Available online at www.sciencedirect.com

SCIENCE DIRECT®

Neural Networks 19 (2006) 812–816

Neural Networks

www.elsevier.com/locate/neunet

2006 Special Issue

SOM's mathematics

J.C. Fort

Laboratoire de Statistique et Probabilités, 118 route de Narbonne, Toulouse, France

Abstract

Since the discovery of the SOM's by T. Kohonen, many results that provide a better description of their behaviour have been found. Most of them are very convincing, but from a mathematical point of view, only a few are actually proved. In this paper, we make a review of some results that are still to be proved and give some framework to formulate various questions.

Keywords: Magnification factor; Grid equilibrium; Topology preservation; Limit theorems

1. Introduction

This paper is an opportunity to review the mathematical results established about the now classical Kohonen algorithm. While the applications of Self-Organizing Maps (SOM's) are numerous, only a few theoretical results are available. Moreover most of them are concerned with the one-dimensional case which is a very particular framework far from the applications. Nevertheless a mathematical study is needed for at least two goals. The first one is to actually prove observed facts, which could lead to a better knowledge of the behaviour of the algorithm. The second one is to better understand what Self Organization is in order to propose some possible new algorithms based on this knowledge.

This paper is "self-organized" as follows. We begin with some analytical results which are well known but require a more rigourous proof. Then we look at the convergence of the learning algorithm. We briefly describe what is proved, which is very limited with respect to what is needed and useful for the applications. At this stage, the question of "what the organization is", is discussed. After these somewhat theoretical aspects, we investigate some more statistical problems.

2. Analytical results

2.1. Magnification factor

We begin with one of the most discussed topics in the framework of the quantization and more recently of SOM's: the magnification factor. It is a fact that, when the SOM units (centroids or code vectors) are equally weighted, then the simplest reconstructed distribution is biased. If we denote by $x^{(n)} = (x_1^{(n)}, \ldots, x_n^{(n)})$ the values of the n units minimizing the square-distortion in the quantization problem (or 0-neighbour Kohonen algorithm) the simplest reconstruction of the data distribution is given by:

$$\mu_{x^{(n)}}^n := \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i^{(n)}}, \quad n \geq 1$$

where $\delta_{x_i^{(n)}}$ is the Dirac measure centered on $x_i^{(n)}$.

Then $\mu_{x^{(n)}}^n$ weakly converges when n goes to ∞, to the distribution

$$\mu^{\infty} := \frac{f^{\frac{\mathrm{d}}{\mathrm{d}+2}}(\xi)}{\int_{\mathbb{R}^d} f^{\frac{\mathrm{d}}{\mathrm{d}+2}}(u)\,\mathrm{d}u}\mathrm{d}\xi$$

if the data distribution has a probability density function (p.d.f.) f over $\mathbb{R}^d$. The so-called magnification factor is then $\alpha = \frac{\mathrm{d}}{\mathrm{d}+2}$.

The same phenomenon occurs in the SOM case as mentioned by Kohonen in 1982 (see Kohonen (1984)).

Of course the more effective reconstruction which is given by (see Fort and Pagès (2002))

$$\widetilde{\mu}_{x^{(n)}}^n := \sum_{1\leq i\leq n} \mu(C_i(x^{(n)}))\delta_{x_i^{(n)}}$$

avoids this problem, where $C_i(x^{(n)})$ is the Voronoï tessellation of $x_i^{(n)}$ and measure μ is defined by $\mu(A) = \int_A f(\xi)\mathrm{d}\xi$.

E-mail address: fort@math.ups-tlse.fr.

0893-6080/$ - see front matter
doi:10.1016/j.neunet.2006.05.025

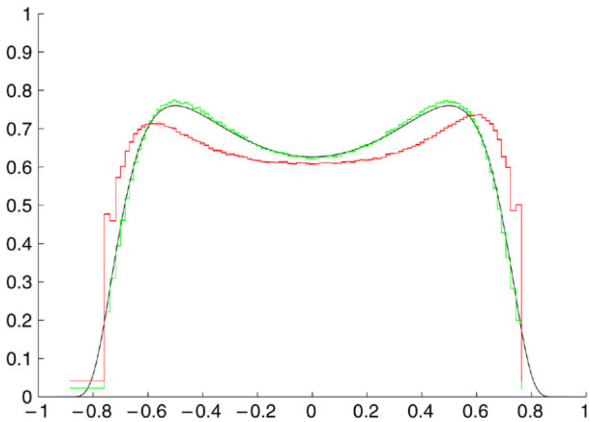

Fig. 1. The p.d.f. reconstruction, with $\mu^n_{x^{(n)}}$ in light gray and with $\widehat{\widetilde{\mu}^n_{x^{(n)}}}$ in dark gray. The true f is superimposed to the dark gray curve.

It is easily seen that $\widetilde{\mu}^n_{x^{(n)}}$ weakly converges towards μ. Moreover one can prove that the "on line" estimation of the μ-mass of $C_i(x^{(n)})$ almost surely (a.s.) converges to the true value. Thus the natural on-line estimate of $\widetilde{\mu}^n_{x^{(n)}}$ (say $\widehat{\widetilde{\mu}^n_{x^{(n)}}}$) of the data distribution converges a.s. to the true distribution. See the illustration in Fig. 1.

As we are able to reconstruct the true distribution in the 0-neighbour case, the question of the magnification factor in $\mu^n_{x^{(n)}}$ becomes purely theoretical. In addition, it needs some mathematical treatment to be established in the case of SOM's (with neighbours).

Ritter and Schulten (1988) and Ritter (1991) have given an asymptotic expansion of α (when n goes to ∞) in the one-dimensional case with the $2k + 1$ neighbours (k neighbours on each side),

$$\alpha \sim \frac{2}{3} - \frac{1}{3(k^2 + (k+1)^2)}.$$

To obtain this expansion, a very strong assumption is needed: $x_i^{(n)}$ is asymptotically a twice differentiable function with respect to the "discrete" variable i.

Thus two facts are to be proved:

1. To justify the asymptotic expansion we need the existence of a smooth map $g: u \in [0,1] \longrightarrow x_u = g(u)$, which is the limit value (for uniform convergence) of $g^{(n)}: \frac{i}{n} \longrightarrow x_i^{(n)}$ for $i = 1, \dots, n$.
2. To connect $\mu^n_{x^{(n)}}$ and $\widetilde{\mu}^n_{x^{(n)}}$ we have to prove that:

$$\sup_{1 \le i \le n} \left| n\,\mu(C_i(x^{(n)})) - f^{1-\alpha}(x_i^{(n)}) \int_{\mathbb{R}^d} f^{\alpha}(u)\,\mathrm{d}u \right| \xrightarrow{n \to +\infty} 0.$$

2.2. Grid equilibrium

The next problem concerns a fact well-known by practitioners (see Cottrell and Fort (1987) and Fort and Pagès (1995)) and illustrated in Figs. 2 and 3: the grid equilibrium is stable, in the case of the uniform distribution over $[0,1]^2$ with the 8-unit neighbourhood.

If we denote by h the mean field in $\mathbb{R}^d$ of the SOM algorithm update (without the ε learning rate), with a neighbourhood function σ, the stability of an equilibrium point depends on the eigenvalues of the gradient of h (see Fort and Pagès (1995)). For the lth component of h_i it reads:

$$\frac{\partial h_i^l}{\partial x_j} = \sum_{k \in I} \sigma(i,k)\mu(C_k(x))\delta_{ij} e_l + \sum_{k \ne j \in I} (\sigma(i,k) - \sigma(i,j)) \int_{\bar{C}_k(x) \cap \bar{C}_j(x)} (x_i^l - \omega^l) \times \left(\frac{1}{2} n_x^{kj} + \frac{1}{\|x_k - x_i\|} \left(\frac{x_k + x_j}{2} - \omega \right) \right) f(\omega) \lambda_x^{kj}(\mathrm{d}\omega),$$

where I is the set of units, $\lambda_x^{kj}(\mathrm{d}\omega)$ is the Lebesgue measure on the Voronoï border between x_k and x_j, i.e. the median hyperplan, n_x^{kj} is the normal unit vector $(\frac{x_j - x_k}{\|x_j - x_k\|})$ to this hyperplan, and $(e_l)_{1 \le l \le d}$ is the canonical basis of $\mathbb{R}^d$.

The three following results are still unproved, though there is empirical evidence:

1. In the case of the uniform distribution over $[0,1]^2$ with an 8-neighbour square grid, the product grid equilibrium is stable.
2. More generally in the case of uniform distribution over $[0,1]^d$ with a 2^d-neighbour "square" grid, the product grid equilibrium is stable.
3. This last result can be generalized to the case of the uniform distribution being replaced by the product of symmetrical distributions.

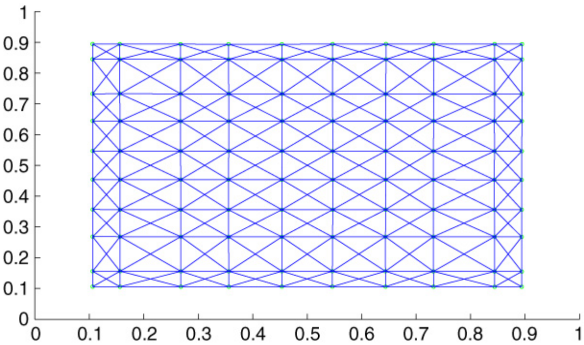

Fig. 2. The equilibrium obtained for an 8-unit neighbourhood function is a product grid equilibrium.

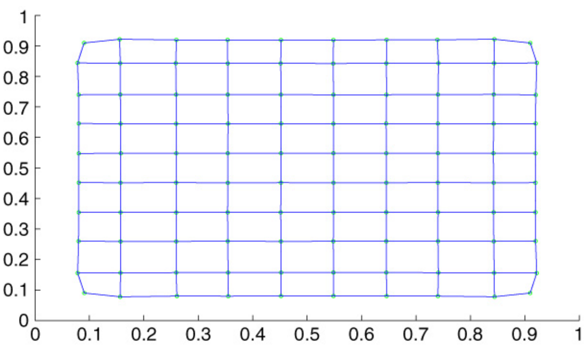

Fig. 3. The equilibrium obtained for a 4-unit neighbourhood function is not a product grid (see the corners).