Rich Prior Knowledge in Learning for Reducing Annotation Cost

Speakers:
- Joao Graca (L2F, Inesc-ID),
- Gregory Druck (Yahoo! Research),
- Kuzman Ganchev (Google Inc.)

Time:
- afternoon session of Monday, 21 May 2012.

Motivation:

We possess a wealth of prior knowledge about most prediction problems, and particularly so for many of the fundamental tasks in natural language processing. Clearly we should be able to use such information to reduce annotation cost. Though we could use prior knowledge to build rule-based systems, machine learning allows generalization. Unfortunately, it is often difficult to make use of prior knowledge during learning, as it typically does not come in the form of labeled examples, may be difficult to encode as a prior on parameters in a Bayesian setting, and may be impossible to incorporate into a tractable model. Instead, we usually have prior knowledge about the values of output variables. For example, linguistic knowledge or an out-of-domain parser may provide the locations of likely syntactic dependencies for grammar induction. Motivated by the prospect of being able to naturally leverage such knowledge, four different groups have recently developed similar, general frameworks for expressing and learning with side information about output variables. The tutorial surveys this work.

Outline:
Introduction (30 minutes):
- Introduction to different types of prior knowledge about NLP problems
- Limitations of previous methods for incorporating prior knowledge, including Bayesian and heuristic approaches
- Motivation for constraining the output variables directly

Interactive Examples and Demos (15 minutes)

Recent Frameworks for Learning with Prior Knowledge (15 minutes):
- Brief theoretical overview of and discussion of connections between:
  - Learning from Measurements (University of California, Berkeley)
  - Generalized Expectation (University of Massachusetts, Amherst)
  - Posterior Regularization (University of Pennsylvania)
  - Constraint Driven Learning (University of Illinois, Urbana-Champaign)

Coffee Break (15 minutes)

Applications (90 minutes):
- Unstructured problems:

- Document Classification: labeled features, multi-view learning
- Sequence problems:
  - Information Extraction: labeled features, multi-view learning, long-range dependencies
  - Word Alignment: bijectivity, symmetry
  - POS Tagging: posterior sparsity
- Tree problems:
  - Dependency Parsing: linguistic knowledge, noisy labels, posterior sparsity
- Active / interactive training: active feature labeling for sequence problems, interactive training

Coffee Break (15 minutes)

Implementation (45 minutes):
- Guidance on implementation
- Description and walk-through of existing software packages