



HOW A PEAK CAN BE SEARCHED FOR IN AN ALMOST EVERYWHERE FLATLAND OF ALTITUDE ZERO? — TINY FLAT ISLAND IN HUGE LAKE

Akira Imada

Brest State Technical University
Moskowskaja 267, Brest 224017 Republic of Belarus
akira@bstu.by, <http://neuro.bstu.by/ai/akira.html>

Abstract: *We are exploring a weight configuration space searching for solutions to make our neural network with spiking neurons do some tasks. For the task of simulating an associative memory model, we have already known one such solution — a weight configuration learned a set of patterns using Hebb's rule, and we guess we have many others which we have not known so far. In searching for such solutions, we observed that the so-called fitness landscape was almost everywhere completely flatland of altitude zero in which the Hebbian weight configuration is the only unique peak, and in addition, the sidewall of the peak is not gradient at all. In such circumstances how could we search for the other peaks? This paper is a call for challenges to the problem.*

Keywords: – *Associative Memory, Spiking Neuron, Evolutionary Computations, Fitness Landscape, Needle in haystack, Random Hill-climbing, Baldwin Effect, Artificial Immune System.*

1. INTRODUCTION

With the goal being a realization of *associative memory* by a neural network with spiking neurons, we explore the weight space of a neural network in which some weight configurations are assumed to give the network a function of associative memory. Hyper-planes defined on those spaces are sometimes called *fitness landscapes* when we fictitiously plot a *measure of goodness*, or equivalently, a *fitness* value on all the possible points of configuration assuming altitude of the hyper-plane constructs a landscape, and hence the location of peaks implies the solution of our problem. In our experiment of associative memory, when we were exploring the fitness landscape to try to find those peaks exhaustively, we noticed that the landscape was a very unusual one. That is, the landscape is almost everywhere a flat-land of fitness zero and the shape of the peak whose location is already known is more like a mesa than a peak. The top is not a pin-point due to a synaptic plasticity of the neural network and the sidewall is very steep. Therefore, evolutionary computations which usually recombine points on the hyper-plane as candidate solutions selecting those points which perform better than others, would not work in this fitness landscape of almost everywhere flat-land of fitness zero. This reminds us a classical but a seminal experiment by Hinton & Nowlan [1] which was proposed to find *a peak like a needle in*

haystack.

In short, assuming that we have many peaks in a huge landscape of almost everywhere completely flat-land in which only a few of the peaks are already known, our goal is to find a computational method that has a capability to search for those unknown peaks by employing an information of already known peaks.

In the following three sections, we describe *Associative Memory, Fitness Landscape, and Hinton & Nowlan's experiment* more in detail. Then we propose a test-function and some results of exploring it.

2. ASSOCIATIVE MEMORY

How does anyone stop thinking of something? Accidentally. Accidental thoughts. All thoughts are accidental. — from "Key to Rebecca" by Ken Follett.

Associative memory is a memory system in which we can store information and recall it later from its partial and/or imperfect stimuli. An information is stored as a number of stable states with a domain of attraction around each of the stable states. If the system starts with any stimulus within the domain it will converge to the attractor following a trajectory, hopefully a short one. This models human memory in the sense that, e.g., we can recognize our friend's face even without meeting for a long time, or we can recall a song immediately after listening to a very beginning

part of the song. Hopfield [2] proposed a fully connected neural network model of associative memory in which a set of patterns is stored distributedly among neurons as attractors. Since then the model had been fairly well studied for more than a decade, and we now know it is not so practical, partly due to its small storage capacity, and we study another model using spiking neurons instead of the McCulloch-Pitts [3] neurons like in the Hopfield model, with the goal being to overcome those problems and, more importantly, to look for more biologically plausible models of human memory.

Some regions in our brain such as *neocortex* or *hippocampus* are said to be made up of two categories of neurons, that is, *pyramidal cells* and *interneurons*. Typically, the pyramidal cells communicate with each other via *excitatory synapses* (positive influences), while interneurons send signals to pyramidal cells via *inhibitory synapses* (negative influences). As Wilson [3] wrote in his book, Marr [5] was one of the first to propose this *hippocampal model* involving both recurrent excitation via Hebbian [6] synapses and inhibition. In his book, Wilson [4] wrote that a single neuron which emits spike train when it receives an external stimulus $P(t)$ could be modeled by

$$\frac{d(R(t))}{dt} = \frac{1}{\tau}(-R(t) + S(P(t))) \quad (1)$$

where Wilson [4] proposed to employ, among many alternatives, Naka-Rushton [7] function:

$$S(P) = \begin{cases} M \cdot P^n / (\sigma^n + P^n) & \text{if } P \geq 0 \\ 0 & \text{if } P < 0 \end{cases} \quad (2)$$

M and σ are called *saturation* and *semi-saturation constant*, respectively, and n is an integer parameter for its graph to fit a phenomenon. Here we assume N pyramidal cells and implicit number of interneurons. We simulate these pyramidal cells by spiking neurons which interact with each other using electric current via plastic synapses. Pyramidal cells are also interacted by interneurons by global inhibition. To be more specific, stimuli to one pyramidal cell are given from all the other pyramidal cells via synaptic strength, as well as interneuron cells whose number is reduced to only one here for the sake of simplicity.

The synaptic strength from pyramidal cell j to i is denoted as w_{ij} and all the inhibitory synapses from interneuron are assumed to have a value g . Then stimulus to the i -th pyramidal cell P_i is described as

$$P_i = \left(\sum_{j=1}^N (\omega_{ij} \cdot R_j - g \cdot G)^2 \right)_+ \quad (3)$$

where $(\cdot)_+$ means that we use the value if and only if inside the parentheses is positive and zero otherwise. Following Wilson [4], we experimented with $\sigma = 10$, $M = 100$, and $n=2$ in Eq.(2).

Thus, our equation of spiking ratio of the i -th pyramidal cell R_i with the spiking ratio of the interneuron G is given as

$$\tau_R \frac{dR_i}{dt} = -R_i + \frac{100 \left(\sum_{j=1}^N \omega_{ij} R_j - 0.1G \right)_+^2}{100 + \left(\sum_{j=1}^N \omega_{ij} R_j - 0.1G \right)_+^2} \quad (4)$$

where, τ_R is set to 10. Note that w_{ii} ($i = 1, \dots, N$) should be set to all zero.

In order to encode N -bit binary patterns using N spiking neurons, we use *firing-rate* of a neuron within certain time window which expresses binary number according to whether the rate exceeds a threshold or not.

In what he calls *CA3 network* in his book, Wilson [4] employed 256 pyramidal cells so that these cells represent a pattern constructed by 16×16 array of pixels. The network also incorporates one interneuron cell to provide pyramidal cells a feedback inhibition. The task of the network is to recognize four given patterns from its noisy input. Each of the four patterns is represented by 32 active cells plus other 224 quiet cells. Network has learned to recognize these four patterns by modifying the synapses according to the following what might be called Hebb's [6] rule.

$$w_{ij} = k \cdot \text{sgn}(R_i - 0.5M) \cdot \text{sgn}(R_j - 0.5M) \quad (5)$$

where k is set to 0.016, M is a saturation level in Eq.(1), and $\text{sgn}(x)$ is equal to 1 if $x > 0$ and 0 otherwise. The equation is called Hebb's rule in the sense that w_{ij} will be modified if and only if both the neuron i and j should be activated. Also note that Eq.(5) is applied only if the previous value of w_{ij} is 0, otherwise, w_{ij} will remain intact.

A noisy input of a pattern is constructed by randomly picking up about one-third of the active cells of the selected pattern with adding them other 20 quiet cells, also chosen at random, after turning them active. Then one of these four patterns is given to the network, that is, network starts the dynamics with the pattern as the initial configuration of its neurons' state. Network updates the state according to Eq.(4). The dynamics is observed during a total of 100 ms (assuming step of dt of dr/dt to be 1 ms), with the

noisy input being continued to be fed for the first 20 ms.

3. FITNESS LANDSCAPE

The hill on the south side of the town sloped steeply to the river, here. On the west there was a long ridge which fell gently to the plain. — from "Pillar of the Earth" by Ken Follett.

The concept of the fitness landscape was first introduced by Wright [8] to study biological evolutionary processes. Since then, this concept has been used not only in evolutionary biology but also in chemistry, physics, computer science and so on.

In chemistry, for example, a molecule can be represented as a string of N letters with each letter being chosen from an alphabet of size k (see Macken et al. [10]). Twenty amino acids ($k = 20$) for proteins or four nucleotides ($k = 4$) for nucleic acids can be considered as examples of the alphabet. The k^N possible combinations of the letters construct a configuration space of the string. Then, for example, the free energy of RNA folding into secondary structures (see Fontana [11]) or the ability of peptides to bind to a particular substrate to catalyze a specific reaction (see Maynard Smith [12]) is assigned as a fitness value to each configuration.

In physics, the Hamiltonian energy of Ising spins defines a fitness landscape on the configuration space of N spins, where each spin takes the value either 1 or -1 ($k=2$). Bray and Moore [9] argued about the number and distribution of meta-stable states (local optima) of the Hamiltonian energies.

To explore these fitness landscapes, we need a rule by which a point in the space moves to one of its neighbors. Then, consecutive movements of a point to the neighbors form a *walk* on the landscape. Macken et al. [10] used *random point mutation* that changes a single letter in the string to specify neighbors of the string. Then, by sampling points along an "*evolutionary walk*" in which point moves to the *firstly found fitter* neighbor, they studied the statistical properties of the landscape defined by the chemical affinity of antibody for antigen in immune response. Weinberger [11] used two different walks: "*gradient walk*" in which the walker steps to the *best* of its neighbors and "*random adaptive walk*" in which the next step is chosen *at random* from the set of better neighbors, to investigate the Kauffman's NK landscape [13] which is a model formulated in more general form.

We extend the concept of the discrete fitness landscape to a continuous one. Namely, a capability of a fully-connected neural network to store a set of bipolar patterns (each bit is either 1 or -1) as associative memory assigns fitness on the real-valued

synaptic weight configuration space ($k = \infty$). A walker moves to its neighboring point determined by Gaussian random mutation.

4. NEEDLE IN HAYSTACK

One day I'd cooked soba. Great! I tried teuchi udon the next day and it was also great. It is really easy to make such delicious dishes just from flour or buckwheat flour. The main thing is sauce, however. It is either tasty or not. Nothing in between. — Edward Venskovich (Personal Communication).

The problem Hinton & Nowlan [1] proposed is to search for only one configuration of 20 bits of one and zero, that is, the search space is made up of 2^{20} points all of which except for one point are assigned fitness zero. Only exactly one point, for example, (1111111110000000000) is assigned fitness one. That is why this is called search for a needle in a haystack. See Fig. 1 below.

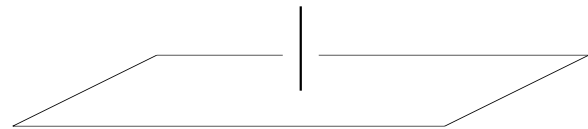


Fig. 1 A fictitious sketch of fitness landscape of a needle in a haystack. The haystack here is drawn as a two-dimensional flat plane of fitness zero.

It seems impossible to solve this if we use a simple genetic algorithm, since usually it recombines two genotypes whose phenotypes are a little better than others, and in our circumstance almost all genotypes perform equally badly. Any hill-climbing would not seem to work. Hinton & Nowlan [1], however, exploited *lifetime learning* of each individual. That is, chromosome is made up of genes of which about 25% are "1", 25% are "0", and the rest of the 50% are "?". Within one generation all the "?" positions are assigned one or zero at random and fitness is evaluated, which is called *lifetime learning* of each individual. Each individual repeats the learning 1000 times in its lifetime. If it reaches the point of fitness one at the n -th trial, then the *degree to which learning succeeded* is calculated as:

$$1 + 19 \cdot (1000 - n) / 1000. \quad (6)$$

Hinton & Nowlan's model is a sort of *gedanken-experiment* to study how the *lifetime learning* affects an evolution, that is, the *Baldwin effect*. The location of the unique solution (whose fitness is one, while all others' are zero) is assumed to be known before a run, though it is not of the case in real world problems.

5. TINY FLAT ISLAND IN HUGE LAKE

It's Spring Cleaning time again ... for my brain. Below, a cluster of random follow-ups, postscripts and observations to tide you over until the spring weather actually turns springlike. — David Pogue from New York Times, 10 Apr. 2003.

Hinton & Nowlan's experiment is valid under an assumption, as they wrote, that phenotype can recognize when it has achieved the perfect fitness during its lifetime learning. This usually does not hold when it is applied to solve our real world problems. If the phenotype recognizes that it reaches the solution, all we need is to check its genotype to know the parameter configuration which give the phenotype to achieve its goal and no need for the computation to proceed. Without the assumption, we could not explore this specific fitness landscape searching for the goal. Hence, here, we call for challenges to this type of problems and we proposed a test function for the purpose. This is essentially identical to the Hinton & Nowlan's fitness landscape, but more flexible to control its complexity.

Test-function (*Tiny Flat Island in Huge Lake*)

Assuming a n -dimensional hypercube all of whose coordinate x_i ($i=1, \dots, n$) lies $[-1, 1]$, find an algorithm to locate a point in the region A whose coordinates all lie $[0, a]$ ($a \leq 1$).

The target of the search is a hypercube in the n -dimensional Euclidean space, and the size of the hypercube and complexity of the search are controlled by changing a and n , respectively. When we see the search from the fitness landscape point of view, this is like a search for a *tiny flat-land of altitude one in a huge flat-land of altitude zero*. When $n=20$ and $a=1$ it is equivalent to Hinton & Nowlan's needle in haystack, and if necessary, we can make the needle tinier by decreasing the value of a .

Or, if we have, for example, *multi-agent system* or *artificial immune system* in mind to make them this kind of search, we might modify this test-function as giving the agents an information of, say, a *hyper-cube whose coordinates all lie $[-0.1, 0.1]$ and then asking them to search for two regions each of whose coordinates all lie $[-0.9, -1]$ and $[0.9, 1]$ respectively.*

6. EXPERIMENTS

Needle in Haystack. We were forced to modify the Hinton & Nowlan's experiment because when individuals are created at random, they usually did not achieve fitness one during 1000 times of lifetime learning. Hence, we create individuals one by one at random and each time we make it learn 1000 times,

and if it reaches the fitness one we put it in the population of the first generation, and this is repeated until those individuals fill the population. In other words, a run starts with a population of individuals who are within 1000 steps from the needle. One example of run shows we have to try 118,499 times randomly to obtain such a population of 100 individuals.

The result mentioned above is only within the first generation. If we proceed the evolution under the condition that individual knows whether it reaches the fitness one while we observer cannot know what is going on to genes whose allele is "?", we expect the number of "?" genes decreases as the evolution proceeds, and eventually we obtain the target chromosome which is made up genes "1" and "0" alone. See Fig. 2 bellow.

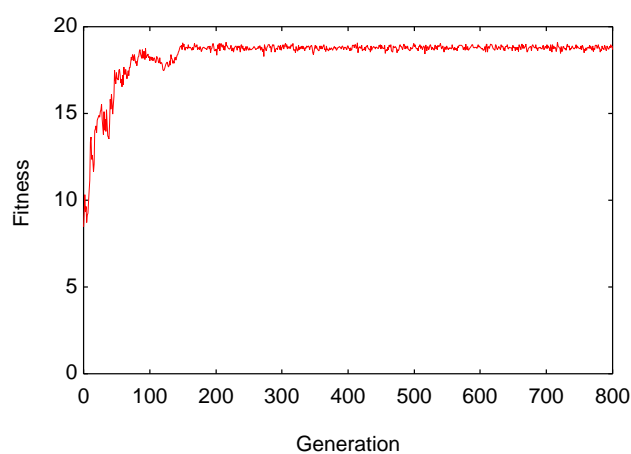


Fig. 2 An evolution of search for a needle in a haystack from the experiment of Hinton & Nowlan [1].

Tiny Flat Island in Huge Lake. Thus, we now know that although the Hinton & Nowlan's experiment is elegant method to see lifetime learning enhances the genetic search, that is, the Baldwin effect works in our evolutionary computations, while in reality we have not found so far an algorithm to solve this type of a needle hidden in a haystack. Then we proposed a test-function in the previous Section, and here we show results of applying both a *simple random search*, which is not even a random hill-climbing, and the *lifetime learning*, the one proposed by Hinton & Nowlan but only within one generation, to the test function.

Simple random search. We set $a = 1$ and study if a randomly created chromosome with length n will be in the domain A or not (See Eq.(4), that is to say, a random search looking for points in A . As n becomes large, search becomes difficult and eventually when $n = 20$ we cannot find any such point within a reasonable time, say, in 24 hours. No wonder Hinton & Nowlan adopted the chromosome of length 20! Fig. 4 shows the result of how many chromosomes were on A out of 10,000 randomly created ones.

Lifetime learning. Here, we also create a chromosome at random one by one, but we study if each of them reaches the domain A after 1000 times of learning. In Fig. 4 we plot how many chromosomes we have to create until we find the individual who reaches the goal within 1000 times of learning. We see the results are a little better than the above mentioned random search.

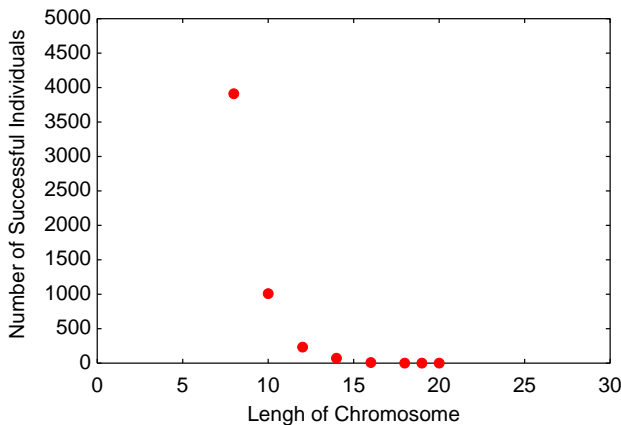


Fig. 3 Number of individuals who happen to be in a point in the target region A out of randomly created 10,000 individuals.

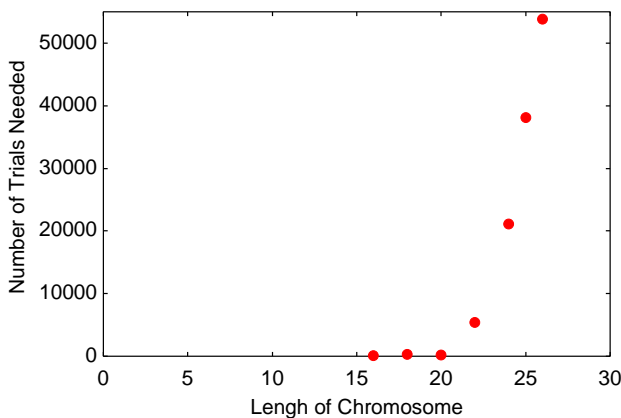


Fig. 4 Average Number of random individuals needed to find a one who succeeded in reaching the target via 1000 times of lifetime learning.

7. SUMMARY

We came across a very difficult problem while we made experiments with our associative memory system using a neural network with spiking neurons. We have already known one weight configuration which give the network a function of associative memory — Hebbian weights. We conjecture we have other such weight configurations. Then we explored a fitness landscape defined on weight space in which the Hebbian weights locates, and we observed that the Hebbian peak in the fitness landscape is like a tiny

flat island in a huge lake. The lake is too huge to get a bird's eye view of the whole lake and we have never been able to see other islands. Hence we have proposed in this paper a test-function which is a simplified version of our problem and we can easily control the difficulty of the problem with the structure being essentially the same. This reminds us of the Hinton & Nowlan's classical experiment of searching for a needle in a haystack in which individual's lifetime learning was employed to learn if the Baldwin effect works in our computational evolution. We have found that the lifetime learning also somehow works in the proposed test-function if we compare it to a simple random search. However, we still doubt more or less if we can apply the Baldwin effect as it is to a real world problem. So, this paper is a call for challenge proposals of the methods to solve our test function.

In short, not so short though, in a huge landscape of almost everywhere completely flat-land, assuming we have many peaks only a few of which we know, our goal is to find a computational method that searches for the unknown peaks, by employing an information of those already known peaks.

As a candidate of these methods, we now thinking of *anomaly detection* by *artificial immune system*.

REFERENCES

- [1] G. E. Hinton and S. J. Nowlan. (1987) How Learning can Guide Evolution. *Complex Systems*, 1, pp. 495–502.
- [2] J. J. Hopfield (1982) Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences, USA*, 79, pp. 2554–2558.
- [3] W. S. McCulloch and W. Pitts (1943) A Logical Calculus of Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5, pp. 115–113.
- [4] H. R. Wilson (1999) Spikes, Decisions, and Actions: The Dynamical Foundations of Neuroscience. *Oxford University Press*.
- [5] D. Marr (1971) A Theory for Archicortex. *Philosophical Transactions of the Royal Society of London*, B262, pp. 23–81
- [6] D. O. Hebb (1949) *The Organization of Behavior*. Wiley.
- [7] K. I. Naka and W. A. Rushton (1966) S-potentials from Colour Units in the Retina of Fish. *Journal of Physiology*, 185, pp. 584–599.
- [8] S. Wright (1932) The Roles of Mutation, Inbreeding, Crossbreeding and Selection in Evolution. *Proceedings of the 6th International Congress of Genetics*, pp. 356–366.
- [9] A. Bray, and M. Moore (1980) Metastable States in Spin Glasses. *Journal of Physics C: Solid State Physics*, 13, pp. L469–L476.

- [10] C. Macken, P. Hagan, and A. Perelson (1991) Evolutionary Walks on Rugged Landscapes. *Society for Industrial and Applied Mathematics: Journal of Applied Mathematics* 51, pp. 799–827.
- [11] E. D. Weinberger (1990) Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference. *Biological Cybernetics* 63, pp. 325–336.
- [12] W. Fontana, and P. Schuster (1987) A Computer Model of Evolutionary Optimization. *Biophysical Chemistry* 26, pp. 123–147.
- [13] J. Maynard Smith (1970) Natural Selection and the Concept of a Protein Space. *Nature* 225. pp. 563–564.
- [14] S. A. Kauffman, and S. Levin (1987) Towards a General Theory of Adaptive Walks on Rugged Landscapes. *Journal of Theoretical Biology* 128, pp. 11–45.



Akira Imada is a Professor of Computer Science in the Brest State Technical University, Republic of Belarus, with primary research interests in Neural Networks with Spiking Neurons, Evolutionary Computations, and recently, Artificial Immune System. He has been working in the University with the Department of Intelligent Information Technology since March 2002 after working in Turkey with Anatolia University, Eskisehir for one year and Dokuz Eylul University, Izmir for two years. He completed his Ph.D. in March 1999 at the Nara Institute of Science and Technology, Japan in the area of Artificial Intelligence with the topic of Associative Memory by Evolving Neural Networks. He has authored so far around 50 papers in workshops, conferences and journals in the areas of Neural Networks, Evolutionary Computations and so on.