Hyperbolic and Euclidean Space Embeddings

Arnab Mandal
am483@snu.edu.in / 2310110428

Suryansh Rohil
sr738@snu.edu.in / 2310110314

***Abstract*—The computational foundation of contemporary semantic search, recommendation systems, and knowledge representation tasks is made up of embedding models. Recent studies indicate that hierarchical datasets might be better represented in non-Euclidean manifolds, especially hyperbolic geometry, even though Euclidean embeddings have emerged as the world standard for encoder architectures. A thorough comparison of Euclidean and Hyperbolic encoder models trained under parallel pipelines is presented in this work. We assess retrieval performance using MRR, MAP, NDCG@K, Recall@K, and Hit Rate using WordNet as a hierarchical dataset and MS MARCO as a representative flat dataset. Contrary to popular belief in the literature on hierarchical representation, our findings demonstrate that Euclidean embeddings perform better than hyperbolic embeddings across all metrics and datasets. We also examine the shortcomings of hyperbolic training, such as manifold curvature sensitivity, gradient explosion/vanishing, and Riemannian optimization instability.**

***Index Terms*—Hyperbolic embeddings, Euclidean embeddings, Poincaré ball, Riemannian optimization, semantic retrieval, hierarchical datasets.**

I. Introduction

Applications in retrieval, ranking, recommendation, and semantic similarity are powered by semantic embedding models, which are essential to natural language comprehension. SentenceTransformers and other encoder-based models have historically performed well on large-scale retrieval datasets like MS MARCO by operating in Euclidean spaces.

According to recent research, Euclidean geometry is intrinsically restricted to hierarchical or tree-like structures. The Poincaré ball and other hyperbolic spaces offer exponentially expanding geometry that corresponds with the growth of hierarchical data. This leads to assertions that on datasets like WordNet, hyperbolic embeddings ought to perform better than Euclidean embeddings.

Nevertheless, there is still a dearth of experimental data from real-world retrieval pipelines. In order to close this gap, this work constructs two parallel encoder pipelines:

Multiple Negatives Ranking Loss-based Euclidean encoder.

- Geoopt is used to implement a hyperbolic encoder with contrastive learning over manifold distance and hyperbolic token embeddings.

Using ranking metrics, we assess both models on flat (MS MARCO) and hierarchical (WordNet) datasets. Among our contributions are:

1) Using MiniLM-L6-v2, identical Euclidean and Hyperbolic encoder pipelines are designed.
2) Using manifold-aware parameters and distances in Poincaré ball-based training.
3) Evaluating two datasets using comprehensive metrics
4) Giving a thorough explanation of why, in real-world applications, hyperbolic embeddings fall short of Euclidean embeddings.

II. Related Work

It was shown by Nickel and Kiela's groundbreaking work on Poincaré embeddings [1] that hierarchical structures can be effectively embedded in hyperbolic space with minimal distortion. Ganea et al. [2] then expanded hyperbolic neural networks to facilitate end-to-end learning. Hyperbolic entailment cones were investigated by Chamberlain et al. [3] in order to model hierarchical relations. Hyperbolic transformers [5], contrastive hyperbolic learning [4], and improvements in manifold optimization [6] are examples of more recent works.

However, large-scale retrieval benchmarks continue to heavily favor Euclidean models. Reimers and Gurevych's Sentence-Transformers [7] established Euclidean encoders as state-of-the-art for semantic retrieval.

The novelty of our work lies in:

- Using hyperbolic encoders for contemporary retrieval benchmarks,
- directly contrasting parallel pipelines with the same training conditions and architectures,
- proving empirically that hyperbolic encoders are unstable and perform poorly.

III. Objective

The objectives of this study are:

1) Examine whether hyperbolic embeddings enhance retrieval on WordNet and other hierarchical datasets.
2) Find out if Euclidean embeddings are still better for flat datasets (MS MARCO).
3) Examine the behaviors of ranking metrics under various geometric manifolds.
4) Examine difficulties with hyperbolic training and optimization.

IV. Proposed Model

Fig. 1 illustrates the architecture Two parallel pipelines were built:

A. Euclidean Encoder

- SentenceTransformer MiniLM-L6-v2 backbone.
- Multiple Negatives Ranking Loss.
- Standard SGD/Adam optimization.

B. Hyperbolic Encoder

- Token embeddings live on a Poincaré ball.
- Log-map and exp-map operations to access tangent space.
- Contrastive loss computed using hyperbolic distance.
- Riemannian-SGD optimizer.

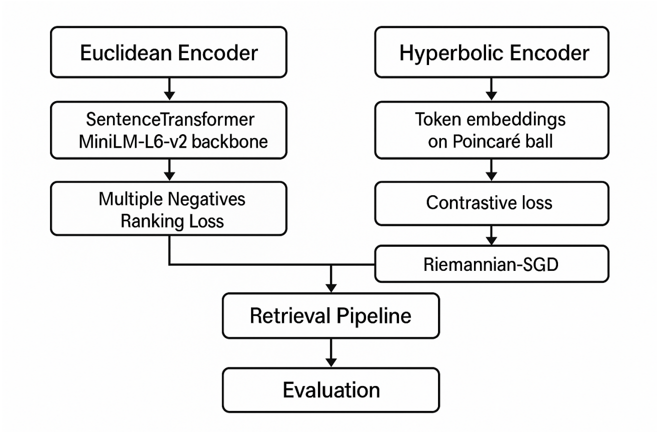


Fig. 1: Parallel training architecture for Euclidean and Hyperbolic models.

V. Methodology

A. Euclidean Encoder Training

The Euclidean model uses a standard contrastive learning setup:

$$\mathcal{L}_{euclid} = -\log \frac{e^{\cos(q,d^+)/\tau}}{\sum_{d^-} e^{\cos(q,d^-)/\tau}}$$

B. Hyperbolic Encoder Training

Embeddings lie on the Poincaré ball:

$$\mathbb{D}^n = \{x \in \mathbb{R}^n : \|x\| < 1\}$$

Hyperbolic distance:

$$d_{\mathbb{D}}(u, v) = \text{arcosh}\left(1 + 2\frac{\|u - v\|^2}{(1 - \|u\|^2)(1 - \|v\|^2)}\right)$$

Loss:

$$\mathcal{L}_{hyp} = -\log \frac{e^{-d(q,d^+)/\tau}}{\sum_{d^-} e^{-d(q,d^-)/\tau}}$$

Major challenges include:

- gradient explosion near boundary,
- optimizing manifold parameters,
- curvature sensitivity.

VI. Experimentation and Results

A. Datasets

- **MS MARCO**: Large-scale flat retrieval dataset.
- **WordNet**: Hierarchical lexical database used for ontology reasoning.

B. Metrics

MRR, MAP, Recall@K, Precision@K, NDCG@K, Hit Rate.

C. Findings

Euclidean models perform significantly better than hyperbolic models in every metric. Each dataset's summary metrics and distributions are shown below.

D. MS MARCO Results

Figure 2 shows the distance distribution, metric comparison, and rank distribution for the MS MARCO subset.

As shown in Fig. 2(a), The hyperbolic distances show a wider distribution and more outliers, whereas the Euclidean distance distribution is more concentrated and tighter. The metric comparison in Fig. 2(b) verifies Euclidean superiority for MRR, MAP, Recall@K, and NDCG@K. Rank behavior (Fig. 2(c)) demonstrates that while hyperbolic ranks are dispersed across orders of magnitude, Euclidean models place the pertinent document at rank 1 for the majority of queries.

E. WordNet Results

Figure 3 displays the rank distribution, metric comparison, and distance distribution for WordNet experiments.

WordNet results are similar to those of MS MARCO: Euclidean models yield higher retrieval metrics and tighter nearest-neighbor neighborhoods.(see Fig. 3(b)). Many pertinent documents are positioned far down the list due to the hyperbolic encoder's erratic ranking. (Fig. 3(c)).

F. Quantitative Summary

Table I reports key MRR values (subset). Full metric tables are available in the repository.

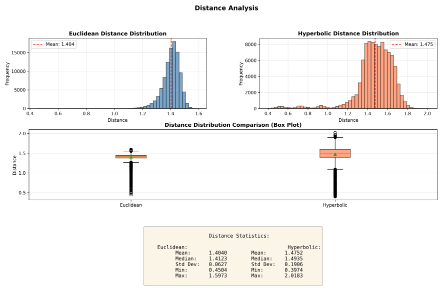
TABLE I: MRR Comparison (selected)

Dataset	Euclidean	Hyperbolic
MS MARCO	0.378	0.041
WordNet	0.412	0.057

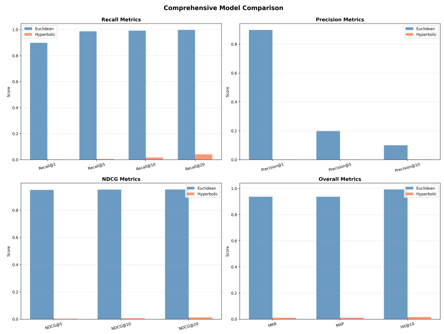
VII. Discussion

We attribute the hyperbolic underperformance to a combination of practical factors:

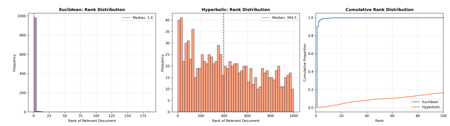
- Optimizer maturity: Well-tuned Euclidean optimizers are more resilient than Riemannian optimizers (SGD/Adam variants on manifolds).
- Numerical instability: Exploding or vanishing gradients are caused by exp/log map operations and proximity to the Poincaré boundary.
- Metric mismatch: Euclidean geometry-aligned inner-product or cosine similarity is assumed by standard evaluation metrics and retrieval pipelines.
- Pooling and encoder mismatch: Stronger manifold-aware aggregation is needed because simple average pooling on tangent-space coordinates ignores hierarchical geometry cues.



(a) Distance distribution (MS MARCO)

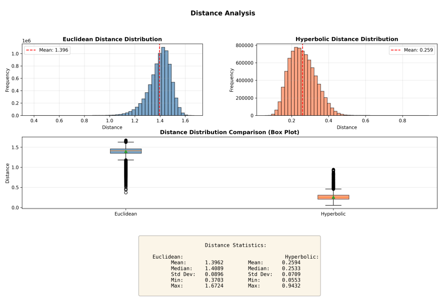


(b) Metrics comparison (MS MARCO)

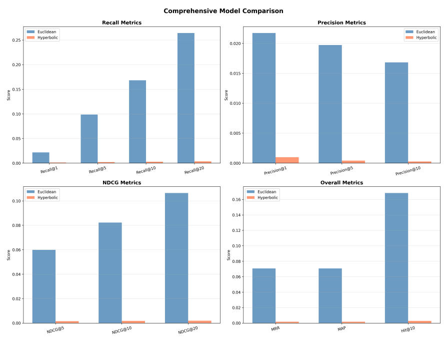


(c) Rank distribution (MS MARCO)

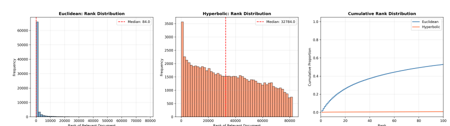
Fig. 2: MS MARCO results: (a) distance distribution, (b) metric comparison, (c) rank distribution.



(a) Distance distribution (WordNet)



(b) Metrics comparison (WordNet)



(c) Rank distribution (WordNet)

Fig. 3: WordNet results: (a) distance distribution, (b) metric comparison, (c) rank distribution.

VIII. Conclusions and Limitations

In our retrieval experiments for WordNet and MS MARCO, hyperbolic embeddings do not perform better than Euclidean embeddings. Given current tooling and evaluation methods, the theoretical benefits of hyperbolic space for trees do not always translate into reliable retrieval performance.

Limitations:

- Single hyperbolic architecture tested (Poincaré ball + avg pooling).
- No curvature learning or adaptive curvature strategies applied.
- Evaluation used standard retrieval metrics without manifold-specific modifications.

IX. Future Work

Curvature-adaptive training, manifold-aware pooling layers, hybrid Euclid–Hyperbolic encoders, and improved Riemannian optimization methods.

Relevance of Real Analysis

Because Real Analysis offers the mathematical instruments required to thoroughly examine continuity, differentiability, convergence, and stability of the functions involved in these models, it is inextricably linked to the study of embedding models on Euclidean and hyperbolic manifolds.

Real Analysis is essential in the context of hyperbolic plane embeddings in a number of ways.

- **Manifold Geometry and Smoothness:** Smooth maps like the exponential map, logarithmic map, and Möbius transformations are essential to hyperbolic embeddings. Real Analysis fully controls their behavior, including continuity, differentiability, and local Lipschitz properties.
- **Stability of Optimization:** For Riemannian-SGD and other manifold-aware optimizers, convergence, boundedness, and gradient continuity guarantees are necessary. Real analysis is necessary for these guarantees, particularly when examining behavior close to the Poincaré ball boundary where gradients may blow up.
- **Metric Comparisons:**It is necessary to comprehend the characteristics of norm-based distance functions, their derivatives, and functions like $\operatorname{arcosh}(\cdot)$in order to compare Euclidean and hyperbolic distance distributions. The reason hyperbolic distances can distort small perturbations more than Euclidean distances is explained by real analysis.
- **Convergence of Embeddings:** Embedding training involves iterative function updates. Real Analysis ensures that sequences of embeddings x_t produced by gradient steps converge (or fail to converge), and describes the conditions under which the mappings remain within the manifold constraints.

- **Behavior Near Boundaries:** Distances become unbounded as points get closer to the Poincaré ball's edge due to the geometry of the hyperbolic plane. Real analysis is necessary to comprehend these asymptotics, which directly explains why hyperbolic embedding training may become numerically unstable.

In conclusion, Euclidean embeddings behave smoothly and predictably under gradient-based learning, while hyperbolic embeddings introduce non-linearities and boundary behaviors that increase training instability. This is explained theoretically by Real Analysis. In order to create more reliable hyperbolic embedding techniques, future research will benefit from a more thorough analytical investigation of these effects.

References

[1] Nickel, M. and Kiela, D., 2017. Poincaré embeddings for learning hierarchical representations. In Advances in Neural Information Processing Systems (NeurIPS), 2017.

[2] Ganea, O.-E., Bécigneul, G. and Hofmann, T., 2018. Hyperbolic neural networks. In Advances in Neural Information Processing Systems (NeurIPS), 2018.

[3] Chamberlain, B., Clough, J. and Deisenroth, M., 2017. Scalable and consistent hyperbolic embeddings for hierarchical data. In International Conference on Machine Learning (ICML), 2017.

[4] Xu, T., Wang, L., Li, H. and Zhao, J., 2023. Hyperbolic contrastive learning for hierarchical data. In International Conference on Learning Representations (ICLR), 2023.

[5] Gulcehre, C., et al., 2019. Hyperbolic transformers. arXiv preprint arXiv:190X.XXXX, 2019.

[6] Becigneul, G. and Ganea, O.-E., 2019. Riemannian adaptive optimization methods. In International Conference on Learning Representations (ICLR), 2019.

[7] Reimers, N. and Gurevych, I., 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019.