LSA 2015: Intro to Computational Linguistics

Class schedule (tentative)

Sharon Goldwater
Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh

Week 1

What is computational linguistics?

- Computational linguistics vs natural language processing

- Core methods and core questions

Probability and estimation

- Generative models and inference

- Estimating model parameters

- Examples

Language models

- What are language models and what are they used for?

    - NLP applications
    - linguistic and psycholinguistic models

- *n*-gram language models

- Entropy and perplexity

Lab: probability estimation

Week 2

Smoothing

- add-one smoothing

- add-λ smoothing

- optimizing λ and the train/dev/test split

- problems with add-λ smoothing

- other smoothing methods

Hidden Markov models and part-of-speech tagging

- The tagging task

- Definition of an HMM

- Training an HMM

- Finding the best tag sequence

Algorithms for HMMs

- Viterbi algorithm
- forward-backward algorithm and EM (high-level view)

Lab: tagging

Week 3

Parsing

- Parsers and recognizers
- Uses in NLP and psycholinguistics
- Recursive descent parsing
- Shift-reduce parsing

Chart parsing

- Chart data structure
- Chart parsing algorithm

Broad-coverage parsing

- treebanks and treebank grammars
- probabilistic context-free grammars
- ambiguity

Lab: recursive descent parsing

Week 4

Probabilistic parsing

- best-first parsing and evaluation
- problems with PCFGs
- types of solutions (high-level view)

Distributional semantics

- Context vectors
- mutual information
- alternative association measures
- vector similarity
- uses in NLP and psycholinguistics

Maximum entropy (logistic regression) models

- classification tasks
- discriminative models and logistic regression
- example: parsing
- example: MaxEnt phonology

Lab: sentiment analysis