

Statistische Maschinelle Übersetzung

Teil VII - Fortgeschrittene Themen

insbesondere Baumbasierte Verfahren

Thomas Schoenemann

Heinrich-Heine-Universität Düsseldorf

Sommersemester 2012

Überblick

1. Modelle für starke Änderungen in der Wortordnung
(Inversion Transduction Grammars)
2. Hierarchische Übersetzung

Rückblick: Umordnungsmodelle

- Für Word Alignment
 - HMM: $p(a_j | a_{j-1}, I)$
 - IBM-4: $p_{=1}(d_{i,1} - \odot_{\text{prev}(i)}), p_{>1}(d_{i,k} - d_{i,k-1})$
- Für phrasenbasierte Übersetzung
 - $p(\text{start}_k - \text{end}_{k-1})$
- Gut geeignet zwischen europäischen Sprachen
(etwa Deutsch \leftrightarrow Englisch): relativ gleiche Wortstellungen
- Bei Englisch \leftrightarrow Japanisch, Englisch \leftrightarrow Chinesisch etc:
 - stark unterschiedliche Wortordnungen
 - globaleres Umordnungsmodell nötig

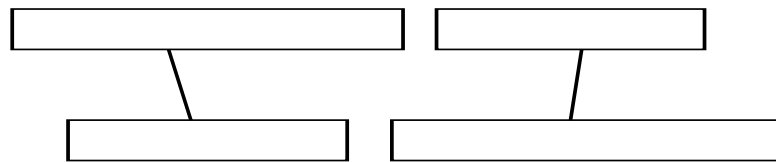
Inversion Transduction Grammars (ITG)

z.B. für Word Alignment, gegeben f_1^J, e_1^I :

das Spannenpaar $(1, J) - (1, I)$ lässt sich auf zwei Weisen zerlegen:

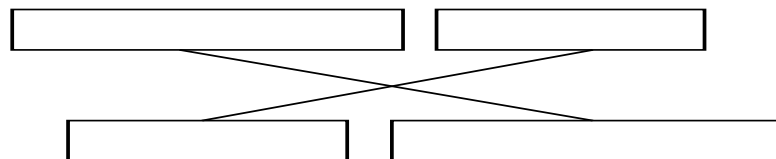
- in beiden Fällen, wähle ein $j < J$ und ein $i < I$.
- 1. Zerlegung (monoton):

$(1, j) - (1, i)$ und $(j + 1, J) - (i + 1, I)$



- 2. Zerlegung (Invertierung):

$(1, j) - (i + 1, I)$ und $(j + 1, J) - (1, i)$



Die entstehenden Spannen können nun ihrerseits zerlegt werden (\rightarrow Rekursion)

ITG: beste Zerlegung

Mit Wahrscheinlichkeiten p_M, p_I für monoton/Inversion
 ($p_M + p_I = 1$):

Wk. der wahrscheinlichsten Zerlegung:

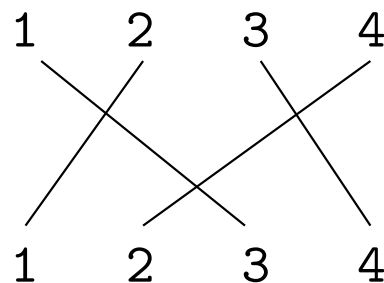
$$Q(j_1, j_2, i_1, i_2) = \max_{\substack{j_1 \leq j < j_2, \\ i_1 \leq i < i_2}} \max \left\{ p_M Q(j_1, j, i_1, i) Q(j+1, j_2, i+1, i_2), \right. \\ \left. p_I Q(j_1, j, i+1, i_2) Q(j+1, j_2, i_1, i) \right\},$$

wobei $|j_2 - j_1| \geq 2$, $|i_2 - i_1| \geq 2$, sonst geeignete Basiswk.s

ITG: Unmögliche Umordnungen

Nicht alle Umordnungen lassen sich durch ITG ausdrücken.

Insbesondere unmöglich:



Hierarchische Übersetzung

bei phrasenbasierter Übersetzung: **konsekutive Wortfolgen** können als einzelne Einheit übersetzt werden.

Eigentlich gewünscht: auch Phrasen mit Lücken.

Etwa in: `macht die Tür auf` → `opens the door`

Gewünschte Regel: `macht X auf` → `opens X`

Zwei neue Aufgaben:

- Extraktion von hierarchischen Regeln aus Word Alignments
- Übersetzung mit solchen Regeln
(ähnlich wie Parsing, hier nicht betrachtet)

Beispiel

	ich	werde	ihnen	die	entsprechenden	anmerkungen	aushändigen
i							
shall							
be							
passing							
on							
to							
you							
some							
comments							

regulär extrahierbare Phrasenpaare sind u.a.:

werde ihnen die entspr. anmerk. aushändigen
 -shall be passing on to you some comments

ihnen die entsprechenden anmerkungen
 -to you some comments

Konstruiere daraus:

werde X aushändigen

- shall be passing on X

Wiederholung: Extraktion regulärer Phrasen aus Word Alignments

Gegeben:

- (i) zwei Sätze \mathbf{f}, \mathbf{e} der Längen J und I .
- (ii) ein Word Alignment $\mathbf{a} \subseteq \{1, \dots, J\} \times \{1, \dots, I\}$

Das Phrasenpaar $f_{j_1}^{j_2}, e_{i_1}^{i_2}$ mit $j_1 \leq j_2, i_1 \leq i_2$ kann extrahiert werden, wenn diese drei Bedingungen gelten:

- $\exists i, j$ mit $i_1 \leq i \leq i_2, j_1 \leq j \leq j_2$, sodass $(j, i) \in A$
- $\forall i \in \{i_1, \dots, i_2\} : \nexists j$ sodass $j \notin \{j_1, \dots, j_2\}$ und $(j, i) \in A$
- $\forall j \in \{j_1, \dots, j_2\} : \nexists i$ sodass $i \notin \{i_1, \dots, i_2\}$ und $(j, i) \in A$

Beachte: hier kein Limit für die Längen $|j_2 - j_1|$ und $|i_2 - i_1|$.

Extraktion von hierarchischen Phrasen

Für gegebene $\mathbf{f}, \mathbf{e}, \mathbf{a}$:

1. Initialisiere die Menge \mathcal{P} mit allen regulär extrahierbaren Phrasen
2. Iteriere bis keine Änderungen:
Wenn $(\bar{e}, \bar{f}) \in \mathcal{P}$ und $(\bar{e}_{\text{SUB}}, \bar{f}_{\text{SUB}}) \in \mathcal{P}$ und $\bar{e} = \bar{e}_{\text{PRE}} \circ \bar{e}_{\text{SUB}} \circ \bar{e}_{\text{POST}}$
und $\bar{f} = \bar{f}_{\text{PRE}} \circ \bar{f}_{\text{SUB}} \circ \bar{f}_{\text{POST}}$ und $\bar{e} \neq \bar{e}_{\text{SUB}}$ und $\bar{f} \neq \bar{f}_{\text{SUB}}$,
dann ist auch $(\bar{e}_{\text{PRE}} + X + \bar{e}_{\text{POST}}, \bar{f}_{\text{PRE}} + X + \bar{f}_{\text{POST}}) \in \mathcal{P}$.

Optional: Limits für $|\bar{e}|$ und $|\bar{f}|$ (häufig: 15)

Mehrere Variablen

Regel bisher: werde X aushändigen - shall be passing on X

wobei $X =$ ihnen die entspr. anmerkungen bzw. $X =$ to you some comments

Linguistisch sinnvoller:

werde $X_1 X_2$ aushändigen - shall be passing on $X_1 X_2$

wobei $X_1 =$ Ihnen bzw. $X_1 =$ to you,

$X_2 =$ die entspr. anmerkungen bzw. $X_2 =$ some comments

Das obige Extraktionsverfahren lässt sich auf solche Regeln ausweiten.

Beispiele mit mehreren Variablen

z.B. Regel für eine Tasse Kaffee | a cup of coffee:

$X_1 X_2 \rightarrow X_1 \text{ of } X_2$