

Parsing

Homework 10 (PCFG Viterbi and inside and outside computation), due 28 July 2021

Laura Kallmeyer

SS 2021, Heinrich-Heine-Universität Düsseldorf

Question 1 (PCFG, viterbi) Consider the following CFG G with non-terminals $N = \{S, T, A, B\}$, terminals $T = \{a, b\}$, start symbol S and productions

- | | | | |
|-------------------|--------------------|-------------|--------------------|
| 0.3 (-0.5) | $S \rightarrow SX$ | 1 (0) | $X \rightarrow BC$ |
| 0.04 (-1.4) | $S \rightarrow ZS$ | 1 (0) | $Z \rightarrow AB$ |
| 0.05 (-1.3) | $S \rightarrow a$ | 1 (0) | $A \rightarrow a$ |
| 0.01 (-2) | $S \rightarrow b$ | 1 (0) | $B \rightarrow b$ |
| 0.6 (-0.2) | $S \rightarrow c$ | 1 (0) | $C \rightarrow c$ |

Preceding each production, its probability is given together with the \log_{10} value of the probability in parentheses.

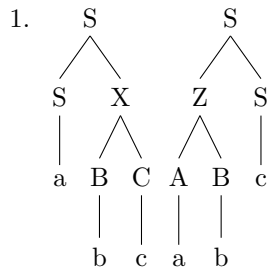
We consider the input $w = abc$.

1. Give the two parse trees for w together with the respective \log_{10} value of their probabilities.
2. Give the viterbi chart one obtains when using this PCFG in a CYK-parsing of the input abc . Calculate with the log values instead of the probabilities. This means that instead of multiplying probabilities, you just add their log values.

Explain how you computed the value for category S and indices $i = 1$ and $l = 3$.

Note that the higher a probability, the higher the log value and the better the item. (We are using $\log_{10}(p)$, not $|\log_{10}(p)|$.)

Solution:



with respective weights -1.8 and -1.6 .

2. Chart:

l				
3		$-1.6:S$		
2		$0:Z$	$0:X$	
1		$0:A, -1.3:S$	$0:B, -2:S$	$0:C, -0.2:S$
		1	2	3
				i

There are two possibilities for $S, 1, 3$: a) $S \rightarrow SX$, $\log(p)$ is $-0.5 - 1.3 = -1.8$ and b) $S \rightarrow ZS$, $\log(p)$ is $-1.4 - 0.2 = -1.6$. The second is better since $-1.6 > -1.8$.

Question 2 (Inside and Outside probabilities)

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

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

Compute the inside and outside charts for the strings $w_1 = c$ and $w_2 = aaaa$.

Solution:

Inside values:

		<i>aaaa:</i>				
		<i>j</i>				
<i>c:</i>	<i>j</i>	4	(0.08,S)	(0.8,S)	(1,X)	(1,A)
1	(0.1,S)	3	(0.8,S)	(1,X)	(1,A)	
	1	2	(1,X)	(1,A)		
		1	(1,A)			
		1	2	3	4	<i>i</i>

Outside values:

		<i>aaaa:</i>				
		<i>j</i>				
		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)	
<i>c:</i>	<i>j</i>	2	(0.01,S)	(0.08,A),		
1	(1,S)	1	(0.03,X)	(0.005,X)		
	1		(0.8,A)			
		1	(0.001,S),			
			(0.08,A),			
			(0.003,X)			
		1	2	3	4	<i>i</i>