

Maschinelles Übersetzen natürlicher Sprachen

1. Übungsblatt

2016-10-20

Aufgabe 1

Let X be an arbitrary set, and $f: X \rightarrow \mathbb{R}$ and $g: \mathbb{R} \rightarrow \mathbb{R}$ be mappings such that g is strictly increasing, i.e., $\forall x, y \in \mathbb{R}: x < y \implies g(x) < g(y)$. Show that

1. $\forall x, y \in \mathbb{R}: g(x) < g(y) \implies x < y$ and
2. $\operatorname{argmax} f = \operatorname{argmax} g \circ f$.

Aufgabe 2 (Wörterbuch erstellen – intuitiv)

Versuchen Sie intuitiv 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 3 (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 4 (IBM-Modell 1)

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

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

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

(b) Sei $\varepsilon(l | m) = 2^{-l}$ wenn $l > 0$ und 0 sonst. Bestimmen Sie $\operatorname{argmax}_e P(e | f)$ für die folgenden Sätze.

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

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.

Aufgabe 5

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 6

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 7

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 8

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

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

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

Trainieren Sie mit dem in der Vorlesung vorgestellten Algorithmus ausgehend vom Korpus d ein Wörterbuch t , welches Vokabeln aus V_F in Vokabeln aus V_E probabilistisch übersetzt. Beginnen Sie mit gleichverteilten Übersetzungswahrscheinlichkeiten.