

Maschinelles Übersetzen natürlicher Sprachen

1. Übungsblatt

2014-10-16

Aufgabe 1 (Wörterbuch erstellen – intuitiv)

Versuchen Sie anhand des folgenden Korpus ein Wörterbuch aufzustellen. Beachten Sie, dass einige Worte mehrdeutig sein können und somit mehrere mögliche Übersetzungen haben.

	<i>Language 1</i>	–	<i>Language 2</i>
<i>Pair 1</i>	kra ban las gha	–	du su ur fur
<i>Pair 2</i>	mar las ban	–	ur du su
<i>Pair 3</i>	ban fab mar kra	–	mux su fur du
<i>Pair 4</i>	mar kra ad	–	krus du fur lub

Welche Annahmen haben Sie zum Erstellen des Wörterbuches getroffen? Sind die Annahmen realistisch? Fallen Ihnen bessere Ansätze ein?

Aufgabe 2 (Summen und Produkte)

Sei $n \in \mathbb{N}$. Für alle $i \in \{1, \dots, n\}$ sei A_i eine Menge und $f_i: A_i \rightarrow \mathbb{R}$ eine Abbildung. Zeigen Sie, dass unter diesen Voraussetzungen folgende Gleichung gilt:

$$\sum_{a_1 \in A_1} \cdots \sum_{a_n \in A_n} \prod_{i=1}^n f_i(a_i) = \prod_{i=1}^n \sum_{a \in A_i} f_i(a).$$

Welche Rolle spielt diese Beziehung für IBM-Modell 1?

Aufgabe 3 (IBM-Modell 1)

Betrachten Sie das folgende Wörterbuch (die Worte „du“, „su“, ... sind „Englische“ Worte).

	kra	ban	las	gha
du	0.2	0.4	0.4	0
su	0.0	0.1	0.8	0.1
ur	0.3	0.4	0.25	0.25
fur	0.4	0.3	0.1	0.2

(a) Bestimmen Sie $P(f | e)$ für die folgenden Satzpaare auf Basis von IBM-Modell 1. Nutzen Sie dazu einen sinnvollen Parameter ε .

- $f = \text{kra las gha}, e = \text{du su ur}$,
- $f = \text{kra las gha}, e = \text{du su}$,
- $f = \text{gha gha}, e = \text{su du su}$,

(b) Sei $\varepsilon(l | m) = 2^{-m}$. Bestimmen Sie $\operatorname{argmax}_f P(f | e)$ für die folgenden Sätze.

- $e = \text{du}$,
- $e = \text{du su ur}$,
- $e = \text{fur du fur su fur}$,

Nehmen Sie nun an, $\varepsilon(l | m) = 1$ wenn $l = m$ und ansonsten $\varepsilon(l | m) = 0$. Versuchen Sie die beste Übersetzung für die oben genannten Englischen Sätze zu bestimmen.