Topics in Natural Language Processing

Shay Cohen

Institute for Language, Cognition and Computation

University of Edinburgh

Lecture 6

Administrativia

- Some people have sent in their presentation preferences
- Deadline to send papers/topics: Monday 2/2 5pm. If you have questions, please resolve them with me beforehand.

Last class

Context-free grammars

$S \to NP \ VP$

N - non terminal set

V - terminal set

R - rules $A \to \alpha$ $A \in N$ $\alpha \in (N \cup V)^*$

$S \in N$ start symbol

Probabilistic context-free grammars

$$p(A \to \alpha \mid A) \geq 0 \qquad \sum_{A \to \alpha} p(A \to \alpha \mid A) = 1$$

CNF : $A \to B \ C \mid a$

Chomsky normal form

every CFG can be converted to CNF (subtle issue : ε-rules)

How to estimate a PCFG?

Treebank $\rightarrow$ A set of rules and a probability for each rule

$$\hat{p}(A \rightarrow \alpha) = \frac{\text{count}(A \rightarrow \alpha, TB)}{\text{count}(A, TB)}$$

(To extract the grammar, we consider parent and immediate children)

Today's class

Inference in natural language processing

- What is inference?
- The CKY algorithm
- The inside algorithm
- Weighted logic programs and semirings
- Hypergraph algorithms

Estimation

We learned how to do estimation:

- Maximum likelihood estimate $\hat{\theta} \arg\max_{\theta} L(\theta, w_1 \ldots w_n)$
- Bayesian posterior summarisation $\hat{\theta} \arg\max_{\theta} p(\theta | w_1 \ldots w_n)$
- ... There are many other ways

What's next?

Inference

What's inference?

Given a statsical model, find probable structure, classification, etc. for the input

Inference

Our Ω was usually a cross-product of inputs and outputs

$$\Omega = \{ (\text{input}, \text{output}) \}$$

Now, given an input, we need to find the correct output

$$\text{output}^* = \underset{\text{output}}{\text{argmax}}\ p(\text{output} \mid \text{input}) =$$

$$= \underset{\text{output}}{\text{argmax}}\ \frac{p(\text{output}, \text{input})}{p(\text{input})} = \underset{\text{output}}{\text{argmax}}\ p(\text{input}, \text{output})$$

Inference

Our Ω was usually a cross-product of inputs and outputs

Now, given an input, we need to find the correct output

$$\arg\max_{\text{output}} p(\text{output}|\text{input})$$

Linear Score Function

Consider a model which is a PCFG. $p(r_i)$

Probability of a tree: $t = (r_1 \ldots r_n)$

$$p(t) = \prod_{i=1}^{n} p(r_i) = \prod_{r \in t} p(r)^{freq(r,t)} =$$

$$= \prod_{r \in R} p(r)^{freq(r,t)}$$

"Best" tree y given sentence x:

$$t^*(x) = \arg\max_{\substack{t \\ yield(t) = x}} p(t)$$

sentence

Linear Score Function

“Best” tree given sentence x:

$$y^* = \arg \max_{y:\ \text{yield}(y)=x} \underbrace{\prod_{r \in y} p(r)^{\text{freq}(y,r)}}_{p(t)} =$$

$$\arg\max_{y,\ \text{yield}(y)=x} \sum \big(\underbrace{\log p(r)}_{w(r)}\big) \times \text{freq}(y,r) =$$

$$= \arg\max_{y,\ \text{yield}(y)=x} \sum_{r \in R} w(r) \times \text{freq}(y,r)$$

parameters

"refers" to the structure

$$\arg\max_{y} \theta^T f(y,x)$$

w

The CKY Algorithm

$$y^* = \arg \max_{y:\ \text{yield}(y)=x} \sum_{r \in y} w(r) \times \text{freq}(y, r)$$

$\alpha(A, i, j)$ - the maximum score for a tree that has head A, and spans $x_i \ldots x_j$

$$\alpha(A, i, j) =$$

$$= \max_{k=i}^{j-1} \max_{A \to B\ C} \alpha(B, i, k) + \alpha(C, k+1, j) + w(A \to B\ C)$$

Multiplicative version of the CKY algorithm

$$y^* = \arg \max_{y:\ \mathrm{yield}(y)=x} \prod_{r \in y} p(r)^{\mathrm{freq}(y,r)}$$

$\alpha(A, i, j)$

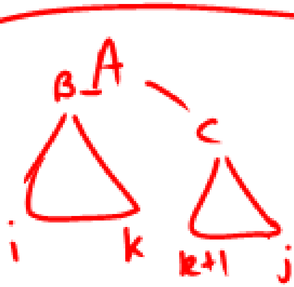
$$\alpha(A, i, j) = \max_{i \le k < j} \max_{A \to B C} \underline{\alpha(B, i, k)} \times \underline{\alpha(C, k+1, j)} \times \boxed{\underline{p(A \to B\ C \mid A)}}$$

The Inside Algorithm

$$p(x) = \sum_{y:\ \text{yield}(y)=x} \prod_{r \in y} p(r)^{\text{freq}(y,r)}$$

$\alpha(A,i,j)$ ~~ws~~ the sum over all derivations spanning $x_i \ldots x_j$ headed by A.

$$\alpha(A,i,j) = \sum_{k=i}^{j-1} \sum_{A \to B\,C} \alpha(B,i,k) \times \alpha(C,k+1,j) \times p(A \to B\,C \mid A)$$



$k+1 = j \Rightarrow k = j-1$

$$p(\text{tree}, x) = \prod_{\text{rule}} p(\text{rule})$$

$$p(x) = \sum_{\text{tree}} p(\text{tree}, x)$$ marginalizing the tree

Inside and CKY

What is the connection between the inside algorithm and CKY?

CKY:

Version 1:

$$\alpha(A,i,j) = \max_{i \leq k \leq j-1} \max_{A \rightarrow B\,C} p(A \rightarrow B\,C|A)\alpha(B,i,k)\alpha(C,k+1,j)$$

Version 2:

$$\alpha(A,i,j) = \max_{i \leq k \leq j-1} \max_{A \rightarrow B\,C} w(A \rightarrow B\,C) + \alpha(B,i,k) + \alpha(C,k+1,j)$$

Inside:

$$\alpha(A,i,j) = \sum_{k=i}^{j-1} \sum_{A \rightarrow B\,C} p(A \rightarrow B\,C|A)\alpha(B,i,k)\alpha(C,k+1,j)$$

Semirings

What is a semiring?

An algebraic structure over $\mathbb{R}$

$\otimes$ $\qquad$ $a \oplus b$

$\oplus$ $\qquad$ $a \otimes b$

Semirings

What is a semiring?

- A set R $\quad \mathbb{R} \quad [0,1]$
- Two operations: $\oplus$ and $\otimes$
- Identity element $\bar{1}$ for $\otimes$ $\quad \bar{1} \otimes a = a$
- Identity element $\bar{0}$ for $\oplus$ $\quad \bar{0} \oplus a = a$
- (... and a few more important properties)

CKY and Semirings

CKY:

$$\alpha(A, i, j) = \max_{i \leq k \leq j-1} \max_{A \to B\, C} p(A \to B\, C | A) \alpha(B, i, k) \alpha(C, k+1, j)$$

What is the semiring?

$\oplus$ $a \oplus b = \max\{a, b\}$

$\otimes$ $\times$

$\bar{1}$ 1

$\bar{0}$ 0

CKY and Semirings

CKY:

$$\alpha(A,i,j) = \max_{i \le k \le j-1} \max_{A \to B\,C} w(A \to B\,C) + \alpha(B,i,k) + \alpha(C,k+1,j)$$

What is the semiring?

$\oplus$ $\quad a \oplus b = \max\{a, b\}$

$\otimes$ $\quad a \otimes b = a + b$

$\overline{1}$

$\overline{0}$

Inside and Semirings

Inside:

$$\alpha(A,i,j) = \sum_{k=i}^{j-1} \sum_{A \to B\,C} p(A \to B\,C|A)\alpha(B,i,k)\alpha(C,k+1,j)$$

What is the semiring?

$\oplus$ $a \oplus b = a + b$

$\otimes$ $a \otimes b = a \times b$

$\overline{1}$ 1

$\overline{0}$ 0

Parsing as Weighted Logic Programming

$$\overset{\alpha}{\text{constit}}(a, i, j) \oplus = \overset{\alpha}{\text{constit}}(b, i, k) \otimes \overset{\alpha}{\text{constit}}(c, k+1, j) \otimes \overset{p(r)}{\overset{w}{\text{rule}}}(a \rightarrow b\, c)$$

$$\text{constit}(a, i, i) \oplus = \text{rule}(a \rightarrow w)$$ (bottom tree conditions)

Goal: $\text{constit}(S, 0, n)$

$R = \mathbb{R} \times \text{TREES}$

$a + b = (\max\{a_1, b_1\},$ the corresponding combination of those two subtrees)

" (a_1, a_2)

number tree

$$\begin{array}{ccc} & A & \\ a_2 & & b_2 \end{array}$$