EuroVis 2013 Workshop on Visual Analytics using Multidimensional Projections (2013)
M. Aupetit and L. van der Maaten (Editors)

Short Papers

ClassiMap : a Supervised Mapping Technique for Decision Support

Sylvain Lespinats[1], Michaël Aupetit[2]

[1] sylvain.lespinats@cea.fr – CEA, LITEN, Le bourget du lac, F-73375, France
[2] michael.aupetit@cea.fr – CEA, LIST, Gif-sur-Yvette, F-91191, France

Abstract
Dimensionality reduction algorithms may be of great help as decision support, representing the information as a map which summarizes the data similarities. When data come with an assigned class label, such a map can be used to check the quality of the labeling detecting class outliers or data near decision boundary, or to evaluate the relevance of the similarity measure used for the mapping from which to derive a good classification space. However, state-of-the-art mapping techniques are either unsupervised, not considering the class labels, or supervised, considering it but putting too much emphasis on the class information. The result is that well separated classes can be mapped as overlapping with the unsupervised techniques, while overlapping classes can be mapped as clearly separated with the supervised techniques, so none of these maps tends to show the truth about the inter-class and between-class high-dimensional structure. We designed ClassiMap, a supervised mapping technique which come over these limits by exploiting the unavoidable tears and false neighborhoods mapping distortions to preserve at best the class structure through the mapping. We compare it to other supervised mapping techniques in labeled data visual exploration tasks.

Paper type: Optimizing embeddings for visual analysis

1. Introduction

1.1 Context

Decision support systems based on automatic classification techniques [5] rely on classification scores like precision and recall, to help the user in their daily data classification task. However users faced with such a black box system, are prone either to blind trust or to mistrust these scores. In the former case, the user relies too much on the system which comes with its own risks, while in the latter the system is not used at all and so is useless.

We expect that visualization techniques of multidimensional labeled data (*i.e.* data with an assigned class label) could help the user to trust the decision support system, better understanding why the system provides this classification score for a new data point, by positioning the new point on a map relative to the global class structures of the data.

1.2. Multidimensional scaling and mapping distortions

Multidimensional scaling techniques are used for visual analysis of high-dimensional datasets. Original high-dimensional data are mapped as points into a 2-dimensional vector space (the map) by attempting to preserve at best some similarity measure between the data points. The label information is usually color-coded on the map. Unsupervised mapping techniques do not consider the class information while supervised mapping techniques do. Whether supervised or not, any mapping of a dataset from a high-dimensional (**HD**) vector space to a lower dimensional (**map**) vector space necessarily comes with distortions. Let the boolean function $N_s(x,y)$ be true if vectors x and y are neighbors in the space s and false otherwise, then *tears* (**T**) and *false neighborhoods* (**FN**) are the two types of distortions (Figure 1) which are relative to such a neighborhood function [1][4]. Tears and false neighborhoods can occur simultaneously relatively to the same data.

Unsupervised mapping techniques do not consider the class information, so due to the above mapping distortions, they can map well separated classes as overlapping ones. In contrast, state-of-the-art supervised mapping techniques take into account

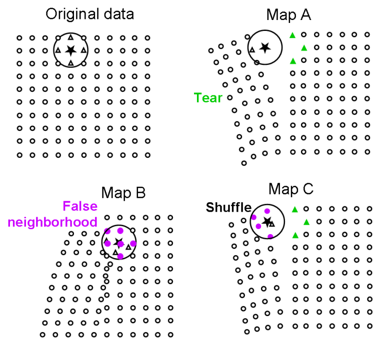

Figure 1: *Mapping distortions relative to the star-shaped data point. The original dataset (upper left insert) is a regular grid of points on the plane. Triangles are neighbors of the star. The large circle around the star defines its neighborhood in both the original and projection spaces. Map A (upper right insert) shows a tear where some of the original neighbors do not remain neighbors on the map (green triangles). Map B (down left insert) shows a false neighborhood where some neighbors on the map (magenta circles) are not original neighbors of the star. Please notice that tears and false neighborhoods can occur simultaneously as on Map C (down right insert).*

DOI: 10.2312/PE.VAMP.VAMP2013.017-020

both the original data similarities and the data class labels, but they are designed as visual classifiers putting too much emphasis on the class information, so they can map overlapping classes as clearly separated ones. Finally, none of these mapping techniques tends to show the truth about the inter-class and between-class high-dimensional structure likely to help the user to trust the classifier of the decision support system.

1.3. Purposes of supervised mapping techniques

Supervised mapping techniques aim to import in the visual field typical analytic tasks which are usually performed on high-dimensional labeled data with data mining techniques: namely data classification [10] and labeled data exploration [11]. These tasks can be transposed in the visual domain as:

- *(i) visual data classification*: a new unlabeled data point projected onto the map is visually assigned to the majority class among its labeled neighbors;
- *(ii) mislabeled data detection*: a labeled data whose neighbors on the map are assigned to a different class is assumed to be mislabeled;
- *(iii) class-boundary data detection*: a labeled data whose neighbors on the map are mixed between its own class and other classes is assumed to lie near the HD decision boundary of the classifier;
- *(iv) similarity measure evaluation*: a map showing much class overlapping implies a similarity measure which is not relevant to classification;
- *(v) class structure summarization*: the map is used to provide a visual global summary of inter-class and between-class HD structures.

The *visual data classification* task *(i)* implies that data with different labels should be mapped as separated points, and data with the same label should be mapped as neighbors, in order to increase the probability that a point has the same label as its neighbors on the map. In this case, the labels are taken for granted and the mapping technique is designed as a supervised classifier which aims to separate the class and incidentally to preserve the HD data similarities. The *mislabeled* and *class-boundary data detection* tasks *(ii)* and *(iii)* assume that the similarity is well preserved on the map and correlated to the class labels, so a point with a label different from the majority one of its neighbors is assumed to be mislabeled or near the decision boundary in the HD space. The *similarity measure evaluation* task *(iv)* assumes that the similarity is well preserved on the map and the class labels are all correct, so if points with different labels appear in the same area of the map, that means they are similar in the HD space and so the similarity measure is not correlated to the class labels. The *class structure summarization* task *(v)* assumes that data similarities are all preserved at best.

The task *(i)* is related to data classification [10] which takes the class labels for granted and attempts to provide a classifier which in a probabilistic framework is based on a model of the distribution of the classes given the data similarities. The tasks *(ii)* to *(v)* are related to labeled data exploration [11] where the class-labels come as a supplementary characteristic of the data, so the analysis focuses on the joint distribution of both classes and data similarities.

In this work we focus on the labeled data exploration tasks *(ii)* to *(v)* through a multidimensional projection, what we call Visual Exploration (VE) tasks in the following.

1.4. Limits of existing supervised mapping techniques

Most of the state-of-the-art supervised mapping techniques have been designed to comply with the visual data classification task *(i)*. For instance, the Supervised Locally Linear Embedding (S-LLE) [2], the Supervised Isomap (S-Isomap) [3] and the S-NeRV [4]. Indeed, these techniques explicitly modify the HD similarity measure according to the class assignment of the data. Usually, the similarity is decreased between data assigned to different classes, while it is increased between data with the same label. The mapping technique "Map" appearing in the name of the supervised mapping technique "S-Map", is unsupervised and used to map the data with the class-modified HD similarity measure (*e.g.* S-Isomap uses Isomap to map the data based on class-modified similarities). The Linear Discriminant Analysis (LDA) [6] is a linear projection technique which also provides a map enhancing class separation.

The VE tasks strongly rely on the good preservation of the HD data similarities on the map to get reliable inference about the original HD class structures. So the above state-of-the-art supervised mapping techniques do not comply with the VE tasks as they modify the original HD similarity measure. Moreover modifying the original similarity measure corrupts the primary available information generating distortions which add up to the subsequent unsupervised mapping distortions in an unpredictable way, making VE inference harder and distance-based mapping quality evaluation tools useless [7]. Another issue is that class separation is enforced even when classes strongly overlap in the HD space.

This tendency to artificially separate the different classes is illustrated with a simple example in the figure 2: original data points are located at the nodes of a 2-dimensional square grid with random and balanced class assignment. The original space and the projection space are both 2-dimensional planes so that it is obvious that the optimal mapping is the distortion-free identity function which maps each data onto itself. We used the original data points as the initial positioning of their image through the mapping and then start the optimization process of each technique. This simple experiment is a sanity check: we expect the mapping of the original data and the original data themselves to be strictly identical because no dimension reduction occurs. Figure 2 shows that the above supervised techniques designed for visual data classification do not preserve the original pairwise distances and tend to separate the classes while both classes obviously overlap in the original space. In contrast, ClassiMap, our proposed supervised mapping technique provides a perfect mapping as well as DD-HDS [8], an unsupervised mapping technique.

It seems that unsupervised mapping techniques could be used as well to provide maps for the VE tasks as they aim at preserving the HD similarities. However we expect that taking into account the class labels during the mapping process can be useful to preserve the most important information about the class structures while dealing with the unavoidable mapping distortions.

In the following we present ClassiMap, a supervised mapping technique dedicated to labeled data Visual Exploration tasks for decision support systems, which exploit the unavoidable mapping distortions to better preserve the class structure and to enable more reliable inference about it from the map. The main idea is to drive the unavoidable mapping distortions where they are the less detrimental to the global class structure visualization, namely driving tears between the classes and false neighborhoods within the classes.

Table 1: *Notation summary*

d_{ij}	Distance between items i and j in the original HD space.
d_{ij}'	Distance between data points i and j on the map.
F	Weighting function which emphasizes small distances.
$S_m(i,j)$	Stress function between items i and j for the mapping technique *m*.
A_{ij}	Class co-membership of data points i and j (1 if their classes are identical, 0 if they belong to two different classes).

2. The ClassiMap technique

2.1. The ClassiMap stress function

Stress-driven mapping techniques optimize the position of the data points on the map so as to make the distances as close as possible to the ones in the original HD space according to a given cost function called the *global stress*). The global stress takes the general form of a sum over all pairs of points (2) of *local stress* functions $S_m(i,j)$ between points *i* and *j* (1):

$$S_m(i,j) = \left|d_{ij} - d_{ij}'\right|^p \times F_m(.) \quad (1) \qquad \text{(usually, } p \text{ equals 1 or 2)}$$

$$E_m = \sum_{i,j} S_m(i,j) \quad (2)$$

Generally, F is a decreasing function that depends on the distances in the HD or/and on the map. It is well-known that stress-driven mapping techniques tend to commit false neighborhoods (respectively tears) when F depends on HD distances (respectively on map distances) [1]. Except in trivial cases (as in the case shows in the figure 2), all the original pairwise distances cannot be preserved [1] and the choice of the function F strongly influences the type of distortions which are released on the map. Therefore F gives us a chance to control the type of distortions and to attempt to drive them in the less detrimental place.

ClassiMap relies on this characteristic: in order to take into account the class information, ClassiMap adapts the weighing function F rather than modifying the original distances. As a direct consequence, distortions are avoided each time it is possible as in unsupervised mapping techniques. The optimization process is designed so as to drive the unavoidable distortions where they are the less detrimental, that is to say tears are driven between classes possibly separating data from different classes, and false neighborhood within classes possibly gathering data assigned to the same class.

In that purpose, the weighing function F depends on either the HD or the map distances depending of the class co-membership A_{ij} of the pair *(i,j)* of data points considered (3). The ClassiMap stress function is then defined as:

$$S_{ClassiMap}(i,j) = \left|d_{ij} - d_{ij}'\right| \times \left[A_{ij} \times F(d_{ij}) + (1 - A_{ij}) \times F\left(d_{ij}'\right)\right] \quad (3)$$

where A is the class co-membership $n \times n$ matrix of the n data points. The value A_{ij} located in the i^th row and the j^th column quantifies the class similarity between data points *i* and *j*. We set $A_{ij} = 1$ if the data points *i* and *j* belong to the same class, and $A_{ij} = 0$ otherwise. Many functions can be used for F. In the following, we use the one proposed in [8] for its robustness to the concentration of measure phenomenon:

$$F(x) = 1 - \int_{-\infty}^{x} f(u,\mu,\theta)\,du \quad (4)$$

where

$$\mu = \underset{i,j}{mean}(d_{ij}) - 2(1-\gamma)\underset{i,j}{std}(d_{ij})$$

$$\theta = -2\gamma \underset{i,j}{std}(d_{ij})$$

and $f(u,\mu,\theta)$ is the Normal distribution with mean μ and standard deviation (std) θ. The parameter γ controls the influence of the neighborhood. It decreases linearly from 0.9 to 0.1 during the mapping optimization process.

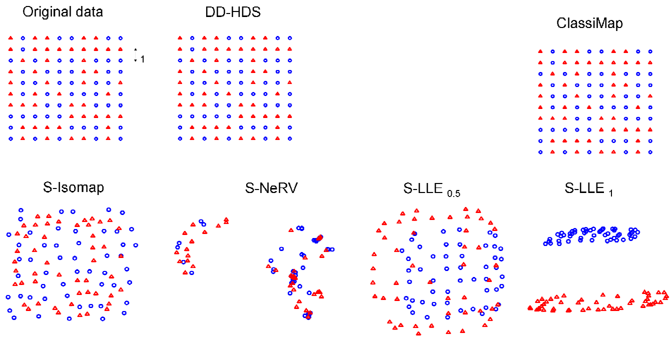

Figure 2: *Original 2-dimensional data (top left) with randomly assigned class labels (red triangles and blue circles) are mapped on a plane with unsupervised and supervised (name in a frame) mapping techniques. DD-HDS which preserve distances give exactly the same data distribution as the original one. Conversely, S-Isomap, S-NeRV and S-LLE, supervised mapping techniques which first modify original distances based on the class labels, tend to separate the classes unduly. In this case, the proposed ClassiMap supervised mapping technique gives the same correct map as distance-preserving unsupervised mapping techniques.*

2.2. ClassiMap optimization process

The minimization of $E_{ClassiMap}$ is achieved thanks to a force directed placement technique [9] similar to the one used in [8] where only changes the stress function. Indeed, a force is defined between each pair of points:

$$\Phi(i,j) = \frac{\left(d_{ij}' - d_{ij}\right)}{\left|d_{ij}' - d_{ij}\right|} \times S_{Classimap}(i,j) \times \vec{u}_{ij} \quad (5)$$

where $\vec{u}_{ij}$ is the unitary vector oriented from point *i* to point *j* on the map. The points image of the data on the map move due to these forces until they reach an equilibrium state, local minimum of the global stress function.

At first, a subset of the data points [8] is mapped. The value of the γ parameter at this step is set to 0.9 and positions are optimized. The result allows providing a rough map of the global structures. Then remaining data points are progressively considered and the γ parameter is decreased simultaneously towards its final value. This process is used [8] in an unsupervised setting. Such an iterative process allows progressive focusing on local structures while accounting for the global structures.

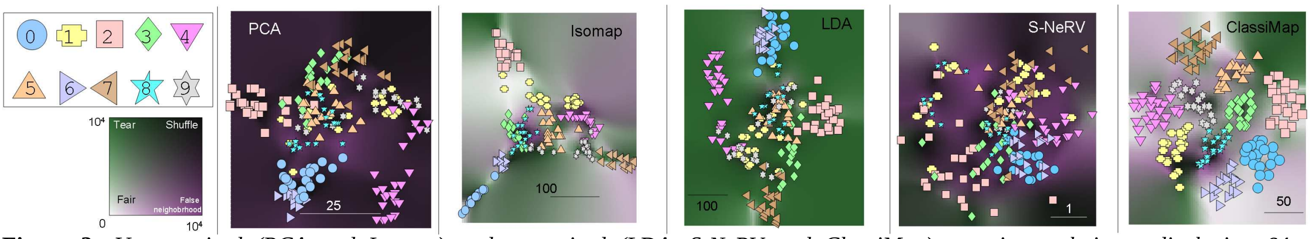

Figure 3: *Unsupervised (PCA and Isomap) and supervised (LDA, S-NeRV and ClassiMap) mapping techniques displaying 64-dimensional Digit data (pictures of handwritten digits 0 to 9) in a plane. ClassiMap emphasizes class separation while preserving classes' topology. Local distortion evaluation is provided by coloring the background using the CheckViz technique [1] (σ is set to 35); light colored background account for fairly mapped areas; green ones account for torn areas, purple ones account for false neighobrhoods and black ones account for areas where both types on distorition occur simultaneously (the colorcode is similar as in figure 1).*

4. Example with Digit data

ClassiMap as well as several other supervised and unsupervised mapping techniques has been tested on the popular optical recognition of handwritten digits dataset: a set of black and white pictures (8 × 8 pixels) of handwritten digits (numbers from "0" to "9") from the UCI machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/).

The resulting maps are shown in the figure 3. We can observe that, conversely to other methods, ClassiMap provides a map where classes can be easily distinguished. Moreover, using the CheckViz technique, we see that tears (green background) appear preferentially between classes and false neighborhoods (purple background) within classes. Finally, the ambiguous data points appear on the map in grey or black areas. For example, several "3" and "8" are very close (green diamonds and blue stars). Figure 4 shows some of the digits which are mapped very close while they are assigned to different classes. These points at the boundaries of the classes deserve special attention from the user as they are either the ones which define the classes' boundaries if the labels are taken for granted, or the ones more likely to be mislabeled otherwise.

4. Conclusion

As far as we know, ClassiMap is yet the only one supervised mapping technique that does not distort the original distances. It is seems able to untangle classes while fairly preserving the main HD structures of the classes. Such features are very suitable for decision support systems because it shows proximity relationships between data on a more readable map.

Acknowledgment

The authors would like to thank Bjørn Sand Jensen (Technical University of Denmark) for kindly providing codes (and support) to calculate the supervised distances used in S-NeRV and D. de Ridder (Delft University of Technology) for his codes for Supervised LLE.

References

[1] Lespinats S. and Aupetit M. "CheckViz: Sanity Check and Topological Clues for Linear and Non-Linear Mappings", Computer Graphics Forum, 30:113-125, 2011.
[2] de Ridder, D., Kouropteva, O., Okun., O., 2003. Supervised locally linear embedding - Lecture Notes in Artificial Intelligence, 2714:333–341.
[3] Weng, S., Zhang C., Lin, Z., 2005. Exploring the structure of supervised data by discriminant isometric mapping, Pattern Recognition 38, 599–601.
[4] Venna, J., Peltonen, J., Nybo, K., Aidos, H., Kaski S., 2010. Information retrieval perspective to nonlinear dimensionality reduction for data visualization The Journal of Machine Learning Research 11, 451−490.
[5] Mercier, D., Gaillard, P., Aupetit, M., Maillard, C., Quach, R., Muller, J.-D. 2006 How to help seismic analysts to verify the French seismic bulletin? Engineering Applications of Artificial Intelligence, Volume 19, Issue 7, Pages 797-806, Elsevier.
[6] Fisher R.A., 1936. The Use of Multiple Measurements in Taxonomic Problems, Annals of Eugenics, n° 7, p. 179-188.
[7] Fukunaga. K., Introduction to Statistical Pattern Recognition, second ed. Academic Press, 1990.
[8] Lespinats S., Verleysen M., Giron A., Fertil B., 2007. DD-HDS: a tool for visualization and exploration of highdimensional data, IEEE Trans. Neural Netw., vol. 18, no. 5, pp. 1265-1279.
[9] Eades, P., 1984 A heuristic for graph drawing. in Proc. 13th Manitoba Conf. Numer. Math. Comput. (Congressus Numerantium), D. S. Meek and G. H. J. V. Rees, Eds., Winnipeg, MB, Canada, vol. 42, pp. 149–160.
[10] Bishop, C.M., 2011 Pattern Recognition and Machine Learning. 740 p. 2nd edition, Springer.
[11] Aupetit, M., Catz, T., 2005 High-dimensional labeled data analysis with topology representing graphs. Neurocomputing 63, 139-169, Elsevier.

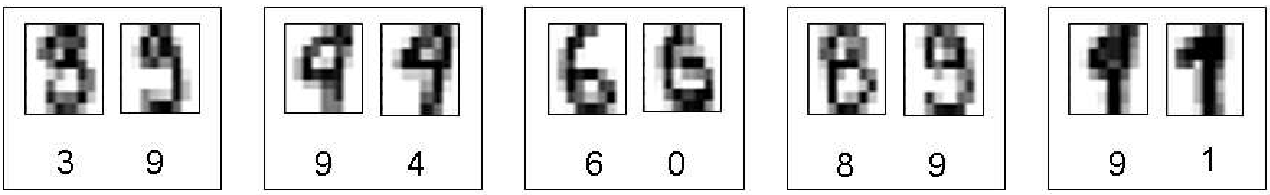

Figure 4: *Several examples of digit data with ambigous labels. Each picture is the digit as drawn by the writer (and corresponds to a point in the figure 3). The above examples arepairs of points lying between different classes on the map provided by classiMap. As we can observe, many of these examples are ambigous for the reader as well.*