

Sprachtechnologie

4. Übung

Prof. Tobias Scheffer
Uwe Dick

Sommer 2015

Ausgabe am: 11.05.15
Besprechung am: 18.05.15

Aufgabe 1

Viterbi-Algorithmus

Angenommen Sie hätten die folgenden Wahrscheinlichkeiten aus einem Korpus geschätzt:

t \ t+1	JJ	VBG	VBP	PRP	VBD	π	schlau	werden	wir
JJ	0.3	0.0	0.1	0.3	0.3	0.3	0.2	0.0	0.0
VBG	0.2	0.2	0.2	0.2	0.2	0.2	0.0	0.3	0.0
VBP	0.2	0.3	0.1	0.3	0.1	0.1	0.0	0.5	0.0
PRB	0.0	0.4	0.3	0.1	0.2	0.3	0.0	0.0	0.2
VBD	0.2	0.1	0.2	0.2	0.3	0.1	0.4	0.0	0.0

Zeichnen Sie das entstandene HMM als Automat und bestimmen Sie mit Hilfe des Viterbi-Algorithmus die wahrscheinlichste Zustandsfolge zu der Beobachtung „*Wir werden schlau*“. Visualisieren Sie anschließend Ihre Berechnung und Lösung im Trellis. Geben Sie eine probabilistische reguläre Grammatik an, welche die gleiche Wahrscheinlichkeitsverteilung modelliert.

Aufgabe 2

PCFG

Gegeben sei der folgende Auszug einer probabilistischen kontextfreien Grammatik mit dem Startsymbol S :

Nicht-Terminale				P	Terminale				P
S	→	P	VP	0.8	P	←	ich	0.3	
S	→	NP	AK	0.2	Q	←	im	0.2	
AK	→	PP	N	0.6	N	←	Raketenauto	0.8	
QP	→	Q	N	0.75	V	←	fahre	0.4	
PP	→	V	Q	0.5					
VP	→	V	QP	0.4					
NP	→	P		0.3					

Visualisieren Sie alle Parsebäume, die den Satz „*Ich fahre im Raketenauto*“ generieren und berechnen Sie die Wahrscheinlichkeit, dass die obige Grammatik diesen Satz generiert.

Aufgabe 3

Trigramm-Tagger

Bei der Wortarterkennung mit HMMs stellt jeder Zustand ein POS-Tag dar. Um größere Kontextabhängigkeiten zu erkennen, müssen die Zustände mehrere POS-Tags kodieren. Wie groß ist der Zustandsraum für einen Trigramm-Tagger bei m möglichen POS-Tags?