



The Challenges of Optimizing Machine Translation for Low Resource Cross-Language Information Retrieval



Constantine Lignos, Daniel Cohen, Yen-Chieh Lien, Pratik Mehta, W. Bruce Croft, Scott Miller

University of Southern California Information Sciences Institute, University of Massachusetts Amherst

Summary

- **Goal:** Improve cross-language IR (CLIR) for lower-resourced languages by improving IR on MT output
- **Problem:** Unclear how to optimize MT-IR pipeline
- **Approach:** Simulate low resource conditions using higher-resourced languages and evaluate MT/IR
- **Key Finding:** End-to-end tuning, especially of BPE, is critical as optimizing MT in isolation can hurt IR

Background

- Larger goal is to improve CLIR for languages such as those of IARPA MATERIAL program: Bulgarian, Lithuanian, Somali, Swahili, Tagalog, and more
- MT enables easy CLIR, but low resource MT-IR is understudied, especially so for neural models

Data

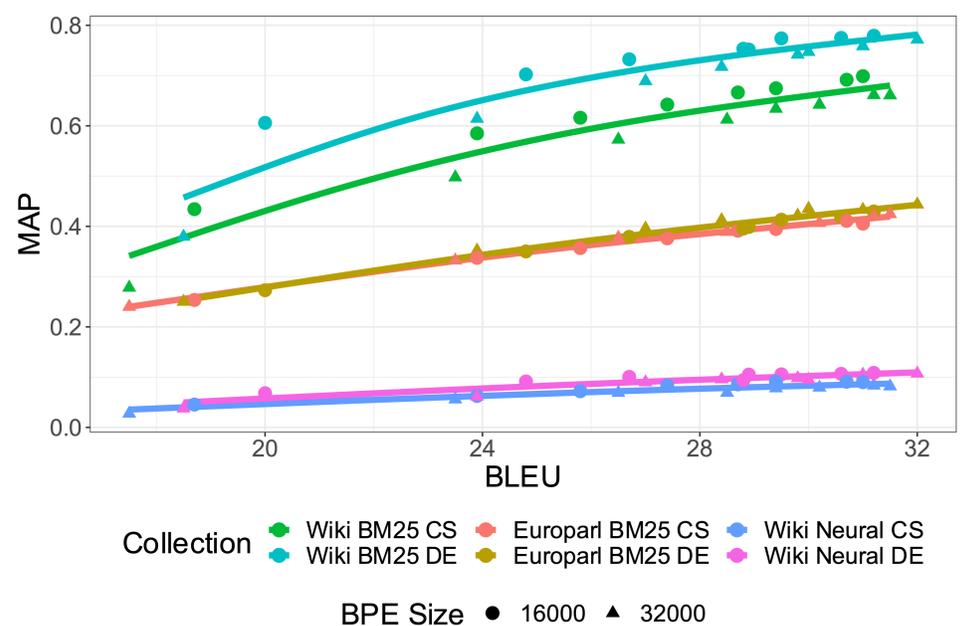
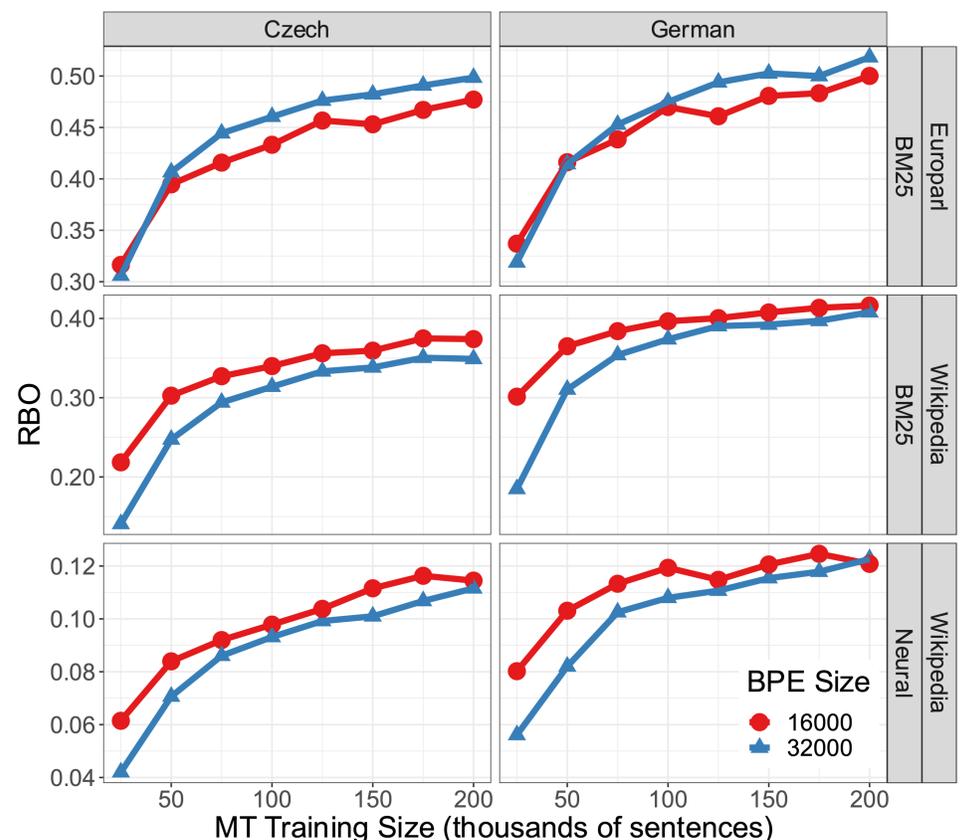
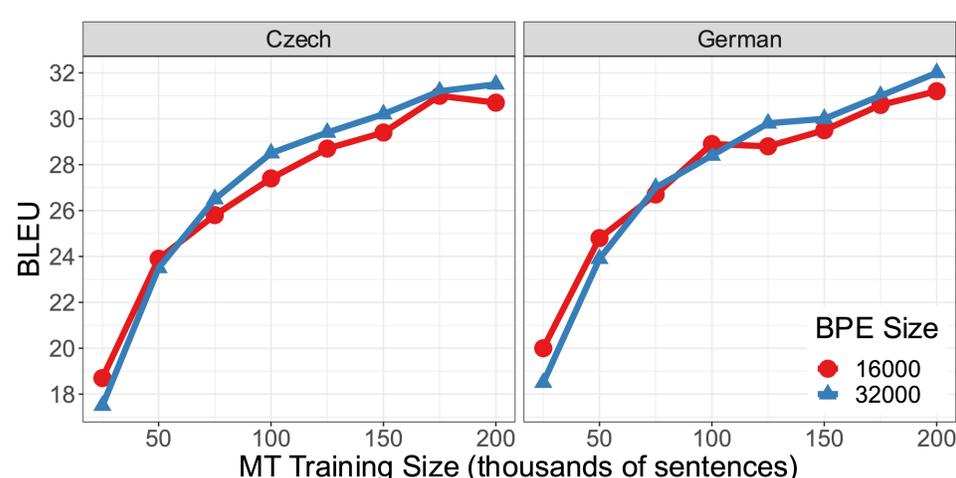
- Ideally: parallel, relevance-annotated collections in a variety of lower-resourced languages
- Absent such resources, we use Europarl and Wikipedia to create the best collections we can
- Europarl relevance: BM25 for GOV2 queries
- Wikipedia: title as query, article is relevant document
- Selected German and Czech due to comparable MT training data and sufficient Wikipedia articles

Models

- MT: fairseq fully convolutional model trained on 25k-200k sentences of the News Commentary dataset
- IR: Okapi-BM25, Duet (neural) trained on Wikipedia

MT and IR Performance

- Used RBO (similar to Kendall's Tau) to assess correlation between retrieval over MT and English
- Optimal BPE tuning for IR varies across collections and only partially matches optimal tuning for MT
- Possible cause: better match between MT training data and Europarl, smaller vocabularies support better generalization ([Senrich and Zhang, 2019](#))
- Neural IR performance on MT is very poor



Negative Results

- Failed to improve BLEU-RBO correlation by preprocessing translation before BLEU to match IR conditions (stemming, punc. removal)
- Could not identify hyperparameters for transformer that were stable across training data sizes

Conclusions

- Generally, better MT leads to better CLIR, but optimizing MT in isolation can reduce IR quality
- No substitute (yet) for end-to-end tuning
- MT substantially degrades neural retrieval performance, further work needed in this area to enable usable neural IR over MT output

Acknowledgments

Thanks to Jonathan May and Thamme Gowda for enlightening discussions. This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract FA8650-17-C-9116. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.