Distributional Semantics

Machine Learning: Jordan Boyd-Graber
University of Maryland
SLIDES ADAPTED FROM YOAV GOLDBERG AND OMER LEVY

What's wrong with PMI?

- PMI-based methods prefer rare words
- E.g., closest to "king"

- Jeongjo (Koryo), Adulyadej (Chakri), Coretta (MLK)
- Hard to scale
- Doesn't work as well?

Hyperparameters Matter

- Preprocessing (word2vec)
  - Dynamic Context Windows
  - Subsampling
  - Deleting Rare Words
- Postprocessing (GloVe)
  - Adding Context Vectors
- Association Metric (SGNS)
  - Shifted PMI
  - Context Distribution Smoothing

Hyperparameters Matter

- Preprocessing (word2vec)
  - Dynamic Context Windows
  - Subsampling
  - Deleting Rare Words
- Postprocessing (GloVe)
  - Adding Context Vectors
- Association Metric (SGNS)
  - Shifted PMI
  - Context Distribution Smoothing

Dynamic Context Windows

saw a furry little wampimuk hiding in the tree

word2vec:	1/4	2/4	3/4	4/4	4/4	3/4	2/4
GloVe:	1/4	1/3	1/2	1/1	1/1	1/2	1/3
Aggressive:	1/8	1/4	1/2	1/1	1/1	1/2	1/4

The Word-Space Model *(Sahlgren, 2006)*

Adding Context Vectors

- Skip-Gram Negative Sampling creates word vectors w
- ... and context vectors c
- Pennington et al. (2014) use $w + c$ to represent word
- Levy et al. (2015) find that data size and preprocessing account for most (if not all) of difference

Smoothing

- Introduced in word2vec for negative sampling ($\alpha = 0.75$)

$$\hat{P}_{\alpha}(c) = \frac{\#(c)^{\alpha}}{\sum_{c'} \#(c)^{\alpha}} \tag{1}$$

- For PMI, helps remove bias toward rare words

Smoothing

- Introduced in word2vec for negative sampling ($\alpha = 0.75$)

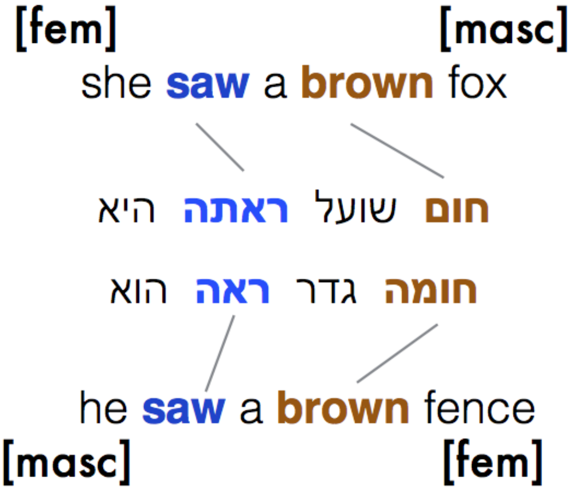
$$\hat{P}_\alpha(c) = \frac{\#(c)^\alpha}{\sum_{c'} \#(c)^\alpha} \tag{1}$$

- For PMI, helps remove bias toward rare words
- And makes it about as good as word2vec

Rant on Evaluation

- Analogy and Similarity aren't that useful
- Find a real-world task and optimize for that
- Innovation is still possible
- Just getting better word vectors is a fruitless cottage industry
- Always tune baseline hyperparameters (and recognize what the hyperparameters are)

Other Languages are Harder



Other Languages are Harder

וכשמהבית

and when from the house

בצל

in shadow

בצל

onion

Other Languages are Harder

ספר

book(N). barber(N). counted(V). tell!(V). told(V).

חומה

brown (feminine, singular)
wall (noun)
her fever (possessed noun)

Takeaway

- Word representations very important
- Future: continuous representations in more complicated models
- Future: document representations