Superparamagnetic Maps: Self-organising Low-Dimensional Embedding with Superparamagnetic Clustering

Thomas Ott[1], Robert Rohrkemper[1], and Markus Christen[2]

[1] ZHAW Zurich University of Applied Sciences, Switzerland
{ottt,roro}@zhaw.ch
[2] Institute of Biomedical Ethics, University of Zurich, Switzerland
christen@ethik.uzh.ch

Abstract. Visualisation of high-dimensional data by means of a low-dimensional embedding is an important technique in data analysis. We present a novel approach to this problem based on a heuristic using superparamagnetic clustering. Our method can deal with nonlinear structures since it is essentially local. Moreover, the nonparametric characteristics and the robustness of the superparamagnetic self-organisation approach allow discrimination between different clusters as well as between clusters and background noise. Hence, cluster structures can be boosted in the low-dimensional representation. The advantages provided by these capabilities are illustrated by three applications.

Keywords: Low-dimensional Embedding, Multidimensional Scaling, Superparamagnetic Clustering, Self-organisation

1 Introduction

Visualisation of data plays a key role in explorative data analysis since it makes data structures easily accessible to vision, our most powerful sense. When dealing with complex data sets, visualisation requires dimensionality reduction to display meaningful structures of high-dimensional data in a low-dimensional space. The classical approaches to dimensionality reduction, principal component analysis PCA and multidimensional scaling MDS, aim to represent the data structure on a linear subspace of the original data space. While PCA performs a projection onto the most important axes of a subspace, i.e. the axes with maximal data variance, the goal of MDS is to find a low-dimensional embedding that preserves the interpoint distances. Although the essentially linear nature of these methods is theoretically advantageous, the methods often perform poorly or can even fail when applied to nonlinear data structures. Furthermore, for many real-world applications, data vectors are not available. Instead, researchers are faced with similarity or proximity data, in which the ordering is correct, but the exact numbers are unreliable [1].

Various approaches exist to overcome these problems. The latter problem is addressed by non-metric MDS [5]. This approach aims to prevent the ordering of

the original proximities in the low-dimensional embedding, while the magnitudes of the proximities can be rescaled by a monotonic transformation. In order to overcome the problems of nonlinearites, specialised nonlinear methods such as the Isomap algorithm [10] have been invented.
Isomap requires constructing a k nearest neighbour graph to represent the structure of a data manifold. This enables a more correct description of proximities between points of a folded lower-dimensional manifold embedded into a higher-dimensional space. Along with the publication of Isomap there has been some debate about the distraction by noise, highlighting the importance to address the noise issue.

In this contribution, we present a novel approach that is able to deal with nonlinear structures in data space and that includes a mechanism to reduce the distraction by noise. Our self self-organisation-based heuristic bears some similarity with self-organising maps [4], but is based on superparamagnetic clustering [2, 7, 8], a nonparametric clustering approach based on local spin interactions. It takes advantage of a graph-based approach akin to Isomap. Moreover, it incorporates the non-metric MDS idea of applying a transformation to the proximities. With reference to these origins, we will call our method *(self-organising) superparamagnetic maps*.

2 Self-organising Superparamagnetic Maps

To explain our idea, we recall the underlying goal of (non-metric) MDS. Given a $n \times n$ symmetric matrix of pairwise proximities g_{ij} (with $g_{ii} = 0$), the goal is to find a low-dimensional embedded representation of n points such that a cost (or stress) function is minimised.

$$s = \sum_{(i,j)} (d_{ij} - f(g_{ij}))^2 \tag{1}$$

is a possible simple choice (d_{ij} indicates the Euclidian distance in the $p-$dimensional embedding space). The monotonic function f that is applied to the original 'distances' g_{ij} does not change the order of the values and allows for rescaling. Non-metric MDS is implemented as a two-step procedure: Starting with an arbitrary point configuration and corresponding distances d_{ij}, iteratively, the configuration is optimised for fixed f, and f is optimised for a fixed configuration. A gradient descent method can be used to optimise the point configuration with respect to s (and for fixed f). Gradient descent methods implement the idea that two points are moved towards each other on their connecting line if s can be minimised by decreasing the distance d_{ij}, or the points are moved apart on their connecting line if s can be minimised by increasing d_{ij}, respectively.

We will adopt this idea of moving points along their connecting lines, although – in contrast to non-metric MDS – we will not optimise f. Instead, we will integrate information about the local data structure via a transformation similar to f. For this, we use superparamagnetic clustering (SC) that is based

on a principle of self-organization.
SC translates original proximities g_{ij} into a weighted graph with n nodes, each holding a spin variable. The edges represent coupling strengths: the larger d_{ij} is, the stronger the corresponding coupling J_{ij} between node i and node j is. The couplings are restricted to k nearest neighbour interactions. The resulting inhomogeneous spin system is treated in the formalism of the canonical ensemble.
Depending on the system temperature T, groups of aligned spins can emerge that can be identified as data clusters. T gives rise to a clustering hierarchy: For small values of T, the system is in the ferromagnetic phase, where spins are likely to be aligned. In an intermediary T-range, a superparamagnetic phase occurs: strongly coupled spins tend to be aligned, whereas weakly coupled spins behave independently. A further increase of T generally leads to a continued breaking up of these clusters into smaller clusters. For high T, the system enters the paramagnetic phase where any order disappears and only singleton clusters remain.
Clusters are identified by means of the pair correlation, G_{ij}, which provides us with a robust measure of the likelihood of two points belonging to the same cluster structure or to the noisy background [8]. Instead of choosing a single temperature T, we calculate a temperature average G_{ij}^T along a pre-defined temperature interval, yielding a well-balanced likelihood (see below).

In order to generate a low-dimensional representation of the data, we use G_{ij}^T instead of g_{ij} as input similarities. The map $g_{ij} \rightarrow G_{ij}^T$ can be interpreted as a nonlinear transformation, f, that boosts cluster structures in the original data. In practice, the pair correlation G_{ij}^T is calculated using a Markov chain Monte Carlo MCMC procedure. This can be understood as a discrete time simulation of the spin system, where the correlation of two spins is estimated as a time average and the temperature average is incorporated by a slow increase of T during this process.
The information about G_{ij}^T is transferred to the low-dimensional data representation by connecting the time simulation with a data self-organisation principle: starting from an arbitrary point configuration, two points are moved towards each other if the corresponding spins are coupled and aligned, otherwise they are moved apart. Over the course of time, we can observe a "steady state distance" d_{ij} between point i and j which reflects G_{ij}^T and hence the cluster structure of the original data (since G_{ij}^T must be viewed as a similarity measure, larger values lead to smaller distances d_{ij}).

In the following sections, we will first describe SC briefly (details can be found in [2, 8]), and then give a detailed description of our algorithm.

2.1 Superparamagnetic Clustering

For data clustering with SC, we map a data set onto a spin system as follows: Each data item is represented by a Potts spin variable s_i with possible values

in $\{1, ..., q = 10\}$[3]. Each spin is coupled to its k nearest neighbours[4], where for given proximity values $g_{ij} = g_{ji}$, the couplings between spins are determined according to

$$J_{ij} = J_{ji} = \frac{1}{k} \exp\left(\frac{-g_{ij}^2}{2a^2}\right) \tag{2}$$

a is the average distance between neighbours. Each spin configuration s is associated via the Boltzmann distribution with the probability

$$p(s) = \frac{1}{Z} \exp(-H(s)/T) \tag{3}$$

with the Hamiltonian $H(s) = \sum_{(i,j)} J_{ij}(1-\delta_{s_i s_j})$ and the normalization constant Z . The parameter T represents the system temperature. At a given T, the pair correlation $G_{ij} = \sum_s p(s)\delta_{s_i s_j}$ is calculated. SC uses G_{ij} to identify the clusters (we are mainly interested in G_{ij} itself). G_{ij} is approximately calculated by

$$G_{ij} = \frac{1}{M} \sum_{t=1}^{M} \underbrace{\delta_{s_i^t s_j^t}}_{G_{ij}(t)} \tag{4}$$

where the Swendsen-Wang algorithm [9] has been used to generate the series of states.
In summary, SC is nonparametric clustering approach for which neither the number of clusters nor the shape has to be predefined. Clusters are formed in a robust self-organising process, which allows us to distinguish them from a noisy background or halo [2, 8].

2.2 The Algorithm of Superparamagnetic Maps

A $\mathbb{R}^2-$embedding of a matrix of proximities g_{ij} is constructed by setting up the superparamagnetic clustering framework and performing the following steps:

1. Choose a random point configuation $(\boldsymbol{x_1^0}, ..., \boldsymbol{x_n^0})$ with $\boldsymbol{x_i^0} \in \mathbb{R}^2$
2. Choose a random spin configuration s^0
3. Set the temperature $T = T_{min}$ and ΔT (see below)
4. For T, calculate a new spin configuration s^{t+1} (according to Swendsen-Wang)
5. Calulate the actual pair correlations $G_{ij}(t+1) = \delta_{s_i^{t+1} s_j^{t+1}}$
6. For each pair of points, do:
   - If $G_{ij}(t+1) = 1$ and $J_{ij} > 0$ then

$$\boldsymbol{x_i^{t+1}} = \boldsymbol{x_i^t} + \alpha \cdot (\boldsymbol{x_j^t} - \boldsymbol{x_i^t}) \tag{5}$$

$$\boldsymbol{x_j^{t+1}} = \boldsymbol{x_j^t} + \alpha \cdot (\boldsymbol{x_i^t} - \boldsymbol{x_j^t}) \tag{6}$$

[3] q can be chosen nearly arbitrarily and is not related to the number of clusters [2].
[4] Again, the choice of k is largely uncritical. We used $k = 10$ as default value.

– else

$$\boldsymbol{x_i^{t+1}} = \boldsymbol{x_i^t} + \beta \cdot e^{-d_{ij}^t} \dot{(}\boldsymbol{x_i^t} - \boldsymbol{x_j^t}) \tag{7}$$

$$\boldsymbol{x_j^{t+1}} = \boldsymbol{x_j^t} + \beta \cdot e^{-d_{ij}^t} \dot{(}\boldsymbol{x_j^t} - \boldsymbol{x_i^t}) \tag{8}$$

where $d_{ij}^t = |\boldsymbol{x_i^t} - \boldsymbol{x_j^t}|$.

7. Set $T = T + \Delta T$ and go back to 4 as long as $T < T_{max}$

The choice of parameters is as follows:

* For the temperature range $[T_{min}, T_{max}]$ the optimal choice is the superparamagnetic phase since it provides information about the cluster structures. This range can differ for each data set, but the differences are usually small. The range could be estimated in advance by a mean-field calculation (similar to the one presented in [8]). For simplicity, we chose a fixed range of $[0, 0.1]$, which is – according to our experience – a reasonable first approximation.
* ΔT is related to the number of Monte Carlo steps M: $\Delta T = [T_{min}, T_{max}]/M$. According to our experience, $M = 150$ gives stable results.
* $0 < \alpha < 0.5$ controls the attraction (speed) of two points whose spins are correlated.
* $0 < \beta$ controls the repulsion (speed) of two points whose spins are uncorrelated.
* Superparamagnetic maps do not offer unique solutions, which highlights the importance of the parameters involved. Experiments show that α and β strongly determine the scaling of the final point configuration. α mainly affects the intra-cluster distances and β mainly affects the inter-cluster distances. For all the experiments, we used the values $\alpha = 0.1$ and $\beta = 0.01$. This is a canonical choice that seems to balance inter-and intra-cluster distances.
* The additional factor $e^{-d_{ij}^t}$ makes sure that the point configuration remains bounded (other factors can be used as well).

3 Examples

3.1 The Noisy Ring Problem

In [6], a benchmark data set was introduced, showing two interlocked rings on a noisy background (750 points in total, 250 points for each ring and the background, see Fig.1 a)). The problem cannot be solved by most clustering algorithms [6]. In contrast, our approach is capable of generating a 2D image since superparamagnetic clustering can solve the problem with ease (Fig.1 b)). The scaling in this image, however, does not directly reflect the scaling in the original data. The loop in one of the rings is interesting and is a consequence of the dimensionality reduction.

Note that the approach includes the possibility of inherent noise cleaning. In Fig.1 c) we took the results from superparamagnetic clustering into account, which allows discrimination between clusters and background noise. Hence the background noise can be subtracted from the image at the end.

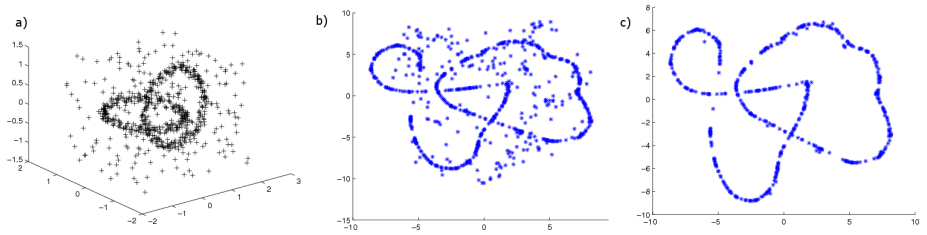

Fig. 1. a) Original data set with two rings on a noisy background b) Superparamagnetic map solution c) Superparamagnetic maps offer inherent noise cleaning

3.2 The PCA-Mixer Problem

This data set consists of two Gaussian clusters with 200 points each and means $\mu_1 = (0, 0, 0)$ and $\mu_2 = (0, 0, 2)$ (Fig 2 a). The standard deviations are $\sigma_1^x = \sigma_2^x = 4.5$, $\sigma_1^y = \sigma_2^y = 1.5$ and $\sigma_1^z = \sigma_2^z = 0.05$. While the two clusters can clearly be distinguished in 3D, they are invisible to PCA in 2D because the extension in the $x-$ and $y-$ direction is larger than in $z-$direction (Fig 2 b). For SM, this is no problem (Fig 2 c).

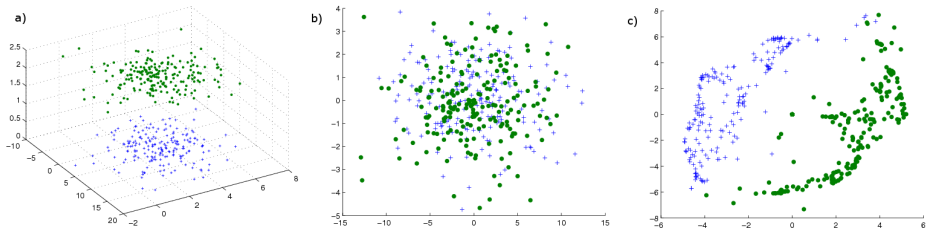

Fig. 2. a) Original data set with two clusters b) PCA solution c) Superparamagnetic map

3.3 A Political Landscape

In our last example we use data from the "smartvote" project[5], an internet-based tool that allows citizens in Switzerland to compare their own political opinions with the opinions of candidates of national elections. As part of the project, all political candidates in running for the last Swiss general elections were invited to take part in a survey of about 70 questions, designed to elicit the candidates political position on a broad range of issues (see [3] for details). We applied SM to the candidates of the five largest Swiss parties from the canton of Berne: SVP

[5] See http://www.smartvote.ch/index.php

(Swiss People's Party), SPS (Social-democrats), FDP (Free Democratic Party), CVP (Christian-democrats) and the GPS (Green Party). The resulting political landscape is shown in Fig. 3, where the Manhattan distance has been used to calculate the input g_{ij}. The resulting map corresponds to the known structure of the Swiss political landscape, although no external classification criteria along 'right' or 'left' politics had to be applied in the analysis.

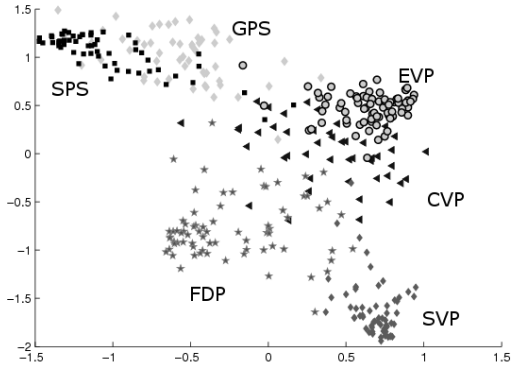

Fig. 3. The political landscape of the canton of Berne including the 5 largest Swiss parties (see text).

4 Discussion

We have introduced a novel algorithm for finding low-dimensional embedded representations of a set of proximities g_{ij}, called self-organising superparamagentic maps (SPM). The algorithm is based on a heuristic using superparamagnetic clustering. The main idea is that clustering provides the possibility to incorporate crucial information about cluster structures in the original data. Using this information, our superparamagentic maps generate a low-dimensional image of the data. This approach has two main advantages. First, it is capable of bringing out nonlinear structures that are invisible to classical techniques such as PCA. Second, due to its robust nonparametric characteristics, it is able to distinguish between clusters and background noise. On the downside, the procedure is more time-consuming than other methods since it involves a spin system simulation. However, applying SPM also yields a clustering of the original data without an extra effort, providing an additional benefit.

Although the heuristic algorithm was successful in several applications, questions remain, regarding the theoretical understanding of superparamagnetic maps: How can we quantify the role of the parameters α and β? How can the theoretical connections to other methods such as (nonmetric) multidimensional scaling be elaborated? Which other rules or clustering methods could be used to generate

the low-dimensional representation? Can we also use the technique to find out the true dimensionality of higher-dimensional data structures? Answers would clarify why the heuristic works so well.

Acknowledgments. We thank Daniel Schwarz (Institute of Political Science of the University of Berne) for providing us with data of the Smartvote project.

References

1. Agarwal, S., Wills, J., Cayton, L., Lanckriet, G., Kriegman, D., Belongie, S.: Generalized Non-metric Multidimensional Scaling. AISTATS, San Juan, Puerto Rico (2007)
2. Blatt, M., Wiseman, S., Domany, E.: Superparamagnetic Clustering of Data. Phys.Rev.Lett. 76, 3251–3254 (1996)
3. Christen, M., Ott, T., Schwarz, D.: When do incoherent parties split? A new measure of party coherence based on the collective behavior of magnetic parts in physics (submitted) (2011)
4. Kohonen,T.: Self-organizing Maps. Springer, New York (1997)
5. Kruskal, J.B..: Nonmetric Multidimensional Scaling: A Numerical Method. Psychometrika 29, 115-129 (1964)
6. Landis, F., Ott, T., Stoop, R.: Hebbian Self.organizing Integrate-and-Fire Networks for Data Clustering. Neural Computation 22, 273-288 (2010)
7. Ott, T.,Kern, A., Schuffenhauer, A., Popov, M., Acklin, P., Jacoby, E., Stoop, R.: Sequential Superparamagnetic Clustering for Unbiased Classification of High-dimensional Chemical Data. Journal of Chemical Information and Computer Sciences 44(4), 1358-1364 (2004)
8. Ott, T.,Kern, A., Steeb, W.-H., Stoop, R.: Sequential Clustering: Tracking Down the Most Natural Clusters. Journal of Statistical Mechanics: theory and experiment, P11014 (2005)
9. Swendsen, R.H., Wang, S.: Non-universal Critical Dynamics in Monte Carlo Simulations. Phys.Rev.Lett 58, 586-88 (1987)
10. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290, 2319-2323 (2000)