

Language Technology

7. Übung

Prof. Tobias Scheffer
Uwe Dick

Sommer 2016

Ausgabe am: 07.06.16
Besprechung am: 13.06.16

Aufgabe 1

Phrase-based Model

Gegeben sei ein (sehr kleiner) paralleler Corpus mit zwei alignierten Sätzen.

(Ich habe das kleine Haus für die Katze gekauft, I've bought the small house for the cat)

Außerdem habe sie Schätzungen von Wort-zu-Wort-Übersetzungswahrscheinlichkeiten gegeben. Im folgenden ein Ausschnitt:

Wahrscheinlichkeiten sind beispielhaft angegeben als $P(I|ich) = 0,7$.

ich	I	0.7	Haus	house	0.6
ich	me	0.2	Haus	Gebäude	0.1
habe	have	0.3	Haus	shell	0.1
habe	've	0.3	gekauft	bought	0.8
habe	got	0.1	gekauft	believe	0.05
habe	own	0.1	für	for	0.9
das	the	0.8	für	favor	0.05
das	that	0.1	die	the	0.6
kleine	small	0.4	die	that	0.2
kleine	little	0.4	Katze	cat	0.8

Ihre Aufgabe ist es, ein phrasenbasiertes Übersetzungsmodell für die Übersetzung von Deutsch nach Englisch zu lernen. Dafür nehmen wir an, dass Sie bereits ein englisches Sprachmodell haben.

1. Als erstes müssen Sie also ein Wort-zu-Wort-Alignment erstellen.
2. Anschließend sollten Sie, ausgehend vom Alignment, Phrasenpaare extrahieren.
3. Als letzten Schritt müssen noch Phrasenübersetzungswahrscheinlichkeiten berechnet werden. (Aufgrund des kleinen Corpus sind diese ziemlich eindeutig)

Aufgabe 2

Hierarchische Phrasenübersetzung

Bestimmen Sie nun, ausgehend von den in 1.1 erstellten Phrasenpaaren, alle Regeln für die hierarchische Phrasenübersetzung.

Aufgabe 3*Synchrone PCFG*

Gegeben Sei folgende synchrone PCFG für Sätze in deutscher und gnarfischer Sprache, wobei S das Startsymbol ist:

Regel	Wahrscheinlichkeit
$S \rightarrow A_{\boxed{1}}A_{\boxed{2}}/A_{\boxed{2}}A_{\boxed{1}}$	0,5
$S \rightarrow C_{\boxed{1}}B_{\boxed{2}}/C_{\boxed{1}}B_{\boxed{2}}$	0,5
$A \rightarrow C_{\boxed{1}}A_{\boxed{2}}/A_{\boxed{2}}C_{\boxed{1}}$	0,2
$A \rightarrow D_{\boxed{1}}A_{\boxed{2}}/A_{\boxed{2}}D_{\boxed{1}}$	0,3
$A \rightarrow \text{Übung/flarg}$	0,3
$A \rightarrow \text{Spaß/yarg}$	0,2
$B \rightarrow B_{\boxed{1}}D_{\boxed{2}}/D_{\boxed{2}}B_{\boxed{1}}$	0,3
$B \rightarrow \text{Übung/schnarg}$	0,4
$B \rightarrow \text{Spaß/hjarb}$	0,3
$C \rightarrow C_{\boxed{1}}B_{\boxed{2}}/C_{\boxed{1}}B_{\boxed{2}}$	0,3
$C \rightarrow \text{Die/warb}$	0,7
$D \rightarrow \text{macht/bralb}$	1

Übersetzen Sie mit Hilfe der Grammatik den Satz “Die Übung macht Spaß” vom Deutschen in die Gnarf-Sprache, indem Sie zuerst den wahrscheinlichsten Parsebaum bestimmen.