Geodesic vs. Euclidean Dense Retrievers in Cross-Lingual Transfer Performance

Assignee Research

June 3, 2026

Abstract

This report synthesises findings from 1 peer-reviewed paper addressing the following research question: How does the cross-lingual transferability of geodesic distance-based dense retrievers compare to Euclidean-based models when evaluated on multilingual BEIR benchmarks. 6 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Beyond the Unit Hypersphere: Embedding Magnitude in Contrastive Learning. Research question: How does the cross-lingual transferability of geodesic distance-based dense retrievers compare to Euclidean-based models when evaluated on multilingual BEIR benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 1 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.8/10.

3 Results

1 papers retrieved. 6 claims extracted; 6 independently verified. Quality review score: 8.8/10.

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
Cosine similarity normalizes both sides; dot product normalizes neither.	✓	0.27
We propose a 2x2 framework that independently controls query-side and document-side normalization, exposing two intermed	✓	0.35
On retrieval with four encoders, evaluated in-domain on MS MARCO and out-of-domain on BEIR, BRIGHT, and multi-hop QA, th	✓	0.47
Cross-evaluation reveals the mechanism: document magnitude scales inference scores while query magnitude modulates train	✓	0.41
We then classify tasks by functional symmetry, defined as whether the aggregate scoring procedure treats Q and C as inte	✓	0.33
On five additional task families (semantic textual similarity, CLIP, knowledge graph completion, few-shot classification	✓	0.57

References

- `https://openalex.org/W7128648559`