Lexical Semantics

Ling571
Deep Processing Techniques for NLP
February 22, 2016

Roadmap

- Lexical semantics
  - Motivation & definitions
  - Word senses
  - Tasks:
    - Word sense disambiguation
    - Word sense similarity
  - Distributional similarity

What is a plant?

There are more kinds of **plants** and animals in the rainforests than anywhere else on Earth. Over half of the millions of known species of **plants** and animals live in the rainforest. Many are found nowhere else. There are even **plant**s and animals in the rainforest that we have not yet discovered.

The Paulus company was founded in 1938. Since those days the product range has been the subject of constant expansions and is brought up continuously to correspond with the state of the art. We’ re engineering, manufacturing, and commissioning world-wide ready-to-run **plants** packed with our comprehensive know-how.

Lexical Semantics

- So far, word meanings discrete
  - Constants, predicates, functions
- Focus on word meanings:
  - Relations of meaning among words
    - Similarities & differences of meaning in sim context
  - Internal meaning structure of words
    - Basic internal units combine for meaning

Terminology

- **Lexeme**:
  - Form: Orthographic/phonological + meaning
  - Represented by lemma
    - **Lemma**: citation form; infinitive in inflection
      - Sing: sing, sings, sang, sung,...
- **Lexicon**: finite list of lexemes

Sources of Confusion

- Homonymy:
  - Words have same form but different meanings
    - Generally same POS, but unrelated meaning
    - E.g. bank (side of river) vs bank (financial institution)
      - bank^1 vs bank^2
    - Homophones: same phonology, diff't orthographic form
      - E.g. two, to, too
    - Homographs: Same orthography, diff't phonology
- Why do we care?
  - Problem for applications: TTS, ASR transcription, IR

Sources of Confusion II

- Polysemy
  - Multiple RELATED senses
    - E.g. bank: money, organ, blood,...
  - Big issue in lexicography
    - # of senses, relations among senses, differentiation
    - E.g. serve breakfast, serve Philadelphia, serve time

Relations between Senses

- Synonymy:
  - (near) identical meaning
  - Substitutability
    - Maintains propositional meaning
  - Issues:
    - Polysemy – same as some sense
    - Shades of meaning – other associations:
      - Price/fare; big/large; water H_2O
    - Collocational constraints: e.g. babbling brook
    - Register:
      - social factors: e.g. politeness, formality

Relations between Senses

- Antonyms:
  - Opposition
    - Typically ends of a scale
      - Fast/slow; big/little
    - Can be hard to distinguish automatically from syns
- Hyponomy:
  - Isa relations:
    - More General (hypernym) vs more specific (hyponym)
      - E.g. dog/golden retriever; fruit/mango;
  - Organize as ontology/taxonomy

Word Sense Disambiguation

- Application of lexical semantics
- Goal: Given a word *in context,* identify the appropriate sense
  - E.g. plants and animals in the rainforest
- Crucial for real syntactic & semantic analysis
  - Correct sense can determine
    - Available syntactic structure
    - Available thematic roles, correct meaning,..

Robust Disambiguation

- Learning approaches
  - Supervised, Bootstrapped, Unsupervised

- Knowledge-based approaches
  - Dictionaries, Taxonomies

- Widen notion of context for sense selection
  - Words within window (2,50,discourse)
  - Narrow cooccurrence - collocations

There are more kinds of **plants** and animals in the rainforests than anywhere else on Earth. Over half of the millions of known species of **plants** and animals live in the rainforest. Many are found nowhere else. There are even **plants** and animals in the rainforest that we have not yet discovered.

Biological Example

The Paulus company was founded in 1938. Since those days the product range has been the subject of constant expansions and is brought up continuously to correspond with the state of the art. We' re engineering, manufacturing and commissioning world-wide ready-to-run **plants** packed with our comprehensive know-how. Our Product Range includes pneumatic conveying systems for carbon, carbide, sand, lime and many others. We use reagent injection in molten metal for the…

Industrial Example

Label the First Use of "Plant"

Disambiguation Features

- Key: What are the features?
  - Part of speech
    - Of word and neighbors
  - Morphologically simplified form
  - Words in neighborhood
    - Question: How big a neighborhood?
      - Is there a single optimal size? Why?
  - (Possibly shallow) Syntactic analysis
    - E.g. predicate-argument relations, modification, phrases
  - Collocation vs co-occurrence features
    - Collocation: words in specific relation: p-a, 1 word +/-
    - Co-occurrence: bag of words..

WSD Evaluation

- Ideally, end-to-end evaluation with WSD component
  - Demonstrate real impact of technique in system
  - Difficult, expensive, still application specific
- Typically, intrinsic, sense-based
  - Accuracy, precision, recall
  - SENSEVAL/SEMEVAL: all words, lexical sample
- Baseline:
  - Most frequent sense
- Topline:
  - Human inter-rater agreement: 75-80% fine; 90% coarse

Word Similarity

- Synonymy:
  - True propositional substitutability is rare, slippery
- Word similarity (semantic distance):
  - Looser notion, more flexible
  - Appropriate to applications:
    - IR, summarization, MT, essay scoring
      - Don’t need binary +/- synonym decision
      - Want terms/documents that have high similarity
        - Differ from relatedness
- Approaches:
  - Distributional
  - Thesaurus-based

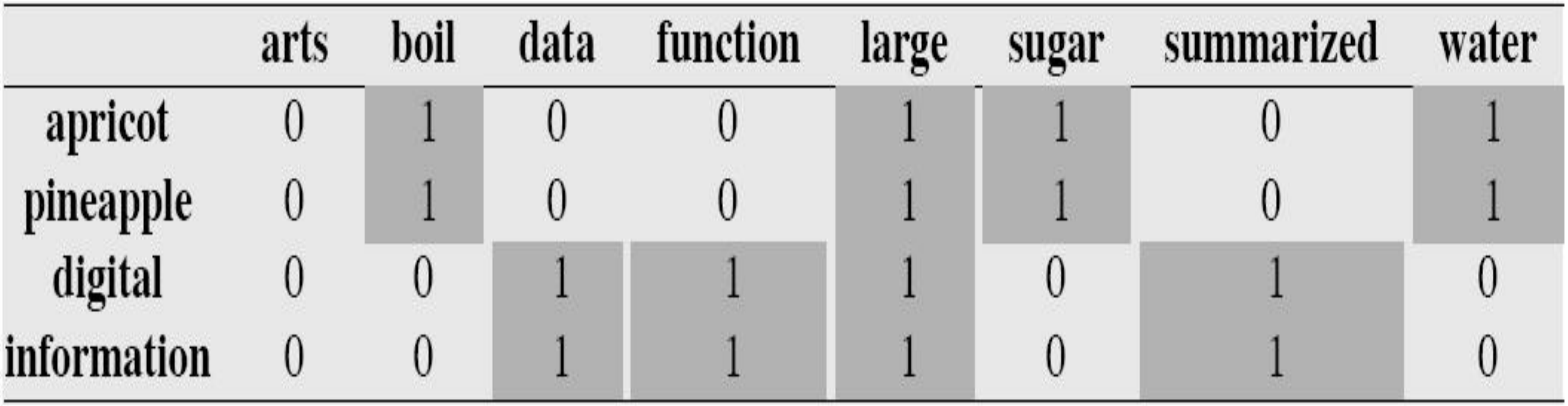
Distributional Similarity

- Unsupervised approach:
  - Clustering, WSD, automatic thesaurus enrichment
- Insight:
  - “You shall know a word by the company it keeps!”
    - (Firth, 1957)
  - A bottle of *tezguino* is on the table.
  - Everybody likes *tezguino*.
  - *Tezguino* makes you drunk.
  - We make *tezguino* from corn.
- Tezguino: corn-based, alcoholic beverage

Distributional Similarity

- Represent ‘company’ of word such that similar words will have similar representations
  - ‘Company’ = context
- Word represented by context feature vector
  - Many alternatives for vector
- Initial representation:
  - ‘Bag of words’ binary feature vector
  - Feature vector length N, where N is size of vocabulary
    - f_i= 1 if word_i within window of *w*, 0 o.w.

Binary Feature Vector



	arts	boil	data	function	large	sugar	summarized	water
apricot	0	1	0	0	1	1	0	1
pineapple	0	1	0	0	1	1	0	1
digital	0	0	1	1	1	0	1	0
information	0	0	1	1	1	0	1	0

Distributional Similarity Questions

- What is the right neighborhood?
  - What is the context?

- How should we weight the features?

- How can we compute similarity between vectors?

Feature Vector Design

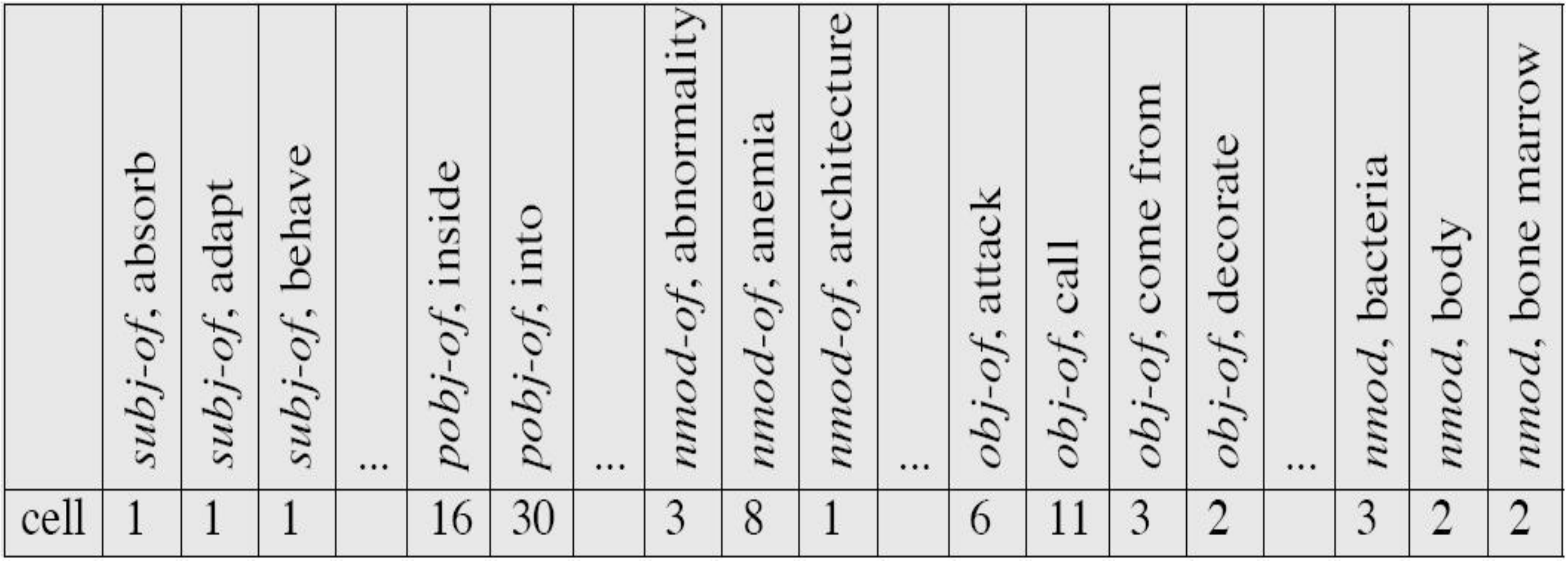
- Window size:
  - How many words in the neighborhood?
    - Tradeoff:
      - +/- 500 words: 'topical context'
      - +/- 1 or 2 words: collocations, predicate-argument
      - Only words in some grammatical relation
        - Parse text (dependency)
        - Include subj-verb; verb-obj; adj-mod
          - NxR vector: word x relation

Context Windows

- Same corpus, different windows
  - BNC
  - Nearest neighbors of “dog”

- 2-word window:
  - Cat, horse, fox, pet, rabbit, pig, animal, mongrel, sheep, pigeon
- 30-word window:
  - Kennel, puppy, pet, terrier, Rottweiler, canine, cat, to bark, Alsatian

Example Lin Relation Vector



	subj-of, absorb	*subj-of*, adapt	*subj-of*, behave	...	*pobj-of*, inside	*pobj-of*, into	...	*nmod-of*, abnormality	*nmod-of*, anemia	*nmod-of*, architecture	...	*obj-of*, attack	*obj-of*, call	*obj-of*, come from	*obj-of*, decorate	...	*nmod*, bacteria	*nmod*, body	*nmod*, bone marrow
cell	1	1	1		16	30		3	8	1		6	11	3	2		3	2	2

Weighting Features

- Baseline: Binary (0/1)
  - Minimally informative
  - Can’t capture intuition that frequent features informative
- Frequency or Probability:

$$P(f \mid w) = \frac{count(f,w)}{count(w)}$$

  - Better but,
  - Can overweight a priori frequent features
    - Chance cooccurrence

Pointwise Mutual Information

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

PMI:

- Contrasts observed cooccurrence
- With that expected by chance (if independent)

- Generally only use positive values
  - Negatives inaccurate unless corpus huge

Lin Association

- Recall:
  - Lin’s vectors include:
    - r: dependency relation
    - w’: other word in dependency relation
- Decomposes weights on that basis:

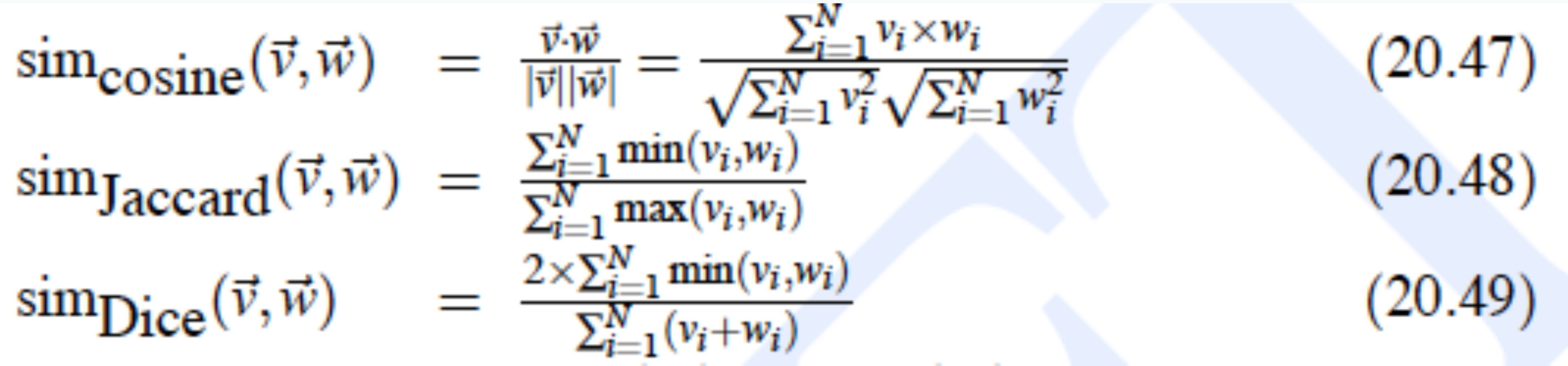
$$\text{assoc}_{\text{Lin}}(w,f) = \log_2 \frac{P(w,f)}{P(w)P(r|w)P(w'|w)}$$

Vector Similarity

- Euclidean or Manhattan distances:
  - Too sensitive to extreme values
- Dot product: $sim_{dot-product}(\vec{v},\vec{w}) = \vec{v} \bullet \vec{w} = \sum_{i=1}^{N} v_i \times w_i$
  - Favors long vectors:
    - More features or higher values
- Cosine: $sim_{\cos ine}(\vec{v},\vec{w}) = \dfrac{\sum_{i=1}^{N} v_i \times w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}}$

Alternative Weighting Schemes

- Models have used alternate weights of computing similarity based on weighted overlap



$$\text{sim}_{\text{cosine}}(\vec{v},\vec{w}) = \frac{\vec{v}\cdot\vec{w}}{|\vec{v}||\vec{w}|} = \frac{\sum_{i=1}^{N} v_i \times w_i}{\sqrt{\sum_{i=1}^{N} v_i^2}\sqrt{\sum_{i=1}^{N} w_i^2}} \tag{20.47}$$

$$\text{sim}_{\text{Jaccard}}(\vec{v},\vec{w}) = \frac{\sum_{i=1}^{N} \min(v_i,w_i)}{\sum_{i=1}^{N} \max(v_i,w_i)} \tag{20.48}$$

$$\text{sim}_{\text{Dice}}(\vec{v},\vec{w}) = \frac{2\times\sum_{i=1}^{N} \min(v_i,w_i)}{\sum_{i=1}^{N} (v_i+w_i)} \tag{20.49}$$

Results

- Based on $\text{Lin}_{\text{assoc}}$
  - Hope (N): optimism, chance, expectation, prospect, dream, desire, fear
  - Hope (V): would like, wish, plan, say, believe, think

  - Brief (N): legal brief, affidavit, filing, petition, document, argument, letter
  - Brief (A): lengthy, hour-long, short, extended, frequent, recent, short-lived, prolonged, week-long