LIMITATIONS ON COMPLEXITY OF RANDOM LEARNING NETWORKS

FRANKLIN OFFNER

From the Bio-Medical Engineering Center, The Technological Institute, Northwestern University, Evanston

ABSTRACT Randomly connected networks can be made adaptive, and thus able to "learn." Signal-to-noise considerations are shown to limit the maximum initial complexity which can learn. A higher order of complexity may be possible in multilayered structures which learn layer-by-layer; or if learning is possible during construction. Perception-like devices would appear not to be operative if of a high order of complexity.

Randomly connected matrices composed of excitable elements can at least to some extent be made adaptive, so that "learning" is possible. The hypothesis has been made that the nervous system has features in common with such a system. While it might seem that an increase in complexity of randomly connected nets would increase adaptivity, it appears that there is a maximum complexity of the initial random connections, beyond which the ability to learn will be lost. This is the result of a degradation of the signal-to-noise ratio below a useable value.

Consider first two matrices. The first contains R "receptor" elements: the second, A "associative" elements. It is desired that a given stimulus pattern falling on the R matrix should eventually (after sufficient "teaching") cause the excitation of a single A element. If all R elements are to be effective, all, or substantially all, must be connected to each A element. These connections will initially be randomly distributed, with both excitatory and inhibitory connections.

Learning would presumably be accomplished in the following manner: the pattern to be recognized is exposed to the R matrix. One, several, or perhaps many A elements received an excitation above a threshold[1] value. The excitatory connections to these elements are strengthened, and the inhibitory ones weakened. The process is repeated as additional examples of the same pattern are presented. For the system to even start to be operative, it is obviously necessary that it be possible to detect a differential rise in excitation between elements. The conditions for this will now be examined.

[1] The argument is not affected if graded, rather than threshold, response is considered.

To each A element, there are on an average a total of $\bar{N}$ connections. It has been assumed that $\bar{N}$ is of the same order as R, the number of elements in the first (receptor) matrix, but this assumption is not necessary to the argument. There are, on an average, $\bar{E}$ excitatory connections and $\bar{I}$ inhibitory on each A element. $\bar{E} + \bar{I} = \bar{N}$.

If all A elements were connected in the same manner, there would be no ability to distinguish between patterns exposed to the R matrix: the discriminatory ability is due to the departure from the average.

To determine the ability of the matrix to discriminate one pattern from another, consider first the simplest task: to distinguish illumination of the right half of the field from the left. This means that some (or at least one) A elements must have an excitation level (E connections stimulated, minus I connections stimulated[2]) significantly different from the average. 32 per cent will have a difference in level greater than the standard deviation of $\bar{E} - \bar{I}$ in the illuminated half:

$$\sigma_{E-I}{}^2 = \sigma_E{}^2 + \sigma_I{}^2$$

But

$$\sigma_E{}^2 = \bar{E}$$

$$\sigma_I{}^2 = \bar{I}$$

Therefore

$$\sigma_{E-I} = \sqrt{\bar{E} + \bar{I}} = \sqrt{\bar{N}}$$

The question now must be examined, as to what constitutes a "significant" difference; *i.e.*, that which could produce a detectable response. Two limitations on the minimum value of excitation ($E - I$) are apparent. The first is the precision with which the A element can determine the difference in excitation produced by the pattern, as compared with the background, which is $\bar{E} - \bar{I}$. Thus, if the excitation produced by illuminating the whole field (or the opposite half) is 1000 units, and that produced by the pattern is 1001, the A element must have a threshold precision of 0.1 per cent. While this could be accomplished by digital computer simulation of a learning network, it would certainly be well beyond the reliable capability of a neural element.

This criterion of detectability is then the standard deviation of the excitation level, divided by the total excitation:

$$D_1 = \sqrt{(\bar{E} + \bar{I})}/(\bar{E} - \bar{I}) = \frac{1 + k}{1 - k}\frac{1}{\sqrt{\bar{N}}}$$

where $k = \bar{I}/\bar{E}$. Evidently on this basis, if $k = 1$, *i.e.* if on an average there is no

---

[2] There is evidence that inhibition in the nervous system may be a proportioning, rather than a subtractive, process. This would affect only minor details of this analysis.

net excitation from a pattern, then the detectability would not be decreased by increasing $\bar{N}$. However, any net average surplus of **E** (or **I**) connections would place a limit on $\bar{N}$.

As a numerical example, if it be assumed that the threshold of the **A** element is stable to 10 per cent, so that the minimum value for $D_1 = 0.1$, and that $k = 0.8$, then

$$\bar{N} = \left(\frac{1.8}{0.2 \times 0.1}\right)^2$$

and the maximum value for $\bar{N}$ is 8100.

A second, and probably more important, limitation is the stability of the excitatory and inhibitory signals furnished by the **E** and the **I** connections. If the response transmitted by such connections is variable with an RMS fractional variability v, the standard deviation of E is

$$\sigma_e = vE$$

and

$$\sigma_i = vI,$$

the standard deviations of excitatory and inhibitory connection signals to any one **A** element. The standard deviation of the resultant net excitatory (or inhibitory) signal is then

$$\sigma_{e-i} = v\sqrt{E^2 + I^2} = v\bar{E}\sqrt{1 + k^2} = v\bar{N}\frac{\sqrt{1 + k^2}}{1 + k}$$

The criterion for detectability based on variability of response is thus

$$D_2 = \frac{\sigma_{E-I}}{\sigma_{e-i}} = \frac{\sqrt{\bar{N}}}{v\bar{N}}\frac{1 + k}{\sqrt{1 + k^2}} = \frac{1}{v\sqrt{\bar{N}}}\frac{1 + k}{\sqrt{1 + k^2}}$$

Since in this expression the second factor can only vary between unity and $\sqrt{2}$, for the rough accuracy of the present treatment it will be neglected. Then if $v = 0.1$, for $D_2 = 1$ (unit "signal-to-noise" ratio), $\bar{N}$ may not exceed 100; that is, each **A** element may on an average have not more than a total of 100 connections to **R** elements, distributed between excitatory and inhibitory.

The above estimates are based on having any **A** element within one standard deviation-width able to respond. Some improvement results if we only require that fewer elements be probably able to respond. If one element in a thousand is to respond, $\bar{N}$ may be increased by a factor of 3.3; if one in a hundred million, the factor remains less than 5.

It thus appears that if the elements of the two matrices are to in any respect simulate neurons, so that there will be a definite limit to the constancy of threshold and of impinging impulses, the number of random connections to each of the associative elements will be limited to a relatively low number, of the order of less than a few

thousand. However, to distinguish complicated patterns, there must be eventual connection to millions of receptor elements.

How can this be accomplished? One way is by learning during the process of development. If only a few connections are functioning when the system is first activated, these may be partially organized by learning before additional complexity occurs due to continued development. This then reduces the randomicity of connections so that the above derived limitations do not apply.

A second way is by employing a succession of matrices ("layers"). Each associative **A** matrix is randomly connected to the preceding matrix. The multiplicity of resulting interconnections increases the number of elements in the **R** matrix connected to each element of the $\mathbf{A}_i$ matrix approximately as $\bar{N}^i$, where as before $\bar{N}$ is the average number of connections to each of the A_i elements. More specifically, if F_1 is the fraction of the R elements not connected to one of the A_1 elements of the first associative matrix,

$$F_1 = (1 - \bar{N}/R)^{A_1}$$
$$= \exp(-A_1\bar{N}/R) \qquad (\bar{N}/R \ll 1)$$

The fraction F_n of R elements not connected with *each* of the $\mathbf{A}_n$ elements in the nth A matrix is then:

$$F_n = \exp - B_1(1 - \exp - B_2(1 - \exp - B_3(1 - \cdots \exp - B_{n-1}\bar{N}_n/A_{n-1})))\cdots)$$

where $B_i = \bar{N}_iA_i/A_{i-1}$, with $A_{-1} = R$, the first matrix being considered the receptor (R) matrix.

If $\bar{N}_i \gg A_i/A_{i-1}$ and $A_n \gg \bar{N}_n$, which appears to be probably true in practical examples, the above simplifies to

$$F_n = \exp(-B_1B_2B_3 \cdots B_{n-1}\bar{N}_n/A_{n-1})$$
$$= \exp(-\bar{N}_1\bar{N}_2\bar{N}_3\bar{N}_4 \cdots \bar{N}_n/R)$$

Thus the probability of an element in the final layer being associated with an element in the receptor matrix depends to the degree the approximations made herein hold, only on the number of layers, and the average number of connections to each element of each layer, but not on the number of elements in any layer after the first (receptor).

To find an order of magnitude, consider there are 4×10^8 **R** elements. (This is the order of magnitude in the human retina.) Consider that there are eight matrices successively joined, corresponding to the visual system, to the last layer of the striate cortex. Then if on the average each **A** element connects to twenty or more elements of the preceding layer, each element of the last matrix will on an average be associated with 95 per cent of the elements of the **R** matrix.

Using this value in the expressions for detectability, for a detectability of unity

(representing a signal-to-noise ratio of approximately one, and therefore probably the lower limit of operability),

$$D_2 \approx 1/v\sqrt{\bar{N}} = 1/v\sqrt{20} = 1$$

$$v \approx 0.22$$

With such a small number of connections, the criterion of precision of threshold of detection would probably not be significant: even with $k = 0$ (all connections excitatory), the threshold stability need be only 22 per cent to achieve operability.

CONCLUSIONS

Considerations of detectability appear to limit the maximum initial complexity of a randomly connected system. Practical considerations indicate that the maximum operable degree of initial complexity in a "perceptron"-like system(1) would be insufficient to provide useful perception. Two ways out of this impasse are apparent. The first is learning during the developmental process of the system. The second is a layered structure, so that even after development, complexity becomes of a high order only in deeper layers. Dependence on such a structure for avoiding the problem of excess complexity implies that learning must be possible in early layers, without reference to the signal produced in later layers.

It is apparent that if no considerations of variability of response, or of sensitivity or stability of threshold detection were involved, the criteria of detectability here treated would disappear. Such considerations do not appear in digital computer simulations of learning networks, unless explicitly included.

The detectability criteria presented should give an indication of the upper limit to the initial complexity of a random network that could learn. No statement is made that it can in fact learn; the extent to which learning is possible must be determined by other analyses.

From the considerations presented herein, it appears that perceptron-like analog devices, in which decisions are made only on the basis of the output of the last matrix, will not be operative beyond a rather low order of complexity. Increasing the elaboration of such devices would therefore not increase their learning ability (as seems to have been implicitly assumed), but rather the contrary.

A number of approximations have been made in the development, which are not even in accord with the simple known features of an actual perceptual neural system. The initial strengths of excitatory and inhibitory connections have been assumed uniform. A random distribution would leave the net excitation randomly distributed, and should not affect the conclusions. The assumption of only layer-to-adjacent layer connections is certainly false: there are collateral connections in all aspects of the nervous system, and at least in the cortex the layer structure is only an approximation, so that connections both penetrate more than one layer, and also

proceed retroactively ("feedback"). It appears that departures from the layered structure would tend to lower the allowable complexity.

The nervous system may not, in fact, be initially random in connection. Hubel and Wiesel(2) find that there is at least partial organization in the cortex of kittens who have had no previous visual experience.

This work was supported in part by National Institutes of Health contract FR-00018-03.

Received for publication, March 9, 1964.

REFERENCES

1. ROSENBLATT, F., Principles of Neurodynamics, Washington, D. C., Spartan Books, 1962.
2. HUBEL, D. H., and WIESEL, T. N., *J. Neurophysiol.*, 1963, **26**, 994.