

Einführung in die Computerlinguistik

Hausaufgabe zu PCFG, Abgabe 09.07.2018

Laura Kallmeyer

Sommer 2018, Heinrich-Heine-Universität Düsseldorf

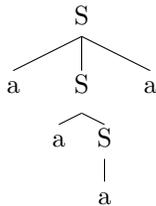
Aufgabe 1 Betrachten Sie folgende PCFG:

$$G = \langle \{S\}, \{a\}, \{0.6 : S \rightarrow aSa, 0.1 : S \rightarrow aS, 0.3 : S \rightarrow a\}, S \rangle$$

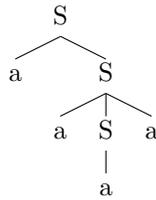
1. Geben Sie für die Eingabe $w = aaaa$ alle Parsbäume mit ihrer jeweiligen Wahrscheinlichkeit an.
2. Wie ist demnach die Wahrscheinlichkeit $P(w)$ des Wortes $w = aaaa$?
3. Was ist der beste (i.e., wahrscheinlichste) Parsbaum für $w = aaaa$? Geben Sie den Baum und seine Wahrscheinlichkeit an.

Lösung:

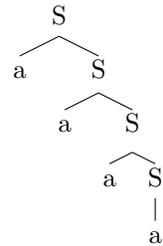
1. T_1 :



T_2 :

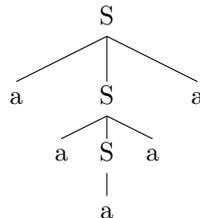


T_3 :



Wahrscheinlichkeiten: $P(T_1) = P(T_2) = 0.6 \cdot 0.1 \cdot 0.3 = 18 \cdot 10^{-3} = 0.018$
 $P(T_3) = 0.1 \cdot 0.1 \cdot 0.1 \cdot 0.3 = 3 \cdot 10^{-4} = 0.0003$

2. $P(aaaa) = 0.018 + 0.018 + 0.0003 = 0.0363$
3. Die erste S-Produktion ist besser als die zweite, außerdem erzeugt sie zwei a's, was mit der zweiten nur in zwei Schritten möglich ist. Daher ist es am besten, soweit möglich nur die erste Produktion zu verwenden. \Rightarrow



Wahrscheinlichkeit $0.6 \cdot 0.6 \cdot 0.3 = 108 \cdot 10^{-3} = 0.108$

Aufgabe 2 Betrachten Sie die PCFG G mit Nichtterminalen $\{S, A, B\}$, Terminalen $\{a, b\}$, Startsymbol S und Produktionen

$$\left\{ \begin{array}{l} 0.5 \quad S \rightarrow AS, \\ 0.2 \quad S \rightarrow SB, \\ 0.3 \quad S \rightarrow AB, \\ 0.8 \quad A \rightarrow AA, \\ 0.2 \quad A \rightarrow a, \\ 0.8 \quad B \rightarrow BB, \\ 0.2 \quad B \rightarrow b \end{array} \right\}$$

(Vor jeder Produktion steht ihre Wahrscheinlichkeit.)

- Geben Sie die Inside Chart für die Eingabe $w = aab$.
Vorsicht: Spaltenindex = Anfangsposition, Zeilenindex = Endposition des jeweiligen Spans.
Geben Sie insbesondere den Rechenweg für das Feld mit Indizes 1,3 an.
- Geben Sie die Viterbi chart für ein probabilistisches CYK parsing der Eingabe $w = aab$.
(Spaltenindex = Anfangsposition, Zeilenindex = Länge!)

Solution:

1. Inside Chart:

3	(S, 0.00312)	(S, 0.012)	(B, 0.2)
2	(A, 0.032)	(A, 0.2)	
1	(A, 0.2)		
	1	2	3

Feld 1,3: zwei Möglichkeiten: $0.5 \cdot 0.2 \cdot 0.012 + 0.3 \cdot 0.032 \cdot 0.2 = 0.0012 + 0.00192 = 0.00312$

2. Viterbi Chart:

3	0.00192 : S		
2	0.032 : A	0.012 : S	
1	0.2 : A	0.2 : A	0.2 : B
	1	2	3

Aufgabe 3 Betrachten Sie folgende PCFG:

$G = \langle \{N, A\}, \{n_1, n_2, a_1, a_2\}, P, N \rangle$ mit den folgenden Produktionen in P :

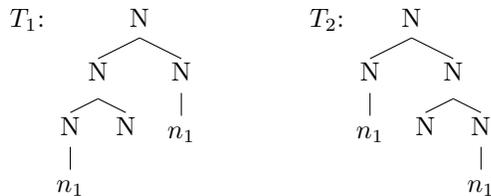
$0.5 : N \rightarrow NN$ $0.2 : N \rightarrow n_1$ $0.6 : A \rightarrow a_1$
 $0.2 : N \rightarrow AN$ $0.1 : N \rightarrow n_2$ $0.4 : A \rightarrow a_2$

Betrachten Sie die Eingabe $w = n_1 a_1 n_2 n_1$.

Geben Sie die Outsidewahrscheinlichkeit von N mit Indizes 2, 3 bezüglich dieser Eingabe an.

Lösung:

Außerhalb von der Teilkette der Eingabe mit Indizes 2, 3 ($= a_1 n_2$) gibt es davor ein n_1 und danach ebenfalls ein n_1 . Die Outsidewahrscheinlichkeit ergibt sich also als Summe aller Parsbäume mit Wurzel N (Startsymbol) und String $n_1 N n_1$ in den Blättern.



$$P(N \xrightarrow{*} n_1 N n_1) = 2 \cdot (0.5 \cdot 0.5 \cdot 0.2 \cdot 0.2) = 200 \cdot 10^{-4} = 0.02$$