CO-SNE: Dimensionality Reduction & Visualization for Hyperbolic Data

Yunhui Guo
Haoran Guo
Stella X. Yu
Poincaré Ball of Hyperbolic Space

- **Points** in hyperbolic space:

\[ \mathbb{B}^n = \{ \mathbf{x} \in \mathbb{R}^n : \| \mathbf{x} \| < 1 \} \]

- **Lines** in hyperbolic space:

- **Hyperbolic distance**:

\[
d_{\mathbb{B}^n}(\mathbf{u}, \mathbf{v}) = \text{arcosh}\left(1 + 2 \frac{\| \mathbf{u} - \mathbf{v} \|^2}{(1 - \| \mathbf{u} \|^2)(1 - \| \mathbf{v} \|^2)}\right)
\]
Visualizing 2D Hyperbolic Space is Easy

- Embedding WordNet hierarchy in hyperbolic space
Interpreting High-dimensional Hyperbolic Data
Interpreting High-dimensional Hyperbolic Data

- Mammals
  - Rodent
  - Carnivore
- Squirrel
- Muskrat
- Feline
- Canine

High-Dimensional Poincaré Embeddings

`t-SNE: No global hierarchy`
Interpreting High-dimensional Hyperbolic Data
Interpreting High-dimensional Hyperbolic Data
t-SNE for Visualizing High-Dimensional Euclidean Data

- Given high-dimensional data $\mathbf{x}$, seek low-dimensional data $\mathbf{y}$ such that their distance-based joint probabilities $P$ and $Q$ match:

$$
C = KL(P \| Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}
$$

- A normal distribution is used for $P$:

$$
p_{j|i} = \frac{\exp(-d(\mathbf{x}_i, \mathbf{x}_j)^2/2\sigma_i^2)}{\sum_{k \neq i} \exp(-d(\mathbf{x}_i, \mathbf{x}_k)^2/2\sigma_i^2)}
$$

- A heavier tailed $t$-distribution is used for $Q$ to make up for smaller volumes in low-dimensional space at the same distance:

$$
q_{ij} = \frac{(1 + d(\mathbf{y}_i, \mathbf{y}_j)^2)^{-1}}{\sum_{k \neq i} (1 + d(\mathbf{y}_k, \mathbf{y}_l)^2)^{-1}}
$$
Extending t-SNE to Hyperbolic Space

Hyperbolic Normal Distribution

\[ \mathcal{N}_B^n(x|\mu, \sigma^2) = \frac{1}{Z} \exp\left(-\frac{d_B^2(\mu, x)}{2\sigma^2}\right) \]

Hyperbolic Student’s t-distribution

\[ f_B^n(t; t_0) = \frac{1}{\pi(1 + d_B^2(t, t_0)^2)} \]
Extension of t-SNE: The Wrong Way

<table>
<thead>
<tr>
<th></th>
<th>Metric</th>
<th>Low-dimensional Distribution</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-SNE</td>
<td>Euclidean</td>
<td>t-distribution</td>
<td>KL-divergence</td>
</tr>
<tr>
<td>HT-SNE</td>
<td>Hyperbolic</td>
<td>t-distribution</td>
<td>KL-divergence</td>
</tr>
</tbody>
</table>

Hyperbolic Student’s t-distribution is not heavy-tailed
Extension of t-SNE: The Wrong Way

<table>
<thead>
<tr>
<th></th>
<th>Metric</th>
<th>Low-dimensional Distribution</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-SNE</td>
<td>Euclidean</td>
<td>t-distribution</td>
<td>KL-divergence</td>
</tr>
<tr>
<td>HT-SNE</td>
<td>Hyperbolic</td>
<td>t-distribution</td>
<td>KL-divergence</td>
</tr>
</tbody>
</table>

No repulsion between dissimilar high-dimensional points
From t-SNE to CO-SNE

<table>
<thead>
<tr>
<th>Metric</th>
<th>Low-dimensional Distribution</th>
<th>Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-SNE</td>
<td>Euclidean</td>
<td>t-distribution</td>
</tr>
<tr>
<td>HT-SNE</td>
<td>Hyperbolic</td>
<td>t-distribution</td>
</tr>
<tr>
<td>CO-SNE</td>
<td>Hyperbolic</td>
<td>Cauchy</td>
</tr>
</tbody>
</table>

Strong repulsion between dissimilar high-dimensional points
Losses in CO-SNE

- **Similarity Loss**
  \[
  C = KL(P || Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}
  \]
  → Preserve local similarity!

- **Distance Loss**
  \[
  H = \frac{1}{m} \sum_{i=1}^{m} (\|x_i\|^2 - \|y_i\|^2)^2
  \]
  → Preserve global hierarchy!
Visualizing Synthetic Point Clouds

- Visualize data in a five-dimensional hyperbolic space
  - Sample data from a mixture of hyperbolic Gaussians with means as,
    - \([0, 0, 0, 0, 0]\) -- origin
    - \([0.1, 0, 0, 0, 0]\]
    - \([0, -0.2, 0, 0, 0]\]
    - \([0, 0, 0.9, 0, 0]\]
    - \([0, 0, 0,-0.9, 0]\]
Visualizing Synthetic Point Clouds
Why Not Existing Methods?

The standard \( t \)-SNE and UMAP:
- Underestimates the distance between points that are close to the boundary of the Poincaré ball.

PCA and HoroPCA:
- As a linear dimensionality method, PCA cannot reduce high-dimensional data to two dimensions in a meaningful way for visualization.

HT-SNE:
- No repulsion between dissimilar high-dimensional points.
Visualizing High-dimensional Poincaré Word Vectors

Visualizing Poincaré Variational Auto-Encoder Features

Visualizing Hyperbolic Neural Net Features

Impact of t-SNE Loss and Distance Loss

Objective function of CO-SNE: $\mathcal{L} = \lambda_1 \cdot t\text{-SNE Loss} + \lambda_2 \cdot \text{Distance Loss}$

Maintaining global hierarchy

<table>
<thead>
<tr>
<th>$\lambda_1$ = 0</th>
<th>$\lambda_2$ = 0.0</th>
<th>$\lambda_2$ = 0.01</th>
<th>$\lambda_2$ = 0.05</th>
<th>$\lambda_2$ = 0.1</th>
<th>$\lambda_2$ = 0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_2$ = 0.0</td>
<td><img src="image1.png" alt="Diagram" /></td>
<td><img src="image2.png" alt="Diagram" /></td>
<td><img src="image3.png" alt="Diagram" /></td>
<td><img src="image4.png" alt="Diagram" /></td>
<td><img src="image5.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>

Maintaining local similarity
Impact of t-SNE Loss and Distance Loss

Objective function of CO-SNE: $\mathcal{L} = \lambda_1 \cdot \text{t-SNE Loss} + \lambda_2 \cdot \text{Distance Loss}$

Maintaining global hierarchy

Maintaining local similarity
Impact of t-SNE Loss and Distance Loss

Objective function of CO-SNE: \[ \mathcal{L} = \lambda_1 \cdot \text{t-SNE Loss} + \lambda_2 \cdot \text{Distance Loss} \]

Maintaining global hierarchy

<table>
<thead>
<tr>
<th>(\lambda_1)</th>
<th>(\lambda_2 = 0.0)</th>
<th>(\lambda_2 = 0.01)</th>
<th>(\lambda_2 = 0.05)</th>
<th>(\lambda_2 = 0.1)</th>
<th>(\lambda_2 = 0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_1 = 0)</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
<td><img src="image5" alt="Diagram" /></td>
</tr>
<tr>
<td>(\lambda_1 = 5)</td>
<td><img src="image6" alt="Diagram" /></td>
<td><img src="image7" alt="Diagram" /></td>
<td><img src="image8" alt="Diagram" /></td>
<td><img src="image9" alt="Diagram" /></td>
<td><img src="image10" alt="Diagram" /></td>
</tr>
<tr>
<td>(\lambda_1 = 10)</td>
<td><img src="image11" alt="Diagram" /></td>
<td><img src="image12" alt="Diagram" /></td>
<td><img src="image13" alt="Diagram" /></td>
<td><img src="image14" alt="Diagram" /></td>
<td><img src="image15" alt="Diagram" /></td>
</tr>
</tbody>
</table>
CO-SNE: Dimensionality Reduction & Visualization for Hyperbolic Data
Yunhui Guo    Haoran Guo    Stella X. Yu

Contributions

Visualizing Two-Dimensional Hyperbolic Space is Easy

CO-SNE Uses Hyperbolic Cauchy Distribution

CO-SNE Produces Stronger Repulsion Force

Visualizing Hyperbolic Features

Hyperbolic Space

Contributions

Losses in CO-SNE

Embedding WordNet in a Two-dimensional Poincaré Ball

Visualizing Hyperbolic Features

Hyperbolic Distance

\[ d_{H}(u, v) = \arccosh \left( 1 + 2 \frac{||u - v||^2}{1 - ||u||^2} \right) \]

- Non-Euclidean space with constant negative curvature
- Can embed tree-like data continuously with low distortion

CO-SNE: A novel visualization method designed specifically for high-dimensional hyperbolic data

HoroPCA

CO-SNE

Total Loss: \[ \mathcal{L} = \lambda_1 \mathcal{C} + \lambda_2 \mathcal{H} \]

- t-SNE Loss:
  \[ \mathcal{C} = KL(P || Q) = \sum_{i} \sum_{j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \]
  Maintaining local similarities

- Distance Loss:
  \[ \mathcal{H} = \frac{1}{m} \sum_{i=1}^{m} (||x_i||^2 - ||y_i||^2)^2 \]
  Maintaining global hierarchy

The gradients as a function of high-dimensional and low-dimensional hyperbolic distance

CO-SNE uses a small \( \gamma \)

Hyperbolic Normal

\[ K_\gamma(x, \mu, \sigma^2) = \frac{1}{2} \exp \left( -\frac{d_H(x, \mu)^2}{2\sigma^2} \right) \]

Hyperbolic Cauchy

\[ f(t; t_0, \gamma) = \frac{1}{\tau \gamma} \exp \left( -\frac{d_H(t, t_0)^2}{\tau^2 + \gamma^2} \right) \]

CO-SNE produces much better visualization than HoroPCA