AdKDD 2020

NAVER CLOVA

Multi-Manifold Learning for Large-Scale Targeted Advertising System

Kyuyong Shin[1], Young-Jin Park[2], Kyung-Min Kim[1], Sunyoung Kwon[1]

[1] *Clova AI Research, NAVER Corp.*

[2] *Naver R&D Center, NAVER Corp.*

INTRODUCTION

INTRODUCTION

- Messenger advertisements (ads) give direct personal user experience.
- However, an ad for broad random users without precise user targeting cannot resonate with their potential audience playing as annoying spam.
- In this paper, since the accurate representation of the users and ads is a necessary for targeted advertising system, we present the deep representation learning scheme using hyperbolic geometries.

Figure 1: LINE messenger advertisement system.

What is hyperbolic space?

and

Why is hyperbolic manifold well-suited space for hierarchical structure?

INTRODUCTION

- Hyperbolic space is a non-Euclidean space with a constant negative Gaussian curvature.

- Hyperbolic space is often associated with Minkowski space time in special relativity.

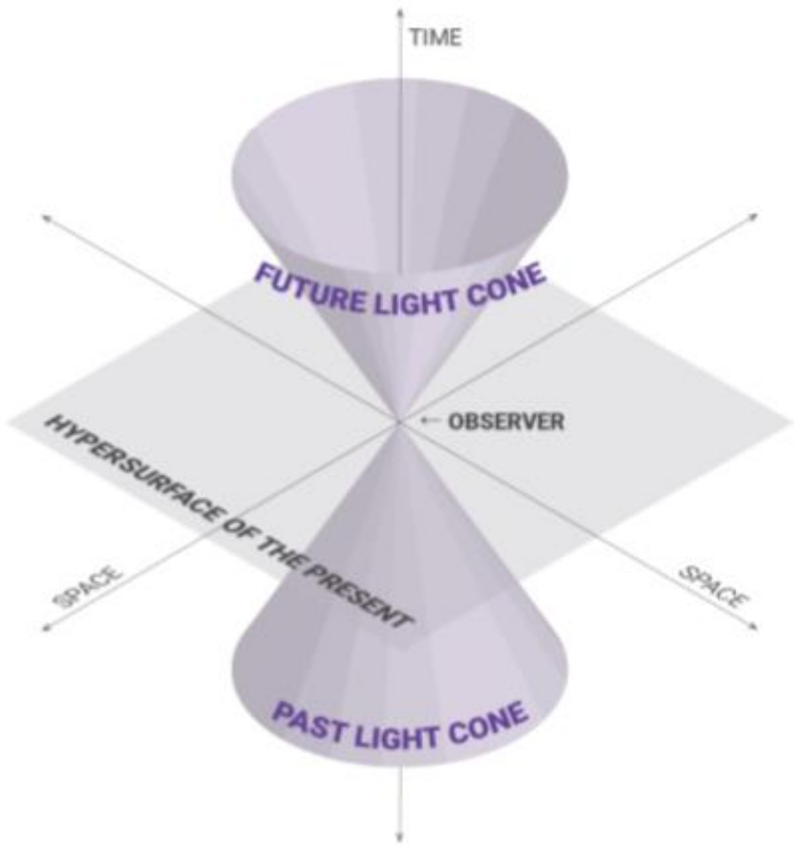


Figure 2: Minkowski space.

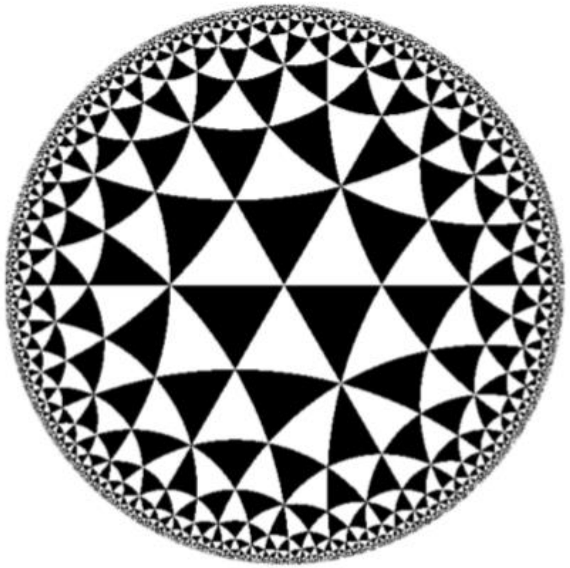


Figure 3: M.C. Escher style illustration of the Poincaré disk model.

Each triangle has constant area in hyperbolic space, but in Euclidean space, it rapidly shrinks at the boundary.

TARGETED ADVERTISING SYSTEM

$$\textit{User attributes } X_u \in \mathbb{R}^{N_u \times F_u}, \qquad \textit{MLP Layer for users } f_u: \mathbb{R}^{F_u} \rightarrow \mathbb{R}^H$$

$$\textit{Ads attributes } X_a \in \mathbb{R}^{N_a \times F_a}, \qquad \textit{MLP Layer for ads } f_a: \mathbb{R}^{F_a} \rightarrow \mathbb{R}^H$$

$$Z_u = f_u(X_u) \in \mathbb{R}^{N_u \times H}, \qquad Z_a = f_a(X_a) \in \mathbb{R}^{N_a \times H}$$

Click history matrix C into $Z_h = f_h(C) \in \mathbb{R}^{N_u \times H}$ where $c_{i,j}$ is one if there is positive interaction between the i-th user and j-th advertisement and is zero otherwise.

f_h can be any neural networks model (e.g, Transformer, LSTM, MLP etc..)

$$P_{u,a} = Decision(dist(Z_u + Z_h, Z_a))$$

Where $dist(p,q)$ is the distance between two points p and q on the given manifold.

User preference scores are sorted for each ad, and the top k users with the highest scores are selected as the targeted users for the ad

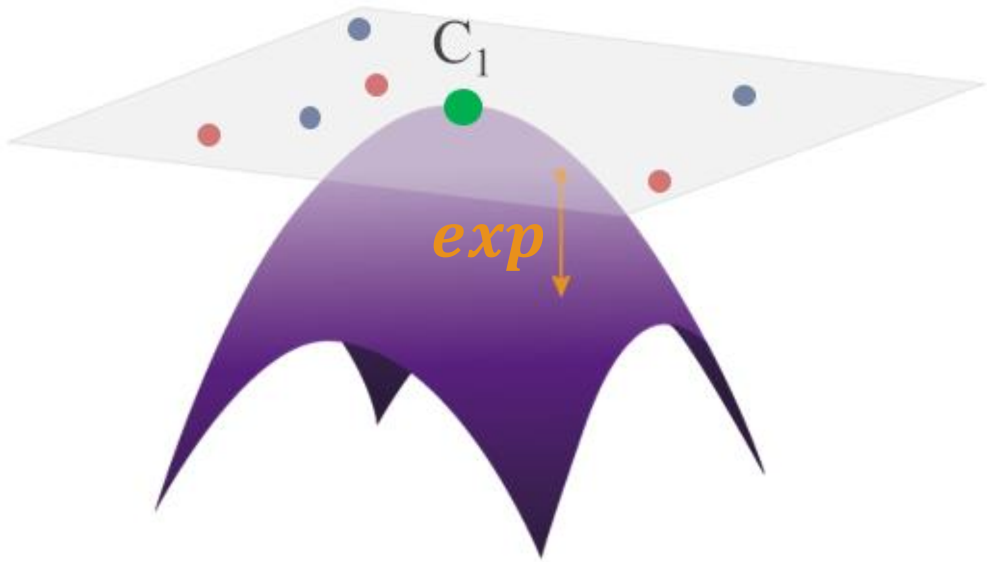
HYPERBOLIC GEOMETRY

- Exponential map & Logarithm map
- Diffeomorphism

HYPERBOLIC GEOMETRY

- Exponential map & Logarithm map
- Diffeomorphism

Exponential map & Logarithm map



$$exp_p^K(v) = \cosh(\frac{||v||_{g_M}}{\sqrt{K}})p + \sqrt{K}\sinh(\frac{||v||_{g_M}}{\sqrt{K}})\frac{v}{||v||_{g_M}} \quad (3)$$

$$log_p^K(q) = d_g^K(p,q)\frac{q + \frac{1}{K} < p,q >_{g_M} p}{||q + \frac{1}{K} < p,q >_{g_M} p||}, \quad (4)$$

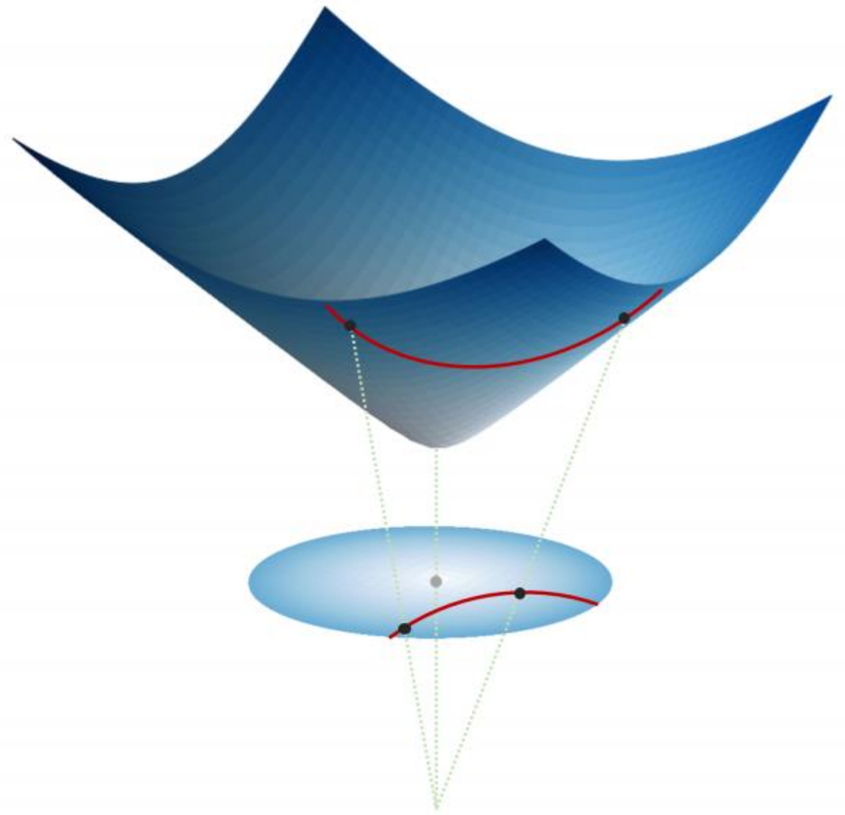
where $||v||_{g_M} = \sqrt{< v, v >_{g_M}}$ denotes norm of $v \in T_pH^K$ and $d_g^K(p,q) = \sqrt{K}\text{arcosh}(- < p,q >_{g_M} /K)$ denotes geodesic distance between p and q. Above expressions assume that $\gamma_{p\to v}^K(t) = \cosh(\frac{t}{\sqrt{K}})p + \sqrt{K}\sinh(\frac{t}{\sqrt{K}})v$, when t is small enough and tangent vector v is unit-speed, i.e. $< v, v >_{g_M} = 1$.

HYPERBOLIC GEOMETRY

- Exponential map & Logarithm map
- Diffeomorphism

Diffeomorphism

$$\Psi(x_0, x_1 \ldots, x_d) = \frac{\sqrt{K}(x_1, x_2 \ldots, x_d)}{x_0 + \sqrt{K}},$$



Stereographic projection of Hyperboloid
to Poincaré disk[1]

[1]Chami et al. 2019. https://arxiv.org/pdf/1910.12933.pdf

MULTI-MANIFOLD LEARNING

MULTI-MANIFOLD LEARNING

In a real-world, large-scale advertising system, there exist various groups of users with different preference characteristics, and it may not be valid to assume that every user and advertisement entity can be expressed by using single geometry.

MULTI-MANIFOLD LEARNING

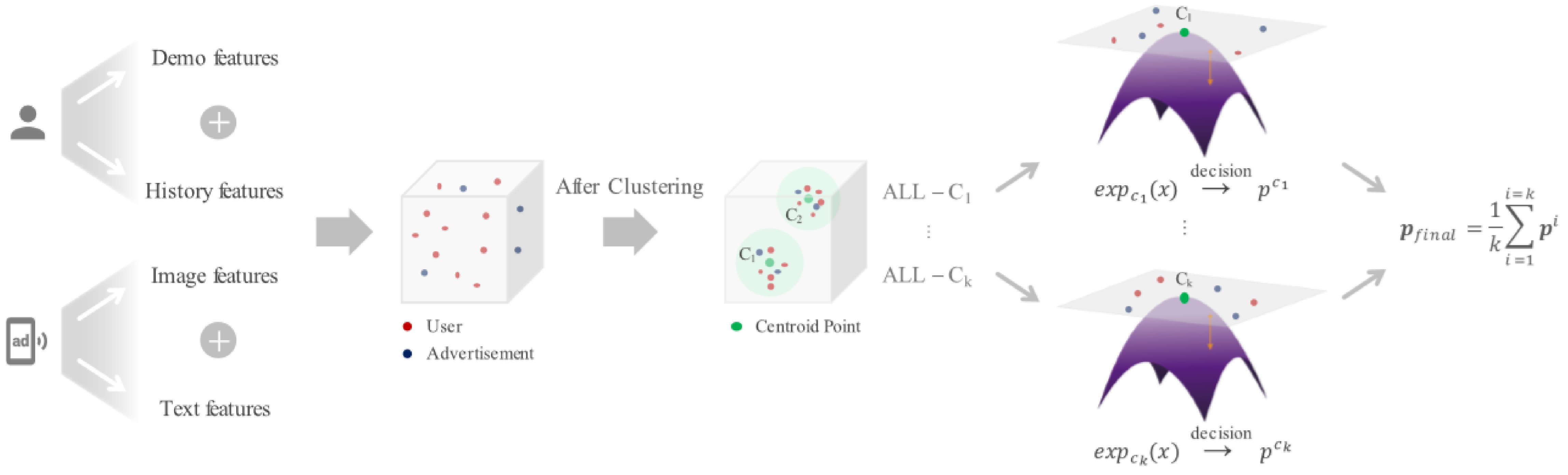


Figure 3: Conceptual scheme of our proposed method.

MULTI-MANIFOLD LEARNING

$$E_u^t = (Z_u + Z_h) - c_t, \quad E_a^t = Z_a - c_t \tag{6}$$

$$exp_{\mathbf{o}}^K(E) = \left(\sqrt{K} cosh(\frac{\|E\|_{Euc}}{\sqrt{K}}), \sqrt{K} sinh(\frac{\|E\|_{Euc}}{\sqrt{K}}) \frac{E}{\|E\|_{Euc}}\right), \tag{7}$$

$$p_{u,a}^t = (1 + \exp^{(d_g^K(E_{u,\ h}^t, E_{a,\ h}^t) - s)/b})^{-1}, \quad E_h = exp_{\mathbf{o}}^K(E) \tag{8}$$

$$\mathcal{L} = \sum \mathrm{BCE}(P_{u,a}, Y), \tag{9}$$

EXPERIMENT

- Dataset
- Performance comparison
- Effects of the number of clusters
- Embedding visualization

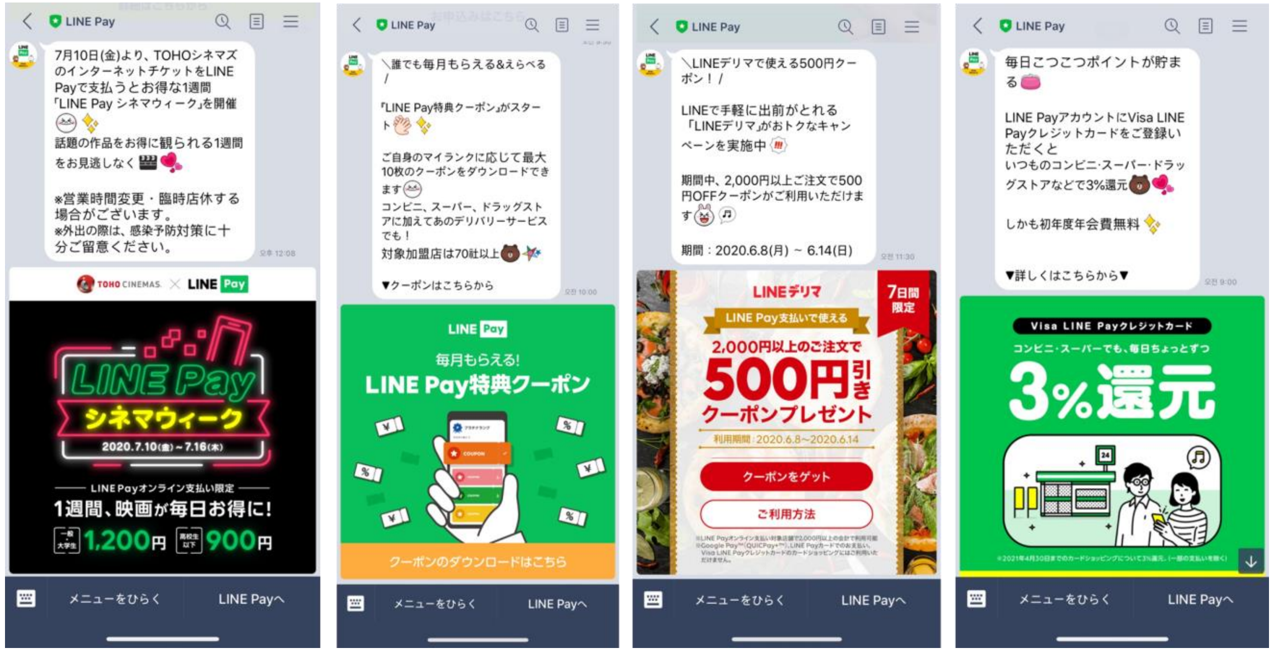
EXPERIMENT

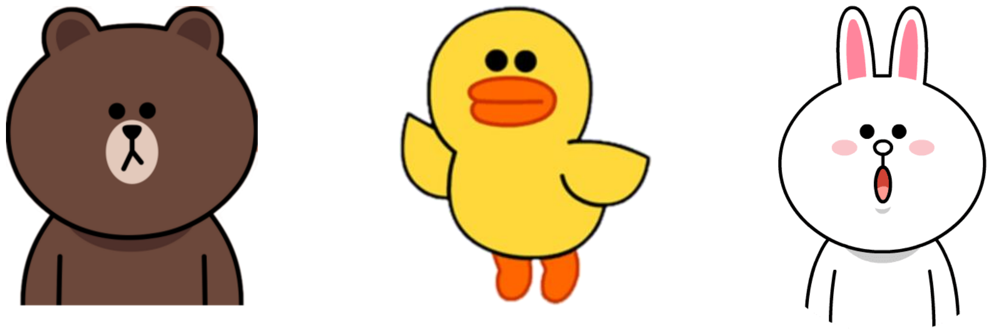
- Dataset
- Performance comparison
- Effects of the number of clusters
- Embedding visualization

Dataset

50+ Ads

200M+ Users





EXPERIMENT

- Dataset
- Performance comparison
- Effects of the number of clusters
- Embedding visualization

Performance comparison

Table 1: Model Performance on LINE messenger advertisement system

	RocAuc	Accuracy	Average Precision	Shannon Entropy
CF	0.756	0.673	0.786	14.431
MLP	0.770	0.681	0.775	14.197
HNN	0.778	0.753	0.841	14.451
Multi-Manifold	**0.818**	**0.765**	**0.846**	**14.567**

Table 2: Model Performance on public benchmark MovieLens dataset

	MovieLens - 1M		MovieLens - 100K	
	RocAuc	Average Precision	RocAuc	Average Precision
CF	60.3	67.4	60.5	61.1
MLP	57.4	66.3	61.7	62.0
HNN	**61.7**	69.0	68.0	67.8
Multi-Manifold	61.5	**69.8**	**68.3**	**68.5**

Table 1: Our method shows the best prediction performance for all accuracy metrics, as well as the diversity metric.

Table 2: Our model shows the best or second-best performance, demonstrating its generality not overfitted to a certain dataset.

EXPERIMENT

- Dataset
- Performance comparison
- Effects of the number of clusters
- Embedding visualization

Effects of the number of clusters

Table 3: Model performance comparison as the number of cluster increases on LINE messenger dataset.

# of Clusters	RocAuc	Accuracy	Average Precision	Shannon Entropy
1-cluster	0.798	0.715	0.817	13.871
3-cluster	0.805	0.753	0.840	14.146
5-cluster	**0.818**	**0.765**	**0.846**	14.567
10-cluster	0.813	0.753	0.841	14.653
15-cluster	0.810	0.753	0.841	**14.665**

Table3: shows that the overall performance improves as the cluster grows, and the best performance is obtained at $T = 5$. After $T = 5$, the prediction accuracy converges while the diversity improves further.

EXPERIMENT

- Dataset
- Performance comparison
- Effects of the number of clusters
- Embedding visualization

Embedding visualization

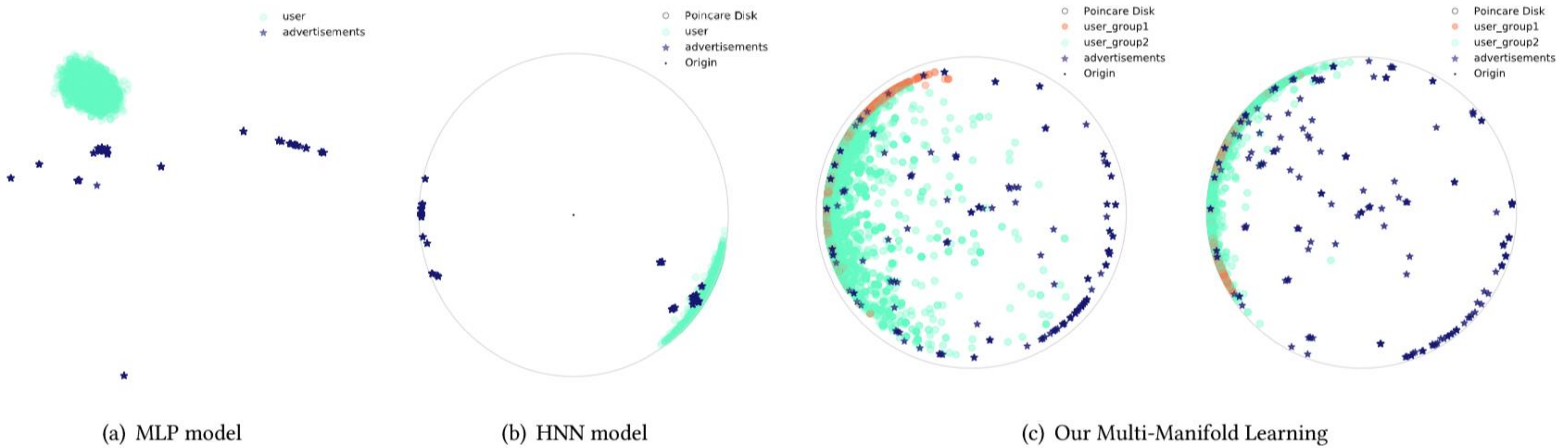


Figure 5: Visualization in embedding representations of users and advertisements. (a) embedding of Euclidean MLP, (b) Poincaré disk visualization of HNN, (c) Poincaré disk visualization of our Multi-Manifold Learning with two clusters on each manifold. We visualize them by using diffeomorphism between Hyperbolic space and Poincaré space as described in Section 3.2.

CONCLUSION

CONCLUSION

Traditional targeted advertising systems struggle with data representation capabilities because of the inherent limitation of Euclidean space. To tackle this issue, we present Multi-Manifold Learning, a well-designed technique to learn better representation of users and advertisements. Experimental results show the proposed scheme improves the targeted advertising quality in terms of both accuracy and diversity. As the future directions, we will develop a Multi-Manifold Learning scheme in terms of diffeomorphism learning.

Multi-Manifold Learning for Large-Scale Targeted Advertising System

Kyuyong Shin[1], Young-Jin Park[2], Kyung-Min Kim[1], Sunyoung Kwon[1]

[1] *Clova AI Research, NAVER Corp.*

[2] *Naver R&D Center, NAVER Corp.*

Thank you

Paper url QR

NAVER CLOVA