

Einführung in die Computerlinguistik Vorbereitung der Abschlussklausur

18.01.2012

Laura Kallmeyer

WS 2011/2012, Heinrich-Heine-Universität Düsseldorf

Erlaubte Hilfsmittel in der Klausur: Eine Din-A4 Seite mit Notizen. Kein Taschenrechner.

Aufgabe 1 1. Betrachten Sie die folgende Merkmalsstrukturen, formuliert als Attribut-Wert Matrix. Geben Sie den entsprechenden Graphen an.

$$S_1 = \left[\begin{array}{l} \textit{phrase} \\ \text{HEAD} \left[\begin{array}{l} \textit{syn} \\ \text{CAT} \quad \textit{v} \\ \text{SUBCAT} \left[\begin{array}{l} \textit{subcat} \\ \text{NP1} \left[\begin{array}{l} \textit{phrase} \\ \text{CAT} \quad \textit{noun} \end{array} \right] \\ \text{NP2} \left[\begin{array}{l} \textit{phrase} \\ \text{CAT} \quad \textit{noun} \\ \text{CASE} \quad \textit{acc} \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{SUBJ} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{nom} \end{array} \right] \\ \text{OBJ} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{acc} \end{array} \right] \end{array} \right]$$

2. Betrachten Sie folgende Merkmalsstrukturen. *cooking* ist ein Untertyp von *event*.

$$S_2 = \left[\begin{array}{l} \textit{sign} \\ \text{PHON} \quad \text{"cooks"} \\ \text{SYN} \left[\begin{array}{l} \textit{syn} \\ \text{CAT} \quad \textit{v} \\ \text{SUBCAT} \left[\begin{array}{l} \textit{subcat} \\ \text{NP1} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{nom} \end{array} \right] \\ \text{NP2} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{acc} \end{array} \right] \end{array} \right] \end{array} \right] \\ \text{SEM} \left[\begin{array}{l} \textit{cooking} \\ \text{ARG1} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{nom} \end{array} \right] \\ \text{ARG2} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{acc} \end{array} \right] \end{array} \right] \end{array} \right]$$

$$S_3 = \left[\begin{array}{l} \textit{sign} \\ \text{SUBCAT} \left[\begin{array}{l} \textit{subcat} \\ \text{NP1} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{nom} \end{array} \right] \\ \text{NP2} \left[\begin{array}{l} \textit{phrase} \\ \text{CASE} \quad \textit{dat} \end{array} \right] \end{array} \right] \end{array} \right]$$

Angenommen, die Wahrscheinlichkeit, dass ein N am Satzanfang steht ist $2 \cdot 10^{-1}$, die, dass ein V am Satzanfang steht $3 \cdot 10^{-1}$. Die, dass ein Satzende auf Adj folgt, ist $1 \cdot 10^{-1}$.

1. Geben Sie die Viterbi Matrix an, die sich bei diesen Wahrscheinlichkeiten für die Eingabe learning changes thoroughly ergibt. Es reicht, die Einträge anzugeben, die $\neq 0$ sind.
2. Welche POS Tags ermittelt der Tagger für die Eingabe?

Lösung:

q_F				$216 \cdot 10^{-13}$, Adv
Adv			$216 \cdot 10^{-12}$, V	
V	$9 \cdot 10^{-4}$, q_0	$36 \cdot 10^{-8}$, V		
N	$2 \cdot 10^{-4}$, q_0	$108 \cdot 10^{-8}$, V		
	1 learning	2 changes	3 thoroughly	

2. Die beste POS-TAG Folge ist V V Adv.

Aufgabe 4 Betrachten Sie die folgende PCFG: $N = \{S, T, A, B, X, Y\}$, $T = \{a, b, c\}$, Startsymbol S .

Produktionen mit Wahrscheinlichkeiten:

$$6 \cdot 10^{-1}: S \rightarrow AT \quad 6 \cdot 10^{-1}: X \rightarrow AX \quad 5 \cdot 10^{-1}: Y \rightarrow YB$$

$$3 \cdot 10^{-1}: S \rightarrow AX \quad 4 \cdot 10^{-1}: X \rightarrow YB \quad 5 \cdot 10^{-1}: Y \rightarrow c$$

$$1 \cdot 10^{-1}: S \rightarrow c$$

$$1: T \rightarrow SB \quad 1: A \rightarrow a \quad 1: B \rightarrow b$$

1. Wieviele Linksableitungen gibt es für $w = aacbb$? Geben Sie sämtliche Ableitungen und deren Wahrscheinlichkeiten an.
2. Wie hoch ist die Inside Wahrscheinlichkeit von T und $acbb$, gegeben die Gesamteingabe w , das heißt mit einem Teilstring vom 2. bis zum 5. Terminalen in w ?
3. Wie hoch ist die Outside Wahrscheinlichkeit von T und $acbb$, gegeben die Gesamteingabe w , das heißt mit einem Teilstring vom 2. bis zum 5. Terminalen in w ?
4. Wie hoch ist die Wahrscheinlichkeit dafür, dass sich ein Parsbaum für w ergibt, der einen Teilbaum mit Wurzelsymbol T enthält, der genau den Teilstring vom 2. bis zum 5. Terminalen in w abdeckt?

Lösung:

1. Es gibt drei Linksableitungen:

$$S \Rightarrow AT \Rightarrow aT \Rightarrow aSB \Rightarrow aATB \Rightarrow aaTB \Rightarrow aaSBB \Rightarrow aacBB \Rightarrow aacbB \Rightarrow aacbb$$

$$\text{Wahrscheinlichkeit } 6 \cdot 10^{-1} \cdot 6 \cdot 10^{-1} \cdot 1 \cdot 10^{-1} = 36 \cdot 10^{-3}$$

$$S \Rightarrow AT \Rightarrow aT \Rightarrow aSB \Rightarrow aAXB \Rightarrow aaXB \Rightarrow aaYBB \Rightarrow aacBB \Rightarrow aacbB \Rightarrow aacbb$$

$$\text{Wahrscheinlichkeit } 6 \cdot 10^{-1} \cdot 3 \cdot 10^{-1} \cdot 4 \cdot 10^{-1} \cdot 5 \cdot 10^{-1} = 360 \cdot 10^{-4} = 36 \cdot 10^{-3}$$

$$S \Rightarrow AX \Rightarrow aX \Rightarrow aAX \Rightarrow aaX \Rightarrow aaYB \Rightarrow aaYBB \Rightarrow aacBB \Rightarrow aacbB \Rightarrow aacbb$$

$$\text{Wahrscheinlichkeit } 3 \cdot 10^{-1} \cdot 6 \cdot 10^{-1} \cdot 4 \cdot 10^{-1} \cdot 5 \cdot 10^{-1} \cdot 5 \cdot 10^{-1} = 1800 \cdot 10^{-5} = 18 \cdot 10^{-3}$$

2. Es gibt zwei mögliche Linksableitungen $T \xrightarrow{*} acbb$ (siehe die Teilableitungen der ersten beiden Ableitungen), nämlich

$T \Rightarrow SB \Rightarrow ATB \Rightarrow aTB \Rightarrow aSBB \Rightarrow acBB \Rightarrow acbB \Rightarrow acbb$, Wahrscheinlichkeit $6 \cdot 10^{-1} \cdot 1 \cdot 10^{-1} = 6 \cdot 10^{-2}$

und

$T \Rightarrow SB \Rightarrow AXB \Rightarrow aXB \Rightarrow aYBB \Rightarrow acBB \Rightarrow acbB \Rightarrow acbb$, Wahrscheinlichkeit $3 \cdot 10^{-1} \cdot 4 \cdot 10^{-1} \cdot 5 \cdot 10^{-1} = 60 \cdot 10^{-3} = 6 \cdot 10^{-2}$

Die Wahrscheinlichkeit ist demnach bei beiden $6 \cdot 10^{-2}$, insgesamt ergibt sich also

$$6 \cdot 10^{-2} + 6 \cdot 10^{-2} = 12 \cdot 10^{-2} = 0.12$$

3. Es gibt nur eine mögliche Linksableitung $S \xrightarrow{*} aT$, mit Wahrscheinlichkeit

$$6 \cdot 10^{-1} = 0.6$$

4. Der Wert ergibt sich als Produkt von Inside und Outside:

$$12 \cdot 10^{-2} \cdot 6 \cdot 10^{-1} = 72 \cdot 10^{-3} = 0.072$$

Aufgabe 5 Betrachten Sie die folgende PCFG: $N = \{S, T, A, B, X, Y\}$, $T = \{a, b, c\}$, Startsymbol S .

Produktionen mit Wahrscheinlichkeiten:

$$6 \cdot 10^{-1}: S \rightarrow AT \quad 7 \cdot 10^{-1}: X \rightarrow AX \quad 5 \cdot 10^{-1}: Y \rightarrow YB$$

$$3 \cdot 10^{-1}: S \rightarrow AX \quad 3 \cdot 10^{-1}: X \rightarrow YB \quad 5 \cdot 10^{-1}: Y \rightarrow c$$

$$1 \cdot 10^{-1}: S \rightarrow c$$

$$1: T \rightarrow SB \quad 1: A \rightarrow a \quad 1: B \rightarrow b$$

(Identisch wie in der vorhergehenden Aufgabe, abgesehen von zwei Wahrscheinlichkeiten.)

1. Geben Sie die Chart inkl. Produktionen und Zwischenlängen an, die sich bei einem Parsen von acb mit dieser Grammatik ergibt.

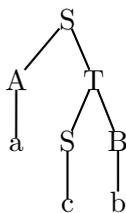
2. Wie sieht der beste Parsbaum aus? Welche Wahrscheinlichkeit hat er?

Lösung:

1. Chart:

3	$6 \cdot 10^{-2}, S \rightarrow AT, 1$			
2		$1 \cdot 10^{-1}, T \rightarrow SB, 1$ $15 \cdot 10^{-2}, X \rightarrow YB, 1$ $25 \cdot 10^{-2}, Y \rightarrow YB, 1$		
1	$1, A \rightarrow a, -$	$5 \cdot 10^{-1}, Y \rightarrow c, -$ $1 \cdot 10^{-1}, S \rightarrow c, -$	$1, B \rightarrow b, -$	
1	a	2	3	i
		c	b	

2. Bester Parsbaum:



Wahrscheinlichkeit $6 \cdot 10^{-2}$.