A BIOLOGICAL APPROACH TO PLASTICITY IN ARTIFICIAL NEURAL NETWORKS

G. Orlandi (*), F. Piazza, A. Uncini, A. Ascone (*)

Dip. di Elettronica ed Automatica, Università di Ancona, Via Brecce Bianche - 60131 Ancona, Italy
(*) Dip. INFOCOM, Università di Roma "La Sapienza", Via Eudossiana 18 - 00100 Roma, Italy

ABSTRACT

The plasticity of brain, i. e. the capability of changing neural connections over time, is an extensively studied phenomenon. Recently, it has been proposed that the network changes not only the synaptic strengths of connection but also partially its internal topology, according to either the received external stimuli and the pre-existent connection layout. We have found that this idea can be applied also to simple artificial neural networks. In the paper a new method is presented to dynamically adapt the topology of a neural network using only the information of the learning set. The method, although simply eliminates connections from an initial fully connected network, presents characteristics which can resemble some biological behaviour.

INTRODUCTION

It is well known that Artificial Neural Networks can solve a large variety of problems. In order to reach this goal, usually a formal model for the neuron is chosen and used to build a network with a given topology. Then this network is trained, adapting the synaptic weights of every neuron, by using a suitable learning algorithm.

The network topologies which are extensively used, are chosen among a set of known fixed models that are characterized by the presence of dense connection matrices among all neurons (fully connected network) or among groups of neurons, as in the well known Multi Layer Perceptron (MLP) [1]. These topologies however are very different from those found in biological neural systems, where the networks are often much less regular. Moreover, since usually a particular problem can be optimally solved by a network with a non-regular connection topology, the task of transforming an oversized regularly connected network into such optimal non-regular network is completely committed to the learning algorithm. Most of known learning algorithms, however, are not able to produce a reduced topology by zeroing some synaptic weights, instead they typically produce different structures spreading nonvanishing weights all over the network. The efficiency can result poor and several dimensions of a chosen configuration must be tried to yield an acceptable solution. In fact, if the network is too small, the input-output mapping cannot be learned with satisfactory accuracy. Conversely, if the network is too large, after a long training phase it will learn the given set correctly, but will badly generalize due to learning data overfitting [2,3]. Therefore, the goal when training a network is to find a topology large enough to learn the mapping and as small as possible to generalize correctly.

Two possible approach to produce networks with correct topologies have been proposed:
1) start with a small network and grow additional synapses and/or neurons until the desired behaviour is reached;
2) start with a large network and prune off synapses and/or neurons until the desired behaviour is mantained.

The first approach, although it could be efficient since most of the job is performed on networks of small size, can be difficult to realize and control in order to reach network topologies which are "optimal" in some senses. In [4] a single, non-optimal procedure to add neurons to a MLP is reported.

The second approach is much more studied and consists of finding a subset of network synaptic weights that, when set to zero, lead to the smallest increase of an error measure at the output. Several methods has been proposed in the last years in particular for the MLP. In [5] an entire neuron is pruned off if it does not change state or replicates another neuron after the presentation of the entire training data. The idea of weight decay is experimented in [6], where the synaptic weights with low influence on the output error during learning are pruned off. A sensitivity measure of the error function to the elimination of a unit or a synaptic weight is proposed in [7,8], and a-posteriori used to reduce the size of the network. A similar pruning procedure, based on the second derivative information, is reported in [9]. The proposed methods can be slow, requiring periodic retraining of the network, and often require a "supervisor", a non-local algorithm which performs to pruning task. Moreover, some of them can require a difficult fine-tuning of the pruning coefficients in order to work correctly.

Therefore it is highly desiderable to find new methods which are able to dynamically adapt the network topology to the problem. In this paper, using some observations on the biological network plasticity, an idea is introduced to develop a procedure for varying the network topology by simply pruning off connections in an initial overdimensioned fully connected network.

This work was supported in part by the Consiglio Nazionale delle Ricerche of Italy and in part by the Ministero della Pubblica Istruzione of Italy.

0-7803-0164-1/91/0000-0583301.00©1991 IEEE

NEURAL PLASTICITY

The plasticity of brain, i. e. the capability of changing neural connections over time, is an extensively studied phenomenon. In this field the work of Hubel and Wiesel on measuring changes in the ocular dominance statistics of a large neuron sample of the visual cortex in mammalians (particularly cat and monkey) are a fundamental milestone. They found that monocular deprivation drivers most neurons of the cortex to prefer the open eye, provided that the animal is still sufficiently young [10]. In particular it is now clear that the plasticity varies over the course of ontogenetic development and reaches a sharp maximum during the so-called critical period, during which a strong change of the brain structure is driven by external environmental stimuli.

Recently [11], it has been proposed that, during the critical period, the network changes not only the synaptic strengths of connection but also partially its internal topology. These changes are driven by either the received external stimuli and the pre-existent connection layout.

We have found that this idea can be applied also to simple artificial neural networks. By using only the external stimuli (i. e. input and output patterns) and the knowledge of the synaptic weights it is possible to adapt the network to the problem, changing concurrently the network synapses and the connection topology.

A simple procedure to prune off connection from a fully connected network is presented here. No new connections are created, although a similar approach could be followed to derive a procedure to add synapses to the network. In particular a neural network is considered in which there are input, hidden and output neurons and where all possible connections among these units are allowed, specifically including recursive connections and self- feedback connections. The algorithm for training the recursive network is derived in [12,13], where the authors have shown that the error signal as defined by the generalized delta rule, when calculated on the basis of the equilibrium state of the recursive network and then applied at the output of the network and allowed to backpropagate recursively, provides the proper final error signal at each unit for correcting the weights in a gradient descend sense. These fully interconnected networks are found to have often a greater learning capacity and better error characteristics than the corresponding feedforward networks.

In order to verify if a particular connection is necessary or not to solve a particular problem, a measure of the importance of the connection is necessary. According to the biological behaviour, this measure must be proportional either to the external stimuli and to pre-existent synaptic strength. A very simple formula can be derived by defining the following quantity:

- the synaptic activity of the connection from the i-th neuron to the j-th neuron relative to the p-th training pattern

$$a_{ij}^{(p)} = [\, w_{ij} \ Act_j^{(p)} \,]^2 \tag{1}$$

where w_{ij} is the synaptic weight of the connection between the i-th and j-th neurons, and $Act_j^{(p)}$ is the activation of the j-th neuron corresponding to the p-th input pattern. This activity can be averaged on the whole training set

$$\hat{a}_{ij} = \frac{\sum_{p=1}^{N_p} a_{ij}^{(p)}}{N_p} \tag{2}$$

where N_p is the number of the input training patterns. Since each neuron is driven by several different synapses, an average total synaptic activity relative to the i-th neuron can be defined

$$A_i = \sum_{j=1}^{N_i} \hat{a}_{ij} \tag{3}$$

where N_i is the total number of inputs to the i-th neuron. A measure of importance of the synapsis can be obtained by comparing its averaged activity with respect to the the total average activity relative to the neuron which the synapsis belongs to. Therefore the Percentage Average Synaptic Activity (PASA) can be defined as follows

$$PASA_{ij} = \frac{100\, \hat{a}_{ij}}{A_i} \tag{4}$$

During the learning phase the previous quantities vary until they reach stable values. The idea, used for changing the network topology, consists in dynamically eliminating all the connections which present a PASA lower than a threshold value varying according to a predefined curve.

In the forward step of the learning phase, in each neuron the synaptic activity of every connection relative to the current training pattern, is computed and accumulated. At the end of the training set all the PASA's can be computed. These values are compared with a threshold Th(p) which is a function of the epochs in a way depending on the particular problem. The connections with a PASA lower than the threshold value are pruned off.

It can be noted that this algorithm is local in the sense that each neuron requires only local quantities to compute the PASA's. Moreover, the computational cost is low, since:

- the calculation of the synaptic activity per training pattern requires one multiplication/accumulation operation per synapsis (the multiplication can be avoided by using the absolute value instead of the square value in (1);
- the calculation of the PASA's of each connection relative to the i-th neuron at the end of one training set requires (N_i-1) sums, one division and one multiplication per synapsis.

PRELIMINARY EXPERIMENTAL RESULTS

The proposed algorithm has been tested on several simple applications. In particular the experimental results relative to the XOR and the binary 2x2 bits multiplication problems are reported.

XOR:

The initial fully connected network was composed of 3 neurons with a total of 18 parameters (weights of the connections and offsets). The PASA's has been further averaged on 10 epochs in the learning phase. The threshold Th(p) varied as shown in Fig. 1a., while the learning rate was varied from 2 to 5.

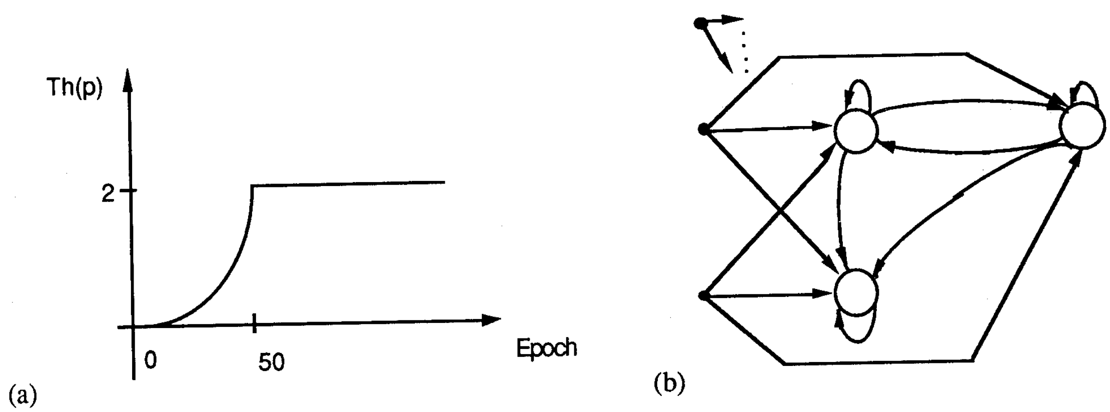

Fig. 1 (a) Threshold function Th(p) used in the XOR experiment. If the PASA of a connection is lower than the threshold at the p-th epoch, the connection is eliminated. (b) Network topology obtained in the most of the cases of the XOR problem. The lower left neuron is useless since it has no output connections.

Several trials were carried out using different initial sets of weights and offsets. In the most of the cases (about 70%) the topology in Fig. 1b was obtained. Between the two left neurons a winner is established (the upper one) and the other results to have no output connections, therefore the canonical 2 neuron topology was reached. In the remaining cases, the algorithm obtains feedforward networks by eliminating the connections between the two hidden neuron and the feedbacks between the output neuron and these neurons.

BINARY MULTIPLICATION:

The initial fully connected network used in this case was composed of 5 neurons with a total 50 parameters. The PASA's have been further averaged on 10 epochs during the learning phase. Several trials were carried out using different initial sets of weights and offsets and various threshold curves have been tested for pruning off. Good results, in terms of convergence rate and pruning capability, have been obtained by using the curve in Fig. 2 which closely resembles the critical period of the biological neural networks. The proposed algorithm is able to eliminate up to 40% of the number of synapses depending on the initial sets of weights and offsets and the learning phase converges in the large majority of the cases.

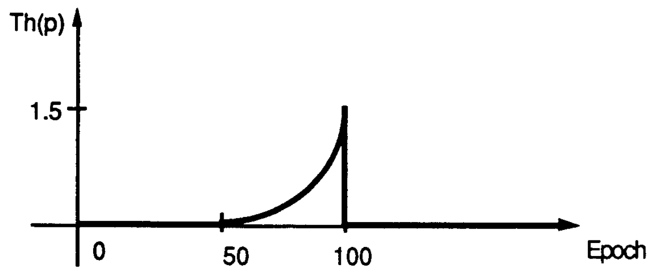

Fig. 2 Threshold function Th(p) used in the binary multiplication experiment. Note the resemblance with the biological "critical period".

Two sets of experiments were carried out after the proposed algorithm had reduced the original network by pruning off several connections. In the first the obtained topology was tested by performing a new learning phase on pruned networks loaded with small random set of weights and offsets. The performance of these networks were at least as good as those of the original fully connected network and often they worked better. On the contrary, in the second set of experiments several networks were built from the initial fully connected network by pruning off the same number of connections eliminated by the proposed algorithm but at random positions. These networks, during the learning phase, gave very poor results, often they did not converge at all, with respect to the equivalent network obtained by the proposed algorithm.

It is possible also to iterate the algorithm, by using an already pruned network as initial network, to further reduce the number of synaptic connections. Using this technique, we have obtained a network with 28 parameters which contains 8 parameters less than the canonical 4 neuron network used to solve the 2 bit multiplication problem.

CONCLUSIONS

In the paper a new method has been presented to dynamically adapt the topology of the neural network using only the information of the learning set. The method, although simply eliminates connections from an initial fully connected network without creating new synapses, presents characteristics which can resemble some biological behaviours like the so-called critical period. The future work is addressed to obtain a deeper understanding of the algorithm, also in terms of optimality, robustness and generalization capability. An application to a real-world problem will show better the performance of the method with respect to other pruning approaches. Finally, we want to extend the same approach also to the creation of new synapses.

REFERENCES

[1] D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, Parallel Distributed Processing (PDP): Exploration in the Microstructure of Cognition (Vol. 1), MIT Press, Cambridge (MA), 1986.
[2] E.B. Baum, D. Haussler, "What size net gives valid generalization?", Neural Computation, No. 1, 1989 , pp. 151-160.
[3] K.A. Huyser, M.A. Horowitz, "Generalization in digital functions",Neural Networks,No.1,1988,pp. 101.
[4] T. Ash , "Dynamic node creation in backpropagation networks", ICS Report 8901, UCSD Feb. 1989.
[5] J. Sietsma and R.J.F. Dow, "Neural net pruning - Why and how?", in Proc. IEEE Int. Conf. on Neural Network, vol. 1, San Diego, CA, 1988, pp. 325-332.
[6] S.J. Hanson and L.Y. Pratt, "Comparing biases for minimal network construction will back-propagation", in Advances in Neural Inf. Proc. 1, D.S. Tourezky, Ed. Morgan Kaufman, 1989, pp. 177-185.
[7] M.C. Mozer and P. Smolensky, "Skeletonization: A technique for trimming the fat from a network via relevance assessment", in Advances in Neural Information Processing 1, D.S. Tourezky, Ed. Morgan Kaufman, 1989, pp. 107-115.
[8] E.D. Karin, "A Simple Procedure for Pruning Back-Propagation Trained Neural Networks", IEEE Trans. on Neural Networks, Vol. 1 No. 2, June 1990.
[9] Y. Le Cunn, J.S. Denker, S.A. Solla,"Optimal Brain Damage", Computer and System Science.
[10] D.H. Hubel, T.N. Wiesel, "The period of susceptibility to the physiological effects of unilateral eye closure in kittens",Journal of Physiology, No. 206, pp. 419-436.
[11] C. Aoki, P. Siekevitz, "Ontogenetic changes in the Cyclic Adenosine 3',5'-Monophosphate-stimulatable phosphorylation of cat visual cortex proteins, particularly of microtubule-associated protein 2 (MAP2): effects of normal and dark rearing and of the exposure to light", Jou. of Neuroscience, No. 9-5, Set. 1985.
[12] F.J. Pineda, "Recurrent backpropagation and the dynamical approach to adaptive neural computation", Neural Computation, No. 1, 1989, pp. 161-172.
[13] L.B. Almeida, "A learning rule for asynchronous perceptrons with feedback in a combinatorial environment", Proc. Int. Conf. on Neural Network, San Diego 1987, pp. II.609-II.618.