2016 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2016, Yugawara, Japan, November 27th-30th, 2016

Clustered Multidimensional Scaling with Rulkov Neurons

Thomas Ott†, Martin Schüle†, Jenny Held‡, Carlo Albert‡ and Ruedi Stoop*

†ZHAW Zurich University of Applied Sciences
8820 Wädenswil, Switzerland
‡EAWAG Swiss Federal Institute of Aquatic Science and Technology
8600 Dübendorf, Switzerland
* Institute of Neuroinformatics, ETH/UZH Zurich
Winterthurerstrasse 190, 8047 Zurich, Switzerland
Email: thomas.ott@zhaw.ch, scli@zhaw.ch, Jenny.Held@eawag.ch, Carlo.Albert@eawag.ch, ruedi@ini.phys.ethz.ch

Abstract—When dealing with high-dimensional measurements that often show non-linear characteristics at multiple scales, a need for unbiased and robust classification and interpretation techniques has emerged. Here, we present a method for mapping high-dimensional data onto low-dimensional spaces, allowing for a fast visual interpretation of the data. Classical approaches of dimensionality reduction attempt to preserve the geometry of the data. They often fail to correctly grasp cluster structures, for instance in high-dimensional situations, where distances between data points tend to become more similar. In order to cope with this clustering problem, we propose to combine classical multi-dimensional scaling with data clustering based on self-organization processes in neural networks, where the goal is to amplify rather than preserve local cluster structures. We find that applying dimensionality reduction techniques to the output of neural network based clustering not only allows for a convenient visual inspection, but also leads to further insights into the intra- and inter-cluster connectivity. We report on an implementation of the method with Rulkov-Hebbian-learning clustering and illustrate its suitability in comparison to traditional methods by means of an artificial dataset and a real world example.

1. Introduction

Visual inspection of scatterplots is a fast and common way to interpret data. Yet, high-dimensional data can be difficult to interpret. Hence, dimensionality reduction is often performed, aiming to achieve a compact representation of the data in two or three dimensions. A plethora of different techniques for dimensionality reduction has been proposed (see e.g. [1] for an overview). The most common methods are linear techniques such as principal component analysis PCA and multidimensional scaling (MDS). These methods often have difficulties with the representation of real world data, as in many cases high-dimensional data is generated by non-linear processes, resulting in highly non-trivial structures in the space of measurements [2]. Hence, during the last two decades, effort has been put in the development of nonlinear dimensionality reduction techniques in order to map nonlinear manifolds, e.g, kernel PCA, Isomap, Locally Linear Embedding, Diffusion Maps or t-SNE [3], to name but a few.

Typically, these approaches attempt to preserve the geometry of the data, at least at a local scale. That is, the distances between data points in a local neighborhood shall be preserved in the low-dimensional representation. Less attention has been paid to mapping cluster-like structures of a possibly complex (e.g. nonconvex) shape that can be intrinsically high-dimensional. Here, classical approaches often struggle because of unclear cluster boundaries that might even be obscured by data noise or due to the curse of dimensionality, notably the fact that distances between data points tend to become more equal with growing dimensionality. Real examples of such data situations are for instance encountered when dealing with flow cytometry data. In order to highlight the prevalent intrinsic cluster structures in the low-dimensional representation, we suggest to no longer stick to the goal of preserving the local geometry. Our idea is to combine self-organizing clustering processes with multi-dimensional scaling in order to enhance local cluster structures. To this end, we employ the Rulkov-Hebbian-learning clustering algorithm (RHLC) that has recently been introduced in [2] and can be considered an efficient cortex-inspired clustering method. The method presented here however also works with similar algorithms such as HLC with integrate-and-fire neurons [4]. In the following we briefly discuss the two ingredients of our approach, MDS and RHLC, and then present their application to two example data sets.

2. Multi-Dimensional Scaling and Extensions

We are given high-dimensional data vectors $x_i \in \mathbb{R}^d$ ($i \in \{1, ..., n\}$) that form the rows of a matrix X. The goal of classical multidimensional scaling MDS is to find low-dimensional (often 2-dimensional) reconstruction vectors y_i (or Y as a matrix) that minimize the following cost function [1]

$$\Phi(y) = \sum_{(i,j)} (d_{ij} - ||y_i - y_j||)^2 \qquad (1)$$

where $d_{ij} = d_{ji}$ is the Euclidian distance between the original data points and $\|y_i - y_j\|$ denotes the Euclidian distance in the reconstruction space. The minimum of (1) can be calculated based on the eigendecomposition of the Gram matrix $K = XX^T$ which can be obtained by double-centering the distance matrix $D = (d_{ij})$ [1]. The $m = 2$ coordinates of y_i are then given by $Y = E_m \Lambda_m^{1/2}$, where Λ_m is the diagonal matrix with the m largest eigenvalues and E_m is the matrix of the corresponding eigenvectors of K. For Euclidian distances, principal component analysis (PCA) is identical to MDS due to the fact that the eigenvectors of the Gram Matrix and the eigenvectors of the covariance matrix are directly related. MDS can also be applied for any other distance matrix, a fact that is exploited by the Isomap approach with the goal to better account for describing the neighborhood of datapoints on curved manifolds. In Isomap, the distances are calculated by first constructing a graph, in which every point is connected to its k nearest neighbors. The distance between two points is then set to be the length of the shortest path in the graph [5].

In general metric MDS, the goal of (1) is relaxed by replacing d_{ij} with $\delta_{ij} = f(d_{ij})$, where $f()$ leads to a new symmetric nonnegative (i.e. positive semi-definite) dissimilarity matrix. Isomap and similar methods can be interpreted as a kernel PCA method with the advantage of having a method that also works for datapoints that are not in the training sample.

3. From Rulkov Clustering to RHLC-MDS/Isomap

Rulkov-Hebbian-learning clustering (RHLC) is based on self-organization processes in a network of Rulkov neurons, letting clusters arise from the interplay between neural activity and changes in the network connectivity. Hebbian learning-based clustering (HLC) has been introduced in previous work as a remedy for the intrinsic shape biases introduced by standard clustering algorithms [2],[6]. Recently, RHLC has been developed, functioning on the same Hebbian-learing principle but making use of the more efficient map-based Rulkov neuron dynamics [7]. Generally, in HLC, every data item is interpreted as a dynamical unit with node dynamics, which are allowed to interact via a k-nearest neighbors graph. The pair-wise interaction strength between nodes is weighted so as to represent the local distances between the data items. In an iterative process, using Hebbian learning, the network structure is updated such that the weights between dynamically similar nodes are strengthened, while a counter-acting mechanism aiming to preserve the level of activity in the network causes the weights between less similar nodes to decrease. The final graph structure can thus be represented by the weights $w_{ij}^{\infty} \in [0, 1]$, where strongly connected nodes will have a large coupling strength. The connectivity of nodes across a cluster can in this way easily be represented without the need for direct interaction, and thus without shape bias. For clustering, all weights below a threshold are deleted and the remaining (sub)graphs define the final clusters.

Here, for the purpose of data visualization, we are not using a hard cluster assignment. Rather, we are interested in the final weight matrix produced by RHLC as it reflects an amplified similarity between the data items. On the whole, RHLC performs a mapping

$$d_{ij} \rightarrow \delta_{ij} = f(d_{ij}) = 1 - w_{ij}^{\infty}. \quad (2)$$

The matrix defined by (δ_{ij}) can be interpreted as a 'clustered distance matrix' and can serve as input for MDS, giving rise to our RHLC-MDS method. As an alternative method, we use (δ_{ij}) as input for Isomap. This RHLC-Isomap method is motivated by the observation that the basic connection matrix of RHLC reflects a k nearest neighbor graph as it is also used for Isomap.

4. Experiments and Results

Datasets: We test our methods on the basis of two datasets.

1. The first dataset is an artificial dataset in 3D, where two clusters cannot be discriminated by comparing inter- and intra-cluster distances. This is due to the fact that the inter-cluster distance in $V3-$direction is smaller than the clusters' extension in $V1-$ and $V2-$direction (see Fig.1 a)) and can be considered a low-dimensional simulation of the curse of dimensionality. The situation is reminiscent of many real-life data sets, where the data may stretch multiple scales in different dimensions. Additionally, the clusters are embedded into a noisy background of datapoints. The clusters are colored in Fig.1 b).

2. The second dataset contains 2443 flow cytometry measurements from 3 different phytoplankton species, reduced to 8 descriptors. The scatterplots for a pairwise selection of variables are shown in Fig.1 c), where the species clusters are color-coded.

Compared Methods and Evaluation: We compare 4 methods: standard MDS and Isomap and our clustered versions, RHLC-MDS and RHLC-Isomap. In order to assess the convenience of reading out the (by virtue of construction) expected clusters or classes from the 2D scatterplot we use the ratio of the mean inter-class distances and the mean of the distances between the expected classes A as an indicator, i.e.

$$r = \frac{< d_{within} >}{< d_{between} >}, \quad (3)$$

where the distances are meant to be the Euclidian distances in the two-dimensional projections. For compact and clearly separated clusters, a small r is expected.

RHLC as well as Isomap involve parameters to be tuned. For the following, the tuning was made manually based on a visual inspection of the results. Hence, they reflect optimal solutions at a pragmatic level.

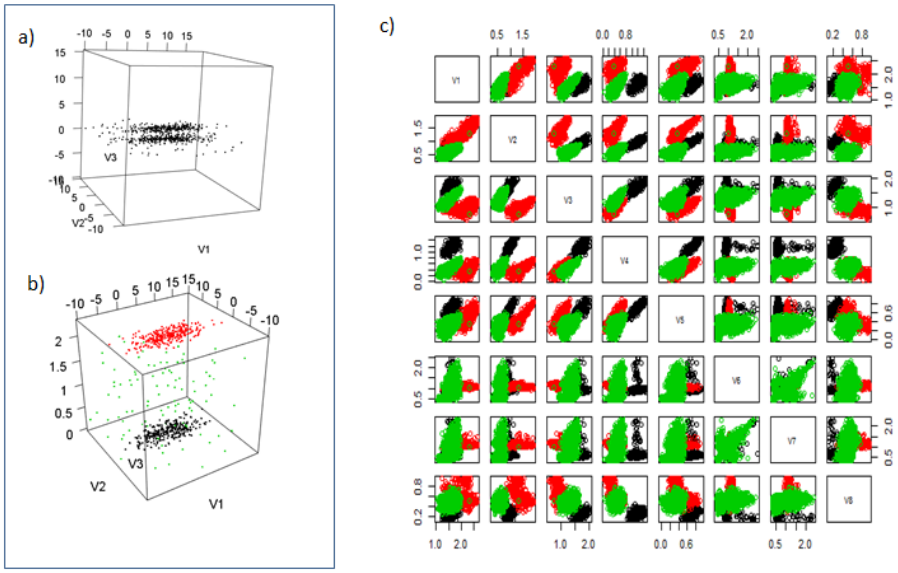

Figure 1: Datasets for evaluation. a)/b) artifical dataset, c) phytoplankton dataset. The expected clusters are color-coded.

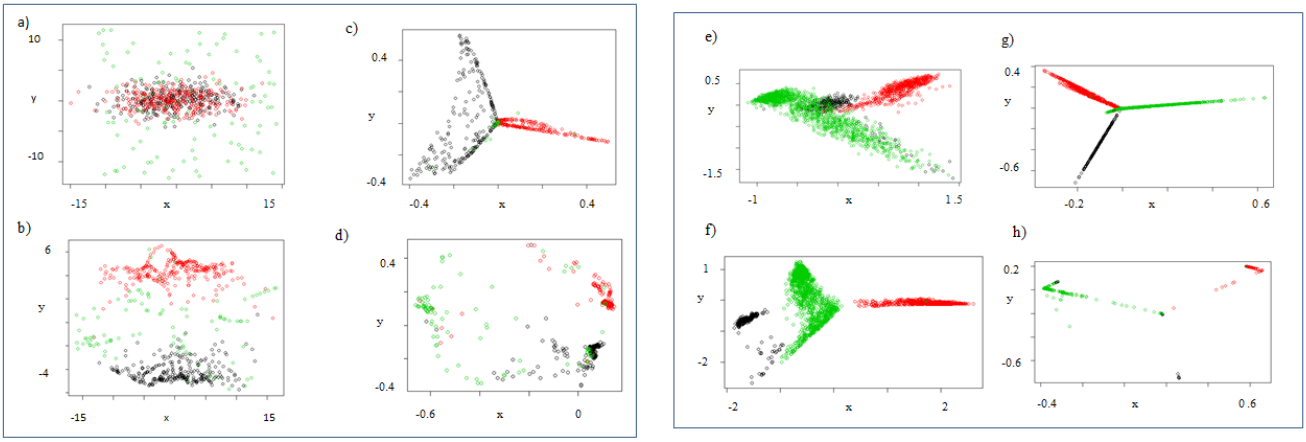

Figure 2: Visualization results for 1) artificial dataset with a) MDS, b) Isomap, c) RHLC-MDS, d) RHLC-Isomap, 2) phytoplankton dataset with e) MDS, f) Isomap, g) RHLC-MDS, h) RHLC-Isomap

Results: The visualization results for both datasets and all the compared methods are subsumed in Fig. 2. We observe by first focussing at the results for the artificial dataset (Fig. 2 a)-d)) that normal MDS fails to display the clusters separately. While normal Isomap seems to do a satisfying mapping job regarding the internal structure of the data, the RHLC versions of MDS and Isomap both better display the cluster structure of the data. However, RHLC-MDS splits the black cluster into two subunits and, generally, tends to arrange the points in linear chains with an overall center of mass in the center of the coordinate system. For RHLC-MDS, the background points (noise) are concentrated in the center (Fig. 2 c)). In contrast, RHLC-Isomap separates the noise as an independent cluster (Fig. 2 d)).

Similar observations regarding the visual output characteristics of the methods can be made for the phytoplankton dataset (Fig. 2 e)-h)) with the initial difference that the clusters are more clearly separable and a noisy background is absent.

The evaluation using the indicator r confirms that RHLC-Isomap is clearly superior to the other methods (smallest r) regarding the capability of highlighting the overall cluster structures (Table 1). For RHLC-MDS the r measure indicates a performance that is in the range of normal Isomap,

but better than normal MDS. To some extent, the results can be explained by the chain-like shape of the clusters that results from RHLC-MDS. Take the example of the phytoplankton dataset: while the form makes it convenient for humans to grasp the group structure of the data, the inter-class distances become rather large and hence r is increased. In the case of the artificial dataset, the aspect that RHLC-MDS splits one cluster into two units also leads to an increased r. At the same time, it hints at the existence of an internal cluster structure. In fact, a closer inspection of this cluster reveals two different areas, where the neighbor density of points reaches a maximum. These slight inhomogeneities are amplified by RHLC and are made visible as two branches in the RHLC-MDS plot. The observation is illustrated in Fig. 3. The neighbor density was defined as the number of neighbors within a ball of radius $R = 0.8$ divided by the volume of the ball. Fig. 3 b) shows the neighbor density as a function of the points of the split cluster in the $V_1 - V_2-$projection.

	artifical dataset	phytoplankton dataset
MDS	4.00	1.91
Isomap	1.07	0.70
RHLC-MDS	1.32	0.49
RHLC-Isomap	0.246	0.05

Table 1: r values for artificial dataset and phytoplankton dataset.

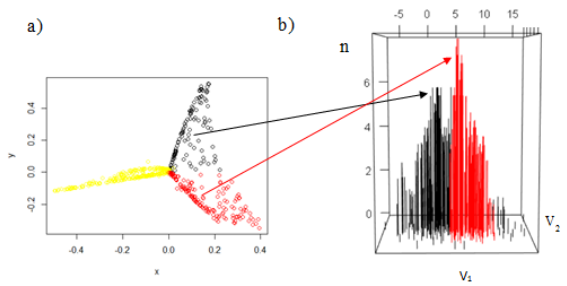

Figure 3: The two branches in the RHLC-MDS plot a) reflect regions of high neighbor density n in the corresponding cluster.

5. Conclusion and Outlook

Neural network based clustering algorithms such as RHLC allow for an unbiased detection of local cluster structures on the basis of self-organization. In this process, the neighborhood structure of the data is encoded as a weighted network that evolves in such a way that inhomogeneities are amplified. Hence, the emergence of clear cluster structures is possible even in cases, where the detection of clusters is very challenging, e.g. when facing high-dimensional measurements with non-linear cluster characteristics. Here, we demonstrated that dimensionality reduction techniques such as MDS and Isomap allow for a low-dimensional representation of the evolved clustering network, shedding light on both the intra- and the inter-cluster structure. While RHLC-Isomap separates clusters more clearly, RHCL-MDS elucidates the internal structure of clusters. This also allows for a more robust determination of the most natural number of clusters by means of a quick visual inspection.

The method employs clustering as a preprocessing step for a dimensionality reduction (DR) step, which switches the role of the steps in comparison to the standard data analysis procedure. Thus, for future research, the results suggest a clustering method that iteratively applies DR and clustering techniques. Alternatively, DR and self-organized clustering can be hybridized in one method, following the idea developed in [8].

References

[1] L. van der Maaten, E. Postma, J. van den Herik, "Dimensionality reduction: a comparative review," *Tech. rep., MICC, Maastricht University*, 2008.

[2] R. Stoop, K. Kanders, T. Lorimer, J. Held, C. Albert, "Big data naturally rescaled," *Chaos Solitons and Fractals*, doi: 10.1016/j.chaos.2016.02.035 2016.

[3] L. van der Maaten, G. E. Hinton, "Visualizing High-Dimensional Data Using t-SNE," *Journal of Machine Learning Research*, 9, pp. 2579-2605, 2008.

[4] F. Landis, T. Ott, R. Stoop, "Hebbian Self-organizing Integrate-and-Fire Networks for Data Clustering," *Neural Computation* 22, pp. 273–288, 2010.

[5] J. B. Tenenbaum, V. de Silva, J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, 290(5500), pp. 2319-2323, 2000.

[6] F. Gomez, R.L. Stoop, R. Stoop, "Universal dynamical features preclude standard clustering in a large class of biochemical data," *Bioinformatics* 20, pp. 2486–2493, 2014.

[7] N.F. Rulkov, "Modeling of spiking-bursting neural behavior using two-dimensional map," *Phys. Rev. E* 65, p.041922, 2002.

[8] T. Ott, T. Eggel, M. Christen, "Generating Low-Dimensional Denoised Embeddings of Nonlinear Data with Superparamagentic Agents," *Proceedings of the 2014 International Symposium on Nonlinear Theory and its Applications (NOLTA2014)*, 2014.