Maschinelles Übersetzen natürlicher Sprachen

4. Übungsblatt

2015-11-12

Aufgabe 1

Using the context-free grammar in Figure 1, give a leftmost derivation, and the corresponding parse tree and abstract syntax tree for the sentence "I fly to Alaska".

S	→	NP VP
NP	→	Pronoun | Proper-Noun | Det Nominal
Nominal	→	Noun Nominal | Noun
VP	→	Verb | Verb NP | Verb NP PP | Verb PP
PP	→	Preposition NP
Noun	→	flight | breeze | trip | morning | ...
Verb	→	is | prefer | like | need | want | fly
Pronoun	→	me | I | you | it | ...
Proper-Noun	→	Alaska | Baltimore | Los Angeles | Chicago | ...
Det	→	the | a | an | this | these | that | ...
Preposition	→	from | to | on | near | ...

Figure 1: Productions of a context-free grammar [JM09, Figs. 9.2. and 9.3, p. 330].

Aufgabe 2

Give a context-free grammar over the terminal alphabet $\{(,), [,]\}$ which represents the well-braced strings over this alphabet (Dyck language). For example, ([ ])[ ] is well-braced, while ([ )] is not.

Aufgabe 3

Let Σ be an alphabet and $t \in \mathrm{U}_\Sigma$. Formally define the set of positions of t, denoted by $\mathrm{pos}(t)$. A position is a sequence of integers greater or equal than one. Such a sequence describes a (partial) path through the tree starting at the root. The integers determine with which sub-tree to proceed.

Let $p \in \mathrm{pos}(t)$. Formally define the label of t at p, denoted by $t(p)$, and the sub-tree of t at p, denoted by $t|_p$.

Aufgabe 4

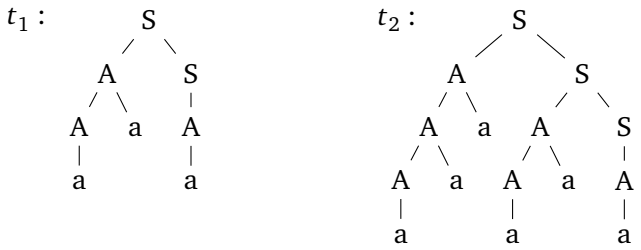
Let $q \in [0, 1]$ and let (G, p) be a probabilistic context-free grammar with start symbol S and the following rules and probabilities:

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

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 5
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 6
Supervised training of pcfg results in proper and consistent pcfgs. Give intuitive arguments why these pcfgs are consistent.

References

[JM09] Daniel Jurafsky and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Upper Saddle River, NJ, USA, second edition edition, 2009.