Proceedings of 1993 International Joint Conference on Neural Networks

LEARNING IN BINARY NEURAL NETS WHICH HAVE FUZZY REASONING

José Alberto Baptista Tomé
Instituto Superior Técnico
INESC
R.Alves Redol nº9,1000 Lisboa, Portugal
E-mail jbt@eniac.inesc.pt

ABSTRACT A binary neural model is presented, which allows for a natural way of learning and reasoning. Learning is a two phase process where first connections are established and then functions defined. An important fact is that this learning is completely unsupervised and based only in neural activity induced by experiments. It is shown that fuzzy reasoning is naturally induced in this model without any centralized commands or explicitation. Learning may be performed in a mixed way, that is more then one rule at a time, and with any degree of uncertainty about the rules at the moment of learning.

1. Introduction

Binary neural nets have been used for different purposes such as pattern recognition (1,2), image processing(3) and universal logic function construction (4).A simple binary neural model has been presented (6) which allows non supervised learning and reasoning of the fuzzy type. It is partially inspired on some views for the thinking as having the biological basis of great number of connections between cards in the brain(5). Here the learning of these nets is addressed. It is shown that learning automatically develops connections and establish logic functions between areas of neurons which may be interpreted as fuzzy predicates, if the density of active neurons represents the membership grades of those predicates. It may be stated that fuzzy reasoning is the natural behaviour of the net, even if it was not taught as such!

2.The Model

2.1 Macro-level organization

Neurons are organized in layers, each layer with different areas (similar to brain cards) and each area divided into sub-areas. One associates to each area the concept of variable, and to each sub-area the concept of predicate of a variable. These predicates are taken as fuzzy predicates (e.g. *low,medium, high* for the variable *temperature*). This is true if the *density of active neurons in each sub-area represents the membership grade of the associated predicate*. See Figure 1 to visualize the concept. There, three sub-areas with different densities, try to define the membership grades of the variable "age" to the fuzzy sets "child","young" and "old" for a person with 20 years of age.

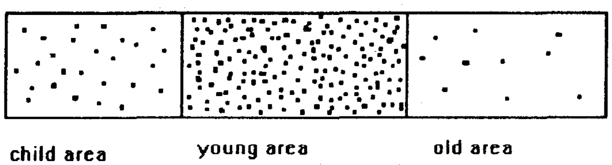

Figure 1

If reasoning is expressed by expressions of the type:

IF A_1^1 **AND** A_2^1 **AND .. THEN** C

the Ai and C are the above predicates . For the implementation of a given rule, as above, a great number of random connections is established from sub-areas A_i (the antecedent areas) to sub-area C (the consequent area) . Moreover, if a set J of different rules have the same consequent variable predicate then connections to sub-area C come from every A_i^j, for all $j \in J$.

2.2 Micro level Organization

Neurons are binary and the connections among them, from sub-areas to sub-areas, are random in nature. Operations performed at each neuron are Boolean , as follows:

For each neuron in the consequent rule sub-area an input is taken from a randomly chosen neuron from each of the antecedent sub-areas present in that rule and the **Boolean AND** operation is performed. If more than one rule with the same consequent is being processed then the **Boolean OR** operation is performed over the results of the above andings. If more than one rule is activated the neurons become Universal Logic Functions (ULF). These micro operations ***induce at the macro level the fuzzy AND and fuzzy OR operations***, if the fuzzy operations are defined as follows:

1. Fuzzy Complement:
   c(p)=1-p if p is a probability

2. Fuzzy AND (t norm) :
   pa.pb, that is the probability product.
3. Fuzzy OR (s norm)
   pa+pb-pa.pb, the probability union.

Therefore, the logic function performed by each neuron in the sub-area C (that is the sub-area of the consequent of the rule) may be defined as:

$$\Sigma_r (\Pi_{ar}\, o_{ar}^{ri})$$

where Σ_r is the ***ORING*** over every rule r which has C as consequent, Π_{ar} is the ***ANDING*** over every antecedent sub-area ***ar*** of rule r, and o_{ar}^{ri} is the output of a random neuron from sub-area ***ar***.

In order to visualize at a macro-level, figure 2 shows the connections between sub-areas for the implementation of the following rules:

IF A2 AND B1 THEN C2
IF A3 AND B1 THEN C2
IF A4 AND B4 THEN D3

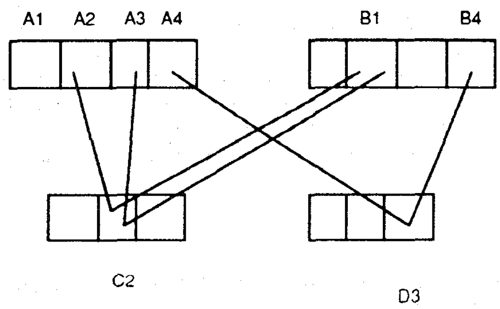

Fig. 2

3.Learning

Two completely different learning mechanisms are supported in this model: the first, a centralized one, which may be denominated by *cognitive learning*, since fuzzy rules are taught as such; and the second *the generic or fuzzy learning*, with more interest, where fuzzy rules are taught by experience and possibly in mixed form.

3.1 Cognitive Learning

This is straightforward. If a given fuzzy rule is to be taught :

IF A_1^1 **AND** A_2^1 **AND** ..A_j^1 **THEN** C,

every neuron in antecedent sub-areas A_i^1 and in the consequent area C is activated by the supervisor, that is, the rule is in some way explicitly stated. The net automatic learning, following this neuron activation, is a process by which to each one of the neurons in area C, a set of j randomly chosen neurons - one from each sub-area A_1^1 - is assigned to the j inputs of that neuron.

Such a topology allows ,in a reasoning stage (implementation of the rule), the following neuron operation:

$\Pi_{ar}\, o_{ar}^{ri}$, where Π_{ar} is the ***ANDING*** over antecedent sub-areas ***ar*** of the rule r, and o_{ar}^{ri} is the output of a random neuron from sub-area ***ar***. The individual neuron reduces to an AND.

3.2 Fuzzy learning

In this more generic type of learning, two stages may be considered:

1. A variable selection phase (also sub-area selection), when a "filtering" process takes place in order to eliminate antecedent variables which are not relevant for some consequent. In this phase connections are chosen.
2. A phase of *n-tupple* teaching when some memorization takes place, in order to establish the logical function of each neuron.

3.2.1 Selection Phase

A situation which is supposed to be relevant for the non explicit learning of the rules which relates antecedent variables with a given consequent variable, is denominated a *shot*. For each shot no knowledge is supposed to exist, that is, the antecedent or consequent predicates are not known, and neural activation is the only basis for learning.

This configuration or topologic learning behaves as follows:

For each shot, and for each activated neuron $\mathbf{n_k}$ in the consequent area, n random neurons are taken from each of the areas associated with possible antecedent variables. If any set of these n neurons has every neuron at "1" (that is the logical AND is also "1") an input from one of these neurons is established to $\mathbf{n_k}$.
If the non-explicit fuzzy rule being taught refers mostly to one predicate variable as antecedent then this will be apparent on the learning shots by activating highly that sub-area on these shots and not substantially activating the other sub-areas. As a result the AND of n neurons will give "1" if the n-tupple comes all from that sub-area with high probability and "0" to other sub-areas , also with high probability.
That is, a connection between an antecedent sub-area *A* and a consequent generalized sub-area *C* for a given shot is established with probability:

$p_i=p_A{}^n.p_C.k_i$, where p_A and p_C are the probabilities of a given neuron in those generalized sub-areas to be active for that learning shot and ki is a weighting factor (<1). Notice the effect of n : the greater it is, more "noise filtering" is realized, in the sense that sub-areas which are not intensively activated are disregarded.
This same concept can be applicable if a set of non explicit fuzzy rules are taught and not only one.
For example, if the selection phase uses experiments which are equivalent to a set of fuzzy rules:

IF $A_1{}^1$ AND $A_2{}^1$ AND ..$A_j{}^1$ THEN C

....

IF $A_1{}^k$ AND $A_2{}^k$ AND ..$A_j{}^k$ THEN C

neurons in generalized sub-area C are expected to have inputs from neurons in generalized sub-areas $A_1{}^1$, $A_2{}^1$ ...$A_1{}^k$...$A_j{}^k$.

3.2.2 N-Tupple Teaching

It is supposed that in each experiment the antecedent and consequent areas may be activated but not completely. This induces the concept of *credibility* of a rule in a teaching experiment (or in a sequence of experiments):
Credibility *is the membership grade of the consequent obtained by the application of the rule, when the antecedent sub-areas are completely (100%) activated and after having taught the net with that experiment (or set of experiments).*
As the membership grade is the probability that a random neuron is activated,the expressions for credibilities are easily derived.
Two processes may be used for teaching:
1. During the teaching experiment, and for each consequent neuron (Nc), the ULF bit associated with the n-tupple configuration for that neuron is set to "1" if that neuron is activated. If Nc=0 the ULF is unchanged. This is a *monotonic* type of teaching!
2. The ULF bit is set to the logical value of Nc. It is a *non-monotonic* teaching!

The credibilities are then:

Monotonic:

If sets of *mutually exclusive rules* (the antecedents of any one are not a sub set of the antecedents of any other) the credibility of such a set of rules in experiment E is:

$P_E{}^{set}= \Sigma_i P_E{}^i$, where i is the enumeration index of rules
with the credibility of rule i:
$P_E{}^i= \Pi_j p_{Aj}{}^i \cdot \Pi_k (1-p_{bk}{}^i) \cdot p_{nc}$ or $P_E{}^i=Q_E{}^i \cdot p_{nc}$

where $p_{Aj}{}^i$ is the probability that a random neuron in antecedent sub-area $A_j{}^i$ of rule i, is activated, $p_{bk}{}^i$ has the same meaning but for antecedents of other rules which are not antecedents of rule i and p_{nc} the same for the consequent area C. Σ and Π represent normal algebric sums and products. The individual operation "." is also the algebric product.
For a sequence of experiments one obtains the credibilities:
$P_{ACUM(k+1)}{}^{set}= P_{ACUM(k)}{}^{Set} +P_E{}^{set}(k+1) - P_{ACUM(k)}{}^{Set} . P_E{}^{set}(k+1)$
$P_{ACUM}{}^i(k+1) = P_{ACUM}{}^i(k) + P_{E(k+1)}{}^i - P_{ACUM}{}^i(k) . P_{E(k+1)}{}^i$
and where k is the sequence experiment index.

Non-Monotonic

The expressions are similar, with the following difference:
$P_{ACUM}{}^i(k+1) = P_{ACUM}{}^i(k).(1-(Q_E{}^i(k+1)).(1-Pcn))+(1-P_{ACUM}{}^i(k)).P_{E(k+1)}{}^i$
and similarly for Set expression.
If each experiment affects only a fraction β of the consequent neurons, and if these fractions are mutually exclusive, the final credibility is:

$P_{Global}{}^{Set}=\Sigma\ \beta\ P^{Set} = \beta\ \Sigma\ P^{Set}$, where sum is over all β
The expression is valid if index Set is substituted by index i.
This means that an average is made of the fraction credibilities. This feature may be important if there are some wrong teaching experiments.

It should be stressed that in the consequent sub area, every possible rule compatible with the antecedents is simultaneously present because each neuron at this layer is associated with an ULF. That is, the density of activated i^{th} bits in the consequent layer is the membership grade of the consequent fuzzy predicate if rule i is applied with its antecedent sub areas activated at 100%. Moreover in the *utilization phase* of the net the overall density of activated neurons at the consequent sub area may be divided into disjoint densities of bit1,bit2,...,bit 2^n (since for each neuron only one of its ULF bits is activated). Thus the *overall membership grade of predicate C is the addition of parcelar membership grades of each one of the rules in the set.*

Example

To clarify consider a net with a variable A with 2 predicates (sub areas A1 and A2) and a variable B with only one predicate. Consider the monotonic teaching:

If every possible sub area is considered 8 different mutually exclusive rules may be considered:

R0: IF NOT A1 AND NOT A2 AND NOT B THEN C

R1: IF NOT A1 AND NOT A2 AND B THEN C

....

R7: IF A1 AND A2 AND B THEN C

Consider also that 3 experiments have been made to teach the net with the following activations:

pA1=.98 pA2=.02 pB=.95 pC=.98

pA1=.95 pA2=.2 pB=.9 pC=.9

pA1=.02 pA2=.97 pB=.02 pC=0.95

Obviously, no cognitive learning is being done (if this was the case an experiment with 100% of sub areas activation should be enough), but one may detect, with some uncertainty, the objective of teaching the rules:

Rule A: IF A1 AND B THEN C

Rule B: IF A2 AND NOT B THEN C,

although the second experiment could indicate also some teaching of rule

IF A2 AND B THEN C (Rule C) with little confidence.

Applying expressions above the teaching would produce the following credibilities for the exclusive rules:

Rule 5:~ .96 ; Rule 7~.17 ; Rule 6 ~.03 ; Rule 2~.89 ; Rule 3 ~.025. Every other rule has a very low credibility.

Credibility of rule A is the sum of credibilities of exclusive rules 5 and 7 minus their product, and the credibility of rule B is the sum of credibilities of exclusive rules 2 and 6 minus their product. The credibility of rule C should be the sum of credibilities of rules 7 and 3 also minus their product, which is, as expected, a "low" credibility.

Finally, using the net after the above learning, if antecedent sub areas are activated with the following rates:

A1: .1 ; A2: .93 ; B: .01,

the activations of neurons in the consequent sub area are due mostly to ULFs bit 2 (exclusive rule 2) with 73.8% and bit 6 with 2.7% of activations. All the other ULF bits have a very low activation, and then it may be stated that Rule B has operated almost exclusively, as expected!

4. CONCLUSIONS

These nets present great potentialities for moddeling natural reasoning and learning. Their topology is similar to natural neural topologies: neurons are aggregated in cards at a macro level and connected by dense meshes of random connections at a micro level. Altough it is a very simple neural model it has enough power to learn sets of non-declared fuzzy rules in a supervised or non supervised way. It may be concluded that the natural behaviour of these neural nets is the fuzzy reasoning, even if it the teaching was not done as such.

REFERENCES

1.ALEKSANDER;I: "The logic of Connectionist Systems"- personal communication

2.TOME,J., "Low order arrays for geometric pattern detection" Proc. of IEEE Computer Architecture for Pattern analysis and Data base Management,Miami 1985

3.DUFF,M.,Fountain,T., "Cellular Logic Image Processing", Ac. Press 1986

4.VIDAL,J., "Implementing neural nets with programmable Logic" IEEE Transactions on Acout.,Speech and Signal Processing, July1988

5. EDELMAN,G., "Biologie de la Conscience" , Editions Odile Jacob , 1992.

6. TOME,J., A Simple Neural Model for Fuzzy Reasoning; Proceedings of the Second International Conference on Fuzzy Systems, pp624-628.

7.ZADEH,L.A.,"Fuzzy Sets", Information and Control,8,1965

8.ZADEH,L.A.,"Fuzzy Logic and Approximate Reasoning",Synthese, 30,1975

9.KAUFMANN,A.,Gupta,M., Introduction to Fuzzy Arithmetic : Theory and Applications, New York: Van Nostrand, 1985

10.KLIR,G.,FOLGER,T., Fuzzy Sets,Uncertainty and Information, Englewood Cliffs, Prentice Hall, 1988

11.MAMDANI,E.H.,GAINES,B.R., Fuzzy Reasoning and Applications, Academic Press, 1981

12. WITOLD,P., Fuzzy Control and Fuzzy Systems, Research Studies Press Ltd., 1989

13.TERANO,T.,SUGENO,M.,MUKAIDONO,M.SHIGEMASU,K., Fuzzy Engineering toward Human Friendly Systems, Proceedings of the International Fuzzy Enginnering Symposium '91, Yokohama,1991