Influencia de un Procedimiento de Desaprendizaje en Memorias de Corto Término

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

Universidad Nacional de San Luis. Facultad de Ciencias Físico-Matemáticas y Naturales.
Departamento de Física. Instituto de Matemática Aplicada San Luis (IMASL).
Ejército de los Andes 950 - (5700) San Luis - Argentina.
e-mail: jhoras@unsl.edu.ar

Se estudia la capacidad como memoria asociativa de una red neuronal tipo Hopfield que es alimentada indefinidamente con nuevos patrones (*aprendizaje dinámico*), en la cual las conexiones sinápticas son modificadas mediante un procedimiento de desaprendizaje. De este modo se obtiene una "working memory" o memoria de corto término.
Se ha realizado un estudio numérico de la capacidad dinámica tanto para el caso de "recency" (permanencia de los patrones mas recientes) como para el de "primacy" (permanencia de los mas antiguos), mostrándose resultados en ambos casos, para distintos modos de operación.

We study the capacity as associative memory of a Hopfield-like neural network, which is continuously feeded with new patterns (dinamic learning). The sinaptic weights are modificated throught an unlearning process obtaining a "working memory" or short term memory.
A numeric study of the dinamic capacity was done in both the "primacy" case (staying the older patterns) and the "recency" case (staying the more fresh patterns), showing results for different operation modes.

I. Introducción

Redes de neuronas formales proveen modelos para memorias asociativas y ha habido mucho progreso [1,2], tanto analítico como numérico, especialmente en el modelo de Hopfield.

El algoritmo de aprendizaje usado por Hopfield tiene un principal inconveniente: uno no puede alimentar a la red indefinidamente puesto que si el número q de patrones almacenados cruza un límite bien definido, $q_c = \alpha_{cH} N$, donde N es el número de neurodos completamente conectados en la red y $\alpha_{cH} \approx 0.14$, entonces la memoria colapsa abruptamente y ya no es posible recuperar la información almacenada. Esto puede evitarse si no se permite que los pesos sinápticos crezcan mas allá de ciertos límites: $|J_{ij}| \leq A$. Entonces los patrones aprendidos mas recientemente van borrando gradualmente a los mas antiguos, y la memoria se convierte en una especie de "pila" en donde los patrones "frescos" están arriba y los "viejos" cada vez mas deteriorados están abajo. Una memoria organizada de esta manera es llamada palimpsest [3]. El precio a pagar es que, si bien puede evitarse la catástrofe de la memoria al alimentarla indefinidamente, se produce una fuerte disminución en la capacidad dinámica: $\alpha_{cD} \approx 0.05 \ll \alpha_{cH} \approx 0.14$.

El objetivo central de este trabajo es mostrar que el procedimiento de desaprendizaje, ya aplicado en el caso de "memorias estáticas" [4], es también útil en memorias dinámicas.

Se muestran diversas simulaciones en que queda establecido que el procedimiento de desaprendizaje sirve efectivamente para producir la progresiva degradación de los patrones viejos, y lograr mantener así un cierto número de patrones según se estudie el caso "recency" (permanencia de los patrones mas recientes) o el "primacy" (permanencia de los mas antiguos).

II. Arquitectura y Procedimiento

La arquitectura básica empleada, que se muestra en la figura 1, es:

RH: consiste en una red de Hopfield en la que cada nodo se conecta con todos los demás, que se opera asíncronamente y sigue la dinámica de Glauber a temperatura cero:

$$h_i(t) = \sum_{j=1}^{n} J_{ij} s_j(t), \qquad s_i(t+1) = \text{signo}\left(h_i(t)\right)$$

P: una colección de N perceptrones independientes, embebidos en **RH**

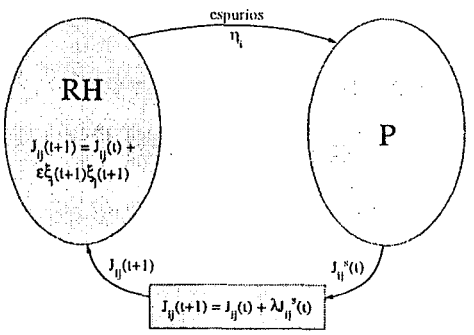

Figura 1: arquitectura empleada, donde t es el tiempo (discreto) que se incrementa en 1 por cada ciclo completo en el diagrama.

Como resultado de la anterior arquitectura existen dos *fases* que son repetidas indefinidamente. La primera, de un solo paso, que consiste en grabar, mediante la regla de

* Autor a quién debe dirigirse la correspondencia.

Hebb, un nuevo patrón a memorizar en la red de Hopfield según el procedimiento estandard. La segunda, el desaprendizaje que consiste en modificar las conexiones sinápticas de la red de Hopfield según la regla del perceptrón inversa.

El procedimiento usado (que es iterativo y local) es:

1) grabación del siguiente patrón ξ_i (i=1..N) (al azar entre −1 y 1) por medio de la regla de Hebb en **RH**:

$$J_{ij}(t+1) = J_{ij}(t) + \varepsilon \xi_i(t+1)\xi_j(t+1),$$

2) obtención de estados metaestables espurios η_i (i=1..N) en **RH**,

3) desestabilización de los estados espurios con los N perceptrones independientes entrenados por la regla del perceptron inverso. Las nuevas conexiones $J^S_{ij}(t)$ a que esto da lugar, son usadas para corregir las conexiones de **RH** según el esquema: $J_{ij}(t+1) = J_{ij}(t) + \lambda J^S_{ij}(t)$,

4) iterar en el punto 1).

III. Simulaciones y Resultados

Se han realizado diversas pruebas de capacidad dinámica, mostrándose los resultados en las fig. 2, 3 y 4.

Los parámetros estudiados han sido λ y ε. También fueron estudiados el número de veces que fue aplicado el procedimiento de desaprendizaje (entre patrones grabados) y el número de patrones grabados antes de aplicar el procedimiento por primera vez.

La figura 2 muestra la capacidad dinámica (número de patrones con m > 0.95) vs. la "intensidad" de grabación ε, para dos valores de la intensidad del procedimiento de desaprendizaje, λ.

En la figura 3 se muestra con tonos de grises, para cada instante de tiempo (ordenada), el overlap $m\left(=1/N\sum_{i=1}^{N}\xi_i s_i\right)$ vs. la "antigüedad" (absisa), mostrando la capacidad dinámica para distintas condiciones de operación. s_i es el estado de la red después de relajar a partir de ξ_i.

La figura 4 es similar a la 3 pero correspondiente ahora a los pesos $J^S_{ij}(t)$ en iguales condiciones que lo mostrado para $J_{ij}(t)$. Se observa en estos casos que el overlap máximo es alcanzado para una antigüedad del orden de la capacidad dinámica obtenida. Obsérvese que el signo del parámetro λ determina si se mantienen los últimos o los primeros patrones grabados. Se puede notar también que en t=50 el estado estacionario ya es alcanzado para prácticamente todos los casos.

IV. Conclusiones

De este trabajo se puede concluir:

1) El procedimiento de desaprendizaje es apto también para producir memorias de corto término o palimpsesticas.

2) Se ha mostrado que mediante el procedimiento de desaprendizaje se obtienen resultados similares a los de otros modelos para la capacidad crítica en el caso recency y mejoras para el caso primacy.

3) La máxima capacidad dinámica obtenida, tanto para recency como para primacy, se obtiene para distinto número de veces que se aplica el procedimiento de desaprendizaje, y parece no depender del incremento de éste.

Referencias

[1] J. Hertz, A. Krogh and R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, 1991.

[2] E. Domany, J. L. van Hemmen and K. Shulten, *Models of Neural Networks*, Springer-Verlang, 1st Edition, 1991.

[3] M. Mezard, J. P. Nadal and G. Toulouse, J. Physique **47** (1986) 1457-1462.

[4] J. Horas, P. M. Pasinetti, Anales AFA, Vol. 8, 1996 (en prensa).

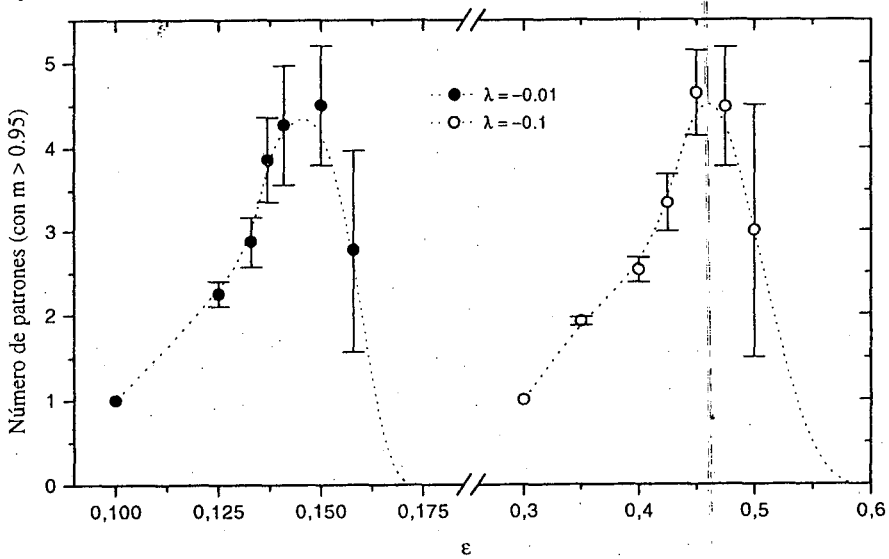

Figura 2: Capacidad dinámica vs. "intensidad" de grabación (ε). Cada punto es un promedio sobre al menos 10 redes mostrándose la dispersión en cada uno de ellos.

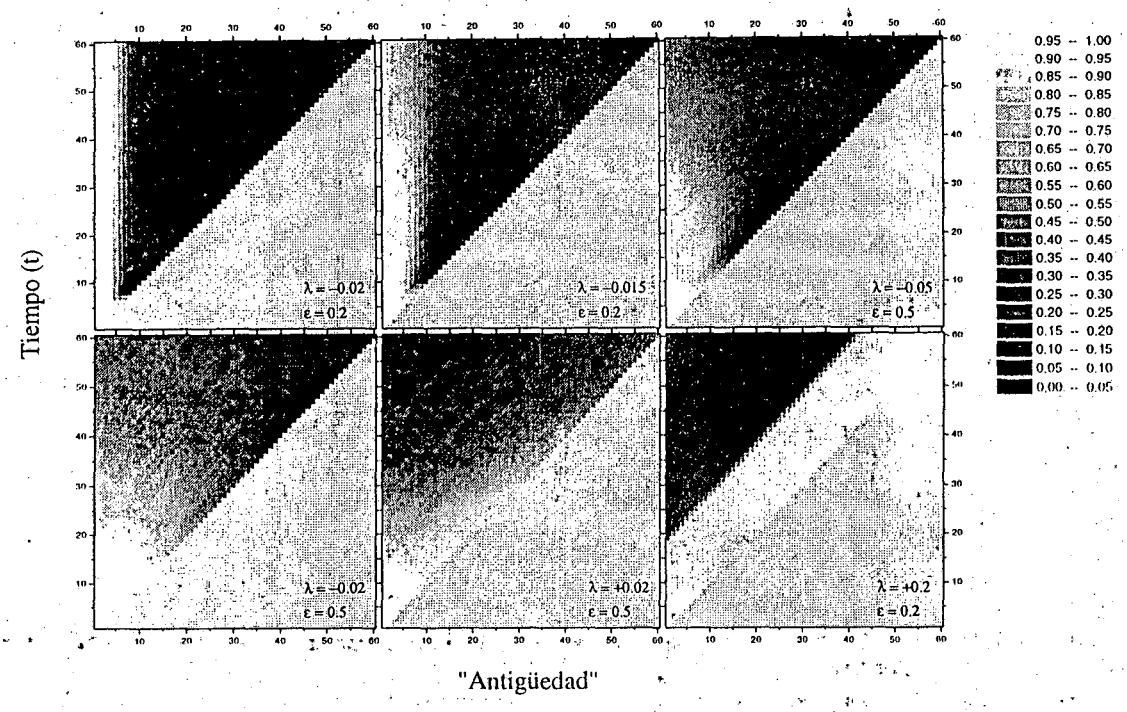

Figura 3: Capacidad dinámica. "Antigüedad" de un patrón se entiende como el tiempo transcurrido desde que fue grabado. Cada gráfica es el promedio sobre 100 redes.

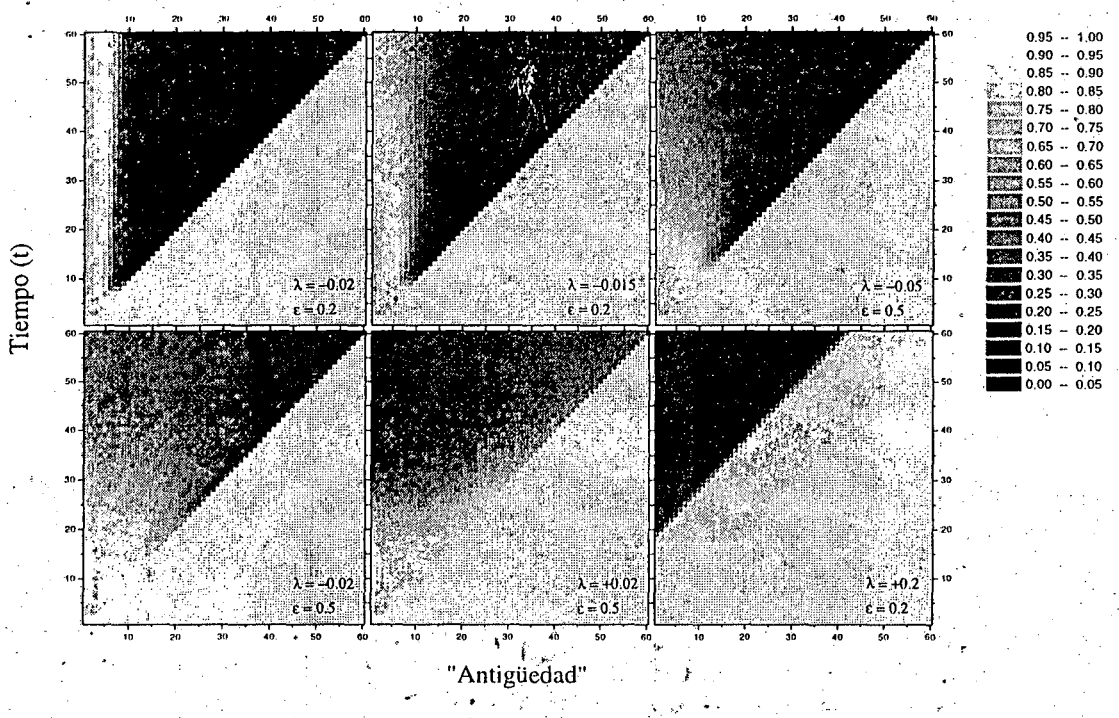

Figura 4: ídem a figura 2, pero para J^s.