ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 9-14

About the Kohonen algorithm: Strong or weak self-organization?

Jean-Claude FORT
Univ. Nancy I & Samos, Univ. Paris 1

Gilles PAGÈS
Labo. de Proba. Paris 6 & Univ. Paris 12

Abstract

The question of self-organization for the Kohonen algorithm is investigated. First the notions of organized states, weak and strong self-organizations are defined. Then we prove that the Kohonen string has not the strong self-organization property at least in two well-known cases: the string and the grid in $[0,1]^2$.

Introduction

The Kohonen algorithm in the early 80's is wellknown as a Self-Organizing Process. The very definition of what self-organizing exactly is turns out to be still unclear. It involves two different notions: "what is an organized state?" and "how to get it?"

So, we first have to properly define what an organized state is. So far, the only really satisfactory example was provided by the standard 1-dimensional Kohonen algorithm for which the organized states are the weight vectors having monotoneous components. In [3] is developed both a general definition of a "topology preserving map" and a measure of organization level with respect to such a topology for a given state. By somewhat weakening this definition we will introduce the concept of "organized state".

The second point to be discussed is what is as a self-organizing process. Once again, the only known example of a both rigourous, satisfactory – and possibly fulfilled – definition is the 1-dimensional standard Kohonen algorithm provides. Following this, it is most natural to claim that self-organization occurs whenever the set $\mathcal{O}$ of organized states satisfies

- $\mathcal{O}$ is an absorbing set (*i.e.* when the process reaches $\mathcal{O}$ it stays in it forever),
- the hitting time of $\mathcal{O}$ is almost surely finite for every starting weight vector.

These features will make up the *strong self-organization property.*

Some way round is to study the associated *ODE* that describes the mean behaviour of the algorithm. Self-organization occurs when the algorithm is con-

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 9-14

verging and all the possible attracting equilibrium points of the *ODE* lie in the organized state set. We may call this behaviuor *weak self-organization.* Unfortunately, it is usually not possible to prove that such a convergence holds.

Our aim is to show that, unlike the 1-dim case, the nearest neighbour multi-dim Kohonen algorithm has not the strong self-organization property.

1 About the notion of organized state

1.1 The topology

The general definition of organization involves two ingredients: a *weight* or *stimuli* space $\Omega \subset \mathbb{R}^d$ whose elements are denoted by ω, a *unit* set identified to a finite non empty set $\mathbf{I} \subset \mathbf{Z}^k$, endowed with a "topology".

Definition 1 *Let* $\mathbf{I}$ *be a unit set i.e. a (nonempty) subset of* $\mathbf{Z}^k$. *A family* $\mathcal{N} := \{\mathcal{N}(i), i \in \mathbf{I}\}$ *of subsets of* $\mathbf{I}$ *is a topology on* $\mathbf{I}$ *if it satisfies the following two assumptions:* $\begin{cases} (a) & \forall i \in \mathbf{I}, i \in \mathcal{N}(i) \text{ and } \mathcal{N}(i) \neq \mathbf{I}, \\ (b) & \forall i, j \in \mathbf{I},\ j \in \mathcal{N}(i) \Longleftrightarrow i \in \mathcal{N}(j). \end{cases}$
$\mathcal{N}(i)$ *is called the neighbourhood of unit* i.

Of course the standard "r-nearest neighbourhoods" satisfies this definition. One easily derives from the above definition that, if $\mathcal{N}$ and $\mathcal{M}$ are two topologies on $\mathbf{I}$, then their intersection $\mathcal{M} \cap \mathcal{N} := \{\mathcal{N}(i) \cap \mathcal{M}(i),\ i \in \mathbf{I}\}$ is a topology too.

1.2 A definition for the organized states

Our first task is to clearly define what everybody calls the dimension of the unit set $\mathbf{I} \subset \mathbf{Z}^k$ (which may be not k). Let $\mathcal{L}_\mathbf{I}$ be the lattice (or $\mathbf{Z}$-modulus) spanned by $\mathbf{I}$: this is the smallest $\mathbf{Z}$-linear subset of $\mathbf{Z}^k$ containing $\{i - i_0,\ i \in \mathbf{I}\}$ (where i_0 is any fixed element of $\mathbf{I}$). We recall the well-known result:

Theorem(-Definition) 2 $\mathcal{L}_\mathbf{I}$ *has at least one* $\mathbf{Z}$*-basis and every basis of* $\mathcal{L}_\mathbf{I}$ *is finite (with at most* k *elements). All* $\mathbf{Z}$*-basis have the same number of elements called the* $\mathbf{Z}$*-dimension of* $\mathcal{L}_\mathbf{I}$. *The dimension of* $\mathbf{I}$ *is the* $\mathbf{Z}$*-dimension of* $\mathcal{L}_\mathbf{I}$.

At this stage, we cannot distinguish the basis of $\mathcal{L}_\mathbf{I}$. However, in $\mathbf{Z}^2$ a 4-nearest neighbour topology related to the basis $\{(1,0),\ (0,1)\}$ is obviously not the same as that related to $\{(1,0),\ (1,5)\}$. So $\mathbf{I}$ must be endowed by a metric to take into account the intuitive notion of organization induced by a topology on $\mathbf{I}$.

Definition 3 *Let* $\|\,.\,\|$ *be a norm on* $\mathbb{R}^k$. *A basis* $(e_1, \cdots, e_k)$ *of* $\mathcal{L}_\mathbf{I}$ *is* $\|\,.\,\|$*-minimal if it achieves:* $\min\left\{\max_{1 \leq \ell \leq k} \|e_\ell\|,\ (\varepsilon_1, \cdots, \varepsilon_k)\ \mathbf{Z}\text{-basis of } \mathcal{L}_\mathbf{I}\right\}$.

Examples: • In the 1-dimensional case, whenever $\mathbf{I} \neq \{i_0\}$, $\mathcal{L}_\mathbf{I} = \mathbf{Z}$ and there are only 2 minimal basis (with respect to any norm): $\{1\}$ and $\{-1\}$.
• It becomes more intricated in higher dimension *e.g.*, if $\|x\|_p := (\sum_{\ell=1}^k |x_\ell|^p)^{\frac{1}{p}}, p \in [1, +\infty[$, $\|x\|_\infty := \max_{1 \leq \ell \leq k} |x_\ell|$, then the $\|\,.\,\|_\infty$-minimal basis of $\mathbf{Z}^2$ are $\{(0, \pm 1), (\pm 1, 0)\}$ and $\{(0, \pm 1), (\pm 1, \pm 1)\}$ while the $\|\,.\,\|_p$-minimal, $p \in [1, +\infty[$, are $\{(0, \pm 1), (\pm 1, 0)\}$.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 9-14

Now, we make a connection between topologies and metric properties.

Definition 4 *Assume that* $\dim_{\mathbb{Z}} \mathbf{I} = k$. *A topology is a* $\|\;\|$*-(admissible) topology if for any* $i \in \mathbf{I}$ *and any* $\|\;\|$*-minimal basis* $(e_1, \cdots, e_k)$ *of* $\mathcal{L}_{\mathbf{I}}$,

$$\{i + e_\ell,\ 1 \le \ell \le k\} \cap \mathbf{I} \subset \mathcal{N}(i).$$

Thus, in the 1-dimensional case a topology $\mathcal{N}$ is a norm-topology (for any norm) iff, for every $i \in \mathbf{I}$, $\{i-1, i, i+1\} \cap \mathbf{I} \subset \mathcal{N}(i)$.

Notice that for any basis $\mathcal{B}$ of $\mathcal{L}_{\mathbf{I}}$ there exists a norm $\|.\|_{\mathcal{B}}$ for which $\mathcal{B}$ is minimal: any $\|.\|_{p,\mathcal{B}}$ p-norm related to the $\mathcal{B}$-coordinates will work.

Example: Any standard r-nearest neighbour topology, $r \geq 8$, on $\mathbf{Z}^2$ is an ℓ^p-topology, $p \in [1, +\infty]$, while the standard 4-nearest neighbour only is for $p < +\infty$.

Similarly, we can extend such a definition to neighbourhood functions:

Definition 5 *(a)* $\sigma : \mathbf{I} \times \mathbf{I} \to [0,1]$ *is a* $\|.\|$*-neighbourhood function if* $\sigma(i,j) = \sigma(j,i)$ *and* $\sigma(i,i) = 1$ *and*

$$\mathcal{N}(i) := \{j \in \mathbf{I} \,/\, \sigma(i,.) > \alpha\},\ i \in \mathbf{I},\ \textit{is a } \|.\|\textit{-topology}.$$

The notion of organization itself relies once again on the Voronoï tessellation.

Definition 6 *(a) Let* $x := \{x_i,\ i \in \mathbf{I}\}$ *be a set of* pairwise distinct *points in* $\mathbb{R}^d$, *endowed with the Euclidean norm denoted* $|.|_2$. *The Voronoï tessellation (of* $\mathbb{R}^d$*) induced by* x *is the family* $\{C(i),\ i \in \mathbf{I}\}$ *of open sets in* $\mathbb{R}^d$ *given by:*

$$C(i) := \{u \in \mathbb{R}^d \,/\, |u - x_i|_2 < |u - x_j|_2,\ j \neq i\}.$$

(b) Two units $i, j \in \mathbf{I}$ *are Voronoï-neighbour if* $\overline{C}(i) \cap \overline{C}(j) \neq \emptyset$ *where* $\overline{C}(i)$ *denotes the closure of* $C(i)$. *Let* $\mathcal{V}(i)$ *denote the Voronoï-neighbourhood of* i.

Proposition 1 $\{\mathcal{V}(i),\ i \in \mathbf{I}\}$ *makes up a topology on* $\mathbf{I}$.

Definition 7 *(a) A state* $(x_i)_{i \in \mathbf{I}}$ *is* $\|.\|$*-organized if the associated Voronoï topology is an* $\|\;\|$*-(admissible) topology.*
(b) Let $\mathcal{O}$ *be the set of the* $\|.\|$*-organized states for at least one norm.* $\mathcal{O}$ *is called the set of organized states.*

2 The Kohonen algorithm

We exclusively deal here with the constant gain parameter – say ϵ – version of the algorithm and an i.i.d sequence of Ω-valued stimuli with distribution μ. Let $\{X_i(t), i \in \mathbf{I}\}$ denote the set of the weight vectors at time $t \in \mathbb{N}$, i.e the state of the process at time t. Then the algorithm is recursively defined by:

$$\begin{cases} i_0(\omega^{t+1}) & = \operatorname{argmin} \left\{|\omega^{t+1} - X_j(t)|_2,\ j \in \mathbf{I}\right\}, \\ X_j(t+1) & = X_j(t) + \epsilon\sigma(i_0(\omega^{t+1}), j)\big(\omega^{t+1} - X_j(t)\big),\ j \in \mathbf{I}. \end{cases}$$

Definition 8 *Assume that* $\mathbf{I}$ *is endowed with a* $\|.\|$*-topology. The sequence of weights* $(X(t)_i,\ i \in \mathbf{I})_{t \in \mathbb{N}}$ *is said to have the strong self-organizing property if there exists some* $t_0 \in \mathbb{N}$ *s.t., for every* $t \geq t_0$, *the state* $(X(t)_i,\ i \in \mathbf{I})$ *is organized for some norm* $\|.\|'$. *If* $\|.\| = \|.\|'$ *the organization is said to be optimal.*

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 9-14

Generally speaking, an optimal organization in the above sense cannot be obtained using the Kohonen algorithm. When $p=\infty$ and $\mathbf{I}$ is a rectangle of $\mathbf{Z}^2$, it has been previously noticed in [3] (see fig.1).

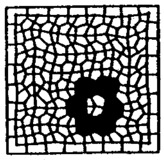

fig.1

A contrario, if a rectangle of $\mathbf{Z}^2$ is endowed with one of the r-nearest neighbour topology, organization occurs as soon as the Voronoï neigbourhoods $\mathcal{V}(i)$ contains the 4-nearest neighbours (if in $\mathbf{I}$).

From now on we will restrict to the standard r-nearest neighbour setting *i.e.* $\mathbf{I}$ is an (hyper-)rectangle and $\mathcal{N}(i)$ is a standard (hyper-)rectangle surrounding i. So $\mathcal{L}_\mathbf{I} = \mathbf{Z}^2$. We consider the "continuous case" when the stimuli space is a non empty (d-dim) compact domain Ω of $\mathbb{R}^d$. So $K(\epsilon;\, n;\, d,\, k,\, r)$ will denote a Kohonen algorithm with gain parameter ϵ, n units, some d-dimensional i.i.d. stimuli, a k-dimensional unit set endowed with a r-nearest neighbour topology.

The result sketched by Kohonen and then proved in [2] and [1] is that the algorithm $K(\epsilon;\, n;\, 1,\, 1,\, 2)$ reaches the absorbing set of monotoneous weight vectors at an almost surely finite time. This result holds as soon as the stimuli distribution μ locally has a continuous density (see [1] for improvements).

For more than ten years it has been an open question to find out such a non trivial absorbing set when d is greater than 1. Here we investigate this question for the Kohonen string in $\mathbb{R}^2$ $K(\epsilon;\, n;\, 1,\, 2,\, 2)$, and the Kohonen grid with 8 neighbours in $\mathbb{R}^2$ $K(\epsilon;\, n;\, 2,\, 2,\, 8)$. These results can easily be generalized to higher dimensional settings.

We close this section by the following important remark on the distribution μ. Any event related to the Kohonen algorithm which is true with a positive $\mathbb{P}_x$-probability for every $x \in [0,1]^d$ when μ is uniformly distributed over the hypercube $[0,1]^d$ still has positive probability for any stimuli measure μ (locally) having a continuous density on $\mathbb{R}^d$. Following this result, we may restrict our computations and experiments to $\Omega := [0,1]^d$ and $\mu := U([0,1]^d)$ in the proofs.

3 A mathematical result on the Kohonen string

Theorem 1 *The set $\mathcal{O}$ of organized states is not an absorbing set for the Kohonen string $K(\epsilon;\, n;\, 1,\, 2,\, 2)$ as soon as $n \geq 6$: the Kohonen string has not the strong self-organizing property.*

Recall that $\mathbf{I} := \{1, \cdots, n\}$ and – forcedly – $\mathcal{N}(i) = \{i-1,\, i,\, i+1\} \cap \{1, \cdots, n\}$.

Proof of the theorem: Let $\mathbb{P}_x$ denote the probability distribution of the algorithm starting from x. What we have to prove is that there exists some state $x \in \mathcal{O}$ and a finite time t such that: $\mathbb{P}_x(X(t) \notin \mathcal{O}) > 0$.

Step 1: We begin with a simple and general lemma which states that we only need to find a deterministic path from x to $\mathcal{O}^c$ (the outside of $\mathcal{O}$).

Lemma 2 *(Local continuity lemma) Set $C := \cup_{i \in \mathbf{I}} C_i$ where $C_i := \{(\omega, x) \,/\, |x_i - \omega|_2 < |x_j - \omega|_2,\, j \neq i\},\, , i \in \mathbf{I}$. Let $x \in \Omega^\mathbf{I}$ be a starting weight vector and $\omega^1, \cdots, \omega^t$ be an Ω-valued finite sequence satisfying*

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 9-14

$$(*)_t \equiv \qquad (\omega^{s+1}, X(s)) \in C, \quad 0 \leq s \leq t-1$$

where $X(s)$ denote the state of the Kohonen algorithm at time s starting from x, with stimuli sequence $\omega^1, \cdots, \omega^s$. Then, for any $X(t)$-centered open ball B, there exists an x-centered open ball $\tilde{B}$ and a sequence of ω^s-centered open balls $\mathbb{B}^s$, $s = 1, \cdots, t$ such that, for any $\tilde{x} \in \tilde{B}$ and any $\tilde{\omega}^s \in \mathbb{B}^s$, $s = 1, \cdots, t$, the corresponding vector $\tilde{X}(t)$ of the Kohonen algorithm at time t belongs to B.

A straightforward consequence of this lemma is that: $\mathbb{P}_x(X(t) \in B) > 0$.

Step 2: it remains to build a path going from $x \in \mathcal{O}$ into the interior of $\mathcal{O}^c$. It will be provided by an *ad hoc* sequence $(\omega^1, \cdots, \omega^t)$ satisfying assumption $(*)_t$. Lemma 2 will complete the proof. So, let $x \in \mathcal{O}$

defined by (see fig.2.1): $\begin{cases} x := (x_1, \cdots, x_n),\ x_i := (x_i^1, x_i^2), \\ x_i^2 := \frac{1}{2},\ x_1^1 < x_2^1 < \cdots < x_5^1 < \frac{a}{2} < 4a < x_6^1 < \cdots < x_n^1 \end{cases}$.

Now we consider 3 geometric values for the stimuli: $\omega_1 := (0, \frac{1}{2} - a)$, $\omega_2 := (a, \frac{1}{2} - a)$, $\omega_3 := (\frac{1}{2} + \rho, 2a + \rho)$ (see fig.2.1). Now put $\omega^1 = \cdots = \omega^{t_1} := \omega_1$, $\omega^{t_1+1} = \cdots = \omega^{t_1+t_2} := \omega_2$, $\omega^{t_1+t_2+1} = \cdots = \omega^{t_1+t_2+t_3} := \omega_3$. We set ρ, t_1, t_2, t_3 in such a way that at time $t_1 + t_2 + t_3$, $X(t_1 + t_2 + t_3)$ belongs to the interior of $\mathcal{O}^c$.

Notice that while $\omega^t = \omega_1$ the value of $i_0(\omega^t)$ is 1. So, let $\eta > 0$ (small). Set t_1 so that, picking up $\omega^t = \omega_1$ until time t_1, one has $|x_1(t_1) - \omega_1|_2$ and $|x_2(t_1) - \omega_1|_2 < \eta$. This is possible since $1 - \varepsilon > 0$ and the only modified components of x are x_1 and x_2. From now on, one may assume w.l.g. that $\eta < a\frac{\sqrt{5}-2}{2}$. While $t \in \{t_1+1, \cdots, t_1+t_2\}$, one sets $\omega^t := \omega_2$ and, since the previous homotheties have left unchanged the original direction of the mediatrix of $[x_1, x_2]$, the value of $i_0(\omega^t)$ is always 2. Carrying the process until time $t_1 + t_2$ yields:

$$|X_1(t_1 + t_2) - \omega_2|_2 < \eta,\ |X_2(t_1 + t_2) - \omega_2|_2 < \eta,\ |X_3(t_1 + t_2) - \omega_2|_2 < \eta.$$

Now set $\omega^t := \omega_3, t \in [t_1+t_2+1, \cdots t_1+t_2+t_3]$. If $\eta < \min(\rho, \frac{2}{5}\frac{a(a+2\rho)}{\sqrt{(2a+\rho)^2+\rho^2}+\sqrt{2}\sqrt{a^2+\rho^2}})$ – and $\rho < a\sqrt{2}$ so that ω_3 stays closer to x_3 than to x_6 – we check that $i_0(\omega^t) = 3$.

So, for large enough t_3: $\begin{array}{l} |X_1(t_1 + t_2 + t_3) - \omega_2|_2 < \eta,\ |X_2(t_1 + t_2 + t_3) - \omega_3|_2 < \eta, \\ |X_3(t_1 + t_2 + t_3) - \omega_3|_2 < \eta,\ |X_4(t_1 + t_2 + t_3) - \omega_3|_2 < \eta \end{array}$.

Again we notice that the mediatrix of $[x_2, x_3]$ and $[x_3, x_4]$ and $[x_2, x_4]$ keep the same direction during the "dragging" between times $t_1 + t_2 + 1$ and $t_1 + t_2 + t_3$. Thus we reach the final state showed in fig.2.2

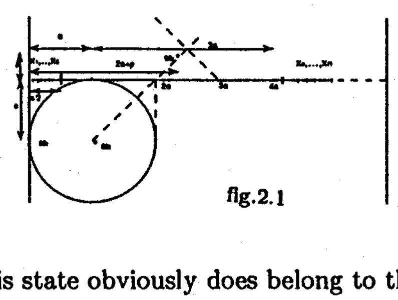
fig.2.1

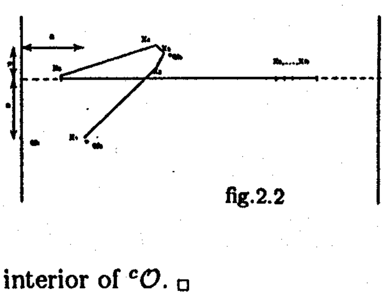
fig.2.2

This state obviously does belong to the interior of ${}^c\mathcal{O}$. □

The result still holds starting from other weights (see below).

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 9-14

4 Some simulation results

The Kohonen string: in fig.3 below we present 4 states of a $K(\varepsilon;\, n;\, 1,\, 2,\, 2)$. The black square indicates (approximately) the subset of $[0,1]^2$ in which the stimuli are drawn to obtain the next state. We took $\varepsilon := 0.01$ and $n := 30$.

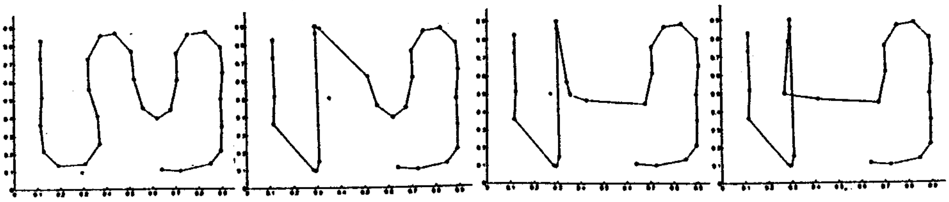

fig.3

The 8 neighbour Kohonen grid: the mathematical approach exposed in section 2 can be extended to the simulation of much more general cases, namely the $K(\varepsilon;\, n;\, 2,\, 2,\, 8)$: the idea is to first transform an initial organized state (*e.g.* a grid) in a string-like state (probably still organized) and then to apply the previous method. In fact this is very critical to realize. We display (see fig.4) a sequence of states obtained by such a technique.

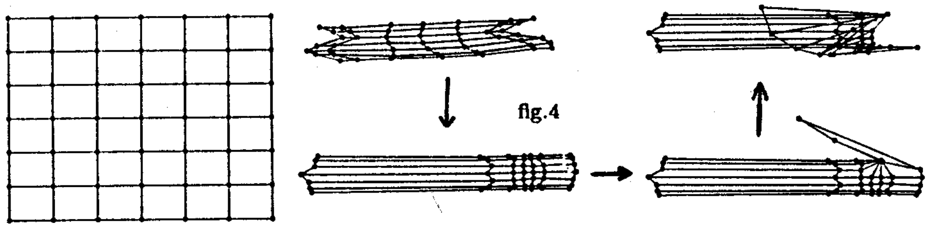

fig.4

Starting from a regular square grid we slowly make this grid thinner and thinner by drawing some stimuli in a thinner and thinner horizontal stripe centered on the line $y = 1/2$. Then, still working within a thin stripe around the line $y = 1/2$, we strech the pseudo-string obtained. To this end, we alternately pick up the stimuli at the right and left part of this small stripe. Now we make a "loop" with this streched pseudo-string, using a similar method to that described section 2. The final state is obviously not organized.

References

[1] C.Bouton, G.Pagès, Self-org. and convergence of the 1-dim. Kohonen algo. with non uniformly distributed stimuli, *Stoch. Proc. & their Appl.*, 47, 1993, pp.249-274.

[2] M. Cottrell, J.-C. Fort, Etude d'un algorithme d'auto-organisation, *Annales de l'Institut Henri Poincaré*, 23, 1987, n°1, pp.1-20.

[3] Th. Villmann, R. Der, Th. Martinez, A novel approach to measure the topology preservation of feature maps, Proc. of the ICANN 94 (Sorrento), Springer-Verlag, M. Marinaro & P.G. Morasso eds.,1994 pp.298-301.