

Discriminación, generalización y clasificación mediante autómatas neuronales con ruido que induce depresión sináptica

Joaquín Marro

Dept. de Electromagnetismo y
Física de la Materia
Facultad de Ciencias
Universidad de Granada
18071, Granada (SPAIN)
jmarro@ugr.es

Jesús M. Cortés

Dept. de Electromagnetismo y
Física de la Materia
Facultad de Ciencias
Universidad de Granada
18071, Granada (SPAIN)
jesus.m.cortes@gmail.com

Joaquín J. Torres

Dept. de Electromagnetismo y
Física de la Materia
Facultad de Ciencias
Universidad de Granada
18071, Granada (SPAIN)
jtorres@onsager.ugr.es

Resumen

En una serie de trabajos recientes [15, 7, 21, 8, 16, 9] se estudian varias realizaciones de redes de neuronas artificiales que tienen aspectos notables. Se trata de algoritmos que, además de mostrar una memoria asociativa eficaz, son capaces de detectar similitudes (generalización) y diferencias (discriminación). Se sigue su potencialidad para *restaurar*, esto es, identificar el original del que procede una copia fuertemente degradada y, quizás más importante, para clasificar patrones complejos incluyendo, por ejemplo, fotografías de alta resolución o el conjunto de cualidades que identifican con precisión una gran corporación. Los algoritmos aquí formulados y estudiados están basados en la hipótesis de que el “ruido” sináptico rápido que se observa en sistemas naturales es esencial para que los procesos de información sean eficaces. Más concretamente, incorporan fluctuaciones sinápticas que, acopladas a la actividad neuronal, producen estados de *depresión* postsináptica que mimetizan lo que se observa en algunos sistemas neurobiológicos. Se sigue una estrategia que mejora notablemente otros métodos familiares como, por ejemplo, el uso de atractores tipo Hopfield [13, 4] que, a pesar de su evidente interés teórico, presentan indeseables estados mezcla, evoluciones muy lentas, capacidad estrictamente limitada y errores estadísticos inadmisibles. En esta comu-

nicación ilustraremos estos hechos, y revisaremos así nuevos resultados en el modelado de propiedades cooperativas que, de acuerdo con evidencia experimental reciente, podrían determinar en parte el procesado de información en el cerebro.

1. Introducción

Aunque la estrategia del cerebro en relación con el problema de discriminación, generalización y clasificación de memorias es todavía desconocida, hay ciertas indicaciones que, siguiendo a otros autores, vamos a tomar aquí como hipótesis básicas para la formulación de nuevos paradigmas de redes neuronales capaces de resolver dichas tareas. Admitimos, en primer lugar, que el procesado de patrones o memorias es el resultado de la cooperación entre muchas unidades. También es hoy evidente que la memoria asociativa real es consecuencia de lo que podríamos llamar una *dinámica globalmente estocástica*. Esto es, usando un lenguaje familiar, la dinámica del sistema no sólo induce tendencia estocástica hacia un atractor *puro*, o “punto fijo”, como en el modelo de Hopfield [13, 4]. También ocurren saltos estocásticos entre un conjunto de atractores que suele constituir una *clase* de las varias en las que puede dividirse el conjunto de todos los atractores. De hecho, de acuerdo con la observación de que podemos identificar una

cara familiar antes de recordar su identidad, parece que el cerebro reconoce antes la clase que el elemento concreto [6]. En consecuencia, tendríamos que exigir a nuestros modelos que fuesen capaces de clasificar los patrones almacenados de acuerdo con algún criterio, generalmente el grado de correlación, y de visitar los elementos en cada clase antes de pasar a otra, tal como se ha informado que ocurre en la actividad cortical de monos [2] y en la del bulbo olfativo de insectos [14]. Numerosos experimentos sugieren hoy que esta *dinámica globalmente estocástica* es consecuencia de fluctuaciones, generalmente rápidas, en las intensidades sinápticas [1]. Es decir, a pesar de que la actividad neuronal ya es intrínsecamente estocástica, requiere de la competencia continua de un “ruido” sináptico rápido para que la transmisión de información y otros procesos de computación sean suficientemente eficaces [3, 24]. En consecuencia, siguiendo esfuerzos recientes [5], hemos considerado interesante investigar el efecto que pueden tener las sinapsis estocásticas en los puntos fijos o atractores de redes de neuronas. Pero el ruido sináptico (rápido) de nuestros modelos no es arbitrario, sino modelado de forma que, acoplado con la actividad neuronal, produce el fenómeno de *depresión* a tiempos cortos que se ha detectado recientemente en los sistemas naturales. Más concretamente, se observa que periodos de elevada actividad presináptica disminuyen la liberación de neurotransmisores y, por tanto, se *deprime* la respuesta postsináptica [5, 23, 19, 20, 1].

2. Modelos y Resultados

Vamos a ilustrar las introducidas en la sección anterior en un caso sencillo. Sea un conjunto de neuronas que cambian su estado estocásticamente con el tiempo reguladas mediante un parámetro “temperatura”, T_0 . Cada neurona está conectada a las demás mediante sinapsis cuyos pesos también cambian estocásticamente con el tiempo, pero reguladas por otro parámetro “temperatura”, T_1 . Este tipo de competición ha sido estudiado antes en Refs.[10, 11, 22, 17, 18]. Si

$\mathbf{S} = \{s_i = \pm 1; i = 1, \dots, N\}$ representa el estado del conjunto de las neuronas en un instante y $w_{ij} \in \mathfrak{R}; i, j = 1, \dots, N$ son los pesos sinápticos, interesa la probabilidad de $\Omega = (\mathbf{S}, \mathbf{W} = \{w_{ij}\})$ en cualquier tiempo t . El requerimiento de estocasticidad implica que Ω satisface la ecuación maestra familiar [18]. Ésta se complementa suponiendo que la correspondiente probabilidad de transición es, con probabilidad p , igual a $c^{\mathbf{W}}(\mathbf{S} \rightarrow \mathbf{S}')$ con \mathbf{W} dado y, con probabilidad $1 - p$, igual a $c^{\mathbf{S}}(\mathbf{W} \rightarrow \mathbf{W}')$ con \mathbf{S} dado. Para completar el modelo, elegimos las probabilidades de transición elementales $c^{\mathbf{W}}$ y $c^{\mathbf{S}}$ buscando la máxima sencillez. Vamos a suponer que la primera factoriza en funciones de $2(s_i/T_0)h_i^{\mathbf{W}}(\mathbf{S})$, donde $h_i^{\mathbf{W}}(\mathbf{S}) = \sum_{j \neq i} w_{ij}s_j$ es la corriente neta presináptica (o campo local) sobre la neurona i . Los pesos w_{ij} se suponen determinados en un proceso previo de aprendizaje, por ejemplo usando la regla de Hebb [12] a partir de M patrones.

El caso $p \rightarrow 0$ es particularmente interesante puesto que corresponde a ruido muy rápido. Además, el modelo puede resolverse entonces reescalando el tiempo con p y suponiendo una distribución estacionaria para el ruido. El resultado es que el sistema no admite una descripción termodinámica usual, pues sus estados estacionarios no son de equilibrio [18] debido a la competición entre las dos dinámicas estocásticas. Esta situación se analizó originalmente en las Refs.[11, 22, 17, 15].

Un hecho nuevo notable es el diagrama de fases para la versión “autómata neuronal”, esto es, cuando el sistema se actualiza globalmente (*dinámica de Little*), en lugar de actualizar secuencialmente una neurona en cada paso. Además de una fase desordenada y otra con memoria asociativa, se observan regiones en el espacio (T_0, T_1) en las que se producen continuos saltos de un atractor a otro. Dependiendo del valor de los parámetros, este proceso de saltos puede ser complejo [7], como se ilustra en Fig. 1. Nótese que no aparecen los estados mezcla que limitan severamente la utilidad del modelo de Hopfield [4, 17]. Además, en general, los cambios de fase son aquí de primer orden, lo que permite diseñar procesos

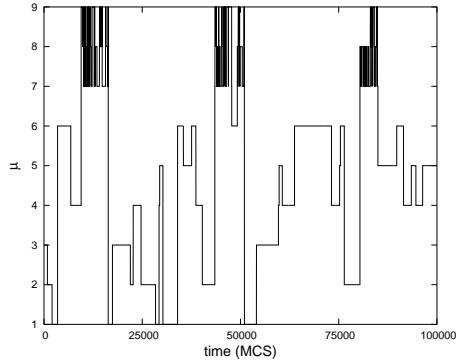


Figura 1: Evolución con el tiempo del orden μ del patrón más parecido al estado del sistema en una simulación Monte Carlo (MC). Esto ilustra cómo el autómata constantemente salta entre los patrones almacenados. Los saltos son aparentemente aleatorios pero, como muestra la parte alta del gráfico, detecta la familia formada por los patrones 7, 8 y 9 que están más correlacionados entre sí que con el resto. La red consiste aquí de 10^4 neuronas y $M = 9$ patrones correlacionados con $T_0 = 0,05$ y $T_1 = 1,25$ (en unidades de las temperaturas críticas respectivas).

de reconocimiento con error prácticamente despreciable [15]. El mismo modelo ha sido estudiado recientemente variando la topología de la red de conexiones sinápticas [21]. Se ha visto así, por ejemplo, que la memoria asociativa de una red aleatoriamente diluida mejora sensiblemente si la topología es “scale free” con la misma densidad de conexiones.

Es notable que los saltos entre patrones realizan la llamada “family discrimination” [6] para una elección adecuada de los detalles del modelo. Supongamos, por ejemplo, que hacemos $w_{ij} = \bar{w}_{ij}x_j$, donde \bar{w}_{ij} se fija en un proceso previo de *aprendizaje*, como antes, y x_j es una variable estocástica. Supongamos también que, en el límite $p \rightarrow 0$, la distribución estacionaria para esta variable es un producto de funciones $P(x_j|\mathbf{S}) = \zeta(\bar{\mathbf{m}}) \delta(x_j + \Phi) + [1 - \zeta(\bar{\mathbf{m}})] \delta(x_j - 1)$. Denotamos $\bar{\mathbf{m}} = \bar{\mathbf{m}}(\mathbf{S})$ al vector cuyos componentes son las *intersecciones* $m^\nu(\mathbf{S}) \equiv \frac{1}{N} \sum_i s_i \xi_i^\nu$, donde $\xi^\nu = \{\xi_i^\nu = \pm 1, i = 1, \dots, N\}$ representan los M patrones almacenados y $\zeta(\bar{\mathbf{m}})$ es una función de $\bar{\mathbf{m}}$ por

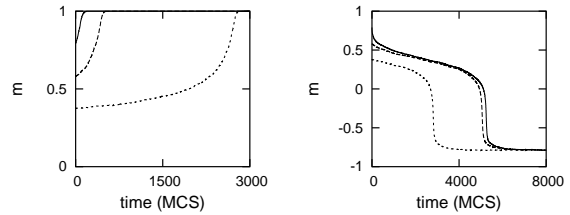


Figura 2: Evolución con el tiempo de la intersección entre el estado del sistema y el único patrón almacenado en una simulación MC con 3600 neuronas y $T = 0,1$. Cada trayectoria corresponde a un estado inicial distinto. Los dos casos son para $\delta = 0,3$ y $\Phi = 1$ (dibujo de la derecha; caso de Hopfield, sin ruido depresor) y $\Phi = -1$ (izquierda).

determinar. Esta elección hace que el modelo muestre depresión sináptica acorde con observaciones experimentales [23, 19, 9].

Este modelo, incluso en el caso de una dinámica secuencial [8], presenta en el espacio (T, Φ) , donde Φ es el parámetro que controla en cierto modo la intensidad del “ruido”, una fase con memoria asociativa. Es interesante que aparece un punto crítico para $(T_c, \Phi_c) = (1, -4/3)$, de modo que la transición a esta fase es de segundo orden sólo si $\Phi > -4/3$, que incluye el caso Hopfield $\Phi = -1$ (sin depresión). En otro caso, la transición es discontinua, lo que favorece un eficaz funcionamiento muy cerca por debajo de la temperatura de transición. También es interesante que el modelo puede mostrar una alta sensibilidad a un estímulo externo, de modo que se adapta a cambios de éste, como ocurre de hecho en sistemas naturales. Puede simularse un estímulo en nuestros modelos añadiendo $-\delta \xi_i$, $\forall i$, con $0 < \delta \ll 1$ a cada campo local [5]. Para un patrón, $M = 1$, este estímulo intenta mover el estado del sistema desde el *atractor*, ξ , hacia el “antipatrón n”, $-\xi$. La Fig. 2 ilustra el interesante comportamiento que se sigue en este caso [8].

Esta figura ilustra la importancia que puede tener el ruido depresor en la sensibilidad

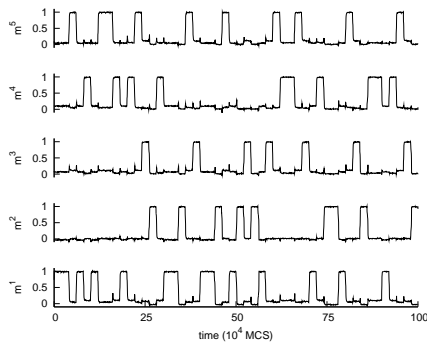


Figura 3: Evolución temporal de las intersecciones en una simulación MC con 400 neuronas, $M = 5$ patrones ortogonales, $\Phi = 0,05$, y $T = 0,1$. La perturbación en este caso consiste en un estímulo $-\delta\xi^\nu$, $\delta = 0,3$, con el índice ν del patrón cambiando cada 20000 pasos MC. El mismo experimento con $\Phi = -1$ (caso de Hopfield sin ruido depresor) produce una señal constante.

del sistema frente a estímulos externos. Se obtienen situaciones similares al variar el parámetro δ y otros detalles. También ocurre que los saltos entre patrones no son críticamente sensibles al tipo de éstos. Vemos este hecho en la Fig. 3, que corresponde al caso de patrones ortogonales, de modo que sus intersecciones o solapamientos mutuos son $m^{\nu,\mu} \equiv 1/N \sum_i \xi_i^\nu \xi_i^\mu \approx 0$, que puede compararse con el caso mostrado en la Ref.[8] para patrones bastante solapados entre sí. En el caso de Hopfield no se observa sensibilidad, independientemente del tipo de patrones; esto es, la señal es plana a pesar del estímulo (a menos que δ tome valores no razonables). Las cuencas de atracción en este caso son grandes e impiden saltos para δ pequeño, excepto suficientemente cerca del punto crítico $T_c = 1$, donde se produce la típica divergencia de las fluctuaciones. Esto es, el “ruido” biológicamente motivado que hemos introducido arriba generalmente desestabiliza el punto fijo en el caso interesante de δ pequeño lejos de criticalidad.

3. Conclusiones

En definitiva, se puede concluir acerca del interés del “modelo de las dos temperaturas” expuesto anteriormente [18], una red de neuronas con dos parámetros principales (“temperaturas”) que controlan la dinámica estocástica de las neuronas y de las sinapsis, respectivamente. Hemos mostrado que, incorporando ciertos detalles que sugieren recientes experimentos en neurobiología, el modelo resulta eficaz en labores de discriminación, generalización y clasificación. Desde un punto de vista más formal, hemos estudiado las consecuencias de ruido sináptico rápido en los atractores de un sistema que almacena un número finito de patrones (aunque los resultados disponibles se refieren al caso $\alpha = M/N \rightarrow 0$, estudios preliminares sugieren un comportamiento similar para un número macroscópico de patrones, esto es, $\alpha \neq 0$). Entre los casos estudiados, el comportamiento más interesante ocurre cuando el ruido depende de la corriente total presináptica que llega a la neurona posináptica. Aparte de algunos resultados analíticos obtenidos con la teoría de campo medio, hemos realizado series de simulaciones Monte Carlo. Los resultados numéricos sugieren que la eficacia del sistema mejora en el caso de actualización global, esto es, para un *autómata neuronal*.

Se concluye también que el ruido presináptico rápido induce una condición de noequilibrio que sensibiliza notablemente a la red frente a estímulos externos. En concreto, el ruido desestabiliza el atractor, y el estado evoluciona hacia otro punto fijo. Esto implica una mejora en la habilidad de la red para detectar cambios en el exterior. De hecho, hemos demostrado que el proceso de saltos puede ser caótico, como sugiere el exponente de Lyapunov en Fig. 4 que se sigue de un cálculo tipo campo medio.

Terminamos comentando que hay ciertos indicios para pensar que algunos procesos naturales de clasificación podrían seguir una estrategia similar a la de nuestros modelos. Es decir, diferentes atractores corresponderían a distintos objetos, y una dinámica convenientemente perturbada por ruido rápido induciría

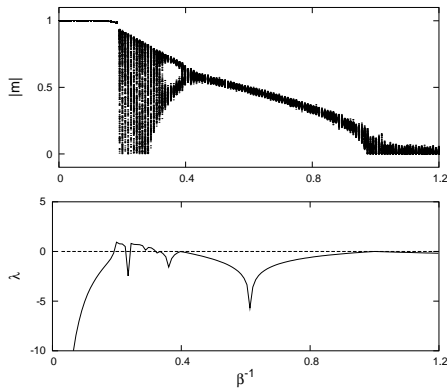


Figura 4: Valor absoluto de la intersección, para $M = 1$, en función de la temperatura neuronal; $\Phi = 0,05$. Esto ilustra, en una simulación MC con 10^4 neuronas, comportamiento irregular y la transición a $T_C = 1$ a una fase con memoria asociativa. El gráfico de abajo muestra el exponente de Lyapunov en una descripción campo medio, demostrando comportamiento caótico donde $\lambda > 0$.

visitas a los atractores en una determinada clase caracterizada por un cierto grado de correlación entre sus elementos. De hecho, como se indica arriba, hay evidencias de que un mecanismo de este tipo podría ocurrir en varios fenómenos naturales.

4. Agradecimientos

Este resumen describe resultados del trabajo continuado realizado por los autores en colaboración con Pedro L. Garrido, Hilbert J. Kappen, Cristian Morillas y David Navidad. También se agradecen ayudas económicas específicas recibidas del MCyT-FEDER (project No. BFM2001-2841 y contrato Ramón y Cajal) y del MADOC.

Referencias

[1] L.F. Abbott and W.G. Regehr, *Nature* **431**, 796 (2004)

[2] M. Abeles, H. Bergman, I. Gat, I. Meilijson, E. Seidemann, N. Tishby and E. Vaadia, *Proc. Natl. Acad. Sci. USA* **92**, 8616 (1995)

[3] C. Allen and C. F. Stevens, *Proc. Nat. Acad. Sci. USA* **91**, 10380 (1994)

[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys.* **173**, 30 (1987)

[5] D. Bibitchkov, J.M. Herrmann, and T. Geisel, *Network: Comput. Neural Syst.* **13**, 115 (2002)

[6] R. Bogacz and M.W. Brown, *Hippocampus* **13**, 494 (2003)

[7] J.M. Cortes, P.L. Garrido, J. Marro, and J.J. Torres, *Neurocomputing* **58-60**, 67 (2004)

[8] J.M. Cortes, J.J. Torres, J. Marro, P.L. Garrido, and R. Kappen, "Effects of fast presynaptic noise in attractor neural networks", *Neural Computation*, to be published.

[9] J.M. Cortes *et al.*, "Algorithms for identification and categorization", AIP Proceedings, in press.

[10] P.L. Garrido and J. Marro, *J. Phys. A: Math. Gen.* **25**, 1453 (1992)

[11] P.L. Garrido and J. Marro, *Lecture Notes in Computer Science* (Springer-Verlag) **540**, 25 (1992)

[12] D.O. Hebb, *The Organization of Behavior*, Wiley, New York 1949.

[13] J.J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**, 2554(1982)

[14] G. Laurent, M. Stopfer, R. Friedrich, M. Rabinovich, A. Volkovskii, and H. Abarbanel, *Annu. Rev. Neurosci.* **24**, 263 (2001)

[15] J. Marro, J.J. Torres, and P.L. Garrido, *J. Stat. Phys.* **94**, 837 (1999)

[16] J. Marro, J. J. Torres, and J. M. Cortes, "Chaotic hopping between the attractors in stochastic neural automata" *Phys. Rev. E*, to be published.

- [17] J. Marro, P.L. Garrido, and J.J. Torres, *Phys. Rev. Lett.* **81**, 2827 (1998)
- [18] J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models*, Cambridge Univ. Press, Cambridge 1999.
- [19] L. Pantic, J.J. Torres, and H.J. Kappen, *Neural Comp.* **14**, 2903 (2002)
- [20] A.M. Thomson, A.P. Bannister, A. Mercer, and O.T. Morris, *Phil. Trans. R. Soc. London B* **357**, 1781 (2002)
- [21] J.J. Torres, M.A. Muñoz, J. Marro, and P.L. Garrido, *Neurocomputing* **58-60**, 229 (2004)
- [22] J.J. Torres, P.L. Garrido, and J. Marro, *J. Phys. A: Math. Gen.* **30**, 7801 (1997)
- [23] M. Tsodyks, K. Pawelzik, and H. Markram, *Neural Comp.* **10**, 821 (1998)
- [24] A. Zador, *J. Neurophysiology* **79**, 1219 (1998)