ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 209-214

The Self-Organising Map, Robustness, Self-Organising Criticality and Power laws

John A. Flanagan

Helsinki University of Technology
Neural Networks Research Center
Finland

Abstract. Observations of complex dynamical systems operating at Self-Organising Criticality (SOC) have shown them to be inherently robust to fluctuations in their environment. This SOC has the signature spectrum $S(f) \propto 1/f^{\beta}, \beta \approx 1$ for some variable f in the system. The observation of power laws in the spectrum of the updates of the neuron weights in the SOM are reported. Such a signature is shown in the SOM for certain types of neighbourhood functions, which are intuitively robust. Other neighbourhood functions have different spectrums which are presented, but their meaning remains to be explained.

1. Introduction

The Self-Organising Map (SOM) as developed by Kohonen [4], and based on the biological process of retinotopy, has been used in many different applications with success. One of the reasons for the widespread application of the algorithm is its robustness, that is, with an unknown training data set the neuron weights will generally self-organise for a broad range of network parameters (i.e. gain, neighbourhood function). Relatively little fine tuning of the parameters is required to produce this organised state, although for the convergence phase this may not be so. The question to be addressed here is the robustness of the self-organising phase.

In the natural and man-made world there exist many non-linear dynamical systems with both spatial and temporal degrees of freedom, which can be classified as dissipative coupled systems and exhibit complex behaviour. The SOM, during the training phase, can be viewed as such and one of the characteristics of systems belonging to this class are their inherent robustness. A phenomena associated with such systems is the formation of spatial structures which are scale invariant and have self-similar structures (fractals) [6]. A second phenomena in such systems is the existence of power laws, which have been observed in systems as diverse as semiconductors, traffic flow [5], written text [7], occurrence of earthquakes, size of cities, living systems [1]. The best known example is probably $1/f$ noise in semiconductor devices [3]. Here the noise caused by a current flowing through a silicon sample has a frequency spectrum proportional

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 209-214

to $1/f$ where f is the frequency the noise is measured at. Such systems evolve towards, what has been called a self-organised critical (SOC) [2] state.

Although the existence of power laws in nature are well know the origin of these power laws is not well understood. Despite the fact that for example the existence of fluctuations in a semiconductor with a $1/f$ spectrum are considered as noise, now it may be seen more as the macroscopic signature of the existence of a system operating at a self-organised critical state.

In this work by the use of simulations the existence of SOC in the SOM during the training phase will be investigated. The approach is to show that there exists fluctuations in the SOM which follow a power law over certain ranges. It is hoped that this will give a new perspective to the analysis of self-organisation in the SOM.

In what follows the SOM to be analysed is presented, with a brief description of power laws and where they exist in the SOM. This will be followed by the presentation of simulation results and conclusions.

2. The SOM and Power Laws

For the pourpose of simulation a one dimensional SOM is used. There are a total of N neurons with neuron weights x_i for neuron i. At each time t there is an input $\omega(t)$ and the winner neuron $v(t)$ is such that

$$v(t) \;=\; \min_i |x_i(t) - \omega(t)|$$

Each weight is then updated as

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

The gain function $0 \leq \alpha(t) < 1$, and the neighbourhood function is defined such that $h(v,v) = 1$ and $h(v,|v \pm j|) \geq h(v,|v \pm i|)$ for $i > j$. For a broad range of parameters and input signals the weights will go to one of two organised states $x_1 < x_2 < \ldots < x_N, x_N < x_{N-1} < \ldots < x_1$ which are both absorbing. It is known that for $h(v,|v \pm j|) = 0, j > 0$ that the neuron weights will not self-organise.

Before discussing power laws in the SOM a few words about power laws themselves. A power law implies a functional relationship of the following form,

$$g(y) \;=\; \frac{K}{y^{\beta}} \tag{2}$$

for some constants β and K. In most examples of SOC systems the value of β is most often close to 1, hence $1/f$ noise in semi-conductors. Its actual value is usually related to specific parameters of the system. In using graphs to display power law relationships of measured data it is most common to plot log-log graphs. The resultant curve is linear with a slope of $-\beta$

$$\log_{10}(g(y)) \;=\; -\beta\log_{10}(y) \;+\; \log_{10}(K)$$

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 209-214

This form of plot is used here for presentation of simulation results with a least squares curve fit to determine the slope $-\beta$.

The parameter y and function $g(y)$ to be analysed in the case of the SOM is the occurrence of values of the updates of the weight values. Thus at each iteration the update of each neuron is given from equation 1 by,

$$\Delta x_i(t) \;=\; h(i,v)|(\omega(t) - x_i(t))| \qquad (3)$$

The $\alpha(t)$ will be kept constant and is considered as a scaling factor and not included here as part of the update. The function $g(y)$ used here will be the spectrum of the Δx_i or in the discrete case, as here, a histogram.

3. Simulations

The approach taken in the simulations is to use SOMs with a large number of neurons and to perform many iterations and then average over all the neurons. Unless otherwise stated in the simulations presented below $N = 500$ and 10^7 iterations are performed. The spectrum S is calculated by dividing the range of possible updates (e.g. $0-1$) into j intervals with f_j the midpoint of interval j. At each iteration the update of each neuron is calculated and the counter $S(f_j)$ of each interval is updated for each neuron having an update in the interval j. Thus the number of updates in each interval are averaged over all the neurons. Figure 1 shows the result for an SOM with a neighbourhood function $h(v,i) = \exp^{(\frac{4|v-i|}{N})^2}$ with a uniformly distributed input ω. It is seen that at higher values of f the linear behaviour of the curve is lost. In the literature mentioned previously the effects of finiteness are sometimes used to explain deviations from linearity. However later on it is seen that here the curve could be piecewise linear having different regimes for different values of f. As it turns out it is difficult to classify this behaviour. In figure 2 four decades of f from $10^{-6.5}$ to $10^{-2.5}$ of the graph in figure 1 are shown. The linearity is quite obvious. The dotted line least squares approximation for equally spaced points on the log scale gives a $\beta = 0.815$. To see the variation in the curve for a different type of input distribution, figure 3 shows the spectrum of updates for a normal distribution (i.e. a sum of uniformly distributed inputs). All other parameters of the SOM remain the same as the previous case. As seen the spectrum is also linear with a slope of -0.804 for the dotted line approximation. Thus the slope of the curve does not change significantly with respect to the probability distribution of the input signal. To compare these results with the case of a neighbourhood function where $h(v,v) = 1$ and $h(v,j) = 0, \forall j \neq v$ (i.e. no neighbourhood), which as stated already is a case known not to self-organise. The spectrum of the updates is shown in figure 4 a). It is quite different from the previous curves. For values of $\log_{10}(f)$ greater than -3 the slope is zero. In the transition for $\log_{10}(f)$ between -3 and -1.5 the slope is approximately -6.0.

Figure 4 b) shows the spectrum of updates for the neighbourhood function $h(v,|v \pm j|) = 1, j \leq 1$ and $h(v,|v \pm j|) = 0, \forall j > 1$ (i.e. one neighbour) which

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 209-214

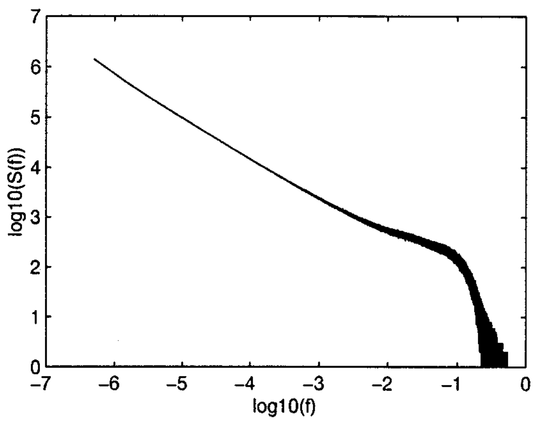

Figure 1: Spectrum of the updates for a uniformly distributed input.

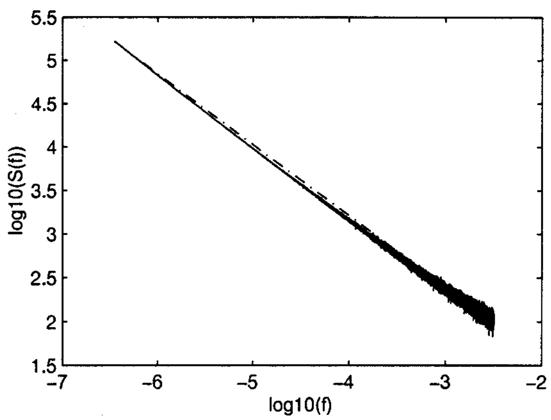

Figure 2: Spectrum of the updates for a uniformly distributed input. Dotted line is a linear approximation of slope -0.815.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 209-214

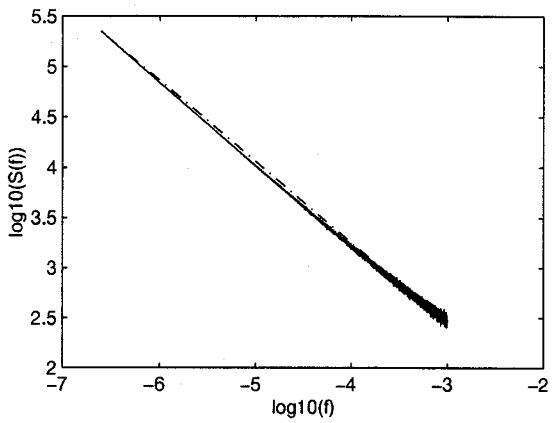

Figure 3: Spectrum of the updates for a normally distributed input. Dotted line is a linear approximation of slope -0.804.

should lead to an organised state of the weights. While the shape is similar to that of a), the slope of the curve in the transition region for $\log_{10}(f)$ between -2.5 and -1.2 is approximately $\beta = 2.5$.

The question now is to interpret these results. First it is clear that the critical factor for self-organisation is the neighbourhood function. For the cases of the extended neighbourhood function for lower values of f there is a linear relationship with a $\beta \approx 0.8$, which is close to 1 the value associated with SOC. This would suggest a certain robustness of the system. However there is a range for higher values of f, still linear, where $\beta \approx 4$. In the case of zero neighbourhood function a very different characteristic is obtained which is dominated by a uniform spectrum with a sharp transition region at higher f. Introducing a neighbourhood function of width 1, leads to a similar characteristic but with a more gentle transition which has $\beta \approx 2.5$ and decreases as the width of the function increases. This transition however, has been observed to be sharper for smaller numbers of neurons. However while decreasing the number of neurons changes β the ratio of the βs for the two cases, of no neighbourhood and one neighbour, remains the same. Intuitively and from experience an SOM has a better chance of reaching an organised state for broader neighbourhood functions, thus the slope of the spectral curves could be taken as an indication of whether a system can reach SOC, and hence a measure of the system's robustness.

4. Conclusions

The existence of power laws in the spectrum of the updates of the neuron weights in one dimensional SOMs has been demonstrated. The form of the power law depends on the type of neighbourhood function. In classical SOC

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 209-214

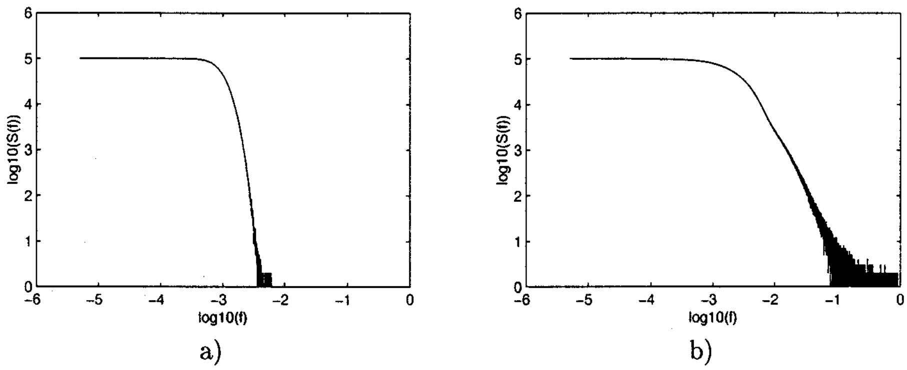

Figure 4: Spectrum of the updates for a uniformly distributed input with a neighbourhood function a) $h(v,v) = 1$ and $h(v,j) = 0, \forall j \neq v$, b) $h(v, |v \pm j|) = 1, j \leq 1$ and $h(v, |v \pm j|) = 0, \forall j > 1$.

theory the robustness of a system is considered greatest for exponent values of β close to 1. Here this has been observed for decreasing neighbourhood functions extending over large numbers of neurons. For no neighbourhood a different spectrum has been observed, consisting of a flat spectrum and a sharp transition. By introducing a neighbourhood the slope of this transition is reduced. However this effect depends on the number of neurons, further studies are required to determine if this is a numerical effect or an inherent effect of the SOM.

References

[1] C. Adami. Self-organised criticality in living systems. *Physics Letters A*, 203:29–32, 1995.

[2] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organised criticality. *Physical Review A*, 38(1):364–374, 1988.

[3] F. N. Hooge, T. G. M. Kleinpenning, and L. K. J. Vandamme. Experimental studies on 1/f noise. *Reports on Progress in Physics*, 44:480–531, 1981.

[4] Teuvo Kohonen. *Self-Organizing Maps.* Springer, Berlin, Heidelberg, 1995.

[5] Kai Nagel and Maya Paczuski. Emergent traffic jams. *Physical Review E*, 51(4):2909–2918, 1995.

[6] Manfred Robert Schroeder. *Fractals Chaos, Power Laws.* W.H. Freeman and Compnay, New York, 1990.

[7] George Kingsley Zipf. *Human Behavior and the Principle of Least Effort.* Addison Wesley, Cambridge MA, 1949.