School of CIT –
Social Computing
Research Group

Technical
University
of Munich

Natural Language Processing
IN2361

Prof. Dr. Georg Groh

Chapter 18
Word Senses and WordNet

- content is based on [1]
- certain elements (e.g. equations or tables) were taken over or taken over in a modified form from [1]
- citations of [1] or from [1] are omitted for legibility
- errors are fully in the responsibility of Georg Groh
- BIG thanks to Dan and James for a great book!

(Computational) Lexical Semantics

- in previous chapters: sentence semantics by relative occurrence of otherwise not further questioned "atomic" "words": now: look deeper into semantics of individual words

- Lemma (citation form): grammatical form of word used to represent it in dictionaries and thesaurus. (base form)
  carpet: lemma for *carpets* (word-form),
  sing: lemma for *sing, sang, sung* (word-forms)

- Word-senses: bank (side of river) vs bank (financial institution): Homonyms.
  bank: polysemous word → Word-Sense-Disambiguation

- "*financial institution that accepts deposits and channels the money into lending activities*": gloss defining a word sense

Word Senses

- lemma *bank* has two senses (word senses).
  sense (word sense): discrete representation of one aspect of meaning of a word.

- denote word-senses: $bank^1$, $bank^2$

- $bank^1$ ("financial institution"), $bank^2$ ("sloping mound");
  bat^1 ("club for hitting a ball"), bat^2 ("nocturnal flying animal"):
  homonyms and homographs (same writing)

- *write - right*; *piece - peace:* homophones. (←→ spelling errors)

- homographs that are not homophones:
  $bass^1$ ("fish") - $bass^2$ ("instrument")
  (←→ speech synthesis errors)

Word Senses

- sense bank^3: "biological repository": blood bank, sperm bank:
  - bank^1 and bank^3: "repositories for entities" (bank^1: money, bank^3: biological)
  - bank^1, bank^3: polysemy

polysemic word: one word ("one" in the sense of e.g. etymology, **rough** meaning) that has different subtle meaning variations (meaning aspects). **homonymous** words: different words (each with its own meaning), that have the same spelling

---

- sense bank^4: "building of financial institution":
  - bank^4 ←→ bank^1: systematic relationship: BUILDING ←→ ORGANIZATION;
  - bank^1, bank^4 : metonymy (subtype of polysemy): using one aspect of a concept or entity to refer to other aspects of the entity or to the entity itself
  - other examples:
    Author (*Jane Austen wrote Emma*) ←→ Works of Author (*I really love Jane Austen*)
    Tree (*Plums have beautiful blossoms*) ←→ Fruit (*I ate a preserved plum yesterday*)

Word Senses

- criteria for deciding whether differing uses of a word should be represented as distinct senses:
  - independent truth conditions
  - different syntactic behaviour
  - independent sense relations
  - exhibit antagonistic meanings etc.

[1] Subcategorization categories for a verb: which phrasal structures (NP, PP, etc.) a verb takes (e.g. as objects or prepositional refinements)

- example:
  1. *They rarely serve red meat, preferring to prepare seafood.*
  2. *He served as U.S. ambassador to Norway in 1976.*
  3. *He might have served his time, come out and led an upstanding life.*
     - (1), (3) different truth conditions and presuppositions;
     - (2) distinct subcategorization[1] structure: ←→ *serve **as** <NP>*

- Zeugma: decide problem "experimentally": example:
  1. *Which of those flights serve breakfast?*
  2. *Does Midwest Express serve Philadelphia?*
  3. *?: Does Midwest Express serve breakfast and Philadelphia?*

Relations Between Senses: Synonymy

see also [3]

- (nearly) identical senses of different lemmas: synonymy:
  - substitutable one for the other in any sentence without changing the truth conditions of the sentence (same propositional meaning)
  - *couch/sofa vomit/throw up filbert/hazelnut car/automobile*

- synonymy: actually between senses of words:
  synonyms may replace one another in a sentence:
  - example: big / large replaceable (big = big[1]):
    - *How big is that plane?*
    - *Would I be flying on a large or small plane?*
  - big / large not replaceable (big = big[2]):
    - *Miss Nelson became a kind of big sister to Benjamin.*

- synonyms: principle of contrast: different word forms → at least SLIGHTLY different meaning / different pragmatics.
  examples: water, H_20 ; car, automobile

Relations Between Senses: Hyponymy, Meronymy, and Antonymy

see also [3]

- Hyponmy (subordinate): sub-class - class relation;
  Hypernymy (superordinate): super-class – class relation;
  - sense A hyponym of sense B ("B subsumes A" "A is-a B") if $\forall x\colon A(x) \rightarrow B(x)$
  - examples:

Hypernym:	vehicle	fruit	furniture	mammal
Hyponym:	car	mango	chair	dog

- Meronym: part-of - whole relation;
  Holonym: whole – part-of relation;
  - examples:

Meronym:	leg	leg	wheel	CPU
Holonym:	chair	human	car	computer

- Antonymy:
  - *long/short big/little fast/slow cold/hot* (opposite ends of a scale)
  - *rise/fall up/down in/out* (reversed direction (reversives))

WordNet: A Database of Lexical Relations

- database(s) of lemmas + senses and relations. WordNet 3.0: 117,798 nouns, 11,529 verbs, 22,479 adjectives, and 4,481 adverbs

The noun “bass” has 8 senses in WordNet.
1. bass^{1} - (the lowest part of the musical range)
2. bass^{2}, bass part^{1} - (the lowest part in polyphonic music)
3. bass^{3}, basso^{1} - (an adult male singer with the lowest voice)
4. sea bass^{1}, bass^{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass^{1}, bass^{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass^{6}, bass voice^{1}, basso^{2} - (the lowest adult male singing voice)
7. bass^{7} - (the member with the lowest range of a family of musical instruments)
8. bass^{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

The adjective “bass” has 1 sense in WordNet.
1. bass^{1}, deep^{6} - (having or denoting a low vocal or instrumental range)
“a deep voice”; “a bass voice is lower than a baritone voice”; “a bass clarinet”

- Synsets: sets of synonyms: {bass^{1}, deep^{6}}, {bass^{6}, bass voice^{1}, basso^{2}}, {chump^{1}, fool^{2}, gull^{1}, mark^{9}, patsy^{1}, fall guy^{1}, sucker^{1}, soft touch^{1}, mug^{2}}

WordNet: A Database of Lexical Relations

- noun relations (between synsets):

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Instance Hypernym	Instance	From instances to their concepts	$Austen^1 \rightarrow author^1$
Instance Hyponym	Has-Instance	From concepts to their instances	$composer^1 \rightarrow Bach^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivation		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

- verb relations (between synsets):

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From events to subordinate event	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Semantic opposition between lemmas	$increase^1 \Longleftrightarrow decrease^1$

WordNet: A Database of Lexical Relations

- Supersenses (higher level categories) for synsets) (26 for nouns, 15 for verbs etc.)

Category	Example	Category	Example	Category	Example
ACT	*service*	GROUP	*place*	PLANT	*tree*
ANIMAL	*dog*	LOCATION	*area*	POSSESSION	*price*
ARTIFACT	*car*	MOTIVE	*reason*	PROCESS	*process*
ATTRIBUTE	*quality*	NATURAL EVENT	*experience*	QUANTITY	*amount*
BODY	*hair*	NATURAL OBJECT	*flower*	RELATION	*portion*
COGNITION	*way*	OTHER	*stuff*	SHAPE	*square*
COMMUNICATION	*review*	PERSON	*people*	STATE	*pain*
FEELING	*discomfort*	PHENOMENON	*result*	SUBSTANCE	*oil*
FOOD	*food*			TIME	*day*

WordNet: A Database of Lexical Relations

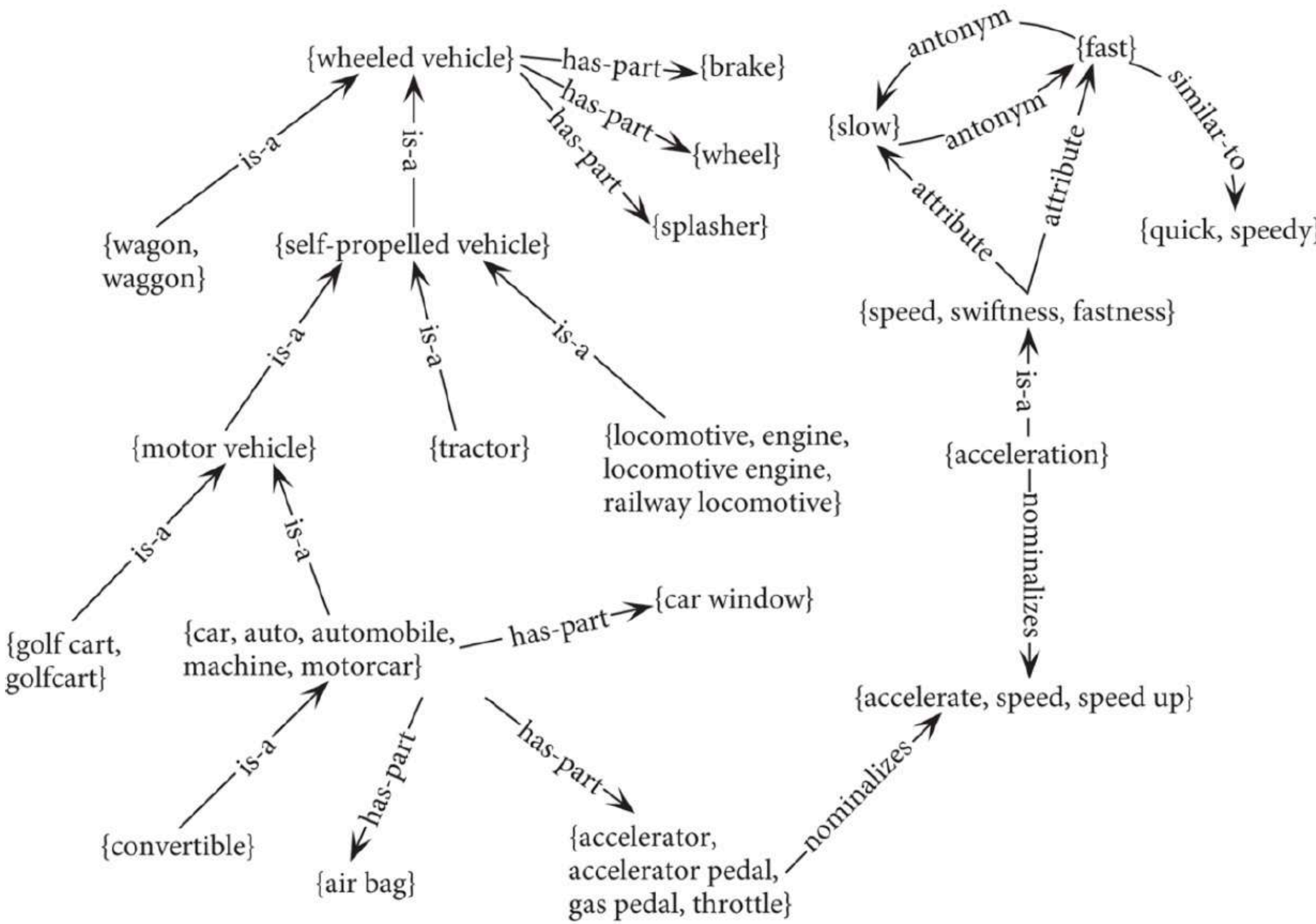
- Subsumption (hyponymy) chains in WordNet:

```
Sense 3
bass, basso --
(an adult male singer with the lowest voice)
=> singer, vocalist, vocalizer, vocaliser
   => musician, instrumentalist, player
      => performer, performing artist
         => entertainer
            => person, individual, someone...
               => organism, being
                  => living thing, animate thing,
                     => whole, unit
                        => object, physical object
                           => physical entity
                              => entity
               => causal agent, cause, causal agency
                  => physical entity
                     => entity

Sense 7
bass --
(the member with the lowest range of a family of
musical instruments)
=> musical instrument, instrument
   => device
      => instrumentality, instrumentation
         => artifact, artefact
            => whole, unit
               => object, physical object
                  => physical entity
                     => entity
```

WordNet: A Database of Lexical Relations

- WordNet as graph:



Word-Sense Disambiguation

- **WSD**: maps words to sense tags.
  sense tags: senses or equivalence classes of senses (e.g. for machine translation:)

WordNet Sense	Spanish Translation	WordNet Supersense	Target Word in Context
bass^4	lubina	FOOD	...fish as Pacific salmon and striped **bass** and...
bass^7	bajo	ARTIFACT	...play **bass** because he doesn't have to solo...

- lexical sample task: choose small set of target words along with an inventory of senses for those word from some lexicon: disambiguate those in a text

- all-words task: entire text + lexicon with inventory of senses: disambiguate every word in the text; similar to POS tagging, except with much larger set of tags;

  excellent baseline: choose most probable sense.
  alternative baseline guideline: one sense per discourse

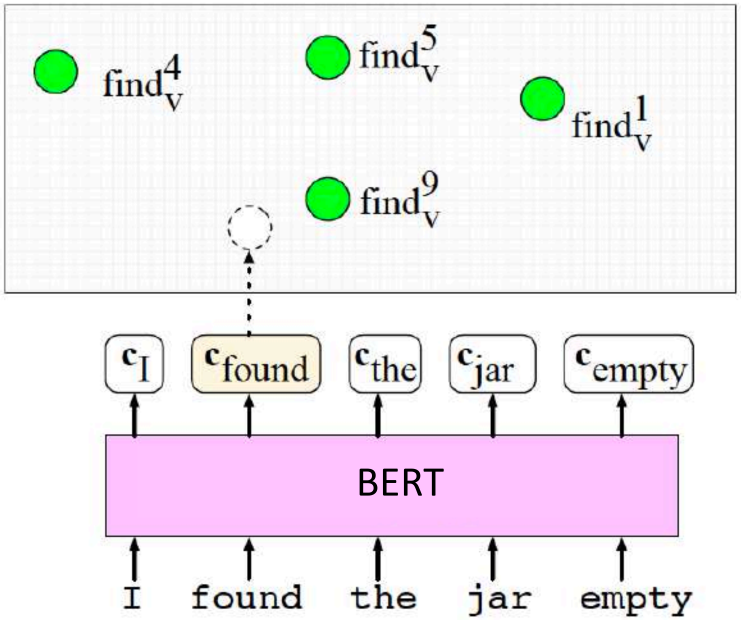
WSD with Contextual Embeddings

- ELMo, BERT: Contextual Embeddings: embeddings that depend on context (e.g. the sentence) that a word is in.

- Very simple, but SOTA: 1-NN classifier with sense embeddings:
  - training time: word-sense-labeled corpus (e.g. SemCore): for each token instance c_i of one word sense c compute contextual embeddings $\boldsymbol{c}_i$ → sense embedding $\boldsymbol{v}_s$:

$$\mathbf{v}_s = \frac{1}{n}\sum_i \mathbf{c}_i$$

  - test time: compute contextual embedding $\boldsymbol{t}$ for word, classify with 1-NN (using cosine similarity $\boldsymbol{t}^{\mathrm{T}}\boldsymbol{v}_s$)



WSD with Contextual Embeddings

- imputation of unseen senses $\hat{s}$ (of words s not present in sense-tagged corpus but present in WordNet):

$$
\begin{aligned}
&\text{if} \quad |S_{\hat{s}}| > 0, \quad \mathbf{v}_{\hat{s}} = \frac{1}{|S_{\hat{s}}|}\sum \mathbf{v}_s, \forall \mathbf{v}_s \in S_{\hat{s}} \\
&\text{else if} \quad |H_{\hat{s}}| > 0, \quad \mathbf{v}_{\hat{s}} = \frac{1}{|H_{\hat{s}}|}\sum \mathbf{v}_{syn}, \forall \mathbf{v}_{syn} \in H_{\hat{s}} \\
&\text{else if} \quad |L_{\hat{s}}| > 0, \quad \mathbf{v}_{\hat{s}} = \frac{1}{|L_{\hat{s}}|}\sum \mathbf{v}_{syn}, \forall \mathbf{v}_{syn} \in L_{\hat{s}}
\end{aligned}
$$

known sense embeddings of words in synset $\hat{s}$ of s

known sense embeddings of hyponym synsets of synset $\hat{s}$ of s

known sense embeddings of synsets that have the same super-sense as synset $\hat{s}$ of s

Supervised Word Sense Disambiguation

WSD: supervised machine learning task

- corpora for lexical sample tasks:
  - *line-hard-serve* corpus (1993): 4,000 sense-tagged examples of noun *line*, adjective *hard* and verb *serve*
  - Interest Corpus (1994): 2,369 sense-tagged examples of noun *interest*
  - SemEval: SENSEVAL-1 (2000): 34 target words, SENSEVAL-2 (2001): 73 target words and SENSEVAL-3 (2001): 57 target words

- corpora for all word tasks (semantic concordance: corpus with each open-class word labeled with its word sense from a specific dictionary or thesaurus):
  - Sem-Cor (1998), subset of Brown Corpus: 234,000 words manually tagged with WordNet senses
  - SENSEVAL-3 (2001) all-words: 2081 word tokens from WSJ and Brown corpora

Supervised WSD: Features

- isolated words: impossible to disambiguate → choose minimum bi-directional window of size 2N around word

---

- features from sentence: POS tags, lemmatization, syntactic parsing revealing headwords and dependency relations etc.

- collocational features from window: for each word: relative position, word itself, its root form, POS, etc.
  - *An electric guitar and bass player stand off to one side, not really part of the scene*

$$[w_{i-2}, \mathrm{POS}_{i-2}, w_{i-1}, \mathrm{POS}_{i-1}, w_{i+1}, \mathrm{POS}_{i+1}, w_{i+2}, \mathrm{POS}_{i+2}, w_{i-2}^{i-1}, w_{i}^{i+1}] \Longrightarrow$$

`[guitar, NN, and, CC, player, NN, stand, VB, and guitar, player stand]`

- distance weighted average of embeddings in symmetric window around word

WSD Evaluation

- extrinsic, task-based, or end-to-end evaluation: does new WSD approach improve end-to-end application like question answering or machine translation.

- intrinsic evaluation: as any other ML problem.
  baseline: for each word, choose most frequent (=first) sense from WordNet

WSD: Dictionary and Thesaurus Methods

- hand-labeling corpora expensive → indirect supervision from dictionaries and thesauruses or similar knowledge bases (distant supervision / weak supervision): knowledge based WSD

Dictionary and Thesaurus Methods: Lesk Algorithm

Simplified Lesk: idea: choose word-sense, whose dictionary gloss or dictionary definition has highest word-overlap with context window of target word in sentence:

The bank can guarantee deposits that will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.

bank[1]	Gloss:	a financial institution that accepts deposits and channels the money into lending activities	2 non stopword overlaps
	Examples:	"he cashed a check at the bank", "that bank holds the mortgage on my home"	
bank[2]	Gloss:	sloping land (especially the slope beside a body of water)	0 non stopword overlaps
	Examples:	"they pulled the canoe up on the bank", "he sat on the bank of the river and watched the currents"	

- possible upgrades:
  - IDF weighting of overlapping word's counts
  - use embeddings and cosine to compute similarity instead of overlap counts

Dictionary and Thesaurus Methods: Lesk Algorithm

Original Lesk: choose word-sense, whose dictionary definition has highest word-overlap with dictionary definitions of other words in context window of target word in sentence:

... pine cone ...

pine	1	kinds of evergreen tree with needle-shaped leaves	
	2	waste away through sorrow or illness	
cone	1	solid body which narrows to a point	0 non stopword overlaps
	2	something of this shape whether solid or hollow	0 non stopword overlaps
	3	fruit of certain evergreen trees	2 non stopword overlaps

- usually performs worse than Simplified Lesk
- reason for not so spectacular performance of both: dictionary definitions are too short

Word-in-Context Evaluation

- system is given two sentences (e.g. WordNet glosses), each with a possibly different meaning of a word: decide whether meaning are the same (T) or not (F)

F	There's a lot of trash on the **bed** of the river — I keep a glass of water next to my **bed** when I sleep
F	**Justify** the margins — The end **justifies** the means
T	**Air** pollution — Open a window and let in some **air**
T	The expanded **window** will give us time to catch the thieves — You have a two-hour **window** of clear weather to finish working on the lawn

- baseline algorithm: use contextual embeddings (e.g. BERT) for each of the two instances of the word and a threshold on cosine similarity to decide between T and S

Supervised WSD: Wikipedia as Training Data

use Wikipedia or DBpedia or other (more or less formal) online encyclopedias:

- Wikipedia:
  - each link on a word in a Wiki-article links to a Wikipedia page specific for the corresponding word-sense (target page)
  - → annotate the linked word in the sentence with the word-sense (e.g. from WordNet) that is associated most with the target page
  - → preprocessing task: for each Wikipedia page find the most likely word-sense in WordNet

In 1834, Sumner was admitted to the **[[bar (law)|bar]]** at the age of twenty-three, and entered private practice in Boston.

It is danced in 3/4 time (like most waltzes), with the couple turning approx. 180 degrees every **[[bar (music)|bar]]**.

Jenga is a popular beer in the **[[bar (establishment)|bar]]**s of Thailand.

Unsupervised Word Sense Induction

- clustering based WSI algorithm by Schütze:
  required: distance measure between context vectors + clustering algorithm (e.g. agglomerative hierarchical clustering)

  1. For each token (occurrence) w_i of word w in a corpus, compute a context vector c_i
  2. cluster the context vectors $\{c_i\}$ into J clusters
  3. sense vector $s_j(w)$= centroid of cluster j

- WSD: compute context vector c of token t of w, determine closest $s_j(t)$

Using Thesauruses to Improve Embeddings

- static embeddings (e.g. GloVe, Word2Vec): often antonyms (*expensive, cheap*) are too cosine-similar

- light-weight idea: retrofitting / counterfitting: use thesaurus with synonymy and antonymy information to push synonyms closer and antonyms further apart

	Before counterfitting			After counterfitting		
east	west	north	south	eastward	eastern	easterly
expensive	pricey	cheaper	costly	costly	pricy	overpriced
British	American	Australian	Britain	Brits	London	BBC

Figure 19.12 The nearest neighbors in GloVe to *east*, *expensive*, and *British* include antonyms like *west*. The right side showing the improvement in GloVe nearest neighbors

- more fundamental idea: neural training of embeddings with modified loss functions, e.g. incorporating antonymy-, synonymy- or super-senses (e.g. in Word2Vec instead of just using “positive” or “negative” examples

Word Similarity: Older Thesaurus Methods

- path-length-based similarity btw. word-senses : use subsumption (hyponym / hypernym) hierarchy from e.g. WordNet

$$\text{sim}_{\text{path}}(c_1, c_2) = \frac{1}{\text{pathlen}(c_1, c_2)}$$

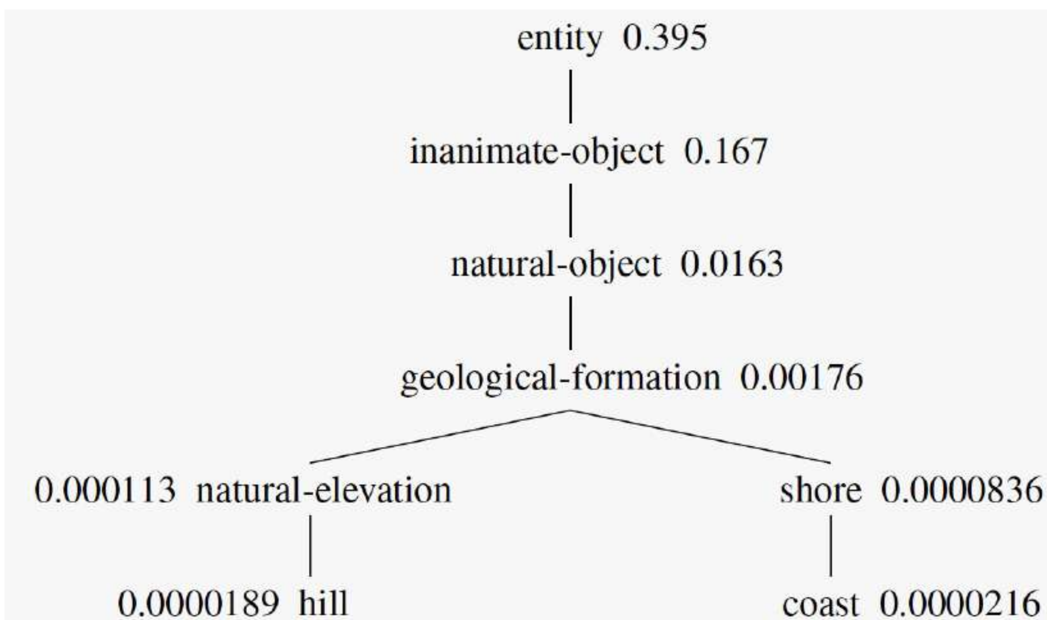
- word similarity incorporating all possible senses:

$$\text{wordsim}(w_1, w_2) = \max_{\substack{c_1 \in \text{senses}(w_1) \\ c_2 \in \text{senses}(w_2)}} \text{sim}(c_1, c_2)$$

Word Similarity: Older Thesaurus Methods

- path-lengths: other idea: incorporate edge-weights (semantic distances btw. concepts in hierarchy not uniform): information-content word-similarity:
  - $P(c) := P(w \in c)$: Estimate with N word corpus and e.g. WordNet:

$$P(c) = \frac{\sum_{w \in \text{words}(c)} count(w)}{N}$$



Word Similarity: Older Thesaurus Methods

- two additional elements:
  - Information content of a concept c:

    $$\text{IC}(c) = -\log P(c)$$

  - Lowest Common Subsumer LCS (least common ancestor LCA)

    $LCS(c_1, c_2)$: lowest node in the hierarchy that subsumes (is a hypernym of) both c_1 and c_2

→ Resnik similarity measure: information content of the lowest common subsumer:

$$\text{sim}_{\text{resnik}}(c_1, c_2) = -\log P(\text{LCS}(c_1, c_2))$$

Word Similarity: Older Thesaurus Methods

- Lin's similarity measure: "similarity between A and B is measured by the ratio between the amount of information needed to state the commonality of A and B and the information needed to fully describe what A and B are"

$$\text{sim}_{\text{Lin}}(c_1, c_2) = \frac{2 \times \log P(LCS(c_1, c_2))}{\log P(c_1) + \log P(c_2)}$$

$$\text{sim}_{\text{Lin}}(\text{hill}, \text{coast}) = \frac{2 \times \log P(\text{geological-formation})}{\log P(\text{hill}) + \log P(\text{coast})} = 0.59$$

- Jiang Conrath distance:

$$\text{dist}_{\text{JC}}(c_1, c_2) = 2 \times \log P(\text{LCS}(c_1, c_2)) - (\log P(c_1) + \log P(c_2))$$

Bibliography

(1) Dan Jurafsky and James Martin: Speech and Language Processing (3rd ed. draft, Jan 2023); Online: https://web.stanford.edu/~jurafsky/slp3/ (URL, Oct 2023) (this slide set is especially based on chapter 23)

Recommendations for Studying

- minimal approach:

  work with the slides and understand their contents! Think beyond instead of merely memorizing the contents

- standard approach:

  minimal approach + read the corresponding pages in Jurafsky [1]

- interested students

  == standard approach