Progressive Attractor Selection in Latent Attractor Networks

Simona Doboli
Complex Adaptive Systems Laboratory
ECECS Department
University of Cincinnati
Cincinnati, OH 45221

Ali A. Minai
Complex Adaptive Systems Laboratory
ECECS Department
University of Cincinnati
Cincinnati, OH 45221

Abstract

Latent attractor networks are recurrent neural networks with weak embedded attractors. The attractors bias the network's response to external inputs without becoming fully manifested themselves. Latent attractor networks have been used to model context-dependent spatial representations in the hippocampus [5], and to encode context-dependent stimuli in neural networks [3]. In the current model, the selection of the biasing attractor occurred in response to an initial triggering stimulus indicating the context signal. For example, the sign on a door may set the context for the representation of a room. However, in many realistic situations, context is set by a set of cues rather than a single cue, and these cues are typically seen sequentially, though not in a particular order. The problem addressed here is: How can a latent attractor network progressively select an attractor in response to a sequence of context patterns?

1 Introduction

In many situations the meaning of identical inputs depends on the context specified in the past. Depending on how far in the past the context information is given *relative to the current time*, we can distinguish two types of context-dependent problems. The first one — which may be called near-term context-dependence — refers to situations where the output at time t depends on inputs presented in the immediately preceding interval of time. Examples of such problems appear in dynamical systems, as well as in speech processing and word recognition, and are embodied in autoregressive models or finite-state machines and can be learned by recurrent neural networks [10, 12, 11].

The second case — termed episodic context dependence — refers to problems where the information about context is given transiently at a particular time in the past. Examples of such context are found in social situations, in recognition of spatial environments, task specifications, etc., where the context is typically specified at the beginning of the episode and continues to be in force for its entire duration — even if the episode lasts a long time. In general, artificial neural networks have difficulties in representing inputs in a context-dependent way. Problems involving near-term context dependence are usually solved with recurrent neural networks, where past states are fed back to the input [10, 12, 11]. Episodic context dependence is more difficult to address in the same way, since the information on context is specified at a fixed time in the past, which could be very remote from the present. This requires the system to "latch" information which is difficult even in recurrent networks [1, 11]. For a concrete example, consider a situation when a set of identical rooms (e.g., in a hotel) are distinguished only by the number on the door. Looking at this number then sets the context for the subsequent episode as one enters a room.

Experiments with rodents suggest that the hippocampal region of the brain, which is critical to spatial tasks and episodic memory, can construct distinct representations of similar or even identical environments depending on episodic context and this has been a prime motivation for our work. We have proposed a class of networks called *latent attractor networks* as a paradigm for dealing with episodic context-dependent situations without the use of off-line or external biasing [17, 5, 3, 6]. Latent attractor networks have been used to model context-dependent spatial representations in the hippocampus [17, 4, 5, 6], and to encode context-dependent mappings of stimuli in neural networks [3]. They are recurrent neural networks with competitive firing that embed patterns of activity as attractors using associative learning. However, the recurrent connections are not strong enough to stabilize the patterns autonomously. In the original formulation, each attractor is associated with a specific external stimulus pattern called the context pattern. If the context pattern is presented to the network, it tends to disproportionately activate neurons in the active set of the associated attractor. This sets up a stable bias onto this set via the recurrent connections, with the result that subsequent external stimuli — not explicitly associated with any particular attractor — also produce response patterns whose activity lies largely in the active set of the chosen attractor. Thus, the system produces responses conditioned by the original context pattern long after the context pattern itself is gone. This situation lasts until an external stimulus associated with another context/attractor is presented to the network [5].

In the paradigm described above, the context patterns represent the stimuli that set the context for an episode (e.g., looking at the number on a door). However, in typical situations, context is not set by a single stimulus, but by a conjunction of stimuli. For example, upon entering an environment, one may recognize it by a set of characteristic objects or the presence of certain stimuli. Furthermore, it is possible that any one stimulus be part of the combination indicating several contexts; it is the combination which denotes the specific context. Thus, as a cognitive system scans over the context-setting stimuli, a unique context would only emerge gradually rather than instantaneously. Indeed, until the unique context is identified, the system's response may be compatible with several choices.

In our studies of latent attractor networks, we have not, so far, focused on how latent attractors can be activated gradually by a combination of stimulus patterns rather than by just one pattern instantaneously. The aim of this paper is to investigate this issue. An important point to note here is that the problem here is somewhat different from standard sequence recognition, where stimuli presented in a particular order are to be recognized and differentiated. In the case of context setting, the order of stimuli is not necessarily important, but the stimuli are encountered sequentially — possibly in different order each time. The question addressed here is how a latent attractor network could slowly settle its activity into one of the attractors, as context inputs are sequentially presented and recognized by the network.

The problem of encoding episodes — in this case contexts — by a set of inputs presented sequentially to a neural system is relevant in other situations as well. For example in the case of storing information hierarchically in a neural network, several pieces of information at a lower level define a single entity at a higher level. Whenever the higher concept — for example, the title of a movie — cannot be recalled immediately, the sequential presentation, or recall of lower level concepts — like topic, actors, type, etc. — can slowly direct the recollection of the higher category.

Another example is the case of speech recognition, where words are represented by sequences of phonemes [13, 2]. In the system by Grossberg and Myers [13], which recognizes spoken words, phonemic entities are activated in order, and they in turn, stimulate word entities in another module. Word representations compete for activation and they also stimulate the active phonemes, so that after all components of have been seen by the network, only one module representing the recognized word will be active. This case, however, differs from the general one described earlier in that the order of stimuli is important.

Several neural networks models have been developed for encoding and recovering hierarchical information [8, 9]. There are modular networks, where the lower level concepts are stored in different modules, and the higher level categories are represented by the activation of all the component modules. Hierarchical networks learn patterns on multiple levels by representing them through specially designed patterns where the common part is retained in the high level concepts [8, 9].

2 Problem Definition

The system's task is to progressively recognize a set of context inputs presented sequentially in a random order.

The network is presented with M different external stimulus sequences in discrete time. Each sequence

$$S^q = C_1^q C_2^q ... C_p^q R_{p+1}^q R_{p+2}^q ... R_n^q; \quad q = 1, 2, ..., M$$

consists of p context patterns followed by $n-p$ regular patterns. The context patterns, C_i^q are drawn from the context set ,$\{C_k\}$, and the regular patterns are drawn from the set, $\{R_k\}$, C_k, $R_k \in I$, where I is the input space of dimension N_i.

The set of context patterns, $C^q = \{C_1^q C_2^q ... C_p^q\}$, for sequence S^q is selected randomly without repetition from the context set C and presented at the network input in an arbitrary order. The context pattern sets for different episodes may share patterns. The regular stimulus patterns for each episode are also chosen randomly from R, and may overlap between episodes.

At the beginning of an episode, as context patterns are presented, the network response has to become confined gradually to a state associated with the particular context pattern set. At the beginning of the sequence, the network's state overlaps with all candidate states that are activated by the current context input. Competition between simultaneously stimulated network states and subsequent evidence in the form of more context patterns makes neurons less likely to be active in the correct attractor stop firing, and the ones which are continuously reinforced to increase their activity level. By the end of the context pattern sequence, the network activity should be concentrated completely in the active set of the specified attractor.

3 Method

The approach uses a two-layer latent attractor network. The stimulus layer, L_S, has N_S neurons that project the input patterns to the response layer, L_R. The response layer, with N_R neurons, also receives a disynaptic recurrent connection through the intermediary layer, L_H with N_H neurons. The connections from L_S to L_R are set randomly with probability p_S of connection. Only K_S neurons in the input layer are active at any one time.

The latent attractors are stored in the recurrent connections between the L_R and L_H layers. There are M attractors, each comprising two binary patterns, one in layer L_R and the other in layer L_H. The patterns are sparse, with G_R and G_H active neurons, respectively in L_R and L_H. The sets of G_R and G_H neurons that would be active if the attractor were fully manifested are termed the *active sets* of the attractor. The connections between L_R and L_H layers are chosen randomly with probability of connections p_R (L_H to L_R) and p_H (L_R to L_H). The attractors are embedded in the connections by setting the weights according to a clipped binary Hebbian rule first proposed by Willshaw: The connections between neurons active in the two patterns of any attractor are set to high values, while the rest are set to low values. In this way, the M pairs of patterns are set as attractors or fixed points in the in the 2-layer network. The attractors are called latent because they are not allowed to become fully active at any time. The activity in the network is determined in the following way. The excitation to a layer L_R neuron, i, at time t is given by:

$$y_i(t) = \sum_{j \in L_S} w_{ij} x_j(t) + g_i(t) sum_{j \in L_H} w_i j z_j(t-1)$$

where w_{ij} denote connection weights, $x_j(t)$ is the jth bit of the external stimulus patterns at time t, $z_j(t)$ is the output of neuron j in layer L_H, nd $g_i(t)$ is the (modiafible) recurrent gain of neuron i.

The excitation to a layer L_H neuron, i, is given by:

$$y_i(t) = \sum_{j \in L_R} w_i j z_j(t)$$

where $z_j(t)$ is the output of $j \in L_R$.

Once the excitation to all L_R and L_H neurons has been determined, firing in both layers is competitive: The output of the K_R (K_H) most excited neurons in L_R (L_H) at time t is set to 1, while the rest of the neurons output 0. This corresponds to a K-winner take all competitive firing rule. The values of K_R and K_H respectively are much smaller than G_R and G_H, the respective sizes of the attractors active sets in each layer.

Latent attractors are associated with stimulus sequences as follows: The connections between L_S and L_R layers are modified such that context patterns in each sequence (C^q) stimulate mostly neurons in the active set of the corresponding attractor in the L_R layer.

The recurrent gain parameter g_i is an important one, since it controls the stability of attractors by determining the strength of the recurrent projection to L_R relative to the external stimulus. For a latent attractor to be persistent, neurons in its active set must have a minimum value of g_i [6].

When the regular patterns of a stimulus sequence are presented at the input, the system has to be in a stable regime: The network's state has to vary smoothly with the current input, but, at the same time to stay within the neurons of the attractor activated initially following the context patterns [3]. The requirement of smooth dependence between the regular stimuli and the corresponding responses is important because the context-dependence representations generated by the network should not lose any similarity information inherent in the timulus set. For example, if the stimuli are sensory representations in a moving animal, similarity of these stimuli might indicate spatial proximity, and this information should not be lost. We have shown previously that achieving the twin (and apparently incompatible) objectives of preserving stimulus information and sustaining latent attractors requires careful (but not critical) management of the recurrent gain, g_i [5], and also that 2-layer recurrent networks are better able to support this than 1-layer networks [6].

The key requirement for correct context selection is that when the context patterns of a stimulus sequence are presented, the network's state should not settle until it has received enough context patterns to uniquely identify the sequence. This is achieved in our system through the strategy of *incremental competitive positive feedback*. The stability of any attractor in the network is affected decisively by the recurrent gains, g_i. When g_i are small (relative to the strength of the external stimulus), attractor dynamics is dominated by the impact of the feed-forward stimulus. If the stimulus is selectively associated with particular sets of neurons, these are likelier to fire since total activity is constrained to be no more than K_R neurons in L_R. If the g_i are large, the recurrent path dominates and the network is forced to choose between attractors due to competitive firing in L_R and L_H.

In our system, all g_i are set to a small value at the beginning of an episode, leaving the choice of activity in L_R to the initial context stimuli. Thus, attractors that are associated with the early context patterns are likely to be activated a bit more than others due to feed-forward association. As the presentation of context patterns proceeds, g_i for neurons that belong to the active sets of attractors with more current activity is increased gradually, priming these attractors for possible persistence if reinforced by subsequent context stimuli. Thus, at each stage, activity is distributed among those attractors that are consistent with the context stimuli received thus far. As each new context stimulus is presented, some of these "hypothesized" candidate attractors are reinforced further at the expense of others until, finally, only one left.

The equation governing the modulation of recurrent gain is:

$$g_i(t) = g_i(t-1) + \alpha[a_l(t) - a_l(t-1)]/K_R$$

where α is a rate parameter, l is the index of the attractor for which i is in the active set, and a_l is the total number of currently firing neurons in the L_R active set of attractor l. If i belongs to the active set of several attractors, the largest value of g_i chosen over all these attractors is used.

The modulation of recurrent gain on individual neurons is motivated by several biological considerations:

1. There is extensive evidence that projections to neurons in most cortical regions are segregated on the dendritic tree, making the selective modulation of gain on projections from individual sources quite feasible. Indeed, such modulation has been suggested as the basis of efficient associative learning [15].

2. It is well known that, in the hippocampal region, which is the basis for our model, animals are especially attentive at the beginning of an episode, as indicated by the change in the EEG theta rhythm. This leads to, for example, greater spike synchronization, lower firing latency, ad other phenomena.

3. In the granule cells of the dentate gyrus, which, we hypothesize, corresponds roughly to our layer L_R, there is both anatomical and physiological evidence [14, 16] of an intricate and highly specific system of excitability modulation based on motivational and attentional circumstances.

Our model represents a simple first-cut attempt to understand how these mechanisms may help support gradual context selection in the hippocampus and artificial neural networks motivated by the hippocampus.

4 Simulation Results

Simulation were done using a two layer latent attractor network with the following parameters: $N_S = 400$, $K_S = 40$, $p_S = 0.4$, $N_R = 2000$, $G_R = 200$, $K_R = 40$, $p_R = 0.4$, $N_H = 500$, $G_H = 50$, $K_H = 45$, $p_H = 0.8$. There are $M = 10$ attractors embedded in the connections between L_R and L_H layers. The modulation rate for recurrent gain g_i is $\alpha = 0.1$.

The context set C has 20 distinct patterns, from which 5 context sets are selected. Each C^q consists of $p = 5$ distinct patterns picked randomly from C. The context patterns in different C^q are not mutually exclusive. Each C^q set of context patterns is associated with a randomly chosen attractor: The connections from L_S to L_R layers are potentiated off-line such that patterns in C^q are more likely to activate neurons in the active set of the associated attractor.

The patterns in each context set are presented in a random order at the network input layer. At the beginning of each sequence, the recurrent gain for all attractors is set to a low value. Depending on how many context groups are simultaneously stimulated by the incoming context patterns from a C^q, the activity in the latent attractor layer is distributed among the excited attractors. The recurrent gain of the potentiated attractors goes up, while the one of the rest of the attractors is

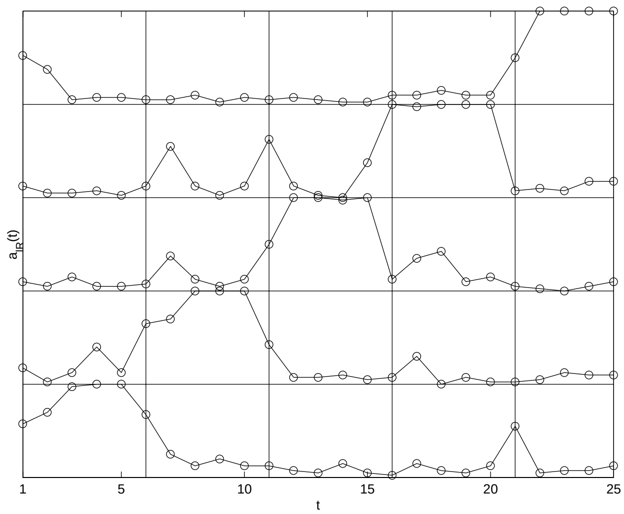

Figure 1: The activity level in the selected attractors in the L_R layer with respect to time. Every $p = 5$ time steps a different context sequence starts. The activity is normalized with K_R.

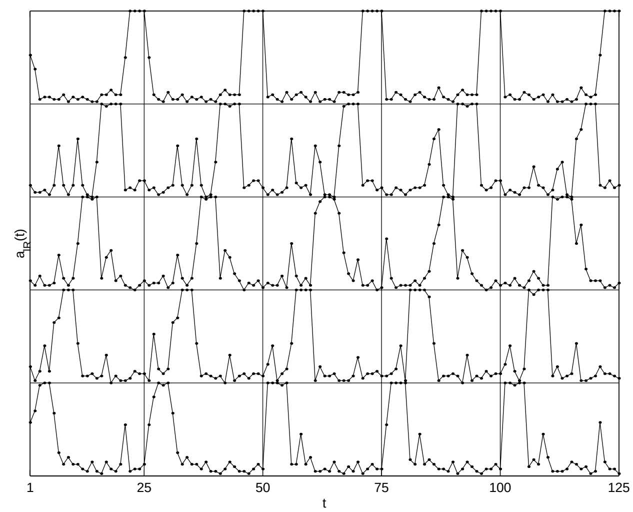

Figure 2: Several repeats of the run in Figure 2, with the same context pattern sets but different pattern order in each set each time.

decreased. At the end of a context sequence, only one attractor is consistently stimulated by the current context patterns. This, together with the modulation of the recurrent gain, results in a continuous increase in the amount of activity in the selected attractor's active set.

Figure 1 shows the result of a single network simulation, when context sequences are presented at the input. The sequences are presented in order from C^1 to C^5 over 5 time steps each. Each graph represents the normalized activity within the active set of an attractor pattern in the L_R layer. It can be seen that, for each context sequence (every $p = 5$ time steps), the activity in only one of the attractors goes up steadily. In all other attractors the activity might increase for a few time steps, but it finally shuts down. Figure 2 shows the results when the simulation is repeated with the same 5 contexts but with the context patterns presented in different order each time. As can be seen, this has no real effect on performance. Figure 3 shows a repeat of the same process, but this time withy 10 regular stimulus patterns following each context set. It is clear that the activity remains confined within the chosen attractor even though the regular patterns have no association with any attractor. We do note, however, that the incorrect attractor was chosen in the second episode. This will happen occasionally in small networks with limited capacity.

In Figure 4, the activity in all context attractors is averaged over 10 different networks and, for each network, over five random presentations of the context patterns in each set. At the beginning of a context sequence, the activity level is low, with a high variance, due to the difference in the number of attractors simultaneously excited. The activity becomes progressively confined into one of the attractors, as more context patterns are seen, reaching a single attractor with probability close to 1 by step 5.

5 Conclusions

We have proposed a mechanism by which an attractor in a latent attractor network can be activated progressively by a set of inputs presented sequentially in a random order, rather than by a single cue. The system is able to slowly select and activate the right attractor, even though, each individual input pattern can be associated with one or more attractors. Importantly, the response is invariant to the order in which the context patterns within a set are presented.

The process of associating a set of inputs to a single episode has strong implications in the recognition of spatial environments, or the formation of hierarchical memory systems.

References

[1] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is diffi-

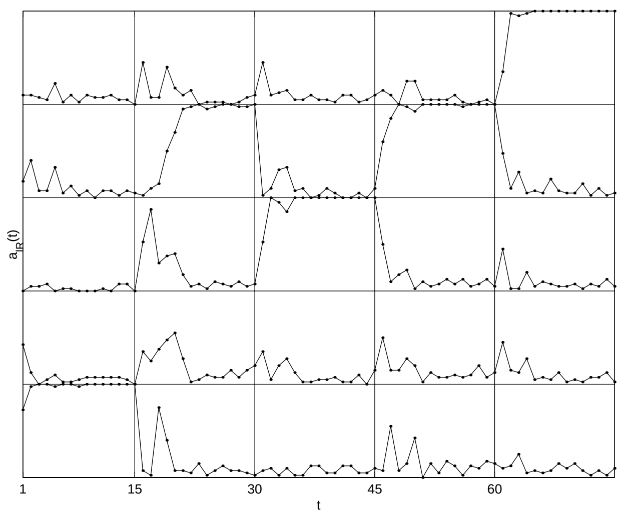

Figure 3: A repeat of the run in Figure 2, but with 10 regular patterns following the presentation of each context set.

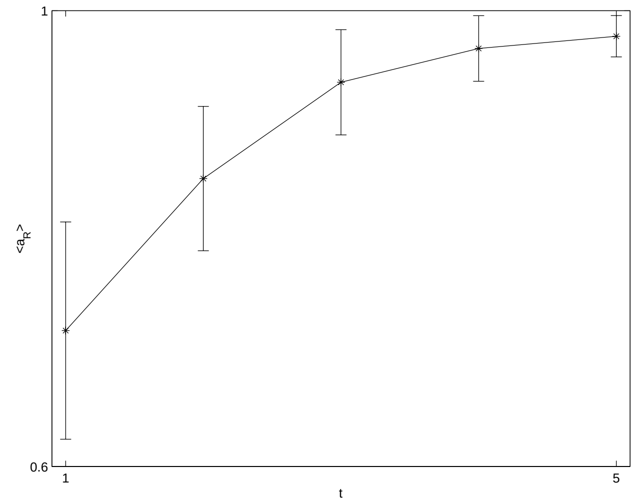

Figure 4: The mean activity level of all context patterns with respect to 10 different networks and five different presentation orders of the context patterns. The time axis corresponds to the size of the context set ($p = 5$).

cult," *IEEE Trans. on Neural Networks*, vol. Vol. 5, No. 2, pp. 157–166, 1994.

[2] G. Bradski, G.A. Carpenter and S. Grossberg. STORE working memory networks for storage ad recall of arbitrary temporal sequences. *Biological Cybernetics* 71:469–480, 1994.

[3] S. Doboli, A.A. Minai and P.J. Best. Generating smooth context-dependent representations. *Proc. of IJCNN'1999*

[4] S. Doboli, A.A. Minai, and P.J. Best. A latent attractors model of context-selection in the dentate gyrus-hilus system. *Neurocomputing* 26-27:671–676, 1999.

[5] S. Doboli, A.A. Minai and P.J. Best. Latent attractors: a model for context-dependence place representations in the hippocampus. *Neural Computation* 12:1009–1043, 2000.

[6] S. Doboli and A.A. Minai. Network capacity for network attractor computation. *Proc. IJCNN'2000* 222-228, 2000.

[7] S. Doboli, A.A. Minai, and P.J. Best, "A comparison of context-dependent hippocampal place codes in 1-layer and 2-layer recurrent networks," *Neurocomputing*, 3-33:353–358, 2000.

[8] D.R.C. Dominguez. Information capacity of a hierarchical neural network. *Phys. Rev. E* 58:4811–4815, 1998.

[9] V.S. Dotsenko. Hierarchical model of memory. *Physica A*, 410–415, 1986.

[10] J.L. Elman, "Finding structure in time," *Cognitive Science.*, vol. 14, pp. 179–211, 1990.

[11] P. Frasconi and M. Gori, "Computational capabilities of local-feedback recurrent networks acting as finite-state machines," *IEEE Trans. on Neural Networks*, vol. Vol. 7, No. 6, pp. 1521–1525, 1996.

[12] C.L. Giles, C.B. Miller, D. Chen, H.H. Chen, G.Z. Sun, and Y.C. Lee, "Learning and extracting finite state automata with second-order recurrent neural networks," *Neural Computation*, vol. 4, pp. 393–405, 1992.

[13] S. Grossberg and C. W. Myers. The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review*, in press.

[14] Z.-S. Han, E.H. Buhl, Z. Lörinczi, and P. Somogyi, A high degree of spatial selectivity in the axonal and dendritic domains of physiologically identified local-circuit neurons in the dentate gyrus of the rat hippocampus. *Eur. J. Neurosci.* 5 (1993) 395–410.

[15] M.E. Hasselmo, E. Schnell, and E. Barkai. Dynamics of learning and recall at excitatory recurrent synapses and cholinergic modulation in hippocampal region CA3. *J. Neurosci.*, 15:5249–5262, 1995.

[16] M.B. Jackson and H.E. Scharfman, Positive feedback from hilar mossy cells to granule cells in the dentate gyrus revealed by voltage-sensitive dye and microelectrode recording. *J. Neurophysiol.* 76 (1996) 601–616.

[17] A.A. Minai and P.J. Best. Encoding spatial context: A hypothesis on the function of the dentate gyrus-hilus system. *Proc. of IJCNN'1998* 587-598, 1998.