



LECTURE 8: LEXICAL SEMANTICS, DISTRIBUTIONAL SEMANTICS AND WORD EMBEDDINGS

Mark Granroth-Wilding

EXERCISE FROM BEFORE

- We did this exercise
- Assume you:
 - are a computer
 - have database of logical/factual world knowledge
 - have lots of rules/statistics about English
- What steps are involved in:
 - analysing this textual input?
 - extracting & **encoding relevant information?**
 - answering the question?

Meaning representation

A robotic co-pilot developed under DARPA's ALIAS programme has already flown a light aircraft.

What agency has created a computer that can pilot a plane?

1

EXERCISE

In small groups (3-4)

A robotic co-pilot developed under DARPA's ALIAS programme has already flown a light aircraft.

What agency has created a computer that can pilot a plane?

- What types of meaning / meaning representation that we've seen are useful here?
- What other aspects of meaning do we need?

LEXICAL SEMANTICS

Question

What is a *good* way to *remove* wine stains?

Text available

Salt is a *great* way to *eliminate* wine stains

- Different words may have similar/related meanings
- Relationships between words can be highly complex
- Essential for using information from text
- Difficult for **broad-domain** systems

3

LEXICAL SEMANTICS

- Representing **word meaning**
- Some difficult issues:
 - Representation
 - Acquiring broad-domain knowledge
 - Polysemy: same word, multiple meanings
 - Multiword expressions

EXERCISE

In small groups (3-4)

A **robotic** co-pilot **developed** under DARPA's ALIAS programme has already flown a light **aircraft**.

What **agency** has created a **computer** that can pilot a **plane**?

- What aspects of meaning of these **words** must we capture?
- How might these be represented computationally?
- Think about difficulties in:
 - formalizing meaning
 - specifying/acquiring/learning representations

5

THESAURUS

- Dictionary of **synonyms** and closely related words
- **WordNet**: large machine-readable thesaurus (and more)
- Complex network of relationships between words
- Type of **network MR**



6

SEMANTIC RELATIONS: HYPONYMY

- Relationship between meanings of words
- **Hyponymy: Is-A** relation

dog is-a *animal*

- *dog* hyponym of *animal*
- *animal* hypernym of *dog*
- Forms a taxonomy
- Mainly works for **concrete** nouns
democracy, thought, sticky?

7

SEMANTIC RELATIONS: MERONYMY

- Another relationship: part-whole

air bag part-of *car*

- *air bag* meronym of *car*

8

WORDNET: SYNSETS

- **Synset**: 'synonym set' \simeq *sense*
- Unit of organization in WordNet
- **Synonyms** grouped into synset:
 - car* .n.01 = {*car*, *auto*, *automobile*}
- **Polysemous** words belong to multiple:
 - car* .n.04 = {*car*, *elevator car*}
- Connected by semantic relations: **hyponymy**, **meronymy**, ...

9

WORDNET

<https://wordnet.princeton.edu/>

- Hand-built ontology
- 117k synsets
- Primarily for English
- *Global WordNet*
 - Lists WN-like resources in different languages
 - Currently 77
 - Varying quality, size, licenses
- *Open Multilingual WordNet*
 - Linked/merged WNs for different languages
 - Currently 34

10

PROBLEMS WITH WORDNET(S)

- Mainly English: limited other languages
- Hand-built: large, but still limited coverage
 - Many **multiword expressions**, but far from complete:
 - take a break*, *pay attention*
 - Language changes: **neologisms**
 - greenwashing*, *Brexit*, *barista*
 - **Names**
 - Mikkeli*, *Microsoft*
- Fine-grained senses

11

PROBLEMS WITH WORDNET(S)

- Mainly English: limited other languages
- Hand-built: large, but still limited coverage
 - Many **multiword expressions**, but far from complete:
 - take a break*, *pay attention*
 - Language changes: **neologisms**
 - greenwashing*, *Brexit*, *barista*
 - **Names**
 - Mikkeli*, *Microsoft*
- Fine-grained senses

• S: (n) **top** (the upper part of anything) "the mower cuts off the tops of the grass"; "the title should be written at the top of the first page"
 • S: (n) **top**, **top side**, **upper side**, **upside** (the highest or uppermost side of anything) "put your books on top of the desk"; "only the top side of the box was painted"
 • S: (n) **peak**, **crest**, **top**, **tip**, **summit** (the top or extreme point of something (usually a mountain or hill)) "the view from the peak was magnificent"; "they clambered to the tip of Monadnock"; "the region is a few molecules wide at the summit"
 • S: (n) **top**, **top of the inning** (the first half of an inning; while the visiting team is at bat) "a relief pitcher look over in the top of the fifth"
 • S: (n) **acme**, **height**, **elevation**, **peak**, **pinnacle**, **summit**, **superlative**, **meridian**, **top**, **top** (the highest level or degree attainable; the highest stage of development) "his landscapes were deemed the acme of beauty"; "the artist's gifts are at their acme"; "at the height of her career"; "the peak of perfection"; "summer was at its peak"; "...catapulted Einstein to the pinnacle of fame"; "the summit of his ambition"; "so many highest superlatives achieved by man"; "at the top of his profession"
 • S: (n) **top** (the greatest possible intensity) "he screamed at the top of his lungs"

11

EXERCISE: WORD SENSES

In small groups

1 thing. She was talking at a party thrown at Daphne's restaurant in
 2 turned it into the hot dinner-party topic. The comedy is the
 3 the 1983 general election for a party which was uncertain of its policies
 4 to attack the Scottish National Party, who look set to seize Perth and
 5 had been passed to a second party who made a financial decision
 6 by-pass there will be a street party. "Then," he says, "we are going
 7 political tradition and the same party. They are both Anglophilic
 8 he told Tony Blair's modernised party they must not retreat into "warm
 9 "Oh no, I'm just here for the party," they said. "I think it's terrible
 10 A future obliges each party to the contract to fulfil it by
 11 signed by or on behalf of each party to the contract." Mr David N

- How many senses here?
- Can you define them?

12

WORD SENSES

thing. She was talking at a party thrown at Daphne's restaurant in
 turned it into the hot dinner-party topic. The comedy is the
 the 1983 general election for a party which was uncertain of its policies
 to attack the Scottish National Party, who look set to seize Perth and
 had been passed to a second party who made a financial decision
 by-pass there will be a street party. "Then," he says, "we are going
 political tradition and the same party. They are both Anglophilic
 he told Tony Blair's modernised party they must not retreat into "warm
 "Oh no, I'm just here for the party," they said. "I think it's terrible
 A future obliges each party to the contract to fulfil it by
 signed by or on behalf of each party to the contract." Mr David N

- How many senses? Difficult decision
- WordNet is quite fine-grained

13

TRANSLATION SENSES

- WordNet senses defined by hand
- Another way to define senses:
 - different translations → evidence for two senses
- E.g. *interest* ⇒ German:
 - *Zins*: charge for a loan
 - *Anteil*: stake in company
 - *Interesse*: other senses
- Senses of German word may be distinguished in English
- Or in other languages

14

WORD SENSE DISAMBIGUATION

thing. She was talking at a **party**₀ thrown at Daphne's restaurant in the 1983 general election for a **party**₁ which was uncertain of its policies

- Which senses are used here?
- Often useful to distinguish
- But, useful/important:
 - **Search**: only interested in one sense
 - **Topic model**: different senses in different topics
 - Etc...

15

WORD SENSE DISAMBIGUATION (WSD)

- Have predefined senses for words
 - E.g. from WordNet
- **Word sense disambiguation** (WSD) becomes supervised classification task
- Features: context words, other clues seen
- Classes: known senses for word

17

REMINDER: MEANING REPRESENTATIONS

- Formal MRs: logical expressions over entities
 - Focus on logical relations/properties $\exists e, y \text{ Having}(e) \wedge \text{Haver}(e, \text{Speaker})$
 - Mostly assume representations of words, entities, ... given
- Network-based representations: *semantic nets*
 - Network of semantic *relations* between concepts
 - E.g. **WordNet**, **ConceptNet**
- Feature-based representations
 - *Attribute-value* pairs for *objects* or concepts
 - Words, objects, facts, events, ...
- Latent feature-based representations
 - **Vector representations**
 - E.g. **distributional word vectors**, **word embeddings**
 - E.g. **deep learning-based reprs**

ConceptNet
An open, multilingual knowledge graph



19

DISTRIBUTIONAL HYPOTHESIS

- Can infer **meaning** of word from *contexts* it typically occurs in
- Is word meaning anything more than contexts it occurs in?
- **Distributional hypothesis**: words that tend to appear in *similar contexts* have similar meanings

You shall know a word by the company it keeps! – Firth (1957)

21

CLUES TO WORD SENSE

thing. She was talking at a **party**^{beer, invite, host} thrown at Daphne's **restaurant** in the 1983 general **election** for a **party**^{political, minister, manifesto} which was uncertain of its **policies**

Some clues that might help:

- **Surrounding** words
- **Connected** words (grammatically, same doc, ...)
- Document **topic**
- (Disambiguated) **sense of same word** used in same doc

Most also clues to word **meaning**: see later!

16

WORD MEANING

(senses)

- **WordNet**: attempt to encode **meaning** of words in machine-usable form
- Can discover/represent word meaning in other ways
- Statistical approach: **distributional semantics**
- Learn meaning from word's use in text
- No reliance on annotated resources
- Broad coverage
- Apply to any language, domain, ...

18

MEANING FROM CONTEXT

A bottle of **tezgin** is on the table
Everybody likes **tezgin**
Tezgin makes you drunk
We make **tezgin** out of corn

- What is **tezgin**?
- You can probably work it out

20

WORD CONTEXT

Words that tend to appear in *similar contexts* have similar meanings

- Measure **word similarity** by **context similarity**
 - What is a word's **context**?
 - How to measure *typical* contexts?
 - How to *compare* typical contexts?
- Lots of possible answers
- Some capture different aspects of word meaning
- A couple of examples here
 - (1) Context = nearby words
 - (2) Context = syntactically related words

22

COUNTING CONTEXTS

Context = nearby words

53 billion in 1985 the **company** said it had been able acquisition by bond of the **company** s existing bank loans and be repaid by the mining **company** in gold exploration and development of the **company** s gold properties in the mln vs 6 9 mln **company** name is bowater industries plc

celanese canada inc hoechst celanese **corporation** and trade mountain pipe line to iraq the commodity credit **corporation** has accepted a bid for of adjusting the commodity credit **corporation** s ccc discount and premium jordan usda the commodity credit **corporation** ccc accepted eight bonus offers iraq usda the commodity credit **corporation** ccc accepted a bid for

COUNTING CONTEXTS

Context = nearby words, ignoring stopwords

Word	Count
said	2306
dtrs	573
mln	539
pct	397
stock	306
year	289
1986	276
would	272
shares	261
share	257

loss 1 53 billion 1985 **company** said able cut losses id negotiations continuing acquisition bond **company** existing bank loans s repaid mining **company** gold on development **company** gold properties cer mln vs 6 9 mln **company** name bowater indu

Word	Count
credit	59
commodity	57
ccc	49
tax	30
mln	28
said	28
accepted	21
bonus	20
state	20
pct	19

hoechst celanese **corporation** trade mountain p nmodity credit **corporation** accepted bid exp nmodity credit **corporation** ccc discount premium schedules improve nmodity credit **corporation** ccc accepted eight bonus offers nmodity credit **corporation** ccc accepted bid export bonus

COLLECTING CONTEXT COUNTS

- **Window:**
 - Small number of words (5, 10?)
 - Large number (100)
 - Whole document
- Exclude **stopwords**?
- Weight words? Closer → higher?
- Normalize by word?
- More than counts?
 - *said* was common for *company*
 - But also for most other words
 - Frequency *relative to expectation*
 - One way: **mutual information** (e.g. PPMI)

WORD VECTORS

Contexts – features

Words	89	0	202	0	0	45	0	0
	52	0	110	64	0	0	178	0
	0	87	0	71	84	51	0	20
	77	67	0	0	73	0	0	0

- Context word counts → *features* to characterize word
- Construct **vector** for word
- Can compare *words* by comparing vectors
- Word **embedded** in vector space: also called **word embeddings**

WORD MATRIX

Feature (context) counts

Words	89	0	202	0	0	45	0	0
	52	0	110	64	0	10	178	0
	0	87	0	71	84	0	0	20
	77	67	0	0	73	0	0	0

Sparse: mostly 0s

WORD MATRIX

Very high-dimensional

	sea	melody	sail	shanty	lyric	oar	roger	chorus
boat	89	0	202	0	0	45	0	0
pirate	52	0	110	64	0	10	178	0
song	0	87	0	71	84	0	0	20
serenade	77	67	0	0	73	0	0	0

Sparse: mostly 0s

VECTOR SPACES

Features (N)

Data points	89	0	202	0	0	45	0	0
	52	0	110	64	0	0	178	0
	0	87	0	71	84	51	0	20
	77	67	0	0	73	0	0	0

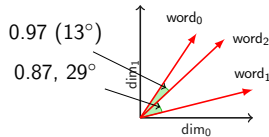
- Method treats **words** as **vectors** (in a matrix)
- **Vector space model** (VSM)
- Representation method extremely general
- Used to formulate many representation tasks
- Commonly applied to represent many things: next lecture

WORD SIMILARITY

Compare angles

- Compare two words to *boat*
- Imagine we just look at all *boat*'s contexts twice
- Knowledge of meaning not changed
- **Angle** compares relative values on dimensions

COSINE SIMILARITY



Cosine similarity: \cos of angle between vectors

$$\cos(\vec{word}_0, \vec{word}_1) = \frac{\vec{word}_0 \cdot \vec{word}_1}{|\vec{word}_0| |\vec{word}_1|} = \frac{\sum_{i=1}^N \vec{word}_{0,i} \vec{word}_{1,i}}{\sqrt{\sum_{i=1}^N \vec{word}_{0,i}^2} \sqrt{\sum_{i=1}^N \vec{word}_{1,i}^2}}$$

Commonly used measure of vector similarity

31

EXERCISE: WORD ASSOCIATION

Individually

<https://presemo.helsinki.fi/nlp2020>



- Go to Presemo
- For each word, enter associated words that come into your mind
 - No specific relation
 - ~10 words
- NB: Words invisible when you've entered them!

- (1) Duck
- (2) Submarine
- (3) Aeroplane

32

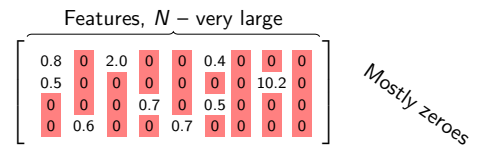
VECTORS IN NEURAL NETWORKS

Aside

- Feed-forward NNs: all representations are vectors
- **Node** contains numeric value
- "Layer" of nodes: vector
- **Latent representations** are vectors
- Can sometimes inspect and compare in this way
- Sometimes capture these intuitions about e.g. similarity
- Example later: *word embeddings* trained with NNs

33

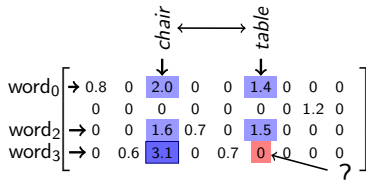
SPARSE VECTORS



- Vectors seen up to now: **sparse**
 - **Problem:**
 - similar words may appear in *related* contexts
 - but have few exact contexts in common
- friend, friends, buddy* ↗
- Want to capture more **abstract** features
 - Single dimension for these

34

SPARSE VECTORS



- One possibility: exploit *correlations* between features
 - *chair* and *table* often contexts for same word
 - Then word seen with *chair*, but not *table*
 - Probably still quite *table-chairy*?
 - Maybe *table* unobserved by chance?
- **Dimensionality reduction:** removes redundancy in columns
- Exploit these correlations
- Correlated dims collapsed

35

DIMENSIONALITY REDUCTION

- **Sparse** matrix → **dense** matrix
- Fewer columns – features
- Mixtures of original dims mapped to new dims
- Apply to word-context matrix
 - Standard methods: e.g. SVD
 - Each word now has vector
 - Might lose some important distinctions (*chair* vs *table* features)
 - Often worth it for generalization

36

DIMENSIONALITY REDUCTION ON COUNTS

- Count vectors are **sparse**
- Top words seem somewhat characteristic
- But, not much actual overlap
- *Generalize* sparse dimensions with **dimensionality reduction**

<i>company</i>		<i>corporation</i>	
Word	Count	Word	Count
said	2306	credit	59
dllrs	573	commodity	57
mln	539	ccc	49
pct	397	tax	30
stock	306	mln	28
year	289	said	28
1986	276	accepted	21
would	272	bonus	20
shares	261	state	20
share	257	pct	19

37

COMPARING WORD VECTORS



- Apply dimensionality reduction
- Compare vectors using **cosine similarity**
- Similar contexts → similar features
 - similar vectors
 - similar meaning

38

VECTOR \neq MEANING

- Distributional word vectors capture *some* aspects of meaning
- Exactly what depends on:
 - features
 - weighting
 - dimensionality reduction
 - comparison metric, ...
- *The meaning of a word is given by its usage* \leftarrow ?
- Vectors also depend on factors in corpus:
 - culture
 - linguistic register
 - document types

39

PROBLEMS WITH COUNT-BASED VECTORS

- Want very large vocabulary
 - *Oxford English Dictionary*: 600k words
 - More if separate vectors for *eat*, *eating*, *ate*, ...

Words (V) many!	Context features						
	0.89	0	2.02	0	0	0.45	0
0.52	0	1.10	0.64	0	0	1.78	
0	0.87	0	0.71	0.84	0.51	0	
0.77	0.67	0	0	0.73	0	0	

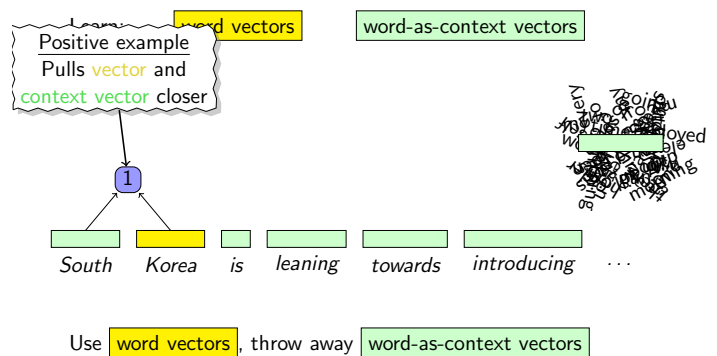
- Dimensionality reduction slow for many words
- Memory inefficient
- Hard to add new words

40

DON'T COUNT, PREDICT!

- Random embedding initialization
- Iterate over huge corpus
- Update vectors iteratively to **predict** context
- Prediction with **neural network**: embedding = layer
- Update using **gradient descent**: adjust embeddings for each observation
- word2vec and many others since

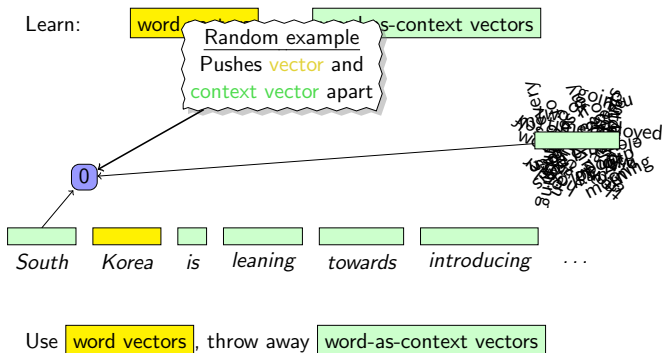
WORD2VEC



41

42

WORD2VEC



42

WORD2VEC

- Training is *fast*: simple network
- Train on many examples: 100B words news text¹
- Vectors for many words: 3M vectors

¹Pre-trained vectors from authors

43

USING WORD EMBEDDINGS

- Exploit word similarity measures, e.g.:
 - Query expansion with related words: nearest vector neighbours
 - Matching similar words: *good* \approx *great*
- Use word representation directly
 - Use **vectors** instead of words as input
 - Benefit: related words look similar to system

44

USING EMBEDDINGS: EXAMPLE

Classifier for named-entity recognition

- Choose NER label for each word
- Classes: O, IPERS, ILOC, ...
- Features: word, previous word, previous label, ...
- Word: *boat*, *pirate*, *ship*, ... V
- Represented as V independent binary features **Num dims**
- But, *boat* \approx *ship*
- Use embedding as input: N real-valued features

45

