SI485i : NLP

Set 10
Lexical Relations

slides adapted from Dan Jurafsky and Bill MacCartney
Three levels of meaning

1. Lexical Semantics (words)

2. Sentential / Compositional / Formal Semantics

3. Discourse or Pragmatics
   - meaning + context + world knowledge
The unit of meaning is a *sense*

- One word can have multiple meanings:
  - *Instead, a* bank *can hold the investments in a custodial account in the client’s name.*
  - *But as agriculture burgeons on the east bank, the river will shrink even more.*
- A **word sense** is a representation of one aspect of the meaning of a word.
- **bank** here has two senses
Terminology

- **Lexeme**: a pairing of meaning and form
- **Lemma**: the word form that represents a lexeme
  - *Carpet* is the lemma for *carpets*
  - *Dormir* is the lemma for *duermes*

- The lemma *bank* has two **senses**:
  - *Financial institution*
  - *Soil wall next to water*

- A **sense** is a discrete representation of one aspect of the meaning of a word
Relations between words/senses

- Homonymy
- Polysemy
- Synonymy
- Antonymy
- Hypernymy
- Hyponymy
- Meronymy
Homonymy

- Homonyms: lexemes that share a form, but unrelated meanings

- Examples:
  - *bat* (wooden stick thing) vs *bat* (flying scary mammal)
  - *bank* (financial institution) vs *bank* (riverside)

- Can be homophones, homographs, or both:
  - Homophones: *write and right, piece and peace*
  - Homographs: *bass and bass*
Homonymy, yikes!

Homonymy causes problems for NLP applications:

- Text-to-Speech
- Information retrieval
- Machine Translation
- Speech recognition

Why?
Polysemy

- **Polysemy**: when a single word has multiple related meanings (bank the building, bank the financial institution, bank the biological repository)

- Most non-rare words have multiple meanings
Polysemy

1. The **bank** was constructed in 1875 out of local red brick.
2. I withdrew the money from the **bank**.

- Are those the same meaning?
  - We might define meaning 1 as: “The building belonging to a financial institution”
  - And meaning 2: “A financial institution”
How do we know when a word has more than one sense?

• The “zeugma” test!

• Take two different uses of *serve*:
  • *Which flights serve breakfast?*
  • *Does America West serve Philadelphia?*

• Combine the two:
  • *Does United serve breakfast and San Jose? (BAD, TWO SENSES)*
Synonyms

• Word that have the same meaning in some or all contexts.
  • couch / sofa
  • big / large
  • automobile / car
  • vomit / throw up
  • water / H₂O
Synonyms

• But there are few (or no) examples of perfect synonymy.
  • Why should that be?
  • Even if many aspects of meaning are identical
  • Still may not preserve the acceptability based on notions of politeness, slang, register, genre, etc.

• Example:
  • Big/large
  • Brave/courageous
  • Water and H₂O
Antonyms

- Senses that are opposites with respect to one feature of their meaning

- Otherwise, they are very similar!
  - dark / light
  - short / long
  - hot / cold
  - up / down
  - in / out
Hyponyms and Hypernyms

- **Hyponym**: the sense is a subclass of another sense
  - *car* is a hyponym of *vehicle*
  - *dog* is a hyponym of *animal*
  - *mango* is a hyponym of *fruit*

- **Hypernym**: the sense is a superclass
  - *vehicle* is a hypernym of *car*
  - *animal* is a hypernym of *dog*
  - *fruit* is a hypernym of *mango*
WordNet

- A hierarchically organized lexical database
- On-line thesaurus + aspects of a dictionary
  - Versions for other languages are under development

http://wordnetweb.princeton.edu/perl/webwn

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>117,097</td>
</tr>
<tr>
<td>Verb</td>
<td>11,488</td>
</tr>
<tr>
<td>Adjective</td>
<td>22,141</td>
</tr>
<tr>
<td>Adverb</td>
<td>4,601</td>
</tr>
</tbody>
</table>
WordNet “senses”

• The set of near-synonyms for a WordNet sense is called a **synset (synonym set)**

• Example: *chump* as a noun to mean

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

  *gloss*: *(a person who is gullible and easy to take advantage of)*

• Each of these senses share this same gloss
WordNet Hypernym Chains

Sense 3
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

Sense 7
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
Word Similarity

- Synonymy is binary, on/off, they are synonyms or not
- We want a looser metric: word similarity
- Two words are more similar if they share more features of meaning
Why word similarity?

- Information retrieval
- Question answering
- Machine translation
- Natural language generation
- Language modeling
- Automatic essay grading
- Document clustering
Two classes of algorithms

- **Thesaurus-based algorithms**
  - Based on whether words are “nearby” in Wordnet

- **Distributional algorithms**
  - By comparing words based on their distributional context in corpora
Thesaurus-based word similarity

- Find words that are connected in the thesaurus
  - Synonymy, hyponymy, etc.
  - Glosses and example sentences
  - Derivational relations and sentence frames

- Similarity vs Relatedness
  - Related words could be related any way
    - *car, gasoline*: related, but not similar
    - *car, bicycle*: similar
Path-based similarity

Idea: two words are similar if they’re nearby in the thesaurus hierarchy (i.e., short path between them)
Tweaks to path-based similarity

- $\text{pathlen}(c_1, c_2) =$ number of edges in the shortest path in the thesaurus graph between the sense nodes $c_1$ and $c_2$
- $\text{sim}_{\text{path}}(c_1, c_2) = - \log \text{pathlen}(c_1, c_2)$
- $\text{wordsim}(w_1, w_2) = \max_{c_1 \in \text{senses}(w_1), c_2 \in \text{senses}(w_2)} \text{sim}(c_1, c_2)$
Problems with path-based similarity

- Assumes each link represents a uniform distance
- *nickel* to *money* seems closer than *nickel* to *standard*
- Seems like we want a metric which lets us assign different “lengths” to different edges — but how?
From paths to probabilities

- Don’t measure paths. Measure probability?
- Define $P(c)$ as the probability that a randomly selected word is an instance of concept (synset) $c$
- $P(\text{ROOT}) = 1$
- 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.

- More formally:

\[
P(c) = \frac{\sum_{w \in \text{words}(c)} \text{count}(w)}{N}
\]
Concept probability examples

WordNet hierarchy augmented with probabilities $P(c)$:

```
entity  0.395
  inanimate-object  0.167
    natural-object  0.0163
      geological-formation  0.00176
        0.000113  natural-elevation  shore  0.0000836
        0.0000189  hill  coast  0.0000216
```
Information content: definitions

- Information content:
  - $\text{IC}(c) = -\log P(c)$

- Lowest common subsumer
  - $\text{LCS}(c_1, c_2) =$ the lowest common subsumer
    i.e., 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
Information content examples

WordNet hierarchy augmented with information content IC(c):

- entity 0.403
  - inanimate-object 0.777
    - natural-object 1.788
      - geological-formation 2.754
        - natural-elevation 3.947
        - shore 4.078
        - hill 4.724
        - coast 4.666
Resnik method

• The similarity between two words is related to their common information

• The more two words have in common, the more similar they are

• Resnik: measure the common information as:
  • The information content of the lowest common subsumer of the two nodes
  • \( \text{sim}_{\text{resnik}}(c_1, c_2) = - \log P(\text{LCS}(c_1, c_2)) \)
Resnik example

\[ \text{sim}_{\text{resnik}}(\text{hill}, \text{coast}) = ? \]

- entity: 0.403
  - inanimate-object: 0.777
    - geological-formation: 2.754
      - natural-elevation: 3.947
      - shore: 4.078
    - hill: 4.724
  - natural-object: 1.788
    - coast: 4.666
Let's examine how the various measures compute the similarity between gun and a selection of other words:

<table>
<thead>
<tr>
<th>w2</th>
<th>IC(w2)</th>
<th>lso</th>
<th>IC(lso)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resnik</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gun</td>
<td>10.9828</td>
<td>gun</td>
<td>10.9828</td>
</tr>
<tr>
<td>weapon</td>
<td>8.6121</td>
<td>weapon</td>
<td>8.6121</td>
</tr>
<tr>
<td>animal</td>
<td>5.8775</td>
<td>object</td>
<td>1.2161</td>
</tr>
<tr>
<td>cat</td>
<td>1.2161</td>
<td>[ROOT]</td>
<td>0.0000</td>
</tr>
<tr>
<td>water</td>
<td>11.2821</td>
<td>entity</td>
<td>0.9447</td>
</tr>
<tr>
<td>evaporation</td>
<td>13.2252</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

IC(w2): information content (negative log prob) of (the first synset for) word w2

lso: least superordinate (most specific hypernym) for "gun" and word w2.

IC(lso): information content for the lso.
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* for $1 + 4 = 5$
Recap: thesaurus-based similarity

\[ \text{sim}_{\text{path}}(c_1, c_2) = -\log \text{pathlen}(c_1, c_2) \]

\[ \text{sim}_{\text{Resnik}}(c_1, c_2) = -\log P(\text{LCS}(c_1, c_2)) \]

\[ \text{sim}_{\text{Lin}}(c_1, c_2) = \frac{2 \times \log P(\text{LCS}(c_1, c_2))}{\log P(c_1) + \log P(c_2)} \]

\[ \text{sim}_{\text{JC}}(c_1, c_2) = \frac{1}{2 \times \log P(\text{LCS}(c_1, c_2)) - (\log P(c_1) + \log P(c_2))} \]

\[ \text{sim}_{\text{eLesk}}(c_1, c_2) = \sum_{r,q\in\text{RELS}} \text{overlap}(\text{gloss}(r(c_1)), \text{gloss}(q(c_2))) \]
Problems with thesaurus-based methods

• We don’t have a thesaurus for every language

• Even if we do, many words are missing
  • Neologisms: *retweet*, *iPad*, *blog*, *unfriend*, …
  • Jargon: *poset*, *LIBOR*, *hypervisor*, …

• Typically only nouns have coverage

• **What to do??**  Distributional methods.