DPR Retrieval Performance Decline with Reduced Context Window Sizes on SQuAD 2.0

Assignee Research

May 29, 2026

Abstract

Dense Passage Retrieval (DPR) typically relies on Euclidean or cosine distance to measure query-passage relevance in embedding space, which is effective when embeddings lie on a linear manifold. However, our experiments across DPR benchmarks suggest that embeddings often lie on lower-dimensional, non-linear manifolds, especially in out-of-distribution (OOD) settings, where cosine and Euclidean distance fail to capture semantic similarity. To address this limitation, we propose a manifold-aware distance metric for DPR (MA-DPR) that models the intrinsic manifold structure of passages using a

1 Introduction

This paper examines: MA-DPR: Manifold-aware Distance Metrics for Dense Passage Retrieval. Research question: How does DPR retrieval performance (Recall@5) on SQuAD 2.0 vary when context window size is reduced from 4096 to 1024 tokens?.

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 7.2/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 7.2/10.

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.

References

- http://arxiv.org/abs/2509.13562v1
- http://arxiv.org/abs/2108.06279v2
- http://arxiv.org/abs/2503.17876v1