Left-to-Right Target Generation for Hierarchical Phrase-based Translation

Marzieh Razavi
Maryam Siahbani
Ravikiran Vadlapudi
Introduction

• Problem with Hierarchical phrase-based translation: **Cost of Decoding**
  – the number of extracted rules from initial phrases would be numerous
  – The integration with language model especially when incorporating with higher order n-grams:
  – $O\left(n^3 |T|^{4(m-1)}\right)$ (T: English terminal alphabet)
Introduction

• Idea: **Target-normalized** Hierarchical Phrase-based Translation
  
  – Restrict the target side to have GNF-like structure:
    • Sequence of terminals followed by non-terminals
  
  ➡️ Reduce the number of extracted rules from the bilingual corpus

  ➡️ Integration with n-gram language model would be straightforward
    • The target side is generated in Left-to-right order
Translation Model

\[ X \rightarrow < \gamma, \overline{b}\beta, \sim > \]

- \( X \): non-terminal
- \( \gamma \): strings of terminals and non-terminals for source
- \( \overline{b} \): strings of terminals or phrase for target
- \( \beta \): string of non-terminals for target
- \( \sim \): 1-1 correspondence between non-terminals
Rule Extraction

1. Identifying initial phrase pairs

2. Extracting rules:
   I. Use initial phrases \((\overline{f}, \overline{e})\):

   \[ X \rightarrow < \overline{f}, \overline{e} > \]

   II. Choose a rule \(X \rightarrow < \gamma, \alpha >\) and a phrase \((\overline{f}, \overline{e})\) s.t.
   \(\gamma = \gamma' f \gamma''\) and \(\alpha = \overline{e'} \overline{e} \beta\)

   \[ X \rightarrow < \gamma' X \gamma'', \overline{e'} X \beta > \]
The International terrorism also is a possible threat in Japan.
The international terrorism is also a possible threat in Japan.
The International terrorism is also a possible threat in Japan.
The International terrorism is also a possible threat in Japan.
The International terrorism also is a possible threat in Japan.
International terrorism is also a possible threat in Japan.
The International terrorism is also a possible threat in Japan.
The International terrorism is also a possible threat in Japan.

\[ X \rightarrow \langle X \text{ でも起こりうる脅威である} \rangle \]

\[ also \text{ is a possible threat } X \rangle \]

\[ X \rightarrow \langle X \text{ でも } X, \text{ also } X \rangle \]
The International terrorism is also possible threat in Japan.
Phrase-based Rules

• From step 1 of rule extraction: $X \rightarrow < \bar{f}, \bar{e} >$

• We add the following rules to avoid data sparseness:
  • $X \rightarrow < \bar{f}X, \bar{e}X >$
    
  • $X \rightarrow < X\bar{f}, \bar{e}X >$
    
  • $X \rightarrow < X\bar{f}X, \bar{e}XX >$
    
  • $X \rightarrow < X\bar{f}XX, \bar{e}XX >$
Example

The international terrorism also is a possible threat in Japan.
国際テロは日本でも起こりうる脅威である

Rules

Stack

\[ [1, 11] \]

\[ X: [1, 11] \rightarrow \langle X_1: [1, 2] は X_2: [4, 11], The X_1 X_2 \rangle \]
$X : [1, 11] \rightarrow \langle X_1 : [1, 2] \text{ is } X_2 : [4, 11], \text{ The } \underline{X_1} \underline{X_2} \rangle$

$X : [1, 2] \rightarrow \langle \text{国際 } X_1 : [2, 2], \text{ international } \underline{X_1} \rangle$
$X : [1, 11] \rightarrow \langle X_1 : [1, 2] \text{ は } X_2 : [4, 11], \text{ The } X_1 X_2 \rangle$

$X : [1, 2] \rightarrow \langle \text{ 国際 } X_1 : [2, 2], \text{ international } X_1 \rangle$

$X : [2, 2] \rightarrow \langle \text{ テロ, terrorism} \rangle$
Rules

$X : [1, 11] \rightarrow \langle X_1 : [1, 2] \text{ は } X_2 : [4, 11], \text{ The } X_1 X_2 \rangle$

$X : [1, 2] \rightarrow \langle \text{国際 } X_1 : [2, 2], \text{ international } X_1 \rangle$

$X : [2, 2] \rightarrow \langle \text{テロ, terrorism} \rangle$

$X : [4, 11] \rightarrow \langle X_2 : [4, 5] \text{ も } X_1 : [7, 11], \text{ also } X_1 X_2 \rangle$
<table>
<thead>
<tr>
<th></th>
<th>Rules</th>
<th>Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>X: [1, 11]</td>
<td>( \langle X_1: [1, 2] \text{ は } X_2: [4, 11], \text{ The } X_1 X_2 \rangle )</td>
<td></td>
</tr>
<tr>
<td>X: [1, 2]</td>
<td>( \langle \text{ 国際 } X_1: [2, 2], \text{ international } X_1 \rangle )</td>
<td>[2, 2]</td>
</tr>
<tr>
<td></td>
<td>( \langle \text{ テロ, terrorism} \rangle )</td>
<td></td>
</tr>
<tr>
<td>X: [2, 2]</td>
<td>( \langle X_1: [4, 5] \text{ も } X_1: [7, 11], \text{ also } X_1 X_2 \rangle )</td>
<td>[7, 11]</td>
</tr>
<tr>
<td>X: [4, 11]</td>
<td>( \langle X_2: [4, 5] \text{ である, is a } X_1 \rangle )</td>
<td></td>
</tr>
</tbody>
</table>
### Rules

\[ X: [7, 11] \rightarrow \langle X_{\Box}: [7, 9] \text{ である, is a } X_{\Box} \rangle \]

\[ X: [7, 9] \rightarrow \langle \text{起こりうる } X_{\Box}: [9, 9], \text{ possible } X_{\Box} \rangle \]
$X : [7, 11] \rightarrow \langle X_{[1]} : [7, 9] \text{ is a } X_{[1]} \rangle$

$X : [7, 9] \rightarrow \langle \text{起こりうる } X_{[1]} : [9, 9], \text{ possible } X_{[1]} \rangle$

$X : [9, 9] \rightarrow \langle \text{脅威, threat} \rangle$
Rules

$X : [7, 11] \rightarrow \langle X_1 : [7, 9] \text{ である, is a } X_1 \rangle$

$X : [7, 9] \rightarrow \langle \text{起こりうる } X_1 : [9, 9], \text{ possible } X_1 \rangle$

$X : [9, 9] \rightarrow \langle \text{脅威, threat} \rangle$

$X : [4, 5] \rightarrow \langle X_1 : [4, 4] \text{ で, in } X_1 \rangle$
1 2 3 4 5 6 7 8 9 10 11

Rules

$X : [7, 11] \rightarrow \langle X_{11} : [7, 9] である, is a X_{11} \rangle$

$X : [7, 9] \rightarrow \langle 起こりうる X_{11} : [9, 9], possible X_{11} \rangle$

$X : [9, 9] \rightarrow \langle 脅威, threat \rangle$

$X : [4, 5] \rightarrow \langle X_{11} : [4, 4] で, in X_{11} \rangle$

$X : [4, 4] \rightarrow \langle 日本, Japan \rangle$

Stack

$\left[ \begin{array}{c} [7, 9] \\ [4, 5] \end{array} \right]$
$X : [1, 11] \rightarrow \langle X_{11} : [1, 2] \text{ is } X_{4 \text{ }} : [4, 11], \text{ The } X_{11} X_{2} \rangle$

$X : [1, 2] \rightarrow \langle \text{international } X_{1}, [2, 2] \rangle$

$X : [2, 2] \rightarrow \langle \text{terrorism} \rangle$

$X : [4, 11] \rightarrow \langle X_{4} : [4, 5] \text{ also } X_{11} : [7, 11], \text{ also } X_{4} X_{2} \rangle$

$X : [7, 11] \rightarrow \langle \text{is a } X_{7}, [7, 9] \rangle$

$X : [7, 9] \rightarrow \langle \text{possible } X_{7}, [9, 9] \rangle$

$X : [9, 9] \rightarrow \langle \text{threat} \rangle$

$X : [4, 5] \rightarrow \langle \text{in } X_{4}, [4, 4] \rangle$

$X : [4, 4] \rightarrow \langle \text{Japan} \rangle$
Example of derivation tree
Model

• General log-linear model over derivations $D$

$$
\arg\max_{e_1^I} \frac{\exp(\lambda_{lm} p_{lm}(e_1^I) + \sum_{m=1, m \neq lm}^M \lambda_m h_m(e_1^I | f_1^J, D))}{\sum_{e_1^{I'}} \exp(\lambda_{lm} p_{lm}(e_1^{I'}) + \sum_{m=1, m \neq lm}^M \lambda_m h_m(e_1^{I'} | f_1^J, D))}
$$

- $h_m(e_1^I | f_1^J, D)$: feature functions
- $\lambda_m$: weights
Feature functions

• Likelihood of two sentences $f_1^J$ and $e_1^I$
  
  $h_\phi(e_1^I|f_1^J, D) \cdot h_\phi(f_1^J|e_1^I, D)$

  $h_\phi(f_1^J|e_1^I, D) = \log \prod_{(\gamma, \alpha) \in D} \phi(\gamma|\alpha)$

• Lexical weights: $h_w(e_1^I|f_1^J, D), h_w(f_1^J|e_1^I, D)$
  
  – How well the words in $e_1^I$ translate the words in $f_1^J$

• Language model
Reordering features

- Two features to limit the reordering
Reordering features

\[ h_n(e_1^I, f_1^I, \mathcal{D}) = \sum_{\mathcal{D}_i \in \text{back} \,(\mathcal{D})} \text{height}(\mathcal{D}_i) \]

\[ h_w(e_1^I, f_1^J, \mathcal{D}) = \sum_{\mathcal{D}_i \in \text{back} \,(\mathcal{D})} \text{width}(\mathcal{D}_i) \]
Length-based features

• Three length-based features:
  – Length of English sentence: \( h_l(e_1^I) = I \)
    • Control the length of output
  – No. of extracted rules: \( h_r(D) = rule(D) \)
  – No. of phrase-based rules: \( h_p(D) = phrase(D) \)
    • Control whether to choose a extracted rule of phrase-based rules in D
Experiments

<table>
<thead>
<tr>
<th></th>
<th># rules/phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase</td>
<td>5,433,091</td>
</tr>
<tr>
<td>Normalized-2</td>
<td>6,225,630</td>
</tr>
<tr>
<td>Normalized-3</td>
<td>6,233,294</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>12,824,387</td>
</tr>
</tbody>
</table>

- No. of Rules with 2 non-terminals are slightly larger than phrase pairs
- Including 3 non-terminals did not change the grammar size
Experiments

- At most two non-terminals
- 3-gram/5-gram language models

<table>
<thead>
<tr>
<th></th>
<th>BLEU [%]</th>
<th>NIST</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phrase</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-gram</td>
<td>7.14</td>
<td>3.21</td>
</tr>
<tr>
<td>5-gram</td>
<td>7.33</td>
<td>3.19</td>
</tr>
<tr>
<td><strong>Normalized-2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-gram</td>
<td>10.00</td>
<td>4.11</td>
</tr>
<tr>
<td>5-gram</td>
<td>10.26</td>
<td>4.20</td>
</tr>
</tbody>
</table>

Results for Japanese-English
Questions ??