PHYSICAL REVIEW A VOLUME 37, NUMBER 7 APRIL 1, 1988

Potts-glass models of neural networks

Ido Kanter*
Department of Physics, Bar-Ilan University, 52100 Ramat-Gan, Israel
(Received 3 November 1987)

The theory of neural networks is extended to include discrete neurons with more than two discrete states. The dynamics of such systems are studied. The maximum number of storage patterns is found to be proportional to $Nq(q-1)$, where q is the number of Potts states and N is the size of the network. The properties of the Potts neural network are compared with the Ising case, and the similarity between the Potts neural network and a diluted multineuron interacting Hopfield model is discussed.

Neural networks which exhibit features of learning and associative memory can be modeled[1-8] by a system of Ising spins with an energy function

$$H=-\frac{1}{2}\sum_{i\neq j}^{N}J_{ij}S_iS_j \ . \qquad (1)$$

The two states $S_i=\pm 1$, represent the two main levels of activity of the ith neuron, and N is the total number of neurons. The bonds J_{ij} are the synaptic efficacies of a pair of neurons. They are assumed to be modified by learning in a manner which ensures the dynamic stability of certain configurations. In the Hebb learning rules[9] the accumulated effect of learning on the synaptic connection between the pair (i,j) can be represented by a matrix

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu \ . \qquad (2)$$

The p patterns $\{\xi_i^\mu\}$ constitute the embedded memories. They are assumed to be random, with equal probabilities for $\xi_i=\pm 1$. The changes ΔJ_{ij} induced by the addition of a new pattern has a fixed magnitude $\Delta J_{ij}=\xi_i\xi_j/N=\pm 1/N$.

The statistical mechanics of Hopfield's model [Eqs. (1) and (2)] have been recently studied[3-4] in the limit of $N\rightarrow\infty$. Three classes of metastable states have been found. Retrieval states, each of which has a large overlap with a signal pattern, exist when $\alpha\equiv p/N<\alpha_C\simeq 0.14$. These states are the most important ones for retrieval of memory. The overlap of each state with the corresponding pattern is $R=N^{-1}\sum_i\xi_i^\mu S_i$. Its value at maximum capacity (and zero temperature) is $R\simeq 0.97$. The slight reduction of R from unity is due to the small internal static noise which is generated by the random overlaps among the patterns. For a finite α there are also spin-glass states that have overlaps which are of $O(1/\sqrt{N})$ with the patterns. In addition, at sufficiently small α, mixture states exist, which have finite overlaps with a small number of patterns.

In this paper, we consider a generalization of the Ising neural network, with only two-state neurons, to the Potts neural network. In this model each neuron σ_i is viewed as a Potts spin[10] with q possible discrete states: Each state may represent a color or shade of grey of each pixel in the pattern. The state of the network of N such neurons is defined as the instantaneous configuration of all the spin variables at a given time. The dynamic evolution of the system, in the phase space of p^N states, is determined by the interactions among the neurons. The neurons are interconnected by a synaptic *matrix* of strength J_{ij} which determines the contribution of the jth presynaptic neuron to the potential of the postsynaptic neuron i. The potential h_{σ_i} on neuron i, which is in state σ_i, is the sum of all postsynaptic potentials delivered to it in an integrating period of time, i.e.,

$$h_{\sigma_i}=-\sum_j\sum_{k,l=1}^{q}J_{ij}^{kl}m_{\sigma_i,k}m_{\sigma_j,l} \ , \qquad (3)$$

where $m_{\sigma_i,r}$ is an operator which obeys the Potts symmetry constraints and is given by

$$m_{\sigma_i,r}=q\delta_{\sigma_i,r}-1 \ , \qquad (4)$$

q is the number of Potts states and σ_i is a q-state Potts variable, which could be in q different states $\sigma_i=1,2,\ldots,q$. From Eq. (3) it is clear that the potential of the postsynaptic neuron i depends both on the state of neuron i and its neighbors and on their synaptic efficacies.

The dynamics of the Potts neural network is very different from the Ising case. In the Ising case, the state of the neuron in the next time step and at zero temperature is equal to the *sign* of the induced local field, which is calculated by the neuron itself. In the Potts case, by using heat bath dynamics,[11] each neuron makes a more complicated decision. It first calculates the induced local field for each of the q Potts states $\{h_\sigma\}$. At zero temperature the state of the neuron in the next time step is fixed to be the state which minimized the induced local field. Therefore, the dynamics in the Potts case is more complicated and slower due to the calculation of q induced local fields and their minima.

The stable states of the system will be those configurations in which every Potts spin variable σ_i is in a Potts state which gives a minimum value to $\{h_{\sigma_i}\}$.

It will be assumed throughout this paper that the J_{ij}'s are symmetric, i.e., $J_{ij}^{kl}=J_{ji}^{lk}$. In such a case the above mentioned stability[12] condition is equivalent to the requirement that the configurations $\{\sigma_i\}$ by local minima (i.e., stable to single-spin flips) of the anisotropic Potts Hamiltonian

$$H=-\frac{1}{2}\sum_{i\neq j}^{N}\sum_{k,l=1}^{q}J_{ij}^{kl}m_{\sigma_{i,k}}m_{\sigma_j,l} \ . \qquad (5)$$

In the presence of noise there is a finite probability of having configurations other than the local minima. This can be taken into account by introducing an effective temperature $1/\beta$,[3] characterizing the level of noise in the system. The probability of neuron i to be in the next step in state σ, is given by $e^{-\beta h(\sigma)}/\mathrm{Tr}\, e^{-\beta h(\sigma_i)}$.

For the network to have a capacity for learning and memory, its stable configurations must be correlated with certain configurations, which are determined by the learning process. This is achieved by choosing the interactions to be given by

$$J_{ij}^{kl}=(q^2N)^{-1}\sum_{\mu=1}^{p} m_{k_i^\mu,k}m_{k_j^\mu,l} . \tag{6}$$

The p sets of $\{k_i^\mu\}$ are certain configurations of the network which were fixed by the learning process. The k_i^μ are taken to be quenched random variables, assuming the values $1,2,\ldots,q$, with equal probability.

There are two reasons for the choice of randomly anisotropic interactions. First, the gauge symmetry in (5) ensures the absence of a ferromagnetic order,[13,14] which means that all spins are occupied by the same Potts state. Second, in order to achieve a higher capacity, except the Potts symmetry, it seems that it is preferrable to choose more complicated interactions and not just a higher number of spin states. This result is due to the fact that the information is embedded in the synaptic strength.

The model (3)–(6) will have the capacity of storage and retrieval of information if indeed the Monte Carlo dynamics yield stable configurations $\{\sigma_i\}$ which are correlated with the learned memory $\{k_i^\mu\}$. This question is the main point of this paper.

The first case to be discussed here is the case of finite p, and we will concentrate the discussion mostly on the limit $T\rightarrow 0$. Extending the method of Ref. 3, one can show that the ensemble averaged free-energy density is given by

$$f=\tfrac{1}{2}\mathbf{R}^2-\frac{1}{\beta}\left\langle\!\!\left\langle \ln\left[\mathop{\mathrm{Tr}}_{\{\sigma\}}\exp\left(\beta\sum R_\mu m_{k^\mu,\sigma}\right)\right]\right\rangle\!\!\right\rangle , \tag{7}$$

where $\beta\equiv 1/T$. The notation $\langle\langle\ldots\rangle\rangle$ stands for the average over the distribution of $\{k_i^\mu\}$. The order parameter R is determined by the saddle-point equations $\partial f/\partial R_\mu=0$ and gives

$$R_\mu=\langle\langle\langle m_{k^\mu,\sigma}\rangle\rangle\rangle , \tag{8}$$

where $\langle\ldots\rangle$ stands for thermal average. The order parameter R_μ defines the overlap with the pattern μ. In a random configuration $R_\mu=0$ and when $\sigma_i=k_i^\mu$ $(i=1,\ldots,N)$, $R=q-1$. In the low-temperature limit the order parameter, and the ground-state energy are given by

$$R_\mu=\langle\langle m_{k^\mu,\sigma_0}\rangle\rangle , \tag{9}$$

$$E=-\mathbf{R}^2/2 , \tag{10}$$

where σ_0 is the σ which maximizes the term $(\sum m_{k^\mu,\sigma}R_\mu)$. It is easy to show from Eqs. (9) and (10) that $\mathbf{R}^2\leq(\mathbf{R}^2)^{1/2}$, which implies that $\mathbf{R}^2$ is less than $(q-1)^2$ for any states except the Mattis states. To study the symmetric solutions in which all l nonzero components are equal in magnitude near $T=0$, we use Eqs. (9) and (10) to obtain

$$R_l=\frac{1}{l}\left\langle\!\!\left\langle \max_\sigma\sum_\mu^l m_{k^\mu,\sigma}\right\rangle\!\!\right\rangle , \tag{11}$$

$$E_l=-\tfrac{1}{2}lR_l^2 . \tag{12}$$

The local stability of the saddle points of f [Eq. (7)] is determined by the eigenvalues of the matrix $A^{\mu\delta}\equiv\partial^2 f/\partial R_\mu\partial R_\delta$. The matrix $\underline{A}$ has three groups of eigenvalues $\lambda_1=\gamma$, $\lambda_2=\gamma+(l-1)\delta$, and $\lambda_3=\gamma-\delta$ where the quantities γ and δ are given by

$$\gamma=1+q^2\beta\left\langle\!\!\left\langle\left(\frac{Z_{k^\mu}}{\Sigma Z_a}\right)^2-\frac{Z_{k^\mu}}{\Sigma Z_a}\right\rangle\!\!\right\rangle , \tag{13}$$

$$\delta=-\beta q^2\left\langle\!\!\left\langle\frac{\delta_{k^\mu,k^\nu}Z_{k^\mu}}{\Sigma Z_a}-\frac{Z_{k^\mu}Z_{k^\nu}}{(\Sigma Z_a)^2}\right\rangle\!\!\right\rangle , \tag{14}$$

where $Z_a\equiv\exp(\beta\sum_l m_{k_l,a}R_l)$. One should notice that in the symmetric solutions there is a finite probability that $\delta\propto-\beta$. Therefore all the symmetric solutions, except the Mattis solution, are unstable in the limit $T\rightarrow 0$. In the same way one can show that any solution which has two or more equal overlaps is unstable at low temperature. These results indicate that, unlike the Ising case,[3] in the Potts neural network any stable solution must have only *one dominant macroscopic overlap*. Another difference between the Ising and the Potts neural network is that in the Potts system there are only p Mattis states for the absence of inversion symmetry in the system.

The second case is a finite $\alpha\equiv p/N$. The average free-energy per spin $f=-\langle\langle\ln\mathrm{Tr}\exp(-\beta H)\rangle\rangle/N\beta$ of the Hamiltonian (3)–(6) is calculated by the replica method. Our discussion will be within the replica symmetric theory[15] and will concentrate on the most important ferromagnetic solutions, which are characterized by only one macroscopic overlap with a single pattern. Extending the method of Ref. 4, one can show that the free energy of the Mattis state in the replica symmetric theory is given by

$$f=\frac{R^2}{2}+\frac{\alpha}{2}\left[\beta r(q-1)(1-Q)+\beta^{-1}\ln[1-\beta(q-1)(1-Q)]-\frac{(q-1)Q}{1-(q-1)\beta(1-Q)}\right]$$
$$-\beta^{-1}\left\langle\!\!\left\langle\ln\mathrm{Tr}\exp\left\{\beta\left[Qr\left(\frac{q-1}{q}\right)\right]^{1/2}\sum_l m_{\sigma,l}Z_l+\beta Rm_{\sigma,k_1}\right\}\right\rangle\!\!\right\rangle_{k_1,Z_l} , \tag{15}$$

where the symbol $\langle\langle\ldots\rangle\rangle$ stands for average over k_1 and over q Gaussian variables Z_l with zero mean and unity variance. The free-energy equation (15) is a function of three order parameters. The macroscopic overlap with one of the patterns $R=N^{-1}\sum m_{\sigma,k_1}$, the Edwards-Anderson order parameter[16] $Q=(Nq)^{-1}\sum_{i,l}\langle m_{\sigma_i,l}\rangle^2$, and the total mean-square random overlap with $p-1$ rest patterns $r=\alpha^{-1}\sum_\mu\langle\langle(R_\mu)^2\rangle\rangle$. From Eq. (15) one can see that the local field consists of two parts. A ferromagnetic part Rm_{σ,k_1}, resulting from the condensed overlap, and a spin-glass part $[Qr(q-1)/q]^{1/2}$ generated by the random overlap with the rest of the $p-1$ patterns.

In the zero-temperature limit the saddle-point equations for the order parameters are

$$R=-1+q\int_{-\infty}^{\infty}\frac{dz}{\sqrt{\pi}}e^{-z^2}\left(\frac{1+\mathrm{erf}(z+R\sqrt{q/2ar})}{2}\right)^{q-1}, \tag{16}$$

$$R=q(1-c)^{-2}, \tag{17}$$

$$c=\frac{2q}{\pi r a}\int_{-\infty}^{\infty}dz\,e^{-z^2}z\left[\left(\frac{1+\mathrm{erf}(z+R\sqrt{q/2ra})}{2}\right)^{q-1}+(q-1)\left(\frac{1+\mathrm{erf}(z)}{2}\right)^{q-2}\left(\frac{1+\mathrm{erf}(z-R\sqrt{q/2ra})}{2}\right)\right], \tag{18}$$

where $c\equiv\beta(q-1)(1-Q)$. The saddle point equations (16)–(18) always have the solution $R=0$. This is a spin-glass solution that has no macroscopic overalps with any of the patterns.

Numerical solution of Eqs. (18)–(20) gives the following results. The maximum capacity for $q=3$, 4, 5, and 9 is given by $a_c(q)=0.415$, 0.82, 1.37, and 4.8. The retrieval $\bar{R}=R/(q-1)$ at the maximum capacity is given, respectively, by 0.956, 0.941, 0.93, 0.92. One can notice that the maximum capacity is very close to the formula

$$a_C\simeq\frac{q(q-1)}{2}0.138, \tag{19}$$

and the retrieval at the maximum capacity is a decreasing function of q. (The maximum capacity in the Ising case and within the replica symmetry solution is $a_C=0.138$.[4]) A more exact way to estimate the error in the retrieval state is the generalization of the Shanon formula[17] to the Potts case

$$S=-\frac{1}{q}[1+(q-1)\bar{R}]\ln\left(\frac{1+(q-1)\bar{R}}{1-\bar{R}}\right)+\ln[q/(1-\bar{R})]. \tag{20}$$

Equation (20) gives for $q=2$, 3, 4, 5, and 9 at the maximum storage $S=0.084$, 0.15, 0.23, 0.28, and 0.40, respectively. This result indicates that the retrieval of information is a decreasing function of q. The solution in the large q limit is still unknown. Therefore, it is still unknown if the retrieval decreases to zero or to a finite value in the large q limit.

The higher capacity in the Potts case, compared to the Ising case, is partially due to the fact that in the Potts neural network each synapse contains more information. From Eq. (6) each synapse between a pair of neurons is a matrix containing q^2 elements. For each of the patterns, the embedded information between a pair of neurons (i,j) is fixed together by the variables k_i and k_j, which have q^2 different possibilities. This explanation shows that each synapse contains q^2 bits, which makes the results of Eq. (19) more logical.

Another aspect which increases the capacity is the fact that each pattern in the Potts case contains $N\log_2 q$ bits. The reason that $\log_2 q$ does not appear in Eq. (19) is perhaps due to the fact that the "Hebb's learning rule" in the case of the Potts neural network is $\Delta J_{ij}=(q-1)^2$, $-(q-1)$, 1 which must be represented also by $\log_2 q$ bits.

Another aspect of the Potts neural network is the mapping between this model and a corresponding highly diluted network with multineuron interactions. Each Potts neuron σ_i, which could be in $p=2^n$ states, $n=1,2,\ldots$, should be replaced by a block of n Ising neurons, $S_i^1,S_i^2,\ldots,S_i^n$. For the cases where $2^{n-1}<p<2^n$ one should replace a Potts neuron by a block of n Ising neurons, and add constraints or field in order to avoid the forbidden states.[18] Using the identity that $\delta_{s_is_j}\equiv(1+s_is_j)/2$, one can rewrite the interaction between the pair (i,j) in the following form:

$$\sum_{k,l}J_{ij}^{kl}m_{\sigma_i,k}m_{\sigma_j,l}=\sum_{\{\rho\}=\pm1}\left(\prod_{l=1}^{n}(\rho_i^l\xi_i^l+1)/2-1\right)\left(\prod_{l=1}^{n}(\rho_j^l\xi_j^l+1)/2-1\right)\times\left(\prod_{l=1}^{n}(\rho_i^lS_i^l+1)/2-1\right)\left(\prod_{l=1}^{n}(\rho_j^lS_j^l+1)/2-1\right), \tag{21}$$

where ξ_i^l is a random number which could be ±1 with equal probability, and the summation is over 2^{2n} possibilities of the ρ's. It is obvious that only terms which contain an even number of each of the ρ's remain after the summation. These terms have the form of multineuron interactions of order $2,3,\ldots,2\log_2 q$. Each term must contain at least one two-state neuron from each of the two blocks of n neurons. For the case $p=4$, for example, the remaining terms are

$$S_i^kS_j^l\xi_i^k\xi_j^l,\quad S_i^kS_i^lS_j^k\xi_i^k\xi_i^l\xi_j^k,\quad S_i^kS_i^lS_j^kS_j^l\xi_i^k\xi_i^l\xi_j^k\xi_j^l,$$

where in the second and third terms $k\neq l$. The second term, which includes three multineuron interactions, is responsible for the breaking *of the inversion symmetry.* The first term with k equal to l is exactly an interaction of the Hopfield type, but this term with $k \neq l$ represents interactions which prefer neurons to be parallel to the k pattern in the ith block and to the l pattern in the jth block. Such types of interaction may be used to represent correlations between different parts of the pattern.

In the representation of the Potts neural network as a network of highly-diluted multineuron interactions, the capacity per bond is proportional to $\log_2 q$. This huge capacity (compared to the fully connected case[19]) is due to the fact that the number of synapses in the system is proportional to N^2q^2, but on the other hand, the number of retrieval bits is proportional to $N^2q^2\log_2 q$. The huge capacity per synapse is perhaps due to the fact that most of the synapses are of the $\log_2 q$ multineuron type. The slower dynamics in the multineuron representation is due to the calculation of the induced local field, which is mostly a summation of $\log_2 q$ multineuron synapses.

The nature of the Potts neural network in the full phase space and the solution in the large q limit are still open questions. The relevance of such dynamics in biological systems is still unclear.

It is a pleasure to thank Professor D. J. Amit and Professor H. Sompolinsky for numerous stimulating discussions.

*Present address: Department of Physics, Princeton University, Princeton, New Jersey 08544.

[1]J. J. Hopfield, Proc. Nat. Acad. Sci. U.S.A. **79**, 2554 (1982); **81**, 3088 (1984).

[2]W. A. Little, Math. Biosci. **19**, 101 (1974); W. A. Little and G. L. Shaw, *ibid.* **39**, 281 (1978).

[3]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).

[4]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) **173**, 30 (1987).

[5]G. Toulouse, S. Dehaene, and J. P. Changeux, Proc. Nat. Acad. Sci. U.S.A. **83**, 1695 (1986).

[6]M. Mezard, J. P. Nadal, and G. Toulouse, J. Phys. (Paris) **47**, 1457 (1986).

[7]T. Kohonen, *Self Organization and Associative Memory* (Springer-Verlag, Berlin, 1984).

[8]P. Peretto, Biol. Cybern. **50**, 51 (1984).

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

[10]F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).

[11]N. Metropolis, A. Rosenblatt, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[12]R. J. Glauber, J. Math. Phys. **4**, 294 (1963); K. Binder, in *Fundamental Problems in Statistical Mechanics V,* edited by E. G. D. Cohen (North-Holland, Amsterdam, 1980), p. 21.

[13]D. J. Gross, I. Kanter, and H. Sompolinsky, Phys. Rev. Lett. **55**, 304 (1985).

[14]D. Elderfieid and D. Sherrington, J. Phys. C **16**, L497 (1983).

[15]S. Kirkpatrick and D. Sherrington, Phys. Rev. B **17**, 4384 (1978).

[16]S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).

[17]B. Shanon, in *The Role of Representation in Cognition,* edited by J. Bishop, J. Lockheed, and D. N. Perkins (Erlbaum, Hillsdale, 1987).

[18]I. Kanter, J. Phys. C **20**, L257 (1987).

[19]E. Gardner (unpublished).