

Fortgeschrittene Methoden der statistische maschinelle Übersetzung

Übung 4.2

Abgabe (pdf, Zip) an kaeshammer@phil.uni-duesseldorf.de
bis Montag, 19.01.2015, 13:00

1

Zur automatischen Evaluierung der Qualität von maschineller Übersetzung wird häufig der BLEU-Score verwendet. Er ist wie folgt definiert:

$$\text{BLEU-}n : \min \left(1, \frac{H}{R} \right) \prod_{k=1}^n \frac{\#(k\text{-grams present in } \mathbf{h} \text{ and } \mathbf{r})}{\#(k\text{-grams present in } \mathbf{h})}$$

- Erklären Sie, wie BLEU funktioniert. (Was sind die verschiedenen Bestandteile des Scores? Was sind \mathbf{r} und \mathbf{h} bzw. R und H ?)
- Warum macht es keinen Sinn BLEU für einzelne Sätze auszurechnen, und was macht man stattdessen?
- Wenn man das Beispiel geschickt wählt, kann man auch für nur einen Satz BLEU berechnen. Berechnen Sie BLEU-3 für die folgende Systemausgabe gegeben die Referenzübersetzung:

System: I want not but go home !

Referenz: but I don't want to go home !

2

Die folgenden Eigenschaften werden manchmal als Vorteile von klassischen regelbasierten (also manuell entwickelten) *oder* statistischen (z.B. phrasenbasierten oder auch hierarchischen) Übersetzungssystemen aufgeführt. Welche werden im Allgemeinen als positiv für welche Seite angesehen? Erklären Sie jeweils in wenigen Sätzen warum.

- Robustheit
- Breite Abdeckung