Announcements

- Quiz #2
  - Ends tonight
- Assignment 7 / Final Project
  - Submit your choice in email today!
  - Questions?

Objectives

- Review the idea behind joint, generative models, as well as their benefits and limitations.
- Discuss the challenge of word sense disambiguation and approach it as a classification problem.
- Explore knowledge sources (i.e., “features”) that might help us do a better job of disambiguating word senses.
- Motivate the need for conditional models, trained discriminatively.
- Prepare to see maximum entropy models.

Joint / Generative Models

- E.g., Naive-Bayes, class-conditional LMs, HMMs, GMMs, LDA
- These are joint models == Generative models:
  - Break complex structure down into derivation steps
  - “Factors” or “local model”
  - Often in our models, each step is a categorical, multinomial, or Dirichlet draw, conditioned on specified context
  - How to estimate parameters from labeled data?
  - Collect counts and smooth
  - How to estimate parameters from unlabeled data?
  - EM, MCMC –

Today & next time

- Conditional models
- Trained discriminatively
- Motivated by the problem of word sense disambiguation

Word Senses

- [http://www.youtube.com/watch?v=algmKOzTSGE](http://www.youtube.com/watch?v=algmKOzTSGE)
### Word Senses
- Multiple distinct meanings of words are called senses:
  - *plant*: living plant, manufacturing plant, ...
  - *title*: name of a work, ownership document, form of address, material at the start of a film, ...
- Many levels of sense distinctions
  - Homonymy: totally unrelated meanings (river bank, money bank)
  - Polysemy: related meanings (star in sky, star on tv)
  - Systematic polysemy: productive meaning extensions (organizations to their buildings) or metaphor
- Sense distinctions can be extremely subtle (or not)
- Granularity of senses needed depends a lot on the task
- Why is it important to model word senses?
  - Translation, parsing, information retrieval?

### Word Sense Disambiguation
- Example: *living plant* vs. *manufacturing plant*
  - The *plant* which had previously sustained the town's economy shut down after an extended labor strike.
- How do we tell these senses apart?
  - “context”
  - Is this just text classification using words from nearby text, where each word sense represents a class?
  - Possible solution: run our naïve-bayes classifier?
- Naïve Bayes classification works OK for noun senses
  - 90% on classic, shockingly easy examples (line, interest, star)
  - 80% on senseval-1 nouns
  - 70% on senseval-1 verbs – harder!

### Others’ Approaches to WSD
- Supervised learning
  - Most WSD systems do some kind of supervised learning
  - Many competing classification techniques perform about the same
  - It’s all about the knowledge sources – “features” – you rely on
  - Problem: limited labeled training data available
- Unsupervised learning
  - Bootstrapping (Yarowsky 95): “one sense per document”
  - Clustering
- Indirect supervision
  - From thesauri
  - From WordNet
  - From parallel corpora

### Resources: WordNet
- Hand-built (but large) hierarchy of word senses
  - Basically a hierarchical thesaurus

### Resources
- SemEval & SemEval
  - An ongoing WSD competition
  - Training/test sets for a wide range of words, difficulties, and parts-of-speech
  - Some problems where lots of lab try many competing approaches
- FrameNet
  - A large database of English that is both human- and machine-readable
  - Annotated examples of how words are used in actual texts
  - Student: dictionary of more than 10,000 word-senses, most of them with annotated examples
  - Show the meaning and usage
- SRL: more than 170,000 manually annotated sentences provide a unique training dataset for semantic role labeling
- Lexical: a selection of dictionaries, with uniquely detailed evidence for the combinatorial properties of a core set of the English vocabulary
- Sentential frame: a description of a type of event, relation, or entity and the participants in it.
- Other Resources
  - “Open-Mind Common Sense”: open source fact base
  - Open-domain Information extraction from the web
  - Parallel corpora
  - Flat thesauri (e.g., Roger’s)

### Back to Verb WSD
- Why are verbs harder?
  - Verbal senses are less topical
  - More sensitive to structure, argument choice
- Verb Example: “Serve”
  - The tree stump **served** as a table.
  - The scandal **served** to increase his popularity.
  - We **serve** meals for the homeless.
  - He **served** his country.
  - He **served** six years for embezzlement.
  - It was Agassi’s turn to **serve**.
  - He was **served** by the sheriff.
Verb Example: “Serve”

(function) The tree stump serves as a table.
[means] The scandal served to increase his popularity.
[dish] We serve meals for the homeless.
[military] He served his country.
[jail] He served six years for embezzlement.

It was Agassi’s turn to serve.
He was served by the sheriff.

(function) The tree stump serves as a table.
[means] The scandal served to increase his popularity.
[dish] We serve meals for the homeless.
[military] He served his country.
[jail] He served six years for embezzlement.

It was Agassi’s turn to serve.
He was served by the sheriff.

(function) The tree stump serves as a table.
[means] The scandal served to increase his popularity.
[dish] We serve meals for the homeless.
[military] He served his country.
[jail] He served six years for embezzlement.

It was Agassi’s turn to serve.
He was served by the sheriff.

(function) The tree stump serves as a table.
[means] The scandal served to increase his popularity.
[dish] We serve meals for the homeless.
[military] He served his country.
[jail] He served six years for embezzlement.

It was Agassi’s turn to serve.
He was served by the sheriff.
Back to Verb WSD

- Why are verbs harder?
  - Verbal senses are less topical
  - More sensitive to structure, argument choice

Verb Example: “Serve”

- [function] The tree stump serves as a table.
- [means] The scandal served to increase his popularity.
- [dish] We serve meals for the homeless.
- [military] He served his country.
- [jail] He served six years for embezzlement.
- [tennis] It was Agassi’s turn to serve.
- [legal] He was served by the sheriff.

Next

- Designing some features to help disambiguate / classify.
- Building a classifier that accommodates all of our features.