Rethinking the Distribution Law in t-SNE by using a Pearson type VII distribution

Bomans Edouard, Frénay Benoit[0000-0002-7859-2750], and Libert Anne-Sophie[0000-0003-2682-2793]

University of Namur, 5000 Namur, Belgium

Abstract. High-dimensional data visualization is a fundamental challenge in modern data analysis. Among the numerous methods developed, t-distributed Stochastic Neighbor Embedding (t-SNE) has become a standard tool for projecting complex data into a low-dimensional space while preserving local structure. Despite its widespread use, the assumption of a Student's t-distribution in the low-dimensional space has rarely been questioned. This work explores alternative distribution laws for t-SNE and introduces P7-SNE, which provides greater flexibility in controlling local and global interactions. We also analyze the underlying forces in these methods and illustrate how distribution choice affects attraction and repulsion in the projected space.

Keywords: Dimension Reduction · Embedding · t-SNE

1 Introduction

Visualizing high-dimensional data is crucial in many domains, from image recognition to genomics. t-SNE [7] is widely used to reduce dimensionality while preserving local neighborhood structure. However, the choice of the Student's t-distribution as the low-dimensional similarity function has rarely been revisited. Our work investigates whether other distributions can improve visualization quality both quantitatively and qualitatively. We also focus on understanding the forces that drive points in the projection, as these forces determine cluster formation, separation, and crowding effects.

2 t-SNE Overview

t-SNE matches two distributions: the high-dimensional similarities $P = \{p_{ij}\}$ and the low-dimensional similarities $Q = \{q_{ij}\}$. In standard t-SNE, those are given by

$$p_{ij} = \frac{\exp(-||x_i - x_j||^2)}{\sum_{k \neq l} \exp(-||x_k - x_l||^2)},$$

$$q_{ij} = \frac{(1+||y_i - y_j||^2)^{-1}}{\sum_{k\neq l}(1+||y_k - y_l||^2)^{-1}},$$

where the x_i and the y_i represents the point i in the high dimensional space and in the low dimensional space. The algorithm minimizes the Kullback-Leibler divergence

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

whose gradient is

$$\frac{\partial C}{\partial y_i} = 4\sum_j (p_{ij} - q_{ij}) \frac{y_i - y_j}{1+||y_i - y_j||^2}.$$

Intuitively, p_{ij} represents the probability that points i and j are neighbors in high-dimensional space, and q_{ij} is the similarity in low-dimensional space. The gradient can be interpreted in terms of forces: an *attractive force* pulling similar points together, and a *repulsive force* pushing dissimilar points apart. The Student's t-distribution is characterized by a heavy tail, which limits the crowding effect and allows distant points to exert a small repulsive influence.

3 Pearson type VII-based method

P7-SNE replaces the Student's t-distribution with alternative distribution, the Pearson VII distribution [4]. This distribution is a more flexible generalization of the Student's t law. Indeed it can be tuned to be exactly the t law but it can also tend to be a Gaussian or to have heavier tails than this modification in Van der Maaten's algorithm [6] changes the tail behavior of q_{ij} given by

$$q_{ij}^{(P7)} = \frac{\left(1+\left(\frac{||y_i - y_j|| - \lambda}{\alpha}\right)^2\right)^{-m}}{\sum_{k\neq l}\left(1+\left(\frac{||y_k - y_l|| - \lambda}{\alpha}\right)^2\right)^{-m}},$$

and the gradient's KL divergence is

$$\begin{aligned}\frac{\partial C}{\partial y_i}^{(P7)} =& 4\sum_j (p_{ij} - q_{ij})(y_i - y_j) \\ & \left(\frac{m(\|y_i - y_j\| - \lambda)\left(\left(\frac{\|y_i - y_j\| - \lambda}{\alpha}\right)^2 + 1\right)^{-1}}{\alpha^2}\right).\end{aligned}$$

In the rest of the article we will use the notation $Coeff_{ij}^{P7}$ for the last term of the gradient. If we select $\alpha = 1$, $m = 1$ and $\lambda = 0$, P7-SNE reduces to t-SNE. If we let m tend to $+\infty$ the Pearson VII becomes a Gaussian and P7-SNE reduces to classical SNE [2]. By tuning distribution parameters, P7-SNE can increase or decrease the range of interactions, affecting both attraction and repulsion. Stronger long-range repulsion can prevent crowding and produce more globally structured embeddings, while local attraction ensures that clusters remain tight.

4 Analysis of Forces

The behavior of points in the low-dimensional embedding can be understood through the lens of forces [1]. For classical t-SNE, the pairwise attractive and repulsive forces are

$$F_{ij}^{\text{attr}} = p_{ij} \frac{y_i - y_j}{1 + \|y_i - y_j\|^2} \quad \text{and} \quad F_{ij}^{\text{rep}} = q_{ij} \frac{y_i - y_j}{1 + \|y_i - y_j\|^2}.$$

The total force on point i is then

$$F_i = \sum_j F_{ij}^{\text{attr}} - \sum_j F_{ij}^{\text{rep}}.$$

Here, attraction is strong between nearby points (due to p_{ij} being large for neighbors) and diminishes rapidly with distance. Repulsion acts globally but weakly, as q_{ij} decreases slowly with distance due to the long tails of the Student's t-distribution. This reduces the *crowding effect* in low dimensions.

For P7-SNE, the forces are generalized to account for the Pearson VII distribution and become

$$\begin{aligned} F_{ij}^{attr} &= p_{ij} Coeff_{ij}^{P7}(y_j - y_i), \\ F_{ij}^{rep} &= q_{ij} Coeff_{ij}^{P7}(y_i - y_j). \end{aligned}$$

Figure 1 illustrate how modifying the distribution affects the range and strength of forces. P7-SNE can produce embeddings where clusters remain compact locally while exhibiting better global separation, thanks to tunable long-range repulsion.

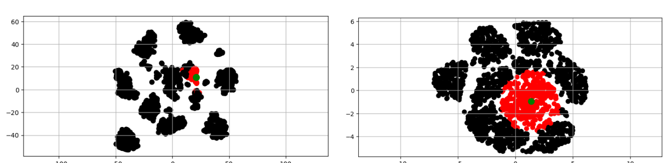

Fig. 1. Visualization of the influence of a single point (green point) on the visualization (red if influenced and black if not) in t-SNE (top) and P7-SNE (bottom).

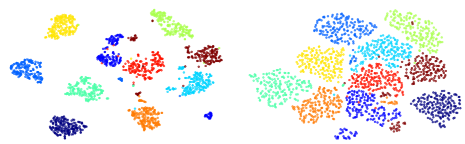

Fig. 2. Visualization of the *Digits* dataset with t-SNE (top) and P7-SNE (bottom).

5 Experimental Results

We evaluated classical t-SNE and P7-SNE on several datasets, but we will only present those on *Digits* with metrics AUC-Log-RNX [3](neighborhood preservation), Distance Consistency (DSC) [5](cluster separation) and Distribution Consistency (DC) [5](crowding evaluation). Visual inspection complements quantitative measures. Notice how clusters are well formed in both visualizations as we can see in Fig. 2, but less cluttered and easier to read with P7-SNE.

Metrics	t-SNE	P7-SNE
AUC-Log-RNX	0.5281	0.6739
DSC	0.9588	0.9271
DC	0.9985	0.9969

Table 1. Scores of the three metrics for both t-SNe and P7-SNE on the *Digits* dataset.

Results in Table 1 indicate that the P7-SNE method as a better AUC-Log-RNX on the *Digits* dataset while the other metrics are alike for both methods. Visually we can see that the visualization produced by the P7-SNE algorithm is easier to read than the one produced by the t-SNE's.

6 Discussion and Conclusion

Adjusting the tail of the distribution in SNE variants modifies the balance between attraction and repulsion, leading to different embedding structures. Our work demonstrates that rethinking the underlying distribution law in t-SNE can improve visualization quality. Further research is needed to establish precise connections between force dynamics, distribution parameters, and performance metrics.

References

1. Delchevalerie V., Mayer A., Bibal A., and Frenay B., *Accelerating t-SNE using Fast Fourier Transforms and the Particle-Mesh Algorithm from Physics*, Proceedings of the International Joint Conference on Neural Networks, vol. 2021-July, 2021.

2. Hinton G., and Roweis S., *Stochastic neighbor embedding*, Advances in Neural Information Processing Systems, 2003.
3. Lee J. A., Peluffo-Ordóñez D. H., and Verleysen M., *Multi-scale similarities in stochastic neighbour embedding: Reducing dimensionality while preserving both local and global structure*, Neurocomputing, vol. 169, pp. 246–261, 2015.
4. Nadarajah S., *Pearson type VII ratio distribution*, Empirical Economics, vol. 37, no. 1, pp. 219–229, 2009.
5. Sips M., Neubert B., Lewis J. P., and Hanrahan P., *Selecting good views of high-dimensional data using class consistency*, Computer Graphics Forum, vol. 28, no. 3, pp. 831–838, 2009.
6. Van Der Maaten L., *t_SNE*, 2008. Available : https://lvdmaaten.github.io/tsne/.
7. Van Der Maaten L., and Hinton G., *Visualizing data using t-SNE*, Journal of Machine Learning Research, vol. 9, pp. 2579–2625, 2008.