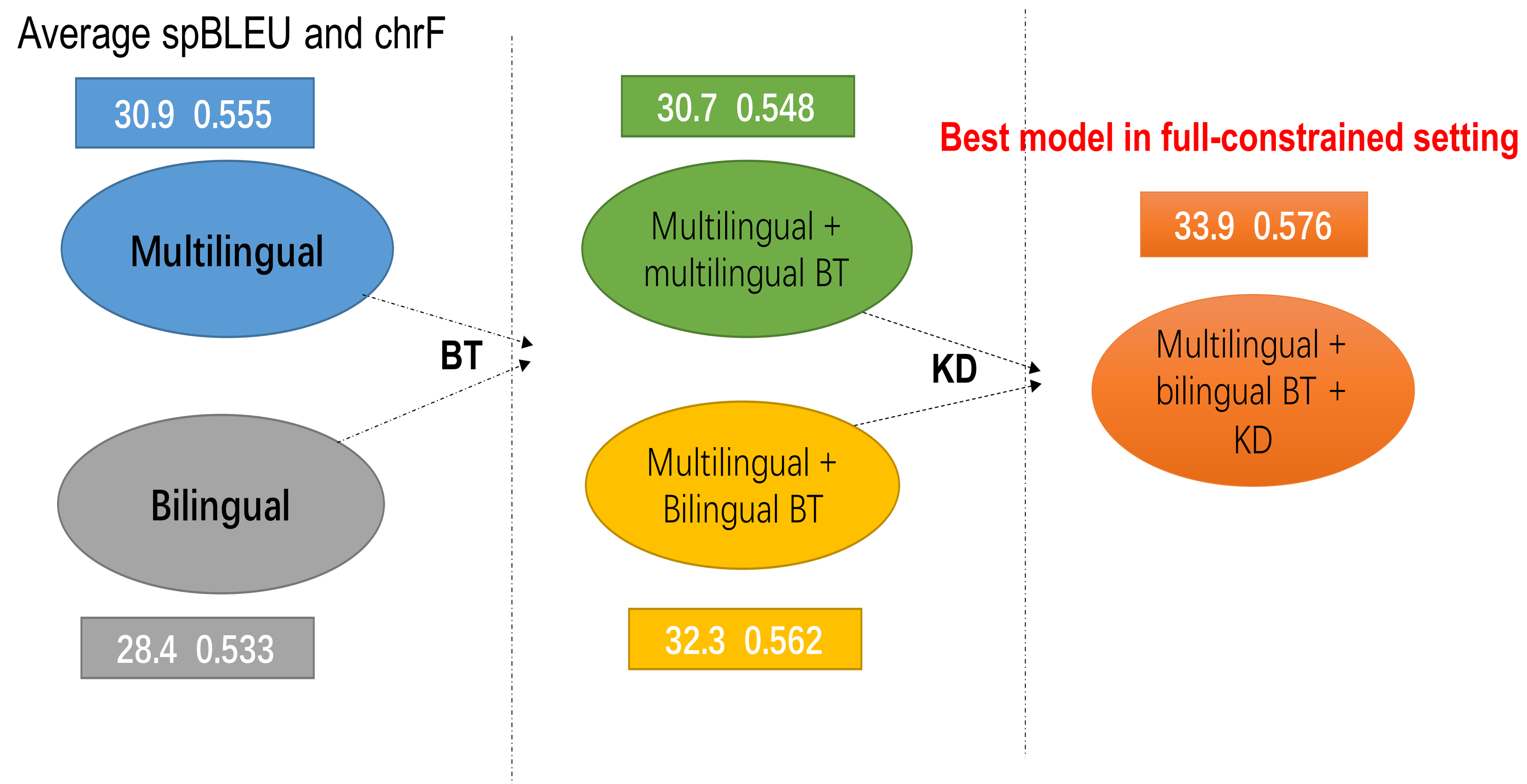
The LMU Munich System for the WMT 2021 Large-Scale Multilingual Machine Translation Shared Task

Wen Lai, Jindřich Libovický and Alexander Fraser

Center for Information and Language Processing, LMU Munich, Germany

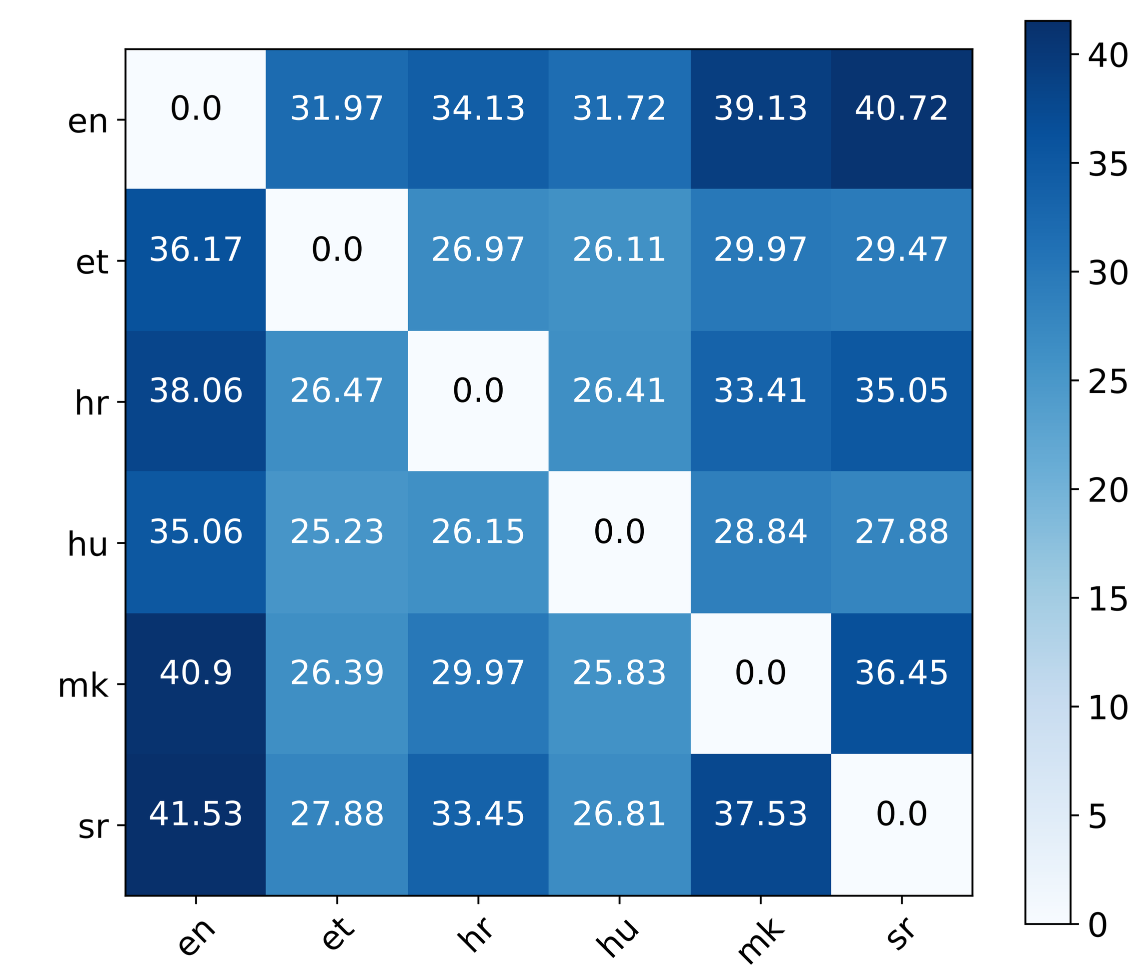
Introduction

- **MNMT** aims to handle the translation from multiple source languages into multiple target languages with singled unfied 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.



Overview of the best-performing system of our proposed approach. we report spBLEU scores and chrF on the provided devtest set.

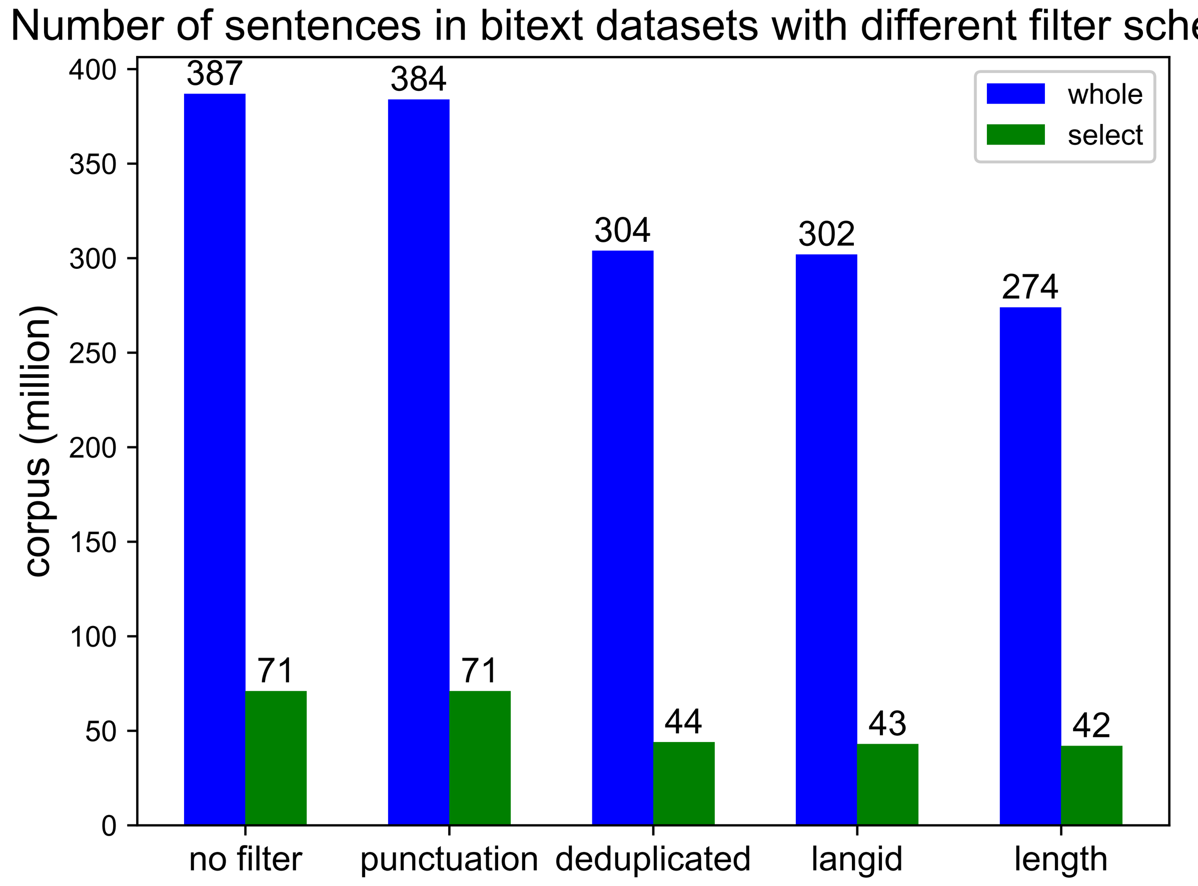
flores small1 devtest Performance Grid (BLEU score)



Data

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

Number of sentences in bitext datasets with different filter schemes



Contact Information

Wen Lai

Ph.D. Candidate at LMU Munich
lavine@cis.lmu.de

Results

#	Systems	spBLEU	chrF	BERTScore	BEST BLEU
0	Flores	28.0	0.528	0.867	sr-mk (36.0)
1	Bilingual$_{whole}$	21.1	0.477	0.831	en-mk (31.3)
2	Bilingual$_{select}$	28.4	0.533	0.863	sr-en (40.6)
3	Multilingual$_{whole}$	16.7	0.431	0.827	sr-en (26.1)
4	Multilingual$_{select}$	30.9	0.555	0.874	sr-en (40.0)
5	Multilingual$_{select}$ + TaggedBT(Multilingual$_{select}$)	30.7	0.548	0.873	sr-en (40.5)
6	Multilingual$_{select}$ + TaggedBT(Bilingual$_{select}$)	**32.3**	**0.562**	**0.879**	sr-en (41.5)
7*	Multilingual$_{select}$ + TaggedBT(Bilingual$_{select}$) + KD$_{batch}$	33.2	0.572	0.883	sr-en (42.0)
8*	Multilingual$_{select}$ + TaggedBT(Bilingual$_{select}$) + KD$_{global}$	**33.9**	**0.576**	**0.887**	sr-en (42.4)

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*).

Acknowledgements

We would like to thank the Flores team for the organization and for the grant of computer credits 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 #640550). This work was also supported by the DFG (grant FR 2829/4-1).

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.

References

1. Johnson, Melvin, et al. Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics (2017).
2. Fan, Angela, et al. Beyond english-centric multilingual machine translation. Journal of Machine Learning Research 22.107 (2021).
3. Sennrich, Rico, et al. Improving neural machine translation models with monolingual data.In proceeding of ACL.
4. Isaac Caswell et al. Tagged back-translation. In Proceedings of WMT.
5. Biao Zhang et al. Improving massively multilingual neural machine translation and zero-shot translation. In Proceedings of ACL.
6. Yoon Kim et al. Sequence-level knowledge distillation. In Proceedings of EMNLP.
7. Fusheng Wang et al. Selective knowledge distillation for neural machine translation. In Proceedings of ACL.
8. Naman Goyal et al. The flores-101 evaluation benchmark for low-resource and multilingual machine translation. arXiv preprint arXiv:2106.03193.