Introduction

- **MNMT** aims to handle the translation from multiple source languages into multiple target languages with singled unified model (Johnson et al., 2017; Fan et al., 2021).
- **Back Translation (BT)** (Sennrich et al., 2016; Caswell et al., 2019) is a simple and effective data augmentation technique, which makes use of monolingual corpora and has proven to be effective for MNMT (Zhang et al., 2020).
- **Knowledge Distillation (KD)** (Kim and Rush, 2016) is a commonly used technique to improve model performance. We follow a recent approach to KD (Wang et al., 2021), which uses selection at the batch level and global level to choose suitable samples for distillation.

Data

- Moses truercased and punctuation normalized.
- Sentences containing 50% punctuation are removed.
- Duplicate sentences are removed.
- Using 1angid filter out sentences with mixed language.
- SentencePiece was used to produce subword units.
- Remove sentences with more than 250 subwords.

Analysis

- Our multilingual model (#4) performs competitively with the Flores strong baseline (Model #0).
- Although the overall performance of the Multilingual model (#4) is better than the Bilingual model (#2), back-translation using the Bilingual model (model #6) is better than back-translation using the Multilingual model (model #5).
- Knowledge Distillation further improves performance slightly (Model # 7 and Model # 8).

Results

<table>
<thead>
<tr>
<th># Systems</th>
<th>spBLEU</th>
<th>chrF</th>
<th>BERTScore</th>
<th>BEST BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Flores</td>
<td>28.0</td>
<td>0.528</td>
<td>0.867</td>
<td>sr-mk (36.0)</td>
</tr>
<tr>
<td>1 Bilingual</td>
<td>21.1</td>
<td>0.477</td>
<td>0.831</td>
<td>en-mk (31.3)</td>
</tr>
<tr>
<td>2 Bilingual</td>
<td>28.4</td>
<td>0.533</td>
<td>0.863</td>
<td>sr-en (40.6)</td>
</tr>
<tr>
<td>3 Multilingual</td>
<td>16.7</td>
<td>0.431</td>
<td>0.827</td>
<td>sr-en (26.1)</td>
</tr>
<tr>
<td>4 Multilingual</td>
<td>30.9</td>
<td>0.555</td>
<td>0.874</td>
<td>sr-en (40.0)</td>
</tr>
<tr>
<td>5 Multilingual + TaggedBT (Multilingual)</td>
<td>30.7</td>
<td>0.548</td>
<td>0.873</td>
<td>sr-en (40.5)</td>
</tr>
<tr>
<td>6 Multilingual + TaggedBT (Bilingual)</td>
<td>32.3</td>
<td>0.562</td>
<td>0.879</td>
<td>sr-en (41.5)</td>
</tr>
<tr>
<td>7* Multilingual + TaggedBT (Bilingual) + KD</td>
<td>32.2</td>
<td>0.572</td>
<td>0.883</td>
<td>sr-en (42.0)</td>
</tr>
<tr>
<td>8* Multilingual + TaggedBT (Bilingual) + KD</td>
<td>33.9</td>
<td>0.576</td>
<td>0.887</td>
<td>sr-en (42.4)</td>
</tr>
</tbody>
</table>

Conclusion

- Our goal is to investigate the impact of bilingual systems on multilingual systems.
- Our best translation system scores 5 to 6 BLEU higher than a strong baseline system provided by the organizers (Goyal et al., 2021).
- Our model only has 313M parameters, which is smaller than the other submissions.
- Our submission is the only fully constrained submission that uses only the corpus provided by the organizers and does not use any pre-trained models.

Contact Information

Wen Lai
Ph.D. Candidate at LMU Munich
wenlai@cis.lmu.de

Credits


Acknowledgements

We would like to thank the Flores team for the organization and for the grant of computer ondles that we used in our experiments. This work was supported by funding to Wen Lai from LMU-CSC (China Scholarship Council) Scholarship Program (CSC, 202006390016). This work has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement #715624). This work was also supported by the DFG (grant FR 2192/4-1).