

Statistisches Parsing

Detlef Prescher

3. Juli 2007

Hausaufgabe 10

(i) Beschreibe einige Vorteile, aber auch einige Nachteile von DOP gegenüber PCFGs.

(ii) In DOP ist jeder Baum auch ein Fragment. Beschreibe den Unterschied von Baum- und Fragment-Wahrscheinlichkeit. Musst Du fuer Baum- und Fragment-Wahrscheinlichkeit zwei unterschiedliche Notationen benutzen?

(iii) Denk Dir einen Satz aus, der zwei Syntaxbäume hat. Liste nun alle DOP-Fragmente auf (Nimm kein zu triviales Beispiel!) Welches ist der wahrscheinlichere Syntaxbaum, wenn Du das DOP1-Modell benutzt, um beide Syntaxanalysen zu bewerten?

(iv) Zeige, dass man relativ effizient die wahrscheinlichste DOP-Derivation eines Satzes finden kann. Hinweis: Versuche eine Codierung zu finden, die jedes Fragment einer gegebenen DOP-Grammatik in eindeutiger Weise als eine (oder mehrere) Regel(n) einer neuen CFG kodiert. Zeige weiter, dass der wahrscheinlichste CFG-Baum des Satzes sich nun auf die wahrscheinlichste DOP-Derivation des Satzes abbilden lässt.

(v) Denk Dir ein Verfahren aus, den wahrscheinlichsten DOP-Baum eines Satzes zu berechnen. Ist Dein Verfahren effizient?

Abgabetermin: Dienstag, 10. Juli

Besprechung: Donnerstag, 12. Juli