CS474 Natural Language Processing

- Last class
  - Intro to lexical semantics

- Today
  - Lexical semantic resources: WordNet
  - Word sense disambiguation
    » Dictionary-based approaches
    » Supervised machine learning methods

- New classroom!!! Hollister 206

CS Colloquium of Possible Interest

Synonymy

- Lexemes with the same meaning
- Invoke the notion of substitutability
  - Two lexemes will be considered synonyms if they can be substituted for one another in a sentence without changing the meaning or acceptability of the sentence
    » How big is that plane?
    » Would I be flying on a large or small plane?
    » Miss Nelson, for instance, became a kind of big sister to Mrs. Van Tassel’s son, Benjamin.
    » We frustrate ‘em and frustrate ‘em, and pretty soon they make a big mistake.
  - Also issues of register
    » Social factors that surround the use of possible synonyms, e.g. politeness, group status.

WordNet

- Handcrafted database of lexical relations
- Three separate databases: nouns; verbs; adjectives and adverbs
- Each database is a set of lexical entries (according to unique orthographic forms)
  - Set of senses associated with each entry

<table>
<thead>
<tr>
<th>Category</th>
<th>Unique Forms</th>
<th>Number of Senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noun</td>
<td>94174</td>
<td>116317</td>
</tr>
<tr>
<td>Verb</td>
<td>10319</td>
<td>22066</td>
</tr>
<tr>
<td>Adjective</td>
<td>20170</td>
<td>29881</td>
</tr>
<tr>
<td>Adverb</td>
<td>4546</td>
<td>5677</td>
</tr>
</tbody>
</table>
Sample entry

The noun “bass” has 8 senses in WordNet:
1. bass - (the lowest part of the musical range)
2. bass, bass part - (the lowest part in polyphonic music)
3. bass, basse - (an adult male singer with the lowest voice)
4. sea bass, bass - (fish of lean-fleshed saltwater fish of the family Serranidae)
5. freshwater bass, bass - (any of various North American lean-fleshed freshwater fishes especially of the genus Micropterus)
6. bass, bass voice, basso - (the lowest adult male singing voice)
7. bass - (the member with the lowest range of a family of musical instruments)
8. bass - (more technical name for any of numerous edible marine and freshwater spiny-finned fishes)

Distribution of senses

- Zipf distribution of senses

![Graph showing Zipf distribution of senses]

WordNet relations

- **Nouns**
  - Hyponym: From concepts to subtypes
  - Hypernym: From concepts to superordinates
  - Member-of: From groups to their members
  - Part-of: From wholes to parts
  - Attrib: From properties to their adjectives

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponym</td>
<td>From concepts to superordinates</td>
<td></td>
</tr>
<tr>
<td>Hypernym</td>
<td>From concepts to subtypes</td>
<td></td>
</tr>
<tr>
<td>Member-of</td>
<td>From groups to their members</td>
<td></td>
</tr>
<tr>
<td>Part-of</td>
<td>From wholes to parts</td>
<td></td>
</tr>
<tr>
<td>Attrib</td>
<td>From properties to their adjectives</td>
<td></td>
</tr>
</tbody>
</table>

- **Verbs**
  - Hyponym: From events to subevents
  - Hypernym: From events to superordinate events
  - Entails: From events to the events they entail

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyponym</td>
<td>From events to subevents</td>
<td></td>
</tr>
<tr>
<td>Hypernym</td>
<td>From events to superordinate events</td>
<td></td>
</tr>
<tr>
<td>Entails</td>
<td>From events to the events they entail</td>
<td></td>
</tr>
</tbody>
</table>

- **Adjectives/adverbs**
  - Antonym: Opposites

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antonym</td>
<td>Opposites</td>
<td>heavy ⇔ light</td>
</tr>
<tr>
<td>Adverb</td>
<td>Opposites</td>
<td>quickly ⇔ slowly</td>
</tr>
</tbody>
</table>

CS674 Natural Language Processing

- **Last class**
  - Intro to lexical semantics

- **Today**
  - Lexical semantic resources: WordNet
    - Word sense disambiguation
      - Dictionary-based approaches
      - Supervised machine learning methods
      - Issues for WSD evaluation
Word sense disambiguation

- Given a fixed set of senses associated with a lexical item, determine which of them applies to a particular instance of the lexical item
- Two fundamental approaches
  - WSD occurs during semantic analysis as a side-effect of the elimination of ill-formed semantic representations
  - Stand-alone approach
    » WSD is performed independent of, and prior to, compositional semantic analysis
    » Makes minimal assumptions about what information will be available from other NLP processes
    » Applicable in large-scale practical applications

Dictionary-based approaches

- Rely on machine readable dictionaries
- Initial implementation of this kind of approach is due to Michael Lesk (1986)
  - Given a word \( W \) to be disambiguated
    » Retrieve all of the sense definitions, \( S \), for \( W \) from the MRD
    » Compare each \( s \) in \( S \) to the dictionary definitions of all the remaining words in the context
    » Select the sense \( s \) with the most overlap with (the definitions of) these context words

Example

- Word: cone
- Context: pine cone
- Sense definitions
  - \textit{pine}: 1 kind of evergreen tree with needle-shaped leaves
  - \textit{pine}: 2 waste away through sorrow or illness
  - \textit{cone}: 1 solid body which narrows to a point
  - \textit{cone}: 2 something of this shape whether solid or hollow
  - \textit{cone}: 3 fruit of certain evergreen trees

- Accuracy of 50-70% on short samples of text from \textit{Pride and Prejudice} and an AP newswire article.

Machine learning approaches

- Machine learning methods
  - Supervised inductive learning
  - Bootstrapping
  - Unsupervised

- Emphasis is on acquiring the knowledge needed for the task from data, rather than from human analysts.
Inductive ML framework

Examples of task
(features + class)

description of context

ML Algorithm

Novel example
(features)

Classifier
(program)

correct word sense

class

Feature vector representation

- **target:** the word to be disambiguated
- **context:** portion of the surrounding text
  - Select a “window” size
  - Tagged with part-of-speech information
  - Stemming or morphological processing
  - Possibly some partial parsing
- Convert the context (and target) into a set of features
  - Attribute-value pairs
    » Numeric or nominal values

Running example

An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.

1. Fish sense
2. Musical sense
3. ...

Collocational features

- Encode information about the lexical inhabitants of specific positions located to the left or right of the target word.
  - E.g. the word, its root form, its part-of-speech
  - **An electric guitar and bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
  - [guitar, NN1, and, CJC, player, NN1, stand, VVB]
Co-occurrence features

- Encodes information about neighboring words, ignoring exact positions.
  - **Attributes**: the words themselves (or their roots)
  - **Values**: number of times the word occurs in a region surrounding the target word
  - Select a small number of frequently used content words for use as features
    - 12 most frequent content words from a collection of bass sentences drawn from the WSJ: fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band
    - Co-occurrence vector (window of size 10) for the previous example:
      \[ [0,0,0,1,0,0,0,0,0,1,0] \]

Inductive ML framework

- Examples of task (features + class)
  - ML Algorithm
  - Novel example
  - **Classifier** (program) → class
    - Learn one such classifier for each lexeme to be disambiguated

Decision list classifiers

- Decision lists: equivalent to simple case statements.
  - Classifier consists of a sequence of tests (usually on a single feature) to be applied to each input example/vector; returns a word sense.
  - Continue only until the first applicable test.
  - Default test returns the majority sense.

Decision list example

- Binary decision: fish **bass** vs. musical **bass**

<table>
<thead>
<tr>
<th>Rule</th>
<th>Sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>fish within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>striped bass</td>
<td>bass¹</td>
</tr>
<tr>
<td>gilthead within window</td>
<td>bass²</td>
</tr>
<tr>
<td>bass player</td>
<td>bass¹</td>
</tr>
<tr>
<td>plano within window</td>
<td>bass²</td>
</tr>
<tr>
<td>zenor within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>red bass</td>
<td>bass¹</td>
</tr>
<tr>
<td>plau'v bass</td>
<td>bass¹</td>
</tr>
<tr>
<td>silver within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>white within window</td>
<td>bass²</td>
</tr>
<tr>
<td>zebrew within window</td>
<td>bass¹</td>
</tr>
<tr>
<td>on bass</td>
<td>bass¹</td>
</tr>
<tr>
<td>bass are</td>
<td>bass¹</td>
</tr>
</tbody>
</table>
Learning decision lists

- Consists of *generating* and *ordering* individual tests based on the characteristics of the training data
- **Generation**: every feature-value pair constitutes a test
- **Ordering**: based on accuracy on the training set
  \[
  \text{abs} \left( \log \frac{P(\text{Sense}_i \mid f_i = v_j)}{P(\text{Sense}_2 \mid f_i = v_j)} \right)
  \]
- Associate the appropriate sense with each test