J. Indian Inst. Sci., July–August 1998, 78, 249–256.

Neural dynamical systems—From a mathematical model to a theory of dreams

K. VISWANATH
Department of Mathematics and Statistics, University of Hyderabad, Hyderabad 500 046, India.

Received on May 7, 1998; Revised on May 25, 1998.

Abstract

This paper gives an overview of the mathematical aspects of artificial neural networks theory presenting it broadly as a problem of constructing dynamical systems with given properties. It is also shown how computer simulations on a mathematical model have suggested a theory of dreams to the two biologists F. Crick and G. Mitchison.

1. Introduction

The purpose of this paper is to present the basic ideas of the subject of artificial neural networks from a purely mathematical point of view. For the sake of highlighting the mathematical structures without distractions we relegate to the background all biological motivations, hardware/software implementations, computer science considerations and application needs. Also we concentrate on the simplest possisble mathematical formulation. But, it is amazing how even such simple-minded models can lead to nontrivial insights into real phenomena. As an illustration we explain how the Crick–Mitchison theory of dreams arises from simulation studies on a certain neural dynamical system.

2. Neural maps

An *activation* is a function $g\colon \mathbb{R} \to \mathbb{R}$ which is piece-wise continuous and has only Type I discontinuities. For any activation g and for any positive integer n, we define $\tilde{g}\colon \mathbb{R}^n \to \mathbb{R}^n$ by $\tilde{g}\colon (x_1, x_2, \ldots, x_n) = (g(x_1), g(x_2), \ldots, g(x_n))$.

Two standard examples of activations are the signum and the sigmoid functions. The signum function is defined by

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

The sigmoid function is defined by $g(x) = \dfrac{1}{1+e^{-cx}}$ where $c > 0$ for all $x \in \mathbb{R}$.

A map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is called a basic neural map with activation g if f is of the form $\tilde{g} o T$ where $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is affine linear and $\tilde{g}\colon \mathbb{R}^m \to \mathbb{R}^m$. Compositions of basic neural maps are called neural maps. A composition of k basic neural maps is called a neural map of level k. If all the basic neural maps involved in the composition of a neural map f have the same activation g, we say that f is a neural map with activation g.

Thus, a typical neural map $f: I\!R^{n_0} \to I\!R^{n_k}$ of level k and activation g is of the form:

$f = \tilde{g} \circ T_k \circ \tilde{g} \circ T_{k-1} \circ \ldots \circ \tilde{g} \circ T_1$ where $T_1: I\!R^{n_0} \to I\!R^{n_1}, \ldots, T_k; I\!R^{n_{k-1}} \to I\!R^{n_k}$. The integers $n_0, \ldots, n_k$ are called the dimensions of f and $n_1, \ldots, n_{k-1}$ are called the hidden dimensions of f.

If the range of a neural map f is $I\!R$, we call f a neural function. Clearly, if a map $f: I\!R^n \to I\!R^m$ is a neural map then each component is a neural function.

Often neural functions are defined in terms of state spaces on directed graphs. However, directed graphs are only devices which help us to identify the neural map to be used in a given context. This is a detail which need not concern us in the bird's eye-view picture presented here.

It is tempting to trace the lineage of neural functions through Kolmogorov to Hilbert. But this is disputed pedigree (see Hassoun[1] for a discussion). A theorem similar in spirit to Kolmogorov's theorem and indisputably pertinent is Cybenko's theorem[2–4] which assures us of the relative abundance of neural functions.

Cybenko's Theorem: Let σ be the distribution function of any continuous probability measure on $I\!R$ or a Dirac distribution function. Then any continuous function $h: K \to I\!R$ (where $K \subset I\!R^n$ is compact) can be approximated uniformly as closely as necessary by a neural function of level 3 with activation σ.

The main theme in the study of artificial neural networks is the construction of dynamical systems having given properties. We recall some basic definitions below.

Definition: (a) A (discrete) dynamical system is a triple (X, d, f) where (X, d) is a metric space and $f: X \to X$ is not necessarily continuous.

(b) (X, d, f) is a neural dynamical system if $X \subset I\!R^n$, d is a metric on X and $f: X \to X$ is a neural map.

Suppose (X, d, f) is a dynamical system.

(c) For $x \in X$ the sequence $\{x, f(x), f^2(x) = f(f(x)), \ldots\}$ is called the *orbit* of x.

(d) f is said to be globally stable if every orbit is convergent. In this case we define $f^\infty(x) = \lim_{n\to\infty} f^n(x)$.

(e) $A \subset X$ is an *invariant* set if $f(A) \subset A$.

(f) If A is a closed invariant set the *basin* of A is defined by $bas(A) = \{x: d(f^n(x), A) \to 0\}$.

The central problem in the study of neural networks may be formulated as follows:

Central problem: Given

(i) $X \subset \mathbb{R}^n$ and a metric d on X

(ii) pair-wise disjoint closed subsets $A_1, A_2, \ldots, A_p$ of X, and

(iii) pair-wise disjoint subsets $B_1, B_2, \ldots, B_p$ of X such that $A_i \subset B_i$ for all i,

to find a globally stable neural map $f: X \to X$ such that

(a) $f(A_i) \subset A_i$ for all i, and

(b) $B_i \subset bas(A_i)$ for all i.

We show below that two standard problems in the study of neural networks can be seen to be special cases of the Central problem stated above.

Example: Multiple layer perceptron (MLP) problem

This is essentially an interpolation problem, but there are subtle and important questions of 'overfitting' and 'underfitting' which are hard to model mathematically and because of which it is more of an art than science to find the appropriate interpolation function. In the barest mathematical details the problem is as follows.

Suppose $C_1, \ldots, C_p$ are pair-wise disjoint finite subsets of $\mathbb{R}^n$. Let $a_1, \ldots, a_p \in \mathbb{R}^m$. It is required to construct a neural map $F: \mathbb{R}^n \to \mathbb{R}^m$ such that $F(x) = a_i$ for all $x \in C_i$, $i = 1, 2, \ldots, p$. More precisely, given an activation g we need to choose appropriate hidden dimensions $n_1, \ldots, n_{k-1}$, and basic neural maps

$$F_1: \mathbb{R}^n \to \mathbb{R}^{n_1},$$

$$F_2: \mathbb{R}^{n_1} \to \mathbb{R}^{n_2},$$

$$\ldots$$

$$F_k: \mathbb{R}^{n_{k-1}} \to \mathbb{R}^m$$

with common activation g such that, for $F = F_k \circ \ldots \circ F_k$, it is true that $F|c_i = a_i$, for all $i = 1, \ldots, p$.

This is equivalent to a special case of the Central problem if

$$N = n_0 + n_1 + \ldots + n_k$$

$$X = \mathbb{R}^N \cong \mathbb{R}^{n_0} \oplus \cdots \oplus \mathbb{R}^{n_k}$$

$$d = \text{the Euclidean metric on } \mathbb{R}^N$$

$$B_i = \left\{ x = (x_0, \cdots, x_k) \in \mathbb{R}^N \mid x_0 \in C_i \right\}$$

$$A_i = \left\{ x = (x_0, \cdots x_k) \in \mathbb{R}^N \mid x_0 \in C_i, x_k = a_i \right\}.$$

If $f: \mathbb{R}^N \to \mathbb{R}^N$ is given by

$$f(x_0, x_1, \cdots, x_k) = (x_0, F_1(x_0), F_2(x_1), \cdots, F_k(x_{k-1}))$$

then f is a solution to the problem. Note that here $f^\infty = f^k$.

Example: The Hopfield problem

Let X be the 'bipolar' space $\{-1,1\}^N$ of dimension N and let $d(x, y) = \#\{i: x_i \neq y_i\}$ be the Hamming distance. For x in X and $r > 0$ let $B(x, r)$ denote the closed ball of radius r centred at x. Choose and fix $a_1,\cdots, a_p$ in X and $r_1,\cdots, r_p$ such that the closed balls $B(a_i, r_i)$ are pair-wise disjoint. Choose the activation to be the signum function. The problem then is to find a globally stable basic neural map f such that $f(a_i) = a_i$ for all i and $B(a_i, r_i) \subset \text{bas}(a_i)$ for all i. Equivalently, such that $f^\infty(B(a_i, r_i)) = \{a_i\}$ for all i. (Note here that $f^\infty = f^{2^N}$.) A variation is to require that $f(B(a_i, r_i)) = \{a_i\}$ for all i. In this case, the Hopfield problem is equivalent to a single-layer Perceptron problem.[5]

The Hopfield problem is a very simple-minded model of the way the brain stores memories. If every neuron in the brain can be either 'on' or 'off' then the state space of the set of all neurons can be modelled by X and any memory can be modelled by a particular pattern of 'on's and 'off's, i.e. an element of X. So each a_i above can be considered to be a 'memory'. It is reasonable to think that any stimulus associated to the memory and which gives rise to the exact recall of the memory will be close to that memory in terms of the Hamming distance. So for modelling the activity of the brain by f we require the elements in $B(a_i, r_i)$ to be carried to a_i by the map f. f is called an associative memory because f associates elements of $B(a_i, r_i)$ with a_i.

A graphic metaphor for an associative memory is to visualize a surface on X which has local minima at the memories and for which the valleys associated with the local minima are the basins of the memories.

3. Learning

There is a general method of approaching the Central problem of neural network theory which we now explain. Given (X, d) and the property P which the neural dynamical system is required to satisfy, we first choose and fix the activation, the level of the neural map and the hidden dimensions. Then the only variability left for the dynamical system is in the real coefficients of the affine linear maps defining the system. Thus, we have essentially a family of dynamical systems parametrized by elements of some Euclidean space, say W. Each $w \in W$ defines a dynamical system on X and w is called the weight vector of that system.

Let W_0 be the set of all dynamical systems satisfying the property P. The problem now reduces to locating an element of W_0, assuming W_0 is nonempty. This is sought to be done by constructing a map $\Lambda : W \to W$ (not necessarily a neural map) which is globally stable and is such that for any $w \in W$, $\lim_{n \to \infty} d(\Lambda^n(w), W_0) = 0$. Then, $\Lambda^\infty(w) \in \overline{W_0}$. (Note that this is also a case of the Central problem.) However, such an ideal Λ is only rarely met in practice. One generally writes down a plausible map Λ or a finite or infinite sequence of Λ^n's with the hope that starting from some carefully chosen w_0 and defining $w_n = \Lambda_n(w_{n-1})$, w_n is close to W_0 for some sufficiently large n. Such an algorithm is called a learning algorithm or a learning rule. The standard learning rule for the multiple layer Perceptron problem is called 'back-propagation' and the standard learning rule for the Hopfield problem is called 'Hebb's rule'.

4. Generalizations

Keeping the Central problem stated above in focus one can consider several possible generalizations.

(i) Instead of affine maps in the formation of neural maps one can consider quadratic or even general polynomial maps.[6]

(ii) In the place of discrete dynamical systems one can consider continuous/stochastic dynamical systems.[7–9]

(iii) The condition of global stability may be generalized by considering systems with periodic or even chaotic attractors.[10]

(iv) The real line $\mathbb{R}$ may be replaced by a topological ring R.[11]

(v) Instead of deterministic learning rules as above one can consider fuzzy or stochastic algorithms.[12, 13]

A rich theory has grown along all the generalizations suggested above in the last few years. Some of these results are really breathtaking. For example, the EM-algorithm of statistics can be identified, under certain conditions, with a suitable learning rule defined in terms of geodesics in information geometry![14]

5. Hebbian learning and unlearning

We now return to concentrate on the Hopfield problem. Let $X = \{-1, 1\}^N$. Let $\mathcal{S}$ denote the space of $N \times N$ real symmetric matrices (which are identified with linear endomorphisms of $\mathbb{R}^N$). For each $T \in \mathcal{S}$, define $\hat{T}: X \to X$ by $\hat{T} = \hat{\text{sgn}}\, o T$. Each $\hat{T}$ is a basic neural map on X. Note that if $c > 0$ then $\left(\widehat{cT}\right) = \hat{T}$.

Let now a finite set $A = \{a_1, a_2, \cdots, a_p\} \subset X$ be given. We are asked to find $T \in \mathcal{S}$ such that $\hat{T}$ is globally stable and each element of A is a memory (i.e. a fixed point) of $\hat{T}$.

Hebb's learning rule is given by Λ_k, $k = 1, 2, \cdots, p$ where Λ_k: $\mathcal{S} \to \mathcal{S}$, $\Lambda_k(T) = T + a_k a_k^{tr}$ (here, a_k^{tr} denotes the transpose of the column vector a_k). Λ_k is called the rule for learning a_k.

Set $H = \Lambda_p\, o \cdots o\, \Lambda_1(0) = \Sigma_{k=1}^{p} a_k a_k^{tr}$.

H is said to be given by Hebb's rule. It is not difficult to show that if the a_ks are pair-wise orthogonal then each a_k is a memory of $\hat{H}$. In the general case, a_ks need not be fixed points of $\hat{H}$. However, in practice one usually starts with H as above (because, Hebb's rule is motivated by neurobiological considerations) and tries to modify H to get a solution to Hopfield's problem.

It is known that $\hat{H}$ is always globally stable. The trouble with $\hat{H}$ is that apart from the fact that a_ks need not be memories of $\hat{H}$, $\hat{H}$ has, in general, fixed points which are not in A. These are called spurious memories. The problem now is to modify H so that true memories are learnt and spurious memories are unlearnt.

A paper of Hopfield *et al.*[15] (which does not seem to have been given sufficient recognition in literature) suggests a procedure which may not eliminate spurious memories but reduces their ill effects. To be precise, for $f: X \to X$ define for each fixed point a of f, the *accessibility* of a as the ratio $\# bas(a)/2^N$.

Consider now the 'unlearning' procedure given below, which is a special case of the classical method of stochastic approximation.

1. Let $H_0 = H$.

2. Given H_n, choose $x_n \in X$ at random (with equal probabilities) and $\varepsilon_n > 0$ 'small'. Let $H_{n+1} = H_n - \varepsilon_n \left(\hat{H}_n^{\infty}(x_n)\right)\left(\hat{H}_n^{\infty}(x_n)\right)^{tr}$.

Then for large n, $\hat{H}_n$ gives low accessibility to spurious memories and equalizes the accessibilities of the true memories.

In terms of the graphic metaphor suggested earlier according to which we visualize the memories as the local minima of a surface defined on X, the unlearning procedure can be understood in the following way.[16] Imagine the true memories to have deeper valleys as compared to the spurious ones. Throw a marble at random on the surface and let it run into a valley. Put a bucketful of sand in that valley. This has the effect of making the valley less deep. Keep throwing marbles as above. Then after some time, the valleys corresponding to spurious memories get filled up, because they are shallower, leaving only the true memory valleys. (Actually, the true valleys tend to become equal in size also, according to mathematical procedure. But you can push a metaphor only so far).

6. The Crick–Mitchison theory of dreams

The contents of the last section may be summarized as follows:

The brain's problems of storing and recalling memories is modelled, however, crudely and inadequately, by Hopfield's problem. Since we expect the brain to function according to Hebb's rule, we initialize the solution to Hopfield's problem at $H_0 = H$. Simulations suggest that 'unlearning' leads to an improved solution.

This viewpoint suggested a natural (and at the same time bold) conjecture to Crick and Mitchison:[17, 18] *Perhaps the brain too needs to unlearn? Is that what happens in dreams?*

To explain the basic idea of their theory, we first present some biological observations. All quotations are from Crick and Mitchison.[18]

REM (rapid eye movement) sleep has been found to be strongly associated with dreaming in human subjects*. REM sleep (and by inference, dreaming) has been observed not only in-new-born human babies but also in "almost all mammals and in most birds". So REM sleep probably has some "important function and this function is biological in nature and not specifi-

*Dreams can apparently occur in non-REM sleep also. So readers having their own pet theories of dreams need not be overly disconcerted by the Crick–Mitchison theory.

cally human". Further, "it has been known for many years that during REM sleep a series of impulses, called PGO (ponto-geniculo-occipital) waves, appear in the brain".

Considering everything together Crick and Mitchison suggest that perhaps the PGO waves implement the randomization mechanism of 'unlearning' and that the resultant brain activity is dreaming. This will have the effect of minimizing spurious memories and equalizing the accessibility of true memories, thus making the brain more efficient in storing memories. Calling a spurious memory a 'fantasy' and a true memory with a disproportionately large basin an 'obsession', they sumarize their theory with the slogan "We dream to reduce fantasy and obsession''.

In the references cited they examine the case for the plausibility of their theory. But irrespective of that, the point we we wish to emphasize is that the "idea did not come from an explicit consideration of REM sleep and dreams but from *theoretical studies* on the way large groups of neurons might interact together" (emphasis added).[18] Thus, the theory is a beautiful example of the use of mathematical modelling and computer simulation at its best.

7. Conclusion

We have outlined the theory of artificial neural networks theory from a mathematical view point and have also explained briefly an interesting application to the biology of sleep. It is hoped that this will be of help to the mathematically oriented readers to get an idea of what the study of neural systems is all about and why it is so exciting.

Acknowledgements

The author thanks the University Grants Commission (UGC) for its Special Assistance Programme under the Departmental Research Support Scheme. The paper is based on a talk given by the author on January 1, 1998, at the Conference on Recent Advances in Statistics and Probability organized by the Indian Statistical Institute and The Bernoulli Society for Mathematical Statistics and Probability during December 29, 1997–January 1, 1998 at Calcutta, India.

References

1. HASSOUN, M. H. — *Fundamentals of artificial neural networks*, MIT Press, 1995.

2. CYBENKO, G. — *Continuous valued neural networks with two hidden layers are sufficient*, Technical Report, Department of Computer Science, Tufts University, Medford, MA, 1988.

3. CYBENKO, G. — Approximation by superpositions of a sigmoidal function, *Math. Control, Signals, Systems*, 1989, **2**, 303–314.

4. VISWANATH, K. — *Mathematical elements of artificial neural networks*, Technical Report No. 5/97, Stat-Math Unit, Indian Statistical Institute, Calcutta, 1997.

5. CHANDRU, V., VIDYASAGAR, M. AND VINAY, V. — Tractable theories for the synthesis of neural networks. In *Interfaces of computer science and operations research*, R. S. Barr, R. V. Helgason and J. E. Kennington (eds), Kluwer, 1997.

6. KAMP, Y. AND HASLER, M. *Recursive neural networks for associative memory*, Wiley, 1990.

7. COHEN, M. A. The construction of arbitrary stable dynamics in nonlinear neural networks, *Neural Networks*, 1992, **5**, 83–103.

8. GROSSBERG, S. Nonlinear neural networks: Principles, mechanisms and architectures, *Neural Networks*, 1988, **1**, 17–61.

9. HIRSCH, M. W. Convergent activation dynamics in continuous time networks, *Neural Networks*, 1989, **2**, 331–349.

10. CHAPEAU-BLONDEAU, F. AND CHAUVET, G. Stable, oscillatory and chaotic regimes in the dynamics of small neural networks with delay, *Neural Networks*, 1992, **5**, 735–743.

11. TEH, H. H. *Neural logic networks: a new class of neural networks*, World Scientific, 1995.

12. CHENG, B. AND TITTERINGTON, D. M. Neural networks: A review from a statistical perspective, *Statist. Sci.*, 1994, **9**, 2–54.

13. ROJAS, R. *Neural networks: a systematic introduction*, Springer, 1996.

14. AMARI, S. Information geometry of EM and em algorithms for neural networks, *Neural Networks*, 1995, **8**, 1379–1408.

15. HOPFIELD, J. J., FENSTEIN, D. I. AND PALMER, R. G. 'Unlearning' has a stabilizing effect in collective memories, *Nature*, 1983, **304**, 158–159.

16. BRUNAK, S. AND LAUTRUP, B. *Neural networks: Computers with intuition*, World Scientific, 1990.

17 CRICK, F. AND MITCHISON, G. The function of dream sleep, *Nature*, 1983, **304**, 111–114.

18. CRICK, F. AND MITCHISON, G. REM sleep and neural nets, *J. Mind Behaviour*, 1986, **7**, 229(99)–249(119).