

Maschinelles Übersetzen natürlicher Sprachen

2. Übungsblatt

2014-10-23

Aufgabe 1

Nehmen Sie an, ein Wörterbuch t und ein Längenmodell ε seien gegeben, und wir verwenden einen Dekoder $h: F \rightarrow E$ mit

$$h(f) = \operatorname{argmax}_{e \in E} P(e | f, t, \varepsilon)$$

für jedes $f \in F$. Wir haben in der Vorlesung bereits gesehen, dass eine Definition, die argmax verwendet, im Allgemeinen nicht eindeutig ist, weil die Funktion hinter dem argmax mehrere Maxima haben kann. Daher oben die Formulierung „einen Dekoder“. Streng mathematisch müssten wir schreiben

$$h(f) \in \operatorname{argmax}_{e \in E} P(e | f, t, \varepsilon).$$

Es obliegt dem Implementierer, h geeignet zu präzisieren. Wie wird dies in der Praxis Ihrer Ansicht nach passieren?

Wir wissen an dieser Stelle nicht, wie der Implementierer vorgegangen ist. Daher können wir nur aus der Gleichung $h(f) = e$ nur schließen, dass

$$e \in \operatorname{argmax}_{e' \in E} P(e' | f, t, \varepsilon).$$

Welche (naheliegenden) Veränderungen können wir an e vornehmen, ohne dass diese Bedingung verletzt wird? Anders gefragt: Welche Übersetzungen von f lassen sich mit dem IBM-Modell 1 nicht unterscheiden?

Aufgabe 2

Für das Trainieren des Längenmodells haben wir $\hat{\varepsilon}$ in Abhängigkeit von einem parallelen Korpus d in Form eines Quotienten definiert. Dabei haben wir den Zähler und den Nenner umgangssprachlich formuliert. Formulieren Sie diese Größen formal. Beweisen Sie, dass $\hat{\varepsilon}$ tatsächlich eine bedingte Wahrscheinlichkeitsverteilung ist, dass also für jedes m die Gleichung $\sum_l \hat{\varepsilon}(l | m) = 1$ gilt.

Aufgabe 3

Zum Trainieren des Wörterbuchs wird ein iteratives Verfahren verwendet. Nach welchen Kriterien würden Sie ein solches Verfahren beurteilen? Ist es in jedem Fall ratsam, die Likelihood zu maximieren? Welche Abbruchbedingungen fallen Ihnen für die äußere Schleife ein? Geben Sie je eine Summenformel für den Wert von s in Zeile 10 und den Wert von $c(u | v)$ in Zeile 12 des Algorithmus an.

Aufgabe 4

Es seien $V_E = \{\text{children, let's, play}\}$ und $V_F = \{\text{Kinder, lasst, spielen, uns}\}$. Des Weiteren sei ein bilingualer Korpus $d = (f_1, e_1)(f_2, e_2)$ gegeben, wobei

$$(f_1, e_1) = (\text{Kinder spielen, children play})$$

$$(f_2, e_2) = (\text{lasst uns spielen, let's play})$$

Trainieren Sie mit dem in der Vorlesung vorgestellten Algorithmus ausgehend vom Korpus d ein Wörterbuch $t: V_E \times V_F \rightarrow [0, 1]$. Beginnen Sie mit gleichverteilten Übersetzungswahrscheinlichkeiten.

Aufgabe 5

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.