WSOM 2001 proceedings - Workshop on Self-Organizing Maps, Lincoln (United Kingdom), 13-15 June 2001
Advances in Self-Organising Maps, N. Allinson, H. Yin, L. Allinson, J. Slack eds., Springer Verlag, 2001, ISBN 1-85233-511-4, pp. 67-72

Recursive learning rules for SOMs

John A. Lee, Nicolas Donckers, and Michel Verleysen

Université catholique de Louvain, Laboratoire de Microélectronique,
Bâtiment Maxwell, Place du Levant, 3, B-1348 Louvain-la-Neuve, Belgium

Summary. Three extensions of the traditional learning rule for Self-Organizing Maps are presented. They are based on geometrical considerations and explore various possibilities regarding the norm and the direction of the adaptation vectors. The performance and convergence of each rule is evaluated by two criteria: topology preservation and quantization error.

1 Introduction

Self-Organizing Maps (SOMs, [1,2]) are well known in the domain of Vector Quantization (VQ). Unlike other VQ methods, the neurons (or prototypes) used for the quantization are given a position in a grid, which is often one- or two-dimensional. This predefined geometrical organization, combined with a well chosen learning rule, generates a self-organizing behavior, useful in numerous areas like nonlinear projection and data representation. More technically, learning rules for VQ can be classified into two sets, according to the number of neurons which are adapted at each stimulation of the network:

- 'Winner Takes All' (WTA) rules, like for Competitive Learning (CL, [3]), where only one neuron is adapted;
- 'Winner Takes Most' (WTM) rules, like for Neural Gas (NG, [4]), where all neurons are adapted.

In order to observe self-organization, learning rules for SOMs have to fulfill two conditions: they belong to the WTM set and they use information given by the position of each prototype in the grid. Other characteristics of the rules are more or less unconstrained. Section 2 studies these degrees of freedom and presents four different learning rules (three alternative rules in addition to the traditional one). Next, section 3 describes some experiments and criteria to test the four rules. Section 4 gathers results and discusses them. Finally, conclusion (section 5) shows that one of the proposed rules improves convergence with respect to the traditional SOM rule.

2 Four learning rules

First of all, let's write the traditional learning rule for SOMs. Suppose that the map is defined by:

1. matrix $\boldsymbol{W}$, of which rows $\boldsymbol{w}_r$ give the weights in the feature space;
2. function $d(q, r)$, measuring the distance between neurons q and r in the grid space.

At learning time t, a vector $\boldsymbol{x}_i$ of the feature space stimulates the map (see also Figure 2, left):

$$\boldsymbol{w}_r^{t+1} = \boldsymbol{w}_r^t + \mathit{\Delta}\boldsymbol{w}_r^t = \boldsymbol{w}_r^t + \gamma_r^t(\boldsymbol{x}_i - \boldsymbol{w}_r^t). \tag{1}$$

In this equation, $\gamma_r^t = \alpha^t e^{-0.5(d(r,*)/\lambda^t)^2}$ is the learning rate for neuron r, computed as the product of the global learning rate α^t (between 0 and 1, decreasing in time) and the neighborhood factor depending on the neighborhood width λ^t (between 0 and ∞, decreasing in time). Finally, $*$ is the index of the 'winning' neuron, such that $\|\boldsymbol{x}_i - \boldsymbol{w}_*\| \leq \|\boldsymbol{x}_i - \boldsymbol{w}_s\|$ for all neurons s.

From a geometrical point of view, the traditional SOM rule adapts neurons radially around the stimulating vector $\boldsymbol{x}_i$. Now, consider that the self-organizing map is like a fishing net crumpled on the ground and that the neurons are like small balls linked together by short pieces of elastic cord. Consider also that the stimulating vector is like the hand of the fisherman, taking a node of the net between two fingers. In this case, the neurons do not move directly towards the hand, i.e. radially, but instead they are pulled by neighboring neurons. Formally, it means that the neurons have to be adapted towards neuron q, where index q indicates the neuron preceding neuron r on the shortest path [5] between neuron r and winning neuron $*$. This idea is illustrated on the right of Figure 2 and leads to the following learning rule:

$$\mathit{\Delta}\boldsymbol{w}_r^t = \gamma_r^t(\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t), \tag{2}$$

where neuron r has to be adapted *after* neuron q. Note that the neighborhood factor cannot be suppressed, although the new rule presents an interesting property without the neighborhood factor. Indeed, suppose that the map is reduced to a string in a one-dimensional feature space and that its current state is such that:

- $\boldsymbol{w}_r^t = r$, with $\boldsymbol{w}_*^t = * = 0$;
- $\boldsymbol{w}_r^t$ is linked with $\boldsymbol{w}_{r-1}^t$;
- the stimulating vector is $\boldsymbol{x}_i = 0$.

If λ^t is set to $+\infty$, then the neighborhood has no limit, γ_r^t degenerates to α^t and one can rewrite the traditional rule into:

$$\boldsymbol{w}_r^{t+1} = \boldsymbol{w}_r^t + \alpha^t(\boldsymbol{x}_i - \boldsymbol{w}_r^t) = (1 - \alpha^t)r, \tag{3}$$

while the fisherman's rule leads to a recurrence:

$$\boldsymbol{w}_r^{t+1} = \boldsymbol{w}_r^t + \alpha^t(\boldsymbol{w}_{r-1}^{t+1} - \boldsymbol{w}_r^t) = (1 - \alpha^t)r + \alpha^t\boldsymbol{w}_{r-1}^{t+1}, \tag{4}$$

solved by $\boldsymbol{w}_r^{t+1} = (1 - \alpha^t)\sum_{i=0}^{r} i(\alpha^t)^{r-i}$.

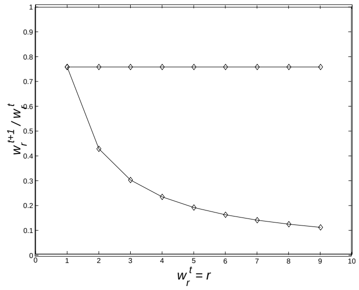

Fig. 1. Weight ratio after and before adaptation, with $\alpha^t = 0.25$, for the traditional SOM rule (constant horizontal line) and for the fisherman's rule (decreasing curve)

Figure 1 is a plot of the weight ratio $\boldsymbol{w}_r^{t+1}/\boldsymbol{w}_r^t$, showing that even without the neighborhood factor, the fisherman's rule induces an attenuation of the adaptation, as the distance to the winning neurons grows. This constatation does not imply that the neighborhood factor is useless in the fisherman's rule. Actually, it would be like running the traditional rule with λ^t set as constant: convergence is not guaranteed.

3 Description of the experiments

Comparing the traditional learning rule and the fisherman's one is almost impossible because their nature is totally different. In order to make them comparable, one has to list their differences one by one and try each combination. Actually, there are two differences between both learning rules: the traditional rule is non-recursive and purely radial, while the fisherman's one is recursive but not radial. This leads to four combinations showed in equations 5a to 5d, table 1 and figures 2 and 3. The four rules can be written more or less the same way:

$$\Delta \boldsymbol{w}_r^t = \gamma_r^t \quad \|\boldsymbol{x}_i - \boldsymbol{w}_r^t\| \quad \frac{\boldsymbol{x}_i - \boldsymbol{w}_r^t}{\|\boldsymbol{x}_i - \boldsymbol{w}_r^t\|}, \tag{5a}$$

$$\Delta \boldsymbol{w}_r^t = \gamma_r^t \, \|\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t\| \quad \frac{\boldsymbol{x}_i - \boldsymbol{w}_r^t}{\|\boldsymbol{x}_i - \boldsymbol{w}_r^t\|}, \tag{5b}$$

$$\Delta \boldsymbol{w}_r^t = \gamma_r^t \quad \|\boldsymbol{x}_i - \boldsymbol{w}_r^t\| \quad \frac{\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t}{\|\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t\|}, \tag{5c}$$

$$\Delta \boldsymbol{w}_r^t = \gamma_r^t \, \|\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t\| \quad \frac{\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t}{\|\boldsymbol{w}_q^{t+1} - \boldsymbol{w}_r^t\|}. \tag{5d}$$

In each of these four equations, one finds from left to right the learning rate, the norm of the adaptation and its direction. To evaluate these four rules, two criteria are defined. The first one is the relative quantization error, written as:

$$E_{VQ} = \sqrt{\sum_i (\boldsymbol{x}_i - \boldsymbol{w}_*)^2} \;/\; \sqrt{\sum_i (\boldsymbol{x}_i - \mu \boldsymbol{x}_i)^2}, \tag{6}$$

WSOM 2001 proceedings - Workshop on Self-Organizing Maps, Lincoln (United Kingdom), 13-15 June 2001
Advances in Self-Organising Maps, N. Allinson, H. Yin, L. Allinson, J. Slack eds., Springer Verlag, 2001, ISBN 1-85233-511-4, pp. 67-72

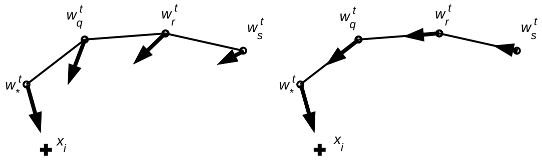

Fig. 2. Traditional SOM rule (left) and fisherman's rule (right)

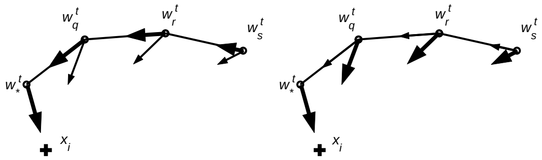

Fig. 3. Hybrid rules: non radial non recursive (left) and recursive radial (right)

Table 1. Rules

	Non-recursive	Recursive
Radial	Rule 5a (Figure 2, left)	Rule 5b (Figure 3, right)
	Traditional SOM Rule	Hybrid
Non-radial	Rule 5c (Figure 3, left)	Rule 5d (Figure 2, right)
	Hybrid	Fisherman's rule

where index i traverses the whole learning set, $\boldsymbol{w}_*$ without time index t is the winning neuron after learning and $\mu\boldsymbol{x}_i$ is the averaged stimulus. The second criterion measures the topology preservation:

$$E_T = \sum_s \sharp \left\{ r | r \neq s, r \notin N(s), \|\boldsymbol{w}_r - \boldsymbol{w}_s\| < \max_{q \in N(k)} \|\boldsymbol{w}_q - \boldsymbol{w}_s\| \right\}, \quad (7)$$

where $\sharp$ is the set cardinality operator and $N(k)$ is the set of direct neighbors of neuron s. This error criterion is specially designed for honeycombed maps and counts the number of prototypes which are badly positioned on the map. More precisely, E_T approximates each hexagonal cell with a circle and counts all prototypes that are inside the circle while they should be outside. If the map is thoroughly unfolded, then the circles approximate very well the hexagonal cells and E_T gives a value close to zero. On the contrary, when the map is crumpled, hexagonal cells are distorted, circles grow and so does E_T. Unlike other criteria [6], E_T easily detects when maps are twisted, as shown in Figure 4.

The experiments were conducted on honeycombed maps with 100 neurons (see Figure 4). The training set was always the same: 2500 samples drawn from a uniform distribution of width 1 and length 4, so that the map can fit

WSOM 2001 proceedings - Workshop on Self-Organizing Maps, Lincoln (United Kingdom), 13-15 June 2001
Advances in Self-Organising Maps, N. Allinson, H. Yin, L. Allinson, J. Slack eds., Springer Verlag, 2001, ISBN 1-85233-511-4, pp. 67-72

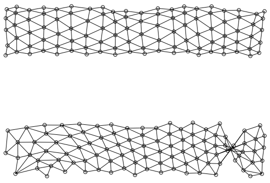

Fig. 4. Rectangular honeycombed maps with 100 neurons: if parameters are well chosen, then convergence is fast (first map), and E_T is below 20; on the contrary, if λ^t decreases too fast, then convergence is slower, twist effects may occur (second map) and E_T around 100

perfectly. The maps were randomly initialized with 100 of the 2500 samples and convergence was stopped after 5 epochs on the data set. Different values were tested for parameters α^t and λ^t, the initial value being always ten times larger than the fifth one (last epoch); the three intermediate values are computed with an exponential decrease from the initial to the final value. Each parameter configuration was repeated 300 times.

4 Results and Discussion

Table 2 and 3 summarizes some results for the four rules (λ^t is given relatively to the largest grid distance). A first good result is that all four rules perform well when α^t and λ^t values are well chosen. But, when parameters are too low or too high, some differences appear. The main difference exists between radial rules and non-radial ones, which are rapidly outperformed. Two rules remain: the traditional one and the recursive radial. When λ^t is low, the traditional rule wins, but shortly. On the contrary, when λ^t is high, the recursive rule is clearly better. Generally, when λ^t is high, twist effects tend to disappear with the recursive rule. An intuitive explanation for this unexpected but good performance may be the following: after initialization, the map is strongly crumpled and therefore the path along the links from any neuron r to the winning neuron $*$ can be much longer than the distance between stimulus and neuron r. This leads to stronger adaptation norms and better unfolding. Moreover, the recursive rule can move neurons beyond the stimulus (this occurs when the last neuron on the shortest path is further than stimulus). The poor performance of non radial rules is probably also due to initialization: when the map is crumpled, the adaptation vectors have a strong tangential part and a weak radial component, so that neurons move a lot, but without coherence in their direction.

Table 2. Mean values of E_{VQ} for various parameter settings

α^1	0.5	0.5	0.5	0.5	0.7	0.7	0.7	0.7
λ^1	0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9
5a	0.0732	0.0871	0.1076	0.1317	0.0725	0.0869	0.1076	0.1320
5b	0.0736	0.0851	0.0988	0.1145	0.0727	0.0843	0.0981	0.1132
5c	0.0800	0.1118	0.1569	0.1957	0.0774	0.1093	0.1516	0.1918
5d	0.0790	0.1016	0.1265	0.1495	0.0772	0.0984	0.1240	0.1480

WSOM 2001 proceedings - Workshop on Self-Organizing Maps, Lincoln (United Kingdom), 13-15 June 2001
Advances in Self-Organising Maps, N. Allinson, H. Yin, L. Allinson, J. Slack eds., Springer Verlag, 2001, ISBN 1-85233-511-4, pp. 67-72

Table 3. Mean values of E_T for various parameter settings

α^1	0.5	0.5	0.5	0.5	0.7	0.7	0.7	0.7
λ^1	0.3	0.5	0.7	0.9	0.3	0.5	0.7	0.9
5a	0.88	5.67	16.96	35.07	3.27	9.53	24.67	44.89
5b	1.51	2.95	5.71	16.46	0.70	5.36	7.37	16.44
5c	62.22	137.64	264.68	421.46	42.53	131.15	238.25	398.88
5d	32.46	79.94	152.38	236.06	24.79	64.36	137.83	223.80

5 Conclusion

This study has shown that the recursive radial rule is more robust than the traditional SOM rule when the neighborhood is wide. Actually, recursiveness fastens the convergence during the first epochs. Therefore, an idea for future work consists in combining both rules: the recursive rule for map initialization and rough unfolding, and the traditional one for fine tuning.

Acknowledgement. This work was realized with the support of the 'Ministère de la Région wallonne', under the 'Programme de Formation et d'Impulsion à la Recherche Scientifique et Technologique'. Michel Verleysen is Research Associate of the Belgian National Fund for Scientific Research (FNRS).

References

1. T. Kohonen. Self-organization of topologically correct feature maps. *Biological Cybernetics*, 43:59-69, 1982.
2. T. Kohonen. Self-organization and Associaative Memory. Springer-Verlag, Berlin, 3rd edition, 1989.
3. A. Ahalt, A.K. Krishnamurthy, P. Chen and D.E. Melton. Competitive learning algorithms for vector quantization. *Neural Networks*, 3:277-290, 1990.
4. T. Martinetz and K. Schulten. A neural gas network learns topologies. In T. Kohonen *et al.*, editor, *IEEE International Conference on Artificial Neural Networks, Espoo, Finland*, volume 1, 397-407. Elsevier, 1991.
5. E.W. Dijkstra. A note on two problems in connection with graphs. *Numerical Math.*,1:69-271, 1959.
6. H.U. Bauer and K.R. Pawelzik. Quantifying the neighborhood preservation of self-organizing maps. *IEEE Transactions on Neural Networks*, 3:570-579, 1992.