Short Communication

Categorization and generalization in the Hopfield model

Miguel C. Branchtein and Jeferson J. Arenzon

Instituto de Física, Universidade Federal do Rio Grande do Sul, C.P. 15051, 91501-970, Porto Alegre, RS, Brazil

Abstract. — The capability of the Hopfield model in generalizing concepts from examples is studied through numerical simulation. Noisy examples are taught and we test if the network grasps the underlying concepts. Several parameters are modified along the simulation: the number of concepts and examples and the initial noise in the examples. It is obtained the first order transition from the retrieval phase to the categorization phase predicted by the mean field theory.

Nowadays the study of neural networks is focused mainly in the ability of recalling stored patterns that are attractors (stable fixed points) of the network dynamics. However, analogously to real systems, other properties deserve to be studied, among them, the capability of generalizing concepts from examples. When a given set of examples is learned, we must be able to recognize the essential information embedded in each example, learning the underlying concept exemplified by them. This categorization process improves both the way memories are stored in real systems and the handling velocity of information. As stressed by Virasoro [1], this is not a choice but rather a necessity.

To define the generalization problem, let us denote each of the P concepts by an N-dimensional vector $\{\xi^\mu\}$ with components ± 1 and the s examples of each concept by $\{\xi^{\mu\nu}\}$ where $\mu = 1,\ldots,P$ and $\nu = 1,\ldots,s$. Each example is created from one concept introducing a certain amount of noise, measured by r. The noise parameter r is defined as the probability that the i-th neuron in an example is different from the corresponding neuron in the concept, i.e. $P(\xi_i^{\mu\nu} = -\xi_i^\mu) = r = (1-b)/2$, while b is the overlap between a concept and its examples. The concepts are randomly chosen uncorrelated patterns with probability 0.5 of each neuron state be ± 1. Remark that only the examples are learned, not the concepts.

Symmetric spurious states, that disturb the dynamics when other properties are studied, play an important role in the categorization process: in the generalization phase, the system extracts the common information from the examples of a given concept μ and the state $\mathbf{C}^\mu$ (or

one near it) that lies in the "center of mass" of all examples and is given by

$$C_i^\mu = \text{sgn} \left(\sum_{\nu=1}^{s} \xi_i^{\mu\nu} \right) \quad , \tag{1}$$

becomes a stable fixed point. This state overlaps on average equally with all s examples and is the effectively learned concept. In this context, the desired situation occurs when these states are absent if the patterns are not correlated but this s-symmetric central state appears when similar patterns (the examples) are stored [2].

In the Hopfield model, that may be described by the energy function

$$E = -\frac{N}{2} \sum_{\mu=1}^{P} \sum_{\nu=1}^{s} m_{\mu\nu}^2 \quad , \tag{2}$$

where $m_{\mu\nu}$ is the overlap with the examples

$$m_{\mu\nu} = \frac{1}{N} \sum_i^{N} \xi_i^{\mu\nu} S_i \quad , \tag{3}$$

the particular examples are no longer stable when the system enters in the so called generalization phase if the number of concepts is finite (if it is extensive, they are never stable). As a consequence the network cannot recognize a given realization of a concept (an example) but just the concept itself. One possible measure of the generalization error is the Hamming distance ϵ between the original concept (not taught) and the generalized state (effectively learned) for a certain number s of learned examples:

$$\epsilon = \frac{1 - m_\mu}{2} \quad , \tag{4}$$

where m_μ is the overlap with the μ-th concept

$$m_\mu = \frac{1}{N} \sum_i \xi_i^\mu S_i \tag{5}$$

The generalization in the Hopfield model for $\alpha' \neq 0$ ($\alpha' \equiv$ number of concepts/number of neurons) was previously studied through numerical simulation by Miranda [4]: above a critical number of examples s_c, the generalization error decays rapidly following a power law ($\epsilon \sim s^{-\theta}$, where θ decreases linearly with α'). Miranda also found that s_c grows exponentially with the noise level and linearly with α'. The transition to the generalization regime is continuous, while analytical calculations [5] predicted it to be discontinuous.

The $T = 0$ simulation was performed for network sizes up to 8192 neurons for $\alpha' = 0$ and 0.05. The algorithm follows the steps:

(i) a set of P concepts is generated;

(ii) for each concept, one example ($s = 1$) is generated with noise b and taught to the network;

(iii) the initial state is set in one of the original concepts (or one of the examples) and a neuron is flipped whenever this lowers the energy until a stable state is reached. The generalization error ϵ is then measured;

(iv) another example is created ($s \rightarrow s + 1$) and steps (ii) and (iii) are repeated.

The step (iii) is repeated for many initial states and for several different sets of concepts, yielding the averaged generalization error ϵ. An important difference from reference [4] is that there only the stability of the concepts were tested, while here the initial state can also be set in one of the examples since the transition found in analytical calculations (for $\alpha' = 0$) refers to the situation when the individual examples stop being stables. The multispin coding algorithm [7] was used in order to save computer memory and decrease computational time.
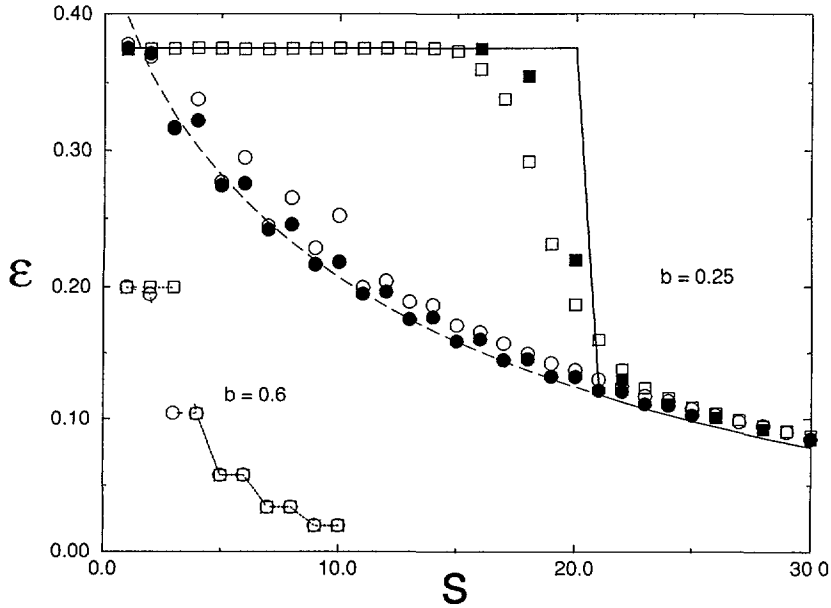

Fig.1. — Average generalization error ϵ *versus* the number of examples s for $\alpha' = 0$ ($P = 5$) and $b = 0.25$ and 0.6. The network sizes are $N = 4096$ (empty symbols) and $N = 8192$ (filled symbols). Note that for $s < s_c$ we have two values of ϵ, depending on whether the initial state is the concept (circles) or the example (boxes). For $s > s_c$, both curves merge. The analytical prediction corresponding to the case when the initial state is the example is indicated by the full line while for the concepts it is the dashed line (after [5]).

Figure 1 shows the generalization error ϵ versus the number of examples s in the limit $\alpha' \rightarrow 0$ (not studied by Miranda [4]) and two different values of b. This limit is achieved by holding fixed the number of concepts (in this case 5) and increasing the size of the network (up to 8192). The averages were taken over 5 distinct sets of concepts and, for each set, all concepts and up to a maximum of 20 examples were tested. The main new result is that the discontinuous transition predicted by the mean field theory [5] shows up. The step-like form of the simulation curve ϵ, clearly seen for small s, demonstrates that an odd number of examp les is required to correctly extract the concept, while doubts are created when the number is even: the network cannot decide between two situations. When the number of examples is even, some values C_i^μ, equation (1), are zero, the correspondent state of the i-th neuron in the concept is not defined and the generalization error does not decrease when an even example is learned by the network.

These steps do not appear in the generalization error curve in previous calcula tions [5] because the substitution of binomial (discrete) distributions by a Gaussian (continuous) one masks this effect smoothing the curves. This approximation can also contribute for the small discrepancy between the predicted critical number of examples [5] and the one found in the simulations (it may also be due to replica symmetric unstable solutions). The actual shape of the curves, obtained with discrete distributions [6], shows these fine details. Note that if the initial state is set in one of the examples, the simulated curves fit very well the analytical prediction [5] (solid line) for large networks. On the other hand, when one of the concepts is chosen as the initial state of the dynamics, we find that for $s < s_c$ the net stabilizes in a different state, yielding another value for the error ϵ: both examples and concepts are stable and have distinct basins of attraction. As predicted in reference [5], for $s > s_c$, the generalization error obbeys the equation

$$\epsilon = \frac{1}{2}\mathrm{erfc}\left[\left(\frac{sb^2}{2(1-b^2)}\right)^{1/2}\right] \tag{6}$$

Here we found that this solution also holds for $s < s_c$, as can be seen in fig.1 (dashed line) if the initial state of the dynamics is one of the concepts. Moreover, the conclusions of Miranda [4] do not hold in the $\alpha' = 0$ limit: from equation(6), the generalization error, for great s, no longer follows a power law but has an exponential decay. The model was also studied for $\alpha' = 0.05$
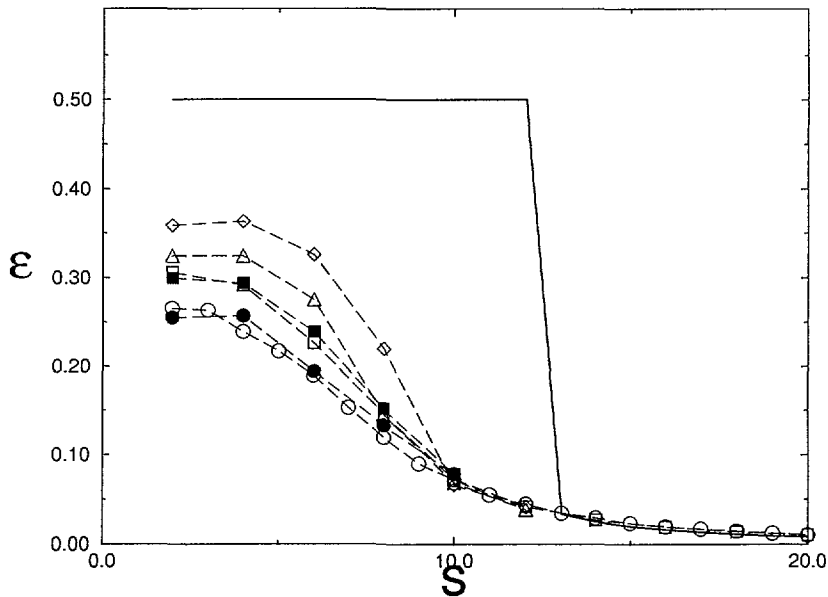

Fig.2. — Average generalization error versus the number s of examples for $\alpha' = 0.05$, $b = 0.6$ and $N = 1024$ (○), 2048 (□), 4096 (△) and 8192 (◇). For $s < s_c$ the curves, for both concepts (empty symbols) and examples (filled symbols) as initial states, tend to be a flat plateau in $\epsilon = 0.5$ as N increases. The full line is the mean field prediction (after [5]).

(Fig. 2). Differently from the previous case, here the plateau must no longer be at the value of r but at 0.5 [5]: below a certain critical number of examples, the network will always evolve

towards a spin glass state that has no macroscopic overlap with the concepts. In this limit the dependence on the size of the network is too strong: the 0.5 plateau is not achieved for small networks but, as N increases, the curves slowly approach their asymptotic behavior. Also, there is no dependence on the initial state, the curves for examples and concepts are almost the same. Again the curves match the analytical predictions [5] and, comparing with them (solid line), it may also be argued that the continuous transition found by Miranda was just an effect of the finite size of his simulations, although his other conclusions seems to remain valid (at least for the sizes used here).

Virasoro [1] pointed out the analogy between the lack of ability in recognizing examples belonging to a certain class (although not the class itself) in a synaptic dilution context and a mental disorder known as prosopagnosia: when synapses are cut, the system first loose the capability of discerning the examples and further the concepts (agnosia). The same effect appears in the Hopfield model when too many examples are taught (for a finite number of concepts): the basins of attraction of the examples are ruled out when their number is greater than a critical value. If one has an extensive number of concepts, the examples are never stable and the system is always in a prosopagnosia state. This is not relevant when the task proposed to the system is to classify a given pattern in a certain class or category and it is conceptually correct to denominate the phase appearing for $s > s_c$ in the Hopfield model as a categorization phase. However, when the purpose is to simulate systems that must preserve the identity of the particular examples, as well as extract the common information between them, the denomination of "generalization phase" is somewhat imprecise. Also, when the system stores a given hierarchy of patterns, including classes inside classes several times, it is no longer possible to control with temperature the desired level on the hierarchy [8]. This would only be possible if the system preserved the individual basins of attraction. Thus, the ideal situation consists of a big basin of attraction centered at the concept with smaller basins related to each example lying inside the big basin, around its center: general ideas are easier to remember than particular examples of such ideas. This situation is achieved by other models, for instance, the RS model [3].

Summarizing, we presented results for the Hopfield model and Hebb learning rule storing sets of correlated patterns with each set consisting of some particular examples of a given concept. The results for $\alpha' = 0$ and $\alpha' \neq 0$ and several degrees of correlation match very well the analytical predictions [5]. Nevertheless, it would be interesting to study the size of the basins of attraction of examples and concepts as their number increases [3] as well the mean convergence time, in order to verify whether there is or not a change of behavior for some value of s. In particular, to test if the twofold solution found for $\alpha' = 0$ and $s < s_c$ has similar basins of attraction or not. It would also be interesting to investigate systems for which a new symmetric phase (or more than one, depending on the number of levels of hierarchy) appears when sets of correlated patterns are embedded in the network [2, 3]. The Hopfield model presents such a phase (for all values of b) when examples are learned in pairs [9] and, in this case, the desired retrieval level can be controlled with temperature. As a final remark, a suitable and more general definition of the critical number of examples s_c to start generalization should not make reference to the stability of the examples or the mere existence of the s-symmetric state as, for instance, the one given by Miranda [4] that considers a change in the generalization error decaying velocity.

References

[1] Virasoro M., *Phys. Rep.* **184** (1989) 300.

[2] Arenzon J.J., de Almeida R.M.C. and Iglesias J.R., *J. Stat. Phys.* **69** (1992) 385
Arenzon J.J., de Almeida R.M.C., Iglesias J.R., Penna T.J.P. and de Oliveira P.M.C., *J. Phys. I France* **1** (1992) 55.

[3] Branchtein M.C., J.J. Arenzon, R.M.C. de Almeida and J.R. Iglesias, in preparation.

[4] Miranda E.N., *J. Phys. I France* **1** (1991) 999.

[5] Fontanari J.F., *J. Phys. I France* **51** (1990) 2421.

[6] Krebs P.R., private communication

[7] Penna T.J.P. and de Oliveira P.M.C., *J. Phys* **A22** (1989) L719.
de Oliveira P.M.C., Computing Boolean Statistical Models, (World Scientific Publishing, 1991).

[8] We thank P.R. Krebs for an interesting discussion on this point.

[9] Tamarit F.A. and Curado E.M.F. *J. Stat. Phys.* **62** (1991) 473.