

Tutorium zu „Einführung in die Computerlinguistik“

Maschinelle Übersetzung

Maschinelle Übersetzung

Basic Translation Model:

$$P(f|e) \propto \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(\langle a_1, \dots, a_m \rangle) \prod_{j=1}^m P(f_j | e_{a_j})$$

>> Problem: Wort für Wort-Übersetzung erkennt vieles nicht richtig

Lösung: **Language Model** >> wie wahrscheinlich kommt dieser Satz so in der Zielsprache vor?

$\operatorname{argmax}_{\text{word-sequence}}$	$=$	$P(\text{evidence} \text{word-sequence})$	$P(\text{word-sequence})$
		translation model	language model
		Übersetzung	Wortfolge

>>> für Wörter der Zielsprache, die nicht zugeordnet werden können: empty cept e_0 (fiktives Wort in Ausgangssprache)

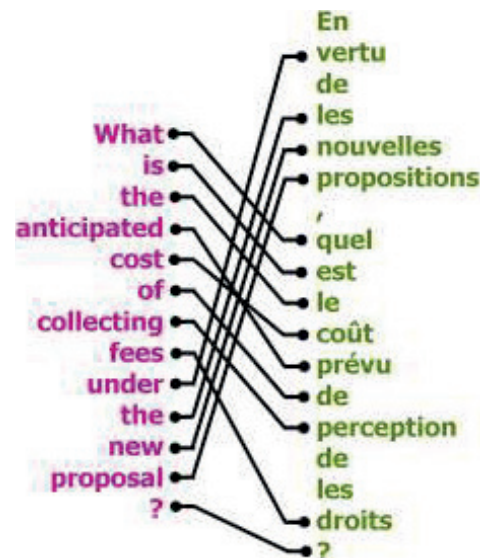
z.B. Tafelsalz —> sel **de** table

drita —> **das** Licht

Relative Häufigkeit & Übergangswahrscheinlichkeit

Wahrscheinlichkeit für eine Übersetzung >> Rel. Häufigkeit

(siehe auch Naive Bayes – Priors)



$$P(e_j | \text{,nouvelles'}) = 1$$

>> 1 zu 1-Zuordnung

$$P(f_j | \text{,the'}) = \text{je } 0,5$$

>> 2 ,the'-Alignments

$$P(f_j | e_0) = \text{je } 0,2$$

>> 5 e_0 -Alignments

$$P(\text{,en'} | e_0) = 0,2$$

>> 1 ,en' / 5 e_0 -Alignments

$$P(\text{,de'} | e_0) = \text{je } 0,4$$

>> 2 ,de' / 5 e_0 -Alignments

Wahrscheinlichkeit für eine Wortfolge >> Übergangswahrscheinlichkeit

(siehe auch Markov-Model 1. Ordnung)

Bert gönnt sich morgens gerne eine schöne heiße Tasse Kaffee. Susi dagegen mag lieber eine schöne heiße Tasse Tee. Susi mag keinen Kaffee, nicht mal eine Tasse Kaffee.

ohne Smoothing:

$$P(\text{Kaffee} | \text{Tasse}) = 2 / 3$$

$$P(\text{Tee} | \text{Tasse}) = 1 / 3$$

mit Smoothing:

Vokabular = 20 (mit Satzz.)

$$P(\text{Kaffee} | \text{Tasse}) = (2+1) / (3+20)$$

$$P(\text{Kakao} | \text{Tasse}) = (0+1) / (3 + 20)$$

Übung

They left the building

e	f	P(f e)
sie	they	0,8
verließen	left	0,4
links	left	0,6
das	the	0,7
Gebäude	building	0,6
bauen	building	0,5

w_t w_{t-1}	w_{t+1} w_t	$P(w_{t+1} w_t)$ $P(w_t w_{t-1})$
sie	verließen	0,5
verließen	das	0,6
das	Gebäude	0,7
sie	links	0,2
links	das	0,3
das	bauen	0,4

Mögliche deutsche Übersetzungen:

- A: sie links das Gebäude
- B: sie verließen das bauen
- C: sie links das bauen
- D: sie verließen das Gebäude

Übung

I like cheese cake

e	f	P(f e)
I	ich	1
like	wie	0,5
like	mag	0,6
cheese	Käse	0,7
cake	Kuchen	0,6
cheese cake	Käsekuchen	0,5

w_t w_{t-1}	w_{t+1} w_t	$P(w_{t+1} w_t)$ $P(w_t w_{t-1})$
ich	mag	0,6
ich	wie	0,3
mag	Käse	0,6
mag	Käsekuchen	0,6
Käse	Kuchen	0,1

Mögliche deutsche Übersetzungen:

A: ich mag Käse Kuchen

B: ich mag Käsekuchen