Search with Partial Information: Data-Driven Word Sense Disambiguation

Andrei Alexandrescu
andrei@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."

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.”
- 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”, “fileserver” 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

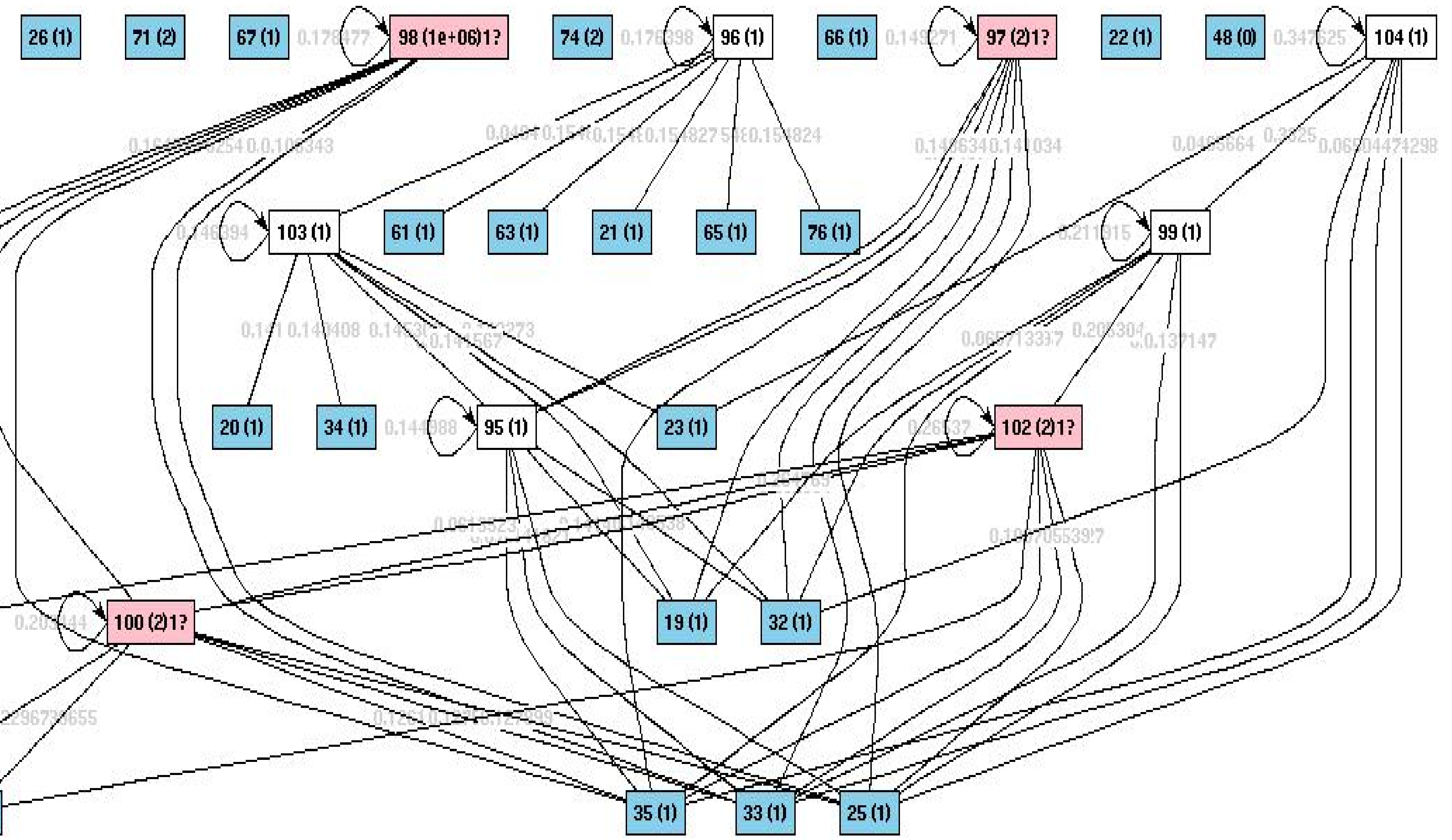
- ***Semi-supervised learning:*** Test features are visible during training on known data

Label Propagation (Zhu, 2005)

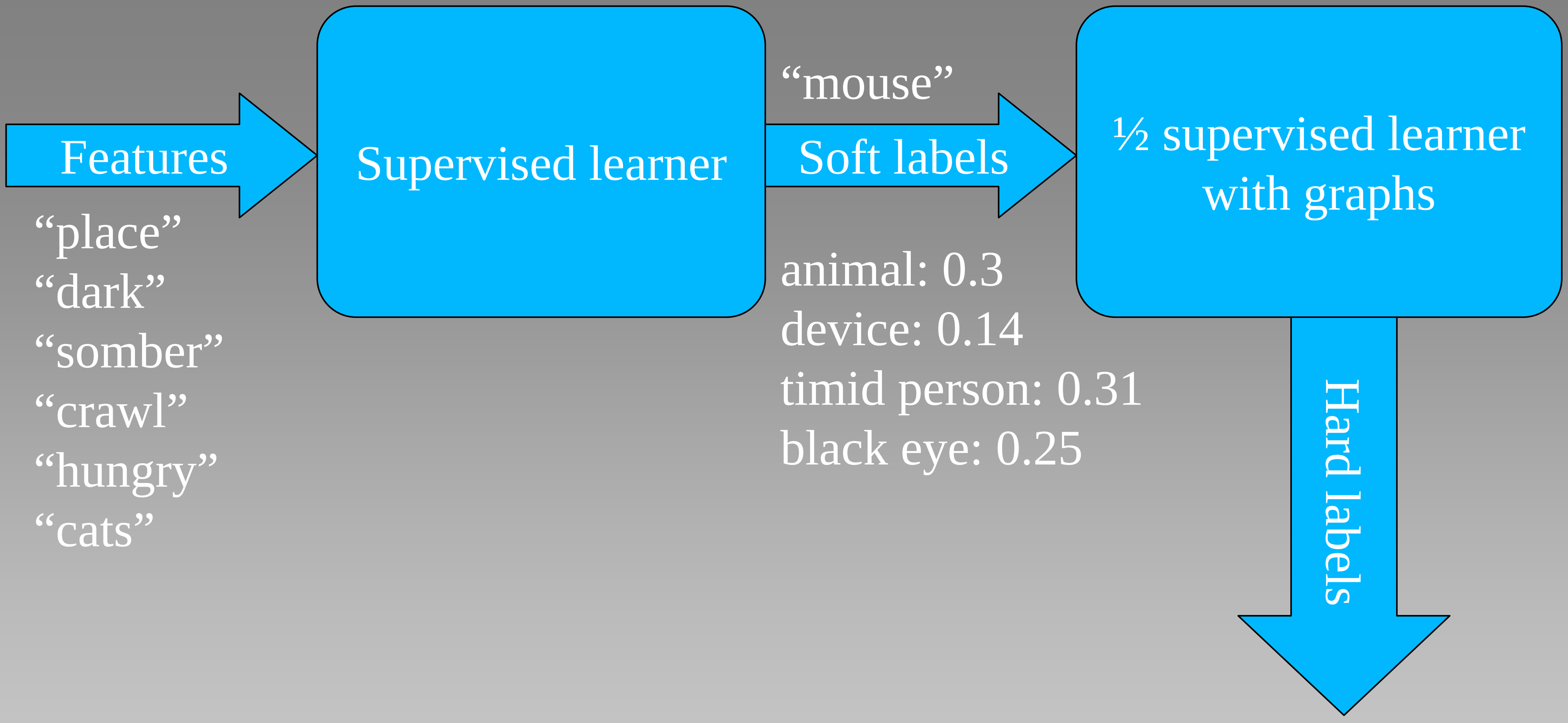
- Construct a graph from all features (train + test)
  - Each sample is a vertex in the graph
  - Edges reflect similarity among samples
- Assign labels to known data
- Propagate labels using a random walk

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

Example


26 (1)
71 (2)
67 (1)
98 (1e+06)1?
74 (2)
96 (1)
66 (1)
97 (2)1?
22 (1)
48 (0)
104 (1)
103 (1)
61 (1)
63 (1)
21 (1)
65 (1)
76 (1)
99 (1)
20 (1)
34 (1)
95 (1)
23 (1)
102 (2)1?
100 (2)1?
19 (1)
32 (1)
35 (1)
33 (1)
25 (1)

Our idea: Data-driven graph construction

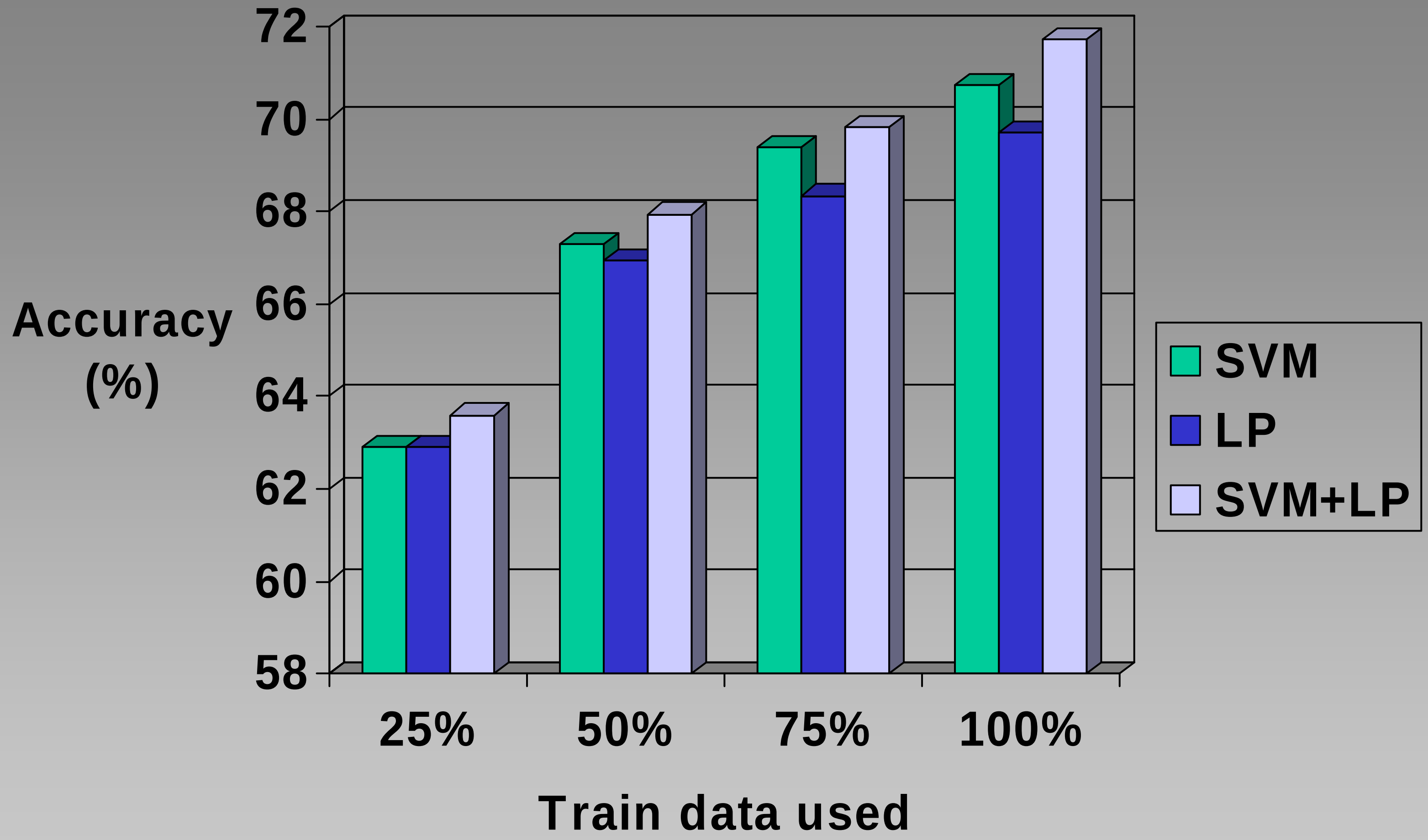


✓ **They meant "mouse" as "animal"**

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



72
70
68
66
64
62
60
58
Accuracy (%)
25%
50%
75%
100%
Train data used
SVM
LP
SVM+LP

Conclusions

- Better graphs
  - Closer to optimal features by using a separate learner
  - Non-uniform $\rightarrow$ uniform features
- Better performance
  - Significantly better than label propagation using the initial features
- Simplified graph construction step
  - Domain-specific knowledge not needed