Hyperbolic Entailment Cones for Learning Hierarchical Embeddings

Octavian-Eugen Ganea, Gary Bécigneul, Thomas Hofmann
Department of Computer Science, ETH Zürich, Switzerland

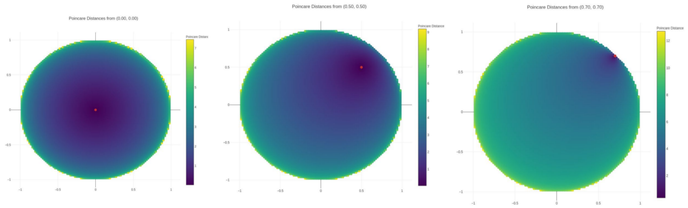
Problem Definition & Goals

Entailment:

- is-a or is-subconcept relations
- asymmetric relations/edges in a directed acyclic graph (DAG)

Goal 1: model entailment in the embedding space - learn ordered representations that respect the partial order between objects
Goal 2: use HYPERBOLIC embedding space instead of Euclidean

- exponential growing distances
- mathematically best choice for embedding hierarchies and complex networks

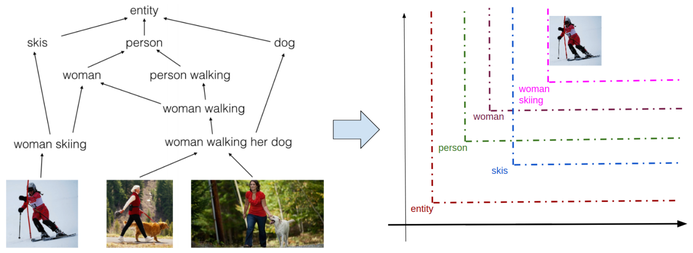


Prior Work & Baselines

1) **Poincaré embeddings** (M. Nickel and D. Kiela, 2017):

- Loss: $\mathcal{L}(\theta) = \sum_{(u,v)\in\mathcal{E}} \log \frac{e^{-d(u,v)}}{\sum_{v'\in\mathcal{N}(u)} e^{-d(u,v')}}$
- does not explicitly model asymmetric relations
- relies on heuristic: $score(is-a(u,v)) = (1+\alpha(\|u\|-\|v\|))\cdot d(u,v)$

2) **Order embeddings (OE)** (I. Vendrov et al, 2016):



- defines a partial order on $\mathbb{R}^d$ via simple **entailment cones**
- drawbacks: capacity scales linearly with d, heavy cone intersections, bounded disjoint areas

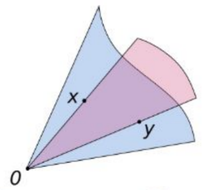
Riemannian Entailment Cones

- generalize **convex entailment cones** to any Riemannian manifold:

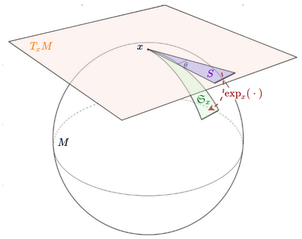
$$(u,v) \in \mathcal{E} \iff v \in \mathfrak{S}_u$$

- induce a **partial order** in the embedding space
- expand in all space directions (increased disjoint capacity)
- apply to both Euclidean and hyperbolic geometries

EUCLIDEAN CONVEX CONES
$v_1, v_2 \in S, \forall \alpha, \beta \geq 0$
$\Rightarrow \alpha v_1 + \beta v_2 \in S$



RIEMANNIAN CONVEX CONES
$S \subseteq T_x\mathcal{M},\ \mathfrak{S}_x := \exp_x(S) \subseteq \mathcal{M}$



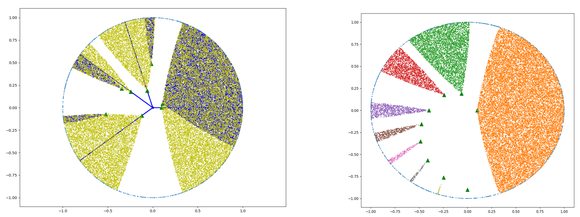
Angular Entailment Cones

We require 4 properties:
1) **Axial symmetry** (cone depends on aperture angle $\psi(x)$).

$$S_x^{\psi(x)} := \{v \in T_x\mathbb{D}^n : \angle(v, \bar{x}) \leq \psi(x)\}, \quad \mathfrak{S}_x^{\psi(x)} := \exp_x(S_x^{\psi(x)}).$$

2) **Continuous cone aperture function ψ.**
3) **Rotation invariance** - ψ depends only on $\|x\|$, i.e. $\psi(x) = \tilde{\psi}(\|x\|)$.
4) **Transitivity** of nested angular cones (partial order property).

$$\forall x, x' \in \mathbb{D}^n \setminus \{0\}: \quad x' \in \mathfrak{S}_x^{\psi(x)} \implies \mathfrak{S}_{x'}^{\psi(x')} \subseteq \mathfrak{S}_x^{\psi(x)}.$$



Theorem 1. *Cones cannot have an aperture ψ larger than $\pi/2$.*
Consequence: One has to leave an ϵ ball out of the domain of ψ.
Theorem 2. *The above properties imply in the hyperbolic space* $\mathfrak{S}_x^{\psi(x)} = \{y \in \mathbb{D}^n | \Xi(x,y) \leq \arcsin\left(K\frac{1-\|x\|^2}{\|x\|}\right)\}$, *where*

$$\Xi(x,y) := \arccos\left(\frac{\langle x,y\rangle(1+\|x\|^2)-\|x\|^2(1+\|y\|^2)}{\|x\|\cdot\|x-y\|\sqrt{1+\|x\|^2\|y\|^2-2\langle x,y\rangle}}\right)$$

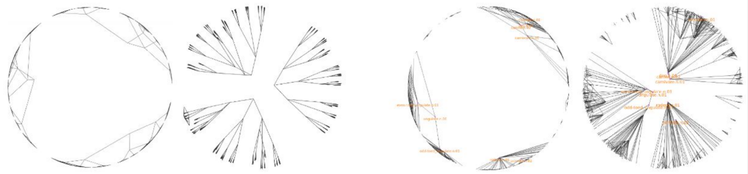
Learning with Entailment Cones

- Learning from positive and negative pairs in a DAG.
- Loss: $\mathcal{L} = \sum_{(u,v)\in P} E(u,v) + \sum_{(u',v')\in N} \max(0, \gamma - E(u',v'))$
- Penalty: $E(u,v) = \max(0, \Xi(u,v) - \psi(u))$
- Riemannian optimization:

$$u \leftarrow \exp_u(-\eta \nabla_u^R \mathcal{L}),$$

- w/ Riemannian gradient: $\nabla_u^R \mathcal{L} = (1/\lambda_u)^2 \nabla_u \mathcal{L}, \quad \lambda_u = \frac{2}{1-\|u\|^2}$
- closed form expression of exponential map (see our derivation)

Experiments (Qualitative)



Left deck: uniform tree of depth 7 and branching factor 3.
Right deck: WordNet mammal subtree (4.2K edges, 1.1K vertices).

Experiments (Quantitative)

Task: Predict edges in the transitive closure of WordNet hierarchy (82K vertices, 660K edges).

	EMBEDDING DIMENSION = 5				EMBEDDING DIMENSION = 10			
	PERCENTAGE OF TRANSITIVE CLOSURE (NON-BASIC) EDGES IN TRAINING							
	0%	10%	25%	50%	0%	10%	25%	50%
SIMPLE EUCLIDEAN EMB	26.8%	71.3%	73.8%	72.8%	29.4%	75.4%	78.4%	78.1%
POINCARÉ EMB	29.4%	70.2%	78.2%	83.6%	28.9%	71.4%	82.0%	85.3%
ORDER EMB	**34.4%**	70.2%	75.9%	81.7%	**43.0%**	69.7%	79.4%	84.1%
OUR EUCLIDEAN CONES	28.5%	69.7%	75.0%	77.4%	31.3%	81.5%	84.5%	81.6%
OUR HYPERBOLIC CONES	29.2%	**80.1%**	**86.0%**	**92.8%**	32.2%	**85.9%**	**91.0%**	**94.4%**

Euclidean cones: Similar derivation of Euclidean angular cones.

Resources

code : https://github.com/dalab/hyperbolic_cones

Website / blog: www.hyperbolicdeeplearning.com

References

- Nickel, Maximillian, and Douwe Kiela. "Poincaré embeddings for learning hierarchical representations." NIPS 2017.
- Vendrov, Ivan, et al. "Order-embeddings of images and language." arXiv preprint arXiv:1511.06361 (2015), ICLR 2016.