Re-training Monolingual Parser Bilingually for Syntactic SMT

Reporter Xiang Li

†Shujie Liu, ‡Chi-Ho Li, ‡Mu Li and ‡Ming Zhou

† Harbin Institute of Technology
Outline

• Inconsistent Training Problem
• Re-training Method
  • Targeted Self-Training with Frontier Set Based Evaluation
  • Forced Alignment-based Parser Re-Training
• Symmetrization Method: IDSG
• Experiment
Inconsistent Training Problem

• Syntax based SMT methods start with word alignment and a parser tree, and obtain translation rules driven by the syntax

• During word alignment, it is not considered whether a structure in the source sentence can project to a legal structure in the target sentence

• Parser is a monolingual activity targeted for a specific task, and not necessarily well studied for MT

• Independent training causes problems during rule extraction

• Rule generalization problem

• Violation of syntactic correspondences
Basic concepts

- **Span**: the minimal contiguous source string that covers all the source words reachable from a node

- **Complement span**: the union of spans of all the nodes that are neither descendants nor ancestors of a node

- **Frontier node**: a node of which the span and the complement span do not overlap

- A minimal rule can be extracted from one frontier node

  $$X_1 X_2 \rightarrow \text{NP( the NN}(X_1) \text{ POS}(X_2))$$

- The more frontier nodes, the more minimal rules can be extracted
First Problem

- The incorrect alignment (6-5) and (6-7) links may violate the syntactic structures
-毡房 $\rightarrow$ NNS(yurts) can not be extracted
-牧民，的 $\rightarrow$ NP(DT(the),NN(herdsmen),POS('s))
Second Problem

- The incorrect POS tagging of the word "lectures" causes a series of parsing errors, including the absence of the noun phrase NP → (NN(propaganda), NN(lectures))
- 宣讲 → NP(NN(propaganda), NN(lectures))
Re-training Framework

• Two Solutions
  • Frontier set based parse tree selection
  • Forced alignment based parser selection
Targeted Self-Training with Frontier Set Based Evaluation

• Frontier Set Based Parse Tree Selection
  • Given word alignment, n-best parse trees
  • The tree which can generate the largest frontier set is selected as the most consistent tree

\[ \text{Tree} = \arg \max (\text{FrontierSetSize A, Tree'}) \text{ Tree' } \epsilon \text{Treenbest} \]

• Targeted Self-Training with Frontier Set Based Evaluation (TST-FS)
  • Unlike the standard self-training which select the best parse tree for the re-training, we select the parse tree which can generate the largest frontier set
Targeted Self-Training with Frontier Set Based Evaluation

Tree selected by TST-FS
Forced Alignment-based Parser Re-Training

- Parse tree from a monolingual parser may be not appropriate enough for translation
- It seems reasonable to consider using the parse tree produced by an syntactic SMT system to re-train the parser
- Naive idea: run an syntactic SMT system over some source sentences and retrieve the by-product target language parse trees for re-training the monolingual parser
  - Translation by an MT system is often a weird target sentence
  - Associated parse tree is of little use in improving the parser
- Forced alignment of bilingual data is a much more promising approach
Forced Alignment-based Parser Re-Training

- Given a bilingual sentence, Forced alignment performs phrase segmentation of the source side, parsing of the target side, and word alignment of the bilingual sentence, using the full translation system as in decoding.

- How: During decoding, any translation candidate which is not the sub-span of the target sentence is discarded, so that we can generate the exactly target sentence as final translation result.

A large number of people coming to listen to their propaganda lectures
Forced Alignment-based Parser Re-Training

- Parse tree generated by forced alignment is in the search space of decoder, which may be more suitable for parser re-training

- Due to the pruning of translation model training and decoding, forced alignment is not guaranteed to get the target parse tree

- To generate as many target parse tree as possible
  - **keep all the rules** during rule extraction
  - **enlarge the stack size** of each cell during decoding
  - both the source and target span can be **aligned to null** during decoding
Forced Alignment-based Parser Re-Training

Tree selected by FA-PR

来1 听2 他们3 宣讲4 的5 人6 很多，
Intersect Diag Syntactic Grow

- Parser re-training aims to improve a parser with alignment matrix
- How to use parse tree to improve alignment?
- Most word aligner are directional aligners, with a symmetrization method (such as Intersect-Diag-Grow: IDG) to combine both directions alignment result
- We try to take parse tree into consideration during IDG.

  - step1: Generate all the candidate links $A_{candi}$ using IDG.
  - step2: Select the one which can generate the biggest frontier set:
    $$ l = \arg\max_{l' \in A_{candi}} (\text{frontierSize}(A \cup l', \text{Tree})) $$
  - step3: Add $l$ to $A$, and repeat step 1, until no new link can be added.

The link leading to the maximum number of frontier nodes is added first.
Intersect Diag Syntactic Grow

IDG

IDSG
Combining TST-FS/FA-PR and IDSG

- Two alternatives of the combination
  - Improve alignment matrix by IDSG and then re-train parser with the better alignment
  - Re-train parser and then improve alignment matrix with better syntactic information
  - Empirically found that only one round of parser re-training before or after only one round of IDSG is already enough
Experiment

• The parser we used is Berkeley parser, with grammar trained on WSJ corpus
• The decoder is an in-house implementation of string-to-tree decoder
• The feature weights are tuned with MERT
• 5-gram language model are trained from the Xinhua section of the Gigaword corpus
Experiment

• Data Setting: IWSLT
  • Training: 81K sentence pairs, 655K Chinese words and 806K English words.
  • Dev: devset8+dialog
  • Test: devset9

• Data Setting: NIST
  • Training: 354K sentence pairs, 8M Chinese words and 10M English words
  • Dev: NIST 2003 data set
  • Test: NIST 2005 and 2008 data set
## Results for Re-Training

<table>
<thead>
<tr>
<th></th>
<th>dev8+dialog</th>
<th>dev9</th>
<th>#rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>50.58</td>
<td>49.85</td>
<td>515K</td>
</tr>
<tr>
<td><strong>SST</strong></td>
<td>52.04</td>
<td>51.26</td>
<td>574K</td>
</tr>
<tr>
<td><strong>FST-FS</strong></td>
<td>52.75</td>
<td>52.51</td>
<td>572K</td>
</tr>
<tr>
<td><strong>FA-PR</strong></td>
<td>53.31</td>
<td>52.8</td>
<td>591K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NIST’03</th>
<th>NIST’05</th>
<th>NIST’08</th>
<th>#rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>37.57</td>
<td>36.44</td>
<td>34.87</td>
<td>3,376K</td>
</tr>
<tr>
<td><strong>SST</strong></td>
<td>37.98</td>
<td>36.79</td>
<td>25.30</td>
<td>3,462K</td>
</tr>
<tr>
<td><strong>FST-FS</strong></td>
<td>38.42</td>
<td>37.39</td>
<td>25.79</td>
<td>3,642K</td>
</tr>
<tr>
<td><strong>FA-PR</strong></td>
<td>38.74</td>
<td>37.69</td>
<td>25.89</td>
<td>3,976K</td>
</tr>
</tbody>
</table>
Result for IDSG

<table>
<thead>
<tr>
<th></th>
<th>dev8+dialog</th>
<th>dev9</th>
<th>#rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDG</td>
<td>50.58</td>
<td>49.85</td>
<td>515K</td>
</tr>
<tr>
<td>IDSG</td>
<td>52.71</td>
<td>51.80</td>
<td>626K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NIST’03</th>
<th>NIST’05</th>
<th>NIST’08</th>
<th>#rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDG</td>
<td>37.57</td>
<td>36.44</td>
<td>34.87</td>
<td>3,376K</td>
</tr>
<tr>
<td>IDSG</td>
<td>38.15</td>
<td>37.07</td>
<td>25.67</td>
<td>4,109K</td>
</tr>
</tbody>
</table>
## Result for Combined

<table>
<thead>
<tr>
<th></th>
<th>dev8+dialog</th>
<th>dev9</th>
<th>#rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>50.58</td>
<td>49.85</td>
<td>515K</td>
</tr>
<tr>
<td><strong>IDSG</strong></td>
<td>52.72</td>
<td>51.8</td>
<td>626K</td>
</tr>
<tr>
<td><strong>FA-PR</strong></td>
<td>53.31</td>
<td>52.8</td>
<td>591K</td>
</tr>
<tr>
<td><strong>IDSG + FA-PR</strong></td>
<td>53.64</td>
<td>53.32</td>
<td>602K</td>
</tr>
<tr>
<td><strong>FA-PR+IDSG</strong></td>
<td>53.81</td>
<td>53.26</td>
<td>597K</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>NIST’03</th>
<th>NIST’05</th>
<th>NIST’08</th>
<th>#rules</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Baseline</strong></td>
<td>37.57</td>
<td>36.44</td>
<td>34.87</td>
<td>3,376K</td>
</tr>
<tr>
<td><strong>IDSG</strong></td>
<td>38.15</td>
<td>37.07</td>
<td>25.67</td>
<td>4,109K</td>
</tr>
<tr>
<td><strong>FA-PR</strong></td>
<td>38.74</td>
<td>37.69</td>
<td>25.89</td>
<td>3,976K</td>
</tr>
<tr>
<td><strong>IDSG + FA-PR</strong></td>
<td>38.97</td>
<td>37.95</td>
<td>26.74</td>
<td>4,557K</td>
</tr>
<tr>
<td><strong>FA-PR+IDSG</strong></td>
<td>38.90</td>
<td>37.94</td>
<td>26.52</td>
<td>4,478K</td>
</tr>
</tbody>
</table>
Conclusion

• Reduce the errors introduced by the mutual independence between monolingual parser and word aligner

• Two retraining methods are proposed
  • Targeted self-training with frontier set based evaluation
  • Forced alignment-based parser re-training

• Alignment matrices are improved by a new symmetrization method

• Our method can improve Syntactic SMT performance significantly
Thanks