Influencia de un Procedimiento de Desaprendizaje sobre la performance como Memoria Asociativa de Redes Neuronales

J. A. Horas y P. M. Pasinetti

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

Se estudia la performance de una red neuronal tipo Hopfield en la cual las conexiones sinápticas se modifican también mediante un procedimiento de desaprendizaje. Se estudia la influencia de este sobre la capacidad y sobre el tamaño de las cuencas de atracción.
Se obtiene el número de óptimo de veces en que debe ser aplicado el procedimiento de desaprendizaje y su dependencia con parámetros relevantes.

We study the performance of a Hopfield-like neural network with sinaptic strengths that are modified through an unlearning procedure. We analize its influence on the capacity and the size of the attraction basis.
We obtain the optimum number or times that the unlearning procedure must be applied and its dependence with the relevant parameters.

Introducción

Una de las capacidades mas notables del cerebro humano es la de actuar como memoria direccionable por el contenido (CAM). Ello motiva al intenso estudio de redes neuronales artificiales que se comportan como memorias asociativas.

Un aspecto importante consiste en identificar y entender el mecanismo mediante el cual la red neuronal de la corteza cerebral es capaz de mantener y aun incrementar un importante numero de recuerdos en forma confiable. A este respecto se destaca la hipótesis formulada algún tiempo atrás por Crick y Mitchison [1] en el sentido de que el propósito del sueño de rápidos movimientos oculares (REM) es minimizar ciertos modos indeseables (patrones espurios) en la red neuronal de la corteza cerebral. A partir de allí numerosos estudios en redes neuronales artificiales (RNA) han tratado de validar tal hipótesis [2,3]. Todos ellos en la dirección de que la modificación de la intensidad de las sinápsis se realiza mediante un proceso denominado de "desaprendizaje" [2,3], en que se hace hincapié en el número de veces en que debe aplicarse dicho proceso, determinándose su óptimo.

En este trabajo, a diferencia de lo anterior, la modificación de las conexiones sinápticas se lleva a cabo a través de la desestabilización de estados metaestables espurios haciendo hincapié ahora en la determinación del número óptimo de espurios a desestabilizar.

El objetivo del presente trabajo es:

a) aplicar este procedimiento dando un nuevo contexto para la hipótesis de Crick y Mitchison,

b) describir su funcionamiento y la determinación de los parámetros fundamentales,

c) mostrar que se logran importantes mejoras en la performance de la memoria asociativa.

Respecto a esto último, los criterios mas importantes que se usan para comparar la eficiencia de redes neuronales artificiales usadas como CAM, son:

1) capacidad: la RNA debe ser capaz de grabar la mayor cantidad de patrones,

2) tamaño de cuencas de atracción: la red neuronal debe ser tolerante a errores,

3) la cantidad de estados metaestables espurios debe minimizarse al igual que la presencia de ciclos límite,

4) la recuperación de las memorias debe ser la mas rápida posible.

Mostraremos que el procedimiento propuesto de desestabilización de estados metaestables espurios importa una mejora de los cuatro puntos anteriores.

Arquitectura y Procedimiento

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

RH: consiste en una red tipo 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:

$$h_i(t) \equiv \sum_{j=1}^{N} J_{ij} s_j(t) \qquad (1)$$

$$s_i(t+1) = \text{signo}\left(h_i(t)\right) \qquad (2)$$

P: una colección de N perceptrones independientes, embebidos en RH, que son entrenados por la regla del perceptrón inverso, cuya operación puede definirse algorítmicamente como:

1) comenzar con el conjunto de conexiones existentes en RH, $K_{ij}^{0} = J_{ij}$,

2) chequear con el espurio η_i (i=1..N), si la condición de estabilidad para el neurodo i es satisfecha:

$$\eta_i^\mu \sum_j K_{ij}\eta_j > c \qquad (3)$$

donde $c\,(\geq 0)$ es la cota de estabilidad. Mientras la condición (3) sea cierta, cambiar cada uno de los pesos de acuerdo a la regla:

$$K_{ij} \rightarrow K_{ij} - \varepsilon\eta_i^\mu\eta_j^\mu \qquad (i,j=1...N;\quad j\neq i) \qquad (4)$$

donde ε es el tamaño de la modificación hecha a los pesos en cada paso de perceptrón.

3) repetir 2) para cada neurodo.

Una descripción completa del procedimiento usado (que es iterativo y local) puede expresarse según:

a) grabación de patrones (al azar, entre -1 y +1) por medio de la regla de Hebb en RH,

b) obtención de estados metaestables espurios en RH, según el siguiente procedimiento: i) "random shooting": la red es iniciada en una configuración al azar, ii) relajación hacia un punto fijo: se le permite al sistema relajar hasta una configuración estacionaria, η_i (i=1..N).

c) desestabilización de los estados espurios con los N perceptrones independientes entrenados por la regla del perceptrón inversa. Las nuevas conexiones, K_{ij}^M, a que ésto da lugar, son usadas para corregir las conexiones de RH según:

$$J_{ij}^M = J_{ij}^{M-1} + \lambda K_{ij}^M \qquad (5)$$

d) iterar en el punto b).

Como resultado de lo anterior, existen dos *fases de aprendizaje*. La primera que consiste en grabar, mediante la regla de Hebb, los patrones a memorizar en la red de Hopfield según el procedimiento estándar (a). La segunda consiste en modificar las conexiones sinápticas de la red de Hopfield original (b, c y d).

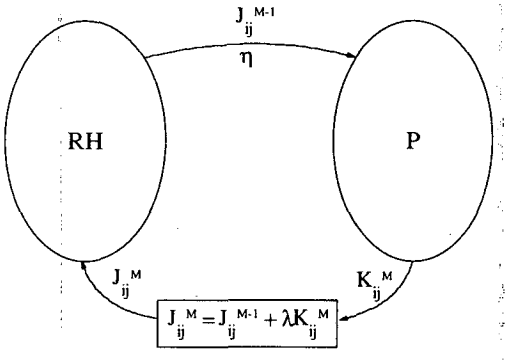

Figura 1: arquitectura empleada. J_{ij}^{M-1} son los pesos resultantes de la iteración anterior, K_{ij}^M son los pesos que desestabilizan a η_i (i=1..N) y J_{ij}^M son los pesos resultantes de la iteración actual.

Es de central importancia la determinación del número óptimo de estados metaestables espurios a desestabilizar, M*. Existen múltiples formas de obtención de tal óptimo, entre ellas elegimos la que asegura la máxima estabilidad de los patrones ξ_i^μ (i=1..N). Esto último se justifica puesto que existe una fuerte indicación [5] de que ello implica la maximización de las cuencas de atracción.

En acuerdo a lo expresado la estabilidad de los patrones se define según:

$$\Delta_{min}^M = \min\left\{\xi_i^\mu \sum_j \xi_j^\mu J_{ij}^M \text{ con } i = 1..N \text{ y } \mu = 1..q\right\} \qquad (6)$$

y su máximo:

$$\Delta_{min}^{M^*} = \max\left\{\Delta_{min}^M \text{ con } M = 0..M^{max}\right\} \qquad (7)$$

donde J_{ij}^M son los pesos que surgen de haber operado el perceptrón inverso sobre un número M de estados metaestables espurios. Esto en nuestro caso equivale a un número M de iteraciones. En la figura 2 se muestran resultados de estabilidad de los patrones.

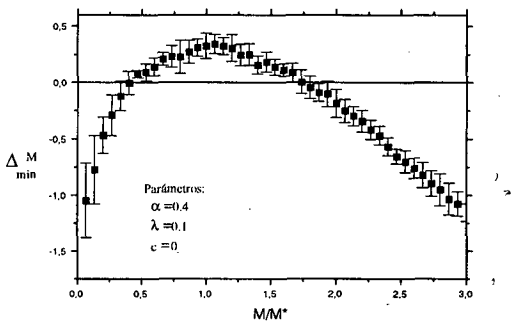

Figura 2: estabilidad de los patrones (ver texto) versus número de pasos de desestabilización. El máximo se encuentra en M*=500. Cada punto es un promedio sobre al menos 8 redes, mostrándose también la dispersión de cada uno de ellos.

Para el número óptimo M* proponemos la siguiente funcionalidad:

$$M^*(\lambda,q,N,c) = f^\lambda(\lambda)\cdot f^q(q)\cdot f^N(N)\cdot f^c(c) \qquad (8)$$

donde λ (Ec. 5) es la medida de la corrección hecha a las conexiones de RH con los pesos K_{ij}^M que desestabilizan estados espurios. q es el número de patrones, N es el número de neurodos y c es la cota de estabilidad del algoritmo del perceptrón inverso. La suposición de una dependencia funcional multiplicativa con los parámetros surge de la independencia entre ellos, hecho que fue corroborado numéricamente. La expresión final es:

$$M^*(\lambda,q,N,c) = \frac{1}{\lambda}\cdot q\cdot\left(\frac{\kappa_1}{N}+\kappa_2\right)\exp(\kappa_3\cdot c) \qquad (9)$$

donde $\kappa_1 = 16.684 \pm 2.22$,
$\kappa_2 = 1.5 \pm 0.047$,
$\kappa_3 = 0.91 \pm 0.02$.

Resultados

Es particularmente instructivo comparar un dado modelo con el modelo de Hopfield [4]. En consecuencia nuestros resultados serán contrastados con el esquema de Hopfield.

1) *Capacidad*: la figura 3 muestra una curva típica de capacidad mostrando los resultados del modelo propuesto versus la red de Hopfield. La ordenada muestra el overlap del estado estacionario final ζ_i (i=1..N) alcanzado al relajar desde el patrón ξ_i (i=1..N) para distintas cargas de la red:

$$m = \frac{1}{N}\sum_i \xi_i \zeta_i \qquad (10)$$

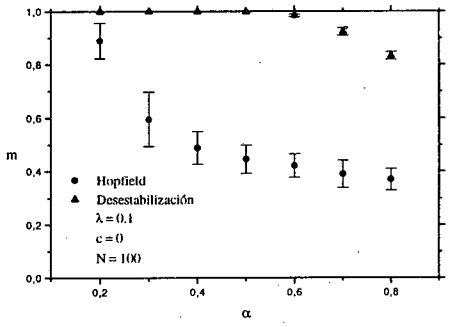

Figura 3: overlap m versus α (=p/N). Cada punto es un promedio sobre al menos 8 redes.

2) *Tamaño de cuencas de atracción*: en la figura 4 se muestra, para cierto valor de la carga (α=q/N), el tamaño de las cuencas alrededor de los patrones. La red es iniciada en configuraciones con overlap inicial m_o respecto a un patrón de referencia. La ordenada muestra el overlap de este patrón con el estado estacionario final alcanzado después de la relajación. Resultados similares se han obtenido para tamaños de red de hasta N=200 y para otros valores de c.

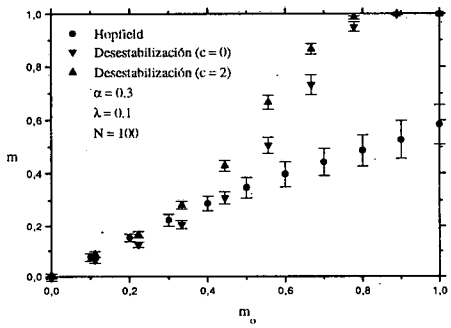

Figura 4: overlap final m versus overlap inicial m_o. Promedio sobre mas de 10 redes.

3) *Minimización de los estados metaestables espurios y de ciclos límites si los hubiera*. Esto es verificable en la figura de tamaño de cuencas (figura 4) y de tiempo de relajación (figuras 5a, b y c).

4) *Tiempo de relajación para alcanzar un punto fijo*, a partir de un overlap inicial m_o respecto de un patrón de referencia. En las figuras 5a, b y c se representan con cruces (x) los casos que relajan acercándose al patrón (m_f>0.95), y con círculos (o) los que se alejan (m_f<0.95). Según se muestra, la recuperación de las memorias es comparativamente mas rápida (a cargas similares) que el modelo de Hopfield y se mantiene aún a cargas superiores, donde el modelo de Hopfield deja de funcionar como CAM.

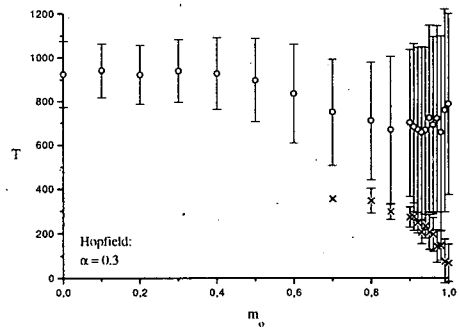

Figura 5a: tiempo de relajación T (o número de pasos de relajación) para alcanzar punto fijo, partiendo de overlap inicial m_o respecto de un patrón de referencia (ver texto).

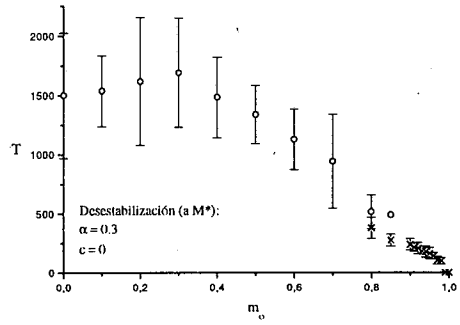

Figura 5b: idem a 5a (ver texto).

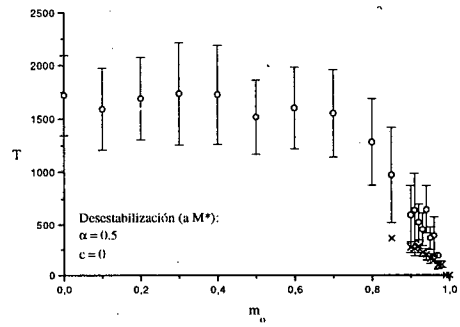

Figura 5c: idem a 5a (ver texto).

Conclusiones

En este trabajo se han mostrado los siguientes puntos:

- resultados relevantes de un extenso estudio numérico realizado para entender mejor el procedimiento de desestabilización,
- que el procedimiento de desestabilización de los estados espurios utilizada es una explicación alternativa y válida para la hipótesis de Crick y Mitchison,
- que la aplicación de este procedimiento implica un claro mejoramiento de los cuatro criterios que permiten comparar la performance de una CAM.

Agradecimientos

Los autores agradecen a un referee anónimo cuyas sugerencias y comentarios permitieron mejorar este trabajo. Así mismo agradecen la ayuda económica de la Secretaría de Ciencia y Técnica de la U.N. de San Luis. J.A.H. es investigador del CONICET y P.M.P. es becario de iniciación del CONICET.

Referencias

1 - F. Crick and G. Mitchison, (1993) Nature, **304**, 11.
2 - J. Van Hemmen, L. B. Ioffe and R. Kuhn, Physica A (1990), **163,** 386-392.
3 - J. J. Hopfield, D. I. Feinstein and R. G. Palmer, (1983) Nature, **304**, 158-159.
4 - J. J. Hopfield, Proc. Natl. Acad. USA, **79**, 2554 (1982).
5 - W. Krauth, J. Nadal and M. Mézard, (1988) J. Phys. A: Math Gen. **21**, 2995.