WHAT AFFECTS PATIENT (DIS)SATISFACTION?
ANALYZING ONLINE DOCTOR RATINGS WITH A JOINT TOPIC-SENTIMENT MODEL

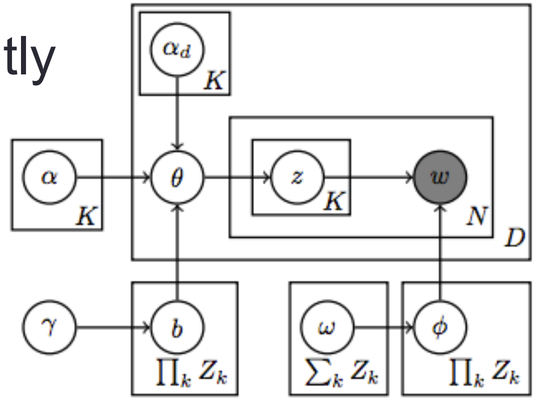
Michael Paul, Johns Hopkins University
Byron Wallace, Brown University
Mark Dredze, Johns Hopkins University

Online Doctor Reviews

- Online reviews of doctors are increasingly widespread and widely used
  - An estimated 80% of doctors have reviews written about them
  - About 20% of Internet users have looked up doctor reviews in the past year
- Can also give us insights into what patients look for and care about in a health provider
- Analyzing these reviews at a large scale requires some automation

Our Approach: Topic Models

- Provides coarse overview of large text collection
  - Can highlight the main themes that appear in doctor reviews
- Previous work used LDA
  - Brody and Elhadad (2010)
- This work: **Factorial LDA**
  - Paul and Dredze (2012)
  - Model both **topic** and **sentiment** together, jointly
- Will explain in a few minutes



Data Source: RateMDs.com

- Online site for doctor reviews
- We collected **52,226** reviews from the United States
- Each review includes 1–5 rating for various aspects
  - Helpfulness
  - Knowledgeability
  - Staff

Data Source: Labeled Reviews

- 846 doctor reviews hand-coded with aspect and sentiment information
  - López et al. (2012)

- Three main aspects
  - Interpersonal manner
  - Technical competence
  - Systems issues

Systems issues	
positive	*negative*
friendly staff, short wait times, convenient location	difficult to park, rude staff, expensive

- Aspects roughly map to the RateMDs.com ratings:
  - Interpersonal manner : Helpfulness
  - Technical competence: Knowledgeability
  - Systems issues : Staff

Topic Modeling

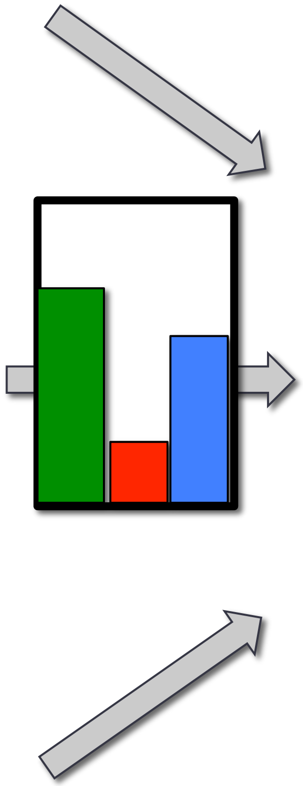
- Probabilistic model of text generation
  - e.g. Latent Dirichlet Allocation (Blei et al, 03)
- Each document has a distribution over *topics*
- Each topic has a distribution over words
- Each word token is associated with a latent topic variable

Topic Modeling

football	0.03
team	0.01
hockey	0.01
baseball	0.005
…	…

charge	0.02
court	0.02
police	0.015
robbery	0.01
…	…

congress	0.02
president	0.02
election	0.015
senate	0.01
…	…



Jury Finds Baseball Star Roger Clemens Not Guilty On All Counts

A **jury** found **baseball** star **Roger Clemens** not **guilty** on six **charges** against. **Clemens** was **accused** of **lying** to **Congress** in 2008 about his use of **performance** enhancing **drugs**.

Factorial LDA (f-LDA)

- **Multi-dimensional** topic model
  - Paul and Dredze (2012)
- Word tokens are associated with a *vector* of latent variables instead of a single topic variable
  - Can jointly model pairs of concepts like topic and sentiment
- Instead of a distribution over topics, each document has distribution over *pairs*
- Each pair is associated with its own word distribution

Factorial LDA (f-LDA)

- We use f-LDA to model topic and sentiment
- Each (topic,sentiment) pair has a word distribution
- e.g. (Systems/Staff, Negative):

office
time
doctor
appointment
rude
staff
room
didn't
visit
wait

Factorial LDA (f-LDA)

- We use f-LDA to model topic and sentiment
- Each (topic,sentiment) pair has a word distribution
- e.g. (Systems/Staff, Positive):

dr
time
staff
great
helpful
feel
questions
office
really
friendly

Factorial LDA (f-LDA)

- We use f-LDA to model topic and sentiment
- Each (topic,sentiment) pair has a word distribution
- e.g. (Interpersonal, Positive):

dr
doctor
best
years
caring
care
patients
patient
recommend
family

Model Parameters

- Why should the word distributions for pairs make any sense?

- Parameters are tied across the priors of each word distribution
  - The prior for (Systems,Negative) shares parameters with (Systems,Positive) which shares parameters with the prior for (Interpersonal,Positive)

Model Parameters

Systems	Positive
staff	recommend
time	wonderful
office	highly
questions	knowledgeable
wait	professional
helpful	kind
nice	great
feel	dr
great	best
appointment	helpful
nurse	amazing

Each dimension has a weight vector over the vocabulary

Model Parameters

exp(Systems + Positive) =

Systems	Positive	
staff	recommend	dr
time	wonderful	time
office	highly	staff
questions	knowledgeable	great
wait	professional	helpful
helpful	kind	feel
nice	great	doctor
feel	dr	questions
great	best	office
appointment	helpful	friendly
nurse	amazing	really

Model Parameters

Systems

Positive

dr
time
staff
great
helpful
feel
questions
office
really
friendly
doctor

multinomial parameters
sampled from Dirichlet

dr
time
staff
great
helpful
feel
doctor
questions
office
friendly
really

Model Parameters

- Where did the weight vectors come from?
- Parameter optimization
  - We learn from the data
- Overall algorithm
  - E-step: Gibbs sampling of variable assignments (same as LDA)
  - M-step: Gradient ascent on weight vectors
- We would probably not what we care about with zero supervision
  - **Semi-supervised** approach using informed priors

Semi-Supervision

- We incorporate prior knowledge into the model from two sources
- Prior knowledge from **labeled data**
  - Based on the Lopez dataset
  - Influence the **word distributions**
- Prior knowledge from **user ratings**
  - Based on the 1–5 scores in the RateMDs dataset
  - Influence the **(topic,sentiment) distributions**

Prior Knowledge from Labeled Data

- Learn a similar but simpler log-linear model on the labeled data
  - Details in paper; similar to Paul and Dredze (2013)
- The weights learned by training the log-linear model serve as a **Gaussian prior** over the weights in our f-LDA model

Prior Knowledge from Labeled Data

- Learn a similar but simpler log-linear model on the labeled data
  - Details in paper; similar to Paul and Dredze (2013)
- The weights learned by training the log-linear model serve as a **Gaussian prior** over the weights in our f-LDA model

Systems

staff
time
office
questions
wait
helpful
nice
feel
great
appointment
nurse

$\sim N($

office
receptionist
staff
appointment
friendly
waiting
make
minutes
plus
wait
sick

$, \sigma^2)$

Prior Knowledge from Labeled Data

- Learn a similar but simpler log-linear model on the labeled data
  - Details in paper; similar to Paul and Dredze (2013)
- The weights learned by training the log-linear model serve as a **Gaussian prior** over the weights in our f-LDA model

Positive $\sim N($ thorough, great, best, dr, ive, caring, friendly, hes, knowledgeable, listens, professional $, \sigma^2)$

Positive (prior)	Positive (mean)
recommend	thorough
wonderful	great
highly	best
knowledgeable	dr
professional	ive
kind	caring
great	friendly
dr	hes
best	knowledgeable
helpful	listens
amazing	professional

Prior Knowledge from User Ratings

- We incorporated the 1–5 ratings into the Dirichlet prior over the (topic,sentiment) pair distribution
- Defined as:

$$\exp\left(\underbrace{\alpha^{(B)} + \alpha_{t_1}^{(\mathcal{D},\text{top.})} + \alpha_{t_1}^{(d,\text{top.})} + \alpha_{t_2}^{(\mathcal{D},\text{sen.})} + \alpha_{t_2}^{(d,\text{sen.})}}_{\substack{\text{Prior in standard f-LDA} \\ \text{Details in paper}}} + \underbrace{\rho\, r_{\vec{t}}^{(d)}}_{\textbf{New}}\right)$$

  - r is user rating variable – given as input in the dataset
  - ρ is scaling parameter – automatically optimized for likelihood

Experiments

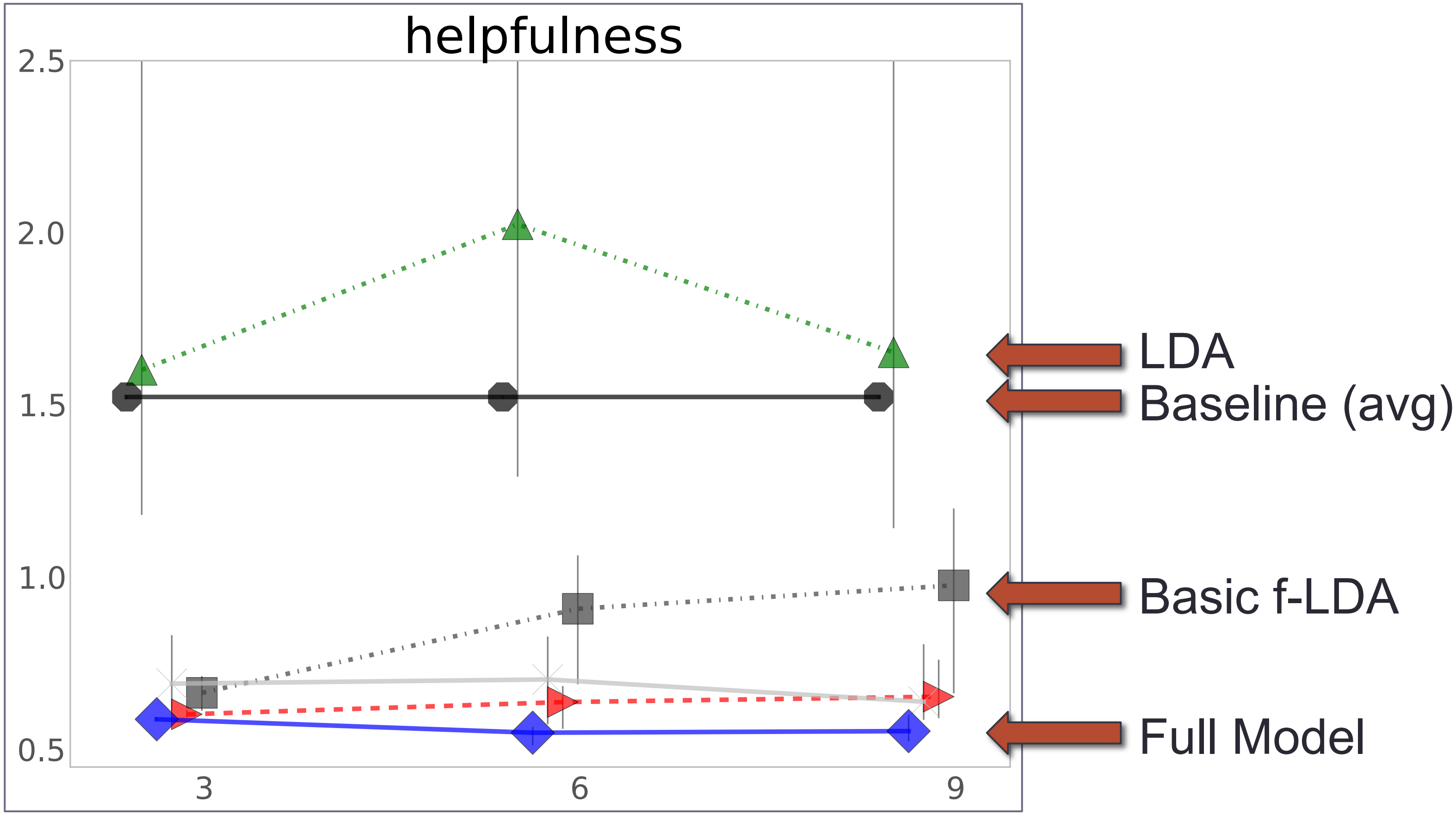
- Task: predict the user ratings from the (topic, sentiment) distributions
  - Measures how predictive the model is

- Method: ordinal regression
  - f-LDA distributions only
  - f-LDA distributions + bag of words

- Compared against f-LDA variants without the two modifications with prior knowledge

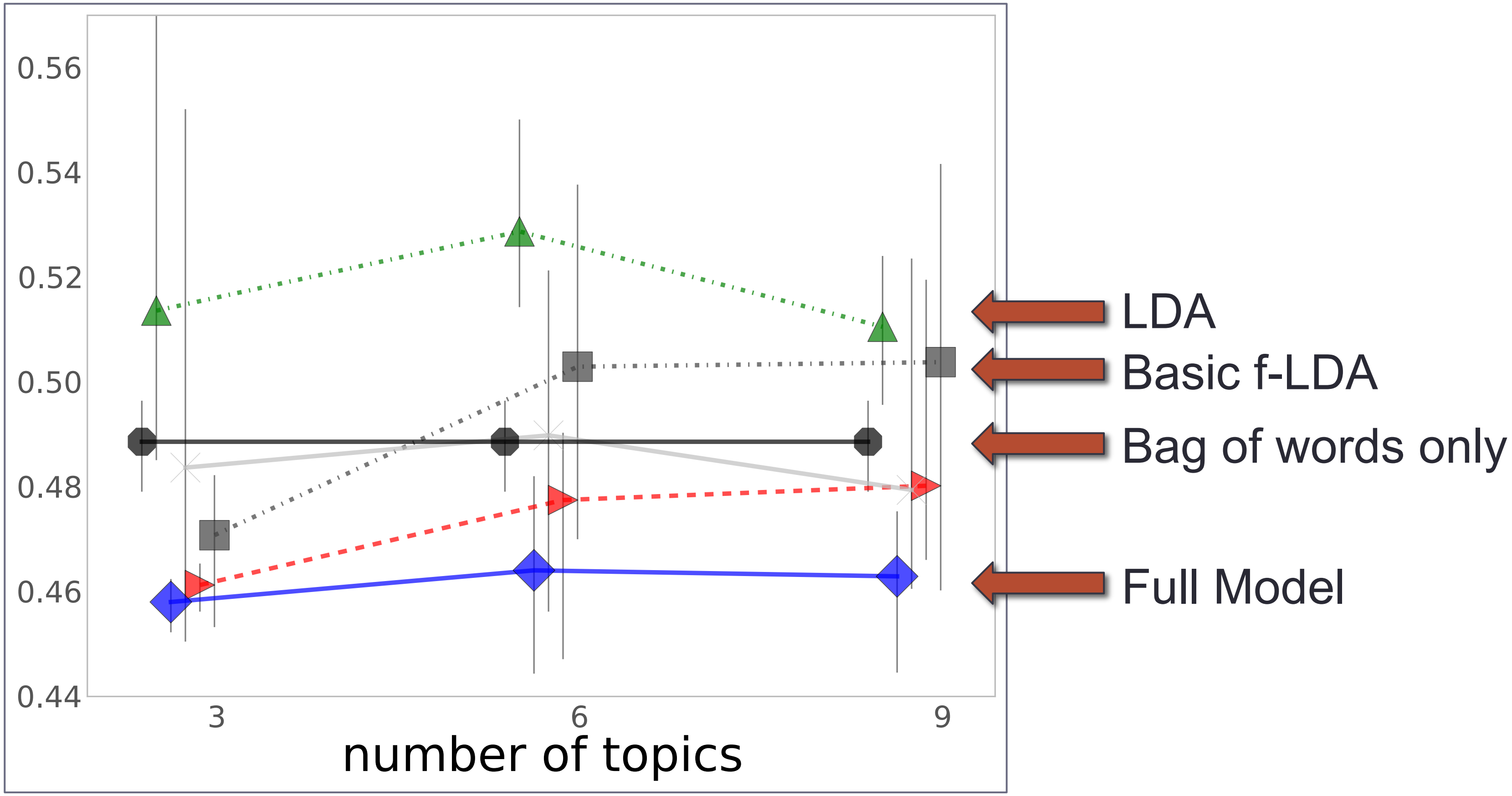
Experiments

- Mean absolute error; distributions only:



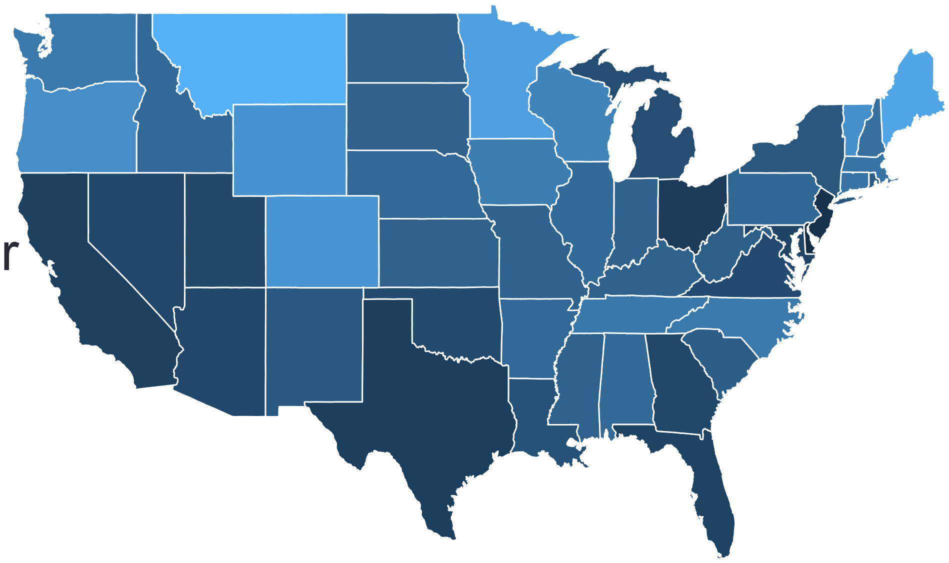
Experiments

- Mean absolute error; distributions + bag of words:



What's Next?

- Goal: validate model on existing data
- Next steps: use model to discover new correlations in topic/sentiment and health statistics
- Interested in geographic analysis, comparing US states
    - Early result: (topic, sentiment) distributions predict hospital post-discharge event rates better than user ratings alone
    - Right: darker states have higher proportion of the Systems topic



Questions?

- Our data sets are freely available
  - `http://www.cebm.brown.edu/static/dr-sentiment.zip`

- Thanks to Andrea Lopez and Urmimala Sarkar for sharing their labeled data