Overlapping Spaces for Compact Graph Representations

Kirill Shevkunov
Yandex, MIPT
Moscow, Russia
shevkunov.ks@phystech.edu

Liudmila Prokhorenkova
Yandex Research, MIPT, HSE University
Moscow, Russia
ostroumova-la@yandex.ru

NeurIPS 2021

Table of Contents

Summary

- We introduce Overlapping Spaces (OS) for embedding structured data
- Main idea: subsets of coordinates can be shared between spaces of different types (Euclidean, hyperbolic, spherical, etc.)
- OS automatically learns optimal combination of spaces
- Due to complex geometry, OS allows for more compact representations

Embeddings

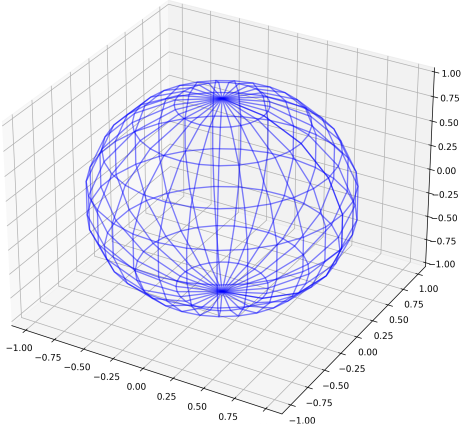
- Vector representations of objects are needed for many applications
- Distance between vector representations should preserve ‘distance’ between objects

Embeddings

- Vector representations of objects are needed for many applications
- Distance between vector representations should preserve 'distance' between objects

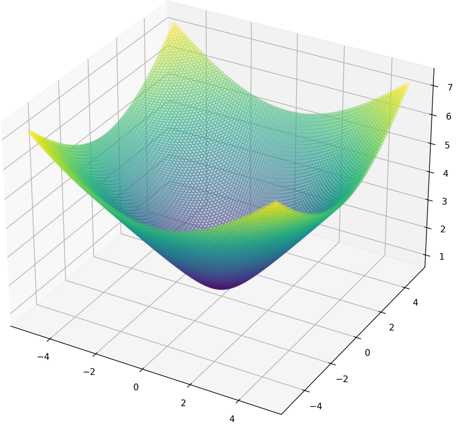
- Besides Euclidean space, what other options do we have?

Spherical space



$$\mathbb{S} = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$$
$$d_S(x, y) = \arccos(x^T y)$$

Hyperbolic space: Lorentz Model



$$\langle x, y\rangle_H = x_1 y_1 - \sum_{i=2}^{n+1} x_i y_i$$

$$\mathbb{H} = \{x \in \mathbb{R}^{n+1} | \langle x, y\rangle_H = 1, x_1 \geq 0\},\ d_H(x, y) = \operatorname{arccosh}(\langle x, y\rangle_H)$$

Product space

Definition

Given metric spaces $(M_1, d_{M_1}), \dots, (M_n, d_{M_n})$,
product space is defined as $M_1 \times \dots \times M_n$ equipped with distance function $d_{l1} = \sum_{i=1}^{n} d_{M_i}$ or $d_{l2}^2 = \sum_{i=1}^{n} d_{M_i}^2$.

Examples: $\mathbb{S}_5 \times \mathbb{S}_5$, $\mathbb{S}_5 \times \mathbb{H}_5$, $\mathbb{S}_2 \times \mathbb{S}_2 \times \mathbb{H}_2 \times \mathbb{H}_2 \times \mathbb{E}_2$, ...

Product space

Definition

Given metric spaces $(M_1, d_{M_1}), \dots, (M_n, d_{M_n})$,
product space is defined as $M_1 \times \dots \times M_n$ equipped with
distance function $d_{l1} = \sum_{i=1}^{n} d_{M_i}$ or $d_{l2}^2 = \sum_{i=1}^{n} d_{M_i}^2$.

Examples: $\mathbb{S}_5 \times \mathbb{S}_5$, $\mathbb{S}_5 \times \mathbb{H}_5$, $\mathbb{S}_2 \times \mathbb{S}_2 \times \mathbb{H}_2 \times \mathbb{H}_2 \times \mathbb{E}_2$, ...

Problem: the best combination of spaces depends on the dataset.

Gu, Sala, Gunel, Ré, "Learning mixed-curvature representations in product spaces", ICLR 2018.

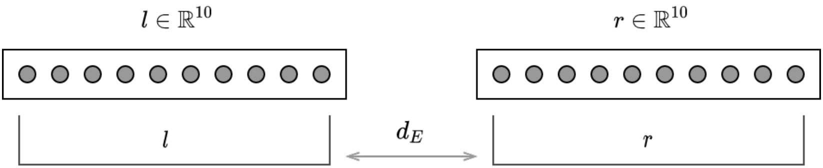
Our approach

Overlapping spaces generalize product spaces

Importantly, they do not require signature brute-forcing

From single to overlapping space

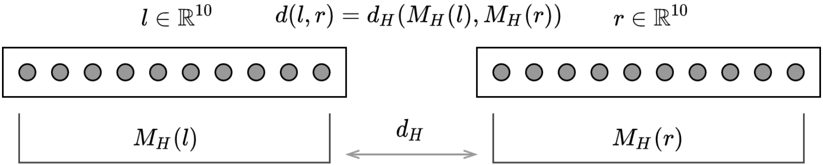
For two vectors we can compute the Euclidean distance ($\mathbb{E}^{10}$):



$$d_E^2(l, r) = \sum_{i=1}^{10} (l_i - r_i)^2$$

From single to overlapping space

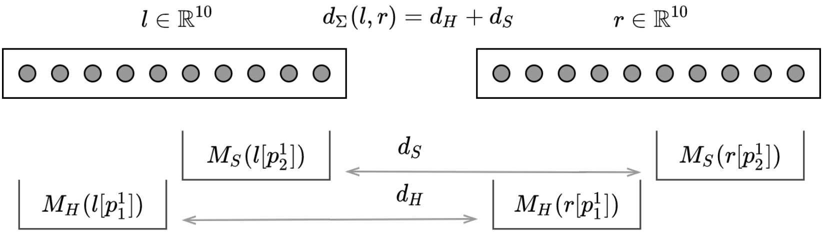
For spherical and hyperbolic spaces, we can use differentiable mappings $M_S(x) : \mathbb{R}^{10} \to \mathbb{S}^{10}$ and $M_H(x) : \mathbb{R}^{10} \to \mathbb{H}^{10}$:



With such parameterization, we can use any conventional optimizer

From single to overlapping space

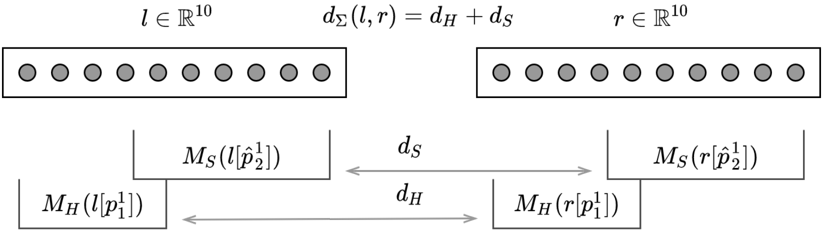
In a similar way, we may construct a product space, e.g., $\mathbb{H}_5 \times \mathbb{S}_5$:



Here $p_1^1 = [1..5]$, $p_2^1 = [6..10]$

From single to overlapping space

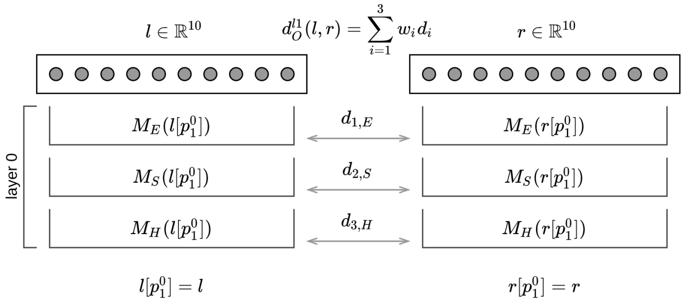
Let us allow the mappings **overlap**:



We constructed a subspace of product space $\mathbb{H}_5 \times \mathbb{S}_6$ with only 10 parameters per item

From single to overlapping space

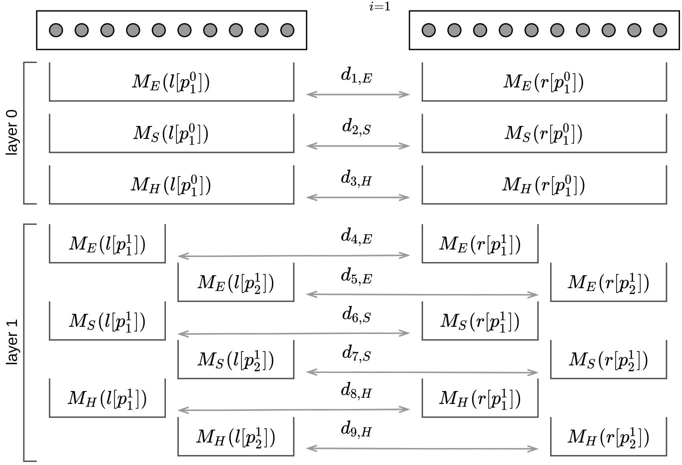
Proposed implementation, layer 0:



From single to overlapping space

Layer 1:

$l \in \mathbb{R}^{10}$ $\qquad d_O^{l1}(l, r) = \sum_{i=1}^{9} w_i d_i$ $\qquad r \in \mathbb{R}^{10}$



$l[p_1^1] = (l_1 \dots l_5)$ $\qquad r[p_2^1] = (r_6 \dots r_{10})$

OS-Mixed

Statement

Overlapping space is a metric space.

OS-Mixed

Statement

Overlapping space is a metric space.

Extension: add shifted weighted inner product (WIPS) to OS

Shifted WIPS: $d_{WIPS}(x, y) = c - \sum w_i x_i y_i$

$d_{OS-Mixed} = d_{OS} + d_{WIPS}$

Experiments

Time for some experiments!

Graph reconstruction

Table: Graph reconstruction

	UCSA312	CS PhDs	Power	Facebook	WLA6	EuCore
Nodes	312	1025	4941	4039	3227	986
Edges	48516 (weighted)	1043	6594	88234	3604	16687

Graph reconstruction

Table: Graph reconstruction

	UCSA312	CS PhDs	Power	Facebook	WLA6	EuCore
Nodes	312	1025	4941	4039	3227	986
Edges	48516 (weighted)	1043	6594	88234	3604	16687

Quality measures: distortion D_{avg} and mean average precision (mAP)

$$D_{avg} = \frac{2}{|V|(|V|-1)} \sum_{(v,u)\in V^2, v\neq u} \frac{|d_U(e(v), e(u)) - d_G(v,u)|}{d_G(v,u)}$$

Here d_G — original distance, d_U — distance between vector representations

Graph reconstruction (distortion)

Table: Distortion graph reconstruction, top results are highlighted, top metric results are underlined.

Signature	UCSA312	CS PhDs	Power	Facebook	WLA6	EuCore
E_{10}	**0.00318**	0.0475	0.0408	0.0487	0.0530	0.1242
H_{10}	0.01104	0.0443	0.0348	0.0483	0.0279	0.1144
S_{10}	0.01065	0.0519	0.0453	0.0561	0.0608	0.1260
$H_5^2 \equiv H_5 \times H_5$	0.00573	0.0345	0.0255	0.0372	0.0279	0.1106
$S_5 \times S_5 \equiv S_5^2$	0.00700	0.0501	0.0438	0.0552	0.0584	0.1251
$H_5 \times S_5$	0.00541	0.0341	0.0254	0.0346	0.0310	0.1195
H_2^5	0.00592	0.0344	0.0273	0.0439	0.0356	0.1163
S_2^5	0.00604	0.0464	0.0416	0.0512	0.0543	0.1244
$H_2^2 \times E_2 \times S_2^2$	0.00537	0.0344	0.0302	0.0406	0.0437	0.1193
$O_{l1}, t = 0$	**0.00324**	0.0368	0.0281	0.0458	0.0286	0.1141
$O_{l1}, t = 1$	0.00325	**0.0300**	**0.0231**	0.0371	0.0272	0.1117
$O_{l2}, t = 1$	0.00530	0.0328	**0.0246**	0.0324	0.0278	0.1127
$c - \text{dot}$	0.04005	0.0412	0.0461	0.0236	0.0296	0.1085
$c - \text{wips}$	0.06468	0.0358	0.0442	**0.0161**	**0.0238**	**0.1016**
$ce^{-\text{dot}}$	0.08142	0.0424	0.0505	0.0192	0.0270	0.1048
$O_{mix-l1}, t = 1$	**0.00277**	**0.0243**	**0.0235**	**0.0172**	**0.0187**	**0.1026**
$O_{mix-l2}, t = 1$	0.00464	**0.0220**	0.0258	**0.0163**	**0.0198**	**0.1028**

Graph reconstruction (mAP)

Table: mAP graph reconstruction, top results are highlighted, top metric results are underlined.

Signature	UCSA312	CS PhDs	Power	Facebook	WLA6	EuCore
E_{10}	0.9290	0.9487	0.9380	0.7876	0.7199	0.6108
H_{10}	0.9173	0.9399	0.9385	0.7997	0.9617	0.6670
S_{10}	0.9183	0.9519	0.9445	0.7768	0.7289	0.6037
H_5^2	0.9247	0.9481	0.9415	0.8084	$\underline{0.9682}$	$\underline{0.6783}$
S_5^2	0.9316	0.9600	0.9482	0.7790	0.7307	0.6116
$H_5 \times S_5$	0.9397	0.9538	0.9505	0.7947	$\underline{0.9751}$	$\underline{0.6847}$
H_2^5	0.9364	0.9671	0.9508	0.7979	0.8597	0.6611
S_2^5	0.9439	0.9656	0.9511	0.7800	0.7358	0.6169
$H_2^2 \times E_2 \times S_2^2$	0.9519	0.9638	0.9507	0.7873	0.7794	0.6492
$O_{l_1}, t = 0$	$\underline{\mathbf{0.9538}}$	$\underline{0.9879}$	$\underline{0.9728}$	$\underline{0.8093}$	0.6759	0.6580
$O_{l_1}, t = 1$	$\underline{\mathbf{0.9522}}$	$\underline{\mathbf{0.9904}}$	$\underline{0.9762}$	$\underline{0.8185}$	0.9598	0.6691
$O_{l_2}, t = 1$	$\underline{\mathbf{0.9522}}$	$\underline{\mathbf{0.9938}}$	$\underline{0.9907}$	$\underline{0.8326}$	$\underline{0.9694}$	$\underline{0.7078}$
c − dot	**1**	**1**	**0.9983**	**0.8745**	**0.9990**	0.7409
c − wips	**1**	**1**	**1**	0.8704	**1**	**0.7742**
$O_{mix-l_1}, t = 1$	**1**	**1**	**0.9994**	**0.8806**	**0.9997**	**0.7860**
$O_{mix-l_2}, t = 1$	**1**	**1**	**1**	**0.9021**	**1**	**0.8405**

Weight analysis

For many datasets, more than half of the weights are near-zero, so unnecessary components can be removed.

For Power dataset:

$$d_{O_{l1},t=1}(l,r) \propto 0.1 d_E(l_1^0, r_1^0) + 0.5 d_R(l_1^0, r_1^0) + 0.4 d_H(l_1^1, r_1^1),$$

$$\text{where } l_1^0 = l[0..5], l_1^1 = l[6..10]$$

Information Retrieval with DSSM

We trained classical DSSM with different distance functions on Wikipedia search dataset:

Table: Search query examples

Query	Web site
Kris Wallace	en.wikipedia.org/wiki/Chris_Wallace
1980: Mitsubishi produces one million cars...	en.wikipedia.org/wiki/Mitsubishi_Motors
code napoleon	en.wikipedia.org/wiki/Napoleonic_Code

Information Retrieval with DSSM

Table: DSSM results, dimension 10, top three results are highlighted

Signature	Test mAP
E_{10}	0.4459
H_{10}	0.4047
S_{10}	0.4364
H_5^2	0.4492
S_5^2	**0.4573**
$H_5 \times S_5$	0.3295
H_2^5	0.3681
S_2^5	**0.4616**
$H_2^2 \times E_2 \times S_2^2$	0.3526
$c - \text{dot}$	0.4194
$O_{l1}, t = 0$	**0.4562**
$O_{l1}, t = 1$	0.4498
$O_{l2}, t = 1$	0.4456
$O_{mix-l1}, t = 1$	0.4447
$O_{mix-l2}, t = 1$	0.4483

To sum up

1. We propose overlapping spaces that allow for parameter sharing between different sub-spaces
2. OS automatically learns optimal combination of combined spaces
3. OS allows for compact representations in various setups
4. The proposed method can be combined with non-metric similarities leading to even more flexibility