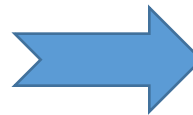
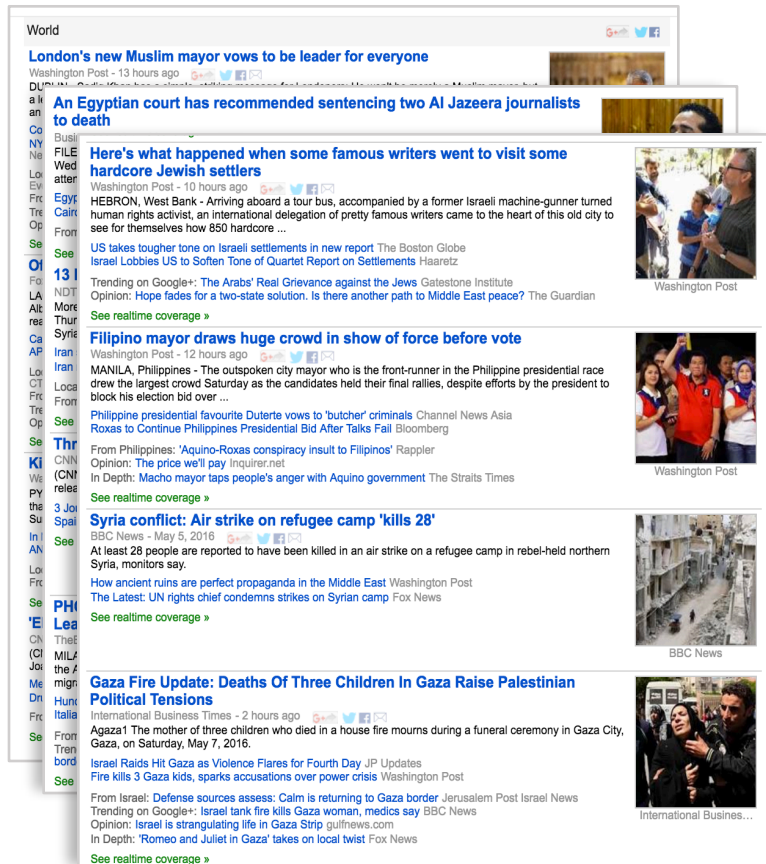


Interpretable Machine Learning: Lessons from Topic Modeling

Michael Paul
University of Colorado Boulder

CHI HCML
May 8, 2016

Topic Modeling



attorney
federal
charged
filed
court
fraud
indictment
investigation

air
flight
plane
airlines
pilots
eastern
airline
airport

united
states
trade
nations
world
countries
european
international

voters
vote
votes
campaign
democratic
candidate
state
election

nicaragua
government
rebels
contras
sandinista
ortega
sandinistas
chamorro

sales
percent
billion
share
quarter
earnings
last
first

...

...

...

from: <http://www.cs.princeton.edu/~blei/lda-c/ap-topics.pdf>

Topic Modeling

Topics

gene 0.04
dna 0.02
genetic 0.01
...

life 0.02
evolve 0.01
organism 0.01
...

brain 0.04
neuron 0.02
nerve 0.01
...

data 0.02
number 0.02
computer 0.01
...

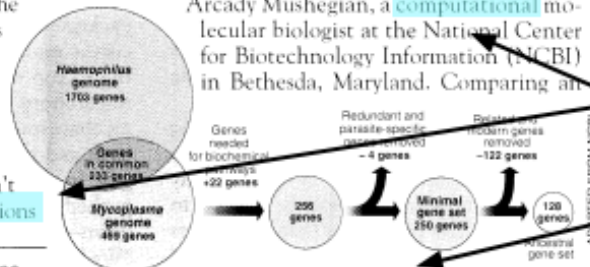
Documents

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**

"are not all that far apart," especially in comparison to the 75,000 **genes** in the human genome, notes Siv Andersson at 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** molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

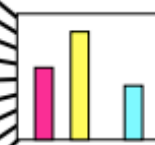


* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

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

SCIENCE • VOL. 272 • 24 MAY 1996

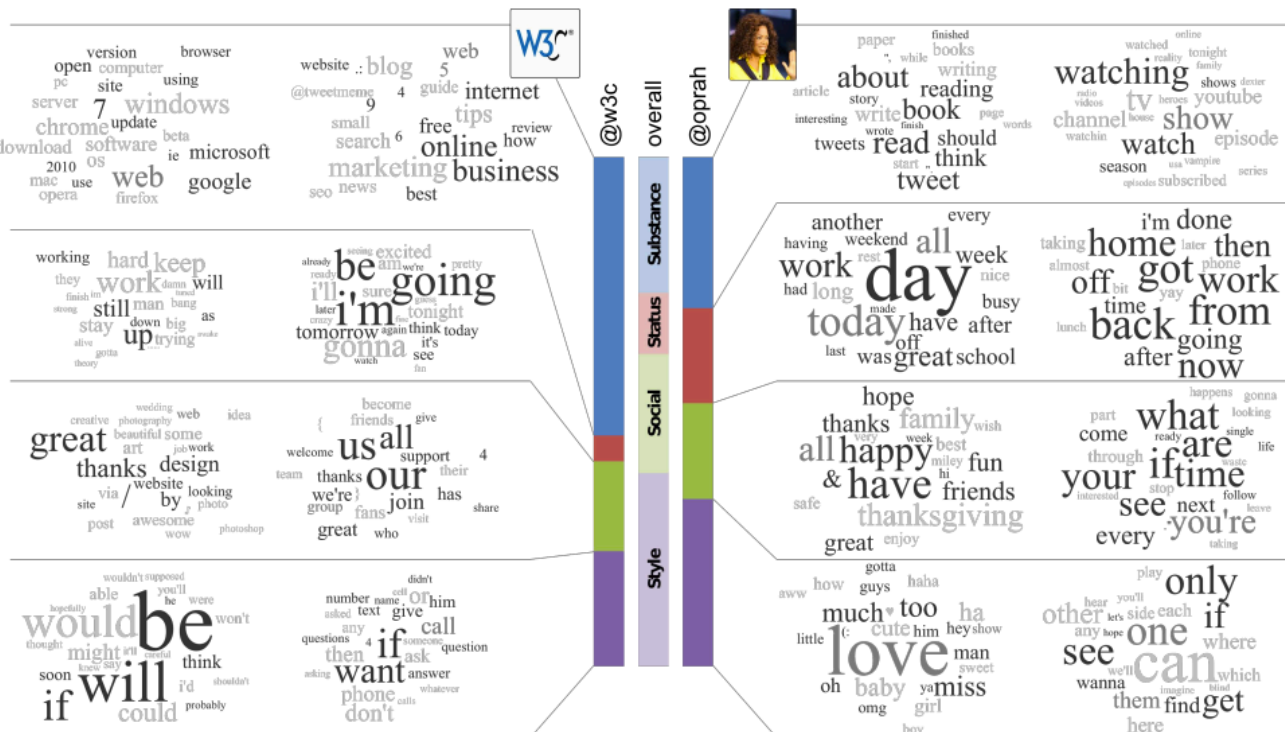
Topic proportions and assignments



Topic Modeling: Uses

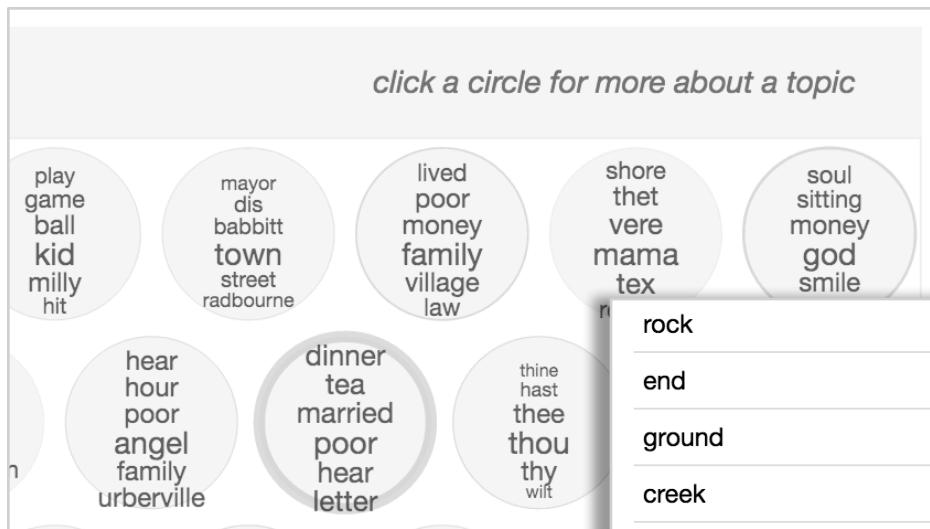
- Identifying main themes in text

“What do people talk about on Twitter?”



Topic Modeling: Uses

- Identifying main themes in text
- Exploring/navigating corpora



from: <http://jgoodwin.net/htb/>

rock	
end	
ground	
creek	
fire	
canyon	
hear	
gun	
lost	

Top documents

Document

Hart, William S. *Told under a white oak tree*. Boston: and New York, Houghton Mifflin company;1922.

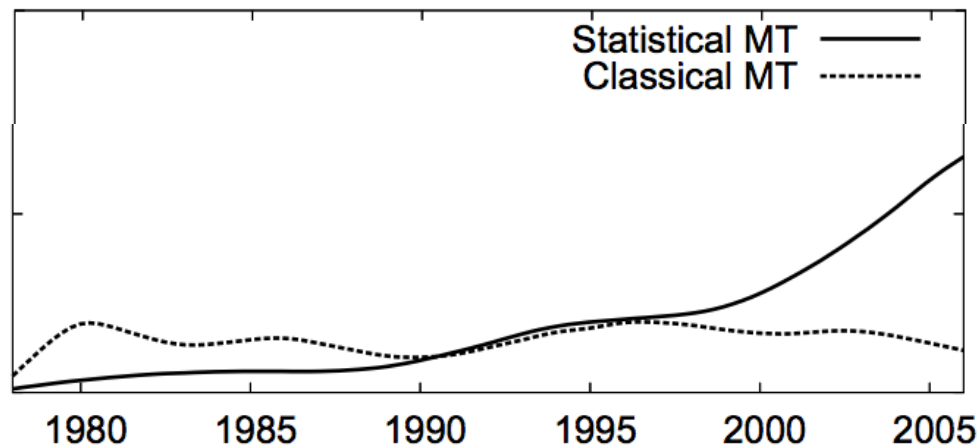
Thompson, Ruth. *Comrades of the desert*. San Francisco, Calif.: Harr Wagner Pub. Co., 1922.

Coolidge, Dane. *Wunpost*. New York: Grosset & Dunlap, 1920.

Bennet, Robert Ames. *Bloom of cactus*. New York: Burt, c1920

Topic Modeling: Uses

- Identifying main themes in text
- Exploring/navigating corpora
- Revealing trends in content



Hall, Jurafsky, Manning (2008) Studying the History of Ideas Using Topic Models. *EMNLP*.

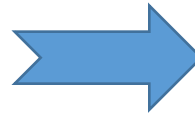
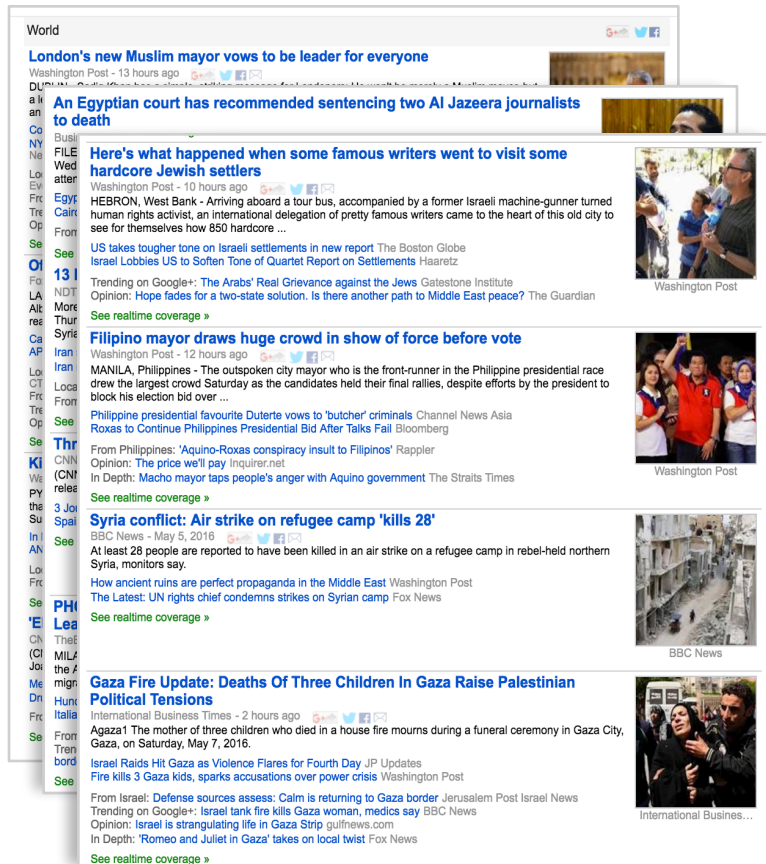
Topic Modeling: Uses

- Identifying main themes in text
- Exploring/navigating corpora
- Revealing trends in content

These uses assume topic models are human-interpretable

Interpretability in Topic Models

Topics aren't always easy to decipher...



door
peace
table
i
california
day
hand
fast

...

hes
i
frank
years
time
good
tree
calls

...

ms
thompson
day
jackpot
hospital
eight
pm
time

...

from: <http://www.cs.princeton.edu/~blei/lda-c/ap-topics.pdf>

Interpretability in Topic Models

How does the topic modeling community deal with interpretability?

Two key areas:

- **Evaluation**
- **Training**

Interpretability in Topic Models: Evaluation

How to evaluate the quality of a topic?

Commonly used concept: *coherence*

Coherent	
space	health
earth	disease
moon	aids
science	virus
scientist	vaccine
light	infection
nasa	hiv
mission	cases
planet	infected
mars	asthma

Incoherent	
dog	king
moment	bond
hand	berry
face	bill
love	ray
self	rate
eye	james
turn	treas
young	byrd
character	key

Newman, Lau, Grieser, Baldwin (2010) Automatic Evaluation of Topic Coherence. *NAACL*.

Interpretability in Topic Models: Evaluation

How to evaluate the quality of a topic?

- Coherence

- Human judgments: *word intrusion*

- Spot the out-of-place word:
 - nasa
 - mission
 - planet
 - dog
 - mars

Chang, Gerrish, Wang, Boyd-Graber, Blei (2009) Reading Tea Leaves: How Humans Interpret Topic Models. *NIPS*.

Interpretability in Topic Models: Evaluation

How to evaluate the quality of a topic?

- **Coherence**
- Human judgments: *word intrusion*
- Metrics based on co-occurrence statistics
 - Similarity of all pairs of words in a topic

$$C(t; V^{(t)}) = \sum_{m=2}^M \sum_{l=1}^{m-1} \log \frac{D(v_m^{(t)}, v_l^{(t)}) + 1}{D(v_l^{(t)})}$$

Mimno, Wallach, Talley, Leenders, McCallum (2011)
Optimizing Semantic Coherence in Topic Models. *EMNLP*.

Interpretability in Topic Models: Training

How to train topic models in a way that will be more interpretable?

- Priors to incorporate human preferences/constraints

Andrzejewski, Zhu, Craven (2009) Incorporating Domain Knowledge into Topic Modeling via Dirichlet Forest Priors. *ICML*.

- Priors to encourage co-occurrence patterns

Mimno, Wallach, Talley, Leenders, McCallum (2011) Optimizing Semantic Coherence in Topic Models. *EMNLP*.

- Interactive topic modeling

Hu, Boyd-Graber, Satinoff, Smith (2013) Interactive Topic Modeling. *Machine Learning*.

Beyond Topic Modeling

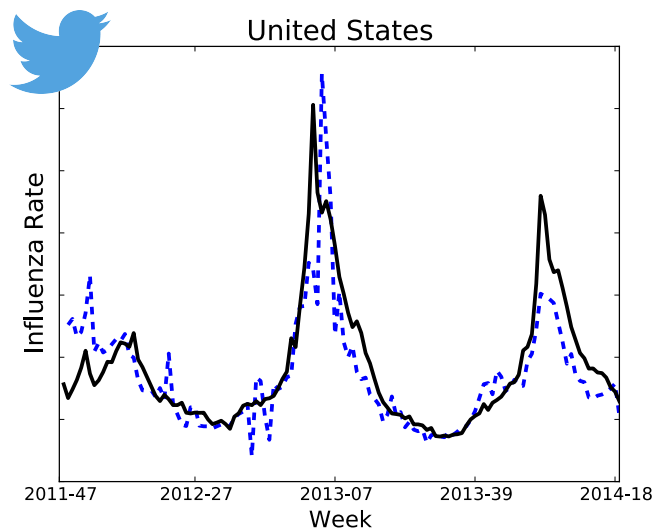
Can other areas of machine learning benefit from the topic modeling approach to interpretability?

Is coherence a desirable property in other types of models?

Beyond Topic Modeling

Poor coherence = poor predictive performance?

Example: Modeling flu prevalence from Twitter



christmas
sick
flu
strong
processing
snow
new
want
hard
better
body
best
coughing
festivities
eve

Top predictors of
regression model

Conclusion

Takeaway:

The topic modeling community has a growing body of research on making models more interpretable

This research should be incorporated into the larger context of interpretable machine learning

What's next?

- Coherence as an evaluation metric for other machine learning models
- Training models to be coherent