

# Performance como Memoria Asociativa de Redes Neuronales Interactivas

J. A. Horas\* y P. M. Pasinetti.

Universidad Nacional de San Luis. Facultad de Ciencias Fisico, Matematicas y Naturales.  
Departamento de Fisica. Instituto de Matematica Aplicada (IMASL).  
Ejército de los Andes 950 - (5700) San Luis - Argentina.  
e-mail: jhoras@unsl.edu.ar

Estudiamos las propiedades como memoria asociativa de dos redes neuronales que interactúan entre sí formando un sistema. Ambas subredes son del tipo Hopfield. Cada región de este sistema tiene alta conectividad mientras que la interacción entre ellas es de baja conectividad. El aprendizaje para todas las conexiones, tanto inter como intra-subred es Hebbiano.

La intensidad de las conexiones es modificada mediante un procedimiento de desaprendizaje también Hebbiano, estudiándose su influencia sobre la recuperación de los patrones globales. La aplicación de este procedimiento y la dependencia con parámetros del sistema son analizadas para diversos casos.

We study the properties as associative memory of two neural networks interacting to form a system. Both sub-networks are of the Hopfield type. Each region of this system has high connectivity while the interaction between them is of a low connectivity. The learning for all the connexions, inter as well as intra-subnetwork is Hebbian. The connexions' intensity is modified through an unlearning procedure, also Hebbian, its influence on the recovery of the global patterns is analyzed. The application of this procedure and the dependance on system parameters are analyzed for different cases.

## Introducción

Una muy importante capacidad del cerebro humano es su eficiente comportamiento como memoria asociativa. Esta propiedad es estudiada con intensidad<sup>1,2,3</sup>. En este trabajo prestamos atención a dos aspectos con fuerte motivación biológica:

a) Hay clara evidencia de que la información asociada a un evento es grabada en diferentes áreas del cortex. La memoria global de ese evento puede ser recuperada aun a partir de información difusa o incompleta de solo parte de él<sup>4</sup>.

b) Se cree fuertemente que el proposito del sueño con rápidos movimientos oculares (REM) es minimizar o aun eliminar ciertos modos indeseables en la red neuronal de la corteza cerebral<sup>5</sup>.

En este trabajo intentamos analizar estos dos aspectos usando un modelo simplificado. Este no necesariamente describe el proceso real que ocurre en el cerebro.

Deseamos entonces analizar solamente un modelo donde las memorias globales son grabadas en diferentes regiones de la red, aplicando también un proceso de desaprendizaje tanto en la red global como en alguna de sus partes. Este proceso mejora la performance de la red facilitando el acceso a las memorias reales y minimizando la cantidad de memorias espúreas.

Es razonable asumir que una red neuronal puede estar constituida de regiones interactuantes con distinta conectividad mas que como un sistema con conectividad homogénea como se considera usualmente.

En la figura 1 mostramos esquemáticamente la red global a analizar conformada por:

Red total = subred a  
+ subred b  
+ conexiones ab

## El modelo

La Red total consiste de N neurodos dispuestos en subredes a y b. Solo por simplicidad consideramos igual numero de neurodos en cada subred y que estas son solo dos.

Las subredes son tratadas como dos redes tipo Hopfield: cada neurodo esta conectado con todos los otros dentro de la misma subred y las conexiones entre las subredes solo conectan un neurodo de una con solo uno de la otra.

La red global mostrada en la figura 1 se puede ver entonces como una red de Hopfield usual, en la que se ha incluido un tipo particular de dilución solo en las conexiones entre una y otra subred. El hamiltoniano general que describe el modelo adoptado puede ser escrito como

$$H = -\frac{1}{2} \sum_{i,j=1}^{N/2} J_{ij}^a s_i s_j - \frac{1}{2} \sum_{i,j=1}^{N/2} J_{ij}^b \sigma_i \sigma_j - \frac{1}{2} \sum_{i=1}^{N/2} J_{ii}^{ab} s_i \sigma_i \quad (1)$$

\* Investigador del CONICET.

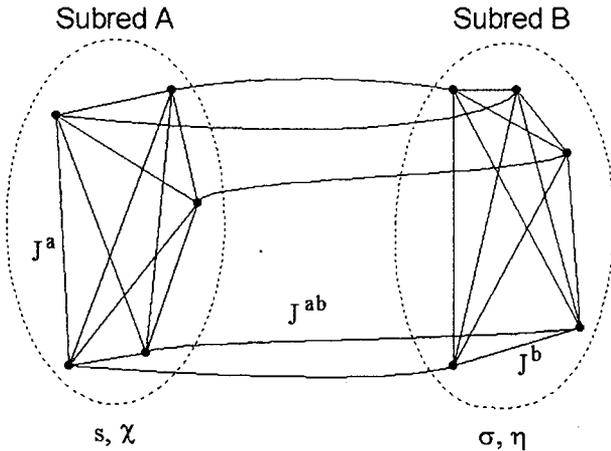


Figura 1: esquema de la red global y cada una de las subredes.

Los acoplamientos sinápticos se determinan siguiendo la regla de Hebb según

$$J_{ij}^a = \frac{1}{N/2} \sum_{\mu=1}^p \chi_i^\mu \chi_j^\mu \quad (i \neq j) \quad (2)$$

$$J_{ij}^b = \frac{1}{N/2} \sum_{\mu=1}^p \eta_i^\mu \eta_j^\mu \quad (i \neq j) \quad (3)$$

$$J_{ij}^{ab} = \frac{J_0}{N/2} \sum_{\mu=1}^p \chi_i^\mu \eta_j^\mu \quad (i \neq j) \quad (4)$$

Se estudia el caso de "temperatura cero" y la actualización es asíncrona. Los neurodos son modelados por espines tipo Ising con  $s_i = +1$  ó  $s_i = -1$  si el neurodo  $i$ -ésimo dispara o no. Acordemente la dinámica del sistema esta descrita por

$$s_i(t+1) = \text{sgn}(h_i^a(t)) \quad (5)$$

$$\sigma_i(t+1) = \text{sgn}(h_i^b(t)) \quad (6)$$

donde los campos locales  $h_i^a(t)$  ó  $h_i^b(t)$  estan dados por

$$h_i^a = \sum_{j=1}^{N/2} J_{ij}^a s_j(t) + J_{ii}^{ab} \sigma_i(t) \quad (7)$$

$$h_i^b = \sum_{j=1}^{N/2} J_{ij}^b \sigma_j(t) + J_{ii}^{ab} s_i(t) \quad (8)$$

El modelo se completa con el proceso de desaprendizaje que se realiza tanto en la red global como en las subredes, según corresponda, de acuerdo al siguiente proceso:

1) Inicialización al azar (random shooting).

2) Relajación: el sistema debe relajar puesto que las conexiones son simétricas, a una configuración estacionaria denotada por  $X_i$ .

3) Desaprendizaje: despues de un sueño las conexiones son actualizadas según

$$J_{ij}^z \rightarrow J_{ij}^z - \frac{\epsilon}{N} X_i X_j \quad (9)$$

donde el supraíndice  $z$  indica en que región se realiza el procedimiento (a; b; a ↔ b; ab).

El procedimiento es local, iterativo y puramente hebbiano. Existe un número óptimo de veces,  $D_{op}$ , en que debe repetirse este proceso de desaprendizaje. Puede obtenerse <sup>6</sup>

$$D_{op} = \epsilon^{-1} \left( \frac{c_1}{N} + c_2 \right) p \quad (10)$$

donde  $c_1 = 22 \pm 10$  y  $c_2 = 0.56 \pm 0.08$ .

### Resultados numéricos

Interesa contar con un parámetro que mida el overlap entre dos subpatrones, en el rango desde subpatrones identicos hasta totalmente no correlacionados. Este tipo de correlación puede escribirse como

$$q^\mu = \frac{1}{N/2} \sum_{i=1}^{N/2} \chi_i^\mu \eta_i^\mu \quad (11)$$

cada componente  $\chi_i^\mu$  es elegida al azar y puede tomar valores iguales a +1 o -1 con la misma probabilidad (no-biased). También se consideran casos (biased) donde la probabilidad no es la misma.

Para el caso de patrones "no-biased" es conveniente introducir un overlap de recuperación  $m^\mu$ , que es una medida de la correlación entre la configuración final alcanzada por el sistema y el patrón global grabado  $\mu$ . Esto puede escribirse

$$m^\mu = \frac{1}{2} (m_1^\mu + m_2^\mu) \quad (12)$$

El overlap de recuperación mide la capacidad de la red para reconocer un patrón dado  $\mu$ . Para cada subred esta dado por

$$m_1^\mu = \frac{1}{N/2} \left\langle \sum_{i=1}^{N/2} \chi_i^\mu s_i(t \rightarrow \infty) \right\rangle \quad (13)$$

$$m_2^\mu = \frac{1}{N/2} \left\langle \sum_{i=1}^{N/2} \eta_i^\mu \sigma_i(t \rightarrow \infty) \right\rangle \quad (14)$$

donde los corchetes denotan promedios sobre condiciones iniciales y sobre distintas redes.

A fin de analizar el comportamiento del sistema, presentamos resultados numéricos para ambos casos:

**Caso 1:** patrones al azar con la misma probabilidad de tomar los valores +1 o -1.

**Capacidad:** se han realizado diversas pruebas cuyos resultados se muestran en las figuras 3, 4 y 5 en donde se grafica el overlap  $m$  vs.  $\alpha$  ( $=p/N$ ) mostrando la capacidad para las distintas condiciones de operación.

**Tamaño de cuencas de atracción:** se muestra cualitativamente la variación del tamaño de las cuencas de atracción para diversos casos según se explicita en las figuras 6, 7 y 8 donde se grafica el overlap final  $m$  vs. el inicial  $m_0$ .

**Caso 2:** patrones con distinta probabilidad de tomar los valores +1 o -1.

### Conclusiones y perspectivas

El efecto del desaprendizaje es apreciable en los casos estudiados tanto en lo que hace a la capacidad como al tamaño de las cuencas de atracción.

El efecto del procedimiento de desaprendizaje decrece según se aplique a *todas las interacciones sinápticas de la red global*, a las de *una de las subredes* y las interacciones con la otra y por último solo a las interacciones entre ellas.

La figura 2 muestra resultados, especialmente en lo que hace a capacidad, para el caso de patrones altamente correlacionados.

Los resultados obtenidos impulsan a profundizar este trabajo en al menos dos direcciones:

1) Arquitectura: es interesante modificar las conexiones intersubred no usando necesariamente aprendizaje hebbiano.

2) Desaprendizaje: es importante cuantificar el número óptimo de sueños a aplicar y su dependencia paramétrica de una nueva manera.

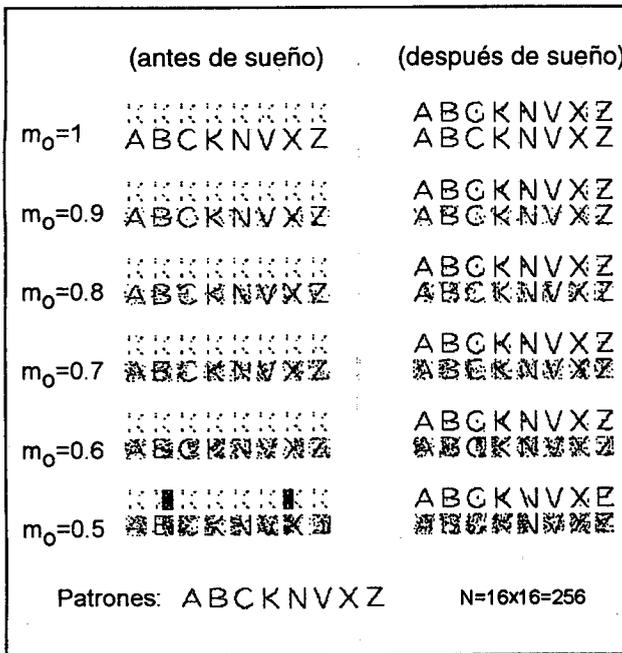


Figura 2: resultados para el caso de patrones biased.

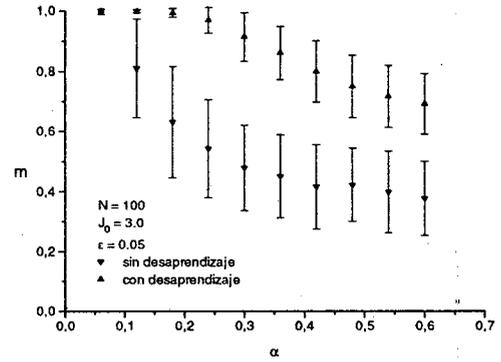


Figura 3: desaprendizaje en ab.

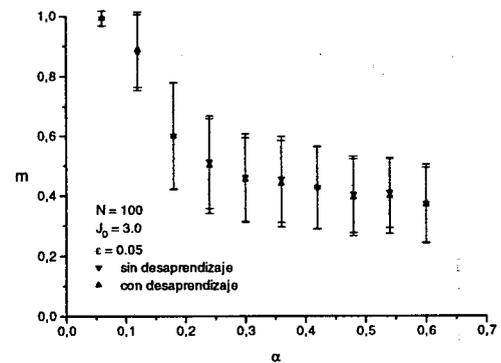


Figura 4: desaprendizaje en inter-ab.

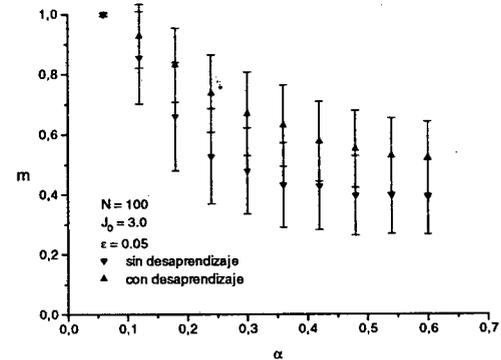


Figura 5: desaprendizaje en inter-ab + a.

Curvas de capacidad ( $m$  vs.  $\alpha=p/N$ ) para los distintos modos de operación. Cada punto es un promedio sobre al menos 5 redes, mostrándose también la dispersión de cada uno de ellos.

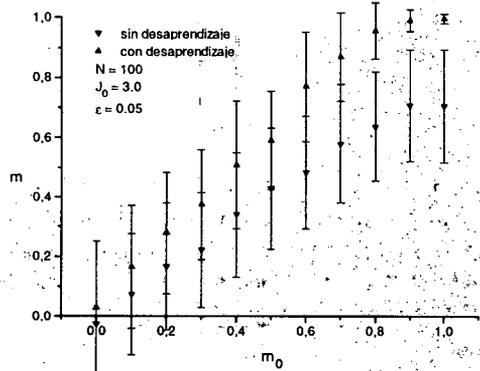


Figura 6: desprendizaje en ab.

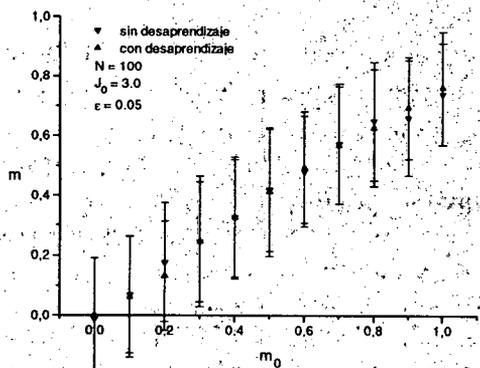


Figura 7: desprendizaje en inter-ab.

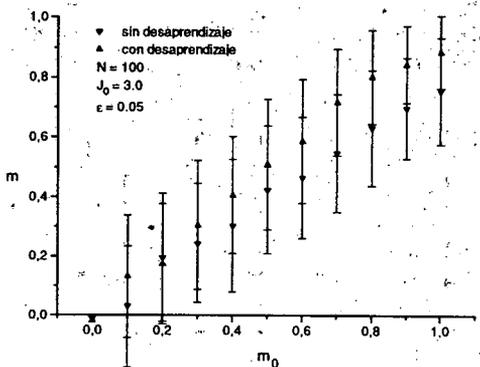


Figura 8: desprendizaje en inter-ab + a.

Variación del tamaño de las cuencas de atracción según el modo de operación (los puntos son un promedio sobre al menos 5 redes).

### Referencias

- 1 - J.J.Hopfield: Proc. Acad. USA, **79**, 2554 (1982).
- 2 - J.Hertz, A.Krogh and R.G.Palmer, *Introduction to the theory of Neural Computation* (Addison-Wesley, New York, 1991).
- 3 - Monoranjan P. Singh, Zhang Chengxiang and Chandan Dasgupta, Phys. Rev. E **52**, 1995, 5261.
- 4 - Márcia M. Ochi and O.L.T. de Menezés, Int. Journal of Neural Systems, Vol. 3, No. 4 (1992) 389-393.
- 5 - F.Crick and G.Mitchison, (1983) Nature, **304**, 11.
- 6 - E.Domany, J.L.Hemmen and K.Shulten, *Models of Neural Networks*, Springer-Verlang, 1st. Edition, 1991.