

Schriftliche Prüfung zur Vorlesung
Statistische Sprachverarbeitung
SS 2014
Helmut Schmid

Aufgabe 1) Wie lautet das Zipfsche Gesetz und was sagt es aus? (2 Punkte)

Aufgabe 2) Nennen Sie 3 Beispielanwendungen von Sprachmodellen in der Computerlinguistik. (1 Punkt)

Aufgabe 3) Wie wird die Maximum-Likelihood-Schätzung der bedingten Wahrscheinlichkeit $p(b|a)$ des Wortpaares (a,b) aus Häufigkeiten $f(a, b)$ berechnet? Welches Problem kann hierbei auftreten und welche Lösungen existieren dafür? (3 Punkte)

Aufgabe 4) Der Inside-Outside-Algorithmus kann zum Training von probabilistischen kontextfreien Grammatiken (PCFG) auf unannotierten Daten benutzt werden.

Was benötigen Sie außer dem Trainingskorpus sonst noch für das Training? (1 Punkt)

Wie lauten die Formeln zur Berechnung der Inside- und Outside-Wahrscheinlichkeiten? (3 Punkte)

Wie berechnen Sie die erwartete Häufigkeit der Parsewaldregel $A \rightarrow B_1 \dots B_m$, wenn die Inside- und Outside-Wahrscheinlichkeiten bereits berechnet wurden? (3 Punkte)

Was wird im M-Schritt des EM-Algorithmus getan? (1 Punkt)

Aufgabe 5) Erklären Sie die Grundidee des Berkeley-Parsers von Petrov und Klein. Wie wird der Parser trainiert? (3 Punkte)

Aufgabe 6) Wie wird der F-Score zur Messung der Genauigkeit von Parsern berechnet? (3 Punkte)

Aufgabe 7) Erklären Sie, wie der Vorzeichentest (Binomialtest) funktioniert, den man zur Berechnung der Signifikanz beim Vergleich zweier Wortart-Tagger benutzen kann. (3 Punkte)

Aufgabe 8) Warum stellen unbekannte Wörter beim Wortart-Taggen und Parsen ein Problem dar? Wie kann man mit unbekannten Wörtern umgehen? (3 Punkte)

Aufgabe 9) Erklären Sie, wie der Perzeptron-Algorithmus funktioniert und geben Sie die Formel für die Gewichts Anpassung an. (4 Punkte)

(30 Punkte insgesamt)

Viel Erfolg!