

Does incorporating multilingual intermediate tasks in addition to English tasks further improve zero-shot cross-lingual transfer

Assignee Research

July 26, 2026

Abstract

Previous work mainly focuses on improving cross-lingual transfer for NLU tasks with a multilingual pretrained encoder (MPE), or improving the performance on supervised machine translation with BERT. However, it is under-explored that whether the MPE can help to facilitate the cross-lingual transferability of NMT model. In this paper, we focus on a zero-shot cross-lingual transfer task in NMT. In this task, the NMT model is trained with parallel dataset of only one language pair and an off-the-shelf MPE, then it is directly tested on zero-shot language pairs. We propose SixT, a simple yet effective

1 Introduction

This paper examines: Zero-shot Cross-lingual Transfer of Neural Machine Translation with Multilingual Pretrained Encoders. Research question: Does incorporating multilingual intermediate tasks in addition to English tasks further improve zero-shot cross-lingual transfer performance on XTREME-R compared to English-only intermediate training, as measured by average accuracy across all languages?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.7/10.

3 Results

11 papers retrieved. 10 claims extracted; 7 independently verified. Quality review score: 6.7/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The Europarl-v7 German and English dataset is used as the training set for the ZeXT task.	✓	0.19
The cross-lingual transfer abilities of NMT models are evaluated on languages from different language groups: German gro	✓	0.39
A concatenation of Fr-En and Cs-En validation datasets is used as the validation dataset for all any-to-English translat	✓	0.25
The checkpoint is selected by zero-shot cross-lingual transfer performance on the validation set for all experiments.	✓	0.22
The baseline models include Vanilla Transformer, +XLM-R fine-tune encoder, +XLM-R fine-tune all, +XLM-R as encoder embed	✓	0.18
The proposed SixT model achieves an average BLEU score of 18.3 across various languages.	×	0.12
Strategy (10) with a big decoder and initialized decoder embedding and all encoder parameters with XLM-R, and training t	✓	0.23
The performance of SixT drops by 1.8 average BLEU when using a vanilla Transformer with the same size as Strategy (7) (w	✓	0.28
The SixT model achieves a BLEU score of 33.8 for German, 54.7 for Dutch, 30.1 for Spanish, 33.9 for Italian, 43.0 for Ro	×	0.04
The SixT model uses 0.04B sentences for training.	×	0.05

References

- <http://arxiv.org/abs/2104.08757v2>

- <http://arxiv.org/abs/2005.00633v1>
- <http://arxiv.org/abs/2005.13013v2>