The MLLP-UPV Spanish-Portuguese and Portuguese-Spanish Machine Translation Systems for WMT19 Similar Language Translation Task

Pau Baquero-Arnal  Javier Iranzo-Sánchez  Jorge Civera  Alfons Juan
www.mllp.upv.es

INTRODUCTION

• Neural Machine Translation (NMT) system created for the WMT19 Similar Language Translation shared task (ES→PT)
• Standard NMT approach to similar language translation
• 2 NMT architectures explored:
  – Multi-headed self-attention (Transformer)
  – 2D Alternating RNN
• Domain adaptation carried out using fine-tuning

2D ALTERNATING RNN

- First grid: concatenate combinations of all source/target positions
- Block: two recurrent layers, one for each direction
- Output: concatenation of each layer
- Context vector from a row of vectors with attention mechanism

FINE-TUNING

• Significant domain mismatch between train and test data
• Fine-tuning (after training converges) on a subset of dev data

<table>
<thead>
<tr>
<th>System</th>
<th>Portuguese → Spanish test</th>
<th>test-hidden</th>
<th>Spanish → Portuguese test</th>
<th>test-hidden</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer</td>
<td>57.4</td>
<td>51.9</td>
<td>51.3</td>
<td>45.5</td>
</tr>
<tr>
<td>+ fine-tuned</td>
<td>72.4</td>
<td>66.6</td>
<td>70.7</td>
<td>64.7</td>
</tr>
<tr>
<td>2D altern. RNN</td>
<td>55.1</td>
<td>49.7</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>+ fine-tuned</td>
<td>64.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- First fine-tuning epochs are the most useful. After that, we get diminishing returns until the BLEU curve flattens.

SYSTEM EVALUATION

<table>
<thead>
<tr>
<th>Team</th>
<th>Portuguese → Spanish BLEU TER</th>
<th>Spanish → Portuguese</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLLP</td>
<td>66.6 19.7</td>
<td></td>
</tr>
<tr>
<td>NICT</td>
<td>59.9 25.3</td>
<td></td>
</tr>
<tr>
<td>U. Helsinki</td>
<td>58.4 25.3</td>
<td></td>
</tr>
<tr>
<td>Kyoto U.</td>
<td>56.9 26.9</td>
<td></td>
</tr>
<tr>
<td>BSC</td>
<td>54.8 29.8</td>
<td></td>
</tr>
<tr>
<td>UBC-NLP</td>
<td>52.3 32.9</td>
<td></td>
</tr>
<tr>
<td>MLLP-2D</td>
<td>49.7 32.1</td>
<td></td>
</tr>
<tr>
<td>UPC-TALP</td>
<td>62.1 23.0</td>
<td>MLLP</td>
</tr>
<tr>
<td>NICT</td>
<td>53.3 29.1</td>
<td>UPCTALP</td>
</tr>
<tr>
<td>U. Helsinki</td>
<td>52.0 29.4</td>
<td>BSC</td>
</tr>
<tr>
<td>UBC-NLP</td>
<td>46.1 36.0</td>
<td></td>
</tr>
<tr>
<td>BSC</td>
<td>44.0 37.5</td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSIONS

• Generalist approach to similar language translation
• For this task, fine-tuning with in-domain data was critical
• We introduced a novel NMT architecture still in development to test against other participants

Acknowledgments

The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement no. 761758 (X5gon); the Government of Spain’s research project Multisub, ref. RTI2018-094879-B-I00 (MCIU/AEI/FEDER, EU); and the Universitat Politècnica de València’s PAID-01-17 R&D support programme.