Neural Networks 66 (2015) 22–35

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Hierarchical neural networks perform both serial and parallel processing

Elena Agliari [a], Adriano Barra [a], Andrea Galluzzi [b], Francesco Guerra [a,c], Daniele Tantari [b], Flavia Tavani [d,*]

[a] Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy
[b] Dipartimento di Matematica, Sapienza Università di Roma, P.le A. Moro 2, 00185, Roma, Italy
[c] Istituto Nazionale di Fisica Nucleare, Sezione di Roma, Roma, Italy
[d] Dipartimento SBAI (Ingegneria), Sapienza Università di Roma, Via A. Scarpa 14, 00185, Roma, Italy

ARTICLE INFO

Article history:
Received 5 September 2014
Received in revised form 18 February 2015
Accepted 22 February 2015
Available online 2 March 2015

Keywords:
Multitasking associative networks
Serial processing
Parallel processing

ABSTRACT

In this work we study a Hebbian neural network, where neurons are arranged according to a hierarchical architecture such that their couplings scale with their reciprocal distance. As a full statistical mechanics solution is not yet available, after a streamlined introduction to the state of the art via that route, the problem is consistently approached through signal-to-noise technique and extensive numerical simulations. Focusing on the low-storage regime, where the amount of stored patterns grows at most logarithmical with the system size, we prove that these non-mean-field Hopfield-like networks display a richer phase diagram than their classical counterparts. In particular, these networks are able to perform serial processing (i.e. retrieve one pattern at a time through a complete rearrangement of the whole ensemble of neurons) as well as parallel processing (i.e. retrieve several patterns simultaneously, delegating the management of different patterns to diverse communities that build network). The tune between the two regimes is given by the rate of the coupling decay and by the level of noise affecting the system.

The price to pay for those remarkable capabilities lies in a network's capacity smaller than the mean field counterpart, thus yielding a new budget principle: the wider the multitasking capabilities, the lower the network load and vice versa. This may have important implications in our understanding of biological complexity.

1. Introduction

Statistical mechanics constitutes a powerful technique for the understanding of neural networks (Amit, 1992; Coolen, Kuhn, & Sollich, 2005; Sollich, Tantari, Annibale, & Barra, 2014), however overcoming the mean-field approximation is extremely hard (even beyond neural networks). Basically, the mean-field approximation lies in assuming that each spin/neuron S_i in a network dialogs with *all* the other spin/neurons with the same strength.[1] For instance, if we consider a ferromagnetic model, once introduced N spins $S_i = \pm 1,\ i \in (1, \dots, N)$, we have the two extreme scenarios of a nearest-neighbor model like the Ising lattice, whose Hamiltonian can be written as

$$H_{\text{Ising}} = -\sum_{\langle i,j\rangle} J S_i S_j, \tag{1}$$

where, crucially, the sum runs over all the couples $\langle i, j\rangle$ of *adjacent* sites, and the mean-field Curie–Weiss model, whose Hamiltonian can be written as

$$H_{\text{Curie–Weiss}} = -\sum_{i<j}^{N,N} J S_i S_j, \tag{2}$$

where the sum runs over *all* the $N(N-1)/2$ spin couples irrespective of any notion of distance; this is equivalent to think of spins interacting through nearest neighbor prescriptions but as they were embedded in an N-dimensional space. Clearly, solving the

* Corresponding author.
E-mail address: flavia.tavani@sbai.uniroma1.it (F. Tavani).

[1] Notice that this situation corresponds to a system embedded in a fully-connected (i.e. complete graph) topology. However, situations where we introduce some degree of dilution (e.g. Erdös–Rényi graph), yet preserving the homogeneity of the structure and an extensive coordination number, can be looked and treated as mean field models.

http://dx.doi.org/10.1016/j.neunet.2015.02.010
0893-6080/

statistical mechanics of the latter model is much simpler with respect to the former. The main route toward finite-dimensional descriptions has been paved by physicists in the study of condensed matter.[2] Indeed, incredible efforts have been spent from the 1970s in working out the renormalization-group (Wilson, 1971a), namely a technique which allows inferring the properties of three-dimensional ferromagnets starting from mean-field descriptions, but a straight solution of the Ising model in dimensions 3 is still out of the current mathematical reach.[3]

Actually, in the last decade some steps forward toward *more realistic* systems have been achieved merging statistical mechanics (Ellis, 1985; Gallavotti & Miracle-Sole', 1967; Mezard, Parisi, & Virasoro, 1987) and graph theory (Albert & Barabasi, 2002; Bollobas, 1998; Watts & Strogatz, 1998). In particular, mathematical methodologies were developed to deal with spin systems embedded in random graphs, where the ideal, full homogeneity among spins is lost (Agliari, Annibale, Barra, Coolen, & Tantari, 2013a, 2013b). Thus, networks of neurons arranged according to Erdös–Rényi (Barra & Agliari, 2008), small-world (Agliari & Barra, 2011), or scale-free (Perez-Castillo et al., 2004) topologies were addressed, yet finite-dimensional networks were still out of debate.

Focusing on neural networks, it should be noted that, beyond the difficulty of treating non-trivial topologies for neuron architecture, one has also to cope with the complexity of their coupling pattern, meant to encode the Hebbian learning rule. The emerging statistical mechanics is much trickier than that for ferromagnets; indeed neural networks can behave either as ferromagnets or as spin-glasses, according to the parameter settings: their phase space is split into several disconnected pure states, each coding for a particular stored pattern, so to interpret the thermalization of the system within a particular energy valley as the spontaneous retrieval of the stored pattern associated to that valley. However in the high-storage limit, where the amount of patterns scales linearly with the number of neurons, neural networks approach pure spin-glasses (losing retrieval capabilities at the blackout catastrophe Amit, 1992) and, as a simple Central Limit argument shows (Barra, Genovese, Guerra, & Tantari, 2012), when the amount of patterns diverge faster that the amount of neurons they become purely spin glasses. For the sake of exhaustiveness we also stress that, even in the retrieval region, neural networks are *exactly* linear combinations of two-party spin glasses (Barra, Contucci, Mingione, & Tantari, 2015; Barra, Genovese, & Guerra, 2010, 2012; Barra, Genovese, Guerra, Tantari et al., 2012; Barra, Genovese, Guerra, & Tantari, 2014): due to the combination of such difficulties, neural networks on a finite dimensional topology have not been extensively investigated so far.

However, very recently, a non-mean-field model, where a topological distance among spins can be defined and couplings can be accordingly rescaled, turned out to be, to some extent, treatable also for complex systems such as spin-glasses (Castellana & Parisi, 2011; Monthus & Garel, 2014). More precisely, spins are arranged according to a hierarchical architecture as shown in Fig. 1: each pair of nearest-neighbor spins form a "dimer" connected with the strongest coupling, then spins belonging to nearest "dimers" interact each other with a weaker coupling and so on recursively (Mukamel, 2008). In particular, the Sherrington–Kirkpatrick model for spin-glasses defined on the hierarchical topology has been investigated in Castellana, Decelle, Franz, Mezard, and Parisi (2010): despite a full analytic formulation of its solution still lacks, renormalization techniques, (Castellana & Parisi, 2011; Monthus & Garel, 2013), rigorous bounds on its free-energies (Castellana, Barra, & Guerra, 2014) and extensive numerics (Metz, Leuzzi, & Parisi, 2014; Metz, Leuzzi, Parisi, & Sacksteder, 2013) can be achieved nowadays and they give extremely sharp hints on the behavior of systems at large size defined on these peculiar topologies.

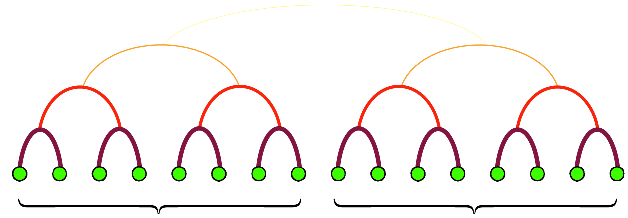

Fig. 1. Schematic representation of the hierarchical topology, that underlies the system under study: green spots represent nodes where spins/neurons live, while different colors and thickness for the links mimic different intensities in their mutual interactions: the brighter and thinner the link, the smaller the related coupling. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Remarkably, as we are going to show, when implementing the Hebb prescription for learning on these hierarchical networks, an impressive phase diagram, much richer than the mean-field counterpart, emerges. More precisely, neurons turn out to be able to orchestrate both serial processing (namely sharp and extensive retrieval of a pattern of information), as well as parallel processing (namely retrieval of different patterns simultaneously).

The remaining of the paper is structured as follows: in the next subsections we provide a streamlined description of mean-field serial and parallel processors, and we introduce the hierarchical scenario. Then, we split in three sections our findings according to the methods exploited for investigation: statistical mechanics, signal-to-noise technique and extensive numerical simulations. All these approaches consistently converge to the scenario outlined above. Seeking for clarity and completeness, each technique is first applied to a ferromagnetic hierarchical mode (which can be thought of as a trivial one-pattern neural network and acts as a test-case) and then for a low-storage hierarchical Hopfield model.

1.1. Mean-field processing: serial and parallel processors.

Probably the most famous model for neural networks is the Hopfield model presented in his seminal paper dated 1982 (Hopfield & Tank, 1987), counting nowadays more than twenty-thousand citations (Scholar). This is a mean-field model, where neurons are schematically represented as dichotomic Ising spins (state +1 represents firing while state −1 stands for quiescence) interacting via a (symmetric rearrangement of) the Hebbian rule for learning as masterfully shown by the extensive statistical–mechanical analysis that Amit, Gutfreund and Sompolinsky performed on the model (Amit, 1992; Amit, Gutfreund, & Sompolinsky, 1985).

More formally, once introduced N neurons/spins S_i, $i \in (1, \dots, N)$, and p quenched patterns ξ_μ, with $\mu \in (1, \dots, p)$, whose entries are drawn once for all from the uniform distribution

$$P(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1), \tag{3}$$

the Hopfield model is then captured by the following Hamiltonian $H_{\mathrm{Hopfield}}(S|\xi)$:

$$H_{\mathrm{Hopfield}}(S|\xi) = -\frac{1}{N}\sum_{i<j}^{N}\left(\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu\right)S_iS_j. \tag{4}$$

[2] In that context the long-range interactions are unacceptable because the involved couplings are of electromagnetic nature, hence displaying power-law decay with the distance.

[3] It is worth mentioning that the Wilson–Kadanoff renormalization equations (Wilson, 1971b, 1972, 1974) turn out to be exact in models with power law interactions as those built on the hierarchical lattice that we are going to consider.