

# DISCRIMINACIÓN DE ATRACTORES NOMINALES Y ESPURIOS EN REDES NEURONALES ASIMÉTRICAS

Jorge A. Horas y Edgar A. Bea

Universidad Nacional de San Luis – Departamento de Física – Instituto de Matemática Aplicada (IMASL)  
Ejército de los Andes 950 – (5700) San Luis – Argentina  
e-mail: jhoras@unsl.edu.ar, ebea@unsl.edu.ar

Se aplica un proceso de desaprendizaje sobre las conexiones de una red asimétrica. Este no solo establece un mejoramiento general en la performance de recuperación sino que permite la separación a priori del resultado final de la relajación. Se discrimina así convenientemente entre atractores nominales, que son las memorias grabadas, y los atractores espurios. Se estudia también la influencia del proceso de desaprendizaje aplicado sobre diferentes tipos de patrones.

An unlearning process is applied over connections of an asymmetric neural network. The form reach a global improvement in performance retrieval and also allow the a priori separation of the relaxation final result. So the stored memories in the spurious attractors are clearly distinguished. We also study the unlearning influence on some patterns forms.

## Introducción

Una de las capacidades mas importantes y notables del cerebro es la de actuar como memoria asociativa. Existe cierto tipo de redes neuronales artificiales que tratan de imitar este comportamiento, con diverso grado de éxito.

En este trabajo, continuando con contribuciones anteriores, se usa un proceso de desaprendizaje sobre las conexiones sinápticas. La novedad aquí es que tal proceso es aplicado a una red totalmente asimétrica en sus conexiones, que permite grabar y recuperar secuencias temporales de patrones.

Según se mostrará, el procedimiento de desaprendizaje logra también en este caso los ya conocidos beneficios de:

- Aumentar la capacidad de almacenamiento;
- Incrementar el tamaño de las cuencas de atracción y
- Reducir el número de atractores espurios.

A los cuales debe adicionarse:

- Permitir la discriminación entre atractores nominales y atractores espurios.
- Recuperar secuencias de patrones correlacionados.

## El Modelo

El modelo consiste en una red con conexiones asimétricas con la misma arquitectura y ley dinámica del modelo de Hopfield [1] al cual se le aplica un procedimiento de desaprendizaje [2,3].

El proceso de memorización consiste en hacer que los patrones  $\xi^{(\mu, \nu)}$  de una secuencia o ciclo  $\mu$  sea idéntica a uno de los ciclos límite de la dinámica [4].

Los patrones a ser memorizados (ver fig. 1) son usados para formar la matriz de acoplamientos sinápticos entre neuronas

$$J_{ij}^{(a)} = \sum_{\mu=1}^K \sum_{\nu=1}^L \xi_i^{(\mu, \nu+1)} \xi_j^{(\mu, \nu)} \quad i \neq j \quad (1)$$

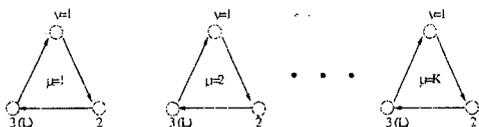


Figura 1: Ejemplo de ciclos bajo la condición  $\xi^{(\mu, L+1)} = \xi^{(\mu, 1)}$ . Cada producto entre pares se representa mediante una flecha.

El sistema evoluciona según una dinámica síncrona

$$S_i(t+1) = \text{sgn} \left[ \sum_{j=1}^N J_{ij}^{(a)} S_j(t) \right] \quad (2)$$

El procedimiento de desaprendizaje puede ser fácilmente generalizado para el caso de ciclos límite:

- 1) *Random Shooting*: el sistema es iniciado al azar.
- 2) *Relajación*: el sistema evoluciona iterando  $t$  veces la ec. (2), y alcanza una configuración final  $\chi^{(d, t)}$ .
- 3) *Desaprendizaje*: la matriz de conexiones sinápticas es modificada según

$$J_{ij}^{(d)} = J_{ij}^{(d-1)} - \frac{\epsilon}{N} \chi_i^{(d, t)} \chi_j^{(d, t-1)} \quad (3)$$

donde  $\epsilon$  es el parámetro de desaprendizaje y  $d$  denota el número de iteración del proceso, que se completa al iterar  $D$  veces.

## Simulaciones

Para estudiar la performance de recuperación se analizan la capacidad de almacenamiento, las características de las cuencas de atracción, el tiempo de convergencia y la distribución de atractores espurios.

Se muestran también resultados de recuperación de secuencias de patrones correlacionados.

En cada uno de los ciclos atractores nominales se hacen pruebas de recuperación iniciando la red con un

overlap especificado  $m^{(\mu, \nu)}(0) = \frac{1}{N} \sum_{i=1}^N S_i(0) \xi_i^{(\mu, \nu)} = m_0$ ,

luego el sistema evoluciona hasta alcanzar un ciclo límite estacionario en un tiempo  $t_r$ , o hasta alcanzar una cantidad preestablecida  $t$  de pasos de tiempo. A continuación se determina:

- 1) El *overlap de recuperación* ( $m_r$ ), de forma tal que si se recupera perfectamente el ciclo  $m_r = 1$

$$m_r = \left\langle m^{(\mu)} \right\rangle_{\text{patrones}(L)} \quad (4)$$

- 2) El valor medio del overlap de recuperación ( $\bar{m}_r$ ) es

$$\bar{m}_r = \left\langle \left\langle m_r \right\rangle_{\text{muestras}} \right\rangle_{\text{ciclos}} \quad (5)$$

Los resultados mostrados corresponden a una red de 200 neuronas en los cuales se han almacenado 40 patrones organizados en ciclos. Las barras de error en las curvas, representan la desviación standard del overlap de recuperación.

### Resultados

Se analizan varias formas de organización en ciclos de los patrones usando  $L = fr \cdot p$  (el modelo de autocorrelación  $fr=L=1$ ).

#### a. Capacidad de almacenamiento

Las curvas de capacidad se efectúan iniciando las pruebas de recuperación desde los ciclos nominales. En la figura 2 se muestra el mejoramiento obtenido al aplicar desaprendizaje.

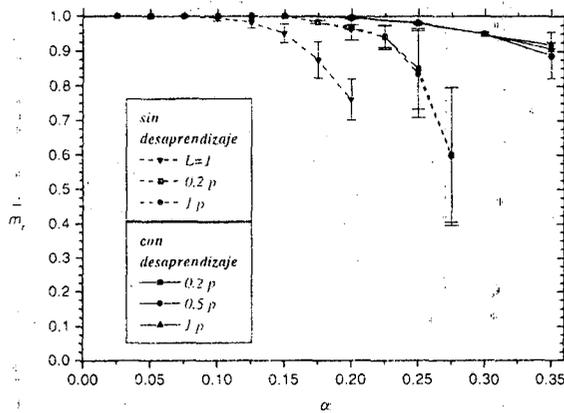


Figura 2: overlap medio de recuperación ( $\bar{m}_r$ ) vs. capacidad ( $\alpha = \frac{KL}{N}$ ). Curvas de capacidad sin y con desaprendizaje. Estadística: 10 redes.

#### b. Incremento del tamaño de las cuencas de atracción

Una estimación del tamaño de las cuencas de atracción y su incremento al aplicar el proceso de desaprendizaje se muestra en la figura 4. En ella se grafica el overlap medio de recuperación, iniciando la red desde posiciones, dadas por  $m_0$ , cada vez más alejadas del patrón nominal ( $m_0=1$ ).

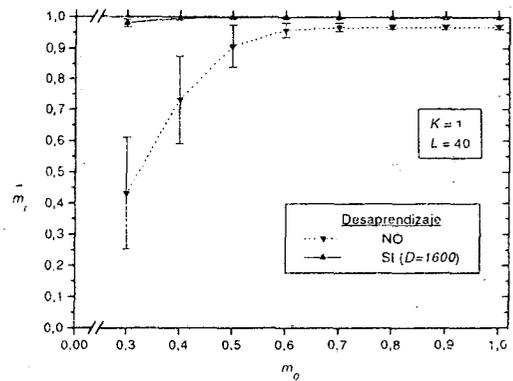
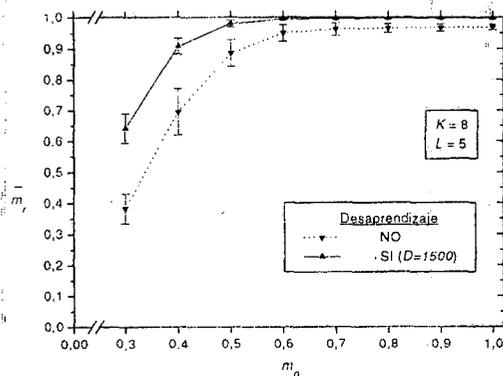


Figura 3: overlap medio de recuperación ( $\bar{m}_r$ ) vs. overlap inicial ( $m_0$ ), 40 patrones, 20 redes y un muestreo de 200.

#### c. Reducción del número de atractores espurios

Con el objeto de mostrar la notable reducción en el número de atractores espurios, al aplicar desaprendizaje se muestran en la figura 5 histogramas de overlap de recuperación donde se muestran con columnas los atractores. Se observa en esta figura no solo una clara separación entre atractores nominales (alrededor de  $m_0=1$ ) y espurios (por debajo de  $m_0=0.4$ ) sino la drástica disminución de estos últimos.

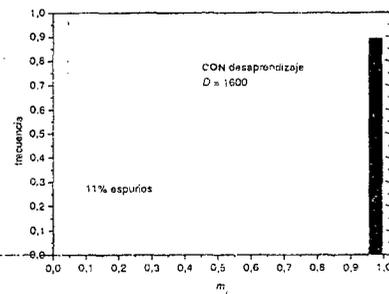
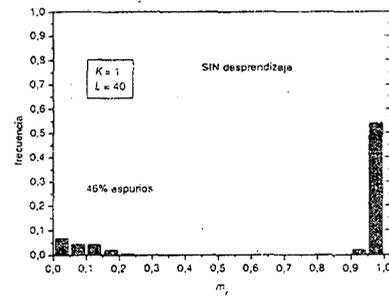


Figura 4: histogramas de overlaps de recuperación, 40 patrones. La estadística es la que corresponde a la fig. 3. El overlap inicial es 0.4.

#### d. Distinción entre patrones y espurios

En la figura 6 se muestran histogramas de tiempo de relajación para 40 patrones. La flecha en cada gráfica señala el tiempo discriminante ( $T_D$ ) que permite separar la convergencia a patrones o espurios respectivamente.

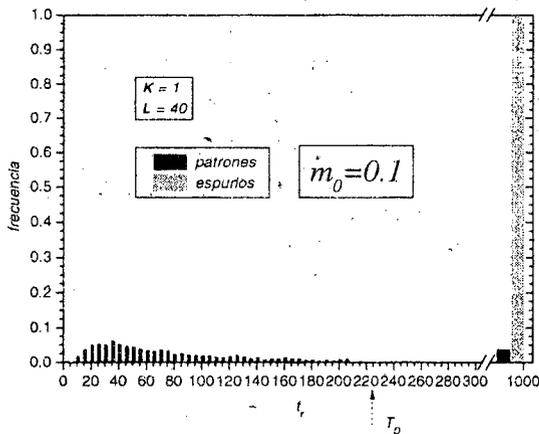
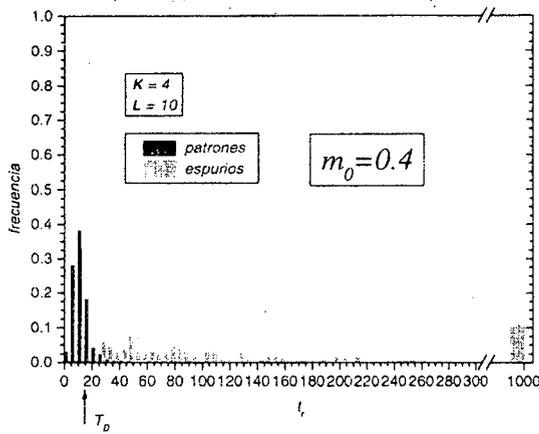


Figura 6: estadística: 20 redes y un muestreo de 200.

### e. Recuperación de secuencias

Se muestra la recuperación de secuencias de patrones correlacionados. Esta tarea es extremadamente difícil con este tipo de patrones donde  $prob(+1) \neq prob(-1)$ . Para mostrar este resultado se usan palabras para su fácil visualización.

#### a. PATRONES GRABADOS EN 1 CICLO



#### Recuperación

SIN DESAPRENDIZAJE

INICIO



CON DESAPRENDIZAJE

INICIO



#### b. PATRONES GRABADOS EN 2 CICLOS



#### Recuperación

SIN DESAPRENDIZAJE

INICIO



INICIO



CON DESAPRENDIZAJE

INICIO



INICIO



#### Conclusiones

En este trabajo se ha mostrado:

- La forma de aplicar un procedimiento de desaprendizaje que actúa sobre una red con conexiones sinápticas asimétricas.
- Que el procedimiento de desaprendizaje aplicado, permite obtener resultados positivos en la performance de recuperación de patrones organizados en ciclos.
- Que el número de atractores espurios se reduce efectivamente también en redes asimétricas mediante la aplicación de dicho proceso.
- Que es posible discernir, a través del tiempo de relajación, si la respuesta de la red corresponde a un atractor nominal o espurio.
- Que pueden recuperarse secuencias de patrones correlacionados aún partiendo de versiones ruidosas.

#### Referencias

- [1] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982), and **81**, 3088 (1984).
- [2] J. van Hemmen, L.B. Ioffe and R. Kuhn, Physica A **163** (1990), 386.
- [3] J.A. Horas and P.M. Pasinetti, J. Phys. A: Math. Gen. **31** (1998), L463.
- [4] Y. Mori, P. Davis and S. Nara, J. Phys. A: Math. Gen. **22** (1989), L525.