

Stochastisches Parsing

PCFGs lassen sich nicht nur zum Parsen von Sätzen verwenden, sondern auch um Phrasen zu erkennen. Gegeben sei die folgende Grammatik mit Startsymbol NP:

NP	=>	Det N	0.6
NP	=>	Det N N	0.4
N	=>	N N	0.4
N	=>	train	0.1
N	=>	conductor	0.2
N	=>	training	0.3
Det	=>	the	1.0

1. Berechne die Wahrscheinlichkeit der NP „the train conductor training“ mit Hilfe des naiven Algorithmus, d.h. durch Aufzählen aller Parsebäume.
2. Berechne die Wahrscheinlichkeit der NP mit Hilfe des Inside Algorithmus (bitte machen Sie eine Tabelle, wie in der Vorlesung bzw. M&S'99, Table 11.3 auf Seite 194).
3. Berechne die Wahrscheinlichkeit des besten Parsebaums mit Hilfe des geeigneten Algorithmus.
4. Gib die Viterbi-Backtraces an und verfolge den Pfad durch sie zurück, um die Struktur des wahrscheinlichsten Parsebaums anzuzeigen.
5. Berechnen Sie die Outside Wahrscheinlichkeiten für die NP. Dazu benutzen Sie wieder eine Tabelle, diesmal fangen Sie rechts oben an und arbeiten sich vor zur Diagonalen.

N.B. Da die Grammatik ternär ist, muss sie entweder in Chomsky-Normalform umgeformt werden, oder die Algorithmen können intuitiv durch weitere Terme erweitert werden.