Multilingual Model Alignment and Zero-Shot Cross-Lingual Text Classification Performance

Assignee Research

July 8, 2026

Abstract

The introduction of pretrained cross-lingual language models brought decisive improvements to multilingual NLP tasks. However, the lack of labelled task data necessitates a variety of methods aiming to close the gap to high-resource languages. Zero-shot methods in particular, often use translated task data as a training signal to bridge the performance gap between the source and target language(s). We introduce XeroAlign, a simple method for task-specific alignment of cross-lingual pretrained transformers such as XLM-R. XeroAlign uses translated task data to encourage the model to generate sim

1 Introduction

This paper examines: XeroAlign: Zero-Shot Cross-lingual Transformer Alignment. Research question: How does the alignment of multilingual language models with human feedback impact zero-shot cross-lingual performance in text classification tasks compared to alignment using English-only feedback?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.2/10.

3 Results

9 papers retrieved. 18 claims extracted; 15 independently verified. Quality review score: 8.2/10.

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
XeroAligned XLM-R achieves state-of-the-art scores on three task-oriented XNLU datasets.	✓	0.18
For MTOP, the intent classification accuracy (+1.1) and slot filling F-Score (+2.4) averaged over 5 languages improved o	✓	0.43
For MultiATIS++, XLM-RA shows an improved intent accuracy (+1.1) and slot F-Score (+3.2) over 8 languages, as compared t	✓	0.41
For MTOD, the classification accuracy (+1.3) and slot tagging F-Score (+5.0) on average improved on XLM-R-Large with tra	✓	0.46
XLM-RA-base model outperforms (albeit marginally) XLM-RA-large on the MTOD dataset.	✓	0.26
XLM-RA scores marginally higher (+0.1 accuracy) than VECO (Luo et al., 2020) on the adversarial paraphrase task (PAWS-X)	✓	0.29
XLM-RA scores marginally lower (-0.2 accuracy) than FILTER (Fang et al., 2020) on the adversarial paraphrase task (PAWS-	✓	0.29
XeroAligned XLM-R exceeds the intent classification accuracy of XLM-R trained with labelled data, averaged across three	✓	0.41
XeroAlign improves intent classification by $\sim$5-10 points (larger for XLM-R-base).	✓	0.20
XLM-RA accuracy exceeds both Target and the Translate-Train averages by over 1 point and by almost 3 points over the zer	×	0.13
XeroAlign uses translated task data to encourage the model to generate similar sentence embeddings for different languag	✓	0.36
XLM-RA shows strong improvements over the baseline models to achieve state-of-the-art zero-shot results on three multili	✓	0.43
XLM-RA's text classification accuracy exceeds that of XLM-R trained with labelled data and performs on par with state-of	✓	0.42
XeroAlign pursues a simplified, task-specific model alignment instead of large-scale model alignment with general parall	×	0.14
XeroAligned XLM-R model (XLM-RA) achieves SOTA scores on three XNLU datasets.	✓	0.21
XLM-RA exceeds the text classification performance of XLM-R trained with labelled data.	✓	0.18
XLM-RA performs on par with SOTA models on an adversarial paraphrasing task.	✓	0.17
XeroAlign is evaluated on 4 datasets that cover 11 unique languages.	×	0.11

References

- http://arxiv.org/abs/2105.02472v2
- http://arxiv.org/abs/2406.01771v1
- http://arxiv.org/abs/2402.18120v3