Computational Linguistics

11. More on WordNet

Xiaojing Bai

Tsinghua University

https://bxjthu.github.io/CompLing

More on WordNet

- What is it?
- How does it look like?
- How to use it?

WordNet: What is it?

- WordNet 3.1
- Thesaurus: a structured list of words organized by meaning
- A lexical database of English
- Three databases: 1) nouns, 2) verbs, 3) adjectives and adverbs
- Defining a sense through its relationship with other senses
- The most commonly used resource for English sense relations

Noun relations in WordNet

Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$\textit{breakfast}^1 \rightarrow \textit{meal}^1$
Hyponym	Subordinate	From concepts to subtypes	$\textit{meal}^1 \rightarrow \textit{lunch}^1$
Instance Hypernym	Instance	From instances to their concepts	$\textit{Austen}^1 \rightarrow \textit{author}^1$
Instance Hyponym	Has-Instance	From concepts to concept instances	$\textit{composer}^1 \rightarrow \textit{Bach}^1$
Member Meronym	Has-Member	From groups to their members	$\textit{faculty}^2 \rightarrow \textit{professor}^1$
Member Holonym	Member-Of	From members to their groups	$\textit{copilot}^1 \rightarrow \textit{crew}^1$
Part Meronym	Has-Part	From wholes to parts	$\textit{table}^2 \rightarrow \textit{leg}^3$
Part Holonym	Part-Of	From parts to wholes	$\textit{course}^7 \rightarrow \textit{meal}^1$
Substance Meronym		From substances to their subparts	$\textit{water}^1 \rightarrow \textit{oxygen}^1$
Substance Holonym		From parts of substances to wholes	$\textit{gin}^1 \rightarrow \textit{martini}^1$
Antonym		Semantic opposition between lemmas	$\textit{leader}^1 \Longleftrightarrow \textit{follower}^1$
Derivationally Related Form		Lemmas w/same morphological root	$\textit{destruction}^1 \Longleftrightarrow \textit{destroy}^1$

Verb relations in WordNet

Relation	Definition	Example
Hypernym	From events to superordinate events	$\textit{fly}^9 \rightarrow \textit{travel}^5$
Troponym	From events to subordinate event (often via specific manner)	$\textit{walk}^1 \rightarrow \textit{stroll}^1$
Entails	From verbs (events) to the verbs (events) they entail	$\textit{snore}^1 \rightarrow \textit{sleep}^1$
Antonym	Semantic opposition between lemmas	$\textit{increase}^1 \Longleftrightarrow \textit{decrease}^1$
Derivationally Related Form	Lemmas with same morphological root	$\textit{destroy}^1 \Longleftrightarrow \textit{destruction}^1$

More details about WordNet

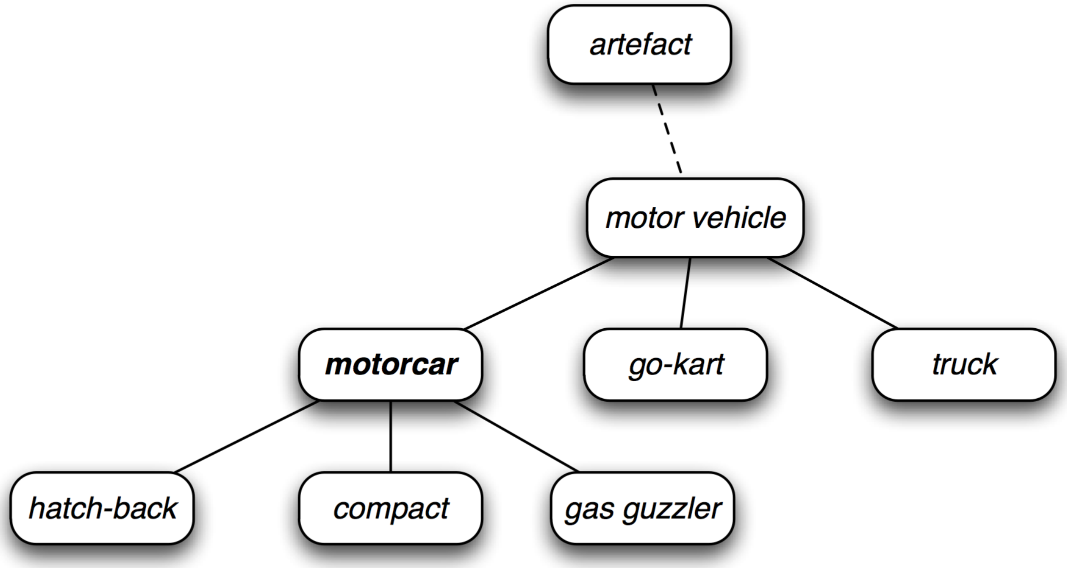
- WordNet online
- Database statistics
- A glossary of WordNet terms
- Five Papers on WordNet
- Frequently Asked Questions
- Wordnet with NLTK

WordNet: How does it look like?

- [WordNet 3.1 database files]()

  Samples: [index.noun](); [data.noun]()

- WordNet concept hierarchy



WordNet: How to use it?

- [Related projects]
- Measuring word similarity with WordNet
  - A fundamental task for semantic models is to predict how similar two words’ meanings are
  - Thesaurus methods
  - Goal: To measure how close the two target words are within the hierarchy

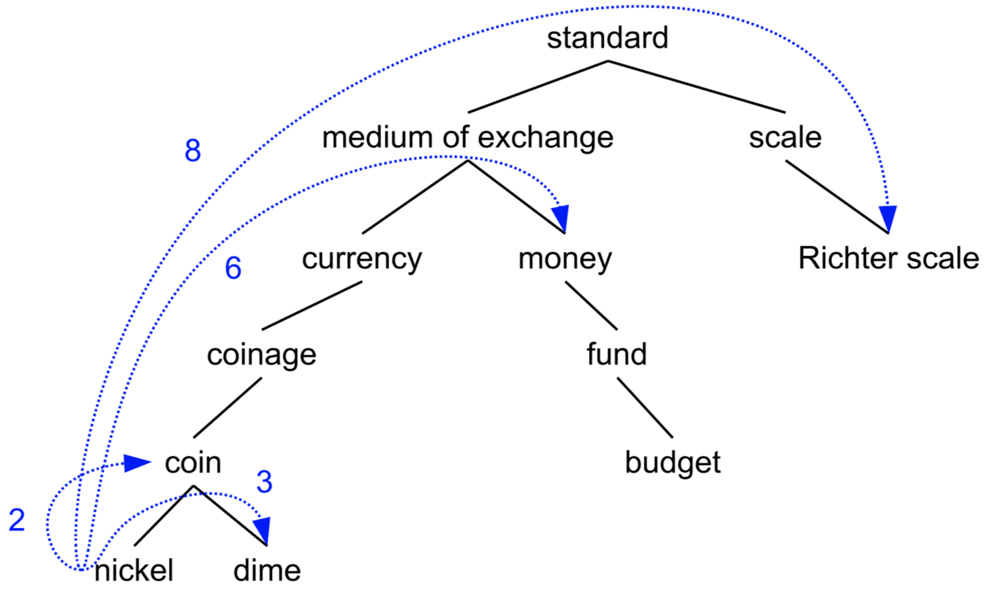
Word similarity: path-length based algorithm

- Basic intuition
- Implicit assumption

pathlen(c_1, c_2) = 1 + the number of edges in the shortest path in the thesaurus graph between the sense nodes c_1 and c_2

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

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



Word similarity: refined algorithms

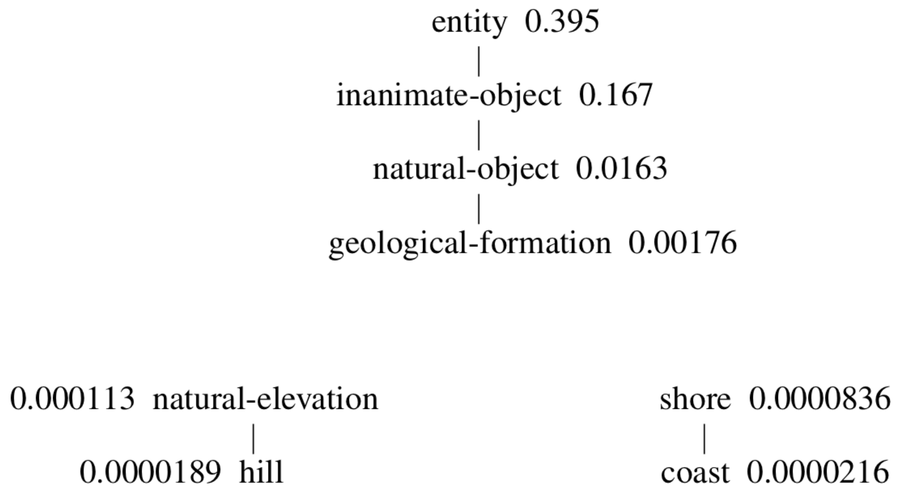
- More fine-grained metrics, e.g., information-content word similarity
    - Structure of the thesaurus
    - Probabilistic information derived from a corpus

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

$P(c)$: the probability of encountering an instance of concept c in the corpus;

$words(c)$: the set of words subsumed by concept c;

N: the total number of words in the corpus that are also present in the thesaurus.



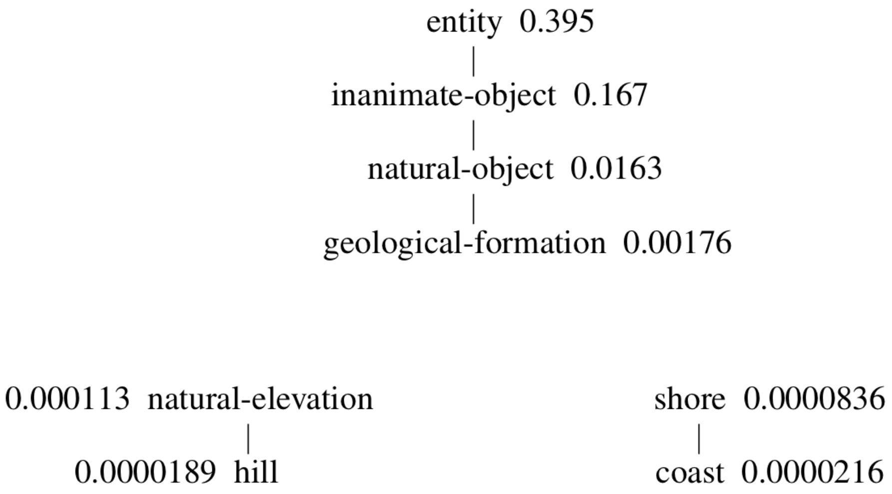
Word similarity: refined algorithms

- More fine-grained metrics, e.g., information-content word similarity
  - Structure of the thesaurus
  - Probabilistic information derived from a corpus

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

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

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



Logarithms

- Logarithms can be used to simplify calculations, as the addition and subtraction of logarithms is equivalent to multiplication and division, though the use of printed tables of logarithms for this has declined with the spread of electronic calculators… The base of a common logarithm is 10, and that of a natural logarithm is the number e (2.71828 …) (Source: Oxford dictionary)

- The logarithm of x to base b is denoted as logb (x) (or, without parentheses, as logb x, or even without explicit base as log x, when no confusion is possible). (Source: Wiki)

- 在自信息量（self-information）$I(x) = \log(\frac{1}{p(x)}) = -\log p(x)$的定义中，log对数操作既可以以2为底，也可以以e为底，但它们没有本质区别，只差一个固定的倍数。当讨论信息编码的时候，采用以2为底比较方便，这时候$I(x)$的单位是bit；而讨论机器学习的时候，以e为底比较方便。(source: http://zhangtielei.com/posts/blog-deep-learning-foundation.html)

Word similarity: five thesaurus/dictionary-based similarity measures

$$
\begin{aligned}
\mathrm{sim}_{\mathrm{path}}(c_1,c_2) &= \frac{1}{\mathrm{pathlen}(c_1,c_2)} \\
\mathrm{sim}_{\mathrm{Resnik}}(c_1,c_2) &= -\log P(\mathrm{LCS}(c_1,c_2)) \\
\mathrm{sim}_{\mathrm{Lin}}(c_1,c_2) &= \frac{2\times\log P(LCS(c_1,c_2))}{\log P(c_1)+\log P(c_2)} \\
\mathrm{sim}_{\mathrm{JC}}(c_1,c_2) &= \frac{1}{2\times\log P(\mathrm{LCS}(c_1,c_2))-(\log P(c_1)+\log P(c_2))} \\
\mathrm{sim}_{\mathrm{eLesk}}(c_1,c_2) &= \sum_{r,q\in\mathrm{RELS}} \mathrm{overlap}(\mathrm{gloss}(r(c_1)),\mathrm{gloss}(q(c_2)))
\end{aligned}
$$

- WordNet::Similarity
- Computing semantic similarity with NLTK

Evaluating thesaurus-based similarity

- NLP evaluation

  "As the engineering branch of computational linguistics, natural language processing is concerned with the creation of artifacts that accomplish tasks. The operative question in evaluating an NLP algorithm or system is therefore the extent to which it produces the results for which it was designed. "
  (Resnik & Lin, 2010)

- Evaluation metrics

  - Correlation coefficient:
    (1)human-labeled datasets; (2)TOEFL datasets; (3)contextual datasets

  - End-application embedment

Related projects

- Wordnets in the World
- Semantic networks - in languages other than English
- NLTK-HOWTO: WordNet Interface
- Chinese Object-Oriented Lexicon (introduced by Zheng Hua)

Recap: Algorithms to measure word similarity

- Thesaurus-based algorithms
- Distributional algorithms

  Cosine for measuring word similarity

  - Word similarity as vector similarity
  - Vector similarity as the cosine of the angle between the vectors

Recap: Word similarity vs. word relatedness

Mug, cup, coffee, croissant

- Word similarity - a subcase of word relatedness
- Not distinguished in the following measuring algorithm

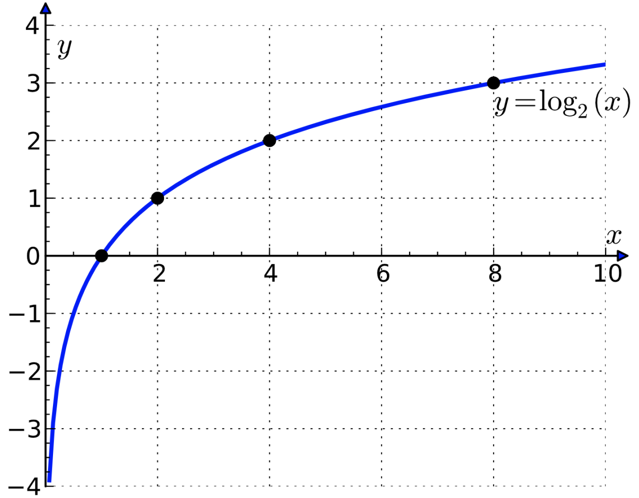
Recap: Measure of the association between words

Positive Pointwise Mutual Information (PPMI)

How much more are the two words co-occurring in our corpus than we would expect if they were independent?

$$PMI(w,c) = \log_2 \frac{P(w,c)}{P(w)P(c)}$$

$$PPMI(w,c) = max(\log_2 \frac{P(w,c)}{P(w)P(c)}, 0)$$



Recap: Word sense disambiguation (WSD)

- Lexical ambiguity and an avalanche of competing interpretations
- WSD: the task of selecting the correct sense for a word
  - Input: a word in context along with a fixed inventory of potential senses
  - Output: the correct word sense for that use

  Reports said the plant was likely to close in December, leaving many jobless.

  plant 1: leafy green organism
  plant 2: equipment and fixtures for manufacturing

- Applications

Voluntary research tasks: WSD algorithms

- Supervised

  We know the answers for many examples and can use them to learn from their (automatically determinable) characteristics. We then apply the learned model to a comparable set of examples (not the same ones).

- Weakly supervised (knowledge-based)

  We start with no known answers, but we use secondary texts (dictionary glosses) to infer underlying relationships through the Lesk algorithm.

- Semi-supervised

  We know the answers for a small number of examples, and can gain more examples from the data by finding similar cases and inferring the answers they should have through bootstrapping.

- Unsupervised

  We start with no known answers, and no predefined senses. The set of “senses” is created automatically from the instances of each word in the training set.

Homework

- Read/Review (Quiz 9 on Dec. 12, 2018)
  - J+M_6 (6.1-6.7)
  - J+M_C (C.1-C.4)
- Practice
  - http://www.nltk.org/book/ch02.html#wordnet

Next session

Semantic Role Labeling