Word embedding

(Some slides come from Geoffrey Hinton’s lectures)

Outline

- Motivation
- What is word embedding?
- A semantic task
- Bias in word embedding

N-gram LM

- Given a sentence $w_1\ w_2 \dots w_n$, how to estimate $P(w_1 \dots w_n)$?

- The Markov independence assumption:
  $P(w_n \mid w_1, \dots, w_{n-1})$ depends only on the previous k words.

- $P(w_1 \dots w_n)$
  $= P(w_1) * P(w_2|w_1) * \dots P(w_n \mid w_1, \dots, w_{n-1})$
  $\approx P(w_1)*P(w_2|w_1)* \dots P(w_n \mid w_{n-k+1}, \dots, w_{n-1})$

- 0^{th} order Markov model: unigram model
- 1^{st} order Markov model: bigram model
- 2^{nd} order Markov model: trigram model
- …

Limitation of n-gram LM

- It does not understand the similarities between words:
  - Ex: cat/dog, garden/yard, Friday/Monday, Seattle/LA, king/queen

  ➔ Represent each word as a feature vector

- We cannot use a bigger context (i.e., large n) because there are too many parameters to store and most ngrams will be unseen.

  ➔ Represent the context as a vector

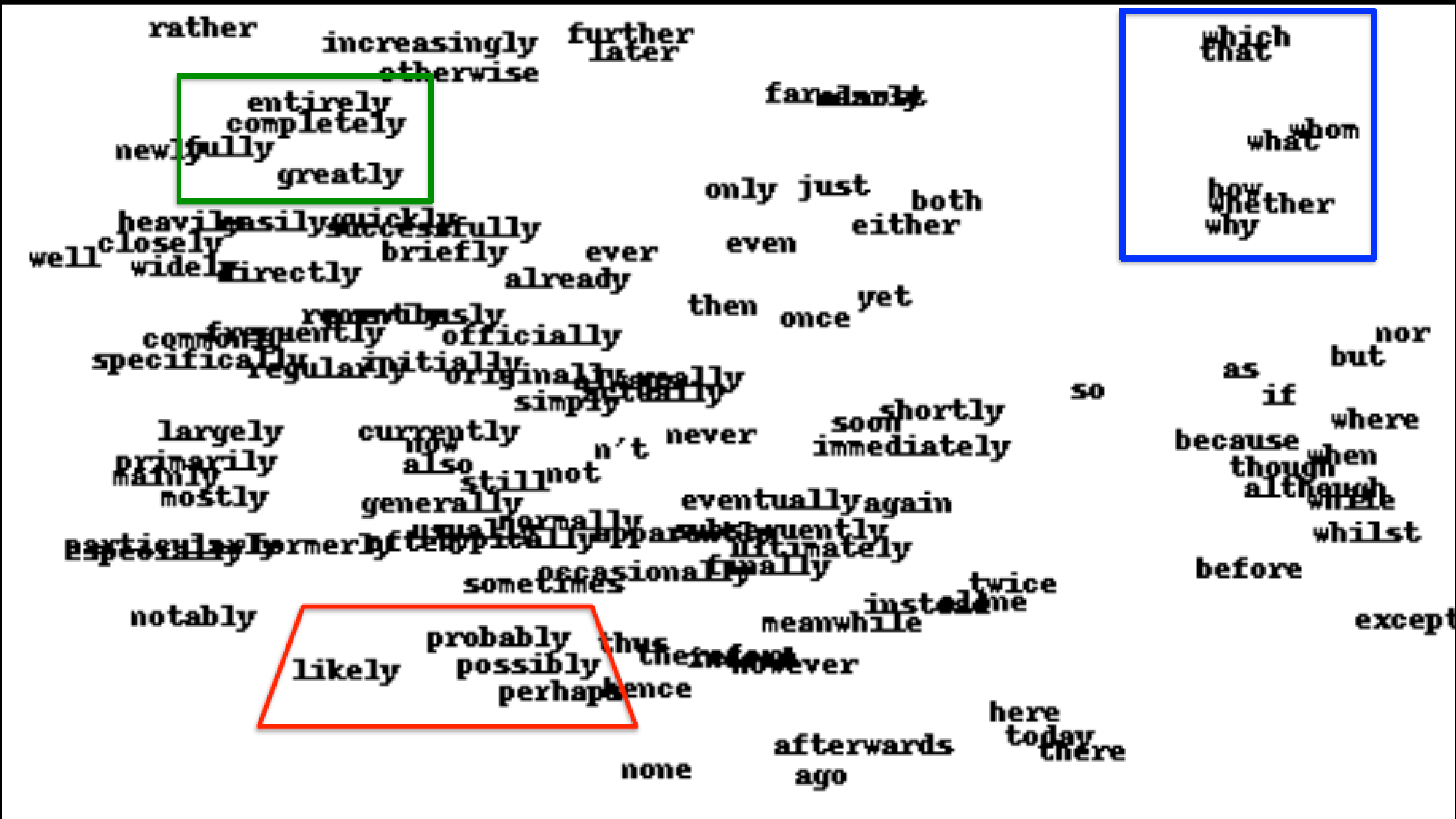
- Sentences have structures.

- …

What is word embedding?

- Word embedding: Represent a word as a vector
  - Similar words should have similar embeddings.
- Pretrained word embeddings: word embeddings learned in one task and then used for other tasks.
  - Pretrained vs. learned from scratch: Learning word embedding requires a large amount of data and can be very expensive.
- Transfer learning: transferring the learnings of one task to another
  - Learning could be weights or embedding.
  - Pretrained word embeddings are a form of Transfer Learning

Part of a 2-D map of the 2500 most common words


rather
increasingly
further
later
otherwise
which
that
entirely
completely
newly
fully
greatly
whom
what
only
just
both
how
whether
why
either
heavily
easily
quickly
successfully
closely
well
widely
directly
briefly
ever
even
already
then
once
yet
officially
specifically
regularly
initially
nor
but
as
so
if
simply
shortly
soon
where
largely
currently
now
also
n't
never
immediately
because
when
though
although
while
primarily
mainly
still
not
mostly
generally
eventually
again
whilst
normally
ultimately
finally
occasionally
sometimes
before
twice
notably
instead
meanwhile
except
probably
likely
possibly
perhaps
hence
thus
however
here
today
there
afterwards
none
ago

Many studies on neural network

- Early studies: (Hinton 1986), (Pollack 1990), (Elman 1991), etc.
- Feed-forward networks: (Bengio et al., 2003; 2006)
- Recurrent neural networks: (Mikolov et al., 2010; 2011; 2013)
- Now: tons of papers in 2014-now

Where is the word embedding used?

- Answer semantic questions: e.g., A:B is like C:D
- LM
- Classification
- Sequence labeling: POS tagging, chunking, NER, etc.
- Structure prediction: parsing
- Question answering
- …

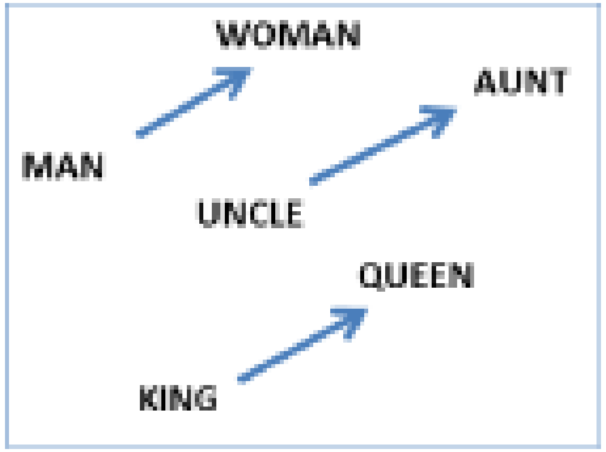
Outline

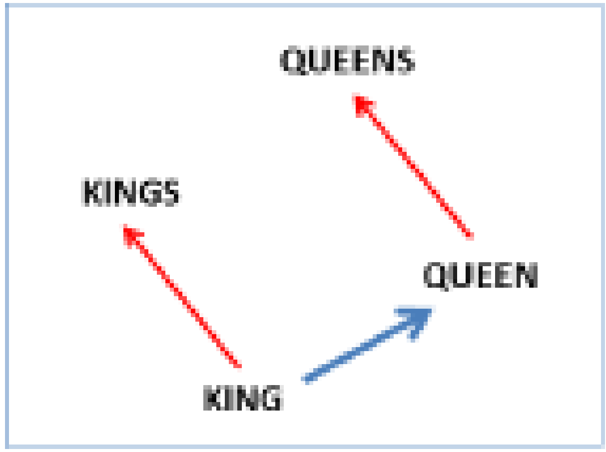
- Motivation
- What is word embedding?
- A semantic task
- Bias in word embedding

A semantic task (Mikolov et al., 2013)

- Task: “A:B C:D”
  - Training: given a corpus of text, learn word embedding
  - Test: given A, B, and C, find D
  - Evaluation: percentage of examples with the correct D.

- Examples:
  - Good:better   rough:___
  - Year:years   law:___
  - See:sees   return:___
  - come:go   borrow:___





Algorithm

- A:B is like C:D

  ➔ $x_b - x_a = x_d - x_c$

  ➔ $x_b - x_a + x_c = x_d$

- Represent each word w as a word vector x_w
- Compute $y = x_b - x_a + x_c$
- Find $w^* = \arg\max_w \text{sim}(x_w, y)$

Results

Method	Adjectives	Nouns	Verbs	All
LSA-80	9.2	11.1	17.4	12.8
LSA-320	11.3	18.1	20.7	16.5
LSA-640	9.6	10.1	13.8	11.3
RNN-80	9.3	5.2	30.4	16.2
RNN-320	18.2	19.0	45.0	28.5
RNN-640	21.0	25.2	54.8	34.7
RNN-1600	**23.9**	**29.2**	**62.2**	**39.6**

Bias in word embedding

- (Bolukbasi et al., 2016): Man is to computer programmer as woman is to homemaker? Debiasing word embeddings

- Ex: man : woman = king : queen
- But man : woman = programmer : homemaker

- Many studies on the bias in word embedding. For more, see the Bias/Discrimination papers at https://faculty.washington.edu/ebender/2021_575/

Extreme *she*	**Extreme *he***
1. homemaker	1. maestro
2. nurse	2. skipper
3. receptionist	3. protege
4. librarian	4. philosopher
5. socialite	5. captain
6. hairdresser	6. architect
7. nanny	7. financier
8. bookkeeper	8. warrior
9. stylist	9. broadcaster
10. housekeeper	10. magician

	Gender stereotype *she*-*he* analogies	
sewing-carpentry	registered nurse-physician	housewife-shopkeeper
nurse-surgeon	interior designer-architect	softball-baseball
blond-burly	feminism-conservatism	cosmetics-pharmaceuticals
giggle-chuckle	vocalist-guitarist	petite-lanky
sassy-snappy	diva-superstar	charming-affable
volleyball-football	cupcakes-pizzas	lovely-brilliant
	Gender appropriate *she*-*he* analogies	
queen-king	sister-brother	mother-father
waitress-waiter	ovarian cancer-prostate cancer	convent-monastery

Summary

- Word embedding is to represent a word as a vector, and it is often used in the input layer of neural networks.

- Pretrained word embeddings are used in many NLP systems; however, be aware of the bias in the embeddings.

- There are many algorithms for learning word embeddings.