Project info

- A very useful learning algorithm:
  ≈ Allows you to incorporate arbitrary features into a sequence model with a very simple algorithm.

- A good solo (or two person?) project:
  ≈ Predict presence of “digraphs”

Semantics (Meaning)

- Lexical semantics (words)
  next, today, Thursday

- Reference (linking to the world)
  today

- Compositional semantics (sentences/trees)
  next Tuesday
What's a Word?

- Lemma (base form)
  - “animals” becomes animal
  - “brought” becomes bring
- Sometimes more than one word
  - “throws up” → throw up
  - “New York” → New York
  - “Multiword expression”, Idiom. How much of language?

Word Sense

- 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.
- While some banks furnish sperm only to married women, others are much less restrictive.
- The bank is near the corner of Forbes and Murray.

Four Meanings of “Bank”

- bank¹ = “financial institution”
- bank² = “sloping mound”
- bank³ = “biological repository”
- bank⁴ = “building where a bank¹ does its business”

- The connections between these different senses vary from practically none (homonymy) to related (polysemy).
- The relationship between the senses bank⁴ and bank¹ is called metonymy.
How Many Senses?

- This is a hard question.
- Considerations:
  - Truth conditions (serve meat / serve time)
  - Syntactic behavior (serve meat / serve as senator)
  - Zeugma test
    - Does United serve breakfast and Pittsburgh?
  - She poaches elephants and pears.

Related Phenomena

- Homophones (would/wood, two/too/to)
- Homographs (bass/bass)

Word Senses and Dictionaries
Word Senses and Dictionaries

Word Sense

- Defining word senses is a really hard problem!
- Nonetheless there is reason to believe we need to **model** this effect if we want intelligent NL understanding and generation.
- For now, assume we know the set of possible senses for each word.

Word Sense Disambiguation

- Given a word, choose its sense.
- This is a classification problem ... what techniques do you already know for classification?

  Two versions:
  - Fixed set of words to focus on (i.e., for research)
  - All-words (disambiguate everything)
Data for WSD

- As usual, corpora!
- SENSEVAL competitions
- SemCor (234K words tagged with WordNet senses)

The Standard Approach

- For a given word token, extract features:
  - words nearby in specific relative positions (collocation)
  - POS
  - lemmas
  - syntactic dependencies
  - “bag of words”

Decision List WSD

- For each feature and sense, calculate \( p(\text{sense}_i \mid \text{feature}_j) \)
- Rank features:
  \[
  \log \left( \frac{p(\text{sense}_a \mid \text{feature}_j)}{p(\text{sense}_b \mid \text{feature}_j)} \right)
  \]
- Sequence of rules: “if feature, then sense”
### Decision List for “Bass”

<table>
<thead>
<tr>
<th>Rule</th>
<th>Sense</th>
</tr>
</thead>
<tbody>
<tr>
<td>fisch within window</td>
<td>bass</td>
</tr>
<tr>
<td>carved bass</td>
<td>bass</td>
</tr>
<tr>
<td>guitar within window</td>
<td>bass</td>
</tr>
<tr>
<td>bass player</td>
<td>bass</td>
</tr>
<tr>
<td>piano within window</td>
<td>bass</td>
</tr>
<tr>
<td>tower within window</td>
<td>bass</td>
</tr>
<tr>
<td>sea bass</td>
<td>bass</td>
</tr>
<tr>
<td>playIN bass</td>
<td>bass</td>
</tr>
<tr>
<td>roar within window</td>
<td>bass</td>
</tr>
<tr>
<td>sniff within window</td>
<td>bass</td>
</tr>
<tr>
<td>scream within window</td>
<td>bass</td>
</tr>
<tr>
<td>on bars</td>
<td>bass</td>
</tr>
</tbody>
</table>

*Figure 0.3: An abbreviated decision list for finding the full sense of bass from the music sense (Turney, 1997).*

---

### Lesk Algorithm

- **Idea:** use a dictionary!
- **Pick the sense who dictionary definition overlaps the most with the word’s context.**
- **Extensions:**
  - Use dictionary definitions of context words
  - Add context words from labeled corpus sentences
  - Put weights on the words (e.g., IDF) or stoplist

---

### Selectional Restrictions and Preferences

- If context words have strong constraints on their arguments, we can exploit that. What’s “washable” or “stir-fryable”?
- In our house, everybody has a career and none of them includes *washing dishes*.
- In her tiny kitchen at home, Ms. Chen works efficiently, *stir-frying* several simple *dishes*, including braised pig’s ears and chicken livers with green peppers.
- Problems: negation, idioms, out-of-vocabulary...
**Bootstrapping**

- “Semisupervised” learning: some labeled data or initial, high-precision seeds, plus a large unlabeled corpus.
- Not just for WSD, but WSD was the first problem for which this idea really hit a home run.

**One Sense Per Discourse**

- This is a fancy way of saying that, within a discourse (e.g., document), ambiguous tokens of the same type tend to be correlated.
- WSD decisions about tokens of the same type should not be made independently!

<table>
<thead>
<tr>
<th>Word</th>
<th>Senses</th>
<th>Accuracy</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>plant</td>
<td>living/factory</td>
<td>99.8 %</td>
<td>72.8 %</td>
</tr>
<tr>
<td>tank</td>
<td>vehicle/contnr</td>
<td>99.6 %</td>
<td>50.5 %</td>
</tr>
<tr>
<td>poach</td>
<td>steal/boil</td>
<td>100.0 %</td>
<td>44.4 %</td>
</tr>
<tr>
<td>palm</td>
<td>tree/hand</td>
<td>99.8 %</td>
<td>38.5 %</td>
</tr>
<tr>
<td>axes</td>
<td>grid/tools</td>
<td>100.0 %</td>
<td>35.5 %</td>
</tr>
<tr>
<td>sake</td>
<td>benefit/drink</td>
<td>100.0 %</td>
<td>33.7 %</td>
</tr>
<tr>
<td>bass</td>
<td>fish/music</td>
<td>100.0 %</td>
<td>38.8 %</td>
</tr>
<tr>
<td>space</td>
<td>volume/outer</td>
<td>99.2 %</td>
<td>67.7 %</td>
</tr>
<tr>
<td>motion</td>
<td>legal/physical</td>
<td>99.9 %</td>
<td>49.8 %</td>
</tr>
<tr>
<td>crane</td>
<td>bird/machine</td>
<td>100.0 %</td>
<td>49.1 %</td>
</tr>
</tbody>
</table>

**Average**

- 99.8 %
- 50.1 %
One Sense Per Collocation

- Certain features of the context are very strong predictors for one sense or another.
  - ... power plant ...
  - ... palm of ...
  - ... the park ...
- This is a fancy way of saying that (some) collocations are excellent features.

The Yarowsky Algorithm

Given: ambiguous word type \( w \), lots of text
1. Choose a few seed collocations for each sense and label data in those collocations.
The Yarowsky Algorithm

Given: ambiguous word type \( w \), lots of text
1. Choose a few seed collocations for each sense and label data in those collocations.
2. Train a supervised classifier on the labeled examples. (Yarowsky used a decision list.)
3. Label all examples. Keep the labels about which the supervised classifier was highly confident (above threshold).
   - Optionally, exploit one-sense-per-discourse to “spread” a label throughout the discourse.
4. Go to 2.
Whence Seeds?

- Yarowsky suggests:
  - dictionary definitions
  - single defining collocate (e.g., from WordNet)
  - label extremely common collocations

Evaluating WSD

- Classification accuracy
- Straw-man baseline: “most frequent sense”
- Strong-man baseline: Lesk algorithm
- Ceiling: human inter-annotator agreement
- Pseudoword evaluation

Experimental Results
What's Wrong?

- Assumptions about sense identification, availability of corpora
- Is intrinsic evaluation meaningful?
- What if word sense disambiguation isn't required for IR, translation, QA, etc.?

Coreference Resolution

The Problem

Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based financial-services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.
Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based financial-services company's president. It has been ten years since she came to Megabucks from rival Lotsabucks.
The Problem

Victoria Chen, Chief Financial Officer of Megabucks Banking Corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37-year-old also became the Denver-based financial-services company’s president. It has been ten years since she came to Megabucks from rival Lotsabucks.

Different Kinds of NPs

- Indefinite NPs: a smart professor, some cheesecake
- Definite NPs: the New York Times, the friend I was telling you about
- Pronouns: she
- Demonstratives: that one, this, those students
- Names: IBM, Barack Obama, UW-Madison

General Idea

- Parse.
- For every pair of NPs, carry out binary classification.
- Need a corpus of training examples (positive and negative)
- Need features.
- Naive Bayes is one option; there are others.
- Ensure consistency by enforcing transitive closure heuristically or with graph cut algorithms.
Coreference Features

- Edit distance between the two NPs
- Are the two NPs of the same NER type, and if so, are they plausibly "aliases" of each other (same date, or same names modulo titles, or same acronym for organizations)?
- Appositive syntax: \( \ldots \text{NP}_1, \text{NP}_2, \ldots \)
- Proper/definite/indefinite/pronoun
- Gender
- Number
- Distance in sentences
- Number of NPs between
- Grammatical role (subject, object, etc.)

Final Thought: Gottlob Frege

- "Sense" and "reference" - two elements of meaning
- Reference is the thing a name refers to
- Sense is the cognitive significance, or mode of presentation.
  - Samuel Clemens and Mark Twain
  - the greatest integer