Sparse Features for Reordering

(Final Report)

Barry Haddow
Lexi Birch
Han Dan
(Lane Schwartz)

14th September, 2013

Lexicalised Reordering Models in PBMT

Max-Likelihood [Tillmann, 2004], [Koehn et al, 2005,2007]

- Count orientations in training data
- Maybe smooth
- "Standard Model"

Maxent e.g. [Zens and Ney, 2006]

- Build a classifier to predict M, S or D
- Use lexical features, part-of-speech etc.

Lexicalised Reordering Models in PBMT

Max-Likelihood [Tillmann, 2004], [Koehn et al, 2005,2007]

- Count orientations in training data
- Maybe smooth
- "Standard Model"

Maxent e.g. [Zens and Ney, 2006]

- Build a classifier to predict M, S or D
- Use lexical features, part-of-speech etc.

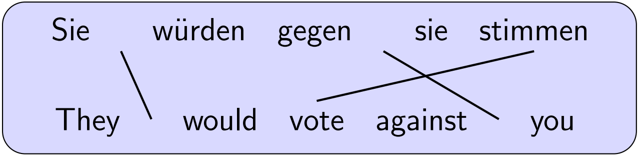
Add small number of features (e.g. 6) to translation system

Sparse Features for Reordering [Cherry, 2013]

- We now have methods to train (i.e. tune) translation systems with thousands or even millions of features
- Normally predicates on hypotheses
- Replace maxent model (indirect objective) ...

  ... with sparse feature model (direct objective, e.g. BLEU)
- Cherry showed improvements on large zh-en and ar-en models
  - → Over a maxent baseline
  - → In addition to Tillman-style model
  - → Lattice batch MIRA better than *k*-best batch MIRA

Sparse Features for Reordering – Example



Template	**Features**
src.left × *orientation*	`sl_sie_M`, `sl_stimmen_D`, `sl_gegen_S`
src.right × *orientation*	`sr_wurden_M`, `sr_stimmen_D`, `sr_sie_S`
tgt.left × *orientation*	`tl_they_M`, `tl_vote_D`, `tl_against_S`
tgt.right × *orientation*	`tr_would_M`, `tr_vote_D`, `tr_you_S`

Lexicalised Reordering in Hiero Models

- Zens&Ney Reordering (maxent)
  [Huck et al, 2012]
- Tillman Reordering (max-like)
  [Huck et al, 2013]
- Latter shown to perform better – both beat baseline

Sparse Reordering Features in Hiero Models

- Suppose we apply a rule:

$$X \rightarrow a\ X_1\ b\ X_2\ \mid\ X_1\ z\ X_2\ y$$

- With X_1 covering c and X_2 covering de
  - $\rightarrow$ Add features indicating that words c are *monotone* with respect to de
- Or if the rule is:

$$X \rightarrow a\ X_1\ b\ X_2\ \mid\ X_2\ p\ X_1$$

  - $\rightarrow$ Add features indicating that words c are *swapped* with respect to de

Sparse Reordering Features in Hiero Models

- Suppose we apply a rule:

$$X \rightarrow a\ X_1\ b\ X_2 \ \mid\ X_1\ z\ X_2\ y$$

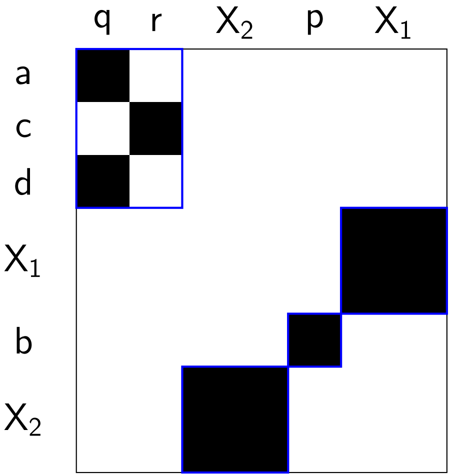
- With X_1 covering c and X_2 covering de
  - $\rightarrow$ Add features indicating that words c are *monotone* with respect to de
- Or if the rule is:

$$X \rightarrow a\ X_1\ b\ X_2 \ \mid\ X_2\ p\ X_1$$

  - $\rightarrow$ Add features indicating that words c are *swapped* with respect to de

Sparse Reordering in Hiero Models – Example

$$X \rightarrow acd\ X_1\ b\ X_2 \ \mid\ qrX_2\ p\ X_1$$



X_1 covers ef, X_2 covers g

```
src_left_d_mono
src_right_e_mono

src_left_f_swap
src_right_b_swap

src_left_b_swap
src_right_g_swap
```

Results

- Small de-en model (news commentary $\approx$ 130k sentences)
- Use `src.left` variant, and top 100 words.

Model	Tune	Test
Baseline	26.83	27.71
Sparse Reorder	27.15	27.90

(Baseline trained 25 iterations, Sparse reordering 10)

Feature weights

Feature	Weights
und_swap	-0.159
,_swap	-0.111
ein_mono	-0.046
sind_mono	-0.045
⋮	⋮
sich_mono	0.039
,_mono	0.057
OTHER_mono	0.062
die_mono	0.100