Exploiting Lexical Information and Discriminative Alignment Training in Statistical Machine Translation

Patrik Lambert

(supervisor: Rafael E. Banchs)
(tutor: Núria Castell)

TALP Research Center
Barcelona, Spain

DCU Internal MT Workshop, July 2008
1 Multi-word Expression Grouping

2 Alignment Minimum-Translation-Error Training
   - Introduction
   - Alignment System
   - Alignment Model Weight Optimisation

3 Further Work
Multi-word Expressions Grouping

- Method to extract automatically bilingual MWEs
- 2 experiments suggest that SMT improved by grouping MWEs before alignment (Giza ++):
- Verbmobil corpus:
Multi-word Expressions Grouping

- EPPS corpus:
  - extraction method probably too noisy: not only non-compositional MWE grouped
  - not quantitative improvement overall
  - detailed error analysis: when MWE grouped were non-compositional, grouping helped for their correct translation
Alignment Minimum-Translation-Error Training: Motivation

- In current SMT system, word alignment is performed independently from the MT system ⇒ alignment adapted to the MT system
Related Work

- IBM model 4:
  - not easily adaptable
  - large computational resources
- discriminative approach (log-linear combination of models):
  - easily adaptable to MT system (not done in practise)
  - some of them [Moore 2005] don’t require heavy training
  - inadequate model weight optimisation
- problem with model weight optimisation:
  - need of data annotated with word alignment can be a limitation
  - main difficulty: absence of adequate metric (bad correlation AER-MT)
- new metrics proposed [Fraser and Marcu 2006; Ayan and Dorr 2006]:
  - more informative for translation units extraction, but indirect method
Proposed Method

- Discriminative word alignment framework with some models designed based on the characteristics of our SMT system
- Optimise alignment model weights without annotated data, directly in a Minimum-translation-error Training scheme: use automated translation metrics as minimisation criterion.
- For large corpora, weights are optimised on a small part of the corpus and used to align whole corpus
BIA: Bilingual word Aligner

- Beam-search decoder minimising cost of linear combination of models (some of them similar to [Moore 2005])
- Two issues of N-gram MT can be dealt with at the alignment stage:
  - tuples with NULL source sides cannot be allowed
  - only one monotonic segmentation of each sentence pair is performed
    - long reorderings produce long and sparse tuples
- Features designed to help the N-gram SMT system:
  - distinct source and target unlinked word penalties
  - embedded word position penalty: penalises situations like this one:
    - link bonus: because N-gram model prefers higher recall alignments
BIA (ii)

Other basic features:
- two distortion features: number and amplitude of crossing links
- word association model(s)

Second pass features:
- Word association model with relative link probabilities [Melamed 00]
- Fertility model: probability to have zero, one, two, three or four and more links
Procedure

- Optimal coefficients were estimated on small subset as follows:

![Diagram of the procedure]

\[ \{\lambda_1, \ldots, \lambda_v\} \]
Procedure

- Optimal coefficients were estimated on small subset as follows:

- Once a set of optimal weights has been obtained:
  - align whole training corpus with optimal weights
  - extract translation units
  - train full SMT system: translation model + target language model, word bonus model and two lexical models.
AR→EN United Nations Results

- alignment weights tuned on 50k subset
- resulting SMT system was compared to identical system trained from grow-diag-final combination of source-target and target-source GIZA++ (50 cl, 1-4 H-5 4-4) alignments
- Results shown are the average and standard error (in parentheses) of 3 SMT model weight optimisations:

<table>
<thead>
<tr>
<th></th>
<th>BLEU</th>
<th>NIST</th>
<th>WER</th>
<th>METEOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results on the 50k subset (difference between systems lies in alignment)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIA</td>
<td>23.8 (0.1)</td>
<td>7.77 (0.1)</td>
<td>67.1 (1.3)</td>
<td>53.2 (0.2)</td>
</tr>
<tr>
<td>Giza++ GDF</td>
<td>22.5 (0.7)</td>
<td>7.41 (0.2)</td>
<td>70.7 (2.4)</td>
<td>52.9 (0.4)</td>
</tr>
<tr>
<td>Results on the full corpus (weights still tuned on 50k subset)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIA</td>
<td>27.0 (0.2)</td>
<td>8.15 (0.07)</td>
<td>64.8 (0.8)</td>
<td>54.4 (0.2)</td>
</tr>
<tr>
<td>Giza++ GDF</td>
<td>26.9 (0.3)</td>
<td>8.09 (0.1)</td>
<td>64.9 (1.0)</td>
<td>54.4 (0.1)</td>
</tr>
</tbody>
</table>
Further Work

Multi-word Expressions:
- refine extraction method
- use manually built resources (wordnet?)
- compare this work to Yanjun’s word packing experiments

Discriminative Alignment Training:
- improve the alignment system
- analyse the differences between alignments produced after optimisation in function of AER or MT metrics
- study the effect of the size of the subcorpus used to train alignment parameters, as well as the effect of the sentences selection method.
Thank you for your attention!