NTT Statistical Machine Translation for IWSLT 2006

Taro Watanabe, Jun Suzuki, Hajime Tsukada and Hideki Isozaki

NTT Communication Science Labs.
{taro,jun,tsukada,isozaki}@cslab.kecl.ntt.co.jp
Overview

- Hierarchical Phrase-based SMT achieved by:
  - Target normalized form
  - Earley-style top-down parsing
  - Reranking via voted perceptron

Diagram:

- Foreign text → Hierarchical Phrase-based SMT
- n-best list → Reranking
- English text
Hierarchical Phrase-based SMT

X1

X2
は

X3
国際

X9
テロ

X6
である

X7
脅威

X8
も

X4
international

X5
terrorism is a threat in Japan

X1
The

X2
also

X3
in

X4
possible

X5
Japan

X6
possible
Simplified Grammar

- Target normalized form
- Phrase-prefixed structure for target-side (GNF-like structure)
- Arbitrary structure for source-side
- Constrained rule-extraction

\[
\begin{align*}
X_1 & \text{ 份子} \quad \mid\mid\mid \quad \text{members of } X_1 \\
\text{定将 } X_1 \quad \mid\mid\mid \quad \text{decided to } X_1 \\
X_1 \text{ 是 } X_2 \quad \mid\mid\mid \quad \text{is } X_1 X_2 \\
X_1 \text{ 不再 } X_2 \quad \mid\mid\mid \quad \text{no longer } X_1 X_2 \\
X_2 \text{ 一个最 } X_1 \quad \mid\mid\mid \quad \text{One of the most } X_1 X_2 \\
X_2 \text{ 的 } X_1 \text{ 表示 哀悼} \quad \mid\mid\mid \quad \text{condolences to the } X_1 X_2
\end{align*}
\]
Decoding by Top-down parsing

- Earley-style parsing on source-side
- Straight-forward intersection with ngram
- Similar to a phrase-based decoding algorithm
Log-linear Approach

\[ \hat{e}^I_1 = \arg \max_{e^I_1} \frac{\exp \left( \sum_{m=1}^{M} \lambda_m h_m(e^I_1, f^J_1) \right)}{\sum_{e'^I_1} \exp \left( \sum_{m=1}^{M} \lambda_m h_m(e'^I_1, f^J_1) \right)} \]

- Mixed-case 5-gram
- Rule translation probabilities
- Lexical weights
- Insertion/deletion penalties
- Backtrack penalties
- # of words/# of rules
Reranking by Voted Perceptron

• Ranking Voted Perceptron with BLEU-based updates

• Features
  • SC: Scores from the baseline decoder
  • AL: Word-pairs from IBM Model Viterbi alignment
  • RU: Rules & Rule pattern

Please write down your address here
Reranking Algorithm

\[ D = \{ D^1, \ldots, D^M \} \]: Development set  
\[ C^m = \{ c^m_1, \ldots, c^m_N \} \]: The original N-best list of \( D^m \)  
\[ X^m = \{ x^m_1, \ldots, x^m_N \} \]: (reordered) N-best list of \( D^m \)  

\[ \text{Ranking}(W, C^m) \]: returns N-best list of \( C^m \) reordered based on the score, \( s^m_n = \langle W, \phi(c^m_n) \rangle \)

\[
\text{for } t = 1, \ldots, T \text{ do } \\
\text{  for } m = 1, \ldots, M \text{ do } \\
\text{    } X^m \leftarrow \text{Ranking}(W, C^m) \\
\text{    for } i = 1, \ldots, |X^m| \text{ do } \\
\text{      for } j = i + 1, \ldots, |X^m| \text{ do } \\
\text{        if } \text{BLEU}(x^m_j) > \text{BLEU}(x^m_i) \& \text{WER}(x^m_j) \leq \text{WER}(x^m_i) \text{ then } \\
\text{          } W = W + (\text{BLEU}(x^m_j) - \text{BLEU}(x^m_i)) \times (\phi(x^m_j) - \phi(x^m_i)) \\
\text{        end if } \\
\text{      end for } \\
\text{    end for } \\
\text{  end for } \\
\text{  } V_t = W \\
\text{end for } \\
\text{end for }
\]

Update all incorrect ranking pair through pair-wise comparison
Approximated BLEU

- Very frequent updates required:
  - Computation of doc-set BLEU is impossible
- Sentence-wise BLEU?
  - Bad objective: 27.78 to 25.95 in MTEval 2006
- Approximated BLEU:
  - doc-set BLEU of 1-best
  - Compute difference for each segment
Tasks

• ASR’s 1-best translation
• Case-restoration/punctuation-insertion required
• Preprocessing:
  • Case/punctuation-preserved English-side + lower-cased/punctuation-removed source-side
  • Induce multiple alignments from differently preprocessed corpora (punct-removed, etc.)
  • Aggregate rules from differently aligned corpora
## Official Results

<table>
<thead>
<tr>
<th>Language Pair</th>
<th>Type</th>
<th>BLEU</th>
<th>NIST</th>
<th>METEOR</th>
<th>mPER</th>
<th>mWER</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar-en</td>
<td>spoken</td>
<td>20.71 (5th)</td>
<td>4.84</td>
<td>43.97</td>
<td>64.67</td>
<td>56.65</td>
</tr>
<tr>
<td></td>
<td>text</td>
<td>22.65 (5th)</td>
<td>5.33</td>
<td>47.76</td>
<td>62.79</td>
<td>54.15</td>
</tr>
<tr>
<td>it-en</td>
<td>spoken</td>
<td>27.69 (7th)</td>
<td>6.70</td>
<td>56.07</td>
<td>57.00</td>
<td>48.13</td>
</tr>
<tr>
<td></td>
<td>text</td>
<td>34.49 (5th)</td>
<td>7.83</td>
<td>64.31</td>
<td>50.79</td>
<td>41.57</td>
</tr>
<tr>
<td>ja-en</td>
<td>spoken</td>
<td>19.84 (2nd)</td>
<td>5.48</td>
<td>45.00</td>
<td>71.08</td>
<td>55.12</td>
</tr>
<tr>
<td></td>
<td>text</td>
<td>22.03 (2nd)</td>
<td>5.91</td>
<td>48.77</td>
<td>69.02</td>
<td>52.17</td>
</tr>
<tr>
<td>zh-en</td>
<td>spontaneous</td>
<td>15.59 (6th)</td>
<td>4.18</td>
<td>39.46</td>
<td>70.20</td>
<td>59.72</td>
</tr>
<tr>
<td></td>
<td>spoken</td>
<td>18.34 (5th)</td>
<td>4.53</td>
<td>42.15</td>
<td>68.44</td>
<td>57.71</td>
</tr>
<tr>
<td></td>
<td>text</td>
<td>21.35 (5th)</td>
<td>5.13</td>
<td>47.43</td>
<td>65.47</td>
<td>53.70</td>
</tr>
</tbody>
</table>

Remarks: reranked with only SC features
Results on Hierarchical Phrase-based SMT

- Phrase (spontaneous)
- Phrase (spoken)
- Phrase (text)
- Rule (spontaneous)
- Rule (spoken)
- Rule (text)

**Languages**:
- ar-en
- it-en
- ja-en
- zh-en
Results on Reranking

- l-best (spontaneous)
- l-best (spoken)
- l-best (text)
- SC (spontaneous)
- SC (spoken)
- SC (text)
- ALL (spontaneous)
- ALL (spoken)
- ALL (text)

Chart showing performance across different languages and conditions.
Conclusion

• Better than non-hierarchical translation
• Benefit from sparse features (RU, AL) in reranking
• Hierarchical Phrase-based SMT as a baseline
  • Target normalized form + top-down parsing
• Reranking by Voted Perceptron
  • BLEU-based updates + Approximated-BLEU