Improving Word Alignment for Statistical Machine Translation

Yanjun Ma

National Centre for Language Technology
School of Computing
Dublin City University

NCLT/CNGL MT Workshop
July 23, 2008
Outline

Word Alignment

Two Novel Approaches

System Integration

Further Work
Word Alignment and SMT

Word alignment is a fundamental component for all SMT systems

- Word-based SMT
- Phrase-based SMT (PB-SMT): conditional v.s. joint model
- Formal syntax-based SMT: hiero-style v.s. ITG
- Syntax-based SMT: Tree-to-String, String-to-Tree, Tree-to-Tree

Approaches to word alignment

- Generative word alignment
- Discriminative word alignment
- Heuristics-based word alignment
- A mixture of above mentioned approaches
Challenges in Word Alignment

Evaluation

- Intrinsic evaluation on gold-standard: AER, F-measure
- Extrinsic evaluation on MT task: influence on BLEU

Further improvement in alignment quality

- GIZA++ can produce fairly good alignment

Correlation between word alignment quality and translation quality (cf. Patrik’s talk)

- AER is shown not to correlate with BLEU, F-measure is better, but not good enough
Outline

Word Alignment

Two Novel Approaches

System Integration

Further Work
Exploiting Word Segmentation and Syntax for Word Alignment

Word segmentation

- Question the *monolingual* segmentation of languages, i.e. Chinese

Syntax

- Make use of monolingual processing technology (parsing)
Word Packing [Ma et al., 2007]

Motivation

- *Monolingual* segmenters bring noise into alignment task
- 1-to-n alignment is one of the most difficult cases for word alignment

| 白葡萄酒: white wine | closest: 最近 |
| 百货公司: department store | fifteen: 十五 |
| 抱歉: excuse me | fine: 很好 |
| 报警: call the police | flight: 次航班 |
| 杯: cup of | get: 拿到 |
| 必须: have to | here: 在这里 |
Work Flow

- Use a word aligner to extract candidate $1\sim n$ alignments
  - Bilingual sentence
    我想要一杯茶。
    I'd like a cup of tea.
  - Candidate extracted: $a_i = <c_i, E_i>$
    想要: 'd like
    杯: cup of

- Estimate the reliability of the candidates $a_i = <c_i, E_i>$
  - Co-occurrence frequency: $COOC(c_i, E_i)$
  - Alignment confidence: $AC(a_i) = \frac{C(a_i)}{COOC(c_i, E_i)}$

- Pack the reliable candidates with a single token

- Re-iterate word alignment $k$ times and stop when (i) there is no more packing to perform, or (ii) no improvement in translation is seen on the development set
Improved MT on IWSLT 2006 and IWSLT 2007 Chinese–English translation

<table>
<thead>
<tr>
<th></th>
<th>IWSLT 2006</th>
<th>IWSLT 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.1855</td>
<td>0.2897</td>
</tr>
<tr>
<td>Iteration=1</td>
<td>0.1902</td>
<td>-</td>
</tr>
<tr>
<td>Iteration=2</td>
<td>0.1945</td>
<td>0.3000</td>
</tr>
</tbody>
</table>

- Statistically significant \((p < 0.05)\) improvement
- On a relatively small training data (40k sentence pairs)
Word Alignment using Syntactic Dependencies
[Ma et al., 2008]

Motivation

- Incorporate syntactic dependencies into word alignment

```
我 1 打 2 网球 3 时 4 扭伤 5 的 6 。 7

I twisted 2 it 3 playing 4 tennis 5 。 6

我 1 愿意 2 要 3 二 4 号 5 。6

I 'll 2 have 3 the 4 number 5 two 6 。7
```
General model

\[ p(A|c_1^J, e_1^I) = \frac{1}{Z} \cdot p_\epsilon(A_\Delta|c_1^J, e_1^I) \cdot \prod_{j \in \Delta} p(a_j|c_1^J, e_1^I, a_1^{j-1}, A_\Delta) \]

- Decomposed into an anchor alignment model and a syntax-enhanced model
- Syntax-enhanced model

\[ \hat{a}_j = \arg\max_{a_j} \left\{ p_{\lambda_1^M}(a_j|c_1^J, e_1^I, a_1^{j-1}, A_\Delta) \right\} \]

\[ = \arg\max_{a_j} \left\{ \sum_{m=1}^{M} \lambda_m h_m(c_1^J, e_1^I, a_1^j, A_\Delta, T_c, T_e) \right\} \]

- Assuming a set of anchor alignment can be obtained, the syntactic dependencies can be transformed into features in a discriminative word alignment model
Improved MT on IWSLT 2008 Chinese–English translation

<table>
<thead>
<tr>
<th></th>
<th>CHALLENGE</th>
<th>BTEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>0.3194</td>
<td>0.3595</td>
</tr>
<tr>
<td>Syntax-Enhanced</td>
<td>0.3452</td>
<td>0.3823</td>
</tr>
</tbody>
</table>

- Statistically significant ($p < 0.05$) improvement
- Especially effective with a small training data (20k sentence pairs)
Outline

Word Alignment

Two Novel Approaches

System Integration

Further Work
Word Alignment, MATrEx and MT Evaluation

Word alignment modules in MATrEx (cf. John’s talk)

- Giza++, MTTK
- Word packing and its parallelisation
- Syntax-enhanced word alignment (not integrated)

Evaluation Campaign

- IWSLT 2007: Arabic–, Chinese–, Japanese–English
  [Hassan et al., 2007]
Outline

Word Alignment

Two Novel Approaches

System Integration

Further Work
Further Work

Error analysis for word alignment

- GIZA++
- Word packing
- Word alignment with syntactic dependencies
- Different errors when optimised with different criteria (AER v.s. BLEU)

Word packing incorporating syntactic dependencies

- Direct packing using dependencies
- Incorporate dependency (packing) information into HMM alignment models
- Comparison with fertility-based models

Optimising word alignment for SMT
Open Research Topics

Word alignment and MT

• Improving generative word alignment models
• Improving discriminative word alignment models
• Impact of segmentation models on word alignment models
• Impact of syntactic information on word alignment models
• Gold-standard construction: guidelines for manual word alignment
• Evaluation of word alignment: intrinsic v.s. extrinsic evaluation
• Correlation between internal and external word alignment quality
• Optimising word alignment for MT
• Word alignment and EBMT
• Evaluating word alignment using other applications besides MT
Thanks for your attention

yma@computing.dcu.ie
http://www.computing.dcu.ie/~yma
References

