

VIDROS DE SPINS EM REDES NEURAIS

Aldo TANCREDO*

José Roberto CAMPANHA*

- RESUMO: A conexão vidros de spins – redes neurais é apresentada procurando interpretar os parâmetros de ordem resultantes da aplicação dos métodos de mecânica estatística.
- UNITERMOS: Redes neurais; vidro de spins.

Introdução

Nos últimos anos tem aumentado o interesse dos físicos teóricos, principalmente daqueles ligados à área de mecânica estatística, no estudo de redes neurais. Isso se deve ao fato de que as redes neurais tornarão possível, num futuro próximo, a criação de modelos físicos que simulem as várias funções do cérebro humano. Outro fato muito importante para o desenvolvimento das redes neurais foi a descoberta de sistema de spins, com uma fase extremamente interessante e singular denominada de vidros de spins^{3,4}. Os vidros de spins têm excitado e desafiado os físicos por quase uma década a apresentarem um modelo teórico que reproduzisse suas propriedades físicas^{1,3}. Desse desafio surgiram novos problemas^{1,1}, novos métodos e novas soluções para velhos problemas.

O propósito de nosso trabalho é mostrar de maneira simples a semelhança entre os sistemas de spins que apresentam a fase de vidro de spins e as redes neurais.

Redes neurais

Existem aproximadamente no cérebro humano¹² 10^{12} células denominadas neurônios. Os neurônios são constituídos de corpo, axônio e dendritos. Operacionalmente

* Departamento de Física – Instituto de Geociências e Ciências Exatas – UNESP – 13500-230 – Rio Claro – SP.

os neurônios são receptores e transmissores de estímulos elétricos. O corpo e os dendritos servem para receber e transmitir os estímulos a outros neurônios, e o axônio atua como condutor desses estímulos.

Individualmente e de forma bem simples, os neurônios respondem a estímulos de forma binária e seguem a regra do *tudo ou nada*, ou seja, podem estar ativos ou inativos. Essa condição possibilita associarmos a cada neurônio dois estados: $S_i = +1$ se estiver ativo e $S_i = -1$ se estiver inativo.

A dinâmica da rede neural é muito simples: o neurônio recebe estímulos elétricos de outros neurônios que podem excitá-lo ou inibi-lo, e, se a soma resultante atingir um certo valor limiar (-70 mV), ele dispara e o estímulo propaga-se a outro neurônio através das sinapses. As sinapses são as interconexões entre os neurônios, sendo que um neurônio pode ter até 10^4 sinapses.

Com estas informações podemos definir uma rede neural como um conjunto de neurônios conectados uns aos outros através das sinapses e atuando coletivamente¹⁰.

Vidros de spins

Algumas ligas metálicas, como por exemplo AuFe (1% a 8% em Fe), apresentam uma brusca mudança na susceptibilidade magnética a uma determinada temperatura T_f , que depende da concentração de Fe na liga.

Tal fato sugere a existência de uma transição de fase termodinâmica. Assim, para explicar essa transição, supõe-se que abaixo da temperatura T_f os spins S_i , $i=1..N$, dos átomos de Fe, orientados aleatoriamente, ficam numa configuração $\{S_i(t)\}$, a qual permanece inalterada no decorrer do tempo, dando origem a uma fase termodinâmica da liga metálica denominada vidro de spins. Um dos resultados dos modelos teóricos propostos foi um parâmetro de ordem temporal definido por:

$$q = \lim_{t \rightarrow \infty} \ll S_i(0) S_i(t) \gg \quad (1)$$

onde o conjunto de chaves interno representa a média térmica, enquanto o conjunto externo representa a média sobre todos os spins.

Devido às interações entre os spins dos átomos de Fe serem do tipo ferromagnéticas e antiferromagnéticas e haver uma competição entre essas interações, ocorre o fenômeno da frustração magnética¹⁴, o qual causa o aparecimento de um grande número de configurações, todas com a mesma energia.

Um dos modelos propostos para explicar as propriedades dos vidros de spin foi desenvolvido por Kirkpatrick & Sherrington⁸, os quais aplicaram métodos de mecânica estatística, usando a teoria do campo médio e o método de réplicas⁵. O modelo prevê,

através de equações autoconsistentes para os parâmetros de ordem m (magnetização) e q , as seguintes fases: a) $m = 0$, $q = 0$ fase paramagnética, b) $m \neq 0$, $q = 0$ fase ferromagnética e c) $m = 0$, $q \neq 0$ fase de vidro de spins.

Dos vidros de spins às redes neurais

A semelhança entre um sistema de spins e redes neurais apareceu pela primeira vez em 1974, em um trabalho de Little⁶. Ele comparou os neurônios a um sistema de spins de Ising, onde o neurônio i é representado por um spin S_i e assume os valores $+1$ se ativo e -1 se inativo.

A analogia mais profunda e consequente foi feita em 1982 quando J. J. Hopfield⁷ sugeriu a semelhança entre uma rede neural e os vidros de spins. Posteriormente, em 1985, Amit et al.¹ aplicaram no modelo de redes neurais de Hopfield os métodos de mecânica estatística usados em vidros de spins. O sucesso desse trabalho despertou na comunidade de físicos o interesse pelo problema das redes neurais.

Para um sistema de N neurônios, onde cada neurônio S_i pode ter os valores aleatórios $S_i = \pm 1$, a energia é dada pela equação

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

que corresponde a mesma equação de um sistema de spins que apresenta a fase de vidro de spins.

Na equação (2), J_{ij} representa o termo de acoplamento nos vidros de spins e, na rede neural, corresponde ao valor das sinapses entre o neurônio i e o neurônio j .

O estado de uma rede neural, definido como os valores assumidos por todos os neurônios, evolui no tempo através de uma dinâmica proposta por J. J. Hopfield⁷, ou

$$S_i(t+1) = \text{sign} \left(\sum_{j=1}^N J_{ij} S_j(t) \right) \quad (3)$$

onde a função $\text{sign}(x) = +1$ se $x > 0$ e $\text{sign}(x) = -1$ se $x < 0$.

Assim, à rede neural associamos uma superfície de energia (Figura 1), na qual um padrão inicial, $\{S_i(t)\}$ $i=1..N$, evoluirá no tempo seguindo uma trajetória que, a cada instante, reduz a sua energia indo finalmente alojar-se num ponto de mínimo dessa superfície. Esses pontos de mínimos podem atuar como locais, nos quais as informações estariam armazenadas na rede neural.

onde p é o número de padrões ensinados e os ξ_i^μ são as componentes desses padrões que assumem os valores aleatórios ± 1 . Dessa forma, as informações são codificadas por configurações dos ξ_i^μ e ficam armazenadas nos mínimos locais da superfície de energia da rede neural.

Dessa forma teríamos um modelo de memória na qual a rede aprenderia p padrões $\{\xi_i^\mu\}$, e esses padrões poderiam ser recuperados a partir de uma informação parcial que seguisse os passos estabelecidos pela dinâmica da rede. Esse tipo de memória é chamado de memória de acesso pelo conteúdo ou memória associativa.

A qualidade da recuperação de um padrão armazenado na rede é medida pela sobreposição denotada por $m^\mu(t)$, entre um estado da rede neural $\{S_i(t)\}$ e o padrão $\{\xi_i^\mu\}$, definido como

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

Também podemos definir uma medida da distância entre o padrão $\{\xi_i^\mu\}$ e o estado $\{S_i(t)\}$, o qual é denominado de distância de Hamming, ou

$$d^\mu = 1/2 (1 - m^\mu) \quad (6)$$

Mecânica estatística de redes neurais

Considerando uma rede neural com N neurônios interagindo entre si, a Hamiltoniana desse sistema pode ser escrita de maneira idêntica à dos vidros de spins no modelo de Kirkpatrick & Sherrington⁸, ou

$$H = - 1/2 \sum_{i,j=1}^N [1/N \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu] S_i S_j \quad (7)$$

onde p é o número de padrões ensinados à rede. Com a equação (7) e a função de partição do sistema, $Z = \text{Tr} \exp(-\beta H)$, calculamos a energia livre média de Helmholtz, usando o método das réplicas, no qual $\langle \ln Z \rangle$ é substituído por

$$\langle \ln Z \rangle = \lim_{n \rightarrow 0} \frac{\langle Z^n \rangle - 1}{n} \quad (8)$$

A função energia livre no limite termodinâmico, no caso do número de padrões permanecer finito, é expressa por

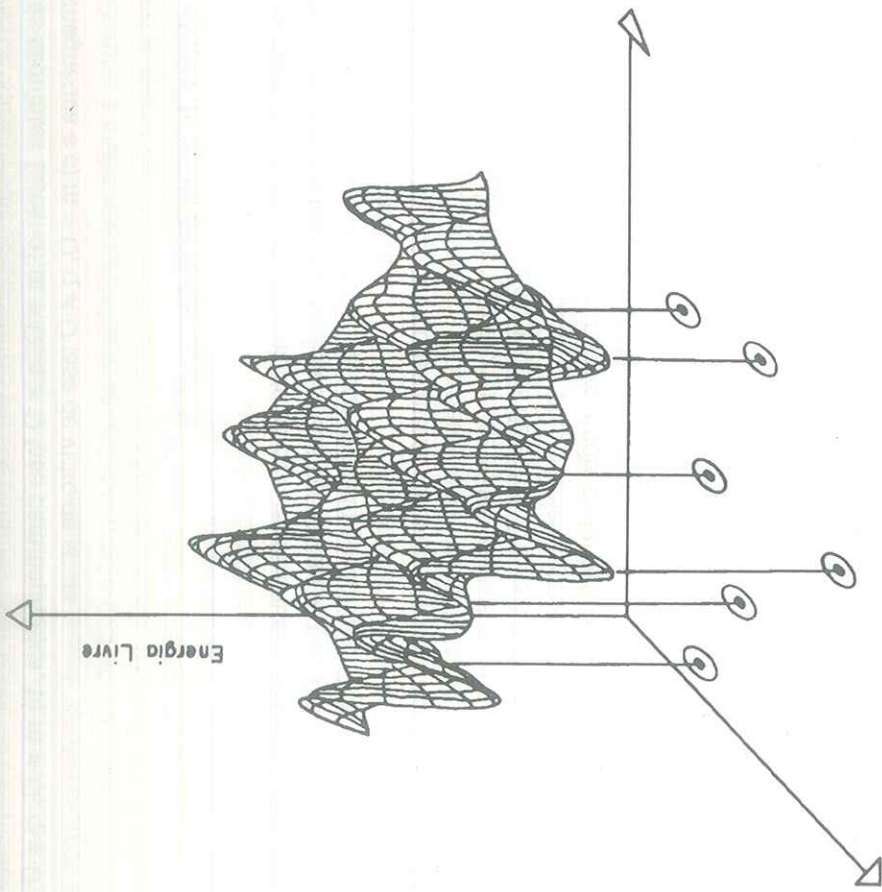


FIGURA 1 - Representação tridimensional da superfície da energia livre. Os \odot correspondem aos mínimos locais da energia, onde estão localizadas as memórias ξ_i^μ .

Outro fato importante é que nas redes neurais o termo J_{ij} , ou coeficiente sináptico, possui uma forma para sustentar uma hipótese mais realista devida a Hebb⁶, na qual: "Quando os neurônios estão ativos em seus padrões específicos, suas atividades induzem mudanças em seus coeficientes sinápticos, de maneira a reforçar a estabilidade daquele padrão de atividade". Essa hipótese estabelece uma regra de aprendizado traduzida na equação abaixo,

$$J_{ij} = \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu \quad (4)$$

$$\zeta(\beta) = 1/2 m^2 - \frac{1}{\beta N} \sum_{i=1}^N \ln [2 \cosh (\beta m \cdot \xi_i)] \quad (9)$$

$$\xi_i = (\xi_i^1, \xi_i^2, \xi_i^3, \dots, \xi_i^p) \quad (10)$$

onde $\beta = 1/T$ e m é o parâmetro de ordem, cujas componentes m^μ correspondem àquelas dadas na equação (5). O parâmetro m é obtido encontrando-se os pontos de mínimo da energia livre ou

$$m = 1/N \sum_{i=1}^N \xi_i \operatorname{tgh} (\beta m \cdot \xi_i) \quad (11)$$

No limite de N tendendo ao infinito, e fazendo a média sobre os $\{\xi_i\}$, obtemos

$$m = \langle \xi \operatorname{tgh} (\beta m \cdot \xi) \rangle > \quad (12)$$

Na interpretação das componentes m^μ , verificamos que as mesmas correspondem à sobreposição média entre $\langle S_i \rangle$, que nos vidros de spins é a magnetização local, e os ξ 's. Assim temos,

$$m^\mu = \langle \xi_i^\mu \langle S_i \rangle \rangle > \quad (13)$$

onde $\langle S_i \rangle = \operatorname{tgh} (\beta m \cdot \xi_i)$.

Outro parâmetro de ordem que pode ser definido por analogia aos vidros de spins é o parâmetro de ordem q ,

$$q = 1/N \sum_{i=1}^N \langle S_i \rangle^2 \quad (14)$$

Substituindo $\langle S_i \rangle$ e fazendo a média sobre os ξ_i , a equação (14) fica,

$$q = \langle \langle \xi \operatorname{tgh}^2 (\beta m \cdot \xi) \rangle \rangle > \quad (15)$$

Podemos definir a capacidade de armazenamento da rede pela relação $\alpha = p/N$. Um modelo² mais completo é desenvolvido considerando que a capacidade de armazenamento permanece finita, enquanto o número de neurônios N tende ao infinito. Nesse caso ocorre o que denominamos de saturação da rede.

Na saturação, a desvantagem é a introdução de um ruído na rede descrito por um parâmetro r , e com essas alterações as equações dos parâmetros de ordem ficam

$$m^\mu = \langle \langle \xi_i^\mu \operatorname{tgh} (\beta \langle m \cdot \xi \rangle + (\alpha r)^{1/2} z) \rangle \rangle > \quad (16)$$

$$q = \langle \langle \operatorname{tgh}^2 (\beta \langle m \cdot \xi \rangle + (\alpha r)^{1/2} z) \rangle \rangle > \quad (17)$$

$$r = 1/\alpha < \sum_{\mu > s} (m^\mu)^2 \quad (18)$$

onde na equação (18) o parâmetro r é definido a partir de $\mu > s$, pois corresponde aos padrões que não condensam, ou seja, não são recuperados, provocando flutuações e introduzindo ruídos na rede neural. Nestas equações, $m \cdot \xi$ corresponde à fase ferromagnética, representando os padrões memorizados pela rede, e $(\alpha r)^{1/2} z$ corresponde à fase de vidro de spins.

Dependendo da temperatura e da capacidade de armazenamento, o modelo fornece o diagrama de fase (Figura 2) de uma rede neural do tipo Hopfield.

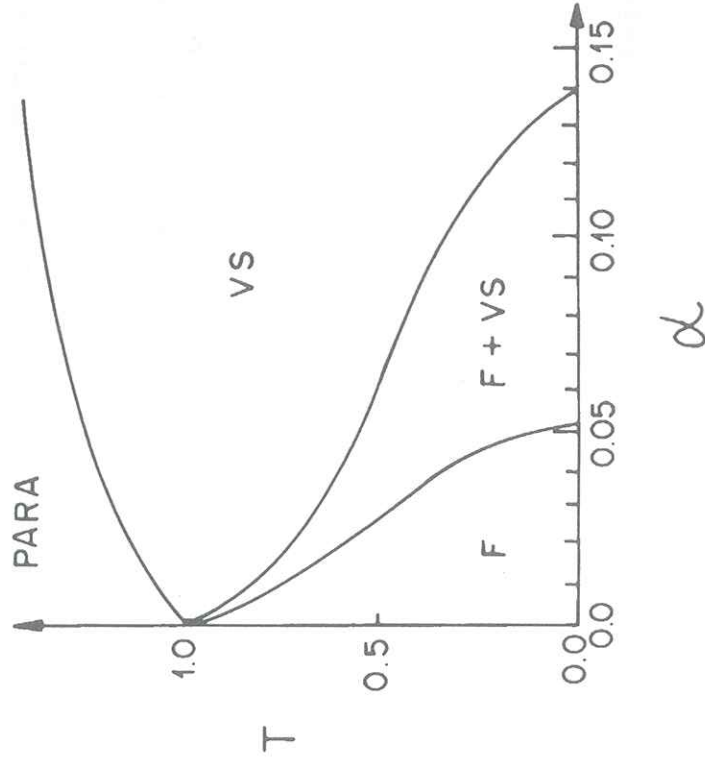


FIGURA 2 - Diagrama de fase de uma rede neural do tipo Hopfield. As fases são PARA-paramagnética, F-ferromagnética, VS-vidro de spins.

Analisando o diagrama de fase, a baixa temperatura ($T < T_c$), a região F é a fase ferromagnética, na qual os padrões armazenados são integralmente recuperados e representa propriamente a memória do sistema. No entanto, quando o número de padrões ensinados aumenta, passamos para uma região F+VS, onde temos uma mistura de fase ferromagnética com fase de vidro de spins, e a recuperação torna-se mais difícil devido ao ruído ocasionado pelo vidro de spins. Finalmente podemos atingir uma região VS de puro vidro de spins, onde nenhum padrão é recuperado.

A alta temperatura, em ($T > T_c$), somente persiste a fase paramagnética que corresponde ao caos da memória e nenhum padrão pode ser ensinado ou recuperado.

Conclusões

A rede neural, estudada com a aplicação dos métodos de mecânica estatística usados em vidro de spins, mostra resultados extraordinários. Primeiro, pela analogia com um sistema de spins, o qual apresenta fase de vidro de spins, cujos parâmetros de ordem determinam as diversas fases e estabelece fatores limitantes ao aprendizado e recuperação de padrões. Segundo, pela rica estrutura topológica apresentada pela superfície de energia, onde as memórias constituem pontos estáveis de mínimo local, ou melhor, as memórias são pontos atratores da rede neural.

Uma análise mais detalhada e rigorosa possibilita-nos também interpretar o que ocorre próximo às bacias de atração dos padrões, além de estados de misturas de padrões, padrões espúrios etc.

O estudo de redes neurais, no enfoque da mecânica estatística, está apenas começando e constitui um dos mais promissores ramos da física atualmente.

TANCREDO, A., CAMPANHA, J. R. Spin glass on neural networks. *Ecl. Quím.* São Paulo, v. 18, p. 19-27, 1993.

■ **ABSTRACT:** Our goal is to show the analogy spin glass - neural network and to interpret the order parameters that results from Statistical Mechanics methods.

■ **KEYWORDS:** Neural networks; spin glass.

Referências bibliográficas

1. AMIT, D. J., GUTTFREUND, H., SOMPOLINSKY, H. *Phys. Rev.*, v. 32A, p. 1 007, 1985.
2. ————. *Ann. Phys.*, v. 173, p. 30, 1987.

3. CANNELLA, V., MYDOSHI, J. A., BUDNICK, J. I. *J. Appl. Phys.*, v. 42, p. 1 689, 1971.
4. CHAKRAVARTY, A. S. Introduction to the properties of solids, New York: John Wiley, 1980.
5. EDWARDS, S. F., ANDERSON, P. W. *J. Phys. F.*, v. 5, p. 965, 1975.
6. HEBB, D. O. The Organization of Behaviour, New York: Wiley, 1949.
7. HOPFIELD, J. J. *Proc. Nat. Acad. Sci. USA*, v. 79, p. 2 554, 1982.
8. KIRKPATRICK, S., SHERRINGTON, D. *Phys. Rev.*, v. 17B, p. 4384, 1978.
9. LITTLE, W. A. *Math. Biosci.*, v. 19, p. 101, 1974.
10. McCULLOCH, W. S., PITTS, W. A. *Bull. Math. Biophys.*, v. 5, p. 115, 1943.
11. MEZARD, M., PARISI, G., VIRASORO, M. A. Spin glass theory and beyond. Ed. World Scientific, 1988.
12. SMITH, G. M. *The Brain*, G. P. New York: Putmanns, 1970.
13. TANCREDO, A., CAMPANHA, J. R. *Ecl. Quím.*, v. 15, p. 51, 1990.
14. TOULOUSE, G. *Commun. Phys.*, v. 2, p. 115, 1977.

Recebido em 4.9.1992.
Aceito em 14.12.1992.