ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 15-20

Topological Interpolation in SOM by Affine Transformations

Josef Göppert, Wolfgang Rosenstiel
University of Tübingen, Sand 13, D-72076 Tübingen

Abstract. The calculation of virtual neurons in a self-organizing map by interpolation allows to reduce calculation time and memory space in training and retrieval. In this paper, a new interpolation method, based on affine coordinates of a local system is presented and discussed. An example in the domain of colour mixture is used to show the properties and compare this method to others.

1 Introduction

The self-organizing map (SOM) [Koh82] provides a general data approximation method which is suitable for several application domains such as robotics, evaluation of sensory data and visualisation of process states. Recent works [Spe94] have explored the training data distribution and have provided tools for the definition of the number of dimensions of the SOM architecture. It had been shown, that several data sets need higher-dimensional maps for topology preserving mapping.

But a serious problem occurs: The number of neurons and consequently the calculation time needed for finding the winner, as well as the memory for storing the weights increases exponentially. This effect are drastically reduced by applying local linearization and interpolation. In previous works [Göp93] different methods, based on the projection of error vectors onto the distance vector to further winners have been presented. Different applications have shown the aptitude of these methods and the properties compared to other neural network techniques. A method with a similar aim, but another strategy is the parametrized self-organizing map (PSOM) [Rit94].

In this paper we present an improved method for the interpolation in a SOM. This method is based on the affine coordinates of a local coordinate system. An application example shows the properties of this new method.

2 Set of winners

In the standard SOM algorithm only one neuron, the winner, is activated. All other neurons are inhibited. This principle is called "winner-takes-all" (WTA). It can be replaced by another one, called "winner-takes-most", which means, that winning is shared by a set of neurons. This leads to a smoother variation of the output by interpolation.

Two different strategies can be applied to find the set of winners:

1. Selection of the nearest neighbours in the input space: These are the neurons which have the smallest distance to the input vector.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 15-20

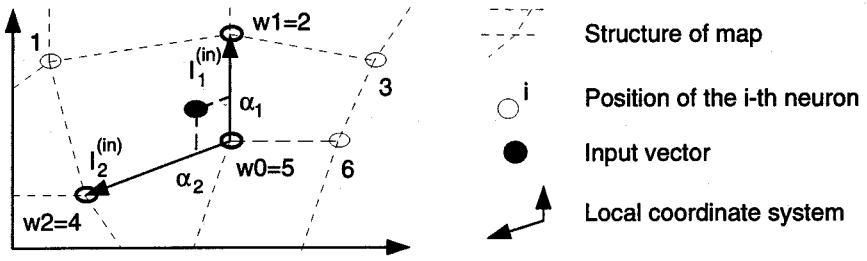

Figure 1: Linearisation by a local coordinate system.

2. Selection of the topological neighbours of the winning neuron: These are the neurons which are placed in adjacent grid positions to the winner. This has the advantage of the ordered structure of the self-organizing map can be used.

The choice of topological neighbours is an advantage, because the structure of the interpolation can be predefined. On the other hand topological defects lead also to big errors in the interpolation.

3 Projection and Interpolation

The first winner (index $w0$) and a list of k further winning neurons (index $wi \in w1 \ldots wk$) are found. The input vector $\mathbf{X}$ and the codebook vectors $\mathbf{W}_{wi}^{(in)}$ have n components. The grid position of the neuron is defined by $\mathbf{W}_{wi}^{(grid)}$. Further an output vector $\mathbf{W}_{wi}^{(out)}$ can be associated with these neurons.

The first approximation of the input vector is the codebook vector of the first winner $\mathbf{W}_{w0}^{(in)}$. From this base point the distance vector $\mathbf{l}_i^{(in)}$ to the codebook vectors of the further winners are used to form the directions of a k-dimensional local coordinate system $\mathbf{L}^{(in)}$ (see figure 1). $\mathbf{X}^l$ is the residual error in this system:

$$\mathbf{l}_i^{(in)} = \mathbf{W}_{wi}^{(in)} - \mathbf{W}_{w0}^{(in)} \qquad i = 1 \ldots k \qquad (1)$$

$$\mathbf{X}^l = \mathbf{X} - \mathbf{W}_{w0}^{(in)} \qquad (2)$$

$$\mathbf{L}^{(in)} = \left[\mathbf{l}_1^{(in)} \mathbf{l}_2^{(in)} \ldots \mathbf{l}_k^{(in)}\right] \qquad (3)$$

The local system in the grid ($\mathbf{L}^{(grid)}$) and the output space ($\mathbf{L}^{(out)}$) are calculated accordingly. The base directions of the coordinate system are supposed to be linearly independent, but not orthogonal. So the affine coordinates are calculated by a transformation matrix $\mathbf{T}$ (See appendix):

$$\alpha_i = \sum_{j=1}^{n} T_{ij} x_j^l \qquad i = 1 \ldots k \qquad (4)$$

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 15-20

$$\vec{\alpha} = \mathbf{T}\mathbf{X}^l \tag{5}$$

$$\mathbf{T} = \left(\mathbf{L}^{(in)T}\mathbf{L}^{(in)}\right)^{-1}\mathbf{L}^{(in)T} \qquad T: \textit{transpose matrix} \tag{6}$$

In general case the coordinates $\vec{\alpha}$ describe the best approximation of the input vector in the local system:

$$\tilde{\mathbf{X}} = \mathbf{W}_{w0}^{(in)} + \mathbf{L}^{(in)}\vec{\alpha} \tag{7}$$

The approximation $\tilde{\mathbf{X}}$ is considered to be the codebook vector of the virtual winner. The residual error vector is orthogonal to the local coordinate system, and it has therefore the smallest residual distance. The grid position of the virtual winner and the interpolated output vector $\mathbf{Y}$ is calculated accordingly:

$$\tilde{\mathbf{W}}^{(grid)} = \mathbf{W}_{w0}^{(grid)} + \mathbf{L}^{(grid)}\vec{\alpha} \tag{8}$$

$$\mathbf{Y} = \mathbf{W}_{w0}^{(out)} + \mathbf{L}^{(out)}\vec{\alpha} \tag{9}$$

Notice that the virtual grid position is only meaningful, if the k winners are topological neighbours of the first winner in the SOM grid.

4 Matrix inversion and dimensions

The matrix $\mathbf{L}^{(in)T}\mathbf{L}^{(in)}$ is square but may be singular if the rank is smaller than k. This is especially the case if the number of dimensions of the input space (n) is smaller than the number of dimensions of the local interpolation system (k equals the number of winners not counting the first one). But similar problems occur, if some axes ($\mathbf{l}_i^{(in)}$) of the local system are linearly dependent or zero.

Furthermore the linear equation may be ill-conditioned if the solution is sensitive to small changes in the data. With regard to the application different conditions can be identified:

- Supress zrero axes on the local system: Neurons should not be adapted to the same or similar position in the input space. This occures if the number of training patterns is too small. Therefore the number of training vectors has to be much bigger than the number of neurons.
- Suppress linear dependent axes: The number of inherent dimensions of the data set should not be smaller than the number of dimensions of the interpolation. In general it can be noticed that the dimensions of the map, the dimensions of the interpolation and the inherent dimensions of the data set should coincide as best as possible. For further informations see [Spe94].
- Define the structure of the interpolation: In a two-dimensional map the second and third winner can be the right and the left neighbour of the first winner (compare figure 1: Neuron 6 and 4). Both neurons represent

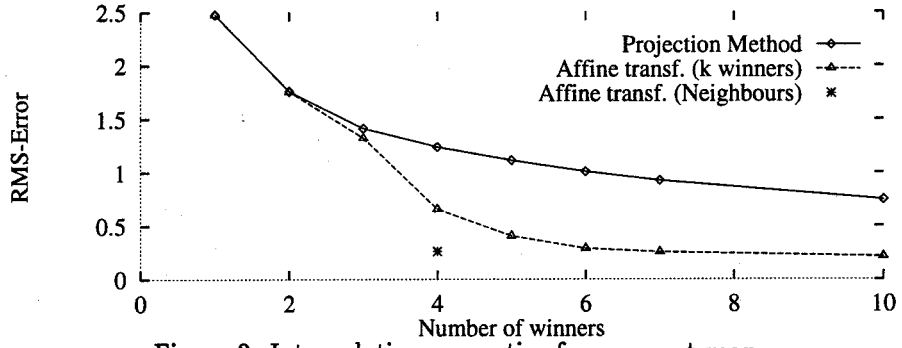

Figure 2: Interpolation properties for a pre-set map.

mainly one parameter change; other directions which are orthogonal to these directions get lost. It is supposed that the best solution is a topological choice of the winners. In this paper we are taking only the left or the right neighbour in each dimension (the one with the smallest distance).

5 Application Example

In order to show the properties of this method, we use a self-organizing map to predict the base colour concentrations of a colour mixture. The training data set was acquired by mixing the red, green and blue colour in the range of 0% ... 30% in steps of 2 %. The remaining volume to 100 % was filled up with white colour. This leads to $16 \times 16 \times 16$ (4096) mixtures. The spectra of the mixture were measured in the range of visible light by a spectrometer (16 components). The task of the self-organizing map is to associate the spectra with the base colour concentrations. Due to the properties of the data acquisition it can be stated, that the data set has 3 inherent dimensions. So the map is also choosen to be three-dimensional with $6 \times 6 \times 6$ neurons.

In this example the structure of the data set points out one possible organization. A good map can be achieved, if one dimension is representing the change of the red colour, the second dimension the change of green and the last dimension a change of the blue colour. Maybe this is not the best map for optimum approximation properties, but it represents a very good topology preservation. In order to eliminate the training effects of the SOM this knowledge can be used to bypass the training by pre-setting a map. All spectra of 0%, 6%, 12%, 18%, 24% and 30% are taken to pre-set the neuron codebooks ($\mathbf{W}^{(in)}$) in a topologically correct order. The same method is used to pre-set the output values ($\mathbf{W}^{(out)}$). This allows to concentrate mainly on the interpolation properties of a well-organized map.

Figure 2 shows a comparison of the affine transformation with the projection method [Göp93]. In both cases an increase in the number of winners leads to a decrease in the root-mean-square-error (RMS). The affine transformation behaves much better, especially for 4 to 10 winners. A topological choice of

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 15-20

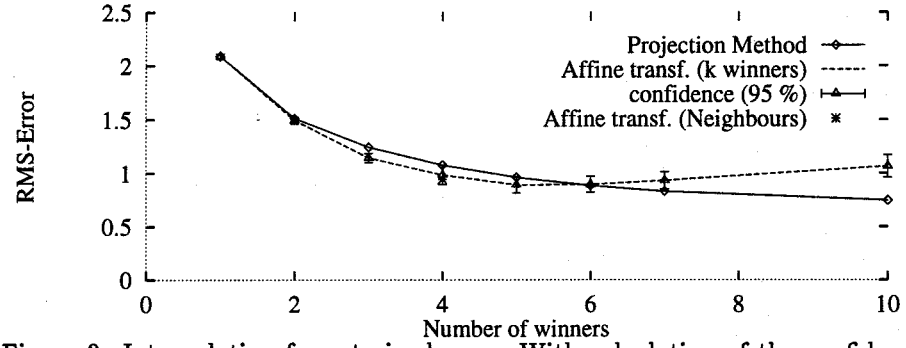

Figure 3: Interpolation for a trained map. With calculation of the confidence region (95%) for 50 trained maps.

the winners can be considered to be best. A very small error confirms that the method of local linearization is a suitable strategy for this data set. The difference between the affine transformation of the 4 best winners and the 4 topological winners can be explained by a better choice of the winners.

The next step is, to examine the behaviour with trained maps. The training was performed by combining input and output values to one training vector [Rit94]. In the recall, only the input components are used to calculate the distance and the interpolation. We have performed 50 independent training cycles. Afterwards, the mean of the RMS-error and the confidence region was calculated. Figure 3 shows that in this case the difference between projection and affine transformation is smaller. For 3 and 4 winners, the affine transformation is better than the projection at a significance level of more than 95%. An interesting aspect is also the fact that for higher-dimensional interpolation the error of the projection method decreases more and more, while it increases for the affine transformation. The reason is, that the projection is less sensitive to topological defects and high-dimensional affine transformation is sensitive to ill-conditioned matrices. Nevertheless, if the inherent topology of the data set, the topology of the map and the structure of the interpolation agree, the affine transformation is better.

In comparison with the pre-set map (figure 2) it can be noticed, that with up to 3 winners the training behaves better, but for more than 3 winners the well-defined topology of the map leads to better results. The reason of this is that the training of the SOM minimises the WTA-distance. So it seems to be promising to modify the training towards a better aptitude for interpolation.

6 Conclusion

A new method for the interpolation of self-organizing maps was presented. This method allows especially a reduction of the number of neurons, the memory size and the calculation time for finding the winner. An application example shows the properties of this method. It turns out, that this method needs well organized

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 15-20

maps. Standard SOM training minimises the distance to the winning neuron. Thus further research will deal with the improvement of the training, in order to support a minimisation of the interpolated error and topologically correct organisation. It has been shown, that in the case of a good organisation of the map, this method shows quite good results and allows it to reduce the residual error of real-valued output vectors of the SOM in a function approximation task.

Appendix

The aim is to transform the input coordinates $\mathbf{X}^l$ into the affine coordinates $\vec{\alpha}$ of the local system of $\mathbf{L}^{(in)}$ (Equ. 5). This transformation is characterised by the transformation of the base vectors:

$$\mathbf{I} = \mathbf{L}^{(in)}\mathbf{T} \qquad \mathbf{I}: \textit{ identity matrix of size } n \tag{10}$$

If $\mathbf{T}$ is a square matrix of rank k, the system can be solved by the inverse matrix $\mathbf{T} = \left(\mathbf{L}^{(in)}\right)^{-1}$. In general case the matrix $\mathbf{L}^{(in)}$ is not square, so the least-square solution is used. This leads to the normal equation of $\mathbf{I} = \mathbf{L}^{(in)}\mathbf{T}$ and the quasi-inverse of matrix $\mathbf{L}^{(in)}$:

$$||\mathbf{I} - \mathbf{L}^{(in)}\mathbf{T}||^2 = \min \tag{11}$$

$$\frac{\partial}{\partial T_{ij}}||\mathbf{I} - \mathbf{L}^{(in)}\mathbf{T}||^2 = 0 \tag{12}$$

$$2\mathbf{L}^{(in)T}\left(\mathbf{I} - \mathbf{L}^{(in)}\mathbf{T}\right) = 0 \qquad \textit{normal equation} \tag{13}$$

$$\mathbf{T} = \left(\mathbf{L}^{(in)T}\mathbf{L}^{(in)}\right)^{-1}\mathbf{L}^{(in)T} \tag{14}$$

References

[Göp93] J. Göppert and W. Rosenstiel. Topology-preserving interpolation in self-organizing maps. In *Proceedings of NeuroNimes 93*, pages 425–434, Nanterre, France, 10 1993. EC2.

[Koh82] T. Kohonen. Self-organized formation of topology correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.

[Rit94] Helge Ritter. Parametrized self-organizing-maps for vison learning tasks. In *Proc. of ICANN'94, Sorrento, Italy*, pages 803 – 810, London, 5 1994. Springer.

[Spe94] H. Speckmann, G. Raddatz, and W. Rosenstiel. Improvement of learning results of the self-organizing map by calculating fractal dimensions. In *Proc. of ESANN'94, Brussels, Belgium*, 4 1994.