ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 261-266

Self-Organisation in the SOM with a Finite Number of Possible Inputs

John A. Flanagan
Helsinki University of Technology
Neural Network Research Center
PO Box 5400 FIN-02015 HUT
Finland

Abstract. Given a one dimensional SOM with a monotonically decreasing neighbourhood and an input distribution which is not Lebesque continuous, a set of sufficient conditions and a Theorem are stated which ensure probability one organisation of the neuron weights. This leads to a rule for choosing the number of neurons and width of the neighbourhood to improve the chances of reaching an organised state in a practical implementation of the SOM.

1. Introduction

Kohonen's Self-Organising Map (SOM) algorithm [7] was originally conceived as a heuristic description of the process of self-organisation which occurs during learning in parts of the brain, especially the cortex. It finds widespread application in many different areas such as clustering, vector quantization, and data mining where different topologies of the SOM are used, with different types of input data. Analysing the self-organising mechanism of the SOM has proved difficult, and for the most part the only results available are for the one dimensional case.

In the one dimensional SOM there are a total of N neurons and with each neuron i is associated a neuron weight $x_i \in R$. The neuron weight vector is denoted by $\mathbf{X} = (x_1, x_2, \ldots, x_N)$. The input is assumed to be a random variable $\omega \in R$ with a probability distribution $F(\omega)$. At iteration t a *winner* neuron $v(t)$ is chosen such that

$$v(t) = \arg\min_{1 \leq i \leq N} |\omega(t) - x_i(t)|. \tag{1}$$

Each neuron weight $x_i(t),\ i = 1, 2, \ldots, N$ is then updated as

$$x_i(t+1) = x_i(t) + \alpha(t)h(|i - v(t)|)(\omega(t) - x_i(t)), \tag{2}$$

where the gain $\alpha(t)$, with $0 < \alpha(t) < 1$ during the training phase is normally a decreasing function with time. In what follows it is assumed that $\alpha(t) \geq \alpha_m > 0, \forall\ t$. The function $h(|i - v(t)|)$ with $0 \leq h(|i - v(t)|) \leq 1$ is referred

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 261-266

to as the *neighbourhood* and $h(j)$ decreases with increasing j. In what follows $h(|i - v(t)|)$ will be written as $h(i, v(t))$. In general h satisfies the following four conditions *a)* $h(i,i) = 1$, *b)* $h(i, i \pm W) = h_m > 0$, *c)* $h(i,j) = 0$ when $|i-j| > W$, *d)* $h(i,j) \leq h(i,k)$, for $|i-j| > |i-k|$, where W is the *width* of the neighborhood and h_m its minimum non zero value. In practise under a wide variety of conditions, that is for different F, α, and h it has been found that the weights will converge to the ordered configuration

$$D = \{\mathbf{X} : x_1 < x_2 < \ldots < x_N\} \bigcup \{\mathbf{X} : x_1 > x_2 > \ldots > x_N\}. \quad (3)$$

No general analysis of the SOMs self-organising ability exists and little is known about the essential conditions required for self-organisation. In the one dimensional case the organised configuration D of the neuron weights is *absorbing*, and in [2] it has been shown that from any initial condition where $x_i \neq x_j, i \neq j$, $W = 1$, and a uniform F that the weights will almost surely converge to D. This result was further generalised in [1], [6], [3], [4] and [8]. All of the latter consider $\mathbf{X}(t)$ as a Markov process and to prove self-organisation it is shown that $\exists\ T < \infty$ and $\delta > 0$ for which,

$$\pi_{\mathbf{X}(0)}(\{\psi : \ \tau_D \leq T\}) \ \geq \delta \qquad \forall\ X(0), \quad (4)$$

or that the probability $\pi_{\mathbf{X}(0)}$, of finding sets of samples $\psi = \{\omega(0), \omega(1), \ldots\}$ which take the neuron weights $\mathbf{X}$ from all initial conditions $\mathbf{X}(0)$ to the organised configuration in a finite time τ_D is non zero. In [2], [1] and [6] either a uniform F or a diffuse F has been assumed. The generalisation of these results is limited by the existence of situations where there is a tie for winner. An example of when this may occur is $x_i(t) = x_j(t), i \neq j$ and

$$i, j = \arg\min_{1 \leq k \leq N} |\omega(t) - x_k(t)|. \quad (5)$$

In this situation Sadeghi [8] defined the winner in such a way that the instability resulting from this tie no longer existed and was able to prove self-organisation for a one dimensional SOM with any form of decreasing neighbourhood function, and a quite general distribution of the input. The only restriction on the distribution of the input was that it must be continuous with respect to the Lebesque measure, which means for any set A if,

$$\mu_L(A) = 0, \text{ then } F(A) = 0, \quad (6)$$

where $\mu_L(A)$ is the Lebesque measure of A. From a theoretical point of view this is not such a big restriction, but for the practical implementation of the SOM it is limiting. A result of Sadedghi's analysis based on Lebesque continuity of F is that the weights organise independent of the number of neurons N and width W of the neighbourhood if $W > 0$.

Consider what happens in practise where the support $S_M = \{\omega : f(\omega) > 0\}$ of the input consists of a finite number of points M, let S_M be given by,

$$S_M = \{\omega_1, \ldots, \omega_M\}, \qquad \omega_i < \omega_j, \ i \neq j. \quad (7)$$

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 261-266

F is not Lebesque continuous on S_M. Consider the very simple case of $M = 2$, $W = 1$ and $\mathbf{X}(0)$ such that,

$$\omega_1 < x_1(0) < x_2(0) < x_3(0) < \omega_2 < x_4(0) < x_5(0) < x_7(0) < x_6(0). \tag{8}$$

For $\omega(t) = \omega_1$ then $v(t) = 1$ and for $\omega(t) = \omega_2$ then $v(t) = 3, 4$, this is true $\forall\ t \geq 0$, hence $x_6(t) = x_6(0)$ and $x_7(t) = x_7(0)$ and the weights can never reach an organised configuration. In [3] it was shown for this example if $W \geq 7$ (i.e. $W \geq N$) then the weights will organise with probability one. From this simple example it is obvious in the case where F is not Lebesque continuous that self-organisation depends at least on the width W of the neighbourhood and the number of neurons N. It should be emphasised that this is always the case in practise. In what follows a set of conditions on the support of the input distribution, which is not Lebesque continuous, is described along with a Theorem which specifies the relationship between N, W, M which are sufficient to ensure probability one self-organisation of the neuron weights. This result is a generalisation of the cases analysed in [3], [4], and apply for a neighbourhood where $h(i) \geq h(j) + \phi$, for $i < j$ and $\phi > 0$ is the smallest difference between any two values of the neighbourhood.

2. Conditions on the Support of F

Given the support S_M of equation (7) of the distribution F the question is to define conditions that it must satisfy to ensure probability one self-organisation of the neuron weights. These conditions are defined in terms of intervals $A_n, n = 0, 1 \ldots$ based on S_M. Begin by defining a basic interval $A_0 = [\mu_0, \nu_0]$ around $\omega_i \in S_M$ such that,

$$\mu_0 \ < \ \omega_i \ < \ \nu_0, \tag{9}$$

where the distance $\gamma_0 = \nu_0 - \mu_0$ can be made arbitrarily small. The definition of A_0 is completely independent of the SOM, and an interval A_0 can be defined for every $\omega_i \in S_M$. To avoid unnecessary complication of the notation the intervals A_0 are not indexed for each ω_i. Consider two A_0 intervals, the first one the A_0 interval of ω_i the second the A_0 interval of ω_j. Assume that $i < j$ and denote the A_0 interval on the lower valued part of the real line by A_0^l (i.e. the A_0 interval of ω_i) and the A_0 interval on the higher valued part of the real line by A_0^h (i.e. the A_0 interval of ω_j). All parameters associated with A_0^l, A_0^h are superscripted with l, h respectively. Further assume that they satisfy the following condition,

$$\frac{\max(\gamma_0^l, \gamma_0^h)}{\mu_0^h - \nu_0^l} \ < \ \frac{\alpha_m \phi}{1 - \alpha_m h_m}, \tag{10}$$

and $\mu_0^h - \nu_0^l > 0$ is the distance between the intervals, α_m is the minimum value of $\alpha(t)$, h_m the minimum non-zero value of h and ϕ the smallest difference

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 261-266

between two values of h. If this condition is satisfied call the smallest interval containing A_0^h, A_0^l by $A_1 = [\mu_1, \nu_1]$, where $\mu_1 = \mu_0^l$ and $\nu_1 = \nu_0^h$. Figure 1 shows an illustration of this interval. Note that for S_M there can be many pairs ω_i, ω_j satisfying this condition and hence many intervals can be referred to as an interval A_1. In a similar way a general interval $A_{n+1} = [\mu_{n+1}, \nu_{n+1}]$ can be defined in terms of two A_n intervals. As before $A_n^l = [\mu_n^l, \nu_n^l]$ is the A_n interval on the lower valued part of the real line and $A_n^h = [\mu_n^h, \nu_n^h]$ is the A_n interval on the higher valued part of the real line. Again parameters associated with A_n^l, A_n^h are superscripted with l, h respectively with $\mu_n^l = \mu_{n+1}$ and $\nu_n^h = \nu_{n+1}$, and to form an interval A_{n+1} they satisfy the following condition,

$$\frac{\max(\gamma_n^l, \gamma_n^h)}{\mu_n^h - \nu_n^l} < \frac{\alpha_m \phi}{1 - \alpha_m h_m}, \tag{11}$$

where $\mu_n^h - \nu_n^l > 0$ is the distance between the intervals. By replacing the subscripts 0 by n and 1 by $n+1$ in Fig. 1 an illustration of an interval A_{n+1} is obtained. Figure 2 shows possible intervals $A_n, n = 1, 2, 3$ for an example support S_{18}. It is easily deduced that an interval A_n is composed of 2^n, A_0 intervals, which of course means defining A_n on S_M is only possible if $M \geq 2^n$.

In the next section a theorem of self-organisation is stated which shows the conditions specified in the definition of an interval A_n are sufficient to prove probability one self-organisation of the weights.

3. A Proof of Self-Organisation

As stated in the introduction the most successful proofs of self-organisation in the SOM are based on finding sequences of inputs, which take the neuron weights from all initial conditions to an organised configuration in a finite time with a positive probability. The following Theorem states the conditions on the SOM for which such sequences can be found, given an interval A_n. The theorem is then discussed and an indication given as to how it is proved.

Theorem 1 *Given an interval A_n and $N \leq nW$, there exists $T < \infty$ and $\delta > 0$ for which,*

$$\pi_{\mathbf{X}(0)}(\{\psi \ : \ \tau_D \leq T\}) \geq \delta, \tag{12}$$

$\forall \ \mathbf{X}(0)$*, where τ_D is the first entry time of $\mathbf{X}$ into D.*

What this Theorem states is that for an interval A_n any SOM, with a decreasing neighbourhood function, will almost surely reach an organised state if the total number of neurons N and the width W of the neighbourhood are such that,

$$N/W \leq n. \tag{13}$$

In the case analysed by Sadeghi [8] where the support of F is Lebesque continuous it is shown that self-organisation occurs independent of N and W if

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 261-266

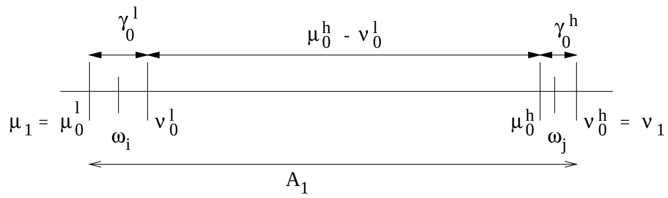

Figure 1: Illustration of an interval A_1 with the various parameters labelled.

$W \geq 1$. In this case when the support of F is not Lebesque continuous the sufficient conditions for self-organisation depend very much on N and W. Such a result has meaning in a general practical case ; to define an interval A_n requires at least that $M \geq 2^n$ thus it would seem in a practical situation given S_M that N, W should be chosen as,

$$N/W \;\; \leq \;\; \log_2 M. \tag{14}$$

The proof of this Theorem, fully described in [5] is conceptually quite simple being based on an inductive arguement. The first step is to assume that for an interval A_n with $N \leq nW$ and for all $\mathbf{X}(0)$ that by choosing $\omega(t) \in A^l_{n-1}$ there is $t_1 < \infty$ for which $x_i(t_1) \in A^l_{n-1}$, $\forall\ i$. Given the weights in this state an assumption is made on A^h_{n-1} where for $\omega(t) \in A^h_{n-1}$ there is $t_2 < \infty$ for which $x_i(t_2) \in A^h_{n-1}$, $\forall\ i$ and either $x_N(t_2) < x_i(t_2), \forall\ i \neq N$ or $x_1(t_2) < x_i(t_2), \forall\ i \neq 1$. Using once again the assumption on A^l_{n-1} means for $\omega(t) \in A^l_{n-1}$ there is some $t_3 < \infty$ for which $x_i(t_3) \in A^l_{n-1}$. Based on the fact that at t_2 either neuron N or 1 is the winner it can be shown that at t_3 the neurons are in an organised configuration. Thus the assumptions made on A^l_{n-1}, A^h_{n-1} are sufficient for self-organisation when $N \leq nW$. The next step of the inductive arguement is to show that when these assumptions are true then an interval A_{n+1} is sufficient for self-organisation when $N \leq (n+1)W$. Finally it is shown that an interval A_2 is sufficient for self-organisation when $N \leq 2W$. The result of the Theorem follows from the normal inductive arguement.

4. Conclusion

Given F which is not Lebesque continuous with support S_M. If it is possible to define a set of conditions on S_M in terms of an interval A_n then for an SOM with a monotonically decreasing neighbourhood function almost sure convergence to an organised configuration occurs for all $\mathbf{X}(0)$ if $N/W \leq n$. This theoretical result can be stated in a form significant in practise, in that to increase the likelihood of the SOM reaching an organised configuration N, W should be chosen such that $N/W \leq \log_2 M$.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 261-266

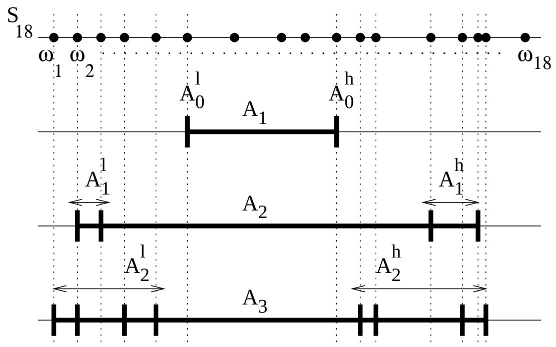

Figure 2: A support S_{18} and possible intervals $A_n, n = 1, 2, 3$ indicated by dark horizontal lines. Vertical dark lines signify an interval A_0 associated with the ω_i to which it is joined by the vertical dotted lines.

References

[1] Catherine Bouton and Gilles Pagès. Self-organization and a. s. convergence of the one-dimensional Kohonen algorithm with non-uniformly distributed stimuli. *Stochastic Processes and Their Applications*, 47:249–274, 1993.

[2] Marie Cottrell and Jean-Claude Fort. Étude d'un processus d'auto-organisation. *Annales de l'Institut Henri Poincaré*, 23(1):1–20, 1987. (in French).

[3] John A. Flanagan. Self-organization in Kohonen's SOM. *Neural Networks*, 9:1185–1197, 1996.

[4] John A. Flanagan. Sufficient conditions for self-organisation in the one dimensional SOM with a reduced width neighbourhood. *Neurocomputing*, 21:51–60, 1998.

[5] John A. Flanagan. Self-organisation in the one dimensional SOM with a decreasing neighbourhood. Revised and resubmitted to Neural Networks, 1999.

[6] Jean-Claude Fort and Gilles Pagès. On the a.s. convergence of the Kohonen algorithm with a general neighbourhood function. *Annals of Applied Probability*, 5(4):1177–1216, 1995.

[7] Teuvo Kohonen. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, 1995. (Second Extended Edition 1997).

[8] A.A. Sadeghi. Self-organisation and convergence of the one dimensional Kohonen algorithm. In *Proceedings ESANN98*, pages 173–178, Brussels, 1998. D Facto ed.