Resolution Enhancement for the ViSOM

Hujun Yin
University of Manchester Institute of Science of Technology (UMIST)
Dept. of Electrical Engineering and Electronics
Manchester, M60 1QD, UK
h.yin@umist.ac.uk

Keywords: Data visualisation, Nonlinear mapping, ViSOM, Principal curves/surfaces

Abstract—The topology perserving self-organising maps (SOM) has become a powerful and useful tool for visualising high dimensional data. However it requires a colouring scheme to imprint the inter-neuron distances on the map. The recently proposed Visualisation induced SOM (ViSOM) is able to directly preserve the distance information, as well as topology, on the map. It has been proved to be a discrete principal curve/surface. The resolution of the map is inversely proportional to the size of the grid. Large maps require more computation thus long training time. This paper proposes two simple methods for enhancing the resolution of the ViSOM so avoiding using large maps. The first interpolates a trained map, while the second incorporates local linear projections into the projection step when mapping the data on a trained map. Experiments and results are given to demonstrate the advantages.

1 Introduction

Data projection and visualisation has become a major application area for neural networks, in particular the self-organising map (SOM) [5], as its topology preserving property is unique among other neural models. Good projection and visualisation methods help to identify clustering tendency, to reveal the underlying functions and patterns, and to facilitate decision support. A great deal of research has been devoted to this subject and a number of methods have been proposed. A recent review on this subject has been given in [16].

Classic linear projection methods include the subspace projection and principal component analysis (PCA). The subspace projection reveals data relationships and distribution for every pair of variables, which are either original or transformed, in attempt to find a meaningful subspace. PCA projects the data onto its linear principal directions, which are represented by the orthogonal eigenvectors of the covariance matrix of the data set. It de-correlates the data variables and finds the dominating features in the data set. It is the optimal linear projection in the sense of minimum mean-square-error between the original data points and the projected ones on the principal subspace. But the PCA’s linearity has limited its power for practical datasets, as it cannot capture nonlinear relationships defined by higher than second order statistics. If the input dimensionality is much higher than two, the projection onto a linear plane may provide limited visualisation power. Extending the PCA to nonlinear domain, in principle, could model the practical data set better. Various methods have indeed been proposed, e.g. the principal curve and surface [1, 7], the auto-associative network [6], the generalised PCA [3] and the kernel PCA [10].

The principal curves/surfaces are the primary nonlinear extension of the PCA. They are defined as smooth and self-consistent curves or surfaces, which pass through the middle of the data set and do not intersect themselves. There are two steps in the definition of the principal curves/surfaces: the projection step and expectation (smoothing) step. The HS algorithm directly iterates these two steps [1]. In the projection step a data point is mapped to the closest point on the curve/surface. While in the smoothing step, the curve/surface at a particular index is defined as the mean (or kernel smoothing) of all the data points mapped to that index. The HS algorithm is a batch operation.

Multidimensional scaling (MDS) is nonlinear projection and projects data points onto a two-dimensional plane or plot by preserving as close as possible the inter-point metrics. The objective is thus to minimise the difference of the (usually weighted) inter-point distances between the original space and mapped plane. The MDS relies on an optimisation method to minimise the defined objective function. Such a mapping can capture the nonlinear relationship of the data and reveal the overall structure of the data. Sammon [9] mapping is a well-known example of the MDS. It uses an intermediate weighting scheme so maintaining a balance of good local distributions and a clear view of the global data structure. However, like other MDS methods, the Sammon algorithm is a point-to-point mapping, which does not provide the explicit mapping function and cannot naturally
accommodate new data points (unless triangulation is used and new data points are far fewer than original data set). Sammon mapping is efficient and effective when the data set is small. It may become impractical for many applications where data arrives sequentially or the quantity of the data is huge.

The self-organising map (SOM) has been widely used as a visualisation tool for dimensionality reduction [e.g. 2,4,5,11]. The SOM’s unique topology preserving property can be used to visualise the relative mutual relationships among the data. However, the SOM does not directly apply to scaling, which aims to reproduce proximity in (Euclidean) distance on a low visualisation space, as it has to rely on a colouring scheme (e.g. the U-matrix method [11]) to imprint the distances crudely on the map. Often the distributions of the data points are distorted on the map.

The recently proposed visualisation induced SOM (ViSOM) [13, 14] constrains the lateral contraction force between the neurons in the SOM and hence regularises the inter-neuron distances with respect to a scaleable parameter that defines and controls the resolution of the map. It preserves the data structure as well as the topology as faithfully as possible. The ViSOM provides a direct visualisation of both the structure and distribution of the data. For a fine visualisation, a large ViSOM is usually required – this can be computational demanding. In this paper, two simple method have been proposed to overcome this limit so that a map can be enhanced after the training so having the equal resolution and effect as large maps.

2 SOM, ViSOM and Manifold

The SOM uses either a rectangular or hexagonal grid of nodes and adopts a neighbourhood learning mechanism in an unsupervised learning of the data space. As a result, the grid or map possesses topological ordering properties as similar data points are mapped closely on the map. Thus the map can be used to show the relative relationships among data points. However, the SOM does not directly show the inter-neuron distances on the map. When the SOM is used for visualisation, some colouring scheme is needed to imprint the inter-neuron distances so that the clusters and boundaries can be marked. The colour or grey tone of a node or a region between nodes is proportional to the mean or median of the distances between that node and its nearest neighbours. Such a colouring method has been used in many data visualisation applications, e.g. WEBSOM [2] and World Welfare Map [4]. The colouring methods indeed enhance the visualisation ability of the SOM. However, the cluster structures and distribution of the data shown on the map often are not apparent and appear in distorted and unnatural forms [13, 14].

The SOM has been linked with the principal curve/surface algorithm [8]. However differences remain not only in the projection process but also in the smoothing step. In the SOM, data are projected to the nodes rather than onto the curve/surface. This is the discrete nature of the SOM and this difference is minor when the number of nodes is large. The key difference, however, is in the smoothing step [15]. The neighbourhood smoothing in the SOM is governed by the indexes of the neurons, while in the principal curves/surfaces the smoothing is performed entirely in the data space. As the grid indexes are uniform integers (coordinates), the smoothing effect from uniform neighbourhood indexes does not reflect that from the neighbouring data points, which can be highly non-uniformly distributed.

The SOM can serve as a visualisation map to show the relative closeness and relationships among data and clusters. In many cases, a direct and faithful display of structure shapes and distributions of the data is more desirable in visualisation. The ViSOM has been proposed to directly preserve distances on the map [13-15]. For the map to capture the data structure naturally and directly, the distance quantity must be preserved on the map, along with the topology. The nodes should be approximately uniformly and smoothly placed in the nonlinear manifold of the data space. To achieve this in the ViSOM, the lateral contraction force presented in SOM’s updating force remain not only in the projection process but also in the smoothing step. In the SOM, data are projected to the curves/surfaces the smoothing is performed entirely in the data space. As the grid indexes are uniform integers (coordinates), the smoothing effect from uniform neighbourhood indexes does not reflect that from the neighbouring data points, which can be highly non-uniformly distributed.

\[ \Delta w_i = \alpha \eta (v,c) [(x-w_i)+\beta (w_i-w_c)] \] (1)

where \( \beta \) is the constraint. The simplest constraint is \( \beta = d_w/(\Delta w, \lambda) - 1 \), with \( d_w \) the distance of the winning node \( v \) and its neighbouring node \( c \) in the input space, \( \Delta w \) the corresponding distance on the map, and \( \lambda \) a resolution constant. More complicated constrain has been proposed in [15], together with details regarding the choices of resolution parameters.

The ViSOM regularises the contraction force so that the distances between the nodes on the map are analogous to the distances of their weights in the data space. The aim is to adjust inter-neuron distances on the map in proportion to those in the data space, i.e. \( \Delta w_i = d_w \) within the neighbourhood. When the data points are eventually projected on a trained map, the distance between data point \( i \) and \( j \) on the map is
proportional to that of the original space, subject to the quantisation error (the distance between a data point and its neural representative). When the map is trained and data points mapped, the distances between mapped data points on the map will resemble approximately those in the original space (subject to the resolution of the map). This makes visualisation more direct, quantitatively measurable, and visually appealing.

In the ViSOM, the smoothing effect is similar to that in the principal curves/surfaces [15] and is regarded as the discrete principal curve/surface. The scaling in the ViSOM is approximately uniform, so that is appealing to the multivariate visualisation problems. While the scaling in SOM is non-uniform but density related.

3 Resolution Enhancement

The resolution parameter controls the accuracy of the visualisation and eventually the size of the map. Large maps require longer training time, and this can be costly in some cases. However as the nodes in the ViSOM are uniformly distributed in the data space, some simple post-processing can be introduced to enhance the resolution of an original not so large map.

3.1 Interpolation

As the nodes in the ViSOM are evenly placed on the manifold of the data space. The use of uniform interpolation of new rows and columns of nodes between every two nearest rows or columns on the grid is justified and the new map will retain the equal distance constrain between adjacent nodes. If a $N \times N$ map is trained and $k$ new rows and columns of nodes are inserting between two adjacent rows or columns, the new map is of $(N+k(N-1))^2$ size. Interpolation can also be achieved by re-training a larger map based on a trained map as in the iSOM [12]. The advantage in this case is that the new map can have an arbitrary number of nodes.

Here we use the linear interpolation method as it is simple and quick to implement. For directly inserting $k$ new nodes linearly between two closest nodes on the grid, their weights can be calculated from the weights of these two nodes, i.e.,

$$w_{ci} = w_c + \frac{i}{k+1}(w_c - w_{c'})$$

where $w_c$ and $w_{c'}$ are two reference vectors and are the weight vectors of the two adjacent nodes $c$ and $c'$ on the original map and $w_{ci}$ are the newly inserted nodes between them (on the original rows or columns). For the new nodes inserted on the newly inserted rows and columns (but not on the original rows and columns), two reference vectors are the inserted nodes on either original rows or columns.

3.2 Local Linear Projection

When a map is trained, the data points are then projected on the map by placing them on to the best matching units (i.e. the winner nodes). When the map is small, large quantisation error will be introduced in this projection step. Here we propose to add a further local projection in this step, so that the data points are mapped on the manifold (the piece-wise hyperplanes) represented by the map grid. The projection index is then,

$$r(x) = \arg\max_c [(x - w_c)] + LLP(x, v)$$

where the first term refers to the index of the winning node, i.e. $v$, and the second term is the local linear projection, i.e. the offset, defined as,

$$LLP_{x,v} = \max_{c'=v,x} \left\{ \frac{(x - w_{c'}) \cdot (w_v - w_{c'})}{\| w_v - w_{c'} \|^2} \right\}, 0$$

where the dot product, `$\cdot$` denotes the dot product, and `$\| . \|$` is the norm.

Each non-boundary node has four branches or edges formed by this node and its four nearest neighbouring nodes. Boundary nodes have fewer edges. The local projection treats the edges as orthogonal and projects a data point, which is closest to the node, onto the span by the two closest neighbouring edges, i.e. the projections of the data point on these edges are the (positive) largest. This process is shown in Fig. 1.

![Figure 1: Local linear projection. $x$ is the data point and node $v$ is the winner. $x'$ is its projection on the map.](image-url)
4 Results and Conclusions

A 20×20 and a 100×100 ViSOMs have been trained on the Iris data set. The usual visualisation results of both are shown in Fig. 2. As can be seen that the resolution of the smaller map is much lower than the larger one as many points fall onto the same nodes. Fig. 3 shows the resolution enhancements by the two proposed methods. Fig. 3(a) is the result of interpolating four rows and columns of new nodes between two adjacent rows and columns, making the new map of size 96×96. Fig. 3(b) is the result of adding local linear projection. Both enhanced maps resemble the original 100×100 map, but with a significant reduction, about 25 times, in computational cost.

For large data sets, fine maps are often required. Otherwise the projected data points on the map will have the grid effect as shown in Fig. 4(a), in which a 40×40 ViSOM was trained to visualise 3924 4-dimensional tuberculosis gene expression profiles. Training a larger map, say 400×400 would take at least 100 times longer time. Instead, the 40×40 trained map was enhanced by the proposed local projection method and the result is shown in Fig. 4(b). The objective of this microarray data analysis was to identify and visualise if any new clusters have emerged after the infection. The result shown clearly demonstrates that no significant new cluster is formed.

The proposed resolution enhancement methods can be also applied to the SOM easily to provide an enhance visualisation and to reduce the training time.

If the local linear projection method is also incorporated in the ViSOM’s projection step, the ViSOM becomes equivalent to the HS principal curve/surface algorithm, in which the projection index is also defined as the arc length on the piece-wise linear lines or hyperplanes.
References


Figure 4: Visualisation of 3924 4-dimensional gene profiles, (a) 40×40 rectangular ViSOM, (b) enhanced map using the local linear projection method.