


## Editorial

### **Integrating Information Visualization and Dimensionality Reduction: A pathway to Bridge the Gap between Natural and Artificial Intelligence**

 Diego H. Peluffo-Ordóñez<sup>1,2</sup>

<sup>1</sup>Mohamed VI Polytechnic University, Ben Guerir-Morocco,  
diego.peluffo@um6p.ma

<sup>2</sup>Corporación Universitaria Autónoma de Nariño, Pasto – Colombia  
diego.peluffo@aunar.edu.co

By importing some natural abilities from human thinking into the design of computerized decision support systems, a cross-cutting trend of intelligent systems has emerged, namely, the synergetic integration between natural and artificial intelligence [1]. While natural intelligence provides creative, parallel, and holistic thinking, its artificial counterpart is logical, accurate, able to perform complex and extensive calculations, and tireless. In the light of such integration, two concepts are important: controllability and interpretability. The former is defined as the ability of computerized systems to receive feedback and follow users' instructions, while the latter refers to human-machine communication. A suitable alternative to simultaneously involve these two concepts—and then bridging the gap between natural and artificial intelligence—is bringing together the fields of dimensionality reduction (DimRed) and information visualization (InfoVis).

### **Dimensionality reduction (DimRed)**

DimRed is a key tool for artificial intelligence tasks—more specifically machine learning—that involve high dimensional data sets. The aim of DimRed approaches is to extract lower dimensional, relevant information (called embedded data) from high-dimensional input data so that the performance of a pattern recognition system is improved and/or the data visualization becomes more intelligible. Principal component analysis (PCA) and classical multidimensional scaling (CMDS) are two classical DimRed approaches based on variance and distance preservation criteria, respectively [2]. The modern focus of DimRed approaches relies on more developed criteria, which are aimed at preserving the data topology. In particular, data topology is involved in the formulation of the problem through pairwise similarities between data points. Therefore, these approaches can be readily understood from a graph-theory point of view because data are represented in a non-directed and weighted graph where data points denote the nodes, and a non-negative similarity (also affinity) matrix holds the pairwise edge weights. Two pioneer methods incorporate similarities, i.e., Laplacian eigenmaps (LE) [3] and locally linear embedding (LLE) [4], which are spectral approaches.

Likewise, since the rows of the normalized similarity matrix can be interpreted as probability distributions, methods based on divergences have emerged. Due to its probabilistic connotation, the most representative among such methods is called stochastic neighbor embedding (SNE) [5]. SNE and its variants have shown to be suitable for obtaining high-quality embedded data since they preserve similarities in both low- and high-

dimensional spaces during the optimization process. Indeed, according to some studies [6],[7],[8], SNE-like approaches are the most effective in terms of agreement rate between neighbors [9]. Naturally, some alternatives to SNE and improvements have been proposed.

For instance, in [10], a mixture of divergences is proposed. More recent approaches [11] have focused on free-parameter alternatives derived from a multi-scale SNE.

## **Information visualization (InfoVis)**

Recent analyses have indicated that DimRed should reach two goals: (1) ensure that data points that are neighbors in the original space remain neighbors in the embedded space and (2) guarantee that two data points are shown as neighbors in the embedded space only if they are neighbors in the original space. In the context of information retrieval, these two goals can be seen as precision and recall measures, respectively. Although clearly conflictive, the compromise between precision and recall defines the performance of the DimRed method.

Furthermore, since DimRed methods are often developed under predetermined design parameters and pre-established optimization criterion, they still lack properties such as user interaction and controllability. Such properties are characteristic of information visualization (InfoVis) procedures. The field of InfoVis aims to develop graphical ways to represent data so that information can be more usable and intelligible for users [12]. As a result, based on the premise that DimRed can be improved by importing some properties of InfoVis methods, a research area that integrates the two has emerged.

## **Integration between InfoVis and DimRed**

The main goal of this research area is to link the field of DimRed with that of InfoVis to harness the special properties of the latter within DimRed frameworks. Therefore, controllability and interactivity properties are of great interest because they may make the DimRed outcomes significantly more understandable and tractable for (not necessarily expert) users. Particularly, these two properties provide users with leeway to explore and select the best ways to represent data. In other words, the goal of the aforementioned integration is to develop a DimRed framework that facilitates an interactive and quick visualization of data representations that make DimRed outcomes more intelligible and allow users to modify data views according to their needs in an affordable fashion [13]. Such goal is of special interest for current artificial intelligence communities, but also greatly challenging.

Future studies in the field of InfoVis-DimRed integration should address the following open issues:

- Designing a unified or generalized framework for DimRed methods (UFDR) so that the embedding approach can be readily and quickly selected according to user needs.
- Determining a clear and definite relationship between performance and design parameters (weights determining the compromise between recall and precision, and regularization or smoothness parameters for outlier detection) according to UFDR settings.
- Based on UFDR, designing new or more general divergence-based methods where the design parameters are independent and their roles are clearly identified.
- Designing faster and more stable implementations for DimRed methods in such a way that sensitivity to starting parameters is avoided.
- Developing DimRed approaches that can be successfully incorporated into an interface in which users can interact with the parameters in an efficient and interactive way.

## REFERENCES

- [1] J. C. Alvarado-Pérez; D. H. Peluffo-Ordóñez; R. Therón, “Bridging the gap between human knowledge and machine learning,” *ADCAIJ Adv. Distrib. Comput. Artif. Intell. J.*, vol. 4, no. 1, pp. 54-64, Oct. 2015. <https://doi.org/10.14201/ADCAIJ2015415464>
- [2] I. Borg; P. Groenen, *Modern multidimensional scaling: Theory and applications*. Springer, 2005.
- [3] M. Belkin; P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, Jun. 2003. <https://doi.org/10.1162/089976603321780317>
- [4] S. T. Roweis; L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000. <https://doi.org/10.1126/science.290.5500.2323>
- [5] G. E. Hinton; S. T. Roweis, “Stochastic neighbor embedding,” in *Advances in neural information processing systems*, pp. 833–840, 2002. [URL](#)
- [6] J. A. Lee; M. Verleysen, *Nonlinear dimensionality reduction*. Springer, 2007.
- [7] D. H. Peluffo; J. A. Lee; M. Verleysen, “Recent methods for dimensionality reduction: A brief comparative analysis,” *2014 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2014)*, Bruges, 2014. [URL](#)
- [8] D. H. Peluffo-Ordóñez; J. A. Lee; M. Verleysen, “Short Review of Dimensionality Reduction Methods Based on Stochastic Neighbour Embedding,” in *Advances in Intelligent Systems and Computing*, 2014, pp. 65–74. [https://doi.org/10.1007/978-3-319-07695-9\\_6](https://doi.org/10.1007/978-3-319-07695-9_6)
- [9] J. A. Lee; M. Verleysen, “Quality assessment of dimensionality reduction: Rank-based criteria,” *Neurocomputing*, vol. 72, no. 7, pp. 1431–1443, Mar. 2009. <https://doi.org/10.1016/j.neucom.2008.12.017>
- [10] L. der Maaten; G. Hinton, “Visualizing data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579-2605, Nov. 2008. [URL](#)
- [11] J. A. Lee; D. H. Peluffo-Ordóñez; M. Verleysen, “Multi-scale similarities in stochastic neighbour embedding: Reducing dimensionality while preserving both local and global structure,” *Neurocomputing*, vol. 169, pp. 246–261, Dec. 2015. <https://doi.org/10.1016/j.neucom.2014.12.095>
- [12] M. Zastrow, “Data visualization: Science on the map,” *Nature*, vol. 519, no. 7541, pp. 119–120, Mar. 2015. <https://doi.org/10.1038/519119a>
- [13] M. C. Ortega-Bustamante *et al.*, “Introducing the Concept of Interaction Model for Interactive Dimensionality Reduction and Data Visualization,” *International Conference on Computational Science and Its Applications*. Springer, 2020. [https://doi.org/10.1007/978-3-030-58802-1\\_14](https://doi.org/10.1007/978-3-030-58802-1_14)