Multi-Pass Decoding With Complex Feature Guidance for Statistical Machine Translation

Benjamin Marie \textsuperscript{1,2} \& Aurélien Max \textsuperscript{1,3}

(1) LIMSI-CNRS, (2) Lingua et Machina, (3) Université Paris-Sud

**Highlights**

Using complex features in SMT

- some complex features are not easily integrable during decoding:
  - need of a complete hypothesis (e.g. syntactic features)
  - computational cost (e.g. neural network models)
  - not available during a first decoding (e.g. word posteriors)
- such features are usually used through single-pass reranking

**Main ideas**

- exploit a reranking pass result during decoding
- isolate in separate translation tables the possibly misused bi-phrases to better optimize their feature weights

**IN and out translation tables**

- **IN**: contains bi-phrases of the reranking 1-best not in the decoder 1-best
- **OUT**: contains bi-phrases of the decoder 1-best not in the reranking 1-best
- extracted bi-phrases are removed from the original translation table
- IN or OUT (or both) are added to the next decoding pass

**Experimental settings**

- Moses tuned with kb-mira
- reranking system (Rerank) trained with kb-mira using the Moses 1,000-best translation table
- complex features used during reranking:
  - MosesNorm
  - NeuralNet
  - POSTag
  - POSTagRatio
  - Syntax
  - IBM1

**Conclusion & future work**

**Conclusion**

- strong and consistent improvements for all configurations
- simple criterion to iteratively partition the translation table
- makes a better use of complex features than single-pass reranking

**Future work**

- add features to the new translation tables to use more informations about the reranking result during decoding
- use the result of a rewriting system (Marie and Max, 2014) to guide the decoder
- enhance the diversity in the \(n\)-best list (Chatterjee and Canciedda, 2010; Gimpel et al., 2013) to train a better reranking system
- add more complex features to the reranking pass

**Multi-pass decoding results for all configurations**

<table>
<thead>
<tr>
<th>Configuration</th>
<th>medical</th>
<th>medical</th>
<th>medical</th>
<th>news</th>
<th>news</th>
</tr>
</thead>
<tbody>
<tr>
<td>test (\rightarrow) Fr</td>
<td>test (\rightarrow) Fr</td>
<td>test (\rightarrow) En</td>
<td>test (\rightarrow) En</td>
<td>test (\rightarrow) En</td>
<td>test (\rightarrow) En</td>
</tr>
<tr>
<td>Moses</td>
<td>38.8</td>
<td>37.1</td>
<td>31.1</td>
<td>28.6</td>
<td></td>
</tr>
<tr>
<td>Rerank</td>
<td>38.9</td>
<td>37.2</td>
<td>31.1</td>
<td>28.7</td>
<td></td>
</tr>
<tr>
<td>OUT</td>
<td>41.8</td>
<td>40.5</td>
<td>3</td>
<td>31.8</td>
<td></td>
</tr>
<tr>
<td>IN</td>
<td>43.2</td>
<td>41.0</td>
<td>3</td>
<td>32.4</td>
<td></td>
</tr>
<tr>
<td>IN and OUT</td>
<td>42.4</td>
<td>40.6</td>
<td>3</td>
<td>32.1</td>
<td></td>
</tr>
</tbody>
</table>

**Increasing of the \(n\)-best quality**

- analysis for the IN configuration (medical En \(\rightarrow\) Fr):

- quickly reduces the gap between Moses and Rerank BLEU scores
- 1,000-best average BLEU increases over the iterations