Intermediate Task Selection and Zero-Shot Cross-Lingual Transfer in XTREME-R

Assignee Research

June 21, 2026

Abstract

Intermediate-task training—fine-tuning a pretrained model on an intermediate task before fine-tuning again on the target task—often improves model performance substantially on language understanding tasks in monolingual English settings. We investigate whether English intermediate-task training is still helpful on non-English target tasks. Using nine intermediate language-understanding tasks, we evaluate intermediate-task transfer in a zero-shot cross-lingual setting on the XTREME benchmark. We see large improvements from intermediate training on the BUCC and Tatoeba sentence retrieval tasks a

1 Introduction

This paper examines: . Research question: How does the choice of intermediate language-understanding task (e.g., NLI, QA, classification) influence the zero-shot cross-lingual transfer accuracy in XTREME-R when the intermediate task is from a non-English language?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.7/10.

3 Results

14 papers retrieved. 9 claims extracted; 9 independently verified. Quality review score: 8.7/10.

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
Intermediate-task training improves model performance substantially on language understanding tasks in monolingual Engli	✓	0.39
English intermediate-task training is investigated for its effectiveness on non-English target tasks.	✓	0.27
Nine intermediate language-understanding tasks are used to evaluate intermediate-task transfer in a zero-shot cross-ling	✓	0.36
Large improvements are observed from intermediate training on the BUCC and Tatoeba sentence retrieval tasks.	✓	0.27
Moderate improvements are observed from intermediate training on question-answering target tasks.	✓	0.25
MNLI, SQuAD, and HellaSwag achieve the best overall results as intermediate tasks.	✓	0.30
Multi-task intermediate offers small additional improvements.	✓	0.28
Using the best intermediate-task models for each target task, a 5.4 point improvement over XLM-R Large on the XTREME ben	✓	0.41
Continuing multilingual MLM during intermediate-task training and using machine-translated intermediate-task data do not	✓	0.41

References

- https://openalex.org/W3116343068
- https://doi.org/10.18653/v1/2022.findings-acl.196
- https://doi.org/10.18653/v1/2022.acl-long.62