HAL open science

MODELOS MATEMATICOS EN CIENCIAS HUMANAS : MODELOS DE MEMORIA Y DE COALICIONES

Mirta B. Gordon

To cite this version:

Mirta B. Gordon. MODELOS MATEMATICOS EN CIENCIAS HUMANAS : MODELOS DE MEMORIA Y DE COALICIONES. InCognito, 2009, 3 (3), pp.39-61. hal-00404241

HAL Id: hal-00404241

https://hal.science/hal-00404241

Submitted on 15 Jul 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MODELOS MATEMATICOS EN CIENCIAS HUMANAS
MODELOS DE MEMORIA Y DE COALICIONES

Mirta B. GORDON

Laboratoire TIMC-IMAG – UMR 5525
Université de Grenoble – Faculté de Médecine
Bât. Jean Roget – Domaine de La Merci – 38706 La Tronche - France
Mél: mirta.gordon@imag.fr

Resumen

Presentamos de manera pedagógica los modelos de memoria asociativa de Little y de Hopfield. Damos ejemplos del interés que presentan los modelos matemáticos, extendidos a otras disciplinas. Finalmente presentamos un modelo de interacciones sociales y deducimos consecuencias inéditas en el campo social, gracias a la equivalencia con el modelo de Hopfield.

Palabras-clave: memoria asociativa,modelos, interacciones sociales, coaliciones.

Abstract

MATHEMATICAL MODELS IN HUMAN SCIENCES. MODELS OF MEMORY AND COALITIONS

We present the models of associative memory of Little and Hopfield. We give examples of the interest of transposing mathematical models to other disciplines. Finally we present a model of social interactions and deduce novel results thanks to its equivalence to Hopfield's model.

Keywords: associative memory, models, social interactions, coalitions.

Résumé

MODÈLES MATHÉMATIQUES EN SCIENCES HUMAINES. MODÈLES DE MEMOIRE ET DE COALITIONS

Nous présentons de manière pédagogique les modèles de mémoire associative de Little et Hopfield. Nous donnons des exemples de l'intérêt des modèles mathématiques, transposables à d'autres disciplines. Finalement nous présentons un modèle d'interactions sociales et déduisons des conséquences inédites en sciences sociales, grâce à l'équivalence de ce modèle au modèle de Hopfield.

Mots-clé: mémoire associative, modèles, interactions sociales, coalitions.

Resumo

MODELOS MATEMÁTICOS EM CIÊNCIAS HUMANAS. MODELOS DE MEMÓRIA E COALIZÕES

Apresentamos de maneira didática os modelos de memória associativa de Little e Hopfield. Damos exemplos do interesse que os modelos matemáticos apresentam quando estendidos a

outras disciplinas. Finalmente, apresentamos um modelo das interações sociais e deduzimos consequências inéditas no campo social, graças a equivalência com o modelo de Hopfield.

Palavras-chave: memória associativa, modelos, interações sociais, coalizões.

Riassunto

MODELLI MATEMATICI NELLE SCIENZE SOCIALI. MODELLI DI MEMORIA E COALIZIONI

Presentiamo in maniera pedagogica i modelli di memoria associativa di Little e Hopfield. Mostriamo degli esempi d'interesse dei modelli matematici, applicabili ad altre discipline. Infine presentiamo un modello di interazioni sociali e deduciamo nuovi risultati grazie alla sua equivalenza con il modello di Hopfield.

Parole chiave: memoria associativa, modelli, interazione sociale, coalizione.

1. Introducción

Los modelos matemáticos en ciencias naturales o sociales parten de hipótesis simples a partir de las cuales tratan de explicar fenómenos complejos. Este tipo de modelos suele aportar nuevas ideas y suscitar preguntas que hacen avanzar nuestra comprensión de los fenómenos estudiados. Como lo han observado muchos investigadores, un modelo matemático puede trascender el problema para el cual fue propuesto originalmente, y ser "recuperado" por otros dominios o disciplinas.

Un caso notorio es el del modelo de Ising, propuesto en 1924 para explicar porqué ciertos cristales forman imanes, es decir, ordenan sus momentos magnéticos espontáneamente. Luego veremos que la comprensión del modelo jugó un papel esencial en los modelos de memoria.

A pesar de que la existencia de los imanes intrigó a la humanidad desde hace miles de años, hubo que esperar hasta la primera mitad del siglo XX para que un modelo propusiera un marco de reflexión fructífero. El modelo de Ising parte de dos suposiciones básicas: que los átomos poseen un momento magnético (un imán microscópico) llamado spin, y que dichos momentos interactúan, es decir, ejercen influencias mutuas. Para simplificar, Ising supuso que los spines pueden apuntar en sólo dos direcciones (norte y sur, que denotamos +1 y -1). Si no hubiera interacciones, cada spin tendría una orientación arbitraria. En promedio la mitad apuntarían hacia el norte y la mitad hacia el sur. La imantación del cristal, que es simplemente la suma de las imantaciones elementales, es nula. Eso es lo que sucede en los materiales no magnéticos. Cuando las interacciones —de origen cuántico– existen y son positivas, cada spin minimiza su energía si adopta la misma orientación que sus vecinos. Si así no fuera, su energía aumentaría (y por consecuencia, aumentaría la energía de todo el sistema). En física, el estado macroscópico de los sistemas compuestos por muchas partículas que interactúan está determinado por un compromiso entre dos tendencias: la minimización de la energía que tiende a estabilizar configuraciones ordenadas, y la maximización de la entropía que tiende a favorecer configuraciones desordenadas. Si la temperatura es suficientemente baja, la minimización de la energía es la que domina: la mayoría de los spines se orienta en la misma dirección, es decir adoptan el mismo estado (todos +1 o todos -1). Las interacciones dan lugar a un estado colectivo: como la mayoría de los spines se orientan en la misma dirección la suma de los estados de spin (la imantación) es del orden del número de spines que componen el cristal. Se dice que

la imantación es una propiedad macroscópica. Si la temperatura aumenta, se manifiestan los efectos de la entropía, que tienden a desordenar el sistema. Los spines tienen una cierta probabilidad de adoptar estados que no minimizan la energía, y esa tendencia es tanto mayor cuanto mayor es la temperatura. Correspondientemente, la imantación disminuye. En el modelo de Ising la imantación desaparece cuando la temperatura aumenta por encima de una temperatura crítica. En ese estado los spines cambian de orientación en función del tiempo de manera no coordinada, de modo que en cada instante una mitad de los spines aproximadamente se encuentran en el estado +1 y la otra mitad en el estado -1. Este cambio cualitativo de las propiedades del sistema cuando un parámetro (en este caso la temperatura) llega a su valor crítico se llama transición de fase.

El modelo de Ising abrió pistas que permitieron explorar distintos aspectos del magnetismo y de las transiciones de fases en general. Varias generalizaciones fueron propuestas para explicar otros fenómenos físicos, como por ejemplo por qué ciertas aleaciones binarias (que son cristales compuestos por dos elementos) se ordenan formando una red cristalina donde las posiciones relativas de los dos tipos de átomos se repiten. A la temperatura crítica se produce une transición orden-desorden. Después, Sherrington y Kirkpatrick (1975) consideraron un modelo de Ising con interacciones positivas y negativas para entender el comportamiento de cristales magnéticos muy desordenados, llamados vidrios de spin. El análisis del modelo resultó ser sorprendentemente complejo, y dió lugar a la invención de nuevos conceptos y técnicas matemáticas que aún están siendo investigadas.

Como ya ha sido mencionado, el modelo de Ising no sólo es un paradigma dentro de la física: diversas generalizaciones han sido aplicadas a la economía, a las ciencias políticas, a la teoría de juegos, etc. Los modelos de memoria e interacciones sociales que presentamos en este artículo son dos ejemplos particularmente interesantes en el marco de las ciencias cognitivas.

El modelo de memoria asociativa de Hopfield (1982), un caso particular del modelo de Sherrington y Kirkpatrick, mostró que la memoria se puede concebir como la emergencia de motivos de actividad neuronal en el cerebro, determinados por aprendizaje. Así, la memoria estaría distribuida. Esta idea se opone a la teoría de la neurona “de la abuela” (ver Gross, 2002), según la cual cada percepto[1] corresponde a la activación de una neurona en particular, que se activaría cada vez que dicho percepto es evocado. Una de las principales consecuencias del modelo de Hopfield es que introdujo una descripción de la memoria completamente novedosa que impregna el vocabulario que se emplea actualmente para describir el funcionamiento del cerebro. Lamentablemente, en los últimos años he podido constatar que dicho modelo prácticamente no se enseña más en las formaciones de ciencias cognitivas, a pesar de su interés.

Este artículo presenta el modelo de Hopfield, que explica cómo a partir de un esbozo o fracción de un percepto memorizado (por ejemplo una parte de una cara), la red neuronal puede reconstruir la representación memorizada, es decir, “recuerda” el percepto. Las propiedades del modelo son presentadas con los detalles necesarios para su comprensión, aunque evitaremos en lo posible los desarrollos matemáticos.

---

[1] Usamos el término poco corriente de “percepto” para indicar la representación interna de una percepción.

Como corolario, y para mostrar con un ejemplo el interés de conocer los modelos existentes, presentaremos un modelo propuesto en ciencias sociales (Axelrod y Bennett, 1996) para explicar la formación de alianzas, coaliciones y otros fenómenos sociopolíticos. Se supone que los entes que forman las coaliciones conocen sus características económicas y socio-culturales recíprocas, y en ese sentido poseen ciertas capacidades cognitivas. Mostraremos que, matemáticamente, el modelo de Axelrod es una declinación del modelo de Hopfield. A partir de las propiedades conocidas de éste último deduciremos conclusiones inéditas en el marco del modelo sociológico.

2. Las redes neuronales

Dado que la actividad del sistema nervioso es eléctrica (Galvani, 1791), y que el sistema nervioso está formado por células conectadas por sinapsis (Ramón y Cajal, 1909-11), la comprensión de cómo se activa una neurona y cómo se propaga el influjo nervioso –crucial para modelizar el comportamiento del sistema– no es suficiente. Por ejemplo, ese conocimiento no nos permite inferir el significado de las señales entre neuronas, es decir, el código neuronal. Uno de los objetivos es comprender cómo se forman los perceptos, qué es lo que cambia cuando aprendemos (la plasticidad del sistema), qué es la memorización, cómo recordamos o reconocemos lo que hemos memorizado, etc.

En los párrafos siguientes presentamos algunas nociones sobre el funcionamiento neuronal así como el modelo simple de neurona utilizado por Hopfield. Pero antes de abordar su modelo de memoria presentaremos algunos desarrollos informáticos que lo precedieron históricamente.

2.1. La neurona binaria

Unos años después de los descubrimientos de Ramón y Cajal, Adrian (1928) descubrió que las señales eléctricas que las neuronas se envían mutuamente son potenciales de acción de unos milisegundos. Estas señales son estímulos de las neuronas pre-sinápticas que contribuyen a modificar el potencial de membrana de la neurona post-sináptica. Cuando éste supera un cierto umbral, la neurona emite a su vez un potencial de acción, en cuyo caso se dice que la célula se activa. Este tipo de funcionamiento todo/nada justifica el modelo de neuronas binarias presentado más adelante. Actualmente, aunque no lo describiremos aquí, se sabe cómo se forman los potenciales de acción, cómo se propagan, etc. Lo más llamativo es que los potenciales de acción de todas las células nerviosas, que son muy variadas, son similares, lo que significa que la información que transportan no está contenida en la forma de la señal. Otro de los descubrimientos de Adrian es que la frecuencia de los potenciales de acción emitidos por una neurona aumenta con la intensidad del estímulo. Estas observaciones sentaron la base de lo que se llama codificación por frecuencia ("*rate coding*"). Finalmente, otra observación crucial de Adrian fue que si una neurona era sometida a un mismo estímulo por períodos cada vez más largos, su respuesta se hacía menos intensa, es decir, la frecuencia de emisión tendía a saturar. Este fenómeno de adaptación o acostumbramiento muestra que la respuesta neuronal depende de la historia, sugiriendo una posible base para el aprendizaje.

Cuando un potencial de acción llega al extremo del axón que está en contacto con la neurona post-sináptica, se produce la apertura de vesículas que contienen

moléculas neurotransmisoras. Estas moléculas migran hasta la dendrita de la neurona post-sináptica, produciendo una despolarización de la membrana de ésta última. Dependiendo del estado de las células, de la historia previa, etc, la sinapsis puede contribuir más o menos eficazmente a la despolarización. Notemos también que todos estos procesos tienen un carácter aleatorio: no siempre se obtiene la misma respuesta ante los mismos estímulos. Resumiendo, los potenciales de acción que llegan a una neurona despolarizan su membrana más o menos eficazmente. Cuando el potencial de membrana debido a la suma de todas esas despolarizaciones supera un cierto umbral, la neurona post-sináptica se activa emitiendo a su vez un potencial de acción.

Investigaciones más recientes mostraron que además de la frecuencia, el instante en que cada potencial de acción llega a la célula es también importante. Por un lado, porque cuando una neurona se activa queda imposibilitada de emitir un nuevo potencial de acción durante un lapso de algunos milisegundos, el período refractario. Durante ese lapso, las señales que pudiera recibir no producen efecto alguno. Por otro lado, hace falta que llegue una mínima cantidad de pulsos más o menos simultáneamente para que, gracias a la adición de sus efectos, la neurona post-sináptica se active. El significado preciso de la organización temporal de los potenciales de acción, es decir, cuál es la información subyacente, es todavía un problema no resuelto.

En este artículo sólo abordaremos los modelos dentro del paradigma de codificación por frecuencia, que más allá de la neurobiología inspiró modelos en informática y en ciencias cognitivas. Esto implica que despreciaremos la forma precisa y la duración de los potenciales de acción, y que sólo nos interesamos por su presencia o por su ausencia. Tales modelos tienen sentido en escalas de tiempo superiores a la centena de milisegundos. Bajo estas condiciones, desde el punto de vista de sus interacciones, una neurona tiene dos estados posibles: es una unidad *binaria*. O está activa y transmite a las otras neuronas la información de que lo está, o está inactiva y no transmite nada. Gracias a estas simplificaciones, McCulloch y Pitts (1943) demostraron en un célebre artículo que todo tipo de razonamiento (en lógica proposicional) se podía implementar por una red de neuronas, es decir, una red de unidades binarias interconectadas. Desde ésa época, la analogía entre cerebro y computadoras da lugar a grandes debates. La posibilidad de que un robot pueda tener aptitudes semejantes a las de un ser humano, y que sea capaz de reemplazarlo inteligentemente, es una cuestión que data de la concepción de las primeras computadoras. Históricamente, la informática y las ciencias cognitivas han dialogado permanentemente, inspirándose mutuamente. Turing (1950) y von Neumann (1958) entre otros escribieron textos muy interesantes al respecto, que son aún de actualidad. Para Turing el problema se reducía fundamentalmente a la capacidad de cálculo, y predecía que antes de fines del siglo XX habría computadoras suficientemente poderosas para comportarse como un ser humano. Von Neumann por su lado notaba que, dada la poca precisión de la transmisión neuronal, probablemente el cerebro use códigos y una lógica muy distintos a los que nosotros, humanos, usamos e implementamos en nuestras computadoras.

2.2. La “neuroinformática”

La informática es una disciplina muy reciente. La primera computadora electrónica fue creada en 1949, pero recién en la década del ‘60 aparecieron las primeras máquinas en el mercado. Poseían memorias del orden de 64 kilo-octetos, y

necesitaban condiciones de temperatura y humedad bien controladas para poder funcionar. Las primeras (micro)-computadoras personales datan de los años 1975-1976, y las actuales PC de 1981. A pesar del progreso extraordinario de las computadoras en cantidad de memoria y rapidez de cálculo, la predicción de Turing parece tanto o más lejana que en 1950.

Hubo muchos intentos de plasmar en realizaciones concretas la analogía entre una red de unidades binarias y el cerebro. Gracias al carácter binario de la transmisión nerviosa, una red de neuronas puede ser considerada como sustrato de un cálculo de lógica proposicional. Frank Rosenblatt (1962) concibió la primera neurona artificial no simulada, que bautizó Perceptrón. Un Perceptrón puede recibir señales del mundo exterior o de otros Perceptrones, por medio de conexiones que juegan el rol de sinapsis. La suma de esas señales, ponderadas por la eficacia de la sinapsis correspondiente, producen un potencial post-sináptico (la despolarización de la membrana). Si dicho potencial supera un cierto umbral, el Perceptrón se activa emitiendo una señal que es transmitida a los otros Perceptrones a los cuales está conectado.

Formalmente, un Perceptrón i, que es un modelo de neurona binaria, puede estar conectado a N fuentes de señal que denotamos k ($1 \leq k \leq N$). Sus estados posibles son $s_i=+1$ (activo) o $s_i=0$ (inactivo). Las señales s_k pueden provenir de otra neurona binaria k, o del mundo exterior. Dichas señales son ponderadas por un coeficiente o peso w_{ik} que representa la eficacia de la sinapsis correspondiente. El signo de w_{ik} indica si la sinapsis es excitante o inhibidora. Consideremos los casos de la Figura 1 en que la neurona i sólo recibe señales de k. El potencial post-sináptico es $w_{ik}s_k$. Supongamos que el umbral de activación de la neurona i es nulo. Entonces, si el potencial post-sináptico es positivo la neurona i se activa ($s_i=+1$); si es negativo o nulo $s_i=0$ (inactiva). Si $s_k=+1$ es interesante notar que si $w_{ik}>0$ (sinapsis excitante) el estado de la neurona i es el mismo que el de k. En cambio, si $w_{ik}<0$ (sinapsis inhibidora) el estado de i es opuesto al de k. Si $s_k=0$ la neurona k no contribuye al potencial post-sináptico de i.

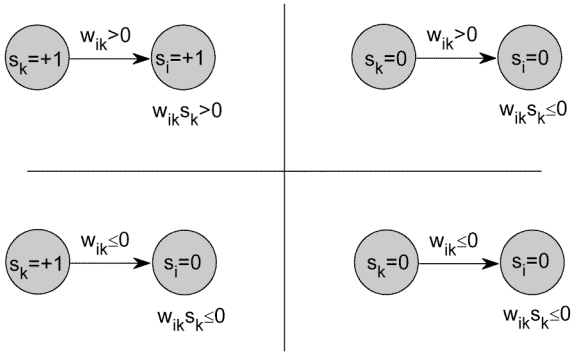

Figura 1 *— Estado de la neurona i en función del estado de la neurona k, bajo la hipótesis de que el umbral de activación de i es nulo. La figura representa los 4 casos posibles, cuando los estados neuronales están codificados por los valores {0,1}.*

En el caso general, el potencial de membrana de la neurona i es la suma de todas las señales provenientes de las neuronas pre-sinápticas, ponderadas por las eficacias de las sinapsis correspondientes. Cuando esta suma supera el umbral de excitación de la neurona i, que denotaremos w_{i0}, ésta se activa (adopta el estado +1). O sea,

$$\text{si} \quad \sum_k w_{ik} s_k > w_{i0} \Rightarrow s_i = 1, \qquad \text{si} \quad \sum_k w_{ik} s_k \leq w_{i0} \Rightarrow s_i = 0 \qquad (1)$$

Rosenblatt mostró que una red de Perceptrones interconectados con pesos w_{ik} adecuados permite reconocer imágenes, como por ejemplo, caracteres escritos. Más aún, Rosenblatt inventó un algoritmo que permite que un Perceptrón "aprenda" los pesos de las conexiones basándose en un conjunto de ejemplos. El algoritmo del Perceptrón establece cómo modificar los pesos paso a paso hasta reconocer perfectamente los ejemplos. Un problema es que el algoritmo converge sólo si los ejemplos poseen una propiedad particular: deben ser linealmente separables. Si no, hay que interconectar varios Perceptrones y en los años 60 no se conocía ningún algoritmo que permitiera "aprender" las eficacias de las sinapsis en ese caso.

A pesar de estos resultados prometedores, durante la década del '70 hubo relativamente poca actividad en este campo, debido en parte a la desaparición prematura de los científicos que hemos citado: en 1950 muere Turing (42 años), en 1957 von Neumann (54 años) y en 1969 desaparecen Pitts (46 años) y Rosenblatt (41 años). Concomitantemente aparece el libro "*Perceptrons*" de Minsky y Pappert (1969), que detalla las limitaciones del Perceptrón de Rosenblatt sin proponer ninguna solución al problema del aprendizaje. El hecho es que la investigación en neuroinformática es abandonada por más de diez años.

Recién en los años '80, con la llegada de una nueva generación de científicos, se reanuda la investigación sobre las redes neuronales. La neurobiología y las ciencias cognitivas se plantean modelizar la percepción y la memoria. En informática se espera que los nuevos modelos servirán como fuente de inspiración para abordar problemas demasiado complejos para los métodos de la inteligencia artificial.

En 1982 fue inventado simultáneamente en varios laboratorios un algoritmo de aprendizaje que permite superar las limitaciones enumeradas en el libro "*Perceptrons*". A partir de entonces, las aplicaciones de los modelos de redes neuronales a la resolución de problemas de inteligencia artificial no cesaron de progresar. Multitud de empresas de informática proponen hoy día soluciones "neuronales" a problemas complejos. Luego de varios intentos de diseñar verdaderas computadoras neuronales, basadas en circuitos de microelectrónica específicos, las aplicaciones se efectúan actualmente simulando el funcionamiento de las redes neuronales en computadoras tradicionales (cuyas capacidades aumentaron extraordinariamente en los últimos años). No detallaremos aquí los desarrollos informáticos, que van de la teoría del aprendizaje a las aplicaciones en clasificación, discriminación de datos, regresión no lineal, etc.

En el mismo año 1982 Hopfield publicó un modelo de funcionamiento de la memoria que provocó una avalancha de estudios teóricos y experimentales. Se trata de un modelo de memoria asociativa, según el cual la memoria está distribuída en la red neuronal y "recordar" corresponde a un fenómeno colectivo de la red. El modelo y sus propiedades más interesantes son descriptos en la sección 3. En la sección 4 presentamos un modelo de ciencias sociales que tiene la misma estructura

matemática que el modelo de Hopfield. Veremos que las propiedades de éste último, conocidas gracias al estudio teórico del modelo, conducen a conclusiones interesantes en el cuadro de las ciencias sociales.

3. Modelos de memoria distribuída

En los animales superiores, la mayor concentración de células nerviosas se encuentra en el cerebro. El cerebro humano contiene del orden de 10^{10} a 10^{11} neuronas. Una manera de aprehender la enormidad de esta cifra, es la siguiente: si tuviéramos que poner una pequeña etiqueta (microscópica) sobre cada neurona y el hacerlo nos llevara sólo 1 segundo por neurona, etiquetar todas las neuronas del cerebro nos llevaría 3000 años (¡sin comer ni dormir!). Otra manera de estimar esta cifra es la siguiente: si pusiéramos las neuronas de un sólo cerebro humano, una al lado de la otra formando un cordón del espesor de una neurona (~50 μm), el mismo mediría 5000 km de largo (¡aproximadamente la distancia Madrid-New York !). En el cerebro, cada neurona puede estar conectada a decenas de miles de otras neuronas, que a su vez tienen decenas de miles de conexiones con otras neuronas. La existencia de tantas neuronas tan interconectadas da lugar a fenómenos emergentes que no se pueden deducir del estudio de neuronas aisladas ni del de pequeños conjuntos de neuronas.

3.1 El modelo de memoria de Little

El primero en presentar el funcionamiento del cerebro como un fenómeno colectivo fue Little (1974). Como físico, Little sabía que una gran cantidad de unidades simples en interacción pueden organizarse de manera muy compleja, impredictible a partir de las propiedades de los componentes del sistema (poco antes P. W. Anderson (1972), premio Nobel de física, había acuñado la frase "*more is different*"). El modelo de Little merece que nos detengamos porque, además de ser un antecedente inmediato del modelo de Hopfield, es un ejemplo de cómo la simplificación nos permite avanzar en nuestra comprensión de los fenómenos naturales.

El modelo de Little tiene en cuenta muchas de las características neuronales descubiertas por Adrian. Como McCullock y Pitts, las neuronas son modelizadas como unidades binarias (llamadas actualmente "neuronas artificiales" pero que en este artículo llamaremos simplemente neuronas[2]). Como ya se mencionó, este modelo binario está basado en el hecho que los potenciales de acción son señales standard, por lo cual se supone que la información transmitida por los axones indica que la neurona emisora está activa. Las neuronas están interconectadas por sinapsis y el estado del sistema, relacionado con los procesos del pensamiento, es descripto dando la lista de cuáles neuronas están activas y cuáles inactivas en cada instante, es decir, por la configuración ("*pattern*") de actividad de la red neuronal.

---

[2] Está claro que se trata de modelos de neuronas, y no de verdaderas neuronas (ver por ejemplo el cuadro de Magritte llamado "Esto no es una pipa"). Obviamente dentro de nuestro contexto no es necesario adjuntar el nombre de "artificial" al término "neurona".

Little considera una red de N neuronas, cuyos estados posibles[3] son σ_i=+1 (activa) y σ_i=-1 (inactiva). Cada par i, k de neuronas está interconectada por sinapsis cuyas eficacias o pesos son w_{ik}. Notemos que si dos neuronas i, k no están conectadas, el valor correspondiente es $w_{ik}=0$. Si, por ejemplo, i tiene una sinapsis sobre k pero k no envía señales a i, entonces $w_{ki}\neq 0$ pero $w_{ik}=0$. Para simplificar supondremos que los umbrales de potencial por encima de los cuales las neuronas se activan son nulos.

Dada una configuración $\sigma=(\sigma_1, \ldots,\sigma_N)$ de la red, el potencial de membrana (que llamaremos simplemente potencial) sobre una neurona i está dado por

$$h_i = \sum_k w_{ik}\sigma_k \qquad (2)$$

Si $h_i>0$, la neurona i se activa, es decir, adopta el estado σ_i=+1. Si $h_i\leq 0$, entonces σ_i=-1. La regla se puede resumir como sigue:

$$\sigma_i(t+1) = \text{signo}\left[h_i(t)\right] \qquad (3)$$

donde $h_i(t)$ es el potencial en el instante t, dado por (2) cuando las neuronas se encuentran en el estado $\boldsymbol{\sigma}(t)$. La figura 2 presenta los casos posibles cuando la neurona i recibe señales de una única neurona : sólo de la neurona k. Nótese que cuando σ_k=-1 y $w_{ik}<0$ el funcionamiento es diferente del de la figura 1. Si $w_{ik}=0$ la neurona k no está conectada a i, y no influye en su potencial de membrana.

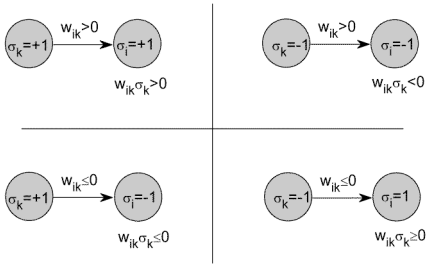

Figura 2 *— Estado de la neurona i en función del estado de la neurona k en los modelos de Little y de Hopfield, bajo la hipótesis de que el umbral de activación de i es nulo. La figura representa los 4 casos posibles, cuando los estados neuronales están codificados por los valores {-1,1}.*

Puede suceder que el nuevo estado de la neurona i al instante t+1 sea distinto que en t. Como las neuronas están interconectadas, un cambio de estado de una de

---

[3] Utilizamos letras griegas cuando los estados neuronales son denotados +1 o -1, y reservamos las letras latinas para los valores 0 o 1. La relación σ = 2s-1 permite transformar las ecuaciones (2) y (3). Nótese que bajo esa transformación, también se modifican los valores de los pesos sinápticos y aparece en la ecuación (2) un término constante que juega el rol de un umbral de activación.

ellas provoca modificaciones en los potenciales post-sinápticos de las otras neuronas, lo cual puede desencadenar una cascada de cambios. La ecuación (3) rige la evolución dinámica de la red.

Little simplificó su modelo al extremo para estudiar su comportamiento de manera analítica: consideró una red de neuronas interconectadas por sinapsis cuyas eficacias tienen valores fijados de manera aleatoria. Dichos valores no cambian durante el tiempo de observación del sistema. Supuso que la red está aislada (ninguna neurona recibe señales del exterior) y que las neuronas están sincronizadas: primero, todos los potenciales post-sinápticos (2) son calculados simultáneamente; a continuación se determinan los estados de todas las neuronas usando (3). Estos nuevos estados difieren generalmente de los precedentes. Cuando las neuronas se encuentran en sus nuevos estados, los potenciales post-sinápticos a su vez se modifican, y como consecuencia, los estados neuronales deberán ser modificados nuevamente. La red evoluciona en función del tiempo: a partir de una configuración inicial $\sigma(0)$, aplicando la ecuación (3) el estado de la red pasará por configuraciones sucesivas $\sigma(1)$, $\sigma(2)$, etc.

Esta dinámica sincrónica, también llamada paralela, es muy adecuada para las simulaciones del modelo pero poco plausible, ya que no existe ningún indicio de que el sistema nervioso tenga un reloj propio que le marque la cadencia. Sin embargo, este tipo de simplificación extrema, muy fructífera en física, permite transformar un problema muy complicado en uno más abordable.

De las simplificaciones del modelo, la hipótesis de sincronización es la más importante porque es la que permitió el estudio analítico del problema, que no detallaremos en este artículo. También se han estudiado las consecuencias de otras simplificaciones adoptadas por Little. Por ejemplo, si una pequeña proporción de neuronas recibe señales del exterior, o si las eficacias de las sinapsis varían lentamente en el tiempo, las conclusiones teóricas no cambian radicalmente.

Notemos que hay una cantidad muy grande de estados posibles de la red; precisamente 2^N. Si N=10000 por ejemplo, hay $2^{10000}=(2^{10})^{1000} \sim 1000^{1000}$ que es un número casi imposible de imaginar (un 1 seguido de 3000 ceros). A partir del estado inicial, el sistema recorrerá un cierto número de esos estados, pero en general no los visitará todos. Little mostró con simulaciones y cálculos analíticos que, a partir de un estado inicial arbitrario, el estado de la red puede cambiar más o menos rápidamente hasta llegar a una configuración “persistente” de actividades. Estos estados colectivos persistentes son configuraciones de actividad donde la mayoría de las neuronas están en un estado coherente con el signo del potencial que actúa sobre ellas de modo que no cambian de estado durante un cierto número de iteraciones. Luego de un cierto tiempo el sistema vuelve a cambiar rápidamente de estado para “visitar” otro estado persistente. Los estados “persistentes” dependen de los valores de los pesos sinápticos. La interpretación de Little es la siguiente: los pesos sinápticos resultan de un aprendizaje; sus valores contienen información de la memoria a largo plazo y determinan cuales serán los estados persistentes. Estos últimos son evocados por las experiencias sensitivas y la memoria a corto plazo.

Nótese que, puesto que a cada iteración las neuronas son capaces de cambiar de estado, la unidad de tiempo del modelo debe ser más grande que el período refractario de las neuronas. La duración de los estados persistentes es pues de varios períodos refractarios.

Este comportamiento se puede visualizar geométricamente. Consideremos un sistema de coordenadas con un eje por cada neurona de la red. En cada eje podemos representar los dos estados de la neurona correspondiente: activa (+1) o inactiva (-1). En ese espacio de dimensión N (el número de neuronas) cada estado de la red está representado por un punto $\boldsymbol{\sigma}$ cuyas coordenadas son los σ_i. Cada vez que una neurona cambia de estado (por ejemplo σ_1 cambia de signo), el punto representativo del estado de la red se desplaza. Los estados persistentes corresponden a pequeños desplazamientos del punto representativo $\boldsymbol{\sigma}$, mientras que cuando la red evoluciona rápidamente (de un estado persistente a otro) $\boldsymbol{\sigma}$ pega grandes saltos.

El motivo de actividad de los estados persistentes depende de los valores particulares de las eficacias sinápticas. Como en las simulaciones de Little los valores de w_{ik} son aleatorios, cada simulación produce estados persistentes diferentes. El modelo no tiene ninguna hipótesis de cómo se determinan los valores de los w_{ik}.

Para completar, es importante mencionar que Little también estudió el modelo teniendo en cuenta que el funcionamiento del sistema no puede ser perfectamente determinista, como lo sugiere la regla (3). Existen numerosas razones en todo sistema vivo para que el comportamiento no siga a la perfección una regla tan precisa. Una razón evidente es que difícilmente el estímulo que provoca el recuerdo (el estado persistente) sea idéntico al estado que se ha memorizado. Por otra parte, desde un punto de vista fisiológico, el mecanismo mismo de la transmisión sináptica es aleatorio. La transmisión se realiza por apertura de vesículas que contienen moléculas neurotransmisoras. Estas migran en el estrecho espacio entre la sinapsis y la membrana de la célula. Tanto la cantidad de moléculas neurotransmisoras expulsadas de las vesículas como la fracción que llega por difusión a la membrana de la célula post-sináptica son procesos aleatorios. La respuesta a un mismo estímulo puede ser cada vez diferente.

En el lenguaje de los modelos se dice que estos procesos introducen ruido en el funcionamiento del sistema. Para dar cuenta del mismo se reemplaza la ecuación (3) por una ecuación que establece con qué probabilidad la neurona i se va a activar cuando el potencial de su membrana vale h_i:

$$P\left[\sigma_i(t+1)=1\right]=p\left(\frac{h_i(t)}{T}\right) \tag{4}$$

donde $p\left(\frac{h_i(t)}{T}\right)$, la probabilidad que el estado de la neurona sea +1, vale

$$p(x)=\frac{e^x}{e^x+e^{-x}} \tag{5}$$

que depende de un único parámetro, T [4]. Estas ecuaciones parecen complicadas, pero son simples de entender. La figura 3 muestra la forma de (5) para tres valores distintos de T. Las curvas tienen una forma sigmóidea. La probabilidad que el estado

[4] Esta manera de introducir el ruido se inspira de la física estadística, donde T representa la temperatura.

de la neurona i sea σ_i=+1 crece con h_i y presenta una variación abrupta en un rango de valores de h que va de –T a +T a ambos lados del origen (h=0). Para h>>T la probabilidad de σ=1 es prácticamente 1; si h<<-T esta probabilidad es casi nula, con lo cual la probabilidad que el estado sea σ=-1 es del orden de 1. Este comportamiento es muy cercano al de la ecuación determinista (3): si el potencial es positivo entonces σ=1, y si es negativo σ=-1. Si el valor absoluto del potencial es del orden de T o menor, la probabilidad de σ=1 es ligeramente mayor si h>0 que si h<0, pero en los dos casos esta probabilidad no es despreciable.

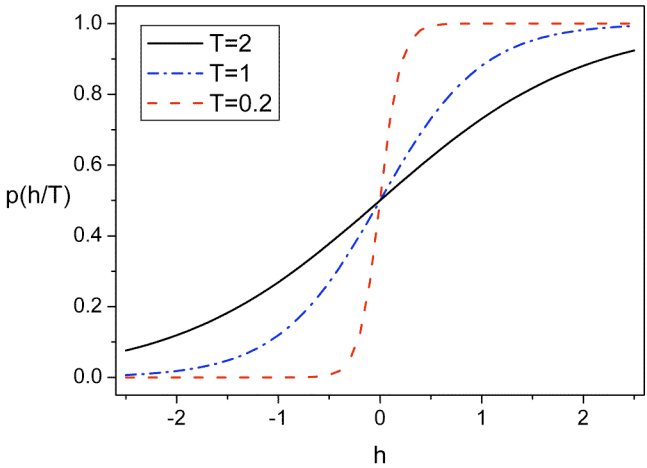

Figura 3 *— Representación de la ecuación (5) para varios valores de T.*

En otros términos, cuanto más pequeño es T, más abrupto es el cambio de p cerca del origen. En el límite T → 0 la ecuación (4) equivale a la (3): el estado de la neurona está dado por el signo del potencial con probabilidad 1. Si T es finito, existe une probabilidad p que σ=1 y 1-p que σ=-1: el estado de la neurona puede no ser igual al signo del potencial. Dado un valor del potencial h, la probabilidad de que la neurona adopte un estado "equivocado" (debido al ruido) es tanto más grande cuanto más grande es T. Es de notar que las neuronas más susceptibles al ruido son aquellas que tienen un potencial h_i pequeño (con respecto a T).

¿Qué modificaciones introduce el ruido en la dinámica de la red? Las neuronas sobre las que el potencial es más fuerte que el ruido (es decir h_i >> T) se comportarán igual que si no hubiera ruido. En cambio el estado de las neuronas con potenciales pequeños fluctuará a lo largo de las iteraciones. Una manera intuitiva de pensar en los estados de la red es imaginar que las neuronas son diodos luminiscentes, que se encienden cuando el estado es +1 y se apagan cuando es -1. A lo largo de las iteraciones, las neuronas con potenciales grandes estarán ya sea encendidas o apagadas permanentemente. En cambio aquéllas que fluctúan estarán alternativamente prendidas o apagadas; las fluctuaciones serán tanto más frecuentes cuanto más pequeño sea el potencial (puesto que entonces las

probabilidades de adoptar uno u otro estado son cercanas a ½). Los estados persistentes se verán más "borrosos" que cuando no hay ruido.

Cuando el parámetro T es pequeño, el punto representativo σ se aleja poco de los estados persistentes en ausencia de ruido. En el caso extremo opuesto, $T \to \infty$, la ecuación (5) vale p=1/2 para todo h. El estado de todas las neuronas tiene probabilidad 1/2 de ser positivo o negativo, independientemente del signo y del valor de h. En ese caso, no existen estados persistentes, el punto representativo σ recorre los 2^N estados posibles de la red en un orden completamente aleatorio. Las neuronas están la mitad del tiempo activas y la otra mitad inactivas, sin que exista correlación entre unas y otras. Con tanto ruido el modelo no funciona para nada como una memoria.

La transición entre los dos tipos de comportamiento (estados persistentes vs. estados con neuronas decorrelacionadas) se produce de manera relativamente abrupta. Por encima de un valor máximo de T (se habla de un valor crítico T_c), desaparecen los estados persistentes. Este fenómeno es similar a una transición de fase de un sistema físico, debida a cambios en la temperatura.

3.2 Modelo de memoria asociativa de Hopfield

El modelo de Little mostró que una red neuronal simplísima presenta estados de actividad persistente. ¿Pero cómo están relacionados esos estados con la memoria? Ya en 1949, Hebb[5] había sugerido que el mecanismo de memorización podría resumirse en un aumento de la eficacia sináptica entre neuronas cuya actividad es persistente. En 1982 Hopfield formuló la regla de Hebb en términos matemáticos, usando el modelo de neuronas binarias y la idea de Little de que cada estado memorizado corresponde a un estado persistente de las actividades neuronales. Supongamos que la red neuronal ha memorizado M objetos (imágenes, palabras, o cualquier otro objeto memorizable), y llamemos "memorias" a los estados correspondientes de la red. Si ξ_i^μ es el estado de la neurona i correspondiente al objeto μ (ξ_i^μ puede tomar los valores +1 o -1), el vector $\boldsymbol{\xi}^\mu = \left(\xi_1^\mu, \ldots, \xi_i^\mu, \ldots \xi_N^\mu\right)$ describe el estado de la red. Si ésta "recuerda" perfectamente el objeto μ su estado será $\boldsymbol{\xi}^\mu$.

La eficacia de la sinapsis entre la neurona i y la neurona k, usando la regla de Hebb tal como la interpretó Hopfield, se escribe así:

$$\begin{aligned} w_{ik} &= \frac{1}{N}\sum_{\mu=1}^{M} \xi_i^\mu \xi_k^\mu \qquad \forall \quad i \neq k \\ w_{ii} &= 0 \end{aligned} \tag{6}$$

El factor 1/N se introduce para que los valores de w_{ik} estén acotados incluso si el número de estados memorizados tiende a infinito (este es el límite usado para estudiar las propiedades del modelo, como lo veremos más adelante). La segunda

---

5 En "*The Organization of Behavior*" (1949): "*When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased*".

línea indica simplemente que las neuronas no tienen conexiones sinápticas sobre ellas mismas. La fórmula (6) merece varios comentarios. Primero, tal como está escrita, las eficacias sinápticas son simétricas: $w_{ik} = w_{ki}$. Es decir, si una neurona k actúa con eficacia w_{ik} sobre la neurona i, entonces la neurona i actúa con la misma eficacia sobre k. Nótese que en el modelo de Little no había ninguna hipótesis sobre la manera en que se formaban las eficacias de las sinapsis: se consideraban dadas (y en las simulaciones, sus valores se fijaban al azar). No sólo *no* se imponía que los w_{ik} sean simétricos sino que, dado que sus valores eran aleatorios, la probabilidad de que fueran simétricos era prácticamente nula. En cambio, la regla de Hebb dada por la ecuación (6) produce eficacias simétricas. Más aún, la razón por la cual se usa la expresión (6) es precisamente porque produce eficacias sinápticas simétricas. Eso permite utilizar toda una serie de resultados de la física estadística para caracterizar el comportamiento de la red. Luego se puede estudiar qué sucede si los w_{ik} no son simétricos, como veremos más adelante. Pero por ahora aceptemos la formulación (6) de Hopfield, y veamos las conclusiones que se pueden extraer sobre el comportamiento del sistema.

Notemos que si sólo se memoriza un estado, M=1 –en cuyo caso el único término de la suma en la ecuación (6) corresponde a μ=1– el valor de w_{ik} es simplemente $w_{ik} = \xi_i^1 \xi_k^1 / N = w_{ki}$. Si en la configuración $\boldsymbol{\xi}^1$ dos neuronas están activas o inactivas simultáneamente, la eficacia de la sinapsis que los conecta es positiva (w_{ik}=1/N). En cambio, si las neuronas se encuentran en estados opuestos, una activa y la otra inactiva, w_{ik}=-1/N. Eficacias negativas corresponden a sinapsis inhibidoras. Es claro que la ecuación (6) no corresponde exactamente a la regla enunciada por Hebb. En particular, cuando dos neuronas están *inactivas* simultáneamente, no hay ninguna razón para que la eficacia de la conexión entre ellas aumente. Nótese que ese "defecto" se podría corregir escribiendo, por ejemplo, $w_{ik}^* = \xi_i^1 \left(\xi_k^1 + 1\right) / N$ de modo que si la neurona pre-sináptica k está inactiva, la eficacia de la conexión no cambia. Un problema de esta definición alternativa es justamente que w_{ik}^* no es simétrica ($w_{ik}^* \neq w_{ki}^*$). El lector interesado por éste y otros puntos mencionados en este artículo puede consultar el libro de P. Peretto (1992).

Cuando hay más de un motivo memorizado (M>1), según (6) cada uno contribuye a la modificación de las eficacias de manera aditiva. Consideremos un par de neuronas i, k. Si en la mayoría de los estados memorizados dichas neuronas están simultáneamente activas o simultáneamente inactivas, la eficacia de la sinapsis que las conecta será positiva w_{ik} . En el caso contrario, w_{ik}<0.

En el modelo de Hopfield, una vez determinados los w_{ik} usando (6), el sistema evoluciona de la misma manera que en el modelo de Little, sólo que en vez de usar una dinámica paralela, se usa una dinámica secuencial. Es decir, en cada instante t se actualiza el estado de una sola neurona, elegida al azar, usando las ecuaciones (2) y (3) (o las ecuaciones (4) y (5) si se quiere tener en cuenta la existencia de ruido). De esta manera se evita el problema del reloj, subyacente a toda dinámica paralela. Los estados de equilibrio estable, que son finalmente los que nos interesan, son comunes a ambas dinámicas (La dinámica paralela puede exhibir, además, estados finales cíclicos de orden 2 : en ellos la configuración del sistema oscila interminablemente entre dos configuraciones).

3.2.1 Propiedades del modelo

Se pueden intuir algunas propiedades del modelo observando su comportamiento en casos extremos. Como antes, denotemos como $\boldsymbol{\sigma}=(\sigma_1, \ldots,\sigma_N)$ el estado de la red, y supongamos que se ha memorizado un único motivo $\boldsymbol{\xi}^1$, de modo que $w_{ik} = \xi_i^1\xi_k^1 / N = w_{ki}$. ¿Qué sucede si el estado inicial de la red es el estado memorizado, $\boldsymbol{\sigma}(0) = \boldsymbol{\xi}^1$ es decir, $\sigma_i(0) = \xi_i^1$ para i=1,...,N? El potencial (2) sobre cada neurona i tiene el mismo signo que ξ_i^1, y si N es grande $h_i(0) \approx \xi_i^1$. De acuerdo con la ecuación (3) ninguna neurona cambiará de estado. El motivo memorizado es estable: se dice que es un punto fijo de la dinámica.

¿Qué sucede si el estado inicial de la red $\boldsymbol{\sigma}(0)$ difiere del (único) estado memorizado $\boldsymbol{\xi}^1$ sobre una fracción ν=n/N de neuronas? Diremos que el estado inicial tiene una fracción ν de errores. Introduciendo (6) en (2) se obtiene que el potencial sobre las neuronas cuyo estado inicial es $\sigma_i(0) = \xi_i^1$ vale $h_i(0) \approx (1 - 2\nu - 1/N)\xi_i^1$. Si $\sigma_i(0) = -\xi_i^1$ el potencial es ligeramente superior, $h_i(0) \approx (1 - 2\nu + 1/N)\xi_i^1$, aunque al orden 1/N la diferencia es despreciable. Como consecuencia, si ν<1/2 los signos de los potenciales son iguales a los de los estados ξ_i^1; cuando el estado de la neurona i (cualquiera sea ésta y cualquiera sea su estado) sea actualizado usando la ecuación (3), adoptará el estado ξ_i^1. Si ya estaba en dicho estado, nada cambia, pero si la neurona estaba en el estado opuesto, no sólo cambia su estado sino que a partir de ese instante la fracción de errores será menor, ya que pasa de ν a ν-1/N. Como consecuencia, en las sucesivas actualizaciones el estado $\boldsymbol{\sigma}(t)$ de la red se irá acercando al estado memorizado $\boldsymbol{\xi}^1$, y terminará convergiendo a éste último, que como hemos visto es un punto fijo en el cual $h_i(0) \approx \xi_i^1$.

Si la fracción de errores inicial es mayor que la mitad de las neuronas (ν>1/2), se puede demostrar que el sistema converge al "anti-estado" $-\boldsymbol{\xi}^1$ (que es algo así como el negativo del estado memorizado: las neuronas activas están inactivas y viceversa). Esta es también una propiedad general del modelo: los puntos fijos existen por pares, a cada estado estable le corresponde otro punto fijo que es su negativo.

Cuando se memoriza más de una sola configuración M>1, el sistema se comporta de manera similar, pero la fracción máxima de errores tolerables ν_{max}=n/N disminuye cuando M aumenta. El valor de ν_{max} da una medida del tamaño de la "cuenca de atracción" de un estado memorizado: todos los estados en el interior de la cuenca, es decir aquéllos que difieren de una configuración memorizada sobre una fracción menor que ν_{max} neuronas son "atraídos" hacia el punto fijo, que es el estado memorizado. Este modo de funcionamiento explica por qué decimos que el modelo de Hopfield es una memoria asociativa : al estado inicial de la red le asocia un estado memorizado.

Si los estados memorizados son ortogonales entre sí, es decir tienen la mitad de las neuronas en el mismo estado (verifican que $\xi_i^\mu \xi_i^\mu = 1$) y la otra mitad en estados opuestos ($\xi_i^\mu \xi_i^\mu = -1$), de modo que satisfacen

$$\frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu = 0 \qquad \forall \quad \mu \neq \nu \tag{7}$$

se puede demostrar que el número de estados recordables es máximo. Un cálculo similar al que hemos hecho más arriba muestra que se pueden memorizar hasta N-1 motivos ortogonales, o sea que se obtiene $M_{max}<N$. El tamaño de las cuencas de atracción de los motivos ortogonales disminuyen con M, y se anula cuando $M=M_{max}$. En este caso, los motivos son puntos fijos, pero si una sola neurona cambia de estado, la red evoluciona hacia otro punto fijo. Vemos pues que la condición de que los estados memorizados sean puntos fijos de la dinámica no es suficiente para garantizar que la red funcione como una memoria asociativa. Hace falta también que las cuencas de atracción de las configuraciones memorizados sean suficientemente grandes.

Es interesante saber cuántas configuraciones arbitrarias, sin elegirlas de manera precisa como en el caso de estados ortogonales, se pueden memorizar. Vamos pues a suponer que los estados neuronales de cada una de las memorias son elegidos de manera aleatoria. Es decir, $\xi_i^\mu = 1$ o $\xi_i^\mu = -1$ con probabilidad ½, para cada valor de i y de μ. La pregunta que nos planteamos es: ¿cuál es la máxima cantidad de motivos, M_{max}, que se puede memorizar usando la regla de Hebb (6) tales que sean puntos fijos de la dinámica (3)? Como ahora los motivos son aleatorios, la respuesta será probabilística. Encontrarla fue un problema muy difícil. Gracias a métodos matemáticos muy elaborados, que permiten obtener resultados en el límite en que la cantidad de neuronas de la red es infinitamente grande, se han podido determinar muchas propiedades interesantes del modelo. En este límite, cuando $N \to \infty$, se pueden utilizar las leyes de los grandes números de la teoría de probabilidades. Pero la dificultad no reside en el pasaje al límite, sino en la necesidad de obtener resultados válidos para cualquier conjunto de estados memorizados elegidos al azar. Para eso hace falta promediar sobre todos los conjuntos posibles de M memorias. Aquí nos limitaremos a presentar los principales resultados.

Como en el caso de configuraciones ortogonales presentado arriba, la máxima cantidad de configuraciones aleatorias memorizables y "recordables" cuando no hay ruido (T=0) es proporcional a la cantidad de sinapsis por neurona de la red (que en el modelo de Hopfield es igual a la cantidad de neuronas). Es decir, $M=\alpha N$. La fracción M_{max}/N se llama capacidad de la red. La capacidad del modelo de Hopfield es

$$\alpha_{max} = \lim_{N\to\infty} \frac{M_{max}}{N} \approx 0.14 \tag{8}$$

Si se introduce ruido en la dinámica del sistema, como en el modelo de Little (ecuaciones (4) y (5)), hay que estudiar los estados persistentes, ya que el estado de

la red $\sigma(t)$ varía con el tiempo. No podemos limitarnos a la noción de punto fijo: en cada actualización las neuronas pueden o no adoptar el estado que corresponde al signo del potencial. Sin embargo, si se observa el sistema durante un período suficientemente largo, se podrá determinar si, en promedio, su estado está cerca o no de un estado memorizado.

La cercanía entre dos estados de la red, $\boldsymbol{\sigma}^1$ y $\boldsymbol{\sigma}^2$, se puede caracterizar por la cantidad siguiente

$$m = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^1\sigma_i^2 \qquad (9)$$

Si los dos estados son idénticos ($\sigma_i^1 = \sigma_i^2 \ \ \forall$ i), m=1. Si los estados son opuestos ($\sigma_i^1 = -\sigma_i^2 \ \ \forall$ i), m=-1. Si son ortogonales, m=0 (comparar con (7)). Llamemos $\langle\sigma_i\rangle$ al valor promedio del estado de la neurona i durante un largo período de observación. El promedio temporal del estado de la red es $\langle\boldsymbol{\sigma}\rangle = \left(\langle\sigma_1\rangle,\ldots,\langle\sigma_N\rangle\right)$. En presencia de ruido, la cercanía entre el estado medio de la red y una configuración memorizada μ es

$$m^\mu = \frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i\rangle \ \xi_i^\mu \qquad (10)$$

Si $m^\mu>0$, la red se encuentra en un estado persistente como los definidos por Little, correlacionado con el motivo memorizado μ. En otros términos, a pesar de que las neuronas cambian de estado de manera aleatoria, dada por las ecuaciones (4) y (5), si la mayoría pasa más tiempo en el estado ξ_i^μ que en el estado $-\xi_i^\mu$, $m^\mu>0$. En ese caso se considera que el motivo μ es "recordado". El estudio del modelo muestra que la capacidad $\alpha_{max}(T)$ disminuye cuando el ruido T aumenta y se anula por encima de un valor crítico que vale $T_c=1$ [6]. Este cambio cualitativo es conocido en la literatura como "pérdida catastrófica de la memoria". Catastrófica porque ni bien α supera α_{max} ninguno de los estados memorizados se puede "recordar". Esta pérdida brutal se puede evitar si se modifica el modelo de aprendizaje o memorización, es decir, si se reemplaza la regla (6), que modeliza la manera en que la memorización modifica las eficacias sinápticas (ver Gordon, 1987).

Las propiedades del sistema se pueden describir en un gráfico con dos ejes: la fracción de configuraciones memorizadas α=M/N en las abscisas y el coeficiente de ruido T en ordenadas. La figura 4 representa de manera esquemática las regiones donde las propiedades del modelo son cualitativamente distintas.

---

[6] Este valor numérico depende del prefactor de la regla de Hebb, que hemos tomado igual a 1/N.

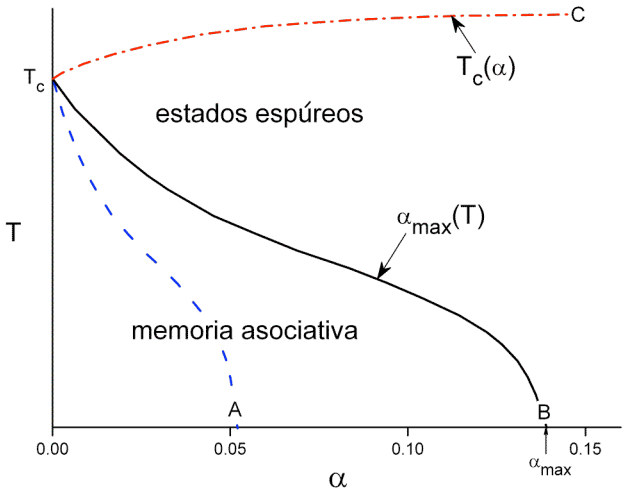

Figura 4 *— Diagrama de fases del modelo de Hopfield.*

Por debajo de la línea T_cB el ruido y la cantidad de motivos memorizados son suficientemente pequeños ($T<T_c$ y $\alpha<\alpha_{max}$) para que existan estados persistentes correlacionados con los estados memorizados. La red es capaz de recordar un motivo memorizado a partir de una fracción de ese motivo. Como ya lo hemos mencionado, se dice que la red constituye una memoria asociativa, ya que asocia a cada estado inicial un estado memorizado, a condición que el estado inicial esté suficientemente cerca del estado memorizado. Este modo de funcionamiento es muy distinto al de una memoria de computadora. En ésta última, para recuperar el contenido de un estado memorizado hace falta conocer la dirección del sitio donde ha sido guardado. Se trata de una memoria "recordable" por su "domiciliación". En cambio, la memoria en el modelo de Hopfield se recupera a partir de una parte de su contenido. En esta región del diagrama también existen otros estados persistentes, llamados espúreos, que no están correlacionados con los motivos memorizados. Si el estado inicial contiene muchos errores, es posible que el sistema converja sobre un estado espúreo. Entre T_cB y T_cA los estados memorizados tienen cuencas de atracción más pequeñas que los estados espurios, en tanto que por debajo de la línea T_cA las cuencas de atracción de los estados memorizados son mayores que las de los estados espúmeos.

Por encima de la línea T_cB el modelo no funciona como memoria asociativa. Los valores $\alpha_{max}(T)$ sobre la línea T_cB corresponden a la capacidad de la red en presencia de una tasa de ruido T. En la región entre las líneas T_cC y T_cB existen estados persistentes, pero no correlacionados con los estados memorizados ($m^\mu=0$ para todo μ). Por encima de la línea T_cC el ruido es tan grande que la dinámica del sistema hace que todos los estados posibles sean recorridos de manera completamente aleatoria, es decir $\langle \sigma_i \rangle = 0$.

El estudio teórico del modelo (ver Amit, Gutfreund y Sompolinsky, 1987) utiliza el concepto de energía de los estados de la red, que no introduciremos aquí. Es posible definir la energía porque los pesos w_{ik} son simétricos, y a partir de ella se pueden usar métodos de la física estadística. Además de las conclusiones mencionadas, dichos métodos permiten calcular otras magnitudes interesantes, como por ejemplo las correlaciones entre los estados persistentes y los estados memorizados para todos los valores de T y α, ciertas propiedades de los estados espurios, etc. También se pueden abordar las propiedades de variantes del modelo y estudiar las modificaciones del diagrama de fases cuando las eficacias de las sinapsis no son perfectamente simétricas, las neuronas no están todas interconectadas, etc. En la medida en que esas perturbaciones al modelo inicial no sean muy importantes, las propiedades que hemos detallado no cambian cualitativamente.

En la década 1985-1995 el modelo de Hopfield fue un paradigma subyacente a muchas investigaciones en neurociencias. Su mérito principal es haber mostrado que era posible explicar el funcionamiento del cerebro como un fenómeno colectivo en el cual el aprendizaje modula las eficacias de las sinapsis y los recuerdos corresponden a estados de actividad de toda la red en su conjunto. Así, no es una neurona en particular la que “recuerda”, sino un gran número de ellas en interacción. Este tipo de funcionamiento es robusto: si hay fallas en unas pocas sinapsis o neuronas, el recuerdo será menos preciso (m^μ será mas pequeño) pero no necesariamente desaparece.

Modelos más recientes consideran escalas de tiempo más pequeñas, incorporando detalles de la respuesta neuronal, y en particular la existencia del período refractario. Esos modelos, mucho más complejos que el modelo de Hopfield, tienen en cuenta la forma de los potenciales de acción y la estadística de los intervalos entre ellos, en un esfuerzo por comprender el código neuronal natural.

4. Modelo de coaliciones de Axelrod - Bennett

El interés de los modelos matemáticos es que son descripciones abstractas que se pueden considerar independientemente del objeto para el cual fueron concebidos. Es por eso que disciplinas muy dispares pueden generar modelos (matemáticamente) similares. Uno de los ejemplos más conspicuos es el modelo de Ising, que hemos descripto brevemente en la introducción, inicialmente concebido para explicar el ferromagnetismo de ciertos cristales. En ciencias humanas, Schelling (1978) describe ciertos modelos de la física que juzga pertinentes para entender fenómenos sociales.

En este capítulo veremos un caso opuesto: presentaremos un modelo propuesto por Axelrod y Bennett (1993) en ciencias políticas y mostraremos que es formalmente equivalente al modelo de Hopfield. Del conocimiento detallado de las propiedades de

este último podremos extraer consecuencias aplicables al campo de las ciencias sociales. La problemática política es la siguiente: consideremos un sistema social cuyos componentes tienen tendencia ya sea a cooperar o a oponerse mutuamente. El sistema social puede consistir en firmas que hacen alianzas para imponer sus propios standards o normas, democracias en las cuales existen clivajes sociales, parlamentos donde los miembros forman coaliciones, etc. Cada par de componentes i, k (firmas, individuos, parlamentarios, etc) tiene una afinidad o propensión a cooperar w_{ik} que depende de sus historias, ideologías, centros de interés, etc. Cuanto más positivo es w_{ik} más grande es la tendencia a colaborar. Supondremos que esas propensiones son simétricas, es decir, $w_{ik}=w_{ki}$, lo cual es plausible porque generalmente las fuentes de conflicto en una interacción social son las mismas para las dos partes que interactúan. Para simplificar, en lo que sigue supondremos que el sistema está constituído por individuos que deben decidir si cooperan o no ante una situación dada. La formulación matemática que presentamos se inspira del trabajo de Cont y Loewe (2003), quienes propusieron (aparentemente de manera independientemente) un modelo similar al de Axelrod.

Supongamos pues que los individuos deben decidir qué actitud adoptar ante una situación estratégica, por ejemplo, votar por o contra una ley en el parlamento. En tales circunstancias, en una democracia ideal, los parlamentarios observarían quiénes tienen intereses afines y formarían alianzas que se plasmarían en votos.

La afinidad, es decir la intensidad de la interacción social, entre un individuo i y un individuo k está representada por los valores de w_{ik}. Para determinarlos supongamos que existen M cuestiones o temas que los miembros del sistema consideran importantes. Por ejemplo, la religión, el nivel de estudios o la universidad a la que asistieron, la raza, el origen social de los padres, el género, la edad, si fuma, etc. Formalmente, cada cuestión es formulada como una pregunta, a la cual cada individuo responde afirmativamente (+1) o negativamente (-1). Por ejemplo, ¿es usted fumador? Toda característica social se puede transformar en una variable binaria. Por ejemplo, para la edad podríamos considerar separadamente diferentes clases: menor de 30 años, entre 30 y 50, etc. con respuestas ±1 a cada clase.

El modelo supone que el valor de la afinidad w_{ik} es proporcional al número de cuestiones en las que i y k tienen una opinión similar *menos* el número en las que i y k tienen opiniones antagónicas. Denotemos μ ($1\leq\mu\leq M$) cada posible tema o cuestión. Sea $\xi_i^\mu \in \{+1,-1\}$ la opinión de i con respecto al tema μ. El producto $\xi_i^\mu\xi_k^\mu$ vale 1 si las opiniones de i y k coinciden, y -1 si son opuestas. Entonces las tendencias a cooperar w_{ik} están dadas por la suma de los productos $\xi_i^\mu\xi_j^\mu$ sobre todos los temas considerados. Esto no es más que otra manera de enunciar la regla de Hebb (6).

Supongamos ahora que se plantea una nueva cuestión estratégica. Bennett y Axelrod se interesaron en las coaliciones entre países durante la segunda guerra mundial, y usaron para calcular los valores de los w_{ik} una serie de criterios, como la religión dominante, la existencia de conflictos de fronteras, la similitud en los tipos de gobierno, etc. Los autores afirman que el modelo permite explicar los alineamientos en países del Eje y Aliados.

En este artículo consideraremos un caso más general. Supongamos por ejemplo que cada individuo debe votar en un referéndum si está a favor (σ_i=+1) o en contra (σ_i=-1) de la constitución europea. Podemos suponer que ante tal pregunta los individuos tienen opiniones a priori. Ese sería el estado inicial del sistema. Luego discuten, pesando las opiniones de unos y otros de acuerdo a los valores de w_{ik}

correspondientes. Por ejemplo, si $w_{ik}<0$ la influencia de la opinión σ_k del individuo k sobre el individuo i tomará la forma $w_{ik}\sigma_k$, de modo que si $\sigma_k=+1$ el individuo i tendrá tendencia a adoptar la opinión opuesta a la de k ($\sigma_i=-1$), y esa tendencia será más fuerte si el valor absoluto de w_{ik} es grande. Teniendo en cuenta todas las influencias sociales, el voto de i tendrá el signo de la suma ponderada dada por la ecuación (2), que es el potencial sobre i en el modelo de Hopfield. Como resultado, la población se polariza en dos grupos: aquellos que adoptan $\sigma=1$ y aquellos que adoptan $\sigma=-1$. La analogía con el modelo de Hopfield es perfecta. Más aún, la simetría en los valores w_{ik} está plenamente justificada, ya que en general las afinidades o tendencias a cooperar son recíprocas.

Se puede tener en cuenta que los individuos no votan necesariamente según lo indica la suma ponderada de las influencias sociales. Para ello introducimos un parámetro de ruido T, como en los modelos de memoria. Si el potencial tiene un valor próximo a cero, la decisión será +1 o -1 con una probabilidad del orden de ½, como lo indica la figura 2, reflejando el hecho que en ese caso el individuo no tiene fuertes razones para tomar una u otra decisión. En cambio, si el potencial tiene un gran valor absoluto, los individuos adoptarán preferentemente la decisión que corresponde al signo del potencial.

Finalmente notemos que las sucesivas actualizaciones de los σ_i a lo largo del tiempo, dadas por la dinámica (4), se pueden interpretar como sucesivos sondeos que permiten a los individuos adaptar sus opiniones a las opiniones de los otros.

Veamos qué podemos deducir sobre el comportamiento social ante una nueva cuestión, a partir de nuestro conocimiento matemático del modelo. Sabemos que si la fracción M/N de temas que determinan los valores de las influencias sociales w_{ik} es menor que la capacidad $\alpha_{max}(T)$, muy probablemente el sistema va a converger a uno de los estados ξ^μ. Las opiniones se polarizarán entre los que están a favor y los que están en contra de la cuestión μ, que no es para nada la cuestión planteada. Esto refleja lo que sucede cuando las decisiones individuales son guiadas, por ejemplo, por razones comunitarias en lugar de ser respuestas guiadas por una verdadera reflexión sobre la nueva cuestión. Es decir, cuando $\alpha<\alpha_{max}$, la respuesta de la sociedad refleja divisiones previas (uno de los estados que sirvieron a determinar las afinidades), en lugar de ser una respuesta ante la nueva pregunta planteada.

En cambio, si $\alpha=M/N>\alpha_{max}(T)$, y si el ruido es inferior a $T_c(\alpha)$ (ver figura 4), las opiniones se van a polarizar sobre uno de los estados espurios $\boldsymbol{\sigma}$, no correlacionado con los estados ξ^μ, pero en el cual la opinión de cada individuo i es coherente con el potencial h_i, es decir, con las opiniones de los otros ponderadas por los coeficientes de afinidad w_{ik}.

Desde un punto de vista social, éste es el comportamiento más deseable. De acuerdo con el modelo, sólo se logra si las interacciones entre individuos están basadas en una cantidad suficientemente grande de temas distintos ($\alpha>\alpha_{max}$). La propiedad de pérdida catastrófica de memoria, indeseable en el modelo de Hopfield, es la que asegura que el sistema social podrá funcionar satisfactoriamente, gracias a la diversidad de los individuos que la componen.

Para concluir, podemos citar el libro de Amin Maalouf (1998), *Les identités meutrières*, traducido al castellano (Identidades asesinas, 1999), que denuncia la

locura que incita a los hombres a matarse entre sí en nombre de una etnia, lengua, religión o color de piel. En su libro, Maalouf escribe: "*Se debería animar a todo ser humano a que asumiera su propia diversidad, a que entendiera su identidad como la suma de sus diversas pertenencias en vez de confundirla con una sola, erigida en pertenencia suprema y en instrumento de guerra.*" Lo cual es exactamente la conclusión del modelo presentado.

5. Agradecimientos

Es un placer agradecer a Sonia Kandel el haberme dado la oportunidad de escribir este artículo en mi lengua materna, y a Luis Kandel y Maribel Chenin por haberme ayudado a corregir los galicismos que se habían deslizado en el texto a pesar mío. Finalmente agradezco a Roberto Calemczuk que, gracias a su lectura implacable, me ayudó a corregir algunas imprecisiones.

6. Bibliografía

Adrian (1928). *The Basis of Sensation*. (ver *Nobel Lectures, Physiology or Medicine* 1922-1941, Amsterdam: Elsevier Publishing Company, 1965).

Amit D.J., Gutfreund H. & Sompolinsky H. (1987). Statistical mechanics of neural networks near saturation. *Annals of Physics*, 173(1) 30-67.

Anderson P. W. (1972). More is different. *Science*, 177, 393-396.

Axelrod R. y Bennett D. S. (1993). A Landscape Theory Of Aggregation. *British Journal of Political Science* 23(2) 211-233.

Ball Ph. (2003). The Physical Modeling Of Human Social Systems. *Complexus* 1, 190-206.

Cont R. & Loewe M. (2003). *Social distance, heterogeneity and social interactions*. Technical Report N° 505, Centre de Mathématiques Appliquées.

Galvani, 1791 (ver http://en.wikipedia.org/wiki/Galvani).

Gordon M.B. (1987) Memory capacity of neural networks learning within bounds. *J.Physique*, 48 2053.

Gross C. G. (2002) Genealogy of the "Grandmother Cell". *Neuroscientist* 8(5), 512-518 (ver también http://en.wikipedia.org/wiki/Grandmother_cell).

Hopfield J. J. (1982). Neural networks and physical systems with emergent collective computational properties. *Proc. Natl. Acad. Sci. USA* 79, 2554-2588.

Ising E. (1924). ver artículo en Wikipedia (http://en.wikipedia.org/wiki/Ernst_Ising).

Little W. A. (1974). The existence of persistent states in the brain. *Mathematical biosciences*, 19, 101-120.

Maalouf A. (1998). *Les identités meurtrières*. Paris: Grasset. (*Identidades asesinas*, 1999. Madrid: Editorial Alianza).

McCulloch W. & Pitts W. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics* 5, 115-133.

Minsky M. & Papert S. (1969). *Perceptrons. An Introduction to Computational Geometry*. Cambridge, Mass: MIT Press.

Peretto P. (1992) *An introduction to the modeling of neural networks*. Cambridge University Press.

Ramón y Cajal S. (1906). The structure and connexions of neurons. *Nobel Lecture*, http://nobelprize.org/nobel_prizes/medicine/laureates/1906/cajal-lecture.html.

Rosenblatt F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. NY: Spartan Books.

Sherrington D. & Kirkpatrick S. (1975). Solvable model of a Spin-Glass. *Physical Review Letters* 35(26), 1792-1796.

Schelling T. (1978). *Micromotives and macrobehavior*. Norton.

Turing (1950). Computing machine and intelligence. *Mind*, 49, 433-460.

von Neumann (1958). *The computer and the brain*. Yale University Press.

La autora

Mirta B. Gordon es Directora de Investigaciones en el CNRS. Física de formación, estudia las redes neuronales y sus capacidades de aprendizaje desde fines de los años 80. Estudió en particular cómo varía la capacidad del modelo de Hopfield cuando la memorización se realiza con diferentes algoritmos de aprendizaje. Usando técnicas de física estadística, contribuyó al estudio de la capacidad de aprendizaje y generalización del Perceptrón, las Máquinas de Vectores Soporte y otros algoritmos de discriminación y clasificación. Actualmente dirige el equipo AMA (Aprendizaje: Modelos y Algoritmos) en el Laboratorio TIMC-IMAG de Grenoble (Francia). Además de los algoritmos de aprendizaje y sus aplicaciones, se interesa en la modelización de sistemas sociales teniendo en cuenta las interacciones y capacidades cognitivas de los individuos. También enseña la modelización de sistemas complejos en las escuelas doctorales de Grenoble y Lyon.