

Sprachtechnologie

9. Übung

Prof. Tobias Scheffer
Uwe Dick

Sommer 2016

Ausgabe am: 20.06.16
Besprechung am: 27.06.16

Aufgabe 1

Maschinelle Übersetzung

Maschinelle Übersetzungssysteme im Internet sind z.B.

- Freetranslation <http://www.freetranslation.com/>
- Babelfish <http://www.babelfish.de/>
- Google Translate <http://translate.google.com/>
- Systran <http://www.systranet.com/>

Wählen Sie einen deutschsprachigen Textausschnitt (ca. 5-10 Zeilen) und übersetzen Sie diesen mit mindestens zwei verschiedenen Systemen in eine von Ihnen gewählte Zielsprache. Betrachten Sie die beiden Ergebnisse unter folgenden Fragestellungen:

1. An welchen Stellen unterscheiden sich die Übersetzungen? Was könnte der Grund dafür sein?
2. An welchen Stellen ist die Übersetzung fehlerhaft und um welchen Fehlertyp handelt es sich hier (z.B. syntaktisch, lexikalisch, morphologisch)?
3. Wie arbeiten die Systeme vermutlich?
4. Welches zusätzliche Wissen über Sprache müssten die Systeme haben um eine korrekte Übersetzung erzeugen zu können?
5. Formulieren Sie den Ausgangstext so um, dass die Systeme eine bessere Übersetzung liefern könnten. Beschreiben Sie die Art der Umformulierungen und die Eigenschaften des resultierenden Textes.

Aufgabe 2

Dekodierung

Ein einfaches Spracherkennungssystem erkennt isolierte Worte. Für jedes Wort liefert das akustische Modell die $k = 3$ höchsten *likelihoods* zurück. Das akustische Modell muss für die Dekodierung eines Signals mit einem 2- Gramm-Sprachmodell kombiniert werden. Simulieren Sie einen Beam-Search-Algorithmus mit Beam-Breite $b = 2$, für ein Signal aus drei Wörtern. Was ist die wahrscheinlichste Wortfolge gegeben das Signal?

Ausgaben des akustischen Modells:

$$P(\text{Signal}_1|\text{doppelte}) = 0.01 \quad (1)$$

$$P(\text{Signal}_1|\text{hoppelte}) = 0.008 \quad (2)$$

$$P(\text{Signal}_1|\text{doppelter}) = 0.007 \quad (3)$$

$$P(\text{Signal}_2|\text{bin}) = 0.013 \quad (4)$$

$$P(\text{Signal}_2|\text{hin}) = 0.013 \quad (5)$$

$$P(\text{Signal}_2|\text{Gin}) = 0.008 \quad (6)$$

$$P(\text{Signal}_3|\text{Honig}) = 0.002 \quad (7)$$

$$P(\text{Signal}_3|\text{Tonic}) = 0.0015 \quad (8)$$

$$P(\text{Signal}_1|\text{sonnig}) = 0.003 \quad (9)$$

Das Sprachmodell sieht wie folgt aus:

1-Gramme	2-Gramme	
$P(\text{bin}) = 0.01$	$P(\text{bin} \text{doppelte}) = 0.001$	$P(\text{Honig} \text{bin}) = 0.001$
$P(\text{hin}) = 0.01$	$P(\text{hin} \text{doppelte}) = 0.001$	$P(\text{Tonic} \text{bin}) = 0.0005$
$P(\text{Gin}) = 0.01$	$P(\text{Gin} \text{doppelte}) = 0$	$P(\text{sonnig} \text{bin}) = 0.001$
$P(\text{doppelte}) = 0.01$	$P(\text{bin} \text{hoppelte}) = 0.0005$	$P(\text{Honig} \text{hin}) = 0$
...	$P(\text{hin} \text{hoppelte}) = 0.02$	$P(\text{Tonic} \text{hin}) = 0$
	$P(\text{Gin} \text{hoppelte}) = 0$	$P(\text{sonnig} \text{hin}) = 0$
	$P(\text{bin} \text{doppelter}) = 0.001$	$P(\text{Honig} \text{Gin}) = 0.001$
	$P(\text{hin} \text{doppelter}) = 0.0005$	$P(\text{Tonic} \text{Gin}) = 0.01$
$6 P(\text{Gin} \text{doppelter}) = 0.02$	$P(\text{sonnig} \text{Gin}) = 0$	

Aufgabe 3

Fragen

Überlegen Sie sich dabei drei Fragen zu den Themen Übersetzung und Spracherkennung. Versuchen Sie möglichst Fragen auszuwählen, auf die nur wenige Studenten eine Antwort haben.