ProdKG: Embedding Knowledge Graphs into Product Spaces

Eric Frankel[1,2], Dhruvik Parikh[2], Tommy Yang[2]

{ericsf, dhruvik, t2yang}@cs.stanford.edu

[1]Department of Mathematics, Stanford University, [2]Department of Computer Science, Stanford University

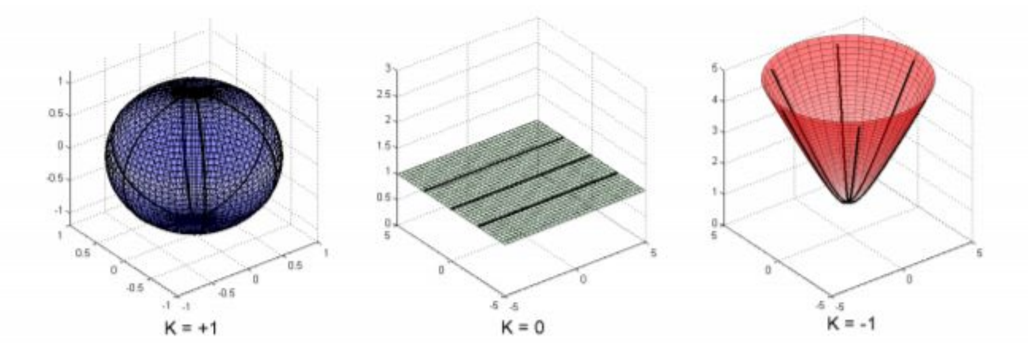
Stanford Computer Science

Abstract

- Knowledge graph (KG) elements are relational triples of the form (**h**ead, **r**elation, **t**ail).

Obama (h) —has nationality (r)→ U.S. (t)

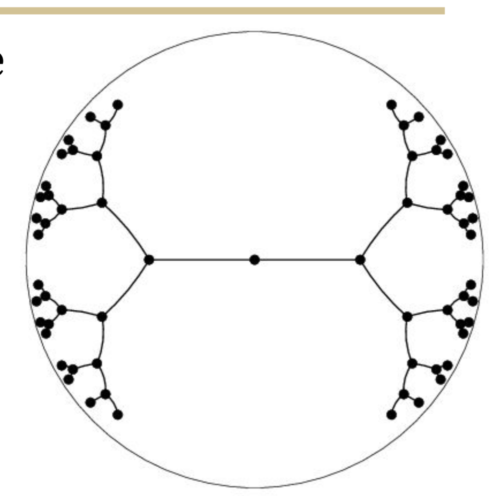
- Predicting **new relations** and returning **correct queries** depends on the **data representation.**



- Embedding graphical data in a **product** of model spaces has led to **decreased** data distortion.
- We embed knowledge graphs on product spaces and assessed **distortion** and **query performance.**
- Embedding KGs in product spaces can **decrease distortion**, but **won't guarantee** all-around improvements on querying tasks.

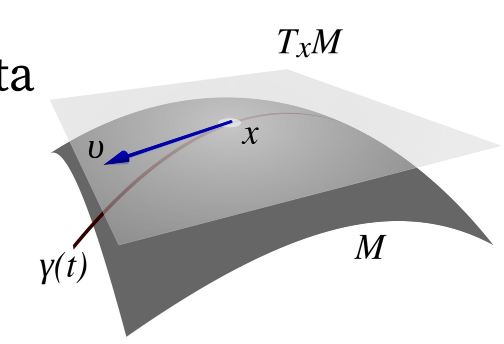
Non-Euclidean Embeddings

- Traditionally, data for machine learning has been represented in **Euclidean space** as tensors
- Three key benefits to non-Euclidean embeddings:
  - Data can have a **better representation**
  - Better **models** built on better geometries
  - **Lower dimensionality** and faster **operations**



Embedding of a tree in hyperbolic space.

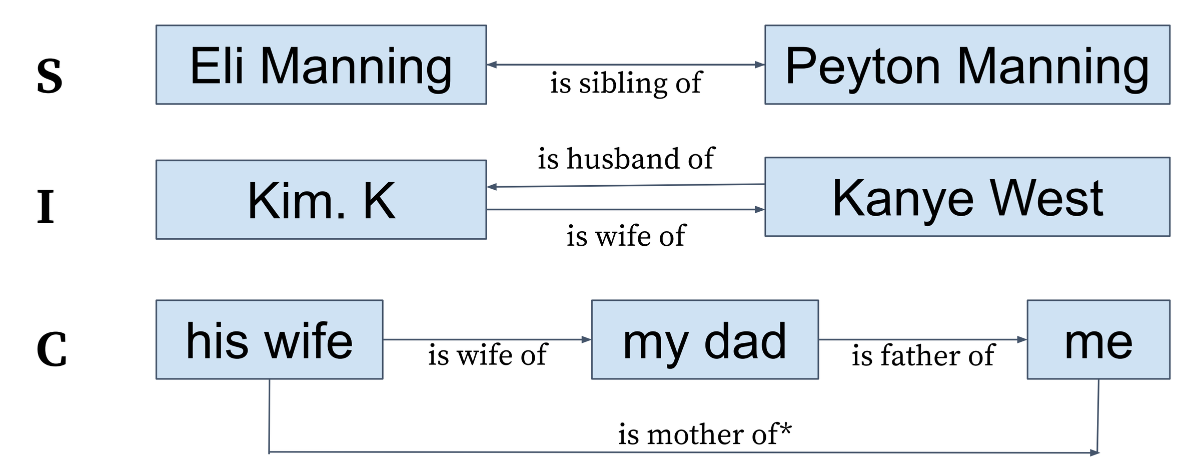
- Different spaces, different data
  - Cycles → **spherical**
  - Tree → **hyperbolic**
- Have to use **non-Euclidean** methods for working over **Riemannian manifolds**
  - Riemannian SGD (RSGD)



Operations in the **tangent space** of a manifold.

KG-Product Space Distortion

- KG embeddings must represent **three relation** patterns: Symmetry, Inversion, Composition

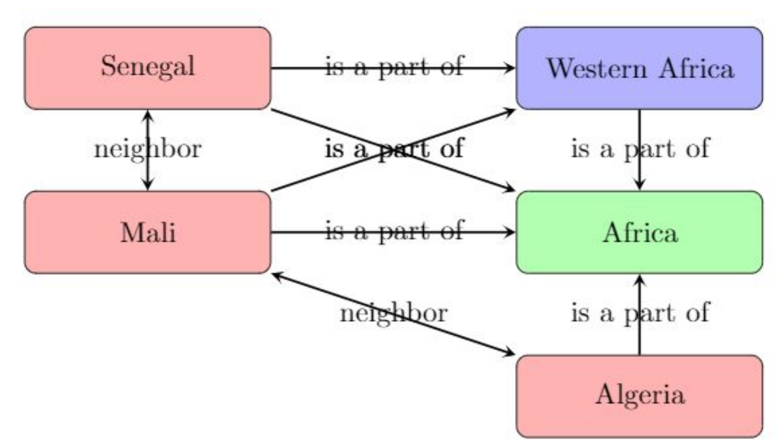


- A **product space** is a Riemannian manifold P defined by $\mathcal{P} = \mathbb{S}^{s_1} \times \cdots \times \mathbb{H}^{h_1} \times \cdots \times \mathbb{R}^{e_1}$
- Goal: embed in a product space that minimizes distortion with using the loss function

$$\mathcal{L}(x) = \sum_{1 \leq i \leq j \leq n} \left| \left(\frac{d_{\mathcal{P}}(x_i, x_j)}{d_G(X_i, X_j)} \right)^2 - 1 \right|$$

optimized using Riemannian SGD.

- Datasets: **Countries**
  - **Countries:** relationships between countries and regions (testing **all three** patterns)



- We embedded the data into **three spaces**, two of which were **exclusively** Euclidean and hyperbolic, and one **product space**, using a constant total dimension.
  - Product space outperformed both Euclidean and Hyperbolic embeddings (lower distortion), suggesting attributes of both types of embeddings were utilized

	Distortion	MAP	Training Loss
$\mathbb{R}^{10}$	.374	.472	.299
$\mathbb{H}^{10}$	.509	.440	.402
$\mathbb{R}^5 \times \mathbb{H}^5$	**.341**	**.498**	**.298**

KG-Product Space Querying

- KG embeddings have a **scoring function** $f_r(h,t)$ that ranks **true triples** the highest.

$$f_{\text{sibling of}}(\text{Eli Manning}, \text{Peyton Manning}) = 143$$
$$f_{\text{sibling of}}(\text{Eli Manning}, \text{Kanye West}) = 2$$
$$f_{\text{sibling of}}(\text{Eli Manning}, \text{Obama}) = -4$$

- We assessed our product space embeddings on the query metrics **mean rank (MR), mean reciprocal rank (MRR),** and **Hits@N.**
- Goal: learn vector representations of triples (h,r,t) to give **high** Hits@N, MRR, **low** MR using the loss function

$$L = -\log \sigma(\gamma - f_r(\mathbf{h}, \mathbf{t})) - \sum_{i=1}^{n} p(h'_i, r, t'_i) \log \sigma(f_r(\mathbf{h}'_i, \mathbf{t}'_i) - \gamma)$$

$$f_r(h, r, t) = -\|h + r - t\|_{\mathcal{P}}$$

where the second loss term is composed of **self-adversarial negative sampling** and the norm is taken with respect to the product space.

- Datasets: **FB15k**
  - **FB15k:** subset of the Freebase knowledge graph, mainly **symmetry, inversion** patterns.
- We embedded the data into **four spaces**, only one of which was purely Euclidean and again keeping the total dimension constant, training for 500 epochs.

	MR	MRR	Hits@1	Hits@3	Hits@10
$\mathbb{R}^{100} \times \mathbb{H}^{500}$	92.5	0.626	**0.609**	0.715	0.810
$\mathbb{R}^{500} \times \mathbb{H}^{100}$	59.8	0.671	0.564	0.752	**0.847**
$\mathbb{R}^{300} \times \mathbb{H}^{300}$	95.97	0.501	0.342	0.622	0.752
$\mathbb{R}^{600}$	**49.4**	**0.684**	0.582	**0.759**	0.845

- **Different spaces** outperformed the others at **different tasks**:
  - Product spaces **outperformed** Euclidean space at Hits@1 and Hits@10 and were close for Hits@3, but **underperformed** at MR and MRR.
- **Takeaway:** product embeddings don't necessarily lead to improved all-around performance.

Conclusion

- Knowledge graph elements are **relational triples** that can be modeled in **non-Euclidean spaces.**
- We embedded knowledge graphs into different product spaces and assessed their **distortion** and **performance** on **query tasks.**
- We found that embedding knowledge graphs in product spaces will **decrease distortion**, but will **not guarantee** improved performance on querying tasks - **some will be improved** while **others worsen.**

Future Work

- This work was severely limited by the computational costs in time and money for training various models. With more resources, we suggest various future experiments:
  - Using **Sarkar's construction** for identifying the optimal embedding size for each dataset.
  - Allow the embeddings to train **for more epochs** could cause meaningful changes in important metrics (**double descent**).
  - Perform more robust **combinatorial analysis** on hyperparameters to tune/improve performance.

References

[1] Chenyan Xiong, Russell Power, and Jamie Callan. Explicit semantic ranking for academic search via knowledge graph embedding. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 1271–1279, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee.

[2] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 353–362, New York, NY, USA, 2016. ACM.

[3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2787–2795. Curran Associates, Inc., 2013.

[4] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes, 2014.

[5] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2019.

[6] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. *CoRR*, abs/1704.08063, 2017.

Acknowledgements

We would like to thank Andrew Ng, Moses Charikar, and Chris Re for teaching CS 229, as well as the rest of the course staff for running it so smoothly. We found the course very interesting and educational and appreciated the opportunity to work on ProdKG. We would also like to thank Fred Sala and Adva Wolf for their help and insight while working on ProdKG.