the difference in sentiment, compare the words in 1 and 5 star reviews**
  - **Find the words most correlated with high/low reviews**
    - Sometimes controlling for demographics ….
  - **Display words based on correlation (size) and frequency (color)**

Yelp - positive

Yelp - negative



Yelp finds 'missing' topics

Topic name	Correlation, Pearson's r	Covered by HCAHPS
YELP TOPICS MOST CORRELATED WITH POSITIVE YELP RATINGS		
Caring doctors, nurses, and staff	0.46	No
Comforting	0.29	No
Clean, private, nice hospital rooms	0.25	Yes
Surgery/procedure and peri-op	0.23	No
Labor and delivery	0.20	No
YELP TOPICS MOST CORRELATED WITH NEGATIVE YELP RATINGS		
Horrible hospital	−0.33	Yes
Rude doctor/nurse communication	−0.29	Yes
Pain control	−0.28	Yes
Insurance and billing	−0.26	No
Cost of hospital visit	−0.26	No

LDA topics

Specific type of medical care	**Pediatric** (son, daughter, hospital, child, kids, time, children's)	" My goddaughter was treated at [hospital] for a congenital heart defect…[The hospital] has gone above and beyond for my family in every respect possible, and its doctors and nurses have… the utmost care and professionalism."
	Labor and delivery (baby, birth, nurses, labor, delivery, experience, nurse)	"Excellent for having natural childbirth, they respected all my wishes. I had a midwife ... and my own doula."
	Surgical management and physical therapy (dr, physical, back, surgery, mri, therapy, year)	"This is one of the best rehab hospitals in the world. First class service by therapists, doctors, nurses, aids dietary, housekeeping. They taught me to walk again after total knee replacement."
	Wait time in ED (room, emergency, er, waiting, hours, wait, time)	"They suck!!!!!! They will make u wait for 3 hrs in the emergency waiting room even though the waiting room is empty!!!!!"
	ED imaging after injury (er, broken, back, accident, x-ray, foot, car)	"Their ER is awesome! I came in with a possible broken toe. They took me in right away, x-rayed my foot and diagnosed it (yup, broken toe)... 1.5 hours from arrival to dismissal. Best ER service Ive ever experienced."

Yelp Domains	Yelp Topic (topic terms)	Example Quote
Cost of hospital visit	**Cost of hospital visit** (insurance, bill, pay, cost, charge, visit, hospital)	"Ask what you will be charged before you allow this place to touch you!... [T]heir charges are excessive. In hospital 18 hours and got a 36k bill for knee surgery that cost 1/3 the amount in Los Angeles."
Insurance and Billing	**Insurance and billing** (insurance, billing, bill, hospital, department, company, paid)	"Billing department is awful. They billed the wrong insurance company twice and then finally had the correct one but billed it incorrectly... Instead of contacting me they sent the bill to a collection agency..."
Ancillary testing	**Ancillary testing and results** (blood, test, doctor, results, tests, lab, work)	"Service way too slow...for relatively standard testing…90 mins so far and sample hasn't even been collected because doc needs lab to tell him what tests need to be ordered."
Facilities	**Facility** (place, big, hand, time, TV, watch, cool)	"This place is awesome....seek out the quite courtyards for some solitude."
	Security and front desk (front, security, desk, room, friend, back, waiting)	"There was no sign that said to check-in, so I didn't bother with going to the information desk…The security guard who told us to check-in was… rude…"
Amenities	**Parking** (parking, hospital, building, free, lot, nice, valet)	"2 hours free parking with validation... I like it how they are building a garden between the parking structure and the hospital."
	Hospital food (food, cafeteria, hospital, good, eat, coffee, order)	"This review is for the cafeteria, prices are super cheap and the burger and nachos were terrific. Chili cheese fries were soggy…."

Latent Dirichlet Allocation (LDA)

D. Blei, A. Ng, and M. Jordan. *Journal of Machine Learning Research*, 3:993-1022, January 2003.

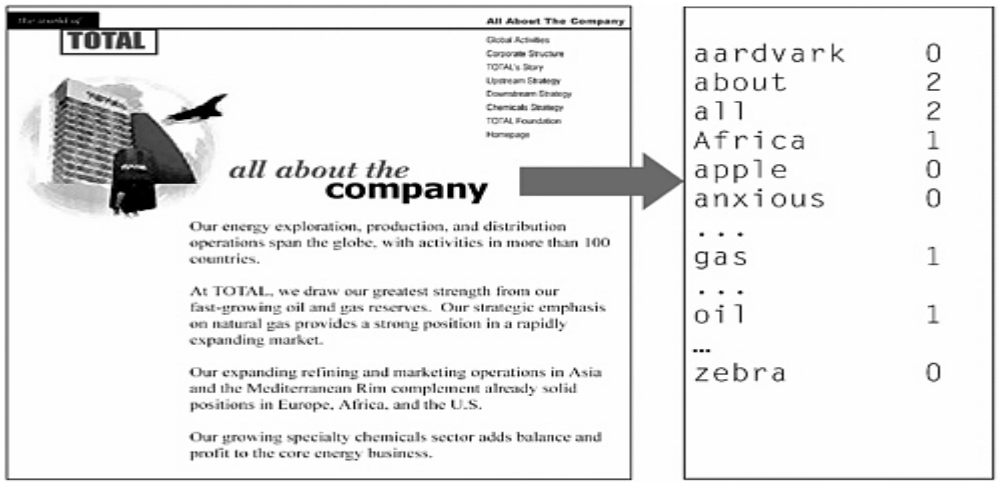
Following slides borrowed ant then heavily modified from:

Jonathan Huang (jch1@cs.cmu.edu)

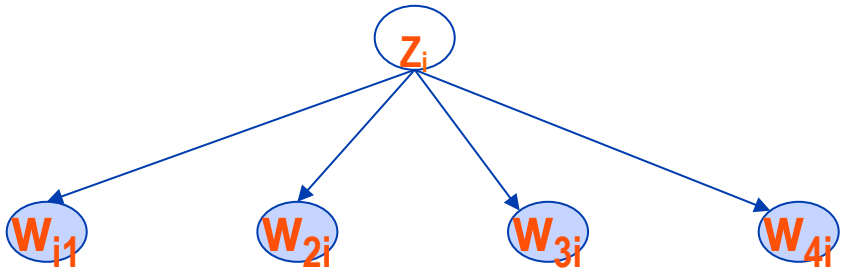
Lyle Ungar, University of Pennsylvania

"Bag of Words" Models

- **Assume that all the words within a document are exchangeable.**
  - The order of the words doesn't matter, just the count



Mixture of Unigrams = Naïve Bayes



Mixture of Unigrams = Naïve Bayes

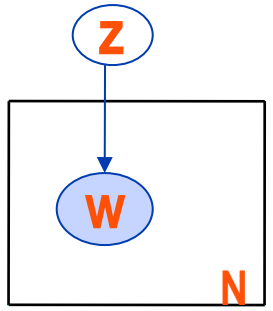


Plate Model

(equivalent)

Model: For each document:

- Choose a topic z_d with $p(topic_i) = \theta$
- Choose N words w_n by drawing each one independently from a multinomial conditioned on z_d with $p(w_n=word_j|topic_i=z) = \beta_z$
  - *Multinomial*: take a (non-uniform prior) dice with a word on each side; roll the dice N times and count how often each word comes up

In NB, we have exactly one topic per document

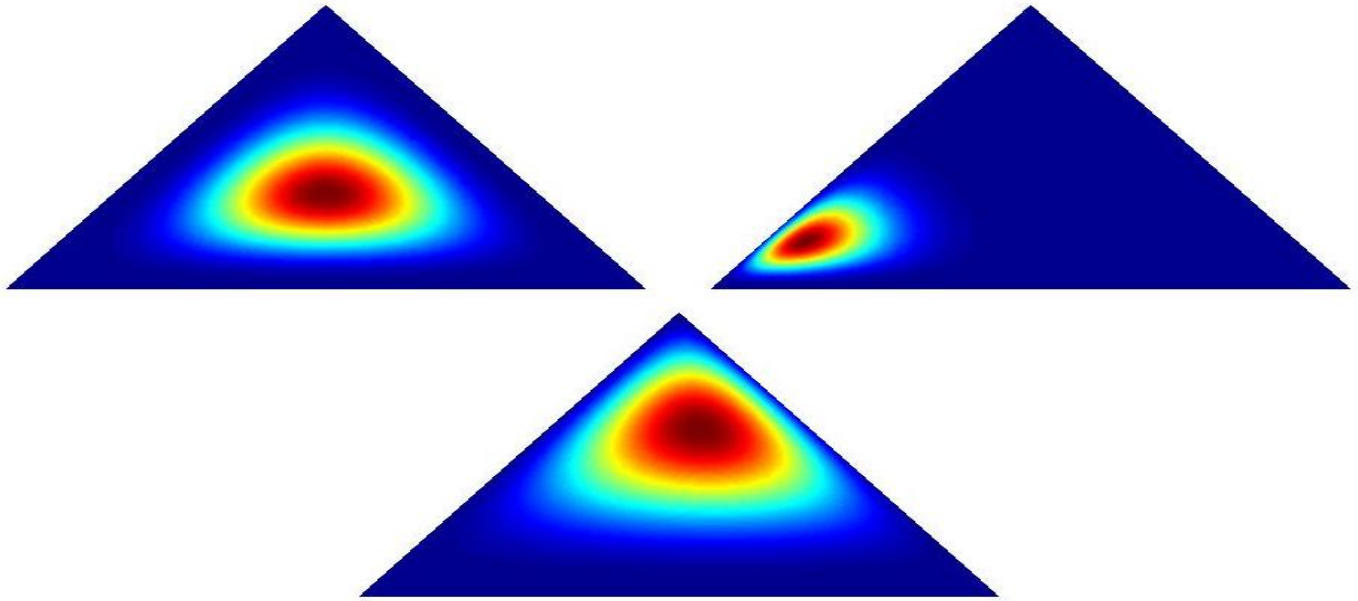
LDA: Each doc is a mixture of topics

- **LDA lets each document be a (different) mixture of topics**
  - Naïve Bayes assumes each *document* is on a single topic
  - LDA lets each *word* be on a different topic
- **For each document,**
  - **Choose a *multinomial distribution* θ_d over topics for that document**
  - **For each of the *N* words w_n in the document**
    - **Choose a topic z_n with *p(topic)* = θ_d**
    - **Choose a word w_n from a multinomial conditioned on z_n with $p(w=w_j|topic=z_n)$**
    - **Note: each topic has a different probability of generating each word**

Dirichlet Distributions

- **In the LDA model, we want the *topic mixture proportions* for each document to be drawn from some *distribution*.**
  - *distribution* = "probability distribution", so it sums to one
- **So, we want to put a prior distribution on multinomials. That is, k-tuples of non-negative numbers that sum to one.**
  - We want probabilities of probabilities
  - These multinomials lie in a (k-1)-*simplex*
    - *Simplex* = generalization of a triangle to (k-1) dimensions.
- **Our prior:**
  - Needs to be defined for a (k-1)-simplex.
  - Conjugate to the multinomial

3 Dirichlet Examples (over 3 topics)



Corners: only one topic
Center: uniform mixture of topics
Colors indicate probability of seeing the topic distribution

Dirichlet Distribution

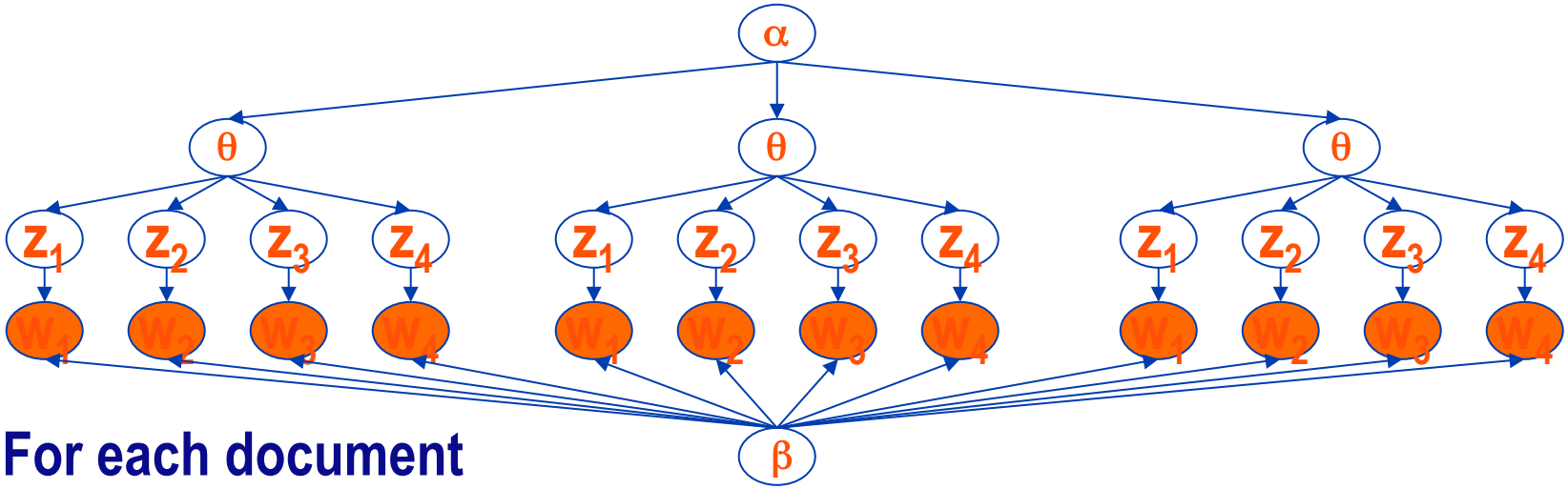
$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \prod_{i=1}^{k} \theta_i^{\alpha_i - 1}$$

- **Dirichlet distribution**
  - is defined over a (k-1)-simplex. I.e., it takes k non-negative arguments which sum to one.
  - is the conjugate prior to the multinomial distribution.
    - I.e. if our likelihood is multinomial with a Dirichlet prior, then the posterior is also Dirichlet
  - The Dirichlet parameter α_i can be thought of as the prior count of the i^{th} class.
- **For LDA, we often use a "symmetric Dirichlet" where all the α are equal;**
  - α is then a "concentration parameter"

Effect of α

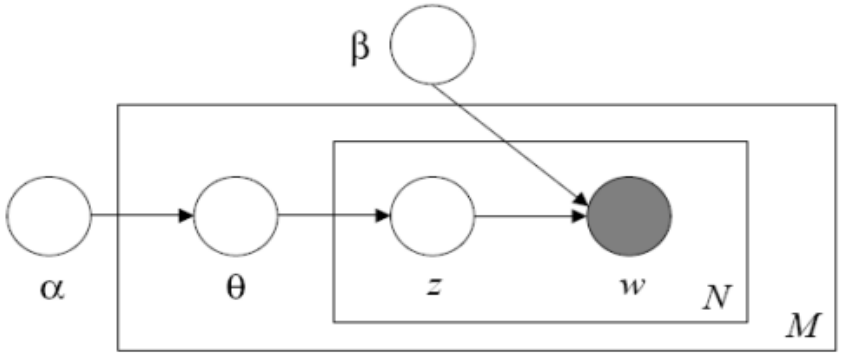
- When α < 1.0, the majority of the probability mass is in the "corners" of the simplex, generating mostly documents that have a small number of topics.
- When α > 1.0, the most documents contain most of the topics.

The LDA Model



- **For each document**
  - Choose the topic distribution $\theta \sim \text{Dirichlet}(\alpha)$
  - For each of the N words w_n:
    - Choose a topic $z \sim \text{Multinomial}(\theta)$
    - Then choose a word $w_n \sim \text{Multinomial}(\beta_z)$
      - Where each topic has a different parameter vector β for the words

The LDA Model: "Plate representation"



- **For each of M documents**
  - Choose the topic distribution $\theta \sim \text{Dirichlet}(\alpha)$
  - For each of the N words w_n:
    - Choose a topic $z \sim \text{Multinomial}(\theta)$
    - Choose a word $w_n \sim \text{Multinomial}(\beta_z)$

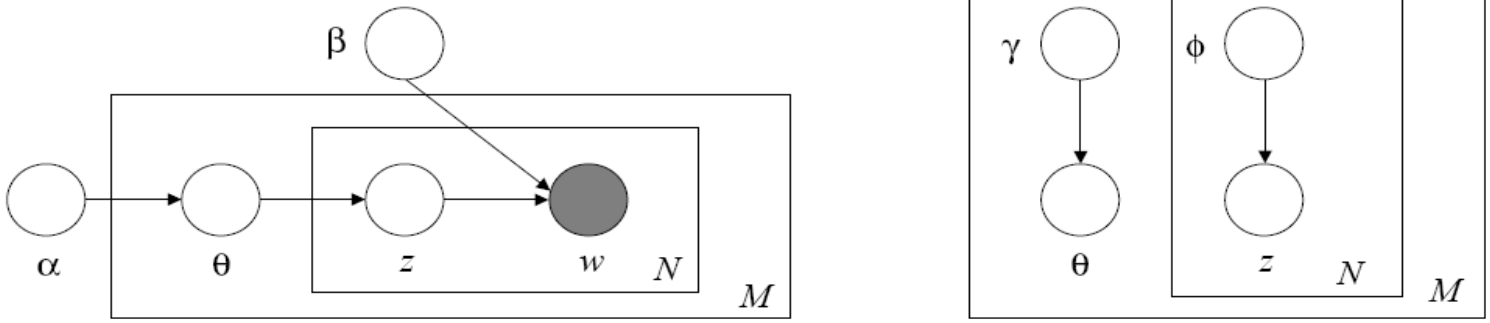
Parameter Estimation

Given a corpus of documents, find the parameters α and β which maximize the likelihood of the observed data (words in documents), marginalizing over the hidden variables θ, z

- **E-step:**
  - Compute $p(\theta,\mathbf{z}|\mathbf{w},\alpha,\beta)$, the posterior of the hidden variables (θ,**z**) given a document **w**, and hyper-parameters α and β.
- **M-step**
  - Estimate parameters α and β given the current hidden variable distribution estimates

    θ: topic distribution for the document,
    z: topic for each word in the document
- **Unfortunately, the E-step cannot be solved in a closed form**
  - So people use a "variational" approximation

Variational Inference



- In variational inference, we consider a simplified graphical model with variational parameters γ, ϕ and minimize the KL Divergence between the variational and posterior distributions.
  - *q* approximates *p*

$$(\gamma^*, \phi^*) = \arg\min_{(\gamma,\phi)} KL(q(\theta, z|\gamma, \phi)||p(\theta, z|w, \alpha, \beta))$$

Parameter Estimation: *Variational EM*

- Given a corpus of documents, find the parameters α and β which maximize the likelihood of the observed data.

- **E-step:**
  - Estimate the variational parameters γ and ϕ in $q(\gamma,\phi;\alpha,\beta)$ by minimizing the KL-divergence to p (with α and β fixed)
- **M-step**
  - Maximize (over α and β) the lower bound on the log likelihood obtained using q in place of p (with γ and ϕ fixed)

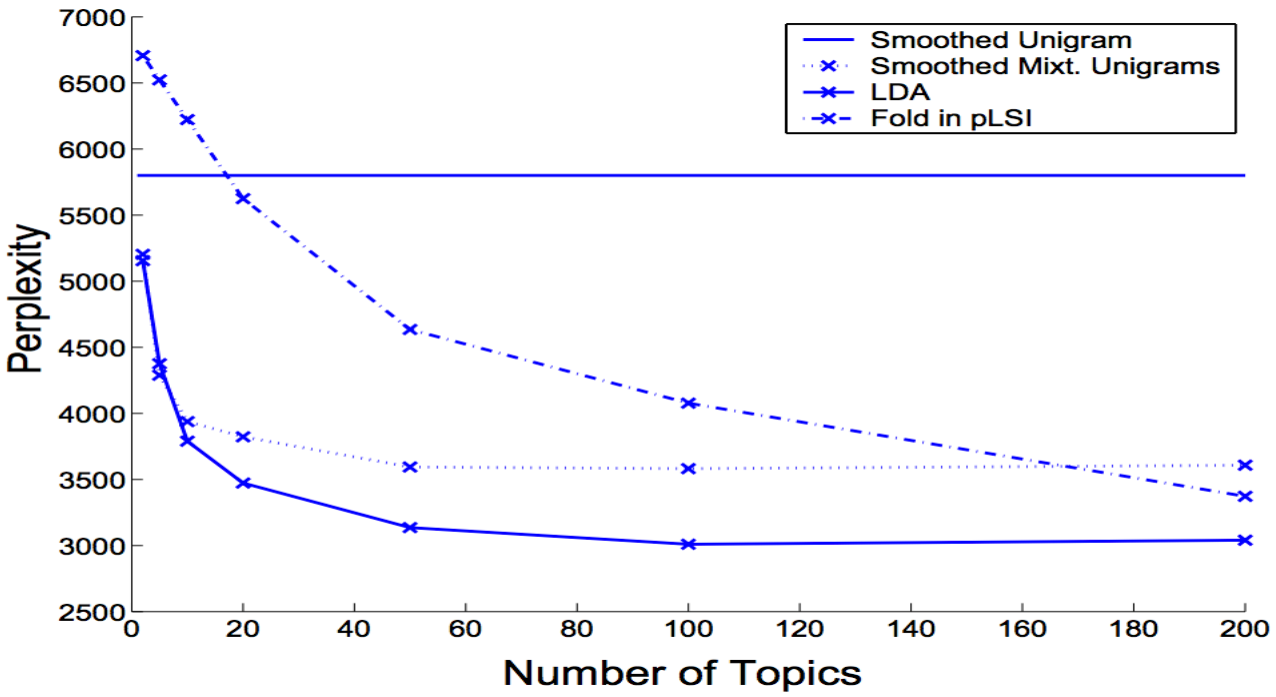
"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	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.

LDA requires fewer topics than NB

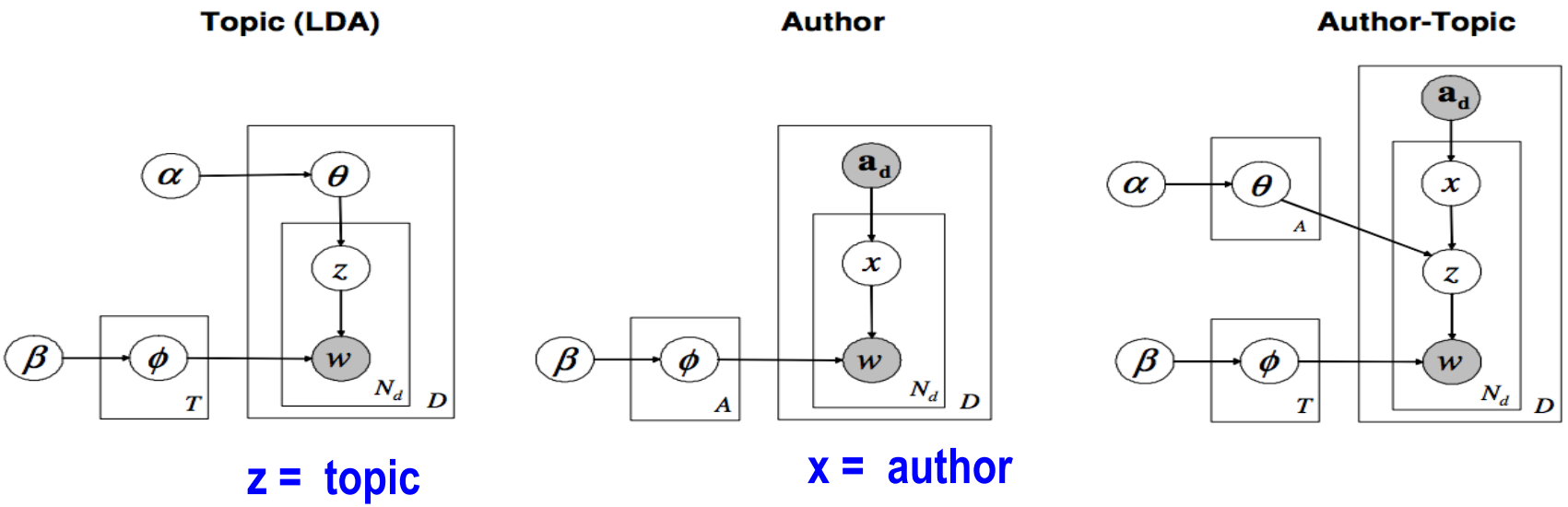
perplexity = $2^{H(p)}$ per word

I.e., $\log_2$(perplexity) = entropy



There are *many* LDA extensions

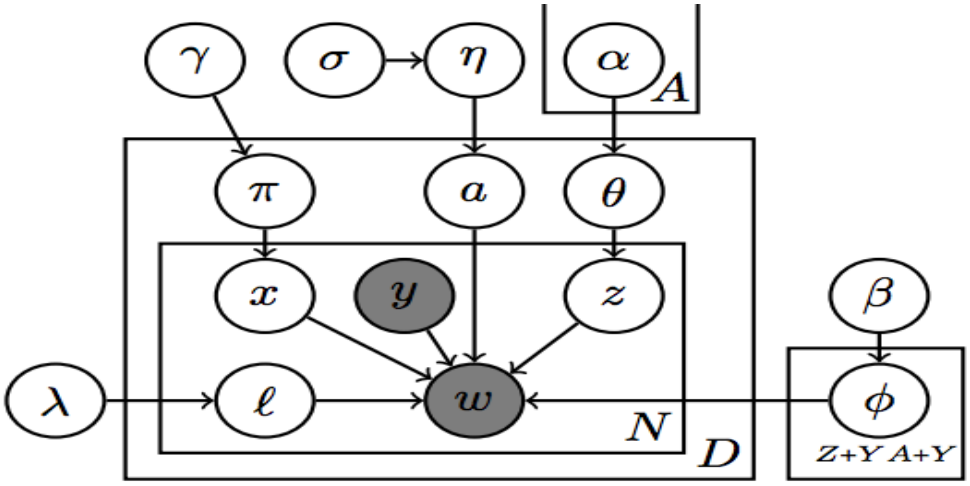
The author-topic model



z = topic

x = author

Ailment Topic Aspect Model



- Set the background switching binomial λ
- Draw an ailment distribution $\eta \sim \text{Dir}(\sigma)$
- Draw word multinomials $\phi \sim \text{Dir}(\beta)$ for the topic, ailment, and background distributions
- For each message $1 \leq m \leq D$:
  - Draw a switching distribution $\pi \sim \text{Beta}(\gamma_0, \gamma_1)$
  - Draw an ailment $a \sim \text{Mult}(\eta)$
  - Draw a topic distribution $\theta \sim \text{Dir}(\alpha_a)$
  - For each word $w_i \in N_m$
    - Draw aspect $y_i \in \{0, 1, 2\}$ (observed)
    - Draw background switcher $\ell \in \{0, 1\} \sim \text{Bi}(\lambda)$
    - If $\ell == 0$:
      - Draw $w_i \sim \text{Mult}(\phi_{B,y})$ (a background)
    - Else:
      - Draw $x_i \in \{0, 1\} \sim \text{Bi}(\pi)$
      - If $x_i == 0$: (draw word from topic z)
        - Draw topic $z_i \sim \text{Mult}(\theta)$
        - Draw $w_i \sim \text{Mult}(\phi_z)$
      - Else: (draw word from ailment a aspect y)
        - Draw $w_i \sim \text{Mult}(\phi_{a,y})$

Observed
- word w
- aspect y = symptom, treatment or other

Hidden
- topic type: background? (l), non-ailment (x)
- topic distribution

Paul & Dredze

LDA generation - example

- Topics = {sports, politics}
- Words = {football, baseball, TV, win, president}
- α = (0.8,0.2)
- β =

	sports	politics
football	0.3	0.01
baseball	0.25	0.01
TV	0.1	0.15
win	0.3	0.25
president	0.01	0.2
OOV	0.04	0.38

LDA generation - example

- **For each document**, d
  - Pick a topic distribution, θ_d using α
  - For each word in the document
    - pick a topic, z
    - given that topic, pick a word using β
- **α = (0.8,0.2)**
- **β =**

	sports	**politics**
football	0.3	0.01
baseball	0.25	0.01
TV	0.1	0.15
win	0.3	0.25
president	0.01	0.2

LDA in practice

- **Be careful what your unit of analysis is**
  - **Obvious: use each review as a document**
    - Reviews are mostly Zipfian
      - Most book reviews are about Harry Potter
    - So most of the topics end up capturing product features
      - Harry, Voldemort, Hermione, …
  - **Often better: group reviews for each product (book, hospital, ..)**

LDA in practice

- **Use a good package**
  - Mallet or Factorie
- **(Usually) build topics for the whole corpus**
  - And observe how topic frequencies differ for different products or sentiments
- **Pick an appropriate number of topics**
  - 2,000 for all of Facebook or Twitter
  - 200 for a specialized domain like hospital reviews

50 Facebook topics

Yields 2 happy topics

happy, Christmas, year, family, friends, hope, merry, Thanksgi
happy, birthday, day, love, wishes, mom, miss, wonderful, dad

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016, August 8). Gaining Insights From Social Media Language: Methodologies and Challenges. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000091

500 Facebook topics

Yields 8 happy topics

day, happy, mothers, mother's, mom, mother, wonderful, moms, mommy, mama
day, happy, valentines, fathers, valentine's, father's, dad, independence, dads, sing
birthday, happy, wishes, bday, wished, b-day, birthdays, present, celebrate, cake
year, happy, 2010, 2011, joy, wishing, bring, happiness, safe, diwali
happy, 4th, July, Halloween, year, fireworks, safe, fourth, holiday, holidays
happy, thanksgiving, Easter, family, thankful, turkey, holiday, bunny, enjoy, eggs
happy, birthday, anniversary, wishing, brother, son, bday, daddy, mommy, celebra
happy, makes, sooo, soo, soooo, Easter, Thanksgiving, camper, ending, sooooo

2000 Facebook topics

Yields 20 happy topics

happy, birthday, mommy, daddy, mama, momma, dearest, bestest, 21st, 18th
happy, birthday, sis, lil, bday, b-day, luv, cousin, 21st, nephew
happy, mothers, mother's, mom, moms, mother, mommy, mom's, mama, mommies
happy, makes, camper, unhappy, extremely, happier, smiling, satisfied, contented, content
happy, diwali, wishing, eid, happiness, mubarak, holi, festival, prosperous, gibran
Easter, happy, bunny, eggs, egg, hunt, holidays, risen, candy, basket
happy, birthday, brother, wishing, 18th, 21st, xxxx, 16th, monthsary, nephew
year, happy, 2010, 2011, chinese, 2009, cheers, prosperous, tiger, rabbit
happy, independence, friendship, valentines, Canada, valentine's, republic, memorial, Aust
year, happy, joy, happiness, bring, 2010, 2011, health, wishing, brings
happy, fathers, father's, dad, dads, father, daddy, dad's, mothers, papa
4th, July, happy, fireworks, fourth, safe, independence, bbq, 5th, quarter
happy, birthday, celebrate, anniversary, celebrating, birthdays, dad's, b-day, b'day, mom's
happy, valentines, valentine's, single, valentine, hump, pi, awareness, singles, v-day
happy, birthday, grandma, mama, aunt, Beth, Mary, anniversary, papa, grandpa

50 Facebook topics

Yields 1 play topic

game, play, win, playing, football, team, won, games, beat, lets

500 Facebook topics

Yields 5 play topics

guitar, play, playing, music, piano, band, bass, hero, practice, played
game, football, play, soccer, basketball, playing, games, team, practice, bas
place, chuck, find, meet, play, birth, Norris, interesting, babies, profile
play, playing, game, games, xbox, halo, Wii, video, Mario, 360
play, playing, game, ball, games, played, golf, tennis, poker, cards

2000 Facebook topics

Yields 25 play topics

golf, played, ultimate, frisbee, mini, ball, balls, golfing, tennis, disc
play, game, let's, role, sims, rules, chess, basketball, plays, poker
words, comment, note, play, wake, jail, copy, paste, sport, fair
black, cod, ops, playing, play, mw2, modern, warfare, ps3, online
game, team, won, win, played, boys, soccer, season, proud, football
soccer, football, game, play, team, basketball, playing, ball, practice, field
kids, park, playing, boys, played, pool, blast, playground, swimming, toys
sand, beach, water, toes, carl, grain, playin, mountain, rocks, desert
music, band, playing, piano, guitar, songs, sound, metal, bass, played
na, stuck, everyday, ki, replay, melody, ami, er, ta, singin
http://www.youtube.com, feature, related, =p, marcus, channel, double, user, nr, youtube_gdat
guitar, bass, drum, playing, amp, drums, string, strings, electric, acoustic
play, guitar, learn, piano, learning, playing, learned, lessons, songs, rules
games, play, playing, game, video, played, card, board, begin, playin
play, playing, starcraft, warcraft, sims, ii, beta, online, nerds, nerd
watchin, sittin, chillin, waitin, doin, havin, gettin, eatin, playin, drinkin
pokemon, playing, mon, shiny, version, pikachu, pok, cards, ds, ash
player, dvd, cd, record, printer, bought, set, mp3, ink, borrow
anime, manga, naruto, bleach, episode, series, cosplay, episodes, alchemist, Japanese

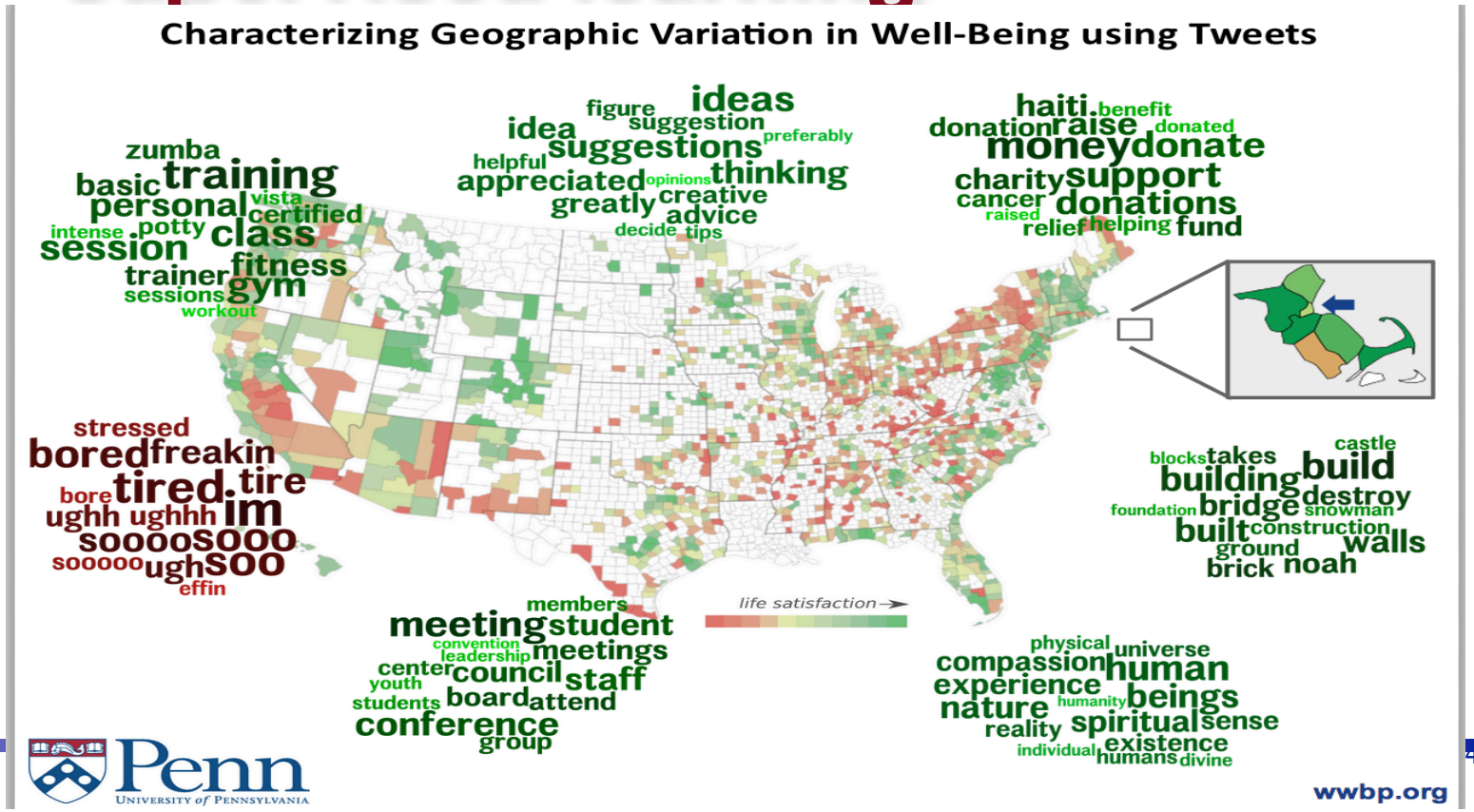
Find how topics correlate with …

- **Positive/negative sentiment**
  - For different product categories

Yelp hospital reviews again

Cost of hospital visit	**Cost of hospital visit** (insurance, bill, pay, cost, charge, visit, hospital)	"Ask what you will be charged before you allow this place to touch you!... [T]heir charges are excessive. In hospital 18 hours and got a 36k bill for knee surgery that cost 1/3 the amount in Los Angeles."
Insurance and Billing	**Insurance and billing** (insurance, billing, bill, hospital, department, company, paid)	"Billing department is awful. They billed the wrong insurance company twice and then finally had the correct one but billed it incorrectly... Instead of contacting me they sent the bill to a collection agency..."
Ancillary testing	**Ancillary testing and results** (blood, test, doctor, results, tests, lab, work)	"Service way too slow...for relatively standard testing…90 mins so far and sample hasn't even been collected because doc needs lab to tell him what tests need to be ordered."
Facilities	**Facility** (place, big, hand, time, TV, watch, cool)	"This place is awesome....seek out the quite courtyards for some solitude."
	Security and front desk (front, security, desk, room, friend, back, waiting)	"There was no sign that said to check-in, so I didn't bother with going to the information desk…The security guard who told us to check-in was… rude…"
Amenities	**Parking** (parking, hospital, building, free, lot, nice, valet)	"2 hours free parking with validation... I like it how they are building a garden between the parking structure and the hospital."
	Hospital food (food, cafeteria, hospital, good,	"This review is for the cafeteria, prices are super cheap and the burger and nachos were terrific. Chili cheese fries were

LDA topics can be used for semi-supervised learning



Unsupervised learning

- **Differential language analysis**
  - Using words
  - Using topics
- **LDA**
  - **Each document is a mixture over topics**
  - **Each topic looks like a Naïve Bayes model**
    - It produces words with some probability
  - **Estimation of LDA is messy**
    - Requires variational EM or Gibbs sampling
  - **Lots of extensions and variations**

Supervised Learning
for SA

Lyle Ungar

Key methods

- **Linear and logistic regression**
  - Requires regularization
- **Deep Learning**
  - **Static** – convolutional neural nets
  - **Dynamic** – GRNNs and LSTMs

Regression for sentiment analysis

- **Given a set of observations with labels, *y***
  - Observations
    - Sentences or reviews labeled with a sentiment score
- **Generate features, *x*, for each observation**
  - E.g. presence or count of a 'word'
- **Learn a regression model to predict y**
  - $y = f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 \ldots$
  - *Most of the w_j are zero.*

Overfitting and Regularization

- With 10,000 labeled tweets and a vocabulary of 40,000 words, you can predict sentiment *really* well

Regularization approaches

- **Use PCA or LDA to to dimensionality reduction**
  - 40,000 words -> 200 topics or factors
    - Often *semi-supervised*
- **Do feature selection**
  - Zero out some words
- **Use Ridge or Lasso or other penalty to "shrink" weights**

A word on words

"I'm sick of dos movies! :("

- **How many words?**
  - 5? 6? 7? More?
- **I'm sick_of dos movies ! :(**
- **Remove punctuation and stopwords?**
- **Normalize spelling?**
  - Depends on the application
    - often yes for product names
    - No for other words

To find multi-word expressions

- Use Pointwise mutual information

$$p(w_1,w_2) / p(w_1)p(w_2)$$

Transform word counts

- **Linear regression assumes linearity (duh!)**
  - $y = f(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 \ldots$
  - But seeing a word 10 vs. 11 times is not the same as seeing it never vs. once.
- **So transform word counts**
  - Easy: square root of count
  - Often better: Anscombe transformation

$$2\sqrt{p(phrase) + \frac{3}{8}}$$

- **Then normalize for document length to get word *frequency***

Now we're ready for regression

- **Linear regression if y is real-valued**
  - But may require transformation
- **Logistic regression if y is binary**

Feature selection for regression

- **Goal: minimize error on a test set**
- **Approximation: minimize a penalized training set error**
  - Argmin_w (Err + λ $|w|_p^p$) where Err = $\Sigma_i (y_i - \Sigma_j w_j x_{ij})^2 = ||y - w^T x||^2$
  - **Different norms**
    - **p = 2 – "ridge regression"**
      - Makes all the *w*'s a little smaller
    - **p = 1 – "LASSO" or "LARS" (least angle regression)**
      - drives some *w*'s to zero; makes all smaller
    - **p = 0 – "stepwise regression"** – the number of features
      - Zeros some *w*'s; leaves rest untouched
- **How to pick λ?**

Note the confusion in the names of the of optimization method with the objective function

Solving with regularization penalties

$\text{Argmin}_w \quad |y - w \cdot x|_2^2 + \lambda |w|_p^p$

- **p = 2 (Ridge)**
  - Closed form: $(X'X + \lambda I)^{-1} X'y$
- **p = 1 (Lasso)**
  - Convex optimization
- **p = 0 (Stepwise regression)**
  - Search

"Elastic net" combines Ridge and Lasso
widely used
convex

Regularization Summary

- **Penalized error approximates test error**
  - L_2, L_1, L_0 penalties
    - Convex and non-convex
  - Use tests set/cross validation to evaluate methods
- **PCA gives roughly similar results to Ridge (L_2)**
- **Highest accuracy from hybrid methods**
  - But requires careful regularization

If you have many more features than observations you need to zero out some of the features