Word senses.[1]

Victor Kitov

v.v.kitov@yandex.ru

[1] With materials used from "Speech and Language Processing", D. Jurafsky and J. H. Martin.

Table of Contents

Lemma, wordform, meaning

- **Lexical semantics** - field of study about meaning of words.
- **Wordform** - word in different forms
  - e.g. carpets; sing, sang, sung
- **Lemma** - word in its standardized form
  - e.g.: carpets->carpet; sing,sang,sung->sing
- The same word can have many **word senses** (meanings):
  - bank - financial institution
    - Instead, a *bank* can hold the investments in a custodial account in the client's name.
  - bank - building where financial institution is located
    - Please turn right after the bank.
  - bank - sloping side of a river
    - But as agriculture burgeons on the east *bank*, the river will shrink even more.
  - bank - repository for various biological entities
    - Taken from the blood bank

Multiple senses

- This sense ambiguity causes problems in:
    - speech synthesys (different meanings may be pronounced differently)
    - information retrieval, question answering (return not what was requested), machine translation
- **Word sense disambiguation** - the task of determining which sense of a word is being used in particular context.
- Bank usages are called homographs (омографы) - different senses but spelled the same.

homonimy vs. polysemy

- Completely unrelated senses called **homonimy** (омонимия)
  - e.g. bank=financial institution and bank=sloping side of a river
- Related senses are called **polysemy** (полисемия, многозначность)
  - e.g. bank=financial institution and bank=building where financial institution is located
- Examples of polysemy:
  - Author (Jane Austen wrote Emma) $\leftrightarrow$ Works of Author (I really love Jane Austen)
  - Tree (Plums have beautiful blossoms) $\leftrightarrow$ Fruit (I ate a preserved plum yesterday)
- Distinguish between homonimy and polysemy by trying to use both senses in one sentence:
  - Which of those flights serve breakfast?
  - Does Midwest Express serve Philadelphia?
  - Test: Does Midwest Express serve breakfast and Philadelphia? - BAD!

Word senses from dictionary

- Dictionary definitions:
    - right - adj. located nearer the right hand esp. being on the right when facing the same direction as the observer.
    - left - adj. located nearer to this side of the body than the right.
    - red - n. the color of blood or a ruby.
    - blood - n. the red liquid that circulates in the heart, arteries and veins of animals.
- Problems with definitions:
    - right - explicit self-reference,
    - left - implicit self-reference
    - red, blood - reference each other
- We can't extract automatically exact meanings, but we can extract meaning relations:
    - left-right - are connected and oppose each other
    - blood is a liquid
    - red is a color, may be applied to blood
- WordNet - free database of sense relations

Same-opposite relation

- **Synonyms** - words with the same meaning
  - couch/sofa, car/automobile, ...
- **Antonyms** - words with opposite meaning
  - adjectives:long/short big/little fast/slow cold/hot dark/light
  - **reversives** (реверсивные смыслы), describe change or movement in opposite directions
    - rise/fall up/down, in/out

General-specific relation

- Hyponim - more specific sense
    - car is a hyponym of vehicle
    - dog is a hyponym of animal
    - mango is a hyponym of fruit
- Hyperonim - more general sense
    - vehicle is a hyperonym for car
    - animal is a hyperonym for dog
    - fruit is a hyperonym for mango

Part-whole relation

- Meronim - part of sense
  - a leg is a meronym of chair
  - a wheel is a meronym of car
- Holonym - contains sense
  - a chair is a holonym of leg
  - a car is a holonym for wheel

Table of Contents

WordNet

- **WordNet** - free database of word senses and their relations
  - three separate databases, one-for nouns, one-for verbs and one-for adjectives and adverbs
- **synset** (for synonym set) - the set of near-synonyms
  - e.g. {fool2, chump1, gull1, mark9, patsy1, fall guy1, sucker1, soft touch1, mug2}
    - number is the sense number for given word

Sense relations in WordNet for nouns

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 concept instances	*composer*^{1} $\rightarrow$ *Bach*^{1}
Member Meronym	Has-Member	From groups to their members	*faculty*^{2} $\rightarrow$ *professor*^{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}
Substance Meronym		From substances to their subparts	*water*^{1} $\rightarrow$ *oxygen*^{1}
Substance Holonym		From parts of substances to wholes	*gin*^{1} $\rightarrow$ *martini*^{1}
Antonym		Semantic opposition between lemmas	*leader*^{1} $\Longleftrightarrow$ *follower*^{1}
Derivationally Related Form		Lemmas w/same morphological root	*destruction*^{1} $\Longleftrightarrow$ *destroy*^{1}

Sense relations in WordNet for verbs

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

Hyperonim path 1

```
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
```

Hyperonim path 2

```
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
```

Table of Contents

Word sense disambiguation

Word sense disambiguation - match sense to word in context.

- lexical sample task: for small set of words find out their senses from lexicon of predefined senses.
  - solved with usual classifiers
- all-words task: disambiguate all words in the text, given lexicon of all senses.
  - similar to POS tagging
  - but-enormous number of classes! individual class set for each word!
    - not sufficient data to fit individual classifier for each word

All words task

- Popular dataset for all-words task - SemCor
  - it is a subset of the Brown Corpus consisting of over 234,000 words
  - each word manually tagged with WordNet senses
- Features for solving this task-based on words around target word in context.
  - **collocational features**: info about words **at specific relative positions** with respect to target word
    - small context used for collocational features +-1,2,3 words.
    - wordform, its root and part-of-speech are usually extracted.
  - **bag-of-words features**: info about occurence of words in context **with no respect to their relative position**
    - vocabulary is pre-selected as some useful subset of words
    - usually ignore stopwords, punctuation, numbers
    - usually only frequent words are accounted for
    - context for baf-of-words features may be wider.

Details

- When labelled data is insufficient, we may use Wikipedia
  - extract words with contexts, for which we hve links to their descriptions
    - need to match descriptions to WordNet senses.
- Evaluation metric is usually accuracy
- Baseline - predict for each word its most common sense

Problems

- Supervised learning is most effective when we have enough data
- Datasets for distinction of meaning for particular words exist:
  - **line-hard-serve corpus** containing 4,000 sense-tagged examples of line as a noun, hard as an adjective and serve as a verb
  - **interest corpus** with 2,369 sense-tagged examples of interest as a noun
- Dataset, where **sense was disambiguated for all words** - **SemCor**.
- For all words labelling data will still be insufficient
  - beause we need to fit individual model for each word
- **Knowledge-based methods** use dictionary or thesaurus to match meaning description and context of each word.

Dictionary methods

```
function SIMPLIFIED LESK(word, sentence) returns best sense of word

  best-sense ← most frequent sense for word
  max-overlap ← 0
  context ← set of words in sentence
  for each sense in senses of word do
    signature ← set of words in the gloss and examples of sense
    overlap ← COMPUTEOVERLAP(signature, context)
    if overlap > max-overlap then
          max-overlap ← overlap
          best-sense ← sense
  end
  return(best-sense)
```

Modifications of Lesk algorithm

- Match descriptions of senses:
  - example: pine *cone*
  - descriptions of senses:

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
	2	something of this shape whether solid or hollow
	3	fruit of certain evergreen trees

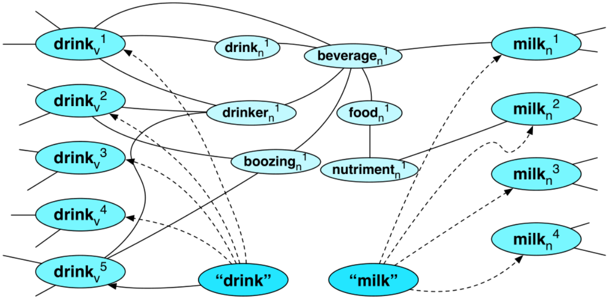
  - for cone 3 is selected (due to overlap in evergreen, pine)

Modifications of Lesk algorithm

- Weight overlapping word scores by
    - $-\ln p(w_i)$ - probability of word in concatenated descriptions & examples
    - $IDF(w_i) = \ln \frac{N_{doc}}{N_i}$ where
        - N_{doc}-total number of documents
        - N_i- number of documents, containing word w_i
        - each document is sense description or usage example sentence.
- Augment description of sense with words from labelled dataset («corpus Lesk»)
- Features, extracted by Lesk (various overlaps), can be added to any other supervised classifier

Graph-based methods

- We build a graph (V, E) for given sentence $w_1 w_2 \ldots w_n$.
  1. Initial nodes V all possible senses for each word in the sentence from WordNet
  2. Expand these senses with V' - WordNet senses that connect V to each other
     - depth-first search with small max depth
- Example subgraph for «drink milk»:



Graph-based methods

- Likelihood of each word sense from V may be measured by some node centrality measure:
  - e.g. degree (the number of edges into the node), PageRank, etc.
- Alternatively we may select the most central subgraph, consisting of senses for each word.

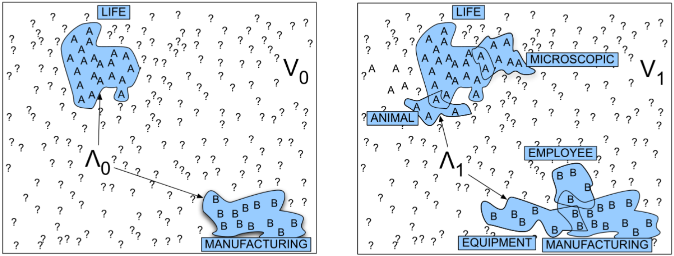
Semi-supervized sense disambiguation

- Supervised and corpus Lesk sense disambiguation algorithms work better when we have larger training set.
- To expand training set we may use semi-supervised learning:

```
hand label small training set TRAIN
while prefomance on validation improves:
   train classifier C with TRAIN
   apply classifier to test data TEST
   subset of test examples S, where C is most confident:
      label with class predictions
      remove S from TEST
      add S to TRAIN
```

Example

- Disambiguation of 2 senses of plant: plant=tree and plant=factory.
  - A=labelled objects with tree, B=labelled objects with factory, ? - test objects.
  - (a) - initial step, (b) next step of semi-supervised learning



(a) (b)

Semi-supervised method: discussion

- Semi-supervised learning works, because word contexts are very versatile, consisting of many words, but presence of each word is very informative.
- Heuristics to assist labelling of training set:
  - for each word select context word very indicative of some sense and label based on presence of this indicative word, e.g.
    - plant together with life most often means «tree»
    - plant together with manufactiring most often means «factory»
  - if word was used in some sense in discourse, then it would typically be used in the same sense in other parts of discourse.

Unsupervised sense induction

- Expensive
  - to create senses thesaurus
  - to label each word
- Unsupervised sense induction: senses for each word are created automatically with clustering.
  - If too many objects - use random subsample for clustering
  - Topic models can be applied instead of clustering..

Alogrithm

- Training:
  1. For each word-position of word w in a corpus, compute a context vector c.
  2. Cluster context vectors c for each word w. Each cluster=possible sense of w.
  3. Compute centroid of each cluster. Each centroid s_j is a sense vector.
- Testing:
  1. For word-position, containing word w in a corpus, compute a context vector c.
  2. Retrieve all sense vectors s_j for w.
  3. Assign word-position to the sense represented by the sense vector s_j most close to c.

Table of Contents

Word sences similarity

- Semantic similarity = how much two senses share in semantic meaning
- Example:
  - *bank*=fin. institution is similar to *fund*
  - bank=riverside is similar to *slope*
- Semantic similarity between words:

$$sim(w_1, w_2) = \max_{\substack{s_1 \in senses(w_1) \\ s_2 \in senses(w_2)}} sim(s_1, s_2)$$

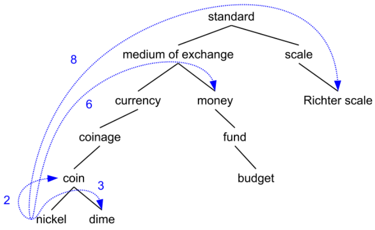
Word similarity applications

- Applications of sematic similarity between words:
    - information retrieavel: find words similar to query
    - question answering: find words relevant to question
    - summarization, machine translation: can we substitute one word with another?
    - automatic essay grading

Minimum path length

- Consider WordNet graph with only hyponim/hyperonim relations
- Define $MinPathLength(c_1, c_2)$-min path length, connecting senses c_1 and c_2 on the thesaurus graph, similarity:

$$sim(c_1, c_2) = \frac{1}{1 + \text{MinPathLength}(c_1, c_2)}$$



Minimum path length

- Each edge has weight 1, but in reality semantic weight should be different
  - upper in the graph - difference in senses is larger
    - lower in the graph - difference in senses is lower
    - we can add weights based on proximity to root of the graph
- Alternatively we can calculate weight of eachstep based on corpus statistics.

Using corpus statistics

- Suppose we have text corpus with labelled sense for each word. We can calculate $p(c)$ for each sense c.
  - if sense appears it counts for itself and all more general senses.
- Define $LCS(c_1, c_2)$ [lowest common subsumer of senses c_1 and c_2] - the most specific common concept-hyperonim both for c_1 and c_2.
- Define information revealed by random event E be $-\ln P(E)$
- Measures for sense similarity:

$$sim(c_1, c_2) = -\ln p(LCS(c_1, c_2))$$

$$sim(c_1, c_2) = \frac{-2\ln p(LCS(c_1, c_2))}{-\ln p(c_1) - \ln p(c_2)}$$

$$sim(c_1, c_2) = \frac{1}{-2\ln p(LCS(c_1, c_2)) + \ln p(c_1) + \ln p(c_2)}$$

Using dictionary definition

- Let $def(c)$ be definition of sense c.
- We may measure similarity between c_1 and c_2 by counting number of words appearing in their definitions:

$$sim(c_1, c_2) = OVERLAP(def(c_1), def(c_2))$$

- each word can contribute nun=uniform weight, say $IDF(w)$
- we may also take into account overlap between:
    - sentences with sense usages
    - words with given sense
- we may calculate overlap between definitions of hyponims and hyperonims of c_1 and c_2.