Lecture 24
Wrapping Up

Nathan Schneider

ENLP | 25 April 2024

In a nutshell

- We have seen **representations**, **datasets**, **models**, and **algorithms** for computationally reasoning about textual language.
  - 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, 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/finite-state methods, e.g. regular expressions
- classification (naïve Bayes, perceptron)
- language modeling (n-gram or neural)
- grammars & parsing
- sequence modeling (HMMs, structured perceptron, LSTM, transformers, seq2seq)
- structured prediction—dynamic programming (Viterbi, CKY)

Models & Learning

- Because language is so complex, most NLP tasks rely on learning from data.
  Multiple paradigms:
  - **supervised learning** with labeled data (classification, tagging, parsing)
  - **self-supervised learning:** e.g. neural embeddings/LMs, where unlabeled text provides the training signal (next word prediction or masked word prediction) & **transfer learning**, applying LMs to induce representations for downstream tasks
  - **unsupervised learning:** inducing explicit clusters or structures without labeled training data (e.g. topic models, word alignment in SMT; see EM algorithm)

Models & Learning

- In designing models/systems, tradeoff of **domain expertise** (e.g. linguistic structure) vs. **scale**.
    - As of 2024, with massive training data and compute, systems like GPT-4 can produce highly fluent and linguistically coherent text. But not always factually coherent/correct, and subject to bias.
    - 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?
    $O(NL^2)$

  - **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? $O(NL^2)$

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

Applications

- Sentiment analysis, machine translation
- Your projects!
- Now that you know the tools in the toolbox, you can

Up Next

- **Poster Session on Tuesday**
  - Make a poster concisely summarizing the key aspects of your project—the task, methods, results
  - Include example inputs/outputs
  - We'll send detailed instructions
- **Project Report** due 5/8
  - Instructions on Canvas assignment
  - Put code on GitHub (public or shared with instructor/TAs)
- **Project Team Peer Evaluations**
- **Midterm 2** opens 5/3 at noon (Canvas, open book, 90 min., 24 hour window)

Other Administrivia

- TA & course evaluations
  https://eval.georgetown.edu/ ★★★★★

- Visit https://gucl.georgetown.edu/ for future talks & courses!

  5/9 | Bigger is not always better: The importance of human-scale language modeling for psycholinguistics (partly about BabyLM challenge)

  5/15 | A Sanity Check on Emergent Properties in LLMs