Archeologia e Calcolatori
Supplemento 6, 2014, 113-130

ARTIFICIAL NEURAL NETWORKS AND COMPLEXITY: AN OVERVIEW

1. Introduction: complex systems and connectionism

Understanding the world around us is usually a difficult task. All dynamically evolving phenomena in the natural world are produced by a strong interaction among a great number of causes of which only few are visible or measurable. Moreover, the phenomena, like the weather evolution, may be so distributed over the space or time that only a small number of measurements can be done, making the understanding of the overall system difficult and approximated. In general, some characteristics of systems can produce a very strange behaviour, even when the elements constituting the system are a small number. All these elements and their mutual interaction can produce the so-called *complexity*.

In order to understand the approach a researcher may use in analysing a system, a very simple metaphor may be adopted: the iceberg. An iceberg is a floating ice mountain in the sea that shows only a small visible part above the waterline. If some specific tools are not used to improve our investigation of the iceberg, all we can describe is the movement of its visible part, the rate of melting, the colour, the transparency, and the like. Any other aspect that belongs to the submerged part is excluded from a direct measurement. Therefore, any hypothesis about the global behaviour of the iceberg can be proved by using only what we are allowed to see directly. Is all this incomplete amount of accessible information enough to fully describe the iceberg and its future evolution? This is a very difficult question to answer. All we could say is that the visible behaviour, in some sense, contains also the occulted information and everything that is out of our sight can be extracted from what is known. Even when no theories or hypotheses are allowable to create a reference framework, complex systems have the characteristic to show an evolution through the mixed actions or interactions of the variables.

In observing natural, social, economical, physical, biological systems, we basically deal with measured data that give us a partial knowledge of the "visible part" of the system. Therefore, data is required to re-build a mathematical or algorithmic framework that could be sufficiently detailed and powerful to describe the fundamental aspects of the system under study, its evolution over time, and its meaningful characteristics. A system is an organised, purposeful structure that consists of interrelated and interdependent elements (components, entities, factors, members, parts, etc.). These elements continually influence one another (directly or indirectly) to maintain their activity and the existence of the system in order to achieve the goal of the system. Although all systems have outputs, which are considered as observable variables that make possible the measurement of what the system is doing at a given time, they may also:

1) have inputs and feedback mechanisms;
2) maintain an internal steady-state (called *homeostasis*) despite a changing external environment;
3) display properties that are different than the whole (called *emergent properties*) but are not possessed by any of the individual elements;
4) have boundaries that are usually defined by the system observer.

Systems underlie every phenomenon and all are part of a larger system. Together, they allow understanding and interpretation of the universe as a meta-system of interlinked wholes and organise our thoughts about the world. If a system has no input variables, it is called *autonomous*; otherwise, if input variables can modify the outcomes of the system, it is called *non-autonomous*. Although different types of systems (from a single cell to the human body, from soap bubbles to galaxies, from ant colonies to nations) look very different on the surface, they have remarkable similarities. At the most basic level, systems are divided into two categories:

1) *Closed systems*: theoretical systems that do not interact with the environment and are not influenced by their surroundings. Only the components within the system are significant. Example: a sealed jar, nothing enters or exits the jar, but whatever is inside can interact.
2) *Open systems*: real-world systems the boundaries of which allow exchanges of energy, material, and information with the larger external environment or system in which they exist. Example: a company where, even if there are separate departments in one organisation, the workers share data and interact with each other on a daily basis.

Some other differences among systems can be found in terms of determinism. Before addressing this aspect in the world of systems, it is necessary to define the system *state*. In a system, the state describes the minimum set of inner variables that are able to uniquely describe any part of the system. When a system returns to a specific state or situation, which it already visited in the past, no differences can be found between the two situations. Therefore, two identical systems with the same state cannot be distinguished. Of course, not all the systems have inner states. If a system has no inner states, it is called a *0*-order system, and the outputs depend only on the input values. Otherwise, the presence of inner states in some way gives the system a sort of memory of the past: what happens now depends on the inputs and also on what the system did previously. These kinds of systems are called *N*-order systems, where *N* is, in some sense, the amount of memory the system beholds. Since a system's evolution over time depends on the inputs and on the past, the future outcomes of the system should also be determined by these two elements. In deterministic systems, the past and the future evolution over time are determined uniquely for a specific input. This means that if the inner state of a system is known and the input sequence in time is given, every future evolution of the system will be known and defined. From the mathematical point of view, a differential equation form represents a continuous time-deterministic system:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}; \boldsymbol{y}(t))$$

where $\boldsymbol{x}$ is a vector containing all state variables of the system and $\boldsymbol{y}$ is a vector describing inputs explicitly depending on time (t). If $\boldsymbol{y}$ is zero, the system is autonomous. The same relation in discrete time is:

$$\boldsymbol{x}_{n+1} = \boldsymbol{G}(\boldsymbol{x}_n; \boldsymbol{y}_n)$$

$\boldsymbol{F}$ (or $\boldsymbol{G}$ in discrete time domain) is the operator linking the rate of variation of the system variables to the present state, and it can be either linear or non-linear.

In general, the evolution in time of linear differential equations is completely determined and can be calculated by means of well-established mathematical techniques. Conversely, non-linear differential equations do not have a general solution mechanism and in most cases do not admit analytical solutions. Anyway, several mathematical and geometrical techniques were developed to define the long-term evolution of this kind of equations and to outline the global behaviour of the differential dynamical system. Autonomous differential systems have steady states if there exists some combination of $\boldsymbol{x}$ variables where $\boldsymbol{F}(\boldsymbol{x})=0$. In these points, also called fixed points, the variation of $\boldsymbol{x}$ is null and the system will keep this steady state until some perturbation is applied from the external environment (*input*). The stability of the fixed points is described by the dynamical behaviour of the surrounding space. The local space can be studied by a linearization of the dynamic system, and the general behaviour of the system around the fixed point can be evaluated by means of the main directions of convergence or divergence (*eigenvectors*) and their associated eigenvalues (Robinson 2004).

Linear differential equations, as dynamic outcomes, can produce only fixed stable or unstable points and oscillations instead of more complex geometric objects (both in two or higher dimensions). Conversely, non-linear differential systems can show a greater amount of time evolutions, some of which are definitely more strange and difficult to deal with. In three or more dimensions, all previous cases can appear but additional behaviours may be added to the geometric taxonomy of *attractors* (Guckenheimer, Holmes 1983; Khalil 2001; Jordan, Smith 2007). An attractor is a set of points in the phase space where all trajectories starting in a sufficiently close state will converge. The set of all points fulfilling this request is called basin of attraction. Therefore, the attractive fixed points and orbits shown in two-dimensional examples are attractors. As mentioned in the previous part, since the dynamic evolution is considered deterministic, two different trajectories cannot intersect each other to preserve the uniqueness of the future system evolution. Starting with this consideration, one may ask what kind of new attractors may emerge from a high-dimensional non-linear system. Around 1970, physicists and computer scientists encountered a special kind of attractors that, even if they were describing a deterministic system, they could not forecast the long-term evolution (or limit behaviour) unless considering a new geometrical object called fractal. This kind of time evolution of a system was named *chaos*. Some examples of chaotic attractors are the Duffing oscillator, the Lorenz system, or the Chua's circuit.

A chaotic attractor shows a geometrical form similar to a ball of thread. Trajectories pass very close to each other but they never intersect, preserving the deterministic nature of the system. It can be proved that trajectories, belonging to the chaotic attractor, do not fill the space in which they are embedded in a uniform way. In previous cases, an attractive fixed point has a dimension equal to zero, an orbit has a dimension equal to one (length), surfaces are two-dimensional, volumes three-dimensional, and so on. Chaotic attractors have a non-integer dimensionality, since they do not fill the space uniformly and densely. For instance, the Lorenz attractor has a geometrical Hausdorff dimension equal to 2.06 (Falconer 1985). It means that the trajectory fills the space more than a 2-dimensional surface, but the density of points is not sufficient to fill the space as a dense volume. This is the reason why these attractors are called strange or fractal.

Another feature characterising the strange attractors is the local divergence of close trajectories. Because of the geometrical aspects of this kind of strange objects, two close initial states are expected to move away from each other with an exponential law of divergence. The rate of local divergence is measured by the so-called Lyapunov exponent (Barreira, Pesin 2007). Therefore, even if the chaotic attractor geometrically describes the global behaviour of the system and the trajectory remains in that part of the space, when the system explores a state, which is close to another one visited in past, its evolution is expected to be very different after some time. The effect of diverging trajectories is called, by using a metaphor, the *Butterfly Effect*. This effect explains the dependence of the system evolution on small indetermination of the initial state. As a matter of fact, the calculation of a dynamic system time course requires infinite precision in the knowledge of the initial state. If either a small perturbation or simply a rounding operation were applied to the initial state, the future evolution of the trajectory would be expected to diverge from the predicted one.

Therefore, the Butterfly Effect describes the fundamental importance of small perturbations in the knowledge of the initial states. The name of the effect, coined by Edward Lorenz, is derived from the theoretical example of a hurricane's formation being contingent on whether or not a distant butterfly had flapped its wings several weeks before (Lorenz 1996). Finally, another feature characterising the chaotic attractors is that a chaotic evolution is neither periodic nor quasi-periodic (i.e., sum of several periodic evolutions the frequencies of which have irrational ratio). Therefore, chaotic evolutions are hardly distinguishable from random evolutions, and the time series coming from chaotic systems may be misinterpreted as unpredictable noise. The power spectrum of chaotic signals reveals continuous dense zones, similarly to noisy and weakly self-correlated systems. According to the existing literature, non-linear dynamic systems are deterministic but manifest their time evolution in a way that is very difficult to describe, analyse, and predict. Long-term prediction is to be fully excluded, even if the deterministic machine gives the possibility to extract some useful and interesting parameters to identify the systems (Ruelle 1989).

Complexity can therefore be summarised by mixing the following factors: high number of dimensions (or descriptive variables), non-linearity in description of differential equation systems, some noise, which may come naturally from environment, from exclusion of any marginal aspect of the system description, or from measurement errors. Complex systems are therefore characterised by strange, non-periodic, unpredictable time evolution, strong inter-relation among variables, sensitivity to initial condition, and difficult discrimination by noisy non-deterministic phenomena. One may ask the reason why it is so interesting to define, identify, analyse and understand complex systems.

The answer lies in the fact that most natural systems are ruled by non-linear differential equations. When these systems are non-autonomous and admit inputs from external stimuli, a very complex evolution may be difficult to define: the amount of chaos may change over time and the understanding of these phenomena becomes difficult. The traditional tools as statistics or classical mathematical approaches can fail to give sufficient information about the nature of what was observed. It has been proved that weather prediction (Lorenz 1963), socio-politic systems (Campbell, Mayer-Kress 1991; Pere *et al.* 2006), economic markets (Guégan 2009), stocks (Levy 1994), currency markets (Chorafas 1994), biological and ecological natural systems (Stone, Ezrati 1996), among others, are ruled by chaotic equations that, even with a small set of variables, can show complex and unpredictable evolution.

Complex systems represent a new approach, which studies how relationships between parts give rise to the collective behaviours of a system and how the system interacts and forms relationships with its environment. The equations from which complex system models are developed generally derived from statistical physics, information theory, and non-linear dynamics, and represent organised but unpredictable behaviours of systems of nature that are considered fundamentally complex. The physical manifestations of such systems cannot be defined; thus, the usual choice is to refer to "the system" as the mathematical information model without referring to the undefined physical subject that the model represents. The key problems of complex systems are difficulties with their formal modelling and simulation. From such a perspective, in different research contexts, complex systems are defined based on their different attributes. Since all complex systems have many interconnected components, the science of networks and network theory are important aspects of the study of complex systems. A consensus regarding a single universal definition of complex system does not yet exist.

For systems that are less usefully represented with equations, various kinds of narratives and methods are used to identify, explore, design and interact with complex systems. Some definitions of complexity focus on the question of the probability of encountering a given condition of a system once characteristics of the system are specified. The complexity of a particular system is the degree of difficulty in predicting the properties of the system, given the properties of the system's parts (Weaver 1948). In Weaver's view, complexity comes in two forms: disorganised complexity and organised complexity. Disorganised complexity results from the particular system having a very large number of parts, say millions of parts, or many more.

Although the interactions of the parts in a disorganised complexity situation can be seen as largely random, the properties of the system as a whole can be understood by using probability and statistical methods. Organised complexity, on the other hand, resides in nothing else than the non-random, or correlated, interaction between the parts. These correlated relationships create a differentiated structure that can, as a system, interact with other systems. The coordinated system manifests properties not carried or dictated by individual parts. The organised aspect of this form of complexity can be said to "emerge" without any "guiding hand". The number of parts does not have to be very large for a particular system to have emergent properties. The properties of a system of organised complexity may be understood through modelling and simulation conducted particularly with computers.

A very important aspect of complexity can be found in the field of *connectionism*. Connectionism comprises a set of approaches in artificial cognition modelling that models mental or behavioural phenomena as emergent processes of interconnected networks of simple units. The key word linking complexity and connectionism is "emergence" because the strange and complex phenomena that may arise from non-linear world are, in some sense, unexpected from the point of view of classical system analysis. For instance, the complex behaviour emerged in Lorenz model of weather was so unexpected that the author himself was convinced that it was an error in the implementation of the algorithm. The non-linear relationships between weather single elements and between the neural cells in the brain have in common the possibility of the emergence of unexpected and extremely interesting behaviour. The interesting part of complexity in brain structures is well known, as it involves the emergence of efficient approaches to solve difficult tasks that traditional algorithmic techniques fail to describe even the simplest cases. In the last years, several problems have been addressed using techniques inspired by natural connectionism: face recognition (Le 2011), language recognition (Cole 1989), automatic robot guidance (Gowdy *et al.* 1991), pattern recognition (Ripley 1996), economic prediction (White 1988), and many others.

Another aspect of connectionism related to complexity is the network of interconnected simple units. Any interconnected structure of dialoguing elements that influence the future is related to the behaviour of some set of neighbour elements of the same kind and is likely to show complex behaviour in its time evolution. Once again, such a complex behaviour is given either by the eventual non-linear relationships among elements, by their inner non-linear dynamics, or by the great amount of elements synchronically evolving in time.

In brain, for example, the complex dynamics can be measured in several cognitive states but, at the same time, some sort of cooperative coherence can be relevant depending on the task that the specific cortex area is performing. Different kind of coherence and different kind of chaotic evolution can relate to different kind of cognitive states and perceptions. According to what previously described, complexity is an attribute of connectionist systems. Therefore, simple non-linear processing units connected to each other according to some defined rule can be considered as the fundamental elements for building a complex system the-behaviour of which may reflect the complexity of a target system under investigation (Sporns *et al.* 2000).

2. Neurons and synaptic connections

The simple units that comprise a neural network are called artificial neurons, whose behaviour is based on the biological neurons by means of the functions performed by the latter operating in their natural environment. What we know about biological neurons is due, among the others, to the pioneering work of Ramón y Cajál (1911) who introduced the idea of neurons as structural constituents of the brain. Typically, neurons are rather slower than silicon logic gates, but the brain compensates the relatively slow rate of operation of a neuron by having a truly staggering number of neurons with massive interconnections between them. It is estimated that there are approximately 10 billion neurons in the human cortex and 60 trillion synapses or connections.

The result is that the brain is an enormously efficient structure. Synapses are elementary structural and functional units that mediate the interactions between neurons. The most common kind of synapse is the chemical synapse. When a presynaptic process liberates a transmitter substance (*neurotransmitter*), it diffuses across the synaptic junction between neurons and then acts on a postsynaptic process. Therefore, a synapse converts a presynaptic electrical signal into a chemical signal and then back into a postsynaptic electrical signal. In terms of physics language, a synapse operates as a one-directional gate in which information or signals may flow in only one direction. A synapse can have excitatory or inhibitory function on the receptive neuron but not both.

The modification of synaptic configuration is called *plasticity* in neurobiology. Plasticity permits the developing nervous system to adapt to its surrounding environment. In an adult brain, plasticity can operate by means of two mechanisms: the creation of new synaptic connections between neurons and the modification of existing synapses. The former part will be implemented in the phase of building the structure of an ANN while the second part will be used in the training phase of a neural system. Bioelectrical signals reach the synaptic zones, flowing into a special transmission line called axon. Axon is the unique output of a neuron, and the signal flowing into it is supported without leakage by the axonal transmitting system until it reaches the synaptic terminals. As mentioned before, a given amount of neurotransmitters is released and by diffusion, the neurotransmitter molecules reach the receptive sites of the postsynaptic neurons in specific neural structures called *dendrites*. The basic mechanisms underlying the functioning of a neuron can be summarised as follows:

1) The external stimuli reach the neuron inputs by means of the synaptic transmission. The efficiency and the nature of every synaptic site determine the amplitude of the signal read by the neuron cell.
2) All the inputs are integrated to define the internal membrane potential.
3) If the membrane potential is greater than a reference threshold potential, an action potential is generated as a sequence of spikes that is transmitted along the axon (output channel).
4) The action potential reaches the terminations where the phenomenon of neurotransmitters diffusion is repeated and the synaptic sites of the postsynaptic neurons can again read the neuronal stimulus at their inputs.

Here, we identify three basic elements of the neuronal model:

1) A set of synapses, or connecting links, each of which is characterised by a weight or connection strength. Specifically, a signal x_j at the input of synapse j connected to neuron k is multiplied by the weight w_{kj}. The first subscript refers to the neuron in question and the second subscript refers to the input end of the synapse to which the weight refers. Unlike a synapse in the brain,

the synapse weight of an artificial neuron may lie in a range that includes negative as well as positive values.
2) An adder (S) for summing (or integrating) the input signals weighted by the respective synapses of the neuron. The operation described here constitutes a linear combiner.
3) An activation function (j) for limiting the amplitude of the output of a neuron. The activation function is also referred to as a squashing function, since it squashes (limits) the permissible amplitude range of the output signal to some finite value. Typically, the normalised amplitude range of the output of a neuron is written as the closed interval [0,1] or alternatively [-1,1].

As stated in Amit (1992), some unexpected perturbations may influence the output of a neuron. Basically, several sources of incoherent mechanisms may be identified in the field of biological processes of neurons. These perturbations may be due to small fluctuations of neurotransmitter densities in synaptic vesicles, by the quantised aspect of neurotransmitter molecules, and by unpredictable fluctuations of biological elements, as for instance hormones, in the area where the neuron is functioning. The total influence of these unpredictable causes of noise follows a Gaussian statistical distribution. Since the amount of activity of a neuron is given by the frequency of spiking pulses, we can say that the number of spikes in the unit of time is proportional to the probability of activation. A spike is transmitted if the activation potential is greater than the threshold; therefore, the activity of a neuron can be formulated in terms of probability depending on the local field. The mathematical relation linking the activation probability and the local field defines a characteristic function, widely used in ANNs, usually called sigmoid or logistic function.

3. ANNs: structure and training

In 1952, Frank Rosenblatt, a psychologist and researcher at the Cornell University, invented an algorithm to perform a simple learning by an artificial neural network (Rosenblatt 1958). Since Rosenblatt attempted to model a sensory system of the brain, this typology of neural network was called *Perceptron*. The basic idea was that human beings learn to enter information and concepts by using common senses (mainly sight and hearing) and store the information in some kind of memory, such that when specific information is recalled, it has to be equal to the original one. If the recalled information were learnt incorrectly, it would be necessary to learn such information again so that the new recall operation would have a higher probability to be correct compared to before. This approach can be repeated until all input information is correctly stored and classified, if possible. A neural network that is able to process such information should have a suitable number of inputs for reading the proposed information and an appropriate number of outputs for describing the class to which it belongs.

A crucial aspect of connectionist models is their ability to learn by experience. Even in the case of Perceptron, Rosenblatt (1958) proposed an algorithm named Delta Rule to define the suitable set of synaptic weights and biases for correctly classifying a set of input-output relations. If the response of an output unit is incorrect, the network can change to produce the correct response the next time that the stimulus is presented. The activity of a neuron is determined by the sum of inputs leading to it and each input is given by the product of the activity of a presynaptic unit multiplied by the weight of the connection between them. This means that any change in connection weights will change the activity level of units in the next layer. Thus, an output unit with activity that is too low can be corrected by increasing the weights of connections from units in the previous layer that provide a positive input to it and by decreasing the weights of connections that provide a negative input. Output units with an activity that is too high can be corrected by the opposite procedure.

The fundamental aspect of the Delta Rule is that, in the case of binary units, it cannot be applied to multilayer networks. In a multilayer network, desired output of hidden units is unknown information since we want to train the network based on the final output values, which are the values of the last layer outputs. Therefore, Delta Rule can be applied only to a single layer Perceptron for linearly separable tasks. Novikoff (1962) proved that the perceptron Delta Rule algorithm converges after a finite number of iterations if the dataset is linearly separable. A more general approach can be pursued by considering the relationship between the overall errors of the network related to the patterns presented at the input. When all P patterns are presented to the network, the overall error can be calculated as:

$$E = \sum_{p=1}^{P} E^{(p)} = \frac{1}{2NP} \sum_{p=1}^{P} \sum_{i=1}^{N} \left(t_i^{(p)} - y_i^{(p)}\right)^2$$

where $t_i^{(p)}$ and $y_i^{(p)}$ are, respectively, the desired and the actual network outcomes for the p-th presented input pattern, and N is the number of outputs. Note that the error E is always positive unless the network outcomes are identically equal to the desired output, and in this case, the error is zero. Since $y_i^{(p)}$ depends on the network synaptic weights, as well as on the presented pattern at input, the general error E may be modified by changing the synaptic inner parameters of the network. In particular, if a given modification of a synaptic weight Δw eliminates the output error when a certain pattern is presented, such modification may be considered as a useful contribution to the training. Thus, the error will be lower the next time the pattern is presented. In order to achieve a better comprehension of the pattern by the network, a positive increment of the weight should be associated with a negative error variation, that is, a decrease in the error. In mathematical terms:

$$\Delta w = -\eta \frac{\partial E}{\partial w}$$

where η is a constant training parameter. This approach is called the Least-Mean-Squares procedure introduced by Widrow and Hoff (1960). According to the presented information, the feed-forward can be trained with Widrow-Hoff procedure, providing that sigmoid activation function model describes the units. Anyway, in a neural network, the error reduction can be achieved only by taking into account the whole set of synaptic weights, unlike the Widrow-Hoff procedure that can treat only the single neural layer case. This issue is avoided by the Back-Propagation training algorithm, which is based on the Widrow-Hoff approach and allows for estimating the expected values of the hidden neurons by reconfiguring the Widrow-Hoff algorithmic technique. Two main steps basically characterise Back-Propagation training algorithm:

1) The computation of the function signal appearing at the output of a neuron, which is expressed as a continuous sigmoid function of the input signal and synaptic weights associated with that neuron.
2) The computation of an estimate of the gradient vector, which is needed for the backward pass through the network.

The objective of the training process is to adjust the free parameters of the network to minimise the average error. To do this minimisation, the Back-Propagation algorithm, introduced by Rumelhart, Hinton and Williams (Rumelhart *et al.* 1986), can be used to exploit the backward error signals to overcome the limit of the Widrow-Hoff approach in case of multilayer networks. Specifically, we consider a simple method of training in which the weights are updated on a pattern-by-pattern basis until one epoch, that is, until one complete presentation of the entire training set has been dealt with. The weights are adjusted in accordance with the respective errors computed for each pattern presented to the network. The Back-Propagation algorithm is used to determine the value of the local gradients of a neuron of a specific hidden layer according to the local gradients of the next layer on the right. The local gradient of every hidden layer can be calculated *backwards* starting with the knowledge of the local gradient of the hidden layer at its right side (obviously the last hidden layer local gradient is calculated on the basis of the output layer). We now summarise the relations derived for the Back-Propagation algorithm. First, the correction Δw_{ij} applied to the synaptic weight connecting neuron *i* to neuron *j* is defined by the delta rule:

(Weight correction Δw_{ij}) = (Learning rate parameter η) × (local gradient δ_j) × (input signal of neuron j y_i).

Second, the local gradient δ_j depends on whether neuron *j* is an output node or a hidden node:

1) If neuron *j* is an output node, δ_j equals the product of the derivative $F'(v_j)$ of the sigmoid function defined in its state, and the error signal e_j, both of which are associated with neuron *j*.
2) If neuron *j* is a hidden node, δ_j equals the product of the associated derivative $F'(v_j)$ and the weighted sum of the δs computed for the neurons in the next hidden layer or output layer that are connected to neuron *j*.

When applying Back-Propagation algorithm, two distinct passes of computation are distinguished. The first pass is referred to as the forward pass, and the second is referred to as the backward pass. In these recent years, the gradient descent approach, proposed by Widrow-Hoff procedure and applied via error Back-Propagation to a multilayer neural network, has undergone numerous improvements and variations aimed at a more reliable searching for the minimum. One of the most efficient modifications of Back-Propagation algorithm is the Levenberg-Marquardt approach (Marquardt 1963), also called Levenberg-Marquardt Back-Propagation (LMBP), which was applied to multilayer neural networks by Hagan and Menhaj (1994).

Recurrent Neural Network is a class of neural networks where connections between units form one or more cycles. This creates an internal state of the network, which allows it to exhibit a dynamic temporal behaviour. Two of the most popular networks belonging to this class are the Elman network and the Jordan network (Cruse 2009). The Elman network is very similar to a multilayer feed-forward neural network, except for the presence of some context units forming a supplementary layer. At their inputs, the context units receive the outcomes of specific hidden layer units and send their outcomes to the inputs of the same hidden units again.

These connections form a cycle of information that flows into the network while the network outputs change in time and the input information remains constant. The network is then characterised by an internal dynamics and the output depends not only on the information presented at the input gates, but also on the internal network state. The outcomes of the Elman network are considered the real network responses when the relaxation process converges to an internal steady state. The Jordan network is similar to Elman networks. The context units are however fed from the output layer instead of the hidden layer. The recurrent neural networks take advantage of the presence of a memory of the recent past (internal recursion) to address complex problems in which the temporal aspect is crucial for achieving excellent results, such as handwriting recognition or spoken recognition.

The Hopfield Network was the first attempt to realise an associative memory by means of a connectionist model. Hopfield proposed a network where each binary unit is connected to all the remaining units in the network (fully connected network), although no auto-connection are allowed (Hopfield 1982). Then, for *N* neurons, the amount of required connections is *N*(*N*-1). The connection weights are defined according to the Hebb's law (Hebb 1949). The Hebb's law is a rule according to which the connection weight between a pre- and a post-synaptic neuron tends to grow when the activity of the neurons is coherent (that is, they are operating in the same state) while it decreases when there is incoherency between the units state. Starting with this aspect, Hopfield stored in this network some input patterns by determining the weights change according to the coherence/incoherence between all units that had to represent the specific "pixel" (or bit) of all the considered patterns.

After defining the network according to its connection weights, when the neurons are initialised to an initial configuration and successively allowed to evolve freely, the network tends to converge its internal dynamics to one of the memorised patterns after a relaxation time. This network was largely used for pattern correction and associative storing of information. Several networks were proposed, starting with the Hopfield original idea to overcome some particular drawbacks of the Hopfield network, such as the poor storing capacity, the patterns of orthogonality constrain, and the presence of undesired memorised phantom patterns.

Self-Organizing Map (SOM) is a neural network that is trained using unsupervised learning to produce a low dimensional (typically two-dimensional), discretised representation of the input space of the training samples called a map. As opposed to the training algorithms used for feed-forward networks, the unsupervised learning does not require an external observing system to define the distance between the correct and the actual output pattern to modify the connection weights; instead, it is based on implicit internal rules usually aimed at performing a suitable representation of the significant features of information. Kohonen (1982) proposed the most popular and widely diffused Self-Organizing Map (SOM) connectionist model. This network is based on the Winner-Takes-All (WTA) mechanism and the local training of the units spatially closed to the winner neuron. These approaches can be very powerful in defining a strategy for finding the closeness between patterns and representing it in a low dimensional space. This characteristic is similar to the result of application of statistical Principal Component Analysis (PCA) but unlike it, the neural process is typically non-linear and the description power of a SOM can be considered as more reliable and robust with respect to the traditional approaches.

4. Representation of the world by ANNs

As stated in the introduction, Artificial Neural Networks have to behold the ability to give a representation of the relations between input and output stimuli in order to assign a sufficiently general model of the system whose data are elaborated. By following the metaphor of the iceberg, data are only the observable, usually dynamic, reduced manifestation of the system, but are supposed to behold the full information about the dynamic temporal and spatial structure of the system. By heuristic observations, simple problems can be solved by neural networks whose structure is, in some sense, as simple as the problem complexity level. Conversely, problems with a high level of complex behaviour can be addressed only if the network reflects such a complexity in its connectivity. This principle is one of the bases of the neural Darwinism stated by Edelman (1987). The main thrust of his theory of neural Darwinism is that the brain is a somatic selection system similar to evolution, and not an instructional system. Here, somatic means that selection is over the time scale of a living body instead of being on the time scale of evolution.

The main difference between an instructional system and a selectional system is that the instructional system uses information from the environment to change the properties of the object in question, but a selectional system has a large and varied population of objects, and the ones that are most fit for the environment are differentially reproduced. According to Edelman, the natural evolution is the most efficient search algorithm in its proper domain, but similar selectional dynamics are present in many other systems, like immune system and natural neural networks. What selectional systems give us that instructional systems do not, is the fact that they do not require any prior knowledge of the environment, and no explicit information transfer from the world.

Whereas with the instructional system, the question of who or what *decides* what is important for the system to learn is in general unresolved. This leads to the endless regression of homunculi, implicitly contained in training algorithms (supervised training algorithms as Back-Propagation where a supervisor decides the modification of the network weights as a result of the comparison between desired and actual output, and unsupervised training algorithms as in SOMs where the algorithm structure itself is externally imposed and based on the similarities between input data and the network prototypes).

The strength of Edelman theory relies on the independence of an external controller deciding *a priori* the structure of the network which defines the ability of the neural system to address the specific task, basing on what reported previously. The complexity of the network, in terms of connectivity, is then ruled by an evolutionary process, which tend to select the best neural structure to optimise the behaviour in the virtual environment defining the problem task. From this point of view, the evolutionary approach fulfils the request that a complex task must be addressed by a sufficiently complex algorithmic process, in this case an artificial neural network, whose structure is generally unknown and depends only on the intuition and experience of the researcher. Similar considerations have been made also in the field of cognitive psychology (Olivetti-Belardinelli 1986). The human mind builds the representation of the world by consequence of the evolutionary process, which led to the actual connectivity of the brain, and the mechanisms governing the learning process of neural networks as the Hebbian synaptic weight modification rule and the brain plasticity.

The concept of a brain, or more simply, a neural network whose connectionist structure is governed by evolution over populations of neural network generations is at the base of several modern efforts in neural network research, attempting to overcome the classical approaches where the network structure is defined, by leaving the evolutionary algorithm evolving in terms of connectivity and, in some cases, weights distribution in order to catch and correctly represent the complexity of the system under analysis (Montana, Davis 1989; Zhang, Mühlenbein 1993; Fiszelew *et al.* 2007). At present, some interesting projects are involved in building a massive neural computer (Migliore *et al.* 2006), containing a hundred billions of artificial dynamic neurons and allowing a connectivity at any range order, whose connectionist structure is either inspired and led by natural brain areas observation or governed by evolutionary algorithms to detect efficient and reliable network structures for the task proposed to the artificial system (Izhikevitch, Edelman 2008).

5. Conclusions

Research on Artificial Neural Networks is still in progress. The evolution of this field, whose foundations were put by McCulloch and Pitts in 1943 (McCulloch, Pitts 1943) and, successively, were carried out by Rosenblatt (1958), Minsky (Minsky, Papert 1969), Rumelhart, Hinton and Williams (Rumelhart *et al.* 1986), was characterised in the last 20 years by explorations towards new ways of conceiving neural networks in terms of connectionist structures, training algorithms, neuron modelling and modern technologies for a physical implementation and usage. Nevertheless, several theoretical aspects have not been still sufficiently analysed and the future of neural networks field will certainly be characterised by the following aspects:

1) ANNs are basically an attempt to model the cortical neural circuits involved in natural cognitive processing. Up to now, a general theory about the development of cognitive functions in Artificial Neural Networks has not yet been formulated, even if several theoretical approaches to artificially emulate some fundamental brain cortex areas (as sensorimotor cortex, primary visual and auditory cortex) were proposed by means of classical artificial structures (Farkaš, Miikkulainen 1999; Boes *et al.* 2012; Adams *et al.* 2013). In this

framework, the involvement of Darwinist approaches, by means of Genetic Algorithms whose genetic information is related to the network structure, allows once more to select suitable neural networks aimed to cognitive tasks.
2) One of the most significant criticisms levelled to classical approaches to ANNs is the so-called Plasticity-Stability Dilemma. All the common training algorithms used to set the synaptic weights in an ANN work in a preliminary (off-line) stage whose finality is only to find the best configuration for the successive usage of the network. During the normal functioning of the ANN, adding some new training data is not recommended since a further training phase which does not take into account the old already trained information could lead to a total substitution of the stored information and the deletion of the original information. Grossberg (1987) tried to address this issue by proposing a particular neural processing called Adaptive-Resonance Theory (ART), inspired to some neurobiological evidence, allowing the coexistence of the two phases and avoiding the destruction of the information already learned. However, ART cannot be extended to the other well-defined ANNs as Perceptron or SOM, and some other heuristic approach is required to address such a specific problem.
3) Some more philosophical issues involved the world of ANNs, with particular reference to one of the most discussed topics of the Artificial Intelligence and the connection between brain and mind. Some authors (Penrose 1989) proposed to address the issue with a reductionist approach, theorising a more complex functioning of individual neurons whose neurobiological mechanisms are affected even by a quantum level. Since the molecular action plays a fundamental role in the neuronal dynamics, the neuron modelling should take into account the quantic nature of such mechanisms in order to build a solid general theory leading to uncover the still mysterious relations between biological matter and conscience.

Alessandro Londei
CeNCA – Centro di Neuroscienze Cognitive
LAA&AAS
Sapienza Università di Roma

REFERENCES

Adams S.V., Wennekers T., Denham S., Culverhouse P.F. 2013, *Adaptive training of cortical feature maps for a robot sensorimotor controller*, «Neural Networks», 44, 6-21.

Amit D. 1992, *Modeling Brain Function: The World of Attractor Neural Networks*, Cambridge, Cambridge University Press.

Barreira L., Yakov P. 2007, *Nonuniform Hyperbolicity. Dynamics of Systems with Nonzero Lyapunov Exponents*, Cambridge, Cambridge University Press.

Boes M., Oldoni D., De Coensel B., Botteldooren D. 2012, *Attention-driven Auditory Stream Segregation using a SOM Coupled with an Excitatory-inhibitory ANN*, in *Proceedings of the International Joint Conference on Neural Networks (IJCNN) (Brisbane, Australia 2012)* (http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=6241467/).

Campbell D., Mayer-Kress G. 1991, *Chaos and Politics: Applications of Nonlinear Dynamics to Socio-Political Issues*, Santa Fe Institute, Paper #: 91-09-032.

Chorafas D.N. 1994, *Chaos Theory in the Financial Markets*, McGraw Hill Professional.

Cole R.A. 1989, *Language Identification with Neural Networks: A Feasibility Study,* Institute of Electrical and Electronics Engineers Communications, Computers and Signal Processing.

Cruse H. 2009, *Neural Networks as Cybernetic Systems*, *Brains, Minds & Media*, 2nd and revised ed., Department of Biological Cybernetics and Theoretical Biology, Stuttgart, Thieme.

Edelman G. 1987, *Neural Darwinism. The Theory of Neuronal Group Selection*, New York, Basic Books.

Falconer K.J. 1985, *The Geometry of Fractal Sets*, Cambridge, Cambridge University Press.

Farkaš I., Miikkulainen R. 1999, *Modeling the Self-Organization of Directional Selectivity in the Primary Visual Cortex*, in *Proceedings of ICANN '99 (Edinburgh, Scotland 1999),* Berlin-New York, Springer, 251-256.

Fiszelew A., Britos P., Ochoa A., Merlino H., Fernández E., García-Martínez R. 2007, *Finding Optimal Neural Network Architecture Using Genetic Algorithms*, «Research in Computing Science», 27, 15-24.

Gowdy J., Pomerleau D.A., Thorpe C.E. 1991, *Combining Artificial Neural Networks and Symbolic Processing for Autonomous Robot Guidance*, «Engineering Applications of Artificial Intelligence», 4/4, 279-291.

Grossberg S. 1987, *Competitive Learning: From Interactive Activation to Adaptive Resonance*, «Cognitive Science», 11, 23-63.

Guckenheimer J., Holmes P. 1983, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, New York-Berlin, Springer.

Guégan D. 2009, *Chaos in Economics and Finance*, «Annual Reviews in Control», 33/1, 89-93.

Hagan M.T., Menhaj M.B. 1994, *Training Feedforward Networks with the Marquardt Algorithm*, in *Institute of Electrical and Electronics Engineers Transactions on Neural Networks*, Vol. 5, Issue 6, 989-993.

Hebb D.O. 1949, *Organization of Behavior*, New York, Wiley.

Izhikevitch E.M., Edelman G.M. 2008, *Large-Scale Model of Mammalian Thalamocortical Systems*, «Proceedings of the National Academy of Sciences of the United States of America», 105, 3593-3598.

Jordan D. W., Smith P. 2007, *Nonlinear Ordinary Differential Equations*, Oxford, Oxford University Press, 4th ed.

Khalil H.K. 2001, *Nonlinear Systems*, Prentice Hall.

Kohonen T. 1982, *Self-Organized Formation of Topologically Correct Feature Maps*, «Biological Cybernetics», 43/1, 59-69.

Le T.H. 2011, *Applying Artificial Neural Networks for Face Recognition*, «Advances in Artificial Neural Systems», Vol. 2011.

Levy D. 1994, *Chaos Theory and Strategy: Theory, Application, and Managerial Implications*, «Strategic Management Journal», 15, 167-178.

Lorenz E. 1963, *Deterministic Non-periodic Flow*, «Journal of the Atmospheric Sciences», 20/2, 130-141.

Marquardt D. 1963, *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, «SIAM Journal on Applied Mathematics», 11/2, 431-441.

McCulloch W.S., Pitts W. 1943, *A Logical Calculus of the Ideas Immanent in Nervous Activity*, «Bulletin of Mathematical Biophysics», 5, 115-133.

Migliore M., Cannia C., Lytton W.W., Markram H., Hines M.L. 2006, *Parralel Network Simulation with NEURON*, «Journal of Computational Neuroscience», 21, 119-129.

Minsky M., Papert S.A. 1969, *Perceptrons*, Cambridge Ma., The MIT Press.

Montana D.J., Davis L. 1989, *Training feedforward neural networks using genetic algorithms*, in *Proceedings at the 11th International Joint Conference on Artificial Intelligence (IJCAI '89) (Detroit 1989)*, Vol. 1, San Mateo California, Morgan Kaufmann, 762-767.

Novikoff A.B. 1962, *On Convergence Proofs on Perceptrons*, in *Proceedings of the Symposium on the Mathematical Theory of Automata*, 12, New York, Polytechnic Institute of Brooklyn, 615-622.

Olivetti-Belardinelli M. 1986, *La costruzione della realtà*, Torino, Bollati Boringhieri.

Penrose R. 1989, *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*, Oxford, Oxford University Press.

Pere Plaza i Font J., Régis D. 2006, *Chaos Theory and its Application in Political Science*, in *IPSA – AISP Congress (Fukuoka 2006)*.

Ramón y Cajál 1911, *Histologie du système nerveux de l'homme et des vertébrés*, Paris, Maloine.

Ripley B.D. 1996, *Pattern Recognition and Neural Networks*, Cambridge, Cambridge University Press.

Robinson J.C. 2004, *An Introduction to Ordinary Differential Equations*, Cambridge, Cambridge University Press.

Rosenblatt F. 1958, *The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain*, «Psychological Review», 65/6, 386-408.

Ruelle D. 1989, *Chaotic Evolution and Strange Attractors*, Cambridge, Cambridge University Press.

Rumelhart D.E., Hinton G.E., Williams R.J. 1986, *Learning representations by back-propagating errors*, «Nature» 323/6088, 533-536.

Sporns O., Tononi G., Edelman G.M. 2000, *Connectivity and Complexity: The Relationship Between Neuroanatomy and Brain Dynamics*, «Neural Networks», 13, 909-922.

Stone L., Ezrati S. 1996, *Chaos, Cycles and Spatiotemporal Dynamics in Plant Ecology*, «Journal of Ecology», 84/2, 279-291.

Weaver W. 1948, *Science and Complexity*, «American Scientist», 36/4, 536-544.

White H. 1988, *Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns*, in *Institute of Electrical and Electronics Engineers International Conference on Neural Networks*, 451-458.

Widrow B., Hoff M.E. 1960, *Adaptive Switching Circuits*, in 1960 IRE WESCON Convention Record, part 4, 96-104.

Zhang B.-T., Mühlenbein H. 1993, *Evolving Optimal Neural Networks Using Genetic Algorithms with Occam's Razor*, «Complex Systems», 7/3, 199-220.

ABSTRACT

Understanding the world around us is usually a hard task. All dynamically evolving phenomena in the natural world are produced by a strong interaction among a great number of causes and, often, only a few amounts of them are visible or measurable. Moreover, the phenomena may be so widely distributed over space and time, like the weather evolution, that only a small number of measurements can be taken, making the understanding of the overall system difficult and approximated. Some characteristics of systems can produce a very strange behaviour, even when the elements constituting the system are a small number. All these elements and their mutual interaction can produce the so-called complexity. Artificial Neural Networks (ANNs) form an interesting class of dynamic systems, as a paradigm of natural and spontaneous computation. ANNs are founded on bases inspired by the neurophysiological nature of neurons and their mutual connectivity. In this paper the historical reasons that led to the former mathematical models of neuron and connectionist topologies will be detailed. Over time, they have evolved through the feed-forward systems, Self-Organizing Maps, the associative memories up to the latest models in artificial cerebral cortex.