Prediction of Structured Objects in NLP

Shay Cohen

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

November 6, 2014

Computer scientist view of NLP

Structured data (trees, graphs, etc.)

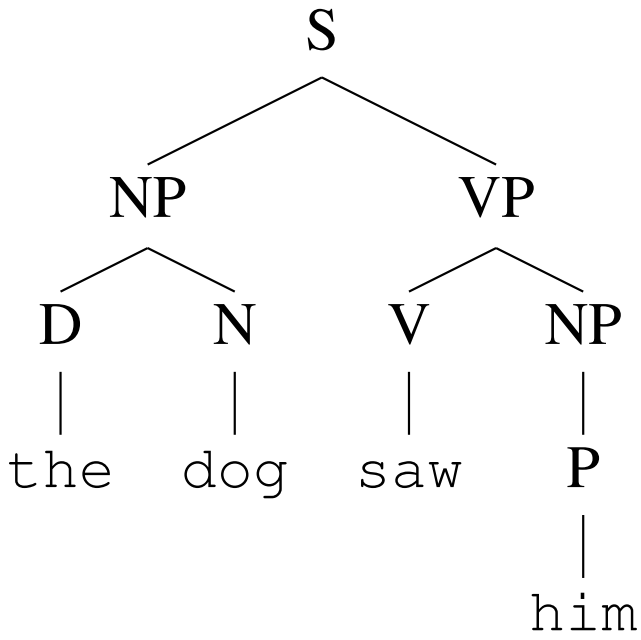
Combinatorial optimization problems (dynamic programming, etc.)

Numerical optimization algorithms (for learning)

Bonus: the language domain is fascinating

Example problem: syntactic parsing

Say we have a tree. The tree has its nodes labeled. For example:



For a given sentence, we can have many trees

Some are right, some are wrong. This is a problem of ambiguity

Example problem: syntactic parsing

We map trees to probabilities. Higher score means “more correct” tree

Questions we need to ask:

- How do we score trees with probabilities in a compositional way?
- How do we choose the tree with the maximal score given a sentence?
- Back to starting point: are such trees the right representation anyway?

Example problem: syntactic parsing

- How do we score trees with probabilities in a compositional way?
  Give a probability to each node. The probability of a whole tree is the product of the probabilities of all nodes.
- How do we choose the tree with the maximal score given a sentence?
  Using dynamic programming.
- Back to starting point: are such trees the right representation anyway?
  Depends on the application or the problem we are trying to solve.

Other questions

We assumed we *know* the probabilities of each node to score the trees

How do we *learn* these weights? – training

- With examples of the trees given? (supervised learning)
- Just from strings (yields of trees)? (unsupervised learning)
- With incomplete data? (latent-variable learning)

Summary

I gave a CS oriented view of NLP

NLP is of course also tied to linguistics

Many of the representations and the models we explore are based on insights from linguistics

Next semester: Topics in NLP (INFR11113)
`http://www.inf.ed.ac.uk/teaching/courses/tnlp/`

Demo: A master's project done by Chiraag Lala:
`http://kinloch.inf.ed.ac.uk/words/`