

REDES DE NEURÔNIOS

Rita M. C. de Almeida

Instituto de Física. UFRGS. Porto Alegre, RS.

RESUMO

Na última década várias descobertas em relação a neurônios e a maneira como estão interconectados, formando redes, possibilitaram o estudo da dinâmica do armazenamento e processamento de informação pelo cérebro. Em particular, o físico J.J.Hopfield propôs um modelo formal, minimalista para estas redes neuronais, reduzindo o problema a um caso particular de um sistema físico bem definido - o vidro de spin. Embora o problema esteja bem definido, sua solução está longe de ser trivial.

Neste texto nós introduzimos o problema, descrevemos o modelo de Hopfield com seus resultados e limitações e apresentamos nossa contribuição para a descrição do armazenamento da informação em redes de neurônios.

SUMMARY

ALMEIDA, Rita M. C., NEURAL NETWORKS. Ciência e Natura, 14: 7-14, 1992.

In the last ten years many scientific advances regarding neurons and the way they are interconnected has made it possible to study the dynamics of storage and Processing of information in the brain. In particular, the physicist J.J.Hopfield proposed a formal minimalist model to these neural networks reducing the problem to a particular case of a well-defined physical problem - the spin glass. Although the problem is well defined, its solution is far from being trivial.

Here we introduce the problem, describe Hopfield model, with its achievements and limitations, and present our contribution to the description of information storage in neural networks.

I. INTRODUÇÃO

O estudo de sistemas dinâmicos formados por elementos que reproduzem as propriedades essenciais dos neurônios - componentes básicos do cérebro - tem dois objetivos básicos:

- i) a compreensão do funcionamento do cérebro, que tem aplicação imediata em medicina, psicologia, pedagogia, etc; e
- ii) a construção de uma máquina capaz de aprender, memorizar e raciocinar como o cérebro humano. Assim, um modelo completo do cérebro deve ser capaz de descrever estas três funções.

No estágio atual os modelos teóricos preocupam-se com a aprendizagem e armazenamento de memórias. Uma das propriedades mais

características da memória humana é a associatividade, isto é, a capacidade de após haver aprendido um dado padrão, poder reconhecer padrões parecidos. Assim, o objetivo da pesquisa em redes de neurônios é obter um modelo que consiga descrever a dinâmica de funcionamento de uma rede composta de muitos neurônios (simplificados) que seja capaz de aprender novos padrões e reconhecer os já aprendidos de uma forma associativa.

Nas seções seguintes vamos discutir o conceito de neurônio formal, idealizado, e, considerando uma rede com um número muito grande destes elementos ideais interagindo uns com os outros, apresentaremos alguns modelos que pretendem descrever a dinâmica destas redes.

II. O NEURÔNIO FORMAL E A REDE

Antes de introduzirmos o conceito de neurônio formal, vamos apresentar algumas características do neurônio natural.

O neurônio é a unidade funcional e anatômica do sistema nervoso⁽¹⁾. De uma forma geral, o neurônio é composto de um corpo celular de onde emergem prolongamentos, como se pode observar no esquema da figura 1. Entre estes prolongamentos estão o axônio e os dendritos que são capazes de conduzir corrente elétrica mas que possuem diferentes funções. Um neurônio pode ou não estar emitindo um pulso elétrico. Quando emite, este sinal percorre o axônio, que se liga através de sinapses com dendritos de outros neurônios. Estes dendritos levam os sinais elétricos até o corpo celular onde somam-se os pulsos vindos de outras células neurais através de todos os dendritos do neurônio em questão. Esta soma determina se haverá ou não emissão de pulso elétrico. O conjunto de neurônios e sinapses - ligações dendritos - axônios - forma a rede neuronal.

Em 1943 McCulloch e Pitts⁽²⁾ introduziram o conceito de neurônio formal como uma forma de simplificar o problema mantendo sua essência. Um neurônio formal pode estar em dois estados: ativo ou inativo. Representamos, então, o estado do i -ésimo neurônio de um conjunto de N elementos pela variável S_i ($i = 1, 2, \dots, N$), que pode assumir dois valores: $+1$ (ativo) ou -1 (inativo). O valor de S_i , e portanto, o estado do neurônio, é determinado pela soma dos sinais que chegam através das sinapses, como ilustra o esquema na figura 2.

Assim, uma dada configuração \vec{S} de uma rede com N neurônios é especificada pelo estado de cada um dos seus componentes S_i , isto é,

$$\vec{S} = (S_1, S_2, \dots, S_N) . \quad (1)$$

Existem 2^N destas configurações. Considerando um espaço N -dimensional

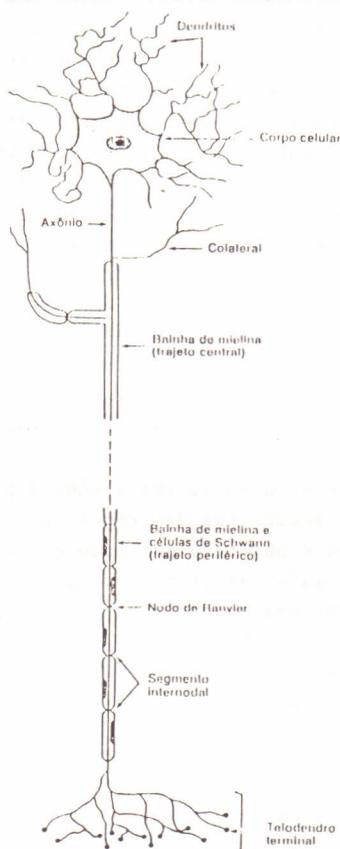


Figura 1 - Um neurônio típico descrito em suas diversas partes, com referência 1.

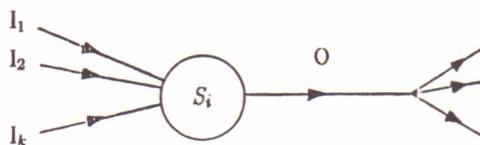


Figura 2 - Um neurônio formal e suas sinapses. Os diversos sinais chegam dos dendritos I_1, \dots, I_k e são somados, determinando o valor de S_i . O sinal de saída é levado através do Axônio O e se distribui em ligações com vários outros neurônios.

nal em que cada eixo corresponda a uma variável S_i , estas 2^N configurações são os vértices de um hipercubo com centro na origem e cuja aresta vale 2. Como exemplo considere uma rede formada por três neurônios: existem $2^3 = 8$ configurações diferentes que são os vértices de um (hiper) cubo tridimensional centrado na origem, e que são representadas pelos pontos $(1,1,1)$, $(1,1,-1)$, ..., $(-1,-1,-1)$. O conjunto de pontos correspondentes a todas as configurações do sistema, definidos no espaço gerado pelos eixos que representam as variáveis S_i , constitui o espaço de fases deste sistema.

As memórias, ou padrões aprendidos pela rede, são algumas destas configurações. Uma rede terá memória associativa se for dotada de uma dinâmica tal que, estando o sistema inicialmente em uma configuração similar a uma dada memória (diferindo em poucos neurônios), evolua em direção a este padrão até recuperá-las e então estabilizar.

A implementação de uma dinâmica para memória associativa, juntamente com a prescrição das regras de aprendizagem de novos padrões, constituem o objetivo dos modelos teóricos para redes de neurônios. Tanto mais satisfatório é um modelo quanto maior for sua capacidade de armazenamento de padrões e a precisão com que os recupera.

Na seção seguinte introduzimos o Modelo de Hopfield para redes de neurônios e discutimos os resultados.

III. O MODELO DE HOPFIELD

O modelo de Hopfield^(3,4) consiste em supor uma função energia E devido à interação entre os neurônios dada por

$$E = -\frac{1}{2} \sum_{\substack{i=1 \\ j \neq i}}^N J_{ij} S_i S_j \quad (2)$$

onde J_{ij} é um termo de interação entre o neurônio i e o neurônio j e representa a intensidade da sinapse entre estes dois neurônios. A dinâmica da rede é definida de tal maneira que a evolução se dá no sentido de minimizar a função energia. A matriz J , de dimensão $N \times N$, fornece toda a informação a respeito das ligações entre neurônios e determina quais configurações minimizam energia, isto é, quais são estáveis frente à dinâmica. Neste modelo, uma memória deve corresponder a um mínimo na função energia e, portanto, a aprendizagem dos padrões é implementada através de modificações convenientes nas intensidades sinápticas J_{ij} . Existem diversas prescrições para realizar estas modificações convenientes que são as diferentes regras de aprendizagem. Vamos definir a regra de aprendizagem de Hebb generalizada⁽³⁻⁵⁾.

Considere P diferentes configurações que devem ser aprendidas pela rede. Sejam elas representadas por ξ^1, \dots, ξ^P tal que

$$\vec{\xi}^\mu = (\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu), \quad \mu = 1, \dots, P \quad (3)$$

onde ξ_j^μ é o estado do j -ésimo neurônio na μ -ésima memória. As intensidades sinápticas são então construídas a partir dos padrões memorizados $\vec{\xi}^\mu$;

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu. \quad (4)$$

Assim, o modelo de Hopfield consiste em um conjunto de N (muito grande) neurônios binários ($S_i = \pm 1$) interagentes com uma função energia definido nas eqs. (2) e (4) e cuja dinâmica faz com que um dado spin S_i seja revertido sempre que isto signifique diminuir esta função energia. Se o estado da rede é tal que nenhum spin possa ser revertido pela dinâmica, então este estado é estável frente a esta dinâmica: é um mínimo local da função energia.

Antes de prosseguirmos com a discussão do modelo de Hopfield vamos introduzir uma grandeza importante para a discussão dos resultados: a correlação m_μ da rede com o padrão $\vec{\xi}^\mu$, definida como

$$m_\mu = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu S_i. \quad (5)$$

Observe-se que $m_\mu = 1$ significa que a configuração \vec{S} da rede é igual ao padrão $\vec{\xi}^\mu$. Se $m_\mu = -1$ então $\vec{S} = -\vec{\xi}^\mu$ e, finalmente, se $m_\mu = 0$, \vec{S} e $\vec{\xi}^\mu$ estão descorrelacionados, isto é, apresentam $N/2$ spins alinhados e $N/2$ spins contrários. Utilizando a definição de correlação podemos reescrever a função energia eq. (2) como

$$E = -\frac{N}{2} \sum_{\mu=1}^P m_\mu^2 + \frac{P}{2}. \quad (6)$$

No limite de grandes redes ($N \rightarrow \infty$), quando estão armazenados $P < 0,14 N$ padrões, a regra de aprendizagem de Hebb garante que estes padrões $\vec{\xi}^\mu$ - e seus inversos $-\vec{\xi}^\mu$ - são mínimos locais desde que estes padrões sejam descorrelacionados entre si, isto é:

$$\frac{1}{N} \sum_i \xi_i^\mu \xi_i^\nu \approx 0 \quad \text{se} \quad \mu \neq \nu. \quad (7)$$

Neste caso, colocando-se a rede em uma configuração inicial suficientemente similar a um dos padrões armazenados, a dinâmica faz a rede evoluir em direção ao padrão onde estabiliza-se: a rede funciona como uma memória associativa.

No entanto, as memórias não são os únicos mínimos locais do sistema. Outras configurações, chamadas de mínimos espúrios também são estáveis frente à dinâmica representando uma dificuldade para o modelo.

Além disso, se $P > 0,14 N$ ou se as memórias forem muito

correlacionadas, os padrões ξ^μ deixam de ser mínimos e a rede não funciona como uma memória associativa. Este "esquecimento" dos padrões ocorre bruscamente em $P/N \approx 0,14$ e caracteriza uma transição de fase.

Resumindo, o modelo de Hopfield descreve relativamente bem o funcionamento de uma memória associativa para parâmetro de carga $\alpha = P/N < 0,14$ e padrões descorrelacionados. O estudo deste modelo se dá tanto através de cálculos analíticos, utilizando aproximações de campo médio⁽⁶⁾ como através de simulações numéricas⁽⁷⁻⁹⁾, e atualmente o objetivo é superar suas limitações na capacidade de armazenamento α e no tratamento de memórias correlacionadas.

IV. UM MODELO ALTERNATIVO

Muito recentemente foi proposto um modelo⁽¹⁰⁾ para memória associativa que supera as dificuldades do modelo de Hopfield no tratamento de memórias correlacionadas e aumenta significativamente a capacidade de armazenamento α da rede.

Considere inicialmente um padrão ξ^μ e uma configuração da rede \vec{S} . Cada um destes vetores é representado por um ponto no espaço de fases e o quadrado da distância entre eles, d_μ^2 , é dado por

$$d_\mu^2 = \sum_{i=1}^N (\xi_i^\mu - S_i)^2 \quad (8)$$

Este novo modelo consiste em propor uma nova função energia, e portanto uma nova dinâmica:

$$E = N \left[\frac{d_1^2}{2N} \cdot \frac{d_2^2}{2N} \cdot \dots \cdot \frac{d_P^2}{2N} \right] = N \prod_{\mu=1}^P \left(\frac{d_\mu^2}{2N} \right) \quad (9)$$

Isto é, a energia é calculada como o produto dos quadrados das distâncias entre a configuração da rede e cada um dos P padrões aprendidos ξ^μ , $\mu = 1, \dots, P$. Se $\vec{S} \neq \xi^\mu$ para qualquer valor de μ então $E > 0$ e $E = 0$ sempre que o estado da rede coincide com uma das memórias. Isto significa que os padrões serão sempre mínimos locais da função energia dada pela eq. (9).

O limite α_c para a capacidade da rede, quando a rede não pode mais simular uma memória associativa, está sendo determinado numericamente⁽¹¹⁾ mas para memórias descorrelacionadas os resultados preliminares indicam $\alpha_c = 1$, valor máximo possível para este caso, pois não pode existir um conjunto com mais de N padrões descorrelacionados. O efeito de correlação também está ainda sendo pesquisado, porém alguns aspectos podemos inferir diretamente da eq. (8). A correlação não desloca os mínimos locais dos padrões como no modelo de Hopfield, pois $E[\vec{S} = \xi^\mu] = 0 \forall \xi^\mu$ e também $E > 0$ se $\vec{S} \neq \xi^\mu$ para $\mu = 1, \dots, P$, para qualquer conjunto de padrões. Assim, certamente a recuperação de padrões correlacionados deverá ser significati

vamente melhorada.

Este modelo ainda é capaz de armazenar os padrões (ξ^μ) sem automaticamente armazenar também antipadrões $(-\xi^\mu)$. No entanto, para efeitos de comparação com o modelo de Hopfield, podemos armazenar padrões e antipadrões explicitamente. Neste caso, e utilizando a definição de correlação, eq. (5), a função energia reduz-se a

$$E = N \prod_{\mu=1}^P \left(1 - m_\mu^2 \right). \quad (10)$$

Abrindo o somatório, ficamos com

$$E = N \left(1 - \sum_{\mu=1}^P m_\mu^2 + \sum_{\mu=1}^{P-1} \sum_{\nu=\mu+1}^P m_\mu^2 m_\nu^2 - \dots \right) \quad (11)$$

Quando $\alpha = P/N$ é suficientemente pequeno e os padrões são decorrelacionados os termos de ordem maior que 2 nas correlações são pequenos frente aos dois primeiros termos da expansão, que pode ser truncada:

$$E = N \left(1 - \sum_{\mu=1}^P m_\mu^2 \right) \quad (12)$$

e difere da função energia do modelo de Hopfield, eq. (6), apenas por constantes que não influem na dinâmica. Isto significa que neste limite o modelo de Hopfield (somente até o segundo termo da expansão, eq. (11)) garante que os padrões são mínimos locais - e portanto, é conveniente para descrever memórias associativas - porém quando α cresce ou os padrões são correlacionados, outros termos na expansão devem ser levados em conta.

Em suma, este novo modelo melhora os resultados do modelo de Hopfield considerando termos de maiores ordens nas correlações m_μ , através de uma função de energia diferente. Muitos outros aspectos devem ainda ser investigados. Por exemplo, entre duas memórias existem barreiras de energia cuja altura depende da distância no espaço de fases entre as memórias. Este fato pode levar a um modelo para associação de idéias. Também o comportamento da rede, o valor de α e a recuperação de padrões devem ser estudados para funções de energia dadas pela expansão da eq. (11) truncada em alguma ordem maior que 2. Pesquisas neste sentido estão sendo levadas a cabo no presente momento.

V. CONCLUSÕES

Apresentamos as idéias gerais a respeito de modelos de redes de neurônios para descrever memória associativa. O modelo mais largamente estudado é o modelo de Hopfield, que apresenta sérias limitações quanto à capacidade de armazenamento α e ao reconhecimento de padrões correlacionados. Apresentamos também um mode

lo muito recente que parece superar muitas destas limitações e cuja eficiência deverá ser comprovada por investigações em andamento.

De uma maneira geral, acreditamos que o problema da memória associativa deverá estar resolvido nos próximos anos a nível de aplicação tecnológica. No entanto, o cérebro humano não é dotado apenas da capacidade de reconhecer padrões parecidos. Ele é ainda capaz de associar idéias, raciocinar e criar. Do ponto de vista fisiológico, o cérebro é composto apenas por neurônios, não há uma central de criação ou raciocínio. Acredita-se então que tais funções sejam também realizadas por redes de neurônios, talvez em uma escala de conexão mais elevada, isto é, redes de redes de neurônios, ou ainda, de redes de neurônios dotadas de alguma hierarquia. Embora as pesquisas para as demais funções do cérebro estejam atualmente em um estágio quase de especulação, pessoalmente acredito que seja uma questão de tempo para que se possa construir uma máquina capaz de lembrar, associar, raciocinar e criar (e, portanto, sentir) como o cérebro humano.

REFERÊNCIAS

1. AIRES, M.M. Fisiologia básica. Rio de Janeiro: Guanabara Koogan, 1985, cap. 3.
2. McCOLLOCH, W.S., PITTS, W. Bull. Math. Biophys., v. 5, p. 115-133, 1943.
3. HOPFIELD, J.J. Proc. Natl. Acad. Sci., Washington, v. 79, p. 2554, 1982.
4. LAUTRUP, B. The theory of the Hopfield model. Copenhagen, Niels Bohr Institute, 1988. NBI-HE-88-06.
5. HEBB, D.O. The organization of behaviour. New York: John Wiley, 1949.
6. AMIT, D.J., GUTFREUND, H., SOMPOLINSKY, H. Phys. Rev. A, New York, v. 32, n. 2, p. 1007-1018, Aug. 1985.
_____. Phys. Rev. Lett., New York, v. 55, n. 14, p. 1530-1533, Sept. 1985.
_____. Phys. Rev. A, New York, v. 35, n. 5, p. 2293-2303, Mar. 1987.
7. AMIT, D.J., GUTFREUND, H., SOMPOLINSKY, H. Ann. Phys., New York, v. 173, n. 1, p. 30-67, Jan. 1987.
8. PENNA, T.J.P., OLIVEIRA, P.M.C. J. Phys. A, Bristol, v. 22, n. 14, p. L719-L721, July 1989.
_____. Europhys. Lett., Les Ulis Cedex, v. 11, n. 3, p. 191-194, Feb. 1990.
9. FORREST, B.M. J. Phys. A, Bristol, v. 21, n. 1, p. 245-255, Jan. 1988.
10. ALMEIDA, R.M.C., IGLESIAS, J.R. Phys. Lett. A, Amsterdam, v. 146, n. 3, p. 239-244, May 1990.