FNLP Lecture 23b
Wrapping Up

Nathan Schneider

7 December 2016

In a nutshell

- We have seen **representations**, **datasets**, **models**, and **algorithms** for computationally reasoning about textual language ***in a data-driven fashion***.
  - Persistent challenges: Zipf's Law, ambiguity & flexibility, variation, context
- **Core NLP tasks** (judgments about the language itself): tokenization, POS tagging, syntactic parsing (constituency, dependency), word sense disambiguation, word similarity, semantic role labeling, coreference resolution
- **NLP applications** (solve some practical problem involving/using language): spam classification, language/author identification, sentiment analysis, spelling correction, named entity recognition, question answering, machine translation
- Which of these are generally easy, and which are hard?

Language complexity and diversity

- **Ambiguity** and **flexibility** of expression often best addressed with corpora & statistics
  - Treebanks and statistical parsing
- Grammatical forms help convey meaning, but the relationship is complicated, motivating **semantic** representations
  - proposed by linguists, or
  - induced from data
- Typological variation: Languages vary extensively in **phonology**, **morphology**, and **syntax**

Methods useful for more than one task

- annotation, crowdsourcing
- rule-based algorithms, e.g. regular expressions
- classification (naïve Bayes, perceptron, SVM, MaxEnt)
- n-gram language modeling
- grammars & parsing
- sequence modeling (HMMs, structured perceptron)
- structured prediction—decoding as search:
  greedy vs. exact; dynamic programming (Viterbi, CKY)

Models & Learning

- Because language is so complex, most NLP tasks benefit from statistical learning.

- In this course, mostly **supervised learning** with *labeled* data. Exceptions:

  - **unsupervised learning:** the EM algorithm (e.g. for word alignment, topic models)

  - n-gram models: supervised learning, but no extra labels necessary.

- In NLP research, a tension between building a lot of linguistic insights into models vs. learning almost purely from the data.

  - Current research on neural networks tries to bypass hand-designed features/ intermediate representations as much as possible.

  - We still don't quite know how to capture "deep" understanding.

Generative and discriminative models

- Assign probability to language AND hidden variable? Or just score hidden variable GIVEN language?

- Independence assumptions: how useful/harmful are they?
  - "**all models are wrong**, but **some are useful**"
  - bag-of-words; Markov models
  - combining statistics from different sources, e.g. Noisy Channel Model

- Avoiding overfitting (smoothing, regularization)

- Evaluation: gold standard? sometimes difficult

Dynamic Programming Algorithms

- Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

Dynamic Programming Algorithms

- Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

- In a sentence of length *N*, what is the asymptotic runtime complexity of:

  - **Viterbi** (in a first-order HMM), with *L* possible labels?

Dynamic Programming Algorithms

- Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

- In a sentence of length *N*, what is the asymptotic runtime complexity of:

  - **Viterbi** (in a first-order HMM), with *L* possible labels? $O(N^2L)$

  - **CKY**, with a grammar of size *G*?

Dynamic Programming Algorithms

- Allow us to search a combinatorial (exponential) space efficiently by reusing partial results.

- In a sentence of length *N*, what is the asymptotic runtime complexity of:

  - **Viterbi** (in a first-order HMM), with *L* possible labels? $O(N^2L)$

  - **CKY**, with a grammar of size *G*? $O(N^3G)$

Applications

- Question answering, information retrieval, machine translation

- Your projects!

- Now that you know the tools in the toolbox, you can build all kinds of cool things!

The Final Exam

- Tuesday, 4:00-6:00
- Largely similar in style to the midterm & quizzes, but with content covering the entire course.
- …and more short answer questions. For each major concept or technique, be prepared to define it, explain its relevance to NLP, discuss its strengths and weaknesses, and compare to alternatives.
  - E.g.: “Why is smoothing used? For a model covered in class, describe two methods for smoothing and their pros/cons.”
- Study guide will be posted.

Other Administrivia

- Grading is ongoing
- Peer evaluations for the final project
- Course evaluation https://eval.georgetown.edu/
- James will hold usual office hours on Friday.
- Office hour tomorrow?