In: Computational Neuroscience: Trends in Research, 1997. Proceedings of the Annual Computational Neuroscience Conference. (J. M. Bower, Ed.) New York:Plenum Press, pp. 705-709, 1997.

PREDICTING NOVEL PATHS TO GOALS BY A SIMPLE, BIOLOGICALLY INSPIRED NEURAL NETWORK

William B. Levy and Xiangbao Wu

Department of Neurological Surgery
University of Virginia Health Sciences Center
Charlottesville, Virginia 22908
wbl@virginia.edu
xw3f@virginia.edu

1. ABSTRACT

Although recurrent networks can be used as content addressable memories, they can also be used as sequence prediction systems. Because problem solving can often be viewed as a sequence prediction problem, we hypothesize that such networks can be used as problem solvers. There are many aspects to problem solving. Here we concentrate on a single but important aspect, goal finding without search. Using a highly simplified, clearly prototypical version of this problem, a sparsely connected recurrent network successfully predicts novel paths to reach a goal.

2. INTRODUCTION

Problem solving is often recognized as a sequence completion or sequence prediction problem (e.g., Newell & Simon 1972). In the artificial intelligence approach to problem solving, logical sequences to reach goals are typically investigated goal-by-goal and sequence-by-sequence. This approach becomes rapidly computationally infeasible. Here, we present a neural network model that uses attractors to find goals and even novel paths, without search.

3. THE NETWORK

The network (see Fig. 1a and 1b), inspired by a small piece of hippocampal region CA3, has a sparse (10%) recurrent connectivity, c_{ij}, between its 512 excitatory neurons

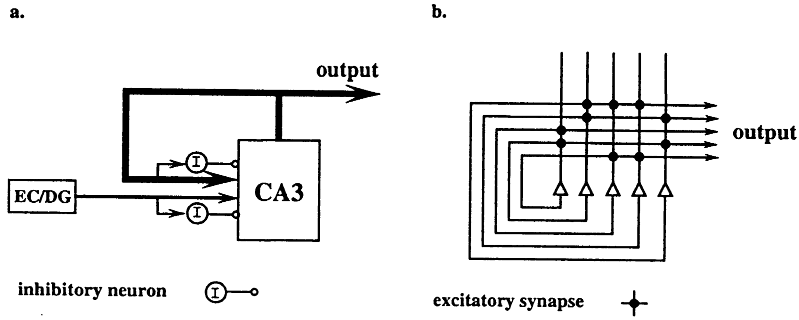

Figure 1. The Model. a) Simplified Hippocampal model. In the model the input layer is a combination of the entorhinal cortex and dentate gyrus. Accompanying this feedforward excitation is a proportional feedforward inhibition. The strong excitation of the network results from the recurrent connections, which is also accompanied by a feedback inhibition. The output of the network is the state of the excitatory CA3 cells themselves, and this is decoded by a simple cosine comparison. b) The recurrent excitatory synapses are sparse and randomly placed.

and a paucity (two) inhibitory neurons that maintain control over net activity levels in a crude, imprecise way (Minai & Levy 1993) via fixed parameters K_I and K_R. The simultaneously updated McCulloch-Pitts neurons transmit $z_i(t) \in \{0,1\}$ every increment of time, t through synaptic weights w_{ij}.

The excitation y_j of CA3 neuron j is given by:

$$y_j(t) = \frac{\sum_i w_{ij} c_{ij} z_i(t-1)}{\sum_i w_{ij} c_{ij} z_i(t-1) + K_I \sum_i x_i(t) + K_R \sum_i z_i(t-1)}$$

and its output of neuron j is

$$z_j(t) = \begin{cases} 1 & \text{if } y_j(t) \geq \theta \text{ or if } x_j(t) = 1; \\ 0 & \text{otherwise} \end{cases}$$

where each neuron j receives a single external input $x_j(t) \in \{0,1\}$.

Synaptic modification proceeds adaptively via a self-supervised, local modified Hebbian process of the form $w_{ij}(t+1) = w_{ij}(t) + 0.01 \cdot z_j(t)\,(z_i(t-1) - w_{ij}(t))$. The time delay, similar to a proposal of Amari (1972), is qualitatively consistent with physiological observations (Levy & Steward 1979, 1983).

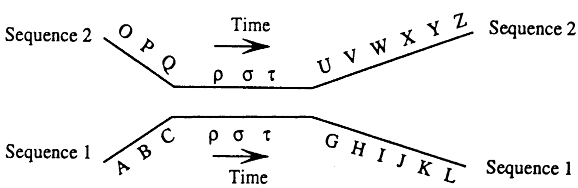

Figure 2. Pictorial representation of the enviroment. There are two sequences of 12 input patterns to learn. Each sequence contains three orthogonal segments. The two sequences share a subsequence of three patterns.

The cosine function for comparing two vectors is used to understand the output of the network during testing. Specifically, after learning, the network states that evolve in response to the test input are compared with the states produced by a complete sequence of input patterns. By definition, the largest value at each time step is the decoded network state at that moment. Although other techniques (e.g., maximum likelihood) would produce superior decoding, such methods would obscure our object of our study, i.e., what the CA3-like network is actually doing.

4. RESULTS

4.1. Sequence Completion under Ambiguous Conditions (The Disambiguation Problem)

Here we look at an analog of a useful skill in problem solving — the ability to combine separately learned inferences. In this problem the network learns two deterministic sequences of 12 patterns (see Fig. 2). Each sequence contains three orthogonal segments. Within each segment, the successive external inputs share 7 of 8 active neurons. The two sequences share a common subsequence [ρ,σ,τ] that is three patterns long. We initially studied this problem of shared subsequences (Minai et al 1994 and again in Levy et al. 1995, Wu et al. 1996) to prove that the network, as a model of the hippocampus, can learn and can use context. Here, we start by studying the disambiguation problem. As before, the network successfully disambiguates two related sequences based only on context. After learning, when the network is given pattern O, it produces a sequence culminating in pattern Z. Similarly (see Fig. 3a), giving the network pattern A produces a sequence culminating in pattern L. Once again, we emphasize that there is no explicit spectrum of delays (e.g. Kleinfeld 1986) nor is there any microscopic time spanning properties greater than one step so that the solution derives from the dynamic code created by the sparsely interconnected network itself (Levy 1989).

4.2. Goal Finding Via Novel Sequence Generation

In addition to solving sequence prediction under ambiguous conditions, we now want to show goal finding by the network. This is a compositional problem involving a recomposition of subsequences. Therefore, in addition to giving the network an initial starting point as an input, we also give it a partial description of a goal that requires a novel path. Thus, the input is initially pattern A as before, but now two neurons of pattern Z are continuously activated — as if the model has some partial idea about what the goal looks like. In this case the network hypothesizes the path [A,B,C,ρ,σ,τ,U,V,W,X,Y,Z] which reaches the appropriate goal pattern. Figure 3b helps show how the network succeeds. Thus, the model decomposed the two learned sequences and recomposed the pieces as appropriate to solve the problem.

To avoid unrepresentative results, here we present only robust results where robust is defined as replicable in 4 out of 5 randomly constructed networks. For each network there are four test cases, i.e. given pattern A, given pattern O, given pattern A and two neurons of pattern Z, and given pattern O and two neurons of pattern L. Out of the 5 networks we tried, the problems were correctly solved between 80–90% of the time with an average of 85%.

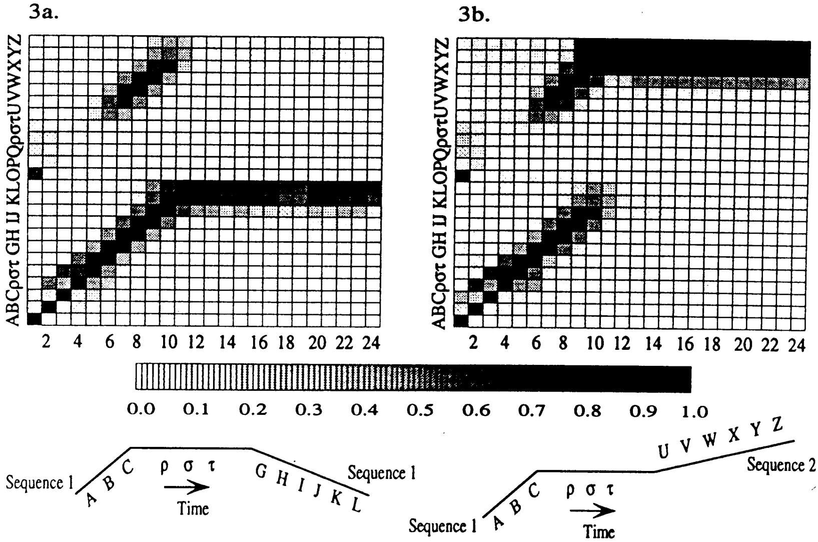

Figure 3. Sequence prediction after learning the two partially overlapping sequences illustrated in Fig. 2. These similarity matrices use the cosine function. They compare the network states generated by each full pattern sequence after learning (ordinate) with the network states generated over time during testing (abscissa). Decode the output during testing by finding the darkest square in each column; the letter on the ordinate is the decoded answer. 3a. Sequence completion in the disambiguation problem is made difficult by the shared subsequence [ρστ]. Here we show the similarity values (used for decoding) when the network is transiently given pattern A. Appropriately enough for the learning and for the starting point, the states go to pattern L, a pattern essentially orthogonal to the representation of the other learned goal pattern, Pattern Z. 3b. When the same network is given the same transient input but two neurons of goal Z are also turned on, the network produces a sequence of representations leading to this partially specified goal. To create this path, the network must produce a novel sequence that appropriately combines its knowledge of the two separately learned sequences.

5. DISCUSSION

Because the network solves the context-dependent disambiguation problem, representations of ρ through τ must differ depending on where the network starts its sequence of representations. Even so, there is enough similarity between each set of the ρ-τ states for the network to follow either path depending on the relative strength of the two attractors at the end of each sequence. In addition to the network's immediate representation of τ, the winning attractor is a function of experience and of external biasing.

Hopfield's research (Hopfield 1982, 1984) has inspired a great amount of analytical work on the properties of recurrent networks. As a result, we now know that even networks with a time spanning Hebb rule can form, at the very least, transient attractors (e.g. Sompolinsky & Kanter 1986, Wu & Liljenström 1994, Liljenström & Wu 1995). However, using the dynamics of such networks to solve sequence prediction problems with novel goals seems a new use of such networks.

In sum, the hippocampal model studied here is capable of goal finding, albeit a highly simplified form of this problem.

6. ACKNOWLEDGMENTS

This work was supported by NIH MH48161 and MH00622, by EPRI RP8030-08, and Pittsburgh Supercomputing Center Grant BNS950001 to WBL, and by the Department of Neurosurgery, Dr. John A. Jane, Chairman.

7. REFERENCES

Amari, S.I. (1972) *IEEE Trans. on Computers* **C-21**, 1197–1206.
Hopfield, J. J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2554–2558.
Hopfield, J. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3088–3092.
Kleinfeld, D. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 9469–9473.
Levy, W. B (1989) in *Computational Models of Learning in Simple Neural Systems*, eds. Hawkins, R. D. & Bower, G. H. (Academic Press, New York), pp. 243–305.
Levy, W. B & Steward, O. (1979) *Brain Res.* **175**, 233–245.
Levy, W. B & Steward, O. (1983) *Neurosci.* **8**, 791–797.
Levy, W. B, Wu, X. B., & Baxter, R. A. (1995) *Intl. J. Neural Syst.* **6 (Supp.)**, 71–80.
Liljenström, H. & Wu, X. B. (1995) *Intl. J. Neural Syst.* **6**, 19–29.
Minai, A. A., Barrows, G., & Levy, W. B (1994) *World Congress on Neural Networks* **IV**-176–181
Minai, A. A. & Levy, W. B (1993) In: *Advances in Neural Information Processing Systems*, eds. Giles, C. L., Hanson, S. J., & Cowen, J. D. (Morgan Kaufmann, San Mateo, CA), pp. 556–563
Newell, A. & Simon H. (1972) *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.
Sompolinsky, H. & Kanter, I. (1986) *Phys. Rev. Ltrs.* **57**, 2861–2864.
Wu, X. B., Baxter, R. A., & Levy, W. B (1996) *Biol. Cybern.* **74**, 159–165.
Wu, X. B. & Liljenström, H. (1994) *Network: Comput. Neural Syst.* **5**, 47–60.