Categorization by a three-state attractor neural network

D. R. C. Dominguez* and D. Bollé†
Instituut voor Theoretische Fysica, Katholicke Universiteit Leuven, B-3001 Leuven, Belgium
(Received 16 July 1997)

The categorization properties of an attractor network of three-state neurons, which infers three-state concepts from examples, are studied. The evolution equations governing the parallel dynamics at zero temperature for the overlap between the state of the network and the examples, the state of the network, and the concepts, as well as the neuron activity, are derived in the limit of extreme dilution. The transition from the retrieval region to the categorization region occurring when the number of examples or their correlations are increased is discussed as a function of the zero-activity threshold of the neurons. In particular, the differences with models for binary concepts are highlighted. [S1063-651X(97)05312-9]

PACS number(s): 87.10+e, 64.60.Cn

Some years ago a minimal modification of the Hopfield model was suggested such that categorization of patterns emerges naturally from an encoding stage structured in layers [1]. The network is spatially homogeneous, but the patterns are hierarchically ordered. Related models have been examined in [2–5]. These models lead to the appearance of stable states besides those corresponding to the original patterns, e.g., the ancestors of the categories to which those patterns belong.

Shortly after, a simple Hebbian rule was proposed in order to study the performance of a network in learning an extensive number of ancestor patterns in such a hierarchical ordering, given that the learning takes place with groups of a finite number of (correlated) patterns situated on a lower level of the hierarchical tree [6]. In other words, the problem of categorizing examples (i.e., the correlated patterns) into classes defined by concepts (i.e., the ancestors) is studied. It turns out that such a network loses its ability to retrieve the examples when a critical number of them is presented during the learning stage, but it then gets the ability to categorize the concepts [6–10].

Recently, models with multistate and analog neurons have been introduced in the study of categorization problems [11,12]. By using analog neurons [11], fewer (binary) examples are needed in order to start categorization. However, the generalization error, i.e., the Hamming distance between the microscopic state of the network and the (binary) concepts is larger than in the corresponding two-state model. A further improvement is obtained by using low-activity examples, from which (binary) full-activity concepts can be inferred, even if the number of examples is small [12]. This must be due to the fact that mixture states of patterns can be inherently stable, allowing the network to ultimately form higher-activity patterns out of smaller ones, just as happens in the retrieval regime for both highly diluted [13] and fully connected three-state networks [14].

In this Brief Report we extend these models by allowing the concepts themselves to be three state. As in [7], we do not require full symmetry of the retrieval overlaps of the examples.

*Electronic address: david@tfdec1.fys.kuleuven.ac.be
†Electronic address: desire.bolle@fys.kuleuven.ac.be

Consider a network of N three-state neurons. At time t and zero temperature the neurons $\{\sigma_{i,t}\}$ are updated in parallel according to the rule

$$\sigma_{i,t+1}=F_\theta(h_{i,t}), \quad h_{i,t}=\sum_{j(\neq i)} J_{ij}\sigma_{j,t}, \quad i=1,\ldots,N \tag{1}$$

where $h_{i,t}$ is the local field of neuron i at time t. The input-output relation F_θ is, in general, a monotonic function and will later on be chosen as the three-state steplike function

$$F_\theta(x)=\begin{cases} \mathrm{sgn}(x) & \text{if } |x|>\theta \\ 0 & \text{if } |x|<\theta, \end{cases} \tag{2}$$

where θ is the zero-activity threshold parameter of the neurons. The synaptic couplings J_{ij} are determined through the Hebbian learning of s three-state examples, $\eta_i^{\mu\rho}\in\{0,\pm 1\}, \mu=1,\ldots,p, \rho=1,\ldots,s$, of p three-state concepts, $\xi_i^\mu\in\{0,\pm 1\}$. The examples have zero mean and variance $A=1/N\Sigma_i(\eta_i^{\mu\rho})^2$, which is a measure for their activity. The concepts ξ_i^μ are chosen to be independent identically distributed random variables (IIDRV) with mean zero and activity equal to the activity A of the examples. At site $i=1,\ldots,N$, each set of examples is built from the concept through the process

$$\eta_i^{\mu\rho}=\xi_i^\mu\lambda_i^{\mu\rho}, \quad \lambda_i^{\mu\rho}\in\{\pm 1\}, \quad \rho=1,\ldots,s. \tag{3}$$

The variables $\lambda_i^{\mu\rho}$ are also taken to be IIDRV with a bias towards the value $+1$ such that they are given by the probability distribution

$$\mathrm{p}(\lambda_i^{\mu\rho})=b_+\delta(\lambda_i^{\mu\rho}-1)+b_-\delta(\lambda_i^{\mu\rho}+1), \tag{4}$$

with $b_\pm=(1\pm b)/2$. The parameter b describes the correlation between the example $\eta_i^{\mu\rho}$ and its concept ξ_j^μ, $\langle\eta_i^{\mu\rho}\xi_j^\mu\rangle=bA\delta_{ij}$, and the correlation between two different examples of the same concept $\langle\eta_i^{\mu\rho}\eta_j^{\mu\sigma}\rangle=b^2A\delta_{ij}$.

We remark that the binary categorization model [6,7] is recovered by setting $A=1$ and $\theta=0$; the standard three-state neuron model [13] is obtained by taking $s=1$ and $b=1$.

In order to measure the quality of retrieval (categorization) of the examples (concepts) we introduce the Hamming distance between the examples (concepts) and the microscopic state of the network [16]

$$D_t^{\mu\rho}=\frac{1}{N}\sum_i\ [\eta_i^{\mu\rho}-\sigma_{i,t}]^2=A-2Am_{N,t}^{\mu\rho}+Q_{N,t}, \qquad (5)$$

$$E_t^{\mu}=\frac{1}{N}\sum_i\ [\xi_i^{\mu}-\sigma_{i,t}]^2=A-2AM_{N,t}^{\mu}+Q_{N,t}. \qquad (6)$$

The quantity E_t^{μ} can be considered as the generalization error in this context (see, e.g., [17]). Equations (5) and (6) define the retrieval overlap between the microscopic state of the network and the ρth example of the μth concept, the overlap between the microscopic state of the network and the concept ξ_i^{μ} itself, and the neural activity

$$m_{N,t}^{\mu\rho}=\frac{1}{NA}\sum_i\ \eta_i^{\mu\rho}\sigma_{i,t}, \quad M_{N,t}^{\mu}=\frac{1}{NA}\sum_i\ \xi_i^{\mu}\sigma_{i,t}, \qquad (7)$$

$$Q_{N,t}=\frac{1}{N}\sum_i\ |\sigma_{i,t}|^2. \qquad (8)$$

In obtaining the second equalities in Eqs. (5) and (6) we have used the fact that the activities of the concepts and of the examples are taken to be equal. The $m_{N,t}^{\mu\rho}$ are normalized order parameters within the interval $[-1,1]$, which attain the maximal value $m_N^{\mu\rho}=1$ whenever $\sigma_i=\eta_i^{\mu\rho}$.

We now consider an extremely diluted asymmetric version of this model in which each neuron is connected, on average, with C other neurons,

$$J_{ij}(C)=\frac{C_{ij}}{CA}\sum_{\mu=1}^{p}\sum_{\rho=1}^{s}\ \eta_i^{\mu\rho}\eta_j^{\mu\rho}. \qquad (9)$$

Here, the $C_{ij}\in\{0,1\}$ are IIDRV with probability $\Pr\{C_{ij}=1\}=C/N, C>0$. This allows an exact solution of the parallell dynamics [15].

We take an initial network configuration correlated with only one concept meaning that only the retrieval overlaps for the s examples of that given concept, say the first one, are macroscopic, i.e., of order $O(1)$ in the thermodynamic limit $N\to\infty$. In order to study the retrieval of a particular example we single out the component $\rho=1$. We furthermore assume that all other components are the same, i.e., $m_{N,t}^{1\rho}=m_{N,t}^{1s}$ for all $\rho>1$. We call this property of the examples quasisymmetry.

The dynamics of this model is studied following standard methods involving a signal-to-noise analysis (see, e.g., [15,16]). At this point we recall that the retrieval overlaps have to be considered over the diluted structure and the loading α is defined by $p=\alpha C$. Splitting the local field in Eq. (1) into a signal and noise part gives

$$h_{i,t}=\eta_i^{11}m_{C,t}^{11}+\sum_{\rho>1}^{s}\ \eta_i^{1\rho}m_{C,t}^{1\rho}+\sum_{\mu>1,\rho}^{p,s}\ \eta_i^{\mu\rho}\sum_{j\neq i}^{N}\frac{C_{ij}}{CA}\,\eta_j^{\mu\rho}\sigma_{j,t},$$

$$m_{C,t}^{1\rho}\equiv\sum_{j(\neq i)}^{N}\frac{C_{ij}}{CA}\,\eta_j^{1\rho}\sigma_{j,t}. \qquad (10)$$

In the thermodynamic limit we then obtain in a standard way [12,15,16]

$$m_{t+1}^{11}=\langle\langle\langle\lambda^{11}F_\theta(\tilde{h}_t)\rangle_{\lambda^{11}}\rangle_{x_s}\rangle_{\omega_t}, \qquad (11)$$

$$m_{t+1}^{1s}=\langle\langle\langle x_sF_\theta(\tilde{h}_t)\rangle_{\lambda^{11}}\rangle_{x_s}\rangle_{\omega_t}, \qquad (12)$$

$$M_{t+1}^{1}=\langle\langle\langle F_\theta(\tilde{h}_t)\rangle_{\lambda^{11}}\rangle_{x_s}\rangle_{\omega_t}, \qquad (13)$$

$$Q_{t+1}=A\langle\langle\langle F_\theta^2(\tilde{h}_t)\rangle_{\lambda^{11}}\rangle_{x_s}\rangle_{\omega_t}+(1-A)\langle F_\theta^2(\omega_t)\rangle_{\omega_t}, \qquad (14)$$

with

$$\tilde{h}_t\doteq\lambda_i^{11}m_t^{11}+(s-1)x_sm_t^{1s}+\omega_t, \qquad (15)$$

$$\omega_t=[\alpha rQ_t]^{1/2}\mathcal{N}(0,1). \qquad (16)$$

Here $\doteq$ indicates that this relation is valid in distribution, $x_s=[1/(s-1)]\Sigma_{\rho>1}^{s}\lambda^{1\rho}$, $r=s[1+(s-1)b^4]$, and the quantity $\mathcal{N}(0,1)$ is a Gaussian random variable with mean zero and variance unity. The angle brackets denote the further averages over both λ^{11} and x_s, and over ω_t. The average over λ^{11} has to be done according to the distribution (4). For the average over x_s we employ

$$p_b(j)=\binom{s-1}{j}b_+^jb_-^{s-1-j}, \qquad (17)$$

with $2j=(s-1)(x_s+1)$. In the case that we have many examples per concept we use for x_s the Gaussian approximation $x_s\doteq b+z_s\sqrt{(1-b^2)/(s-1)}$ with $z_s=\mathcal{N}(0,1)$ independent of ω_t. Furthermore, we have averaged already over ξ^1.

The first term in the expression (15) is the signal coming from the first example of the first concept, while the second term represents the signal of the other examples of the first concept. It has a strength factor x_s. The third term is the noise caused by the examples of the $(p-1)$ residual noncondensed concepts.

Because of the extremely diluted structure of the network, Eqs. (11)–(14) give a complete description of the dynamics for a general monotonic input-output function F_θ. Choosing F_θ to be the three-state function (2) we now discuss this dynamics by solving numerically the fixed-point equations given by Eqs. (11)–(14).

Besides the zero solution Z determined by $m^{11}=M^1=Q=0$ we have the following different types of solution: the retrieval solutions R defined by $m^{11}>M^1>0$ and $Q>0$, the categorization solutions G defined by $0<m^{11}<M^1$ and $Q>0$ and the self-sustained activity solutions S with $Q>0$ but $m^{11}=M^1=0$. For the retrieval, respectively, categorization solution we impose the further condition $D<0.1$, respectively, $E<0.1$. The idea is to guarantee a minimal retrieval, respectively, generalization quality of the network.

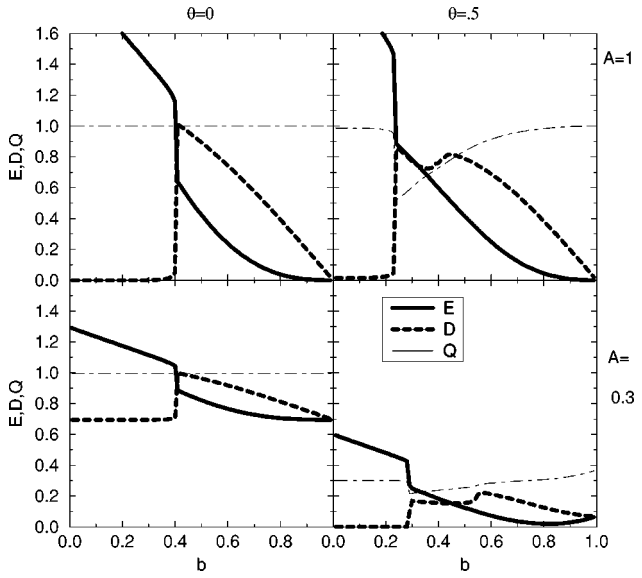

FIG. 1. The Hamming distance $D=D_{\infty}^{11}$ (dashed line), the categorization error $E=E_{\infty}^{1}$ (full line), and the neuron activity $Q=Q_{\infty}$ (thin dashed-dotted line) as a function of the correlation b. The number of examples $s=5$, the loading rate $\alpha=0.01$, the activity A takes the values 1 in the upper part and 0.3 in the lower part, and the threshold θ is 0 in the left part and 0.5 in the right part.

Since there are many parameters to be considered in the discussion of the numerical results we only show in Figs. 1-4 the properties of the network we believe to be typical and important.

Figure 1 shows that for binary patterns ($A=1$) the use of three-state neurons ($\theta=0.5$) does not affect the overall behavior of the network. However, for three-state patterns ($A=0.3$) both the retrieval and the categorization abilities are improved. The transition from an R phase to a G phase is clearly present in all data. For a critical value of the correlation, b_c, there is a crossing between the D and E lines. Here we remark that in the case of three-state patterns and binary neurons ($\theta=0$) the conditions for good retrieval and categorization behavior of the full patterns are, of course, not satisfied. But as the curves indicate, e.g., one finds the best possible retrieval of the active sites ($D=0.7$ for $b<b_c$). Furthermore, we also note the existence of a plateau for D in the case of three-state patterns and three-state neurons, where the Hamming distance is not very small but it still satisfies $D<E$. Finally, in all cases there exists a minimal value for E, meaning that the categorization is optimal for the corresponding network parameters. This does not always happen for $b=1$, the reason being that although the neuron activity Q becomes high for large b, the pattern activity A may be so small that σ_i cannot match ξ_i. This is in agreement with Eq. (6).

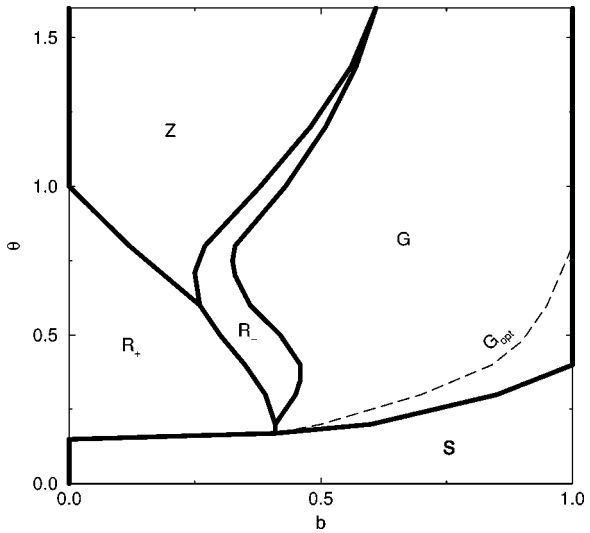

FIG. 2. The (θ,b) phase diagram with $A=0.1$, $s=5$, and $\alpha=0.01$. The thin dashed line G_{opt} indicates optimal categorization.

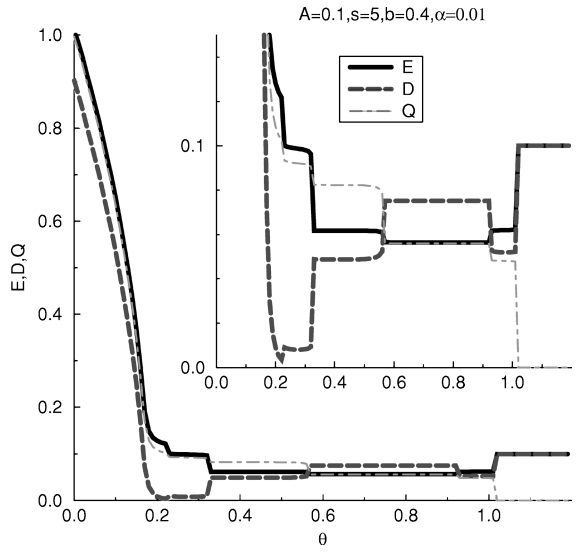

FIG. 3. The Hamming distance D (dashed line), the categorization error E (full line), and the neuron activity Q (thin dashed-dotted line) as a function of the threshold θ in the region R_- of Fig. 2.

This behavior is further illustrated in a typical (θ,b) phase diagram (Fig. 2). We have divided the R phase in two regions, one of them (R_+) referring to the region where D is

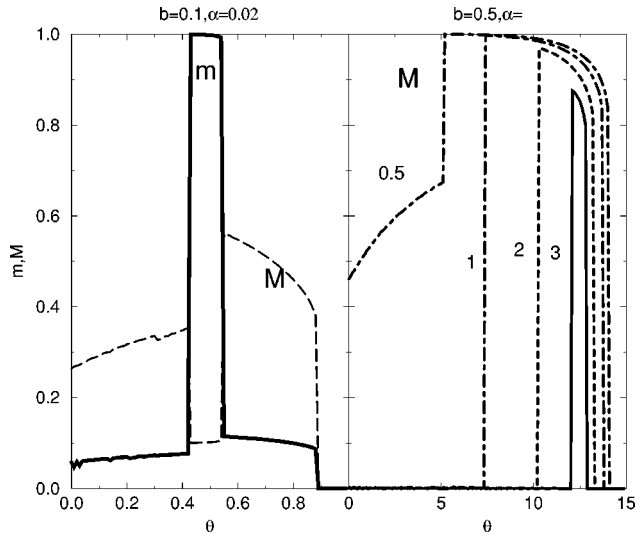

FIG. 4. Left: The overlaps m (full line) and M (dashed line) as a function of θ for $A=0.01$, $s=80$, $b=0.1$, and $\alpha=0.02$. Right: The overlap M as a function of θ for $A=0.01$, $s=80$, $b=0.5$, and $\alpha=0.5$ (dashed-dotted line), $\alpha=1$ (dashed line), $\alpha=2$ (dotted line), and $\alpha=3$ (full line).

almost zero, the other one (R_-) indicating the region where D has already jumped to the plateau seen in Fig. 1. This R_- phase has not been seen before in the literature on the models with binary concepts. The typical behavior in this phase is represented in Fig. 3. When increasing θ starting from zero, both E and D sharply drop simultaneously from a maximum value to a small value (almost zero in the case of D). Then D jumps to a higher plateau, but stays smaller than E (the R_- phase). Afterwards, D jumps further to become bigger than E (the categorization phase). Finally it attains the value $0.1 = A$. To a certain extent one could say that the network allows for both retrieval and categorization in this R_- region. However, optimal categorization occurs along the thin dashed line G_{opt} of Fig. 2.

In Fig. 4 we plot the behavior of the network for a smaller A. For a network in the R phase an appropriate choice of θ leads to a retrieval overlap $m \approx 1$ while the overlap with a concept M becomes small. Thereby we note that although the concept storage seems to be rather small ($\alpha = p/C = 0.02$), the example storage is large for correlated patterns ($\alpha_s = s\alpha = 1.6$). For a network in the G phase and increasing values of θ until $\theta = \theta_{\mathrm{opt}}(\alpha)$ the overlap with a concept becomes larger, indicating that the categorization ability improves. For $\theta > \theta_{\mathrm{opt}}(\alpha)$ this categorization ability slowly decreases and at $\theta = \theta_Z(\alpha)$, M falls abruptly to zero. This illustrates that for a carefully tuned θ, $M \approx 1$, implying that categorization stays successful even for a concept loading larger than $\alpha = 3$. So compared with the categorization properties of the binary concept model (see Fig. 2 of [12]), this indicates that by using three-state concepts the categorization error is smaller and that a greater number of concepts can be categorized.

Finally, we have also studied, for comparison, m and M as a function of θ for an analog input-output relation $F_\theta = \tanh(x/\theta)$. The same network parameters are used as in Fig. 4. Concerning retrieval of the examples a similar behavior is found with a slightly smaller m. Concerning categorization, however, although an analogous nonmonotonic behavior in θ is seen, M does not come close to 1 for larger α. It demonstrates that the gain parameter of a continuous input-output relation does not play the role the zero-activity threshold does for the three-state case. The reason is that in the three-state case this threshold switches off the neurons whose field h_i is not large enough such that the σ_i can match the three-state patterns ξ_i.

In conclusion, we have studied the retrieval and categorization properties of an extremely diluted three-state neural network through the solution of its parallel dynamics. In comparison with existing models in the literature the concepts are allowed to be three state. The important parameters governing the transition from the retrieval to the categorization phase are the number of examples per concept and their correlations. By choosing appropriately the zero-activity threshold of the neurons we find, in comparison with models for binary concepts, that there exists a region where both the Hamming distance and categorization error stay small and that a much greater number of concepts can be categorized.

This work has been supported in part by Cnpq/Brazil, the Universidad Autonoma de Madrid, and the Research Fund of the K.U. Leuven (Grant No. OT/94/9). One of us (D.B.) is indebted to the Fund for Scientific Research–Flanders (Belgium) for financial support.

[1] N. Parga and M. A. Virasoro, J. Phys. (France) **47**, 1857 (1986).
[2] M. V. Feigelman and L. B. Ioffe, Int. J. Mod. Phys. B **1**, 51 (1987).
[3] S. Bös, R. Kühn and J. L. van Hemmen, Z. Phys. B **71**, 261 (1988).
[4] A. Krogh and J. A. Hertz, J. Phys. A **21**, 2211 (1988).
[5] H. Gutfreund, Phys. Rev. A **37**, 570 (1988).
[6] J. F. Fontanari and R. Meir, Phys. Rev. A **40**, 2806 (1989).
[7] J. F. Fontanari, J. Phys. (France) **51**, 2421 (1990).
[8] E. Miranda, J. Phys. I (France) **1**, 999 (1991).
[9] P. R. Krebs and W. K. Theumann, J. Phys. A **26**, 3983 (1993).
[10] R. Crisogono, A. Tamarit, N. Lemke, J. Arenzon, and E. Curado J. Phys. A **28**, 1593 (1995).
[11] D. A. Stariolo and F. A. Tamarit, Phys. Rev. A **46**, 5249 (1992).
[12] D. R. C. Dominguez and W. K. Theumann, J. Phys. A **29**, 749 (1996).
[13] J. S. Yedidia, J. Phys. A **22**, 2265 (1989).
[14] C. Meunier, D. Hansel, and A. Verga, J. Stat. Phys. **55**, 859 (1989).
[15] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[16] B. Bollé, G. M. Shim, B. Vinck, and V. A. Zagrebnov, J. Stat. Phys. **74**, 565 (1994).
[17] T. L. H. Watkin, A. Rau, and M. Biehl, Rev. Mod. Phys. **65**, 499 (1993).