Maschinelles Übersetzen natürlicher Sprachen

7. Übungsblatt

2016-12-08

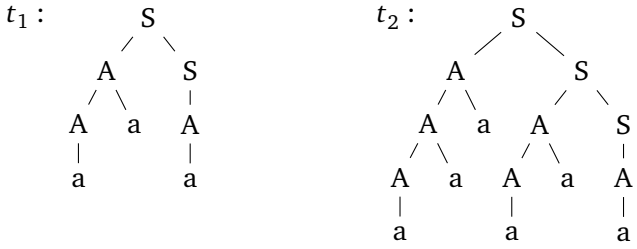
Aufgabe 1
Let $q \in [0,1]$ and let (G,p) be a probabilistic context-free grammar with start symbol S and the following rules and probabilities:

$$S \to SS \quad \# \, q$$
$$S \to a \quad \# \, 1-q$$

1. Find (recursive) definitions for the number of derivations for a^n and the probability $P(a^n \mid (G,p))$ where $n \geq 1$.

2. Show that (G,p) is consistent iff $q \leq 0.5$. (*Hint:* Don't get distracted by the solution of the first task.)

Aufgabe 2
Consider the following trees.



Let C be a corpus with $C(t_1) = 2$ and $C(t_2) = 1$. Train a pcfg on C.

Aufgabe 3
Supervised training of pcfg results in proper and consistent pcfgs. Give intuitive arguments why these pcfgs are consistent.

Aufgabe 4
Let (G,p) be a pcfg with the following rules and probability assignment.

$$S \to S \quad \# \, 0.25$$
$$S \to SS \quad \# \, 0.25$$
$$S \to a \quad \# \, 0.25$$
$$S \to \varepsilon \quad \# \, 0.25$$

Determine the reduct $a \lhd G$; you may ignore rules which are not useful. Approximate the probability of a.

Aufgabe 5
Let (G,p) be a pcfg with $G=(N,\Sigma,S,R)$. Let *in* be defined as in the lecture.

$$in\colon N \to \mathbb{R}_{\infty}^{\geq 0}\colon A \mapsto \sum_{d\in D_G(A,\Sigma^*)} P(d)$$

Show that the following equation holds, as stated in the lecture.

$$\forall A\in N\colon in(A) = \sum_{\substack{k\in\mathbb{N},\ w_0,\ldots,w_k\in\Sigma^*,\\ A_1,\ldots,A_k\in N\colon\\ r=A\to w_0A_1w_1\ldots A_kw_k\in R_A}} p(r)\cdot\prod_{i=1}^{k} in(A_i)$$

Aufgabe 6
Let (G,p) be a proper pcfg with $G=(N,\Sigma,S,R)$. Using your knowledge about *in*, show the following.

$$\forall A\in N\colon \sum_{d\in D_G(A,\Sigma^*)} P(d) \leq 1$$