Search with Partial Information: Data-Driven Word Sense Disambiguation

Andrei Alexandrescu
andre@cs.washington.edu

Katrin Kirchhoff
katrin@ee.washington.edu
What is Word Sense Disambiguation?

- “He was moving as fast as he could, blade in one hand, mouse in the other.”
What is Word Sense Disambiguation?

- “He was moving as fast as he could, blade in one hand, mouse in the other.”

- “The place was dark, somber, and crawling with hungry cats.”
What is Word Sense Disambiguation?

- “He was moving as fast as he could, blade in one hand, mouse in the other.”

  - “The place was dark, somber, and crawling with hungry cats.”

  - “He was repairing a fileserver blade with the help of a graphical program.”
“He was moving as fast as he could, blade in one hand, mouse in the other.”

- “The place was dark, somber, and crawling with hungry cats.”

- “He was repairing a fileserver blade with the help of a graphical program.”

Goal: we’d like to assign a sense tag to words in a sentence
  - A classic labeling problem
Terminology

- **Features**: Any visible input that can help a decision – e.g. “cats”, “filesaver” etc.

- **Soft prediction**: “Sample has label X with probability 0.6, Y with probability 0.3, Z with 0.1”

- **Supervised learning**: Train on known data, then test on unseen data

- **Semi-supervised learning**: Test features are visible during training on known data
Label Propagation (Zhu, 2005)

- Construct a graph from all features (train + test)
  - Each sample is a vertex in the graph
  - Edges reflect similarity among samples
- Assign labels to known data
- Propagate labels using a random walk

- Problems:
  - graph construction process is very domain-specific
  - critically depends on the chosen similarity measure
  - NLP features are often discrete or mixed
Our idea: Data-driven graph construction

Features
“place”
“dark”
“somber”
“crawl”
“hungry”
“cats”

Supervised learner

“mouse”
Soft labels

animal: 0.3
device: 0.14
timid person: 0.31
black eye: 0.25

½ supervised learner with graphs

Hard labels

✓ They meant “mouse” as “animal”
Characteristics

- **Particularities**
  - Special training for supervised classifier

- **Advantages**
  - Uniform range and type of features
  - Facile feature postprocessing
  - Optimized class separation

- **Risks**
  - Overspecialization of first-pass classifier
    - Confident but wrong predictions
Results

Accuracy (%)

Train data used

- SVM
- LP
- SVM+LP
Conclusions

- Better graphs
  - Closer to optimal features by using a separate learner
  - Non-uniform → uniform features

- Better performance
  - Significantly better than label propagation using the initial features

- Simplified graph construction step
  - Domain-specific knowledge not needed