A Fashion Item Recommendation Model in Hyperbolic Space (Supplementary Material)

Ryotaro Shimizu[1,2], Yu Wang[1], Masanari Kimura[3],
Yuki Hirakawa[2], Takashi Wada[2], Yuki Saito[2], Julian McAuley[1]
[1]University of California, San Diego [2] ZOZO Research [3] The University of Melbourne
{r2shimizu, yuw164}@ucsd.edu, m.kimura@unimelb.edu.au,
{yuki.hirakawa,takashi.wada,yuki.saito}@zozo.com, jmcauley@ucsd.edu

1. Methodology

Out of five isometric models for modeling hyperbolic space [1, 2, 6, 15]: the Poincaré ball model, the Poincaré half-space model, the Lorentz model, the Klein model, and the Hemisphere model, this paper adopts the Poincaré ball model, one of the most common models in computer vision [6, 8, 9, 13].

The Poincaé ball in n-dimensional hyperbolic space is described by the manifold $\mathbb{D}^n := \{\boldsymbol{q} \in \mathbb{R}^n : \|\boldsymbol{q}\|^2 < 1\}$ and n-dimensional Riemannian metric $g^{\mathbb{D}}$, which are denoted as $(\mathbb{D}^n, g^{\mathbb{D}})$. Here, the Reimannian metric $g^{\mathbb{D}}$ is given by $g^{\mathbb{D}}_q = \lambda_{\boldsymbol{q}}^2 g^E$, where $\lambda_{\boldsymbol{q}} := \frac{2}{1-\|\boldsymbol{q}\|^2}$ is the conformal factor at a point $\boldsymbol{q}$ and $g^E = \boldsymbol{I}_n$ is the Euclidean metric tensor. In other words, the hyperbolic metric tensor is conformal to the Euclidean one.

The induced geodesic distance (length of the shortest path) between two points $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n$ is given as follows:

$$d_{\mathbb{D}}(\boldsymbol{x}, \boldsymbol{y}) = \cosh^{-1}\left(1 + 2\frac{\|\boldsymbol{x}-\boldsymbol{y}\|^2}{(1-\|\boldsymbol{x}\|^2)(1-\|\boldsymbol{y}\|^2)}\right). \quad (1)$$

Note that the geodesic is a straight line connecting two points in the Euclidean space.

Because hyperbolic space is based on curved geometry, standard vector operations such as vector addition or multiplication are difficult to define in hyperbolic space. This is because it needs to be modified appropriately to match the curvature of the hyperbolic space and its properties. Therefore, to facilitate vector operations, the algebraic formalism provided by the Möbius gyrovector space [19] is introduced for the Riemannian geometry of the Poincaré ball model of the hyperbolic space [7].

Consequently, in Möbius gyrovector space, the Poincaré ball is defined as $\mathbb{D}^n_c := \{\boldsymbol{q} \in \mathbb{R}^n : c\|\boldsymbol{q}\|^2 < 1\}$ with the curvature $-c$ for $c \geq 0$. The comfort factor is defined as $\lambda^c_{\boldsymbol{q}} = \frac{2}{1-c\|\boldsymbol{q}^2\|}$. Note that $\mathbb{D}^n_c = \mathbb{R}^n$ when $c = 0$. In addition, the Möbius addition of the two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{D}^n_c$ is defined as follows:

$$\boldsymbol{x} \oplus_c \boldsymbol{y} := \frac{(1 + 2c\langle \boldsymbol{x}, \boldsymbol{y}\rangle + c\|\boldsymbol{y}\|^2)\boldsymbol{x} + (1 - c\|\boldsymbol{x}\|^2)\boldsymbol{y}}{1 + 2c\langle \boldsymbol{x}, \boldsymbol{y}\rangle + c^2\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2}. \quad (2)$$

Note that when $c = 0$, the addition defined in Eq. 2 recovers the Euclidean addition of two vectors in $\mathbb{R}^n$.

With Möbius gyrovector space operations, the induced geodesic distance between $\boldsymbol{x}$ and $\boldsymbol{y}$ in the hyperbolic space $\mathbb{D}^n_c$ is defined as follows:

$$d_c(\boldsymbol{x}, \boldsymbol{y}) := \frac{2}{\sqrt{c}} \tanh^{-1}\left(\sqrt{c}\|(-\boldsymbol{x}) \oplus_c \boldsymbol{y}\|\right), \quad (3)$$

In addition, Eq. 4 below allows us to navigate a point $\boldsymbol{z} \in T_{\boldsymbol{q}}\mathbb{D}^n_c$ from a n-dimensional vector space $T_{\boldsymbol{q}}\mathbb{D}^n_c$, which is a tangent space of $\mathbb{D}^n_c$ at a point $\boldsymbol{q} \in \mathbb{D}^n_c$, to the hyperbolic space $\mathbb{D}^n_c$.

$$h(\boldsymbol{z}) = \exp^c_{\boldsymbol{q}}(\boldsymbol{z}) := \boldsymbol{q} \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda^c_{\boldsymbol{q}}\|\boldsymbol{z}\|^2}{2}\right)\frac{\boldsymbol{z}}{\sqrt{c}\|\boldsymbol{z}\|}\right). \quad (4)$$

Furthermore, Eq. 5 bellow allows us to project a point $\boldsymbol{z}' \in \mathbb{D}^n_c$ to $T_{\boldsymbol{q}}\mathbb{D}^n_c$.

$$\log^c_{\boldsymbol{q}}(\boldsymbol{z}') := \frac{2}{\lambda^c_{\boldsymbol{q}}\sqrt{c}} \tanh^{-1}(\sqrt{c}\| - \boldsymbol{q} \oplus_c \boldsymbol{z}'\|)\frac{-\boldsymbol{q} \oplus_c \boldsymbol{z}'}{\| - \boldsymbol{q} \oplus_c \boldsymbol{z}'\|}. \quad (5)$$

Note that it holds $\log^c_{\boldsymbol{q}}(\exp^c_{\boldsymbol{q}}(\boldsymbol{z})) = \boldsymbol{z}$.

2. Experiments

2.1. Settings and Implementation Details

- **Bayesian Personalized Ranking Matrix Factorization (BPRMF)** [16] is a standard recommendation model that learns interactions between users and items using BPR loss.

- **Hyperbolic Bayesian Personalized Ranking (HBPR)** [21] learns interaction information between users and items based on BPR loss using hyperbolic distance. This is the first study that incorporates hyperbolic geometry into a recommendation model. The Poincaré ball model is used as an isometric model to model hyperbolic space.

- **Hyperbolic Recommender System (HRec)** [3] uses the Lorentz model to model hyperbolic space. Although it has a simple structure using BPR training with hinge loss, it easily scales on large data.

- **Lorentzian Factorization Machine (LFM)** [23] models the user-item interactions by using a triangle inequality using the Lorentz distance instead of the distance between two points. It works well with a small number of parameters.

- **Hyper Metric Learning (HML)** [22] is a metric learning recommendation model that learns user-item interaction using hinge loss in hyperbolic space. It also calculates the loss based on the Euclidean distance, and we adopt this approach in our proposed model. The Poincaré ball model is used to model hyperbolic space.

- **Visual Bayesian Personalized Ranking (VBPR)** [10] is a standard visually-aware recommendation model that incorporates a latent content-based preference factor, which is built on BPRMF.

- **Deep Visual Bayesian Personalized Ranking (DVBPR) [11]** is an end-to-end model that trains a visual feature extractor, which is built on VBPR. Due to our limited computational resources, we used a model that excludes the GAN structure. This simplified structure is also adopted in [14] and achieves good accuracy.

- **DeepStyle** [12] is a model based on VBPR that learns user preferences using both visual feature vectors and category vectors. Since we focus only on user-item interaction and visual information, we do not use category information and instead employ trainable vectors that cross all categories.

- **Attentive Collaborative Filtering (ACF)** [4] utilizes user preference vectors based on visual information, and its structure inspires our proposed model. Due to its computational complexity, we only use the item-level attention mechanism.

We chose all the models (VBPR, DVBPR, DeepStyle, and ACF) listed in Section 3.1 "visually-aware model-based collaborative filtering" in the fashion recommendation survey paper [5] as our baselines with item visual features. In addition, we chose all the models (HBPR, HRec, LFM, and HML) listed in the hyperbolic "recommender systems" in Chapter 4 of the hyperbolic deep neural networks survey paper [15] as our baselines trained in hyperbolic space.

2.2. Ablation Studies

- $d_c(\cdot) \Rightarrow d_{\mathrm{euc}}(\cdot)$ is a model trained using the Euclidean distance instead of the hyperbolic one. This means that $\mathcal{L}_{\mathrm{adj}}$ always becomes zero.

- **w/o** $\mathcal{L}_{\mathrm{adj}}$ is a model that removes $\mathcal{L}_{\mathrm{adj}}$ from the overall multi-task loss and only includes $\mathcal{L}_{\mathrm{hyp}}$. In other words, it is equivalent to setting $\gamma = 0$ and only using hyperbolic distance.

- **w/o aggregation** is a model without the neighbor-attentive aggregation mechanism. In other words, it is a model trained only with the user vector $\boldsymbol{u}_i$ and the item vectors $\boldsymbol{v}_j$, $\boldsymbol{v}_k$.

- **w/o attention** is a model that calculates attention scores based on the uniform distribution.

- **attention w/o** $E(\boldsymbol{I}_l)$ is a model that does not take image features $E(\boldsymbol{I}_l)$ as input in the visually-aware attention mechanism.

- **attention w/o** $\boldsymbol{v}_l$ is a model in which the item vector $\boldsymbol{v}_l$ is removed in the visually-aware attention mechanism.

- **attention w/o** $\boldsymbol{p}_l$ is a model in which the auxiliary information vector $\boldsymbol{p}_l$ is removed in the visually-aware attention mechanism.

3. Discussion

3.1. Impacts of Hyperparameters

Figure 1 shows that our model performance slightly changes with different values for the hyperparameter L.

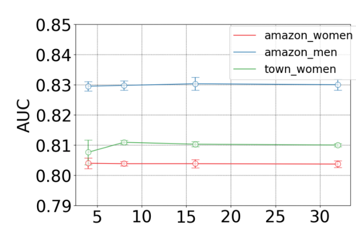


Figure 1. Our model performance with different values for the hyperparameter L (= #neighbor items).

Figure 2. Examples of user preference attention scores on Amazon Women.

3.2. Analysis of Preference Profile

Figure 2 shows examples of a list of: (1) items purchased by a certain user; (2) the contribution (attention) scores obtained by our proposed model for each item; and (3) the top recommended items during inference. Items that were actually purchased in the test data are surrounded by a red frame.

For example, given the purchase history and the preference scores of User 1, we can interpret that our proposed model pays close attention to tight silhouette tops and tight spats that User 1 purchased in the past, and these types of items are actually included in the list of recommended items.

3.3. Analysis of Embeddings

Figures 3–6 show the embedding-norm distributions and 2D projections of the user and item embeddings trained on the Amazon Men and TOWN Women data sets.

4. Limitations

In this study, we adopted the most common Poincaré ball model to incorporate hyperbolic geometry into a visual recommendation model, but it remains unclear how well our model performs with other hyperbolic-space models. Our model uses purchase history and item visual data as features but does not consider text information such as item tags and descriptions, which can provide important information as shown in previous work [17, 18].

References

[1] Eugenio Beltrami. Teoria fondamentale degli spazii di curvatura costante. *Annali di Matematica Pura ed Applicata (1867-1897)*, 2:232–255, 1868. 1

[2] James W. Cannon, William J. Floyd, Richard Kenyon, and Walter R. Parry. Hyperbolic geometry. *Flavors of Geometry*, 31:59–115, 1997. 1

[3] Benjamin Paul Chamberlain, Stephen R Hardwick, David R Wardrope, Fabon Dzogang, Fabio Daolio, and Saúl Vargas. Scalable hyperbolic recommender systems. *arXiv preprint arXiv:1902.08648*, 2019. 2

[4] Jingyuan Chen, Hanwang Zhang, Xiangnan He, Liqiang Nie, Wei Liu, and Tat-Seng Chua. Attentive collaborative filtering: Multimedia recommendation with item- and component-level attention. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 335–344, 2017. 2

[5] Yashar Deldjoo, Fatemeh Nazary, Arnau Ramisa, Julian McAuley, Giovanni Pellegrini, Alejandro Bellogin, and Tommaso Di Noia. A review of modern fashion recommender systems. *ACM Computing Surveys*, 2023. 2

[6] Pengfei Fang, Mehrtash Harandi, Trung Le, and Dinh Phung. Hyperbolic geometry in computer vision: A survey. *arXiv preprint arXiv:2304.10764*, 2023. 1

[7] Octavian Ganea, Gary Becigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, volume 31, 2018. 1

[8] Songwei Ge, Shlok Mishra, Simon Kornblith, Chun-Liang Li, and David Jacobs. Hyperbolic contrastive learning for visual representations beyond objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6840–6849, 2023. 1

[9] Yunhui Guo, Xudong Wang, Yubei Chen, and Stella X. Yu. Clipped hyperbolic classifiers are super-hyperbolic classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1–10, 2022. 1

[10] Ruining He and Julian McAuley. Vbpr: Visual bayesian personalized ranking from implicit feedback. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 144–150, 2016. 2

[11] Wang-Cheng Kang, Chen Fang, Zhaowen Wang, and Julian McAuley. Visually-aware fashion recommendation and design with generative image models. In *Proceedings of the IEEE International Conference on Data Mining*, pages 207–216, 2017. 2

[12] Qiang Liu, Shu Wu, and Liang Wang. Deepstyle: Learning user preferences for visual recommendation. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 841–844, 2017. 2

[13] Rongkai Ma, Pengfei Fang, Tom Drummond, and Mehrtash Harandi. Adaptive poincaré point to set distance for few-shot classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1926–1934, 2022. 1

[14] Denis Parra, Antonio Ossa-Guerra, Manuel Cartagena, Patricio Cerda-Mardini, and Felipe del Rio. Visrec: A hands-on tutorial on deep learning for visual recommender systems. In *Proceedings of the International Conference on Intelligent User Interfaces - Companion*, pages 5–6, 2021. 2

[15] Wei Peng, Tuomas Varanka, Abdelrahman Mostafa, Henglin Shi, and Guoying Zhao. Hyperbolic deep neural networks: A

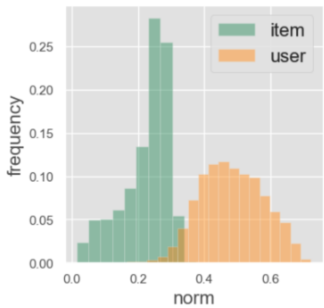
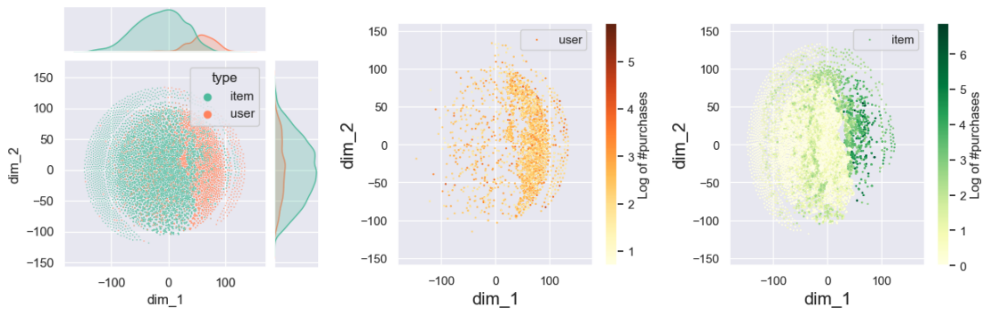


Figure 3. The embedding-norm distribution on Amazon Men.



(a) User and item representations (b) User representations (c) Item representations

Figure 4. User and item 2D-representations compressed by tSNE [20] on Amazon Men.

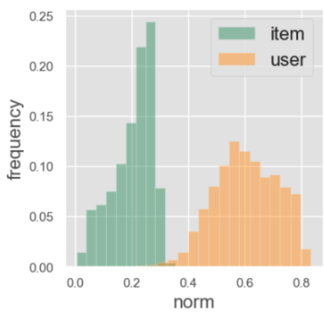
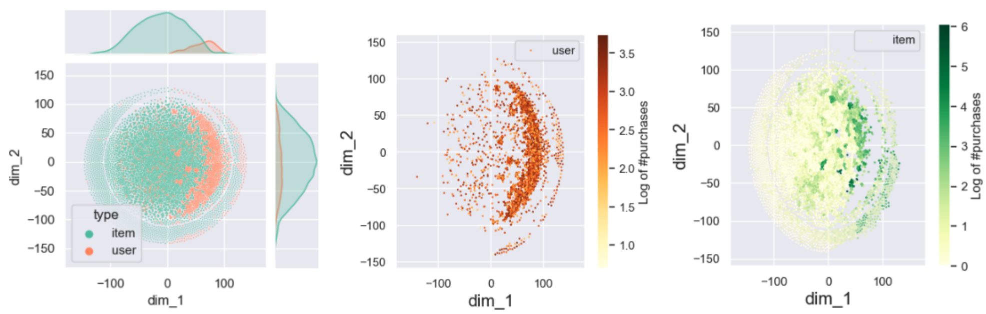


Figure 5. The embedding-norm distribution on TOWN Women.



(a) User and item representations (b) User representations (c) Item representations

Figure 6. User and item 2D-representations compressed by tSNE [20] on TOWN Women.

survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10023–10044, 2022. 1, 2

[16] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 452–461, 2009. 1

[17] Ryotaro Shimizu, Takuma Nakamura, and Masayuki Goto. Fashion-specific ambiguous expression interpretation with partial visual-semantic embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 3497–3502, 2023. 3

[18] Ryotaro Shimizu, Yuki Saito, Megumi Matsutani, and Masayuki Goto. Fashion intelligence system: An outfit interpretation utilizing images and rich abstract tags. *Expert Systems with Applications*, page 119167, 2022. 3

[19] Abraham Ungar. Hyperbolic trigonometry and its application in the poincaré ball model of hyperbolic geometry. *Computers & Mathematics with Applications*, 41:135–147, 2000. 1

[20] Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008. 4

[21] Tran Dang Quang Vinh, Yi Tay, Shuai Zhang, Gao Cong, and Xiao-Li Li. Hyperbolic recommender systems. *arXiv preprint arXiv:1809.01703*, 2018. 2

[22] Lucas Vinh Tran, Yi Tay, Shuai Zhang, Gao Cong, and Xiaoli Li. Hyperml: A boosting metric learning approach in hyperbolic space for recommender systems. In *Proceedings of the International Conference on Web Search and Data Mining*, pages 609–617, 2020. 2

[23] Canran Xu and Ming Wu. Learning feature interactions with lorentzian factorization machine. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 6470–6477, 2020. 2