Hybrid Machine Translation

Coupling Systems

Daniel Nagel

ISCL 2018
Recap
Vauquois pyramid

Diagram showing the process of translation through an intermediate language (Interlingua) and the methods of analysis, transfer, and generation.
Vauquois pyramid
SMT: Statistical machine translation
RBMT: Rule-based machine translation
SMT: Statistical machine translation
RBMT: Rule-based machine translation
Rule-based machine translation is machine translation systems based on linguistic information about source and target languages basically retrieved from dictionaries and grammars covering the main semantic, morphological, and syntactic regularities of each language respectively.
Input Sentence – English Language

Stanford Parser to Generate POS

Sentence Simplifier

Stanford Parser

Reordering

Check for word Availability in Bilingual DB

If Found?

Y

Bilingual Corpus

N

Transliterate using Unicode DB

Translate input data from Data Corpus

Morphology Generator (Case Marker, Tense Marker, PNG Marker)

Translated target sentence
SMT: Statistical machine translation
RBMT: Rule-based machine translation
Statistical machine translation is a machine translation paradigm where translations are generated on the basis of statistical models whose parameters are derived from the analysis of bilingual text corpora.
Source Language Text

Preprocessing

Global search

\[ e^* = \text{argmax}_e p(e | f) \]

Language model 
\[ p(e) \]

Translation model 
\[ p(f | e) \]

Preprocessing

Target Language Text
...and of course...
Hybrid machine translation systems
Main types of architectures
Main types of architectures

- coupling of systems
  - serial or parallel
Main types of architectures

- coupling of systems
  serial or parallel
- architecture adaptations
  integrating novel components into SMT or RMT architectures, either by pre/post-editing, or by system core modifications
Main types of architectures

- coupling of systems
  serial or parallel
- architecture adaptations
  integrating novel components into SMT or RMT architectures, either by pre/post-editing, or by system core modifications
- genuine hybrid systems
  combining components of different paradigms
Main types of hybrid machine translation architectures

- coupling of systems
  - serial or parallel
- architecture adaptations
  - integrating novel components into SMT or RMT architectures, either by pre/post-editing, or by system core modifications
- genuine hybrid systems
  - combining components of different paradigms
Coupling of systems

Serial coupling
Coupling - Serial Coupling

- Serial Coupling:
  - SMT + RBMT: Syntactic Selection
  - RBMT + SMT: Statistical Post-Editing
Coupling - Serial Coupling

“Hybrid Systems”

1) Syntactic selection
   Source Text → SMT-engine(s) → Hypotheses → Selection → Target Text

2) Stochastic selection
   Source Text → Rule-based MT engines → Hypotheses → Selection → Target Text

3) SMT feeds rule-based MT
   Source Text → Rule-based MT engine → MT Lexicon → Phrase-Table → Manual Validation → Linguistic Processing → Parallel Corpus

4) SMT has the last word
   Parallel Corpus → Alignment, Phrase Extraction → Phrase-Table
   Hypotheses → Rule-based MT engine → SMT Decoder → Target Text

5) SMT corrects RBMT output
   Parallel Corpus → Rule-based MT engine → Source Text
   Hypotheses → SMT Decoder → Target Text

6) Rule-based transfer architecture interleaved with stochastic ranking
   Source Text → Rule-based analysis → Stochastic ranking → Rule-based transfer → Stochastic ranking → Rule-based generation → Stochastic ranking → Target Text
Coupling - Serial Coupling “Hybrid Systems”

1) Syntactic selection

2) Stochastic selection

3) SMT feeds rule-based MT

4) SMT has the last word

5) SMT corrects RBMT output

6) Rule-based transfer architecture interleaved with stochastic ranking
Coupling - Serial Coupling
Hybrid Systems (Eisele 2008)
Coupling - Serial Coupling
Hybrid Systems (Eisele 2008)

• Merging multiple MT results via a SMT decoder
Coupling - Serial Coupling Hybrid Systems (Eisele 2008)

- Merging multiple MT results via a SMT decoder
- Feeding SMT phrases into a rule-based MT system
Merging multiple MT results via SMT decoder

1) Syntactic selection
   Source Text → SMT-engine(s) → Hypotheses → Selection → Target Text

2) Stochastic selection
   Source Text → Rule-based MT engines → Hypotheses → Selection → Target Text

3) SMT feeds rule-based MT
   Source Text → MT Lexicon → Manual Validation → Linguistic Processing → Parallel Corpus

4) SMT has the last word
   Parallel Corpus → Alignment, Phrase Extr. → Phrase-Table → SMT-Model → Target Text

5) SMT corrects RBMT output
   Source Text → Parallel Corpus → Rule-based MT engine → Hypotheses → SMT Decoder → Target Text

6) Rule-based transfer architecture interleaved with stochastic ranking
   Source Text → Rule-based analysis → Stochastic ranking → Rule-based transfer → Stochastic ranking → Rule-based generation → Stochastic ranking → Target Text
Merging multiple MT results via a SMT decoder
giza++

And
the
programme
has
been
implemented

Le
programme
a
été
mis
en
application
Merging multiple MT results via a SMT decoder
Merging multiple MT results via a SMT decoder

In-domain

Out-of-domain
Merging multiple MT results via a SMT decoder

<table>
<thead>
<tr>
<th></th>
<th>Europarl</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>NewsCommentary</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hybrid</td>
<td>27.85</td>
<td>20.75</td>
<td>28.12</td>
<td>28.82</td>
<td>33.15</td>
<td>32.31</td>
<td>17.36</td>
<td>13.57</td>
<td>17.66</td>
<td>20.71</td>
</tr>
<tr>
<td>RBMT1*</td>
<td>13.34</td>
<td>11.09</td>
<td></td>
<td>17.19</td>
<td></td>
<td>18.63</td>
<td>14.90</td>
<td>12.34</td>
<td></td>
<td>15.11</td>
</tr>
<tr>
<td>RBMT2</td>
<td>16.19</td>
<td>12.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16.66</td>
<td>13.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RBMT3</td>
<td>16.32</td>
<td>10.88</td>
<td>18.18</td>
<td>20.38</td>
<td>19.32</td>
<td>20.89</td>
<td>16.88</td>
<td>12.53</td>
<td>17.20</td>
<td>18.82</td>
</tr>
<tr>
<td>RBMT4</td>
<td>15.58</td>
<td>12.09</td>
<td>19.00</td>
<td>22.20</td>
<td>18.99</td>
<td>21.69</td>
<td>17.41</td>
<td>13.93</td>
<td>17.73</td>
<td>20.85</td>
</tr>
<tr>
<td>RBMT5</td>
<td>15.58</td>
<td>9.54</td>
<td>21.36</td>
<td>12.98</td>
<td>18.47</td>
<td>20.59</td>
<td>15.99</td>
<td>11.05</td>
<td></td>
<td>18.65</td>
</tr>
<tr>
<td>RBMT6</td>
<td>13.96</td>
<td>9.44</td>
<td>17.16</td>
<td>18.91</td>
<td>18.01</td>
<td>19.18</td>
<td>15.08</td>
<td>10.41</td>
<td>16.86</td>
<td>17.82</td>
</tr>
</tbody>
</table>

**Table 1:** Performance of baseline SMT system, our system and RBMT systems (BLEU scores)
<table>
<thead>
<tr>
<th>Source</th>
<th>Darüber hinaus gibt es je zwei Feiertage zu Ostern, Pfingsten, und Weihnachten.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>In addition, Easter, Pentecost, and Christmas are each two-day holidays.</td>
</tr>
<tr>
<td>Moses</td>
<td>In addition, there are two holidays, pfingsten to Easter, and Christmas.</td>
</tr>
<tr>
<td>Hybrid</td>
<td>In addition there are the two holidays to Easter, Pentecost and Christmas.</td>
</tr>
<tr>
<td>RBMT1</td>
<td>Furthermore there are two holidays to Easter, Pentecost and Christmas.</td>
</tr>
<tr>
<td>RBMT2</td>
<td>Furthermore there are two holidays each at Easter, Pentecost and Christmas.</td>
</tr>
<tr>
<td>RBMT3</td>
<td>In addition there are each two holidays to Easters, Whitsun, and Christmas.</td>
</tr>
<tr>
<td>RBMT4</td>
<td>In addition, there is two holidays to Easter, Pentecost, and Christmas.</td>
</tr>
<tr>
<td>RBMT5</td>
<td>Beyond that there are ever two holidays to Easter, Whitsuntide, and Christmas.</td>
</tr>
<tr>
<td>RBMT6</td>
<td>In addition it gives two holidays apiece to easter, Pentecost, and Christmas.</td>
</tr>
</tbody>
</table>
Feeding SMT phrases into a rule-based MT system

1) Syntactic selection
   - Source Text
   - SMT-engine(s)
   - Hypotheses
   - Selection
   - Target Text

2) Stochastic selection
   - Source Text
   - Rule-based MT engines
   - Hypotheses
   - Target Text

3) SMT feeds rule-based MT
   - Source Text
   - MT-Lexicon
   - Manual Validation
   - Linguistic Processing
   - Target Text

4) SMT has the last word
   - Parallel Corpus
   - Alignment, Phrase Extraction
   - Phrase-Table
   - Rule-based MT engines
   - Hypotheses
   - Dyn. PT
   - SMT Decoder
   - Target Text

5) SMT corrects RBMT output
   - Parallel Corpus
   - Rule-based MT engine
   - Hypotheses
   - SMT Decoder
   - Target Text

Transfer architecture interleaved with stochastic ranking

- Source Text
- Rule-based analysis
- Stochastic ranking
- Rule-based transfer
- Stochastic ranking
- Rule-based generation
- Stochastic ranking
- Target Text
Feeding SMT phrases into a rule-based MT system
Feeding SMT phrases into a rule-based MT system
Feeding SMT phrases into a rule-based MT system
Coupling of systems

Parallel coupling
Coupling - Parallel Coupling (Ren)

- Parallel Coupling:
  MT1, ..., MTn -> select best output
  Works on full sentences or smaller segments
Coupling - Parallel Coupling
Coupling - Parallel Coupling

- Use parallel processing to attack ambiguities to ensure the correct translation is preserved in all the procedures.
Coupling - Parallel Coupling

- Use parallel processing to attack ambiguities to ensure the correct translation is preserved in all the procedures.
- Use a coordination mechanism to combine the candidate translations such that the combined result is better than any of the candidate translations.
Coupling - Parallel Coupling
Coupling - Parallel Coupling
Coupling - Parallel Coupling? (Callison-Burch)
Coupling - Parallel Coupling?

\[ P(\text{I like snakes that are not poisonous}) \sim \]
\[ b(I \mid \text{start-of-sentence start-of-sentence}) \ast \]
\[ b(\text{like} \mid \text{start-of-sentence I}) \ast \]
\[ b(\text{snakes} \mid \text{I like}) \ast \]
\[ \ldots \]
\[ b(\text{poisonous} \mid \text{are not}) \ast \]
\[ b(\text{end-of-sentence} \mid \text{not poisonous}) \ast \]
\[ b(\text{end-of-sentence} \mid \text{poisonous end-of-sentence}) \]
Instead of

\[
b(z \mid x \ y) = \frac{\text{number-of-occurrences(“xyz”)}}{\text{number-of-occurrences(“xy”)}}
\]

it uses

\[
b(z \mid x \ y) = 0.80 \times \frac{\text{number-of-occurrences(“xyz”)}}{\text{number-of-occurrences(“xy”)}} + 0.14 \times \frac{\text{number-of-occurrences(“yz”)}}{\text{number-of-occurrences(“z”)}} + 0.099 \times \frac{\text{number-of-occurrences(“z”)}}{\text{total-words-seen} + 0.001}
\]
Coupling - Parallel Coupling?

1: *Almost Perfect* – the sentence is a fluent English sentence. It seems like it was written by a native speaker.

2: *Understandable* – the sentence is understandable but may have (slightly) strange word choice, or contain some minor grammatical errors, such as an incorrect preposition or determiner.

3: *Barely Understandable* – the sentence contains several grammar and/or vocabulary errors and can only be understood with great effort on the part of the reader.

4: *Incomprehensible* – the meaning of the sentence cannot be derived.
### Coupling - Parallel Coupling?

<table>
<thead>
<tr>
<th>Multi-engine Tool</th>
<th>All (100 sets)</th>
<th>At least Barely Understandable (94)</th>
<th>Understandable (86)</th>
<th>Almost Perfect (57)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engine A</td>
<td>58%</td>
<td>55%</td>
<td>54%</td>
<td>61%</td>
</tr>
<tr>
<td>Engine B (baseline)</td>
<td>70%</td>
<td>68%</td>
<td>69%</td>
<td>66%</td>
</tr>
<tr>
<td>Engine C</td>
<td>27%</td>
<td>22%</td>
<td>21%</td>
<td>19%</td>
</tr>
<tr>
<td>Engine D</td>
<td>40%</td>
<td>36%</td>
<td>35%</td>
<td>39%</td>
</tr>
</tbody>
</table>

**Table 1: Japanese → English chat translations**

<table>
<thead>
<tr>
<th>Multi-engine Tool</th>
<th>All (154 sets)</th>
<th>At least Barely Understandable (146)</th>
<th>Understandable (118)</th>
<th>Almost Perfect (38)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engine E (baseline)</td>
<td>76%</td>
<td>75%</td>
<td>70%</td>
<td>68%</td>
</tr>
<tr>
<td>Engine F</td>
<td>58%</td>
<td>56%</td>
<td>52%</td>
<td>45%</td>
</tr>
</tbody>
</table>

**Table 2: French → English web page translations**

<table>
<thead>
<tr>
<th>Multi-engine Tool</th>
<th>All (51 sets)</th>
<th>At least Barely Understandable (50)</th>
<th>Understandable (44)</th>
<th>Almost Perfect (34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engine E</td>
<td>53%</td>
<td>52%</td>
<td>48%</td>
<td>47%</td>
</tr>
<tr>
<td>Engine F</td>
<td>49%</td>
<td>48%</td>
<td>41%</td>
<td>47%</td>
</tr>
<tr>
<td>Engine G</td>
<td>45%</td>
<td>44%</td>
<td>36%</td>
<td>32%</td>
</tr>
<tr>
<td>Engine H</td>
<td>51%</td>
<td>50%</td>
<td>45%</td>
<td>44%</td>
</tr>
<tr>
<td><strong>Engine I</strong></td>
<td><strong>63%</strong></td>
<td><strong>62%</strong></td>
<td><strong>57%</strong></td>
<td><strong>62%</strong></td>
</tr>
</tbody>
</table>
References

• Andreas Eisele et al.: Hybrid Architectures for Multi-Engine Machine Translation, 2008
• Andreas Eisele et al.: Hybrid Machine Translation Architectures within and beyond the EuroMatrix project, 2008
• Andreas Eisele et al.: Using Moses to Integrate Multiple Rule-Based Machine Translation Engines into a Hybrid System, 2008
• Chris Callison-Burch, Raymond S. Flournoy: A Program for Automatically Selecting the Best Output from Multiple Machine Translation Engines, 2001
• Fuji Ren, Hongchi Shi: Parallel Machine Translation: Principles and Practice, 2001