Lexical Semantics - WordNet

Pawan Goyal

CSE, IIT Kharagpur

Week 8, Lecture 2

WordNet

https://wordnet.princeton.edu/wordnet/

- A hierarchically organized lexical database
- A machine-readable thesaurus, and aspects of a dictionary
- Versions for other languages are under development

part of speech	no. synsets
noun	82,115
verb	13,767
adjective	18,156
adverb	3,621

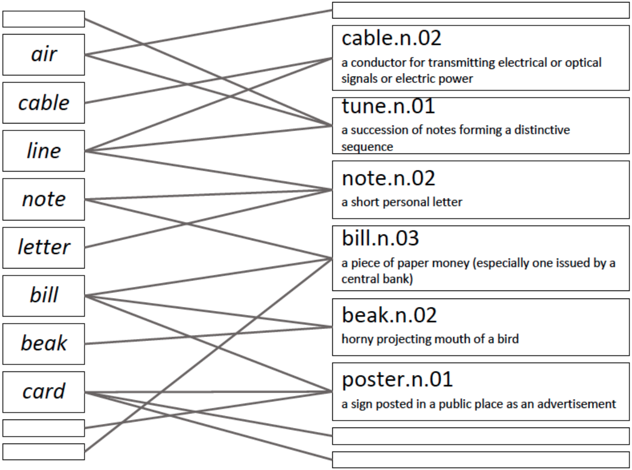
Synsets in WordNet

- A **synset** is a set of synonyms representing a sense
- Example: chump as a noun to mean ‘a person who is gullible and easy to take advantage of’

  {chump^1, fool^2, gull^1, mark^9, patsy^1, fall guy^1, sucker^1, soft touch^1, mug^2}

- Each of these senses share this same gloss.
- For WordNet, the meaning of this sense of chump is this list.

lemma vs. synsets



All relations in WordNet

```
searchtype is at least one of the following:
        -ants{n|v|a|r}          Antonyms
        -hype{n|v}              Hypernyms
        -hypo{n|v}, -tree{n|v}  Hyponyms & Hyponym Tree
        -entav                  Verb Entailment
        -syns{n|v|a|r}          Synonyms (ordered by estimated frequency)
        -smemn                  Member of Holonyms
        -ssubn                  Substance of Holonyms
        -sprtn                  Part of Holonyms
        -membn                  Has Member Meronyms
        -subsn                  Has Substance Meronyms
        -partn                  Has Part Meronyms
        -meron                  All Meronyms
        -holon                  All Holonyms
        -causv                  Cause to
        -pert{a|r}              Pertainyms
        -attr{n|a}              Attributes
        -deri{n|v}              Derived Forms
        -domn{n|v|a|r}          Domain
        -domt{n|v|a|r}          Domain Terms
        -faml{n|v|a|r}          Familiarity & Polysemy Count
        -framv                  Verb Frames
        -coor{n|v}              Coordinate Terms (sisters)
        -simsv                  Synonyms (grouped by similarity of meaning)
        -hmern                  Hierarchical Meronyms
        -hholn                  Hierarchical Holonyms
        -grep{n|v|a|r}          List of Compound Words
        -over                   Overview of Senses
```

Wordnet noun and verb relations

Relation	Also called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Member Meronym	Has-Member	From groups to their members	$faculty^2 \rightarrow professor^1$
Has-Instance		From concepts to instances of the concept	$composer^1 \rightarrow Bach^1$
Instance		From instances to their concepts	$Austen^1 \rightarrow author^1$
Member Holonym	Member-Of	From members to their groups	$copilot^1 \rightarrow crew^1$
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		Opposites	$leader^1 \rightarrow follower^1$

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From a verb (event) to a specific manner elaboration of that verb	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Opposites	$increase^1 \Longleftrightarrow decrease^1$

WordNet Hierarchies

```
Synonyms/Hypernyms (Ordered by Estimated Frequency) of noun mouse

4 senses of mouse

Sense 1
mouse
       => rodent, gnawer
           => placental, placental mammal, eutherian, eutherian mammal
               => mammal, mammalian
                   => vertebrate, craniate
                       => chordate
                           => animal, animate being, beast, brute, creature, fauna
                               => organism, being
                                   => living thing, animate thing
                                       => whole, unit
                                           => object, physical object
                                               => physical entity
                                                   => entity
```

```
Sense 4
mouse, computer mouse
       => electronic device
           => device
               => instrumentality, instrumentation
                   => artifact, artefact
                       => whole, unit
                           => object, physical object
                               => physical entity
                                   => entity
```

Word Similarity

- Synonymy is a binary relation
  - Two words are either synonymous or not

Word Similarity

- Synonymy is a binary relation
  - Two words are either synonymous or not
- We want a looser metric
  - Word similarity or
  - Word distance

Word Similarity

- Synonymy is a binary relation
  - Two words are either synonymous or not
- We want a looser metric
  - Word similarity or
  - Word distance
- Two words are more similar
  - If they share more features of meaning

Word Similarity

- Synonymy is a binary relation
  - Two words are either synonymous or not
- We want a looser metric
  - Word similarity or
  - Word distance
- Two words are more similar
  - If they share more features of meaning
- Actually these are really relations between **senses**:
  - Instead of saying “bank is like fund”
  - We say
    - ★ Bank^1 is similar to fund^3
    - ★ Bank^2 is similar to slope^5

Word Similarity

- Synonymy is a binary relation
  - Two words are either synonymous or not
- We want a looser metric
  - Word similarity or
  - Word distance
- Two words are more similar
  - If they share more features of meaning
- Actually these are really relations between **senses**:
  - Instead of saying "bank is like fund"
  - We say
    - Bank^1 is similar to fund^3
    - Bank^2 is similar to slope^5
- We will compute similarity over both words and senses

Two classes of algorithms

Distributional algorithms

By comparing words based on their distributional context in the corpora

Thesaurus-based algorithms

Based on whether words are “nearby” in WordNet

Thesaurus-based Word Similarity

- We could use anything in the thesaurus:
  - Meronymy, hyponymy, troponymy
  - Glosses and example sentences

Thesaurus-based Word Similarity

- We could use anything in the thesaurus:
  - Meronymy, hyponymy, troponymy
  - Glosses and example sentences
- In practice, “thesaurus-based” methods usually use:
  - the is-a/subsumption/hypernymy hierarchy
  - and sometimes the glosses too

Thesaurus-based Word Similarity

- We could use anything in the thesaurus:
  - Meronymy, hyponymy, troponymy
  - Glosses and example sentences
- In practice, “thesaurus-based” methods usually use:
  - the is-a/subsumption/hypernymy hierarchy
  - and sometimes the glosses too
- Word similarity vs. word relatedness
  - Similar words are near-synonyms
  - Related words could be related any way
    - ★ car, gasoline : related, but nor similar
    - ★ car, bicycle: similar

Path-based similarity

Basic Idea

- Two words are similar if they are nearby in the hypernym graph

Path-based similarity

Basic Idea

- Two words are similar if they are nearby in the hypernym graph
- $pathlen(c_1, c_2)$ = number of edges in shortest path (in hypernym graph) between senses c_1 and c_2

Path-based similarity

Basic Idea

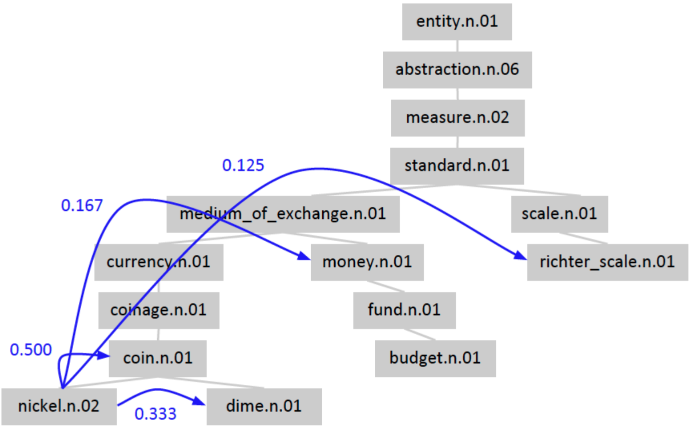
- Two words are similar if they are nearby in the hypernym graph
- $pathlen(c_1, c_2)$ = number of edges in shortest path (in hypernym graph) between senses c_1 and c_2
- $sim_{path}(c_1, c_2) = \frac{1}{1+pathlen(c_1,c_2)}$

Path-based similarity

Basic Idea

- Two words are similar if they are nearby in the hypernym graph
- $pathlen(c_1, c_2)$ = number of edges in shortest path (in hypernym graph) between senses c_1 and c_2
- $sim_{path}(c_1, c_2) = \frac{1}{1+pathlen(c_1,c_2)}$
- $sim(w_1, w_2) = max_{c_1 \in senses(w_1), c_2 \in senses(w_2)} sim(c_1, c_2)$

Shortest path in the hierarchy



Leacock-Chodorow (L-C) Similarity

L-C similarity

$$sim_{LC}(c_1, c_2) = -log(pathlen(c_1, c_2)/2d)$$

d: maximum depth of the hierarchy

Leacock-Chodorow (L-C) Similarity

L-C similarity

$$sim_{LC}(c_1, c_2) = -log(pathlen(c_1, c_2)/2d)$$

d: maximum depth of the hierarchy

Problems with L-C similarity

- Assumes each edge represents a uniform distance
- ‘nickel-money’ seems closer than ‘nickel-standard’

Leacock-Chodorow (L-C) Similarity

L-C similarity

$$sim_{LC}(c_1, c_2) = -log(pathlen(c_1, c_2)/2d)$$

d: maximum depth of the hierarchy

Problems with L-C similarity

- Assumes each edge represents a uniform distance
- ‘nickel-money’ seems closer than ‘nickel-standard’
- We want a metric which lets us assign different “lengths” to different edges - but how?

Concept probability models

Cencept probabilities

- For each concept (synset) c, let $P(c)$ be the probability that a randomly selected word in a corpus is an instance (hyponym) of c

Concept probability models

Cencept probabilities

- For each concept (synset) c, let $P(c)$ be the probability that a randomly selected word in a corpus is an instance (hyponym) of c
- $P(ROOT) = 1$
- The lower a node in the hierarchy, the lower its probability

Concept probability models

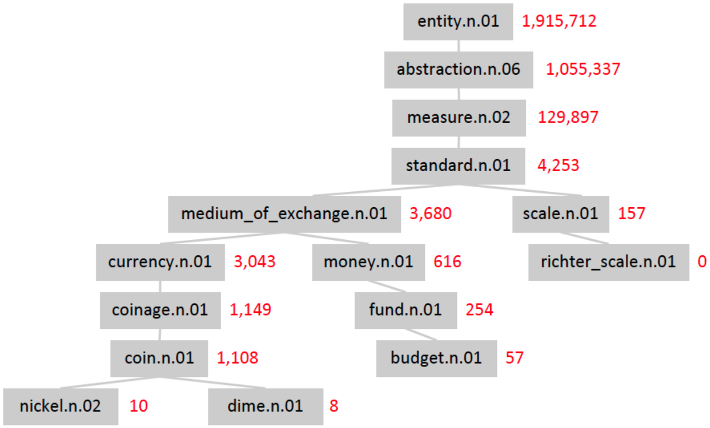
Concept probabilities

- For each concept (synset) c, let $P(c)$ be the probability that a randomly selected word in a corpus is an instance (hyponym) of c
- $P(ROOT) = 1$
- The lower a node in the hierarchy, the lower its probability

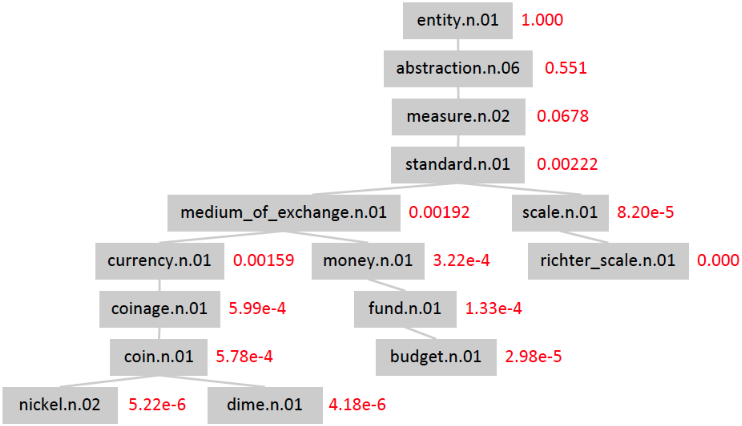
Estimating concept probabilities

- Train by counting "concept activations" in a corpus
- Each occurrence of *dime* also increments counts for *coin, currency, standard*, etc.

Example : concept count



Example : concept probabilities

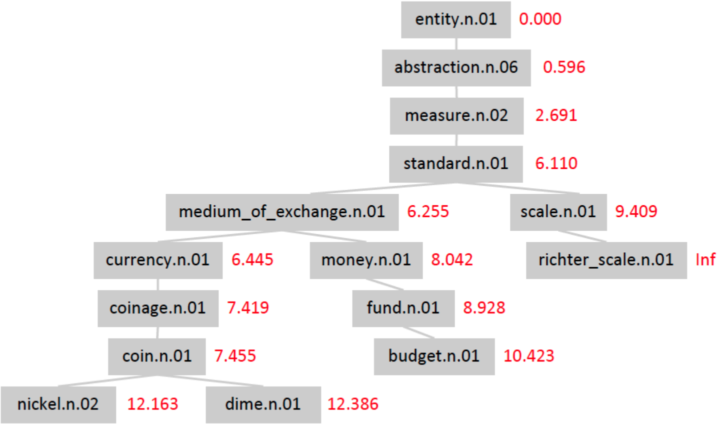


Information content

Information content

- Information content: $IC(c) = -logP(c)$
- Lowest common subsumer : $LCS(c_1, c_2)$: the lowest node in the hierarchy that subsumes (is a hypernym of) both c_1 and c_2
- We are now ready to see how to use information content (IC) as a similarity metric.

Example : Information content

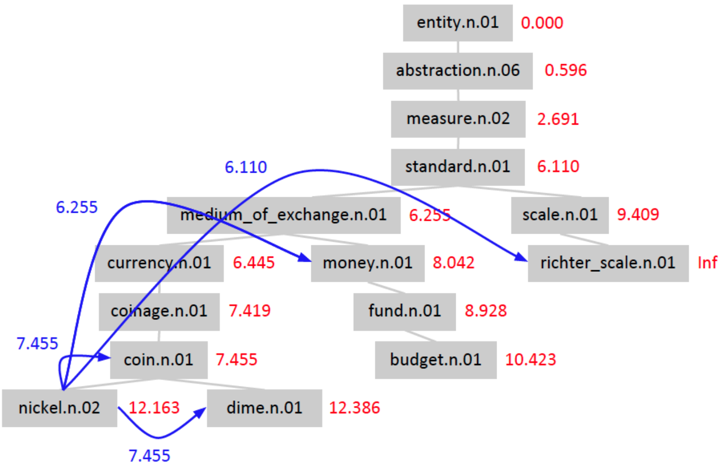


Resnik Similarity

Resnik Similarity

- Intuition: how similar two words are depends on how much they have in common
- It measures the commonality by the information content of the lowest common subsumer
- $sim_{resnik}(c_1, c_2) = IC(LCS(c_1, c_2)) = -logP(LCS(c_1, c_2))$

Example: Resnik similarity



Lin similarity

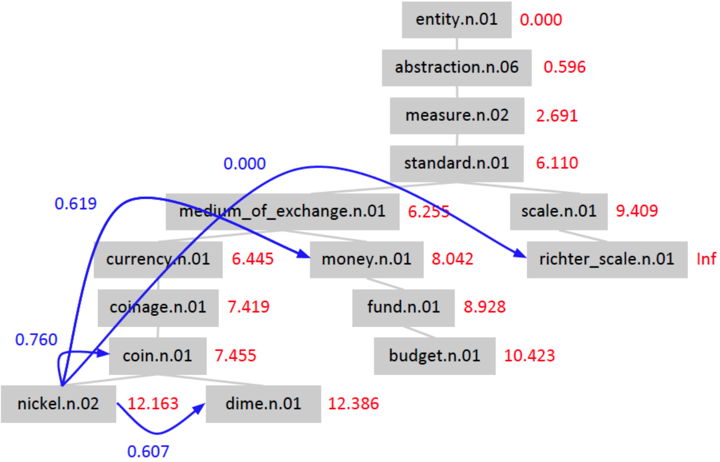
Proportion of shared information

- It's not just about commonalities - it's also about differences!
- **Resnik:** The more information content they share, the more similar they are
- **Lin:** The more information content they don't share, the less similar they are
- Not the *absolute* quantity of shared information but the *proportion* of shared information

$$sim_{Lin}(c_1,c_2) = \frac{2logP(LCS(c_1,c_2))}{logP(c_1)+logP(c_2)}$$

The information content common to c_1 and c_2, normalized by their average information content.

Example: Lin similarity



Jiang-Conrath distance

JC similarity

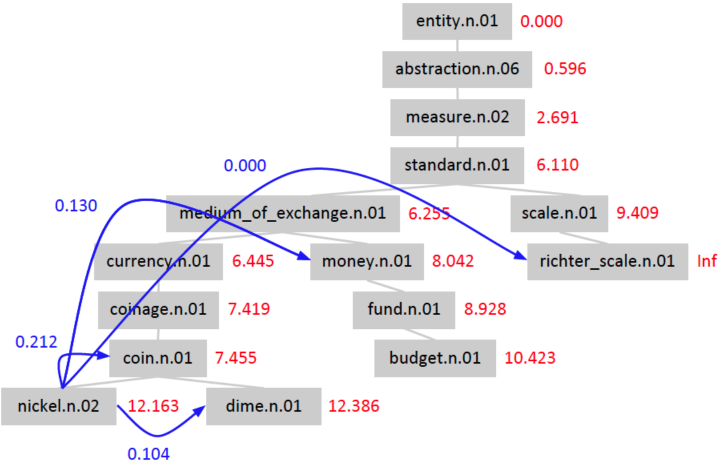
We can use IC to assign lengths to graph edges:

$$dist_{JC}(c, hypernym(c)) = IC(c) - IC(hypernym(c))$$

$$\begin{aligned} dist_{JC}(c_1, c_2) &= dist_{JC}(c_1, LCS(c_1, c_2)) + dist_{JC}(c_2, LCS(c_1, c_2)) \\ &= IC(c_1) - IC(LCS(c_1, c_2)) + IC(c_2) - IC(LCS(c_1, c_2)) \\ &= IC(c_1) + IC(c_2) - 2 \times IC(LCS(c_1, c_2)) \end{aligned}$$

$$sim_{JC}(c_1, c_2) = \frac{1}{IC(c_1) + IC(c_2) - 2 \times IC(LCS(c_1, c_2))}$$

Example: Jiang-Conrath distance



The (extended) Lesk Algorithm

- Two concepts are similar if their glosses contain similar words
  - *Drawing paper: **paper** that is **specially prepared** for use in drafting*
  - *Decal: the art of transferring designs from **specially prepared paper** to a wood or glass or metal surface*

The (extended) Lesk Algorithm

- Two concepts are similar if their glosses contain similar words
  - *Drawing paper: **paper** that is **specially prepared** for use in drafting*
  - *Decal: the art of transferring designs from **specially prepared paper** to a wood or glass or metal surface*
- For each n-word phrase that occurs in both glosses, add a score of n^2

The (extended) Lesk Algorithm

- Two concepts are similar if their glosses contain similar words
  - *Drawing paper: **paper** that is **specially prepared** for use in drafting*
  - *Decal: the art of transferring designs from **specially prepared paper** to a wood or glass or metal surface*
- For each n-word phrase that occurs in both glosses, add a score of n^2
- **paper** and **specially prepared** $\rightarrow 1+4=5$

Problem in mapping words to wordnet senses

I saw a man who is 98 years old and can still walk and tell jokes

Ambiguity is rampant!