Impact of Hybrid Batch Training on Zero-Shot Cross-Lingual Retrieval Accuracy in BEIR

Assignee Research

June 28, 2026

Abstract

Information retrieval across different languages is an increasingly important challenge in natural language processing. Recent approaches based on multilingual pre-trained language models have achieved remarkable success, yet they often optimize for either monolingual, cross-lingual, or multilingual retrieval performance at the expense of others. This paper proposes a novel hybrid batch training strategy to simultaneously improve zero-shot retrieval performance across monolingual, cross-lingual, and multilingual settings while mitigating language bias. The approach fine-tunes multilingual lang

1 Introduction

This paper examines: Synergistic Approach for Simultaneous Optimization of Monolingual, Cross-lingual, and Multilingual Information Retrieval. Research question: How does the hybrid batch training strategy impact zero-shot cross-lingual retrieval accuracy on the BEIR benchmark compared to single-objective contrastive fine-tuning?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.5/10.

3 Results

13 papers retrieved. 18 claims extracted; 17 independently verified. Quality review score: 8.5/10.

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 approach fine-tunes multilingual language models using a mix of monolingual and cross-lingual question-answer pair b	✓	0.41
Experiments on XQuAD-R, MLQA-R, and MIRACL Datasets.	×	0.12
XQuAD-R and MLQA-R are question-answering datasets with parallel questions and passages in 11 languages and 7 languages,	✓	0.20
The evaluation of the models is conducted on datasets that are completely separate and distinct from the ones used for t	✓	0.24
The models have not encountered any data samples, whether from the training or testing splits, of the evaluation dataset	✓	0.24
We report the mean average precision (mAP) for XQuAD-R and MLQA-R.	✓	0.16
We conduct retrieval using the queries with XQ language against the corpus with XC language and report the macro-average	✓	0.43
In XQuAD-R (MLQA-R), we have 11 and 7 parallel languages; thus, there are 110 (42) and 11 (7) cross-lingual and monoling	✓	0.24
For multilingual (denoting Mul.) retrieval, we conduct retrieval using the queries with XQ language against all the para	✓	0.26
X-X and X-Y sampling only perform well in monolingual and cross-lingual retrieval settings, respectively.	✓	0.23
Optimization for either monolingual or cross-lingual retrieval alone may come at the expense of the other.	✓	0.19
Hybrid batch sampling achieves the best performance in multilingual retrieval settings.	✓	0.29
Hybrid batch sampling is better than the other two baseline batch sampling methods when using XLM-R and LaBSE as initial	✓	0.26
Hybrid batch training achieves superior results in zero-shot retrieval across various languages and retrieval tasks comp	✓	0.41
Hybrid batch training substantially reduces language bias in multilingual retrieval compared to monolingual training.	✓	0.37
The proposed approach enables learning language-agnostic representations that enable strong zero-shot retrieval performa	✓	0.34
The approach fine-tunes multilingual language models using a balanced mix of monolingual and cross-lingual question-answ	✓	0.31
The approach collects a diverse set of English question-answer datasets and uses machine	✓	0.23

References

- http://arxiv.org/abs/2408.10536v1
- http://arxiv.org/abs/2506.15415v1
- http://arxiv.org/abs/2305.19840v2