

Einführung in die Computerlinguistik

Hausaufgabe zu PCFG, Abgabe 11.07.2022, 8.30 Uhr

Laura Kallmeyer

Sommer 2022, Heinrich-Heine-Universität Düsseldorf

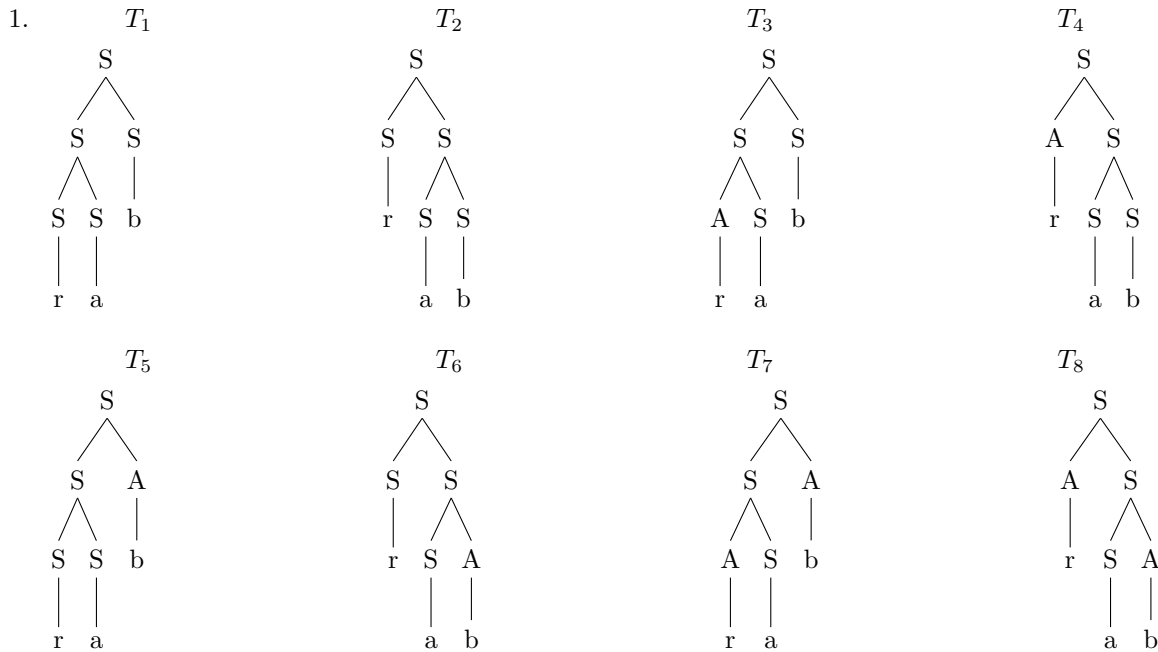
Aufgabe 1 Betrachten Sie folgende PCFG:

$G = \langle \{S, A\}, \{a, r, b\}P, S \rangle$ mit den folgenden Produktionen in P :

$0.3 : S \rightarrow SS$ $0.1 : S \rightarrow a$ $0.1 : S \rightarrow r$ $0.6 : A \rightarrow b$
 $0.2 : S \rightarrow AS$ $0.2 : S \rightarrow SA$ $0.1 : S \rightarrow b$ $0.4 : A \rightarrow r$

1. Geben Sie für die Eingabe $w = rab$ alle Parsbäume mit ihrer jeweiligen Wahrscheinlichkeit an.
2. Wie ist demnach die Wahrscheinlichkeit $P(w)$ des Wortes $w = rab$?

Lösung:



$$P(T_1) = P(T_2) = 0.3 \cdot 0.3 \cdot 0.1 \cdot 0.1 \cdot 0.1 = 9 \cdot 10^{-5}$$

$$P(T_3) = P(T_4) = 0.3 \cdot 0.2 \cdot 0.4 \cdot 0.1 \cdot 0.1 = 24 \cdot 10^{-5}$$

$$P(T_5) = P(T_6) = 0.3 \cdot 0.2 \cdot 0.6 \cdot 0.1 \cdot 0.1 = 36 \cdot 10^{-5}$$

$$P(T_7) = P(T_8) = 0.2 \cdot 0.2 \cdot 0.6 \cdot 0.1 \cdot 0.4 = 96 \cdot 10^{-5}$$

$$2. P(rab) = \sum_{i=1}^8 P(T_i) = 2(9 + 24 + 36 + 96) \cdot 10^{-5} = 2 \cdot 165 \cdot 10^{-5} = 330 \cdot 10^{-5}$$

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

$\{$ 0.5 $S \rightarrow AS,$
 0.2 $S \rightarrow SB,$
 0.3 $S \rightarrow AB,$
 0.8 $A \rightarrow AA,$
 0.2 $A \rightarrow a,$
 0.8 $B \rightarrow BB,$
 0.2 $B \rightarrow b$ $\}$

(Vor jeder Produktion steht ihre Wahrscheinlichkeit.)

1. 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.

2. 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 :

$0.5 : N \rightarrow NN$ $0.2 : N \rightarrow n_1$ $0.6 : A \rightarrow a_1$

$0.2 : N \rightarrow AN$ $0.1 : N \rightarrow n_2$ $0.4 : A \rightarrow a_2$

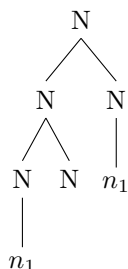
Betrachten Sie die Eingabe $w = n_1 a_1 n_2 n_1$.

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

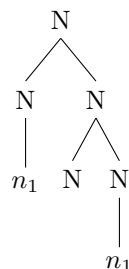
Lösung:

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.

T_1 :



T_2 :



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