

Programmierkurs Python II – SS 2013

Übung 8

Probabilistische Regeln als Paare nach dem Schema (LHS, (prob, [RHS1,RHS2,...])) repräsentiert werden, die Lexikon-Regeln als dictionary nach dem Schema `wort : [(prob1, POS1), (prob2,POS2)]` (siehe Template).

1 Inside Probabilities & CNF (2 Punkte)

Schreibe eine Funktion `binarize(rules)`, die Grammatik-Regeln mit mehr als 2 Nichtterminalen auf der rechten Seite binarisiert und den modifizierten Regeln korrekte Wahrscheinlichkeitswerte zuweist. Die Methode soll eine Grammatik im obigen Format als Eingabe nehmen und eine äquivalente binarisierte Grammatik wieder ausgeben. (Ketten- und Tilgungsregeln können für dies Aufgabe ignoriert werden.)

2 Inside Probabilities & CYK (2 Punkte)

Erweitere die Implementierung des CYK-Erkenners vom letzten Übungsblatt so, dass er Eingaben die entsprechenden „inside probabilities“ zuweist; dafür muss für jeden Chart-Eintrag die Wahrscheinlichkeit zur Verfügung stehen. Implementiere die Methode `recognize(start,rules,lexicon,sentence)` so, dass sie ein Paar aus dem Erkennen-Wahrheitswert (True, wenn der Satz von der Grammatik erkannt wird) und der Wahrscheinlichkeit des Satzes zurück gibt. (2 Punkte)

3 Viterbi scores (4 Punkte)

Passe die CYK-Parsers dem letzten Übungsblatt 1 so an, dass der wahrscheinlichste Ableitungsbaum und dessen Wahrscheinlichkeit zurückgegeben werden. Implementiere dafür die Methode `parse_cyk(start,rules,lexicon,sentence)` so, dass sie ein Paar aus dem Wahrscheinlichsten Satzbaum für den Satz (`Tree`) und dessen Wahrscheinlichkeit zurückgibt.

Abgabe bis Donnerstag, 27.06.2012, 11:00 Uhr