Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Topic Models

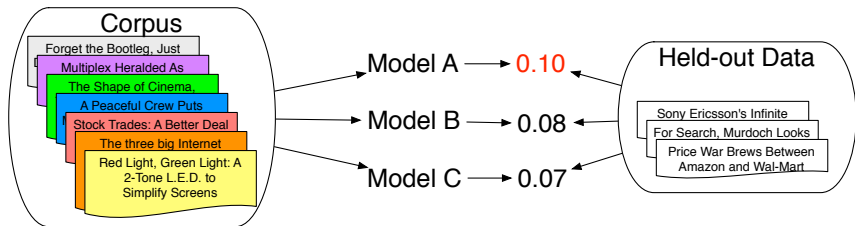Advanced Machine Learning for NLP
Jordan Boyd-Graber
EVALUATION

Evaluation
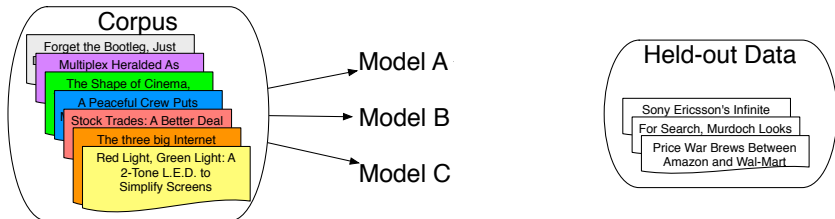
$$P(\boldsymbol{w} \mid \boldsymbol{w}', \boldsymbol{z}', \alpha\boldsymbol{m}, \beta\boldsymbol{u}) =$$

$$\sum_{\boldsymbol{z}} P(\boldsymbol{w}, \boldsymbol{z} \mid \boldsymbol{w}', \boldsymbol{z}', \alpha\boldsymbol{m}, \beta\boldsymbol{u})$$

How you compute it is important too

Measures predictive power, not what the topics are

$$P(\boldsymbol{w} \mid \boldsymbol{w}', \boldsymbol{z}', \alpha \boldsymbol{m}, \beta \boldsymbol{u}) =$$
$$\sum_{\boldsymbol{z}} P(\boldsymbol{w}, \boldsymbol{z} \mid \boldsymbol{w}', \boldsymbol{z}', \alpha \boldsymbol{m}, \beta \boldsymbol{u})$$

How you compute it is important too

TOPIC 1

computer,
technology,
system,
service, site,
phone,
internet,
machine

TOPIC 2

sell, sale,
store, product,
business,
advertising,
market,
consumer

TOPIC 3

play, film,
movie, theater,
production,
star, director,
stage

1 Take the highest probability words from a topic

Original Topic

dog, cat, horse, pig, cow

1 Take the highest probability words from a topic

Original Topic

dog, cat, horse, pig, cow

2 Take a high-probability word from another topic and add it

Topic with Intruder

dog, cat, apple, horse, pig, cow

Word Intrusion

1 Take the highest probability words from a topic

Original Topic

dog, cat, horse, pig, cow

2 Take a high-probability word from another topic and add it

Topic with Intruder

dog, cat, apple, horse, pig, cow

3 We ask users to find the word that doesn't belong

Hypothesis

If the topics are interpretable, users will consistently choose true intruder

Word Intrusion

Word Intrusion

1 / 10					
crash	accident	board	agency	tibetan	safety

2 / 10					
commercial	network	television	advertising	viewer	layoff

3 / 10					
arrest	crime	inmate	pitcher	prison	death

4 / 10					
hospital	doctor	health	care	medical	tradition

1 / 10		Reveal additional response			
crash	accident	board	agency	tibetan	safety

2 / 10					
commercial	network	television	advertising	viewer	layoff

3 / 10					
arrest	crime	inmate	pitcher	prison	death

4 / 10					
hospital	doctor	health	care	medical	tradition

- Order of words was shuffled
- Which intruder was selected varied
- Model precision: percentage of users who clicked on intruder

New York Times, 50 LDA Topics
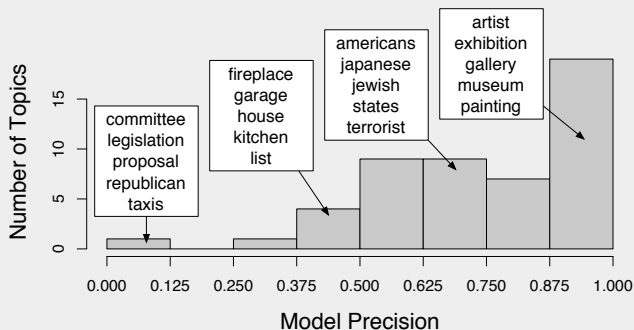
Model Precision: percentage of correct intruders found
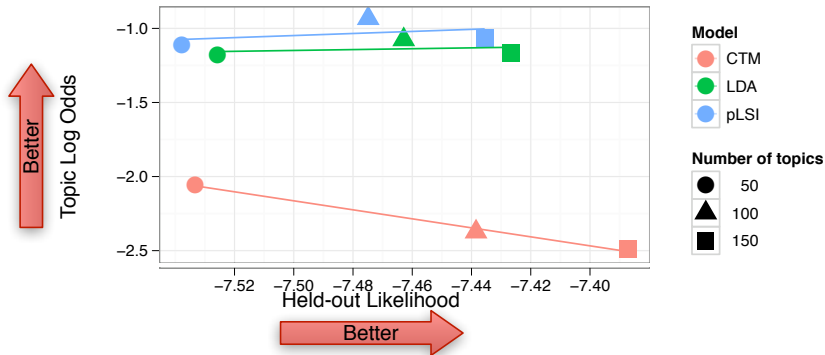
Model Precision on New York Times
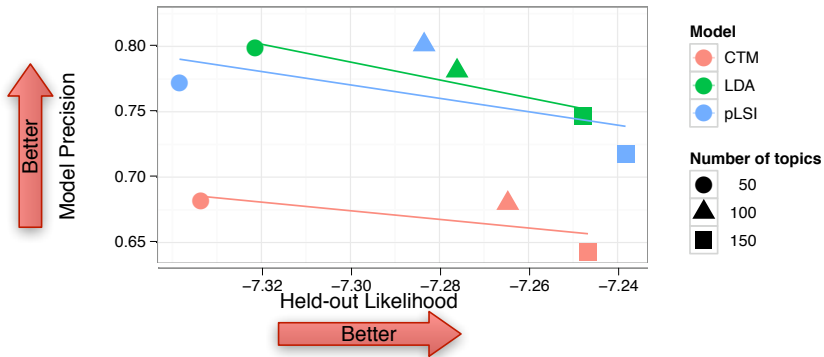
within a model, higher likelihood $\neq$ higher interpretability

Topic Log Odds on Wikipedia

across models, higher likelihood $\neq$ higher interpretability

- Classification
- Machine Translation
- Political Polarization/Framing

- Classification
- Machine Translation
- Political Polarization/Framing