Improving Tree-to-Tree Translation with Packed Forests

Yang Liu, Yajuan Lü, and Qun Liu

Institute of Computing Technology,
Chinese Academy of Sciences
{yliu,lvyajuan,liuqun}@ict.ac.cn
Tree-to-Tree Translation

- Tree-to-Tree (e.g., Eisner 2003, Zhang et al., 2008)

- "bushi yu shalong juxing le huitan" -> "Bush held a talk with Sharon"

- Tree-to-Tree structure diagrams for the English and Chinese sentences.
Tree-to-Tree Translation

- Tree-to-Tree (e.g., Eisner 2003, Zhang et al., 2008)

- Bush held a talk with Sharon
Challenges

- Tree-to-tree approaches face two major challenges:
  - most vulnerable to parsing error
  - poorest rule coverage
Parsing Error

both trees can be ill-formed!
Rule Coverage

Bush held a talk with Sharon.
This Work

- We replace 1-best trees with packed forests to alleviate the two problems:
  - parsing error
  - rule coverage
- Our approach outperforms the tree-based system dramatically (+3.6) and achieves comparable performance with Moses.
Packed Forest

Tomita, 1985; Billot and Lang, 1989
Bush held a talk with Sharon

Zhang et al., 2008; Lavie et al., 2008
Forest-based Rule Extraction

Bush held a talk with Sharon
Forest-based Rule Extraction

- Following GHKM (Galley et al., 2004), our extraction method involves three steps:
  - identify the correspondence between nodes
  - identify minimal rules
  - get composed rules
Frontier Nodes

Bush held a talk with Sharon.
Identify Tree Pairs

Bush held a talk with Sharon

Bush:

NP-B

NP

VBD

DT

NN

NP

VP

PP

NP-B

NP

IN

NNP

with

IN

NNP

with

NP

PP

NP-B

NP

yu shalong with Sharon

yu

shalong

with

Sharon

with

Sharon

PP

NP-B

NP

NN

NR

talk

juxing le huitan

le

huitan

juxing

le

huitan

juxing

le

huitan

juxing
Decomposition

\[\text{PP with } \text{Sharon} = \text{NP-B with } \text{Sharon} + \text{IN with } \text{Sharon} + \text{PP with } \text{Sharon}\]
Frontier Trees and Tree Pairs

- frontier tree
- minimal frontier tree
- frontier tree pair
- minimal frontier tree pair
Identify Tree Pairs for the Node PP

step 1: identify corresponding frontier nodes

Bush held a talk with Sharon
Identify Tree Pairs for the Node PP

step 2: identify frontier trees for each node

Bush held a talk with Sharon.
Identify Tree Pairs for the Node PP

step 2: identify frontier trees for each node

Bush held a talk with Sharon
Identify Tree Pairs for the Node PP

step 2: identify frontier trees for each node
Identify Tree Pairs for the Node PP
step 3: build frontier trees pairs
Identify Tree Pairs for the Node PP

step 4: identify minimal frontier trees pairs
Difficulty in Finding Tree Pairs

A minimal frontier tree pair is not necessarily a pair of minimal frontier trees.
Difficulty in Finding Tree Pairs

Bush held a talk with Sharon.
Difficulty in Finding Tree Pairs

minimal frontier tree

non-minimal frontier tree

minimal frontier tree pair
Get Composed Rules

\[ \text{P} \rightarrow \text{IN} \rightarrow \text{PP} \rightarrow \text{NP-B} \]

\[ \text{P} \rightarrow \text{yu} \rightarrow \text{with} \rightarrow \text{IN} \]

\[ \text{P} \rightarrow \text{yu} \rightarrow \text{with} \rightarrow \text{IN} \rightarrow \text{PP} \rightarrow \text{NP-B} \]
Experiments

- Chinese-to-English translation
- Training set: 31K sentence pairs with 840K Chinese words and 950K English words
- Language Model: 3-gram trained on the English side
- Development set: NIST 2002
- Test set: NIST 2005
- Metric: case-insensitive BLEU4
## Tree-based Vs. Forest-based

<table>
<thead>
<tr>
<th>p</th>
<th>avg. trees</th>
<th># of rules</th>
<th>BLEU 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>74K</td>
<td>20.21</td>
</tr>
<tr>
<td>2</td>
<td>238.94</td>
<td>105K</td>
<td>21.65</td>
</tr>
<tr>
<td>5</td>
<td>5.78M</td>
<td>348K</td>
<td>23.36</td>
</tr>
<tr>
<td>8</td>
<td>65.9M</td>
<td>574K</td>
<td>23.73</td>
</tr>
<tr>
<td>10</td>
<td>105M</td>
<td>743K</td>
<td>23.85</td>
</tr>
</tbody>
</table>
## Extraction and Decoding Time

<table>
<thead>
<tr>
<th>p</th>
<th>avg. trees</th>
<th>extraction</th>
<th>decoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1.26</td>
<td>6.76</td>
</tr>
<tr>
<td>2</td>
<td>238.94</td>
<td>2.35</td>
<td>8.52</td>
</tr>
<tr>
<td>5</td>
<td>5.78M</td>
<td>6.34</td>
<td>14.87</td>
</tr>
<tr>
<td>8</td>
<td>65.9M</td>
<td>8.51</td>
<td>19.78</td>
</tr>
<tr>
<td>10</td>
<td>105M</td>
<td>10.21</td>
<td>25.81</td>
</tr>
</tbody>
</table>

extraction: milliseconds / sentence pair
decoding: seconds / sentence
Comparison with Moses

<table>
<thead>
<tr>
<th>training</th>
<th>Moses</th>
<th>this work</th>
</tr>
</thead>
<tbody>
<tr>
<td>840K +950K</td>
<td>23.66</td>
<td>23.85</td>
</tr>
<tr>
<td>7.39M +9.41M</td>
<td>30.43</td>
<td>30.59</td>
</tr>
</tbody>
</table>
Conclusion

- Packed forests help alleviate two problems that tree-to-tree approaches face:
  - negative impact of parsing mistakes on translation quality
  - poor rule coverage
Thanks!