Lexical Semantics & WSD

Computertaalkunde
December 8, 2014

Walter Daelemans Walter.Daelemans@uantwerpen.be
Guy De Pauw Guy.Depauw@uantwerpen.be

Exam

- January 12, 2014, 9.00-17.00
- Python exercise (to be decided tomorrow)

Program

- Introduction to lexical semantics
  - Senses, relations between senses
- Word Sense Disambiguation (WSD) and word similarity
  - Task description
  - 3 types of approaches
  - Applications
- Assignment

Lexical Semantics

Word senses

- Wordforms and lemmas
  - appeltjes appel
  - lopen lopen (V)
  - lopen loop (N)
- Lemmas have lexical meaning
- One lemma can have many different (word) senses
  - Discrete representation of aspects of a word lemma's meaning
- Senses, rather than words, are important in NLP systems:
  - Machine translation: bank → bank or oever
  - Text categorization: python → snake or programming language
  - Text to speech: bass → music or fishing

Distinguishing senses

- Word can have many senses, see WordNet:
  - Bank as noun: 10 senses
  - Bank as verb: 8 senses
- Sometimes subtle differences:
  - Bank: sloping land
  - Bank: a slope in the turn of a road or track
- Rule of thumb:
  - Different truth conditions, syntactic behavior
  - Zeugma

Relations between senses

- Homonymy in case of same form but unrelated meaning
  - bank^1 : financial institution
  - bank^2: sloping mound
- Polysemy if there is a systematic relation: bank^1 and bank^3
  - bank^3: biological repository
- Metonymy: systematic polysemy
  - E.g. Building – Organization, Author – Work of the author
  - *Jane Austen is on the top shelf – Jane Austen wrote Emma*

Relations between senses

- Synonymy (synonyms mean the same in all contexts, same propositional meaning, same truth conditions): *couch/sofa*
- Perfect synonymy is rare
  - big car, large car
  - big sister, large sister ?
- Synonymy is a relation between senses rather than between words

Relations between senses

- Antonymy:
  - Different ends of a scale: *long/short*; *dark/light*
  - Reversives: *up/down*
- Hyponymy: *car/vehicle* (*x is subordinate, hyponym of y*) (*y is superordinate, hypernym of x*)
- Hyponymy mostly associative
  - Grape is hyponym of Fruit, Fruit is hyponym of Edible Things
  - Grape is hyponym of Edible Things
- Classes and instances
  - Relation between instance and class versus relation between classes
  - ISA-hierarchy, AKO-hierarchy
  - Antwerp ISA city, city AKO location

Relations between senses

- Meronymy: *wheel/car* (*x is-part-of y*) (*y is holonym of x*)
- Semantic field
  - *Reservation, flight, travel, buy, price, cost, fare, rates, plane*

Structured lexical resources

- Dictionaries available in machine-readable form
  - Contains list of senses, definitions for all senses, typical usage examples for most senses
  - E.g. Oxford English Dictionary, Collins, Longman Dictionary of Ordinary Contemporary English
- Thesaurus
  - Contains explicit semantic relation information between word senses
  - E.g. Roget's Thesaurus
- Lexical database
  - Contains relations between senses, definitions, etc.
  - E.g. WordNet, EuroWordNet

WordNet (Fellbaum 1998)

- Combination of dictionary, thesaurus & semantic network
- Database of lexical relations in 3 parts: nouns, verbs, adjectives & adverbs
- Word : senses
  - Sense : gloss, synset (= set of near-synonyms)
- Downloadable resource
- Web interface
  - http://wordnetweb.princeton.edu/perl/webwn

Word to search for: be [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) beryllium, **Be**, glucinium, atomic number 4 (a light strong brittle grey toxic bivalent metallic element)

Verb

- S: (v) **be** (have the quality of being; (copula, used with an adjective or a predicate noun)) *"John is rich"; "This is not a good answer"*
- S: (v) **be** (be identical to; be someone or something) *"The president of the company is John Smith"; "This is my house"*
- S: (v) **be** (occupy a certain position or area; be somewhere) *"Where is my umbrella?"; "The toolshed is in the back"; "What is behind this behavior?"*
- S: (v) exist, **be** (have an existence, be extant) *"Is there a God?"*
- S: (v) **be** (happen, occur, take place) *"I lost my wallet; this was during the visit to my parents' house"; "There were two hundred people at his funeral"; "There was a lot of noise in the kitchen"*
- S: (v) equal, **be** (be identical or equivalent to) *"One dollar equals 1,000 rubles these days!"*
- S: (v) constitute, represent, make up, comprise, **be** (form or compose) *"This*

NLTK

```
>>> from nltk.corpus import wordnet as wn
```

Which synsets does a word have (of particular POS: VERB, NOUN, ADJ, ADV)

```
>>> wn.synsets('dog')

[Synset('dog.n.01'), Synset('frump.n.01'),
Synset('dog.n.03'), Synset('cad.n.01'),
Synset('frank.n.02'), Synset('pawl.n.01'),
Synset('andiron.n.01'), Synset('chase.v.01')]

>>> wn.synsets('dog', pos=wn.VERB)

[Synset('chase.v.01')]
```

NLTK

Properties of Synsets

```
>>> dog = wn.synset('dog.n.01')

>>> dog.hypernyms()

[Synset('domestic_animal.n.01'), Synset('canine.n.02')]

>>> dog.hyponyms()

[Synset('puppy.n.01'), Synset('great_pyrenees.n.01'),
Synset('basenji.n.01'), Synset('newfoundland.n.01'),
Synset('lapdog.n.01'), Synset('poodle.n.01'), Synset('leonberg.n.
01'), Synset('toy_dog.n.01'), Synset('spitz.n.01'), Synset('pooch.n.
01'), Synset('cur.n.01'), Synset('mexican_hairless.n.01'),
Synset('hunting_dog.n.01'), Synset('working_dog.n.01'),
Synset('dalmatian.n.02'), Synset('pug.n.01'), Synset('corgi.n.01'),
Synset('griffon.n.02')]
```

Word Sense Disambiguation

Extreme cases of ambiguity

Drunk Gets Nine Years In Violin Case

Farmer Bill Dies In House

Prostitutes Appeal To Pope

Stolen Painting Found By Tree

Red Tape Holds Up New Bridge

Include Children When Baking Cookies

Miners Refuse To Work After Death

Problems & solutions

Drunk Gets Nine Years In Violin Case

Farmer Bill Dies In House

Prostitutes Appeal To Pope

Stolen Painting Found By Tree

Red Tape Holds Up New Bridge

Include Children When Baking Cookies

Miners Refuse To Work After Death

Lexical, syntactic, referential ambiguity

World Knowledge
Fixed Expressions

Word Sense Ambiguity

- Most of the time no problem for humans, except in some extreme cases
- Computers need help to disambiguate even the ‘simplest’ of cases

Computationally explosive problem

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

26 senses

11 senses

4 senses

8 senses

5 senses

4 senses

10 senses

8 senses

3 senses

43,929,600 senses

How big is the problem?

- Most words in English have only one sense
  - 62% in Longmans Dictionary of Contemporary English
  - 79% in WordNet
- Average number of senses per word
  - 3.83 in LDOCE vs. 2.96 in WordNet
- But … ambiguous words are used more frequently!
  - BNC (British National Corpus): 84% of words have more than one sense
  - Some senses are more frequent than others

Word Sense Disambiguation (WSD)

- = automatically identify the intended sense of a word *in context*
- Assumes a fixed inventory of senses that you can select the right one from
- Can be seen as a categorization task (cf. POS-tagging)
  - Senses = classes
  - Context = features

Relevance

- Important aspect of many NLP applications
- Relevant for all languages
- Needed in
  - *Machine translation*: select the right sense to translate
  - *Information retrieval*: resolve ambiguity in query
  - *Information extraction*: accurate analysis of text

Upper bound and baseline

- Human performance as an upper bound
  - Fine-grained sense inventories: 75-80% human agreement
  - Coarser-grained inventories: 90% human agreement possible
- Predict the most frequent sense in a given lexical resource (‘MFS baseline’)
  - *bank* 97.20%
  - *bar* 47.38%

Evaluation of WSD

- Internal: measure accuracy of sense selection compared to gold standard
- External: integrate WSD in MT or IR system and evaluate
- Test data
  - Lexical sample: the occurrences of a small sample of target words need to be disambiguated
  - All-words: all words in running text need to be disambiguated
  - Cf. SensEval competitions http://www.senseval.org

Development of research in WSD

- Noted as problem for Machine Translation (Weaver, 1949)
- Bar-Hillel (1960) declared it unsolvable, left the field of MT
  - The box is in the pen. The pen is in the box.
- 1970s-80s Rule-based approaches
- 1990s Corpus-based approaches

  Dependence on sense-tagged training texts
- 2000s Hybrid Systems

  Unsupervised learning

  Taking advantage of the Web

Approaches to WSD

Knowledge-based : External lexical resources

Supervised : Labeled training data

Semi-supervised

Unsupervised : Large collections of raw text

1. Knowledge-based approaches

WSD from sense definitions

- LESK algorithm (Lesk, 1986)
  - Retrieve from dictionary all sense definitions of the word to be disambiguated
  - Determine the overlap between each sense definition and definitions of words in the current context
  - Choose the sense that leads to highest overlap

LESK algorithm example

e.g. Pine cones hanging in a tree

Pine[1]	kind of evergreen tree with needle-shaped leaves
Pine[2]	waste away through sorrow or illness
Cone[1]	solid body which narrows to a point
Cone[2]	something of this shape whether solid or hollow
Cone[3]	fruit of certain evergreen trees

Pine 1 ∩ Cone 1 = 0
Pine 2 ∩ Cone 1 = 0
Pine 1 ∩ Cone 2 = 0
Pine 2 ∩ Cone 2 = 0
Pine 1 ∩ Cone 3 = 2
Pine 2 ∩ Cone 3 = 0

Problems with LESK algorithm

- Problems
  - Very sensitive to the exact wording of definitions: absence of a certain word can radically change the results
  - Dictionary glosses tend to be fairly short; often not sufficient vocabulary to relate fine-grained sense distinctions

LESK variants

- Simplified LESK
  - Retrieve all sense definitions of target word
  - Compare with *context* instead of sense definitions of the context
  - e.g. Pine cones hanging in a tree

Pine 1 ∩ Sentence = 1
Pine 2 ∩ Sentence = 0

- Corpus-based LESK
  - Add context words from sense tagged corpus to definitions
  - Weight words by inverse document frequency (IDF)
  - Gloss is the document
  - $IDF(w) = -\log (d_w/D)$ (function words have low IDF)
  - Best-performing LESK variant, baseline in SensEval competitions

2. Supervised approaches

Supervised learning

- Last 15-20 years: shift from manually crafted systems to automated classification methods
- Basic steps
  - Collect a set of examples that illustrate the various possible classifications or outcomes of an event
  - Identify patterns in the examples
  - Generalize those patterns into rules
  - Apply the rules to classify a new event

Supervised WSD

- Resources
  - Sense-tagged text (*unstructured*)
  - Dictionaries, thesauri, semantic networks (*structured*)
  - Syntactic Analysis (POS tagger, chunker, parser, etc.)
- WSD as a classification problem
  - target word is assigned the most appropriate sense
  - from a given set of possibilities
  - based on the context in which it occurs
  - = word expert approach

Sense-tagged corpora

- SemCor [Miller et al. 1993]: 352 texts tagged with around 234,000 sense annotations
- MultiSemCor [Pianta et al. 2002]: English-Italian parallel corpus annotated with WordNet senses
- line-hard-serve corpus [Leacock et al. 1993]: 4000 sense-tagged examples
- interest corpus [Bruce and Wiebe 1994]: 2369 sense-labeled examples of noun interest
- DSO corpus [Ng and Lee 1996]: 192,800 sense-tagged tokens of 191 words from the Brown and WSJ corpora
- Open Mind Word Expert corpus [Chklovski and Mihalcea 2002], 288 nouns semantically annotated by Web users in a collaborative effort
- Senseval / Semeval data sets ⇒ Nearly all annotated with different versions of the WordNet sense inventory

Sense-tagged corpora (2)

e.g.

Bonnie and Clyde are two really famous criminals, I think they were bank/1 robbers.

My bank/1 charges too much for an overdraft.

I went to the bank/1 to deposit my check and get a new ATM card.

The University of Minnesota has an East and a West Bank/2 campus right on the Mississippi River.

My grandfather planted his pole in the bank/2 and got a great big catfish!

The bank/2 is pretty muddy, I can't walk there.

Simple supervised system

- Extract bags of words from sense-tagged text
  - #1 financial-bank

    a an and are ATM Bonnie card charges check Clyde criminals deposit famous for get I much My new overdraft really robbers the they think to too two went were
  - #2 river-bank

    a an and big campus cant catfish East got grandfather great has his I in is Minnesota Mississippi muddy My of on planted pole pretty right River The the there University walk West

Simple supervised system (2)

Given a sentence S containing <u>bank</u>

For each word W_i in S

If W_i is in *financial-bank* then Sense#1 = Sense#1 + 1

If W_i is in *river-bank* then Sense#2 = Sense#2 + 1

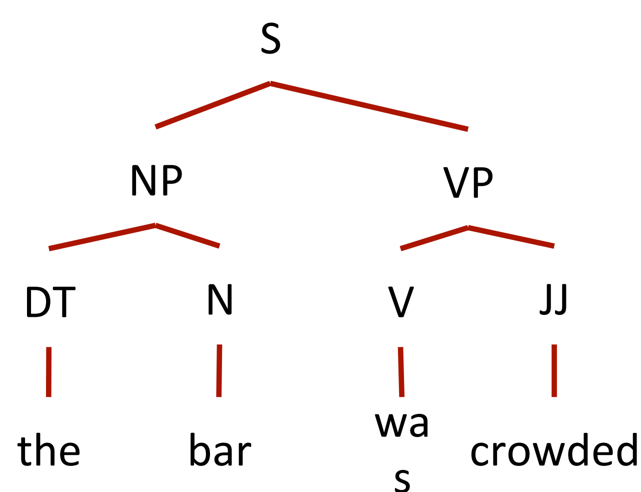
If Sense 1 > Sense 2 then print “Financial”

else if Sense 2 > Sense 1 then print “River”

else print “Financial” (majority sense)

Supervised methodology

1. Tokenization (The, bar, was, crowded)
2. POS tagging (DT, NN, VBD, JJ)
3. Lemmatization (The, bar, be, crowded)
4. Chunking (DT+NN/NP, VBD+JJ/VP)
5. Parsing



Supervised methodology (2)

- Features for WSD retrieved from preprocessing information
  - local features : local context of word usage (e.g. POS, lemma, etc.)
  - topical features : general topic of a text or discourse
  - syntactic features : syntactic cues and argument-head relations between the target word and other words
  - semantic features : representing semantic information, e.g. previously established senses of words in context, domain indicators, etc.

Representing context (1)

- Using these features, convert each word occurrence into a feature vector

My father used to fish along the banks/SHORE of the river

The bank/FINANCE issued a check for the amount of interest

P-1	P+1	P+2	Fish	Check	River	Interest	SENSE TAG
det	prep	det	1	0	1	0	SHORE
det	verb	det	0	1	0	1	FINANCE

Representing context (2)

- Which context words are taken into account?
  - No function words
  - Only words that are in a specific grammatical relation
  - Size of the window
- How are they represented in the vector?
  - Binary: present/not present
  - Continuous: Relative frequency, mutual information
- How is the similarity between vectors measured?

Supervised methodology (4)

- Use any supervised learning algorithm
  - Lazy learners
    - e.g. k-Nearest Neighbor Classifiers
  - Eager learners
    - e.g. Support Vector Machines, Decision Trees, naïve Bayes
- Training data to train and validate the machine learner
- Procedure: *n*-fold cross-validation
- Hold-out test data to test the resulting classifier

Shortcomings

- Supervised approaches to WSD achieve best results, but
  - heavily rely on large sense-tagged corpora
  - fixed sense inventory: often arbitrary divisions of word meanings into dictionary senses
  - low inter-annotator agreements on sense tagging
- WSD should be integrated in real applications such as MT or multilingual IR (cf. extrinsic evaluation)

3. Unsupervised and semi-supervised approaches

Motivation

- Supervised yields highest performance, but…
  - Limited to words whose senses are tagged
  - *Corpus Annotation Bottleneck*
- Solutions: raw corpora instead of sense-tagged text or lexical resources

Unsupervised / minimally supervised

- = learning sense classifiers from annotated data, with minimal or no human supervision
- Examples
  - Sense clustering
  - Automatically bootstrap a corpus starting with *a few human annotated examples*
  - Cross-lingual evidence
  - Use Wikipedia as sense-tagged text

Sense clustering

- Also *word sense induction/discrimination*
- Cluster words on similarity of context (using distributions and similarity metrics)
- Hypothesis:
  - Words with similar meanings tend to occur in similar contexts (Miller and Charles, 1991)
  - Cf. ‘You shall know a word by the company it keeps’ (Firth, 1957)

Bootstrapping

- Build sense classifiers with little training data
  - Expand applicability of supervised WSD
- Components
  - (Some) labeled data
  - (Large amounts of) unlabeled data
  - (One or more) basic classifiers
- Output
  - Classifier that improves over the basic classifiers

Bootstrapping algorithm

- Bootstrapping algorithm (Yarowsky, 1995)
  - Start from small seed set of hand-labeled data Λ_0
  - Learn decision-list classifier from Λ_0
  - Use learned classifier to label unlabeled data V_0
  - Move high-confidence examples in V_0 to Λ_1
  - Repeat until low training error or no longer confident tagging
- 2 heuristics to automatically select Λ_0
  - One sense per collocation: bass/fish & bass/play
  - One sense per discourse: within a text or discourse, you will find either bass/fish or bass/play, not both

Selectional restrictions

- Constrain the possible meanings of words in a context
- Constraints on the semantic type that a word sense imposes on the words with which it combines in sentences – using grammatical relationships like agent, patient, instrument etc.

e.g. *eat*(*x*,y) (verb)

x : animate entity as subject – 'agent'

y : edible entity as direct object – 'theme'/'edible-thing'

- wash a dish versus serve a dish
- Problem: selectional restrictions are often violated
  - But it fell apart in 1931, perhaps because people realized that you can't eat gold for lunch if you're hungry.

WSD using cross-lingual evidence

- Corpus-based approach: using translations from a parallel corpus instead of human-defined sense labels
- Advantages
  - easier to integrate in real applications
  - implicitly deals with granularity problem
  - language-independent approach
- Hypothesis: different sense distinctions are often lexicalized across languages

Cross-lingual WSD (SemEval 2010)

- living on the bank of the river

Dutch	oever/dijk
French	rives/rivage/bord/bords
German	Ufer
Italian	riva
Spanish	orilla

- money supply is of direct interest to any bank

Dutch	bank/kredietinstelling
French	banque/établissement de crédit
German	Bank/Kreditinstitut
Italian	banca
Spanish	banco

Wikipedia as sense-tagged corpus

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]].

(Mihalcea, 2007)

- For most investigated words, performance using Wikipedia improves over MFS and LESK baselines
- Advantages
  - Size of Wikipedia is growing
  - Wikipedia is available for about 200 different languages

Word sense	Labels in Wikipedia	Wikipedia definition	WordNet definition
bar (establishment)	bar_(establishment), nightclub gay_club, pub	a retail establishment which serves alcoholic beverages	a room or establishment where alcoholic drinks are served over a counter
bar (counter)	bar_(counter)	the counter from which drinks are dispensed	a counter where you can obtain food or drink
bar (unit)	bar_(unit)	a scientific unit of pressure	a unit of pressure equal to a million dynes per square centimeter
bar (music)	bar_(music), measure_music musical_notation	a period of music	musical notation for a repeating pattern of musical beats
bar (law)	bar_association, bar_law law_society_of_upper_canada state_bar_of_california	the community of persons engaged in the practice of law	the body of individuals qualified to practice law in a particular jurisdiction
bar (landform)	bar_(landform)	a type of beach behind which lies a lagoon	a submerged (or partly submerged) ridge in a river or along a shore
bar (metal)	bar_metal, pole_(object)	-	a rigid piece of metal or wood
bar (sports)	gymnastics_uneven_bars, handle_bar	-	a horizontal rod that serves as a support for gymnasts as they perform exercises
bar (solid)	candy_bar, chocolate_bar	-	a block of solid substance

Table 1: Word senses for the word *bar*, based on annotation labels used in Wikipedia

Taken from Mihalcea 2007 ‘Using Wikipedia for Automatic Word Sense Disambiguation’

Applications of WSD

Information Retrieval

- Word sense ambiguity is one of the reasons for poor performance of IR systems / Search engines

e.g. Find all Web Pages about <u>cricket</u>

Machine Translation

- = the use of computers to conduct large-scale translation operations

Translate logiciel Cordial from French to English

Google Translate: *software friendly*

Information Extraction

- Automatically extract structured information
- Typical subtasks of IE
  - named entity recognition (NER)
  - terminology extraction
  - relationship extraction

Other applications

- Content analysis: the analysis of the general content of a text in terms of its ideas, themes, etc.
- Lexicography: WSD can help provide empirical sense groupings and statistically significant indicators of context for new or existing senses.
- Semantic Web: needs domain-oriented and unrestricted sense disambiguation to deal with the semantics of (Web) documents, and enable interoperability between systems, ontologies, and users

Conclusions

- Introduction to lexical semantics & WSD
  - Problem of polysemy
  - What makes it difficult for computers?
- (Roughly) 3 types of approaches to automatically identify the right sense of the word in context
  - Knowledge-based approaches are dependent on the source
  - Supervised approaches are often limited to a specific domain and require considerable human effort
  - Semi-supervised/unsupervised approaches are promising when looking at effective large-scale, up-to-date WSD

Assignment

Assignment by 29 Dec 2014

1. For three ambiguous verbs
   - Google the words and take the top-*5* snippets and web pages
     - Using sense definitions from WordNet and an English-language dictionary, explain how original/simplified/corpus-based LESK work
     - Is any of these successful at WSD?
     - Which technique(s) would be more successful and why?
2. For the noun CLUB
   - Compute word similarity with three words: football, golf, country.
   - Find out which types of word and sense similarity are available.
   - Write down the shortest paths between the words using WordNet (http://wordnetweb.princeton.edu/perl/webwn or nltk)

- Send answers (walter.daelemans@uantwerpen.be)
  - For 1: report with all data used, for 2: paths and similarities