Manifold-Aware Distance Metrics and Inference Efficiency in Large-Scale Dense Retrieval

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the impact of manifold-aware distance metrics like MA-DPR on the inference efficiency of dense retrievers when scaling to large-scale document collections (e.g., MS MARCO or BEIR benchmarks). 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. 14 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MA-DPR: Manifold-aware Distance Metrics for Dense Passage Retrieval. Research question: What is the impact of manifold-aware distance metrics like MA-DPR on the inference efficiency of dense retrievers when scaling to large-scale document collections (e.g., MS MARCO or BEIR benchmarks) compared to traditional cosine-based DPR?.

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 3.8/10.

3 Results

13 papers retrieved. 14 claims extracted; 1 independently verified. Quality review score: 3.8/10.

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
The system specifications include CPU—Intel(R) Core(TM) i7-14700HX and GPU—NVIDIA GeForce RTX 4070 Laptop GPU.	×	0.00
Average CPU utilization during measurement is $\sim$5%.	×	0.01
All codes and results are available online at github.com/QianfengWen/Manifold_Distance_Retrieval.git.	×	0.06
The experiments aim to evaluate the effectiveness of MA-DPR dManifold against several baselines including DPR with dEucl	×	0.09
DPR benchmarks include MS MARCO, NFCorpus, SciDocs, and ANTIQUE.	×	0.09
Two embedding models are used: msmarco-distilbert-base-tas-b (tas-b) and SciNCL.	×	0.04
MS MARCO is the in-distribution dataset for tas-b and SciDocs is the in-distribution dataset for SciNCL.	×	0.14
All embeddings are 2-normalized.	×	0.03
Empirical evaluation assesses Recall, Mean Average Precision (MAP), and Normalized Discounted Cumulative Gain (nDCG) for	×	0.03
The experiments address five key research questions: RQ1, RQ2, RQ3, RQ4, and RQ5.	×	0.02
RQ1 aims to validate the manifold hypothesis in dense embedding spaces by examining the relationship between dManifold a	×	0.09
In Figure 2, for each ground truth relevant query q and passage p pair (orange dots) and irrelevant pair (blue dots), dE	×	0.07
In a perfectly linear embedding space, the manifold-aware distance induced by dKNN_Euclidean+cDC should closely align wi	✓	0.20
In the presence of non-linear structure, the two distances are expected to diverge.	×	0.07

References

- http://arxiv.org/abs/2509.13562v1
- http://arxiv.org/abs/2205.02870v2

- http://arxiv.org/abs/2412.08329v1