Word embeddings

CS 485, Fall 2023
Applications of Natural Language Processing
https://people.cs.umass.edu/~brenocon/cs485_f23/

Brendan O'Connor
College of Information and Computer Sciences
University of Massachusetts Amherst

- Distributional semantics:
  - a word's meaning is related to how it's used
  - we approximate that from its context distribution in a corpus
  - **word embeddings**: we can reduce this dimensionality into, say, 100 latent dimensions of meaning (matrix factorization: LSA or SGNS)

- Today: So what do you get from word embeddings / distributional info?
  - Lookup similar words (with what function?)
  - Automatically cluster words by syntax?/topic?/meaning?
  - "Bag of embeddings" model for text classification
  - Exploratory analysis of both docs and words

Euclidean Distance

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

Issue: Vector length depends on frequency. More frequent words will have longer vectors.

Cosine Similarity

$$s(x, y) = \frac{x \cdot y}{|x||y|}$$

Only depends on vector angle

Range:

Non- negative vectors & cosine similarity

Pre-trained embeddings

- Demo!
- Widely useful. But make sure you know what you're getting!
  - Examples: GLOVE, fasttext, word2vec, etc.
  - Is the corpus similar to what you care about?
  - Should you care about the *data*?

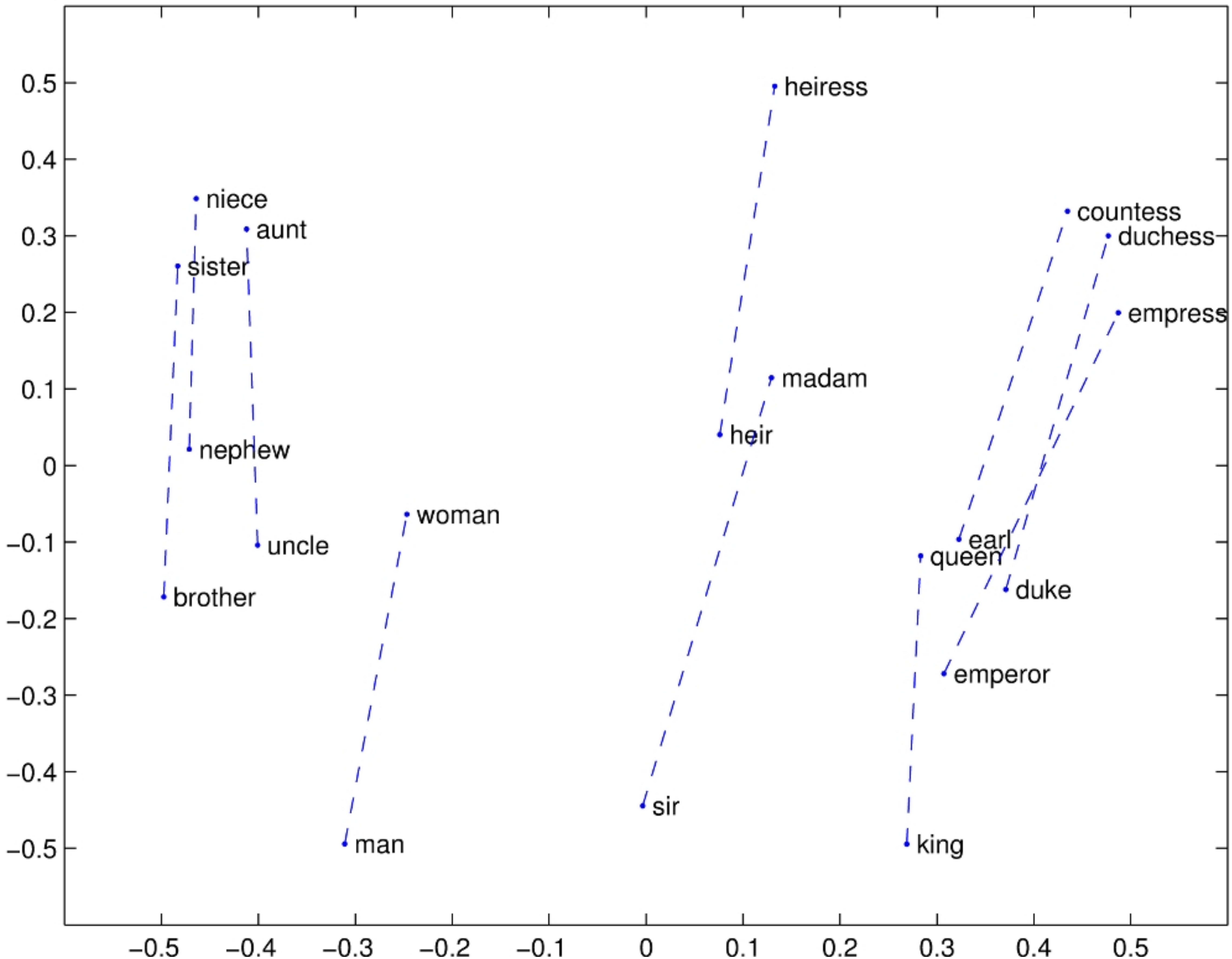
Alternate/mis- spellings

- Distributional methods are really good at this
- Twitter-trained word clusters: http://www.cs.cmu.edu/~ark/TweetNLP/cluster_viewer.html
    - See also: GLOVE website has Twitter-trained word embeddings

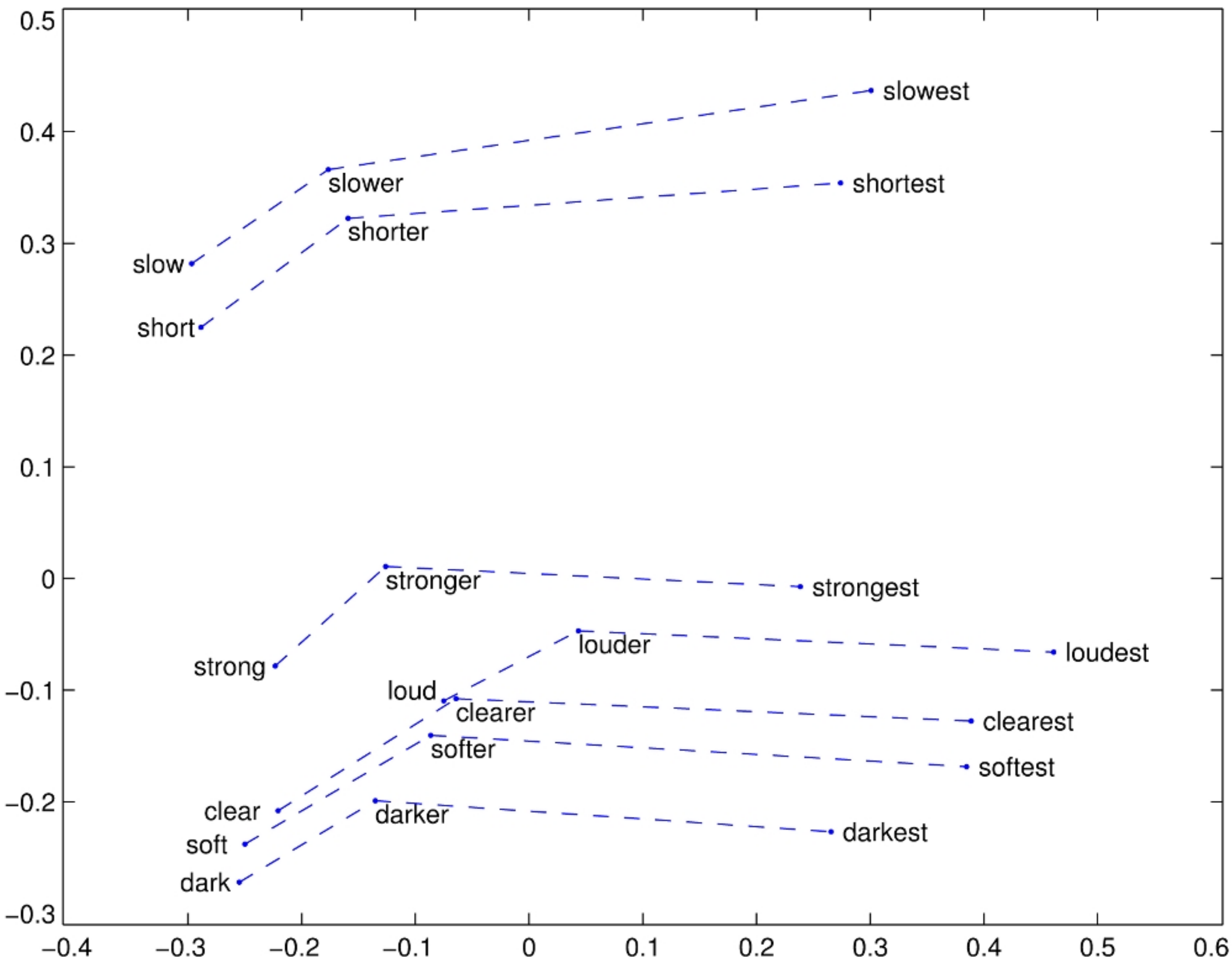
Evaluating embeddings

- Intrinsic evaluations
    - Compare embeddings' word pair similarities to human judgments
        - TOEFL: "*Levied* is closest to *imposed*, *believed*, *requested*, *correlated*"
        - Numerical similarity judgments (e.g. Wordsim-353)
    - Attempt to look at structure of the embedding space, such as analogies
        - Controversial; see Linzen 2016
- Extrinsic evaluation: use embeddings in some task

https://nlp.stanford.edu/projects/glove/



https://nlp.stanford.edu/projects/glove/



Application: keyword expansion

- I have a few keywords for my task. Are there any I missed?
- Automated or semi-automated new terms from embedding neighbors

- Other non-embedding lexical resources can do this too (e.g. WordNet), but word embeddings typically cover a *lot* of diverse vocabulary

Application: document embedding

- Instead of bag-of-words, can we derive a latent embedding of a document/sentence?
  - "Bag of embeddings" or "averaged word embeddings" representation
  - You can use it just like a BOW logistic regression - it's just a different type of feature vector
  - Pros/cons?

See: Arora et al. 2017

Transfer learning

- Sparsity problems for traditional bag-of-words
- Labeled datasets are small … but *unlabeled* data is much bigger!

Exploratory usage

- Example: tweets about mass shootings (Demszky et al. 2019)
  1. Average word embeddings => tweet embeddings
  2. Cluster tweets (kmeans)
  3. Interpret clusters' words (closest to centroid)

Topic	**10 Nearest Stems**
news (19%)	break, custodi, #breakingnew, #updat, confirm, fatal, multipl, updat, unconfirm, sever
investigation (9%)	suspect, arrest, alleg, apprehend, custodi, charg, accus, prosecutor, #break, ap
shooter's identity & ideology (11%)	extremist, radic, racist, ideolog, label, rhetor, wing, blm, islamist, christian
victims & location (4%)	bar, thousand, california, calif, among, los, southern, veteran, angel, via
laws & policy (14%)	sensibl, regul, requir, access, abid, #gunreformnow, legisl, argument, allow, #guncontolnow
solidarity (13%)	affect, senseless, ach, heart, heartbroken, sadden, faculti, pray, #prayer, deepest
remembrance (6%)	honor, memori, tuesday, candlelight, flown, vigil, gather, observ, honour, capitol
other (23%)	dude, yeah, eat, huh, gonna, ain, shit, ass, damn, guess

Table 1: Our eight topics (with their average proportions across events) and nearest-neighbor stem embeddings to the cluster centroids. Topic names were manually assigned based on inspecting the tweets.

Embeddings reflect cultural bias

Bolukbasi, Tolga, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." In *Advances in Neural Information Processing Systems*, pp. 4349-4357. 2016.

Ask "Paris : France :: Tokyo : x"

- x = Japan

Ask "father : doctor :: mother : x"

- x = nurse

Ask "man : computer programmer :: woman : x"

- x = homemaker

huge concern for NLP systems deployed in the real world that use embeddings!

Occupations		Adjectives	
Man	Woman	Man	Woman
carpenter	nurse	honorable	maternal
mechanic	midwife	ascetic	romantic
mason	librarian	amiable	submissive
blacksmith	housekeeper	dissolute	hysterical
retired	dancer	arrogant	elegant
architect	teacher	erratic	caring
engineer	cashier	heroic	delicate
mathematician	student	boyish	superficial
shoemaker	designer	fanatical	neurotic
physicist	weaver	aimless	attractive

Table 7: Top occupations and adjectives by gender in the Google News embedding.

Changes in framing: adjectives associated with Chinese

Garg, Nikhil, Schiebinger, Londa, Jurafsky, Dan, and Zou, James (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644

1910	1950	1990
Irresponsible	Disorganized	Inhibited
Envious	Outrageous	Passive
Barbaric	Pompous	Dissolute
Aggressive	Unstable	Haughty
Transparent	Effeminate	Complacent
Monstrous	Unprincipled	Forceful
Hateful	Venomous	Fixed
Cruel	Disobedient	Active
Greedy	Predatory	Sensitive
Bizarre	Boisterous	Hearty