

Einführung in die statistische Sprachverarbeitung

Detlef Prescher

12. Juli 2007

Hausaufgabe 11

(i) Was bezeichnet man als Language Modeling? Gib drei interessante Beispiele an.

(ii) Was ist die sogenannte Markov-Annahme? Wie ist ein $(N - 1)^{th}$ -order Markov Model definiert? Wie ist die Wahrscheinlichkeit eines Korpus bzgl. eines solchen Modells definiert? Berechne die Wahrscheinlichkeit des Korpus "*ich liebe es und liebe es und es liebt auch mich*" mit dem 0th-, 1st-, 2nd-, und 3rd- oder Markov-Model (trainiere diese Modelle vorher auf diesem Korpus). Gib auch für jedes Modell einen weiteren Korpus mit einer Null-Wahrscheinlichkeit an. (Benutze möglichst nur die Worte *ich, mich, es, und, auch, liebe, liebt.*)

(iii) Beweise, dass sich die $(N - 1)^{th}$ -order-Wahrscheinlichkeiten der Korpora einer festen Länge l auf 1 aufsummieren.

(iv) Was ist die Chain-Rule? Diskutiere N-gram-Modelle im Kontext dieser Regel...

(v) Warum benötigen wir eigentlich Smoothing? Zeige, dass Add-One-Smoothing und Lineare Interpolation wieder zu echten Wahrscheinlichkeitsmodellen führen.

Abgabetermin: Dienstag, 17. Juli

Besprechung: Mittwoch, 18. Juli