

Deakin Research Online

This is the published version:

Keith-Magee, Russell, Venkatesh, Svetha and Takatsuka, Masahiro 1999, The importance of neighbourhood size in self organising systems, in *ICONIP'99 : Proceedings of the 6th International Conference on Neural Information Processing*, IEEE, [Washington, D. C.], pp. 267-272.

Available from Deakin Research Online:

<http://hdl.handle.net/10536/DRO/DU:30044824>

Reproduced with the kind permissions of the copyright owner.

Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

Copyright : 1999, IEEE

The Importance of Neighbourhood Size in Self Organising Systems

Russell Keith-Magee, Svetha Venkatesh
School of Computing
Curtin University of Technology
Perth
Western Australia
{keithmag,svetha}@cs.curtin.edu.au

Masahiro Takatsuka
Department of Geography
The Pennsylvania State University
302 Walker Building, University Park
Pennsylvania
masa@gis.psu.edu

Abstract

In recent times, analysis of SOM performance has concentrated on optimising gain decay, rather than the size, form, and decay of the neighbourhood function. We propose that the size, form, and decay of region size plays a much more significant role in the learning, and especially the development, of topographic feature maps.

In this paper, a biologically derived SOM model is presented. This model is able to select a single winning neuron, and form Gaussian outputs about this winner, without the need for a meta-level decision making structure to artificially select a winner and fit a Gaussian output about that winner. Using this model, some fundamental characteristics of the relationship between neighbourhood size and SOM output states are demonstrated.

1. Introduction

The theory underlying self organisation was originally grounded in a biological process - the topological organisation of stimuli observed in the cortex. Self organising behaviour can be found in a number of cortical groups, in a wide range of species. While these biological systems are interesting in themselves, it is the cause of this phenomenon that is of more interest to those wishing to emulate biological intelligence in a computational framework. Why is the cortex organised in this way, and through what mechanism do neurons fall into this manner of organisation?

The work of [6] provided an algorithmic model of the mechanism by which this topographical organisation can occur. This model was biologically motivated; however, in recent times, the focus of SOM research has

moved away from the biological origins of self organising processes, to concentrate on largely empirical studies of the performance of Kohonen's computational approximation of the underlying biological process [5]. While Kohonen's model is a useful computational tool, it is difficult to optimise the performance of Kohonen's algorithm without a grounding in the theory of the underlying self organising process.

To date, analysis of SOM performance has concentrated on optimising gain decay, rather than the size, form, and decay of the neighbourhood function, within Kohonen's algorithm. For example, [5] suggests that a neighbourhood set N_c , or plasticity control kernel $h_{C(x),i}$ of the same order of magnitude as half the largest dimension of the neuron array should be sufficient, especially if the neighbourhood function is time invariant. Alternatively, [2] states "the theoretical study in the one-dimensional case is nearly complete. The convenient decreasing rate to ensure the ordering is still to be found", but the proof presented relies upon a time invariant, or 0 neighbour lateral interaction function.

We propose that the size, form, and decay of region size plays a much more significant role in the learning, and especially the development, of topographic maps. In this paper, an investigation into neighbourhood size is performed using a biologically derived SOM model. Using this model, some fundamental characteristics of the relationship between neighbourhood size and SOM output states are demonstrated.

1.1. Malsberg's SOM Model

At the time at which [3] performed their work, genetic predetermination was the best explanation that could be offered for cortical topological organisation. Malsberg believed that the complete genetic predetermination of connections was implausible, and so, proposed an alternative mechanism. Malsberg proposed that rather than being predetermined, the topological organisation of connection weights is an algorithmic method which is fully functionally dependent on the provided stimuli. The self organising plane consists of excitatory E-cells, and inhibitory I-cells, in equal numbers. The connection between cells is modeled by a function $f(x)$, where x is the distance between cells. This function decreases monotonically with increasing distance - for example, a Gaussian function centered at $x = 0$. This function can be characterised by its maximum amplitude A , and the range R over which the connections are significant. The connections between $E \rightarrow E$, $E \rightarrow I$, $I \rightarrow E$, and $I \rightarrow I$ cells vary according to the following relationship:

- $f_{EE}(x)$ has $A_{EE} > 0$, range R_{EE} ,

- $f_{EI}(x)$ has $A_{EI} > 0$, range $R_{EI} = R_{EE}$,
- $f_{IE}(x)$ has $A_{IE} < 0$, range $R_{IE} > R_{EE}$,
- $f_{II}(x)$ has $A_{II} = 0$,

When superimposed, these weights yield a net Laplacian interconnection between cells. Using these lateral connections, and randomly instantiated synaptic weights, each cell is allowed to relax into a stable state. During relaxation, the excitation level of the cells oscillates; however, these oscillations dampen, approaching a stable value.

Learning was then performed using a Hebbian update rule - synapses are increased in strength proportional to the level of excitation. The gain used in this learning remained constant throughout the experiment. Normalisation across cells is also performed to prevent saturation. Malsberg reached four key conclusions regarding self organising processes [8]:

1. Maps develop in a step-by-step, orderly fashion,
2. Lateral connections within the map are initially widespread, but over the learning process they decay in extent,
3. The orientation of the map is established early in the self organising process; the final pattern of connections takes longer to develop,
4. Appropriate starting conditions are essential to the formation of a good map.

The conclusions reached by Malsberg are theoretical generalisations, rather than experimental results [8, Sec. 6]. In particular, the second conclusion — that decaying region size is a requirement for topological learning — was untested. However, two important qualities are demonstrated by Malsberg's experiments.

Firstly, learning was achieved without the need for a decaying learning gain parameter. In most theoretical and experimental studies since, variation in gain has been considered the most significant factor controlling learning.

Secondly, global topographical organisation was only observed on small maps; in large map simulations, local topographical organisation occurred, but global topology was erratic. The cause of this result was theoretically demonstrated in [1]. In this proof, it was demonstrated that a single peak of excitation will result, provided the domain of the lateral connection function spans the entire map; a lateral connection function with a smaller range will result in multiple peaks occurring on a map. During Malsberg's experiments, the range of the

lateral connection function was only a small number of neurons, and was not decayed. Consequently, on small maps, the connection function spanned the entire map, and global organisation was observed; on larger maps, the small connection function did not span the map, and many areas of excitation occurred, resulting in erratic global topology.

2. Investigation of Neighbourhood Size

In this section, an iterative mechanism with a Laplacian lateral interaction kernel is used to provide a Gaussian fitted about a local region of good fit. This model will be used to demonstrate some key features of region size change during the learning and development process of topographical self organising maps.

2.1. Implementation

The algorithm used to implement the SOM model is very similar to that used by Malsberg. The learning process consists of two stages; an internal iterative loop, which allows lateral connections to relax the output of each neuron, and an external loop which controls the Hebbian update based upon output values. Given a set of input vectors \vec{X} , this algorithm learns a topographical mapping for the set of weights w_{ij} . Lateral connections are defined in a kernel L , defined as:

$$L_{ij} = (1 + a) \exp\left(\frac{-\|i - j\|^2}{2r^2}\right) - a \exp\left(\frac{-\|i - j\|^2}{2br^2}\right) \quad (1)$$

For the purposes of this experiment, $a = 0.3$ and $b = 3$. The size of weight update for neuron i is proportional to the output of neuron i in response to the training pattern; this update is drawn from [4]. Each cycle of the outer Repeat-Until loop (Algorithm 1) is an epoch of the training regime.

The algorithm for the inner loop, the function **Evaluate_Map_Output**, is not as straightforward. This is a result of the same algebraic explosion problem observed in previous sections. To counter this algebraic explosion, a process of normalisation is required. This normalisation can be performed using a number of methods. Two versions of the inner iterative loop are presented here.

In the first version (Algorithm 2), algebraic normalisation is performed at the end of each iteration. At the conclusion of the iteration process, the normalised output is passed through a sigmoid function to provide maximum and minimum saturated outputs. In the second version, there is no explicit normalisation process - the process of normalisation is performed on each iterative loop by

Mainline**repeat**

Evaluate changes in learning parameters α and σ
while \exists a pattern which has not been presented this epoch **do**
 Randomly select a training pattern \vec{X} which has not been presented this epoch
 Evaluate_Map_Output \vec{v} for training pattern
 Update weights according to rule:

$$w_{ij}(t+1) = \frac{w_{ij}(t) + \alpha(t)\vec{X}v_i}{\|w_{ij}(t) + \alpha(t)\vec{X}v_i\|}$$

end while**until** finished training

Algorithm 1: Mainline algorithm to implement SOM ordering.

Evaluate_MapOutput ($L, \vec{u}, \beta, W, \vec{X}, N$): $\vec{u} = \vec{0}$ **for** $C = 1 \dots N$ **do** $\vec{u} = W\vec{X} + \beta L\vec{u}$ $\vec{u} = \frac{\vec{u}}{\|\vec{u}\|}$ **end for** $\vec{v} = \sigma(\vec{u})$ return \vec{v}

Algorithm 2: Computationally normalised version of the Evaluate_Map_Output algorithm.

a sigmoid function tuned to the dynamic range of the iterative process.

This model is much easier to control. Given an input data vector \vec{X} and a weight matrix W , the algorithm calculates the output \vec{v} . The strength of feedback is controlled by the efficacy parameter β , and the strength of lateral connections is modeled by L , a function of $|i - j|$. At each iteration, the internal voltage is dynamically scaled to prevent algebraic explosion. The final output value is the result of passing a sigmoid over the internal voltage \vec{u} ; parameters for this sigmoid are custom selected to suit the dynamic range of \vec{u} .

This process of algebraic normalisation guarantees a consistent dynamic range on output, permitting the use of a fitted sigmoid for the final transformation of output. However, this model is somewhat artificial, as normalisation is a meta-level function, requiring knowledge of the raw output of all other neurons in the system.

The second model (Algorithm 3) is biologically more plausible; however, it requires significantly more effort to tune. Given an input data vector \vec{X} and a weight matrix M , the algorithm calculates the output \vec{v} . The

Evaluate_MapOutput ($L, \vec{u}, \beta, W, \vec{X}, N$): $\vec{u} = \vec{0}$ $\vec{v} = \vec{0}$ **for** $C = 1 \dots N$ **do** $\vec{u} = W\vec{X} + \beta L\vec{v}$ $\vec{v} = \sigma(\vec{u})$ **end for**return \vec{v}

Algorithm 3: Biologically plausible version of the Evaluate_Map_Output algorithm.

strength of feedback is controlled by the efficacy parameter β , and the strength of lateral connections is modeled by L , a function of $|i - j|$. At each iteration, the output value is the result of passing a sigmoid over the internal voltage \vec{u} ; parameters for this sigmoid are selected to suit the dynamic range of \vec{u} . Note that in this model it is the output voltage \vec{v} , not the internal voltage \vec{u} that is used for feedback.

Changes in lateral interaction kernels, training patterns, or any other features of the network rapidly lead to saturation or depletion states in the output neurons. In a biological system, this sort of dynamic tuning could be performed by intra-neuron monitoring of input, output and weight conditions; however, in the computational framework as it stands, this kind of tuning is not at our disposal.

Aside from philosophical differences, the differences between the two algorithms are largely negligible. The output states obtained from the two systems are largely analogous; slight differences can be observed, but this can be attributed to slight differences in the tuning of the sigmoid functions.

In the following experiments, the first, algebraically normalised version is utilised. This is done for no reason other than convenience. The biological version is presented for completeness; to demonstrate that the self organising process *could* be performed using a completely biologically plausible model, without the need for meta-level functions.

2.2. Experimental Results

In these experiments, a 13 input, 100 neuron SOM was used. These 100 neurons were arranged into a ring, so neurons on one end were attached to the other. This was achieved by wrap around in the lateral interaction kernel. The Laplacian L of Equation 1 was used for this kernel, with $a = 0.3$, $b = 3$, and σ varying through the experiment. This kernel was normalised. 10 iterations were used to evaluate the neuron output values. In the learning algorithm, the gain parameter was set at a con-

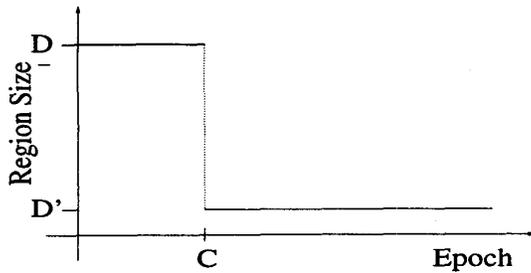


Figure 1: Region decay scheme for experiment. Region size is kept at a large value D for C epochs, then decayed to D' for the remainder of training.

stant $\alpha(t) = 0.3$, to minimize the degrees of freedom in the experiment. Lateral efficacy β was set at 1.

The training set consisted of patterns simulating pointillistic sensations on a 1D skin surface; each training vector was of the form $\vec{X} = [0 \dots 0 \ 0.5 \ 1.0 \ 0.5 \ 0 \dots 0]^T$. The first and last data vectors were $\vec{X} = [0.5 \ 1.0 \ 0.5 \ 0 \dots 0]^T$ and $\vec{X} = [0 \dots 0.5 \ 1.0 \ 0.5]^T$ respectively. These patterns were each normalised.

Two similar experiments were performed, to establish the effect of a reduction in region size on the map. In both experiments, the SOM was first trained for 50 epochs with a region size of $\sigma = 20$. The output states resulting from this initial training can be seen in Figures 2(a)–2(c). At the conclusion of this initial training, the map can be seen to converge into a stable state, with good topological organisation. Each curve in Figure 2(c) represents the output response of a single training pattern; the maxima of these curves are seen to be evenly distributed across the 1D SOM.

After this initial training, the region size was reduced. The scheme used to decay the region size is shown graphically in Figure 1. In the first experiment, the region size was reduced to $\sigma = 4$. As a result of this reduction, the neuron output responses can be observed to sharpen significantly (compare Figures 2(c) and 2(d)). This highlights the benefit of a reduction in region size to precision on a map.

In the second experiment, the region size was reduced to $\sigma = 2$. However, the result of this reduction was not as positive as in the first experiment. Instead of sharpening the output states, each output state is split into multiple peaks. The effect is mild at first, but the splitting behaviour becomes more severe as training with the smaller σ continues.

To clarify the role played by the period of training with a large region size, an attempt was made to train the

same network without the initial training period with a $\sigma = 20$ kernel. Instead, all training was performed with a kernel of size $\sigma = 4$. However, this training regime was unable to form a global topographical organisation - instead, a large number of zones of local organisation formed. This indicates that a period of training with a large region size is essential to the formation of global topology; small region size kernels can be used to improve precision on an existing map, but not to establish a map from scratch.

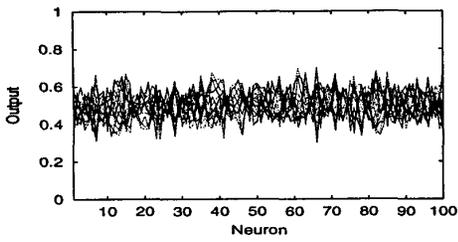
2.3. Discussion

This experiment highlights a number of key features of SOM training. Firstly, it highlights the importance of region size to the formation of topographic maps. In order to produce a global topographical organisation, an initially wide spanning - empirically, it would seem reasonable to hypothesize that the 6σ diameter (3σ , or 99.5% of area, both sides of the mean) of the lateral interaction kernel must slightly exceed the size of the map. This concurs with the experimental results of [6] and [8], and one of the conclusions of [1].

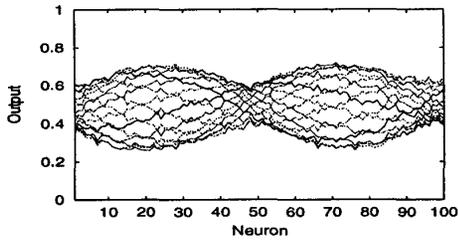
This experiment also highlights the role to be played by a reduction in region size during training. The use of a large region size may guarantee global topology, but this global topology comes at the price of local precision. By reducing the region size of the lateral interaction kernel once global topology has been established, it is possible to significantly improve the precision of the output response of the SOM to each pattern. This concurs with the suggestion of [5].

However, the amount by which region size is decayed is significant. If the region size is decayed by too large an amount, local topologies will emerge. This is not a particularly significant result if a single 'winner', the maximum of the output response, is to be used as the sole reporting feature of the SOM. It is also not of concern in algorithms such as Kohonen's, where the artificial selection of a maximum prevents the formation of local topologies. However, if we intend to use the entire output response, the increase precision brought about by training with a reduced region size is an essential feature. Care must therefore be taken to not decay neighbourhood size by too large an amount. This concurs with the results of [9].

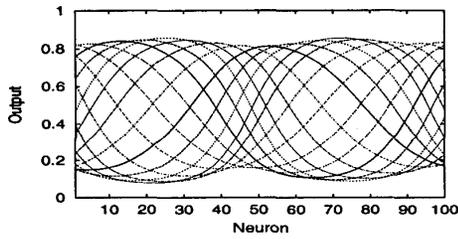
The optimal decay for the kernel region size remains to be determined. However, it is reasonable to hypothesize that just as there is a relationship between the 6σ diameter of the kernel and the size of the map, there is a relationship between the 6σ diameter of the output response, and the 6σ diameter of the lateral interaction kernel. In other words, during initial training, the span of



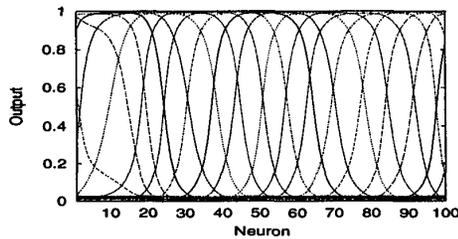
(a) 5 Epochs



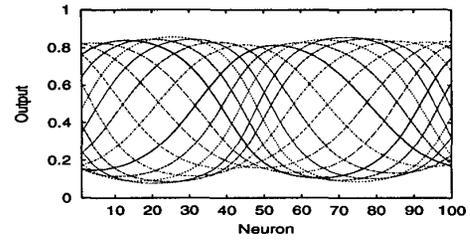
(b) 20 Epochs



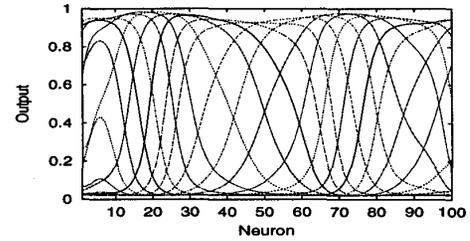
(c) 40 Epochs



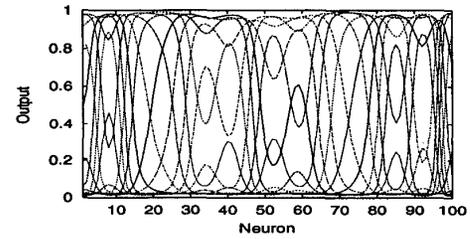
(d) 100 Epochs



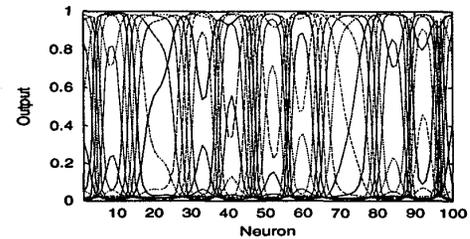
(a) 40 Epochs



(b) 60 Epochs



(c) 80 Epochs



(d) 100 Epochs

Figure 2: Output response over time for a non-WTA network. Each line is the output response for a single training pattern. $\sigma = 20$ for epochs 1–50; $\sigma = 4$ for epochs 51 onwards.

Figure 3: Output response over time for a non-WTA network. Each line is the output response for a single training pattern. $\sigma = 20$ for epochs 1–50; $\sigma = 2$ for epochs 51 onwards.

the lateral interaction kernel must be equivalent to the size of the map; after training, the size of the kernel can be reduced to span only the range of activated outputs. It is also interesting to note that these self organising systems converge *without* the need for a sophisticated scheme of learning gain decay. This seems to contradict the vast majority of literature (For example, [5, 2, 7], which emphasizes the importance of learning gain, leaving neighbourhood size as a peripheral concern.

3. Conclusion

In recent times, analysis of SOM performance has concentrated on optimising gain decay, rather than the size, form, and decay of the neighbourhood function. In this paper, the importance of neighbourhood size to the learning and development process was established. A mathematical model was presented, consolidating the biological model described by Malsberg, and the Kohonen's computational model. This model is able to select a single winning node, and form Gaussian outputs about this winner, without the need for a meta-level decision making structure to artificially select a winner and fit a Gaussian output about that winner. This model can be used as an empirical tool for investigating novel lateral connection strategies.

Using a biologically inspired computational model, it was shown that the range of the lateral interaction kernel can have a significant effect on the speed of convergence, and the precision of representations; region size also determines whether the map will converge at all. Further study of region size decay schemes and lateral connection functions are required.

4. Future Research

These results suggest three avenues for future research. Firstly, the results obtained in the previous section suggest further theoretical or empirical studies of region size decay schemes. This paper has demonstrated the roles played by large and small kernel sizes to the learning process; a study of the balance between the two extremes should lead to the development of optimized training regimes for self organising systems.

Secondly, an analysis of lateral connection functions other than the Laplacian is warranted. The topologies formed using the Laplacian kernel represent a simple topological mapping; by using other kernels (such as an inverted Gaussian), it may be possible to obtain topological maps representing features other than simple spatial relationships. This could open a wide range of new uses for self organising algorithms.

Lastly, the work presented in this paper should be adapted for use in Kohonen's SOM algorithm. In this paper, a continuous model of self organisation derived from the work of Malsberg was used to test the importance of the form and size of the lateral connection function. Kohonen's algorithm is a discrete approximation of this type of process which is in common use as a result of its computational efficiency. Whilst the two algorithms are fundamentally similar, some fine tuning will be required to adapt the findings of this paper to Kohonen's algorithm.

References

- [1] Shun-Ichi Amari. Topographic organisation of nerve fields. *Bulletin of Mathematical Biology*, 42:339–364, 1980.
- [2] M. Cottrell, J.C. Fort, and G. Pags. Theoretical aspects of the SOM algorithm. *Neurocomputing*, 21:119–137, 1998.
- [3] D. H. Hubel and T. N. Wiesel. Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol*, 26:994–1002, 1963.
- [4] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43:59–69, 1982.
- [5] Teuvo Kohonen. *Self Organising Maps*. Springer Series in Information Sciences. Springer-Verlag, 2nd edition, 1997.
- [6] Chr von der Malsberg. Self-organisation of orientation sensitive cells in the striate cortex. *Kybernetik*, 14:212–225, 1973.
- [7] Filip Mulier and Vladimir Cherkassky. Learning rate schedules for self-organising maps. In *Proceedings fo the 12th IAPR International Conference on Pattern Recognition*, volume 2, pages 224–8, 1994.
- [8] D.J. Willshaw and C. von der Malsberg. How patterned neural connections can be set up by self-organisation. *Proc R. Soc. Lond*, B(194):431–445, 1976.
- [9] A. Zell. *Simulation Neuronaler Netze*. Addison-Wesley, Bonn, Germany, 1994.