Introducing the Concept of Interaction Model for Interactive Dimensionality Reduction and Data Visualization

M. C. Ortega-Bustamante[1,2], W. Hasperué[2], D. H. Peluffo-Ordóñez[4,5,6(✉)], M. Paéz-Jaime[5], I. Marrufo-Rodríguez[5], P. Rosero-Montalvo[1,3], A. C. Umaquinga-Criollo[1], and M. Vélez-Falconi[4,5]

[1] Universidad Técnica del Norte, Ibarra, Ecuador
[2] III-LIDI, Facultad de Informática, Universidad Nacional de La Plata, La Plata, Argentina
cosme.ortegab@info.unlp.edu.ar
[3] Instituto Superior Tecnológico 17 de Julio, San Miguel de Urcuquí, Ecuador
[4] Yachay Tech University, San Miguel de Urcuquí, Ecuador
[5] SDAS Research Group, Ibarra, Ecuador
diego.peluffo@sdas-group.com
https://www.sdas-group.com
[6] Corporación Univeristaria Autónoma de Nariño, Pasto, Colombia

Abstract. This letter formally introduces the concept of interaction model (IM), which has been used either directly or tangentially in previous works but never defined. Broadly speaking, an IM consists of the use of a mixture of dimensionality reduction (DR) techniques within an interactive data visualization framework. The rationale of creating an IM is the need for simultaneously harnessing the benefit of several DR approaches to reach a data representation being intelligible and/or fitted to any user's criterion. As a remarkable advantage, an IM naturally provides a generalized framework for designing both interactive DR approaches as well as readily-to-use data visualization interfaces. In addition to a comprehensive overview on basics of data representation and dimensionality reduction, the main contribution of this manuscript is the elegant definition of the concept of IM in mathematical terms.

Keywords: Dimensionality reduction · Interaction model · Kernel functions · Data visualization

1 Introduction

Very often, dimensionality reduction (DR) is an essential building block to design both machine learning systems, and information visualization interfaces [1,2]. In simple terms, DR consists of finding a low-dimensional representation of the

This work is supported by Yachay Tech and SDAS Research Group (http://www.sdas-group.com).

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12250, pp. 193–203, 2020.
https://doi.org/10.1007/978-3-030-58802-1_14

original data (said to be high-dimensional) by keeping a criterion of either data structure preservation, or class-separability ensuring. Recent analysis has shown that DR should attempt to reach two goals: First, to ensure that data points that are neighbors in the original space should remain neighbors in the embedded space. Second, to guarantee that two data points should be 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. In spite of being clearly conflicting, the compromise between precision and recall denes the DR method performance. Furthermore, since DR methods are often developed under determined design parameters and pre-established optimization criterion, they still lack of properties such as user interaction and controllability. These properties are characteristic of information visualization procedures. The eld of data visualization (DataVis) is aimed at developing graphical ways of representing data so that information can be more usable and intelligible for the user [3]. Then, one can intuit that DR can be improved by importing some properties of the DataVis methods. This is in fact the premise on which this research is based.

This emergent research area can be referred as interactive dimensionality reduction for visualization. Its main goal is to link the field of DR with that of DataVis, in order to harness the special properties of the latter within DR frameworks. In particular, the properties of controllability and interactivity are of great interest, which should make the DR outcomes significantly more understandable and tractable for the (no-necessarily-expert) user. These two properties allow the user to have freedom to select the best way for representing data. Then, in other words, it can be said that the goal of this project is to develop a DR framework that facilitates an interactive and quick visualization of data representation to make more intelligible the DR outcomes, as well as to allow users modifying the views of data according to their needs in an affordable fashion.

In this connection, this letter formally introduces the concept of interaction model (IM) as a key tool for both interactive DR and DataVis. Even though the term interaction model has been referred directly or tangentially in previous works [4–8], it has not been formally defined. This paper aims to fill that void. In general terms, the concept of IM refers to a mixture of dimensionality reduction (DR) techniques within an interactive data visualization framework. The rationale of creating an IM is the need for simultaneously harnessing the benefit of several DR approaches to reach a data representation being intelligible and/or fitted to any user's criterion. As a remarkable advantage, an IM naturally provides a generalized framework for designing both interactive DR approaches as well as readily-to-use data visualization interfaces. That said, the main contribution of this manuscript is the elegant definition of the concept of IM in mathematical terms. Also, it overviews some basics of data representation and dimensionality reduction from matrix algebra point of view. A special interest is given to spectral and kernel-based DR methods, which can be generalized by kernel principal component analysis (KPCA) and readily incorporated into a linear IM.

The remaining of this manuscript is organized as follows: Sect. 2 states the mathematical notation for main variables and operators. Section 3 presents a short overview on basic concepts and introductory formulations for DR, with a special interest in KPCA. In Sect. 4, we formally define the concept of IM as well as its particular linear version. Also, the use of DR-based DataVis is outlined. Finally, some concluding remarks are gathered in Sect. 5.

2 Mathematical Notation

For introducing further concepts and describing some mathematical developments, let us consider the following notation: Let $\boldsymbol{X} \in \mathbb{R}^{N \times D}$ be the input data matrix (also called observed space) holding N data points or samples described by a D-dimensional feature set, such that:

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1l} & \cdots & x_{1D} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & & x_{il} & & x_{iD} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nl} & \cdots & x_{ND} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1^\top \\ \vdots \\ \boldsymbol{x}_i^\top \\ \vdots \\ \boldsymbol{x}_N^\top \end{pmatrix}, \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^D$ denotes the i-th data point, $i \in \{1, \ldots, N\}$ and $l \in \{1, \ldots, D\}$. Likewise, let us consider a lower-dimensional matrix, so-named embedded space, as

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1\ell} & \cdots & y_{1d} \\ \vdots & \ddots & \vdots & & \vdots \\ y_{i1} & & y_{i\ell} & & y_{id} \\ \vdots & & \vdots & \ddots & \vdots \\ y_{N1} & \cdots & y_{N\ell} & \cdots & y_{Nd} \end{pmatrix} = \begin{pmatrix} \boldsymbol{y}_1^\top \\ \vdots \\ \boldsymbol{y}_i^\top \\ \vdots \\ \boldsymbol{y}_N^\top \end{pmatrix}, \tag{2}$$

with $\boldsymbol{y}_i \in \mathbb{R}^d$, such that $d < D$ and $\ell \in \{1, \ldots, d\}$. Also, let us suppose that there exists an unknown high dimensional representation space $\boldsymbol{\Phi} \in \mathbb{R}^{D_h \times N}$ such that $D_h \ggg D$, in which calculating the inner product should improve the representation and visualization of resultant embedded data in contrast to that obtained directly from the observed data. Hence, the need for a kernel representation arises to calculate the dot product in the unknown high dimensional space. Let $\phi(\cdot)$ be a function that maps data from the original dimension to a higher one, such that:

$$\begin{aligned} \phi(\cdot): \quad & \mathbb{R}^D \to \mathbb{R}^{D_h} \\ & \boldsymbol{x}_i \mapsto \phi(\boldsymbol{x}_i). \end{aligned} \tag{3}$$

Therefore, the matrix $\boldsymbol{\Phi}$ can be expressed as:

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_1^\top \\ \vdots \\ \boldsymbol{\Phi}_N^\top \end{pmatrix} \tag{4}$$

being $\boldsymbol{\Phi}_i = \phi(\boldsymbol{x}_i)$ its i-th column vector.

3 Overview on Dimensionality Reduction and Data Visualization

3.1 Data Representation

Data representation is a wide-meaning term coined by some authors to refer by-and-large to either data transformation or feature extraction. The former consists of transforming data into a new version intended to fulfill a specific goal [9]. The latter, meanwhile, is somewhat as a data remaking in such a manner that input data undergo a morphological deformation, or projection (also called rotation) by following a certain transformation criterion [10]. Also, data representation may be referred to yielding a new data representation -just as a new data matrix being an alternative to the original given one. For instance, a dissimilarity-based representation of input data [11]. An exhaustive review on data representation is presented in [12].

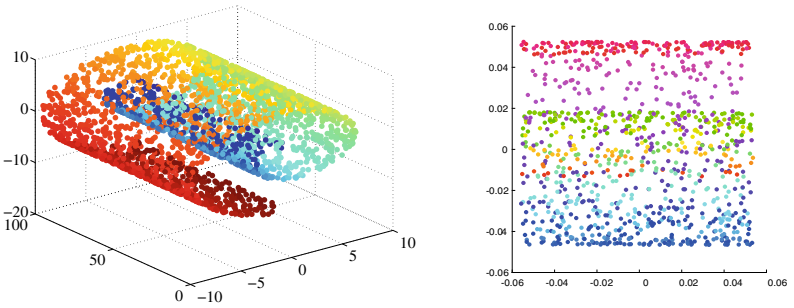
3.2 Dimensionality Reduction

Dimensionality reduction (DR) [13] is the most widely-used data representation approach. Broadly, DR is intended to embed a high-dimensional space $\boldsymbol{X}$ into a lower-dimensional space $\boldsymbol{X}$. More technically, DR can be seen as a data transformation through a certain operator $\mathcal{T}\{\cdot\}$ in the form:

$$\begin{aligned} \mathcal{T}(\cdot): \quad & \mathbb{R}^{N\times D} \to \mathbb{R}^{N\times d} \\ & \boldsymbol{X} \quad \mapsto \mathcal{T}\{\boldsymbol{X}\}, \end{aligned} \tag{5}$$

such that $\boldsymbol{Y} = \mathcal{T}\{\boldsymbol{X}\}$. Such a transformation operator can be ruled by different criteria. Figure 1 depicts the effect of a toy dataset, so-called `Swiss Roll` ($N = 3000$ data points, and $D = 3$ dimensions).

DR Approaches: For instance, pioneer approaches such as principal component analysis (PCA) or classical multidimensional scaling (CMDS) optimize the reduction in terms of variance and distance preservation criterion, respectively [14]. More sophisticated methods attempt to capture the data topology through a non-directed and weighted data-driven graph, which is formed by nodes located at the geometrical coordinates pointed out by the data points represent the nodes, and a non-negative similarity (also called affinity, or Gram) matrix holding the pairwise edge weights. Such data-topology-based criteria has been addressed by both spectral [15] and divergence-based methods [16]. Similarity matrix can represent either the weighting factor for pairwise distances as happens in Laplacian eigenmaps and locally linear embedding [17,18], or a probability distribution as is the case of methods based on divergences such as stochastic neighbour embedding [19]. In this letter, we give especial interest to the spectral approaches -more specifically, to the so-named kernel PCA.



(a) `Swiss Roll` dataset. (b) Lower-dimensional (2D) representation.

Fig. 1. Effect of unfolding a 3D `Swiss Roll`–like toy dataset into a two-dimensional representation through a DR approach based on kernel principal component analysis (KPCA).

Kernel PCA (KPCA): Now, let us recall the high-dimensional space $\boldsymbol{\Phi}$ defined in Sect. 2, let us also consider a lower-rank reconstructed matrix $\widehat{\boldsymbol{\Phi}}$, and a dissimilarity function $\delta(\cdot,\cdot)$. Matrix $\widehat{\boldsymbol{\Phi}}$ is obtained from a lower-rank base whose full-rank version spans $\boldsymbol{\Phi}$, and minimizes $\delta(\boldsymbol{\Phi},\widehat{\boldsymbol{\Phi}})$. To keep all the KPCA conditions, following developments are done under the assumption that $\boldsymbol{\Phi}$ is centered. Let us define a d-dimensional base $\boldsymbol{W} \in \mathbb{R}^{D_h \times d}$, such that $\boldsymbol{W} = (\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(d)})$ and $\boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_d$, where $\boldsymbol{w}^{(\ell)} \in \mathbb{R}^{D_h}$, $\ell \in \{1, \ldots, d\}$, and $\boldsymbol{I}_d$ is a d-dimensional identity matrix. Since $d < D$, we can say that the base $\boldsymbol{W}$ is lower-rank. Given this, the low-dimensional space can be calculated by means of a linear projection, as follows:

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{W}. \tag{6}$$

Thus, from Eq. 6, we can write that:

$$\widehat{\boldsymbol{\Phi}} = \boldsymbol{Y}\boldsymbol{W}^\top. \tag{7}$$

If we define $\delta(\boldsymbol{\Phi},\widehat{\boldsymbol{\Phi}}) = ||\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}||_F^2$, where $||\cdot||_F$ denotes Frobenius norm. By design, a squared distance is used so that the problem has a dual version that can be readily expressed as a quadratic form. Accordingly, we can pose the following primal optimization problem:

$$\begin{aligned} \min_{\boldsymbol{W}} \quad & \|\boldsymbol{\Phi} - \widehat{\boldsymbol{\Phi}}\|_F^2 \\ \text{s. t.} \quad & \boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_d, d < D, \\ & \boldsymbol{Y} = \boldsymbol{W}^\top \boldsymbol{\Phi}, \end{aligned} \tag{8}$$

which can be expressed as a dual problem in the form:

$$\begin{aligned} \max_{\boldsymbol{W}} \quad & \text{tr}(\boldsymbol{W}^\top \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \boldsymbol{W}) \\ \text{s. t.} \quad & \boldsymbol{W}^\top \boldsymbol{W} = \boldsymbol{I}_d, d < D, \\ & \boldsymbol{Y} = \boldsymbol{W}^\top \boldsymbol{\Phi}, \end{aligned}$$

having as a feasible solution for $\boldsymbol{W}$ and $\boldsymbol{Y}$ the eigenvectors associated to the d largest eigenvalues of $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top$ and $\boldsymbol{\Phi}^\top\boldsymbol{\Phi}$, respectively. This theorem is known as optimal low-rank representation widely discussed and demonstrated in [15].

Kernel Trick: Furthermore, by following the Mercer's condition or the so-called kernel trick, we can introduce a kernel function $k(\cdot, \cdot)$, which estimates the inner product $\phi(\boldsymbol{x}_i)^\top \phi(\boldsymbol{x}_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_i)$. By gathering all the pairwise kernel values, we can write a kernel matrix $\boldsymbol{K} = [k_{ij}]$ as:

$$\boldsymbol{K} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi}, \tag{9}$$

where $k_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and $i, j \in \{1, \ldots, N\}$. As mentioned above, such a kernel matrix captures the data topology and may be considered as a similarity matrix.

3.3 DataVis via Interactive DR

Quite intuitively, one can infer that the premise underlying the use of DR for DataVis purposes is to making directly intelligible the information of a high-dimensional dataset by displaying it into a representation in 3 or less dimensions.

Besides, the incorporation of interactivity into the DR technique itself or DR-based DataVis interfaces enables the users (even non-expert ones) to select a method or tune parameters thereof in an intuitive fashion.

4 Concept of Interaction Model (IM) for DR

4.1 Definition of IM

Herein, the interaction is considered as the ability to incorporate in a readily manner the user's criterion into the stages of the data exploration process. In this case, the DR is the stage of interest. Particularly, we refer to interactivity to the possibility of tuning parameters or selecting methods within an interactive interface. As traditionally done in previous works [8], the interactivity consists of a mixture of functions or elements representing DR techniques. In the following, we formally define the concept of IM:

Definition 1 (Interaction model (IM)). *Under a certain topological space $\mathcal{V}$ and given a set of M functions or elements ($f_m \in \mathcal{V}$) representing M different dimensionality reduction techniques ($\boldsymbol{f} = \{f_1, \ldots, f_M\}$) and a set of weighting*

factors $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_M\}$*, an IM is defined as any possible mixture* $\tilde{f} \in \mathcal{V}$ *of such a set, in the form:*

$$\begin{aligned} \mathrm{IM}\{\cdot\}: \quad & \mathcal{V}^M \times \mathbb{R}^M \to \mathcal{V} \\ & \boldsymbol{f},\ \boldsymbol{\alpha} \quad \mapsto \mathrm{IM}_{\boldsymbol{\alpha}}\{\boldsymbol{f}\}, \end{aligned} \tag{10}$$

such that $\widetilde{f} = \mathrm{IM}_{\boldsymbol{\alpha}}\{\boldsymbol{f}\}$.

Then, IM can be itself understood as a suitable building block for powering data analysis systems with interactive dimensionality-reduction-driven visualization stages, as depicted in 2.

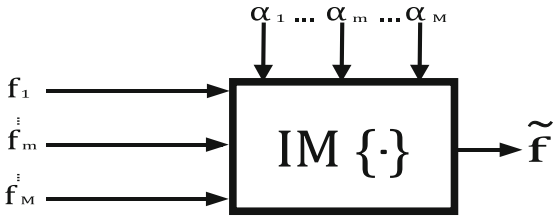


Fig. 2. Graphical explanation of the interaction model (IM) concept aimed at producing a final data representation $\widetilde{f}$ in terms of a set of functions $\{f_1, \ldots, f_M\}$ and their corresponding weighting factors $\{\alpha_1, \ldots, \alpha_M\}$.

4.2 Linear IM

As a particular case of the Definition 1, we can define the linear IM as follows:

Definition 2 (Linear IM (LIM)). *Under a certain vector space* $\mathbb{E}$ *and given a set of* M *functions or elements* ($f_m \in \mathcal{V}$) *representing* M *different dimensionality reduction techniques* ($\boldsymbol{f} = \{f_1, \ldots, f_M\}$) *and a set of weighting factors* $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_M\}$*, a LIM is defined as the weighted sum* $\tilde{f} \in \mathcal{V}$ *in the form:*

$$\begin{aligned} \mathrm{LIM}\{\cdot\}: \quad & \mathbb{E}^M \times \mathbb{R}^M \to \mathbb{E} \\ & \boldsymbol{f},\ \boldsymbol{\alpha} \quad \mapsto \mathrm{LIM}_{\boldsymbol{\alpha}}\{\boldsymbol{f}\}, \end{aligned} \tag{11}$$

such that $\widetilde{f} = \mathrm{LIM}_{\boldsymbol{\alpha}}\{\boldsymbol{f}\} = \sum_{m=1}^{M} \alpha_m f_m$.

4.3 DR-techniques Representation

As explained [20], spectral DR methods are susceptible to be represented as kernel matrices. Also, in [15] is demonstrated that, when incorporated into a KCPA algorithm, such kernel matrices reach the same low-dimensional spaces

as those obtained by the original DR methods. Let us consider the following kernel representations for three well-known spectral DR approaches:

Classical Multi-dimensional Scaling (CMDS): CMDS kernel can be expressed as the double centered, squared Euclidean distance matrix $\boldsymbol{D} \in \mathbb{R}^{N \times N}$ so

$$\boldsymbol{K}_{CMDS} = -\frac{1}{2}(\boldsymbol{I}_N - \boldsymbol{1}_N\boldsymbol{1}_N^\top)\boldsymbol{D}(\boldsymbol{I}_N - \boldsymbol{1}_N\boldsymbol{1}_N^\top), \tag{12}$$

where the ij entry of $\boldsymbol{D}$ is given by $d_{ij} = ||\boldsymbol{x}_i - \boldsymbol{x}_j||_2^2$.

Laplacian Eigenmaps (LE). Since KPCA is a maximization of the high-dimensional covariance represented by a kernel, LE can be represented as the pseudo-inverse of the graph Laplacian $\boldsymbol{L}$:

$$\boldsymbol{K}_{LE} = \boldsymbol{L}^\dagger, \tag{13}$$

where $\boldsymbol{L} = \boldsymbol{\mathcal{D}} - \boldsymbol{S}$, $\boldsymbol{S}$ is a similarity matrix and $\boldsymbol{\mathcal{D}} = \text{Diag}(\boldsymbol{S}\boldsymbol{1}_N)$ is the degree matrix.

Locally Linear Embedding (LLE): A kernel for LLE can be approximated from a quadratic form in terms of the matrix $\boldsymbol{\mathcal{W}}$ holding linear coefficients that sum to 1 and optimally reconstruct the original data matrix $\boldsymbol{X}$. Define a matrix $\boldsymbol{M} \in \mathbb{R}^{N \times N}$ as $\boldsymbol{M} = (\boldsymbol{I}_N - \boldsymbol{\mathcal{W}})(\boldsymbol{I}_N - \boldsymbol{\mathcal{W}}^\top)$ and λ_{max} as the largest eigenvalue of $\boldsymbol{M}$. Kernel matrix for LLE is in the form

$$\boldsymbol{K}_{LLE} = \lambda_{max}\boldsymbol{I}_N - \boldsymbol{M}. \tag{14}$$

4.4 Use of IM in Interactive DR

To explain how to use the IM for DR purposes, consider the kernel matrices from Sect. 4.3, by setting $M = 3$, as well as $f_1 = \boldsymbol{K}^{(1)} = \boldsymbol{K}_{CMDS}$, $f_2 = \boldsymbol{K}^{(2)} = \boldsymbol{K}_{LE}$ and $f_3 = \boldsymbol{K}^{(3)} = \boldsymbol{K}_{LLE}$. Then, the linear IM applied to spectral DR methods using kernel representations can be expressed as follows:

$$\widetilde{\boldsymbol{K}} = \widetilde{f} = \text{LIM}_{\boldsymbol{\alpha}}\{\boldsymbol{f}\} = \text{LIM}_{\boldsymbol{\alpha}}\{\boldsymbol{K}^{(1)}, \ldots, \boldsymbol{K}^{(3)}\} = \sum_{m=1}^{M} \alpha_m \boldsymbol{K}^{(m)}. \tag{15}$$

To add the effect of interactivity, the values of weighting factors $\boldsymbol{\alpha}$ can be provided through a user-friendly and intuitive-to-use interface[1], as those reviewed in [8]. Finally, the low-dimensional space $\boldsymbol{Y}$ is calculated by applying KPCA on $\widetilde{\boldsymbol{K}}$, so:

$$\boldsymbol{Y} = \text{KPCA}(\widetilde{\boldsymbol{K}}), \tag{16}$$

where KPCA$(\cdot)$ means that the KPCA problem stated in Sect. 3 is solved with the kernel matrix in its argument. The first notion of this approach (even though with no formal definition) is presented in [4]. Figure 3 graphically explains the use of LIM for DR techniques (DRT) following a KPCA approach.

[1] Some IM-based interfaces are available at https://sdas-group.com/gallery/.

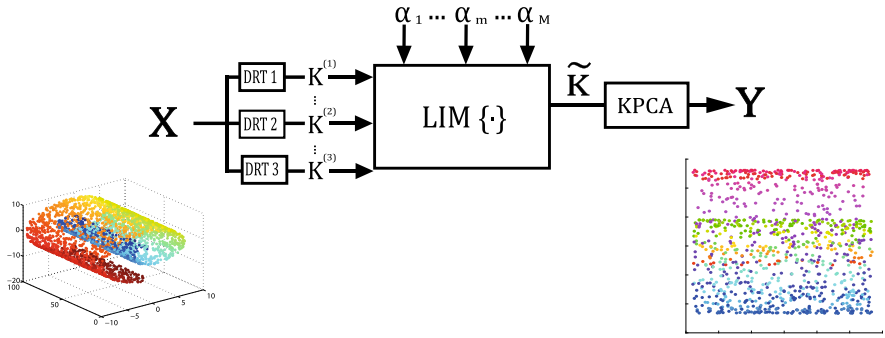


Fig. 3. Graphical explanation of the use of the linear interaction model (LIM) for a KPCA-based DR framework. Specifically, a linear combination of kernel matrices representing DR techniques (DRTs) is considered.

5 Final Remarks

In this work, we have elegantly defined the concept of the so-named interaction model (IM). Such a definition open the possibility of developing more formally new interactive data visualization based on a mixture of dimensionality reduction techniques.

In future works, we will explore and/or develop novel kernel representations arising from other dimensionality reduction methods as well as IM approaches enabling users to readily incorporate their knowledge and expertise into data exploration and visualization.

Acknowledgment. The authors acknowledge to the research project "Desarrollo de una metodología de visualización interactiva y eficaz de información en Big Data" supported by Agreement No. 180 November 1st, 2016 by VIPRI from Universidad de Nariño.

As well, authors thank the valuable support given by the SDAS Research Group (www.sdas-group.com).

References

1. Gou, J., Yang, Y., Yi, Z., Lv, J., Mao, Q., Zhan, Y.: Discriminative globality and locality preserving graph embedding for dimensionality reduction. Expert Syst. Appl. **144**, 113079 (2020)
2. Lee, J.A., Peluffo-Ordóñez, D.H., Verleysen, M.: Multi-scale similarities in stochastic neighbour embedding: reducing dimensionality while preserving both local and global structure. Neurocomputing **169**, 246–261 (2015)
3. Ward, M.O., Grinstein, G., Keim, D.: Interactive Data Visualization: Foundations, Techniques, and Applications. CRC Press (2010)
4. Peluffo-Ordónez, D.H., Alvarado-Pérez, J.C., Lee, J.A., Verleysen, M., et al.: Geometrical homotopy for data visualization. In: European Symposium on Artificial Neural Networks (ESANN 2015). Computational Intelligence and Machine Learning (2015)

5. Salazar-Castro, J., Rosas-Narváez, Y., Pantoja, A., Alvarado-Pérez, J.C., Peluffo-Ordóñez, D.H.: Interactive interface for efficient data visualization via a geometric approach. In: 2015 20th Symposium on Signal Processing, Images and Computer Vision (STSIVA), pp. 1–6. IEEE (2015)
6. Rosero-Montalvo, P., et al.: Interactive data visualization using dimensionality reduction and similarity-based representations. In: Beltrán-Castañón, C., Nyström, I., Famili, F. (eds.) CIARP 2016. LNCS, vol. 10125, pp. 334–342. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-52277-7_41
7. Rosero-Montalvo, P.D., Peña-Unigarro, D.F., Peluffo, D.H., Castro-Silva, J.A., Umaquinga, A., Rosero-Rosero, E.A.: Data visualization using interactive dimensionality reduction and improved color-based interaction model. In: Ferrández Vicente, J.M., Álvarez-Sánchez, J.R., de la Paz López, F., Toledo Moreo, J., Adeli, H. (eds.) IWINAC 2017. LNCS, vol. 10338, pp. 289–298. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59773-7_30. (Cited by 8)
8. Umaquinga-Criollo, A.C., Peluffo-Ordóñez, D.H., Rosero-Montalvo, P.D., Godoy-Trujillo, P.E., Benítez-Pereira, H.: Interactive visualization interfaces for big data analysis using combination of dimensionality reduction methods: a brief review. In: Basantes-Andrade, A., Naranjo-Toro, M., Zambrano Vizuete, M., Botto-Tobar, M. (eds.) TSIE 2019. AISC, vol. 1110, pp. 193–203. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-37221-7_17
9. Amin, A., et al.: Cross-company customer churn prediction in telecommunication: a comparison of data transformation methods. Int. J. Inf. Manag. **46**, 304–319 (2019)
10. Peluffo, D., Lee, J., Verleysen, M., Rodríguez-Sotelo, J., Castellanos-Domínguez, G.: Unsupervised relevance analysis for feature extraction and selection: a distance-based approach for feature relevance. In: International Conference on Pattern Recognition, Applications and Methods-ICPRAM (2014)
11. Cao, H., Bernard, S., Heutte, L., Sabourin, R.: Dissimilarity-based representation for radiomics applications. CoRR abs/1803.04460 (2018)
12. Zhong, G., Wang, L.N., Ling, X., Dong, J.: An overview on data representation learning: from traditional feature learning to recent deep learning. J. Finance Data Sci. **2**(4), 265–278 (2016)
13. Lee, J.A., Verleysen, M.: Nonlinear Dimensionality Reduction. Springer, Heidelberg (2007). https://doi.org/10.1007/978-0-387-39351-3
14. Borg, I., Groenen, P.J.: Modern Multidimensional Scaling: Theory and Applications. Springer, Heidelberg (2005). https://doi.org/10.1007/0-387-28981-X
15. Peluffo-Ordóñez, D.H., Lee, J.A., Verleysen, M.: Generalized kernel framework for unsupervised spectral methods of dimensionality reduction. In: 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp. 171–177. IEEE (2014)
16. Peluffo-Ordóñez, D.H., Lee, J.A., Verleysen, M.: Short review of dimensionality reduction methods based on stochastic neighbour embedding. In: Villmann, T., Schleif, F.-M., Kaden, M., Lange, M. (eds.) Advances in Self-Organizing Maps and Learning Vector Quantization. AISC, vol. 295, pp. 65–74. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07695-9_6
17. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 1373–1396 (2003)
18. Zhang, Z., Wang, J.: MLLE: modified locally linear embedding using multiple weights. In: Advances in Neural Information Processing Systems, pp. 1593–1600 (2007)

19. Hinton, G.E., Roweis, S.T.: Stochastic neighbor embedding. In: Advances in Neural Information Processing Systems, pp. 857–864 (2003)
20. Ham, J., Lee, D.D., Mika, S., Schölkopf, B.: A kernel view of the dimensionality reduction of manifolds. In: Proceedings of the Twenty-First International Conference on Machine Learning, p. 47. ACM (2004)