

Ergänzungen zum maschinellen Übersetzen natürlicher Sprachen

3. Übungsblatt

2015-06-09

In dieser Übung werden wir zusätzlich zwei kleine Beispiele aus der Vorlesung nachholen.

Aufgabe 1

Let Σ be an alphabet and (G, p) with $G = (Q, q_S, R)$ a probabilistic regular tree grammar over Σ . Assume (G, p) is *consistent*, i.e., $\llbracket (G, p) \rrbracket$ is a probability distribution. Show that the expected count of a rule $\rho = \langle q_1, \dots, q_k, \sigma, q_0 \rangle \in R$ in a derivation of this grammar is $\alpha(q_0) \cdot p(\rho) \cdot \prod_{i=1}^k \beta(q_i)$ where α and β represent the outside and inside weights w.r.t. (G, p) , respectively.

Aufgabe 2

Imagine the following game: Two coins are thrown and you win, if both coins land on the same side. You are only told, if you won. Assume that one coin is rather thick and may land on the edge. We represent the possible outcomes with the set $Y = \{M_1, M_2, M_3\} \times \{N_1, N_2\}$. You win with the outcomes (M_1, N_1) and (M_2, N_2) .

After playing the game several times, you won 6 times and lost 18 times. Instantiate the corpus-based EM algorithm with this scenario and calculate one EM step. Start with the probability of $2/5$ for M_1 and M_2 , and $1/3$ for N_1 .