

Einführung in die statistische Sprachverarbeitung

Detlef Prescher

Zwischenklausur, Mittwoch, 4. Juli 2007: Für die Klausur sind fünf der sechs Aufgaben zu lösen. Wird für jede der sechs Aufgaben ein Lösungsversuch abgegeben, so wird der Lösungsversuch der Aufgabe (vi) nicht gewertet. Viel Erfolg!

(i) Definiere die Begriffe: Korpus, Wahrscheinlichkeitsverteilung, Modell, Korpus-Wahrscheinlichkeit, sowie Maximum-Likelihood-Estimierung (MLE). Erstelle eine Abbildung, die MLE hinreichend genau beschreibt.

(ii) Beschreibe drei Probleme mit dem MLE und gib jeweils ein Beispiel an.

(iii) Beweise, dass die Maximierung der Korpuswahrscheinlichkeit $L(f, \cdot)$ und die Minimierung der relativen Entropie $D(\cdot || \tilde{p})$ stets zu denselben Lösungen führen.

(iv) Wann genau stimmen RFE und MLE für ein beschränktes Modell überein? Erstelle zu Deiner Aussage eine kleine Abbildung. Beweise Deine Aussage. (Hinweis: Nutze die Aussage in (iii) aus, sowie die Informationsungleichung der Informationstheorie.)

(v) Wie ist Charniaks Baumbank-Grammatik definiert? Ist sie stets das RFE auf der Baumbank? Ist sie stets ein MLE auf der Baumbank? Gib jeweils ein Gegenbeispiel oder eine kleine Beweisskizze an.

(vi) Erstelle eine Abbildung, die die Prozedur des EM-Algorithmus hinreichend genau beschreibt. Erstelle eine weitere Abbildung, die den M-Schritt des EM-Algorithmus hinreichend genau beschreibt.

(Ende der Zwischenklausur)