

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*.

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

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

$$S_4 = \left[\begin{array}{l} \text{sign} \\ \\ \\ \text{SUBCAT} \\ \\ \\ \text{SEM} \end{array} \left[\begin{array}{l} \text{subcat} \\ \text{NP1 } \boxed{11} \left[\begin{array}{l} \text{phrase} \\ \text{CASE nom} \end{array} \right] \\ \text{NP2 } \boxed{12} \left[\begin{array}{l} \text{phrase} \\ \text{CASE dat} \end{array} \right] \\ \text{NP3 } \boxed{13} \left[\begin{array}{l} \text{phrase} \\ \text{CASE acc} \end{array} \right] \\ \text{event} \\ \text{ARG1 } \boxed{11} \\ \text{ARG2 } \boxed{12} \\ \text{ARG3 } \boxed{13} \end{array} \right] \right]$$

Berechnen Sie $S_2 \sqcup S_3$, $S_2 \sqcup S_4$ und $S_3 \sqcup S_4$.

Aufgabe 2 Nehmen Sie folgende Gruppe von Personen an: Anna, Emil, Achim, Doris, Eva, Carla, Alexander, Adelheid. Nehmen Sie weiterhin an, dass Zettel mit den Namen dieser Personen in einem Behälter sind, und dass daraus gezogen wird.

1. Betrachten Sie die Ergebnismenge $\Omega = \{A, E, D, C\}$ (für "Name mit A", "Name mit E", ...). Berechnen Sie die Wahrscheinlichkeiten für die einelementigen Teilmengen von Ω , vorausgesetzt, alle Zettel werden mit gleicher Wahrscheinlichkeit gezogen.
2. Berechnen Sie die Wahrscheinlichkeit dafür, dass bei einmaligem Ziehen ein Name gezogen wird, der weder mit A noch mit D beginnt.
3. Berechnen Sie die Wahrscheinlichkeit dafür, dass bei zweimaligem Ziehen a) ohne Zurücklegen und b) mit Zurücklegen zwei Namen mit gleichem Anfangsbuchstaben gezogen werden.

Aufgabe 3 Nehmen Sie an, Sie haben einen HMM-POS Tagger, mit dem folgende Eingabe getaggt werden soll: learning changes thoroughly. Dem Tagger liegen folgende Wahrscheinlichkeiten zugrunde:

Emissionswahrscheinlichkeiten:

$$P(\text{learning}|V) = 3 \cdot 10^{-3} \quad P(\text{changes}|V) = 4 \cdot 10^{-3} \quad P(\text{thoroughly}|Adv) = 2 \cdot 10^{-3}$$

$$P(\text{learning}|N) = 1 \cdot 10^{-3} \quad P(\text{changes}|N) = 3 \cdot 10^{-3}$$

Alle anderen Emissionswahrscheinlichkeiten für unsere Eingabewörter seien 0.

Relevante Übergangswahrscheinlichkeiten:

$$P(N|N) = 1 \cdot 10^{-1} \quad P(N|V) = 4 \cdot 10^{-1} \quad P(Adv|V) = 3 \cdot 10^{-1}$$

$$P(Adv|N) = 1 \cdot 10^{-1} \quad P(V|N) = 3 \cdot 10^{-1} \quad P(V|V) = 1 \cdot 10^{-1}$$

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?

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

symbol S .

Produktionen mit Wahrscheinlichkeiten:

$$\begin{array}{lll} 6 \cdot 10^{-1}: S \rightarrow AT & 6 \cdot 10^{-1}: X \rightarrow AX & 5 \cdot 10^{-1}: Y \rightarrow YB \\ 3 \cdot 10^{-1}: S \rightarrow AX & 4 \cdot 10^{-1}: X \rightarrow YB & 5 \cdot 10^{-1}: Y \rightarrow c \\ 1 \cdot 10^{-1}: S \rightarrow c & & \\ 1: T \rightarrow SB & 1: A \rightarrow a & 1: B \rightarrow b \end{array}$$

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?

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

Produktionen mit Wahrscheinlichkeiten:

$$\begin{array}{lll} 6 \cdot 10^{-1}: S \rightarrow AT & 7 \cdot 10^{-1}: X \rightarrow AX & 5 \cdot 10^{-1}: Y \rightarrow YB \\ 3 \cdot 10^{-1}: S \rightarrow AX & 3 \cdot 10^{-1}: X \rightarrow YB & 5 \cdot 10^{-1}: Y \rightarrow c \\ 1 \cdot 10^{-1}: S \rightarrow c & & \\ 1: T \rightarrow SB & 1: A \rightarrow a & 1: B \rightarrow b \end{array}$$

(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?