Maschinelles Übersetzen natürlicher Sprachen

2. Übungsblatt

2015-10-29

Aufgabe 1

1. Extend the bigram model from the lecture to a trigram model, i.e. a language model where the probability of a word in a sentence depends on two preceding words.
2. Define a general n-gram model.

Aufgabe 2 (Bigrams)
Let $E = \{\text{du}, \text{su}, \text{ur}, \text{fur}, \text{mu}\}$ and $F = \{\text{kra}, \text{ban}, \text{las}, \text{gha}, \text{ra}\}$. Consider the following dictionary:

		f				
$t(f \mid e)$		kra	ban	las	gha	ra
	du	1	0	0	0	0
	su	0	1	0	0	0
e	ur	0	0	1	0	0
	fur	0	0	0	1	0
	mu	0	0	0	0	1

Let $\varepsilon(m \mid l) = 1$ if $l = m$ and $\varepsilon(m \mid l) = 0$ otherwise. Decode the sentence "kra ban las gha ra" using the following bigram model:

		e'					
$b(e' \mid e)$		du	su	ur	fur	mu	#
	du	0	0.2	0.1	0.3	0.2	0.2
	su	0.3	0	0	0.3	0.3	0.1
e	ur	0.1	0	0	0.7	0	0.2
	fur	0.3	0.2	0.2	0	0	0.3
	mu	0.2	0.5	0	0	0	0.3
	#	1	0	0	0	0	0

Aufgabe 3
Let $V_E = \{a, b\}$ and $V_F = \{\alpha, \beta, \gamma\}$ be an English and French vocabulary, respectively. Consider the following bigram model, length model, and dictionary.

$b(\rightarrow \mid \downarrow)$	#	a	b
#	2^{-1}	2^{-1}	0
a	0	2^{-1}	2^{-1}
b	2^{-1}	0	2^{-1}

$$\epsilon(m \mid l) = \begin{cases} 2^{-1} & \text{if } m = l \\ 2^{-2} & \text{if } |m - l| = 1 \\ 0 & \text{otherwise} \end{cases}$$

$t(\rightarrow \mid \downarrow)$	α	β	γ
a	2^{-1}	0	2^{-1}
b	0	2^{-1}	2^{-1}

Decode the sentence $\beta\gamma$ using the algorithm from the lecture. Annotate the hypotheses with a target length to make the calculation feasible.

Zusatzaufgabe 1
Write a small program to solve the previous task.