Parsing

Homework 11 (PCFG parameter estimation with EM, weighted deductive parsing), due 05 July 2021

Laura Kallmeyer

SS 2021, Heinrich-Heine-Universität Düsseldorf

Question 1 (PCFG parameter estimation with EM)

Consider the following PCFG:

$G = \langle \{S, A, X\}, \{a, c\}, P, S, p\rangle$ *with* P *and* p *as follows:*

0.5: $S \to AX$ *0.3:* $S \to XA$ *0.1:* $S \to SA$ *0.1:* $S \to c$ *1:* $X \to AA$ *1:* $A \to a$

Assume that we have training data consisting of 5 times the string $w_1 = c$ *and twice the string* $w_2 = aaaa$.

The inside and outside values for these strings are as follows:

Inside values:

c:

j	
1	(0.1,S)
	1 i

aaaa:

j				
4	(0.08,S)	(0.8,S)	(1,X)	(1,A)
3	(0.8,S)	(1,X)	(1,A)	
2	(1,X)	(1,A)		
1	(1,A)			
	1	2	3	4 i

Outside values:

c:

j	
1	(1,S)
	1 i

aaaa:

j					
4	(1,S)	(0.5,X)	(0.8,A)	(0.08,A)	
3	(0.1,S) (0.3,X)	(0.5,X) (0.8, A)	(0.8, A)		
2	(0.01,S) (0.03,X) (0.8,A)	(0.08,A), (0,005,X)			
1	(0.001,S), (0.08,A), (0.003,X)				
	1	2	3	4	i

Assume that the above probabilities are our starting probabilities for a parameter estimation using EM, with, as mentioned, training data consisting of 5 times the string $w_1 = c$ *and twice the string* $w_2 = aaaa$.

Perform the two steps only for the S productions. (The other productions will stay with probability 1 anyway.)

1. *E-step: Compute the new counts* $C_c(A \to \alpha)$ *and* $C_{aaaa}(A \to \alpha)$ *and, based on these, the new frequency* $f(A \to \alpha)$ *for all* $A \to \alpha \in P$.
2. *M-step: Compute the new probabilities* $\hat{p}(A \to \alpha)$ *for all* $A \to \alpha \in P$*, based on the previous frequencies.*

Solution:

1. $C_c(S \to c) = \frac{(\beta_{S,1,1})p(S \to c)}{\alpha_{S,1,1}} = 1$

   For all other productions $A \to \alpha$, we have $C_c(A \to \alpha) = 0$.

(For the following: due to the α chart, A's can only span exactly one terminal, X and S have to span at least two.)

$$C_{aaaa}(S \to SA) = \frac{\beta_{S,1,3}\alpha_{S,1,2}\alpha_{A,3,3}p(S\to SA)}{\alpha_{S,1,4}} + \frac{\beta_{S,1,4}\alpha_{S,1,3}\alpha_{A,4,4}p(S\to SA)}{\alpha_{S,1,4}} + \frac{\beta_{S,2,4}\alpha_{S,2,3}\alpha_{A,4,4}p(S\to SA)}{\alpha_{S,1,4}}$$
$$= \frac{0+1\cdot 0.8\cdot 1\cdot 0.1+0}{0.08} = 1$$

$$C_{aaaa}(S \to AX) = \frac{\beta_{S,1,3}\alpha_{A,1,1}\alpha_{X,2,3}p(S\to AX)}{\alpha_{S,1,4}} + \frac{\beta_{S,1,4}\alpha_{A,1,1}\alpha_{X,2,4}p(S\to AX)}{\alpha_{S,1,4}} + \frac{\beta_{S,2,4}\alpha_{A,2,2}\alpha_{X,3,4}p(S\to AX)}{\alpha_{S,1,4}}$$
$$= \frac{0.1\cdot 1\cdot 1\cdot 0.5+0+0.8\cdot 1\cdot 1\cdot 0.5}{0.08} = \frac{0.05}{0.08} = 0.625$$

$$C_{aaaa}(S \to XA) = \frac{\beta_{S,1,3}\alpha_{X,1,2}\alpha_{A,3,3}p(S\to XA)}{\alpha_{S,1,4}} + \frac{\beta_{S,1,4}\alpha_{X,1,3}\alpha_{A,4,4}p(S\to XA)}{\alpha_{S,1,4}} + \frac{\beta_{S,2,4}\alpha_{X,2,3}\alpha_{A,4,4}p(S\to XA)}{\alpha_{S,1,4}}$$
$$= \frac{0.1\cdot 1\cdot 1\cdot 0.3+0+0.8\cdot 1\cdot 1\cdot 0.3}{0.08} = \frac{0.03}{0.08} = 0.375$$

$C_{aaaa}(S \to c) = 0$

$f(S \to SA) = 2 \cdot 1 = 2$

$f(S \to AX) = 2 \cdot 0.625 = 1.25$

$f(S \to XA) = 2 \cdot 0.375 = 0.75$

$f(S \to c) = 5 \cdot 1 = 5$

2. $\hat{p}(S \to SA) = \frac{2}{2+1.25+0.75+5} = \frac{2}{9} = 0.22$

   $\hat{p}(S \to AX) = \frac{1.25}{9} = 0.14$

   $\hat{p}(S \to XA) = \frac{0.75}{9} = 8.33 \cdot 10^{-2}$

   $\hat{p}(S \to c) = \frac{5}{9} = 0.56$

Question 2 (Weighted deductive parsing)

Consider the PCFG G with non-terminals $\{S, A, B\}$, terminals $\{a, b\}$, start symbol S and productions

$$\{ \quad \begin{array}{rl} 0.5\ (0.3) & S \to AS, \\ 0.3\ (0.5) & S \to SB, \\ 0.2\ (0.7) & S \to AB, \\ 1\ (0) & A \to a, \\ 1\ (0) & B \to b \end{array} \quad \}$$

The numbers preceding the productions are the corresponding probabilities, followed by the $|log(p)|$ in brackets.

As input consider $w = aabb$.

1. *Show the weighted deductive CYK-Parsing with chart and agenda using this grammar and input with weights as described on slide 7.*

   Notate chart and agenda in a table, the first column for the chart items, and the second for the agenda items. Concerning the chart column, it is enough to list only new items in each row. (This is different from the agenda where items are not only added but also removed and reordering depending on weights takes place.)

2. *The log used here is log_{10}. Compute the probability of the best parse tree from the weight of the goal item. (Note that the log numbers are rounded, the probability is not exactly the one we obtain from the rule probabilities.)*

Solution

1.

chart	*agenda*
	$0 : [A, 1, 1]$, $0 : [A, 2, 2]$, $0 : [B, 3, 3]$, $0 : [B, 4, 4]$
$0 : [A, 1, 1]$	$0 : [A, 2, 2]$, $0 : [B, 3, 3]$, $0 : [B, 4, 4]$
$0 : [A, 2, 2]$	$0 : [B, 3, 3]$, $0 : [B, 4, 4]$
$0 : [B, 3, 3]$	$0 : [B, 4, 4]$, $0.7 : [S, 2, 3]$
$0 : [B, 4, 4]$	$0.7 : [S, 2, 3]$
$0.7 : [S, 2, 3]$	$1 : [S, 1, 3]$, $1.2 : [S, 2, 4]$
$1 : [S, 1, 3]$	$1.2 : [S, 2, 4]$, $1.5 : [S, 1, 4]$
$1.2 : [S, 2, 4]$	$1.5 : [S, 1, 4]$ *(same weight, no update) goal reached*

2. *Probability of best parse:* $10^{-1.5} = 0.032$.