
6th International Conference on Nanotechnologies and Biomedical Engineering
Proceedings of ICNBME-2023, September 20–23, 2023, Chisinau, Moldova
Volume 2: Biomedical Engineering and New Technologies for Diagnosis, Treatment, and Rehabilitation

# Unsupervised Knowledge Extraction from Biomedical Data

**Basarab Matei, Petru Alexandru Vlaicu, Nicoleta Rogovschi, Nistor Grozavu**

https://doi.org/10.1007/978-3-031-42782-4_27

## Abstract

In this paper we introduce a study on the use of the unsupervised representation learning on biomedical data i.e. on Growth weight data and Wisconsin Diagnostic Breast Cancer obtaining good performances in terms of clustering In this study, we propose an adaptation of the unsupervised topological learning to deals with biomedical datasets based on a new approximation strategy to visualize high dimensional datasets. In data containing high-dimensional data manifold, the level of the discrepancy changes depending on the dimension of intrinsic data manifold. Then the strength of the repelling power is dependent of dataset. The proposed approach is based on t-SNE (Stochastic Neighbor Embedding) dimensionality reduction method with a different inhomogenous approximation strategy of the t-Distribution. In order to avoid the exponential computation we propose an inhomogenous approximation of the t-Distribution having the precision order of $10^{-3}$. By using this inhomogenous approximation we allow to optimize approximately the t-Distribution with respect to the number of degree of freedom and also to reduce the computational time. We illustrate the power of the proposed approach with two bio-medical real datasets and the obtained results outperform classical SNE and t-SNE methods.

*Keywords: data visualization, dimensional reduction, clustering*


**6th International Conference on Nanotechnologies and Biomedical Engineering
Proceedings of ICNBME-2023, September 20–23, 2023, Chisinau, Moldova
Volume 2: Biomedical Engineering and New Technologies for Diagnosis, Treatment, and Rehabilitation**

## References

1. Andreopoulos, B., An, A., Wang, X.: Bi-level clustering of mixed categorical and numerical biomedical data. Int. J. Data Min. Bioinform. **1**(1), 19–56 (2006)
2. Asuncion, A., Newman, D.J.: UCI Machine Learning Repository (2007)
3. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 1373–1396 (2003)
4. Cieslak, M.C., Castelfranco, A.M., Roncalli, V., Lenz, P.H., Hartline, D.K.: t-Distributed stochastic neighbor embedding (t-SNE): a tool for eco-physiological transcriptomic analysis. Mar. Genomics **51**, 100–123 (2020)
5. Hinton, G., Roweis, S.: Stochastic neighbor embedding. Adv. Neural. Inf. Process. Syst. **15**, 833–840 (2003)
6. Hubert, L., Arabie, P.: Comparing partitions. J. Classif. **2**(1), 193–218 (1985)
7. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Inc., Upper Saddle River (1988)
8. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. ACM Comput. Surv. **31**(3), 264–323 (1999)
9. Johnson, N.L., Kotz, S., Balakrishnan, N.: Distributions in Statistics: Continuous Univariate Distributions, 2nd edn. Wiley, New York (1999)
10. Khan, S.S., Kant, S.: Computation of initial modes for K-modes clustering algorithm using evidence accumulation. In: IJCAI, pp. 2784–2789 (2007)
11. Kitazono, J., Grozavu, N., Rogovschi, N., Omori, T., Ozawa, S.: t-Distributed stochastic neighbor embedding with inhomogeneous degrees of freedom. In: Hirose, A., Ozawa, S., Doya, K., Ikeda, K., Lee, M., Liu, D. (eds.) ICONIP 2016. LNCS, vol. 9949, pp. 119–128. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46675-0 14
12. Lafon, S., Lee, A.B.: Diffusion maps and coarse-graining: a unified framework for dimensionality reduction, graph partitioning, and data set parameterization. IEEE Trans. Pattern Anal. Mach. Intell. **28**(9), 1393–1403 (2006)
13. Li, B., de Moor, B.: A corrected normal approximation for student's t distribution. Comput. Stat. Data Anal. **29**, 213–216 (1999)




14. Platzer, A.: Visualization of SNPs with t-SNE. PLOS ONE **8**(2), 1–6 (2013)

15. Rand, W.M.: Objective criteria for the evaluation of clustering methods. J. Am. Stat. Assoc. **66**, 846–850 (1971)

16. Rogovschi, N., Kitazono, J., Grozavu, N., Omori, T., Ozawa, S.: t-Distributed stochastic neighbor embedding spectral clustering. In: 2017 International Joint Conference on Neural Networks, IJCNN 2017, Anchorage, AK, USA, 14–19 May 2017, pp. 1628–1632. IEEE (2017)

17. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)

18. Tenenbaum, J.B., De Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)

19. Van der Maaten, L.J.P.: Learning a parametric embedding by preserving local structure. In: International Conference on Artificial Intelligence and Statistics, JMLR, W&CP, vol. 5 (2008)

20. Van der Maaten, L.J.P., Hinton, G.E.: Visualizing high-dimensional data using t-SNE. J. Mach. Learn. **9**, 2579–2605 (2008)

21. Vladymyrov, M., Carreira-Perpinan, M.: Entropic affinities: properties and efficient numerical computation. In: Proceedings of the 30th International Conference on Machine Learning, pp. 477–485 (2013)

22. Vlaicu, P.A., Untea, A.E., Panaite, T.D., Turcu, R.P.: Effect of dietary orange and grapefruit peel on growth performance, health status, meat quality and intestinal microflora of broiler chickens. Ital. J. Anim. Sci. **19**(1), 1394–1405 (2020)

23. Zhou, B., Jin, W.: Visualization of single cell RNA-Seq data using t-SNE in R. In: Kidder, B.L. (ed.) Stem Cell Transcriptional Networks. MMB, vol. 2117, pp. 159–167. Springer, New York (2020). https://doi.org/10.1007/978-1-0716-0301-7 8

24. Zhou, H., Wang, F., Tao, P.: t-Distributed stochastic neighbor embedding method with the least information loss for macromolecular simulations. J. Chem. Theory Comput. **14**(11), 5499–5510 (2018)

25. Zhu, W.,Webb, Z.T., Mao, K., Romagnoli, J.: A deep learning approach for process data visualization using t-distributed stochastic neighbor embedding. Ind. Eng. Chem. Res. **58**(22), 9564–9575 (2019)