Sprachtechnologie

3. Übung

Prof. Tobias Scheffer
Paul Prasse
Michael Großhans
Sebastian Arzt
Philipp Schmidt

Sommer 2014

Ausgabe am: 05.05.14
Besprechung am: 12.05.14

Aufgabe 1 *T9*

Die naive Implementierung der Eingabehilfe T9 (siehe Übung 2), bei der alle möglichen Eingaben $w_1, \dots w_T$ durchsucht werden, hat eine exponentielle Laufzeit in T. In der Vorlesung wurde erklärt, dass man durch eine geeignete Dekodierung eine Laufzeit von $\mathcal{O}(V^N)$ erreichen kann. Geben Sie einen konkreten Algorithmus mit dieser Laufzeit an.

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 *ML-Schätzer*

Zeigen Sie, dass für den ML-Schätzer des n-Gramm-Modells

$$\theta^{ML}_{v_{i_n}|\dots v_{i_1}} = \frac{x_{v_{i_n}\dots v_{i_1}}}{x_{v_{i_{n-1}}\dots v_{i_1}}}$$

gilt. (Siehe Folie 21 der Sprachmodell-Vorlesung)

Hinweis: Leiten Sie die log-Likelihood von $\theta^{ML}_{i|j} = \arg\max_{\theta_{i|j}} \prod_{g=v_1\dots v_1}^{v_k\dots v_k} \left(\prod_{k\in V} \theta^{x_{kg}}_{k|g}\right)$ ab und betrachten Sie gegebenfalls nur den Fall $K = 2$.