Neural networks and the brain : associative learning and/or self-organisation ?

Dedeurwaerdere Tom, Université Catholique de Louvain and National Foundation for Scientific Research, Chaussée de Wavre 434, 1370 Lathuy

Abstract

Experimental evidence suggests that modification of synaptic strength in the brain does not depend on co-activation of two connected neurons, as is assumed in most theoretical work since the proposals of Hebb (Hebb, 1949). Instead, through independent post- and presynaptic rules multiple modifications occur simultaneously at various sites in the nervous system.
To account for this data, various researchers (Edelman, Fuster, ...) propose an extension of the self-organising PDP approach to populational thinking. However, as in the PDP approach, the selection rules they propose only account for dynamical evolution of the system towards point attractors. The learning strategy of the networks is therefore still a purely bottom-up strategy.
Experiments on visual perception seem to indicate that even low level visual processes can converge to more than one attractor (ambiguous figures, binocular rivalry), to limit cycles (oscillatory behaviour) or low-dimensional chaotic attractors. I argue to extend the neural network models of perceptual categorization to dynamical attractors and to include the multiplicity of forms created by the autonomous, nonlinear brain dynamics as a complementary source of variation on which constraints of higher cognitive processes can act.

Keywords : self-organisation, dynamical attractors, neural networks

1. Introduction

In this paper we would like to analyse the possibility to use the principles of neural network computing to construct a theoretical model of the brain. For this we will compare some of the basic hypotheses of neural networks both to actual brain research and experimental evidence from cognitive psychology. We will see that neural networks - as far as we want to use them as a model of the brain - are still too much emprisoned in a purely empirical philosophy of mind and that some of its basic hypotheses needs revision.

2. Associative learning : from individuals to populations

2.1. Neural networks, some general principles

A neural network is given by a set of nodes, the formal neurons, and a set of connections between the nodes (Blayo, Verleysen, 1996). The formal neurons are artificial units of the kind depicted in figure 1.

International Journal of Computing Anticipatory Systems, Volume 1, 1998
Ed. by D. M. Dubois, Publ. by CHAOS, Liège, Belgium. ISSN 1373-5411 ISBN 2-9600179-1-9

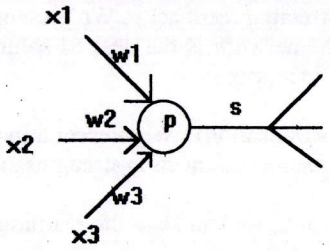

fig. 1. Neuron unit : x_i the inputs from the other neurons, w_i the synaptical weights or transmission efficacies, p the activity level of the neuron, s the output.

The activity level p of the neuron unit is the sum of the inputs x_i multiplicated by the respective synaptical weights w_i (here also called connection strenghts or transmission efficacies) :

$$p(t) = \sum w_i x_i(t) = W.x(t)$$

If the level of activation p exceeds a certain threshold θ the neuron "fires". For sake of modelisation one can suppose the output of the neuron unit to be a binary function, answering -1 (no activity) or 1 (the neuron fires) to a configuration of inputs. One can suppose the transition from one state to another to be a smooth one (ex. sigmoid output function), a linear one or abrupt.

In figure 2 we give an example of an output graph of a neuron unit with a linear transition function. We see that the neuron operates as a linear classificator. The space of possible inputs is classified in two sets, the one correponding to the firing of the neuron (y = 1), the other to the rest state (y = -1).

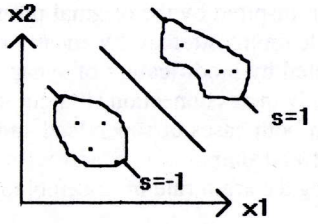

fig. 2. Output graph of a linear neuron unit

So the neuron "recognises" a certain input and classifies it in one of the two classes. Applying a set of data to two neuron units will give us four output possibilities (1,1), (1,-1), (-1,1), (-1,-1) ; a network of N units will give 2^N possibilities.

The neuron behaves as a small information processing unit. The synaptical weights determine how the inputs will be processed into the output. However normally we are confronted with the reverse problem : we have a certain idea of the processing function and we would like to know which synaptical weights can realise this function. To solve this problem the following two strategies for the modification of the synaptic weights have been proposed :

1. **Supervised learning** (error correcting feedback) : We present a list of examples $y = f(x)$ to the network, if the output s of the network is the desired solution y, or sufficiently close to this solution, connection strength increases.

2. **Unsupervised learning** (self-organisation) : We present a list of data x to the network and if they are correlated in a certain manner connection strengths are increased.

As an example of supervised learning we can state the learning rule for the Adaptive Linear Element or Adaline (B. Widrow, 1960) :

unitary transition function : $s(t) = \sum W.x(t)$

gradient descent in weight/error space : $W_i(t+1) = W_i(t) + \alpha(t)(y(t) - s(t))x(t)$ (delta rule)

Other examples are Rosenblath's Perceptron and the Multi-layered Perceptron with backpropagation of error (cfr. PDP group of Rumelhart & co.)

An example of unsupervised learning is given by the learning rule for the network of Kohonen and Von der Marlsburg (Kohonen, 1982) :

maximisation of non-correlation between input data (= statistical analyses in main components) :

1. we search the neuron i minimizing d(x(t), Wi)
2. for i and its closest neighbours : $W_i(t+1) = W_i(t) + \alpha(t)(x(t) - W_i(t))$

Other examples are the learning rule of the Hopfield's network and the Boltzmann-machine of Hinton and Sejnowski or the Hérault-Jutten model for seperating mixed up signal from independent sources

These learning strategies are both inspired by the original propositions of D. Hebb (D. Hebb, 1949). He proposed a general learning strategy by means of association of neurons in a network. The association is created by modification of synaptic weights : if two neurons are frequently activated simultaneously their connection strength increases, if not it decreases.
We find back this general rule in both cases of supervised and unsupervised learning. In first case we associate a list of input and output data x and y. In the second input we associate input data to each other, following a certain rule of ressemblance.

2.2. Comparison with actual brain research

Since the 1970's much experimental work has been done to look for support for Hebb's rule. As it turns out the experimental evidence suggests that co-activation of two connected neurons is neither a necessary, neither a sufficient condition for modification of synaptic strength. Instead, through independent post- and presynaptic rules multiple modifications occur at different sites in the nervous system (for a review of different mechanisms see for ex. Fuster, 1995, ch. 3).

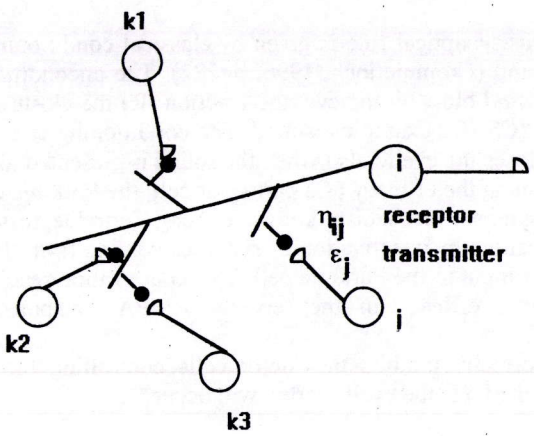

fig. 3. Schematic representation of inputs k1, k2, k3,.j received by a neuron i ; ε_j presynaptic efficacy and η_{ij} postsynaptic efficacy.

In figure 3 we represented axon-dentrite connections from one neuron unit to another. We see that there are two variables specifying the transmission efficacy : the presynaptic efficacy and the postsynaptic efficacy. Presynaptic efficacy ε_j is defined by the amount of transmitter released by cell j for a given depolarization. Postsynaptical efficacy η_{ij} is the local depolarization produced at postsynaptic processes for a given amount of released transmitter (Edelman, 1987, ch. 7).

We can now state two modification rules derived from the available experimental data (Edelman, 1987, p.183) :

1. **Presynaptic rule** : If the long-term average (over times of the order of 1 sec.) of the instantaneous presynaptic efficacy as determined by transmitter release exceeds a threshold, baseline presynaptic efficacy is modified :

$$\langle \varepsilon \rangle_{\approx 1\,\mathrm{sec}} \geq \theta \Rightarrow \varepsilon_{ref}\ modified$$

The long-term average of ε_j is a function of a large population of neurons connected with j.

2. **Postsynaptic rule** : Modification of the postsynaptic efficacy η_{ij} is a function of the stimulation to other synapses on the same neuron (heterosynaptic modification), coactivated heterosynaptical inputs to a neuron will alter η_{ij} :

$$\Delta\eta_{ij} = f(\varepsilon_{k1},\ldots,\varepsilon_{kn})$$

An example of presynaptic modification is given by research on learning in mollusks (Crommelinck, 1996, p. 210). In two forms of nonassociative conditioning of the gill-withdrawal reaction of the *Aplysia* mollusk - namely, sensitization and habituation - changes in ion conductance (concentration of Ca^{++}) lead respectively to the increase and decrease of neurotransmitter release from presynaptic neurons.

An example of the postsynaptical rule is given by classical conditioning experiments on the eyelid reflex of the rabit (Crommelinck, 1996, p. 222). The unconditioned stimulus (US) in this experiment is a wind blow on the eye, the reaction (R) the closure of the eyelid and the conditioned stimulus (CS) for example a sound. The conditioning creates an association US-CS so that the rabit closes the eyelid also when the sound is presented alone.
During these conditioning the efficacy of a particular cell, the Purkinje cell in the cerebellum, diminishes by a postsynaptic mechanism known as Long Term Depression (LTD). This Long Term Depression is caused by coactivation of synapses coming from the prarallel fiber input and the climbing fiber input to the Purkinje cell. This coactivation results in a decrease of the postsynaptic transmitter efficacy of the receptor AMPA responsible for the impulse transmission (figure 4).
The Purkinje cell normally inhibits the motor cells controlling the eyelid reflex. If the inhibition is lifted by the LTD the eyelid reflex will occur.

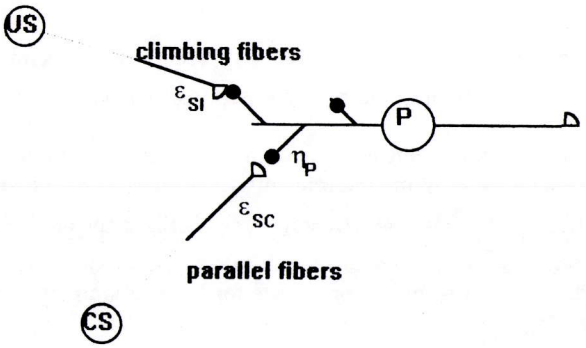

fig. 4. Heterosynaptic coactivation of synapses comming from the conditionned CS and unconditionned stimulus (US) ; P Purkinje Cell, ε et η pre- and postsynaptic efficacy respectively.

Using the formalisation given supra the post-synaptical rule for the Purkinje cell becomes :

$$\eta_P(AMPA) = f(\varepsilon_{CS}, \varepsilon_{US})$$

Finally, for sake of completeness, we have to state the numerous studies on Long Term Potentiation (LTP) in the hippocampus of the rabit. By artificial synchronous depolarization at high frequency (~ 100 Hz.) of both sides of the same synaps one observes an increase in the synaptic efficacy which can last from several hours, to days and even weeks. It's a mechanism instantiating Hebb's rule, by coactivation of two neurons one increases the synaptic weight, even if - due to the artificial caracter of the experiment - the link to the learning and memorisation capacities of the brain is difficult to state univocally (Crommelinck, 1996, p. 236-239).

2.3. Generalisation of Parallel Distributed Processing

Several researchers propose to extend the association rules in order to account for the experimental data. Essentially the suggestions turn around the pre- and postsynaptical rules one can observe in experiment. Fuster, for example, (Fuster, 1995) coins different terms for the association we can observe in the postsynaptical rule. It can be seen as a rule for sensory-

sensory association ("*Hebb's second rule*"), a rule for *synchronous convergence* of information or also to coincide with Miller's *principle of cooperativity* (Miller, 1991).

Here we will briefly discuss the propositions of G. Edelman. Especially the implications of the dual character of the two rules and the claim that the unit of selection/association is not the single neuron or synaptical connection, but a population of neurons or neuronal group.

The pre- and postsynaptical rule, together eventually with a Hebb-like rule or even other mechanisms, operate in a cooperative manner within the brain. If we combine both rules, we realise that synaptic alterations of a neuron i are not governed by correlated firing with one single neuron, but with a large population of other neurons. This is observed directly for the postsynaptical rule, but it is also valid for the presynaptical rule. Indeed, the long-term average of the presynaptical efficacy depends on the activity of all the neurons connected with the firing unit.
As a consequence a slight modification of the presynaptic efficacy in a certain group of neurons will cause a hierarchy of subsequent short-term modifications among various groups. We can say that multiple synaptic modifications occur simultaneously at various sites in the network. These multiple modifications are caused by one single synaptic modification, for example operating on the presynaptical level. This is the fundamental difference with classical parallel distributed processing, where the multiple modifications are caused by multiple parallel operating coactivations.
The degeneracy in the synaptical modifications of the network is of course transitory, if not the brain would give different answers at once to the same stimuli. In the theory of Edelman, where the brain dynamics follows the principles of natural selection, this degeneracy fulfills the need for a continual source of variation. After repeated interaction with the environment the most apt answer will be "reinforced" and selected on behalf of the others. In this the extension of the PDP approach that Edelman proposes is a refinement of the classical scheme of operant conditionning for animals, combined with a theory of percepetual categorisation (cfr. Edelman, 1987, p. 297).

3. From unsupervised learning to dynamical attractors

The self-organisation approach in the sense of unsupervised associative learning as we found it in the networks of Kohonen, Fuster and Edelman tries to satisfy as closely as possible the available empirical evidence on the brain. In this the generalised approaches seem to offer a more or less proper theoretical model of the functionning of the brain.
However a second constraint has to be satisfied, this time on the behavioural or cognitive level. The neural network not only wants to simulate mechanisms of neuronal transmission but also to reproduce some behavioural and cognitive properties, in particular learning, memorisation and pattern recognition.

The learning strategy adopted, in the classical as well as in the generalised approaches we discussed, is one of associative learning. It's a well-known principle in philosophy of mind suggested already by Aristotle. It was strongly developped within the anglo-saxon philosophy from Hobbes to J.S. Mill, before it was introduced in the emerging experimental psychology and neurophysiology (Boring, 1950). The basic tenet of the proponents of associative learning is one of empiricism, trying to derive all knowledge from associations between elementary sensations (Meyer, 1994, for a more detailed review). It is this tentative of the emprical philosophy that we find back in the neural networks :

"The most fascinating hypothesis, the most tempting dream is the one of a completely general system, capable of learning anything at all by examples. The properties of discrimination, generalisation and robustness of such a system should provide it with the capabilities, given an arbitrary series of examples, to develop a sensible "theory" about the domain. [...] The beauty of this hypothesis lies in the fact that it is based exclusively upon self-organisation properties induced by examples." (Serra, Zanarini, 1990, p.186).

However their is a serious flow in this dream. Human learning operates not only through presentation of examples, but also by explicit transmission of concepts, methods and techniques. This is of course a quite obvious remark and leaves open the real question, the estimation of the respective importance of the different learning mechanisms.
By presenting some experiments on visual perception, we would like to show that the associationist approach only is insufficient to account for perceptual categorisation. By formulating our examples in the language of dynamical systems theory we propose an account of perception which allows both for exemplar and conceptual constraints to operate in category formation. Another way to account for both is the construction of heterogeneous systems using both connectionist and symbolic approaches towards Artificial Intelligence (Serra, Zanarini, 1990, p. 192).

3.1. Dynamical systems theory

The modification rules for synaptical weights w in neural networks are stated in evolution equations $w(t+1) = F(w(t))$ or $dw/dt = F(w)$, with w taking real values. We briefly recapitulate some results from the general study of systems governed by equations of this type.

Starting from an initial value of the varables, in our example w, the system can evolve towards some time indendent behaviour, the steady state. This steady state is also called the attractor of the system and the set of initial values for which the system evolves to this attractor the basin of attractor. If the steady state is a fix value we talk about a point attractor, if it shows oscillatory behaviour the attractor is called a limit cycle. By extension one also speaks of an attractor of a system when the system evolves - after a transitory period - towards an unsteady solution, showing seemingly randomlike behaviour, which is called a chaotic attractor (Drazin, 1992, p.3). So we obtain :

- point attractor : fix value of w, equilibrium solution
- limit cycle : w oscillates between two values, periodic solution
- chaotic attractor : w has a seemingly random behavior with stationary statistical properties, aperiodic unsteady solution

It is easy to show that if $F(w)$ is a linear function the steady state is necessary a point attractor (it is sufficient to calculate w so that $w = F(w)$, which is obtained, if the solution exists, by inversing the matrice of the coefficients of F). So a necessary condition to have limit cycles and chaotic attractors is to have non-linear evolution equations.
Such a nonlinear system may regarded as a system with a feedback loop in which the output of an element is not proportional to its input (Drazin, 1992, p.1). As a typical example one can think of autocatalytic chemical systems where the product of a reaction occurs in its proper synthesis (Prigogine, 1979, p. 217).

3.2. Nonlinear brain dynamics

The unsupervised associative learning networks of Kohonen, Fuster and Edelman all account for an evolution towards point attractors. Hebb's rule or the extension to the pre- and postsynaptical association rules are linear principles (no feedback or feedforward loops). However, experimental evidence on visual perception seem to indicate that even low level visual processes can converge to more than one attractor, to limit cycles or even to chaotical attractors.

A first case is the case of binocular rivalry. Consider that by some optical trick, your right eye is shown something quite different from your left eye. What happens is that after a brief period the percepts start alternating at regular intervals, changing every few seconds (fi. 5.). The brain allows you to perceive only one of them at a time. This is called binocular rivalry. This phenomena was already described by Helmholtz in his *Physiological Optics* but it's only recently that we start to have a idea of the corresponding "rival pathways" in the brain (Crick, 1996 and Leopold & Logothetis, 1996).

fixed stimulus : right eye : / left eye : -
alternating perception : - / - / - t →

fig. 5. : example of a binocular rivalry experiment

As Leopold & Logothetis show, the rivalry is not simply occuring at an early stage in visual processing, when the images from the eyes are kept somewhat seperate. Instead evidence suggests that the rivalry is between alternative stimulus representations that are encoded in the activity of many neurons in different visual areas.
In any case, with binocular rivalry, we have a clear example of a perceptual change without any change in the stimulus. Other examples with a similar temporal dynamics of rivalry are obtained when viewing ambiguous figures, such as the Necker cube and other depth reversals (Leopold, Logothetis, 1996, p. 552)

An example of a limit cycle in visual processing is given by Zeeman using a sequence of 8 gradual changing pictures causing suddenly a change in perception (Zeeman, 1988). Here the stimulus is changing, first without any change in perception (in the experiment we observe a man's face), and then showing a sudden change to another percept (producing a kneeling woman), the bifurcation occuring when showing the 7th picture. When the series of pictures is showed again in the other sense, from picture n° 8 to picture n° 1, one observes the same phenomenon, but the change in perception occurs not at the same picture. Instead it occurs halfway at picture n° 4.
Zeeman explains this difference by constructing two different models of pattern recognition. In the first experiment we have an initial recognition and the brain behaves as a passive dynamical system. When passing the pictures a second time in the reverse order we have already some preliminary knowledge and the brain behaves as a cognitive system able to optimise its choice between the ambiguous perceptions. It makes an active choice to judge the likiest hypothesis.

Finally we mention the low-dimensional chaotic attractors one can observe by nonlinear time series analys of electroencephalogram recordings in some very particular cases of brain activity. On basis of their observations on chaotic brain dynamics Babloyantz and her colleagues constructed a model of a chaotic categorizer. Starting from a dynamical system

they use the unstable periodic orbits contained in a chaotical attractor as coding devices for incoming information (Babloyantz, Lourenco, 1994).

Thus, experimental evidence suggests to extend the linear associationist account of visual perception to non-linear dynamics (ex. through feedback and feedforward loops) in order to include the multiplicity of forms created by the autonomous brain dynamics. Indeed, as we have seen, perceptual change can occur in an autonomous manner without any change in the stimulus and even without being able to find a unique optimal solution, as in the case of binocular rivalry. To decide between the ambiguous perceptions an active intervention of a higher cognitive level is needed, which suggests a closer interaction between associationist and concept guided recognition processes.

References

Babloyantz A., Lourenco C. (1994), Proc. Natl. Acad. Sci. USA, Vol. 91, p.9027.

Blayo F., Verleysen M. (1996), Les réseaux de neurones artificiels, Presses Universitaires de France.

Boring E.G. (1950), A history of experimental psychology, 2ième édition revue, New York : Appleton-Century-Crofts.

Churchland Paul M. (1989), A neurocomputational perspective, The MIT Press.

Crick F. (1996), Visual perception : rivalry and consciousness, Nature, vol. 379, pp.485-486.

Crommelinck M., Boisacq-Schepens, N. (1996), Neuro-psycho-physiologie : tome 2, Masson.

Drazin P.G. (1992), Nonlinear systems, Cambridge University Press.

Edelman G.M. (1987), Neural Darwinism, Basic Books.

Fuster J.M. (1995), Memory in the cerebral cortex, MIT Press.

Hebb D.O. (1949), The organization of behaviour : a neuropsychological theory. New-York : Wiley.

Kohonen T. (1982), Self-organized formation of topologically correct feature maps, Biological cybernetics, 43, 59-69.

Leopold D.A., Logothetis N.K. (1996), Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry, Nature, vol. 379, pp.549-553.

Meyer M. (1994), La naissance de l'empirisme, in : Meyer, M. (1994) (ed.), La philosophie anglo-saxonne, PUF.

Serra R., Zanarini G. (1990), Complex systems and cognitive processes, Springer-Verlag.

Widrow B., Hoff M.(1960), Adaptive switching circuits, Proc. of the 1960 WESCON, 4, 96-140.

Zeeman Christopher (1988), Sudden changes of perception, in Petitot J. (ed.), Logos et Théorie des Catastrophes, Patino, Genève.