IS REPORT-1

Bipasha Garg (2022111006)

January 2025

1 Core Idea of the Paper

1. **The Problem:** High-dimensional data is difficult to transform into 2D or 3D visualizations. When we try to squeeze a complex, high-dimensional dataset into a lower dimension, data points tend to crowd together and lose the underlying structure(and information)-*"crowding problem."*

2. **SpaceMAP's Approach: Space Expansion:** SpaceMAP proposes that instead of just fitting data into a lower dimension, it aims to *match the capacity* of high- and low-dimensional spaces. It does this by transforming the distances between data points using a concept called "Equivalent Extended Distance" (EED). Essentially, the algorithm "expands" the low-dimensional space to better reflect the high-dimensional space, and in the process, reduces the effects of crowding.

3. **Hierarchical Modeling:** SpaceMAP doesn't treat all data points the same way. It takes into account that real-world data often has hierarchical structures (like sub-manifolds within larger ones). It models similarity between data points differently depending on whether they are in a near, middle, or far-field.

4. **Connection to Contrastive Learning:** The paper shows a connection between SpaceMAP (and UMAP, another similar method) and contrastive learning. Essentially, the attractive and repulsive forces at play in both SpaceMAP and UMAP act similarly to how contrastive learning methods function.

2 Key Concepts in the Paper

1. **Space Capacity:** The authors define capacity as a measure of volume in high-dimensional space (Equation 3) [volume of hyper-sphere to explain crowding].

2. **Equivalent Extended Distance (EED):** It's a mathematical transformation that allows distances in a low-dimensional space to effectively mir-

ror the volume/capacity of distances in a high-dimensional space (Equation 4). The EED is derived from the concept of matching the Hausdorff measure of volume in the high and low dimensional spaces.

3. **Intrinsic Dimensionality:** SpaceMAP estimates the *intrinsic dimension* of the data using maximum likelihood estimation (MLE) (Equations 5,6). This is important because the "real" complexity of the data might be much lower than its observed dimension.

4. **Hierarchical Manifold Modeling:** SpaceMAP considers data in near, middle, and far fields. Near field, middle field and far-field data relationships are treated separately, with specific distance transformation (Equation 8).

5. **Similarity Function:** SpaceMAP uses a Gaussian-like similarity function in high-dimensional space (similar to SNE/t-SNE) and exponential in the low dimensional space (Equation 9). **But what distinguishes spaceMap is that its high dimensional distance is transformed using EED first, before being modeled with Gaussian.**

6. **Loss Function:** The loss function is very similar to that of UMAP, using a contrastive learning framework (Equation 10).

3 Usage in BEADS

1. **Addressing Crowding:** SpaceMap does address crowding but right as my focus is on the spatial position this may come as a secondary problem. And as very high dimensional are very highly unlikely to fall in the same quadrant, crowding may not a be an issue as discussed previously.

2. **Hierarchical Structure in BEADS Data:** In case of hierarchical data, SpaceMap thinking could be considered.

4 Benefits of SpaceMap

1. **More Meaningful Visualizations:** Reduced crowding and better preservation of both local and global structures could lead to visualizations that are more accurate, insightful, and easier to interpret.

2. **Application of DR:** Can use the SpaceMAP algorithm as a dimensionality reduction method to transform high-dimensional data to lower dimensions if it is required at any point.

In summary, SpaceMAP provides a mathematical approach to visualize high-dimensional data, making it a valuable tool. Its ability to handle both local and global structures, combined with its link to contrastive learning, may help you create more effective visualizations, uncover hidden patterns, and better interpret data as it effectively reduces crowding.

5 Refrence

SpaceMAP: Visualizing High-dimensional Data by Space Expansion, Xinrui Zu, Qian Tao.