Generating smooth context-dependent neural representations *

Simona Doboli
Complex Adaptive Systems Laboratory
ECECS Department
University of Cincinnati
Cincinnati, OH 45221-0030
Email: sdoboli@ececs.uc.edu

Ali A. Minai
Complex Adaptive Systems Laboratory
ECECS Department
University of Cincinnati
Cincinnati, OH 45221-0030
Email: Ali.Minai@uc.edu

Phillip J. Best
Laboratory of Cognitive Psychobiology
Department of Psychology
Miami University
Oxford, OH 45056
Email: BESTPJ@muohio.edu

Abstract

In many cognitive situations, context is specified globally — encompassing the entire duration of an episode, but explicitly identified only at its initiation. Examples of such contexts are social situations, spatial environments, task specifications, etc. In many such cases, a neural system needs to respond differently to the same immediate stimulus depending on the global context. Clearly, the system must be able to switch into different response regimes and remain there while the context remains fixed. Simultaneously, it must remain sensitive to external stimuli, and respond to them in a reliable and informative manner. In this paper, we propose a neurally plausible approach for handling these requirements.

Introduction

The generation and use of context-dependent neural representations is an important issue in many cognitive situations. Most connectionist studies of context have focused on what may be termed "immediate context" — the dependence of system response on recent history in addition to external input. This is important in sequential tasks such as trajectory learning and speech recognition, and is typically mediated by explicit feedback of the system's previous state [11, 2, 3, 6, 5], or by slowly decaying correlation of activity [9]. However, in many situations, the context is more global — encompassing the entire duration of an episode, but explicitly identified only at its initiation. Examples of such contexts are social situations, spatial environments, task specifications, etc. In many such cases, a system needs to respond differently to the same immediate stimulus depending on the global context. Clearly, the system must be able to switch into different response regimes and remain there while the context remains fixed. Simultaneously, it must remain sensitive to external stimuli, and respond to them in a reliable and informative manner. In this paper, we report on an approach that allows a system to sustain a stable response regime, while remaining completely sensitive to external stimuli.

One obvious way to achieve the required behavior is to make the system modular, with one module for each context. The appropriate module can then be enabled by an external context-recognition system. This would, of course, require the external recognizer to sustain the enabling signal for the duration of an episode, suggesting that the recognizer/enabler might be a recurrent attractor network. A further refinement would be to allow the modules to overlap, essentially defining them as subsets chosen from a large pool of neurons. This would allow more efficient use of resources with a small additional risk of error. In this paper, we describe a single attractor network system which integrates the functionality of the recognizer/enabler subsystem and the modular response subsystem. This approach, in addition to being parsimonious, is also biologically plausible

*Proc. 1999 Int. Joint Conf. on Neural Networks

— being inspired by the architecture of the hippocampal formation [8, 4, 10].

Formal Task Description

The task for the system is to produce reliable, structure-preserving, and context-dependent codes for stimulus patterns when the context is indicated only by a transient initial stimulus. This task is formally defined as follows:
The system stimulus consists of binary patterns drawn from two sets:

1. The set of *context patterns*, $\mathcal{C} = \{C^k\}$, $C^k \in \mathcal{I}$,
2. The set of *regular patterns*, $\mathcal{X} = \{X^k\}$, $X^k \in \mathcal{I}$,

where $\mathcal{I}$ is the input space of dimension N_I. The system receives *stimulus sequences* of the form $S^q = s_0^q s_1^q ... s_n^q; \quad q = 1, 2, 3, ..., M$, where $s_0^q \in \mathcal{C}$ and $s_j^q \in \mathcal{X}\ \forall j \geq 1$.
In response to the stimulus sequence, the system produces a *response sequence*, $R^q = r_0^q r_1^q ... r_n^q; \quad q = 1, 2, 3, ..., M$, with a one-to-one correspondence between the stimulus and response patterns. Each r_j^q is a binary pattern of neural activity.
In the ideal case, the response patterns must satisfy the following requirements:

1. Reliability: *In the same context*, the same stimulus pattern must produce the same response regardless of its position in the sequence or trial initial conditions:

$$d(s_j^q, s_i^q) = 0 \implies d(r_j^q, r_i^p) = 0$$

where $d(x, y)$ is a normalized measure of distance between x and y. Since this ideal specification is never met, we measure reliability as follows:

$$R = 1 - \langle d(r_i^q, \rho(i, q)) \rangle$$

where $\rho(i, q)$ is defined as the maximum likelihood response of the system to all stimuli identical to r_i^q, and angular brackets represent averaging over i and q.

2. Structure Preservation: *In the same context*, the similarity between pairs of response patterns must depend monotonically on the similarity between the corresponding stimulus patterns:

$$d(s_j^q, s_i^q) > d(s_i^q, s_k^q) \iff d(r_j^q, r_i^q) > d(r_i^q, r_k^q) \ \ \forall i, j, k \geq 1$$

3. Context-dependence: *In different contexts*, similar — even identical — stimulus patterns should produce maximally distinct responses:

$$d(r_j^q, r_i^p) = d_{MAX} \ \ \forall q \neq p, \ i, j \geq 1$$

where d_{MAX} is the maximum possible distance between two patterns.

The goal is to produce neural codes that preserve the similarity structure of stimuli within each context, and, at the same time, enhance discrimination between similar stimuli in different contexts.
To make the situation more concrete, consider a system which codes spatial locations in two distinct but visually similar (or even identical) environments. A neural system would, in this case, use distributed patterns of activity to code for different locations, using the visual (and possibly other) input as the driving stimulus. This is precisely what the hippocampal system appears to do [8, 7, 1, 13, 14]. If the purpose of coding locations is to use it for spatial tasks (e.g., route planning), it is clearly desirable to code nearby locations more similarly than distant locations, and a simple feature-driven neural system would do that easily through smooth interpolation. However, it is also essential to code visually similar locations in the two environments in a distinct way. A neural system simply interpolating between visual stimuli cannot do this. Another mechanism at a higher level must, therefore, be posited, which switches the system between different representational repertoires for the two contexts. Within each context, then, the system would interpolate as usual, assuring smooth mapping from input features to output representations. Between contexts, the system would switch, ensuring that the representation codes are completely different. Similar examples can be imagined in tasks such as social behavior, language recognition, etc.

Method

The approach is to use a two-layer recurrent attractor network, which receives external input on one of the layers. Figure 1 shows the network's architecture. Layer I, with n_I elements, is the input layer, which simply provides external stimulus to the system (e.g., visual information). The A layer, with n_A neurons, is the primary processing layer of the system, and its state represents the system output. Layer B, with n_B neurons, has the function of providing recurrent stimulation to A, which allows the stabilization of context as described below.
The connections between layers A and B are set such that the $A - B$ network has N attractors (stable patterns) embedded in it. Each attractor, α_k, is defined by

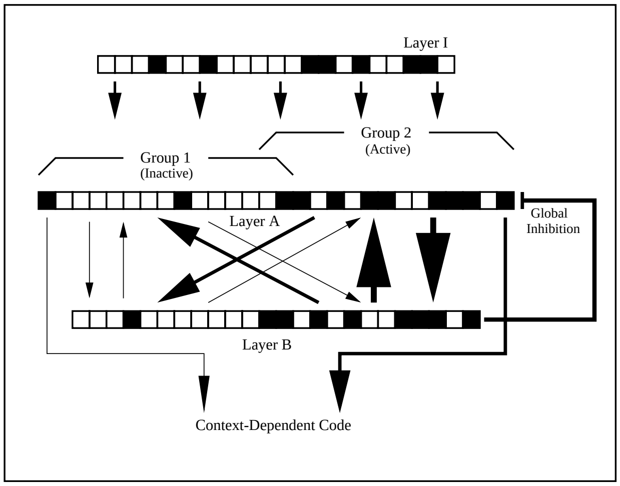

Figure 1. System achitecture.

a subgroup, G_k^A, of neurons in layer A and subgroup G_k^B in layer B. THe subgroups are chosen randomly, with those in layers A and B comprising g_A and g_B neurons each, respectively. The initial connectivity between A and B is specified randomly. However, the values of the connections are set as follows. For neuron j in A and i in B, the weight, w_{ij}, from j to i is set to 1 if $j \in G_k^A$ and $i \in G_k^B$ for any $k = 1, 2, 3, ..., N$. Otherwise, the weight is set to 0. This is a Hebbian style learning rule first studied by Willshaw et al. [12]. The weights from B to A are set in the same way. In addition, layer B also projects a global inhibition onto all neurons of layer A. The net effect is to make each neuron group, $G_k^{A/B}$, an attractor for the $A - B$ system, which would be completely activated if the attractor were to be fully manifested. However, firing in A is competitive such that, at any time, only a fraction, γ of the neurons in layer A can be active, where $\gamma n_A << g_A$. When layer A is stimulated by an external input from layer I, the response is focused primarily in the active group, since its neurons are biased for firing by recurrent excitation while the rest are depressed by recurrent inhibition. However, the precise identity of *which* neurons within the active group fire depends on the external stimulus. The activity in the group is sufficient to keep the chosen attractor stable in the presence of continued external stimulation without ever becoming fully active. In essence, the attractor serves to *channel* the response to external stimulus, and is termed a *latent attractor*.

When a stimulus sequence, S^q, is presented to the system, the initial (context) pattern, s_0^q, preferentially excites the layer A neurons of one subgroup $G_{k_q}^A$, placing the system on attractor α_{k_q}. Subsequent stimuli then produce responses which activate neurons primarily within G_{k_q}, but the precise pattern of activity reflects the stimulus patterns. Similar stimuli produce

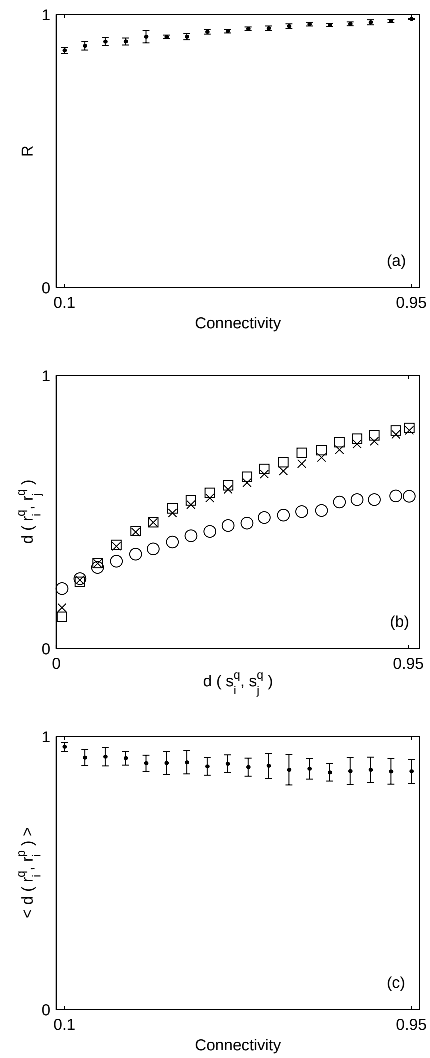

Figure 2. Simulation results: (a) Reliability; (b) Structure preservation ($B \rightarrow A$ $= 0.1(\circ)$, $0.4(\times)$, $0.9(\square)$); (c) Context-dependence (see text for details).

similar responses, thus preserving the structure of the stimulus ensemble. However, the same (or similar) stimuli in another context — with a different attractor — produce totally different responses. Essentially, the context pattern confines future activity to subsets of layers A and B, until a major disruption — e.g., the beginning of a new episode/task — can disrupt the situation and switch the system to another attractor.

One way to see the $A-B$ system is as a set of overlapping, mutually competitive recurrent subnetworks. Instead of an external selector, however, the appropriate subnetwork/group is stabilized by the connectivity structure of the system.

Results

We evaluated the performance of the system using a network with $n_I = 400$, $n_A = 2000$, $n_B = 500$, $N = 10$, $g_A = 200$, and $g_B = 50$. The connectivity from B to A was varied systematically, while the others were set as $I \rightarrow A = 0.4$ and $A \rightarrow B = 0.5$.

To avaluate reliability, R, we generated 10 random stimulus patterns and constructed a stimulus sequence of 300 patterns by choosing randomly from the pattern set. All responses to the same pattern (in various sequential positions) were collected and the maximum likelihood (ML) response determined by averaging the responses bitwise and thresholding this vector. For each pattern, the mean distance between the ML response and the actual responses was calculated. These were averaged for all stimulus patterns, and 1 minus this quantity was used as an estimate of R. Figure 2(a) shows the results for various $B \rightarrow A$ connectivities.

To evaluate structure preservation and context-dependence, we generated a set of 200 stimulus patterns with various degrees of correlation between pattern pairs. Two sequences, each of length 100 (plus the initial context patterns), were then generated such that they differed only in the context pattern:

$$S^1 = s_0^1 s_1 s_2 s_3 s_4 \ldots s_{100}$$

$$S^2 = s_0^2 s_1 s_2 s_3 s_4 \ldots s_{100}$$

The response of the system to the presentation of each sequence was determined. To measure structure preservation, the inter-stimulus and inter-response distances were averaged over five different trials (with different networks) at various $B \rightarrow A$ connectivities. These are plotted in Fig. 2(b), and show a monotonic relationship, as desired, in all cases. To measure context-dependence, the distance between the responses to corresponding patterns in S^1 and S^2 were averaged over all trials and patterns. This is plotted against $B \rightarrow A$ connectivity in Fig. 2(c). A high degree of context-dependence is apparent.

Conclusions

This paper has presented a simple, elegant model for achieving the conflicting goals of context-dependence and smoothness in neural representations.

Acknowledgement: This research was supported by Grant Nos. IBN-9634424, IBN-9808664, and IBN-9816612 from the National Science Foundation, and by the University Research Council, University of Cincinnati.

References

[1] C. Barnes, B. McNaughton, S. Mizumori, B. Leonard, and L.-H. Lin. Comparison of spatial and temporal characteristics of neuronal activity in sequential stages of hippocampal processing. *Progress in Brain Research*, 83:287–300, 1990.

[2] J. Elman. Finding structure in time. *Cog. Sci.*, 14:179–211, 1990.

[3] C. Giles, C. Miller, D. Chen, H. Chen, G. Sun, and Y. Lee. Learning and extracting finite state automata with second-order recurrent neural networks. *Neural Computation*, 4:393–405, 1992.

[4] W. Levy. A computational approach to hippocampal function. In R. Hawkins and G. Bower, editors, *Computational Models of Learning in Simple Neural Systems*, pages 243–305. Academic Press, San Diego, CA, 1989.

[5] W. Levy and X. Wu. The relationship of local context cues to sequence length memory capacity. *Network*, 7:371–384, 1996.

[6] A. Minai, G. Barrows, and W. Levy. Disambiguation of pattern sequences with recurrent networks. In *Proc. WCNN, San Diego*, volume IV, pages 176–180, 1994.

[7] R. Muller, J. Kubie, and J. Ranck, Jr. Spatial firing patterns of hippocampal complex-spike cells in a fixed environment. *J. Neurosci.*, 7:1935–1950, 1987.

[8] J. O'Keefe and L. Nadel. *The Hippocampus as a Cognitive Map.* Clarendon Press, Oxford, UK, 1978.

[9] M. Reiss and J. Taylor. Storing temporal sequences. *Neural Networks*, 4:773–787, 1991.

[10] E. Rolls. The representation and storage of information in neuronal networks in the primate cerebral cortex and hippocampus. In R. Durbin, C. Miall, and G. Mitchison, editors, *The Computing Neuron*, pages 125–159. Addison-Wesley, Reading, MA, 1989.

[11] R. Williams and D. Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1:270–280, 1989.

[12] D. Willshaw, O. Buneman, and H. Longuet-Higgins. Non-holographic associative memory. *Nature*, 222:960–962, 1969.

[13] M. Wilson and B. McNaughton. Dynamics of the hippocampal ensemble code for space. *Science*, 261:1055–1058, 1993.

[14] K. Zhang, I. Ginzburg, B. McNaughton, and T. Sejnowski. Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal cells. *J. Neurophysiol.*, 79:1017–1044, 1998.