

Spatiotemporal Novelty Detection Using Resonance Networks

Benjamin Rowland[†] and Anthony S. Maida^{†‡}

[†] Institute of Cognitive Science

[‡] Center for Advanced Computer Studies

University of Louisiana at Lafayette

Lafayette, LA 70504-44330

Email: rowland@louisiana.edu , maida@cacs.louisiana.edu

Abstract

We present a single-layer recurrent neural network that implements novelty detection for spatiotemporal patterns. The architecture is based on the structure of region CA3 in the hippocampus, which is believed to implement this function. Through analysis and numerical simulation we generate theorems that constrain the operation of this network. We show that once a pattern has been encoded, it will never be miscategorized as novel. We also show that the upper bound on the network capacity is equal to the number of connections. We discuss the tradeoff between generalization and total performance in the architecture.

Introduction

We present a single-layer recurrent neural network that implements novelty detection for spatiotemporal patterns. The architecture and its behavior are based on region CA3 (*cornu ammoni 3*) of the hippocampus, which is believed to implement this function. We present a set of theoretical results that constrain this and more elaborate models, including an analysis of the tradeoffs between generalization and total performance. We conduct a set of simulations to investigate scenarios with random pattern selection. We conclude with a comparison to other work and some ways to increase the sophistication of the architecture.

A general problem in the design of intelligent agents with memory is the determination of which events in the environment should be stored for future reference. It has long been recognized that the memory system of humans and other animals is selective in nature (James 1890). In particular, organisms preferentially select to encode new or unexpected stimuli (von Restorff 1933). The goal of this selection is to maximize the accuracy of the organism's internal model of the external world by encoding those events that the internal model does not already predict (Sokolov 1960). The advantage of a selective memory can also be phrased statistically as reducing sampling artifacts. The frequency of an event in the natural world does not necessarily correlate with its importance to survival, and the repeated encoding of familiar events skews distributions used for decision-making away from more rare but potentially more relevant events

(Marsland 2001). Equivalently, the repeated encoding of familiar data points leads to the formation of large, degenerate representations in neural networks (Kali & Dayan 2000), an example of "catastrophic forgetting".

In our search for optimal artificial architectures that implement selective memory, it is natural to investigate the ways in which organisms may implement it. In this paper we focus on one component of a selective memory, the detection of novel stimuli. It is important to note that the physiological responses associated with the presentation of *unexpected* stimuli are the same as for completely new stimuli (see review in Ranganath & Rainer, 2003). Novelty detection is a dynamic procedure: rather than passively categorizing stimuli as new or old, animals actively generate internal predictions for external events and evaluate whether or not they are accurate. When a mismatch occurs, novelty is detected and the animal encodes the new information.

Cellular recordings from region CA3 in the rabbit hippocampus reveal changes in the spike rate contingent upon the presentation of novel stimuli (see review in Vinogradova, 2001). Specifically, most cells in this region decrease their firing rate when a novel stimulus is presented. It is believed that this region is where the novelty computation takes place, since areas "upstream" to it (e.g. entorhinal cortex) do not respond differently to novel stimuli and areas "downstream" from it (e.g. the septum) do respond differently. The hippocampus is generally recognized as implementing a memory system in a "loop" of which CA3 is one of the initial stages (O'Keefe & Nadel 1978). CA3 directly and indirectly projects to areas of the brain typically associated with increases in arousal (Swanson 1978). Thus, CA3 seems ideally suited both physiologically and anatomically to act as a novelty detector for a selective memory. The most pronounced architectural feature of CA3 is that the principal cells exchange multiple recurrent connections with one another. It is known that these connections implement temporally-asymmetric learning rules, i.e., connections are strengthened when a postsynaptic cell fires after a presynaptic cell, but are weakened if the reverse relationship occurs (Bi & Poo 1998).

To a first approximation CA3 can be abstracted as a single-layer recurrent neural network, commonly referred to as a "Hopfield" or "Willshaw" network (Willshaw, Buneman, & Longuet-Higgins 1969; Hopfield 1982). However,

CA3 neurons can exchange multiple connections with variable transmission delays. A single-layer recurrent neural network architecture generalized to include multiple connections with different delays has been referred to as a “resonance network”. When a learning algorithm strengthens connections with delays matching the firing delays of the units they connect (Equation 3 in this paper), the network is trained to *resonate* with input patterns (Toulouse, Dehaene, & Changeux 1986; Dehaene, Changeux, & Nadal 1987). In a recent theoretical study it was shown that temporally asymmetric learning rules of the sort found in CA3 can approximate the learning algorithms used by resonance networks (Senn, Schneider, & Ruf 2002). Thus, resonance networks may offer a better approximation of CA3 than Hopfield / Willshaw networks.

We will demonstrate how resonance networks can implement novelty detection in a manner consistent with the physiology of CA3.

Network Design

The network is very simple in structure. This simplicity will support a set of theorems that give key intuitions applicable to more elaborate architectures. We conduct a set of simulations that reveal the expected performance of the network under conditions of random pattern selection.

The architecture consists of a single layer of interconnected units. Each unit computes a unit step function denoted $\Psi(\dots)$:

$$\Psi(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (1)$$

We refer to the set of all N units in the network as K . Units in the network receive binary “feedforward” input from an external population. We denote the external input to unit k in the network at time t by $i_k(t) \in \{0, 1\}$. We identify a pattern presented to the network by the set of network units that receive positive external input at different times. Pattern F_1 of length T consists of a set of snapshots $F_1(1), F_1(2), \dots, F_1(T)$; where $F_1(t) = \{k \in K | i_k(t) = 1\}$. We impose the constraint that the size of each snapshot is the same, i.e., that the input spike density is some constant $X = \sum_{k \in K} i_k(t)$ for every moment in time t .

To compute on the time domain, a network needs an internal temporal structure. We provide this structure in the transmission delays of connections between units. Every unit in the network projects to every other unit S connections, each with a unique transmission delay $s \in [s_{lo}, s_{hi}]$, with $s_{lo} = 1$. We write the strength (weight) of the connection from unit j to unit k with delay s at time t as $w_{kjs}(t)$. Weights only take values of -1 or 0 and can be modified during training. All weights are initialized to -1 . Figure 1 illustrates the network architecture.

The output of unit k at time t is calculated according to Equation 2:

$$a_k(t) = \Psi \left(i_k(t) - 1 + \sum_s \sum_{j \in K} w_{kjs}(t) a_j(t-s) \right) \quad (2)$$

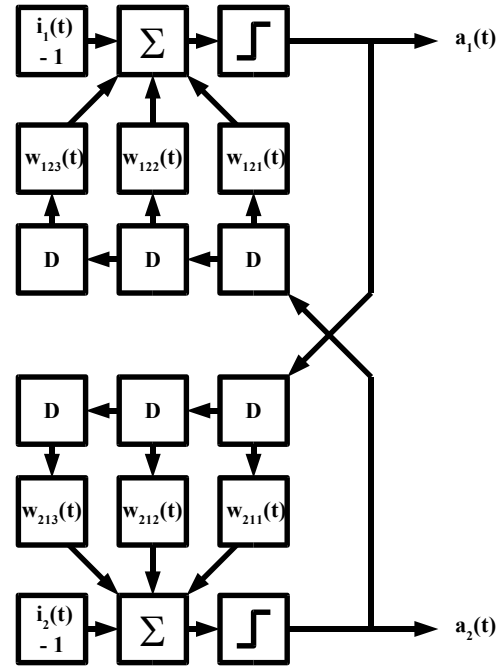


Figure 1: The network architecture. Depicted are units 1 and 2 with $s_{hi} = 3$. Each unit k receives an external input at time t denoted $i_k(t)$. Each unit output is computed by summing the external input, a bias of -1 , and the recurrent input and passing the result through a unit step function. Every unit in the network projects S connections to every other unit in the network including itself. Each delay box D imposes a delay of one timestep. Self-coupling connections (e.g., from unit 1 to 1) are not shown, but take the same form as cross-coupling connections.

Thus, the output of unit k at time t is 1 (unit k is active) if and only if it receives positive external input and every unit with positive output at time $t - s$ projects a connection with a delay s and a weight equal to 0 to k . We separate the operation of the network into two phases: encoding and evaluation. During an encoding phase we calculate weight changes according to Equation 3:

$$w_{kjs}(t+1) = w_{kjs}(t) + i_k(t) \cdot i_j(t-s) \quad (3)$$

With the constraint that weights saturate at a maximum of 0. Weight changes are applied at the end of the encoding phase. During an evaluation phase, weights do not change. We interleave encoding and evaluation phases to evaluate network performance (patterns correctly categorized as novel / familiar) as a function of the number of stored patterns.

The network categorizes a pattern as novel when $\sum_{k \in K} a_k(t) < X$ at any time t , i.e., the actual number of units active is less than the expected number X . As in the hippocampus, novelty is signaled by a decrease in the number of units that are active.

Analysis

There are two theorems that constrain the behavior of these networks. The first theorem proves that after a pattern is

stored, it will never be miscategorized as novel. The second theorem proves that for no generalization to occur, each pattern must be different from all stored patterns by at least one three-tuple of a projecting unit, receiving unit, and delay.

Let $G = \{F_1, F_2, \dots, F_p\}$ be the set of distinct patterns that have been stored in the network.

Definition A new pattern $F_h \notin G$ is correctly categorized as novel if at some time t there is at least one unit k where $k \in F_h(t)$ and $a_k(t) = 0$.

Definition A pattern $F_i \in G$ is correctly categorized as familiar if for all times t there is no unit k where $k \in F_i(t)$ and $a_k(t) = 0$.

We first show that all stored patterns will be correctly categorized as familiar.

Theorem 1 *If pattern F_i has been stored in the network, then it will never be categorized as novel.*

Proof. Assume for contradiction that pattern F_i has been stored in the network but is categorized as novel. Then there must be some time t and unit k where $k \in F_i(t)$ but $a_k(t) = 0$. For $a_k(t)$ to equal 0, either $i_k(t) = 0$ or there exists some j and s such that $a_j(t-s) = 1$ and $w_{kjs}(t) = -1$. $k \in F_i(t)$ implies that $i_k(t) = 1$, so the latter must be true. $a_j(t-s) = 1$ implies that $i_j(t-s) = 1$. Then because $i_k(t) = i_j(t-s) = 1$ and pattern F_i has been stored, $w_{kjs}(t)$ must equal 0, which contradicts the earlier conclusion that $w_{kjs}(t) = -1$. ■

As is the case with most neural networks, this simple design generalizes across patterns (i.e., it may categorize a novel pattern as familiar). To illuminate where and when generalization occurs in our network design, we present the conditions that must be met for no generalization to occur.

Theorem 2 *Pattern $F_h \notin G$ is (correctly) categorized as novel if and only if there is at least one three-tuple (k, j, s) unique to pattern F_h such that $k \in F_h(t)$ and $j \in F_h(t-s)$ for some time t and there does not exist a pattern $F_i \in G$ and time t' such that $k \in F_i(t')$ and $j \in F_i(t'-s)$.*

Proof. To prove the forward implication, assume for contradiction that $F_h \notin G$ is categorized as novel but there does not exist a distinct three-tuple meeting the above criteria. Since F_h is categorized as novel, there is a t and unit k such that $k \in F_h(t)$ but $a_k(t) = 0$. It follows that there must be some unit j and delay s such that $j \in F_h(t-s)$ but $w_{kjs}(t) = -1$. But because no three-tuple exists, there must be some time t' and pattern $F_i \in G$ such that $k \in F_i(t')$ and $j \in F_i(t'-s)$. However, this fact implies that $w_{kjs}(t) = 0$, a contradiction of the conclusion that $w_{kjs}(t) = -1$.

To prove the reverse implication, assume for contradiction that at least one distinct three-tuple exists but $F_h \notin G$ is categorized as familiar. For F_h not to be categorized as novel, for every time t and unit $k \in F_h(t)$ it must be the case that $a_k(t) = 1$. For the output of these units to be 1, it must be the case that $w_{kjs}(t) = 0$ for every delay s and unit $j \in F_h(t-s)$, that is, there must be some time t' and pattern $F_i \in G$ such that $k \in F_i(t')$ and $j \in F_i(t'-s)$. However, the assumption that a distinct three-tuple exists implies that for at least one three-tuple a pattern F_i and time t' cannot be found that satisfies this relation, a contradiction. ■

$$\begin{aligned} F_1(1) &= \{1, 2\} & F_1(2) &= \{1, 2\} \\ F_2(1) &= \{1, 2\} & F_2(2) &= \{3, 4\} \\ F_3(1) &= \{3, 4\} & F_3(2) &= \{1, 2\} \\ F_4(1) &= \{3, 4\} & F_4(2) &= \{3, 4\} \end{aligned}$$

Table 1: Only four patterns of length 2 are necessary to set all of the weights to 0 when $N = 4$ and $s_{hi} = 1$. After these four patterns are stored, all 36 possible patterns will be categorized as familiar.

Thus, for every pattern to be categorized correctly before and after it is stored it must include at least one distinct three-tuple of (k, j, s) , that is, it must change at least one weight when it is stored. There are two corollaries that are generated from Theorems 1 and 2:

Corollary 3 *The upper bound on the number of patterns that can be stored in the network while no generalization occurs is the number of weights in the network, i.e., SN^2 .*

We will refer to this number as the upper bound on the network capacity. Intuitively, for no generalization to occur patterns must be orthogonal in an SN^2 -dimensional space defined by weights and their delays. The mapping between the population producing the external input and the network population does not affect the upper bound. Thus, making the feedforward connections adaptable cannot improve performance past this point. However, adaptable feedforward weights and preprocessing layers can help performance approach the upper bound on capacity. It is important to note that such an augmentation would be necessary to maximize performance. Theorem 2 states that any pattern to be stored must use a distinct three-tuple, and Corollary 3 concludes that only SN^2 unique three-tuples can exist. Clearly $SN^2 \leq \binom{N}{X}^T$, the number of possible patterns of length T when inputs are independent of one another. A preprocessor is necessary to ensure all unstored patterns will map to a (potentially not unique) unallocated three-tuple prior to storage. Intuitively, once a pattern is stored it removes at least 1 three-tuple from the SN^2 we have available to allocate. Once we have stored SN^2 patterns, we will run out of unique three-tuples to represent any unstored pattern.

Corollary 4 *A minimum of $(N/X)^2$ patterns of length $s_{hi} + 1$ need to be stored to maximize generalization.*

Generalization is maximized when all weights are set to 0 by the minimum number of stored patterns. We refer to the strategy that accomplishes this goal as an ‘‘adversary’’. The strategy is to divide the network units into unique clusters of size X and store patterns only using these clusters. For example, if $X = 2$ and there are four units in the network, the adversary could parse units 1 and 2 and units 3 and 4 together. All weights in the network can be set to 0 for $s_{hi} = 1$ by storing the four patterns listed in Table 1.

$(N/X)^2$ patterns are required to change the weights of every connection with the maximum delay, which is the limiting factor. At most X^2 of these can be changed for each pattern, but for weights with lesser delays more than X^2 can be changed per pattern. Once all weights have been changed, all $\binom{N}{X}^{S+1}$ possible patterns will be categorized as familiar.

We can control generalization in the network by controlling the input spike density (i.e., X). This feature is not unique to the function of novelty detection or our particular architecture (cf. McClelland, McNaughton, & O'Reilly, 1995). It is a general property of neural networks: sparse patterns are more likely to be orthogonal in the SN^2 -dimensional weight space. However, as X is decreased, the number of possible patterns also decreases (assuming $X < N/2$). Thus, a network using sparse patterns has a lower initial performance (total number of patterns correctly categorized) than a network using distributed patterns. The advantage of distributed patterns is that they allow us to encode a larger set of external events. So if generalization is highly undesirable and relatively few external events need to be encoded, X should be relatively low. However, if generalization is desirable and many external events need to be encoded, X should be relatively high.

Numerical simulations are conducted to confirm our analyses and investigate network performance under conditions of random and adversarial pattern selection for patterns of length $S + 1$. In applications where pattern length is greater than $S + 1$, a single pattern of length T is equivalent to storing $1 + T - (S + 1)$ (potentially unique) patterns of length $S + 1$. After each pattern is stored, we test the network on the range of all possible patterns and record how many are correctly categorized. We plot this number as a function of the total number of patterns stored to visualize the importance of different network parameters and pattern selection conditions. Thus, we pair a single encoding phase (for each new pattern) with $\binom{N}{X}^{S+1}$ evaluation phases (all possible patterns).

Figure 2 displays the results for $N = 6$, $N = 8$, and $N = 10$, where X is either 1 or 2. The right plots are magnifications of the bottom portions of the left plots. The six plots illustrate the tradeoff between larger values of X and miscategorizations of unstored patterns as familiar. The number of possible patterns is $\binom{N}{X}^{S+1}$, and we are interested in the biologically plausible range where $X \ll N$ and $N \gg 1$. In this range an increase in either X or N produces an exponential increase in the number of possible patterns, and therefore the number of events that can be presented to the network (in our simple example where $X \in \{1, 2\}$ it is only quadratic with respect to X). However, increasing X also increases the average number of weights that are set to 0 by each pattern. This fact leads to a quicker decrease in the number of patterns accurately categorized with respect to the number of stored patterns. The intersection point between the $X = 1$ and $X = 2$ conditions gives the point of the tradeoff, where selecting $X = 1$ would produce better results if more than this many patterns must be stored in the network. With larger values of N the $X = 1$ curve becomes increasingly shallow, making the intersection point more favorable to the $X = 1$ condition. However, with larger values of N the potentially exponential scaling of the number of patterns vs. X becomes more significant.

Because patterns are encoded regardless of whether they are categorized as novel, performance (temporarily) increases linearly from minima as patterns that were (erro-

neously) categorized as familiar are stored. Performance (permanently) increases linearly once all weights are set to 0.

Discussion

The primary advancement of this study is the design of a simple model that gives the basic intuitions for how a particular biological structure implements its function. However, we have also provided a foundation on which a series of artificial neural networks that dynamically compute the novelty of spatiotemporal patterns can be built. We begin this section by comparing and contrasting our design with existing solutions to the problem of novelty detection. We conclude the section with some possible elaborations of the design that either improve performance and/or make the model more biologically realistic.

The Kohonen-Oja filter is an early example of a novelty detector (Kohonen & Oja 1976). It consists of a feedforward architecture with single linear unit (i.e., a linear perceptron) that uses an anti-Hebbian or decorrelative rule to train the feedforward weights. Because weights decrease when a pattern is presented, the output is positive if and only if a new pattern differs significantly from all previous patterns. However, these networks are not designed to use spatiotemporal patterns. Adaptive Resonance Theory (ART) networks also contain components designed to detect novelty in input patterns (e.g., Carpenter & Grossberg, 1987; 1990). They use a module that determines the distance between a new pattern and the closest matching stored pattern. However, ART networks have not been designed to compute on spatiotemporal patterns. There have been several proposals of networks of Self-Organizing Feature Maps (SOFMs) designed for novelty detection (e.g., Muruzábal, 1998). These networks use competitive algorithms to train individual units in the network to recognize areas of the input space (where each input unit is a dimension). Novelty is detected when a new pattern is not recognized by any existing unit in the network. The Grow-When-Required (GWR) variant of the SOFM can compute on spatiotemporal input patterns (Marsland 2001). However, it does not use recurrent connections between network units to dynamically compute novelty in the way we have designed, which we see as more closely approximating the behavior of biological novelty detectors.

The most similar approach to our design is from Bogacz, Brown, and Giraud-Carrier, who use the Lyapunov energy of a Hopfield network as a measure of the novelty of a pattern (Bogacz, Brown, & Giraud-Carrier 1999). Intuitively, energy is proportional to the summed mismatches of unit states and predictions of those states based on states of other units in the network and the connecting weights. The number of units active in our network is proportional to the summed matches of unit states and the predictions of those states based on unit states at times $t - s$ and the connecting weights with delay s . Thus, a decrease in the total number of units active in the network is intuitively equivalent to a rise in energy. Our architecture can be considered a variant of this proposal where energy is computed *internal to the network*, i.e., without using some peripheral device with access to all weights and states that computes the energy. Biologically,

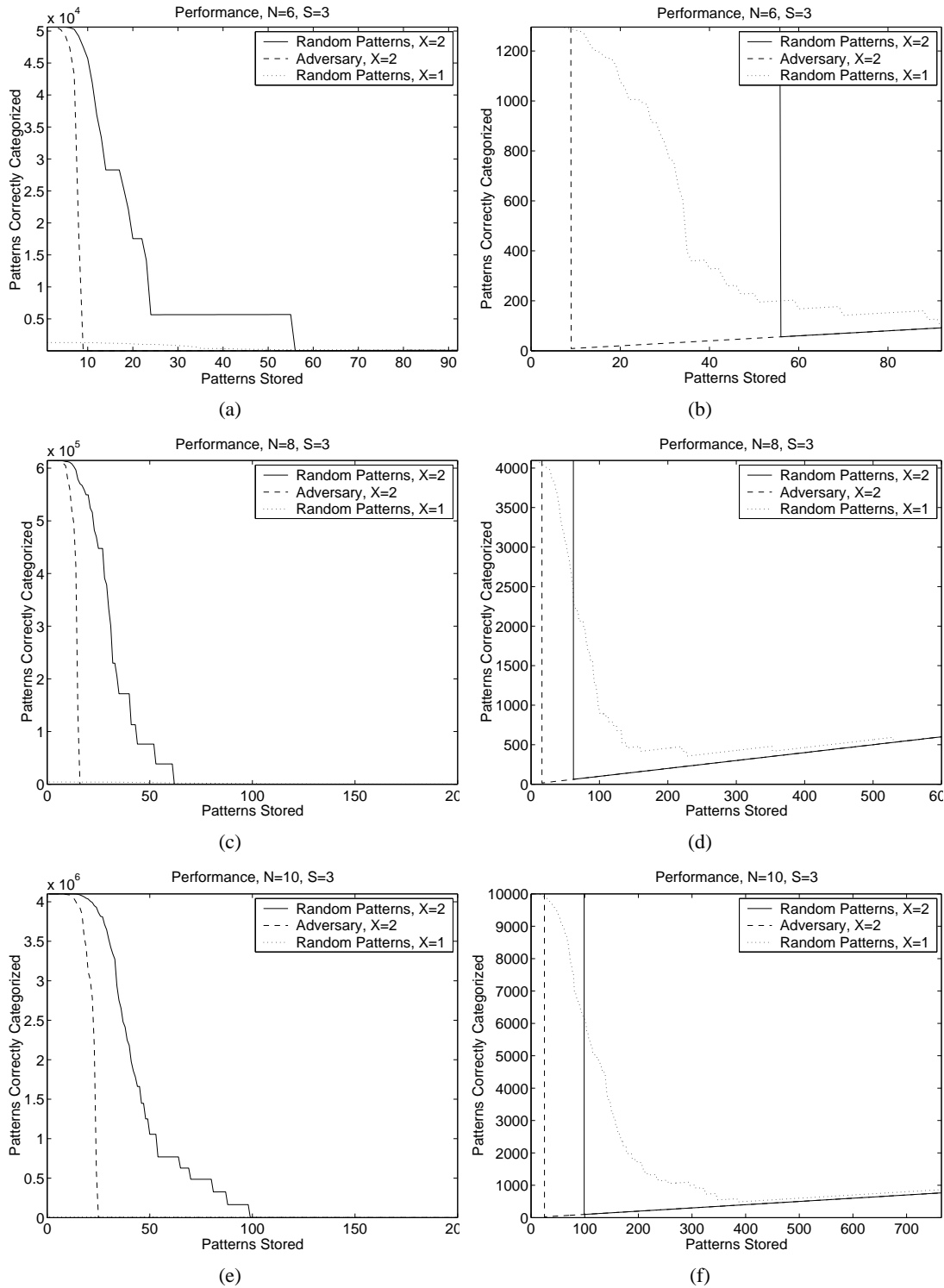


Figure 2: Results for $N \in \{6, 8, 10\}$, $X \in \{1, 2\}$, and $s_{hi} = 3$. The independent variable is the number of stored patterns, the dependent variable is the number of patterns of length $S + 1$ that are categorized correctly by the network (familiar if they are stored, novel if they are not). In the Adversary condition, patterns are selected to set all of the weights to 0 as quickly as possible (worst-case performance). In the Random Pattern conditions, patterns were randomly selected from a uniform distribution. The right plots magnify the bottom portions of the left plots, showing more clearly the difference between the $X = 1$ and $X = 2$ conditions. Note that each curve begins at the maximum accuracy, which is $\binom{N}{X}^{S+1}$. The initial difference between the $X = 1$ and $X = 2$ curves is attributable to there being fewer possible patterns in the $X = 1$ case.

the strength of a synapse is a local and not global variable, and so our design computes novelty in a more biologically realistic way.

We conclude with two ways to make the architecture more sophisticated and give intuitions as to how they will change the performance of the network.

We have shown that the upper bound on the capacity of the network is SN^2 . Randomly generated patterns do not reach this upper bound because it is rare that each pattern changes at most one weight. However, in natural applications, the encoding of a pattern in the network will be contingent upon whether or not it has been detected as novel. This would allow us to construct a preprocessor that orthogonalizes patterns in the network weight space *provided* that they have been detected as novel. Once novelty is detected, the goal of the preprocessor is to identify at least one unused three-tuple of k , j , and s and allocate it to the pattern when it is presented again for encoding. An ideal candidate area for such a preprocessor in the hippocampus is the dentate gyrus, which receives direct and indirect output from CA3 (communicating when a pattern has been categorized as novel) and is also upstream from CA3 (i.e., closer to the inputs of the hippocampus).

The primary cost of the network in implementation is the number of weights. It may be the case that some connections in the network can be removed without severely harming performance. In the present design, a removed connection is equivalent to setting the weight equal to 0. Thus, we would expect networks with sparse connectivity to have increased generalization and therefore a lower capacity. Alternatively, we might separate each connection into two connections, one that is excitatory and one that is inhibitory. If we fix inhibitory weights to -1 and allow excitatory weights to range from 0 to 1, the new architecture produces equivalent output to the current model but does so in a more biologically realistic way (inhibitory connections representing the output of interneurons in the hippocampus). Changing the architecture in this way also would allow us to remove the excitatory connections and leave the inhibitory connections intact.

References

- Bi, G. Q., and Poo, M. M. 1998. Activity-induced synaptic modifications in hippocampal culture: dependence on spike timing, synaptic strength, and cell type. *Journal of Neuroscience* 18:10464–10472.
- Bogacz, R.; Brown, M. W.; and Giraud-Carrier, C. 1999. *High capacity neural networks for familiarity discrimination*. Proceedings of the International Conference on Artificial Neural Networks. 773–778.
- Carpenter, G., and Grossberg, S. 1987. A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, graphics, and image processing* 37:54–115.
- Carpenter, G., and Grossberg, S. 1990. ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks* 3(23):129–152.
- Dehaene, S.; Changeux, J. P.; and Nadal, J. P. 1987. Neural networks that learn temporal sequences by selection. *Proceedings of the National Academy of Sciences* 84:2727–2731.
- Hopfield, J. J. 1982. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences* 79:2554–2558.
- James, W. 1890. *Principles of Psychology*. New York, NY: Henry Holt.
- Kali, S., and Dayan, P. 2000. *A familiarity-based learning procedure for the establishment of place fields in area CA3 of the rat hippocampus*. Proceedings of the CNS, Brugge, BE.
- Kohonen, T., and Oja, E. 1976. Fast adaptive formation of orthogonalizing filters and associative memory in recurrent networks of neuron-like elements. *Biological Cybernetics* 25:85–95.
- Marsland, S. 2001. *On-line novelty detection through self-organisation, with application to inspection robotics*. Ph.D. Dissertation, University of Manchester.
- McClelland, J. L.; McNaughton, B.; and O'Reilly, R. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102:419–457.
- Muruzábal, A. M. J. 1998. Self-organising maps for outlier detection. *Neurocomputing* 18:33–60.
- O'Keefe, J., and Nadel, L. 1978. *The Hippocampus as a Cognitive Map*. Oxford, UK: Clarendon Press.
- Ranganath, C., and Rainer, G. 2003. Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience* 4:193–202.
- Senn, W.; Schneider, M.; and Ruf, B. 2002. Activity-dependent development of axonal and dendritic delays, or, why synaptic transmission should be unreliable. *Neural Computation* 14:583–619.
- Sokolov, E. N. 1960. *Neuronal models and the orienting reflex*. New York, NY: Josiah Macy, Jr. Foundation.
- Swanson, L. W. 1978. *The anatomical organization of septo-hippocampal projections*. Amsterdam, NE: Elsevier. 25–43.
- Toulouse, G.; Dehaene, S.; and Changeux, J. P. 1986. Spin glass model of learning by selection. *Proceedings of the National Academy of Sciences* 83:1695–1698.
- Vinogradova, O. 2001. Hippocampus as a comparator: Role of the two input and two output systems of the hippocampus in selection and registration of information. *Hippocampus* 11:578–598.
- von Restorff, H. 1933. Über die Wirkung von Bereichsbildungen im Spurenfeld. *Psychologie Forschung* 18:242–299.
- Willshaw, D. J.; Buneman, O. P.; and Longuet-Higgins, H. C. 1969. Non-holographic associative memory. *Nature* 222:960–962.