Understanding X, Part 2: Word Embeddings

Elliott Ash

Max Planck Summer School 2017

Word Embeddings

- Word embeddings:
  - a hot topic in NLP since arrival of Word2Vec in 2013.
    - refers to a class of statistical models that **represent words or phrases as points in a vector space**.
- The key idea is to represent the meaning of words by the neighbor words – their **contexts**.
- I tend to use "word embeddings" and "word2vec" interchangeably, although word2vec technically refers to the most famous example of a word embedding model.

Word Embeddings

- Word embeddings:
  - a hot topic in NLP since arrival of Word2Vec in 2013.
  - refers to a class of statistical models that **represent words or phrases as points in a vector space**.
- The key idea is to represent the meaning of words by the neighbor words – their **contexts**.
- I tend to use "word embeddings" and "word2vec" interchangeably, although word2vec technically refers to the most famous example of a word embedding model.

Word Embeddings

- Word embeddings:
  - a hot topic in NLP since arrival of Word2Vec in 2013.
  - refers to a class of statistical models that **represent words or phrases as points in a vector space**.
- The key idea is to represent the meaning of words by the neighbor words – their **contexts**.
- I tend to use "word embeddings" and "word2vec" interchangeably, although word2vec technically refers to the most famous example of a word embedding model.

Word Embeddings

- Word embeddings:
  - a hot topic in NLP since arrival of Word2Vec in 2013.
  - refers to a class of statistical models that **represent words or phrases as points in a vector space**.
- The key idea is to represent the meaning of words by the neighbor words – their **contexts**.
- I tend to use "word embeddings" and "word2vec" interchangeably, although word2vec technically refers to the most famous example of a word embedding model.

Outline

Outline

Why word vectors?

- Once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
  - More intriguingly, word2vec algebra can depict conceptual, analogical relationships between words.
    - Consider the analogy: **man is to king as woman is to ____**
    - With word2vec, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

Why word vectors?

- Once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
  - More intriguingly, word2vec algebra can depict conceptual, analogical relationships between words.
    - Consider the analogy: **man is to king as woman is to ____**
    - With word2vec, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

Why word vectors?

- Once words are represented as vectors, we can use linear algebra to understand the relationships between words:
  - Words that are geometrically close to each other are similar: e.g. "student" and "pupil."
  - More intriguingly, word2vec algebra can depict conceptual, analogical relationships between words.
    - Consider the analogy: **man is to king as woman is to ____**
    - With word2vec, we have

$$vec(king) - vec(man) + vec(woman) \approx vec(queen)$$

How are word embeddings different from topic models?

- Ben Schmidt:
  - Topic models reduce words to core meanings to understand documents more clearly.
  - Word embedding models ignore information about individual documents to better understand the relationships between words.

How are word embeddings different from topic models?

- Ben Schmidt:
  - Topic models reduce words to core meanings to understand documents more clearly.
  - Word embedding models ignore information about individual documents to better understand the relationships between words.

Word2Vec in Social Science?

- If you type "word2vec economics" in google scholar, you get nothing relevant.
- I will show you some interesting examples later in the slides.
- This is still a very new, under-used tool with a lot of potential for social-science applications

Word2Vec in Social Science?

- If you type "word2vec economics" in google scholar, you get nothing relevant.
- I will show you some interesting examples later in the slides.
- This is still a very new, under-used tool with a lot of potential for social-science applications

Outline

Word Function $\longleftrightarrow$ Word Neighbors

- "The meaning of a word is its use in the language"

  - Ludwig Wittgenstein, *Philosophical Investigation*, 1953

- "You shall know a word by the company it keeps"

  - J.R. Firth, Papers in Linguistics,1957

Word Function $\longleftrightarrow$ Word Neighbors

- "The meaning of a word is its use in the language"

  - Ludwig Wittgenstein, *Philosophical Investigation*, 1953

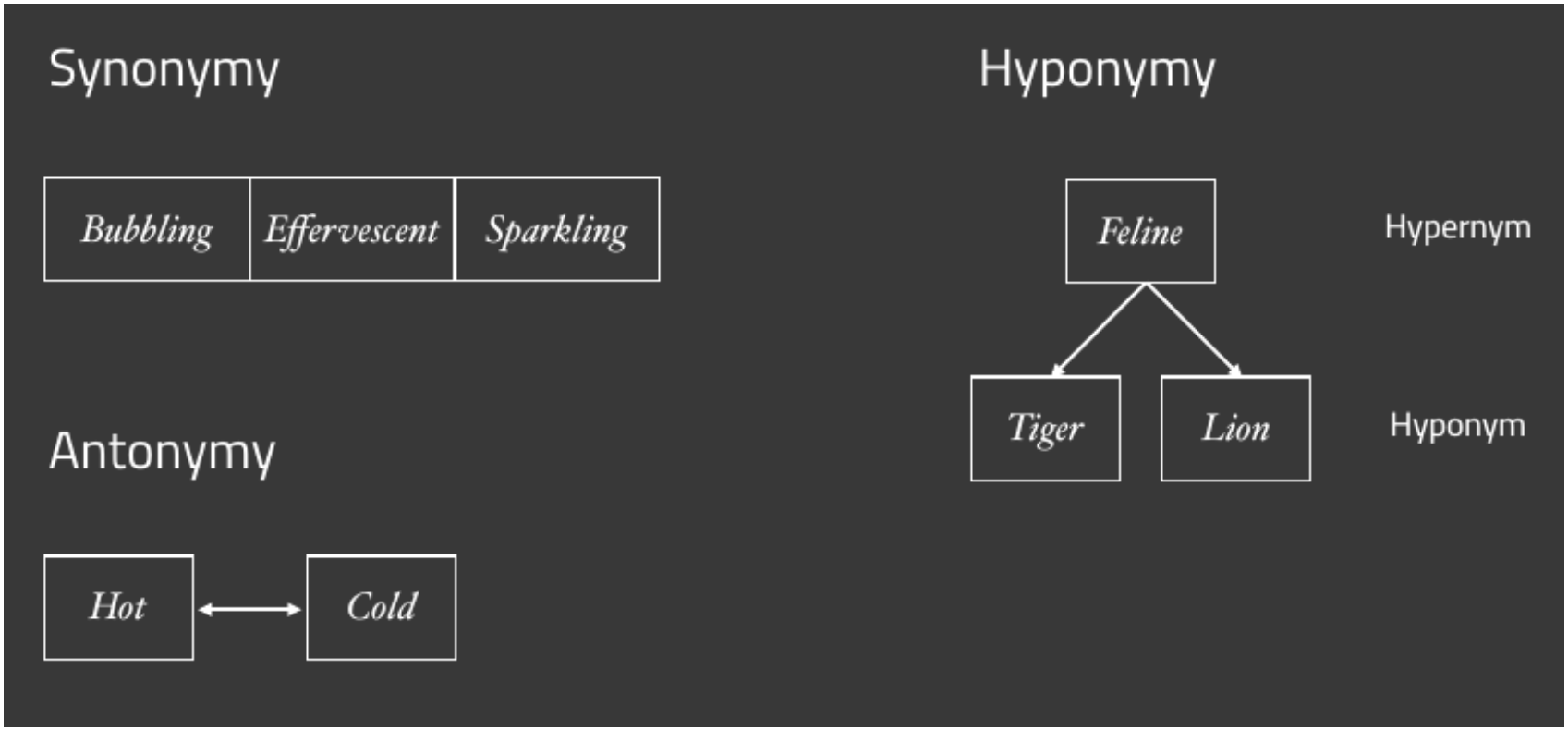
- "You shall know a word by the company it keeps"

  - J.R. Firth, Papers in Linguistics,1957

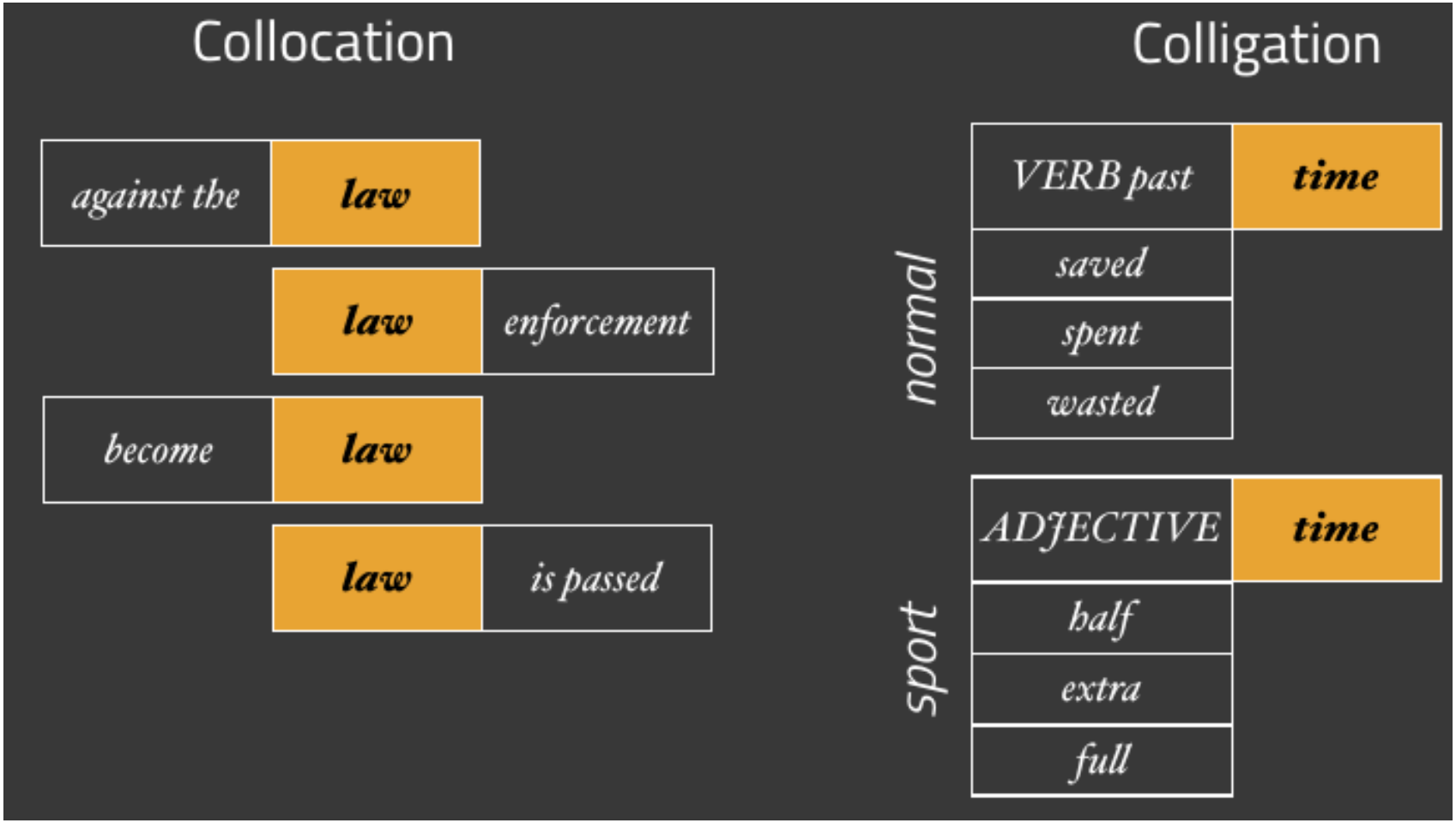
I've never seen this word before, but...

- He filled the **wampimuk**, passed it around and we all drunk some
- We found a little, hairy **wampimuk** sleeping behind the tree

Linguistic Relations



Collocational Relations



Similarity vs. Relatedness

- Semantic **similarity**: words sharing salient attributes / features
  - synonymy (car / automobile)
  - hypernymy (car / vehicle)
  - co-hyponymy (car / van / truck)
- Semantic **relatedness**: words semantically associated without necessarily being similar
  - function (car / drive)
  - meronymy (car / tire)
  - location (car / road)
  - attribute (car / fast)

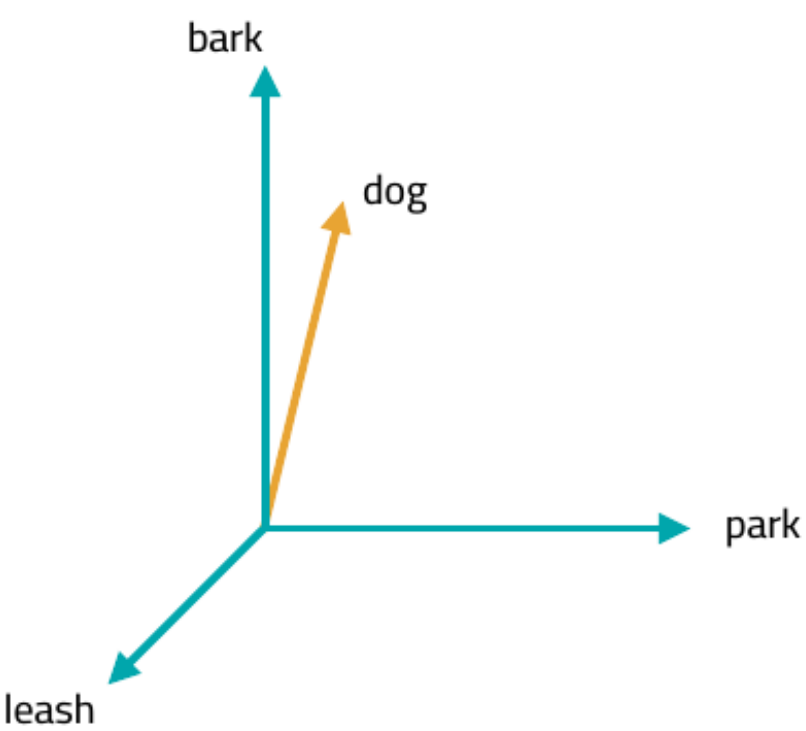
(Budansky and Hirst, 2006)

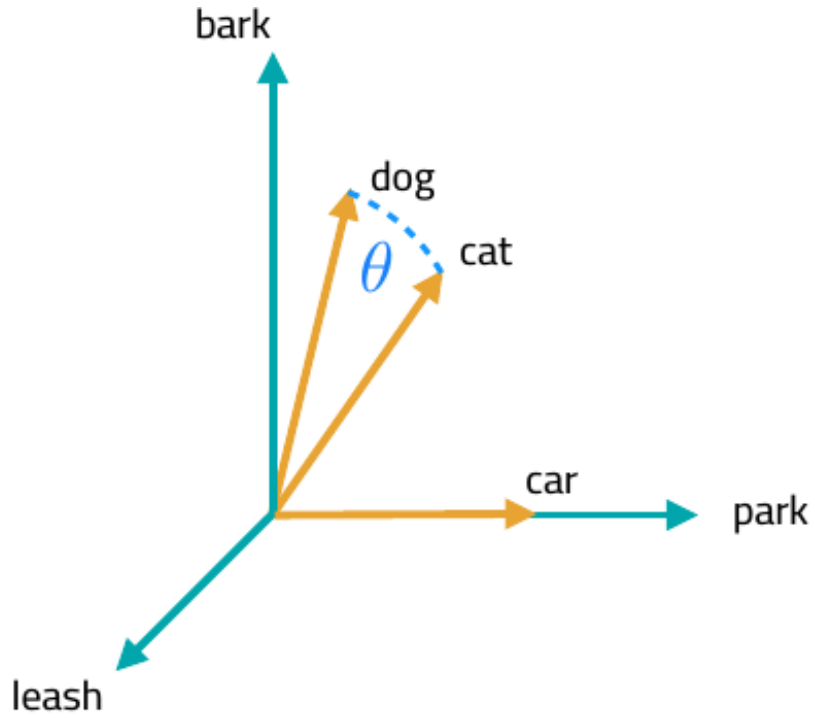
Similarity vs. Relatedness

- Semantic **similarity**: words sharing salient attributes / features
  - synonymy (car / automobile)
  - hypernymy (car / vehicle)
  - co-hyponymy (car / van / truck)
- Semantic **relatedness**: words semantically associated without necessarily being similar
  - function (car / drive)
  - meronymy (car / tire)
  - location (car / road)
  - attribute (car / fast)

(Budansky and Hirst, 2006)

Words as Vectors





- Use cosine similarity as a measure of relatedness:

$$\cos\theta = \frac{v_1 \cdot v_2}{||v_1||||v_2||}$$

Outline

SGNS: Skip-gram with negative sampling

- When people mention "word2vec", they are usually talking about SGNS: "skip gram with negative sampling."
- This is a particular word-embedding model with good performance on a range of analogy and prediction tasks.
- How does it learn the meaning of the word "fox":

  The quick brown **fox** jumps over the lazy dog

- Word2Vec reads in every example of the word "fox", and learns its relation to nearby words in the context window.
  - A lot of technical details, I can provide references.

SGNS: Skip-gram with negative sampling

- When people mention "word2vec", they are usually talking about SGNS: "skip gram with negative sampling."
- This is a particular word-embedding model with good performance on a range of analogy and prediction tasks.
- How does it learn the meaning of the word "fox":

  The quick brown **fox** jumps over the lazy dog

- Word2Vec reads in every example of the word "fox", and learns its relation to nearby words in the context window.
  - A lot of technical details, I can provide references.

SGNS: Skip-gram with negative sampling

- When people mention "word2vec", they are usually talking about SGNS: "skip gram with negative sampling."
- This is a particular word-embedding model with good performance on a range of analogy and prediction tasks.
- How does it learn the meaning of the word "fox":

  The quick brown **fox** jumps over the lazy dog

- Word2Vec reads in every example of the word "fox", and learns its relation to nearby words in the context window.
  - A lot of technical details, I can provide references.

SGNS: Skip-gram with negative sampling

- When people mention "word2vec", they are usually talking about SGNS: "skip gram with negative sampling."
- This is a particular word-embedding model with good performance on a range of analogy and prediction tasks.
- How does it learn the meaning of the word "fox":

  The quick brown **fox** jumps over the lazy dog

- Word2Vec reads in every example of the word "fox", and learns its relation to nearby words in the context window.
  - A lot of technical details, I can provide references.

SGNS: Skip-gram with negative sampling

- When people mention "word2vec", they are usually talking about SGNS: "skip gram with negative sampling."
- This is a particular word-embedding model with good performance on a range of analogy and prediction tasks.
- How does it learn the meaning of the word "fox":

  The quick brown **fox** jumps over the lazy dog

- Word2Vec reads in every example of the word "fox", and learns its relation to nearby words in the context window.
  - A lot of technical details, I can provide references.

Outline

gensim implemention

- gensim's implementation of word2vec has the same benefits as the LDA implementation:
  - intuitive, streaming, and fast/parallellized
  - [demo_code.py]

Pre-trained word embeddings

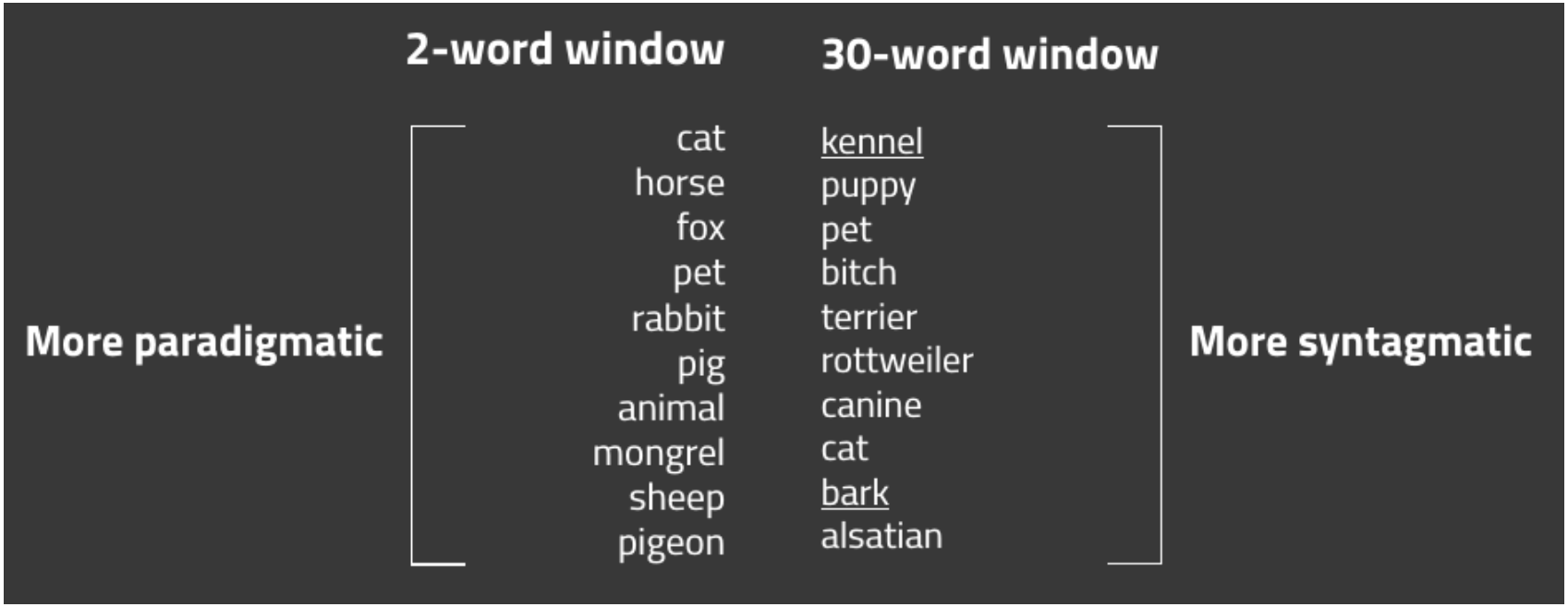
- Good vectors require a big corpus.
- If your corpus is small, you might want to use a pre-trained model.
- spaCy's English model has:
  - one million vocabulary entries
  - 300-dimensional vectors
  - trained on the Common Crawl corpus
  - using the GloVe algorithm
- spaCy has vector models available for other languages, included German.
  - at command line, run "`python -m spacy download de`"
- [demo_code.py]

Pre-trained word embeddings

- Good vectors require a big corpus.
- If your corpus is small, you might want to use a pre-trained model.
- spaCy's English model has:
  - one million vocabulary entries
  - 300-dimensional vectors
  - trained on the Common Crawl corpus
  - using the GloVe algorithm
- spaCy has vector models available for other languages, included German.
  - at command line, run "`python -m spacy download de`"
- [demo_code.py]

Outline

Most similar words to dog, depending on window size



- Small windows pick up substitutable words; large windows pick up topics.

Evaluation of Word Embeddings

- Intrinsic:
  - evaluate word-pairs similarities $\rightarrow$ compare with similarity judgments given by humans
  - evaluate on analogy tasks ("Paris is to France as Tokyo is to ___")
- Extrinsic:
  - use the vectors in a downstream task (classification, translation, ...) and evaluate the final performance on the task

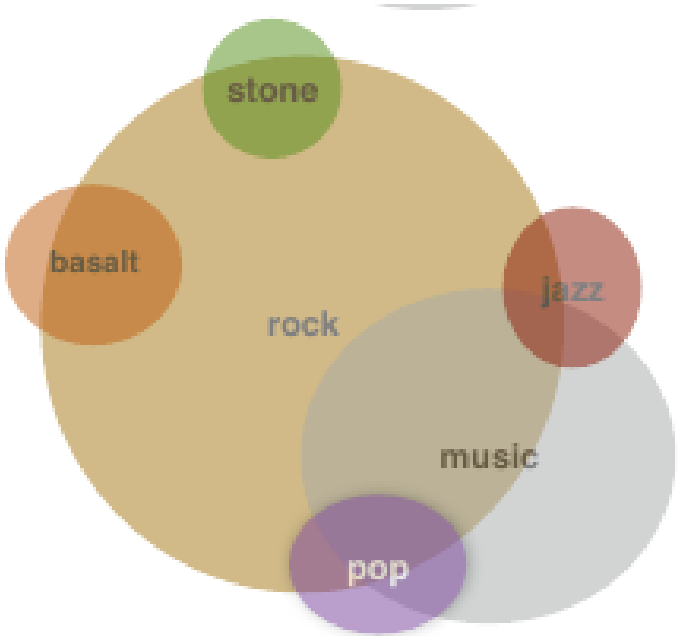
Evaluation of Word Embeddings

- Intrinsic:
  - evaluate word-pairs similarities $\rightarrow$ compare with similarity judgments given by humans
  - evaluate on analogy tasks ("Paris is to France as Tokyo is to ___")
- Extrinsic:
  - use the vectors in a downstream task (classification, translation, ...) and evaluate the final performance on the task

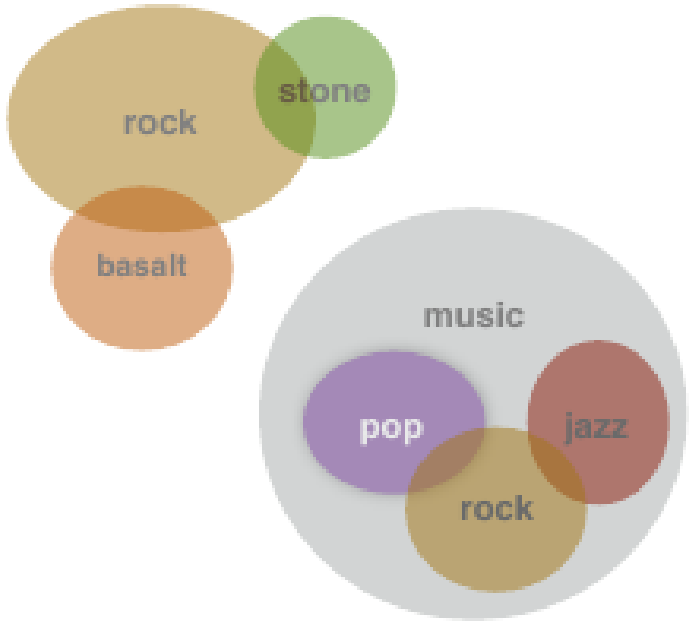
The Future: Words as distributions (rather than points)

- Instead of representing words as points, represent them as distributions
  - Mean and variance in every dimension
- Multimodal mixes a fixed number of gaussian distributions.

Gaussian Embeddings



Multimodal Distributions



Outline

Outline

- We replicated a spectrum of known biases, as measured by the Implicit Association Test, using a widely used, purely statistical machine-learning model trained on a standard corpus of text from the World Wide Web. . .

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...
- Attribute words:
  - man, male, ...
  - woman, female, ...
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...
- Attribute words:
  - man, male, ...
  - woman, female, ...
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Word Embedding Association Test

- Target words:
  - programmer, engineer, scientist, ...
  - nurse, teacher, librarian, ...
- Attribute words:
  - man, male, ...
  - woman, female, ...
- WEAT Test:
  - Compute similarities between all target words and all attribute words
  - Compute mean target-attribute clustering

Example Stimuli

- Targets:
  - **Flowers**: aster, clover, hyacinth, marigold, poppy, azalea, crocus, iris, orchid, rose, bluebell, daffodil, lilac, pansy, tulip, buttercup, daisy, lily, peony, violet, carnation, gladiola, magnolia, petunia, zinnia.
  - **Insects**: ant, caterpillar, flea, locust, spider, bedbug, centipede, fly, maggot, tarantula, bee, cockroach, gnat, mosquito, termite, beetle, cricket, hornet, moth, wasp, blackfly, dragonfly, horsefly, roach, weevil.
- Attributes:
  - **Pleasant**: caress, freedom, health, love, peace, cheer, friend, heaven, loyal, pleasure, diamond, gentle, honest, lucky, rainbow, diploma, gift, honor, miracle, sunrise, family, happy, laughter, paradise, vacation.
  - **Unpleasant**: abuse, crash, filth, murder, sickness, accident, death, grief, poison, stink, assault, disaster, hatred, pollute, tragedy, divorce, jail, poverty, ugly, cancer, kill, rotten, vomit, agony, prison.

Example Stimuli

- Targets:
  - **Flowers**: aster, clover, hyacinth, marigold, poppy, azalea, crocus, iris, orchid, rose, bluebell, daffodil, lilac, pansy, tulip, buttercup, daisy, lily, peony, violet, carnation, gladiola, magnolia, petunia, zinnia.
  - **Insects**: ant, caterpillar, flea, locust, spider, bedbug, centipede, fly, maggot, tarantula, bee, cockroach, gnat, mosquito, termite, beetle, cricket, hornet, moth, wasp, blackfly, dragonfly, horsefly, roach, weevil.
- Attributes:
  - **Pleasant**: caress, freedom, health, love, peace, cheer, friend, heaven, loyal, pleasure, diamond, gentle, honest, lucky, rainbow, diploma, gift, honor, miracle, sunrise, family, happy, laughter, paradise, vacation.
  - **Unpleasant**: abuse, crash, filth, murder, sickness, accident, death, grief, poison, stink, assault, disaster, hatred, pollute, tragedy, divorce, jail, poverty, ugly, cancer, kill, rotten, vomit, agony, prison.

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Results

- Pleasant vs. Unpleasant?
  - Flowers vs. Insects
  - Musical instruments vs. weapons.
  - European-American names vs. African-American names
- Male names vs. Female names?
  - Career words (e.g. professional, corporation, ...) vs. family words (e.g. home, children, ...)
  - Math/science words vs arts words
  - Gender-associated careers

Bolukbasi et al (NIPS 2016)

- Geometrically, **gender bias is first shown to be captured by a direction in the word embedding.**
- Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding.
- Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words receptionist and female, while maintaining desired associations such as between the words queen and female.

Bolukbasi et al (NIPS 2016)

- Geometrically, **gender bias is first shown to be captured by a direction in the word embedding**.
- Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding.
- Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words receptionist and female, while maintaining desired associations such as between the words queen and female.

Bolukbasi et al (NIPS 2016)

- Geometrically, **gender bias is first shown to be captured by a direction in the word embedding**.
- Second, gender neutral words are shown to be linearly separable from gender definition words in the word embedding.
- Using these properties, we provide a methodology for modifying an embedding to remove gender stereotypes, such as the association between the words receptionist and female, while maintaining desired associations such as between the words queen and female.

Outline

Which laws are close to "sales tax"?

Outline

Rudolph and Blei (2017)

- Train word embeddings on the U.S. Congressional Record, 1858-2009.
- Dynamic word embeddings model:
  - Captures how the meaning of words evolves over time.
  - The innovation is to include "year" in the embedding model, and allow word vectors to drift over time (following a random walk).

Rudolph and Blei (2017)

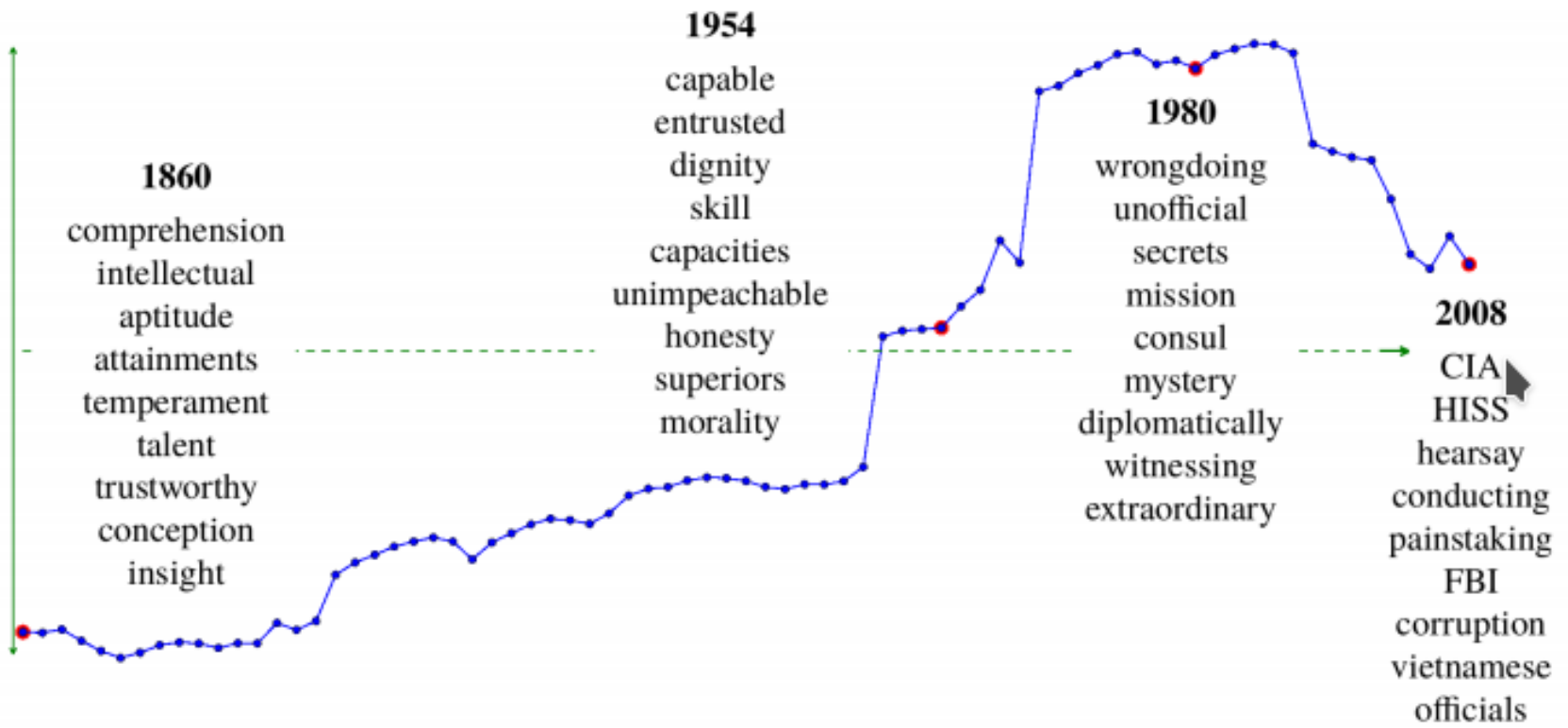
- Train word embeddings on the U.S. Congressional Record, 1858-2009.
- Dynamic word embeddings model:
  - Captures how the meaning of words evolves over time.
  - The innovation is to include "year" in the embedding model, and allow word vectors to drift over time (following a random walk).

Meaning Changes

computer	
1858	**1986**
computer	computer
draftsman	software
draftsmen	computers
copyist	copyright
photographer	technological
computers	innovation
copyists	mechanical
janitor	hardware
accountant	technologies
bookkeeper	vehicles

bush	
1858	**1990**
bush	bush
barberry	cheney
rust	nonsense
bushes	nixon
borer	reagan
eradication	george
grasshoppers	headed
cancer	criticized
tick	clinton
eradicate	blindness

Drift in word "intelligence"



Drift in word "prostitution"

prostitution				
1930	**1945**	**1962**	**1988**	**1990**
prostitution	prostitution	prostitution	harassment	prostitution
punishing	indecent	indecent	intimidation	servitude
immoral	vile	harassment	prostitution	harassment
bootlegging	immoral	intimidation	counterfeit	intimidation
riotous	induces	sexual	illegal	trafficking
forbidden	incite	vile	trafficking	harassing
anarchists	abortion	counterfeit	indecent	apprehended
assemblage	forbid	anarchists	disregard	killings
forbid	harboring	mobs	anarchists	labeled
abet	assemblage	lawbreakers	punishing	naked

Extensions

- Rudolph and co-authors are working on other covariates, for example legal jurisdictions.